Machine Learning Break Down

Data Carpentry (Data Munging/Data Wrangling):

- Get the data
- Clean the data
- Make data numeric
    - One hot encoding, make it binary, or integers somehow, etc

EDA (Exploratory Data Analysis):

- See it, visualize it, use charts, histograms, correlations, word clouds, etc
- Do at least 10+ ways to look at data

Feature Selection and Feature Extraction

- Dimensionality Reduction
- PCA, FA, LDA
- Transform the data

Make a Model

- Train, Test, Split for validation (do the 80% train and 20% test split)
- Assign model
- Fit the model
- Make ensemble

**** You want a high score, F-1

****** if low go back to last step

Save Model

- Make a Pipeline to save the model and then you can call it back
- Now Apply to new dataset. Just change the parameters
- Test for good scores