TODO list

Outline:

Introduction/Background/Scenario modelling Data Dataset Exploratory Data Analysis Data preparation Feature Engineering Analysis Type - binary classification with multivariate Learning Algorithm Selection and Discussion Evaluation Metrics Results and Discussion Limitations, Future Work

* [] write intro / brief / business problem
* ~~[x] write up data processing section - long~~
  + ~~use v1\_2 for decision making process file - include graph and EDA from working data~~
  + ~~use v1\_1 for just code loading -> fix ASSESSMENT section~~
  + ~~rename appropriately~~
  + ~~remove working\_data~~
  + ~~remove data\_v1~~
* ~~[x] check data processing functions~~
* ☒ do eda plots
* ☐ eda long write up
* ☐ write eda section
* ☐ modelling / analysis working section

NEXT - VERSION 2:

* ☐ ~~V2\_data\_initial~~
  + include correlations
  + ~~keep brief~~
  + ~~decisions re data ~~
* ☐ ~~V2 data prep code with narrative~~
* ☐ V2 EDA
  + This section should be tight, but comprehensive.
  + It is the exploration of the data produced in the the initial prep section:
    - include correlations
    - pair plots
    - histograms
    - distributions
    - It will have reasons which spill into the Feature enginering section - like why not to include a feature, eg. region, etc.
* ☐ V2 Feature Engineering
  + Features to exclude - all the student biographical data - the reason why (not relevant to the problem)
    - ANOVA and t-tests
    - but it could be interesting to do cluster analysis or more indepth analysis with this data - but not for now. keep it tight!
    - include deletion of region, etc.
  + include splitting, scaling - as initial transformations
  + Dimension reduction:
    - include PCA
    - include VIF, feature reduction
* ☐ V2 data prep code only:
  + include splitting, scaling, dimension reduction -> as needed for modelling
* ☐ V2 optional (do later) - clustering - make a note but this won’t happen now, link to V1 attempts
  + Kmeans
  + Hclust
* ☐ V2 Prediction modelling section
  + keep it tight