Analyzing House Sales in a Northwestern County Using Multiple Linear Regression Modeling

By Jeff Kiarie

Overview

In this project, we will apply statistical analytic methods in this case Multiple Linear Regression Modeling to comprehend the variables affecting home sales in a certain county in the northwest.

Overall, this analysis aims to advance knowledge and comprehension of the northwest county's housing market by illuminating the variables that have a significant impact on sales and possibly assisting various stakeholders in streamlining their strategies and decision-making procedures.

Business and Data Understanding

The primary objective of a real estate company that specializes in helping homeowners buy and sell homes is to offer beneficial services that aid homeowners in maximizing the value of their properties. A significant business problem to address is to be able to advise clients on house modifications and their potential impact on the assessed worth of their homes.

We are provided a file ‘kc_house_data.csv’ which contains dataset consisting of house prices from King County an area in the US State of Washington.The dataset contains 21597 entities with the following columns: 'id', 'date', 'price', 'bedrooms', 'bathrooms', 'sqft_living', 'sqft_lot', 'floors', 'waterfront', 'view', 'condition', 'grade', 'sqft_above', 'sqft_basement', 'yr_built', 'yr_renovated', 'zip code', 'lat', 'long', 'sqft_living15', 'sqft_lot15'.

Modeling

We will be using multiple linear regression modeling because the coefficients in linear regression models have clear interpretations. They represent the change in the dependent variable associated with a one-unit change in the corresponding independent variable, holding other variables constant. This makes it easier to explain the impact of different factors on the outcome.

For example we can have the model:

Salary =$30,000 + $2,000 * Experience + $5,000 * Education + ε

Now, let's interpret the coefficients:

Intercept (β0): The intercept represents the expected salary when both "Experience" and "Education" are zero. In this example, the intercept is $30,000. This means that even with zero years of experience and education, a person can expect a salary of $30,000.

Experience coefficient (β1): The coefficient for "Experience" is $2,000. This indicates that for each additional year of experience, we can expect an increase in salary of $2,000, assuming education level remains constant. It suggests that experience has a positive and linear relationship with salary.

Education coefficient (β2): The coefficient for "Education" is $5,000. This implies that for each additional level of education (e.g., completing another degree or certification), we can expect an increase in salary of $5,000, assuming years of experience remain constant. It suggests that higher education has a positive impact on salary.

Interpreting these coefficients allows us to make specific predictions and understand the relative importance of each variable. For example, suppose we have a person with 5 years of experience and a master's degree (education level is 2, representing higher education). Plugging in these values into the model, we can predict their salary as follows:

Salary = $30,000 + $2,000 * 5 + $5,000 * 2 = $30,000 + $10,000 + $10,000 = $50,000

So, according to the model, a person with 5 years of experience and a master's degree can expect a salary of $50,000.

In our dataset when we pass the price as the predictor variable(salary) with sqm_above,sqm_basement,floors,bathrooms yr_built,yr_renovated as the independent variables(experience,education) we get the following model:

price = 6,006,000 + 2934.7*sqm_above + 2799.72*sqm_basement + 55,510*floors + 49,230*bathrooms- 3154.97*yr_built + 28.64*yr_renovated

(we’ve converted the metric from square foot to square meters to enhance relatability)

Regression results

price = 6,006,000 + 2934.7*sqm_above + 2799.72*sqm_basement + 55,510*floors + 49,230*bathrooms- 3154.97*yr_built + 28.64*yr_renovated

With the above model we can deduct the following results:

The intercept term is 6,006,000. This represents the estimated price when all the independent variables are zero.

The coefficient for "sqm_above" is 2934.7. It suggests that, on average, for every square meter increase in above ground living area, the price is expected to increase by $2934.7, holding other variables constant.

The coefficient for "sqm_basement" is 2799.72. It indicates that, on average, for every square meter increase in basement area, the price is expected to increase by $2799.72 , holding other variables constant.

The coefficient for "floors" is 55,510. This implies that, on average, each additional floor in the house is associated with an increase of $55,510 in price, assuming other variables remain constant.

The coefficient for "bathrooms" is 49,230. It suggests that, on average, for each additional bathroom in the house, the price is expected to increase by $49,230 , holding other variables constant.

The coefficient for "yr_built" is -3154.97. It indicates that, on average, for every year increase in the age of the house, the price is expected to decrease by $3154.97 , holding other variables constant.

The coefficient for "yr_renovated" is 28.64. It suggests that, on average, for every year increase in the age of the renovation, the price is expected to increase by $28.64 , holding other variables constant.

Recommendations

1. **Living Area (sqm_above)**: Increasing the living area of the home by one square meter is associated with an estimated increase in price of $2,934.7. For every additional square meter increase in the living area of a house relative to the mean of the living area of the nearest 15 houses, the price is expected to increase by $3,376. To maximize the estimated value of the property, consider expanding the above ground living space through additions or remodeling.

2. **Basement Area (sqm_basement)**: Increasing the basement area of the home by one square meter is associated with an estimated increase in price of $2,799.72. If your property has a basement or potential for one, renovating or expanding it could positively impact the estimated value of the home.

3. **Number of Floors (floors)**: Each additional floor in the house is associated with an estimated increase in price of $55,510. If feasible and within zoning regulations, consider adding additional floors to the property to potentially increase its value.

4. **Year of Renovation (yr_renovated)**: The year of renovation also affects the price of a house. For every year since the mean year of renovation of the nearest 15 houses, the price is expected to increase by $118.53. This implies that more recently renovated houses tend to have higher prices. If you are considering renovating your house, it's worth keeping in mind that a more recent renovation might contribute to an increased selling price.

Next Steps

1.Evaluate the costs associated with each recommendation against the expected increase in the property's value

2.Conduct thorough market research to understand the demand and preferences of potential buyers in the area

3.Consult with professionals: Seek advice from professionals such as real estate agents, property appraisers, or architects who specialize in renovations and property value enhancement.

4.Obtain necessary permits and approvals: If you decide to proceed with any construction or renovation work, ensure compliance with local zoning regulations and obtain any necessary permits or approvals from the appropriate authorities.

5.Create a detailed budget for the recommended renovations or additions, considering both construction costs and associated expenses such as design fees, permits, and materials.

6.If you choose to hire contractors or professionals for the renovations, conduct thorough research and obtain multiple quotes from reputable contractors

7.Throughout the renovation process, closely monitor the progress and regularly assess the impact on the property's estimated value

Thank you