The Battle of the Neighborhood
Coursera Capstone Project

Yannik Moll

Introduction:

In Cologne the rents are very high and it is difficult to know if the rent is reasonable or too high when looking for an apartment.
So with this Project we can help people to see:

1. If the rent demanded by the owner is too high for the spesific District
2. Which district is the cheapest district
3. which area is the best regarding the venues

This project helps people who want to move to Cologne or people who already live in Cologne but are looking for a new apartment.

Data Section:

The data on apartments (size, number of rooms, address, and price) is collected by scraping a local website with apartment listings (Immoscout24.de).

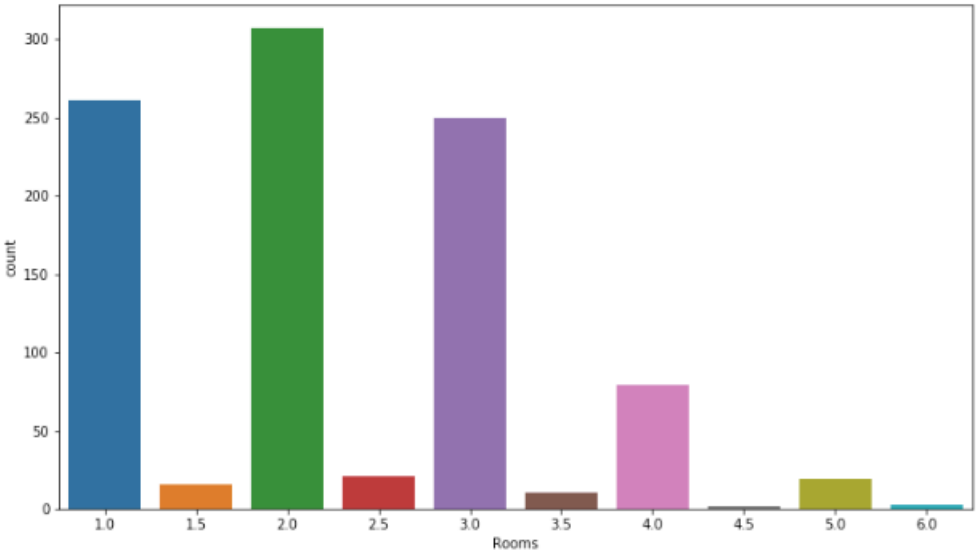
ID	Haus/Wohnung	ID	Miete/Kauf	address	balcony	builtInKitchen	city	floorplan	from	garden	...	livingSpace	lon	numberO
117826677	Wohnung	117826677	Miete	Braunfelsweg 1, Seeberg, Köln	true	false	Köln	false	001.685264	false	...	53	6.90357	
47462593	Wohnung	47462593	Miete	Im Falkenhorst 12, Urbach, Köln	true	false	Köln	true	001.3187698	false	...	73.66	7.08184	
97006573	Wohnung	97006573	Miete	Trenkebergerstr. 15, Meschenich, Köln	false	false	Köln	false	001.528711	false	...	27.52	6.92922	
117096394	Wohnung	117096394	Miete	Trenkebergerstr. 15, Meschenich, Köln	false	false	Köln	false	001.528711	false	...	42.86	6.92922	
83540344	Wohnung	83540344	Miete	Alte Fischenicher Straße 31, Meschenich, Köln	true	false	Köln	false	001.473881	true	...	75	6.92389	

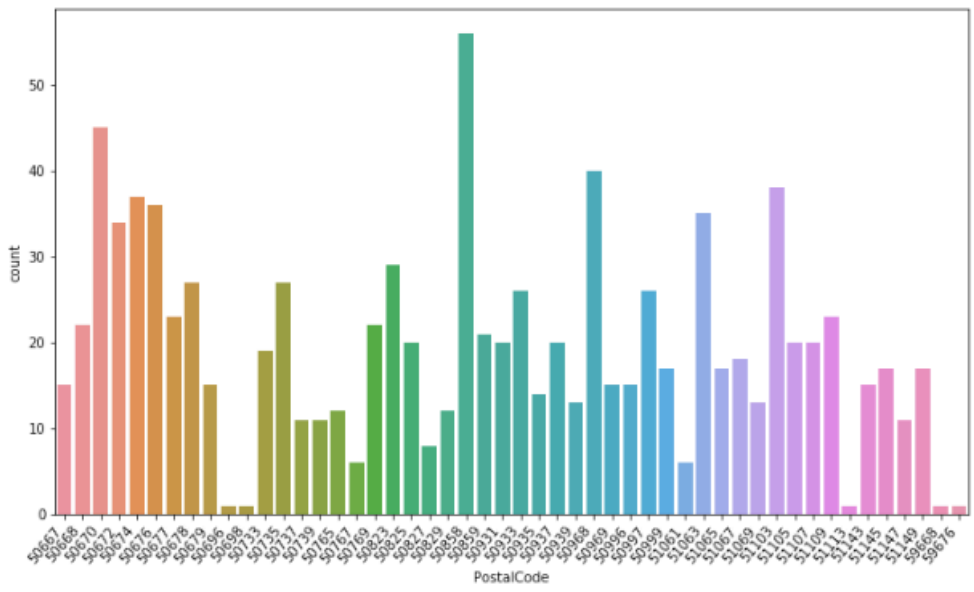
For further analyzing we cleaned the data. We removed all not necessary columns and renamed them. Furthermore, we checked if there are any NAN values and changed the datatype so we can process the data later. This is the final data frame after cleaning:

	PostalCode	District	Size	Rooms	Price	Price/m2
0	50769	Seeberg	53	2	329.16	6.21057
1	51145	Urbach	73.66	3	494.25	6.70988
2	50997	Meschenich	27.52	1	198.14	7.19985
3	50997	Meschenich	42.86	2	308.59	7.19995
4	50997	Meschenich	75	3	560	7.46667

After this we removed all outliers especially luxuries apartments where excluded.

So, we can create the following charts.



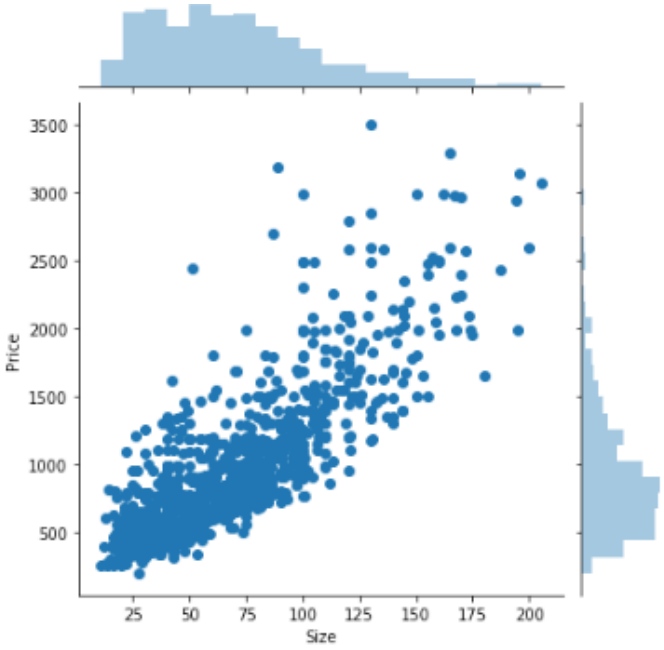


Count apartments sorted by rooms and count of apartments sorted by postcode

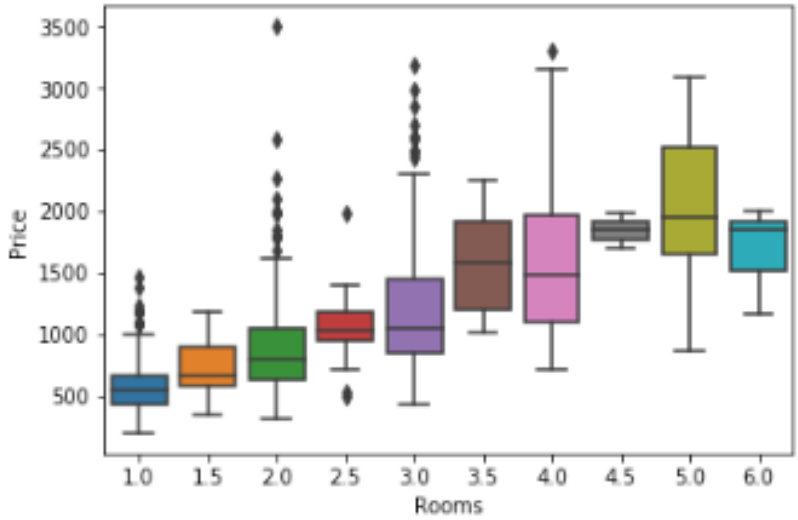
Here you can see that 1, 2 and 3 room apartments are the most commonly ones.

Methodology:

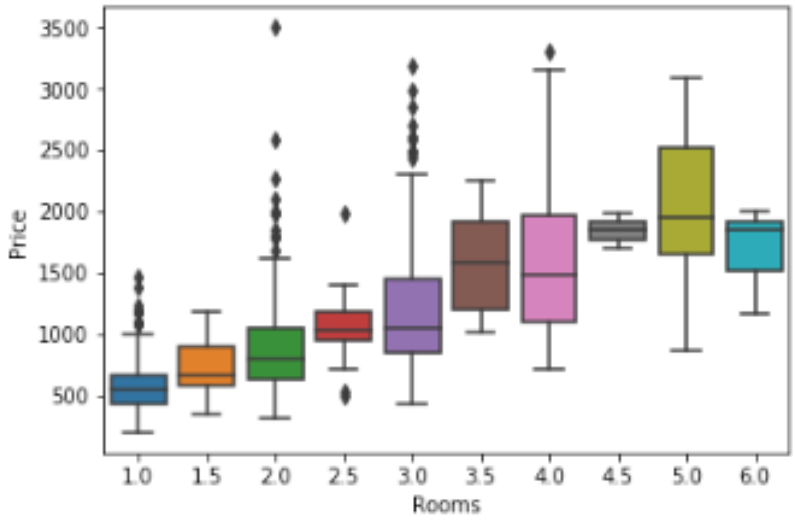
We can also see a strong correlation between price and apartment size:



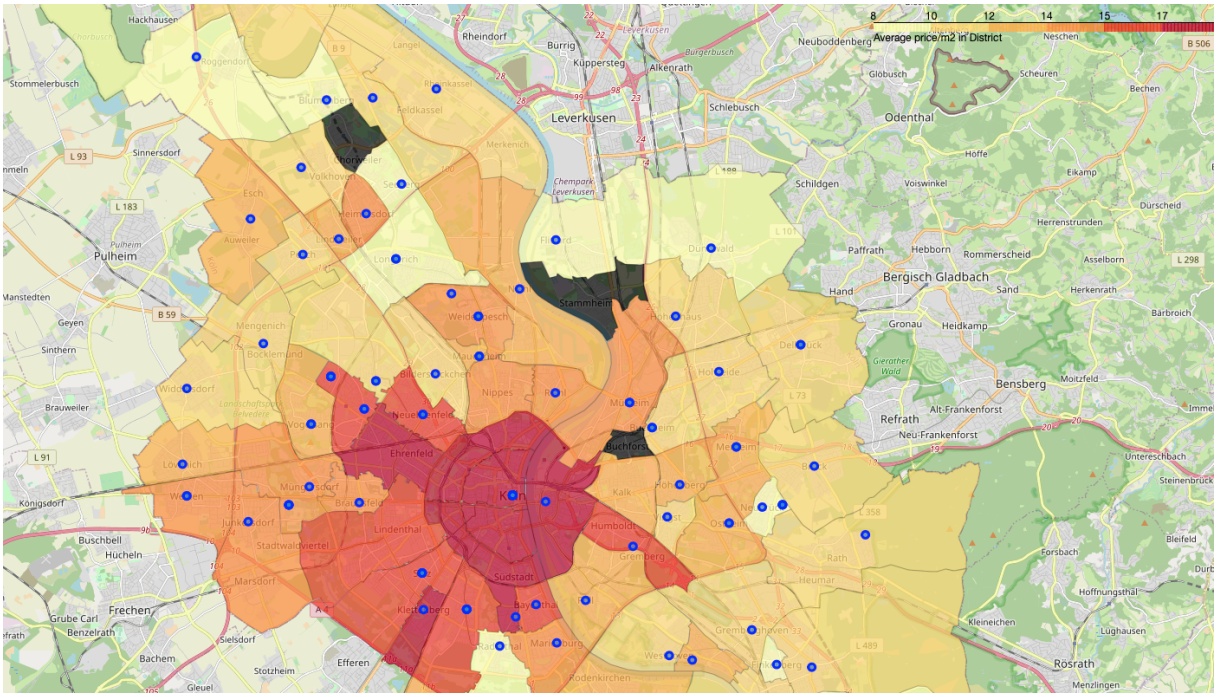
As expected, the prices for the apartments are increasing as the number of rooms is increasing. Interesting to see, is that the price for the 6 room apartments are decreasing.



In this chart we can see, that all apartments are approx. in the same range except the 1 and 1.5 room apartments. These are far more expensive than the others

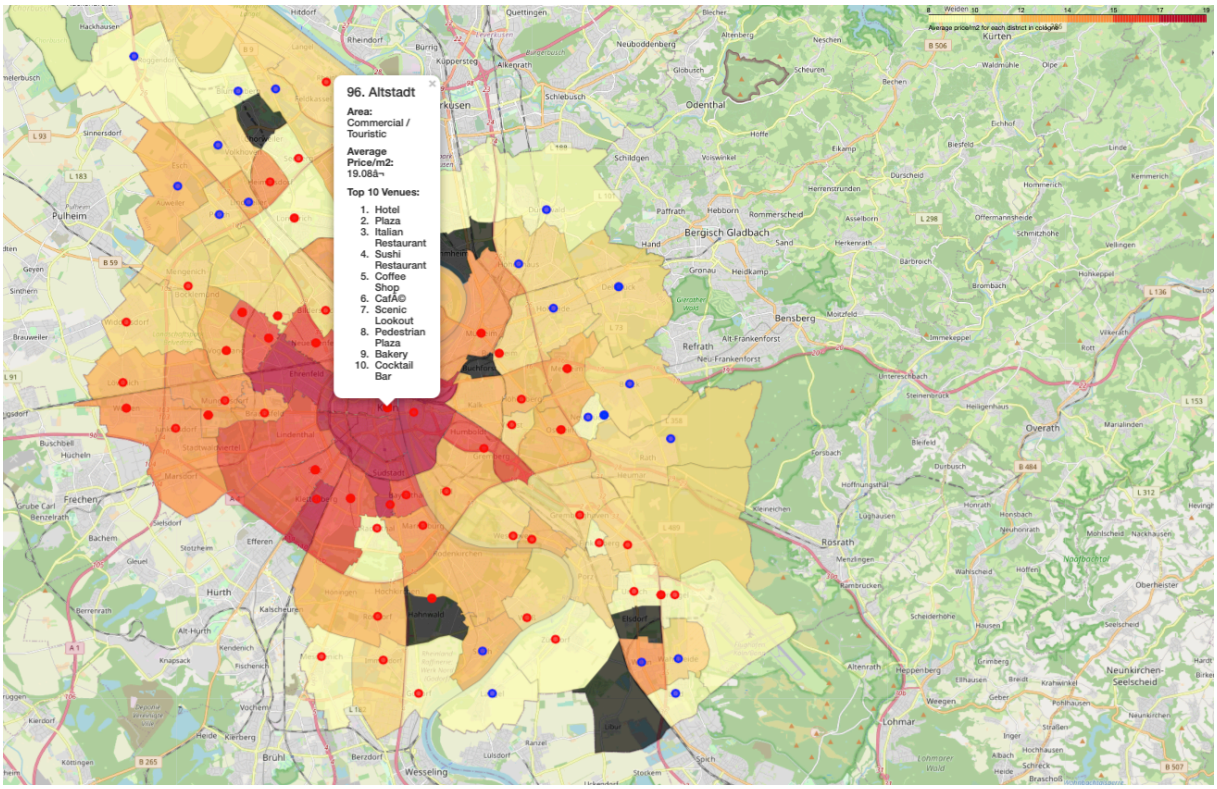


Using the cleaned apartment data and the district coordinates from the json file we can visualize the average price/m2 for each district in cologne.



Results:

After the collection of the foursquare data we run the k-mean clustering to cluster the venues in two categories. So, we have a commercial and a more touristic cluster (Red: touristic, blue: commercial). Combined with the average price/m2 we created a new map. In this map you can see which is a cheap or expensive district. He markers give further data on ech district such as the average price/m2 and the top10 venues for each district.



Discussion & Conclusion:

With this map, one could determine for example that the district "Altstadt" is the most expensive district to live in, however by clustering we determined that there are several more similar districts where the price/m2 is significantly lower. Therefore, if someone wants to rent an apartment but cannot afford to live in the district " Altstadt", they could look for apartments in Longerich or Raderthal which is similar in venues but has much lower price for renting apartments.¶

It should be noted that this analysis was carried out on the basis of random sample results (results of one Immoscout.de query). This means that the results are critical and should be observed over a longer period of time in order to achieve a realistic result in terms of the price.