1. Introduction/Business Problem

Vilnius is the capital city of Lithuania. It produces ~50% of Lithuania GDP and attracts largest amount of newcomers from smaller towns around the country. Currently Vilnius has ~600 000 inhabitants and together with surroundings creates largest hub for local companies and foreign investments.

The goal of this assignment will be to analyze all 21 Vilnius districts in order to find the most attractive areas to buy apartment for newcomers coming to this city. The main criteria for this assignment will be to analyze apartment prices per sq.m. and also venues in every district. All information would be consolidated clustering neighborhoods into several clusters.

Currently there are no such tool which would clearly consolidate and analyze Vilnius real estate market, which should help to make decisions regarding settlement in various Vilnius city districts. Such tool would also save a lot of time analyzing various data in 3-4 different local portals in order to find optimal decision before starting to search target apartments in Vilnius city.

The **target audience** for such tool could be divided into 3 groups:

1. **Vilnius city residents** who are searching for new opportunities for reallocation in better parts of the city.
2. **Lithuanian residents** who searching for a new job opportunity in Vilnius and also for long term settlement in capital city of Lithuania
3. **Foreigners** who are planning long term investments into real estate market in Vilnius or planning to reallocate into Lithuania's capital.

2. Data

For data section we will use data from **3 main sources**:

Vilnius Municipality open API -> data will be used in order to get information about locations of Vilnius neighborhoods. Vilnius Municipality open API consist of various data including all districts geographical coordinates. There are also lots of additional data including streets, traffic data, utilities companies activities, air pollution and etc. For this project will be used a dataset of Vilnius district's boundaries https://data-vplanas.opendata.arcgis.com/datasets/vilniaus-miesto-seni%C5%ABnij%C5%B3-ribos

The main features that can be extracted from this dataset are:

- District name
- District coordinates

Biggest real estate portal in Lithuania www.Aruodas.lt -> using web scrapping we will use Vilnius apartment data which are currently on sale for last 30 days. The main features that can be web scrapped from this dataset using Beautifulsoup are:

- Apartment district
- Apartment street
- Apartment area
- Apartment price
- Apparent room numbers

This data will be used to calculate average prices and average area in each district.

Foursquare data of Vilnius city venues -> using Foursquare API we will use data which will show most attractive districts with largest amount of venues in all Vilnius districts.

Foursquare API data will be used for clustering districts according to venues.

3. Short description of the process

I. Methodology

1. Data from Vilnius Municipality API will be used in order to make municipality maps and **get district coordinates**.
2. Data from aruodas.lt webpages using **web scrapping** will be moved to CSV file.

3. CSV file will be analyzed **plotting charts** and getting more information about situation in Vilnius real estate market.
4. A map of average price per sq. m. in all Vilnius municipalities will be **plotted using folium.**
5. **Foursquare API data** will be used in order to analyze Vilnius venues.
6. **Elbow method** will be used in order to decide optimal amount of clusters.
7. **Clustering will be performed** using data from Foursquare venues and Vilnius Municipality API.

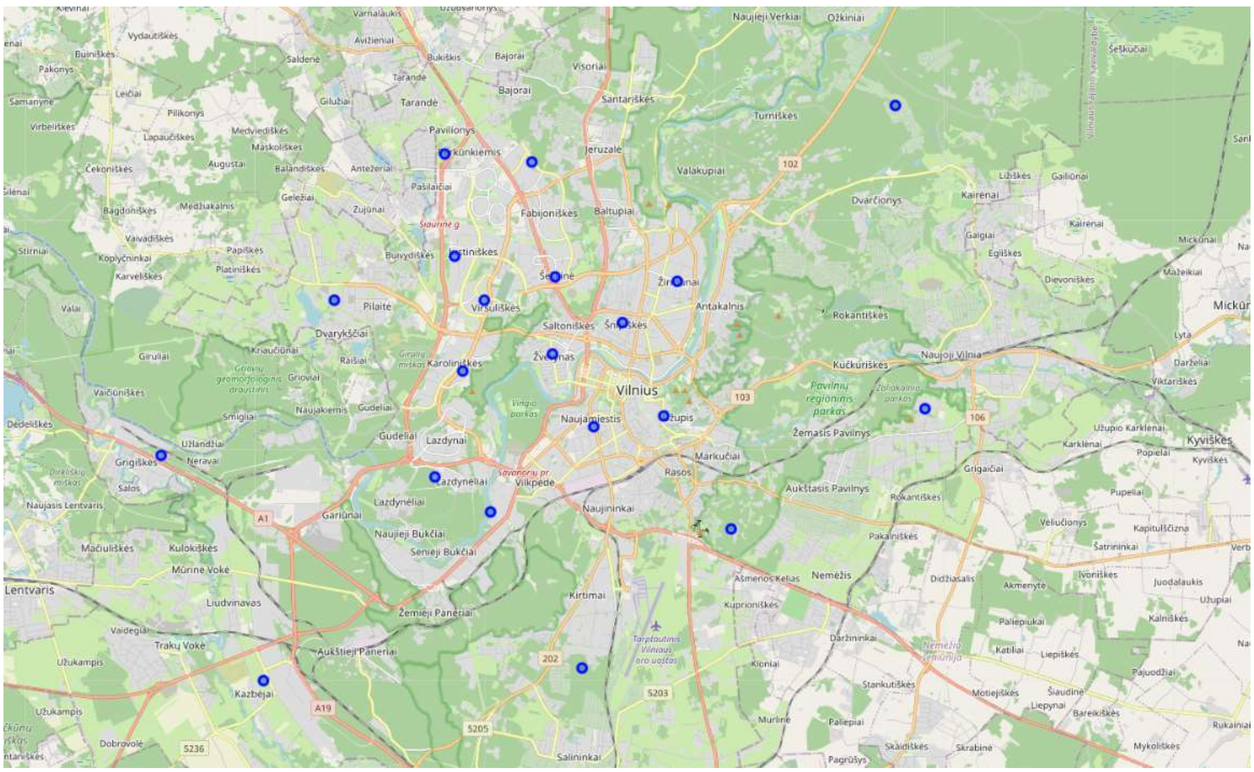
II. Final result as map of clusters will be provided together with detailed information about each cluster.

III. Discussion and conclusion section will provide further insights into the project.

4. Methodology

4.1. Vilnius districts

Vilnius consist of 21 districts with total number of 544 38 inhabitants according to Wikipedia. For data analysis we need to get geographical data of all Vilnius districts/neighborhoods including district name and its geographical boundaries within the city. For this data we use Vilnius Municipality open API, which also contains data of Vilnius districts. Still those data does not have coordinates of centroids of each districts so with the help of geopandas we get centroid coordinates in the map of each Vilnius district. The main outcome for this process is Vilnius map with centroid coordinates of all 20 districts.



4.2. Web scrapping

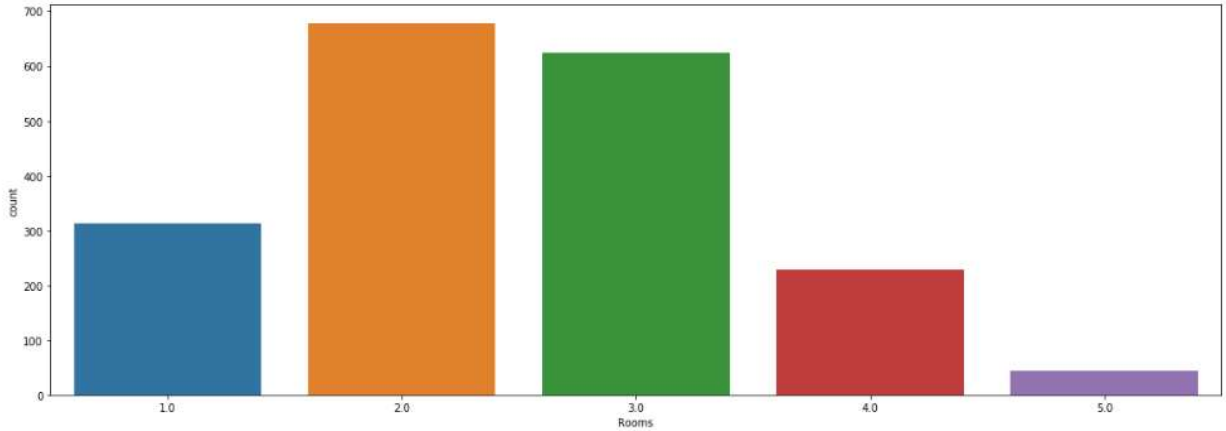
Using Beautifulsoup we web scrapping data from largest real estate portal in Lithuania aruodas.lt. For next analysis we use data of apartments on sale for last 30 calendar days. This data are filtered in aruodas.lt portal. Data are moved to pandas dataframe and lately into csv file. After simple analysis we got for 1896 apartments in Vilnius. With no filtering general distribution of total 1896 apartments according to number of rooms are:

Rooms	# of appartments on sale

2.0	678
3.0	625
1.0	313
4.0	228
5.0	45
6.0	5
7.0	1
20.0	1

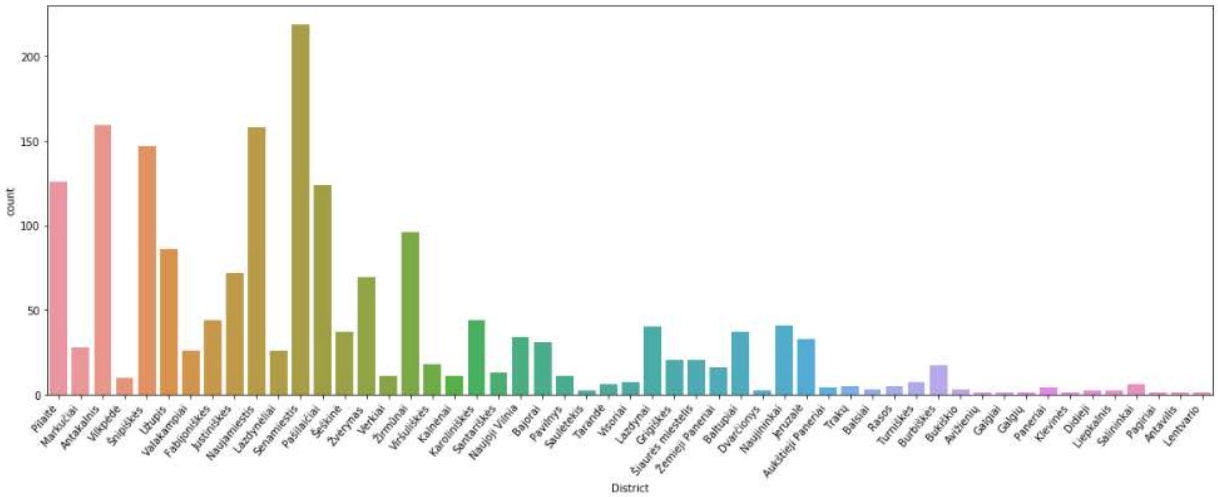
4.3. Plotting charts / apartment data analysis

Before starting analysis and looking on previous table in the report we remove outliers that are apartments with room number equal or larger than 5. Next we could split all dataset according to the number of rooms and plot a chart.



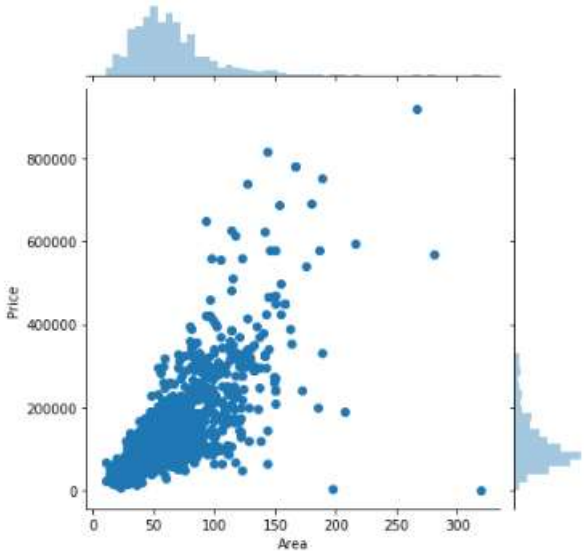
From the chart we se the same information as it was provided in the table, that largest amount of apartments have 2 and 3 rooms. This amount is around 650-700 apartments in each category. Such type of apartments is most popular for young families as well as possible investment for rent market.

Having all dataset we can analyze which Vilnius districts have largest number supply of apartments.

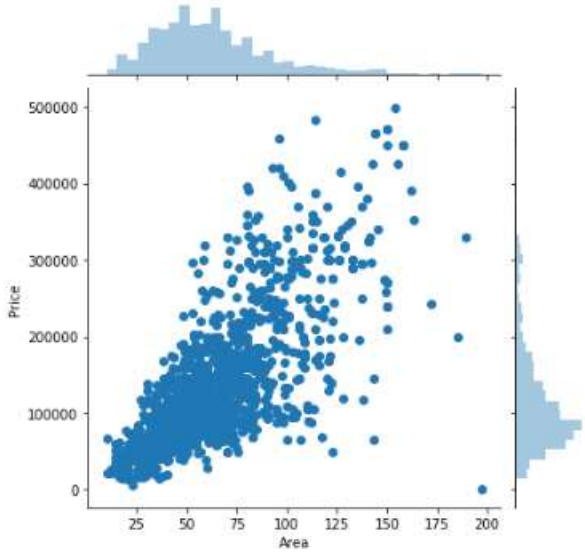


The chart shows that largest supply of apartments are in Senamiestis (Old Town), Naujamiestis (New Town) districts. However those parts of the city are one of the most expensive, so quite large set of apartments are supplied in so cold “sleeping districts” such as Pasilaiciai, Snipiskes.

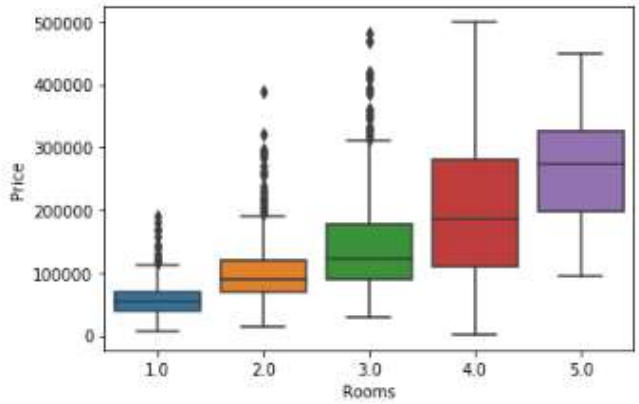
For further analysis we take all apartment data to further identify outliers which will be seen from joint plot.



As we see from the chart plot there are few apartments that probably are outliers with very large area and extremely rage prices. From the dataset we remove apartments with prices larger than 500 000 EUR and areas larger than 200. Such type of properties are more luxurious and do not fit into general analysis.
Removing outliers we get such distribution of dataset.

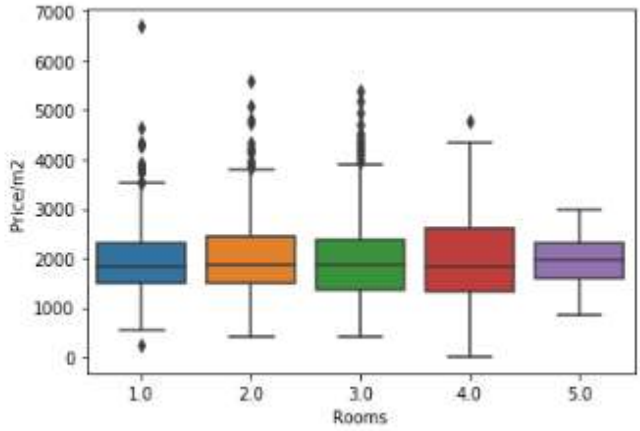


As we go further to analysis, we can plot the table to show price differences between different size of apartments in Vilnius.



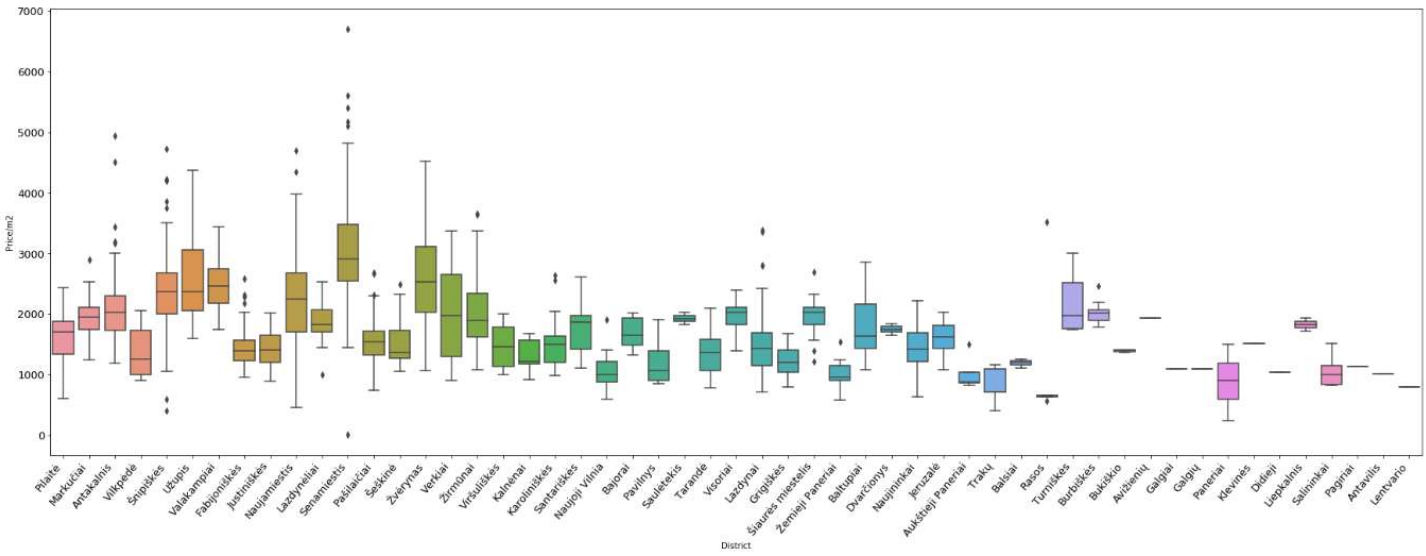
Data analysis shows no big surprises - the larger apartment is the higher is median price. However, the gap between median of 2 room and 3 room apartments is smaller comparing to other types of apartments.

In case absolute prices does not show the real situation, further we use price per square meter as indicator instead of absolute price.



From the chart above we see that median price of apartment in Vilnius is 1850 EUR/m2. There are no big differences between apartment types. The lowest quartile end at around 1440 EUR/m2, while the top quartile starts at 2400 EUR/m2. So the general range of apartment prices in Vilnius fluctuates between 1400-2400 EUR/m2.

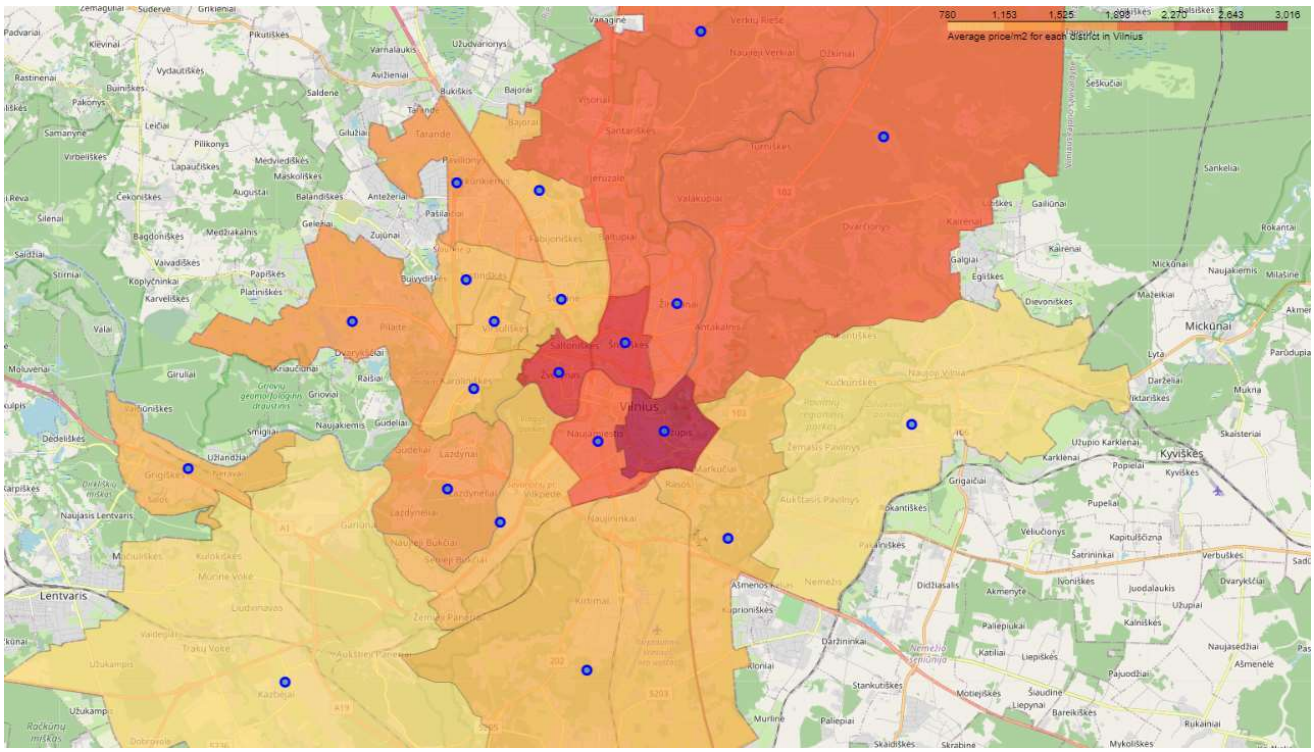
We also can analyze prices in different districts in Vilnius to see which parts of city are more expensive than others.



From this chart we see that Senamiestis (Old Town) and Zverynas are most expensive parts of the city, those are the most centrally located districts. Still there are much cheaper alternatives which lately can be seen on map, but for locals its already seen that close to center districts such as Naujamiestis (New Town), Uzupis, Snipiskes are more economic similarities that are close.

4.4. Map of apartments market in Vilnius

That is one of the final outcomes of analysis. Using data from real estate portal, and geographical coordinates of Vilnius districts we can plot a map and show price per sq. m. of each Vilnius district apartments.



The data shows already previously presented information, that most expensive part of the city is Old Town and nearby located districts such as Zverynas and Snipiskes. But the New Town (Naujamiestis) seems a cheaper alternative as well as Zirmunai. From this map it is also seen that southern part of

the city is less expensive due to airport and industrial activities and northern part, especially north-east part is more attractive due to recreational activities, riverside, parks and forests located in this area.

4.5. Foursquare API data

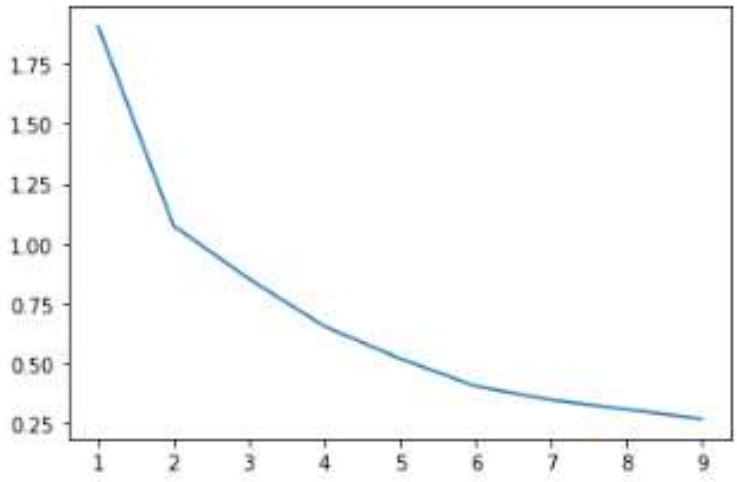
Using Foursquare API we get Vilnius venues on each district. From the table listed below we see that most venues (we take maximum limit of 100 venues in API) are in districts Senamiestis, Naujamiestis, Seskine, Zirmunai, Zverynas and Snipiskes. The least amount of venue results are in Verkiai and Antakalnis districts which are further from the center part of the city.

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Antakalnis	6	6	6	6	6	6
Fabijoniškės	57	57	57	57	57	57
Grigiškės	7	7	7	7	7	7
Justiniškės	46	46	46	46	46	46
Karoliniškės	51	51	51	51	51	51
Lazdynai	24	24	24	24	24	24
Naujamiestis	100	100	100	100	100	100
Naujininkai	6	6	6	6	6	6
Naujoji Vilnia	9	9	9	9	9	9
Paneriai	8	8	8	8	8	8
Pašilaičiai	22	22	22	22	22	22
Pilaitė	17	17	17	17	17	17
Rasos	23	23	23	23	23	23
Senamiestis	100	100	100	100	100	100
Verkiai	6	6	6	6	6	6
Vilkpėdė	19	19	19	19	19	19
Viršuliškės	47	47	47	47	47	47
Šeškinė	100	100	100	100	100	100
Šnipiškės	100	100	100	100	100	100
Žirmūnai	100	100	100	100	100	100
Žvėrynas	100	100	100	100	100	100

4.6. Elbow method

Using elbow method, we would like to know what is the optimal amount of clusters in order to organize data into similar clusters. Elbow method shows when is no more rational to make more clusters. The point where the curve starts to become horizontal is the mark of number of clusters in the model.

Elbow method points out such line:



From the graph we see that there are 2 braking points where the curve starts to become horizontal – one at 2 clusters and one at 6 clusters. For further analysis we choose 6 clusters as optimal amount of analyzing Vilnius districts.

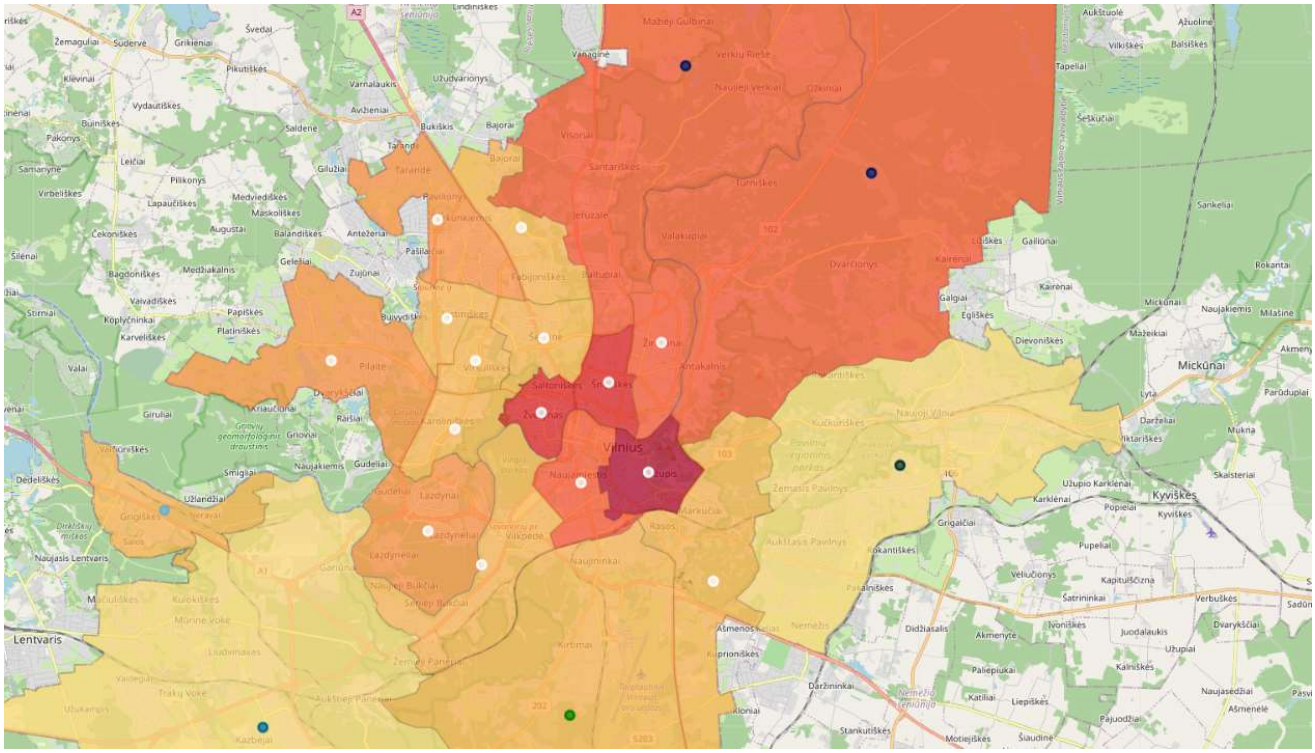
4.7. Clustering

Using clustering we divide all 21 Vilnius districts into 6 clusters. Analysis gives such outcome:

Cluster	Number of districts
0	15
3	2
1	1
2	1
4	1
5	1

We see that there is one dominating cluster of 15 districts, while the rest 6 districts are spread across 5 clusters. So only taking this data we see that there will be a lot of alternatives in cluster Number 0, and also we would analyze what are the differences in other 5 clusters.

5. Results



As the final result of this analysis we provide map of Vilnius city districts and 6 clusters which are spread across the town.

Also detailed information about each cluster is provided in the tables below.

Cluster 0.

	District(eng)	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Rasos	0	Hotel	Ski Area	Park	Brewery	Shopping Mall	Shoe Store	Eastern European Restaurant	Dog Run	Supermarket	Diner
1	Naujamiestis	0	Coffee Shop	Gym	Restaurant	Hotel	Bar	Café	Dessert Shop	Plaza	Pizza Place	Museum
2	Senamiestis	0	Coffee Shop	Bar	Park	Restaurant	Dessert Shop	Beer Bar	Museum	Café	Bakery	Pub
3	Vilkpede	0	Gym	Restaurant	Sporting Goods Shop	Convenience Store	Pizza Place	Bookstore	Shipping Store	Fast Food Restaurant	Hotel	Cosmetics Shop
4	Lazdynai	0	Park	Café	Fast Food Restaurant	Grocery Store	Bus Station	Restaurant	Chinese Restaurant	Campground	Bus Stop	Food Court
10	Pasilaiciai	0	Supermarket	Grocery Store	Pizza Place	Bar	Burger Joint	Burrito Place	Kebab Restaurant	Cafeteria	Beer Store	Bed & Breakfast
11	Fabijoniskes	0	Grocery Store	Gym / Fitness Center	Fast Food Restaurant	Pizza Place	Food & Drink Shop	Bus Station	Supermarket	Electronics Store	Bed & Breakfast	Market
12	Zirmunai	0	Coffee Shop	Pizza Place	Clothing Store	Park	Gym	Gym / Fitness Center	Bakery	Supermarket	Burger Joint	Shoe Store
13	Pilaite	0	Supermarket	Forest	Hotel	Restaurant	Fast Food Restaurant	Coffee Shop	Lake	Kebab Restaurant	Steakhouse	Beach
14	Justiniskes	0	Supermarket	Gym / Fitness Center	Grocery Store	Pizza Place	Flea Market	Coffee Shop	Gym	Bed & Breakfast	Fast Food Restaurant	Soccer Stadium
15	Seskine	0	Clothing Store	Coffee Shop	Gym / Fitness Center	Pizza Place	Supermarket	Cosmetics Shop	Shopping Mall	Fast Food Restaurant	Italian Restaurant	Electronics Store
17	Virsuliskes	0	Grocery Store	Pizza Place	Gym / Fitness Center	Coffee Shop	Supermarket	Bed & Breakfast	Bowling Alley	Fast Food Restaurant	Flea Market	Lake
18	Snipiskes	0	Clothing Store	Coffee Shop	Pizza Place	Italian Restaurant	Bar	Plaza	Hotel	Park	Brewery	Café
19	Zverynas	0	Coffee Shop	Gym / Fitness Center	Café	Park	Hotel	Clothing Store	Restaurant	Dance Studio	Modern European Restaurant	Italian Restaurant
20	Karoliniskes	0	Park	Pizza Place	Restaurant	Grocery Store	Gym	Office	Diner	Café	Soccer Field	Shop & Service

Cluster No. 0 can be described as **touristic, commercial and residential** cluster. This cluster gives the greatest comfort of living in apartment due to great amount of venues and close proximity to the center. However price per sq.m. is different and vary among districts that are in this cluster.

Cluster 3.

	District(eng)	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Verkiai	3	Lake	Grocery Store	Campground	Garden	Outdoors & Recreation	Yoga Studio	Donut Shop	Flea Market	Fishing Store	Fast Food Restaurant
9	Antakalnis	3	Food Truck	Campground	Lake	Garden	Park	Food & Drink Shop	Department Store	Electronics Store	Flower Shop	Flea Market

Cluster No. 3 can be described **as recreational**. Those districts are quite far from the center part of the city and also more suitable for houses.

Cluster 2.

	District(eng)	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
16	Naujoji Vilnia	2	Park	Mountain	Market	Convenience Store	Supermarket	Soccer Field	Yoga Studio	Eastern European Restaurant	Fishing Store	Fast Food Restaurant

Cluster No. 2 has only one district. This area has the lowest price per sq. m. and also has **a potential** for the future to become residential district.

Cluster 1.

	District(eng)	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Naujininkai	1	Airport Service	Airport Terminal	Hardware Store	Fast Food Restaurant	Soccer Field	Donut Shop	Flower Shop	Flea Market	Fishing Store	Farmers Market

This is **airport** zone. This area is mostly industrial and due to airport traffic is unattractive for apartments.

Cluster 4.

	District(eng)	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	Paneriai	4	Business Service	Diner	Market	Fast Food Restaurant	Furniture / Home Store	Eastern European Restaurant	Train Station	Food & Drink Shop	Dessert Shop	Department Store

Cluster No. 4 can be described **as industrial** part of Vilnius municipality. This is highly covered by industrial buildings, factories and also commercial property.

Cluster 5.

	District(eng)	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Grigiskes	5	Bus Stop	Fast Food Restaurant	Pharmacy	Grocery Store	River	Soccer Field	Shopping Mall	Yoga Studio	Farmers Market	Electronics Store

Cluster No. 5 can be treated as **Vilnius suburb**. This part of Vilnius has the shortest history being as Vilnius city part and long years was treated as Vilnius nearby town.

6. Discussion and conclusions

Several conclusions can be done from clustering analysis.

1. Cluster No.0 is the most attractive cluster out of all 6 clusters. Top 3 districts that are most expensive are in this cluster (Senamiestis, Zverynas and Snipiskes). However this cluster has more alternatives that are close to the city center and are cheaper alternatives. The best comparing to mentioned 3 districts is Naujamiestis which has 15 perc. cheaper price and the same proximity as the most expensive alternatives.
2. In south-west part of the city we can find even more cheaper alternatives like Seskine, Virsuliskes, Karoliniskes that are also lie in the same cluster. Those parts of the city prices are 60 perc. lower than Vilnius central part.

Taking everything into account that alternatives mentioned in conclusion No. 2 are the best choices looking from price and convenience perspective considering new apartment acquisition in Vilnius.

Further discussion

1. Analysis was done during the Covid-19 crisis and there are possibilities for price correction in real estate market due to consequences in economy.
2. Analysis did not separated newly constructed and old apartments, this would definitely give more insights into real estate market in Vilnius.
3. Also this analysis have been done only taking apartment sector and not analyzing house market which sometimes can be as attractive alternative.
4. For further analysis it would be logical to take historical price development in Vilnius, also there can be done alternative analysis taking other amount of clusters.