Capstone Project - The Battle of Neighborhoods

"Finding the best neighborhoods in Downtown San Diego to open a French Restaurant"

Report prepared by - Ganesh K

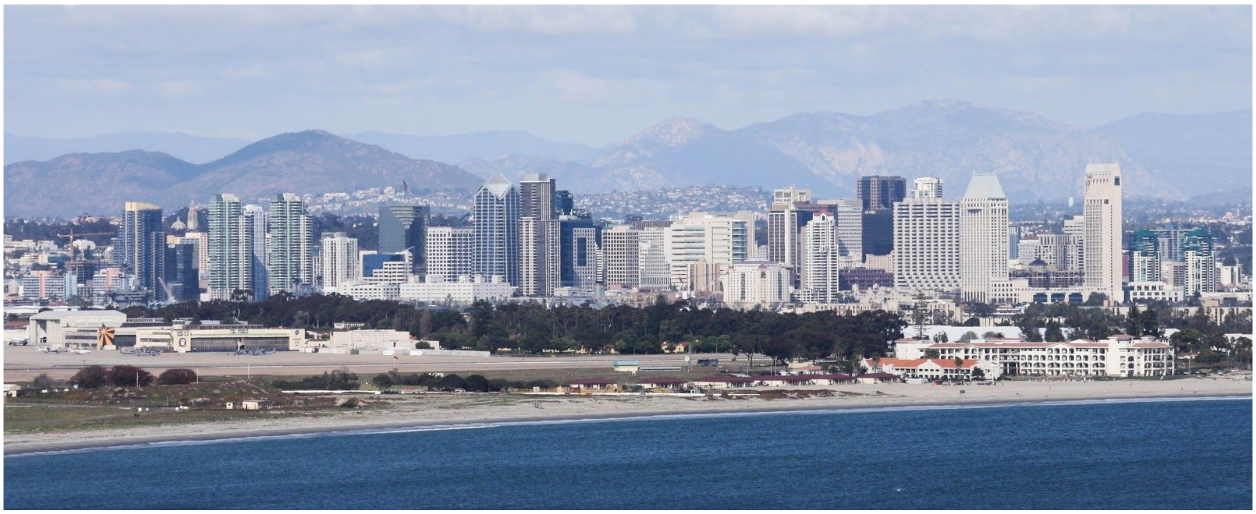


Table of Contents

1. Introduction and Business problem statement

Our client "**XYZ French food chain**" is interested in opening a **new French Restaurant** in **Downtown San Diego** neighborhood. For that they have asked us to do research of Downtown San Diego area and suggest probable locations for opening a new French restaurant.

Downtown San Diego is the city center of San Diego, California, the eighth largest city in the United States. Downtown San Diego serves as the cultural and financial center and central business district of San Diego, with more than 4,000 businesses and following neighborhoods

Columbia - the west district of downtown. Located between the Marina and Little Italy, west of Columbia Street
Civic Core - District, the central business district of downtown
Cortez Hill - the northeast district of downtown
East Village - the east district of downtown, which is home to Petco Park and the surrounding Ballpark Village
Gaslamp - Quarter, a two- by ten-block nightlife district in central Downtown
Little Italy - the northwest district of downtown
Marina - the southwest district of downtown, which is home to Seaport Village and Pantoja Park

Downtown San Diego encompasses seven thriving neighborhoods, each with its own unique identity.

Tourism is a major industry owing to the city's climate, beaches, and tourist attractions such as Balboa Park, Belmont amusement park, San Diego Zoo, San Diego Zoo Safari Park, and SeaWorld San Diego. As per report, San Diego makes Forbes' travel guide's 14 top destinations of 2019 list. A major highlight for San Diego as it pertains to the list is its rise as an up-and-coming foodie city. San Diego's thriving dining scene is a big draw for foodies who, according to Yelp, rank it the No. 5 U.S. Destination for Food Lovers. With a year-round growing season and plenty of sunshine, San Diego produces a wealth of farm fresh goods. Downtown San Diego is a diverse, vibrant community with roughly 35,000 residents, 80,000 employees who work downtown, 137,000 jobs connected to downtown businesses and millions of year-round visitors. Downtown is a regional employment center, a public transportation hotspot and the number one space for innovation firms and startup growth in the region. The largest age group in downtown is adults in their early 30s. The median age in downtown is 40 years old. Downtown is filled with high-earning professionals. Downtown residents have an average annual income of $73,756. The majority of residents dine out three or more times a week, with most residents typically visiting restaurants located downtown (80%). Residents reported spending an average of $10 to $49 per meal. Downtown San Diego is dominated by millennials and young professionals. Roughly one third (31%) of downtown's population are millennials. They live, work and play in downtown's coffee houses, restaurants, bars, lounges and alternative work spaces where ideas can be shared and creativity can be nourished over craft cocktails and custom lattes. Sixty-seven percent of residents enjoy nightlife on a regular basis. The majority (90%) frequent concerts, bars and clubs located downtown. Downtown San Diego is home to 113 Zagat rated restaurants and is quickly becoming a national foodie destination. This is yet another element of the millennial attraction to downtown and, in fact, several of downtown's restaurants have caused such a buzz, they are now expanding to other locations both inside and outside of San Diego. Downtown is a public transportation hotspot with more than 25 bus routes, four Rapid bus routes, three trolley lines and nearly 20 miles of bike paths/routes.

Hence in this project, will use the Foursquare APIs to extract the venues of Downtown San Diego neighborhoods to analyze the different neighborhoods, overall Restaurant business, in specific French Restaurant locations and probable neighborhoods where a new French restaurant can be started as specified by our client. Will use the Data Analysis and Machine Learning techniques for accomplishing the given task.

Target Audience - Our client "**XYZ French food chain**" , who is interested in opening a new French Restaurant in Downtown San Diego neighborhood, is the target audience for this project.

2. Data set

The data to be used in this project is extracted from below sources

- Foursquare data - It is a local search and discovery service which provides information on different types of entertainment, drinking and dining venues. Foursquare has an API that is used to query their database and find information related to the venues, such as location, overall category, reviews and tips. In this project will use Foursquare API to extract the venues details from Downtown San Diego neighborhoods.
  https://developer.foursquare.com/

- Downtown San Diego Neighborhood names -
  Downtown San Diego has seven neighborhoods i.e. Civic Core, Columbia, Cortez Hill, Gaslamp, Little Italy, Marina and East Village
  https://en.wikipedia.org/wiki/List_of_communities_and_neighborhoods_of_San_Diego
  https://www.luxurylivingsd.com/guide/san-diego/

- Geographic coordinates of Downtown San Diego neighborhoods -
  Following links contains the Latitude and Longitude coordinates of Downtown San Diego neighborhoods
  http://sqldbpros.com/2011/11/free-zip-code-city-county-state-csv/
  http://download.geonames.org/export/dump/
  https://www.latlong.net/convert-address-to-lat-long.html

  Here are the details of how we will use each data source in our project

2.1. Foursquare API data

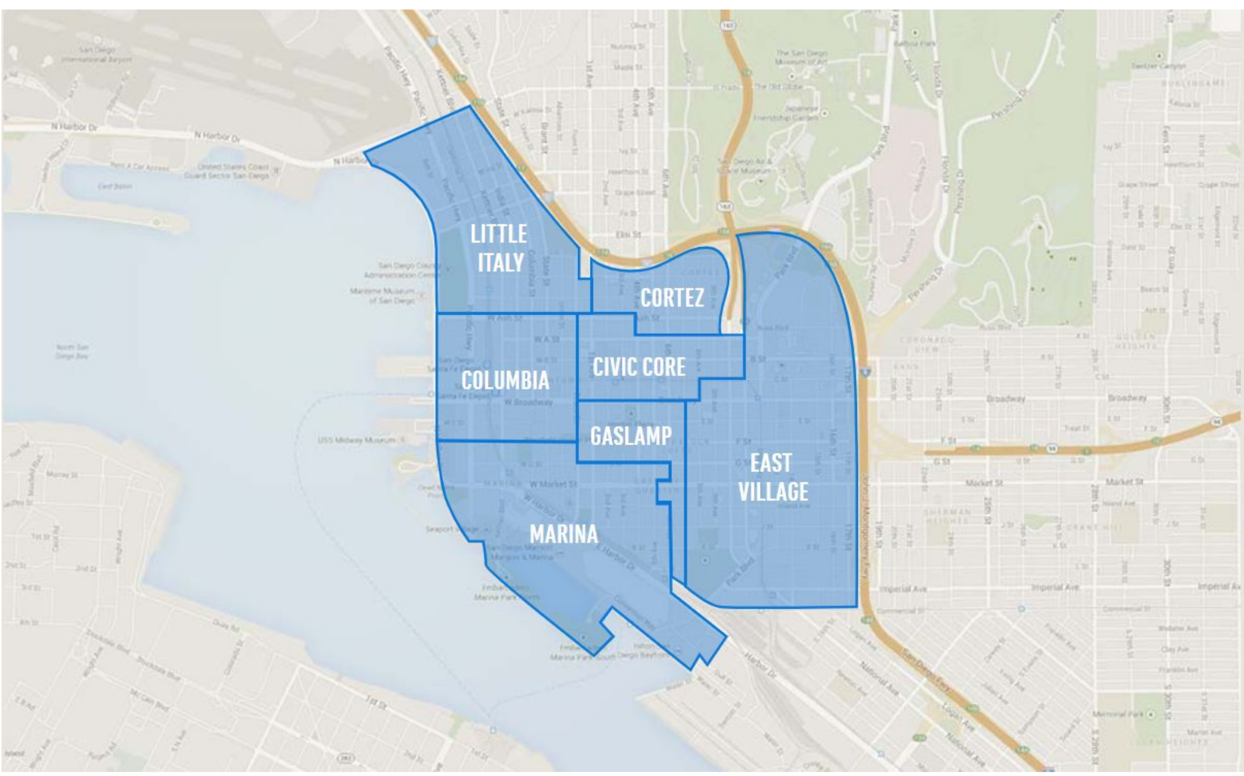
For this project we will use the Foursquare Places API. One of the features of this API is to provide a list of venues within a specific location, based on the Latitude and Longitude coordinates and a radius. In order to obtain a list of venues within a specified area, we use the "explore" endpoint from the API. By passing the proper parameters via an HTTP request to the explore endpoint, we get a JSON object with the information shown in the table below

Field	Description
id	A unique string identifier for this venue.
name	The best known name for this venue.
location	An object containing none, some, or all of `address` (street address), `crossStreet`, `city`, `state`, `postalCode`, `country`, `lat`, `lng`, and `distance`. All fields are strings, except for `lat`, `lng`, and `distance`. Distance is measured in meters. Some venues have their locations intentionally hidden for privacy reasons (such as private residences). If this is the case, the parameter `isFuzzed` will be set to true, and the `lat`/`lng` parameters will have reduced precision.
categories	An array, possibly empty, of categories that have been applied to this venue. One of the categories will have a `primary` field indicating that it is the primary category for the venue. For the complete category tree, see categories.

The location object contains the coordinates of each venue, which will be used to associate it with its respective neighborhood. The categories array will be used to categorize the neighborhood. Basically, we will count how many venues from all available categories are found on each neighborhood, and then use that information to compare neighborhoods

2.2. Downtown San Diego Neighborhood names and coordinates

Downtown San Diego has following seven neighborhoods i.e. Civic Core, Columbia, Cortez Hill, Gaslamp, Little Italy, Marina and East Village



Which are extracted from following links
https://en.wikipedia.org/wiki/List_of_communities_and_neighborhoods_of_San_Diego
https://www.luxurylivingsd.com/guide/san-diego/

Now to extract the Latitude and Longitude coordinates of these neighborhoods, will take help of following links where this info is present
http://sqldbpros.com/2011/11/free-zip-code-city-county-state-csv/
http://download.geonames.org/export/dump/
https://www.latlong.net/convert-address-to-lat-long.html

After extracting both the neighborhood names and its location coordinates, they are saved in below table(in csv format)

Borough	**Neighborhood**	**Latitude**	**Longitude**	**City**
Downtown SanDiego	CIVIC_CORE	32.7159	-117.1595	San Diego
Downtown SanDiego	COLUMBIA	32.7178	-117.1673	San Diego
Downtown SanDiego	CORTEZ HILL	32.7214	-117.1598	San Diego
Downtown SanDiego	GASLAMP	32.7101	-117.1601	San Diego
Downtown SanDiego	LITTLE ITALY	32.7234	-117.1682	San Diego
Downtown SanDiego	MARINA	32.7108	-117.1701	San Diego
Downtown SanDiego	EAST VILLAGE	32.7137	-117.1536	San Diego

So will be using the above location coordinates in Foursquare API to get the list of venues in each of the Downtown San Diego neighborhoods.

3. Methodology

3.1 Data collection and preparation

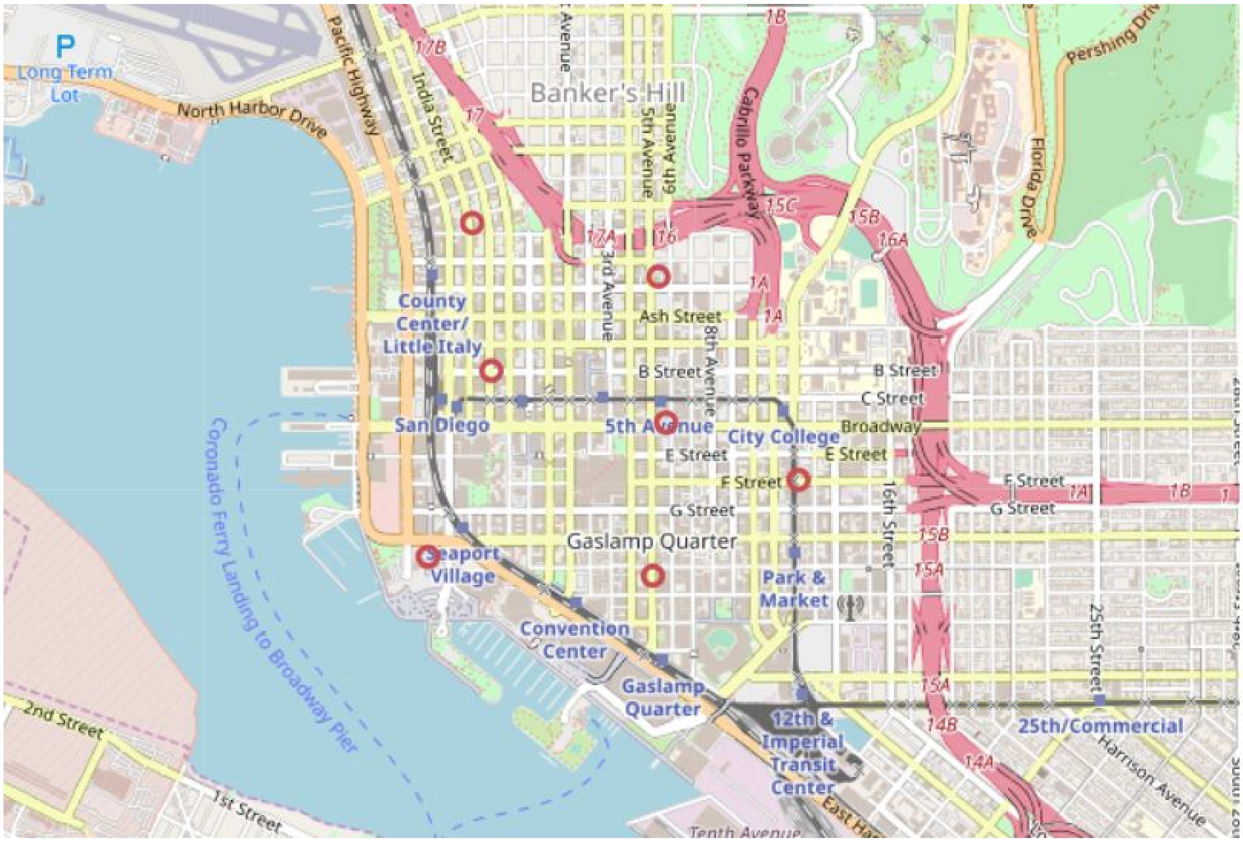
Using the weblinks mentioned in section 2, we extracted the seven Neighborhoods names of Downtown San Diego. Then extracted the corresponding location coordinates i.e. Latitude and Longitude for each of the neighborhood location. This information was saved in a csv file and in our project we read this file and the location info into Pandas dataframe.

```
[2]: #Read the preprocessed csv file which has Downtown San Diego neighborhood names and its location info
df = pd.read_csv('SD1_LatLong_Dataset.csv')
df
```

[2]:

	Borough	Neighborhood	Latitude	Longitude	City
0	Downtown SanDiego	CIVIC_CORE	32.7159	-117.1595	San Diego
1	Downtown SanDiego	COLUMBIA	32.7178	-117.1673	San Diego
2	Downtown SanDiego	CORTEZ HILL	32.7214	-117.1598	San Diego
3	Downtown SanDiego	GASLAMP	32.7101	-117.1601	San Diego
4	Downtown SanDiego	LITTLE ITALY	32.7234	-117.1682	San Diego
5	Downtown SanDiego	MARINA	32.7108	-117.1701	San Diego
6	Downtown SanDiego	EAST VILLAGE	32.7137	-117.1536	San Diego

Now that we have the Neighborhood names and its coordinates, lets draw the map of Downtown San Diego with neighborhoods superimposed using folium



Now using Foursquare APIs lets extract the venues info for all the neighborhoods

```
[27]: #Get venues for all neighborhoods in our dataset
sd_venues = getNearbyVenues(names=df['Neighborhood'],
                                   latitudes=df['Latitude'],
                                   longitudes=df['Longitude'])
```

```
Extracting the venues from Neighborhood -  CIVIC_CORE
Extracting the venues from Neighborhood -  COLUMBIA
Extracting the venues from Neighborhood -  CORTEZ HILL
Extracting the venues from Neighborhood -  GASLAMP
Extracting the venues from Neighborhood -  LITTLE ITALY
Extracting the venues from Neighborhood -  MARINA
Extracting the venues from Neighborhood -  EAST VILLAGE
```

Now we are ready with data.

3.2 Exploratory Data Analysis(EDA)

Lets check the dimension(num of rows and columns) of the generated dataframe, names of columns and the data types stored in each column

[29]:
```
#check the paramters(columns) in generated dataframe
sd_venues.head()
```

[29]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CIVIC_CORE	32.7159	-117.1595	House of Blues San Diego	32.716370	-117.159528	Music Venue
1	CIVIC_CORE	32.7159	-117.1595	THE US GRANT, A Luxury Collection Hotel, San D...	32.716149	-117.161527	Hotel
2	CIVIC_CORE	32.7159	-117.1595	Vin de Syrah Wine Parlor	32.714723	-117.160040	Wine Bar
3	CIVIC_CORE	32.7159	-117.1595	The Taco Stand Downtown	32.717749	-117.158497	Taco Place
4	CIVIC_CORE	32.7159	-117.1595	Donut Bar	32.717619	-117.158757	Donut Shop

[75]:
```
#check the names of columns in dataframe and its type
sd_venues.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 699
Data columns (total 7 columns):
Neighborhood              700 non-null object
Neighborhood Latitude     700 non-null float64
Neighborhood Longitude    700 non-null float64
Venue                     700 non-null object
Venue Latitude            700 non-null float64
Venue Longitude           700 non-null float64
Venue Category            700 non-null object
dtypes: float64(4), object(3)
memory usage: 38.4+ KB
```

Also lets check the number of venues per neighborhood

[77]:
```
#check the number of venues per neighbourhood
sd_venues['Neighborhood'].value_counts()
```

[77]:
```
LITTLE ITALY    100
MARINA          100
EAST VILLAGE    100
CIVIC_CORE      100
GASLAMP         100
COLUMBIA        100
CORTEZ HILL     100
Name: Neighborhood, dtype: int64
```

[78]:
```
#Number of venues per neighborhood
sd_venues.groupby('Neighborhood').count()
```

[78]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
CIVIC_CORE	100	100	100	100	100	100
COLUMBIA	100	100	100	100	100	100
CORTEZ HILL	100	100	100	100	100	100
EAST VILLAGE	100	100	100	100	100	100
GASLAMP	100	100	100	100	100	100
LITTLE ITALY	100	100	100	100	100	100
MARINA	100	100	100	100	100	100

Lets check number of unique venue categories and number of venues per category

```
[79]: #Number of unique venue categories
      print('There are {} uniques categories.'.format(len(sd_venues['Venue Category'].unique())))

      There are 129 uniques categories.
```

```
[80]: #print out the list of categories
      sd_venues['Venue Category'].unique()[:150]
```

```
[80]: array(['Music Venue', 'Hotel', 'Wine Bar', 'Taco Place', 'Donut Shop',
             'Beer Store', 'Lounge', 'Sandwich Place', 'American Restaurant',
             'Theater', 'Seafood Restaurant', 'Brazilian Restaurant', 'Bar',
             'Fondue Restaurant', 'Theme Restaurant', 'New American Restaurant',
             'Coffee Shop', 'Italian Restaurant', 'Hot Dog Joint', 'Café',
             'Concert Hall', 'Liquor Store', 'Mexican Restaurant',
             'Chocolate Shop', 'Mediterranean Restaurant', 'Burger Joint',
             'Brewery', 'Breakfast Spot', 'Ramen Restaurant', 'Comedy Club',
             'French Restaurant', 'Jewelry Store', 'Speakeasy',
             'Mobile Phone Shop', 'Sushi Restaurant', 'Hookah Bar',
             'Vietnamese Restaurant', 'Pizza Place', 'Empanada Restaurant',
             'Lingerie Store', 'Health Food Store', 'Business Service',
             'Nightclub', 'Piano Bar', 'Arcade', 'Turkish Restaurant',
             'Cocktail Bar', 'Steakhouse', 'Pub', 'Bakery',
             'Russian Restaurant', 'Massage Studio', 'Dessert Shop',
             'Accessories Store', 'Irish Pub', 'Gastropub', 'Gym',
             'Falafel Restaurant', 'Tiki Bar', 'Farmers Market',
             'Ice Cream Shop', 'Convenience Store', 'Park', 'Hostel',
             'Neighborhood', 'Juice Bar', 'Pilates Studio', 'Pier', 'BBQ Joint',
             'Museum', 'Boat or Ferry', 'Roof Deck', 'Arts & Crafts Store',
             'Plaza', 'Vegetarian / Vegan Restaurant', 'History Museum',
             'Event Space', 'Tour Provider', 'South American Restaurant',
             'Middle Eastern Restaurant', 'Spa', 'Dog Run', 'Whisky Bar',
```

[81]:

```python
#lets check num of venues per category
sd_venues['Venue Category'].value_counts()
```

[81]:

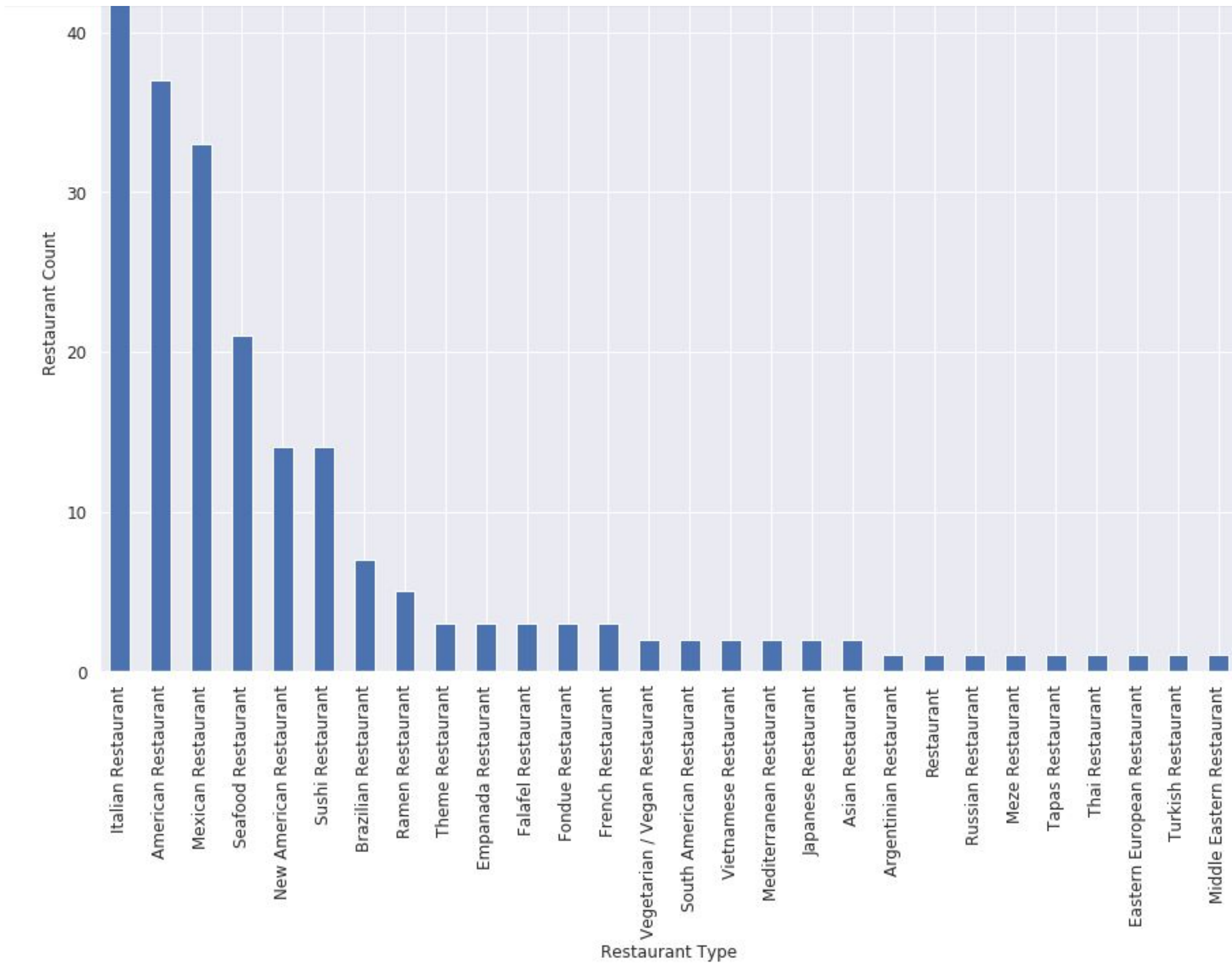
```
Hotel                        58
Italian Restaurant           44
American Restaurant          37
Mexican Restaurant           33
Bar                          29
Coffee Shop                  27
Café                         22
Seafood Restaurant           21
Steakhouse                   16
New American Restaurant      15
Pizza Place                  14
Brewery                      13
Sushi Restaurant             13
Wine Bar                     13
Park                         13
Burger Joint                 11
Breakfast Spot               11
Taco Place                   11
Theater                      10
Lounge                       10
Gastropub                     9
Gym                           8
Music Venue                   8
Dessert Shop                  8
Museum                        8
Ice Cream Shop                8
Brazilian Restaurant          7
Convenience Store             7
Sandwich Place                6
Pub                           5
```

Lets check the restaurants types and counts

[39]:

	Venue Category	Counts
9	Italian Restaurant	44
0	American Restaurant	37
12	Mexican Restaurant	33
19	Seafood Restaurant	21
15	New American Restaurant	14
21	Sushi Restaurant	14
3	Brazilian Restaurant	7
16	Ramen Restaurant	5
24	Theme Restaurant	3
5	Empanada Restaurant	3
6	Falafel Restaurant	3
7	Fondue Restaurant	3
8	French Restaurant	3
26	Vegetarian / Vegan Restaurant	2
20	South American Restaurant	2
27	Vietnamese Restaurant	2
11	Mediterranean Restaurant	2
10	Japanese Restaurant	2

Lets plot the bar chart of types of Restaurants and its count



Now lets see the French restaurants in neighborhoods

Now lets analyze Each Neighborhood specifically for French Restaurants

[83]:
```
# check if the results contain "French Restaurants"
"French Restaurant" in sd_venues['Venue Category'].unique()
```

[83]: True

[84]:
```
#lets get the details of French Restaurant in neighborhoods
sd_venues[sd_venues['Venue Category'] == 'French Restaurant']
```

[84]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
43	CIVIC_CORE	32.7159	-117.1595	Le Fontainbleau - The Westgate Hotel	32.716629	-117.162721	French Restaurant
132	COLUMBIA	32.7178	-117.1673	Le Fontainbleau - The Westgate Hotel	32.716629	-117.162721	French Restaurant
232	CORTEZ HILL	32.7214	-117.1598	Le Fontainbleau - The Westgate Hotel	32.716629	-117.162721	French Restaurant

So there are three French Restaurant in neighborhoods CIVIC_CORE, COLUMBIA and CORTEZ HILL

Lets see top 5 venues in each neighborhood

```
[89]: #Let's print each neighborhood along with the top 5 most common venues
num_top_venues = 5

for hood in to_grouped['Neighborhoods']:
    print("----"+hood+"----")
    temp = to_grouped[to_grouped['Neighborhoods'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----CIVIC_CORE----
                venue  freq
0  Mexican Restaurant  0.08
1               Hotel  0.07
2 American Restaurant  0.06
3                Café  0.05
4                 Bar  0.05


----COLUMBIA----
                venue  freq
0  Italian Restaurant  0.11
1               Hotel  0.08
2 American Restaurant  0.06
3         Coffee Shop  0.05
4            Wine Bar  0.04
```

Top 10 venues in each neighborhood

```
neighborhoods_venues_sorted
```

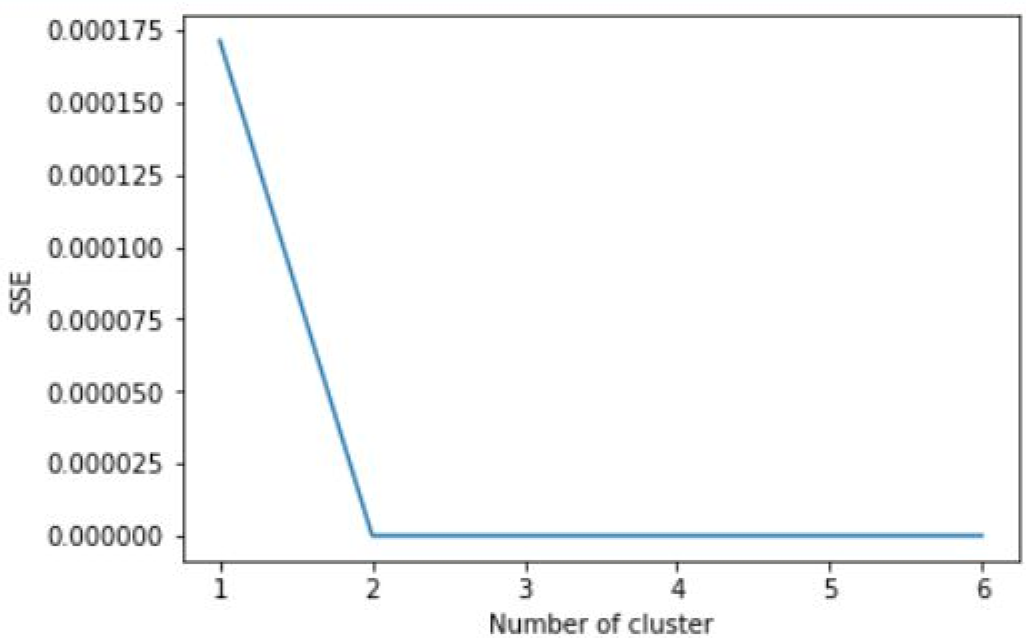
[95]:

	Neighborhoods	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	CIVIC_CORE	Mexican Restaurant	Hotel	American Restaurant	Café	Bar	Italian Restaurant	Sushi Restaurant	Seafood Restaurant	Coffee Shop	Ramen Restaurant
1	COLUMBIA	Italian Restaurant	Hotel	American Restaurant	Coffee Shop	Wine Bar	Park	Mexican Restaurant	Museum	Steakhouse	Seafood Restaurant
2	CORTEZ HILL	Hotel	American Restaurant	Italian Restaurant	Coffee Shop	Mexican Restaurant	Bar	New American Restaurant	Café	Pizza Place	Taco Place
3	EAST VILLAGE	Hotel	Mexican Restaurant	American Restaurant	Café	Bar	Italian Restaurant	Lounge	Gastropub	Seafood Restaurant	Burger Joint
4	GASLAMP	Hotel	Bar	Steakhouse	Mexican Restaurant	Café	American Restaurant	Italian Restaurant	Lounge	Seafood Restaurant	Breakfast Spot
5	LITTLE ITALY	Italian Restaurant	Coffee Shop	American Restaurant	Hotel	Wine Bar	Pizza Place	Convenience Store	New American Restaurant	Brewery	Japanese Restaurant
6	MARINA	Hotel	Seafood Restaurant	Coffee Shop	American Restaurant	Mexican Restaurant	Park	New American Restaurant	Steakhouse	Gift Shop	Italian Restaurant

3.3 Machine Learning algorithm for clustering

Now lets use the machine learning algorithm to generate the clusters of neighborhoods which may have the French restaurants. For this will use the K Means clustering algorithm of machine learning. First we need to get the right value of K

```
""""
```




```
As seen in above graph the k value to be used is 2
```

After applying the K Means clustering algorithm with K =2 will get the following result

```
[101]: # create a new dataframe that includes the cluster as well as the top 10 v
to_merged = to_french.copy()

# add clustering labels
to_merged["Cluster Labels"] = kmeans.labels_
```

```
[102]: to_merged.rename(columns={"Neighborhoods": "Neighborhood"}, inplace=True)
to_merged
```

[102]:

	Neighborhood	French Restaurant	Cluster	Cluster Labels
0	CIVIC_CORE	0.01	1	1
1	COLUMBIA	0.01	1	1
2	CORTEZ HILL	0.01	1	1
3	EAST VILLAGE	0.00	0	0
4	GASLAMP	0.00	0	0
5	LITTLE ITALY	0.00	0	0
6	MARINA	0.00	0	0

From above table its clear that Cluster Label 1 contains the Downtown San Diego neighborhoods of CIVIC_CORE, COLUMBIA and CORTEZ HILL which have at least one French Restaurant in it. Whereas Cluster Label 0 contains the Downtown San Diego neighborhoods of MARINA, GASLAMP, LITTLE ITALY and EAST VILLAGE which has no French Restaurant.

Generated cluster info

```
[108]: #get the neighborhood names per cluster
to_merged.groupby('Cluster Labels')['Neighborhood'].unique()
```

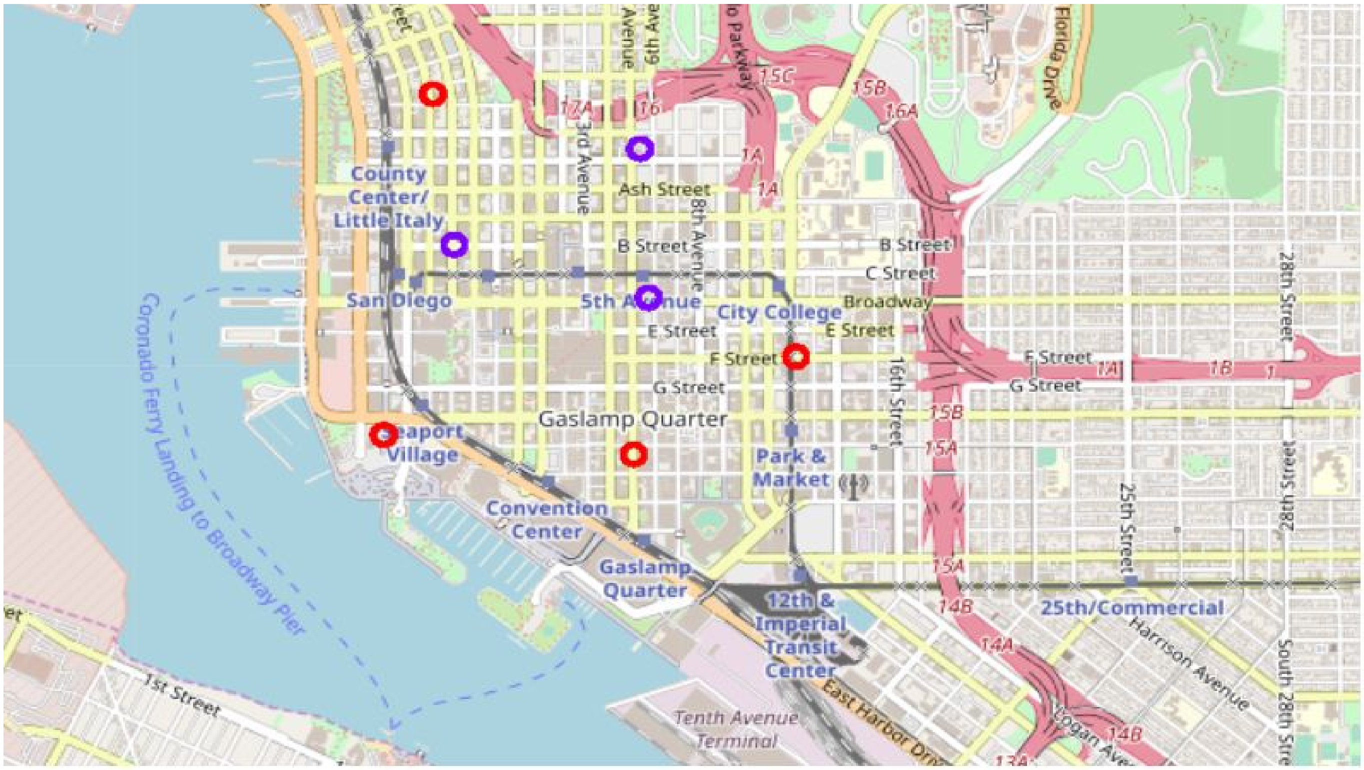
```
[108]: Cluster Labels
0    [EAST VILLAGE, GASLAMP, LITTLE ITALY, MARINA]
1               [CIVIC_CORE, COLUMBIA, CORTEZ HILL]
Name: Neighborhood, dtype: object
```

```
[111]: #lets check the French Restaurant details with cluster info
to_merged[to_merged['Venue Category'] == 'French Restaurant']
```

[111]:

	Neighborhood	French Restaurant	Cluster	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	COLUMBIA	0.01	1	1	32.7178	-117.1673	Le Fontainbleau - The Westgate Hotel	32.716629	-117.162721	French Restaurant
0	CIVIC_CORE	0.01	1	1	32.7159	-117.1595	Le Fontainbleau - The Westgate Hotel	32.716629	-117.162721	French Restaurant
2	CORTEZ HILL	0.01	1	1	32.7214	-117.1598	Le Fontainbleau - The Westgate Hotel	32.716629	-117.162721	French Restaurant

Now lets visualize the generated cluster neighborhoods using folium map

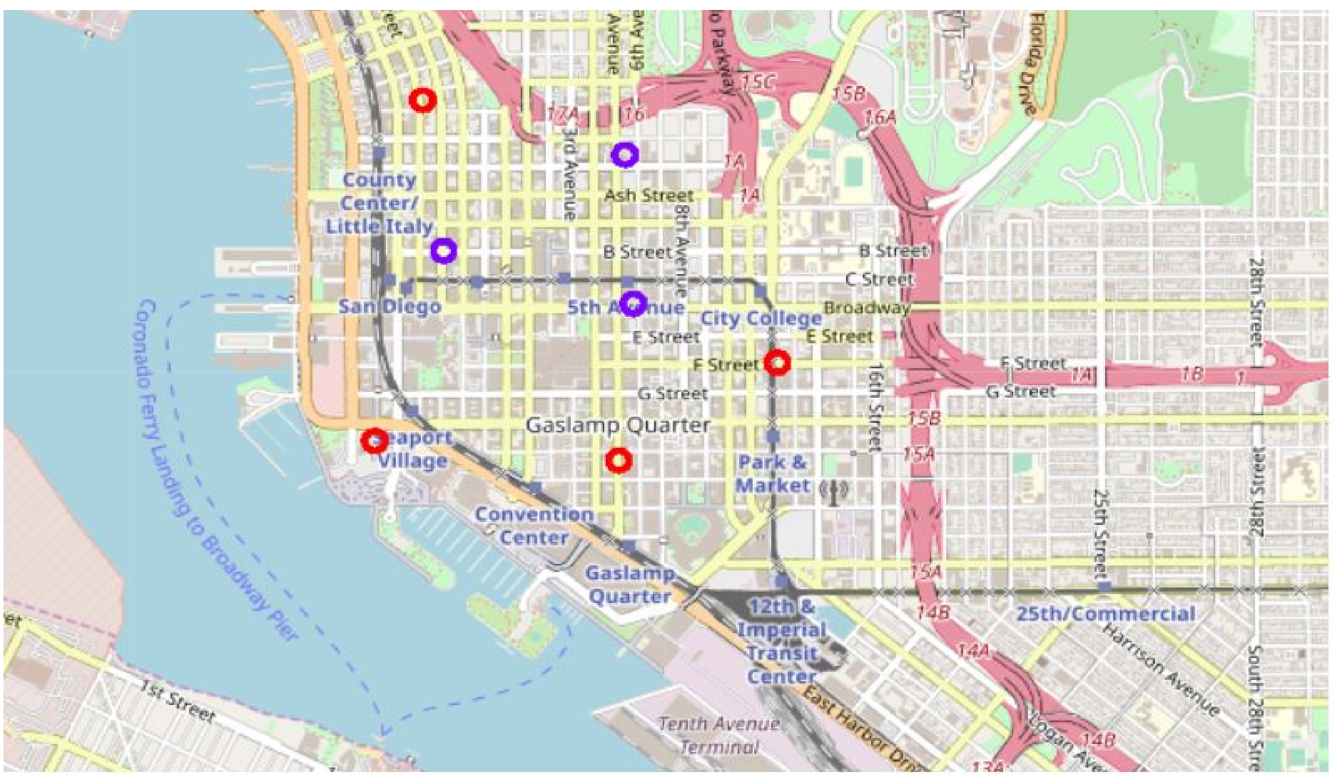


4. Results

As per our preprocessed data set there are seven Neighborhoods in Downtown San Diego. Using Foursquare API, we extracted venues in all the neighborhoods. We got total of 700 venues in seven neighborhoods and there are 129 unique venue categories in generated data. There are 28 different types of Restaurants categories and total number of Restaurant venues are 211.

When we further explored the Restaurants data specifically for French Restaurants, we found that there are three French Restaurants in three different neighborhoods i.e. CIVIC_CORE, COLUMBIA and CORTEZ HILL. K Means clustering algorithm further confirms this fact with below clustering details

- **Cluster 1** - contains the Downtown San Diego neighborhoods of **CIVIC_CORE, COLUMBIA and CORTEZ HILL** which have at least one French Restaurant
- **Cluster 0** - contains the Downtown San Diego neighborhoods of **MARINA, GASLAMP, LITTLE ITALY and EAST VILLAGE** which has no French Restaurant



5. Discussion

While working on this project I faced following challenges and my observations

- I was looking for a common database of all the major cities in the world with its Borough/Districts and Neighborhoods/localities with zip code and geocodes(Latitude and Longitude). I observed that this info is not freely available and getting the most accurate geocodes at neighborhood level for all venues is a challenge
- Foursquare API is good and contains rich venue data for US and other few countries but when I tried looking for India data, I found that its not complete
- In current project I limited the scope with Foursquare APIs data for Downtown San Diego Neighborhood selection. But along with venues data, even real estate pricing, cost of living, population, earning capabilities/opportunities, public transport, commute time to work, pollution levels, crime rate/safety, local government policies etc can be considered to arrive at proper neighborhood choices

6. Conclusion

In this project the given problem statement was, to identify the **Neighborhoods in Downtown San Diego** to open a **new French Restaurant** as enquired by our client **"XYZ French food chain**". We started with data collection of Downtown San Diego Neighborhood names and its geocodes. Then using Foursquare API we extracted the list of venues in each of the neighborhoods. Then we performed the exploratory data analysis to check the type of Restaurant venues in different neighborhoods. After that we performed the EDA, especially for French Restaurants in all the neighborhoods. Then we used the machine learning technique K Means clustering algorithm to create clusters of neighborhoods with availability of French Restaurants. The results of EDA and clustering algorithm shows that there are two distinct clusters in Downtown San Diego i.e.

- **Cluster 1** - contains the Downtown San Diego neighborhoods of **CIVIC_CORE, COLUMBIA and CORTEZ HILL** which have at least one French Restaurant
- **Cluster 0** - contains the Downtown San Diego neighborhoods of **MARINA, GASLAMP, LITTLE ITALY and EAST VILLAGE** which has no French Restaurant

Hence we would like to recommend our client **"XYZ French food chain**" to check the possibility of opening a new French Restaurant in **Cluster 0** i.e. **Downtown San Diego neighborhoods of MARINA, GASLAMP, LITTLE ITALY or EAST VILLAGE** which has no French Restaurant as per Foursquare venue data. The final call of opening a new French Restaurant at above mentioned neighborhoods rests with our client senior management, based on various business considerations.

7. References

Here are the list of references used to prepare this report

1. Foursquare API - https://developer.foursquare.com/
2. Downtown San Diego Neighborhood names - https://en.wikipedia.org/wiki/List_of_communities_and_neighborhoods_of_San_Diego
3. https://www.luxurylivingsd.com/guide/san-diego/
4. https://downtownsandiego.org/neighborhoods/
5. Geographic coordinates of Downtown San Diego neighborhoods - http://sqldbpros.com/2011/11/free-zip-code-city-county-state-csv/
6. http://download.geonames.org/export/dump/
7. https://www.latlong.net/convert-address-to-lat-long.html
8. Ref report - http://www.zinkohlaing.com/data-science/using-machine-learning-to-find-locations-to-open-a-burmese-restaurant-in-toronto-ibm-capstone-project/
9. Google maps - https://www.google.com/maps
10. My GIT hub - https://github.com/kamatgc/Coursera_Capstone