

KING COUNTY HOUSING DATA ANALYSIS





Overview

The price of a home is affected by several factors that are important for real estate agents and homeowners to understand.

This presentation will cover the key factors that influence the price of a home, including bathrooms, living area and lot size, floors, condition and grade, age and renovation, waterfront view, and season.

We will explore each of these factors in detail and provide data, examples, and recommendations for how to price and market homes based on these factors.

By the end of this presentation, you will have a better understanding of what drives the price of a home and how to maximize its value in the market.



Business Understanding

As a data scientist analyzing the King County housing market, my business understanding is that the real estate industry is a crucial sector that plays a significant role in the economy. The success of a real estate transaction depends on several factors, including the location, the size of the property, the condition of the property, the amenities, and the current market conditions. The housing market is subject to various external factors such as interest rates, economic conditions, and government policies that can impact the demand and supply of properties.

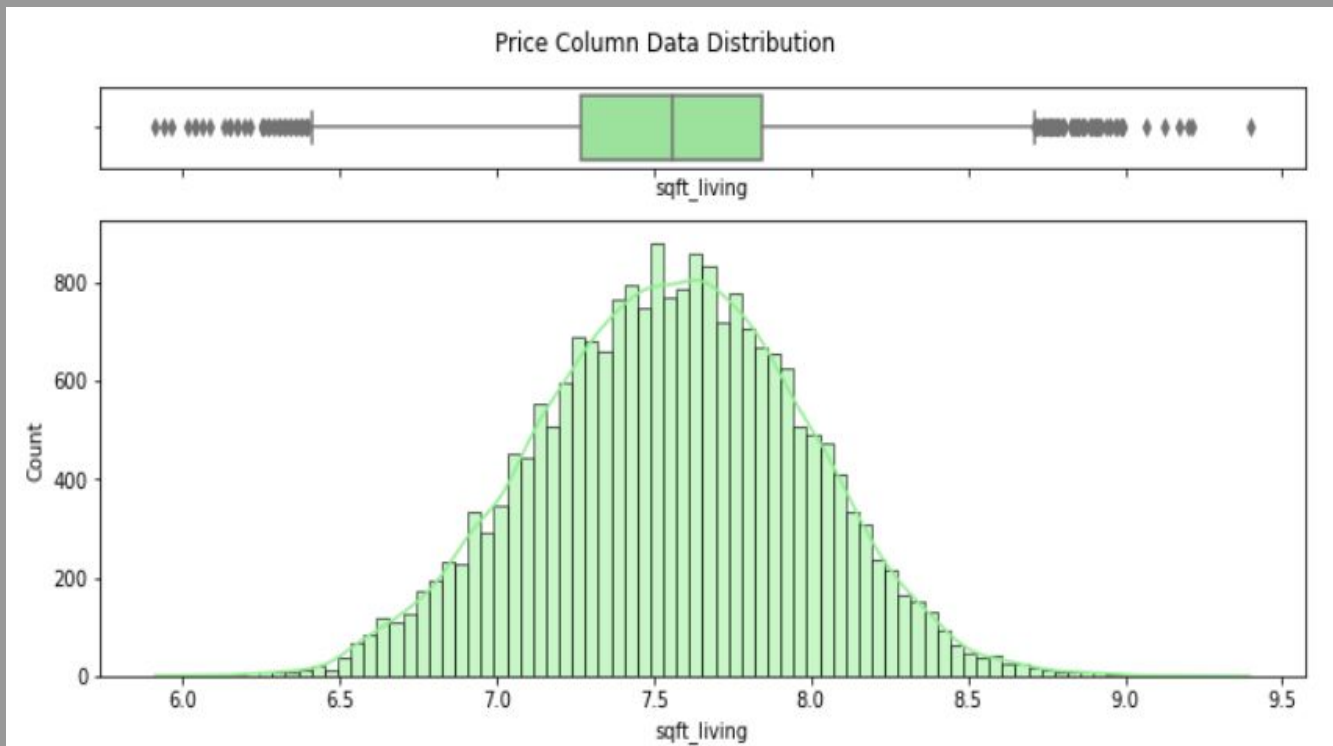
A real estate agency from King County, Seattle hired us for a project to analyse how different factors affect prices of homes. The aim of this analysis is to build a multiple linear regression model that predicts the prices of houses in King County, Seattle.



Data Understanding

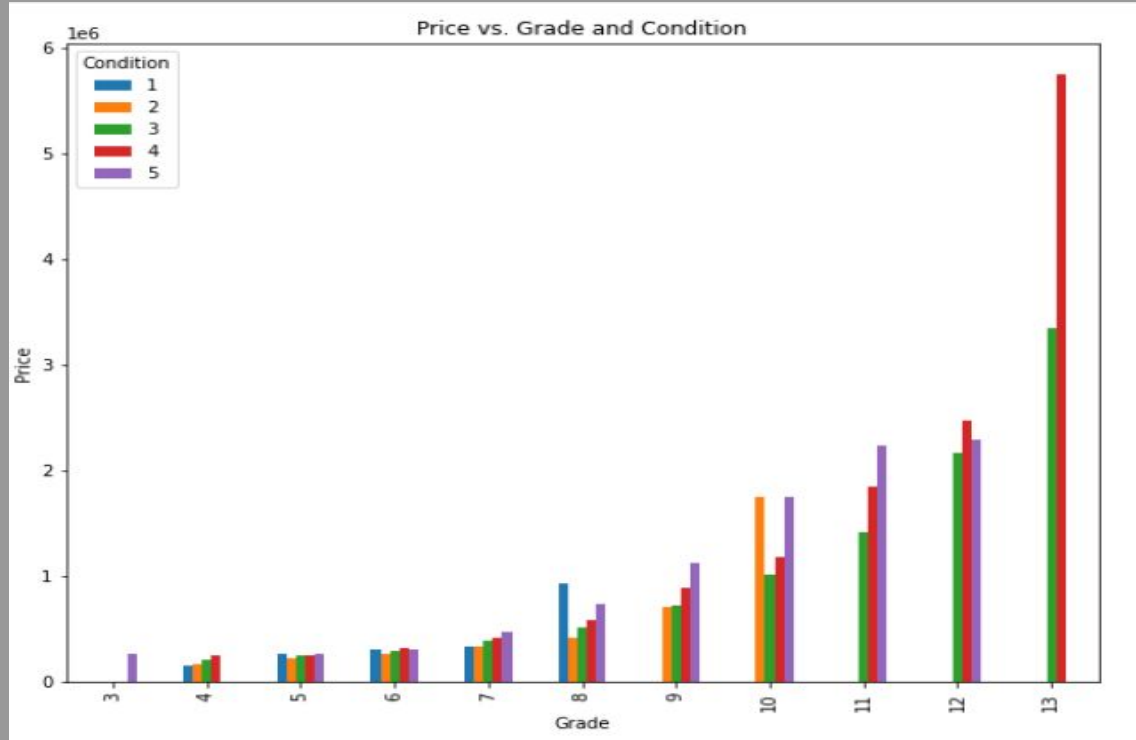
The dataset contains information about the houses in King County, Seattle. The dataset has 21 variables including the price, number of bedrooms, bathrooms, square footage of the living area, and other variables. The dataset contains 21,597 observations. The scope of this analysis is limited to the data provided. We will use feature engineering techniques such as imputation, normalization, and one-hot encoding to preprocess the data. We will use multiple linear regression model. We will evaluate the performance of the model using metrics such as mean squared error, mean absolute error, and R-squared.

Univariate Analysis



The living area column distribution is not skewed as the and is normally distributed

Bivariate analysis



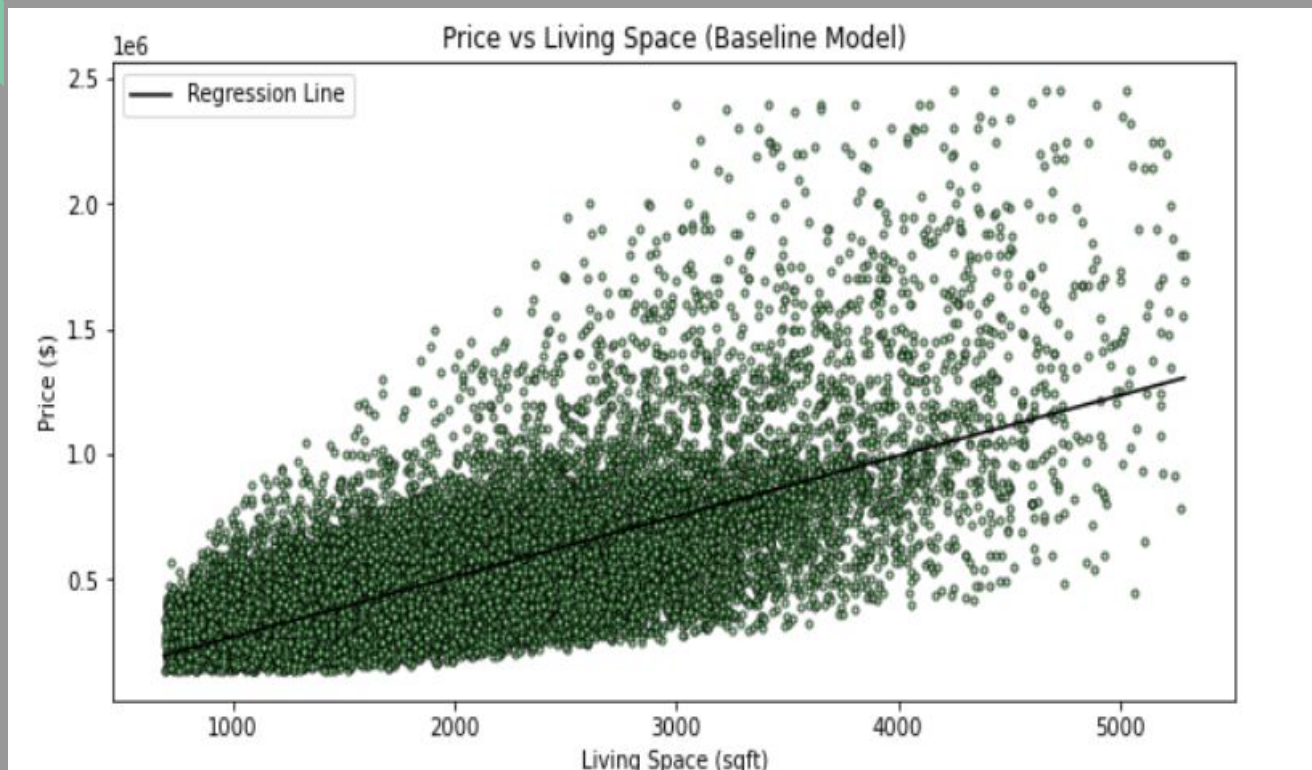
As grade and condition increase, the higher the price, especially for those with a condition of 5. The worse the condition and grade the lower the price.

Multivariate Analysis



the correlations
between multiple
variables['price',
'bedrooms',
'bathrooms',
'sqft_living'] at
once

Simple Regression Model



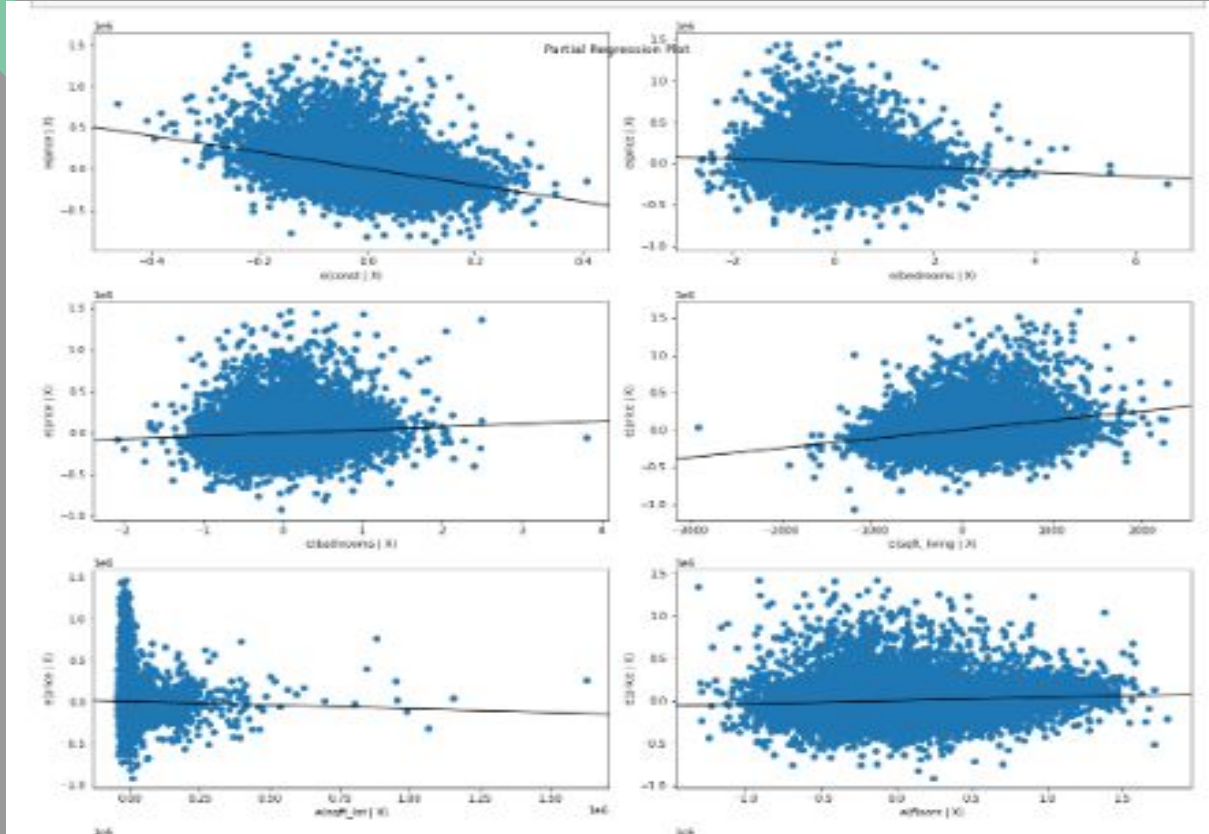
From this model, there appears to be a strong relationship between living space and price.



Simple Linear Regression Results

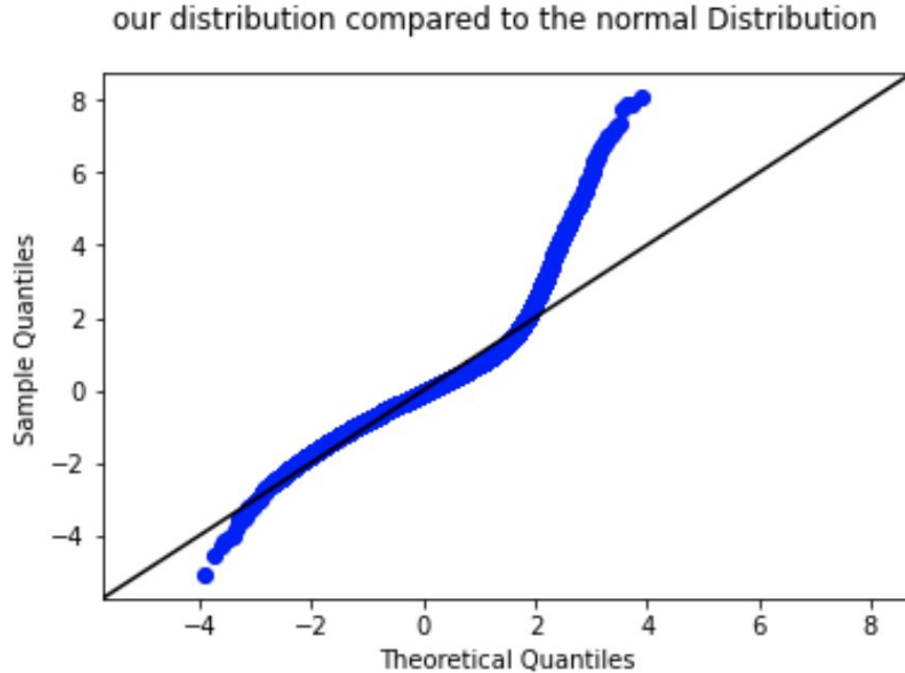
- approximately 44.3% of the variance in housing prices can be explained by the square footage of the living area
- for each one-unit increase in square footage of the living area, the housing price increases by \$240.99

Multiple Linear Regression



a partial regression plot for our model for each predictor variable

Multiple Regression Model:



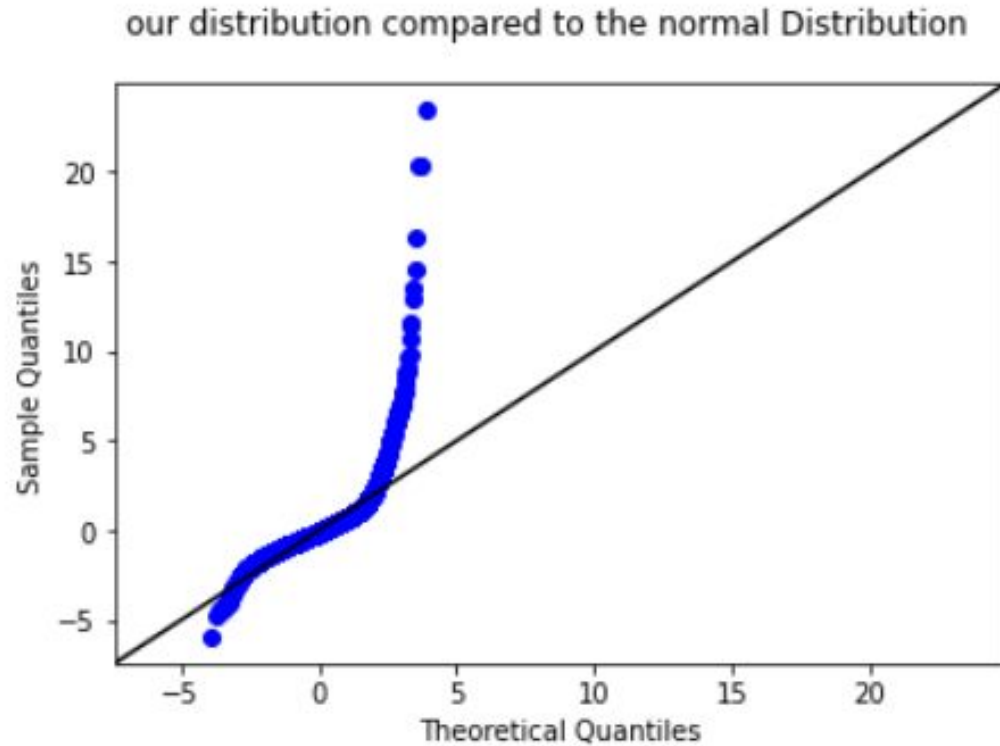
A Q-Q plot, of the model compared to normal distribution.



Multiple Regression Results:

- The model explains about 63.4% of the variance in the target variable (price). This is a significant improvement over the previous model which had 44.3%.
- We can also see that some variables have a stronger effect than others. For example, the coefficient for the waterfront view variable is much larger than the coefficients for the other variables, indicating that having a waterfront view is a very significant factor in determining the price of a house.
- Overall, this model provides a more comprehensive understanding of the factors that affect the price of a house, and can be used to make more accurate predictions of house prices based on the characteristics of the house

Log Transformed Regression Model





Log Transformed Model Results

- The log-transformed model explains 57.1% of the variance, indicating that the model is worse than the previous model.




Recommendations:

1. Bathrooms: More bathrooms equals higher price.
2. Living Area and Lot Size: Emphasize living area, be mindful of lot size.
3. Floors: Multi-story homes are priced higher.
4. Condition and Grade: Higher ratings equal higher price.
5. Age and Renovated: Consider age and renovation when pricing and marketing.
6. Waterfront View: Waterfront properties are priced significantly higher.
7. Season: Spring sells for higher prices than fall.



Next steps

- 
- Develop and compare multiple regression models to predict house prices with high accuracy.
 - provide data or statistics to back up your claim. This could include research studies, market data, or industry reports.
 - Use examples or case studies to demonstrate how each factor affects the price of a home



THANK YOU

With regards:

Isaac Muturi,

Henry Mativo,

Nelson Odhiambo,

Mark Ngige,

Marwa Osman,

Fiona Kung'u