R3TR0Quan / KC_Realestate_Analysis_and_Modelling Public

King Country House Sales Analysis and Modelling ~ Moringa Phase-2 Project

1 star 5 forks

Star Watch

Code Issues Pull requests Actions Projects Wiki Security Insig

main

R3TR0Quan Merge pull request #19 from R3TR0Quan/Edna ... now 48

View code

README.md

KC_Realestate_Analysis_and_Modelling

Analysis and Modelling of the King Country House Sales Dataser ~ Moringa Phase-2 Project

1. Project Overview

Stakeholders Name: ALPHA TENNENT

In our analysis, we explored the data provided by Alpha Tennent Stakeholders and build a multiple linear regression model with some of the features stipulated in the dataset.

Problem Statement

We will be reviewing building grade, square-footage of living space, and location-related factors such as proximity to schools, coffee shops, parks, and scientology churches to determine which factors are highly correlated with home sale prices.

2. The Data

The main dataset we are using comes from the King County housing dataset that contains information on house sales between May 2014 and May 2015 consist of the following variables:

- date: Date of house sale
- price: The price which the house sold for
- bedrooms: How many bedrooms the house has
- bathrooms: How many bathrooms the house has
- sqft_living: How much square footage the house has
- sqft_lot: How much square footage the lot has
- floors: How many floors the house has
- waterfront: Whether the house is on the - - - - waterfront. Originally contained 'YES' or 'NO', converted to 0 or 1 for comparative purposes
- view: Whether the house has a view and whether it's fair, average, good, or excellent. Converted to numberical (0-4) for comparative purposes
- condition: overall condition of the house: Poor, Fair, Average, Good, Very Good
- grade: Numerical grading for house
- sqft_above: How much of the houses square footage is above ground
- sqft_basement: How much of the square footage is in the basement
- yr_built: Year the house was built
- yr_renovated: Year the house was renovated, if applicable
- zipcode: House zipcode
- lat: House's latitude coordinate
- long: House's longitude coordinate
- sqft_living15: Average size of living space for the closest 15 houses
- sqft_lot15: Average size of lot for the - - closest 15 houses

3. EDA

We did **Univariate**, **Bivariate** and **Multivariate** on the data.

- The categorical columns used were `Condition` and `Grade`
- The numerical columns include: `sqft_living` , `sqft_above` , `yr_built` , `yr_renovated`

4. Modelling

The base model used is a Log-Transformed model that gave an R^2 value of 0.503 We then proceeded to create some Multiple Regression models to try and improve our R^2 score

5. Challenges

Some of the challenges we faced were:

- Skeweness
- Fitting the model

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 4


R3TR0Quan Joshua Rwanda


leah-katiwa


FirenziaF Miss Nguuni

Languages

- **Jupyter Notebook** 100.0%