
Professor Alain Kornhauser
Kyle Marocchini
Tyler Roth
Shirley Zhu

Max Bressler
Thomas Byrne
Chad Cowden
Artur Filipowicz
Keith Gladstone
Naman Jian
Kathryn Jones
Kate Ju
Ian Kinn
Alicia Lamb
Kevin Manyara
Ellie McDonald
Trevor Osborne

Bryan Osln
Ebo Quansah
Jordan Radke
Jack Rogers
Tyler Rudolph
Marlon Sabo
Aaron Schwartz
Ryan Slattery
Ryan Siiro
Alex Singleton
Daniel Toro
Joe Yates
Jessica Zou

autonomousTaxi for New Jersey, 3 passenger imbalance visualization

January 12th, 2016
Orf 467F15 Report 15-3
Intentionally Blank
# Table of Contents:

Introduction........................................................................................................3

Simple ATaxi Management (Fleet Size)..........................................................4

Camden, Gloucester, and Salem Counties.................................................16

Cape May, Cumberland, and Atlantic Counties....................................43

Burlington and Ocean Counties.................................................................76

Monmouth, Somerset and External Counties.......................................104

Middlesex County.......................................................................................176

Sussex, Warren, and Hunterdon Counties.............................................206

Essex and Morris Counties......................................................................239

Bergen and Passaic Counties...................................................................259

Union and Hudson Counties....................................................................291

Simple aTaxi Management Strategies..................................................310

Optimal aTaxi Management Strategies..................................................323

Stochastic aTaxi Management Strategies............................................346
Introduction

In our project, the ORFE 467 Class of 2015 was tasked with achieving the following:

1. From a basis of OTrip and NNTrip files, generate a set of theoretical “taxi” trips maximizing consumer utility and following three Level of Service Strategies – Common Destination 3, Max Circuity 20% and Departure Delay 300 Seconds.

2. From these generated trip files, generate a simplistic hypothetical fleet that could service consumer demands by dividing vehicles into capacity limits.

3. Analyze the unique constraints presented by each county and examine areas for improvement in the current fleet.

4. Develop strategies to deal with imbalances at the beginning and end of the day in counties (as discovered in 3.), test on select counties and examine their effectiveness.

The following pages document our class’s attempts to analyze these issues and pinpoint solutions. We acknowledge the tremendous help of Professor Kornhauser in helping guide us in completing this report, as well as working with us to identify and rectify errors in our analysis. Without your help, we would have not been able to complete this report.

This work represents our own work in accordance with University Honor Regulations.

The ORFE 467 Class of 2015
This page intentionally left blank
Simple aTaxi Management Strategies
Hana Ku and Shirley Zhu

Introduction
Our analysis of simple aTaxi management strategies is the first in a series of analyses with the common goal of determining the aTaxi fleet size-management strategy combination with the lowest cost while still servicing all the demand for trips across New Jersey and its neighboring external points. This report explores the minimum fleet size needed for the simplest aTaxi management strategy: an aTaxi travels during the day only when it is filled with passengers, and fleet-wide repositioning occurs once per day. We chose this repositioning to occur at midnight. We have developed a method for determining the minimum fleet size for this management strategy, visualized the need for fleet repositioning for system sustainability, and calculated cost metrics that can be compared against other simple and optimal fleet management strategies. Original data files can be found at: [http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F14/aTaxiDeparturesByCounty_3p-300-20/](http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F14/aTaxiDeparturesByCounty_3p-300-20/)
[http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F15/aTaxiArrivalsByCounty_3p-300-20_2015/](http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F15/aTaxiArrivalsByCounty_3p-300-20_2015/)

Minimum Fleet Sizing Method
We analyzed one day of hypothetical aTaxi trips (total of 11,481,856 trips) in NJ that were created based on US Census data, and other data sources that describe the types of people traveling in NJ and the number and types of trips they take. We assume trip data for all of NJ to be a closed system: the number of departures equals the number of arrivals. This implies the number of cars that exist in the system at the beginning of the day should all be accounted for at the end of the day. The only mechanism that can add cars to the system is the ‘super-source,’ which we view as the aTaxi dealership. Our goal is to find the minimum number of aTaxis needed at the beginning of the day - and where to put them throughout NJ - so that we do not need to add aTaxis from the supersource during the day. The supersource is limited to being a component of the minimum fleet sizing method, and is not a component that we expect to exist once “the system is in motion.”

The fundamental assumption of our fleet sizing model is that an aTaxi only travels during the day when it is in transit with passengers; there is no repositioning of an aTaxi during the day. An aTaxi will wait at the pixel’s central station until a passenger arrives at that station in need of a ride somewhere else, or the end of the day, whichever comes first. At the end of our algorithm, we return an array of pixel coordinates and the number of aTaxis at that pixel at the start of the day (00:00:00), and an array of pixel coordinates and the number of aTaxis at that pixel at the end of the day (23:59:59). The number of cars that we present as the minimum fleet size will satisfy all demand throughout the day without needing to add cars from the supersource. Since the fleet sizing method is applied to the simplest management strategy, the number we present serves as the upper bound for the minimum fleet size required of the system.

The Algorithm
For each 3, 6, 15, 50 passenger cars
- Extract relevant arrivals and departures statewide
- Condense arrivals and departures by minute
- For each pixel
View sequence of arrivals and departures in chronological order
All pixels start at 0 current supply. All pixels start at 0 beginning of day supply
If arrival occurs
   No change for pixel’s beginning of day supply
   Increment pixel’s current supply
If departure occurs when pixel’s current supply = 0
   Increment pixel’s beginning of day supply
   No change for pixel’s current supply
If departure occurs when pixel’s current supply > 0
   No change for pixel’s beginning of day supply
   Decrement pixel’s current supply

Cautions
Analyzing arrivals and departures in chronological order is crucial to obtaining the correct minimum fleet size. Simply calculating the difference between total number of arrivals and total number of departures for a pixel is not sufficient: minimum fleet sizing is path dependent. If we model the arrivals and departures of cars at a pixel as a random walk starting at 0, with each arrivals as +1, and each departure as -1, then the minimum number of cars needed at a pixel at the beginning of the day equals the absolute value of the most negative value of the random walk over the entire day. The number of cars at a pixel at the end of the day should equal the last value of the random walk plus the aforementioned number of cars needed at the beginning of the day.

Results
As described, our method outputs a matrix of NJ state pixels with the beginning of the day supply minimums, a matrix of NJ state pixels with the end of the day supplies. The sum across the entirety of either matrix equals the minimum number of aTaxis of that type for the entire state. We calculate the number of trips per car type to get a sense of the average number of times a car is reused without repositioning. We calculate the number of cars that are “out-of-position” at the end of the day, and need to be moved to the appropriate pixels so the system operate day over day without adding additional cars. Finally, we calculate an imbalance matrix, which is the beginning of the day supply matrix subtracted by the end of the day supply matrix. This yields a matrix of pixels of NJ with positive values indicating that the pixel needs more cars in order to satisfy the next day’s demand, and negative values indicating that the pixel has more cars than necessary to satisfy the next day’s demand. Again, we bring up the assumption that NJ is a closed system, so all the cars needed to satisfy demand of operating the system another day are somewhere in the state, but they are simply not in the correct pixels in the correct quantities throughout the state. We will revisit the implication of out of position cars when we calculate repositioning costs.

<table>
<thead>
<tr>
<th></th>
<th>3 passenger car</th>
<th>6 passenger car</th>
<th>15 passenger car</th>
<th>50 passenger car</th>
</tr>
</thead>
<tbody>
<tr>
<td># aTaxi trips</td>
<td>10,091,739</td>
<td>1,202,538</td>
<td>169,835</td>
<td>17,744</td>
</tr>
<tr>
<td>minimum # cars</td>
<td>1,548,368</td>
<td>391,787</td>
<td>93,031</td>
<td>13,443</td>
</tr>
<tr>
<td>avg trips per car</td>
<td>6.5</td>
<td>3.1</td>
<td>1.8</td>
<td>1.3</td>
</tr>
<tr>
<td># out of position cars</td>
<td>938,678</td>
<td>216,348</td>
<td>58,894</td>
<td>10,140</td>
</tr>
</tbody>
</table>

Figure 1: Number of Trips, cars needed, average trips per car, and cars that need to be repositioned in Early Morning Repositioning method
While 3 passenger cars might need the most absolute number of cars, we see that even without repositioning, cars can be used multiple times per day. This is not the case with 15 and 50 pass cars. The demand for these trips do not often start where another trip ended, so we need close to 1 car for every trip. The goal of other simple and optimal management strategies should include increasing the average trips per car per day statistic, while not compromising on costs for repositioning.

**Visualizations**

In order to visualize the spatial distribution of the out of position cars that need to be repositioned at midnight, as well as the distribution of the fleet at the beginning and end of the day, we created interactive maps to show these three values at each pixel. One map was generated for each type of car (3, 6, 15, 50 passengers) and each map had an overlay of the beginning of day fleet at each pixel, the end of day fleet at each pixel, and the imbalance, i.e. out of position cars, at each pixel. The scale of each type of map was created with the following percentiles of total cars: 0%, 25%, 50%, 75%, 90%, 95%, 100%. For the imbalances, the scale was created with these percentiles for positive and negative values.

Clicking on each pixel displays a popup with information about the particular pixel’s statistic. Red pixels, or negative imbalances, meant that cars had piled up there at the end of the day and needed to be moved out of that pixel. Blue pixels, or positive imbalances, meant that cars needed to be moved to that pixel to satisfy the demand. Again, since New Jersey is a closed system, the sum of all of the imbalances should equal zero and working under the assumption the trips for the second day are exactly the same, just moving these cars should allow the system to run fully for a second day. The visualizations can be viewed [here](#) and an example is shown in Figure 2.

![Figure 2: 3 passenger car visualization example](image)
Midnight Optimal Repositioning

The metric to compare fleet sizes and management strategies includes initial capital investment for fleet size and repositioning costs. We make the assumption that repositioning cost is determined solely on distance of empty car travel for repositioning.

**Linear Programming**

We explore two models for executing optimal repositioning. The first is writing a linear program that minimizes distance traveled constrained to satisfying pixels that need cars with cars from pixels with cars in excess. We attempt to solve this using CVX in Matlab.

\[
\begin{align*}
\text{min.} & \quad \sum_{i \in N} \sum_{j \in N} D_{i,j} \times M_{i,j} \\
\text{s.t.} & \quad M \geq 0 \\
& \quad \sum_{j \in N} M_{i,j} = P_i, \quad \forall i \in N \\
& \quad \sum_{i \in N} M_{i,j} = A_j, \quad \forall j \in N
\end{align*}
\]

Figure 3: Linear Program to solve repositioning problem

This linear program requires a scalar N as number of active pixels for a car type, P as a Nx1 production vector representing end of day supply at a pixel, A as a Nx1 attraction vector representing beginning of day demand, and D as a NxN symmetric matrix of distances between two pixels. The decision variable M is a NxN nonnegative elementwise matrix of number of cars to move from pixel in row i to pixel in column j, (i,j in N). Unfortunately this is too computationally intensive to run for all NJ trip data at once, even for just 50 passenger cars. CVX limits the number of variables and constraints to run at a time, and solving optimal positioning for 50 passenger cars already puts us above the variable maximum, at 25 million variables to optimize. It is possible to run this linear program for a single county at a time.

**Gravity Model**

The second attempt for optimal repositioning leads us back to the gravity model.

\[
T_{i,j} = \frac{A_j F_{i,j} K_{i,j}}{\sum_{x \in \text{zones}} A_x F_{ix} K_{ix}} P_i
\]

Figure 4: Formula for Gravity Model

P is a Nx1 production vector of end of day supply of cars, A is a Nx1 attraction vector of beginning of day demand, D is a NxN symmetric matrix of distances between two pixels, F is a NxN matrix of disutility of travel (which we equate to be 1/D_{i,j}^2), and K is omitted from this model.
### Reference Data from Fall 2014 files

<table>
<thead>
<tr>
<th></th>
<th>3 passenger car</th>
<th>6 passenger car</th>
<th>15 passenger car</th>
<th>50 passenger car</th>
</tr>
</thead>
<tbody>
<tr>
<td># aTaxi trips</td>
<td>8,659,171</td>
<td>1,526,478</td>
<td>257,149</td>
<td>36,584</td>
</tr>
<tr>
<td>total passenger miles (PM)</td>
<td>365,531,087</td>
<td>83,065,248</td>
<td>20,439,721</td>
<td>6,015,532</td>
</tr>
<tr>
<td>Total loaded vehicle miles (VM&lt;sub&gt;L&lt;/sub&gt;)</td>
<td>230,464,820</td>
<td>35,956,356</td>
<td>5,698,202</td>
<td>737,901</td>
</tr>
<tr>
<td>Average trip length miles</td>
<td>26.6</td>
<td>23.6</td>
<td>22.2</td>
<td>20.2</td>
</tr>
</tbody>
</table>

Figure 5: Table of summary statistics from Fall 2014. Access data here: [http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F14/aTaxiDeparturesByCounty_3p-300-20/](http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F14/aTaxiDeparturesByCounty_3p-300-20/)

### Our Findings

<table>
<thead>
<tr>
<th></th>
<th>3 passenger cars</th>
<th>6 passenger cars</th>
<th>15 passenger cars</th>
<th>50 passenger cars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total empty vehicle miles (VM&lt;sub&gt;E&lt;/sub&gt;)</td>
<td>20,211,573</td>
<td>4,497,714</td>
<td>967,450</td>
<td>108,470</td>
</tr>
<tr>
<td>Average empty miles per car</td>
<td>21.5</td>
<td>20.8</td>
<td>16.4</td>
<td>11.1</td>
</tr>
<tr>
<td>ERR = VM&lt;sub&gt;L&lt;/sub&gt;/VM&lt;sub&gt;E&lt;/sub&gt; (VM&lt;sub&gt;E&lt;/sub&gt;/VM&lt;sub&gt;L&lt;/sub&gt;)</td>
<td>11.4 (.056)</td>
<td>7.99 (.12)</td>
<td>5.9 (.16)</td>
<td>6.8 (.15)</td>
</tr>
<tr>
<td>AVO = PM/VM&lt;sub&gt;L&lt;/sub&gt;</td>
<td>1.6</td>
<td>2.3</td>
<td>3.6</td>
<td>8.2</td>
</tr>
<tr>
<td>AVO’ = PM/(VM&lt;sub&gt;L&lt;/sub&gt; + VM&lt;sub&gt;E&lt;/sub&gt;)</td>
<td>1.5</td>
<td>2.1</td>
<td>3.1</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Figure 6: Table of summary statistics from our analysis

Interestingly, from Figure 6, we see that the average empty vehicle miles per car is smaller for 50 passenger cars than for 15 passenger cars. This may be because the origins and destinations that require 50 passenger cars are all relatively concentrated spatially, so they do not have to travel as far to reposition. However, as we saw in the data in the last section, 50 passenger cars are not reused often throughout the day, so while the pixels are relatively concentrated spatially, supply and demand does not occur at the same exact pixel. Of course, the average vehicle occupancy is much higher for 50 passenger cars given that there are always more passengers in these vehicles. We also see that the Early Morning Repositioning cost (ERR), which we have defined to be vehicle miles empty / vehicle miles loaded, is higher for 50 passenger cars, even though the average empty miles per car is lower. This indicates that the 15 passenger cars are being reused more throughout the day, whereas the 50 passenger cars will not and will tend to make fewer trip, and this is also backed by data explored earlier.

For 6 passenger cars, the ERR is higher than both 15 passenger cars and 50 passenger cars. This tells us that 6 passenger cars are traveling far from where they started during the day and the spatial distribution of their use at the beginning and end of day is very different, hence, many of them need to be repositioned very far. The cost of repositioning drastically decreases for 3 passenger cars, even though the average empty miles per car is about the same as 6 passenger cars.
Since these strategies were implemented without enforcing an integer number of cars moved, this would serve as a lower bound to the actual empty vehicle miles required to implement the strategy. This is already the most optimal solution for this strategy and adding an integer constraint cannot decrease the empty vehicle miles.

We have calculated the absolute cost of repositioning cars statewide at only one time in a day, however it is not enough to compare these metrics against different empty car management strategies. Transportation is necessary to improving our time-space utility, so we want to maximize service of passengers and minimize costs at the same time. The AVO’ and ERR metrics scales the absolute cost incurred against the benefit to society. Ideal empty car management systems should strive to have high AVO’ and high ERR.

Cautions:
This model assumes instantaneous repositioning in one second before midnight from the end of day supplies to beginning of day demands. Obviously this is unrealistic. More intense optimization models can account for repositioning time as a function of distance to travel. This caveat further emphasises that the proposed cost for aTaxi management in this manner is a lower bound, or the minimum fleet size is actually larger, and the true cost of empty vehicle adjustments for day over day operation is likely higher with more real world constraints for operation.

Single County Analysis
Running these repositioning optimizations models for all the data in NJ is computationally intensive. Since the data has already been split for county by county, we can run each county individually, keeping in mind that analyzing data county by county leads to an open system. An open system implies the number of cars in the county at the end of the day does not necessarily have to equal the number of cars at the beginning of the day. A county can be deemed ‘an exporter’ of cars, when more cars leave the county than enter throughout the day, or ‘an importer’ of cars, where more cars enter the county than leave throughout the day.

Summary
Although the results obtained through this method can seem high, this simple aTaxi management system provides a baseline for smarter and more efficient aTaxi repositioning methods. This allows us to quantitatively measure how well a different repositioning algorithm will perform in terms of empty vehicle miles, average vehicle occupancy, and cost, which can determine if the additional effort to implement a more complicated system is worth it. In the next section, we can see a comparison to a more complicated repositioning system where repositioning occurs locally throughout the day as well as at midnight.
Camden, Gloucester, Salem and the Trip Generation Algorithm Explained

Final Report
ORF 467

Tyler Roth, Ian Kinn, Kyle Marochini

**Introduction:** This paper investigates the aTaxi ride sharing opportunities within the counties of Camden, Gloucester and Salem as well as explains the overall algorithm by which our class generated data for aTaxi trips. The purpose of this paper is to try to understand why the distribution of taxis trips looks the way it does in these counties through some statistical analysis of aTaxi trips that we have modeled in class in the last couple of weeks.

**Data Generation: Complexities and the Algorithm Explained**
In completing the overall ORF467 2015 aTaxi analysis, our group was tasked with the process of creating an algorithm that took generated files and assigned taxi trips to them. We were given NN Files, a set of linked lists...
that contained all of the information on the where a given person travels throughout the day (including pixel origin, destination, and time) and oTrip files that essentially split up every link in the NN files into specific trips organized by type (train, aTaxi, etc) and county. We relied almost exclusively on the oTrip files for our algorithm as we found that approach easier to deal with. Future groups may opt for using the NN trip files, as they contain more information and allow one to carry through more detailed analysis on trip miles and repositioning.

**Level of Service Strategies and Critiques for the Future**

To implement our strategy, we found it necessary to first examine several level of service strategies. The three we ended up using were Max Circuity, Common Destination and Departure Delay. Max Circuity essentially measures the maximum additional distance that any given person would be willing to accept to get to their destination. For example, under the assumption of a Max Circuitry of 50%, someone who would normally travel along a route of 1 mile would be assumed indifferent (or at least, willing) between travelling the old route of 1 mile and the alternative route of 1.5 miles. Common Destination measures the number of stops a person would be willing to accept to get to their destination. A Common Destination of 5, for example, would mean a person would be assumed indifferent (or at least, willing) to stop 5 times before arriving at their destination. Finally, Departure Delay measures the amount of time a person would be willing to wait from when they enter a car until the car itself departs. A departure time of 30 seconds, for example, would mean that a person would be at most willing to wait 30 seconds between entering the car and departing to their set of destinations.

Although Common Destination and Departure Delay seem to be quite relevant towards the direct utility of the rider, Max Circuitry seems to be more relevant towards the person paying for the maintenance and fuel of the aTaxi than the actual rider. A far more relevant indicator would be maximum ride time. We believe that the max circuitry may have been an actual approximation of this, but we see no reason for substituting an approximation when the real strategy was computable. One could easily compare current trip time for each person in a given route with their original trip times in the NN Files to see if it met the criterion or not in the route. This gives credence towards our earlier argument in including the NN Files in early analysis, as they allow for more accurate analysis of the scenario at hand.

**The Algorithm Explained**

In computing our data, we used a Departure Delay of 300 seconds, a Common Destination of 3 and a Max Circuitry of 20%. To start, we examined a given county and looked at departures within all pixels within that country to create the trip data for that country. Our algorithm proceeds as follows:

1. Select a pixel to examine and order every departure with respect to time from earliest to latest departure in a virtual queue.
2. Apply appropriate Departure Delay strategy and segment this queue into X (in our case, 300) second intervals and examine the riders within these segments.
3. Within this subset, group people into shared destinations, e.g., if two people were going to New York City, one person to Trenton, one person to Princeton, we would then create a set of 3 groups – New York City, Trenton and Princeton.
4. Within these groups, examine all combinations and apply appropriate level of service strategies to find the most optimal route.
5. After finding this most optimal route, remove the groups from the original list of groups and continue a new search for the next most optimal route with appropriate level of service strategies through all destination groupings.
6. If there are groups left over that cannot be satisfied in any possible way using level of service strategies, then these destination groupings must ride in their own route and are assigned a car uniquely for them.

7. Repeat through all pixels to assign every rider to a car on a route for your county.

This algorithm guarantees that every rider will be assigned to a car and nobody is left over. Its implementation (which our group did via MATLAB) can be done efficiently through the implementation of logical indexing, which significantly reduces the number of for loops used and speeds up computation greatly. A more direct approach is available with Python’s list structures.

Complexities in determining Most Optimal Path
The biggest concern in the algorithm is thus determining the most optimal path. A brute force approach works as follows:

1. Examine every possible combination of routes possible for your highest common destination parameter.
2. Select the most optimal route based off of this and remove it from the listing of groups.
3. Continue to apply this until there are no more possible routings for the highest possible parameter.
4. Continue to the (n-1)th value of Common Destination and apply the above until at the lowest value of common destination, at which point you assign appropriate routings for the leftovers as in the case of the general algorithm.

However, the issue with this is that the arrays examining the total number of permutations becomes truly massive when examining large time subsets. Generating this array, either piecewise or completely, is extremely slow, inefficient and generally impossible for most computers to handle (even using rank preserving transformations and unsigned 8bit integers, we ended up with arrays greater than 32gb in size). Thus, we aimed to instead examine permutations of no greater than 4! at a time (in general, (CD + 1)! comparisons) and assign routes based off of this. By adopting this simplification we were able to more efficiently and quickly solve routes, to such an extent that our algorithm could be easily applied to any real life aTaxi routing system.

However, in doing so, we remove the possibility that a possible combination including values not within our search further down the queue could work. As such, this simplification does not guarantee the most optimal routing. Although we believe that further approximations could be made towards the most optimal solution, there may indeed be no way of computing the most optimal solution. This problem at its core resembles an even harder version of the famous Travelling Salesman Problem. As travelling salesman can only be approximated, we believe that the next step can only be more accurate approximations.

County Overviews
Camden County, New Jersey
Camden County is the eighth-most populous county in New Jersey with a population size of 513,678 people spread out over 221.26 miles of land. This gives Camden a population density of 2,321.5 people per square mile. The largest city in Camden County is also named Camden, and it is home to 77,344 residents. The next largest cities are Cherry Hill and Gloucester Township. The following table describes the amount and types of each trip in Camden County:

<table>
<thead>
<tr>
<th>Intrapixel Trips</th>
<th>aTaxi Trips</th>
<th>Train Trips</th>
<th>Walking Trips</th>
<th>Total Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>22,199</td>
<td>1,979,902</td>
<td>83,709</td>
<td>112,523</td>
<td>2,198,333</td>
</tr>
</tbody>
</table>

From Fig. 2, it is clear to see that the vast majority of trips (90%) come from autonomous taxis. This probably has to do in large part with the nearby city of Philadelphia, which is very close to Camden’s Western boundary. Of the 83,709 train trips, a large portion of those are likely derived from the River Line, which travels up the Western Coast of New Jersey, and the New Jersey Transit line which runs from Pennsauken to Philadelphia 30th Street Station. As for the walking and intrapixel trips, Camden County’s Western boundary is located on the Delaware River, along which are many attractions, including the Adventure Aquarium, the waterfront, and the USS New Jersey, the most decorated American battleship in history. It is logical that people who travel to the waterfront for tourist attractions will walk up and down along the river or travel very short distances from one attraction to another. The waterfront also harbors many of Camden’s larger cities, including the city of Camden, which helps to account for the large number of walking trip

**Gloucester County, New Jersey**
Gloucester County is located just to the south of and adjacent to Camden County. It is the fourteenth-most populous county in New Jersey, with a population size of 288,288 people. Covering 322.01 square miles of land, Gloucester is less dense than Camden and has a population density of 895.3 people per square mile. The most populated city in Gloucester, Washington Township, is smaller than each of the three largest cities in Camden County and only contains 48,599 people. The next largest cities, Monroe Township and Deptford, contain 36,129 and 30,561 people respectively. Fig. 4 describes the amount and types of each trip in Gloucester County:

<table>
<thead>
<tr>
<th>Intrapixel Trips</th>
<th>aTaxi Trips</th>
<th>Train Trips</th>
<th>Walking Trips</th>
<th>Total Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>11,912</td>
<td>1,427,224</td>
<td>0</td>
<td>51,934</td>
<td>1,491,070</td>
</tr>
</tbody>
</table>

Again, an incredible amount of trips come from the transportation of autonomous taxis (96%). As Gloucester is just to the south of Camden, it is also situated close to Philadelphia, creating a large demand for taxis to transport people to Philadelphia or a train station that can take them to Philadelphia. Of particular interest is the fact that there are no train trips, which stems from the fact that there are no New Jersey Transit routes connecting Gloucester to other counties. There is, however, a proposed Glassboro-Camden Line which would connect Gloucester County to Camden County and the previously mentioned River Line along the coast.

Again, there are a reasonable number of walking and intrapixel trips which are generated primarily from tourist and pleasure-seeking activities along the Delaware River which are located in close proximity to one another, such as the Tinicum Rear Range Lighthouse, which has continued to light up the Delaware River since 1880.
Another popular attraction is Scotland Run Park, which consists of over 1,000 acres of land and promotes both walking and intrapixel trips to nearby destinations.

**Salem County, New Jersey**

![Salem County Map](image)

Salem County is located directly to the southwest of Gloucester County. Consisting of 66,083 individuals, it is the least-populous county in all of New Jersey. In addition, Salem contains 331.9 square miles of land, giving it a population density of 199.1 people per square mile. This also gives Salem County the lowest population density per square mile in New Jersey. The largest city in Salem, Pennsville Township, consists of only 13,409 people, and it is the only city in Salem which has greater than 10,000 people. The actual city of Salem contains just over 5,000 people. Fig. 6 describes the amount and types of each trip in Salem County:

<table>
<thead>
<tr>
<th>Intrapixel Trips</th>
<th>Taxi Trips</th>
<th>Train Trips</th>
<th>Walking Trips</th>
<th>Total Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,841</td>
<td>184,127</td>
<td>0</td>
<td>12,351</td>
<td>199,319</td>
</tr>
</tbody>
</table>

In a similar fashion to the previous two counties, the vast majority (92%) of trips comes from autonomous taxis. Salem is not as close to Philadelphia as Camden and Gloucester, but the lack of New Jersey Transit train stations connecting it to the rest of the state helps to explain the abundance of autonomous taxi trips. Bridges across the Delaware River and the New Jersey Turnpike, along with many other roads, also run through Salem County. Even more so than Camden and Gloucester, Salem runs along a large portion of the Delaware River. Many of its largest cities, including Pennsville and Salem reside along the coastline. The remaining intrapixel and walking trips can easily be explained by the attractions in close proximity to one another along the Delaware River and near the coast, including the Salem Oak Vineyards, the Salem County Historical Museum, and the Cowtown Rodeo.

**County Comparison**
Overall, Camden County has 707,623 more trips and 225,390 more people than Gloucester County. However, despite this, Gloucester County averages almost a full trip (0.89) more per person. This could be related to the fact that Gloucester covers a greater area of land but has less people. As a result, the population is less clustered and they have to travel more (and further) on average to see people or to perform everyday activities.

In contrast, Salem County has greater than a million fewer trips than Gloucester County and almost two million fewer trips than Camden County. This is easily explained by its extremely low population. However, an interesting result is that the average number of trips per person in Salem is also significantly lower than both Camden and Gloucester Counties. Another reason for this lack of trips in comparison to nearby counties may relate to the fact that Salem is significantly farther away from Philadelphia than both Camden and Gloucester, which are both relatively close and allow for easy transportation to and from Philadelphia.

**Gloucester:**

Fig 8 above represents the trip of aTaxis from each pixel in Gloucester County over the course of one day. The larger the circle the more aTaxi activity a pixel had over the course of that day. In calculating the activity at each pixel we just look at the sum of the total departure and arrivals in a given pixel over the time period.

<table>
<thead>
<tr>
<th>County</th>
<th>Population</th>
<th>Total Trips</th>
<th>Trips per Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camden</td>
<td>513,678</td>
<td>2,198,333</td>
<td>4.28</td>
</tr>
<tr>
<td>Gloucester</td>
<td>288,288</td>
<td>1,491,070</td>
<td>5.17</td>
</tr>
<tr>
<td>Salem</td>
<td>66,083</td>
<td>199,319</td>
<td>3.02</td>
</tr>
</tbody>
</table>
Fig. 9 represents the total number of people moved in over the course of the day in each county. So the size of each of these circles represents sum of the people who departed as opposed to the people who left during a given day.

In looking at these two graphs above there are couple things that would be useful to compare. The first thing that stands out is the large dark cluster of points that sit in the top center of both of these graphs. This group of points happens to fall right next to Philadelphia which makes sense because one would expect to have much greater populations in this areas which would usually lead to more travel.

In addition, I believe there is one more interesting things that you can take from these two graphs. The difference between these two displays is that one measures taxi activity and the other measures person activity and the interesting thing to note is that they look almost exactly the same. This points to the fact that AVO across all pixels in this county end up being very consistent or close to the same. This result is not necessarily trivial. One may think that areas of higher trip density would have higher AVO’s because there are more opportunities to match people together, but in comparing these graphs this does not look to be the case. No matter the amount of trips the AVO across all of these look to be very similar.
Camden:

Fig. 10 – Camden County aTaxi Activity – scaled to maximum

Fig. 10 represents that total aTaxi activity in each pixel over a days’ worth of time just as was used in the Gloucester discussion.

Fig. 11 – Camden County Person Activity – scaled to maximum

In looking to discuss similarities and differences between these two graph and the same ones from Gloucester there are two things that stand out as the same. One is that both the aTaxi activity and the person activity look to be very similar between the two graph and for this reason one should think that AVO is similar across all pixels independent of population of number of trips. Secondly, as you can see in both graphs the left side of the county has many more trips than the right side of the county due to the fact that Philadelphia falls just off the left side of the county.

Furthermore, there is one more interest facet to this Camden County that did not take place in Gloucester. This is that idea that there are over 86,000 train trips that go through go from Camden County each day. The interesting part of this is that one can almost see the two train routes that go through the county in
looking at the two graphs above. As one can see this is most a line or group of larger and darker dot that run along the top and bottom of the graphs. I believe these are the train routes and station for all of the people riding the train into the city for work each day and since each person has to drive to a stop and drive home from their stop there pixels end up getting a lot of activity during a given day.

Salem:

Fig 12 depicts the aTaxi vehicle activity in Salem county by pixel over a day’s worth of time.

Fig 13 depicts the person activity in Salem county by pixel over the course of a given day’s time.

As one can see there is one similar result to Gloucester and Camden that you can see in looking at the comparison of these two graphs. One again it looks to be that there is pretty consistent AVO across all of the
Salem because as you can see in the figures above the places of high person activity also have very high vehicle activity. Once again you can also see the large draw that live close to Philadelphia has in this part of NJ. Philadelphia sits just to northwest of Salem County and as you can see a large amount of the vehicle and person activity also falls to the north west. One of the other cools things you can tell from Salem county that one count not see from the Gloucester and Camden figure is where the other large towns in the county are. You can see Woodstown which is the group of about 7 bright blue pixels in the Northern middle of Salem county. In addition, as you look east you can see Elmer which is the group of three larger pixels falling on the eastern side of the county and one can see Salem itself if you look to the south west area of the figures where there are a couple brighter pixels. Looking at Fig 14 below might help one understand what I am explaining.

![Salem County Towns](image)

**Fig. 14 – Salem County Towns**

**Ride Sharing imbalances:**
The three graphs below represent ride sharing imbalances throughout time in Gloucester (Fig 15), Camden (Fig 16) and Salem counties (Fig 17) in a given day. They were calculated by using departures and arrivals to find the net number of vehicles in each pixel throughout the time of the day. Then we averaged the values of the largest 20 pixels in each county to find a single continues line as the ones shown below in each graph.

![Measure of Taxi Imbalance in Gloucester County](image)

**Fig. 15 – Averaged Taxi Imbalance for the Top Twenty Most Active Pixels in Gloucester County**
There are some very interesting and surprising things one can find from looking at the first two of these graphs. One is that both of them have very similar shape to them and this could be due to a couple of reason. One is that maybe this shape is standard among all counties in NJ and the second and what is believe is more realistic is the idea that people liking in these two counties live a very similar lifestyle. As you can see both of these graphs kind of tell a story. Earlier in the day at around 400 minutes in to the day both of these graphs are very low meaning there are more aTaxis leaving then entering which makes sense because many people who work in the city are leaving for work early each morning to get into the city.

In addition, one sees the most defining part of the graph which is two large spikes that fall right around 7AM to 8AM in each of the graphs. My explanation for these is that both Camden and Gloucester fall very close to Philadelphia and are industrial hub for manufacturers in the area. Because of this many labor-intensive industrial jobs are held in these two counties. The large spikes up in the imbalances graphs represent the large number of people coming to work each day in these job on a strict bell schedule. Also, I thought it was very surprising that after the two large spikes there was so much consistency in the two graphs. You do not see a sign of rush hour in the afternoon after seeing it in the morning. This may be because everyone coming back home that lives in these counties and everyone leaving the counties after work balances each other out. Though this is hard to believe it is only thing that explains this kind of consistency in the two graphs.
Though figures for Camden (Fig. 15) and Gloucester County (Fig. 16) look very similar it is very odd to me that Salem doesn’t look similar as well. Salem is in the same area of NJ in that it is close to Philadelphia and many people living in this county should be commutes to Philly which should make it look very similar to the figures for Camden and Gloucester but it is not at all. The one thing you can tell from the Salem imbalance figure is when the times of high volume trips are taken place. As you can see at the beginning and end of the day as one would expect there is relatively less volatility in the imbalance because there are very few trips being taken at these times of the data. Also, during morning and afternoon rush hour is when you see high volatility in Salem imbalances which is what you would expect to see. However, the great amount of volatility in imbalances amongst the top 20 pixels in Salem is in great contrast to the rest of the counties. I suspect it may have something to do with Salem’s relatively small size, in that population movements have a more pronounced effect (as opposed to Camden or Gloucester).

EMR:
In this section of the report we look at Early Morning Repositioning (EMR) of cars in ours 3 counties. The first graph for each county represents the demand for aTaxis of capacity 3 or less at the beginning of the day. The second represents where aTaxis will be left at the end of the day in each of the counties and the last figure represents the imbalance between the cars are needed in the morning and where they are left at the end of the day. This will be helpful because it help those who control the taxi services best and most efficiently be able to serve its clients. If there is demand in the morning in a certain area they will want to move many of their taxis to these areas overnight instead of leaving them where they are at the end of the night.

Camden:
Fig. 19 – Camden End of Day (EOD) Imbalance in EMR

Fig. 20 – Camden Combined Imbalance in EMR
Fig. 21 – Gloucester Beginning of Day (BOD) Imbalance in EMR

Fig. 22 – Gloucester End of Day (EOD) Imbalance in EMR
Fig. 23 – Gloucester Combined Imbalance in EMR

Fig. 24 – Salem Beginning of Day (BOD) Imbalance in EMR
Fig. 25 – Salem End of Day (EOD) Imbalance in EMR

Fig. 26 – Salem Combined Imbalance in EMR
There are a couple of interesting things to note when looking at these counties. One is that most of the taxis at the beginning and end of the day are closer to the Philadelphia area in each of these counties. In this area of NJ the big draw for work is Philly and this is why much of the imbalance activity happens closest to the Philly area in these counties. In addition there are some interesting things you can tell about the populations in these counties. Just by taking a look at the amount of color on each of the three imbalance figures for the counties you can tell how much smaller Salem County (Fig. 26) is in both population and trips. Another idea worth noting is where the red and blue fall for each county on their imbalance figures. In both Gloucester and Camden counties most of the bright red and blue areas fall for the most part pretty close to each other or at least in the same area of the county. This will lead quick and cheap reposition of vehicles overnight in both Camden and Gloucester counties. But, in Salem there is quite a different picture. A lot of the areas lacking vehicles in Salem county fall on the west side of the county whereas much of areas that have too many vehicle fall on the east side. This will result in much larger repositioning cost because the taxis will have to move further to reposition.
Cape May County
Population: 97,265
Area: Land - 251.42 sq mi
       Water - 368.99 sq mi
Number of Households: 40,812
Geography: Flat and coastal

Atlantic County
Population: 274,549
Area: Land - 555.70 sq mi
       Water - 116.12 sq mi
Number of Households: 102,847
Geography: Low-lying and flat
Figure 3 depicts the cumulative distribution function of all the trips based in Cape May County. There seems to be a steady, increasing trend beginning at 6 AM with a subtle, slight peak at around 9 AM - 10AM. It then flattens out and picks back up again around 5 PM. This correlates with people going to and from school or work. The total distance covered in a day by all trips sums up to around 510,000 miles.

Figure 4: CDF of Available Vehicles from size 6:15 in Cape May Count
Figure 5 depicts the net imbalance of vehicles for a fleet size of 6 - 15 passengers by the time of day. At the beginning of the day there is a net imbalance, which lasts until about 8 AM. This could be due to all the departures for work and school, which create a net imbalance of -2500. The net imbalance is reduced substantially and brought back down to 0 at around 9. Steady fluctuations continue for another 7 hours until there is another huge deficit between departures and arrivals beginning at 3 PM and reaching a peak imbalance of -1100 around 5 PM. This can be explained by the departures to get back home from work or school. By the end of the day the imbalance is restored to 0.

Figure 6: Available Vehicles in Cape May Count of size less than 3 by time of day
For vehicles with a fleet size less than 3, there is a huge availability of vehicles. These vehicles could be available due to possibly taxis dropping off people at work and returning back to their original position. It could also be people arriving home from work that originally took a bus with many people and then took a taxi to get back home.

![Available Vehicles of Passenger Size 3:6](image)

Figure 7: Available Vehicles of Size 3:6 by time of day in Cape May County

Figure 7 depicts the net imbalances of vehicles by time of day for fleet size of 3 - 6 passengers. There seems to be an income of 1000 cars at the beginning of the day until about 7 AM. At about 7 AM the net imbalance reaches about -3900. The net imbalance is restored to 0 around 9 AM. However, after 9 AM there is a huge deficit between departures and arrivals that continues until about 9 PM. The net imbalance reaches a maximum of about -6200 and by the end of the day is reduced to about -3000. At the beginning of the next day there will be a net imbalance of -2000 since there is 1000 arrivals at the beginning of each day.
For fleet sizes greater than 15, the maximum net imbalance is about -5000 and it is experienced around 9 AM. It is eventually restored to a number that fluctuates between -1000 and -1500. It never reaches equilibrium and this will also be a problem at the beginning of the next day. Also there is no additional flow of extra arrivals in the morning so aTaxis will have to be pulled from another source.

Figure 8: Available Vehicles in Cape May County of size >15 by time of day
Figure 9: CDF of available Vehicles of size >15
Figure 10 is a heat map depicting the net balance of arriving and departing vehicles in Cape May from about 6 AM to rush hour which is around 5 PM. A yellow color indicates “arrived” vehicles or “net-positive” and a dark blue color indicate “departed” vehicles or “net-negative”. This heat map aligns with the graphs we generated previously as we can see bright yellow and orange pixels indicating the arriving vehicles in the early morning that create a net positive. It can also be observed that at the peak of rush hour there is a significant amount of departures and almost the entire heat map is a blue color. There is a huge net-negative imbalance and this could be due to the fact that more people are driving in from bordering counties to work at Cape May. Vehicles would need to be brought in from bordering counties that have excess vehicles or a super source.

Atlantic County
Figure 11: Atlantic County Trip CDF

This CDF is very similar to the CDF of Cape May in terms of the slope. However the total distance of the trips is roughly 10 time greater than the sum of Cape May trips. This makes sense since there is a greater population in Atlantic County. At rush hour, the slope of the CDF increases significantly.
Figure 12: Atlantic County available vehicles of size <3 by time of day

Figure 12 depicts the net imbalance of vehicles for a fleet size that is less than 3. At the beginning of the day up until about 9 AM, there is heavy income of arrivals that totals to a net positive of 20,000. However, after 9 AM there is a significant amount of departures that cause a maximum net-negative of -30,000. At the end of the day there is a net imbalance of -20000 but this will be restored to 0 by the beginning of the next day due to all the arrivals that come flooding in.
The net imbalance of vehicles with a fleet size of 3-6 can be observed above. It can be seen that there are greater departures than arrivals in the morning at 9 and at 5 PM. By the end of the day there is a net-positive of about 2500 vehicles. This information is very useful and we can use these excess vehicles to supply other pixels that have a net-negative imbalance.
Figure 15: Atlantic County available vehicles of size 3:6
The net imbalance for vehicles with a fleet size of 6 - 15 passengers can be seen above. There is a net-negative imbalance of -800 at around 9 AM. After 9 AM, the net imbalance is restored to 0 by 10 AM and continues to fluctuate, ultimately converging to a value of about 200 by the end of the day. This could reflect buses that take people to work that once they drop off their cargo they return to the original position.

Figure 16: Atlantic County size 6:15 available vehicles by time of day
Figure 17: Atlantic County size 6:15 available vehicles

Figure 18: Atlantic County size >15 available vehicles by time of day
This graph depicts the net imbalance of vehicles with a fleet size greater than 15 passengers. Already, it can be observed that the number of trips made by such a fleet size is significantly less than all other fleet sizes. This means that there are not many people taking the train or buses to work/school. At 9 AM, a maximum net-negative imbalance of -150 is experienced. By 10 AM, the net imbalance is restored to 0 and stays there until the end of the day.

Figure 19: Atlantic County size >15 available vehicles

This section analyzes the AVO levels in both Atlantic and Cape May counties. The AVO’s are representative of the highest activity pixels in both counties, and are graphed in terms of increasing DD and CD.

- ATL person miles = 22,539,342
- CAP person miles = 5,244,365
- MER person miles = 13,530,118
- CUM person miles = 11,535,263
In the Atlantic p2p study, it seems that increasing DD has very little to zero effect on AVO. Conversely, with increasing CD we observe a dramatic increase in AVO, from CD = 1 having an AVO = 1, to CD = 4 approaching AVO = 3. With increasing CD, AVO is improved 3x.

These results mirror the findings in the AVO report. Increasing DD has negligible effect, whereas increasing CD dramatically reduces vehicle miles. The improvement is of the magnitude of 80,000 vehicle miles.
In the Atlantic p2Sp study, it seems that increasing DD also has very little to zero effect on AVO. With increasing CD we observe a dramatic increase in AVO, from CD = 1 having an AVO = 1, to CD = 4 approaching AVO = 3. With increasing CD, AVO is improved 3x.

These results are essentially identical to the findings in the ATL p2Sp AVO report. Increasing DD has negligible effect, whereas increasing CD dramatically reduces vehicle miles. The improvement is of the magnitude of 80,000 vehicle miles.
In the Cape May p2p study, it seems that increasing DD has very little to zero effect on AVO, just like with the Atlantic county AVO analysis. By analyzing increasing CD, we note a dramatic increase in AVO even greater than in Atlantic County, from CD = 1 having an AVO = 1, to CD = 4 approaching AVO = 3.5. With increasing CD, AVO is improved 3.5x, which is a .5X aggregate increase in AVO compared to Atlantic County levels.

These results mirror the findings in the AVO report. Increasing DD has negligible effect, whereas increasing CD dramatically reduces vehicle miles. The improvement is of the magnitude of 250,000 vehicle miles.
In the Cape May p2Sp study, it seems that increasing DD also has very little to zero effect on AVO. With increasing CD we observe a dramatic increase in AVO, from CD = 1 having an AVO = 1, to CD = 4 approaching AVO = 4. With increasing CD, AVO is improved 4x.

These results are essentially identical to the findings in the Cape May p2Sp AVO report. Increasing DD has negligible effect, whereas increasing CD dramatically reduces vehicle miles. The improvement is of the magnitude of 280,000 vehicle miles.
NOTE: X axis is displayed in seconds.
(20,000 seconds corresponds to 5:30am, 40,000 corresponds to 11:00am, 60,00 to 4:30 pm. 80,000 to 10:00pm)
Y axis is displayed in number of trips.

Figure 28: Atlantic County trip departure times
Figure 28 depicts the frequency of departure times in Atlantic County. Intuition would suggest that people make relatively few trips between the hours of midnight to 6:00am. Furthermore, intuition would suggest that most people make trips when making their commute to work, and their commute after work. This graph details how a relatively large amount people make a trip starting at around 7am - 9am, which suggests a large morning commute to work. There is also a peak from around 4:30pm - 8 pm. This could be indicative of both a commute back home and/or to other locations like restaurants and shops.

Figure 29 depicts the frequency of departure times in Cape May County. Cape May departure times appear to follow the Atlantic county times, with a morning commute at 7-9am, and an after work commute at 4:30pm. However, there is a third peak just around noon. This could indicate a rush to lunch that is not as clearly defined in Atlantic County.
Figure 30 depicts the distance of trips in Atlantic county. The largest percentage of trips fall between the 0-5 mile range of distance. However, the relative amount of trips of distance 5-65 miles are fairly uniformly distributed between 40,000 - 60,000 trips. It seems strange a priori that the amount of people traveling from 5-10 miles are less than the amount of people traveling 50-55 miles, however the data is clear about this fact.
Figure 31: Cape May County trip distances

Figure 31 depicts the amount of trips originating in Cape May county, arranged by distance. A statistically significant fact of this graph is that a lot fewer trips are made compared to Atlantic county. No bar in between distance intervals reaches much above the 30,000 trip mark, while almost all of the bars in Atlantic county are larger than 30,000 trips. We also observe that there are two “peaks” of frequency, one occurring in the 0-5 mile interval, and one occurring in the 35-45 mile intervals. This could be the result of Cape May being an isolated county, where interesting places could either only be very close or very far away, although more research is necessary before any definitive conclusions could be made.

Figure 32: Atlantic County trip source map
Figure 32 is a plot of the ox and oy pixel locations of all trips originating in Atlantic county. Clearly, the trips form a rough outline of the county, yet pockets in the center exist where trips are not being made.

Figure 33: Atlantic County map
A similar graph of ox and oy pixels is made for Cape May county, and we can see that the trips also roughly fill out an image of the county. Like Atlantic county, we also see several gaps where no trips are originated at all. One particularly large gap is a large swath of land that runs parallel to the atlantic coastline yet is some distance away from the shore. Everywhere else seems to be the origin of a significant amount of trips.
Figure 36 represents the destination pixels of trips that originate from Atlantic county. By inspection it appears that trips from Atlantic County travel mostly to locations in south jersey and the coast, yet make hardly any trips to south-central jersey of North Jersey.

![Figure 36: Atlantic County trip destination map](image)

Figure 37 represents the destination pixels of trips that originate from Cape May county. There are clearly much fewer destination pixels as Cape May hosts dramatically fewer origin trips than Atlantic county. By inspection it also appears that trips from Cape May County travel mostly to locations in south jersey and the coast, yet like Atlantic County Cape May trip makers make hardly any trips to south-central jersey of North Jersey.

![Figure 37: Cape May County trip destination map](image)
Ocean and Burlington Counties are two counties in central New Jersey. Ocean County has a population of 586,301 and an area of 916 square miles. Its beaches are extremely popular among New Jersey residents, and the coastal city of Seaside Heights is the location of the popular reality television show “Jersey Shore”. Burlington County is the largest county in NJ by area, with an area of 827 square miles. It has a population of 448,734, and is the second-largest cranberry producing county in the United States.

In this chapter, we will be performing various analyses of the daily trips in both Ocean and Burlington counties, using data generated by past sections of ORF 467: Transportation Systems Analysis. As expected, Burlington and Ocean counties have a different distribution of daily trips:
Fig. 1 – Burlington and Ocean Trip Overviews

Interestingly, Burlington County has more daily trips, even though it has a lower population.

### ANALYSIS OF TOTAL VEHICLE MILES AND AVERAGE VEHICLE OCCUPANCY (AVO)

<table>
<thead>
<tr>
<th>County</th>
<th>aTaxi</th>
<th>Intra-pixel</th>
<th>Train</th>
<th>Walking</th>
<th>Total Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burlington</td>
<td>1,416,580</td>
<td>11,023</td>
<td>0</td>
<td>61,139</td>
<td>1,488,742</td>
</tr>
<tr>
<td>Ocean</td>
<td>1,262,382</td>
<td>10,969</td>
<td>15,820</td>
<td>63,967</td>
<td>1,343,138</td>
</tr>
</tbody>
</table>

Fig. 2 – Burlington County Pixel to Pixel Vehicle Mile for various LoS
Fig. 2 and Fig. 3 represent the total vehicle miles and average vehicle occupancy in Burlington County with varying Departure Delay (DD) times in seconds and Circuitry Distance (CD). As expected, the total vehicle miles decrease and the AVO increases as we increase the CD. We see a range of AVO of about 1.5 for CD = 1 to 3.5 for CD = 3. However, the values for total vehicle miles seem unrealistically large, and may be orders of magnitude off from the true number of vehicle miles. What is unexpected is that the DD seems to have little to no impact on our AVO and total vehicle miles. This is surprising, and is more likely a result flaw in our method of computation than being an accurate reflection of the distribution of trips in Burlington County based on differing Departure Delay.
The AVO and total vehicle miles in Ocean County are again consistent with our expectations with increasing AVO and decreasing number of vehicle miles with increasing CD, but again inconsistent with our expectations based on the minimal to no impact of DD on our results. Again, this is likely due to an error in our computation script, and may not accurately reflect the true impact of increased DD in Ocean County. Similarly, while our estimates for AVO seem reasonable, ranging form 1.1 for CD = 1 to 2.5 for CD = 4, our estimates for total vehicle miles seem orders of magnitude too high.
INTRA VS EXTRA COUNTY TRIPS

As visualized in the pie chart above, inter-county trips, which refer to trips between Ocean and other counties account for more than 60% of the total trips, totaling at 79%. This result is limited to only aTaxi trips, and doesn’t account for other modes of transport such as rail, or walk. Consequently, Intra-county taxi trips (trips within ocean county), account for 21%.

Fig. 6 – Ocean County Trip Proportionalities

Burlington County, as by the pie chart above has 87% inter county aTaxi trips, and only 13% intra county trips.

Fig. 7 – Burlington County Trip Proportionalities
ARRIVALS & DEPARTURES

Departures

The following histograms show departures and arrivals by time for Burlington and Ocean counties.

As detailed in the above histograms, we can see a large spike at around 9pm, which signals a large morning rush which later distributes in the evening to signal trips towards home or recreational centers. Additionally, we see an increase in departures immediately before and after five, which represents the commute
home from work, and addition post-work trip departures. Both Burlington and Ocean counties display relatively similar aTaxi departure distribution over the course of a day.

**Arrivals**

![Histogram of Arrivals by Time in Burlington County](image1)

Fig. 10 – Burlington County Arrivals Histogram by Time

![Histogram of Arrivals by Time in Ocean County](image2)

Fig. 11 – Burlington County Arrivals Histogram by Time

The pattern of arrivals mirrors the pattern of departures – we see a large number of arrivals centered around 9:00 A.M., which represents arrivals to work and school, and a large number of arrivals from 5:00 P.M. to 8:00 P.M., which represents arrivals home from work, and arrivals at destinations from post-work trips.

**ANALYSIS OF TAXI IMBALANCES**
The following graphs detail daily aTaxi imbalances in both Burlington and Ocean counties. The graphs have been indexed on the x-axis by hour of day, summing up to 24hrs. The columns of the graphs therefore represent the aTaxi imbalance per hour of day, and the height of the column represents the level of imbalance (either negative or positive).

The graphs are divided into:
- Cumulative net imbalance of arrivals
- Net imbalance density of arrivals

For both graphs, a negative value implies shortage of aTaxis (more departures than arrivals), and a positive value implies an excess, or rather surplus of aTaxis at a given time. Both graphs take into account all vehicle sizes.

**Cumulative Net Imbalances of Arrivals**

![Cumulative Net Imbalance of Arrivals - Departures in Burlington County](image)

**Fig. 12 – Burlington County Departure-Arrival Imbalance**
In both counties, we see that for aTaxi arrivals, daily cumulative imbalances start at approximately 6am. This makes sense as we expect high demand of aTaxis from people moving from their homes to work in the morning. As such, we have more cumulative departures and fewer cumulative arrivals for the rest of the day, creating an imbalance over the course of the day. This imbalance stresses the need for a repositioning model.

**Net Imbalance Density of Arrivals**
Again, as we expect, these graphs mirror a normal daily schedule of a work day, with net departures taking place in the morning, net arrivals taking place around 9:00 A.M., a large rush of departures during lunch, departures from work around 5:00 P.M, and net arrivals afterwards as people arrive back at their homes. We see
this overall pattern in both Ocean and Burlington counties, though the magnitude of net imbalance of arrivals and departures at each time is greater in Ocean County, likely due to its greater populations.

**CUMULATIVE DISTRIBUTION FUNCTIONS OF aTaxi TRIP LENGTHS**

Fig. 16 shows the cumulative distribution function of aTaxi trip lengths in Burlington County. We can see that 50% of trips throughout the day are less than 15 miles in length. According to the Federal Highway Administration, the average automobile commute is 12.6 miles, so it would appear as though the many daily commutes in Burlington County are greater than this national average. Interestingly, there seems to be a dip in the slope of the CDF around 30 miles, and an increase again around 40 miles which indicates that there a greater number of trips around 40 miles in length. About 1 in 5 trips are below 5 miles. As expected, the slope of the CDF decreases as the number of miles increases. It seems as thought the longest trip is in the range of 60 miles.
The cumulative distribution of aTaxi trips in Ocean County is similar to that of Burlington County. Here, about 50% of trips are below 17 miles, and there are a few trips greater than 60 miles in length. Unlike the Burlington County CDF, there is only the slightest decrease in slope at around 35 miles. Here, about one in five trips are below 3 miles.
As expected, the cumulative distribution of a Taxi trip durations mirrors the distribution of trip lengths, since trip length is predominately a function of distance (other factors may include road choice, congestion, etc.). There are no trips greater than two hours in length, and about 50% of trips are less than 30 minutes in lengths. This is longer than would be expected – the commute data provided by the Federal Highway Authority claims that the average automobile commute is 22.8 minutes. Over 70% trips are less than one hour.
The cumulative distribution function of a Taxi trip duration in Ocean County, again, follows the trip length distribution very closely. In Ocean County, about 50% of trips are below 35 minutes in length, and there are a couple of trips greater than two hours in duration.

**CUMULATIVE DISTRIBUTION FUNCTIONS FOR END OF DAY ATA XI IMBALANCE**
Here, we examine the cumulative net end of day taxi imbalance in Burlington County. As we can see, the large majority of pixels have an imbalance between -100 and 100 vehicles. With a greater number of pixels having a large taxi surplus at the end of the day (+100 taxis at the end of the day) than a large taxi deficit (-100 taxis at the end of the day). As expected, this distribution has a mean of approximately zero, and most of the pixels have a small taxi imbalance.
The cumulative net end of day taxi imbalance in Ocean County is very similar to that of Burlington County. One of the most noticeable differences is that there are more pixels with a large vehicle deficit at the end of the day (-100 vehicles) than in Burlington County.

**DISTANCE TRIPS PER VEHICLE SIZE BY TIME OF DAY**

In order to calculate the distance travelled by the different vehicle sizes (3,6,15,50), we created an algorithm that implemented the following steps:

- Read in data from the County aTaxi departures csv file
- Divide the huge file into the various subsets of vehicle size
- Compute distance such that,
  1. For each entry (i), subtract the departure time(i) from the arrival time(i)
  2. Convert the time difference into minutes
  3. Multiply the difference by 0.5 miles per minute (30mph, constant speed)

However, this algorithm proved to be very slow for vehicle occupancy of 3, where we have more than a million trips per day. We however, have the data for vehicle occupancy 15&50.

The two CDF’s illustrated below show the cumulative distance for vehicles 15, 50 in Burlington county.
For max occupancy = 15, we see that about 50% of the Taxi Trips distances are for the day are less than 20 miles. For max occupancy = 50, about 50% of the trip distances are less than 18 miles.
In Ocean county, the trend seems nearly similar to Burlington for taxi with max occupancy = 50, with 50% of trip distances per day being less than 20 miles.

For Max occupancy = 15, the trend is also similar to Burlington, with 50% of the daily aTaxi distance travelled being less than about 22 miles.

**VISUALIZING ATAXI SURPLUS TIMES**
The time of day that a pixel begins accrues a taxi surplus that lasts until the end of the day will be of great help in empty vehicle repositioning, as this marks the time in which taxis from this pixel can be repositioned to other pixels in need of aTaxis. We wrote a script which determines and visualizes the time of day that this occurs. If a pixel does not have a surplus of taxis that last until the end of the day, then the script sets the time of this pixel to 1440 minutes, or the end of the day. If a pixel has a vehicle surplus time at the very beginning of the day, then the pixel’s time value is set to 0 minutes. Due to limitations in the visualization software, pixels that are outside of the county were also mapped to a “0” values. Below are the results of this analysis and visualization for Burlington County:

![Visualization of taxi surplus for Burlington County](image)

**Fig. 26 – Burlington County Visualization for Capacity**

We can see that there are very few pixels that have end of day aTaxi surpluses for three passenger vehicles. We can also see a group of pixels on the middle-left edge of the county with an early/ mid-day taxi surplus. The six passenger vehicles seem to have the greatest number of pixels with a taxi surplus out of the four different sizes of taxis. For 15 passenger vehicles, there seem to be a number of pixels which have taxi surpluses before the end of the day, but many pixels which do not require any 15 person taxis. There are few pixels that require 50+ passenger vehicles, and the image produced is a subset of the top half of the county.
We see similar trends with daily taxi surplus times in Ocean County than we say in Burlington County. There are few pixels with end of day taxi surpluses in for 3 passenger vehicles, but many pixels with 6 passenger vehicle surpluses sometime during the day, many of which are towards the end of the day. Of the pixels that require 15 passenger vehicles, many of then have a surplus of taxis y the middle of the day, with few pixels arriving at a surplus later in the day. While only a modest subset of pixels in Ocean County require 50 passenger vehicles, it seems as though a third of these pixels arrive a vehicle surplus that lasts until the end of the day.

VISUALIZING ATAXI ARRIVAL AND DEPARTURE TIMES

We developed a script in order to visualize arrival and departure times in Ocean and Burlington Counties. These videos will be uploaded with the submission of this document. The analysis performed on the daily arrival/departure histograms and CDFs above holds for each video, in which one can visualize the daily morning and evening commutes, afternoon trips, and activity predominately in the daytime. Each video is 16 seconds long, with each second representing 20 minutes later in the day.

ADKNOWLEDGEMENTS
We would like to thank Professor Kornhauser for providing us the relevant background and understanding of aTaxi systems in order perform appropriate analysis, as well as developing the datasets that were used. We would like to acknowledge our collaboration with Keith Gladstone, Ellie McDonald, Tom Byrne, Alexander Singleton, Trevor Osborne, and Joe Yates, all of whom worked with us to write code and share ideas.

We would like to acknowledge the use of data from the U.S. Census Bureau for county populations and sizes, the Federal Highway Bureau for information about average commute times and distances (https://www.fhwa.dot.gov/policy/2010cpr/execsum.cfm) and information from the Burlington County Library System for facts about Burlington County (http://www.bcls.lib.nj.us/about-burlington-county).
New Jersey County Analysis Final Report:
“Central Jersey Suburbs”

Monmouth County,
Somerset County,
and External Locations

ORF 467: Transportation Systems Analysis

By: Ellie McDonald ’17
    Joe Yates ’17
    Tom Byrne ’17
Preface: Compiling Data

Before any cumulative county data could be analyzed, the individual county arrival and departure files needed to be compiled and reorganized. The county arrival files originally included the following fields: county of origin, x and y pixel coordinates for said county, the time of arrival, the occupancy at dispatch, and the county of destination. The county departure files had the same layout. These files were compiled for the following level of service:

- DD (delayed departure) = 500
- CD (common destinations) = 3
- MaxCircuity (how far out of the way a vehicle is willing to travel, relative to the distance associated with each passenger’s individual trip) = 0.2

From these county arrival and departure files, six types of files were created:

- Non-Cumulative Arrivals
- Non-Cumulative Departures
- Non-Cumulative Net Taxis
- Cumulative Arrivals
- Cumulative Departures
- Cumulative Net Taxis

The Non-Cumulative Arrival files took all of the county arrival files and created one document that recorded the arrivals in every pixel in every county for every minute of the day. The pixels without arrivals were not included in the file. It was found that 21,273 pixels had arrivals throughout the 24 hour day analyzed. The columns of this file were: x-pixel coordinate, y-pixel coordinate, county code, and then a column for every minute of the day. Within each minute, a one was added to the cell for every arrival in that pixel in that minute. The number in each column represented how many arrivals existed in that minute alone. The Non-Cumulative Departures file had the same set up but instead of recording a +1 for each arrival, it recorded a -1 for each departure from a pixel. Again the columns were as follows: x-pixel coordinate, y-pixel coordinate, and a column for each minute of the day. The Non-Cumulative Net Taxis file combined the Non-Cumulative Arrival and Non-Cumulative Departure files. For each minute the Non-Cumulative Net Taxis file added the number of number and the negative number of departures per minute. So for each minute column, the cell had the total number of taxis that left/arrived in said minute. If this number was positive there were more arrivals than departures during that minute; if this number was negative there were more departures than arrivals during that minute. The Cumulative Arrival, Departure, and Net-Taxi files took the stated Non-Cumulative files and made the arrivals, departures, and net number of taxis and made them cumulative. This means that if one taxi arrived in the first minute and another arrived in the second minute, then the second minute column of Cumulative Arrivals would have a two to represent that two taxis had arrived. For the Cumulative Net Taxis file, the number in each cell for each minute for each pixel represented the number of taxis available in that pixel at the minute. If the number was negative, there was more demand for taxis than there were actual taxis in the pixel at the minute.

The final step of organizing data involved thinning the Cumulative Arrival, Cumulative Departure, and Cumulative Net Taxis files by the number of occupants in the taxi. There were four categories to distinguish
size of the taxi. These categories were as follows: 1-3 passengers, 4-6 passengers, 7-15 passengers, 16+ passengers. These different groups represent sedan taxis, larger minivan size taxis, Small van taxis, and buses. The columns of these files were identical to the columns for the Cumulative Arrival, Departure, and Net Taxis files. These vehicle size files were important for two reasons. First, you cannot pick up 15 people with a sedan size taxi and it would not be space or fuel efficient to pick up 3 people in a bus. Consequently, we could look at arrivals and departures for such groups respectively and figure out the proper fleet size. Second, by dividing the arrivals, departures, and net taxis into these categories, the analysis could better spot aTaxi sharing trends for different sectors of the market because the different groups followed different patterns throughout the day.
Monmouth and Somerset Counties

By: Ellie McDonald, Joe Yates, and Tom Byrne
Distribution of Trip Times and Trip Lengths:

The cumulative distributions of taxi trip times in Monmouth County and Somerset County were pulled from the oTrip files, specifically by taking the difference between the dTime and oTime fields. We see that the cumulative distributions of trip times for both Monmouth and Somerset (Figures MON.1 and SOM.1) can be generalized as concave curves, affirming the intuitive notion that taxi passengers have a preference for relatively shorter trips. That is to say, passengers will generally set their attractions (e.g. work, school, recreation) to be closer to their homes or their current locations.

Figure MON.1: Monmouth County trips CDF
A speed constant was used to transform the taxi trip times into taxi trip lengths, shown in Figures MON.2 and SOM.2 on the following page. Consequently, the taxi trip lengths are distributed identically in comparison to the taxi trip times and will thus fail to provide new or groundbreaking insight into the data. That being said, looking at the distribution of trip lengths as opposed to trip times can offer us a different perspective in terms of passenger preferences. See Figures MON.2 and SOM.2 and the corresponding analyses on the following pages.
Figure MON.2: Monmouth County Trip Length CDF

Figure SOM.2: Somerset County Trip Length CDF
For the trip length distributions in both Monmouth and Somerset counties (Figures MON.2 and SOM.2), there is an observable concavity throughout trip lengths of distances 0 miles to 5 miles. From 10 miles to 30 miles, the CDF appears approximately linear. Additionally, the frequency of trips does not drop off drastically until taxi trip lengths exceeding approximately 30 miles. We can conclude from the behavior of these curves that the relative marginal utility associated with increasing the length of a trip matters more when a passenger is travelling shorter distances.

This seems to be a reasonable conclusion. The difference between a twenty mile trip and a twenty-two mile trip is negligent in comparison to the difference between a two mile trip and a four mile trip.
A Closer Look at the Largest Volume Pixels in Monmouth and Somerset:

Freehold Borough:
As a town, Freehold Borough is among the highest in population density in Monmouth County (see Figure MON.3.1 on following page). We can thus postulate that many of the trips originating in and departing from Freehold Borough are home-based work trips. Additionally, attractions in Freehold Township include Freehold Raceway and Freehold Raceway Mall. Consequently, we can assume that there are a substantial amount of recreational trips made to and from the municipality as a whole.

<table>
<thead>
<tr>
<th>Municipality (w/ map index)</th>
<th>Municipal type</th>
<th>Population</th>
<th>Housing units</th>
<th>Total area</th>
<th>Water area</th>
<th>Land area</th>
<th>Pop. density</th>
<th>Housing density</th>
<th>Unincorporated communities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asbury Park (11)</td>
<td>city</td>
<td>16,116</td>
<td>8,076</td>
<td>1.60</td>
<td>0.18</td>
<td>1.42</td>
<td>11,319.5</td>
<td>5,672.4</td>
<td></td>
</tr>
<tr>
<td>Shrewsbury Township (48)</td>
<td>township</td>
<td>1,141</td>
<td>648</td>
<td>0.10</td>
<td>0.00</td>
<td>0.10</td>
<td>10,877.7</td>
<td>6,177.7</td>
<td></td>
</tr>
<tr>
<td>Keansburg (30)</td>
<td>borough</td>
<td>10,105</td>
<td>4,318</td>
<td>16.79</td>
<td>15.72</td>
<td>1.07</td>
<td>9,452.3</td>
<td>4,039.1</td>
<td>Tiltons Corner</td>
</tr>
<tr>
<td>Bradley Beach (10)</td>
<td>borough</td>
<td>4,298</td>
<td>3,180</td>
<td>0.63</td>
<td>0.02</td>
<td>0.61</td>
<td>7,023.6</td>
<td>5,196.6</td>
<td></td>
</tr>
<tr>
<td>Red Bank (26)</td>
<td>borough</td>
<td>12,206</td>
<td>5,381</td>
<td>2.16</td>
<td>0.42</td>
<td>1.74</td>
<td>7,019.1</td>
<td>3,094.4</td>
<td></td>
</tr>
<tr>
<td>Lake Como (6)</td>
<td>borough</td>
<td>1,759</td>
<td>1,115</td>
<td>0.27</td>
<td>0.01</td>
<td>0.25</td>
<td>6,943.6</td>
<td>4,401.4</td>
<td></td>
</tr>
<tr>
<td>Highlands (28)</td>
<td>borough</td>
<td>5,005</td>
<td>3,146</td>
<td>1.37</td>
<td>0.60</td>
<td>0.77</td>
<td>6,522.8</td>
<td>4,100.1</td>
<td>Waterwitch</td>
</tr>
<tr>
<td>Freehold Borough (35)</td>
<td>borough</td>
<td>12,052</td>
<td>4,249</td>
<td>1.95</td>
<td>0.00</td>
<td>1.95</td>
<td>6,180.8</td>
<td>2,179.1</td>
<td></td>
</tr>
</tbody>
</table>
Another known fact about Freehold Borough is that it is the hometown of world-renowned Rock N’ Roll musician Bruce Springsteen. Although unlikely, it is possible that New Jersey residents are drawn to the Borough to see the inspiration of some of Springsteen’s hit records including “My Hometown” and “In Freehold.”

As seen before in the analysis of Monmouth County as a whole, the cumulative distribution of trip lengths in Freehold Borough (Figure MON.3) follows a concave curve. Roughly ninety percent of the trips originating in and departing from Freehold Borough are less than twenty miles in length. Given the close attractions of Freehold Raceway and Freehold Raceway Mall, it is plausible to assume that a noticeable proportion of the home-based recreation taxi trips in Freehold Borough are to these local attractions.
North Plainfield:
Corresponding to Freehold in Monmouth County, North Plainfield Borough is the highest-density municipality in Somerset County, and contains the pixel with the largest volume of trip origins in the county. Based on our knowledge of the specific pixel (240, 129) and the surrounding pixels from previous analysis, we know that the borough contains a number of high-density residential suburbs, which generate home-based trips in the morning, which are likely to be commuter trips, whether in-county or to another county. Due to the common departure point of the trips represented in the above graph, and the density of the population, we can assume there are increased ridesharing opportunities from North Plainfield. This seems to be the case especially because we see an increased concavity in Figure SOM.3 compared to the Somerset County CDF of trip distances (SOM.2), which indicates more trips are travelling nearby. Shorter trips are typically more conducive to ridesharing since it is more likely people will travel to nearby destinations than to far-away ones due to convenience. Since North Plainfield borders Union County, some of these trips may be out-of-county, but it may also indicate a large number of intra-county trips within Somerset. North Plainfield also holds a large number of strip malls and superstores which can augment these trips with people departing from these stores after either shopping or completing a workday there.
Intra-County Vs. Intra-County Taxi Trips

The pie charts in Figures MON.4 and SOM.4 show the comparison in frequency between taxi trips that stay within a county at all times (intra-county) and taxi trips that will eventually leave the county to drop off its first, second, or third passenger. Perhaps surprisingly, our computations conclude that a vast majority of taxi trips in both Monmouth County and Somerset County will eventually leave their respective counties. This fact seems contradictory to the distribution of personal trip lengths examined before.
To explain this discrepancy, we postulate that the ride-sharing potential for shorter length taxi trips (zero to twenty miles) is substantially greater than the ride-sharing potential for longer trips (greater than twenty miles). We extend this hypothesis to conclude that the unexpectedly large number of inter-county taxi trips can be explained by a lack of ride-sharing.

In layman’s terms: the farther you are going on your taxi trip, the less likely it is that you are sharing that taxi with someone who is leaving from the same origin. If your destination is outside of your current county, then it is possible that you will require your own taxi. This could offer an explanation for the surprisingly large number of taxi trips labelled as “inter-county.”
Average Vehicle Occupancy Analysis for Taxi Trips:

The following graphs show the average number of occupants per vehicle at departure (AVO) for different levels of service in the top ten most active pixels in both Monmouth and Somerset counties. In Monmouth County, the top ten most active pixels account for 188,102 daily taxi trips. Given that Monmouth County has 1,999,195 total daily taxi trips, the top ten pixels account for approximately 9.4 percent of Monmouth County’s total daily trips. Similarly, in Somerset County, the top ten most active pixels account for 14.6 percent of Somerset County’s total daily trips. When we consider the ridesharing potential for the top ten pixels in Monmouth and Somerset counties, we should acknowledge that the calculated AVOs will perhaps be inflated. This is because we are only considering a fraction of a county’s total trips, all of which involve the most active pixels, which may also have some of the highest potentiality for ride sharing. The figures on the following pages (Figures MON.5-7 and Figures SOM.5-7) are nonetheless relevant because they allow us to see how changing the number of common destinations or the departure delay affects AVO.
Figure MON.5: Analysis for Monmouth County, CD = 3 (Figure MON.5):
The above chart shows that in the top ten pixels, when there were three common destinations and a
departure delay of one minute, the average vehicle occupancy (AVO) in Monmouth County was 2.67 riders.
When there were three common destinations and a departure delay of 2.5 minutes, the average vehicle
occupancy was 2.69 riders. When there were three common destinations and a departure delay of 5 minutes,
the average vehicle occupancy was again 2.69 riders.

These findings indicate that increasing the departure delay while keeping the number of common
destinations fixed increases average vehicle occupancy by almost a negligible amount. These findings also
show that with three common destinations with a departure delay between one and five minutes, we could
potentially take away a little more than half the vehicles currently on the roads. This would cut pollution,
congestion, and fuel consumption in half, thus revolutionizing our means of travel and substantially decreasing
our carbon footprint.
Figure SOM.5: Analysis for Somerset County, CD = 3 (Figure SOM.5):

In a similar chart for the top ten pixels in Somerset County, when there were three common destinations and a departure delay of one minute, the AVO was 2.24 riders. With three common destinations and a departure delay of 2.5 minutes, the AVO became 2.25 riders, and with three common destinations and a departure delay of 5 minutes, the AVO was 2.26 riders.

Here we see AVO values that are less than those of Monmouth County, corresponding with the fact that people are likely travelling further and out-of-county, and as such are travelling together less. This is possibly due to Somerset’s increased proximity to the large, dense urban regions of New Jersey and New York that may draw more commuting workers. Additionally, these findings again show that increasing the departure delay while keeping the number of common destinations fixed does not meaningfully increase AVO. However, the findings maintain the fact that with these level-of-service conditions, we could potentially take away a little more than half the vehicles currently on the roads in Somerset County as well, leading to the environmental and economic benefits listed above in another county.
Comparing the average vehicle occupancy chart for Monmouth County when there are three common destinations (Figure MON.5) and the AVO chart for Monmouth County when there are four destinations (Figure MON.6) highlights a significant increase in AVO when there is one extra common destination. When there are four common destinations and a departure delay of 1 minute, the AVO is approximately 3.16. When there are four common destinations and a departure delay of 2.5 minutes, the AVO is approximately 3.18. When there are four common destinations and a departure delay of 5 minutes, the AVO comes out to be 3.19.

These averages show that again, increasing the departure delay almost does not affect the average vehicle occupancy. Comparing these averages to the averages for a common destination of 3 (from Figure MON.5) however shows an increase of approximately 0.6 riders per trip, and indicates that increasing the number of common destinations, even by one, significantly increases the average vehicle occupancy.
Comparing AVO for three common destinations and four destinations highlights a significant increase in AVO when there is one extra common destination. When there are four common destinations (Figure SOM.6) and a departure delay of 1 minute, the AVO is approximately 3.16. When there are four common destinations and a departure delay of 2.5 minutes, the AVO is approximately 3.18. When there are four common destinations and a departure delay of 5 minutes, the AVO comes out to be 3.19.

These averages show that again, increasing the departure delay almost does not affect the average vehicle occupancy. Comparing these averages to the averages for a common destination of 3 (Figure SOM.5) however shows an increase of approximately 0.6 riders per trip, and indicates that increasing the number of common destinations, even by one, significantly increases the average vehicle occupancy.
The Effects of Increasing the Number of Common Destinations:

The following figures (MON.7 and SOM.7) put together a range of common destinations (the lowest being 1 and the highest being 4) for a consistent delayed departure of 5 minutes. Looking at this data side by side it is easy to see that increasing the number of common destinations significantly increases the average number of vehicle occupants. With a common destination of 1, there is an average vehicle occupancy of approximately 1.2 in both Monmouth and Somerset counties versus an average vehicle occupancy of approximately 3.2 in Monmouth and 2.7 in Somerset when there are 4 common destinations. In summary, by having four common destinations, aTaxis could cut the number of cars on the road by roughly one third, leading to significant decreases in congestion, pollution, and fuel consumption.

Figure MON.7: Monmouth County top 10 pixels AVO Analysis
Figure SOM.7: Somerset County top 10 pixels AVO analysis
End-of-Day Imbalances by Vehicle Size:

The following figures (Figures MON.8-11 and Figures SOM.8-11) display the cumulative distribution by pixel of the end-of-day imbalances of aTaxis by vehicle size. The x-axis is measured in terms of the end-of-day imbalance of aTaxis, given in the number of vehicles. A positive number denotes an end-of-day surplus and a negative number denotes an end-of-day deficit. The y-axis gives the cumulative percentage of pixels for which the end-of-day imbalance is at a particular level.

At a glance, we notice that each cumulative distribution function has roughly equal area to the left and to the right of the value zero on the x-axis, with the exception of 1-3 passenger vehicles. This implies that for every vehicle size greater than 1-3 passengers, there is generally speaking not much of an end-of-day imbalance of aTaxis of that size across Monmouth and Somerset counties. However, for vehicles of size 1-3 passengers, we see that roughly eighty percent of pixels have an end-of-day imbalance less than zero (i.e. an end-of-day deficit of vehicles). We can conclude that if we were to orchestrate early morning repositioning, vehicles of size 1-3 passengers would have to be brought to Monmouth and Somerset counties overnight.

We can also notice that the cumulative distribution functions appear more and more like vertical lines as the vehicle size increases. We must first acknowledge that this trend occurs because there are less total vehicles available as vehicle size increases. Therefore, we can expect vehicle imbalances of any sort to be smaller for larger vehicle sizes. If we look further into this trend, we can postulate that the demand for larger vehicles is potentially more balanced than the demand for 1-3 passenger vehicles. This might be the case if vehicles are going to and from the same locations within a given county (e.g. schools, large offices, train stations).
Figure MON.8: Monmouth County EOD imbalance CDF

Figure SOM.8: Somerset County EOD imbalance CDF
Figure MON.9: Monmouth County EOD imbalance CDF for vehicle size of 6

Figure SOM.9: Somerset County EOD imbalance CDF for vehicle size of 6
Figure MON.10: Monmouth County EOD imbalance CDF for vehicle size of 15

Vehicle Size = 7 to 15

Figure SOM.10: Somerset County EOD imbalance CDF for vehicle size of 15
Histograms of Departures, by Hour:

The hourly departures of vehicles in Monmouth and Somerset counties (Figures MON.12 and SOM.12) generally follow the same trends. We observe peaks in vehicle departures from roughly 6 AM to 9 AM, from 11
AM to 1 PM, and from 5 PM to 8 PM. These periods correspond to morning rush hour, lunch time, and evening rush hour, respectively.

**Figure MON.12**: Monmouth County departure times
Histograms of Arrivals, by Hour:

The hourly arrivals of vehicles in Monmouth and Somerset counties seen on the following page (Figures MON.13 and SOM.13) experience the same rush hour trends as did the departures (Figures MON.12 and SOM.12). Thus there is little new information given by Figures MON.13 and SOM.13 as stand-alone figures.

However, we see that the peaks in the histograms of arrivals for both counties occur on average one hour later than in the histograms of the departures, which we will call the “next-hour” effect. This implies that many of the rush hour trips run into the next hour. Looking back to MON.1 and SOM.1, we see that roughly half of the trips in Monmouth County are less than 30 minutes, and that roughly half of the trips in Somerset County are less than 20 minutes. Given the “next-hour” effect noticed between the departures and arrivals histograms, we can conclude that many of the trips in Monmouth and Somerset counties during rush hour must be longer than the average trip durations for each county in order to produce such a noticeable effect.
**Figure MON.13:** Monmouth County Arrivals

**Figure SOM.13:** Somerset County Arrivals
Net Imbalance Figures (Arrivals Less Departures): Densities and CDFs

The following figures (Figures MON.14-15 and Figures SOM.14-15) show the Net Imbalance Density of arrivals less departures and the Cumulative Net Imbalance Density of Arrivals and Departures, respectively. The x-axis breaks up the day in hour long segments and the y-axis is dependent on the graph. These graphs convey the overall surplus or deficit of taxis in the county by hour. The Non-Cumulative Net Imbalance Density graphs allow for one to clearly see the morning rush hour, the noon lunch break, and the evening rush hour. During these times, there are deficits of available taxis in the county suggesting that in the case of early morning repositioning, Monmouth would require quite a few taxis imported from outside the county to ensure no deficits throughout the day. The Cumulative Net Imbalance graphs better illustrate just how many taxis would have to be brought into the county at midnight to prevent against a deficit at any point throughout the day. These graphs show the net number of taxis available based on how many taxis have entered and left the county earlier in the day. The extreme deficit of taxis at the end of the day, approximately 10,000 taxis, makes sense for Monmouth County because of the high number of intercounty taxi trips throughout the day (70% of all taxi trips are intercounty). The highest deficit of taxis, approximately 40,000, occurs at 6 pm, the height of rush hour. The highest surplus of taxis, approximately 20,000, occurs at 5am before early morning rush hour.

The Somerset County graphs show similar trends with lower surplus and deficit numbers. This is simply because there were fewer trips coming and going from Somerset County. The Non-Cumulative Net Imbalance graph shows the same morning rush hour, lunch break, and evening rush hour trends but with longer windows of time indicating that rush hours and lunch breaks last longer in Somerset County. This may be due to the county’s central location within the state indicating that people travel to a wider range of places causing a wider range of travel times. Interestingly, the Cumulative Net Imbalance graph for Somerset County indicates that unlike Monmouth County, they keep a relatively surplus until evening rush hour. Monmouth County has a deficit from 11am onwards. This might be due again to the central location of Somerset.

Net Imbalance Densities, by Hour:
Figure MON.14: Monmouth County Net Imbalances
Figure SOM.14: Somerset County Net Imbalances
Cumulative Net Imbalances, by Hour:

Figure MON.15: Monmouth County Cumulative Net Imbalances
Figure SOM.15: Somerset County Cumulative Net Imbalances
Available Vehicles, by Time of Day:

Figures MON.16 and SOM.16 on the following pages show plots of the available number of vehicles for Monmouth and Somerset counties by time of day, by vehicle size. The plots were generated by taking the difference between the cumulative arrivals and cumulative departures.

1-3 Passenger Vehicles:

For both Monmouth and Somerset counties, we see a general trend throughout the day for the number of available vehicles to decrease as a typical day progresses. We see that the county as a whole generally reaches a vehicle deficit that is never gained back at around mid-day (12 PM to 1 PM). We can extrapolate that continuous vehicle repositioning throughout a typical day would call for 1-3 passenger vehicles to be sent to Monmouth and Somerset counties at this time.

4-6 Passenger Vehicles, 7-15 Passenger Vehicles, and 16+ Passenger Vehicles:

Morning rush hour initially causes substantial deficits in 4-6, 7-15, and 16+ passenger vehicles in both Monmouth and Somerset counties. However, the deficit of 4-6 passenger vehicles in both counties is gained back at the end of morning rush hour (8:30 AM to 10 AM). A possible explanation for this phenomenon could be that there exists a late morning rush hour in these counties consisting of home-based school trips and industries that start their jobs later in the morning.

Early morning workers presumably leave the central Jersey suburbs to commute to many jobs which are outside of their respective counties, whereas school children and late morning workers enter the central Jersey suburbs and make up this deficit later in the morning. This process accounts for great ride-sharing potential as thousands of 7-15 passenger vehicles and hundreds of 16+ passenger vehicles are deployed for the morning commutes.
Figure MON.16: Monmouth County available vehicles size 3, 6, 15, and 50
Figure SOM.16: Somerset County available vehicles of size 3, 6, 15, and 50
Histograms of Available Vehicles, by Time of Day:

Figures MON.17 and SOM.17 on the following pages represent the frequency at which a particular range of available vehicles is available. That is to say, a large value for a particular bin in one of the histograms implies that the corresponding range of available vehicles occupies a large percentage of time for a particular day. The x-axis is measured in terms of the number of available aTaxis whereas the y-axis measures the frequency at which a particular range is recorded throughout a typical day, which is recorded in minutes (each histogram’s frequencies sum to 1440).

1-3 Passenger Vehicles (Light Blue Histograms):

We notice that both Monmouth and Somerset register prolonged periods of large deficits of 1-3 passenger vehicles at some point throughout a typical day. Referring back to Figures MON.16 and SOM.16, we see that this prolonged deficit occurs in the evenings of a typical day. Monmouth County experiences a deficit of 3500 to 4000 in 1-3 passenger vehicles for a duration of nearly 3 hours, which in itself highlights the need for vehicle repositioning.

4-6 Passenger Vehicles (Light Green Histograms):

Monmouth and Somerset counties appear to be much more balanced with regards to avoiding long durations of significant deficits or surpluses of 4-6 passenger vehicles. The only time of day for which large deficits of vehicles are ran is for a short window during morning and evening rush hours.

7-15 Passenger Vehicles and 16+ Passenger Vehicles (Orange and Red Histograms):

The central Jersey suburbs are even more balanced in their supply and demand for large vehicles (7-15 passengers and 16+ passengers). The morning rush hour deficit self corrects.
Figure MON.17: Monmouth County available vehicle histogram
Beginning Time of Daily Vehicle Surplus in Monmouth County:

The following figures (Figures MON.18-21 and SOM.18-21) are color gradient maps indicating the time of day a pixel within Monmouth or Somerset County have a surplus of taxis for the rest of the day. As can be seen by referencing the color gradient scale to the right of the mapping, dark blue represents a pixel has a surplus of cars from very early on in the day until the end of the day, while green indicates a pixel has a surplus of cars from halfway through the day until the end of the day, and a yellow pixel represents that the pixel does not maintain a surplus of taxis until the end of the day if at all. These charts are important when looking to implement early morning repositioning within a county. Early morning repositioning relies on the ability to move cars multiple miles in less than a second to prepare the county for the rest of the day. Obviously, this is an impossible task. By knowing when certain pixels start maintaining a surplus, taxis could begin to be repositioned throughout the day, making repositioning more realistic and perhaps more efficient in regards to how many hours a driver must be on the clock.

These gradient maps were divided by the average vehicle occupancy (1-3 passengers, 4-6 passengers, 7-15 passengers, and 16+ passengers) because repositioning a vehicle that could only hold three passengers to a pixel that needs a 6-passenger vehicle would be illogical.
Three-Passenger Vehicles

The gradient maps for 3 passenger vehicles (Figures MON.18 and SOM.18) indicate that few pixels maintain surpluses from early in the day. This means that it would be difficult to start repositioning taxis before midnight. Furthermore, 47,573 vehicles would have to be imported into Monmouth County from another county at the beginning of the day to offset deficits and 21,215 vehicles would need to be brought into Somerset County.

Figure MON.18: Monmouth County BOD vehicle size of 3 surplus
Six-Passenger Vehicles

The six passenger vehicle gradient maps (Figures MON.19 and SOM.19) shows many more opportunities to begin repositioning earlier in the day. This can be seen by the fact that many parts of the counties are blue, green, or orange. For Monmouth County, 7,154 vehicles would need to be imported from another county at the beginning of the day to insure the county never suffered a deficit. For Somerset County, 2,871 vehicles would need to be imported from another county.
Figure SOM.19: Somerset County BOD vehicle size of 6 surplus
Fifteen-Passenger Vehicles

For the 15 passenger vehicles (Figures MON.20 and SOM.20), it can be seen that more pixels maintain a surplus starting early in the day while fewer maintain a surplus starting in the early evening. This suggests that larger vehicles are used during morning rush hour and some during evening rush hour but then not used often during the later hours of the day. For Monmouth County, 3,037 vehicles would need to be imported from other counties at the start of the day to insure the county never suffered a deficit. For Somerset County, 1,486 vehicles would need to be imported from other counties.

Figure MON.20: Monmouth County BOD vehicle size of 15 surplus
Figure SOM.20: Somerset County BOD vehicle size of 15 surplus
Sixteen-or-More Passenger Vehicles

For 16+ passenger vehicles it can be seen that most can be repositioned starting early in the day (Figures MON.21 and SOM.21). This again indicates that 16+ passenger vehicles are used in the morning for rush hour transportation and some for evening rush hour but not needed otherwise. The larger need for bus sized vehicles during morning rush hour versus evening rush hour can be explained by the smaller morning rush hour window, leading to more ridesharing.

Night time rush hour is more spread out due to more diverse release times from work whereas most people begin work around the same time. For Monmouth County, 272 vehicles would need to be imported from another county at the beginning of the day to insure the county never suffered a deficit. For Somerset County, 180 vehicles would need to be imported from another county.

Figure MON.21: Monmouth County BOD vehicle size of 50 surplus
**Summary Table:**

<table>
<thead>
<tr>
<th>County</th>
<th>Number of taxis that need to imported from outside the County for EMR</th>
<th>Number of taxis that can be repositioned within the county at the end of the day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monmouth</td>
<td>3 passenger: 42,573</td>
<td>3 passenger: 193</td>
</tr>
<tr>
<td></td>
<td>6 passenger: 7,154</td>
<td>6 passenger: 796</td>
</tr>
<tr>
<td></td>
<td>15 passenger: 3,074</td>
<td>15 passenger: 416</td>
</tr>
<tr>
<td></td>
<td>16+ passenger: 272</td>
<td>16+ passenger: 187</td>
</tr>
<tr>
<td>Somerset</td>
<td>3 passenger: 21,215</td>
<td>3 passenger: 116</td>
</tr>
<tr>
<td></td>
<td>6 passenger: 2,871</td>
<td>6 passenger: 491</td>
</tr>
<tr>
<td></td>
<td>15 passenger: 1,486</td>
<td>15 passenger: 280</td>
</tr>
<tr>
<td></td>
<td>16+ passenger: 180</td>
<td>16+ passenger: 130</td>
</tr>
</tbody>
</table>

*Figure SOM.21: Somerset County BOD vehicle size of 50 surplus*
External Locations
By: Ellie McDonald, Joe Yates, and Tom Byrne

❖ Rockland County, NY
❖ Bucks County, PA
❖ North
❖ South
❖ West

❖ New York City
❖ Philadelphia
❖ International (trips to Newark Airport)
Overview of External Destinations:

The external destinations analyzed were New York City, Philadelphia, International (trips to Newark Airport), Rockland County, NY, Bucks County, PA, a cumulative point north of the state of New Jersey, a cumulative point south of the state of New Jersey, and a cumulative point west of the state of New Jersey. These external points provided approximately 811,174 trips, but only 257,551 of these trips were taxi trips. We made the assumption that all of the trips from NYC, Newark Airport, and Philadelphia were train trips due to the large number of commuters that already travel from these destinations using trains and the fact that they have extraordinarily well-developed railroad systems from all three points into multiple New Jersey train stations. For the other external locations, Rockland, Bucks County, North, South, and West, we simplified our analysis and data by assuming that all taxi trips from these locations left and returned to the same pixel. This would represent in each external location a single aTaxi stand that sent taxis into New Jersey, possibly owned by the New Jersey autonomous taxi company and brought aTaxis back to the same stand essentially functioning as a train station in this regard.
We see that the cumulative distribution of taxi trip times to and from the external locations is approximately linear at least until the trip time reaches seventy minutes. We do not see a noticeable concavity early in the curve as we saw in both Monmouth and Somerset counties. It is plausible that people who choose to live outside the state in which they work do not have the same utility and preferences associated with shorter commutes. In other words, if you work at Princeton University and then choose to buy a house in Bucks County, PA, it is probable that making short commutes is generally not that important to you.
Figure EXT.2 displays the average vehicle occupancy for the external locations that supply and demand taxi trips. The average vehicle occupancy for 3 common destinations and a 2.5 minute delay departure varied based on the external location but stayed within a 0.24 range. The average vehicle occupancy for Bucks County was 2.81. The average vehicle occupancy for the Northern Hub was 2.89. The average vehicle occupancy for Rockland County was 2.65. The average vehicle occupancy for the Southern Hub was 2.85. The average vehicle occupancy for the Western Hub was 2.75. The overall average vehicle occupancy for external locations was 2.79. Rockland might have the lowest average vehicle occupancy while the Northern Hub had the highest because they both come from the same geographic area. Some of the Northern Hub’s passengers may have been able to also use the Rockland stand, but chose not to.
Figure EXT.3: External locations AVO analysis

Figure EXT.3 shows the average vehicle occupancy for all five external locations that have taxi trips when the number of common destinations goes from 1 to 4 and the departure delay remains 2.5 minutes. As is apparent from the graphic, the number of common destinations dramatically increases the average vehicle occupancy. When there is a single common destination there is an average vehicle occupancy of approximately 1.114 across all externals. When there are 2 common destinations there is an average vehicle occupancy of approximately 1.854 across all externals. When there are 3 common destinations there is an average vehicle occupancy of approximately 2.79 across all externals. When there are 4 common destinations there is an average vehicle occupancy of approximately 3.326 across all externals. This illustrates the same upward trend seen in Monmouth County and Somerset County: when the common destination increases the average vehicle occupancy increases dramatically. The average vehicle occupancy for the external locations is slightly higher than the average vehicle occupancy for Monmouth County or Somerset County because it has a single aTaxi stand, forcing more ridesharing. Furthermore, almost all trips from these external locations are work related, meaning they depart and arrive at similar times of day which lends itself to ridesharing.

Histograms of Departures and Arrivals Concerning the External Locations:

The following page contains Figures EXT.4 and EXT.5, which plot histograms of the daily departures and arrivals aggregated amongst all of the external locations. Aggregating all of the external locations assumes that each external location individually follows the same trends in trip departures and arrivals as one another. Given the assumptions made to construct the trips to and from the external locations up to this point, we believe that the assumption holds and that the following figures paint a picture that can be applied to each external location.
The takeaway from Figures EXT.4 and EXT.5 is that they are roughly mirror images of one another. We see in the departures from the external locations that there is a large spike in departures throughout the entirety of morning rush hour, from 5 AM to 9 AM. However, most of these dispersed departures arrive at the same hour, between 7 AM and 8 AM. An opposite trend is seen during evening rush hour, when it is the departures which are bunched together and the evening arrivals which are dispersed between the hours of 7 PM and 11 PM.

We suspect that the bunching of arrivals during morning rush hour and departures during evening rush hour illustrate the tendency of workers to arrive at the office just before their shifts start and to leave the office immediately when their days end.
Figure EXT.4: External Departure Times

Figure EXT.5: External Arrival Times
Figure EXT.6: External Net Imbalances

Figure EXT.6 shows the net imbalance density of arrivals and departures across external locations. As can be seen from the above graph, there is very little activity until early morning rush hour and then again until evening rush hour. This demonstrates that most external trips are work based and hence mostly travel during rush hours. The departure spike in the morning represents the commuters who live outside New Jersey leaving for work and the arrival spike in the morning represents the commuters who come from New Jersey to work outside the state. The departure spike at night represents these New Jersey residents returning home at the end of the day and the arrivals at the end of the day represent people coming back from New Jersey at the end of the day.
The cumulative net imbalance of vehicles shown in Figure EXT.7 effectively represents the current number of available vehicles in the external locations at the end of each hour. We see that a deficit of vehicles is created during morning rush hour, and that this deficit is more or less maintained until the beginning of evening rush hour, when the deficit doubles from approximately 1,200 vehicles to approximately 2,400 vehicles at roughly 5 PM.

As evening rush hour continues through 6 PM, 7 PM, and then 8 PM, we see that there is a significant amount of vehicles entering the external locations. We assume that workers who live in the external locations leave their jobs at relatively the same time as New Jersey residents. However, the external residents necessarily have a longer commute home from work. Thus, the large jump in arrivals of vehicles in the external locations later in the evening represents people getting home from work between roughly the hours of 6 PM to 9 PM.
Acknowledgements

Contributions to code and figure generation were made by:

❖ Chad Cowden ’17
❖ Kyle Marocchini ’18
❖ Keith Gladstone ’17
❖ Alicia Lamb ’17
❖ Kevin Manyara ’17
❖ Trevor Osborne ’17
❖ Aaron Schwartz ’17
❖ Alexander Singleton ’17
aTaxi Analysis in Middlesex County

Naman Jain & Tyler Rudolph
ORF 467 Transportation System Analysis, Fall 2015-2016
Professor Alain Kornhauser
Introduction

As the second most populous county right in the middle of New Jersey, Middlesex County is very much the beating heart of the state. Its county seat is New Brunswick and it’s most populous town is Edison, both iconic towns that represent everything it is to be from New Jersey. This county is vital economically to the state for several reasons. First and foremost it is the home of America’s medical industry. The Robert Wood Johnson Foundation, many prominent research hospitals, and large drug companies boast headquarters in this county. It is also home to Rutgers, New Jersey’s largest college campus and source of ambitious educated youth. The city of Edison, birthplace of the light bulb and many other very famous inventions from the mind of Thomas Edison remains a research powerhouse and home to a variety of prominent labs and facilities. Only 45 minutes from downtown Manhattan, it also holds sway as a major commuting hub for New York City office workers looking to live in a quieter neighborhood.

With such an educated and mobile workforce and with access to a major metropolitan area and numerous international airports, it is critical that Middlesex have a state of the art transportation system. The goal of this report is to thoroughly investigate the possibility for the establishment of an aTaxi system in the county. With the rapid advances in autonomous car technology over the past decade, planning for impact on infrastructure and developing feasibility studies is critical for municipalities looking to stay with the times. This report will investigate county characteristics and flesh out actual supply and demand available for an aTaxi system, with the hope that when the technology is fully realized New Jersey and Middlesex will be ready to capitalize on it.

aTaxi Vehicle Management

Analysis of aTaxi Imbalances

Supply and Demand Characteristics

Before we delve into the deeply technical analysis, let’s overview a few characteristics of the county’s supply and demand that will put all of this in context.

Demand in Middlesex County varies extremely. There are pixels/regions where demand is only a few aTaxis a day, but there are also regions where demand is in the tens of thousands. This is a critical consideration as we consider potential fleet positioning of our aTaxis. Our highest demands are concentrated in pixels (125:137) and (220:228), which makes sense as these both represent very active urban areas where people are living, working, and moving around with considerable regularity.

Supply in Middlesex, is much steadier. The majority of aTaxis supplied originate in adjacent counties or neighborhoods. We see that levels of available aTaxis remain almost entirely constant from 8AM through the day and into the early evening. At which point they drop for the night.
What we will see as a recurring theme throughout Middlesex however is the fact that it has a surplus of aTaxis. It is a net exporter as a result can serve as an effective hub for the region.

**Inter vs. Intra County Characteristics**

![Fig. 1](image)

Fig. 1 – Comparison of Departure and Arrivals in Middlesex County

Fig. 1 is critical for us understanding how Middlesex interacts with other counties in the region. We can see that the vast majority of departures go outside of the county and the vast majority of arrivals come from outside of the county as well. This supports the idea of Middlesex being a major regional hub. Its traffic comes from commuters going into New York as well as workers entering the region to work in the major research centers.

This is not your typical sleepy and self-contained neighborhood. Middlesex is a regional player that is tied in to the rest of Jersey and the states around it.

**Daily Taxi Imbalance**

The core of our supply and demand analysis rests on determining the daily aTaxi imbalances in a given pixel. The following graphs (Figs. 2-5) provide a visual examination of the imbalances present in Middlesex County. Each column represents a single pixel in that county. The column height represents the amount of imbalance. A negative imbalance means that there are more departures than arrivals and visa versa. The county was broken down into sub-regions and then at the end compiled into one final aggregate analysis. This gives us a sense of
how many cars are sitting around and available at a given moment. This is critically important, because if there are massive imbalances more cars will have to be provided from other counties or locations.

Fig. 2 – Middlesex Subregion 1 Imbalance

Fig. 3 – Middlesex Subregion 2 Imbalance
We can in Fig. 5 that we have a vehicle surplus present for about 65% of all the pixels in Middlesex County. This also means that we have a shortage for over a third of the county. The most encouraging thing about this
data however, is that the vast majority of the pixels have very small absolute imbalances. There are a few outliers on the end with massive discrepancies, but the majority of the county is very well served. This means we will have to engage in very little importation or exportation of aTaxis to meet demand.

**Imbalances throughout the day**

The second part of our analysis examines how supply and demand both change over a given day. Fig. 6 shows demand (departures) and supply (arrivals) of trips in all of Middlesex County. The horizontal difference between the red and the blue lines shows the imbalance between demand and supply at a particular time. If the red line is to the left of the blue line, there is over supply but if the red line is to the right of the blue line then we have a shortage. From the graph below we do see that the two lines mostly follow each other pretty closely throughout the day. We see some oversupply early in the day (morning commute) but it is limited.

![Fig. 6 – Supply and Demand Imbalance](image)

To nuance this analysis we looked at the supply and demand analysis for those vehicles with just two passengers (Fig. 7). On the graph below we see a similar pattern when we look at the demand (departures) and supply (arrivals) of just those vehicles with two passengers. There is some oversupply and inefficiency in the morning, which oscillates before noon such that the absolute imbalance shrinks as we move through the day.
The pixel (125, 220) was by far the most active pixel with more than 50,000 person trips (and the most vehicle trips as well as seen from the horizontal axis of the graph above) originating and ending within the bounds of the pixel throughout the course of the day. We see from Fig. 8 that though there is a significant oversupply of vehicles in this pixel as we have almost 3000 more vehicles arriving on a daily basis than departing on aggregate. The imbalance is even larger during the afternoon/evening rush hour around 4-5pm when the oversupply is north of 5000 vehicles.
Next, we provide further analysis on the pixel with the most significant undersupply in the county i.e. pixel (137,228) with a total imbalance of -2942 by the end of the day. We look at how this imbalance evolves through the day below. We can see that there are slightly more departures than arrivals all day, which supports our initial point that Middlesex is a strong supplier of taxis to other counties.

![Graph showing supply and demand imbalance in pixel (137,228).]

Fig. 9 – Supply and Demand Imbalance, Pixel (137,228)
aTaxi Utilization Curves

Utilization curves give us an idea of exactly how many vehicles and people are on the road at a given time. For an aTaxi program to be effective, aTaxi’s must be able to totally match all of the current demand on the roads. Otherwise, the program will leave people stranded and will fall out of use.

Utilization curves have been created for the total number of vehicles and the total number of people on each road. These were also broken down further to examine the number 3, 6, 15, and 50 passenger vehicles on the road at a given moment. This only deals with departure data because we simply need to know what when there is demand for the system. Demand for departures is a complete picture of utilization data; it will also provide us with the minimum aTaxi fleet size, as we will only be happy when fully servicing the demand. As the Middlesex County dataset is so large, each of these is broken down into sub-regions. There are four subregions in Middlesex.

A few observations:
1. We can see a sharp size for morning rush hour followed by mid afternoon peaks and valleys. The late night hours and early mornings are relatively quiet.
2. One interesting aspect to note is that the rush hour peak starts slightly later than would be expected, I suspect this is because of the strong influence of research facilities and universities which operate on a more relaxed schedule than a corporate environment would. There is also a fair amount of usage late at night which I am hypothesizing comes from a strong college bar scene.

3. The values that are at “0” are not truly zero; they are simply the result of a rounding we were required to do to make the data processing possible.

Utilization Curves by Total Vehicle Numbers
This is a simple breakdown of exactly how many vehicles are on the road at a given time.
Fig. 10 – Middlesex Subregion 1 Total Vehicles vs. Time

Fig. 11 – Middlesex Subregion 2 Total Vehicles vs. Time
Utilization Curves by Total People
This is a simple breakdown of exactly how many people are on the road at a given time.
Fig. 14 – Middlesex Subregion 2 Total People vs. Time
Fig. 15 – Middlesex Subregion 3 Total People vs. Time

Fig. 16 – Middlesex Subregion 4 Total People vs. Time
Utilization Curves by Vehicle Size

Capacity 3:

Fig. 15 – Middlesex Subregion 1 Total Vehicle vs. Time, Capacity 3

Fig. 16 – Middlesex Subregion 2 Total Vehicle vs. Time, Capacity 3
Fig. 17 – Middlesex Subregion 3 Total Vehicle vs. Time, Capacity 3

Fig. 17 – Middlesex Subregion 4 Total Vehicle vs. Time, Capacity 3
Capacity 6:

Fig. 18 – Middlesex Subregion 1 Total Vehicle vs. Time, Capacity 6

Fig. 19 – Middlesex Subregion 2 Total Vehicle vs. Time, Capacity 6
Fig. 20 – Middlesex Subregion 3 Total Vehicle vs. Time, Capacity 6

Fig. 21 – Middlesex Subregion 4 Total Vehicle vs. Time, Capacity 6
Fig. 22 – Middlesex Subregion 1 Total Vehicle vs. Time, Capacity 15

Fig. 23 – Middlesex Subregion 2 Total Vehicle vs. Time, Capacity 15
Fig. 24 – Middlesex Subregion 3 Total Vehicle vs. Time, Capacity 15

Fig. 24 – Middlesex Subregion 4 Total Vehicle vs. Time, Capacity 15
Capacity 50

Fig. 25 – Middlesex Subregion 1 Total Vehicle vs. Time, Capacity 50

Fig. 26 – Middlesex Subregion 2 Total Vehicle vs. Time, Capacity 50
Fig. 27 – Middlesex Subregion 3 Total Vehicle vs. Time, Capacity 50

Fig. 28 – Middlesex Subregion 4 Total Vehicle vs. Time, Capacity 50
Minimum aTaxi Fleet Size

Without the use of extensive algorithmic estimation or a variety of other data sets, we can rely on figuring out the minimum aTaxi fleet size necessary by using brute force. This is as easy as looking at each of the graphs above and determining the maximum number of vehicles in use at any one time. This provides the max for each sub-county and each vehicle type. These maximum peaks are all timed at roughly the same point during the day, so we simply add these up and determine the minimum number of vehicles required to service the max utilization at any one time. This is broken down below by car capacity and by sub-region.

This information is essential because it allows us to estimate total investment required as well as to determine what sort of other infrastructure we need to build around the aTaxi system. Without accurate estimates of fleet size, Middlesex is left in the dark as to how to build up and prepare for this technological advance.

<table>
<thead>
<tr>
<th></th>
<th>Sub-county 1</th>
<th>Sub-county 2</th>
<th>Sub-county 3</th>
<th>Sub-county 4</th>
<th>Total Per Vehicle Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity 3</td>
<td>3900</td>
<td>4493</td>
<td>4505</td>
<td>3020</td>
<td>15918</td>
</tr>
<tr>
<td>Capacity 6</td>
<td>592</td>
<td>798</td>
<td>922</td>
<td>508</td>
<td>2820</td>
</tr>
<tr>
<td>Capacity 15</td>
<td>85</td>
<td>272</td>
<td>93</td>
<td>66</td>
<td>516</td>
</tr>
<tr>
<td>Capacity 50</td>
<td>4</td>
<td>25</td>
<td>1</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td><strong>Total per Sub-county</strong></td>
<td><strong>4581</strong></td>
<td><strong>5579</strong></td>
<td><strong>5521</strong></td>
<td><strong>3597</strong></td>
<td><strong>19287 Total Vehicles</strong></td>
</tr>
</tbody>
</table>

Fig. 29 – Middlesex EMR Fleet Sizes

Early Morning Repositioning Analysis

Early morning repositioning is an optimization method based on moving cars around at night so they are best positioned to respond to the next days demand. This reduces wait time and ensures that demand can be met in each region.

This graphic (Fig. 30) represents how this works on a daily basis.
The EMR cumulative data was used to produce the following estimates for fleet size if an EMR repositioning system is used. This obviously is meant to include a large number of vehicles that have access to another county and is not the isolated case study that we use above. As a result the numbers are significantly higher. We have also provided an estimate of the total percentage of the entire NJ fleet that would be used to service Middlesex County (Fig. 31). This is useful from a state planning basis and give them a sense of where and how resources should be allocated.

<table>
<thead>
<tr>
<th></th>
<th>EMR Fleet Size</th>
<th>% of Total Fleet</th>
<th>EoD Fleet Size</th>
<th>% of total fleet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity 3</td>
<td>128835</td>
<td>8.3%</td>
<td>74345</td>
<td>4.8%</td>
</tr>
<tr>
<td>Capacity 6</td>
<td>32688</td>
<td>8.3%</td>
<td>15157</td>
<td>3.9%</td>
</tr>
<tr>
<td>Capacity 15</td>
<td>6618</td>
<td>7.1%</td>
<td>3819</td>
<td>4.1%</td>
</tr>
<tr>
<td>Capacity 50</td>
<td>805</td>
<td>6.0%</td>
<td>783</td>
<td>5.8%</td>
</tr>
<tr>
<td>Totals</td>
<td>168946</td>
<td>8.3%</td>
<td>94104</td>
<td>7.7%</td>
</tr>
</tbody>
</table>
Hunterdon, Sussex, and Warren
Chad Cowden and Ryan Siiro

Hunterdon County
Trip Type Comparison
Before looking at the Taxi trips individually, we wanted to get a sense of what is actually going on in Hunterdon so we compared the Taxi trips to the other trips in the county.

<table>
<thead>
<tr>
<th>County</th>
<th>Walk</th>
<th>IntraPixel</th>
<th>Train</th>
<th>Taxi</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunterdon</td>
<td>14,734</td>
<td>2,338</td>
<td>2,498</td>
<td>291,632</td>
<td>311,202</td>
</tr>
</tbody>
</table>

Figure 1: Hunterdon County trips table

Figure 2: Hunterdon County Trips pie chart

As in most of the counties, Hunterdon is dominated by Taxi trips. There is also a good spread of other trips mixed in as seen in the pie chart if you compare it to other counties as you will see with Sussex and Warren. The presence of train trips in the county give hope for a possible aTaxi station because of the concentration of people in a specific location.

Trip Generation Modeling
Using the data generated to represent individuals in NJ and their trip needs throughout the day we were able to develop code that could model the ride-sharing potential that these trip needs could generate. We tested ride-sharing under different circumstances of function for the aTaxi system. Namely, we tested 3 different variables: Departure Delay, Common Destinations, and Pixel/Super Pixel. Departure delay tested to see how a delay after a person arrived to take a taxi would affect the potential for others to show up to the stand within the delay time and potentially ride along with the first passenger. We tested 3 different departure delays, listed as DD: 60, 150, and 300 seconds. Along with departure delay we tested common destinations. A common destination that a vehicle could handle meant that it did not necessarily have to only go to the destination of the
first passenger but could also take another passenger in the taxi as long as adding their final destination would not make the first’s trip any longer than a max circuity constant, which we set at 1.2. We tested common destinations of 1, 2, 3, and 4. Pixel destination meant that only the exact pixel of the destination could be counted as the same destination between 2 passengers, whereas superpixel meant that if two destinations were within a location of + or - 1 pixel up or down, left or right could count as the same destination.

With all of these different variables we had 24 different test cases to calculate total vehicle miles accumulated by the aTaxis and the ride-sharing potential that each combination could serve. The few following pages display this data and briefly analyze it.

![AVM TOTAL]

Figure 3: Vehicle Miles totals

From figure 3 of Average Vehicle Miles we can clearly see the correlation between departure delay and common destinations and their effects on vehicle miles. As DD and CD go up, vehicle miles drop. The difference between the effects of pixel and super pixel had negligible effects on Vehicle Miles, so its benefits would certainly not outweigh the annoying cost of not being dropped off at your exact desired location pixel. One interesting thing to notice is the change between common destination 3 and 4 seems to have less of an
impact in dropping vehicle miles than it did changing from 1 to 2 or 2 to 3. This suggests that having CD at 3 may be optimal in terms of minimizing cost to customer travel but also maximizing efficiency.

Figure 4: Hunterdon County AVO Chart
Like the vehicle miles charts, the AVO can be correlated to the change in variables. As CD and DD increase the AVO increases. The effects of the super pixel were negligible. As in the vehicle miles the change between CD 3 and 4 was almost non-existent, leaving CD = 3 as the best option for maximizing efficiency and cost. These AVO’s for the entire county were pretty low, which should be expected considering that Hunterdon is not a very dense county. Most people here travel singularly and do not have many common destinations.
To get a better sense of how AVO can be affected by change in variables, we decided to take a closer look at the pixels which would be most affected by these changes. We looked at the 10 most active pixels, meaning the pixels which had the most trip departures throughout the county. These charts show us that we can reach an AVO of above 2 in the more popular pixels. This means that the possibility of implementing aTaxi stands at locations at which more demand arises can be a better way to integrate the aTaxi system. It would allow for a more efficient allocation of resources as far as taxi utilization is concerned.

Cumulative aTaxi Trips over a Day
In this plot you can right around 7 o’clock in the morning there is originally a steep increase in the departures from people leaving for work but then it is quickly followed by a steep increase in arrivals due to workers arriving to their jobs in Hunterdon. The rest of the day seems to be very standard which means that the jobs that people are working do not have specific end times because there are no steep increases after the increase at around 7.

This graph shows the imbalance of taxi size arrivals and departures for the whole day.

This graph shows the imbalance of taxi size arrivals and departures for the whole day.

Cumulative Trips in a Day of Different Vehicle Size
Vehicle Size 3
Similar to the total trips, in the size 3 vehicle the steepest increase for both the departures and arrivals is around 7 am and then it is at a steady increase for the most part after that. Looking at this plot through the lens of empty vehicle repositioning, there would need to be some sort of supply of aTaxis already there because the departures are so much greater than the arrivals.

This graph shows the net imbalance for all pixels for taxi size 3 for the day.

**Vehicle Size 6**
Taking a look at the trips of size 6 vehicles there is a change with the imbalance of arrivals and departures. Due to this change in imbalance, there would be little need to have aTaxis of size 6 in Hunterdon county because the arriving cars will account for almost all of the departing trips. A few aTaxis will be needed at the rush hour around 7am.

This graph shows the net imbalance for all pixels for taxi size 6 for the day.

**Vehicle Size 15**
The majority of these trips are between 5 and 10 am where people are arriving to work. Similar to the analysis of aTaxis of size 6, if there are enough size 15 aTaxis to cover the morning rush, then the rest of the day is covered with all of the other arrivals.

This graph shows the net imbalance for all pixels for taxi size 15 for the day.

Vehicle Size 50

Figure 12: Hunterdon County imbalance CDF for vehicle size of 15

Figure 13: Hunterdon County imbalance by time of day for vehicle size of 15

Figure 14: Hunterdon County imbalance CDF for vehicle size of 50
Similar to the last two plots, around 5 aTaxis of size 50 will need to be in stock for the morning rush and then after around 8 there are mainly only arrivals and those arriving cars can for the most part be used other places because of the small amounts of departures in the afternoon and night.

![Figure 15: Hunterdon County imbalance by time of day for vehicle size of 50](image)

This graph shows the net imbalance for all pixels for taxi size 50 for the day.

**Sussex County**

**Trip Type Comparison**

Before looking at the Taxi trips individually, we wanted to get a sense of what is actually going on in Sussex so we compared the Taxi trips to the other trips in the county as we did with Hunterdon.

<table>
<thead>
<tr>
<th>County</th>
<th>Walk</th>
<th>IntraPixel</th>
<th>Train</th>
<th>Taxi</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sussex</td>
<td>14,556</td>
<td>2,645</td>
<td>0</td>
<td>390,369</td>
<td>407,570</td>
</tr>
</tbody>
</table>

![Figure 16: Sussex County trip distribution pie chart](image)
Similar to Hunterdon, Sussex County is dominated by Taxi trips. However, in Sussex there are no train trips. It seems as though the train trips that may have been there is there was a station are instead seen as Taxi trips which is why the percentage of Taxi trips in Sussex County is greater than the percentage in Hunterdon.

**Cumulative aTaxi Trips over a Day**

![Cumulative Imbalances of aTaxis versus Time in Sussex County](image1)

Figure 17: Sussex County trip CDF

In this plot you can right around 7 o’clock in the morning, similar to Hunterdon, there is a steep increase in the departures and arrivals for the morning rush. The rest of the day seems to be at a steady increase for both the departures and arrivals. The imbalance between the departures and arrivals seem to be consistent throughout the day after 10am. This could lead to an easier empty vehicle repositioning strategy then a county in which the departures and arrivals are not increasing at the same rate.

![SUS NET](image2)

Figure 18: Sussex County trips by time of day

This graph shows the net imbalance for all pixels for all taxi sizes for the day.

**Cumulative Trips in a Day of Different Vehicle Size**

**Vehicle Size 3**
In the morning before the work rush, the departures and arrivals are roughly the same and at a relatively shallow increase. However, at around 7 am the arrivals and departures split and continue to get further and further apart as the day goes on which tells us that more people leave Sussex County than arrive. Knowing this, there is good reason why there are no train trips because Sussex County must not be a spot where many people need to go and therefore the demand for a train station is not high enough to go through the trouble of building one.

This graph shows the net imbalance for all pixels for taxi size 3 for the day.

**Vehicle Size 6**
Similar to Hunterdon County, there are more arrivals of aTaxis of size 6 than departures. This is true throughout most of the day besides what seems to be the work day. For any sort of empty vehicle repositioning process Sussex County would need some vehicles of size 6 during the work day in order to meet the demand, however, the morning and the night are covered by the arriving vehicles.

This graph shows the net imbalance for all pixels for taxi size 6 for the day.

**Vehicle Size 15**
For an aTaxis of size 15, there are two main times in Sussex County when they are needed and they are at the beginning and the end of the work day. Also, with more departures than arrivals, there needs to be aTaxis of size 15 in stock for what seems to be the whole day because the amount of arriving aTaxis will not be enough to cover the demand of the departures.
Vehicle Size 50

For aTaxis of size 50, there needs to be 8 vehicles stored overnight in Sussex County. With the relative abundance of arrivals towards the end of the day and no departures after 9 am, having enough vehicles for the morning should be very easy to accomplish. Since all of the departures in the morning take place before the arrivals, there will not be a way to get any vehicles from that day and there is no point in taking the vehicles from somewhere else if they will already be there the night before.

Figure 25: Sussex County imbalance by time of day for vehicle size of 50
This graph shows the net imbalance for all pixels for taxi size 50 for the day.

**Warren County**

**Trip Type Comparison**

Before looking at the Taxi trips individually, we wanted to get a sense of what is actually going on in Warren so we compared the Taxi trips to the other trips in the county as we did with Hunterdon and Sussex.

<table>
<thead>
<tr>
<th>County</th>
<th>Walk</th>
<th>IntraPixel</th>
<th>Train</th>
<th>Taxi</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warren</td>
<td>19,656</td>
<td>3,580</td>
<td>5,725</td>
<td>272,046</td>
<td>301,007</td>
</tr>
</tbody>
</table>

![Warren County trip distribution pie chart](image)

Figure 26: Warren County trip distribution pie chart

Contrary to Sussex and Hunterdon County, Warren has a high percentage of train and walking trips. The relative demand for aTaxis in Warren County is smaller than Sussex and Hunterdon because of the walking trips and the availability of a popular train station. If there were only a certain amount of counties that could have aTaxis, Warren should not be included because of the low demand compared to the other counties.

**Cumulative aTaxi Trips over a Day**
Warren’s imbalances are very similar to Sussex’s in the fact that the arrivals and departures are very equal up until about 7 am and then there are more departures when the morning rush begins. After the morning rush, the arrivals and departures seem to grow at roughly the same rate. This once again could be an easy imbalance to correct with having a stock at the beginning of the day that will cover the differences that are seen after 7 am.

This graph shows the net imbalance for all pixels for all taxi sizes for the day.

**Cumulative Trips in a Day of Different Vehicle Size**

**Vehicle Size 3**
Figure 29: Warren County imbalance CDF for vehicle size of 3
Once again, this plot shows that for vehicle size 3, the imbalances are very similar to Sussex in the fact that before the morning rush there is almost no imbalance, however around 7 am there is an imbalance that continually grows throughout the day. A simple stock of vehicles of size 3 would need to be present around 7 am and either it would have to grow throughout the day or be as large as the largest imbalance of the day.

Figure 30: Warren County imbalance by time of day for vehicle size of 3
This graph shows the net imbalance for all pixels for taxi size 3 for the day.

Vehicle Size 6
For aTaxis with a size of 6, there are more arrivals in the morning and at night and then during the workday there are more departures. Similar to Sussex County, if there is a way to get a fleet of aTaxis of size 6 available during the workday then the imbalance problem would be solved. Looking at the graph closer maybe the arriving aTaxis from the night before and the morning would be enough to solve the problem without having to have any aTaxis in stock.

This graph shows the net imbalance for all pixels for taxi size 6 for the day.

Vehicle Size 15
Comparing the aTaxis of vehicle size 15 in Warren County to those in Sussex county there seems to once again some similarities in the plots. The morning rush and the drive home from work times seem to have steep increases for both the departures and arrivals. However, the arrivals for Warren County are much greater after 10 am and there will be no problem finding aTaxis of vehicle size 15 if the strategy of keeping the cars overnight is used.

This graph shows the net imbalance for all pixels for taxi size 15 for the day.

**Vehicle Size 50**
The departures of the aTaxis of size 50 are very indicative of a workday. At around 7-8 am there is a large increase of trips and the same goes for 7-8pm. Similar to the aTaxis of size 15, there is a relative abundance of arriving trips which once again makes it easier for to figure out what to do with the empty vehicles. Warren county does not need to borrow any empty vehicles and can send some to other needy counties to meet the demand.

Vehicle Relocation Modeling for Hunterdon, Sussex, Warren

In order to get a better sense of how we would manage taxi movement and relocation in our counties we chose to implement a simple matching and moving strategy to model empty vehicle relocation. To do this we implemented the following algorithm:
if (demand3(x,y,t) > 0)
    time = t-1;
    count = size(data,1) + 1;
    a = x;
    b = y-1;
    if (a >= 0 && a <= 199 && b >= 0 && b <= 399)
        while (time >=1 && demand3(x,y,t) > 0)
            if (supply3(a,b,time) > 0)
                if (supply3(a,b,time) >= demand3(x,y,t))
                    data(count,1) = demand3(x,y,t);
                    data(count,2) = t;
                    data(count,3) = a;
                    data(count,4) = b;
                    data(count,5) = x;
                    data(count,6) = y;
                    count = count +1;
                    supply3(a,b,time) = supply3(a,b,time) - demand3(x,y,t);
                    demand3(x,y,t) = 0;
                else
                    data(count,1) = supply3(a,b,time);
                    data(count,2) = t;
                    data(count,3) = a;
                    data(count,4) = b;
                    data(count,5) = x;
                    data(count,6) = y;
                    count = count +1;
                    demand3(x,y,t) = demand3(x,y,t) - supply3(a,b,time);
                    supply3(a,b,time) = 0;
            end
        end
    time = time - 1;
end
end

What this algorithm is meant to do is if a pixel has a demand at a certain minute of the day, it will check to see if any taxis are remaining at that taxi stand to suffice the demand. If there are no taxis at the current station then it will check local pixels at their taxi stands to see if empty taxis are remaining there that can be moved in time
in order to satisfy demand. The taxi stand will first check pixels 1 pixel away, then 2 pixels, then 3 pixels, and the farthest they can check is 4 pixels away because this would take 4 minutes of empty travel.

Shown above is the first box of pixels checked and the final box of pixels checked.

If no pixels can be found in proximity for taxi relocation then the stand will have to get the supply from the Super Source, an imaginary source with infinite supply. This will tell us how many taxis that this method will produce and move to satisfy demand. This algorithm runs for all pixels, at all minutes of the day t = 1:1440.
Above is a sample of the data produced by the analysis. Column 1 is amount of moved, Column 2 is the time of day in minutes at relocation occurs, Column 3 and 4 is the x and y pixel that the supply is coming from, Column 5 and 6 is the x and y of the pixel that has the demand. These files will tell you at what times taxis are being relocated and from how far away. It will also enable us to know at what time of day the Super Source must be drawn from.

### Data Analysis

#### Hunterdon

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Sussex

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Warren

As vehicle size increases the percent need usage of the super source also increases, which is to be expected because there are less vehicles being used as vehicle size increases. As to be expected more taxis are drawn from the current pixel station given that it checks there first and then as the farther you get from the demand pixel the less amount of taxis have been drawn from that distance. This data enables us to realize the...
potential for the taxis to be relocated throughout the day in a way which allows us to minimize the fleet size. This along with EMR (Early Morning Relocation) would be a sufficient way to handle to allocation and relocation of empty taxis throughout the day.
This page intentionally left blank
Essex County & Morris County
Keith Gladstone and Alicia Lamb

Essex County

Fig. 1 – Essex County

The county of Essex in northern New Jersey stretches westward from the ports of Newark Bay, bordering the counties of Union and Morris to its south and west, respectively. Essex County is dominated by the City of Newark, which contains Newark Liberty International Airport, The Prudential Center -- home to the National Hockey League’s New Jersey Devils -- as well as Newark-Penn Station, a major rail hub connecting New York and New Jersey. Outside the urban centers lie a variety of suburbs that vary in socioeconomic status, many of which are classified as “commuter towns,” as residents typically travel to other towns and cities to work during the day. These towns include Montclair, the Oranges, Maplewood, and Millburn.

Morris County

Fig. 2 – Morris County

The county of Morris occupies a large, landlocked suburban and rural region in northern New Jersey, sharing its eastern border with Essex, its southern border with Union, Somerset, and Hunterdon, its western border with Warren, and its northern border with Passaic and Sussex. Its county seat is Morristown, a faster-paced town that resembles a small city that spreads out to wider-spaced quieter towns surrounded by forests and farms. Many residents of this county commute to the east, working in Newark (Essex) or as far as New York City. This is made possible in part due to the New Jersey transit rail lines that run from Manhattan and
highways like I-80 and I-287 that anchor the county to interstate road systems. The closest airport to Morris County is Newark Airport (Essex) to the east.

**Trip Lengths**

![Cumulative Distribution of Trip Lengths, Essex County](image1)

The median trip length for Essex County is approximately 7 miles, as depicted in the graph above. Because it is a more densely populated county that is closer to New York, the high density and likely resulting congestion can often translate into shorter trip lengths. Morris County has a median trip length of 12 miles, as shown in the graph below. This greater median trip length is consistent with Morris being more widespread.

![Cumulative Distribution of Trip Lengths, Morris County](image2)

**Trip Times**

184
The graph above shows the empirical cumulative distribution of trip times originating in Essex County. The median trip time is approximately 15 minutes. 90% of these taxi trips take less than 33 minutes. Hence, the majority of these trips do not take very long, and can serve as reasonable taxi requests for users of an automated fleet system. The graph below shows the distribution of trip times for Morris County. Although nearly 100% of the trips in both counties are under 100 minutes, Essex County generally creates shorter trip time lengths, which makes sense, as the distances of the trips are shorter as well.
Inter- vs. Intra- County Taxi Trips

Fig. 7 – Proportions of Intra vs Inter Trips in Morris and Essex County
The pie charts above depict the proportion of taxi trips that are within a county or travel out-of-county. The chart on the left corresponds to Essex County, with 19% of all trips being intra-county trips. The chart on the right corresponds to Morris County, with only 9% of all trips being intra-county trips. There are a greater proportion of intra-county trips in Essex County as it holds a number of predominant cities that are large sites of attraction. In Morris County, a larger proportion of trips are inter-county, meaning the trips go out of the county, indicative of its suburban population comprising people who leave Morris County more. A possible explanation for this can be attributed to the realization that lots of jobs are located in major cities, for example Newark, which is in Essex County. Hence, more people are likely to travel from where they live to Essex County than to Morris County to go to work.

**Essex County: Ride-Sharing Vehicle Miles and Average Vehicle Occupancy**

The series of orange graphs below depict ride-sharing vehicle miles and average vehicle occupancy metrics for different variants of vehicle fleet. The variants we considered were Common Destination (CD = {1,2,3,4}), Departure Delay (DD = {60, 150, 300 seconds}), and Pixel-to-Pixel (P2P) transit vs. Pixel-to-Superpixel (P2SP) transit. As the state of New Jersey was divided into the array of pixels we have used to analyze transportation thus far, we define a “superpixel” as an amalgam of adjacent pixels. Practically speaking, these superpixels would be centrally-located transit stations where taxis would drop off users, as opposed to dropping users off at their specific end destinations. This strategy, as shown in the graphs, causes lower Vehicle Miles (VM) and more efficient (higher) Average Vehicle Occupancy (AVO).

As the VM graphs indicate, increasing CD and DD (essentially sending larger vehicles with more intermediate stops) is a more efficient solution for the overall transportation system we implement. While it may inconvenience users who want small cars with personal space and perfectly direct routes, this strategy clearly keeps fewer vehicles on the road, which is better for the transportation costs, the environment and for safety. The decrease in VM is steeper when CD is increased than when DD is increased, perhaps due to the idea that having the vehicle delay its departure by a matter of minutes without increasing the number of stops doesn’t actually appeal to many more users. Rather, adding more stops on the routes to accommodate more users has a more pronounced effect on fleet efficiency (with lower VM and higher AVO). Varying from P2P to P2SP increases efficiency, as more users can share rides to the destination transit centers, rather than to specific personal destinations.
**Vehicle Miles: Sample Pixel in Essex County, NJ**

Fig. 8 – Vehicle miles in Sample Pixel for various LoS, Essex County

**Superpixel Vehicle Miles: Sample Pixel in Essex County, NJ**

Fig. 9 – Super Pixel Vehicle miles in Sample Pixel for various LoS, Essex County
Morris County: Ride-Sharing Vehicle Miles and Average Vehicle Occupancy

The series of graphs in blue depict the analogous VM and AVO dataset for Morris County. Compared to Essex County, the VM statistics are all generally higher, and the AVO is generally lower. This makes sense, as Morris County is larger, less dense and therefore dispersed in user destinations. The nature of the county makes it more difficult to share rides, and when users do, their trips are longer anyway. We do, however, see the same trends of greater efficiency when CD and DD are increased, as well as when we adjust P2P to P2SP.
Fig. 12 – Vehicle miles in Sample Pixel for various LoS, Morris County

Fig. 13 – Super Pixel Vehicle miles in Sample Pixel for various LoS, Morris County
**Essex County: Net Vehicle Animation**

This section deals with “net vehicles” in a given pixel. Arriving vehicles add to this total as departing vehicles subtract from it. Yellow indicates a higher magnitude of “arrived” vehicles or “net-positive” and dark blue indicates a higher magnitude of “departed” vehicles or “net-negative”.
The above figure shows a heat-map representing the net vehicles in Essex County at 6AM. As there is a prevalence of dark blue, that means that there are few accumulations of arrived vehicles (except for the two yellow pixels). This means that there is an abundance of departures at this time, which makes sense, as the morning commute is just beginning. The upper yellow pixel is likely in the town of Montclair, a relatively busy commuter town, and the lower yellow pixel is around the town of West Orange – also a commuter town. These towns have local train stations that take individuals into Newark and New York City.
The above figure shows a pattern that takes place during evening rush hour in Essex County. A higher concentration of yellow indicates net-positive vehicle arrivals (meaning that empty vehicles are occupying these spaces). Background knowledge of Essex County provides insight that this Southeast corner of the county is the City of Newark. This part of the county has the highest population density, and thus an accumulation of empty vehicles in this area makes sense, as more people translates to more rides.

This graphic appears very differently than the 6AM graphic that precedes it. This is likely due to the fact that accumulation throughout the day builds up in certain areas as the hours progress (as there is not much action in Essex County prior to 6AM, but there is plenty before 6PM). Below in the non-cumulative departures animation for the entire state, a pattern of higher action appears during this evening rush hour – so Essex County is consistent with the expectation for net vehicle accumulation in a busy time of day.

The same animation and analysis was conducted for Morris County. The first image (Fig. 18) below shows the county at 6 AM. There are rumblings of activity scattered throughout the region, but not much. The second image (Fig. 19) depicts Morris County at 6 PM during rush hour, when there are clear accumulations of arrivals.
more concentrated toward the center of the county in Morristown and Dover.
Beginning of Daily Surplus

When creating the following graphs, we considered Early Morning Repositioning (EMR): the idea of transporting vehicles in our fleet to different locations at midnight to reset the supply to meet each day’s demand. The following graphs of this variety depict the earliest times during the day when there are surpluses of vehicles at specific locations on the county map. As arrivals and departures occur throughout the day, we refer to net-positive arrivals as a vehicle surplus. The opposite (more departures than arrivals) would thus be called a vehicle shortage. It is useful to know when there are vehicle shortages so that we as managers can replenish the fleet stations where needed. As there are many locations throughout the state that have surpluses (as shown in the below two images), we can pull vehicles from these locations, earlier in the day than midnight, to begin to replenish the areas with shortages. Essex County has vehicle surpluses on the outskirts of the county in the suburbs, as does Morris County. To implement EMR, it would be wise to start dispatching vehicles from these surplus locations when they accumulate to serve the higher-demand areas of Newark and Morristown (both in darker blue on the respective maps).
End-of-Day Imbalances
The graphs above depict the cumulative distribution of the end-of-day imbalances in Essex County (Fig. 21) and Morris County (Fig. 22). The median end-of-day imbalance, which is the net arrival of vehicles minus the departures, is -120 in Essex County. This implies that more vehicles have departed than are currently there, signaling there is a shortage of vehicles, which leads to unfulfilled demand. In Morris County, on the other hand, the median end-of-day imbalance is only -50, so there is a greater number of vehicles that remain in use on the road.

Cumulative Arrivals and Departures by Size of Vehicle
The graphs above illustrate the cumulative arrivals and departures by size of vehicle in Essex County. The different colors are indicative of four different sized vehicles. As evident by the general slopes of the graphs, there is surplus of “blue” vehicles, the smallest size, and a shortage of all other sized vehicles. This means we can dispatch the smaller vehicles at the end of the day to serve the needs of other counties when needed.
Analogous to the graphs for Essex County, the above four graphs depict the cumulative arrivals and departures by size of vehicle in Morris County. An opposite trend in this particular county is evident. There is a shortage of “blue” vehicles, or small vehicles, but a surplus of all other sized vehicles. In order to correct this imbalance, an exchange between Morris County and Essex County can be orchestrated, and both would benefit as their respective surpluses and shortages would balance out.

### Early Morning Repositioning

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>274</td>
<td>2702</td>
<td>1</td>
<td>134</td>
<td>254</td>
<td>1303</td>
<td>3</td>
<td>154</td>
<td>248</td>
<td>1301</td>
</tr>
<tr>
<td>97</td>
<td>274</td>
<td>2702</td>
<td>3</td>
<td>135</td>
<td>259</td>
<td>1304</td>
<td>2</td>
<td>157</td>
<td>252</td>
<td>1302</td>
</tr>
<tr>
<td>97</td>
<td>276</td>
<td>2703</td>
<td>3</td>
<td>147</td>
<td>257</td>
<td>1303</td>
<td>2</td>
<td>157</td>
<td>265</td>
<td>1305</td>
</tr>
<tr>
<td>101</td>
<td>277</td>
<td>2703</td>
<td>4</td>
<td>149</td>
<td>269</td>
<td>1306</td>
<td>3</td>
<td>158</td>
<td>262</td>
<td>1304</td>
</tr>
<tr>
<td>103</td>
<td>273</td>
<td>2702</td>
<td>4</td>
<td>150</td>
<td>264</td>
<td>1305</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>281</td>
<td>2703</td>
<td>1</td>
<td>154</td>
<td>248</td>
<td>1301</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>276</td>
<td>2703</td>
<td>4</td>
<td>157</td>
<td>252</td>
<td>1302</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>262</td>
<td>2701</td>
<td>1</td>
<td>157</td>
<td>265</td>
<td>1305</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>269</td>
<td>2702</td>
<td>1</td>
<td>158</td>
<td>262</td>
<td>1304</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>158</td>
<td>263</td>
<td>1305</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 24 – EMR Fleet Statistics, Essex (Left), Morris (Right)**

The tables above correspond to the pixels in each county that have surpluses, with Essex County on the left and Morris County on the right. A linear program was implemented to find optimal fractions of vehicles to relocate to pixels with shortages. Ellie McDonald ’17 and Aaron Schwartz ’17 completed the code for this, and comma-separated files of the results of this program have been generated and saved for reference.
aTaxi Analysis for Bergen and Passaic Counties

Kathryn Jones and Jessica Zou

Introduction:
We begin the report by with a brief overview of the trips in Bergen and Passaic Counties. We gather information on the four different methods of trips: walk trips, intra-pixel trips, taxi trips, and train trips. Next, we take a look at a NJ autonomous taxi system and its implementation in Bergen and Passaic Counties. We look at distribution of vehicles on the road as well as the imbalances throughout the day. Based on our analysis in earlier parts, we look at the minimum fleet size needed for each county using only early morning repositioning.
Analysis of aTaxi Trips

<table>
<thead>
<tr>
<th>Country</th>
<th>Total aTaxi trips</th>
<th>3-Passenger vehicles</th>
<th>6-Passenger Vehicles</th>
<th>15-Passenger Vehicles</th>
<th>50+ Passenger vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergen</td>
<td>1,458,872</td>
<td>1,212,099</td>
<td>214,992</td>
<td>31,577</td>
<td>204</td>
</tr>
<tr>
<td>Passaic</td>
<td>792,407</td>
<td>685,600</td>
<td>93,480</td>
<td>13,293</td>
<td>34</td>
</tr>
</tbody>
</table>

Trips per county by aTaxi Vehicle Size

Bergen has a population of about 1,000,000 people while Passaic has a population size of about 500,000 people. The size differences can be seen when analyzing the number of daily aTaxi trips in each county. Each county has about 1.5 trips per person, however, these may not be an accurate representation of the aTaxi trips. It can be assumed that roughly 40 percent of the population, children and the elderly, does not use taxis on a daily basis. Several of the aTaxi trips may be accounted for Bergen by train stations because Bergen County has several NJ transit stations, and there is a Bergen County Line for NJ Transit. Bergen County is also the home of MetLife Stadium where the New York Giants and Jets play. The games played in the stadium should account for a large number of the aTaxi trips. Another explanation for the difference in trips between the two modes is convenience. There is a gradual rise in the number of taxi trips taken during the day as each hour between 7 am and 8 pm have a similar amount of taxi trips.

Average Vehicle Occupancy and Total Vehicle Miles

The graphs below show the average vehicle occupancies for different levels of service. P2P = pixel to pixel
P2SP = pixel to superpixel (superpixel includes the 9 adjacent pixels surrounding the destination pixel)
CD = common destinations (eg. CD = 3 means a vehicle makes at most three stops for a single trip)
DD = departure delay
From this graph, we can see that the average vehicle occupancies for Bergen County are not very high. Generally the p2p AVOs are just slightly above 1, which means that even when we try to allow vehicles to make several stops and wait, we still can’t increase the ride sharing potential much. Even with the most relaxed setting of p2sp, CD=3, DD=300, the AVO is only slightly above 1.5. Also worth noting is that increasing common distance to 4 does not increase AVO at all, because the number of rides qualifying for CD=4 is very few.

Corresponding to the AVO graph, we can see in the chart above that the total vehicle distance does not decrease much for all the p2p strategies in Bergen county. We do see a drop in total vehicle miles for p2sp strategies, especially when the common distance is increased to 3 and departure delay is relaxed to 5 minutes.
The Passaic county graph also shows that for pixel-to-pixel strategies, the average vehicle occupancies are around 1.5 at best. Interesting to note is that, for pixel-to-pixel strategies, the most important factor in increasing AVO is departure time, rather than number of common destinations. For point-to-superpixel strategies, we can see a steep increase in AVOs as we increase the number of common destinations to 3. This result makes intuitive sense. Passaic county probably has high density destination areas. Trips may not necessarily all end up in one pixel, but they will end up in the same superpixel if the area has high trip attraction.
The total vehicle miles graph further illustrates the effect of different ride sharing strategies. We can see the p2sp strategies produce significantly smaller total vehicle miles. Therefore, they provide great ride sharing potential.

Intra- and Inter-County Trips Analysis

From this chart, we can see the breakdown of inter-county trips and intra-county trips in Bergen county and Passaic county. For both counties, there are many more inter-county trips than intra-county trips. Therefore,
when considering increasing ride sharing potential and planning for aTaxi management, it is important to pay attention to inter-county trips that make up for the bulk of trips in these counties.

This chart shows the AVOs for Bergen county and Passaic county, broken down by intra-county trips and inter-county trips. From this graph, we can see that intra-county trips on average have a higher AVO than inter-county trips. Surprisingly, for these two counties, the AVOs for intra-county trips are the same and the AVOs for inter-county trips are very similar as well. The fact that intra-county trips have higher AVOs makes sense, because within county trips are more likely to start and end in the same pixel. Trips leaving the originating county are much more likely to end up in different counties and different pixels.

Combining the two graphs, we can draw the conclusion that, for Bergen county and Passaic county, there are more inter-county trips that have smaller AVOs, and less intra-county trips that have larger AVOs. In terms of tackling ride sharing and aTaxi fleet management, it will be important to keep these big picture trends in mind in order to effectively optimize our strategies.
Above are the cumulative distribution graphs of departure time for inter-county trips and intra-county trips in Bergen county and Passaic county. We can see that the trends for inter-county and intra-county trips are similar across counties. For inter-county trips, there is little activity until around 6 am, when there is a steep rise in the number of trips taken. This sharp increase could be attributed to morning rush hour traffic. For intra-county trips, the cumulative distribution shows that trips are taken throughout the day. This makes sense because intra-county trips are mostly short distance, and are not subjected to the effect of morning rush hour as much.

Supply / Demand Characteristics

We will explore the spatial imbalance of supply and demand for vehicle size of 3, 6, 15, and 50. The spatial imbalance is the net supply or demand in a pixel at the end of the day. The graphs are plotted by sorting the pixels from the most net demand to the most net supply. Extreme points are removed to better show the general trends.

In Bergen County, most of the pixels end up with a net demand for 3-passenger vehicles. This result makes sense because 3-passenger vehicles are the smallest and most flexible option. There are a lot of 3-
passenger vehicles and they are needed to transport passengers to various places in the county. There are many pixels that have a large net demand, as illustrated by the left-hand side of the graph. This shows that there are a few areas in Bergen county that people generally leave from, causing the net deficit of cars at the end of the day.

Passaic county’s 3-passenger vehicle spacial imbalance graph looks similar to Bergen country for similar reasons. However, Passaic county does not have the same kind of large net demand, showing that trips that end up in Passaic are either less or they are more likely to originate from the same places.
Bergen county and Passaic county show similar spacial imbalance trends for 6-passenger vehicles. The trend is different from that of 3-passenger vehicles probably because 6-passenger vehicles are more likely to end up in places where they are not needed. Therefore, without repositioning during the day, they end up sitting idle, resulting in net surplus in many pixels.
For 15-passenger vehicles, the spacial imbalance graphs for both counties also show there are more surpluses in general. Again, this is probably due to the fact that large vehicles are more likely to end up in pixels where they are not needed. Therefore, they result in over supply in certain pixels.
There are much fewer 50-passenger vehicles in Bergen county and Passaic county. This is because not many people go from the same origin to destination within a certain delay time. The difficulty to reuse 50-passenger vehicles once they arrive at their destinations is illustrated by either large net demands or net supplies on both ends of the spectrum.

**Imbalances Throughout the Day**

There are varying needs for vehicles of different sizes throughout each country. The following graphs are plots of the arrivals/departures throughout the day. Following each cdf of arrivals and departures, is the cumulative imbalance in the county throughout the day. Here are the arrivals/departures and imbalances throughout the day for all vehicles in Bergen County, followed by vehicles of size 3, 6, 15, and 50+ passengers. Arrivals are in red and departures are in blue. Bergen County had 739,173 arrivals and 616,034 departures. The end of day imbalances for total vehicles and vehicles of each size were about:

- Total vehicles: ~ -15,000
- 3-Passenger: ~ +1203
- 6-Passenger: ~ +10,000
- 15-Passenger: ~ +1100
- 50+-Passenger: ~ +50
We have also compiled breakdowns of arrivals/departures for each of the top ten pixels in Bergen county. It can be seen which pixels tend to have the most activity (Pixel 273, 160), and the ebbs and flows of travels in each
pixel. For example, it is fairly easy to discern the heavy work areas vs. the residential areas. Pixels with the red
departure curve rising in the later day can be assumed to be work areas. This is true for many of the top pixels in
the county. The only different pixel is pixel 264, 174 where there are more departures than arrivals throughout
much of the day.

**Passaic County**

Here is the same information for Passaic Counties. Arrivals are in red and departures are in blue. Passaic
County had 39,952 arrivals and 395,931 departures. The end of day imbalances for total vehicles and each
vehicle size were about:

- Total vehicles: ~ +2000
- 3-Passenger: ~ +9300
- 6-Passenger: ~ -6000
- 15-Passenger: ~ -1100
- 50+-Passenger: ~ 0

Unlike Bergen County Passaic County has an imbalance close to zero at the end of the day, and ends the day
with roughly the same amount of vehicles as it starts with. It turns out that there this balance is not as promising
as it seems. The vehicle breakdowns for imbalances for each vehicle size are not breakeven. This can be seen in
the numbers above and figures below.

![Taxi Arrivals/Departures](image-url)
**Taxi Imbalance Accumulation Vehicle Size 6**

![Graph of Taxi Imbalance Accumulation Vehicle Size 6](image)

**Taxi Arrivals/Departures Vehicle Size 15**

![Graph of Taxi Arrivals/Departures Vehicle Size 15](image)
We have also compiled breakdowns of arrivals/departures for each of the top ten pixels in Passaic county. It can be seen which pixels tend to have the most activity (Pixel 278, 155), and the ebbs and flows of travels in each pixel. For example, it is fairly easy to discern the heavy work areas vs. the residential areas. Most of the pixels either have large surpluses or large deficits of taxis throughout the day. The imbalances seen in Passaic County are larger than those seen in Bergen Country. This is true for many of the top pixels in the county. Pixel 271, 160 is an interesting pixel in that it rarely has a negative imbalance.

**Early Morning Repositioning and aTaxi Fleet Size**

The charts below summarize the number of active aTaxi stands for different vehicle sizes, the minimum fleet size needed to serve an entire day’s demand, and the number of aTaxis that needs to be repositioned at the end of the day in order to start the next day anew.

<table>
<thead>
<tr>
<th>aTaxi Capacity</th>
<th>Active aTaxi Stands (pixels)</th>
<th>EMR Fleet Size</th>
<th>EoD EMR, #</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Passenger</td>
<td>917</td>
<td>117,197</td>
<td>61,163</td>
</tr>
</tbody>
</table>
We can see that for Bergen county, there are many more 3-passenger aTaxis than other types of taxis. Almost half of the each kind of vehicle needs to be repositioned at the end of the day. This is a large number and shows that the EMR strategy results in a lot of wasted resources with aTaxis ending up in places where they are not needed at the end of the day.

<table>
<thead>
<tr>
<th></th>
<th>6 Passenger</th>
<th>15 Passenger</th>
<th>50+ Passenger</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMR Fleet Size</td>
<td>27,475</td>
<td>8,042</td>
<td>2,043</td>
</tr>
<tr>
<td>EoD EMR, #</td>
<td>10,618</td>
<td>3,862</td>
<td>1,884</td>
</tr>
</tbody>
</table>

Similarly, Passaic county also has the most number of 3-passenger vehicles, and about half of the aTaxi fleet needs to be repositioned for all vehicle sizes. However, the minimum fleet sizes for all types of vehicles are half of those for Bergen county. This is because Bergen county has a denser population and has more trip activity than Passaic county.
Union & Hudson Counties
Daniel Toro and Bryan Oslin

This paper/project represents our own work in accordance with University regulations.
Daily aTaxi Imbalances

The two graphs below display the daily imbalances between arrivals and departures in the two counties we studied. In doing so, we aggregated data for each pixel within the county. In the graphs, each column represents a single pixel, with the y-axis accounting for the magnitude of imbalance in the pixel. A negative imbalance indicates that more departures than arrivals are taking place and we face a vehicle shortage, while a positive imbalance hints at a vehicle oversupply. Furthermore, we aggregated all sub-counties such that we only display a graph per county, which makes analyzing data more practical.

Figure 1.a: Union County Net Imbalances

Looking at Union county in Figure 1.a, we see that imbalances are more heavily weighted toward vehicle oversupply than shortage, with a heavier tail on the right side of the graph. As such, it is more common for pixels within Union to see more arrivals than departures taking place. Looking at the graph above, one can estimate that significant negative imbalances take place in around 200 pixels, relative balanced pixels number around 50, and the remaining 170 exhibit positive imbalances. Also, negative imbalances fall below -500, while positive imbalances fall north of 2,500 at the extreme, which indicated that the oversupply of vehicles is more acute than the shortage.
In Hudson County in Figure 1.b, we see that imbalances are more evenly weighted between vehicle oversupply and shortage than for Union. Looking at the graph above, one can estimate that significant negative imbalances take place in around 60 pixels, relative balanced pixels number around 60, and the remaining 80 exhibit positive imbalances - roughly a third each. Also, unlike Union, extreme negative imbalances are at around 9,000, while positive imbalances fall north of 10,000 at the extreme. Not only are both of the magnitudes substantially higher than those at the extremes for Union, but the even magnitudes hint at vehicle shortage and oversupply being similarly acute in Hudson. We hypothesize that larger imbalances are a result of a less dense county, where around 20% fewer trips results in a less liquid market for aTaxis that results in large imbalances.

**Imbalances During The Day**

We used the following links:

http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F15/aTaxiDeparturesByCounty_3p-300-20_2015/

http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F15/aTaxiArrivalsByCounty_3p-300-20_2015/

The two graphs below display the imbalances as the day progresses, hinting at the trajectory of the imbalances that build up to arrive at the net results we saw above. In doing so, aggregated all data for each county to arrive at a single graph for each that is more helpful in drawing conclusion than attempting to conclude based on multiple graphs representing different parts of the county. The x-axis represent the percentage of the daily imbalance that has built up at that point in time, and the y-axis tracks the time of day in hours. First, we
combined all 2, 6, 20, and 50 passenger vehicles together to obtain an overall cumulative imbalance of aTaxis for each county.

In both Figure 2.a and Figure 2.b, it looks like the fractions of daily arrivals and departures taking place prior to 6am is very small. Furthermore, it is also the period during which we see wider spread between the two lines. Although this does not hint at total imbalance, it does speak to county-specific activity. For example, the proportion of Hudson county morning departures to total departures is larger than that for arrival, which is not clear in Union. One can interpret this as more people living in Hudson but working elsewhere (i.e. they depart early in morning, but less people arrive) than in Union, where more people might come to work in the morning balancing the large departures. More residential counties likely exhibit larger spread in the morning hours. Beyond that, however, it seems like the trajectory of the cumulative arrivals and departures is very close in both counties, with a very tight spread towards the end of the day as they both approach 1.
The following graphs, Figure 2.c, Figure 2.d, Figure 2.e, and Figure 2.f, show the cumulative imbalance of aTaxis throughout the day for vehicle sizes 2, 6, 20, and 50 for Union County. To get these vehicle sizes, we manually went into the file and used vehicle size 2 to represent taking 1 and 2 passengers. For vehicle size 6, we used passenger sizes of 3, 4, 5, and 6. We did similar allotting for vehicle sizes 20 and 50.

Figure 2.c: Union County trips CDF for vehicle size of 2
Figure 2.d: Union County trips CDF for vehicle size 6

Figure 2.e: Union County trips CDF for vehicle size 20
From the above figures 2.c-2.f, one of the first trends that stands out is how the larger vehicle sizes exhibit much less activity in the early hours, with the imbalance proportion close or at zero. It also seems to be the case that larger vehicles, particularly in the case of departures, are done earlier in the day (i.e. the last graph hits a proportion of 1 for departures at around 8pm, versus around 11pm for the first graph). This is not surprising, as large vehicles are mobilized during more active hours, with fewer people needing transportation late at night or early in the morning.

The following graphs, Figure 2.g, Figure 2.h, Figure 2.i, and Figure 2.j, show the cumulative imbalance of aTaxis throughout the day for vehicles sizes 2, 6, 20, and 50 for Hudson County.
Figure 2.g: Hudson County trips CDF for vehicle size 2

Figure 2.h: Hudson County Trips CDF for Vehicle Size 6
The same trend described in Union applies for Hudson, the largest difference being the gap between arrivals and departures in large vehicle sizes. The bulk of departures of large vehicle sizes happens in the morning hours,
likely representing all the people who go to work. This is congruent with our previous conclusion that it is likely that more people live in Hudson on work elsewhere than the case for Union county, where maybe more people work in the county or come from other counties to work at. They display very different demographic behaviors in that sense, and the above trend helps build and strengthen the previous conclusion.

**Pixel Analysis**

Using the same data above for each county, we analyzed the pixels to find the one at any point in time that either had the largest net supply, largest net demand, or the largest most balanced.

**Hudson County**

Shown in Figure 3.a, pixel (156,257) had the largest net supply of all the pixels in Hudson County at 2,646.

![Figure 3.a: Hudson County largest supply at any point in time](image)

Shown in Figure 3.b, pixel (167, 252) had the largest net demand of all the pixels in Hudson County at 739.

![Figure 3.b: Hudson County largest demand at any point in time](image)
Figure 3.b: Hudson County largest demand at any point in time

Shown in Figure 3.c, pixel (170, 259) had the largest most balanced pixel of all the pixels in Hudson County.

Figure 3.c: Hudson County most balanced

Union County

Shown in Figure 4.a, pixel (128,237) had the largest net supply of all the pixels in Union County at 2,789.
Figure 4.a: Union County largest supply at any point in time

Shown in Figure 4.b, pixel (151, 243) had the largest net demand of all the pixels in Union County at 472.

Figure 4.b: Union County largest demand at any point in time

Shown in Figure 4.c, pixel (144, 246) had the largest most balanced pixel of all the pixels in Union County at 0.
Lastly, we wanted to graph the activity in terms of vehicles circulating at different times of the day. Below, the graphs help visualize that activity, which is fairly similar in both counties. In Figures 5.a and 5.b, it seems like the hours between 1am and 5am are least active, with 4pm to 9pm being most active. This makes sense, as the evenings should show most congested roads as people go back home from work. One could argue similarly in the morning, but this is spread out over two different modes given that people arrive at work at different times. There is an early-bird population going to work between 6 and 9am, and another spike in activity between 11am and 1pm when people go to lunch and arrive at later work shifts - helping explain the midnight activity. As stated above, the conclusions on both counties are very similar, which is expected given similar demographics.
EMR Analysis

Using the following data
(http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F15/EarlyMorningRepositioning_Orf467F15/) and the Level of Service strategy of common destination 3, departure delay 300 seconds, and maximum circuity of 20%, early
morning repositioning algorithm produced results that ultimately show how to move these aTaxis from their final position to another position. Therefore, we were able to see for each county and for each of the four types of vehicle sizes how many of each were either needed to be imported or exported into/out of each county for the following day.

The following table shows that necessary import/export number for each type of vehicle size for each county where the negative number represents the number of vehicles of that size that can be exported to another county that needs vehicles and the positive number represents the number of vehicles of that sizes that needs to be imported into each county from others/super source for the following day’s activities.

<table>
<thead>
<tr>
<th></th>
<th>Union</th>
<th>Hudson</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 PAX</td>
<td>Active aTaxi Stands</td>
<td>EMR Fleet Size</td>
</tr>
<tr>
<td></td>
<td>422</td>
<td>65,904</td>
</tr>
<tr>
<td>6 PAX</td>
<td>438</td>
<td>23,528</td>
</tr>
<tr>
<td>15 PAX</td>
<td>409</td>
<td>5,289</td>
</tr>
<tr>
<td>50 PAX</td>
<td>275</td>
<td>562</td>
</tr>
</tbody>
</table>

Table 1: EMR analysis for Union and Hudson County

From Table 1, Union County needs to start the day with 65,904 3-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 85,907 3-passenger aTaxis sitting around at the end of the day. Therefore, it needs to export 20,003 3-passenger aTaxis to be able to serve a repeated day without wasting resources and allowing other counties to do the same.

From Table 1, Union County needs to start the day with 23,528 6-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 25,976 6-passenger aTaxis sitting around at the end of the day. Therefore, it needs to export 2,448 6-passenger aTaxis to be able to serve a repeated day without wasting resources and allowing other counties to do the same.

From Table 1, Union County needs to start the day with 5,289 15-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 8,188 15-passenger aTaxis sitting around at the end of the day. Therefore, it needs to export 2,899 15-passenger aTaxis to be able to serve a repeated day without wasting resources and allowing other counties to do the same.

From Table 1, Union County needs to start the day with 562 50-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 709 50-passenger aTaxis sitting around at the end of the day. Therefore, it needs to export 147 50-passenger aTaxis to be able to serve a repeated day without wasting resources and allowing other counties to do the same.

From Table 1, Hudson County needs to start the day with 30,672 3-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 68,052 3-passenger aTaxis sitting around at the end of the day. Therefore, it needs to export 37,380 3-passenger aTaxis to be able to serve a repeated day without wasting resources and allowing other counties to do the same.

From Table 1, Hudson County needs to start the day with 43,452 6-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 12,748 6-passenger aTaxis sitting around at the end of the day. Therefore, it needs to import 30,704 6-passenger aTaxis to be able to serve a repeated day.

From Table 1, Hudson County needs to start the day with 6,511 15-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 8,027 15-passenger aTaxis sitting around at the end
of the day. Therefore, it needs to export 1,516 15-passenger aTaxis to be able to serve a repeated day without wasting resources and allowing other counties to do the same. From Table 1, Hudson County needs to start the day with 3,290 50-passenger aTaxis in order to make it through the day without having to reposition any, but ends up with 2,589 50-passenger aTaxis sitting around at the end of the day. Therefore, it needs to import 701 50-passenger aTaxis to be able to serve a repeated day.

Further EMR Analysis:

<table>
<thead>
<tr>
<th>Union County</th>
<th>Hudson County</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Passenger Car</td>
<td>5 Passenger Car</td>
</tr>
<tr>
<td>Total Empty Vehicle Miles</td>
<td>47,565</td>
</tr>
<tr>
<td>Average Empty Miles Per Car</td>
<td>2.38</td>
</tr>
<tr>
<td>Total Loaded Vehicle Miles</td>
<td>592,473.31</td>
</tr>
<tr>
<td>Average Loaded Vehicle Miles</td>
<td>6.90</td>
</tr>
<tr>
<td>3 Passenger Car</td>
<td>5 Passenger Car</td>
</tr>
<tr>
<td>Total Empty Vehicle Miles</td>
<td>748,957</td>
</tr>
<tr>
<td>Average Empty Miles Per Car</td>
<td>20.30</td>
</tr>
<tr>
<td>Total Loaded Vehicle Miles</td>
<td>1,020,797</td>
</tr>
<tr>
<td>Average Loaded Vehicle Miles</td>
<td>15,000,242.99</td>
</tr>
</tbody>
</table>

Simple aTaxi Management Strategies

Jack Rogers & Ryan Slattery

Introduction
In the previous chapter, Zhu & Ku detailed their methodology for sizing a statewide aTaxi fleet for New Jersey. A natural progression from this step, however, is to ascertain if there are vehicle management strategies that one can employ to lower transportation costs and increase the operational efficiency of a statewide aTaxi fleet. More specifically, we will address the following question in our work: Are there any simple, “back-of-the-envelope” strategies that can be implemented to get an initial idea of how one should reposition vehicles over the course of a day (i.e. between 0:00 and 23:59)? The first part of this chapter will introduce the simplest matching algorithm used in our analysis. The second part will detail the results.

Simplest Matching Algorithm

Introduction of Algorithm
A couple of assumptions are made in the employment of our simplest matching algorithm. Spatially, the state of New Jersey is broken up into a width of 200 pixels and length of 400 pixels (i.e. spatial coordinates such that 0 ≤ x ≤ 199 and 0 ≤ y ≤ 399). Temporally, time is broken up into one-minute increments between 0:00 and 23:59.

The goal of this algorithm is to assign available supply to serve demand, by progressively assigning demand at time t. Two possible cases exist for assigning aTaxis to pixels. First, if an aTaxi lies within a specific proximity of a pixel within a specific timeframe before the demand can be met, then aTaxis are assigned to that pixel to satisfy demand. If this demand cannot be satisfied by aTaxis in the neighboring area, however, aTaxis must be drawn from the “superSource” in order to satisfy demand.
Algorithm
The simplest matching strategy involves looking at pixel (x, y) at time t, and observing the supply and demand of aTaxis at that time. If there exists a demand for an aTaxi, then one examines the current pixel to see if there existed a supply of aTaxis at times \( t - 4 \), \( t - 3 \), \( t - 2 \), and \( t - 1 \) that could be assigned to the pixel’s demand at time \( t \). If demand is still not satisfied, then one checks the 3 x 3 perimeter of pixels surrounding the given pixel to see if supply from \( t - 4 \), \( t - 3 \), \( t - 2 \), and \( t - 1 \) can meet demand at pixel (x, y) at time \( t \). If demand is not met, then one observes the 3 x 3 perimeter around pixel (x, y) to see if supply from \( t - 4 \), \( t - 3 \), and \( t - 2 \) can meet demand at pixel (x, y) at time \( t \). If demand is still not met, then one observes the 3 x 3 perimeter around pixel (x, y) to see if supply from \( t - 4 \), \( t - 3 \), \( t - 2 \), and \( t - 1 \) can meet demand at pixel (x, y) at time \( t \). If demand is still not met, then one observes the 5 x 5 perimeter around pixel (x, y) to see if supply from \( t - 4 \), \( t - 3 \), and \( t - 2 \) can meet demand at pixel (x, y) at time \( t \). If demand is still not met, then one observes the 5 x 5 perimeter around pixel (x, y) to see if supply from \( t - 4 \) and \( t - 3 \) can meet demand at pixel (x, y) at time \( t \). Finally, if demand is still not met, then one observes the 7 x 7 perimeter around pixel (x, y) to see if supply from \( t - 4 \) can meet demand at pixel (x, y) at time \( t \). If demand is still not met after these steps, then aTaxis are called from the superSource to satisfy the remaining demand at pixel (x, y) at time \( t \). A visualization of the above algorithm is shown below, courtesy of Professor Kornhauser:
Furthermore, the order of movement through pixels for each perimeter are shown above:

There are two important notes to mention about the order of checking pixels. First, the order is based on the cost functions related to moving aTaxis from one pixel to another, and are shown in the above visualization of the algorithm. Second, the order in which the corners are checked is due to the order in which the simplest matching algorithm runs through the pixels. Starting with the pixel at the very bottom left of the 200 x 400 pixel representation of New Jersey, the algorithm works through the pixels by moving from left to right, and then moving to the left-most pixel of the above row and moving left to right again.

Results

Empty Vehicle Repositioning Graphs

The first order of business is to determine a minimum fleet size to service supply and demand across New Jersey pixels on a minute-by-minute basis over the course of a day. Assuming an initial fleet size of zero at 0:00, we used the simple matching algorithm described above to determine the number of empty vehicles that would be required to accomplish repositioning over the course of a day. Since the fleet size is assumed to be zero at the start of Day 1, the simple matching algorithm provides us with the number of aTaxis that would be called from the superSource on Day 1, based on the supply at the end of Day 1.

After running the simplest matching algorithm for the 50-passenger, 15-passenger, and 6-passenger data, we first generated graphs to get an idea for the patterns in which total aTaxis would be called for repositioning throughout the day, versus those aTaxis called specifically from the superSource. In the graphs below, time is on the x-axis (from 0 to 24 hours), and the number of empty vehicles repositioned is on the y-axis. The red line denotes the total cars called for repositioning at any given time, and the blue line denotes the cars specifically called from the superSource:
Fig. 3 – 50 Passenger Scenario, Repositioning

The above graph demonstrates that an enormous spike in total aTaxis occurs between 7 am and 8 am, which is reasonable given morning commutes and trips to school. These 50-passenger aTaxis likely stay in the vicinity of pixels that require such services for the rest of the day, which likely explains why there is no significant spike in aTaxis called from the superSource for the duration of the day.
Once again, a large spike in total cars and cars from the superSource is evident between 7 am and 8 am. More noticeable, however, are the spikes in total cars and superSource cars in the afternoon between 2 pm and 6 pm. It is probable that some of these trips are accounted for by pixels where academic, recreational, and entertainment-based services are found (as observed in the MyCity Project).
6-Passenger Case

The 6-passenger case is similar to the 50-passenger and 15-passenger cases in terms of the spike of total cars and superSource cars called during the morning rush period of 7 am to 8 am. The heavy spikes between 4 pm and 8 pm, however, stand out noticeably from the previous cases. Similar to the 15-passenger scenario, this trend is likely explained by trips that originate from SRE (school-recreation-entertainment) locations. It is more realistic to see a family of six or fewer going out to dinner and taking a ride home, or six or fewer teammates sharing a ride home from soccer practice, than it would be for a group of seven to fifteen people to share a ride from SRE originations.

Comparison of EMR and Local Approaches

The goal of this chapter is to determine whether the local approach that is employed via the simplest matching algorithm leads to lower transportation costs than the early morning repositioning approach taken by Zhu & Ku. We determined the necessary fleet size by applying Day 1 supply/demand pixel data to the simplest matching strategy. This algorithm also enabled us to calculate the total number of repositioning miles required over the course of a day. We determined the end-of-day repositioning of cars using the gravity model implemented by Zhu & Ku, using the EoD supply (where the cars were located at the end of the day) and the BoD demand, measured by the calls to the superSource throughout the day. The EoD repositioning miles were calculated by multiplying our generated trip array (which tells us the trips that occur between all pixels in NJ) by the distance
between respective pixels. Essentially, this distance signifies the distance of all aTaxis from where they were in the EoD supply production vector to wherever they were moved in the attraction vector.

Summary results for the 50-passenger, 15-passenger, and 6-passenger cases are shown in the tables below. “EMR” denotes the results from Zhu & Ku’s analysis in the previous chapter, and “Local” denotes the results from analysis performed in this chapter. The addition of fleet size, during the day repositioning miles, and EoD repositioning miles/car can give us an idea of whether the strategy employed in this chapter is superior to the one employed by Zhu & Ku in the previous chapter:

### 50-Passenger Case

<table>
<thead>
<tr>
<th></th>
<th>Fleet Size</th>
<th>During the Day Repositioning (miles)</th>
<th>End-of-Day Repositioning (miles)</th>
<th>End-of-Day Repositioning (cars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMR</td>
<td>13,443</td>
<td>0</td>
<td>108,470</td>
<td>10,140</td>
</tr>
<tr>
<td>Local</td>
<td>8,179</td>
<td>8,250</td>
<td>86,550</td>
<td>6,670</td>
</tr>
</tbody>
</table>

### 15-Passenger Case

<table>
<thead>
<tr>
<th></th>
<th>Fleet Size</th>
<th>During the Day Repositioning (miles)</th>
<th>End-of-Day Repositioning (miles)</th>
<th>End-of-Day Repositioning (cars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMR</td>
<td>93,031</td>
<td>0</td>
<td>967,450</td>
<td>58,894</td>
</tr>
<tr>
<td>Local</td>
<td>53,136</td>
<td>53,343</td>
<td>735,509</td>
<td>35,067</td>
</tr>
</tbody>
</table>

### 6-Passenger Case

<table>
<thead>
<tr>
<th></th>
<th>Fleet Size</th>
<th>During the Day Repositioning (miles)</th>
<th>End-of-Day Repositioning (miles)</th>
<th>End-of-Day Repositioning (cars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMR</td>
<td>391,787</td>
<td>0</td>
<td>4,497,714</td>
<td>216,348</td>
</tr>
<tr>
<td>Local</td>
<td>589,263</td>
<td>1,301,119</td>
<td>13,597,243</td>
<td>450,915</td>
</tr>
</tbody>
</table>

A sum of costs from fleet size, during-the-day repositioning, and EoD repositioning demonstrates whether or not the Local strategy wins out over the EMR strategy in each particular case. There is some fixed cost for each car pulled from the superSource, so calculating costs from fleet size is a trivial matter. Without a formal cost function of repositioning miles, however, we are unable to sum during-the-day and EoD repositioning costs to fleet size costs to calculate a total cost. We can eyeball the results in the above tables, however, to get an idea of how the two methodologies compare. For the 50-passenger and 15-passenger cases, it is clear that local repositioning results in a smaller overall fleet size and fewer total repositioning miles (summing during-the-day and EoD miles) than early morning repositioning, which confirms that local repositioning is a superior strategy. The six-passenger case complicates the picture, however, as the required fleet size and the repositioning miles in the local repositioning strategy exceed the corresponding results under the early morning repositioning strategy. We conclude that this result is plausible, however, due to the sheer number of trips that occur in a 6-passenger scenario versus the 50-passenger and 15-passenger cases. In the 6-passenger case, trips can leave from far more pixels, which will inevitably require a significantly higher number of cars to be called from the superSource. With this larger number of cars, far more opportunities will also arise to implement the repositioning strategy over the course of the day, which leads to about 1.3 million repositioning miles in a day. Even with the greater number of cars and during-the-day repositioning miles, though, the vast number of pixels
from which 6-passenger trips can originate (homes, school/work, SRE, etc.) leads to larger EoD repositioning costs as well. Further exploration into the simple matching algorithm versus early morning repositioning might look to see at what size passenger vehicles begin to favor one of these approaches over the other, for the sake of further exploration and potential implementation in the future.

Optimal aTaxis Management Strategies

Jay Karandikar
Jordan Radke
Kate Ju
Marlon Sabo

Objective:
In this project, we will focus on the effectiveness of search depth in optimizing aTaxis allocation in New Jersey. Search depth is referring to the length of time or data we will search to allocate the future aTaxis in demand. Our input data is organized by every minute and we will test for four different search depths: 4, 6, 8, and 10. That is, we will consider, say for search depth 4, past four minutes of demand and supply in allocating future aTaxis in demand. We will determine the minimum number of vehicles needed to meet the demand. Our hypothesis is that as we increase the search depth, the number of total vehicles needed will decrease. Furthermore, we believe that there is a “threshold” search depth, an optimal search depth that will result in optimized fleet size in a reasonable optimizing time. In other words, we believe that as we increase the search depth, the program run time or the optimizing time or cost will also increase; the decrease in total fleet size from increasing search depth will be minimal at a certain search depth.

Algorithm and Method:
This algorithm reads in arrays that describe the supply and demand for taxis at each point in time and space in New Jersey. That is, at each point (i,j,t) in the pixelization of New Jersey throughout the day (divided into 1440 minute intervals), the supply array records how many taxis are incoming to a pixel, and the demand array records how many are outgoing. The goal of our algorithm is to take this input and meet the demand at each pixel throughout the day with the fewest number of taxis possible.

First, for each coordinate (i,j) in the New Jersey pixelization at a given time t, we can create an array that records how many taxis there are currently at a stand at each coordinate (i,j,t). Then, we travel through the pixelization at each time interval and meet the demand there. But, before drawing from the supersource to do it--that is, generating a new taxi and sending it to that pixel from some general taxi yard--we call a function called TaxiSearch. TaxiSearch tries to find unoccupied taxis that have already fulfilled their trip, and are currently just sitting at another stand. Since our pixels are ½ mile wide and we assume taxis can travel at 30 m.p.h, if we
search r time steps into the past, we can feasibly draw from any taxi stand within an r-pixel radius of (i,j,t). If this still doesn’t satisfy demand at our pixel, then we draw from the supersource. We ran the algorithm for varying levels of search radius to see if the optimal fleet sizes converged to any minimum as we increased the search’s depth.

By recording when and where we have to draw from the supersource, we can find out: a lower bound on the size of the taxi fleet we need; a distribution in time of the number of taxis we’re drawing from the supersource throughout the day; and the number of taxis each specific pixel for the whole day. With this last piece of information, we can modify and reposition the fleet at, say, midnight so that each pixel has the taxis it needs to get through the next day.

**Results and Analysis:**

<table>
<thead>
<tr>
<th>Search Depth</th>
<th>3 Passengers</th>
<th>6 Passengers</th>
<th>15 Passengers</th>
<th>50 Passengers</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>7,831,886</td>
<td>4,249,046</td>
<td>94,925</td>
<td>11,851</td>
</tr>
<tr>
<td>6</td>
<td>7,830,027</td>
<td>4,216,423</td>
<td>88,789</td>
<td>10,572</td>
</tr>
<tr>
<td>8</td>
<td>7,829,146</td>
<td>4,200,182</td>
<td>85,534</td>
<td>9,574</td>
</tr>
<tr>
<td>10</td>
<td>7,828,181</td>
<td>4,189,716</td>
<td>82,827</td>
<td>9,045</td>
</tr>
</tbody>
</table>

Table 1: Number of Total aTaxis Supersourced for a Day by Various Fleet Sizes

**Case #1: 3 Passengers**

![Figure 1: Total Vehicle Size for 3 Passengers by Search Depth](image)

Total number of vehicles needed for a day in NJ for search depth 4, 6, 8, and 10 are 7,831,886, 7,830,027, 7,829,146, and 7,828,181, respectively. Despite of the increase in the runtime of the code, the number of aTaxis decreased by 1,859, 881, and 965 in the order of increasing search depth. Compared to the
total number of vehicles, the decrease in vehicles is close to zero percent. That is, despite of the increase in search depth, aTaxis allocation did not improve significantly for 3 Passengers vehicles.

The graph below overlays the vehicle distribution from the supersource over the day for four different search depths. (Individual vehicle distribution graphs follow afterwards.) Due to a relatively small difference in total vehicles, the distributions of the four different search depths are very similar. The difference among the four graphs is only visible from a very close look. Notice that the y-axis represent the number of vehicles from the supersource at the moment. The y-axis values represent the number of vehicles added to the aTaxis system because currently occupied vehicles cannot meet the future demand on time due to the speed constraint and their past locations. The area under the curve represents the total number of vehicles in the aTaxis system.

Since this is a 3 passenger vehicle, there is demand throughout the day from the midnight. At the beginning of the day around midnight, approximately 2,600 vehicles are added to the system. In the morning, including the morning rush hour period, the number of new vehicles added to the system from the supersource remains as low as approximately 900 and as high as 7,500. The demand for aTaxis increase during the mid-day. During the evening rush hour, the number of vehicles from the supersource increases greatly. Interestingly, it also has five peaks in the late afternoon and evening rush hour. Possible cause of the peaks can be surge of demand from unpopular areas during the rush hours. Another interesting pattern is shown at night when more vehicles are keep added to the system. Considering that the demand after 8 p.m. is as large as during even rush hours, most of the vehicles from the supersource is because current vehicles cannot make it to the future demand on time. Considering that a large number of vehicles are still added during the low demand period, there may be a pattern in demand over time and a great potential for repositioning to optimize the fleet allocation further.
Figure 3: 3-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 4)

Figure 4: 3-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 6)
Case #2: 6 Passengers

Figure 5: 3-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 8)

Figure 6: 3-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 10)

Figure 7: Total Vehicle Size for 6 Passengers for a Day by Search Depth
The graphs of the number of vehicles called from the supersource are below. They start out at about 1000 aTaxis at midnight. There is then a spike of about 600 around the morning rush hour, followed by a smaller spike of about 2000. This followed by a fairly consistent period of demand of around 4000-6000 aTaxis around lunchtime. Finally, there are five spikes between 10,000 and 18,000 during the evening commute. This pattern is similar to the pattern followed by the 3-passenger vehicles.
Figure 10: 6-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 6)

Figure 11: 6-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 8)
Case #3: 15 Passengers

Total number of vehicles needed for a day in NJ for search depth 4, 6, 8, and 10 are 94,925, 88,789, 85,534 and 82,827 respectively. Despite of the increase in the runtime of the code, the number of aTaxis decreased by a decreasing number in the order of increasing search depth, showing that eventually it will converge onto a value that would be optimal for all depths of search. Compared to the total number of vehicles, the decrease in vehicles is close to zero percent due to their inverse relationship.
The graph above is a superposition of all the graphs below which show the number of vehicles that are used from the Supersource by minute throughout the day. aTaxis are pulled from the Supersource when there aren’t enough aTaxis on the road to meet the current demand at the respective time of day given our speed constraint with how quickly aTaxis should respond. The distributions of the 4 different search depths do not vary by extreme values and so the superpositioned graph looks like one even curve. If you take the area under each of the curves you will get total number of aTaxis that are current in use at the given time.

For all the 15 passenger aTaxi search depths the demand starts at exactly 0 at midnight and doesn’t really pick up until around 6am where it spikes up to its max size over the entire day at around 1500 vehicles being pulled from the Supersource at around 7:30am which makes sense because all the people are on their morning commute to work. Then it drops to just above 0 till around 2 where it spikes a couple times within a span of 4-5 hours to highs of around 550 vehicles being drawn from the Supersource to meet demands. Later throughout the day after around 8pm the draw from the source gradually decreases back to 0. You can see that the larger aTaxis are in highest demand at the beginning of the work day and at the end of the work day since larger amounts of people will be going to and from work at the same time and in larger groups.

Compared to the 3 and 6 passenger aTaxis, the larger aTaxis are more in demand for the beginning and end of the work day compared to mostly being in demand throughout the end of the day and later into the evening.
Figure 15: 15-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 4)

Figure 16: 15-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 6)

Figure 17: 15-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 8)

Figure 18: 15-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 10)

Case #4: 50 Passengers
Total number of vehicles needed for a day in NJ for search depth 4, 6, 8, and 10 are 11,851, 10,572, 9,574 and 9,045 respectively. Despite of the increase in the runtime of the code, the number of aTaxis decreased by a decreasing number in the order of increasing search depth, showing that eventually it will converge onto a value that would be optimal for all depths of search. Compared to the total number of vehicles, the decrease in vehicles is close to zero percent due to their inverse relationship. The 50 passenger vehicles were extremely low in numbers compared to the rest of the taxis due to the rarity of their demand.

The graph above is a superposition of all the graphs below which show the number of vehicles that are used from the Supersource by minute throughout the day. aTaxis are pulled from the Supersource when there aren’t enough aTaxis on the road to meet the current demand at the respective time of day given our speed constraint with how quickly aTaxis should respond. The distributions of the 4 different search depths do not vary by extreme values and so the superpositioned graph looks like one even curve. If you take the area under each of the curves you will get total number of aTaxis that are currently in use at the given time.
For all the 50 passenger aTaxi search depths the demand starts at exactly 0 at midnight and doesn’t really pick up until around 6am where it spikes up to its max size over the entire day at around 370 vehicles being pulled from the Supersource at around 7:30am which makes sense because all the people are on their morning commute to work. Then it drops to just above 0 till around 2 where it spikes a couple times within a span of 4-5 hours to highs of around 40 vehicles being drawn from the Supersource to meet demands. Later throughout the day after around 8pm the draw from the source gradually decreases back to 0. You can see that the larger aTaxis are in highest demand at the beginning of the work day and at the end of the work day since larger amounts of people will be going to and from work at the same time and in larger groups.

The difference between the 15 passenger aTaxi’s and the 50 passenger aTaxi’s is obviously the demand in numbers drawn from the Supersource but for the 50 passenger aTaxi the demand is only high for around 7:30am then for the rest of the day the 50 aTaxi is mostly non-existent except for a handful being sent out for the end of the work day.

Figure 21: 50-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 4)

Figure 22: 50-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 6)
Figure 23: 50-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 8)

Figure 24: 50-Passenger Vehicles Traveling in NJ by Time of Day (Search Depth 10)
Conclusion and Next Steps:

In our project, we tried to estimate the effectiveness of the search depth in optimizing the allocation of aTaxis with New Jersey data. We found that increasing the search depth does not decrease the fleet size by much. We also found that changing the search depth does not change much the temporal distribution of when new aTaxis are called from the supersource. We also could not confirm the “threshold” search depth theory. While for all passenger types, the decrease in total fleet size decreased as the search depth increased close to 10. We do not know our computational capability and costs to estimate the “threshold” search depth, but there seem to be room for optimization as we increase the search depth; we still expect the decrease in fleet size to decrease even more as we increase search depth beyond 10.

There are several ways that next year’s class can build upon our results.

1) Increase the search depth beyond 10. Yet, as we have seen from our modeling, the effectiveness of increasing search depth will be less influential than other factors such as repositioning and costs.
2) Treat the empty vehicle repositioning problem as a linear programming problem and to use a linear programming solver, such as CVX in MATLAB or AMPL, to find the optimal solution. In this repositioning problem, include an analysis of the costs and only reposition an empty the vehicle if the benefits outweigh the costs.
3) Experiment with different kinds of repositioning paradigms. For example, they could reposition every hour, or every day at midnight. We suspect that doing this would significantly reduce the fleet size.
4) Include the cost constraint in the cost of the vehicle and the storage cost overnight or unused time. In our project, we added a new vehicle from the supersource, if current vehicles in the system cannot make it to the future demand on time under the speed constraint. Including the cost constraint of adding a new vehicle or parking costs overnight or during unoccupied time can make our model optimize the vehicle distribution in a realistic fashion.
5) Analyze how fleet size would be affected if aTaxis were only allowed to serve passengers from their own county.
6) Vary the speed of aTaxis. In our analysis, we assumed that the aTaxis would be travelling at 30 miles per hour. While safety is one of our main concern in aTaxis system, 30 miles per hour is a fairly low speed. Changing this parameter will expand the distance that a vehicle can meet demand and decrease the number of total fleet size.
Stochastic Empty aTaxi Management

Artur Filipowicz and Max Bressler

In this project, we model the distribution of New York City taxi trips with a real dataset. Then we implement four strategies for aTaxi vehicle management, with important results and insights for an urban aTaxi system.

This paper represents our own work in accordance with University regulations

/s/ /s/
Artur Filipowicz Max Bressler

Overview

This project contributes to the New Jersey statewide aTaxi system design by analyzing real taxi data. An autonomous taxi system must account for uncertain demand, and we use a taxi dataset to observe spatial and temporal distributions of taxi demand. Armed with this data, we are able to design stochastic taxi management strategies that reduce the aTaxi fleet size, allow aTaxis to make smarter routing decisions, and in general model what an autonomous taxi system would look like in the real world. In this report, we first analyze and identify various characteristics of taxi demand. Later in the report, we act upon this information to deliver aTaxi management strategies that can account for real-time, stochastic demand. The analysis and aTaxi system implementation serve as an important building block for a nationwide aTaxi system.
Description of Dataset

In order to model stochastic trip demand, we use a dataset containing 1.1 billion New York City taxi trips from January 2009 to June 2015. The dataset and accompanying resources are available at https://github.com/toddwschneider/nyc-taxi-data. The subset of this data most relevant to our analysis is Yellow Taxi trips, which are focused on the Manhattan area. This part of the dataset contains both spatial and temporal columns for the purpose of constructing a stochastic model of trip demand. Temporally, each trip contains a pickup time and a dropoff time, with high precision (at the second level). Spatially, each trip contains the longitude and latitude of pickup and dropoff, allowing for an analysis of taxi movement across the city. In addition, the data contains the number of passengers traveling in a taxi. The only missing ingredient from this dataset is the “medallion” of each taxi. With this information, we can analyze the movement of individual taxis in the New York City area, which would give us direct access to individual vehicle patterns that already exist in Yellow Taxis. However, with the data that we have, we are still able to model the distribution of trip departures and trip arrivals, both temporally and spatially. With the number of passengers per taxi, we can also model total trip demand at each location and at each time.

The biggest utility of the New York City taxi dataset for a Vehicle Management Process is that it allows for the modeling of taxi trip demand, which in turns leads to strategies for empty vehicle repositioning. One may argue that the New York City taxi trip distribution is not representative of statewide or nationwide taxi trip distributions. However, the New York City dataset allows us to formulate a process for a vehicle management system, from modeling stochastic demand to implementing corresponding strategies that match this demand with a fleet of aTaxis. By analyzing this dataset, we are able to find out what kinds of scenarios and challenges that an aTaxi system must face. In the next sections, we examine how we can use this dataset to inform us about different parts of a taxi management system, the first of which includes placement of taxi stands.

New York Taxi Trips Infrastructure

Before proceeding, we need to briefly describe the mechanism underlying the trip data. The original data comes in a comma separated value format. There is one file for each month. Since each month has around 15 million trips, the size of the file is about 1 Gb. This is too cumbersome to work with since load time is long and it is easy to run out of memory when working with more than one month. Separating the data into smaller files is also problematic, because loading many file takes a long time as well.

Our solution was to use SQLite, a local database. This increase our read and write time and allowed us to do queries and sorts on the data more easily. The database if made up of 168 tables, one table for every hour. Data from the original datafiles is sorted into the tables. The column headers are listed below for reference. One potential improvement in this area would be to use a standalone SQL server. This might improve performance even more since data queries would be executed by another process.

TripStartTime
Pickup_longitude
Pickup_latitude

260
The first step in designing an autonomous taxi system is the placement of taxi stands. In New York City’s taxi system, there are no taxi stands - taxis can drop off and pick up passengers at many places along a street, essentially allowing “infinite” dropoff and pickup locations. An autonomous taxi system, however, includes the use of a finite number of “taxi stands” where departing passengers will congregate and arriving passengers will be dropped off. The organization of taxi stands allows for places for autonomous taxis to be serviced, and the distribution of taxi stands allows a vehicle management system to better keep track of demand over the course of a day.

Pixelated Stands

One potential organization of aTaxi stands is the placement of stands at the centroids of a pixelated grid. A map of New York City is pixelated into various regions of the same size, one-half mile by one-half mile (we
use this pixel size for comparison with the New Jersey taxi system). Passenger arrivals can then be modelled by the pixel that they arrive in, as passengers will go to the taxi stand in the pixel in which they arrive. Of course, the limiting factor in a pixelated system is the distance that passengers are willing to walk to a location. In a half-mile-by-half-mile system, especially in New York City, passengers may not be willing to walk the distance to the taxi stands. A better alternative for New York City would be pixels of a tenth-of-a-mile wide. However, as part of the goal of this project is for comparison to simple and optimal matching strategies, we use the convention of square pixels with half of a mile on each side. Below, we see the arrangement of passengers pickups (i.e. passengers getting picked up in a taxi in the New York City taxi dataset). We observe that much of the demand is concentrated in a relatively small number of pixels. We can model demand from other pixels as white noise, as most of the demand is concentrated in a small number of pixels in the center of the grid. The diagram on the bottom provides a more granular image of taxi trips in this pixelated system.. The locations of the centroids of each pixel, where the taxi stand is supposed to be, are also drawn on the image to give a sense of the distance that customers have to travel to the taxi stand. In many pixels in this image, passenger arrivals are distributed uniformly around the pixels; the ideal case for a pixelated taxi stand distribution is for passenger arrivals to be located near the center of pixels.

Figure 1: Taxi grid system map

If this kind of taxi grid system is to be used, then, it will require passengers to change their spatial arrival distributions. In other words, passengers will have to get picked up somewhere else than where they are used to getting picked up. Overall, while the grid system is well-defined and allows for an easier vehicle management system, customers will have to adapt to changes in taxi pickup and dropoff locations.
Figure 2: ATaxi stand map

Figure 3: New York Map
K-means Stand Positioning

Taxi departures appear to be concentrated in a few pixels, and even within pixels, arrivals appear to be uniformly distributed throughout the pixel. Taxi customers, then, would have to adjust to a pixelated taxi by getting used to being picked up elsewhere. As a result, these customers may not be amenable to such a system. In order to account for this spatial characteristic of taxi demand, then, we use a learning algorithm for K-means to find the optimal positions of the taxi stands. Below is the K-means algorithm, where DL refers to the spatial distribution of departures, SL refers to the stand locations:

Algorithm 1 Algorithm for aTaxi Stand Positioning (Stand Limited)

Input: $DL$, $n$
Output: $SL$
1: $SL \leftarrow KMEANS(DL, n)$

Algorithm 2 Algorithm for aTaxi Stand Positioning (Distance Limited)

Input: $DL$, $d$
Output: $SL$
$n \leftarrow 0$
2: repeat
$n \leftarrow n + 1$
4: $SL \leftarrow KMEANS(DL, n)$
until $\|DL_i - SL_{DL_i}\| < d \ \forall i$

K-means is a clustering algorithm where data points are partitioned into a set of clusters. In the context of taxi stand positioning, passenger arrivals are matched with a cluster that they are near to. The goal of the k-means clustering algorithm, then, is to station taxi stands where New York City taxi demand is the greatest. Because passengers have to travel less far, the use of K-means stand positioning is an improvement for aTaxi customers - customers are more likely to travel to stands and use the aTaxi service if they have to walk a shorter distance. However, for vehicle management strategies and repositioning, the K-means system poses a more difficult challenge, as taxi stands are arranged according to the dataset rather than neatly arranged according to a pixel grid. Below, the red squares indicate the positioning of taxi stands for the New York City area, where we use 1000 stands as the parameter for the K-means clustering. In addition, the arrivals are color coded according to the taxi stand to which they are assigned. A couple of key trends appear in this image; first, the K-means algorithm appears to be fitting the noisy departures (they represent a large cost to the algorithm if a taxi stand is not near them). Second, since some street corners and areas have so many departures, many taxi stands/clusters are assigned to those areas. Refer to the second image below to see many stands along street corners; a better alternative would be to assign an aTaxi “hub” to those areas rather than having many stands. In addition, a better solution to the third image below is to remove the noise from departures that are far away, or to not even service these areas: a couple of departures for an aTaxi stand will likely not make economic sense. In the image two pictures below, notice how the clusters seem to be along the streets, which is fitting for New York City stands.

264
Figure 4: Taxi stand visualization

Figure 4: aTaxi Activity Visualization
Hybrid Systems

Clearly, there are advantages and disadvantages to taxi stand positioning that is completely K-means-based, or that is completely pixel-based. The best alternative, then, might be to follow a hybrid approach to a taxi stand location strategy. On one hand, a K-means approach allows taxi stands to be located where the demand is, which reduces travel time to the stand. Indeed, a comparison of mean travel distance to a taxi stand shows that the K-means algorithm is much more effective in this respect. According to the below diagram, passengers have to travel less than three times the distance to get to stands distributed according to K-means than to pixels. Of course, some passengers have to travel an extensive amount to get to the taxi stands.

<table>
<thead>
<tr>
<th>K-means trip dist (miles)</th>
<th>Pixel distance to stand (miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The mean is 0.05</td>
<td>The mean is 0.18</td>
</tr>
<tr>
<td>The median is 0.03</td>
<td>The median is 0.19</td>
</tr>
<tr>
<td>The std is 0.18</td>
<td>The std is 0.07</td>
</tr>
<tr>
<td>The min is 0.00</td>
<td>The min is 0.00</td>
</tr>
<tr>
<td>The max is 16.04</td>
<td>The max is 0.35</td>
</tr>
<tr>
<td>The 10th percentile is 0.01</td>
<td>The 10th percentile is 0.08</td>
</tr>
<tr>
<td>The 20th percentile is 0.01</td>
<td>The 20th percentile is 0.12</td>
</tr>
<tr>
<td>The 30th percentile is 0.02</td>
<td>The 30th percentile is 0.14</td>
</tr>
<tr>
<td>The 40th percentile is 0.02</td>
<td>The 40th percentile is 0.17</td>
</tr>
</tbody>
</table>
Analyzing Taxi Demand

A large majority of the taxi demand comes from a very small number of pixels. Just the top 37 pixels account for 80% of the departures in the 4 analyzed months. To account for 90% of the departures, 53 pixels are needed. Almost all departures (99%) come from 175 pixels. All of these pixels are also very concentrated specially. As seen in the bar graph of departures below, most pixels with significant number of departures are located in the center. This also corresponds to the Island of Manhattan. There is also a smaller “bump” beyond
the island. This corresponds to John F. Kennedy International Airport.

The concentration of departures can be further seen in the cumulative trip graphs, as almost all of the 60 million trips are cumulated within the first hundred pixel (when the pixels are ordered by number of departures).
Figure 8: CDF of trips by pixel over 4 months

Figure 9: Trips by Pixel CDF
Unsurprisingly, the same fact is seen in the cumulative percent of trips. It is interesting to compare the cumulative percent of arrivals. This graph shows the percent of arrivals cumulated over pixels ordered by the number of arrivals. While the cumulative percentage increases quickly, reaching 99% within 500 pixels, it does not increase as quickly as the corresponding percentage of departures. This suggests, which will be later confirmed, is that there are more destination pixel than departure pixels. In other words, there are a few sources of trips and more sinks. With in context of New York City is not all surprising. One would imagine that many taxis depart from transportation hubs such as JFK Airport, Laguardia Airport, Penn Station New York, and Grand Central Station. One the other one, one would imagine that the numerous business, residential, and tourist attractions are spread throughout the city.

![Cumulative Trips By Pixel Over 4 Months](image)

Figure 10: CDF % of trips over 4 months by pixel
Zooming in on the cluster of pixels with large number of departures reveals more intricate structure. There are 3 distinct rows of departures. The orientation of the graph is looking east from New Jersey across the Hudson River. Photographs are included for reference. The middle peak corresponds to Midtown. The peaks on the right are from Downtown.
Pixel Imbalances

A critical component of taxi management is knowing the imbalances of trips at different location as vehicles must be repositioned from areas of where there are more arrivals of vehicles to areas where there are more departures. The special distribution of these imbalances determines the difficulty of the repositioning problem. The examination of cumulative trip distribution already indicated that there are more variance in the destination of trips than in the origins. It is still necessary to know how this difference translates into imbalances.

For imbalance analysis we looked at the total arrivals and departures from a pixel. While this is a very coarse measure which omits temporal patterns in imbalances, that is not the current focus. When we develop and simulate our repositioning strategies, we use hourly imbalances. We will also examine the temporal distribution of trips next. Thus, for now this measure suffices to reveal which pixels are trips generating and which pixels tend to be a destination.
To calculate the imbalance we subtracted the total number of departures from a pixel from the total number of arrivals. The map above shows the imbalances for the New York City area. Red circles indicate a net deficit of taxis and green circle indicate a net surplus of taxis. Pixel 55,31 had the largest deficit of 590,000 trips. This Manhattan pixel contains many apartment buildings and border pixels with net surplus of taxis. This suggests potential for easy repositioning. More broadly, we can see John F. Kennedy Airport and Laguardia Airport are the pixels outside of Manhattan with the largest deficits, as expected.
Zooming in on Manhattan we see an interesting pattern of imbalances. Many of the central pixels have a net deficit. Many of the surrounding pixels have a net surplus. The surpluses are much smaller than the deficits. It is important to recall that unlike the trips used for New Jersey, these trips are not based on paths people take throughout the day. These trips are occurring because someone needed to get from one pixel to another. However, that need is stochastic. There are temporal patterns in the trips, both on the hourly and daily scale, but there are no explicitly fixed travelers. With that in mind, in Midtown and center of Manhattan, we can imagine many tourists, businessmen, and New Jersey commuters grabbing a cab when getting out of Penn Station or Grand Central Station. Those cabs may return to that transportation hub several times without bringing a person there. Additionally, each of the Manhattan pixels encompasses several subway stations. Even if someone took a cab to get out of pixel, they may return to that pixel in another mode of transit. These are only speculations. The purposes of the trips would be necessary better understand the imbalance pattern. With information we could figure out if there is a group of people, like tourists. In such case may be the initial trip to the hotel is with a taxi and the rest of travel they do with the subway. However, from the present data it is rather hard to explain. However, one could map the destination coordinate to an address and assign trip purpose that way.

The following maps show clearly the imbalances across Manhattan. These maps indicate that a lot of
repositioning may be necessary as there are many pixel with many departures clustered together and even though there are pixel with surplus near by, those surpluses are relatively small. Therefore, taxis need to be brought in from further way. Of course, this is an estimate, as the temporal resolution is 4 months.

Figure 14: ATaxi imbalances south Manhattan
Figure 15: Ataxis imbalance in Union Square area
Figure 16: Ataxi imbalance in midtown area
For the last part of special distribution analysis, we will look at individual pixels. More specifically, we identified and provided the aerial views of five pixels with the most total departures over the 4 month period.

First the most traveled path, route between two pixels, is between 2 of the top 5 pixels. There were 192,363 trips between these places over the 4 months. The origin pixel (left) contains Penn Station - it is also the pixel with the most departures - and the destination (right) pixel contains Times Square and the Theater District. Most likely these trips are mostly made by tourists.
The top 5 pixels, see together below, each have over two million departures. As mentioned before, the top pixel contains Penn Station. The fifth pixel contains Times Square. The other pixels contain many office skyscrapers and those departures are probably business trips. These are a few places which did not make to the top five list. The Port Authority Bus Terminal, while near by, does not have enough departures to be in the top five. Both John F. Kennedy Airport and Laguardia Airport have under two million departures, however since they span several pixels, the departure coordinates (probably to do GPS noise) may have been spread among several pixels. Additionally, about half of all passenger at JFK use the AirTrain (2014 Taxicab Fact Book). Infact, by looking at the imbalances we do see several red circles for both airport. The total departures may have been larger. Grand Central Station is located just below the border of two pixels. While not being in any of the pixels, it’s proximity suggests that most of the departures for it are made in one or both of the pixels.
(55, 34) 2,676,423 departures

(57, 35) 2,384,065 departures
(56, 36) 2,309,576 departures

(57, 36) 2,098,186 departures
Hourly Departure Rates Distribution

The main benefit of using the New York City taxicab trip data is access to trips outside of the one day period on which the 2010 Census was taken. For temporal analysis we decided to look at a week long period of departure rates for each hour, 168 hours from midnight on Monday to 11 PM on Sunday. We limited the horizon to 1 week and resolution to 1 hours for two reasons. First, we expected to see the most patterns and the least noise at these parameters; perhaps using a 30 or 15 minute resolution would have been interesting. Second, is a concern over the number of parameters to be estimated. We are working with 60 million trips because of limited computational resources, although up to a billion trips are available. The number of parameters to be estimated for 168 hours for 11,250 pixels is 1.89 million. That is about 30 trips for every parameter. A full month would require over 8 million parameters. Using 15 minute intervals for a week would require 7.5 million parameters. In both cases, the amount of data we are working with seemed too small to estimate all of those parameters.

Looking at the 168 hourly departure rates averaged over the 4 months, we see that the one day New Jersey data is missing a lot of important information. The graph below shows the departure rates for each hour grouped by day. Zero corresponds to Monday which is very similar to Tuesday, Wednesday and Thursday. Then Friday and Saturday have similar distributions. Lastly, Sunday is similar to Monday. We cannot explain why Friday and Saturday look so similar. One would expect such difference for the weekend, but we’ve check the code for an off by one error. Additionally, if we assume an incorrect numbering, then the offset is in the wrong direction. As the error would imply that Tuesday starts the week, that is not the case in international standards used in Python. But, even if there is a code error we cannot identify, the numbering of the days is arbitrary for the purposes of taxi repositioning and the fact remains that there are significant difference between the days.
Individual day graphs are reproduced below for closer inspection. In general, everyday exhibits the least trips at 3 and 4 in the morning, followed by a sharp increase during morning commute. The middays are very similar. It seems that one New York City gets going it keeps going until late in the night. Most of the days have a spike in taxi departures late at night, from 10 PM to midnight.
Figure 19: Monday Hourly trip volumes

Figure 20: Tuesday hourly trip volumes
Figure 21: Wednesday hourly trip volumes

Figure 22: Thursday hourly trip volumes
Figure 22: Friday hourly trip volume

Figure 23: Saturday hourly trip volume
Figure 24: Sunday Hourly trip volume

Variance of in the hourly departures highlight what can already be gleaned from the graphs. Early morning (3 AM to 5 PM) is predictable across the seven days, as is midday (9 AM to 5 PM). That is not to say there are few trips. In fact, most of the trips happen during the midday period. The most variance occurs in the morning commute and at night. Most of the variance during midnight, 1 AM and 2 AM is due to the increase in trips on Friday and Saturday. Based on the result no single day can be used for departure prediction.
Combining spatial and temporal distributions of departures, we looked at the variance of trip departures for every hour over a week for the pixel with the most departures, the pixel containing Penn Station. The graph below shows the variances. The most variance occurs during the morning commute. To explore this further we looked at trip departures from that pixel at 6 AM on every day of the week.
The number of departures for each day is about 20,000 for each day except for two, when the number drops by almost 15,000 departures. That is a significant number of departures. This graph demonstrates that for repositioning one needs to consider the special and temporal distribution of departures, at least with a week horizon week and to the hour level.

**Vehicle Management System**

Armed with this analysis of the taxi dataset and the resulting distributions, we are now ready to implement a vehicle management system. Specifically, there are two aspects of a vehicle management system that we cover in this section. The first aspect is vehicle fleet size - how many taxis it takes to cover a geographic region in a certain amount of time. The second aspect is the vehicle repositioning strategies that we will employ - where will we move empty taxis throughout the day in order to cover demand. In this section, we implement four vehicle repositioning strategies, each attempting to make an improvement on the other. We discuss the rationale behind each of these strategies. Then we simulate the four strategies across a forty-eight hour period, which contains just under 800,000 trips. The simulation allows us to see how the strategies respond to the changing demand characteristics that we observed in the previous section, day-to-day and even hour-to-hour.

**Responding to Stochastic Demand**

In our vehicle management simulation, the fleet of taxis must respond to taxis in real-time. Therefore, the fleet of taxis must be well-positioned so that it doesn’t miss a large wave of expected demand. The main probabilistic concept that we use to model the demand comes from the earlier analysis of the taxi trips. In that analysis, we found that the demand varied by hour, by day, and even by pixel. Therefore, for each pixel, we represent the demand of each hour of each day as an inhomogeneous Poisson process. In other words, for each pixel the arrival rate of customers to a taxi stand changes throughout the day. We use 4 months of yellow taxi
data, corresponding to about 60 million trips, to compute the rates of the inhomogeneous Poisson process for each pixel, for each day of the week, for each hour of the day. We end up with a distribution of 168 data points for all 150x75 = 11,250 pixels, where one taxi stand is in each pixel. Each data point represents our expectation of arrivals for that pixel in that hour, which is critical in helping us meet the taxi demand. We now move on to formulating strategies for meeting stochastic demand.

**Strategy 1: End-of-Day Bulk Repositioning**

The first strategy that we implement is a bulk repositioning of taxis at one time during the day. This represents a *system-wide* repositioning of the strategies. Rather than having each taxi act greedily to pick up the next passenger, this strategy positions taxis such that the overall positioning of the fleet of taxis captures the expected demand. The first step in this process is to find the optimal hour of the day to reposition the taxis, as well as the optimal hour that we want to tune the taxis to (i.e. the optimal hour at which we want to meet the expected demand). Using the distributions that we computed in the analysis above, we find that the best time for repositioning occurs in the morning at 2 AM, when there is the least expected demand, on average throughout the day. Of course, a taxi system can also alter this hour based on the day of the week. However, we find that this is the time of the day where we will miss the least number of trips from Monday night to Tuesday morning, which is when the simulation takes place.

The next step of the process is to put the problem in the form of a network flow problem. Specifically, the vehicle repositioning strategy is the solution to the minimum-cost flow problem, which outputs a method of sending flow throughout a network (the problem is described in further detail at [http://perso.ens-lyon.fr/eric.thierry/Graphes2010/amaury-pouly.pdf](http://perso.ens-lyon.fr/eric.thierry/Graphes2010/amaury-pouly.pdf)). The network is a set of nodes and arcs (which connect nodes to each other). Each arc contains a capacity (i.e. how much flow can pass through the network) as well as a cost (i.e. how expensive it is to pass flow through that arc). Each node contains a net demand (a negative value means that the node is a net supplier of flow and a positive value means that the node is a net importer of flow). Finally, the sum of the total demands must be zero in order for the problem to be feasible.

To start to formulate the vehicle repositioning problem as a minimum-cost network flow problem, we first compute the expectation of the demand for each pixel in the hour that we want to meet the expected demand (7 AM in our simulation, which is the start of the weekday morning rush). For each pixel, we subtract the number of taxis in the pixel at 2 AM from the expected demand at 7 AM. This gives us an imbalance for each pixel. Below is an example of what the imbalances would look like for each pixel. A negative value at a pixel means that there is an excess of taxis at a pixel, and a positive value at a pixel means that there is an excess of passenger demand expected at a pixel. A zero value means that there is no imbalance.
We group the pixels into two sets: excess taxis and excess demands. In the below picture, we create a digraph, where excess taxi pixels are in the green partition (the “supply” nodes), and excess demand pixels are in the red partition (the “demand” nodes). Each supply node is connected to each demand node because we want to find out which set of demand nodes to send the supply node. Since the supply and expected demand do not usually match up, we include a third partition where all of the excess flow goes. If there are excess taxis, then this node is to the left and connected to all supply nodes. If there is excess demand, then the node is in the opposite position. Next, the cost on the arcs from supply to demand is 1.2 times the Cartesian distance from the supply to the demand, and the cost from the supply nodes (or demand nodes, in the opposite case) must be the same for each supply node (so that all of the supply from one node doesn’t go to the excess node if the node is relatively farther away from demand than the other supply nodes). We make this cost zero. This problem is thus feasible, and its solution gives us the number of taxis to send from each supply node to each demand node. It turns out that this problem is computationally expensive to solve, and so bulk repositioning occurs once per day. The rationale for this strategy is that at the end of the day, a lot of imbalance is generated for taxis, and so this strategy tries to correct imbalance this at the end of the day, when demand for aTaxis is less frequent.
The results of using this strategy over a two day period are displayed below. Rarely is more than 20 percent of the fleet moving or occupied. Most of the taxis are making one trip during the day. The increase in the size of the fleet on the second day is much smaller, around 30,000 vehicles compared to 50,000 on the first day. Note the spike in the number of moving taxis around midnight as the vehicles are repositioned.

Figure 28: % of fleet moving by time
Figure 29: Total number of taxis with repositioning

Figure 30: Trips per taxi Distribution
Strategy 2: Real-Time Individual Repositioning

The second strategy is real-time individual repositioning. As opposed to the first strategy, which is a system-wide strategy, this strategy is an individual repositioning strategy. The rationale for this strategy is due to the asymmetry in the dataset that we observed earlier. On average, taxi trips that travel from Point A to Point B do go back in the direction of Point A. In other words, throughout the day there is a “flow” of trips from one spatial location to another. Taxis that are not redistributed before the end of the day, however, might therefore
be idle and make only one trip. This is evidenced by the fact that about 10,000 taxis in the first strategy made only one trip.

This strategy works by sending a recently unoccupied taxi to a different location if its dropoff location is expected to have little demand. In this case, the strategy computes a value for all pixels within a certain radius of the taxi. The value for each pixel is the number of expected arrivals in the current hour, minus the number of taxis already at that location (to prevent taxis from congregating at one pixel), minus the distance from the current pixel to that pixel (to prevent the taxi from making wasteful, long-distance trips). In simple representation of the second strategy in the image below, the taxi has to “decide” between going to the closer pixel with a lot of expected demand but some taxis and the farther pixel with a lot of expected demand and no taxis.

The results of using this strategy over a two day period are displayed below. On average about 30% of the fleet is moving at any point in time. The taxis are also making many more trips, about 60 per day. This increase leads to more people being picked up by fewer taxies as the total fleet size at the end of the second day is about half of the total fleet size of the first strategy. The increase in the size of the fleet on the second day is much smaller, around 7,500 vehicles, and only 30,000 vehicles were added on the first day. We can see a large spike in the movement of empty taxis from midnight to 3 am. This is due to the low expected demand for most pixels during this time, and the strategy repositioning the vehicles for morning commute.
Figure 33: Trips per Taxi Distribution

Figure 34: % of fleet occupied by time
Strategy 3: Real-Time Departure Pickup

The third strategy involves a tradeoff between longer passenger wait time and smaller fleet size. Rather than generate a new taxi if there is no taxi to service that demand at the moment, this strategy searches for available taxis within an approximate five-minute radius of the passenger. In the below representation of the third strategy, we search for other taxis in the twenty-four closest pixels to the arrival, spanning a five-by-five grid around the current passenger. While we used this wait time for passengers in New Jersey, this wait time might not be appropriate for customers in New York City, however. The reason that we think this change will reduce the fleet size is partly because we will have to generate less taxis, and partly because we think that getting taxis in motion is another way to generate higher taxi usage. This is because that after this taxi drops off its passenger, it will have another opportunity to move to another place with high expected demand, using the individual policy of strategy two. This third strategy is demand-centric - it looks at individual demands and assigns a taxi to it, without looking at the overall system and state of the taxi fleet and demand.
The results of using this strategy over a two day period are displayed below. On average about 40% of the fleet is moving at any point in time. The taxis are also making many more trips, about 85 per day. The total fleet size is only 25,000 vehicles with 7,000 vehicles added on the second day and only 18,000 vehicles were added on the first day. Also, the rate at which taxis are added on the first day is much smoother than with strategy 2. This is due to around 75% of the fleet being utilized during that time as opposed to 35%. Over all, the taxes are making more occupied trips, 31 per day on average.

Figure 36: % of fleet used by time
Figure 37: Total taxis repositioned

Figure 38: Trips by time of day
Strategy 4: Hourly Expectation Repositioning

In our fourth and final strategy, we implement bulk repositioning on a larger scale. However, recognizing the amount of time required to solve the minimum-cost network flow problem, we perform a computation that does not optimally meet the demand with minimum distanced travelled, but rather satisfies the nodes with the highest expected demand. At the end of each hour, we compute the expected demand and imbalance for each pixel for the next hour. We then send taxis from negatively-imbalanced nodes to the most positively-imbalanced supply nodes. The rationale for this strategy is that we do not want to imbalance to accumulate until the end of the day; instead, we want to continually deal with the imbalance of supply and demand throughout the day. The goal of this approach is to end up with less of an imbalance at the end of the day than with a one-time repositioning per day.
The results of using this strategy over a two day period are displayed below. The results are very similar to the ones from the previous strategy. On average about 45% of the fleet is moving at any point in time. The taxis are also making about 87 trips per day, and on average 30 of those are occupied. The total fleet size is about 26,000 vehicles with 8,000 vehicles added on the second day and only 18,000 vehicles were added on the first day. The utilization during morning commute is around 90%.

Figure 41: % of fleet moving by time
Figure 42: Total number of taxies Repositioned

Figure 43: Trips per taxi distribution
Figure 44: Occupied trips per taxi distribution

Figure 45: Average Trips per Taxi per day: 30
Results

The use of K-means to position taxi stands improves the mean walking distance to the stand from 0.18 miles for pixels to 0.05 miles. The maximum walking distance is unfortunately unbounded, unlike with the pixel approach. In our test the maximum was 16 miles. An unacceptable distance to travel. Therefore a hybrid approach should be used where stands in areas with many departures are positioned by K-means and all other areas are pixelated. We also proposed an algorithm for using K-means with increasing number of stands until all departures are within a certain distance of a stand. That still produced undesirable results in low density areas as stands would be positioned in the exact location of the departure, assuming the algorithm is allowed unlimited number of stands.

Specially, trips departures are very heavily concentrated. Most trips depart from Manhattan. 175 of the 11,250 pixels generate 99% of all of the trips. These is more variety in the destinations than the origins. This manifests itself in imbalances across the New York City area. The significant regions of net deficit of taxis include JFK Airport, Laguardia Airport, and central Manhattan. Areas of net surplus of taxis are most prominent around Manhattan.
Various times of the week have different departure rates. While the days can be divided into two groups in terms of similar distributions, there are still daily variations. Most stable part of the day across the week is early morning and midday, 10 AM to 5 PM. We demonstrated the need for a hourly resolution and weekly horizon for estimating departure rates based on the Penn Station pixel.

Using the learned departure rates we created 4 repositioning strategies and simulated them on two days of trips. Each strategy improves upon the previous one, with strategies 3 and 4 yielding very similar results. The use of better repositioning strategies reduced the fleet size needed for the two days from 79,000 to 25,000. As the number of needed taxis decreased so did the average miles traveled for repositioning. The average number of occupied trips also improved from 9 to 31 trips per day.

Several other characteristics improved across the four strategies. While strategy 1 used about 20% of the taxis at any point in time, strategy 2 improved that to 30% and strategies 3 and 4 increase that to 40% and 45%. Strategy 1 needed around 50,000 vehicles on the first day, while strategies 2, 3, and 4 needed around 30,000, 25,000 and 26,000. This is due partially to better taxi utilization during the morning commute which increased from 20% for strategy 1 to 30% for strategy 2 to 75% for strategies 3 and 4. The number of taxis added in the second day also improved. Strategy one added around 30,000 vehicles while the other 3 strategies added around 7,500.
<table>
<thead>
<tr>
<th></th>
<th>Strategy 1</th>
<th>Strategy 2</th>
<th>Strategy 3</th>
<th>Strategy 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Repositioning Trip Miles</td>
<td>2.225</td>
<td>1.114</td>
<td>0.621</td>
<td>0.678</td>
</tr>
<tr>
<td>Average Occupied Trip Miles</td>
<td>2.023</td>
<td>2.023</td>
<td>2.023</td>
<td>2.023</td>
</tr>
<tr>
<td>Repositioning Trips</td>
<td>13,593</td>
<td>804,202</td>
<td>916,281</td>
<td>911,259</td>
</tr>
<tr>
<td>Occupied Trips</td>
<td>795,019</td>
<td>795,019</td>
<td>795,019</td>
<td>795,019</td>
</tr>
<tr>
<td>Fleet size after 48 hours</td>
<td>79,505</td>
<td>36,956</td>
<td>25,171</td>
<td>26,137</td>
</tr>
</tbody>
</table>

Similar to New Jersey, the average vehicle occupancy (AVO) for New York City is rather low. For all routes, the AVO is 1.73. There are 9,944 have with AVO greater than one 1. The table below list the AVO for best routes for ride sharing. All of these AVO are at 5 minute departure delay and one common destination. While the results should improve with more common destinations, it is still surprising to see the AVO being so small in such a dense city.

<table>
<thead>
<tr>
<th>Top Routes</th>
<th>AVO</th>
<th>People Carried (per week)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>5.69</td>
<td>76,000</td>
</tr>
<tr>
<td>100</td>
<td>4.1</td>
<td>486,000</td>
</tr>
<tr>
<td>500</td>
<td>2.92</td>
<td>1,380,000</td>
</tr>
<tr>
<td>1000</td>
<td>2.52</td>
<td>1,900,000</td>
</tr>
</tbody>
</table>

According to the 2014 Taxicab Fact Book, there were 13,437 taxi medallions in 2014. Our best strategy requires 25,171 cars, which is almost double the number of taxis our best strategy needs. There are several reasons for why our strategies underperform. We may not be doing enough individual taxi policies. In other words we are not making the taxis selfish enough to keep going back to areas with many departures. This would...
leave taxis stranded outside Manhattan and thus new taxis would need to be called up. For all of the strategies, the percent of taxis with passengers keeps falling throughout the day. From the fleet size graphs, we can also see that there is a large increase in new taxis after the number of taxis reaches 13,000. This corresponds point also corresponds to 5 PM. This is also the point at which the departure rate increases in the distribution for Monday. This would naturally drive up the demand for taxis. Our strategies may be keeping taxis at pixels for too long in the expectation for arrival. Additionally, our methodology may have made this comparison invalid. For the simulation we select random trips for each hour of the week. However, these trips were pick from a month worth of data, meaning trips for 1 am on Monday were picked from 1 am across all Mondays. Therefore, while each trip was service by a taxi in the real world, our random group of trips may not be serviceable by the same number of taxis.

The implication of our research is that for realistic results of a minimum fleet size in New York City, New Jersey, and the United States of America, the temporal distribution of trips must be considered with a horizon greater than one day.

Future Endeavors

In future research several things should be tried. The pixel could be decreased to 0.1 miles as opposed to 0.5 miles. Although only 5% of trips are less than 1/2 mile (2014 Taxicab Fact Book), therefore, half mile pixels still capture many trips.

This research should be used as groundwork for a much larger study. That is to say, a longer horizon and more time intervals should be considered. The simulation of policies should also be done on several weeks or even months so that seasonal effects can be captured. As noted in Taxicab Factbook 2014, “Average daily taxi usage is typically highest in the spring months and lowest in the summer months”.

Lastly, we found data on the road network in New York City. It can be found on http://www.nyc.gov/html/dcp/html/bytes/meta_lion.shtml. Using QGIS software, we processed the data and plotted the roads in Python. The results can be seen in maps below. It would be interesting to use this network in simulation and flow taxi trips along actual streets.
References for Data

All students started off with the OTrip files and NNTrip files found on Professor Kornhauser’s website. Links at (1), (2), respectively.

1. http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F14/oTrips_ByCountyByMode/
2. http://orfe.princeton.edu/~alaink/NJ_aTaxiOrf467F14/%5bABC%5dModule7NN_New/

Consult Chenyi Chen (chenyi@princeton.edu) for header explanations and methods.

For the data generation, files were used from the OTrip files exclusively. Outputs for cumulative arrival and departure distributions can be found in (3), (4) respectively.


Individual report sections on data usage can be found within each section.

MATLAB, Python, R and Java (for animation purposes) comprised the majority of code used in the creation of this final report.

Any additional data sections (specific animations used, etc) can be found in the link at (5):

This final report can be viewed and downloaded in its entirety at (5) as well.