Investigating Uber Usage in New York by Region

Citadel Datathon

University of Waterloo

May 13, 2017

Nolan Dey

Kevin Wang

Kevin Zhang

Christopher Zhu

Introduction

Over the last few years, Uber and other ride-sharing services have risen in popularity and disrupted the for-hire vehicle (FHV) industry. This is especially apparent in New York where taxis are popular due to the dense concentration of people and cars. Yet, with New York's classic yellow cabs still extremely relevant, Uber needs to adapt to achieve a greater market share.

Topic Question

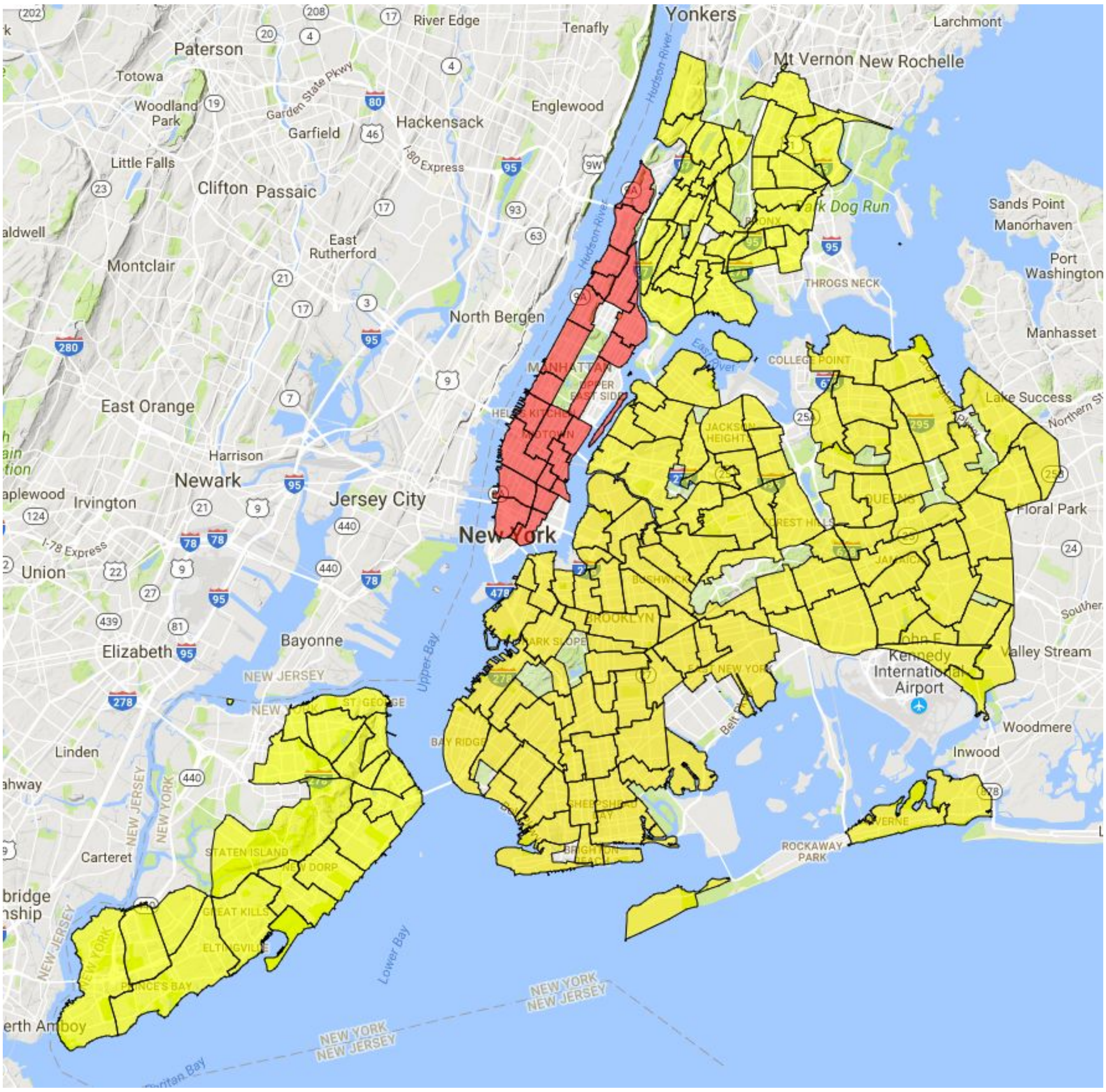
The main focus of this report was to visualize the Uber ridership statistics in New York separated by neighbourhood tabulation areas (NTAs). A further, more analytical goal, of this report was to identify why some areas have low ridership, and to suggest methods of how Uber can increase their appeal to these areas in order to increase their revenue.

In order to determine the number of Uber cars that were requested in 2015 with respect to NTA_Code, the "*uber_trips_2015*" and "*zones*' datasets were joined on the column 'pickup_location_id' in "*uber_trips_2015"* and "location_id" in "*zones*" .

One of the ways that uber can achieve a greater market share is by expanding their service within New York to specific regions. If Uber can identify regions which have a high potential for growth, then the company can maximize the effectiveness of their campaigns. The question that this report wants to answer is "How popular is Uber in each region of New York?"

Non-Technical Executive Summary

Figure 1: Annual Uber Rides / Person



In the heatmap above, the more red a region, the more annual Uber rides/person in 2015. The more yellow a region, the less annual Uber rides/person in 2015. The subdivisions in this heat map denote each NTA region. While usage data by NTA was collected, we found that it was more strongly grouped by borough so we focussed our analysis on boroughs.

Table 1. Summary of Annual Uber Rides by Borough

Borough	Annual Uber Rides/Person (per Borough)
Queens	0.60
Bronx	0.16
Manhattan	6.55
Staten Island	0.01
Brooklyn	0.93

Using the heat map we generated, in addition to Table 1, we were able to gain some insights into Uber usage in each borough. We found that Manhattan has very high usage when compared to the other boroughs. This makes intuitive sense since Manhattan is the most populous, dense, and affluent borough in New York. It is also intuitive that Staten Island has very low Uber usage because it is a suburban borough. The Bronx, Queens, and Brooklyn all have moderate Uber use.

After examining the Uber usage data in each New York borough, we found that Manhattan was the most popular, followed by Brooklyn, Queen, Bronx, and Staten Island (in decreasing order). This analysis lays the foundation for this usage data to be compared to taxi usage data to get a sense of which boroughs have potential for growth. If Uber usage is lower than taxi usage in any region, it is a region which Uber can focus their business development efforts on. Unfortunately we did not have enough time to complete this analysis in the time allotted but the work we completed lays the groundwork for further, richer analysis to be completed.

Technical Executive Summary

In order to produce the heatmap, the annual Uber ride per NTA_Code was determined by the process described previously. First, the necessary datasets were imported to a python dataframe using Pandas. Then, the "*uber_trips_2015*" dataset was joined to the "*zones*" dataset on the common "*pickup_location_id*" field. Finally, this table was aggregated on the borough and NTA_Code fields with a count on the "zone" field. The "*demographics*" dataset was also used in order to determine the population of each NTA region in order to divide the amount of Uber rides per NTA by its corresponding population. This is to make sure that the data is not skewed by the population of each NTA, i.e. more populated NTA should naturally have more Uber rides, but Uber rides per population should be similar.

Table 2. Summary of Annual Uber Rides by NTA_Code (First 6 Rows)

Borough	NTA_Code	Annual Uber Rides/Population
Bronx	BX01	0.138168479
Bronx	BX03	0.088970652
Bronx	BX05	0.120904162
Bronx	BX06	0.334319527
Bronx	BX07	0.11342788
Bronx	BX08	0.102967639

The heatmap itself was drawn using Google Maps Javascript APIs for Polygon Arrays integrated inside of an HTML webpage. Using the *"geographic"* dataset, all the vertices for each NTA was parsed out into JSON Arrays before being combined with the derived metric data from the other datasets. This combined data was used to draw each NTA with corresponding colors from the heatmap's color scale. The heatmap color scale is a 64-step linear gradient corresponding to the minimum metric and maximum metric values (figure 2).

Figure 2. 64 step gradient used in heat map

1	#F2FF00	#F2FF00	#F2FF00	1
2	#F2FA00	#F2FA00	#F2FA00	2
3	#F2F600	#F2F600	#F2F600	3
4	#F2F200	#F2F200	#F2F200	4
5	#F2EE00	#F2EE00	#F2EE00	5
6	#F3EA00	#F3EA00	#F3EA00	6
7	#F3E600	#F3E600	#F3E600	7
8	#F3E200	#F3E200	#F3E200	8
9	#F3DE00	#F3DE00	#F3DE00	9
10	#F3DA00	#F3DA00	#F3DA00	10
11	#F4D600	#F4D600	#F4D600	11
12	#F4D200	#F4D200	#F4D200	12
13	#F4CE00	#F4CE00	#F4CE00	13
14	#F4CA00	#F4CA00	#F4CA00	14
15	#F4C600	#F4C600	#F4C600	15
16	#F5C200	#F5C200	#F5C200	16
17	#F5BE00	#F5BE00	#F5BE00	17
18	#F5BA00	#F5BA00	#F5BA00	18
19	#F5B600	#F5B600	#F5B600	19
20	#F5B200	#F5B200	#F5B200	20
21	#F6AE00	#F6AE00	#F6AE00	21
22	#F6AA00	#F6AA00	#F6AA00	22
23	#F6A500	#F6A500	#F6A500	23
24	#F6A100	#F6A100	#F6A100	24
25	#F69D00	#F69D00	#F69D00	25
26	#F79900	#F79900	#F79900	26
27	#F79500	#F79500	#F79500	27
28	#F79100	#F79100	#F79100	28
29	#F78D00	#F78D00	#F78D00	29
30	#F78900	#F78900	#F78900	30
31	#F88500	#F88500	#F88500	31
32	#F88100	#F88100	#F88100	32
33	#F87D00	#F87D00	#F87D00	33
34	#F87900	#F87900	#F87900	34
35	#F97500	#F97500	#F97500	35
36	#F97100	#F97100	#F97100	36
37	#F96D00	#F96D00	#F96D00	37
38	#F96900	#F96900	#F96900	38
39	#F96500	#F96500	#F96500	39
40	#FA6100	#FA6100	#FA6100	40
41	#FA5D00	#FA5D00	#FA5D00	41
42	#FA5900	#FA5900	#FA5900	42
43	#FA5500	#FA5500	#FA5500	43
44	#FA5000	#FA5000	#FA5000	44
45	#FB4C00	#FB4C00	#FB4C00	45
46	#FB4800	#FB4800	#FB4800	46
47	#FB4400	#FB4400	#FB4400	47
48	#FB4000	#FB4000	#FB4000	48
49	#FB3C00	#FB3C00	#FB3C00	49
50	#FC3800	#FC3800	#FC3800	50
51	#FC3400	#FC3400	#FC3400	51
52	#FC3000	#FC3000	#FC3000	52
53	#FC2C00	#FC2C00	#FC2C00	53
54	#FC2800	#FC2800	#FC2800	54
55	#FD2400	#FD2400	#FD2400	55
56	#FD2000	#FD2000	#FD2000	56
57	#FD1C00	#FD1C00	#FD1C00	57
58	#FD1800	#FD1800	#FD1800	58
59	#FD1400	#FD1400	#FD1400	59
60	#FE1000	#FE1000	#FE1000	60
61	#FE0C00	#FE0C00	#FE0C00	61
62	#FE0800	#FE0800	#FE0800	62
63	#FE0400	#FE0400	#FE0400	63
64	#FF0000	#FF0000	#FF0000	64