**SCHOOL OF COMPUTING**

**Programming for Data Science**

**Self-Reflection (CA1)**

|  |
| --- |
| **Instructions:**   1. Submit the reflection as an item of your CA1 submission 2. Name your file “YourModuleClass-YourStudentID-YourName\_reflection.docx” |

|  |  |
| --- | --- |
| **Name** | Nabil Ridhwanshah Bin Rosli |
| **Student ID** | 2007421 |
| **Module Class** | EL/EP0302/FT/01 |

# QUESTION 1: CHALLENGES - SELF-REFLECTION FOR CA1

Provide a brief reflection of the challenges you have faced in this assignment.

|  |
| --- |
| One of the biggest challenges I faced was getting used to the syntax of Python, which is quite different from JavaScript. For example, I had to learn how to use indentation to define code blocks, and I had to get used to using colons and white space in a way that is different from JavaScript.  The second challenge I faced was learning the ins and outs of NumPy, a powerful library for scientific computing in Python. I had to learn how to use NumPy arrays, how to perform mathematical operations on these arrays, and how to use NumPy's many functions for working with arrays and reading data from csv files. This was a steep learning curve for me, but I was able to overcome it with practice and by seeking help from google (along with YouTube and  DataCamp).  The third challenge I faced was while working with data, was determining which type of graph was appropriate for each dataset. With so many different types of graphs available, it was often difficult to know which one would best represent the data and provide the most meaningful insights.  Through trial and error, I learned which types of graphs were most effective for different types of data, and I became more confident.  The fourth challenge I faced during this project was data cleaning. The dataset I was working with had a categorical column with commas in it, which caused the CSV file to have extra rows for some data points. This made it difficult to accurately process and analyze the data. I had to spend some time identifying the problem and finding a solution.  The last challenge I faced during this project was ensuring that the data was correctly formatted and processed. One of the key factors in achieving this was being careful about the dtype argument when reading the CSV files. I discovered that if I passed the wrong dtype, some of the numbers in the dataset would turn out to be negative, even though they were supposed to be positive. This was because the dtype I passed was too "small" to accurately represent the data, which caused it to overflow into the negative range. I learned that it is crucial to carefully consider the dtype argument when reading CSV files, and to make sure that it accurately reflects the data in the dataset. By doing so, I was able to avoid this issue and ensure that the data was correctly processed.  Through this experience, I learned that it is important to be persistent and to not give up when faced with challenges. I also learned the value of reaching out to google and YouTube for help and guidance when needed. Overall, I am glad that I took on the challenge of learning Python and I am excited to continue exploring its many capabilities. |

# QUESTION 2: ACHIEVEMENTS - SELF-REFLECTION FOR CA1

Provide a brief reflection of what you think you have personally achieved in this assignment or the knowledge or skills you have found satisfaction in learning / acquiring. Indicate all the online courses you have taken.

|  |
| --- |
| In my opinion, the biggest achievement in this assignment was that I was able to answer the questions I had at the beginning of the project using data analysis and visualization.  By creating and interpreting graphs and plots, I was able to gain insights and identify trends and patterns in the data that helped me answer my questions.  This process of exploration and discovery is a crucial aspect of data science, and I am proud of the progress I made in this assignment.  I took a series of lectures on numpy and pandas and was able to successfully apply what I learned by working on a hobby project (and assignment) where I plotted graphs using matplotlib.  One of the graphs I created was based on the relationship between a movie's budget and its rating, and I used datasets from Kaggle to clean and organize the data.    *There was in fact, no correlation between movie production cost and ratings*  I am constantly amazed by the capabilities of the matplotlib library for creating visualizations. Despite its quirky syntax, matplotlib allows me to quickly and easily generate high-quality graphs and plots. Its powerful features enable me to perform exploratory data analysis and gain insights from my data in a visually appealing way.  I am grateful for the opportunity to learn about data science and the Python programming language, and I am confident that these skills will continue to benefit me in my professional and personal endeavors.  These are stuff I managed to pick up via DataCamp by completing the “Introduction to Python” and “Intermediate Python” lessons and YouTube tutorials and videos by Rob Mulla who posts videos on data science driven by curiosity (it was also the inspiration for me to look for datasets on Kaggle)  Overall, I am grateful for the opportunity to learn about data analysis and visualization, and I am excited to continue using these tools and techniques in my future work. |

**-- End of Self-Reflection --**