Coursera: IBM Data Science Professional Certificate

Course 3: Data Science Methodology

Final Peer-graded Assignment

In this Assignment, you will demonstrate your understanding of the data science methodology by applying it to a given problem. Pick one of the following topics to apply the data science methodology to: emails/hospitals/credit cards.

You will have to play the role of the client as well as the data scientist to come up with a problem that is more specific but related to these topics.

Q1: Which topic did you choose to apply the data science methodology to?

My project will revolve around launching a new credit card programme designed for young professionals, age range from 20 to 30 years old.

Q2: 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. You are required to 1) describe the problem, related to the topic you selected, and 2) phrase the problem as a question to be answered using data.

1) Problem:

We (the bank) are planning a new credit card programme with young professionals (age 20-30) being the target customers. We aim to 1) grow our revenue deriving from the credit card transaction commission fee by increasing the number of the new credit card users, and 2) increase our brand awareness in the market among this particular age group.

2) Question to be solved:

Can we design a credit card scheme based on the general consumption pattern/trend of the young professionals who make payments with their debit cards already?

Q3: 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.

1) Analytic Approach:

The question stated seems to be able to be best answered with an explicitly designed credit card scheme with prescriptive award programmes.

2) Data Requirements:

The following data may be required to formulate the best credit card scheme: age, gender, marital status, profession, yearly income, monthly spending, and expenditure categories.

3) Data Collection:

Data may be collected through the database of personal information registered when the individual opened an account at the bank. It can also be collected from debit card payment histories, where the bank is able to categorize different types of expenditures (e.g. restaurants, bars, supermarkets, education, leisure, tourism, transportation, telecommunication, utilities bill...).

4) Data Understanding and Preparation:

A preliminary understanding of the data may be provided using visualisation showing consumption trend/patterns and feature correlations. This may help us figure out if additional data (e.g. type of living status: living in parents' house/living in their own properties/renting)) is still required (whether at the moment of examining or later during date modelling stage) in order to better tackle the initial problem. Redundant data may also be eliminated at this stage.

The data collected must be prepared/cleansed before it is ready to be processed to modelling. This part may involve tasks such as correcting typos, unifying terms, eliminating items with missing data, finding out missing data, updating old data, re-interpreting data from/into numerical terms, and formatting data.

5) Modeling and Evaluation:

In this stage, different algorithms would be used to explore the best models with the greatest potential to increase the bank's revenue from the commission fee. For example, an attractive cash back scheme which is aimed for supermarket frequenters who prefer immediate cash back, or a mileage accumulation scheme for travel lovers.

The outcome models should then be compared with reference to the initial question. In addition to increasing revenue, the bank also wants to grow its brand awareness in the market. Then a mileage scheme should be more promising in this

regard when the bank discovers travelling has become an increasingly hot topic among the age group in concern.