

# Lebanon: Small Country Large Variety

**IBM Professional Data Science Certificate**

**Course 9: Applied Data Science Project**

**Robert Matta**



## Introduction

Lebanon is one of the most culturally diverse countries in the Middle East having over 300 breathtaking touristic sites just within a 10452 km<sup>2</sup> area. Today, Lebanon is facing a very challenging economic crisis, and obviously there is no better way to aid this crisis than by investing in tourism, what was once one of the main sources of income in the country. Whether it is museums, old towns, castles, night-clubs, restaurants or even ecotourism, a tourist can only wonder where to begin in order to enjoy the Lebanese experience to the fullest during his short stay.

Roads in Lebanon are poorly maintained to the extent where traffic is part of the population's daily lives. So a tourist that is unfamiliar with the country could waste a lot of time just to find a place to drink a simple cup of coffee. Nevertheless, the tourist can use some guidance based on their preferences and can be the judge of where to start or stop if a map that states what every town in the country is famous for was provided to them and therefore save time to explore even more places which would be an enjoyable experience for the tourist and a profitable one for Lebanon.

Our project aims to give tourists a better idea about Lebanon and where to go for specific preferences using Data science and machine learning tools by clustering areas based on what are the most common venues. This way a tourist can immediately know where to go for a drink, or for a hike based on his current location.

## Data Description

To consider the problem we can list the datas as below:

- The resource from where I got my Data Set is the following:  
<https://data2.unhcr.org/en/documents/details/45209>
- Initially the data contained some useless columns like municipality ID which was a reference for the UNDP so I cleaned it till it became 5 columns comprising:

Figure 1: Raw Data

	Municipality_ID	Municipality Name_EN	UNION_ID	Union_Full	UOM_Name	Pcode	Location_Name_En	Latitude	Longitude	Governorate	UN_AreaOfOperation	District	CAS_CODE	CAD_CODE	CAS_NAME	Cadastral_Type	Google_M
0	M1	Beirut	NaN	NaN	NaN	LBN11081	Marfaa	33.89690	35.51240	Beirut	Beirut & Mount Lebanon	Beirut	10450	10004	Marfaa	CF	<a href="http://maps.google.com/q=33.8969,35.5124">http://maps.google.com q=33.8969,35.5124</a>
1	M2	Jbail	281.0	Union of Municipalities Caza Jbeil	UOM Caza Jbeil	LBN34120	Jbail	34.12042	35.64711	Mount Lebanon	Beirut & Mount Lebanon	Jbeil	28111	28020	Jbayl	CF	<a href="http://maps.google.com/q=34.12042,35.6471">http://maps.google.com q=34.12042,35.6471</a>
2	M3	Eddé	NaN	NaN	NaN	LBN34085	Eddé	34.13681	35.66345	Mount Lebanon	Beirut & Mount Lebanon	Jbeil	28135	28048	Eddéh Jbayl	CF	<a href="http://maps.google.com/q=34.13681,35.6634">http://maps.google.com q=34.13681,35.6634</a>
3	M4	Ehmej	281.0	Union of Municipalities Caza Jbeil	UOM Caza Jbeil	LBN34086	Ehmej	34.12178	35.78399	Mount Lebanon	Beirut & Mount Lebanon	Jbeil	28211	28030	Ehmej	CF	<a href="http://maps.google.com/q=34.12178,35.7839">http://maps.google.com q=34.12178,35.7839</a>
4	M5	Beje	NaN	NaN	NaN	LBN34035	Beje	34.18018	35.71358	Mount Lebanon	Beirut & Mount Lebanon	Jbeil	28278	28068	Bejeh	CF	<a href="http://maps.google.com/q=34.18018,35.7135">http://maps.google.com q=34.18018,35.7135</a>

Figure 2: Data of interest

	Municipality Name_EN	Latitude	Longitude	Governorate	District
0	Beirut	33.89890	35.51240	Beirut	Beirut
1	Jbail	34.12042	35.64711	Mount Lebanon	Jbeil
2	Edde	34.13981	35.66345	Mount Lebanon	Jbeil
3	Ehmej	34.12178	35.78399	Mount Lebanon	Jbeil
4	Bejje	34.18018	35.71356	Mount Lebanon	Jbeil

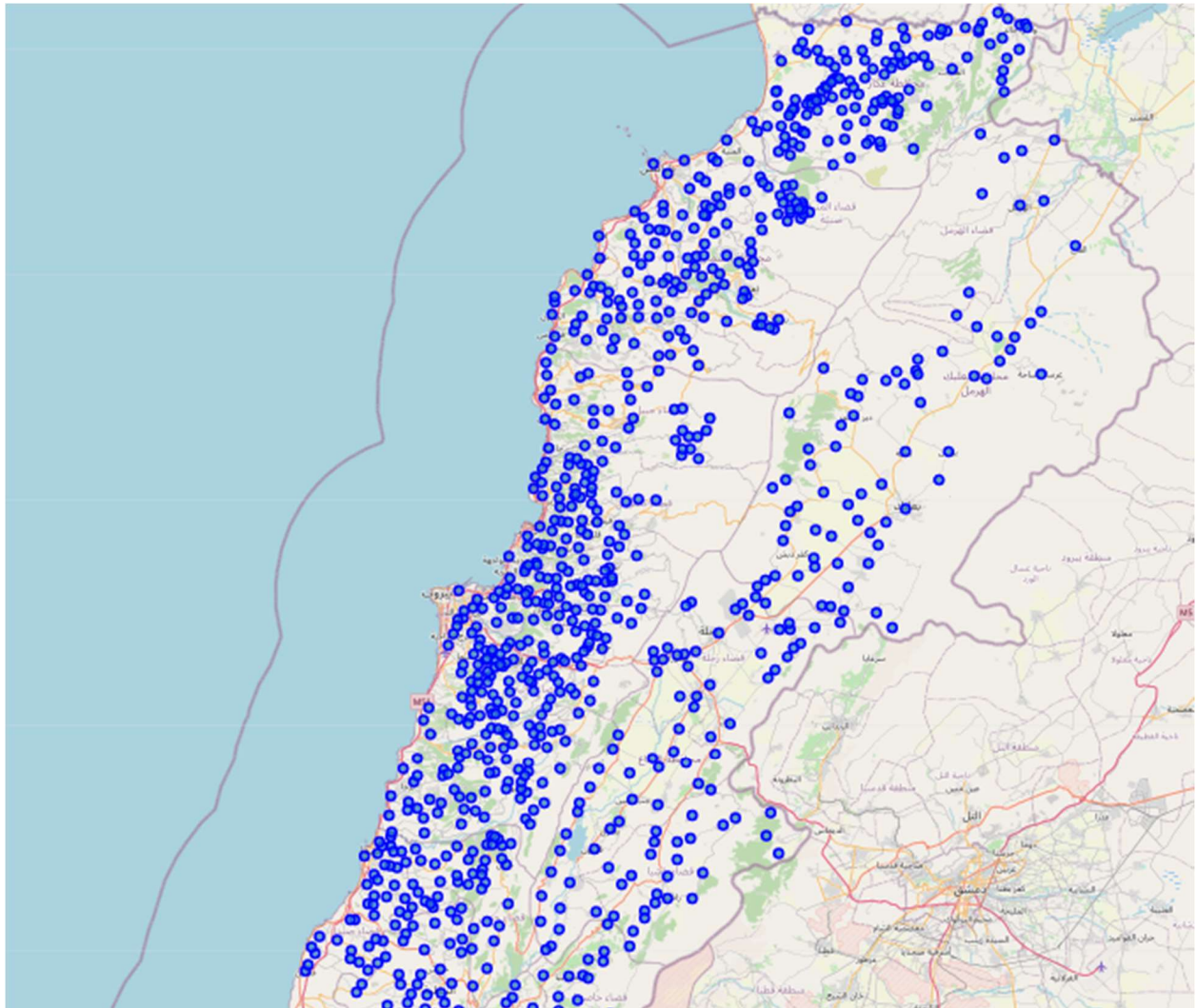
- I used **Forsquare API** to get the most common venues of Municipalities of Mount Lebanon and Beirut.
- Lebanon is a small country so there is no Boroughs. It is rather divided into 8 Governorates (like states in the USA) and those Governorates are then divided into 26 Districts (like Boroughs in the USA) and in each district are many Municipalities (1005 in total). So I cleaned the data accordingly

# Methodology

## 1. Data Visualization

After reducing the dataset to what is necessary, I used python folium library to visualize the geographical location of all municipalities in Lebanon. I used latitude and longitude values to get the visual as below:

Figure 3: Municipalities on the map of Lebanon



Then to get an idea about each governorate and district in Lebanon, I plotted bar charts that show how many municipalities are in every governorate and district

Figure 4: Visualization of the number of municipalities per Governorate (States in the USA)

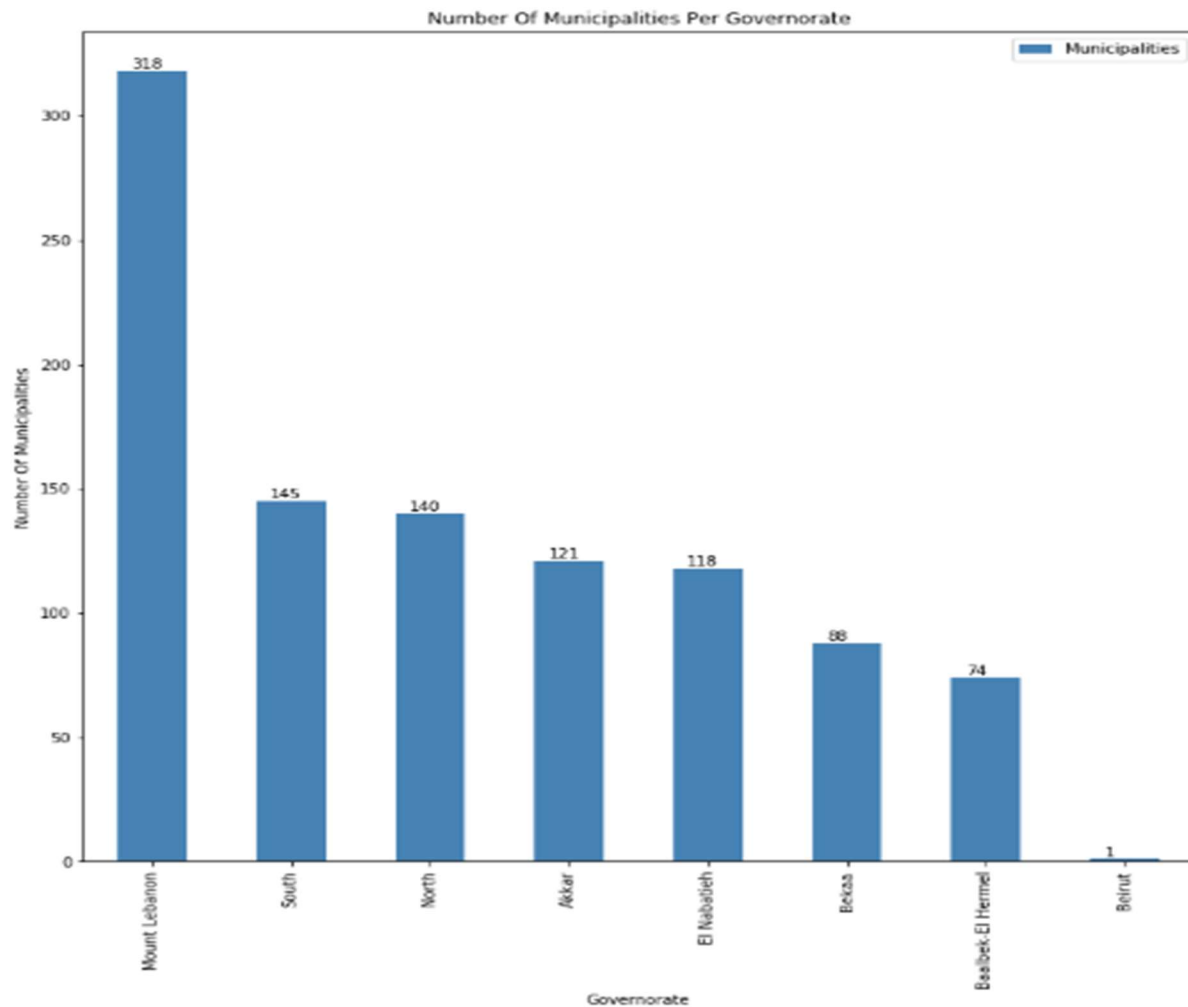
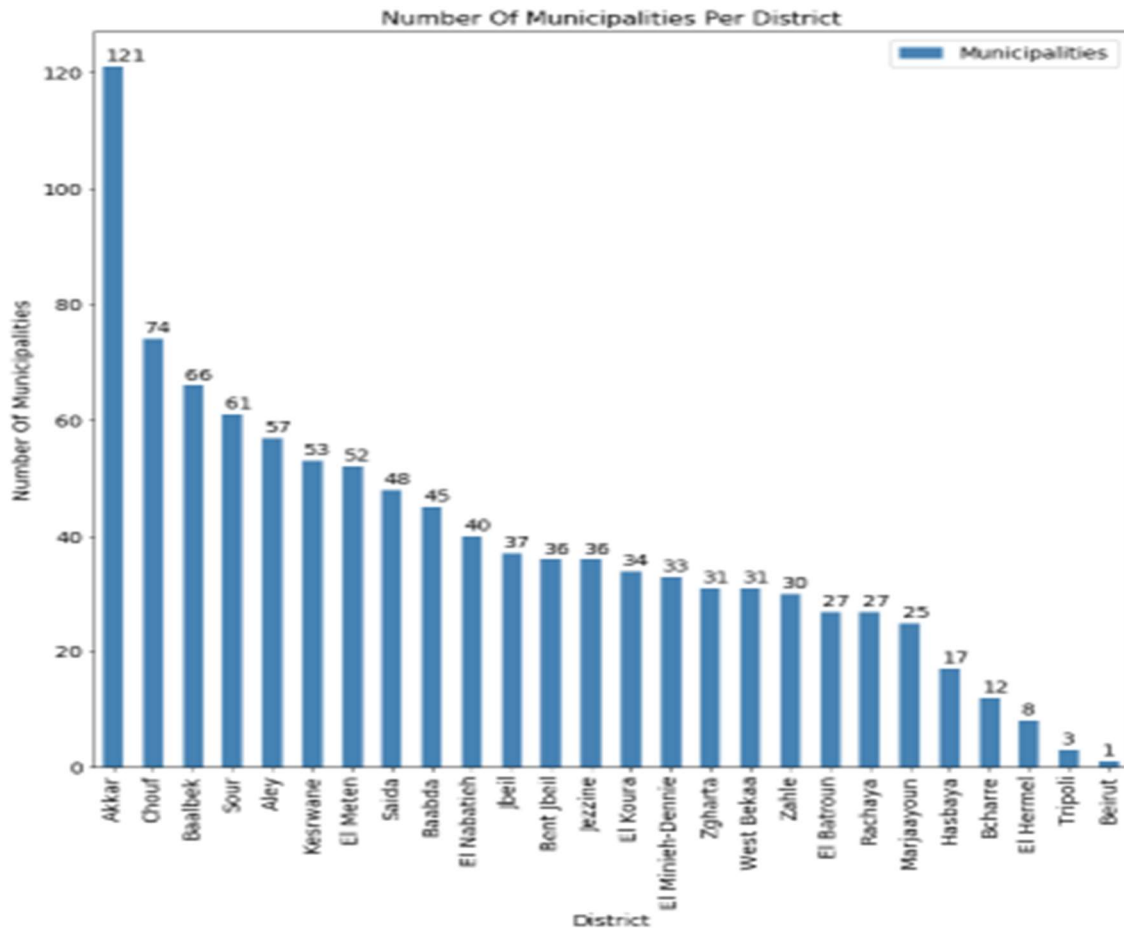


Figure 5: Visualization of the number of municipalities per District (Boroughs in the USA)



## 2. Data Cleaning

Mount Lebanon turned out to be the biggest governorate and Beirut the capital of Lebanon so I narrowed down my data set into those 2 Governorates because of the 500 API calls per day limit. Then used Foursquare API to extract the venues for each neighborhood (municipality in that case) and obtained the following data set

Figure 6: Venues data set

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Beirut	33.8989	35.5124	Lux	33.898216	35.511064	Diner
1	Beirut	33.8989	35.5124	La Mezcaleria	33.897481	35.511032	Mexican Restaurant
2	Beirut	33.8989	35.5124	Freddy's Hotdog	33.897188	35.511741	Hot Dog Joint
3	Beirut	33.8989	35.5124	Le Petit Gris	33.896019	35.510799	French Restaurant
4	Beirut	33.8989	35.5124	4:29	33.896833	35.509955	Pub

Then I decided to plot the top 20 municipalities with most venues and found that some municipalities had less than 5 venues which means that their clustering won't be significant afterwards so I decided to rule them out of my analysis.

Figure 7: Number of venues per Neighborhood

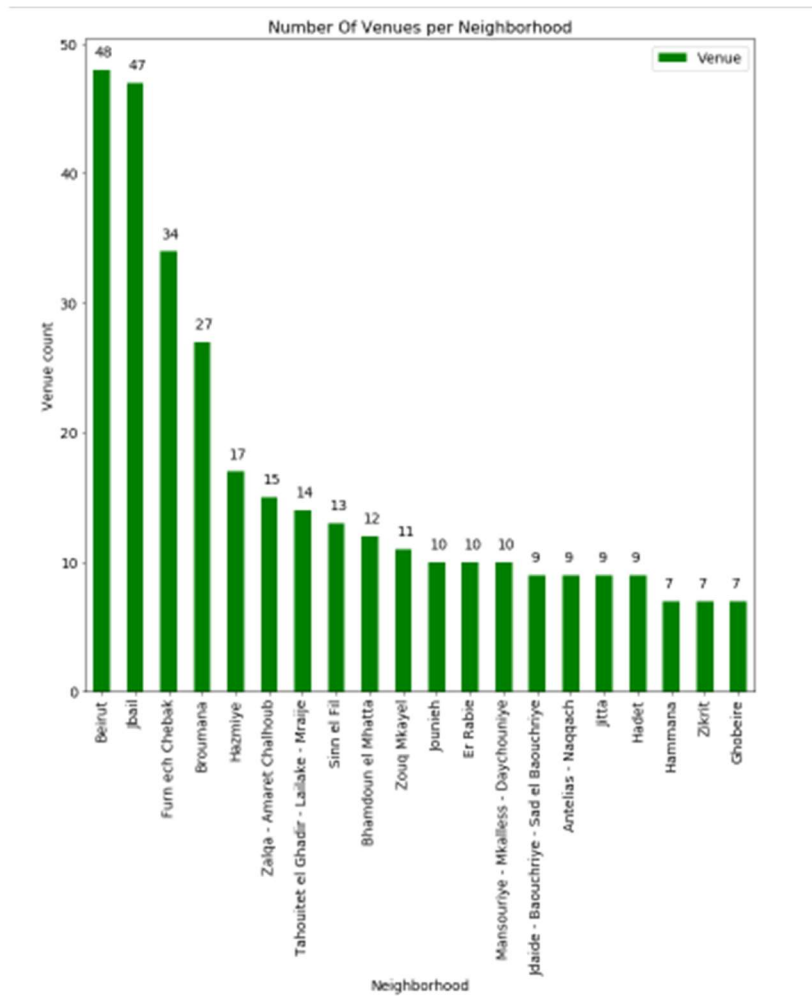




Figure 8: Neighborhoods with less than 5 venues DataFrame

```

In [80]: MLB_waste.head()
Out[80]:
      Neighborhood  Number of Venues
0      Deir el Qamar                4
1  El Aaqaybe Kesrouane                4
2  Bqennaya - Jall Ed Dib                4
3      Kfar Him                      4
4      Kfar Qatra                    4

In [82]: MLB_waste.shape
Out[82]: (175, 1)

```

So there was a total of 174 municipalities with less than 5 venues that were ruled out of the main dataset that had a final number of municipalities equal to 145 and a total number of venues equal to 465 venues.

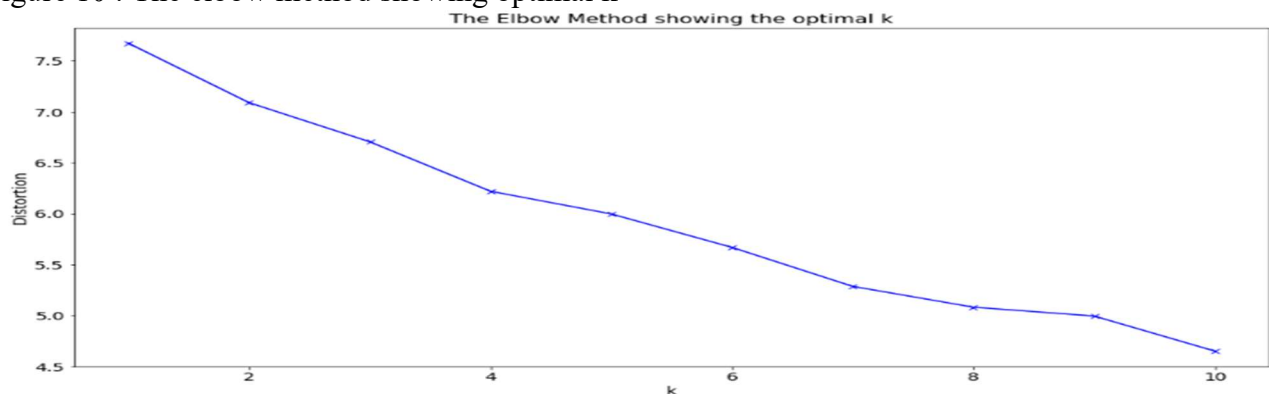
Figure 9: 5 most common venues per Neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Antelias - Naqqach	Gourmet Shop	Mediterranean Restaurant	Road	Karaoke Bar	Bakery
1	Arayia	Campground	Gym	Japanese Restaurant	Mountain	Diner
2	Baaqline	Bakery	Burger Joint	Department Store	Bookstore	Women's Store
3	Baissour	Gift Shop	Tourist Information Center	American Restaurant	Athletics & Sports	Arts & Crafts Store
4	Balloune	Sandwich Place	Pizza Place	Gift Shop	Sushi Restaurant	Pharmacy

### 3. Using K-means clustering

Before starting clustering based on venue categories, one should find the optimal k to have an idea on how many clusters to use by using the elbow method. It turned out to be k=7

Figure 10 : The elbow method showing optimal k





After applying the clustering algorithm using the sci-kit library I added the cluster labels to the cleaned data set as follows:

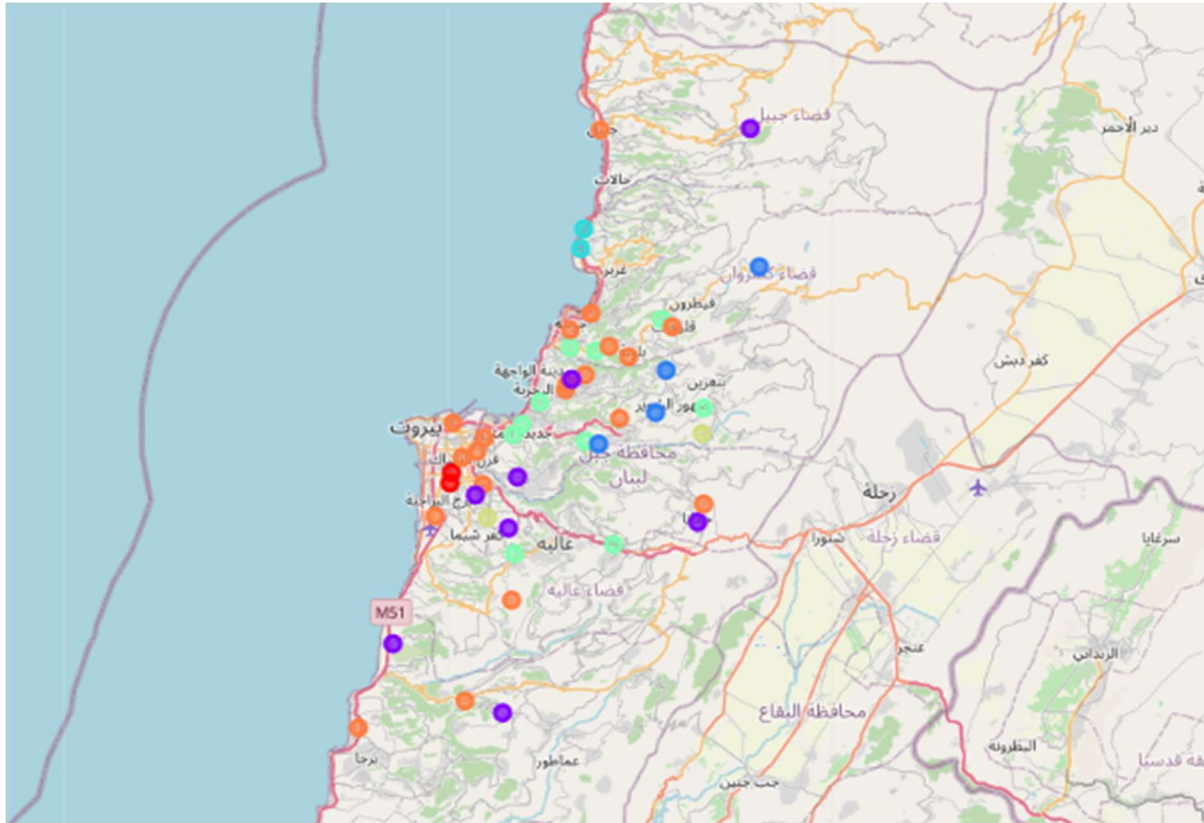
Figure 11: Final dataset after clustering

	Municipality Name_EN	Latitude	Longitude	Governorate	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Beirut	33.89890	35.51240	Beirut	Beirut	6	Pub	Bar	Cocktail Bar	Restaurant	Café
1	Jbail	34.12042	35.64711	Mount Lebanon	Jbeil	6	Mediterranean Restaurant	Pub	French Restaurant	Seafood Restaurant	Bar
2	Ehmej	34.12178	35.78399	Mount Lebanon	Jbeil	1	Mountain	Food Court	Plaza	Bakery	Women's Store
3	Jounieh	33.98231	35.63818	Mount Lebanon	Kesrwane	6	Pizza Place	Surf Spot	Racetrack	Bar	Bakery
4	Balloune	33.94877	35.67224	Mount Lebanon	Kesrwane	6	Gift Shop	Pharmacy	Sushi Restaurant	Pizza Place	Sandwich Place

## Results

After clustering the data the final mapping of the clusters of municipalities is as follows:

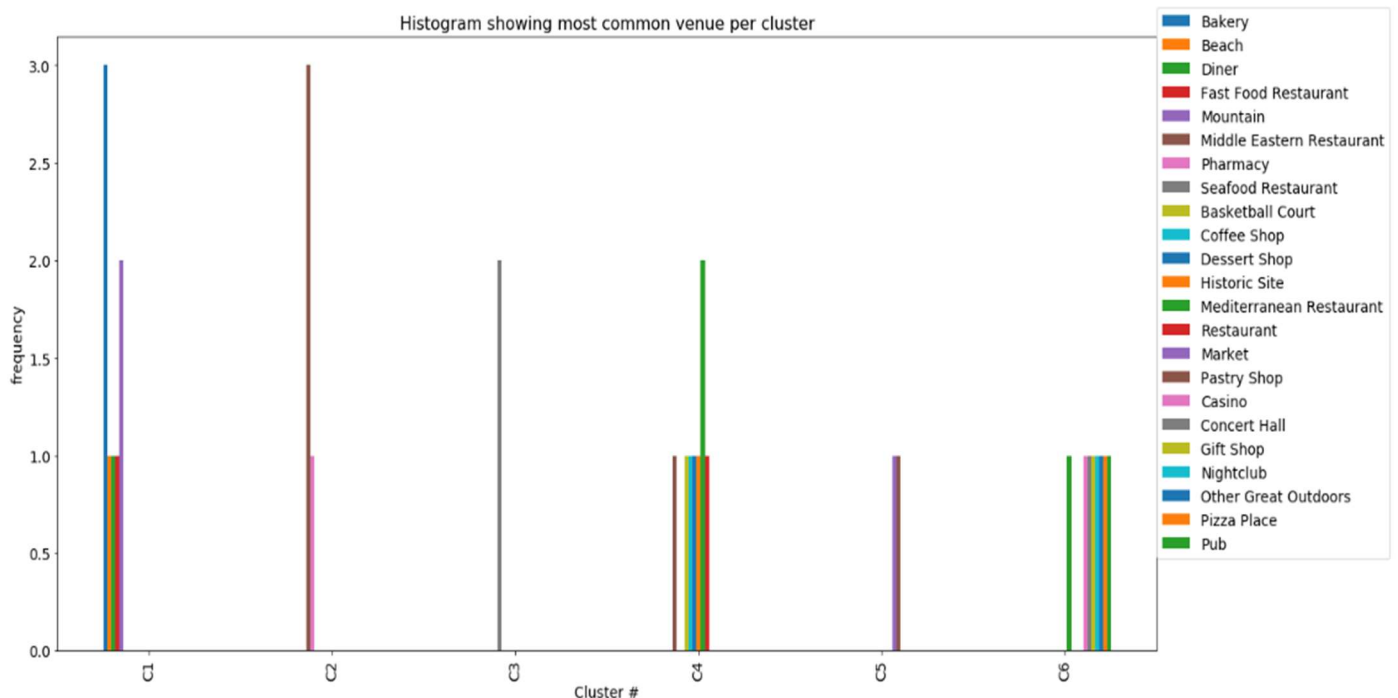
Figure 12: Clusters on the map



- The orange clusters are mostly on the coast line
- The green clusters are concentrated on the coastline of baabda and kesserwan districts
- The purple clusters are concentrated around Beirut
- The blue clusters are mountainous terrains
- The turquoise clusters are on the coastline of kesserwan district
- 2 yellow dots situated a bit randomly

And the most important result which is the meaning of each cluster based on the 1<sup>st</sup> most common venue of each municipality

Figure 13: Histogram showing most common venue per cluster



We have found that:

Cluster 1 (C1): is the necessities cluster: markets, restaurants, bakeries

Cluster 2 (C2): has and middle eastern restaurants







Cluster 3 (C3): has seafood restaurants

Cluster 4 (C4): has a lot of variety concerning nightclubs, pubs and restaurants

Cluster 5 (C5): has markets

Cluster 6 (C6): has also a lot of variety like C4 concerning markets and restaurants but less nightlife

## Discussion

-  Cluster 1: the necessities cluster as discussed in the results are in fact located around Beirut which makes sense given that living around the capital is a good place to live given there is less traffic and you are still near the capital where all the big businesses are concentrated
-  Cluster 2: Mountain cluster, the location was as expected on the map
-  Cluster 3: Seafood restaurants are obviously on the coastline and are apparently famous on the kesserwan coastline
-  Cluster 4: is a cluster with a lot of variety concerning restaurants, nightlife etc... and are abundant in Beirut and around it which is natural given that Lebanon is known for its nightlife
-  Cluster 5 is a bit random however it was found that markets are abundant in those random areas
-  Cluster 6: has as much variety as cluster 4 but less nightlife. It is located in the south of the capital mostly. This may be due to the presence of some quiet neighborhoods in Beirut where people prefer to live to encounter less noise.

## Conclusion

Lebanon is known as a small country with a lot of activities. As we can notice from our research, that taking into consideration only 2 governorates, we were able to notice necessity clusters, mountainous clusters, restaurant clusters and last but not least nightlife clusters. All of this to say that in such a small area it would take a tourist not more than 20 minutes to switch from a cluster to another given that they are almost divided into different zones. More specifically, we have provided a map where one can identify in what cluster he is situated and where to go based on his needs.

For future research, I would recommend to use different API sources more familiar with Lebanese restaurant and nightlife venues alone, because one limitation of this research was the lack of necessary API calls, which lead us to conduct a more general clustering. In other words, future research can tackle the regions where to find specific types of restaurants for hangouts and tourism.