NYU CUSP

Urban Science Intensive

DO UBER/LYFT REDUCE PARKING VIOLATIONS IN NEW YORK CITY?

Team Members:

Junjie Cai, Junru Lu, Pranay Anchan, Shijia Gu, Yuxuan Wang

Mentor:

Zhan Guo

July 24, 2019

1 Abstract

The fast expansion of Uber and Lyft result in people suspecting these companies' effect on our urban system, for example, on public transportation and city congestion. However, due to the lack of open data and scientific research, most of its influence has not been proven yet.

This capstone project aimed at exploring one potential Uber & Lyft impact: whether daily Uber & Lyft trips affect parking violations. NYC daily Uber & Lyft trips and parking tickets data were collected and correlated by taxi zones. Three technical models, Fixed Effects, Difference in Difference (DID), and Bayesian Network, were applied on the prepared data. The results of these models showed a negative correlation and causal effect between the number of Uber & Lyft trips and parking tickets, suggesting that Uber & Lyft help in reducing parking violations in NYC. Given the controversial issues around TNC, this capstone project could assist in understanding the impact of Uber & Lyft and offer policy insight to the TNC regulation.

2 Introduction

2.1 Problem Definition

Rideshare companies, like Uber & Lyft, are taking over marketing shares from Taxis (Richter, 2018). They together have earned a revenue of $13.46 billion in 2018 (Grocer, 2019), while the revenue in U.S ride-hailing segment is roughly $44.8 billion (statista, 2019). Nevertheless, the impact of Uber & Lyft is not limited to the For-Hire vehicles (FHV) market alone. As a critical part of the transportation system, they are reshaping our urban settings; although these effects are being questioned. On the one hand, Uber & Lyft have created thousands of new jobs with an hourly salary between $8.55 and $11.77 on average (Reed, 2018). On the other hand, they are suspected of increasing major city congestion significantly (Brinklow, 2018), and even damage the public transit system of United States (Schmitt, 2019).

However, due to the limited availability of Uber & Lyft trip data (Uber & Lyft only have public trip data available in NYC) and related scientific research, most of the Uber & Lyft effects have not been uncovered and proven. To impose effective regulations on rideshare companies, it is critical for city administrators to understand them comprehensively. So, the goal of this project was to investigate one potential Uber & Lyft impact: Do Uber & Lyft increase/reduce parking violations in NYC? To explore this, the project leveraged the power of data science to model and empirically prove the causal effect of Uber & Lyft rides on parking violations.

2.2 Literature Review

Because researches on how ridesharing services affect parking violations are rarely available, other related work are reviewed. Henao, in a 2013 paper, used a self-collected dataset containing information of 311 surveys to ridesharing passengers. He asked a few questions in the survey, including the driving frequency and the purpose of taking a ridesharing trip. By analyzing the dataset, he found out there are lots of frequent drivers, who take

TNC rides to avoid parking (Henao & Marshall, 2017). This could be a potential factor in the decrease in parking violations in the city. However, due to the size and nature of his dataset, his analysis cannot statistically prove this effect. Looking at patterns of parking violations in NYC, 'tickets for stopping or parking in illegal zones, blocking traffic, tend to be most in common in areas with a greater commercial concentration of traffic.' The violations usually occur as a 'result of a driver needing to stop temporarily in a busy area where parking is scarce; they try to park in a loading zone or double park for a few minutes to avoid having to search for a parking spot.' (Ackerman & Moustafa, 2011)

For the methodology, the team focused on Causal Inference, as proving causation was the basis of this project. Causal inference is the process of drawing conclusions about a causal connection based on the conditions of the occurrence of an effect. In this project, specifically three models - Bayesian Network, Fixed Effects and Difference-in-Differences (DID) were adopted.

Heckerman and Breese show in their 1996 study how the 'use of causal independence in a Bayesian network could greatly simplify probability assessment as well as probabilistic inference' (Heckerman & Breese, 1996). They used a few models on several artificial and real-world Bayesian networks to better understand general Bayesian networks. In each model, they used the noisy-MAX model to encode all parent-child relationships(Kim & Pearl, 1983), and transformed Bayesian networks into an annotated undirected tree where each clique (node) corresponds to a set of nodes in the original network. The results indicate the 'use of decomposition can decrease inference complexity substantially when nodes have many states and many parents.'

Fixed effects models provide a way to estimate causal effects in analyses where units are measured repeatedly over time. It can eliminate the effects of confounding variables without measuring them or even knowing exactly what they are, if they remain constant over time. One significant drawback with this model is that 'a great deal of information can be lost by focusing only on variation within individuals, thereby ignoring the variation across individuals.' Since this model removes the effects of all time-invariant causes, the standard fixed effects model is unable to estimate the effects of time-invariant measured causes (Glenn Firebaugh & Massoglia, 2013).

Difference-in-differences (DID) is another mainstream statistical technique widely applied in econometrics and social sciences for causal inference. DID attempts to use observational data to mimic the design of experimental research, studying the differential effect of a treatment. (Angrist & Pischke, 2008) Alberto Abadie from Harvard University is one of the representatives who applied the DID model to the field of sociology. He used this idea in 2010 to effectively estimate the negative impact of California's Proposition 99 on local tobacco sales (Abadie, Diamond, & Hainmueller, 2010).

3 Data

3.1 Parking Ticket Data

The parking violations data were obtained from NYC Open Data, and this project used data from 2014 to 2018. Four columns, Issue Date, Violation Code, Street Name, and Borough, and 80 types of parking tickets, which related to private vehicle, are used.

Since the dataset only provides borough and street name for each parking ticket, Google API was used to geocode the given borough and street name for each observation to an approximate location with longitude and latitude and then mapped on to according taxi zones. Eventually, the dataset was grouped by date and taxi zone to show the number of parking tickets issued in each taxi zone in New York for each day for the past five years.

3.2 Uber & Lyft Ride Data

The Uber & Lyft trip data was extracted from the NYC For-Hire Vehicle (FHV) dataset that is available on the Taxi & Limousine Commission (TLC) website. Dispatching base number, which is provided for each observation, is different for different FHV companies. So it was used to identify which trips were conducted by Uber & Lyft.

The collected dataset for Uber & Lyft trips includes fuzzy location by taxi zone (there are 263 taxi zones in New York City, and the mean partition granularity is 3 km^2) and time of pick up and drop off for each trip. Spark was used to group the dataset by trip date and pick-up and drop-off taxi zones. Each row of the prepared dataset indicates the number of trips conducted by Uber & Lyft from one taxi zone to another taxi zone for each day in the four-year window.

3.3 Other Data

Five additional spatial data were included to be used in the DID as clustering variables and in the Bayesian Network as potential causation for parking violation:

- **ACS data** was extracted from the American Community Survey 2015 5-year estimates. 17 statistical attributes, including population density, poverty rate, etc., were used to add the socioeconomic information for the model.
- **NYC crime data** is accessible on NYC Open Data. Three variables: number of felonies, number of violations, and number of misdemeanors, were mapped to taxi zone level and used in the model.
- **SAT result data** is hosted on NYC Open Data. The average score of SAT reading, math, and writing sections was included to measure the education level of each zone.
- **Transportation accessibility data** is available on NYC Open Data. This dataset includes subway entrance and bus stop coordinates. The number of subway entrances and the number of bus stops represent the transportation accessibility.
- **Parking capacity data** was collected from NYC Open Data, including meter parking and parking lot. They indicate the parking capacity of taxi zones.

4 Methodology

Three models - Bayesian Network, Fixed effects, and Difference in Differences (DID) were developed. The combination of the three methods was expected to comprehensively explain the causal effect of Uber & Lyft on parking violations.

4.1 Bayesian Network: Causal Structure Learning

The key idea of Bayesian Network is that one can distinguish correlation from causation if independent causes can be observed (Pearl, 2009). This project integrated 12 observational independent features of taxi zones, such as parking condition, ACS census, and education (Figure 7). So, a complete Bayesian Network can be learned to reveal causation of Uber & Lyft on parking violations in NYC.

Network structure learning estimates a directed acyclic graph (DAG) that captures the dependencies between the variables. To learn a causal structure, both score-based and constraint-based structure learning algorithms were used. Score-based Structure Learning approach construes model selection as an optimization task. It has two building blocks: first apply a 'scoring function' $s_D \colon M \to \mathbb{R}$ that maps models to a numerical score, based on how well they fit to a given data set D; second perform 'search strategy' to traverse the search space of possible models M and select a model with optimal score. Constraint-based Structure Learning attempts to capture the directionality of causal relationships. It has two building blocks: identifying independencies in the data set using χ^2 conditional independence tests, and then constructing a DAG according to identified independence.

4.2 Fixed Effects Model

Fixed effects model (FEM) is widely used to control for unobserved variables when performing causal inference with panel data (Imai & Kim, 2019). FEM fits this study well, which aims to explore the causality using panel datasets shown in Figure 8. The datasets were standardized before modeling in order to improve the optimization training process and measure variable importance. To develop a FEM, dummy variables of each taxi zone were added to a standard OLS model to evaluate the fixed effects of the number of Uber & Lyft trips across different taxi zones. Considering the model for taxi zones $i = 1, \ldots,$ N which is observed at certain time periods $t = 1, \ldots, T$:

$$y_{it} = \alpha + \beta \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} + \gamma \sum_{i=1}^{N} z_i + u_{it} \tag{1}$$

where y_{it} and x_{it} are, respectively, the number of parking tickets (dependent variable) and the number of Uber & Lyft trips (independent variables) in the i-th taxi zone at time t; z_i are unobserved, time-invariant dummy variables in different taxi zones; β and γ represent coefficients of the independent variables and dummy variables; α and u_{it} refers to the intercept and the error term. The coefficient, β, of the independent variables would indicate the causal effect exerted by Uber & Lyft trips on parking violations.

4.3 DID: Difference in Differences

Scientifically, two NYCs, one with TNC and one with not, were needed to prove the causal effect of Uber & Lyft. However, such ideal ’treatment vs. control groups’ are impossible to be found. Consequently, DID model was applied in this project to solve this problem. Figure 1 shows the general idea of DID and how it works. The ’treatment object’ and ’control object’, P and S, have status P_1 and S_1 on time period 1, and status P_2 and S_2 on time period 2. Based on the DID’s most critical assumption, parallel trend assumption, that any existing factor would have same effect on P and S, shown by the parallel S_1 to S_2 line and P_1 to Q dotted line DID, the difference between Q and S_2 on the second time period equals the difference between P_1 and S_1 on the first time period, and the difference between P_2 and Q represents the effect of the outside factor or treatment.

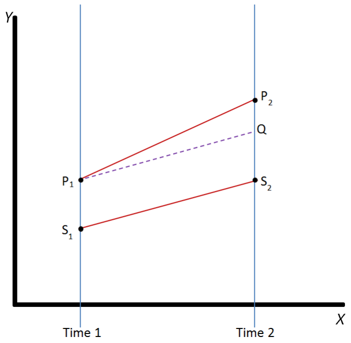


Figure 1: Illustration of DID model, by Danni Ruthvan

Average causal effect can be expressed by formula (2):

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\Delta Outside}{(P_2 - S_2) - (P_1 - S_1)} \tag{2}$$

where $\Delta Outside$ refers to the difference between the average amount of Uber & Lyft trips in 2015-2018 between the ’treatment taxi zone’ P, and the ’control taxi zone’ S, for every i. P_1, P_2, S_1, and S_2 refer to the average amount of parking tickets in 2014 and 2015-2018 between P and S. n refers to the number of ’treatment vs. control’ pairs.

To meet the parallel trend assumption, cross-clustering was applied to find multi-dimensionally homogeneous taxi zones to be used as ’treatment group’ and ’control group.’ The taxis zones were clustered based on two dimensions separately, extracting those zones always belonging to the same clusters: the number of parking tickets for each taxi zone in 2014, which, shown by data, had not been affected by Uber & Lyft; and 5 additional datasets (introduced in section 3.3, shown in Figure 6), which summarizing different aspects of taxi zones. In each cluster, the remaining zones were paired with each other, which generated over 900 pairs of homogenous taxi zones.

5 Results

- The causal network structure generated by Bayesian Network indicates Uber & Lyft having a direct impact on parking tickets. According to Figure 2, there are four variables that directly influence the parking ticket, and the number of Uber & Lyft trips is one of them. However, there are limitations and potential biases with Bayesian Network. Initially, it only states causality instead of explaining the impact. So, whether the Uber & Lyft impact is positive or negative and how strong the impact is, remains unknown. Moreover, Bayesian Network is based on many hypotheses, including causal Markov, causal faithfulness, causal sufficiency, and acyclicity, but it is difficult to meet all of them in real world. Finally, unobserved intermediate factors may be missed between nodes.

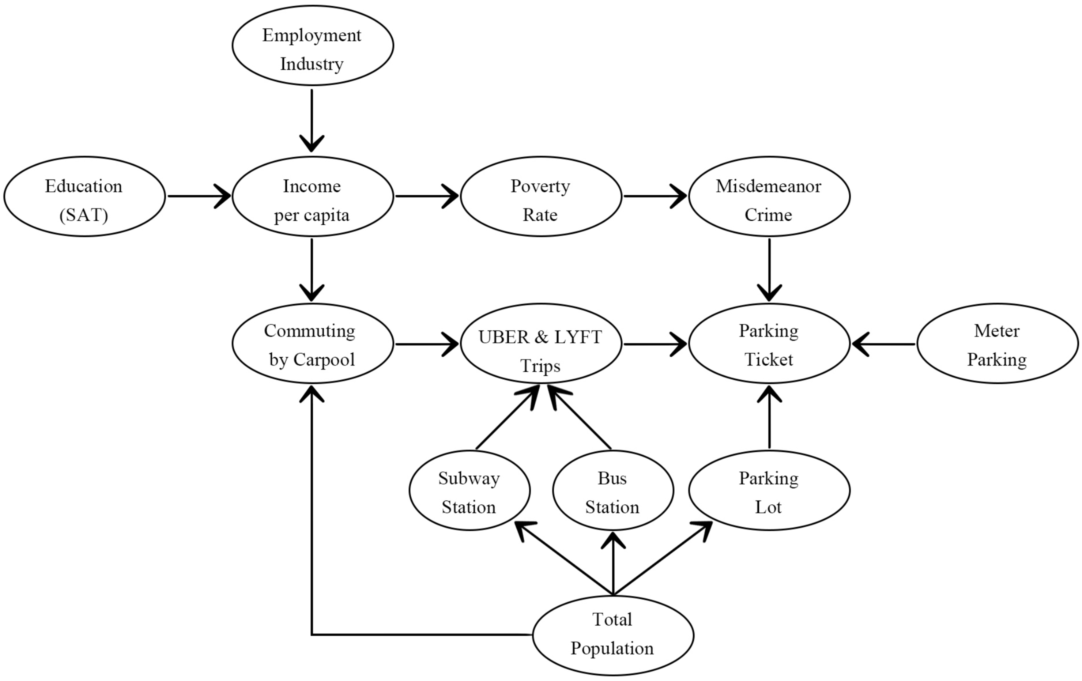
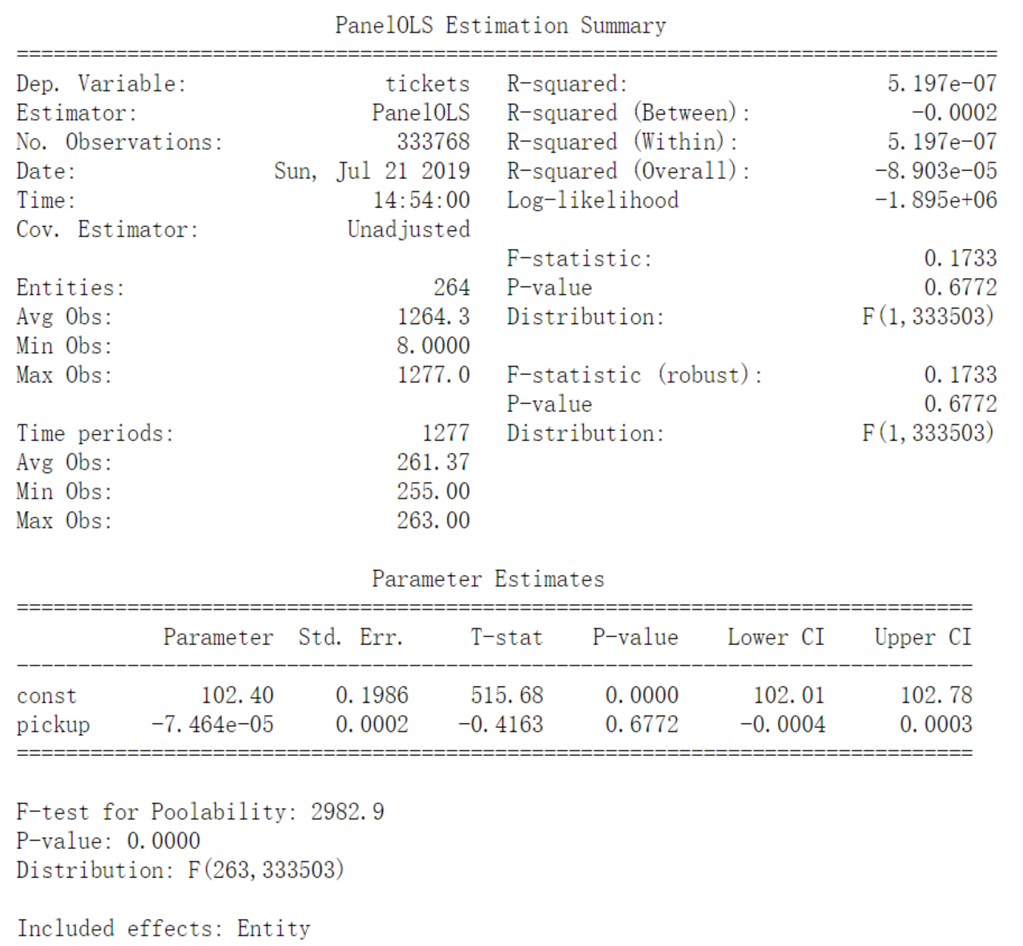


Figure 2: Bayesian Network DAG Graph

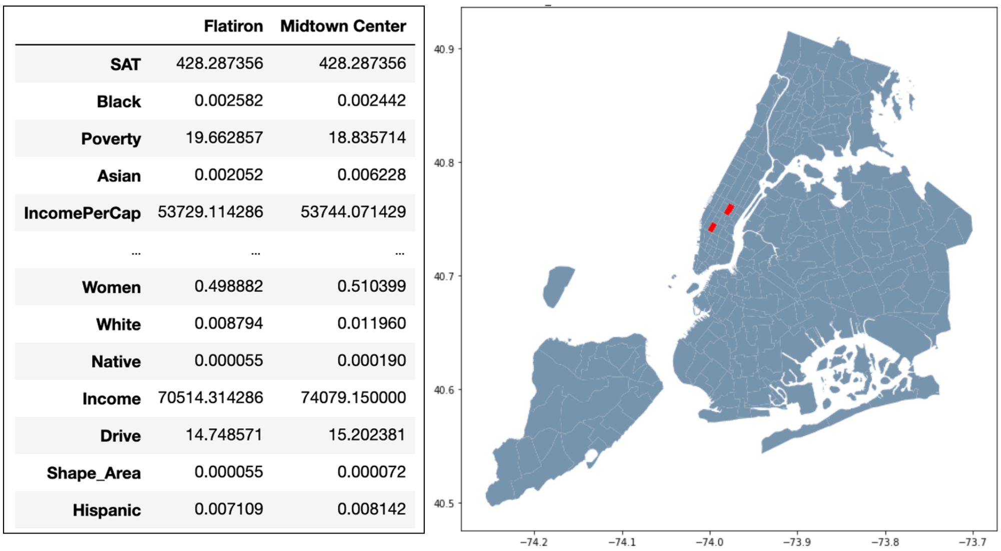
- The fixed effects model (FEM) controls for the unobserved factors. According to Figure 3, R-squared of the Uber & Lyft regression is low since the number of parking tickets cannot be explained by Uber & Lyft trips solely. But after adding dummy variables in regression, the model achieves an R-squared of 0.70 (Figure 9), well explaining the variation of number of parking tickets. The coefficient of Uber & Lyft trips (pickup) is -0.0007, indicating Uber & Lyft trips reduce NYC parking violations slightly. However, this modeled relationship is not statistically significant, because the p-value of 0.677, is much larger than the significance level, 0.05, making the result of FEM less reliable.



```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:                tickets   R-squared:                     5.197e-07
Estimator:                   PanelOLS   R-squared (Between):             -0.0002
No. Observations:              333768   R-squared (Within):            5.197e-07
Date:                Sun, Jul 21 2019   R-squared (Overall):          -8.903e-05
Time:                        14:54:00   Log-likelihood                -1.895e+06
Cov. Estimator:            Unadjusted
                                        F-statistic:                      0.1733
Entities:                         264   P-value                           0.6772
Avg Obs:                       1264.3   Distribution:                F(1,333503)
Min Obs:                       8.0000
Max Obs:                       1277.0   F-statistic (robust):             0.1733
                                        P-value                           0.6772
Time periods:                    1277   Distribution:                F(1,333503)
Avg Obs:                       261.37
Min Obs:                       255.00
Max Obs:                       263.00

                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
const          102.40     0.1986     515.68     0.0000      102.01      102.78
pickup     -7.464e-05     0.0002    -0.4163     0.6772     -0.0004      0.0003
==============================================================================

F-test for Poolability: 2982.9
P-value: 0.0000
Distribution: F(263,333503)

Included effects: Entity
```

Figure 3: Fixed Effects Model Result

- A pair of homogeneous taxi zones in Figures 4 and 5 show how the parallel trend assumption was met. The map in Figure 4 indicates that Flatiron and Midtown Center are similar. Most of their attributes, displayed in the table, are highly close. Their parking violation distribution in 2014 is also similar, shown in blue in the first two line charts in Figure 5. However, the distribution of the number of Uber & Lyft trips and the parking tickets after 2014 of these two areas are different, shown by the orange and green lines on the line charts in Figure 5.



	Flatiron	Midtown Center
SAT	428.287356	428.287356
Black	0.002582	0.002442
Poverty	19.662857	18.835714
Asian	0.002052	0.006228
IncomePerCap	53729.114286	53744.071429
...	...	...
Women	0.498882	0.510399
White	0.008794	0.011960
Native	0.000055	0.000190
Income	70514.314286	74079.150000
Drive	14.748571	15.202381
Shape_Area	0.000055	0.000072
Hispanic	0.007109	0.008142

Figure 4: Features and Locations of a pair of highly homogeneous taxi zones

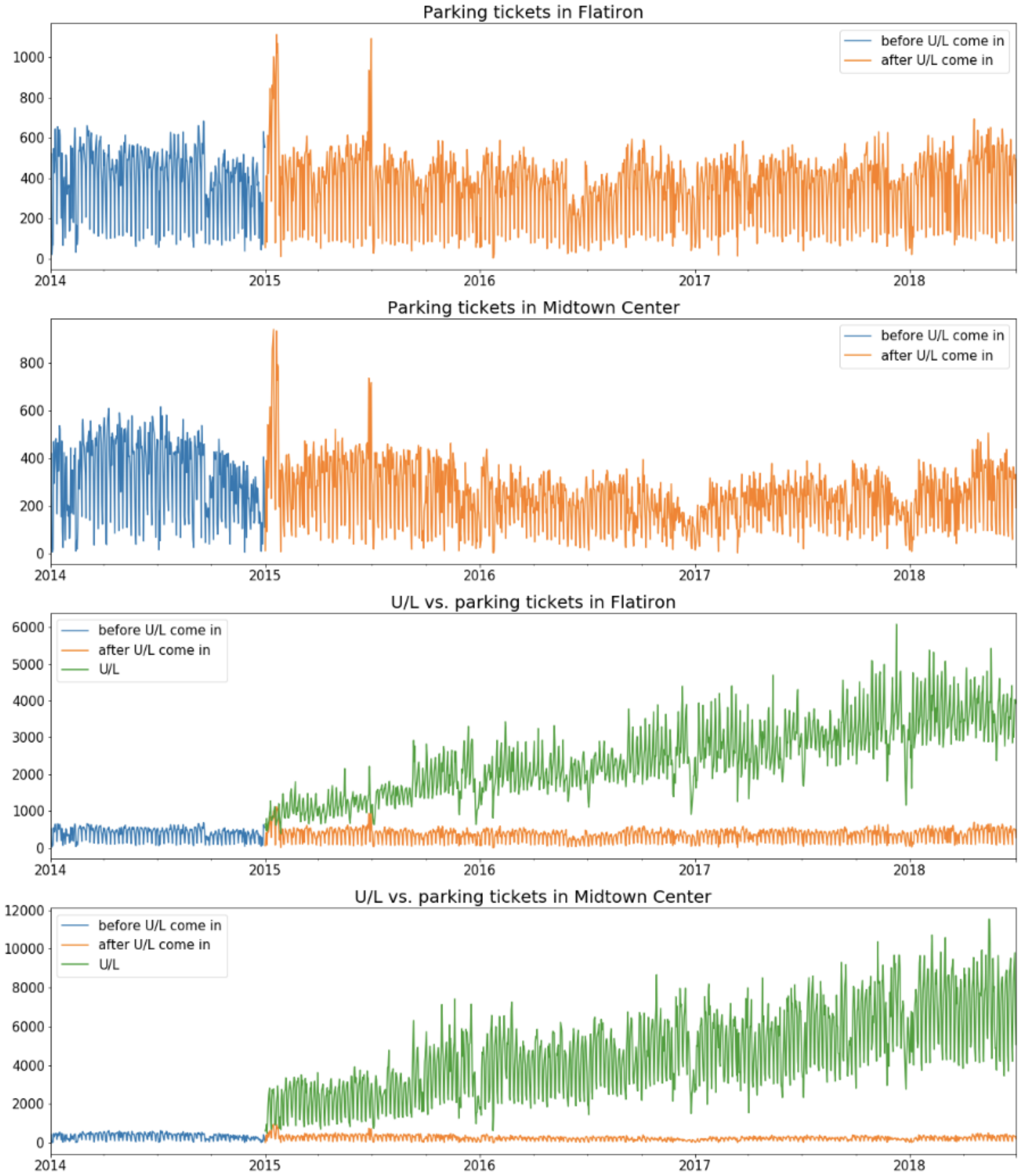


Figure 5: Tickets and TNC Trips of a pair of highly homogeneous taxi zones

Based on all qualified pairs, the DID model gave an average causal effect intensity of 224.2, which implies that an increase of 224.2 Uber & Lyft pickups is associated with a decrease of 1 parking ticket. So, the DID model suggests that Uber & Lyft rides can reduce parking violations in NYC significantly, and this result is quite different from the FEM's output.

6 Conclusions

The main goal of this project was to explore the causal effect between Uber & Lyft and parking violations: determining whether Uber & Lyft increases or reduces the number of parking violations in NYC, and how strong the effect is. After a series of analyses, the results demonstrated that Uber & Lyft have negative causal effect on NYC parking violations, suggesting Uber & Lyft are able to reduce parking violation. However, the extent of the impact has not been fully confirmed yet, due to the difference between the result of the FEM and DID. Future improvements could be made to further modeling the scale of the effect. Besides, validations, such as a hypothesis test, of the Bayesian Network and DID, are also necessary to be included.

With more than 100 million users in the world (Niu, 2019), Uber & Lyft has far-reaching impacts; some yet to be identified and explored. This project serves as an example to apply machine learning methods on the available data to study and prove unnoticed yet potential effects of ride-share companies on cities. Reducing parking violation can be regarded as a positive impact, and also have other effects such as the reduction of parking enforcement pressure for the city. In this aspect, Uber & Lyft could be beneficial to NYC, and potentially to other major cities in the world. However, this is only one aspect of Uber & Lyft's influence. More research could be done in this area to further explore other aspects of Uber & Lyft influence in the city.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, *105*(490), 493–505.

Ackerman, S. S., & Moustafa, R. E. (2011). *Red zone, blue zone: Discovering parking ticket trends in new york city.* Retrieved 2019-07-22, from `https://sites.temple.edu/samackerman/files/2012/10/NYC_parking_Samuel_Ackerman5.pdf`

Angrist, J. D., & Pischke, J.-S. (2008). Mostly harmless econometrics: An empiricist's companion. In (Vol. 4, p. 227-243). Princeton University Press.

Brinklow, A. (2018). *Lyft, uber increase traffic 180 percent in major cities, says report.* Retrieved 2019-06-21, from `https://www.sf.curbed.com/2018/7/27/17622178/uber-lyft-cause-traffic-streets-congestion-bruce-schaller-tnc-report`

Glenn Firebaugh, C. W., & Massoglia, M. (2013). Handbook of causal analysis for social research. In (p. 113). Springer.

Grocer, S. (2019). *How uber and lyft compare, in four charts.* Retrieved 2019-06-21, from `https://www.nytimes.com/2019/04/11/business/dealbook/uber-vs-lyft-ipo-financials.html`

Heckerman, D., & Breese, J. S. (1996). Causal independence for probability assessment and inference using bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, *26*, 826-831.

Henao, A., & Marshall, W. (2017). A framework for understanding the impacts of ridesourcing on transportation. In *Disrupting mobility* (pp. 197–209). Springer.

Imai, K., & Kim, I. S. (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data. *American Journal of Political Science*, *63*(2), 467–490.

Kim, J. H., & Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference engines. *Proceedirzgs /JCAf-83*, 190–193.

Niu, E. (2019). *Uber has nearly 5 times more users than lyft.* Retrieved 2019-04-12, from `https://www.fool.com/investing/2019/04/12/uber-has-nearly-5-times-more-users-than-lyft.aspx`

Pearl, J. (2009). Causal inference in statistics: An overview. *Statist. Surv. 3*, 96-146. doi: 10.1214/09-SS057

Reed, E. (2018). *How much do uber and lyft drivers make in 2018?.* Retrieved 2019-06-21, from `https://www.thestreet.com/personal-finance/education/how-much-do-uber-lyft-drivers-make-14804869`

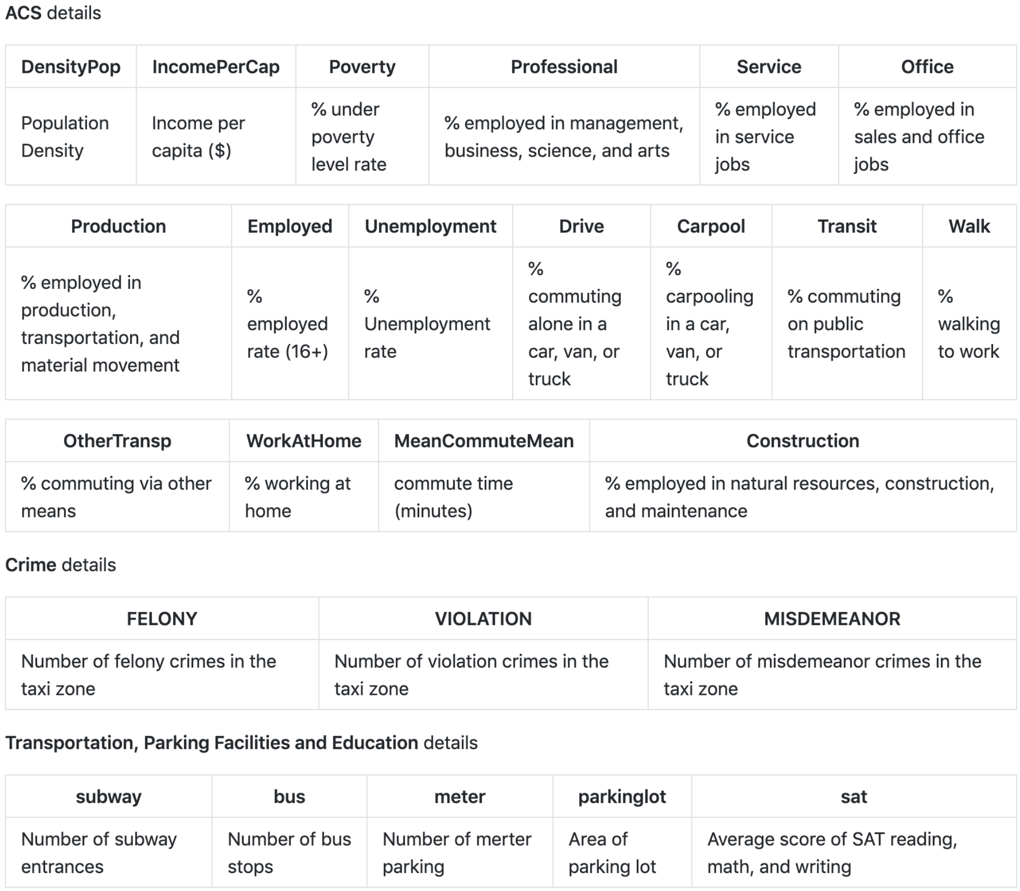
Richter, W. (2018). *Uber and lyft are gaining even more market share over taxis and rentals.* Retrieved 2019-06-21, from `https://www.businessinsider.com/uber-lyft-are-gaining-even-more-market-share-over-taxis-and-rentals-2018-7`

Schmitt, A. (2019). *Study: Uber and lyft caused u.s. transit decline. streetsblog.* Retrieved 2019-06-21, from `https://wwwusa.streetsblog.org/2019/01/22/study-uber-and-lyft-are-responsible-for-u-s-transit-decline`

statista. (2019). *Ride hailing.* Retrieved 2019-06-21, from `https://www.statista.com/outlook/368/109/ride-hailing/united-states`

Appendices

A Dataset Notes



ACS details

DensityPop	IncomePerCap	Poverty	Professional	Service	Office
Population Density	Income per capita ($)	% under poverty level rate	% employed in management, business, science, and arts	% employed in service jobs	% employed in sales and office jobs

Production	Employed	Unemployment	Drive	Carpool	Transit	Walk
% employed in production, transportation, and material movement	% employed rate (16+)	% Unemployment rate	% commuting alone in a car, van, or truck	% carpooling in a car, van, or truck	% commuting on public transportation	% walking to work

OtherTransp	WorkAtHome	MeanCommuteMean	Construction
% commuting via other means	% working at home	commute time (minutes)	% employed in natural resources, construction, and maintenance

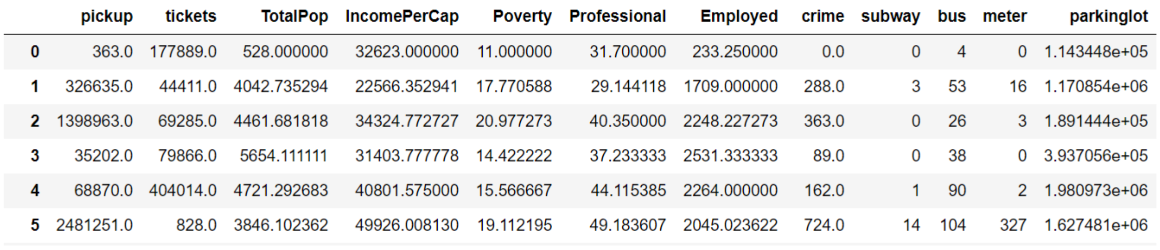
Crime details

FELONY	VIOLATION	MISDEMEANOR
Number of felony crimes in the taxi zone	Number of violation crimes in the taxi zone	Number of misdemeanor crimes in the taxi zone

Transportation, Parking Facilities and Education details

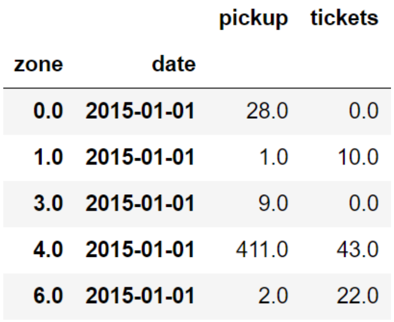
subway	bus	meter	parkinglot	sat
Number of subway entrances	Number of bus stops	Number of merter parking	Area of parking lot	Average score of SAT reading, math, and writing

Figure 6: All factors used in five additional spatial data



	pickup	tickets	TotalPop	IncomePerCap	Poverty	Professional	Employed	crime	subway	bus	meter	parkinglot
0	363.0	177889.0	528.000000	32623.000000	11.000000	31.700000	233.250000	0.0	0	4	0	1.143448e+05
1	326635.0	44411.0	4042.735294	22566.352941	17.770588	29.144118	1709.000000	288.0	3	53	16	1.170854e+06
2	1398963.0	69285.0	4461.681818	34324.772727	20.977273	40.350000	2248.227273	363.0	0	26	3	1.891444e+05
3	35202.0	79866.0	5654.111111	31403.777778	14.422222	37.233333	2531.333333	89.0	0	38	0	3.937056e+05
4	68870.0	404014.0	4721.292683	40801.575000	15.566667	44.115385	2264.000000	162.0	1	90	2	1.980973e+06
5	2481251.0	828.0	3846.102362	49926.008130	19.112195	49.183607	2045.023622	724.0	14	104	327	1.627481e+06

Figure 7: Spatial Data Sample for Bayesian Network



		pickup	tickets
zone	date		
0.0	2015-01-01	28.0	0.0
1.0	2015-01-01	1.0	10.0
3.0	2015-01-01	9.0	0.0
4.0	2015-01-01	411.0	43.0
6.0	2015-01-01	2.0	22.0

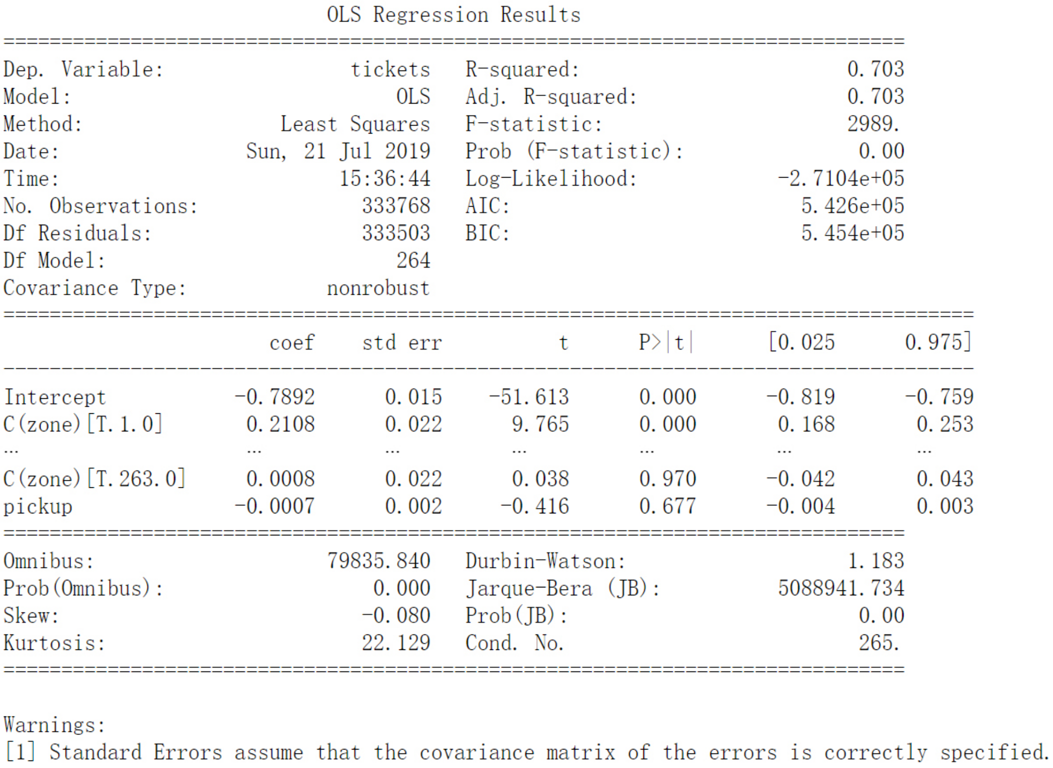
Figure 8: Panel Data Sample for Fixed Effects Modeling

B Team Collaboration Statement

All five members of the team have worked together in this project collaboratively; attending weekly meetings to help in brainstorming and work-split. Looking at specific roles and contributions:

- Junjie Cai: Focused on data processing and model exploration; prepared FHV and other related dataset; completed model exploration and developed Fixed Effect and Bayesian Network Models.
- Junru Lu: Focused on data collection and model exploration; completed model exploration and developed Difference in Difference Model.
- Pranay Anchan: Helped in the data preparation (Geocode FHV data with google API), helped in preparing the report and presentation.
- Shijia Gu: Helped in data collection and preparation, model development, and project report.
- Yuxuan Wang: Focused on the preparation of written report and presentation; Developed the project website; Helped in data collection and model exploration.

C Supplement of Fixed Effect Model



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                tickets   R-squared:                       0.703
Model:                            OLS   Adj. R-squared:                  0.703
Method:                 Least Squares   F-statistic:                     2989.
Date:                Sun, 21 Jul 2019   Prob (F-statistic):               0.00
Time:                        15:36:44   Log-Likelihood:            -2.7104e+05
No. Observations:              333768   AIC:                         5.426e+05
Df Residuals:                  333503   BIC:                         5.454e+05
Df Model:                         264
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
Intercept           -0.7892      0.015    -51.613      0.000      -0.819      -0.759
C(zone)[T.1.0]       0.2108      0.022      9.765      0.000       0.168       0.253
...                     ...        ...        ...        ...         ...         ...
C(zone)[T.263.0]     0.0008      0.022      0.038      0.970      -0.042       0.043
pickup              -0.0007      0.002     -0.416      0.677      -0.004       0.003
==============================================================================
Omnibus:                    79835.840   Durbin-Watson:                   1.183
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          5088941.734
Skew:                          -0.080   Prob(JB):                         0.00
Kurtosis:                      22.129   Cond. No.                         265.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 9: Fixed Effects Model Result with Dummy Variables