Exploring the Food Vibe of LA neighborhoods

1. Introduction

1.1. 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.

1.2. 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

2. 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. The longitude and latitude of all areas in California can be obtained through OpenDataSoft.

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.

	Zip Code	Neighborhood	City	Latitude	Longitude
0	90001	South Los Angeles	Los Angeles	33.972914	-118.24878
1	90002	Southeast Los Angeles, Watts	Los Angeles	33.948315	-118.24845
2	90003	South Los Angeles, Southeast Los Angeles	Los Angeles	33.962714	-118.27600
3	90004	Hancock Park, Rampart Village, Virgil Village,...	Los Angeles	34.077110	-118.30755
4	90005	Hancock Park, Koreatown, Wilshire Center, Wils...	Los Angeles	34.058911	-118.30848

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

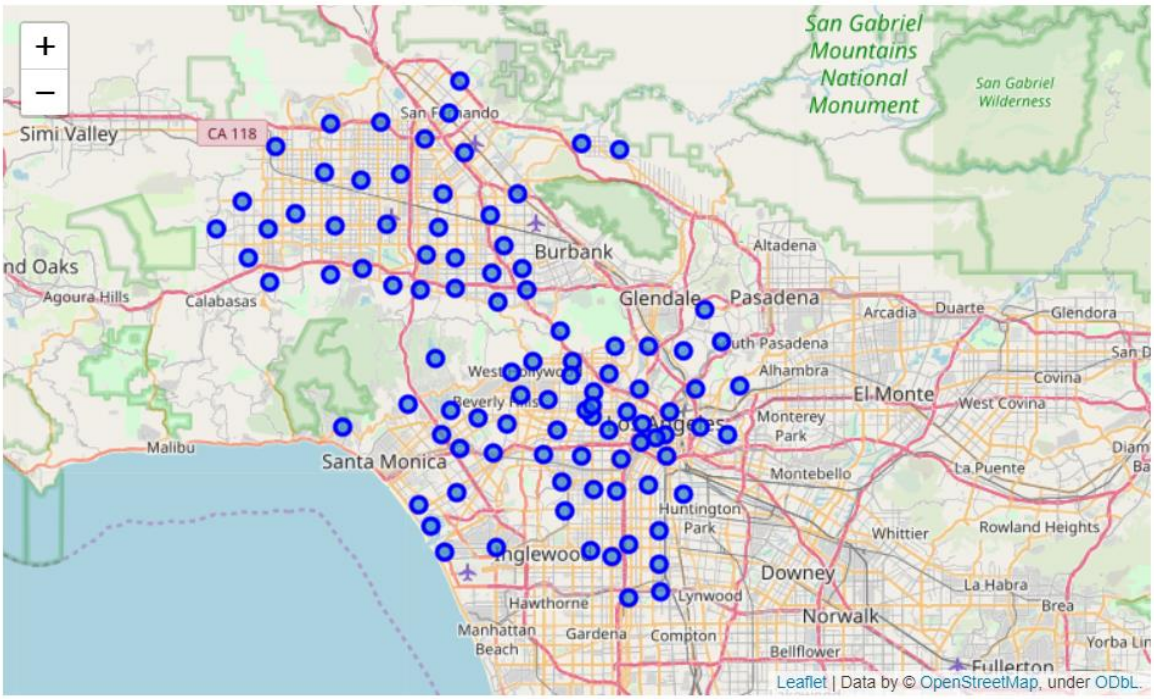


Figure 1. Folium map marking the 97 neighborhoods in Los Angeles area in blue circles.

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.

3. 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

	highest percentage in neighborhood	best neighborhood
American Restaurant	54.5	Hyde Park, View Park, Windsor Hills
Asian Restaurant	28.6	Panorama City
BBQ Joint	33.3	Granada Hills
Bakery	50.0	Granada Hills
Breakfast Spot	10.0	Lake View Terrace, Sylmar
Café	33.3	Fairfax, Melrose, Miracle Mile, Park La Brea, ...
Chinese Restaurant	33.3	Downtown Civic Center, Chinatown, Arts Distric...
Diner	14.3	Southeast Los Angeles
Filipino Restaurant	42.9	Panorama City
Food Court	11.1	Downtown Fashion District, Downtown Southeast
Food Truck	64.5	Venice
French Restaurant	9.5	Eagle Rock
Hawaiian Restaurant	11.1	Downtown Fashion District, Downtown Southeast
Indian Restaurant	10.0	Baldwin Hills, Crenshaw, Leimert Park
Italian Restaurant	14.3	Eagle Rock
Japanese Restaurant	41.5	Downtown Central, Downtown Fashion District
Korean Restaurant	56.5	Hancock Park, Wilshire Center, Windsor Square
Kosher Restaurant	15.6	West Fairfax
Latin American Restaurant	22.2	Byzantine-Latino Quarter, Harvard Heights, Kor...
Mediterranean Restaurant	14.3	Shadow Hills, Sunland
Mexican Restaurant	60.0	Lincoln Heights, Montecito Heights
Middle Eastern Restaurant	16.7	Granada Hills
Seafood Restaurant	18.2	Hyde Park, View Park, Windsor Hills
Spanish Restaurant	11.1	Byzantine-Latino Quarter, Harvard Heights, Kor...
Steakhouse	9.7	Downtown Bunker Hill, City West, South Park-North
Thai Restaurant	18.8	North Hollywood, Studio City, Toluca Lake, Wes...
Vegetarian / Vegan Restaurant	12.5	North Hollywood, Studio City, Toluca Lake, Wes...
Vietnamese Restaurant	33.3	Reseda
pub	6.7	Southeast Los Angeles, Univerity Park

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.

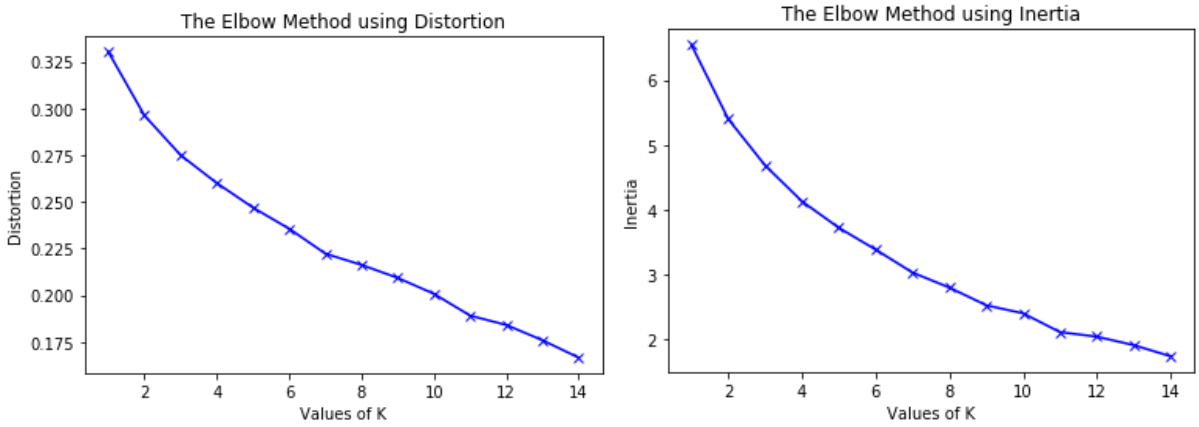
```
----Reseda----
                        venue  freq
0      Vietnamese Restaurant  0.33
1            Thai Restaurant  0.11
2         Chinese Restaurant  0.11
3         Seafood Restaurant  0.11
4  Middle Eastern Restaurant  0.11


----Sawtelle, West Los Angeles----
                        venue  freq
0        Japanese Restaurant  0.39
1  Middle Eastern Restaurant  0.13
2         Chinese Restaurant  0.09
3         Mexican Restaurant  0.09
4        American Restaurant  0.04
```

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

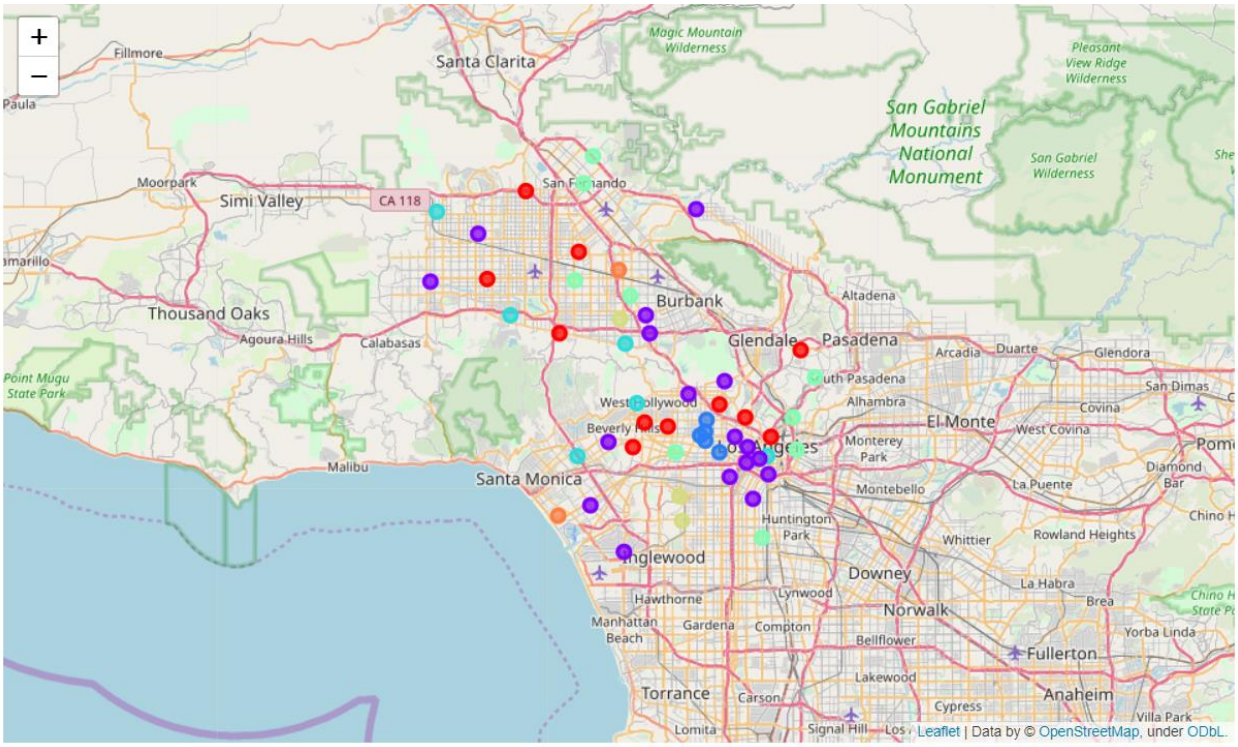
K-mean clustering was used to cluster the 53 neighborhoods into k different groups. The features used were the frequencies of the cuisine in each neighborhood. To pick the appropriate k to use, I tried using the elbow method, which is to plot the distortion or inertia against a range of k.



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.

4. 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.

	Zip Code	Estimated Median Income	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
6	90004	46581	Hancock Park, Rampart Village, Virgil Village,...	Los Angeles	34.077110	-118.30755	2	Korean Restaurant	Japanese Restaurant	Seafood Restaurant	American Restaurant	Mexican Restaurant
7	90005	32461	Hancock Park, Koreatown, Wilshire Center, Wils...	Los Angeles	34.058911	-118.30848	2	Korean Restaurant	Café	Japanese Restaurant	Asian Restaurant	Italian Restaurant
8	90006	33790	Byzantine-Latino Quarter, Harvard Heights, Kor...	Los Angeles	34.048351	-118.29430	2	Korean Restaurant	Latin American Restaurant	Bakery	Spanish Restaurant	Diner
11	90010	47115	Hancock Park, Wilshire Center, Windsor Square	Los Angeles	34.062709	-118.31481	2	Korean Restaurant	Japanese Restaurant	Italian Restaurant	Asian Restaurant	Bakery
20	90020	42407	Hancock Park, Western Wilton, Wilshire Center,...	Los Angeles	34.066460	-118.30863	2	Korean Restaurant	Café	Bakery	Japanese Restaurant	Asian Restaurant

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

	Zip Code	Estimated Median Income	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
15	90013	22808	Downtown Central, Downtown Fashion District	Los Angeles	34.044662	-118.24255	3	Japanese Restaurant	Asian Restaurant	Korean Restaurant	Food Truck	Seafood Restaurant
22	90025	78713	Sawtelle, West Los Angeles	Los Angeles	34.045006	-118.44527	3	Japanese Restaurant	Middle Eastern Restaurant	Mexican Restaurant	Chinese Restaurant	pub
43	90069	86403	Hollywood, Melrose	West Hollywood	34.090975	-118.38130	3	Japanese Restaurant	American Restaurant	Italian Restaurant	Café	French Restaurant
47	91311	82738	Chatsworth	Chatsworth	34.259052	-118.59426	3	Japanese Restaurant	Mexican Restaurant	Thai Restaurant	Breakfast Spot	Chinese Restaurant
48	91316	68720	Encino	Encino	34.168753	-118.51636	3	Japanese Restaurant	American Restaurant	Middle Eastern Restaurant	Bakery	Italian Restaurant
62	91604	99745	North Hollywood, Studio City	Studio City	34.143856	-118.39429	3	Japanese Restaurant	Mexican Restaurant	American Restaurant	Breakfast Spot	Food Truck

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

	Zip Code	Estimated Median Income	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	90001	35660	South Los Angeles	Los Angeles	33.972914	-118.24878	4	Mexican Restaurant	American Restaurant	Chinese Restaurant	Bakery	Seafood Restaurant
19	90019	50671	Arlington Heights, Country Club Park, Mid-City	Los Angeles	34.048411	-118.34015	4	Mexican Restaurant	Chinese Restaurant	American Restaurant	BBQ Joint	Japanese Restaurant
30	90031	41126	Lincoln Heights, Montecito Heights	Los Angeles	34.078710	-118.21610	4	Mexican Restaurant	Bakery	Seafood Restaurant	Japanese Restaurant	Vietnamese Restaurant
31	90033	31683	Boyle Heights	Los Angeles	34.050411	-118.21195	4	Mexican Restaurant	Food Truck	Thai Restaurant	Bakery	Seafood Restaurant
36	90042	55596	Highland Park	Los Angeles	34.114558	-118.19233	4	Mexican Restaurant	Bakery	Latin American Restaurant	Italian Restaurant	Breakfast Spot
53	91340	54703	Mission Hills, Pacoima	San Fernando	34.284935	-118.43821	4	Mexican Restaurant	Breakfast Spot	Mediterranean Restaurant	Café	Japanese Restaurant
54	91342	68600	Lake View Terrace, Sylmar	Sylmar	34.307877	-118.42904	4	Mexican Restaurant	Chinese Restaurant	Thai Restaurant	Seafood Restaurant	Breakfast Spot
59	91405	41283	Valley Glen	Van Nuys	34.199704	-118.44724	4	Mexican Restaurant	Korean Restaurant	Breakfast Spot	Latin American Restaurant	Café
64	91606	44773	North Hollywood, Valley Glen	North Hollywood	34.185698	-118.38836	4	Mexican Restaurant	American Restaurant	Latin American Restaurant	Thai Restaurant	Bakery

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

	Zip Code	Estimated Median Income	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	90008	36641	Baldwin Hills, Crenshaw, Leimert Park	Los Angeles	34.009754	-118.33705	5	American Restaurant	Asian Restaurant	Seafood Restaurant	Mexican Restaurant	Chinese Restaurant
37	90043	41812	Hyde Park, View Park, Windsor Hills	Los Angeles	33.987463	-118.33400	5	American Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	BBQ Joint	Mexican Restaurant
65	91607	60772	North Hollywood, Sherman Village, Valley Glen,...	Valley Village	34.165706	-118.39986	5	American Restaurant	Chinese Restaurant	Asian Restaurant	Café	Middle Eastern Restaurant

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

	Zip Code	Estimated Median Income	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
44	90291	88078	Venice	Venice	33.992411	-118.46531	6	Food Truck	American Restaurant	Café	Mexican Restaurant	Hawaiian Restaurant
63	91605	43004	North Hollywood	North Hollywood	34.208142	-118.40110	6	Food Truck	Mexican Restaurant	Bakery	pub	Indian Restaurant

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.

	Zip Code	Estimated Median Income	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
14	90012	38786	Downtown Civic Center, Chinatown, Arts Distric...	Los Angeles	34.061611	-118.23944	0	Chinese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Bakery	Asian Restaurant
23	90026	54342	Echo Park, Silver Lake	Los Angeles	34.078510	-118.26596	0	Food Truck	Café	Mexican Restaurant	Thai Restaurant	Asian Restaurant
29	90029	37379	East Hollywood	Los Angeles	34.090259	-118.29455	0	Bakery	Asian Restaurant	Mexican Restaurant	American Restaurant	Middle Eastern Restaurant
33	90035	76677	West Fairfax	Los Angeles	34.052660	-118.38531	0	Kosher Restaurant	Middle Eastern Restaurant	Italian Restaurant	Bakery	Japanese Restaurant
34	90036	74279	Fairfax, Melrose, Miracle Mile, Park La Brea, ...	Los Angeles	34.070360	-118.34926	0	Café	Thai Restaurant	Bakery	Chinese Restaurant	Food Truck
35	90041	73250	Eagle Rock	Los Angeles	34.137557	-118.20765	0	Italian Restaurant	Thai Restaurant	Asian Restaurant	Mexican Restaurant	French Restaurant
39	90048	85396	Mid-City West	Los Angeles	34.073759	-118.37376	0	Seafood Restaurant	Mexican Restaurant	Chinese Restaurant	Café	Middle Eastern Restaurant
52	91335	55580	Reseda	Reseda	34.200104	-118.54099	0	Vietnamese Restaurant	Mexican Restaurant	Chinese Restaurant	Asian Restaurant	Thai Restaurant
55	91344	86460	Granada Hills	Granada Hills	34.277648	-118.50047	0	Bakery	BBQ Joint	Middle Eastern Restaurant	pub	Indian Restaurant
56	91402	41669	Panorama City	Panorama City	34.223753	-118.44322	0	Filipino Restaurant	Asian Restaurant	Seafood Restaurant	Café	pub
57	91403	89552	Sherman Oaks, Van Nuys	Sherman Oaks	34.152513	-118.46396	0	Chinese Restaurant	Mexican Restaurant	Café	Japanese Restaurant	Italian Restaurant

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

	Zip Code	Estimated Median Income	Neighborhood	City	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	90007	23070	Southeast Los Angeles, Univerity Park	Los Angeles	34.026448	-118.28290	1	Italian Restaurant	Food Truck	American Restaurant	Hawaiian Restaurant	Latin American Restaurant
12	90011	33824	Southeast Los Angeles	Los Angeles	34.007063	-118.25868	1	Mexican Restaurant	Food Truck	American Restaurant	Seafood Restaurant	Diner
16	90014	37619	Downtown Historic Core, Arts District	Los Angeles	34.042912	-118.25193	1	Mexican Restaurant	American Restaurant	Japanese Restaurant	Italian Restaurant	Mediterranean Restaurant
17	90015	32979	Dowtown Fashion District, South Park-South	Los Angeles	34.038993	-118.26516	1	Food Truck	Mexican Restaurant	Café	American Restaurant	Japanese Restaurant
18	90017	28638	Downtown Bunker Hill, City West, South Park-North	Los Angeles	34.052561	-118.26434	1	Café	Food Truck	Thai Restaurant	Steakhouse	Mexican Restaurant
21	90021	12864	Downtown Fashion District, Downtown Southeast	Los Angeles	34.029313	-118.24182	1	Mexican Restaurant	Food Truck	Italian Restaurant	Hawaiian Restaurant	Asian Restaurant
24	90027	58344	Griffith Park, Hollywood, Los Feliz	Los Angeles	34.111208	-118.28842	1	American Restaurant	Food Truck	Japanese Restaurant	Italian Restaurant	Mexican Restaurant
25	90028	40068	Hollywood	Los Angeles	34.098859	-118.32745	1	Mexican Restaurant	Japanese Restaurant	American Restaurant	Breakfast Spot	Café
38	90045	90399	Los Angeles International Airport, Westchester	Los Angeles	33.960041	-118.39490	1	Food Truck	Mexican Restaurant	Japanese Restaurant	Thai Restaurant	Bakery
40	90057	31337	Westlake	Los Angeles	34.061911	-118.27687	1	Mexican Restaurant	American Restaurant	Thai Restaurant	Bakery	Seafood Restaurant
41	90066	75209	Mar Vista	Los Angeles	34.002011	-118.43083	1	Mexican Restaurant	Food Truck	Japanese Restaurant	Bakery	Café

5. 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.

6. 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.