Mohamed Ibrahim

Project Results

## Introduction

I have analyzed “Real estate and housing in kazakhstan”by gathering Data from krisha.kz as the main website of my project. I have already learned how to parse this site, so it will be a big plus for me.

## Krisha.kz

Krisha is the main website of my project. Krisha is the main website that residents of Kazakhstan use for renting and selling Real estate.

## The main idea of the project and the importance of analyzing real estate

This analysis will help you understand the current housing market, how much properties similar to yours are worth. The information gathered through an analysis of the main website in **Krisha.kz** about real estate and housing. It helps the seller choose a price for their house and helps buyers see if the asking price is too high, low or reasonable. By comparing similar properties on the market, you will be able to accurately put a price on a home.

It also helps to compare prices in different cities or in different districts in one city. In order to give an answer for these kinds of demands,I wrote main questions to analyze using krisha.kz.

## Main questions:

1. The districts with the most expensive and cheapest housing on average?
2. The most expensive/cheapest apartments for each type of apartment.
3. The largest/smallest apartments for each type of apartment.

## Analyzing and visualizing the parsed data

In this section I will explain how I analyzed the obtained data and get informative and interesting results.

After parsing the website, I got .csv files with 5285 records from Almaty and 3734 records from Astana, and about a thousand records from each other's districts. In general I have focused on housing in Almaty.

From the results we can get all apartments for rent in Almaty and we have features like type (number of rooms), price, area, address, district. Here we have 5285 records and 6 columns.

| city | name | price | area | address | distinct |
|---|---|---|---|---|---|
| almaty | 2-комнатная квартира | 130 000 ₸ | 60 м² | Сатпаева | Бостандыкский р-н |
| almaty | 2-комнатная квартира | 270 000 ₸ | 62 м² | Самал-1 29 | NaN |
| almaty | 3-комнатная квартира | 380 000 ₸ | 180 м² | Луганского — Сатпаева | Медеуский р-н |
| almaty | 2-комнатная квартира | 125 000 ₸ | 42 м² | мкр Аксай-2 11 | Ауэзовский р-н |
| almaty | 2-комнатная квартира | 190 000 ₸ | 53 м² | мкр Казахфильм | Бостандыкский р-н |
| almaty | 3-комнатная квартира | 600 000 ₸ | 110 м² помесячно | Аль-Фараби 5к3А — Козыбаева | Бостандыкский р-н |
| almaty | 3-комнатная квартира | 320 000 ₸ | 90 м² помесячно | Аскарова Асанбая 21/20 | Бостандыкский р-н |

Firstly, I checked the whole data and found some outliers like rare cases and wrong datas. Also, we have some columns which we will use as numbers, so we must convert them first. So, I decided to clean the data. I deleted some rows and converted columns like price and area to int and float type.I used a new dataframe to get correct results. After preprocessing the data, I had 7 types of apartments (from 1 room until 7 rooms) from each of the 8 districts of Almaty.

I used several types of libraries in Python to analyze and visualize the results. Firstly, I used groupby() for dataframe, to group the data by district and found the number of apartments in each district.

| | distinct | price |
|---|---|---|
| 3 | Бостандыкский р-н | 1656 |
| 5 | Медеуский р-н | 1281 |
| 1 | Алмалинский р-н | 935 |
| 2 | Ауэзовский р-н | 557 |
| 0 | Алатауский р-н | 165 |
| 7 | Турксибский р-н | 116 |
| 4 | Жетысуский р-н | 107 |
| 6 | Наурызбайский р-н | 94 |

In this table you can see the number of apartments for rent in each of the 8 districts of Almaty. Here we can say that in districts like Bostandyk(1656), Medeu(1281) and Almaly(935) which are central areas in Almaty we have many apartments for rent(shown in brackets). It means that in these areas we can easily find the apartments that we want and we will have many choices. It is also good for renters to analyze the differences in each district. For example, if they have an apartment in districts like Bostandyk, Medeu or Almaly, they can quickly rent out, because in these areas there is a lot of demand.

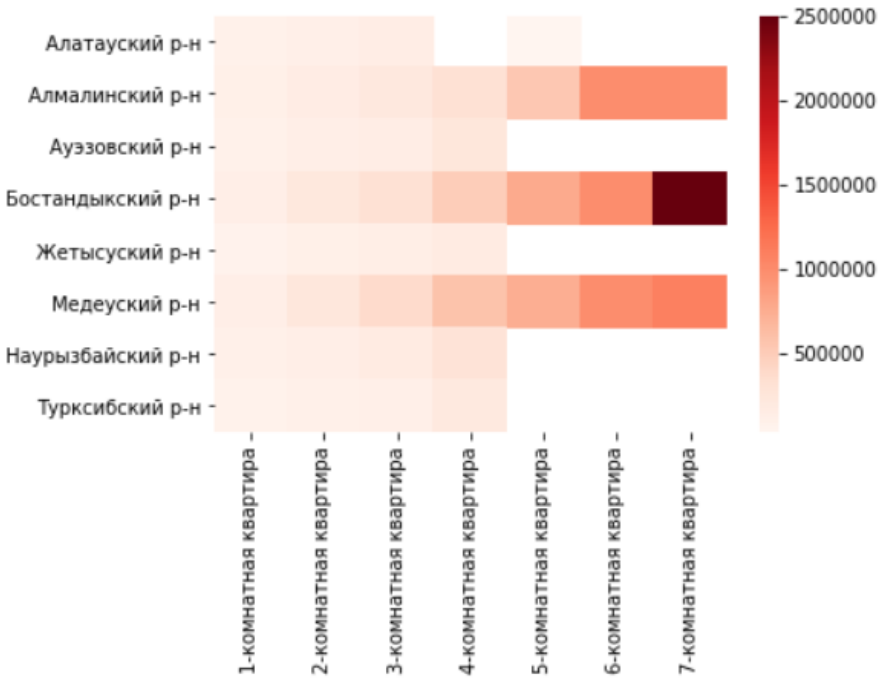

In the pie chart above we can easily define the percentage of each type of apartment. Diagram shows that there are many apartments with 2 rooms(39.7%).

Afterwards, I made a table for the average price per apartment type in each district. After using the pivot_table() function of the pandas library I changed the table to pivot table which is very comfortable to use. For example, in the table below you can see that the average price for 1-room apartments in district Alatau is 90,000 tg.

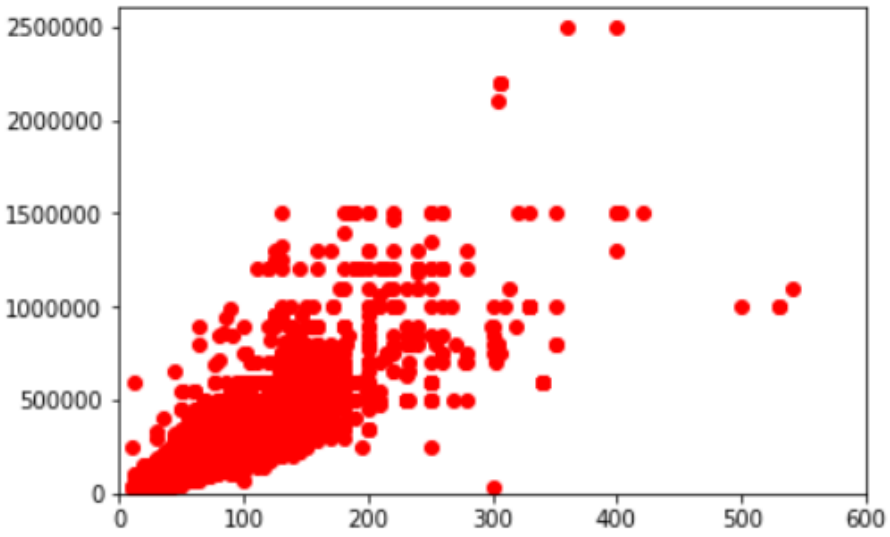
From this table we notice that the number of apartments with 5-7 rooms is small and even in some areas we can not find these kinds of apartments. It means that there is a lot of demand for apartments with 1-4 rooms, and if you want to sell or rent your apartment, it is desirable to have these kinds of apartments.

| name | 1-комнатная квартира | 2-комнатная квартира | 3-комнатная квартира | 4-комнатная квартира | 5-комнатная квартира | 6-комнатная квартира | 7-комнатная квартира |
|---|---|---|---|---|---|---|---|
| distinct | | | | | | | |
| Алатауский р-н | 90000.0 | 120000.0 | 150000.0 | NaN | 30000.0 | NaN | NaN |
| Алмалинский р-н | 115000.0 | 160000.0 | 220000.0 | 320000.0 | 550000.0 | 1000000.0 | 1000000.0 |
| Ауэзовский р-н | 95000.0 | 130000.0 | 150000.0 | 250000.0 | NaN | NaN | NaN |
| Бостандыкский р-н | 130000.0 | 230000.0 | 320000.0 | 500000.0 | 775000.0 | 1000000.0 | 2500000.0 |
| Жетысуский р-н | 75000.0 | 110000.0 | 140000.0 | 180000.0 | NaN | NaN | NaN |
| Медеуский р-н | 130000.0 | 240000.0 | 380000.0 | 580000.0 | 750000.0 | 1000000.0 | 1100000.0 |
| Наурызбайский р-н | 100000.0 | 130000.0 | 175000.0 | 300000.0 | NaN | NaN | NaN |
| Турксибский р-н | 80000.0 | 100000.0 | 120000.0 | 210000.0 | NaN | NaN | NaN |

I made a visualization with a heatmap in the seaborn library. It shows apartments' average price for each type and district. Using this map we can see that the highest prices are usually in Bostandyk and Medeu districts. The lowest prices are in Alatau, Turksib districts.




Afterwards I made a correlation between price and area of the apartment. The result was p = 0.8 with Pearson correlation. It seems that it is a good result, but in each district we have their own average prices and in some districts like Bostandyk or Medeu the price does not depend on the area of the apartment, it will be higher because of the central location. Also I showed this result in the plot using matplotlib.pyplot library.



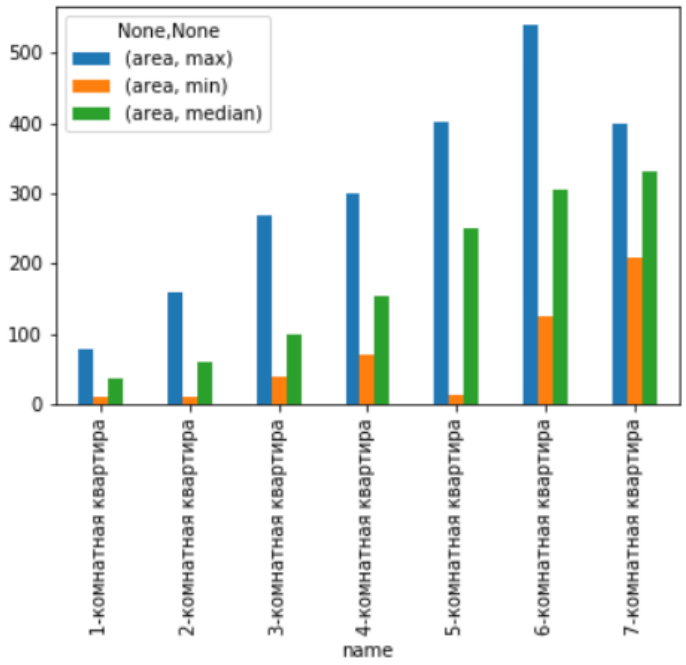

| | price |
|---|---|
| distinct | |
| Туркисибский р-н | 95000 |
| Алатауский р-н | 100000 |
| Жетысуский р-н | 100000 |
| Наурызбайский р-н | 117500 |
| Ауэзовский р-н | 125000 |
| Алмалинский р-н | 160000 |
| Бостандыкский р-н | 250000 |
| Медеуский р-н | 300000 |

The next table shows the average price for each district. So, it shows for all the types of apartments, and in this table we can see that in districts like Turksib(95 000 tg), Alatau(100 000 tg), Zhetysu(100 000 tg) we have the lowest prices and in Bostandyk and Medeu districts we have highest prices.

In this table we have the maximum,minimum and average prices for each type of the apartment. For example, the maximum and minimum price for 1-room apartments are 450 000 and 15 000 tg. Using this table we can see the huge difference between the prices in each district even though the number of rooms are the same. Using this table, we can define the average

| | price | | |
|---|---|---|---|
| | max | min | median |
| name | | | |
| 1-комнатная квартира | 450000 | 15000 | 110000 |
| 2-комнатная квартира | 990000 | 25000 | 180000 |
| 3-комнатная квартира | 1500000 | 40000 | 300000 |
| 4-комнатная квартира | 1500000 | 120000 | 500000 |
| 5-комнатная квартира | 2200000 | 30000 | 750000 |
| 6-комнатная квартира | 4200000 | 300000 | 1000000 |

price for each room, in order to know the value of our own apartments.




Here the smallest and largest and middle apartments with different rooms. Using columns max and min, we notice that there can be huge differences in the area of apartments even if the number of the rooms are the same. Also, using the median of the area we can define our apartment's size and difference between the average.

## Conclusion

The aim of this project was to understand the current housing market in Kazakhstan and find a way to analyze the data gathered by parsing data.I could set research questions and I answered them by using my skills of python.

I learned how it is possible to parse all data from a website using only computer and Python skills. This experience taught me not only the programming language, but also the frameworks such as BeautifulSoup or Selenium. Most likely, this knowledge will be vital or at least useful in the future. By automating browser surfing with Selenium I will be able to accomplish lots of my ideas. This experience could be my base for future projects.

## References

- Pyplot tutorial — Matplotlib 3.1.2 documentation
- pandas documentation — pandas 1.1.4 documentation (pydata.org)
- Seaborn heatmap tutorial (Python Data Visualization) - Like Geeks
- Finding correlation coefficient between columns of a pandas dataframe | Pythontic.com