

Team # 2 Correlation One DataThon Report

Jeffrey Adams, Satyapriya Krishna, Chaitanya Patel, Guillermo Perez

Question: What are the differences in certain key usage features between Uber and traditional Taxi services in in the Manhattan area of New York City.

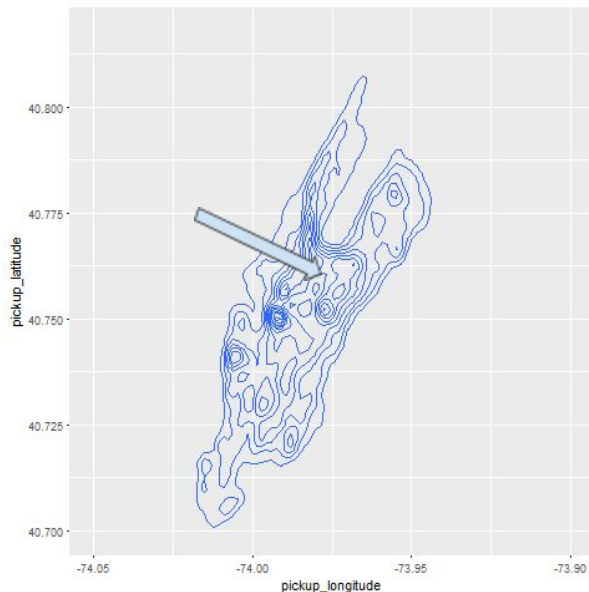
Non-Technical Executive Summary

- Uber rides and traditional taxi services rides pick up locations have different distributions through space.
- There is no difference in the time distribution between Uber and other taxi services. The usage throughout the day is very similar for both types of services
- We find that Uber rides start in locations where Yellow rides earn the least (lower total_amount for the ride). Using the yellow cab data as a proxy we believe that Uber drivers do rides with lower cost than yellow cab drivers.
- We find that neighborhoods with more 25-35 year olds and senior citizens initiated more more Uber rides

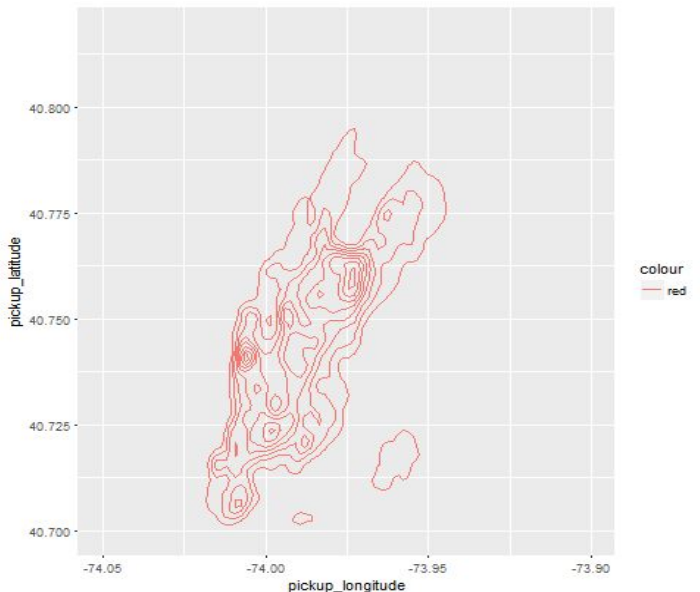
Technical Executive Summary

We restrict our data to rides initiated to the Manhattan area. We analyze the spatial density of trips to understand the areas of dominance by Uber and Yellow cab (these being the top two competitors in the ride industry).

Yellow rides density



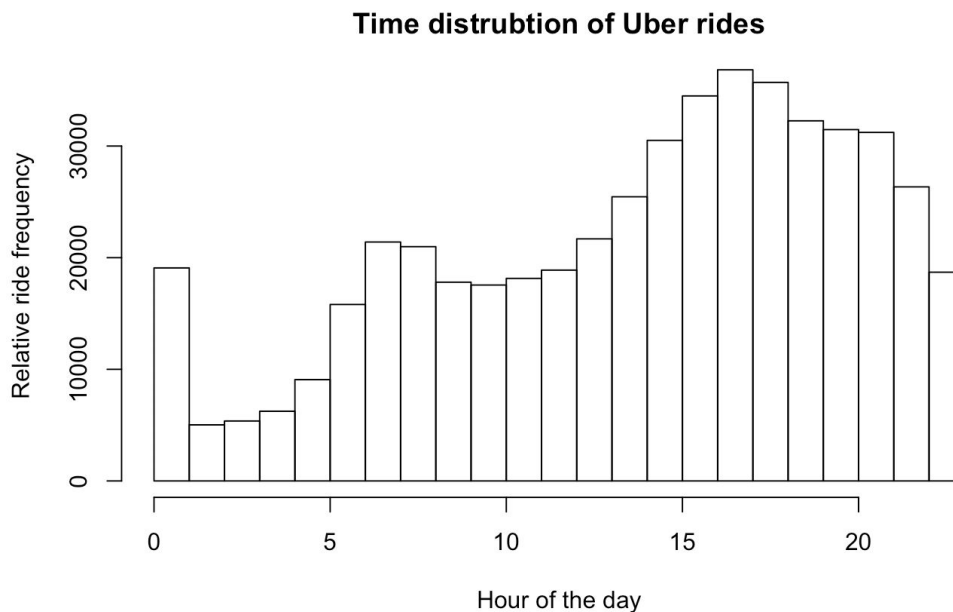
Uber rides density

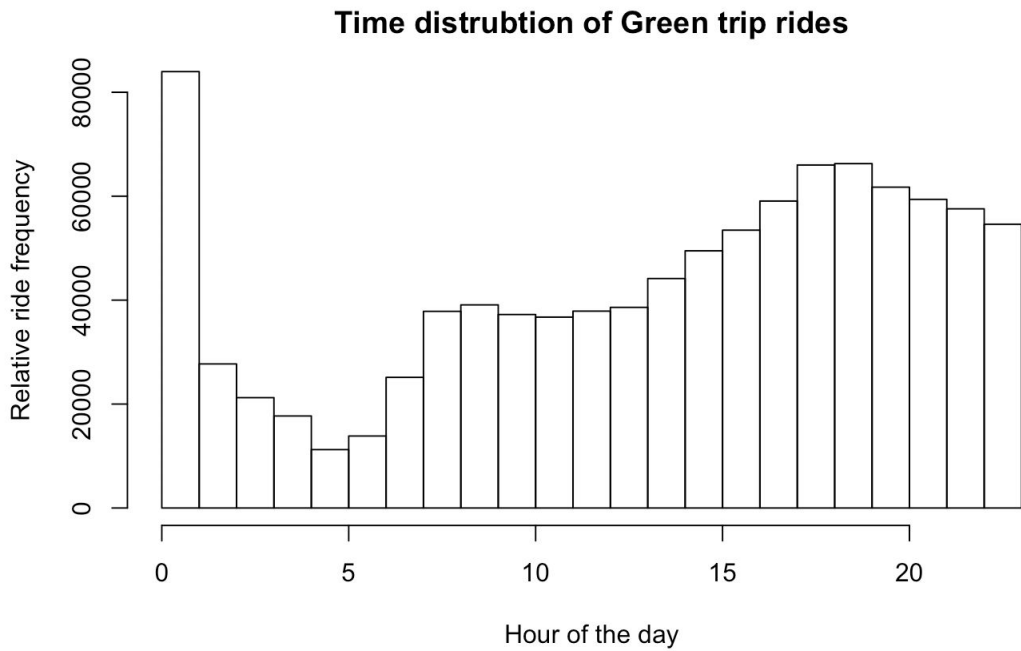


The graph on the left depicts the yellow ride density in the Manhattan area. The right graph depicts the same for the Uber ride density. We see that there are a few places where the Uber density is high and some places where the Yellow cab density is high. As a company, one can make use of such a graph to understand where the company can focus their attention in order to capture more potential/ For example, the yellow cabs could focus their presence in areas where Uber ride density is not as high in order to maximize their chances of matching with passengers.

Usage throughout the day

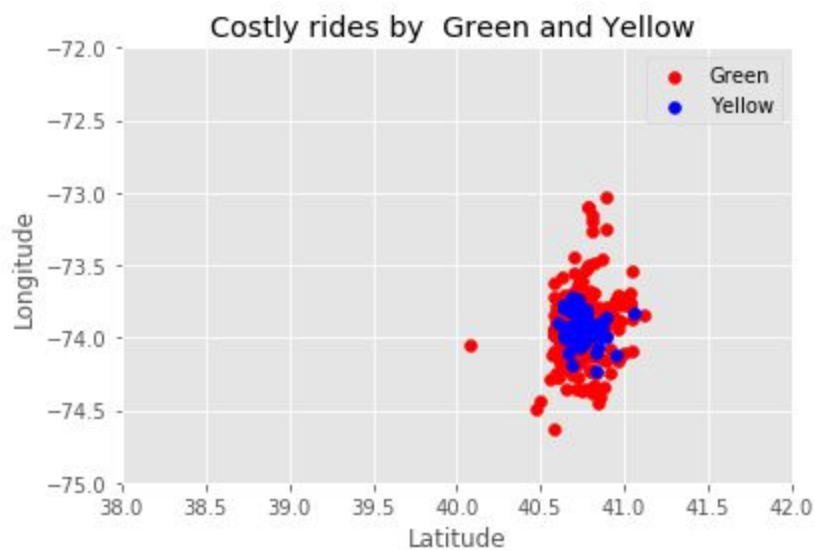
Now, we move our focus to understand the proportional ride density of Yellow, Green, and Uber rides throughout the day. This is to say that we will show what proportions of the respective cab services are used on an hourly basis. The graph below shows that proportion of Uber rides are high from 2PM to 10PM. This means that most of the Uber rides happen from 2PM to 10PM on a typical day. On the other hand, most Green trips are taken from 5AM to 11PM with a small increase after 2PM. There is an unusually high demand for Green trips during midnight for some specific regional reason. This means that the Green trip frequency stay almost constant throughout the day even when the “2PM traffic” comes into the picture. This further supports the claim that Green trips are not as effective in capturing the passengers once the ride demand increases during the afternoon period.



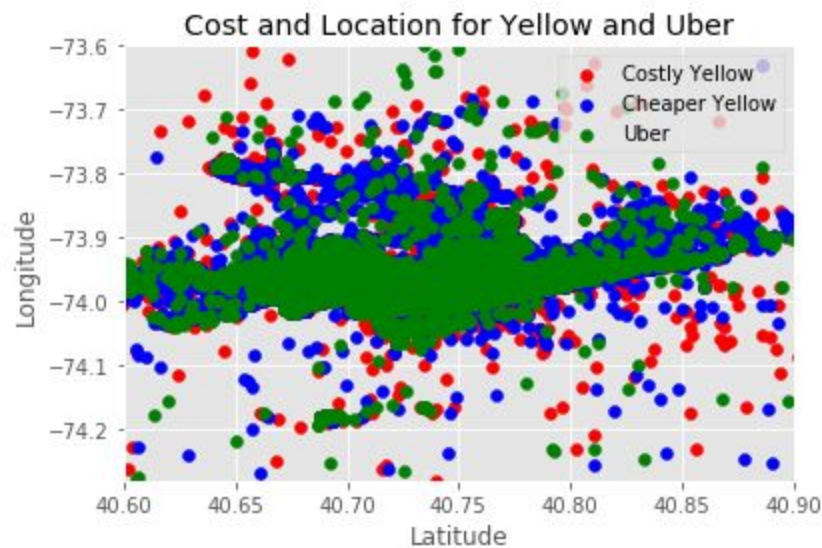


Cost comparison of Uber and Yellow

This graph below shows the rides which are “costly” (cost above the average in the data set) based on the dropoff latitude and longitude where we can see that Green reach out to farther places and hence cost high compared to yellow.



In the description below, we tried to find the cost of Uber rides with the given information.



This is a plot which shows all the rides by yellow and uber based on the pick-up latitude and longitude. The red dots are the rides which cost above the average and the blue signifies the cost below average. This graph is used to infer that uber has more low cost rides compared to Yellow because the density of green is more than yellow in the region with less red.

Demographics

After matching which region each uber pickup took place in, we found that only a small number of the 195 NTA regions have a large number of uber pickups. We compared the age distributions in these NTA regions as compared to all other regions, and found that there was a disproportionately small number of teens and children, and a disproportionately large number of 25-35 year olds and senior citizens. This could indicate that the principal users of Uber in New York are in the 25-35 and 65+ age ranges. A table containing the number of residents in each demographic in the high-use areas, compared with the general demographics for the city as a whole, is reproduced below:

15-19	.061
20-24	.109
25-29	.146
30-34	.137
35-39	.123

40-44	.112
45-49	.105
50-54	.104
55-59	.114
60-64	.125
65+	.132