Final Assignment: Data Science Methodology

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

   I have chosen as topic for this task the application of data science in the field of credit cards. The reason behind this choice is that it is related with my background in Consulting, especially in Financial Services.

2. *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?"

   The main problem for Banks regarding credit cards is that they must create a model to know to whom they can provide them. Certain clients will not be feasible, as they do not have the economic strength to back up this service. And it can take long time for Banks to do the approval process for their prospective clients.
   Hence, our question would be " Can we automatically determine if a client is suitable to obtain a credit card?

3. *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. Modelling 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:
      As mentioned above in the Problem Statement, the problem requires a yes/no answer we will use a classification model.
   2. Data Requirements:
      To create the classification model we will require information regarding the bank clients. This info should include personal data of the client and should include the ones that defaulted and the one that paid.
   3. Data Collection:
      We would use techniques like descriptive statistics and data evaluation should be implemented in this phase to make sure that we have useful data for our model.
   4. Data Understanding and Preparation:

In this step we need to evaluate the different variables of our data in order to understand it better. For example, we would calculate univariate statistics, such as mean or median and the correlation between variables. So we need to evaluate the quality of the data. In the data preparation phase, we have to prepare the data in a specific way depending on the model.

5. Modelling and Evaluation:

Lastly, we create a classification model, evaluate the outcome and perform the corresponding changes untill we have a suitable model.