This data contains data about date and time, latitude and longitude, and a Base column that contains code affiliated with the uber pickup.

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
data = pd.read_csv("uber-raw-data-sep14.csv")
```

In [3]:
```
data.head()
```

Out[3]:

	Date/Time	Lat	Lon	Base
0	9/1/2014 0:01:00	40.2201	-74.0021	B02512
1	9/1/2014 0:01:00	40.7500	-74.0027	B02512
2	9/1/2014 0:03:00	40.7559	-73.9864	B02512
3	9/1/2014 0:06:00	40.7450	-73.9889	B02512
4	9/1/2014 0:11:00	40.8145	-73.9444	B02512

In [4]:
```
data["Date/Time"] = data["Date/Time"].map(pd.to_datetime)
```

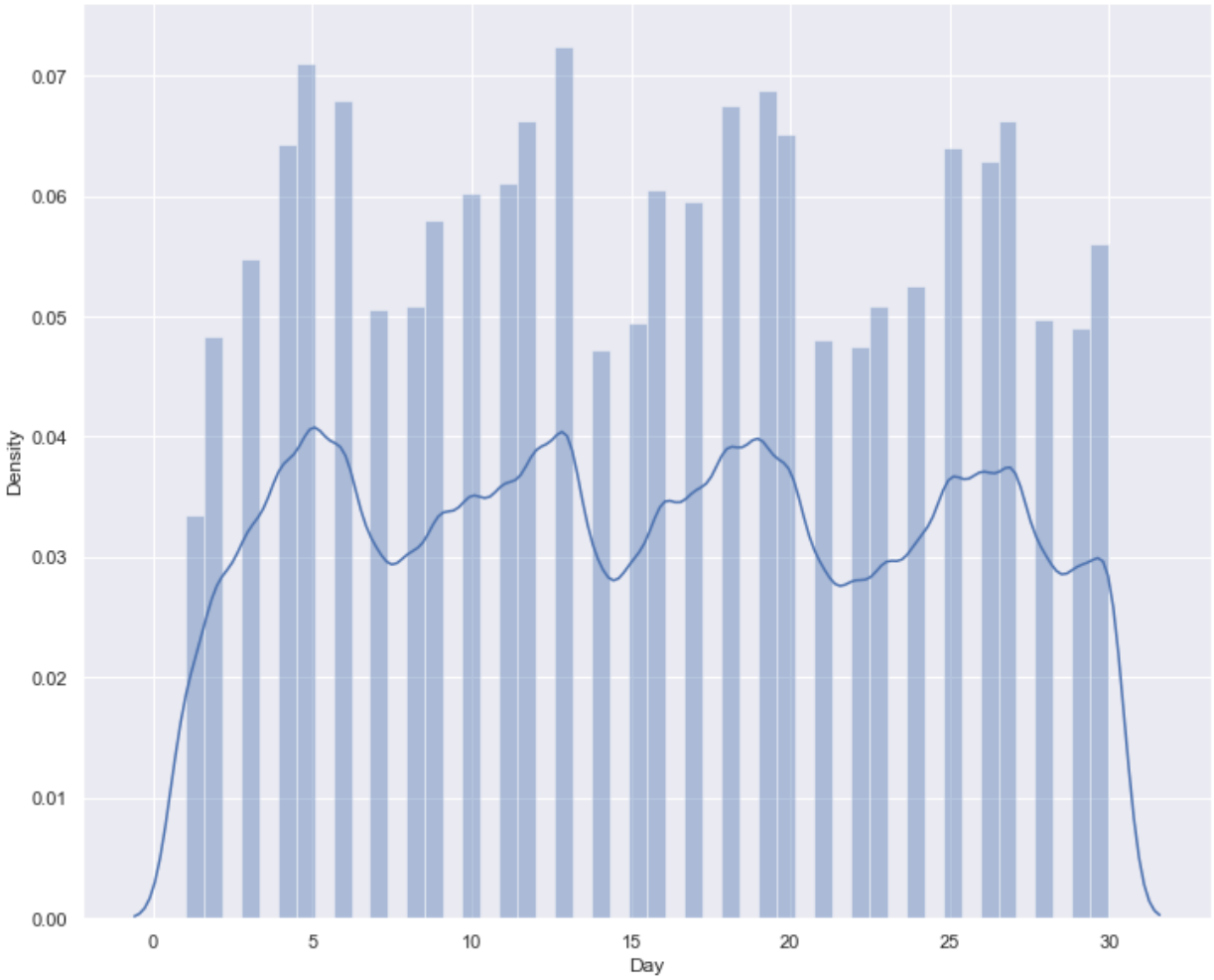
In [5]:
```
data["Day"] = data["Date/Time"].apply(lambda x: x.day)
data["Weekday"] = data["Date/Time"].apply(lambda x: x.weekday())
data["Hour"] = data["Date/Time"].apply(lambda x: x.hour)
print(data.head())
```

```
            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
```

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

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

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

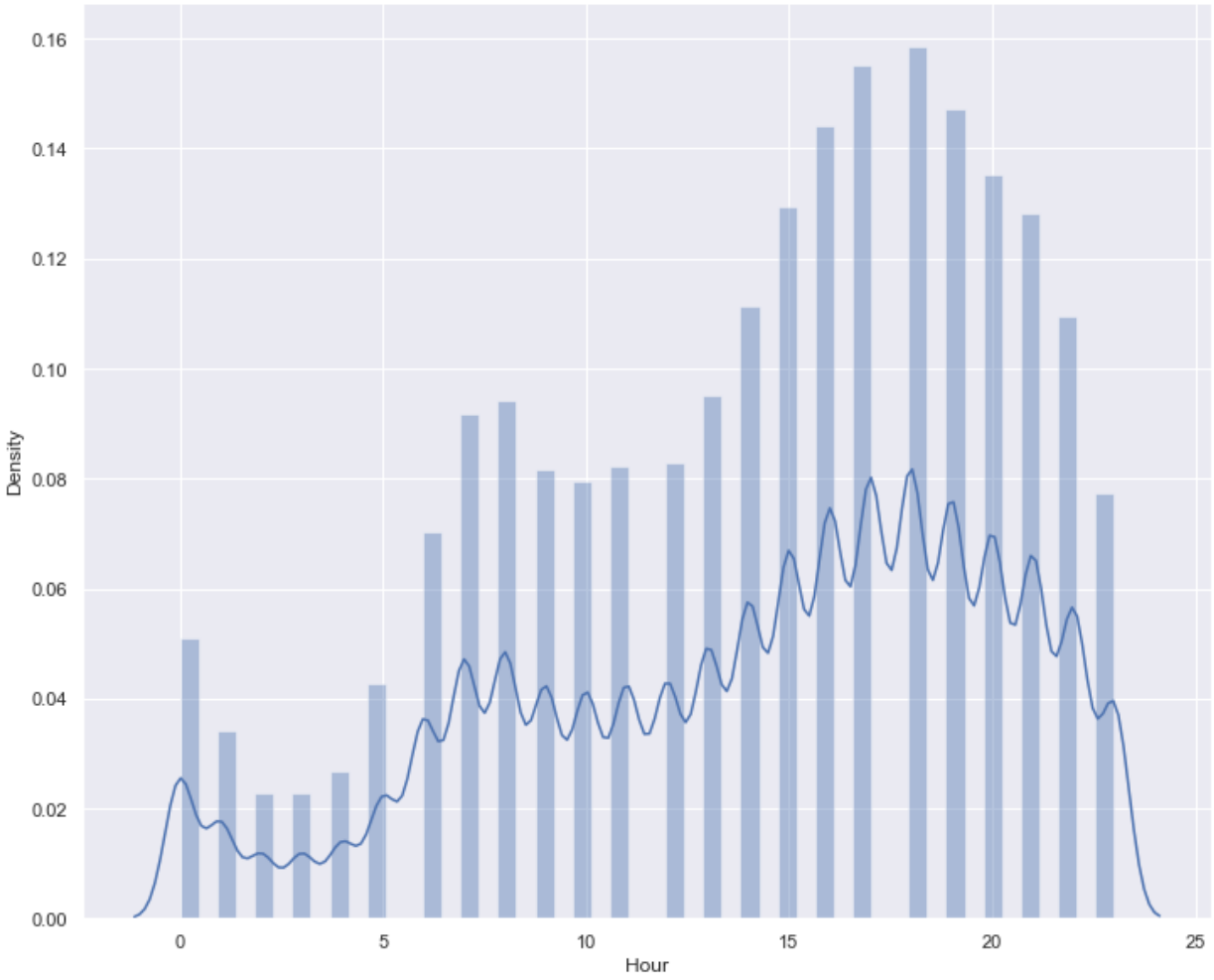


By looking at the daily trips we can say that the Uber trips are rising on the working days and decreases on the weekends. Now let's analyze the Uber trips according to the hours:

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

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

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

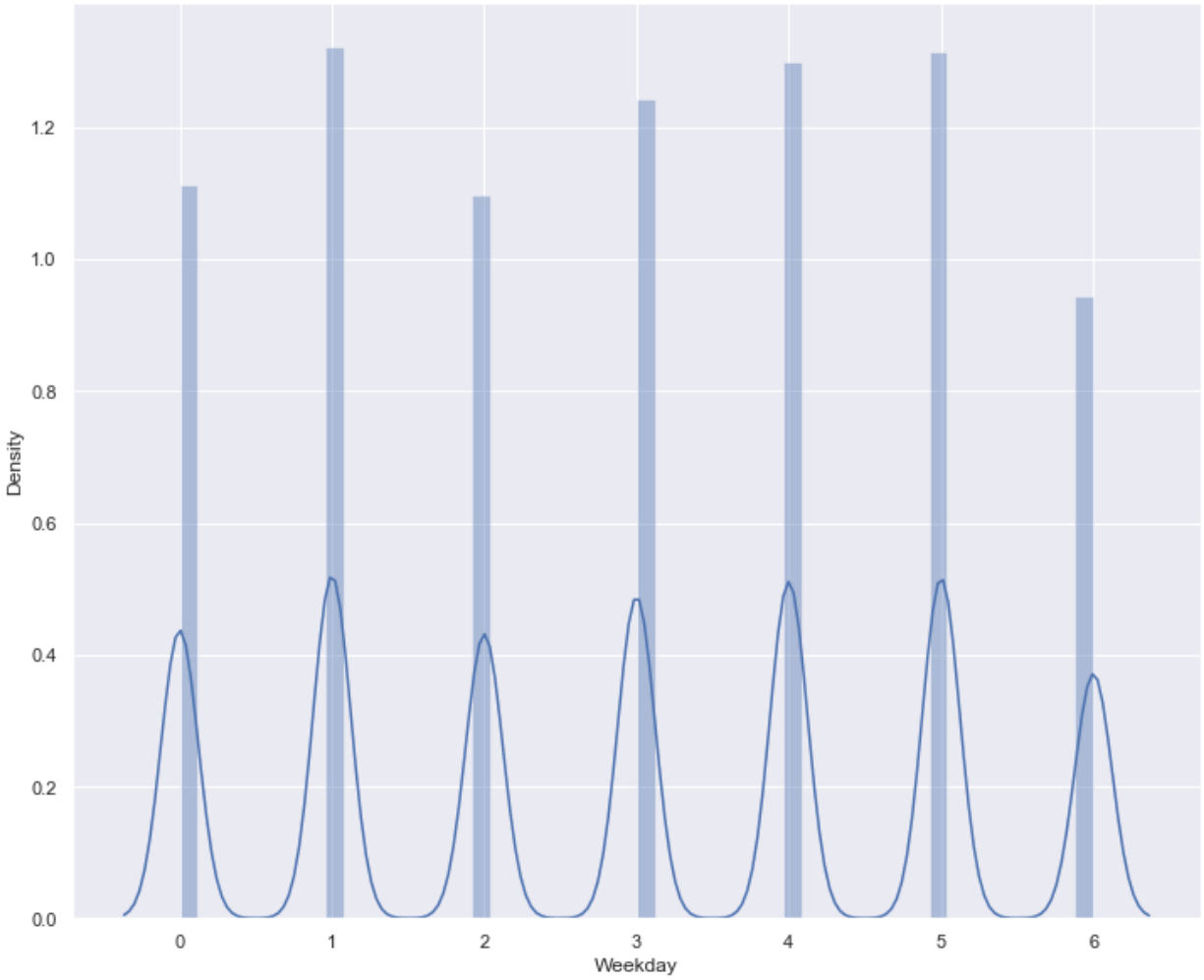


According to the hourly data, the Uber trips decreases after midnight and then start increasing after 5 am and the trips keep rising till 6 pm such that 6 pm is the busiest hour for Uber then the trips start decreasing. Now let's analyze the Uber trips according to the weekdays:

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

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

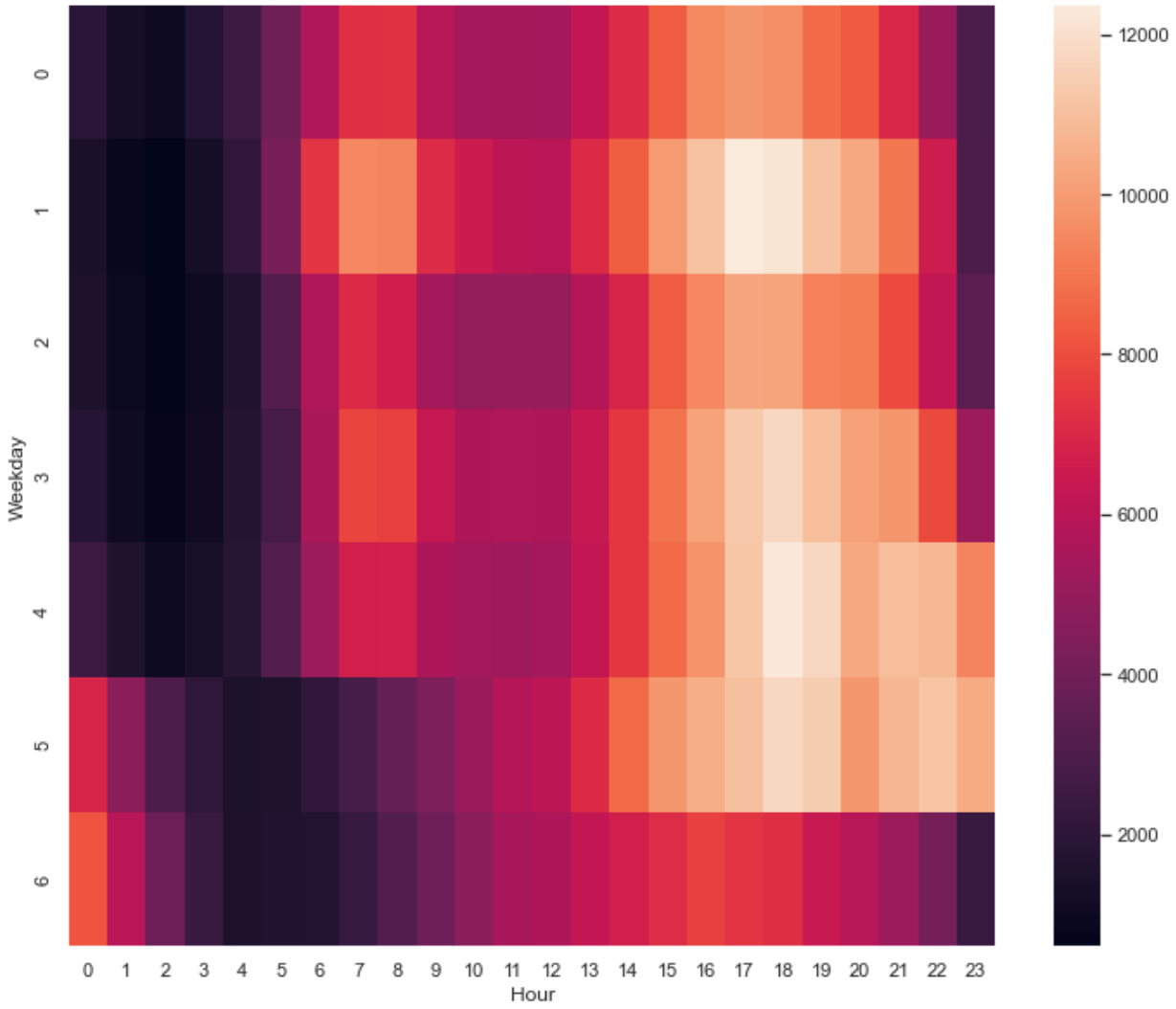
Out[8]:
```
<AxesSubplot:xlabel='Weekday', ylabel='Density'>
```



In the above figure 0 indicates Sunday, on Sundays the Uber trips and 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'>
```



In [10]:
```
# As we are having the data about longitude and latitude so we can also plot the densi

data.plot(kind='scatter', x='Lon', y='Lat', alpha=0.4, s=data['Day'], label='Uber Trip
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 avoide
d as value-mapping will have precedence in case its length matches with *x* & *y*.  P
lease 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.
```

In [11]:
```
# So this is how we can analyze the Uber trips for New York City. Some of the conclusi

# Monday is the most profitable day for Uber
# On Saturdays less number of people use Uber
# 6 pm is the busiest day 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.
```