• ~~Code target variable into binary (0 and 1) and make sure there are enough data points in each outcome~~

~~• Deal with missing values~~

~~• Drop columns:~~

~~V18q1 – number of tablets owned~~

~~Rez\_esc – years behind in schooling~~

~~Dependency – dependency rate~~

~~• Change columns:~~

~~Edjefe – change all “no” to 0~~

~~Edjefa – change all “no” to 0~~

Determine if Target binary counts are OK (6k vs 3.5k)

Handle NaN values in first column

• Exploratory Analysis w/ visuals

• Correlation matrix

• Distribution of outcomes

• Etc.

• Feature selection

• Split test/train data

• Run Statistical Model on training data

• Run statistical model on test data

• Graph of predictions vs. actual

• Run machine Learning Models on training data

• Cross validation

• Predict on actual test data

• Graph of predictions vs. actual