1. EDA
   1. Data cleaning
      1. Null values in each column, remove columns which are 100% useless
      2. Research what to do with null values in acceptable and try multiple methods => choose which seems the best
      3. Duplicate rows
      4. Make sure all columns are of the correct type
      5. Discrete variable = apply OHE
      6. Target variable convert to 0/1 values
      7. Component breaking direction columns (WindDir3pm', 'WindDir9am', 'WindGustDir)
   2. Plots :
      1. Correlation matrix (heatmap)
      2. Outliers
      3. Boxplots
   3. PCA + plot => linearly/non-linearly separable?
2. Models
   1. Naive Bayes - base model. All other models after NB. Save the results for future comparisons.
   2. Linear models
      1. Logistic regression
      2. SVM
   3. Decision trees, random forests, AdaBoost

