Establishment of a Japanese restaurant in San Francisco

In the safest neighborhood and least restaurant

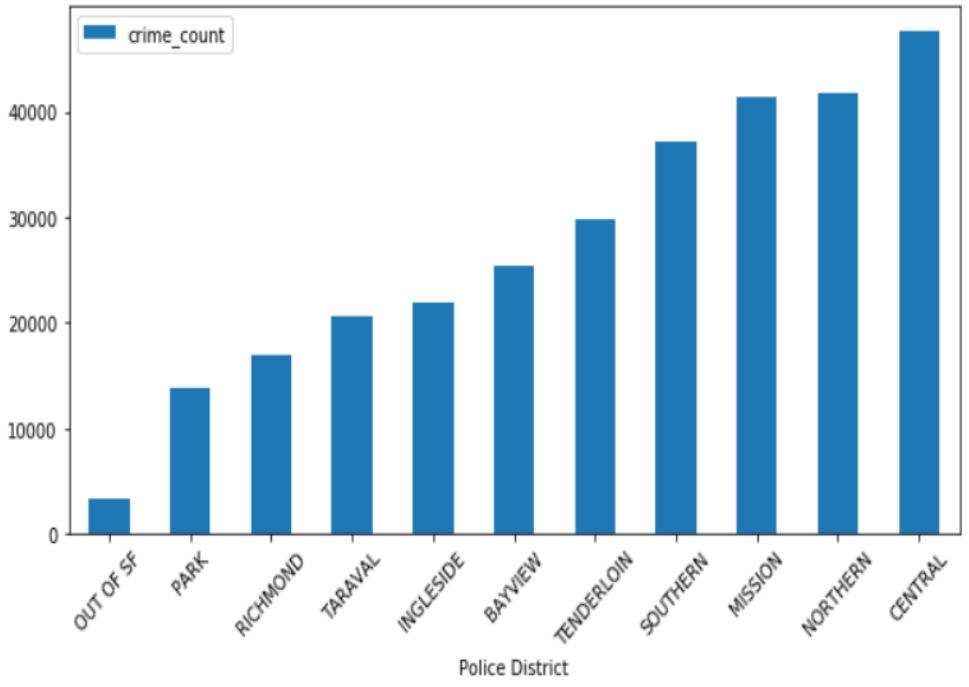
Finding the safest neighborhood is important for our employer.

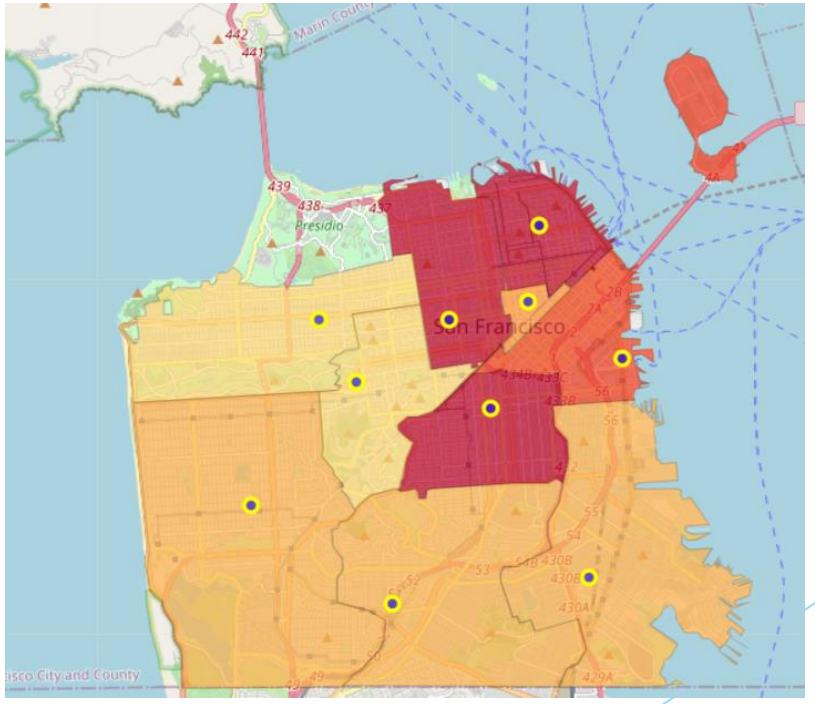
- We have a customer who wants to open a Japanese restaurant in San Francisco.
- This customer asks us to find a suitable place for him. The customer explained to us that he had failed several times due to the selection of inappropriate places and had been forced to close his restaurant.
- The biggest 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 acquisition and cleaning

- San Francisco crime dataset (2018 to present) from https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783
- In total 339K rows and 26 Columns in raw dataset miss values and non-cime values is dropped.
- Clean data contains 299K rows and 4 columns.
- San Francisco Realtor Neighborhoods (August 2010) from https://data.sfgov.org/Geographic-Locations-and-Boundaries/Realtor-Neighborhoods/5gzd-g9ns/data

The safest areas of San Francisco with the location of the police station on it.

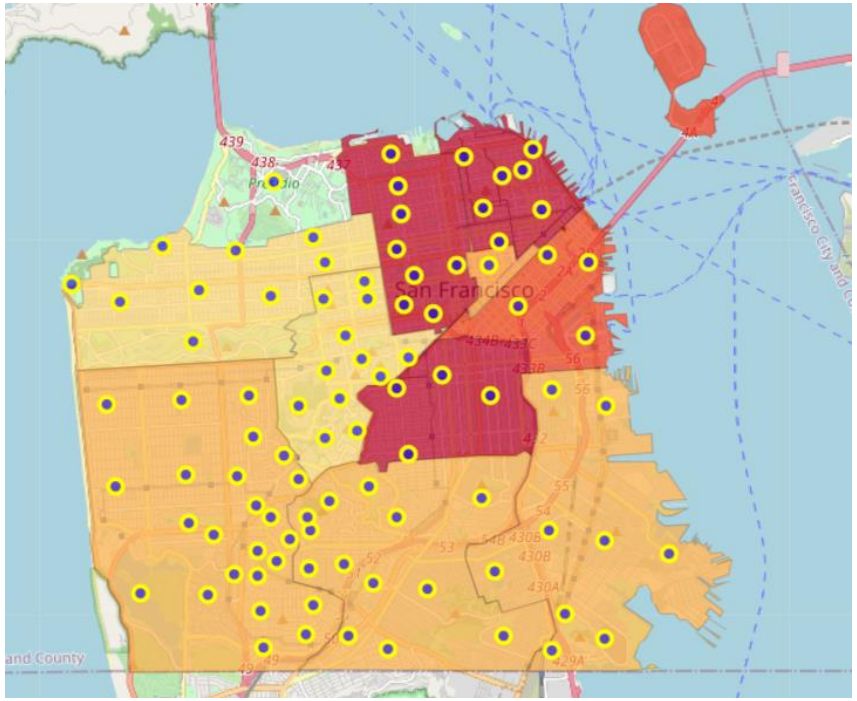




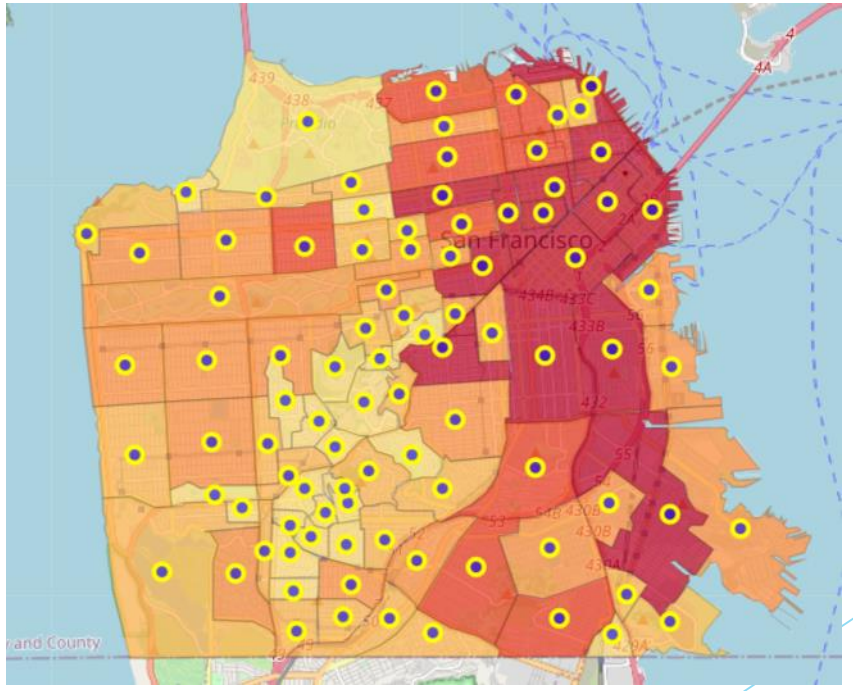
Map of Neighborhood area and District Area

(a lighter color indicates a lower crime)

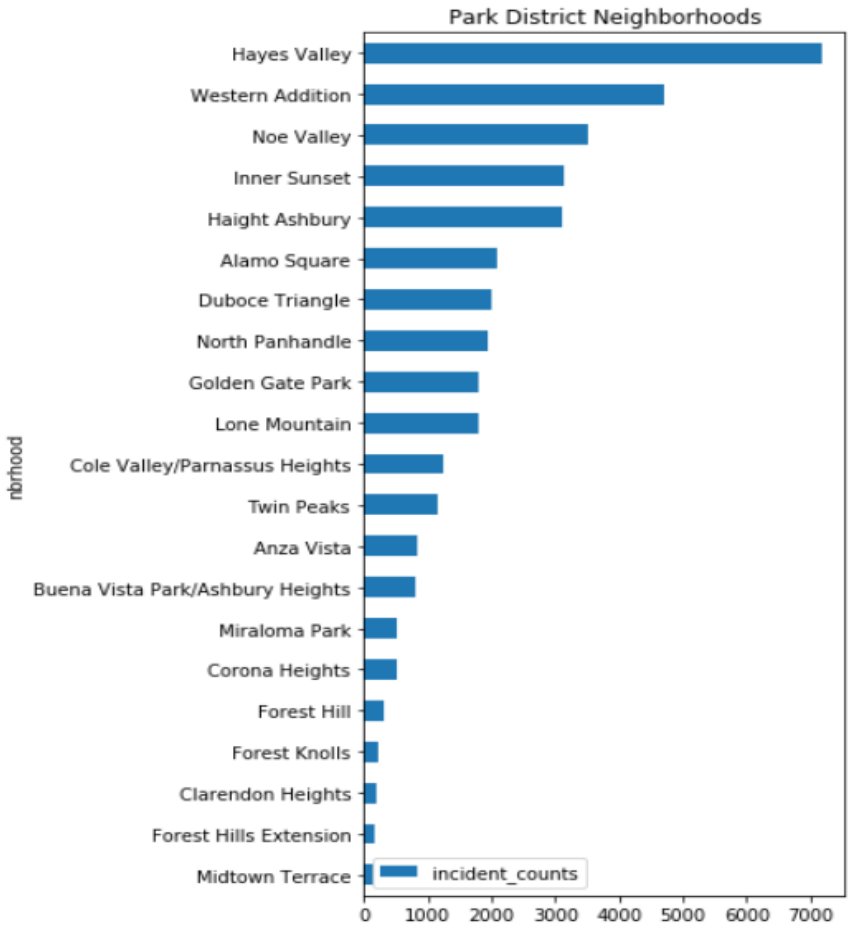
District area and neighborhood points

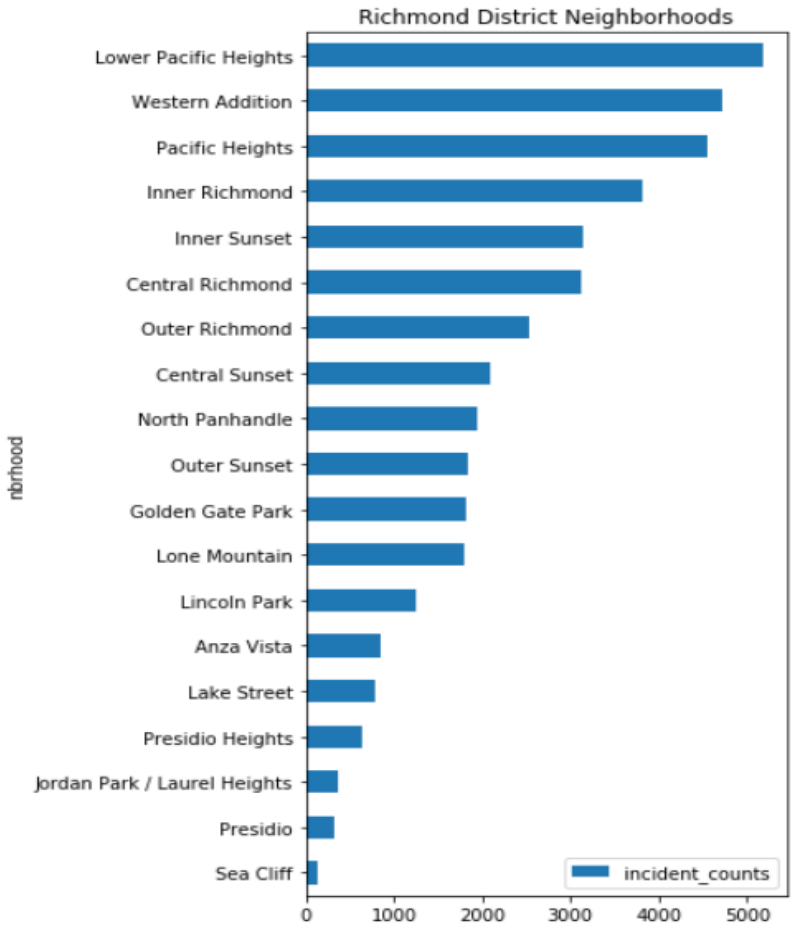


Neighborhoods area



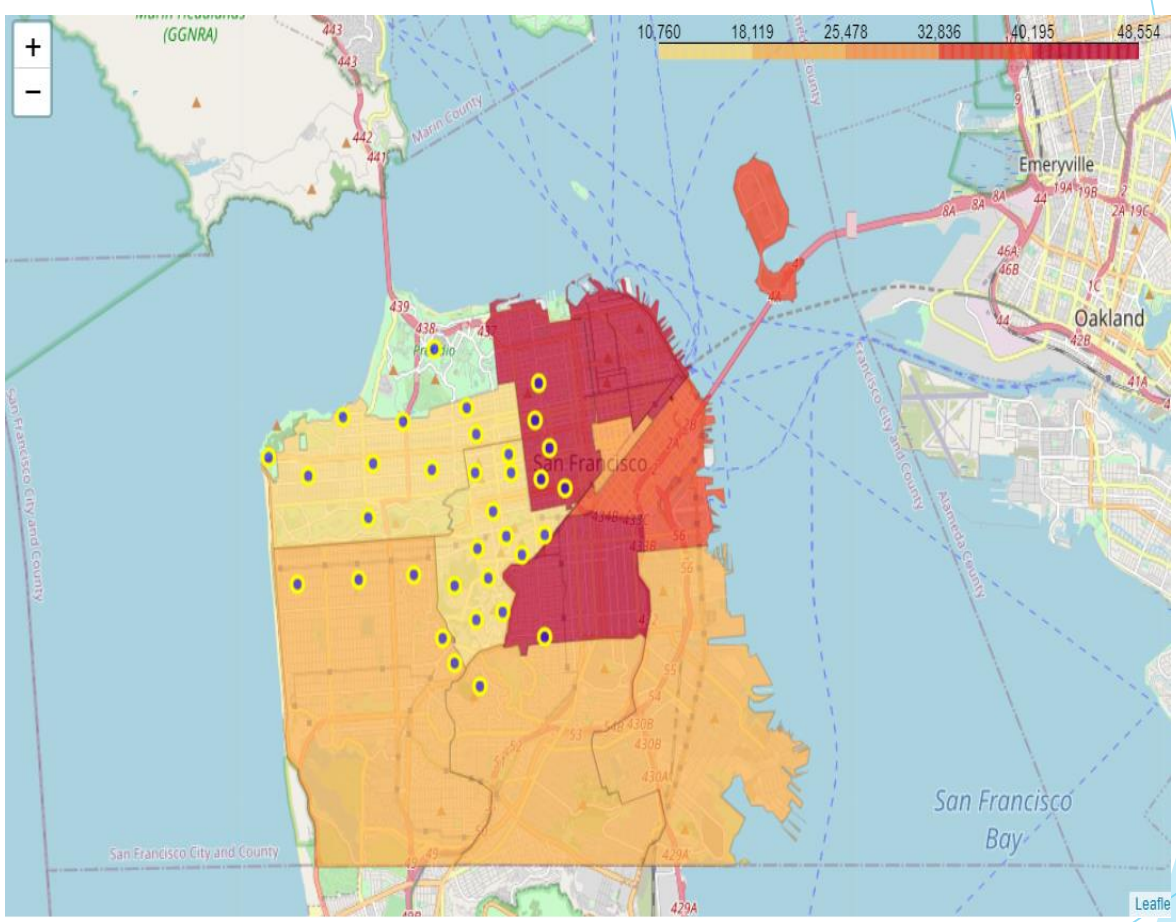
The number of crimes committed in the neighborhoods of the two selected districts



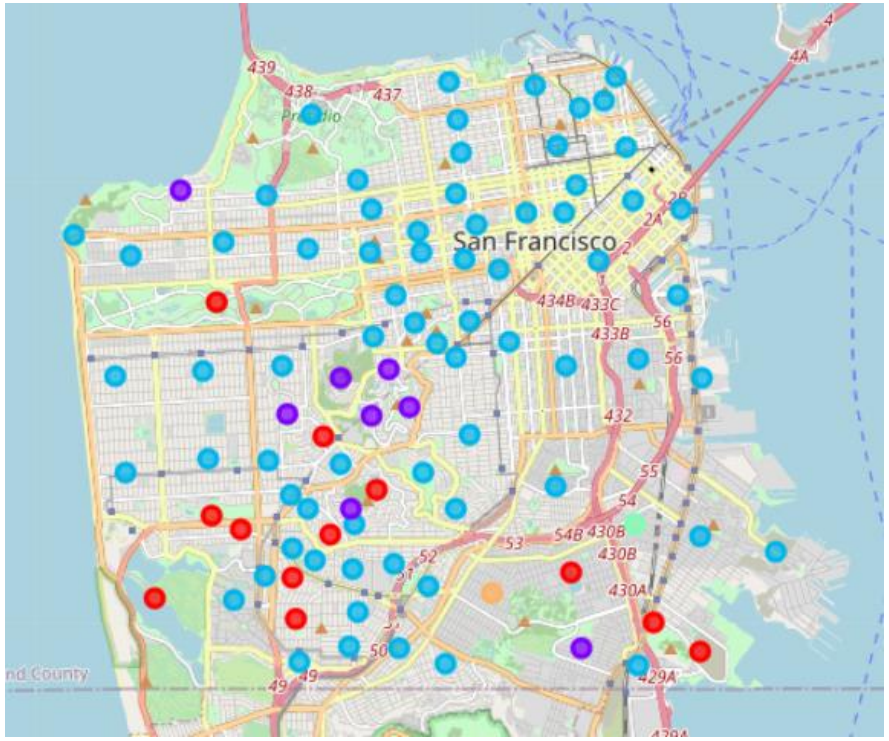


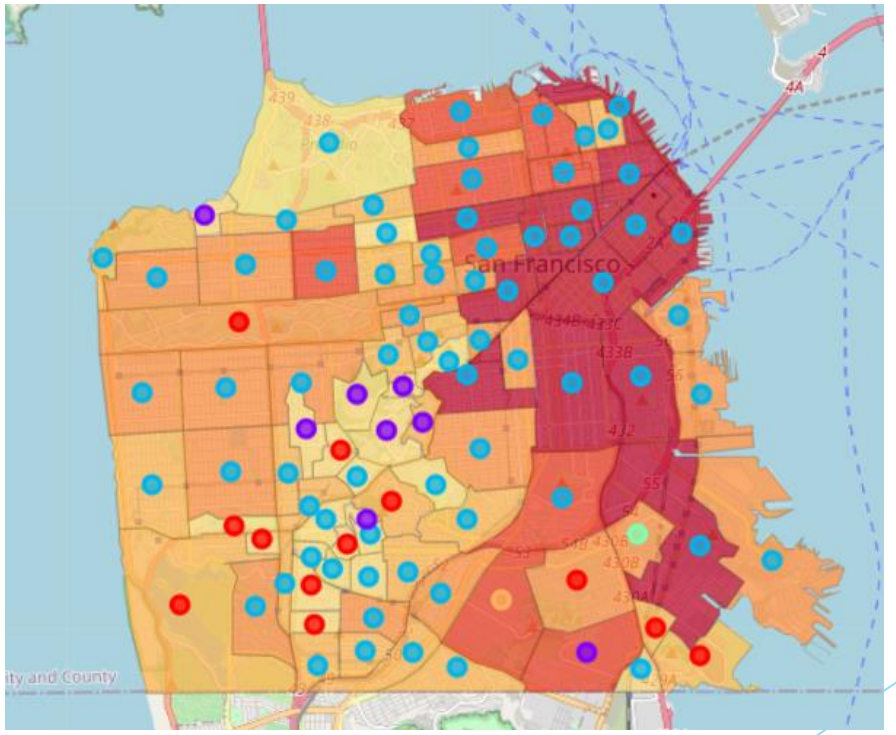
Some neighborhood areas is in two or more district.

We can see Park and Richmond district whit there neighborhoods point.



Five clusters and neighborhood areas





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 decission 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.