Segmenting and Clustering Jordan

YOUSEF MAJADBI

The scenario of this Capstone project.

Say you live in **Irbid** city. You love your neighborhood, mainly because of all the great amenities and other types of venues that exist in the neighborhood, a great company in **Amman** city. However, given the far distance from your current place such as gourmet fast food joints, pharmacies, parks, graduate schools and so on. Now say you receive a job offer from you unfortunately must move if you decide to accept the offer. It would be great if you can find the most convenient neighborhood in Amman, both in terms of the lowest distance from the company headquarters and in terms of the similarity of the amenities in your home neighborhood

Data acquisition and cleaning:

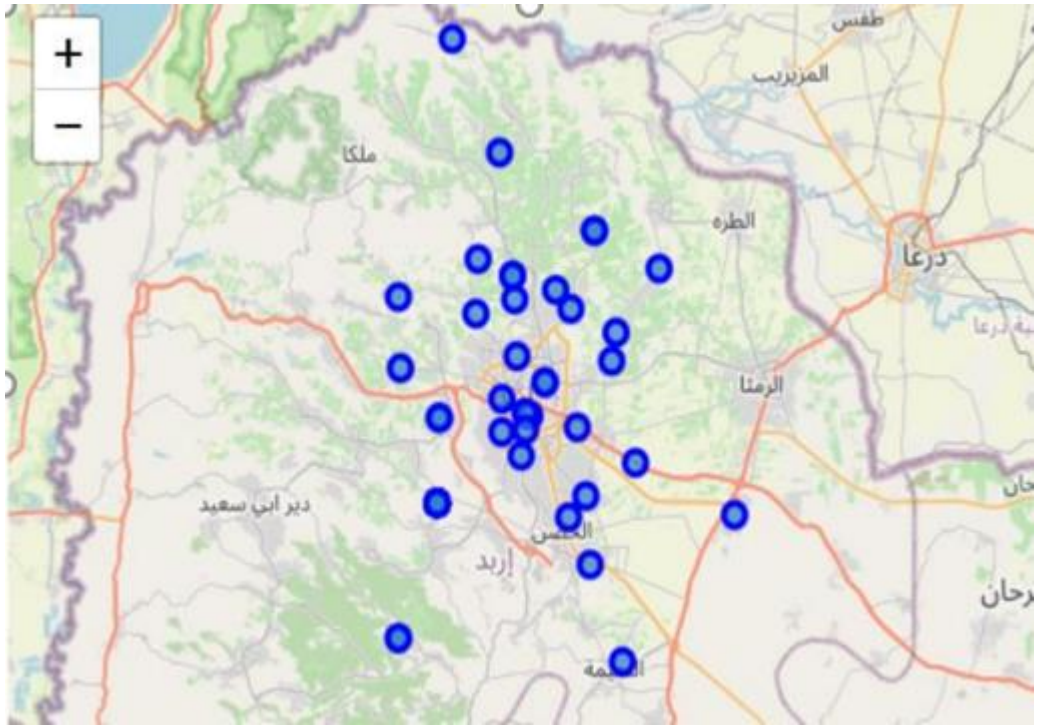
Data Sources:

- List Of All ZIP/POSTAL Codes In JORDAN: The following page was scraped to pull out all the necessary information: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canad a:_M all if the information required for our project are listed in the above link which contains: City name, Neighborhood name, Zip Code for each neighborhood, geographical coordinates (Latitude, Longitude) for each neighborhood. The information obtained i.e. the table of postal codes was transformed into a pandas data frame for further analysis.
- Foursquare API: to collect information about the venues in the neighborhoods of Amman and Irbid.

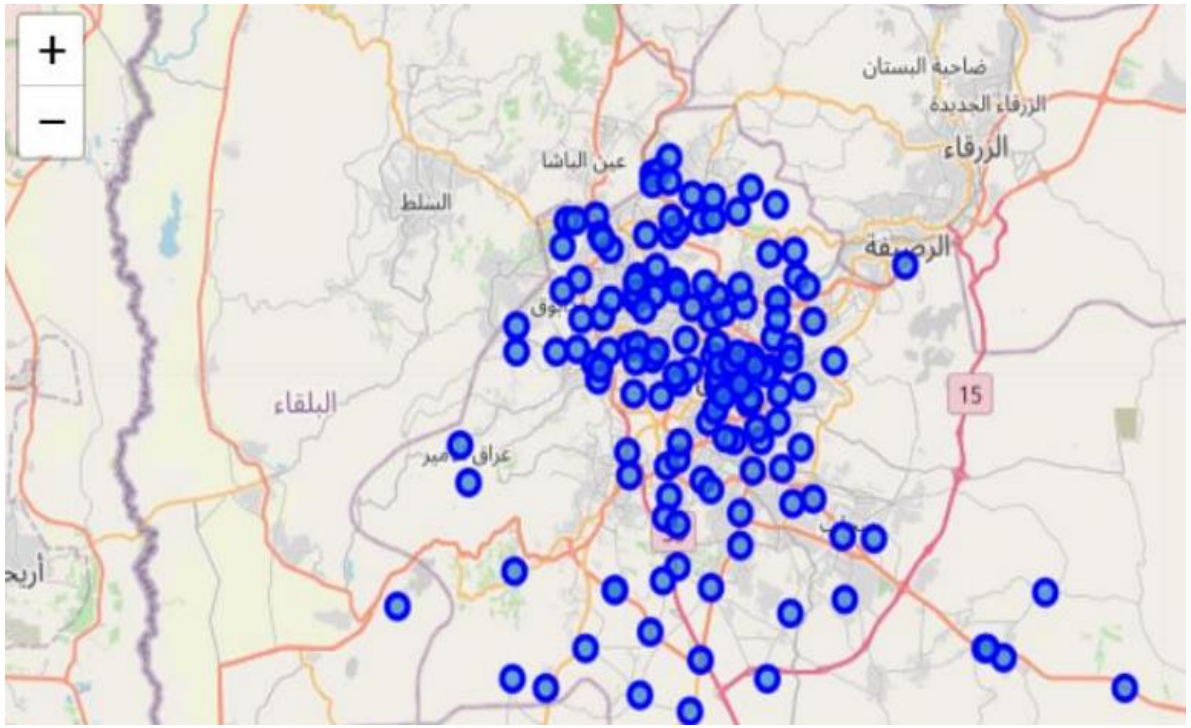
After cleaning the data we have a total rows 459 with 7 columns

Using Folium package for Maps visualization

Irbid Map



Amman Map



Using Foursquare API to analyze neighborhoods venues

Irbid top venues:

Venue Category	
Café	9
Coffee Shop	4
Middle Eastern Restaurant	4
Fast Food Restaurant	4
Asian Restaurant	3
Farm	2
Soccer Field	2

Amman top venues:

Venue Category	
Café	184
Middle Eastern Restaurant	80
Hotel	48
Coffee Shop	47
Dessert Shop	39
Fast Food Restaurant	35
Restaurant	32
Bakery	31
Ice Cream Shop	30
Burger Joint	26

K-Means cluster

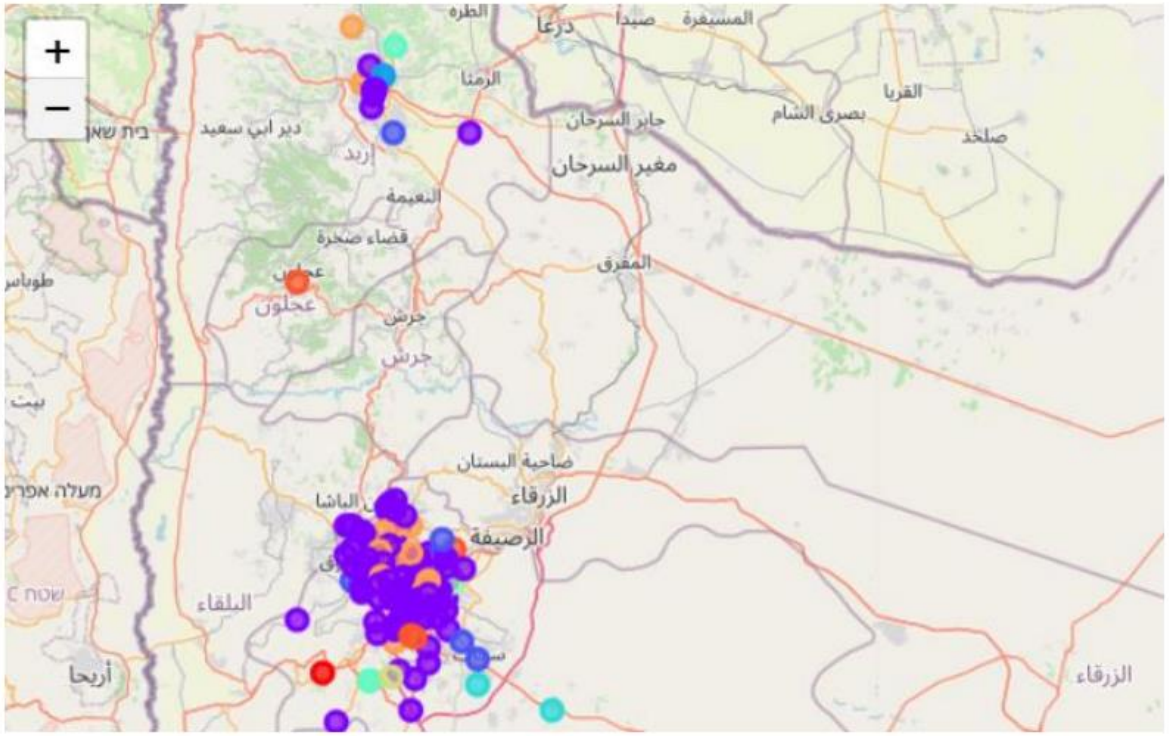
We used the K-Means cluster modeling to create 10 clusters for venues categories.

And by adding the clusters labels to the Data:

	City	City_Arabic	Neighborhood	Neighborhood_Arabic	ZipCode	Latitude	Longitude	Cluster Labels	1th Most Common Venue	2th Most Common Venue	3th 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
0	Amman	عمان	Abdoun Al Janobi (S)	عبدون الجنوبي	11183	31.942011	35.881741	1	Café	Restaurant	Burger Joint	Wings Joint	Gym	Breakfast Spot	Middle Eastern Restaurant	Ice Cream Shop	Salad Place
1	Amman	عمان	Abdoun Alshamali (N)	عبدون الشمالي	11183	31.948469	35.893509	1	Café	Restaurant	Ice Cream Shop	Fast Food Restaurant	Lounge	Burger Joint	Middle Eastern Restaurant	Donut Shop	Coffee Shop
2	Amman	عمان	Abu Alanda	ابو علندا	11592	31.905396	35.960555	1	Department Store	Snack Place	Plaza	Cheese Shop	Supermarket	Dessert Shop	Eastern European Restaurant	Duty-free Shop	Donut Shop
3	Amman	عمان	Abu Alya	أبو علیا	11946	32.001043	35.970014	9	Coffee Shop	Theme Park Ride / Attraction	Diner	Falafel Restaurant	Electronics Store	Eastern European Restaurant	Duty-free Shop	Donut Shop	Doner Restaurant

Clusters Visualization

We used Folium package to create a map visualization for the clusters as below:



Discussion

The goal of this project was to find the most similar and nearest neighborhood to the company headquarter Since the home neighborhood in Irbid is Yarmouk University and its Cluster is 1 We can see that most of the neighborhoods in Amman are on the same Cluster. and the most nearest neighborhood in Amman close to the company headquarter with the same amenities is Jabal Al Hussein Al Gharbi Lets have a look to the common venues in Cluster 1

	Cluster Labels	Neighborhood	1th Most Common Venue	2th Most Common Venue	3th 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
75	1	Jabal Al Hussein Al Gharbi	Café	Ice Cream Shop	Dessert Shop	Shopping Mall	Hookah Bar	Turkish Restaurant	Hotel	Mediterranean Restaurant	Intersection	Eastern R
135	1	Yarmouk University	Café	Donut Shop	Pizza Place	Dessert Shop	Fried Chicken Joint	Middle Eastern Restaurant	Fast Food Restaurant	Burger Joint	Bagel Shop	

Conclusion and References

Conclusion:

Data Science is widely used field, which can be used in a vary real world problems. such as the above one where we used the data to cluster neighborhoods in Jordan country based on the most common venues in those neighborhoods.

References

LIST OF ALL ZIP/POSTAL CODES IN JORDAN: https://gpsarab.com/shop11/en/content/12-zip-code-injordan

Foursqaure API