DATA CLEANING

Data cleaning

- Unfortunately, real world datasets can’t be directly used for making predictions.
- They need some data cleaning steps in order to process them and make it suitable for machine learning.
- Each dataset will be different and it will require different data cleaning steps

Data cleaning

- Standardize column names
- Standardize units of a single column
- Deleting and rearranging columns
- Working with data types (setting the right types)
- Filtering data
- Removing duplicates (redundant, they don’t add info)
- Correct typos
- Replace/drop missing values (most of models are not able to handle them)

IRON
HACK