

Coursera Capstone Project

Restaurant Analysis in San Francisco

Final Report

Beygan Sundaramurthy



2020

Business Problem

The target audience for this project are the people looking to find the best spot in San Francisco, California, US to setup a restaurant or fast food and feel confident that they will have enough customer to run their business.

City of San Francisco is one of the busy cities in California state and a big city. There are so many restaurants and fast foods available in the city. That also doesn't mean every part of the city has restaurants or all types of cuisine is available in every parts of city. There may be neighborhoods where there are no restaurants or less restaurants with more people living or walking around the place or there may be a specific cuisine type not available in a neighborhood. These neighborhoods are our areas of interest. The goal of this project is to find the best neighborhood to setup a restaurant of a specific cuisine type.

In this project we will specifically focus on setting up a Chinese/Asian cuisine restaurant in San Francisco, California, US.

Data and Cleaning

San Francisco Neighborhood

The list of neighborhood in San Francisco will be extracted from the Wikipedia page https://en.wikipedia.org/wiki/List_of_neighborhoods_in_San_Francisco using the BeautifulSoup library for web page scrapping. The neighborhood list will be extracted one by one from the web page and saved as a list. Using the geocoder library from Nominatim, we will get the latitude and longitude detail for each neighborhood in the list. If location information is not available for any neighborhood, we will drop those neighborhoods for our analysis.

Restaurants/Fast foods in San Francisco

Using the Foursquare API, we will get the list of all restaurants/fast foods using the Food category and their sub-categories in each neighborhood to understand how many restaurants are available and the type of cuisine of each restaurant. Separate out the Chinese/Asian cuisine from that list to understand which neighborhoods have this type of cuisine and how many are there.

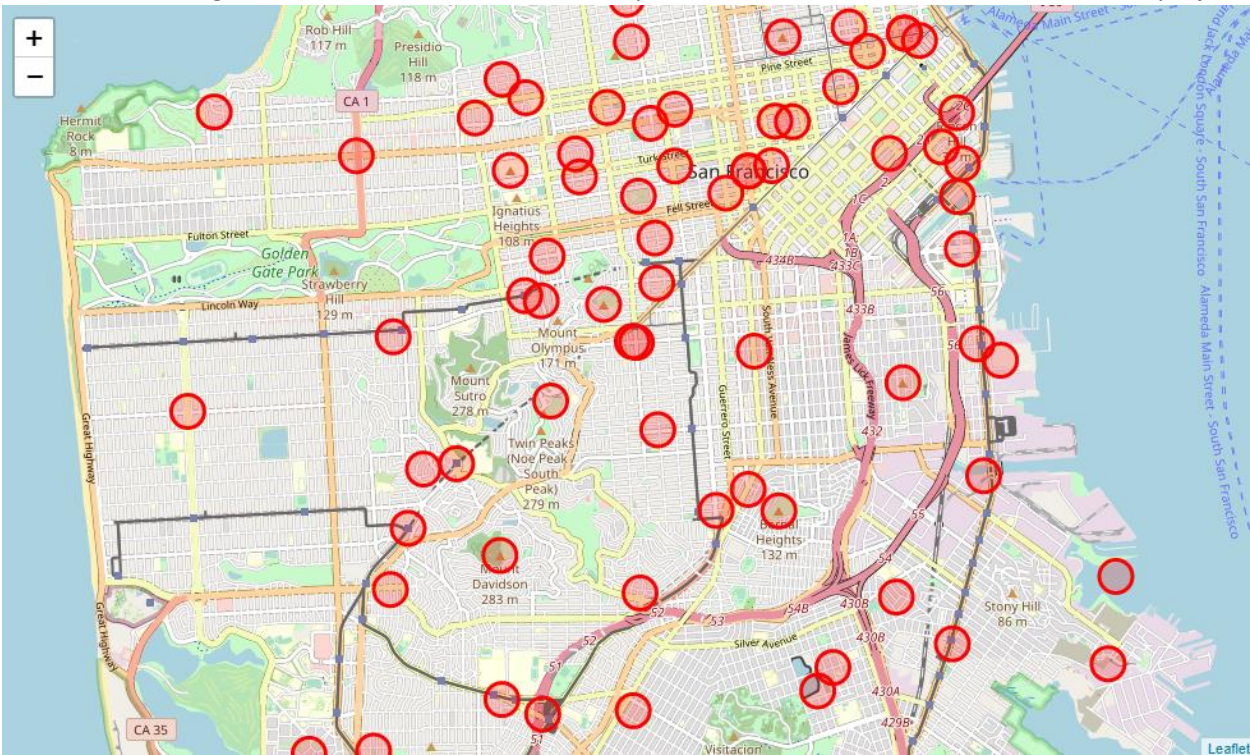
People Venues

Get the list of venues in each neighborhood where we expect more people to be there and see a high chance of those people using the restaurants/fast foods. For example, if there are lot of shops or parks, then there is high possibility that people can visit the restaurants in those locations. If there are lot of night clubs, then there is low possibility of people visiting the restaurants in that location.

Methodology

Neighborhood Location and Size

The neighborhoods of San Francisco (SFO) is explored and determined to consider 200 meters radius for each neighborhood for exploration of this project.



Food Venue Exploration

Using the Foursquare API Get Venue Recommendations, retrieve all the food category venues detail in each neighborhood of SFO. The required key attributes for each venue are venue category and location coordinates. The venue category explains the type of eat out and the type of cuisine. The neighborhoods of SFO had 88 different types of venue categories to explore.

	Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Category
0	Alamo Square	37.776360	-122.434689	Lady Falcon Coffee Club	37.776303	-122.433174	Food Truck
1	Anza Vista	37.780836	-122.443149	Kaiser Permanente Cafeteria - 1st Floor	37.782563	-122.443595	Cafeteria
2	Balboa Park	37.721427	-122.447547	AJ's Barbeque Cafe	37.720757	-122.446280	BBQ Joint
3	Balboa Park	37.721427	-122.447547	Crazy Pepper	37.720832	-122.446663	Asian Restaurant
4	Balboa Park	37.721427	-122.447547	Tai Shan Restaurant	37.721012	-122.445922	Food

To make the exploration easier, the venue categories are converted to table columns and the count of venues in each category is summed up at the neighborhood level. Each neighborhood now has the detail of number of food venues under each category. All the food venues in each category is accumulated to get the total number of food venues in each neighborhood. Since our focus is setting up a Chinese/Asian

restaurant, we determine the number of Chinese/Asian restaurants in each neighborhood. There were 2 categories to consider such as Chinese Restaurants and Asian Restaurants for this analysis.

	Neighborhood	Latitude	Longitude	Total Restaurants	Chinese Restaurants
0	Alamo Square	37.776360	-122.434689	1.0	0.0
1	Anza Vista	37.780836	-122.443149	1.0	0.0
2	Balboa Park	37.721427	-122.447547	8.0	1.0
3	Bayview	37.728889	-122.392500	7.0	0.0
4	Belden Place	37.791744	-122.403886	35.0	0.0
5	Bernal Heights	37.742986	-122.415804	0.0	0.0
6	Buena Vista	37.806532	-122.420648	17.0	0.0
7	Castro	37.760856	-122.434957	35.0	0.0
8	China Basin	37.776330	-122.391839	12.0	1.0
9	Chinatown	37.794301	-122.406376	26.0	11.0

Other Venues Exploration

Now it's time learn more about each neighborhood to know how busy the neighborhood is. From the list of categories available in Foursquare, we pick few categories where there will be more people living or walking around in those venues. This gives us an idea on how busy the neighborhood is, and probability of customers being attracted towards restaurants. For this analysis we used the following categories...

- Stadiums
- Colleges
- Parks & Recreations
- Shops
- Apartments/Residence
- Hotels

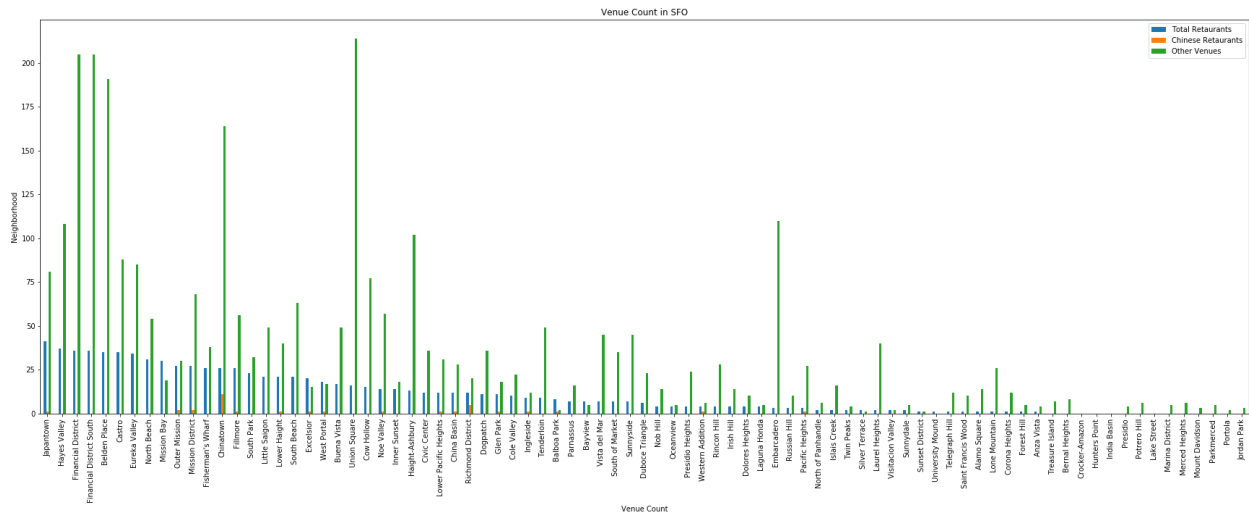
Retrieve the counts of these busy venues in each neighborhood and combine with the restaurant data. This provides all the data we need to use for our analysis to determine the best neighborhood to setup a restaurant.

Exploration of all data

We will be using the following data for each neighborhood for further analysis...

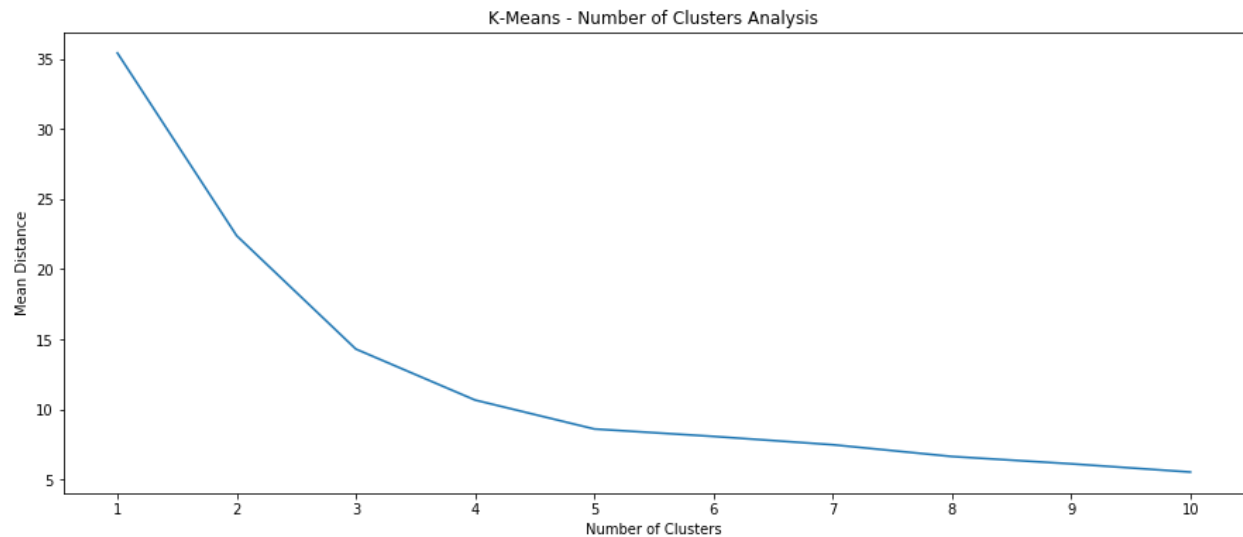
- Total Restaurants count
- Chinese Restaurants count
- Busy venues count

Each neighborhood is explored with these data to get an idea on how each neighborhood looks.

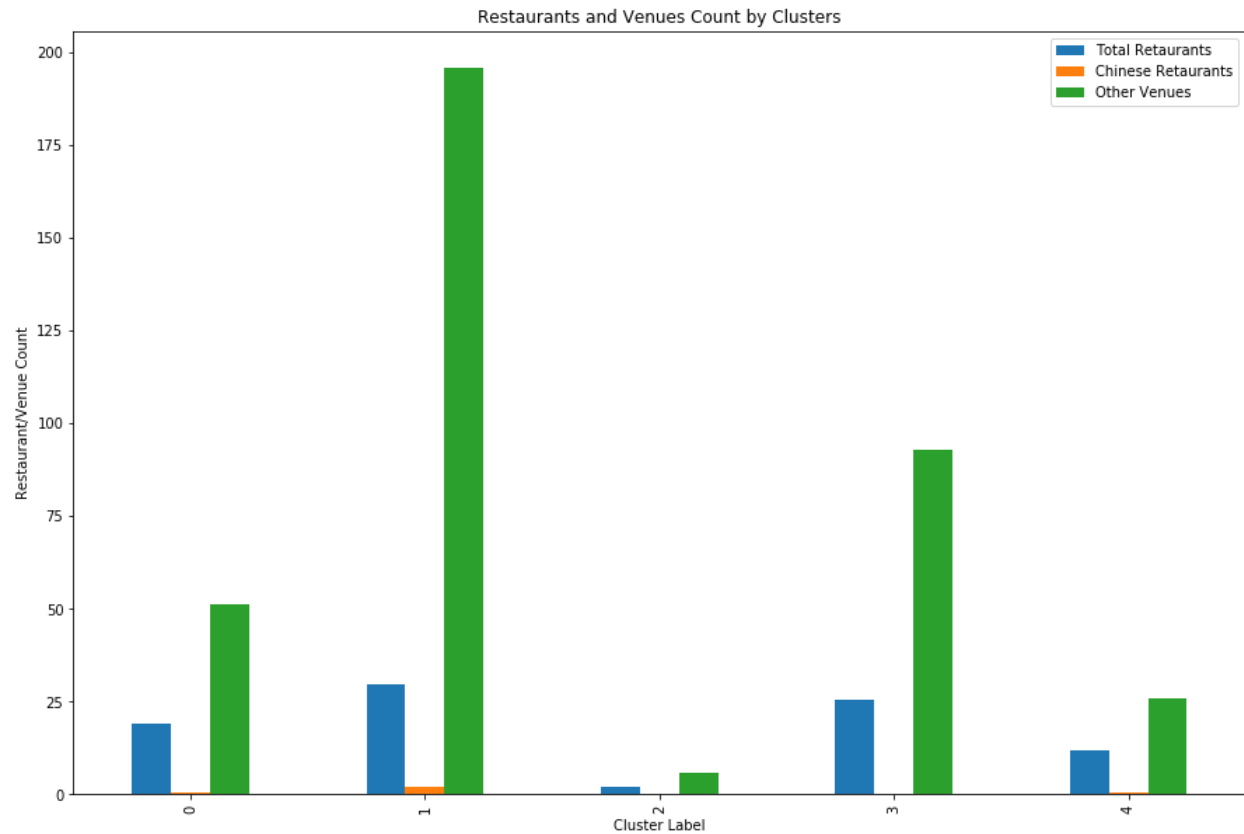


Cluster Analysis

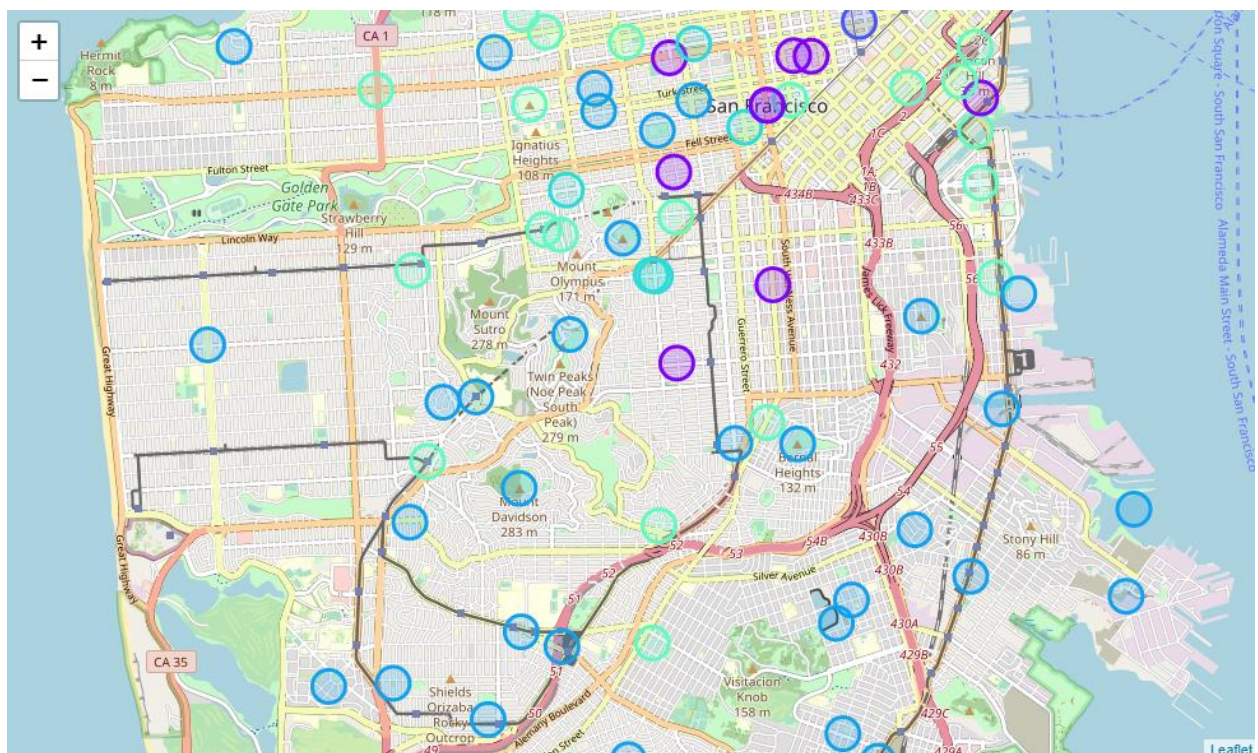
For this analysis we will be using the K-Mean clustering to cluster these neighborhoods based on the data Total Restaurants count, Chinese Restaurants count, and Busy venues count we have for each neighborhood. The data set is tested with varying number of clusters to determine the correct and minimum number of clusters required. The testing was done starting from 1 cluster to 10 clusters. For each iteration of testing with different number of clusters, the mean distance for data points to centroid is calculated and plotted against the K value (number of clusters). The elbow point where the reduction in mean distance slows down is picked as the optimal number of clusters for this analysis. Based on the graph, the optimal number of clusters is determined as 5.



In the final step each cluster is analyzed to determine which cluster is the best one to setup a new Chinese/Asian Restaurant. The bar graph is plotted using the mean values for each cluster to give a visual indicator on which clusters seems close to the expectation. The cluster is analyzed in detail to find the good neighborhood for setting up the new restaurant.



Cluster classification in the SFO map:

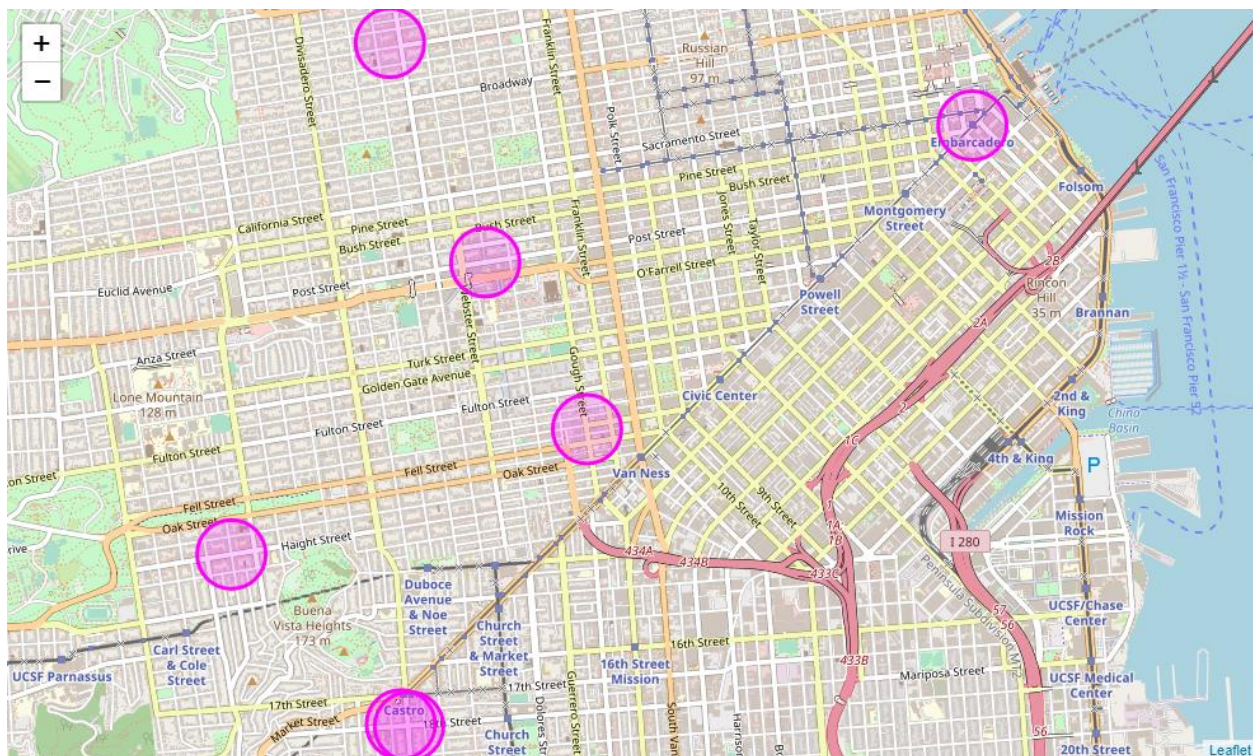


Results & Discussion

Based on the K-Means clustering mean distance analysis, the neighborhoods are grouped into 5 clusters. Analyzing these 5 clusters, the cluster with the neighborhoods **Castro, Cow Hollow, Embarcadero, Eureka Valley, Haight-Ashbury, Hayes Valley, Japantown** is selected as one of the best clusters to setup a Chinese/Asian Restaurant.

Neighborhood	Total Restaurants	Chinese Restaurants	Other Venues
Castro	35	0	88
Cow Hollow	15	0	77
Embarcadero	3	0	110
Eureka Valley	34	0	85
Haight-Ashbury	13	0	102
Hayes Valley	37	0	108
Japantown	41	1	81

This cluster has neighborhoods with lot of busy venues where lot of people moving around. At the same time there are no Chinese/Asian restaurants in these neighborhoods. The interesting part of this cluster is, all the neighborhoods has many restaurants except for one. This gives an idea that restaurants in these neighborhoods are a good business and people are interested in visiting the restaurants in these neighborhoods. These neighborhoods can be strongly considered to setup a Chinese/Asian restaurant.



Conclusion

This project was done to identify the best locations in the San Francisco city, California to setup a Chinese/Asian Restaurant and run a successfully business. This analysis was done using the neighborhood data from Wikipedia and the venue data from Foursquare. The venue data has been classified as all restaurants, Chinese/Asian restaurants and other busy venues to characterize a specific neighborhood. Based on these characters we formed clusters of neighborhoods. These clusters will be analyzed to determine the perfect neighborhood to setup the restaurant. This project is flexible, and the same process can be used to setup any type restaurant by tweaking the customizable parameters and utilizing the same process.