


 **D-EKALE / KC-Housing-Project-2** Public

 View license

 0 stars  0 forks

Star

Watch

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings



 main ▾

...



Your main branch isn't protected
Protect this branch from force pushing or deletion, or require status checks before merging. [Learn more](#)

Protect this branch

Dismiss

 **D-EKALE** Added Files ... 2 minutes ago  80

[View code](#)

 **README.md** 

Predicting Home Sale Prices in King County.

This project uses linear regression models to analyze the factors that influence home prices in the King County area. The goal of this project is to provide actionable insights and recommendations to stakeholders in the real estate industry.

Business Problem

Our stakeholder is homeowners who are looking to renovate their homes and want to estimate the impact of these renovations on the value of their home. Our business problem is to identify which home features are most important in determining a home's sale price and estimate how much value can be added by improving these features.

Objectives

- Identify the most significant home features that contribute to a home's sale price.
- Estimate the value added to a home by improving these significant features.
- Develop a methodology for accurately estimating the impact of home renovations on home value.
- Provide homeowners with an easy-to-use tool for estimating the value added by renovating specific home features.

- Provide actionable recommendations to homeowners looking to renovate their homes to maximize their home value.

Data Understanding

The data used in this project is the King County House Sales dataset, which can be found in `kc_house_data.csv` in the data folder. The dataset contains 21,597 records and 21 columns. The description of the column names can be found in `column_names.md` in the same folder. It contains 21 columns with each being either numerical or categorical data.

Data Processing

Data cleaning techniques are used to handle any missing values, duplicates as well as encoding categorical features as necessary.

Data Analysis

Exploratory data analysis is performed to visualize the distribution of each feature, as well as any relationships between features and the target variable. We can easily do this using plots and summary statistics tables.

Modeling

To be able to predict how these variables affect the target variable (price), we need to build a model that will help us make predictions. In this context, a linear regression was a good fit for this model. Both simple and multiple regression models are tested, as well as models with regularization to handle multicollinearity and overfitting.

Interpretations and Recommendations

The coefficients of the final model are examined to identify the features that have the strongest effect on home prices. These features are used to make recommendations to stakeholders about how to increase the value of their homes.

Conclusion

The linear regression models provide valuable insights into the factors that influence home prices in the King County area. These insights can be used to make informed business decisions and to provide valuable advice to homeowners. It is important to note that the models have limitations and may not capture all the factors that influence home prices. Additional factors should be considered when making decisions about buying or selling a home.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 5



D-EKALE Daniel Ekale



Eddie-254 Eddie

Languages

● Jupyter Notebook 100.0%