What topic did you chose to apply the data science methodology to?

The topic I chose for my data science methodology is credit cards. The question I want to answer is who qualifies for a credit card?

This question requires a client's personal information such as their age, if they have provided a social security number, their access to income, and any potential credit card history they may have had. A classification model can then be built using this information that determines who qualifies and may be issued or denied a credit card.

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
   1. The business requirements for the model are the client's personal information and any credit history they may have had and as it is a yes or no question of if they qualify for a credit card a classification model will be used for the analytical approach.
   2. The data requirements are the client's personal information like age, valid social security, income, and credit history. The company's guidelines will be used to define what age, income, or credit history/score a client must have to be issued a credit card.
   3. For data collection the data source will be acquired from the banking corporation as they have the necessary information which would allow us to help them clients qualified for a credit card. This data should be available and descriptive statistics can be used to get an understanding of how clients vary for these variables and verify that all of the necessary information is given by each client. If there is missing data on a client, we could more easily discern the missing information and reach out to them to acquire it.
   4. For the data understanding and preparation stage the descriptive statistics used earlier still apply. We could verify that each variable is recorded for each client and look at averages, minimums, and median age, credit score, and income. This will let me better understand if I expect most of the clients to qualify or not before moving into model evaluation.
   5. For the model it would be a classification model. To evaluate its effectiveness, we could look at its classification accuracy and break it down with a confusion matrix. This would let us see the number of misclassifications or the false positives and negatives. The model that results in the best accuracy and rate after being tested and trained on any past data would be our final model chosen.