

This is my report of the **FullStory** Data Science Challenge.

## Goal

How someone would approach maximizing his/her income as a taxi driver if he/she decides to drive a taxi for 10 hours each week to earn a little extra money.

## Process

### 1. Understanding the Data

The given data contained information about trips taken by Yellow taxis in June 2017.

There were 17 variables: VendorID, tpep\_pickup\_datetime, tpep\_dropoff\_datetime, passenger\_count, trip\_distance, RatecodeID, store\_and\_fwd\_flag, PULocationID, DOLocationID, payment\_type, fare\_amount, extra, mta\_tax, tip\_amount, tolls\_amount, improvement\_surcharge, total\_amount.

For the analysis, I decided to use the 'total\_amount' column as a response variable.

### 2. Data Preprocessing

- **Checking Missing Values**

Fortunately, there was no missing value.

- **Delete Outliers**

There was a large amount of outliers in the data.

- a) There were negative **total\_amount** values. Since these can be considered as refunds, they have been removed from the data.
- b) The **tpep\_pickup\_datetime** and **tpep\_dropoff\_datetime** variables indicated when a taxi meter was engaged and disengaged for a particular ride. In other words, the difference between the two variables tells us a travel time of a particular ride. Some samples showed a zero difference between the two variables, even though the actual trip distance was displayed. In addition, despite taxis traveled short distances, the travel times of certain samples showed significant differences, for example, a day. All of these samples were considered outliers and removed.
- c) **PULocationID** indicated pick-up locations. Some values of this feature do not provide location information. They have been deleted as outliers.
- d) **trip\_distance** represents a mileage of a particular ride. Some samples had positive **fare\_amount** values even though mileages were zero. These samples were also removed.
- e) The purpose of this test is to optimize the extra income while working as a taxi driver for about 10 hours a week. As a result, it is virtually impossible to provide a taxi service with mileage greater than 200 miles. (It may be possible if someone spends 10 hours a day to work as a taxi driver, but it is also very difficult to meet customers who require to travel long distances.)

- **Manipulate Target Variable**

Since **tolls\_amount** has no effect on income, I subtracted **tolls\_amount** from **total\_amount**.

- **Create new features**

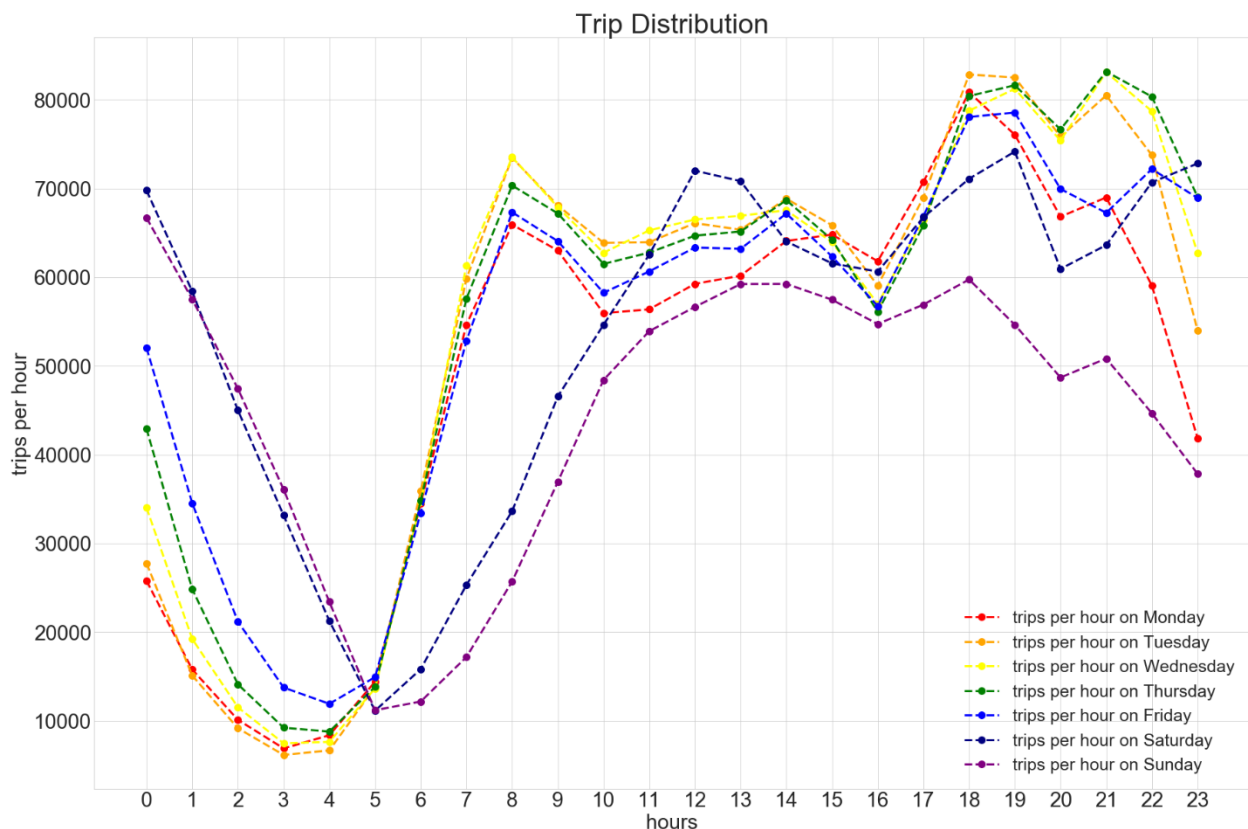
I created several new variables.

- a) **time\_difference** : differences between **tpep\_pickup\_datetime** and **tpep\_dropoff\_datetime** variables show travel times of taxi rides.
- b) **pu\_dates** : converted date-time value shows when a passenger was picked up.
- c) **do\_dates** : converted date-time value shows when a passenger was dropped off.
- d) **day\_of\_week** : indicate a day a passenger was picked up.

### 3. Analyzing the Data

I assumed that demands for taxis would be different hourly, daily, and weekly.

The below graph shows the difference in hourly and daily taxi demand.

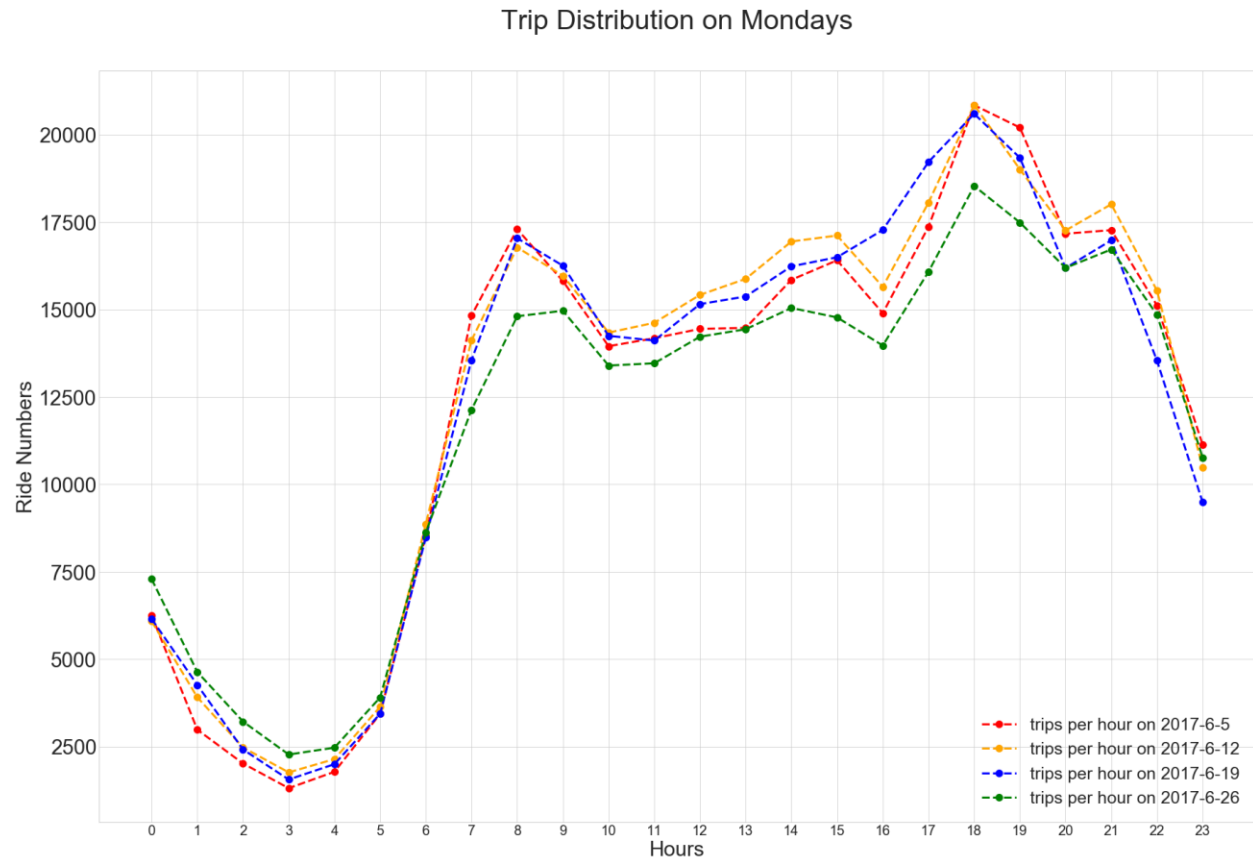


For reference, the month of June, 2017, consists of five Thursdays and Fridays, and four Mondays, Tuesdays, Wednesdays, Saturday and Sundays. So I only used the second, third, fourth and fifth weeks of June to draw the graph to ensure that all days were included in the data four times evenly. That is, the values of 6/1/2017 (Thursday) and 6/2/2017 (Friday) were not included in this graph. However, these values were used when performing other analyses.

It can be seen that demands of taxis differ by each day of a week and by time of a day. For example, on Saturdays and Sundays, the demand for taxis between 0:00 and 3:00 am is much higher than on other days. Also, there are a large number of taxi customers on weekdays between 7 am and 9 am when they go to work, and between 6 pm and 8 pm when people leave their offices.

### 3.1. Data Analysis by each Day of Week

#### 1) Monday



This graph shows the number of customers picked up by taxis on every Monday hourly. A similar pattern can be found on each Monday in June. For example, the demand for taxi would increase sharply during commute time. More specifically, demand increases between 8:00 am and 9:00 am, and between 6:00 pm and 8:00 pm.

So, let's look at the places where the demand for taxi was the greatest at these times.

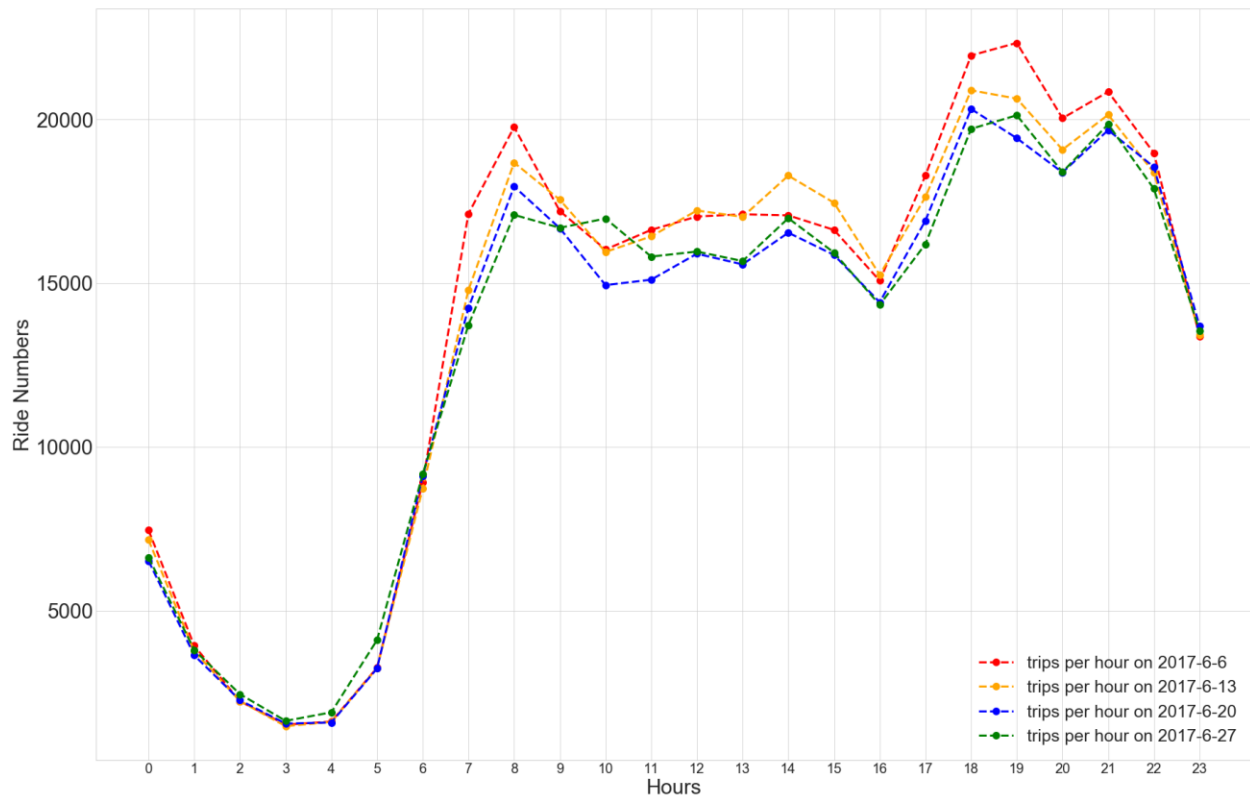
8am - 10 am				6pm - 8pm			
PULocationID	Count	Borough	Zone	PULocationID	Count	Borough	Zone
236	6562	Manhattan	Upper East Side North	161	8166	Manhattan	Midtown Center
237	6446	Manhattan	Upper East Side South	162	7471	Manhattan	Midtown East
186	6046	Manhattan	Penn Station/Madison Sq West	237	7464	Manhattan	Upper East Side South
170	5355	Manhattan	Murray Hill	234	6394	Manhattan	Union Sq
138	5166	Queens	LaGuardia Airport	230	6142	Manhattan	Times Sq/Theatre District
162	4503	Manhattan	Midtown East	236	6087	Manhattan	Upper East Side North
161	3979	Manhattan	Midtown Center	163	5698	Manhattan	Midtown North
48	3889	Manhattan	Clinton East	170	5658	Manhattan	Murray Hill
141	3541	Manhattan	Lenox Hill West	186	4931	Manhattan	Penn Station/Madison Sq West
142	3361	Manhattan	Lincoln Square East	142	4851	Manhattan	Lincoln Square East
140	3292	Manhattan	Lenox Hill East	138	4286	Queens	LaGuardia Airport
107	3259	Manhattan	Gramercy	164	4084	Manhattan	Midtown South
229	3193	Manhattan	Sutton Place/Turtle Bay North	48	4037	Manhattan	Clinton East
239	3106	Manhattan	Upper West Side South	239	3992	Manhattan	Upper West Side South
100	3096	Manhattan	Garment District	141	3483	Manhattan	Lenox Hill West
234	2946	Manhattan	Union Sq	132	3390	Queens	JFK Airport
238	2908	Manhattan	Upper West Side North	107	3384	Manhattan	Gramercy
68	2788	Manhattan	East Chelsea	229	3332	Manhattan	Sutton Place/Turtle Bay North
230	2725	Manhattan	Times Sq/Theatre District	68	3214	Manhattan	East Chelsea
164	2636	Manhattan	Midtown South	238	3119	Manhattan	Upper West Side North
163	2559	Manhattan	Midtown North	100	3083	Manhattan	Garment District
263	2549	Manhattan	Yorkville West	231	2997	Manhattan	TriBeCa/Civic Center
79	2541	Manhattan	East Village	90	2959	Manhattan	Flatiron
231	2245	Manhattan	TriBeCa/Civic Center	140	2927	Manhattan	Lenox Hill East
90	2232	Manhattan	Flatiron	113	2824	Manhattan	Greenwich Village North
10pm - 12am				3am - 5am			
PULocationID	Count	Borough	Zone	PULocationID	Count	Borough	Zone
230	5537	Manhattan	Times Sq/Theatre District	48	1127	Manhattan	Clinton East
161	5286	Manhattan	Midtown Center	79	926	Manhattan	East Village
138	5166	Queens	LaGuardia Airport	230	771	Manhattan	Times Sq/Theatre District
132	4310	Queens	JFK Airport	68	711	Manhattan	East Chelsea
162	3908	Manhattan	Midtown East	249	617	Manhattan	West Village
234	3787	Manhattan	Union Sq	186	587	Manhattan	Penn Station/Madison Sq West
79	3484	Manhattan	East Village	148	433	Manhattan	Lower East Side
142	3336	Manhattan	Lincoln Square East	141	413	Manhattan	Lenox Hill West
48	3193	Manhattan	Clinton East	90	381	Manhattan	Flatiron
163	3012	Manhattan	Midtown North	246	325	Manhattan	West Chelsea/Hudson Yards
186	2857	Manhattan	Penn Station/Madison Sq West	164	324	Manhattan	Midtown South
249	2825	Manhattan	West Village	237	322	Manhattan	Upper East Side South
164	2588	Manhattan	Midtown South	107	314	Manhattan	Gramercy
237	2542	Manhattan	Upper East Side South	100	308	Manhattan	Garment District
170	2529	Manhattan	Murray Hill	170	289	Manhattan	Murray Hill
68	2298	Manhattan	East Chelsea	234	262	Manhattan	Union Sq
231	2274	Manhattan	TriBeCa/Civic Center	114	260	Manhattan	Greenwich Village South
107	2238	Manhattan	Gramercy	132	256	Queens	JFK Airport
239	2166	Manhattan	Upper West Side South	142	252	Manhattan	Lincoln Square East
114	2113	Manhattan	Greenwich Village South	263	248	Manhattan	Yorkville West
100	1940	Manhattan	Garment District	163	246	Manhattan	Midtown North
238	1849	Manhattan	Upper West Side North	162	226	Manhattan	Midtown East
148	1804	Manhattan	Lower East Side	161	222	Manhattan	Midtown Center
236	1676	Manhattan	Upper East Side North	50	222	Manhattan	Clinton West
229	1661	Manhattan	Sutton Place/Turtle Bay North	158	218	Manhattan	Meatpacking/West Village West

Above tables show 25 most popular places customers want to take a taxi by hourly. The area with the largest number of guests on Monday morning is the **Upper East** area, and the area with the highest taxi demand in each Monday evening is the **Midtown** area. And it can be seen that the number of taxi trips decreases as the time passes after 8 pm.

Let's see if this pattern is visible on other days.

## 2) Tuesday

Trip Distribution on Tuesdays



8am - 10am			
PULocationID	Count	Borough	Zone
236	7598	Manhattan	Upper East Side North
237	7483	Manhattan	Upper East Side South
170	5581	Manhattan	Murray Hill
186	5558	Manhattan	Penn Station/Madison Sq West
162	5031	Manhattan	Midtown East
161	4637	Manhattan	Midtown Center
138	4367	Queens	LaGuardia Airport
48	4323	Manhattan	Clinton East
141	4064	Manhattan	Lenox Hill West
142	3811	Manhattan	Lincoln Square East

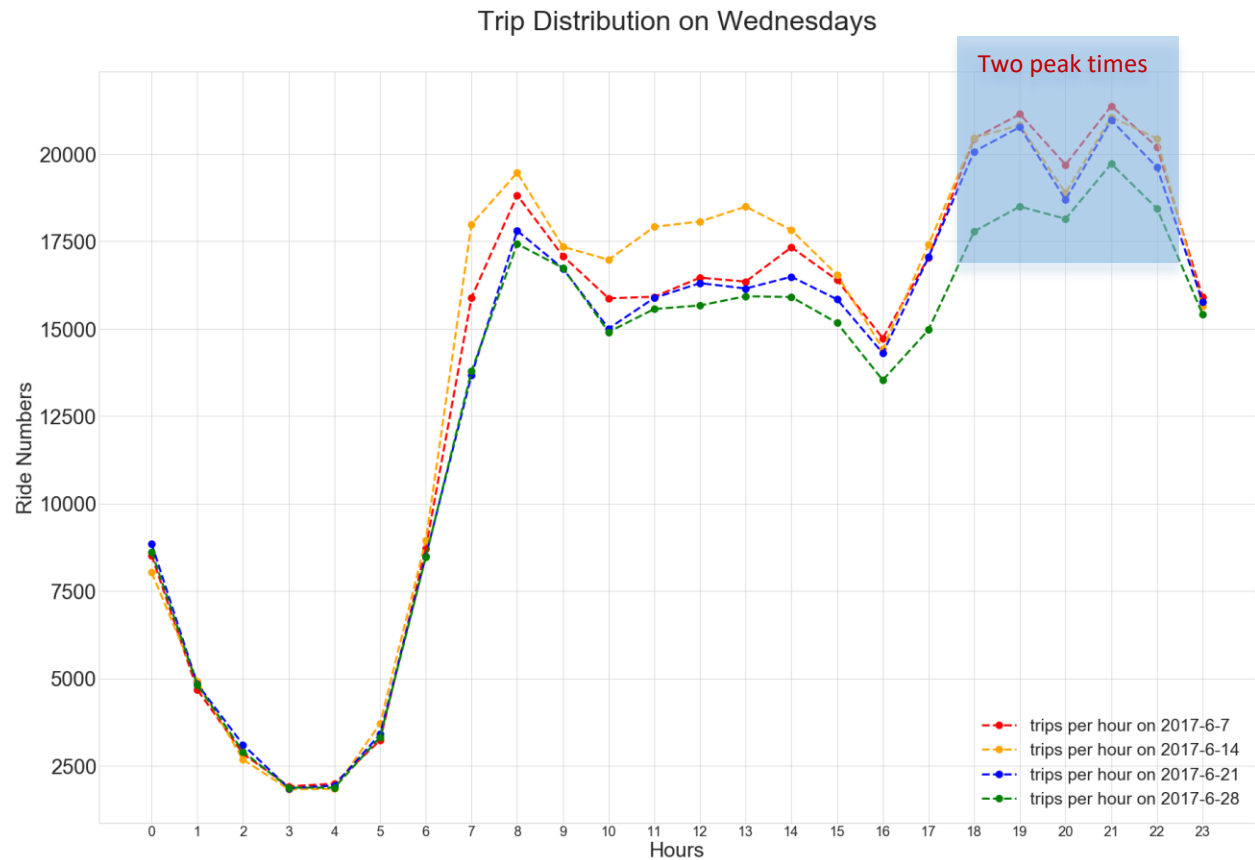
10pm - 12am			
PULocationID	Count	Borough	Zone
161	6220	Manhattan	Midtown Center
230	6156	Manhattan	Times Sq/Theatre District
234	5230	Manhattan	Union Sq
79	5156	Manhattan	East Village
142	4820	Manhattan	Lincoln Square East
48	4812	Manhattan	Clinton East
162	4800	Manhattan	Midtown East
138	4756	Queens	LaGuardia Airport
163	3948	Manhattan	Midtown North
249	3779	Manhattan	West Village

6pm - 8pm			
PULocationID	Count	Borough	Zone
237	8382	Manhattan	Upper East Side South
161	7978	Manhattan	Midtown Center
162	7433	Manhattan	Midtown East
234	6883	Manhattan	Union Sq
236	6586	Manhattan	Upper East Side North
230	6161	Manhattan	Times Sq/Theatre District
163	6044	Manhattan	Midtown North
170	5978	Manhattan	Murray Hill
142	5216	Manhattan	Lincoln Square East
186	4922	Manhattan	Penn Station/Madison Sq West

3am - 5am			
PULocationID	Count	Borough	Zone
48	811	Manhattan	Clinton East
230	772	Manhattan	Times Sq/Theatre District
79	757	Manhattan	East Village
132	620	Queens	JFK Airport
68	608	Manhattan	East Chelsea
249	435	Manhattan	West Village
148	369	Manhattan	Lower East Side
186	361	Manhattan	Penn Station/Madison Sq West
161	329	Manhattan	Midtown Center
164	319	Manhattan	Midtown South

The pattern on Tuesdays is similar to that of Mondays.

### 3) Wednesday



The Wednesdays' feature is different from Mondays or Tuesdays. The graph shows that there were two peak times after 6 pm. Sometimes the demand for a taxi peaks at around 9 o'clock.

8am - 10am			
PULocationID	Count	Borough	Zone
236	7389	Manhattan	Upper East Side North
237	7241	Manhattan	Upper East Side South
186	5741	Manhattan	Penn Station/Madison Sq West
170	5614	Manhattan	Murray Hill
162	4717	Manhattan	Midtown East
48	4460	Manhattan	Clinton East
161	4332	Manhattan	Midtown Center
141	4161	Manhattan	Lenox Hill West
142	3820	Manhattan	Lincoln Square East
138	3787	Queens	LaGuardia Airport

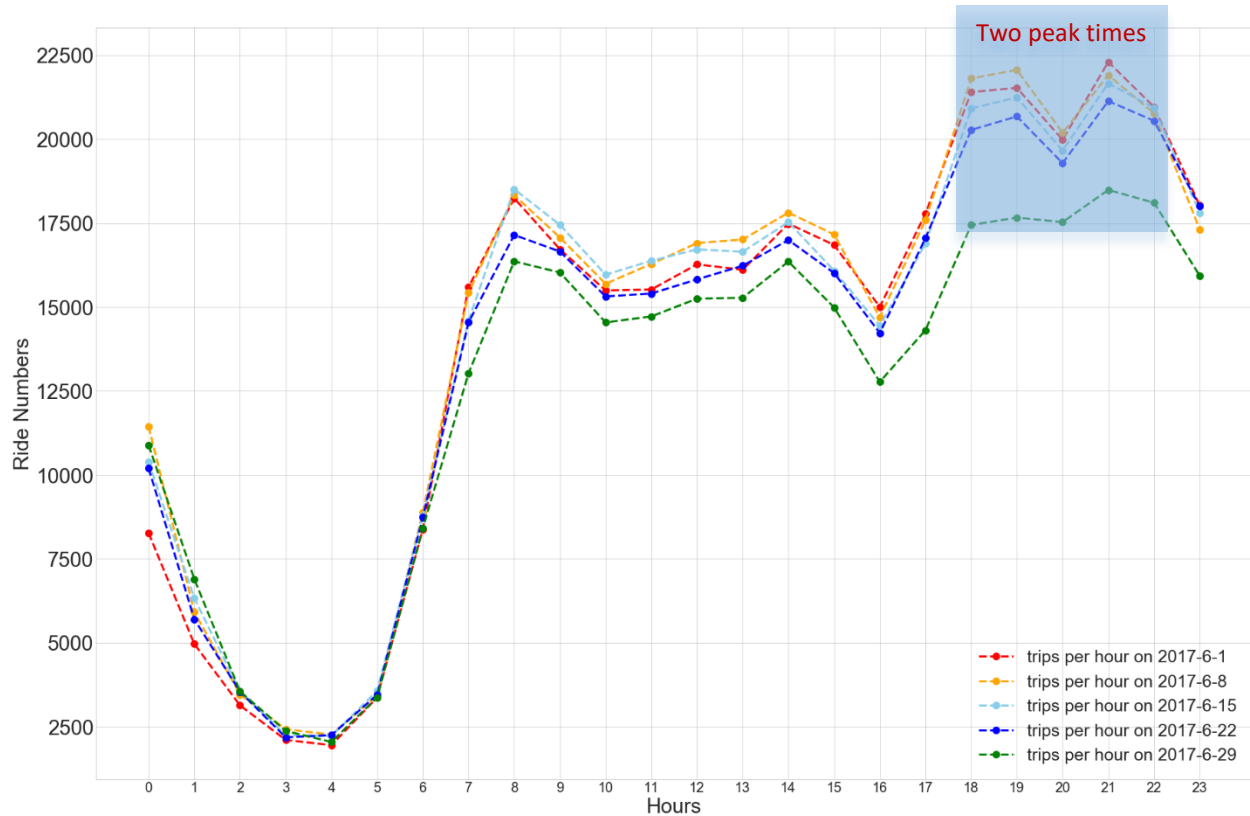
9pm - 11pm			
PULocationID	Count	Borough	Zone
161	7352	Manhattan	Midtown Center
162	6904	Manhattan	Midtown East
230	6690	Manhattan	Times Sq/Theatre District
234	6035	Manhattan	Union Sq
142	5855	Manhattan	Lincoln Square East
79	5823	Manhattan	East Village
237	5809	Manhattan	Upper East Side South
48	5778	Manhattan	Clinton East
170	5242	Manhattan	Murray Hill
163	4843	Manhattan	Midtown North

6pm - 8pm			
PULocationID	Count	Borough	Zone
237	8249	Manhattan	Upper East Side South
161	7461	Manhattan	Midtown Center
162	7178	Manhattan	Midtown East
234	6520	Manhattan	Union Sq
236	6430	Manhattan	Upper East Side North
170	6281	Manhattan	Murray Hill
163	5785	Manhattan	Midtown North
230	5380	Manhattan	Times Sq/Theatre District
142	4923	Manhattan	Lincoln Square East
138	4853	Queens	LaGuardia Airport

3am - 5am			
PULocationID	Count	Borough	Zone
79	946	Manhattan	East Village
230	903	Manhattan	Times Sq/Theatre District
48	881	Manhattan	Clinton East
68	755	Manhattan	East Chelsea
249	684	Manhattan	West Village
148	657	Manhattan	Lower East Side
164	444	Manhattan	Midtown South
158	416	Manhattan	Meatpacking/West Village West
186	389	Manhattan	Penn Station/Madison Sq West
161	373	Manhattan	Midtown Center

#### 4) Thursday

Trip Distribution on Thursdays



8am - 10am			
PULocationID	Count	Borough	Zone
236	9065	Manhattan	Upper East Side North
237	8923	Manhattan	Upper East Side South
170	7014	Manhattan	Murray Hill
186	6950	Manhattan	Penn Station/Madison Sq West
162	6082	Manhattan	Midtown East
161	5582	Manhattan	Midtown Center
48	5228	Manhattan	Clinton East
141	5103	Manhattan	Lenox Hill West
142	4602	Manhattan	Lincoln Square East
107	4487	Manhattan	Gramercy

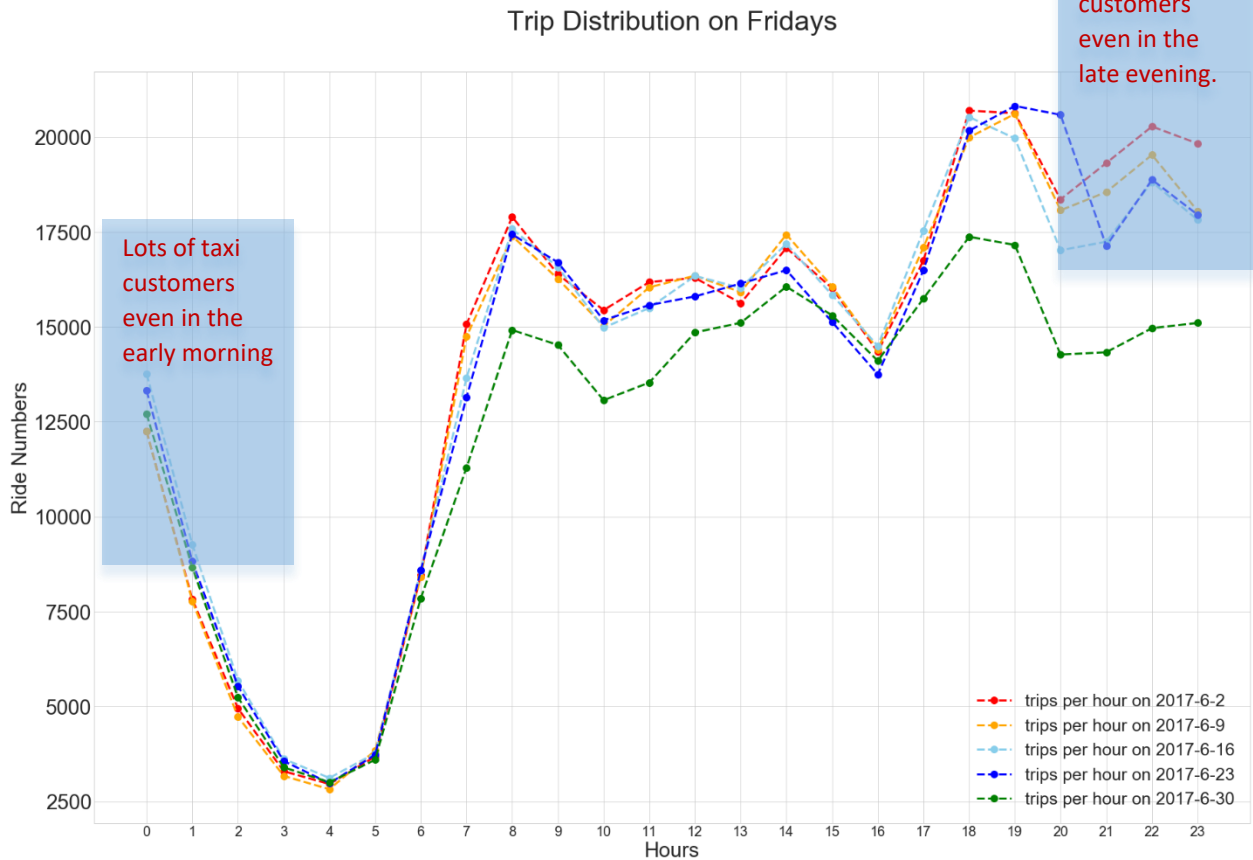
9pm - 11pm			
PULocationID	Count	Borough	Zone
162	8850	Manhattan	Midtown East
161	8705	Manhattan	Midtown Center
142	8194	Manhattan	Lincoln Square East
48	8113	Manhattan	Clinton East
79	7855	Manhattan	East Village
234	7552	Manhattan	Union Sq
230	7173	Manhattan	Times Sq/Theatre District
237	7112	Manhattan	Upper East Side South
170	6956	Manhattan	Murray Hill
138	6179	Queens	LaGuardia Airport

6pm - 8pm			
PULocationID	Count	Borough	Zone
237	9954	Manhattan	Upper East Side South
161	9106	Manhattan	Midtown Center
162	9025	Manhattan	Midtown East
234	8318	Manhattan	Union Sq
236	8045	Manhattan	Upper East Side North
170	7688	Manhattan	Murray Hill
163	7315	Manhattan	Midtown North
230	6902	Manhattan	Times Sq/Theatre District
138	6462	Queens	LaGuardia Airport
142	6182	Manhattan	Lincoln Square East

3am - 5am			
PULocationID	Count	Borough	Zone
79	1278	Manhattan	East Village
48	1245	Manhattan	Clinton East
230	1207	Manhattan	Times Sq/Theatre District
148	1185	Manhattan	Lower East Side
68	881	Manhattan	East Chelsea
246	842	Manhattan	West Chelsea/Hudson Yards
249	826	Manhattan	West Village
164	637	Manhattan	Midtown South
231	633	Manhattan	TriBeCa/Civic Center
186	554	Manhattan	Penn Station/Madison Sq West

As on Wednesdays, we can see that there were usually two peak times in the evening on each Thursday as well.

## 5) Friday



8am - 10am			
PULocationID	Count	Borough	Zone
236	8269	Manhattan	Upper East Side North
237	7657	Manhattan	Upper East Side South
170	6321	Manhattan	Murray Hill
186	6034	Manhattan	Penn Station/Madison Sq West
48	5205	Manhattan	Clinton East
162	5091	Manhattan	Midtown East
141	4960	Manhattan	Lenox Hill West
239	4880	Manhattan	Upper West Side South
238	4599	Manhattan	Upper West Side North
142	4567	Manhattan	Lincoln Square East

9pm - 11pm			
PULocationID	Count	Borough	Zone
79	7852	Manhattan	East Village
230	7470	Manhattan	Times Sq/Theatre District
48	7213	Manhattan	Clinton East
234	6679	Manhattan	Union Sq
142	6568	Manhattan	Lincoln Square East
161	5988	Manhattan	Midtown Center
249	5561	Manhattan	West Village
170	5510	Manhattan	Murray Hill
186	5499	Manhattan	Penn Station/Madison Sq West
162	5121	Manhattan	Midtown East

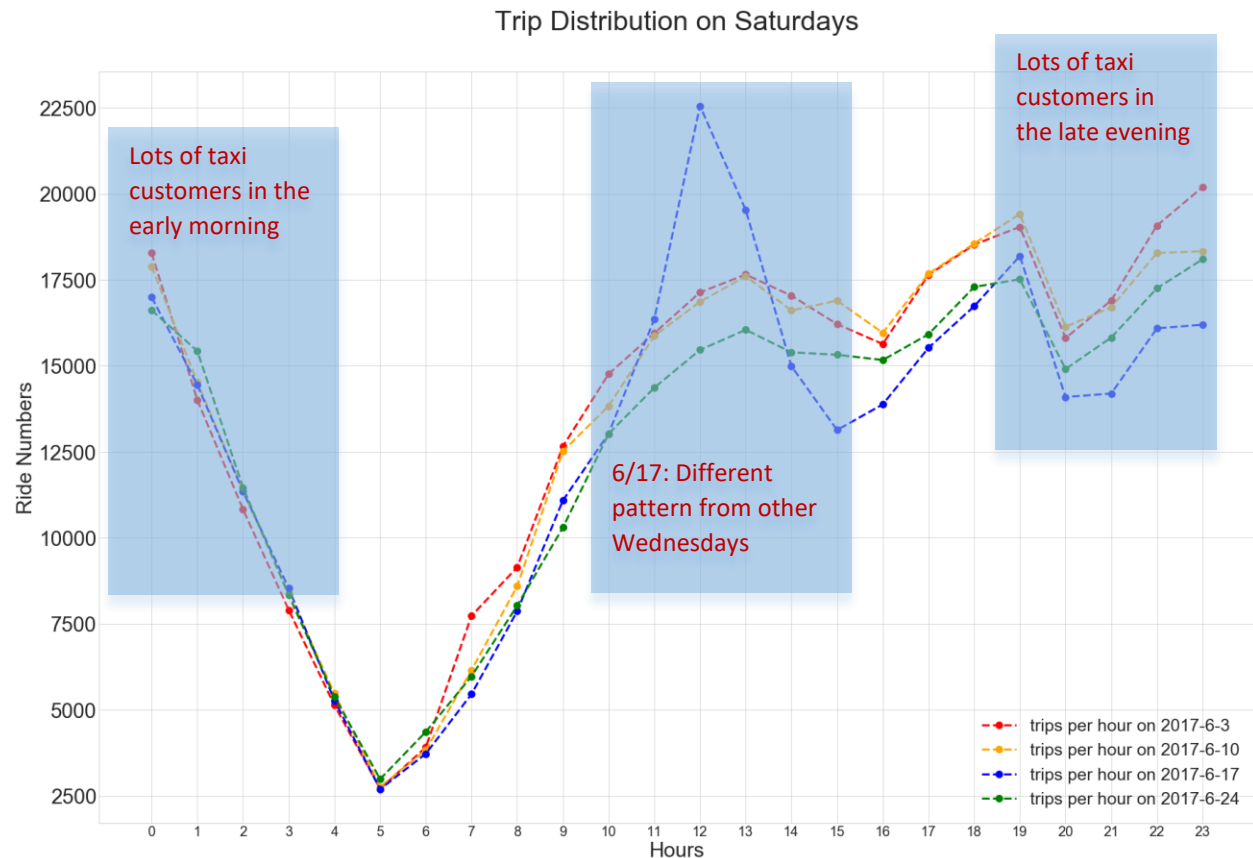
8pm - 9pm			
PULocationID	Count	Borough	Zone
161	9333	Manhattan	Midtown Center
162	8326	Manhattan	Midtown East
237	7876	Manhattan	Upper East Side South
170	7655	Manhattan	Murray Hill
234	7460	Manhattan	Union Sq
186	6827	Manhattan	Penn Station/Madison Sq West
236	6478	Manhattan	Upper East Side North
163	6451	Manhattan	Midtown North
230	6071	Manhattan	Times Sq/Theatre District
142	5772	Manhattan	Lincoln Square East

0am - 2am			
PULocationID	Count	Borough	Zone
79	6418	Manhattan	East Village
148	6018	Manhattan	Lower East Side
230	4748	Manhattan	Times Sq/Theatre District
249	4697	Manhattan	West Village
48	4223	Manhattan	Clinton East
234	4057	Manhattan	Union Sq
114	3525	Manhattan	Greenwich Village South
164	3512	Manhattan	Midtown South
68	3349	Manhattan	East Chelsea
161	3336	Manhattan	Midtown Center

The difference between Fridays and other weekdays is that there were a large number of taxi passengers in the early morning hours (0am-2am) and late evening hours (9pm - 0am).



## 6) Saturday



6pm - 8pm			
PULocationID	Count	Borough	Zone
186	5168	Manhattan	Penn Station/Madison Sq West
161	5135	Manhattan	Midtown Center
237	5113	Manhattan	Upper East Side South
142	5039	Manhattan	Lincoln Square East
236	4558	Manhattan	Upper East Side North
234	4547	Manhattan	Union Sq
48	4411	Manhattan	Clinton East
230	4394	Manhattan	Times Sq/Theatre District
239	4334	Manhattan	Upper West Side South
163	4334	Manhattan	Midtown North

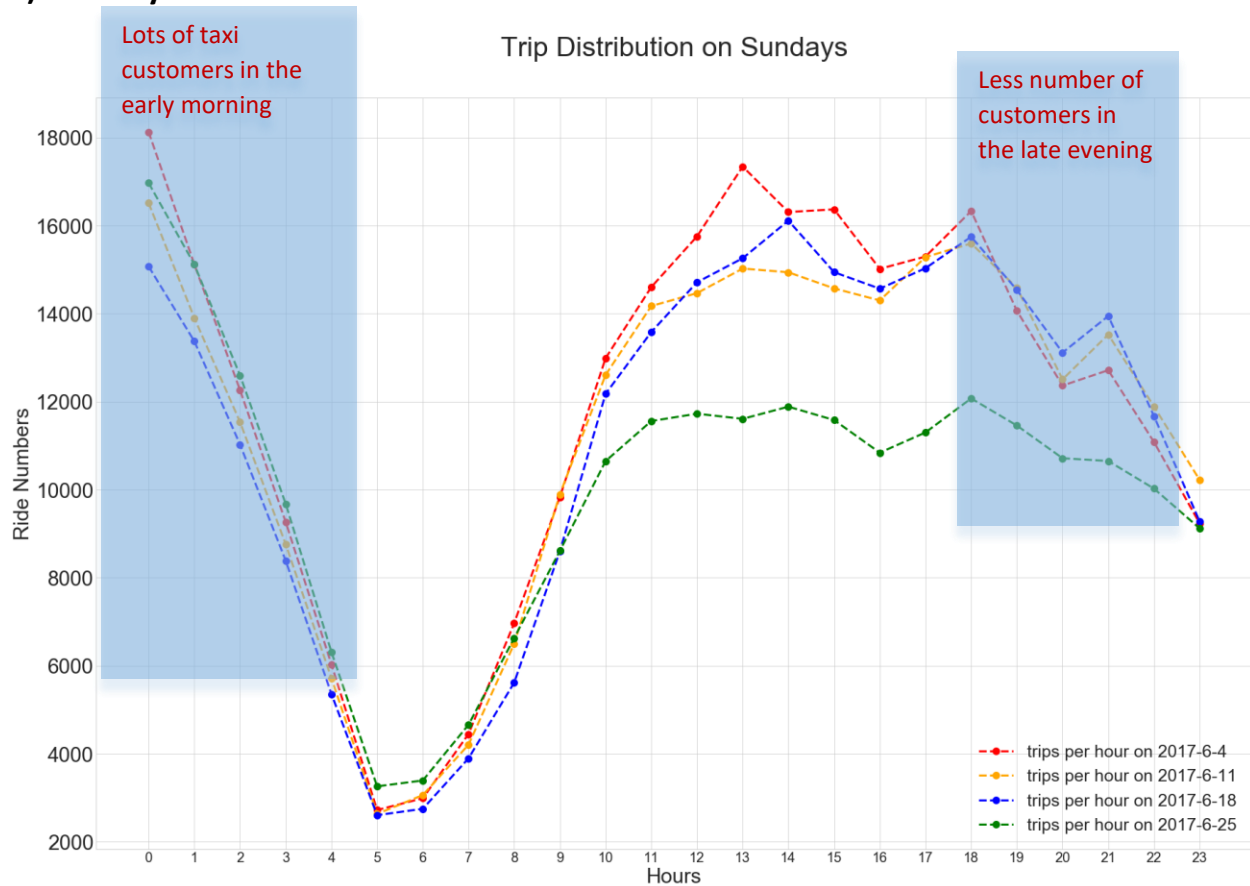
0am - 2am			
PULocationID	Count	Borough	Zone
79	11497	Manhattan	East Village
148	8035	Manhattan	Lower East Side
249	6032	Manhattan	West Village
114	5471	Manhattan	Greenwich Village South
48	5328	Manhattan	Clinton East
234	4521	Manhattan	Union Sq
230	4048	Manhattan	Times Sq/Theatre District
68	3917	Manhattan	East Chelsea
164	3726	Manhattan	Midtown South
158	3722	Manhattan	Meatpacking/West Village West

10pm - 0am			
PULocationID	Count	Borough	Zone
79	7378	Manhattan	East Village
230	6267	Manhattan	Times Sq/Theatre District
48	6114	Manhattan	Clinton East
249	5400	Manhattan	West Village
142	5391	Manhattan	Lincoln Square East
234	5014	Manhattan	Union Sq
186	4428	Manhattan	Penn Station/Madison Sq West
170	4228	Manhattan	Murray Hill
161	4122	Manhattan	Midtown Center
148	4089	Manhattan	Lower East Side

2am - 4am			
PULocationID	Count	Borough	Zone
79	8982	Manhattan	East Village
148	7751	Manhattan	Lower East Side
249	3733	Manhattan	West Village
114	3708	Manhattan	Greenwich Village South
68	3634	Manhattan	East Chelsea
48	3275	Manhattan	Clinton East
158	2585	Manhattan	Meatpacking/West Village West
144	2567	Manhattan	Little Italy/NoLiTa
164	2377	Manhattan	Midtown South
234	2280	Manhattan	Union Sq

Saturdays had a completely different pattern from other weekdays. First, the largest number of taxi trips occurred between 12am and 4 am. To be specific, on each Saturday, the **Lower East** area had the most demand for taxis early in the morning. In the evening, taxi demand also increased as time goes by.

## 7) Sunday



0am - 2am			
PULocationID	Count	Borough	Zone
79	10362	Manhattan	East Village
148	7726	Manhattan	Lower East Side
249	5765	Manhattan	West Village
48	5607	Manhattan	Clinton East
114	4641	Manhattan	Greenwich Village South
68	4203	Manhattan	East Chelsea
230	4127	Manhattan	Times Sq/Theatre District
158	3908	Manhattan	Meatpacking/West Village West
234	3856	Manhattan	Union Sq
164	3609	Manhattan	Midtown South

2am - 4am			
PULocationID	Count	Borough	Zone
79	9636	Manhattan	East Village
148	8223	Manhattan	Lower East Side
249	4041	Manhattan	West Village
68	3862	Manhattan	East Chelsea
48	3686	Manhattan	Clinton East
114	3583	Manhattan	Greenwich Village South
158	3040	Manhattan	Meatpacking/West Village West
144	2613	Manhattan	Little Italy/NoliTa
230	2402	Manhattan	Times Sq/Theatre District
164	2350	Manhattan	Midtown South

1pm - 3pm			
PULocationID	Count	Borough	Zone
236	4486	Manhattan	Upper East Side North
239	4347	Manhattan	Upper West Side South
237	4322	Manhattan	Upper East Side South
186	4254	Manhattan	Penn Station/Madison Sq West
138	4134	Queens	LaGuardia Airport
132	4106	Queens	JFK Airport
142	3970	Manhattan	Lincoln Square East
48	3252	Manhattan	Clinton East
230	3250	Manhattan	Times Sq/Theatre District
234	3248	Manhattan	Union Sq

6pm - 8pm			
PULocationID	Count	Borough	Zone
186	5432	Manhattan	Penn Station/Madison Sq West
138	5033	Queens	LaGuardia Airport
132	4633	Queens	JFK Airport
142	4180	Manhattan	Lincoln Square East
162	4116	Manhattan	Midtown East
237	3946	Manhattan	Upper East Side South
161	3779	Manhattan	Midtown Center
230	3610	Manhattan	Times Sq/Theatre District
170	3603	Manhattan	Murray Hill
48	3581	Manhattan	Clinton East

Like Saturdays, most taxi trips occur on Sundays at Lower East Side areas early in the morning (0am – 4am). But in the evening there were the smallest number of taxi demands compared to other days.

We found the times with the highest taxi demands on taxi by each day. The highest taxi demands occurred in the commute times during weekdays, such as between 8am and 10 am, and between 6pm and 8pm. If so, as taxi demand increases at these times, is it possible to maximize income by driving a taxi around these times?

My answer is **NO!**

Zone	Avg Total Fare	Avg Trip Distance	Avg Trip Time	Avg Fare/mile	Avg Fare/minute	Avg Dist/minute
Mondays (8am - 10am)						
Upper East Side North	12.8274779	2.072715635	12.19866403	6.19	<b>1.05</b>	0.17
Upper East Side South	10.80245113	1.578879926	9.958488468	6.84	<b>1.08</b>	0.16
Penn Station/Madison Sq West	13.34560536	1.808579226	14.39583747	7.38	<b>0.93</b>	0.13
Murray Hill	12.72227824	1.902250233	12.21673514	6.69	<b>1.04</b>	0.16
LaGuardia Airport	44.97425281	10.53727449	45.53974706	4.27	<b>0.99</b>	0.23
Midtown East	13.14890073	2.061867644	12.86570435	6.38	<b>1.02</b>	0.16
Midtown Center	12.96822317	2.011731591	12.67091396	6.45	<b>1.02</b>	0.16
Clinton East	13.16425302	2.008297763	13.34732579	6.55	<b>0.99</b>	0.15
Lenox Hill West	12.4614318	2.026167749	11.53677398	6.15	<b>1.08</b>	0.18
Lincoln Square East	12.80829515	2.018857483	12.36814936	6.34	<b>1.04</b>	0.16
<b>mean</b>	<b>15.92231693</b>	<b>2.802662174</b>	<b>15.70983396</b>	<b>6.324</b>	<b>1.024</b>	<b>0.166</b>
Tuesdays (8am - 10am)						
Upper East Side North	13.67592882	1.905977572	14.68042418	7.18	<b>0.93</b>	0.13
Upper East Side South	11.37797888	1.483533365	11.56385022	7.67	<b>0.98</b>	0.13
Murray Hill	12.56034137	1.620254351	12.98162651	7.75	<b>0.97</b>	0.12
Penn Station/Madison Sq West	14.3268406	1.768627591	16.49474625	8.1	<b>0.87</b>	0.11
Midtown East	12.95562865	1.820233918	13.35478801	7.12	<b>0.97</b>	0.14
Midtown Center	12.81468971	1.734084839	13.09573187	7.39	<b>0.98</b>	0.13
LaGuardia Airport	47.79660461	10.8989539	51.54024823	4.39	<b>0.93</b>	0.21
Clinton East	13.05190196	1.906235294	13.46599673	6.85	<b>0.97</b>	0.14
Lenox Hill West	13.08060039	1.896643701	13.41315617	6.9	<b>0.98</b>	0.14
Lincoln Square East	12.77152929	1.848838133	13.28992056	6.91	<b>0.96</b>	0.14
<b>mean</b>	<b>16.44120443</b>	<b>2.688338266</b>	<b>17.38804887</b>	<b>7.026</b>	<b>0.954</b>	<b>0.139</b>

The above chart is an analysis of taxi trips between **8 am and 10 am on Mondays and Tuesdays** morning. Especially, the tables show 10 areas with the highest number of taxi customers on Mondays and Tuesdays morning.

Driving a taxi at these times in these areas, someone would earn \$15.9 and drive approximately 2.82 miles per trip, and it would take 15.7 minutes per ride on Monday. Similar results are obtained on the Tuesday chart.

The most important part here is how much someone could earn per minute if he/she works as a taxi driver from 8:00 am to 10:00 am on Mondays and Tuesdays.

**According to the charts above, driving a taxi in these areas on Monday will result in \$1.024 per minute and \$0.954 on Tuesday.**

This can also be interpreted in the following manner: on Monday morning, a driver can travel 0.166 miles per minute and earn \$ 6.324 per mile in these areas. So multiplying these two shows that a taxi driver could earn about \$1.05 per minute, which is similar to \$1.024.

Let's see if we can get similar results on Wednesday, Thursday and Friday data.

Zone	Avg Total Fare	Avg Trip Distance	Avg Trip Time	Avg Fare/mile	Avg Fare/minute	Avg Dist/minute
Wednesdays (8am - 10am)						
Upper East Side North	13.47450304	1.920588235	14.11142664	7.02	0.95	0.14
Upper East Side South	11.75089724	1.539311424	11.93577639	7.63	0.98	0.13
Penn Station/Madison Sq West	13.94166005	1.779444444	15.49652778	7.83	0.9	0.11
Murray Hill	12.57983536	1.648460988	12.84009783	7.63	0.98	0.13
Midtown East	13.68985	1.986458333	14.04840278	6.89	0.97	0.14
Clinton East	13.45777778	1.931657407	14.2471142	6.97	0.94	0.14
Midtown Center	13.63577839	1.953021978	13.72828144	6.98	0.99	0.14
Lenox Hill West	13.02653996	1.861276803	13.28594867	7	0.98	0.14
Lincoln Square East	13.73010941	1.930196937	14.80406637	7.11	0.93	0.13
LaGuardia Airport	46.25358946	10.76925357	47.05334797	4.29	0.98	0.23
mean	16.55405407	2.731967012	17.15509901	6.935	0.96	0.143
Thursdays (8am - 10am)						
Upper East Side North	12.94380585	1.996392719	12.75216768	6.48	1.02	0.16
Upper East Side South	11.15625126	1.564770817	10.74100639	7.13	1.04	0.15
Murray Hill	13.25613345	1.850804106	13.36960603	7.16	0.99	0.14
Penn Station/Madison Sq West	14.5763554	1.852814388	16.4449976	7.87	0.89	0.11
Midtown East	13.83617889	2.107602762	13.71881508	6.56	1.01	0.15
Midtown Center	13.83087961	2.066986743	13.70858115	6.69	1.01	0.15
Clinton East	13.6008837	1.980334736	14.00557256	6.87	0.97	0.14
Lenox Hill West	12.95616108	2.000125416	12.65667255	6.48	1.02	0.16
Lincoln Square East	12.91362451	1.962416341	12.87739027	6.58	1	0.15
Gramercy	13.28326053	1.804673501	13.69754104	7.36	0.97	0.13
mean	13.23535343	1.918692153	13.39723503	6.918	0.992	0.144
Fridays (8am - 10am)						
Upper East Side North	13.16755714	2.004944975	13.17673036	6.57	1	0.15
Upper East Side South	11.52485308	1.621721301	11.2189391	7.11	1.03	0.14
Murray Hill	13.0093577	1.789587091	13.17450825	7.27	0.99	0.14
Penn Station/Madison Sq West	14.11261352	1.870449122	15.69334328	7.55	0.9	0.12
Clinton East	13.22055908	1.98348511	13.43599744	6.67	0.98	0.15
Midtown East	14.79577686	2.31429778	14.68283572	6.39	1.01	0.16
Lenox Hill West	13.05022177	2.00984879	12.86326949	6.49	1.01	0.16
Upper West Side South	13.54529713	2.236104508	13.03538251	6.06	1.04	0.17
Upper West Side North	13.64532942	2.317975647	13.14241864	5.89	1.04	0.18
Lincoln Square East	12.8303788	1.978416904	12.61809722	6.49	1.02	0.16
mean	13.29019445	2.012683123	13.3041522	6.649	1.002	0.153

As seen above, regardless of days, places where customers were picked up were generally similar. Also, fares that could be earned per minute were not much different, such as \$ 0.96, \$ .992, and \$ 1.002 on Wednesday, Thursday, and Friday respectively.

Assuming that a taxi driver can pick up customers and drive for two hours without a break (even if this assumption is not realistic), then this taxi can approximately earn up to \$122.88, \$ 114.48, \$115.2, \$119.04, and \$120.24 on Monday, Tuesday, Wednesday, Thursday, and Friday respectively.

So, is it a wise choice to drive a taxi in these areas at these times? Let's look at the following charts.

Zone	Avg Total Fare	Avg Trip Distance	Avg Trip Time	Avg Fare/mile	Avg Fare/minute	Avg Dist/minute
Wednesdays (6pm - 8pm)						
Upper East Side South	13.0103855	1.69874409	12.04254253	7.66	1.08	0.14
Midtown Center	16.1535384	2.145290176	16.73571014	7.53	0.97	0.13
Midtown East	15.5208122	2.144311786	15.3151226	7.24	1.01	0.14
Union Sq	14.33704908	1.85697546	13.94637781	7.72	1.03	0.13
Upper East Side North	12.64911664	1.728384137	11.15123121	7.32	1.13	0.15
Murray Hill	15.45398503	2.12789524	15.11353288	7.26	1.02	0.14
Midtown North	16.15313397	2.006305964	17.21545952	8.05	0.94	0.12
Times Sq/Theatre District	16.62145911	2.177598513	17.41318463	7.63	0.95	0.13
Lincoln Square East	14.16891326	1.827083079	14.00404902	7.75	1.01	0.13
LaGuardia Airport	40.33834123	9.586536163	34.67905076	4.21	1.16	0.28
mean	20.54716652	3.545083792	19.68505536	6.98	1.016	0.16
Wednesdays (9pm - 11pm)						
Midtown Center	14.11014001	2.42177706	11.93731826	5.83	1.18	0.2
Midtown East	14.24174873	2.563672892	11.61353518	5.56	1.23	0.22
Times Sq/Theatre District	14.79873819	2.663796554	12.64345933	5.56	1.17	0.21
Union Sq	14.10026236	2.445845027	12.31273134	5.76	1.15	0.2
Lincoln Square East	14.03925439	2.527178363	12.09496832	5.56	1.16	0.21
East Village	14.54878567	2.612163238	12.76087149	5.57	1.14	0.2
Upper East Side South	12.31130695	2.076792566	9.902867706	5.93	1.24	0.21
Clinton East	14.55799081	2.63973736	12.56661195	5.51	1.16	0.21
Murray Hill	14.09346923	2.471876923	11.71212821	5.7	1.2	0.21
Midtown North	14.84956956	2.619561875	12.90521394	5.67	1.15	0.2
mean	14.07222444	2.484026393	11.96953866	5.676	1.178	0.206
Thursdays (6pm - 8pm)						
Upper East Side South	13.01689472	1.720971469	12.01355736	7.56	1.08	0.14
Midtown Center	15.85923347	2.108152866	16.29787869	7.52	0.97	0.13
Midtown East	15.64954349	2.160299169	15.4598024	7.24	1.01	0.14
Union Sq	14.53806083	1.875245251	14.17409634	7.75	1.03	0.13
Upper East Side North	12.95066252	1.803771287	11.55200746	7.18	1.12	0.16
Murray Hill	15.43492976	2.134631894	15.01185397	7.23	1.03	0.14
Midtown North	15.56771702	1.98806972	16.15809979	7.83	0.96	0.12
Times Sq/Theatre District	16.23674008	2.231525645	16.58001545	7.28	0.98	0.13
LaGuardia Airport	40.92291705	9.542780873	36.78440885	4.29	1.11	0.26
Lincoln Square East	13.79507279	1.888691362	13.13147579	7.3	1.05	0.14
mean	20.39147534	3.557139899	19.53317077	6.786	1.026	0.158
Thursdays (9pm - 11pm)						
Midtown East	14.75396384	2.517733333	12.95993032	5.86	1.14	0.19
Midtown Center	14.78369098	2.506542217	13.40653839	5.9	1.1	0.19
Lincoln Square East	14.18982304	2.472856969	13.00676104	5.74	1.09	0.19
Clinton East	14.96712683	2.625616911	13.6427914	5.7	1.1	0.19
East Village	15.13074475	2.565352005	14.19334182	5.9	1.07	0.18
Union Sq	14.78647643	2.474540519	13.81398526	5.98	1.07	0.18
Times Sq/Theatre District	14.90345183	2.549220689	14.05489103	5.85	1.06	0.18
Upper East Side South	12.744491	2.144492407	10.8558188	5.94	1.17	0.2
Murray Hill	14.60681857	2.486889017	13.0187512	5.87	1.12	0.19
LaGuardia Airport	38.655687	9.721377246	29.31509953	3.98	1.32	0.33
mean	19.13938497	3.875303975	16.21170916	5.524	1.148	0.216

First of all, the areas in the above tables are similar regardless of day and time. In other words, in the evening, Midtown Manhattan had the highest demand for taxis on Wednesday and Thursday evening.

We have seen in the graphs of Wednesdays and Thursdays presented earlier that there were two peak times of taxi demands on each Wednesday and Thursday evening. Thus, the number of taxi rides between 6 pm and 8 pm and the number of taxi trips between 9 pm and 11 pm were not significantly

different on Wednesdays and Thursdays. However, each average income-per-minute seems to be different between the two evening times.

For example, average earning per minute between 6 pm and 8 pm is \$1.016, but a driver can earn \$1.178 per minute if he/she drives a taxi between 9 pm and 11 pm on Wednesday evening. This means a driver can earn up to \$121.92 if the driver works from 6:00 pm to 8:00 pm, but if he/she works from 9:00 to 11:00, the driver can earn up to \$141.36 on Wednesday. Similarly, on Thursday, someone can have up to \$123.12, or earn up to \$ 137.76 depending on when this person works as a taxi driver for two hours.

This means that selecting times with the biggest taxi demand and driving a taxi at the times cannot guarantee the maximum income of a taxi driver. In other words, **it indicates that it is important to find times when taxis are in high demand with low traffic congestion so that a taxi driver can pick up more customers at a given time.**

## 4. Strategies to maximize revenue

To find the best strategies, I selected 55 areas with the highest taxi demands by time and by day. Let me explain why I selected only top 55 locations.

First, just using the top 55 locations, around 95% of all taxi records can be described by the selected information. (For example, on Fridays between 0am – 2am, total 106,653 taxi rides occurred, of which 101,505 trips took place in the top 55 areas, which is 95.17%.)

Secondly, information of areas with a small number of taxi customers at the times could be outliers.

Saturdays (0am - 4am)									
PULocationID	Count	Borough	Zone	Avg Total Fare	Avg Trip Distance	Avg Trip Time	Avg Fare/mile	Avg Fare/minute	Avg Dist/minute
Five areas with high taxi demand									
79	19933	Manhattan	East Village	14.65182059	2.769072569	12.12506692	5.29	1.21	0.23
148	15906	Manhattan	Lower East Side	15.77651909	3.076257384	13.09258506	5.13	1.2	0.23
249	9784	Manhattan	West Village	15.32358074	2.931663074	13.21836738	5.23	1.16	0.22
48	9249	Manhattan	Clinton East	14.67223752	2.710786627	12.6907462	5.41	1.16	0.21
114	8205	Manhattan	Greenwich Village South	15.69474477	3.235201495	12.62350971	4.85	1.24	0.26
Five areas with very low taxi demand									
214	1	Staten Island	South Beach/Dongan Hills	70.8	0.01	6.15	7080	11.51	0
192	3	Queens	Queensboro Hill	16.3	0.95	3.266666667	17.16	4.99	0.29
8	1	Queens	Astoria Park	60.72	3.7	7.183333333	16.41	8.45	0.52
155	2	Brooklyn	Marine Park/Mill Basin	6.8	1.1	3.1	6.18	2.19	0.35
194	3	Manhattan	Randalls Island	67.71	8.775	15.75833333	7.72	4.3	0.56

Let's look at the two highlighted samples. Of the four Saturdays in June, 2017, a single taxi trip took place at **South Beach/Dongan Hills** at **Staten Island** between 0 am and 4 am. This record is almost meaningless compared to around 20,000 taxi demand in Manhattan's East Village over the same period. Also, looking at this record closely, it took more than six minutes to drive just 0.01 miles, and the cost was over \$70. Hence it indicates that this sample is an obvious outlier.

Similarly, the second highlighted sample provides similar information. Only the taxi moved 3.7 miles and \$60.72 was charged to the customer. It could be possible that the overall fare increased since the taxi customer tipped the taxi a large amount, but this case is rare.

Even if a low number of taxi rides occur in a certain area, we cannot assure that taxi records in this location are outliers, but the probability of these piece of information are outliers is relatively high.

Finally, there were usually more than 200 selected areas for each hour and day, and it is impossible to create a table to show information analyzing more than 200 regions at a glance.

Therefore, I used only 55 areas with high taxi demand for analysis.

Saturdays (0am - 4am)							
PULocationID	Count	Borough	Zone	PULocationID	Count	Borough	Zone
55 areas with high taxi demand				55 areas with very low taxi demand			
79	19933	Manhattan	East Village	57	4	Queens	Corona
148	15906	Manhattan	Lower East Side	126	4	Bronx	Hunts Point
249	9784	Manhattan	West Village	35	3	Brooklyn	Brownsville
48	9249	Manhattan	Clinton East	98	3	Queens	Fresh Meadows
114	8205	Manhattan	Greenwich Village South	220	3	Bronx	Spuyten Duyvil/Kingsbridge
68	8034	Manhattan	East Chelsea	67	3	Brooklyn	Dyker Heights
158	6933	Manhattan	Meatpacking/West Village West	32	3	Bronx	Bronxdale
230	6505	Manhattan	Times Sq/Theatre District	47	3	Bronx	Claremont/Bathgate
234	5987	Manhattan	Union Sq	242	3	Bronx	Van Nest/Morris Park
164	5942	Manhattan	Midtown South	20	3	Bronx	Belmont
144	5510	Manhattan	Little Italy/NoLiTa	123	3	Brooklyn	Homecrest
107	5388	Manhattan	Gramercy	102	3	Queens	Glendale
170	4848	Manhattan	Murray Hill	191	3	Queens	Queens Village
246	4583	Manhattan	West Chelsea/Hudson Yards	192	3	Queens	Queensboro Hill
90	4462	Manhattan	Flatiron	178	3	Brooklyn	Ocean Parkway South
186	4446	Manhattan	Penn Station/Madison Sq West	194	3	Manhattan	Randalls Island
163	3803	Manhattan	Midtown North	6	2	Staten Island	Arrochar/Fort Wadsworth
231	3638	Manhattan	TriBeCa/Civic Center	63	2	Brooklyn	Cypress Hills
113	3554	Manhattan	Greenwich Village North	240	2	Bronx	Van Cortlandt Park
100	3198	Manhattan	Garment District	183	2	Bronx	Pelham Bay
263	2939	Manhattan	Yorkville West	254	2	Bronx	Williamsbridge/Olinville
141	2876	Manhattan	Lenox Hill West	81	2	Bronx	Eastchester
137	2830	Manhattan	Kips Bay	53	2	Queens	College Point
50	2809	Manhattan	Clinton West	201	2	Queens	Rockaway Park
142	2744	Manhattan	Lincoln Square East	155	2	Brooklyn	Marine Park/Mill Basin
162	2725	Manhattan	Midtown East	51	2	Bronx	Co-Op City
229	2620	Manhattan	Sutton Place/Turtle Bay North	150	2	Brooklyn	Manhattan Beach
239	2491	Manhattan	Upper West Side South	149	2	Brooklyn	Madison
161	2454	Manhattan	Midtown Center	218	2	Queens	Springfield Gardens North
211	2158	Manhattan	SoHo	139	1	Queens	Laurelton
237	1905	Manhattan	Upper East Side South	21	1	Brooklyn	Bensonhurst East
255	1814	Brooklyn	Williamsburg (North Side)	253	1	Queens	Willets Point
125	1709	Manhattan	Hudson Sq	250	1	Bronx	Westchester Village/Unionport
132	1607	Queens	JFK Airport	171	1	Queens	Murray Hill-Queens
4	1602	Manhattan	Alphabet City	175	1	Queens	Oakland Gardens
238	1506	Manhattan	Upper West Side North	46	1	Bronx	City Island
233	1464	Manhattan	UN/Turtle Bay South	8	1	Queens	Astoria Park
87	1416	Manhattan	Financial District North	214	1	Staten Island	South Beach/Dongan Hills
256	1397	Brooklyn	Williamsburg (South Side)	185	1	Bronx	Pelham Parkway
232	1288	Manhattan	Two Bridges/Seward Park	122	1	Queens	Hollis
236	1223	Manhattan	Upper East Side North	182	1	Bronx	Parkchester
80	959	Brooklyn	East Williamsburg	215	1	Queens	South Jamaica
140	907	Manhattan	Lenox Hill East	121	1	Queens	Hillcrest/Pomonok
74	853	Manhattan	East Harlem North	118	1	Staten Island	Heartland Village/Todt Hill
75	852	Manhattan	East Harlem South	117	1	Queens	Hammels/Arverne
7	839	Queens	Astoria	115	1	Staten Island	Grymes Hill/Clifton
41	830	Manhattan	Central Harlem	11	1	Brooklyn	Bath Beach
181	780	Brooklyn	Park Slope	26	1	Brooklyn	Borough Park
88	775	Manhattan	Financial District South	111	1	Brooklyn	Green-Wood Cemetery
166	750	Manhattan	Morningside Heights	203	1	Queens	Rosedale
112	746	Brooklyn	Greenpoint	94	1	Bronx	Fordham South
261	721	Manhattan	World Trade Center	205	1	Queens	Saint Albans
45	661	Manhattan	Chinatown	207	1	Queens	Saint Michaels Cemetery/Woodside
151	631	Manhattan	Manhattan Valley	210	1	Brooklyn	Sheepshead Bay
143	595	Manhattan	Lincoln Square West	38	1	Queens	Cambria Heights
Total # trips	194384			Total # trips	102		

< Information of the only top 55 regions is enough to represent the entire data. >

## 4.1. Selecting times to work as a driver

I grouped the given data by each day, and then divided each grouped data by time. I calculated and checked all average incomes per minute for each split data. Finally, I found **times with high demands for taxi, but low traffic jam.**



i. 12am – 2am on Friday

Pick-up Zone	Avg Total Fare	Avg Trip Distance	Avg Trip Time	Avg Fare/mile	Avg Fare/minute	Avg Dist/minute
East Village	14.93373144	2.818397686	12.85209994	5.3	1.16	0.22
Lower East Side	16.01869355	3.037583736	13.87869522	5.27	1.15	0.22
Times Sq/Theatre District	15.56911647	3.022168675	13.30332911	5.15	1.17	0.23
West Village	15.31680256	2.861013874	13.42954465	5.35	1.14	0.21
Clinton East	16.07914964	3.197475059	13.48832937	5.03	1.19	0.24
Union Sq	15.14850384	2.867089423	13.04021138	5.28	1.16	0.22
Greenwich Village South	15.41680591	2.849013924	13.54596003	5.41	1.14	0.21
Midtown South	15.48679154	2.956928796	13.15560004	5.24	1.18	0.22
East Chelsea	15.27901949	2.907796102	13.14761119	5.25	1.16	0.22
Midtown Center	14.33000601	2.654433764	11.76270151	5.4	1.22	0.23
Penn Station/Madison Sq West	16.37036445	3.188478261	14.44661125	5.13	1.13	0.22
Meatpacking/West Village West	15.64384337	2.879266206	14.03566963	5.43	1.11	0.21
Murray Hill	14.58570988	2.7693107	11.49394719	5.27	1.27	0.24
Midtown North	14.97682133	2.822077562	12.47555979	5.31	1.2	0.23
JFK Airport	51.50173773	15.64424579	26.59823241	3.29	1.94	0.59
Gramercy	13.73955478	2.550797522	11.20043877	5.39	1.23	0.23
Midtown East	14.16614015	2.704107627	11.26413086	5.24	1.26	0.24
Little Italy/NoLiTa	15.07718531	2.728605035	13.34433897	5.53	1.13	0.2
TriBeCa/Civic Center	17.7796978	3.484931319	15.57307692	5.1	1.14	0.22
Flatiron	15.14915026	2.906428229	12.8218675	5.21	1.18	0.23
Garment District	16.83077071	3.44227842	14.2859666	4.89	1.18	0.24
Sutton Place/Turtle Bay North	14.06095314	2.829483493	10.82018992	4.97	1.3	0.26
Lincoln Square East	15.28604945	3.063806786	12.44344451	4.99	1.23	0.25
LaGuardia Airport	38.10141853	9.681789137	24.87149095	3.94	1.53	0.39
Upper East Side South	13.31423077	2.549328553	10.57447849	5.22	1.26	0.24
Greenwich Village North	14.91480808	2.800356902	12.81444444	5.33	1.16	0.22
SoHo	15.090468	2.727123193	13.28752007	5.53	1.14	0.21
Lenox Hill West	13.95199436	2.803157153	10.75057552	4.98	1.3	0.26
Kips Bay	13.71027437	2.674043321	10.55993983	5.13	1.3	0.25
Yorkville West	13.98356112	2.867691724	10.48420653	4.88	1.33	0.27
Upper West Side South	14.82270465	2.910481541	11.69415463	5.09	1.27	0.25
West Chelsea/Hudson Yards	15.93666358	3.080203892	13.1353877	5.17	1.21	0.23
Hudson Sq	15.61689293	2.843059273	13.68334927	5.49	1.14	0.21
UN/Turtle Bay South	14.75337411	2.905178389	11.44904859	5.08	1.29	0.25
Upper West Side North	14.20942589	2.796169102	11.74485038	5.08	1.21	0.24
Upper East Side North	12.80524123	2.427631579	9.158205409	5.27	1.4	0.27
Financial District North	22.3331407	4.701457286	18.60538107	4.75	1.2	0.25
Clinton West	14.81689873	2.902962025	12.19067511	5.1	1.22	0.24
Lenox Hill East	15.36886179	3.177772358	11.91422764	4.84	1.29	0.27
Williamsburg (North Side)	16.27202147	3.287191413	13.746452	4.95	1.18	0.24
Financial District South	23.44928429	5.082564612	19.30649437	4.61	1.21	0.26
Central Park	13.82746888	2.63120332	10.85674274	5.26	1.27	0.24
World Trade Center	22.81577495	5.06059448	19.57600849	4.51	1.17	0.26
Battery Park City	21.42437367	4.61403397	16.54278132	4.64	1.3	0.28
Alphabet City	16.06391209	3.131538462	13.75164835	5.13	1.17	0.23
Two Bridges/Seward Park	17.56551282	3.448538462	15.08901709	5.09	1.16	0.23
Lincoln Square West	14.01289063	2.81390625	11.79270833	4.98	1.19	0.24
East Harlem South	12.26687166	2.58473262	8.501336898	4.75	1.44	0.3
East Harlem North	12.42659517	2.627882038	9.315951743	4.73	1.33	0.28
Williamsburg (South Side)	14.48536986	2.807890411	11.79296804	5.16	1.23	0.24
Chinatown	17.64945087	3.50234104	15.96806358	5.04	1.11	0.22
Manhattan Valley	13.92162242	2.777522124	11.4280236	5.01	1.22	0.24
Central Harlem	12.95109034	2.566573209	10.41048806	5.05	1.24	0.25
Morningside Heights	13.42435737	2.598746082	10.79817137	5.17	1.24	0.24
Park Slope	17.00093248	3.46244373	14.21275456	4.91	1.2	0.24
mean	16.58243794	3.400578648	13.31663824	5.06	1.234181818	0.247636364

< Information of the top 55 regions on Fridays between 0am and 2am. >

This table is an analysis of the taxi service records from 0:00 am to 2:00 am on Fridays. I selected the 55 high taxi demand regions and computed the relevant values. Most areas are located in Manhattan. As seen above, the average income-per-minute is \$ 1.234. Hence someone can earn up to \$148.1 if the driver works for two hours on Friday between 0 am and 2 am.

ii. 12am – 4am on Saturday

Pick-upZone	Avg Total Fare	Avg Trip Distance	Avg Trip Time	Avg Fare/mile	Avg Fare/minute	Avg Dist/minute
East Village	14.65182059	2.769072569	12.12506692	5.29	1.21	0.23
Lower East Side	15.77651909	3.076257384	13.09258506	5.13	1.2	0.23
West Village	15.32358074	2.931663074	13.21836738	5.23	1.16	0.22
Greenwich Village South	14.67223752	2.710786627	12.6907462	5.41	1.16	0.21
Clinton East	15.69474477	3.235201495	12.62350971	4.85	1.24	0.26
East Chelsea	14.68049302	2.827743522	12.39640753	5.19	1.18	0.23
Union Sq	14.36210511	2.748953351	12.1653356	5.22	1.18	0.23
Meatpacking/West Village West	15.15923211	2.795149444	13.5945151	5.42	1.12	0.21
Midtown South	14.72654946	2.864211305	12.18626082	5.14	1.21	0.24
Times Sq/Theatre District	15.07698292	3.098573561	12.52005861	4.87	1.2	0.25
Little Italy/NoLiTa	14.10656587	2.595872437	11.73699455	5.43	1.2	0.22
Gramercy	13.65255024	2.549601754	10.94571916	5.35	1.25	0.23
Murray Hill	14.1279692	2.736850103	11.08844285	5.16	1.27	0.25
Flatiron	14.4410444	2.778778468	12.04270761	5.2	1.2	0.23
Penn Station/Madison Sq West	15.75784706	3.151618824	12.90671765	5	1.22	0.24
TriBeCa/Civic Center	15.85307572	3.215454308	13.09238903	4.93	1.21	0.25
Greenwich Village North	14.96817502	2.848559956	12.63783347	5.25	1.18	0.23
West Chelsea/Hudson Yards	14.76535058	2.920114079	11.94530699	5.06	1.24	0.24
Midtown North	14.80608218	2.99463831	12.15366512	4.94	1.22	0.25
Kips Bay	13.13836182	2.477204128	10.11475331	5.3	1.3	0.24
Garment District	15.77947954	3.289325696	12.85205674	4.8	1.23	0.26
Sutton Place/Turtle Bay North	14.53110301	3.014600861	11.2076129	4.82	1.3	0.27
Yorkville West	13.97269793	2.912402311	10.13104542	4.8	1.38	0.29
JFK Airport	51.33499821	15.61242229	27.89474812	3.29	1.84	0.56
Lenox Hill West	14.00305764	2.878768349	10.37093925	4.86	1.35	0.28
Midtown East	14.22778745	2.854613711	11.07305042	4.98	1.28	0.26
Clinton West	14.0328739	2.768929619	11.41356305	5.07	1.23	0.24
Midtown Center	14.93886551	2.974139745	12.11336088	5.02	1.23	0.25
Lincoln Square East	14.61242449	2.919035995	11.8007861	5.01	1.24	0.25
Upper West Side South	14.89577108	3.064521625	11.5970948	4.86	1.28	0.26
SoHo	14.36497129	2.70915311	12.34916268	5.3	1.16	0.22
Hudson Sq	14.34807616	2.684683482	11.81535608	5.34	1.21	0.23
LaGuardia Airport	37.13724084	9.618701571	26.00367365	3.86	1.43	0.37
UN/Turtle Bay South	14.32632513	2.941247853	11.1452967	4.87	1.29	0.26
Upper East Side South	13.88879445	2.811651597	10.72332731	4.94	1.3	0.26
Alphabet City	15.6009199	3.012991239	13.06110763	5.18	1.19	0.23
Williamsburg (North Side)	16.89312583	3.486854305	14.19731788	4.84	1.19	0.25
Upper West Side North	14.10614384	2.872794521	10.87847032	4.91	1.3	0.26
Upper East Side North	13.00756778	2.601402014	9.531525949	5	1.36	0.27
Williamsburg (South Side)	15.90876818	3.210248075	12.90115483	4.96	1.23	0.25
Two Bridges/Seward Park	16.67003442	3.342590361	13.64447791	4.99	1.22	0.24
Financial District North	19.85992248	4.543049096	15.41752799	4.37	1.29	0.29
Lenox Hill East	14.51666667	3.018844697	10.60175189	4.81	1.37	0.28
East Harlem South	13.8998568	3.008544153	10.41113763	4.62	1.34	0.29
East Harlem North	15.64302885	3.543533654	12.01560497	4.41	1.3	0.29
Astoria	13.17032609	2.802355072	9.976932367	4.7	1.32	0.28
Central Harlem	12.56703846	2.549935897	9.59534188	4.93	1.31	0.27
World Trade Center	21.00515504	4.870542636	16.48891042	4.31	1.27	0.3
East Williamsburg	15.47120787	3.060519663	12.51718165	5.06	1.24	0.24
Financial District South	20.47365385	4.792100592	15.83000493	4.27	1.29	0.3
Chinatown	16.54724884	3.357681607	13.47279753	4.93	1.23	0.25
Park Slope	16.02188854	3.371934985	12.93343653	4.75	1.24	0.26
Manhattan Valley	14.81811166	3.148784893	11.87509579	4.71	1.25	0.27
Greenpoint	16.99764901	3.524271523	13.814266	4.82	1.23	0.26
Battery Park City	19.65335714	4.447660714	14.95449405	4.42	1.31	0.3
mean	16.27213501	3.435402586	12.76158173	4.912363636	1.261454545	0.260545455

< Information of the top 55 regions on Saturdays between 0am and 4am. >

This table is an analysis of the taxi service records from 0:00 am to 4:00 am on Saturdays. The selected 55 high taxi demand regions are very similar to those of Fridays. Most areas are located in Manhattan. As seen above, the average income-per-minute is \$1.26. Hence someone can earn up to \$151.32 if the driver works for two hours, and \$302.64 if he/she works for four hours on Saturday.

### iii. 12am – 4am on Sunday

Pick-up Zone	Avg Total Fare	Avg Trip Distance	Avg Trip Time	Avg Fare/mile	Avg Fare/minute	Avg Dist/minute
East Village	14.53430944	2.793587017	11.81526531	5.2	1.23	0.24
Lower East Side	15.80018358	3.110328178	12.94709544	5.08	1.22	0.24
West Village	15.28903618	2.964518602	12.97596416	5.16	1.18	0.23
Clinton East	15.46823548	3.16015353	12.55279309	4.89	1.23	0.25
Greenwich Village South	14.86466057	2.816460695	12.56711355	5.28	1.18	0.22
East Chelsea	14.74890341	2.820634802	12.43770019	5.23	1.19	0.23
Meatpacking/West Village West	15.23875379	2.841599596	13.64181451	5.36	1.12	0.21
Times Sq/Theatre District	15.33411683	3.182381245	12.79985396	4.82	1.2	0.25
Union Sq	14.52398864	2.843814932	12.04407049	5.11	1.21	0.24
Midtown South	14.49751767	2.803381016	11.96140188	5.17	1.21	0.23
Little Italy/NoLiTa	14.06846279	2.594065336	11.66228675	5.42	1.21	0.22
Gramercy	13.55103192	2.572418337	10.65000309	5.27	1.27	0.24
Murray Hill	13.65705652	2.648545792	10.57614136	5.16	1.29	0.25
West Chelsea/Hudson Yards	14.30482653	2.724876718	11.65613863	5.25	1.23	0.23
Flatiron	14.38014792	2.794697445	11.86089571	5.15	1.21	0.24
Penn Station/Madison Sq West	15.81689609	3.157764283	13.106429	5.01	1.21	0.24
Midtown North	14.73348935	2.908927163	11.95746779	5.06	1.23	0.24
TriBeCa/Civic Center	15.79224849	3.220673447	12.92466099	4.9	1.22	0.25
Greenwich Village North	14.52609454	2.801457513	12.09222003	5.19	1.2	0.23
Garment District	14.9275985	3.068758599	12.2583177	4.86	1.22	0.25
Yorkville West	13.89438925	2.903433141	10.2702223	4.79	1.35	0.28
Lenox Hill West	13.8267733	2.836265647	10.28234237	4.87	1.34	0.28
Kips Bay	13.59426148	2.681007067	10.43041814	5.07	1.3	0.26
Clinton West	15.179644	3.157807049	12.16782366	4.81	1.25	0.26
Lincoln Square East	14.60515671	2.936694606	11.78699587	4.97	1.24	0.25
Midtown East	13.51503853	2.615144954	10.64432416	5.17	1.27	0.25
Sutton Place/Turtle Bay North	13.59693511	2.769465649	10.41026081	4.91	1.31	0.27
Upper West Side South	14.49195102	2.973833802	11.06773719	4.87	1.31	0.27
Midtown Center	14.12579462	2.776764466	11.5565743	5.09	1.22	0.24
SoHo	14.88627433	2.870185357	12.5833565	5.19	1.18	0.23
Upper East Side South	12.87805774	2.484304462	10.00247594	5.18	1.29	0.25
Williamsburg (North Side)	17.78626792	3.733340684	15.09340316	4.76	1.18	0.25
Hudson Sq	14.7061732	2.810269163	12.00145309	5.23	1.23	0.23
JFK Airport	47.49448662	14.81583074	26.68029454	3.21	1.78	0.56
Alphabet City	15.17120474	2.94886392	12.19241573	5.14	1.24	0.24
Upper West Side North	13.69349934	2.809820717	10.30752545	4.87	1.33	0.27
UN/Turtle Bay South	14.41362022	2.920887978	10.84201958	4.93	1.33	0.27
Financial District North	19.84833333	4.639103107	15.27239878	4.28	1.3	0.3
Williamsburg (South Side)	16.77093772	3.423979957	14.04223336	4.9	1.19	0.24
Two Bridges/Seward Park	16.8225854	3.413400621	13.71924172	4.93	1.23	0.25
Upper East Side North	12.81684383	2.583565004	9.276914691	4.96	1.38	0.28
East Williamsburg	17.04353493	3.542168926	13.59871394	4.81	1.25	0.26
Lenox Hill East	13.79523705	2.788566703	9.780558618	4.95	1.41	0.29
East Harlem North	13.03315358	2.867889801	9.884017194	4.54	1.32	0.29
East Harlem South	13.49308685	2.9557277	10.36858372	4.57	1.3	0.29
Astoria	13.5118236	2.92647199	10.58077076	4.62	1.28	0.28
Central Harlem	12.59089157	2.505349398	9.972028112	5.03	1.26	0.25
Park Slope	17.01588462	3.676487179	14.00318376	4.63	1.22	0.26
Financial District South	20.83812903	5.093729032	15.3524086	4.09	1.36	0.33
Morningside Heights	15.71145333	3.249733333	12.93224444	4.83	1.21	0.25
Greenpoint	17.06101877	3.58349866	13.65784182	4.76	1.25	0.26
World Trade Center	19.39918169	4.434660194	15.41440129	4.37	1.26	0.29
Chinatown	15.33452345	3.013252648	12.43136662	5.09	1.23	0.24
Manhattan Valley	14.92364501	3.152155309	11.50467512	4.73	1.3	0.27
Lincoln Square West	14.57144538	3.018487395	11.56030812	4.83	1.26	0.26
mean	15.60906901	3.268021647	12.29380304	4.919090909	1.262181818	0.259636364

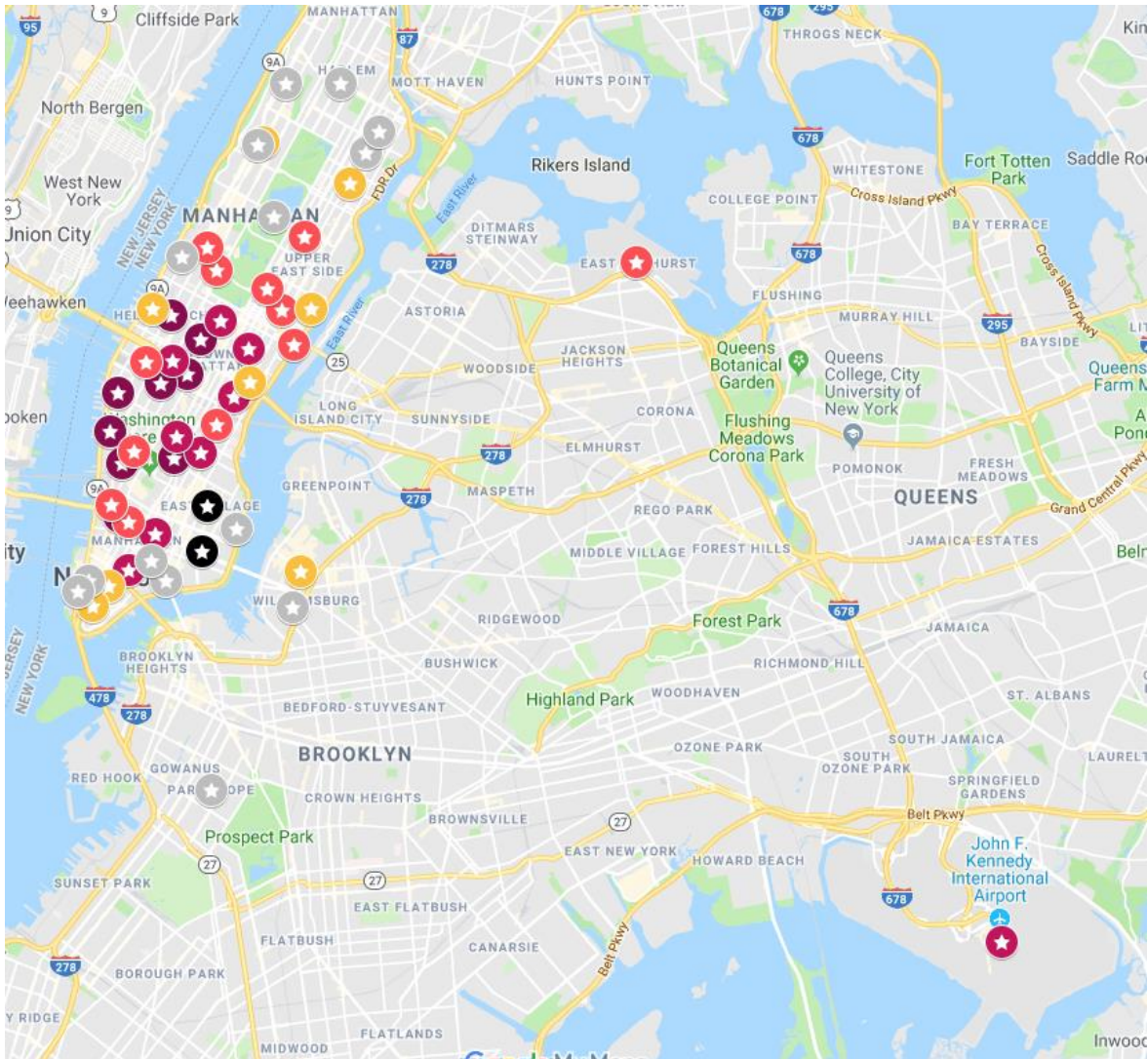
< Information of the top 55 regions on Saturdays between 0am and 4am. >

Last, the above table is an analysis of the taxi service records from 0:00 am to 4:00 am on Sundays. The selected 55 high taxi demand regions are very similar to those of Fridays and Saturdays. Most areas are located in Manhattan. As seen above, the average income-per-minute is \$1.262. Hence someone can earn up to \$151.44 if the driver works for two hours, and \$302.88 if he/she works for four hours on Sunday.

## 4.2. Selecting locations to pick-up customers

We were able to find out at what times someone can make the most income. Then, where can the person drive a taxi to maximize profit? The answer is areas between Lower Manhattan and Midtown Manhattan, especially Lower East regions of Manhattan.

### i. 0am – 2am on Fridays



< Pickup locations on Fridays between 0:00 and 2:00 am >

- = more than 6000
- = between 3000 and 6000
- = between 2000 and 3000
- = between 1000 and 2000
- = between 500 and 1000
- = less than 500



Fridays 0am - 2am			
PULocationID	Count	Borough	Zone
79	6418	Manhattan	East Village
148	6018	Manhattan	Lower East Side
230	4748	Manhattan	Times Sq/Theatre District
249	4697	Manhattan	West Village
48	4223	Manhattan	Clinton East
234	4057	Manhattan	Union Sq
114	3525	Manhattan	Greenwich Village South
164	3512	Manhattan	Midtown South
68	3349	Manhattan	East Chelsea
161	3336	Manhattan	Midtown Center
186	3134	Manhattan	Penn Station/Madison Sq West
158	3048	Manhattan	Meatpacking/West Village West
170	2940	Manhattan	Murray Hill
163	2892	Manhattan	Midtown North
132	2814	Queens	JFK Airport
107	2592	Manhattan	Gramercy
162	2590	Manhattan	Midtown East
144	2430	Manhattan	Little Italy/NoLiTa
231	2194	Manhattan	TriBeCa/Civic Center
90	2095	Manhattan	Flatiron
100	2086	Manhattan	Garment District
229	1892	Manhattan	Sutton Place/Turtle Bay North
142	1748	Manhattan	Lincoln Square East
138	1567	Queens	LaGuardia Airport
237	1544	Manhattan	Upper East Side South
113	1496	Manhattan	Greenwich Village North
211	1454	Manhattan	SoHo
141	1425	Manhattan	Lenox Hill West
137	1390	Manhattan	Kips Bay
263	1330	Manhattan	Yorkville West
239	1250	Manhattan	Upper West Side South
246	1083	Manhattan	West Chelsea/Hudson Yards
125	1048	Manhattan	Hudson Sq
233	984	Manhattan	UN/Turtle Bay South
238	963	Manhattan	Upper West Side North
236	920	Manhattan	Upper East Side North
87	803	Manhattan	Financial District North
50	796	Manhattan	Clinton West
140	620	Manhattan	Lenox Hill East
255	559	Brooklyn	Williamsburg (North Side)
88	506	Manhattan	Financial District South
43	485	Manhattan	Central Park
261	474	Manhattan	World Trade Center
13	472	Manhattan	Battery Park City
4	455	Manhattan	Alphabet City
232	391	Manhattan	Two Bridges/Seward Park
143	385	Manhattan	Lincoln Square West
75	379	Manhattan	East Harlem South
74	374	Manhattan	East Harlem North
256	366	Brooklyn	Williamsburg (South Side)
45	347	Manhattan	Chinatown
151	341	Manhattan	Manhattan Valley
41	323	Manhattan	Central Harlem
166	321	Manhattan	Morningside Heights
181	316	Brooklyn	Park Slope

The map above is created using the next table values. The points on the map indicate that the darker colors, the higher demand for taxis. If we look at the darkest points of the map, East Village and Lower East Side had the most demands for the taxis on Fridays early in the morning. There were also a significant number of taxi demands between Lower Manhattan and Midtown Manhattan.

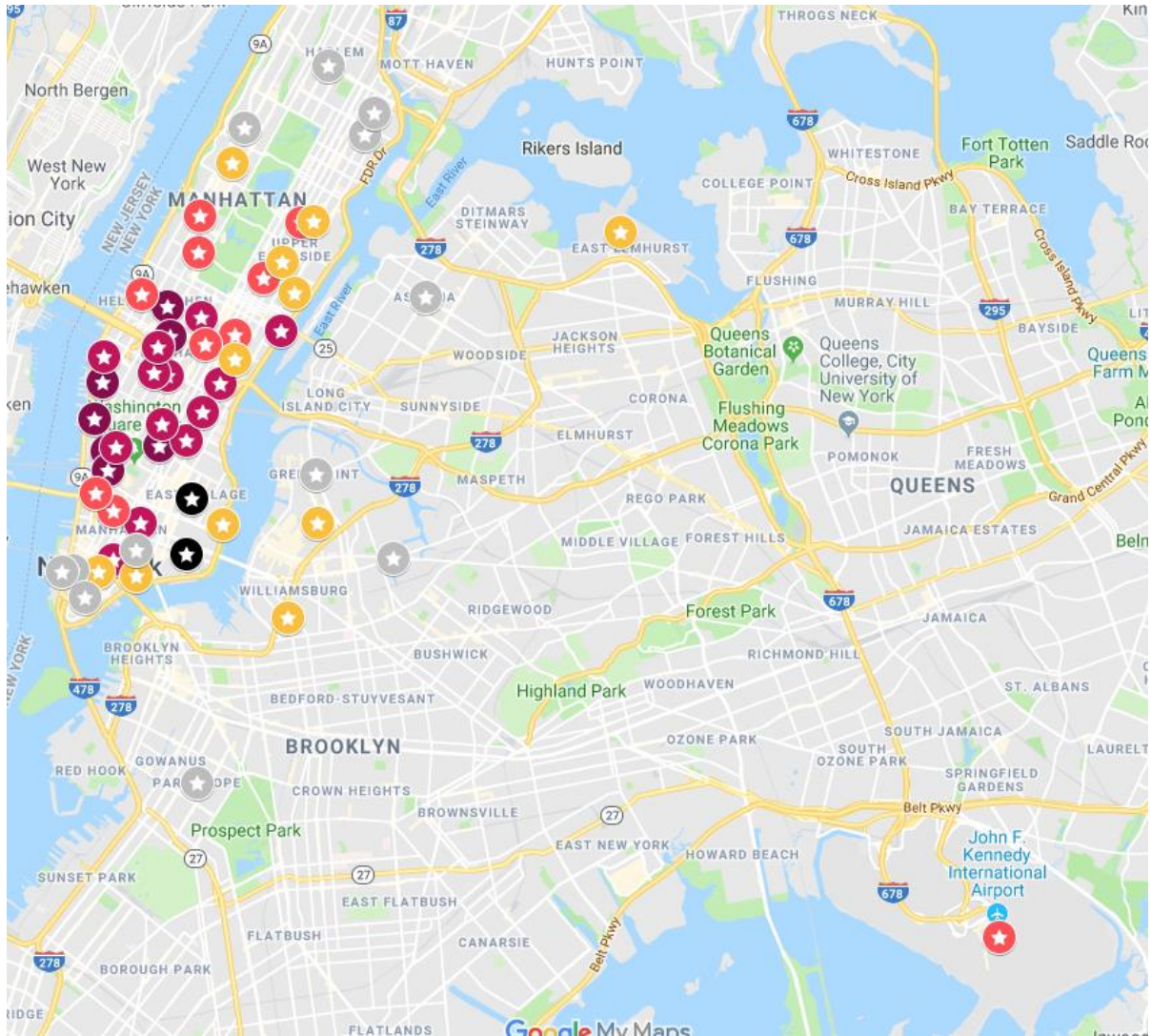
Also, if we look at the table, we can see that all taxi trips took place in Manhattan except for just four places, JFK airport, La Guardia airport and Williamsburg, Brooklyn.

If a taxi drivers wants to pick up the first customers **NOT** in the lower-central Manhattan areas, JFK airport or LaGuardia airport in Queens could be another choice.

✂ I would like to note that taxi records *from 10 pm to 0 am on Tuesdays night* also give similar results compared to the taxi records on Fridays between 0 am and 2 am. For example, average income per minute on Tuesdays between 10 pm and 0 am was \$1.25. Furthermore, the number of passengers and the areas with high taxi demands were very similar.

So if a person has a difficulty with working on Friday, he/she can drive a taxi on Tuesday night instead of Friday morning.

ii. 0am – 4am on Saturdays



< Pickup locations on Saturdays between 0:00 and 4:00 am >

- = more than 10000
- = between 6000 and 10000
- = between 3000 and 6000
- = between 2000 and 3000
- = between 1000 and 2000
- = less than 1000

Saturdays 0am - 4am			
PULocationID	Count	Borough	Zone
79	20422	Manhattan	East Village
148	15743	Manhattan	Lower East Side
249	9741	Manhattan	West Village
114	9153	Manhattan	Greenwich Village South
48	8561	Manhattan	Clinton East
68	7525	Manhattan	East Chelsea
234	6774	Manhattan	Union Sq
158	6290	Manhattan	Meatpacking/West Village West
164	6086	Manhattan	Midtown South
230	6029	Manhattan	Times Sq/Theatre District
144	5754	Manhattan	Little Italy/NoLiTa
107	5474	Manhattan	Gramercy
170	4870	Manhattan	Murray Hill
90	4347	Manhattan	Flatiron
186	4250	Manhattan	Penn Station/Madison Sq West
231	3830	Manhattan	TriBeCa/Civic Center
113	3611	Manhattan	Greenwich Village North
246	3594	Manhattan	West Chelsea/Hudson Yards
163	3456	Manhattan	Midtown North
137	3101	Manhattan	Kips Bay
100	3055	Manhattan	Garment District
229	3019	Manhattan	Sutton Place/Turtle Bay North
263	2943	Manhattan	Yorkville West
132	2799	Queens	JFK Airport
141	2793	Manhattan	Lenox Hill West
162	2757	Manhattan	Midtown East
50	2728	Manhattan	Clinton West
161	2662	Manhattan	Midtown Center
142	2417	Manhattan	Lincoln Square East
239	2289	Manhattan	Upper West Side South
211	2090	Manhattan	SoHo
125	2022	Manhattan	Hudson Sq
138	1910	Queens	LaGuardia Airport
233	1747	Manhattan	UN/Turtle Bay South
237	1659	Manhattan	Upper East Side South
4	1598	Manhattan	Alphabet City
255	1510	Brooklyn	Williamsburg (North Side)
238	1460	Manhattan	Upper West Side North
236	1291	Manhattan	Upper East Side North
256	1169	Brooklyn	Williamsburg (South Side)
232	1162	Manhattan	Two Bridges/Seward Park
87	1161	Manhattan	Financial District North
140	1056	Manhattan	Lenox Hill East
75	838	Manhattan	East Harlem South
74	832	Manhattan	East Harlem North
7	828	Queens	Astoria
41	780	Manhattan	Central Harlem
261	774	Manhattan	World Trade Center
80	712	Brooklyn	East Williamsburg
88	676	Manhattan	Financial District South
45	647	Manhattan	Chinatown
181	646	Brooklyn	Park Slope
151	609	Manhattan	Manhattan Valley
112	604	Brooklyn	Greenpoint
13	560	Manhattan	Battery Park City

Likewise, the above map was created using the next table values (Saturdays between 0 – 4 am).

Compared to Fridays, **the importance of the lower central Manhattan areas increases significantly.**

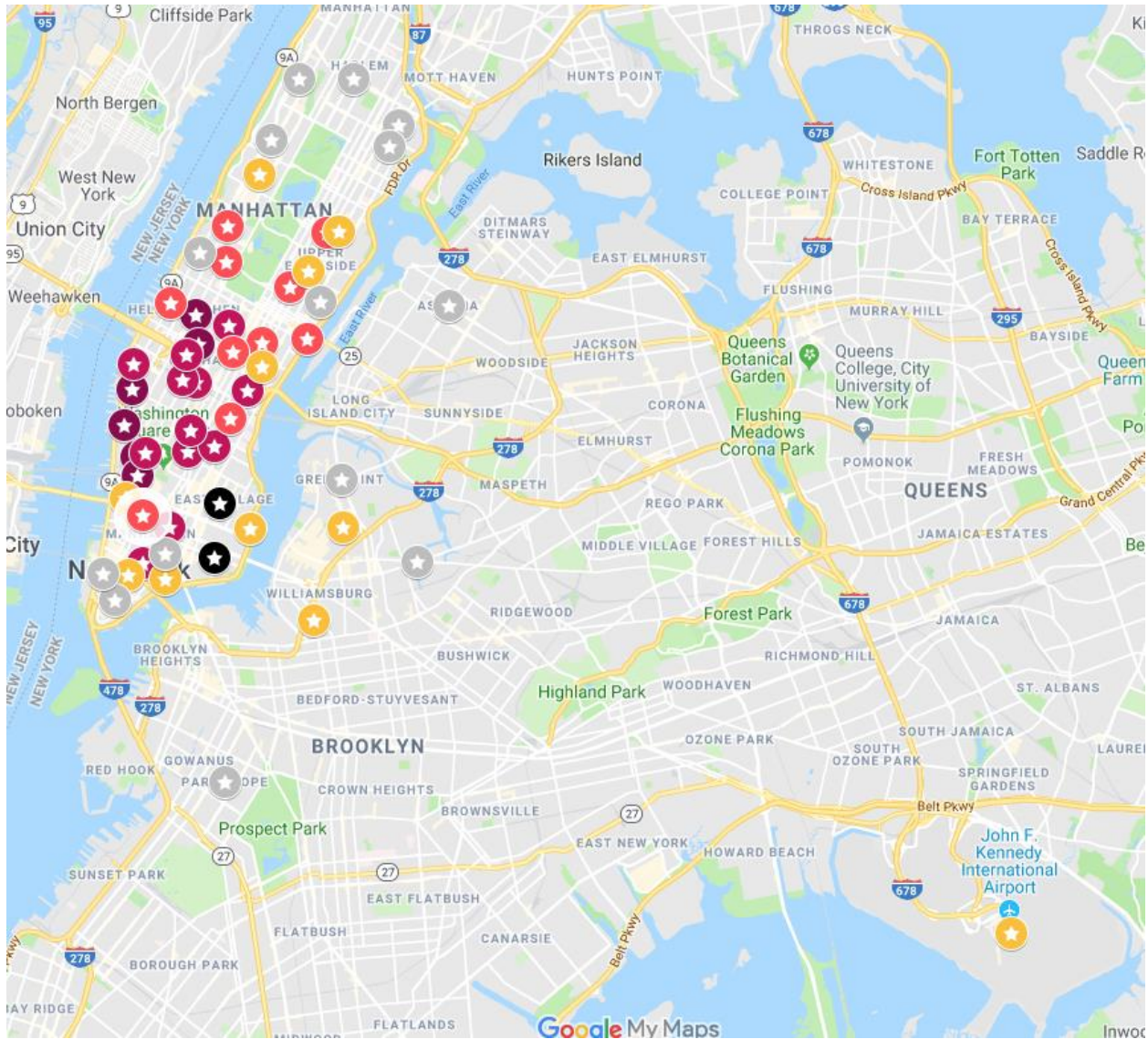
Look at the table. Of the total 205,488 trips on Saturdays, 144,685 rides occurred in the lower center of Manhattan. (Highlighted in yellow on the table.) This corresponds to 70% of the total taxi demand.

By expanding the range to all areas of Manhattan, the demand for taxis will increase to 188,857. This amounts to approximately 92% of total demand on Saturdays between 0 am and 4 am.

Therefore, it is advisable to drive a taxi in Manhattan early in the morning on Saturday.



iii. 0am – 4am on Sundays



< Pickup locations on Sundays between 0:00 and 4:00 am >



Sundays 0am - 4am			
PULocationID	Count	Borough	Zone
79	19933	Manhattan	East Village
148	15906	Manhattan	Lower East Side
249	9784	Manhattan	West Village
48	9249	Manhattan	Clinton East
114	8205	Manhattan	Greenwich Village South
68	8034	Manhattan	East Chelsea
158	6933	Manhattan	Meatpacking/West Village West
230	6505	Manhattan	Times Sq/Theatre District
234	5987	Manhattan	Union Sq
164	5942	Manhattan	Midtown South
144	5510	Manhattan	Little Italy/NoLiTa
107	5388	Manhattan	Gramercy
170	4848	Manhattan	Murray Hill
246	4583	Manhattan	West Chelsea/Hudson Yards
90	4462	Manhattan	Flatiron
186	4446	Manhattan	Penn Station/Madison Sq West
163	3803	Manhattan	Midtown North
231	3638	Manhattan	TriBeCa/Civic Center
113	3554	Manhattan	Greenwich Village North
100	3198	Manhattan	Garment District
263	2939	Manhattan	Yorkville West
141	2876	Manhattan	Lenox Hill West
137	2830	Manhattan	Kips Bay
50	2809	Manhattan	Clinton West
142	2744	Manhattan	Lincoln Square East
162	2725	Manhattan	Midtown East
229	2620	Manhattan	Sutton Place/Turtle Bay North
239	2491	Manhattan	Upper West Side South
161	2454	Manhattan	Midtown Center
211	2158	Manhattan	SoHo
237	1905	Manhattan	Upper East Side South
255	1814	Brooklyn	Williamsburg (North Side)
125	1709	Manhattan	Hudson Sq
132	1607	Queens	JFK Airport
4	1602	Manhattan	Alphabet City
238	1506	Manhattan	Upper West Side North
233	1464	Manhattan	UN/Turtle Bay South
87	1416	Manhattan	Financial District North
256	1397	Brooklyn	Williamsburg (South Side)
232	1288	Manhattan	Two Bridges/Seward Park
236	1223	Manhattan	Upper East Side North
80	959	Brooklyn	East Williamsburg
140	907	Manhattan	Lenox Hill East
74	853	Manhattan	East Harlem North
75	852	Manhattan	East Harlem South
7	839	Queens	Astoria
41	830	Manhattan	Central Harlem
181	780	Brooklyn	Park Slope
88	775	Manhattan	Financial District South
166	750	Manhattan	Morningside Heights
112	746	Brooklyn	Greenpoint
261	721	Manhattan	World Trade Center
45	661	Manhattan	Chinatown
151	631	Manhattan	Manhattan Valley
143	595	Manhattan	Lincoln Square West

Likewise, the above map was generated using the next table values (Sundays between 0 – 4 am).

Like Saturdays, **the lower central Manhattan zones seem to be very important in picking up taxi customers.**

The sum of the taxi demand in the areas marked yellow on the table is 139,908. The entire taxi rides that occurred on Sundays early in the morning were 206908. That is, in the lower central areas, approximately 68% of the total taxi demand happened.

By expanding the range to all areas of Manhattan, the demand for taxis will increase to 191469. This represents more than 92% of total demands on Sundays between 0am and 4am.

Therefore, it is also advisable to drive a taxi in Manhattan early in the morning on Sunday.

There is one thing to consider.

If a driver pick up a passenger in Manhattan, how likely is it that the driver will be asked to move to a location other than Manhattan?

If the probability of a trip from Manhattan to another region is relatively high, a taxi driver will have to leave Manhattan with a high possibility, and this will not help maximize taxi income. This is because taxi drivers are more likely to maximize their income by picking up and dropping off customers in Manhattan for a given period of time.

Fridays 0am - 2am		
Drop-offs in Manhattan	Total Trips	Percent
83826	106182	<b>78.95%</b>
Pick-ups in Manhattan	Total Trips	Percent
97473	106182	91.80%
Saturdays 0am - 4am		
Drop-offs in Manhattan	Total Trips	Percent
164874	205488	<b>80.23%</b>
Pick-ups in Manhattan	Total Trips	Percent
188857	205488	91.91%
Sundays 0am - 4am		
Drop-offs in Manhattan	Total Trips	Percent
167432	206908	<b>80.92%</b>
Pick-ups in Manhattan	Total Trips	Percent
190763	206908	92.19%

Fortunately, the odds of getting a customer in Manhattan and dropping off in Manhattan are about 79%, 80%, and 80% on Friday, Saturday, and Sunday respectively. Also, if we consider cases that dropping customers in Queens or Brooklyn areas which are within 10 miles to Manhattan, then the probability of getting off passengers in areas that are within 15 miles of central Manhattan increases to about 90%.

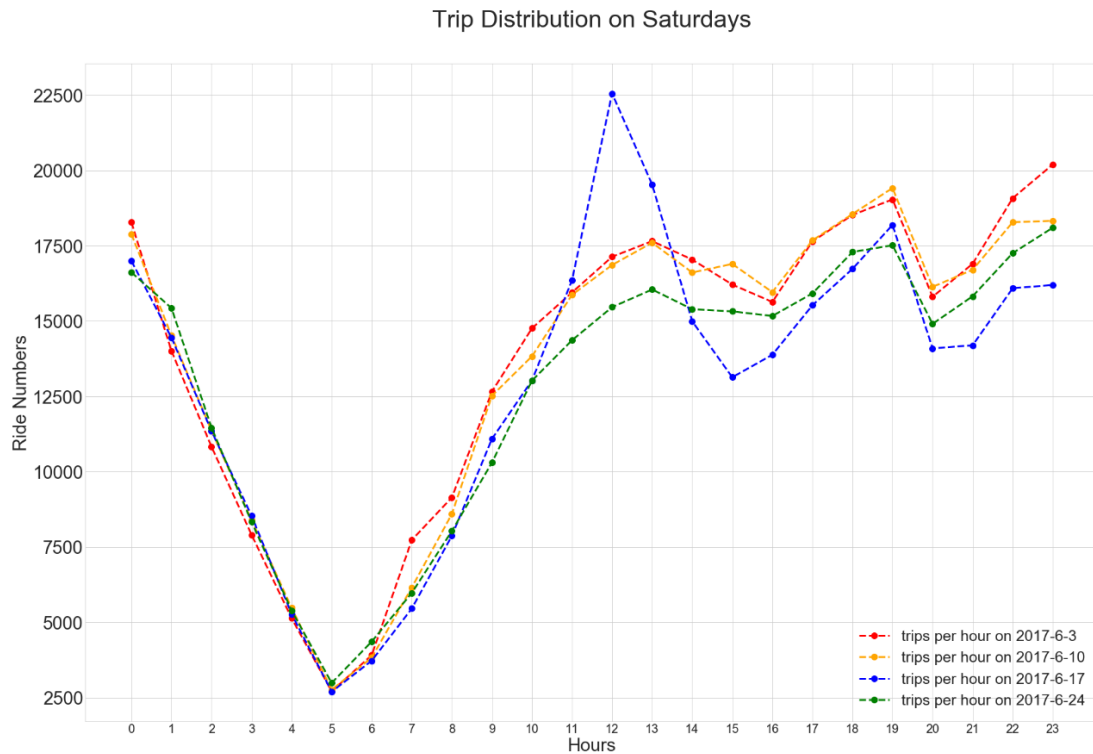
In conclusion, people who take a taxi in Manhattan are very likely to get off at other locations in Manhattan or areas near Manhattan.

So, if someone works 10 hours a week as a taxi driver, I strongly recommend the taxi driver to **operate a taxi between Lower (East) and Midtown Manhattan areas from 0 am to 2 am on Friday, from 0 am to 4 am on Saturday and from 0 am to 4 am on Sunday to maximize his/her income.** (Working on Friday between 0 am and 2 am can be replaced by driving on Tuesday between 10 pm and 0 am.)

# Question 1

*If you could enrich the dataset, what would you add?*

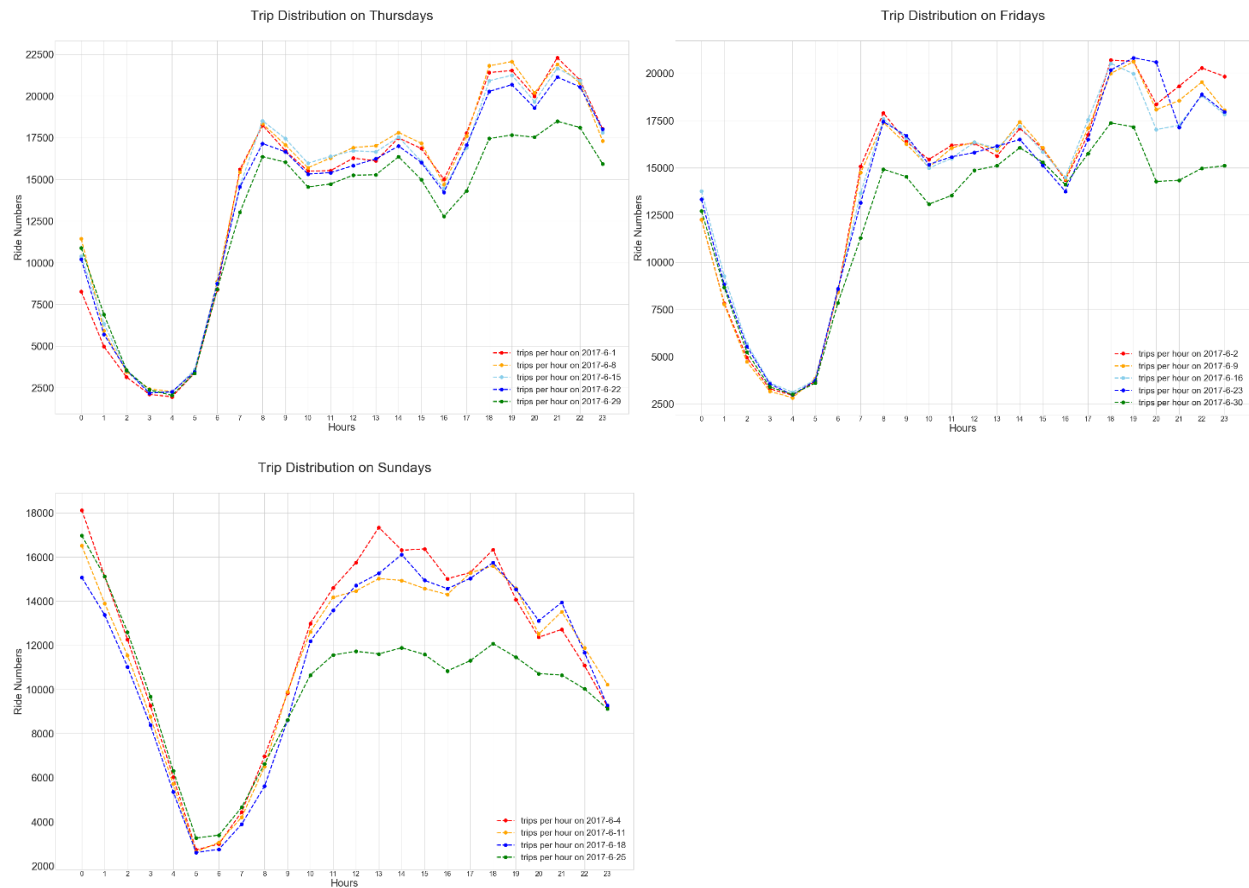
First, I think weather data is essential. People are more likely to take taxis when it rains or snows. Or when weather is too hot or cold, a number of demands for taxis will increase.



For example, the sudden rise of the blue line (6-24-2017) at 12 o'clock in the above graph may be due to bad weather.

Therefore, a taxi driver can maximize his/her income by predicting a number of demands of taxis according to weather.

Secondly, I would like to add some data that show when a particular event was held over a certain period of time.



In the above graphs, the green lines show different patterns than the other lines' shapes. In other words, taxi demands on those days (6/25, 6/29, 6/30) were considerably smaller. We need a data that can explain why there were different patterns.

Finally, I think it would be better if coordinates of areas where actual customers took and left taxis were presented, instead of information of the grouped areas such as East Village or Lower East Side. This is because, even if there is an area with a high demand for taxi, there could be a sub-place where customers do not ride in a taxi many times, or vice versa. Hence if exact coordinates would be given, it would help to maximize a taxi driver's income.

## Question 2

*Is there anything in the dataset that you don't find especially useful?*

I developed a Random Forest model to find out the importance of the variables associated with taxi fare.

(A). VendorID : TPEP provider that provided the taxi record. It has no significance as a variable related to taxi fare.

(B). store\_and\_fwd\_flag : indicates whether the trip record was held in vehicle memory before sending to the vendor. It also does not have significance as a variable related to taxi fare.

(C). tolls\_amount : Total amount of all tolls paid in trip. Taxi drivers can charge passengers only for toll expenses incurred during trips, which do not help increase their income. Therefore, this cost should be subtracted from total\_amount.

The following variables may help predict taxi drivers' earnings, but they do not seem to help taxi drivers maximize their income.

According to the Random Forest model, the payment\_type, RatecodeID, and DOLocationID variables are important variables in predicting taxi drivers' incomes. However, these variables are values that taxi drivers cannot select. For example, with regard to payment\_type, taxi drivers cannot ask a customer for a credit card payment or force a cash payment. The same is true of RatecodeID. Taxi drivers cannot ask customers for their desired cost rate. Only a fixed cost rate can be applied depending on destinations. In addition, taxi drivers should drive taxis to the destinations of passengers as requested by passengers.

The values of these features cannot be selected by taxi drivers. Therefore, these variables are not useful to achieve the goal of this test, 'How can a taxi driver maximize his/her income?'