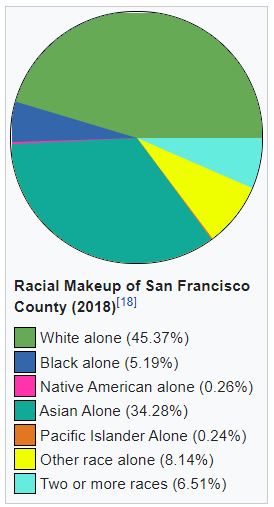
**New Indian restaurant in San Francisco Bay Area**

1. **Business Understanding and Goals**
   1. **Introduction**

The **San Francisco Bay Area**, popularly referred to as the **Bay Area** or simply the **Bay**, is a populous region surrounding the [San Francisco](https://en.wikipedia.org/wiki/San_Francisco_Bay), [San Pablo](https://en.wikipedia.org/wiki/San_Pablo_Bay), and [Suisun Bay](https://en.wikipedia.org/wiki/Suisun_Bay) [estuaries](https://en.wikipedia.org/wiki/Estuary) in [Northern California](https://en.wikipedia.org/wiki/Northern_California).  [Northern California](https://en.wikipedia.org/wiki/Northern_California)'s nine-county Bay Area contains many cities, towns, airports, and associated regional, state, and [national parks](https://en.wikipedia.org/wiki/National_park), connected by a complex multimodal transportation network It is a cultural and financial center of the western United States. It has a diverse population with 2.7% of Indian Americans among 34.28% of Asian population in San Francisco city alone. Annually it attracts about 25 million visitors. This diversity and tourism provide opportunity to experience different culture and food. These statistics provides an opportunity to explore possible neighborhood in San Francisco Bay area to open an Indian restaurant.

* 1. **Goal**Although there are lot of opportunity exists, competitions are also there. In Bay area, many Indian restaurants are already in battle. To open a new restaurant is very challenging. There are many features to be considered before opening an Indian restaurant like location, indoor, outdoor dining, delivery, transport options etc. The location of the restaurant plays an important role. It should be in a busy locality as well as less competitive. The goal of this project is to identify the suitable locations to open an Indian restaurant in San Francisco Bay Area Neighborhoods
  2. **Target Audiences**

The targeted audiences for this report are people looking to open a new Indian restaurant and anyone looking to expand their Indian restaurant in San Francisco Bay Area neighborhood.

1. **Analytic Approach**

* Collect the San Francisco bay area neighborhood data.
* Collect its neighborhood geolocation information.
* Using Foursquare API we will get all the nearby venues for each neighborhood.
* Filter out the Indian Restaurants venue for each neighborhood.
* Visualize the Data in map using folium library.
* Calculate the mean of all venue groups by their neighborhoods.
* Clustering using K-means Neighborhoods.
* Examine Clusters to find the suitable neighborhood to start an Indian Restaurant.
* Discuss about the results and conclude.

1. **Data Requirements**

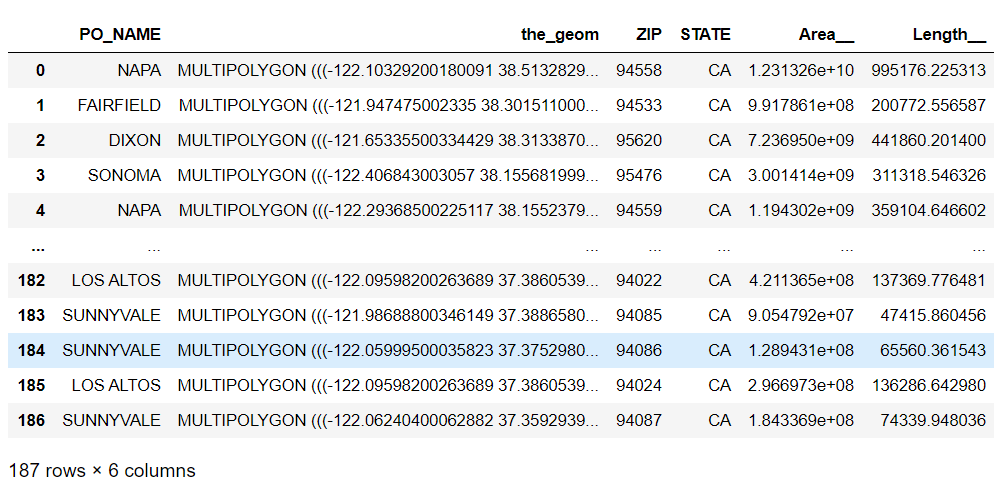
Thedata that will be required to analysis the neighborhood of San Francisco Bay Area to open an Indian restaurant are

* List of neighborhoods in San Francisco Bay Area
* Geographical location of the neighborhoods
* List of location of Indian restaurants in these neighborhoods

1. **Data Collection**

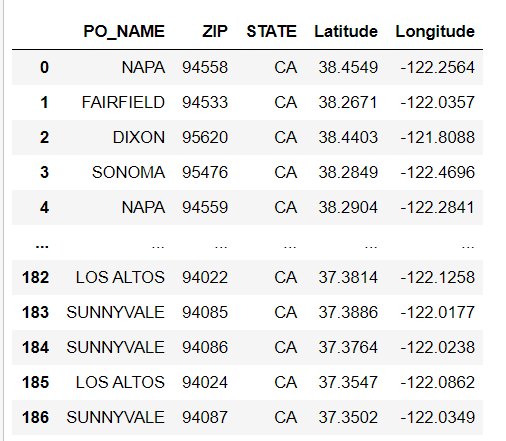
List of neighborhoods of San Francisco Bay Area neighborhood can be downloaded from

<https://data.sfgov.org/api/views/f9wk-m4qb/rows.csv?accessType=DOWNLOAD>



The downloaded information from above link provides the postal code, location name, state information. When accessing the link, the data is downloaded as CSV and parse through pandas library in python to derive the required information.

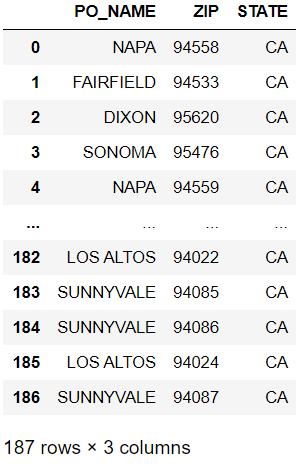
For each of the neighborhood from the previous CSV file, the corresponding geolocation is derived using pgeocode python library by passing their zip code.



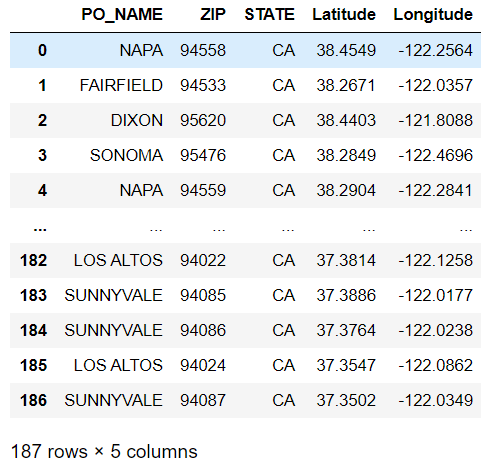
1. **Data Preparation**

The data from the required sources are collected and loaded into pandas dataframes. Then data is cleansed and prepared for clustering and segmenting.

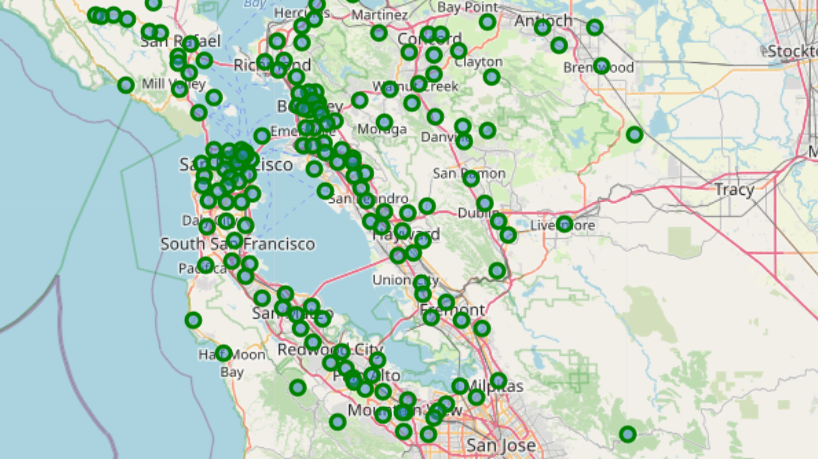
In the list of neighborhoods dataframe, the unwanted columns, the geonm, area and length, are dropped. Same POName has many zip codes.

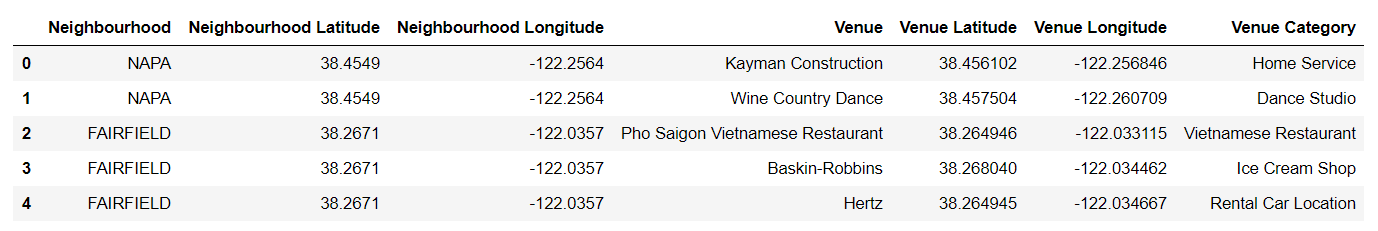


Using pgeocode, the latitude and longitude of the bay area zip codes are found. The collected geo information is merged with the bayareacode dataframe.



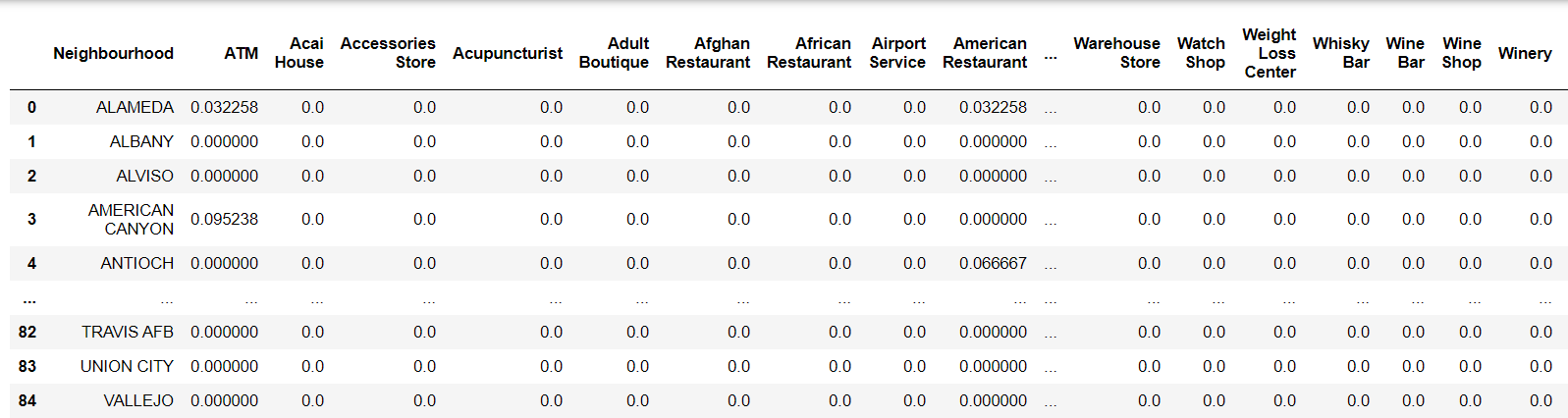
Using foursquare API, 100 venues are gathered for each zipcode’s geolocation, within the radius of 3000 meters. There are 359 unique venue categories from this data set, across 87 bay area neighborhoods.



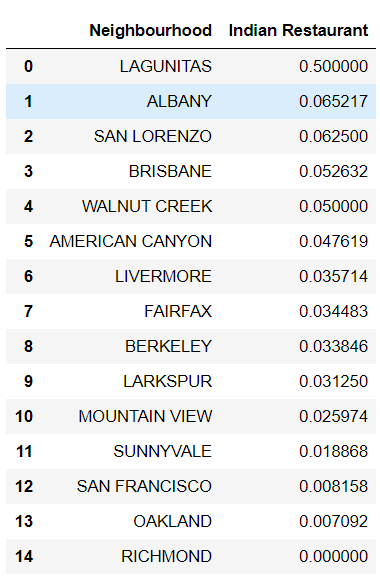


1. **Modeling**

The mean value for the venue category for each neighborhood is calculated.



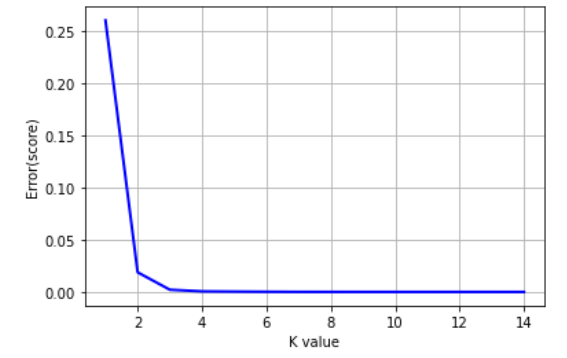
The Indian Restaurants, venue category is alone are filtered out.



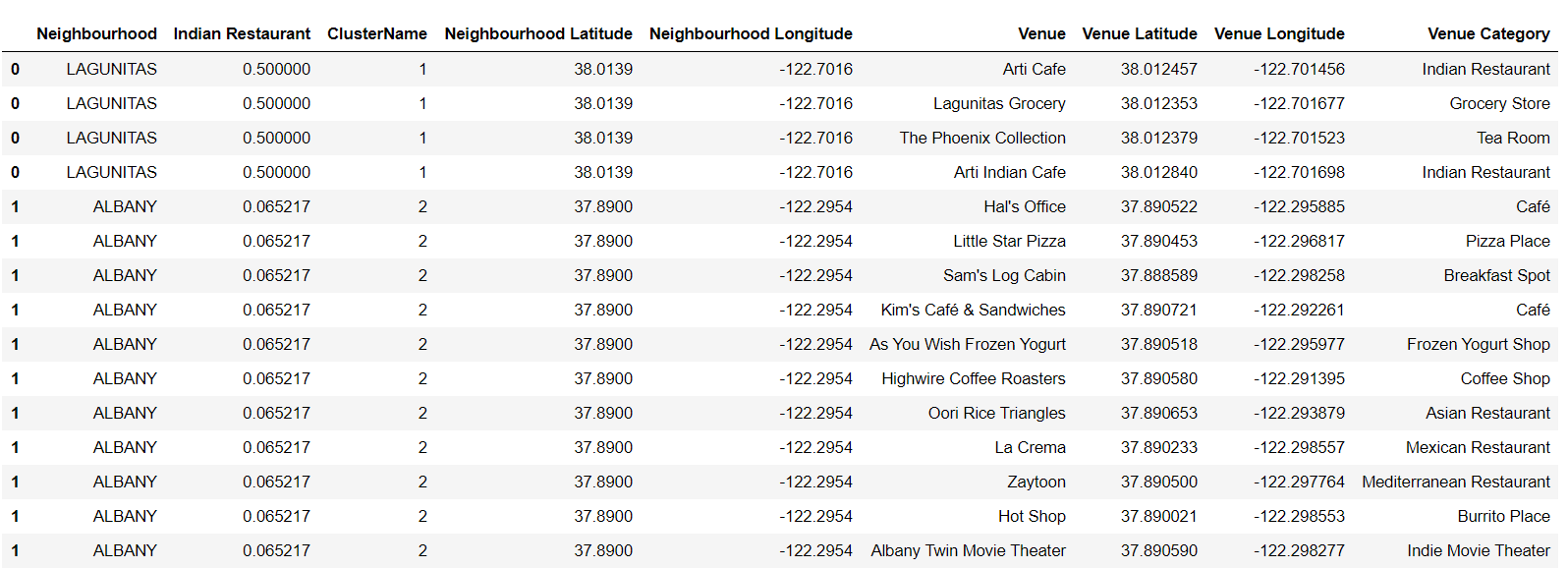
* 1. **K-Means clustering**

**K**-**means clustering** is one of the simplest and popular unsupervised machine learning algorithms. It aims to partition *n* observations into **k clusters** in which each observation belongs to the **cluster** with the nearest **mean.** We use Kmeans clustering to cluster the neighborhoods and find the best place for opening the Indian restaurant.

To find the best K value, 15 iterations are done and the error value are jotted in the graph as below. The best value for k is 3 based on this graph.

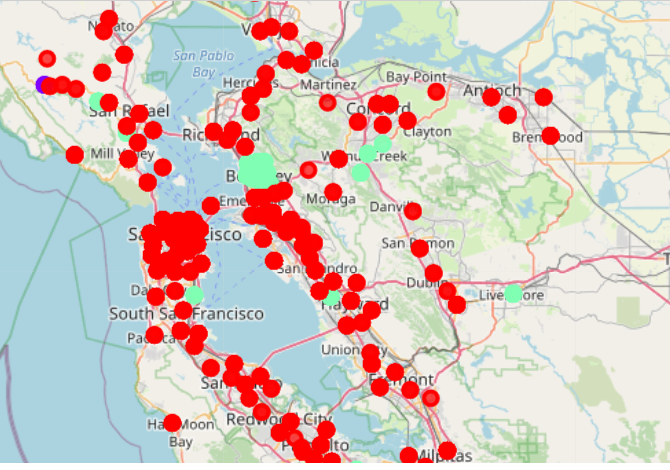


The 3 cluster labels are assigned to the neighborhood and its corresponding Indian restaurants.



1. **Evaluation and Conclusion**

Totally there are 40 Indian restaurants in 87 neighborhoods of bayarea. These 87 neighborhoods are divided into 3 clusters.



Cluster 0 has 76 neighborhoods with 15 Indian restaurants, which makes an average 0.197

Cluster 1 has 1 neighborhood with 2 Indian restaurants, which makes average of 2.

Cluster 2 has 10 neighborhoods with 23 Indian restaurants, which makes average 2.3

So, **cluster 0** has the least average Indian restaurant and hence it is the best neighborhood to open an Indian restaurant. This analysis considered only the foursquare data for top 100 venues. Although to open a restaurant, we need to consider the real estate cost, crime rate etc., these were not considered in the scope of this project.

1. **References**

[**https://en.wikipedia.org/wiki/San\_Francisco\_Bay\_Area**](https://en.wikipedia.org/wiki/San_Francisco_Bay_Area)

[**https://en.wikipedia.org/wiki/Demographics\_of\_San\_Francisco**](https://en.wikipedia.org/wiki/Demographics_of_San_Francisco)

[**https://en.wikipedia.org/wiki/K-means\_clustering**](https://en.wikipedia.org/wiki/K-means_clustering)