

DEMYSTIFYING HOUSE SALES ANALYSIS WITH REGRESSION MODELING IN A NORTHWESTERN COUNTY



BUSINESS UNDERSTANDING

This project emphasizes the critical role of real estate in a region's well-being and seeks to empower informed decision-making through a deeper understanding of house price determinants in this specific northwestern county.



CHALLENGES OF A FLUCTUATING REAL ESTATE MARKET



Agent Dilemma:
Market swings hinder agents' ability to price houses and advise on offers.



Buyer Uncertainty: In a volatile market, rapid price changes pose challenges for homebuyers, risking overpayment.



Seller's Gamble: Timing the market for maximum profit is a stressful gamble with low odds and high pressure.



Developer's Conundrum: High land prices and buyer affordability challenge builders in volatile markets





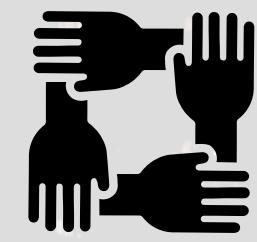
PROBLEM STATEMENT

We aim to determine the factors driving house prices in a specific northwest US county, explore methods to quantify their impact (e.g., bedrooms, location), and develop a predictive tool for house prices based on these features.

CONCLUSION

We're crafting a predictive tool for house prices to aid real estate agents in navigating the volatile market in a northwest county

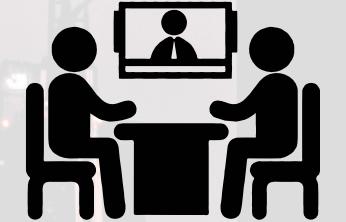
OBJECTIVES



Develop a robust multiple linear regression model to predict house prices accurately in the chosen northwest county.



identifying key factors influencing property value.



Offer valuable insights into regional housing market dynamics, benefiting buyers, agents, and stakeholders

DATA UNDERSTANDING

Our analysis leverages the King County House Sales data collected from May 2014 to May 2015?

it contains over 21,500 records and 20 distinct features/columns

Review and editing

Presentation and sharing

DATA PREPARATION

This includes cleaning the data and handling missing values, removing the errors and correcting the duplicates



Creating visual representation of the data we are working with



Investigate relationships between variables, identify outliers, and understand the underlying structure of the data..



EXPLANATORY DATA ANALYSIS

we will perform exploratory data analysis (EDA) to understand the data better and discover any patterns, trends using univariate,bivariate and multivariate analysis

Univariate Analysis

Univariate analysis involves the examination of single variables. We focus in the summary statistics of target variable - price to help us understand the distribution and skewness of house prices

Bivariate Analysis

we will perform bivariate analysis to examine the relationship between the target variable - price and the other numeric and continuous features in the data.

Multivariate Analysis

we will perform multivariate analysis to examine the relationship between the target variable - price and multiple features in the data

RECOMMENDATIONS

Under univariate analysis, consider the house prices since most of the prices are relatively low.

Under bivariate analysis, most independent variables such as sell year, house age, bathrooms and bedrooms influence the house price.

Under multivariate analysis, bathrooms and square foot living are the most independent variables that affect the house price.

REGRESSION MODELLING

Why did we do regression?

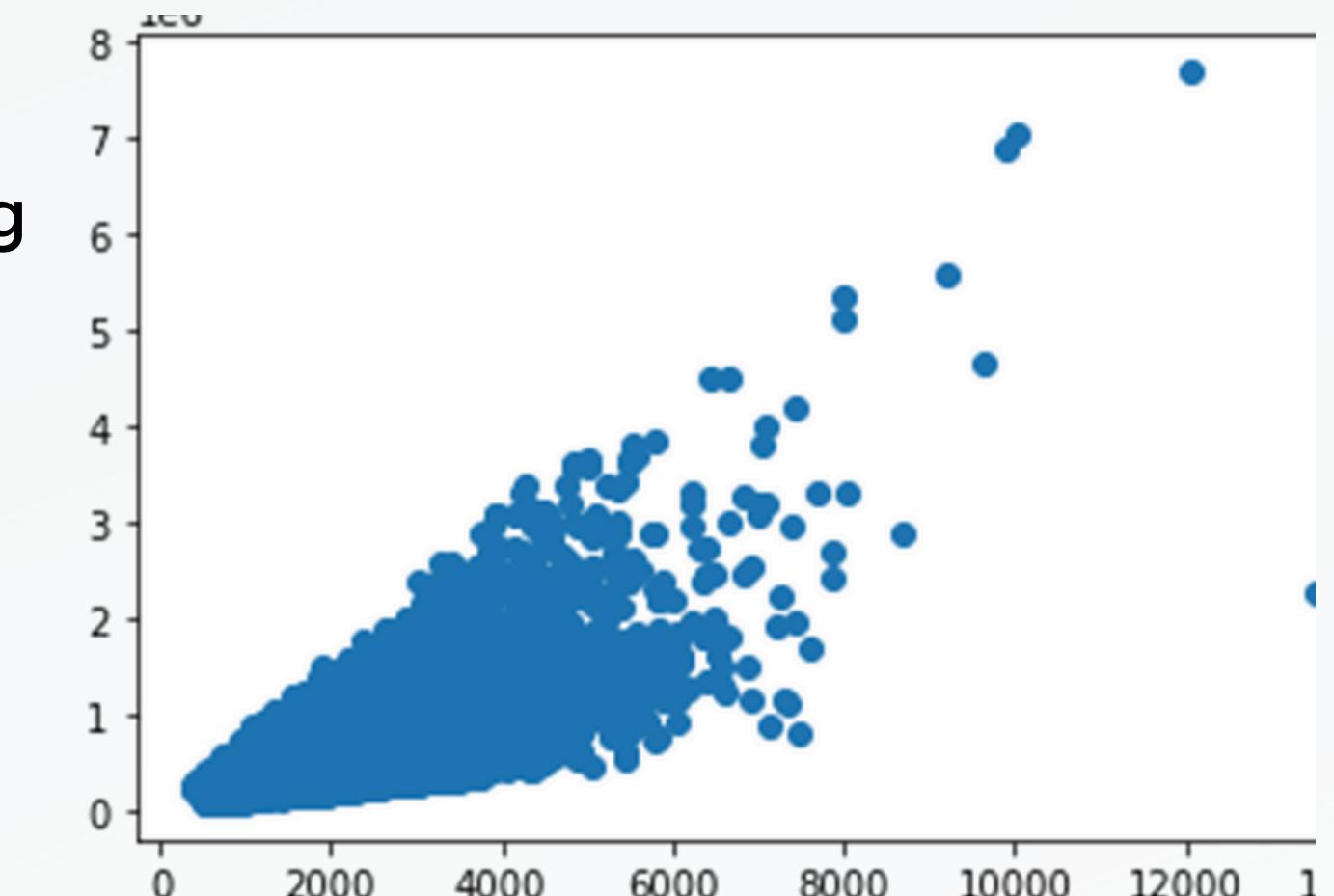
- Regression is done to build models that will help us predict house prices given some house features for a stakeholder.
- It will also allow a renovator know what to improve on to increase the value of their house

SIMPLE LINEAR REGRESSION

What is simple linear regression?

Simple linear regression is building models from one independent variable or predictor

We built our model based off of the sqft_living column as it has the highest correlation to price(0.704441)



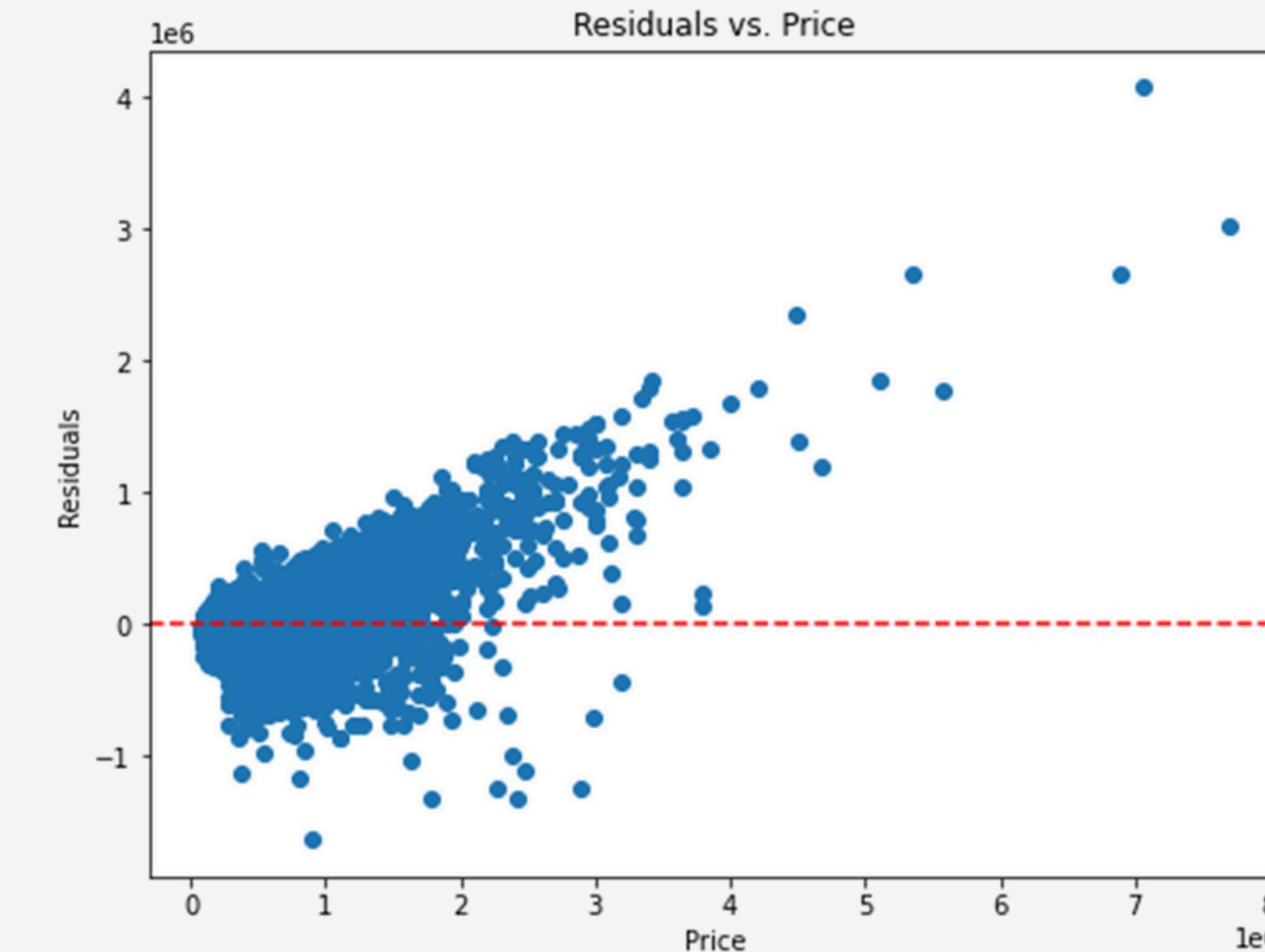
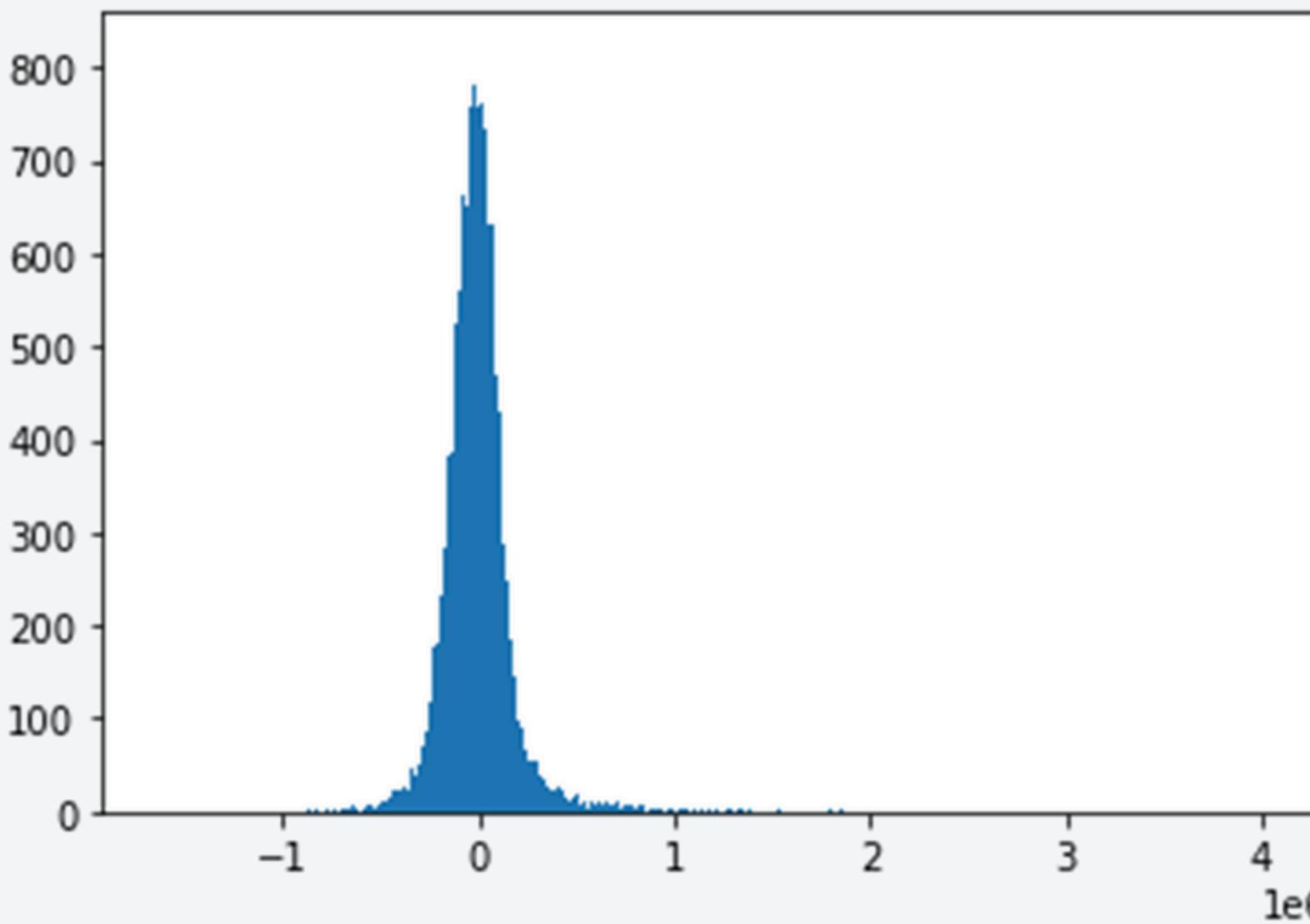
MULTIPLE LINEAR REGRESSION

We built this model to hopefully fill the gaps that the simple model left

For the first model we used the entire dataset meaning we had to use some dummies on our categorical data and drop one of the dummy columns to remove perfect multi collinearity

We got an r-squared adj of 72.8% and a probability statistic below 5% which is much better than what we had before

Checking for LINE features



We have very few linear distributions and a terrible mse of 1.9281434833350915e+17

We will do better with this in the next model

MODEL 3

The predictors we used are:

Bathrooms, sqft_living, grade , sqft_above, sqft_living' as they have better correlation thus a more linear relationship with price

MODEL SUMMARY

We now have a new mse of 230179.555569266 which is much better

We now have an r-squares adj of 87%

From the calculated MAE and RMSE we can tell that the third model is way accurate compared to the previous 2 models and we recommend using 3rd model in predicting the price of a house.

RECCOMENDATIONS

- Consider the square footage of the house if you are buying a house for resale
- If renovating and buying as well an increase in the number of bathrooms will increase the value off the house

STAKE HOLDERS

- Handle all outliers for better models
- Use linearly related data for better models

MODEL DEVELOPERS





OUR TEAM

- Calvine Dasilver
- Jack Otieno
- Salahudin Salat
- Hellen Samuel
- Sandra Kiptum

THANK
YOU