HW 1
Deadline – Friday, April 5

Go to the link https://github.com/JWarmenhoven/ISLR-python/tree/master/Notebooks/Data and download all available datasets as "csv" files: Advertising, Auto, Boston, Caravan, Carseats, Credit, Default, Heart, Hitters (all versions), Khan (all versions), NCI60, SMarket, USArrests, wage.

For each dataset prepare the following:

1. Read the description of each dataset to understand the problem and identify the inputs (features, variables) and the output (response).
2. For each dataset find the number of observations and variables.
3. Which dataset corresponds to regression, classification or clustering problem.
4. Which variables are numerical, and which are categorical. For numeric variables find the classical statistical measures. For categorical data find the number of different levels.
5. Draw pairwise scatterplots for numeric variables vs the response. Add also the linear regression line. Is there any relationship? Is the relationship linear or non-linear?
6. Draw boxplots for categorical variables vs the response. Is there any relationship between the levels and the response?
7. Which dataset contains missing values? How to find them and exclude from the datasets? Do it.