

# A Recommender System for Commercial Warehouse

---

Applied Data Science Capstone

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 live and thrive. Many businesses want to open a venue here but pay top price to have their presence made. 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.

There is a Commercial Cold Storage Warehouse provider looking to open a new Warehouse in Los Angeles. The deciding factor includes whether there are Restaurants and Grocery Stores in surrounding communities that will pull new business.

# Introduction and Business Problem

---

The analysis will be of interest to the following groups: 1. First time entrepreneurs, who want to start their first business. Below dataset will give a comprehensive insight into where best to open a new venue, to maximise the value for money. 2. People who already run a business and want to branch out. Given the extra information, it may provide valuable information before decision making.

# Outline

---

- **1: Problem Description**

We will need geo-locational information about Los Angeles, specifically Beverly Hills and environs.



Los Angeles / Coordinates

34.0522° N, 118.2437° W

# Outline

---

- **2: Steps to Take**
  - Download and Explore Dataset
  - Explore Neighborhoods in LA
  - Analyze Each Neighborhood
  - Cluster Neighborhoods
  - Examine Clusters

# Download and Explore Dataset

---

- 3: Search for then Download the appropriate dataset

In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the neighborhoods that exist as well as the latitude and longitude coordinates of each neighborhood.

```
[4]: la_data
[4]: {'type': 'FeatureCollection',
      'features': [{'type': 'Feature',
                    'properties': {'external_id': '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',
                                  'name_1': None,
                                  'kind': 'L.A. County Neighborhood (Current)',
                                  'type': 'unincorporated-area'},
                    'geometry': {'type': 'MultiPolygon',
                                  'coordinates': [[[-118.20261747920541, 34.53898972076929],
                                                    [-118.18946958918568, 34.5385546636616],
```

# Download and Explore Dataset

- 3: Data Required

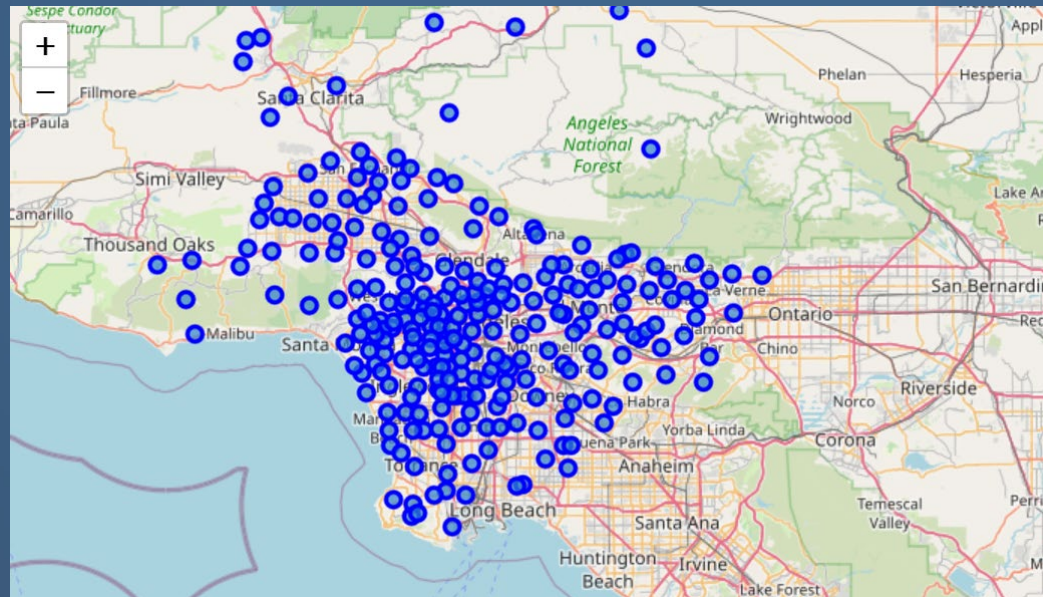
We will need data about different venues in different neighborhoods of Los Angeles. In order to gain that information we will use "Foursquare" locational information. A typical request from Foursquare will provide us with the following information:

	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

# Download and Explore Dataset

- 3: Map the data

Use Folium to Visualize the Data on a map:





# How the problem will be solved

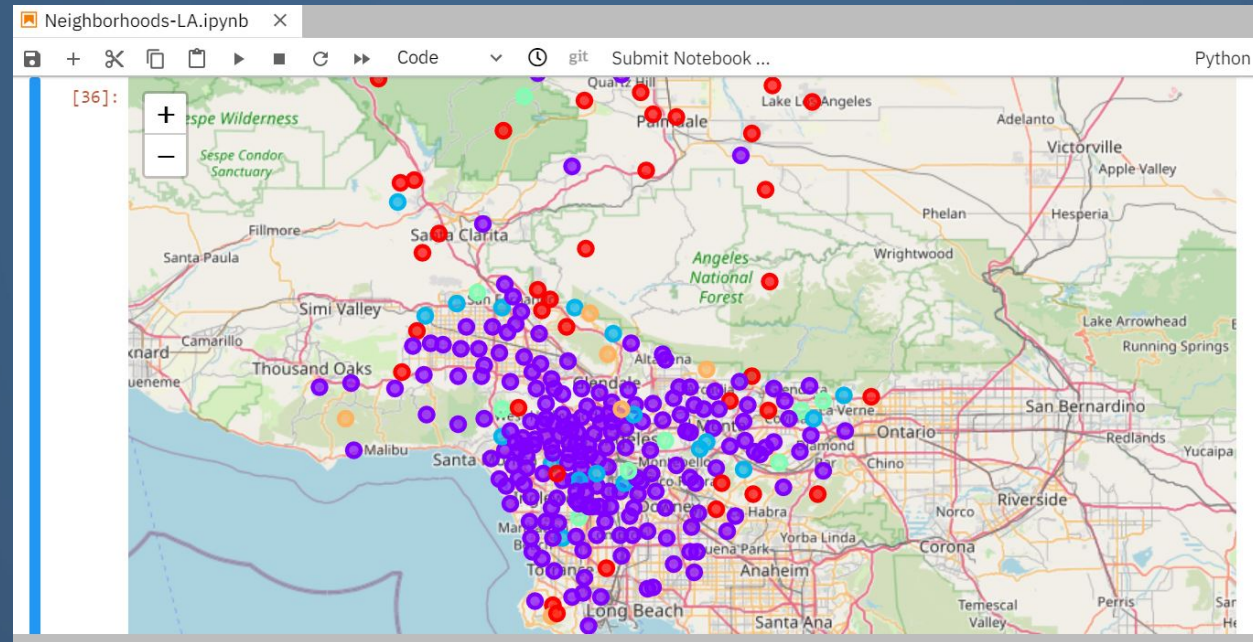
- 4: Exploratory Data Analysis

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:

	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

# How the problem will be solved

- 4: Visualize the Clusters using Folium



# Results and Decision Made

- 5: Conclusion

Most of the Grocery Stores and Restaurants are concentrated in the central area of LA, with the highest number in cluster 2 and moderate number in cluster 4.

Cluster 2									
[45]: la_merged.loc[la_merged['Cluster Labels'] == 1, la_merged.columns[[0] + list(range(4, la_merged.shape[1]))]]									
[45]:									
	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
1	Adams-Normandie	Sushi Restaurant	Yoga Studio	Playground	Park	Taco Place	Grocery Store	Bookstore	Gas Station
2	Agoura Hills	Fast Food Restaurant	Chinese Restaurant	Breakfast Spot	Hotel	Burger Joint	Bakery	Thai Restaurant	Sushi Restaurant
3	Agua Dulce	Airport	Yoga Studio	Farm	Electronics Store	Empanada Restaurant	English Restaurant	Escape Room	Ethiopian Restaurant
4	Alhambra	Convenience Store	Sporting Goods Shop	Construction & Landscaping	Bagel Shop	Pet Store	Breakfast Spot	Video Store	Pizza Place
6	Artesia	Indian Restaurant	Chinese Restaurant	Vietnamese Restaurant	Bubble Tea Shop	Pizza Place	BBQ Joint	Tea Room	Bar
7	Altadena	Home Service	Food	Campground	Pharmacy	Notary	Yoga Studio	Fabric Shop	Empanada Restaurant

# Results and Decision Made

---

- 5: Conclusion

The neighborhoods in cluster 2 are the most preferred locations to open a new Commercial Cold Storage Warehouse.

Cluster 2 has the best potential as there are the highest concentration of Grocery Stores and Restaurants. This would mean that it is likely that food businesses would use the Cold Storage facilities being offered.