Project Name: Zillow Regression Project
Project Type: Required partner project

Project Goals:
Construct a ML classification model that predicts tax value.
Deliverables: Presentation slide deck, Main Notebook for Pipeline process, Additional Information notebook with specific location and tax rate calculation information, readme.md, acquire.py, prepare.py
5 min Slide Deck presentation with partner

Stage	Tools	Brief Description of Process	Challenge Resolution
Plan	• Visual Studio	• Set up new GitHub organization for group work - CY Data Services • In Visual Studio create a new readme.md file to outline project plan • Import key elements and deliverables from curriculum requirements	• Corey used {{cookiecutter}} to create repo organization structure and default elements
Acquire	• Visual Studio • MySQL • .py script • SQL query • Query function	• In Visual Studio create a new acquire.py file to obtain data from Codeup Zillow database • Corey primarily responsible for this section • Corey researched “single property” definition and certain categories were excluded on this basis from SQL query • Query also limited to transactions between requested dates of May and June 2017	• No unusual challenges in this section

Prepare	• Visual Studio • .py script • Jupyter Notebook • Matplotlib	• In Visual Studio Corey created a new prepare.py file • I took the Additional Tax Rate info notebook and incorporated the features I needed to retain with the prepare.py file • Added an additional tax_rate column for the requested tax rate by property • Noted many outliers skewing the data • Found outside research to support dropping values with a calculated tax rate above 10% • Corey set up prepare.py to return multiple dataframes ready for use in different sections of the pipeline	• Debugging prepare.py • Sharing file via GitHub and resolving merge conflicts
Explore	• Jupyter Notebook • Seaborn • Scipy.stats • Matplotlib	• Created requested additional information visualizations using seaborn • Used pandas to find IQR and summarize results • Corey created Explore content in Main Notebook	• No unusual challenges
Model	• Jupyter Notebook • Sklearn	• Corey created content and iterations in Main Notebook	• No unusual challenges in this section
Evaluate	• Jupyter Notebook	• Corey provided model evaluation • I wrote Executive Summary	• No unusual challenges in this section

	• Sklearn		
Model Explanation	**How does your algorithm work?**	• Polynomial regression worked best • Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y • Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear. For this reason, polynomial regression is considered to be a special case of multiple linear regression	• No unusual challenges in this section
Delivery	• Google Slides	• Created slide deck in Google slides • Used graphics where possible to limit text information	• Originally planned to use Tableau but had to switch to Google slides