NEWARK CRIME DATA

* **Abstract**

We have chosen Newark crime dataset for our project. We will be working on visualizing in which type area is the highest crime and in which time the crime is happening .We the given dataset we will be using shape file to plot the neighbourhoods to understand the crime.

* **Business Objective:** To help the Police Department be aware of the Crime Statistics in every area of Newark.
* **Problem Statement:** To understand the pattern of crime in Newark city.
* **Dataset Description**

Datasets that are used in the approach are:

* Newark Crime Dataset: The dataset has 160 rows and 12 columns
* Abandoned Properties: The dataset has 820 rows and 9 columns
* 4311 Data of 2018: The dataset has 5780 rows and 11 columns
* **Tools and Techniques**
* **Tools:**

We plotted and visualized the data we had gathered from the datasets of Newark crime data, Abandoned properties and 4311.

* **Packages & Files:**

Leaflet, Shapefile, JSON file

* **Tools:**

R language and R Studio

* **Results and Discussion :**
* In below graph analysis we are considering the ‘Newark Crime Dataset’.
* The dataset is from the month of May, and we can see that most crime prone neighborhood is ‘Central Business District’ which yields the highest average of 44 crimes a month.
* Followed by ‘Forest Hill’ with 16 and ‘Dayton’ with 15 crimes.

* In the below graph, a day is divided into 4 parts namely: Afternoon, Evening, Morning, Night.
* And we can observe that in the ‘Afternoon’ time the crime rate is very high.
* So, we can conclude that in Newark city the most crime prone neighborhood is ‘Central Business District’ and the most common time of crime is Afternoon which ranges from 12PM to 4PM.

* Some of the crimes we consider in this Dataset are:
* Aggravated Assault
* Auto Theft
* Burglary
* Murder
* Robbery
* Theft
* Theft F/A

A close up of a map

Description automatically generated

* **Consider the Abandoned Properties dataset :**

**A close up of a map

Description automatically generated**

* In this graph, we can see that neighbourhoods of ‘Dayton’, ‘Central Business District’ have a lot of abandoned properties.
* If we analyse the ‘Newark Crime Data’ and ‘Abandoned Properties Data’ together we can see that there is some relationship between them.
* We can infer that in neighbourhoods where there are a lot of abandoned properties, a higher crime rate is also observed.
* For example, consider ‘Dayton’ and ‘Central Business District’ neighbourhoods. We can observe that there is a high rate of abandoned properties along with a high Crime rate (According to Crime data).
* **Consider the 4311 dataset:**
* To generate this graph, ‘4311 Data of 2018’ is considered.
* In this graph, we can see that neighbourhoods of ‘Dayton’ and ‘Central Business District’ have a lot of distress calls.
* If we analyse the ‘Newark Crime Data’ and ‘4311 Data’ together we can see that there is some relationship between them.
* We can infer that in neighbourhoods where there are a lot of 4311 calls, a higher crime rate is also observed.
* For example, consider ‘Dayton’ and ‘Central Business District’ neighbourhoods. We can observe that there is a high rate of 4311 calls along with a high Crime rate (According to Crime data).

**A close up of a map

Description automatically generated**

* **Problems Faced:**
* The Datasets that were obtained were vague, incomplete and had many missing values.
* Thus, making the data cleaning procedure very strenuous.
* The addresses very incomplete and did not have any Zip-code values, which made it hard to generate the Geo-codes that are used to plot the addresses in the map.
* To summarise, the Data cleaning, Zip-code and Geo-code Generation.
* **Conclusion :**

**From the results of ‘Newark Crime Data’ , ‘Abandoned Properties Data’ and ‘4311 Data of 2018’ we can infer that:**

* There is a relation between the locations of abandoned properties, 4311 data and the number of crimes that are being committed in or around those neighbourhoods.
* Consider ‘Central Business District’ neighbourhood. It records the highest number of crimes in the city along with a high number of abandoned buildings and alarmingly high number of 4311 distress calls.
* **Future work :**

The above approach can be used to generate:

* + Time-Series
  + Weight Sampling
* **Business objective is accomplished as we can predict the most crime prone neighbourhoods, some of the factors influencing the crimes and the timing of most of the crimes that are being committed.**

**By: Sanket Gohel, Sruthi Boddapati, Tjiexn**