

Coursera Capstone Final Project

Introduction

This project is the final requirement towards earning the IBM Data Science Professional Certificate on Coursera. The purpose of the project is to demonstrate the use of tools and techniques learned in this data science course. The project has to utilize the FourSquare location data via their API as its primary data source along with the other statistical data acquired from other sources such as Wikipedia.

Business Problem

A Mediterranean restaurant owner known for their fast food style restaurants on the East coast wants to expand his business to the West coast. They would like to identify a location around Los Angeles County to open their first upscale restaurant and bar. This will not be a fast food type restaurant but a full service sit-in dining experience.

The place must meet the following criteria:

- Taking into account the higher price point at which the restaurant would operate, it must be in an upscale location.
- Must have low crime rates
- Must be near entertainment and shopping malls to gain more visibility and foot traffic
- There should be no other Mediterranean restaurant within a 2 miles radius

We will use unsupervised machine learning to create clusters of cities or areas in Los Angeles County that will provide us the insight to pick the right location.

Data Requirements

- List of cities in Los Angeles county and their geo coordinates
- Cities demographics including per capita income, populate and crime rate
- Venue data such as restaurants types, ratings, etc.

Data Sources:

- List of cities:
[List of cities in Los Angeles County, California - Wikipedia](#)
- Cities demographics: (Population and income)
https://en.wikipedia.org/wiki/List_of_California_locations_by_income
- Crime rate by city:
https://en.wikipedia.org/wiki/California_locations_by_crime_rate
- Venues data: Foursquare
- Geo coordinates: Python Geocoders library

Methodology

Step 1: Data will be acquired from Wikipedia as HTML tables and saved into CSV files for further processing. This will be done using Pandas Read function.

Step 2: The acquired datasets (data frames) consisting of Cities demographic data will be consolidated into one single data frame. The resulting consolidated data frame will contain city name, population, per capita income, and violent crime rate.

Step 3: All rows with county other than Los Angeles will be removed from the data frame.

Step 4: Add geospatial data to the data frame using the Nominatim function in the Conda-forge Geopy package.

Step 5: Using the Folium package, create a map of cities in LA County.

Step 6: Using Foursquare API, retrieve data for all restaurants within 1 mile of each city. This will include the restaurant categories, such as Chinese, Indian, Middle Easter, Mediterranean, etc.

Step 7: Transform the data frame with restaurant venues/categories to one-hot encoding (0/1).

Step 8: Use grouping to show the frequency of each restaurant category in each city.

Step 9: Using the info from step 8 above, create a data frame with most common restaurant categories for each city.

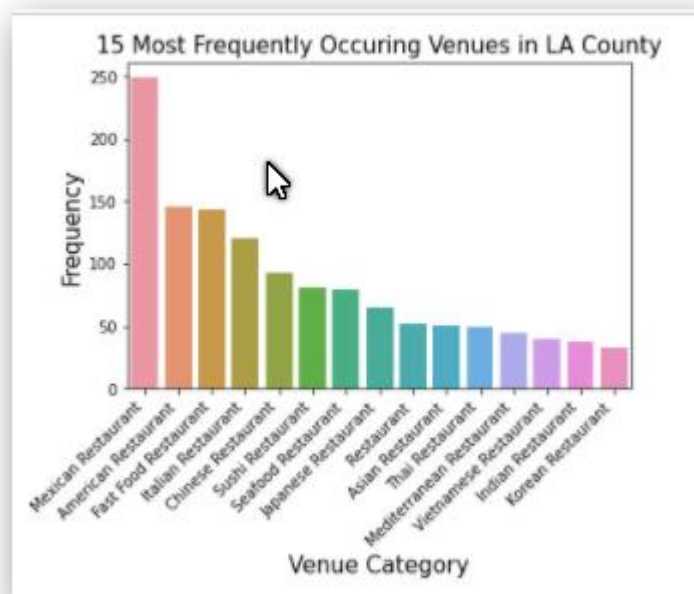
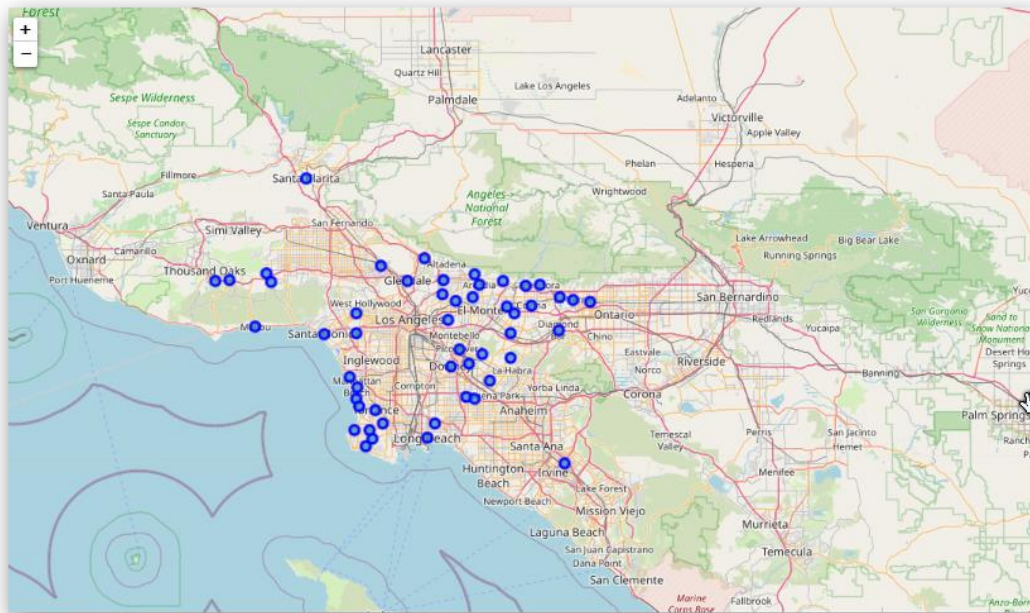
Step 10: Use silhouette score to identify the optimal number of parameter for the number of clusters.

Step 11: Using the silhouette score for number of clusters, run a **k-means clustering algorithm** (unsupervised learning) from the Scikit-learn package.

Final Step: Review the clusters to pick the one with most venues and make recommendation to client.

Results

The population, crime and income data for all the cities in LA county was consolidated, cleaned and filtered for low income and high crime rates. The Foursquare API was used to get the geo coordinates, mapped for visual presentation. Five data clusters were created using k-Means algorithm and analyzed for final recommendation.



	Neighborhood	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	Agoura Hills	Mexican Restaurant	Fast Food Restaurant	Chinese Restaurant	Indian Restaurant	Italian Restaurant	Restaurant	Japanese Restaurant	Sushi Restaurant	French Restaurant	Middle Eastern Restaurant
1	Alhambra	Spanish Restaurant	Tapas Restaurant	Restaurant	Italian Restaurant	Seafood Restaurant	Fast Food Restaurant	Mediterranean Restaurant	American Restaurant	Japanese Restaurant	Dim Sum Restaurant
2	Arcadia	Mexican Restaurant	Sushi Restaurant	American Restaurant	Fast Food Restaurant	Dumpling Restaurant	Chinese Restaurant	Seafood Restaurant	Restaurant	Cantonese Restaurant	Hotpot Restaurant
3	Artesia	Indian Restaurant	Thai Restaurant	Korean Restaurant	Chinese Restaurant	Italian Restaurant	American Restaurant	Japanese Restaurant	Seafood Restaurant	Mediterranean Restaurant	Filipino Restaurant
4	Azusa	Mexican Restaurant	American Restaurant	Fast Food Restaurant	Restaurant	Thai Restaurant	Italian Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Dumpling Restaurant	Hotpot Restaurant
5	Baldwin Park	Mexican Restaurant	Fast Food Restaurant	Italian Restaurant	Asian Restaurant	Latin American Restaurant	Hawaiian Restaurant	Dumpling Restaurant	Indian Restaurant	Indian Chinese Restaurant	Hotpot Restaurant
6	Beverly Hills	American Restaurant	Italian Restaurant	Mediterranean Restaurant	Restaurant	Sushi Restaurant	New American Restaurant	Mexican Restaurant	Seafood Restaurant	Moroccan Restaurant	French Restaurant
7	Bradbury	Mexican Restaurant	Fast Food Restaurant	Italian Restaurant	American Restaurant	Thai Restaurant	Mediterranean Restaurant	Sushi Restaurant	New American Restaurant	Chinese Restaurant	Restaurant
8	Burbank	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Sushi Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Asian Restaurant	Cuban Restaurant	Empanada Restaurant	Indian Restaurant
9	Calabasas	Italian Restaurant	Sushi Restaurant	Seafood Restaurant	Fast Food Restaurant	New American Restaurant	American Restaurant	Brazilian Restaurant	Dumpling Restaurant	Indian Chinese Restaurant	Hotpot Restaurant
10	Cerritos	Indian Restaurant	Mexican Restaurant	American Restaurant	Chinese Restaurant	Filipino Restaurant	Italian Restaurant	Asian Restaurant	Korean Restaurant	Fast Food Restaurant	French Restaurant
11	Claremont	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Thai Restaurant	Italian Restaurant	Restaurant	Chinese Restaurant	Japanese Restaurant	Sushi Restaurant
12	Covina	Mexican Restaurant	American Restaurant	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Asian Restaurant	Cajun / Creole Restaurant	Korean Restaurant	Latin American Restaurant	Peruvian Restaurant
13	Culver City	Sushi Restaurant	New American Restaurant	Mexican Restaurant	Indonesian Restaurant	Mediterranean Restaurant	American Restaurant	Japanese Restaurant	Cuban Restaurant	French Restaurant	Vegetarian / Vegan Restaurant
14	Diamond Bar	Fast Food Restaurant	Chinese Restaurant	Vietnamese Restaurant	Mexican Restaurant	Japanese Restaurant	Korean Restaurant	American Restaurant	Sushi Restaurant	Indian Restaurant	Seafood Restaurant
15	Downey	Mexican Restaurant	American Restaurant	Seafood Restaurant	Italian Restaurant	Fast Food Restaurant	Japanese Restaurant	Sushi Restaurant	Chinese Restaurant	Peruvian Restaurant	Mediterranean Restaurant
16	El Segundo	Mexican Restaurant	Italian Restaurant	American Restaurant	Fast Food Restaurant	Restaurant	Seafood Restaurant	Mediterranean Restaurant	Cuban Restaurant	Sushi Restaurant	Japanese Restaurant
17	Glendale	Japanese Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Asian Restaurant	Ramen Restaurant	Fast Food Restaurant	Italian Restaurant	Seafood Restaurant	Vietnamese Restaurant	Argentinian Restaurant
18	Glendora	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Argentinian Restaurant	Chinese Restaurant	Italian Restaurant	Japanese Restaurant	Vietnamese Restaurant
19	Hemosa Beach	Mexican Restaurant	Italian Restaurant	American Restaurant	Seafood Restaurant	Sushi Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	Peruvian Restaurant	French Restaurant	Fast Food Restaurant
20	Hidden Hills	Fast Food Restaurant	Sushi Restaurant	Seafood Restaurant	American Restaurant	New American Restaurant	Italian Restaurant	Mexican Restaurant	Chinese Restaurant	Hotpot Restaurant	Hawaiian Restaurant

Discussion and Conclusion

After reviewing the clusters generated for most common venues by each city, Agoura Hills is recommended as it meets the following client's criteria:

- Low crime rate of 0.82 per 1000 population
- Higher median income of \$107,268
- No Mediterranean restaurants in the vicinity

Link to Notebook

<https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/6c5f3210-437a-4476-af1a-7041d543d521?projectid=5edb56c5-477f-4b3a-8e42-2df03f09180d&context=cpdaas>