

ndgigliotti / dsc-phase-2-project

forked from [learn-co-curriculum/dsc-phase-2-project](#)


 View license

☆ 0 stars  103 forks

☆ Star

👁 Watch ▾

 Code


 Pull requests

 Actions

 Projects

 Wiki

 Security


 Insights

 Settings

 main ▾

...

This branch is 25 commits ahead of learn-co-curriculum:main.

 Pull request

 Compare



ndgigliotti resid scatters ...

1 hour ago

 34

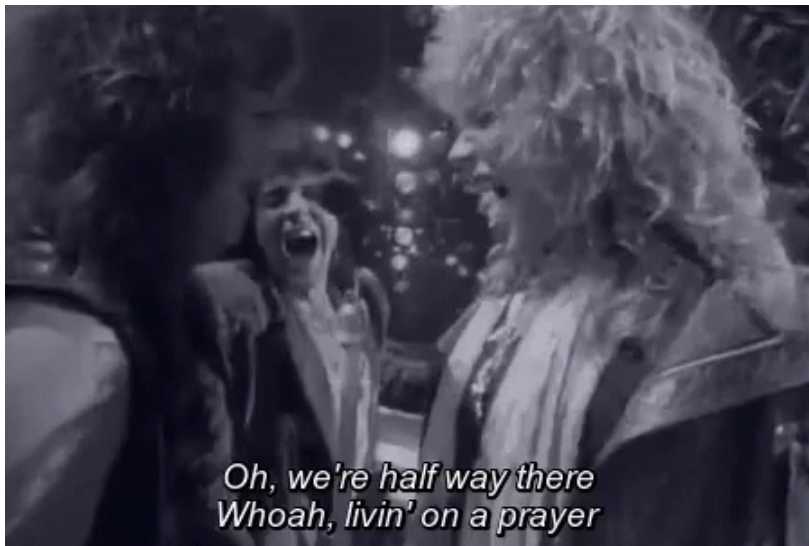
[View code](#)

☰ README.md



Phase 2 Project

Another module down--you're almost half way there!



All that remains in Phase 2 is to put our newfound data science skills to use with a large project! This project should take 20 to 30 hours to complete.

Project Overview

For this project, you will use regression modeling to analyze house sales in a northwestern county.

The Data

This project uses the King County House Sales dataset, which can be found in `kc_house_data.csv` in the data folder in this repo. The description of the column names can be found in `column_names.md` in the same folder. As with most real world data sets, the column names are not perfectly described, so you'll have to do some research or use your best judgment if you have questions about what the data means.

It is up to you to decide what data from this dataset to use and how to use it. If you are feeling overwhelmed or behind, we recommend you ignore some or all of the following features:

- date
- view
- sqft_above
- sqft_basement
- yr_renovated
- zipcode
- lat
- long
- sqft_living15
- sqft_lot15

Business Problem

It is up to you to define a stakeholder and business problem appropriate to this dataset.

If you are struggling to define a stakeholder, we recommend you complete a project for a real estate agency that helps homeowners buy and/or sell homes. A business problem you could focus on for this stakeholder is the need to provide advice to homeowners about how home renovations might increase the estimated value of their homes, and by what amount.

Deliverables

There are three deliverables for this project:

- A **GitHub repository**
- A **Jupyter Notebook**
- A **non-technical presentation**

Review the "Project Submission & Review" page in the "Milestones Instructions" topic for instructions on creating and submitting your deliverables. Refer to the rubric associated with this assignment for specifications describing high-quality deliverables.

Key Points

- **Your deliverables should explicitly address each step of the data science process.** Refer to [the Data Science Process lesson](#) from Topic 19 for more information about process models you can use.
- **Your Jupyter Notebook should demonstrate an iterative approach to modeling.** This means that you begin with a basic model, evaluate it, and then provide justification for and proceed to a new model. After you finish refining your models, you should provide 1-3 paragraphs discussing your final model - this should include interpreting at least 3 important parameter estimates or statistics.
- **Based on the results of your models, your notebook and presentation should discuss at least two features that have strong relationships with housing prices.**

Getting Started

Start on this project by forking and cloning [this project repository](#) to get a local copy of the dataset.

We recommend structuring your project repository similar to the structure in [the Phase 1 Project Template](#). You can do this either by creating a new fork of that repository to work in or by building a new repository from scratch that mimics that structure.

Project Submission and Review

Review the "Project Submission & Review" page in the "Milestones Instructions" topic to learn how to submit your project and how it will be reviewed. Your project must pass review for you to progress to the next Phase.

Summary

This project will give you a valuable opportunity to develop your data science skills using real-world data. The end-of-phase projects are a critical part of the program because they give you a chance to bring together all the skills you've learned, apply them to realistic projects for a business stakeholder, practice communication skills, and get feedback to help you improve. You've got this!

Releases

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

Packages

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

Languages

● HTML 65.2% ● Jupyter Notebook 34.8%