

# **CRIME DATA ANALYSIS IN BALTIMORE CITY**

## **ABSTRACT**

Crime is high in certain cities, and it varies based on the city's economic status and laws. To analyse various crimes and identify hotspots in the city, crime data analysis is prominent. Crime data analytics help personnel prevent further occurrences. In Baltimore, the city with the highest crime rate, crime data is analysed to provide in-depth insight into locations with the highest crime rates as well as specify the day and time where each type of crime profoundly transpired using proposed visualisation tools and an interactive visualisation model. This has been achieved by classifying crimes based on their spatial, or spatial and temporal, aspects.

## **INTRODUCTION**

Baltimore City, otherwise known as "Charm City," is one of the most dangerous cities to live in America. Baltimore has one of the highest murder rates per capita in the United States. Due to the urban sprawl and increased crime rate, the city has really become dangerous to live and substantially, the population growth has been reduced.

The goal of this project is to present an analysis on crimes and visualise the crimes with distinguished factors in Baltimore city.

## **DATA**

The data set on which the analysis has been performed is Crime Data of Baltimore City. This data set has been referenced from the official government website of Baltimore

"<https://data.baltimorecity.gov/datasets/baltimore::part-1-crime-data-/explore>".

We have loaded this data set as a raw csv file into our system. The csv file contained around 543617 rows and 23 columns which includes 'Location', 'CrimeDateTime', 'Gender', 'Inside\_Outside', 'Age', 'Longitude', 'Latitude' etc. The data set defines the location and characteristics of major crime against the persons such as homicide, shooting, robbery, aggravated assault etc., within the city of Baltimore.

## **DB SCHEMA**

The raw csv file has been loaded into our system. We have identified the unnecessary columns from the csv file and excluded them from our database schema. Normalized the data and formed 5 tables called are 'CrimeCode' which recognized different crime descriptions and their unique crime codes, 'Location' table which contains the information about the location where the crime has been occurred such as street name, district and neighbourhood, 'Race' table which contains the race and ethnicity information of the criminal, 'Criminal' table contains the criminal's age, gender, 'Crime' table is the most important table and this table use references of all the tables, it contains the information of the Crime Date and time, the weapon used for the crime, if the crime has performed inside or outside the house.

## **ANALYSIS**

Sqlite3 was used to extract the data from the normalised database by considering various scenarios like the number of crimes by location, types of crimes and their respective codes, types of crimes and the number of crimes per each type, top crime committing race, highest crime committing age groups, and more detailed analysis on the location which has highest crimes which is '1500 Russell Street' where the number of crimes in each type is recorded. This extracted data is further transformed into a graph to view the analysis and its flow.

## **RESULTS**

After getting the results from the queries using Sqlite3 on the normalised database, Python Packages matplotlib, Seaborn, Folium are used to convey the raw results of the queries on graphs for better understanding of the analysis. Different types of graphs have been plotted to analyse the data completely and extract specific results. Initially, the Seaborn's count plot represents the count of different crimes that have taken place in the city of which Larceny and Common Assault are found to be most common crime types. Secondly, an interactive heat map from folium package has been plotted to mark the hot zones of crime in the city. After that, a tabular representation has been drawn that consists of crimes along with their codes associated. A stack plot from matplotlib is pictured to represent various crimes in the most dangerous region of Baltimore, i.e., 1500 Russell Street where we found Larceny is the most committed crime. Later, a scatter plot constitutes the location of the crime, i.e., inside and outside of the house/building from which the information drawn is the reduction in crime during the pandemic in the year 2020. Further on, a heat map is depicted which represents the age of the criminals where it is found that most of the people involved in crime are of age or around 34 yo and all these results are achieved by implementing the Python libraries.

## **CONCLUSION**

The purpose of this project was to utilise effective python strategies to deal with large data sets and to formulate results based on the prepared database schema. Based on the analysis conveyed, it can be concluded that there are various crimes such as 'Rape', 'Robbery', 'Aggravated Assault' etc., performed in Baltimore city, '1500 Russell Street' was identified as the most dangerous street in Baltimore city, tabulated various crime codes for different crimes, age category 34 is the most common age to perform crime, race 'American\_Indian\_Alaska\_Native' performed the most number of crimes and other results.

## **REFERENCES**

- [https://matplotlib.org/stable/plot\\_types/basic/index.html](https://matplotlib.org/stable/plot_types/basic/index.html)
- <https://seaborn.pydata.org/generated/seaborn.countplot.html>
- [https://pandas.pydata.org/docs/user\\_guide/index.html](https://pandas.pydata.org/docs/user_guide/index.html)
- <https://www.sqlite.org/lang.html>
- <https://docs.python.org/3/library/re.html>
- <https://python-visualization.github.io/folium/>

## **A PROJECT BY :**

**Geethika Sukamanchi (UB ID : 50477946)**

**Venkata Geethik Kumar Rachaputi (UB ID :50476154)**

**Sai Sahiti Chittem (UB ID: 50478898)**

**Sai Kumar Nalla (UB ID: 50481156)**

**Lalith Sai Padakula (UB ID : 50461348)**