1. Introduction

1.1. Description & Discussion of the Background

Baku is the capital and largest city of Azerbaijan, as well as the largest city on the Caspian Sea and of the Caucasus region. There are more than 2 million people live in the Capital of Azerbaijan and it is officially about 25 percent of all inhabitants of the country.

Baku is the sole metropolis in Azerbaijan and we use it for our project. The city is divided into twelve administrative raions and 48 townships. Baku is the scientific, cultural, and industrial centre of Azerbaijan. Many sizeable Azerbaijani institutions have their headquarters there. The Baku International Sea Trade Port is capable of handling two million tons of general and dry bulk cargoes per year.

Our project data analysis will be helpful for those who plans to move to Baku for business or living and to invest into properties and venues in the capital of Azerbaijan. Taking into consideration dynamic urban life, the main criteria for selection will be developed public transport infrastructure, variety and density of venues and social places in the neighbourhoods which are under consideration.

As a city resident who plans to buy apartment for instance, the objective might be to choose the regions where real estate values are lower taking into account main criteria mentioned here above. Similarly, investors expect the districts where there is a lower real estate cost and the type of business, they want to establish is less intense.

Despite it is difficult to obtain information that will guide investors and property buyers in this direction, proper data selection and representation will certainly be useful for solving the problem.

We will create a map and information chart where the real estate prices per sqm distributed according to the neighbourhoods and subway stations and each district is clustered according to the venue density.

1.2. Data Description

I have used following data to work for our analysis:

- I found administrative division and geographic coordinates for divisions and subdivisions from Wikipedia and using Google Map, 'Search Nearby' option to get the centre coordinates of the each districts (administrative division) [1]
- I used Foursquare API to get the most common venues of given districts of Baku [2]
- I collected data about apartments' prices from real estate web-site [3]
- I used information about transportation hub – list of Baku metro stations and their coordinates from Wikipedia [4]

2. Methodology

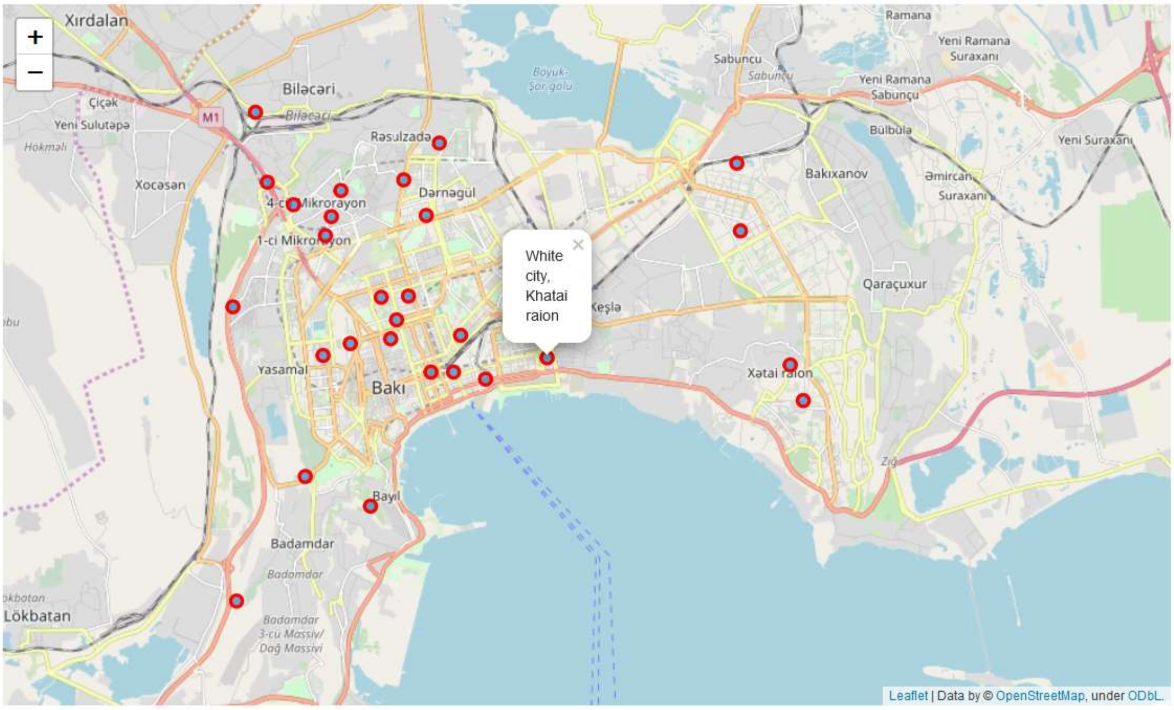
I collected data about divisions called raions and subdivisions(neighbourhoods) from Wikipedia and used geographical location which was provided there, and missing ones <u>were found</u> using google map. Then created table in Excel which contains the list of division, subdivision and their latitude and longitude. (Dataframe1)

Dataframe 1

	Raions	Subdivisions	Latitude	Longitude
0	Nasimi raion	2nd microdistrict	40.413341	49.816159
1	Nasimi raion	3th microdistrict	40.409066	49.814411
2	Nasimi raion	4th microdistrict	40.415929	49.805075
3	Binagadi raion	6th microdistrict	40.419114	49.818975
4	Binagadi raion	7th microdistrict	40.429367	49.847434
5	Binagadi raion	8th microdistrict	40.421250	49.836913
6	Binagadi raion	9th microdistrict	40.420800	49.797520
7	Nasimi raion	Zorge park	40.390614	49.835017
8	Yasamal raion	Izmir park	40.385238	49.821588
9	Khatai raion	White city	40.382046	49.878552

Using geopy library I found geographical coordinate of Baku, generated map using folium library and superimposed label of neighbourhoods on it. (Map 1)

Map 1



Apartments prices

When somebody plan to invest in apartment, location should be always considered in addition to apartment specification.

Is it near to transport hub? Can we commit conveniently? What is the regular price per sqm in selected neighbourhood?

Let's start our analysis.

I loaded information from real estate site https://bina.az (using requests and BeautifulSoup, ran relevant codes) and prepared excel table with data about apartments which includes price per square meter, location, nearest metro station in selected raions and districts. Let's load this table in Pandas dataframe and make some exploratory data analysis.(Dataframe 2)

Dataframe 2

	price	raion	district	nearest_metro	category
0	1600.0	Narimanov raion	NaN	Narimanov metro	new
1	1580.0	Binagadi raion	9th microdistrict	Nasimi metro	old
2	1580.0	Binagadi raion	9th microdistrict	Nasimi metro	old
3	1610.0	Nasimi raion	Nasimi bazar	NaN	old
4	1610.0	Nasimi raion	Nasimi bazar	NaN	old
5	1350.0	Nizami raion	8th kilometer	Neftchilar metro	new
6	1070.0	Yasamal raion	Yeni Yasamal	NaN	old
7	1070.0	Yasamal raion	Yeni Yasamal	NaN	old
8	1690.0	Nasimi raion	NaN	28 may metro	old
9	1690.0	Nasimi raion	NaN	28 may metro	old

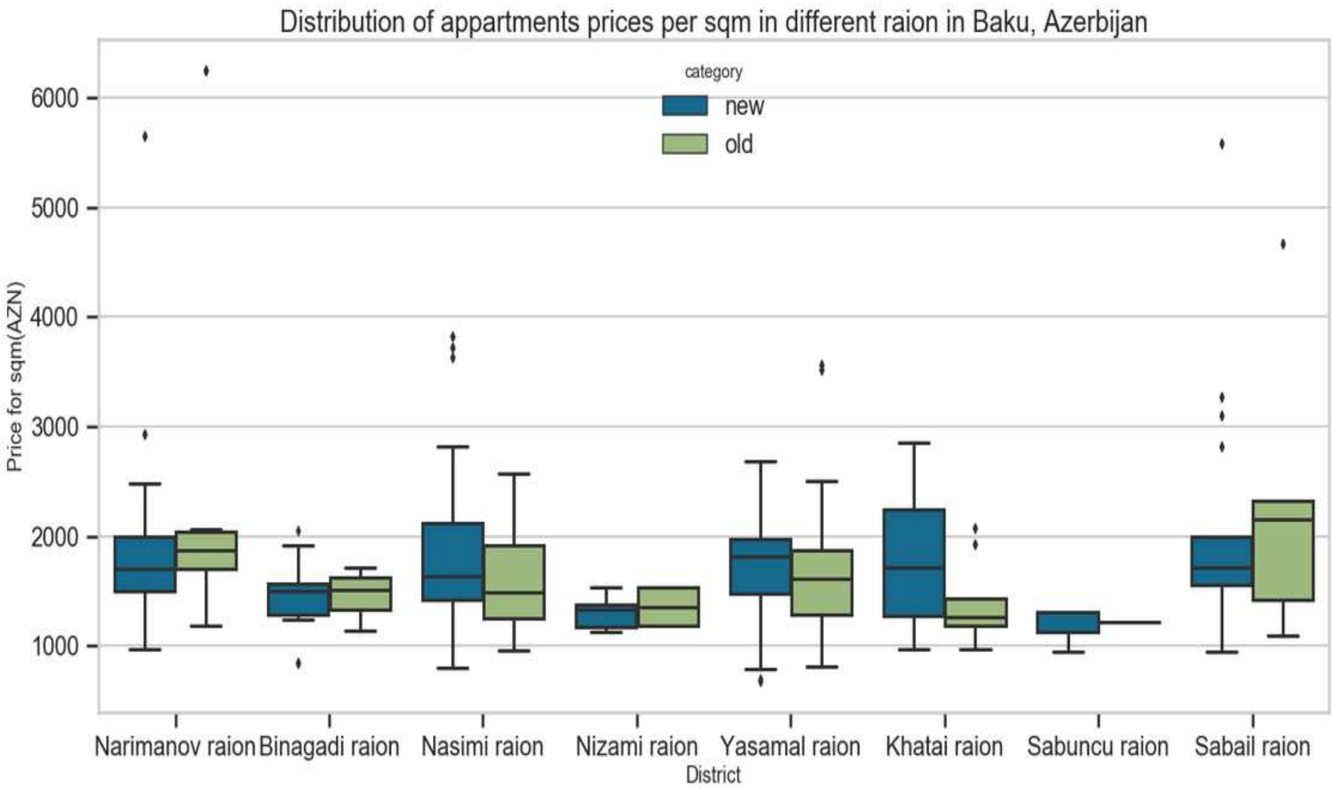
Count mean price per square meter in different raions (districts). (Dataframe 3)

Dataframe 3

	price
raion	
Binagadi raion	1466.628571
Khatai raion	1713.715686
Narimanov raion	2011.757895
Nasimi raion	1735.909605
Nizami raion	1272.045333
Sabail raion	2089.423077
Sabuncu raion	1207.285714
Yasamal raion	1749.547945

Let's have a look at distribution of prices for each raion, and main statistics as first quartile, median, third quartile and outliers.
The boxplot1 below represents prices per sqm of the old and new buildings, where new stands for the buildings constructed during last 20 years, and the rest categorized as old.

Boxplot 1

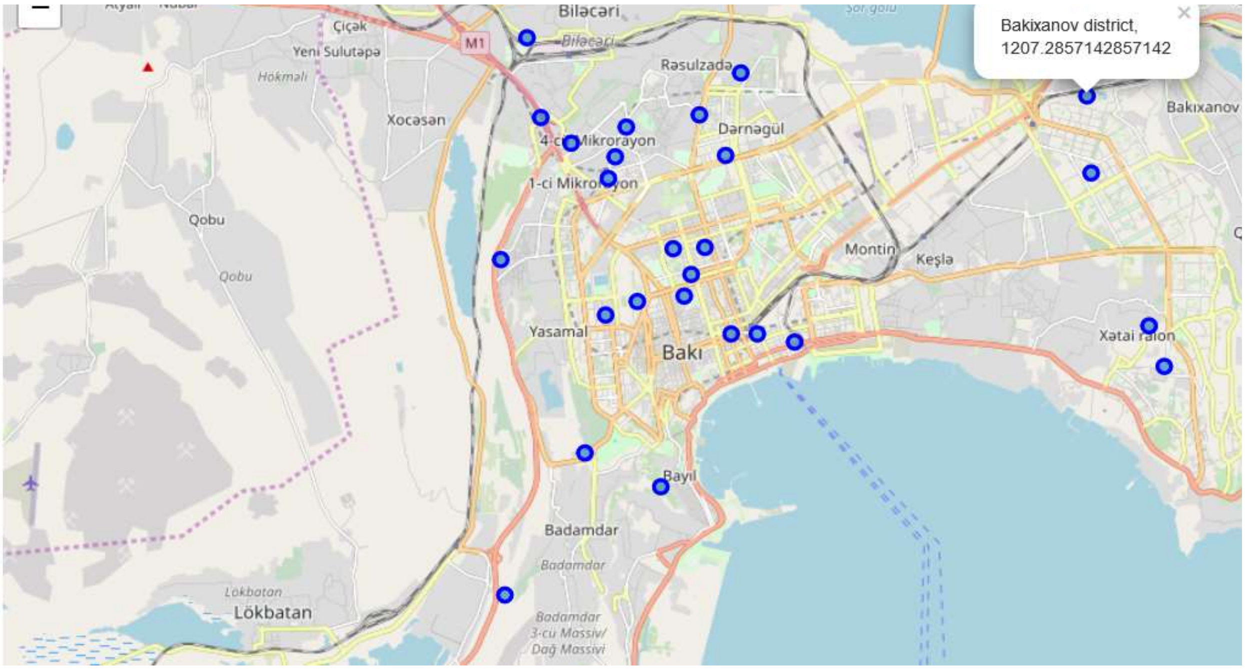


Merging data frames 1 and 2 into dataframe 4 here below will be required to superimpose labels into the map containing information about average prices in each neighbourhood according to coordinates. (Map2)

Dataframe 4

	Raions	Subdivisions	Latitude	Longitude	price
0	Nasimi raion	2nd microdistrict	40.413341	49.816159	1060.000000
1	Nasimi raion	3th microdistrict	40.409066	49.814411	1353.636364
2	Nasimi raion	4th microdistrict	40.415929	49.805075	1461.333333
3	Binagadi raion	6th microdistrict	40.419114	49.818975	1344.285714
4	Binagadi raion	7th microdistrict	40.429367	49.847434	1499.090909
5	Binagadi raion	8th microdistrict	40.421250	49.836913	1640.454545
6	Binagadi raion	9th microdistrict	40.420800	49.797520	1401.176471
7	Nasimi raion	Zorge park	40.390614	49.835017	1862.500000
8	Yasamal raion	Izmir park	40.385238	49.821588	2418.000000
9	Yasamal raion	Beshmartaba	40.386244	49.833450	2420.000000

Map 2

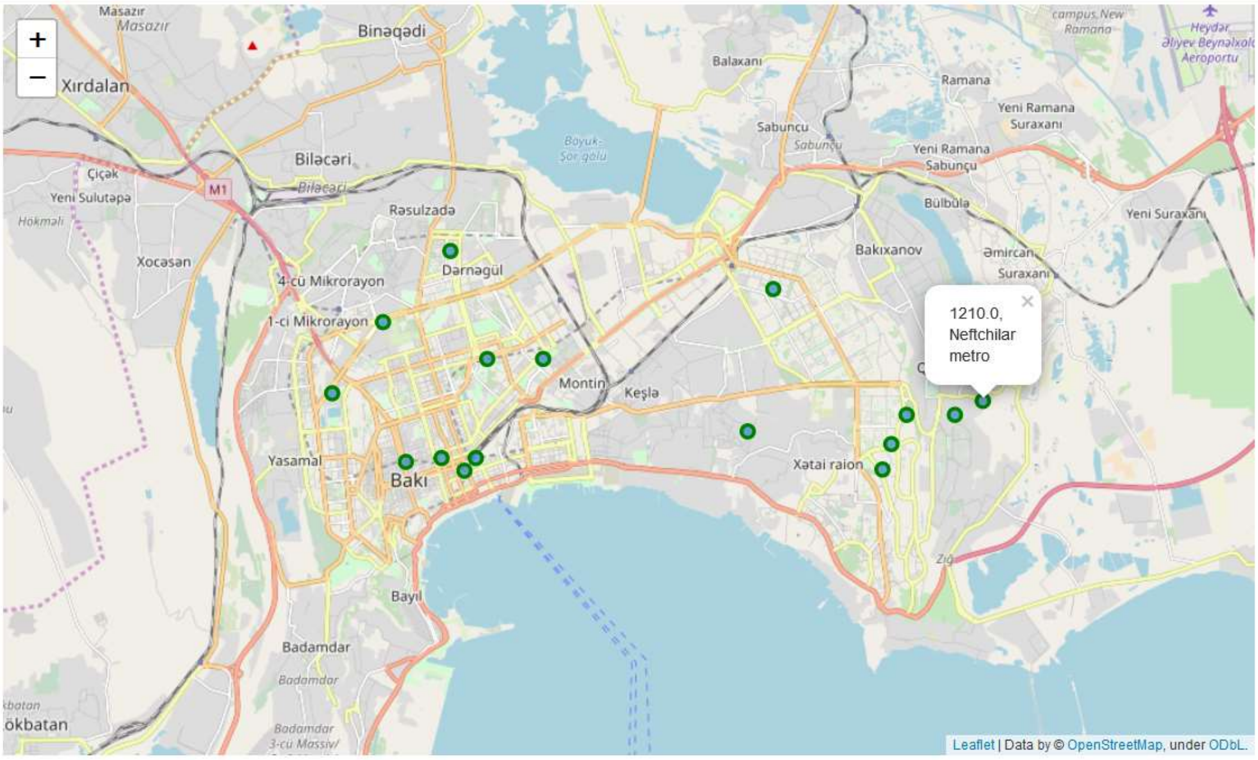


One of the main criteria for apartment's location, availability of nearby transportation network. The transportation networks i.e. bus stations and taxi stands in Baku city are mainly located near metro stations, therefore we select apartments near metro stations and do exploratory analysis on average price per sqm near each metro station. This information represented in data frame 5 and map3.

Dataframe 5

nearest_metro	price
20 yanvar metro	1658.800000
28 may metro	2029.761905
Akhmedli metro	1309.230769
Azadlig metro	1466.666667
Elmlar Akademiyasi metro	2167.272727
Ganjlik metro	2568.400000
Gara Garayev metro	1286.250000
Hazi Aslanov metro	1249.666667
Inshaatchilar metro	1456.696970
Khalglar Dostlughu metro	1406.666667
Khatai metro	1970.000000
Memar Ajami metro	1354.800000
Narimanov metro	1887.592593
Nasimi metro	1360.000000
Neftchilar metro	1210.000000
Nizami metro	1928.421053
Old City metro	2170.833333
Sahil metro	2290.000000
Ulduz metro	1039.000000

Map 3



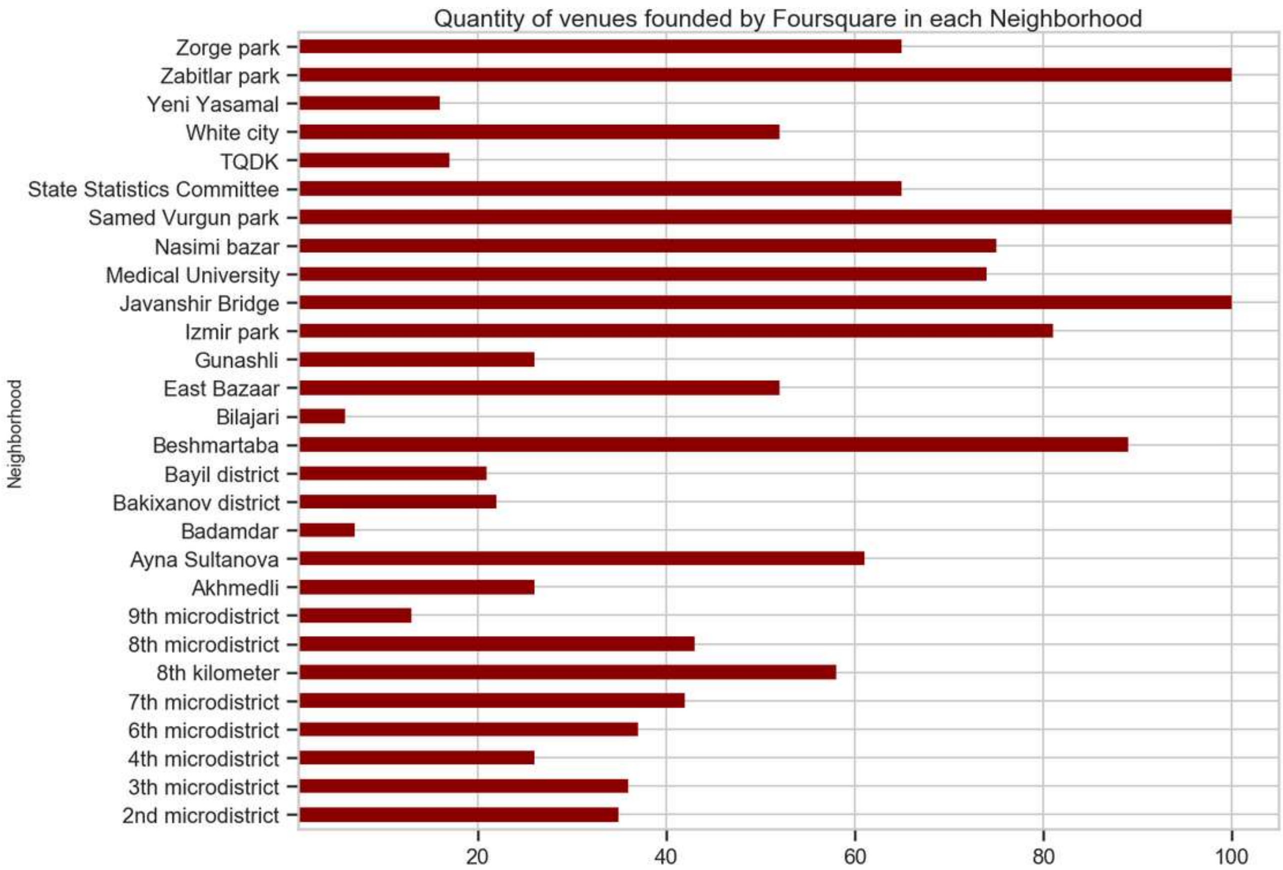
I utilized the Foursquare API to explore the neighbourhoods and segment them. I designed the limit of **100 venues** and the radius of **900 meters** for each neighbourhood from their given latitude and longitude information. Dataframe 6 represents head of the list retrieved from Foursquare API with Venues, category, coordinates at each neighbourhood.

Dataframe 6

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	2nd microdistrict	40.413341	49.816159	Papa John's	40.409949	49.814276	Italian Restaurant
1	2nd microdistrict	40.413341	49.816159	Edem Fitness & Spa	40.407561	49.811484	Gym Pool
2	2nd microdistrict	40.413341	49.816159	AS Lounge	40.411878	49.821344	Karaoke Bar
3	2nd microdistrict	40.413341	49.816159	McDonald's	40.410264	49.813984	Fast Food Restaurant
4	2nd microdistrict	40.413341	49.816159	Ideal Parfumery & Cosmetics	40.411005	49.817553	Cosmetics Shop
5	2nd microdistrict	40.413341	49.816159	Grace Beauty Complex	40.410646	49.815447	Gym / Fitness Center
6	2nd microdistrict	40.413341	49.816159	Favorit Market	40.418813	49.810967	Supermarket
7	2nd microdistrict	40.413341	49.816159	Saraclı Gürcü Mətbəxi	40.419235	49.818224	Restaurant
8	2nd microdistrict	40.413341	49.816159	New Life	40.419921	49.817416	Café
9	2nd microdistrict	40.413341	49.816159	Spar Supermarket	40.406810	49.812777	Market
10	2nd microdistrict	40.413341	49.816159	pioner arena	40.413596	49.811071	Athletics & Sports

Subsequently bar chart 1 below represents quantity of the venues located at each neighbourhood

Barchart 1



The result might not represent all possible venues in the neighbourhoods it will depends on given Latitude and Longitude, in our case we run single point coordinate for each neighbourhood with the certain radius of 900 meters.

For further processing data collected from Foursquare I created dataframe, that contains information about 10 common venues in each neighbourhood. Head of it represented in dataframe 7.

Dataframe 7

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd 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	2nd microdistrict	Restaurant	Fast Food Restaurant	Lounge	Hookah Bar	Café	Italian Restaurant	Steakhouse	Eastern European Restaurant	Soccer Stadium	Market
1	3th microdistrict	Restaurant	Fast Food Restaurant	Athletics & Sports	Tea Room	Convenience Store	Sports Club	Breakfast Spot	Lounge	Eastern European Restaurant	Market
2	4th microdistrict	Bus Station	Restaurant	Supermarket	Steakhouse	Hookah Bar	Shopping Mall	Eastern European Restaurant	Grocery Store	Asian Restaurant	Athletics & Sports
3	6th microdistrict	Restaurant	Café	Hookah Bar	Park	Diner	Middle Eastern Restaurant	Lounge	Karaoke Bar	Spa	Department Store
4	7th microdistrict	Restaurant	Café	Park	Metro Station	Gym	Market	Pub	Department Store	Residential Building (Apartment / Condo)	Buffet
5	8th kilometer	Restaurant	Hotel	Tea Room	Pub	Café	Metro Station	Park	Optical Shop	Middle Eastern Restaurant	Women's Store
6	8th microdistrict	Restaurant	Park	Café	Eastern European Restaurant	Metro Station	Lounge	Electronics Store	Food & Drink Shop	Diner	Cupcake Shop
7	9th microdistrict	Bus Station	Clothing Store	Asian Restaurant	Restaurant	Shopping Mall	Board Shop	Jewelry Store	Comfort Food Restaurant	Hotel	Grocery Store
8	Akhmedli	Restaurant	Bakery	Park	Comfort Food Restaurant	Big Box Store	Pub	Snack Place	Eastern European Restaurant	Bookstore	Fast Food Restaurant
9	Ayna Sultanova	Restaurant	Hotel	Gym	Middle Eastern Restaurant	Clothing Store	Eastern European Restaurant	Lounge	Caucasian Restaurant	Soccer Stadium	Men's Store
10	Badamdar	Shopping	Women's	Furniture /	Eastern European	Bridal Shop	Men's Store	Cupcake	Department	Cosmetics	Electronics

Using this information, we can apply unsupervised machine learning algorithm K means clustering, which creates clusters with for neighbourhoods where the same venue categories are located.

3. Results

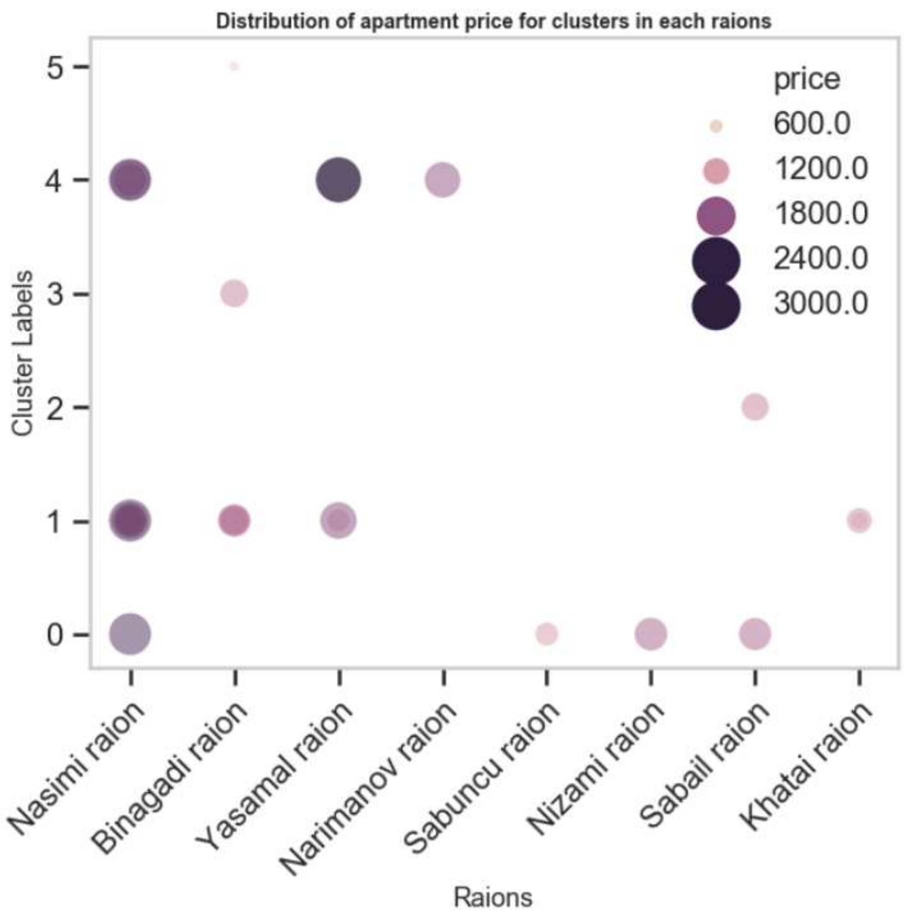
We got cluster label for each neighbourhood, and now we can merge all our data in one table. In which all information about neighbourhoods (subdivisions) is shown. (Dataframe 8)

Dataframe 8

	Raions	Subdivisions	Latitude	Longitude	price	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Mo Comm Venu
0	Nasimi raion	2nd microdistrict	40.413341	49.816159	1060.000000	1	Restaurant	Fast Food Restaurant	Lounge	Hookah Bar	Café	Italian Restaurant	Steakhou
1	Nasimi raion	3th microdistrict	40.409066	49.814411	1353.636364	1	Restaurant	Fast Food Restaurant	Athletics & Sports	Tea Room	Convenience Store	Sports Club	Breakfa Sp
2	Nasimi raion	4th microdistrict	40.415929	49.805075	1461.333333	1	Bus Station	Restaurant	Supermarket	Steakhouse	Hookah Bar	Shopping Mall	Easte Europe Restaura
3	Binagadi raion	6th microdistrict	40.419114	49.818975	1344.285714	1	Restaurant	Café	Hookah Bar	Park	Diner	Middle Eastern Restaurant	Loun
4	Binagadi raion	7th microdistrict	40.429367	49.847434	1499.090909	1	Restaurant	Café	Park	Metro Station	Gym	Market	P

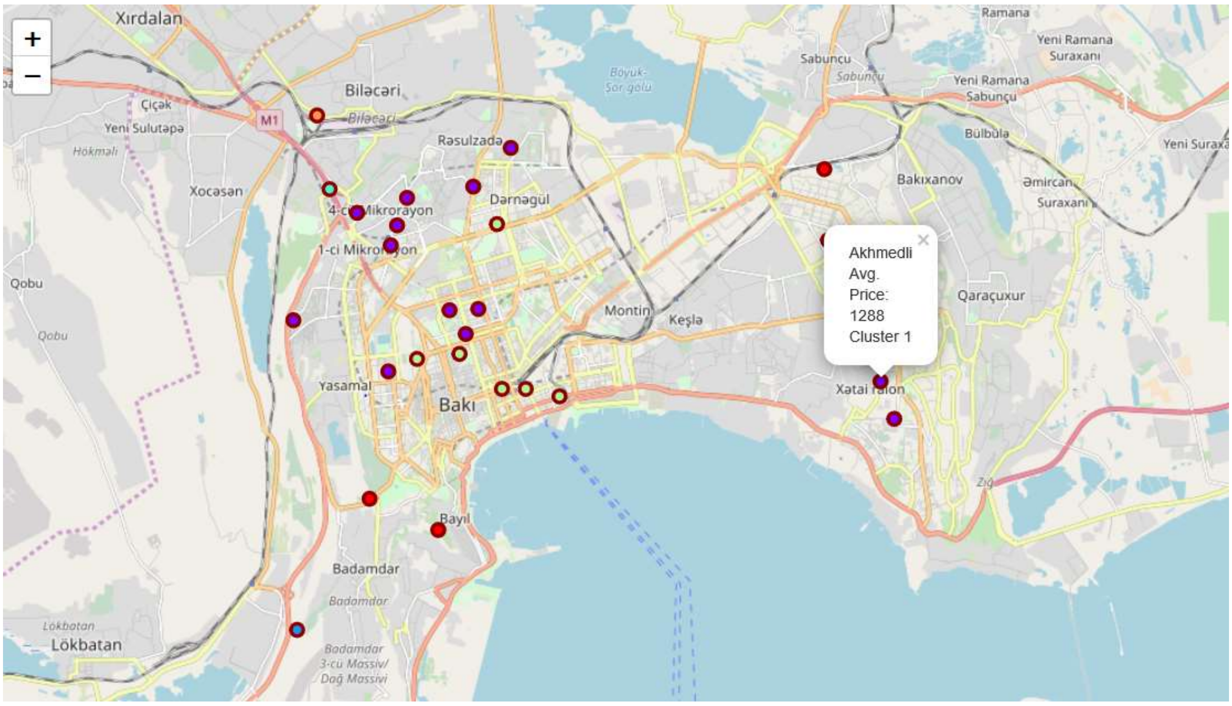
Now we can visualize Dataframe 8 in scatter plot. Where x axis it's raions, y axis clusters, and circles shows average price for each cluster in every raion. (scatterplot 1)

scatterplot 1



Similar information can be represented on Baku city map as well (map 4)

Map 4



4. Discussion

As it was mentioned before, Baku is the sole metropolis in Azerbaijan and city is divided into twelve administrative raions and 48 townships. I have taken into account 8 raions which are covered by Baku city metro network in order to analyse apartment prices at respective raions and nearby respective metro station. Then I used the Kmeans algorithm as part of clustering study to define equivalent neighbourhoods located in various parts of the city in terms of venue category and their density and then assigned prices per sqm for apartments in considered raions with respect to relevant cluster.

It is necessary to highlight that more detailed and accurate guidance can be achieved since data set can be expanded and the details of the neighbourhoods or streets can be also drilled down.

The result of this exercise will be stored in the GitHub for future similar or more detailed studies and open for public access.

My study ended with data visualisation and clustering information on the map of Baku city.

5. Conclusion

As a result, such visual representation of the necessary data superimposed on location map might be very helpful for people to achieve various project objectives effectively.

It is worth to highlight that similar data analysis might be useful for city governors such as municipalities, town planning, department of transport and etc. in their project initiatives related to development of various neighbourhoods of the city and transport hubs.

Best Regards

Rahila Iman-zada

References:

- [1] https://en.wikipedia.org/wiki/Baku
- [2] Baku Administrative divisions
- [3] Foursquare API
- [4] Google Map