

HOME SWEET HOME

Coursera Capstone Project

Prepared by: Abdurahman Alrobaie

Date: 12/7/20

TABLE OF CONTENTS

1	INTRODUCTION: BUSINESS PROBLEM	3
2	DATA UNDERSTANDING.....	3
2.1	Data Sources	3
2.1.1	National Address API.....	3
2.1.2	GeoJson.io	4
2.1.3	Foursquare API.....	4
2.2	Retrieving Data.....	4
2.2.1	Neighborhoods Data	4
2.2.2	Foursquare Data.....	6
3	EXPLORATORY DATA ANALYSIS (EDA).....	7
3.1	Commercial Density	8
3.2	Outdoor Places Density.....	8
3.3	Industrial Density	8
4	METHODOLOGY	9
4.1	Preparing the Data	9
4.2	Clustering the Neighborhoods	10
4.2.1	Two Clusters.....	11
4.2.2	Three Clusters	12
4.2.3	Four Clusters	13
4.2.4	Five Clusters	14
5	RESULTS.....	15
5.1	First Cluster	15
5.2	Second Cluster	15
5.3	Third Cluster.....	16
5.4	Fourth Cluster	17
6	DISCUSSION.....	17
7	Conclusion.....	17

1 INTRODUCTION: BUSINESS PROBLEM

According to a statistical study by the General Authority for Statistics in Saudi Arabia, about 38% of the population do not own a permanent residence (Aawsat, 2019). The decision of owning a home is a critical action to many families in Saudi Arabia in terms of effort, time, and most importantly money. One point that stands as a crucial element in this process is the location of your house. Such a decision is crucial to determine many of your future life activities including the quality of living in a specific destination within the city and proximity to various services. This project addresses this issue by performing an analysis of Khobar City, SA, and clustering the promising neighborhoods according to the following factors:

1. Proximity to the city center
2. Services and venues within the neighborhood including:
 - a. Supermarkets
 - b. Clothing stores
 - c. Gyms
 - d. Mosques
 - e. Restaurants
 - f. Coffee Shops
 - g. Parks
 - h. Schools
3. Distance from factories and industrial areas

Consequently, a Machine Learning (ML) clustering model is fitted to cluster the neighborhoods based on the venues nearby them. The segmentation of neighborhoods can cluster similar neighborhoods for an easy process of finding your future home's location. The ML does not rate the neighborhoods but rather it shows the similarity and dissimilarity between neighborhoods. For example, a neighborhood located within the edges of the city with few services would be in a different cluster than a neighborhood close to the center with many services. Additionally, the model would place neighborhoods of dense commercial areas together which can help both customers and workers to choose a convenient location.

2 DATA UNDERSTANDING

2.1 Data Sources

The data required to cluster the neighborhoods are retrieved from different sources. Followingly, they are combined to carry on Exploratory Data Analysis (EDA), wrangling, and fitting the ML model. The data includes several sources as the following:

2.1.1 *National Address API*

National Address is a system that was issued by the Saudi Post Corporation as a unified national post addressing system. It covers several levels varying from regions and cities to neighborhoods and buildings. On the other hand, the National Address API is a tool for providing agencies with data services. The API is used to address Khobar's neighborhoods' coordinates.

2.1.2 GeoJson.io

GeoJson.io is a website for creating and showing Geojson files either manually or by loading geo data files. The Khobar's Neighborhoods' coordinates and names are obtained through National Address API as mentioned earlier. Consequently, the Boundaries are obtained through Google Earth as a KML file which is converted to a Geojson file by GeoJson.io. By using Shapely library, the center of each polygon is identified and distances to several venues are calculated.

2.1.3 Foursquare API

Foursquare API is used to request results about venues within each neighborhood. The results include the Universal Transverse Mercator (UTM) coordinates along with the name and category of each venue. Based on a preliminary search, Foursquare showed that it is capable of giving a sufficiently accurate count of the venues within Khobar's neighborhood. Additionally, the venues are separated into different categories with a unique ID for each to group similar venues. Although the results of Foursquare depends largely on users' inputs, which could lead to inaccurate information, the large number of venues should overcome these discrepancies. The data are checked in terms of format, repetition, and misfit of categories through the EDA. All previous steps can give the safe assumption of data to be accurate for fitting the ML model.

2.2 Retrieving Data

2.2.1 Neighborhoods Data

Firstly, the neighborhoods' names and coordinates are retrieved from the National Address API. The UTM coordinates are then converted into the World Geodetic System (WGS 84). This step is important to find the Euclidean distance as the UTM system relies on latitudes and longitudes, both in degrees, rather than X and Y coordinates as in WGS 84. Table 1 shows the resulted data frame of 10 Khobar's neighborhoods. Later, Geojson.io was utilized to generate a geojson file that includes the boundaries of each neighborhood such that visualization maps could be generated like a choropleth map of neighborhoods' distances from Khobar City center as shown in Figure 1.

Table 1 Sample of Khobar's neighborhoods data frame

	Neighborhood	Latitude	Longitude	X	Y	Distance
0	THUQBAH	26.274439	50.191364	419260.957823	2.906329e+06	1466.819086
1	SOUTH THUQBAH	26.273737	50.205405	420662.446460	2.906243e+06	106.211710
2	ISKAN	26.257636	50.208391	420949.699536	2.904457e+06	1709.884239
3	HAMRA	26.226879	50.204523	420542.507086	2.901053e+06	5100.947741
4	TAAWIN	26.224262	50.186858	418776.065391	2.900774e+06	5716.547726
5	KHUZAMA	26.208595	50.186968	418776.199502	2.899039e+06	7372.194020
6	DUGHEITHER VILLAGE	26.274842	50.213668	421488.268917	2.906360e+06	798.957141
7	SAHIL	26.255873	50.217583	421866.439607	2.904257e+06	2216.201050
8	NORTH KHOBAR	26.290838	50.213636	421495.859179	2.908132e+06	2127.867270
9	MADINAT UMAL	26.293403	50.203582	420493.793127	2.908422e+06	2281.309626

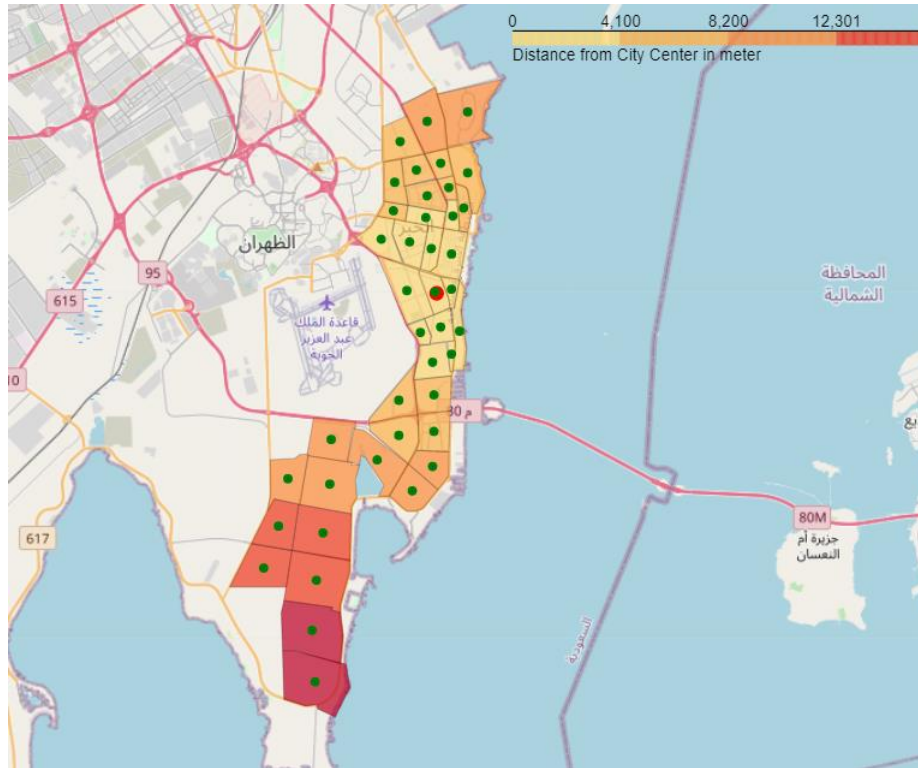


Figure 2 Choropleth map of neighborhoods' distances from Khobar City center

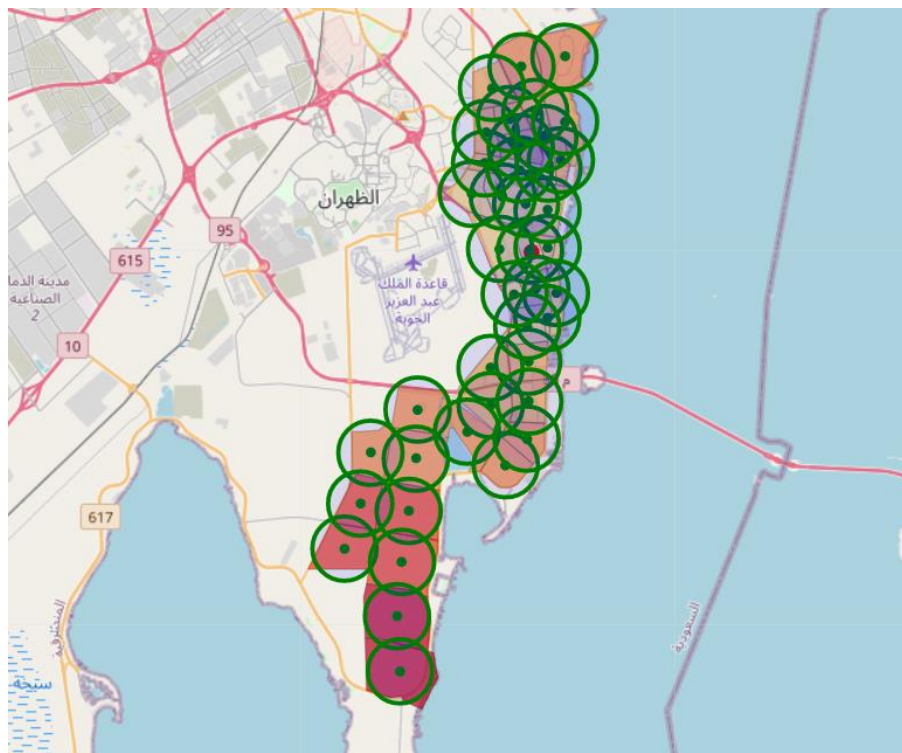


Figure 1 The search area coverage of each neighborhood

2.2.2 Foursquare Data

Based on the project objective, the data retrieved from the Foursquare API must contain venues that are important or relatively important in the decision of the location. The selected venue categories along with the category ID as per the Foursquare API website are shown in Table 2. The table includes additionally factories which could influence the decision as these areas are usually avoidable and unsuitable due to their polluted environment and noises.

Table 2 The selected venue categories and their Foursquare API IDs

Category	ID	Category	ID
Supermarket	52f2ab2ebc57f1066b8b46	Elementary School	4f4533804b9074f6e4fb0105
Middle School	4f4533814b9074f6e4fb0106	High School	4bf58dd8d48988d13d941735
University	4bf58dd8d48988d1ae941735	Food	4d4b7105d754a06374d81259
Parks	4bf58dd8d48988d163941735	Mosque	4bf58dd8d48988d138941735
Gym & Fitness Center	4bf58dd8d48988d175941735	Art & Entertainment	4d4b7104d754a06370d81259
Beach	4bf58dd8d48988d1e2941735	Factory	4eb1bea83b7b6f98df247e06
Government Building	4bf58dd8d48988d126941735	Library	4bf58dd8d48988d12f941735
Medical Center	4bf58dd8d48988d104941735	Auto Workshop	56aa371be4b08b9a8d5734d3
Clothing Store	4bf58dd8d48988d103951735	Grocery Store	4bf58dd8d48988d118951735
Shopping Mall	4bf58dd8d48988d1fd941735	Shopping Plaza	5744ccdf4b0c0459246b4dc

The algorithm to request calls about venues was to select a radius of 1,500m centered at each neighborhoods' center. Such a method could sufficiently cover the area of each neighborhood with repeated venues from other nearby search areas being skipped. In each search area, Foursquare is requested to give 100 calls per each category. Figure 3 shows the coverage of each search area.

The figure shows almost a full coverage of Khobar's City along with some areas of Dhahran City, located directly west of Khobar's City. Although some of the resulted calls are outside of Khobar's city, they are, however, contributing services to the neighborhoods next to Dhahran City. The resulted calls are retrieved as a JSON file which was filtered and read into a data frame as shown in Table 3. Since Foursquare is relatively unreliable in addressing the neighborhood of each venue, Khobar's Geojson file was used to add an additional column about the neighborhood.

Table 3 Sample of filtered Foursquare API calls data frame

	ID	name	lats	lons	cat	sub	Neighborhood
0	577e08ba498ee027e6fb1727	سوبرماركت القرني	26.273430	50.186268	Supermarket	Supermarket	THUQBAH
1	4f413350e4b0740e7b8c7f02	Farm Supermarket (أسواق المزرعة)	26.289101	50.192048	Supermarket	Supermarket	AQRABIYAH
2	571623c8498e0d3bcf4d6465	Meed (ميد)	26.264993	50.192548	Supermarket	Supermarket	THUQBAH
3	57706fb0cd100106d4872c5c	بقالة ماجد	26.264034	50.191147	Supermarket	Grocery Store	THUQBAH
4	5cc7ff1c65211f002c271991	أسواق التلحة العالمية	26.285603	50.193726	Supermarket	Supermarket	AQRABIYAH
...
2694	536b9b03498e1f325ba4336a	موجي	26.264544	50.191902	Food	Burrito Place	THUQBAH
2695	5cc8012416ef67002ca4cf44	مطعم رجال ألمع	26.267555	50.200341	Food	Asian Restaurant	THUQBAH
2696	52dae6a3498e8778e0e7a965	كافيتيريا السندباد	26.267265	50.195027	Food	Burrito Place	THUQBAH
2697	5db153f0fe7ae00008189b6d	بوفيه شواطئ الود	26.261724	50.190409	Food	Fast Food Restaurant	THUQBAH
2698	50a61760e4b0cd103aac5e2b	مطعم ومطبخ نور الحجاز	26.268066	50.197292	Food	Middle Eastern Restaurant	THUQBAH

3 EXPLORATORY DATA ANALYSIS (EDA)

The data frame of venues includes 42 neighborhoods, 20 categories, 211 sub-categories, and 2736 venues. Table 4 shows the count of retrieved venues according to their category. The venues within the food category are highest in the venues data frame. Although some values seem reasonable, others do not clearly represent the density of venues in the neighborhoods. For example, the number of medical centers is high since Foursquare API users tend to register departments within a medical center as separate medical centers. However, the effect of this error does not harm the cause of the clustering neighborhoods since this helps distinguishing medical centers based on their number of provided services. The university category holds the same comment about medical centers. Table 5 shows the number of venues within each neighborhood. The retrieved venues outside Khobar City constitute the highest number among other neighborhoods. This is due to venues search radius being larger than the neighborhood and having Dhahran City next to this search radius. Some neighborhoods have a very low number of venues like 1 or even 10. Although the reality is different, it is accurate that they have far less than the rest of other neighborhoods. Since the objective is only to cluster them, the model will not be affected severely by the lack of an accurate number.

Table 4 Number of retrieved venues within each neighborhood

Count	
Neighborhood	
Outside Khobar	374
NORTH KHOBAR	252
OLAYA	196
AQRABIYAH	153
THUQBAH	128
SOUTH RAKAH	122
YARMOUK	95
JISR	95
SOUTH THUQBAH	94
HIZAM THAHABI	89
HIZAM AKHDAR	83
KHUZAMA	82
KURNAISH	72
MADINAT UMAL	71
DUGHEITHER VILLAGE	71
QURTOBAH	66
BANDARIYAH	65
ANDALUS	59
SINAYIAT THUGBA	47
ISKAN	46
TAAWIN	45
SAWARI	44
RAWABI	40
SAHIL	40
TAHLIYAH	38
JAWHARA	35
SHERAA	33
IBN SINA	25
HADA	25
BUHAIRAH	24
AZIZIYAH	23
BAHAR	20
AMWAJ	17
BUSTAN	16
HAMRA	15
SINAYIAT FAWAZIYA	14
AQIQ	10
RAJA	6
KAWTHAR	3
MAHA	1
LULU	1
KHOBAR HOUSING PROJECT	1

Table 4 Number of venues within each category

Count	
Category	
Food	795
Medical Center	310
Clothing Store	298
Art & Entertainment	256
Gym & Fitness Center	211
Mosque	194
Government Building	90
Auto Workshop	86
Grocery Store	66
Supermarket	65
Factory	61
Beach	59
Shopping Mall	57
Parks	53
High School	44
Elementary School	25
University	23
Middle School	19
Shopping Plaza	15
Library	9

3.1 Commercial Density

To further understand the data, Khobar City is plotted as a heat map of venues. Firstly, the following categories are chosen to represent the venues of commercial nature:

- | | |
|--------------------|-------------------------|
| 1. Shopping Malls | 5. Gym & Fitness Center |
| 2. Clothing Stores | 6. Art & Entertainment |
| 3. Restaurants | 7. Grocery Store |
| 4. Food | 8. Shopping Plaza |

Figure 4.a shows the density in Khobar City. The neighborhoods close to the center in the north of Khobar City show a dense concentration of commercial venues.

3.2 Outdoor Places Density

Outdoor places are considered as they enhance the quality of the residential areas. Parks and beaches are considered within the heat map. As shown in Figure 4.b., The density is concentrated on the north seafront.

3.3 Industrial Density

Industrial places include car workshops and factories. Due to Foursquare not having data about industrial places, Figure 4.c does not precisely specify the industrially concentrated places. However, based on an internet search, the industrial places are more in South Thuqbah and Safaa neighborhoods which is the south slightly dense area in Figure 4.c. Car workshops may not be considered as industrial places but usually, the noise generated from such venues is annoying. Additionally, having both venues on the same category can direct the clustering model to include certain neighborhoods like Safaa together.



Figure 3 Heat maps of venues density.
a): Commercial density, b): Outdoor places density, c): Industrial density

4 METHODOLOGY

4.1 Preparing the Data

To cluster the neighborhoods, a data frame of statistical information about the venue categories in each neighborhood must be created. One approach is to list the frequency of venues of a specific category within each neighborhood. For example, the numbers of venues within each category represent the features of the new data frame while the neighborhoods are the rows. However, since some neighborhoods can be small and within reach of neighborhoods of dense venues, the frequency is not sufficient as a clustering criterion. As an example, the Jawharah neighborhood has a low number of venues but it is very close to Andalus, which has many venues, making the Jawharah neighborhood similar to any empty neighborhood on the edge of Khobar City. Therefore, the average distance to a venue of a specific category is considered. The average distance is calculated by counting the total distance to each venue of a specific category within a radius of 1,500 m from the neighborhood center and dividing it by their number. Such a method takes into account the convenience of reach, demonstrated by the average distance, and the density of these venues, demonstrated by the frequency. Also, The distance from the city center is added to demonstrate the closeness of every part of the city. Table 6 shows the resulted data frame with some of the features and neighborhoods. The resulted data frame has 41 rows and 42 columns.

Table 5 The neighborhood venue categories statistical information data frame

	Neighborhood	Distance from Center	Supermarket Average Distance	Supermarket Frequency	Elementary School Average Distance	Elementary School Frequency	Middle School Average Distance	Middle School Frequency	High School Average Distance	High School Frequency	...	Auto Workshop Average Distance
0	THUQBAH	1466.819086	1029.408236	4	791.430308	5	402.364304	1	851.421823	2	...	1147.817005
1	SOUTH THUQBAH	106.211710	1184.318876	7	672.761062	2	711.390339	2	1077.097359	4	...	962.801319
2	ISKAN	1709.884239	983.416303	3	590.450156	3	316.844829	1	592.979182	4	...	1022.764287
3	HAMRA	5100.947741	0.000000	0	638.814175	1	739.968224	1	0.000000	0	...	637.740252
4	TAAWIN	5716.547726	629.110848	1	0.000000	0	0.000000	0	0.000000	0	...	992.774912
5	KHUZAMA	7372.194020	1110.316188	5	132.557197	1	987.541636	3	413.681532	1	...	943.830309
6	DUGHEITHER VILLAGE	798.957141	623.373109	7	722.587404	1	604.910463	2	353.292399	1	...	1123.816092
7	SAHIL	2216.201050	999.981005	2	895.776229	2	246.341498	1	747.093206	3	...	1048.845419
8	NORTH KHOBAR	2127.867270	1090.853670	11	992.661004	5	1376.832495	5	1381.526813	5	...	1017.753156
9	MADINAT UMAL	2281.309626	905.666583	11	963.600881	7	1379.840574	8	1096.519383	8	...	1249.392836
10	AQRABIYAH	2926.142816	1196.124896	13	919.855511	8	620.351132	8	973.601705	11	...	888.898673
11	OLAYA	3900.760462	891.326634	11	955.131121	3	1083.238316	3	621.746471	5	...	936.887281
12	HADA	4683.326229	1294.790910	11	819.830277	4	633.224874	1	1151.732968	5	...	0.000000
13	HIZAM AKHDAR	3907.359596	1038.639868	9	1314.728669	5	768.208074	5	1262.010087	9	...	481.135444
14	BANDARIYAH	3993.892337	1337.926831	6	1062.084882	3	1275.604550	5	1145.253954	4	...	1220.930701
15	YARMOUK	4524.535883	715.377808	2	745.324436	1	0.000000	0	898.862459	3	...	728.893998
16	BUSTAN	5371.697446	661.844407	5	0.000000	0	0.000000	0	627.875924	2	...	257.115863
17	HIZAM THAHABI	4941.291149	721.789293	5	0.000000	0	0.000000	0	580.008213	5	...	697.049152
18	ANDALUS	5977.472548	965.243428	3	680.568442	1	0.000000	0	1006.713124	4	...	0.000000
19	JAWHARA	6307.257599	1005.320665	3	0.000000	0	0.000000	0	866.922876	4	...	1028.897749
20	RAWABI	6562.698142	807.841703	4	0.000000	0	0.000000	0	1266.571620	3	...	680.956015
21	KURNAISH	6306.033946	838.924512	3	0.000000	0	0.000000	0	0.000000	0	...	0.000000
22	BAHAR	9262.098061	0.000000	0	0.000000	0	0.000000	0	0.000000	0	...	0.000000
23	SOUTH RAKAH	8693.777436	1310.119863	6	421.036366	1	779.019993	2	881.616999	4	...	903.065587
24	KHOBAR HOUSING PROJECT	9064.195999	0.000000	0	0.000000	0	0.000000	0	0.000000	0	...	0.000000

Additionally, a data frame is created by showing the 10 most common venues within each neighborhood. This data frame helps in understanding the result of the clustering and what categories can represent the nature of venues within that neighborhood. However, the clustering model is fitted only by the processed data in Table 6.

Table 6 The 10 most common venue categories within data frame sample

	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	THUQBAH	Food	Medical Center	Clothing Store	Mosque	Gym & Fitness Center	Art & Entertainment	Grocery Store	Auto Workshop	Elementary School	Government Building
1	SOUTH THUQBAH	Food	Medical Center	Clothing Store	Gym & Fitness Center	Art & Entertainment	Mosque	Government Building	Shopping Mall	Grocery Store	Supermarket
2	ISKAN	Food	Auto Workshop	Factory	Medical Center	Mosque	Art & Entertainment	High School	Elementary School	Supermarket	Clothing Store
3	HAMRA	Food	Grocery Store	Mosque	Auto Workshop	Medical Center	Gym & Fitness Center	Clothing Store	Elementary School	Middle School	Factory
4	TAAWIN	Food	Factory	Auto Workshop	Mosque	Art & Entertainment	Gym & Fitness Center	Supermarket	Clothing Store	Grocery Store	Government Building
5	KHUZAMA	Food	Art & Entertainment	Gym & Fitness Center	Mosque	Medical Center	Factory	Government Building	Supermarket	Clothing Store	Auto Workshop
6	DUGHEITHER VILLAGE	Food	Medical Center	Clothing Store	Art & Entertainment	Gym & Fitness Center	Mosque	Government Building	Shopping Mall	Supermarket	Grocery Store
7	SAHIL	Food	Auto Workshop	Factory	Mosque	Art & Entertainment	Clothing Store	Medical Center	High School	Elementary School	Parks
8	NORTH KHOBAR	Clothing Store	Food	Medical Center	Art & Entertainment	Government Building	Mosque	Gym & Fitness Center	Shopping Mall	Grocery Store	Supermarket

4.2 Clustering the Neighborhoods

The chosen clustering model is K-means which is a method of vector quantization that based on an initial number of clusters, the observations are clustered. The data are clustered since there are a lot of features and observations which makes grouping such observations without using machine learning difficult and possibly inaccurate. Clustering is an unsupervised ML technique which provides no classes to test the accuracy of the model. However, the number of desired clusters is set to several values and each result is analyzed to choose the best number of clusters. The factors affecting the number of clusters include the average Silhouette Coefficient and the agreement of previous EDA. The Silhouette Coefficient is a value for each sample that is calculated by the following formula:

$$s = \frac{b - a}{\max(a, b)}$$

Such that:

s : Silhouette Coefficient

b : The average distance between the observation and other observations within the same cluster

a : The average distance between the observation and other observations of other clusters

4.2.1 Two Clusters

The average Silhouette coefficient is 0.29658 which among the others is the highest. The Silhouette Coefficient of neighborhoods shows some deviation from the average coefficient value of each cluster indicating a dissimilarity between like-cluster observations. Additionally, it seems to be clustering only based on commercial density which does distinguish neighborhoods clearly based on other features as shown in Figure 4.c. The colored map of the clustering result is shown in Figure 4.d.

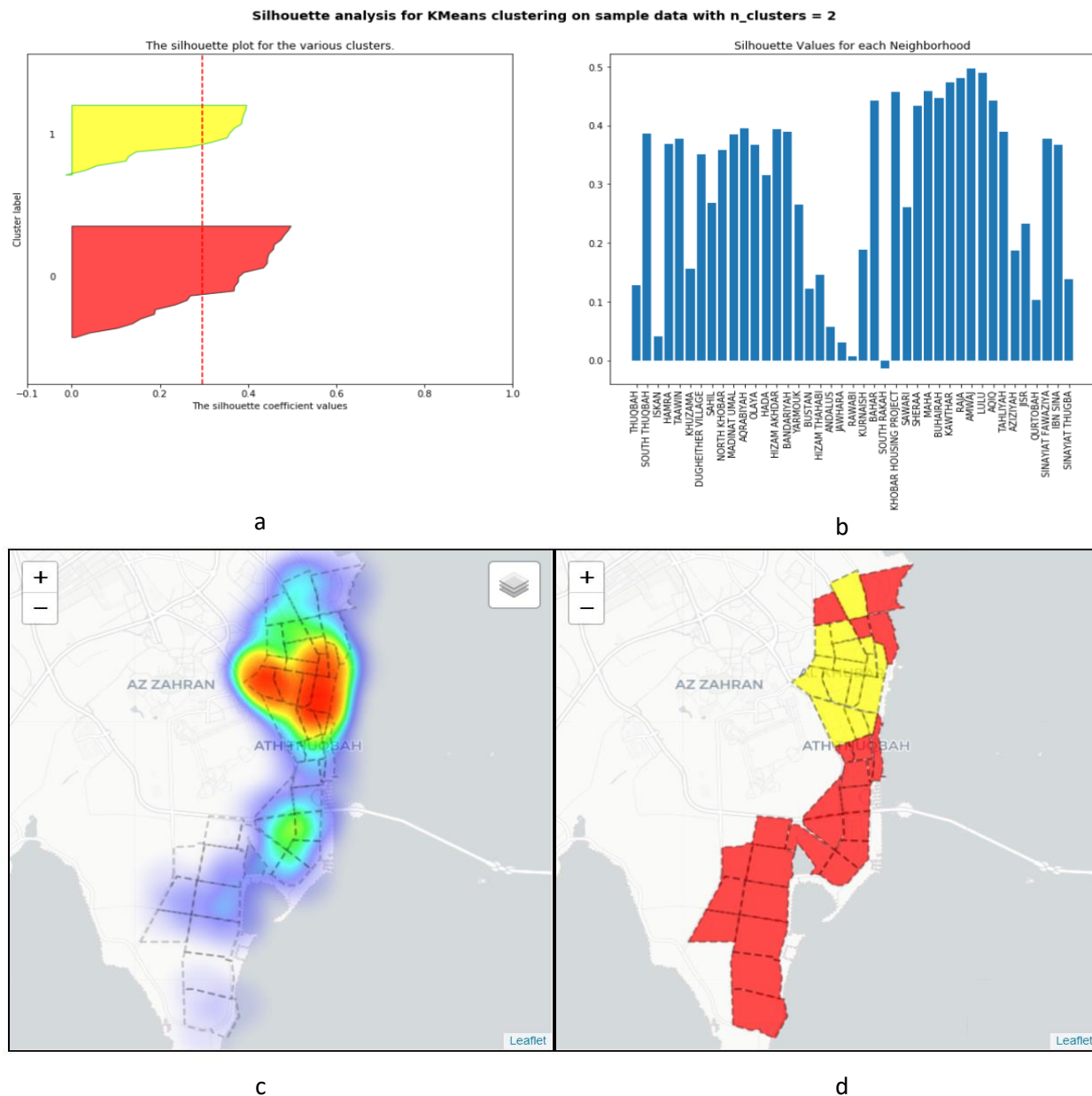


Figure 4 Two clusters option results a): Cluster Silhoutte plot, b): Neighborhoods Silhouette bar graph, c): Commercial density heat map, d): Clusters colored map

4.2.2 Three Clusters

The average Silhouette coefficient is 0.25307 which is relatively less than the previous number of clusters option. The three clusters option provides extra information and fits the heat map of commercial density in Figure 4.c better than the two clusters option. However, it clusters the southern neighborhoods with neighborhoods of higher industrial density as shown in Figure 5.c. The Silhouette Coefficient of neighborhoods is shown in Figure 5.b.

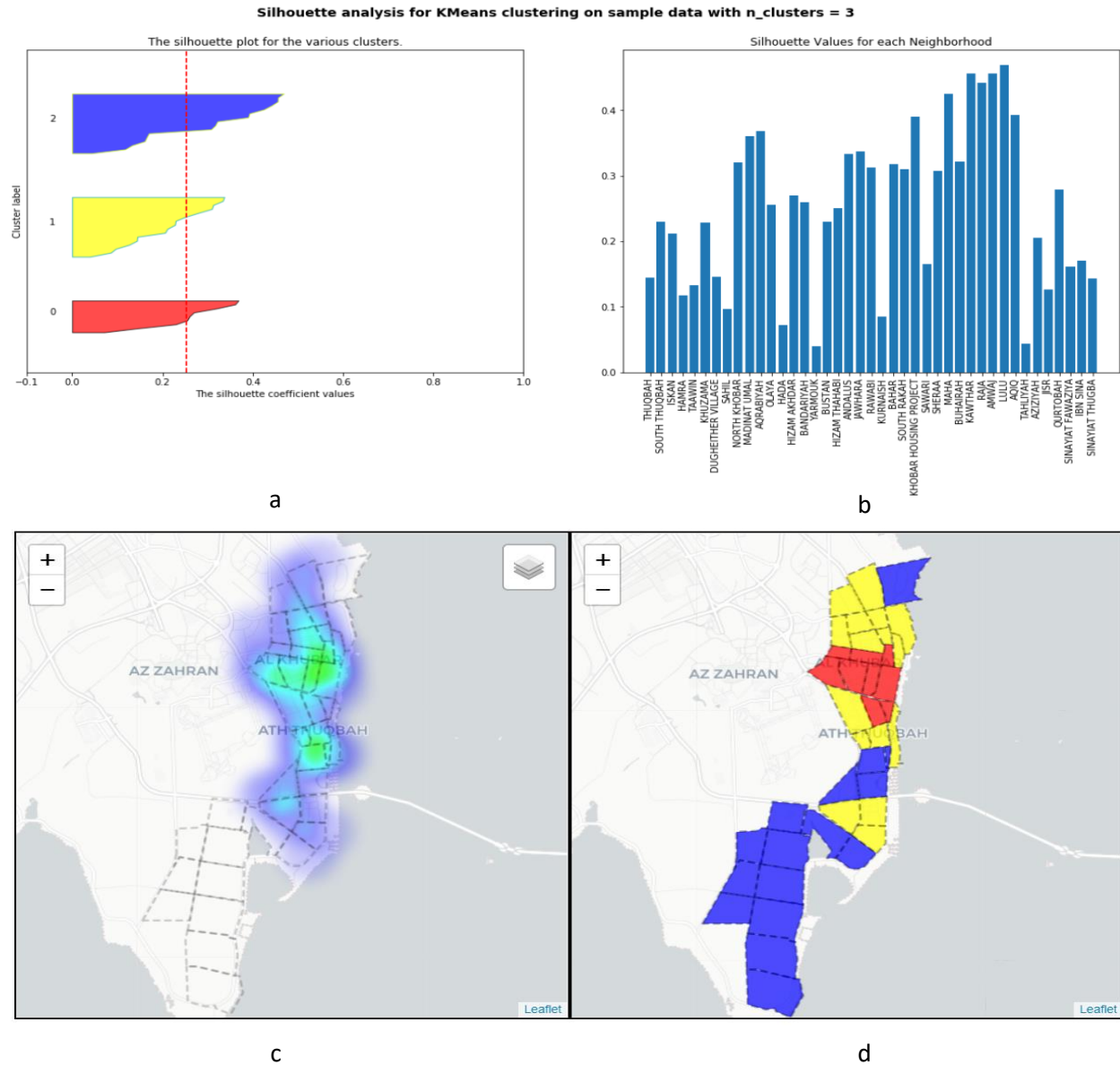


Figure 5 Three clusters option results a): Cluster Silhouette plot, b): Neighborhoods Silhouette bar graph, c): Industrial density heat map, d): Clusters colored map

4.2.3 Four Clusters

The average Silhouette coefficient is 0.230571 which is less than the previous two options. Although the coefficient is less, the clustering visual map fits both the industrial and commercial heat maps. The clustering distinguishes between commercially dense, moderate, and light neighborhoods as shown in green, red, and yellow colors, respectively. Moreover, the blue-colored neighborhood shares a similar higher level of industrial activities. The Silhouette coefficient value for each neighborhood plot shows a variation with few neighborhoods having low values and only one below zero. This option balances between the agreement with previous EDA and having a relative reduction in the average Silhouette coefficient.

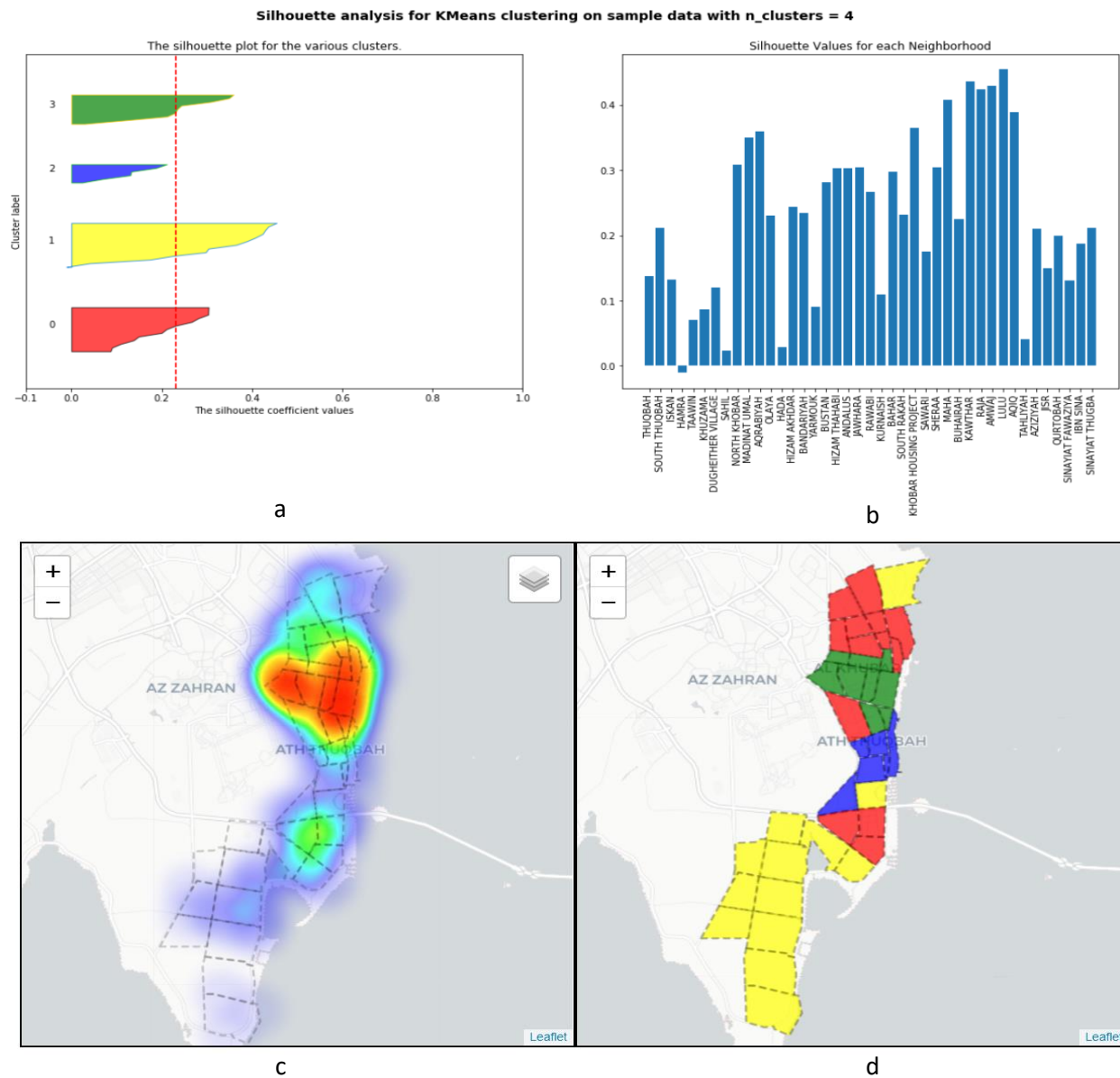


Figure 6 Four clusters option results a): Cluster Silhouette plot, b): Neighborhoods Silhouette bar graph, c): Commercial density heat map, d): Clusters colored map

The average Silhouette coefficient is 0.17970 which is far less than the highest result. Furthermore, many neighborhoods have negative coefficient values, as shown in Figure 7.b, unlike the previous options. The clustering colored map shows a disagreement with the EDA industrial density heat map. Clearly, from and after five clusters, the model would result in lower and lower average Silhouette coefficient values.

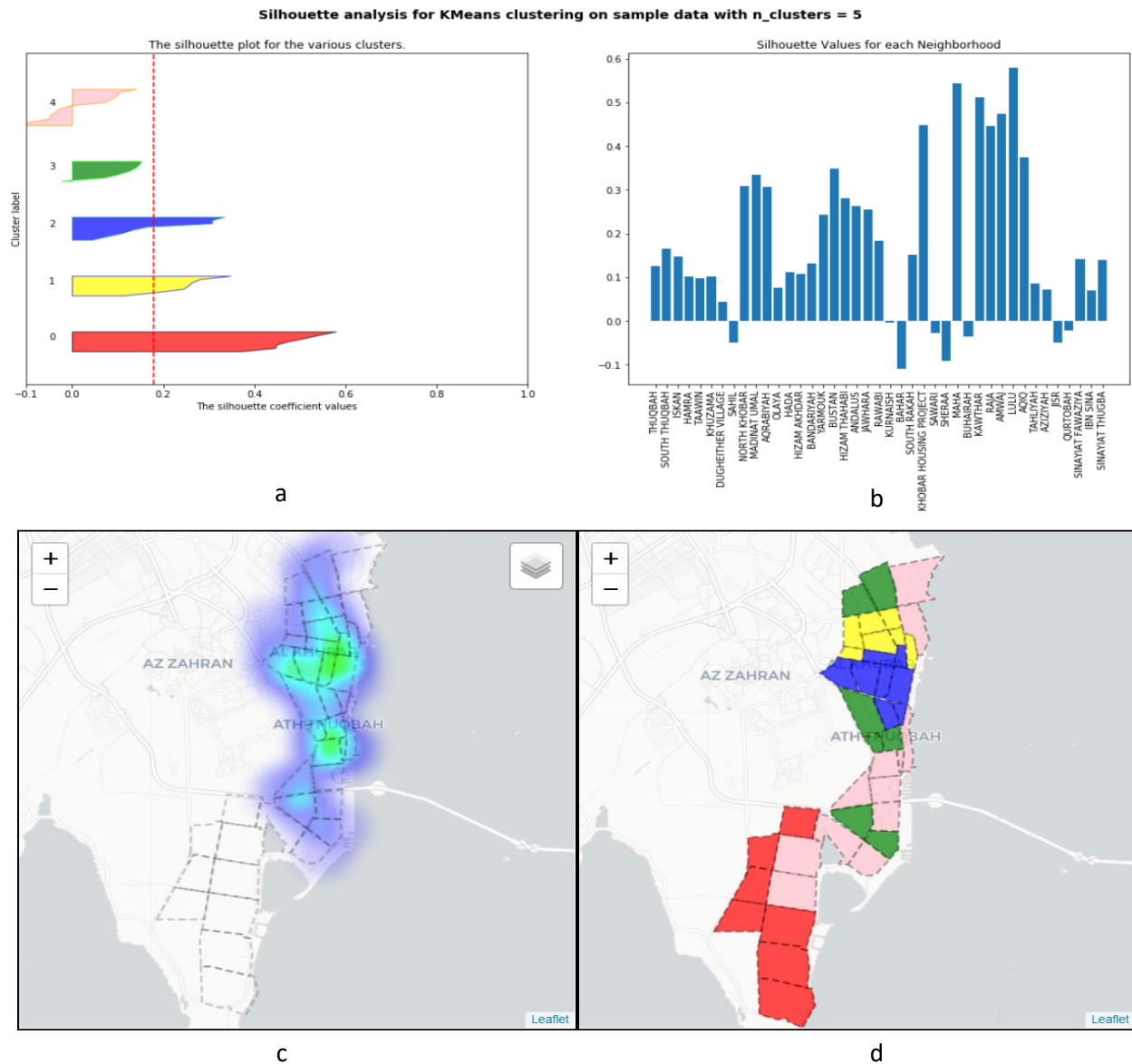


Figure 7 Five clusters option results a): Cluster Silhouette plot, b): Neighborhoods Silhouette bar graph, c): Industrial density heat map, d): Clusters colored map

5 RESULTS

Based on the previous analysis, Khobar City's neighborhoods are clustered into 4 clusters. This option shows the best description of each cluster and matches with the heat map density analysis in the EDA. The following lists the clusters and the neighborhoods within each one:

5.1 First Cluster

The first cluster encompasses 13 neighborhoods that share a moderate commercial density and relatively close to the city's center. Table 8 shows that the neighborhoods do not share a similar common category of venues. Instead, they share the absence of industrial and large commercial venues.

Table 7 First cluster most common venues

	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	Class
0	THUQBAH	Food	Medical Center	Clothing Store	Mosque	Gym & Fitness Center	Art & Entertainment	Grocery Store	Auto Workshop	Elementary School	Government Building	0
5	KHUZAMA	Food	Art & Entertainment	Gym & Fitness Center	Mosque	Medical Center	Factory	Government Building	Supermarket	Clothing Store	Auto Workshop	0
15	YARMOUK	Food	Clothing Store	Medical Center	Art & Entertainment	Gym & Fitness Center	Mosque	Parks	Shopping Mall	Government Building	Grocery Store	0
16	BUSTAN	Food	Medical Center	Gym & Fitness Center	Clothing Store	Mosque	Art & Entertainment	Parks	Shopping Mall	Government Building	Supermarket	0
17	HIZAM THAHABI	Food	Medical Center	Gym & Fitness Center	Art & Entertainment	Clothing Store	Mosque	Parks	Shopping Mall	Government Building	High School	0
18	ANDALUS	Food	Gym & Fitness Center	Art & Entertainment	Mosque	Medical Center	Clothing Store	Government Building	High School	Parks	Supermarket	0
19	JAWHARA	Gym & Fitness Center	Food	Art & Entertainment	Medical Center	Mosque	Clothing Store	Parks	Factory	Government Building	High School	0
20	RAWABI	Food	Gym & Fitness Center	Clothing Store	Medical Center	Art & Entertainment	Mosque	Government Building	Factory	Supermarket	Grocery Store	0
21	KURNAISH	Food	Medical Center	Art & Entertainment	Clothing Store	Gym & Fitness Center	Mosque	Beach	Parks	Government Building	Supermarket	0
23	SOUTH RAKAH	Food	Mosque	Medical Center	Auto Workshop	Art & Entertainment	Gym & Fitness Center	Government Building	Supermarket	Grocery Store	High School	0
35	AZIZIYAH	Food	Mosque	Medical Center	Art & Entertainment	Gym & Fitness Center	Clothing Store	Government Building	High School	Supermarket	Grocery Store	0
36	JISR	Food	Mosque	Medical Center	Gym & Fitness Center	Art & Entertainment	Government Building	Grocery Store	Clothing Store	Supermarket	High School	0
37	QURTOBAH	Food	Art & Entertainment	Gym & Fitness Center	Medical Center	Mosque	Auto Workshop	High School	Government Building	Supermarket	Factory	0

5.2 Second Cluster

This cluster includes 13 neighborhoods with relatively low density of any type of venues and remote location from the city's center. Table 9 shows some similarities with neighborhoods from the first cluster, but they are not similar since the table shows only frequency. Although having similar most common venues, they still are low in density to other neighborhoods in the first cluster.

Table 8 Second cluster most common venues

	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	Class
3	HAMRA	Food	Grocery Store	Mosque	Auto Workshop	Medical Center	Gym & Fitness Center	Clothing Store	Elementary School	Middle School	Factory	1
22	BAHAR	Food	Medical Center	Art & Entertainment	Beach	Grocery Store	Gym & Fitness Center	Mosque	Elementary School	Middle School	High School	1
24	KHOBAR HOUSING PROJECT	Medical Center	Shopping Plaza	Gym & Fitness Center	Elementary School	Middle School	High School	University	Food	Parks	Mosque	1
25	SAWARI	Food	University	Gym & Fitness Center	Art & Entertainment	Beach	Mosque	Supermarket	Government Building	Factory	Library	1
26	SHERAA	Food	Art & Entertainment	Mosque	Gym & Fitness Center	Beach	Government Building	Parks	Supermarket	Medical Center	Factory	1
27	MAHA	University	Shopping Plaza	Shopping Mall	Elementary School	Middle School	High School	Food	Parks	Mosque	Gym & Fitness Center	1
28	BUHAIRAH	Food	Art & Entertainment	Mosque	Auto Workshop	Medical Center	Beach	Elementary School	Middle School	High School	University	1
29	KAWTHAR	Food	University	Shopping Plaza	Shopping Mall	Elementary School	Middle School	High School	Parks	Mosque	Gym & Fitness Center	1
30	RAJA	Food	Art & Entertainment	Medical Center	University	Gym & Fitness Center	Elementary School	Middle School	High School	Parks	Mosque	1
31	AMWAJ	Food	Supermarket	Elementary School	Beach	Gym & Fitness Center	Middle School	High School	University	Parks	Mosque	1
32	LULU	Food	Shopping Plaza	Shopping Mall	Elementary School	Middle School	High School	University	Parks	Mosque	Gym & Fitness Center	1
33	AQIQ	Beach	Food	Shopping Plaza	Gym & Fitness Center	Elementary School	Middle School	High School	University	Parks	Mosque	1
34	TAHLIYAH	Food	Medical Center	Mosque	Supermarket	Grocery Store	Government Building	Gym & Fitness Center	Elementary School	High School	Art & Entertainment	1

5.3 Third Cluster

The third cluster contains 6 neighborhoods with dense industrial spaces as shown in Table 10. The first, second, and third most common venues have either auto workshops or factories as the most common venues. The table shows a presence of other categories as well like schools and medical centers but this is due to the search radius being extended relatively outside all neighborhoods.

Table 9 Third cluster most common venues

	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	Class
2	ISKAN	Food	Auto Workshop	Factory	Medical Center	Mosque	Art & Entertainment	High School	Elementary School	Supermarket	Clothing Store	2
4	TAAWIN	Food	Factory	Auto Workshop	Mosque	Art & Entertainment	Gym & Fitness Center	Supermarket	Clothing Store	Grocery Store	Government Building	2
7	SAHIL	Food	Auto Workshop	Factory	Mosque	Art & Entertainment	Clothing Store	Medical Center	High School	Elementary School	Parks	2
38	SINAYIAT FAWAZIYA	Food	Auto Workshop	Factory	Art & Entertainment	Elementary School	Medical Center	Beach	Mosque	Middle School	High School	2
39	IBN SINA	Auto Workshop	Food	Factory	Art & Entertainment	Mosque	Elementary School	Middle School	Gym & Fitness Center	Shopping Plaza	Medical Center	2
40	SINAYIAT THUGBA	Food	Auto Workshop	Factory	Mosque	Medical Center	Elementary School	Supermarket	Grocery Store	High School	Beach	2

5.4 Fourth Cluster

The fourth cluster has 9 neighborhoods with the highest commercial density among other neighborhoods. Table 11 demonstrates this fact by having commercial categories being in the first 5 most common venues. Additionally, shopping malls are included within the 10 most common venues in almost all of them.

Table 10 Fourth cluster most common venues

	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	Class
1	SOUTH THUQBAH	Food	Medical Center	Clothing Store	Gym & Fitness Center	Art & Entertainment	Mosque	Government Building	Shopping Mall	Grocery Store	Supermarket	3
6	DUGHEITHER VILLAGE	Food	Medical Center	Clothing Store	Art & Entertainment	Gym & Fitness Center	Mosque	Government Building	Shopping Mall	Supermarket	Grocery Store	3
8	NORTH KHOBAR	Clothing Store	Food	Medical Center	Art & Entertainment	Government Building	Mosque	Gym & Fitness Center	Shopping Mall	Grocery Store	Supermarket	3
9	MADINAT UMAL	Medical Center	Clothing Store	Food	Art & Entertainment	Mosque	Gym & Fitness Center	Grocery Store	Government Building	Supermarket	Shopping Mall	3
10	AQRABIYAH	Medical Center	Clothing Store	Food	Art & Entertainment	Gym & Fitness Center	Mosque	Grocery Store	Shopping Mall	Supermarket	Government Building	3
11	OLAYA	Clothing Store	Food	Medical Center	Art & Entertainment	Gym & Fitness Center	Shopping Mall	Mosque	Supermarket	Government Building	Parks	3
12	HADA	Food	Medical Center	Clothing Store	Gym & Fitness Center	Art & Entertainment	Mosque	Supermarket	Shopping Mall	Government Building	Parks	3
13	HIZAM AKHDAR	Medical Center	Food	Clothing Store	Gym & Fitness Center	Art & Entertainment	Mosque	Grocery Store	Parks	High School	Supermarket	3
14	BANDARIYAH	Food	Medical Center	Clothing Store	Art & Entertainment	Gym & Fitness Center	Mosque	Parks	Shopping Mall	Grocery Store	Government Building	3

6 DISCUSSION

Usually, perfect residential areas are located within a neighborhood that is neither high nor low in commercial density. This gives the convenience of reaching them easily while not being annoyed by the large traffic and presence near them. Moreover, they are relatively close to the city center which makes commuting to different parts of the city not difficult. The industrial presence is also a crucial factor that could greatly affect the quality of living. These factors based on the previous clustering result are clearly noticed in the first cluster which makes a group of the most promising neighborhoods for your future house.

7 Conclusion

In conclusion, the most likely suitable neighborhood for a future house is one of the first cluster neighborhoods. Although there are numerous factors for selecting that location, whether personal or financial, the first to consider are included in the first cluster. Consequently, a more in-depth analysis takes place where you can select from different parts of the neighborhood. It is worth mentioning that the analysis does not take into account the availability of residential building permits within all places. Therefore, the analysis was focused more on neighborhoods rather than residential areas. Overall, this can be a good kick-off for the first step into the pursuit of having your dream house.