House Price Prediction with Multiple Regression Analysis

BY ANDRINNY SOEMITRO

Introduction

Objective:

Develop an accurate model for predicting house prices.

Focus:

Provide a strategic edge for house sellers in the real estate market.

Challenge:

House sellers struggle with optimal pricing for attracting buyers and ensuring profitability.

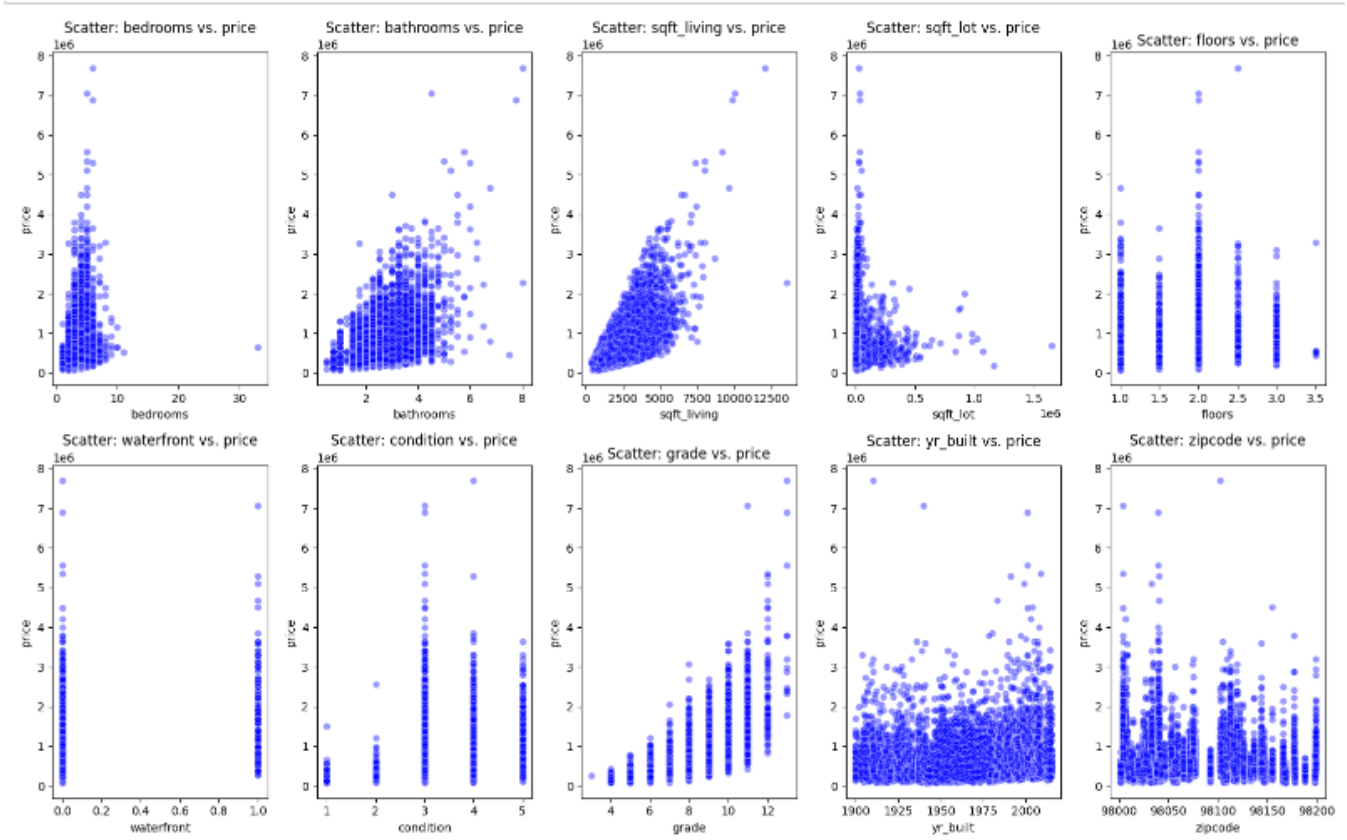
Business Questions

1. What are the significant factors influencing house prices?
2. How do different zip codes impact average property prices, and what factors contribute to these variations?
3. How can this knowledge be effectively utilized to establish an optimal price prediction for sellers?

Data Understanding

- Dataset from 'House Sales in King County, USA
- Data preparation involved handling missing values, encoding categorical variables, and scaling numerical features for optimal model performance.
- Methodologies = Regression predictive analysis.

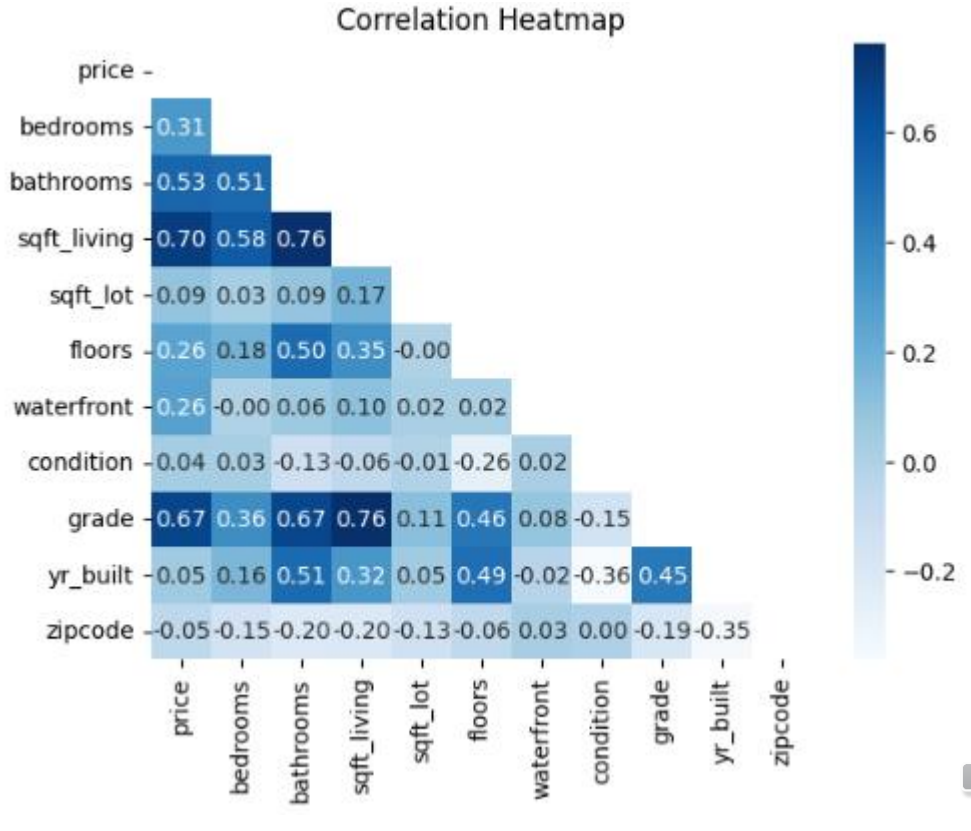
Data Modelling



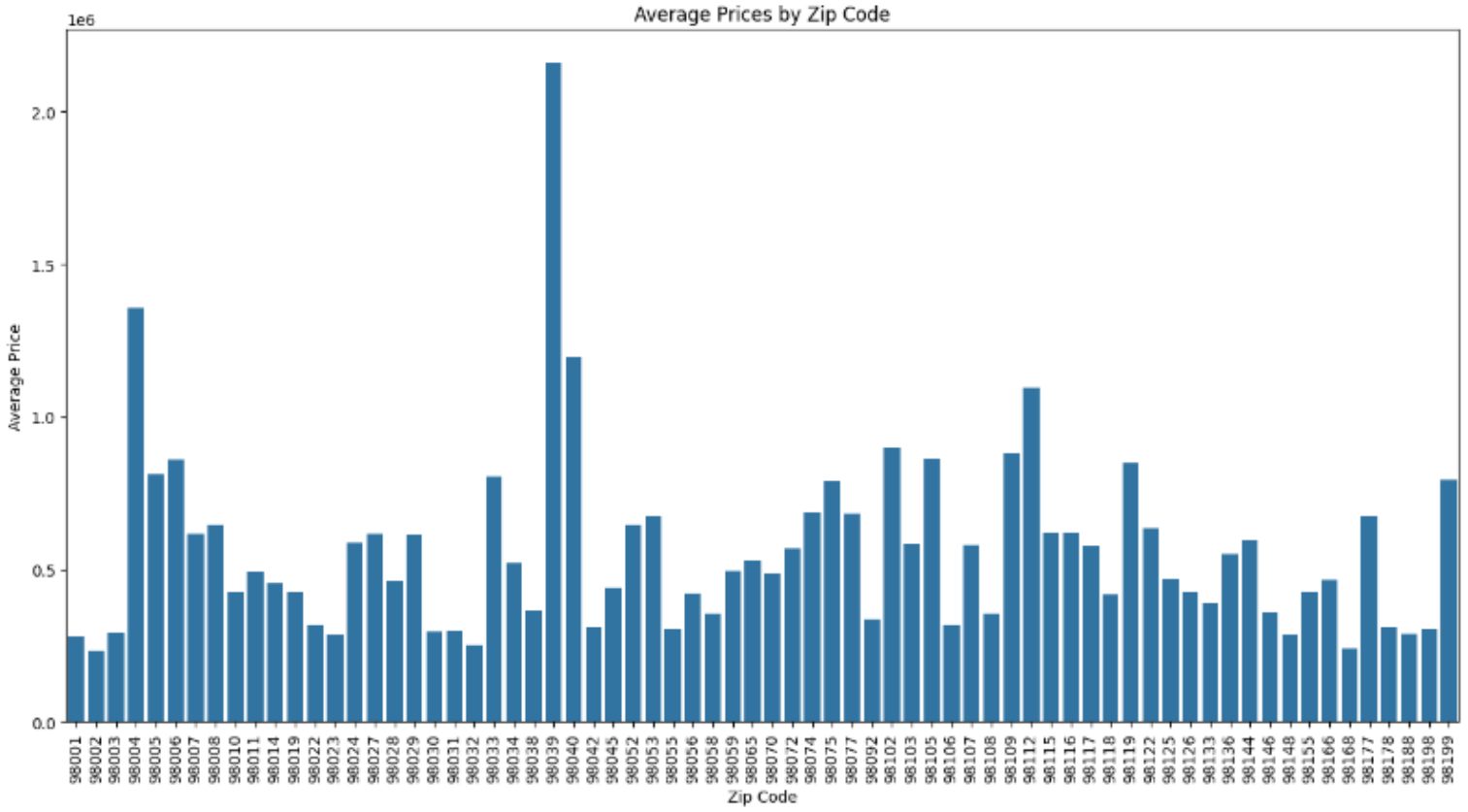
Identifying categorical and continuous variables with scatterplot

Significant Factors Influencing Price

- Correlation closer to 1.
- Positively correlated with price: sqft_living(0.70), grade(0.67), and bathrooms(0.53).
- Larger living spaces, higher quality, and more bathrooms tend to correlate with higher property values.



Effect of Zipcode on Average Property Price



The top 3 highest average prices are within the zipcodes 98039, 98004, and 98040.

Effect of Zipcode on Average Property Price (Contd.)

Despite similar features, houses exhibit variations in pricing in different zipcodes.

	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	condition	grade	yr_built	zipcode
0	775000	3	1.00	1175	10454	1.0	0	4	6	1949	98004
1	650000	3	1.00	1520	10227	1.0	0	4	6	1951	98004
2	650000	3	1.00	920	6750	1.0	0	4	7	1951	98004
3	875000	3	1.00	1220	8119	1.0	0	4	7	1955	98039
4	940000	3	1.00	1220	8119	1.0	0	4	7	1955	98039
5	937500	3	1.00	1320	8500	1.0	0	4	7	1954	98039
6	571500	3	1.00	1300	6710	1.0	0	4	6	1952	98040
7	1200000	3	1.75	1560	8078	1.5	1	4	6	1928	98040
8	604000	3	1.00	1440	13824	1.0	0	4	7	1957	98040

House Price Prediction

```
({'bedrooms': [7],
  'bathrooms': [6],
  'sqft_living': [9500],
  'sqft_lot': [6000],
  'floors': [2],
  'waterfront': [1],
  'condition': [5],
  'grade': [8],
  'yr_built': [2014],
  'zipcode': [98004]})
```

Estimated price for a 7-bed, 6-bath, 9,500 sqft home in the 98004 zipcode is US$2,607,762.39

```
# Feature scaling (applied to the new data)
new_data_scaled = sc_X.transform(AB)

models = [linear_model]
for i in models:
    # Making predictions with the scaled new data
    predicted_output_scaled = i.predict(new_data_scaled)
    predicted_output = sc_y.inverse_transform(predicted_output_scaled.reshape(-1, 1)).flatten()
    print("Predicted Output:", predicted_output[0])
```

```
Predicted Output: 2607762.3871373176
```

Conclusion (Findings)

- Square footage of living space (sqft_living), grade, and the number of bathrooms exhibit strong correlations with property prices.
- Despite similar features, homes in the top 3 zip codes show varied prices, highlighting the strong impact of location on housing market dynamics and property values.

Conclusion (Limitations and Future Steps)

Limitations:

- Data limitation
- Analysis of external factors not performed
- Dynamic market of real estate ⇨ frequent updates for predictive modelling to stay relevant

Future Steps:

- Analysis of external factors
- Perform time series analysis

THANK YOU