**Coursera IBM Data Science Professional Certificate Capstone Project - Report**

Alessandro Bax

**Introduction:**

This project aims to better understand the crime in the city of San Francisco (CA) in order to help the decision makers (Politicians, Authorities and Citizens) target the right solutions for each area of the city, based on what kind of problem they present.

For example, if a region in the city have more incidence of a car stealing and murderers in other area, they demand different kind of action, like placing traffic cameras and increasing the number of police.

**Data:**

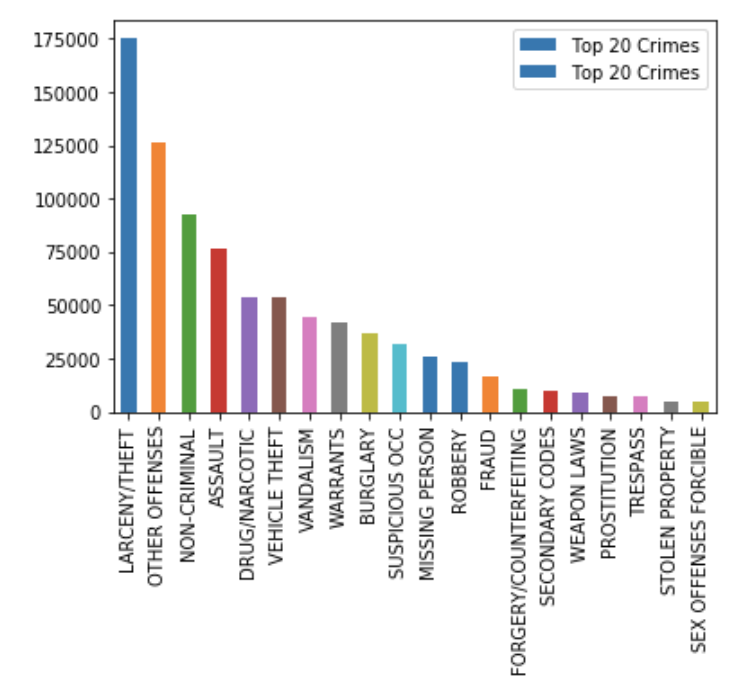
The data used in this project was from a Kaggle Competition, and contains incidents derived from SFPD Crime Incident Reporting system, as listed below:

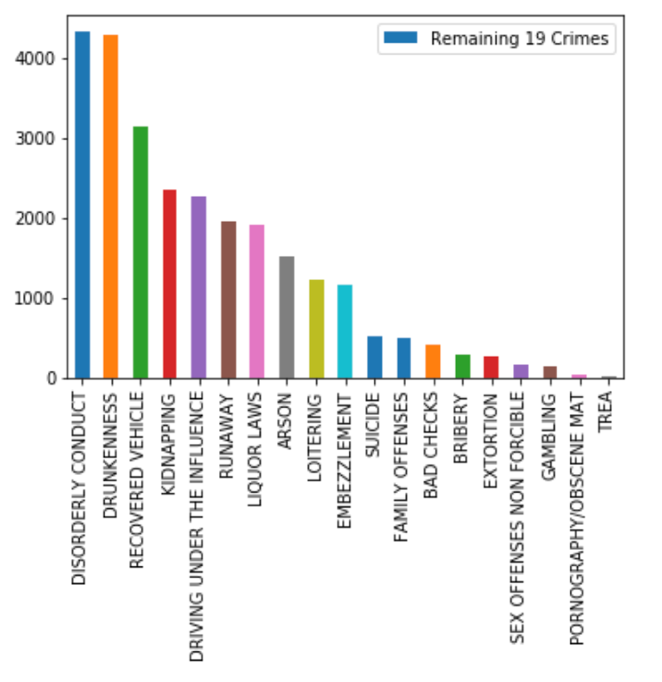
* **Dates:** timestamp of the crime incident
* **Category:** category of the crime incident
* **Descript:** detailed description of the crime incident
* **DayOfWeek:** the day of the week
* **PdDistrict:** name of the Police Department District
* **Resolution:** how the crime incident was resolved
* **Address:** the approximate street address of the crime incident
* **X:** Longitude
* **Y:** Latitude

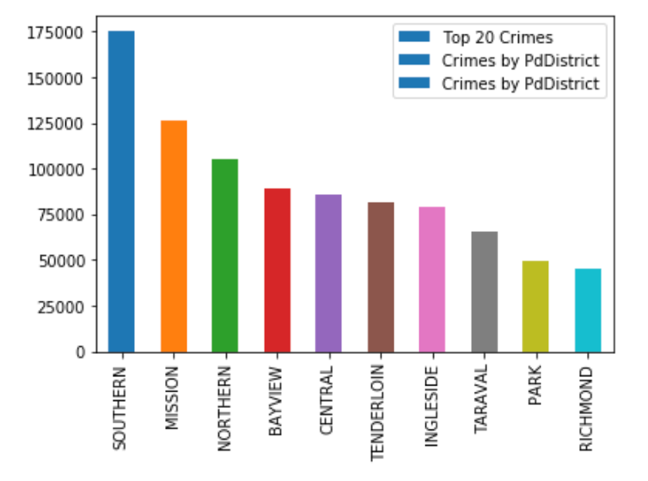
This dataset is composed by 9 columns and 878049 observations with data from crimes between 2003 and 2015.

To solve the project objectives with this dataset, the data is grouped by the category of the crime incident and by region in order to better understand the problem in the city.

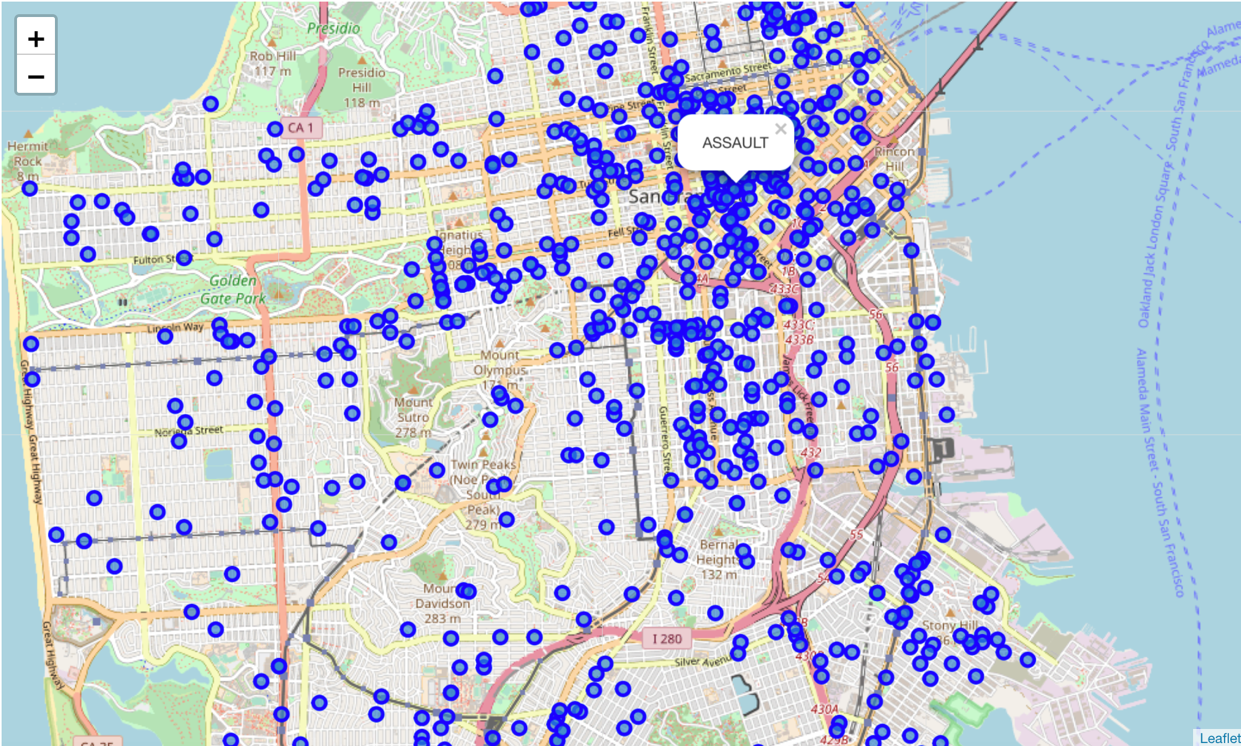
**Methodology:**

The Python Jupyter was used to analyze this dataset, with the libraries *numpy*, *pandas* and *matplotlib*. The first step was to import the dataset into a pandas dataframe. Then generates some basic visualizations to better understand the data.





To plot the data points in the map, and get a better understanding of the crime distribution in the city of San Francisco, it was used the library *folium* and *geopy.geocoders*.



**Results:**

With this analysis it was possible to identify first the most frequent category of crime incidents in San Francisco, as shown below:

1. Larceny/Theft
2. Assault
3. Drug/Narcotic
4. Warrants
5. Vandalism
6. Vehicle Theft

Also, the analysis included data from the Foursquare API, which took the latitude and longitude from the crime record, to identify the closest Venue Category to each crime, and then grouped the data, to identify the main places nearby the crime incident location, ranked below:

1. Bar
2. Vietnamese Restaurant
3. Bakery
4. Art Gallery
5. American Restaurant
6. Wine Bar
7. Yoga Studio

**Discussion:**

The relation between the main category of the crime incidents and the places nearby are consistent and indicates that the crowded places, like bars and restaurants tends to have crime incidents like theft and assault. Also, the map indicates that the northeast region concentrates more crime incidents than the other regions in the city.

**Conclusion:**

The Foursquare API was important to aggregate more information for the analysis, bringing new insights along with the crime dataset. It was possible to correlate the crime category with the main venues category, and understand the relation between the kind of places and the crimes in the city. With this information, it is possible to start discussing a few solutions to increase the security in the city of San Francisco.