**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** | Yu Dong En |
| **Your Student ID** | P1936348 |
| **Your Class** | DIT2A02,EP0302\_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?

|  |
| --- |
| The challenges I faced was trying to optimize the program for execution. I was used to doing for-loops and if statements to sort data. However, this is inefficient as the pandas package uses C for its functions leading to a faster time to execute if I used the functions. I used some of the tricks from the previous assignment to help me sort/clean data without for-loops. I did not really pay this much mind in CA1, however, in CA2, I prioritized this a lot more, leading to a improvement in my code’s efficiency.  I also researched more on how to do certain functions without for-loops.  One other challenge I faced was cleaning the datasets of world bank. All of the Data had a header that pandas seem to ‘lock’ onto and not display any of the actual data. I used some parameters in read\_csv to skip the header and get past this. I did not face this problem in the CA1 as all the datasets were cleanly formatted as they were from the data.gov.sg, however, other websites had different formats which I had to get around.  Another challenge was trying to make my graphs easy to read while providing a accurate representation. Most of the graphs were alright. However, a big challenge was the large pie-chart. I wanted to showcase the info of a year as one large pie chart. I had to use a lot of sorting functions and research to find a good solution.  Then at the end I tinkered around with the size to make it readable. This issue was not a large concern in CA1. In CA2 however, it was a big concern. |

# 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.

|  |
| --- |
| I feel like I have put in a lot of effort to clean and sort data. I also feel that for my last pie-chart, it displays information clearly and looks nice. I had to change its legend and research a way to make it look like the way it is now. One thing I would like to highlight is I grouped some segments together to help with visibility as I thought it would be too messy, for example I combined all of the different “Watching TV” segments into a single watching TV segment.  I feel that I achieved a lot. Looking back at my CA1 work, I feel that my CA2 is a major improvement compared to it. I did a lot more complex functions such as renaming, capitalizing using lambda functions, combining and v stacking(Vertical Stacking) series. I also did multiple techniques to achieve the same result, Boolean indexing and the isin() function. I also focused more on efficiency in CA2, relying more on the pandas functions instead of for-loops and others.  I also felt like I made an effort to make my plots look better. I picked out high contrast colors in my plots without it being too harsh on the eyes. I changed the size of the pie chart to be more easier to read. |

# 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  State your justifications here  I think that I have done all of the necessary requirements and spent quite a bit of effort on the notebook.  I feel that I have put in more effort compared to CA1 generally, especially the readability and cleaning.  I feel like I have gained a lot of knowledge about the Pandas, NumPy and Matplotlib Package and that any task related to data cleaning and sorting is not too hard for me.  I feel like I have learnt a lot overall. |

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