Istanbul & the arcades:

An analysis of the current situation and future business possibilities through ML and Geographical representation

Coursera capstone project

IBM Data Science professional certificate

Piero Abbondanza

17.03.2021

Table of contents

1. Introduction

The game sector, part of the broader entertainment industry, reach everyday new user and possible buyers. This continuous prospect of growth is at the base of the creation of gaming exchange traded funds and an increasing trust in the solidity of such business.

In fact, the entertainment sector brought over the year solid returns of investments. The same sector has shown very resilient in its upward trend even during the covid pandemic. Yet there there is still some scepticism toward a business that rely on the expenditure of young generations and that is very high-tech, therefore also costly to innovate.

Business case/study case

Istanbuls offers a different perspective with respect to the rich-countries behaviour: most of the gamers don't own a console, rather they visit Arcade/ playstation café. This place can be defined as a location that for a meager payment (low-cost) offers the experience of playing with friends, online or alone any game. This is often connected to the consumption of food, beverages and related merchandise.

Thus, it happens that an intermediary, that can lower the cost through a sharing/renting of the good and overcome the limitation of the initial high-cost of consoles and games, exists. Consequently the intermediary can address a large crowd of user, with little regard to their current economic condition. An hypothesis we are going to test further in the analysis is as follows: low average income parts of the city will have more arcades.

With this hypothesis in mind we can infer and then research for example:

- High density of population (population/square kilometer) could mean more arcades;
- High density of population also means lower annual or monthly income, which in turn means less console ownership and more arcades
- Does geographical location and annexed characteristics affect the number of arcades?

Here the aim is to test such questions regarding the arcade business in istanbul in a first draft of a more complex analysis. Finally, this study case and projections, when complete, could be useful for replicating the phenomenon in other locations or offer services/partnerships that can complete it. It is not by chance that I could visit some arcade places in very different location myself after leaving Istanbul

2. The Data-frame

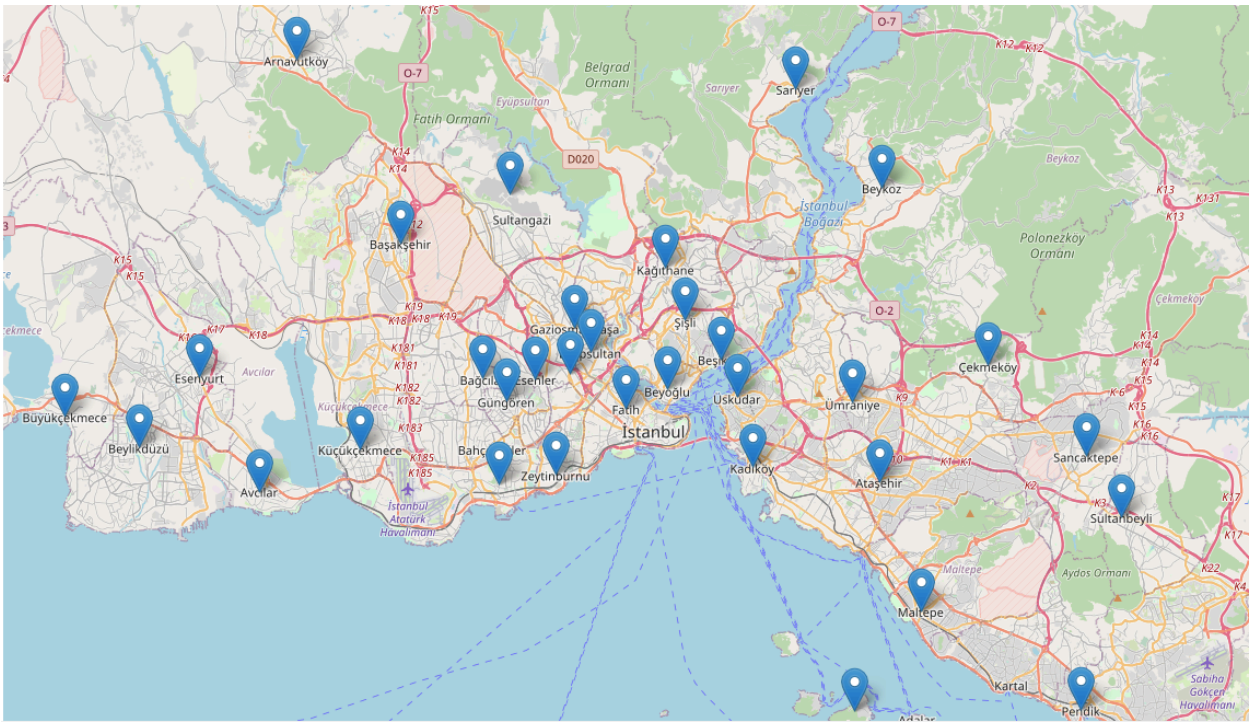
Structure of the data

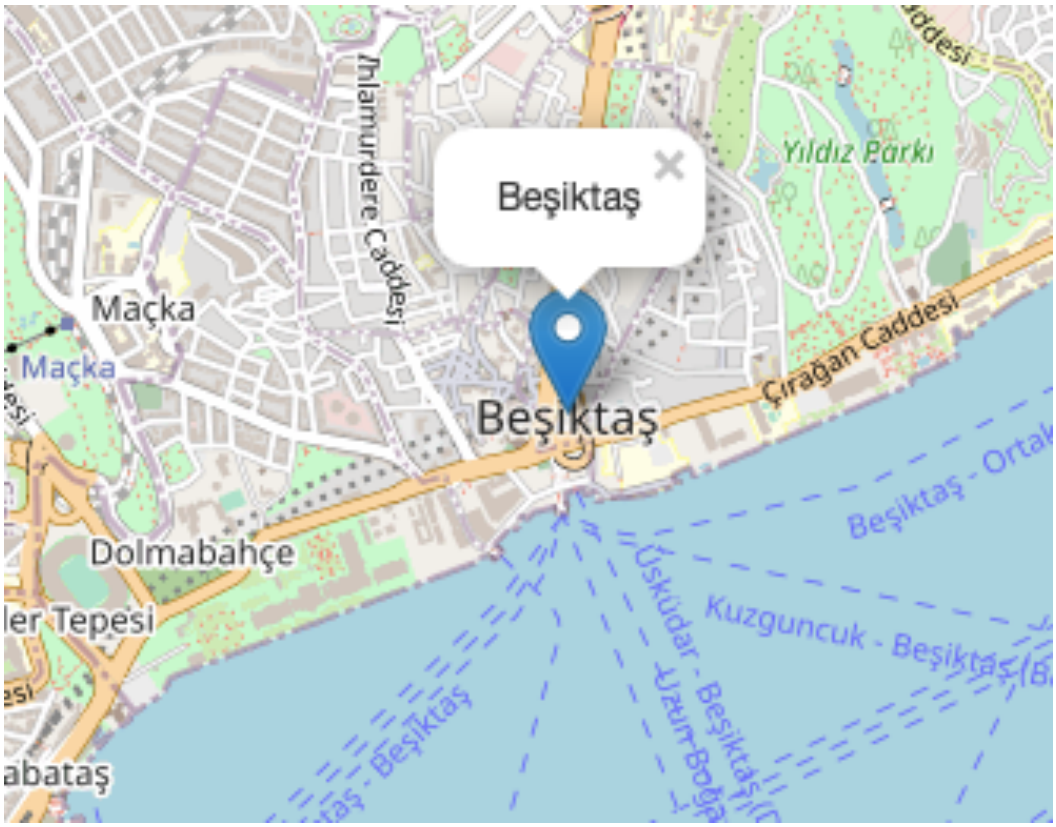
For this analysis I proceeded and build our data sources through the Coursera exercises. First of all the first data set was the Districts in Istanbul from the Wikipedia website. Since they are in a table format (html element) it has been a fast and direct download and import in the code of the data. Worth of mention here are two aspects of this method:

- This option is not available for all cities on the wikipedia server. However there are many alternative options here;
- Once downloaded and wrote the script on the original data frame, I had to save it since the Wikipedia webpage is constantly updated and this could cause problem if I run the code in a later point in time.

Next the table of districts and related data for Istanbul:

#	District	pop2020	area (km^2)	density	monthly_income	annual_income
0	Adalar	16033,00	11,05	1451,00	$ 918,00	$ 10.978,00
1	Arnavutköy	296709,00	450,35	659,00	$ 279,00	$ 3.350,00
2	Ataşehir	422594,00	25,23	16750,00	$ 904,00	$ 10.854,00
3	Avcılar	436897,00	42,01	10400,00	$ 503,00	$ 6.064,00
4	Bağcılar	737206,00	22,36	32970,00	$ 441,00	$ 5.295,00
5	Bahçelievler	592371,00	16,62	35642,00	$ 645,00	$ 7.741,00
6	Bakırköy	226229,00	29,64	7633,00	$ 1.220,00	$ 14.650,00
7	Başakşehir	469924,00	104,3	4506,00	$ 622,00	$ 7.474,00
8	Bayrampaşa	269950,00	9,61	28091,00	$ 480,00	$ 5.764,00
9	Beşiktaş	176513,00	18,01	9801,00	$ 1.457,00	$ 17.490,00
10	Beykoz	246110,00	310,36	793,00	$ 509,00	$ 6.116,00
11	Beylikdüzü	365572,00	37,78	9676,00	$ 597,00	$ 7.166,00
12	Beyoğlu	226396,00	8,91	25409,00	$ 658,00	$ 7.905,00
13	Büyükçekmece	257362,00	139,17	1849,00	$ 506,00	$ 6.079,00
14	Çatalca	74975,00	1115,1	67,00	$ 293,00	$ 3.524,00
15	Çekmeköy	273658,00	148,09	1848,00	$ 483,00	$ 5.801,00
16	Esenler	446276,00	18,43	24215,00	$ 392,00	$ 4.715,00
17	Esenyurt	957398,00	43,13	22198,00	$ 417,00	$ 5.008,00
18	Eyüpsultan	405845,00	228,42	1777,00	$ 644,00	$ 7.735,00
19	Fatih	396594,00	15,59	25439,00	$ 728,00	$ 8.747,00
20	Gaziosmanpaşa	487778,00	11,76	41478,00	$ 416,00	$ 5.000,00
21	Güngören	280299,00	7,21	38876,00	$ 467,00	$ 5.611,00
22	Kadıköy	481983,00	25,09	19210,00	$ 1.245,00	$ 14.948,00

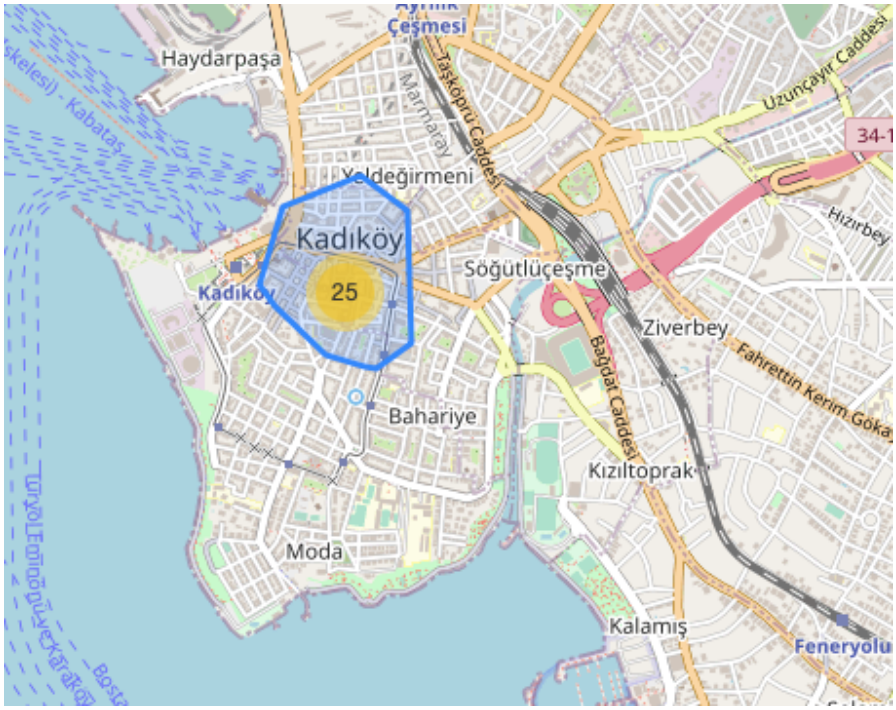




23	Kağıthane	442415,00	14,87	29752,00	$	578,00	$	6.937,00
24	Kartal	474514,00	38,54	12312,00	$	568,00	$	6.824,00

25	Küçükçekmece	789633,00	37,54	21034,00	$	492,00	$	5.908,00
26	Maltepe	515021,00	52,97	9723,00	$	796,00	$	9.559,00
27	Pendik	726481,00	179,99	4036,00	$	421,00	$	5.060,00
28	Sancaktepe	456861,00	62,42	7319,00	$	363,00	$	4.361,00
29	Sarıyer	335298,00	175,39	1912,00	$	1.008,00	$	12.104,00
30	Silivri	200215,00	869,52	230,00	$	327,00	$	3.928,00
31	Sultanbeyli	343318,00	29,14	11782,00	$	299,00	$	3.597,00
32	Sultangazi	537488,00	36,3	14807,00	$	301,00	$	3.622,00
33	Şile	37904,00	781,72	48,00	$	342,00	$	4.111,00
34	Şişli	266793,00	10,71	24911,00	$	1.079,00	$	12.955,00
35	Tuzla	273608,00	123,63	2213,00	$	470,00	$	5.643,00
36	Ümraniye	713803,00	45,31	15754,00	$	502,00	$	6.023,00
37	Üsküdar	520771,00	35,33	14740,00	$	964,00	$	11.572,00
38	Zeytinburnu	283657,00	11,59	24474,00	$	502,00	$	6.036,00

table I delete rows most



From the original had the the N/A and the distant districts could be that they enough to be relevant analysis.

since it said are not central

for the This statement is confirmed by the fact that when combining the foursquare data, no arcade or interesting location was found belonging to such distant districts. The total number of rows deleted were 4 out of 42.

To sum up, after downloading the raw data, formatting them correctly as table through pandas and cleaning the data frame we have our starting point for the analysis and can proceed by integrating the geographical reference. The last point has been done with the help of the python library geopy.

Geographical representation

Today most of the data visualisation involves a geographical representation of data and analysis. The geographical element has a very high explanatory power in such cases. For the scope of this research it is also the case of using geographical data visualizations that can explain our data, their distribution and hypothesis.

It follows that through the call of data for Istanbul, their combination with the foursquare "Arcade" category and the clustering method (supported by the library selenium for python), we obtained a very clear graphical representation of the data.

Here you can have a brief overview of the resulting maps:

Figure 1: Entire map of Istanbul with districts

Figure 2: Detail pinpoint of a district

Figure 3: Detail cluster in a district after combining the data from foursquare

Figure 4: Detail arcade pinpoint with attached data

Foursquare

By using the data from Foursquare and the proposed geographical representation of them it is possible to research the above mentioned business problem and offer an insight for possible new arcades, new business partnerships, or the offering of a service such as restock of beverages/console parts and gadgets in a lively and highly populated city as Istanbul.

The analysis of the phenomenon could also lead to the idea of opening an arcade chain in location with similar geographical, economical and social conditions/characteristics. In fact, while we talk a new chain opened in Italy near my home!

Next, I show you a summarising table of various grouping of data from Foursquare with relation to the original districts' table:

#	District	pop2020	area (km^2)	density	monthly_income	annual_income	Number Arcades	Number of arcade per 10000 inh.
22	Kadıköy	481983,00	25,09	19210,00	$ 1.245,00	$ 14.948,00	25	0,519
9	Beşiktaş	176513,00	18,01	9801,00	$ 1.457,00	$ 17.490,00	22	1,246
30	Silivri	200215,00	869,52	230,00	$ 327,00	$ 3.928,00	14	0,699
3	Avcılar	436897,00	42,01	10400,00	$ 503,00	$ 6.064,00	12	0,275
27	Pendik	726481,00	179,99	4036,00	$ 421,00	$ 5.060,00	11	0,151
4	Bağcılar	737206,00	22,36	32970,00	$ 441,00	$ 5.295,00	10	0,136
1	Arnavutköy	296709,00	450,35	659,00	$ 279,00	$ 3.350,00	9	0,303
26	Maltepe	515021,00	52,97	9723,00	$ 796,00	$ 9.559,00	8	0,155
36	Ümraniye	713803,00	45,31	15754,00	$ 502,00	$ 6.023,00	8	0,112
20	Gaziosmanpaşa	487778,00	11,76	41478,00	$ 416,00	$ 5.000,00	8	0,164
38	Zeytinburnu	283657,00	11,59	24474,00	$ 502,00	$ 6.036,00	8	0,282
37	Üsküdar	520771,00	35,33	14740,00	$ 964,00	$ 11.572,00	7	0,134
25	Küçükçekmece	789633,00	37,54	21034,00	$ 492,00	$ 5.908,00	6	0,076
19	Fatih	396594,00	15,59	25439,00	$ 728,00	$ 8.747,00	6	0,151
8	Bayrampaşa	269950,00	9,61	28091,00	$ 480,00	$ 5.764,00	5	0,185
11	Beylikdüzü	365572,00	37,78	9676,00	$ 597,00	$ 7.166,00	4	0,109
31	Sultanbeyli	343318,00	29,14	11782,00	$ 299,00	$ 3.597,00	4	0,117
16	Esenler	446276,00	18,43	24215,00	$ 392,00	$ 4.715,00	4	0,090
12	Beyoğlu	226396,00	8,91	25409,00	$ 658,00	$ 7.905,00	3	0,133
14	Çatalca	74975,00	1115,1	67,00	$ 293,00	$ 3.524,00	2	0,267
13	Büyükçekmece	257362,00	139,17	1849,00	$ 506,00	$ 6.079,00	2	0,078
17	Esenyurt	957398,00	43,13	22198,00	$ 417,00	$ 5.008,00	2	0,021
6	Bakırköy	226229,00	29,64	7633,00	$ 1.220,00	$ 14.650,00	2	0,088
34	Şişli	266793,00	10,71	24911,00	$ 1.079,00	$ 12.955,00	2	0,075
21	Güngören	280299,00	7,21	38876,00	$ 467,00	$ 5.611,00	2	0,071
7	Başakşehir	469924,00	104,3	4506,00	$ 622,00	$ 7.474,00	1	0,021
23	Kağıthane	442415,00	14,87	29752,00	$ 578,00	$ 6.937,00	1	0,023
33	Şile	37904,00	781,72	48,00	$ 342,00	$ 4.111,00	0	0,000

10	Beykoz	246110,00	310,36	793,00	$	509,00	$	6.116,00	0	0,000
18	Eyüpsultan	405845,00	228,42	1777,00	$	644,00	$	7.735,00	0	0,000
29	Sarıyer	335298,00	175,39	1912,00	$	1.008,00	$	12.104,00	0	0,000
15	Çekmeköy	273658,00	148,09	1848,00	$	483,00	$	5.801,00	0	0,000
35	Tuzla	273608,00	123,63	2213,00	$	470,00	$	5.643,00	0	0,000
28	Sancaktepe	456861,00	62,42	7319,00	$	363,00	$	4.361,00	0	0,000
24	Kartal	474514,00	38,54	12312,00	$	568,00	$	6.824,00	0	0,000
32	Sultangazi	537488,00	36,3	14807,00	$	301,00	$	3.622,00	0	0,000
2	Ataşehir	422594,00	25,23	16750,00	$	904,00	$	10.854,00	0	0,000
5	Bahçelievler	592371,00	16,62	35642,00	$	645,00	$	7.741,00	0	0,000
0	Adalar	16033,00	11,05	1451,00	$	918,00	$	10.978,00	0	0,000

For the complete analysis and combination of data as well as the relationships obtained through clustering please check the python script.

3. Methodology

For this analysis the most suitable ML Method was Clustering. Clustering is a method of grouping data points into similar clusters. I similar studies or literature clustering is often also called segmentation.

Aim here is not to fig deep into the structure of the algorithm but to present the general characteristics and what we can infer after it. K-clustering library in python was used and the number of cluster chosen has been 5. The number is also suitable since the city of Istanbul can be divided in five main sections around the Bosphorus and 5 is the courser suggested number in the "one hot" coding exercise. With 10 points for example our analysis was not feasible. Our features to group data has been the latitude & longitude combination.

In conclusion let's look at some characteristics of the method:

- non-supervised learning technique because it can be used in unlabelled data;
- clustering is a useful first step for discovering new patterns;
- It requires little prior knowledge about how the data might be structured or how items are related;
- often used for exploration of data prior to analysis with other more predictive algorithms.

With relation to the last mentioned point and as stated at the beginning, this analysis is exploratory for possible development and does not aim at offering projections and hypothesis that may involve more complex ML models / methods.

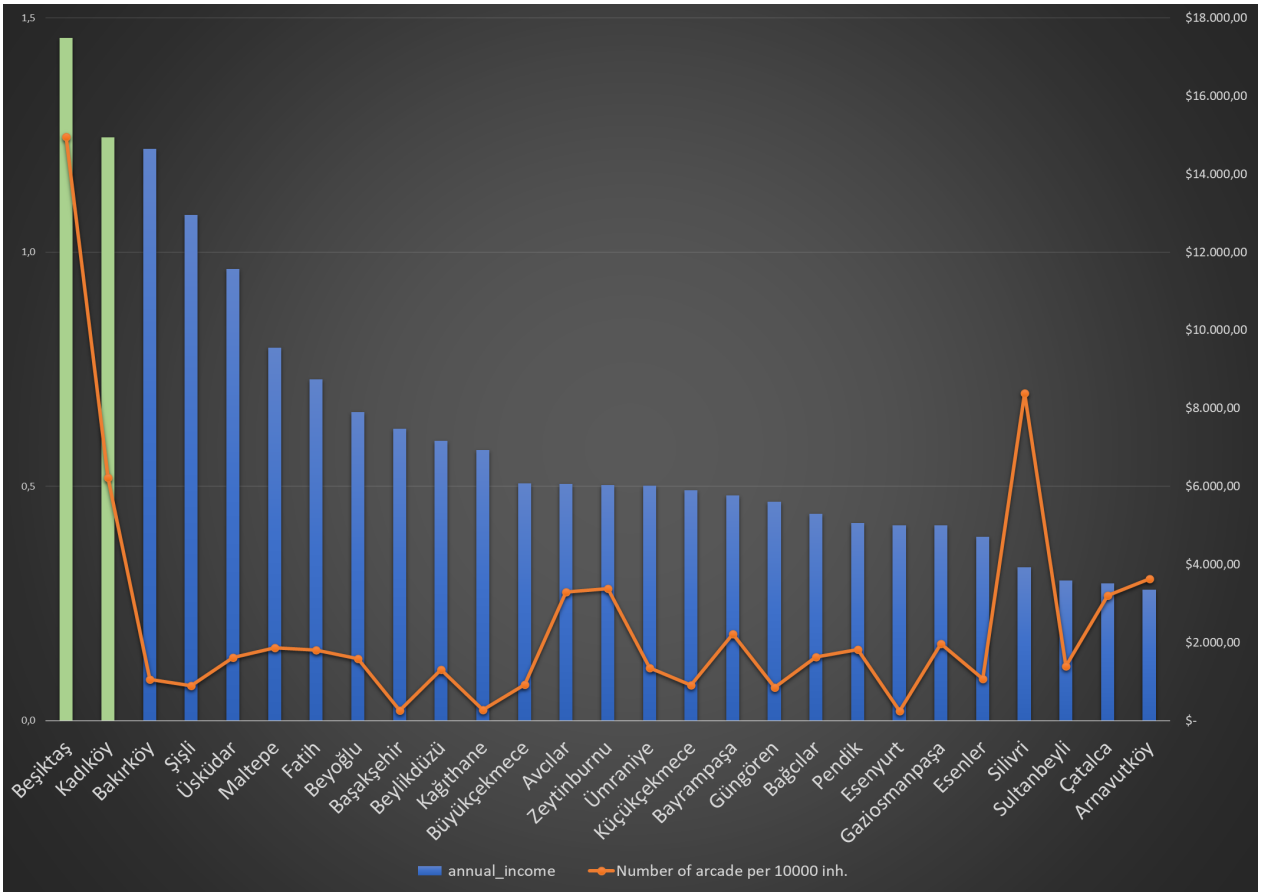
4. Results

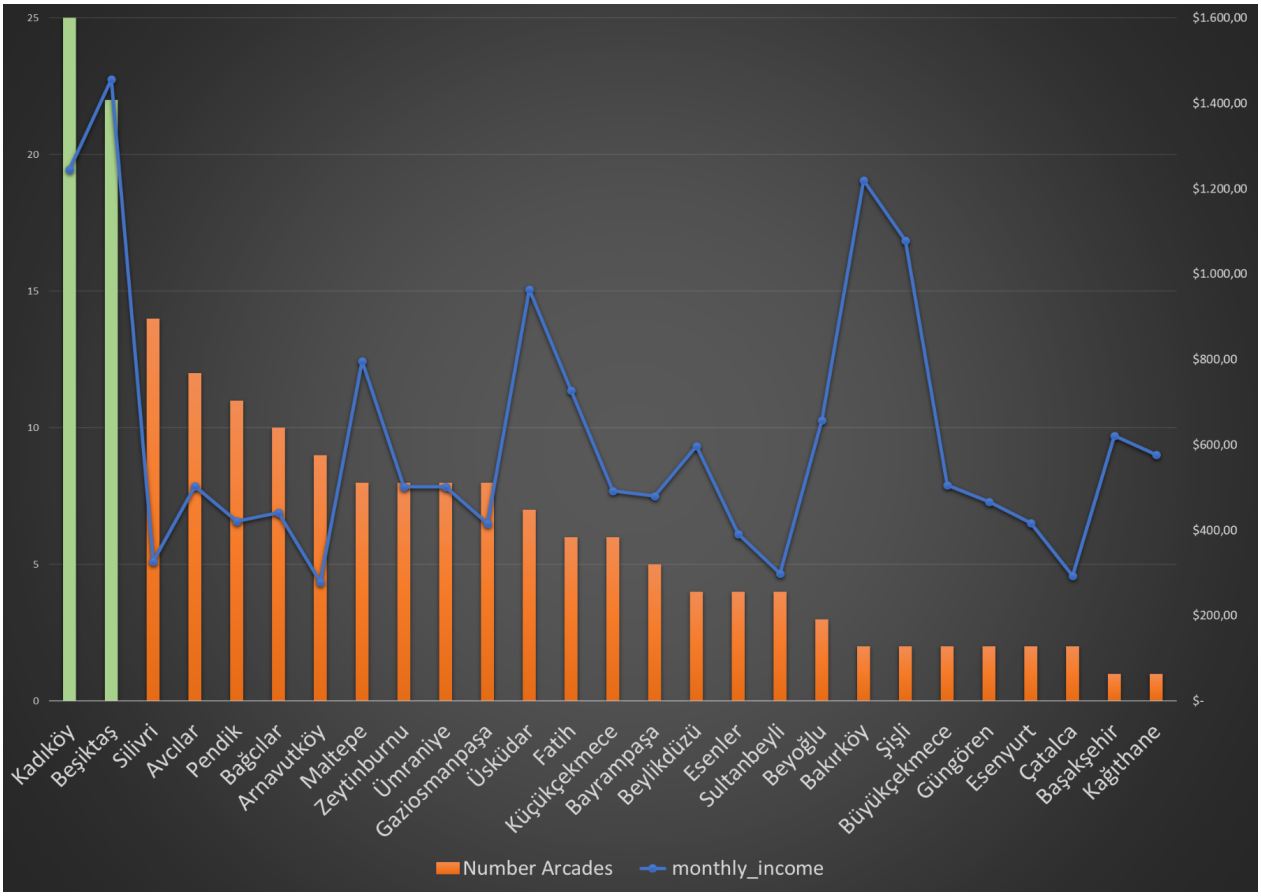
Reaching the result, we can state to be happy with the first steps of the analysis. The application of the clustering was successful and the best number of cluster is 4. This divides the the city into four sections, which corresponds to the most central district of each side, plus north and south reference, of the city. This confirms also my impression of Istanbul and the organisation of the metropolis. Important to understand is the fact that when analysing cities of such demotions it is natural that they are organised as "smaller" cities inside. This confirms my experience as citizen there.

Coming to the business related answer following graphs presents the distribution of the arcades based on the annual income distribution of each district (ordered from higher to lower). This does not confirm our initial hypothesis. The trend is in fact more the opposite. Yet no clear economic patter can be found in the number of arcades per district. This may suggest there are different factors that determine the presence or not of arcade places. Some of theme could be:

- Hub for public transportation;
- Proximity to markets and other places that could symbolise a "city center", such as squares, food center, shopping mall;
- Other behavioural and cultural elements: distance from religious places (more conservative).

To the figures mentioned:


1,5
1,0
0,5
0,0
$18.000,00
$16.000,00
$14.000,00
$12.000,00
$10.000,00
$8.000,00
$6.000,00
$4.000,00
$2.000,00
$-
Beşiktaş
Kadıköy
Bakırköy
Şişli
Üsküdar
Maltepe
Fatih
Beyoğlu
Başakşehir
Beylikdüzü
Kağıthane
Büyükçekmece
Avcılar
Zeytinburnu
Ümraniye
Küçükçekmece
Bayrampaşa
Güngören
Bağcılar
Pendik
Esenyurt
Gaziosmanpaşa
Esenler
Silivri
Sultanbeyli
Çatalca
Arnavutköy
annual_income
Number of arcade per 10000 inh.


25
20
15
10
5
0
$1.600,00
$1.400,00
$1.200,00
$1.000,00
$800,00
$600,00
$400,00
$200,00
$-
Kadıköy
Beşiktaş
Silivri
Avcılar
Pendik
Bağcılar
Arnavutköy
Maltepe
Zeytinburnu
Ümraniye
Gaziosmanpaşa
Üsküdar
Fatih
Küçükçekmece
Bayrampaşa
Beylikdüzü
Esenler
Sultanbeyli
Beyoğlu
Bakırköy
Şişli
Büyükçekmece
Güngören
Esenyurt
Çatalca
Başakşehir
Kağıthane
Number Arcades
monthly_income

5. Discussion

As stated in the results Reston, we cannot confirm our initial hypothesis. Despite this and through some simple steps in writing a code it has been very easy to obtain a geographical representation of the data. This let us space of different hypothesis and have a better look on the whole arcade situation in Istanbul.

From Foursquare we know that there are pretty much 200 arcades and cafe that offer entertainment console services. Such places offer also snacks, beverages, gadgets and require constant restock. This also address the need of updated consoles, new games, accessories.

If walking with the shoes of an entrepreneur there are a broad range of business and return possibilities we can discuss about:

- Restock of supply of beverages/snacks as for bars and restaurants;
- Partnership for supplying consoles and accessories;
- Gather consumer affection and recognition with a chain of arcades (similar to fast food phenomenon);
- Cleaning, repairing & other related services (very important after the Covid 19 pandemic)
- Research of similar condition to open new arcades.

6. Conclusion

To sum up we constructed a database with geographical references (latitude & longitude) on the whole city of Istanbul, choosing to integrate it with foursquare data for the "Entertainment/arcade" category.

The aim was to undertake the first steps for a more complex and comprehensive analysis in the entertainment sector and behaviour in Istanbul, which as explained in the introductions shows very particular characteristics worth analysing. The value in analysing new behaviours is that if may offer business prospects and growth in other cities or more in general locations.

We succeeded in the introductory and explanatory analysis. Also we could test some hypothesis and prepared the ground for more statistical and machine learning analysis. If also the second part was completed then we could present very interesting projections and facts about the business case. Maybe this will be my topic for the next exercise ;)