

Segmenting and Clustering Jordan

Yousef Majadbi

December 19, 2020

A. Introduction

A.1. Background

Amman is the capital of Jordan, with a population of 2,165,925 in 2020 and since it's the capital then most of the jobs opportunities are there and people from different Jordan's cities are coming to work in Amman for that.

Irbid is the 2nd popular city of Jordan with a population of 307,480 in 2020.

A.2. Problem Description:

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, such as gourmet fast food joints, pharmacies, parks, graduate schools and so on. Now say you receive a job offer from a great company in Amman city. However, given the far distance from your current place 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.

A.3 Objective

This project is aiming to analyze the neighborhoods of Amman and Irbid cities and group them into similar clusters. And by analyzing these clusters we can gather meaningful information which will be used to **find out the neighborhoods that are like your current neighborhood**

B. Data acquisition and cleaning:

B.1.1 Data Sources:

We are going to use the below data source to achieve the above objective.

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_Canada:_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.

This first Data Frame is a sample of our data before do any cleaning on it.

	City	Location	Zip Code	Latitude	Longitude
0	Amman عمان	Abdoun Al Janobi (S) عبدون الجنوبي	11183	31.942011	35.881741
1	Amman عمان	Abdoun Alshamali (N) عبدون الشمالي	11183	31.948469	35.893509
2	Amman عمان	Abu Alanda ابو علندا	11592	31.905396	35.960555
3	Amman عمان	Abu Alya أبو عليا	11946	32.001043	35.970014
4	Amman عمان	Abu-Nsair أبو نصير	11937	32.052860	35.876480

B.1.2 Data Cleaning:

Since our data was scraped from one source then it does not have a lot of noise and the below are the only cleaning tasks we did:

- Split the English and Arabic names into individuals columns.
- Rename all columns to a convenient names.

- Remove the additional Spaces at the beginning\ending of values. using strip function

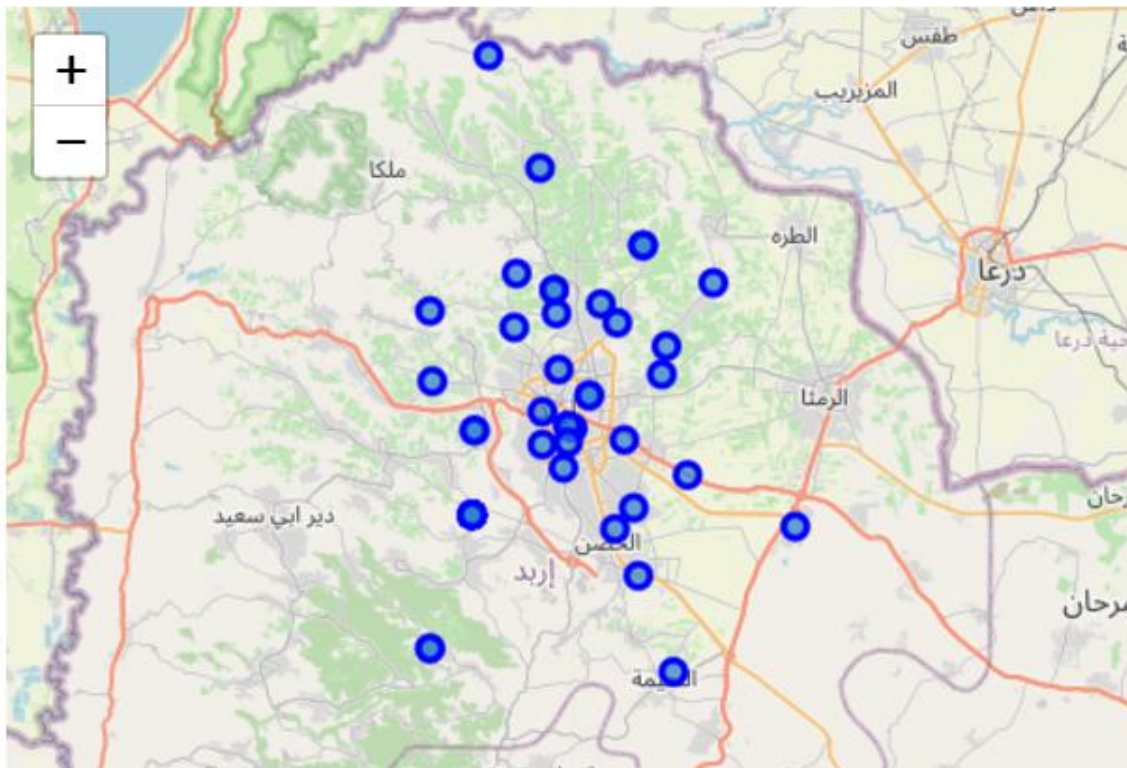
And here is a sample of our data frame after the preprocessing:

	City	City_Arabic	Neighborhood	Neighborhood_Arabic	ZipCode	Latitude	Longitude
0	Amman	عمان	Abdoun Al Janobi (S)	عبدون الجنوبي	11183	31.942011	35.881741
1	Amman	عمان	Abdoun Alshamali (N)	عبدون الشمالي	11183	31.948469	35.893509
2	Amman	عمان	Abu Alanda	ابو علندا	11592	31.905396	35.960555
3	Amman	عمان	Abu Alya	أبو عليا	11946	32.001043	35.970014
4	Amman	عمان	Abu-Nsair	أبو نصير	11937	32.052860	35.876480

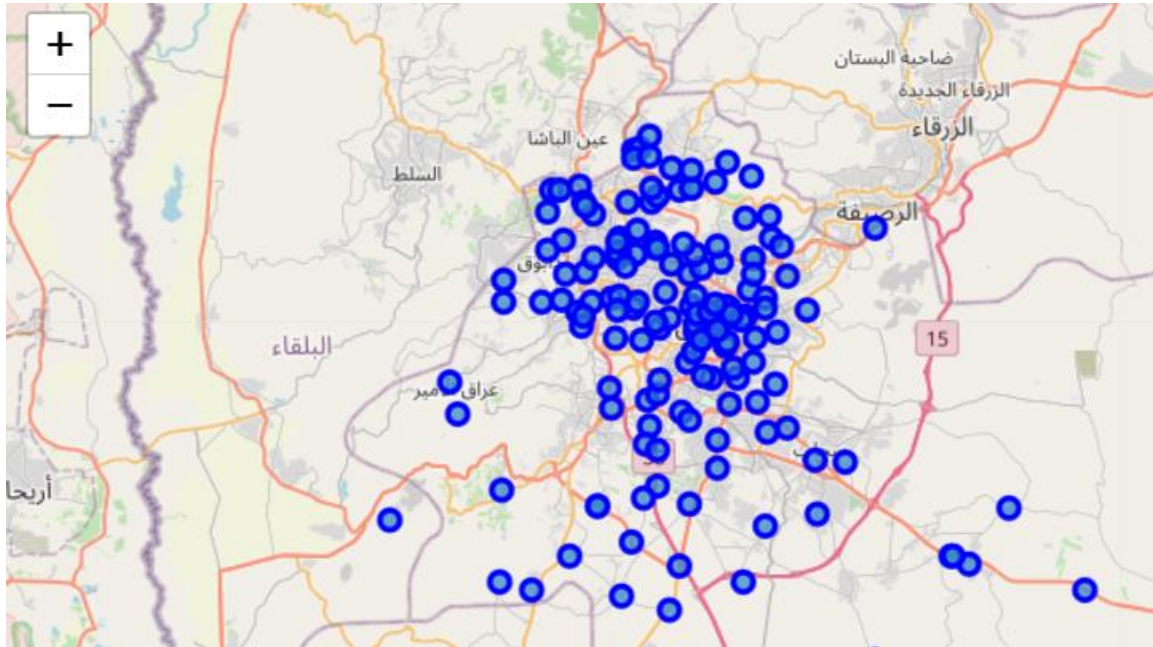
C. Analysis

C.1 First of all let's have a look to the maps of our home city Irbid and the destination city Amman with their neighborhoods:

This map is the Irbid Map with its all neighborhoods.



And this Amman's map with its neighborhoods:



C.2 We used the Foursquare API to pull up the venues of each neighborhood in both cities

This is a sample of the count of each venue in Irbid city sorted by the most frequent:

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

And this is a sample of the count of each venue in Amman city:

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

It's obvious that Cafés category is the top in both cities.

C.3 Now let's prepare our data after getting the venues for each city

We created new DataFrames for each city where this new DataFrame will grouped the Mean of the venue categories for each neighborhood and then to pull the top 10 common frequent categories.

And here is a sample of these DataFrames.

	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
0	Aidun	Pastry Shop	Market	Gym / Fitness Center	Theme Park Ride / Attraction	Farm	Bagel Shop	Burger Joint	Burrito Place
1	Ajloun Community College	Coffee Shop	Campground	Soccer Field	Theme Park Ride / Attraction	Fast Food Restaurant	Bagel Shop	Burger Joint	Burrito Place
2	Al Bariha	Bus Station	Café	Theme Park Ride / Attraction	Fast Food Restaurant	Bagel Shop	Burger Joint	Burrito Place	Campground
3	Al Hai Al Shamali	Sandwich Place	Campground	Theme Park Ride / Attraction	Fast Food Restaurant	Bagel Shop	Burger Joint	Burrito Place	Bus Station

D. Results:

D,1 K-Means Cluster

Now after we analyzed our data and prepared it; clustered the neighborhoods for both cities into 10 clusters using the K-Means Modeling.

D.2 Final DataFrame

After getting the Clusters Labels we added them into the above data frames and then we merged these data frames to the main data frame.

And here is a sample of the final DataFrame.

	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	10th 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	Sandwich 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	Spa
2	Amman	عمان	Abu Alanda	أبو 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	Doner Restaurant
3	Amman	عمان	Abu Alya	أبو عليا	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	Dog Run

D.3 Visualize the Clusters

Below is the **Map** with the **clusters** of venues categories of both neighborhoods

neighborhoods in Jordan country based on the most common venues in those neighborhoods.

G. References:

- LIST OF ALL ZIP/POSTAL CODES IN JORDAN: <https://gpsarab.com/shop11/en/content/12-zip-code-in-jordan>
- Foursqaure API