

Recommender System for Commercial Warehouse Week 1

Applied Data Science Capstone by H. Bhattacharya

Introduction and Business Problem

Los Angeles is a historical city that is diverse, vibrant and full of opportunities. However, because of its popularity and appeal, can be an expensive place to start a business and thrive. Many businesses want to open a venue here but need to ensure that high startup cost is quickly recovered. Within such a competitive market and high real estate prices, it is challenging for upstarts to find a place to establish themselves. My intention is to use Data Science techniques learnt to make this an intelligent choice based on sold data.

In this scenario, there is a Commercial Cold Storage Warehouse provider looking to open a new Warehouse somewhere in Los Angeles. The deciding factor includes whether there are Restaurants and Grocery Stores in surrounding communities that will use this service in a positive trend.

Data Description

As we need to explore, segment, and cluster the neighborhoods in the city of Los Angeles, Los Angeles neighborhood data is key for this project. Fortunately, I was able to find the data in the appropriate structured format here <https://usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr>

```
[4]: la_data

[4]: {'type': 'FeatureCollection',
      'features': [{'type': 'Feature',
                    'properties': {'external_i': 'acton',
                                  'name': 'Acton',
                                  'location': 'POINT(34.497355239240846 -118.16981019229348)',
                                  'latitude': '-118.16981019229348',
                                  'slug_1': None,
                                  'sqmi': '39.3391089485',
                                  'display_na': 'Acton L.A. County Neighborhood (Current)',
                                  'set': 'L.A. County Neighborhoods (Current)',
                                  'slug': 'acton',
                                  'longitude': '34.497355239240846',
```

JSON snip

I was able to explore and cluster the neighborhoods in Los Angeles. The data contained the coordinates for each of the neighborhoods that helped to pull critical venue information for this project using Foursquare API. This was done using these key fields:

1. Neighborhood Name
2. Neighborhood Latitude
3. Neighborhood Longitude

[10]:	Neighborhood	Latitude	Longitude
0	Acton	34.497355	-118.169810
1	Adams-Normandie	34.031461	-118.300208
2	Agoura Hills	34.146736	-118.759885
3	Agua Dulce	34.504927	-118.317104
4	Alhambra	34.085539	-118.136512

Data Features

We will be leveraging on features in a reliable location information provider such as the Foursquare.com to explore the various types of venues and its categories available in each neighborhood. We will also need to understand the trending of these venues in the respective neighborhood. The information obtained per neighborhood will be as such like below and must be in a structured format:

4. Venue Name
5. Venue Category
6. Venue Latitude
7. Venue Longitude

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acton	34.497355	-118.169810	Epik Engineering	34.498718	-118.168046	Construction & Landscaping
1	Acton	34.497355	-118.169810	Alma Gardening Co.	34.494762	-118.172550	Construction & Landscaping
2	Adams-Normandie	34.031461	-118.300208	Orange Door Sushi	34.032485	-118.299368	Sushi Restaurant
3	Adams-Normandie	34.031461	-118.300208	Shell	34.033095	-118.300025	Gas Station
4	Adams-Normandie	34.031461	-118.300208	Little Xian	34.032292	-118.299465	Sushi Restaurant

How the problem will be solved

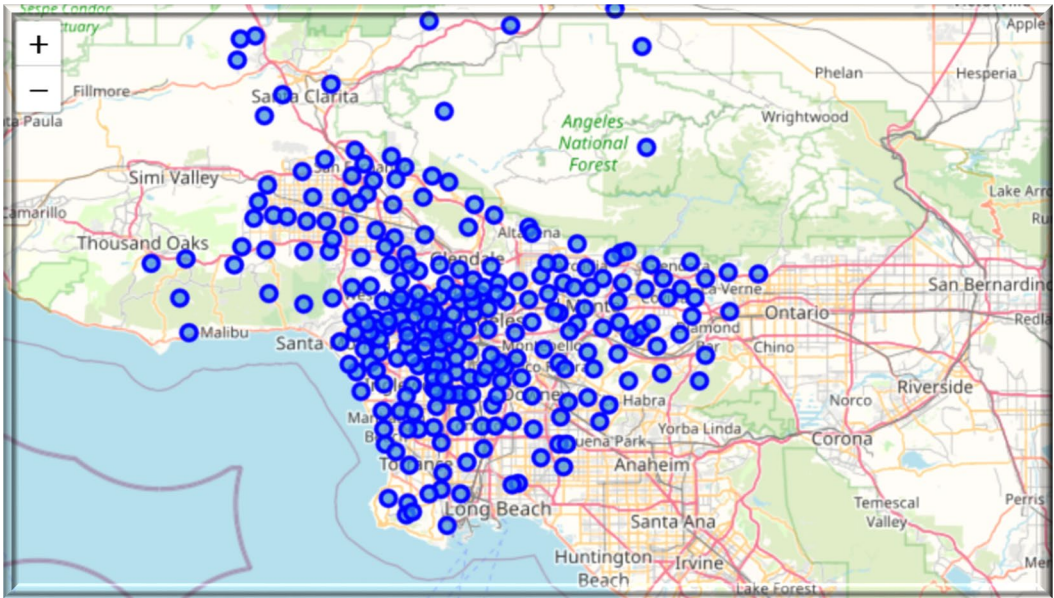
We use K-Clustering techniques to segment and cluster these neighborhoods so that we can group them together to understand their similarities and what best we can do in these types of neighborhoods.

With all these features, techniques and data, we will then be able to come up with a best recommendation for the Commercial Warehouse, that is where is the optimal neighborhood for them to build and base their services. For an example, we will want to enter a neighborhood where there is a high concentration of grocery stores and restaurants; where we know that there will be a higher demand of such storage services in that area.

This is project will make use of many data science tools, working with API (Foursquare), data cleaning, data wrangling, machine learning (K-means clustering) and map visualization (Folium).

	Neighborhood	Latitude	Longitude	Cluster Labels	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
0	Acton	34.497355	-118.169810	0	Construction & Landscaping	Yoga Studio	Falafel Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Escape Room
1	Adams-Normandie	34.031461	-118.300208	1	Sushi Restaurant	Yoga Studio	Playground	Park	Taco Place	Grocery Store	Bookstore
2	Agoura Hills	34.146736	-118.759885	1	Fast Food Restaurant	Chinese Restaurant	Breakfast Spot	Hotel	Burger Joint	Bakery	Thai Restaurant
3	Agua Dulce	34.504927	-118.317104	1	Airport	Yoga Studio	Farm	Electronics Store	Empanada Restaurant	English Restaurant	Escape Room

Dataframe with cluster labels



Folium used to superimpose neighborhoods on map