INTRODUCTION

Problem:

Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred.

Based on these reports, Boston City police wants to determine areas where they can establish their new stations.
They want to cover the areas which have maximum crime rates in order to reduce the crime in the city and so that the police becomes more accessible to approach.

Aproach:

We need the following info-

What types of crimes are most common?
Where are different types of crimes most likely to occur?
Does the frequency of crimes change over the day? Week? Year?

I will be using Boston Crime Rate data set to solve this problem
https://www.kaggle.com/AnalyzeBoston/crimes-in-boston#crime.csv
Along with this foursquare api will also be used to determine the most dense locations of crime and visualize it .
Folium Maps along with foursquare api will be clubbed to visualize the areas so that the police can also understand it easily.

DATA DESCRIPTION

This dataset has 2,60,760 rows and 17 columns.

- INCIDENT_NUMBER:
- OFFENSE_CODE:
- OFFENSE_CODE_GROUP:
- OFFENSE_DESCRIPTION:
- DISTRICT:
- REPORTING_AREA:
- SHOOTING:
- OCCURRED_ON_DATE:
- YEAR:
- MONTH:
- DAY_OF_WEEK:
- HOUR:
- UCR_PART:
- STREET:

- LATITUDE:
- LONGITUDE:
- LOCATION:

The data consisted if many nan values as follows:

```
INCIDENT_NUMBER          0
OFFENSE_CODE             0
OFFENSE_CODE_GROUP       0
OFFENSE_DESCRIPTION      0
DISTRICT              1765
REPORTING_AREA           0
SHOOTING            318054
OCCURRED_ON_DATE         0
YEAR                     0
MONTH                    0
DAY_OF_WEEK              0
HOUR                     0
UCR_PART                90
STREET               10871
Lat                  19999
Long                 19999
Location                 0
dtype: int64
```

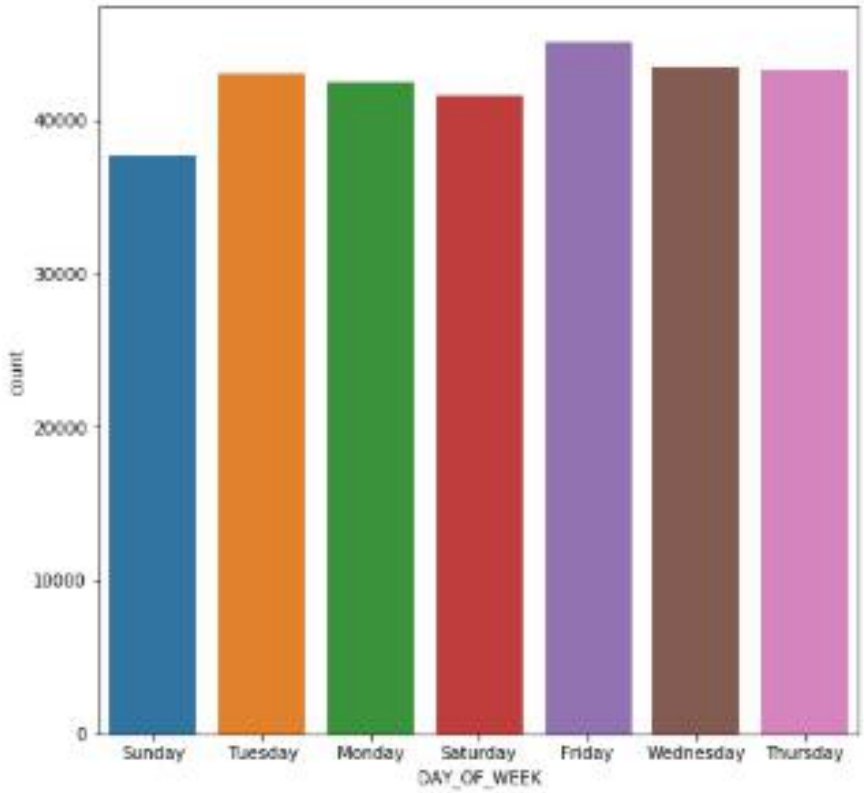
DATA CLEANING

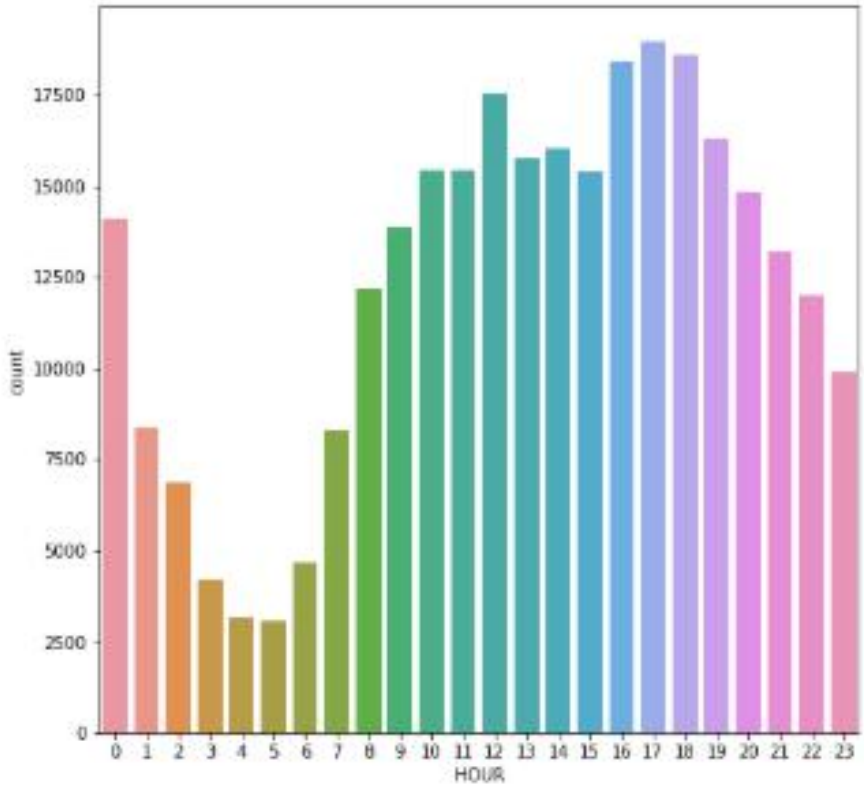
As shooting colojumn had max nan vlaues , we had to drop the coloumn. Also , drop all the rows with nan in them . After that our data statistics are as follows:

```
INCIDENT_NUMBER          0
OFFENSE_CODE             0
OFFENSE_CODE_GROUP       0
OFFENSE_DESCRIPTION      0
DISTRICT                 0
REPORTING_AREA           0
OCCURRED_ON_DATE         0
YEAR                     0
MONTH                    0
DAY_OF_WEEK              0
HOUR                     0
UCR_PART                 0
STREET                   0
Lat                      0
Long                     0
Location                 0
dtype: int64
Shape: (296573, 16)
```

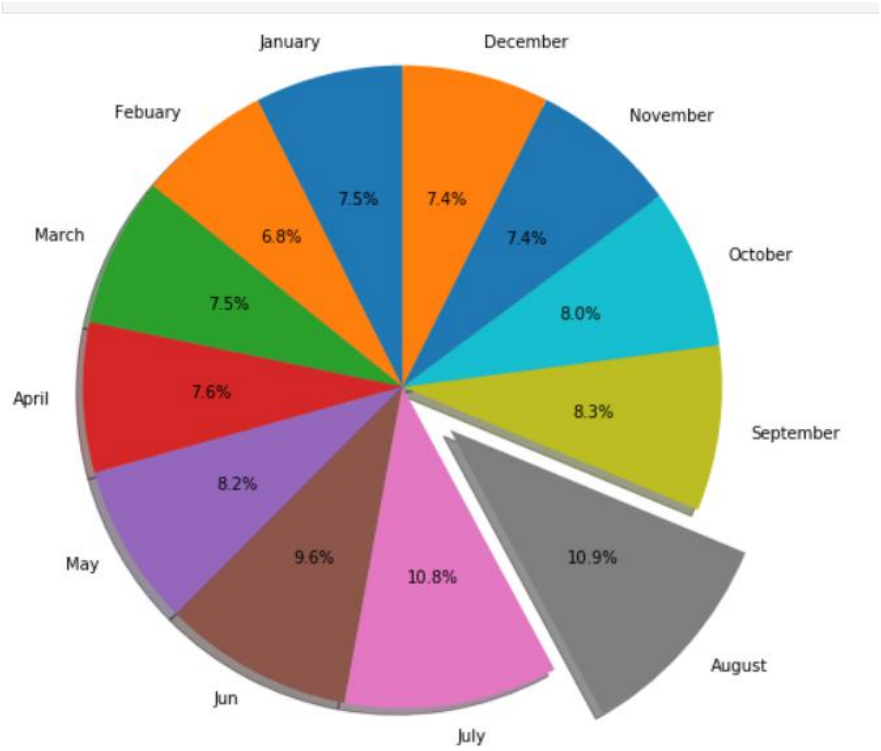
DATA VISUALIZATION

```
plt.show()
```



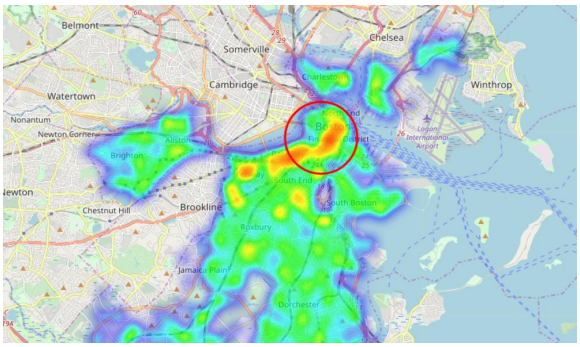


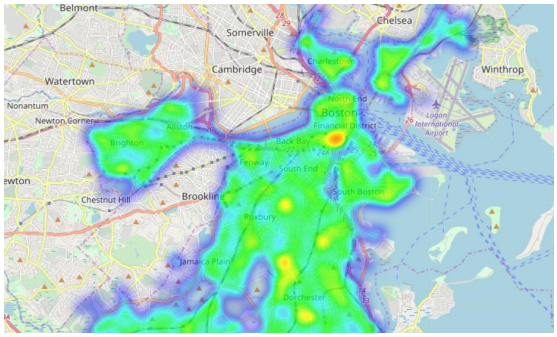

```
plt.figure(figsize=(8,8))    #first size of figure
```

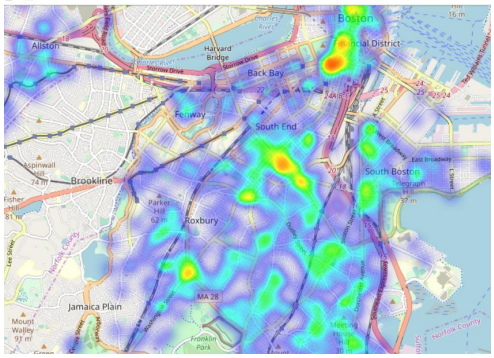


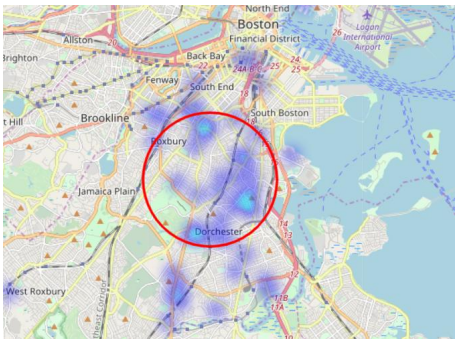
We observe that maximum no. Crimes occurred in:
August
On Fridays
In evening Hours

Model and Predictions









Logistic Regression:

Mean squared error: 0.12%
Variance score: -0.13%
Coefficients: [[-0.0006886 -0.00095075 0.00526656 0.0193502 0.00230191]]
Intercept: [-3.47607224e-06]

KNN:

Best % correct: 30.50%, for n_neighbors=2

Conclusion:

The Boston police can take the analysis from the heatmaps plotted and can decide the areas to setup special statons .
They have to be most active during weekends and evening hours.