Clustering Los Angeles Venues by Categories

Manan Brahmbhatt

April 28, 2020

1. Introduction

1.1 Background

Los Angeles, California is a booming place where many people go to visit. Whether it be people living in the United States or foreigners, anyone going to California visits Los Angeles for its great nightlife, family destinations, gardens, stadiums, and museums. Along with that, Los Angeles also contains some of the most diverse food options in the world.

1.2 Problem

This project aims to cluster the venues in Los Angeles by categories: sports, nature, history, nightlife, family destinations, and food.

1.3 Interest

People traveling to Los Angeles would be interested in this project, because it attempts to streamline the options they have to their interests.

2. Data Acquisition

The data was acquired from here. This data is the list of neighborhoods in the city of Los Angeles. After that, we acquired venue data from Foursquare API.

3. Analysis

3.1 Exploratory Analysis

We see that there are over 2000 venues in the data. However, there were some repeat values, which were dropped. The final number of venues is 2573, and there are 303 unique categories.

3.2 Analysis of the Problem

We then move to the actual problem. We first establish a predetermined set of words that could potentially show up in each category. For example, if we want to classify sports venues, we would look for words such as "Baseball", "Football", etc. in the venue category. We repeat this process for all six categories. We then go through the venues data and split it up based on categories.

After we do this, we clean up the data and put it all into one dataframe. At this point, the data is clustered and has been given a cluster number: 1 for sports, 2 for food, 3 for family, 4 for party, 5 for history, and 6 for nature. We then go on and map the venues.

The map shows that there is not a specific area where a category is more predominant. Food locations, as expected, are abundant, followed by nightlife locations. After that, the other locations have a similar number of locations.

4. Conclusion and Further Directions

This project was to help identify where people traveling to Los Angeles could potentially travel to based on their interests. We first took the data and organized it in terms of the different categories, and finally we mapped them out to make it visually appealing.

The final decision will be made by the people traveling; this only helps make things a little clearer. I would potentially want to look at different areas outside of Los Angeles, and maybe look at specific neighborhoods rather than a broad city.