DS 1001 Project Outline

1. Data Preparation
   1. Synthetic Data (Done): data validation
   2. Data Cleansing:
      1. Change the binary class and multi-class variable feature columns
      2. Change column names
      3. Merge target variables (5 columns to 1 column)
      4. Check missing values
      5. Final validation
2. Feature select:

Data Overview:

1. DefendantID: unique ID for each defendant
2. Target Variable: NewArrestType (last column) multi-class variable
3. Features: 74 in total, including binary class, multi-class and numeric