Data Science Capstone Project

The Battle of the Neighborhoods

Identifying the best location for a new restaurant in Yerevan.

Introduction: Business Problem

Yerevan is the capital and largest city of Armenia as well as one of the world's oldest continuously inhabited cities. Situated along the Hrazdan River, Yerevan is the administrative, cultural, and industrial center of the country. It has been the capital since 1918, the fourteenth in the history of Armenia and the seventh located in or around the Ararat plain. The city also serves as the seat of the Araratian Pontifical Diocese; the largest diocese of the Armenian Apostolic Church and one of the oldest dioceses in the world.

With the growth of the Armenian economy, Yerevan has undergone major transformation. Much construction has been done throughout the city since the early 2000s, and retail outlets such as restaurants, shops, and street cafés, which were rare during Soviet times, have multiplied. As of 2011, the population of Yerevan was 1,060,138, just over 35% of the Republic of Armenia's total population. According to the official estimate of 2016, the current population of the city is 1,073,700. Yerevan was named the 2012 World Book Capital by UNESCO. Yerevan is an associate member of Eurocities.

The objective of this project is to identify the best potential districts where a restaurant can be set up. The analysis and the results of this project would interest stakeholders who are interested in opening a restaurant in Yerevan.

Data

Based on the criteria specified above, the factors that will influence the final decision are:

- Number of existing restaurants in the district
- Number of other places of entertainment, hotels, shops, etc. in the district.
- Distance of the district from the city centre

The following data sources will be needed to extract/generate the required information:

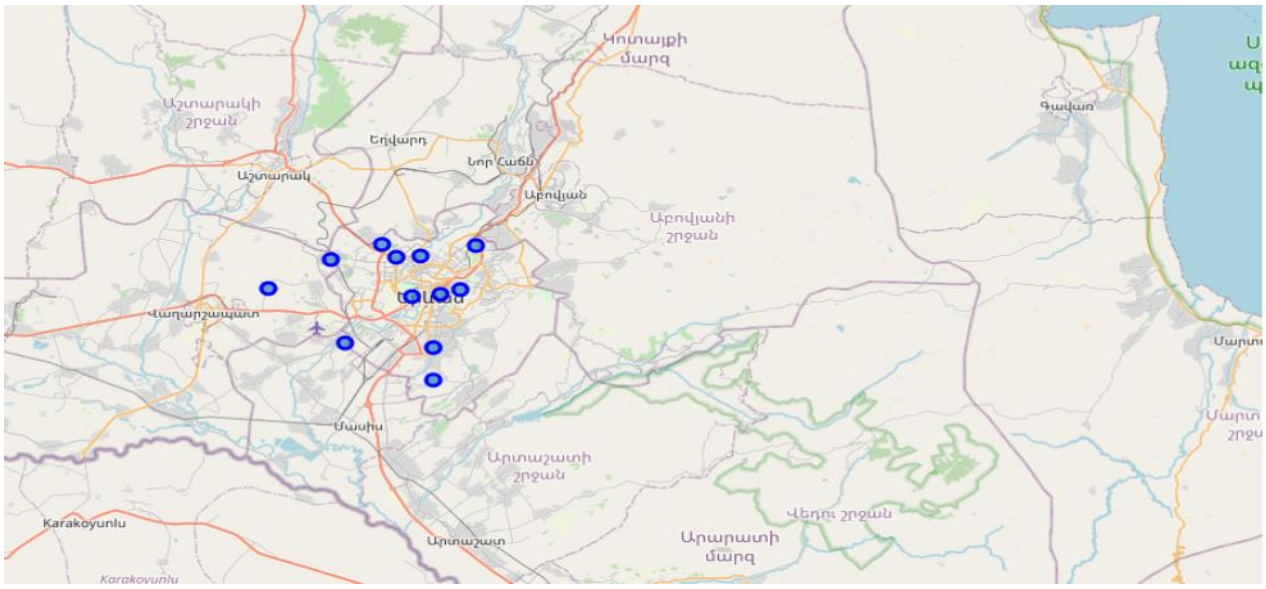
- List of all districts in Yerevan - https://en.wikipedia.org/wiki/Districts_of_Yerevan
- Coordinates of all neighbourhoods and venues - **Google Map** data via csv file.
- Number of restaurants and their location in every district - **Foursquare API** - https://developer.foursquare.com

Analysis

A table of all districts in Yerevan is obtained by scraping the corresponding webpage specified in the Data section. The district coordinates are obtained via csv file, giving rise to a complete district data frame:

	District	Armenian	Population (2011 census)	Population (2016 estimate)	Area (km²)	Latitude	Longitude
0	Ajapnyak	Աջափնյակ	108282	109100	25.82	40.211567	44.416573
1	Arabkir	Արաբկիր	117704	115800	13.29	40.212991	44.480189
2	Avan	Ավան	53231	53100	7.26	40.223600	44.557040
3	Davtashen	Դավթաշեն	42380	42500	6.47	40.224784	44.465119
4	Erebuni	Էրեբունի	123092	126500	47.49	40.131373	44.515156
5	Kanaker-Zeytun	Քանաքեռ-Զեյթուն	73886	74100	7.73	40.214590	44.502848
6	Kentron	Կենտրոն	125453	125700	13.35	40.178103	44.494913
7	Malatia-Sebastia	Մալաթիա-Սեբաստիա	132900	135900	25.16	40.185313	44.355868
8	Nork-Marash	Նորք-Մարաշ	12049	11800	4.76	40.180059	44.522229
9	Nor Nork	Նոր Նորք	126065	130300	14.11	40.183795	44.541669
10	Nubarashen	Նուբարաշեն	9561	9800	17.24	40.102720	44.515233
11	Shengavit	Շենգավիթ	135535	139100	40.60	40.135381	44.430697

A folium map of Yerevan is created:



Foursquare location data is utilized to obtain nearby venues of all districts, which is loaded into a data frame. It makes sense to set up a restaurant in one of the more popular districts so that the restaurant attracts the attention of a larger number of people. Consequently, a list of all the districts with 10 or more venues is acquired. The venues data frame is updated to include only the venues which are in these districts:

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arabkir	40.212991	44.480189	Yerevan Physics Institute Park | Երևանի ֆիզիկա...	40.204106	44.484004	Park
1	Arabkir	40.212991	44.480189	Tashir Pizza	40.218414	44.489403	Pizza Place
2	Arabkir	40.212991	44.480189	Kvartalplatz	40.211204	44.471614	Plaza
3	Arabkir	40.212991	44.480189	Park of Republican hospital "Lich/Լիճ"	40.205533	44.478537	Park
4	Arabkir	40.212991	44.480189	Gevorg Chaush square | Գևորգ Չաուշի հրապարակ	40.208989	44.468457	Plaza

The rows in the above data frame are grouped by district, taking the mean of each category's frequency of occurrence:

	District	American Restaurant	Art Gallery	Asian Restaurant	BBQ Joint	Bakery	Bar	Bed & Breakfast	Bookstore	Border Crossing	...	Tennis Court	Theater	Tram Station	Vegetarian / Vegan Restaurant	Ware
0	Arabkir	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
1	Arabkir	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
2	Arabkir	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
3	Arabkir	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
4	Arabkir	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

5 rows × 85 columns

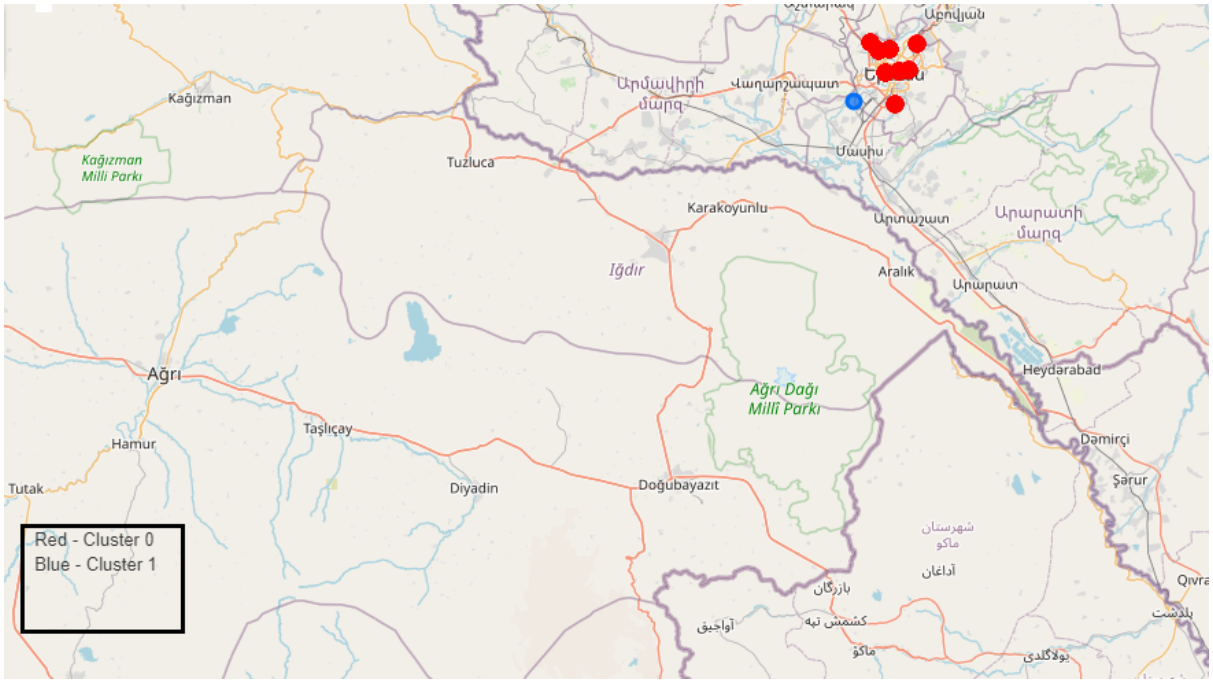
The first step in clustering the districts is to determine the optimal value of K for the dataset. This is carried out using the Silhouette Coefficient Method. It is then discovered that the Silhouette Coefficient is highest when the number of clusters is 2. As a result, the districts shall be grouped into 2 clusters (k=4) using k-means clustering.

A new data frame that includes the cluster, as well as the top 10 venues for each neighbourhood, is created:

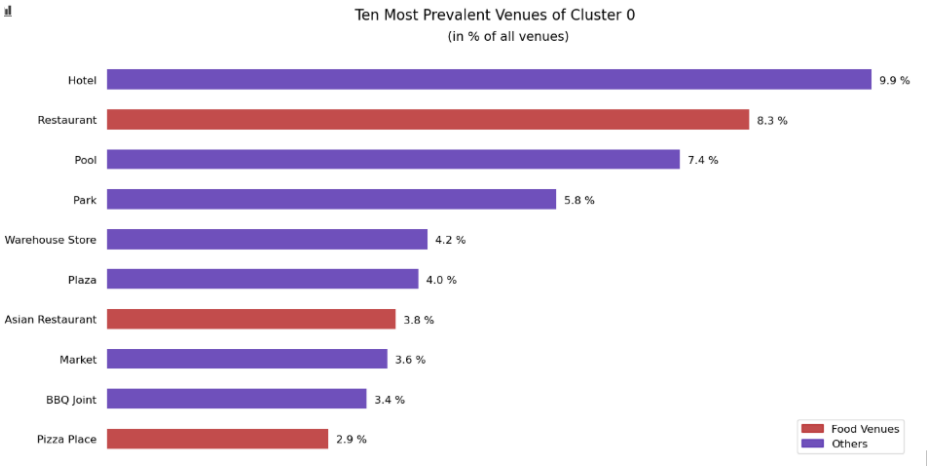
	District	American Restaurant	Art Gallery	Asian Restaurant	BBQ Joint	Bakery	Bar	Bed & Breakfast	Bookstore	Border Crossing	...	Tennis Court	Theater	Tram Station
0	Arabkir	0.000000	0.000000	0.125000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
1	Avan	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
2	Davtashen	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.166667	...	0.000000	0.000000	0.000000
3	Erebuni	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.166667
4	Kanaker-Zeytun	0.000000	0.000000	0.157895	0.052632	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
5	Kentron	0.017241	0.000000	0.017241	0.017241	0.017241	0.017241	0.017241	0.000000	0.000000	...	0.017241	0.017241	0.000000
6	Nor Nork	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	...	0.000000	0.000000	0.000000
7	Nork-Marash	0.000000	0.022472	0.000000	0.000000	0.044944	0.022472	0.011236	0.011236	0.000000	...	0.011236	0.000000	0.000000
8	Shengavit	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000

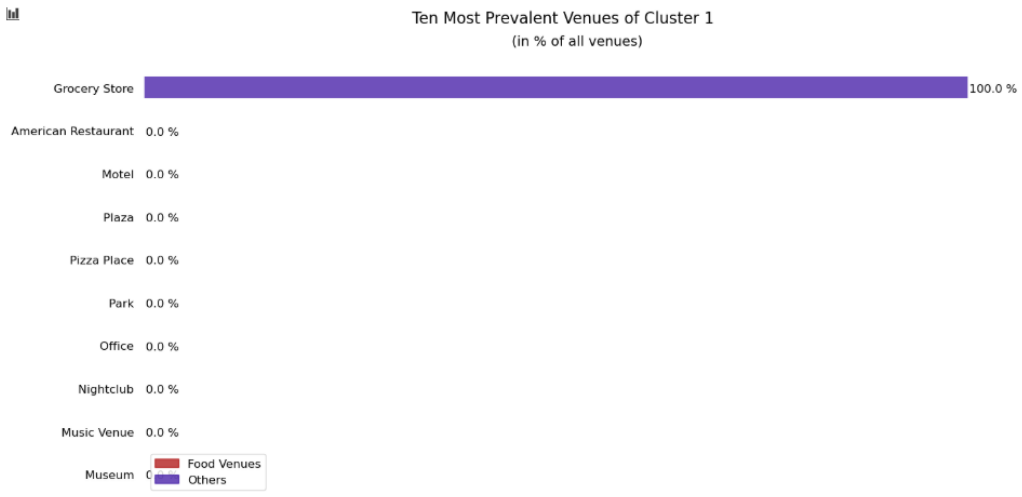
9 rows × 85 columns

The resulting district clusters are also visualized on a map:



A horizontal bar plots are generated that illustrates the top 10 venues in each cluster:





From the results above it is obvious of investigating Cluster 0 and eliminating Cluster 1.

The closest restaurants from each district in Cluster 0 are displayed below, along with the corresponding distances (in meters):

---------- Arabkir ----------

	name	categories	distance	lat	lng
0	Dolma	Restaurant	1222	40.219870	44.491397
1	Latar	Restaurant	1075	40.206126	44.471290
2	Odalen Restaurant	Restaurant	1181	40.215220	44.493776
3	Hrashq Aygi	Restaurant	1394	40.203384	44.469656
4	In Fusion	Restaurant	1443	40.221185	44.493354
5	Genacvale Pandok	Restaurant	1895	40.206289	44.500680
6	Harich	Restaurant	1318	40.215662	44.495301
7	Gandzaqar	Restaurant	1425	40.213074	44.496956
8	At the colonel	Restaurant	1523	40.216705	44.497438

---------- Avan ----------

	name	categories	distance	lat	lng
0	Мельница	Restaurant	942	40.217635	44.564908
1	Ginetun	Restaurant	1285	40.234763	44.560900
2	Restaurant Hatisi Lanjin	Restaurant	1671	40.222936	44.576689
3	Etna	Restaurant	1444	40.229839	44.542132
4	Aygi Hacatun	Restaurant	1787	40.239184	44.562109
5	"Ujqh" hwgwunlu	Restaurant	845	40.224790	44.566860

---------- Davtashen ----------

	name	categories	distance	lat	lng
0	The Vahakni Restaurant	Restaurant	915	40.219349	44.457029
1	Ureni	Restaurant	1060	40.225565	44.452688

From the data above it's clear that districts Avan and Davtashen are closer to the city centre and there is lack of restaurants in these districts.

Conclusion

The objective of this project was to identify the best potential districts in Yerevan where a restaurant can be set up. All the required district data was either scraped of the internet or obtained using Google Map data via csv file. After the districts were visualized on a folium map, their venues were explored using Foursquare location data. Based on the frequency of occurrences of different venue types, the districts were divided into two groups with the help of k-means clustering. The clusters were examined and the best one in which a restaurant could be set up was chosen. The districts were filtered further based on proximity to existing restaurants and distance from the centre of the city. The analysis brought the number of contenders down to two neighbourhoods - **Avan** and **Davtashen**.

This will only serve as a starting point in the overall investigation since there are a lot of other factors that influence such a decision.