Alissa Trujillo

Professor Parajulee

DSC 530

8 August 2022

What Makes a Good Wine? An Exploratory Data Analysis

For my final project in this course, I used a dataset from Kaggle regarding the quality of red wines. I was ultimately able to see which variables were important in determining the quality of a red wine, and which variables were not. I created a regression model that showed volatile acidity, citric acid, density, pH and alcohol content all to have statistically significant effects on a wine’s quality. I was also able to determine that residual sugar did not have a statistically significant effect on a wine’s quality.

I believe this analysis would be richer if there was data on white wine included as well. I am curious to see whether the factors that determine quality are the same for both red wine and white wine. It would be important to create a model for each type of wine separately first before looking at them together, just in case there are negating effects in play. For example, if alcohol content were to have a negative correlation with quality in white wines but a positive correlation with quality in red wines, it would nullify the effect in a model that includes both.

The main challenge that I faced was trying to determine which variables to focus on. The dataset included many other variables that I was not able to dissect and create histograms for. Using the insight I had from the variable descriptions and my own personal knowledge of wine, I was fairly confident in my ability to choose the most potent variables. However, I would love the opportunity to look deeper into other variables in the dataset and add analysis of white wine quality as well.