**SCHOOL OF COMPUTING**

**Programming for Data Science**

**Self-Reflection (CA2)**

|  |
| --- |
| **Instructions:**   1. Submit this together with your other deliverables at Blackboard “Assignments->CA2” folder 2. Name your file “YourModuleClass-YourStudentID-YourName.docx” |

|  |  |
| --- | --- |
| **Your Name** | Chai Pin Zheng |
| **Your Student ID** | P202633 |
| **Your Class** | EL/EP0302/FT/02 |

# QUESTION 1: CHALLENGES - SELF-REFLECTION FOR CA2

Provide a reflection of the challenges you have faced in this assignment (CA2) with a comparison of those you faced in CA1. Did you manage to apply any learning experiences from the previous assignment to improve the way you did this assignment?

|  |
| --- |
| For this assignment, I faced lots of issues with Pandas. Though I had prior experience with Pandas, I was already rusty at this point. As such, I had so much trouble trying to access rows and columns. For instance, when using NumPy, I could access the last row using data[-1]. However, with Pandas, it was completely different. Instead of simple array indexing, I had to use 'iloc'. It wouldn't be the case if I fully have understood how 'iloc' works in the first place. As such, when trying to access the last row, I would keep doing data.iloc[-1: 0], whereas it should be data.iloc[-1]. I was too used to NumPy as I had used it for the entirety of assignment 1. Hence, I was always confused until I watch some tutorials, which managed to answer my questions.  Another challenge I faced is finding relevant datasets, which was the same case for assignment 1. As my topic revolved around finance and education, finding data on education was extremely hard to come by. Moreover, few external datasets were relevant to education, especially when it revolves around Singapore. Hence, a significant amount of time was spent adjusting the thesis statement, as the lack of good data was a significant limiting factor. As such, this caused lots of headaches when researching and selecting datasets, as there were either a lack of a quantitative dataset or the absence of a dataset to support my thesis. However, through hard work and the lesson learnt from assignment 1, I did manage to narrow down a few datasets that support my thesis. I did use the knowledge gained from assignment 1 to help quickly rebound myself.  The last personal challenge I had was when plotting for data visualization. Though I had personally heard of Seaborn before, I had never really used it. Hence, I did not know the hidden quirks of Seaborn, unlike Matplotlib where I had used extensively in assignment 1. As such, using what I had learnt from assignment 1, I quickly browse through StackOverflow forums. It helped me by teaching me how we can customize the graphs. Through this, I have also learnt that figure annotations can be customizable as well. I learnt to bold, underline, and italicize texts. It made the graph I produced have more depth.    Figure of the graph I managed to plot using seaborn |

# QUESTION 2: ACHIEVEMENTS - SELF-REFLECTION FOR CA2

Provide a reflection of what you think you have personally achieved in this assignment. Do not be shy to highlight extra efforts you have put in for this assignment. This is the place for you to impress upon your tutor on your competency in this area.

|  |
| --- |
| A personal achievement I have gained is learning how to use Matplotlib and Seaborn more effectively to produce visually stunning graphs. I felt that the graphs I have plotted for assignment 2 are better in comparison to assignment 1. In assignment 2, I placed more emphasis on creating visually stunning graphs. Hence, I would spend time researching the tips and tricks on how to do so. For instance, I would read articles on 'Medium' to see what the professionals do. Hence, it helped build my fundamental knowledge when it comes to Matplotlib and Seaborn. Moreover, I also got to learn the nitty and gritty details of Matplotlib and Seaborn too!  The article can be found at: <https://medium.com/@dey.mallika/transform-your-graphs-with-seaborn-ea4fa8e606a6>  In addition, I feel more comfortable when coding in Python as compared to assignment 1. In my previous assignment, as my fundamental knowledge of Python was at a novice level, my for loops always relied on indexing. Hence, this made the code very complex and difficult to understand, as mentioned by my tutor, Mr. Qiu. As such, this round of the assignment, as I have become more comfortable with Python, I have managed to reduce the code complexity. Hence, I managed to cut down on the redundant code. Moreover, with the aid of Pandas, I did not have to create so many arrays and dictionaries. Besides, Pandas has a lot of great functions, such as sort\_values(). Hence, I do not have to code from scratch for loops to sort by date or values.  The documentation I am referring to is: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sort_values.html>    Figure of the documentation |

# QUESTION 3: EVALUATE YOUR CA2

What grade do you think you deserve for CA2? Justify the grade you gave yourself with reasons.

|  |
| --- |
| What grade do you think you deserve for CA2?   * A grade   State your justifications here   * Time spent into this module is quite substantial. * Graphing has improved and is well drawn. * Datasets have substantial link and connection to one another, thus allowing a seamless transition from one dataset to another dataset. * External usage of different graphing libraries. |

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