PREDICTION OF SHOOTING INCIDENTS IN BOSTON

Author: HARSH NATVARLAL SINDHWA

NORTHEASTERN UNIVERSITY



PROBLEM STATEMENT

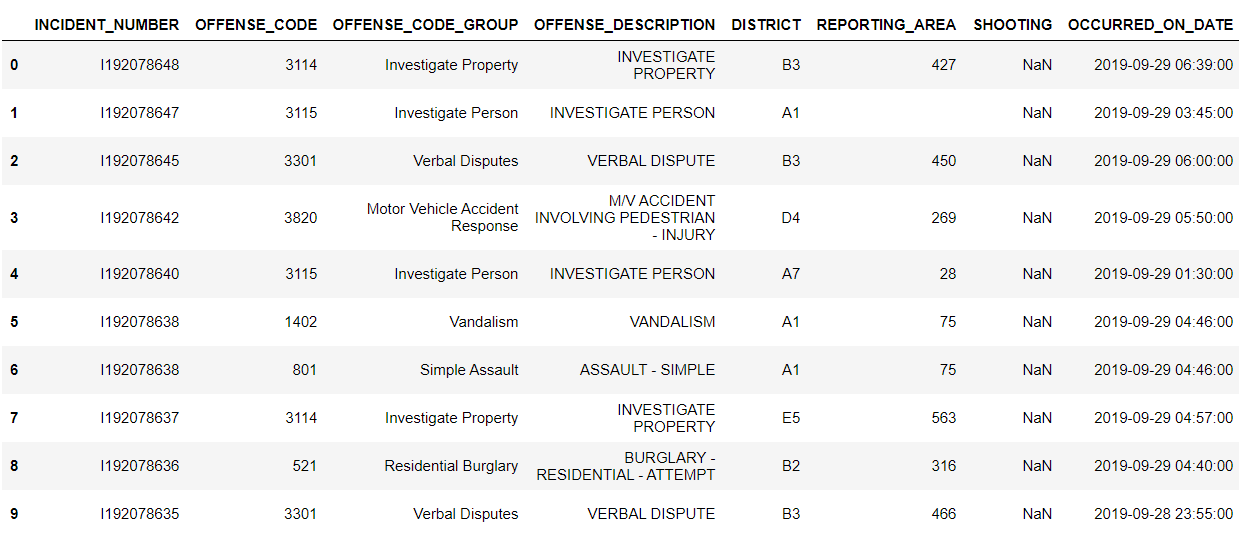
A crime dataset of Boston city was recorded from 2015 till the 9th month of 2019. The crime dataset had many shooting incidents in each year. A model was implemented to try and predict how many shooting incidents could happen in the 2019 year.

DATA EXTRACTION

The data is read using panda’s library and the first 10 records are been shown. The data is shown up as a Data Frame.



Printing the first ten records of dataset

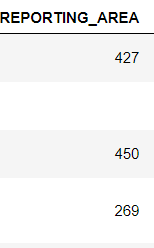


DATA CLEANUP

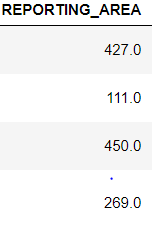
The data cleanup deals with cleaning of the dataset that is filling the missing values, removing some values or selecting a particular subset of dataset.

The dataset has few missing values. So, I initially filled the blank spaces in dataset with ‘NaN’. Later I used imputer function from sklearn library. The imputer function is used to fill the missing values. There are 3 strategies in imputer function viz mean, median and most frequent (mode). I used mode strategy to fill missing values.

Before:

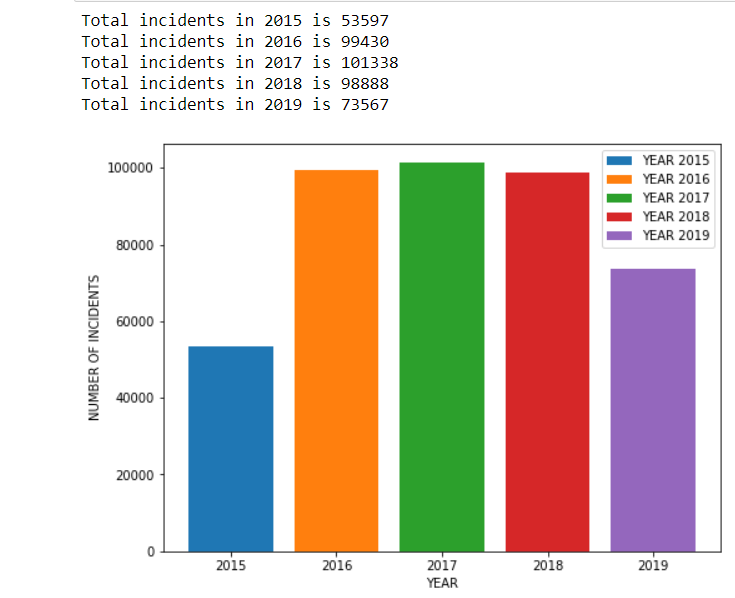


After:

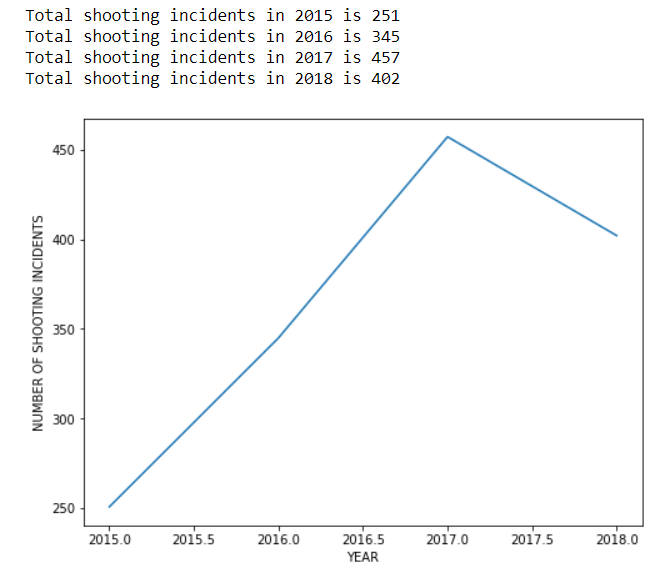


DATA VISUALIZATION

1. A function was created to find the number of incidents happened in each year. The 2019 year had a data up to the 9months. The total number of incidents were plotted versus each year. As, you can see the crime rates increased significantly from 2015 till 2018. The recent incidents till the 2019 were recorded up to 9 months and plotted.



1. The crime incidents had many shooting incidents and plotted the number of shooting incidents vs each year. The shooting incidents have increased from 2015 to 2017 consistently. Now, I will try to predict the shooting incidents in 2019.



DESCRIPTIVE ANALYTICS

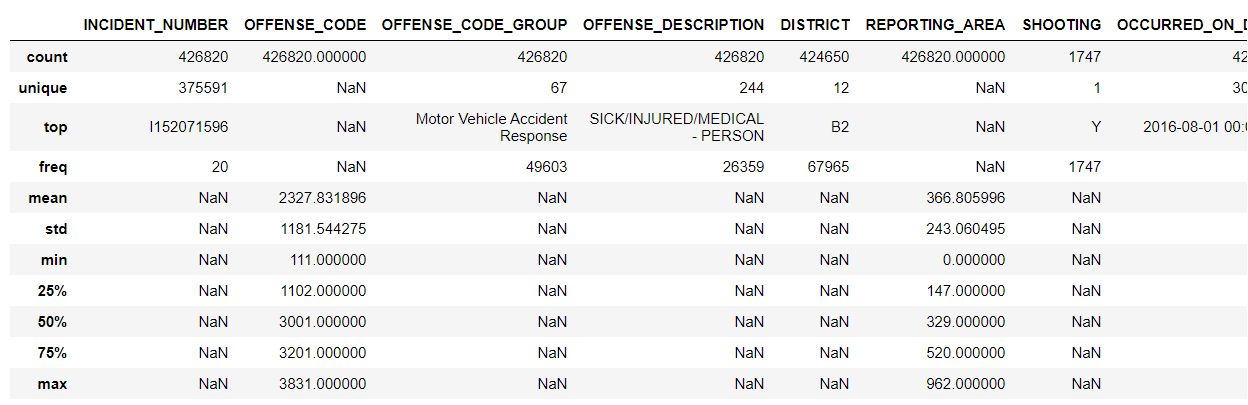
1) The street where the crime incidents happen are seen at the Washington Street.

2) There were about 426820 crime incidents from 2015 to 2019.

3)Out of all crime incidents most crime incidents fall in Motor Vehicle Accident Response.

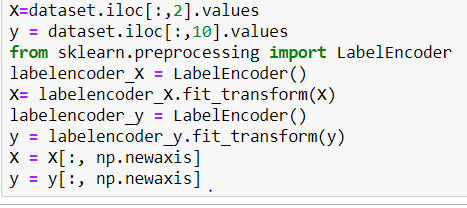
4) Most crime incidents have happened on Friday.

5) Many incidents happened had no record of what hour of the day it had happened.

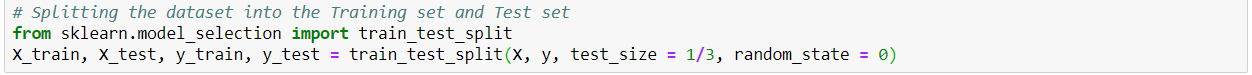


PREDICTIVE ANALYTICS

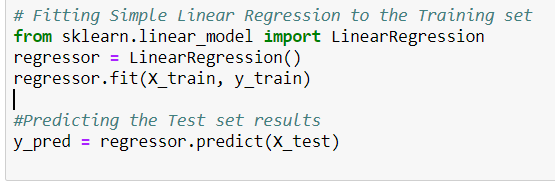
Initially, I tried to show prediction of on what day of week a particular offence can happen. I divided the columns in the variables X (dependent variable) and y (independent variable). X had the column values of Offence Code Group and y had the column values of the day of week.



Here, I converted the days of week i.e seven days in numerical values 0-6. Dataset had 67 unique offence code groups which were numbered from 0-66 using **Label Encoder.**

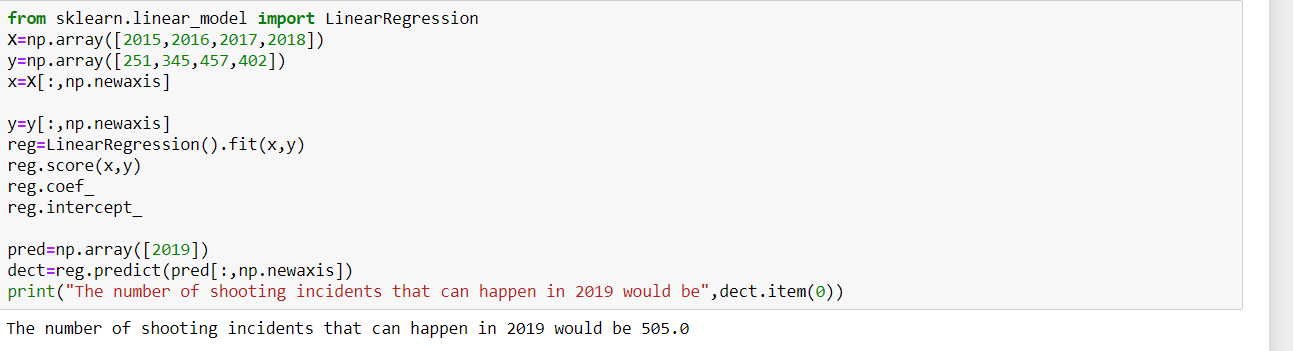


Divided the X and y variables were divided into training and testing variables. The test variables were 33% (i.e 1/3) of the whole variables used.



Applied Linear Regression model from the sklearn library to the training set.

The prediction of number of shooting incidents that can take place in 2019 was done using linear regression model. The values of X were the years 2015,2016,2017,2018. The values of Y were the count of shooting incidents in the years that we calculated in the data visualization part. The pred variable had been assigned as the value of the year which we need to predict that is 2019. The dect variable has the count of shooting incidents that can happen in 2019.



REFERENCES

1. Dataset source: dataset.boston.gov

Click [here](https://data.boston.gov/dataset/6220d948-eae2-4e4b-8723-2dc8e67722a3/resource/12cb3883-56f5-47de-afa5-3b1cf61b257b/download/tmplczp2c4s.csv) to download the dataset