Name: Jaskeerat Singh

Student ID: s4761003

DATA7001: Introduction to Data Science, Semester 2, 2022

<u>Reflective Essay</u>

The most significant learning from this group project is the understanding of implementation of the data science process to the "Data Science jobs" project chosen by the group. The process majorly consists of 5 steps. Seeing how each step affects every part of the project and helped us in defining the work's scope as well as giving a structure to the project was surprising. These five steps helped me understand the deliverables of the project and cemented a process in my mind that should be followed whenever dealing with a data science problem. It described the way by helping us choose relevant data from a reputed location, defining the stakeholders and scope, transforming data, and storytelling by visualising the data. The whole process explained the scenario, which helped us divide the work and gave us various ideas which can used to present the project properly. My part in the process was to ensure that after data collection, it should be ready for exploration by applying various pre-processing techniques and creating visuals for storytelling using tableau.

What went right was the implementation of techniques like cleaning complex data using regex to get individual values, feature selection using correlation for column selection, imputation for missing data after building a linear model, outlier detection and removal using quantiles and creating visuals for storytelling using tableau. Members of the group also helped greatly by discussing the pros and cons of each problem and took small steps in defining the work and its outcomes.

The selection of datasets for developing insights could have been **done differently**, as it contains a high number of nulls in the salary column (60%). Even though the missing values have been imputed, it might give a skewed view of the insights. Other pre-processing techniques could also be used to train the model, which might give a better or provide us with an alternate view of the insights.

Group Dynamics: In the beginning, the communication channel using WhatsApp was set up where regular discussions about the division of work and individual responsibilities were discussed. Regular discussions regarding new ideas and issues during implementation were always performed on a timely basis. All the group members were always prompt in response and never slacked off in helping when someone was stuck. All the members behaved cordially and helped to prepare all the deliverables on time. The members of the group were open to criticism and worked with each other to solve the issues.