* Condensing large categorical variables
  + Non-negative matrix factorization?
  + ~~CatBoost~~, ~~LightGBM~~, ~~H2O GBM, H2O Random Forest~~
  + Combine based on *a priori* knowledge
  + Combine based on frequency of occurrence
  + perform a cluster analysis to identify the levels that behave similarly against the dependent.
  + You can use Factor Analysis as well for dimension reduction
  + let a fairly shallow decision tree handle it

CatBoost, H2O, xgboost kept superseding each other on openssl and/or ca-certificates. May need different environments for them??

1. Finish checking cloned repository w/ local, update from local to clone any differences
2. Move data files outside of cloned directory, change all python files to load data from full path to files outside directory
3. Create new pyenv with python 3.8.3
4. Install tensorflow with new python environment in
5. Install tensorflow mac update into new tf environment
6. Look at installing pytorch
7. Look at installing PlaidML
   * 1. How to condense large categorical variables?
     2. How to group drug codes?
     3. How to group diagnosis codes?
     4. Re-run drug combo analysis and chart events with updated data merging (settle on best way to merge data across files)
     5. Game plan notes analysis
     6. Outline ensemble model format
     7. Install neural network architectures (tensor flow?, pytorch?), ~~xgboost~~, others?
        1. Set up different environments?