

Uber Trips Analysis using Python

Importing the necessary Python libraries and dataset

```
In [4]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv(r"C:/Users/Arvind/Desktop/Uber_Trips_Analysis/Dataset/uber-trips.csv")
data["Date/Time"] = data["Date/Time"].map(pd.to_datetime)
data.head()
```

```
Out[4]:
```

	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

let's prepare the data to analyze the Uber trips according to days and hours:

```
In [5]: 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

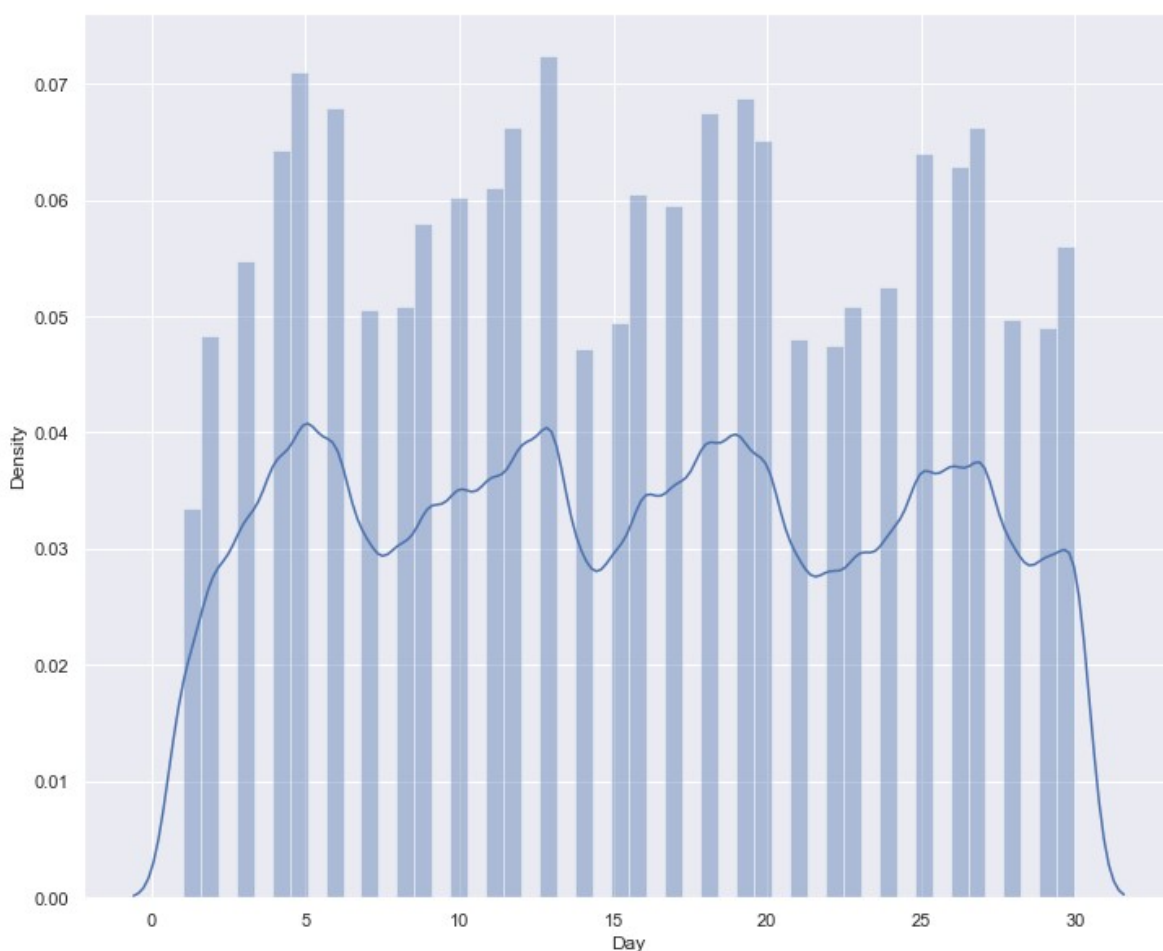
As this dataset contains Uber trips for the September month so let's have a look at each day to see on which day the Uber trips were highest:

```
In [6]: sns.set(rc={'figure.figsize':(12, 10)})
sns.distplot(data["Day"])
```

```
C:\Users\Arvind\anaconda3\lib\site-packages\seaborn\distributions.py:2551:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-le
vel function with similar flexibility) or `histplot` (an axes-level functio
n for histograms).
```

```
warnings.warn(msg, FutureWarning)
```

```
Out[6]: <AxesSubplot:xlabel='Day', ylabel='Density'>
```



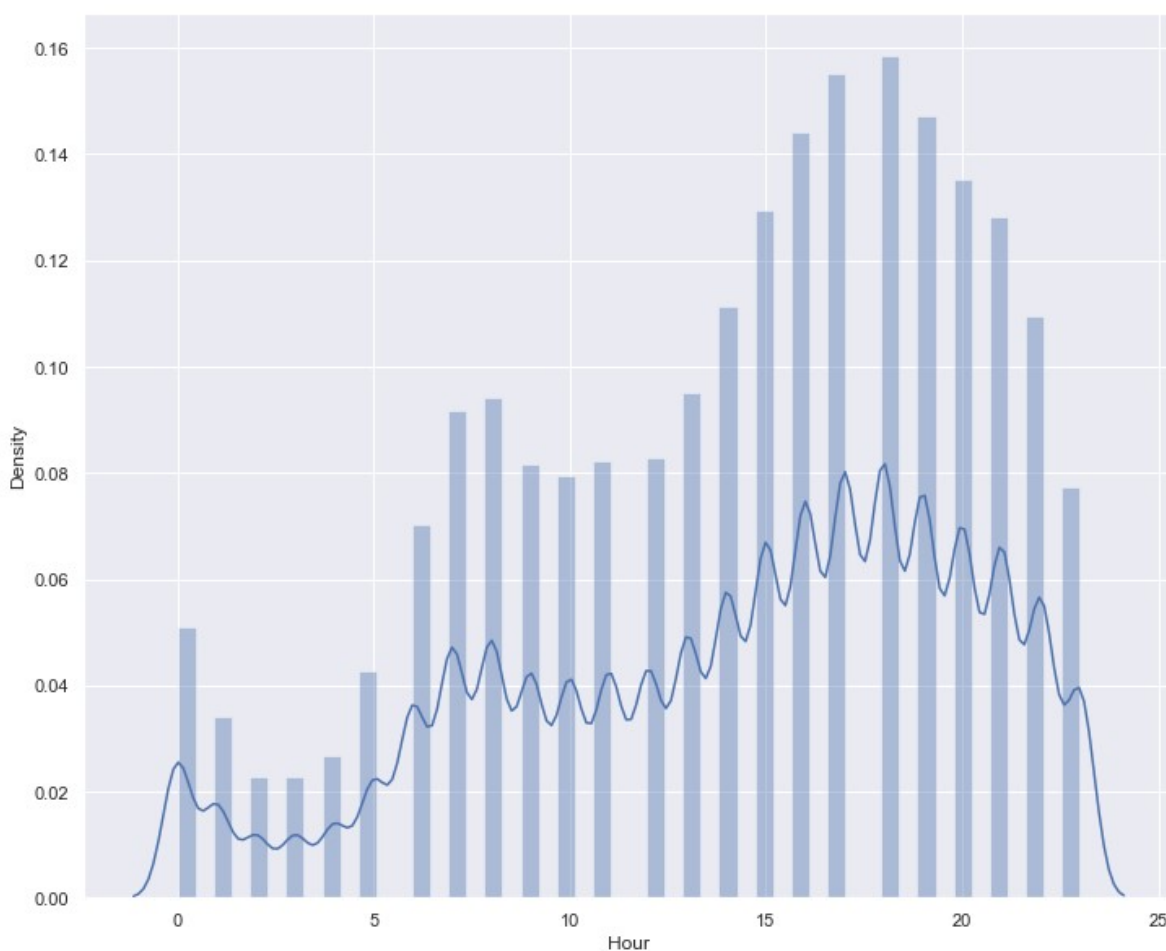
By looking at the daily trips we can say that the Uber trips are rising on the working days and decreases on the weekends. Now let's analyze the Uber trips according to the hours:

```
In [7]: sns.distplot(data["Hour"])
```

```
C:\Users\Arvind\anaconda3\lib\site-packages\seaborn\distributions.py:2551:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-le
vel function with similar flexibility) or `histplot` (an axes-level functio
n for histograms).
```

```
warnings.warn(msg, FutureWarning)
```

```
Out[7]: <AxesSubplot:xlabel='Hour', ylabel='Density'>
```



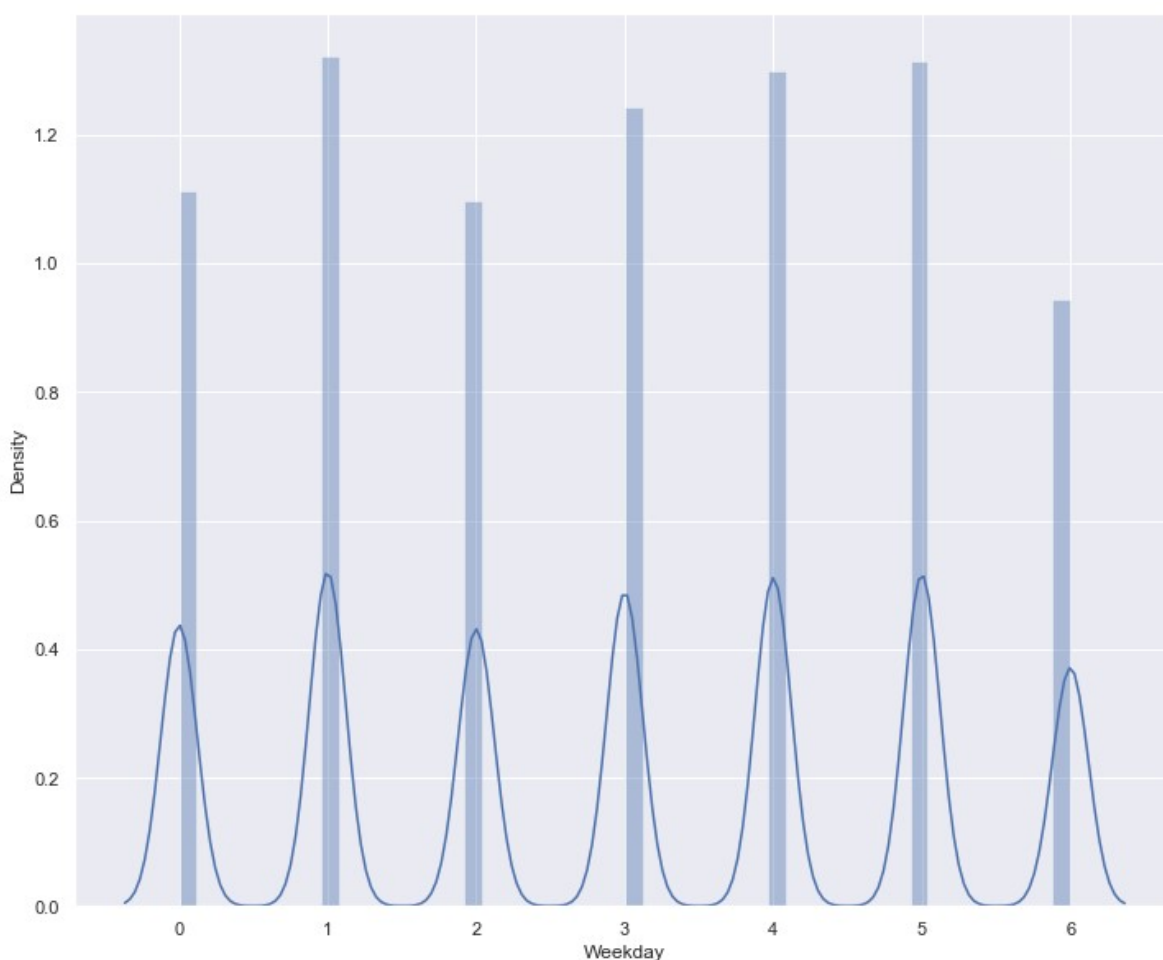
According to the hourly data, the Uber trips decreases after midnight and then start increasing after 5 am and the trips keep rising till 6 pm such that 6 pm is the busiest hour for Uber then the trips start decreasing. Now let's analyze the Uber trips according to the weekdays:

```
In [8]: sns.distplot(data["Weekday"])
```

```
C:\Users\Arvind\anaconda3\lib\site-packages\seaborn\distributions.py:2551:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-le
vel function with similar flexibility) or `histplot` (an axes-level functio
n for histograms).
```

```
warnings.warn(msg, FutureWarning)
```

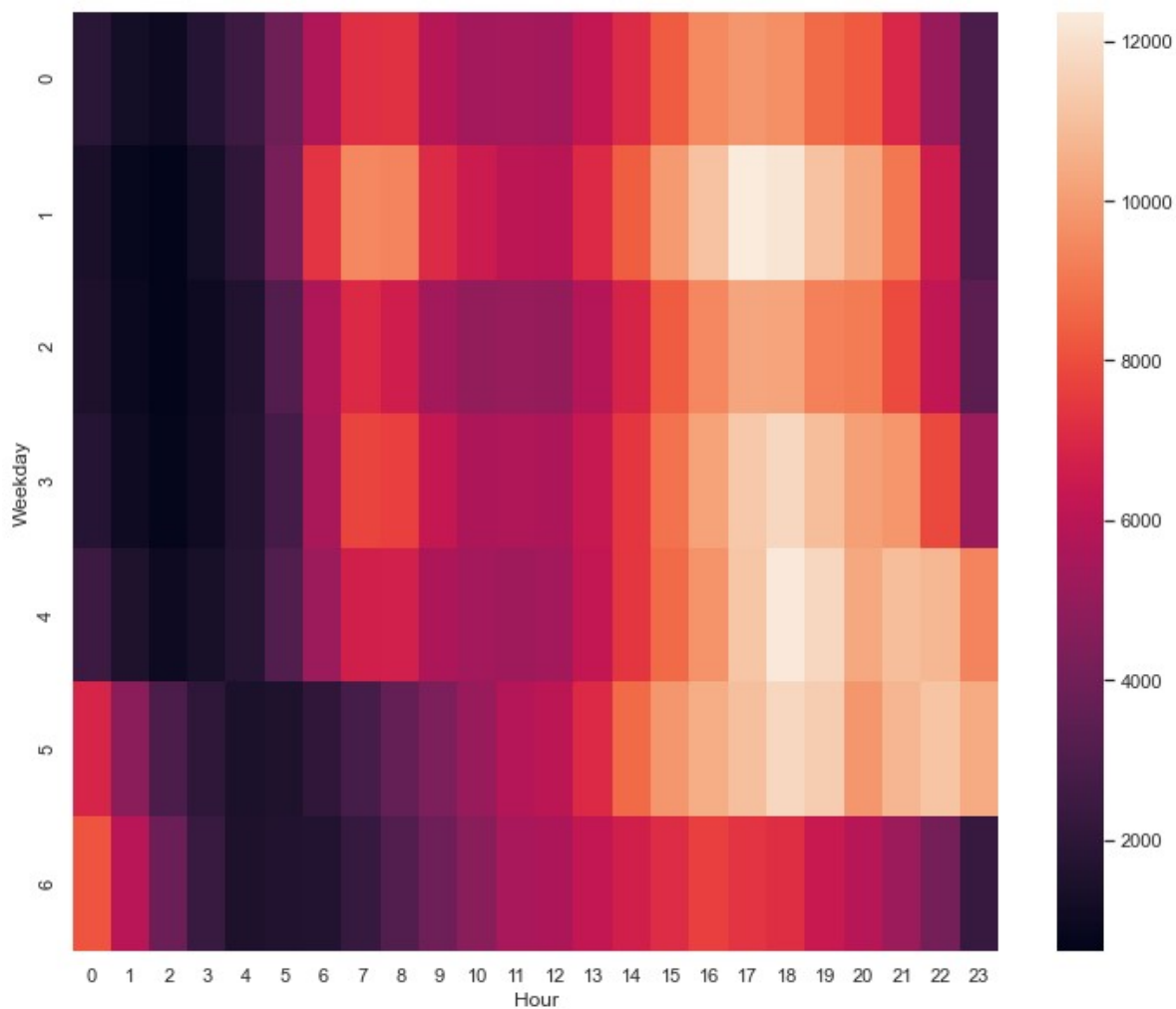
```
Out[8]: <AxesSubplot:xlabel='Weekday', ylabel='Density'>
```



In the above figure 0 indicates Sunday, 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. Now let's have a look at the correlation of hours and weekdays on the Uber trips:

```
In [9]: # Correlation of Weekday and Hour
df = data.groupby(["Weekday", "Hour"]).apply(lambda x: len(x))
df = df.unstack()
sns.heatmap(df, annot=False)
```

```
Out[9]: <AxesSubplot:xlabel='Hour', ylabel='Weekday'>
```



As we are having the data about longitude and latitude so we can also plot the density of Uber trips according to the regions of the New York city:

```
In [10]: data.plot(kind='scatter', x='Lon', y='Lat', alpha=0.4, s=data['Day'], label='Day')
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 value-mapping will have precedence in case its length matches with `*x*` & `*y*`. Please use the `*color*` keyword-argument or provide a 2-D array with a single row if you intend to specify the same RGB or RGBA value for all points.



Summary

So this is how we can analyze the Uber trips for New York City. Some of the conclusions that I got from this analysis are:

1. Monday is the most profitable day for Uber.
2. On Saturdays less number of people use Uber.
3. 6 pm is the busiest day for Uber.
4. On average a rise in Uber trips start around 5 am.
5. Most of the Uber trips originate near the Manhattan region in New York.