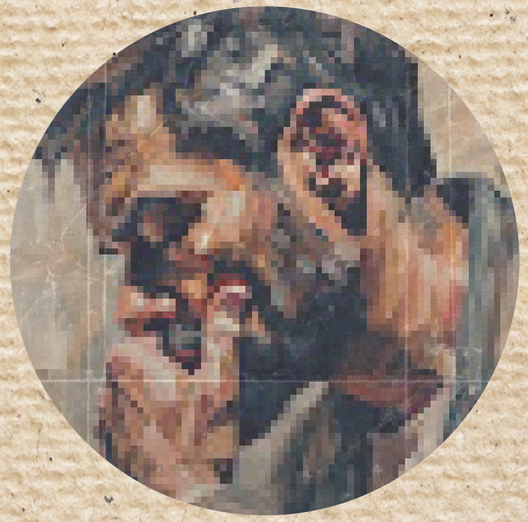
Group 1
Phase 2

HOUSING SALES PROJECT

Members of Group 1

Esther Nyakuiba



Cleve Mwebi

Mark Njagi

Arnold Mochama

Project Overview

This project involved developing a multivariate linear regression model to predict housing prices in King County, Washington based on features like square footage, number of bedrooms/bathrooms, and overall condition. The model aims to help real estate agencies estimate property values. The dataset contained information on 21,597 houses sold between May 2014 - May 2015 sourced from the King Country House Dataset.

Business Understanding

-Key stakeholders: Real estate agencies and clients in King County
-Aim: Develop models that will help in the prediction of house prices from attributes like size, bedrooms, etc.
-Benefits: More accurate valuations for agents and clients; informed investments
-Strategic edge: Identify most influential pricing factors (e.g. bathrooms vs bedrooms)
-Competitive advantage: Enhanced predictive modeling capabilities from data analytics
-Overall goal: Leverage King County housing data to create a regression model that explains and est

-

Objectives

To develop a model to estimate the price of house based on it features

To identify the neighborhoods with the highest sales prices.

To identify how seasonal trends affect sales.

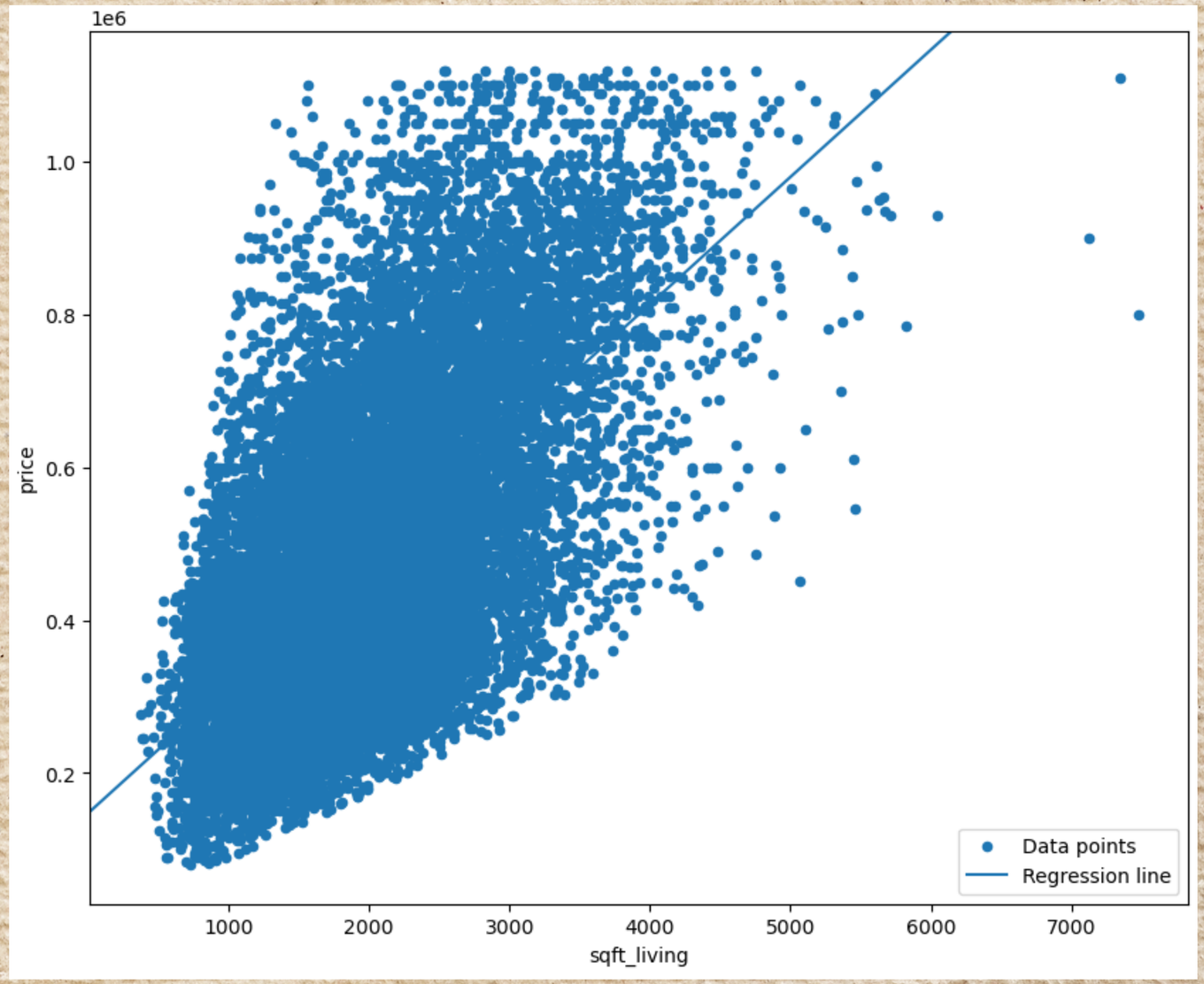
Data Understanding

The core dataset for this project is the King County House Sales dataset.

Key highlights:

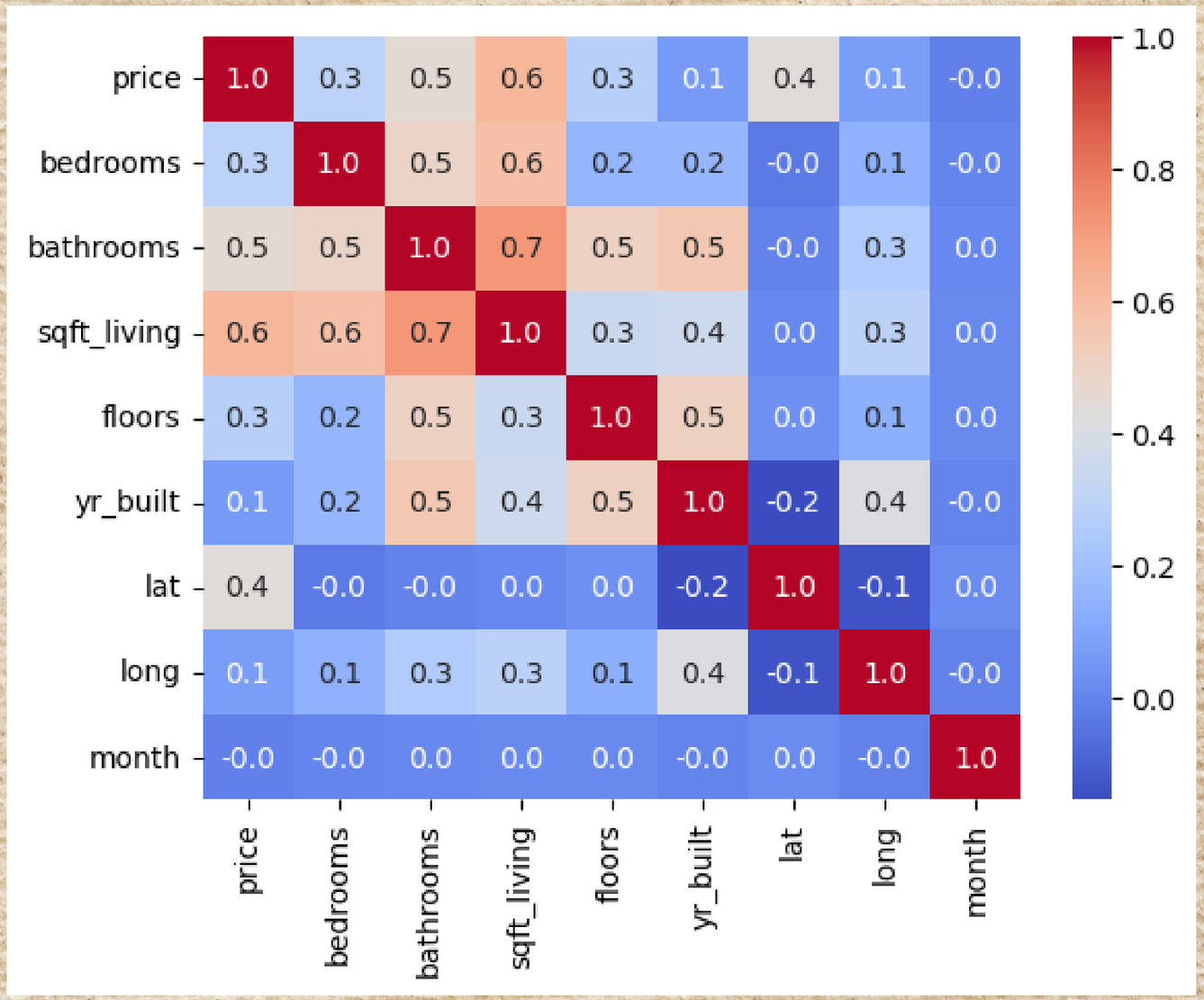
- Source: University of Chicago's Center for Spatial Data Science
- Time period: May 2014 - May 2015
- Format: CSV file
- Rows: 21,597
- Columns: 21 (sale price, bedrooms, bathrooms, sqft, location, date etc.)

Data Visualizations



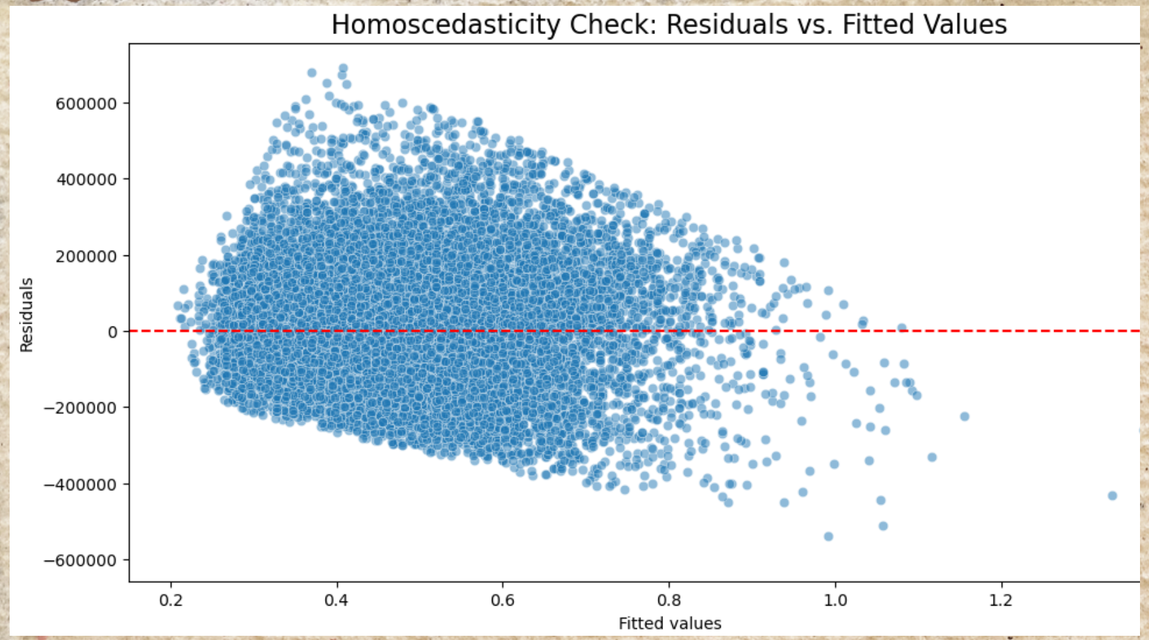
Scatter plot showing a positive corelation between square foot living (sqft_living) and house prices

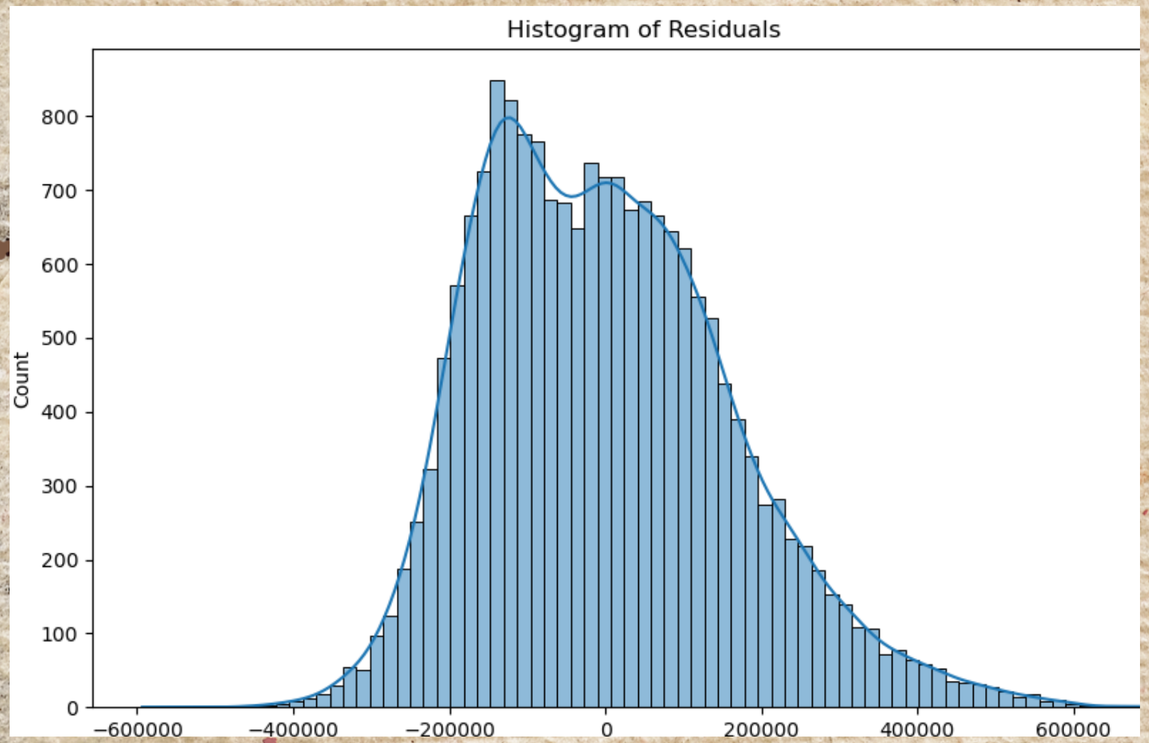
Heatmap showing the correlation of different features.

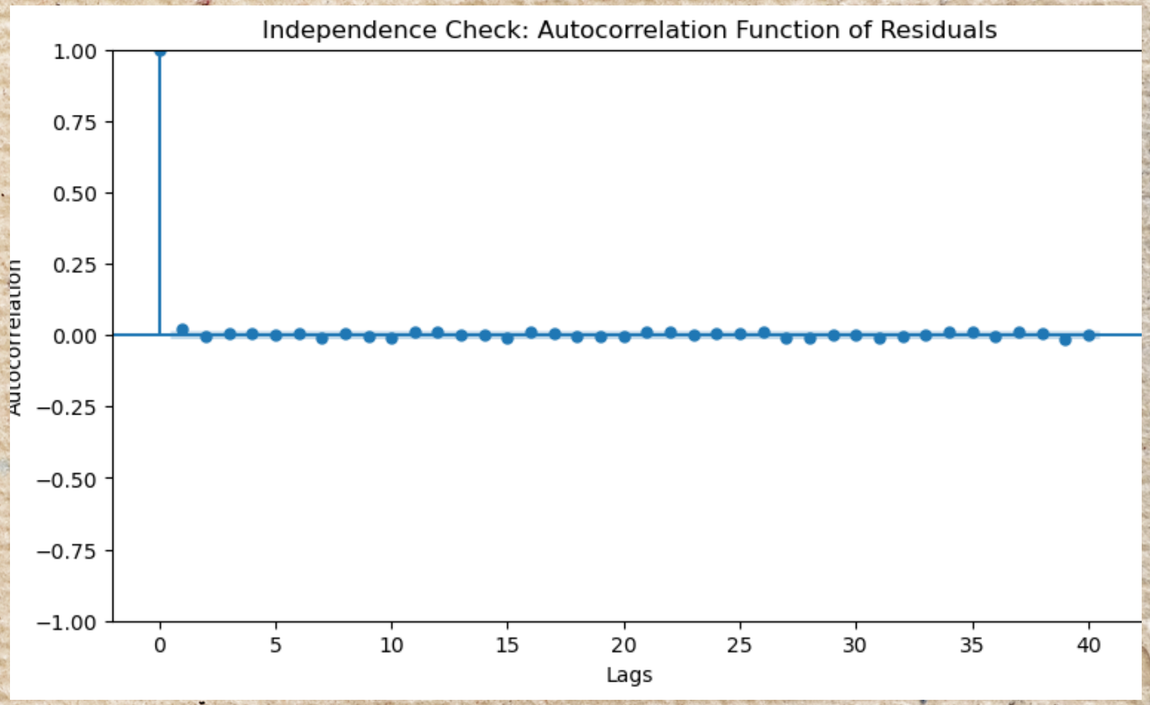


The scatter plot served to check for linearity between the variables 'sqft_living' (predictor variable) and 'price' (target variable). It is clear from the plot that there's a positive linear relation between between the two. As such these will serve to create our baseline linear regression model.

Assumptions for Regression







These three plots are a testament to the regression assumptions having been met: Homoscedasticity, Independence of variables and normal distribution of residuals

Conclusion

Approximately 40.3% of the observed fluctuations in housing prices may be explained by our predictive model. Property values are heavily influenced by key factors such as square footage, bathrooms, bedrooms, and house condition. But it's important to recognize that a lot of dynamic factors affect real estate, which makes predicting house values with 100% accuracy quite difficult. As a reliable tool for the Real Estate Agency and its clients to estimate house prices, our model attempts to explain and provide aid with property values within King County. Nonetheless, to make more precise price judgments, it is advised to supplement the model with extra data.

C. Cont...

-Key Factors Affecting House Prices: The coefficients of the independent variables in the model provide insights into their impact on house prices in the northwestern county.

price = (sqft_living * 176.2193) - (bedrooms * 26,770) + (bathrooms * 15,990) + (condtion * 30560) + 111,300

-Positive Factors: Variables such as the number of bathrooms, square footage of living space (sqft_living), and house condition have positive coefficients. An indication that an increase in these factors is associated with higher house prices.

-

C. Cont...

-Negative Factors: The number of bedrooms (bedrooms) has a negative coefficient implying that an increase in the number of bedrooms with all other factors held constant may lead to lower house prices.

-Customer satisfaction, knowing the above features influence the buyers' decision, the stakeholder can advise on what renovations are most marketable and what properties will have the highest appreciation.

Recommendations

-The size of the living space, as represented by the 'Sqft_living' variable, is a significant driver of house prices. If possible, consider focusing on properties with more substantial living spaces, as they tend to command higher prices.

- While an increase in the number of bedrooms is associated with a reduction in price in your model. However, this should be investigated further. It might be specific to the dataset or region you're working with.

R. Cont...

-The condition of a house significantly impacts its price. If you're looking to sell or invest in properties, it's worthwhile to maintain or improve their condition to increase their market value.

- From our analysis it was found that the number of bathrooms positively correlates with the house prices we recommend that our stakeholder puts more investment/ consideration of this factor be it in the buying, selling or even renovation of houses.

R. Cont..

-Our analysis proves that location is also a huge factor that affects the appreciation or depreciation of house prices. Real estate agents should therefore consider the neighborhood when advising their clients on the property to invest in. In particular, the agents should look out for the safety of the neighborhoods, proximity to key amenities such as hospitals and schools, as well as the affluency of a location when determining the price levels

Next Steps

Conducting additional analysis to pinpoint the factors that are influencing the model's predictions would be beneficial to explore the inclusion of new variables that could enhance the model's accuracy, for example crime data by zip code, school rating data as well as data regarding other social amenities.

Thank You!