Impact of Uber on the traffic in Midtown Manhattan in New York City

Nonie Mathur$^{1}$

**Note:** Throughout this paper, ‘taxis’ include green and yellow taxis and ‘other FHVs’ include all the for hire vehicles other than Uber vehicles and taxis.
**Abstract:** For Hire Vehicles have played an important role in New York City’s transportation. With the increasing number of platforms providing these services, the number of actors in the city’s transportation network have increased, raising a wide range of concerns, including their role played in the city’s traffic congestion.

This project was chosen in light of the debate between Uber and Mayor de Blasio. For my analysis, I used the ‘Aggregate FHV Data’ which was available on FiveThirtyEight’s Github account who have been analyzing the data for the same purposes. This data contains information in the number of pick ups per day by yellow and green taxis, Uber, Lyft and the other ‘For Hire Vehicles’ in Midtown Manhattan.

For my analysis of the research question – Does Uber have an impact on the Traffic congestion of the city, I performed a test of means – Z test to compare the mean of daily pick ups made by Uber vehicles to the mean of daily pick ups made by the other FHVs in Midtown Manhattan. With the calculated Z statistic, I rejected the Null Hypothesis, which proved that Uber vehicles did not lead to traffic congestion in the city (at a significance level of 0.05).

**Introduction:** According to an article published in the blog ‘Hot Air’ in August 2016, when Uber was launched in New York City in the year 2011, the taxi business in the city was booming, increasing the number of medallion licenses being issued. This led to an increase in the number of vehicles on the streets, resulting in traffic congestion.

In Summer 2015, New York City Mayor Bill de Blasio raised his concerns about the increase in traffic congestion due to the increasing number of ride hailing apps, most popular of them being Uber. Mayor de Blasio decided to cap the number of Uber vehicles on the streets in the city implying that the uncapped number of vehicles along with the number of taxis on the streets of the city may lead to ‘urban gridlock’ (FiveThirtyEight, October 2015).

As a result, to study this further, in January 2016, de Blasio administration released the ‘For-Hire Vehicle Transportation Study’ which highlighted that even though the number of Uber vehicles have increased in the city, they are not responsible for the increasing traffic congestion because they are replacing the yellow cabs.

Similarly, a study done by FiveThirtyEight (a website involved in a number of poll analysis in the fields of politics, economics, sports etc.) performed a similar statistical test and came up with the same conclusion as the report by the Mayor’s administration.

Based on the above mentioned studies, I have attempted to answer the following research question:
Does Uber have an impact on the traffic patterns in New York City?
Null Hypothesis:
The average number of Uber pick ups in a day on the streets in Midtown Manhattan is more than the average number of ‘For Hire Vehicles’ and taxi pick ups in a day on the streets in Midtown Manhattan.
Alternate Hypothesis:
The average number of Uber pick ups in a day on the streets in Midtown Manhattan is less than or equal to the average number of ‘For Hire Vehicles’ and taxi trips in a day on the streets in Midtown Manhattan.

The significance level for this analysis is 0.05.

To answer this question, I first specified my null and alternate hypothesis, followed by specifying the significance level. I decided to perform a Z test to answer this research question. The Z test compares the standard deviation of the expected distribution and the observed result. It tells us how many standard deviations from the mean an observation is, under the assumption of normality. The logic behind using this test will be detailed out further in the next section.

**Data:** To answer my research question, I needed the following information:
1.     Number of Uber pick ups
2.     Number of Taxi pick ups
3.     Number of other For Hire Vehicles pick ups
Since Midtown Manhattan is one of the Central Business Districts of New York City, I realized that it would be a good area to analyze traffic in.
I received data for the months of July, August and September 2014.
I got all this information from FiveThirtyEight’s github account which had this information apart from other numbers such as:
·      Average trips per Hour and day of week (Uber, Lyft and the other FHVs)
·      Uber, Lyft pick ups per day within Manhattan core, LaGuardia airport and JFK (2014)
·      Taxi pick ups per day within Manhattan core, LaGuardia airport and JFK (2013 and 2014)
·      Change in daily Uber and Lyft trips in Manhattan Core (Sept 2014)
·      Change in daily yellow taxi trips in Manhattan Core (September 2013 compared with September 2014)
However, I did not need this information to answer my research question, so I dropped these columns while data wrangling.
Since the document had a lot of unfilled columns and rows, I dropped them all so that I could get a clean dataframe which was good for processing the data quickly.

**Methodology:** To make the dataframe easy to understand, I made a new column in the dataframe which added the total number of FHVs (excluding Uber vehicles, yellow and green taxis) on the streets. Similarly, I combined both the green and yellow taxis.
I also grouped the information given in groups of three (using groupby) – for the months of July, August and September – this gave me a clear look at the traffic patterns through the three months in summer 2014.
I also plotted the total daily trips made by Uber, Taxis and other For Hire Vehicles everyday from July 1, 2014 to September 30, 2014 to look at the trends.



 This plot shows that throughout the months of July, August and September in the year 2014, the number of trips made per day by taxis in Midtown Manhattan was higher than the number of trips made by Uber and other FHVs. This plot also shows that the number of trips made per day by Uber is higher than the number of trips made per day by the other FHVs combined.



A t test would also have given us a similar result, however, I chose to perform the z test because the number of observations (length of the dataframe) was high for a t test.
To perform the z test, I calculated the means of the Uber, taxis and other FHVs pick ups and the standard deviation of the taxi and other FHVs pick ups. I also calculate the length of the dataframe and plugged in all the values in the Z test formula to calculate the Z statistic.

**Conclusion:** According to the Z statistic, the area obtained from the Z Table is over 0.9998. This gives me the p-value of 0.0002 (1 – 0.9998). This implies that the probability that I have falsely rejected the null hypothesis is 0.02% which is much lower than specified for this question (5%). Hence, I will reject the Null Hypothesis and conclude that the average number of Uber pick ups in a day on the streets in Midtown Manhattan is not more than the average number of ‘For Hire Vehicles’ and taxi pick ups in a day on the streets in Midtown Manhattan.

This analysis backs up the conclusions drawn by the transportation study done by the Mayor’s administration and by the analysis done by FiveThirtyEight.

However, with this analysis, I could not conclude that the Uber vehicles were replacing taxis and thus were not responsible for the increased congestion in the city.

Another valuable input that this study fails to show is the contribution of the private vehicles and the other modes of transportation used in Midtown Manhattan – such as buses, bicycles etc.

**Future Work:** Now that we have concluded that Uber is not causing traffic congestion in Midtown Manhattan, we need to explore what is causing it. This can be done by looking at the other modes of transportation, and including them in this study. (However, the transportation study done by the Mayor’s office touches upon the fact that it is caused due to the high number of ongoing construction projects in the city).

Links: <https://github.com/noniemathur/PUI2016_nm2565/tree/master/Extra_Credit>
**Bibliography:**
Data source:
<https://github.com/fivethirtyeight/uber-tlc-foil-response>
News articles and background:
<http://hotair.com/archives/2016/08/04/why-is-bill-de-blasio-fighting-so-hard-against-uber-in-nyc/>
<http://www.theverge.com/2015/12/4/9851000/uber-nyc-bill-de-blasio-report-investigation-cap-tax-cuomo>
<http://www.nbcnewyork.com/news/local/uber-de-blasio-mayor-cap-proposal-new-josh-mohrer-construction-traffic-365500281.html>
<http://fivethirtyeight.com/features/uber-is-taking-millions-of-manhattan-rides-away-from-taxis/>
<http://www.theverge.com/2016/1/15/10774878/uber-nyc-bill-de-blasio-traffic-study-failure>
<http://observer.com/2016/01/bill-de-blasios-quest-to-cap-uber-ends-with-a-whimper/>
<http://fivethirtyeight.com/features/is-uber-making-nyc-rush-hour-traffic-worse/>
<http://data.beta.nyc/dataset/uber-trip-data-foiled-apr-sep-2014>
<http://toddwschneider.com/posts/analyzing-1-1-billion-nyc-taxi-and-uber-trips-with-a-vengeance/>
<http://www.nytimes.com/2016/01/16/nyregion/uber-not-to-blame-for-rise-in-manhattan-traffic-congestion-report-says.html>
[http://www1.nyc.gov/assets/operations/downloads/pdf/For-Hire-Vehicle-Transportation-Study.pdf](http://www.nytimes.com/2016/01/16/nyregion/uber-not-to-blame-for-rise-in-manhattan-traffic-congestion-report-says.html)
<http://money.cnn.com/2016/01/15/technology/uber-traffic-study/>
[http://www.streetsblog.org/2015/10/22/23000-cars-on-nyc-streets-and-no-one-is-tracking-uber-traffic-collisions/](http://money.cnn.com/2016/01/15/technology/uber-traffic-study/)
<http://www.wsj.com/articles/bill-de-blasios-progressive-war-on-uber-1470265626>
For Hire Vehicle Transportation Study:
<http://www1.nyc.gov/assets/operations/downloads/pdf/For-Hire-Vehicle-Transportation-Study.pdf>