

Deliverable	Sections	Description	Point Breakdown	Maximum Possible Points	Graded Points
EDA and Inference Presentation	EDA	Your presentations should include at least 3 data visualizations that either help people to better understand the context of the data or the insights that you bring out. For example, you might show a boxplot of the bedrooms and then talk about how it helped you to identify outliers/bad data and led you to clean it up. For another example, you could have a plot of housing prices by grade, to show that it is feature that will help explain the variance in the price. While htis will be graded during the presentation, it should also be present in your final notebook.	1 point for each visualization	3	
	Data Inferences	In your EDA/Inference presentation, you must give at least three actionable insights that you learned from the data. These insights need to be supported with either a visualization, statistical test, or regression analysis. It must also be clear how this insight can be valuable to the audience.	1 Point for each insight	3	
Final Notebook	Executability	The final notebook can be run beginning to end without any errors.	2 points	2	
	Data Prep	The student applied some data cleaning techniques to prepare the data to be modeled. This could take multiple forms including handling extreme/outlier values, turning categorical data into dummies, changing the type of the data (string to int). The process is clearly labeled in the notebook and includes a brief explanation.	1 point for each instance of data preparation ie: 1 point for turning categorical data into dummy 1 point for handling extreme values for a column	3	
	Feature Engineering	The student engineered at least three features to test in your model. Some of these features might not be included in your final model, but we want to see that you at least created them. Turning a categorical column into dummies, does not count as feature engineering. However, creating polynomials and interactions does count for one of them. The process is clearly labeled in the notebook and includes a brief explanation.	1 point for each engineered feature	3	
	Statistical Tests	Three statistical tests where performed on the data and the results of those tests should be interpreted. All of the steps in hypothesis testing should be clearly laid out in the notebook. State the null and alternative hypotheses. Calculate the test statistic and p-value. Make a final decision about the null hypothesis and explain how this could impact your modeling process.	1 point for each test	3	
	Feature Selection	The student implemented at least one feature selection technique into their modeling process.	2 points for each technique you implemented	4	
	Model Evaluation and Selection	Each student should take an iterative approach to arriving at their best model. Their notebook should show that they ran multiple models and evaluated them individually to determine if they should take additional data preprocessing steps and determine their final model to be used on the holdout data.	2 points for checking assumptions and interpreting the output 1 point for implementing train test split 1 point for evaluating the RMSE of multiple models 1 point for explaining how you chose your final model according to the evaluation metrics	5	
Holdout Notebook	Executability	The final notebook can be run beginning to end without any errors.	2 points for creating predicitons	2	
	Holdout Predictions	The major goal of this notebook is to be able to predict the prices of the houses on the holdout set. You will recieve 4 points if the RMSE of those predictions is below \$200,000	4 points for the RMSE being below \$200,000	4	
Github repo	Github	Is your Github repo for this project clearly organized and have all of the necessary components? Read.me , final notebook, holdout notebook,a folder(s) with all supporting data and images.	2 point for Read.me 1 point for organizing files into folders	3	
				35	24