



**PROJECT
PHASE2**

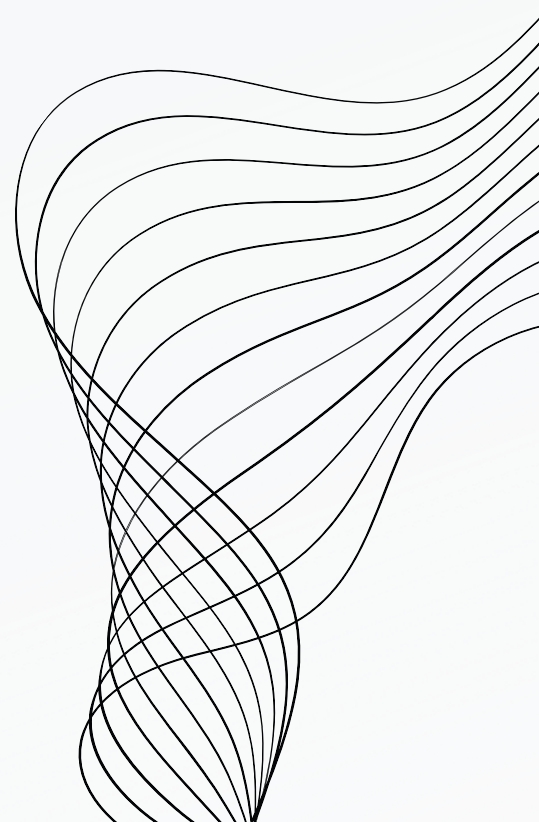
VALUE ANALYSIS FOR HOUSE SALES IN KING COUNTY

JULIE CHEPNGENO



INTRODUCTION

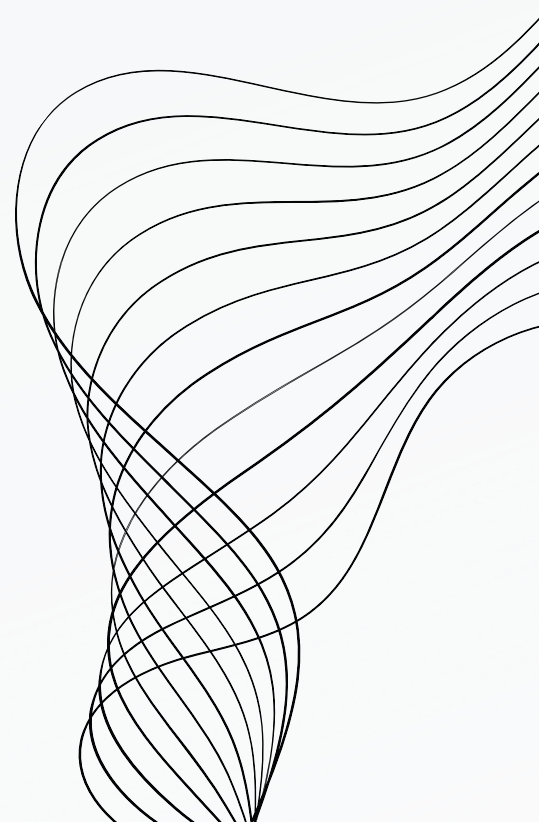
At a Real Estate, buying or selling a home is one of the most significant financial decisions you'll make in your life. Our goal is to make the home buying or selling process as smooth and stress-free as possible for our clients





PROBLEM STATEMENT

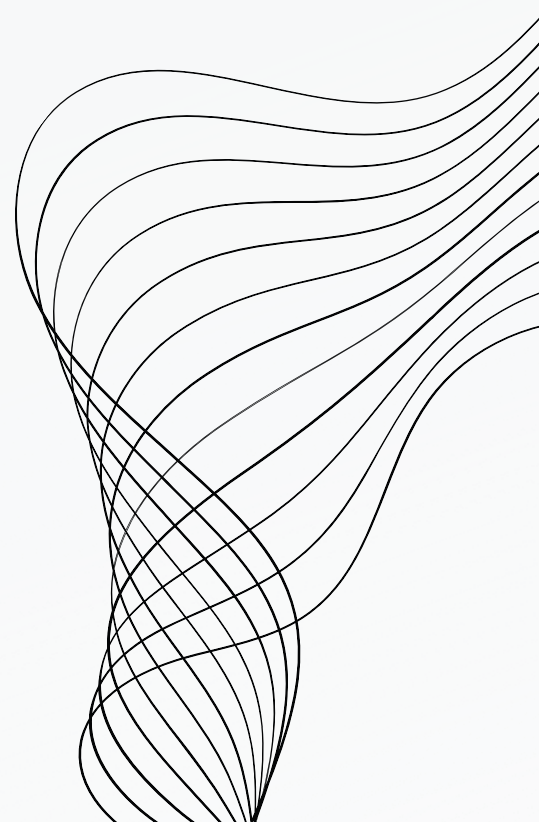
The real estate agency needs to develop a comprehensive advisory framework to assist homeowners in understanding the potential value-adding effects of specific home renovations. This framework should include detailed information on the types of renovations, their estimated costs, and the potential increase in the property's value. By addressing this problem, the agency can empower homeowners to make informed decisions, thereby increasing their satisfaction, trust, and the likelihood of engaging the agency's services.





MAIN OBJECTIVE

The main objective of the real estate agency's initiative is to provide homeowners with accurate and valuable advice regarding home renovations and their potential impact on the estimated value of their properties





SPECIFIC OBJECTIVES

*Develop a Comprehensive Renovation Advisory Framework

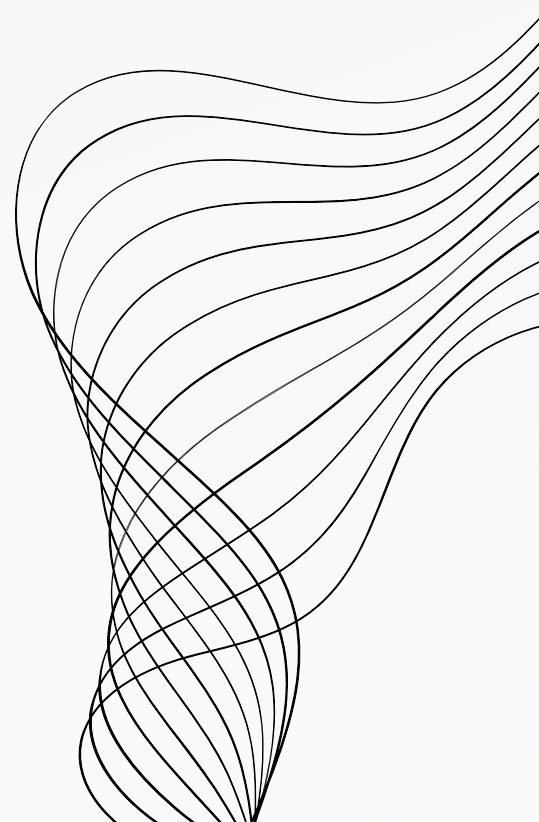
*Track success matrix





DATA UNDERTANDING

In this project, the data was collected from a CSV (Comma-Separated Values) file. The dataset, `kc_house_data.csv`, contains information related to real estate properties in King County, Washington. It includes various attributes that describe the properties, such as their location, size, condition, and sale prices.

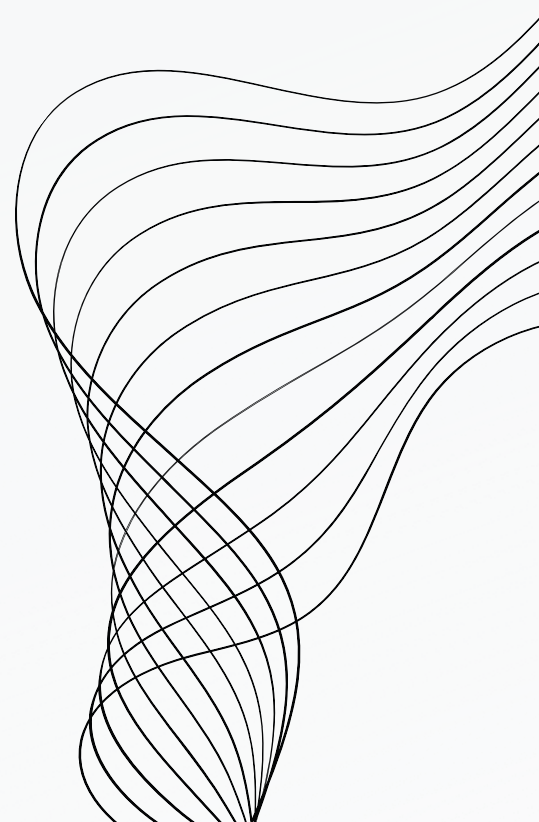




EDA

From our analysis, the condition ratings are converted into a discrete variable.

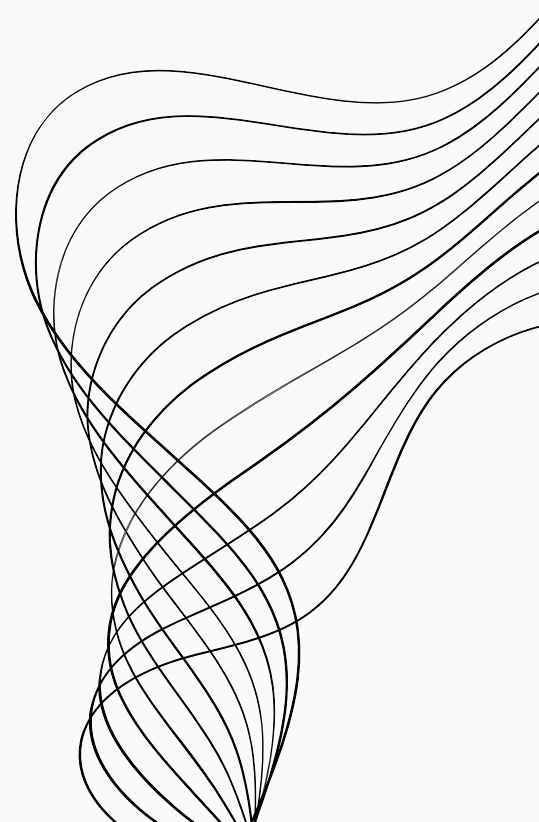
The changes show from 1 to 5, where the lowest has 1 as poor, with the midpoint 3 as average, and the highest 5 as very good

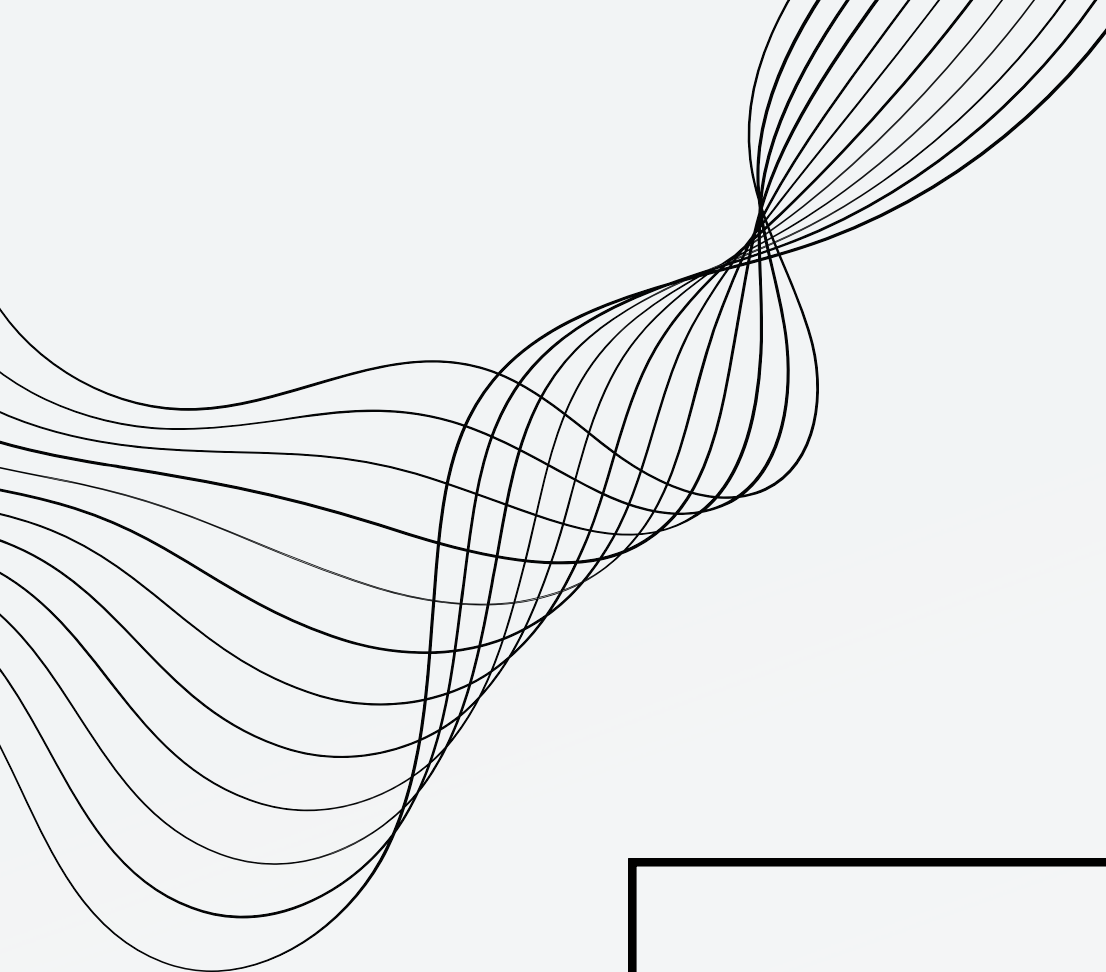




EDA

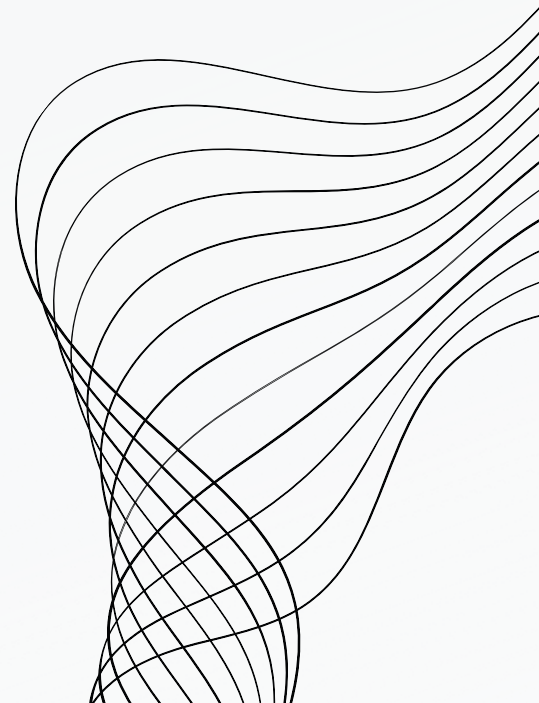
The column 'grade' is split for its values and only the first section is singled out to be converted into numeric type and used as a scale for measuring the grade of the houses / properties Shifting has been done so that the grading parameters may start from 1. From 1 to 11, where the lowest has 1 as poor, with a midpoint of 7 as average and the highest being 11 as mansion





EDA

There is no much difference between the quarters except that the average price in quarter 2 is slightly higher than the rest. There are columns in the dataset that may not be useful in the evaluation and we drop them by initializing the list called dropped and then dropping them from our dataset



1e9 Sum of Prices per quarter



MODELLING

There is a very high correlation between the predictor columns of 'sqft_above', 'sqft_living', 'sqft_living15', 'grade', and 'bathrooms'. Dropping some of these will eliminate multicollinearity features. The columns to drop will be 'sqft_above', 'sqft_living15', 'grade', and 'bathrooms'



1e9

Sum of Prices per quarter



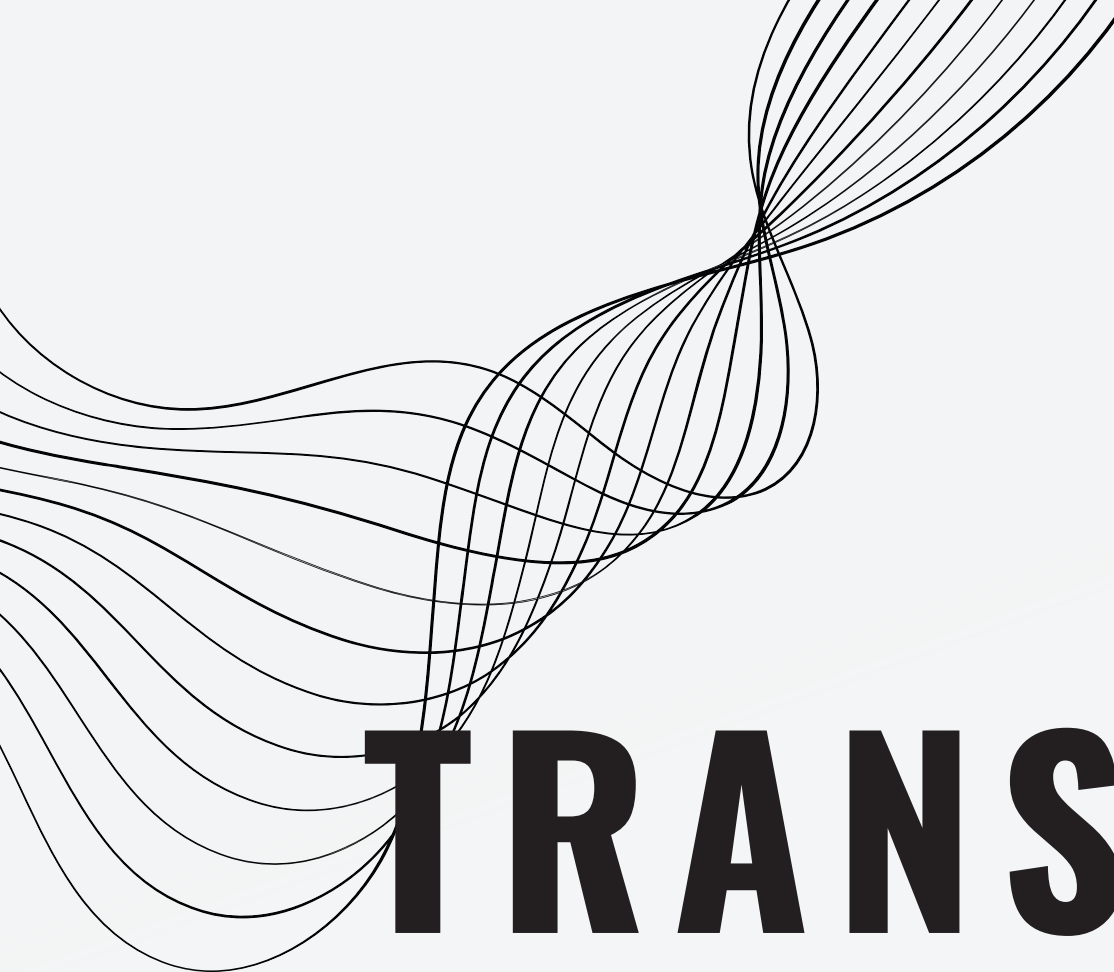
MODELLING

The centred model is statistically significant as standard alpha of 0.05 being 0.0, and still explains 58% of the variance in sale price. The centred model still explains the same amount of variance as the multiple model. The model coefficients are statistically significant except for "sqft_lot". The constant here explains that for an average house that is not renovated, with no waterfront and built since 1900, we would have a sale price of 711,600 dollars.



1e9

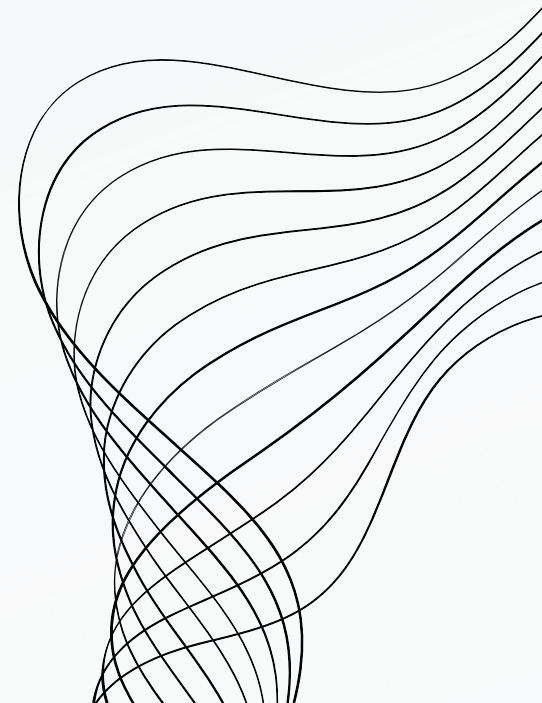
Sum of Prices per quarter



TRANSFORMATIONS

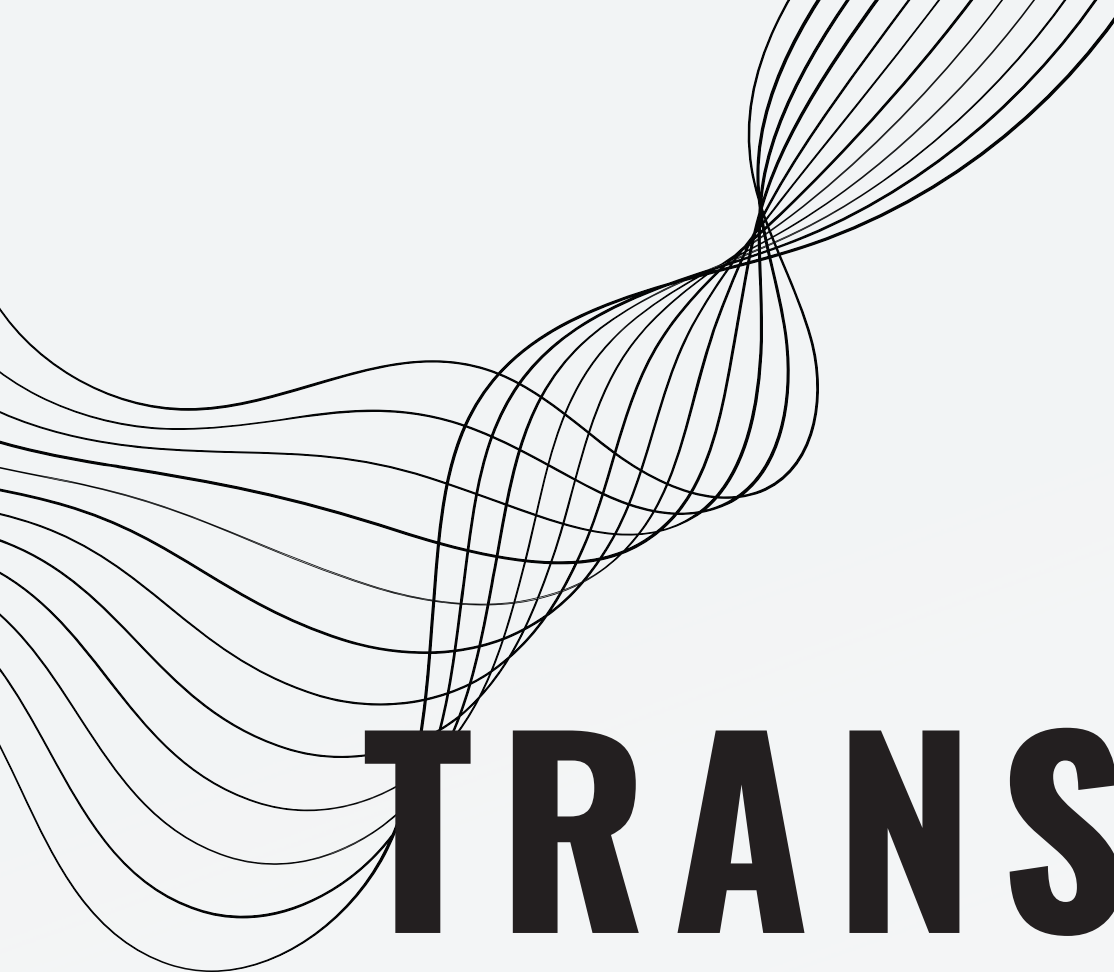
log transformations did not assist in improving improving the model. The adjusted r-squared value remained constant. This led to the need for considering other features affecting the variability in selling price

,



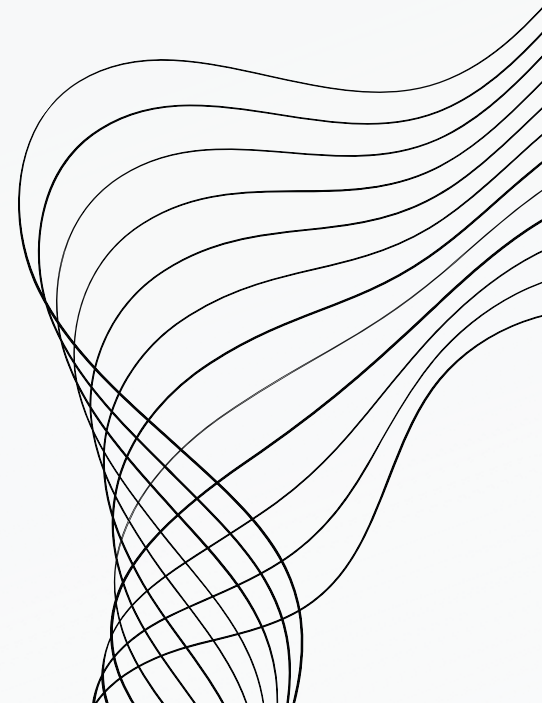
1e9

Sum of Prices per quarter



TRANSFORMATIONS

Standard Errors assume that the covariance matrix of the errors is correctly specified. The condition number is large, $5.47e+06$. This might indicate that there are strong multicollinearity or other numerical problems



1e9

Sum of Prices per quarter



FINDINGS

- * The multiple linear regression model has an R-squared value of 0.79, which indicates that the model can explain 79% of the variance of the market house sale prices which is a good sign that the model is effective in predicting the prices
- * The waterfront view was seen to be most impactful with houses having a waterfront view having value increase of about 834,000 dollars more than those without a waterfront. This can be inferred for the other variables as well



RECOMMENDATIONS

- * We recommend for customers on a budget should look out for houses situated on higher property floors as we expect a price drop on said houses of about 30,000 dollars for every floor increase
- * Development of a comprehensive database the agency can use to track the renovation projects that would improve property value.
- * For customers who would like to sell their properties/houses, it is recommended that they should renovate them first as we see these would be an increase in value by about 48,000 dollars

,



1e9

Sum of Prices per quarter





CONCLUSION

- * The combination of square footage of the living area, bedrooms, floors, and whether the property had a waterfront view or was renovated are the most reliable predictors of a house's price in King County.
- * There were some limitations to the model. To meet regression assumptions, we had to try out log-transformation on certain variables. Therefore, any new data used with the model would require similar preprocessing. Additionally, since housing prices vary regionally, the model's usefulness for data from other counties may be restricted.
- * If you are seeking affordable housing, it may be advisable to compromise on square footage and have no waterfront view.
But, given that many urban residents already do this, it may not be a viable solution for everyone



NEXT STEPS

- * Using datasets from other counties to be able to better advise our customers from comparing the dataset results.
 - * The agency may prepare questionnaires to identify their strengths, weaknesses, opportunities and threats and use this information to prioritize recommendations that would help address their weaknesses and take advantage of their opportunities and strengths.
- * It is also important for the agency to continuously monitor the effectiveness of the strategies they implement and make adjustments as necessary. This could involve tracking metrics like website traffic, this model, social media engagement, and lead generation to assess the impact of their efforts and identify areas for improvement.
 - * More research is required to have a more integrated and informative dataset finding more factors that influence the price

