# TABLE OF CONTENT :

1. **Summary**
   1. **About the data**
   2. **Business Question**
   3. **What are we solving for**
2. **Data Cleaning**
3. **Exploratory Data Analysis**
   1. **Statistical**
   2. **Visual**
4. **Analysis:**
   1. **Decision Tress**
   2. **Random Forest**
   3. **Gradient Boosting**
5. **Conclusion**
6. **References**

**Summary About the Data:**

Here we have a 2022 dataset of the crime that is committed in the city of Boston with respect to the area based on the longitude and the latitude of the that location. The type of crime, the time of the crime and the location of the crime are the key factors that would be considered for our analysis. The data has 63,367 rows and 17 attributes. Link to the data: https://data.boston.gov/dataset/crime-incident-reports-august-2015-

to-date-source-new-system/resource/313e56df-6d77-49d2-9c49-ee411f10cf58

# Business Question:

Here in the main Business question that comes to mind is how we would be able to segregate the different areas in Boston based on the severity of the crimes that are being committed. The data gives us an understanding of time and latitude and longitude, where the crime is being committed and what type of crime is being committed.

# What are we solving for?

Recommendations to the nearby services on how we could reduce these crime rates. What steps should be taken to hinder these crimes that are committed.

# Data Cleaning:

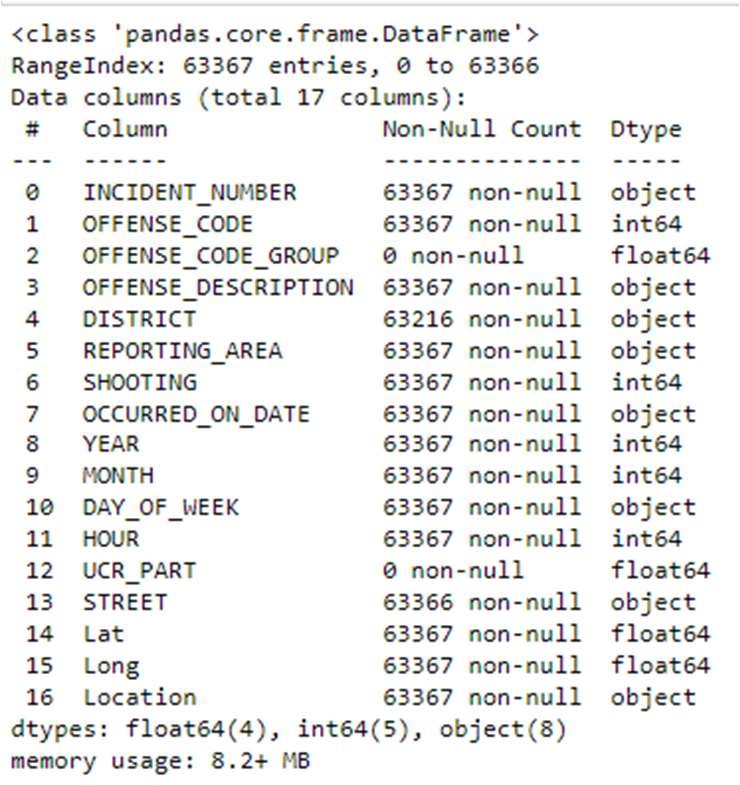


Figure 1 Null Columns

As we can see, the data which we have has a few columns that are null. Since these columns don’t have a huge significance and there won’t be any useful information that we can extract from these columns we can drop them. After dropping these columns, we move on to cleaning the columns individually.

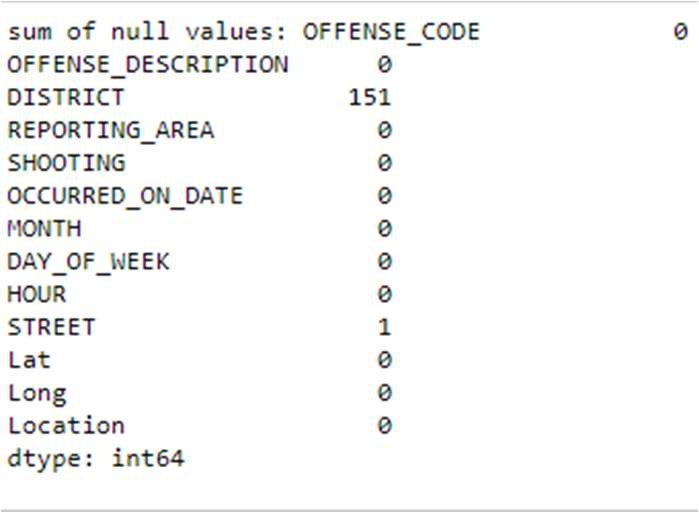


Figure 2 Null Values

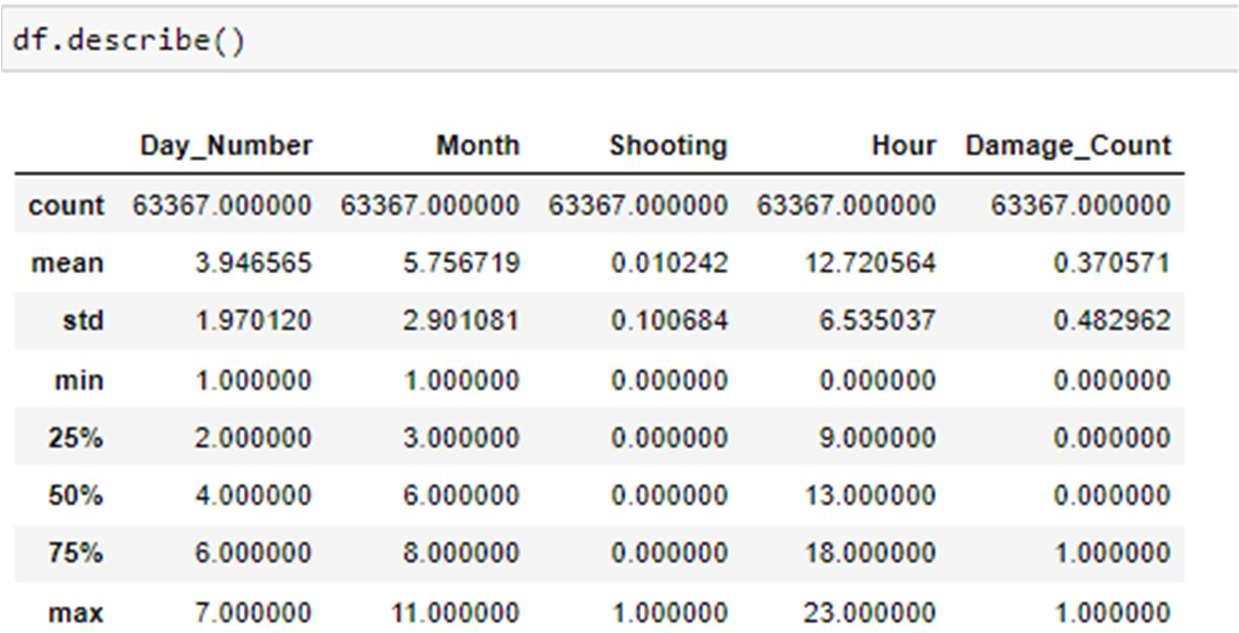
As we can see, the columns ‘Street’ and ‘Reporting Area’ have a few null values in them. These null values must be dealt with as it would give incorrect information regarding how different variables interact with each other. The number of null values is small hence, we can drop these values as they don’t have a significant impact on the

data.

# Analysis

Statistical Analysis:

For understanding the data, we conduct basic statistical and visual analysis. The initial analysis allows us to get an idea on how the data is trending and what we could understand about the data on a preliminary level.



Since the data has mostly categorical variables the inference of those values is better understood visually than statistically. With the one that could be understood statistically, we are able to observe that on average most of the crimes that are committed are in the 5th month of the year during the midday.



Figure 3 Correlation table

The above correlation table gives us an understanding of how closely related the

variables are to each other. This allows us to get a small idea on how these variables might affect each other to an extent that they might influence each other. By visual understanding we observe that there is low correlation between all the variables hence there isn’t any significant effect of the variables of each other. This correlation table is better elaborated with the help of a correlation heatmap.

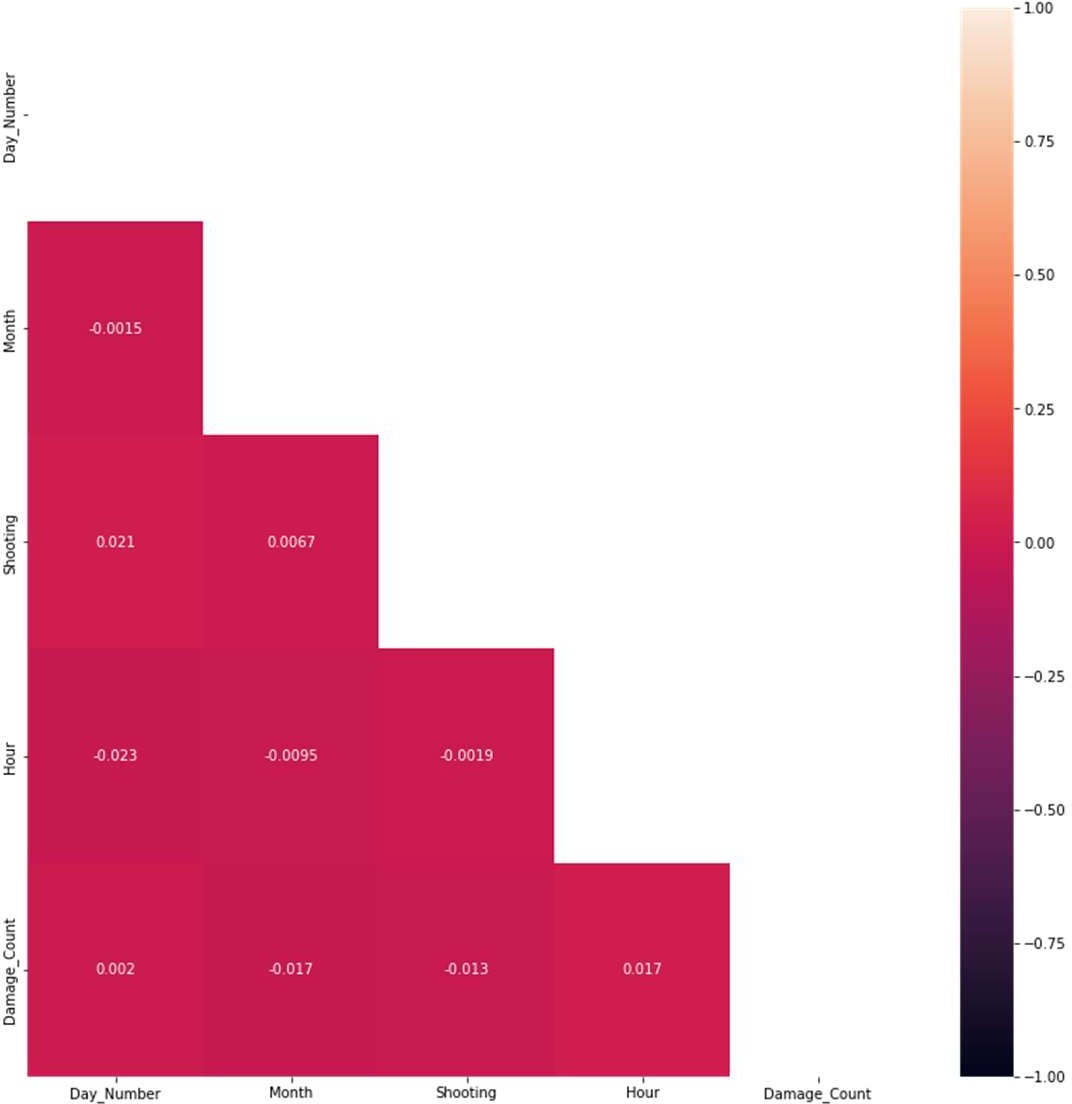


Figure 4 Correlation Heatmap

# Visual Analysis:

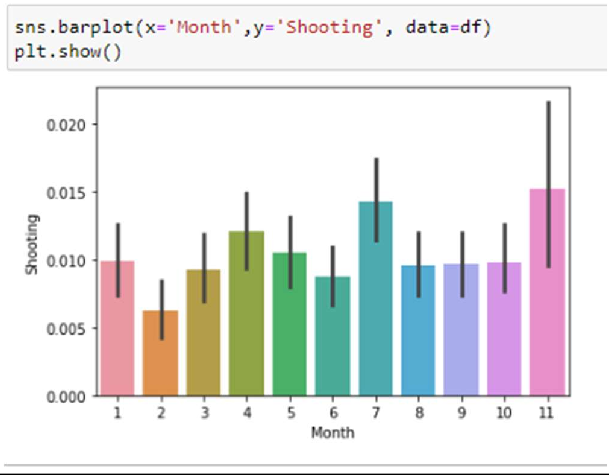


Figure 5 Bar plot for months

The bar plot talks about the number of gun violence that are being committed on monthly bases. The 11th month or November has the highest number of gun violence cases which might be dude to the sale offers which are committed by juveniles due to sales. Such was a case in Dec 8th, 2022, mentioned BPD. Our Data is real time hence the incidences for December are not mentioned.

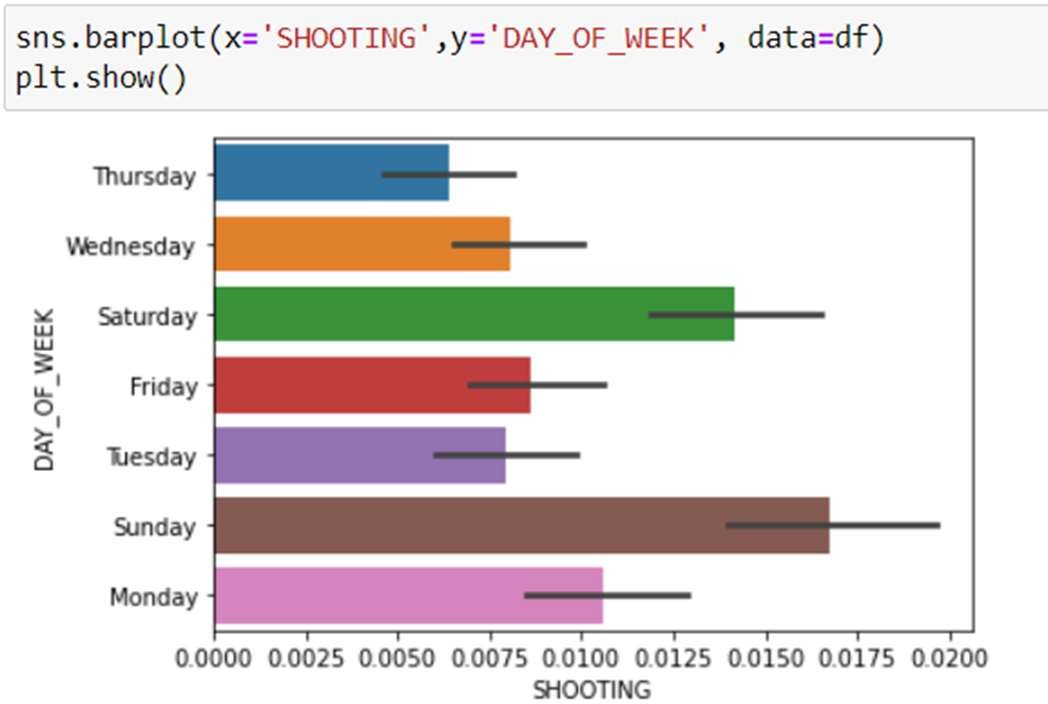


Figure 6 Bar plot for days of the week

The violence that occurs on different days of the week is discussed in the above-bar graphic. Therefore, it is evident from the graph above that most violent incidents have taken place on the weekends. Sunday and Saturday. The cause seems obvious—these are the days of the week when most people do not have any kind of office work, and as a result, most of them are outside their homes, increasing the likelihood that an offense will occur.

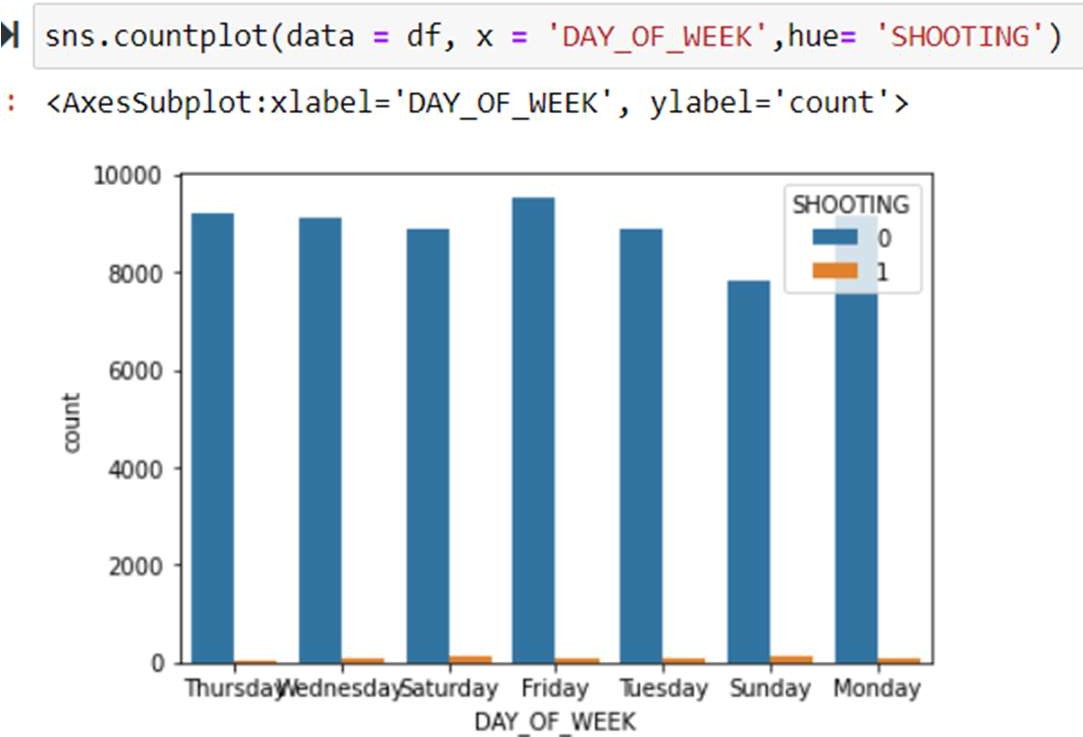


Figure 7 Histogram for days of the week

The relationship between the numbers of shootings that occur in a week is depicted in the above graph. Here, it is evident that the majority of shootings did not take place in

the Boston area. Hence. It is apparent that the majority of crimes that have been committed in Boston's streets do not involve gunshot. The orange bar, which is roughly equal to 0 on every day of the week in 2022, represents all the days where shooting has taken place.

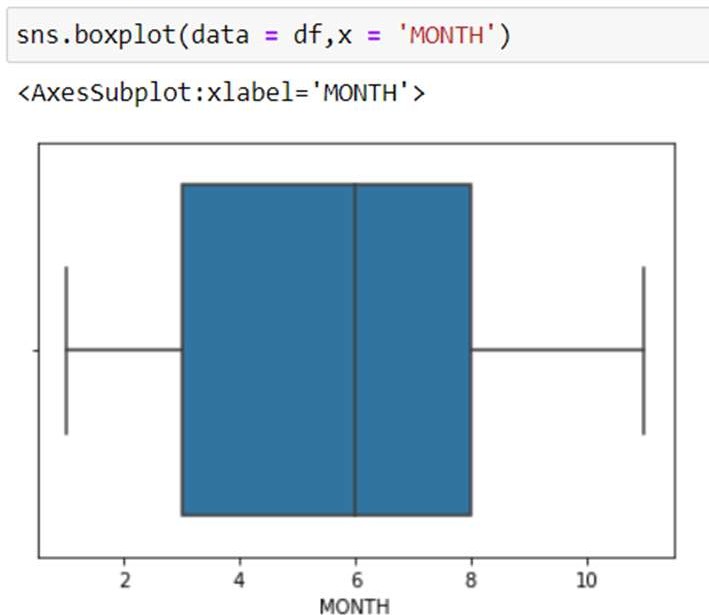
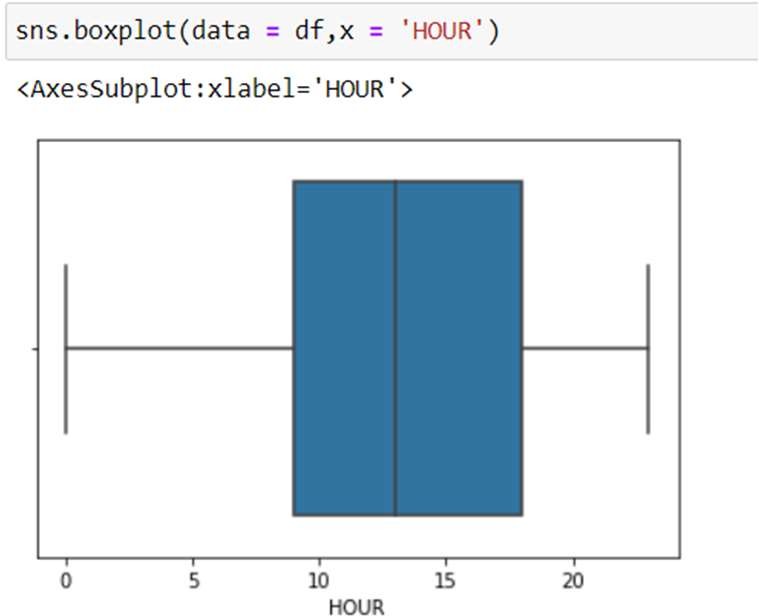


Figure 8 Hours Boxplot Figure 9 Months Boxplot

The two box plots of Hours and Months shown above were created just to look for any data outliers. Therefore, it is evident from the above plots that our dataset is free of outliers. Additionally, the average number of hours was 13 and the hours in which the offense occurred ranged from 9 to 17 in the morning.

Furthermore, it can be deduced from the second boxplot that the majority of the offense is likely to have occurred in June, which is the sixth month of the year. Additionally, the range of months is from the third (March) to the eighth (August) (August).



Figure 10 Top 20 Violation

The above bubble chart gives us an understanding of how the different violation are being committed based on the streets that are in Boston. The reason we chose a bubble chart is that it allows us to see how influentail is a particular violation is based on the streets around it.

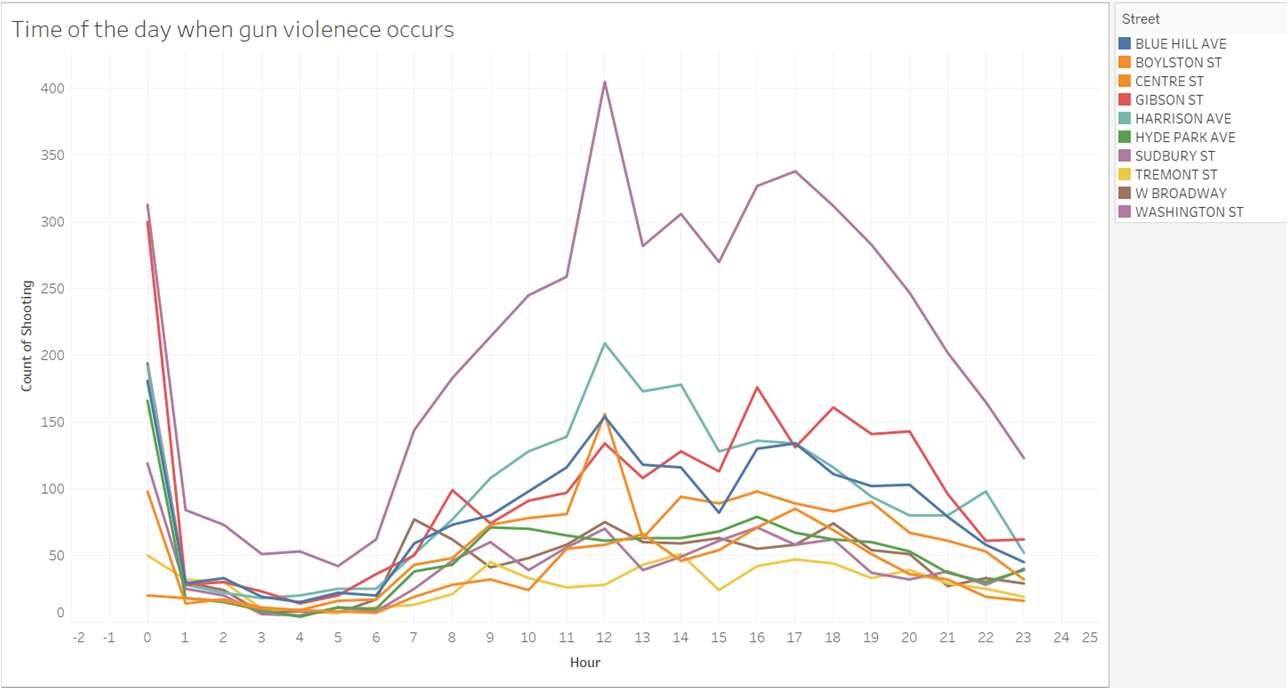


Figure 11 Hour that has most Violations

Here, we have a line chart that provides us with information regarding how gun violence is affecting the citizens of Boston. A line chart is optimal as it allows us to give an understanding of at which hour of the day these violations are occurring and on which street the violations are occurring.

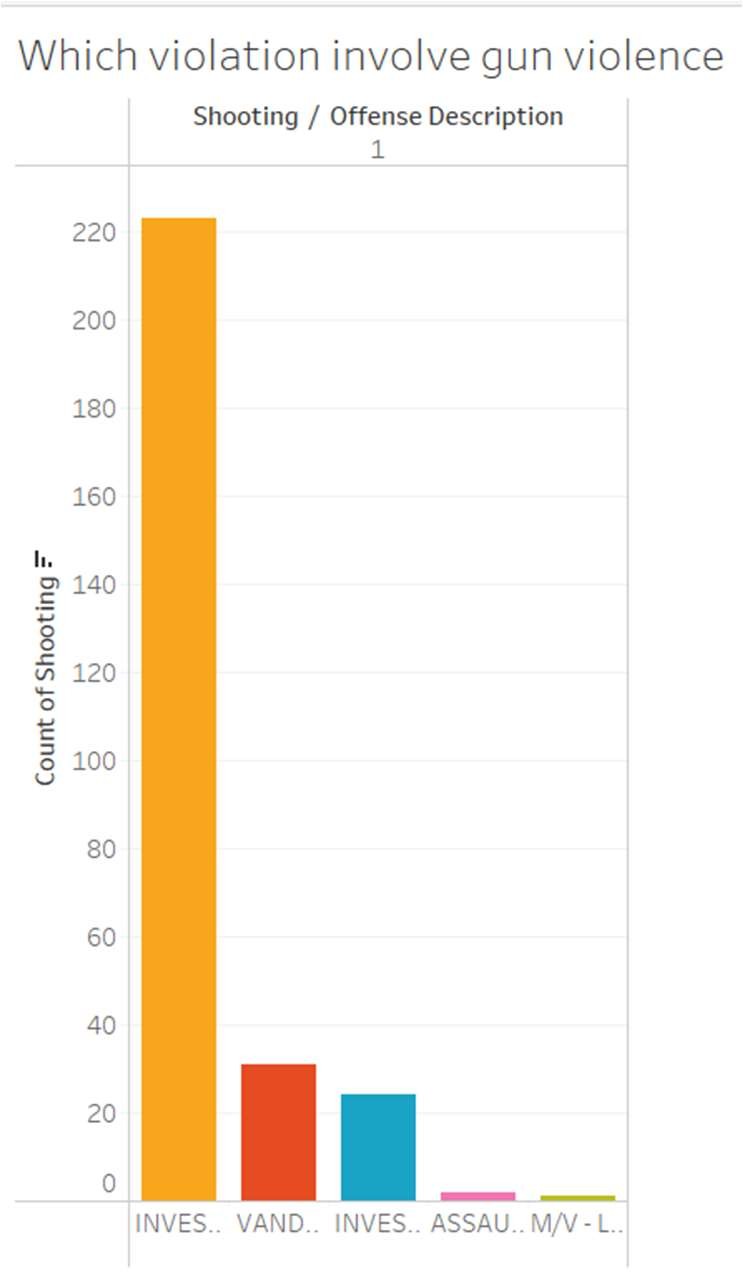


Figure 12 Crimes Involving Gun violence

The graph talks about the different crimes that have the invlovement of gun violence. This allows us to get an understanding of which particular violation has the involvement of gun violence. Here, investigation of the propetry has the highest gun violence.



Figure 13 Map of the streets with gun violence

In the map of Boston, we see that blue hill Ave has the highest gun violence that spread from the north to the south of Boston.

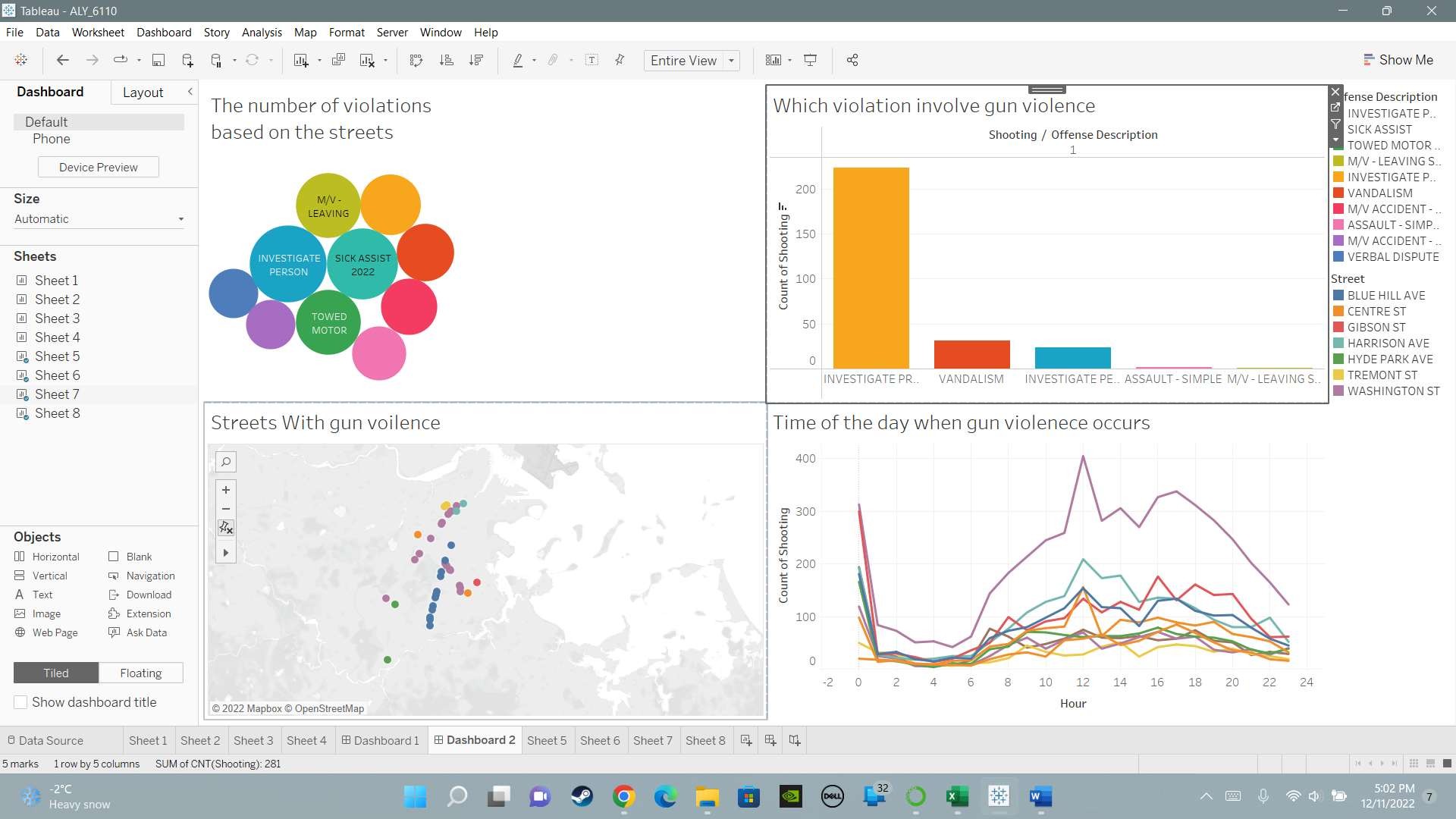


Figure 14 Shooting that have occurred

Here, in the figure we have created a dashboard that talks about the different crime in

Boston that involve shooting. We have created a bubble chart that talks about the number of shootings that have occurred Boston based on the crime that has been

committed. We observe that the shootings that have occurred are highest when there is a form investigation that has been conducted by the police. The next chart talks about the street that has the highest number of cases of gun violence based on the number of streets. Washington street has the highest number of gun violence cases. This might be as Washington street spread over a large distance through Boston. We show this with that we are able make a map that shows better understanding on how the spread of the gun violence is through Boston. The line chart talks about the different times at which these guns violence occur and through our initial analysis we see at the 12th hour of the day there are highest number of gun violence which points to midday.

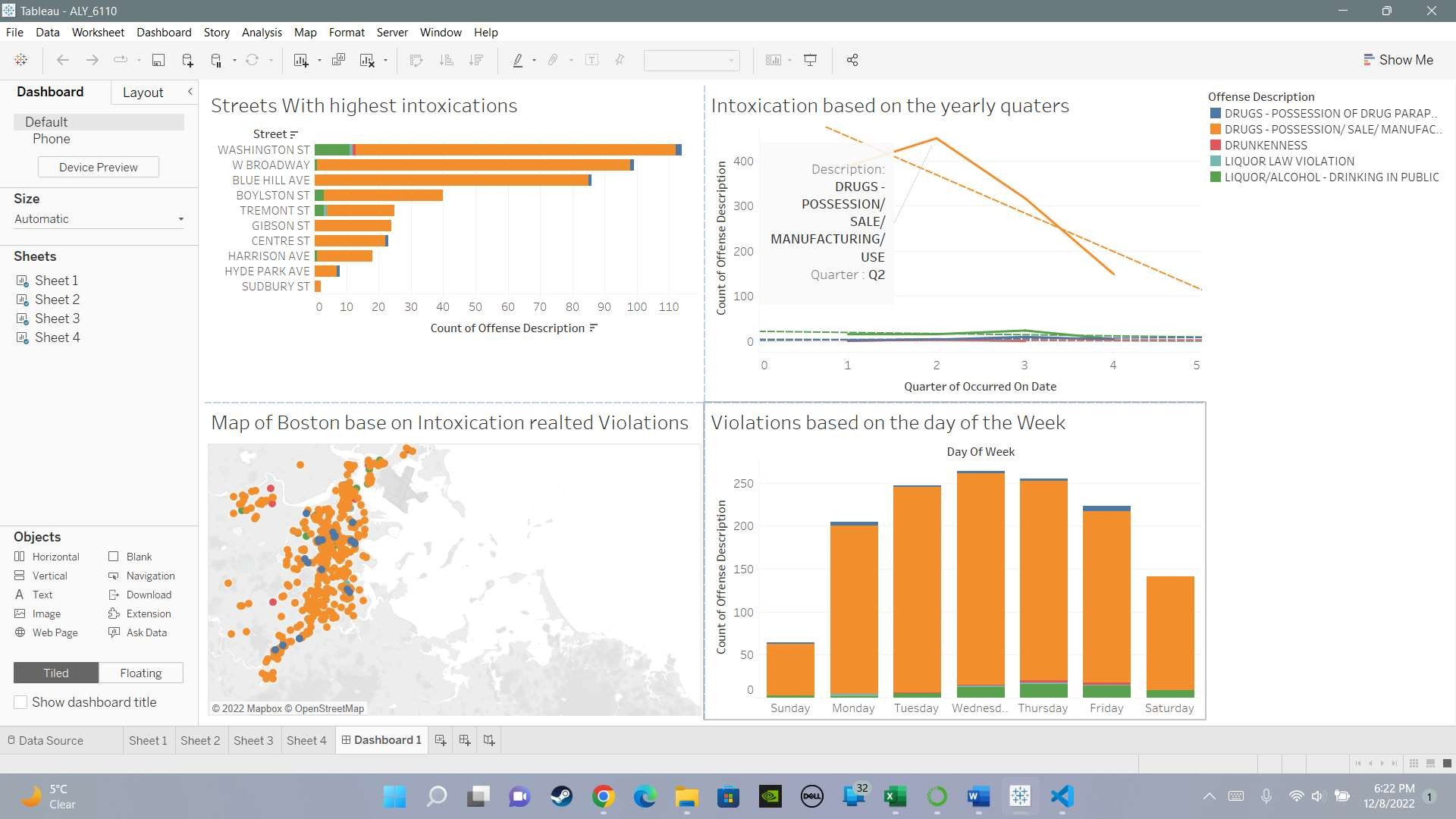


Figure 15 Intoxication Crimes Committed

Hear in the above dashboard we can understand that there are several crimes that are being committed in Boston. The dashboard talks about the different crimes that are being committed related to whether the person is involved with drugs or alcohol. This

is important to understand as Mass is a legal state. This means that there would be a higher chance that other harder drugs can also be caught but the police. Based on our visual analysis we see that Most of the crime that have been recorded are criminal those who are under the possession of drug. As we can see the number people that have been caught by the police are centered towards the 2 quarter of the year which means from April to June. During the spring season the chances of civilian wandering outside is higher as compared to in the months of winter. As the number of people get out after a while there are higher chance of them being caught with certain unauthorized substances. Apart form this a point to keep in mind is that, In the year 2022 the government of Massachusetts lifted its covid restriction where people did not have to wear a mask in public areas. This might be a major reason why there is sharp decline in the succeeding quarter where the number of cases reported have reduce as the months go on. As the crimes a spread across Boston, Washington Street has been a place where the highest number of intoxication related problem have risen. In the western part of Boston, Washington Street begins in Brookline Village, and crosses Brookline, Allston, Brighton, Newton, and Wellesley, ending at Wellesley's boundary with Natick .Apart from this the street covers areas of Downtown Boston ,Roxbury and the Hyde Park neighborhood, this sums up to a total of 36.97 miles. Along this path a majority of drug and alcohol related crime take place then area is filled with student that might be involved in these heinous activities during college life. Another point to understand through our analysis is that Wednesday has shown the highest rate of crime related to any form of intoxication are committed. This is a surprise as

normally any crime related to any form of intoxication are targeted towards the week during get together.

Massachusetts being a legal state would have a large amount drug involvement and with the help of the map we can see that towards the central part of Boston there is a large involvement of crimes that are related to drugs. The spread of these crimes that are committed are over the 4 cardinal direction is uniform.

# Analysis : Decision Tree:

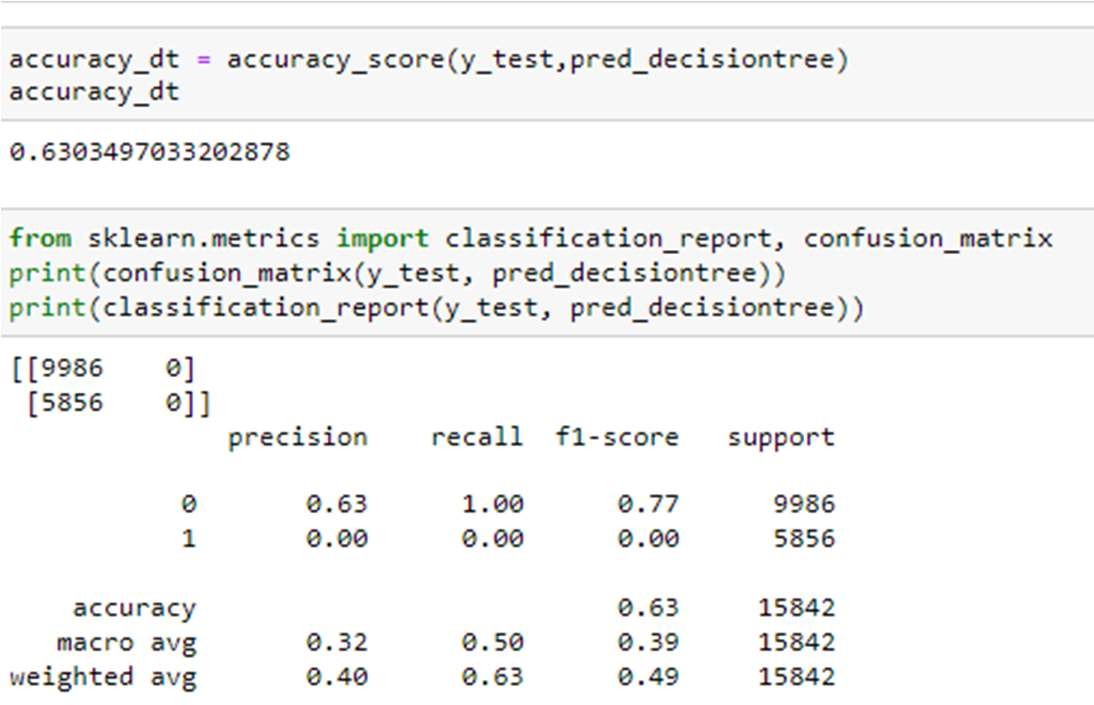


Figure 16 Decision Tree

The Decision Tree assists us in better understanding the model and comparing the variables. This decision tree has a depth of 4, implying that there will be four instances of cross validation. This model has an accuracy of 0.6308.

# Random Forest :

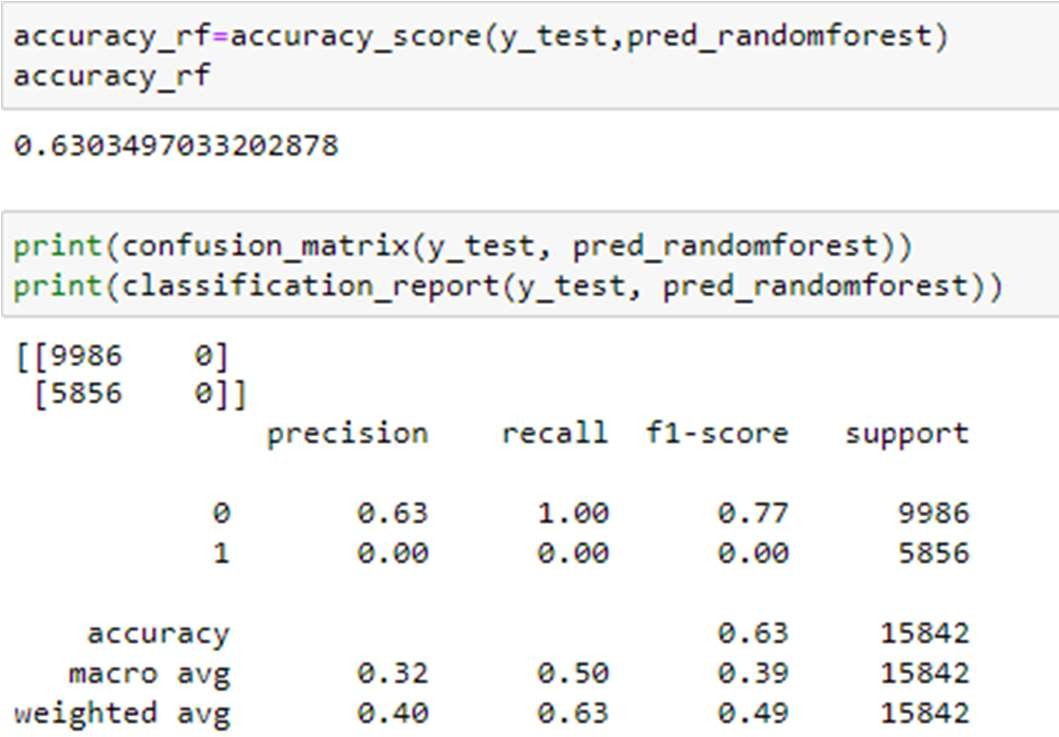


Figure 17 Random Forest

Random forests are more efficient than decision trees since all the branches are formed concurrently rather than individually. The model's depth is 4, which indicates that the model would generate four cross-validation points.

The matrix yields a model accuracy of 0.63.

# Gradient Boost:

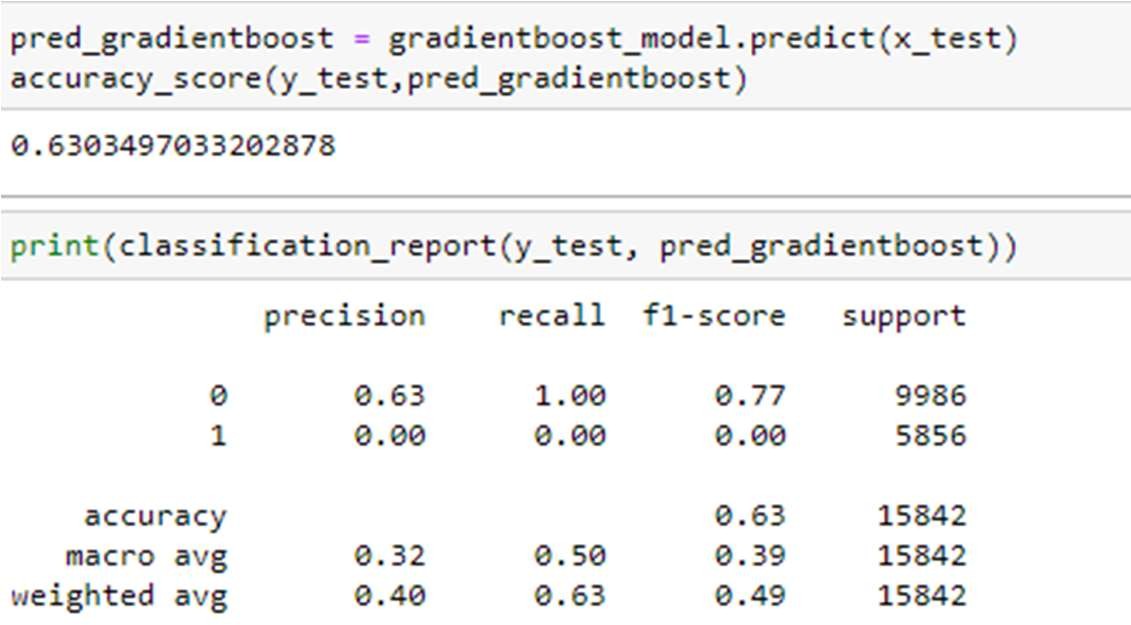


Figure 18 Gradient Boosting

Gradient boosting makes use of a technique called boosting, which systematically combines weak learners such that each new tree corrects the flaws of the prior one. In the matrix, it can be observed that the model has an accuracy of 0.63.

# Feature Importance:

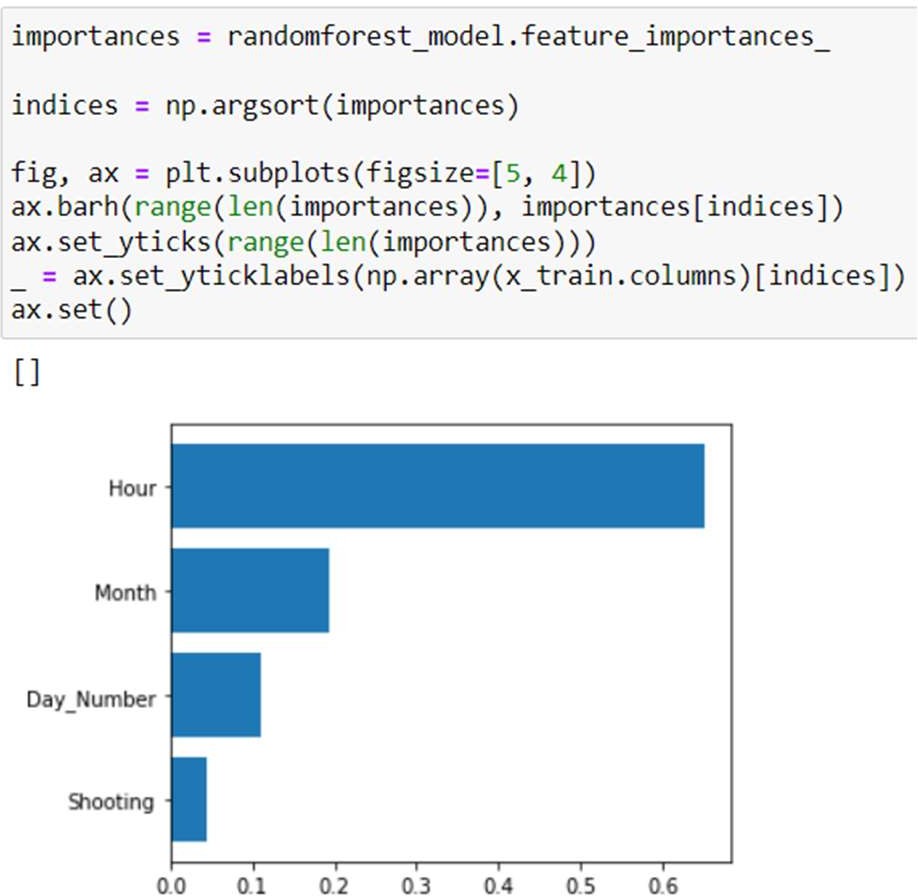
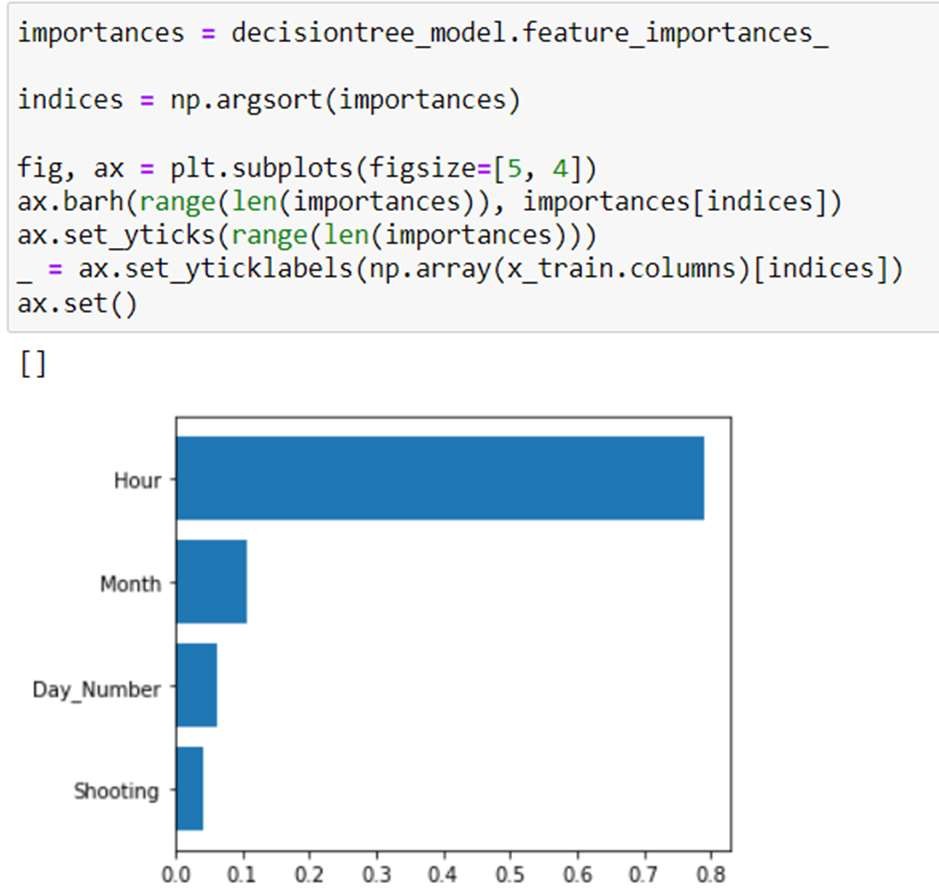


Figure 19: Decision Tree Figure 20: Random Forest



Figure 21: Gradient Boost

The three feature selection graphs created using the models are shown above. The identical variables are shown in all three graphs, and they all, coincidentally, demonstrate that the variable Hour is among the dataset's most crucial elements. It follows that a lot relies on the moment the offense takes place. The month is another characteristic that feature selection identifies as being essential to the dataset. In this

case, the month is quite important, and the weather and holidays may also have a part in the frequency of infractions.

# Conclusion:

To conclude, the issue with crime in a legal state is bound to be high. With our initial understand and analysis we are able get an idea of how this spread of crime could be targeted and reduced by informing the authorities on how these what time the crime is committed, on what day and which month they should consider tighten their security when considering the crime rate in Boston.

# References:

* The Boston Globe. (n.d.). Crime archives. Boston.com. Retrieved December 4, 2022, from https://[www.boston.com/tag/crime/](http://www.boston.com/tag/crime/)
* Queeriesimplified. (2021, September 29). Boston crime. Kaggle. Retrieved December 4, 2022, from

https://[www.kaggle.com/code/queeriesimplified/boston-crime](http://www.kaggle.com/code/queeriesimplified/boston-crime)

# Group Members:

**Karan Gigoo( 001561380) Adarsh Singh Jadon(001097411)**