dan-rl1976 / kc_housing_regression Public

forked from learn-co-curriculum/dsc-phase-2-project

0 stars 155 forks

Star Watch

Code Pull requests Actions Projects Wiki Security Insights

main

This branch is 3 commits ahead of learn-co-curriculum:main. Contribute Fetch upstream

dan-rl1976 finished project ... 3 minutes ago 12

View code

README.md

Title

Authors: Daniel Ross-Leutwyler

Business Problem

I have been hired by a real estate development company in King County, Washington, to model real estate prices based on the King County Housing dataset. My intention is to build a model that will accurately describe housing prices based on the features contained in this data set, and offer advice on the best improvements that can be made to increase house value. Additionally, given the robust location data in these records, I will make suggestions about the best areas to look for housing. This model should ideally be able to assist the real estate company in making informed decisions about what type and where to build new houses.

Data Understanding

The dataset contains records of all recorded house sales in Kings County, WA from May 2014 to May 2015. There is a robust set of 21597 entries to work with. Information about the terms and data contained in this set can be examined at https://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r.

Methods

The data requires minimal cleaning, including scrubbing of outliers, and recasting data types to be suitable for regression modeling. I will also scale continuous data in order to make the coefficient values standard, regardless of the scale of the variables. Categorical variables will be one-hot-encoded using sklearn.

Finally, I will run a series of models, focused on returning a high R-squared value, and check the model residuals for homoscedasticity and a normal distribution.

Results

Through linear regression I have determined that the best variables to predict house price are related primarily to location, based on zip codes and the presence of a waterfront view.

Variables that are under the control of the owner include the total living area, the number of bathrooms, and the grade assignedto the property by the King County Assesor.

Repository Structure

Describe the structure of your repository and its contents:

```
├── README.md                         <- The top-level README for
reviewers of this project
├── final_notebook.ipynb              <- Narrative documentation of
analysis in Jupyter notebook
├── king_co_housing.pdf               <- PDF version of project
presentation
└── data                              <- the KC housing data set and
column descriptions
```

Releases

No releases published

Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%