



HOUSING DATA REGRESSION MODEL.

OVERVIEW.

- With the advent of technology and availability of vast amounts of data, investors are increasingly turning to quantitative analysis and machine learning techniques to determine the factors that affect the sale price of houses in the real estate industry.
- This project aims to determine the specific factors that increase or decrease the sale price of a housing unit by using a range of specific factors within the data provided. The target partners include; real estate agencies, construction companies, home owners who are interested in selling and potential home owners looking to buy.

BUSINESS UNDERSTANDING.

- Determining which factors have the most significant effect on the final selling price of a house is crucial in making informed decisions about whether to invest in specific aspects such as renovations, expansions or even the housing unit as a whole.
- To address this issue, my aim is to use statistical analysis to identify the specific factors that affect the sale price of houses represented in the King County housing dataset. By doing so, I hope to provide an accurate linear regression model that helps the target partners make informed decisions.

OBJECTIVES.

- My aim is to:
- Perform a statistical analysis of the data provided to determine the specific factors that affect the final selling price of a house.
- Formulate an accurate linear regression model that helps stakeholders make informed decisions before investing in houses.
- Provide recommendations based on the regression model.

DATA UNDERSTANDING.

- The King County dataset contains 21597 rows and 21 columns, out of these 21 columns in the dataset, only 12 columns are useful to the process. The other columns are dropped because some of them contain the same data as the columns retained and others have little to no bearing at all on the price of the houses.
- The other operation performed on the dataset is checking for missing values and duplicated records and dropping them.

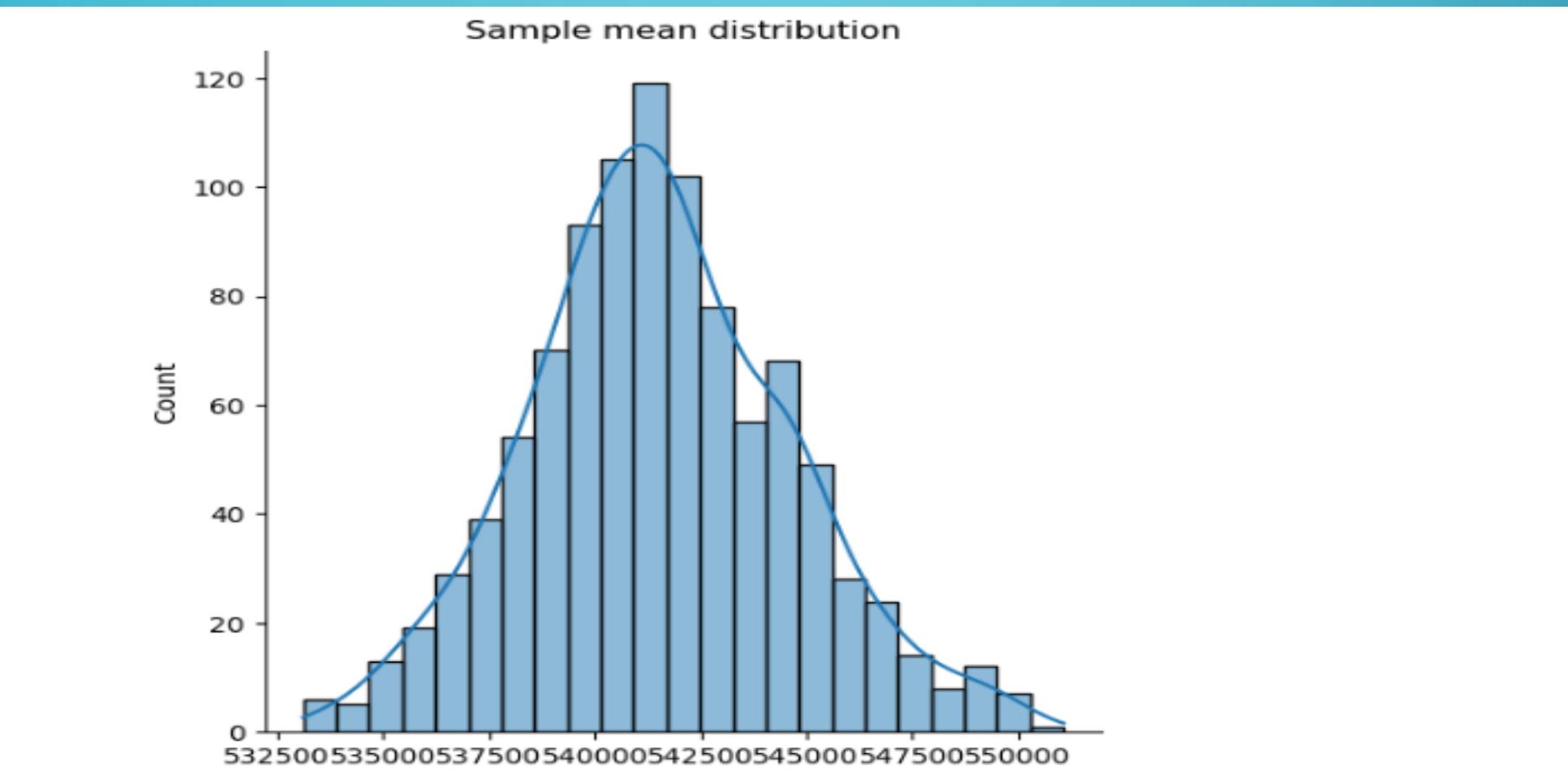
DATA PREPARATION.

- The target variable in this project is the selling price of the houses. Before I came up with a regression model detailing the specific metric that affect my target variable, I first performed multiple statistical analyses on the dataset.
- Here are some of the analyses performed on the price data and the findings.

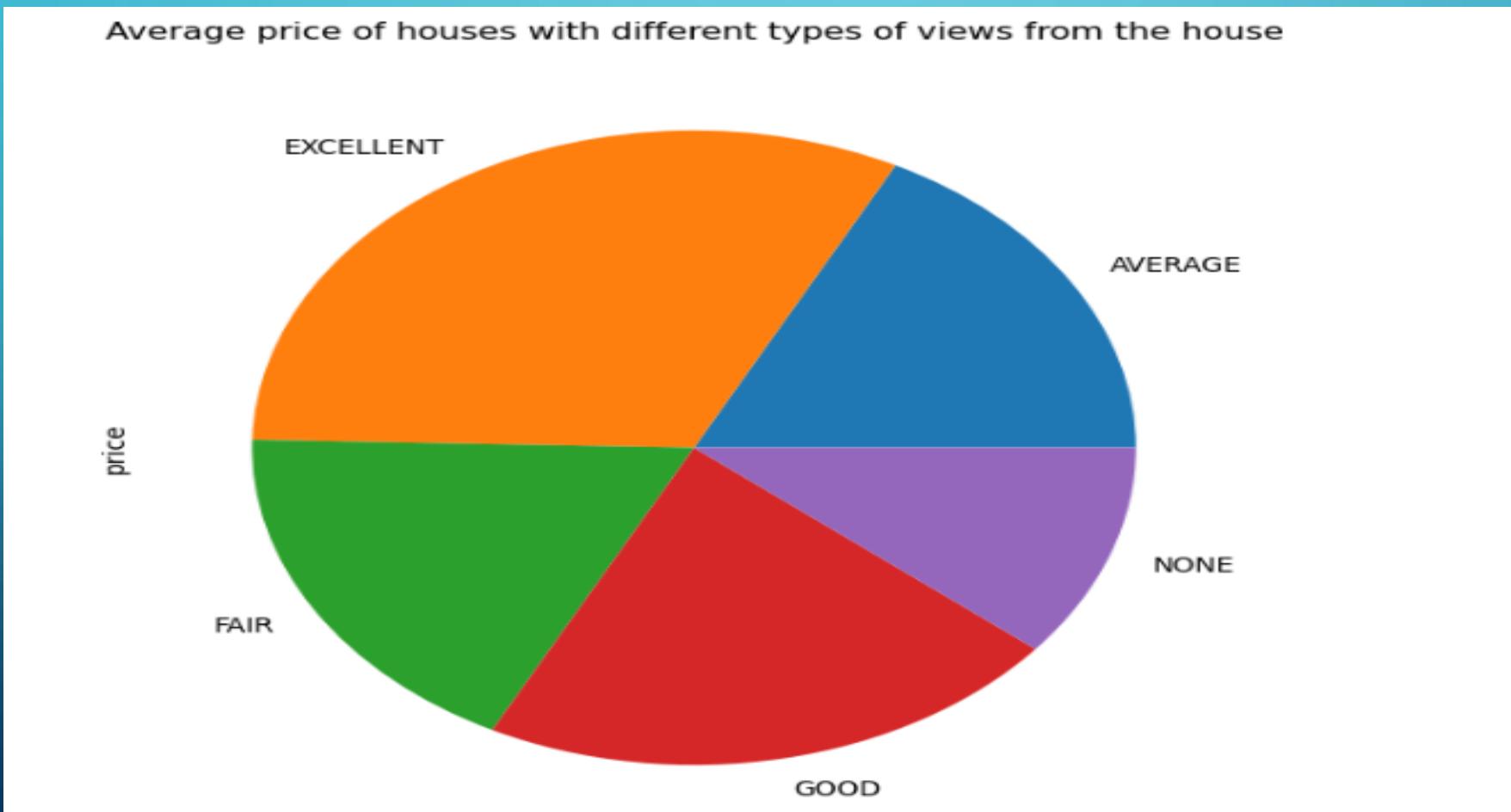
SKEWNESS AND KURTOSIS.

- I measured the skewness and kurtosis of the price data in order to determine the degree of distortion and the length of the tails in the distribution.
- Skewness : 4.08
- Kurtosis : 35.29
- The data on the selling price of the houses had a high positive skew, meaning that there is a long tail in the higher price range. The data also happened to be leptokurtic, meaning that the data had heavy tails and that there were more outliers.

CENTRAL LIMIT THEOREM.



VISUALIZATION OF THE EFFECT OF VIEW TYPE ON THE PRICES.



MODELLING.

- I first created a simple model using the square footage of the living area, and these are the results of the model.
- The model is statistically significant overall and explains about 50% of the variance in price.
- Given that a house has 0 square-feet of living area, the price of the house drops by 48,607 dollars.
- For each increase in 1 square-foot of the living area, we see an associated increase of 283.40 dollars in the price of the house.

FIRST MULTIPLE REGRESSION MODEL RESULTS.

- The model was statistically significant and explained about 57% of the variance in price.
- Given that all other predictors were 0, the price of the house was 6,029,056 dollars.
- When a house had 2 bedrooms as compared to any other number of bedrooms besides 2, 3, 4 and 5, we saw an associated increase of 178,552 dollars in the price of the house.
- When a house had 3 bedrooms as compared to any other number of bedrooms besides 2, 3, 4 and 5, we saw an associated increase of 97,460 dollars in the price of the house.
- When a house had 4 bedrooms as compared to any other number of bedrooms besides 2, 3, 4 and 5, we saw an associated increase of 25,122 dollars in the price of the house.

FIRST MULTIPLE REGRESSION MODEL RESULTS.

- For each increase in 1 bathroom, we saw an associated increase of 66,591 dollars in the price of the house.
- For each increase in 1 square-foot of the living area, we saw an associated increase of 280.71 dollars in the price of the house.
- For each increase in 1 square-foot of the lot area, we saw an associated decrease of 0.34 dollars in the price of the house.
- For each increase of 1 floor, we saw an associated increase of 59,682 dollars in the price of the house.
- For each year newer the house is, we saw an associated decrease of 3,142.8 dollars in the price of the house.
- For each grade higher the house is, we see an associated decrease of 25,624 dollars in the price of the house.

CHANGES TO THE FIRST MODEL.

- In order to get a better model, I had to make a few iterations to the multiple linear regression model.
- First of all, the square footage of the living area explained a lot about the change in the price of a house leaving very little to be explained by the square footage of the lot. Thus, I dropped the square footage of the lot from the modelling dataset.
- Secondly, choosing the bedrooms_other dummy category as the reference category made it harder to interpret the bedrooms variables. I instead made the bedrooms_2 dummy category the reference category.

FINAL MODEL RESULTS.

- The model was statistically significant and explained about 57% of the variance in price.
- Given that all other predictors were 0, the price of the house was 6,259,338 dollars.
- When a house had any other number of bedrooms besides 3, 4 and 5 as compared to a 2 bedroomed house, we saw an associated decrease of 176,091 dollars in the price of the house.
- When a house had 3 bedrooms as compared to a 2 bedroomed house, we saw an associated decrease of 80,027 dollars in the price of the house.
- When a house had 4 bedrooms as compared to a 2 bedroomed house, we saw an associated decrease of 150,833 dollars in the price of the house.

FINAL MODEL RESULTS.

- When a house had 5 bedrooms as compared to a 2 bedroomed house, we saw an associated decrease of 168,757 dollars in the price of the house.
- For each increase in 1 bathroom, we saw an associated increase of 68,080 dollars in the price of the house.
- For each increase in 1 square-foot of the living area, we saw an associated increase of 276.36 dollars in the price of the house.
- For each increase in 1 floor, we saw an associated increase of 61,862 dollars in the price of the house.

FINAL MODEL RESULTS

- For each year newer the house is, we saw an associated decrease of 3,171.93 dollars in the price of the house.
- For each grade higher the house is, we see an associated decrease of 25,418 dollars in the price of the house.

CONCLUSIONS

- Though few, houses with waterfronts fetch more money than houses without waterfronts.
- Houses with excellent views have the highest price average while houses without a view have the lowest price average.
- Houses that are in very good condition fetch the highest amount of money. The unexpected finding here is that houses that are in poor to fair condition almost have the same price average, with poor conditioned houses averaging slightly more money than fair conditioned houses.
- House prices are mostly influenced by the size of the living area.
- From the final model, I can conclude that the number of bathrooms and floors in a house have the highest positive effect on the price of the house.

RECOMMENDATIONS FOR REAL ESTATE AGENCIES.

- The agency should consider listing more houses that have waterfronts and excellent views because these houses are guaranteed to have a higher value than houses without waterfronts or views.
- The agency should also try to steer clear of houses that are in poor condition as these houses fetch little money and are hard to sell.
- The price of a house that has 2 or more floors and 2 or more bathrooms is guaranteed to increase by over 200,000 dollars. Listing more houses like these will lead to increased revenue for the agency.

RECOMMENDATIONS FOR CONSTRUCTION COMPANIES.

- When building houses for sale, a construction company should consider building more houses with spacious living areas. The square footage of the living area should be greater than the square footage of any other area of the house as this has a direct effect on the price of the house.
- The houses should also be fitted with more bathrooms and more floors. These two factors are positively related to the price of a house, which means that the more the number of bathrooms and floors, the higher the value of the house.

RECOMMENDATIONS FOR OWNERS INTERESTED IN SELLING.

- Before putting up the house for sale, home owners should consider making some renovations to the house so as to improve the condition of the house. Putting up the house for sale when its condition is average might lower the value of the house, so it is better to invest in some renovations where needed.

RECOMMENDATIONS FOR PARTIES INTERESTED IN BUYING.

- When looking for a house to buy, a potential home owner should focus on finding a house that has more bedrooms and fewer floors. This type of house is guaranteed to be cheaper as the model has shown that the more the number of bedrooms, the cheaper the house becomes but the more the number of floors, the higher the price of the house.
- Another factor to consider is that, houses with waterfronts and excellent views are very expensive. Potential home owners who have a tight budget should focus on houses without waterfronts and with fair views in order to minimize the cost.
- If one is trying to make a business out of buying and reselling houses, a good tactic would be to buy houses that are in poor conditions and renovate them so as to improve the condition of the houses before selling them at a higher price.



THANK YOU