Using Foursquare API and K-mean Clustering algorithm to choose the right business location

Introduction

For location based businesses especially in the retail sector, the choice of where it operate is critically important. After all **The three most important things in retailing are—location, location and location".** And if they get the location wrong, it can have a serious and often disastrous effect on the business as the the former boss of Marks and Spencer (Lord Sieff) said describing the main success factors in his business.

In this notebook we'll describe the problem and the challenges facing businesses when making location decisions, we'll describe also how can we benefit from data in such situations, we'll have a look next to an example of a data-driven choices in business location and marketing.

Table of Contents

Business problem

Location decisions are usually pretty important—to both large and small businesses. The location decision has a direct effect on an **operation's costs** as well as its **ability to serve customers** (and therefore its revenues).

Also, location decisions, once made, are difficult and costly to undo. The costs of moving an operation are often significant and run the risk of inconveniencing customers and staff. It is always best to get the location decision right first time.

The main aim of choosing a business location is to achieve a balance between three related objectives:

- The costs of the operation
- The customer service that the business wants to provide
- The potential revenues that can be achieved from the location

One of the most important factors that influence the choice of business location is the demand factor which mainly affect customer service and revenues. So in order to choose right we have to consider the following points:

- Customer convenience

Probably the most important factor. Many businesses need to be located where customers find it quick, easy and cheap to access the service being provided. E.g. a fast-food outlet needs to be somewhere close to a strong customer footfall, not hidden away out of sight. Out-of-town retail parks are situated within a convenient short drive from major population centers.

- Site suitability

A site may need to have some particular characteristics to maximize customer satisfaction and revenues. E.g. a luxury restaurant or hotel needs to be located somewhere that customers find attractive—not in the middle of a trading estate.

- Image

This is more intangible, but often important. Some customers associate a product with a certain area and prefer to buy from there (e.g. walking equipment—a business based in the Lake District might enjoy a better perceived reputation)

- Competition

The proximity to other competing businesses could be crucial to the business success. Could they provide a benefit to our business or cause a hindrance? Establishing which competitors are in the selected area and their offering could help guarantee to choose the right location for the business. If there is too much competition then it may be a warning sign to expand our horizons to a new location. There are exceptions to this such as car dealerships who want to be near each other as customers compare and choose the best car deal, hence their close proximity. Likewise, if we have an element in our offering that is unique or offers some kind of new innovation, then choosing an area that already has a ripe market could be the ideal way to pick up customers very quickly and establish a presence in a new area in a relatively short time frame.

- Expansion potential

Future production capacity often has to be taken into account. A location might tick many other boxes, but if it provides limited scope for expansion then it might be rejected. If a location restricts output, then revenues are potentially damaged.

How data can be used to solve the problem

There are 7.7 billion people in the world, with at least 3.5 billion of us online. This means social media platforms are used by one-in-three people in the world, and more than two-thirds of all internet users generating a huge amount of data,

Social media has changed the world and businesses should use it in their favor. The rapid and vast adoption of these technologies is changing how we find partners, how we access information from the news, and how we organize to demand political change.

Individuals increasingly use smartphones to access social media and use location-based services to tell their friends where they are and Businesses are reaching consumers in ways we’ve never seen before. More brands are measuring real world foot traffic patterns. They’re launching targeted mobile advertising campaigns. They’re even analyzing consumer movement patterns to make important business decisions and uncover new market insights. Behind the scenes, brand marketers

are putting an incredible amount of trust in location data providers to give them the most accurate, up-to-date datasets possible.

For new and established businesses alike, Geo-spatial Insights makes it easier to apply location data analysis to business intelligence. Location insights provide the keys to unlocking in-depth understanding of consumers' behaviors and preferences. Monitor competitor performance, analyze site selection, learn about customer offline behavior and more.

Location data providers power the vast majority of mobile targeting strategies that brand marketers implement today. One of the most well-known firms in the location data market is Foursquare, it now runs an independent location data and technology platform for marketers and developers. Using Foursquare's location data, brands can quantify the impact of their media (using validated stops) and find new audiences that they can target with relevant advertising, Foursquare also offers a Places API and a Pilgrim SDK, which enable location discovery inside apps and websites, as well as real-time location detection.

Description of the project

During our analysis we used the neighborhoods names and their geo-spatial coordinates found in the downloaded csv file from https://usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr .

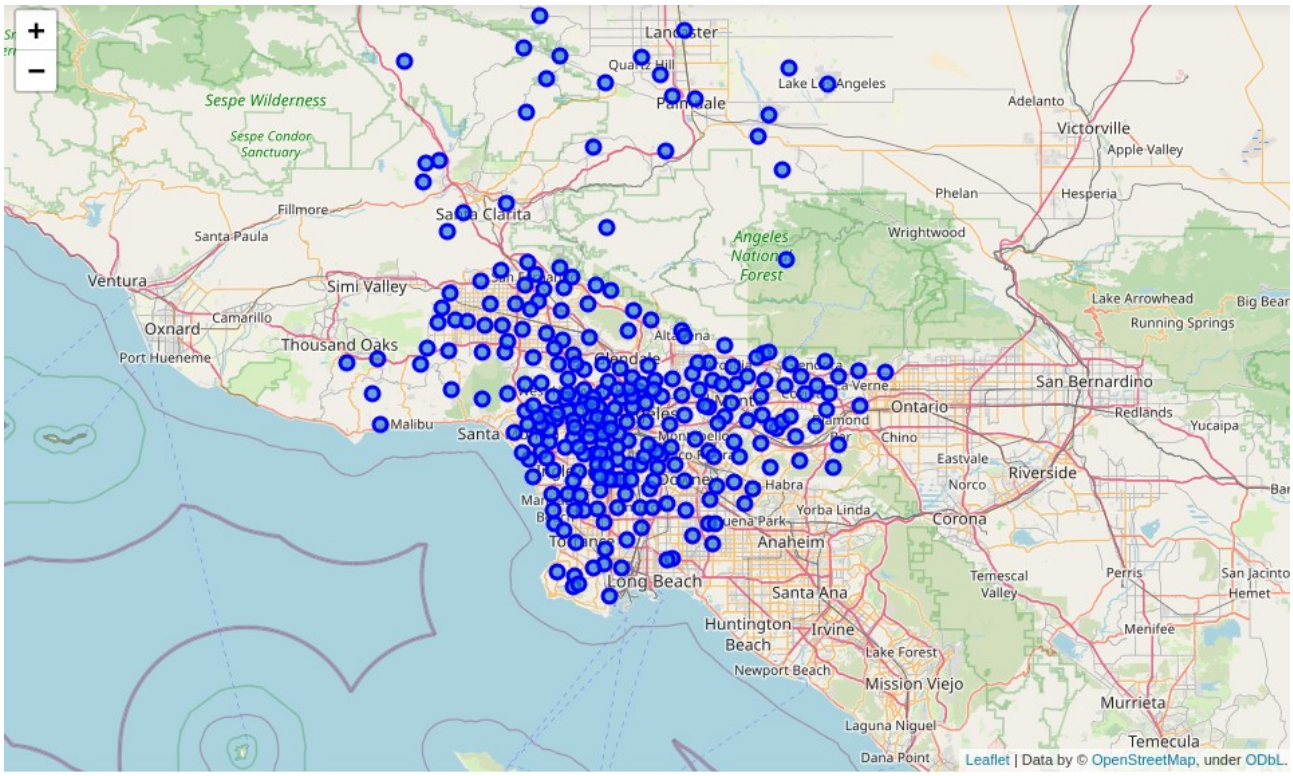
	Neighborhood	Latitude	Longitude
0	Acton	34.497355	-118.169810
1	Adams-Normandie	34.031461	-118.300208
2	Agoura Hills	34.146736	-118.759885
3	Agua Dulce	34.504927	-118.317104
4	Alhambra	34.085539	-118.136512

Let's check the data types and null values

```
la_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 272 entries, 0 to 271
Data columns (total 3 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Neighborhood  272 non-null    object 
 1   Latitude      272 non-null    float64
 2   Longitude     272 non-null    float64
dtypes: float64(2), object(1)
memory usage: 6.5+ KB
```

This data set contain 272 neighborhoods, to explore them We used folium library to visualize the map of Los Angeles city and its neighborhoods .



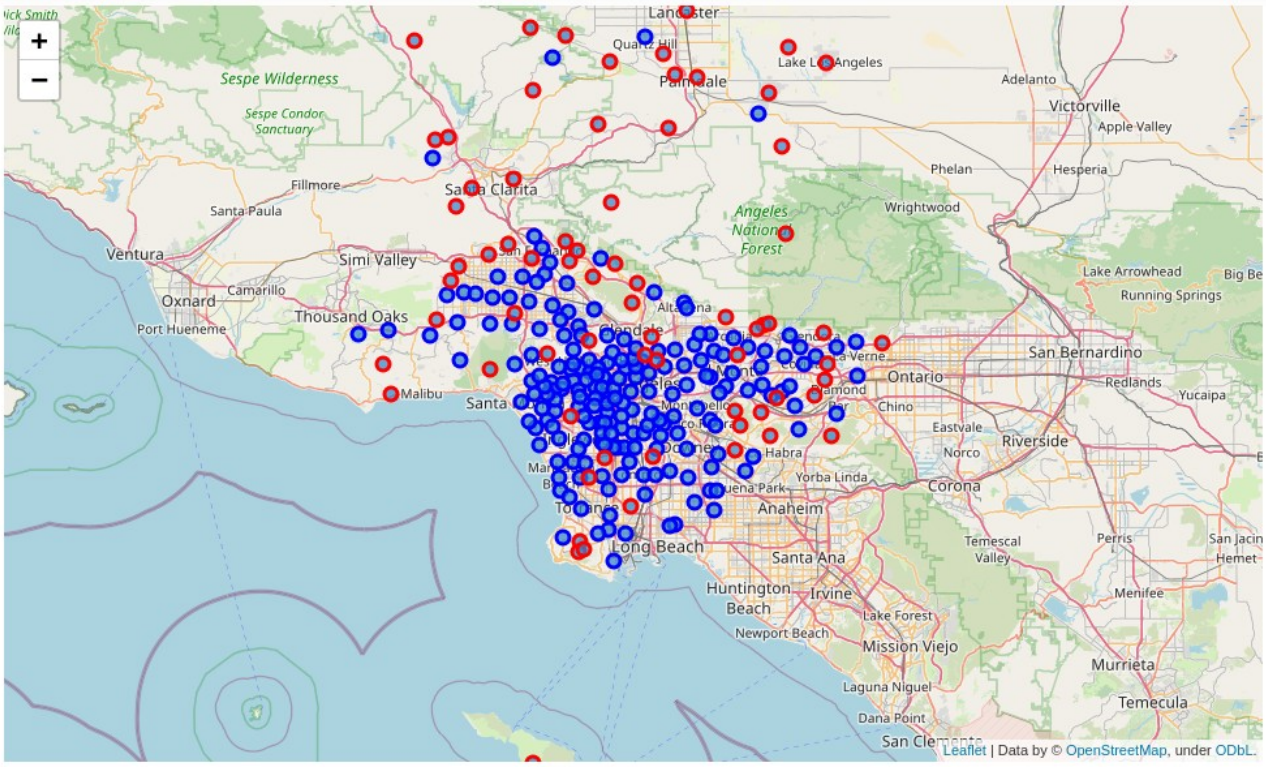
Using the DataFrame containing location coordinates, we wrote a function to leverage the Foursquare API to get info about venues in each neighborhood.

We were interested in data about the 'food' category, within a radius of 500 M, once finished we converted the resulting Json file into a DataFrame that contains 2100 venues and the following columns :

'Neighborhood', 'Neighborhood Latitude', 'Neighborhood Longitude', 'Venue', 'Venue Latitude', 'Venue Longitude', 'Venue Category'

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adams-Normandie	34.031461	-118.300208	Orange Door Sushi	34.032485	-118.299368	Sushi Restaurant
1	Adams-Normandie	34.031461	-118.300208	Sushi Delight	34.032445	-118.299525	Sushi Restaurant
2	Adams-Normandie	34.031461	-118.300208	Little Xian	34.032292	-118.299465	Sushi Restaurant
3	Adams-Normandie	34.031461	-118.300208	Tacos La Estrella	34.032230	-118.300757	Taco Place
4	Adams-Normandie	34.031461	-118.300208	El Rincon	34.032298	-118.299478	Latin American Restaurant

Out of 272, we received results for 201 neighborhoods, the rest (marked with red dots) didn’t had any.



Using One Hot encoding we converted the 95 categories to columns and then grouped the resulting DataFrame by neighborhood.

	Neighborhood	American Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bistro	Brazilian Restaurant	...	Taco Place	Taiwanese Restaurant	Tapas Restaurant	Thai Restaurant
0	Adams-Normandie	0	0	0	0	0	0	0	0	0	...	1	0	0	0
1	Agoura Hills	2	0	0	0	1	0	3	0	0	...	0	0	0	1
2	Alhambra	0	0	0	0	0	1	0	0	0	...	0	0	0	0
3	Altadena	0	0	0	0	0	0	1	0	0	...	0	0	0	0
4	Arcadia	0	0	0	0	0	0	1	0	0	...	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
196	Willowbrook	0	0	0	0	0	0	0	0	0	...	0	0	0	0
197	Wilmington	0	0	0	0	0	0	0	0	0	...	0	0	0	0
198	Windsor Square	0	0	0	0	0	0	0	0	0	...	0	0	0	0
199	Winnetka	0	0	0	0	0	0	1	0	0	...	1	0	0	0
200	Woodland Hills	0	0	1	0	0	0	1	0	0	...	0	0	0	0

201 rows × 96 columns

We used this data to get insights about the top 5 venues in each neighborhood.

```
----Adams-Normandie----
                       venue  freq
0            Sushi Restaurant     3
1                  Taco Place     1
2                        Food     1
3  Latin American Restaurant     1
4   Mediterranean Restaurant     0


----Agoura Hills----
                  venue  freq
```

```
0  Fast Food Restaurant    4
1                Bakery    3
2   American Restaurant    2
3      Sushi Restaurant    2
4    Mexican Restaurant    2


----Alhambra----
                  venue freq
0            Bagel Shop    1
1           Pizza Place    1
2        Breakfast Spot    1
3  Fast Food Restaurant    1
4    Mexican Restaurant    1
```

.

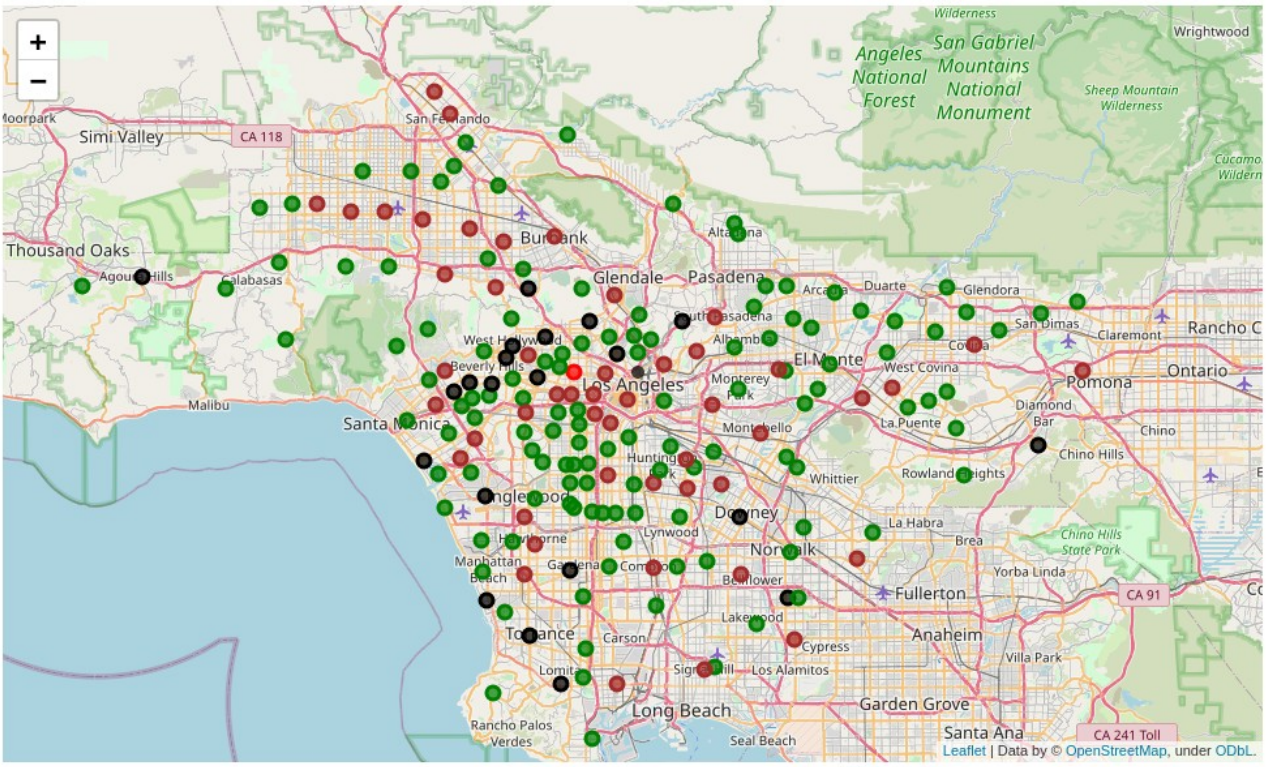
.

.

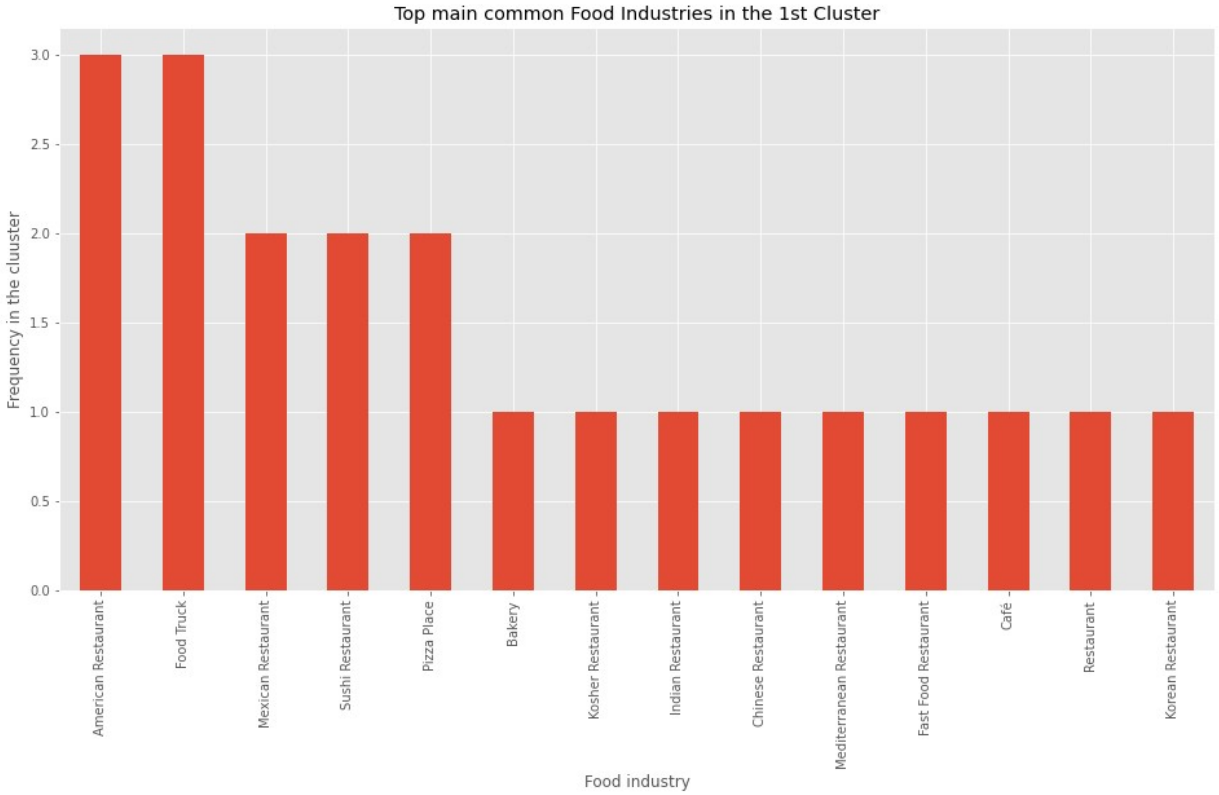
We put the all this in a new DataFrame, but with the top 10 venues instead, the result looks like this:

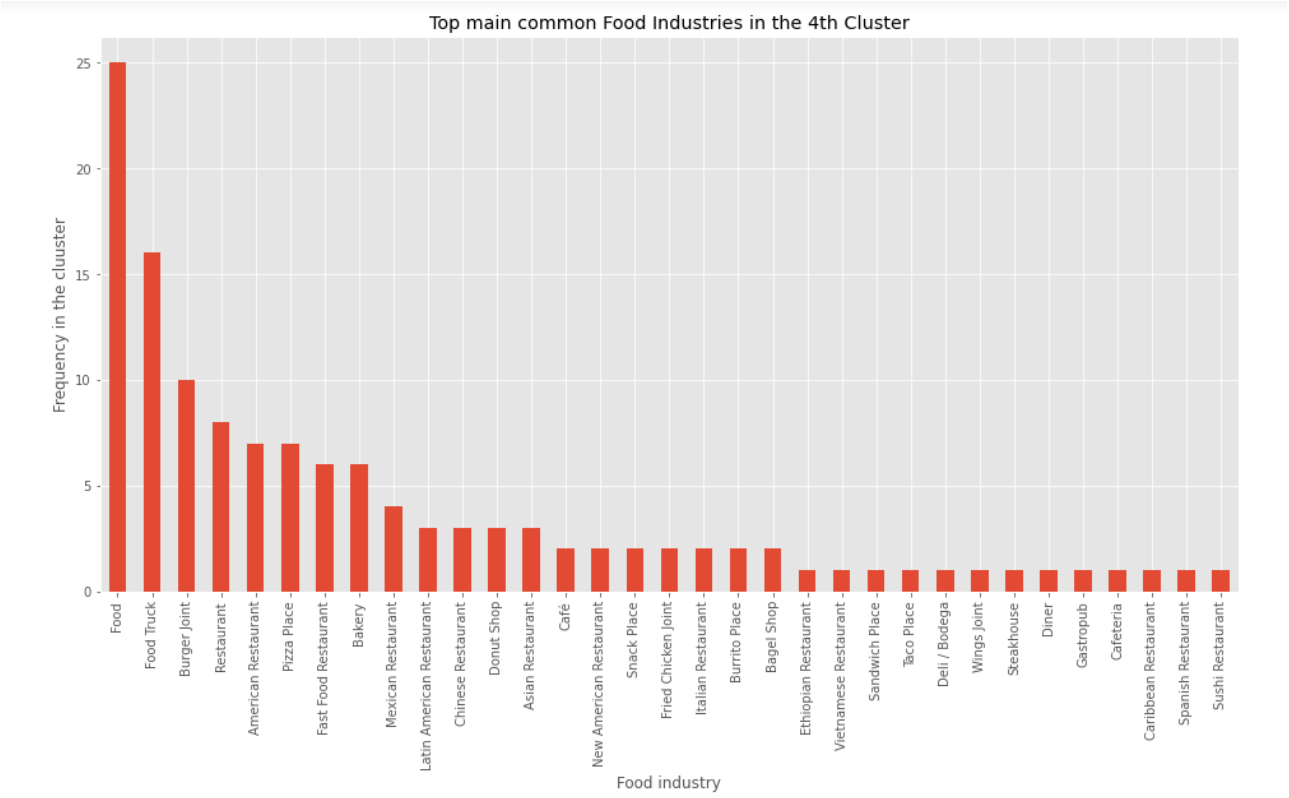
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adams-Normandie	Sushi Restaurant	Food	Taco Place	Latin American Restaurant	Wings Joint	Ethiopian Restaurant	Doner Restaurant	Dongbei Restaurant	Donut Shop	Dosa Place
1	Agoura Hills	Fast Food Restaurant	Bakery	American Restaurant	Breakfast Spot	Chinese Restaurant	Sushi Restaurant	Mexican Restaurant	Snack Place	BBQ Joint	Burger Joint
2	Alhambra	Fast Food Restaurant	Bagel Shop	Mexican Restaurant	Café	Breakfast Spot	Pizza Place	Food Stand	English Restaurant	Dongbei Restaurant	Donut Shop
3	Altadena	Bakery	Food	Sushi Restaurant	Wings Joint	Falafel Restaurant	Doner Restaurant	Dongbei Restaurant	Donut Shop	Dosa Place	Dumpling Restaurant
4	Arcadia	Food	Fast Food Restaurant	Donut Shop	Hotpot Restaurant	Thai Restaurant	Bakery	Chinese Restaurant	Italian Restaurant	Japanese Restaurant	Mexican Restaurant

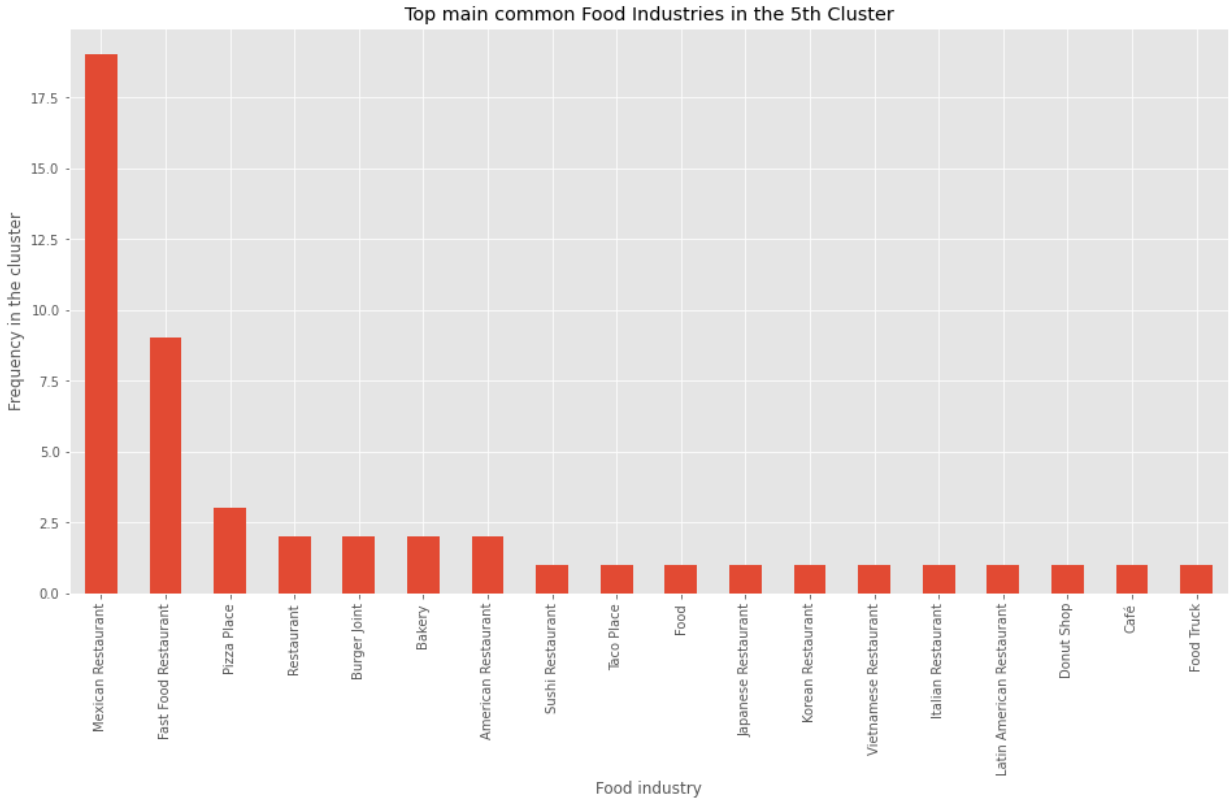
Finally we were ready to use the k-mean algorithm to cluster our neighborhoods into 5 distinct clusters.



We found that there is two small clusters counting one neighborhood each namely china, and korea town where the top venue is their local dishes.
The three other clusters are characterized as follow:







Conclusion

The Asian towns are famous by their local dishes, The rest of neighborhoods where we can find food related businesses are mainly grouped in three big clusters, and each

neighborhood is different than another in way that people visit this neighborhood for a a certain type of food while visiting another one for another type.

Apart from the 71 neighborhoods that should be avoided because they don't have any food related venues, We identified four neighborhoods, each with a distinct food cuisine however Final decision on optimal food industry and location will be made by stakeholders based on specific characteristics of neighborhoods and locations in the three recommended clusters or the four recommended neighborhoods taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.