C9_WK4,5_ASSIGNMENT_RD

TITLE: USING FOURSQAURE APP TO CHOOSE A CITY NEIGHBORHOOD TO RENT IN.

REPORT BY: RYAN DESOUSA

DATE: 19 JUNE 2020

1. INTRODUCTION

1.1 Background and Business Problem

As people relocate between cities for work or otherwise, they often start out in the new location by renting.

As you move to a new city, how would one choose the area of the new city in which to live?

This would depend a lot on the individuals personal choices and lifestyle.

Using rental data that is available online, I would like to explore the use of the Four Square API app to see if the combination can be used to get a better understanding of neighborhoods in a target city and if, based on user input and preference, any neighborhoods in a city are of greater interest to rent in as compared to others.

This target audience are people who are moving to new cities and towns and who want to get a better idea how a different neighborhoods in the city will suit their needs.

The city that I will be exploring will be Los Angeles, CA.

1.2) Data

The average rental rates for different neighborhoods in the city of Los Angeles are available at:

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

This neighborhood and rental data will be extracted to csv and then into a dataframe and cleaned.

I will try to pair this data, with information on the different LA neighbourhoods, obtained through the Four Square API app.

I will use this rental data and the API app to explore different venues in the LA neighborhoods to try get an idea of what are the high traffic offerings in the various neighborhoods.

I will use one hot encoding to try to find the ten most frequently occurring venues in the different neighborhoods.

I will then apply K-Means Clustering Machine Learning algorithm on the acquired neighborhood-venue data to cluster the data into five clusters.

The Neighborhood cluster will then be plotted on a map of LA city using Folium.

I will then filter the Neighborhood-Venue frequency data set, using venues of my choice and look for neighborhoods where these venues are among the top three most frequent venues.

Using the FourSquare API, I will then search any two of these neighborhoods further, with five venues of my choice and try to determine the frequency of occurrence of these venues, within a distance of 2000m of their neighborhood coordinates.

Once I have all this information, I will try to put all this information together to look for similarities/differences in the two neighborhoods and see if any one neighborhood is more preferred to rent in than the other.

The final goal is to see whether this data can be used to differentiate between the neighborhoods and be used by people to make a better decisions on choice of city neighborhood to rent in.

1.3. Import Necessary Libraries.

The necessary libraries were imported.

2 . DATA HANDLING – Location, Acquisition and Treatment.

The average rental rates for different neighborhoods in the city of Los Angeles are available at:

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

This neighborhood and rental data was extracted to csv and then into a dataframe and cleaned.

As I found the data was not scrapable, I manually had to copy the table to excel and save it as a csv.

The csv was converted to a dataframe.

The data was cleaned to remove any missing values and the names LA_Neighborhood column were formatted as 'Neighborhood Name', CA, USA.

This format allowed for more accurate coordinate extraction using the FourSquare API

The data in the AverageMonthlyRent_USD Column had a '$' sign that had to be deleted and the data converted to float from object.

The cleaned dataframe is shown in Figure 2.1

	LA_Neighborhood	AverageMonthlyRent_USD
0	Jefferson Park, CA, USA	1355
1	El Sereno, CA, USA	1396
2	Vermont Vista, CA, USA	1445
3	Vermont Knolls, CA, USA	1445
4	Hyde Park, CA, USA	1484

Figure 2.1: LA Neighborhood rent dataframe la_rent, after treatment.

I then paired neighborhood information, with coordinate information on the different LA neighborhoods, obtained through the Four Square API app to get dataframe LA_lat_long, as shown in Figure 2.2

	LA_Neighborhood	Latitude	Longitude
0	Jefferson Park, CA, USA	34.027234	-118.317576
1	El Sereno, CA, USA	34.081121	-118.177849
2	Vermont Vista, CA, USA	33.941947	-118.285814
3	Vermont Knolls, CA, USA	33.966819	-118.291670
4	Hyde Park, CA, USA	33.980569	-118.330631
...	...	...	...
93	Beverly Grove, CA, USA	34.076034	-118.369972
94	Historic South-Central, CA, USA	34.016230	-118.267308
95	University Park, CA, USA	34.027449	-118.283949
96	Pico, CA, USA	34.040672	-118.266192
97	Santa Monica, CA, USA	34.025072	-118.496513

Figure 2.2: LA Neighborhood and coordinate dataframe La_lat_long.

I then added the rental information to this dataframe to get the dataframe la_data, with all the necessary neighborhood information in one dataframe

	LA_Neighborhood	AverageMonthlyRent_USD	Latitude	Longitude
0	Jefferson Park, CA, USA	1355.0	34.027234	-118.317576
1	El Sereno, CA, USA	1396.0	34.081121	-118.177849
2	Vermont Vista, CA, USA	1445.0	33.941947	-118.285814
3	Vermont Knolls, CA, USA	1445.0	33.966819	-118.291670
4	Hyde Park, CA, USA	1484.0	33.980569	-118.330631

Figure 2.3: LA Neighborhood, rent and coordinate information dataframe, la_data.

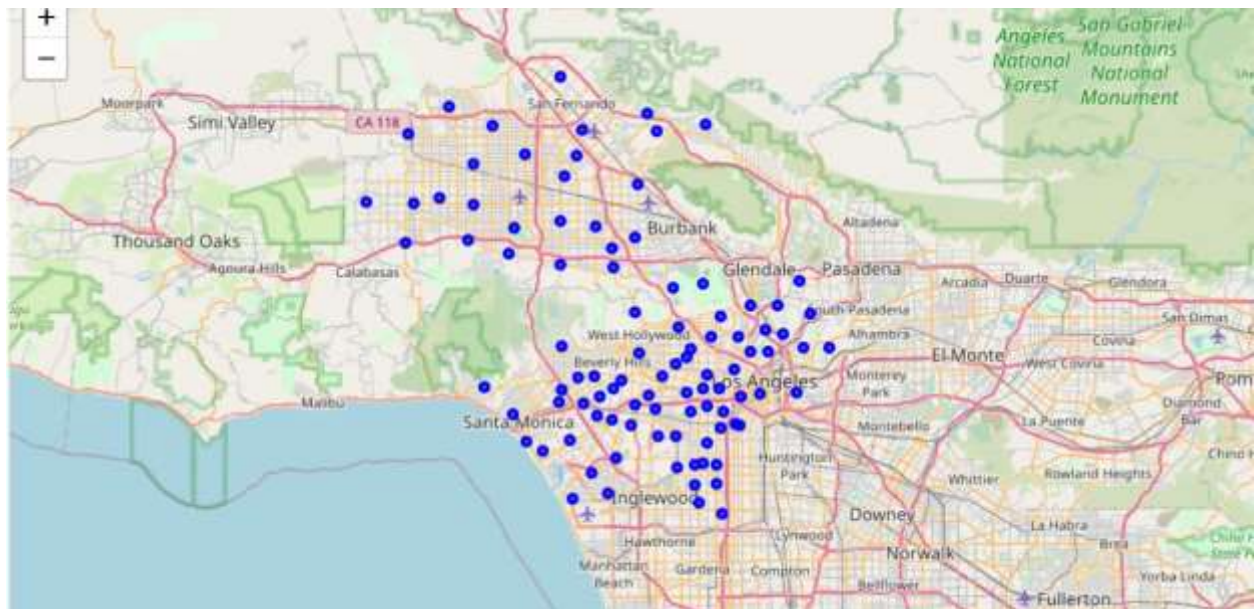The neighborhoods were then viewed on Folium Maps.

Figure 2.4: LA Neighborhoods on Folium map of LA city.

3) BATTLE OF LA NEIGHBORHOODS - METHODOLOGY: EXPLORE VENUES

I used this rental data and the FourSquare API app to explore different venues in the LA neighborhoods to try get an idea of what are the high traffic offerings in the various neighborhoods.

I used one hot encoding to find the ten most frequently occurring venues in the different neighborhoods as shown in figure 3.1

	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	Adams - Normandie, CA, USA	Sushi Restaurant	Taco Place	Gas Station	Wings Joint	Food Service	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Farm
1	Arleta, CA, USA	Movie Theater	Historic Site	Wings Joint	Donut Shop	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Farm	Dive Bar
2	Arlington Heights, CA, USA	Convenience Store	Art Gallery	Restaurant	Café	Latin American Restaurant	Donut Shop	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market
3	Atwater Village, CA, USA	Mexican Restaurant	Bakery	Pizza Place	Italian Restaurant	Farmers Market	Food Service	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant
4	Baldwin Hills - Crenshaw, CA, USA	Trail	Park	Music Venue	Wings Joint	Donut Shop	Flower Shop	Filipino Restaurant	Fast Food Restaurant	Farmers Market	Farm

Figure 3.1: neighborhood_venues_sorted - ten most frequently occurring venues in the different neighborhoods

4) BATTLE OF LA NEIGHBORHOODS - METHODOLOGY: CLUSTER VENUES

I will then applied K-Means Clustering Machine Learning algorithm on the acquired neighborhood-venue data to cluster the data into five clusters.

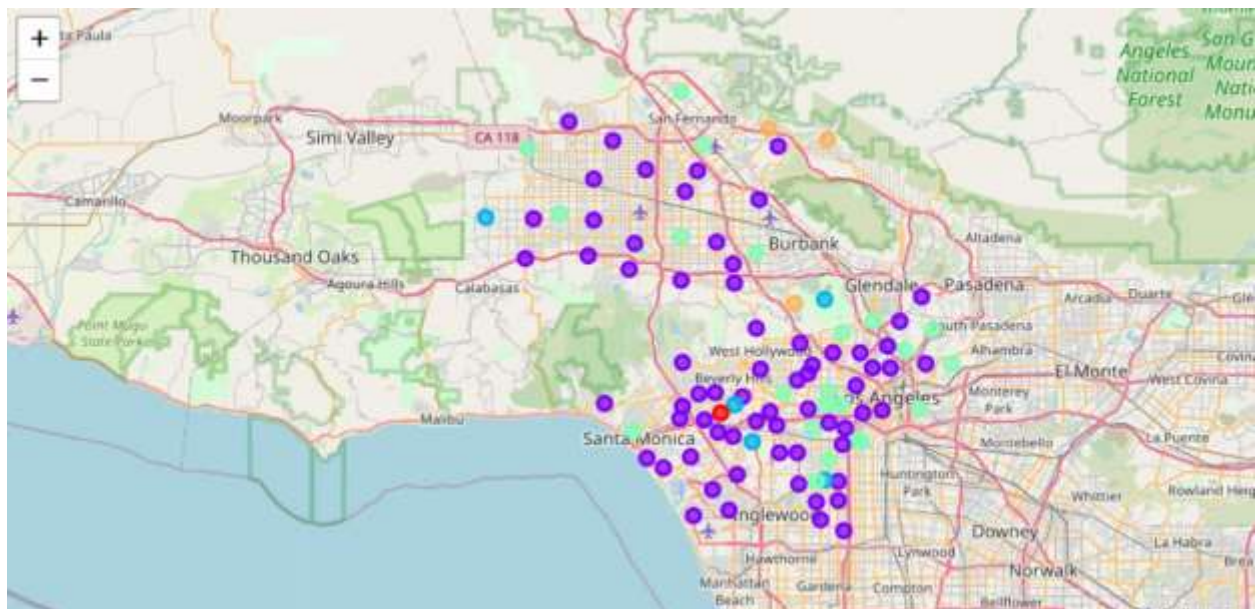The Neighborhood cluster was then plotted on a map of LA city using Folium.

Figure 3.2 Dolium Map of the LA Neighborhood clusters.

5) BATTLE OF LA NEIGHBORHOODS - METHODOLOGY: EVALUATE THESE NEIGHBORHOODS FURTHER, LOOKING FOR SPECIFIC VENUES OF INTEREST

I then filtered the Neighborhood-Venue frequency data set, using venues of my choice, in this case I used venue to be searched as 'Theater'.

I looked for neighborhoods where this venue was among the top three most common venues.

	Neighborhood	AverageMonthlyRent_USD	Latitude	Longitude	Cluster Labels	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 M Com Ve
0	Arleta, CA, USA	1634.0	34.241327	-118.432205	0	Movie Theater	Historic Site	Wings Joint	Donut Shop	Flower Shop	Filipino Restaurant	Fast F Restau
1	Larchmont, CA, USA	2140.0	34.079837	-118.317870	0	Indie Movie Theater	Korean Restaurant	Movie Theater	Park	American Restaurant	Cuban Restaurant	D St
2	Mid-City, CA, USA	2188.0	34.041527	-118.380370	0	Indie Theater	Gym / Fitness Center	Liquor Store	Theater	Food Truck	Dance Studio	I Boo
3	Hollywood, CA, USA	2485.0	34.098003	-118.329523	0	Coffee Shop	Movie Theater	Multiplex	Salon / Barbershop	Farmers Market	Wings Joint	D 5
4	Exposition Park, CA, USA	3522.0	34.013654	-118.287211	0	Science Museum	College Football Field	Movie Theater	Park	Wings Joint	Filipino Restaurant	Fast F Restau

Figure 5.1 df_or showing the neighborhoods where Theater was the top three most common venue.

Using the FourSquare API, I then searched any two of these neighborhoods further (this will be based on user choice), with five venues of my choice (again this will be based on user choice).

The aim was to determine the frequency of occurrence of these five venues, within a distance of 2000m of their neighborhood coordinates.

The five venues that I used here were Pharmacy, Market, Theater, Library and Restaurant.

Once I had all this information, I put all this information together to look for similarities/differences in the two neighborhoods and see if any one neighborhood is more preferred to rent in than the other.

I based this similarity or difference based on the average monthly rent, number of venues found, the number of each type of venue found and the median distance from the neighborhood coordinates.

This information is summarized in the Report Section.

The final goal is to see whether this data can be used to differentiate between the neighborhoods and be used by people to make a better decisions on choice of city neighborhood to rent in.

6. FINAL REPORT - BATTLE OF THE NEIGHBORHOODS

Neighborhood1: Exposition Park, LA, CA, USA.

Neighborhood2: Arleta, LA, CA, USA.

Both these neighborhoods were chosen from a list of neighborhoods that had 'Theatre' as the top three most common venues.

These two neighborhoods were then searched using FourSquare API for presence of the venues Pharmacy, Market, Theater, Library and Restaurant within a radius of 2000 meters.

The data for the two neighborhoods is summarized in Figure 6.1

FINAL NEIGHBORHOODS COMPARISON, SUMMARY by RD :

	LA_Neighborhood	Average_Monthly_Rent_USD	Radius (m)	Location	Found	Median Distance (km)
0	Exposition Park, CA, USA	3522.0	2000	Pharmacy	5	1.10
1	Exposition Park, CA, USA	3522.0	2000	Market	5	0.90
2	Exposition Park, CA, USA	3522.0	2000	Library	5	0.61
3	Exposition Park, CA, USA	3522.0	2000	Theater	4	0.73
4	Exposition Park, CA, USA	3522.0	2000	Restaurant	6	0.50
0	Arleta, CA, USA	1634.0	2000	Pharmacy	5	1.03
1	Arleta, CA, USA	1634.0	2000	Market	4	0.92
2	Arleta, CA, USA	1634.0	2000	Library	3	2.34
3	Arleta, CA, USA	1634.0	2000	Theater	1	1.74
4	Arleta, CA, USA	1634.0	2000	Restaurant	5	1.07

Figure 6.1: Summary of the data found on the two neighborhoods.
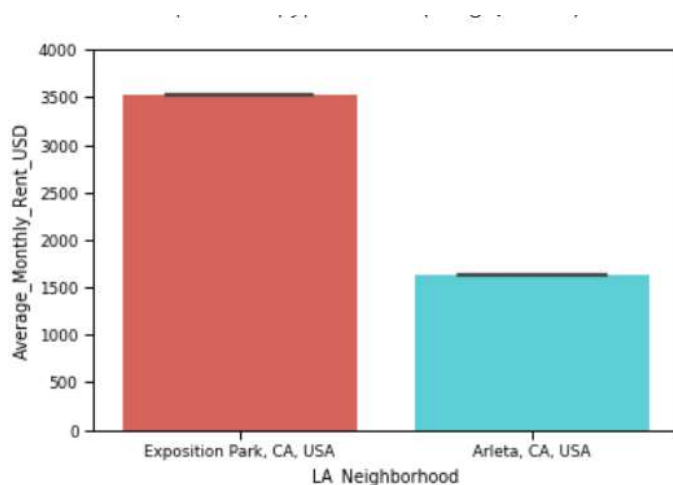
6.1 AVERAGE MONTHLY RENT

Figure 6.1.1: Comparison of the average monthly rent in the two neighborhoods

The Average monthly rent is higher in Exposition Park than in Arleta Neighborhood of Los Angeles.
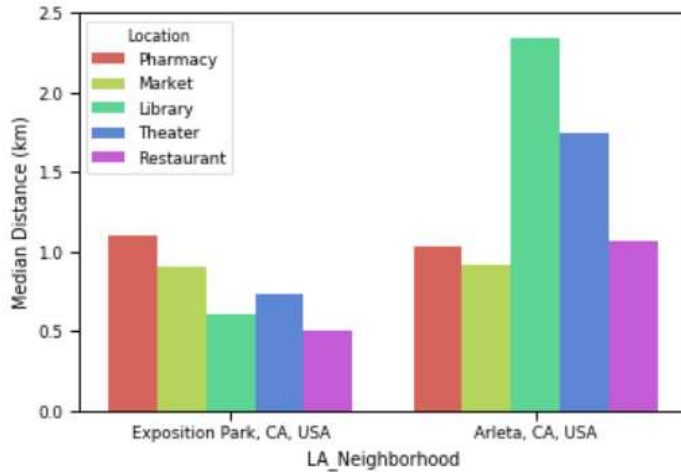
6.2   MEDIAN_DISTANCE

Figure 6.2.1: Comparison of the median distance of each venue from the neighborhood coordinates.

Both Exposition Park and Arleta neighborhoods have all of the of the five venues of interest being represented.

The median distance to the venues present appears to be higher in Arleta than in Exposition Park.
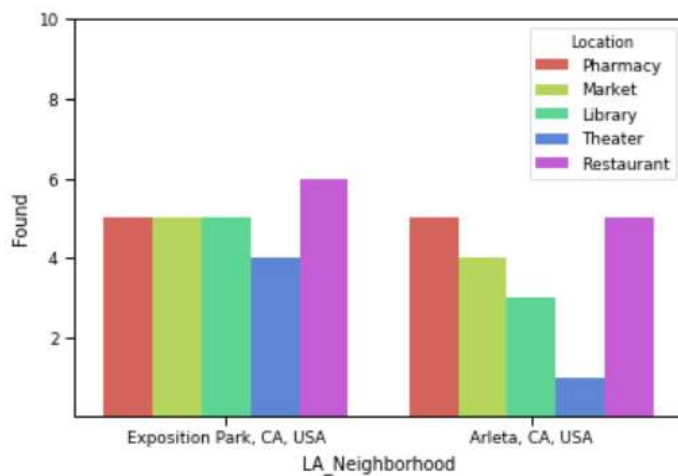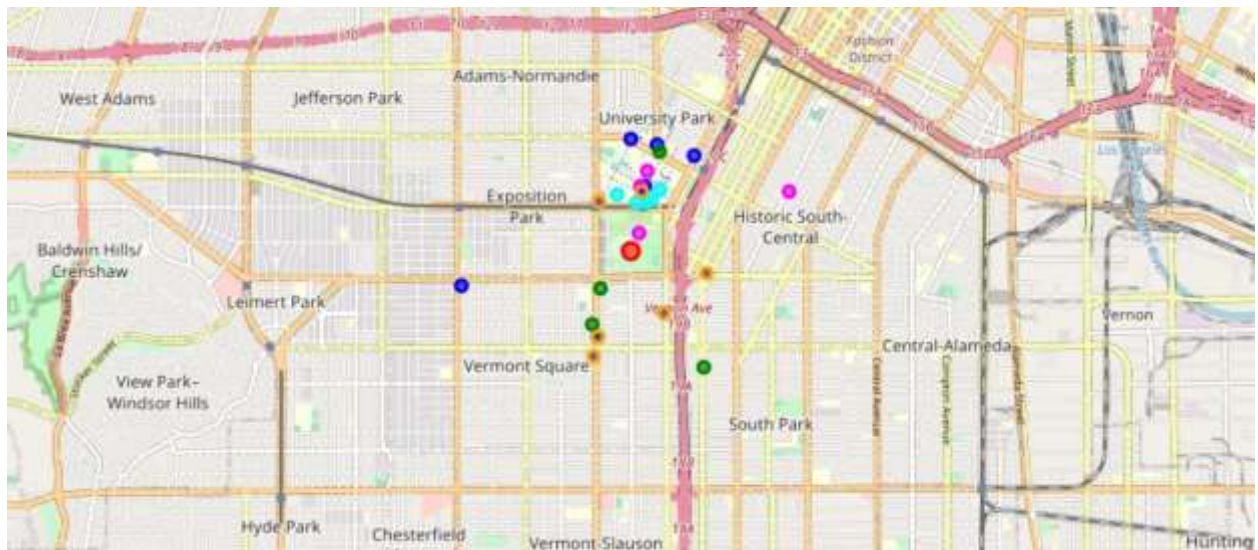
6.3   NUMBER OF VENUES FOUND

Figure 6.3.1: Number of venues found in each neighborhood.

Both Arleta and Exposition Park have all of the five venues present within 2000m of the target co-ordinates.
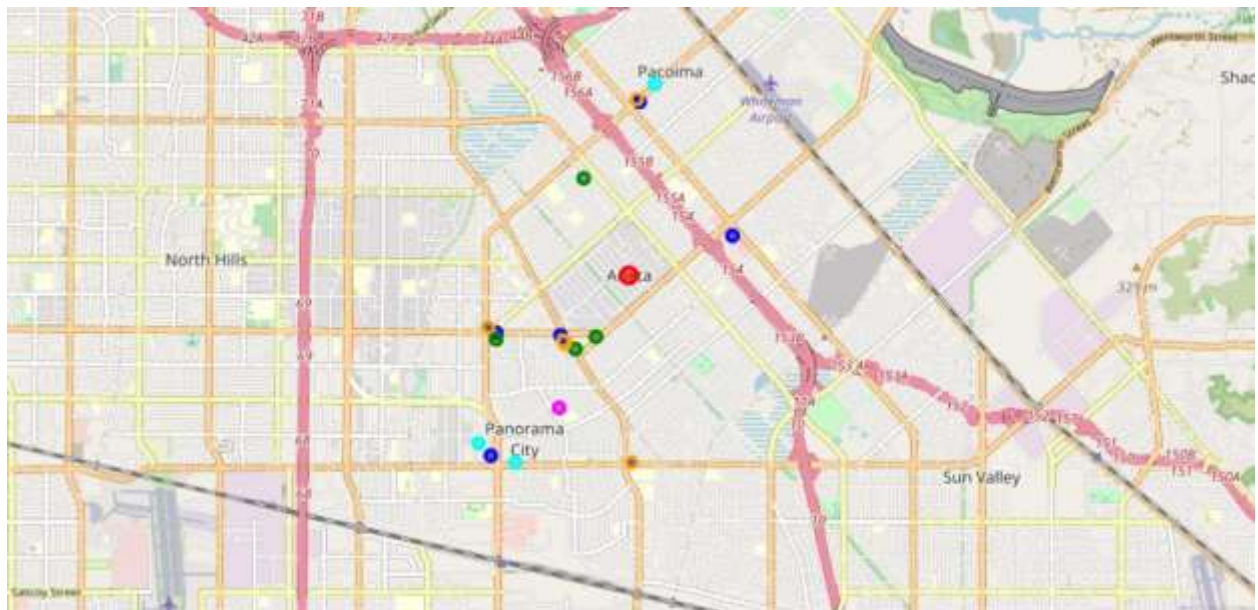
Exposition Park has the same number or more of each venue type found, within 2000m of the target coordinates.
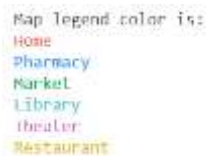
6.4 FOLIUM MAPS

Map legend color is:
Home
Pharmacy
Market
Library
Theater
Restaurant

Figure 6.4.1: Map of found venues in Exposition Park

```
Map legend color is:
Home
Pharmacy
Market
Library
theater
Restaurant
```

Figure 6.4.2: Map of found venues in Arleta

7. CONCLUSION

In this assignment I tried to combine and use rental data that is available online, and the FourSquare API app to explore the neighborhoods in the city of Los Angeles, CA.

The aim was to see if the combination can be used to get a better understanding of neighborhoods in a target city and if, based on user preference and input, are any neighborhoods in a city of greater interest to rent in as compared to others.

The city that I explored is Los Angeles, CA.

The data was treated as discussed in the data section.

K-Means clustering was applied to the data and the 5 clusters plotted on the map of Los Angeles using Folium

I used on hot encoding on this data to find the ten most frequently occurring venues in these neighborhoods.

I then filtered the dataset to find neighborhoods that had 'Theaters' as the top three most frequently occurring venues.

I used the Four Square API to search for the presence of five venues of my choice, in two of these neighborhoods.

These five venues being 'Pharmacy', 'Market', 'Library', Theatre', Restaurant'.

This search returned the number of venues if found and the distance of each of the venues from the neighborhood coordinates.

I used all this information to put together a report that is detailed in the Results section.

7.1 In Conclusion,

I think that it is possible to use the neighborhood data, rental data and the FourSquare API to explore and search neighborhood for venues of ones choice and come to an informed decision on which neighborhood is better to rent in.

This choice will likely be very personal.

7.2 Future Improvements:

More venues can be searched per neighborhood using a bigger radius.

More than two neighborhoods can be compared.

I think that adding crime data for various neighborhoods would add a new dimension to the data, that would help make a more informed decision. (I was unable to locate suitably paired neighborhood - rent - crime data.)