LEARNING ANALYTICS – CRITICAL REFLECTION

I had performed exploratory data analysis to some extent on penguin dataset and breast cancer dataset from MAS8403 and MAS8404 modules respectively, but this is the first time I have engaged in full-fledged data analysis in depth. Unlike those datasets, this dataset is not a single dataset but combination of multiple small datasets on seven different runs of cyber security course from future learn. This gives me an opportunity to dig deeper into the dataset and allow me to connect several small datasets to draw comparison between each run.

Initially, I thought of considering each run as a separate dataset and wanted to compare different runs of the course and produce results which I had obtained from the comparison. But I realised if we consider such analysis, it will not give me a clear and complete picture of the analysis and will not be able to do complex analysis with the dataset. Then, I wanted to analyse every part of the dataset from enrolment to leaving response. But again, it will be clumsy and unorganized if I do analyse every minute detail of the dataset.

I wanted to carry out a deeper analysis rather than a wider one. So, I finalised three sets – enrolment, question response and leaving response to perform data analysis. From these data sets, I can find learner's behaviour from the time they enrolled at to the time they left at. This thrilled me and I started working on this objective.

I did not use the git since I was instructed to do so. I have natural inclination to use git since it gives me an overall view of the project, how I started, how I changed my ideas and what made me to change my idea, how I worked with different approach and so on. Using git for a project makes me to evolve as a developer and gives command over my project. After completing the project with git and if I see the commit log down the memory lane, I can see my growth and involvement with the project.

With the language I have chosen to work on the project is R which will give me multiple libraries to work on the statistical side of the data and I have chosen RStudio as a tool to work with R language which will give me a command over the language than any other IDEs.

I will say my approach has two dimensions – one between runs of the course and one with the complete merge of all runs of the course. I started my analysis with enrolment data set and as first step, I merged all runs of enrolment data set into single frame. With the whole data frame, I compared the enrolled with their age, gender, highest education level and employment status. With different run of enrolment, I compare the number of people enrolled in each run of the course. Then I worked on question response with the number of correct and wrong responses in combined data frame and then with the number of correct and wrong responses with each quiz question in the course. Then with leaving response, I can classify learners who were left with their leaving reason, number of weeks they stayed in the course. These approaches can be accompanied with even better cross examination between enrolment, question response and leaving response.

With the combination of enrolment, question response and leaving response, enrolled people with different age, highest education level and employment status approaches quiz and why they left in between. Questions like “Will they attend the quiz and left the course or left the course without attending any quiz? “was little complex to find and plot. I had more questions in mind which need much more complex analysis, but the background details of enrolled people are not very proper, and many unknowns are there which disrupt the flow of analysis. I recommended the team to set the user background as mandatory to improve such analysis in the future. Likewise, I gave five recommendations to the team at future learn to improve their course.

I have approached this project with CRISP-DM methodology which helps me a lot in designing the flow and analytics of the project. The steps like understanding of business and data, data preparation, modelling, evaluation, and deployment helped me to analyse with much more conscience. But this methodology is not suitable for this kind of big data set since it needed many reworks throughout the project. It made me to redesign the project many times which is a disadvantage and time-consuming work to carry out.

Thus, I have approached the project with CRISP-DM methodology, designed with R using RStudio and analysed little deeper into the data set with two dimensions and produced different plots for the stakeholders to understand the analysis and gave recommendations to them to improve their course and increase the mass appeal of the brand they created all over the world.