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	Lim Chuan Hao
Your Student ID	1922264
Your Class	DIT/FT/1B/11

1   QUESTION 1: CHALLENGES - SELF-REFLECTION FOR CA2

Provide a brief reflection of the challenges you have faced in this assignment (CA1) 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 main challenge I faced in this assignment were trying to come up with the objective I wanted to solve. It was also much harder deciding on a general topic, based on the data I could find on data.gov.sg so as to do analysis on.
As I had to first find my own topic to analyze, find my own datasets and then craft my own objective, it took me quite a while before I was able to finally decide on the topic and objective and even then, I had refine it further to be specific.

Luckily, I learned that it was much easier to look and data and think from the perspectives of people that might want to look at these data to craft my own objective and topic. This thus allowed me to be able to have a more specific and data driven approach to the assignment, allowing me to come up with a meaningful objective, insights and conclusions.

I also faced challenges in learning how to properly clean and manipulate data in pandas. As with pandas, there were a lot more ways I could clean and manipulate data. For example, functions like melt, pivot, groupby. With so many possibilities, I started to think of more complex manipulations that I wanted to do and that meant I had to spend more time to understand what each of these functions does, in order to get the result I wanted at the end.

The previous assignment also helped me with this, as the knowledge I gained from using numpy, indirectly transferred to pandas, as some functions in pandas were built on top of numpy.

2   QUESTION 2: ACHIEVEMENTS - SELF-REFLECTION FOR CA2

Provide a brief 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.

Personally I felt I managed to do some meaningful and complex data cleaning and manipulation in

Programming for Data Science

pandas. Such as effectively using replace to replace strings that were supposed to represent Nans, and cleaning up string values when they were supposed to be floats. Like '1,237' for example.
I also managed to effectively pivot and melt dataframes, when I needed the data to fit specific graphs, in seaborn like the heatmap function.
I also managed to effectively merge different datasets into one large dataframe meaningfully so as to conduct my analysis.

I also think the graphs I created were quite complex, requiring me to plot multiple times, depending on the groups I had. An example would be gross median salary by field of study, which required me to manipulate the dataset first before being able to plot the data.
I also felt I managed to make the graphs created quite clear in what they are trying to show using the seaborn and bokeh library.

3   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?
Distinction

State your justifications here
I think I fulfilled all the requirements and managed to do above and beyond in the anaylsis, mainpulation and cleaning of data.
Such as using pandas functions to clean and manipulate the data to create new columns and merge dataframes together. I also used seaborn and bokeh to create more meaningful graphs so as to draw better insights from the data I had to meet my objectives and reach my conclusions.

-- End of Self-Reflection --