DS 740 Final Proposal

Paul Isaacson

December 2, 2017

I will be analyzing the Wine Reviews data set (<https://www.kaggle.com/zynicide/wine-reviews>). I will be using v2 of the data set that has 130k observations and 13 variables. I will be looking at predicting an unknown wine’s point value (like you would see in Wine Enthusiast) based the 12 (or a reduced amount) of the other variables. I may, depending on results of my data exploration, try to do the analysis based on a particular wine type (e.g. Merlot). I feel like a score is a good indicator of the quality of a wine, but the differences between a white and red are great no matter what the score and the equivalent of comparing apples to oranges. This model could be useful of wine retailers such as Total Wine in building a profitable wine portfolio. They could potentially find a quality obscure wine at a good price.