3.4 Final Project Submission

- Student names: Amos Kipkirui, Brian Muli, Emilly Njue, Swaleh Athuman, Samwel Kagwi,Heddy Berit
- Students pace: Full time
- Scheduled project review date/time: 21/04/2023
- Instructor name:
- Blog post URL: https://github.com/swalehmwadime/dsc-phase-2-project-v2-3.git

BUSINESS UNDERSTANDING

INTRODUCTION

The real estate market is a dynamic and ever-changing industry, and accurate prediction of housing prices is crucial for both buyers and sellers. In order to make informed decisions, stakeholders in the real estate industry need access to reliable and comprehensive data.

The King County House Sales dataset is a valuable resource for understanding the dynamics of the real estate market in King County. This dataset contains detailed information on house sale prices including a wide range of features such as the number of bedrooms, bathrooms, square footage, location, and more.

This dataset allows for in-depth analysis and modeling to understand the factors that influence housing prices in the region, and serves as a valuable resource for developing and testing predictive models for accurate price predictions.

We will provide an overview of the King County House Sales dataset, including its key features, data quality, and potential use cases. We will also highlight the significance of this dataset for evaluating regression models to predict housing prices in King County, and the potential benefits it can offer to stakeholders in the real estate industry.

Business Problem

A real estate agency located in King County is looking to advice homeowners about how home renovations might increase the value of their homes, and by what amount. The agency is looking to use the King County house dataset provided to make recommendations on the best renovations that home owners can undertake.

PROBLEM STATEMENT

To aid in making these recommendations, we will attempt to answer these questions:

1. Determine what are the key factors that significantly impact housing prices in King County?

2. How does the number of bedrooms, bathrooms, grade ,zipcode and square footage of a house correlate with its sale price in King County.
3. Predict the Price of a house given houses specification.

MODEL INTERPRETATION

We build a model that we interpretated as follows

The model explains a `48%` of the variance in `price` which shows an increase compared to the first model which had a variance of `38%` with only one predictor (sqft_living).

The model is `statistically significant` overall, with an F-statistic p-value well below 0.05.

The model coefficients (`const`, `sqft_living`, `numeric_grade`, `bathrooms`, `bedrooms`, `zpcode`) are both statistically significant, with t-statistic p-values well below 0.05

MODEL EVALUATION

We calculated the RMSE value for the final model, which came out to be USD 163011. This means that, on average, our model's predictions are USD 163011 away from the actual prices.

INSIGHTS

1. The size of the house, as measured by sqft_living, is a strong predictor of the price category. A one-unit increase in sqft_living leads to an increase in the log odds of a house being expensive. This suggests that larger homes are more likely to be expensive and buyers may be willing to pay a premium for more living space.

2. The numeric_grade of the property is also an important predictor of price category. As the numeric grade increases, the odds of a house being expensive also increases. This suggests that properties with higher grades (i.e. higher quality) are likely to be more expensive and may be valued by buyers.

3. On the other hand, the number of bathrooms and bedrooms have a negative effect on the odds of a house being in a higher price category. This suggests that more bedrooms and bathrooms may not necessarily be a desirable feature for buyers seeking higher-end properties.

4. Finally, the zipcode variable also plays an important role in predicting the price category. Properties located in certain zipcodes are more likely to be expensive, indicating that location may be an important factor for buyers looking for higher-end properties.

CONCLUSIONS

This project aimed to develop a model to predict housing prices in King County based on various features such as square footage, number of bedrooms and bathrooms, grading and location(zipcode).

Firstly, we performed exploratory data analysis (EDA) and found that the price of houses was positively correlated with the square footage and the grade of the house. We also discovered that the location of the house had a significant impact on the price.

We then developed multiple linear regression models to predict the price of the house, and we found that the model that included square footage, grade, bedrooms, bathrooms and zipcode as the predictors performed reasonably well, explaining about 48% of the variance in price.

We then developed a classification model to predict whether a house is expensive or affordable based on features such as square footage, number of bedrooms and bathrooms, grade and location. We used logistic regression and found that the model performed well, with an accuracy of 73%.

Based on the results, we can conclude that square footage and the numeric grade of the house are significant predictors of housing prices in King County. Additionally, we can use the classification model to predict whether a house is expensive or affordable based on the features of the house.

RECOMMENDATIONS

1. Homeowners should focus on highlighting the size of their homes if they want to attract buyers in higher price categories. This means showcasing the square footage of the property.

2. Buyers who are seeking higher-end properties should pay closer attention to the location and zip code of the properties they are considering. Properties located in certain zip codes are more likely to be expensive, so buyers should research the local real estate market and focus on properties in desirable areas.

3. Houses with a higher "numeric_grade" are more likely to fall into the higher price category. Therefore, it would be recommended for homeowners to consider investing in improvements to their home's grade in order to potentially increase its value.

4. Buyers who are looking for higher-end properties may not necessarily prioritize the number of bathrooms and bedrooms. Therefore, sellers should be aware that adding additional bathrooms or bedrooms may not necessarily increase the value of the property.

NEXT STEPS

Further research and analysis can be done to improve the accuracy of the models and to gain a better understanding of the factors that influence housing prices in King County.