1. Hi Kevin! I hope you are having a great time with your family during the holidays. I am sending you a couple of comments about your project: 1. One of the Section Summary parts has this sentence repeated twice: The data type of the 'quality' column is integer. The data type of the quality column is an integer. 2. In Hypothesis Test: Mean Free Sulfur Dioxide Values and Quality and I think also further, you are checking if the variance between the two groups is equal. It’s very nice that you thought about testing this assumption. However, you check if they are exactly equal, e.g., if one value is 2.5 and another 2.6, then the conclusion is that they are not equal while the difference is negligible. It is enough if they are approximately equal – you can use another statistical test to check this. For example, Leven’s test. 3. I was not sure why you did not include quality in the correlation tables and why you looked at the correlations for each separate score instead of merging them. I thought this would be a bit more in line with your further goal – predicting the quality and alcohol. When building models, the EDA part is often used to see if the features are associated with the outcome variable and allow predicting it. When excluding quality from the table, this information is left out. 4. Scaling of the data can also be included in the pipeline, you had it as a separate function. 5. For alcohol prediction, I would also include RMSE as a performance metric. It indicates how imprecise your predictions are from the real value. R2 only determines the proportion of variance in the dependent variable that can be explained by the independent variable.
2. [10:54 PM]

6. I was curious how did you decide to make a good wine cut-off at value 7? Maybe it was an arbitrary decision. That is fine, but it’s always better to provide some sort of reasoning or maybe back it up with the data. In this case, I think that the cut-off makes a lot of impact on the performance score. If you would have a different cut-off, the score would likely be different and lower. Therefore, it seems like an important decision and worth explaining, in my opinion. Other than that, the project looks great!