CSC8631: Reflection Report

Abdullah Turki H Alshadadi, 190582184

In the course of this data mining endeavour, I encountered several challenges that significantly contributed to my learning experience. One notable difficulty was grappling with writing dynamic code. The programming language's design emphasised the use of variables for creating dataframes, which posed challenges in dynamically selecting and manipulating dataframes. Unlike the flexibility of Python, where functions can dynamically create or rename dataframes, the constraints of the chosen programming language added complexity to the coding process.

As the analysis progressed, a crucial realisation emerged regarding the need for a data exploration phase after the "Data Preparation" step. This phase would enable a holistic view of the dataset, unveiling unforeseen changes and ensuring a comprehensive understanding of the data before delving into subsequent analysis. This iterative approach would have potentially alleviated unforeseen issues encountered during the later stages of the analysis.

During the second cycle, a notable hurdle was encountered in modelling the correlation between the purchase status and answer correctness of participants. Initially considering a simple scatter plot, I faced challenges as the data involved boolean logical types. To overcome this, a strategic shift to employing a heatmap proved instrumental. The heatmap not only accommodated the data's nature but also provided a visually compelling representation of the correlation, offering a nuanced perspective that a scatter plot could not convey.

Given more time, I would have delved deeper into investigating additional factors influencing participants' decisions to purchase certificates. This exploration could have uncovered valuable insights into the complex dynamics that shape participants' choices and potentially informed more targeted strategies for the stakeholders.

Reflecting on the entire experience, it became evident that the journey brought about considerations and insights beyond the technical aspects of data mining. Understanding the business needs and refining the focus of the investigation to a single goal significantly enhanced the quality and relevance of the analysis. The structured approach of first understanding the data, meticulously preparing it, modelling visualisations, and evaluating results against the business goal fostered a more robust and critical thinking process. In conclusion, the experience was overwhelmingly positive and insightful, expanding my understanding of data mining from a technical exercise to a strategic and business-driven endeavour.