How to buy best buy apartment in Zagreb

Natan Bonašin

28.02.2019

1. Introduction/Business Problem

1.Introduction

It is very hard to locate a apartment which will satisfy budget and also social life. For young families its very important so select a districts which will provide high quality infrastructure in form of schools, kinder gartens, public transport. In the same time they don't want to reduce quality of life by reducing access to venues like restaurants, pubs, museums, or other cultural venues.

1.1 Problem

In order to save time and money it is required to analyse huge amounts of data and even then eventual decision is due to lack of available information. Properties tend to get more and more expensive which maks decisio to pick a district even harder and more time consuming. Also sellers because they set their price too high might also wait too much to finish sale.

1.2 Interest

Understanding features which influence the price can help people who are searching for a property to focus on districts which is within their budget, in the same time sellers who have their price too high can in this analysis find a reason why this might happen.

2. DATA COLLECTION

2.1 Data sources

In order to get relevant data for this analysis i had to create geodata.json of city of Zgareb to be able to create cloropleth maps. Some general data about each city district i have scraped from wikipedia, and some data on properties value i have scraped from njuskalo.hr (probably the biggest portal for advertising in Croatia). I will also use data from fousquare.com to analyse number of venues per districts to check if it contributes to the price.

2.2 Data cleaning

As i don't use hige datasets to analyse correlation between few parameters, only dataset related to price of the properties had be cleaned. Some vlaues were not present in the datasets and some values were extreme, therefore i have excluded those to have more relevant results.

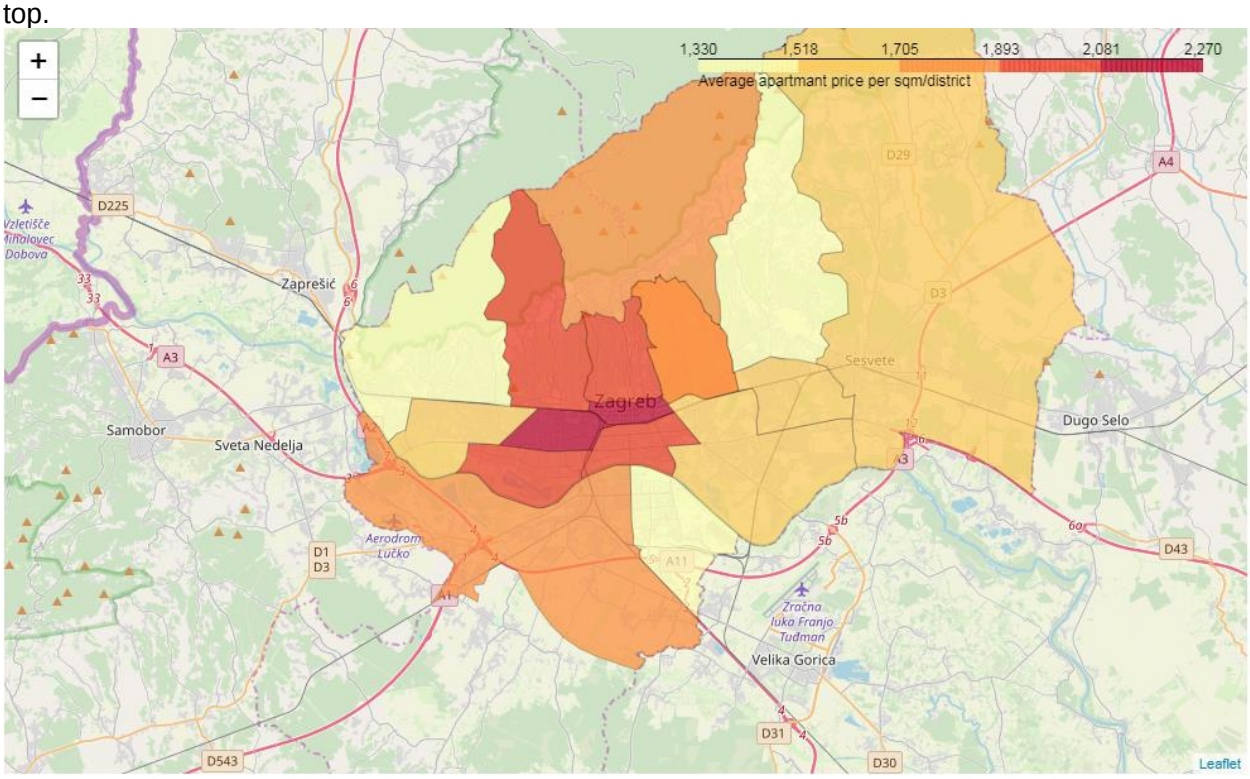
2.3 Feature selection

For this analysis i have decided to use district population, density(inhabitants/sqm),average asset price and number od venues per district.

4. DATA ANALYSIS

As i would like to analyse actual situation and accordingly find any anomaly in the market so can find locate district which has the best proportion between the price per square meter and contents like venues, pubs, restaraunts etc.

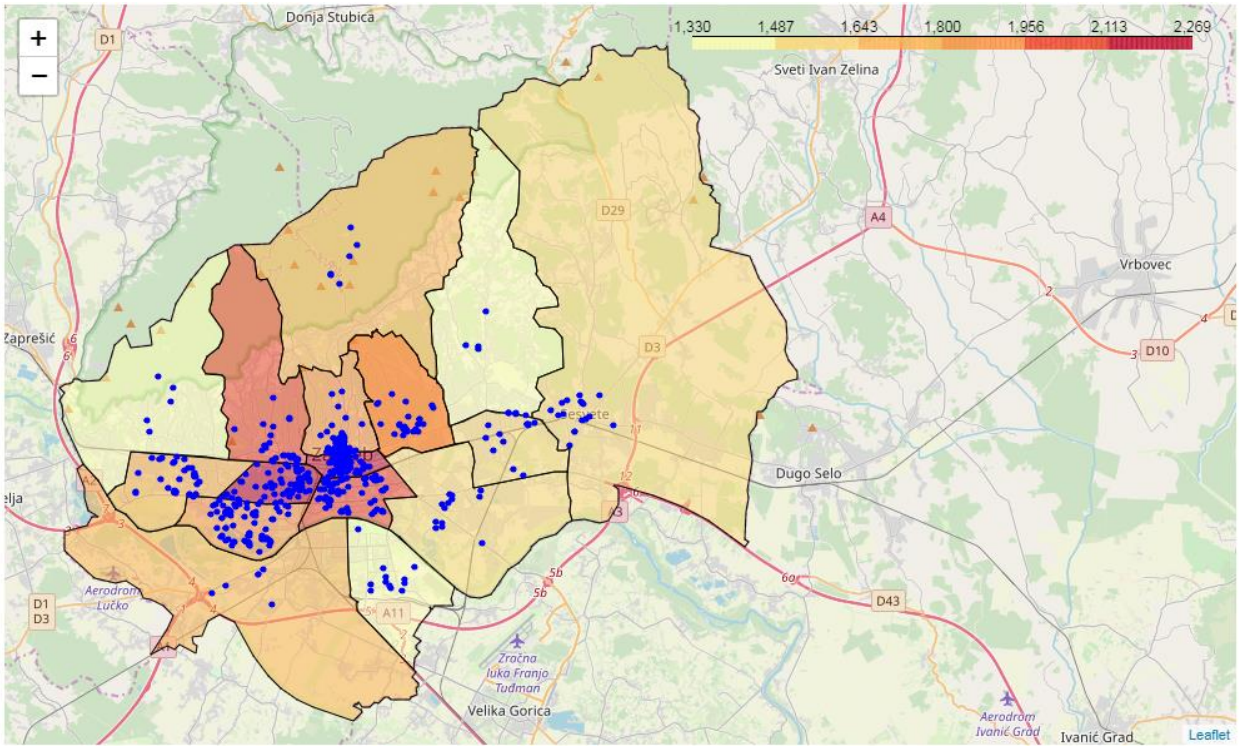
Firstly i am going to visualise actual map of Zagreb and create cloropleth map with average price on top.



Above map is actually as expected, apartment closer to city center on average cost more than those which more distant. it is worth noticing there 3 districts with lowest average price, of which Novi Zagreb - Istok is very well connected also dubrava is very good connected so form above map i would put some focus on those disctricts.

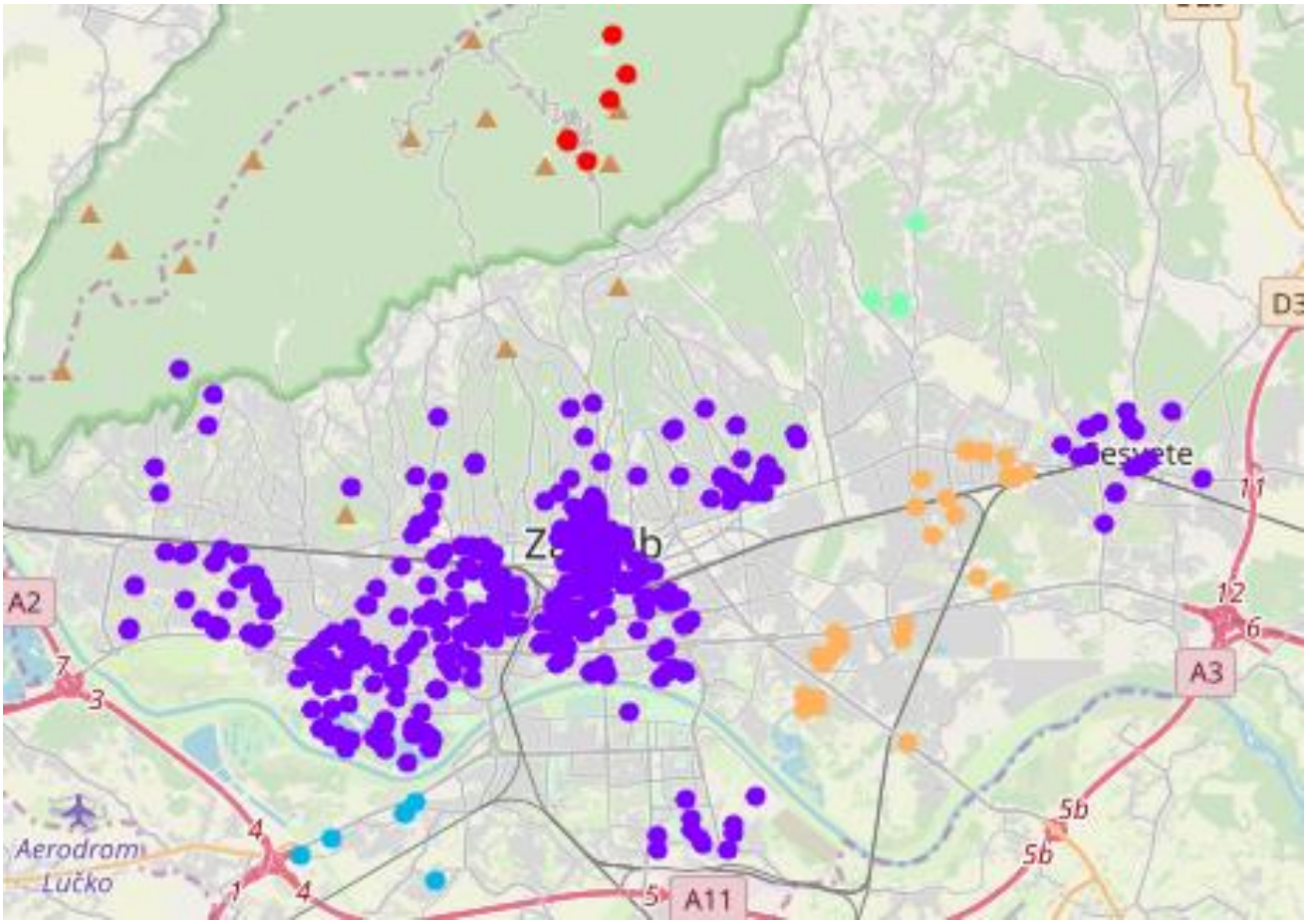
But said in introduction i would like to check also how well this districts stand regarding to content.

So i will use foursquare data to see how many venues are located per each district.



Inspecting above map now with venues on top of map with apartments average prices i can again confirm that those districts which are the most expensive have also more content than thos which are cheaper.

By analyzing each neighborhood and clustering I didn’t find any valuable result.



K-means clustering has helped here only to differentiate between major locations. So options are Zagreb center, Dubrava, Sesvete or top mountain Sljeme. So this ML technique i think is not so much esefull here. So i don't see any need to analyse any further.
Beside of the above findings i would like also to check if there is some correlation between population, density, average price, and size of the disctrict.

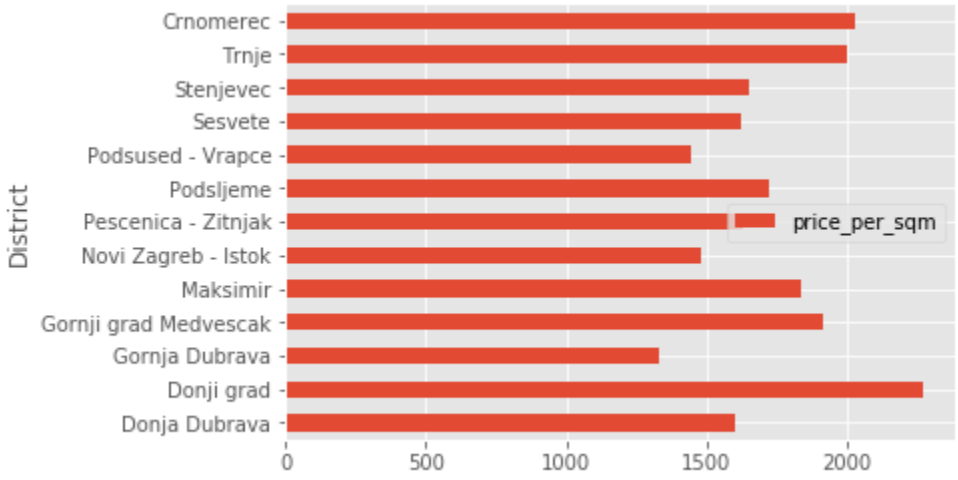
I have created one datasets which combines all collected data. I have tried to if any correlation.

Table 1.

	price_per_sqm	Venue	Lat	Lon	size	population	density
price_per_sqm	1.000000	0.731507	-0.094508	-0.193930	-0.309831	-0.335301	0.663027
Venue	0.731507	1.000000	-0.309355	-0.138885	-0.371247	-0.252046	0.907150
Lat	-0.094508	-0.309355	1.000000	0.031084	0.305993	-0.158676	-0.398806
Lon	-0.193930	-0.138885	0.031084	1.000000	0.476723	0.640320	-0.146984
size	-0.309831	-0.371247	0.305993	0.476723	1.000000	0.457096	-0.486604
population	-0.335301	-0.252046	-0.158676	0.640320	0.457096	1.000000	-0.173559
density	0.663027	0.907150	-0.398806	-0.146984	-0.486604	-0.173559	1.000000

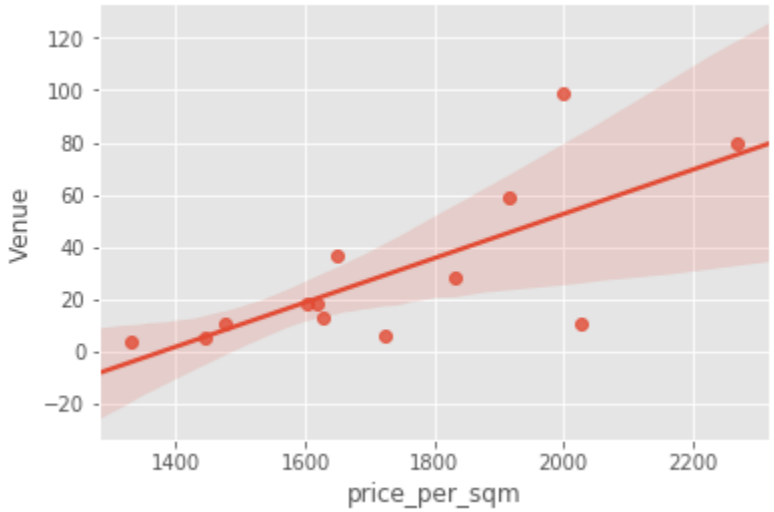
By observing above correlation chart i can notice that closest correlation is on number of venues if district heavily populated. Also number of venues is correlated to price per square meter.

Lets create some bar graph to additionally visualize data.

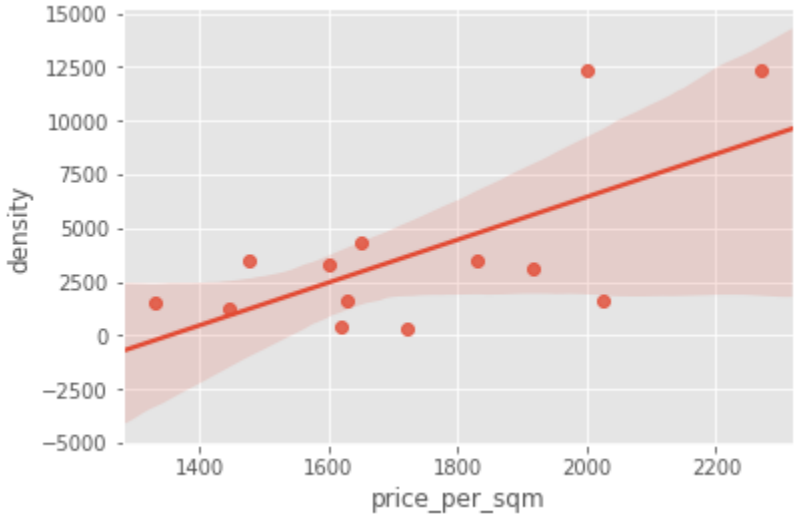


This graph helps to visualize difference in price between each district.

Lets check how compares price_per_sqm vs venue count



Now is a lot easier to notice how as number of venue is bigger also price is higher. Now i will do the same thing with density vs price per sqm.



If I check density(population/district size) it can be concluded that as area is more densely populated price is growing. So for those people who like to live without any parks, green or wood, they will actually have to pay more.

5. Discussion

Above findings are quite interesting and people are very much interested into data which might help them to locate apartmants which are within their budget while also attractive. My sample is too small to be of any significance but still it can present how much is different each district.

It would great if in some future project i could check on more granular level how each city neighborhood relates to other to perhaps notice some anomaly and use this data to find faster appropriate real estate.

I took only few datasets and mostly i have created them alone so i couldn't guess what would be result but still looks interesting. Perhaps some more datasets related to schools, offices, public transport might also help to dive deeper into understanding what is driving eal estate's price so high, or what is degrading its value.

6. Conclusion

Its almost impossibile to find "best buy" appartmant as people differ by huge margins. So some people who have bigger budget won't try to get lower price, while some might not even be able to get anything. From above findings i can conclude that best buy apartmant is definitely either in Novi Zagreb - istok or Gornja dubrava. Both district are in tram area which is very much important as tram travel quite often in zagreb. Both districts have quite low average price per square meter so there is greater chance to get higher quality real estate while not compromising easy access to city content and events.