IMPACT OF APARTMENT'S STOREY ON IT'S PRICE, PREDICTION OF THE PRICE FOR APARTMENT & ANALYSIS OF VENUES IN THE BOROUGHS OF THE CITY NUR-SULTAN

by Asset Alkhanov

June 5 2020

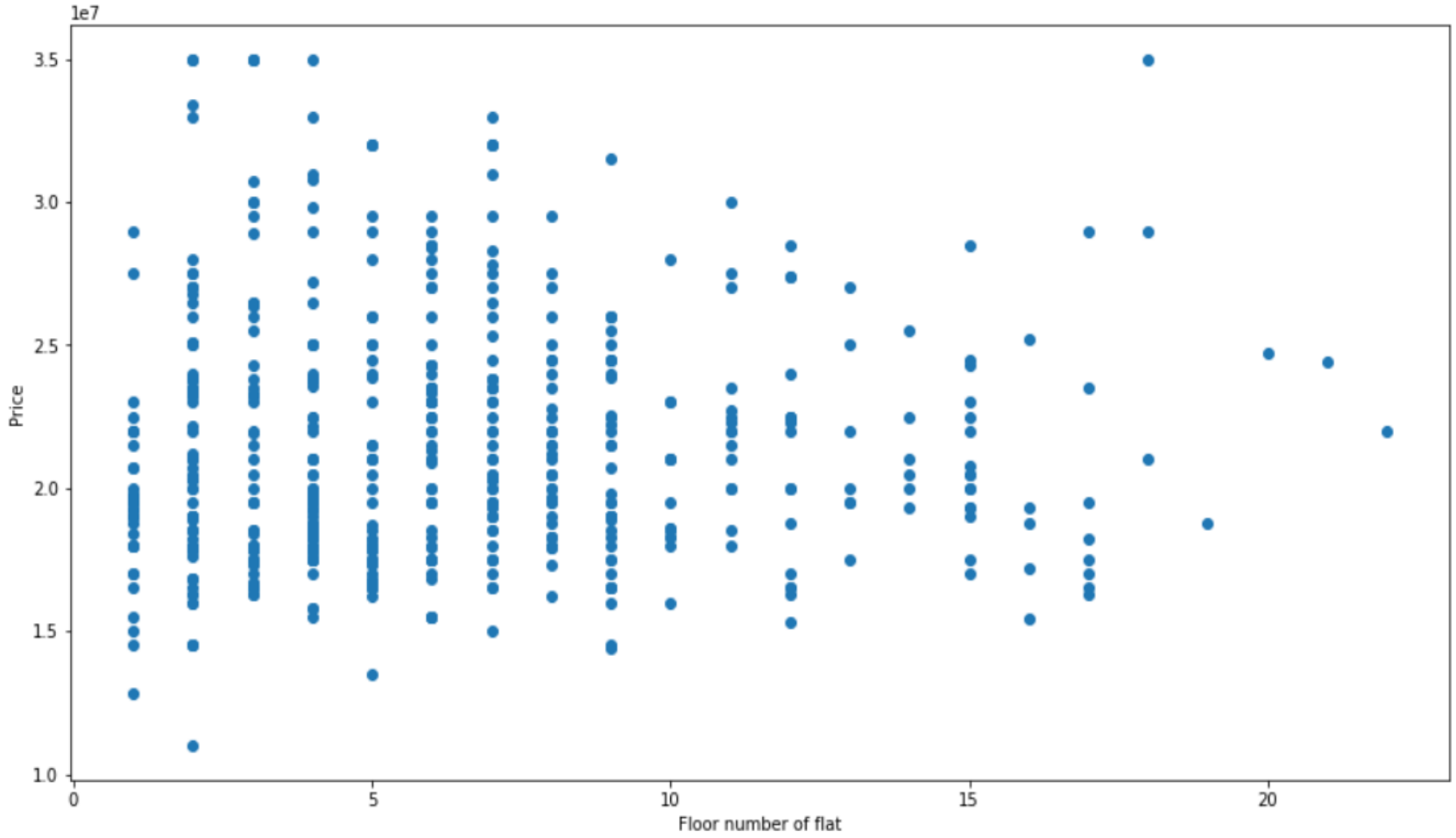
Predicting of the Price for Apartment

- Usually flat prices are predicted by area in square meter, borough, building built year.
- This research reveals how apartment`s storey also affects on the price of the apartment.
- People who work in real estate market are interested in this research as well.
- The study also shows number of venues within a radius of 3000 meters from the center of a particular borough of the city.

Methodology. After scraping and cleaning the data, the following dataframe is created

	Title	Borough	Address	Price	Built year	Sqrms	Flat Floor	Building max floor
0	2-комнатная квартира	Алматы р-н	Нажимеденова	27400000	2018	60.0	12	13
1	2-комнатная квартира	Есиль р-н	Е16 2	19000000	2016	58.0	9	9
2	2-комнатная квартира	Есиль р-н	Керей-жанибек хана 9	19000000	2006	60.0	2	9
3	2-комнатная квартира	Есиль р-н	Мәңгілік Ел 48	27000000	2015	61.0	8	8
4	2-комнатная квартира	Есиль р-н	Акмешит 7 — Ханов Керея и Жанибека	22500000	2010	59.0	1	9
...	...	...	...	...	...	...	...	...
598	2-комнатная квартира	NA	Бараева — Иманбаевой	25000000	2002	59.0	3	4
599	2-комнатная квартира	Есиль р-н	Керей и Жанибек хандар 9	21000000	2008	58.9	5	9
600	2-комнатная квартира	Есиль р-н	К. Мухамедханова 12	20500000	2018	59.0	4	10
601	2-комнатная квартира	Алматы р-н	Кенена Азербаева 12	18000000	2017	58.0	4	5
602	2-комнатная квартира	Сарыарка р-н	Акан серы 16 — Тлендиева	16500000	2016	59.0	9	13

Plotting relationship between 'Price' and 'Flat's floor number'



Application of Linear and Polynomial models shows that neither Linear nor Polynomial models fit to our data set. Data fits only 3% to the Linear, and 1% to the Polynomial approximately.

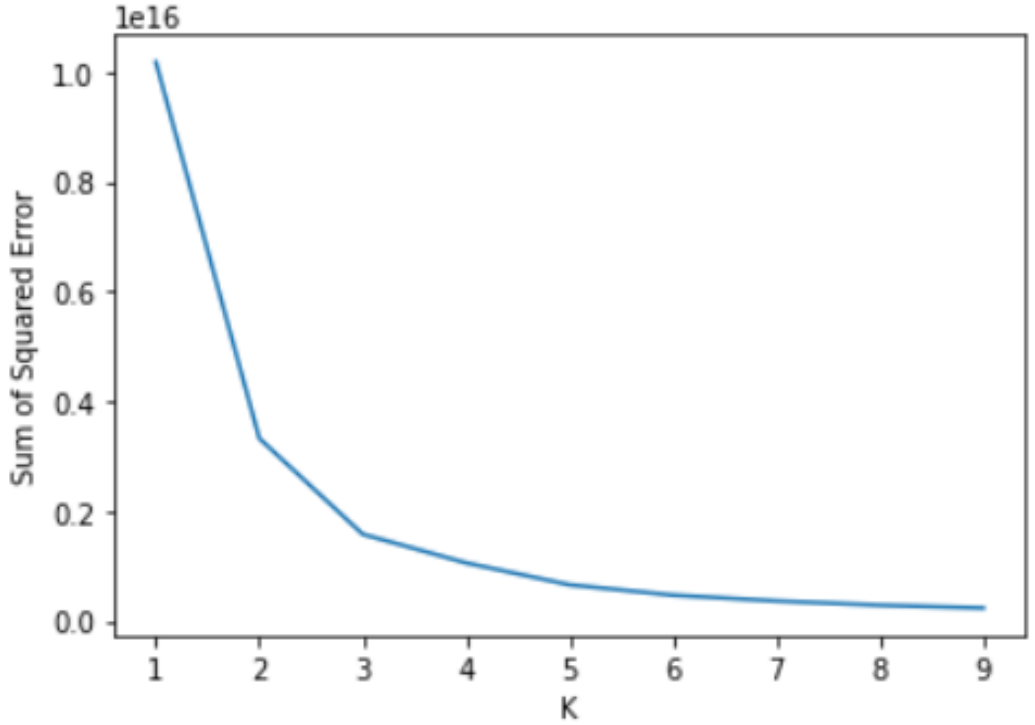
Linear Regression fitting %: 0.028512105701604187

Polynomial Regression fitting %: 0.007722627071624188 5

Predicting a price for a new flat, applying Multiple regression shows rather reasonable result

- To test the prediction model, I passed the following values, 'Flat floor' = 6, 'Borough ID' = 2 (Saryarka), 'Square meters' = 61 and 'Building built year' = 2016.
- **Predicted price of a flat was: 21092984.48722668**

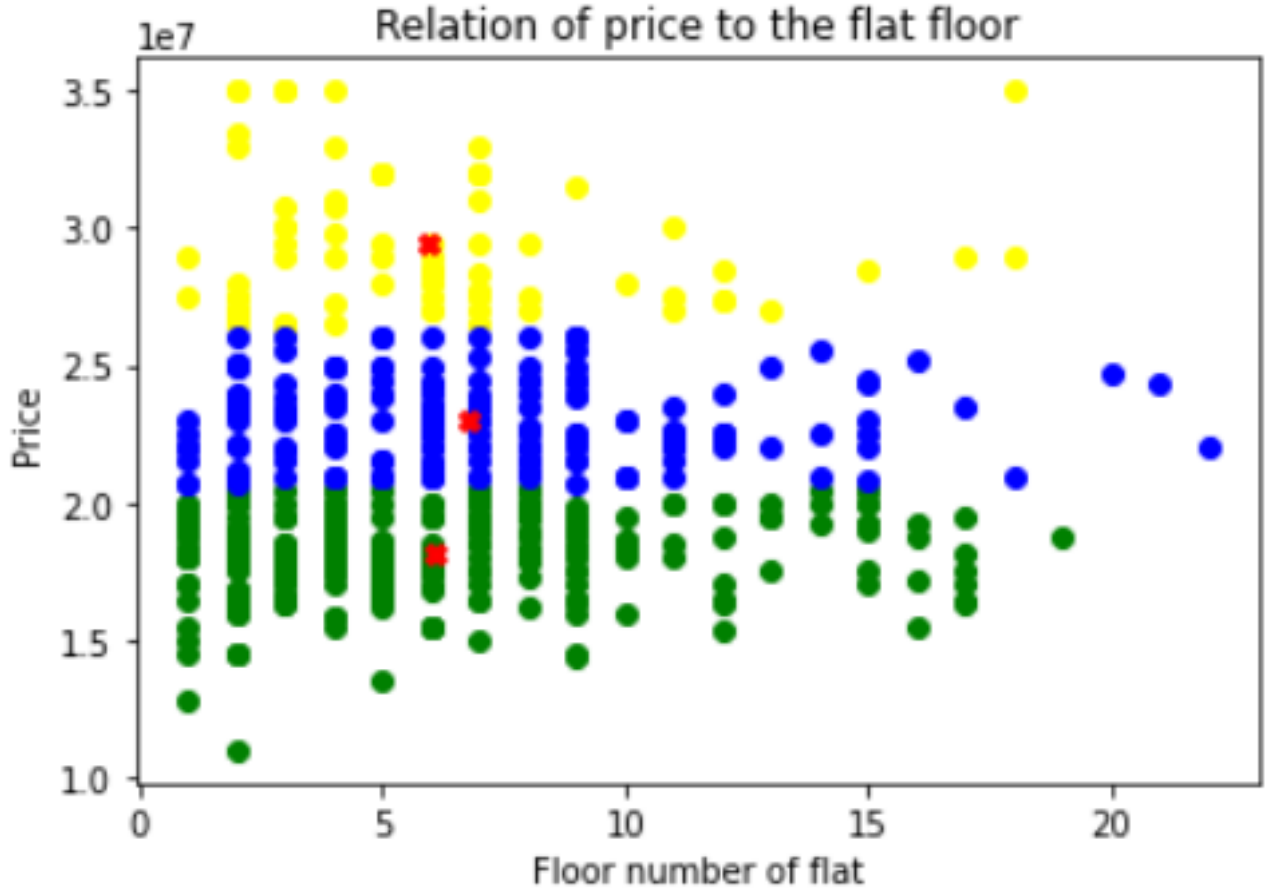
Application of 'Elbow' method to figure out number of clusters for KMeans clustering



The plot shows that 2 or 3 clusters should be applied in the scatter plot.

Results. Plotting three clusters and three red centroids of the clusters on the Scatter plot

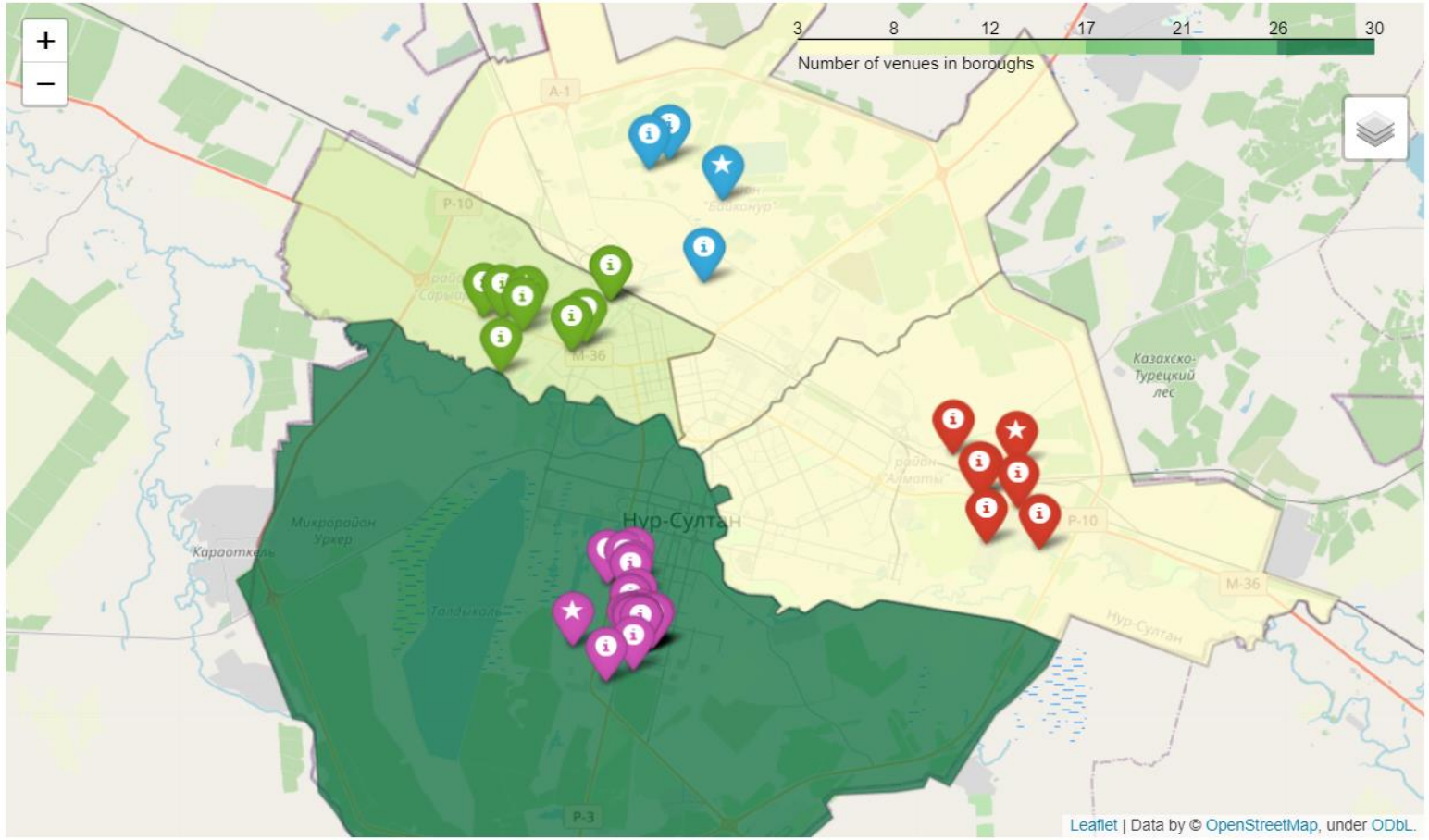
The Scatter plot shows that the most expensive ten flats are located on the storeys lower than 5.



Retrieving location data using Foursquare API and creating a dataframe with borough and venue names

	Borough	Venues
0	Almaty	6
1	Saryarka	8
2	Baikonyr	3
3	Esil	30

Creating a Choropleth map from borough and venue coordinates using Folium library



All the venues are located in radius of 3000 meters from the center of the boroughs. The purple markers are venues located in the 'Esil' borough, green markers are venues in the 'Saryarka', red markers are venues in the 'Almaty' and blue ones are in 'Baikonyr ' boroughs. The markers with a star icon are centers of the boroughs.

Conclusion

- My analysis covered relationship between price of the flat and floor number of the flat. In my analysis I collected data on flats with very similar characteristics, since my main aim was to find the flat floor numbers of the top most expensive flats in the city. The analysis revealed that the 90 % of the top ten the most expensive flats are located below 5th floor. Also I retrieved location data for each borough, in order to see if number of venues has an impact on the price of the flat. The hypothesis was true only for the biggest borough of the city, which is 'Esil'. The other boroughs have numbers close to each other, and they all are below 10. I also used predicting models, to predict prices for a new flats, the model that was applied in my research was a Multiple regression.