1. Introduction

1.1. Description of the problem and discussion of the background

Moving from one part of the country to another has always been exciting for many people. This will provide opportunities to explore new places, make new friends and try new food. However, this excitement often entails some challenges and anxieties. One quite often thinks of which part of a new area to get an apartment, how the rent prices differ across the neighborhoods, which neighborhood has more jobs that fit my abilities and skills. As a future data scientist, I have decided to analyze these questions and help the travelers to make the right decisions.

Los Angeles (LA) is one of the biggest and most populous cities of the USA. Due to its Mediterranean climate, ethnic diversity and Hollywood entertainment industry, every year many people come and enjoy all the benefits the city offers [1]. For this capstone project, I have decided to explore central Los Angeles area which consists of 26 neighborhoods [2]. I have investigated the most common venues in each neighborhood. Then, the neighborhoods have been clustered based on their similar venues. Also, I have incorporated the average rent price of apartments into the analysis. Finally, all the analysis and observations have been depicted in a map which provides visual information for the prospective travelers to central LA area.

2. Description of the data and how it will be used to solve the problem

The names of 26 neighborhoods have been scraped from a web page [2]. The latitude and longitude values of each neighborhood have been extracted from a web site [3] as well as [4]. As for the average rent prices for the neighborhoods, those have been obtained from the internet [5]. The geojson data that defines boundaries of the neighborhoods was not quite readily available [6]. Therefore, the source data have been explored, cleaned and reshaped to meet the requirements of this project. The venues data has been extracted using Foursquare API. To solve the existing problem, the collectedd data was used to cluster the neighborhoods based on their similar venues. This will help the interested parties to choose a particular set of neighborhoods that fit their requirements best. Moreover, the choropleth map was constructed based on the average rent price of the apartments in each neighborhood which will help the travelers to make wise

decision before settling down in a specific neighborhood. Finally, some recommendations were given according to the observed analysis of the project.

2. Methodology

The names of 26 neighborhoods were identified and scraped from the web page [2] using the BeautifulSoup python package.

- Arlington Heights
- Beverly Grove
- Carthay
- Chinatown
- Downtown
- East Hollywood
- Echo Park
- Elysian Park
- Elysian Valley
- Faircrest Heights
- Griffith Park
- Hancock Park
- Harvard Heights
- Hollywood
- Hollywood Hills
- Hollywood Hills West
- Koreatown
- Larchmont
- Los Feliz
- Mid-City
- Mid-Wilshire
- Pico-Union
- Silver Lake
- West Hollywood
- Westlake
- Windsor Square

According to [3], the neighborhood Faircrest Heights is called Fairfax. To be consistent within all the datasets, I have renamed that neighborhood accordingly. After searching the internet, I have found the LA neighborhood data [3]. Here you can see first few rows of it:

LA_Nbhd	Income	Schools	Diversity	Age	Homes	Vets	Asian	Black	Latino	White	Population	Area	Longitude	Latitude
Adams_Normandie	29606	691	0.6	26	0.26	0.05	0.05	0.25	0.62	0.06	31068	0.8	-118.30027	34.03097
Arleta	65649	719	0.4	29	0.29	0.07	0.11	0.02	0.72	0.13	31068	3.1	-118.430015	34.240603

Since I needed only longitude and latitude values, I have removed all unnecessary columns and end up having the following table:

	Neighborhood	Longitude	Latitude
0	Adams_Normandie	-118.300270	34.030970
1	Arleta	-118.430015	34.240603
2	Arlington_Heights	-118.320109	34.043611

The names of the neighborhoods in [2] and [3] were different because one used hyphen and the other did underscore between the words. Therefore, I have modified the names and made them identical. The data in [3] had information for all the neighborhoods in LA city. Because I am exploring only central LA area, I extracted only those neighborhoods that are listed in [2]. The shape attribute of the data frame revealed that resulting table was missing latitude and longitude values for two neighborhoods which were Griffith Park and West Hollywood. Those were obtained from a different source [4] and included into the data table.

Next step was to get rent prices for each neighborhood. To accomplish this task, I have scraped another web page [5] and generated the following data table:

	Neighborhood	Average Rent
0	Jefferson Park	$1,355
1	El Sereno	$1,396
2	Vermont Vista	$1,445

Before merging the above table with the table that has latitude longitude values, I had to again modify neighborhood names so that they match. Here is the final result:

	Neighborhood	Longitude	Latitude	Average Rent
0	Arlington_Heights	-118.320109	34.043611	$1,605
1	Beverly_Grove	-118.372111	34.076009	$3,804
2	Carthay	-118.369000	34.059000	$3,591

Although average rent data was available for all the neighborhoods in central LA area, three neighborhoods generated NaN values. A close inspection revealed that the names of neighborhoods in the table and in the source data were slightly different due to a hyphen and a white space for those three neighborhoods. Thus, NaN values were replaced with the actual values.

The python's Folium package was used to initially visualize the neighborhoods in Los Angeles city with their center markers constructed using latitude and longitude values.
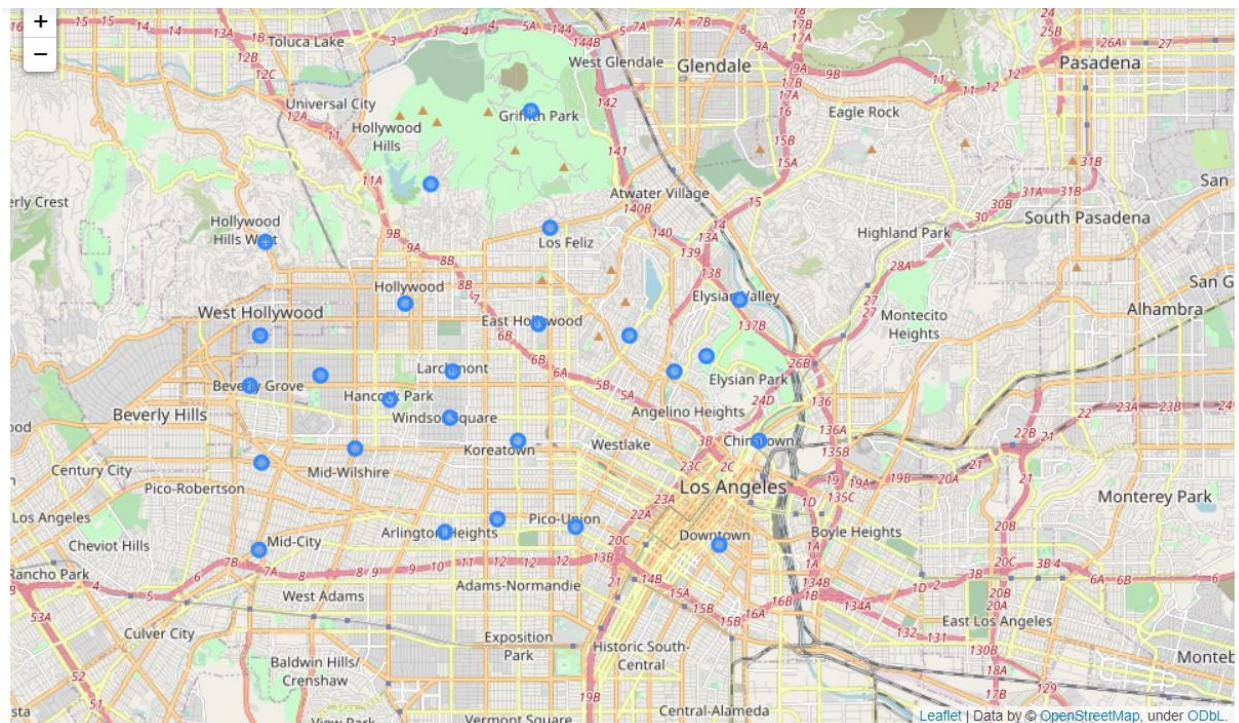

Using Foursquare API top 100 venues for each neighborhood within 500 meters were received. The table below shows the number of venues returned for each neighborhood. The number of unique venue categories turned out to be 183.

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Arlington_Heights	8	8	8	8	8	8
Beverly_Grove	100	100	100	100	100	100
Carthay	15	15	15	15	15	15
Chinatown	44	44	44	44	44	44
Downtown	11	11	11	11	11	11
East_Hollywood	26	26	26	26	26	26
Echo_Park	41	41	41	41	41	41
Elysian_Park	13	13	13	13	13	13
Elysian_Valley	8	8	8	8	8	8
Fairfax	28	28	28	28	28	28
Griffith_Park	4	4	4	4	4	4
Hancock_Park	1	1	1	1	1	1
Harvard_Heights	17	17	17	17	17	17
Hollywood	30	30	30	30	30	30
Hollywood_Hills	6	6	6	6	6	6
Hollywood_Hills_West	6	6	6	6	6	6
Koreatown	89	89	89	89	89	89
Larchmont	6	6	6	6	6	6
Los_Feliz	24	24	24	24	24	24
Mid_City	8	8	8	8	8	8
Mid_Wilshire	57	57	57	57	57	57
Pico_Union	16	16	16	16	16	16
Silver_Lake	34	34	34	34	34	34
West_Hollywood	62	62	62	62	62	62

Then, after manipulating the above table, I have generated the top 10 venues for each neighborhood. It can be seen in the table below.

	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	Arlington_Heights	Restaurant	Seafood Restaurant	Café	Grocery Store	Shop & Service	Pizza Place	Liquor Store	Rental Car Location	Dumpling Restaurant	Fast Food Restaurant
1	Beverly_Grove	Clothing Store	Cosmetics Shop	Juice Bar	Mexican Restaurant	Italian Restaurant	Seafood Restaurant	Sushi Restaurant	Lingerie Store	Café	Gym / Fitness Center
2	Carthay	Ethiopian Restaurant	Historic Site	Fast Food Restaurant	Intersection	Dentist's Office	Video Store	Café	Park	Food Truck	Convenience Store
3	Chinatown	Chinese Restaurant	Vietnamese Restaurant	Bakery	Bar	Café	Coffee Shop	French Restaurant	Bookstore	Grocery Store	Brewery
4	Downtown	Arts & Crafts Store	Southern / Soul Food Restaurant	American Restaurant	Wine Bar	Gay Bar	Neighborhood	Café	Breakfast Spot	Sushi Restaurant	Flower Shop
5	East_Hollywood	Asian Restaurant	Pizza Place	Bakery	Restaurant	Pharmacy	Convenience Store	Fast Food Restaurant	Bus Line	Middle Eastern Restaurant	Coffee Shop
6	Echo_Park	Food Truck	Café	Mexican Restaurant	Bar	Pizza Place	Coffee Shop	American Restaurant	Art Museum	Restaurant	Wine Bar
7	Elysian_Park	Yoga Studio	Grocery Store	Furniture / Home Store	Juice Bar	Arts & Crafts Store	Café	Trail	Gift Shop	Scenic Lookout	Playground
8	Elysian_Valley	Trail	Rock Club	Gym / Fitness Center	Rental Car Location	Burrito Place	Spiritual Center	Intersection	Asian Restaurant	Fast Food Restaurant	Garden Center
9	Fairfax	Italian Restaurant	Mexican Restaurant	Pet Store	Arts & Crafts Store	Coffee Shop	Residential Building (Apartment / Condo)	Mediterranean Restaurant	Dessert Shop	Sandwich Place	Cosmetics Shop

Then, the next step was to cluster the neighborhoods. To do this, I used K-Means clustering algorithm. This algorithm requires from the user to provide the number of clusters to be generated. Rather than guessing, I have used the elbow method to identify what number of clusters would be reasonable.
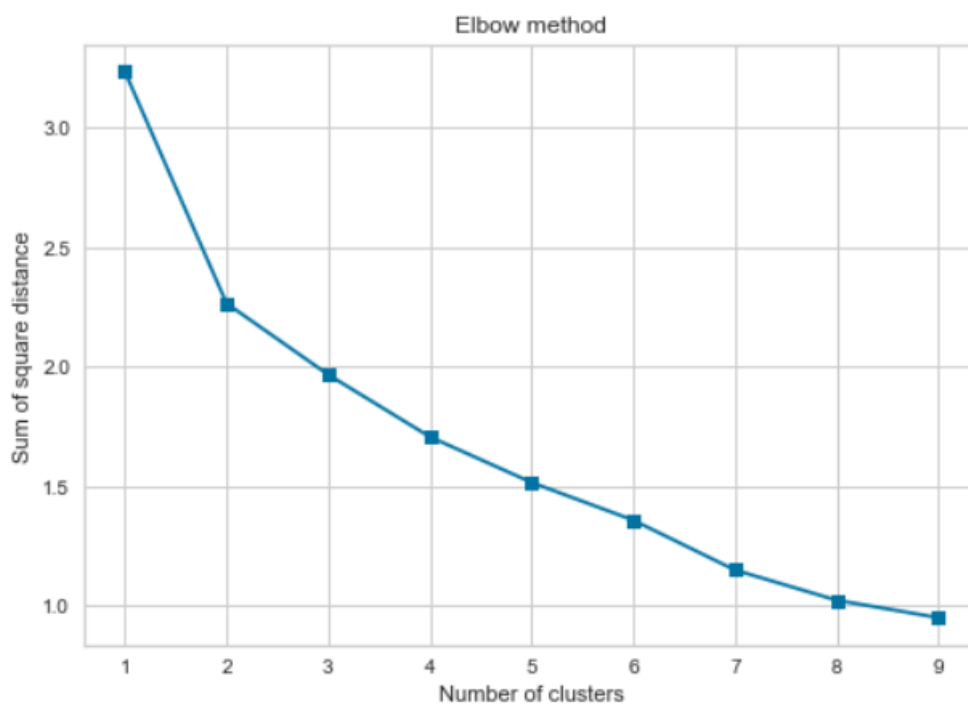
As we can see from the elbow method above, the sum of squares distance within the clusters does not decrease significantly after k = 3. Therefore, I have decided to generate three clusters.

Before visualizing the clusters on the map, I have decided to construct the choropleth map that depicts average rent prices across the neighborhoods. To do this, I needed a geojson file that defines boundaries of the neighborhoods that were being analyzed. Fortunately, there was a json file that contained data for all the neighborhoods in LA city [6]. After exploring and cleaning the file, and removing some unnecessary values, I was able to create json file just for the central LA neighborhoods. Then, the neighborhood clusters were superimposed on top of the choropleth map thus providing more comprehensive information about each neighborhood.

3. Results

The figure below shows the generated choropleth map. As we can see from the color bar, the darker the color the higher the average rent price in that neighborhood.
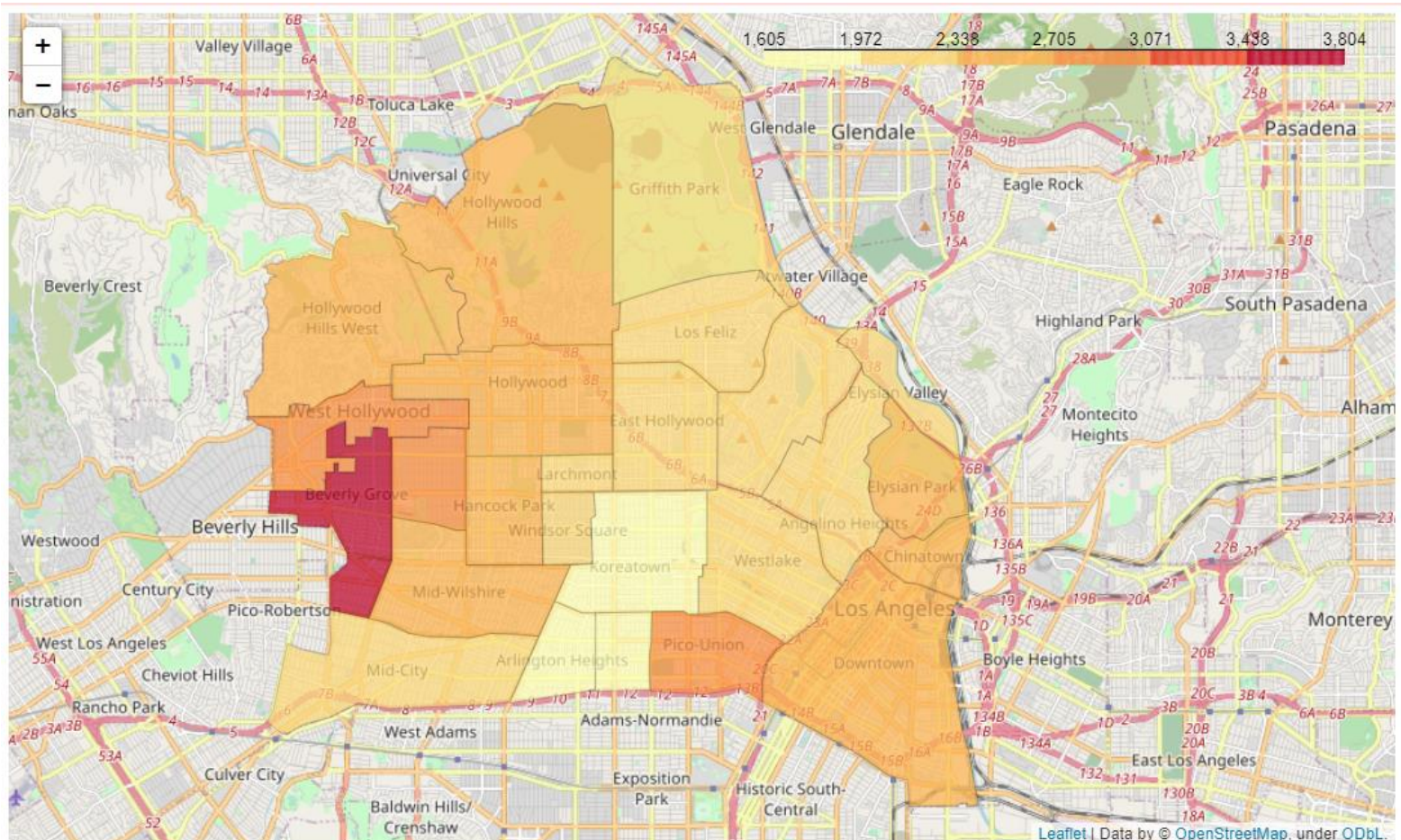
Here, we can see the map of the neighborhoods that has both choropleth map and neighborhood clusters superimposed on top of each other. The clusters are colored in red, blue and green. What is great about this map is that if you click on the circle markers which are the centers of the neighborhoods, they will display the name of the neighborhood, the category of cluster and the average rent price. Is not it cool?
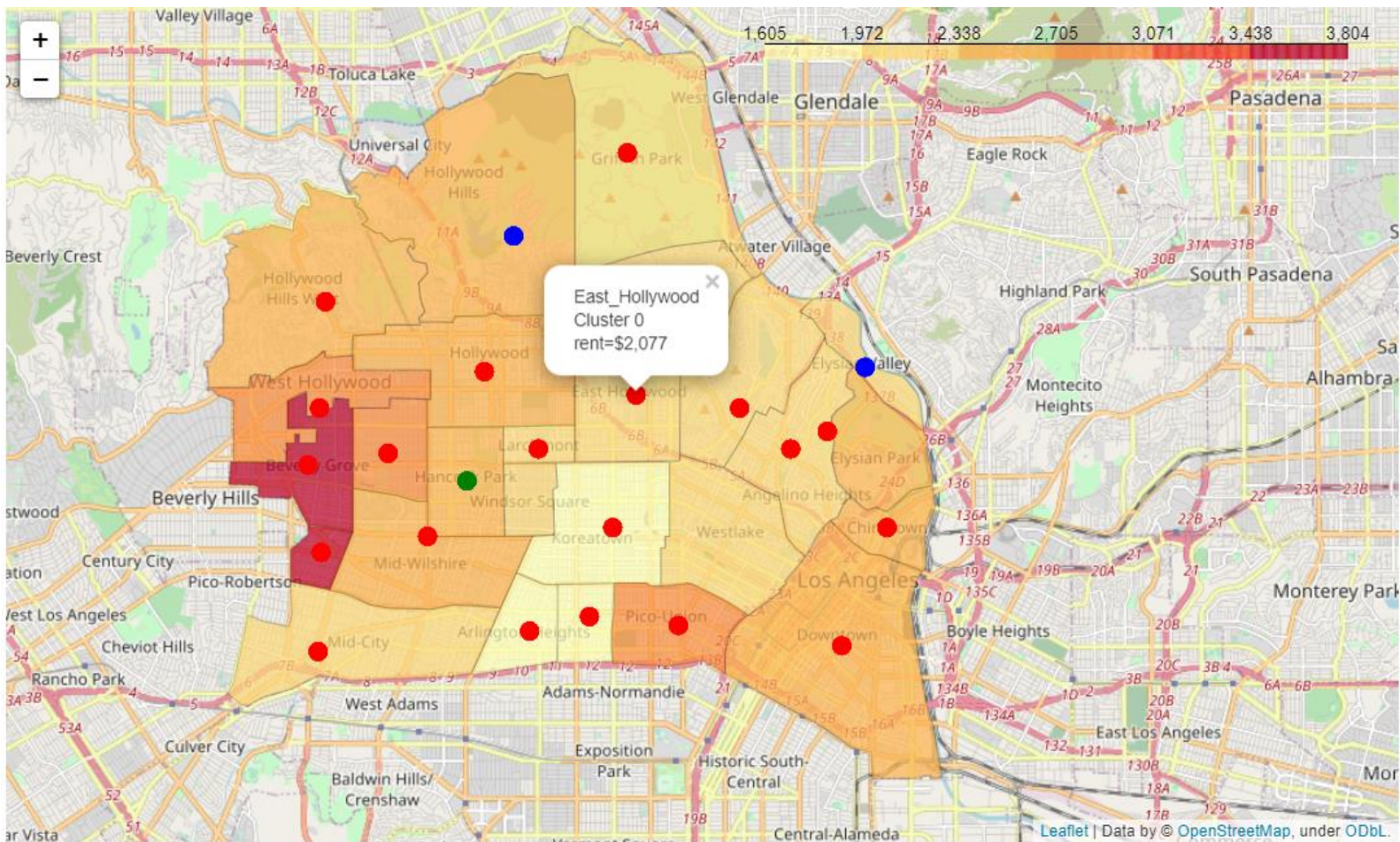
Next, we will present each cluster with their top 10 venues.

Cluster 1

	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	Arlington_Heights	Restaurant	Seafood Restaurant	Café	Grocery Store	Shop & Service	Pizza Place	Liquor Store	Rental Car Location	Dumpling Restaurant	Fast Food Restaurant
1	Beverly_Grove	Clothing Store	Cosmetics Shop	Juice Bar	Mexican Restaurant	Italian Restaurant	Seafood Restaurant	Sushi Restaurant	Lingerie Store	Café	Gym / Fitness Center
2	Carthay	Ethiopian Restaurant	Historic Site	Fast Food Restaurant	Intersection	Dentist's Office	Video Store	Café	Park	Food Truck	Convenience Store
3	Chinatown	Chinese Restaurant	Vietnamese Restaurant	Bakery	Bar	Café	Coffee Shop	French Restaurant	Bookstore	Grocery Store	Brewery
4	Downtown	Arts & Crafts Store	Southern / Soul Food Restaurant	American Restaurant	Wine Bar	Gay Bar	Neighborhood	Café	Breakfast Spot	Sushi Restaurant	Flower Shop
5	East_Hollywood	Asian Restaurant	Pizza Place	Bakery	Restaurant	Pharmacy	Convenience Store	Fast Food Restaurant	Bus Line	Middle Eastern Restaurant	Coffee Shop
6	Echo_Park	Food Truck	Café	Mexican Restaurant	Bar	Pizza Place	Coffee Shop	American Restaurant	Art Museum	Restaurant	Wine Bar
7	Elysian_Park	Yoga Studio	Grocery Store	Furniture / Home Store	Juice Bar	Arts & Crafts Store	Café	Trail	Gift Shop	Scenic Lookout	Playground
9	Fairfax	Italian Restaurant	Mexican Restaurant	Pet Store	Arts & Crafts Store	Coffee Shop	Residential Building (Apartment / Condo)	Mediterranean Restaurant	Dessert Shop	Sandwich Place	Cosmetics Shop
11	Harvard_Heights	Korean Restaurant	Thrift / Vintage Store	Fast Food Restaurant	Japanese Restaurant	Donut Shop	Dentist's Office	Pizza Place	Café	Gas Station	Grocery Store
12	Hollywood	Vegetarian / Vegan Restaurant	Café	Coffee Shop	Pizza Place	Lounge	Park	Movie Theater	Bar	Multiplex	Shipping Store
14	Hollywood_Hills_West	Night Market	Health & Beauty Service	Italian Restaurant	Café	Grocery Store	Hotel	Garden Center	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint
15	Koreatown	Korean Restaurant	Coffee Shop	Bubble Tea Shop	Hotel	Asian Restaurant	Ice Cream Shop	Café	Karaoke Bar	Restaurant	Japanese Restaurant
16	Larchmont	Spa	American Restaurant	Indie Movie Theater	Korean Restaurant	Park	Filipino Restaurant	Gas Station	Garden Center	Furniture / Home Store	Frozen Yogurt Shop

Cluster 2

	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
8	Elysian_Valley	Trail	Rock Club	Gym / Fitness Center	Rental Car Location	Burrito Place	Spiritual Center	Intersection	Asian Restaurant	Fast Food Restaurant	Garden Center
13	Hollywood_Hills	Trail	Movie Theater	Café	Gym / Fitness Center	Yoga Studio	Filipino Restaurant	Garden Center	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint

Cluster 3

	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
10	Hancock_Park	Bookstore	Yoga Studio	Filipino Restaurant	Gas Station	Garden Center	Furniture / Home Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck

4. Discussion

After analyzing the central LA neighborhoods, I have learned the following facts:

The most affordable neighborhoods are Arlington Heights where the average rent is $1,605/month, Harvard Heights where renters pay on average $1,607/month on average and Koreatown where the average rent is $1,970/month.

The most expensive neighborhoods turned out to be Carthay where the average rent is $3,591/month and Beverly Grove where renters pay $3,804/month on average.

Another interesting fact is that the neighborhoods that are located on the eastern part of the area are relatively cheaper than those that are located on the western part.

As for the clusters, because cluster 1 contains most of the neighborhoods, the central LA neighborhoods does not differ to much in terms of the venues they have. As we observed, most of the venues turned out to be the restaurants. Thus, if one has restaurant related skills and knowledge, then the central LA has a lot of jobs to offer.

If one likes working out, doing physical exercise and tries living in a healthy lifestyle, then I would recommend living in neighborhoods that belong to cluster 2. This cluster has lots of trails and gym/fitness centers.

The cluster 3 seems to be suitable for those who likes reading a lot because in this cluster they can find many bookstores.

Finally, for travelers and guests who want to stay or live in central LA, I would recommend living either in Arlington Heights or in Harvard Heights or in

Koreatown. These three neighborhoods are much cheaper than others and the same time they do not differ significantly from other neighborhoods in terms of the type of venues and jobs they have.

5. Conclusion

In this study, I have explored the central Los Angeles area using the spatial data. The Foursquare API was applied to obtain venues around each neighborhood. Then, top 10 venues for each neighborhood were presented. K-Means clustering algorithm was used to cluster the neighborhoods into 3 distinct areas. To show the average rent price across the central LA area, the choropleth map was constructed along with clustered neighborhoods superimposed on top of it. Based on the final outcomes of the project some insights and recommendations were presented for the interested audience. For example, the neighborhoods such as Arlington Heights, Harvard Heights and Koreatown turned out to have much less average rent price compared to the rest of the neighborhoods while being very similar to the rest of the neighborhoods in terms of the types of venues and jobs. This being said, one can save significant amount of money on the rent payment while working in the same jobs that are available in other expensive neighborhoods. Moreover, because the cluster 1 has mostly restaurant venues, it is the best choice for those who has restaurant job experience. The cluster 2 contains may trails and gym/fitness centers for those who tries to stay fit. As for cluster 3, it is a good place for those who likes reading books because these neighborhoods have many bookstores.

6. Bibliography

1. https://en.wikipedia.org/wiki/Los_Angeles

2. https://en.wikipedia.org/wiki/Central_Los_Angeles

3. http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_LA_Neighborhoods_Data

4. https://usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr

5. https://www.rentcafe.com/average-rent-market-trends/us/ca/los-angeles/

6.https://s3-us-west-.amazonaws.com/mappingla.com/downloads/neighborhoods/la_city.json?_ga=2.101578435.35324892.1592446928-1423563668.1590043322