

## Uber Trips Analysis

By analyzing Uber trips, we can draw many patterns like which day has the highest and the lowest trips or the busiest hour for Uber and many other patterns. The dataset I'm using here is based on Uber trips from New York, a city with a very complex transportation system with a large residential community.

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. You can do so much more with this dataset rather than just analyzing it.

```
#ACCESS DATASET FROM GOOGLE DRIVE
```

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
data = pd.read_csv("drive/My Drive/DATA/Uber.csv")
data["Date/Time"] = data["Date/Time"].map(pd.to_datetime)
data.head()
```

	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

```
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())
```

	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

prepared this data according to the days and hours, as I am using the Uber trips for the September month so let's have a look at each day to see on which day the Uber trips were highest:

```
sns.set(rc={'figure.figsize':(12, 10)})
sns.distplot(data["Day"])
```

```
<ipython-input-6-2282722f9d2a>:2: UserWarning:
```

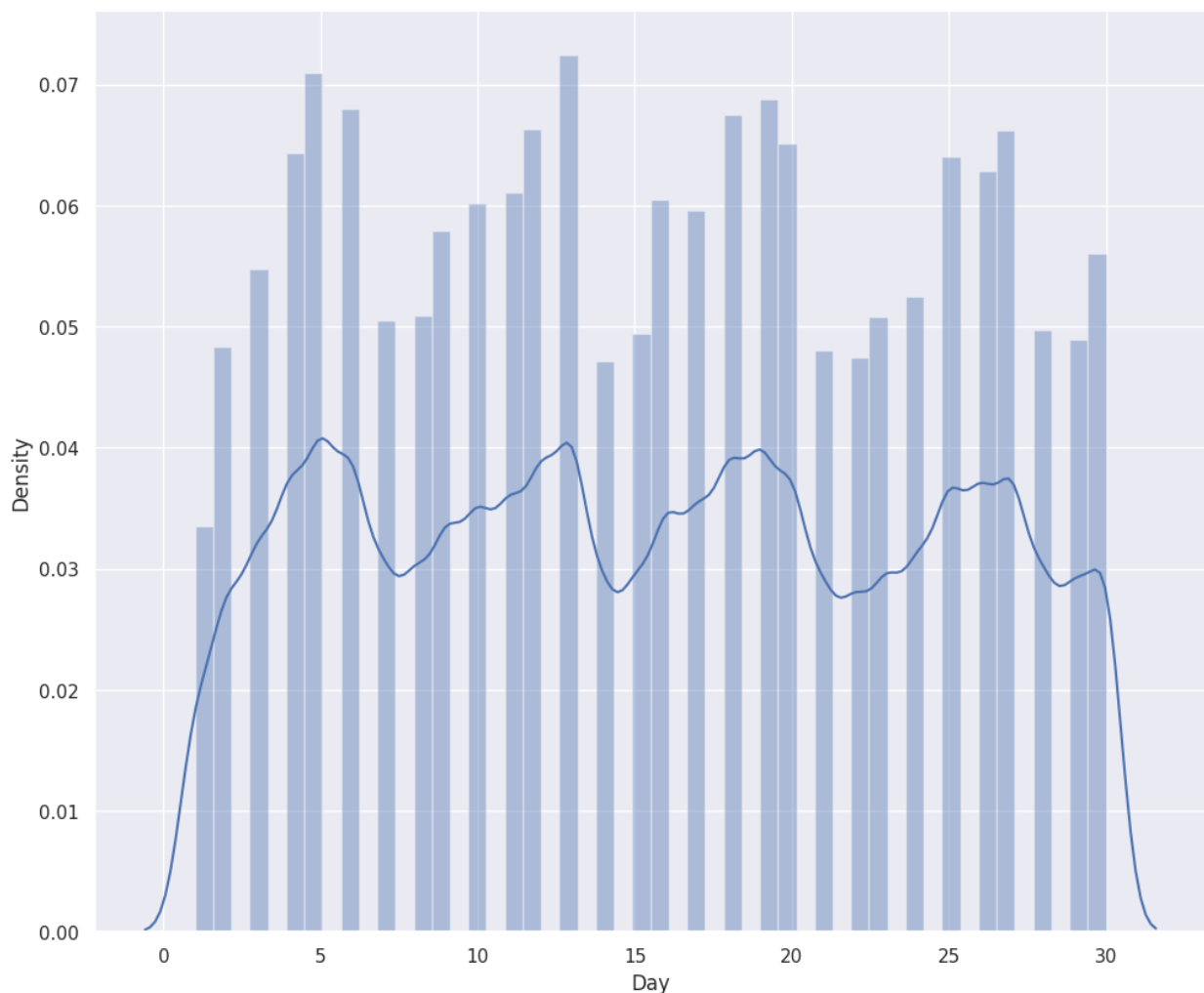
```
`distplot` is a deprecated function and will be removed in seaborn  
v0.14.0.
```

```
Please adapt your code to use either `displot` (a figure-level  
function with  
similar flexibility) or `histplot` (an axes-level function for  
histograms).
```

```
For a guide to updating your code to use the new functions, please see  
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
sns.distplot(data["Day"])
```

```
<Axes: xlabel='Day', ylabel='Density'>
```



By looking at the daily trips we can say that the Uber trips are rising on the working days and decreases on the weekends.

```
sns.distplot(data["Hour"])
```

```
<ipython-input-7-fe964bdaceca>:1: UserWarning:
```

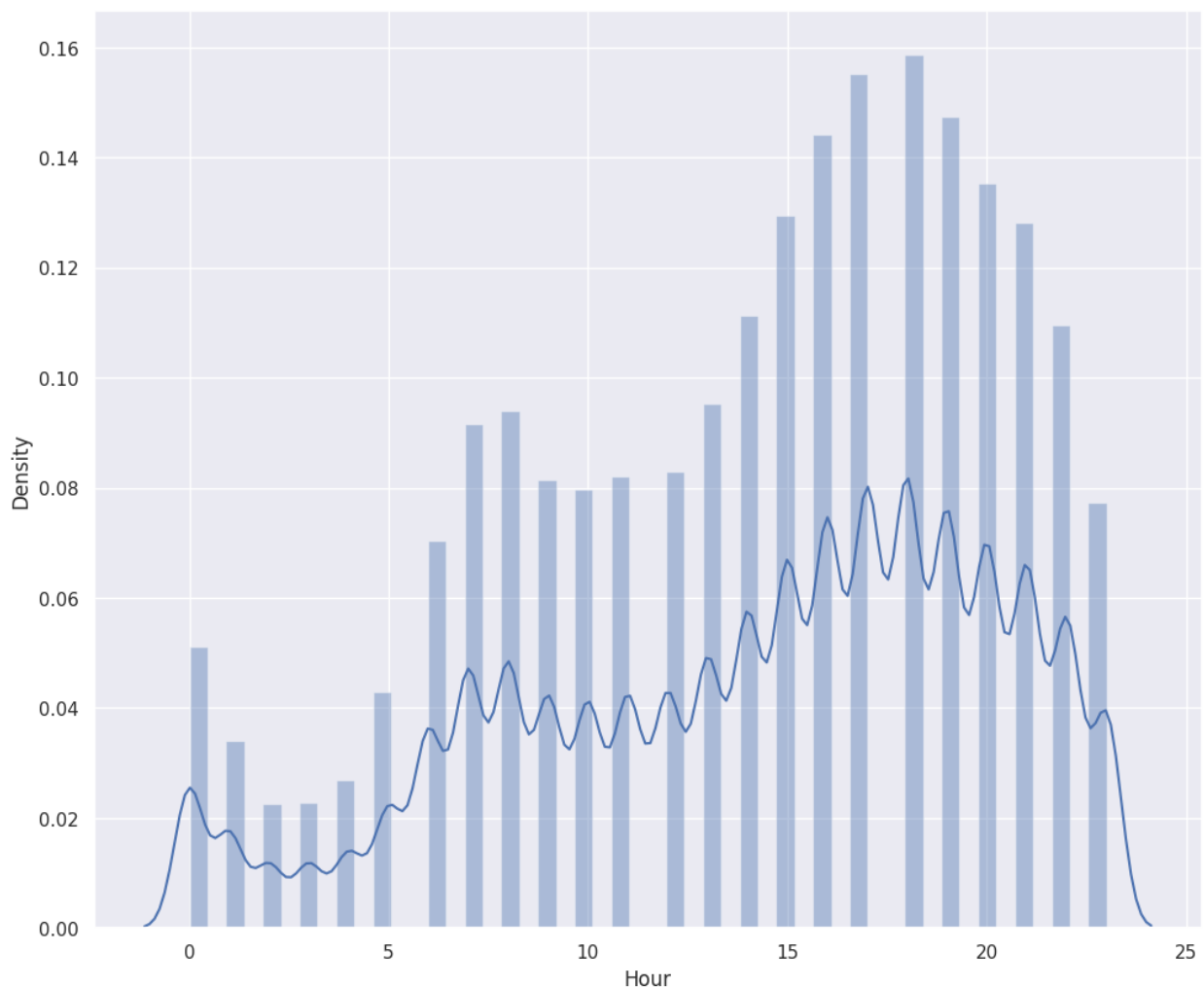
```
`distplot` is a deprecated function and will be removed in seaborn  
v0.14.0.
```

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(data["Hour"])
```

```
<Axes: xlabel='Hour', ylabel='Density'>
```



According to the hourly data, the Uber trips decreases after midnight and then start increasing after 5 am and the trips keep rising till 6 pm such that 6 pm is the busiest hour for Uber then the trips start decreasing.

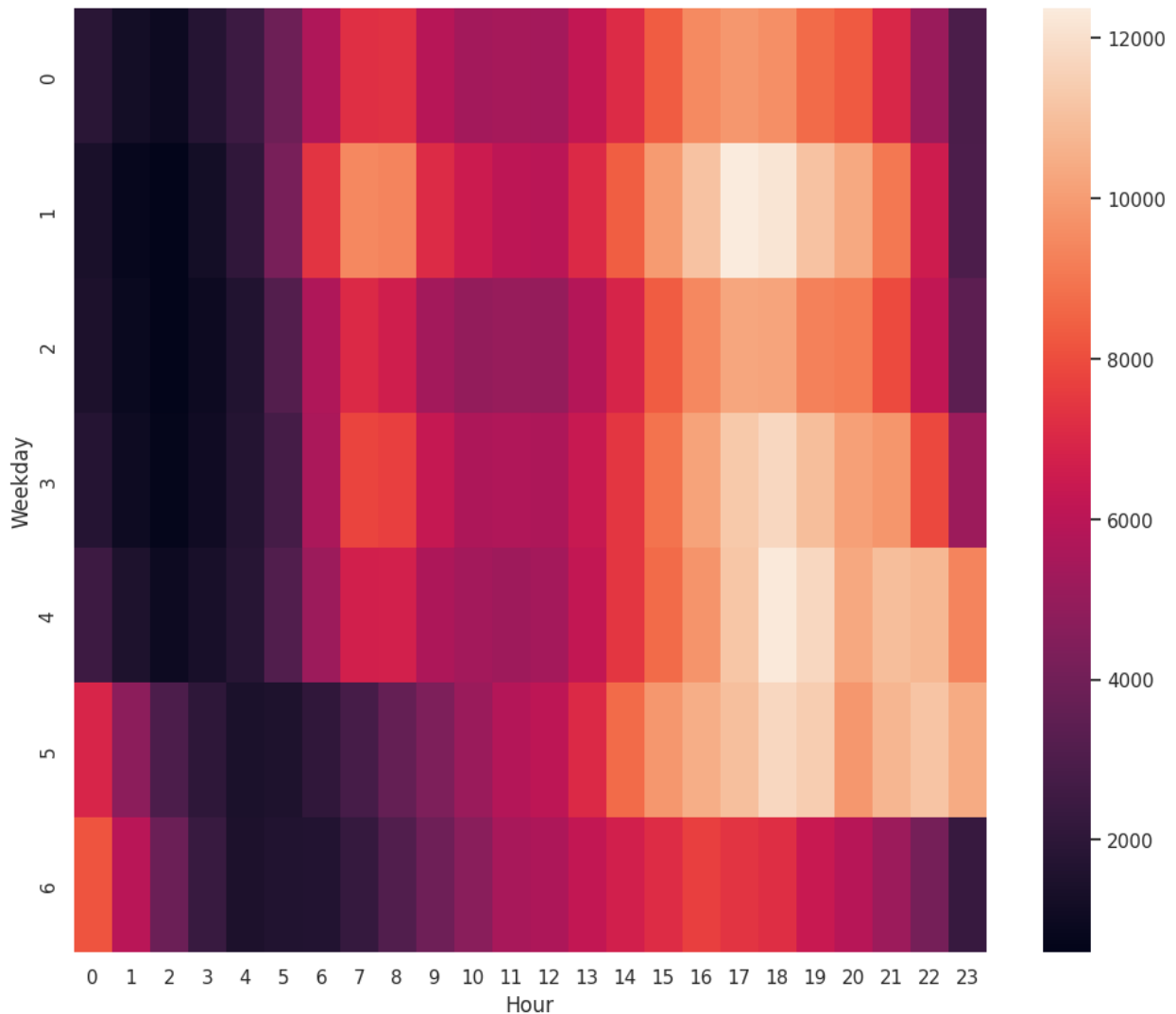
```
sns.distplot(data["Weekday"])
```

on Sundays the Uber trips are more than Saturdays so we can say people also use Uber for outings rather than for just going to work. On Saturdays, the Uber trips are the lowest and on Mondays, they are the highest.

correlation of hours and weekdays on the Uber trips

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

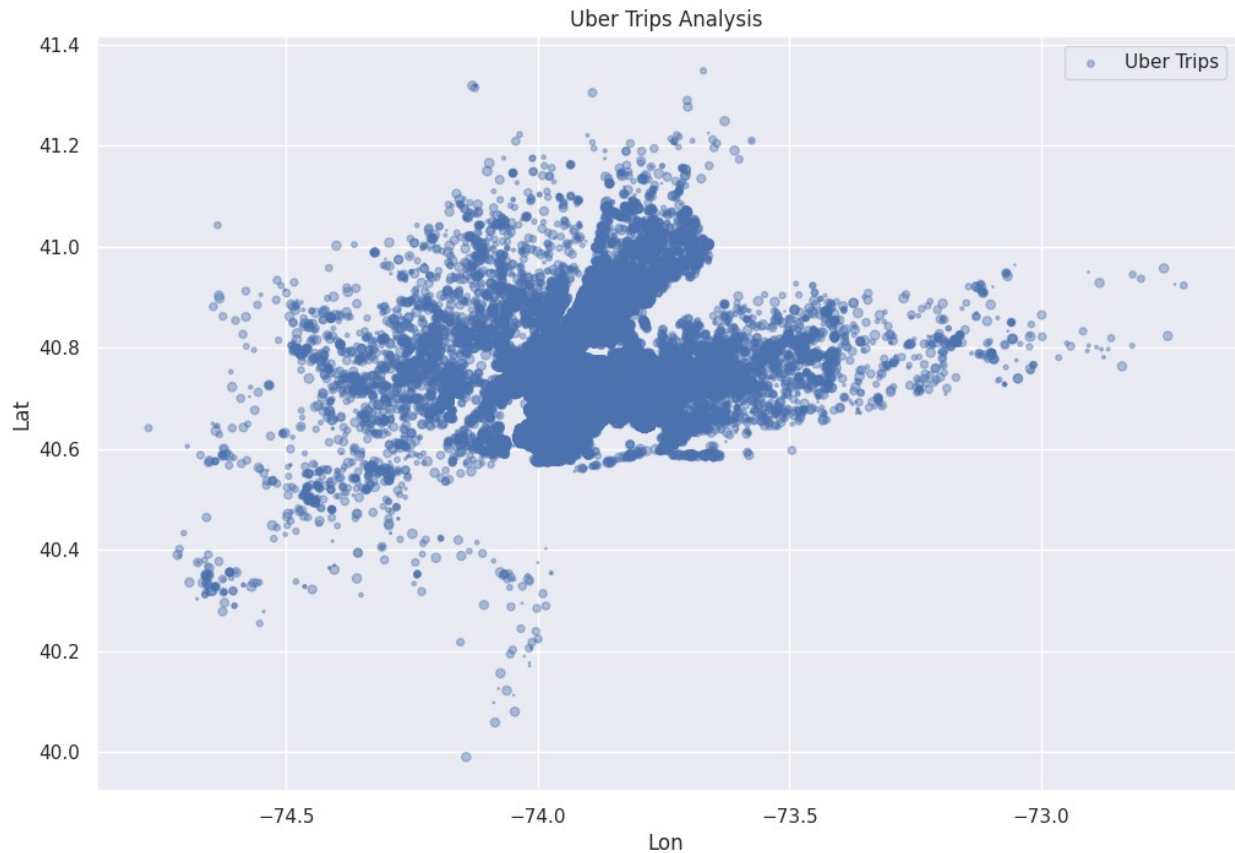
<Axes: xlabel='Hour', ylabel='Weekday'>
```



plot the density of Uber trips according to the regions of the New Your city

```
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()
```

```
/usr/local/lib/python3.10/dist-packages/pandas/plotting/_matplotlib/
core.py:1259: UserWarning: No data for colormapping provided via 'c'.
Parameters 'cmap' will be ignored
  scatter = ax.scatter(
```



## Summary

Monday is the most profitable day for Uber

On Saturdays less number of people use Uber

6 pm is the busiest for Uber

On average a rise in Uber trips start around 5 am.

Most of the Uber trips originate near the Manhattan region in New York.