1. Data Preparation
   1. Original data -> synthetic data
      1. Where did the data come from? (any selection bias)
      2. Why we need to get a synthetic data and how we get that (R code)
      3. Validate that the data we use to build the model represent the original data
   2. Synthetic data to clean data
      1. Data Cleaning
         1. Missing values
            1. What to drop and what cannot be dropped? Why
            2. Some domain knowledge needed to be included
         2. High variance features (drop)
      2. Data Validation
2. LR models
   1. Feature exploration
   2. 1st model
   3. feature selection (2 methods)
   4. 2nd model