Download data from: [Link](https://drive.google.com/file/d/1BBez0hfHQJAA1lmPVEELHiBmOvbJx7xj/view?usp=sharing)

Examine the data on martial arts fights. The data includes columns about the fighters (R and B, the column names start with those letters) and the fight itself. The target variable is a column named ‘winner’. Prepare the data for classification modeling.

* Clean the data
* Handle missing values
* Turn into dummy variables
* Split into train and test
* Save the data into train.csv and test.csv

When you clean the categorical variables, keep the top 70% of the values, and assign the 30% to a new value ‘other’. This way you will not have too many dummy variable columns.

You can take a sample of the columns to work on, if your hardware isn’t sufficient to run the cleaning.

You don’t need to train a model.

Handle missing data for continuous columns only. Assume we will use a model that can work with missing data.

Bin the data for columns ‘R\_Weight’ and ‘B\_Weight’.

Do not include unique indicator columns, such as the name of the fighters. Unique indicator columns don’t give any useful information to the model.

Commit the cleaning codes to GIT.

Use gitignore, in order not to commit data to the GIT repository.

Add a Readme and requirements files to GIT.