Capstone Project

The Battle of Neighborhoods

Table of Contents

Introduction

A chef wants to open his own restaurant in the city of San Francisco. Indian cuisine is his forte. He wants to find a neighborhood with restaurants around indicating there is customer volume. The chef also wants to make sure there are not many Indian restaurants in the area around. So that, there might be interest for a new cuisine.

Objective

We will study and explore the neighborhoods of San Francisco using data from Foursquare. The aim of this project is to identify ideal locations to open a new Indian restaurant based on number of restaurants around, type of cuisines, number of Indian restaurants etc.

Data

1. List of neighborhoods in San Francisco will be scraped from the following Wikipedia Page: https://en.wikipedia.org/wiki/List_of_neighborhoods_in_San_Francisco
2. Google geocoder will be used to get the latitude and longitude for the neighborhoods.
3. Foursquare API will be used to get venue data based on latitude and longitude.
4. K-Means clustering will be used identify the ideal neighborhoods for a new Indian restaurant in San Francisco.

Methodology

Once we had the venue data for neighborhoods in San Francisco, ideal neighborhoods for a new Indian restaurant were identified based on volume of existing restaurants, type of restaurant cuisines available around.

1. Download San Francisco dataset

Used Beautiful Soup library to scrape the list of neighborhoods in San Francisco area from the Wikipedia page. Then converted the data to a dataframe using pandas.

Out[11]:

	Neighborhood
0	Alamo Square
1	Anza Vista
2	Ashbury Heights
3	Balboa Park
4	Balboa Terrace

Next, google geocoder was used to retrieve latitude and longitude information for the neighborhoods.

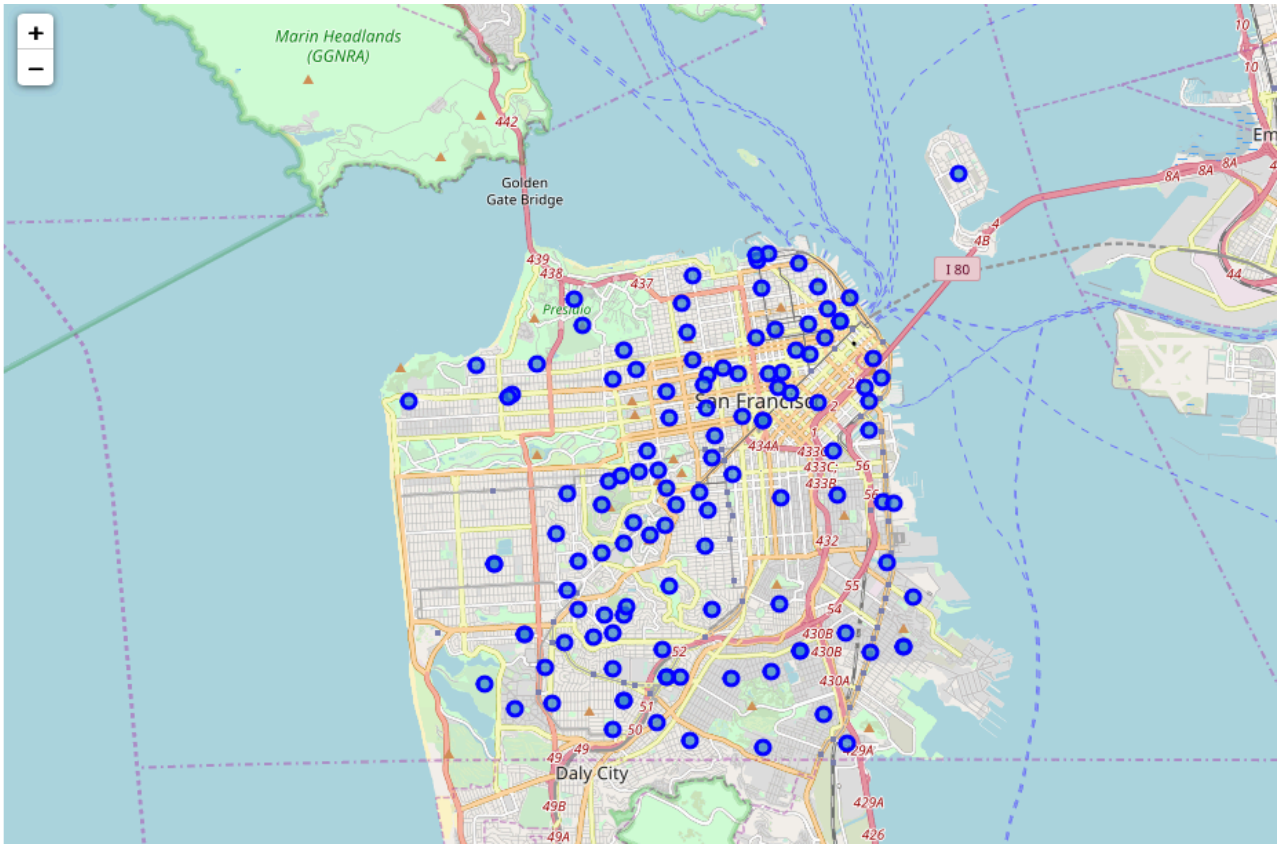
Out[12]:

	Neighborhood	Latitude	Longitude
0	Alamo Square	37.777499	-122.433252
1	Anza Vista	37.780868	-122.443185
2	Ashbury Heights	37.765327	-122.445308
3	Balboa Park	37.724569	-122.443357
4	Balboa Terrace	37.731333	-122.468661

Then, also grabbed the geographical coordinates for San Francisco.

```
The geograpical coordinate of San Francisco are 45.42226195, -75.6397071198492.
```

Once I had the all geographical coordinates for San Francisco and its neighborhoods, I plotted the map using Folium library.



2. Exploring San Francisco neighborhoods

Used Foursquare API to pull 100 venues nearest to the neighborhoods. Cleaned that and structured it into a pandas dataframe.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alamo Square	37.777499	-122.433252	Alamo Square	37.776062	-122.433622	Park
1	Alamo Square	37.777499	-122.433252	Painted Ladies	37.776120	-122.433389	Historic Site
2	Alamo Square	37.777499	-122.433252	Alamo Square Dog Park	37.775878	-122.435740	Dog Run
3	Alamo Square	37.777499	-122.433252	Originals Vinyl	37.775835	-122.431227	Record Shop
4	Alamo Square	37.777499	-122.433252	Kebab King	37.779786	-122.431589	Pakistani Restaurant

```
# Check to see how many Restaurants are there in a neighborhood
sf_venues[sf_venues['Venue Category'].str.contains('Restaurant', case = False)].groupby('Neighborhood').count().head()
```

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Alamo Square	14	14	14	14	14	14
Anza Vista	2	2	2	2	2	2
Ashbury Heights	5	5	5	5	5	5
Balboa Terrace	6	6	6	6	6	6
Belden Place	25	25	25	25	25	25

3. Analyze each San Francisco neighborhood

Looked at the top five venues in each neighborhood:

```
----Alamo Square----
          venue  freq
0           Bar  0.06
1          Café  0.05
2          Park  0.05
3         Hotel  0.03
4  Liquor Store  0.03


----Anza Vista----
                     venue  freq
0                    Café  0.18
1             Bus Station  0.09
2  Health & Beauty Service  0.09
3             Coffee Shop  0.09
4       Mexican Restaurant  0.05


----Ashbury Heights----
               venue  freq
0     Breakfast Spot  0.08
1              Trail  0.05
2        Coffee Shop  0.05
3  Convenience Store  0.05
4               Park  0.05


----Balboa Park----
            venue  freq
0  Cosmetics Shop  0.14
1       BBQ Joint  0.14
2            Park  0.14
3            Pool  0.14
4    Dessert Shop  0.14
```

Next, after exploring, created a dataset with 25 most common venues in each neighborhood.

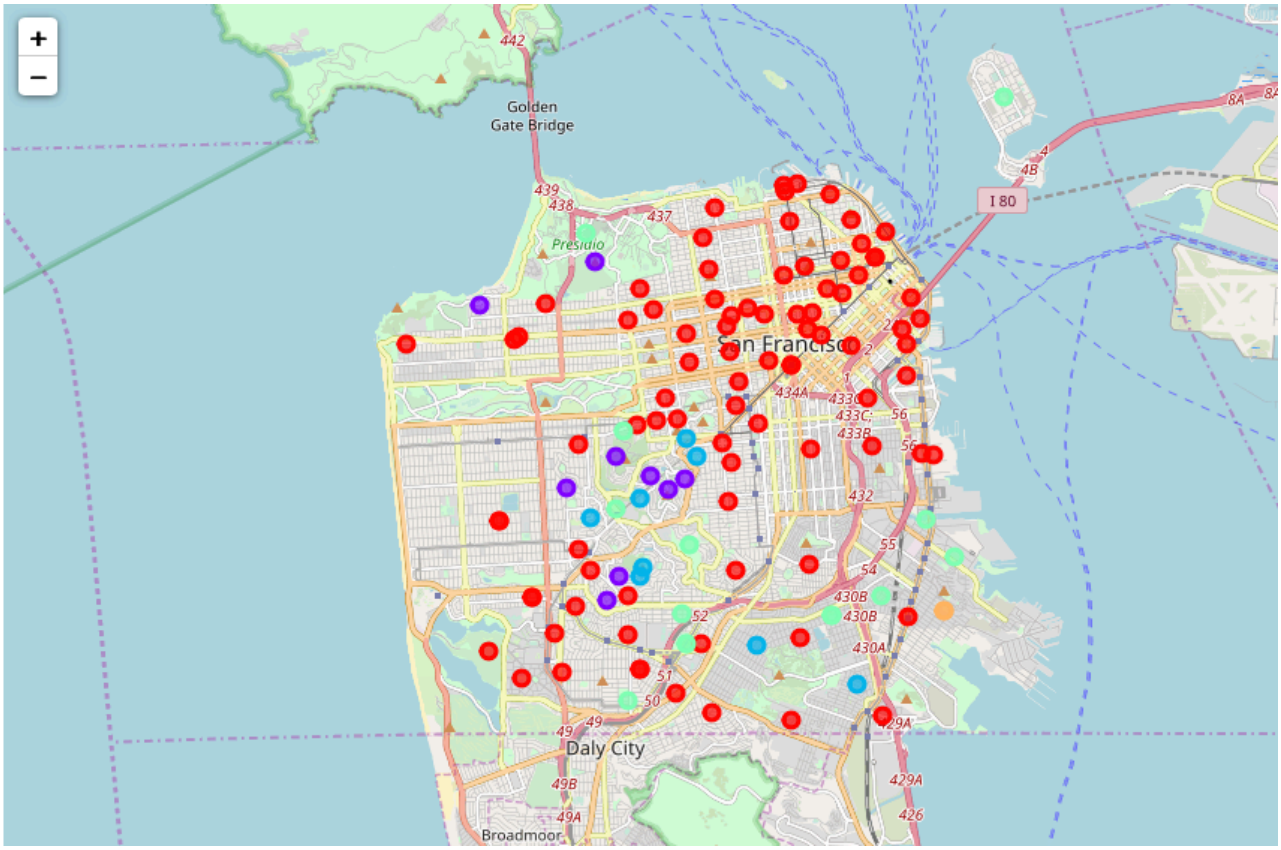
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alamo Square	Bar	Park	Café	BBQ Joint	Seafood Restaurant	Hotel	Sushi Restaurant	Pizza Place	Liquor Store	Marijuana Dispensary
1	Anza Vista	Café	Coffee Shop	Health & Beauty Service	Bus Station	Electronics Store	Tunnel	Big Box Store	Mexican Restaurant	Bus Line	Arts & Crafts Store
2	Ashbury Heights	Breakfast Spot	Convenience Store	Trail	Italian Restaurant	Park	Coffee Shop	Ice Cream Shop	Dog Run	Road	Burger Joint
3	Balboa Park	Cosmetics Shop	Park	BBQ Joint	Sandwich Place	Light Rail Station	Dessert Shop	Pool	Food & Drink Shop	Farm	Ethiopian Restaurant
4	Balboa Terrace	Japanese Restaurant	Bakery	Shoe Repair	Circus	Comic Shop	Playground	Dessert Shop	Pharmacy	Park	Optical Shop

4. Cluster San Francisco neighborhoods

After analyzing, used k-means to cluster San Francisco neighborhoods into 5 clusters.

	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alamo Square	37.777499	-122.433252	0	Bar	Park	Café	BBQ Joint	Seafood Restaurant	Hotel	Sushi Restaurant	Pizza Place	Liquor Store	Marijuana Dispensary
1	Anza Vista	37.780868	-122.443185	0	Café	Coffee Shop	Health & Beauty Service	Bus Station	Electronics Store	Tunnel	Big Box Store	Mexican Restaurant	Bus Line	Arts & Crafts Store
2	Ashbury Heights	37.765327	-122.445308	0	Breakfast Spot	Convenience Store	Trail	Italian Restaurant	Park	Coffee Shop	Ice Cream Shop	Dog Run	Road	Burger Joint
3	Balboa Park	37.724569	-122.443357	3	Cosmetics Shop	Park	BBQ Joint	Sandwich Place	Light Rail Station	Dessert Shop	Pool	Food & Drink Shop	Farm	Ethiopian Restaurant
4	Balboa Terrace	37.731333	-122.468661	0	Japanese Restaurant	Bakery	Shoe Repair	Circus	Comic Shop	Playground	Dessert Shop	Pharmacy	Park	Optical Shop

Next, tried to plot the clusters as follows:



5. Examine San Francisco clusters

After clustering, examined each cluster to identify a suitable one.

Cluster 1

```
sf_merged.loc[sf_merged['Cluster Labels'] == 0, sf_merged.columns[[1] + list(range(5, sf_merged.shape[1]))]].head()
```

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	37.777499	Park	Café	BBQ Joint	Seafood Restaurant	Hotel	Sushi Restaurant	Pizza Place	Liquor Store	Marijuana Dispensary
1	37.780868	Coffee Shop	Health & Beauty Service	Bus Station	Electronics Store	Tunnel	Big Box Store	Mexican Restaurant	Bus Line	Arts & Crafts Store
2	37.765327	Convenience Store	Trail	Italian Restaurant	Park	Coffee Shop	Ice Cream Shop	Dog Run	Road	Burger Joint
4	37.731333	Bakery	Shoe Repair	Circus	Comic Shop	Playground	Dessert Shop	Pharmacy	Park	Optical Shop
6	37.791249	French Restaurant	Gym	Bubble Tea Shop	Hotel	Gym / Fitness Center	Cocktail Bar	Café	Sandwich Place	Boutique

Result

After examining each cluster, I found cluster 4 more suitable or ideal to open an Indian restaurant as it has good number of neighborhood options in it, good number of restaurants in its most common venues – indicating there is demand for restaurants in the area. But, also at the same time, there were not many Indian restaurants in the cluster, indicating there is scope to open a new one.

```
sf_merged.loc[sf_merged['Cluster Labels'] == 3, sf_merged.columns[[1] + list(range(5, sf_merged.shape[1]))]].head(5)
```

	Latitude	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	37.724569	Park	BBQ Joint	Sandwich Place	Light Rail Station	Dessert Shop	Pool	Food & Drink Shop	Farm	Ethiopian Restaurant
22	37.742372	Video Store	Grocery Store	Salon / Barbershop	Athletics & Sports	Trail	Shopping Mall	Dim Sum Restaurant	Bus Line	Coffee Shop
40	37.740435	Food Truck	Thrift / Vintage Store	Brewery	Garden Center	Food	Non-Profit	Women's Store	Filipino Restaurant	Exhibit
45	37.746968	Light Rail Station	Moving Target	Bus Station	Theater	Wine Shop	Music Venue	Furniture / Home Store	Building	Pier
49	37.748976	Light Rail Station	Japanese Restaurant	Hotpot Restaurant	Bus Stop	Event Space	Park	Art Gallery	Lake	French Restaurant

Discussion

Some of the challenging factors that could be considered along with these results are personal preferences of the chef as to the price width of the restaurant he is looking to open. For it, the economic status of the area would also have to be considered into the analysis which was missing here. Also, the area population would also play a part.

Conclusion

So, after analyzing the data, I recommend neighborhoods in cluster 4 as suitable to open new Indian restaurant. Few neighborhoods in cluster 4 are: Balboa Park, Diamond Heights, India Basin, Islais Creek, Laguna Honda.