JMP STATISTICAL DISCOVERY

The Relevance of Measurement Systems Analysis

A Procter & Gamble Case Study on MSA Methodology and Applications

DATE
OCTOBER 10 AND 12

TIME
16:00 CET, 10 am EST

CHRISTIAN NEU
Scientist
Procter & Gamble

JERRY FISH
Systems Engineer
JMP

JASON WIGGINS
Senior Systems Engineer
JMP

Register now

Received: 27 October 2019 | Revised: 19 December 2019 | Accepted: 18 January 2020

DOI: 10.1002/widm.1362

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY WILEY

On the determinants of Uber accessibility and its spatial distribution: Evidence from Uber in Philadelphia

Sina Shokoohyar[1] | **Anae Sobhani**[2] | **Saeed R. Ramezanpour Nargesi**[3]

[1]Erivan K. Haub School of Business, Saint Joseph's University, Philadelphia, Pennsylvania

[2]Department of Human Geography and Planning, Utrecht University, Utrecht, The Netherlands

[3]Department of Civil Engineering, University of Texas at Arlington, Arlington, Texas

Correspondence
Sina Shokoohyar, Erivan K. Haub School of Business, Saint Joseph's University, Philadelphia, PA 19131.
Email: sshokooh@sju.edu

Abstract

This study investigates the impact of socioeconomics and demographic factors (e.g., population density, minority rate, age, gender, education, wealth, and crime) and transportation infrastructure (e.g., walk score, transit score, and bike score) on the accessibility of Uber in the city of Philadelphia. *K*-means clustering is applied for initial data exploration. Based on the spatial model selection diagnostic tests, we developed maximum likelihood spatial lag models with queen contiguity spatial weight matrix. The results show that Uber accessibility is not balanced in different neighborhoods in Philadelphia. Uber is more accessible in denser areas with the high male population, better public transportation access and less access to amenities in the walkable distances. Moreover, we observed that Uber is more accessible in areas with a high crime rate. This observation shows that Uber has made it easier to get out of high crime rate areas. Finally, contribution in the literature on accessibility in ride-sourcing networks is discussed. Findings are additionally used to provide managerial implications to mitigate discrimination in ride-sourcing platforms.

This article is categorized under:

Application Areas > Industry Specific Applications
Algorithmic Development > Spatial and Temporal Data Mining
Commercial, Legal, and Ethical Issues > Social Considerations

KEYWORDS

accessibility, public transportation versus Uber, race, gender, and wealth discrimination, ride-sourcing platforms

1 | INTRODUCTION

Sharing the economy has attracted a great deal of attention in recent decades and many studies have been conducted regarding the usage of sharing economies, especially ride-sourcing platforms such as *Uber* and *Lyft* (B. Cohen & Kietzmann, 2014; Schor, 2016; Zha, Yin, & Yang, 2016). Uber and Lyft are U.S. based ride-sourcing companies. These companies work through mobile applications that connect individuals willing to pay for a ride with independent drivers willing to provide a ride with their privately owned vehicles. In the ride-sourcing business, riders open a mobile application and search for the available rides for a specific route. Then, they can choose to request a ride. If a ride request is sent, then the application calculates the fare according to the time and the distance that will be traveled and bills the rider automatically. In 2016, Uber has been operated in over 503 cities across 77 countries. In the United Kingdom, London alone, more than 1 million Londoners have shared their trips by using Uber Pool services (Rodionova, 2016).

These Uber services have saved more than 700,000 driving miles, 50,000 L of petrol, or 124 metric tons of carbon dioxide emissions (Rodionova, 2016). Lyft also operates in 30 U.S. states (Harding, Kandlikar, & Gulati, 2016). In 2017, Uber and Lyft owned 54 and 37% of the U.S. ride-sourcing market, respectively (Certify, 2017).

In ride-sourcing platforms, drivers are independent contractors, and they can flexibly choose where and when to provide ride services (Rogers, 2015; Woo & Bales, 2017). Therefore, managing the supply of the driver is more challenging in the ride-sourcing platforms compared to the more traditional transportation modes such as buses and taxies. To manage supply of drivers, Uber incentivizes drivers' availability using a different type of promotions (Scheiber, 2017). However, the accessibility of drivers is still a concern and has attracted a lot of attention in the literature (Hughes & MacKenzie, 2016; Jiang, Chen, Mislove, & Wilson, 2018; Thebault-Spieker, Terveen, & Hecht, 2017; Wang & Mu, 2018).

Uber has become a transportation mode for people along with public transit including train, buses, and taxies. Demand in the ride-sourcing platforms in different areas of a city may differ concerning the availability of other transportation modes such as public transportation, availability of biking routes, and walkability of the neighborhood. The strength of the demand may impact the supply of drivers as ride-sourcing platforms apply fare adaptation policies to equilibrate demand with supply (Diakopoulos, 2015; Shokoohyar, 2018a; Uber, 2019b). Areas with higher demand often have a higher surge price compared to the neighborhoods with low demand. A higher surge price, in turn, may attract more drivers and lower the pick-up waiting times.

Socioeconomic and demographic factors may also impact the supply of drivers in the ride-sourcing platforms. Drivers may avoid neighborhood with a high crime rate and unstable demand (Begley, 2015; Rogers, 2016). In turn, people in trouble hot spot areas may face huge waiting times. Discrimination against minorities and female is also a huge concern in ride-sourcing platforms (Ge, Knittel, MacKenzie, & Zoepf, 2016; O'Brien, 2018). In ride-sourcing platforms, drivers are independent contractors and ride-sourcing platforms do not have much power to impose policies against discrimination.

Analyzing the impact of transportation infrastructure as well as socioeconomic and demographic factors on accessibility of rides in ride-sourcing platforms can help transportation policymakers to legislate policies against discrimination. In this study, we developed spatial lag models to investigate the impact of socioeconomics and demographic factors (e.g., population density, minority rate, age, gender, education, wealth, and crime rate) and transportation infrastructure (e.g., walk score, transit score, bike score) on the accessibility of Uber in the city of Philadelphia. The main research questions that we try to answer in this research are (a) “How does accessibility of other transportation modes impact Uber's accessibility?”, (b) “Has Uber made it easier to get out of areas with high crime rate?”, and “Does Uber discriminate against minorities, gender, and wealth?”

To address these research questions, datasets are collected from several resources: Uber estimated pick-up waiting time is collected from Uber API (Application Program Interface) in May and Jun 2019 and in total 2,322,044 data points are collected. The socioeconomic information regarding each zip code is collected from censusreporter.org on June 2019. Crime incidents from the Philadelphia police department are collected on opendataphilly.org website. Walk score, transit score, and bike score are also included in this study.

The results show that Uber accessibility is not balance a different areas of Philadelphia. Uber is more accessible in the neighborhood with better public transportation access and less access to amenities in the walkable distances. Uber is more accessible in denser areas with the high male population and a high crime rate. This observation shows that Uber has made it easier for people to get out of the neighborhood with a high crime rate.

The remainder of this manuscript includes literature review, data collection and methodology, model results, and concluding remarks.

2 | LITERATURE REVIEW

This study mainly focuses on the determinants of Uber accessibility and its spatial distribution. Literature on different aspects including applications of machine learning approach, previous studies on a particular sharing company (Uber) and finally the most related articles concerned about Uber accessibility are reviewed.

Machine learning and *K*-mean clustering is used as methodological approaches in many different studies (Shokoohyar, 2018b; Shokoohyar, Qi, & Katok, 2017). Moreno Izquierdo, Egorova, Peretó Rovira, and Más Ferrando (2018) indicate that using a machine learning approach has helped the companies to enhance their knowledge about the users and optimize their performance. They showed that the estimation process of the neural network approach provides them significantly better results than hedonic models in studying 10,000 Airbnb properties. Georganos et al.

(2019) implemented a machine learning approach named random forest (RF) to geographical random forest (GRF) to be used as a tool for modeling population as a function of remote sensing (RS) covariates. Their proposed technique to understand (a) the relationship between dependent and independent variables and (b) the local variations which permit the modelers to recognize the process which has caused the observed spatial heterogeneity. Another study by Hengl, Nussbaum, Wright, Heuvelink, and Gräler (2018) applied RF for spatial predictions. In this study, the buffer distances from observation points are considered explanatory variables, so the prediction process is affected by geographical process. Rana, Jasola, and Kumar (2011) mentioned that the *k*-means clustering is the most popular clustering algorithm. For instance, Montazeri-Gh and Fotouhi (2011) applied a *k*-means clustering algorithm to identify driving segment clustering to recognize the traffic condition. Therefore, this method is not unorthodox in transportation studies. The presented study also applies a machine learning approach and *k*-means clustering to understand Uber accessibility and its spatial distribution.

As stated in the introduction section, one of the main research questions of this study is "How does accessibility of other transportation modes impact Uber's accessibility?" Confrontation between ride-sourcing and public transportation services in North America in terms of their convenience, cost fare, traffic and their environmental effects has attracted a lot of attentions (Hill, 2018). A comprehensive comparison between Uber, Lyft and taxies regarding supply, demand, price and waiting time in San Francisco and New York City were conducted by Jiang et al. (2018). The outcome of this study showed that the socioeconomic and transportation infrastructure affect the market features of these services. A comparison between traditional taxies and Uber by Rayle, Dai, Chan, Cervero, and Shaheen (2016) reveals that Uber has much shorter waiting time than traditional taxies.

Demand (ride requests) and supply (drivers' availability) have a large fluctuation on the ride-sourcing platforms. Fare adaptation policy (surge pricing) is a strong tool employed by ride-sourcing platforms to equilibrate demand and supply. This policy impacts average pick-up waiting times and average trip durations (M. C. Cohen & Zhang, 2017) which in turn influences the preference of customers to use ride-sourcing services versus traditional taxi or public transit systems (Jiao, 2018).

Fare adaptation policy equilibrates supply and demand by closing the gap between them. This in turn leads to improved outcomes for both riders (lower pick-up waiting times [higher accessibility]) and drivers (higher earnings). Hall, Kendrick, and Nosko (2015) showed that how ride requests drop and drivers' supply jump during the surge times in New York city. Similarly, Chen and Sheldon (2016) studied the responses of Uber drivers to the surge pricing and found that Uber drivers are more available in high earning times and locations. Guda and Subramanian (2019) studied how surge pricing can be used to control the supply of drivers in the ride sourcing platforms. Few studies directly investigated the impact of fare adaptation polices on drivers' strategies. For instance, Malin and Chandler (2017) interviewed 18 Pittsburg-based Uber and Lyft drivers. Shokoohyar (2018a) studied the impact of fare adaptation policies through analyzing drivers' comments on social networks. These studies confirm the significant impact of surge pricing on drivers' strategies.

Weather also might affect requests for using Uber (Shokoohyar, Sobhani, & Sobhani, 2020). Brodeur and Nield (2017) showed that Uber rides in NYC from April to September 2014 and January to June 2015 were higher by 25% which is much more than increase in requesting taxis (only 4%). They also showed that the total number of rides increased by approximately 9% when Uber entered the market and access to a ride in rainy weather became easier in nonrainy weather.

Other important research questions of this study are "Has Uber made it easier to get out of areas with high crime rate?", and "Does Uber discriminate against minorities, gender, and wealth?". Alemi, Circella, Handy, Mokhtarian, and Sperling (2018) used California millennials dataset to investigate the factors which influence the frequency and usage of ride-sourcing services as well as the impacts of such services on other elements of travel behavior. In another study by Fisman and Luca (2016), a discussion about fixing the discrimination issues such as gender or race on service such as Uber, eBay and Airbnb are discussed. Schoenbaum (2016) also investigates the closeness significance of a sharing economy and the implications associated with these economies for sex equality. They argue that the main concern is the intimacy in a sharing economy increases the prominence of sex to conduct in this type of business. Finally, they concluded that sex equality is in question in sharing economy since even laws are not able to reduce sex discrimination due to the intimacy in the sharing economy. Ge et al. (2016) conducted an interesting study by sending out passengers to Seattle and Boston to observe nearly 1,500 rides in specific routes. The results showed that the waiting time for African American passengers was increased by 35% is Seattle. In Boston, also the Uber drivers canceled more passengers with African American sound name. The overall outcome showed that the cancelation for African American sound names were two times more than White people. Another interesting result showed that the male passengers who requested rides in the low dense area have a likelihood of trip cancelation three times more when they had African American name in comparison with White names. The female passengers were taken for longer and more expensive trips in Boston. Overall, this study shows the gender and racial discrimination of a sharing economy more clearly. In another research study by Hughes and MacKenzie (2016), the spatial variability in wait times for a Transportation Network Company vehicle in Seattle is studied to explore if the waiting time for low-income people or other minorities is different than other areas. They concluded that such companies provide more services in dense areas and these services are not particularly offered to White or wealthy people. Wang and Mu (2018) explored the spatial disparities of accessibility of UberX and UberBlack in Atlanta. They also confirmed that wealth and race do not affect Uber accessibility. However, Uber is more accessible in dense areas. They added that where more public transportation stops exist, the accessibility of UberX is better, but the accessibility of UberBlack is worse.

Section 5.1 discusses findings that are consistent with the previous studies and Section 5.2 discusses the contribution of this study on the literature of ride-sourcing accessibility.

3 | DATA COLLECTION AND METHODOLOGY

In this section, data collection procedure and the employed spatial lag model for the analysis purposes are presented in detail.

3.1 | Data collection

This study covers Philadelphia city area in the United State, and it includes 48 zip codes as presented in Figure 1. Each zip code area is considered as one observation. Sample data regarding these zip codes are collected from several sources: Uber API, censusreporter.org, and opendataphilly.org.

First, Uber estimated pick-up waiting time is collected from the Uber API. Uber's API is developed to help third-party developers to incorporate its services into their application. Through Uber API, third-parties can get an estimate of the pick-up waiting time for any given location for all available type of services of the given location. The centroid of

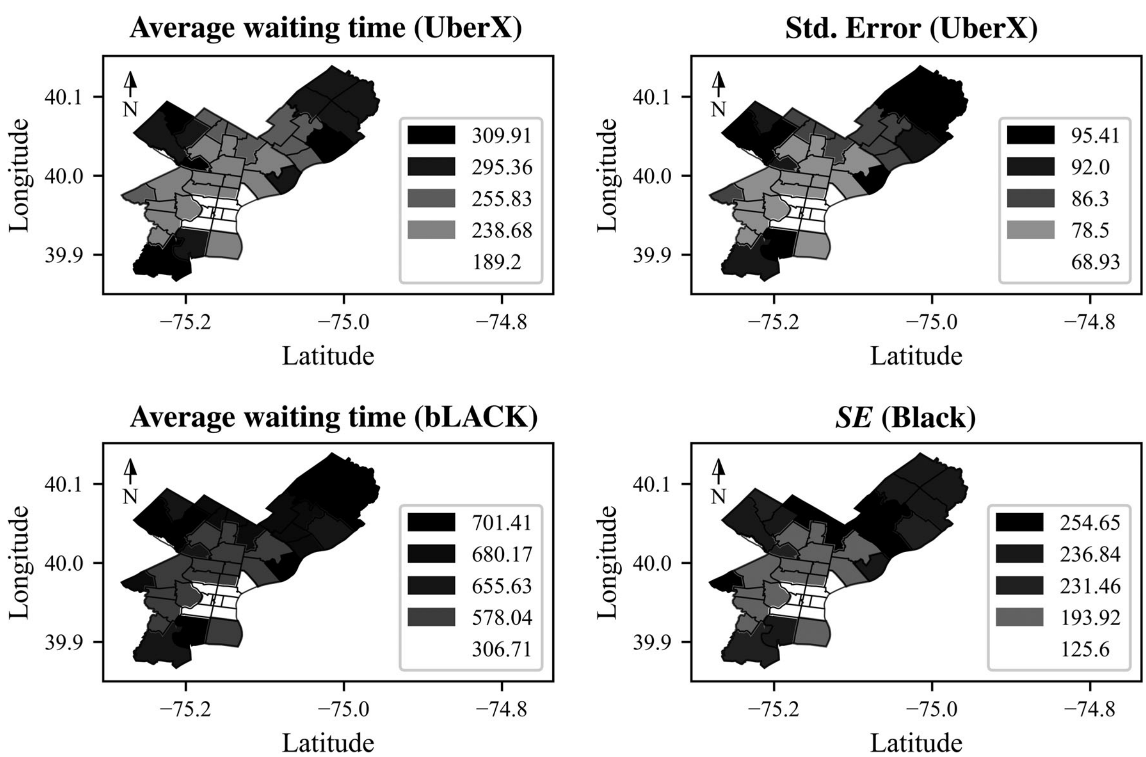

FIGURE 1 Areas clustered based on explanatory variables

each zip code area is considered as the pick-up location. Estimated pick-up waiting times are for hypothetical ride request originated from the centroid of each zip code area. Uber provides pick-up waiting times based on the origin of the trip, and therefore the pick-up waiting times do not depend on the trip destination. Accessing Uber API is free and therefore a very large dataset can be collected free of charge for any pick-up location. Such collected data from Uber and Lyft are extensively used and trusted in literature of ride-sourcing platforms (L. Chen, Mislove, & Wilson, 2015; Jiao, 2018; Wang & Mu, 2018). Pick-up waiting times are estimations based on GPS data collected from their cabs. Estimates can vary based on demand patterns and real-world factors like traffic or road construction. Uber and Lyft claim that these estimates are very close to the actual data, but they are not guaranteed (Golson, 2016; Joseph, 2018; Uber, 2019a). This set of data is collected in May and June 2019. In total 2,322,044 data points are collected. The average and standard error of waiting time of each service type (i.e., UberX and UberBlack in this study) at each location are considered in the analysis.

Second, the socioeconomic information regarding each zip code is collected from censusreporter.org on June 2019. Censusreporter provides useful socioeconomic and demographic facts about every place in America. The variables collected are population, area (in square miles), percent White, percent male, median age, percent bachelor's degree or higher and median household income (in U.S. dollars). The population density in this study is generated from dividing the population by area, and minority is derived by subtracting percent White from 100. Two zip core areas (i.e., 19109 and 19112) are removed from the sample data as some information regarding these two zip codes were not available on censusreporter.org.

Third, crime incidents from the Philadelphia police department are available on opendataphilly.org website. In total, there are 1,048,575 crime incidents in this dataset. The % crime incidents per capita are generated by dividing the number of crime incidents at each zip code by the population size multiplied by 100. The variable facilities are also collected from opendataphilly.org. This variable is generated from summing the number of public facilities such as police and fire department, health centers, and so forth divided by the population size.

Fourth, walk score measures the walkability of neighborhoods. For each zip code, walk score shows how walkable amenities are in that area. Points are awarded based on the distance to amenities in each category. Transit score shows how well an area is supported by the public transportation system. Transit score measures usefulness of nearby transit routes based on the frequency, type of route (rail, bus, etc.), and distance to the nearest stop on the route. Bike score measures how suitable an area is for biking. For a given location, a bike score is generated by measuring bike infrastructure (lanes, trails, etc.), hills, destinations and road connectivity, and the number of bike commuters.

The descriptive statistics of all the variables used in this study are presented in Table 1. The next section will explain how these factors are used in this study to analyze Uber accessibility.

3.2 | Methodology

To explore the impacts of population density, minority, median age, male population ratio, education, household income, crime rate, facilities, walk score, transit score, and bike score on Uber accessibility the following model is developed.

$$\begin{aligned} y = {} & \beta_0 + \beta_1 PopDen + \beta_2 Minority + \beta_3 MedAge + \beta_4 PerMale + \beta_5 PerBacDeg + \beta_6 MedInc \\ & + \beta_7 CriInc + \beta_8 Facilities + \beta_9 WS + \beta_{10} TS + \beta_{11} BS + \epsilon. \end{aligned} \tag{1}$$

The dependent variable y is one of the five variables of interest: average waiting time of UberX, standard error of waiting time of UberX, average waiting time of UberBlack, standard error of waiting time of UberBlack, and the average waiting time of UberBlack divided by average waiting time of UberX.

First, all attributes used in the analysis are logged transformed. Second, the ordinary least squares method is applied to estimate the coefficients. To check for multicollinearity the variation inflation factor is calculated, and the result shows that the model does not suffer from multicollinearity. Third, to test the spatial dependency Moran's I (error), Lagrange Multiplier (lag), Robust Lagrange Multiplier (lag), Lagrange Multiplier (error), and Robust Lagrange Multiplier (error) are applied. To run these tests a row-standardized spatial weight matrix (W) was constructed using queen contiguity with the first-order of neighbors. The results of these tests are provided in Table 3. Fourth, based on the result of these tests the appropriate spatial model is selected following the flowchart of regression model selection

(Anselin, 2013). For a detailed review on spatial models, the readers are referred to Anselin (2001, 2013) and Anselin and Hudak (1992). Summary of the spatial lag regression models including the *direct* and *total effect* of each independent variable and the appropriate model evaluation tests are presented in Tables 4–6. Direct effects describe the average effect of an independent variable at a specific location on the value of the dependent variable at that location. In contrast, indirect effects show the relationship between the value of a dependent variable at a given location and the values of independent variables at neighboring locations. The total effect simply is a sum of a direct and indirect effect.

4 | MODEL RESULTS

This section is organized as follows: first, in Section 4.1 average waiting of UberX and UberBlack is explored via visualization and *K*-mean clustering approach for better understanding of the determinants of Uber accessibility. Then, Section 4.2 provides statistical evidence on how these factors affect Uber accessibility.

4.1 | Exploring accessibility

Table 1 presents descriptive statistics of explanatory variables (i.e., PopDen, Minority, MedAge, PerMale, PerBacDeg, MedInc, CriInc, Facilities, WS, TS, and BS) of the sample data used in this study.

To summarize the information in the explanatory variables, *K*-means clustering approach is applied to categorize the zip codes (areas) in Philadelphia into five categories. Each zip code is then assigned to one, and only one, cluster depending on its features' values. This approach will provide us with a perspective on the types of areas in a city, and on how they are distributed over space. To give equal importance to all features, features are scaled uniformly. *K*-means is a commonly used and simple unsupervised machine learning algorithm to cluster observations into a user-specified number of clusters. It has several applications in transportation (Agard, Morency, & Trépanier, 2006; Ghazanfari, Jafari, & Rouhani, 2011), and customer behavior analysis (Shokoohyar, 2019; Shokoohyar, Shokouhyar, & Naseri, 2020; Shokouhyar, Shokoohyar, Raja, & Gupta, 2020; Shokouhyar, Shokoohyar, & Safari, 2020). For a detailed review on clustering methods, we refer the readers to (Rana et al., 2011; Vora & Oza, 2013). To determine the right number of clusters the *elbow method* is used. For the best number of clusters (i.e., 5 clusters) the sum of squared distances is 9.4558 and the method achieves the positive Silhouette score of 0.3471. Table 2 represents descriptive statistics of the explanatory variables (features) in these five clusters. Note that Table 2 should be read horizontally. The result shows that moving from Cluster 1 to Cluster 5 the value of all features decrease except minority, median age, and facilities. To explore how these features impact the average waiting time and the standard error of waiting time of UberX and UberBlack in these five clusters, the areas in Philadelphia are color-coded in these five clusters based on the average and standard error of UberX and UberBlack waiting time in Figure 1. Note that the average waiting time and the standard error of waiting

TABLE 1 Variables and descriptive statistics

Variable	Description	Min	Average	Max	*SD*
PopDen	Population density (#/square miles)	1887.3000	14,872.5870	35,899.4000	7,449.5745
Minority	Minority rate (%)	12.0600	59.8841	98.1500	28.3950
MedAge	Median age (year)	22.6000	34.8783	44.4000	4.7146
PerMale	Percent male (%)	43.1800	47.5739	53.9900	2.3999
PerBacDeg	Percent bachelor's degree or higher (%)	3.9400	31.5039	85.4100	21.6582
MedInc	Median household income (in USD)	15,232.00	47,201.78	106,823.00	20,256.46
CriInc	% crime incidents per capita	3.6338	12.5501	46.7023	8.3667
Facilities	Police and fire department, health centers, etc.	0.0000	0.0003	0.0010	0.0003
WS	Walk score	11.0000	76.9565	100.0000	20.6440
TS	Transit score	37.0000	65.2391	100.0000	16.5048
BS	Bike score	35.0000	68.8261	93.0000	15.0337

TABLE 2 Descriptive statistics of explanatory variables at each cluster

Variable	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Avg.	*SE*	Avg.	*SE*	Avg.	*SE*	Avg.	*SE*	Avg.	*SE*
PopDen	21,399.44	2,472.46	16,640.03	1,062.65	16,911.81	1,361.68	6,360.76	726.18	4,362.95	1,128.5
Minority	35.29	4	83.12	3.8	76.73	7.59	32.58	6.92	50.96	12.89
MedAge	32.7	0.8	32.15	1.01	36.08	1.47	40.14	1.24	37.33	2.41
PerMale	49.17	0.85	46.68	0.59	46.57	0.63	48.41	0.64	47.54	1.23
PerBacDeg	63.09	4.8	16.65	2.41	19.25	1.93	29.74	4.24	39.34	11.66
MedInc	71,811.6	5,820.43	27,707.6	2,177.16	42,443.67	1800.49	54,506.5	2,956	54,876.75	9,462.49
CriInc	17.99	4.37	15.92	1.03	7.64	0.58	6.82	1.53	8.84	2.1
Facilities	0.0005	0.0001	0.0003	0.0001	0.0001	0	0.0002	0	0.0007	0.0001
WS	93.4	1.96	84.6	3.56	74.33	3.69	50.12	7.97	66.75	12.31
TS	79.3	6.05	71.87	3.23	60.22	1.5	45.88	2.09	55.25	2.95
BS	84.6	2.54	73.87	2.62	65.33	1.43	51.75	4.31	52.5	7.01
# of Obs.	4		12		7		16		7	

time at each cluster are presented in the figure's legend, for example, average waiting time of UberX in Cluster 1 (the white area in the center of top left figure) is 189.2 s.

Figure 1 represents four main points. First, five areas are scattered on rings centered at the center city (white area in the center of all four figures). This observation shows that areas closer to each other share similar socioeconomics features.

Second, moving from the center city to the areas around it, average waiting time and the standard error of waiting time increases. This observation indicates that Uber is more accessible around the center city (that is Cluster 1) and moving from Cluster 1 to Cluster 5 Uber becomes less accessible. This result is also in line with (Shokoohyar, Sobhani, et al., 2020); they observed that Uber becomes less accessible by getting away from the center city.

Third, as observed in Table 2, the population density, percent bachelor degree or higher, median income, crime incident per capita, walk score, transit score and bike score increase from Cluster 1 (the white area shown in Figure 1) to Cluster 5 (black areas shown in Figure 1). That shows that Uber is more accessible in areas with high population density, percent bachelor's degree or higher, median income, crime incident per capita, walk score, transit score, and bike score.

Fourth, as explained in observations 1–4, the dependent variable and independent variables are spatially correlated. Spatial dependence is determined by similarities in position and attributes (Longley, Goodchild, Maguire, & Rhind, 2005). It shows the extent of the similarity between variables that are spatially nearby, closer the distance is, higher dependency they might get (Mennis & Jordan, 2005). Results are largely unsurprising; several studies have confirmed spatial dependency between socioeconomic factors and their impact on sharing economy platforms such as the impact of crime rate on Airbnb rental (Xu, Kim, & Pennington-Gray, 2017), impact of population density, unemployment rate, house value and minority on Uber accessibility (Wang & Mu, 2018), impact of population density and median income on UberX waiting time and TaskRabbit (Thebault-Spieker et al., 2017), Uber and Lyft accessibility, travel time, and fair (Shokoohyar, Sobhani et al., 2020).

To formally test the spatial dependency, Global Moran' *I*, Lagrange Multiplier, and Robust Lagrange Multiplier tests are run on the residual of the OLS model. The results of these tests are presented in Table 3. Three models ordinary least squares (OLS), spatial error model (SEM), and spatial autoregressive model (spatial lag model) are evaluated following the flowchart of regression model selection (Anselin, 2013) to select the best model for this study. First note that, the *p*-value of Moran's *I* test is significant for all models which indicates that the model residuals have a spatial pattern to them that we are failing to control for by only using OLS. Second, in all models Lagrange Multiplier (lag) is significant while the Lagrange Multiplier (error) is not. Note that in UberX SEM, the *p*-value of Lagrange Multiplier (lag) is less than *p*-value of Lagrange Multiplier (error). Therefore, the results show that the spatial lag model should be used for all the four models. To account for the spatial dependence, spatial lag regression (Anselin, 1988) is used; the result of the spatial lag models is presented in Section 4.2. Note that *p*-values that are significant under 5% are given in bold.

4.2 | Spatial modeling

Summary of the spatial lag models of UberX and UberBlack are presented in Tables 4 and 5, respectively. Tables 4a and 5a represent the average waiting time models, and Tables 4b and 5b represent the standard error of waiting time models. To evaluate the models three tests are provided (*F*-test, likelihood ratio test, and Moran's *I*). First, all *F*-statistic is significant under 5% significance level suggesting that the sample data provide sufficient evidence that the regression

TABLE 3 Spatial model selection diagnostics (*p*-value)

Test	UberX		UberBlack	
	Avg.	*SE*	Avg.	*SE*
Moran's *I* (error)	**0.0104**	**0.0003**	**0.0033**	**0.0481**
Lagrange Multiplier (lag)	**0.0029**	**0.0000**	**0.0002**	**0.0002**
Robust Lagrange Multiplier (lag)	**0.0011**	**0.0000**	**0.0003**	**0.0000**
Lagrange Multiplier (error)	0.7442	**0.0465**	0.1563	0.8357
Robust Lagrange Multiplier (error)	0.1743	0.1071	0.2074	**0.0003**
R-squared	0.8829	0.7462	0.8130	0.7862

TABLE 4 Maximum likelihood spatial lag summary of UberX (dependent variables: average waiting time (a) and standard error (b))

Independent variables	(a)				(b)			
	Avg. waiting time				Std. error			
	Direct	*SE*	Total	*p*-Value	Direct	*SE*	Total	*p*-Value
Constant	8.8246	1.7671	13.5134	**0.0000**	5.2016	1.5195	19.1744	**0.0006**
PopDen	−0.1080	0.0272	−0.1654	**0.0001**	−0.0240	0.0243	−0.0883	0.3252
Minority	0.0141	0.0344	0.0216	0.6811	−0.0206	0.0312	−0.0759	0.5088
MedAge	0.0237	0.1075	0.0363	0.8255	−0.0075	0.0967	−0.0275	0.9386
PerMale	−0.6662	0.3216	−1.0202	**0.0383**	−0.7126	0.2904	−2.6267	**0.0141**
PerBacDeg	−0.0246	0.0302	−0.0377	0.4156	−0.0117	0.0271	−0.0431	0.6659
MedInc	−0.0406	0.0566	−0.0621	0.4734	−0.0237	0.0511	−0.0873	0.6432
CriInc	−0.0727	0.0268	−0.1113	**0.0066**	−0.0693	0.0234	−0.2555	**0.0030**
Facilities	0.0052	0.0147	0.0080	0.7228	0.0156	0.0135	0.0574	0.2468
WS	0.1409	0.0441	0.2158	**0.0014**	0.0927	0.0393	0.3418	**0.0184**
TS	−0.2966	0.0782	−0.4542	**0.0001**	−0.1918	0.0694	−0.7072	**0.0057**
BS	−0.1029	0.0817	−0.1576	0.2080	0.0142	0.0731	0.0522	0.8465
ρ	0.3470	0.0969	−0.1576	**0.0003**	0.7287	0.0760		**0.0000**
F-statistic	27.6037			**0.0000**	29.0204			**0.0000**
Likelihood ratio test	10.7636			**0.0010**	43.8038			**0.0000**
Moran's *I*	−0.0577			0.3548	−0.0488			0.3902
R-squared	0.9099				0.9169			
Mean squared error	0.0039				0.0053			
Log likelihood	62.9377				67.7357			

model fits the data better than the model with no independent variables. Second, the likelihood ratio test is also significant indicating that incorporating the spatial lag term improves the models. Additionally, all *R*-squared of the spatial lag models are higher compared to the respective *R*-squared of the OLS models. Third, Moran's *I* test is not significant in all models. This result indicates that the inclusion of the spatial lag term explains away the residual autocorrelation in all models.

Table 4 makes six main points. First, the coefficient for population density in average waiting time model is negative and significant indicating that UberX is more accessible in areas with high population density. The areas with high population density have a higher demand that attracts more Uber driver and in turn decreases the Uber waiting time. Second, the negative and significant coefficient of percent male in both models shows that the male population generates a higher and more stable demand and in turn attract more Uber drivers that the leads to lower waiting time. Third, areas with higher crime incidents, motivate residents to take Uber more often for the safety issues and in turn leads to higher and more stable demand. The higher demand, in turn, attracts more Uber drivers and decreases the Uber waiting time. Note that the coefficient of crime incidents per capita is negative and significant in both models. Uber engineers are studying patterns of ride requests and crimes to predict ride request (Yvkoff, 2011). In denser areas, there is more incidents of prostitution, theft, and alcohol-related crimes; where there are more people, there is also more demand for Uber. Fourth, the positive and significant coefficients of walk score in both models show that Uber is less accessible in areas with a higher walk score. Note that walk score measures the walkability of any area and points are awarded based on the distance to amenities. Accessing amenities in areas with high walk score is simpler for residents and therefore demand for Uber decreases in these areas. Lower demand, in turn, attracts fewer drivers, which leads to higher waiting time. Fifth, the negative and significant coefficients of transit score in both models show that Uber is more accessible in areas with higher transit score. Note that transit score shows how well an area is supported by the transit routes and incorporates the importance of distance to the nearest stop on the route. The area with a better road network is more accessible to Uber drivers, which leads to lower waiting time for rides. This observation also shows that Uber covers similar areas with high access to public transportation. Sixth, the positive and significant coefficient of ρ suggests that there is a positive spatial autocorrelation in both models.

Table 5 provides seven main points. First, the coefficient for population density in the standard error of waiting time model is negative and significant indicating that UberBlack waiting time fluctuates more in areas with high population density. Second, the positive and significant coefficient of walk score in the average waiting time of UberBack model shows that Uber is less accessible in areas with a higher walk score. Third, the negative and significant coefficient of transit score in the average waiting time of UberBack model shows that Uber is more accessible in areas with higher transit score. The first, second, and third observations are similar to that are observed in Table 4 with similar logic behind them. Fourth, the positive and significant coefficient of the minority in the standard error of UberBlack waiting time model shows that UberBlack has a higher fluctuation in areas with higher minority rate. This observation indicates that UberBlack is not a stable service in areas with high minority rate. UberBlack drivers may avoid providing services in areas with high minority rate depending on the time of the day that a ride request is made (daytime vs. night time) which leads to high fluctuation in UberBlack accessibility. Fifth, the negative and significant coefficient of bachelor's degree or higher in the standard error of UberBlack waiting time model shows that UberBlack has a lower fluctuation in areas with higher education. Areas with more educated population generate more stable demand for UberBlack which in turn attract more drivers and a more stable UberBlack service. Note that the coefficient of bachelor's degree or higher in the average waiting time of UberBlack model is also weakly significant with *p*-value of .0705 as well. Sixth, areas with higher crime incidents, motivate resident to take UberBlack more often for the safety issues and in turn leads to more stable demand. Note that this observation is similar to what we observed in Table 4 with a similar logic behind it. Seventh, the positive and significant coefficient of ρ suggests that there is a positive spatial autocorrelation in both models.

TABLE 5 Maximum likelihood spatial lag summary of UberBlack (dependent variables: average waiting time (a) and standard error (b))

Independent variables	(a) Avg. waiting time				(b) *SE*			
	Direct	*SE*	Total	*p*-Value	Direct	*SE*	Total	*p*-Value
Constant	8.6905	3.9049	20.1176	**0.0260**	−0.4698	2.9079	−1.8012	0.8717
PopDen	−0.0972	0.0663	−0.2249	0.1430	−0.1028	0.0508	−0.3940	**0.0431**
Minority	0.0513	0.0822	0.1187	0.5325	0.1295	0.0637	0.4966	**0.0419**
MedAge	0.0270	0.2564	0.0624	0.9163	0.0047	0.1979	0.0182	0.9809
PerMale	−0.8838	0.7681	−2.0459	0.2499	0.3210	0.5918	1.2309	0.5875
PerBacDeg	−0.1333	0.0737	−0.3085	0.0705	−0.1291	0.0578	−0.4951	**0.0255**
MedInc	−0.0674	0.1345	−0.1561	0.6162	0.1214	0.1036	0.4654	0.2415
CriInc	−0.1757	0.0616	−0.4067	**0.0043**	−0.1067	0.0464	−0.4093	**0.0214**
Facilities	−0.0159	0.0373	−0.0367	0.6707	−0.0097	0.0284	−0.0370	0.7341
WS	0.2441	0.1039	0.5650	**0.0188**	0.0272	0.0753	0.1043	0.7177
TS	−0.4933	0.1828	−1.1420	**0.0070**	−0.1271	0.1404	−0.4874	0.3652
BS	0.1128	0.1887	0.2612	0.5499	0.1790	0.1431	0.6862	0.2112
ρ	0.5680	0.1019		**0.0000**	0.7392	0.0847		**0.0000**
F-statistic	20.8774			**0.0000**	22.7898			**0.0000**
Likelihood ratio test	18.9524			**0.0000**	26.1045			**0.0000**
Moran's *I*	0.0047			0.3886	−0.1116			0.1746
R-squared	0.8864				0.8980			
Mean squared error	0.03065				0.0218			
Log likelihood	22.9636				34.8049			

As observed in Figure 1, both UberX and UberBlack are more accessible in the center city and they become less accessible by getting further away from the center city. To analyze how UberX and UberBlack accessibility changes in a different area compared to each other, areas all clustered into five areas using *K*-mean clustering approach. Two features are incorporated for the clustering purpose, the average waiting time of UberX and UberBlack. Figure 2 shows the average waiting time of UberBlack divided by the average waiting time of UberX in these five clusters. This figure shows that in the center city average waiting time for UberBlack is approximately 17% more than average waiting for UberX and getting away from the center city it increases to 190%. To formally analyze the important factors impacting the difference between UberX and UberBlack accessibility, a spatial lag model is run with an independent variable of average waiting time of UberBlack divided by average waiting of UberX. The summary of this model is presented in Table 6.

Table 6 provides three main points. First, the average waiting time of UberBlack is closer to UberX in areas with higher crime incidents. This observation shows that crime incident has a higher impact on UberBlack compared to

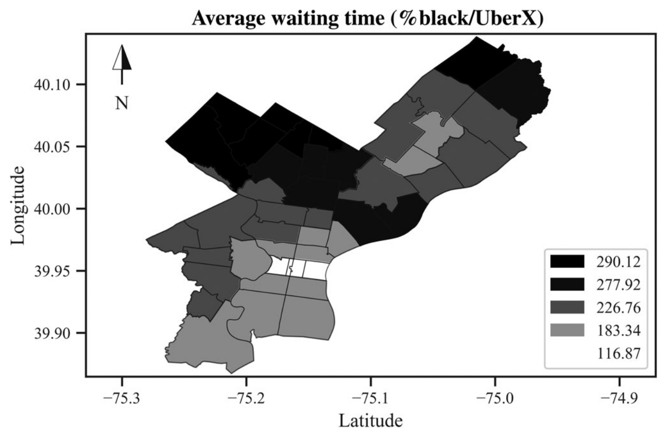

FIGURE 2 Areas clustered based on average waiting time of UberX and Black

TABLE 6 Maximum likelihood spatial lag summary (dependent variables (average waiting time (black divided by UberX)))

Independent variables	Avg. waiting time			
	Direct	*SE*	Total	*p*-Value
Constant	3.9835	3.1222	9.4201	0.2020
PopDen	0.0008	0.0535	0.0019	0.9879
Minority	0.0329	0.0664	0.0779	0.6200
MedAge	−0.0195	0.2075	−0.0462	0.9251
PerMale	−0.2768	0.6209	−0.6546	0.6557
PerBacDeg	−0.1080	0.0600	−0.2554	0.0720
MedInc	−0.0494	0.1086	−0.1167	0.6496
CriInc	−0.1291	0.0471	−0.3053	**0.0062**
Facilities	−0.0235	0.0306	−0.0556	0.4425
WS	0.1484	0.0793	0.3509	0.0613
TS	−0.2483	0.1470	−0.5872	0.0912
BS	0.1577	0.1491	0.3728	0.2904
ρ	0.5771	0.1148		**0.0000**
F-statistic	11.3449			**0.0000**
Likelihood ratio test	14.8889			**0.0001**
Moran's *I*	0.0450			0.2407
R-squared	0.8131			
Mean squared error	0.0141			
Log likelihood	32.6903			

UberX. Note that the coefficient of CriInc is negative and significant in both Tables 4 and 5. This coefficient is −0.1757 for UberBlack which is less than −0.0727 in UberX model. Second, the average waiting time of UberBlack is closer to UberX in areas with a more educated population. In areas with a more educated population, demand for UberBlack is higher, which attracts more UberBlack drivers and in turn decrease UberBlack waiting time (note that the coefficient of PerBacDeg is negative and weakly significant in Table 5). This fact in turn results to closer waiting time between UberBlack and UberX in areas with a population that is more educated (note that coefficient of PerBacDeg is negative and weakly significant in Table 6). Third, the coefficient of WS (walk score) and TS (transit score) is weakly significant indicating that UberBlack waiting time is closer to UberX waiting time in areas with lower walk score and higher transit score.

5 | CONCLUDING REMARKS

Sections 5.1 and 5.2 discuss the contribution of this study on the literature of ride-sourcing accessibility. Then managerial implication of these findings and the future research opportunities are presented in Section 5.3.

5.1 | Findings consistent with previous studies

The findings of this study confirm several results in the literature. First, Uber is more accessible in denser areas (Hughes & MacKenzie, 2016; Jiang et al., 2018; Thebault-Spieker et al., 2017; Wang & Mu, 2018). Second, Uber is not necessarily more accessible in areas with lower minority rate (Hughes & MacKenzie, 2016; Jiang et al., 2018; Wang & Mu, 2018). Third, there is no evidence showing that wealth (median household income) affects Uber waiting time. This result is in line with Jiang et al. (2018) regarding waiting of Uber, additionally, they observed negative and significant coefficient regarding Lyft waiting time indicating that Lyft is more accessible in areas with higher income.

5.2 | Findings different from previous studies

There are several key points that this study contributes to the literature on ride-hailing. First, median age does not affect Uber accessibility. Second, UberX is more accessible in areas with higher male ratio. Third, Jiang et al. (2018) showed that education does not have any impact on Uber waiting time. In contrast, the results of our study show that UberBlack (note that UberBlack is Uber's luxury option compared to UberX) is more accessible in areas with a higher level of education. Fourth, there is significant evidence supporting that Uber is more accessible in areas with the higher crime rate. Higher Uber accessibility in areas with higher crime rate means that Uber is making it easier to get out of trouble hot spots. Uber driver may want to avoid providing service in areas with a high crime rate and discriminate against these areas (Begley, 2015; Rogers, 2016). However, the high ride request in areas with high crime rate leads to a higher surge price which in turn attracts more drivers and leads to lower waiting time (Yvkoff, 2011). This result also partially confirms findings in (Li, 2019). He showed that the *number of Uber pick-ups* increases in the crime rate, indicating that Uber is more accessible in areas with the higher crime rate. Fifth, public facilities such as police and fire department, and health centers do not have any impact on Uber accessibility. Sixth, Uber is more accessible in areas with higher transit score. This result is partially in line with Wang and Mu (2018) which observed that Uber is more accessible in areas with higher road density. Eighth, Uber is more accessible in areas with a lower walk score. Ninth, there is no evidence supporting that bike score has any impact on Uber accessibility.

5.3 | Managerial implication and future research studies

The result of this study confirms that Uber accessibility is not balanced in different areas of Philadelphia. Uber is more accessible in denser areas with the high male population, high crime rate, better public transportation access, and less access to amenities in the walkable distance. Uber surge pricing approach motivates Uber drivers to move to areas with higher surge price and high demand which is more profitable. An unbalanced ride-hailing network may reduce the satisfaction of riders in using Uber services and in turn reduce ride-sourcing market share. Wiser use of surge pricing approach can make Uber accessible more balanced in all areas of the city and improve customer satisfaction. Ride-hailing companies can also balance the network by motivating drivers to be available in different areas of the city by providing promotions for drivers when they accept rides in low demand areas.

Areas with high male population have better access to Uber. This is not necessarily due to gender discrimination. Areas with higher male population may have a stronger demand that is more attractive to drivers which in turn makes Uber more accessible in these areas. A further a study should be conducted to directly analyze gender discrimination by running a survey among both drivers and riders for clearer evidence of gender discrimination.

The result shows that Uber and the public transportation system are covering the same areas of the city. Both Uber and public transportation are more accessible in the city center. This makes getting out of the city center easier and getting into the city center more difficult compared to the areas away from the center city for Philadelphians. With regard to this observation, we make the following implications for policy-makers and regulators. First, Uber and the public transportation system can collaborate to provide a more balanced network throughout the city. Developing a revenue-sharing contract between ride-sourcing platforms and public transportation can significantly impact the unbalanced transportation network. As a promising future research direction, by collecting data from public transit systems and traditional taxi systems, city transportation planners, transportation network company managers, and traditional taxi managers may have better ideas about riding performances in Philadelphia, providing a win-win collaboration that allow Philadelphian in getting better riding services throughout the city. The result of our study shows that, transportation demand decreases in walkability of the neighborhoods, and therefore, city planners may improve walkability in different areas of the city to better manage transportation demand.

This study is limited to Philadelphia during June and May 2019. For future research, more data should be collected in different cities while including other factors, mainly time. Several research studies have shown that Uber accessibility varies throughout the day (e.g., it becomes less accessible during rush hours.). Additionally, other machine learning tools like GRF, neural network, or support vector machine with radial basis kernel can be applied and the results can be compared.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Sina Shokoohyar: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; visualization; writing-original draft; writing-review and editing. **Anae Sobhani**: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; visualization; writing-original draft; writing-review and editing. **Saeed R. Ramezanpour Nargesi**: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; visualization; writing-original draft; writing-review and editing.

ORCID

Sina Shokoohyar https://orcid.org/0000-0002-4271-4736

RELATED WIREs ARTICLE

Ride-hailing's impact on Canadian cities: Now let's consider the long game

REFERENCES

Agard, B., Morency, C., & Trépanier, M. (2006). Mining public transport user behaviour from smart card data. *IFAC Proceedings Volumes*, *39* (3), 399–404.

Alemi, F., Circella, G., Handy, S. L., Mokhtarian, P. L., & Sperling, D. (2018). *What Makes Travelers Use Ridehailing? Exploring the Latent Constructs behind the Adoption and Frequency of Use of Ridehailing Services, and Their Impacts on the Use of Other Travel Modes*. 212.

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, *20*(1), 1–17.

Anselin, L. (2001). Spatial econometrics. In *A companion to theoretical econometrics* (Vol. 310330).

Anselin, L. (2013). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.

Anselin, L., & Hudak, S. (1992). Spatial econometrics in practice: A review of software options. *Regional Science and Urban Economics*, *22*(3), 509–536.

Begley, D. (2015). *For disabled or people in remote areas, Uber service hard to come by*. Retrieved from Houston Chronicle website: https://www.houstonchronicle.com/news/transportation/article/For-disabled-or-people-in-remote-areas-Uber-6534879.php

Brodeur, A., & Nield, K. (2017). *Has uber made it easier to get a ride in the rain?*

Certify. (2017). *Uber declines, Lyft picks up in the Certify SpendSmart™ Report for Q3 2017*. Retrieved from Certify website: https://www.certify.com/2017-10-24-Uber-Declines-Lyft-Picks-Up-in-the-Certify-SpendSmart-Report-for-Q3-2017

Chen, L., Mislove, A., & Wilson, C. (2015). Peeking beneath the hood of uber. In *Proceedings of the 2015 Internet Measurement Conference* (pp. 495–508).

Chen, M. K., & Sheldon, M. (2016). Dynamic pricing in a labor market: surge pricing and flexible work on the Uber platform. In *EC* (p. 455).

Cohen, B., & Kietzmann, J. (2014). Ride on! Mobility business models for the sharing economy. *Organization & Environment*, *27*(3), 279–296.

Cohen, M. C., & Zhang, R. P. (2017). *Coopetition and profit sharing for ride-sharing platforms*.

Diakopoulos, N. (2015). How Uber surge pricing really works. *The Washington Post*, *17*. https://www.washingtonpost.com/news/wonk/wp/2015/04/17/how-uber-surge-pricing-really-works/.

Fisman, R., & Luca, M. (2016, December). Fixing discrimination in online marketplaces. *Harvard Business Review*. Retrieved from https://hbr.org/2016/12/fixing-discrimination-in-online-marketplaces

Ge, Y., Knittel, C. R., MacKenzie, D., & Zoepf, S. (2016). *Racial and gender discrimination in transportation network companies*.

Georganos, S., Grippa, T., Gadiaga, A. N., Linard, C., Lennert, M., Vanhuysse, S., … Kalogirou, S. (2019). Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 1–16. https://doi.org/10.1080/10106049.2019.1595177

Ghazanfari, M., Jafari, M., & Rouhani, S. (2011). A tool to evaluate the business intelligence of enterprise systems. *Scientia Iranica*, *18*(6), 1579–1590.

Golson, J. (2016). *Lyft now shows your fare before you ride, just like Uber*.

Guda, H., & Subramanian, U. (2019). Your Uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science*, *65*(5), 1995–2014.

Hall, J., Kendrick, C., & Nosko, C. (2015). The effects of Uber's surge pricing: A case study. *The University of Chicago Booth School of Business*.

Harding, S., Kandlikar, M., & Gulati, S. (2016). Taxi apps, regulation, and the market for taxi journeys. *Transportation Research Part A: Policy and Practice*, *88*, 15–25.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/10.7717/peerj.5518

Hill, S. (2018). Ridesharing versus public transit. How Uber and Lyfy tend to widen disparities of race and class in urban transportation sytems.

Hughes, R., & MacKenzie, D. (2016). Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators. *Journal of Transport Geography*, *56*, 36–44.

Jiang, S., Chen, L., Mislove, A., & Wilson, C. (2018). On ridesharing competition and accessibility: Evidence from Uber, Lyft, and taxi. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web* (pp. 863–872).

Jiao, J. (2018). Investigating Uber price surges during a special event in Austin, TX. *Research in Transportation Business & Management*.

Joseph, B. B. (2018). *Understanding upfront fares*.

Li, K. (2019). *Investigating the effect crime has on Uber and Yellow Taxi pickups in NYC*.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science*. London, England: John Wiley & Sons.

Malin, B. J., & Chandler, C. (2017). Free to work anxiously: Splintering precarity among drivers for Uber and Lyft. *Communication, Culture & Critique*, *10*(2), 382–400.

Mennis, J. L., & Jordan, L. (2005). The distribution of environmental equity: Exploring spatial nonstationarity in multivariate models of air toxic releases. *Annals of the Association of American Geographers*, *95*(2), 249–268.

Montazeri-Gh, M., & Fotouhi, A. (2011). Traffic condition recognition using the *k*-means clustering method. *Scientia Iranica*, *18*(4), 930–937. https://doi.org/10.1016/j.scient.2011.07.004

Moreno Izquierdo, L., Egorova, G., Peretó Rovira, A., & Más Ferrando, A. (2018). *Exploring the use of artificial intelligence in price maximisation in the tourism sector: Its application in the case of Airbnb in the Valencian Community*. Retrieved from https://ebuah.uah.es/dspace/handle/10017/37170

O'Brien, S. A. (2018). *Uber is under a federal investigation for gender discrimination*.

Rana, S., Jasola, S., & Kumar, R. (2011). A review on particle swarm optimization algorithms and their applications to data clustering. *Artificial Intelligence Review*, *35*(3), 211–222.

Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, *45*, 168–178.

Rodionova, Z. (2016). UberPOOL gets more than 1 million customers. *The Independent*.

Rogers, B. (2015). The social costs of Uber. *University of Chicago Law Review Dialogue*, *82*, 85.

Rogers, B. (2016). *Uber's algorithms could spot crimes in progress. But do we want them to*? Retrieved from Washington Post website: https://www.washingtonpost.com/posteverything/wp/2016/02/24/uber-could-do-more-to-prevent-crimes-in-progress-but-we-might-not-want-it-to/?noredirect=on&utm_term=.f432d5442b73

Scheiber, N. (2017). How Uber uses psychological tricks to push its drivers' buttons. *The New York Times*, *2*. https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html.

Schoenbaum, N. (2016). *Gender and the sharing economy*. Retrieved from https://papers.ssrn.com/abstract=2865710

Schor, J. (2016). Debating the sharing economy. *Journal of Self-Governance and Management Economics*, *4*(3), 7–22.

Shokoohyar, S. (2018a). Ride-sharing platforms from drivers' perspective: Evidence from Uber and Lyft drivers. *International Journal of Data and Network Science*, *2*(4), 89–98.

Shokoohyar, S. (2018b). *Supplier performance management: A behavioral study*. (Doctoral dissertation). University of Texas at Dallas, Dallas, Texas.

Shokoohyar, S. (2019). Determinants of backpackers perceptions of security? A WOM-based approach. *E-Review of Tourism Research*, *16*(4), 249–270.

Shokoohyar, S., Qi, A., & Katok, E. (2017). *Project management under risk-sharing contracts*. Available at SSRN 2889868.

Shokoohyar, S., Shokouhyar, S., & Naseri, S. (2020). Improving internet service providers (ISP) competitiveness: ISP's perception regarding customer satisfaction. *International Journal of Business and Systems Research*.

Shokoohyar, S., Sobhani, A., & Sobhani, A. (2020). Impact of trip characteristics and weather condition on ride-sourcing network: Evidence from Uber and Lyft. *Research in Transportation Economics*.

Shokouhyar, S., Shokoohyar, S., Raja, N., & Gupta, V. (2020). Promoting fashion customer relationship management dimensions based on customer tendency to outfit matching: Mining customer orientation and buying behavior. *International Journal of Applied Decision Sciences*.

Shokouhyar, S., Shokoohyar, S., & Safari, S. (2020). Research on the influence of after-sales service quality factors on customer satisfaction. *Journal of Retailing and Consumer Services*.

Thebault-Spieker, J., Terveen, L., & Hecht, B. (2017). Toward a geographic understanding of the sharing economy: Systemic biases in UberX and TaskRabbit. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *24*(3), 21.

Uber. (2019a). *How do ETAs work*?

Uber. (2019b). *How surge pricing works*.

Vora, P., & Oza, B. (2013). A survey on *k*-mean clustering and particle swarm optimization. *International Journal of Science and Modern Engineering*, *1*(3), 24–26.

Wang, M., & Mu, L. (2018). Spatial disparities of Uber accessibility: An exploratory analysis in Atlanta, USA. *Computers, Environment and Urban Systems*, *67*, 169–175.

Woo, C., & Bales, R. A. (2017). *The Uber million dollar question: Are Uber drivers employees or independent contractors*?

Xu, Y.-H., Kim, J.-W., &Pennington-Gray, L. (2017). *Explore the spatial relationship between Airbnb rental and crime*.

Yvkoff, L. (2011). *Neighborhoods with more crime have more Uber rides*. Retrieved from CNET website: https://www.cnet.com/roadshow/news/neighborhoods-with-more-crime-have-more-uber-rides/

Zha, L., Yin, Y., & Yang, H. (2016). Economic analysis of ride-sourcing markets. *Transportation Research Part C: Emerging Technologies*, *71*, 249–266.

How to cite this article: Shokoohyar S, Sobhani A, Ramezanpour Nargesi SR. On the determinants of Uber accessibility and its spatial distribution: Evidence from Uber in Philadelphia. *WIREs Data Mining Knowl Discov*. 2020;10:e1362. https://doi.org/10.1002/widm.1362