

# Capstone Project - The Battle of Neighborhoods (Week 1):

## -By Arunkumar

### 1. Introduction:

This project's main aim was to determine a suitable area in California to set up a restaurant business.

### 2. Dataset:

The dataset was taken from [https://share.cocalc.com/share/e9d2f604-5c15-48c1-8c69-4d560cf9a933/PythonDataScienceHandbook/notebooks/data/california\\_cities.csv?viewer=share](https://share.cocalc.com/share/e9d2f604-5c15-48c1-8c69-4d560cf9a933/PythonDataScienceHandbook/notebooks/data/california_cities.csv?viewer=share)

### 3. Project Description:

The primary focus of this project was to ensure that the area selected in California was suitable for business or not. Hence some key factors were taken into consideration:

- Area Population
- Longitude
- Latitude
- The area in sq.m

### 4. Data Wrangling:

The dataset was loaded from the URL mentioned above. Some of the critical aspects done were:

- 4.1. Locate all cities in California
- 4.2. Foursquare credentials
- 4.3. Locate all venues in California
- 4.4. Selected all the restaurant's cuisine and located them in California
- 4.5. One hot coding

### 5. Methodology:

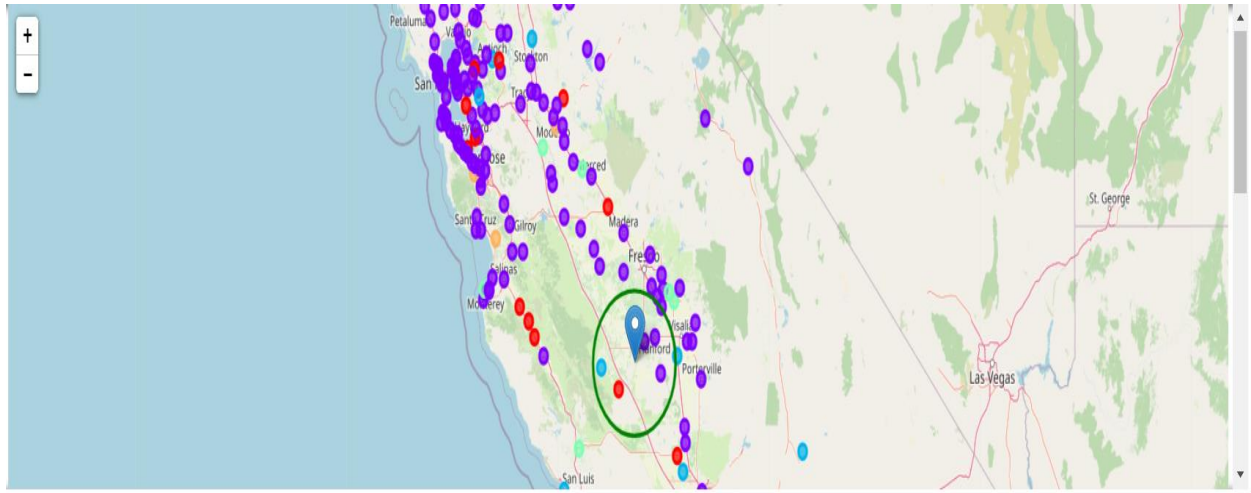
The algorithm selected was K means. K means can be defined as follows:

"K-means clustering is an iterative clustering algorithm where the number of clusters K is

predetermined, and the algorithm iteratively assigns each data point to one of the K clusters based on the feature similarity.”

The k value selected was 5. The clusters were then run, and the data was segmented into clusters, and the labels were generated. The geographic center was also determined.

## 6. Results:



The green circle on the map shows the area where a restaurant can be set up.

### 6.1. Examine Clusters:

Five clusters were examined and depicted in the code.

## 7. Discussion:

A lot of critical factors need to be considered while setting up a restaurant. The area and the population in that area is one of them. The preference of people living in the area is also another critical factor to be considered.

A k- means were also used to locate all the nearby points into a cluster. This is the reason why using K means felt like an ideal choice to me.

Locating the map area to set up a restaurant was one of the primary aims, and I was successful.

## 8. Conclusion:

The area marked on the map is by far the best suitable place to start a restaurant. An improvement I can make to this project is checking the crime rate and other restaurant ratings to check for any competition in the business.