1. Code target variable into binary (0 and 1) and make sure there are enough data points in each outcome
2. 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

1. Exploratory Analysis w/ visuals

* Correlation matrix
* Distribution of outcomes
* Etc.

1. Feature selection
2. Split test/train data
3. Run Statistical Model on training data
4. Run statistical model on test data
5. Graph of predictions vs. actual
6. Run machine Learning Models on training data
7. Cross validation
8. Predict on actual test data
9. Graph of predictions vs. actual