April 22,2020

Setting up a Greek Restaurant in Los Angeles County

Shayan Riyaz

Note: This project is a capstone project and is part of the Coursera IBM Data Science Certification.

Introduction

Background

Recent Demographic surveys have shown that the Los Angeles Metropolitan area is. One of the most densely packed regions in the world. Interestingly due to its popularity, it is also one of the most culturally diverse. With a high number of tourists from all over the world landing at Los Angeles International Airport (LAX) along with the growing number of settlements in the region. Throughout the years we have seen a strong increase in ethnic food in the county itself which correlates with the increase in a diverse population. Studies from the past years have shown that even though the rise in ethnic food is increasing. There is also a disparity in the types of ethnic food [1]. This project aims to introduce a planned method to introduce a Greek restaurant in the Los Angeles County area.

The reason for choosing a Greek Restaurant is mainly because of the low numbers present in the region at this point and also the fact that it stands as a good investment due to it being considered a part of Mediterranean cuisine.

When deciding the most prominent location for a Greek Restaurant is it important to consider some important factors:

Demand

For this problem, we can either choose a region that has seen its fair share of ethnic restaurants or a region that has does not hold a diverse portfolio of Ethic restaurants

Location

To have an effective restaurant business, it is always a beneficial step to set it up in a densely populated location. By observing metropolitan demographics in general. There are strong consensus and statistical evidence that shows population density is higher closer to the downtown area of a city.

Rent Price

To make sure the new business can financially be at an advantage we will choose the are that has the cheapest rent.

April 22,2020

Data and Resources:

To achieve our target, we use the following sources and services:

BeautifulSoup

This Wikipedia page will be the source of identifying all the neighborhoods that exist in the Los Angeles area: List of Districts and Neighborhoods of Los Angeles[1]

GeoPy

We will be using GeoPy to map our neighborhoods. This API allows us to locate the precise coordinates of a location using numerous data sources.

Folium

To map our coordinates on a map that can be visualized, we will be using the folium API.

Foursquare API

The Foursquare API [2] gives us access to the most popular locations in a given area. Using the above-mentioned services and the API we can pin-point the restaurant demographics in the Neighborhood

Rent Data

Los Angeles Rent Prices[3]

Methodology

Setup

The first and foremost step taken for this project is to collect a list of all the Neighborhoods of Los Angeles. To achieve this the most helpful source was Wikipedia itself. Therefore, we used BeautifulSoup to scrape that data. Some neighborhoods which weren't prominent parts of the Los Angeles area were removed from the scraping. Getting the names is enough but the next important step is to be able to visualize those neighborhoods on a map to determine their exact locations. For this, we used GeoPy Nominatim a geolocator that reads the list of Neighborhood names and found their respective geo coordinates. Next, we used folium to map out the exact coordinates.

The next step taken was to analyze the neighborhoods and see the activity present in the respective areas. We took advantage of the Foursquare API. The API allows us to see the most popular venues in a given location. The venues are then analyzed in detail to find existing patterns. For example, coffee shop density and recreational activity venues in a given area.

These patterns are more visible when we group neighborhoods in a cluster. Therefore, we will be using K-Means clustering to find suitable clusters to our goal. The “n” number of clusters will be analyzed, and we will be choosing the cluster that fits our desired goal.
Note: One method of picking out the best cluster is to see the spread of restaurants.

After the cluster is selected, we will further analyze it and find the number of Greek Restaurants in the area. This helps us further shortlist our possible locations. As mentioned above we will be keeping in context the distance from the Center of Los Angeles. As a bonus, to ensure the restaurant can be financially stable we will pick the area with the lowest rent price.

Note: It is very important to understand that when it comes to setting up a restaurant strategically there are a lot of other factors involved. This project serves to target the fundamentals.

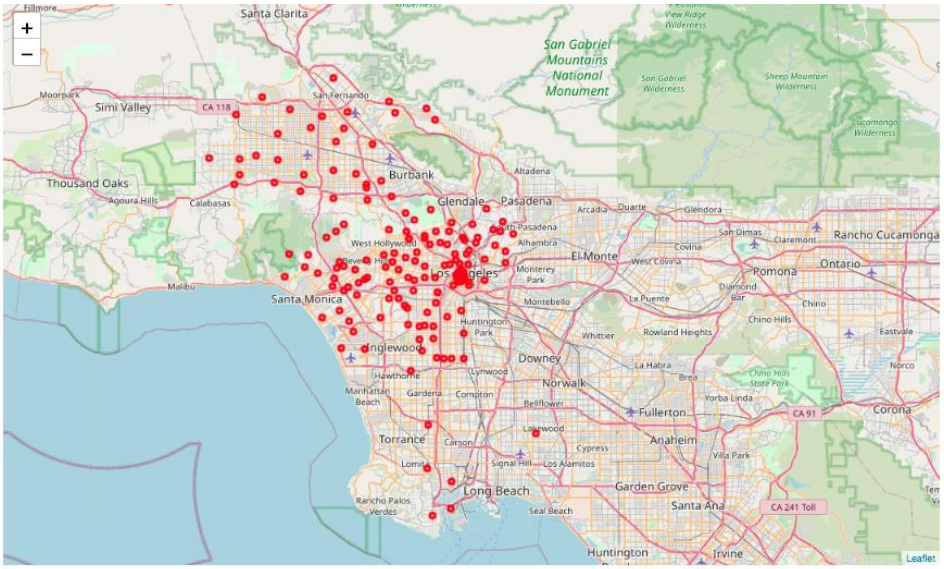
Data Analysis

To get a good list of the Neighborhoods in Los Angeles we will be using Wikipedia, “List of districts and neighborhoods of Los Angeles”[1]. Using BeautifulSoup and GeoPy we get a list of 189 Neighborhoods and their coordinates. This has become our finalized neighborhood Data Frame before we further process it using Foursquare:

	Neighbourhood	Latitude	Longitude
0	Angelino Heights	34.070289	-118.254796
1	Arleta	34.241327	-118.432205
2	Arlington Heights	34.043494	-118.321374
3	Arts District	34.041239	-118.234450
4	Atwater Village	34.118698	-118.262392
5	Baldwin Hills	34.010989	-118.337071
6	Baldwin Village	34.019456	-118.345910
7	Baldwin Vista	0.000000	0.000000

Using the coordinates and Folium we project these coordinates on a geographic map of Los Angeles with the markers in red point towards each location.

April 22,2020



The next step is to use the previously acquired credentials and get access to the Foursquare API. As previously mentioned, the API provides us with location data of all the venues in their respective neighborhoods. We can get access, to restaurant names, the category, and the category counts to understand the commercial demographics better. Over here we can see the most popular venues from the venue ‘Angelino Heights’.

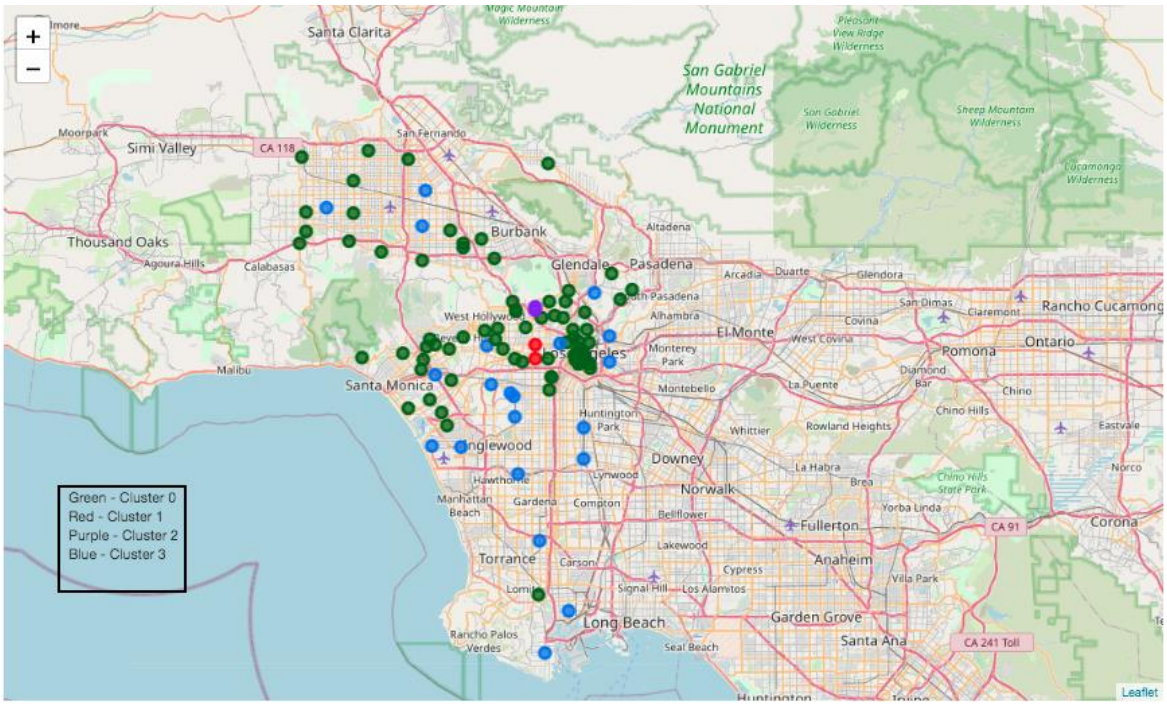
	name	categories	lat	lng
0	Halliwell Manor	Performing Arts Venue	34.069329	-118.254165
1	Guisados	Taco Place	34.070262	-118.250437
2	Eightfold Coffee	Coffee Shop	34.071245	-118.250698
3	Subliminal Projects	Art Gallery	34.072290	-118.250737
4	The Park's Finest BBQ	BBQ Joint	34.066519	-118.254291
5	Michael Jackson's "Thriller" House (and Tree)	Historic Site	34.069557	-118.254599
6	Tsubaki	Japanese Restaurant	34.072938	-118.251298
7	Ototo	Sake Bar	34.074399	-118.254016
8	Bar Henry	Cocktail Bar	34.069062	-118.250465
9	K Bakery	Bakery	34.066124	-118.254059
10	Bob's Market	Convenience Store	34.067812	-118.251939
11	The Park Restaurant	Breakfast Spot	34.074243	-118.252242
12	Guisados	Taco Place	34.071178	-118.250273
13	Domino's Pizza	Pizza Place	34.073528	-118.251187
14	Super 8	Motel	34.072722	-118.250899
15	Knights Inn	Hotel	34.066201	-118.253954
16	Tacos La Movidita (aka Bellevue Steakhouse)	Steakhouse	34.069535	-118.258367

The next step is to create clusters. Using the K-means clustering algorithm we check for the value of K clusters. This was done by using the Silhouette Coefficient Method.

April 22,2020

```
For n_clusters=2, The Silhouette Coefficient is 0.045133170928705 31
For n_clusters=3, The Silhouette Coefficient is 0.063218455674422 77
For n_clusters=4, The Silhouette Coefficient is 0.086515895280830 7
For n_clusters=5, The Silhouette Coefficient is 0.063700797363638 25
For n_clusters=6, The Silhouette Coefficient is 0.074809002828773 87
For n_clusters=7, The Silhouette Coefficient is 0.035968485401956 34
For n_clusters=8, The Silhouette Coefficient is 0.027853774594633 67
For n_clusters=9, The Silhouette Coefficient is 0.037168090701327 254
For n_clusters=10, The Silhouette Coefficient is 0.040589532831130 26
For n_clusters=11, The Silhouette Coefficient is 0.064703513971861 8
```

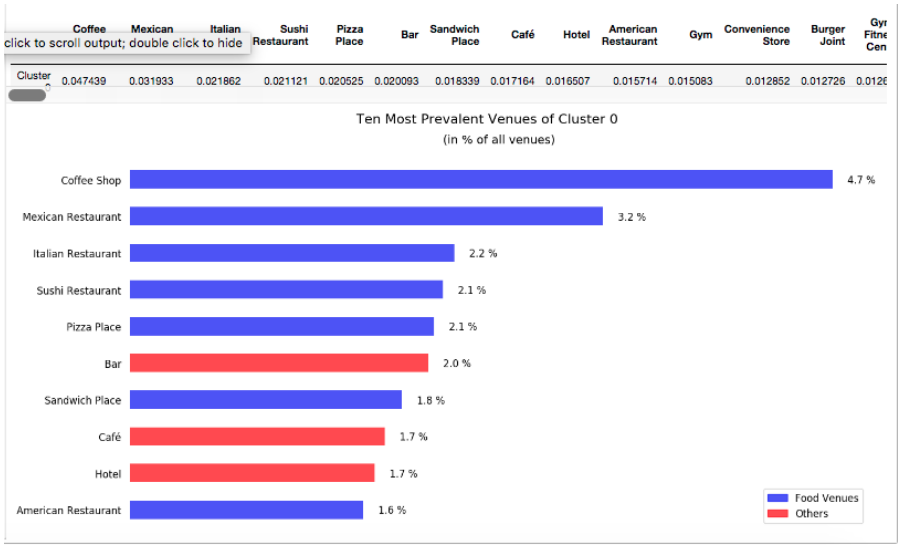
As we can see above K=4 achieved the best coefficient therefore we use this value as our K-cluster. Now we can easily separate the neighborhoods by clusters using Folium and our K value. We will also be able to see the top 10 venues for each neighborhood.



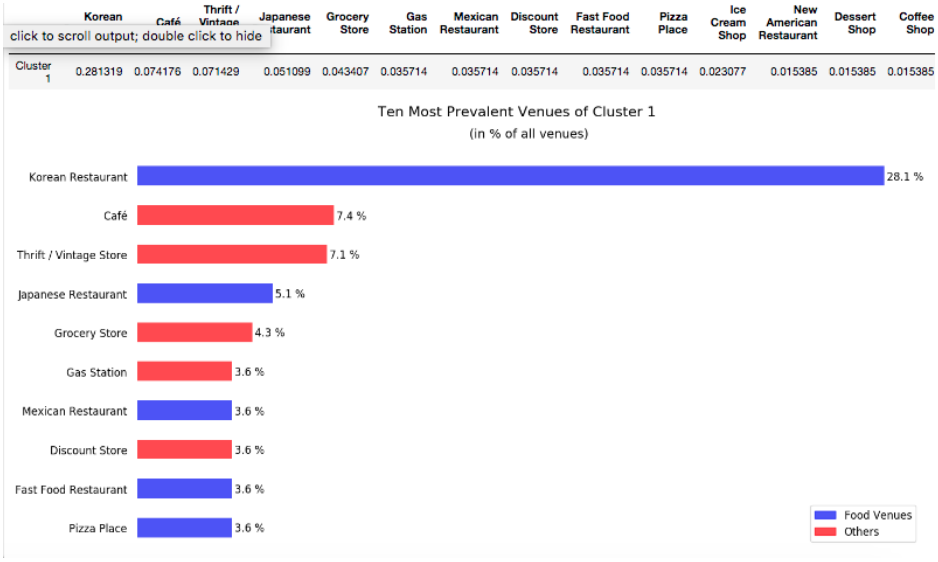
	Neighbourhood	Latitude	Longitude	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Co
0	Angelino Heights	34.070289	-118.254796	0	Taco Place	Hotel	Park	Mexican Restaurant	Breakfast Spot	Boutique	Steakhouse	Market	Clothing Store	Cockt
2	Arlington Heights	34.043494	-118.321374	0	Shop & Service	Latin American Restaurant	Rental Car Location	Grocery Store	Restaurant	Food	Convenience Store	Art Gallery	Pizza Place	
3	Arts District	34.041239	-118.234450	0	Art Gallery	Coffee Shop	Italian Restaurant	Event Space	Pop-Up Shop	Gun Range	Cocktail Bar	Beer Garden	Clothing Store	Mediterr Resta
4	Atwater Village	34.118698	-118.262392	0	Restaurant	Coffee Shop	Liquor Store	Vietnamese Restaurant	Pizza Place	Cosmetics Shop	Nightclub	Donut Shop	Taco Place	San
5	Baldwin Hills	34.010989	-118.337071	3	Fast Food Restaurant	Mexican Restaurant	Lingerie Store	Sandwich Place	Department Store	Southern / Soul Food Restaurant	Supplement Shop	Frozen Yogurt Shop	Fried Chicken Joint	Bre

April 22,2020

To further examine the clusters, we will observe the 10 venues in each cluster using a bar plot. This will help us see what the density of the venues are in a given cluster.



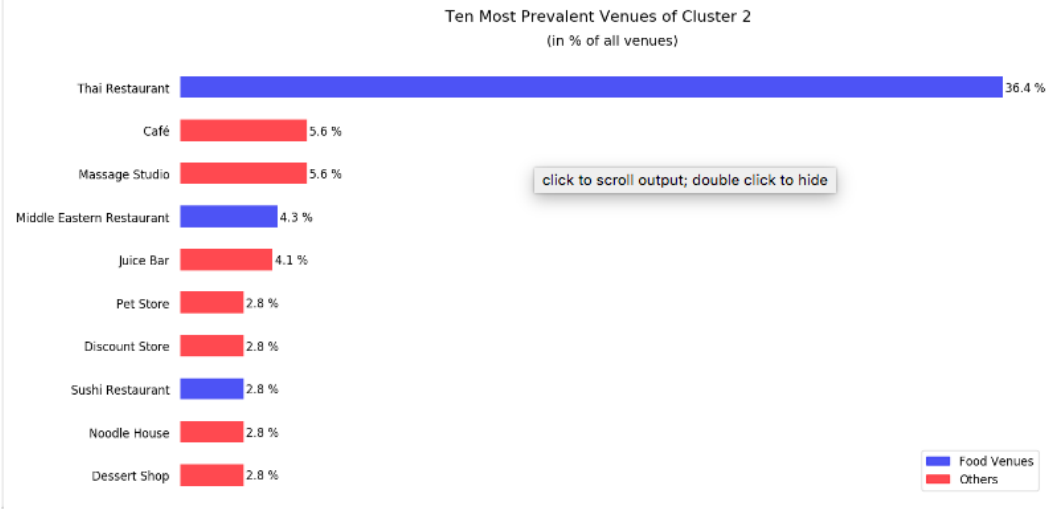
There are around 7 food places in Cluster 0. This concludes that Cluster 0 might not be the best option for a new and upcoming restaurant due to the already saturated food market.



Cluster 1 has 4 food venues with Korean Restaurants making a huge majority. This is expected since Cluster 1 includes Korea Town. However, it can be noted that there is not much variety in terms of food places therefore it can be a suitable option.

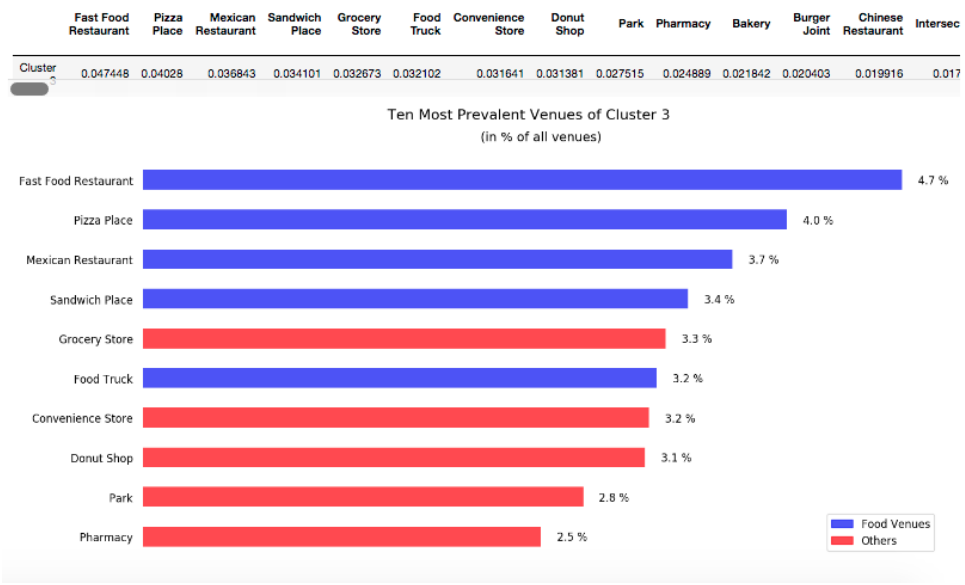
April 22,2020

	Thai Restaurant	Café	Massage Studio	Middle Eastern Restaurant	Juice Bar	Pet Store	Discount Store	Sushi Restaurant	Noodle House	Dessert Shop	Strip Club	Bar	Comedy Club	Marijuana Dispensary
Cluster 2	0.363636	0.055944	0.055944	0.043124	0.040793	0.027972	0.027972	0.027972	0.027972	0.027972	0.027972	0.027972	0.027972	0.027972



Cluster 2 is similar to Cluster 1, with a specific cuisine dominating the area but has a lesser number of food venues in the top 10

	Fast Food Restaurant	Pizza Place	Mexican Restaurant	Sandwich Place	Grocery Store	Food Truck	Convenience Store	Donut Shop	Park	Pharmacy	Bakery	Burger Joint	Chinese Restaurant	Intersec
Cluster 3	0.047448	0.04028	0.036843	0.034101	0.032673	0.032102	0.031641	0.031381	0.027515	0.024889	0.021842	0.020403	0.019916	0.017



Cluster 3 is the least likely of the 4 mainly due to food venues accumulating most of the top 10 %. Therefore, it also among the least likely options

April 22,2020

Cluster 1 proved to have lesser restaurants compared to the other clusters and has a lack of diverse food. Therefore, I chose it as my primary cluster and region for setting a restaurant The next step is to observe the neighborhoods in this cluster. Fortunately, in our case, there are 2 locations.

	Neighbourhood	Latitude	Longitude
0	Harvard Heights	34.047111	-118.305483
1	Koreatown	34.061734	-118.305373

We will also observe the restaurants in these locations.

Harvard Heights

	name	categories	distance	lat	lng
0	Papa Cristo's Catering & Greek Taverna	Greek Restaurant	538	34.047377	-118.299649
1	Ali Ranas Baklava and Gyro cart	Greek Restaurant	310	34.049600	-118.307000
2	Hidden LA's Big Fat Greek Tuesday	Greek Restaurant	530	34.047355	-118.299735

Koreatown

	name	categories	distance	lat	lng
0	Sizzling Kabob	Greek Restaurant	434	34.061834	-118.300668

In this case, We will choose to go with a location that has more Greek Restaurants. This is because the data shows that the population in Harvard Heights is more accustomed to Greek Restaurants and is, therefore, a less risky location. In terms of choosing the Neighborhood closest to the city center, we can observe that there is not much difference in distance for both Neighborhoods.

	Neighbourhood	Latitude	Longitude	Distance from LA center (in km)
0	Harvard Heights	34.047111	-118.305483	5.824201
1	Koreatown	34.061734	-118.305373	5.836395

The last check is to compare the rent prices. Unfortunately, the dataset available does not contain rent rates for Korea Town. However, Harvard Heights has one of the lowest rental prices in all of Los Angeles.

	Neighbourhood	Average Rent
15	Harvard Heights	$1,607

April 22,2020

Results

Discussion

To reach our desired outcome and to have a concise analysis of the datasets used we have to go through numerous steps. The first step was being able to identify the correct neighborhoods and map them. To use the Foursquare API more effectively we limited our Neighborhoods to the ones that have more 10 or more venues. The reason for doing this is because, it would practically be more effective to open a restaurant in an area that is popular and has more venues that people might visit.

One of the main deciding factors for this analysis was the number of clusters we used. In our case we chose K=4 due to its high Silhouette coefficient compared to other k values. It is also important to note a high k value is also not recommended as it would diversify the neighborhood clusters and risks us not being able to identify a suitable cluster correctly. By choosing using K-Means clustering, with K = 4, we were able to carefully examine the unique characteristics of the neighborhoods. For example, in our case we carefully observed the % of food venues present. In our case cluster 1 proved to be the best option. This is because it has a fewer number of food venues and the venues are diverse, therefore the possibility of people trying a different cuisine is higher.

The next step taken was to find the existing Greek restaurants in that location. As we observed Harvard Heights has 3 while Korea Town has 1. All four restaurants are within a 500-meter radius from the center of the cluster. If we are to choose a safer option, it would be to go for Harvard Heights because of the familiarity with the cuisine. To confirm the choice as being the most optimal one we also checked the distance of the two neighborhoods from the city center of Los Angeles. In our case both were approximately at a 6 km distance. Therefore, this factor did not affect our result.

One last step taken to ensure that it is easier for the restaurant to sustain itself was by making sure the rent price in the location is affordable. We did that by scrapping a table [2] containing the average rent of all neighborhoods. Unfortunately, the rent for Korea Town was not included in the database and is yet to be added by the website. However, the rental prices of an area give us a considerable amount on information about the demographics therefore it is useful when planning the setup of the restaurant itself. If we were to open up an affordable restaurant a location such as Harvard Heights is a very suitable location as it has one of the lowest rent prices in the county. However, let us assume the client wants to open a restaurant which leans more towards fine dining. In a case like that it should not be set up in medium to low income neighborhood where it would not perform well. At the same time a restaurant that operates as a casual dining or fast food restaurant would preferably be set up in an area where the rent is low and affordable enough to maintain the restaurants finances. This is particularly due to the low price of food.

April 22,2020

Conclusion

This project focused on highlighting the basic research needed to find potential neighborhoods to open a Greek restaurant in Los Angeles. It follows a few simple procedures which were, finding neighborhood data of Los Angeles. This can be done by either web scrapping or finding a usable *.json* file. The next step conducted was to obtain the coordinates of the respective neighborhoods using a geolocator and then mapping them over a map of Los Angeles using Folium. Our main instrument in this project was the use of Foursquare API which gives us access to the commercial traffic in the region. Commercial traffic, in this case, is all the venues listed in their respective locations and their frequency. The API helps us identify hotspots are more occurring more frequently and plan accordingly. Using the Foursquare API we selected the neighborhoods with at least 10 venues and used K-means clustering (K = 4) to further examine our data frame. According to our needs Cluster, 1 was chosen to be the best option. Finally, our neighborhood choices were trimmed down to two neighborhoods, Harvard Heights and Korea Town. We chose Harvard Heights in the end due to three primary reasons:

- The number of Greek Restaurants was more in Harvard Heights was 3 times more than in Korea Town. Therefore, it can be considered a safe option.
- The distance is approximately 6 km from the City Center.
- The rent price in Harvard Heights is cheaper in comparison to most neighborhoods.

Realistic Constraints

It is important to note that the location of the restaurant was solely done by choosing 3 very simple factors and serve as a point on the road map towards opening a restaurant. In reality, there are many more factors to consider minimizing risk. These include:

- Population Density of the area
- The availability of lot that matches our required size
- The budget of the restaurant itself.
- Safety and crime record of the area.
- Public transport around the location.