**Title**

Data Science Methodology Application in Credit Cards

**Question 1**

I chose the credit cards topic to apply this methodology because I am currently working in the finance sector specifically in the Data Analytics department where we deal with customer financials data on a daily basis.

**Question 2**

Credit card applications has been on the rise in my banking company recently but that would also mean an increase in default risk, and my supervisor does not want the manual way of reviewing applications as it will be time consuming. The question here is "Is there a way to predict whether a credit card customer will default or not?".

**Question 3**

In the Analytic Approach step, since we only want a clear answer of whether to approve the application or not, therefore a binary classification model should be used.

In the Data Requirements Stage, we would require both existing credit card customer data and customer data arising from new but unprocessed applications, the data will include Age, Occupation, Annual Income and Marital Status etc. (Default status will also be included for existing customers only)

In the Data Collection stage, we sourced and collected existing customer data by querying directly from the company database. As for new credit card applications data, we compiled it onto a spreadsheet which could come from both physical and online forms.

In the Data Understanding & Preparation stage, we eliminate irrelevant customer data such as addresses and first & last name as these will unlikely contribute to the performance of the model. We will also conduct feature transformation such as log-transforming right-skewed fields and one-hot encoding categorical variables. We will also use correlation plots to observe the relationship between variables and eliminate independent variables which are highly correlated.

In the Modelling and Evaluation Stage, we try to fit the preprocessed data using our classification model, preferably a Decision Tree Classifier as it allows us to see clearly how the model works, the model of the classifier is evaluated using performance metrics such as Accuracy, Precision, Recall and F1-Score, the model hyperparameters are tuned until we have a model with a balance of both high Accuracy and F1-score.