"San Francisco safe travel and restaurant visits"

YK Woo

1. Introduction

        A. A description of the problem and a discussion of the background.

San Francisco is a fairly large city in the United States, the hub of business, finance, and transportation. This city must be an attractive city for many tourists. Unfortunately, a variety of crimes occurred in this city during 2016. I would like to analyze San Francisco's crime patterns by time and region to introduce safer and better neighborhoods for tourists and to recommend good restaurants to visit.

        B. A description of the data and how it will be used to solve the problem.

The data can be obtained from the following link. This data includes information such as the time of crime, the type of crime, and the area where the crime occurred. By analyzing and visualizing this data, we will use it as proactive information to help us avoid dangerous risks. In addition, we will use the Foursquare API to add the necessary information based on data such as the location and type of restaurants in the San Francisco area.

The Information that can affect data based on the problems we define :

- Time of Crime
- Day of Crime
- Location of Crime
- Location of Restaurants
- Map of San Francisco

2. Import module and data for our project

```
import requests # library to handle requests
import pandas as pd # library for data analysis
import numpy as np # library to handle data in a vectorized manner
import random # library for random number generation

import matplotlib.pyplot as plt # library for plot visualization
%matplotlib inline # Masic code for plot visualization

!conda install -c conda-forge geopy --yes # Install for geopy
from geopy.geocoders import Nominatim # module to convert an address into latitude and
longitude values

# libraries for displaying images
from IPython.display import Image
from IPython.core.display import HTML

# transforming json file into a pandas dataframe library
from pandas.io.json import json_normalize

!conda install -c conda-forge folium=0.5.0 --yes # install folium
import folium # map plotting library
```

2.1 Loading the Crime data of San Francisco

Dataset downloaded and read into a pandas dataframe by URL :
https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/DV0101EN/labs/Data_Files/Police_Department_Incidents_-_Previous_Year__2016_.csv

2.2 Pandas data frame of San Francisco Criminal Issue

	IncidntNum	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y
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
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
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
5	160002869	ASSAULT	BATTERY	Friday	01/01/2016 12:00:00 AM	21:35	NORTHERN	NONE	1700 Block of BUSH ST	-122.426077	37.788019
6	160003130	OTHER OFFENSES	PAROLE VIOLATION	Saturday	01/02/2016 12:00:00 AM	00:04	SOUTHERN	ARREST, BOOKED	MARY ST / HOWARD ST	-122.405721	37.780879

<Each column's explain>

1.IncidntNum: Incident Number

2.Category: Category of crime or incident

3.Descript: Description of the crime or incident

4.DayOfWeek: The day of week on which the incident occurred

5.Date: The Date on which the incident occurred

6.Time: The time of day on which the incident occurred

7.PdDistrict: The police department district

8.Resolution: The resolution of the crime in terms whether the perpetrator was arrested or not

9.Address: The closest address to where the incident took place

10.X: The longitude value of the crime location

11.Y: The latitude value of the crime location

12.Location: A tuple of the latitude and the longitude values

13.PdId: The police department ID

3. Analysis of Data

We examined the number of incidents during the year for the types of crimes from the data presented above. For this purpose, the data that did not constitute a final crime were removed from the record data reported to the police station. Basically, there are 39 kinds of crimes, including up to about 40000 crimes and less than 5000 different crimes.

<Numbers of crime for each category>
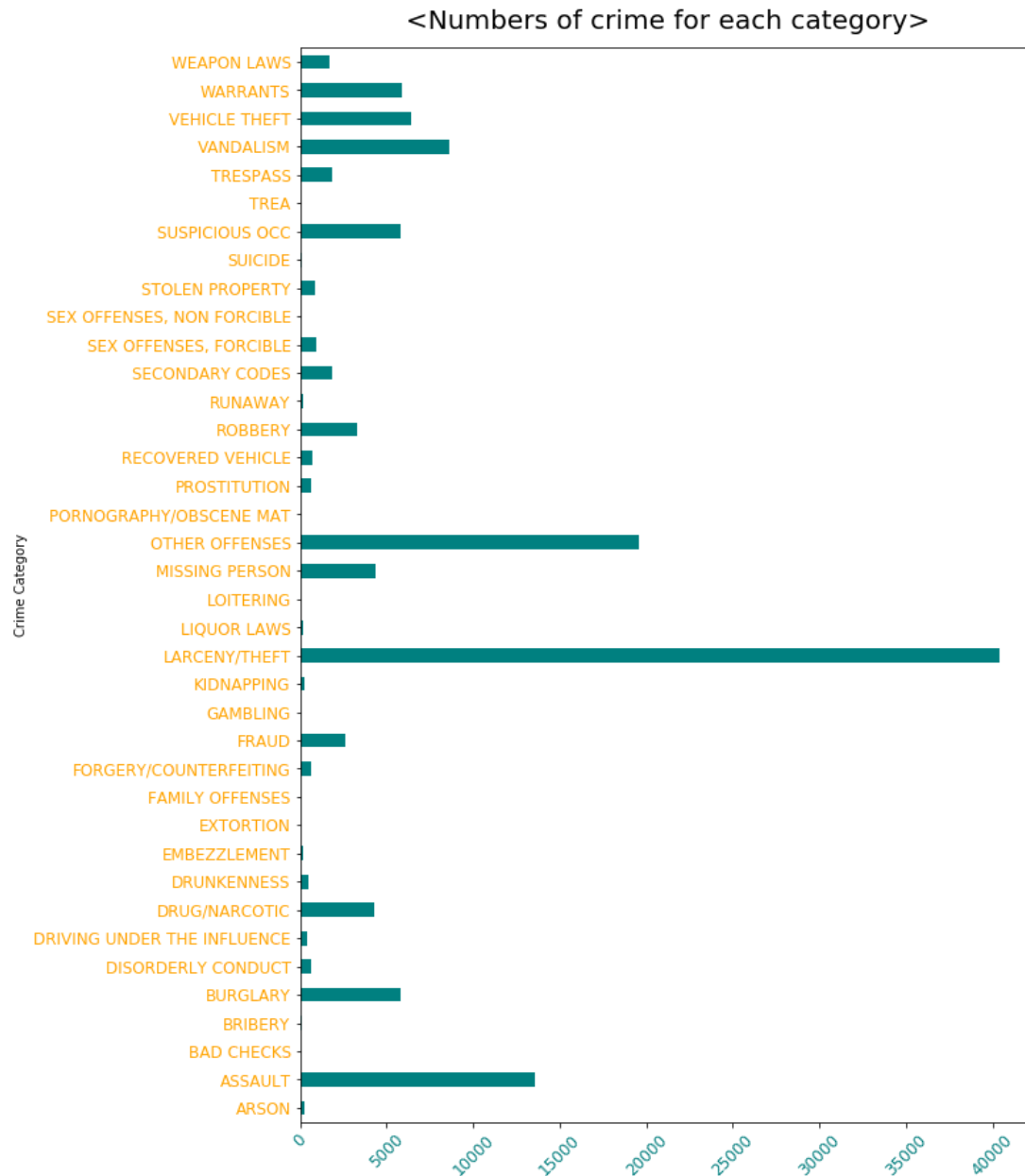
San Francisco's crimes are classified into 39 types. The most common types of crime were LARCENY/THEFT. This type of crime can be considered a dangerous threat for visitors to the city.

\<Top 5 crimes in Friday's crime cases\>
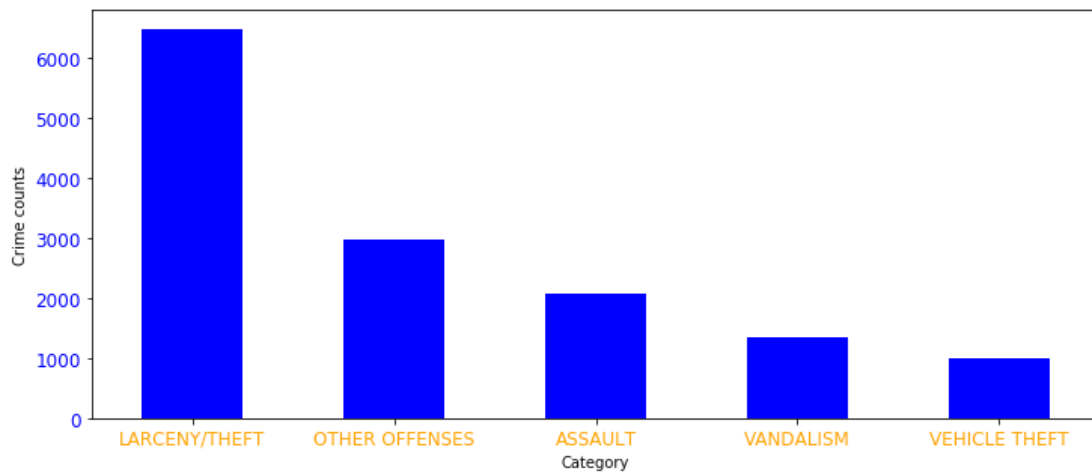
We looked at the top 5 crimes in Friday's crime cases. In the ranking of the top 5 crime types in addition to LARCENY / THEFT, tourists need to be cautious about being exposed to violent situations such as Other offenses, Assault, Vandalism, vehicle theft or renting a car.
We count numbers of crime for each days : From Monday to Sunday


\<Numbers of crime for each days\>
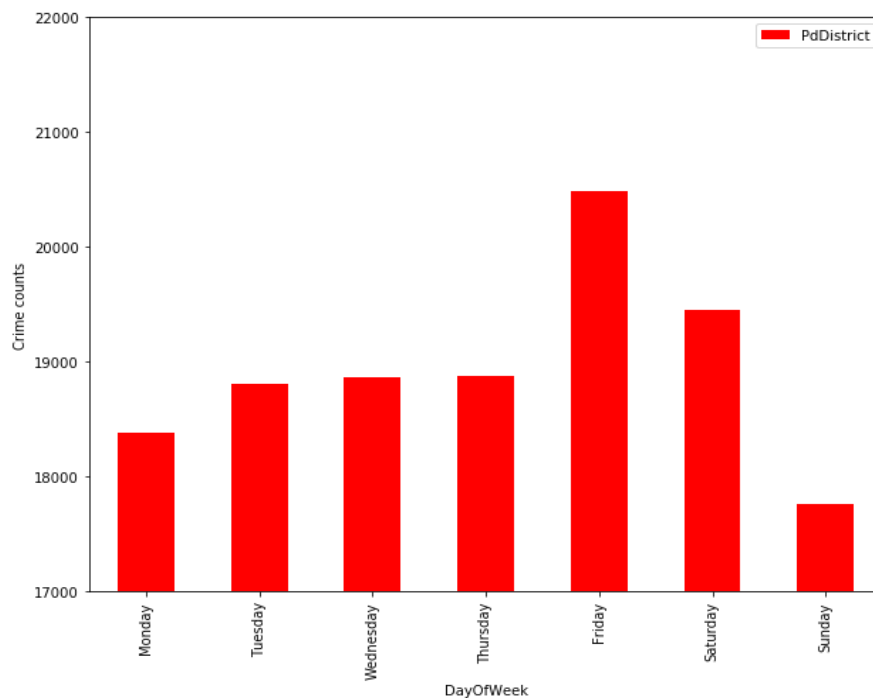
When analyzing the number of crimes per day, many crimes occurred on Friday and Saturday. On days where you enjoy more culture than on weekdays, you may find that there are more threats. This is not good for tourists, but you can see less crime on Sundays.

<Changes in crime by time>
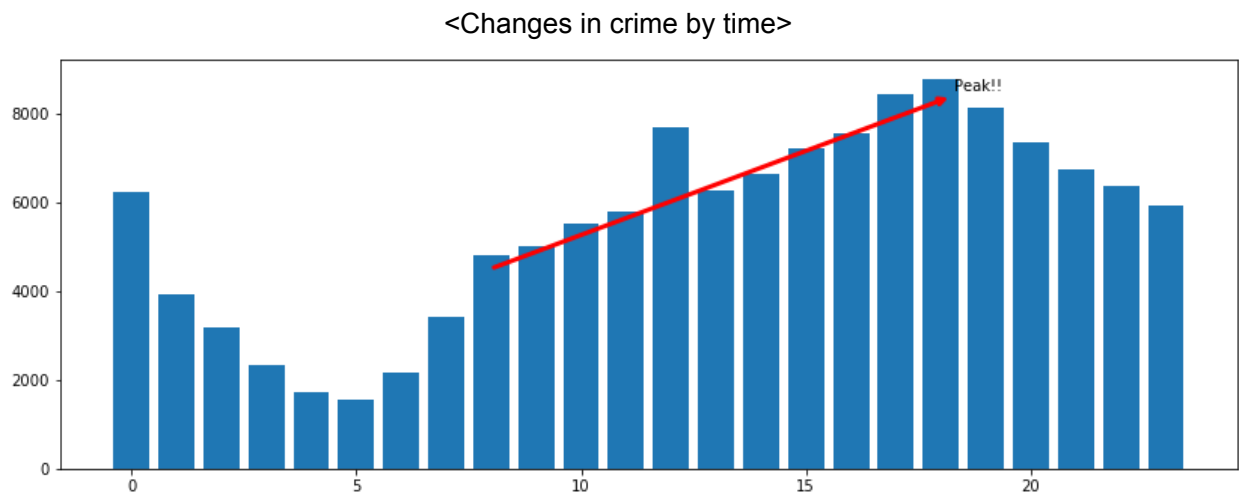
When we observed the change in crime over time, more crimes occurred in the afternoon than in the morning. In particular, at peak time 6 pm when the dinner was active, the most crimes occurred back and forth.

We also analyzed the distribution and pattern of crime occurrence locations. I would like to count the number of crimes per district in San Francisco on the map.

	Neighborhood	Count
0	SOUTHERN	25043
1	NORTHERN	18047
2	MISSION	17242
3	CENTRAL	15356
4	BAYVIEW	12991
5	INGLESIDE	10358
6	TARAVAL	9901
7	TENDERLOIN	8563
8	RICHMOND	7676
9	PARK	7456
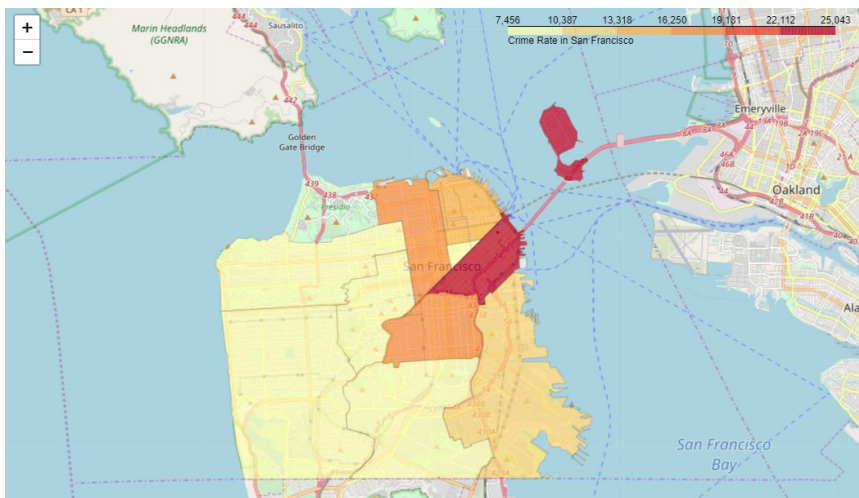
Among the total crime distributions, the pattern of occurrences of the LARCENY / THEFT crime area, which poses the greatest threat to tourists, was more clearly expressed.
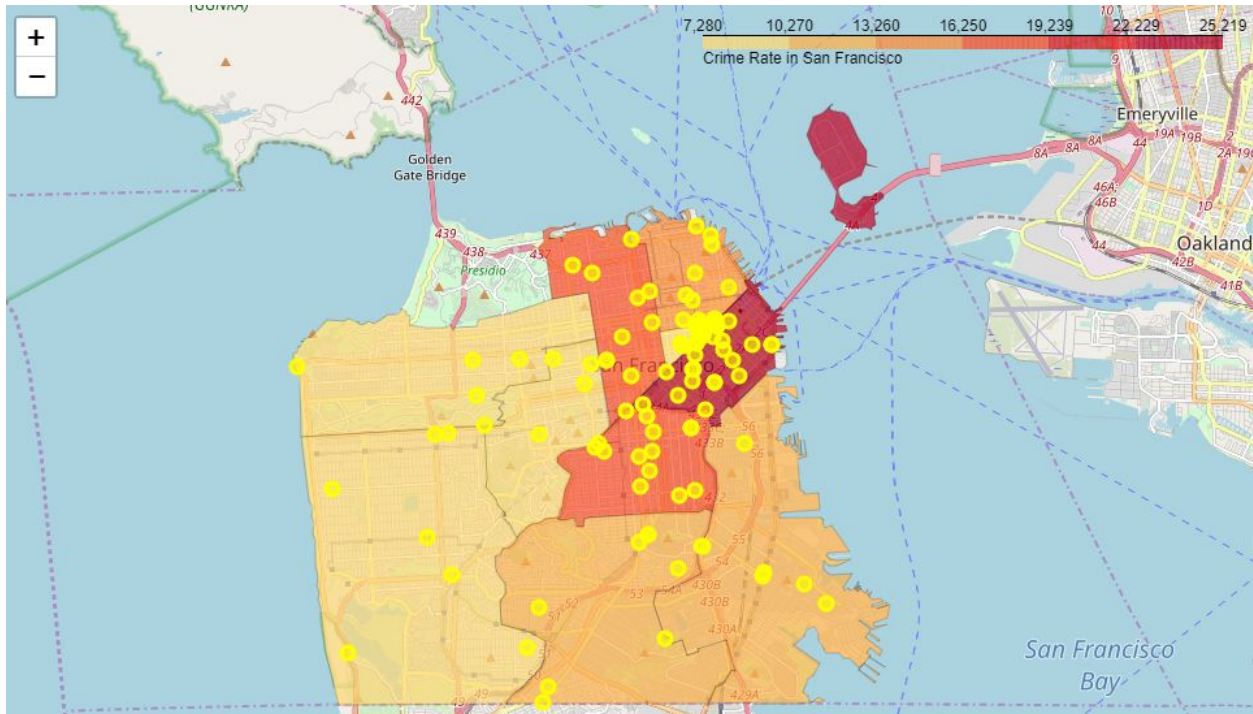
When observed based on the LARCENY / THEFT crime type, it can be seen that the entire crime occurrence is distributed in a pattern similar to the above. In particular, there is a pattern of crimes that are spread evenly in SOUTHERN, NORTHERN, MISSION, and CENTRAL areas that attract a lot of tourists.
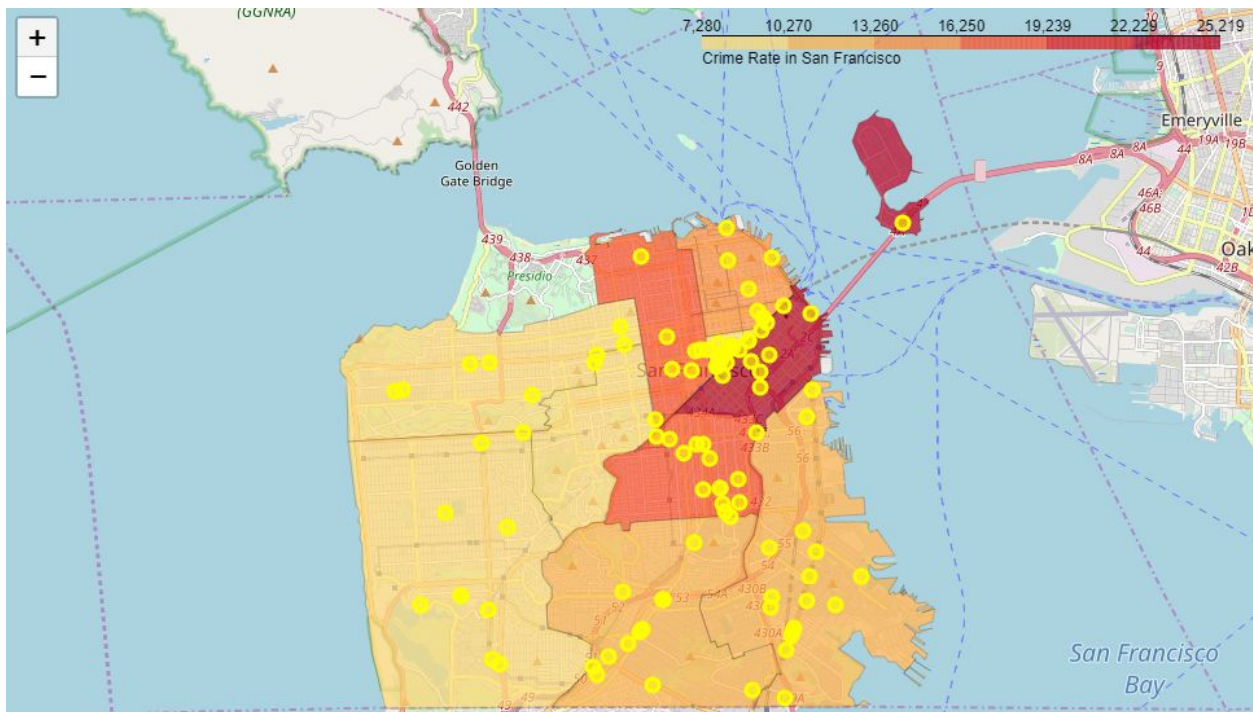
When observed based on the other crime type, it also can be seen that the entire crime occurrence is distributed in a pattern similar to the above. In particular, there is a pattern of crimes that are spread evenly in SOUTHERN, NORTHERN, MISSION, and CENTRAL areas that attract a lot of tourists.

Other crime types also follow the overall distribution. Therefore, we would like to select the area of the state where the threat would be against LARCENY / THEFT, which would be the most threat to tourists.

In addition to information related to crime distribution, restaurants in San Francisco were obtained through the Foursquare API. Up to 50 restaurants can be identified per query. To get more information, you can get a lot of information when you query restaurants of each kind. However, the survey aims to sort the list of high and low risk areas for 50 restaurants.

Foursquare API queries and restaurants were located within a 10 km radius of "the 501 Twin Peaks Blvd, San Francisco, CA", which is the central address of San Francisco.
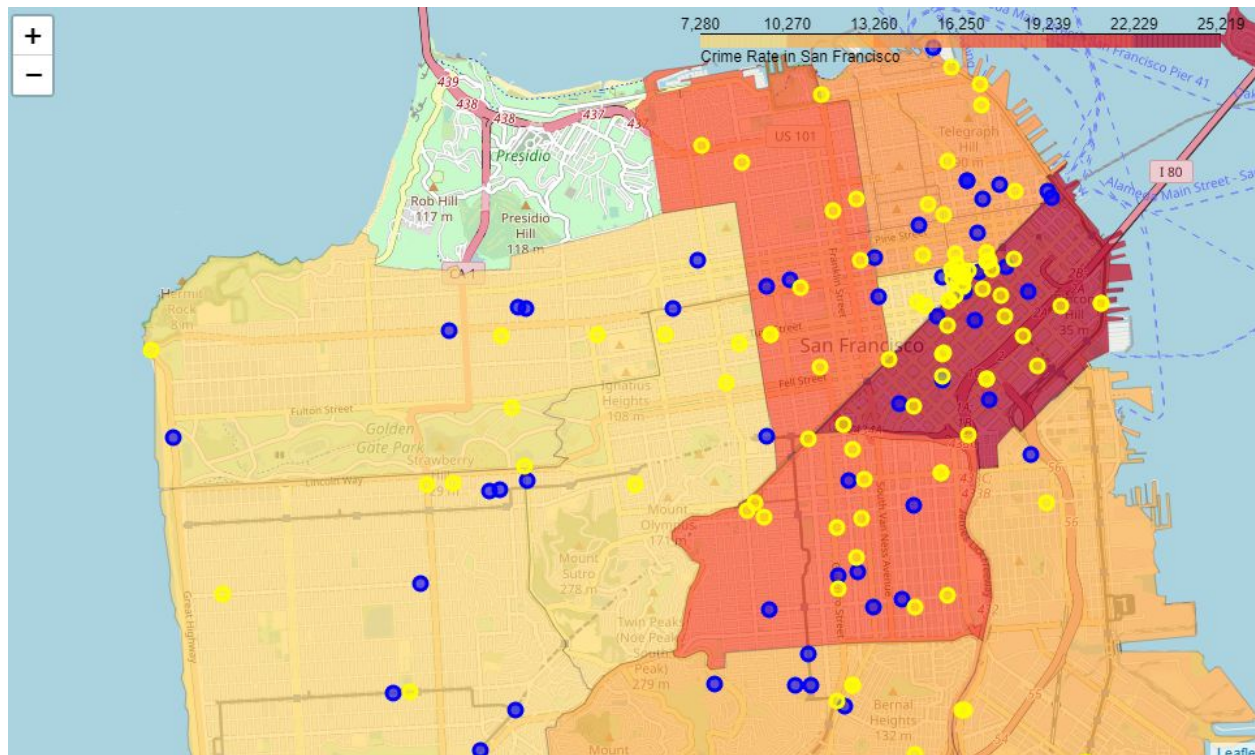
 We searched restaurants location information (blue) in San Francisco using Foursquare Query information. And it is the result of simultaneously drawing LARCENY / THEFT (yellow) which occurred in the vicinity.

The distance between the crime area and the restaurant was calculated based on the Euclidean distance equation based on the X and Y GPS location information. The calculated distance was set

as a boundary value of 0.003 (90-100 meters) based on a GPS error of about 30 meters at 0.001 unit.
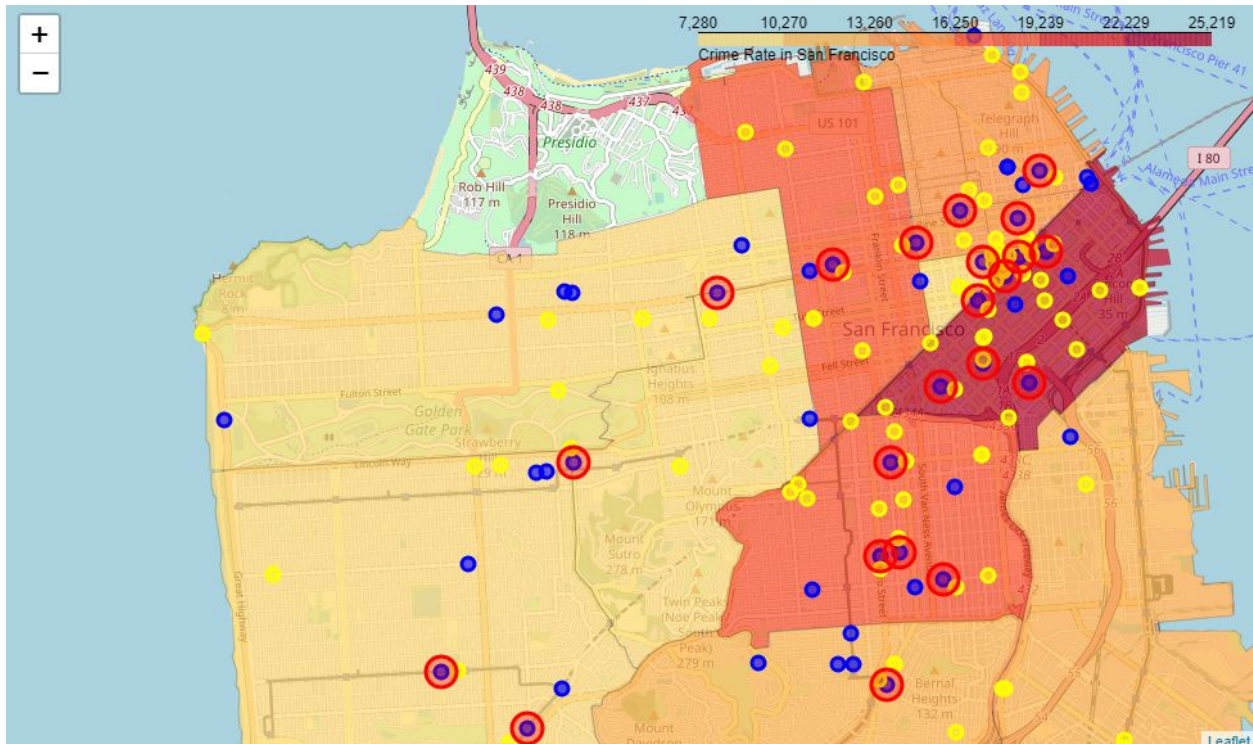
Two dimensions [edit]

In the Euclidean plane, if $\mathbf{p} = (p_1, p_2)$ and $\mathbf{q} = (q_1, q_2)$ then the distance is given by

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}.$$

This is equivalent to the Pythagorean theorem.

Ref, https://en.wikipedia.org/wiki/Euclidean_distance

If the location of the restaurant and the location of LARCENY / THEFT are less than 100 meters, then it is considered a dangerous point. If the restaurant is delicious and famous, tourists should pay attention to theft.

4. Result review

San Francisco crimes are classified into 39 types. The most common type of crime is LARCENY / THEFT. This type of crime can be considered a dangerous threat to city visitors. In the crime case on Friday, we investigated the five major crimes. In addition to LARCENY / THEFT, tourists should be careful to be exposed to crimes such as other violations, assaults, vandalism and theft of vehicles. When analyzing the number of crimes per day, more crimes occurred on Fridays and Saturdays than on weekdays. This is not good for tourists, but less crime on Sundays. Over time, more crimes occurred in the afternoon than in the morning if we observed a change in crime. In particular, the highest number of crimes occurred at 6 pm, dinner time.

Observations based on the LARCENY / THEFT crime type reveal that they are distributed in a pattern similar to the total crime occurrence. There are even patterns of crime spreading evenly in the southern, northern, mission and central areas, which attract many tourists. Other crime types follow the full distribution. Therefore, we want to give the state's information in advance by selecting the state areas that threaten LARCENY / THEFT, which is the most threatening to tourists.

Retrieved restaurant location information in San Francisco using Foursquare Query information. And as a result of simultaneously drawing the LARCENY / THEFTthat occurred nearby, it is considered a dangerous point if the location of the restaurant and the location of the LARCENY / THEFTare less than 100 meters. In these restaurants, tourists should pay attention to theft during dinner time.

5.Conclusion

Using San Francisco's crime occurrence data and Foursquare information, a conclusion is made using San Francisco's restaurant location information. According to the analysis of the data, it is divided into a list of restaurants with a history of theft near 100 meters and a list of other restaurants as follows.

- List of restaurants in the area with high risk of theft
  - Panchita's Pupuseria Restaurant #2
  - Tú Lan Restaurant
  - May Lee Restaurant
  - Tuba - Authentic Turkish Restaurant
  - Restaurant Anzu
  - Basil Thai Restaurant & Bar
  - MKT Restaurant and Bar
  - Tony's Cable Car Restaurant
  - The Restaurant Collection Under The Dome
  - Akiko's Restaurant & Sushi Bar
  - Sun Rise Restaurant
  - Heung Yuen Restaurant
  - Big 4 Restaurant
  - Don Ramon's Mexican Restaurant
  - Mars Bar & Restaurant
  - Ryoko's Japanese Restaurant & Bar
  - San Wang Restaurant
  - Clay Oven Indian Restaurant
  - Nena's Restaurant
  - Sai's Vietnamese Restaurant
  - New Ming's Restaurant
  - 8 Immortals Restaurant

- List of restaurants in the area with low risk of theft
  - San Tung Chinese Restaurant 山東小館 (San Tung Chi...
  - Cafe Bakery & Restaurant
  - Kabuto Restaurant
  - Eric's Restaurant
  - New Eritrea Restaurant & Bar
  - Henry's Hunan Restaurant
  - The Chieftain Irish Pub & Restaurant
  - Turtle Tower Restaurant
  - Evergreen Garden Restaurant
  - Bullshead Restaurant
  - City View Restaurant
  - Pa'ina Lounge & Restaurant
  - Alice's Restaurant
  - Taiwan Restaurant 台灣飯店
  - Kamei Restaurant Supply
  - 750 Restaurant & Bar
  - Trimark Economy Restaurant Fixtures
  - Chava's Mexican Restaurant
  - Z & Y Restaurant
  - Kezar Bar & Restaurant
  - Osha Thai Restaurant & Lounge
  - One Market Restaurant
  - Pier Market Seafood Restaurant
  - Chavita's Restaurant
  - NOVY Restaurant
  - Canton Dim Sum & Seafood Restaurant
  - Kufu-ya Japanese Restaurant

Based on the information from these studies, tourists hope to enjoy safe restaurant visits and enjoy San Francisco more.