

HOUSE SALES PRICING

By:
John Mungai

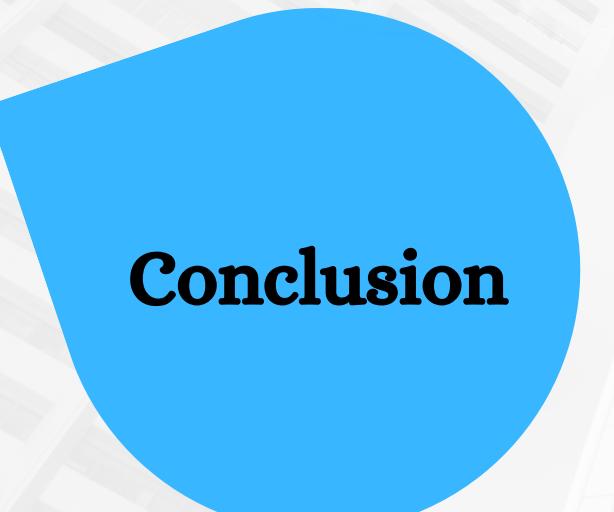
1



Overview

In this analysis, I used various regression methods to predict the house prices. The reason for choosing regression for the analysis is because it gives more insight about the data. In regression we look at the relation between response and explanatory variables. I will start with simple regression model then complex ones and compare the best model for the dataset.

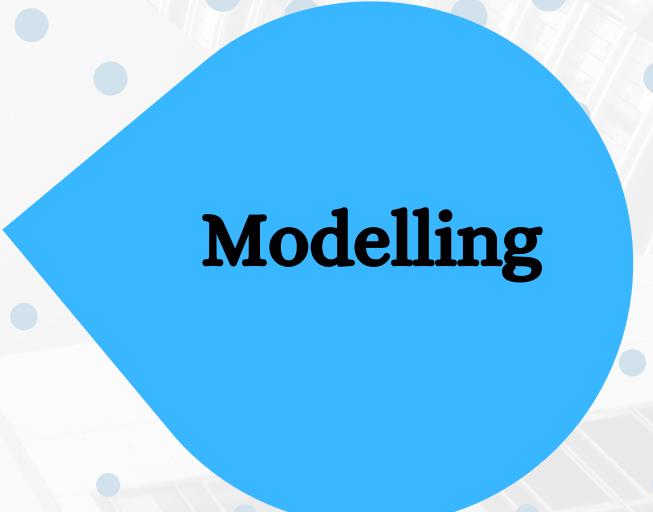




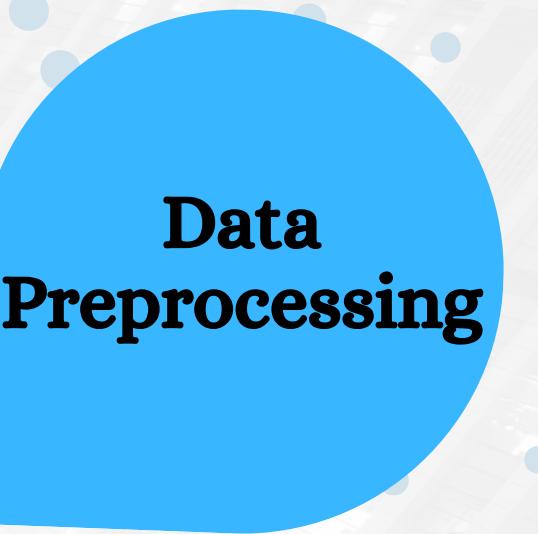
Conclusion



Evaluation



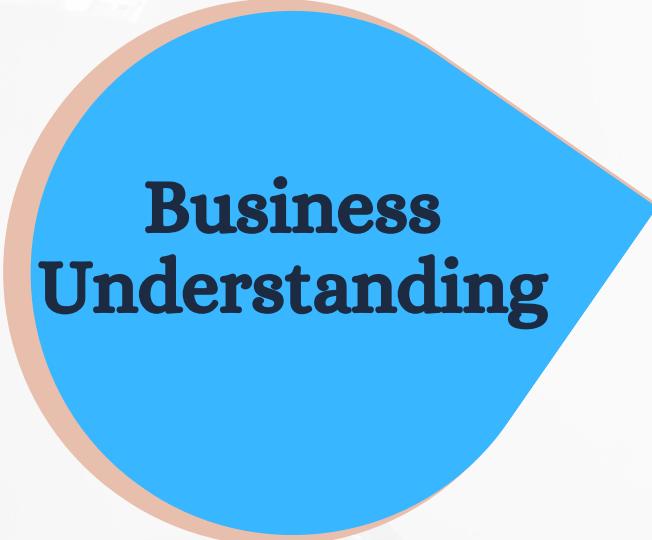
Modelling



Data
Preprocessing



Recommendation



Business
Understanding



Data
Understanding



Data
Preparation



Data
Exploration

Business Understanding

- In the housing market, the understanding of factors that affect house prices is fundamental for the business.
- For real estate in King County, Washington in the United States to advice there clients (home buyers and the home sellers), they need to look at the factors affecting the price of the homes.
- We also provide advice to homeowners about how home renovations might increase the estimated value of their homes, and by what amount.



Data Understanding

- House Sales in King County, USA dataset from Kaggle was used for the analysis.
- The dataset includes homes sold between May 2014 and May 2015.
- It contains 21 columns including price.





Problem Statement

- Identify the best model to fit for predicting the price of the house.
- Identify home renovations might increase the estimated value of their homes, and by what amount.



DATA PREPARATION

Loading Data
Importing libraries, reading and loading of datasets

1

2

3

4

Train-test split

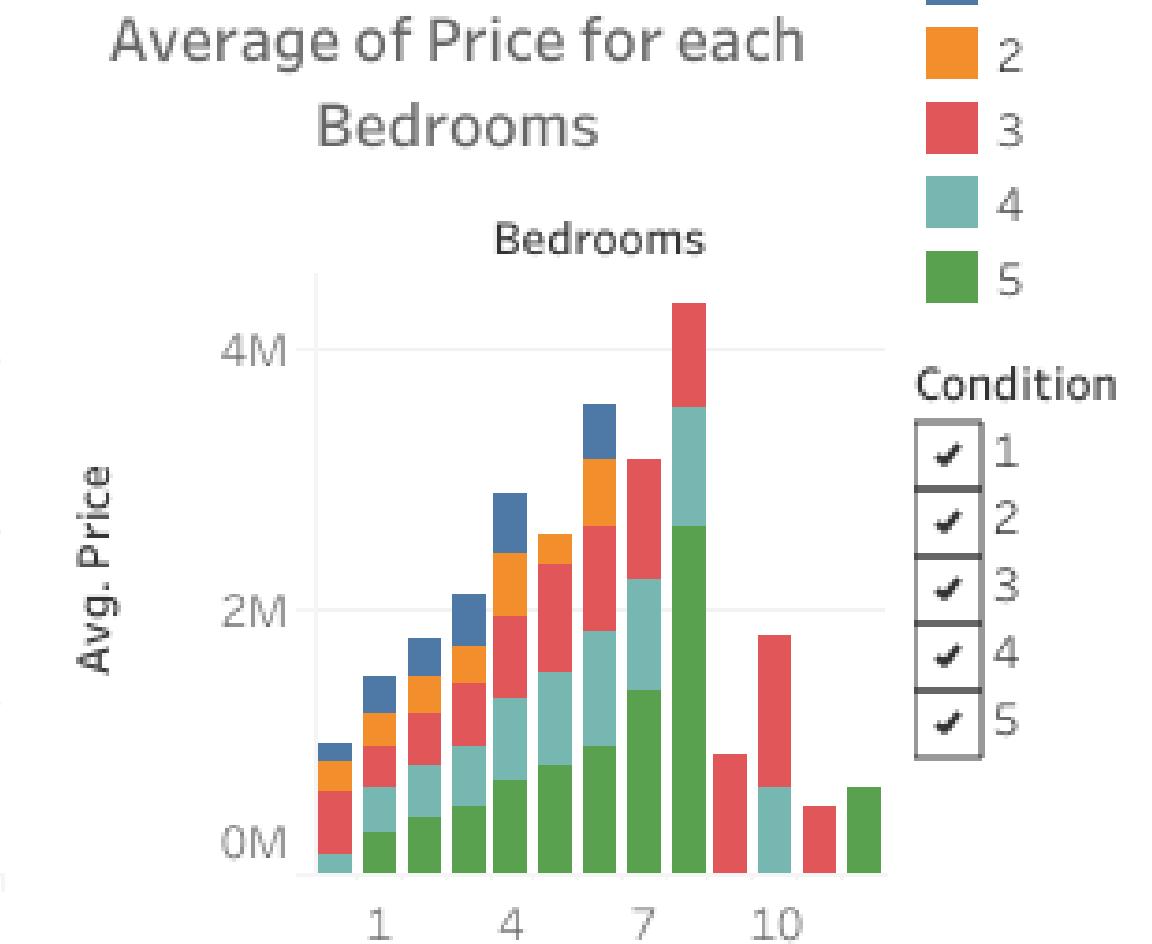
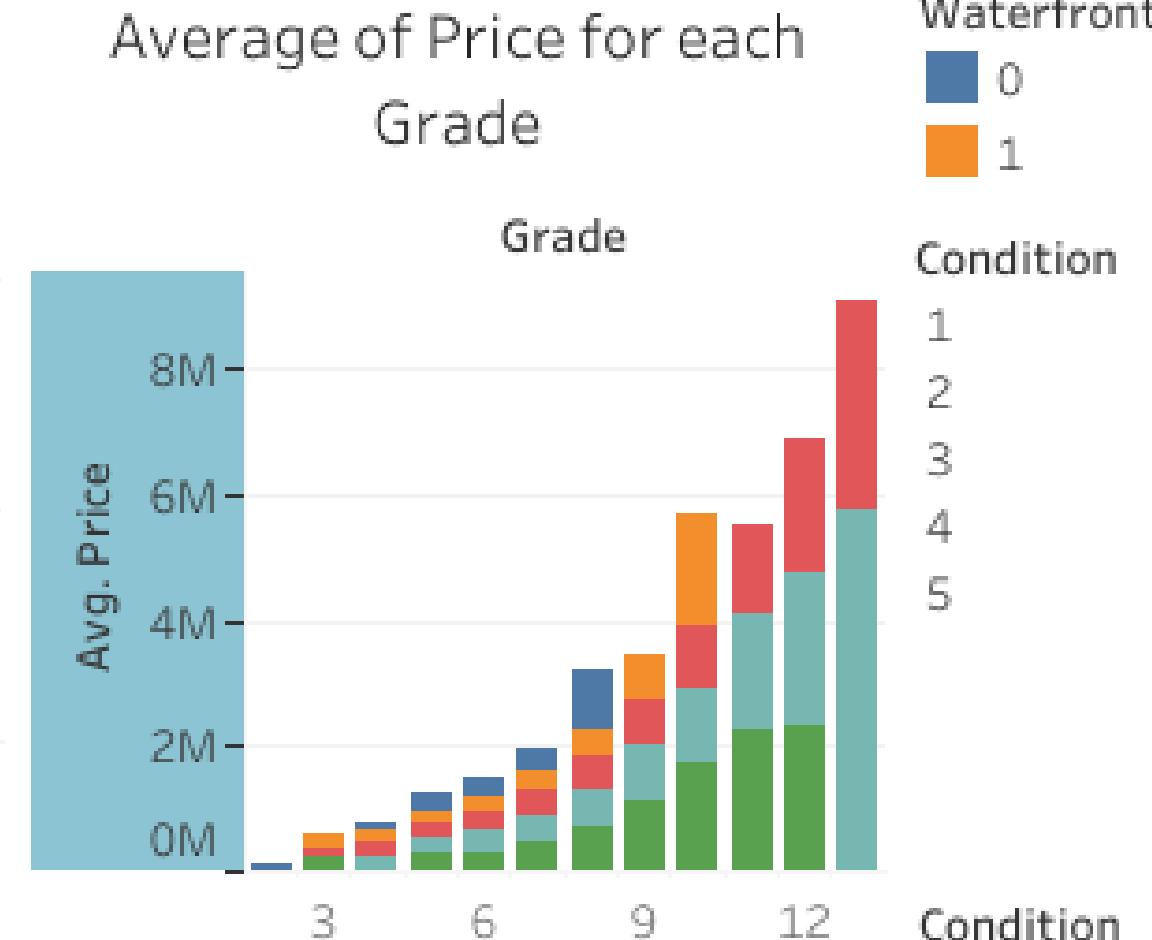
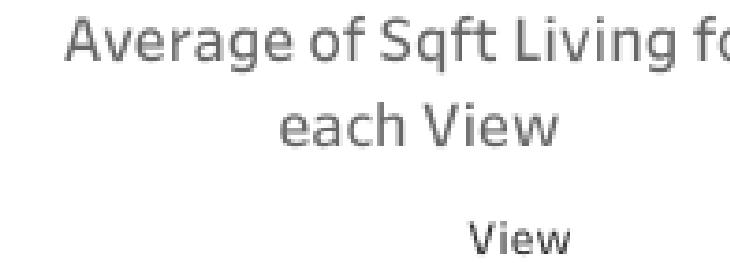
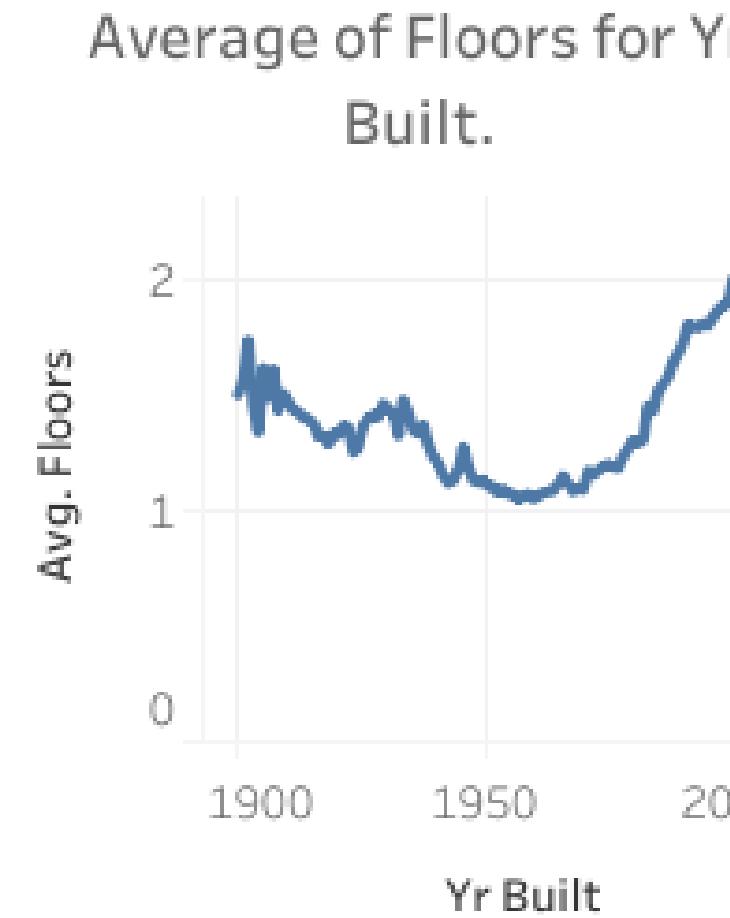
Data Cleaning
Removing null and duplicate values and checking for outliers

Data Analysis
Visualizing how different variables relate to each other and to the target variable

7

VISUALIZATION OF VARIABLES RELATIONSHIP

8



Waterfront
0
1

Condition
1
2
3
4
5

Condition
1
2
3
4
5

Condition
1
2
3
4
5

Modelling

Out of the three models, polynomial model is the best fit model, with 2 degrees and all features. The model explains about 83% of the variance in price.

Model	Adjusted R	MSE
<i>Simple Linear</i>	0.489	0.715
<i>Multiple Linear</i>	0.726	0.523
Polynomial	0.83	0.403

Modelling

Relationship between year home was renovation and the price

Price = 532600 + 115*yr_renovated

Conclusions

- Polynomial model is the best model to fit for predicting the price of houses.



Thank You

12

