Selecting the Best Neighborhood in Los Angeles to open a new Ethiopian Restaurant

Introduction

Running a successful restaurant business in any city is hard. Most restaurants close within a short period of time after opening because they are not profitable. Therefore, considering all the factors that affect the success of the restaurant is important.

One of the important factors that affect the restaurant business is its location. In this project, we will try to find the best location to open a new Ethiopian restaurant in Los Angeles, CA. We consider the density of other restaurants in the neighborhoods, and next the density of Ethiopian restaurants in the neighborhoods. So, we would not have many competitors surrounding our restaurant.

This problem is important to solve for stakeholders, such as any person who is interested in opening a new Ethiopian restaurant in Los Angeles.

Data

We will use three data sources to solve the problem.

- We will get the list of neighborhoods of Los Angeles, CA by scraping a Wikipedia page.
- We will get Latitude and Longitude coordinates of the neighborhoods, to plot maps and use the coordinates to get venue data. We will use Python Geocoder package to get the coordinates.
- We will get venue data, especially restaurants and particularly Ethiopian restaurants, in these neighborhoods using Foursquare API.

We will use this data to perform clustering on the neighborhoods.

Methodology

The first dataset we need for this project is the neighborhoods in Los Angeles. I created an excel table of the names of the neighborhoods in Los Angeles by extracting it from a Wikipedia page and we load it onto python and read it. Then we need to get the coordinates of the neighborhoods and so we use geocoder to get the coordinates. Next, we use foursquare API to explore the first neighborhood in our data frame and find 26 unique venues. We create a function to repeat the same process to all the neighborhoods. We find 366 unique categories. We analyze each neighborhood by creating one hot encoding and group the neighborhood by the frequency of venue category for each neighborhood. We use this grouped one hot data frame to find the top 10 most common venues for each neighborhood; and use to cluster the neighborhoods using k-means clustering algorithm. We visualize the clusters using folium package.

Results

We examine each cluster for the most common venues and the existence and quantity of Ethiopian restaurant.

Cluster 0

This cluster has 5 neighborhoods in it, and its 1st most common venues are 'Food Truck' with 2 counts, 'Convenience Store' with 2 counts, and 'Other Great Outdoors'. Since it does not have food venue other than food truck in its 1st most common venue, we consider it for further analysis to determine if it is the location to open an Ethiopian Restaurant. With regards to Ethiopian restaurant venue, this cluster has this venue as one of the topmost common venue (1st thru 10th) in its neighborhoods except one, Playa Vista.

Cluster 1

This cluster has 4 neighborhoods in it, and its 1st most common venues are 'Restaurant' with 3 counts and one 'Home Service'. Since restaurant is the 1st most common venue in this cluster and it has 'Ethiopian Restaurant' as its 7th most common venue in all the neighborhoods, we will not consider this cluster for further analysis and make the determination that this is not the best location to open an Ethiopian restaurant in Los Angeles.

Cluster 2

This cluster has more than 100 neighborhoods in it and it is the largest cluster by far. The 1st most common venue in this cluster are 21 'Coffee Shops', 18 'Mexican Restaurant', 13 'Pizza Place', 8 'Fast Food Restaurant', and so on. This cluster has many other food venues in its top 10 most common venues. As far as the specific venue 'Ethiopian Restaurant', this cluster has multiple neighborhoods that has this venue in the top most common venue (1st thru 10th), for instance, Little Ethiopia has it as 1st most common venue, Carthay has it as its 6th most common venue. Therefore, we will not consider this cluster as suitable location to open an Ethiopian restaurant.

Cluster 3

This cluster has 3 neighborhoods in it, and its 1st most common venue is 'Trail'. This cluster also has 'Ethiopian Restaurant' venue as one of the top 10 most common venue in all its neighborhoods. The three neighborhoods all have 10 or less venues in them, therefore, because of the existence of the venue 'Ethiopian Restaurant' as top common venue in all of the neighborhoods and the small number venues that exist in total indicating not a central location; we will not consider this cluster.

Cluster 4

This cluster has 12 neighborhoods in it. The 1st most common venues are 'Park' with 8 counts, 'Clothing Store', 'Pizza Place', 'Entertainment Service' with one count each. Even though, no food venue is in the

1st most common venue, 2nd and 3rd most common venues have food venues including restaurants. With regards to the venue 'Ethiopian Restaurant', all 12 neighborhoods in this cluster have this venue in one of the top 10 most common venue. Therefore, we will not consider opening an Ethiopian restaurant.

Based on the analysis of each cluster, cluster 0 is the best location to open an Ethiopian restaurant. Since all the neighborhoods have 'Ethiopian Restaurant' venue in the top most common venue except 'Playa Vista', we will pick 'Playa Vista' neighborhood of Los Angeles to be the best location to open an Ethiopian restaurant.

Discussion

One observation is that some of the neighborhoods in the data set have low number of venues in them, so in the future we should have a threshold value for the number of venues that a neighborhood could have to obtain a balanced clustering and outcome.

Conclusion

In this project, we were able to use location data from foursquare API to find venues within neighborhoods in Los Angeles; use machine learning algorithm k-means clustering to cluster the neighborhoods and apply the criteria that we set for opening a new Ethiopian restaurant, that is, fewer competition.