REGRESSION MODEL
ZAMANI REAL ESTATE

Summary of project

- This project is using provided data to come up with a model that predicts houses that have at least the similar variables.

Business Problem

- Zamani Real Estate have been contracted by different property owners to assist them in selling their houses. They have given me past data of houses sold which i will use to create a model that predicts the price of a house that is to be put on sale

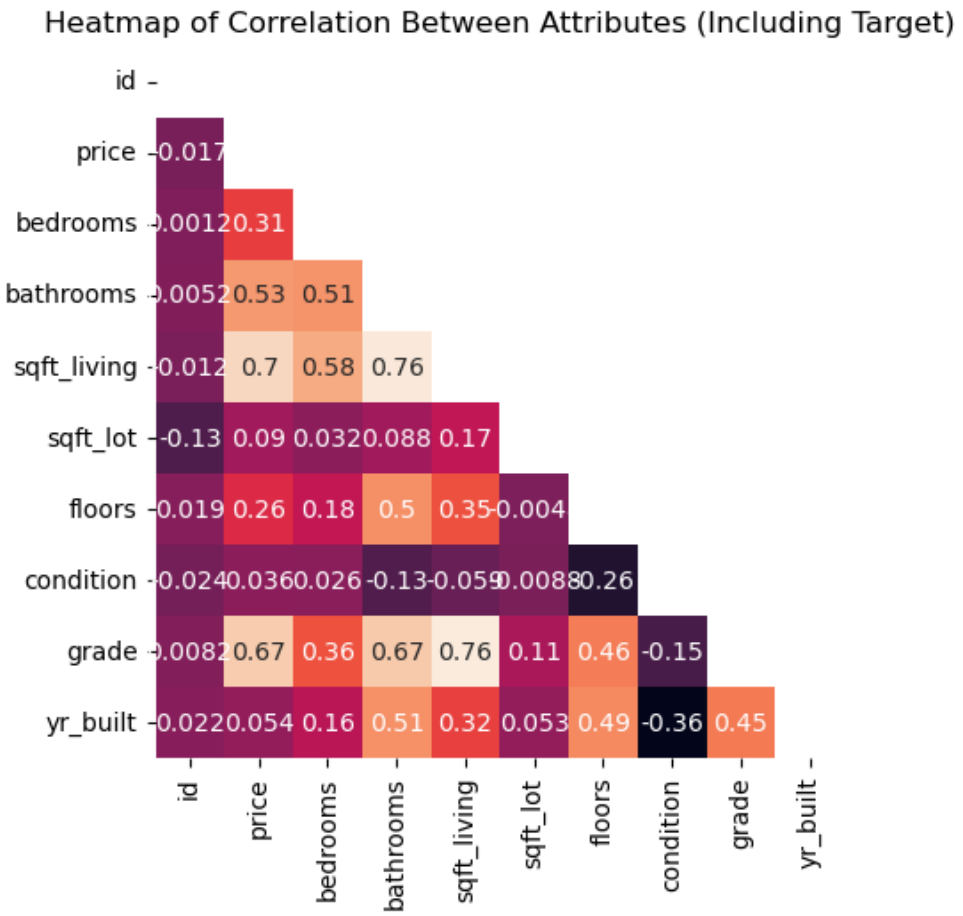
Data

- I picked my data, kc_house_data as I wanted to analyze the problem using profit and other variables that are common in most houses such as bedrooms, number of bathrooms and the size of the living room.
- The variable profit is determined by the other variables which the model would use to predict other house prices.

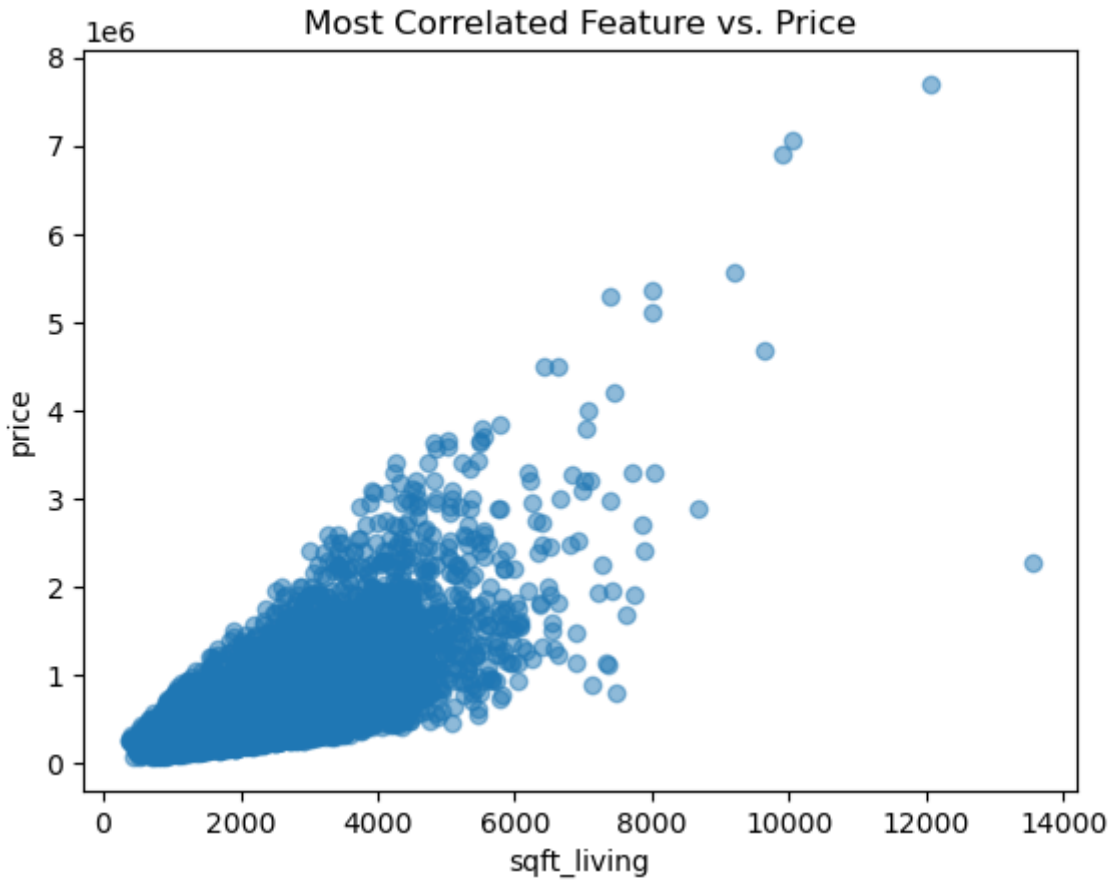
METHODS

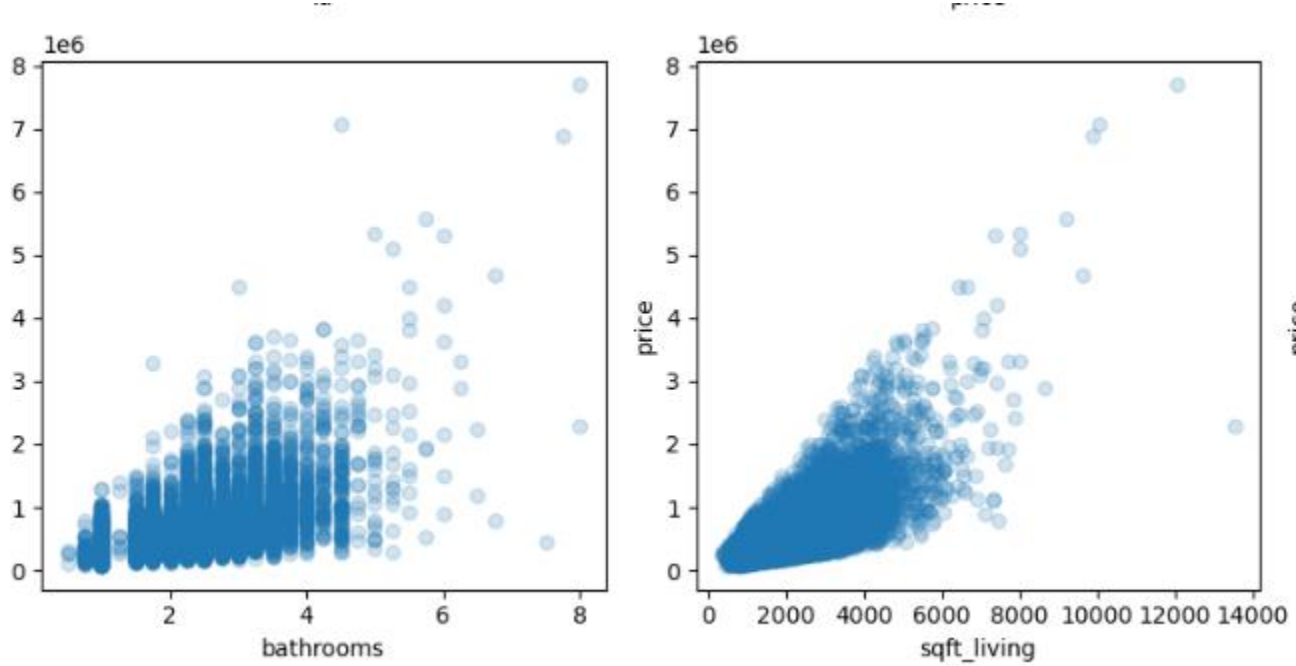
- For me to use my data, I had to to clean my data by removing any unnecessary information.
- For variables that had no data, I filled them accordingly.
- I used standardization/normalization on features of my dataset
- Performed log transformations on different features of my dataset
- I used statsmodels to fit a multiple linear regression model
- Evaluated the linear regression model by using statistical performance metrics pertaining to overall model and specific parameters

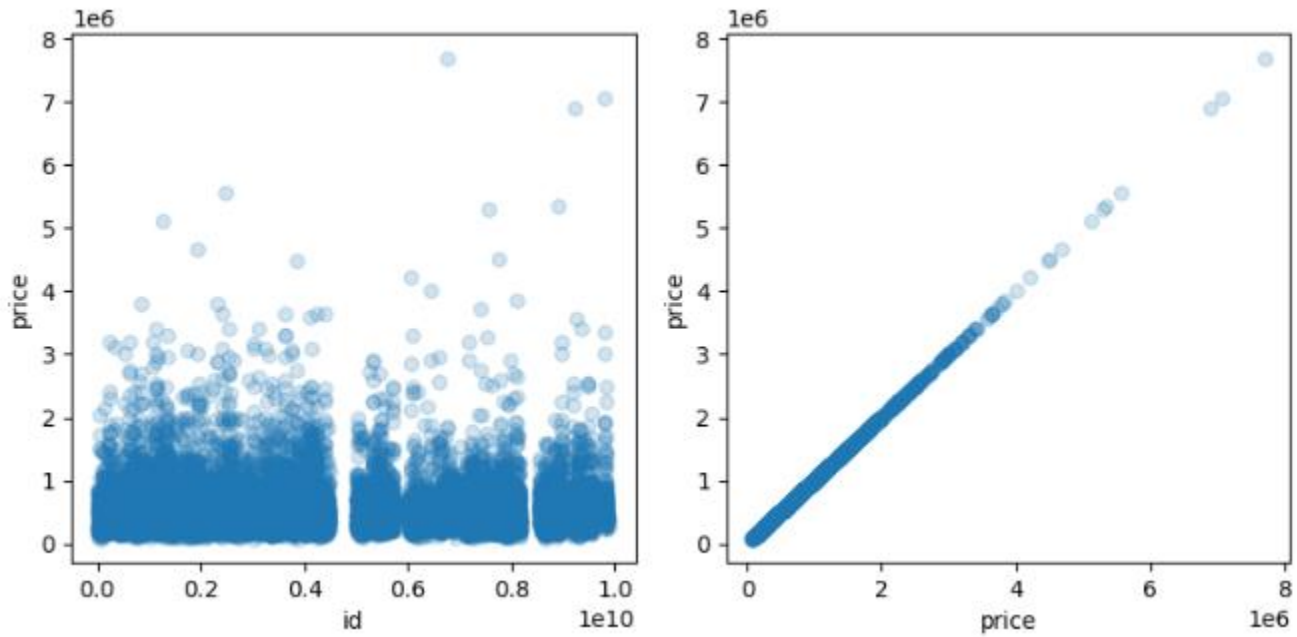
VISUALIZATION RESULTS

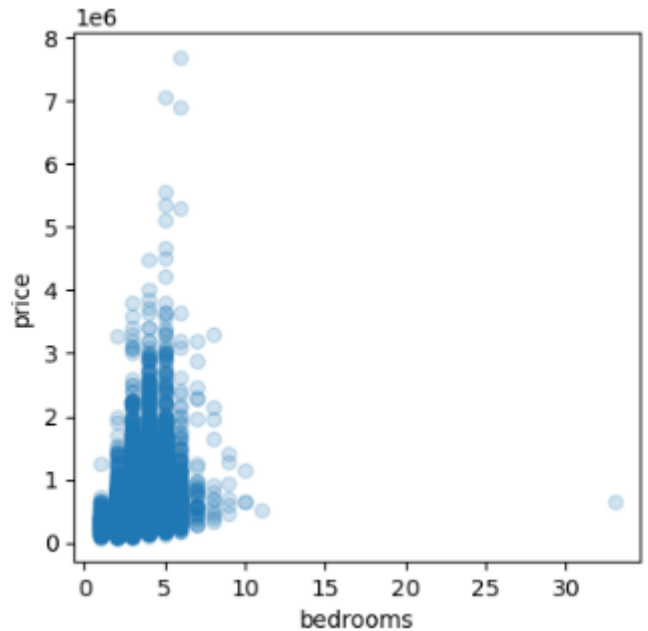


SCATTER PLOTS OF THE VARIABLES vs PRICE

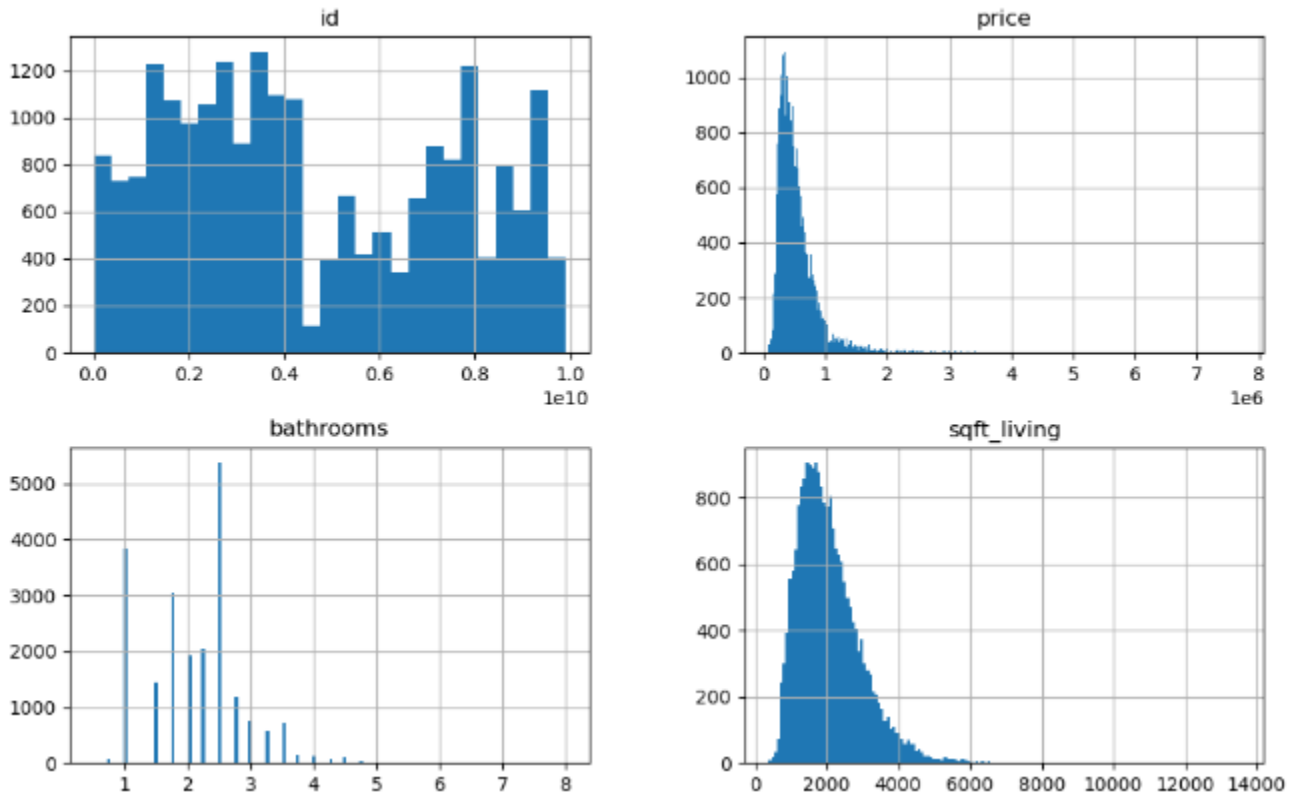



1e6
8
7
6
5
4
3
2
1
0
2
4
6
8
bathrooms
1e6
price
0
2000
4000
6000
8000
10000
12000
14000
sqft_living


1e6
8
7
6
5
4
3
2
1
0
price
0.0
0.2
0.4
0.6
0.8
1.0
1e10
id
1e6
8
7
6
5
4
3
2
1
0
price
0
2
4
6
8
1e6
price


1e6
8
7
6
5
4
3
2
1
0
price
0
5
10
15
20
25
30
bedrooms

Histograms to check distribution of variables



CONCLUSION

- From the data, sqft_living is the most correlated variable with price followed by grade and bathrooms.
- For a developer, they can maximize on the variables while building a new house in order to sell at a higher price.

Thank You!

Email: elizabethbarongo85@email.com

GitHub: @ElizabethOmao

LinkedIn: linkedin.com/in/elizabeth-barongo-98155a19b