Sarah Shupe

DSC530

2/28/23

Wine Quality Analysis with Jupyter Notebook

For my analysis, I looked at a data set on wine quality in Northern Portugal. Many factors were recorded, including my selected variables: pH, fixed acidity, alcohol content, sulfates, and residual sugars. The wine then received an overall quality score, which was the explanatory variable in my analysis. I hypothesized that one of the chemical properties would have a relationship showing correlation to the wine's quality score. Overall I was not able to support my hypothesis with my analysis, as none of the correlations turned out to be statistically significant. I feel like my limited understanding of the concepts was the biggest limiter in this experiment, as I am still learning many of the concepts. Writing the code was less difficult for me than interpreting the results, so that is where I will focus my time in the future. I had a difficult time selecting a data set, as I had originally intended to choose a marine concept such as overfishing or carbon dioxide uptake but there was very limited data available to the public, and even fewer sets with complete data. This makes me feel even more motivated to pursue data science in the realm of marine science and contribute to that data so other people can have access to it. I also struggled a lot this semester with time management as I took on too many responsibilities and could not balance them all, but it has been a lesson and I will continue to improve in the future. Overall I feel my code and graphics were good, my analysis could have been more thorough if I understood the topics better, but I am satisfied with my final product. I have a long way to go but have learned an immense amount since starting this course, and I look forward to seeing where data science takes me.