# Exploring the Food Vibe of LA neighborhoods

1. Introduction

## Background

Los Angeles is well known for having diverse types of food. Often, different cuisine concentrated in different neighborhoods. For example, Sawtelle is known for having many Japanese restaurants, Koreatown has a high concentration of Korean restaurants, and North Hollywood often has lots of food trucks.

## Problem

With so many different neighborhoods in Los Angeles area, foodies need to know which neighborhoods are enriched in what kinds of cuisine and which neighborhoods are similar in their food options.

## 1.3 Target

Restaurateurs looking to open new food venues would be interested in the kind of cuisine represented in the neighborhood. Foodies living near LA or visiting LA would also be interested in this information.

## 1.4 Approach

I am going to use FourSquare API to obtain information about food venues in the different neighborhoods in the downtown Los Angeles area. In order to do that, I need the neighborhood names, longitude, and latitude of the neighborhood. I will present the different

1. Data Acquisition and Cleaning

## 2.1 Data Requirement

For each of the neighborhood in the downtown area, I need the zip code, neighborhood name, longitude, and latitude in order to use the FourSquare API to obtain a list of food venues in that neighborhood. The main information I need from the list of food venues is the type of cuisine the venue belongs to.

## 2.2 Data Source

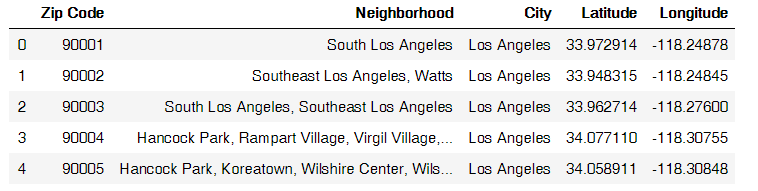
The zip code and name of the neighborhood in Los Angeles County can be obtained through [LA almanac](http://www.laalmanac.com/employment/em12c.php). The longitude and latitude of all areas in California can be obtained through [OpenDataSoft](https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=CA).

## 2.3 Data Cleaning

Data obtained from LA almanac is converted to a pandas dataframe with zipcode, name of neighborhood, and the median income. The median income column was dropped. The neighborhood column has information about whether the neighborhood is in downtown Los Angeles or elsewhere in the county. To focus only on downtown Los Angeles, all rows with neighborhoods outside of the Los Angeles downtown were dropped.

The OpenDataSoft data is converted into a pandas dataframe. We combined it with the neighborhood and zipcode dataframe by using inner join on zip code. The resulting dataframe has zipcode, neighborhood, city, latitude and the longitude.

Table 1. Snapshot of the initial dataset, which contains neighborhood, latitude and longitude information that can be inputted into the FourSquare API. There are 97 rows and 5 columns.



From the table, the neighborhood name, latitude, and longitude are used to make markers on Folium map.

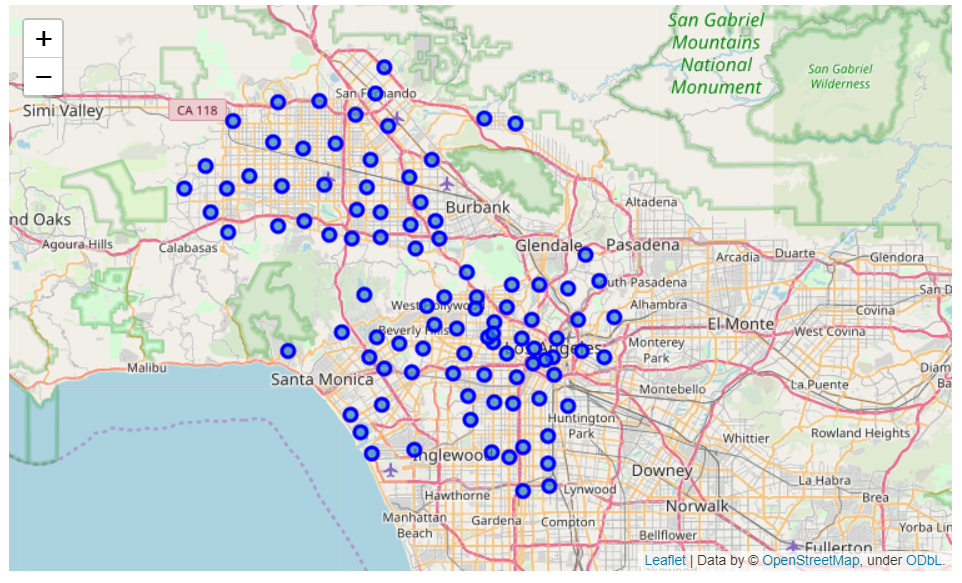


Figure 1. Folium map marking the 97 neighborhoods in Los Angeles area in blue circles.

This information is passed through FourSquare API selecting only for food venues by using category ID of ‘4d4b7105d754a06374d81259’, which encompasses all food venues. I set a limit of 100 per neighborhood and a radius of 500m. A list of up to 100 food venues in each neighborhood is received, containing information about the category of food each belongs to.

There are many categories that are very specific and can be classified into a broader category of cuisine, as the “General Category” column. For example, donburi restaurant, donburi bowl, ramen restaurant, and sushi restaurant all could be classify as Japanese restaurant. Additionally, fast food and similar ubiquitous stores such as donut shops were excluded from the dataset since they do not represent any cuisine.

Neighborhoods with less than 5 venues were dropped from the dataset, and general categories with less than 3 venues were also dropped. This resulted in 76 neighborhoods and 29 general categories of cuisine. The general category was one-hot encoded. Using groupby function, the percentage of each cuisine in the neighborhood was calculated and used as features in the K-mean clustering.

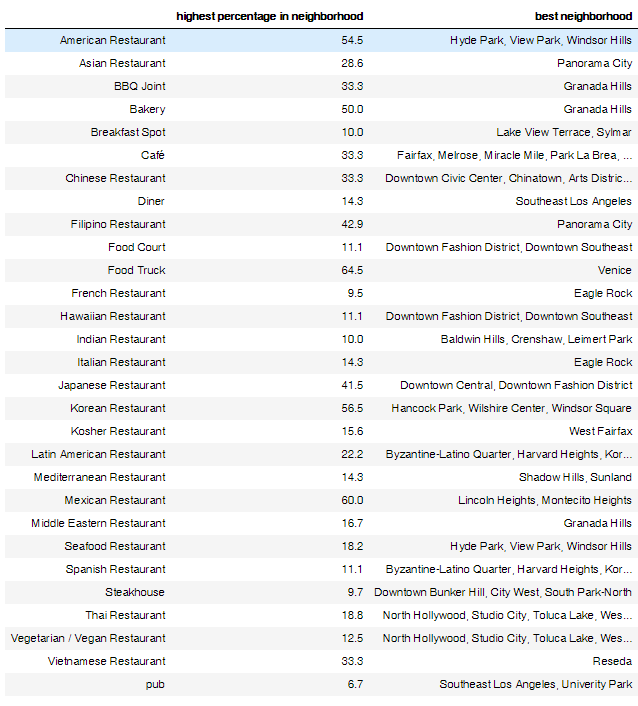
1. Methodology

I performed some exploratory data analysis with obtained results and to check against the prevailing association of certain neighborhood with certain type of cuisine as a sanity check.

## 3.1 The top neighborhood for each food category and the top food category in each neighborhood

First, I found the neighborhood with the highest percentage of certain type of food. For example, as shown in the table below, the highest concentration of food truck is located in Venice, which is well-known for having a large amounts of food trucks along the beach. Also, although Koreatown did not shown up as having the highest percentage of Korean restaurant, Hancock Park, Wilshire Center, and Winsor Square are right next to Koreatown. Another promising result is that Reseda, a small Vietnamese community, indeed has an large percentages of restaurant in the area.

Table 2. The best neighborhood to have certain type of food and the corresponding percentage of that cuisine in the neighborhood



Second, to confirm that the data makes sense, I looked at the top restaurant and the frequency (out of 1) of the top 5 general category of food in each neighborhood. Sawtelle, which known for having lots of hip Japanese restaurant, indeed show that 39% of all food venues are Japanese restaurant.

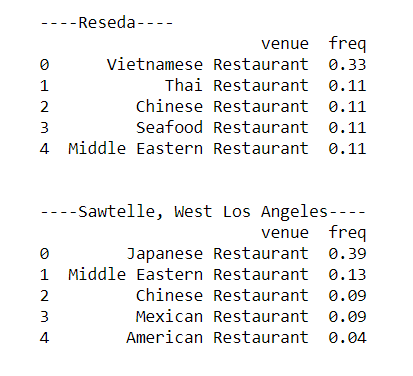
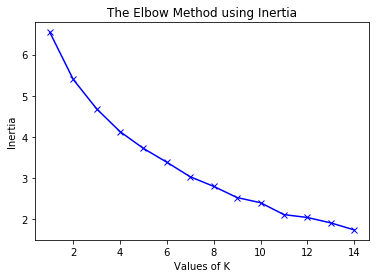
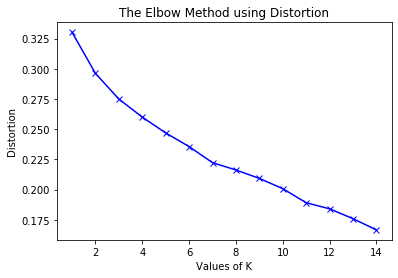


Figure 2. A snapshot of the tables showing the top 5 cuisine in 3 of the 53 neighborhoods.

Third, just to quickly scan through the

## 3.2 K-mean clustering

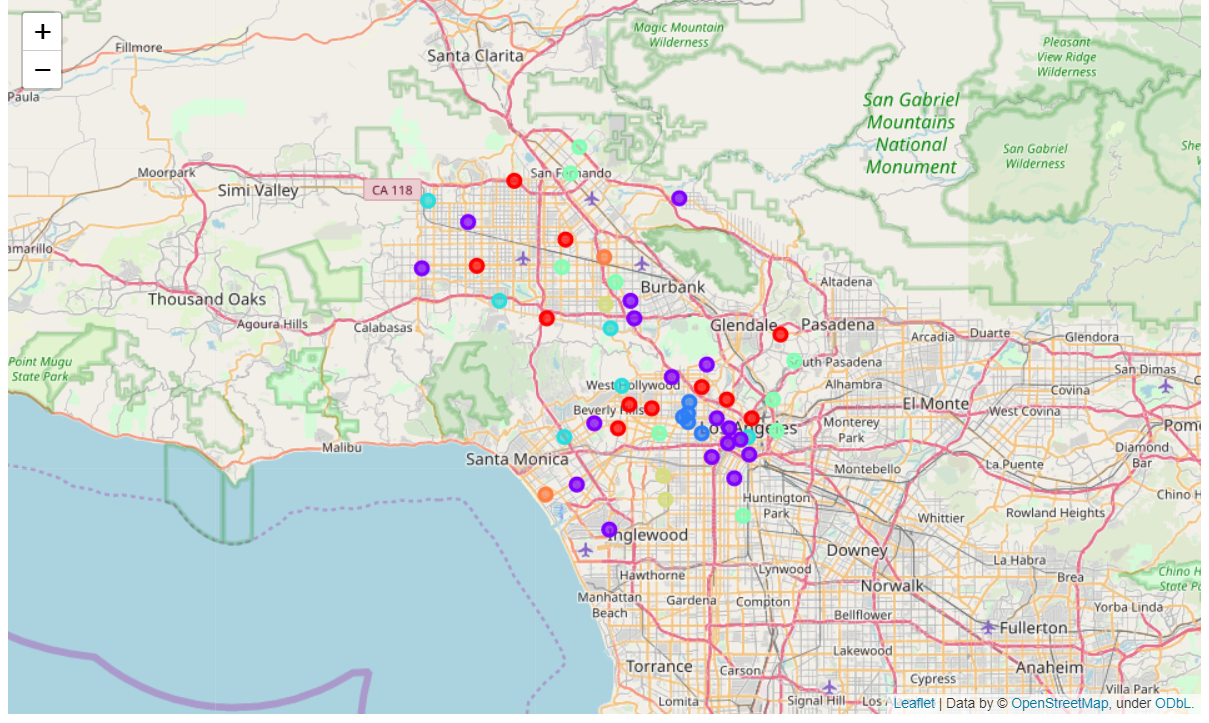
K-mean clustering was used to cluster the 53 neighborhoods into k different groups. The features used were the frequencies of the cuisine in each neighborhood. To pick the appropriate k to use, I tried using the elbow method, which is to plot the distortion or inertia against a range of k.



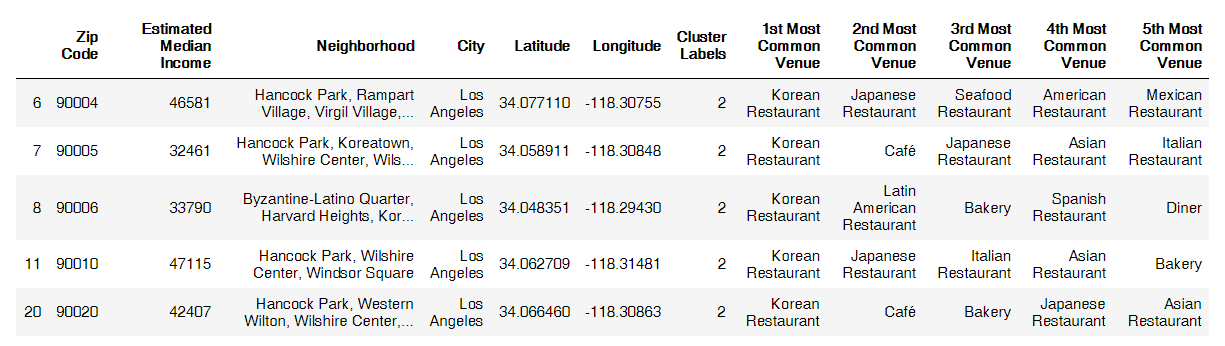
However, it doesn’t look like there is an elbow. Instead, I picked k manually through manual inspection of the clusters. A k of 7 looked to be the most meaningful.

1. Results

I created a Folium map and colored the different clusters with different colors. Visually inspecting the map, there were separation of the clusters regionally. Noticeably, there is a blue cluster in the central area.



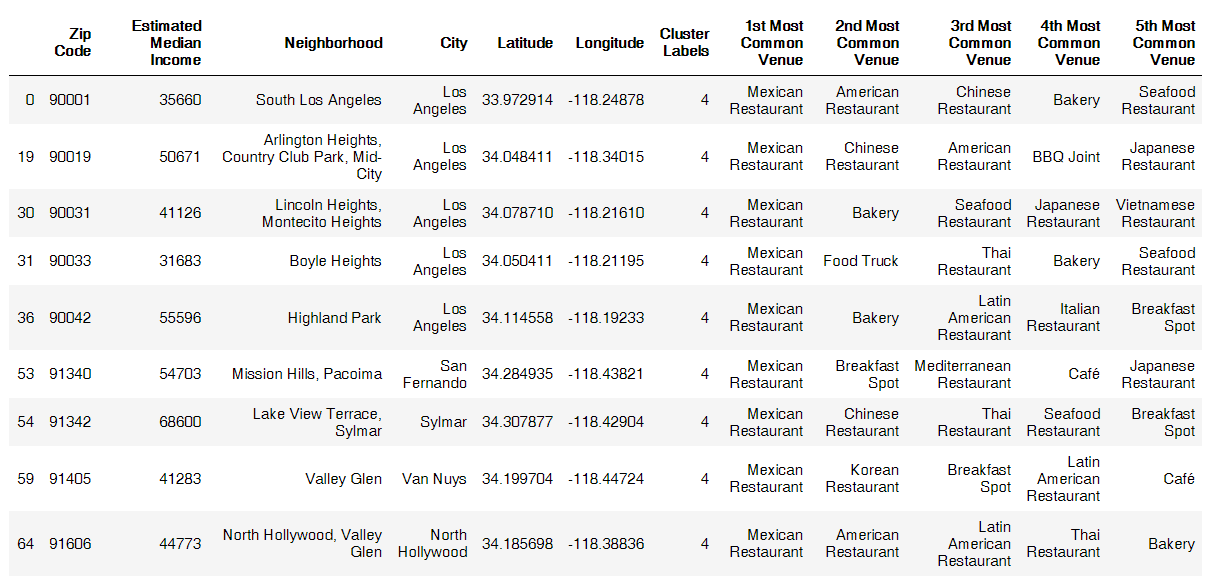
Next, manual inspection of all of the neighborhoods in each cluster is performed. For example, the blue cluster is cluster 2, which is the area around Korea town.



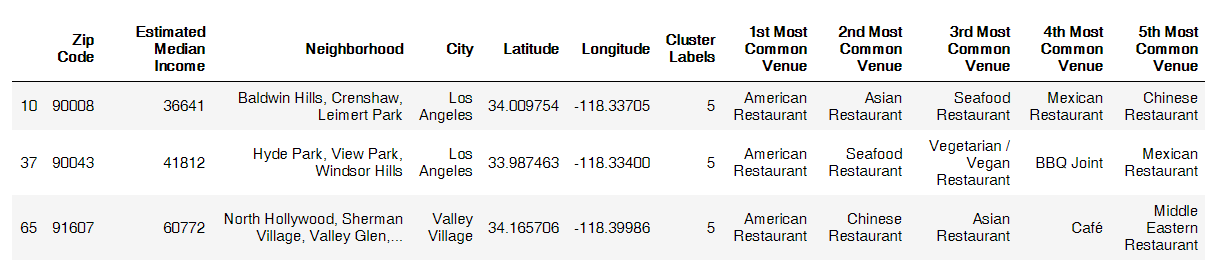
Cluster 3, in teal, which is highly enriched in Japanese restaurant, include Sawtelle.



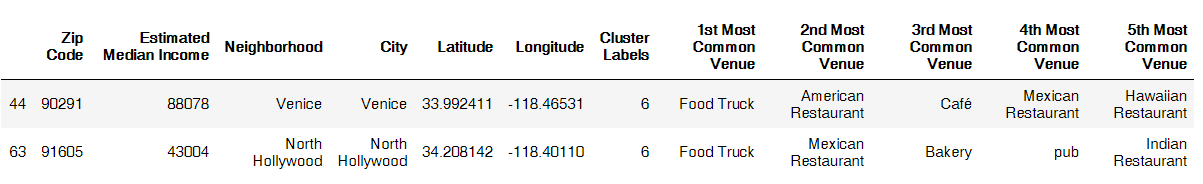
Cluster 4, in green, looks to feature Mexican restaurant in as the most common food venues.



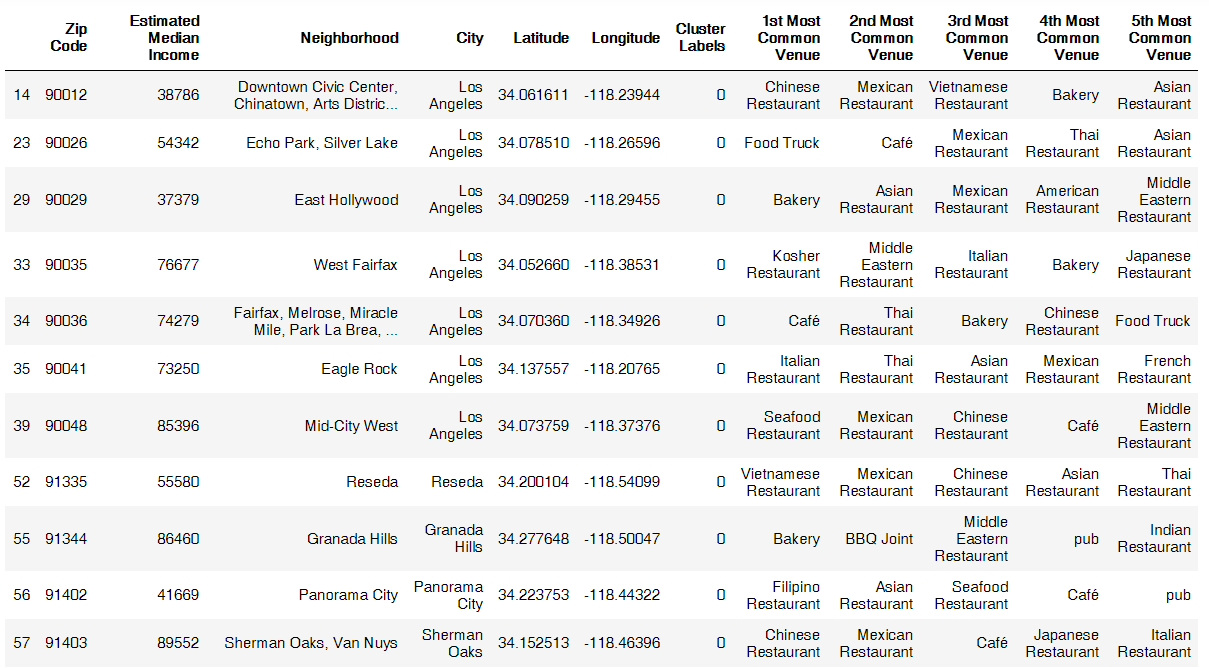
Cluster 5, in brown yellow, is most enriched in American restaurants.



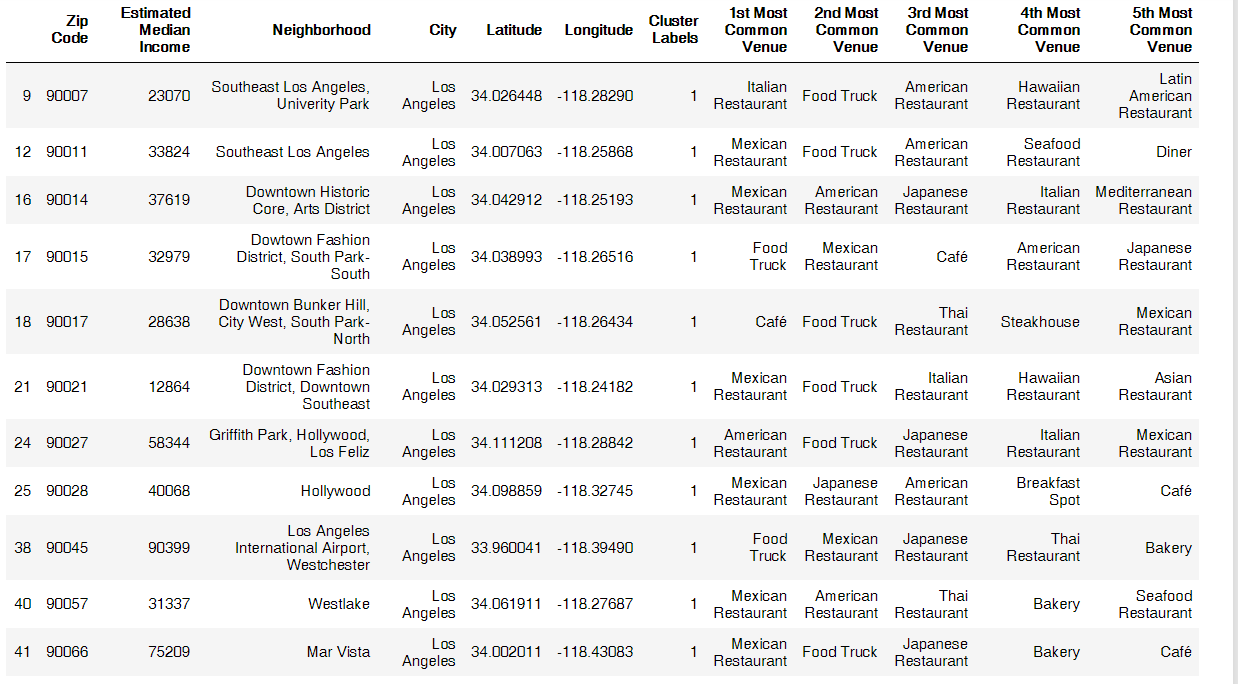
Cluster 6, in orange, is enriched in food truck, which includes Venice.



For cluster 0, in red, it seems like a random mix of neighborhood with different food groups, but featuring certain cuisine in certain neighborhood. For example, Resada with Vietnamese restaurant, Panorama City with Filipino restaurant.



Cluster 1, in purple, like cluster 0, also looked like a random mix of neighborhood with different cuisine combination.



1. Discussion

K-mean clustering clustered the 53 neighborhoods into 7 different clusters. These clusters used the frequency of each cuisine in the neighborhood as features. This results in each clusters having different “food vibe”. For example, the blue clusters features Korean restaurant, which encompasses Koreatown.

More exciting is discovering the food vibe of other neighborhoods that are similar to neighborhood that was well known for certain food.

For example, Sawtelle is known for Japanese restaurant. However, suppose someone had already tried all the restaurant of Sawtelle and wanted to explore a new neighborhood, but with similar “food vibe” as Sawtelle. They could look at the other neighborhood in the teal clusters, which include Encino and Chatsworth. From the red cluster, foodies could also find neighborhoods with enrichment in bakery, such as Granada Hills.

Similarly, a food truck that is doing well in Venice could also explore sending a truck to North Hollywood, which belongs to the same cluster as Venice.

1. Conclusion

In this study, I explored the food atmosphere of neighborhoods in Los Angeles area by collecting data about the type of cuisine in each neighborhood. I clustered 54 neighborhoods into 7 different clusters using K-mean clustering. The 7 clusters represent neighbors featuring Korean, Japanese, Mexican, American, food truck, and two clusters of miscellaneous cuisine. This information could help foodie explore new neighborhood depending on their food taste preference.