

Housing market analysis

Frederic Hatanian





SQL

The questions on SQL were pretty straight forward, so there was no major difficulty, I just had to search on google a couple of time but they allowed me to notice a few things:

- The bathrooms column has a one outlier with 33 bathrooms
- A bunch of columns have a lot of 0 values
- It is way easier to remove outliers with SQL than with python

Interesting but when I tried to export my file from MySQL then import it in python, it did not really work

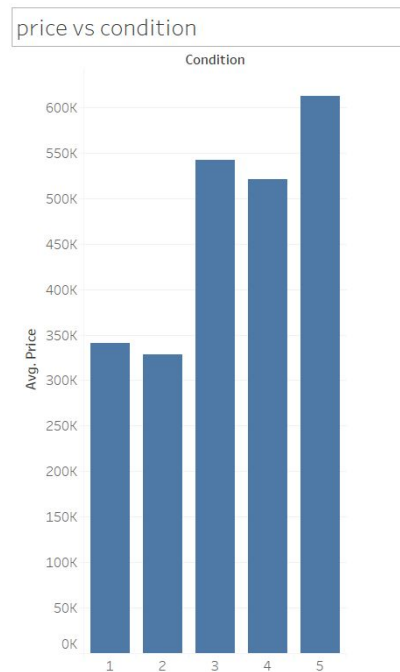


Tableau

That was really interesting to use tableau but I feel like the questions didn't really help to visualize anything particular,

The graph on the right is quite interesting as it shows that there is not much of price difference between the grades 1 and 2 as well as between the grades 3,4 and 5

I would also love it if we could spend more time learning how to create a nice dashboard :)





Model building

- Tried to import my data from MySQL, it did not work so I build my first model without much data cleaning
- Built a linear regression model with an ok R-square of 0.63 but a very high mse and rmse which probably meant that I had a lot of outliers
- Tried removing more of the columns (the ones with lots of 0s) and got a worse R_square
- Tried a log transformation of the price feature and got a slightly better R-square but a way better mse and RMSE