ML Model Outline

1. Datasets

The following data sets are for the following provinces: California, Washington, Bordeaux, Tuscany, Oregon, Burgundy, Northern Spain, Piedmont, Veneto, New York (may need to drop), Alsace, Sicily & Sardinia, Champagne

* 1. Wine reviews
  2. Historical mean temperature
  3. Historical precipitation

1. Features and Target
   1. Features from wine data: Country, Description, Designation, Price, Province, Region\_1, Region\_2, Title, Variety, Winery
   2. Features from weather data: Year, Rainfall, Temperature, Timeseries
   3. Target: Points (wine rating)
2. Types of Data Cleaning
   1. Wine data
      1. Remove rows with provinces that aren’t included in the project
      2. Use regular expressions to pull year from Title field
      3. Perform feature selection (TBD)
      4. Drop rows with null values
   2. Weather data table requires no cleaning
3. Preprocessing
   1. All input data is tabular
   2. Merge Wine data and weather data
   3. Drop rows with null values
   4. Split data into input (X) and output (y)
      1. X – Features defined above
      2. y Target defined above
   5. Split X and y into training and test datasets
4. Train and Evaluate Model
   1. Multiple Linear Regression
      1. Make regression
      2. Fit the model
      3. Predict wine quality
5. Revaluate as necessary