conda install graphviz

pip install gvmagic

2 new features:

* Start with an A
* Up to 4 characters

Process

* Define the problem (+ hypothesis)
* Collect the data
* Tidy the data
* Split train/test
* EDA
  + You can’t ask a machine to understand the data before you do
  + Plots/significance testing/summary statistics
* Modeling
  + Feature engineering (selection)
  + ML algo selection/metric selection
  + Hyperparameter selection?
    - Train/cross-validate the model
  + For hyperparameter1 in (….)
    - For hypparameter2 in (…)
      * Run k-fold(param1, param2, …)
      * Record model
  + Take the model that generalize the best
  + Retrain the model with the entire training set with selected features, and model
  + Are you sure?
  + Test
  + Retrain the model with the entire dataset with selected features, and model
  + Ship!