

San Francisco Police Department Incident Reports – Use Case

AUTHOR : JITHIN PRAKASH K



Introduction

Police crime incident reports are a very important data to understand the history of the crimes based on different crime categories, based on the time that it had happened and the location. Proper analysis of categories of crime occurring at difference places at different time interval gives an idea about the patters of such occurrences if any and can make inferences to prevent such criminal offences. Police department gather incident reports filed by officers and by individuals through self-service online reporting for non-emergency cases. Incident reports filed online will also be reviewed by a supervising officer. Once approved and electronically signed by a Sergeant or Lieutenant, no further information can be added to the initial report. A proper analysis is an essential process to be done to infer the right assumptions and results.

Objectives

With our research we hope to find answers to the following questions.

- Does a criminal data base that contains geographical location & basic details of the criminal activity have enough indicators to predict a type of crime?
- Given just a geographic location and time, how accurately can we classify the crime?
- Explore different techniques to improve the results.
- Analysis can help the Police and Security services to concentrate on major crime incidents based on time and location

Data Description

Police department gather incident reports filed by officers and by individuals through self-service online reporting for non-emergency cases.

The data that will be used to analyse the police records to find the vulnerable area which can be used to predict the best paces and neighbourhood for reducing the criminal activities within San Francisco. The data I have found is collected from 'The office of the chief Data Officer – City and County of San Francisco' (<https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783>). The Polices Department has developed a report of incidents based on different categories, time and location

Data Description

- Incident Date
 - Data and Time of the Incident
- Incident Category
 - Category of the Incident
- Incident Subcategory
 - Subcategory of the incident
- Incident Description
 - Description of the incident
- Resolution
 - Resolution taken against the crime
- Police District
 - Police District of the crime (location)
- Analysis Neighbourhood
 - Neighborhood where the crime happened
- Latitude and Longitude
 - Location Data

Methodology

Collect Police Incidents Data

Explore and Understand Data

Data Preparation and preprocessing

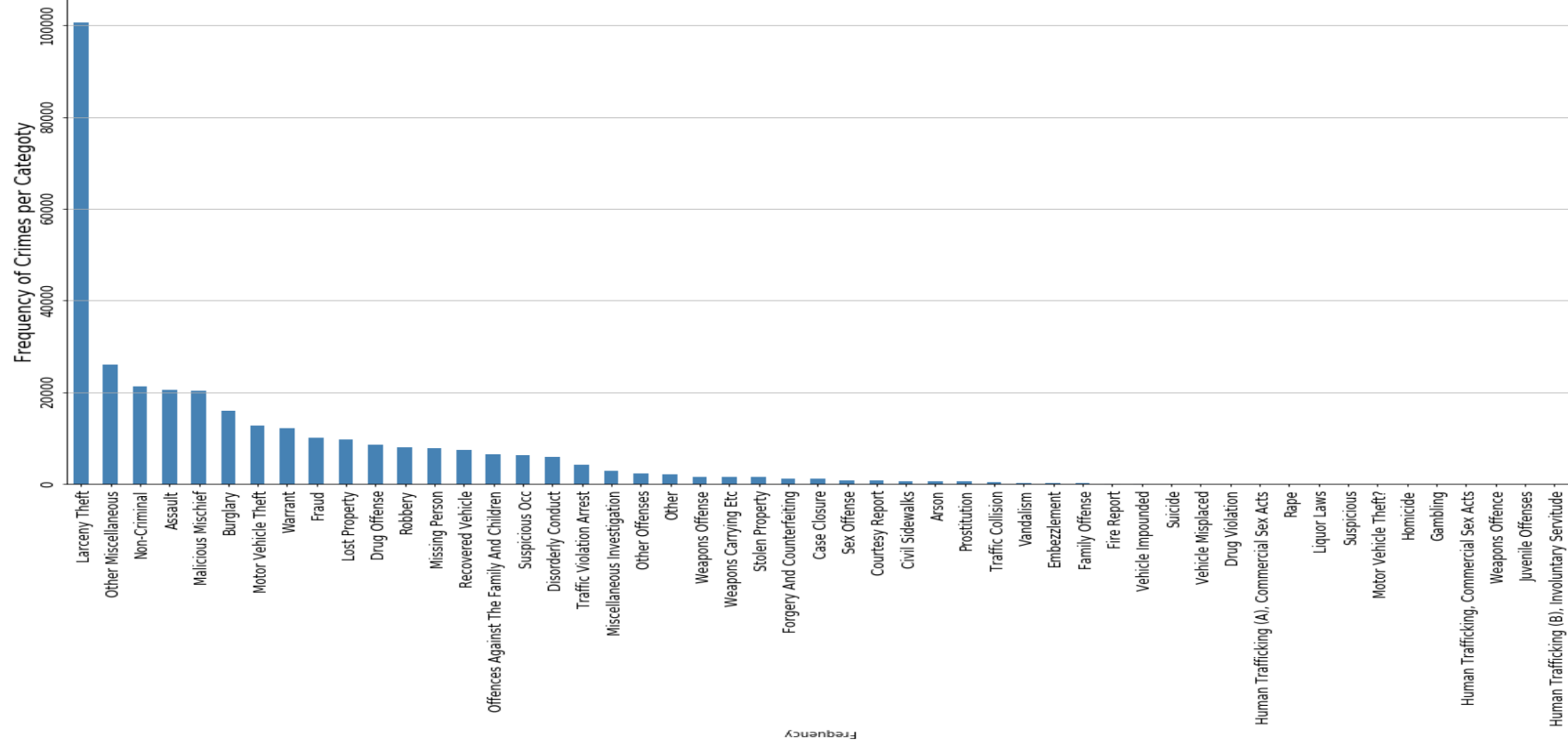
Modelling

Evaluation and testing

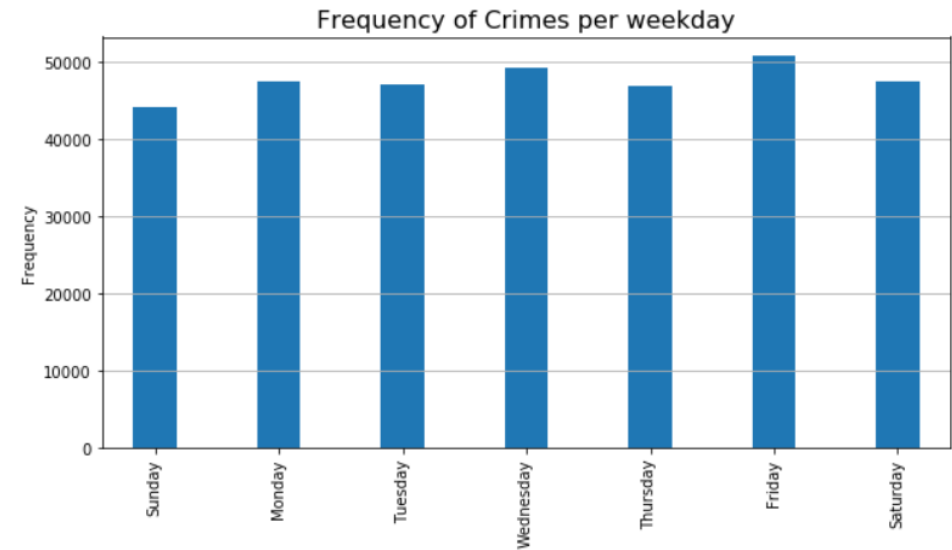
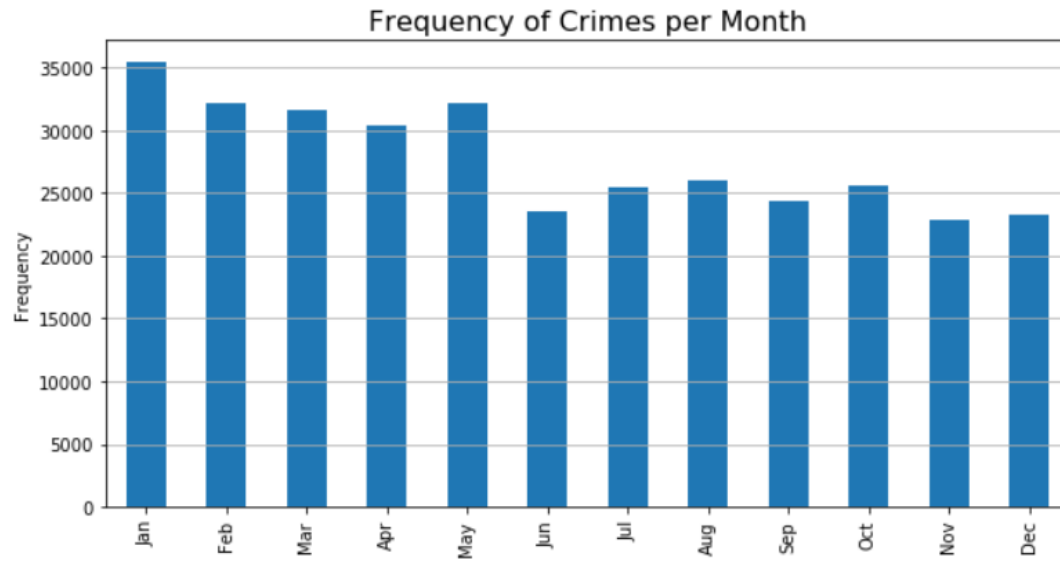
Modelling – Parameters and Results

Classifier	Parameters	Train set Accuracy	Test set Accuracy	F1 Accuracy	Jaccard Index Score	Log Loss
K-Nearest Neighbour	K=9	0.3547	0.2472	0.1714	0.2472	-
Decision Tree	Depth = 80	-	0.2447	-	0.2447	-
Logistic Regression	Log loss = 2.82	0.3033	0.3013	0.1395	0.3013	2.8177

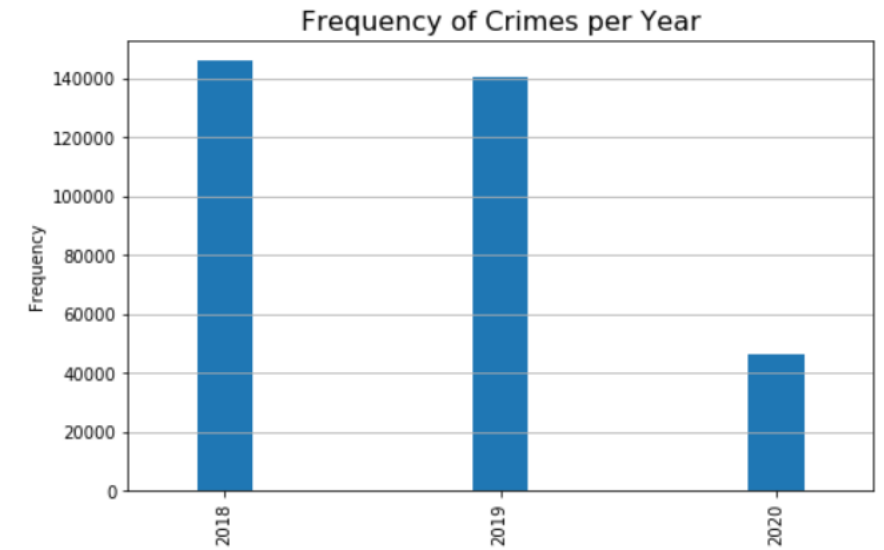
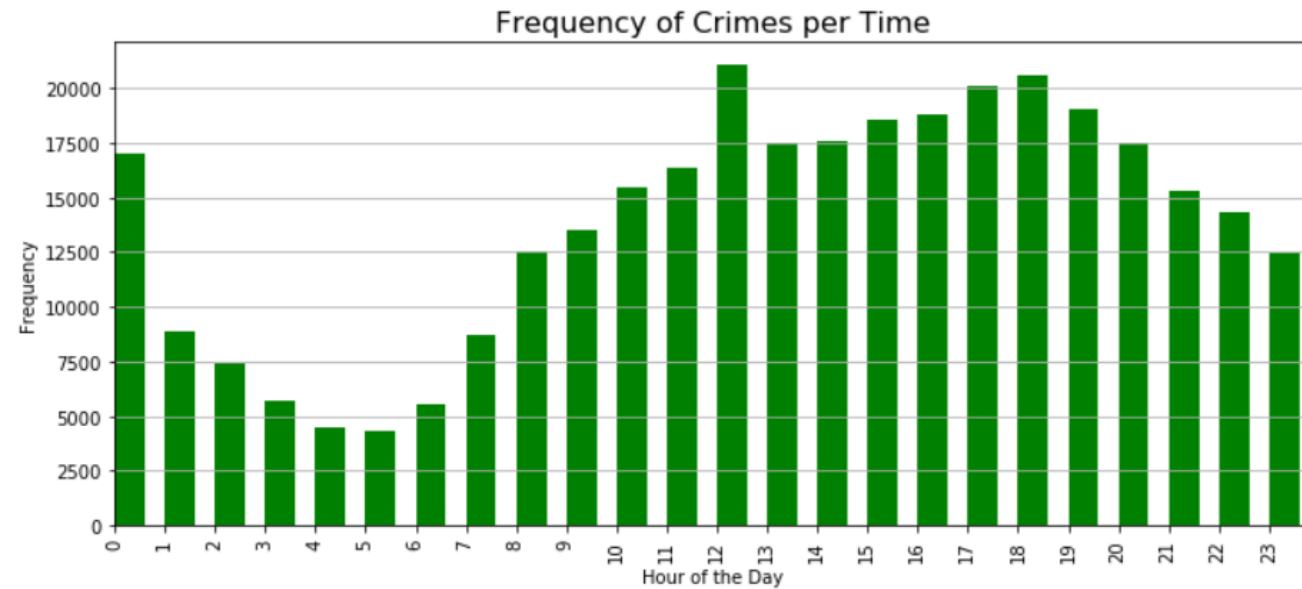
Results : Crimes per Category



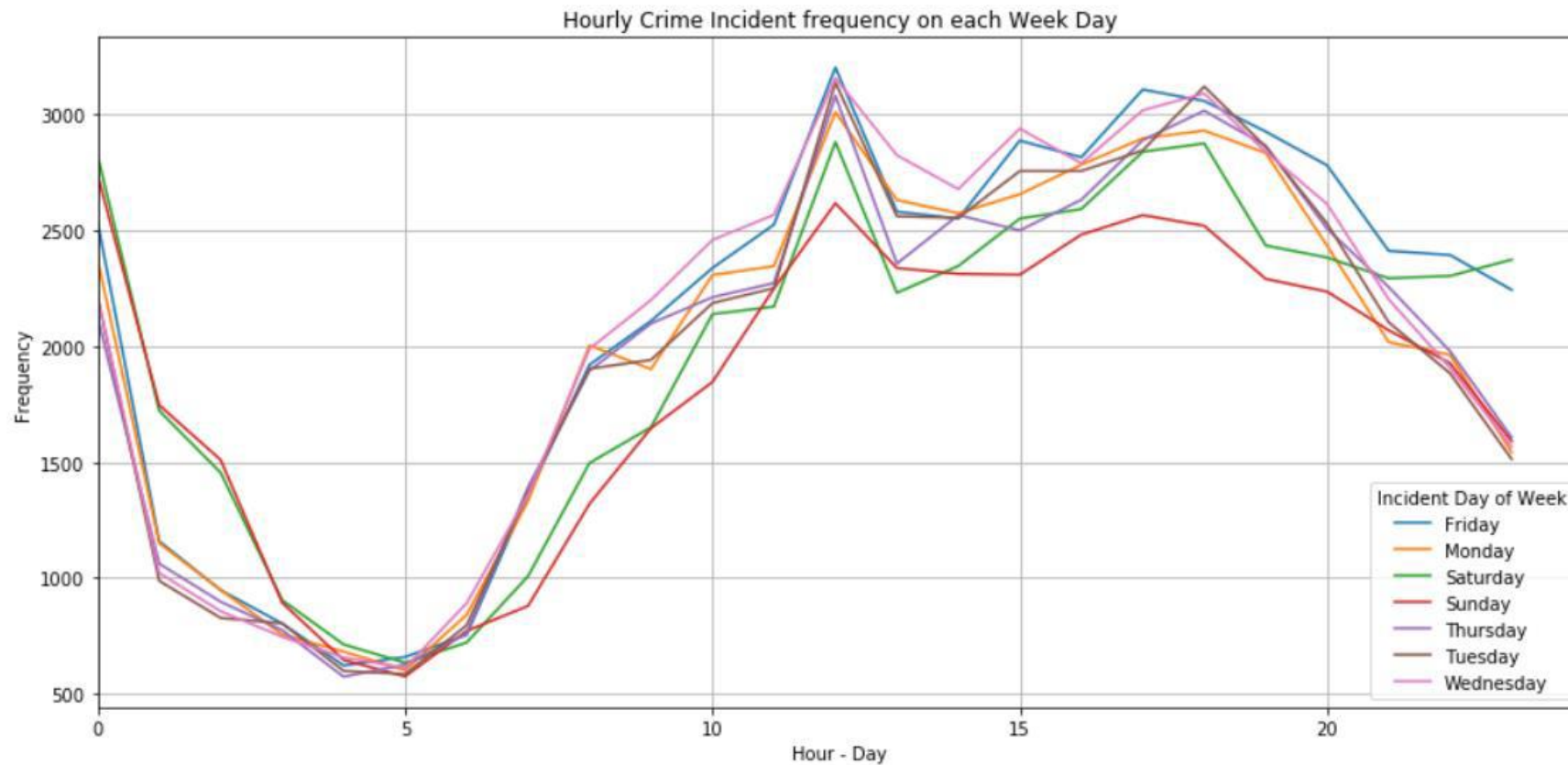
Results : Crime frequency



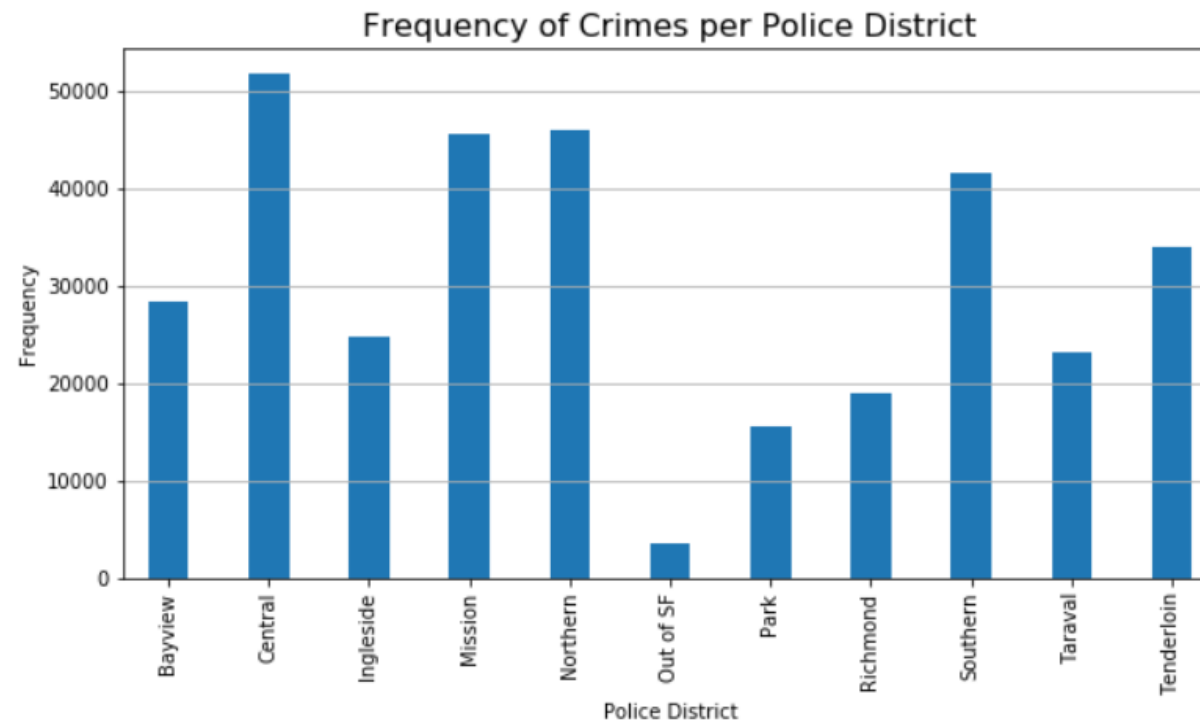
Results : Crime frequency



Results : Crime frequency



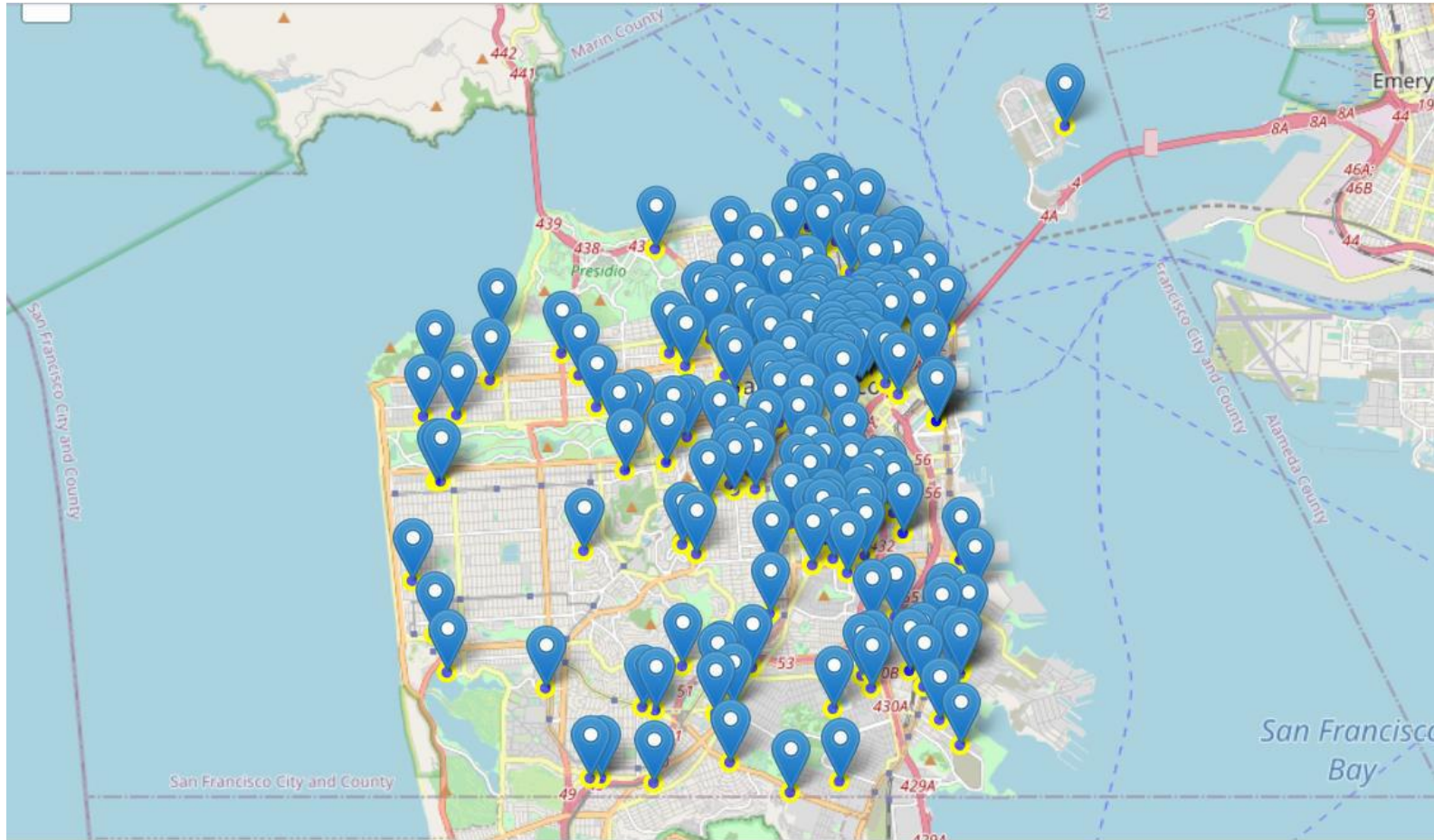
Results : Crime frequency per Police District



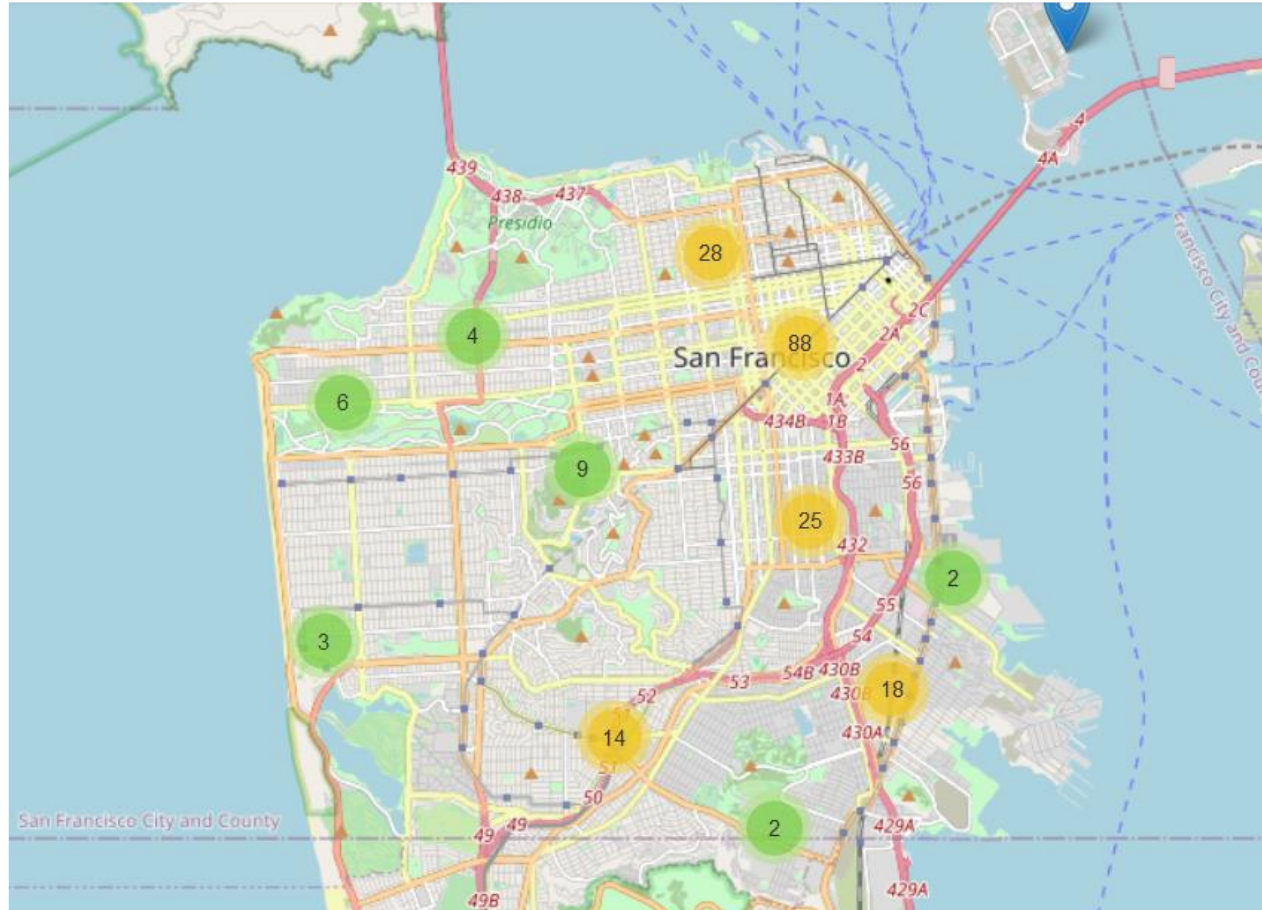
Results : Word-Cloud



Results : Crime Data Plotted on the map



Results : Crime Data Plotted on the map - Cluster



Foursquare – Crime incident and locality



Conclusion

To eliminate the crime occurrence and to bring public safety and peace in place, it is utmost important to find the category of the crime incident and its nature. Police and security service department need to continuously monitor and tighten the measures to eliminate different categories of criminal activities. In public security services, technology can be used to predict a likely critical violation through the use of data analytics instead of inspecting every joint blindly given the lack of enough manpower for this. The data used to predict critical violation include location and time data. Afterward, places data e.g. Foursquare is used to locate the most affected areas and can be worked towards improving the security and surveillance measures.

Thanks
