Project: Multiple linear regression modeling to analyze house sales in a northwestern county.

Ahmed Ali

Business and Data Understanding

Business Understanding

Stakeholder : Customers and Investors searching for affordable houses to buy or investing in .

Business Problem: Price prediction for the houses based on the data provided

Data Understanding

The Data Contains Multiple Features of Houses that are in King County.

Some of them include:

Dates houses were sold and renovated. The prices,bedrooms,bathrooms, their co-ordinates(latitudes and longitudes) and more.

We will be performing data analysis and regression on these.

Modeling for Continuous Data

After preprocessing and analysis of the data.

Here is the final model summary results for the continuous data

R-Squared	***0.468***
Adj. R-squared	***0.468***
F-Statistic	***2378***
Prob(F-statistic)	***0.00***

Regression Diagnostics

R-squared: Suggests that about 47% of the variance in dependent variable (price) is explained by the plot.

The f-statistic and the pvalue of the model suggests that it is statistically significant.

The pvalues of the coefficients indicate independent variables are statistically significant in predicting the dependent variable(reference the jupyter notebook)

`sqft_lot`: For every one-unit increase in square footage , the house price is expected to increase by $0.6107 , holding other variables constant.

`sqft_basement`:For every additional square footage of basement space, the house price is expected to increase by $130.0341, holding other variables constant.

`lat`: For every unit increase in latitude, the house price is expected to increase by $647,500 , holding other variables constant.

`long`:Longitude is a geographical coordinate that represents the east-west position on Earth. For every unit increase in longitude, the house price is expected to decrease by $299,900 , holding other variables constant.

`sqft_living15`: For every additional square foot, the house price is expected to increase by $314.7771 , holding other variables constant.

`sqft_lot15`: For every additional square foot, the house price is expected to decrease by $0.3207 , holding other variables constant.

`yr_built`:For every year increase in the age of the house, the price is expected to decrease by $452.1807 , holding other variables constant.

`numeric_month`:This represents the month when the house was sold . For every month change, the house price is expected to decrease by $2017.3423 , holding other variables constant.

Modelling for Discrete Data

Summary of the model results

R-squared:	***0.634***
Adj. R-squared:	***0.633***
F-statistic:	***409.5***
Prob (F-statistic):	***0.00***

R-squared: The model explains about 63% of variance in dependent variable(price)

The f-statistic and the pvalue suggests the overall model is statistically significant

The low pvalues of the coefficients determine that the inpendent variables are statistically significant in predicting the price

The const coefficeint(intercept) says that the house price is $329600 when all other predictors are 0.

Having 1.5 bathrooms is associated with a decrease in house price by $147,300 in this model.While a house with 8 bathrooms is associated with an increase in $4,358,000

A house with 1.5 floors has an increase of $53,800 and it decreases while floors are added with 3 floors having a decrease of $107,700.

Houses that were sold in 2015 have an increase of $25,160 than those sold in 2014.This could be a trend that house prices increase with year with could be a business opportunity.

Model for the Categorical Data

Summary of the model results

R-squared:	***0.593***
Adj. R-squared:	***0.593***
F-statistic:	***1746.***
Prob (F-statistic):	***0.00***

R-squared: About 59% of the variance in the dependent variable is explained by the model.

The high f-statistic and the low pvalue compared to the previous one before dropping says the model has improved and the model is statistically significant as well as the pvalue.

The high f-statistic and the low pvalue compared to the previous one before dropping says the model has improved and the model is statistically significant as well as the pvalue.

Const: The const(intercept) says the price is $1,109,000 when all the other predictors are 0

Waterfront_YES: Holding all other variables constant, if a house has a waterfront view (waterfront_YES), the predicted price increases by approximately $457,500.

view_EXCELLENT, view_FAIR, view_GOOD, view_NONE: if a house has an excellent view, the predicted price increases by approximately $267,000

condition_Good, condition_Very Good:For example, if the condition is very good, the predicted price increases by approximately $153,500 compared to a good condition.

grade_11 Excellent, grade_12 Luxury, grade_13 Mansion, grade_3 Poor, grade_4 Low, grade_5 Fair, grade_6 Low Average, grade_7 Average, grade_8 Good, grade_9 Better: These variables represent different levels of the house's grade. For instance, if a house is a mansion, the predicted price increases by approximately $2,561,000 compared to a poor grade , holding other variables constant.

numeric_year_2015: This variable represents the year 2015. If a house was sold in 2015, the predicted price increases by approximately $20,750 compared to other years, holding other variables constant.
The low pvalues in of the predictors indicate that they are statistically significant in predicting the target variable

Recommendations

Based on the models

I recommend to potential buyers and investors to invest or buy houses that have additional lot footage and basement footage.

Considering the houses are in king county, for each increase in latitude and longitude the house prices are set to increase by about $648K and $300K respectively.

Houses having 1.5 bathrooms is associated with a decrease in house price by $147,300 in this model.While a house with 8 bathrooms is associated with an increase in $4,358,000. So buy houses with more bathrooms.

Consequently, a house with 1.5 floors has an increase of $53,800 and it decreases while floors are added with 3 floors having a decrease of $107,700.

Houses that have waterfront are increase by approximately $458K than those that dont have waterfront.

If a house has an excellent view , its value is set to increase by about $267K with its value decresing as the view quality decreases.

Thank you for considering tme for the project for future purposes contact me