A RIDESHARING ANALYSIS WITH A HITCHHIKING
MODIFICATION APPLIED TO TAXI TRIPS IN NEW YORK

ALEXANDER J. SCHINDELE
ADVISOR: PROFESSOR ALAIN L. KORNHAUSER

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN ENGINEERING
DEPARTMENT OF OPERATIONS RESEARCH AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY

JUNE 2016
I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

_____________________________
Alexander J. Schindele

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

_____________________________
Alexander J. Schindele
Abstract

The past few years have seen the advent and rise of a number of transportation companies that uses the power of the Internet and mobile devices to fulfill consumers needs for efficient, timely taxi transport. Companies such as Uber and Lyft have leveraged communication networks to grow from small startups to firms exceeding market valuations in the billions of dollars. As these two spheres - the public and the private sectors of taxi companies continue to compete, efficiency in managing large, sprawling transportation networks of taxis will become of paramount importance. The reasons for such policies are numerous. For one, increasing the efficiency of a transportation network reduces the emissions from these vehicles, limiting the impact of the transportation network on the environment. Also, implementing smarter policies regarding taxi usage can reduce the costs of a transportation company by limiting wear and tear on vehicles.

In particular, the policy of ridesharing is one proposed alternative to limiting the distance a taxi drives without any passengers, or in other words increasing the average vehicle occupancy of the taxis in question. The idea of ridesharing rests on the idea that multiple taxi riders can be picked up in the same or similar place and transported one-by-one to each of their different desired destinations. Such an analysis of the potential of ridesharing in New York City has already been investigated in thesis of A.J. Swoboda, who used transportation data from New York’s taxis to determine the feasibility of ridesharing in New York. This thesis supplements this sort of analysis with the addition of a “hitchhiking” policy. With such a policy in place, a taxi would, in addition to the original ridesharing mechanics, pick up passengers en route to its destination or destinations.

To identify the potential for efficiency policies such as ridesharing, one must leverage data and analyze data sets of taxi trips for ways to optimize networks of taxis and their clients. Publicly available data sets from the City of New York describing all taxi rides both through the public Taxi and Limousine Commission as well as those handled by Uber are available and used to determine the aggregate demand for taxis in the five boroughs of New York City. From there, the trip data is analyzed and the potential for ridesharing applicability can be determined or identified.
Acknowledgements

I would firstly like to thank my advisor, Professor Alain Kornhauser, for guiding me through the process of writing this thesis. I came into this topic with a great amount of interest but lacked a clear direction, and it was only with his help that I was able to find a clear line of analysis. I would also like to thank A.J. Swoboda and past Princeton students who have put in huge amounts of work in their own theses which have laid the groundwork for my own thesis.

I would also like to thank my parents, who have shown me support from the beginning. Their love and affection has helped me along my path through Princeton. I don’t think I could have done it without them. I also want to thank my grandmother, who has showered me with affection every time we have the chance to meet. Also a shout out to my sister, who despite not showing it, definitely loves me and appreciates my company.

Finally, I would like to acknowledge my friends, both at Princeton and outside of it. To Joy, thank you for making my life at Princeton so much brighter. To all my teammates on the volleyball, especially Mike, Devin and Zach, thank you for all the great times I have had on the volleyball team. To my friend Chamsi, thank you for commiserating with me during all our ORFE struggles. To Anthony and my other Top Scholar friends, thank you for all the great laughs and amazing food. To all my high school friends from Singapore, especially Oliver, Adharsh and Shreshth, thank you for maintaining this amazing friendship that I hope will last for a long time.
Contents

Abstract ................................................................. iii
Acknowledgements ....................................................... iv
List of Tables ........................................................... ix
List of Figures .......................................................... x

1 Introduction ......................................................... 1
   1.1 Purpose .......................................................... 3
   1.2 Modeling Transportation Demand ............................. 3
   1.3 Private and Public Taxi Companies ........................... 5
      1.3.1 New York City’s Taxi and Limousine Commission .......... 5
      1.3.2 Uber, Lyft and Other Private Taxi Companies .......... 5
   1.4 Efficiency in Transportation Systems ......................... 7
      1.4.1 Average Vehicle Occupancy .............................. 7
      1.4.2 Reasons for Efficiency .................................. 7
   1.5 Ridesharing ................................................... 9
      1.5.1 A Hitchhiking System .................................. 11
      1.5.2 UberPool and Other Implementations of Ridesharing ....... 12
      1.5.3 Self-Driving Cars .................................. 13
   1.6 Summary .................................................. 14

2 Data ............................................................... 15
   2.1 Data Sources ................................................ 15
      2.1.1 The New York City Taxi and Limousine Commission and 2014 Data 15
## C Code

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1  Removal of Columns in TLC Data</td>
<td>72</td>
</tr>
<tr>
<td>C.2  Pixelization of TLC Data</td>
<td>73</td>
</tr>
<tr>
<td>C.3  The Ridesharing Implementation</td>
<td>74</td>
</tr>
</tbody>
</table>
# List of Tables

3.1 The Rise of Uber in 2015 .................................................. 31

5.1 The Most Popular Pickup Areas for TLC Taxis ................. 46
5.2 The Most Popular Dropoff Areas for TLC Taxis ................. 46
5.3 The Most Popular Pickup Areas for Uber Taxis .................. 47
## List of Figures

1.1 Illustration of CD=1 Ridesharing Policy ........................................ 10  
1.2 Illustration of a Hitchhiking System for Ridesharing .................... 11  
1.3 A Diagram of the UberPool Implementation ................................. 12  

3.1 Heatmap of Uber Pickups in New York and Surrounding Areas ......... 22  
3.2 Heatmap of TLC Pickups in New York and Surrounding Areas .......... 22  
3.3 Pickup Times Throughout the Year of 2014 ................................. 24  
3.4 Pickup Times Throughout the Day .............................................. 25  
3.5 The Cumulative Distribution of Trip Distances for Yellow Taxis ....... 27  
3.6 A Linear Regression of Reported Distances vs. Cartesian Distances ... 28  

5.1 An Inset of the Grid Overlaid on Lower Manhattan ..................... 44  

6.1 Example of the Voxel Used In Ridesharing Analysis ..................... 52  
6.2 Sample of a Voxel Environment with Two Voxels Highlighted .......... 52  
6.3 Voxel Environment with Line Drawn Between Origin and Destination ... 54  
6.4 Example Voxel Environment with Time Slices ............................. 55  
6.5 Example Voxel Environment with Time Slices and Trip Boundary “Cone” . 56  

A.1 The Rise of Uber in 2015 ..................................................... 67  
A.2 The Pixel Grid Overlaid Over the New York Area ....................... 68  

B.1 Sample of 2015 Uber Data ................................................... 69  
B.2 Sample of 2014 TLC Data (Continued in Figure B.3) ..................... 70  
B.3 Sample of 2014 TLC Data (Continued from Figure B.2) ................. 70  
Chapter 1

Introduction

The issue of transportation and the sprawling transportation networks of today have drawn the attention of a number of parties, involving professors, logisticians and governmental policymakers. The widespread use of modes of transportation makes this issue one that affects almost all parties in society, whether they be an individual, a corporation or a level of government. Because of the wide-reaching effects of any policies regarding transportation, the issue of transportation has become central to a number of discussions. This is especially true in the United States, where public transportation lags behind other developed countries, and investments for forms of public transportation whether they be tunnels, bridges or new railways have become contested topics. In the face of threats to the environment and the general search for new efficiencies in today's sprawling transportation systems, new avenues for efficiencies are constantly being investigated.

While investments such as public transportation have the potential to fulfill transportation needs and reduce waste and inefficiency in the future, there are possible avenues by which transportation needs can be fulfilled efficiently using policies that utilize existing means of transportation. Instead of constructing new infrastructure entirely in the form of new railroads or subways, some policymakers suggest the utilization of existing networks of roads and rails to fulfill the demands of consumers who require transportation services.

One of the proposed methods to alleviate the demand for transportation is the idea of ridesharing. A ridesharing policy lets riders with common destinations pool together in taxis. For example, two or more people - who are unaffiliated with one another - could
meet at a common point and proceed via a shared taxi to their common destination. By doing so, the number of miles a taxi must drive to satisfy consumer demand is reduced, thus making travel to the respective customers’ destinations more efficient. The mechanics of ridesharing can be adjusted such that no person has to walk more than a certain distance to reach the common meeting area. The implementation of a ridesharing policy was discussed heavily in A.J. Swoboda’s thesis, who used data from the New York City Taxi and Limousine Commission to identify the efficiencies that could be created by implementing such a ridesharing policy (Swoboda, 2015).

However, Swoboda’s analysis, while investigating the dynamics of having multiple destinations for a ridesharing service, did not delve into the specifics of modifying the pickup portion of a ridesharing policy. Instead of having taxi users meet at a common point, for example, one could formulate a policy which involves having a taxi pick up customers along the way to some destination or destinations. In particular, a hitchhiking strategy can be used to further create efficiency within these sprawling taxi networks. A hitchhiking policy allows a taxi with occupants to stop during the trip to their common destination to pick up other riders, who can then go to the same common destination or to a different one entirely. By adding this modification to the existing analyses of taxi ride networks, there is even more potential for finding new avenues of efficiency.

An even more pressing issue in the field of transportation dynamics is the meteoric rise of technological companies that focus on providing forms of transportation, such as Uber. As Uber reaches a valuation of over $65 billion (Newcomer, 2015) and other transportation companies attempt to compete for these new streams of revenue, understanding and analyzing the existing data for these modes of transportation is becoming increasingly important. Furthermore, isolating the effects of Uber and other private companies’ effects on existing transportation networks can provide an avenue into understanding the dynamics of the rise of the sharing economy and its role in a rapidly changing world of transportation. Not only does this data provide avenues of improvement for the private companies themselves, but it also suggests how public taxi companies can adjust to the evolving landscape of transportation.
1.1 Purpose

With the advent of the new sharing economy, as well as the rise of Uber and other taxi companies, the introduction of more ridesharing policies has become a topic of contention. While ridesharing is a service in taxis that many are unused to, the negative and positive possible effects of instituting a ridesharing policy are legion. With possible effects such as a reduction of the carbon footprint and the reduction of automobile injuries or fatalities, the issue of introducing new efficiencies to existing transportation networks continues to be a central concern.

This thesis analyzes the aggregate demand of taxi trips both through public means and through those owned by private companies, then use this analysis to assess the feasibility of a ridesharing policy in these cases. Most importantly, the feasibility of a hitchhiking system will be analyzed. In this hitchhiking system, a taxi with occupants can stop to pick up other riders en route to its destination. Using these various analyses, it is also possible to determine an answer to a fundamental question as to how competition between these various firms both private and public affects the market of taxi trips.

1.2 Modeling Transportation Demand

Various faculty and students of the Operations Research and Financial Engineering department at Princeton University, along with many other top scholars at other institutions, have labored to create a corpus of knowledge that underlies a comprehensive transportation demand model for the entire United States. In particular, the work of Professor Alain Kornhauser and many of the people he has advised have built a framework that seeks to model this unimaginably complex system. This model would be invaluable in understanding the effects of innovations in transportation, whether they be the instituting of a ridesharing policy or the dissemination of autonomous car technologies throughout the United States.

The work of two particular advisees of Professor Kornhauser have drawn attention to the concept of modeling transportation networks and applying a policy to our understanding of these systems. In 2014, Alexander Hill Wyrough created a “national disaggregate transportation demand model for the analysis of autonomous vehicles,” in which he simulated
the roughly 300 million residents of the United States in 2010 and the over 1 billion trips they took using personal travel behavior data that allowed the simulated trip data to mimic the actual trips of each individual person (Wyrough, 2014). In doing so, the synthetic trip data attempted to be a comprehensive understanding of every single one of the over 1 billion trips undergone in the United States in a year.

In 2015, Andrew James Swoboda, working with Professor Kornhauser, assessed the feasibility of a ridesharing policy if applied to public taxi cab trips in New York City (Swoboda, 2015). In his thesis, Swoboda used a comprehensive data set of over 168 million unique taxi trips in New York City in 2013. Through a detailed analysis of the source data, Swoboda was able to calculate the efficiencies produced had a ridesharing policy been implemented among the taxi trips in New York City. Moreover, Swoboda extended his analysis for a number of different forms of ridesharing, modifying the number of common destinations as well as the circuity. In doing so, he was able to isolate the effect of the various policies. He then suggested ways in which ridesharing could be practically be implemented in a city such as New York, recommending that a ridesharing taxi service could start in major transportation hubs such as Penn Station or John F. Kennedy International Airport before expanding the service to a wider level.

This thesis delves into the feasibility of other policies similar to the ridesharing policy proposed by Swoboda in his thesis. Using updated data from the TLC as well as additional sources of data from other services and cities, the results of this thesis can show how further application of ridesharing policies could feasibly affect the transportation dynamics of multiple metropolitan areas. Moreover, this analysis also goes into the dynamic between public and private taxi companies, as the rise of Uber and other taxi services - Uber, notably, has a ridesharing service of its own - has widespread implications for the future of transportation.
1.3 Private and Public Taxi Companies

1.3.1 New York City’s Taxi and Limousine Commission

The City of New York created the Taxi and Limousine Commission in 1971 to oversee the operations of a number of public transportation services in New York City. With over 74,000 vehicles and over 130,000 drivers who operate said vehicles (New York Taxi and Limousine Commission, 2014), the importance of the TLC in New York cannot be understated.

The transportation services the TLC provides are numerous. Medallion taxicabs consist of 13,635 vehicles in the TLCs vehicle pool (New York Taxi and Limousine Commission, 2014). Medallions grant their owners the right to drive the commonly found yellow taxi cab that provides service primarily for Manhattan. In addition to the yellow cabs, the TLC also created a new license for Boro Taxis, which provide service for the boroughs of New York and are known for their green color. A further 58,000 vehicles which the TLC licenses include black cars, luxury limousines and paratransit vehicles, also known as ambulettes. Furthermore, the TLC has also made a push to license commuter vans that can be pre-arranged to move larger groups of passengers. In total, these 78,000 vehicles provide transport for over 1.5 million passengers every day (New York Taxi and Limousine Commission, 2014).

As private companies have exploded into the taxi industry, the TLC has also reacted to this sudden new pressure. Following the rise of companies such as Uber, the TLC implemented its own E-Hail program that allowed riders to hail their taxi through an app on their mobile phone (New York Taxi and Limousine Commission, 2014).

1.3.2 Uber, Lyft and Other Private Taxi Companies

While the rise of the Internet first saw the meteoric rise of technological companies focused on information technologies such as Google and Amazon, the second phase of the Internet and tech boom began almost a decade and a half later in 2010s. Instead of companies producing services such as advertising or search, this new wave of companies relied on the sudden popularity of the mobile device (in particular, iPhones and Android devices). These newer technological companies focused more on what is today called the sharing
economy (Hamari et al., 2015). For example, sites such as Airbnb allow anyone to post a listing for their own living space for others to rent and use for a prescribed amount of time. Technological firms previously limited to the nonphysical space of the Internet have begun to encroach on services that had previously eschewed technological change.

Much of the spotlight in the new technological boom is focused on transportation companies, with Uber being the subject of much discussion and debate. Beginning with Uber in 2010, new transportation companies have sprung up quickly, with many focusing on specific niches.

While public transportation companies have been the larger force in the taxi scene of New York, the public sectors control over the taxi market has eroded since the introduction of these new, innovative transportation companies. The effect of these new transportation companies has been quantified and is elaborated upon in the analysis sections of this thesis. Another question that this thesis seeks to answer is how Ubers rise affects the market for taxi trips in metropolitan areas. To be specific, it is pivotal to understand whether these new Uber trips are taking away market share from public taxi trips, or whether Uber trips represent a new demand from a previously untapped market. Using the data from the TLC and Uber, the Uber effect on the traditional taxi market can be quantified.

The number of new taxi-like service companies is seemingly endless, as startups such as various taxi startups all attempt to carve out a place in the new transportation marketplace. Uber, as the first and largest beneficiary of this new boom, has become the focus of not only its clients and investors but a number of other stakeholders alike. Numerous cities have begun to ban Uber and other services similar to it. Cities such as Amsterdam have enacted laws that prevent Uber from operating in these metropolitan areas (Newcomer, 2015). In the most stunning reaction recently, Parisian taxi drivers rioted in response to the growing presence of Uber drivers in the French city (Chrisafis, 2016). As the tensions between the public and the private sectors of the taxi market continue to mount, the need to determine the balance of the market will become even more critical.

The market for taxi trips, being somewhat constant, will have to contend with the various new companies that are diving into their waters. The question of how this limited number of trips will be segmented among the players in the transportation market will
become increasingly important as Uber and other private services grow.

1.4 Efficiency in Transportation Systems

Metrics in transportation systems are a vital resource in determining the efficiency of said systems. Numerous metrics exist to calculate various facets of a transportation system. The question these metrics attempt to answer is how to compare the efficiency of vastly different transportation systems. Determining and improving the efficiencies of a transportation network can have significant benefits for all stakeholders involved.

1.4.1 Average Vehicle Occupancy

Perhaps the most important metric in defining the efficiency of a transportation system is its average vehicle occupancy (AVO). AVO is essentially a measure of how many miles were moved for all people in a vehicle per the total number of miles driven by that vehicle. In other words, AVO incorporates the idea of the number of people moved for every given vehicle, and can be written as the following:

\[ AVO = \frac{\text{person miles}}{\text{vehicle miles}} \]

The numerator of the AVO figure represents the demand for transportation, or a mobility desire utility, while the denominator can be envisioned as the corresponding supply and its delivery. As a simple example, a car that has four people in it that drives 50 miles will have an AVO of 4, as each person in the car moved 50 miles for a combined total of 200 miles, compared to the 50 miles driven by the vehicle. AVO is one of the most common and relatively simple metrics for determining the efficiency of a mode of transport.

1.4.2 Reasons for Efficiency

Numerous companies and individuals are concerned with increasing the efficiency of transportation systems for a variety of reasons.

One of the most significant reasons is the environmentalist aspect. As climate change
and other environmental crises continue to become more evident on a global scale, the drive to reduce the environmental footprint of the transportation industry has also grown. A total of 3.06 trillion miles were driven in 2015 (Department of Energy, 2015), and the trend of miles driven has grown slowly but almost continuously since the Department of Energy began recording statistics in 1971. 61.6 billion gallons of gasoline are consumed every year in powering these vehicles (U.S. Energy Information Administration, 2015). In the face of impending environmental crises, any meaningful improvement in the efficiency of existing transportation networks could significantly reduce the environmental impact of the transportation systems.

Another reason businesses involved in the transportation sphere might want to increase efficiency in their systems is to raise revenues and decrease costs. By increasing the number of people using a taxi within a given frame of time using a ridesharing policy, the revenue from one given ride - shared by multiple people - can be doubled, tripled or even more. As a given taxi driver can then use the time they saved on multiple trips to serve other customers, there is massive potential for ridesharing to enhance taxi companies’ revenues. Moreover, reduction in miles can reduce the costs associated with operating vehicles, such as those that come through depreciation. In the case of transportation companies, wear-and-tear on vehicles is a large expense that, if reduced, could see increases in a business bottom line. Thus taxi companies could be motivated to adopt ridesharing by the more self-serving goal of profit.

The final reason for the institution of policies to increase efficiency is the decrease in traffic congestion in crowded metropolitan areas. In cities such as New York, and in particular, Manhattan, a reduction in the number of vehicles on the road can obviously lower traffic congestion. The decrease in congestion introduces even more efficiency to these complicated systems, as traffic jams are reduced and travel times are lowered. As citizens of cities such as New York are freed from spending time in taxis, they can be productive in ways that would not be possible without the adoption of efficiencies introduced by ridesharing. Additionally, with fewer cars on the road due to potential transportation policies, accidents with other cars and pedestrians could potentially go down. Traffic reduction will also have other effects such as lowering noise pollution and generally bettering the aesthetic of
these already-crowded cities. Fewer cars on the road would also mean fewer chances for accidents between automobiles as well as between pedestrians and automobiles. This issue of safety is yet another impetus for reducing traffic congestion. More broadly, reducing traffic congestion can cities much more livable.

1.5 Ridesharing

Ridesharing is one of the many proposed ways in which a transportation network could see a marked increase in the efficiency of its transportation network. While ridesharing has the potential to revolutionize transportation networks, there are also legitimate downsides that need to be overcome before implementing such a policy. Customers using taxis with ridesharing will necessarily want benefits to overcome the downsides of ridesharing, such as issues with safety and discomfort.

When a network relies on a ridesharing policy, each unit of transportation in the case of this thesis, usually a taxi picks up multiple riders at a common origin point, then proceeds to each one of the clients destinations one-by-one in an efficient manner. Such a policy can be further supplemented by changing the circuity or the number of common destinations of its passengers.

In his own thesis, A.J. Swoboda analyzed the trips of TLC yellow taxis in New York City and outlined multiple ridesharing policies in an attempt to introduce efficiency into the complex transportation system in this metropolitan area (Swoboda, 2015). The common destinations of these rides were changed to determine how different ridesharing policies would affect the transportation system. For example, the initial analysis of Swoboda focused on rides with only one common destination, which in the ridesharing jargon can be named CD=1. This policy is shown in Figure 1.1.

However, the CD=1 ridesharing policy is not the only policy that can be implemented to improve the efficiency of a transportation system. In tweaking this ridesharing policy, one can change the number of common destinations the riders go toward. For example, the number of common destinations can be increased to 2, 3 or more. However, in doing so, another complication arises, as circuity must now be considered. Circuity refers to the
maximum inconvenience a taxis riders face when a taxi trip is planned. These directional considerations were also accounted for in additional ridesharing policies, as trips with multiple common destinations were constructed with care such that the trip to one destination did not deviate too much from the trips to the other destinations.

Instituting a ridesharing policy in practicality, of course, requires facing some significant obstacles. The most glaring issues are those dealing with the added inconvenience of ridesharing. There is a general mistrust toward the idea of sharing a taxi with potential strangers. While the new ridesharing world could include a re-thinking of a taxi trip as an elevator ride, nevertheless the very idea of sharing a small space such as the inside of a taxi is anathema to many travelers. Moreover, there are valid safety concerns with ridesharing between strangers, and the added liability for passengers and drivers would require changes to insurance policies governing taxis with ridesharing services.
1.5.1 A Hitchhiking System

The most interesting transportation policy which could be used in conjunction with ridesharing is the idea of a “hitchhiking” system. With this particular policy, a taxi will not only pick up occupants at a common origin, but will also pick up further passengers along the way to the current occupants destinations and take those additional passengers to their own desired destinations. This could perhaps be better understood as a “hitchhiking” system.

![Diagram of a Hitchhiking System for Ridesharing](image)

Figure 1.2: Illustration of a Hitchhiking System for Ridesharing

For example, if three passengers are riding in a taxi, then along a route that does not significantly deviate from the path to the passengers destination or destinations, the taxi can pick up another passenger at some mid-trip origin, then take all four occupants to their respective destinations. The destinations of the passengers do not have to be the same either; the taxi could theoretically pick up and drop off customers almost continuously while maintaining a level of circuitry with the intended destinations of all its passengers.

The hitchhiking system, in essence, is one where a customer notifies a central service that then informs a nearby occupied taxi with a similar destination to pick up said customer. With the emergence of Uber and other taxi companies, electronic hailing through mobile
devices has become a cornerstone of the new transportation revolution. Hailing through mobile devices, an innovation that has begun with the widespread use of smartphones, would synergize well with instituting a hitchhiking policy, as a customer’s need to use a taxi could be transmitted quickly to a nearby taxi and allow the taxi - with or without occupants - to satisfy a customer’s transportation needs.

1.5.2 UberPool and Other Implementations of Ridesharing

Beginning in 2014, Uber began offering a new service in a few metropolitan areas, including New York City. This new service, dubbed UberPool, allowed Uber customers to partake in a ridesharing service, where the Uber cab would stop to pick up additional Uber customers along the way to its final destination. Uber promises that such an implementation of a ridesharing policy adds “only a few minutes” (Uber, 2015) to each ride, while saving customers up to 50 percent in taxi fares. In this way, UberPool offsets the downside of having to share a ride with strangers by reducing taxi fares, a powerful incentive to increase the adoption of ridesharing.

UberPool, as shown in Figure 1.3 (Uber, 2015), is the practical implementation of a ridesharing policy that also includes a hitchhiking system. Rather than having all passengers being picked up in the same area, as analyzed in Swoboda’s thesis, UberPool would continuously pick up and drop off passengers as the route to the passengers’ destinations were being traversed. In this system, the route would be altered during the middle of the
trip, but not so significantly that it would cause an adverse delay for passengers already inside the vehicle. Theoretically, then, as can be deduced from Figure 1.3, a single UberPool “ride” could consist of multiple separated rides changed together, if customers happen to be waiting in the exact areas along the route of the continually running UberPool ride.

UberPool’s continued presence in major metropolitan areas and its successful implementation have provided an impetus for other transportation companies - especially the entrenched public groups such as those overseen by New York City’s TLC - to modify their own transportation policies to incorporate the same ideas implemented through UberPool. The analyses presented in this thesis can illustrate the application of the UberPool idea on to the taxi market already existing among yellow taxis in New York.

1.5.3 Self-Driving Cars

The field of autonomous or self-driving cars is rife for progress and innovation. Companies such as Google have pioneered autonomous cars that use a variety of technologies to allow car riders to accede control of the car to a computer, thereby freeing their attention for other tasks. The adoption of this innovation would represent a huge leap forward in efficiency, and would also have the additional benefits of increased safety. Such a technology would create a powerful synergy with a taxi service that also incorporates ridesharing, allowing for a transportation network that leverages numerous innovations to create efficiencies.

Self-driving cars had sprung up in the imagination of automobile companies since at least the 1930s (Weber, 2015). Though the technology underlying such a system were well away at this point in time, as computational power and technology have improved, the dream of a self-driving car has become more and more likely. The technologies behind self-driving cars continue to improve. For example, image processing through LIDAR and other media allow self-driving cars to determine its surroundings and make decisions based on the analyses of these images (Weber, 2015). The accuracy of these image processing algorithms has advanced significantly, and huge leaps have been taken by many other Princetonians in their own theses with regard to the exceedingly complex field of self-driving cars.

Google has been led a significant project in autonomous vehicle technology, with its Self-Driving Car Project having driven over 1 million miles around Mountain View, California,
and Austin, Texas. The car has been involved in only two accidents, both of which were due to human error from other cars (Google, 2016).

The benefits of autonomous vehicles seem to be endless. Apart from the savings in terms of time and money among users of self-driving car technology, the proper implementation of autonomous vehicle technology would see enormous savings in terms of safety. There are still challenges in implementing autonomous vehicles, especially in handling adverse weather conditions and other situations, and the legal system has yet to adjust to determine the idea of accident liability in the world of self-driving cars. Nevertheless, the field of self-driving cars will only continue to grow.

1.6 Summary

The effects of improving the efficiency of our transportation networks cannot be understated. Though innovations in the field of self-driving cars continue to be made, there is still potential for creating new avenues of efficiency through means that already exist at our disposal. Innovations such as Uber’s UberPool and self-driving cars have become applications of the ideas presented in the this chapter. Implementing a ridesharing policy with hitchhiking would provide an additional avenue for making transportation systems in metropolitan areas more efficient, as riders can now hitchhike on to a ride that is already nearby and thus create savings within a transportation system.
Chapter 2

Data

2.1 Data Sources

2.1.1 The New York City Taxi and Limousine Commission and 2014 Data

The New York City Taxi and Limousine Commission oversees the licensing and regulating of New York's yellow taxis, liveries, black cars, commuter vans, ambulettes and luxury limousines (TLC, 2016). In addition to maintaining and ensuring the quality of service of its taxi service, the TLC compiles and provides statistics related to the use of its taxis throughout New York (including the surrounding boroughs of New York in addition to Manhattan).

The TLC's vehicles can be split up into three main groups: yellow taxis, boro taxis and other for-hire vehicles (FHV's) (New York Taxi and Limousine Commission, 2014). Yellow taxis are operated under three different ways: as fleets, run by garages which lease out taxis for drivers to use; driver-owned vehicles, where a driver conditionally owns the taxi but leases the medallion to operate the taxi from an agent; and individual owner-operators, who own both car and medallion and are subject to their own regulations, such as driving at least 210 shifts per year. Boro taxis, which are green, were created in 2012 to provide taxi pickup services to those living in the boroughs of New York excluding Manhattan (New York Taxi and Limousine Commission, 2014). Since 94 percent of all yellow taxi pickups occur in either Manhattan or the John F. Kennedy and LaGuardia Airports, the introduction
of the green boro taxis allow for service in areas that have demand for taxis, though it is considerably less than the demand within Manhattan. Finally, FHVVs are composed of liveries (in other words, car services or community cars), black cars, luxury limousines and commuter vans. Liveries are pre-arranged forms of transport which depart from one of 500 bases spread around New York (New York Taxi and Limousine Commission, 2014). Black cars, on the other hand, provide service for corporate clients. The focus of this thesis revolves around the yellow taxis and the boro taxis, which provide an accurate portrayal of the demand for on-hire modes of transportation. Moreover, yellow and boro taxi data is much more readily available, and can be accessed for free from the TLC’s website.

Pursuant to the Freedom of Information Law, the TLC released trip data for all its yellow and green taxi cabs in 2013, 2014 and selected months in 2015. The trip data accumulated for the TLC’s vehicles encompasses the entirety of the years 2013 and 2014, and portions of the year 2015. This data is freely available at the URL shown in Appendix B. To maintain the privacy of all parties involved, a number of fields were anonymized. The driver ID was anonymized with an MD5 hash. Furthermore, all latitude and longitude coordinates in the data set are at a low precision, which conceals the exact pickup and dropoff locations of the taxis in the data set.

The data contains the following columns: driver ID, date and time of pickup, latitude and longitude of pickup, date and time of dropoff, latitude and longitude of dropoff, and various details regarding fares. An example of a few rows of the data are provided in Appendix B. Though the data regarding fares was detailed and potentially useful, since fares were not relevant to the ridesharing analysis, this data was removed for storage and computational considerations.

2.1.2 Uber and Other For-Hire Vehicle Companies’ Data

The NYC Open Data initiative, created to improve the accessibility, transparency, and accountability of City government, (Open Data, 2016) offers large amounts of data to the public concerning topics from business to social services in New York. The data has proven especially useful in the field of transportation, as curious researchers trawling through the data have identified some interesting findings concerning the transportation networks of
New York City.

Using the Freedom of Information Law, the news website FiveThirtyEight was able to request and obtain trip data from the TLC of Uber pickups in New York City. The data spans from April to September 2015, and contains over 4.5 million trips. Furthermore, 10 other for-hire-vehicle companies’ data were included in the data release, and aggregated data statistics for 329 for-hire vehicle companies were also appended (Bialik et al., 2015). These data were primarily used in four FiveThirtyEight stories which focused on how Uber was serving its customers in New York, and whether Uber was causing congestion problems due to the large number of taxis it runs within the area.

In addition to the data released through FiveThirtyEight’s investigation, a Freedom of Information Law request was put in to obtain data from Uber and other private for-hire tax companies to obtain information regarding the dropoff times and locations. Pursuant to this request, the TLC completed an aggregate trip analysis of all trips from January to June of 2015 affiliated with Uber. The aggregate analysis “includes the dispatching base number, the pickup date, the affiliated base number and pickup location based on taxi zone for each trip completed by a for-hire vehicle affiliated with the aforementioned bases.” In addition, the TLC provided a number of other resources to simplify the process of analyzing its data. The TLC taxi zone shapefile, the TLC taxi zone lookup table and the TLC taxi zone reference map, provided with the data, allowed for the separation of the taxi trips into the zones prescribed within the data itself. This greatly simplified the analysis of the separate boroughs. Notably, the TLC vehicle numbers and the driver IDs for each trip were withheld due to the New York State Public Officers Law.

2.2 Alterations to Source Data

A few notable alterations had to be made to the source data received from these three sources to not only ensure that the input into our model was sanitized, but also to reduce the computation of the approximately 160 million rows present within the data. Any errors that persisted within the dataset could jeopardize the application of a ridesharing policy.
2.2.1 Column Modifications

Firstly, a number of columns were removed from the initial source data. Of the 18 columns present in the dataset, 6 of them were related to the fares of each individual trip. While such information could no doubt be useful in another analysis, for the purposes of ridesharing these numbers only add to the storage burden. Thus, these 6 columns were removed. In addition, the data fields titled “Vendor ID,” “Rate Code,” “Store and Fwd Flag” and “Payment Type” were also removed since they were not relevant to the analysis.

Secondly, the data regarding the pickup and dropoff dates were recorded as a single date-time entry. To more accurately analyze the data, the date and time were split into separate columns for both pickups and dropoffs. This split would be particularly beneficial in analyzing the daily and yearly statistics regarding taxi demand.

2.2.2 Removing Erroneous Data

Provided with the TLC data was a GIS shapefile for the New York region, which provided the entire region where taxi pickups can occur. This constrains the feasible area not only within the confines of the five boroughs of New York, but it also excludes areas such as the ocean and other areas inaccessible by automobiles. Using this GIS shapefile, it was determined that there were trips that fell outside of the feasible set. Each trip’s origin and terminus was checked with the GIS shapefile to identify whether the trip was valid. By checking whether a trip originated or terminated outside the GIS shapefile, some trips which presumably were recorded in error were able to be identified and removed from the data. This simplification ensured that the pixelization and voxelization processes undertaken in Chapters 5 and 6 would be accurately done.

Furthermore, Swoboda noted in his own analysis that the anonymization of the driver identification data led to the inclusion of a number of taxi trips with no driver identification number (Swoboda, 2015). For issues of privacy, the TLC hashed the driver’s unique ID to conceal the identity of the driver in all the data prior to 2014. Notably, the MD5 hash used in this process turns the driver ID marked “0” to “CFCD208495D565EF66E7DFF9F98764DA.” Thus, by identifying and removing all trips marked with a this specific driver ID, the data
set was erased of a number of trips that had been incorporated into the data by accident. It is notable that following 2013, the TLC removed the column of information regarding the driver identification entirely.

In addition to the erroneous trips caused by a lack of driver identification, there were trips in the dataset that were recorded as 0 miles in length. These trips, presumably recorded by human or machine error, clearly have no bearing on an actual trip that took place and were discarded from the data. Moreover, some trips were recorded as having positive length, but their origin and terminus latitude-longitude coordinates were recorded as being the same. While these trips might have been valid, it is impossible to use these instances in the model due to errors in the data collection, and thus these trips were removed as well. The removal of these trips was crucial for the initial data analysis portion of this thesis, which required that trip’s length be non-zero for comparisons.

By removing all the erroneously recorded data, the number of trips was reduced by 3,345,145 entries, from 164,159,098 to 160,813,953 data entries. This represented a 2.04 percent reduction in the dataset, a noticeable decrease which nevertheless shows that most of the data collection was done accurately.
Chapter 3

Data Analysis

The analysis conducted in this thesis eventually requires pixelizing and voxelizing the New York area in question. In these pixelization and voxelization processes, the city is divided into pixels or voxels of a certain size. In the case of this thesis, the spatial dimensions of the pixels and voxels are 0.1 miles by 0.1 miles (Swoboda, 2015). This ensures that no occupant, once reaching their destination, will have to walk more than a tenth of a mile. This also alleviates the computational power required to account for the infinite possible coordinates of latitudes and longitudes that could represent every pickup or dropoff location.

A major component of this thesis includes an expansion on the ideas formulated in previous theses. In particular, the analysis done in this chapter compares and contrasts the findings of previous theses regarding the state of TLC’s transportation systems in past years with the findings from the most recent data from 2014 and 2015. Moreover, with additional data from Uber and other for-hire vehicle companies, the findings from New York in past years can also be understood as just one small facet of a larger picture regarding taxi systems. The results of the analysis of TLC data can be compared with Uber’s data to identify ways in which the TLC could learn lessons from Uber’s recent success.

The focus of this renewed analysis is the idea of the hitchhiking policy framework and how it would lead to gains in efficiency in New York’s taxi networks. To begin understanding how such a system would impact transportation in these various cities and sectors, it is imperative to delve into the base-level insights of the data. These base-level insights revolve around relatively simple analyses of the data such as how the number of trips evolve as a
function of time - both in the short-term (within a day) and the long-term (over an entire year). Further analysis includes an approximate analysis of what areas of New York have the highest volume of pickups, as well as the creation of a cumulative distribution graph of the trips based on trip length. These various calculations and analyses are significant due to their use further on in this thesis.

3.1 Trip Activity

3.1.1 As a Function of Pickup Location

The information regarding where New York taxis pick up their passengers is vital to understanding where to place a greater focus on with regard to transportation. For example, certainly John F. Kennedy sees significantly higher demand than Far Rockaway, and thus the application of transportation policies must reflect the greater importance of transportation hubs such as airports, train stations and other areas of general interest.

Figures 3.1 and 3.2 illustrate the amount of Uber and TLC activity, respectively, based on pickup location using a heatmap overlaid on satellite images of New York and the surrounding areas. This figure was produced by compiling all the data regarding Uber pickups in New York from April to September and using Google Maps’ API to overlay the trip pickups over a map of New York as a heatmap.

The heatmap of Uber’s pickups during this period seem to fall in line with past results regarding the location of pickups in New York. Most obvious is the fact that Manhattan has the vast majority of taxi pickups, as the increased intensity of color on the island shows. While there are some trips originating in Brooklyn (especially the Williamsburg area) based on the shades of yellow and green on the heatmap, it is very clear that most trips begin in Manhattan, with only relatively few originating in any of the other four boroughs. In Manhattan itself, there is heightened activity in a few localized regions. In particular, New York Penn Station in Manhattan has a dark red dot, implying a very high level of pickup activity. As one of the major rail transportation hubs of Manhattan, the incredibly high number of trips originating in this area is hardly surprising. A similar darker red occurs in the area around the World Trade Center in the southern portion of Manhattan, where the
CTA trains transport people between Jersey City and Manhattan. Moreover, Midtown East has a consistent light shade of red. As an area increasingly becoming the central location of
many banks, as well as providing host to Fifth Avenue and all the shoppers that patronize the luxury establishments of the street, Midtown East demand for taxi transportation will only continue to increase.

Notably, the heatmap shows increased activity in a small number of hotspots such as airports. John F. Kennedy International Airport (not pictured), LaGuardia Airport and Newark Liberty International Airport all show increased levels of activity in the generated heatmap of Uber activities. In fact, based on other analyses, trips from airports make up almost 5 percent of all taxi trips in New York. Thus, when considering ridesharing policies to apply to the taxi system of New York, special attention must be given to airports.

The heatmap of TLC taxi pickup activity, on the other hand, is extremely similar to that of the Uber pickup activity. LaGuardia Airport as well as John F. Kennedy International Airport have extremely heightened and localized activity, as seen in the bright red dots visible in the heatmaps. Most of the trip pickups are limited to Manhattan, with high levels of activity in Midtown, especially the areas surrounding Penn Station. In this way, the analysis of taxi pickup activity shows how Uber’s trips seem to be eating away at the market only the TLC had occupied previously. The analysis of trip activity is continued in the analysis of pixels in Chapter 5.

3.1.2 As a Function of Day of Year

The TLC data provides useful insights into how taxi use changes throughout the calendar year as well. Understanding the changes between taxi usage in various months and/or seasons can allow taxi systems to optimize themselves based on these predictable patterns of behavior in taxi demand.

The various peaks and valleys in taxi usage can reveal adverse weather events as well as the impact of holidays and other special days in the calendar year. Near the beginning of 2014, there is an extremely obvious decline in taxi usage. This day turns out to be January 21, 2014, which marked one of the worst winter storms in New York, with the city facing over 7 inches of snow on the day (Weather.Gov, 2014). Noting the impact of weather on transportation demand is another consideration that must be taken into account when investigating the impact of transportation policies.
More generally, there is a consistent rise and fall every week in the number of trips. The day of the week that has the fewest number of trips is by far Sunday, which is indicated by the consistent dip that occurs every week in Figure 3.2. The peaks, on the other hand, are during Saturdays. Furthermore, the usage of taxis clearly is affected by the general weather conditions. During the summer months, the number of taxi trips is significantly lower than during the winter months. As can be seen at the 250th day mark - roughly corresponding to the end of August and the beginning of September - there is a significant rise in taxi trips. During this very time period, the average daily temperature decreased by almost 6 degrees Celsius (Accuweather, 2014).

The general pattern of the number of taxi trips throughout the year is enlightening, as it allows policymakers to provide a larger fleet of taxis when demand is the highest - such as during the winter and on Saturdays. The consistency of weekly taxi usage reveals a regularity that can help predict the demand - or lack thereof - of taxis. Determining these trends accurately ensures that transportation demand is fulfilled at all times without waste.

![Number of Yellow Taxi Trips Throughout 2014](image)

Figure 3.3: Pickup Times Throughout the Year of 2014
3.1.3 As a Function of Time of Day

Figure 3.4: Pickup Times Throughout the Day

The analysis of the Uber and TLC data included visualizing how pickups were spread out through each day. This analysis allowed for the understanding of how the demand for taxi transportation evolved throughout the day. Using the product of this analysis is crucial in selecting a policy that satisfies transportation needs of customers as their demand rises and falls throughout the day.

Both the Uber and TLC datasets follow a similar pattern as the day evolves. Between midnight and 7 a.m., there is a lull in the demand for transportation, with the nadir of transportation demand in the entire day occurring at about 3 a.m. for Uber pickups and just after 5 a.m. for yellow taxi pickups. After 7 a.m., the number of pickups rises as the workday begins, staying steady after 9 a.m.

For the case of Uber, the number of taxi pickups began to rise steadily at 2 p.m., peaking at 5 p.m. The number of pickups then fell only slightly afterward until about 10 p.m., after which pickups quickly decreased. On the other hand, TLC taxi pickups fell between 3 p.m. and 4 p.m. and rose after to a similar proportion as Uber pickups at about 6 p.m.

The graph of pickup times shown in Figure 3.2 reveals one of the most useful insights regarding the TLC’s taxi system and the dynamic between Uber and similar private taxi companies and the taxis overseen by the TLC, which was first identified by the blog iQuantNY.
The decrease in TLC taxis’ demand between 3 p.m. and 4 p.m. is perplexing, especially considering how Uber usage nearly peaks at around that time. In fact, it is a common phenomenon for New Yorkers where finding a taxi during this time period is nearly impossible. The reason for this decrease is in fact a practical consideration involving shift changes between taxi drivers (Wellington, 2015). The majority of taxi drivers who lease their car and medallion from a garage will share their taxi with another driver who will pick up their shift once the original driver’s shift has ended. The decline in the number of TLC taxi pickups at around 4 p.m. represents this shift change, as drivers go back to their garages as their shift ends. Once the other driver picks up the car and begins their shift, the observed demand for taxis increases once again.

Considering the Uber data’s perceived demand during that time period, it seems inefficient for the shift change to occur at around 4 p.m., but the current system of drivers sharing taxis is set up such that the two shifts are split up in such a way that the fares from each shift are relatively equal (Wellington, 2015). Thus, to achieve this even split, the shift change occurs at a time which creates a massive inefficiency in this taxi system.

The consequences of this inefficiency are evident. The number of TLC taxi trips jumps back up to its peak at around 7 p.m., though during the span before it the number of trips is considerably lower. Uber, meanwhile, has managed to turn this inefficiency into a benefit for itself. As seen in the graph of Uber’s pickups, the proportion of Uber pickups goes up significantly in a time when TLC taxi demands seem to decline. This seems to indicate that Uber, with its flexible shifts allowing Uber drivers to operate around 4 p.m. when demand is high, fills in for the lost demand that the TLC taxis forgo by choosing to change shifts at this time. Thus Uber, with its focus on remediating the inefficiencies of existing taxi systems such as this one evident among TLC taxis, was able to capture demand for taxi transportation that had otherwise gone unsatisfied. This example highlights how Uber has taken over the transportation world by storm by identifying and fixing inefficiencies among taxi systems in metropolitan areas.
3.2 Trip Distance

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{cumulative_distribution.png}
\caption{The Cumulative Distribution of Trip Distances for Yellow Taxis}
\end{figure}

In addition to the pickup locations and times of taxi trips, the length of these trips also plays an important part in prescribing a policy for these transportation systems. Using the past data on trip distances can allow for more accurate predictions of demands for future types of trips. This section will analyze the cumulative distributions of trips - both from the data from the TLC and Uber - to understand what kinds of trips have noticeably higher demands.

The idea of cumulative distribution functions is highly applicable to a task such as identifying the proportion of data that falls within a specific subset of the feasible set. Using a cumulative distribution graph in the case of analyzing the distances of taxi trips is appropriate since these graphs excel at displaying the number of trips that fall within a specified subset of distances (Swoboda, 2015).

Thus from Figure 3.4, it can be seen that over half of the yellow taxi trips were less than 2 miles in length. In fact, over 95 percent of New York taxi trips are less than 10 miles in length. These facts are extremely important in understanding the transportation needs of New Yorkers, who, living in the unique city of New York, presumably have different transportation needs than those living in other cities. Considering that the majority of trips
are less than 2 miles, then, it is plain to see that a transportation policy would have to take into account that most trips in New York are relatively short.

3.2.1 Cartesian Distance vs. Reported Distance

An interesting aspect of the data regarding trips is how the Cartesian distance of a trip - in other words, the length of a straight line connecting the origin and endpoint a trip compares to the reported distance of the trip. While the Cartesian distance is the length of the line connecting the beginning and endpoint of a trip, the reported distance is assumed to be the actual distance driven by the taxi from the origin of the trip to its terminus.

The discrepancy between these two numbers is highly valuable. Since obviously most trips do not consist of a straight line, the factor of the reported distance divided by the Cartesian distance produces a figure which highlights the circuity of trip. In other words, the factor illustrates how far from a straight line the trip diverged. This is important because it reveals how circuitous an area’s transportation network is. Generally, this factor hovers at approximately 1.2, which indicates that the reported distance is 1.2 times the length of the Cartesian distance (Kornhauser, A., 2015). However, if this number is abnormally high, it might indicate that an area’s road network is inefficient and could require a rethinking.

![Cartesian and Reported Distances for Random Sampling of New York Taxis](image)

Figure 3.6: A Linear Regression of Reported Distances vs. Cartesian Distances

In analyzing the TLC data, the average factor of all the trips within the TLC dataset
turns out to be 1.2715. This falls within a reasonable distance from the typically seen value of 1.2. A possible reason that the observed value was on the higher end could be due to the fact that, since New York is organized in such a grid-like fashion, the actual routes of the trips varies noticeably from a true straight line connecting the two endpoints of the trip. A random sample of the data is shown in Figure 3.5, with the Cartesian distance on the x-axis and the reported distance on the y-axis. The line of best fit tends toward the factor of 1.27. The fact that trips do not deviate from a straight line more than a factor of 1.27, though, makes the assumption made later on in the analysis portion of this thesis - that all the trips are assumed to be a straight line - more reasonable than if the reported distances were much longer than the Cartesian distances.

### 3.3 Trip Speed

Another point of analysis that becomes necessary during this thesis is the average speed of taxis when they are transporting customers. Due to the time element of modeling a ridesharing policy, it is essential to have a general idea of the speed of a New York taxi when it is transporting its passengers. As it is necessary to compute how the taxi traverses as time passes, having an estimate of the speed - taken from the data, preferably - allows for a better model. In addition to the necessity of calculating the average speed, it is also important to realize the importance of average speed in characterizing automobile transportation in an area. While a city that is sprawled over a larger area might have a higher average speed, the city of New York is the densest city in the United States in terms of population and a significantly constrained road network. Because of the reality of the density of New York, the average speed of New York automobiles are presumed to be considerably lower than other cities in the United States. This difference is highly important in that it means that this analysis relies heavily on a factor that is unique to every city, and thus additional analyses in other cities must take additional care to tailor the average speed to each individual city.

The computation of the average trip speed is fundamental to the eventual analysis of a ridesharing policy with hitchhiking, and is simply done by dividing the reported distance
by the time between pickup and dropoff. While this is a naive approach to determining the true behavior of taxis and transportation in general, the computation of this figure is more to aid the methodology undertaken later in this thesis than to determine the speeds of taxis at all times. Averaging this figure for all trips results in an average speed of 14.73 miles per hour for all TLC trips. The fact that this number is so low is not particularly surprising considering the stop-and-go nature of traffic in New York. This number is of utmost use when determining the exact time element of the ridesharing policy.

3.4 The Rise of Uber

Uber’s data for their trips in 2014, from the outset, shows how the demand for Uber trips is rising at a relatively quick rate. From April to September 2014, the total number of Uber trips rose by 82.13 percent, as the absolute number of Uber trips grew from 564,517 to 1,028,137 trips over a five-month span. Uber saw significant jumps in the number of its trips from April to May, as well from June to July and August to September.

Despite Uber’s emergence in the public eye, these numbers are still dwarfed by the number of trips handled by the TLC. In January of 2014 alone, the TLC has trip data for over 13 million trips, as was visible from the source data. However, considering Uber doubled its number of trips over a six-month span in 2015 to over one million, the rise of private for-hire taxi companies seems to present a significant threat to existing taxi systems. As shown in Section 3.1.3, the fact that Uber was able to pounce upon an existing inefficiency within New York’s taxi system reveals one reason for why Uber continues to become more popular in cities such as New York. As Uber continues to see growth - the number of Uber taxis exceeded yellow taxis in March 2015 - policymakers from the TLC will have to consider how to remain competitive in the taxi market.
### Table 3.1: The Rise of Uber in 2015

<table>
<thead>
<tr>
<th>Month in 2014</th>
<th>Number of Uber trips</th>
<th>Percent rise from previous month</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>564,517</td>
<td>N/A</td>
</tr>
<tr>
<td>May</td>
<td>652,436</td>
<td>15.57%</td>
</tr>
<tr>
<td>June</td>
<td>663,845</td>
<td>1.75%</td>
</tr>
<tr>
<td>July</td>
<td>796,122</td>
<td>19.93%</td>
</tr>
<tr>
<td>August</td>
<td>829,276</td>
<td>4.16%</td>
</tr>
<tr>
<td>September</td>
<td>1,028,137</td>
<td>23.98%</td>
</tr>
</tbody>
</table>

### 3.5 Conclusions

This chapter provided valuable insights on the state of the taxi market as well as a general idea of the entries in each of the data sets. Uber has clearly made a dent in the TLC’s hold on the taxi market, as it has grown exponentially and capitalizing in the inefficiencies of the TLC’s system. This was particularly obvious in the analysis of the trip frequencies over a single day. The fact that Uber was able to capitalize on a supply shortage in taxis in the middle of the afternoon - when demands might be in fact the highest all day - illustrated a deficiency in the existing system of taxi transportation. The substitution of Uber trips in place of this supply shortage led to a spike in the relative number of trips taken by Uber during this crucial late afternoon period. Such realizations have helped Uber grow from a burgeoning young company 5 years ago to the behemoth it is today.

In addition to the insights concerning the taxi market, certain figures calculated in this chapter prove to be of utmost importance in the ridesharing analysis portion of this thesis. The average speed of vehicles is valuable in that it allows for a better segmentation of the time element during the ridesharing analysis. The fact that trips average a speed of 14 miles per hour is crucial in determining accurate “time slices” to divide the trips within the dataset. The circuitry of trips - the relationship between their actual length and the Cartesian length between the origin and terminus points - also will allow for crucial assumptions later on in this thesis concerning the aggregate nature of taxi trips. The fact that trips do not deviate too significantly from a straight line ensures that the assumptions taken to apply certain algorithms are satisfied. As well, the cumulative distribution of trip lengths illustrates that 90 percent of New York trips are shorter than 5 miles, a fact which should
not be too surprising considering the denseness of New York City. The density and relative length of the trips also increases the opportunities for ridesharing, as potential hitchhiking opportunities would crop up more easily in more highly concentrated transportation areas. Moreover, this also illustrates that a practical ridesharing scheme could tackle the most highly trafficked areas first before moving on to the less active areas in terms of trips.
Chapter 4

The Taxi Market

In looking at the data and the results of the analysis of this thesis, another important aspect of this thesis comes into focus. The interaction between the public and private forces of the taxi market is a point of contention between multiple stakeholders in the market, as many different groups compete for market share in a market that has a limited pool of rides. In particular, the recent rise of Uber has drawn interest to the growing competition between the public and private sectors of the taxi market. On one side, there are the public taxi companies such as New York’s Taxi and Limousine Commission (TLC), which are agencies under the purview of local and state governments. Many of these agencies - such as the TLC - have existed for decades (TLC, 2016), and have provided much of the transportation services necessary for cities and municipalities until the dawn of the 21st century. On the other hand, there are also the private taxi services, many of which have cropped up in the age of the sharing economy. Most visible of the private taxi companies as of late is Uber, which provides taxi services in hundreds of cities across the globe. These two sides of the transportation economy, though providing similar or identical services, compete for the same pool of trips which are a function of consumer demand. The public side of the taxi market - operated by the TLC - has already been discussed in previous chapters. This chapter focuses more on the private sphere of the taxi market, which has rapidly gained market share within the past five years. Moreover, this chapter discusses how private taxi companies in recent years have left an indelible mark on the taxi market.
4.1 Uber

Without a doubt, one of the most important players in the rapidly changing taxi market is Uber. Uber uses its app to allow customers to submit requests for trips which are then given to taxi drivers in the surrounding area to fulfill. The company was founded in 2009 by Travis Kalanick and Garrett Camp, and began with an initial seed funding of $200,000 (Chokattu and Crook, 2014). The company quickly attracted the eyes of other investors, as Uber received an additional $1.25 million in seed funding in 2010 (Chokattu and Crook, 2014). By the end of 2015, Uber had managed to accrue a market value of somewhere around $62.5 billion, a staggering rate of growth for the young company (Newcomer, 2015).

Uber’s model focuses on using its proprietary application (also called Uber) to connect customers who want taxi trips and drivers who will provide such services for these customers. The application uses GPS to search for available drivers in the vicinity of a customer and then informs any potential drivers that there is a nearby customer in need of its services (Rosenblat, 2016). This application has become the focal point of Uber’s business model and has provided it with the strategic advantage over traditional taxi services by making the taxi searching process on-demand. This puts power in the hands of the consumer and allows them to create trips at their own whim.

Uber first began offering its services around San Francisco in 2011, but quickly expanded to New York, Chicago and Washington, D.C. (Chokattu and Crook, 2014). The company then began expanding its reach overseas by opening up its Paris operations in December of 2011. Uber has spread around the globe and now operates in hundreds of cities and metropolitan areas. Notable events in Uber’s timeline was its expansion into Asia, first in Singapore in January 2013, as well as into China July of 2014. Uber’s expansion into China, in particular, poses an interesting challenge for the company, where unregulated taxis in many places reign over the taxi market. Another significant expansion was into Africa, when Uber launched in Lagos, Nigeria, in July of 2014 (Chokattu and Crook, 2014). The fact that Uber was able to so quickly establish itself in so many regions across the globe speaks to not only a heavy interest among investors in expanding taxi services in many areas but also that existing taxi markets still held much potential for growth.
Uber’s offerings have also grown significantly since the company’s inception in 2009. While initially Uber vehicles were standard automobiles, since 2009 Uber has added UberX as a budget alternative, as well as UberXL for customers who desire larger vehicles such as SUVs (Chokattu and Crook, 2014). Uber continued to expand its product line by experimenting with the delivery services such as UberFRESH and Uber Rush, which were online food delivery and courier package delivery services, respectively. As mentioned in the introduction to this thesis, perhaps the most interesting product of Uber is UberPool, which began in August 2014 in San Francisco and allowed users to share rides with strangers for a reduction in fare (Uber, 2015). This service most closely approximates the reality of the policy applied to the TLC data in the analysis portion of this thesis.

4.1.1 Controversies

Uber has faced numerous controversies in its stratospheric rise into the business world. The biggest obstacle to Uber in many markets is its controversial legality. Existing taxi systems in many cities operate with much oversight from regulatory boards, and in many cases Uber uses loopholes to expand legally into new areas (Pathe, 2014). The primary concerns from many stakeholders is that Uber is a threat because its drivers might not be insured or licensed. In fact, the introduction of Uber in many cities has led to protests. In Germany, France, India, Italy, Denmark, Canada and many other countries, Uber’s entry into the market has led to mass protests from other taxi drivers who see Uber as a threat to the existing taxi businesses (Chrisafis, 2016). In particular, the fact that Uber drivers - with much less oversight and regulation compared to their public taxi-driving counterparts - do not have to face the same costs in terms of medallions and other expenses that existing taxi drivers had to contend with.

Uber’s surge pricing model, too, has led to significant criticism for its business model (Uber, 2016). Surge pricing is a business model employed by Uber which uses an algorithm to raise prices for taxis in response to an increase in the demand for taxis. This business choice allows supply and demand to match, as the number of drivers rises in response to surge prices and the number of customers who are willing to pay the additional fare falls until the supply accurately meets demand (Uber, 2016). While this allows Uber to
advertise more timely taxi services, it also has led to customer complaints at the significant discrepancy between normal prices and surge prices.

Uber has also been accused of attempting to sabotage the efforts of its competitors. Perhaps the most damning such event occurred in early 2014, when documents were released that showed that Uber employees in New York had sabotaged Gett, a newly established taxi service, by ordering taxi trips in large droves and then cancelling them. This effort helped doom the incipient Gett and helped establish Uber’s dominance in the New York taxi market (Segall, 2014). In addition to this form of sabotage, Uber convinced drivers from other services such as Gett and Lyft to abandon competitors’ platforms and join Uber’s. These documents sparked controversy and led to Uber issuing an official apology for its conduct (Uber, 2014).

4.2 Other Private Taxi Companies

Numerous other private taxi companies inhabit the same space as Uber, and compete with it and other entities for rides. Some of these other companies fulfill niches that Uber does not itself serve.

Lyft is one of the largest private taxi companies apart from Uber, and as of late 2015 was valued at $5.5 billion (Newcomer, 2016). Lyft operates similarly to Uber in that it also utilizes its app to connect people who require trips to drivers who own a car and can provide such a trip. The firm is distinguished from Uber and some other taxi startups in that Lyft emphasizes the community aspect of peer-to-peer ridesharing services (Newcomer, 2016).

Another large competitor in the early 2010s was Sidecar. Sidecar faced many of the same challenges as Uber in its early stages. Though the company received $10 million in Series-A funding and was approved for business in California, by 2015 the company had repositioned itself less as a taxi service and more as a same-day delivery service (Dickey, 2015). By the end of 2015, this service was not enough to maintain the costs of the company, and Sidecar’s assets were sold to General Motors (Dickey, 2015).

Other companies inhabit the the specific niche of ridesharing. Bandwagon, a ridesharing company, began providing its services in New York airports in 2014 (http://bandwagon.io).
Bandwagon, through its app, connected people arriving in these airports and allowed them to share a taxi to offset the taxi fare between the multiple riders. A company offering a similar service, named Hitch (http://hitchdc.com), also entered the market in 2014. Hitch allowed customers with similar destinations to share rides together, much like the ridesharing systems proposed in other theses.

4.3 The Sharing Economy and the Taxi Market

The new “sharing economy” has become a point of interest throughout the business world, as the hope of a new model for a business has piqued the interest of numerous stakeholders. The sharing economy, defined in one way by Juho Hamari, is the hybrid market model between owning and gift giving that allows members of an economy to share their goods between each other (Hamari et al., 2015).

The rise sharing economy has been enabled by a number of relatively new technological innovations (Hamari et al., 2015). Firstly, electronic marketplaces have created new avenues for sellers to find and connect with buyers, and vice versa. Secondly, mobile devices - in particular, the touchscreen phone - have allowed people to be connected to the Internet at all times and places. Such innovations have leveled the playing field, in a sense, as new business models such as those offered by Apple, Uber and other innovators allow nearly anyone to provide their goods and services to customers.

Proponents of the new sharing economy have touted its ability to bring together disparate communities as well as reduce consumption which can help ease the environmental impact of human activities. This new business model also creates flexibility and, in many cases, reduces the costs of things such as fares. However, others have criticized the sharing economy for skirting regulations and for weakening the bargaining power of many employees (Leonard, 2013). In many cases, companies such as Uber and Airbnb have tended to extract profits at the expense of many costs that exist for other businesses such as insurance and taxes. These controversial measures in a way allow many sharing economy startups to reduce their costs, though they also reduce their tax obligation to the government and society. Moreover, the sharing economy has been criticized as a makeshift solution for the
damaged economy following the Great Recession in 2008, and that many workers within this new business model do not have as many benefits as their predecessors (Leonard, 2013).

4.3.1 “Appification” and “Uberification”

The sharing economy as a whole has revealed a new tendency toward the “Appification” or “Uberification” of services. What is meant by this term is the creation of technological environments that allow individuals - rather than corporations - to create and provide their services to customers. In this sort of system, the creator of the environment itself does not have to produce any goods the customer directly consumes, but purely provides the means by which these customers can obtain their goods from developers and suppliers.

“Appification” is most evident in Apple’s creation of the App Store ecosystem soon after it launched the iPhone in 2007. The App Store allowed developers to create their own applications - with little input from Apple - and sell their creations on the App Store. In this system, Apple has the final say in allowing what is on the App Store, and also takes a portion of the revenue from selling the apps. The App Store, then, acts like a middleman between developer and customer that allows the two parties to connect and transmit goods and payment for those goods between each other. A similar phenomenon is visible in other mobile phone operating systems such as Android. Android operates a similar ecosystem as Apple’s, as it hosts the Play Store which allows developers to sell their applications made for the Android operating system.

In the context of Uber and other taxi companies, the “Uberification” of taxi services has become one of the defining aspects of the new sharing economy. While the line separating Uber and its drivers providing services is somewhat blurry, Uber seems to have created its own technological ecosystem that connects drivers and people who require taxis without itself providing the taxi service. This aspect of Uber, however, defines it as a member of the new sharing economy, where corporations such as Apple and Uber merely provide the avenue for individuals to share their goods and services for payment. It is likely that the “Uberification” of other services will continue to occur, as mobile technologies that are ubiquitous and continuously connected to the Internet allow sellers and buyers of all sorts of goods and services to communicate.
4.4 The Interaction of Public and Private Taxi Companies

The rapid rise of private taxi companies in the 2010s has created a vastly new taxi market. The meteoric growth of companies such as Uber have led to changes in how many existing companies operate. Moreover, the changing landscape of taxi transportation has had huge implications for not only taxi companies but in society as well.

In some ways, Uber’s and other private taxi companies’ emergences have illustrated some of the deficiencies of the existing taxi system. The most visible such failure is the reduction of prices for Uber and other private taxi services relative to the taxis operated by public entities in many cities. The reduction in prices for private taxis shows that there are ways to reduce the price for taxis without forfeiting care for the customer. However, such differences in prices could simply reflect the lack of regulation that Uber and other private taxi companies have in relation to their public counterparts. As policymakers change laws to reflect the rise of private taxi companies, the differences in prices between the two spheres could shrink.

Public taxi companies have adopted some of the innovations that Uber and other private taxi companies have pioneered. One of Uber’s most useful innovations has been how central its app has been for its services. The Uber app allows riders to hail taxis on-demand and have a taxi arrive within minutes of the request, making it much more convenient for customers than the traditional model of taxis. In response to the mass adoption of Uber, though, the New York TLC has created its own app, called the E-Hail app, which allows a customer to call for a taxi as long as they are within New York (New York Taxi and Limousine Commission, 2014). In this way, the emergence of private taxi hailing apps has spurred the creation of public ones.

The effects of Uber’s and other companies’ rise are visible in the taxi medallion market as well. The taxi-financing market - which provides loans to taxi drivers so that they can purchase or rent taxi medallions - has dropped 50 percent in value since 2013 (McLannahan, 2015). Moreover, taxi medallions themselves have dropped drastically in price. A New York taxi medallion, once valued at around $1.3 million in 2013, has a value of around $700,000 now. A similar effect occurred in Chicago, where medallion prices dropped by 33 percent,
from $360,000 to $240,000. The presence of companies such as Uber have thus severely disrupted the existing market not only of taxi medallions, but the companies that finance the purchases of these medallions.

In conclusion, the emergence of companies such as Uber have put a competitive strain on existing taxi companies, but they also have spurred innovation that likely would have not occurred had the private sphere not exerted its competitive pressure. As the taxi world continues to evolve, it remains to be seen how the various players in the taxi market will divide the fairly static pool of taxi rides.
Chapter 5

Ridesharing

This thesis focuses on the application on the idea of ridesharing on the datasets provided. However, to fully grasp the extent of impact of ridesharing on taxi networks, it is first necessary to fully understand the language and dynamics underpinning ridesharing. This chapter begins with a summary of the terminology concerning ridesharing, including transportation metrics that allow for the assessment of the savings of ridesharing policies.

Importantly, the idea of the pixelization of the areas in question into pixels that are 0.1 miles by 0.1 miles in size simplifies the analysis of these incredibly complex taxi networks. However, this simplification of the data in question does not come at the expense of too much precision, as the use of the pixel still allows for a good approximation of the origins and terminal points of trips. Moreover, it will be very enlightening to take a cursory look into the most active voxels for yellow taxis operated by the New York TLC. These voxels, which are clustered around transportation hubs such as Penn Station and John F. Kennedy International Airport, have an enormous impact practically on what kind of policy should be implemented.

In addition to this summary of the most active trips from the TLC data set, this chapter includes a comparison with the trip activity for Uber rides. This comparison can be enlightening in a number of ways. Firstly, as shown in Chapter 3 with the comparison of trip activity throughout the day, Uber’s continued operation of taxis at the prime time period of the late afternoon has paid off hugely for the company, as TLC taxi drivers’ tendencies to change shifts at this time had led to a sizable amount of unsatisfied demand
for taxi trips. Similarly, observing the most common pickup areas for Uber rides can show where the TLC’s level of service is lacking. Secondly, a comparison of the two data sets can provide insight into competition between the private and public levels of taxi service for the areas that both TLC and Uber vehicles have high activity in. While the number of Uber rides still pales in the face of the corresponding TLC figures, a detailed analysis of these pixels with high activity can show the effects of increased competition in the transportation market.

This chapter thus focuses not only on enumerating upon the concepts and terminology of ridesharing, but a summary and analysis of the pixels with the most activity in their respective areas.

5.1 Ridesharing Terminology

5.1.1 Policies

This thesis has until now referred to the idea of applying policies to systems. By the definition provided by Warren Powell, a policy is “a rule (or function) that determines a decision given the available information in state $S_t$” (Powell, 2015). A policy, with regard to this definition and the context of this thesis, is then a function that would be applied to the state described by the given data sets. In other words, a ridesharing policy is a policy function approximation that assigns actions to specific instances in the data set. If certain conditions are satisfied for a particular set of taxi customers within the data set, then some change is made affecting the transportation of these particular customers.

5.1.2 Ridesharing Metrics

The savings accrued from applying the specified ridesharing policies to the data can be measured in three different ways: average vehicle occupancy (AVO), the reduction in the usage of taxis ($\%\text{TaxiRed}$), and the vehicle miles saved as a percentage ($\%\text{vMiles}$). As mentioned briefly in the introduction, the AVO is an accurate summary of the approximate efficiency of a transportation system, as this is due to the fact that, if properly implemented, a ridesharing policy with hitchhiking could only have an ADO of one, though the number
of occupants in the taxi will continue to evolve as the taxi picks up and drops off passengers at their respective terminal locations. \( \%\)TaxiRed describes the percentage of fewer taxis needed to satisfy the demands of customers shown in the data. Moreover, as mentioned in the introduction of this thesis, one of the significant benefits of ridesharing is the reduction in number of miles driven by vehicles. \( \%v\)Miles captures this reduction by showing, as a percentage, how many fewer miles are required to satisfy customer demand.

### 5.1.3 Pixelization and Voxelization of Areas

To begin the analysis of the taxi networks in question, it is imperative to simplify the vast amounts of data into more easily manipulated data structures that allow for much quicker computation without sacrificing too much precision. The data structure that is used in much of this thesis is the pixel or the voxel. The concept of pixels and voxels has been used extensively in other applications of transportation dynamics at Princeton University, where it has allowed for an amazing understanding of the world of transportation, as seen in various applications and theses (Swoboda, 2015).

A pixel is a point, an area which is the smallest unit for analyzing images or other areas of interest. The size of the pixel can be varied to add or decrease the complexity of an area of interest. The larger the pixel, the more simplified a pixel area becomes. In the case of this thesis, the pixel is defined as a square area that is 0.1 miles by 0.1 miles (or 528 feet by 528 feet) in size, as done by A.J. Swoboda in his own thesis (Swoboda, 2015). The areas that will be analyzed in this thesis are to be divided into these pixels. The reason for doing so is that, without a data structure such as the pixel, there are a near-infinite number of latitude-longitude coordinates in the areas in question, needlessly creating complexity in our analysis. The size of the pixels is such that it prevents customers of taxi services from walking more than 0.1 miles in any direction to be picked up by a taxi. The size of the pixel is thus a compromise which allows for a much cleaner analysis of the data while also limiting the hassle for the consumer. Moreover, for the purposes of the analysis in this thesis, the taxi is assumed to leave from the centroid of each pixel.

In addition to the pixel, there is also the voxel. The voxel, short for volumetric pixel, is a three-dimensional object as opposed to a two-dimensional pixel. Since this thesis
incorporates the time dimension in addition to the two spatial dimensions, the application of ridesharing with hitchhiking requires the use of three-dimensional voxels (with two spacial dimensions and one temporal dimension). The idea of voxelization, while touched upon in this chapter, is the focus of the ridesharing methodology elaborated upon in the next chapter.

The New York area - including all the five boroughs - constitutes a roughly square region that is shown in Figure A.1.1. A smaller inset of the pixelated New York is shown in Figure 5.1. For the purposes in this chapter, the New York was subdivided into a grid of pixels, with 293 pixels in each row and 290 pixels in each column, for a total of 84,970 pixels. Of these 84,970 pixels, only 46,419 pixels (about 54 percent) contained at least one origin or dropoff point for a trip in 2014. The "origin" of the pixelization was the lat-long coordinate (40.491825, -74.260186), which corresponds to an area off the southwestern coast of Staten Island. Each trip’s origin and destination were thus pixelated using this lat-long coordinate as a reference. The calculation of the x-pixel and y-pixel for each lat-long coordinate was as follows (where ⌈·⌉ indicates the floor operator):

\[
X = \lceil (Long + 74.260186) \times 25 \times 21.7814 \rceil
\]

\[
Y = \lceil (Lat - 40.491825) \times 25 \times 27.64 \rceil
\]
Of these 46,419 active pixels, 23,171 pixels had fewer than 20 pickups or dropoffs over the course of the entire year. In fact the top 1 percent of pickup pixels have over 60.2 percent of rides originating from them, illustrating how trips are concentrated in a few high-activity areas.

5.2 Analysis of Most Active Pixels in New York

An important point of understanding in this thesis is an analysis of the most common areas of pickup in the cities of interest. By analyzing these points of interest, implementations of ridesharing services can, from a practical standpoint, understand what locations can be used as hubs for ridesharing. Continuing from the issue of practicality, a taxi company could feasibly roll out a ridesharing service starting only in these major transportation hubs. This would reduce the complexity of rolling out a ridesharing policy more broadly across the area of interest, especially considering the fact that a small portion of trips originate from locations where no other trips originate.

5.2.1 TLC Data

According to analyses of previous years of data concerning the most active pixels in New York, clear trends began to emerge. As mentioned previously, trip pickups and dropoffs are clustered and the majority of them come from less than 1 percent of the top pixels by trip volume. The TLC dataset from 2014 is no different. Places such as Penn Station are highly active locations for pickups and dropoffs, as they act as transportation hubs for not only automobiles but for trains and buses as well. The following tables contain the pixel number, the latitude and longitude of the pixel, the number of pickups or dropoffs from that pixel and the context of the location.

The data from 2014 shown in Table 4.1 shows similar results as the data from 2013 which was enumerated upon in A.J. Swoboda’s thesis (Swoboda, 2015). The most common pickup locations consistently seem to be the areas around Penn Station and the Port Authority as well as Grand Central Terminal. Furthermore, the airports are also extremely common areas for trip pickups. These areas roles’ as transportation hubs - not only for taxi transport
but for rail and bus as well - illustrate the importance of hubs in the application of any sort of ridesharing policy. If a policy hopes to succeed, it should naturally focus first on these transportation hubs as the most important areas to roll out a ridesharing policy. The fact that the most active pixels have barely changed from the previous year of 2013 also shows that transportation demand does not evolve rapidly, but that many of the same trends visible in 2013 are also evident in 2014’s data as well. If any change did occur, it would be safe to say that it would occur at a relatively slow rate. This is significant as the findings of this thesis are more useful if transportation demand does not drastically change over the next few years.

Table 5.1: The Most Popular Pickup Areas for TLC Taxis

<table>
<thead>
<tr>
<th>X-Pixel</th>
<th>Y-Pixel</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Pickups</th>
<th>Approximate Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>146</td>
<td>178</td>
<td>-73.9921</td>
<td>40.7523</td>
<td>1,450,792</td>
<td>Penn Station 7th Ave. Entrance North</td>
</tr>
<tr>
<td>144</td>
<td>179</td>
<td>-73.9957</td>
<td>40.7537</td>
<td>752,650</td>
<td>Penn Station 8th Ave. Entrance North</td>
</tr>
<tr>
<td>260</td>
<td>105</td>
<td>-73.7827</td>
<td>40.6452</td>
<td>727,942</td>
<td>John F. Kennedy Airport</td>
</tr>
<tr>
<td>146</td>
<td>182</td>
<td>-73.9921</td>
<td>40.7581</td>
<td>697,318</td>
<td>NY Port Authority Bus Terminal</td>
</tr>
<tr>
<td>216</td>
<td>191</td>
<td>-73.8635</td>
<td>40.7711</td>
<td>630,293</td>
<td>LaGuardia Airport</td>
</tr>
</tbody>
</table>

The most common dropoff areas seem to be heavily centralized in the Midtown area of Manhattan. The five most common dropoff areas are located around Penn Station and Grand Central Terminal, pointing to the areas’ significance as a major transportation for not only Manhattan or New York but for the entire Tri-State Area, as Penn Station is an entry point for New Jersey Transit trains, and Grand Central Terminal is also a hub that connects New York to much of suburban New York as well as Connecticut.

Table 5.2: The Most Popular Dropoff Areas for TLC Taxis

<table>
<thead>
<tr>
<th>X-Pixel</th>
<th>Y-Pixel</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Pickups</th>
<th>Approximate Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>146</td>
<td>178</td>
<td>-73.9921</td>
<td>40.7523</td>
<td>1,287,080</td>
<td>Penn Station 7th Ave. Entrance North</td>
</tr>
<tr>
<td>144</td>
<td>178</td>
<td>-73.9957</td>
<td>40.7523</td>
<td>754,586</td>
<td>Penn Station 8th Ave. Entrance South</td>
</tr>
<tr>
<td>146</td>
<td>177</td>
<td>-73.9921</td>
<td>40.7509</td>
<td>551,890</td>
<td>Penn Station 7th Ave. Entrance South</td>
</tr>
<tr>
<td>153</td>
<td>180</td>
<td>-73.9792</td>
<td>40.7552</td>
<td>537,979</td>
<td>Grand Central Southwest</td>
</tr>
<tr>
<td>154</td>
<td>179</td>
<td>-73.9774</td>
<td>40.7538</td>
<td>529,468</td>
<td>Grand Central Northwest</td>
</tr>
</tbody>
</table>
5.2.2 Uber Data

As seen in the heatmap in Chapter 2, the most common areas of pickup for Uber vehicles are John F. Kennedy International Airport and LaGuardia Airport. Strangely, one extremely common pickup point was at Union Station in Brooklyn. The exact significance of this finding is not known. However, the fact that the airports are the most popular places for taxi pickups is corroborated by the data, as shown in Table 4.2.

Table 5.3: The Most Popular Pickup Areas for Uber Taxis

<table>
<thead>
<tr>
<th>X-Pixel</th>
<th>Y-Pixel</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Pickups</th>
<th>Approximate Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td>105</td>
<td>-73.7827</td>
<td>40.6452</td>
<td>37,254</td>
<td>John F. Kennedy Airport</td>
</tr>
<tr>
<td>211</td>
<td>194</td>
<td>-73.8011</td>
<td>40.6365</td>
<td>21,746</td>
<td>LaGuardia Airport</td>
</tr>
<tr>
<td>138</td>
<td>171</td>
<td>-73.9847</td>
<td>40.6754</td>
<td>20,524</td>
<td>Union Station, Brooklyn</td>
</tr>
<tr>
<td>263</td>
<td>106</td>
<td>-73.8011</td>
<td>40.6385</td>
<td>16,337</td>
<td>John F. Kennedy Airport</td>
</tr>
<tr>
<td>216</td>
<td>191</td>
<td>-73.8635</td>
<td>40.7711</td>
<td>15,908</td>
<td>LaGuardia Airport</td>
</tr>
</tbody>
</table>

5.2.3 Comparison of TLC and Uber Data

Though there are a few pixels which are the most common pickup points in New York for both data sets of TLC and Uber, there are still a few notable differences.

Observing the analyses of these two datasets shows significant similarities between the two groups. Both TLC and Uber taxis most commonly originated from a few areas, which include John F. Kennedy International Airport and LaGuardia Airport. The fact that both TLC and Uber vehicles most commonly come from these areas also points to how Uber is taking away a significant portion of the TLC’s most popular pickup areas, and that if Uber’s rise continues, the TLC would continue to see a decrease in the number of pickups from these important locations. Notably, there are a few differences. One of the most common Uber origin points was at Union Station, Brooklyn. This varied from the TLC dataset, as this area was not one of the most popular origin points in the TLC dataset.

The question arises as to what these differences might indicate. For a company like Uber, the differences in the most popular areas could reveal an opportunity to continue to eat away at the hold the TLC once had on the taxi market. If Uber targeted customers who were more likely to require a ride from an airport, then presumably it could steal customers
who would have liked to use a TLC yellow taxi instead. Moreover, providing Uber taxis in the first place in these areas should be extremely important for any taxi company that hopes to compete in the increasingly competitive marketplace for taxi rides. The fact that Uber sees most of its rides at the airport also might hint at how Uber vehicles are more likely to be used in areas that are not as dense as Midtown Manhattan. If the TLC can identify areas which it continues to have high ridership despite Uber’s interference, it can leverage this competitive advantage and focus its efforts in areas that are not as well covered by Uber. In this way, yellow taxis and Uber taxis could coexist in the market by focusing on areas that the respective groups have a comparative advantage in.
Chapter 6

Applying Ridesharing Policies to Data

This chapter begins the focus of this thesis in the application of the ridesharing policy outlined in the previous chapters to the datasets provided. In doing so, the hope is to determine the savings that can be obtained by these policies, as defined by using the metrics enumerated upon in the previous chapter.

While previous theses have very comprehensively applied a simplified ridesharing policy to data, these analyses have tended to gloss over the analysis of the pickup portion of taxi ridesharing. Rather than having one common pickup location for multiple customers of a taxi, this thesis is broader in scope in that it accommodates multiple pickups across a number of different locations. This inclusion of hitchhiking will necessarily require a significant modification of the methodology in previous research concerning the application of ridesharing policies.

This chapter outlines a complex ridesharing policy that incorporates the idea of “hitchhiking” as was previously elaborated upon. The policy will then be applied to the New York taxi dataset provided by the TLC, and the results are analyzed using the efficiency metrics mentioned in previous chapters.
6.1 Methodology

6.1.1 Identifying Ridesharing Opportunities

The crux of applying the specific ridesharing policy described in this thesis is how to handle the multiple pickups that would be necessary with hitchhiking. While previous work began with one common pickup point, the added complication of hitchhiking requires a significant modification of the ridesharing analysis. Unlike with the case of the single pickup, multiple pickups require careful calibration of the time element to incorporate the fact that taxi customers will not want to wait too long along the way to their route to allow other customers to join their trip. The time element also becomes incredibly important in knowing where the taxi is at what point in time, which is crucial in determining whether a customer actually represents an opportunity for ridesharing or if he or she is not in the right place or time to accommodate the current trip.

To begin modeling the new ridesharing policy, it is imperative to determine how to find opportunities for a simulated taxi to pick up additional passengers along its initial route. In other words, how can a policy be created such that a simulated taxi would pick up a passenger whenever its route intersects with the location of the passenger? The division of the data involved into voxels aids this analysis. By subdividing the areas of interest into voxels, it is much simpler to analyze the number of trips originating or terminating around a specific area, with little loss of precision.

Because of the very nature of identifying additional pickup opportunities, time now becomes a significant element of the ridesharing analysis. Time indeed becomes an issue because, as a specified taxi moves from its origin to its terminus, the question arises as to whether a taxi is in some pixel within a certain time frame. The addition of the third dimension of time allows for accounting for this concept, as the line drawn between the centroids of different voxels goes through the time dimension.

A useful productive method of approaching the issue of time is by grouping trips into 30-second intervals. Separating the gigantic trip file into individual 30-second intervals simplifies the problem. The question arises as to how a 30-second timeframe was chosen for the purposes of this analysis. This is where the average speed of a New York taxi comes
into play. Indeed, as seen in Chapter 3, the average speed of a New York taxi is 14.73 miles per hour. And since pixels are squares with side length 0.1 miles, this means that a New York taxi traverses the length of a pixel in approximately 24 seconds. Rounding to a more easily handled number results in the 30 seconds that is used for the rest of this analysis.

6.1.2 Creating a Voxel Environment

The difficult task now is to model the intersections of simulated taxis with customers who could be picked up by a certain simulated taxi. One way of understanding the modeling of such a policy is by having a three-dimensional graph with the x- and y-axes representing the latitude and longitude, respectively, of specific locations, and the z-axis being time. In such a scenario, each customer waiting to be picked up would be represented by a “filament” of a prescribed radius that would span vertically at the customer’s location. The radius of the filament, then, would represent the maximum distance the customer is willing to walk to reach a taxi, and the vertical nature of the filament indicates that the customer stays in the same location as time progresses. Meanwhile, the taxi would be represented by a line spanning three-dimensions. Thus, by imagining the ridesharing simulation by this “filament” data structure, it is possible to find the intersection points which are the opportunities for pickups during a trip. While such an analysis requires tuning from a practical perspective, for the theoretical purposes of this thesis, it provides a good point to begin analysis.

There a few possible ways to tackle the issue of including time. The first potential idea is to model the ridesharing area not just by two dimensions of space, but with the time dimension on the z-axis. Doing so would create the three-dimensional model that was briefly mentioned previously in this chapter. This model, rather than having two-dimensional pixels, would have three-dimensional voxels, with the distance x- and y-axes being the pixel sizes described in the preceding chapters and the z-axis of time having a pixel length that would be described in minutes or seconds. An example of this voxel environment model is shown in Figures 6.1 and 6.2. In the 3-by-3-by-3 voxel environment shown in Figure 6.2, the x- and y-axes are subdivided into 0.1 mile intervals, and the z-axis is subdivided into 30 second intervals.
Figure 6.1: Example of the Voxel Used In Ridesharing Analysis

Figure 6.2: Sample of a Voxel Environment with Two Voxels Highlighted
For the application of a ridesharing policy to this huge amount of data, there are necessarily some simplifications that must take place along the way. A central assumption that will simplify some of the analysis during this portion of the thesis is the simplification of every trip into a straight line between the origin and terminus pixels. While this simplification is wholly unrealistic and takes out a large element of complexity from the data, it is necessary in order to apply the policy to the data. Because requiring the trip to go along the specified route to its destination, a huge amount of computational power would be needed to route each of the approximately 160 million trips in the TLC dataset. In other words, trying to understand the route every taxi took would create such an overwhelming complexity that it would be nigh impossible to model it using the methods described in this thesis. While algorithms exist that would allow for calculations with straight-line trips, such efficient and clean algorithms do not exist for the exceedingly complexity of routed trips.

6.1.3 Bresenham’s Algorithm

The assumption of each trip as a straight line encourages the use of Bresenham’s line algorithm. Bresenham’s line algorithm has a wide range of applications in computer graphics, as it provides the basis for computers’ abilities to draw lines to screens (Flanagan, 2016). The algorithm is useful for the purposes of this thesis because of its ability to determine the pixels that are involved in each straight-line trip. Bresenham’s algorithm could be applied to the previously described voxels to determine which areas were traversed by a taxi and at what time the traversals occurred. To be more specific, in a pixel environment and given two pixels, Bresenham’s algorithm provides all the pixels that are intersected in the straight line that connects the initial two pixels. Clearly, Bresenham’s algorithm seems to provide inroads into how to solve the complex puzzle of ridesharing. By assuming that each trip is a straight line, it would then be possible with Bresenham’s algorithm to determine which pixels a taxi goes through during a trip. Then, if the taxi happens to be in a pixel within a specified timeframe, then a ridesharing opportunity exists.

An example of an application of Bresenham’s algorithm to the voxel environment shown in Figure 6.2 is now shown in Figure 6.3. The grey voxels that were highlighted before
represented either the origin or the destination of the trip. Bresenham’s algorithm identifies all the voxels traversed by the straight line between the initially highlighted voxels. In the case of this simple example, the only newly highlighted pixel is the pixel in the middle of the 3-by-3-by-3 block, as seen in Figure 6.3.

![Figure 6.3: Voxel Environment with Line Drawn Between Origin and Destination](image)

Further modifications were made to make the actual analytical portion of this thesis feasible. To allow for a divide-and-conquer approach to the massive amount of data, present, the dataset was split into 365 separate files each containing all the trips that began on every day of the year. Each trip was also given a flag that indicated whether it was already taken into account within the ridesharing scheme.

### 6.1.4 Computing Angles Between Vectors

To aid in the process of identifying ridesharing opportunities, it becomes necessary to find trips in voxels that have destination pixels close to each other. To insure that this is the case, throughout this analysis finding the angle between two vectors becomes a must. As any student of linear algebra learns, finding the angle $\theta$ between two vectors can be computed
as follows:

\[ \theta = \arccos \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| \cdot |\mathbf{b}|} \]

### 6.1.5 Ridesharing Algorithm

The implementation of the algorithm which finds hitchhiking opportunities thus works as follows. A trip is chosen from the data and its flag was observed to determine whether it had already been taken into account within the ridesharing simulation. If the flag was marked, in other words, if the value was 1, then no further analysis was necessary. If not, when the flag was marked 0, then the algorithm continued by using the Bresenham line algorithm to determine all the voxels that the trip passed through from its origin to its terminus. The methodology of the analysis then involved iterating through each of these voxels and identifying the potential ridesharing opportunities.

![Figure 6.4: Example Voxel Environment with Time Slices](image)

This meant finding all the trips within each voxel which had straight line to its a terminus pixel that deviated no more than 5 degrees from the line to the voxel of the initial trip, which creates a sort of “cone” around the original trip’s path, as shown in Figure 6.5. Any trips that fall within this cone are potential ridesharing opportunities. This computation is made possible by computing the angle between the vectors describing the two trips. The flags of these newly added trips were changed from a 0 to a 1 to indicate the trips were
accounted for, and the algorithm would continue to do this for all the voxels traversed by the original trip. The algorithm would then move on to the subsequent trip until all trips were accounted for.

Figure 6.5: Example Voxel Environment with Time Slices and Trip Boundary “Cone”

6.1.6 Calculating Efficiency

In showing the effects of the ridesharing policy with hitchhiking on the data, it is necessary to reiterate the efficiency metrics used as well as how they are calculated within the analysis.

The focal point of calculating efficiency in this analysis is the average vehicle occupancy (AVO). This figure provides a comprehensive figure of the effect a ridesharing service can have on the efficiency of a transportation system.

The number of vehicles saved (%TaxiRed) is merely calculated by finding how many trips were grouped together. As every grouping of different into a single taxi represents the accumulation of a number of separate trips, each with their own taxi, then the number of taxis was reduced by the difference between one and the number of trips within a grouping. Thus, taking into account the entire dataset, %TaxiRed can be calculated by finding the difference between the number of groupings found by the ridesharing policy and the actual number of trips in the dataset.
6.2 Results

Applying the algorithm to find ridesharing opportunities with hitchhiking revealed promising results. The application of the complicated algorithm to the extremely large dataset, while taking a significant amount of time, produced results that illustrate the effectiveness of using a ridesharing policy in TLC yellow taxis. The application of the ridesharing policy grouped together rides into the same trip, from which the total number of trips could be simply observed by finding the highest trip number. The ridesharing application produced results that were easily understandable through the presentation of the few valuable efficiency metrics.

The first result noticed from the output of the algorithm was the reduction in trips. While the total number of trips is equal to the total number of trips within all the dataset, the total number of trips following the application of ridesharing was much, much lower. The value of %TaxiRed was calculated to be 11.83% over the course of the year. This simple and relatively naive calculation involved observing the difference between the number of trips found through the policy and the total number of trips in the original source data. The total number of trips using a ridesharing policy with hitchhiking was found to be 141,789,662. This is significantly smaller than the total number of trips in the original trip file, which numbered at 160,813,953. Hence, the computed figure of 11.83% fewer taxis shows the potential impact of the ridesharing policy on New York’s taxis.

The calculation of AVO revealed similar results. The AVO for the entire year of taxi trips under the new hitchhiking policy was 1.91. The savings from such efficiency gains are potentially immense. This represents an amazing improvement over the initial AVO which was by definition 1.

6.3 Analysis and Conclusions

The algorithm described in this chapter that encompassed this idea of ridesharing with the addendum of hitchhiking, in theory, has huge potential for creating efficiency within the transportation space of New York. This theoretical idea is backed up by the evidence of applying the ridesharing policy to the actual data describing all the yellow taxi trips taken
in New York in 2014.

The finding that the ridesharing algorithm did have an effect on the efficiency of the system is not altogether unsurprising, but it actualizes the potential of such a policy on real data. The computation of exact figures for %TaxiRed as well as AVO show the vast effects a ridesharing policy with hitchhiking could have on yellow taxi trips. By grouping together these trips by offering riders incentives such as reduced fares, the taxi companies can also benefit by reducing the number of vehicle miles driven by their taxis. With these findings in hand, taxi companies can understand the benefits of ridesharing systems through an exact quantification of the savings made through ridesharing. As Uber continues to offer new services such as UberPool, these findings will be ever more useful in determining exactly how ridesharing can help a transportation system.
Chapter 7

Limitations and Conclusions

7.1 Limitations

Some of the same limitations that have affected previous investigations of ridesharing policies persist in this thesis. However, other new considerations have to be taken into account when observing the results of the analysis presented in this thesis, and these limitations have to be taken into account when observing the results of this thesis.

7.1.1 Assumptions

Undoubtedly the most significant assumption made in this ridesharing analysis is the assumption that the path between pixels is a straight line. This assumption was vital in simplifying the analysis such that it did not require complex routing methods for each of the trips, and because it also allowed for accommodating Bresenham’s line algorithm. However, this massive simplification creates a rift between the ridesharing model and reality. Since obviously almost no taxi trips are straight lines, this simplification necessarily leads to a case where the model diverges significantly from the reality of taxi trips. While no model is a perfect representation of the true series of events it hopes to imitate, assuming all trips are straight lines is nevertheless a significant deviation from the actual nature of taxi trips.

Going along with this assumption was the idea that all trips are moving at a constant average speed which generally approximates the average speed of taxis in New York. This
assumption has a variety of shortcomings, the first being that, of course, taxis - or any form of transportation for that matter - do not move at a constant rate, but their speed and direction varies as a function of time. The added layer of complexity involved in finding routes for each trip, however, was too complex and computationally intensive, and thus the simplification into trips maintaining a constant average speed were utilized.

Yet another assumption made with regard to the model was that the taxis did not have a limit for the number of occupants. Since hitchhiking opportunities occur sparingly relative to the entirety of taxi trips, there were few instances within the ridesharing model where the number of ridesharing opportunities exceeded four or five. Even still, this issue with the data can be framed more as an opportunity rather than a limitation. A ridesharing service could potentially used expanded vehicles - more in the vein of a bus rather than a typical taxi cab - that provides a certain level of service in a ridesharing framework. Nevertheless, in the face of services such as UberPool, it stands to reason that this limitation has to be taken into account considering the results.

While more of a simplification than an assumption, the pixelization process removes some of the granularity in the data. Though pixelization assisted in removing unneeded complexity, it also makes the assumption that each trip originates from the centroid of each pixel, a minor modification which could have larger ramifications. In addition, pixelizing the trip data also makes the assumption that every person is willing to travel a distance of 0.05 miles to reach a taxi, a minor inconvenience that nevertheless will dissuade some people from using a ridesharing service. Aspects of this thesis such as this must be considered in the practical application of any theoretical model.

7.1.2 Issues with Data

The chapter concerning the data used within this thesis revealed that there were a significant number of entries within the data sets that were erroneously recorded. Because of the errors that are inherent in recording hundreds upon millions of different elements of data, there is a concern that not all the errors contained in the data were weeded out. While around 3 percent of the data - which were somehow recorded in error - were removed, there is a chance that other errors persist within the data. These persistent errors have small but
meaningful effects on the results, as trips that are wrongly input into the data can create incorrect impressions within the model.

Considering there was a noticeable percentage of trips which were erroneously recorded, it also stands to reason that there were also trips that occurred but were never recorded at all. This portion of trips - and the lack of them - also could obscure the true picture of taxi transportation dynamics in New York in 2014. Undoubtedly, with the uncertainty coming from the imperfections of the recording devices, there is a significant chance that some trips were left unreported. Accounting for these trips could also skew the results of the data in a different way.

7.1.3 The Changing Landscape of Transportation

With Uber revealing itself to be an extremely disruptive force within the taxi market, there is no doubt that transportation dynamics will continue to change in the coming years. As environmental concerns continue to grow, for example, demand for different forms of transportation will rise and fall. In addition, new forms of transportation can become commonplace. Numerous new innovations in the transportation world have changed how people use transportation. One notable such example is the CitiBike project in many metropolitan areas around the world. The introduction of CitiBike in New York might have had consequences on taxi and automobile demand in unforeseeable ways, and as CitiBike and other transportation innovations appear in New York, demand for taxis - with or without ridesharing - could change in the coming years.

The introduction of other technological innovations such as the self-driving car could also call into question the validity of this thesis’s results. The self-driving car, while still not a complete reality, is becoming less and less of a dream, as Google and other companies have fought to create their own self-driving cars that are becoming more adaptable to various conditions and situations. The introduction of a fleet of self-driving taxis could drastically affect the landscape of transportation demand, and attempting to use the results of this thesis without considering the future context of transportation demand would only result in less than accurate findings.

More generally, societal trends such as migration make the transportation world incred-
ibly dynamic, and thus the trends identified within this thesis could become obsolete even in the near future as people move around. Societal phenomena such as gentrification lead to movement among people that may raise the demand for transportation in specific areas and decrease it in other areas. To account for these changes, the analysis in this thesis must be applied to data sets from future years to observe similarities and changes. Only by reasoning with the changing world of transportation demand and understanding the results within the current context of 2014 can the results of this thesis be most accurately understood.

7.1.4 Extension to other Cities and Environments

While the city in focus for most of this thesis was New York, undoubtedly other cities will have to consider how different transportation policies will affect their own societies. Each city is unique, and New York is no better example of this fact. With New York’s population so densely compacted in such a small area, the findings of this thesis will be much more difficult to apply to other cities with urban sprawl such as Los Angeles. However, other cities such as Tokyo, Japan, have even higher population densities, and whether ridesharing would be worthwhile in these hyper-dense areas also remains to be seen.

New York, with its urban planning system that splits many areas into grids, also differentiates itself from many other cities in the world, where cities sprang up before urban planning became a focus for city builders. The results obtained here must therefore also be understood within the context of New York’s grid plan. Manhattan itself is unique because, as an island, automobile movement to and from Manhattan is restricted to bridges such as the Brooklyn Bridge and the Queensboro Bridge. These bridges, which funnel hundreds of thousands of commuters into New York every morning, are noticeable bottlenecks that differentiate New York from many other cities. While a ridesharing system might succeed in New York based on the results of this thesis, whether these results would occur in other cities would require further analysis.
7.2 Next Steps

Further enhancements can be made to the methods outlined in this thesis to create a more accurate rendition of the results of ridesharing.

7.2.1 Routing

As mentioned previously in the assumptions taken in this thesis, the use of straight line routes, while immensely useful in handling the complexity of the data involved, was not an accurate picture of the reality that the model hoped to analyze. While no doubt exceedingly complex, including routing - in other words, identifying the pixels each trip traverses on its actual route, rather than the pixels on the straight line connecting the pixels - would result in a more accurate analysis of the transportation demand observed within the data.

A ridesharing analysis with routing would determine exactly which pixels a taxi traverses between its actual trip between its origin and terminus. By identifying these pixels, then, ridesharing opportunities can be found by observing if customers are located in the traversed pixels within a certain time frame.

7.2.2 Additional Ridesharing Policies

The ridesharing policy of hitchhiking outlined within this thesis is not the only possible policy that could be applied to New York’s transportation system. In fact, Uber’s own UberPool is a slightly different form of ridesharing with hitchhiking that involves a constant stream of pickups and dropoffs rather than a single pickup and a single dropoff.

A significant change a ridesharing policy with hitchhiking could undertake is allowing for deviation from the route to find additional ridesharing opportunities. Rather than blindly staying on a single route and ignoring nearby pixels with customers waiting, a more sophisticated ridesharing policy would determine if nearby pixels had ridesharing opportunities that could be satisfied.

The ridesharing policy in effect, too, could be changed based on how the transportation demand evolves. Different ridesharing policies could be applied at peak and off-peak hours, for example. Using such a policy could create even more efficiency as the policy would
be tailored to the demand at specific times. More broadly, ridesharing policies could be applied selectively to certain highly active areas, while reducing the need for ridesharing in less active areas.

7.2.3 Ridesharing in Other Cities

For the purposes of this thesis, New York was the focus for ridesharing policies. As mentioned in the limitations of this thesis, the conclusions of this thesis may not necessarily be true in other cities. Due to New York’s proximity and centrality in the East Coast, as well as the fact that the New York’s TLC has the most comprehensive and freely available data concerning taxi trips, New York was chosen as the focus of this thesis. However, further research on ridesharing could examine the effects of ridesharing in other metropolitan areas. Determining whether the conclusions here also apply to other cities would be a veritable next step in expanding the reach of ridesharing.

The very same methods used within this thesis could simply be applied to datasets provided from other cities or countries. By doing so, the effects of the identical policy enumerated upon in this thesis can be observed in other areas to determine if ridesharing is an idea that could be utilized elsewhere, or if New York is unique in such a way that ridesharing would only be applicable there.

7.2.4 Additional Demand Analyses

The emergence of private taxi companies such as Uber has clearly made a dent in the number of trips taken in yellow taxis in New York. The dynamics of taxi demand - and how demand for taxis has changed with Uber’s lower fares and convenience - are still not fully understood. While it stands to reason that demand for taxis has not changed significantly overall despite Uber’s rise, the fact that these new taxi companies provide better or more comprehensive services could point to an evolution in the demand for premium taxi services.
7.3 Conclusions

With the arrival of disruptive companies such as Uber, the landscape of transportation - especially taxi transportation - will continue to change in the foreseeable future. With climate issues looming ever present, too, the issue of reducing the number of vehicles on the road as well as the mileage of these vehicles will become a primary focus for transportation policymakers in the coming years. Identifying solutions to limit the deleterious effects of vehicles will continue to be highly important for transportation policymakers.

Ridesharing would usher in a slew of benefits, both microscopically and macroscopically. On the small scale, ridesharing could save passengers money in fares, while simultaneously increasing the number of passengers taxi drivers could accommodate in a single trip. On a broader scale, the savings in terms of vehicles and vehicle miles would be a boon for cities and countries which are attempting to reduce their environmental impact. The effects of ridesharing are even more pronounced due to the fact that instituting ridesharing policies does not require huge investments in infrastructure that would be necessary for large transportation projects. Additional innovations in transportation - such as self-driving cars - could synergize with the benefits of ridesharing to create a highly streamlined transportation system not only in New York but in cities around the world.

In particular, ridesharing with hitchhiking as described in this thesis is an innovative policy that could further reduce the inefficiencies within the existing taxi system. The results obtained from the application of this policy to the data obtained from the TLC shows that ridesharing opportunities within New York are legion, and the adoption of a ridesharing policy could improve the TLC’s prospects in the face of an emergent Uber. As the average vehicle occupancy of these taxis increases, so do the savings for the taxi companies increase as well.

Ridesharing with hitchhiking, as seen in this thesis, could bring about increased efficiency by significantly reducing the amount of vehicles on the road. The current taxi system still has flaws that ridesharing could help to reduce. As private companies such as Uber continue to innovate with solutions such as UberPool, publicly run taxi companies such as the TLC must consider the benefits of projects such as ridesharing to hope to be
competitive in the rapidly changing landscape of taxi demand. With this in mind, the various players in the world of taxis can hope to adapt to various innovations in the continual search for efficiency.
Appendix A

Additional Figures

Figure A.1: The Rise of Uber in 2015
Figure A.2: The Pixel Grid Overlaid Over the New York Area
Appendix B

Data

B.1 Uber Data

The data regarding Uber pickups from April to September 2015 were obtained in response to a Freedom of Information Law Request from FiveThirtyEight.com. The news and analysis website posted the source files on Github at the following link:

https://github.com/fivethirtyeight/uber-tlc-foil-response

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Lat</th>
<th>Long</th>
<th>Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/1/2014 0:11</td>
<td>40.769</td>
<td>-73.9549</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:17</td>
<td>40.7267</td>
<td>-74.0345</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:21</td>
<td>40.7316</td>
<td>-73.9873</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:28</td>
<td>40.7588</td>
<td>-73.9776</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:33</td>
<td>40.7594</td>
<td>-73.9722</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:33</td>
<td>40.7383</td>
<td>-74.0403</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:39</td>
<td>40.7223</td>
<td>-73.9887</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:45</td>
<td>40.762</td>
<td>-73.979</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 0:55</td>
<td>40.7524</td>
<td>-73.996</td>
<td>B02512</td>
</tr>
<tr>
<td>4/1/2014 1:01</td>
<td>40.7575</td>
<td>-73.9846</td>
<td>B02512</td>
</tr>
</tbody>
</table>

Figure B.1: Sample of 2015 Uber Data
## B.2 TLC Data

The TLC data for yellow and boro taxis in 2013, 2014 and selected months of 2015 are freely available at:


### Table B.1: Sample of 2014 TLC Data (Continued in Table B.2)

<table>
<thead>
<tr>
<th>store_and_fwd_flag</th>
<th>pickup_datetime</th>
<th>dropoff_datetime</th>
<th>passenger</th>
<th>trip_distance</th>
<th>pickup_longitude</th>
<th>pickup_latitude</th>
<th>rate_code</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>-73.960449</td>
<td>40.763995</td>
<td>0.5</td>
<td>0.3</td>
<td>-73.955371</td>
<td>40.736828</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.986626</td>
<td>40.765217</td>
<td>1.5</td>
<td>2.5</td>
<td>-73.982392</td>
<td>40.773382</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.979863</td>
<td>40.77705</td>
<td>0.5</td>
<td>0.3</td>
<td>-73.985857</td>
<td>40.794062</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.984367</td>
<td>40.70529</td>
<td>0.5</td>
<td>2.5</td>
<td>-73.960449</td>
<td>40.763995</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.983974</td>
<td>40.765217</td>
<td>0.5</td>
<td>2.5</td>
<td>-73.982392</td>
<td>40.773382</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.95909</td>
<td>40.773639</td>
<td>0.5</td>
<td>2.5</td>
<td>-73.982392</td>
<td>40.773382</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.94751</td>
<td>40.785975</td>
<td>0.5</td>
<td>2.5</td>
<td>-73.982392</td>
<td>40.773382</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.979023</td>
<td>40.740578</td>
<td>0.5</td>
<td>2.5</td>
<td>-73.982392</td>
<td>40.773382</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.960879</td>
<td>40.777184</td>
<td>0.5</td>
<td>2.5</td>
<td>-73.982392</td>
<td>40.773382</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>-73.983974</td>
<td>40.74205</td>
<td>0.5</td>
<td>2.5</td>
<td>-73.982392</td>
<td>40.773382</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table B.2: Sample of Modified TLC Data (Continued in Table B.3)

<table>
<thead>
<tr>
<th>pickup_date</th>
<th>pickup_time</th>
<th>dropoff_date</th>
<th>dropoff_time</th>
<th>passengers</th>
<th>distance</th>
<th>pickup_lat</th>
<th>pickup_lng</th>
<th>pickup_xPixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/9/2014</td>
<td>20:45:25</td>
<td>1/9/2014</td>
<td>20:52:31</td>
<td>1</td>
<td>0.7</td>
<td>-73.99477</td>
<td>40.736828</td>
<td>144</td>
</tr>
<tr>
<td>1/9/2014</td>
<td>20:44:47</td>
<td>1/9/2014</td>
<td>20:59:46</td>
<td>2</td>
<td>2.3</td>
<td>-73.985857</td>
<td>40.739406</td>
<td>147</td>
</tr>
<tr>
<td>1/9/2014</td>
<td>20:44:57</td>
<td>1/9/2014</td>
<td>20:51:40</td>
<td>1</td>
<td>1.7</td>
<td>-73.960213</td>
<td>40.770464</td>
<td>163</td>
</tr>
<tr>
<td>1/9/2014</td>
<td>20:47:09</td>
<td>1/9/2014</td>
<td>20:53:32</td>
<td>1</td>
<td>0.9</td>
<td>-73.995371</td>
<td>40.717248</td>
<td>144</td>
</tr>
<tr>
<td>1/9/2014</td>
<td>20:45:07</td>
<td>1/9/2014</td>
<td>20:51:01</td>
<td>1</td>
<td>0.9</td>
<td>-73.983811</td>
<td>40.749455</td>
<td>150</td>
</tr>
<tr>
<td>1/9/2014</td>
<td>20:44:04</td>
<td>1/9/2014</td>
<td>21:05:45</td>
<td>1</td>
<td>3.6</td>
<td>-73.984338</td>
<td>40.726317</td>
<td>150</td>
</tr>
<tr>
<td>1/9/2014</td>
<td>20:50:33</td>
<td>1/9/2014</td>
<td>20:58:10</td>
<td>1</td>
<td>2.3</td>
<td>-73.955193</td>
<td>40.765468</td>
<td>166</td>
</tr>
<tr>
<td>pickup_yPixel</td>
<td>pickup_xPixel</td>
<td>dropoff_lat</td>
<td>dropoff_long</td>
<td>dropoff_xPixel</td>
<td>dropoff_yPixel</td>
<td>dropoff_zPixel</td>
<td>flag</td>
<td>trip_number</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>-------------</td>
<td>--------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>------</td>
<td>-------------</td>
</tr>
<tr>
<td>169</td>
<td>1245</td>
<td>-73.982227</td>
<td>40.73179</td>
<td>151</td>
<td>165</td>
<td>1252</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>194</td>
<td>1246</td>
<td>-73.960449</td>
<td>40.763995</td>
<td>163</td>
<td>188</td>
<td>1255</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>171</td>
<td>1244</td>
<td>-73.986626</td>
<td>40.765217</td>
<td>148</td>
<td>188</td>
<td>1259</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>192</td>
<td>1244</td>
<td>-73.979863</td>
<td>40.77705</td>
<td>152</td>
<td>197</td>
<td>1251</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>155</td>
<td>1247</td>
<td>-73.984367</td>
<td>40.720524</td>
<td>150</td>
<td>158</td>
<td>1253</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>178</td>
<td>1245</td>
<td>-73.989747</td>
<td>40.756575</td>
<td>147</td>
<td>182</td>
<td>1251</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>162</td>
<td>1244</td>
<td>-73.962869</td>
<td>40.758443</td>
<td>161</td>
<td>184</td>
<td>1265</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>175</td>
<td>1243</td>
<td>-73.95909</td>
<td>40.773639</td>
<td>163</td>
<td>194</td>
<td>1252</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>184</td>
<td>1243</td>
<td>-73.94251</td>
<td>40.785975</td>
<td>172</td>
<td>203</td>
<td>1254</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>189</td>
<td>1250</td>
<td>-73.979023</td>
<td>40.740578</td>
<td>153</td>
<td>171</td>
<td>1258</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>194</td>
<td>1251</td>
<td>-73.980879</td>
<td>40.777384</td>
<td>152</td>
<td>197</td>
<td>1275</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>177</td>
<td>1248</td>
<td>-73.988359</td>
<td>40.714205</td>
<td>148</td>
<td>153</td>
<td>1261</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure B.5: Sample of Modified TLC Data (Continued from Figure B.4)
Appendix C

Code

Various facets of this thesis required coding efforts in various languages, from Python to R to Matlab. A subset of the code used in the writing of this thesis is shown in this appendix.

C.1 Removal of Columns in TLC Data

```python
import csv

month = 0
for i in range(1, 13):
    if i < 10:
        month = "0" + str(i)
    else:
        month = str(i)
print(month)

initloc = 'E:\Thesis\tlcdata\tlctest.csv'
midloc = 'E:\Thesis\tlcdata\tlctest2.csv'
endloc = 'E:\Thesis\tlcdata\tlctest3.csv'

with open(initloc, "r") as source:
    rdr = csv.reader(source)
    next(rdr, None)
    with open(midloc, "w", newline=\"n\") as result:
        wtr = csv.writer(result)
        for r in rdr:
            if r:
                wtr.writerow((r[1], r[2], r[3], r[4], r[5], r[6], r[9], r[10]))

with open(midloc, "r") as inp, open(endloc, "w", newline=\"n\") as out:
    writer = csv.writer(out)
```

72
for row in csv.reader(inp):
    if row:
        if row[3] != "0":
            writer.writerow(row)

C.2 Pixelization of TLC Data

import csv
import math

initloc = 'E:\\Thesis\\tlcdata\\bla\\tlc2014-3.csv'
endloc = 'E:\\Thesis\\tlcdata\\bla\\tlc2014-4.csv'

with open(initloc, 'r') as inp, open(endloc, 'wb') as out:
    writer = csv.writer(out)
    for row in csv.reader(inp):
        if row:
            nl = []
            nl.append(row[0])
            nl.append(row[1])
            nl.append(row[2])
            nl.append(row[3])
            nl.append(row[4])
            nl.append(row[5])
            nl.append(row[6])
            nl.append(row[7])
            origin_xPixel = math.floor(((float(row[6]) + 74.260186) * 25 * 21.7814))
            origin_yPixel = math.floor(((float(row[7]) - 40.491825)) * 25 * 27.64)
            l = row[1].split(':
            origin_zPixel = math.floor((int(l[0]) * 3600 + int(l[1])
                                      * 60 + int(l[2]))) / 60)
            nl.append(origin_xPixel)
            nl.append(origin_yPixel)
            nl.append(origin_zPixel)
            nl.append(row[8])
            nl.append(row[9])
            dest_xPixel = math.floor(((float(row[8]) + 74.260186)) * 25 * 21.7814)
            dest_yPixel = math.floor(((float(row[9]) - 40.491825)) * 25 * 27.64)
            l = row[3].split(':
            dest_zPixel = math.floor((int(l[0]) * 3600 + int(l[1])
                                      * 60 + int(l[2]))) / 60)
            nl.append(dest_xPixel)
            nl.append(dest_yPixel)
            nl.append(dest_zPixel)
C.3 The Ridesharing Implementation

```python
import csv
import math
import numpy as np

initloc = "E:\Thesis\tlcdata\daysplits\tlc0101.csv"
endloc = "E:\Thesis\tlcdata\bla\ridesharing0101.csv"

# Bresenham's line algorithm, courtesy of Vikas Dhiman at ActiveState Code Recipes
def _bresenhamline_nslope(slope):
    scale = np.amax(np.abs(slope), axis=1).reshape(-1, 1)
zeroslope = (scale == 0).all(1)
scale[zeroslope] = np.ones(1)
normalizedslope = np.array(slope, dtype=np.double) / scale
    normalizedslope[zeroslope] = np.zeros(slope[0].shape)
return normalizedslope

# Courtesy of Vikas Dhiman at ActiveState Code Recipes
def _bresenhamlines(start, end, max_iter):
    if max_iter == -1:
        max_iter = np.amax(np.amax(np.abs(end - start), axis=1))
npts, dim = start.shape
nslope = _bresenhamline_nslope(end - start)

    # steps to iterate on
stepseq = np.arange(1, max_iter + 1)
stepmat = np.tile(stepseq, (dim, 1)).T

    # some hacks for broadcasting properly
bline = start[:, np.newaxis, :] + nslope[:, np.newaxis, :] * stepmat

    # Approximate to nearest int
```

74
return np.array(np.rint(bline), dtype=start.dtype)

# Courtesy of Vikas Dhiman at ActiveState Code Recipes
def bresenhamline(start, end, max_iter=5):
    # Return the points as a single array
    return _bresenhamlines(start, end, max_iter).reshape(-1, start.shape[-1])

# Function that computes the unit vector of an input vector
# Courtesy of David Wolever at Stack Overflow
def unit_vector(vector):
    return vector / np.linalg.norm(vector)

# Function that computes the angle between two unit vectors
# Courtesy of David Wolever at Stack Overflow
def angle(v1, v2):
    v1_u = unit_vector(v1)
    v2_u = unit_vector(v2)
    return np.arccos(np.clip(np.dot(v1_u, v2_u), -1.0, 1.0))

tripNumber = 1
with open(initloc, 'r') as inp, open(endloc, 'wb') as out:
    writer = csv.writer(out)
    for row in csv.reader(inp):
        if row:
            if row[16] == 1:
                next(row, None)
            else:
                pixelsTraversed = bresenhamline(np.array([[int(float(row[8]))],
                                                          int(float(row[9]))],
                                                          int(float(row[10])))),
                np.array([[int(float(row[13]))],
                          int(float(row[14]))],
                          int(float(row[15])))),
                max_iter = -1
                for i in range(1, len(pixelsTraversed)):
                    with open(initloc, 'r') as newinp:
                        for newrow in csv.reader(newinp):
                            origvec = [int(float(row[13]))] -
                            pixelsTraversed[i][0],
                            int(float(row[14])) -
pixelsTraversed[i][1],
int(float(row[15])) –
pixelsTraversed[i][2]
newvec = [int(float(newrow[13])) –
pixelsTraversed[i][0],
int(float(newrow[14])) –
pixelsTraversed[i][1],
int(float(newrow[15])) –
pixelsTraversed[i][2]]
if (int(float(newrow[8])) == pixelsTraversed[i][0] and
int(float(newrow[9])) == pixelsTraversed[i][1] and
int(float(newrow[10])) == pixelsTraversed[i][2] and
−0.17453 <= angle(origvec, newvec) <= 0.17453):
row[16] = 1
row[17] = tripNumber
tripNumber += 1
Bibliography


