

# **Coursera Capstone**

IBM Applied Data Science Capstone

## **Establishment of a Japanese restaurant in San Francisco**

In the safest neighborhood and least restaurant

By Huzaifa Muhammad Abdul Wahab

1/27/2021



## **Introduction:**

We have a customer who wants to open a restaurant in San Francisco.

This customer asks us to find a suitable place for him. The customer explained to us that he had failed several times due to the selection of inappropriate places and had been forced to close his restaurant.

The biggest factors that made a region unsuitable for him were the low security of the region and the large number of competitors in the region.

We asked him to explain to us what was the right place for him and what his characteristics were.

The customer explicitly stated that the best place for him was where he was safe (the least crime was committed) and that the number of Japanese or Asian restaurants was the lowest.

## **Data / Source:**

We use San Francisco crime dataset “Police-Department-Incident-Reports2018-to-Present” from [data.sfgov.org](https://data.sfgov.org).

This dataset includes police incident reports filed by officers and by individuals through self-service online reporting for non-emergency cases. Reports included are those for incidents that occurred starting January 1, 2018 onward and have been approved by a supervising officer.

Disclaimer: The San Francisco Police Department does not guarantee the accuracy, completeness, timeliness or correct sequencing of the information as the data is subject to change as modifications and updates are completed.

In total 338396 rows and 36 Columns in raw dataset miss values and noncriminal values is dropped. Clean data contains 2299574 rows and 4 columns.

For neighborhood areas and name we use “Relator Neighborhoods” from Data.sfgov.org it have 92 rows and 4 columns.

## Methodology

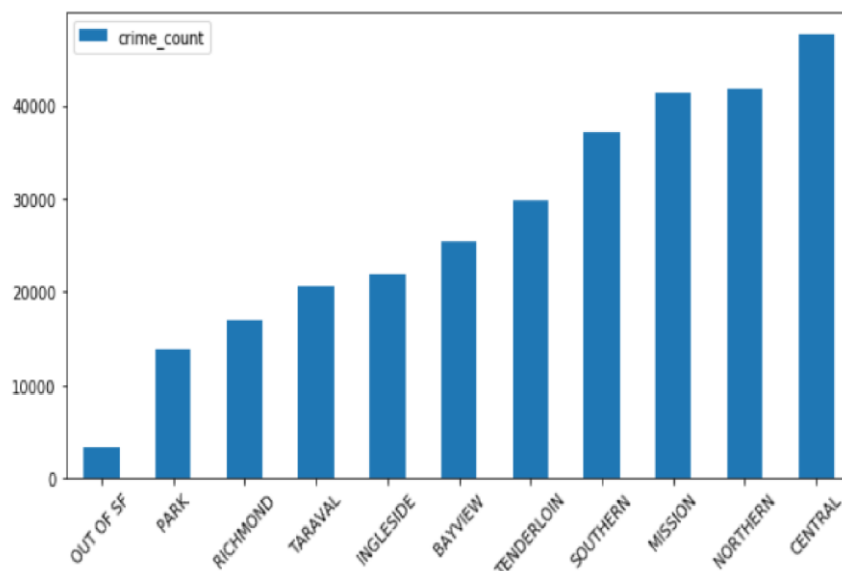
In this project, we are looking for the safest area and safest neighborhood in San Francisco where there are few Japanese restaurants.

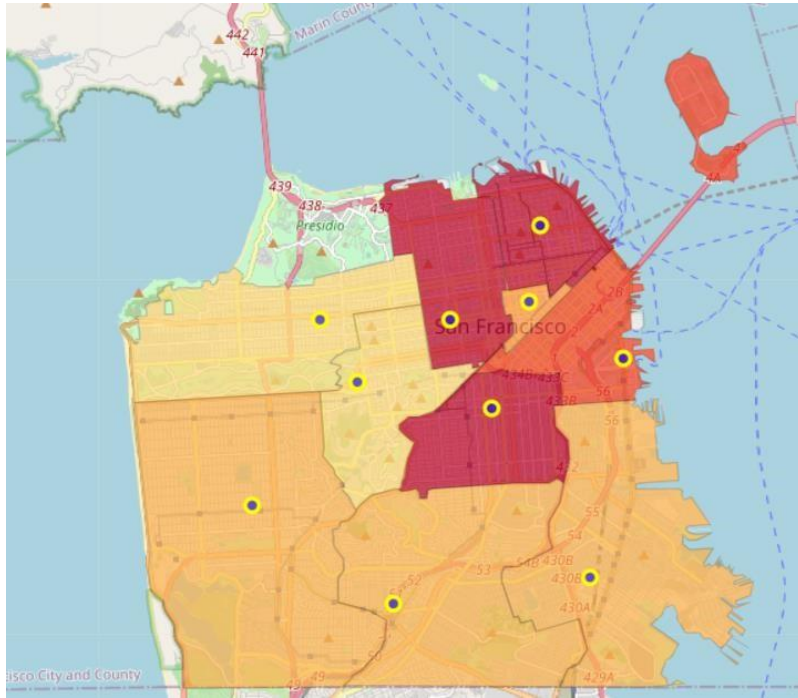
In the first step, using dataset, we separate the crimes in the San Francisco area and select the safest area.

Then, using the neighborhood dataset and the location of each crime, we determine the number of crimes in each neighborhood.

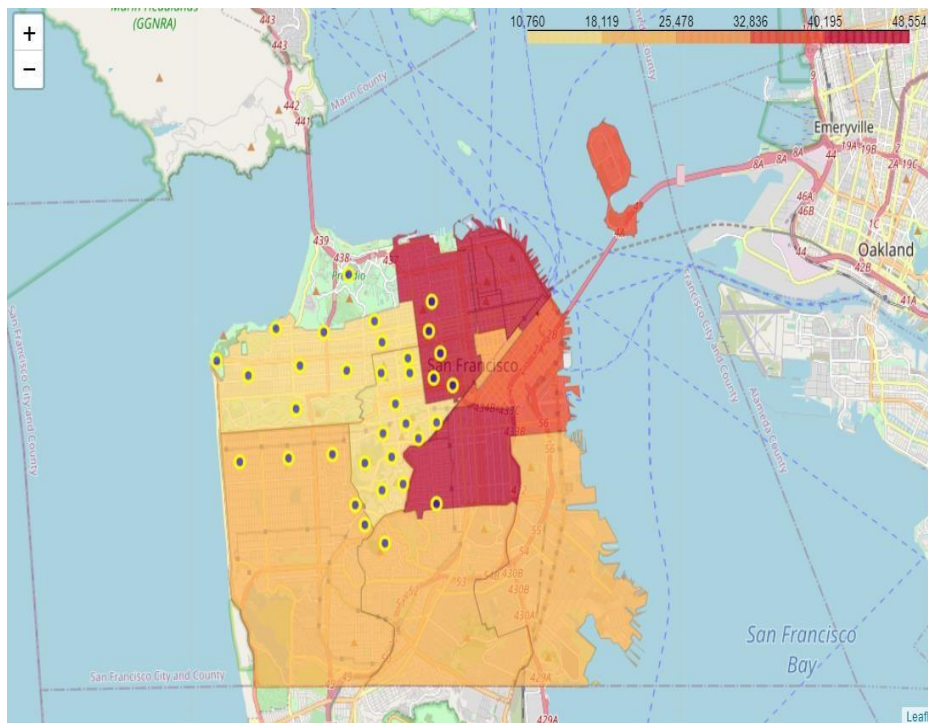
We then use Foursquare and k-mean Clustering to categorize the neighborhoods to see what the distribution of venue in San Francisco is like, and then use the previous information to select the safest neighborhood that has the fewest restaurants, especially the Japanese restaurant.

We find and show the Park and Richmond District is safest.

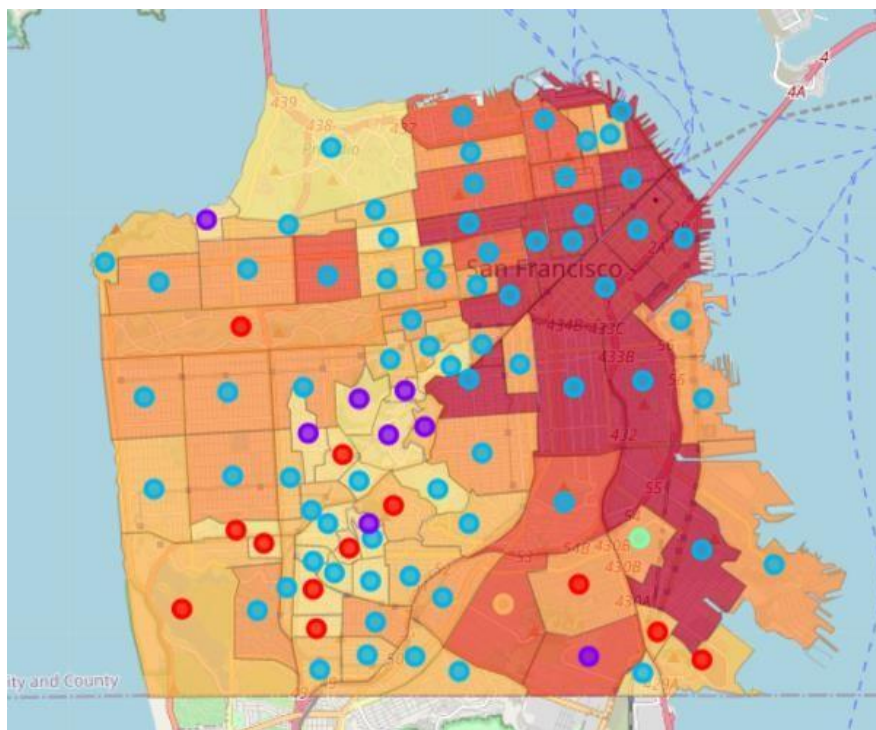




And in Park district, Midtown terrace, Forest Hills Extension, Clarendon Heights Is the safest neighborhoods.



most restaurants, coffee shops, etc. are at the top right and bottom of the map, which continues to the center of the map. (Blue based on cluster number 2)



## **Results and Discussion**

We divided our data into five clusters, the number zero is marked in red, with the park, playground, and gym being the most common venue, followed by restaurants as second place.

In the number one cluster, which is marked in purple, trails, parks and gardens are the most common venue.

number two, which is marked in blue, has the most coffee shops, restaurants, etc.

It has the number three marked in green, has the most Grocery store, soccer field.

Number four marked in orange and has, Convenience Store, lake and scenic lookout.

Based on our analysis of the data obtained, most restaurants, coffee shops, etc. are at the top right and bottom of the map, which continues to the center of the map. (Blue based on cluster number 2)

In the center of the map, the number of these cases decreases. Since our priority was first to select the safest area, we first select the Park district. After this district, we can consider the Richmond district. In the park district, Midtown terrace has the lowest crime rate, which is marked in blue number one cluster. This cluster mostly includes trailers, parks and playgrounds.

After that, we can name Forest Hills Extension neighborhood in cluster number 2 and finally Clarendon Heights neighborhood in cluster number one.

## **Conclusion**

Purpose of this project was to identify the safest areas of San Francisco and then the safest neighborhood in that area with a small number of Asian or Japanese restaurants. To establish a Japanese restaurant. Using clustering, we were able to identify areas where the number of restaurants is low and introduce three desirable neighborhoods in the safest area.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.