Increasing Saudi Females’ Accessibility to Employment via Car-pooling in Riyadh:
Measure the Realistic Commute Cost by Network Computing Methods and
Investigate the Share-ability based on Actual Taxi Trip Data

By
Waishan QIU

M.ARCH. in Urban Design, with Distinction
University College London
London, UK (2014)

B.E. in Urban Planning
Tongji University
Shanghai, China (2013)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Master in City Planning
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

© 2017 Waishan QIU. All Rights Reserved

The author hereby grants to MIT the permission to reproduce and to distribute publicly
paper and electronic copies of the thesis document in whole or in part in any medium now
known or hereafter created.

Signature redacted

Author

Signature redacted

Department of Urban Studies and Planning
(May 24, 2017)

Certified by

Signature redacted

Associate Professor Sarah Williams
Department of Urban Studies and Planning
Thesis Supervisor

Accepted by

Signature redacted

Associate Professor P. Christopher Zegras
Chair, MCP Committee
Department of Urban Studies and Planning
Increasing Saudi Females’ Accessibility to Employment via Car-pooling in Riyadh:
Measure the Realistic Commute Cost by Network Computing Methods and
Investigate the Share-ability based on Actual Taxi Trip Data

by
Waishan QIU

Submitted to the Department of Urban Studies and Planning on May 24, 2017
In Partial Fulfillment of the Requirements for the Degree of
Master in City Planning

Abstract

Due to the driving ban as well as to social restrictions on their movement with male drivers, Saudi women have to rely on either male family members or the employer’s shuttle bus; otherwise they would hire a driver or take the taxi. These few options pose high commute cost on Saudi females, hence their access to economic opportunities is restrained, especially among lower car ownership segments. Such restrictions have negative influence on Saudi females’ employment. The employment rate of Saudi women is only 22%. However, no previous research has quantitatively investigated the commute cost as a financial burden and barrier to job participation for Saudi females. Taking the capital city Riyadh as a case study, this study will (1) develop the method to measure the realistic commute cost (in terms of time and money) by different transportation options for Saudi female residents in different job sectors; (2) examine the spatial mismatch between Saudi females’ concentration and their job markets using the notion of accessibility; (3) demonstrate ridesharing’s capacity of providing greater access for Saudi women based on spatial analysis of the current commute demands and behaviors; (4) and also look at the feasibility of developing ridesharing programs based on network analysis of current taxi trips. This study sheds light on implications for policy makers and ridesharing service companies to reduce Saudi females’ commute cost so as to increase their access to economic opportunities.

Thesis Advisor: Sarah Williams, Associate Professor, DUSP, MIT
Reader: Jawaher Al-Sudairy, Research Fellow, EPoD, Harvard Kennedy School
Key Words

Saudi female, Riyadh, commute cost, employment, accessibility, ridesharing, urban computing, network analysis
Acknowledgements

The completion of this master degree thesis cannot be accomplished without the intellectual inputs, collective insights and generous supports from a group of amazing people I get to know in Cambridge and in my life.

First of all, I would like to thank my thesis advisor Prof. Sarah Williams, because she was the first person brought me to the topic of big data and society, and with the context of Riyadh and Saudi Arabia. Riyadh was an unknown territory and the network computing methods was an unfamiliar approach to me before I took the Big Data and Society class. It was in her class that I started to develop the network method to calculate the travel cost in Riyadh. After I decided to develop it as my thesis, I discussed with her the potential of investigating the relationship between commute cost and economy opportunities for Saudi female. She encouraged me with various instructions particularly in the research question and structure. Many excellent ideas just came out every time after the weekly thesis meeting with her. She guided me to think deeper, suggested me with good ideas, and pushed me to make it the best I could.

I would also like to thank Jawaher Al-Sudairy. Jawaher is a research fellow from Evidence for Policy Design (EPoD) at Harvard Kennedy School. She is not only the PI for the Riyadh research I participate in, but also serves as my reader. She first came to me with the initiative to examine the financial burden as a barrier to jobs for Saudi female. She provided data sources and critical insights on the target population and job sectors in Riyadh. She enables me to complete this study to a depth and within a time frame that a master thesis allows. She also involves me in the progress of negotiating with ride service companies. Such an in-depth cooperative research with her and EPoD greatly broaden my knowledge of how to initiate a feasible and critical research in the future.

I would also like to thank Prof. Joseph Ferreira and Prof. Brent D. Ryan who gave me many critical suggestions in research question and literature review. Many thanks to the TAs of Thesis Preparation, Babak Manouchehrifar (PhD at MIT). He really reshaped my research question during the preliminary stages in the class. And of course I also wish to thank Carlos Sandoval Olascoaga (PhD at MIT) and Michael Foster (GIS specialist at MIT) who provided me with great helps with the computation and visualization part in and outside the class of Big Data Visualization. Meanwhile, I would like to thank the Civic Data Design Lab, Cortni Kerr and the researchers from research partnerships in Riyadh. Since almost every week I attend their meeting, I learn not only enormous ideas to extend the boundary of accessibility, but also the skills of delivering convincing presentation. Additionally, I thank all the other attendees of weekly thesis meeting: Emily Long, Scott Cedric Margeson, Dennis Cameron Harvey, and Erin Kenney.

Finally, I would like to thank my family, who always support me, stand by me and keep giving me the courage to face new challenges.
# Table of Contents

Abstract ........................................................................................................................................... 3

Key Words....................................................................................................................................... 4

Acknowledgements.......................................................................................................................... 5

Chapter 1. Introduction ..................................................................................................................... 8

1.1 Research Question and Objectives ............................................................................................. 10

   1.1.1 Research Question and Hypothesis ....................................................................................... 10

   1.1.2 Objectives .............................................................................................................................. 12

1.2 Thesis Outline ............................................................................................................................. 14

1.3 Work Distribution ....................................................................................................................... 17

Chapter 2. Literature Review ............................................................................................................ 18

2.1 Accessibility and its Measurement .............................................................................................. 19

   2.1.1 Accessibility and Economic Opportunity ............................................................................... 20

   2.1.2 Measurement of Accessibility .............................................................................................. 23

   2.1.3 Positive and Normative Terms ............................................................................................... 26

2.2 Ridesharing and the Measurement of Share-ability ................................................................... 30

   2.2.1 Ridesharing around the World ............................................................................................... 30

   2.2.2 Measure the Share-ability ...................................................................................................... 31

2.3 Limitation and Restrictions on Saudi Females’ Access to Jobs .................................................. 33

   2.3.1 The Driving Ban and Job Limits on Saudi Women ................................................................. 34

   2.3.2 No Public Transportation Exists in Riyadh ........................................................................... 35

   2.3.3 High Commute Cost as a Barrier to Employment ................................................................. 38

2.4 Ridesharing in Saudi Arabia ....................................................................................................... 40

   2.4.1 Long Existence of Informal Ridesharing ............................................................................. 41

   2.4.2 Emerging Formal Ridesharing Service in Saudi Arabia ....................................................... 43

Chapter 3. Data and Methodology .................................................................................................... 45

3.1 Data Selection .............................................................................................................................. 47

   3.1.1 Saudi Females’ Demography ................................................................................................. 48

   3.1.2 Saudi Females’ Employment Opportunities .......................................................................... 48

   3.1.3 Riyadh’s Transportation Network ......................................................................................... 48
Chapter 1. Introduction

In Saudi Arabia, the driving ban, the lack of access to autonomous transportation, the constant need of guardian’s approval, and strict job restrictions that involve gender segregation are the top obstacles for Saudi women to access employment (Beiter, 2016). These deterrents also pose high financial burdens that restrict females’ movement. Take Riyadh, the capital city of Saudi as an example: due to its vehicle-oriented urban form and its year-round high temperatures, most of commuting trips are far beyond the reach of pedestrian. However, the completion of its metro system will not be realized until 2019. To date, only 2% of daily trips rely on public transportation (Zavis, 2015). While most men drive a car to search for jobs and commute to work, women have to rely on their male family members to get to work; otherwise they hire a driver or take the taxi. As Saudi females are restricted to costly commuting options- either by private car or taxi, many women report spending half their salary on their daily commute (Gorney, 2016). Such limitations on commute have direct and negative influence on their economic participation outcomes. The unemployment is exceptionally high among Saudi females. Until recently, Saudi women only comprise about 22% of the employment market (Beiter, 2016).

There is an urgent need to look for ways such as ridesharing to reduce commute cost for Saudi women. Some companies are responding by providing a transportation stipend, or actual transportation services to their female employees (Lori P. Boghardt, 2013). Such benefits are becoming common place following enforcement of the Nitaqat program,
which places quotas on the hiring of Saudi females in private companies (Helen, 2013 and Gorney, 2016). The Saudi Ministry of Labor is also exploring ways for reducing transportation costs to assist women employee on their daily commute (Lori P. Boghardt, 2013). Meanwhile, with the plan of removing gas subsidy, Saudi government is considering options for maintaining affordable commutes through direct subsidies to citizens to offset the increase in gas prices, especially when the metro system is still under construction (Gulf Times, 2010). In order to provide greater access to jobs for Saudi females as well as to increase sustainability, it is of crucial significance to have a realistic measurement of Saudi females’ actual commute costs by different options.

While the research linking commute cost to economy opportunities is varied in America, East Asia and Europe (O’Regan and Quigley, 1997; Blumenberg and Ong, 1998; Allard and Danzoger, 2000; Cervero, Sandova and Landis, 2002), no existing transportation study has focused on the spatial attributes of commute cost and female accessibilities in Saudi Arabia and Riyadh. Financial burden as a barrier to commute and its impact on commute behaviors, preferences and the corresponding employment outcomes have not been examined for Saudi females.

In this context, this paper aims to build knowledge on the role of commute cost as a financial burden impeding Saudi females’ job participation in Riyadh. This research will develop methods to understand the realistic commute cost (measure by time and local currency SAR) of using these systems compared with other transit options for different
female job sectors (manufacturing, retail, healthcare, and education). The results show that while ridesharing could benefit the most population for women’s employment in retail and education, it will also increase female access to manufacturing sector dramatically from as little as 1.70% to a relative high level with 39.76% of the Saudi female population.

The research will also look at the feasibility of developing ridesharing programs based on current analysis of taxi trips. It will quantify the benefits of paired trips on traffic congestion. The integrated method presented in this paper provides an addition to the professional and academic work in the area of urban planning and transportation policy.

1.1 Research Question and Objectives

By summarizing the phenomenon about the limitation and restriction on Saudi females’ movement and describing the facts about their low employment participation, this section explains the motivation, the research question and the objectives of the paper.

1.1.1 Research Question and Hypothesis

On one hand, due to the driving ban and the poor existence of public transportation, Saudi women are subject to commute options of high financial burden (Gulf Times, 2010). On the other hand, the employment rate of Saudi Female is low while the nation is ambitious to increase their economy participation (Gorney, 2016). Previous researches on the relationship between accessibility and economic opportunities worldwide establishes that
limitations on the access to commute affect labor market outcomes significantly. (John Meyer, 1968; O’Regan and Quigley, 1997; Paez, Scott and Morency, 2012).

As a result, with the understanding of Saudi women’s high commuting cost burden, and the brief review of the relationship between accessibility and job opportunities worldwide, the research looks at what is the effect of commuting costs on Saudi females’ economic participation outcomes in Riyadh. Particularly it is questioning that, due to the driving ban and specific religious culture restriction, whether Saudi females’ job participation is limited by the high commute costs?

Meanwhile, there exists the custom of using informal ridesharing services of hiring and sharing a private car driver. Emerging ridesharing service companies such as UBER fits in the existing commute behavior pretty well and are claiming to largely reduce female’s commute cost (Nagy, 2015). As a result, the research also questions a secondary research question that, whether or not ridesharing has the potential to provide greater access to employment for women in Riyadh.

The overarching research hypothesis is that, first of all, due to the prohibition of driving, there is a significant correlation between women’s accessibility and labor force participation in Riyadh, Saudi Arabia. Secondly, the out-of-human-scale urban form and the lack of public transportation make ridesharing the most significant and efficient mean helping women to increase their accessibility. Thirdly, as a result of reduced commute cost,
ridesharing service can increase women’s economic opportunities. To be more specific, the study supposes these consecutive hypotheses:

1. Commute cost is the major barrier for low-income women in Riyadh to potential work opportunities.
2. Ridesharing service can help women in Riyadh to reduce the commute cost and risks, thus becoming the most significant and efficient commute means.
3. By reducing the cost, ridesharing indeed increases women in Riyadh with better accessibility and job opportunities.

1.1.2 Objectives

With the research questions above, the study will reflect on the comprehensive issue of “can ridesharing in Riyadh improve women’s accessibility to employment”. If yes, to what extent do they help? If not, how can ridesharing entrepreneurships work with policy makers and come up with better ride-sharing experimental products? This research also investigates a more methodological research interest, which is how urban computing and network analysis method can help urban planners, policy makers and transportation service enterprises to develop affordable services that improve residents’ accessibility. The methodology deployed in this research is also questioning how information technology as a timely approach help to identify and solve urban problems.

To answer the first research question, it is crucial to have in depth understanding on the realistic terms of commute cost (measured by time and money). As a result, this research
will develop methods to understand the actual commute cost (measure by SAR) of using these systems compared with other transit options for different female job sectors (manufacturing, retail, healthcare, and education) in Riyadh. It is necessary to explore the following:

1. What is the spatial distribution of Saudi female residents?
2. Where are the employment opportunities located for Saudi females?
3. Is Google Direction API reliable to provide realistic commute cost of time and distance?
4. What are the different commute cost by current transportation options for Saudi females?
5. How is the cost of ridesharing compared to them?
6. Is there a spatial mismatch problem between the concentration of Saudi Females and their employment opportunities?
7. Is there a spatial correlation between the accessibility measured by realistic commute cost and the employment rate for Saudi females?
8. What are the implications of such findings for policy and practice, especially for public transportation and urban design department?

Moreover, as the research will also take a brief look at the feasibility of ridesharing service and its benefits based on current analysis of actual trips from the taxi trip data, its investigation of share-ability need to cover following aspects:

1. What is the breakdown of taxi usage by gender in Riyadh?
2. What is the trips pattern of female passengers?

3. Is ridesharing feasible based on current taxi trips?

4. What is the methodology to determine whether or not ridesharing is efficient and to quantify the benefits of it?

5. What is the benefit of ridesharing in terms of city scale sustainability, say how much traffic congestion it can save?

1.2 Thesis Outline

Before the exploration of measuring accessibility to employments and the share-ability based on actual taxi trips for Saudi female in Riyadh, this study will first deliver a comprehensive literature review on the relevant topics in Chapter 2. The literature review covers:

(1) 2.1 Accessibility, its measurement and its relationship with economy opportunities

(2) 2.2 The notion and methods of measuring Share-ability and its benefits

(3) 2.3 Limitation and restriction on the Saudi females’ access to employments

(4) 2.4 Ridesharing and its potentials in Saudi Arabia

Accessibility is one of the most important and continuing concept in transportation research to quantify the joint result of a transportation network and the geographical distribution of activities (Paez, Scott and Morency, 2012). The study will compare different...
measurements of accessibility and then explain the reason why the cumulative approach is selected as well as why the money cost is used as indicator.

Chapter 3 describe the selected dataset and the methods of urban computing and network analysis to quantify the accessibility and measure the share-ability step by step. Starting from a short explanation about dividing the region of Riyadh City to a matrix of geolocation zones by 2KM x 2KM fishnet grids, this study set up the basic geospatial framework for urban computing and spatial investigation. Based on the grids, this study promotes the basic concepts and steps of computing the actual commute time and distance through Google direction API. Following that, it introduces different fare estimation equation offered by different transportation service companies (such as street hailing taxi, a local online hailing taxi company Careem, Uber, and a local car-pooling service company Kasper). The realistic estimation of commute costs become indicators for the actual accessibility measurement. Linear regression model of accessibility and employment outcome in each grid is built. It is worth to mention that, without Civic Data Design Lab and Evidence for Policy Design, it is impossible to acquire the demography dataset and cost estimation equations. The successful authorization of data source is largely due to their good relationship with local municipality and transportation companies.

Chapter 4 starts with describing and illustrating the research findings. It is the interpretation of the result and outcomes of the methodologies developed in Chapter 3.
This section demonstrates three major findings generated from urban network computing and the corresponding spatial investigation:

1. 4.1 The Spatial Mismatch between Concentration of Saudi Female Residents and Their Potential Employment in Riyadh, Using Cumulative Accessibility Model as the Indicator

2. 4.2 The Potential Market of Formal Ridesharing and Its Influence on Metro System

3. 4.3 The Share-ability and Its Benefits in Riyadh

4. 4.4 Limitations and Future Studies

Chapter 5 first concludes the research results in section 5.1. It highlights the findings from previous chapters. It establishes the current spatial mismatch between Saudi female residents and their potential employments in Riyadh. It then advocates the potential market demand of ridesharing and the corresponding benefits of developing formal ridesharing service. In section 5.2, it demonstrates the implication of the computational methodology developed in this study. It asserts the implication of the specific components which measures the realistic terms of commute cost and its corresponding accessibility for future urban and transportation studies.
1.3 Work Distribution

In respect to individual researcher's effort and contribution, it is necessary to clearly state the work distribution between the author Waishan QIU and his cooperator Jawaher Al-Sudairy to explicit their contribution to this research.

Firstly, Waishan is the only author of this thesis paper. For the written part, this thesis is completely written by Waishan. There will be a clear statement at the beginning of any content referred from Jawaher or co-authored by Waishan.

Secondly, Waishan develops the methodology of computing commute cost as early as in the class of Big Data Visualization taught by Prof. Sarah Williams. Prof. Sarah is the advisor of this thesis. The computational method of measuring accessibility and share-ability is also developed by Waishan QIU. The participation of Jawaher comes about after the Big Data class’s final presentation around May 2016, when Waishan has calculated the most part of the commute cost.

Thirdly, Jawaher contributes greatly to the understanding of context and Saudi females’ conditions in Riyadh. She initiated the direction of examining the relationship between commute cost and females’ employment. She negotiated a lot with research partners in Riyadh to acquire CENSUS data and the fare estimation equations. She also initiated to use the proportion of monthly incomes as the price buffers for accessibility measures.
Chapter 2. Literature Review

This chapter reviews the development of the relevant notions and methods that this research takes as basic, namely (1) the notion and measurement of accessibility, (2) the concept and quantitative method of measuring share-ability, (3) the background study of Saudi female’s human right related to mobility and accessibility, and (4) current research on the commute behavior of Saudi females. All component are essential to enable the investigation on the research questions of accessibility and economic opportunities. The literature review also help to develop the innovative urban computing methods established in this paper.

First of all, a brief review on the previous researches exploring the relationship between accessibility and economic opportunities is necessary. It enables the spatial investigation of the specific relationship of commute cost and Saudi females’ employment. Through the literature review, the research narrows down its direction and get inspiration from existing quantitative methods and accessibility models.

Secondly, investigates car-sharing as a mode for daily commuting trips require an overview on the development, trend, benefit and technology of ridesharing. It enables the debate on the positive impacts on the city’s congestion as well as on the commute cost for females. Moreover, as part of this research tend to demonstrate its contribution to the specific technology of urban computing, network analysis and API method, a robust examination
on current urban information technology is essential. Riyadh, as the capital and most populous city in Saudi Arabia, the adequate dataset of taxi trips offers a positive opportunity to test the feasibility of car sharing schemes with its specific urban form (road network and density), trip patterns and traffic condition.

Last but not the least, to get a comprehensive understanding of the social and religious culture in Saudi Arabia, and to look into the impacts of women being banned from driving, previous insights from social political point of view and existing reports on the commute pattern and behavior on Saudi females are prerequisite.

Built on those precedent knowledge, this study will contribute to the specific topic of accessibility and economic opportunities in urban planning and transportation policy studies with Saudi Arabia’s context specifics. It will also shed light on innovative and realistic urban measurement and computing for urban information system methodology.

2.1 Accessibility and its Measurement

This section briefly reviews relevant researches on the relationship of accessibility and economic opportunities. This part of literature review provides preliminary concept, notions and fundamental methods that support the proposed quantitative measurement and spatial investigation. It justifies the core investigation on the relationship of commute cost as a financial burden and the access to employment for Saudi females in Riyadh.
2.1.1 Accessibility and Economic Opportunity

Accessibility is a continuing relevant concept in transportation research. Defined as the potential for reaching spatially distributed opportunities, accessibility is the joint result of a transportation network and the geographical distribution of activities (Paez, Scott and Morency, 2012). It is one of the main outputs as well as indicators of urban spatial development. Almost fifty years ago in 1968, it was a distinct trend in urban studies that people started to pay attention to the link between inadequate transportation and urban poverty (O’Regan and Quigley, 1997). The relationship between accessibility and its employment consequences for low income households has drew widespread attention ever since (John Meyer, 1968). And empirical studies in urban planning and transportation supplies accumulative evidence reinforcing those insights about the effects of urban space upon employment outcomes.

One of the most significant finding is by O’Regan and Quigley in 1997. Based on the evidence of 1970 & 1990 CENSUS data, they establishes that the limitations on the access provided to low income and minority workers affect labor market outcomes in a significant way. After that a variety of cross-sectional analyses based on aggregate census data and, more recently, upon micro data on individual workers has sought to quantify the importance of these linkages. As with most social science research, more sophisticated analyses of access and employment reveal more complexities and ambiguities in their effects. In 2003 Blumenberg and Shiki studied the travel behavior of welfare recipients on
public transit and measured their accessibility to employment outside large urban centers. The result showed that while there are jobs in the rural areas of the county, both jobs and welfare recipients are more dispersed (see Figure 1), making travel much more difficult for those without access to automobiles (see Table 1). Therefore, there is evidence that many rural, transit-dependent welfare recipients have limited access to job opportunities.

Figure 1
Density of Welfare Recipient (Left) & Low-Wage, Feminized Jobs (Right), Fresno County
Source: Blumenberg & Shiki, 2003
Table 1
Welfare Recipients’ Accessibility and Employment Outcomes, Fresno County
Source: Blumenberg & Shiki, 2003

<table>
<thead>
<tr>
<th>Industrial Sector</th>
<th>All Employment</th>
<th>Welfare Recipients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural</td>
<td>19%</td>
<td>15%</td>
</tr>
<tr>
<td>Construction/Mining</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>9%</td>
<td>10%</td>
</tr>
<tr>
<td>Transportation/Communications</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>Wholesale Trade</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>15%</td>
<td>26%</td>
</tr>
<tr>
<td>Finance, Insurance, Real Estate</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>Services</td>
<td>21%</td>
<td>38%</td>
</tr>
<tr>
<td>Public Administration</td>
<td>19%</td>
<td>1%</td>
</tr>
<tr>
<td>Total, all industries</td>
<td>305,822</td>
<td>15,608</td>
</tr>
</tbody>
</table>

Source: Fresno County welfare administrative data and California Employment Development Department, Business Establishment List for first two quarters of 1999 (1999)

*In the total employment data, workers are counted more than once if they were employed at multiple establishments during 1999. For welfare recipients, Fresno County welfare administrative data were matched to the employer in which they earned the highest earnings in 1999.

Table 2: Relative Proximity to Employment by Residential Location
Fresno County, 1999

<table>
<thead>
<tr>
<th>Job Access Quartiles</th>
<th>Working-Age Population</th>
<th>% Welfare Recipients</th>
<th>% Urban Welfare Recipients</th>
<th>% Rural Welfare Recipients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Rich</td>
<td>26%</td>
<td>35%</td>
<td>43%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>29%</td>
<td>33%</td>
<td>41%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>24%</td>
<td>22%</td>
<td>16%</td>
<td>47%</td>
</tr>
<tr>
<td></td>
<td>21%</td>
<td>10%</td>
<td>0%</td>
<td>53%</td>
</tr>
<tr>
<td>Total</td>
<td>390,051</td>
<td>24,974</td>
<td>20,079</td>
<td>4,895</td>
</tr>
</tbody>
</table>

However, notwithstanding these trends, and the importance of transportation, it is worth noting that other factors are more important in affecting the employment of low income and minority workers. Education, training, skills, and the overall health of the economy are all more important in affecting the labor market outcomes of disadvantaged workers than is transportation or access per se (Katherine & John, 1998).
2.1.2 Measurement of Accessibility

According to Handy and Niemeier (1997), and Paez, Scott and Morency (2012), previous typologies of accessibility measures are organized into three types: (1) cumulative opportunities measures, (2) gravity-based measures, and (3) utility-based measures. While they all incorporate both a transportation and an activity element, they differ from each other in the sophistication with travel behavior. Planners must address the specification, particularly the degree of disaggregation; calibration; and interpretation when developing a measurement (Handy, 1997).

(1) Cumulative Opportunities Measures

It is the simplest class of accessibility measures. It counts the number of opportunities that can be reached within a given travel time or distance (Geurs & Van Wee, 2004). Regardless of differences in travel time, all potential destinations within the cutoff time are weighted equally. This measurement demonstrates the number of potential destinations or choices rather than their distance. This measure is popular in urban planning and geographical studies (e.g. Wickstrom, 1971; Wachs and Kumagai, 1973; Gutiérrez and Urbano, 1996; Bruinsma and Rietveld, 1998).

(2) Gravity-based measures

It is more complex than cumulative opportunities approach. It derives from the denominator of the gravity model for trip distribution and have been widely used in urban and geographical studies since the late 1940s (Stewart, 1947; Hansen, 1959; Ingram, 1971;
and Vickerman, 1974). It weights opportunities, usually the quantity of an activity as measured by employment, by impedance, generally a function of travel time or travel cost (Handy, 1997). Accessibility is then measured as

\[ A_i = \sum_{j=1}^{n} D_j e^{-\beta c_{ij}} \quad (1) \]

- \( A_i \): a measure of accessibility in zone \( i \) to all opportunities \( D \) in zone \( j \)
- \( D_j \): activity in zone \( j \)
- \( C_{ij} \): the costs of travel between \( i \) and \( j \),
- \( \beta \): the cost sensitivity parameter.

From equation (1), it is easy to conclude that, the closer the opportunity, or the larger the opportunity, the more it contributes to accessibility.

The cost sensitivity function used has a significant influence on the results of the accessibility measure. Although the impedance function may take many forms, the negative exponential form has been used more often than others in recent studies and is also the most closely tied to travel behavior theory (Handy and Niemeier, 1997). For plausible results, the form of the function should be carefully chosen, and the parameters of the function should be estimated using recent empirical data of spatial travel behavior in the study area.

It is worth mentioning that, a cumulative opportunities measure is actually a specific form of the gravity-based measure, with the impedance function equal to one if the opportunity is within the travel time limit, and zero otherwise (Koenig, 1980).
**Random Utility Based Approach**

The third class of accessibility measures is based on random utility theory, in which the probability of an individual making a particular choice depends on the utility of that choice relative to the utility of all choices. If it is assumed that an individual assigns a utility to each destination (or mode and destination) choice in some specified choice set, \( C \), and then selects the alternative which maximizes his or her utility, accessibility can be defined as the denominator of the multinomial logit model, also known as the log sum (Ben-Akiva and Lerman, 1985; McFadden 1981). Accessibility for an individual \( n \) is then measured as:

\[
A_n = \ln \left[ \sum_{c \in C_n} \exp (V_{n(c)}) \right], \tag{2}
\]

\( A_n \): Accessibility for an individual \( n \);
\( V_{n(c)} \): observable spatial-temporal transportation components of indirect utility of choice \( c \) for person \( n \)
\( C_n \): the choice set for person \( n \)

The logsum serves as a summary measure indicating the desirability of the full choice set (Small, 1992). The specified utility function includes variables that represent the attributes of each choice, reflecting the attractiveness of the destination and the travel impedance that must be overcome to reach the destination, and the socioeconomic characteristics of the individual or household, reflecting individual tastes and preferences (Handy and Niemeier, 1997).
2.1.3 Positive and Normative Terms

Although the use of the term for accessibility has long been a relevance in urban planning and transportation studies, it has not always been translated into performance measures to more concretely direct planning efforts (Handy and Niemeier, 1997). Effective use in planning has been hampered in the past by limited understanding of the measures, definitional issues, and measurement problems. Previous reviews of indicators for measuring accessibility were concerned with various conceptual and computational issues such as translating concepts into operational measures and providing a conceptual framework to facilitate this translation (Handy and Niemeier, 1997); the need to improve land use and transport appraisal (Guers and van Wee, 2004); the comparison of place-based and personal measures of accessibility including their computational and data requirements, as well as the degree of agreement or correlation between measures in practice (Kwan 1998). Paez, Scott and Morency (2012) write a systemic review in addition to the works mentioned above (see Table 2). They also take an alternative perspective to focus on how the utility of accessibility measures are implemented, to illustrate the differences between normative and positive approaches. They present an empirical illustration using Montreal as a case study to shows that it is possible to derive great benefit from the synthesis of both perspectives by comparing accessibility as is to what it ought to be (Figure 2).
Table 2

Examples of accessibility measures: formulation and applications

Source: Paez, Scott and Morency (2012)

<table>
<thead>
<tr>
<th>Accessibility from the origin</th>
<th>Formulation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative opportunity</td>
<td>$W(x)$</td>
<td>1 if $W(x) &lt; 0$ (threshold value); 0 otherwise</td>
</tr>
<tr>
<td>Relative supply: $W(x) = \frac{1}{x}$</td>
<td></td>
<td>1 if $W(x) &lt; 0$ (threshold value); 0 otherwise</td>
</tr>
<tr>
<td>Demand of a service (e.g. number of beds in a hospital)</td>
<td></td>
<td>1 if $W(x) &lt; 0$ (threshold value); 0 otherwise</td>
</tr>
<tr>
<td>Gravity</td>
<td>$E(x)g$</td>
<td></td>
</tr>
<tr>
<td>Specific cases:</td>
<td></td>
<td>1 if $E(x)g &lt; 0$ (threshold value); 0 otherwise</td>
</tr>
<tr>
<td>Cost weighted opportunities: $f = 1$</td>
<td></td>
<td>1 if $E(x)g &lt; 0$ (threshold value); 0 otherwise</td>
</tr>
<tr>
<td>Aggregation: $f = 1$ opportunities exert a more than proportional effect on accessibility</td>
<td></td>
<td>1 if $E(x)g &lt; 0$ (threshold value); 0 otherwise</td>
</tr>
<tr>
<td>Complement: $f = 1$ opportunities exert a less than proportional effect on accessibility</td>
<td></td>
<td>1 if $E(x)g &lt; 0$ (threshold value); 0 otherwise</td>
</tr>
<tr>
<td>Mean travel cost to 4 nearest facilities</td>
<td>$E(x)g &lt; 0$</td>
<td>1 if $E(x)g &lt; 0$ (cost of traveling to 4th nearest facility); 0 otherwise</td>
</tr>
<tr>
<td>Presence of at least one facility within predetermined distance</td>
<td>$E(x)g &lt; 0$</td>
<td>1 if $E(x)g &lt; 0$ (cost of traveling to 4th nearest facility) &lt; 7</td>
</tr>
</tbody>
</table>

Figure 2

Relative accessibility compared to normative (fixed threshold)

Source: Paez, Scott and Morency (2012)

The maps illustrate the positivistic cumulative opportunities indicator for Profiles 1: Female living with a couple and children (left); 2: Female, lone mother (center); and 3: Female, lone mother, no auto or driver license (right).
(1) Actual Commute Cost Measures by API method

Estimation of travel time between a set of origins and a set of destinations through a transportation network is a common task in spatial analysis. Studies in accessibility analysis indicates that an accurate estimate of travel time is essential in spatial analysis (Handy and Niemeier, 1997). Calibrating the O–D travel time matrix in GIS package requires extensive data collection and process to prepare the road network. It also require the adequate and specific knowledge of the software to implementation. Neither can be easily achieved without trivial efforts. Wang & Xu (2011) investigates the potentials of completing the task by calling the Google Maps API. The Python program automates the process by reading the layers of origins and destinations in geographic coordinates, executing a HTTP request to access the Google Maps and calibrate the travel time between each O–D pair, and saving the results in an external ASCII file (Figure 3).

Figure 3
Travel time estimation process by Google Maps API.
Source: Wang & Xu, 2011
Wang and Xu (2011) demonstrate several advantages by API methods compared to GIS or other conventional approaches: (1) there is no need of preparing a road network, (2) it uses more updated road networks, and (3) it accounts for congestion in high-traffic areas and peak hours. They also make visualization comparison between API method and conventional GIS method (see Figure 4).

Figure 4
Accessibility standardized Z score (a) by ArcGIS and (b) by Google Maps API
Source: Wang & Xu, 2011
2.2 Ridesharing and the Measurement of Share-ability

This section investigates the history, development, trend, benefit and technology of ridesharing. It enables the debate on the positive regional impacts on city’s congestion, as well as positive individual impacts on the commute cost side for females in Riyadh. Nevertheless, although ridesharing is always good as it increases the travel options for residents in the city, the feasibility of ridesharing is still a question that depends on specific urban context and transportation patterns. As this research also targets at asserting its specific technology of urban computing, network analysis and API method. To enable this, it also requires a robust examination on current methods of calculating the feasibility of ridesharing.

2.2.1 Ridesharing around the World

Car sharing started in the late 1980s, where new schemes emerged to market to share vehicles to move (Harms and Truffer, 1998). Most of the car sharing schemes was implemented in a small scale at that time. The argument of the principles of car sharing has always been the reduction of ownership and use of cars, reduced fuel consumption and environmental emissions (Shaheen and Elliot, 2006). Fallon et al. (2004) examined impacts of various factors on the potential of car sharing. They investigated (1) the influence of distance, (2) level and intensity of public transport options, (3) the possibility of waiting, and (4) the potentials of cycling. The findings include that, as the distance increased, the probability of using car sharing decreased. Meanwhile, it is more likely to
use car sharing where less public transport exists (Louviere et al., 2000). Fatmi and Habib (2014) demonstrate the urgent need to educate people about the positive externality of ride sharing on the economic, environmental and social benefits. Structure equations modelling (SEM) was adopted to investigate ride mode choice, behavior and attitudes towards car sharing. Al-Atawi (2016) investigated the commuter’s attitude towards ridesharing in unique cultures of Middle East in the population segments of Saudi males and Saudi females, and established that there could be opportunities for car sharing for both male and females.

2.2.2 Measure the Share-ability

Sharing taxi trips is a possible way of reducing the negative impact of taxi services on cities, but this comes at the expense of passenger discomfort quantifiable in terms of a longer travel time and loss of privacy (Dueker, Bair, & Levin, 1977). Meanwhile, due to the limitation of computational ability, taxi sharing has traditionally been approached on small scales, such as within airport perimeters, or with dynamical ad-hoc heuristics (Al-Atawi, 2016). However, a mathematical framework for the systematic understanding of the tradeoff between collective benefits of sharing and individual passenger discomfort is lacking. The use of real-time information allows the monitoring of the urban mobility infrastructure to an unprecedented extent, and opens up new potential for the exploitation of unused capacity (Santi, Paolo, et al, 2014). The group of researchers at MIT SENSEable City Lab introduce the notion of share-ability network to enable the modelling of the collective benefits of sharing as a function of passenger inconvenience. The method computes
optimal sharing strategies on massive datasets (see Figure 5). It translates spatio-temporal sharing problems into a graph theoretic framework that provides efficient solutions (Santi, Paolo, et al, 2014). Applying this method to a dataset of 150 million taxi trips in New York City, its simulations reveal the vast potential of a new taxi system in which trips are routinely shareable while keeping passenger discomfort low in terms of prolonged travel time (see Figure 6). The analysis indicates that, with increasing but still relatively low passenger discomfort, cumulative trip length can be cut by 40% or more (Tachet, Remi, et al, 2016). This benefit comes with reductions in service cost, emissions, and with split fares, hinting towards a wide passenger acceptance of such a shared service.

Figure 5
Share-ability networks method
It translates the spatio-temporal issue of ridesharing feasibility into a graph-theoretic framework so as to provides efficient solutions in a system-optimization way.

Credit: figures and captions adjusted from SENSEable City Lab at MIT
2.3 Limitation and Restrictions on Saudi Females' Access to Jobs

This section reviews the knowledge of current restrictions on the movement and job participation for Saudi females. By relating the social and culture cost of these limitations on human rights with Saudi Arabia and Riyadh’s urban structure, urban fabric, extreme weather\(^1\) and other context specifics, the research aims to establishing the relationship between the regulatory limitations, the physical barriers and the financial burdens for Saudi females on their daily commute choice and employment opportunities.

\(^1\) See Table 3
2.3.1 The Driving Ban and Job Limits on Saudi Women

Women’s rights in Saudi Arabia are limited in comparison to many of its neighbors. The World Economic Forum 2015 Global Gender Gap Report ranked Saudi Arabia 134th out of 145 countries for gender parity (World Economic Forum, 2015). Among the indicators of gender parity and human rights, mobility is a significant consideration. It was since 1957 Riyadh pronounced a ban on women driving (David Commins, 2009). The Saudi ban on women drivers forces women to rely on their husbands, brothers, fathers, sons, and hired male help to take their children to school, run basic errands, and travel to and from jobs if they work. Women’s driving rights constitute one of only a handful of issues that Saudis have come out to the streets to support since the start of the Arab uprisings in early 2011. (Lori P. Boghardt, 2013) Saudi Arabia is the only country in the world where
women are not allowed to drive (Laura Bashraheel, 2009), as a result transportation is
definitely an issue. Women in the kingdom are not allowed to drive and must be veiled in
public. Moreover, Saudi women can’t work in many sectors of the economy, and anything
that involves working with or on men is prohibited. Traditionally, Saudi women have
worked in a limited number of sectors—like education—and currently comprise about 22%
of the employment market (Gorney, 2016). Until recently, they were finally excluded from
industries like retail, hospitality and even law.

According to a report from Arabia News, lack of access to autonomous transportation,
constantly needing guardian approval, and strict job restrictions that involve gender
segregation or gender bias are the top deterrents for Saudi women looking for
employment (Pwc, 2010). While the country is keen on increasing female employment
rates within the kingdom, providing affordable and feasible transportation options for
Saudi female is definitely an urgent task.

2.3.2 No Public Transportation Exists in Riyadh

There are only few alternative options for women on transportation. First of all, although
Riyadh is building its metro system, there are years from completion. According to Arriyadh
Development Authority, until 2019, the proposed public transportation system would be
completed. Moreover, women are generally discouraged from using public transport.
Buses operate on limited routes and are mostly used by men (Alexandra Zavis, 2015). In
early 2010, the government began considering a proposal to create a nationwide women-
only bus system. Activists are divided on the proposal, some saying it will reduce sexual harassment and transportation expenses, while facilitating women entering the workforce (Alexandra Zavis, 2015). Others criticize it as an escape from the real issue of recognizing women's right to drive (Gulf Times, 2010). While public transportation projects are going to be completed in 2019 in Riyadh, the ridership on the public systems is still not clear. Due to Riyadh’s vehicle-oriented urban form and the year-round high temperatures, most of commuting trips are far beyond the reach of pedestrian. As a result, Saudi female’s commute choice is currently limited by the driving ban, the extreme temperature, the vehicle oriented urban structure. They have to rely on commute options that are of high expense due to the comprehensive restrictions from regulatory, religious culture, and physical obstacles. For now, Saudi women are limited to the following options:

(1) Depend on their male family members (husband, father, brother, etc.)

(2) Rely on employer’s own ridesharing program (such as commuting shuttle buses)

(3) Take informal ridesharing service (females from different families co-hire a private driver). The expense is cheaper than taxi, but it is not very flexible

(4) Take private taxi or online hailing taxi service (informal taxi or online service like UBER). It is flexible, cheaper than street hailing taxi however it is more expensive than ridesharing.

(5) Take street-hailing taxi service, which is the most expensive one.

See Figure 7 and Figure 8. Figure 7 demonstrates that the size of Riyadh’s urban block is out of human scale, the minimum block is still as large as around 2KM. The city is still sprawling at this scale. Moreover, in Figure 8 it shows the urban design and street quality in Riyadh. It is a very typical vehicle oriented urban planning.
Riyadh is not a pedestrian friendly city. It is urban grid is at least 2KM. The city is surrounded by desert and with yearly-round high temperature.

Source: from Google map

Figure 8
A Vehicle Oriented City
Source: http://www.thenational.ae/business/property/riyadh-housing-prices-expected-to-rise-after-two-years-of-decline
2.3.3 High Commute Cost as a Barrier to Employment

While the country is keen on increasing female employment rates within the kingdom, providing affordable and feasible transportation options for Saudi female is definitely an urgent task. According to a recent investigation by Arab News, the wages of private drivers have been increasing over the years, sometimes accounting for nearly half the amount of a female employee's monthly income (Jeddah, 2012). Women are up in arms over having to depend on relatives and drivers for travel.

Taxi drivers take advantage of the rush hour and charge more per ride to and from work (Jeddah, 2012). According to the interview of the same investigation, to hire a private driver cost a Saudi female SR 1,500 to SR 2,000 in average in a month. For those women who do not have a private driver and face the daily hassle of hiring a taxi, they are eager for more affordable and convenient solutions to end their transport woes:

“As drivers, our demand has its reasons.” said 44-year-old Indian driver, Mohd. Alawi. “The maintenance of the car costs me around SR500 a month, just enough to demand SR1, 200 per month from my female customers.” He said if a driver abides by road rules, is cordial with his female customers, and above all, maintains a clean, air-conditioned vehicle, then it is only reasonable that he demands between SR1, 500 and SR2, 000.

“Drivers do charge a rather high rate per month,” said Reem Gazzaz, a 47-year-old HR manager. “The company I work for pays SR800 of my transportation fee, so I only have to pay my driver SR400 out of my own pocket. My driver’s income of SR1, 200 per month is pretty low compared to mine.
I’m sure other women only get paid a small sum per month. They might find it difficult to cope with these drivers’ demands.” Gazzas stated that the driver’s demands are climbing and that it is the company’s duty to look after its female employees and provide them with the necessary transport, especially because women are not allowed to drive in Saudi Arabia.

Ahmed Hamar, 32, works part time as a private driver for a Saudi family. He complains he doesn’t earn enough to serve his family back home. “I make only SR1, 500 in my current job,” said Hamar. “Mostly I transport Saudi women to their workplaces. My rates range from SR2, 000 to SR2, 500 per month. If I have to increase, it would only be because they are not punctual. I have another job to attend to and I am sacrificing my time to them.”

Najiba Rashed, a 28-year-old salesgirl said that she has been taking a taxi to work since she started working a month ago. “I am looking to hire a driver but I’m unable to find one for a reasonable fee,” said Rashed. “Several drivers I spoke to request very high amounts, usually SR1, 000 and above, which is almost my salary.” She said that usually when getting home after work she requests a friend, who has her own driver, to drop her home every day. “That is not something I can continue doing and it is very shameful,” she added.

The paragraphs above are all subtracted from the report by Jeddah (2012) on Arabia News. Even though some companies provide their female employees with transportation, most of the rest do not intend to do so. The burden of commute eventually falls on their families who are forced to pay a huge amount to these taxi drivers (Jeddah, 2012). Most of the interviewees expressed their concerns on the issue of high commute burden on taxi or private driver. The fact suggests that, local government should implement public transport...
or start a bus service for working women in the Kingdom. Meanwhile, the vast complains also indicate the potential demand for more affordable transportation service in the market of Saudi Arabia.

2.4 Ridesharing in Saudi Arabia

The sector will examine the ridership of ridesharing and informal taxi service in Saudi Arabia, to demonstrate the potentials and the demand in providing formal ridesharing service to Saudi females. For the Saudi women who do not have a private driver and face the daily hassle of hiring a taxi, they urgently require more affordable and convenient solutions. While women comprise only 13% of the Saudi workforce, they make up a full 60% of the college student population, so this is not an insignificant number of daily trips (Jeddah, 2012). Ridesharing companies like UBER are claiming to help on that (Alexandra, 2015). UBER’s Saudi Arabia general manager says that while his office doesn’t keep precise gender data, observation and anecdotal evidence suggest that 70% to 90% of Saudi Uber riders are women (Jeddah, 2012).

“...... We have some data to show that these women are starting to rely on Uber a lot more for their daily commutes; the proportion of trips that we see in Saudi during the weekday is actually very high relative to other locations. That’s just kind of one indicator to tell us that women are really starting to rely on Uber for their daily commutes to work, or to school, or to university.”
“I use Uber every time I’m [in Riyadh], to get to and from the airport, to go to meetings, or to visit friends,” says Shahd AlShehail, a Saudi national who lives in Dubai but returns home often to visit family and for her work as founder of Just, a platform that helps brands through data and storytelling. “Many of my friends do, as well as family. Usually to go to work, run errands, or go to a social gathering.”

2.4.1 Long Existence of Informal Ridesharing

Travel behavior research indicates that travel decisions are often influenced by accessibility as well as characteristics of the transport systems (Al-Atawi & Saleh, 2014). Factors such as travel times, travel costs, waiting times, walking times have the most significant contributions in mode choice and travel decisions (Alkaabi, 2014). In the case of developing countries however, most influencing factors for travel behavior and decisions are the social factors (Srinivasan and Rogers, 2005). In India, female walking trip shares were 52% higher than men’s in Pune, 61% higher in Bamako, and a 100% higher in Ashgabat (Al-Atawi and Saleh, 2013). In Saudi, there are informal arrangements of car sharing mainly within each household. For example, Al-Atawi (2016) looks at Tabuk city’s ridership and the female residents’ attitude towards ridesharing and other options. He investigates the factors which affect the decisions of car sharing in Saudi Arabia. Tabuk has typical travel characteristics of a medium size Saudi city. Workplaces as well as schools for men are

---

3 Table 4 concludes the current ridership in a middle size city in Saudi Arabia.
4 Table 5 illustrates the attitude towards the potential shifts on ridership.
separate than work places for women. Because of the social factors in Saudi Arabia, most male members of the family drive cars, while the female mainly rely on car sharing, hiring a private driver or using a taxi. The suitability of and attitudes for car sharing as well as alternatives in relation to different trip purposes were investigated (Fatmi and Habib, 2014). The social factors are therefore strongly influencing when it comes to choice of mode of travel (Attiyah and Wafaa, 2014). Female travelers depend on using car sharing more than men. Informal ridesharing service is the most popular substitution option when private vehicle is not available, according to the investigation. Moreover, public transportation is not even in most of the respondents’ consideration.

Table 4
How respondents usually reach their place of work
Source: Alkaabi (2016), Attyiah and Wafaa (2014)

<table>
<thead>
<tr>
<th>Mode</th>
<th>Never</th>
<th>Seldom (1 or less times per week)</th>
<th>Sometimes (2-3 times per week)</th>
<th>Regularly (4-5 times per week)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passenger Informal Car sharing</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.17</td>
<td>0.69</td>
</tr>
<tr>
<td>Passenger car sharing</td>
<td>0.5</td>
<td>0.1</td>
<td>0.0</td>
<td>0.03</td>
<td>0.63</td>
</tr>
<tr>
<td>Contractor driver</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.09</td>
<td>0.69</td>
</tr>
<tr>
<td>Private driver</td>
<td>0.5</td>
<td>0.1</td>
<td>0.1</td>
<td>0.08</td>
<td>0.68</td>
</tr>
<tr>
<td>Drive</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.54</td>
<td>0.75</td>
</tr>
<tr>
<td>Total</td>
<td>1.78</td>
<td>0.44</td>
<td>0.31</td>
<td>0.91</td>
<td>3.43</td>
</tr>
</tbody>
</table>

Table 5
Choice of travel options should current mode becomes unavailable
Source: Alkaabi, 2016

<table>
<thead>
<tr>
<th>Choice</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walk</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Private Bus</td>
<td>22</td>
<td>17</td>
<td>14</td>
<td>29</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>Cycle</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Drive</td>
<td>55</td>
<td>7</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Chauffeur</td>
<td>53</td>
<td>40</td>
<td>31</td>
<td>16</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Contracted driver</td>
<td>44</td>
<td>85</td>
<td>62</td>
<td>17</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Formal car share</td>
<td>14</td>
<td>12</td>
<td>19</td>
<td>11</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Informal car share</td>
<td>147</td>
<td>101</td>
<td>41</td>
<td>19</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Taxi</td>
<td>59</td>
<td>85</td>
<td>54</td>
<td>28</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Totals</td>
<td>413</td>
<td>363</td>
<td>248</td>
<td>144</td>
<td>96</td>
<td>77</td>
</tr>
</tbody>
</table>
2.4.2 Emerging Formal Ridesharing Service in Saudi Arabia

Implementing programs and policies that aim at ridesharing is a very interesting policy measure which can influence car users’ behavior. Car sharing schemes are often constructed with the aim of encouraging car users to reduce their use of the private car and increase the use of shared modes of transport. By having more people using one vehicle, carpooling reduces each person's travel costs such as fuel costs as well as other operating costs (Scott et al., 2003). Carpooling is also seen as a more environmentally friendly and sustainable way to travel as sharing journeys reduces carbon emissions, traffic congestion on the roads, and the need for parking spaces. Authorities and local councils aim at encouraging carpooling, especially during high pollution periods and high fuel prices.

Car sharing popularity has increased in particular in the USA. Scott et al. (2003) reviewed car sharing schemes in more than 30 North American cities and 50 organizations throughout Europe. The authors showed that once car sharing is available, it is much more heavily used than the private automobile when available (Scott et al., 2003). This indicates that not only travelers who do not travel a lot would opt to car sharing, but also passengers from other mode including private vehicles could possibly shift their ridership. In addition, the car sharing concept would encourage a reduction in car ownership and usage.

As the public transportation system need a couple of years for completion and most of the users are men, providing alternative solutions is the only exit. Women are usually driven
around by family members and personal drivers, or are forced to use some other type of private transportation. It is technically forbidden, but unenforced, for women to take taxis or hire private drivers, as it results in “khalwa”, which means illegal mixing with a non-mahram man (IslamQA, 2013). While the private transport is a booming business, the higher the demand, the more expensive the supply becomes. Some companies provide cars and drivers to ferry their women employees for work purposes, but not all companies have the budget to do that. Workingwomen, meanwhile, find it difficult getting to work and are often charged thousands of riyals a month in transportation.

The private sector, however, recognizes the potential of the chauffeur-driven car business. Car Service Company provides transportation and charges by the hour. Data shows that the service started three years ago and expanded due to high demand (Laura Bashraheel, 2009). On the other hand, Smartphone-based ride services are becoming increasingly popular in Saudi Arabia’s major cities, especially among the large number of tech-savvy young people. Customers include foreign businessmen who don’t want to deal with the country’s sometimes chaotic taxi system. Data shows that more than 80% of individual users are women (Alexandra Zavis, 2015).
Chapter 3. Data and Methodology

This study proposes to integrate qualitative, quantitative and spatial analysis methods via novel information technology such as network computing and API query. This chapter will state the data acquisition and the methods for measuring actual commute cost, measuring accessibility, and the cursory analysis of feasibility of ridesharing and its benefits. In particular, the data acquisition and methods developed aim to answer the following questions:

1. What are the commute costs of different transportation options to employment for Saudi females in Riyadh?
2. What are the differences in accessibility measured by commute cost?
3. Is there a spatial relationship between accessibility and employment?
4. Is ridesharing feasible in terms of the current taxi trip patterns and traffic conditions?

Section 3.1 discusses about the selection of research data source on the spatial distribution of Saudi females demography and employment activities, the spatial distribution of public transportation system, the commute cost data querying and the taxi trip data.

Section 3.2 describes the comprehensive methodology of API query, quantitative analysis, spatial analysis and network computing. It consists four parts. The first part provides a step-by-step method that query realistic commute time and distance and then applies fare estimation equations to transfer them to money costs. The second part explains the reason
why this research selects cumulative opportunities approach to measure the accessibility, and then introduce its specific steps. The third sector shows the procedures of spatial analysis and quantitative analysis of the mismatch between female residents’ accessibility and employment opportunities. Last but not the least, the fourth part provides the basic algorithm of calculating share-ability and the method of measuring its benefits.
3.1 Data Selection

This section explains the data selection and data source for this research. The research uses data on the geographic location of Saudi females, employment opportunities for females, planned public transit system and the vehicle road network to examine the spatial difference in females’ access to employment in Riyadh. The quantitative analysis of Saudi females’ demography situation, employment outcomes and economic opportunities would be based on CENSUS data and GIS shape files\(^5\). Data of all taxi trips in 2016 in Riyadh by passenger’s gender\(^6\) profiles would be analyzed. The intention is to qualitatively describe current taxi users’ behavior and their preference. The pattern indicates an unequal spatial distribution and gender distribution as a result of the biased transportation policy as well as the urban spatial structure. The taxi dataset is also deployed to initiate a cursory analysis on the feasibility and benefits of ridesharing.

In general, the dataset covers four categories, namely (1) Saudi females’ demography, (2) Saudi female’s employment opportunities, (3) Riyadh’s transportation network, (4) transportation service fare estimation formula, and (5) taxi trips data.

---

\(^5\) CENSUS data of Riyadh in 2015 is provided by Evidence for Policy Design, as well as the GIS shape files of Riyadh neighborhoods and future metro systems.

\(^6\) Taxi trips data in 2016 in Riyadh is provided by Civic Data Design Lab
3.1.1 Saudi Females’ Demography

1. Geolocation and region of Riyadh’s municipality, sub-municipality and neighborhoods (CENSUS 2016, Riyadh)

   Data format: GIS shapefile
   
   Data source: Evidence for Policy Design, Harvard Kennedy School, required from Arriyadh Development Authority

2. Saudi female’s demography information such as the total population, commuter population, job seeker population, household car ownership, employment rate, etc. Aggregated in neighborhood scale (CENSUS 2016, Riyadh)

   Data format: Excel tables
   
   Data source: Evidence for Policy Design, Harvard Kennedy School, required from Arriyadh Development Authority

3.1.2 Saudi Females’ Employment Opportunities

1. List of Saudi female’s employment opportunities, and their corresponding average income by sectors (Education, Healthcare, Manufactory, and Retail)

   Data format: Excel table
   
   Data source: Evidence for Policy Design, Harvard Kennedy School, required from Arriyadh Development Authority

3.1.3 Riyadh’ Transportation Network

1. Riyadh Road Network
2. Riyadh Public Transportation Plan (Metro, BRT and Buses)

   Data format: GIS shapefile
   Data source: Evidence for Policy Design, Harvard Kennedy School, required from Arriyadh Development Authority

3.1.4 Transportation Services’ Fare Estimation

   1. Fare Estimation Equations by different ride service companies
      Data format: mathematical formulation
      Data source: Evidence for Policy Design, Harvard Kennedy School, required from Arriyadh Development Authority

   2. Length of time (in minutes) and distance (in KM)
      Data format: CSV table
      Data source: Google Direction API

3.1.5 Taxi Trips in Riyadh

   1. Taxi trips in 2016
      Data format: CSV table
      Data source: Civic Data Design Lab, MIT
3.2 Methods

This section will explain the comprehensive methodology consists of API query, spatial analysis, quantitative analysis and network computing approaches step-by-step. The integrated methods are divided to four categories by their purpose and goals.

(1) Geocode the CENSUS data and Saudi Females’ Employer data
(2) Query the actual commute cost through Google Direction API
(3) Spatial analysis of Saudi females’ access to employment through cumulative opportunities measures
(4) Quantitative analysis on the relationship between unbalanced distribution of access to jobs and the employment rate as an outcome
(5) A cursory investigation that quantify ridesharing’s feasibility and its benefits using network analysis method based on taxi data.

To enable the quantitative analysis of correlations between accessibility and economy opportunities, it is crucial to acquire Saudi females’ actual commute costs. Noticing that the commute time and distance query from Google Direction API is reliable for transportation studies (Wang, F., & Xu, Y., 2011), the research proposes to calculate the realistic commute cost in time and distance through API. Apply the first hand commute
time and distance data to different fare estimation functions\(^7\) acquired, the actual commute cost in the form of local currency SAR\(^8\) can be estimated.

The spatial analysis work would be built upon its own urban information system of commute cost calculation. Cost of different transportation options are compared, and the geolocation and distribution of Saudi female’s demography and the corresponding accessibility are correlated. Correlate the travel cost especially with Riyadh’s CENSUS data, regression modeling are feasible.

### 3.2.1 Geocode CENSUS data and Employers’ information

To conduct the analysis, the study draws on data assembled from a variety of sources, including the most updated CENSUS data\(^9\) which contains administrative and demography information for all residents in an aggregated scale cross nationality (Saudi and non-Saudi) and gender (male and female) in Riyadh in 2016. The CENSUS data is joined with GIS shapefile that has neighborhood boundaries assigned with both a map position by neighborhood INDEX number.

---

\(^7\) The fare estimation equations are acquired from different transportation companies by Jawaher. Jawaher is a research fellow at Evidence of Policy Design, Harvard Kennedy School. She is also the reader of this thesis. Taxi companies includes the most common street-hailing taxi service in Riyadh; the local online hailing service company CAREEM, online hailing service UBER original from SF in the US, and local ridesharing service provider KASPER. Currently KASPER does not have service in Riyadh, but it offers carpooling in other cities in Saudi Arabia. Equations can be founded in Table 5.

\(^8\) 1 SAR = 0.267 USD

\(^9\) Provided by EPoD, and original requested from Arriyadh Development Authority.
(1) Spatially join the CENSUS and population projection data

For the CENSUS 2016 data, of the 186 neighborhood records, 90 records (50% percent) were successfully geocoded (see Appendix II, CENSUS 2015). The administrative data were also matched to population projection data in 2030 to identify those neighborhoods where have the most population density growth after the metro project’s completion (see Appendix II, Population Projection 2030).

(2) Geocode the employer data

Saudi females’ employer list is geo-located by Google Geocoding API to determine the concentration of potential job markets by industrial sectors for Saudi female population (see Appendix II, Saudi Female Employers). The names of the Saudi females’ employment opportunities are in the sectors of Education, Healthcare, Retail and Manufacturing. These data are geocoded with longitude and latitude by Google Geocoding API, assigned a map position and a sector category value.

10 The population projection in 2030 GIS file is provide from the Civic Data Design Lab at MIT.
3.2.2 Commute Cost Query via Google Direction API

Riyadh has a varied urban structure, including a medium-sized metropolitan area, new districts and neighborhoods scattered throughout the central and periphery areas. It is also a municipality that is experiencing rapid population growth, has low employment rate in Saudi female fragment. Although Riyadh is the most high density municipality in Saudi Arabia (Saudi Arabia Population, 2016), it still contains a vast expanse of non-urbanized or semi-urbanized land. As a result, distances from the remote areas of Riyadh into the urban area where most female employment opportunities and other social and religious culture activities concentrates can be lengthy since the city area is large, approximately 1,800 square km (Alkaabi, 2014).

(1) Set up the 2KM grids matrix

The average block scale in the central area of Riyadh is around 2 kilometer. As a result, this research sets up its geospatial analysis framework by dividing the region of Riyadh City to a matrix of geolocation zones by 2KM x 2KM fishnet grids. It is the basic geospatial structure for further urban computing and spatial investigation. Based on the grid framework, this study promotes the steps of computing the actual commute time and distance via Google direction API by Python codes in the back end.
Figure 9: Divide the region of Riyadh to a matrix of 2KM x 2KM grids
Credit: by Waishan QIU

(2) Query the commute distance and time via Google Direction API

Secondly, the realistic commute distance and time duration from each of the grids to all job opportunities are estimated via Google Direction API. Figure 11 is a front-end interface to illustrate how a typical query works. The computing of realistic travel time and distance by driving from centroid point of any grid i to any center of grid j is enable by python. Commute time and distance are then recorded in csv tables.
Figure 10: Travel Time and Distance Query in Python
Credit: developed by Waishan, original from GitHub.

```python
# Googlemap query and write csv files for each grid
for i in List_S(9S39I5]
    row = 0
    df_new = pd.DataFrame()
    print 'Start Point: '+str(i)
    start_loc = (df1.Latitude[i], df1.Longitude[i])
    for j in Differences:
        row = row+
        end_loc = (df1.Latitude[j], df1.Longitude[j])
        tmp_key = str(j)+"="+str(i)
        result = gmaps.directions(start_loc, end_loc, mode = 'driving')
        duration = str(result[0]['legs'][0]['duration']['text'])
        distance = str(result[0]['legs'][0]['distance']['text'])
        if i in List_SSE1 and j in List_SSE1:
            if row:
                duration = '0'
                distance = '0'
            else:
                result = gmaps.directions(start_loc, end_loc, mode = 'driving')
                duration = str(result[0]['legs'][0]['duration']['text'])
                distance = str(result[0]['legs'][0]['distance']['text'])
                dict_Result[tmp_key] = (duration, distance)
                duration = convert_to_min(duration)
                distance = convert_to_km(distance)
                df_new.loc[row, 'Distance_km'] = dist
                df_new.loc[row, 'Distance_min'] = dur
                df_new.loc[row, 'Duration_min'] = duration
                df_new.loc[row, 'Duration_km'] = distance
                df_new.loc[row, 'UNIQUE_ID'] = i
                df_new.to_csv('Data_Output/Data/UNIQUE_ID_i.csv', index=False, encoding='utf-8')
            print 'Done starting Grid: '+str(i)+', whose location is: '+str(start_loc)+', with ID '+str(df.FID[i])+'\n'
        else:
            result = gmaps.directions(start_loc, end_loc, mode = 'driving')
            duration = str(result[0]['legs'][0]['duration']['text'])
            distance = str(result[0]['legs'][0]['distance']['text'])
            dict_Result[tmp_key] = (duration, distance)
            duration = convert_to_min(duration)
            distance = convert_to_km(distance)
            df_new.loc[row, 'Distance_km'] = dist
            df_new.loc[row, 'Distance_min'] = dur
            df_new.loc[row, 'Duration_min'] = duration
            df_new.loc[row, 'Duration_km'] = distance
            df_new.to_csv('Data_Output/Data/UNIQUE_ID_i.csv', index=False, encoding='utf-8')
            print 'Done starting Grid: '+str(i)+', whose location is: '+str(start_loc)+', with ID '+str(df.FID[i])+'\n'
```

Figure 11
Front-end User Interface of Google Direction API
Source: Google Direction API
(3) Fare estimation

Following that, different fare estimation equation offered by various transportation service companies (see Table 6) are deployed to calculate the commute cost derive from each grid to all the job locations measured by local currency in the form of SAR. This realistic estimation of commute costs become indicators for the positive accessibility measurement.

Table 6: Fare Estimation of Different Services

<table>
<thead>
<tr>
<th>Company</th>
<th>Fare estimation equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxi (traditional)</td>
<td>Cost_{Taxi} = 5 + 1.7D + 0.7T</td>
</tr>
<tr>
<td>Careem (local ride-sharing)</td>
<td>If Cost_{Careem} &gt; 7, Cost_{Careem} = 3.2 + 1.15D + 0.38T</td>
</tr>
<tr>
<td></td>
<td>If Cost_{Careem} &lt; 7, Cost_{Careem} = 7</td>
</tr>
<tr>
<td>UberX (global ride-sharing)</td>
<td>If Cost_{UberX} &gt; 8, Cost_{UberX} = 3.2 + 0.9D + 0.25T</td>
</tr>
<tr>
<td></td>
<td>If Cost_{UberX} &lt; 8, Cost_{UberX} = 8</td>
</tr>
<tr>
<td>Kasper (Carpooling)</td>
<td>Cost_{kasper} = 0.6 \times (3.2 + 0.9D + 0.25T)</td>
</tr>
</tbody>
</table>

(4) Criteria to define affordable costs for Saudi female commuter

Comparing Riyadh to cities of comparable area size or density, it indicates that, the greater the population density, the lower commute cost tends to be as a proportion of income. A secondary factor that reduces commute cost is city size, as distances become more accessible, costs diminish. According to Table 7, Riyadh is closest to Los Angeles in size and density. The commute cost in LA is 19% of total monthly income, which offers a useful benchmark for determining affordability. As a result, 20% of monthly income is set up as the benchmark for determining affordable trips.
Table 7
Commute Cost (as a Proportion of Income) in Different Cities Worldwide
Data Source: by Jawaher Al Sudairy, Evidence for Policy Design, Harvard Kennedy School

<table>
<thead>
<tr>
<th>City</th>
<th>Area (sqkm)</th>
<th>Density (per sqkm)</th>
<th>Commute Cost (to a proportion of income)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>1,213</td>
<td>10,831</td>
<td>9%</td>
</tr>
<tr>
<td>Boston</td>
<td>232</td>
<td>5,344</td>
<td>11%</td>
</tr>
<tr>
<td>Miami</td>
<td>143</td>
<td>4,770</td>
<td>19%</td>
</tr>
<tr>
<td>Chicago</td>
<td>606</td>
<td>4,447</td>
<td>15%</td>
</tr>
<tr>
<td>Washington DC</td>
<td>177</td>
<td>4,308</td>
<td>10%</td>
</tr>
<tr>
<td>Riyadh</td>
<td>1,798</td>
<td>3,331</td>
<td>Suggest 20% as an appropriate range</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>1,302</td>
<td>3,198</td>
<td>19%</td>
</tr>
<tr>
<td>Detroit</td>
<td>3,350</td>
<td>1,985</td>
<td>21%</td>
</tr>
<tr>
<td>Houston</td>
<td>1,625</td>
<td>1,414</td>
<td>20%</td>
</tr>
<tr>
<td>Austin</td>
<td>704</td>
<td>1,296</td>
<td>28%</td>
</tr>
</tbody>
</table>

(5) Commute Cost Visualization

Figure 12 visualize the result of the spatial difference of commute cost (to the proportion of monthly income) by different transportation services. The trips are derived from the central point of all the grid polygons to the nearest female market location in manufacture sector.
3.2.3 Cumulative Method of Accessibility Measurement

Due to lack of available data on Saudi females' sensitivity to the trade-off between time and money and on the utility of Saudi females' commute choice, the most fundamental approach of accessibility measurement- the cumulative opportunities measures is deployed in this study. This measurement of accessibility counts the number of
opportunities reached within a given travel time (or distance) (Geurs & Van Wee, 2004). Because all potential destinations within the cutoff time are weighted equally, regardless of differences in travel time, this type of measure emphasizes the number of potential destinations or opportunities rather than their distance.

Figure 13
Cumulative Method of Accessibility Measurement
Credit: by Waishan QIU

In this study, the cutoff travel time is replace with cutoff commute cost in SAR, to demonstrate the financial burden of commuting trips as an economic barrier. Additionally, it fits the initial intention of comparing and promoting more innovative and affordable commute options. Figure 13 illustrate the principle and approach of measuring Saudi females’ accessibility to their employment opportunities. The cutoff commute cost is set at 20% of the average monthly income of corresponding sectors. Cumulative accessibility is then measured as the count of Saudi female employment opportunities in the sectors of Education, Healthcare, Manufacturing and Retail that can be reached within the cutoff
commute cost by different traffic modes. For example, $A_{cumulative_i} \_Kasper$ measures the total jobs that can be reached within 20% of the monthly income as commute cost starting from the centroid point location of grid i.

### 3.2.4 Investigation of Spatial Mismatch Problem

The quantitative and spatial analysis draws from a series of maps from which the author pay particular attention to the differences of accessibility between central and periphery areas of Riyadh city.

**1) Dissolve CENSUS data into grids**

Figure 14 depicts Riyadh and its location within Saudi Arabia. The City of Riyadh (around 6 million population) is the largest city and the capital of the nation (Saudi Arabia Population, 2016). Thirty percent of urban residents live in the central area as marked in Figure 15. Of the remaining 70 percent, most of them live in new developed neighborhood scattered around the city center, and the remaining live in large, un-developed neighborhood and rural areas. Like most other resource-based economies, Riyadh’s Saudi female population is characterized by high unemployment rates. It is also noticeable that the average car ownership by household in the neighborhood in or near the city center is higher than the average of neighborhood in the periphery. However, the female commuter density poverty and welfare usage rates. Those demography information in neighborhood level are dissolved into the spatial analysis grids. Results are visualized below in Figure 15.
Figure 14 Riyadh’s Location
Source: Open Street Map
Saudi Arabia and Riyadh in a World Map

Figure 15 Riyadh’s GIS information and Intersect GIS with CENSUS data
Credit: by Waishan QIU, source: Open Street Map, Arriyadh Development Authority, Riyadh CENSUS data 2015
Riyadh’s Population Density and Car Ownership Information in CENSUS format

Intersect the CENSUS information and Dissolve them with the 2KM Grid Matrix in GIS

Landuse Diversity      Public Transportation Coverage      Population Density
Household Car Ownership     Female Commuter Density      Taxi Pickup Density in Jan 2016
Figure 16 Data Analysis Methods, Take Taxi Cost as an Example

Scatter Plot of Accessibility Measured by Taxi (Plot 1)
Credit: by Waishan QIU

Plot 1: Accessibility and Distance to City Center

Accessibility to Education Sector Measured by Different Modes (Plot 2)
Credit: by Waishan QIU

Plot 2: Accessibility to Education Sector Measured by Different Taxi Services
(by Taxi, Careem, UberX and Kasper)
(2) Linear regression model

Linear regression model of the following dependent variable and independent variables are built:

Dependent Variable: Saudi Female’s Job participate rate in a neighborhood ($R_{job}$)

Independent Variable: commute time ($C_{rs\_time}$), commute cost in SAR ($C_{rs\_money}$), resident’s neighborhood’s distance to city center ($D_{dist\_center}$), household car ownership ($Car_{avg}$), and household home ownership ($Home_{avg}$).

Model to be examine:

$$R_{job} = \beta_0 \times C_{rs\_time} + \beta_1 \times C_{rs\_money} + \beta_2 \times D_{dist\_center} + \beta_3 \times Car_{avg} + \beta_4 \times Home_{avg} + \beta_5 \times C_{c\_time} + \beta_6 \times C_{c\_money} + \beta_7 \times W + \beta_8 \times G + \beta_9 \times I_{median} + \epsilon$$

3.2.5 Measuring Share-ability and Its Benefits

Historical taxi trips\textsuperscript{12} enables the analysis examining patterns of taxi users over a period of 12 months in 2016. The qualitative and quantitative understanding of the breakdown of male and female users, their waiting time, commute time and commute cost all support the research hypothesis that, females in Arabia rely much more heavily on taxi to travel, as around 70% percent of the trips are by females. Additionally, females of taxi users have a high commute cost and longer commute time compared to male users. Spatial analysis is

\textsuperscript{12} Taxi trips data provided by Civic Data Design Lab at MIT
operated by dissolving the trips to analysis grids to understand pick up and drop off patterns. It also allows the research to determine the best times and locations where rides can be shared.

(1) Data Cleaning and Preparation

Among the around 36,000 taxi trips in 2016, the study first filter the trips by gender and delete invalid trips whose pick up and drop off location are out of the region of Riyadh. It also delete the records that has no distance, no cost, as well as the suspicious records that has drop off and pick up location nearby or a outlier cost per KM. 23749 records out of 69484 taxi trips are valid to operate the share-ability analysis on. Below is an example of a filter function to clean the trips data enable by python. Figure 17 shows the working phase in python and a quick visualization of cleaned taxi O-D flows.

In [10]:
## 2. Read hourly trip files, and save trip amounts in a dictionary
folder = 'CleanCsv\byHour/'
files = os.listdir(folder)
csv_files = []
for f in files:
    if f.endswith('.csv'):
        csv_files.append(f)
csv_files.sort()
# print csv_files
dict_L = {}
for f in csv_files:
    path = folder + f
df = pd.read_csv(path, encoding = 'utf-8', sep=',', engine='python')
dict_L[f] = len(df)
print('** Trips Amount of Hour Periods:', dict_L)

## 2.1 Read the most relevant trip hours within Riyadh Boundary Scope
df = pd.read_csv('CleanCsv/2016.csv', encoding = 'utf-8', sep=',', engine='python')
df = df.loc[(df['O_x'] > 45.5) & (df['O_y'] > 24.5) & (df['D_x'] > 45.5) & (df['D_y'] > 24.5) & (df['Gender'] == 'female')]
df = df.reset_index(drop=True)

Out[10]:
<table>
<thead>
<tr>
<th>Gender</th>
<th>Cost</th>
<th>Car_Ctgry</th>
<th>Dist_km</th>
<th>O_x</th>
<th>O_y</th>
<th>D_x</th>
<th>D_y</th>
<th>Trip_min</th>
<th>Time_inDay</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>12.25</td>
<td>Economy</td>
<td>2.47</td>
<td>46.685437</td>
<td>24.686311</td>
<td>46.686444</td>
<td>24.682366</td>
<td>6</td>
<td>0.416</td>
</tr>
</tbody>
</table>
Figure 17

Python Scripts Example (data reading, filtering, calculating, and plotting)

Credit: by Waishan QIU

```python
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Read data from a CSV file
data = pd.read_csv('data.csv')

# Filter data based on specific criteria
filtered_data = data[data['column'] == 'value']

# Calculate summary statistics
summary_stats = filtered_data.describe()

# Plot data using a bar chart
plt.figure(figsize=(10, 6))
plt.bar(filtered_data['column'], filtered_data['value'], color='blue')
plt.xlabel('Column')
plt.ylabel('Value')
plt.title('Bar Chart')
plt.show()
```

```
gender, cost, car_opry, dist_km, dist_m, x, y, ph_month, ph_day, ph_min, time_index
female, 102.70, Economy, 2.47, 56.686847, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01
female, 52.90, Economy, 2.46, 56.770535, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01
```
(2) Share-ability Measures

This measure is enabled with the great inspiration from MIT SENSEable City lab\textsuperscript{13}. Their seminal works on share-ability establish a framework to quantify the benefits and the feasibility of ridesharing based on taxi trips data. A two-trip sharing case\textsuperscript{14} model is adopted by this study to measure the feasibility and benefits of carpooling in Riyadh. Although the specific codes are long and tedious, it is worth explaining the underlying principles of the measures.

First of all, the measure starts from the cleaned dataset of trips made by females in Riyadh. Trips are sorted by their initial time, so that the time gap between any two adjacent trips can be efficiently calculated. The measure sets up a maximum waiting time $\Delta T$, which means that for the two passengers that are possible to share a trip, the $\Delta T$ is the longest extra delay they are willing to bear.

Then it comes to the step to measure whether the two trip $T_i$ and $T_j$ can be shared. According to the sorted order of trips records, any two adjacent trips are selected to test whether their interval gap of initial time $\Delta T_1$ is smaller than $\Delta T$. If so, then the network method is implemented to calculate the shortest distance traveled $d_{T_i T_j}$ and time duration


\textsuperscript{14} See the paper ‘Supporting Information for quantifying the benefits of vehicle pooling with shareability networks’ by Santi, Paolo, et al.
that is needed in order to share these two trips Ti and Tj. If the distance \( d_{\_\_ij} \) of the new paired trip Tij is shorter than the total distance of the two original trips’ distance \( d_{\_\_i} + d_{\_\_j} \), the algorithm consider the two trips Ti and Tj can be shared.

To be short, the whole process of calculation can be concluded as following:

- Set up a \( \Delta T \) as the total delayed time tolerance, say 10 minutes
- Set up a \( \Delta T_1 \) as the time interval of any two trips, say 3 minutes, using NetworkX to loop any 2 trips Ti & Tj whose started time is less than \( \Delta T_1 \)
- Set up a \( \Delta T_2 \) as the total delayed time for both of the trips shared due to the ridesharing
  \[
  (\Delta T_2 = t_{\_\_i} \cdot t_{\_\_j} + t_{\_\_j} - t_{\_\_i})
  \]
- If \( \Delta T_1 + \Delta T_2 < \Delta T \), then the two trips Ti and Tj is shareable.

**T**: Trip i, with origin location \( O_i \), destination location \( D_i \), starting time \( T_{start,i} \), duration \( t_i \), distance \( d_i \)

**Tj**: Trip j, with \( O_j, D_j, T_{start,j}, t_j, d_j \)

**\( \Delta T \)**: maximum delay time

**\( \Delta T_1 \)**: the interval/gap of the requesting time of trip T and trip Tj

**\( F_{\text{shortest network}}(P_1, P_2, ..., P_i) \)**: Given a physical road network consists of connected path segments, a function that use weighted shortest network algorithm to calculate and return the minimum total travel time \( t \) and total distance \( d \) traveling from a consecutive OD locations, for example from \( P_1 \) to \( P_2 \), \( P_2 \) to \( P_3 \), ..., \( P_i \) to \( P_{i+1} \). The first location \( P_i \) is the origin, and the second location \( P_{i+1} \) is the destination.

Define \( F_{\text{share-ability}}(O_i, O_j, D_i, D_j) \):

If \( T_{start,j} - T_{start,i} < \Delta T_1 \):

\[
(t1, d1) = F_{\text{shortest network}}(O_i, O_j, D_i, D_j)
\]

\[
(t2, d2) = F_{\text{shortest network}}(O_i, O_j, D_j, D_i)
\]

\[
(t3, d3) = F_{\text{shortest network}}(O_i, O_j, D_i, D_j)
\]
\((t4, d1) = F_{\text{shortest\_network}} (O_p, O_i, D_p, D_i)\)

\(T_{ij} = \min(t_1, t_2, t_3, t_4)\)

If \((T_{ij} - T_i) + \Delta T_i < \Delta T \text{ and } (T_{ij} - T_j) + \Delta T_i < \Delta T:\)

\(T_i \text{ and } T_j \text{ is shareable,}\)

\(\text{saved distance} = d_{T_i} + d_{T_j} - d_{T_{ij}}\)

After iterating the examination of whether any two trips can be shared across the whole dataset, we get a matrix of shareable condition, which can be write as a dictionary format. For example, if trip 1 can be paired with trip 2, trip 3, trip 4, then it can be saved as \(\{T1: T2, T3, T4\}\). As a result, the problem comes to system optimization, which means that as a system, it is beneficial to maximize the total distance saved or in another way, to minimize the total delayed time. In this cursory investigation, the study aims to maximize the total distance saved, and take it as the ultimate goal.
Chapter 4. Research Findings

The statement of the method in previous chapter has delivered the concrete part of the urban network computing and API method as a technical innovation and application in urban and transportation studies. Building on the capacity of computing and geospatial analyzing of the technology, this chapter is trying to extend the innovative methodology to concluding and understanding the spatial mismatch problem, the possibility of ridesharing as a formal service, and the share-ability and its benefits in Riyadh. In the end this chapter also discusses the limitation, the application of the methodology in this research, as well as its implications for future study.

In section 4.1, it demonstrates the findings from the spatial analysis of the relationship between Saudi female residents and their access to potential employments. This chapter investigate the spatial mismatch and spatial inequity in Riyadh’s urban forms. With realistic commute cost as the indicator to measure accessibility, the sector takes a quantitative looking at Saudi females commute cost as a financial burden.

In section 4.2, it continues to investigate the potential market for formal ridesharing service targeted at Saudi female segment by understanding the current usage and behavior of taxi trips and comparing the future metro, BRT and buses plan which would not be completed until 2019. Given the data base from quantitative study and geospatial
analysis, the target female population is quantified, and the possibilities of ridesharing as a supplemental connection to metro system is discussed.

The section 4.3 introduces a cursory description about the feasibility of ridesharing and the corresponding potential benefits via computing the 30,000 taxi trips data in 2016 in Riyadh. It quantifies the share-ability given different time window and measure the benefits on traffic congestion and CO2 emission.

Finally, section 4.4 discusses the major limitations of this research from the data and methodology aspect. To justify the legitimation of accessibility measurement in this study, more delicate sensitivity model should be investigated to measure the realistic utility of time and money cost for various income segments of Saudi female in Riyadh. To support the interpretation of ridesharing’s feasibility and benefits in Riyadh, not only taxi trips but other OD flow information is needed. Additionally the current ridership and the potential shift in ridership after formal ridesharing is available needs to be discussed. None of the future studies have been strictly testified, but proposing these discussions as the extension of a more solid and comprehensive methodology at its preliminary stage is a good start.
4.1 The Spatial Mismatch of Saudi Female and Their Jobs

This analysis is based on CENSUS data and Saudi females’ employment list. The CENSUS data is not comprehensive, and only estimates for a population of 4.6 million while the total Riyadh population is estimated at 5.6 million (Saudi Arabia Population, 2017). Furthermore, due to the failure of joining CENSUS data to the missing neighborhoods index, the geo-located maps in GIS only provide analysis for 95% of the total CENSUS population. As a result, the mapped and valid data size represents 68.66% of the Riyadh population in statistics (see Table 8).

Table 8
Sample Size of the Population in Riyadh

<table>
<thead>
<tr>
<th>Sample Source</th>
<th>Size</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapped Data (valid data after joined to GIS)</td>
<td>4,462,763</td>
<td>68.66%</td>
</tr>
<tr>
<td>Census Data (valid data from 2016 Census)</td>
<td>4,654,677</td>
<td>71.62%</td>
</tr>
<tr>
<td>Total Population (2013 Census Data)</td>
<td>6,499,528</td>
<td>100%</td>
</tr>
</tbody>
</table>

Data Source: Jawaher, EPoD, and Arriyadh Development Authority
4.1.1 The Spatial Location of Saudi Females

(1) Population Density and Growth

This analysis is based on 2015 census data. Population density is greatest in the city center, at a level of around 30-40 thousand per square kilometer. However the central area only constitutes around 30% of the total population. The majority of Riyadh residents (70% of the residents) live in the periphery, with concentration in the northeast and southwest neighborhoods (Figure 18).

Figure 18: Population Density in 2015
Credit: by Waishan QIU, Source: Arriyadh Development Authority
Between 2013 and 2030, neighborhoods in the north and east will become denser, while neighborhoods in the periphery are expected to gain an additional 600,000 people as a result of population growth (see Figure 19).

Figure 19: Population Density in 2030
Credit: by Waishan QIU, Source: Arriyadh Development Authority

From 2015 to 2030, the main population growth would happen in the northeast and southwest neighborhoods such as Dahrat AlBade, West An Naseem, Ar Rawcah and Al Maseef.
(2) Car Ownership Rate

Car ownership is pervasive in Riyadh, with only 3% of households do not own a car, and 40% rely on a single car. Average car ownership is higher in central neighborhoods than the periphery, which is an indicator of level of household income as well as access to economic opportunities (see Figure 20).

![Average Car Ownership in Riyadh](image)

**Figure 20**

Average Car Ownership by Neighborhood

Credit: by Waishan QIU

Source: Arriyadh Development Authority

In average most families owns 2 cars. Only 3% of the households do not own a car. The neighborhoods surrounding the city center see a higher car ownership in average. There are also large neighborhoods on the periphery area such as Khashm Al Ann see a high car ownership rate.

(3) Current Commuters and Job Seekers

Distribution of female employees and students who have to commute every day in Riyadh is highly dense in the central area. However, there is also a quite a large population of female commuters residing in the northeast (West An Naseem & Ar Rawcah) and the southwest (Dahrat Albade). Female commuters are estimated around 472,500 students, and 452,000 employed women. Constituting around 20% of Riyadh’s total census population. Female
Job seekers are estimated around 86,500, constituting 2% of Riyadh's total population, which is quite close to the national data on female employment rate (22%).

Unlike female employees and students, female job seekers are largely distributed in the periphery neighborhoods located in the northeast and southwest. Furthermore, neighborhoods where female job seekers are prevalent, their average car ownership rates are also relatively low. There is a clear spatial correlation between car ownership and access to employment, as females who are job seeking are concentrated in neighborhoods where car ownership is low. This study will investigate the spatial correlation further with regression model in the next sector.

Figure 21
The Concentration of Female Commuters and Job Seekers
Credit: by Waishan QIU, Source: Arriyadh Development Authority
4.1.2 The Spatial Distribution of Saudi Females' Employments

Due to the limitation on Saudi females' job participation, females' employment in Riyadh region is concentrated within specific industries, namely education, healthcare, public administration, retail and manufacturing (Figure 22).

Figure 22
Saudi Females' Employment in Riyadh in 2016
Data source: Jawaher, GASTAT Labor Market Survey, 2016

Education: 27%
Healthcare: 4%
Manufacturing: 9%
Retail: 27%
Construction: 22%
Others: 14%

Figure 23 shows the geographic distribution of Saudi females and their job opportunities in the city. A large portion of Saudi females (50%) live in the highly urbanized area of the city center. Around thirty percent live in the southwest area that are scattered around the county; and 20 percent live outside of city center. The highest concentration of Saudi female population is in the southwest and northeast quadrant of the urbanized area.
Employment is also concentrated in the urbanized area, although slightly less concentrated than Saudi females. Eighty percent of all female jobs are located in the urbanized area around city center. Both retail and healthcare employment are concentrated in the high dense ring area in and around the city center. These areas also see higher density of Saudi
female population. Among those Saudi female commuters and job seekers who living in and around the city center, spatial proximity to employment opportunities in retail, healthcare, and manufacture is relatively high. In the semi-urbanized areas, spatial proximity to retail and healthcare is still surprisingly high even though employment opportunities are fewer in number and more dispersed. Spatial proximity to manufacture is the lowest as the jobs in this sector is concentrated in industrial zone located in the southeast outskirt of Riyadh.

A comparison of the female population and their employment opportunities shows that employment is more concentrated than female population along the north-south highway (Route 65) corridor. Saudi females living outside of the corridor may experience spatial barriers to employment, as Figure 22 shows, female employment opportunities are less concentrated in the periphery relative to the corridor area.

Analysis of the salary range of female jobs in the private sector of manufacture, healthcare, education and retail indicates that, majority of working women (93%) earn less than SAR 4,500 in average monthly (Table 9). The mean salary for female employees in the private sector is estimated at SAR 3,600, which is equivalent to USD 1,000. By comparing Riyadh to other cities of similar size and density indicates that, commute cost tends to be smaller in cities with greater density and smaller areas. In Table 8, it can be concluded that, Riyadh is closest to Los Angeles in size and density. The average commute cost there is 19% of total income, which offers a useful benchmark for determining affordability of commute
cost to a proportion of monthly income in Riyadh. Saudi Labor Law sets the transportation stipend at 10% of base salary, however many low income earner indicate that their commute can go up to 40% (Jawaher, 2017). Estimating the available commute budget based on the mean salary within each sector indicates that maintaining commute costs below 20% of income requires services that cost around SAR 660 - 900 monthly. In other words, any trips that cost less than SAR 22.5 per trip is considered affordable in Riyadh across female job sectors.

Table 9

Affordable Commute Cost (as a Proportion of Income) in Different Sectors

Data Source: by Jawaher Al Sudairy, Evidence for Policy Design, Harvard Kennedy School

<table>
<thead>
<tr>
<th>Employment Sectors</th>
<th>Population (Percent)</th>
<th>Income (Mean)</th>
<th>Commute Cost (SAR) as a Proportion of Income</th>
<th>Affordable Cost Per Trip (SAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service &amp; Finance</td>
<td>13%</td>
<td>4500</td>
<td>450 900 1350 11.25 22.5 33.75</td>
<td></td>
</tr>
<tr>
<td>Social Services</td>
<td>7%</td>
<td>4150</td>
<td>415 830 1245 10.375 20.75 31.125</td>
<td></td>
</tr>
<tr>
<td>Telecommunications</td>
<td>2%</td>
<td>3650</td>
<td>365 730 1095 9.125 18.25 27.375</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>14%</td>
<td>4500</td>
<td>450 900 1350 11.25 22.5 33.75</td>
<td></td>
</tr>
<tr>
<td>Healthcare</td>
<td>4%</td>
<td>3600</td>
<td>360 720 1080 9 18 27</td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>9%</td>
<td>3300</td>
<td>330 660 990 8.25 16.5 24.75</td>
<td></td>
</tr>
<tr>
<td>Retail</td>
<td>27%</td>
<td>3300</td>
<td>330 660 990 8.25 16.5 24.75</td>
<td></td>
</tr>
<tr>
<td>Elect. Gas &amp; Water</td>
<td>1%</td>
<td>3400</td>
<td>340 680 1020 8.5 17 25.5</td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td>22%</td>
<td>3100</td>
<td>310 620 930 7.75 15.5 23.25</td>
<td></td>
</tr>
</tbody>
</table>
4.1.3 Unequal Access to the Planned Metro System

Riyadh’s growth has resulted in a vast urban landscape, pushing the majority of its residents outside of the central district. Hence, all residents have come to rely on the car as their prime mode of transportation, with only 3% of households who do not own a private car. The projective completion of public transportation network consist of metro lines, BRT and buses in 2019 has the potential to improve mobility issues for almost half of Riyadh residents. However several districts in the southwest and northwest will be left out. Those neighborhoods located away from the transit network will require affordable and innovative interventions that response to their level of income and trip density (Figure 24).

By 2030, as neighborhoods in the center increase in density, more people will benefit and organize around the metro and bus network. While the underserved population will grow in size, almost 3.4 million, the overall proportion of population that is beyond 5-minutes’ walking distance to the public transit system is expected to decrease from 48% to 44%. 

Figure 24

Access to Public Transportation Stations in Current Population Density

Credit: by Waishan QIU; Source: EPoD, Havard Kennedy School & Arriyadh Development Authority
(1) Gender Inequity in the Access to Public Transportation

While the metro and bus network will highly improve access for female employees and students living in central neighborhoods, there are neighborhoods left out in the outskirts. The network to be completed in 2019 will cover 48% of the current total population based
on 2015 Hay data. And almost 50% of female commuters will have access to transit stops within 5 minutes’ walk. However in the southwest and northeast neighborhood in the periphery area, there are still noticeable female population left behind. With an in-depth looking at the population that have access to the metro system by gender, there exists an issue of gender inequity (see Table 10). Across all nationality, 53.2% male residents in Riyadh live within the 5-minutes buffer to the future systems, while only 51.0% females can share the benefits. For the current commuters, only 49.8% females can reach the public transportation within walking distance while the ratio in male commuters is 53.7%.

Table 10
Access to Future Metro and Public Transit System by Gender

<table>
<thead>
<tr>
<th>Source: CENSUS 2015 &amp; Metro Plan from EPoD</th>
</tr>
</thead>
</table>

Conclusion: Female in Riyadh in average has less access to the future public transportation system. The situation is especially worse among Saudi female Job Seeker segment and non-Saudi female immigrant worker segments.

<table>
<thead>
<tr>
<th>Population</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Nationality</td>
<td>53.2%</td>
<td>51.0%</td>
</tr>
<tr>
<td>Commuter</td>
<td>53.7%</td>
<td>49.8%</td>
</tr>
<tr>
<td>Job Seeker</td>
<td>49.5%</td>
<td>50.1%</td>
</tr>
<tr>
<td>Saudi</td>
<td>47.4%</td>
<td>47.5%</td>
</tr>
<tr>
<td>Saudi Commuter</td>
<td>47.4%</td>
<td>47.7%</td>
</tr>
<tr>
<td>Saudi Job Seeker</td>
<td>46.2%</td>
<td>49.6%</td>
</tr>
</tbody>
</table>

(2) Spatial Inequity in the Access to Public Transportation

There is also a distinct spatial inequity regarding the coverage of public transportation system. From the plot in Figure 26 it can be concluded that, there are still some outliers whose population density is high (more than 20,000 people per square kilometer) while its access to public system is less than 60% percent. Which means that more than 40% of the population in those outliers’ dense area have difficulty in accessing the metro and bus
stops. For those females cannot afford a private driver, they call for more flexible and affordable options either for commute or for connecting them to the public stations.

Figure 26
Access to Public Transit & Demography Distribution by Neighborhoods
Source: CENSUS 2015 & Metro Plan from EPoD

Figure 26 illustrates the spatial inequity problem across neighborhoods in Riyadh. The top left plot shows the relationship between future metro system service level and the population density. It demonstrates that, there are several neighborhoods with high population density (more than 20,000 residents per km²) but their public transportation service coverage in 2019 is poor (less than 60%).
The top right plot illustrates the relationship between future metro system service and the current daily trips in the across the neighborhoods. It demonstrates that there are several outlier neighborhoods which have large demand of commuting (more than 20,000 trips) however the future metro system is not sufficient enough to support them.

Similarly, the bottom left plot of public transit service area and female ratio across neighborhoods demonstrates the planned metro system cannot fully serve the females due to the spatial distribution of females residents is not balanced. Meanwhile, the public transit plot (bottom right) demonstrates that there are some neighborhoods have low car ownership (less than 2) but the public transit service is also insufficient there. For those families do not own extra cars, improve their public transportation service level would significantly increase their mobility and accessibility.

In general, even though the future metro and buses system would largely increase the affordability of commute trips, there are still questions and spatial inequity underline. First of all, attitudes and preference on metro from residents especially from Saudi female commuter and their families are unknown. Secondly, while the spatial distribution of metro system would perform well, there are still outlier neighborhoods that need specific policy aids to increase their access, such as subsidy or targeted shuttle service. Urban planning strategies should be adjusted to re-balance the spatial concentration of residents or support specific transit service for female in those areas.
4.2 Commute Cost as the Barrier to Access Female Jobs

With spatial investigation and quantitative analysis by regression model, the relationship between commute cost and Saudi females’ employment outcome is examined. This analysis offers quantitative and spatial evidence for addressing current issue of both inadequate access and high commute cost to Saudi females’ employment.

Women’s employment in Riyadh can best be improved by ridesharing programs within Retail, Education and Healthcare sectors that are geographically distributed within and around the city center (see Figure 27). These three sectors are more condense and concentrated within city center. As a result, the benefits of ridesharing service is greater than that of manufacture sector whose location is on the southeast outskirt. Meanwhile, as the manufacture sector is along Route 65, the price buffer shows a strong correlation with the proximity to Route 65, as the most affordable location with a commute cost below SAR 16.25 or 20% of the monthly income are all along Route 65.

The plots in Figure 28 demonstrates that ridesharing has the largest potential to reduce the commute cost. The slopes of ridesharing scatters see the gentlest upslope when the distance to city center increase. Take Education sector as an example, origin from the neighborhoods that are around 20 KM to city center, the commute cost by Taxi can be SAR 45 more expensive than the service by ridesharing per trip. When the distance to city center increases to 40 KM, which is probably the new development area on the outskirt, the commute cost of taxi is SAR 75 higher per trip than carpooling. Commute cost
simulation in other plots indicates that across female job sectors, ridesharing has the largest potential to reduce commute cost, especially for the residents in the periphery.

Figure 27
Commute Price Buffer at 10%, 20%, 30% and 40% of the Monthly Income to Females’ Job Sectors by Different Options

Credit: by Waishan QIU
Figure 28

Commute Cost to Nearest Job Market by Different Options

Credit: by Waishan QOU

It can also be concluded from Figure 29 that, in current situation, Saudi female commuters who rely on taxi and private drivers bear a very high burden to commute. Only residents live in city center’s neighborhoods or certain period along Route 65 would afford taxi service as a daily commute option. Residents along the Route 40 and 65 within the city center would have access to more than 20 job opportunities if taxi or hiring a driver is the
only service, while the rest of residents can access less than 10 jobs within 20% income buffer as commute cost (Figure 30). Online hailing service of Careem and Uber increase the access of 10 to 20 more jobs for 50% female population. Ridesharing can provide affordable access to more than 30 potential jobs to 80% of the female population.

As a result urban planning has an opportunity to reduce the separation between locations of employment and residence. There is demand for transportation services that offer affordable rides below SAR 22.5, or SAR 450 monthly. Ridesharing may offer a real solution for mobility across all sectors.

Figure 29
Saudi Females’ Accessibility to Job Sectors by Different Services

Credit: by Waishan QIU

M.C.P Thesis | Waishan QIU
Figure 30

Saudi Females' Accessibility to Job Sectors by Different Services

Credit: by Waishan QIU
4.3 Potential Market for Formal Ridesharing in Riyadh

This session concludes the quantitative analysis results of the potential Saudi female population that can get benefits from ridesharing services. Those population get benefits in a lower commute cost with a reasonable delay time due to shared trips. Education sector gets the largest benefits regarding the coverage of population while accessibility to manufacture sector is most cost-effective regarding the before and after comparison.

4.3.1 Potential Market and the Increased Accessibility

The size of the female population with access to employment is wide-ranging within each sector and each transportation options. The current analysis of accessibility measured by taxi, Careem, UberX and Kasper in Figure 31 illustrates that, manufacturing sector is accessible to the least proportion of Saudi female population (1.70%) in Riyadh. Healthcare sector’s accessibility is the highest with 65.14% of the Saudi women. Meanwhile, carpooling (measured by Kasper’s cost) provide affordability to the most population across all sectors. This analysis offers multiple opportunities for addressing access to employment. First of all, there is demand for transportation services that offer affordable rides below SAR 650 - 1000 monthly, or under the cost of SAR 25.00 per trips across the employment sectors. Secondly, ridesharing may offer a real solution for mobility across all sectors, as allows for a significant reduction in cost illustrated in Figure 28 thus there population get benefits from ridesharing is large. Take education sector as an example, it can be concluded that by ridesharing, the distance traveled within same budget is longer than by
taxi and online hailing systems. It improves more than 55% percent of the Saudi female population across Riyadh with more affordable commute option.

Figure 31

Potential Market of Ridesharing among Saudi Female Population
Credit: by Waishan QIU

Table 10 establishes the quantitative analysis of the Saudi female population in spatial analysis that can get benefits from formal ridesharing service. In Education sector, more than 1.8 million female gain the affordable access to the nearest female job market, which is 56% growth in percentage. Similarly, Healthcare sector and Retail sector see a 1 million and 1.7 million increase in the female population who are able to commute, with 33% and 52% percentage growth respectively.
Considering the females contribute to 60% students in Riyadh, it is reasonable to suppose that 30% of the female population would get a job. And within the commuters, 50% are willing to take ridesharing\(^{15}\) (see Table 5), then the potential market for the emerging ridesharing service (including online hailing system) in Riyadh is as large as 0.3 million. This is still a very conservative estimation, given the fact that once ridesharing is available, people are going to be relied increasingly on it (Scott et al., 2003).

Table 10

Potential Market of Ridesharing among Saudi Female Population

\(^{15}\) See Table 5- Choice of travel options should current mode becomes unavailable.

The estimation of 50% of the female commuters are willing to use ridesharing is based on the facts showed in Table 5. It counts for the people who are willing to choose taxi, informal car share, formal car share, contracted driver and chauffeur as the target population group. Those people are positively open to ridesharing service and other online hailing programs.
4.3.2 The Empirical Evidence from Taxi Trip Patterns

The analysis on Riyadh’s actual taxi data in 2016 supports some of the findings concluded from commute cost simulation and spatial analysis in the previous chapter.

(1) More than 65% passengers are female

Female passengers contribute to at least 65% of the total taxi trip records including those whose gender cannot be identified. Within the segment of identified and valid taxi trips records, female passengers contributes to 82.25% of the travels. Although the taxi trips are only a very small portion of the commute demand in Riyadh, considering the average taxi trips are only around 200 per day, the empirical statistic still support the fact that female relied more on taxi as they are not allowed to drive a car.

Figure 32
Taxi trips by gender in Riyadh in 2016
Source: Civic Data Design Lab

Table 12
Taxi trips by weekdays in Riyadh in 2016
Source: Civic Data Design Lab

<table>
<thead>
<tr>
<th></th>
<th>Amount</th>
<th>Female</th>
<th>Male</th>
<th>Unknown</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun</td>
<td>6088</td>
<td>1333</td>
<td>2023</td>
<td>9444</td>
<td></td>
</tr>
<tr>
<td>Mon</td>
<td>6283</td>
<td>1359</td>
<td>2064</td>
<td>9706</td>
<td></td>
</tr>
<tr>
<td>Tue</td>
<td>6267</td>
<td>1338</td>
<td>1989</td>
<td>9594</td>
<td></td>
</tr>
<tr>
<td>Wed</td>
<td>6702</td>
<td>1416</td>
<td>2049</td>
<td>10167</td>
<td></td>
</tr>
<tr>
<td>Thu</td>
<td>7464</td>
<td>1673</td>
<td>2361</td>
<td>11498</td>
<td></td>
</tr>
<tr>
<td>Fri</td>
<td>6304</td>
<td>1322</td>
<td>2060</td>
<td>9686</td>
<td></td>
</tr>
<tr>
<td>Sat</td>
<td>6043</td>
<td>1308</td>
<td>2038</td>
<td>9389</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>45151</td>
<td>9749</td>
<td>14584</td>
<td>69484</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage of Trips by Females</th>
<th>Valid Trips</th>
<th>All trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>82.25%</td>
<td>64.98%</td>
<td></td>
</tr>
</tbody>
</table>
(2) Females have Higher Travel Expense

Female passengers contribute to at least 65% of the total taxi trip records including those whose gender cannot be identified. Within the segment of identified and valid taxi trips records, female passengers contributes to 82.25% of the travels. Although the taxi trips are only a very small portion of the commute demand in Riyadh, considering the average taxi trips are only around 200 per day, the empirical statistic still support the fact that female relied more on taxi as they are not allowed to drive a car. The average cost per trip (53 SAR) is far beyond the affordability buffer we estimate for different income level by employment sectors (20% of monthly income by sector, less than 24 SAR per trip).

The bar chart shows that there is a distinct commute cost discrepancy by gender. Females in Riyadh in average have higher commute cost, in specific they pay about 5 more SAR, travel 0.5 more KM, and commute 3 more minutes per trip than males.
(3) Females Do Commute by Taxi

The series of charts in Figure 34 supports the hypothesis that females in Riyadh do commute by taxi service, even though the cost is rather expensive. Firstly, the line chart of the monthly taxi trip (Chart 1) shows a dramatic decrease of taxi usage in female since May 2016 while the trips by male even slowly increases. Though there is no direct evidence, scholars from Riyadh explain the phenomenon as the shift of commute behavior to UBER and other online hailing systems as those companies pushes there promotion in Riyadh during that period. It is reasonable that female commuters shift to online hailing service since their purpose of trips are commuting. The cost is a crucial factor when commuter make choices. It also explains that, while the most male passengers take taxi for non-commute purposes, the activity in male segment has not been affected.

Chart 2 shows the preference of car types. The majority trips are within economy car types (Economy and Sedan), which demonstrate that the purpose of trips are not for business meeting, family traveling or others. Daily traveling goals including commute is the main constitution of taxi trips.

Chart 3 and Chart 4 support the hypothesis with detail travel patterns. In Chart 3, there is distinct peaks in 4 PM and 9 PM. Although it does not show a morning traffic peak of commute, it does support the evening commute peak around 4PM when female go back home from employments. The 9 PM is a dinner time in Riyadh, and the corresponding peak shows that people do take taxi from working space to home or to entertainment and
restaurants. Chart 4 shows that people take taxi more frequently in the end of working days weekly (Friday and Saturday are weekends).

Figure 34

Taxi Trips Pattern in 2016

Credit: Visualized by Waishan QIU
(4) Safety Consideration

Figure 35 shows the difference commute patterns between weekdays and weekends. In general the patterns are similar, however people do take taxis more often in the evening in weekends. It is worth mentioning that, the size of taxi trips by female at night and in the midnight (10:00 PM to 2:00 AM) is still considerable compared to its peak, which supports that online hailing applications do not threaten, and even increase the safety for female to travel and commute in Saudi Arabia.

Figure 35

Comparison between Weekday and Weekend Trip Patterns

Credit: Visualized by Waishan QIU
(5) The Partial Spatio-temporal Patterns

The spatial investigation of commute prices illustrates that people live in the periphery of Riyadh have higher commute cost in average than residents in the city center (Figure 36). This supports our previous simulation that with the distance to city center increase, the financial burden of commute cost increases. Most of taxi trips origin from periphery regions in the southwest and northeast neighborhoods ranges from 60 SAR to 150 SAR per trip in average, which is far beyond the affordability buffer we defined (24 SAR per trip).

Figure 36: Average Cost of Taxi Trips by Pick-up Location

Credit: Visualized by Waishan QIU
4.3.3 The Feasibility and Benefits of Ridesharing in Riyadh

Based on the previous investigation of commute cost estimation, this study figures out that ridesharing service has the potential to largely reduce Saudi female’s commute cost and benefits the retail and manufacture industry the most. However, whether ridesharing service is feasible is still a question, or in another words, whether the service providers are efficient and sustainable to operate car-pooling in Riyadh. According to precedent researches, the feasibility of ridesharing is highly depend on the (1) density of the city, (2) the trips amounts every day, and (3) the congestion situation on the roads (Santi, Paolo, et al, 2014). The research will also look at the feasibility of developing ridesharing programs based on current analysis of taxi trips.

1 Feasibility of Developing Formal Ridesharing Program

The investigation of share-ability\(^\text{16}\) establishes that, 29% of the current taxi trips in Riyadh by female can be shared. The sample trip data selected all happened during morning initiated by female passengers. And the 29% share-ability is just within a sample size of 3000 trips aggregated in 2016. The delay time window is set at the level of within 10 minutes waiting time.

\(^{16}\) The method of calculate share-ability can be seen in Chapter 3. Specific codes developed in Python can be also found in Appendix I- Python Codes.
The interpretation of the cursory calculation of share-ability establishes that, if Saudi female commuters are willing to wait 10 minutes to take affordable trips by carpooling, actually in the morning time based on the yearly commute patterns, around one third of the trips can be shared. The benefits in total distance saved is also illustrated in Figure 37. To justify this finding, interviews on different demography segments of Saudi females are necessary to measure their realistic utility on time and money. Meanwhile, their attitude on carpooling with strange female passengers should be explored more explicitly.

Figure 37
Cursory Investigation on the Share-ability of Saudi Females’ Taxi Trips in Riyadh
Credit: by Waishan QIU; Data Source: Civic Data Design Lab, MIT
Figure 38

Shared Female Taxi Trips in the Morning Commute Time

This figure visualizes Original Trips (L) and the Shared Trips (R)

Credit: by Waishan QIU; Data Source: Civic Data Design Lab, MIT

Figure 38 visualize the shared trips in an aggregated day by the yearly data. The trips pattern indicates that while most trips happens within the ring road, there are demand that origin and terminate in the periphery area such as the airport and neighborhoods such as Twaeeq, Al Azizia and the industrial zone. It can also be concluded that, the trips whose shortest paths are overlapped in the major and secondary roads around the city center and main universities are most likely to be paired.
2 Benefits of Reducing Traffic Congestion and CO$_2$ Emission

In Figure 37, at the sample size of 3000 trips, 3300 KM distance in total is reduced as the result of pairing 426 trips within 5 minutes delay. Regarding the benefits of CO$_2$ reduction, specific information on the congestion of the road network is required. Due to the limitation of data source, further research looking at the benefits in reduction of traffic on the road and greenhouse gas emission is not available in this thesis.

3 Supplemental Connection to Metro System

Although the ridership shift after metro system’s completion is yet to discuss, the current taxi trip pattern indicates that ridesharing has the potential to solve the last kilometer problem in Riyadh for the new metro system. Consider the yearly high temperature\(^\text{17}\), any metro or bus station that is beyond 5-minutes walking distances is actually difficult to access by pedestrian. Given that there are still around 50% of the population that are outside the walking distances buffer\(^\text{18}\), there are quite a large number of population across gender facing the issue of last mile commute. As a result, there is a demand for affordable and comfortable commute options such as ridesharing that connecting passengers to metro and bus stations. By overlaying taxi trips and the metro and bus network,

\(^{17}\) See Table 3 on page 34 in section 2.3. The average high temperature from March to November is more than 27°C.

\(^{18}\) See Table 10. The projected metro plan in 2019 would serve around 50% of the population in Riyadh with a 5-minutes walking distance to the stations. It means that there are the other 50% commuters need other feasible solutions.
neighborhoods in the northeast and southwest do see a distinct demand that is beyond the reach of the planned metro system in 2019. This finding also supports the exploration on the feasibility of ridesharing service even when the ridership shift is not clear.

Figure 39

Saudi Females’ Accessibility to Employment by Different Service

Credit: by Waishan QIU; Data Source: Civic Data Design Lab, MIT

Overlay 2019 Metro Plan with the Taxi Trips Data
4.4 Accessibility and Employment Outcomes

This chapter compares the two different regression models using centrality or accessibility as the independent variable while keep other variables and the dependent variable of female employment rate remained.

4.4.1 Centrality and Accessibility Are Significant to Employment Rate

Linear Regression Model [1]

\[
\text{SaudiFemaleEmploymentRate} = \beta_0 + \beta_1 \times \text{CarOwnership} + \beta_2 \times \text{HomeOwnership} + \beta_3 \times \text{RightAgePopulation} + \beta_4 \times \text{DistanceToCityCenter}
\]

Linear Regression Model [2]

\[
\text{SaudiFemaleEmploymentRate} = \beta_0 + \beta_1 \times \text{CarOwnership} + \beta_2 \times \text{HomeOwnership} + \beta_3 \times \text{RightAgePopulation} + \beta_4 \times \text{Accessibility}
\]

Figure 40 and Figure 41 are the line fit plots for two regression models. While \(X_1\) (Car Ownership), \(X_2\) (Home Ownership), \(X_3\) (Right Age Population) and \(Y\) (Employment Rate) remains the same, model [2] differs from mode [1] with \(X_4\) (Distance to City Center vs Accessibility measured by commute cost cutoff). The purpose is not only to examine the correlation between accessibility measured by cost and economic opportunities outcomes, but also to compare the influence of centrality with the effect of commute cost.
Both results of testing assumptions for linear regression are showed in Table 13 and Table 14. All independent variables in model [1] and [2] are significant, and the adjusted R square for Equation [1] is 0.5 while the R square for Equation [2] is 1 percent higher, at the level of 0.51.

Figure 40
X Variables Line Fit Plots for Regression Equation [1] (Centrality)

Credit: by Waishan QIU
Figure 41

X Variables Line Fit Plots for Regression Equation [2] (Accessibility)

Credit: by Waishan QIU

[Diagrams of line fit plots for Car Ownership (X1), Home Ownership (X2), Right Age POP (X3), and Accessibility (X4)]
Table 13

Summary Output of Regression Analysis

Saudi Females’ Employment Outcomes and Indicators of Centrality

Credit: by Waishan QIU

<table>
<thead>
<tr>
<th>Regression Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Sample Size</td>
</tr>
</tbody>
</table>

Deviation Analysis

<table>
<thead>
<tr>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression Analysis</td>
<td>4</td>
<td>6.049</td>
<td>1.512</td>
<td>195.668</td>
</tr>
<tr>
<td>Residual</td>
<td>760</td>
<td>5.873</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>764</td>
<td>11.922</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>matrix</th>
<th>Coefficients</th>
<th>Standard Deviation</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.045</td>
<td>0.042</td>
<td>-1.066</td>
<td>0.287</td>
<td>-0.128</td>
<td>0.038</td>
</tr>
<tr>
<td>Car Ownership (per household)</td>
<td>0.133</td>
<td>0.008</td>
<td>16.768</td>
<td>0.000</td>
<td>0.117</td>
<td>0.148</td>
</tr>
<tr>
<td>Home Ownership (% own their houses)</td>
<td>-0.148</td>
<td>0.015</td>
<td>-9.645</td>
<td>0.000</td>
<td>-0.178</td>
<td>-0.118</td>
</tr>
<tr>
<td>Right Age Population (% in the age range 20-50)</td>
<td>0.492</td>
<td>0.044</td>
<td>11.225</td>
<td>0.000</td>
<td>0.406</td>
<td>0.578</td>
</tr>
<tr>
<td>Distance to City Center (KM)</td>
<td>-0.002</td>
<td>0.000</td>
<td>-4.292</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.001</td>
</tr>
</tbody>
</table>

As a result, the regression Model of Saudi female’s employments and centrality can be written in the format as equation [1] displayed below:

\[ R_{job} = -0.045 + 0.133 \times \text{Car}_\text{avg} - 0.148 \times \text{Home}_\text{avg} + 0.492 \times \text{I}_\text{age} - 0.002 \times \text{Dist}_\text{center} \]  [1]
Table 14

Summary Output of Regression Analysis

Saudi Females’ Employment Outcomes and Indicators of Accessibility

Credit: by Waishan QIU

<table>
<thead>
<tr>
<th>Regression Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Sample Size</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>Regression Analysis</td>
</tr>
<tr>
<td>Residual</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>matrix</th>
<th>Coefficients</th>
<th>Standard Deviation</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.104</td>
<td>0.035</td>
<td>-2.968</td>
<td>0.003</td>
<td>-0.172</td>
<td>-0.035</td>
</tr>
<tr>
<td>Car Ownership (per household)</td>
<td>0.134</td>
<td>0.008</td>
<td>17.815</td>
<td>0.000</td>
<td>0.119</td>
<td>0.149</td>
</tr>
<tr>
<td>Home Ownership (% own their houses)</td>
<td>-0.148</td>
<td>0.015</td>
<td>-9.740</td>
<td>0.000</td>
<td>-0.178</td>
<td>-0.118</td>
</tr>
<tr>
<td>Right Age Population (% in the age range 20-50)</td>
<td>0.455</td>
<td>0.046</td>
<td>9.978</td>
<td>0.000</td>
<td>0.365</td>
<td>0.544</td>
</tr>
<tr>
<td>Accessibility (measured at the cutoff of 20% monthly income as the commute cost, the total amount of jobs can be reached)</td>
<td>0.001</td>
<td>0.000</td>
<td>4.978</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>

As a result, the regression Model of Saudi female’s employments and accessibility can be written in the format as equation [2] displayed below:

\[ R_{job} = -0.104 + 0.134 \times Car_{avg} - 0.148 \times Home_{avg} + 0.455 \times I_{age} + 0.001 \times A_{20\%count} \quad [2] \]
4.4.2 Implication on Saudi Females’ Employment Rate

Concluded from Table 10 and 11, two regression models are recalled in following format:

\[
R_{job} = -0.045 + 0.133 \times Car_{avg} - 0.148 \times Home_{avg} + 0.492 \times l_{age} - 0.002 \times D_{ist\_center} \tag{1}
\]

\[
R_{job} = -0.104 + 0.134 \times Car_{avg} - 0.148 \times Home_{avg} + 0.455 \times l_{age} + 0.001 \times A_{20\%count} \tag{2}
\]

Since the accessibility formula is already estimated, the study further discuss the predicted discrepancy in employment rate between ridesharing service and taxi.

\[
\text{Accessibility by ridesharing} = -0.0018x^2 - 1.735x + 78.384
\tag{3}
\]

\[
R^2 = 0.82487, \ x: \text{distance to city center}
\]

\[
\text{Accessibility by taxi} = 0.009x^2 - 0.597x + 9.239
\tag{4}
\]

\[
R^2 = 0.39668, \ x: \text{distance to city center}
\]

The discrepancy of accessibility measured by ridesharing and taxi cost equals to [3]-[4]:

\[
\text{Accessibility} \ [\text{ridesharing-taxi}] = 69.145 - 0.0108x^2 - 1.138x
\tag{5}
\]

Based on equation [5] and equation [2], the predicted job participation in the circumstance of ridesharing can be drew. The result is illustrated in Figure 41 on the right.
Some basic conclusions can be drawn from the regression result:

1. While the correlations in both regression models are significant, accessibility in model [2] explains 51.1% of the sample data, which performs better than centrality in model [1] with the R square of 50.7%.

2. For the Car Ownership variable, it is significant in both models, and the coefficient is around 0.13. This suggests that while other variables remain the same, with a unit increase in Car Ownership (in average have one more car by household), the female employment outcome is possible to increase by 13.3%.

3. For the Home Ownership, it is also significant in both models, and the coefficient is around -0.148. This suggests that while other variables remain the same, if the neighborhood has 1% more families own their houses rather than by rental, the female employment rate is going to decrease by 0.148%.
4. For Right Age Population variable, the percentage of females within age range of 20-50 is significantly related to the female employment rate. The regression coefficients in model [1] and [2] are slightly different.

5. Equation [1] shows that holding other variables unchanged, one more percentage of population in the neighborhood within the age of 20 to 50 is associated with 0.492% more job participation in female population.

6. Similarly, Equation [2] shows that holding other variables unchanged, one more percentage of population in the neighborhood within the age of 20 to 50 is associated with 0.455% more job participation in female population.

7. Distance to City Center is significant in model [1]. Holding other variables still, as the distance to city center increase by 1 kilometer, the job participation for female would drop by 0.2%.

8. Accessibility measured by cumulative opportunities within a 20% income cutoff is also significant in mode [2]. While other variables keep the same, Saudi female’s job participation would increase by 1% if there are 10 more potential jobs can be reached.

9. In another words, with lower commute cost options, the female employment rate would see a positive increase at a reasonable level. For example, in Figure 29, it indicates that in the same neighborhood, say 20 KM away from city center, accessibility measured by ridesharing’s cost is 50, while the taxi’s is 5. In this prerequisite, the job participation outcome predicted whose commute cost is
equivalent to ridesharing will be around 4.5% higher than the one being equivalent to the cost by taxi (see Figure 32).

10. From Figure 42, it can be demonstrated that, ridesharing service has the best positive influence on increasing employment rate within 20 KM distance to the city center. Beyond 20 KM distances, the increase in job participation is not clear, which in return indicates that the commute cost reduction is not cost effective by ridesharing if the distance to city center is more than 20 KM.

In general, the regression result quantifies the influence of each variable on the outcome of Saudi females’ job participation rate, differentiating them by distance to city center or by accessibility measures. This provides the operationalized method to quantify the correlation between commute cost and employments. It sheds light on the urban planning and transportation policies in the establishment of future job market growth.
Chapter 5. Conclusion

The research findings from this study establish that in order to improve Saudi females’ economic participation outcomes, it is crucial to offer affordable transportation options that reduce the commute cost. First of all, this paper measures the realistic cost in terms of time and money by car-sharing as a mode compared to other available transit options for Saudi female residents. It then asserts that carpooling service has a large demand based on the spatial analysis of female passengers’ travel patterns. It also initiates a brief network analysis on the feasibility of ridesharing and quantifies the benefits of traffic reduction. Due to the limitation of data source and time scope available, there is further need to investigate the sensitivity between cost and time for different population segments in Saudi female population.

5.1 Research Results

The results from this study establish that reducing Saudi females’ commute cost is crucial to improve their economic participation outcomes. In general, the regression model quantifies the influence of each variable (car ownership, home ownership, right age population, and the distance to city center or accessibility) on the outcomes of Saudi females’ job participation rate. It also differentiates them by distance to city center (centrality) or by accessibility measures. This provides the operationalized method to quantify the correlation between commute cost and employments. Accessibility measured
by cumulative opportunities within a 20% income cutoff is significant and explains more sample than centrality model. In short, while other variables keep the same, the employment rate of Saudi female residents is 1% higher in the neighborhood where people can reach 10 more jobs within same budget than others. In another words, with cheaper commute options, Saudi females would have larger access to job opportunities and the employment participation would see a positive increase at a significant level. To maintain commute cost within 20% of monthly income by sectors requires affordable and convenient services other than metro systems that at a cost of estimated SAR 22.00 per trip. When there is no public transportation, E-hailing services are the most affordable option compared to street hailing taxi services. Meanwhile the preliminarily existence of informal ridesharing service in Saudi Arabia culture indicates the possibility of providing formal programs. As a result there is a large demand market where at least 0.3 million female passengers are potential to shift to women-only carpooling service in Riyadh (at a very conservative estimation).

The quantitative model also establishes that, ridesharing service has the most positive influence in reducing commute cost and increasing accessibility to the job market for Saudi female residents who live within a 20 kilometers distance from the city center. For those who live outside the radius of 20 kilometers from city center, the effect of reducing commute cost and increasing accessibility is not cost-effective. The result indicates an reasonable scope for urban planning, transportation policy and carpooling enterprises to provide affordable services that is of high demand and is highly cost-effective.
The interpretation of regression model also sheds light on comparing different policies in response to the national goal of improving Saudi females’ job participation. The same mechanics may also apply to the Car Ownership factor. By increasing the car ownership, the employment outcome of Saudi females is also highly possible to increase. However, it is easy to foresee that carpooling service is more cost-effective and have larger positive externalities than simply reducing vehicle prices or encouraging more private vehicles on the road. It is because encourage ridesharing programs has the benefits of reducing traffic congestion and CO₂ emission.

The research finding of the spatial mismatch between female population’s concentration and their job opportunities sheds light on urban planning and transportation policies. The result implicates geospatial strategies in the establishment of future job market growth, investments of future transportation infrastructure, and even subsidy to target population segments. For example, by encouraging formal ridesharing enterprises with subsidy or financial aids, Saudi female residents who live in the periphery would benefit from reduction in commute cost.

Since females have greater accessibility to the job market as a result of encouraging ridesharing programs, it is highly possible to see an increased employment rate in female population segment. Additionally, the finding also indicates specific neighborhoods or specific industry sectors that has large demand for ridesharing. Those areas and sectors in
Riyadh can become the pilot places for carpooling enterprises and the transportation department to experiment with their programs or policies. For instance, female employment can be improved the most within sectors that are geographically distributed along the high way (Route 65 & 40) and around the dense city center. Regarding the four sectors investigated in this study, manufacture market is the least accessible one in terms of its geolocation. Meanwhile its salary range does not allow recruitment of women in distant neighborhoods. As a result, it is most cost-effective to provide service to the manufacture sector. Oppositely, healthcare and education sectors are most accessible. It is because of the distribution of the hospitals and schools, as well as employees in these sectors have higher salaries. As a result, ridesharing can provide the largest amount of growth with the affordable access to these sectors.

Lastly, empirical taxi trip data analysis supports that, current financial burden of commute cost across different sectors for Saudi female is higher than males. And the commute time and distance is also longer for female residents in the periphery than those around city center. For those who live in periphery, in average they may have to spend 40% or more of their base salary on commute on taxi. Based on current trip patterns, ridesharing programs is capable to allow for much lower commute cost with a reasonable delay time within 5-10 minutes per trip.
5.2 Research Impacts

This research will shed light on innovative and affordable alternatives that have the potential to reduce Saudi female’s commute cost and promote their job participation rate in Riyadh in Saudi Arabia. First of all, it provides empirical evidence on the commute patterns for women, relating to their commute costs, transit modes, and their spatial distribution of origins and destinations. It also establishes the quantitative relationship between commute cost, centrality to city center and employment outcomes in the segment of Saudi female population for urban planners, public transportation researches and policy makers in Saudi Arabia. It also has implications for Riyadh’s city planning to re-organize employment development and residential areas in order to reduce the separation between the concentration of economic opportunities and the Saudi female residence. Additionally, it demonstrates the feasibility of providing formal ridesharing services for the transportation companies in private sectors in Riyadh and even other Saudi Arabia cities.

Meanwhile, this research aims to contribute to the area of urban information system and urban study methodology. Through the urban computing method that measures the realistic commute time and distance, calculate the accessibility and the feasibility of ridesharing, this research also set up a framework for further urban studies deploying API query, urban computing and network analysis tools.
5.3 Limitations and Future Plan

Section 5.3 discusses the major limitations of this research from the data and methodology aspect. As for the limitation of the models, a major one lies in the selection of different variables. Current selection of variables is limited by the data available. Some other important factors such as education level, income level, sensitivity on the utility of time and money, people’s attitude towards ridesharing and other relevant variables are left out in the regression model. Also, the data could be refined by narrowing the scale of grid matrix the study sets up. For example, by decrease the grid from 2KM to 1KM, higher resolution of commute cost estimation, distance to city center, dissolved CENSUS data such as population density can be achieved. With a more detailed dataset, the relationship between dependent and independent variables can be further revealed, thus providing more resolution to measure the relationship between commute cost and employment outcomes for Saudi female.

To justify the legitimation of accessibility measurement in this study, more delicate model of accessibility measures should be investigated. For instance, it is helpful to measure the realistic utility of time and money for various income segments of Saudi female in Riyadh. With the acquisition of sensitivity of time and money, the accessibility measure would be more comprehensive in space and time dimension. To support the investigation of ridesharing’s feasibility and benefits in Riyadh, not only taxi trip data but other current traffic flows information is needed. With more comprehensive O-D flow data, the share-ability measurement would be more robust and convincing. Meanwhile, knowledge on the
current and potential shifts in the ridership after formal ridesharing programs as well as the metro system’s completion is dispensable to estimate the potential market size for ridesharing. The ridership model will also influence the calculation of share-ability, since in the current research there is only a rough estimation on ridesharing’s portion based on investigation whose sample size is small and the site is no in Riyadh. None of the future studies have been strictly testified, but proposing these discussions as the extension of a more solid and comprehensive methodology at its preliminary stage is a good start.

Regarding the benefits of CO2 reduction, specific information on the congestion of the road network is required. Due to the limitation of data source, further research looking at the benefits in reduction of traffic on the road and greenhouse gas emission is not yet available in this thesis.

Lastly, the research is also restricted by the limited knowledge on the specific culture, religious, economic and income level of Saudi females and their families. The culture and social value of employment for Saudi females is not discussed in-depth in this research due to the research scope. In the future study, it is of great importance to investigate the acceptance, attitudes, and motivation of a woman being working or not being working for different income segments for families in Riyadh.


Appendix I

Python Scripts for API Query, Network Analysis, and Data Visualization

This sector provides all the python scripts developed by Waishan. Some codes are adjusted from open sources such as GitHub. For the algorithm of measuring share-ability, the method is inspired a lot by the SENSEable City Lab at MIT\(^\text{19}\).

Python code for visualizing the taxi trips

```python
### 0. Import libraries
from dateutil import tz
from datetime import datetime, date
import timestring
import time
import os
from collections import Counter
import networkx as nx
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import re
# This allows plots to appear on the IPython notebook.
%matplotlib inline
import matplotlib.cm as cm

### 1. Re-read the Trips Dictionary by json library
Trips = json.load(open("Trips.txt"), "utf-8")

### 1. Construct the Riyadh Road Network graphic RG
## 1. 1 Define an empty undirected graph "RG" to store the road network data
RG = nx.Graph()
## 1. 2 Read dataframe "route_edges_am"
edge_df = pd.read_csv("RiyadhRoad/data/riyadh_route_edges_am.txt", sep=" ", index_col=0)
# Define the values dictionary for the "edge attributes" keys = ['volume', 'capacity', 'voc', 'degree', 'free_travel_time', 'travel_time']
for row in edge_df.itertuples(index=False):
    values = row[2:]
    # We create a dictionary "edge_attributes" with the keys and row values
    edge_attributes = dict(zip(keys, values))
    my_tuple = (row[0], row[1])
    # We add the edge to the graph use "add_edge" function
    RG.add_edge(*my_tuple, attr_dict=edge_attributes)
## 1.3 Read and Add coordinates to nodes of graphic RG
nodes_df = pd.read_csv("RiyadhRoad/data/riyadh_nodes.txt", sep=" ", index_col=0)
nodes_df.head(2)
for node_id in RG.nodes():
    RG.node[node_id]["coordinate"] = (nodes_df["st_x"])[node_id],
                                          nodes_df["st_y"])[node_id])
loc_dict = {}
for node_index in RG.nodes():
    #print node_index
    loc_dict[node_index] = RG.node[node_index]["coordinate"]
## 1.4 Check if the network is connected?
```
print "** Is the road network connected? " + str(nx.is_connected(RG))
print '***',RG.node[1]['coordinate']

taxi_RG=nx.Graph()
edge_count = []
for trip_id in Trips.keys():
    attr= Trips[trip_id]
    route=attr['path']
    #print attr
    for i in range(len(route)):
        if i < len(route)-1:
            o=route[i]
            d=route[i+1]
            taxi_RG.add_edge(*(o,d), attr_dict=attr)
            if o<d:
                edge_count.append((o,d))
            elif o>d:
                edge_count.append((d,o))
loc_taxi={} 
for node_index in taxi_RG.nodes():
    loc_taxi[node_index] = RG.node[node_index]['coordinate']

Count = Counter(edge_count)
Keys = Count.keys()
# print Keys
for key in Keys:
    #print key[0], key[1], key
    taxi_RG[key[0]][key[1]]['Count'] = Count[key]
print taxi_RG[key[0]][key[1]]
edges = taxi_RG.edges()
edge_widths = [(taxi_RG[u][v]['Count']*1) for u,v in edges]
edge_colors = plt.cm.coolwarm(edge_widths)
plt.figure(figsize=(40,40))
nx.draw(RG, pos=loc_dict, node_size=0, width=0.5, node_color='#bccad6',
edge_color = '#bccad6');
nx.draw(taxi_RG, pos=loc_taxi, node_size=0, width=edge_widths,
node_color='#bccad6', edge_color = edge_colors);
Python codes for Cleaning Trips Data

## 0. Import libraries

```python
from dateutil import tz
from datetime import datetime, date
import timestring
import time
import os
from collections import Counter
import networkx as nx
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import re
```

# This allows plots to appear on the IPython notebook.
%matplotlib inline

## 1. Read all hour files and save the trips amount in a dictionary

```python
folder = 'CleanCsv/byHour/'
files = os.listdir(folder)
csv_files = []
for f in files:
    if f.endswith('.csv'):
        csv_files.append(f)
csv_files.sort()

# print csv_files

dict_L={}
for f in csv_files:
    path = folder + f
    df = pd.read_csv(path, encoding = 'utf-8', sep='",', engine='python')
    dict_L[f] = len(df)

print 'Trips Amount of Hour Periods:', dict_L
```

## 2. Read the most relevant trip hours within Riyadh Boundary Scope

```python
df = pd.read_csv('CleanCsv/2016.csv', encoding = 'utf-8', sep='",', engine='python')
df = df.loc[(df['pkWeekday'] != 'Friday') & (df['pkWeekday'] != 'Saturday')
            & (df['Ox'] > 45.5) & (df['Oy'] > 24.5) & (df['Dx'] > 45.5) & (df['Dy'] > 24.5)]
df.drop(['deltaT_min', 'deltaT_min', 'pkHour', 'pkWeekday', 'pkTime_new'], axis=1, inplace=True)
df = df.reset_index(drop=True)

print 'Trips selected: ', len(df)
df.head(2)
```

```python
# df.groupby(['pkMonth', 'pkDay']).agg(['mean', 'count'])

counts = df.groupby(['pkMonth', 'pkDay']).size();
# counts_df = pd.DataFrame(counts).rename({'counts': 'counts'})
counts.head(10)
counts.to_csv("countMonthDayTrips.csv", index=False, encoding='utf-8')
```

```python
df = df.loc[(df['pkMonth'] == 2) & (df['pkDay'] == 1)]
```
df.drop(['pk_Month','pk_Day'], axis=1, inplace=True)

df = df.reset_index(drop=True)
df.head (3)

```python
## 3. Define a Delta Time Window, to calculate the ridersharing possibility

h = 24.000
m = 60.000
DELTA = 10.000
t_delay = round((DELTA/(h*m)),4)
print t_delay
```

```python
## 4. Define an empty undirected graph "RG_taxi" to save the trips data.

RG_taxi = nx.Graph()
RG_Ori = nx.Graph()

# We also define the values dictionary for the "edge attributes"
keys = ['Gender','Cost','Car_Ctgry','Dist_km','O_x', 'O_y', 'D_x',
'D_y', 'Trip_min', 'Time_inDay']
i = 0
loc_taxi = {}
loc_ori = {}

for row in df.itertuples(index=False):
    #print row[0]
i +=1
j = i-1

# to control the nodes amount i
if i < 3000:
    values = row[0:]

    # We create a dictionary with the keys and row values
    edge_attributes = dict(zip(keys, values))

    #print edge_attributes

    # Define the edge as my_tuple
    my_tuple = ('O_'+str(j), 'D_'+str(j))

    #print my_tuple

    # We add the edge to the graph
    RG_taxi.add_edge(*my_tuple, attr_dict=edge_attributes)

    RG_Ori.add_node('O_ '+str(j))

    loc_taxi['O_'+str(j)] = (df['O_x'][j], df['O_y'][j])

    loc_taxi['D_'+str(j)] = (df['D_x'][j], df['D_y'][j])

    loc_ori['O_'+str(j)] = (df['O_x'][j], df['O_y'][j])

#print loc_dict

## 5. Add location to the nodes of RG_taxi

for node_ID in RG_taxi.nodes():
    #print node_ID
    RG_taxi.node[node_ID]['coordinate'] = loc_taxi[node_ID]

    #print RG_taxi.node[node_ID]

    # How to call nodes, edges, and their attributes
    #print "*1. \"+str(RG.edges('O_1'))\" # print all edges with starting point "1"
    #print "*2. \"+str(RG.nodes()[:3])\" # print first 3 nodes
    #print "*3. \"+str(RG.edges(0:3))\" # print first 3 edges
    #print "*4. \"+str(RG['O_1']['D_1'])\" # print attributes of edge(1,2)
    #print "*5. \"+str(RG['O_1']['D_1']['Trip_min'])\"
    #print "*6. \"+str(RG['D_1']['O_1']['Trip_min'])\"
```
# print *7. "+str(RG.node[0.1]) # print attributes of node[1]
# print RG.node[0.1]["coordinate"]
# print RG.node[0.1]
# print 

## 1. Read dataframe "route.edges_am"
edge_df = pd.read_csv("RiyadhRoad/data/riyadh_route_edges_am.txt", sep=" ", index_col=0)
##df["travel_time"]

## 2. Define an empty undirected graph "RG" to store the road network data
RG = nx.Graph()
# We also define the values dictionary for the "edge attributes"
keys = ['volume', 'capacity', 'voc', 'degree', 'free_travel_time', 'travel_time']

for row in edge_df.itertuples(index=False):
    values = row[2:]
    # We create a dictionary "edge attributes" with the keys and row values
    edge_attributes = dict(zip(keys, values))
    my_tuple = (row[0], row[1])
    # We add the edge to the graph use "add_edge" function
    RG.add_edge(*my_tuple, attr_dict=edge_attributes)

## 3. Read and Add coordinates to nodes of graphic RG
nodes_df = pd.read_csv("RiyadhRoad/data/riyadh_nodes.txt", sep=" ", index_col=0)
nodes_df.head(2)
for node_id in RG.nodes():
    RG.node[node_id]["coordinate"] = (nodes_df['st_x'][node_id],
                                       nodes_df['st_y'][node_id])
loc_dict = {}
for node_index in RG.nodes():
    #print node_index
    loc_dict[node_index] = RG.node[node_index]["coordinate"]
print 'Done'

## 4. Analysis the degrees, and centrality of RG_taxi, and give node size, edge colors to it
node_sizes_taxi=[]
degrees_taxi = nx.degree(RG_taxi)
centralities_taxi=nx.degree_centrality(RG_taxi)
for i in RG_taxi.nodes():
    node_sizes_taxi.append(degrees_taxi[i])
node_colors_taxi = range(len(RG_taxi.nodes()))
edges_taxi = RG_taxi.edges()
weights_taxi = [(RG_taxi[u][v]["Cost"]*0.002) for u,v in edges_taxi]
edge_colors_taxi = plt.cm.coolwarm(np.log(np.log(weights_taxi)))

## 5. Analysis the degrees, and centrality of RG, and give node size, edge colors to it
degrees = nx.degree(RG)
centralities= nx.degree_centrality(RG)
# degrees, centralities are dictionary, we need to call their all values as new lists
node_centrality=[round((centralities[u]*10000),2) for u in centralities]
node_degree = [i for i in degrees.values()]
#colors = [G[u][v]['color'] for u,v in edges]
# weights = [G[u][v]('weight') for u, v in edges]
# nx.draw(G, pos, edges=edges, edge_color=colors, width=weights)
node_colors = range(len(RG.nodes()))  # plt.cm.coolwarm() is a function change
node_sizes = node_centrality
edges = RG.edges()
weights = [RG[u][v]['volume']*0.0005+0.5 for u,v in edges]
edge_colors = plt.cm.Greys(np.log(np.log(weights)) # remember to import
the colormap library first

color_gender = []
for u,v in edges_taxi:
    if RG_taxi[u][v]['Gender'] == 'female':
        clr = '#ff6f69'
    else:
        clr = '#80ced6'
    color_gender.append(clr)
pl.figure(figsize=(20,20))
nx.draw(RG, pos=loc_dict, node_size=0, width=1, node_color='#E5E7E9',
edge_color = '#E5E7E9');
nx.draw(RG_Ori, pos=loc_ori, node_size=10, width=1,
node_color='White');
nx.draw(RG_taxi, pos=loc_taxi, node_size=0, edge_color=color_gender,
width=1, node_color=node_colors_taxi);
#nx.draw_networkx_labels(Shortest_RG, pos=loc_dict); # semicolon suppresses output
Python codes for Measuring Accessibility

```
### 0. Import libraries
from dateutil import tz
from datetime import datetime, date
import timestring
import time
import os
from collections import Counter
import networkx as nx
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
import re
# This allows plots to appear on the IPython notebook.
%matplotlib inline
import matplotlib.cm as cm # import the colormap library
import operator
import math
import json
import networkx as nx
## 1. Load Grid Location Data
df_Grid = pd.read_csv('2KMGrid.csv', encoding='utf-8', engine='python')
df_Grid.head(2)
## 2. Load dataframe
```

MCP Thesis | Waishan Qiu
Appendix I | Python Codes
df_Dest = pd.read_csv('Destination_Done.csv', encoding='utf-8', engine='python')
df_Dest.head(1)

```python
### 3. Get Grid ID of Education, Healthcare, Retail, and Manufacture sectors
Edu=[]
Hlth=[]
Retl=[]
Manu=[]
for i in range(len(df_Dest)):
    if df_Dest['Category'][i] == 'Education':
        Edu.append(int(df_Dest.Grid_ID[i]))
    if df_Dest['Category'][i] == 'Healthcare':
        Hlth.append(int(df_Dest.Grid_ID[i]))
    if df_Dest['Category'][i] == 'Retail':
        Retl.append(int(df_Dest.Grid_ID[i]))
    if df_Dest['Category'][i] == 'Manufacturing':
        Manu.append(int(df_Dest.Grid_ID[i]))

print Edu
dict_Edu=Counter(Edu)
print dict_Edu
Edu = list(set(Edu))
print Edu, Hlth, Retl, Manu

### 4. Load Grid Fare Matrix and Calculate the Price Buffer from the Job Destinations
folder = 'Data/Fares/'
files = os.listdir(folder)
csv_files = []
for f in files:
    if f.endswith('.csv'):
        csv_files.append(f)
        csv_files.sort()

RG_Buffer = nx.Graph()
df_cate = pd.DataFrame()
for i in range(len(csv_files)):
    # Read each csv file
    f=csv_files[i]
    path = folder + f
    GridID= (f.split('_')[-1]).split('.')[0]
    df = pd.read_csv(path, encoding='utf-8', sep='", engine='python')
    Careem=[]
    Taxi=[]
    UberX=[]
    Kasper=[]
    Time=[]
    Dist=[]
    # Print 'File:', f, Path, ID,\n    # Loop within the each csv file
    for j in range(len(df)):
        if df['UNQID_E'][j] in Manu:
            S_ID = GridID
            E_ID = df['UNQID_E'][j]
            Careem.append(df['Careem'][j])
            Taxi.append(df['Taxi'][j])
```

MCP Thesis | Waishan Qiu

Appendix I | Python Codes
UberX.append(df['UberX'][j])
Kasper.append(df['Kasper'][j]*1.5)
Time.append(df['Duration_min'][j])
Dist.append(df['Distance_km'][j])
Cr = min(Careem),
Tx = min(Taxi),
Ub = min(UberX),
Kp = min(Kasper),
T = min(Time),
D = min(Dist)

# print Cr[0], Tx[0], Ub[0], Kp[0], T[0], D # Careem, # Kasper

df_cate.loc[i, 'Start_ID'] = GridID
df_cate.loc[i, 'minCareem'] = Cr[0]
df_cate.loc[i, 'minUberX'] = Ub[0]
df_cate.loc[i, 'minTaxi'] = Tx[0]
df_cate.loc[i, 'minKasper'] = Kp[0]
df_cate.loc[i, 'minDist'] = D

def_cate.loc[i, 'minTime'] = T[0]
node_attributes = {'ID': S_ID, 'Careem': Cr[0], 'Duration': T[0], 'Distance': D}

RG_Buffer.add_node(int(S_ID), attr_dict=node_attributes)

def_cate.head(2)

def_cate.to_csv('Data/ManufactureBuffer.csv', index=False, encoding='utf-8')

# 5. Load Grid Fare Matrix and Calculate the numbers of jobs within 30% income buffer from Neighborhoods
folder = 'Data/Fares/'
files = os.listdir(folder)
csv_files = []
for f in files:
    if f.endswith('.csv'):
        csv_files.append(f)
csv_files.sort()

    RG_Buffer = nx.Graph()
df_cate = pd.DataFrame()
SectorList = ['Edu', 'Hlth', 'Retl', 'Manu']
SectorName = ['Edu', 'Hlth', 'Retl', 'Manu']
SectorBuffer = [36.00, 29.25, 25.50, 27.00]
for sector_index in range(4):
    tmpList = SectorList[sector_index]
    tmp_Name = SectorName[sector_index]
    tmp_Buffer = SectorBuffer[sector_index]
    for i in range(len(csv_files)):
        # Read each csv file
        f = csv_files[i]
        path = folder + f
        GridID = int((f.split('_')[-1]).split('.')[0])
        df = pd.read_csv(path, encoding = 'utf-8', sep='", engine='python')
        Careem=[]
        Taxi=[]
        UberX=[]
        Kasper=[]
node_attributes = {'ID':GridID}
RG_Buffer.add_node(GridID, attr_dict=node_attributes)

# Loop within the each csv file
for j in range(len(df)):
    if df['UNQID_E'][j] in tmp_list:
        E_ID = df['UNQID_E'][j]

# Save the costs
Cr = df['Careem'][j]
Tx = df['Taxi'][j]
Ub = df['UberX'][j]
Kp = (df['Kasper'][j]/2)*3
if Cr < tmp_Buffer:
    Careem.append(df['Careem'][j])
if Tx < tmp_Buffer:
    Taxi.append(df['Taxi'][j])
if Ub < tmp_Buffer:
    UberX.append(df['UberX'][j])
if Kp < tmp_Buffer:
    Kasper.append(df['Kasper'][j])

# Save the costs
Cr = df['Careem'][j]
Tx = df['Taxi'][j]
Ub = df['UberX'][j]
Kp = (df['Kasper'][j]/2)*3
if Cr < tmp_Buffer:
    Careem.append(df['Careem'][j])
if Tx < tmp_Buffer:
    Taxi.append(df['Taxi'][j])
if Ub < tmp_Buffer:
    UberX.append(df['UberX'][j])
if Kp < tmp_Buffer:
    Kasper.append(df['Kasper'][j])

dfcate.loc[i, 'Start_ID' ] = GridID
dfcate.loc[i, '3_'+tmp_Name+'Tx' ] = len(Taxi)
dfcate.loc[i, '3_'+tmp_Name+'Cr' ] = len(Careem)
dfcate.loc[i, '3_'+tmp_Name+'Ub' ] = len(UberX)
dfcate.loc[i, '3_'+tmp_Name+'Kp' ] = len(Kasper)
dfcate.loc[i, '2_'+tmp_Name+'Tx' ] = sum(i < tmp_Buffer*2/3 for i in Taxi)
dfcate.loc[i, '2_'+tmp_Name+'Cr' ] = sum(i < tmp_Buffer*2/3 for i in Careem)
dfcate.loc[i, '2_'+tmp_Name+'Ub' ] = sum(i < tmp_Buffer*2/3 for i in UberX)
dfcate.loc[i, '2_'+tmp_Name+'Kp' ] = sum(i < tmp_Buffer*2/3 for i in Kasper)
dfcate.loc[i, '1_'+tmp_Name+'Tx' ] = sum(i < tmp_Buffer*1/3 for i in Taxi)
dfcate.loc[i, '1_'+tmp_Name+'Cr' ] = sum(i < tmp_Buffer*1/3 for i in Careem)
dfcate.loc[i, '1_'+tmp_Name+'Ub' ] = sum(i < tmp_Buffer*1/3 for i in UberX)
dfcate.loc[i, '1_'+tmp_Name+'Kp' ] = sum(i < tmp_Buffer*1/3 for i in Kasper)

RG_Buffer.node[GridID][tmp_Name+'Tx'] = len(Taxi)
RG_Buffer.node[GridID][tmp_Name+'Cr'] = len(Careem)
RG_Buffer.node[GridID][tmp_Name+'Ub'] = len(UberX)
RG_Buffer.node[GridID][tmp_Name+'Kp'] = len(Kasper)
dfcate.to_csv('Data/GridCount_AllSectors.csv', index=False, encoding='utf-8')
dfcate.head(2)

# print RG_Buffer.node[i]
loc_Buffer = {}
for node_index in RG_Buffer.nodes():
    # print node_index
    loc_Buffer[node_index] = (RG_Grid.node[node_index]['X'], RG_Grid.node[node_index]['Y'])
### 1. Construct the Riyadh Road Network graphic RG

#### 1.1 Define an empty undirected graph "RG" to store the road network data

```python
RG = nx.Graph()
```

#### 1.2 Read dataframe "route_edges_am"

```python
tedf =
    pd.read_csv("TaxiData&Script/RiyadhRoad/data/riyadh_route_edges_am.txt",
    sep=" ", index_col=0)
```

# Define the values dictionary for the "edge attributes"

```python
keys = ['volume', 'capacity', 'voc', 'degree', 'free_travel_time',
        'travel_time']
```

```python
for row in edgedf.itertuples(index=False):
    values = row[2:]
    edge_attributes = dict(zip(keys, values))
    my_tuple = (row[0], row[1])
    RG.add_edge(*my_tuple, attr_dict=edge_attributes)
```

#### 1.3 Read and Add coordinates to nodes of graphic RG

```python
nodes_df =
    pd.read_csv("TaxiData&Script/RiyadhRoad/data/riyadh_nodes.txt", sep=" ",
    index_col=0)
```

```python
for node_id in RG.nodes():
    RG.node[node_id]['coordinate'] = (nodes_df['st_x'][node_id],
                                        nodes_df['st_y'][node_id])
```

```python
loc_dict = {}
for nodeindex in RG.nodes():
    print(nodeindex)
    loc_dict[nodeindex] = RG.node[nodeindex]['coordinate']
```

#### 1.4 Check if the network is connected?

```python
print "** Is the road network connected? " + str(nx.is_connected(RG))
```

### Appendix I

#### Python Codes

MCP Thesis | Waishan Qiu

---

**Appendix I**

**Python Codes**

MCP Thesis | Waishan Qiu
clr_Dest.append('#ff6f69')

defin RG_Dest.node[node_index]["Category"] == 'Healthcare':
clr_Dest.append('#f9d5e5')
else:
clr_Dest.append('#ffcc5c')

nodes = RG_Buffer.nodes()
BufferSizes = [5*np.log(RG_Buffer.node[u]["Edu_Cr"]) for u in nodes]
BufferWeights = [5*np.log(RG_Buffer.node[u]["Edu_Cr"]) for u in nodes]
BufferColors = plt.cm.coolwarm(np.log(BufferWeights))
plt.figure(figsize=(15,15))
nx.draw(RG_Buffer, pos=loc_Buffer, nodesize=BufferWeights,
nodecolor=BufferColors, label='Careem Cost Buffer');

nx.draw(RG, pos=loc_dict, node_size=0, width=1, node_color='#E5E7E9',
edge_color = '#E5E7E9');
nx.draw(RG_Dest, pos=loc_Dest, node_size=20, node_color=clr_Dest, label = 'Destination');

plt.legend(loc='upper right')
Python codes for Calculating Commute Cost to the Employments

```python
# reverse_geocoding_result = gmaps.reverse_geocode((40.714224, -73.961452))
# Request directions via public transit
# now = datetime.now()
# directions_result = gmaps.directions("Sydney Town Hall",
# "Parramatta, NSW",
# mode="transit",
# departure_time=now)
'AIzaSyAlYv6FwEd96RSQcAlXg6YkrDwz2KhG7eY8' # My GeoLocation API
'AIzaSyAzJLNlujJrtytANgl7wko844MLQIG28aY' # My GeoCoding API
## 1. Import libraries
import pandas as pd
from collections import defaultdict
import json
import pandas as pd
import time
import os
import numpy as np
from collections import Counter
import googlemaps
import operator
import math
import gpxpy.geo
import networkx as nx
import matplotlib.pyplot as plt
# This allows plots to appear on the IPython notebook.
%matplotlib inline
## 2. Load dataframe and Calculate Longitude, Latitude
df = pd.read_csv('DestinationAddress.csv', encoding='utf-8', engine='python')
print(len(df))
print('Total Destination is ' + str(L) + '
'df.head(2)
## Google GeoCoding Api, convert location name to longtitude and latitude
gmaps = googlemaps.Client(key='AIzaSyAzJLNlujJrtytANgl7wko844MLQIG28aY')
for i in range(len(df)):
    name = df['Name'][i]
    # Geocoding an address
    geocode_result = gmaps.geocode(name+'Riyadh, Saudi Arabia')
    #print geocode_result
    if geocode_result != None:
        geocode_result[0]["geometry"]['location']
        y = geocode_result[0]["geometry"]['location']['lat']
        x = geocode_result[0]["geometry"]['location']['lng']
        #print name, x, y
        df.loc[i, 'Lng_Xl'] = x
        df.loc[i, 'Lat_Yl'] = y
df.to_csv('Destination_Address.csv', index=False, encoding='utf-8')
print "Done"
```
# Load Grid Matrix dataframe and Assign the closest Grid Number

df_Grid = pd.read_csv('2KM_Grid.csv', encoding='utf-8', engine='python')

print df_Grid.head(2)
dict_GID = {} 
for i in range(len(df)): 
x = df['Lng_X'][i] 
y = df['Lat_Y'][i] 
tmp_DIST = {} 
for j in range(len(df_Grid)): 
x1 = df_Grid['Longitude'][j] 
y1 = df_Grid['Latitude'][j] 
import gpxpy.geo
# Calculate the distance by meter 
tmp_DIST[j] = gpxpy.geo.haversine_distance(y, x, y1, x1) 
sorted_DIST = sorted(tmp_DIST.items(), key = operator.itemgetter(1)) 
# print sorted_GID, sorted_DIST 
clst_GID = sorted_DIST[0][0] 
shrtst = round(sorted_DIST[0][1], 2) 
dict_GID[i] = (clst_GID, shrtst) 
df.loc[i, 'Grid ID'] = clst_GID 
df.loc[i, 'Grid Dist'] = shrtst 
print dict_GID

df.to_csv('DestinationAddress.csv', index=False, encoding='utf-8')

## 2. Load dataframe and Assign the Closest Node Number

df = pd.read_csv('Destination Address.csv', encoding='utf-8', engine='python')

print len(df) 
print df.head(1)
dict_NID = {} 
for i in range(len(df)): 
x = df['Lng X'][i] 
y = df['Lat Y'][i] 
tmp_DIST = {} 
for node_id in RG.nodes(): 
x1 = RG.node[node_id]['coordinate'][0] 
y1 = RG.node[node_id]['coordinate'][1] 
import gpxpy.geo
# Calculate the distance by meter 
tmp_DIST[node_id] = gpxpy.geo.haversine_distance(y, x, y1, x1) 
sorted_DIST = sorted(tmp_DIST.items(), key = operator.itemgetter(1)) 
# print sorted_GID, sorted_DIST 
clst_NID = sorted_DIST[0][0] 
shrtst = round(sorted_DIST[0][1], 2) 
dict_NID[i] = (clst_NID, shrtst) 
df.loc[i, 'StreetNodeID'] = clst_NID 
df.loc[i, 'NodeDist'] = shrtst 
print dict_NID 
df.to_csv('DestinationAddress.csv', index=False, encoding='utf-8')
### 1. Construct the Riyadh Road Network graphic RG

#### 1.1 Define an empty undirected graph "RG" to store the road network data

```python
RG = nx.Graph()
```

#### 1.2 Read dataframe "route_edges_am"

```python
def = pd.read_csv("TaxiData&Script/RiyadhRoad/data/riyadh_route_edges_am.txt", sep=" ", index_col=0)
# Define the values dictionary for the "edge attributes"
keys = ['volume', 'capacity', 'voc', 'degree', 'free_travel_time', 'travel_time']
for row in def.itertuples(index=False):
    values = row[2:]
    # We create a dictionary "edge attributes" with the keys and row values
    edge_attributes = dict(zip(keys, values))
    my_tuple = (row[0], row[1])
    # We add the edge to the graph use "add_edge" function
    RG.add_edge(*my_tuple, attr_dict=edge_attributes)
```

#### 1.3 Read and Add coordinates to nodes of graphic RG

```python
nodes_df = pd.read_csv("TaxiData&Script/RiyadhRoad/data/riyadh_nodes.txt", sep=" ", index_col=0)
nodes_df.head(2)
for node_id in RG.nodes():
    RG.node[node_id]["coordinate"] = (nodes_df['st_x'][node_id], nodes_df['st_y'][node_id])
loc_dict = {}
for node_index in RG.nodes():
    #print node_index
    loc_dict[node_index] = RG.node[node_index]["coordinate"]
```

#### 1.4 Check if the network is connected?

```python
print "** Is the road network connected? " + str(nx.is_connected(RG))
```

### 2. Load dataframe

```python
df = pd.read_csv('Destination_Address.csv', encoding='utf-8', engine='python')
df.head(2)
RG_Dest = nx.Graph()
keys = ['Name','Category','X','Y']
for row in df.itertuples(index=False):
    values = row[0:4]
    # We create a dictionary "edge attributes" with the keys and row values
    node_attributes = dict(zip(keys, values))
    my_node = (row[0])
    # We add the edge to the graph use "add_edge" function
    RG_Dest.add_node(my_node, attr_dict=node_attributes)
loc_Dest = {}
clr_Dest = []
for node_index in RG_Dest.nodes():
    #print node_index
    loc_Dest[node_index] = (RG_Dest.node[node_index]["X"], RG_Dest.node[node_index]["Y"])
```
if RG_Dest.node[node_index]['Category'] == 'Education':
    clr_Dest.append('#87bdd8')
elif RG_Dest.node[node_index]['Category'] == 'Retail':
    clr_Dest.append('#ff6f69')
elif RG_Dest.node[node_index]['Category'] == 'Healthcare':
    clr_Dest.append('#f9d5e5')
else:
    clr_Dest.append('#ffcc5c')

plt.figure(figsize=(20,20))
nx.draw(RG, pos=locdict, node_size=0, width=1, node_color='#E5E7E9',
       edge_color = '#E5E7E9', label = 'Road');
nx.draw(RG_Dest, pos=locDest, node_size=100, node_color=clr_Dest,
       label = 'Destination');
nx.draw_networkx_labels(RG_Dest,pos=locDest,font_size=4,font_color='b2b2b2');
plt.legend(loc='upper right')