



shutterstock.com · 1998388166

Project Title: Value analysis for house sale prices in King County

Will your home renovations pay off?

Business overview

- This project aims to provide valuable insights for a real estate agency operating in King County, Washington, USA. Specifically, the agency seeks to provide accurate advice to homeowners on how home renovations can potentially increase the estimated value of their properties and homes, and by how much. This information will help the agency guide their clients towards making informed decisions on home renovations, which can maximize their return on investment when selling their properties.

Introduction

The goal of this project is to conduct extensive research on the real estate market in King County and determine the optimal price range for houses in different neighborhoods on behalf of a real estate agency.

To achieve this objective, we will employ multiple linear regression modeling to analyze house sales data in the King County area. By using statistical techniques, we aim to identify key factors that impact property sales in the region and provide valuable insights to guide our recommendations.

Additionally, we will explore the potential of remodeling specific areas of a property to increase its value and make it more attractive to potential buyers. This approach can help homeowners to add functionality and beauty to their property while simultaneously boosting its resale value.

Problem statement

The problem at hand is to provide homeowners with accurate advice on how specific home renovations can impact the estimated value of their properties and homes, and the amount by which it can increase. This information is crucial for the real estate agency to guide their clients towards making informed decisions on home renovations, which, in turn, can help homeowners

to maximize their return on investment when selling their properties. Therefore, the primary objective of this project is to analyze the impact of home renovations on the estimated value of properties and provide recommendations that can help the real estate agency and their clients to make sound investment decisions.

Data Understanding

- The King County House Sales dataset, available in "kc_house_data.csv," is the primary data source for this project. However, one of the main challenges we may encounter is the ambiguity or incompleteness of the column names in the dataset. Nonetheless, with thorough research and careful judgment, we can extract the necessary insights to make informed decisions about which variables to use in our analysis.

Another challenge we face is ensuring that the multiple linear regression model we develop adds value to our analysis and provides actionable insights that go beyond meeting the project requirements.

The dataset contains information on house sales in King County, including the price, design, square footage, location, and more. A comprehensive list of the column names can be found in the "[Property Schema](#)".

We will also explore the general areas in which renovations are typically undertaken in properties to identify the potential impact on the estimated value of a property.

Objectives

- - To determine the increase in prices of property due to renovations.
- To define a clear stakeholder and business problem that relates to the King County House Sales dataset, and to use this problem to guide the analysis and modeling process.
- To build and evaluate multiple linear regression models using various combinations of the available features in the dataset, with the aim of identifying the most relevant and impactful features for predicting house sale prices.
- To use appropriate statistical techniques to refine and optimize the regression models, and to clearly explain the rationale behind each technique used.
- To demonstrate an iterative approach to modeling, documenting each stage of the process and providing justification for each model modification .
- To provide a final documentation that includes a detailed analysis of the chosen model, including relevant metrics describing overall model performance and the coefficients of the most impactful features, and a clear explanation of how the model adds value to the stakeholder and the appropriate recommendations.

Metric of Success

The final step in evaluating the quality of the model is cross-validation, which gives us an idea of how the model would perform with new data for the same variables. One that the model will be trained on, and another that it will be tested on. By default, the function takes 75% of the data as the training subset and the other 25% as its test subset.

Findings/Results

- * The multiple linear regression model built 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.

- * 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

- * The most impactful features for inferring and predicting house sale prices were found to be the square footage of the living area, bedrooms, floors, and whether the property had a waterfront view or was renovated.

- * The waterfront view proved 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.

- * Our model is off by about 100,000 dollars for each prediction of the price

-

Recommendations

- * 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 see an increase in value by about 48,000 dollars.

- * 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

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.

However, there are 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.

In summary, 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

More research is required to have a more integrated and informative dataset finding more factors that influence the price.

More time would be required to finetune our findings and model results.

Using datasets from other counties to be able to better advice our customers from comparing the dataset results.

The agency may have a questionnaire 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

Student name:

**Joel Omondi,
Aaron Onserio,
James Mbeti,
Victoria Nabea,
Nyokabi Waiganjo,
Kennedy Juma**