

King County House Sales Prediction Project

Linear Regression Modelling



01 Introduction

Project Overview

02 Project Objectives

Project Objectives

03 Data Understanding

Explaining data source and variables

04 EDA and Modelling

Exploratory Data Analysis and Linear Regression Model

05 Conclusions and Recommendations

Drawing conclusions and recommendations from the linear regression model

CONTENTS


INTRODUCTION

Problem Statement:

- *Real estate prices are sophisticatedly linked to our economy*
- *Despite this, we do not have accurate measure of house prices based on the vast amount of data available.*
- *Proper and justified process of predicting prices properties can bring a lot of transparency and trust back to Kings County real estate.*
- *King County situated in the U.S. state of Washington boasts a vibrant real estate market. This project conducts a comprehensive exploration of King County's housing landscape, aiming to unravel the factors that influence property prices. We will use data analytics and predictive modelling techniques to provide homeowners, buyers, and investors with valuable insights for making informed decisions in this thriving real estate market.*



PROJECT OBJECTIVES

- *Analyzing the relationship between house prices and factors such as square footage, number of bedrooms and bathrooms, location, waterfront and view, condition and renovation status.*
 - *Identifying the most important factors that affect house prices.*
 - *Developing predictive models that can be used to estimate the future price of a house.*
 - *We believe that this project will provide valuable insights into the King County housing market and help people make informed decisions about their real estate investments.*
- 

DATA UNDERSTANDING

- *Data source is the King County property appraiser website.*
- *The variables are as explained in the attached ReadMe file.*
- *To allow for regression analysis we encoded the categorical variables*



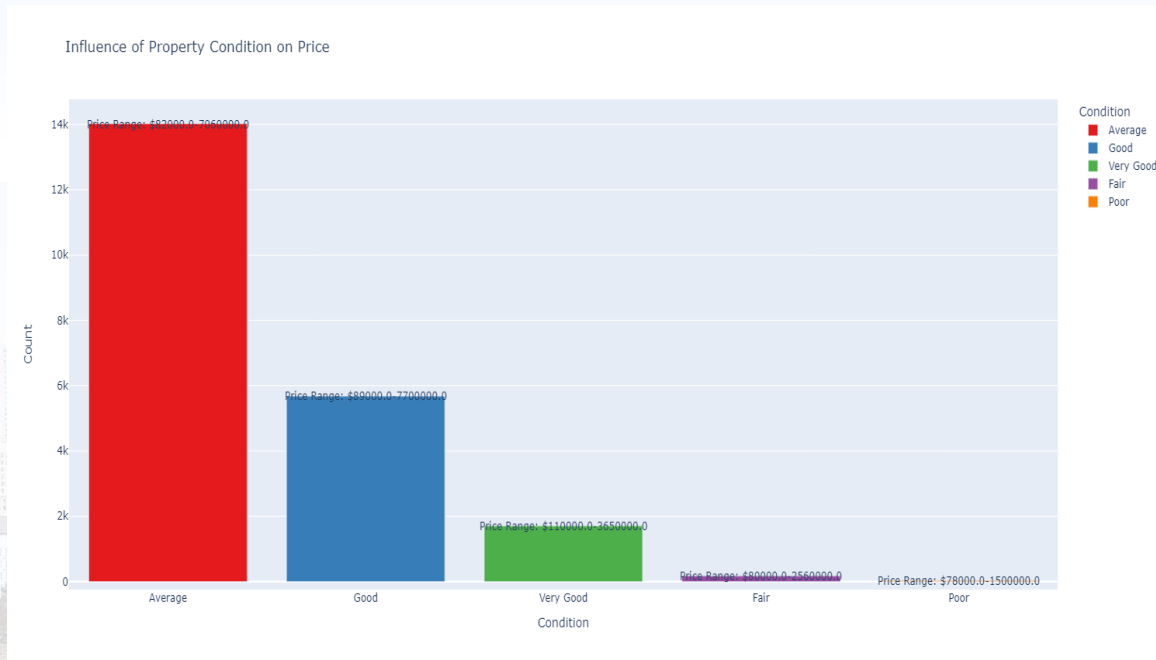
04

Data analysis and Modelling



Data analysis

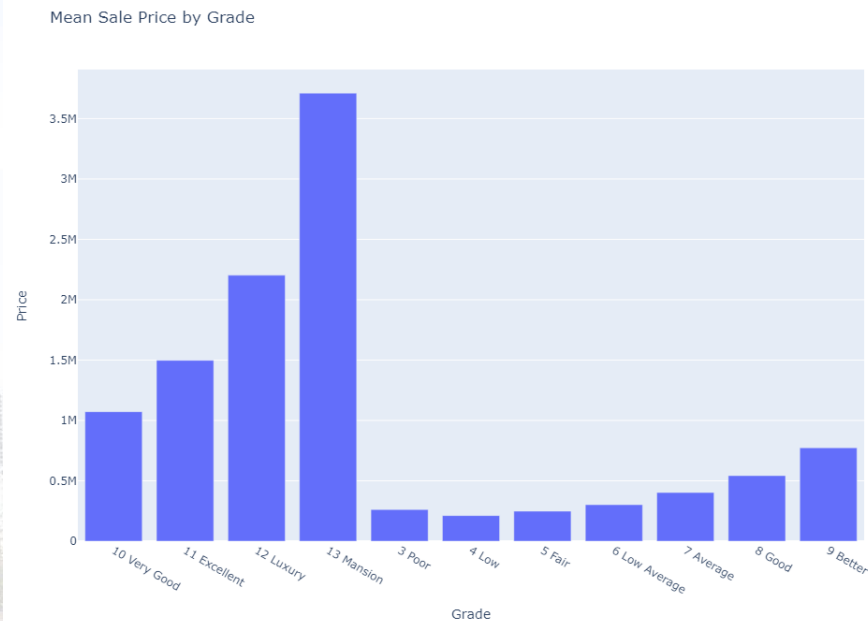
The relationship between properties condition and the price



From the data we can observe that houses with an average rating were the most expensive while the houses with poor conditions went for the least amount.

Data analysis

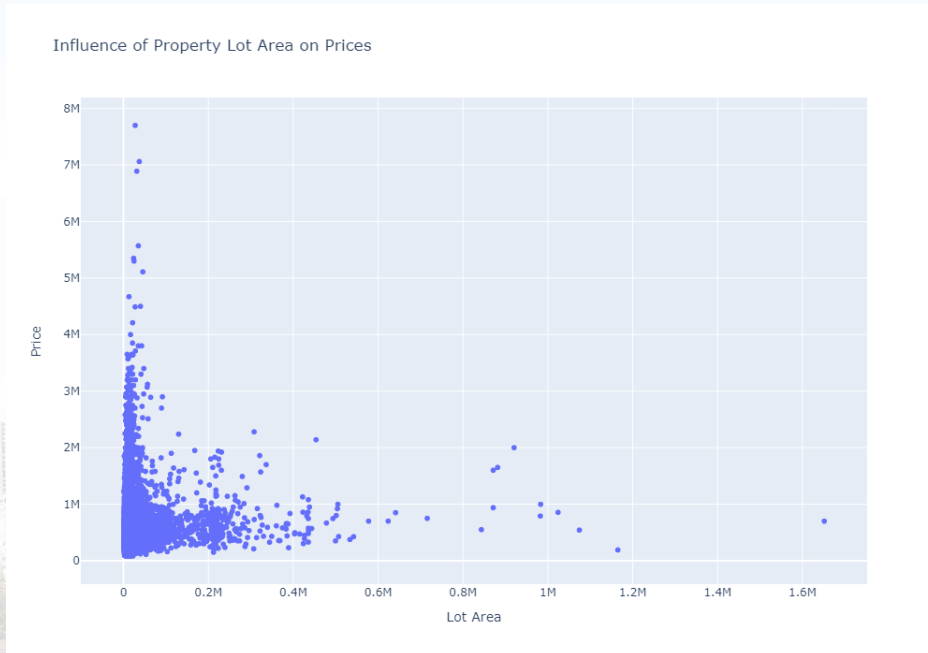
Relationship between mean property sale price and grades



From the data we can observe that houses with mansion grade were the most expensive while the houses with low grades all the way to poor went for the least amount.

Data analysis

The relationship between property lot area and the price

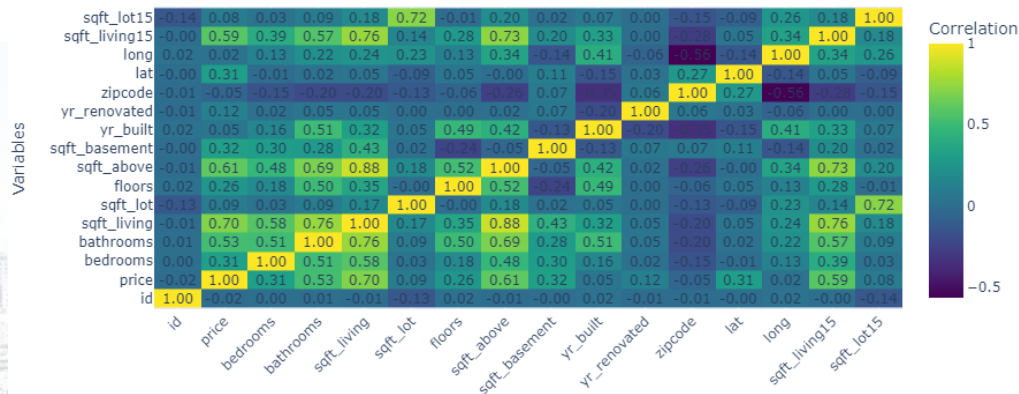


From the data we can observe that there is no linear relationship between property Lot Area and price

Data analysis

Heatmap

Correlation of Numerical Variables



From the heatmap we can observe that most of our independent variables have relatively low multicollinearity. However, others like number of bathrooms and square footage of living area have high correlation as expected since having more bathrooms in a property means the property is bigger. Due to lack of control of such we decided to keep all the variables in the model.

MODELLING

Our second model was statistically significant and yielded the results below:

- intercept -6340563.972160146
- Coefficient for bedrooms: -44814.925573082975
- Coefficient for bathrooms: 55378.911932557334
- Coefficient for sqft_living: 141.01800656747443
- Coefficient for sqft_lot: 0.14583862950126067
- Coefficient for floors: 35157.664833728624
- Coefficient for waterfront: 686388.9088227301
- Coefficient for view: -32554.661057758272
- Coefficient for condition: 17010.52352679578
- Coefficient for grade: -19444.61239366132
- Coefficient for sqft_above: 80.36728154854919
- Coefficient for sqft_basement: 50.597946361994055
- Coefficient for yr_built: -1815.4225049187717
- Coefficient for yr_renovated: 37.9443130354845
- Coefficient for zipcode: -595.7936853375476
- Coefficient for lat: 659864.4644823401
- Coefficient for long: -303984.6489353096
- Coefficient for sqft_living15: 73.24620750372713
- Coefficient for sqft_lot15: -0.5226343925391608
- Mean Squared Error (MSE): 42230927811.607
- Root Mean Squared Error (RMSE): 205501.64917004196
- R-squared (R2): 0.6756862676861248
- F-statistic: 1269.3529411764707
- P-value: 1.1102230246251565e-16

CONCLUSIONS

- *Bedrooms: The number of bedrooms has a negative effect on house prices.*
- *Bathrooms: More bathrooms tend to increase house prices.*
- *Square Footage: An increase in square footage positively impacts house prices.*
- *Waterfront Property: Having a waterfront location significantly increases house prices.*
- *Condition: Better property condition is associated with higher house prices.*
- *Grade: Higher property grades positively affect house prices.*
- *Year Built: Older homes tend to have lower prices, while newer homes have higher prices.*
- *Year Renovated: Renovations positively impact house prices.*
- *Location (Zipcode, Latitude, Longitude): These location-related variables significantly influence house prices.*



RECOMMENDATIONS

- *Property Improvement: Sellers should consider investing in property improvements, such as renovating or upgrading certain features, to potentially increase the property's value.*
- *Location Matters: The location of the property, as indicated by variables like zipcode, latitude, and longitude, plays a significant role in house prices.*
- *Bedrooms and Bathrooms: The number of bedrooms and bathrooms can affect prices. Sellers should highlight these features, and buyers should consider their needs.*
- *Square Footage: Increasing square footage can positively impact property value. Sellers may benefit from maximizing usable living space.*
- *Waterfront Properties: Waterfront properties command higher prices. Buyers interested in such properties should expect premium pricing.*
- *Age of the Property: Older properties tend to have lower prices, so buyers should consider the trade-offs between historic charm and modern amenities.*
- *Condition and Grade: Sellers should focus on property condition and grade to potentially increase market appeal and pricing.*

THANK YOU

- Dominic Mulli
- Allan Ngeiywa
- Celestine Imelda
- Kevin Kagia

