Housing Prices & Venues - Data Analysis of Bucharest

A. Introduction

A.1. Description & Business Problem

Bucharest is the capital and largest city of Romania, as well as its cultural, industrial, and financial center. It is located in the southeast of the country, at 44°25'57"N 26°06'14"ECoordinates: 44°25'57"N 26°06'14"E, on the banks of the Dâmbovița River, less than 60 km (37.3 mi) north of the Danube River and the Bulgarian border. As the largest city in Romania, Bucharest has more than 2 millions of people as inhabitants, and a surface of 228 km2. Is also a city with a very high population density: 8,260/km2, owing to the fact that most of the population lives in high-density communist era apartment blocks.

In 2019 Bucharest was rated by TomTom application as the number 4 city in Europe and as number 14 in the world, based on traffic congestion level (https://www.tomtom.com/en_gb/traffic-index/ranking/) and this situation is strongly connected with the population density, city infrastructure, public transportation and many other factors locally known.

Bucharest is divided in 6 Districts with independent administration and each sector in many neighborhoods. There are a total of 55 neighborhoods and areas of interest in Bucharest which the residents are well-aware of, and the decision of purchasing or renting a house or flat is made depending on these neighborhoods not on the big Districts. At the same time, Bucharest has over 95% of house owners, people preferring to purchase a home and not to rent.

Having all of these in mind, if you need to cross the city every day in order to get to your place of work, it is possible to think one day about actually moving. Beside selling the house you own and buying another in the same price range, the venues around your new location are also very important in order to maintain the same or better life style.

In this analysis I will gather data about the housing prices and about the venues (cultural, educational, and entertaining) in Bucharest and I will cluster all the neighborhoods and areas of interest into categories and show them on the Bucharest map in order to help anyone that needs to switch living accommodation to make a decision.

A.2. Data Description
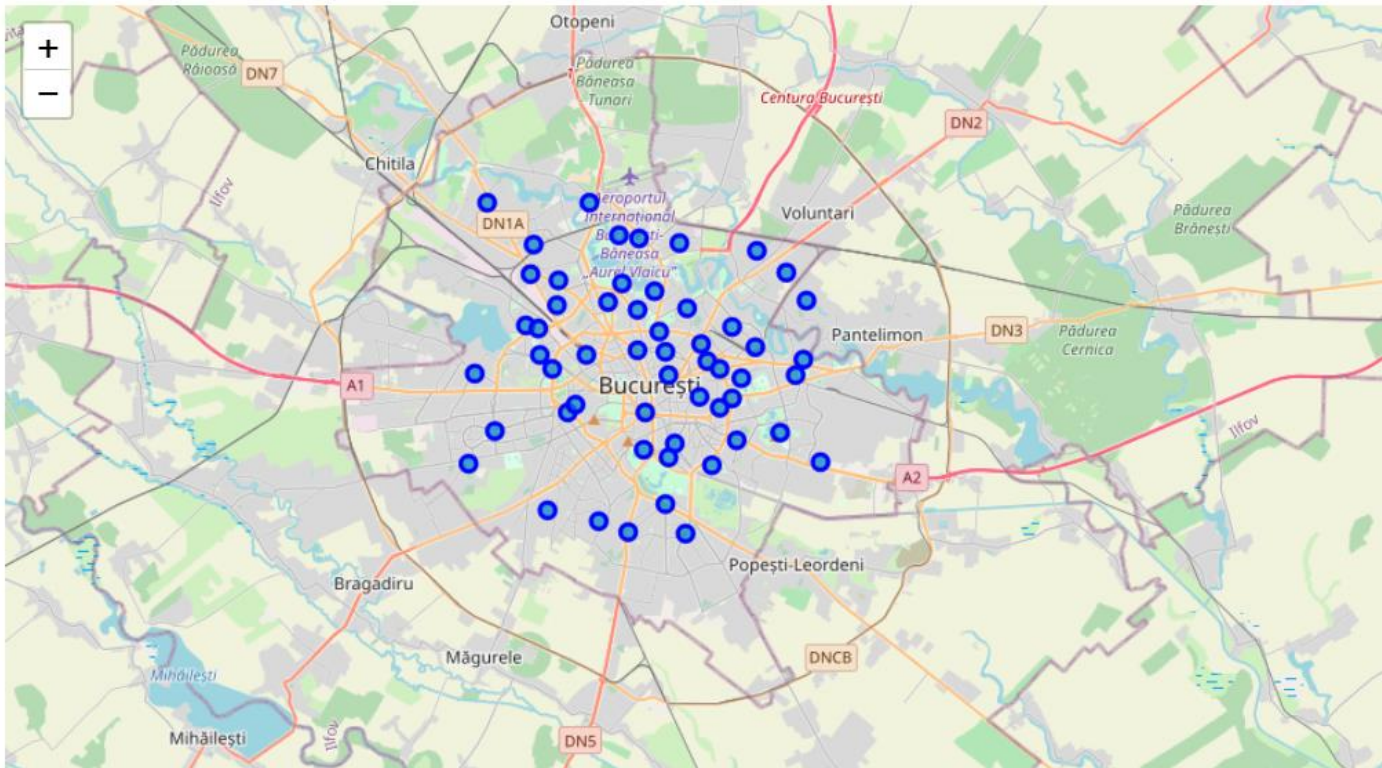
To consider the problem we can list the data as below:

• There are not too many public data related to demographic and social parameters for the city of Bucharest. The housing prices are collected from the "Raport_Q4_2019eng_housing price" on the Romanian portal https://www.analizeimobiliare.ro/, which is part of the largest real estate community in Romania.

• The data of the Bucharest neighborhoods and of the demographic characteristics was collected via Wikipedia web site

• I used Foursquare API to get the most common venues of given Neighborhoods of Bucharest

• Google Map, 'Search Nearby' option was used to get the center coordinates of each neighborhood.

B. Methodology

As a database, I used GitHub repository in my study. My master data has the main components Neighborhood, Average House Price, Latitude and Longitude information of the city of Bucharest

	Neighborhood	Average House Price	Latitude	Longitude
0	1 Mai	1690	44.471100	26.050000
1	13 Septembrie-Panduri	1250	44.427539	26.066529
2	Aviatiei	2060	44.482300	26.098100
3	Andronache	1300	44.478500	26.150100
4	Baicului	990	44.448100	26.149300

I used Python Folium library to visualize geographic details of Bucharest and its Neighborhoods. I created a map of Bucharest with neighborhoods centers superimposed on top. I used latitude and longitude values to get the visual as below:

- With the Foursquare API I explored the neighborhoods in order to get the interests point's venues for each one. The limit used was of 100 venues and the radius 1000 meter for each neighborhood from their given latitude and longitude information's.
- I tested the connection with the Foursquare API for the first neighborhood from the list and the connection returned 13 venues listed in the data frame below:

	name	categories	lat	lng
0	Lebanese Station	Lebanese Restaurant	44.471121	26.050584
1	The Coco Pub	Pub	44.469502	26.057233
2	Mega Image	Supermarket	44.476955	26.044773
3	Mega Image	Supermarket	44.471190	26.051079
4	Ubisoft Gym	Gym / Fitness Center	44.473731	26.057945

- I requested all the venues from the API for all the neighborhoods with the limitation of 1000 meters from the center and with the limit of 100 venues for each neighborhood. The Foursquare API returned 3255 of venues for Bucharest and a sample is listed below:

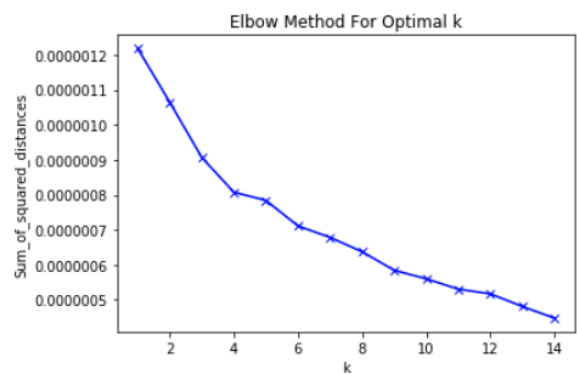|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	1 Mai	44.4711	26.05	Lebanese Station	44.471121	26.050584	Lebanese Restaurant
1	1 Mai	44.4711	26.05	The Coco Pub	44.469502	26.057233	Pub
2	1 Mai	44.4711	26.05	Mega Image	44.476955	26.044773	Supermarket
3	1 Mai	44.4711	26.05	Mega Image	44.471190	26.051079	Supermarket
4	1 Mai	44.4711	26.05	Ubisoft Gym	44.473731	26.057945	Gym / Fitness Center

If we count the number of venues returned by Foursquare API in each neighborhood we observe that in some (Calea Calarasilor, Dacia-Eminescu) we reached the limit but in others (1 Mai, Andronache, Bucuresti Noi) we have a smaller number.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1 Mai	13	13	13	13	13	13
13 Septembrie-Panduri	58	58	58	58	58	58
Andronache	21	21	21	21	21	21
Aviatiei	36	36	36	36	36	36
Aviatorilor	49	49	49	49	49	49
Baicului	54	54	54	54	54	54
Baneasa	38	38	38	38	38	38
Berceni	71	71	71	71	71	71
Bucuresti Noi	9	9	9	9	9	9
Calea Calarasilor	100	100	100	100	100	100
Colentina	22	22	22	22	22	22
Cotroceni	78	78	78	78	78	78
Crangasi	43	43	43	43	43	43
Dacia-Eminescu	100	100	100	100	100	100

- 245 unique categories of venues were returned by Foursquare and then I created a table which shows list of top 10 venue category for each Neighborhood:

	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	1 Mai	Gym / Fitness Center	Supermarket	ATM	Bakery	Lebanese Restaurant	Fast Food Restaurant	Electronics Store	Pizza Place	Pub	Bus Station
1	13 Septembrie-Panduri	Restaurant	Bar	Plaza	Lounge	Supermarket	Café	Romanian Restaurant	Italian Restaurant	Grocery Store	Gym / Fitness Center
2	Andronache	Turkish Restaurant	Café	Restaurant	Bus Station	Fried Chicken Joint	Fast Food Restaurant	Auto Dealership	Furniture / Home Store	Herbs & Spices Store	Middle Eastern Restaurant
3	Aviatiei	Restaurant	Supermarket	Gym / Fitness Center	Hotel	Dessert Shop	Gym	Shopping Mall	Grocery Store	Steakhouse	Beer Garden
4	Aviatorilor	Coffee Shop	Café	Romanian Restaurant	History Museum	Dessert Shop	Middle Eastern Restaurant	Park	Bakery	Mediterranean Restaurant	Restaurant

- I used the Elbow method for KMeans clustering machine learning algorithm in order to determine the best number of clusters to be calculated for the city of Bucharest. The result is 7 for the given data.

Elbow Method For Optimal k

- K-means algorithm is a powerful machine learning tool used mainly to group data based on its similarity and uses unsupervised learning technique. Because the data of Average House Sales Price is not in the same range with the venues data I used Sklearn Python Library normalize algorithm on the entire Data Frame.
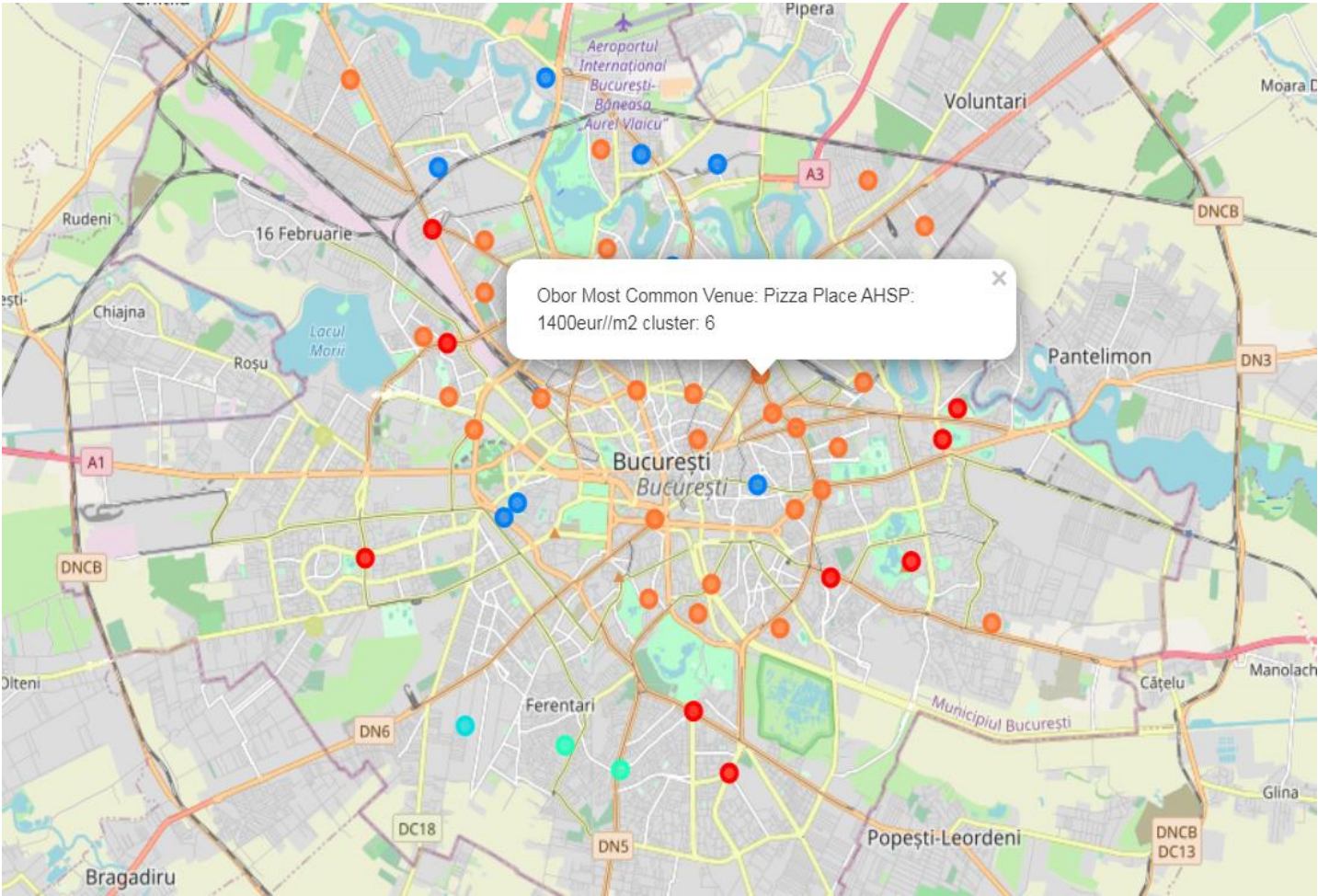
C. Results

After merging the cluster data the neighborhoods table looks like this:

	Neighborhood	Average House Price	Latitude	Longitude	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	Cluster Labels
0	1 Mai	1690	44.471100	26.050000	Gym / Fitness Center	Supermarket	ATM	Bakery	Lebanese Restaurant	Fast Food Restaurant	Electronics Store	Pizza Place	Pub	Bus Station	0
1	13 Septembrie-Panduri	1250	44.427539	26.066529	Restaurant	Bar	Plaza	Lounge	Supermarket	Café	Romanian Restaurant	Italian Restaurant	Grocery Store	Gym / Fitness Center	2
2	Aviatiei	2060	44.482300	26.098100	Restaurant	Supermarket	Gym / Fitness Center	Hotel	Dessert Shop	Gym	Shopping Mall	Grocery Store	Steakhouse	Beer Garden	2
3	Andronache	1300	44.478500	26.150100	Turkish Restaurant	Café	Restaurant	Bus Station	Fried Chicken Joint	Fast Food Restaurant	Auto Dealership	Furniture / Home Store	Herbs & Spices Store	Middle Eastern Restaurant	6
4	Baicului	990	44.448100	26.149300	Supermarket	Park	Café	Fast Food Restaurant	Gym / Fitness Center	Bakery	Department Store	Burger Joint	Soccer Field	Italian Restaurant	6

Over the map of Bucharest generated by Folium Python Library I added markers colored based on their clusters, and if you click on each bullet it shows additional information regarding the first most common venue and the average house sales price in that neighborhood.

Below you will find the tables with the data about the House sales price and the 10th most common venues for each neighborhood grouped by their similarity:

Cluster 1

	Neighborhood	Average House Price	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	Cluster Labels
0	1 Mai	1690	Gym / Fitness Center	Supermarket	ATM	Bakery	Lebanese Restaurant	Fast Food Restaurant	Electronics Store	Pizza Place	Pub	Bus Station	0
6	Berceni	1130	Clothing Store	Café	Pizza Place	Gym	Gym / Fitness Center	Pharmacy	Burger Joint	Supermarket	Coffee Shop	Italian Restaurant	0
15	Doamna Ghica	990	Supermarket	Gym	Department Store	Park	Fast Food Restaurant	Sushi Restaurant	Burger Joint	Shopping Mall	Bus Station	Café	0
16	Dristor-Vitan Vechi	1380	Supermarket	Coffee Shop	Pizza Place	Fried Chicken Joint	Italian Restaurant	Burger Joint	Bakery	Doner Restaurant	Gym / Fitness Center	Restaurant	0
17	Drumul Taberei	1100	Supermarket	Italian Restaurant	Restaurant	Bus Station	Bakery	Pizza Place	Coffee Shop	Clothing Store	Romanian Restaurant	Café	0

Cluster 2

	Neighborhood	Average House Price	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	Cluster Labels
22	Fundeni	1040	Auto Workshop	Food & Drink Shop	Eastern European Restaurant	Gym / Fitness Center	Eye Doctor	French Restaurant	Fountain	Football Stadium	Food Truck	Food	1

Cluster 3

	Neighborhood	Average House Price	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	Cluster Labels
1	13 Septembrie-Panduri	1250	Restaurant	Bar	Plaza	Lounge	Supermarket	Café	Romanian Restaurant	Italian Restaurant	Grocery Store	Gym / Fitness Center	2
2	Aviatiei	2060	Restaurant	Supermarket	Gym / Fitness Center	Hotel	Dessert Shop	Gym	Shopping Mall	Grocery Store	Steakhouse	Beer Garden	2
5	Baneasa	1670	Restaurant	Café	Hotel	Romanian Restaurant	Park	Supermarket	Italian Restaurant	Doner Restaurant	Farmers Market	Lake	2
8	Calea Calarasilor	1700	Café	Restaurant	Lounge	Italian Restaurant	Romanian Restaurant	Hotel	Coffee Shop	Tea Room	Bakery	Bar	2
10	Cotroceni	1700	Restaurant	Plaza	Café	Bistro	Bar	Italian Restaurant	Lounge	Coffee Shop	Supermarket	Tea Room	2

Cluster 4

	Neighborhood	Average House Price	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	Cluster Labels
42	Rahova	990	Light Rail Station	Supermarket	Farmers Market	Grocery Store	Park	Department Store	Fish Market	Fast Food Restaurant	Festival	Women's Store	3

Cluster 5

	Neighborhood	Average House Price	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	Cluster Labels
19	Ferentari	800	Bus Station	Department Store	Food & Drink Shop	Auto Garage	Electronics Store	Farmers Market	Park	Food Truck	Football Stadium	Food	4
26	Giurgiului	1040	Supermarket	Bus Station	Gym	Park	Fast Food Restaurant	Bagel Shop	Light Rail Station	Farmers Market	Electronics Store	Cupcake Shop	4

Cluster 6

	Neighborhood	Average House Price	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	Cluster Labels
24	Ghencea	880	Bus Station	Gym / Fitness Center	Grocery Store	Supermarket	Italian Restaurant	Bakery	Lounge	Romanian Restaurant	Fast Food Restaurant	Athletics & Sports	5
32	Militari	1040	Supermarket	Grocery Store	Pizza Place	Restaurant	Italian Restaurant	Gym	Gym / Fitness Center	Gym Pool	Cosmetics Shop	Café	5

Cluster 7

	Neighborhood	Average House Price	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	Cluster Labels
3	Andronache	1300	Turkish Restaurant	Café	Restaurant	Bus Station	Fried Chicken Joint	Fast Food Restaurant	Auto Dealership	Furniture / Home Store	Herbs & Spices Store	Middle Eastern Restaurant	6
4	Baicului	990	Supermarket	Park	Café	Fast Food Restaurant	Gym / Fitness Center	Bakery	Department Store	Burger Joint	Soccer Field	Italian Restaurant	6
7	Bucuresti Noi	1300	Gym	Soccer Field	Cupcake Shop	Supermarket	Grocery Store	Dessert Shop	Theater	Food & Drink Shop	Falafel Restaurant	Flower Shop	6
9	Colentina	1110	Café	Pizza Place	Gym	Fried Chicken Joint	Gift Shop	Electronics Store	Farmers Market	Park	Gym / Fitness Center	Restaurant	6
11	Crangasi	1210	Supermarket	Park	Grocery Store	Pizza Place	Pub	Climbing Gym	Farmers Market	Café	Gym	Lake	6

D. Discussion

Bucharest is a big city with a high population density and each neighborhood has its unique characteristics. The total number of measurements and population densities of the 55 analyzed neighborhoods can vary. For this very complex city, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same quality results.

I used the Kmeans algorithm as part of this clustering study and the number of clusters was established with the Elbow method, but, as part of setting the optimum k value to 7, was also my the knowledge of the city and its neighborhoods characteristics (For example clusters 2,4,5,6 contain very unique neighborhoods that could not be grouped with any other, and this might be the subject of a different analysis). Only 55 neighborhoods coordinates were used. For more detailed and accurate guidance, the data set can be expanded with the details of the streets.

F. Conclusion

There are 7 types of neighborhoods in Bucharest clustered according with the house sales prices and venues around the geographical location of each one. Taking in consideration the long time commute from a location to another in this city the opportunities to move in a new location that will maintain your life style are very high and as a starting point this study will give anybody some insights for making a decision.

G. References:

- https://en.wikipedia.org/wiki/Bucharest
- https://www.analizeimobiliare.ro/
- https://foursquare.com/
- https://www.google.com/maps