

phase-2-project-read-me

April 10, 2024

1 REAL ESTATE HOUSING PRICES PREDICTION MODEL

1.1 Business Understanding

The selling price of a home is a critical factor in the real estate market, as it directly influences the financial outcome for both buyers and sellers. For homeowners, the selling price represents the return on their investment and can significantly impact their financial well-being. Potential buyers, on the other hand, rely on the selling price to make informed decisions about purchasing a property within their budget and assessing its value relative to similar properties in the market. House price is influenced by a multitude of factors, which can be broadly categorized into three main categories: property-specific factors, market factors, and external factors. Property-specific factors encompass attributes such as location, size, condition, amenities, architectural style, and age of the property. Market factors include supply and demand dynamics, interest rates, mortgage availability, and prevailing economic conditions. External factors can range from neighborhood characteristics, such as school quality and crime rates, to broader influences like government policies, infrastructure development, and demographic trends. Understanding the factors that influence the selling price of residential properties is of paramount importance to various stakeholders involved in the real estate industry. Real estate agents need this knowledge to provide accurate pricing recommendations and effective marketing strategies for their clients. Homeowners can benefit from understanding these factors to make informed decisions when pricing their properties. Investors and developers can leverage this knowledge to identify promising investment opportunities and maximize their returns.

1.2 Problem Statement

Real estate market is highly volatile, influenced by economic conditions, housing demand, and external factors. As such setting inappropriate prices and making uninformed decisions of when to sell a house can be counterproductive. Research is essential to understand market trends and identify the best time to sell a home for maximizing its selling price. Analyzing property characteristics such as location, size, amenities, condition, and recent market trends through research aids in setting an appropriate selling price. By understanding how different home characteristics impact selling prices, the agency can help home owners mitigate the risk of setting inappropriate prices or making poor investment decisions.

1.2.1 Objectives

1. To understand factors that affect price of a house
2. To develop a model that can predict housing price based on various features.
3. To make recommendations on how home owners can optimize selling price of a house ##
Modelling The Baseline model consists of 2 simple linear regression models demonstarting

the relationship between price and two highly correlating variables

1.3 Model evaluation approach

R-squared indicates the proportion of variance in the dependent variable (price) explained by the independent variables included. The Durbin-Watson statistic tests for autocorrelation in the residuals, with values close to 2 indicating no autocorrelation.

1.4 Model comparison approach

Root Mean Squared Error (RMSE): This is the square root of the MSE, which gives a measure of the average magnitude of the errors in the predicted values. Mean Squared Error (MSE): The MSE measures the average squared difference between the actual and predicted values. A lower MSE indicates better model performance, with a value of 0 indicating a perfect fit.

1.5 Baseline model interpretation

With an F-statistic p-value below 0.05, the overall model is statistically significant. The results show that sqft_living explains 41.8% variation in price of houses. The model coefficients (const and sqft_living) are both statistically significant, with t-statistic p-values below 0.05. For 0 sqft_living, the estimated price is estimated to be -43490. An increase of 1 square foot in the living area leads to a price increase of approximately \$283.4564. For the bathroom model, the F-statistic p-value below 0.05, the overall model is statistically significant. The results show that bathrooms explain 29% variation in price of houses. The model coefficients (const and sqft_living) are both statistically significant, with t-statistic p-values below 0.05. For 0 sqft_living, the estimated price is estimated to be 12.2218. An increase of 1 square foot in the living area leads to a price increase of approximately \$0.3935.

1.6 Evaluating Model Performance using Durbin-Watson.

The sqft_living model has a Durbin-Watson of 1.986 which shows that there is no significant autocorrelation. The bathrooms model has a Durbin-Watson of 1.968 which shows that there is no significant autocorrelation.

1.7 Second Model Interpretation(Multiple_Linear_Regression)

The first regression model shows that approximately 59.4% of the variability in house prices can be explained by the independent variables included in the model. These variables include features such as bedrooms, bathrooms, square footage, waterfront status, and others.

The second regression model with an R-squared of 0.594 indicates that approximately 59.4% of the variability in house prices can be explained by the included independent variables. These variables, such as bedrooms, bathrooms, square footage, and others, collectively contribute to predicting house prices, as evidenced by significant coefficients and low p-values. The third regression model (which has variables with the highest correlation with price) has an R-squared of 0.511 suggests that approximately 51.1% of the variability in house prices can be explained by the included independent variables.

1.8 polynomial regression

The first polynomial regression with a degree of 2 on the dataset, scaling the features and splitting the data into train and test sets, resulting in a root mean squared error (RMSE) of approximately 0.35 when predicting the target variable on the test data.

The second polynomial regression with a degree of 3 on the dataset, achieving an RMSE of approximately 0.34 as the evaluation metric. The third polynomial regression with a degree of 4 on the dataset, achieving an RMSE of approximately 0.34 when predicting the target variable on the scaled test data.

1.9 Conclusions

Being on waterfront has highest positive effect followed by having an excellent, good, fair and average view respectively. Number of floors, years since renovated and square footage of living space have moderate positive effect Bedrooms have the highest negative effect on price followed by selling in winter, summer, fall and grade of house ## Limitations The dataset does not include other factors such as location-specific characteristics, neighborhood amenities which determine price of houses The data is collected over a period of time which encompasses different economic situations, changing market conditions, economic trends, and policy changes which were affect house prices but were not captured in the dataset. ## Recommendations Put up houses for sale in spring as it fetches higher prices Renovate houses as it increases value of the property Increase the square footage of the living space as it increases value Do not increase grade of the house as it reduces value of property Do not increase number of bedrooms as it reduces price of the house

[]: