

Vivian-Maiyo /
dsc-phase-2-project-F

<> Code

Issues

Pull requests

Projects

Wiki

Security

Insights

Settings

👁

🔗

★

📄 View license

★ 0 stars

🔗 0 forks

👁 1 watching

🔗 1 Branch

🏷 0 Tags

🔄 Activity

🌐 Public repository

main

1 Branch

0 Tags

🔍 Go to file

Go to file

+

Add file

Code

Vivian-Maiyo

github

1 minute ago

data	revisions for v2.3	3 years ago
images	Add files via upload	yesterday
.canvas	sync .canvas	3 years ago
.gitignore	Initial commit	4 years ago
CONTRIBUTING.md	Initial commit	4 years ago
LICENSE.md	Initial commit	4 years ago
Phase 2 Project Notebook.ipynb	Notebook.ipynb	1 hour ago
Phase 2 Project Notebook.pdf	Notebook.pdf	1 hour ago
Phase 2 Project Presentation.pdf	Presentation	28 minutes ago
README.md	edit	yesterday
halfway-there.gif	Initial commit	4 years ago

README

License

House Price Prediction in King County

Author: Group 12

Project Overview

Multiple linear regression modelling to predict house prices in King County. The project aims to derive data driven recommendations on features that best predict house prices to developers seeking to increase affordable housing units in King County.

Business Problem

There is limited availability of residential units in King County, causing increased costs across different grades. Disparities in the housing market, especially in desirable areas, are leading to unusually high property values. There is need to come up with a model that can accurately predict house prices in King County based on the features of a house to guide developers on what features they ought to focus on to develop affordable housing units.

The Data

This project uses the King County House Sales dataset from Kaggle. The data has 20 variables describing the size, condition, grade, location and other features of a house that influence its sale price. The description of the variable names(column names) can be found in the `column_names.md` folder.

Methods

The data was cleaned and prepared for analysis. Exploratory data analysis was conducted to give insight into the features. Predictor variables were then selected based on the strength of their correlation with the outcome variable(house sale price), while minimizing multicollinearity between predictor variables. The selected variables were used to create a multiple linear regression model to predict house prices.

Results

Our multiple linear regression model consisted 6 predictor variables and an intercept that were all statistically significant with p-values of 0.00. The R-squared value of the model was 0.582, meaning the variables were accounting for 58.2% of the variance in the outcome variable.

Multiple Linear Regression Model Summary		
R-squared value: 0.582	F-statistic: 4404	P-value of F-statistic: 0.00
Variable	Coefficient value	P-value of Co-efficient
Intercept	5.674e+06	0.00
Grade	-1.203e+05	0.00
Bathrooms	2.569e+04	0.00
Bedrooms	-2.228e+04	0.00
Sqft_living	127.79	0.00
Sqft_lot	-5.8597	0.00
Yr_built	-3188.98	0.00

Recommendations

To ensure availability of affordable housing in King County, developers should:

- Prioritize the construction of houses with a an average grade rating. These will be comfortable without experiencing the price inflation associated with higher grade rating.
- Focus more on properties away from the waterfront where price tends to be very high.
- Limit the number of bedrooms and bathrooms to the requirements of an average home buyer in King County. Many bathrooms will increase house prices while many bedrooms will decrease prices as many people do not seem to be searching for luxurious housing.

Further Analysis

▲

⚙

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Contributors 6



Languages

- Jupyter Notebook 100.0%