Thenyathuita /
Kings-House-Price-Prediction-Project

<> **Code**   Issues   Pull requests   Actions   Projects   Wiki   Security   Insights   Set

master ▾   **Kings-House-Price-Prediction-Project**   Go to file   t   Add file ▾   ...

Thenyathuita   Phase 2 Group 3 Project   10 minutes ago

This branch is 2 commits ahead of, 3 commits behind `main`.

Contribute ▾

Name	Name	Last commit date
📁 Group3 Project	Phase 2 Group 3 Project	10 minutes ago
📄 README.md	Create README.md	3 minutes ago

README.md

PROJECT OVERVIEW The primary objective of this project is to assess a data set containing diverse information about residential properties. This data set encompasses a range of crucial variables, each offering insights into distinct aspects of individual houses. By scrutinizing these variables, we can obtain valuable knowledge that serves multiple purposes, such as comprehending real estate market dynamics, making forecasts regarding housing prices, or aiding decision-making processes for both homeowners who plan to do renovations and how it may affect the price of the house if they choose to sell and to a buyer who plans to buy houses that have undergone renovations. ** BUSINESS AND DATA UNDERSTANDING**

BUSINESS; The core of this project lies in gaining a profound understanding of the real estate market and providing valuable insights to various stakeholders. For this project, we will be mainly focusing on homeowners who wish to know how renovations will affect the price of their house(s). **DATA** The King County Housing Data Set contains information about the size, location, condition, and other features of houses in Washington's King County. The data set and variable descriptions can be found on Kaggle.

MODELLING After exploring and preprocessing the data, simple and multiple linear regression models were built in OLS statsmodels, with price as the dependent variable.

RESULTS Together, square footage, grade, bedrooms, and bathrooms are the best predictors of a house's price in King County. These features were included in the final multiple regression model. The model satisfied all multiple regression assumptions and p-values for each predictor variable were below .05. The r-squared value of the model was .489.

Variable Correlations
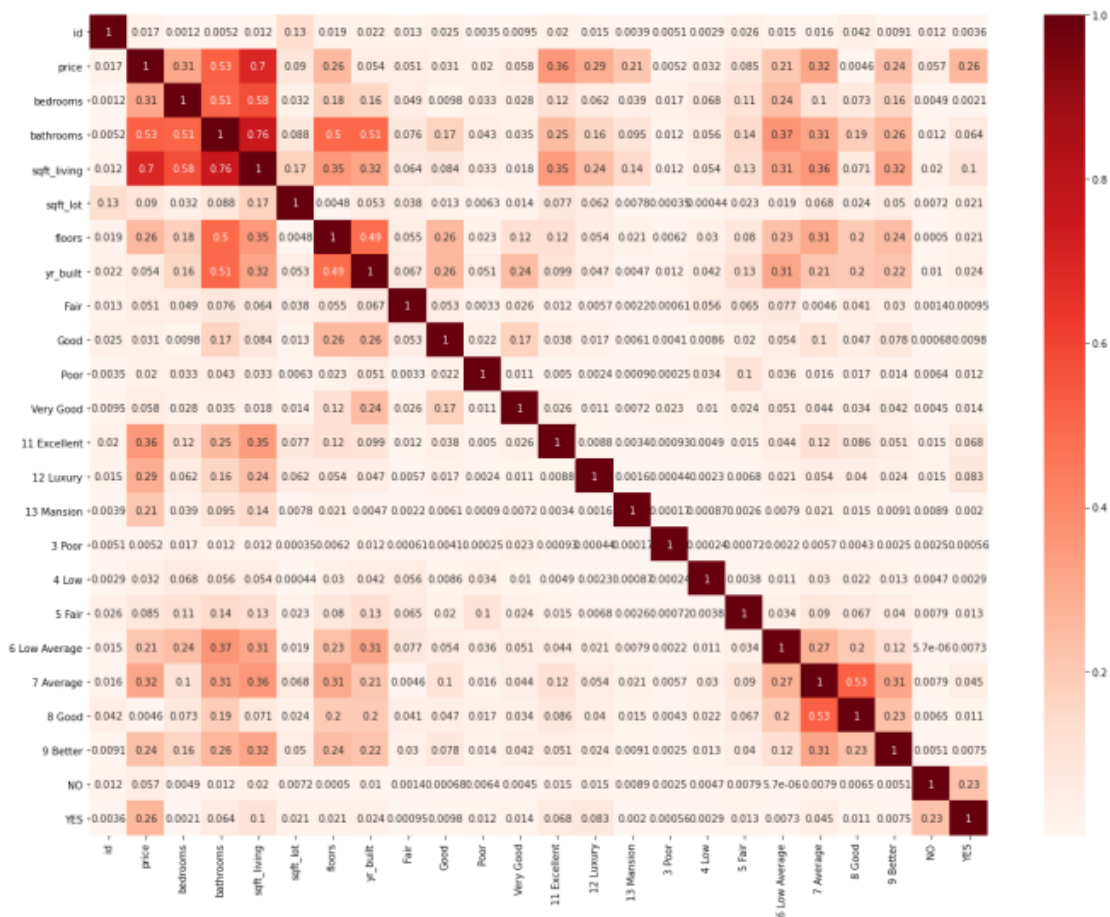
The four used in the modeling process were chosen due to their strong correlation with the price column.

CONCLUSION The best price predictor of house price during renovations with the intent to sell in King's County is sqft_living and grade. As homeowners plan to renovate their house and add more square footage of space they must also make sure their building grade is as close to the best as possible to increase the value of one's house. The model has its limitations as the need to log-transform one of the variables to satisfy the regression assumptions, any new data input needs to undergo the same preprocessing. Without the use of regional data, it must be remembered that house values may differ according to their zipcodes Future analysis should explore better predictors of the prices of homes outside King County.