

RATE OF CRIME IN THE CITY OF BOSTON.

★ **BUSINESS UNDERSTANDING**

The act of committing a crime has led to many being thrown behind bars. Crime is basically the act of committing an offence that is punishable by law. Crime is everywhere, and that does not make Boston an exception. Despite this being the 'to-go-to' city when it comes to the rich culture and historic sites, the likes of Faneuil Hall, it has not fallen short of being in the papers and news for the high rates of crime.

It is estimated, in every 100,000 people, 3,000 are attacked or they are victims of crime. There have been reports in the past where Bostonians have had to give up their walks at night, meet-ups with friends or any social activity due to the sprawling of dead bodies in allies, crimes that may have gone unsolved for over 10 years. An injustice for their families and the dead.

Others have had to secure their homes with double locks due to crimes like rape, forced entry or robbery. This is but to mention a few.

On the bright side, there are towns which are way worse than Boston, the likes of Detroit and Memphis.

People have lived in fear for their lives, but the government is taking initiative to make the city a better place.

★ **BUSINESS OBJECTIVES**

The primary business objective for this report is to understand the areas in Boston that are prone to crime from the year 2018.

There are more objectives to the study and they include;

- Looking for the district with the highest crime rate
- Understanding what was the most occurring type of crime in Boston and how they were committed.
- What time of the year had the highest rate of crime with the highest rates of crime
- What point of the year the majority of them are committed.
- We want to get a clear understanding of what sort of crimes are mostly committed.
- What crimes are most likely to happen
- If any crimes are linked to particular dates or months of a year
- How best to deploy the policing resources of Boston

★ **BUSINESS SUCCESS CRITERIA**

Providing an accurate report of areas with the highest crime rate and the time they were committed.

Also getting an overview of what crimes are most likely to occur.

★ **INVENTORY OF RESOURCES**

For this project we shall be working with our technical mentors who will come in and evaluate the project. In addition we shall work with a dataset we found on the Kaggle website based on the rate of crime in Boston.

We shall also be using Python language and its libraries like Numpy, Pandas for data manipulation ; Seaborn and Matplotlib for data visualization. Google Colab is our code editor. Moreover, we shall use GitHub to post work progress in terms of code. We shall use Sql to manipulate the data and lastly, Trello, which acts as our project management system.

The dataset to be used can be found in the following link:
<https://www.kaggle.com/AnalyzeBoston/crimes-in-boston>

The project completion date is 10/01/2021.

★ **RISK AND CONTINGENCIES**

After careful evaluation of the data, we have concluded that there are no risks and contingencies that come with this project.

★ **DATA MINING GOALS**

The data mining goals that come with the project include:

- Getting the areas that are most prone to criminal activity in Boston in the year 2018.
- How best to deploy the policing resources of Boston.

★ **DATA MINING SUCCESS CRITERIA**

The measure of success for this project will be determined by:

- Being able to identify the areas with the most criminal activity and curbing the problem by facilitating these areas with law enforcement.

★ **DATA COLLECTION, UNDERSTANDING AND DESCRIPTION**

For this project, we used a dataset from the Kaggle website. We decided to focus on a crime related project and thus we chose a dataset from the Boston Police Department. We got two datasets which are:

- Boston Crime dataset- This dataset had a total of 17 columns which showed the districts, streets, latitude, longitude, location, offense code etc all trying to show an overview of crime in the Boston area.
- Offense codes dataset- This is a description dataset showing the name of each offense code.

★ **DATA QUALITY REPORT**

For this project, the team used one dataset that is the Boston Crime dataset. On close examination on the data, we noticed that there had been some missing values in the shooting, district and UCR columns.

★ **DATA PREPARATION**

The data was collected and downloaded from the link provided above. We then selected the data for cleaning and analysis on Google Colab by importing the csv file.

★ **DATA CLEANING**

The cleaning was done in several stages. The first of which was dropping the columns. We did this by selecting the columns that we did not need. This included the shooting column and the latitude and longitude columns. Amongst the others that were repeated. We did this so that the other steps in the cleaning would be faster. As opposed to trying to clean columns, that we are going to drop. This was the theory behind dropping those columns.

Next, we tried to drop the null values. We decided to first check how many null values we had. Then we dropped the rows, where the null values were much less than the number of rows in the dataset. We did this, so that analysis would be easier. As well as that, we can analyse the data with rows that are not null. The only columns that had null values were District & UCR_Part.

Next we changed all the column names, into lowercase. This would allow us to be able to identify the columns easier. As opposed to finding the columns, while they were all in uppercase. It also makes it easier to read them. As reading them in uppercase is hard on the eyes.

Following this, we checked for duplicates in the table. We found that by dropping null values, we were only left with 89 duplicate rows. This allowed us to clean the dataset easier. Now after dropping the duplicate rows. We converted the datatypes of the code columns, to string. So that they are not affected during analysis. This is what we did using both python & sql.

Lastly, we converted the Ucr_part column into numbers. So that we can use it for analysis later on. As if we had left it as a string we would not have been able to analyse it. As it is easier to work with sequential codes like this. If they are numeric. This was our reasoning, in the cleaning process.

★ **ANALYSIS OF DATA**

From the analysis, we can see that the most prone areas to crime are B2, C11, D4, and B3. As they had the most rates of crime during the year 2018. They got over 30,000 incidents of crime in that year. This is why they are in the top list of crime incident districts. Also among these cities there were several top crimes which were committed in this district. This includes Larceny, Motor Vehicle Accidents, and Other crimes.

In terms of Larceny, it shows that those districts had the most amount of crime. This meant that people in this area must have had a lot of property. This meant they were either not guarded properly or they were just very expensive things in that district. Another possibility is that they did not have very expensive property, but they were very poor people there.

Next in terms of Accident response. It probably means that there were many accidents there. So this could have been caused by reckless driving. Another possibility is that people did not follow the traffic lights and rules, or there were not enough police barriers there. Either one could have caused the accident. It could also have been caused by drunk driving. However this would've been sorted by the police having many stops on those roads where accidents usually happen.

Lastly, there are verbal disputes. This could've been caused by several factors. However, it could be by the culture in the community. This could've caused disputes to happen more frequently than they otherwise could've. This is why the police should impose punishments for disputes. So that the community is aware that they should not be encouraged there. This would lower the amount of disputes in the community.

We need to note that all these three crimes happened in each of the four districts listed above. So the police should focus on these areas to reduce crimes in their district so that they would have lower police reports. This would also ensure that the community is a peaceful community. As they would know that the police are listening to disputes in the district.

★ **CONCLUSION**

In conclusion, I believe that the experiment was a success. We were able to find out the main areas where crime happened. We were able to do this, due to the specific objectives that were set. This allowed us to understand the data better. In addition it helped us to understand the main objective in reference to the data. This allowed us to analyse the data and draw conclusions based on it which was written down in the analysis section.

In terms of the data itself, we were able to see that there were five top crimes that happened in the top 4 districts. This allowed us to see where the police force could act in the districts and at what specific time.. It could also allow the police force to know how to handle these types of crimes. Especially in 7 other districts as they would know how to handle it in that area based on our recommendations. This is how we believe our report will help the police to handle their crimes in their areas.

★ **RECOMMENDATIONS**

From our analysis, there are several recommendations that we can deduce for the Boston Police Department. They are outlined as below:

- From the analysis, we saw Motor Vehicle Accident Response was the highest crime in Boston. The main causes for these accidents within the city include: Overspeeding, Inexperienced drivers, drunk driving and Unsafe lane changes. The department should look into these factors to reduce the number of Motor Vehicle Accidents in the area.
- From the analysis, we were able to conclude that the month with the highest level of crimes is August. From our research, in the month of August people are usually on summer holidays thus they will be out more often than not. Due to their celebratory mood, they tend to engage in social activities like drinking. It is confirmed that drinking comes with its own share of problems. Also due to the high temperatures people tend to leave their doors and windows open thus attracting unnecessary attention from the wrong crowd. The police department should be on the lookout during this specific month.
- The District of B2(Roxbury) in Boston had the highest number of crimes. Majority of these crimes occur in this town due to: The education policies are not favourable thus encouraging school dropouts, Also the number of police force deployed in that area is minimal, The lack of neighbourhood and police collaboration is also a major factor, in addition, re-emergence of gang groups due to rising poverty levels has seen many affected by this factor and lastly, ineffective community policing plays a big role. The police department should take these factors into consideration and fix the system.
- At time periods between 22:00 and 0300, security measures need to be improved to make Boston a safe town. During this time, you find most people out in the

streets are probably drunk and it is at that moment they are disoriented and most vulnerable. Gangs will want to take advantage of that moment and commit a crime, most likely rob and batter the drunks. From the above incidence, increasing police patrols would greatly reduce late night crimes.

Link to Trello Board: <https://trello.com/b/enDzddl5/crime-busters>

Link to GitHub: <https://github.com/DevTedd/MoringaPrep-GroupProject>