Which topic did you choose to apply the data science methodology to? **(2 marks)**

**I chose the topic ‘Credit cards’ to apply the data science methodology to determine if a credit card request should be accepted/rejected based on propensity of future fraudulent activity.**

**My reason for choosing this topic is that I work in the Financial Services industry.**

Next, you will play the role of the client and the data scientist.

Using the topic that you selected, complete the Business Understanding stage by coming up with a problem that you would like to solve and phrasing it in the form of a question that you will use data to answer. **(3 marks)**

You are required to:

1. Describe the problem, related to the topic you selected.
2. Phrase the problem as a question to be answered using data.

For example, using the food recipes use case discussed in the labs, the question that we defined was, "Can we automatically determine the cuisine of a given dish based on its ingredients?".

1. Credit cards are one of the most ubiquitous payment tools in the modern world. As millions of people are being brought into the gambit of financial inclusion every month, credit cards as payment instruments are becoming increasingly popular. This gives rise to the incidents of fraud at the same time. The impact of a financial fraud can be devastating to the card-issuing financial institution. Hence, it becomes critically important to understand if a person being issued a credit card can potentially indulge in a fraudulent activity.

2. Question – how to determine if a credit card request by an individual should be declined based on propensity of a fraud/financial crime in future?

Briefly explain how you would complete each of the following stages for the problem that you described in the Business Understanding stage, so that you are ultimately able to answer the question that you came up with. **(5 marks)**:

1. Analytic Approach
2. Data Requirements
3. Data Collection
4. Data Understanding and Preparation
5. Modeling and Evaluation

You can always refer to the labs as a reference with describing how you would complete each stage for your problem.

1. Analytic Approach – The question needs a Yes/No answer so my approach would be predictive analytic model. This approach is used when we need to predict yes/no or stop/go type outcomes, forecasting, what will happen next, determine probabilities of an action.

2. Data Requirements – Decision tree classification requires identifying the necessary data content, formats and sources for initial data collection. The data sources will be as follows: a. Personal credit history b. Loan payment history Data content I will look at: a. Identity confirmation b. Any other credit cards held and for what duration? c. Credit limit requested? d. Income statements via salary and other sources

3. Data Collection a. See if any of the above data content is missing and re-request it. b. Determine if any additional data is required if customer’s repayment capacity is doubtful.

4. Data Understanding and Preparation a. Validate if the data is relevant to decision making? b. Is there any other source that can be utilized? c. Can Feature Engineering be applied here?

5. Modeling and Evaluation a. The data model is predictive in nature. b. Model will build algorithms to make use of the various variables (data points) captured and come up with a decision. c. Utilize a training set is used for predictive modelling before deploying the change. d. Evaluate if the model used really answers the initial question or does it need to be adjusted?