Business Overview

Introduction

Real estate is one of the most important sectors of any economy. Understanding the key drivers of housing prices can provide valuable insights for both buyers and sellers in the market. In this project, we analyze a data set of house sales in a northwestern county to identify the factors that influence housing prices in the area.

Stakeholders: Real Estate Agency

Challenges

1. The real estate agency needs to provide accurate advice to homeowners on how to increase the estimated value of their homes, and by what amount. However, the agency currently lacks a reliable method for predicting the impact of home value. As a result, the agency is unable to provide accurate advice to homeowners about the potential return on investment.
2. Houses that stay longer in the market before being sold depreciate in value over time
3. Most homeowners Improving the value of the houses
4. Housing affordability : some houses were over valued while others were undervalued as compared to the normal market rate.

Proposed Solution

1. Develop a reliable method for predicting the impact of home improvements on the estimated value of a home.
2. To prevent houses from depreciating in value due to longer time in the market, the real estate agency can work with homeowners to set realistic prices based on market trends, property condition, and location. The agency can also help homeowners to prepare their homes for sale by making necessary repairs and improvements to increase the appeal of the property.
3. Provide homeowners with a list of home improvements that can increase the value of their homes, and offer advice on which improvements will provide the greatest return on investment.
4. Conduct market research and analysis to determine the appropriate value for each home based on its location, size, and other factors.

Conclusion

The real estate industry confronts various challenges, including providing accurate advice to homeowners, preventing home value depreciation, improving home value, and ensuring housing affordability. Effective solutions encompass developing reliable methods for predicting the impact of home improvements on value, setting realistic prices, providing a list of home improvements, and conducting market research. These solutions can create a sustainable and prosperous real estate market.

Problem Statement

The real estate agency is unable to accurately identify pricing factors, analyze trends in house prices, and identify undervalued properties, resulting in unreliable information for buyers and sellers.

Objectives (Question format; To ….)

1. To identify features influencing the pricing.
2. To analyze trends in house prices over time (time series analysis) and predict future prices.
3. To identify undervalued properties (outlier detection) and recommend better pricing strategies

Data Understanding

1. We imported different libraries that helped us to load and explore the dataset
2. We are using the kc house_data dataset, that is to be used by real estate companies in helping the homeowners manage and maximize profit from their properties.
3. We used inbuilt python functions to get various insights of our data for example the shape,
4. The dataset provided was sufficient for us hence we didn't source for more data
5. Columns in the dataset contained both numerical and categorical variables

Data Cleaning

- Upon inspection of this dataset we realized there were duplicate ids that needed to be dropped as these would be flagged as fraud since one particular house can't be sold with the same details more than once.
- We dropped missing values because the waterfront and year renovated because they have a high percentages of 11% and 17% respectively.
- We used the outliers to identify undervalued properties (outlier detection) because they recommend better prices that will help the agency maximize on profits.

Data Analysis

- We used a heatmap to establish the correlation between the numerical variables in the dataset
- Then we used a boxplot to establish the categorical variables and price.
- We used scatter plots to visualize the top 5 correlated features with the price.
- We plotted a time series of average house prices over time

Modeling

- We used a baseline model between ‘sqft_ living and price because it had the highest correlation with the price.

- We used Multiple Linear Regression to create a model between several features and the price. We used both to see how the R-squared
- We used the z score to find outliers which were 325
- We also used a baseline model with price as the dependent variable and age of the house as independent variable.
- We decided to opt for RMSE as our metric of success because it is measured in the same units as the response variable.

Conclusion(s)

After analysis in modeling we concluded that:

1. Some of the features that influence the pricing of houses include:
* Square footage of living space in the home: an additional square footage increases the price by $199.09
* waterfront: the presence of a waterfront has an associated increase in price of $70,000
* Condition of the house: houses in good conditions have an associated increase in price of $35,650 compared to houses with average condition.
2. For every additional year in the age of a house, there is an associated decrease in price of $626.09
3. Some of the overvalued properties were found to be older than 50 years of age
* The square footage of interior housing living space for the nearest 15 neighbors influences the pricing of houses, in that, an additional square footage leads to an increase in price by $48.35

Recommendations

We recommend that:

1. There is a need to do further exploration into other variables in order to better understand the determinants of house prices.

2. The agency should consider re-purposing the old houses and targeting business owners rather than homeowners.

3. The agency should consider investing in properties with waterfronts as this could increase their profitability.