KING COUNTY PROPERTY SALES ANALYSIS.

The Team

- Immaculate Mwendwa
- Shalom Irungu
- Stephen Kariuki
- Stella Ndegwa
- Muhsin Ahmed
- Joan Wambua

Contents

- Business Understanding
- Data Understanding and Exploration
- Regression and Modeling
- Outcome
- Recommendations
- Next Steps

Business Understanding

King County, with 2.252 million residents, experiences a competitive housing market due to high demand and limited supply, driven by population growth and increased wages.

The housing crisis, rooted in supply and demand imbalances, guides our role as data scientists in offering insights on home features impacting property values.

Data Understanding and Exploration

The King County House Sales dataset, covering 21,420 property sales from 2014 to 2015, details features like views and square footage.

The process includes descriptive and inferential analysis, employing multiple linear regression to create a model that aids buyers and sellers in identifying price-influencing factors and guides real estate agencies in pricing decisions.

Regression and Modeling

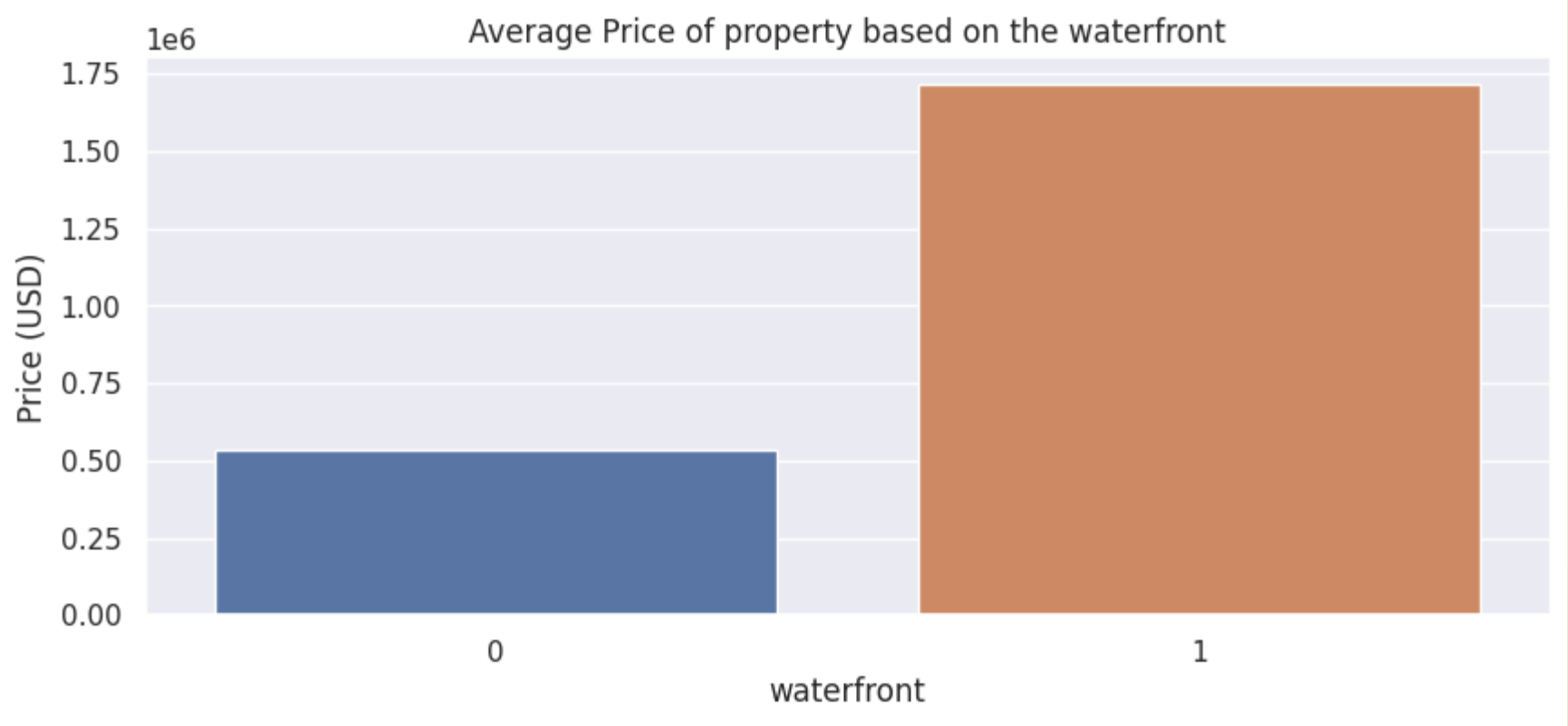
Our initial model was built with the variable 'sqft_living' due to its strongest correlation with Price.

Subsequently, employing an iterative approach, we constructed various models, ultimately arriving at our final regression model.

Grade Vs. Price

Observing the building grade scale, ranging from 3 (low) to 13 (mansion), it can be deduced that as the grade of the structure rises, so does the corresponding price.

Waterfront Vs. Price



In the context where 0 signifies no waterfront and 1 indicates waterfront, it can be deduced that properties with waterfronts possess a higher price value compared to those without.

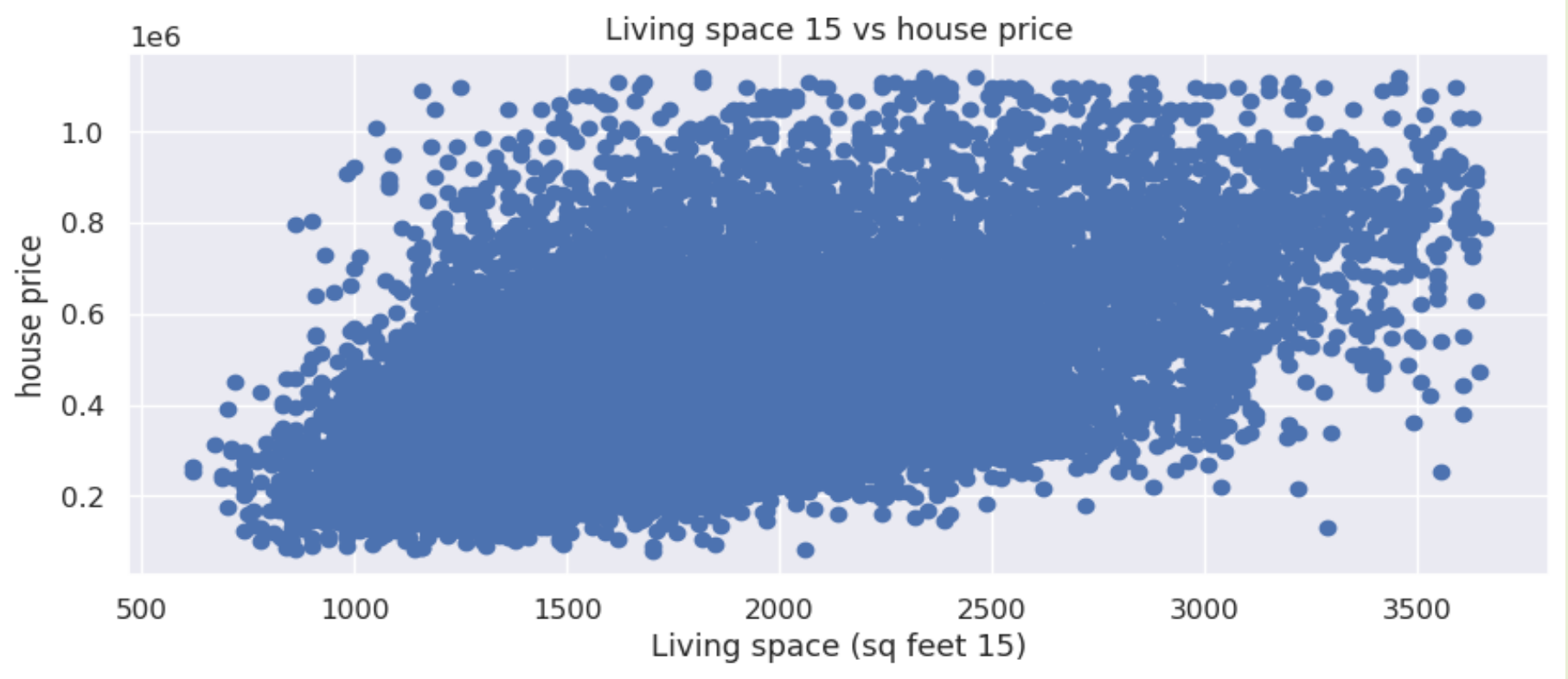
Sqft_living Vs Price

As the living space increases, we see the price also increases. This shows that bigger houses in terms of square feet cost more than small houses.

Condition Vs. Price

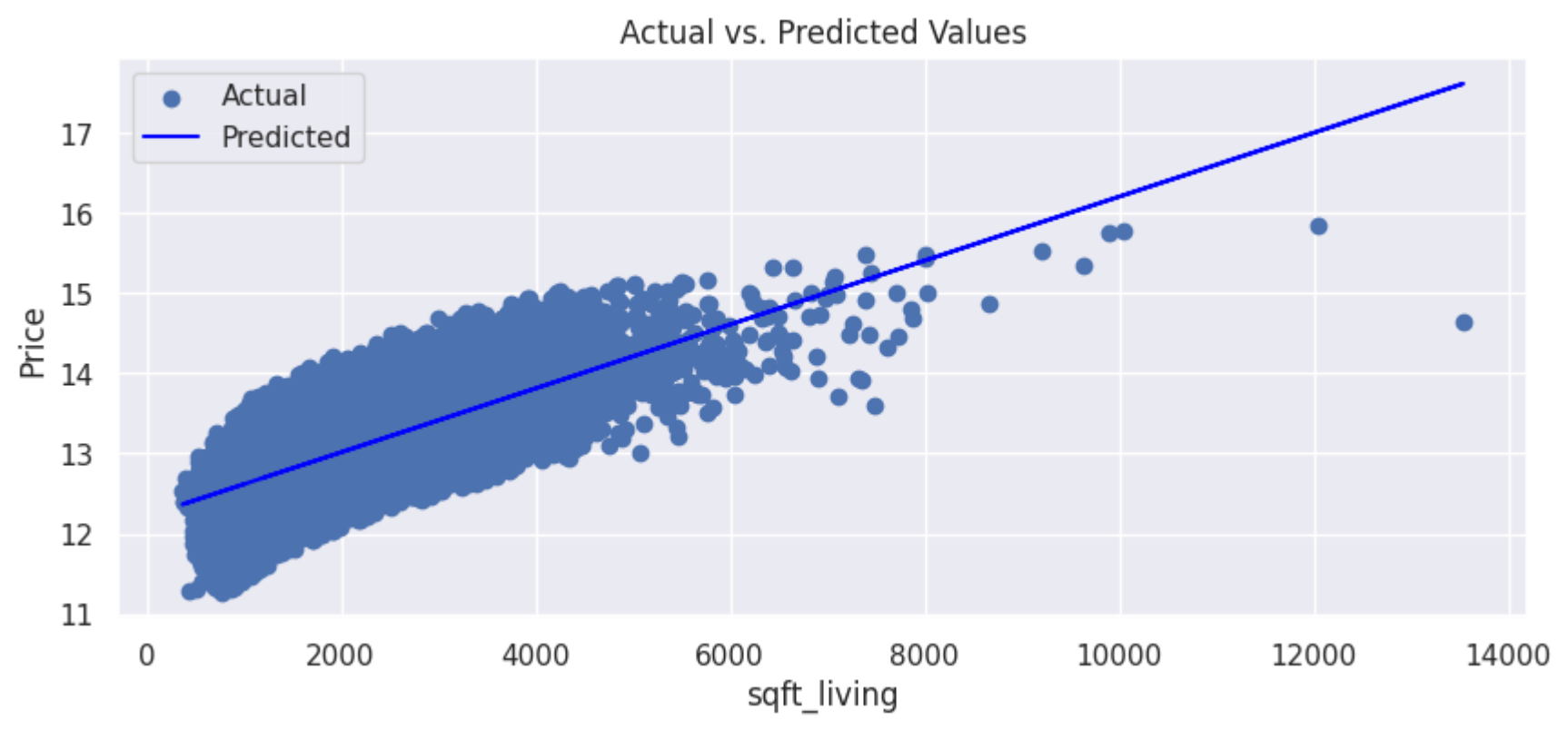
We see that properties with low condition have a lower price value. We can also see that properties with a high condition value have a higher price value.

Sqft_living15 Vs Price



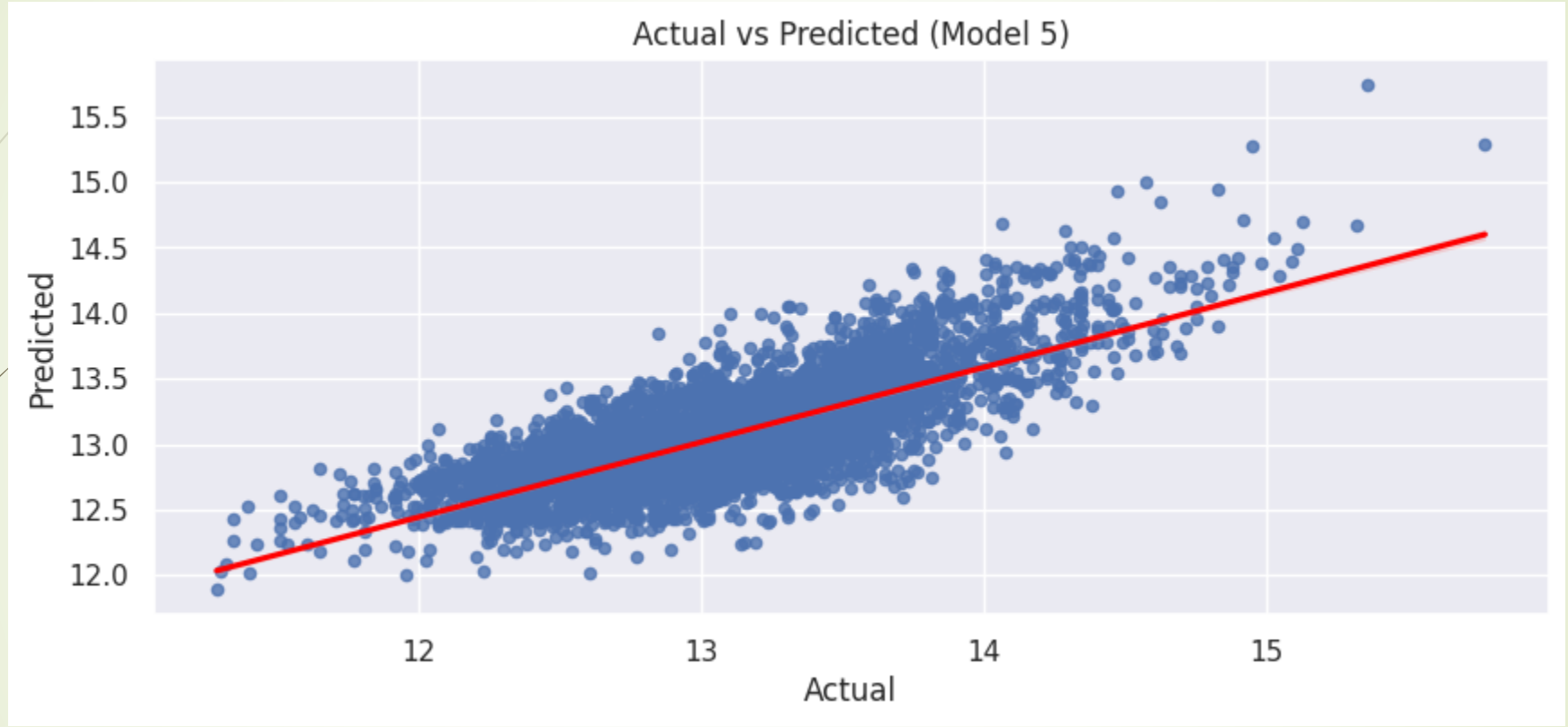
We can infer that as the square footage of interior housing living space for the nearest 15 neighbors increases, the price of the property also increases.

Baseline Model



We built our baseline model using sqft_living because it has the strongest correlation with Price. This models aids to predict house price. The R-squared value is 0.493, indicating that approximately 49.3% of the variance in price can be explained by the sqft_living.

Iterated Model



This model incorporates additional features and categorical data with low p-values. Its higher R-squared (0.590) and adjusted R-squared (0.589) values strike a balance between explanatory power and complexity.

Outcome

Model Validity: While the model explains a significant portion of housing price variance, further investigation is needed to better understand the impact of bathrooms and enhance model validity.

Influential Property Features: Bedrooms, sqft_living, view, condition, and grade play crucial roles in determining housing prices, offering essential considerations for both buyers and sellers.

Model Limitations: The model has limitations, and log transformation was applied to meet linearity assumptions.

Recommendations

1. Prioritize bedrooms, sqft_living, view, condition, and grade for accurate pricing.
2. Enhance property features to elevate grade and communicate financial gains.
3. Address bathroom impact inconsistency by refining variables.
4. Promote affordable properties with fewer bedrooms.
5. Leverage scenic views with visuals to attract buyers.
6. Continuously improve the model with additional features and refinements.

Next Steps

Collecting more data on the homes is necessary to ensure no crucial features are overlooked. Exploring alternative methods beyond regression may be considered to fulfill the assumptions.

THANK YOU