Northwestern County: Real Estate Market Analysis

Table of contents

1. Business Problem

2. Business and Data Understanding

3. Modeling

4. Results

5. Recommendations

6. Limitations

7. Conclusion

1.0 Business Problem

Which property attributes largely collectively influence the estimated value of homes to help the agency make strategic decisions for optimal return on investment for buyers and sellers?

To answer this question, our modeling determines which factors have the biggest positive effects on sale price.

2.0 Business and Data Understanding

The Data:
- King County Housing data
- Includes 20 housing variables ranging from bathrooms to sqft_basement
- We omitted irrelevant variables.
- Categorical variables were dropped, and dummy variables were created
- Regression models were created

Final variables include:
- Sqft_living, bathrooms, sqft_above, sqft_living15, view, waterfront.

3.0 Modeling

Regression modeling fulfills two objectives in relation to our business case:

1. Identifies significant variables by examining the relationship between price and the variables we defined.

2. Offers a factor (referred to as model coefficients) by which pricing fluctuates in response to changes in our variables.

4.0 Model Results

```
                        OLS Regression Results
==============================================================================
Dep. Variable:                    price   R-squared:                       0.506
Model:                              OLS   Adj. R-squared:                  0.506
Method:                   Least Squares   F-statistic:                     2762.
Date:                 Tue, 02 Jan 2024   Prob (F-statistic):               0.00
Time:                         05:19:56   Log-Likelihood:                 -9010.2
No. Observations:                21567   AIC:                          1.804e+04
Df Residuals:                    21558   BIC:                          1.811e+04
Df Model:                            8
Covariance Type:             nonrobust
==============================================================================
                    coef     std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const             6.3165       0.070     89.833      0.000       6.179       6.454
sqft_living       0.5606       0.014     39.339      0.000       0.533       0.589
bathrooms         0.0949       0.016      6.080      0.000       0.064       0.126
sqft_above       -0.0152       0.012     -1.256      0.209      -0.039       0.009
sqft_living15     0.3286       0.012     27.579      0.000       0.305       0.352
view_AVERAGE      0.2063       0.012     16.634      0.000       0.182       0.231
view_EXCELLENT    0.5959       0.021     27.914      0.000       0.554       0.638
view_FAIR         0.2479       0.021     12.034      0.000       0.207       0.288
view_GOOD         0.2797       0.017     16.581      0.000       0.247       0.313
==============================================================================
Omnibus:                       156.856   Durbin-Watson:                    1.971
...
```
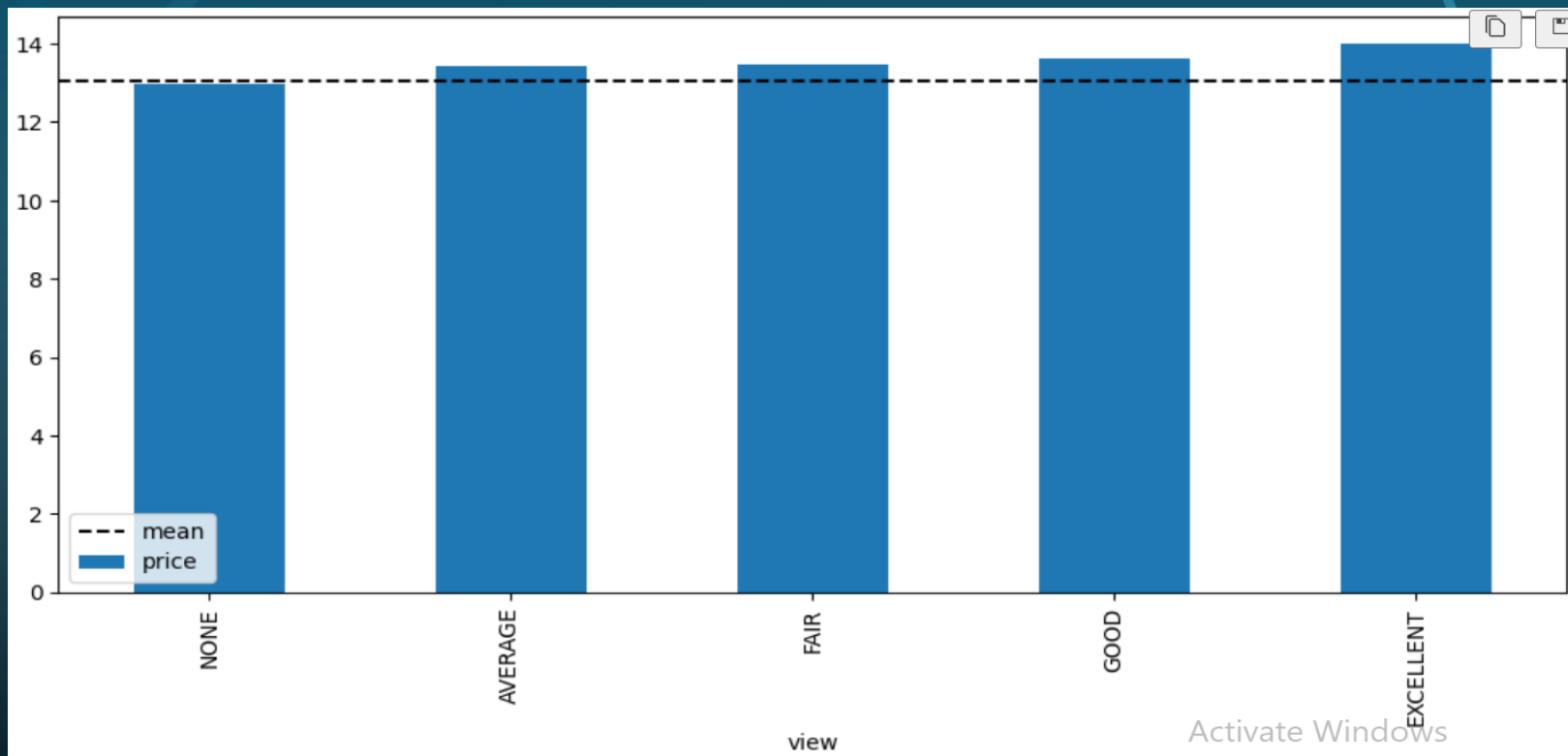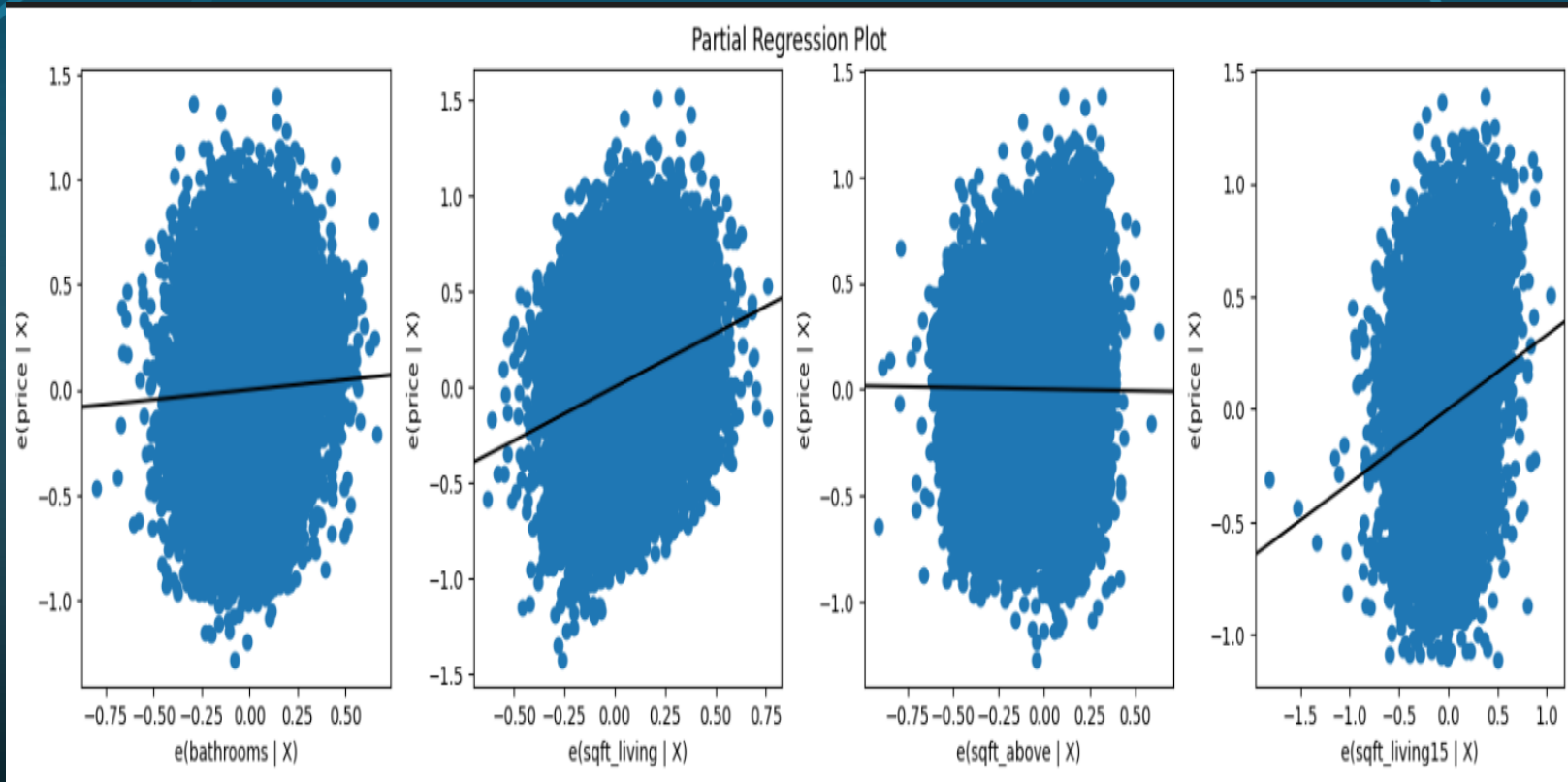
Visualizing to check relationship between different view and house price

Partial regression plot

Partial Regression Plot

4.0 Recommendations

It is highly recommended that the audience focus on the attributes of' sqft_living', 'bathrooms','sqft_above','sqft_living15', and 'view' when deciding on the prices of houses. These variables are the key predictors of house prices in King County. The variables showed strong positive correlations, suggesting that an improvement in the variables increases prices.

When choosing a model, it's important to balance complexity and performance. Depending on your objectives, you might opt for a simpler model with fewer features, especially if interpretability is a key priority.

5.0 Limitation

Variable Selection: The choice of variables included in the models may not be exhaustive. There could be other important factors influencing house prices that are not considered in the current models.

6.0 Conclusion

House properties influence prices of houses. The model suggests 'sqft_living', 'bathrooms', 'sqft_above', 'sqft_living15', and 'view' influence prices of houses in the county.

Thanks!

Any questions?