CSIS3764

Assignment 2 **Final Submission Date: 28 Aug 2023**

1. Pre-processing data for numerical analysis

1.1 Instructions

1.1.1 Download the dataset *netflix.csv*.

1.1.2 Do some Exploratory Data Analysis that includes the following:

- Apply descriptive statistics.
- Inspect the dataset regarding its columns, shape, content, and correlation between the features.
- Plot a graph that indicates the number of subscribers per age group as well as the number of subscribers per country.
- Also create a scatter plot between the subscription type and gender as well as a scatter plot between the subscription type and country.
- Discuss possible conclusions drawn from the graphs.

1.1.3 Apply the concepts and techniques discussed during the sessions on pre-processing to create a Jupyter Notebook that contains the Python code to clean and prepare the dataset so that it can be used for numerical data analysis.

1.1.4 The Jupyter Notebook should be able to receive the data from the *netflix.csv* dataset file and store the final cleaned and pre-processed data in a new dataset file called *clean_ netflix.csv*.

1.1.5 Submit the Jupyter Notebook on Blackboard to complete the assignment.

1.1.6 The final submission date for the assignment is 28 August 2023 at 20h00.

[50]

2. Submitting your assignment

This assignment can be submitted from 18 August 2023 until 28 August 2023 at 20h00. If your assignment is submitted late, you will receive 0%. If your assignment is not submitted, you will receive an incomplete for CSIS3764.

To submit the assignment, compress the Jupyter Notebook into a ZIP file. Name the ZIP file: “Ass2_YourStudentNumber.zip” (Replace “YourStudentNumber” with your actual student number). Then go to Blackboard and submit the assignment.