Coursera Capstone
IBM Applied Data Science Capstone

Best place to open a grocery store in Los Angeles County

by Barane Yoga Natarajan

Introduction:

Considering the dense population of Los Angeles county and it's spread out nature, people have to travel a lot in order to buy groceries and other household necessities. If a neighborhood grocery store is opened in a highly populous area with very little grocery stores, it will attract customers and prevent oversupply of goods.

Business problem:

Ideally, we want to open a grocery store where there is a high population and less grocery stores.This can be done by the use of data science methodologies where we can compare the presence of grocery stores in the cities of Los Angeles.

Data:

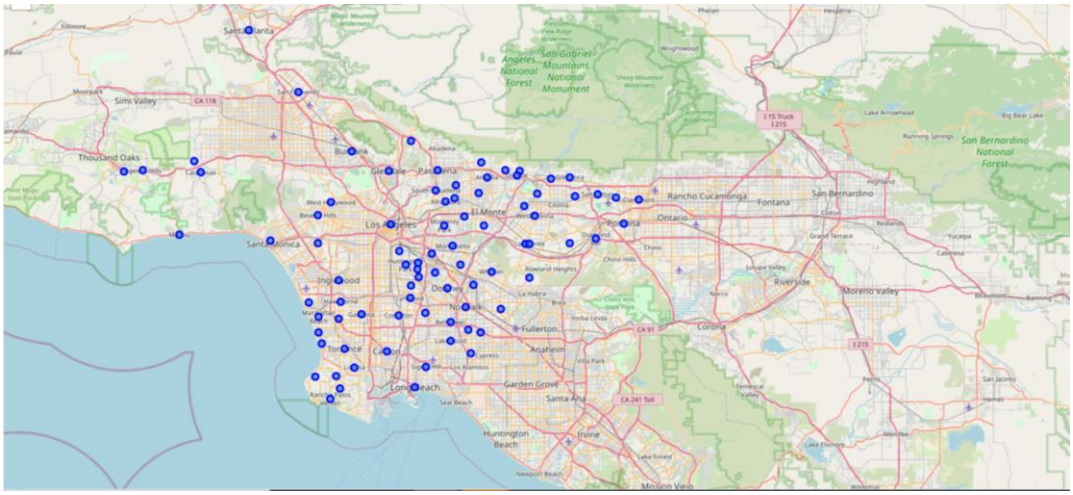
The use of foursquare API to better cluster neighborhoods can give us a good idea of where to open our store. I will do this by using data from wikipedia.
https://en.wikipedia.org/wiki/List_of_cities_in_Los_Angeles_County,_California

Methodology:

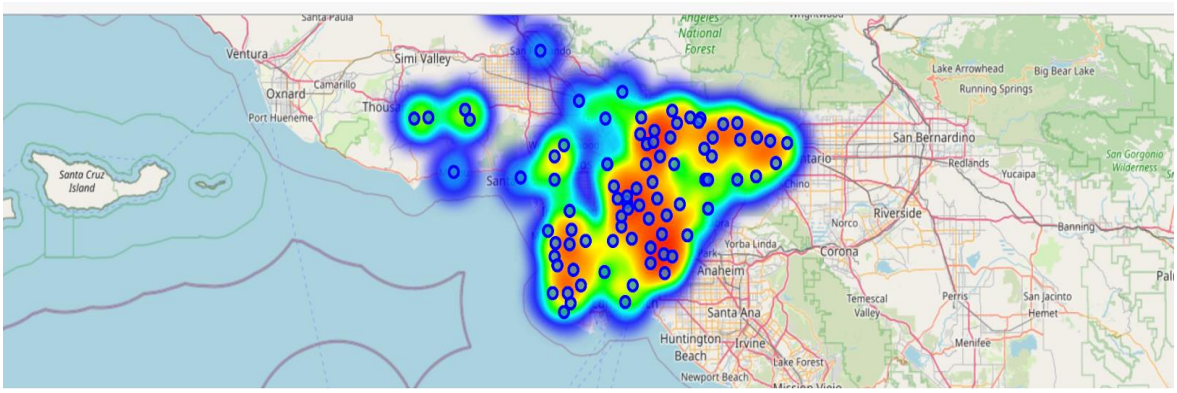
1. First, I scraped data containing the cities in LA county. After cleaning the data, I appended the latitude and longitude of the cities to the original data frame containing the cities in LA using geopandas. Then, I used a folium module to plot the cities on the map to visualize the data.
2. Using the Foursquare API client ID, I collected the venues near Los Angeles county. With this information, I filtered out all the grocery stores across LA county and plot the frequencies of grocery stores in the cities of Los Angeles.
3. Using these frequencies, I made 5 clusters of the grocery stores and cities in LA to identify which cluster had the highest and lowest grocery stores

Results:

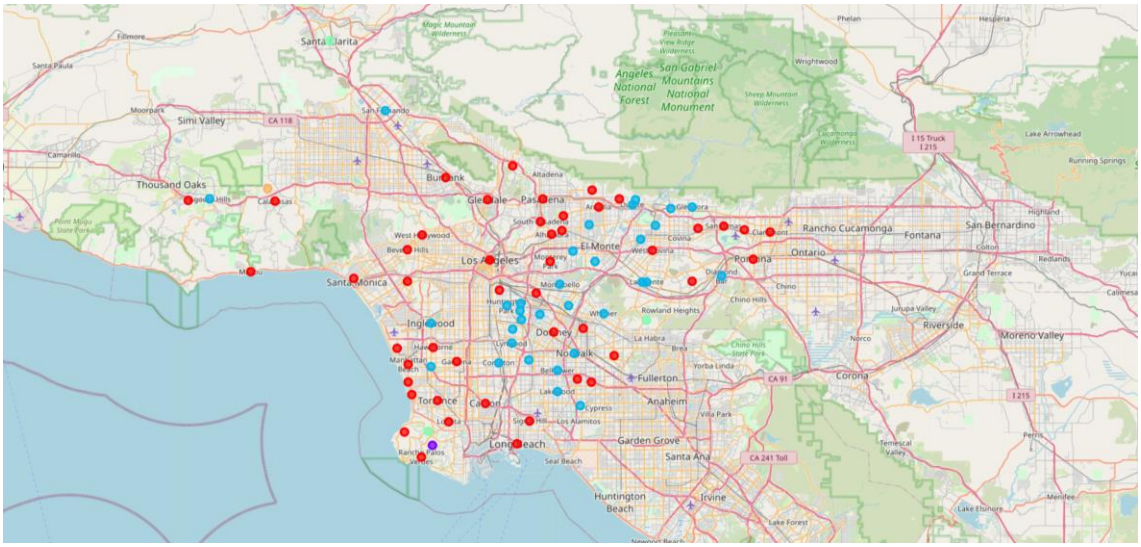
Through clustering, I found that cluster 0 had very little to no grocery stores in their vicinity and cluster 1 and 3 contain some of the highest grocery stores. This is a map of all the cities in Los Angeles county marked with a blue dot.



This is a heat map of the population spread in the cities of Los Angeles county. As we can see, the majority of the population is concentrated near the coast which makes it a good spot for a grocery store.



This is a map of the clusters of grocery stores in Los Angeles county:



Discussion:

With the above analysis, it is clear that cluster 0 is a suitable location for the grocery store given it's high population and low number of existing grocery stores. Cluster 1 and 4 on the other hand have a very high presence of grocery stores, making them an unlikely option for a new grocery store. Also, cluster 0 presents an opportunity to serve tourists as well since it is in close proximity to tourist locations like Santa Monica, Hermosa Beach, etc.

Limitations:

Since the data used in this research was scraped from wikipedia, the authenticity of the population of different cities is not very high. Also, this data may not be sufficient to tell if the location of the store is perfect since other factors like area-to-area taxation, property rent, preferences vary. Since door delivery and online ordering are becoming more and more common, it can mean that the demand in cluster 0 is already being met by other grocery stores in th Los Angeles county.

Conclusion:

In this project, we have gone through the various potential locations for a new grocery store in the Los Angeles county through population analysis and the use of API to determine the location of existing grocery stores. As per our findings, we can conclude that cluster 0 is a suitable location for the store.