**Bare Data Science**

# Data Science Process

1. Develop Problem Statement
   1. What problem are you trying to solve?
   2. Develop scope
2. Collect Data
   1. Directory structure
3. Data Preprocessing
   1. Inspect data
   2. Create a guide – field definitions
   3. Clean data
   4. Missing data
      1. Delete row
      2. Impute missing values
   5. Standardize data
4. Exploratory Data Analysis
   1. Discovery visualization
   2. Univariate analysis
   3. Multivariate analysis
   4. Feature engineering
      1. Categorial variable
         1. Label encoding
         2. One-hot encoding
         3. Get dummy variables
   5. Feature selection
5. Build Model
   1. Separate dependent (y) and independent (X) variables
   2. Separate X and y into training and test sets
      1. You want to include a validation set
   3. Model Selection
      1. Classification
      2. Regression
6. Implement Model
   1. Model Evaluation
      1. Model Fit
         1. Underfit
         2. Overfit
7. Deploy Model
8. Present Data and Findings
9. Evaluate Results and Refine Process