Exercises - Data exploration, cleaning and feature engineering

VIA University College

September 8, 2020

## Exercise 1 - data exploration and feature engineering

Download the dataset "500\_Person\_Gender\_Height\_Weight\_Index.csv".

Explore relationships between variables with FACETS Dive. Do you expect a decision tree to make good predictions?

Now download the notebook "Exercise - feature engineering.ipynb", and solve the exercises posed in the notebook.

## Exercise 2 - text data

Play around with the notebook "Movie Reviews". See if you can improve the accuracies and reduce the amount of overfitting. You can for example try to remove the so-called "stopwords" (see page 334-336 in the book), or try training a random forest or gradient boosted trees instead of a single decision tree.