**King’s County Home Sales dataset analysis**

**Project Overview**

This project aims to analyze the housing sales data and provide information on various factors that influence housing prices. The project uses linear regression to infer what factors influence housing prices and by how much.

**Business Problem**

G-One Limited will analyze the King's County home sales dataset to identify key factors impacting housing prices and help our client choose a home with high resale value. With Insights from the analysis you will be guided in making more informed decisions for profitable real estate transactions.

**Data understanding**

We obtained the dataset for our analysis from the CSV file “kc\_house\_data.csv” containing data on over 21,000 housing units including:

* id - Unique identifier for a house
* date - Date house was sold
* price - Sale price (prediction target)
* bedrooms - Number of bedrooms
* bathrooms - Number of bathrooms
* sqft\_living - Square footage of living space in the home
* sqft\_lot - Square footage of the lot
* floors - Number of floors (levels) in house
* waterfront - Whether the house is on a waterfront Includes Duwamish, Elliott Bay, Puget Sound, Lake Union, Ship Canal, Lake Washington, Lake Sammamish, other lake, and river/slough waterfronts
* view - Quality of view from house Includes views of Mt. Rainier, Olympics, Cascades, Territorial, Seattle Skyline, Puget Sound, Lake Washington, Lake Sammamish, small lake / river / creek, and other
* condition - How good the overall condition of the house is. Related to maintenance of house. See the [King County Assessor Website](https://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r#g) for further explanation of each condition code
* grade - Overall grade of the house. Related to the construction and design of the house. See the [King County Assessor Website](https://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r#g) for further explanation of each building grade code
* sqft\_above - Square footage of house apart from basement
* sqft\_basement - Square footage of the basement
* yr\_built - Year when house was built
* yr\_renovated - Year when house was renovated
* zipcode - ZIP Code used by the United States Postal Service
* lat - Latitude coordinate
* long - Longitude coordinate
* sqft\_living15 - The square footage of interior housing living space for the nearest 15 neighbors
* sqft\_lot15 - The square footage of the land lots of the nearest 15 neighbors

**Method**

We prepared our data for analysis by first importing the dataset and the libraries to facilitate analysis, then examining the contents of the dataset and cleaning the data to eliminate any missing values. The analysis deployed in this project uses multiple and simple linear regression methods to develop a model of house price using a subset of predictor variables.

We first developed a baseline model using a simple linear regression method, with price as our independent variable and sqft\_living as our feature variable. Subsequent iterations of the model built on this baseline model until our final model which utilized multiple linear regression models, with the log-transformed price “price\_log” as the independent variable and xxxx as feature variables.

**Results**

We compared the R-squared score at each model iteration and witnessed an improvement in our R-squared from approximately 48% in our initial model to 65% in our final model.

**Conclusions**

The model provided insights into the various features that affect the price of a house in King's County. Notably, waterfront, grade, and view had the highest impact on the price of houses as houses with waterfronts attracted a price increase of 31.51%,

The following features will improve the pricing of the houses:

* A unit increase in square foot living will increase the price of a house by 0.02%
* A unit increase in the number of bathrooms will increase the price of a house by 7.91%
* A unit increase in the number of floors will increase the price of a house by 7.74%
* The higher the grading of a house, the higher it's price. For instance, a house graded as excellent will attract a price increase of 11.94%, while a house graded as luxury will attract a price increase of 21.27%, and mansion a price increase of 22.91%
* The better the condition of a house, the higher it's price. A house in "good" condition will attract a price increase of 1.9% while a house in "very good" condition will attract a price increase of 8.63%
* Houses without views attract lower prices compared to houses with views. The model demonstrates that a house with a good view attracts a price increase of 3.52%, fair view 8.33%, and excellent view 16.55% increase in price
* Houses with a waterfront attract a price increase of 31.51%

**Recommendations**

G-One Limited therefore has the following recommendations for the family to guide their choice of a house in the King's County neighborhood:

They should consider the number of bathrooms

They should consider the number of floors

They should focus on houses graded as excellent, luxury, or mansion

They should focus on houses whose condition are either good or very good

Houses with a good view will attract a higher price compared to ones without

Houses with a waterfront have the highest price value