Assignment 1

1   Go to https://www.kaggle.com/datasets
2   Download a dataset to your liking (make sure the dataset can be used for a classification task for the later assignments) (5)
3   Identify the different types of attributes (numeric, nominal etc.) (15)
4   Identify the mean, median, mode of the numeric attributes, if any. Also identify which attributes are positively/negatively skewed. (do for at least three numeric attributes) (20)
5   Compute the IQR for the numerical attributes. Based on the IQR, determine the outliers. Decide whether in the context of the dataset, if you are going to keep the outliers or not. Provide arguments. (do for at least three numeric attributes) (20)
6   Use scatter plots to determine if there's correlation between the numeric attributes (do for at least two pairs of numeric attributes) (20)
7   Finally, draw boxplots for the numeric attributes (do for at least three numeric attributes) (20)

What to submit

1. A compressed folder with
    a. The dataset file
    b. The script/code that you used to complete the assignment
    c. A report with the findings (pdf format)
    d. Make sure the report says assignment1_x_y.pdf, where x and y are your name/s
    e. Make sure I can run the script right away without any errors. The folder should be self-contained.

Due Date: February 6, 2022, 11:59 PM