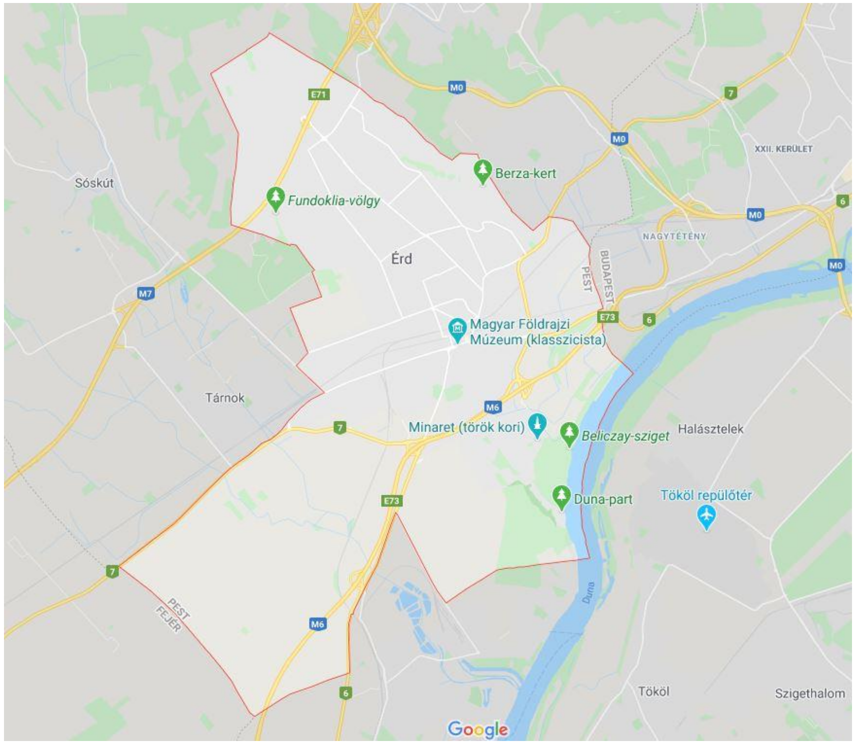
Analyzis of House Sales Prices in Érd, Hungary

Capstone project

Andrásné István Kiss

Érd

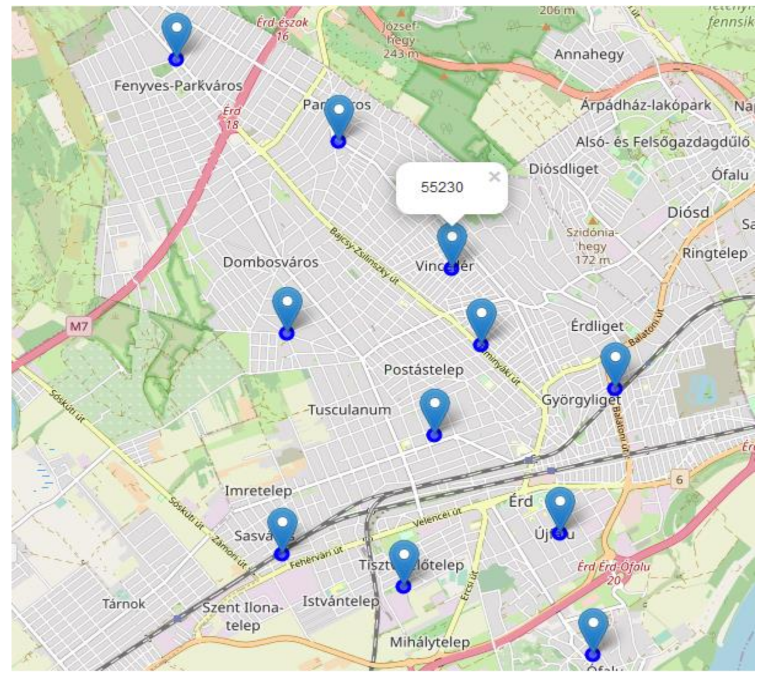
- Small town in Hungary near to the Capital, Budapest
- Its population has grown by 12% in 10 years
- 60% of the inhabitants works in the capital (commuter)
- the number of dwellings increased by 9%



In my project I make an analysis of the real estate prices in this town and the connection to the venue density.

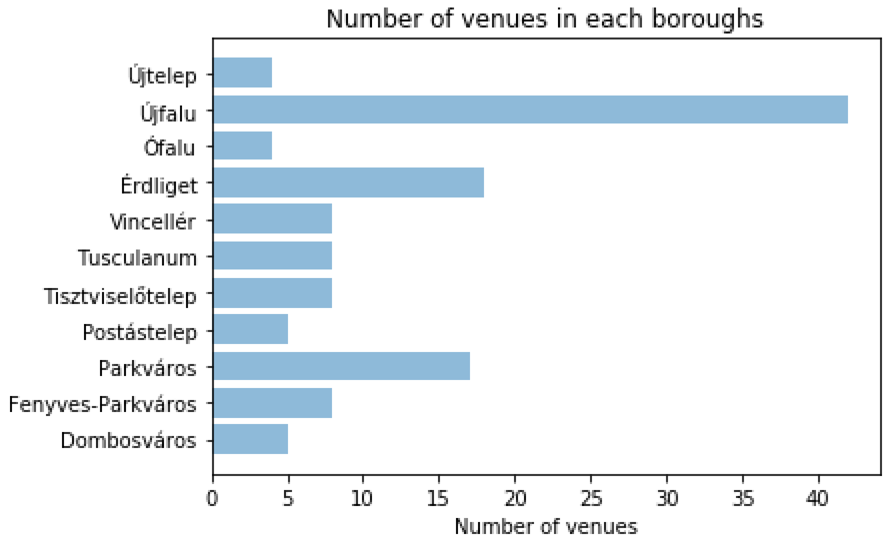
Methodology

	Borough	Price	Latitude	Longitude
0	Újfalu	43660	47.3740	18.9268
1	Tisztviselőtelep	54400	47.3686	18.9050
2	Újtelep	44490	47.3719	18.8881
3	Tusculanum	70960	47.3838	18.9093
4	Dombosváros	56610	47.3940	18.8888
5	Vincellér	55230	47.4006	18.9117
6	Parkváros	60460	47.4133	18.8960
7	Fenyves-Parkváros	71170	47.4216	18.8733
8	Ófalu	40140	47.3617	18.9315
9	Érdliget	63970	47.3885	18.9346
10	Postástelep	52630	47.3929	18.9159



Érd has 11 districts. There is a big difference among the average house prices, the lowest prices are in Ófalu, Újfalu and Újtelep - these are the oldest parts of the town near to the river Danube.

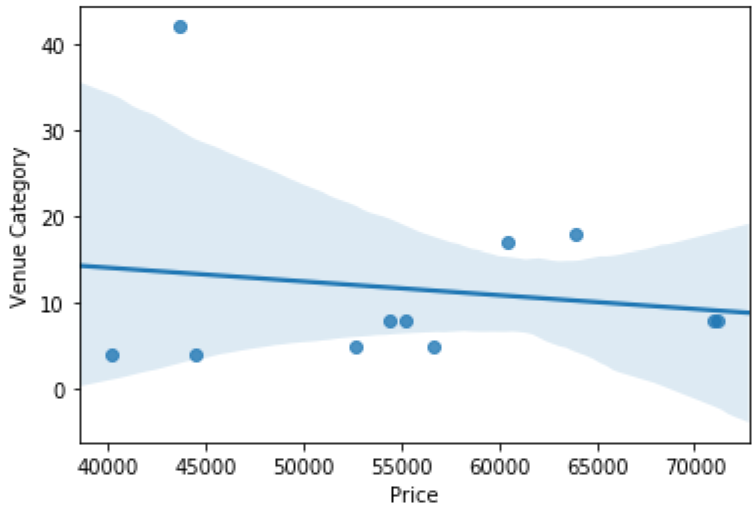
Exploring venues



Foursquare API found 127 venues in the boroughs in 61 unique categories.

The most venue is in Újfalu, Érdliget and Parkváros - these districts are near to the railway station and motorway, and these are the fastest and best developing areas in Érd. The other districts have almost the same number of venues, but under 10.

Connection between prices and venues



I try to find connection between the prices and the number of venues in each districts, so I use a correlation analysis for that (linear regression).

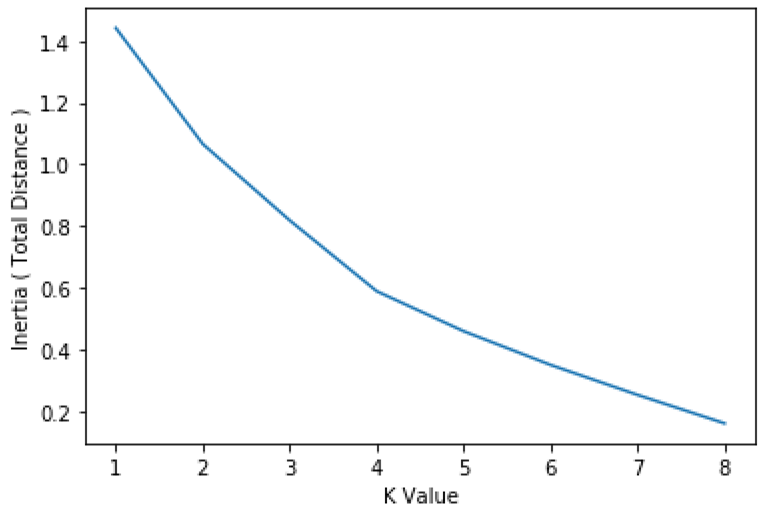
Unfortunately the regression plot shows a weak correlation between the price and number of venues in my town. That means that for people who move into this town lots of stores or venues are not so important, this feature does not increase the housing price.

Clustering

	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	Dombosváros	Bus Stop	Bus Station	Playground	Recording Studio	Train Station	Discount Store	Eastern European Restaurant	Electronics Store	Farm	Cupcake Shop
1	Fenyves-Parkváros	American Restaurant	Bus Stop	Forest	Gym	Intersection	Pharmacy	Cupcake Shop	Fruit & Vegetable Store	Eastern European Restaurant	Athletics & Sports
2	Parkváros	Bakery	Bus Stop	Grocery Store	Pizza Place	Greek Restaurant	Garden Center	Flower Shop	Playground	Pharmacy	Food & Drink Shop
3	Postástelep	Bus Stop	Grocery Store	Bakery	Eastern European Restaurant	Train Station	Dessert Shop	Fruit & Vegetable Store	Forest	Food & Drink Shop	Food
4	Tisztviselőtelep	Stadium	Train Station	Restaurant	Food	Pool	Fruit & Vegetable Store	Bakery	Baseball Field	Dessert Shop	Forest
5	Tusculanum	Fruit & Vegetable Store	Grocery Store	Plaza	Park	Gym Pool	Gym	Restaurant	Bakery	Flower Shop	Food
6	Vincellér	Bus Stop	Greek Restaurant	Grocery Store	Sports Club	Eastern European Restaurant	Pharmacy	Farm	Dessert Shop	Discount Store	Electronics Store
7	Érdliget	Pub	Dessert Shop	Bus Stop	Train Station	Supermarket	Historic Site	Ice Cream Shop	Pizza Place	Carpet Store	Chinese Restaurant
8	Ófalu	Baseball Field	River	Farmers Market	Motel	Train Station	Cupcake Shop	Forest	Food & Drink Shop	Food	Flower Shop
9	Újfalu	Clothing Store	Gym	Shoe Store	Chinese Restaurant	Cosmetics Shop	Fast Food Restaurant	Grocery Store	Pharmacy	Supermarket	Bar
10	Újtelep	Stadium	Food & Drink Shop	Food	Farm	Train Station	Cosmetics Shop	Fruit & Vegetable Store	Forest	Flower Shop	Fast Food Restaurant

I searched for the 10 most common venues in each neighborhood. Among the most common venues we can find transport hubs, social venues as well as stores and eating places.

Clustering



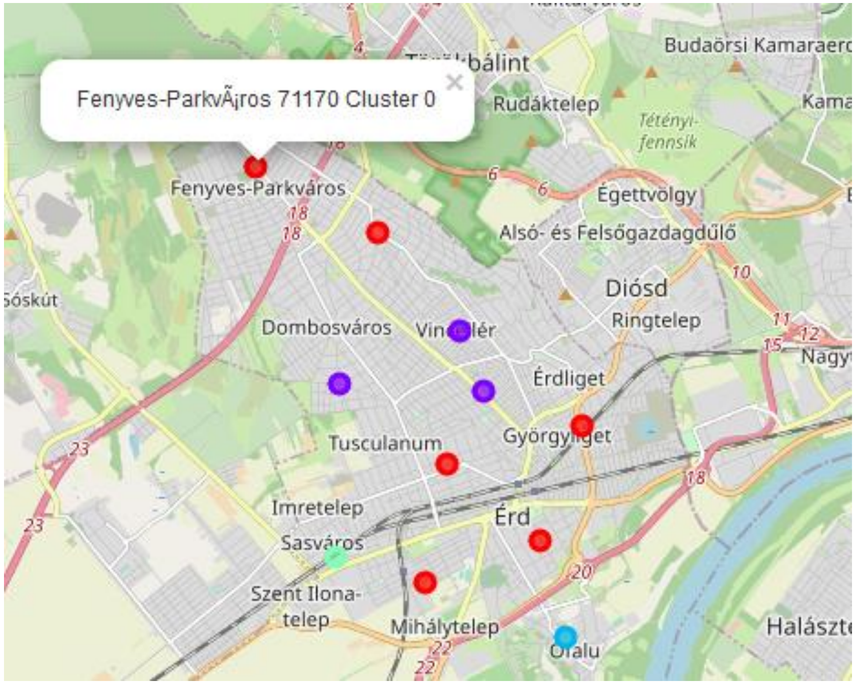
To cluster the town I used unsupervised machine learning technique called K-means algorithm. First I tried to find the best k velue for clustering with elbow method.

There is no sharp shift in the line, but there are two breakings at 2 and 4 K value. 2 is two small value for clustering, so I chose the value 4 for K-means clustering.

Result

The markers shown that the town is divided into well-separeted groups (cluster):

- Cluster 0 (red) = mainly supermarkets, shops and restaurants
- Cluster 1 (blue) = mainly traffic stations and social venues
- Cluster 2 (light blue) = sport field and nature
- Cluster 3 (green) = small shops



Discussion

- I examined the average housing sales prices in a small town in Hungary, Érd and I tried to find connection between the prices and the venues located in the districts and its neighborhoods.
- I used linear regression to find correlation between prices and the number of venues in each boroughs, but the ananlyzis showed the opposite: the number of venues does not influence the price. It is an interesting task to figure out the reasons why people move to Érd in such a big number, but I have no enough data and information to that, and it is out of scope of this study.
- I used the K-means algorithm for clustering the districts. When I tested the Elbow method, I set the optimum k value to 4. All the 11 district coordinates were used in the sample. For more detailed and accurate result, the data set and the radius for searching venues can be expanded, maybe bigger k value can be used.
- Finally I ended my study by visualizing the price and clustering information on the map.

Conclusion

The population of Érd grows year by year, at the same time the number of stores, restaurants and supermarkets also increase. In the last 4 years the housing sales price has increased by almost 50%.

To understand the real causes of this increasing is important for the entrepreneurs and the potential habitants, to find the right place for a new home or a new successful business. This kind of data analyzis can help to bring the supplier and the customer together.