**New Chinese Restaurant in SF Area**

Zuhair Bennasser

March 17, 2020

**1. Introduction**

**1.1 Background**

Opening a restaurant in San Francisco looks like a good investment, the city is busy and a hub for big tech companies. However, this investment is not risk free; NBC BAY AREA reports that in 2019, more than 400 Restaurants closed. Therefore, it is important to do research before opening a new restaurant; Areas of research are for example: type of food, amenities and location.

**1.2 Problem**

The type of new restaurant for this project is Chinese, it is a good approach to research the area for existing Chinese restaurants, to avoid opening near them.

In addition, to research the existing restaurants to compare their rating to the amenities they offer, to find out what is the most needed for the new restaurant.

**1.3 Interest**

This Approach will help to avoid overpaying for amenities that are not essential for a customer satisfaction, and to focus and pay for the amenities that the customer really need.

**2. Data acquisition and cleaning**

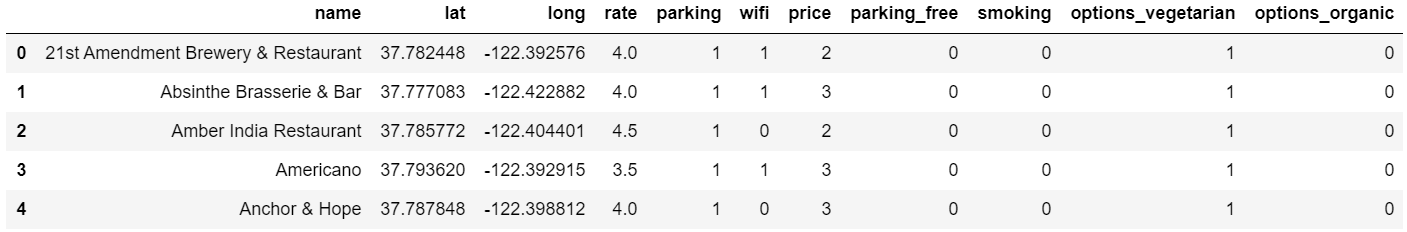
**2.1 Data Sources**

For Customer Rating: Kaggle Data Set <https://www.kaggle.com/jkgatt/restaurant-data-with-100-trip-advisor-reviews-each>.

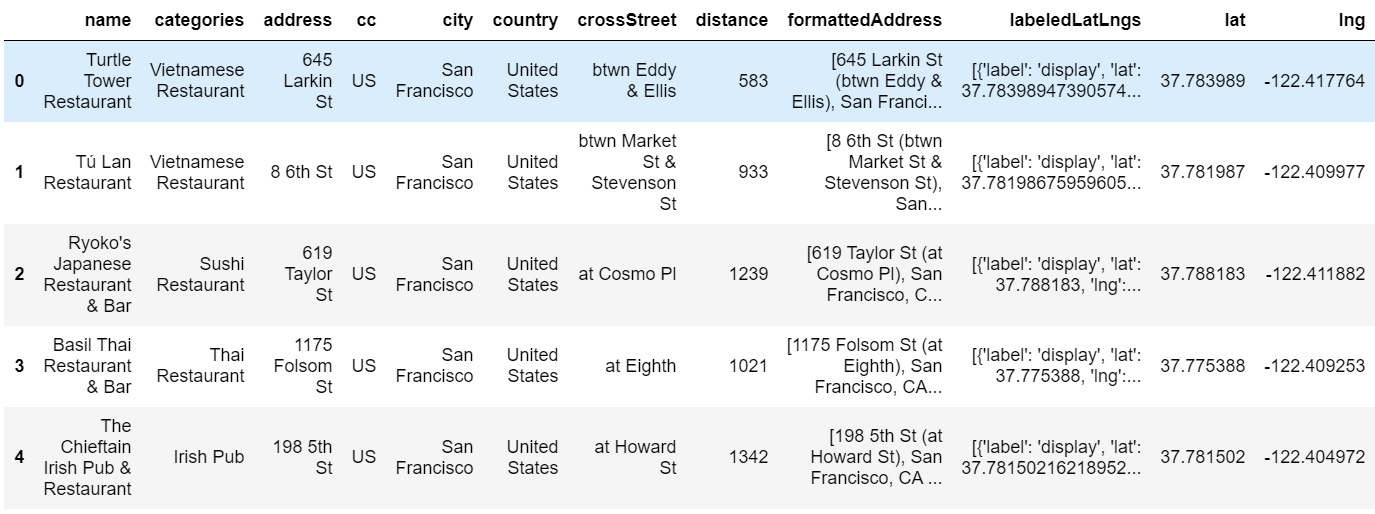
For Restaurants location: Foursquare <https://foursquare.com/>

**2.2 Data Cleaning and Feature selection**

Data format from kaggle is json; converted to data. In addition the features selected are true or false converted to 1, 0.

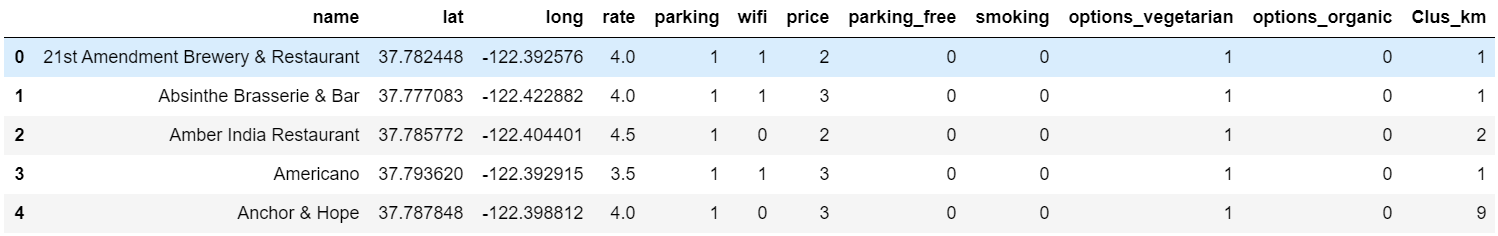


Data format from Foursquare is json, converted to data frame and kept only columns that include venue name and anything that is associated with location.



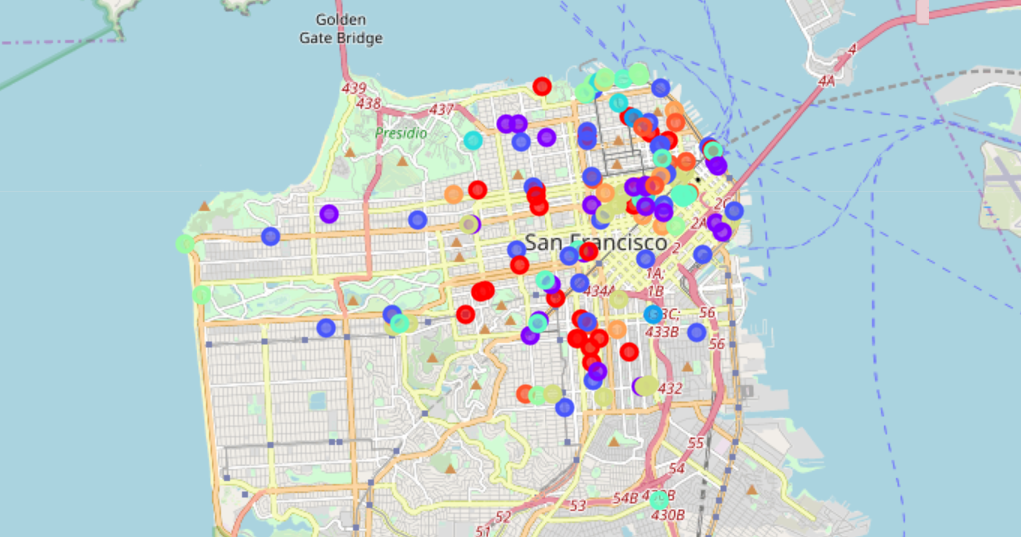
**3. Clustering**

Using Kmeans from Sci kit learn to cluster the data set from kaggle based on the features selected into 10 clusters.



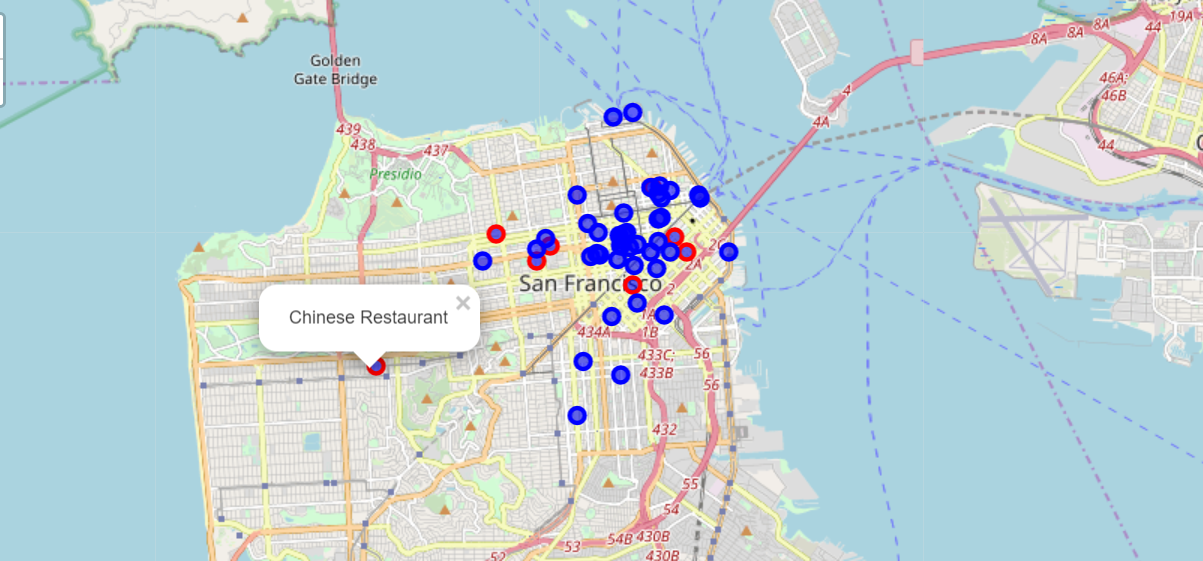
**4. Mapping**

**4.1 Map for Classified restaurants:**



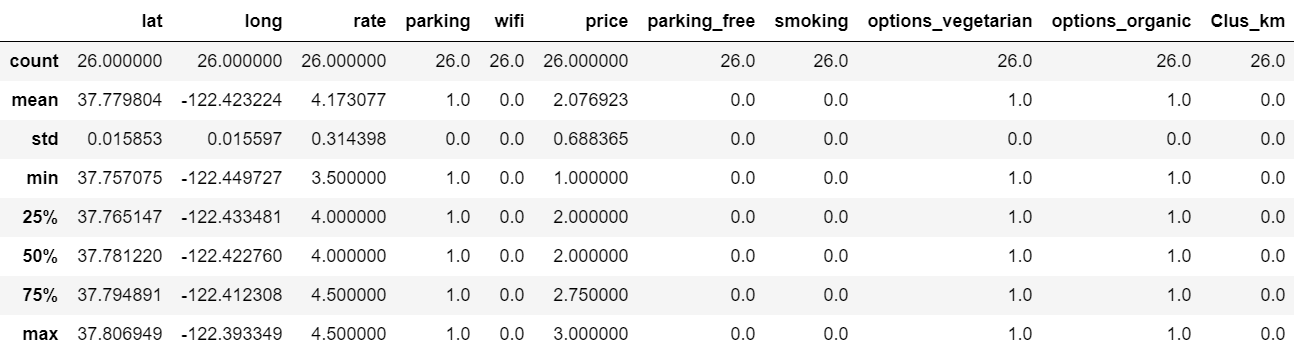
**4.2 Map for restaurants from Foursquare**

Red Are Chinese Restaurants.

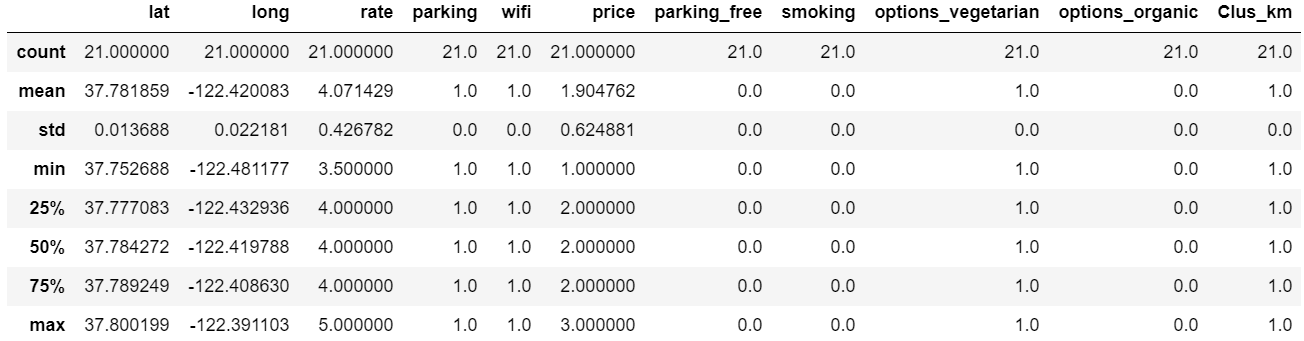


**5. Descriptive analysis for the clusters**

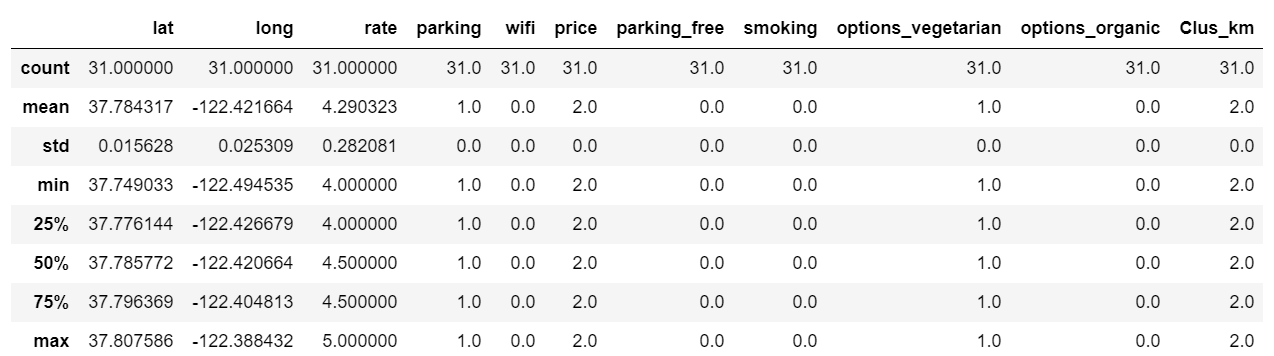
**Cluster 0**

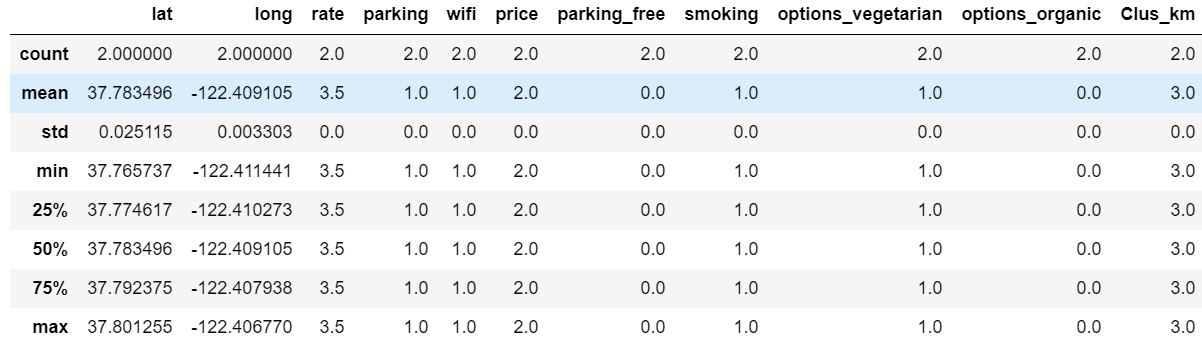


**Cluster 1**

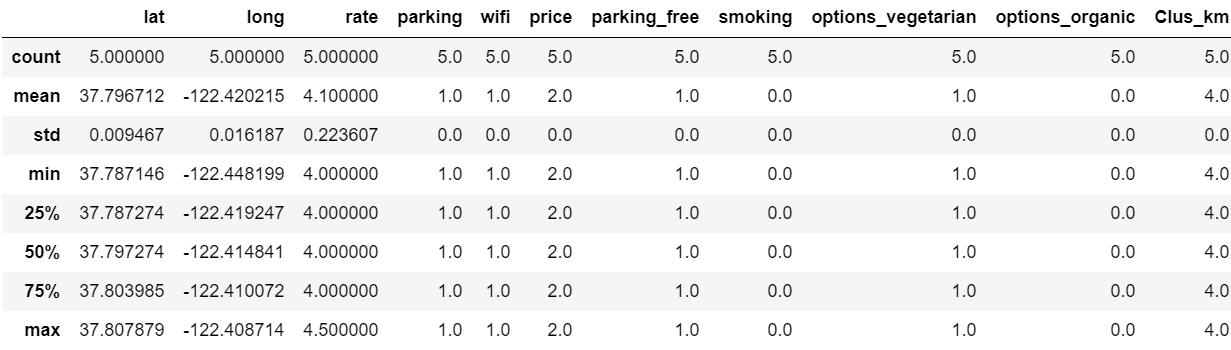


**Cluster 2**

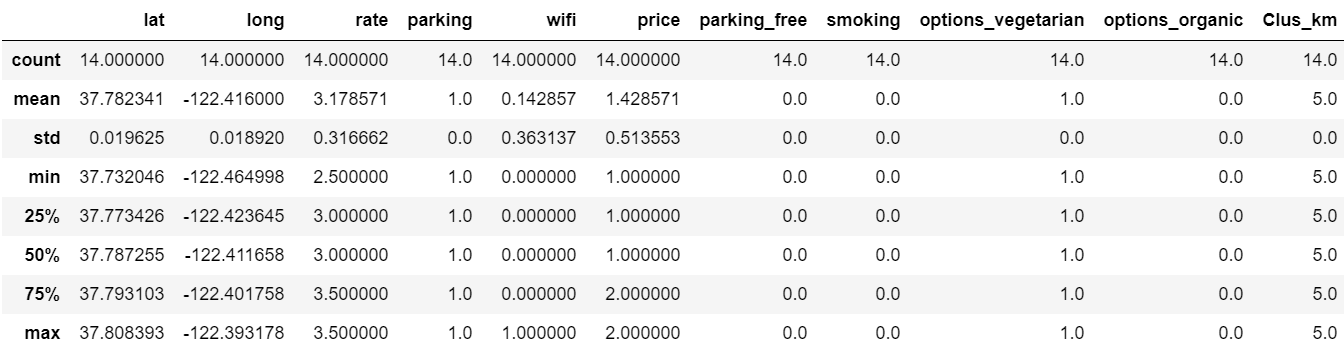


Cluster 3

Cluster 4



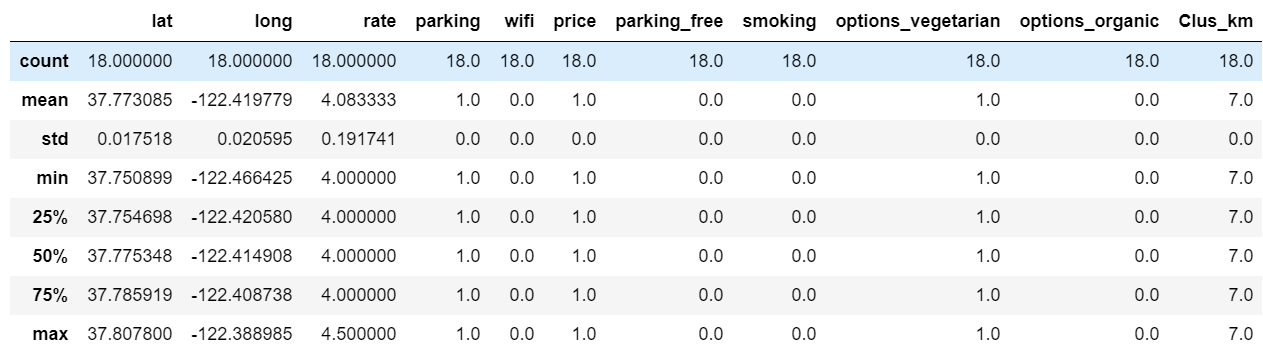
Cluster 5



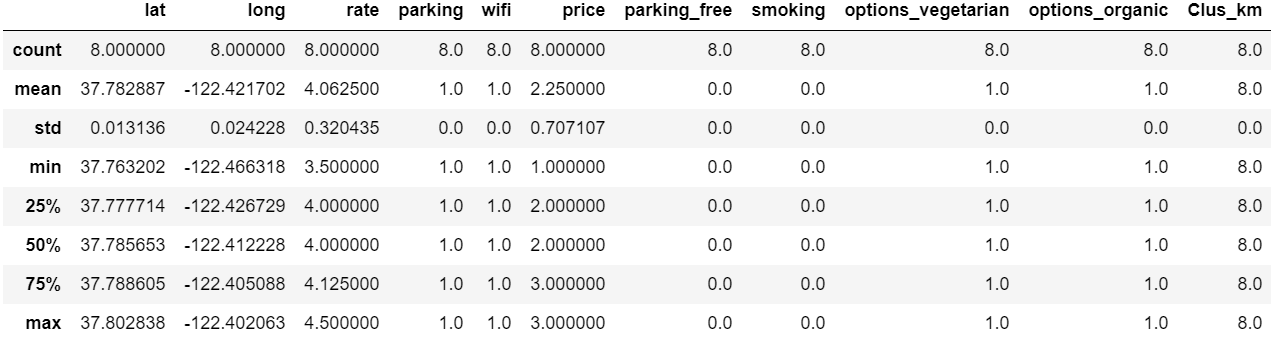
Cluster 6



Cluster 7



Cluster 8



Cluster 9



**6. Analysis**

Looking at 75% at each cluster :

Free parking, options\_vegetarian and options\_organic do not contribute to good rate.

Smoking is main reason for low rate.

Wifi and Price contribute to good rate.

**7. Conclusion**

The Result of the study is that, opening a new Restaurant that has a good price strategy and has wifi, can increase the chances for good rate. Also by exploring the restaurants types from Foursquare and mapping them, we can find a location where there are no Chinese restaurants.