Capstone Project   
Battle of Neighborhoods

Recommend location to open a new restaurant using K-means.

Author: Abhishek Pradhan

Table of Contents

[**1.** **Introduction:** 3](#_Toc1400034)

[**2.** **Data Collection & Python Libraries:** 4](#_Toc1400035)

[**4.** **Results:** 6](#_Toc1400036)

[**5.** **Discussion:** 7](#_Toc1400037)

[**6.** **Conclusion:** 8](#_Toc1400038)

1. **Introduction:**

Sacramento is the farm-to-fork capital of the country. The restaurant scene in Sacramento is maturing every day.

In this project, we will explore Sacramento neighborhoods and fetch data such as venues, its ratings, etc. for each neighborhood. The data analysis application will perform for comparative analysis of the venues in each neighborhood and recommend a location to the business owner to open a restaurant in Sacramento area.

1. **Data Collection & Python Libraries:**
   1. **Data Section:**

Sacramento Neighborhood Names and Co-ordinates collected from Google.

It has 50 neighborhoods.



**Table1**: Sacramento Neighborhood and geographical coordinates.

* 1. **Foursquare API:**

It has a database of more than 105 million places. Foursquare API is used as the primary data source.

* 1. **Python Library Files:**

• Pandas - Library for Data Analysis

• NumPy – Library to numeric functions

• JSON – Library to handle JSON files

• Folium – Map rendering Library

• Matplotlib – Python Plotting Module

• Geopy – To retrieve Location Data

• Requests – Library to handle http requests

• Sklearn – Python machine learning Library

* 1. **Folium:**

Python visualization library is be used to visualize the neighborhoods cluster distribution of Sacramento city over an interactive leaflet map.

1. **Methodology:**

Once the neighborhood data is acquired for any given city, the foursquare API call can be used to acquire the 10 most common ‘Trending’ venues around each neighborhood. The radius was set to 500m with a limit of 100 venues to be returned.

The returned venues are grouped using a hot encoding method to display for top 10 venues for each neighborhood.



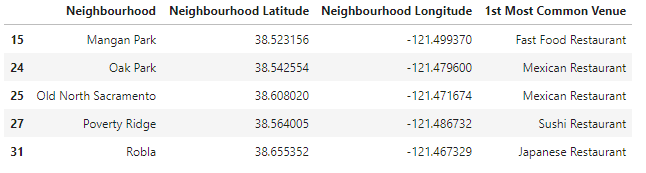
**Table2:** Top 10 venues in neighborhood.

* 1. **Unsupervised machine learning algorithm:**

K-mean clustering is be applied to form the clusters of different categories of places in and around the neighborhoods. Each of them would be analyzed individually and comparatively to derive the best location.

1. **Results:**

The most visited venue is the best location for opening new restaurant. This model identified five best locations to open a new restaurant in Sacramento area.



**Table3:** Top 5 neighborhoods and their 1st Most visited venue/restaurant.

1. **Discussion:**

This model can be used to suggest the best location to open a new restaurant in any city. This model is built on information provided by Foursquare, which could lead to bias. Other features such as population of the city, demographics can be used to improve the results.

1. **Conclusion:**

This model can be applied to any city where neighborhoods are known.This model segments the data into top 5 categories.This model reduces manual work in gathering and analyzing venues information. Businessmen can choose the location for opening a new restaurant much faster.