**1001 Team Project**

1. Task distribution
   1. Business understanding - Ben
      1. Identify and motivate the business problem that you are addressing.
      2. How (precisely) will a machine learning solution address the business problem?
      3. What is the size / magnitude of the opportunity?
      4. What is the current state of the art in solving the problem?
   2. Data understanding
      1. Discussion on variables - Zixiao
      2. EDA -S
      3. Clustering - Ben
   3. Data Prep
      1. Threshold of wine quality -Zixiao
      2. Outliner -Zixiao
      3. Multicollinearity -Zixiao
      4. Red or white together (do both, compare the two, create a new dataset)
      5. Define the target variable -David
      6. Heat map between variables (correlation) - David
      7. Drop NA
      8. Standardize
   4. Modeling and evaluation
      1. Discuss choices for data mining algorithm: what are alternatives, and what are the pros and cons?
      2. Identify an appropriate baseline model and report its performance.
      3. Describe an evaluation framework you will use to improve upon the baseline.
      4. Perform an analysis of possible algorithms and use the data science experimental framework to choose an optimal candidate.
      5. Demonstrate how you were able to improve upon the baseline and document the process of doing so.
      6. Discuss why and how this model should “solve” the business problem (i.e., improve along some dimension of interest to the firm).
      7. Discuss the type of evaluation metric that should be used to choose the best algorithm. How does this metric relate to the business problem?
   5. Deployment
2. Classification or regression

Classification

Regression

1. Red or white together