

Rubric

Your local instructor will evaluate your project (for the most part) using the following criteria. You should make sure that you consider and/or follow most if not all of the considerations/recommendations outlined below while working through your project.

Scores will be out of 27 points based on the 9 items in the rubric.

3 points per section

Score	Interpretation
0	<i>Project fails to meet the minimum requirements for this item.</i>
1	<i>Project meets the minimum requirements for this item, but falls significantly short of portfolio-ready expectations.</i>
2	<i>Project exceeds the minimum requirements for this item, but falls short of portfolio-ready expectations.</i>
3	<i>Project meets or exceeds portfolio-ready expectations; demonstrates a thorough understanding of every outlined consideration.</i>

The Data Science Process

Problem Statement

- Is it clear what the student plans to do?
- What type of model will be developed?
- How will success be evaluated?
- Is the scope of the project appropriate?
- Is it clear who cares about this or why this is important to investigate?
- Does the student consider the audience and the primary and secondary stakeholders?

Score 3/3

Data Cleaning and EDA

- Are missing values imputed appropriately?
- Are distributions examined and described?
- Are outliers identified and addressed?
 - Generally a good idea to look for some outliers.
- Are appropriate summary statistics provided?
- Are steps taken during data cleaning and EDA framed appropriately?
- Does the student address whether or not they are likely to be able to answer their problem statement with the provided data given what they've discovered during EDA?

Score 2/3

Preprocessing and Modeling

- Are categorical variables one-hot encoded?
- Does the student investigate or manufacture features with linear relationships to the target?
- Have the data been scaled appropriately?
- Does the student properly split and/or sample the data for validation/training purposes?
- Does the student utilize feature selection to remove noisy or multi-collinear features?
- Does the student test and evaluate a variety of models to identify a production algorithm (AT MINIMUM: linear regression, lasso, and ridge)?
- Does the student defend their choice of production model relevant to the data at hand and the problem?
 - When comparing scores of several models consider some way to easily look at them all. This will make your argument about what model is best a bit easier to understand.
- Does the student explain how the model works and evaluate its performance successes/downfalls?

Score 2.5/3

Evaluation and Conceptual Understanding

- Does the student accurately identify and explain the baseline score?
 - As above think about a small summary of all scores.
- Does the student select and use metrics relevant to the problem objective?
- Is more than one metric utilized in order to better assess performance?

- Does the student interpret the results of their model for purposes of inference?
- Is domain knowledge demonstrated when interpreting results?
- Does the student provide appropriate interpretation with regards to descriptive and inferential statistics?

Score 2.5/3

Conclusion and Recommendations

- Does the student provide appropriate context to connect individual steps back to the overall project?
- Is it clear how the final recommendations were reached?
- Are the conclusions/recommendations clearly stated?
- Does the conclusion answer the original problem statement?
- Does the student address how findings of this research can be applied for the benefit of stakeholders?
- Are future steps to move the project forward identified?
 - Come up with some discrete next steps. Bullet points are things people will be drawn towards and more likely to read.

Score 2.5/3

Organization and Professionalism

Project Organization

- Are modules imported correctly (using appropriate aliases)?
- Are data imported/saved using relative paths?
- Does the README provide a good executive summary of the project?
- Is markdown formatting used appropriately to structure notebooks?
 - Avoid having everything bolded in your conclusions
- Are there an appropriate amount of comments to support the code?
- Are files & directories organized correctly?
 - Have a directory of all your code
- Are there unnecessary files included?
- Do files and directories have well-structured, appropriate, consistent names?
 - Think about having “eda” or “data cleaning” in your notebooks that do that

Score 2.5/3

Visualizations

- Are sufficient visualizations provided?
- Do plots accurately demonstrate valid relationships?
- Are plots labeled properly?
- Are plots interpreted appropriately?
- Are plots formatted and scaled appropriately for inclusion in a notebook-based technical report?

Score 3/3

Python Syntax and Control Flow

- Is care taken to write human readable code?
- Is the code syntactically correct (no runtime errors)?
- Does the code generate desired results (logically correct)?
- Does the code follow general best practices and style guidelines?
- Are Pandas functions used appropriately?
- Are `sklearn` methods used appropriately?

Score 3/3

Presentation

- Is the problem statement clearly presented?
- Does a strong narrative run through the presentation building toward a final conclusion?
- Are the conclusions/recommendations clearly stated?
- Is the level of technicality appropriate for the intended audience?
 - Try not to explain too much about EDA such as dummies
- Is the student substantially over or under time?
- Does the student appropriately pace their presentation?
- Does the student deliver their message with clarity and volume?
- Are appropriate visualizations generated for the intended audience?
 - Small point but arrows next to your font are hard to see
 - Heatmaps are generally not well received for a non technical audience
- Are visualizations necessary and useful for supporting conclusions/explaining findings?

Score 2/3

In order to pass the project, students must earn a minimum score of 1 for each category.

- Earning below a 1 in one or more of the above categories would result in a failing project.
- While a minimum of 1 in each category is the required threshold for graduation, students should aim to earn at least an average of 1.5 across each category. An average score below 1.5, while it may be passing, means students may want to solicit specific feedback in order to significantly improve the project before showcasing it as part of a portfolio or the job search.

Instructor Feedback

Patrick: I like the appraisal contractor direction. I like that you looked at a histogram of the home prices. I also like that you showed the baseline model. I think you needed to put the actual correlation coefficients in the heatmap. It would have probably been a little bit better if you just showed the features and their correlations with sale price and not with each other. I think you have a nice flow to your presentation. I really love that you included your residual plot. I think that this is fantastic and something that really should be included in all of your future presentations / projects. I would have really liked to see lasso or ridge regression in your model. Overall nice job.

Kelly: I think you took an interesting perspective for this project! It sounds like you give us a good overview of your presentation, too, but it could be helpful to have an overview/agenda slide to support that. Great idea to include the distribution of your target variable. Heatmaps are a great visual for EDA, but not necessarily great for presentations, as they can be hard to interpret or lack specific information you may want to share about correlations. Also try taking only the “sale price” column of the heatmap, since that’s what we’re interested in! I think it’s great how you identified outliers and explained them for a non-technical audience. It’s great that you explained the term “dummify” with an example, and I think it’s awesome that you gave us a visual of our “goal” with creating a model. Is that score your train or test score? What does 90% mean? What recommendations do you have, based on your model, for the house flippers? Overall, great job!

Peer Feedback

What are some positives from the project? ("glows")	What are some areas that could use improvement? ("grows")
Nice pace! Seemed calm, nicely rehearsed! Great looking visuals!	Don't know

Super well designed! Clearly explains graphs and charts. You got TOUGH questions and you handled them well!	Practice to sound more conversational and less like reading.
You did a great job pacing your speech and spoke very articulate	I feel you went a bit too much into explaining statistical analysis
Love how you set the scene. Level of technicality is appropriate to demonstrate your analytical expertise to potential clients.	Heatmap - change the names of the x and y axis to be more readable. Heatmaps in general are a lot to take in and may not be the best way to convey information in presentations. How else might we frame the same information via a visual?
very clear where you can fit in and provide assistance at beginning of the presentation	I think the histogram of distribution got a bit too technical given the audience.
Well laid out problem statement. Well spoken	A little confusing at the end. Maybe go back to the problem statement.
Great plots and problem statement. Strong presentation skills.	Only thing I can think of is that the r^2 score was not explained in the slide to the non-technical audience. Well done!
Nice flow in presentation. Interesting and well prepared.	n/a. Good job!!!
No negatives. Good Job	
Model and data interpretation	slide background too busy
I liked the actual vs predicted graph, I should have made one of those!	I think the heatmaps were underexplained. Maybe reduce the columns to simplify?