FINAL CAPSTONE PROJECT

MAPPING CRIME IN SAN FRANCISCO

By James Njumwa

EXECUTIVE SUMMARY

- Project entails mapping of crime incidences in San Francisco in 2016
- Increased crime rates prompted the need to evaluate most susceptible areas in San Francisco.
- Data upload in csv format and data exploratory analysis done to provide descriptive analysis.
- In-depth analysis conducted using clustering and mapping using cluster markers and drop pins.
- Tyler street highlighted as the most susceptible area including its neighbourhood.
- Further robust analysis recommended for more inferential statistics.

INTRODUCTION

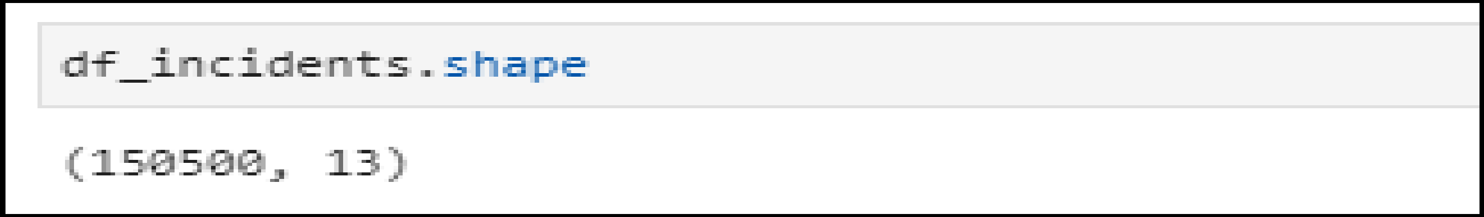
- The final course of the Data Science Professional Certificate consists of a capstone project in which all of the skills and related details gained over the nine rigorous coursework must be implemented on a final project.
- In this project we explore the San Francisco police department incidents using cluster visualization to map out number of crimes in San Francisco.

PROBLEM STATEMENT

- In any developed or developing countries, there exists a marginal gap between the rich and the poor. Based on increased rates of unemployment, the ever growing population tends to lose its moral value and engage in crimes just to overcome the social burdens they are facing individually.
- In 2016, the police department of San Francisco Made numerous arrests include major and minor. This issue rendered the necessity to understand the patterns of criminal behaviors in the state.

DATA PREPARATION

- Data was downloaded and imported into pandas as a csv file. Imported data consisted of 13 variables including longitude and latitude values of crime locations.

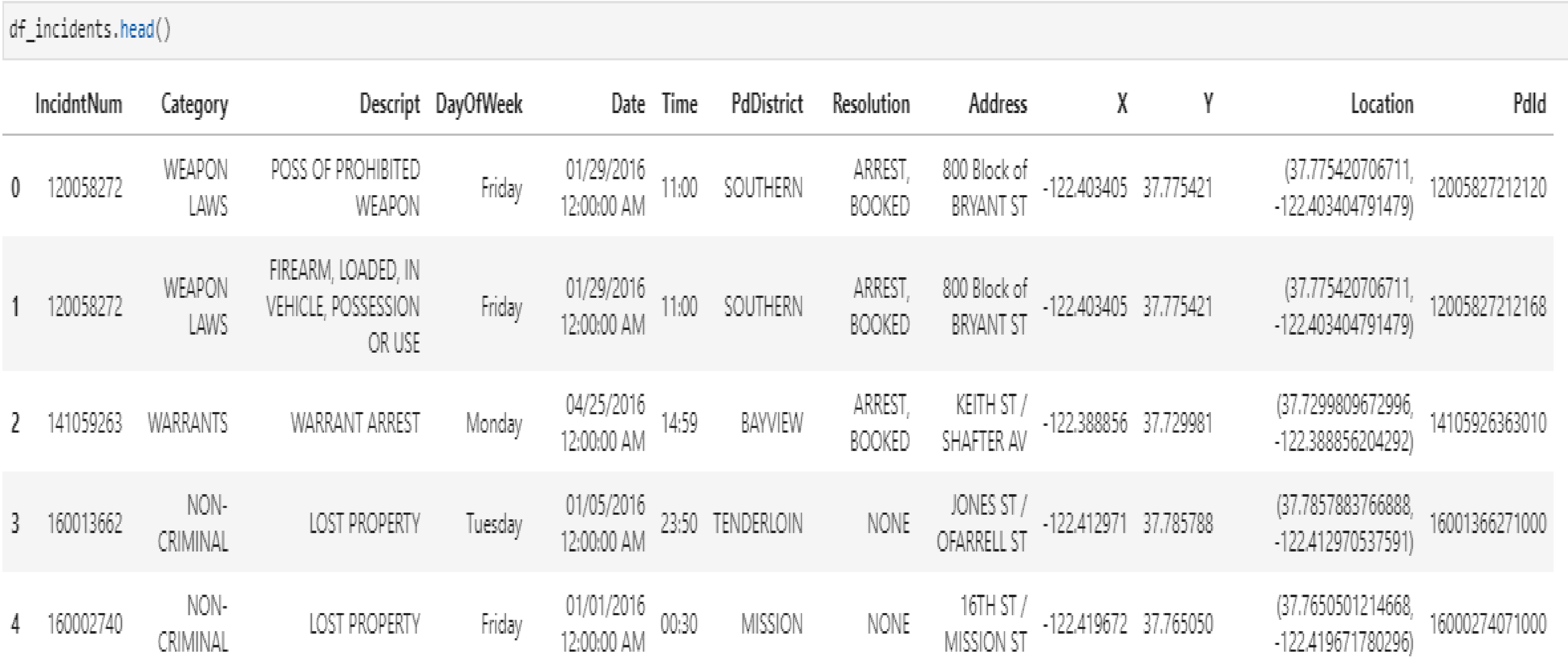


```
df_incidents.shape
```

```
(150500, 13)
```

- The descriptive analysis indicated that there were a total of 150,500 crimes that were reported in San Francisco in 2016.
- The study randomly selected a limit of 100 crimes to work as a representative of the population.

VARIABLES USED

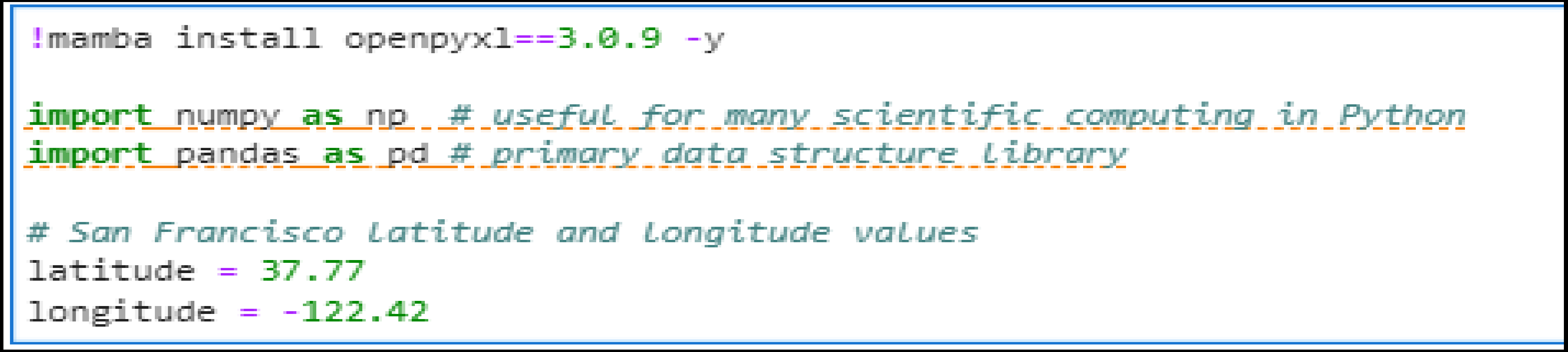


```
df_incidents.head()
```

	IncidntNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	Location	PdId
0	120058272	WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.40340479179)	12005827212120
1	120058272	WEAPON LAWS	FIREARM, LOADED, IN VEHICLE, POSSESSION OR USE	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Block of BRYANT ST	-122.403405	37.775421	(37.775420706711, -122.40340479179)	12005827212168
2	141059263	WARRANTS	WARRANT ARREST	Monday	04/25/2016 12:00:00 AM	14:59	BAYVIEW	ARREST, BOOKED	KEITH ST / SHAFTER AV	-122.388856	37.729981	(37.7299809672996, -122.388856204292)	14105926363010
3	160013662	NON-CRIMINAL	LOST PROPERTY	Tuesday	01/05/2016 12:00:00 AM	23:50	TENDERLOIN	NONE	JONES ST / OFARRELL ST	-122.412971	37.785788	(37.7857883766888, -122.412970537591)	16001366271000
4	160002740	NON-CRIMINAL	LOST PROPERTY	Friday	01/01/2016 12:00:00 AM	00:30	MISSION	NONE	16TH ST / MISSION ST	-122.419672	37.765050	(37.7650501214668, -122.419671780296)	16000274071000

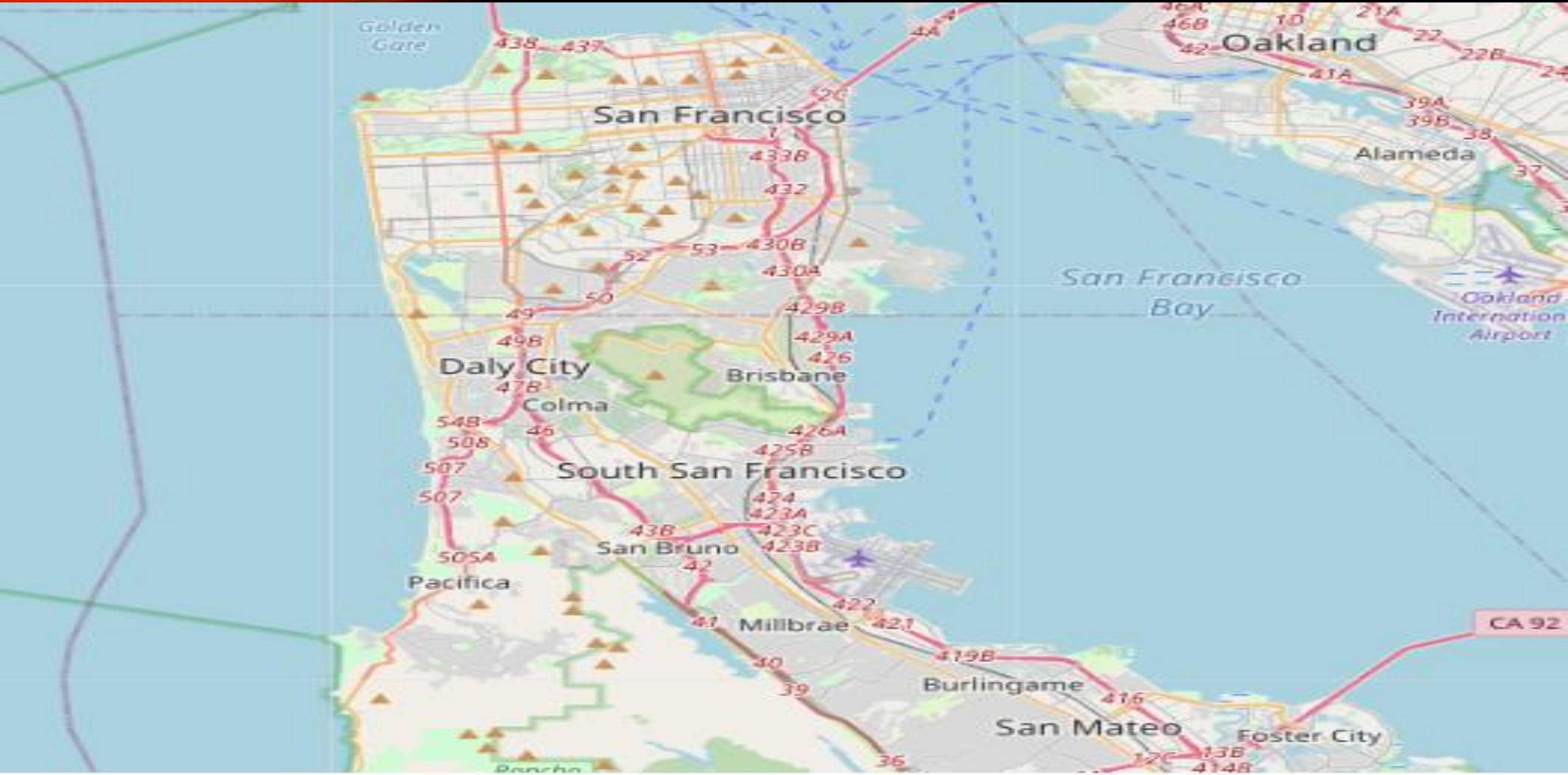
METHODOLOGY

- Data Clustering and mapping was employed in the study after cleaning and sorting the imported data.
- The necessary features including folium maps, mamba, numpy, and pandas were installed with their plugins.

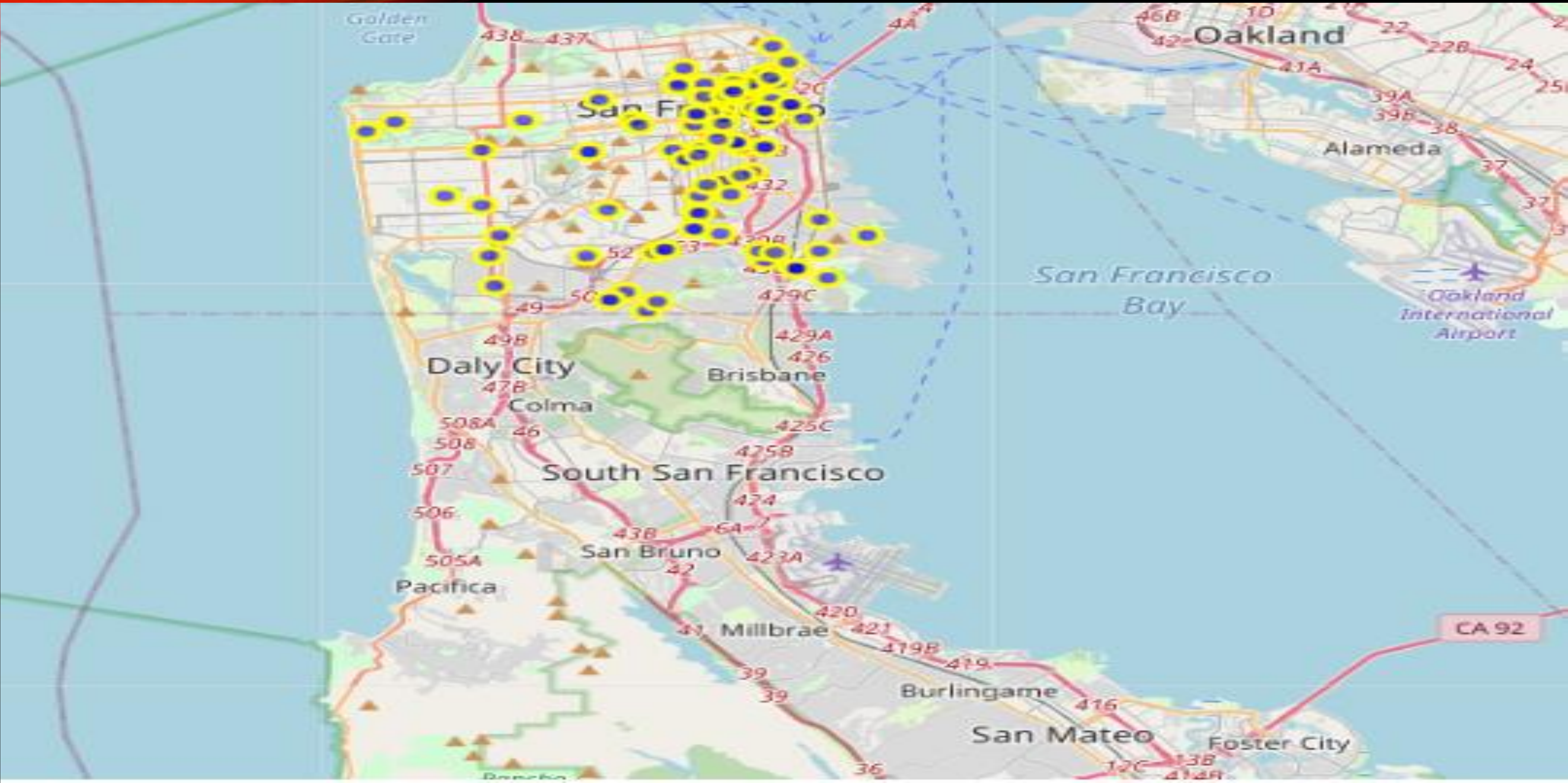


```
!mamba install openpyxl==3.0.9 -y

import numpy as np  # useful for many scientific computing in Python
import pandas as pd # primary data structure library

# San Francisco latitude and longitude values
latitude = 37.77
longitude = -122.42
```

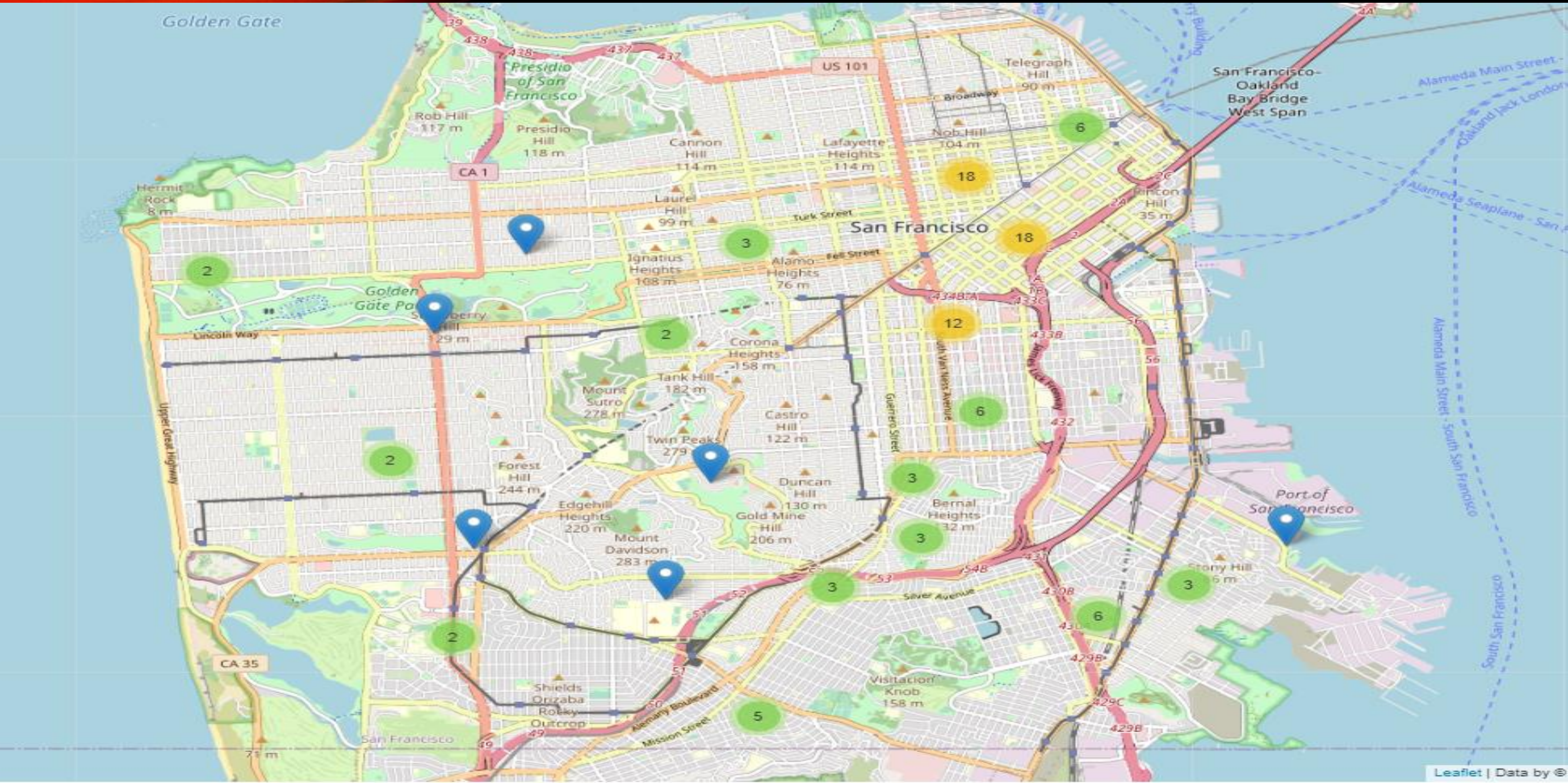
STUDY AREA


Golden Gate
San Francisco
Oakland
Alameda
San Francisco Bay
Oakland International Airport
Daly City
Brisbane
Colma
South San Francisco
San Bruno
Pacifica
Millbrae
Burlingame
San Mateo
Foster City
CA 92
438
437
433B
432
430B
430A
429B
429A
426
426A
425B
424
423A
423C
423B
422
421
419B
416
53
52
50
49
49B
47B
54B
50B
507
505A
46
43B
42
41
40
39
36
4A
46A
46B
10
21A
22
22B
42
41A
39A
39B
38
37

SUPERIMPOSED CRIME LOCATIONS ON THE MAP USING FOLIUM TO CREATE A FEATURE GROUP



GROUPING THE MARKERS INTO CLUSTERS TO REMEDY THE CONGESTION



RESULTS AND DISCUSSION

- The dataset imported was very large. However, it may have posed a problem during the mapping given the 100 sampled were also still congested.
- Data clustering was efficient in minimizing the problem of congestion on the mapped crimes.
- Global clusters applied were able to indicate the number of crimes represented by each cluster on the map upon zooming
- Tylor street and its neighborhood are prone to crimes compared with other areas.
- Burglary and Vandalism were the highest recorded incidences along Tyler street and its neighborhood.

CONCLUSION

- More Police officers should be deployed around Tyler street and its neighborhood to minimize crime incidences.
- The data used was just a sample therefore may not reflect the current situation in the study area.
- The study recommends employment of other robust techniques that would highlight causes of increased crime rates in some areas compared to others in the study area.
- Some codes applied in the python may have changed following recent updates in some apps or plugins on the environment.

ACKNOWLEDGMENT

I would like to thank Coursera and IBM teams for their generous scholarship in studying this professional course and realizing a step into a field I am passionate about. This course was great with practical applications and hands on experience that ensure you understand each lie of code you write and what its execution would result. I encourage anyone interested in data science without any knowledge of programming to take this course as your starter pack.

THANK YOU