

uber data-set

The dataset contains data of about 4.5 million uber pickups in New York City from April to September and 14.3 million pickups from January to June 2015.

Accessing the dataset

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
data = pd.read_csv("E:\\DATA SCIENCE\\Project\\python\\UBER\\uber_tips_analysis.csv")
```

Displaying The TOP10 Row Of Dataset.

```
In [13]: data["Date/Time"] = data["Date/Time"].map(pd.to_datetime)
data.head(10)
```

```
Out[13]:
```

	Date/Time	Lat	Lon	Base
0	2014-09-01 00:01:00	40.2201	-74.0021	B02512
1	2014-09-01 00:01:00	40.7500	-74.0027	B02512
2	2014-09-01 00:03:00	40.7559	-73.9864	B02512
3	2014-09-01 00:06:00	40.7450	-73.9889	B02512
4	2014-09-01 00:11:00	40.8145	-73.9444	B02512
5	2014-09-01 00:12:00	40.6735	-73.9918	B02512
6	2014-09-01 00:15:00	40.7471	-73.6472	B02512
7	2014-09-01 00:16:00	40.6613	-74.2691	B02512
8	2014-09-01 00:32:00	40.3745	-73.9999	B02512
9	2014-09-01 00:33:00	40.7633	-73.9773	B02512

WE FOUND THE DATASET CONTAINS :

1. DATE_TIME,
2. LATITUDE,
3. LONGITUDE,
4. BASE: CODE OF DIFFERENT AFFILIATED PARTNERS

Uber trips according to days and hours:

```
In [18]: data["Day"] = data["Date/Time"].apply(lambda x: x.day)
data["Weekday"] = data["Date/Time"].apply(lambda x: x.weekday())
data["Hour"] = data["Date/Time"].apply(lambda x: x.hour)
print(data.head(10))
```

	Date/Time	Lat	Lon	Base	Day	Weekday	Hour
0	2014-09-01 00:01:00	40.2201	-74.0021	B02512	1	0	0
1	2014-09-01 00:01:00	40.7500	-74.0027	B02512	1	0	0
2	2014-09-01 00:03:00	40.7559	-73.9864	B02512	1	0	0
3	2014-09-01 00:06:00	40.7450	-73.9889	B02512	1	0	0
4	2014-09-01 00:11:00	40.8145	-73.9444	B02512	1	0	0
5	2014-09-01 00:12:00	40.6735	-73.9918	B02512	1	0	0
6	2014-09-01 00:15:00	40.7471	-73.6472	B02512	1	0	0
7	2014-09-01 00:16:00	40.6613	-74.2691	B02512	1	0	0
8	2014-09-01 00:32:00	40.3745	-73.9999	B02512	1	0	0
9	2014-09-01 00:33:00	40.7633	-73.9773	B02512	1	0	0

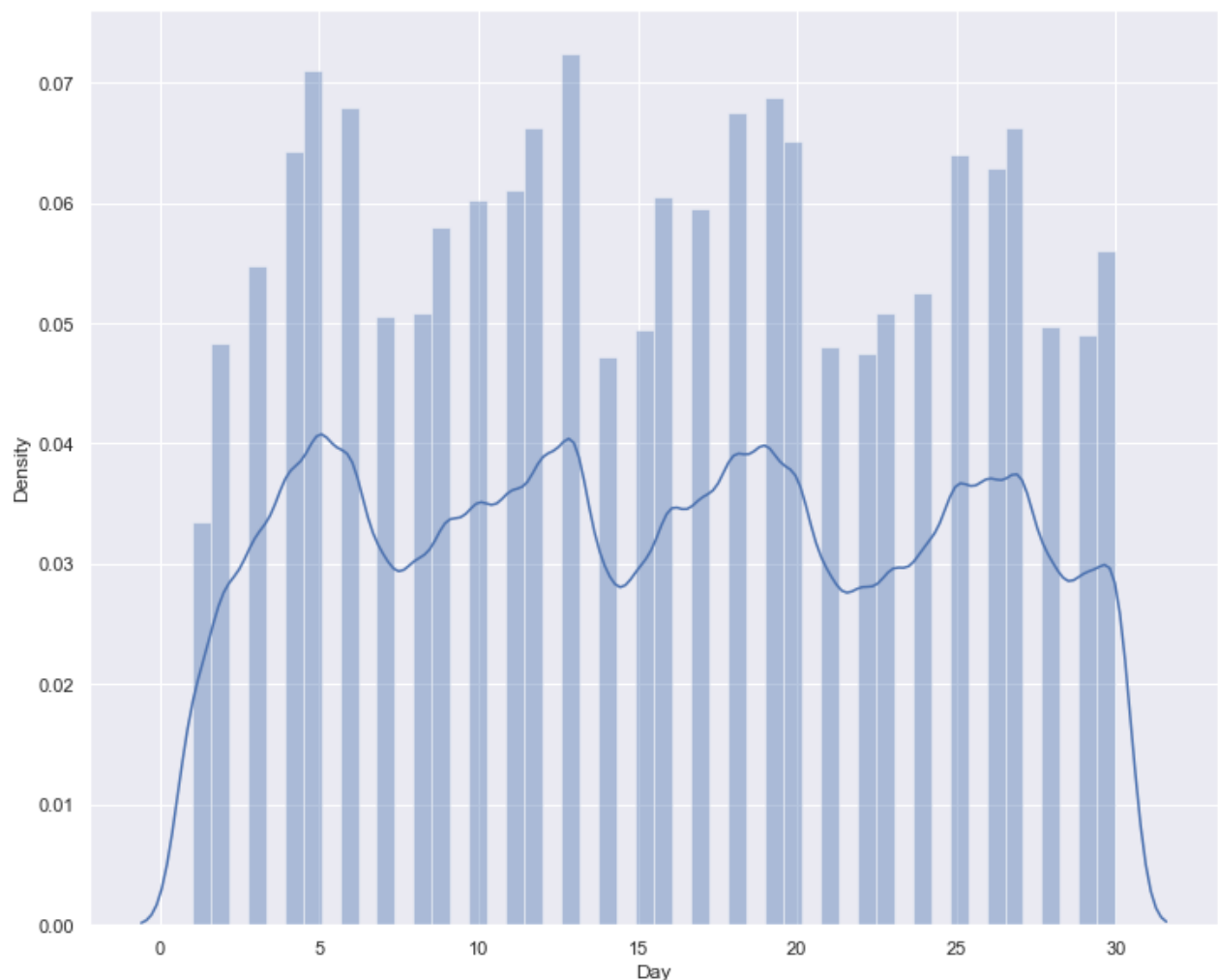
WE FOUND: zero weekdays zero hour

UBER-TRIPS Were Highest on which day?

```
In [24]: sns.distplot(data["Day"])
```

C:\Users\Pooja\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
warnings.warn(msg, FutureWarning)

```
Out[24]: <AxesSubplot:xlabel='Day', ylabel='Density'>
```



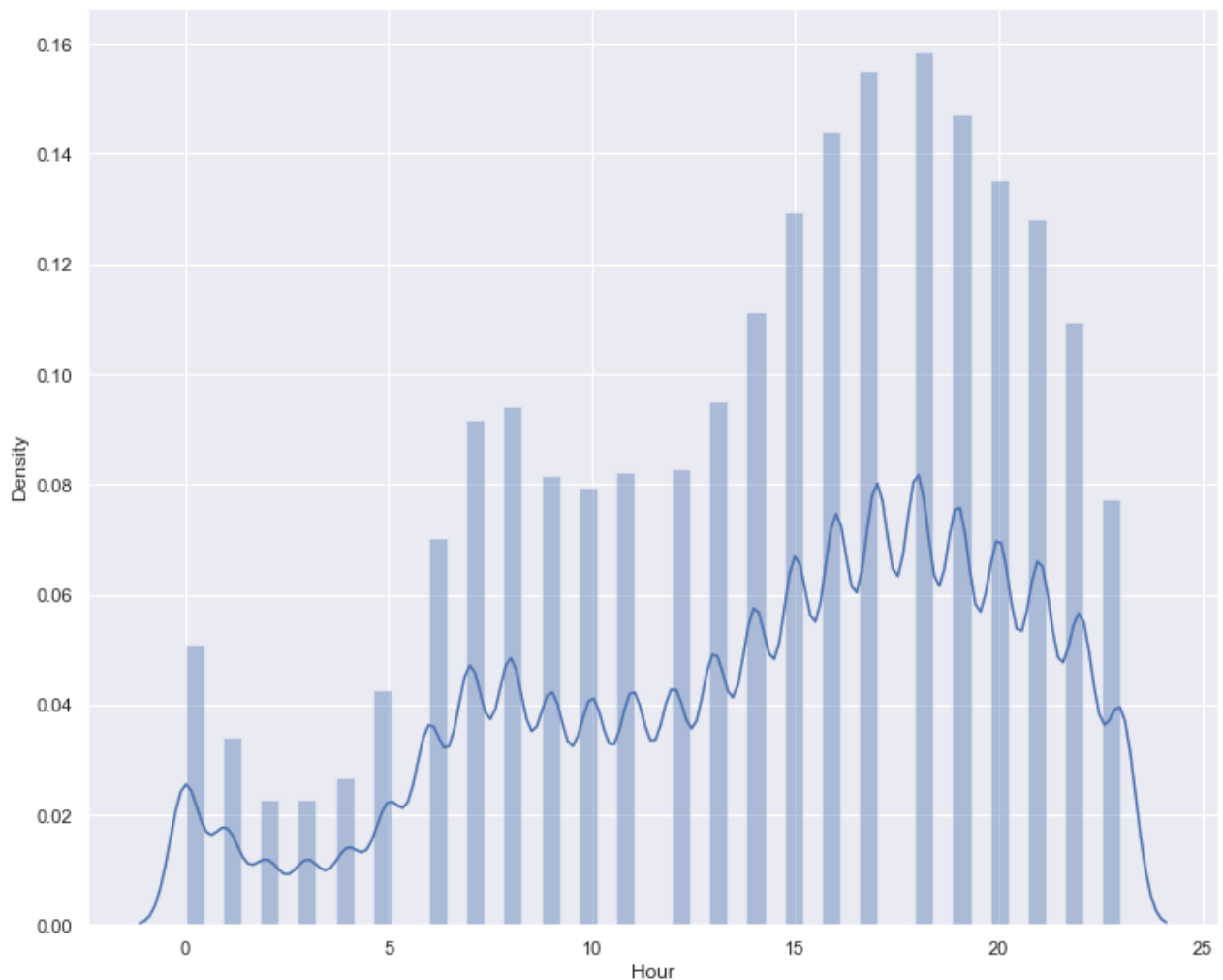
Findings : Uber trips are rising on the working days and decreases on the weekends.

UBER-TRIPS Were Highest on which hours?

```
In [23]: sns.distplot(data["Hour"])
```

C:\Users\Pooja\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
warnings.warn(msg, FutureWarning)

```
Out[23]: <AxesSubplot:xlabel='Hour', ylabel='Density'>
```



1. There is a sharp fall seen in Uber trips after midnight.
2. Then starts increasing after 5 am and the trips keep rising till 6 pm.
3. 6 pm is the busiest hour for Uber.
4. After 6pm the trips start decreasing.

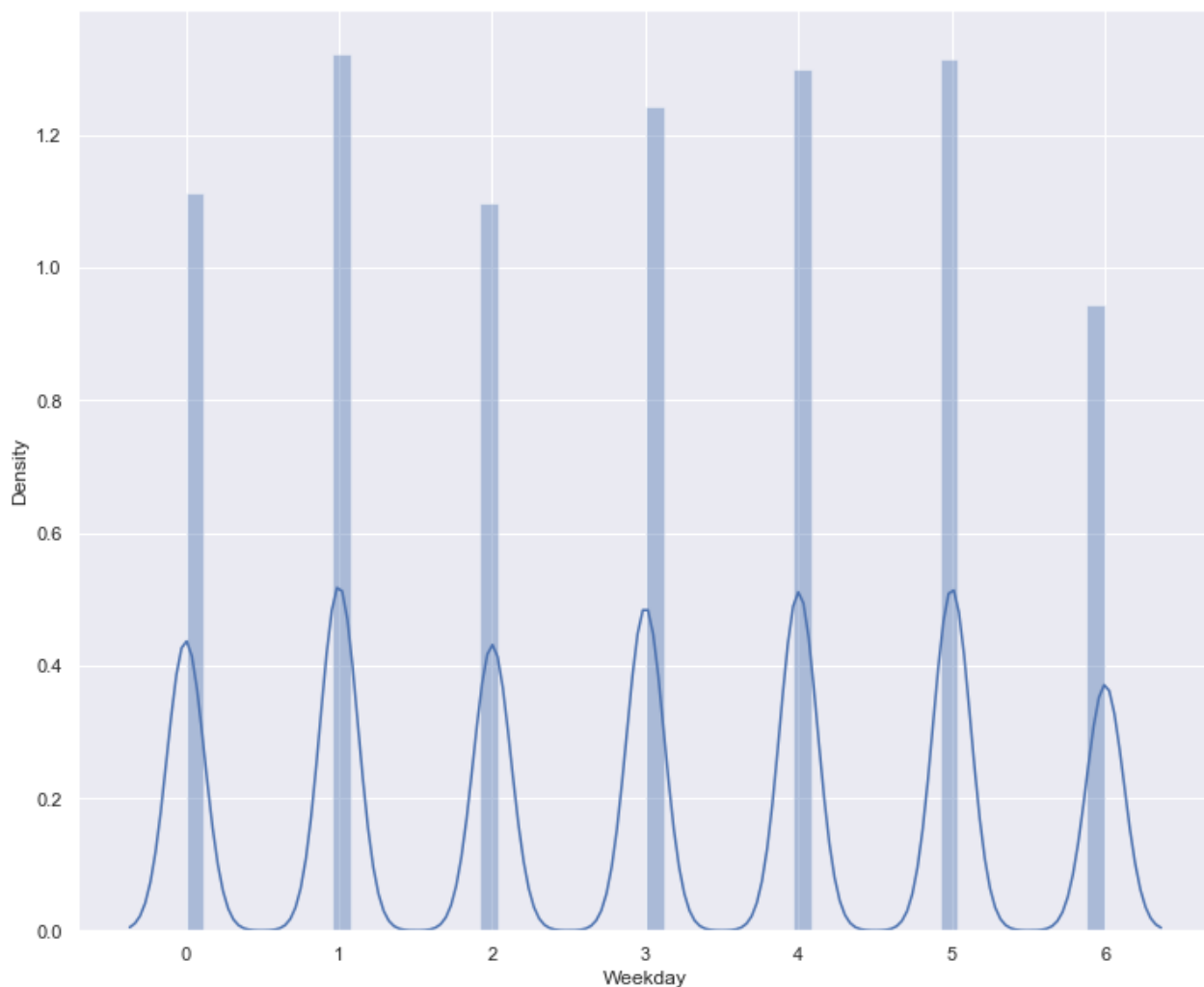
Which weekdays have the highest UBER trips?

```
In [25]: sns.distplot(data["Weekday"])
```

C:\Users\Pooja\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

```
Out[25]: <AxesSubplot:xlabel='Weekday', ylabel='Density'>
```



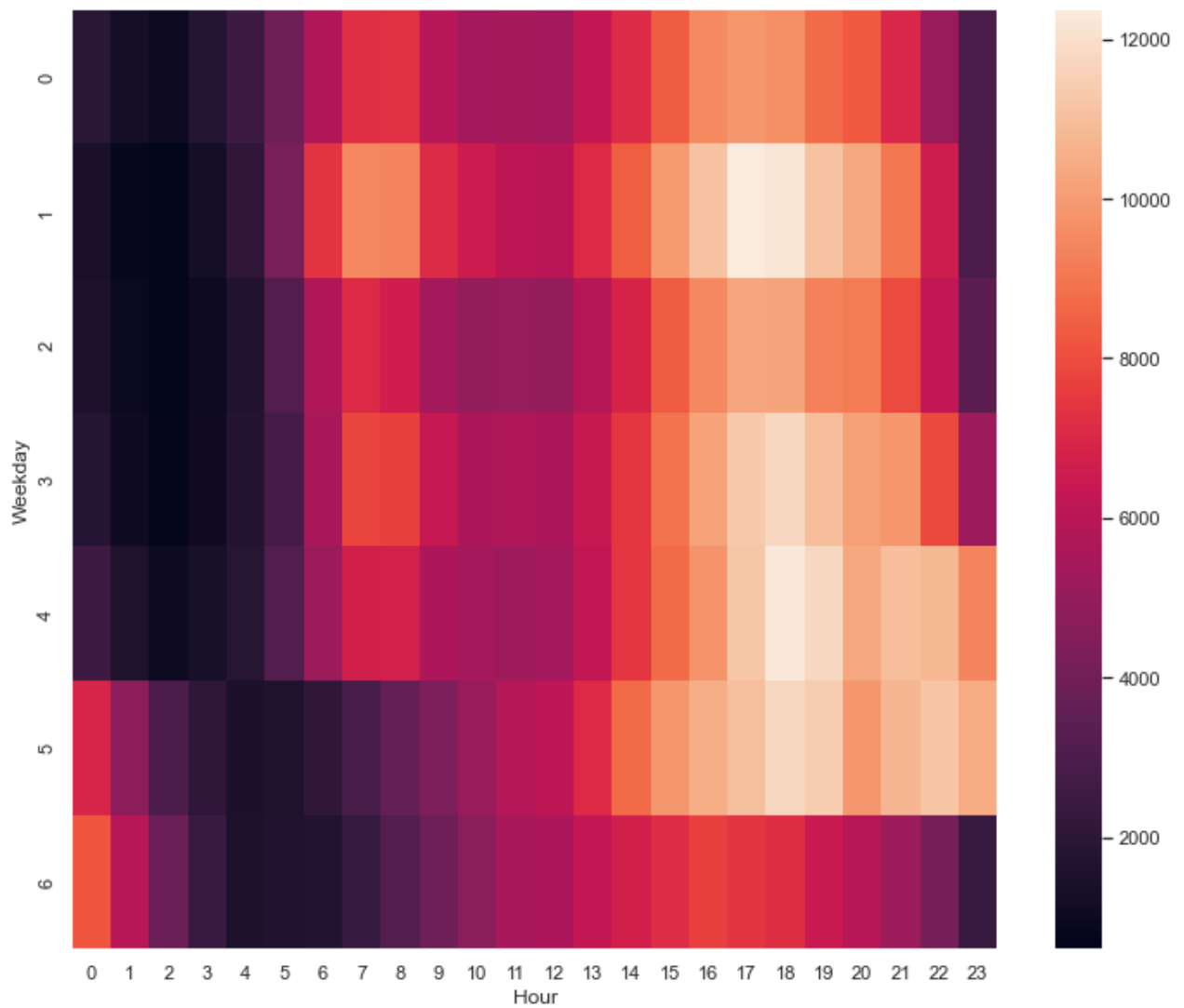
Here 0 indicates Sunday, 1 indicates Monday, 2 Indicates Tuesday, 3 indicates Wednesday..... 6 indicates Saturday. Findings :

1. Highest no of trips : Monday
2. Lowest no of Trips : Saturday
3. No of trips on Saturday is less than Sunday : so, we can conclude that the commuters prefer uber not only for office but also for other purpose also.

Correlation of hours and weekdays on the Uber trips.

```
In [31]: df = data.groupby(["Weekday", "Hour"]).apply(lambda x: len(x))
df = df.unstack()
sns.heatmap(df, annot=False)
```

```
Out[31]: <AxesSubplot:xlabel='Hour', ylabel='Weekday'>
```



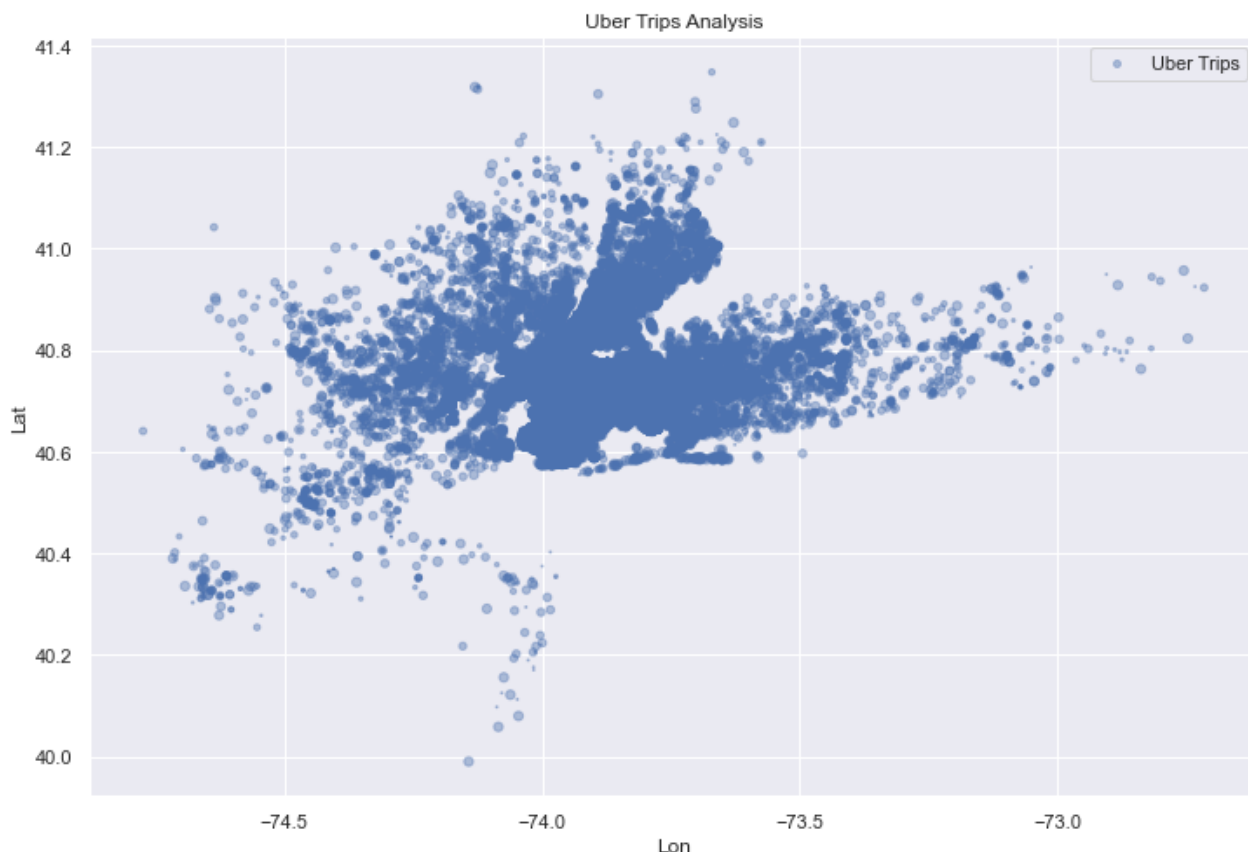
Findings: So

1. The frequency of Uber trips are quite higher during Sunday To Thursday and it continues till 4a.m.
2. The frequency of Uber trips are quite higher in friday - saturday but only during 4am to 7am.

Density of Trips based on regions Of NewYork City.

```
In [34]: data.plot(kind='scatter', x='Lon', y='Lat', alpha=0.4, s=data['Day'], label='Uber Trips',
figsize=(12, 8), cmap=plt.get_cmap('jet'))
plt.title("Uber Trips Analysis")
plt.legend()
plt.show()
```

c argument looks like a single numeric RGB or RGBA sequence, which should be avoided as a value-mapping will have precedence in case its length matches with *x* & *y*. Please use the *color* keyword-argument or provide a 2D array with a single row if you intend to specify the same RGB or RGBA value for all points.



SUMMARY

Some of the conclusions that We got from this analysis are:

1. Monday is the most profitable day for Uber
2. On Saturdays least number of people use Uber
3. 6 pm is the busiest day for Uber.
4. The frequency of Uber trips are quite higher during Sunday To Thursday and it continues till 4a.m.
5. The frequency of Uber trips are quite higher in friday - saturday but only during 4am to 7am.
6. On an average a rise in Uber trips starts around 5 am.
7. Most of the Uber trips originate near the Manhattan region in New York.