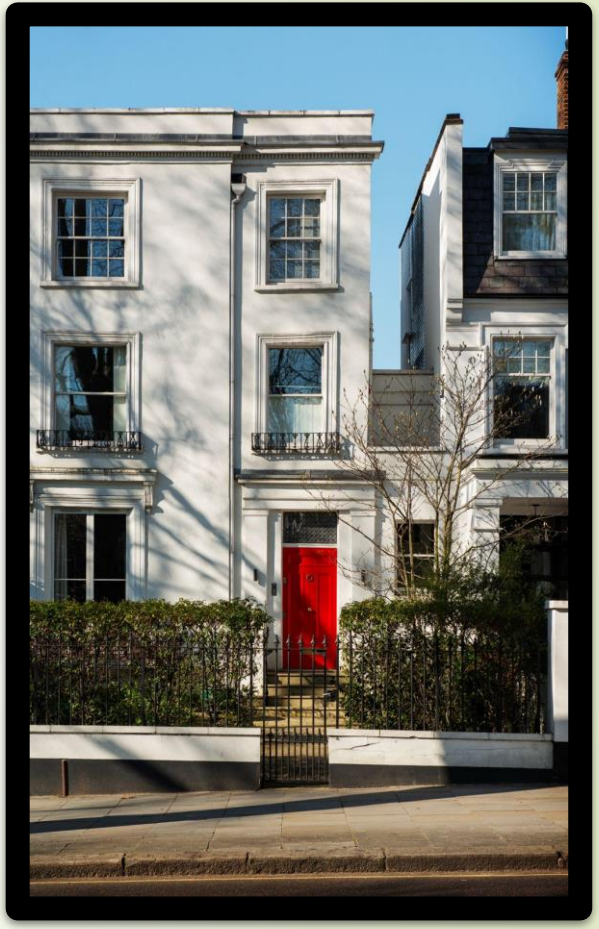
KING COUNTY HOUSING PRICE PREDICTION MODEL

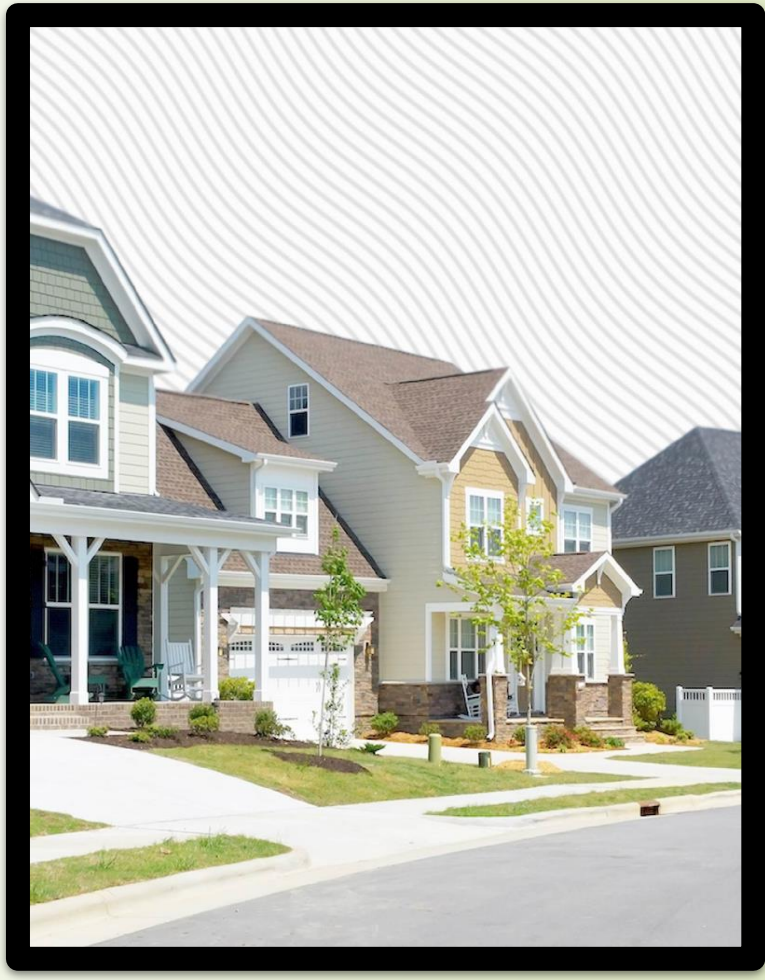
Data Driven solutions for the King County real estate market



OVERVIEW

This project was completed using data for the dataset provided by King County, Washington.

The goal for is to come up with a suitable model for a real estate company in Washington that will be used to predict house prices which will enable them maximize on the profits



Business Understanding

- **A company's real estate agent is curious about the elements that have a major influence on King County home values. This will help in formulating the optimal criterion to follow in order to optimize profit.**

PROBLEM STATEMENT

- Given the complexity of factors that can affect the sale price of a house it is challenging to get a precise assessment

- Data driven analysis can provide key insight to agencies and clientele to help them make informed decisions

Objectives

I. Develop a predictive model for accurate house price estimations
II. Identify and quantify the key determinants that affect house pricing
III. Investigate the impact of waterfront view on price
IV. Evaluate the impact of House condition on price
V. Evaluate the impact of house grade on price

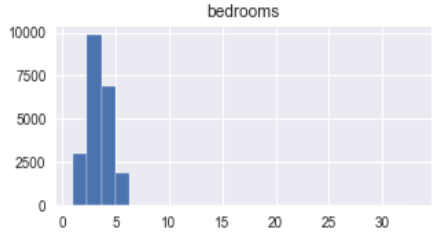
METHODOLOGY AND TECHNIQUES

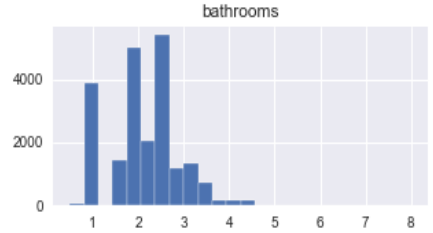
The research used an iterative method to visualize attributes in connection to price, create a multiple linear regression model using Python and Stats Models to predict house sale prices, and discover characteristics that influence King County house prices.

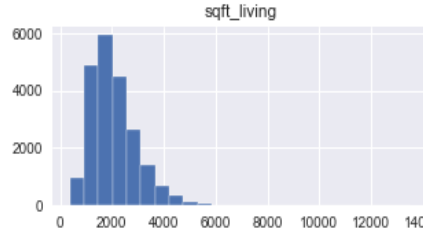
DATA UNDERSTANDING

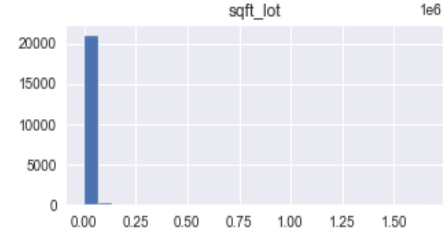
- **There are 21597 records with 20 features in the King County data.**
- **Some null values in the waterfront, view, and year of renovation are present in the data.**
- **The highest-priced home sold for about $800,000, while the average is almost $540,000.**
- **The average house grade is 7, which indicates that the majority o sold homes have grades above average.**
- **In King County, the highest number of floors in a residence is 3.5.**

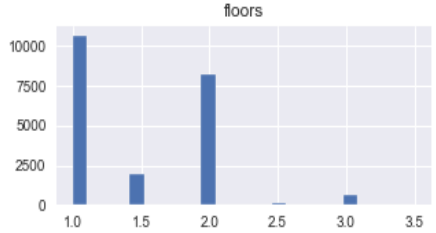
DATA ANALYSIS

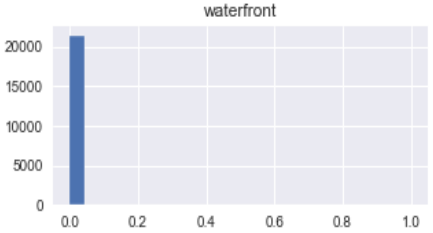


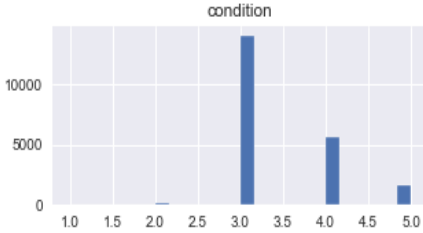


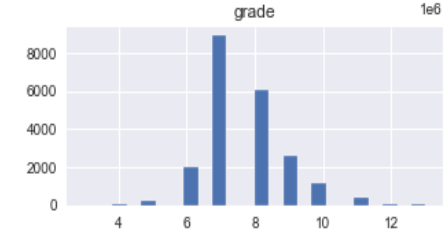


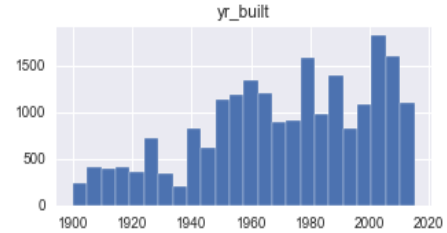


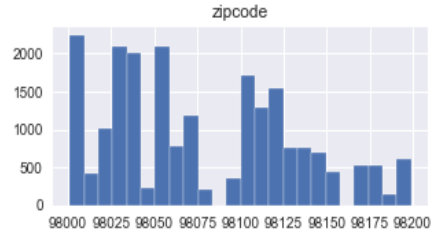


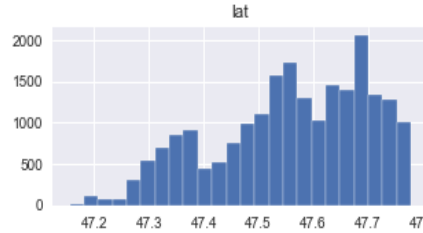


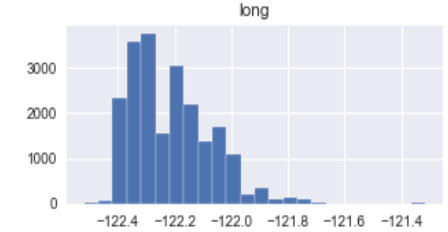


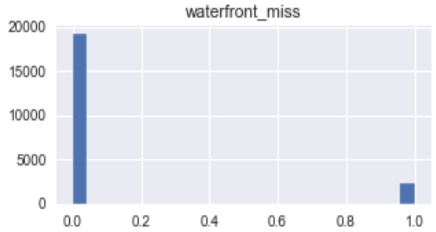


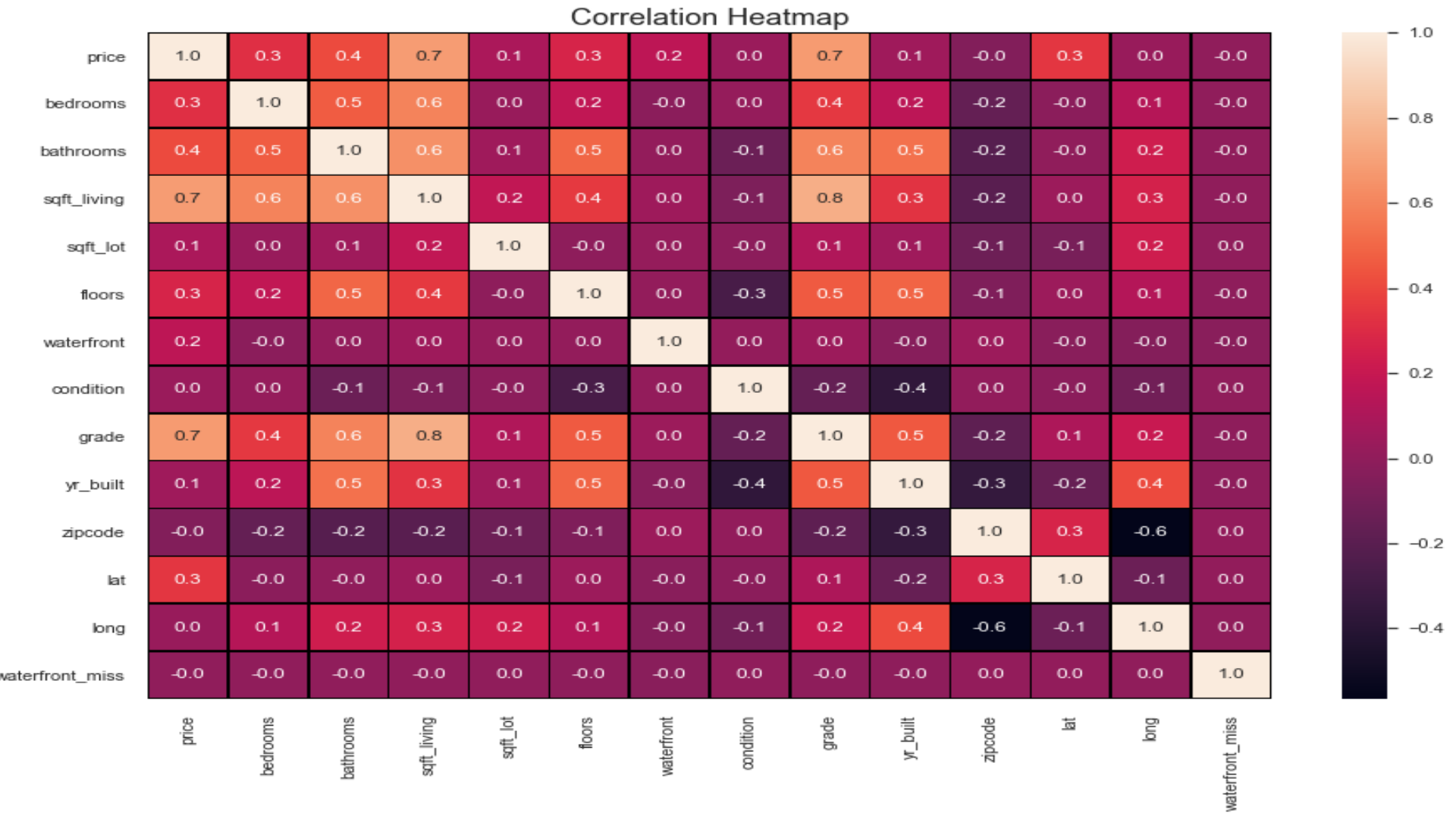


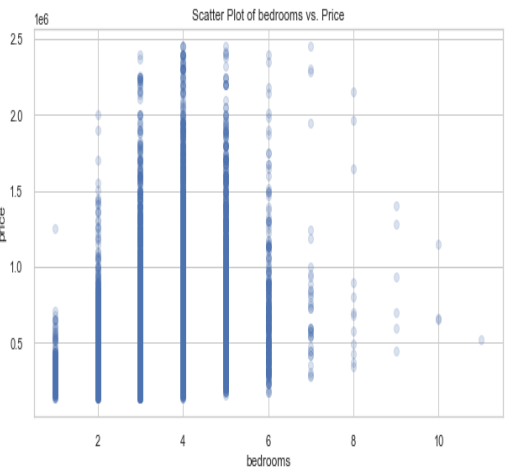


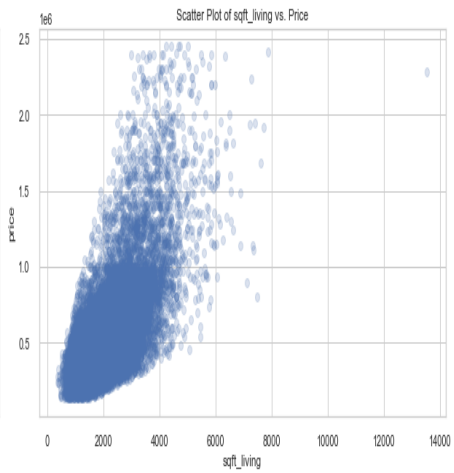


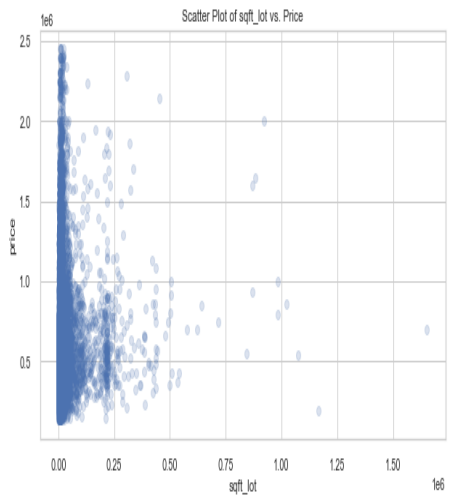


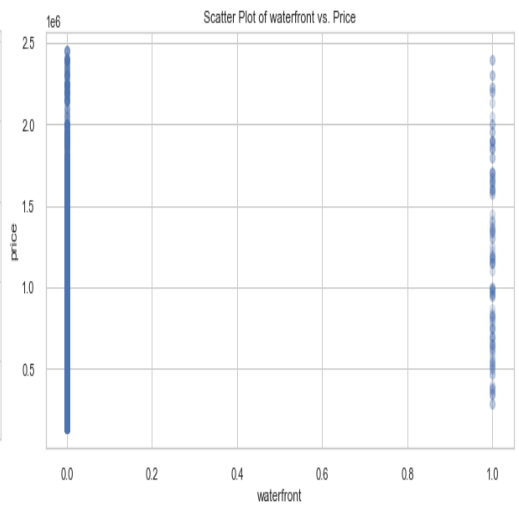


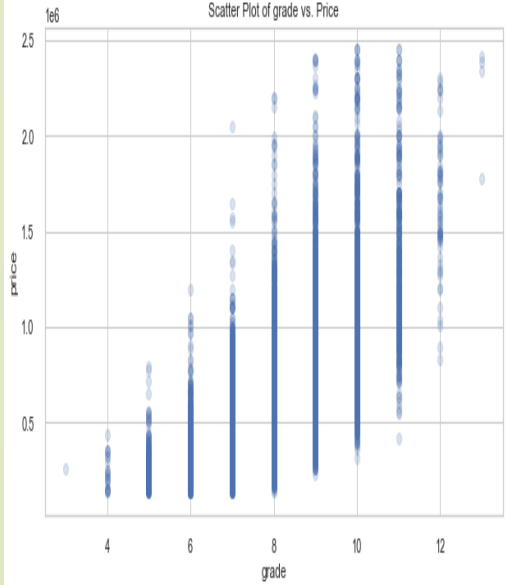

Correlation Heatmap
price 1.0 0.3 0.4 0.7 0.1 0.3 0.2 0.0 0.7 0.1 -0.0 0.3 0.0 -0.0
bedrooms 0.3 1.0 0.5 0.6 0.0 0.2 -0.0 0.0 0.4 0.2 -0.2 -0.0 0.1 -0.0
bathrooms 0.4 0.5 1.0 0.6 0.1 0.5 0.0 -0.1 0.6 0.5 -0.2 -0.0 0.2 -0.0
sqft_living 0.7 0.6 0.6 1.0 0.2 0.4 0.0 -0.1 0.8 0.3 -0.2 0.0 0.3 -0.0
sqft_lot 0.1 0.0 0.1 0.2 1.0 -0.0 0.0 -0.0 0.1 0.1 -0.1 -0.1 0.2 0.0
floors 0.3 0.2 0.5 0.4 -0.0 1.0 0.0 -0.3 0.5 0.5 -0.1 0.0 0.1 -0.0
waterfront 0.2 -0.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 -0.0 0.0 -0.0 -0.0 -0.0
condition 0.0 0.0 -0.1 -0.1 -0.0 -0.3 0.0 1.0 -0.2 -0.4 0.0 -0.0 -0.1 0.0
grade 0.7 0.4 0.6 0.8 0.1 0.5 0.0 -0.2 1.0 0.5 -0.2 0.1 0.2 -0.0
yr_built 0.1 0.2 0.5 0.3 0.1 0.5 -0.0 -0.4 0.5 1.0 -0.3 -0.2 0.4 -0.0
zipcode -0.0 -0.2 -0.2 -0.2 -0.1 -0.1 0.0 0.0 -0.2 -0.3 1.0 0.3 -0.6 0.0
lat 0.3 -0.0 -0.0 0.0 -0.1 0.0 -0.0 -0.0 0.1 -0.2 0.3 1.0 -0.1 0.0
long 0.0 0.1 0.2 0.3 0.2 0.1 -0.0 -0.1 0.2 0.4 -0.6 -0.1 1.0 0.0
waterfront_miss -0.0 -0.0 -0.0 -0.0 0.0 -0.0 -0.0 0.0 -0.0 -0.0 0.0 0.0 0.0 1.0
price bedrooms bathrooms sqft_living sqft_lot floors waterfront condition grade yr_built zipcode lat long waterfront_miss
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4

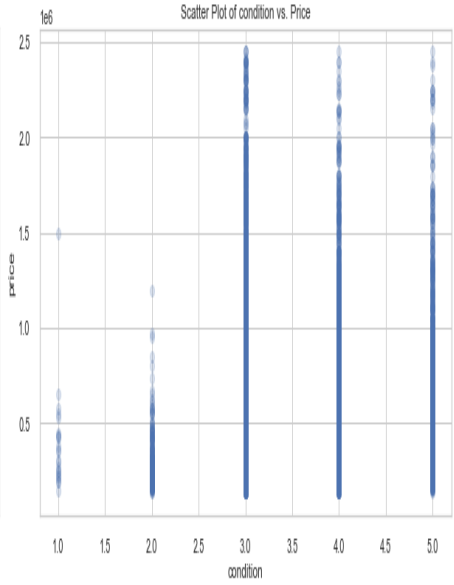

Scatter Plot of bedrooms vs. Price
1e6
price
bedrooms

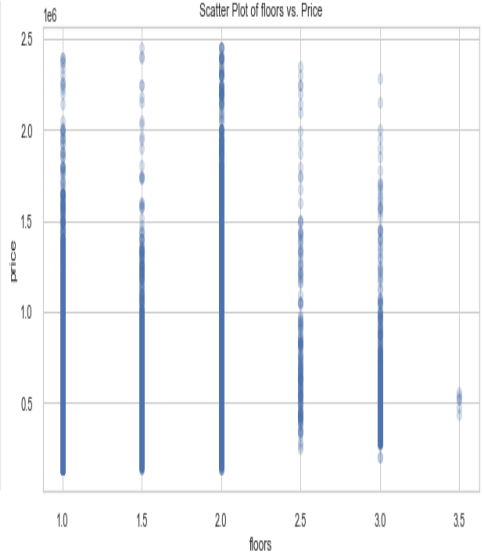

Scatter Plot of sqft_living vs. Price
1e6
price
sqft_living

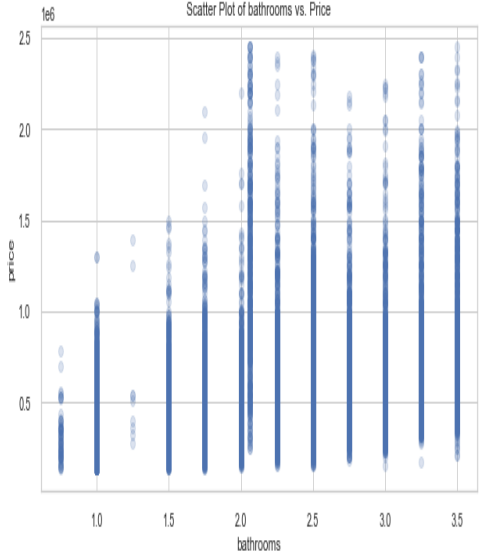

Scatter Plot of sqft_lot vs. Price
1e6
price
sqft_lot
1e6


Scatter Plot of waterfront vs. Price
1e6
price
waterfront

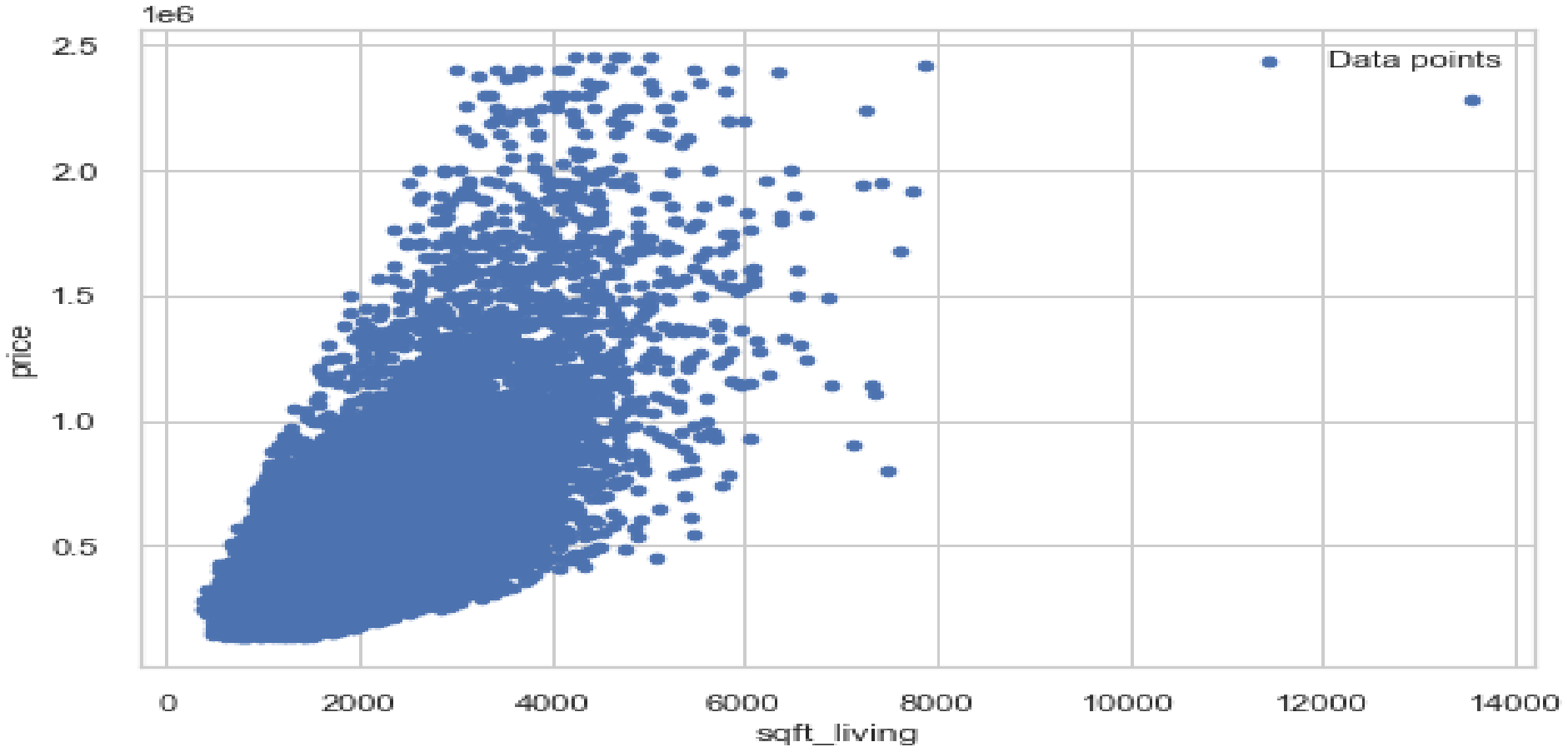

Scatter Plot of grade vs. Price
1e6
price
grade

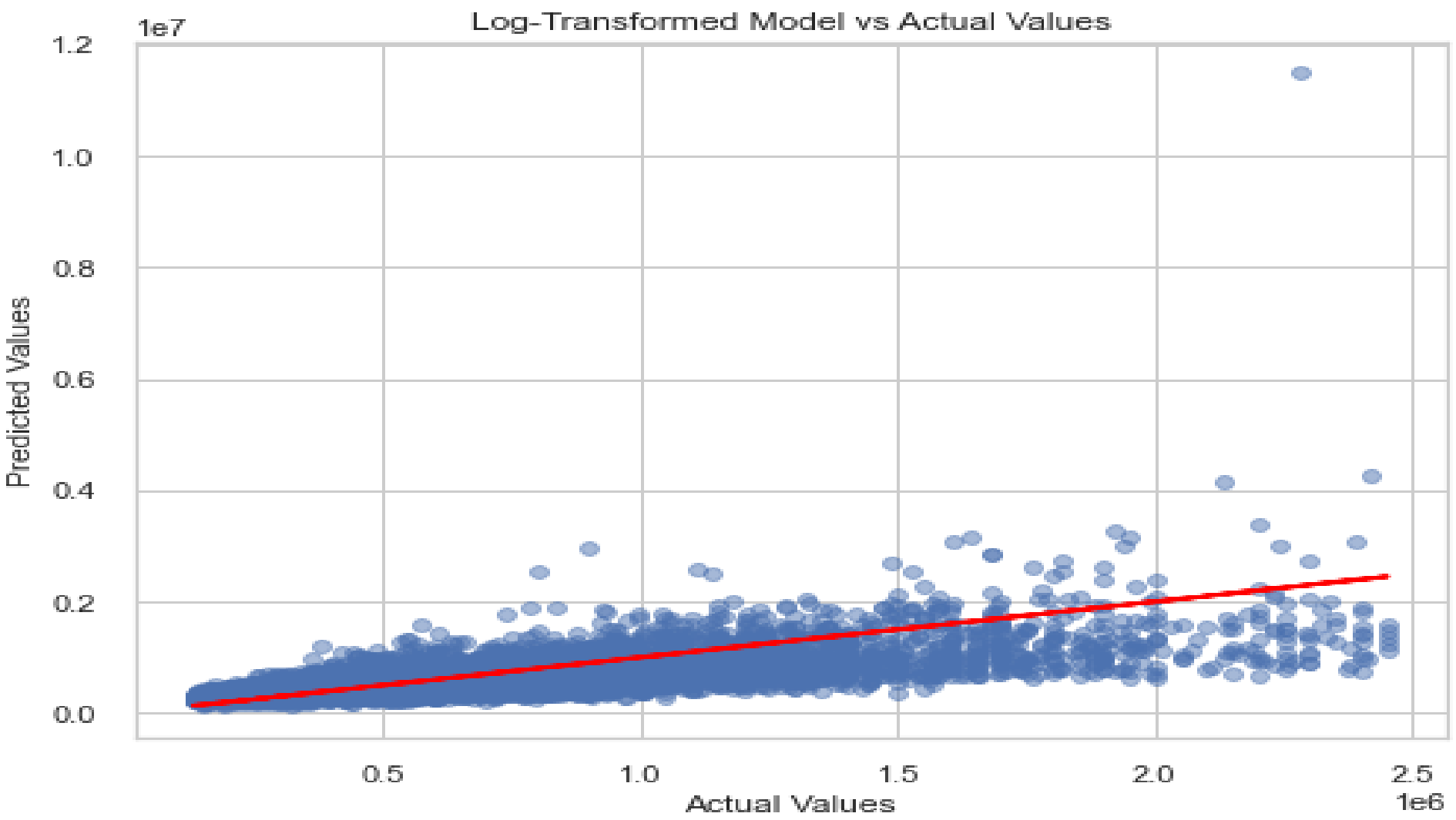

Scatter Plot of condition vs. Price
1e6
price
condition


Scatter Plot of floors vs. Price
1e6
price
floors

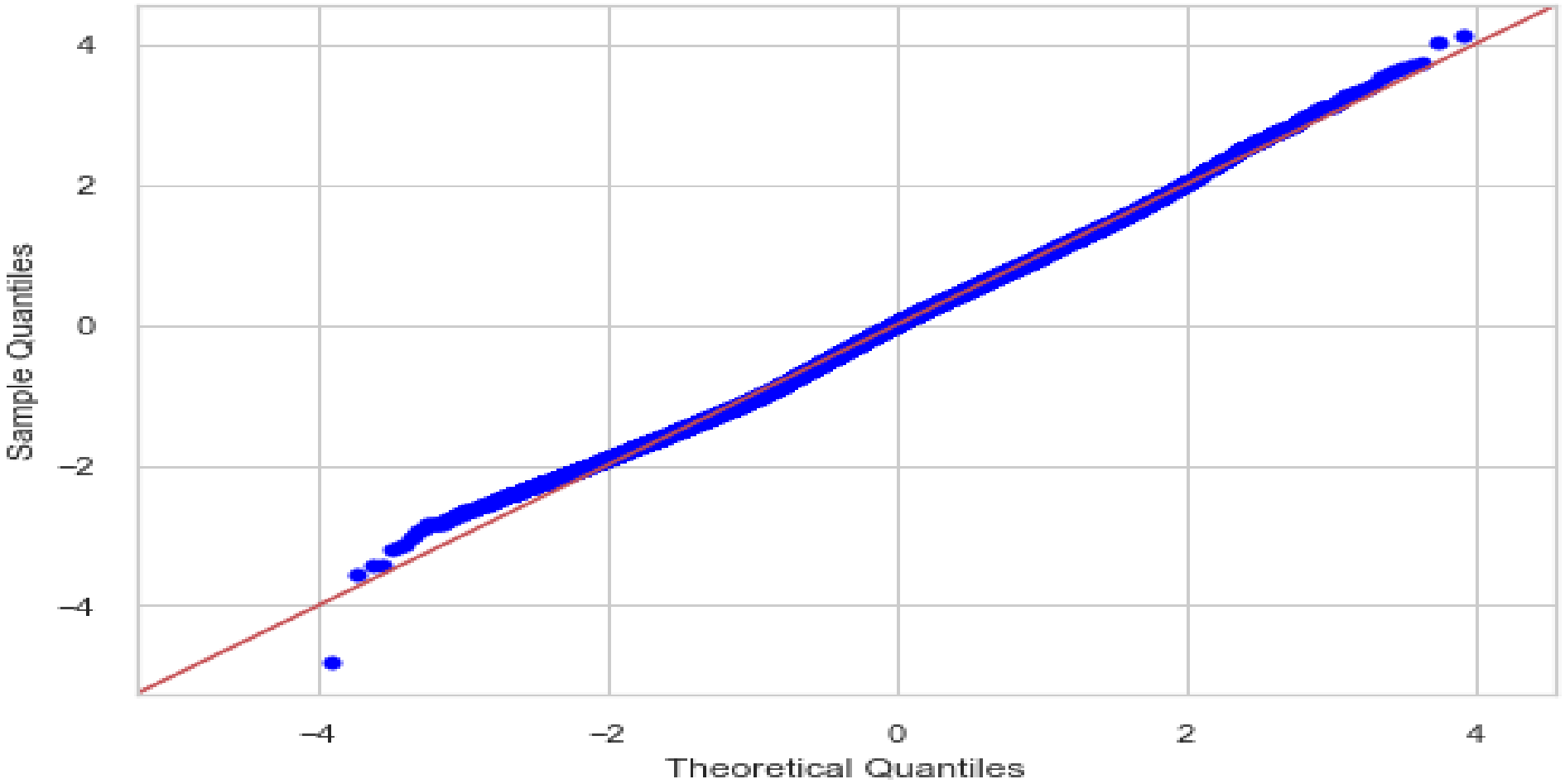

Scatter Plot of bathrooms vs. Price
1e6
price
bathrooms

Baseline Modelling




Log-Transformed Model vs Actual Values
1e7
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Predicted Values
0.5
1.0
1.5
2.0
2.5
1e6
Actual Values

FINAL MODEL QQ-PLOT



CONCLUSION

- The key factors are:
  - Waterfront location
  - Living area square footage
  - Overall grade of the property

- A house with the typical features for the area retails at $528969

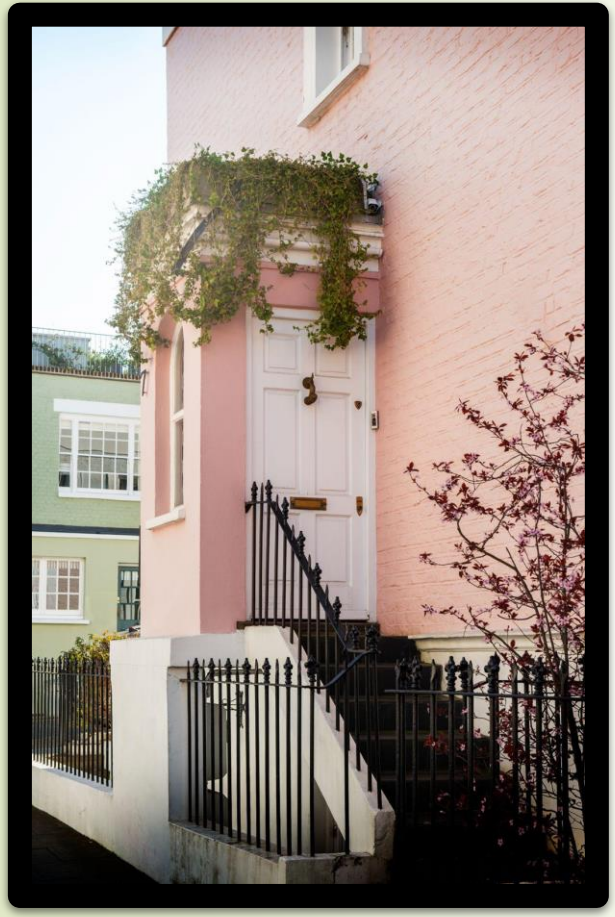
- The data shows most houses of average condition which shows an investment potential for our stakeholders

RECOMMENDATIONS

- Different pricing strategies to emphasize Waterfront properties
- Investment in property upgrades
- Highlight spacious living areas
- Emphasis on Property Grade

NEXT STEPS

- Comparison with other regression models
- Utilize supplementary data categories from the Kings County to improve the model's precision
- Use of current and long term data to understand trends in the market.



Thank You