DATAGURU-5: PHASE 2 PROJECT

Oscar, Eva, Juliet & Wayne

SEPTEMBER 2023

HKPM

DATA-DRIVEN REVENUE MAXIMIZATION FOR HAVEN-KINGS PROPERTY MANAGEMENT

INTRODUCTION

HKPM

Haven-Kings Property Management, operating in King County's dynamic real estate market, seeks to optimize rental pricing. Traditionally, they've relied on Comparative Market Analysis (CMA) and the "1% Rule." This project aims to revolutionize pricing with data-driven, dynamic recommendations using linear regression.

OBJECTIVES

HKPM

1.

***OPTIMIZE HOUSE PRICING*:** DEVELOP A PRICING TOOL CONSIDERING HOUSE CHARACTERISTICS, LOCATION, AND MARKET CONDITIONS

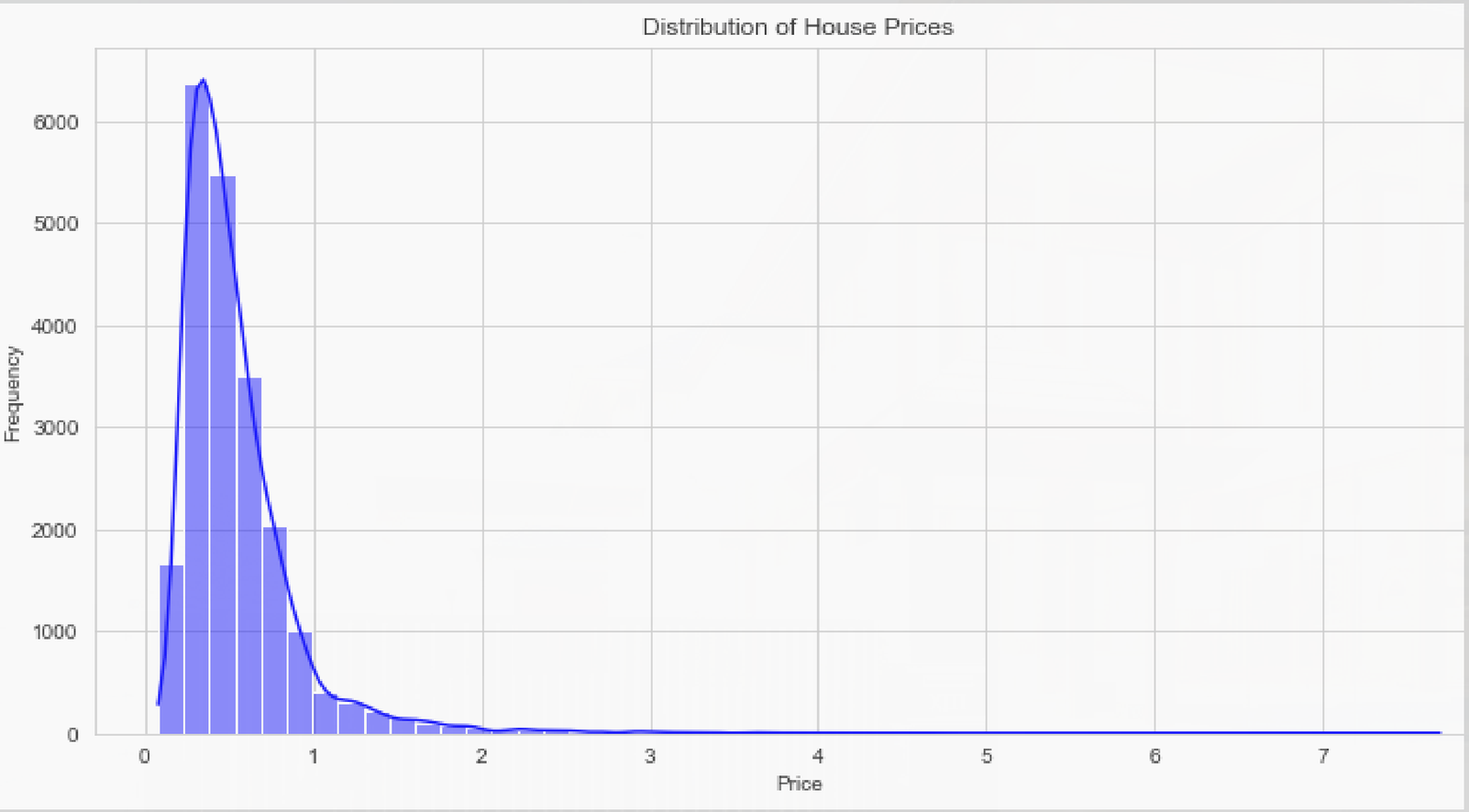
2.

PREDICT RENOVATION NEEDS: UTILIZE HISTORICAL DATA AND PROPERTY-SPECIFIC FEATURES TO PREDICT RENOV PLANNING AND BUDGETING.

3.

DYNAMIC PRICING RECOMMENDATIONS: CREATE DYNAMIC RENTAL PRICING RECOMMENDATIONS USING DATA ANALYTICS

EXPLORATORY DATA ANALYSIS (EDA)



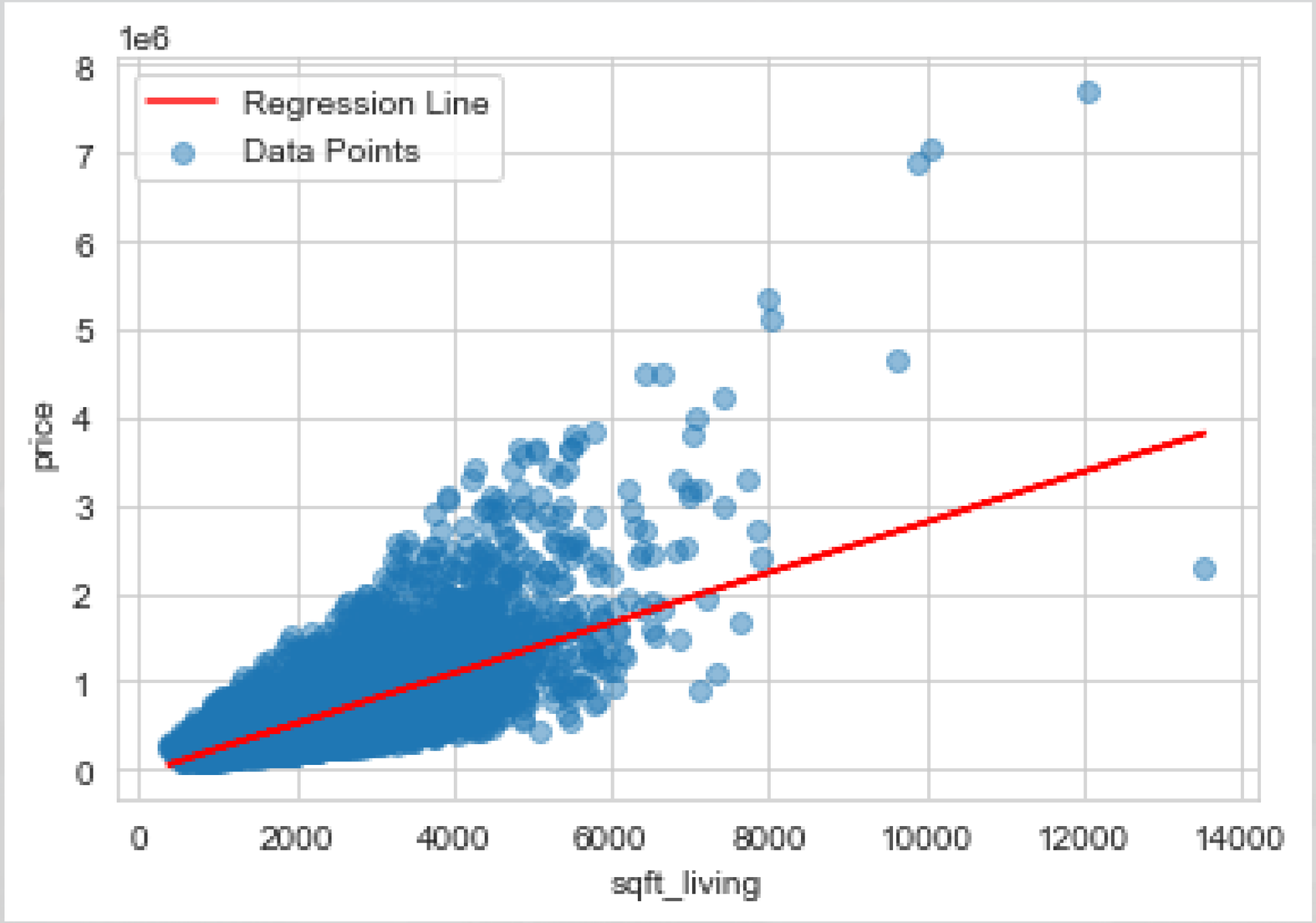
Price being our depandant variable was a column of interest. House price against frequency forms a right-skewed distribution, indicating a few houses with extremely high prices compared to the majority

MODEL 1

Simple Regression

HKPM

- For each additional square foot of living area, the price is estimated to increase
- As the sqft_living value increases our model's performance declines



Price vs. sqft_living

Assumption Checks

The plot shows presence of heteroscedasticity in the residuals plot i.e spread of residuals systematically increases as the predicted values (fitted values) change.

Applying Box-Cox Transformation to address heteroscadasticity

- After the Box-Cox transformation, the residuals are more evenly distributed around zero.

- These improvements are indicative of a more reliable linear regression model that aligns better with the assumptions of homoscedasticity and normality.

MULTIPLE LINEAR REGRESSION

HKPM

MODEL 2 :

PREDICTORS:

- PROPERTY GRADE
- CONDITION
- THE SQUARE FOOTAGE OF LIVING SPACE

MODEL 3 :

PREDICTORS:

- PROPERTY GRADE
- CONDITION
- THE SQUARE FOOTAGE OF LIVING SPACE
- AGE OF HOUSE

MODEL 4 :

PREDICTORS:

- PROPERTY GRADE
- CONDITION
- THE SQUARE FOOTAGE OF LIVING SPACE
- AGE OF HOUSE
- YEARS BETWEEN RENOVATION AND SALE

MODEL 5 :

PREDICTORS:

- PROPERTY GRADE
- CONDITION
- THE SQUARE FOOTAGE OF LIVING SPACE
- AGE OF HOUSE
- YEARS BETWEEN RENOVATION AND SALE
- VIEW

MODEL 2

Model 2 explained more variance in house prices (R-squared 0.548) by adding predictors, but RMSE barely changed. Improved explanation didn't translate to better predictions, emphasizing the balance between explanation and prediction.

MODEL 3

Model 3 Conclusion: Model 3 with house age improved explanation (R-squared 0.601), but RMSE remained similar. Older houses showed higher prices, possibly due to location or historical value, illustrating the value of including this predictor.

MODEL 4

Model 4 introduces "Years_Between_Renovation_and_Sale" to gauge renovation impact on house price. While significant, its effect is minor, with an R-squared of 0.601. Predictive accuracy remains almost the same.

MODEL 5

Model 5 adds "floors" and "view" variables, improving R-squared to 0.616. Houses with better views and more floors command higher prices. However, RMSE remains unchanged, emphasizing the trade-off between explanatory power and prediction accuracy in regression models.

CONCLUSION AND RECOMMENDATIONS

HKPM

In summary, the analysis of various regression models reveals important insights into the factors affecting house pricing. Models that incorporate additional variables, such as floors and view, provide enhanced explanatory power but do not notably improve predictive accuracy. Therefore, when determining house prices, it is crucial to consider factors like size, age, and view, while recognizing that a balance between model interpretability and predictive performance is essential.

HKPM

THANK YOU!