Where Should I Live?

Elie Massoud 13 May 2020

I- Introduction

Moving to a new city or country is challenging. Whether you are relocating for a dream job in another destination, or short business trip or even spending a vacation, it is necessary to find the best location that matches your lifestyle, requirements and comfort. Relocation hard and time consuming since you are required to spend time exploring your new surroundings trying to find satisfying replacements for your old amenities, activity spots and shops. For instance trying to find a nearby satisfying gym, a good barber shop or a good bar nearby. Another factor to consider from a relocating person is his interests. His interests are determined by his social status; is the relocating person married? With kids? Or he is a bachelor where his interest is centered around bar and fast food restaurants? Is he a tourist or a professional which prefers calm areas with day and night activities venues nearby?
It's good to have everything you need or interest you nearby your place, but a relocating person should also consider rental/purchase price of the apartment. He doesn't want to end up paying all his salary to live in a place where all amenities are nearby.

II- Business Problem

The target audience for this project is a relocating person whether he is relocating due to work purposes, immigration, or family obligations, and the problem lies when the relocating person wants to choose his living destination in a city he knows very little or no idea about. As a solution, this project's objective is to provide guidance and explicatory headline on living in certain areas Beirut city Lebanon, depending on apartments average price/rental yields in areas of the city, the relocating person's interest, and the variety of venues available nearby.

III- Data Requirements

For this project I will require the following data along with their sources.
1- Name and coordinates of all areas in Beirut city Lebanon.
2- Average price of square meter of a household in Beirut areas.
3- Venues that are nearby every area in Beirut city.
4- The average percentage price of rent per sqm2 in different areas of Beirut.

IV- Data Acquisition and Wrangling

I- Data Acquisition

1- Names and Coordinates of all areas in Beirut city Lebanon.
  o Names of different areas of Beirut are fetched from BLOM Invest and data bank articles, studying real estate in Lebanon Beirut city. The studies includes a list of all areas in Beirut city.
  o For the coordinates, **Nominatim Geocoder** library was used to fetch the latitude and longitude of all areas per names.
2- Average price of square meter of a household in Beirut areas.
  o Average prices of sqm2 of households in the above areas will be fetched from BLOM INVEST and Data Bank (Credit Libanais bank). BLOM INVEST and Credit Libanais are banks who perform in-depth studies on Lebanese real estate sector attempting to lure investors.
  o **Tabula** library in python will be use to read price tables from the studies.
3- Venues that are nearby every area in Beirut city.
  o Will use **Foursquare** API to get nearby venues data in every area.
4- The average percentage price of rent per sqm2 in different areas of Beirut.
  o I contacted RAMCO which is a leading real estate advisor company in Lebanon and they provided average percentage for rent in Beirut which is annual 5% of total sale price of an apartment. So to calculate average rent price of apartment I took average apartment size of rent which is 100 sqm2 and multiplied by price of sqm2 in the area x 5% and then divided by 12 to get the monthly rent price. **Rent formula** of a 100 meter apartment: (100 meter) x (average price of sqm2 in area) = selling price. Then (selling price of apartment) x 5% (0.05) /12 (months) = average monthly rent of apartment.

II-Data Wrangling

First I got list of names of Beirut areas, from the following pdf article: https://blog.blominvestbank.com/wp-content/uploads/2016/02/In-Depth-Review-of-the-Lebanese-Real-Estate-Sector-in-2015.pdf page 12 in the pdf. The pdf not only include the names by also price of sqm2 of apartments in each area. I used tabula library to read from the pdf get areas of Beirut along with price of meter squared. For more consistency some extra locations or areas along with their average price of household sqm2 were added to the list from the following source http://www.databank.com.lb/docs/Credit%20Libanais%20real%20estate%20report%202008.pdf

	Average Price Per SQM	Average Apartment Size (SQM)	Average Apartment Price
Ramlet el Baida	$ 6,750	414	$ 2,794,500
Manara	$ 6,250	424	$ 2,650,000
Beirut Central District	$ 6,000	333	$ 1,998,000
Ain el Mreisseh	$ 5,750	412	$ 2,369,000
Raouche - Sakiet el Janzir	$ 5,400	352	$ 1,900,800
Sursock	$ 5,150	297	$ 1,529,550
Saifi	$ 4,900	179	$ 877,100
Verdun - Ain el Tineh	$ 4,575	313	$ 1,431,975
Clemenceau - Kantari	$ 4,450	329	$ 1,464,050
Tallet el-Khayat	$ 4,250	319	$ 1,355,750
Furn el Hayek	$ 4,250	373	$ 1,585,250
Georges Haimari - Sassine	$ 4,200	358	$ 1,503,600
Caracas	$ 4,125	228	$ 940,500
Koraytem	$ 4,100	295	$ 1,209,500
Hamra	$ 4,000	217	$ 868,000
Gemmayzeh	$ 3,875	336	$ 1,302,000
Sanayeh - Spears	$ 3,850	222	$ 854,700
Mar Mitr	$ 3,375	199	$ 671,625
Sodeco - Monnot	$ 3,350	287	$ 961,450
Mar Mikhael - Kobayate	$ 3,275	169	$ 553,475
Badaro	$ 3,200	200	$ 640,000
Sioufi - Hay Sirian	$ 3,100	202	$ 626,200
La Sagesse - Geitaoui	$ 3,000	164	$ 492,000
Grand Lycee - Hotel Alexandre	$ 2,975	148	$ 440,300
Rizk Hospital - Nazareth - Corniche du Fleuve	$ 2,950	206	$ 607,700

After fetching Beirut areas and average price of household sqm2, Nominatim Geocoder was used to iterate through names of areas and get their coordinates (latitude and longitude). Finally all area names of Beirut city, average price of household sqm2 and coordinates were merge in a single dataframe.

```
beirut_areas.head()
```

	Area	Average Price per SQM	Average Size of Apparment m2	Average Sale Price	Latitude	Longitude
0	Georges Haimari,Sassine	$4,200	358	$1,503,600	33.8873	35.5178
1	Rizk Hospital,Nazareth,Corniche du Fleuve	$2,950	206	$607,700	33.8853	35.5155
2	Ramlet el Baida	$6,750	414	$2,794,500	33.877	35.4805
3	Manara	$6,250	424	$2,650,000	33.8946	35.4721
4	Beirut Downtown	$6,000	333	$1,998,000	33.9013	35.503

All the areas in Beirut were projected on **Folium map**
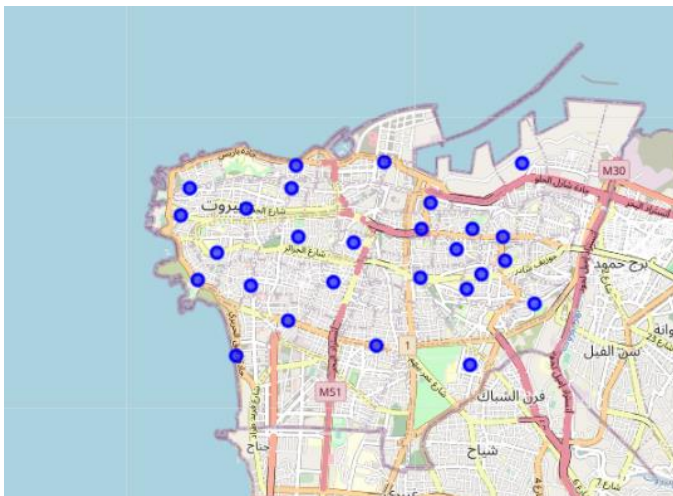
The last part of data acquisition and wrangling is fetching venues around every area in Beirut. So foursquare search call was iterated through all area names returning results of all venues that are nearby each area with 500 meter radius. All the results were stored in a data frame as follows.

[18]:

|	Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	id
0	Georges Haimari,Sassine	33.8873	35.5178	cantina sociale	33.886957	35.518314	Gastropub	4bf58dd8d48988d155941735
1	Georges Haimari,Sassine	33.8873	35.5178	Spinneys Ashrafieh Tilal Beirut	33.888148	35.516906	Convenience Store	4d954b0ea243a5684a65b473
2	Georges Haimari,Sassine	33.8873	35.5178	The Malt Gallery	33.886904	35.515396	Liquor Store	4bf58dd8d48988d186941735
3	Georges Haimari,Sassine	33.8873	35.5178	ABC	33.888474	35.519686	Shopping Mall	4bf58dd8d48988d1fd941735
4	Georges Haimari,Sassine	33.8873	35.5178	Librairie Antoine	33.888588	35.519229	Bookstore	4bf58dd8d48988d114951735
...	...	...	...	...	...	...	...	...
1117	Ras Beirut	33.898091	35.473557	Lamb House	33.895745	35.470536	Restaurant	4bf58dd8d48988d1c4941735
1118	Ras Beirut	33.898091	35.473557	-PitchBlue-	33.901304	35.472828	Nightclub	4bf58dd8d48988d11f941735
1119	Ras Beirut	33.898091	35.473557	Hotel Mediterranee Beirut	33.895496	35.470835	Hotel	4bf58dd8d48988d1fa931735
1120	Ras Beirut	33.898091	35.473557	Corniche El Manara	33.901458	35.474413	Beach	4bf58dd8d48988d1e2941735
1121	Ras Beirut	33.898091	35.473557	Firewall	33.899879	35.477214	Video Game Store	4bf58dd8d48988d10b951735

1122 rows × 8 columns

V-Exploratory Data Analysis and Visualization

As primary analysis, and to understand the data more, a histogram was created to view the total number of venues in every area in Beirut. This histogram portrays which areas has the largest number of venues nearby and which areas have a limited number of venues around. To achieve this visualization, venues fetched from foursquare api should be grouped by area and counted accordingly to get total number of venues by area.

[24]:

	Area	Average Price per SQM	Average Size of Apparment m2	Average Sale Price	Latitude	Longitude	Average Rental Price 100M Appartment	Total No of Venues	Unique Venues
0	Georges Haimari,Sassine	$4,200	358	$1,503,600	33.8873	35.5178	1400	31	25
1	Rizk Hospital,Nazareth,Corniche du Fleuve	$2,950	206	$607,700	33.8853	35.5155	983	32	22
2	Ramlet el Baida	$6,750	414	$2,794,500	33.877	35.4805	2250	7	5
3	Manara	$6,250	424	$2,650,000	33.8946	35.4721	2083	41	23
4	Beirut Downtown	$6,000	333	$1,998,000	33.9013	35.503	2000	100	42
5	Ain el Mreisseh	$5,750	412	$2,369,000	33.9009	35.4896	1916	54	28
6	Raouche,Sakiet el Janzir	$5,400	352	$1,900,800	33.8866	35.4746	1800	27	15
7	Sursock	$5,150	297	$1,529,550	33.893	35.5164	1716	48	29
8	Saifi	$4,900	179	$877,100	33.8929	35.5087	1633	52	26
9	Verdun,Ain el Tineh	$4,575	313	$1,431,975	33.8858	35.4827	1525	47	26
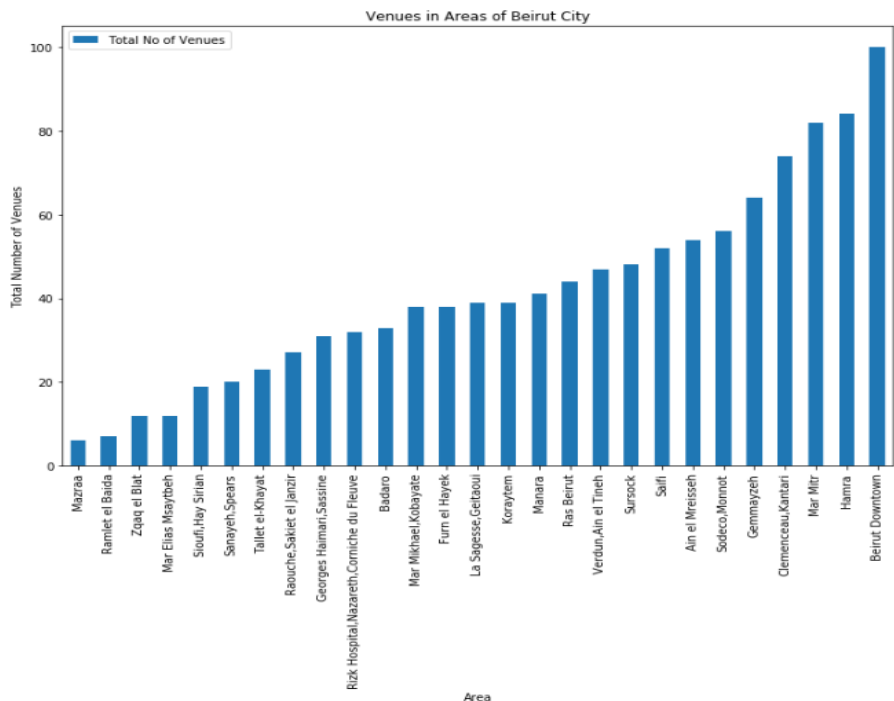
We can notice from the histogram that there is a big difference between areas in Beirut in terms of number of venues located around. Now let check if there is a relation between average prices of household sqm2 in an area and the number of venues located near that area. In other words, are rent prices in an area effected by the number of venues nearby? To achieve this lets fit a regression line and execute the corr function.

```
beirut.corr()
```

|	Average Rental Price 100M Appartment	Total No of Venues	Unique Venues
Average Rental Price 100M Appartment	1.000000	0.311571	0.220643
Total No of Venues	0.311571	1.000000	0.942875
Unique Venues	0.220643	0.942875	1.000000
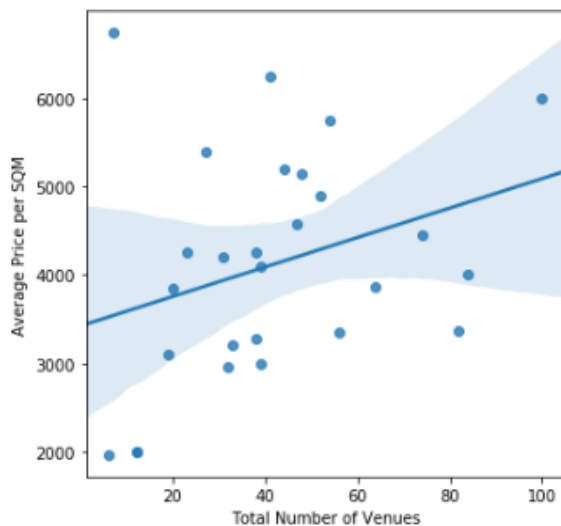
We notice there is a small or insignificant correlation **0.31** between total number of venues and prices of a specific area, as well as you can see that the regression plot is not that 'steep'.
But wait a minute, some areas in Beirut are known of its very high prices of household sqm2 since they are considered the 'Champ-Elise' of Beirut and only the high-class society live there. These areas will be removed from our dataset because the average price of household sqm2 will stay very high no matter what. So now we remove '**Ramlet el Baida**', '**Raouche**', '**Ras Beirut**', and '**Manara**' from our dataset and measure the correlation again.

```
beirut_v.corr()
```

	Average Rental Price 100M Appartment	Total No of Venues	Unique Venues
Average Rental Price 100M Appartment	1.000000	0.623883	0.582063
Total No of Venues	0.623883	1.000000	0.939097
Unique Venues	0.582063	0.939097	1.000000
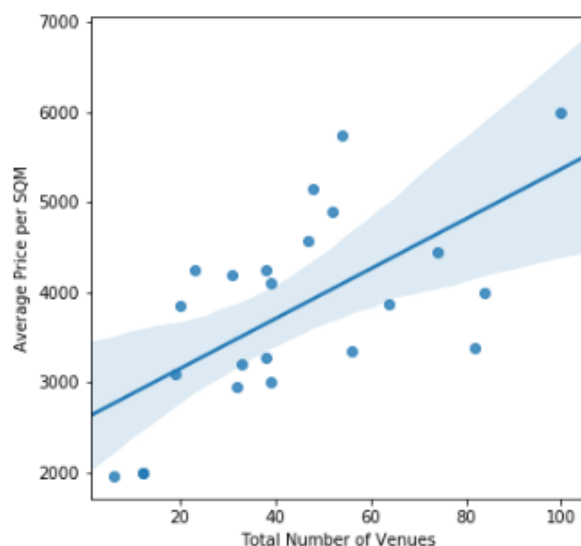
Now we can say that in Beirut, number of venues in an area may positively affect its average price of household sqm2 with a **0.62** correlation.

VI- Modeling and Results

After exploring and understanding the data using visualization, it's time to return to our main problem, where should I live? Earlier we use foursquare to fetch all venues, so we will use these data to create clusters using kmeans clustering algorithm. To achieve this first we create a one-hot matrix of all venues in all categories, then we group by area and calculating the average of each venue.

|	Area	Accessories Store	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	...	Tex-Mex Restaurant	Thai Restaurant	Theater	Toy / Game Store	Turkish Restaurant	Video Game Store	Video Store	Wine Bar	Women's Store	Yoga Studio
0	Georges Haimari,Sassine	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Georges Haimari,Sassine	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	Georges Haimari,Sassine	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Georges Haimari,Sassine	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	Georges Haimari,Sassine	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

```
beirut_grouped = beirut_onehot.groupby('Area').mean().reset_index()
beirut_grouped.head()
```

	Area	Accessories Store	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	...	Tex-Mex Restaurant	Thai Restaurant	Theater	Toy / Game Store	Turkish Restaurant	Video Game Store	Video Store	Wine Bar	Women's Store	Yoga Studio
0	Ain el Mreisseh	0.0	0.000000	0.00	0.000000	0.018519	0.000000	0.018519	0.018519	0.0	...	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
1	Badaro	0.0	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...	0.000000	0.000000	0.030303	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
2	Beirut Downtown	0.0	0.010000	0.01	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
3	Clemenceau,Kantari	0.0	0.013514	0.00	0.000000	0.013514	0.000000	0.000000	0.000000	0.0	...	0.000000	0.013514	0.040541	0.0	0.0	0.0	0.0	0.0	0.0	0.013514
4	Furn el Hayek	0.0	0.000000	0.00	0.026316	0.000000	0.026316	0.000000	0.000000	0.0	...	0.026316	0.000000	0.026316	0.0	0.0	0.0	0.0	0.0	0.0	0.000000

5 rows × 144 columns

We try to cluster the above data but first we should calculate the best k value, so we use the **elbow** and the **silhouette** methods to identify the best k value for clustering

```
For n_clusters = 2 The average silhouette_score is : 0.46397640169564297
For n_clusters = 3 The average silhouette_score is : 0.29352345678008823
For n_clusters = 4 The average silhouette_score is : 0.3106449272189603
For n_clusters = 5 The average silhouette_score is : 0.11515180359875744
For n_clusters = 6 The average silhouette_score is : 0.08024349714739196
For n_clusters = 7 The average silhouette_score is : 0.10658709044791337
For n_clusters = 8 The average silhouette_score is : 0.09591035726852182
For n_clusters = 9 The average silhouette_score is : 0.09664500524393496
```
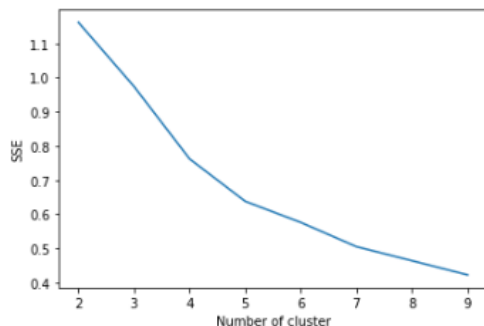

Kmeans clustering doesn't work well on the above dataset since there is no clear 'elbow' in the graph and the silhouette scores are low accordingly.

Difficulty in clustering the areas of Beirut according to the categories of venues around every area, implies that Beirut city is a very diverse city were every area in Beirut contains many amenities and venues of all categories. So whatever the area you want to live in, you'll be partially satisfied. Nevertheless we should cluster Beirut city areas in order to inform the target audience where the best area in Beirut is for them, so we're going to apply a different approach were we will cluster Beirut areas according to the top 10 most common venues only, disregarding other venues. Example if an area has 50 venues around, we will choose only the top 10 most common and delete the others from the dataset to eliminate noise.

To apply this approach first we get the top 10 most common venues for all areas and insert them in a list, then we create a one-hot matrix based on the filtered list, and finally apply the group by area and the mean (average) of venues

category until we get the below dataset. You can notice in below dataset there are only 77 columns (venue category) instead of 144(before filtering), this is because we deleted all categories which are not in the top 10 of any area.

|	Area	Lounge	Bar	Creperie	Yoga Studio	Bank	Bistro	Food Court	Flower Shop	Mediterranean Restaurant	...	Lingerie Store	Food Truck	Garden	Karaoke Bar	Italian Restaurant	Hotel Bar	Gymnastics Gym	Arts & Crafts Store	Beach	Electronics Store
0	Ain el Mreisseh	0.078431	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.0	0.039216	...	0.0	0.0	0.0	0.0	0.058824	0.019608	0.0	0.019608	0.0	0.000000
1	Badaro	0.030303	0.121212	0.0	0.000000	0.0	0.060606	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.030303
2	Beirut Downtown	0.030612	0.000000	0.0	0.000000	0.0	0.000000	0.010204	0.0	0.010204	...	0.0	0.0	0.0	0.0	0.030612	0.000000	0.0	0.000000	0.0	0.000000
3	Clemenceau,Kantari	0.043478	0.000000	0.0	0.014493	0.0	0.000000	0.000000	0.0	0.043478	...	0.0	0.0	0.0	0.0	0.028986	0.014493	0.0	0.000000	0.0	0.000000
4	Furn el Hayek	0.026316	0.026316	0.0	0.000000	0.0	0.000000	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.105263	0.000000	0.0	0.000000	0.0	0.000000

5 rows × 77 columns

Find the k value of KMeans clustering using elbow and silhouette methods.

```
For n_clusters = 2 The average silhouette_score is : 0.678944913617404
For n_clusters = 3 The average silhouette_score is : 0.7423202905714089
For n_clusters = 4 The average silhouette_score is : 0.6987039147276272
For n_clusters = 5 The average silhouette_score is : 0.7120264289289703
For n_clusters = 6 The average silhouette_score is : 0.3678438496987865
For n_clusters = 7 The average silhouette_score is : 0.3671512404444514
For n_clusters = 8 The average silhouette_score is : 0.37488804489230143
For n_clusters = 9 The average silhouette_score is : 0.13817456230371988
```
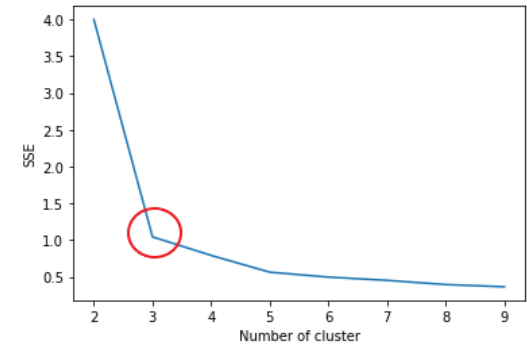

Cleaning the data set from noise and including only the top 10 most common venues in every area showed a huge difference in ability of Kmeans algorithm to cluster the area. Now with the new filtered dataset you can detect a clear value of k=3 based on the obvious elbow and the highest silhouette score. Now it's time to model our dataset, dividing Beirut into 3 cluster. Every cluster includes areas were the common venues of these areas are similar to each other and dissimilar to others.
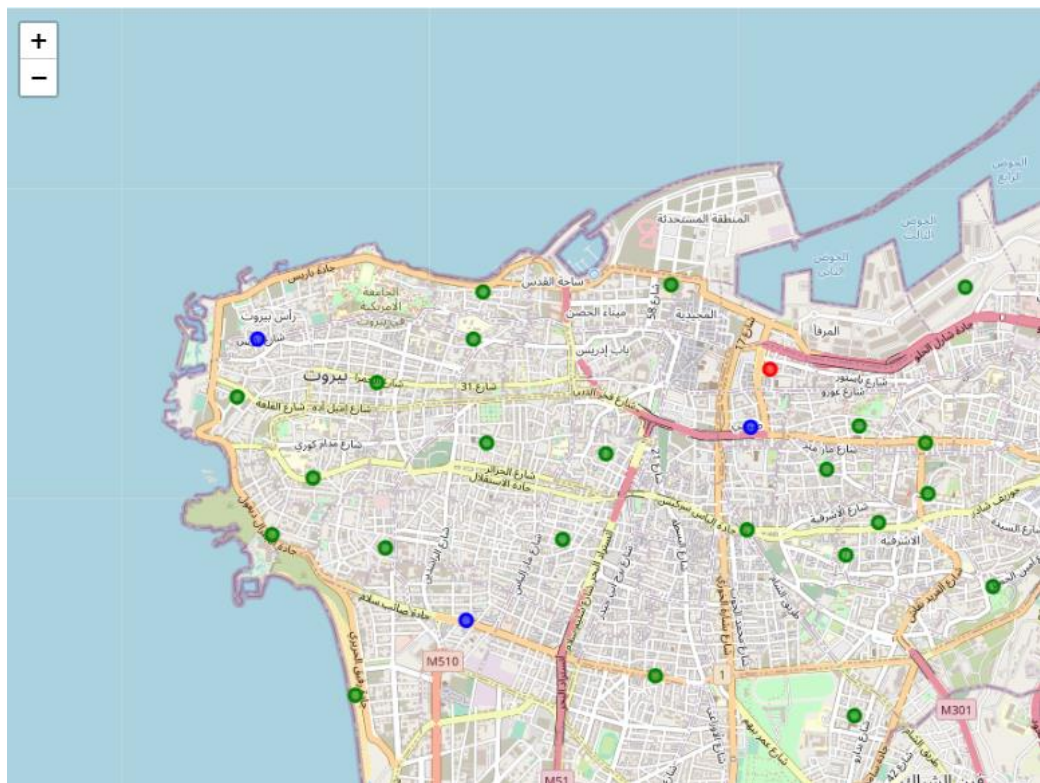
VII- Analyzing Results and Discussions

Kmeans algorithm clustered Beirut areas in 3 categories or clusters. Examining the clusters, we notice that areas in cluster 1 are very diverse, while cluster 3 areas have limited venue categories nearby. Digging deeper into cluster, you can find below venues category in every cluster.

Cluster 1

```
The first cluster venues are
['Lounge', 'Bar', 'Creperie', 'Yoga Studio', 'Candy Store', 'Bank', 'Bistro', 'Food Court', 'Flower Shop', 'Bakery', 'Juice Bar', 'Pool', 'Sandwich Place', 'Bus Station', 'Diner', 'Soccer Field', 'Harbor / Marina', 'Dessert Shop', 'Salon / Barbershop', 'Food Service', 'Gastropub', 'Restaurant', 'Pizza Place', 'Frozen Yogurt Shop', 'Movie Theater', 'Coffee Shop', 'Liquor Store', 'Theater', 'Market', 'Hotel', 'Pedestrian Plaza', 'Gym / Fitness Center', 'Furniture / Home Store', 'Boutique', 'Café', 'Pub', 'Cocktail Bar', 'Fried Chicken Joint', 'Nightclub', 'Steakhouse', 'Art Museum', 'Ice Cream Shop', 'Gourmet Shop', 'Burger Joint', 'Other Repair Shop', 'Shopping Mall', 'Supermarket', 'Grocery Store', 'Music Venue', 'Clothing Store', 'Convenience Store', 'Shop & Service', 'Dance Studio', 'Lingerie Store', 'Food Truck', 'Garden', 'Hotel Bar', 'Arts & Crafts Store', 'Beach', 'Electronics Store']
```

Cluster 2

```
The second cluster venues are
['Dessert Shop', 'Bar', 'Gym', 'Hotel', 'Gym / Fitness Center', 'Café', 'Pub', 'Bistro', 'Restaurant', 'Pizza Place', 'Pharmacy', 'Shopping Mall', 'Coffee Shop', 'Grocery Store', 'Cosmetics Shop', 'Beach']
```

Cluster 3

```
The third cluster venues are
['Bar', 'Café', 'Pub', 'Cocktail Bar', 'Restaurant', 'Pizza Place']
```

We can notice that areas in cluster 1 areas are perfect fit to families settling in Beirut, since all amenities required for daily living are found in areas in cluster 1 such as 'Barber shop', 'Repair shop', 'Furniture/Home store', 'Market', 'Clothing store', 'Electronics Store'…etc. Moreover cluster 1 includes venues for family activities such as 'Pedestrian Plaza', 'Garden', 'Movie Theater', 'Beach', 'Harbor Marina'…etc.

For cluster 2 it matches short visiting/ or business trip professionals professional who are not interested in venues that are required for everyday living, but are interest in some day activities ('Beach', 'Gym/Fitness Centers', 'Shopping Mall') and some night life('Bar', 'Pub', 'Restaurants', 'Bistro'…etc.)

As for cluster 3, we can observe that venues nearby the areas matches tourists who are interested in nightlife (bars, pubs, restaurants). Areas in this cluster doesn't have venues for daily living requirements.

VIII- Comparing Cluster to Prices of sqm2

Cluster 1		Cluster 2		Cluster 3	
Area	Average Price per SQM	Area	Average Price per SQM	Area	Average Price per SQM
Mazraa	$1,960	Tallet el-Khayat	$4,250	Gemmayzeh	$3,875
Mar Elias Msaytbeh	$2,000	Saifi	$4,900		
Rizk Hospital,Nazareth,Corniche du Fleuve	$2,950	Ras Beirut	$5,200		
Zqaq el Blat	$2000				
La Sagesse,Geitaoui	$3,000				
Sioufi,Hay Sirian	$3,100				
Badaro	$3,200				
Mar Mikhael,Kobayate	$3,275				
Sodeco,Monnot	$3,350				
Mar Mitr	$3,375				
Sanayeh,Spears	$3,850				
Hamra	$4,000				
Koraytem	$4,100				
Georges Haimari,Sassine	$4,200				
Furn el Hayek	$4,250				
Clemenceau,Kantari	$4,450				
Verdun,Ain el Tineh	$4,575				
Sursock	$5,150				
Raouche,Sakiet el Janzir	$5,400				
Ain el Mreisseh	$5,750				
Beirut Downtown	$6,000				
Manara	$6,250				
Ramlet el Baida	$6,750				

For cluster 1 '**Mazraa**', '**Mar Elias Msaytbeh**', '**Rizk Hospital,Nazareth,Corniche du Fleuve**' '**Zqaq el Blat**', '**La Sagess, Geitaoui**', '**Sioufi, Hay Siryan**', '**Badaro**' and '**Mar Mikhael**' areas have the cheapest household square meter price among the cluster.

For cluster 2, areas in the cluster are considered as commercial areas and commercial area sqm2 prices for household are usually expensive. But if you have a business trip, short family visit or a conference then you should search in '**Tallet el Khayyat**' area who has the lowest sqm2 price among areas in the same cluster.

As for tourist who are into nightlife, renting an apartment in '**Gemmayze**' is the place for you!!!

VIII- Conclusion

So back to the question, **where should I live?**

- If you are a relocating family or moving permanently then **'Mar Elias Msaytbeh'**, **'Zqaq el Blat'**, **'La Sagess, Geitaoui'**, **'Sioufi, Hay Siryan'** are areas to search for an apartment.
- If you are a professional coming for a short visit or business trip or a conference then '**Mazraa**', '**Rizk Hospital, Nazareth, Corniche du fleuve**', **'Mar Michael'** are your preferred areas to search for an apartment.
- If you are a tourist, then **'Gemayze'** is the sport for you.