Data Science Interview Exercise

Barclays has launched a new wealth product in UK, called ABC. The customer information of this product was aligned with the customer demographic data (a centralised customer information dataset for all Barclays customers). We generated a sample dataset with a mix of Barclays customers who bought this product and the customers didn’t buy this product. This sample dataset is attached with this exercise. The description of this sample dataset is:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| age | Age of customer |
| workclass | Type of work status of customer |
| education | Education level of customer |
| marital-status | Marital status of customer |
| occupation | Occupation of customer |
| relationship | Relationship of customer |
| race | Race of customer |
| sex | Gender of customer |
| capital-gain | Capital gain of customer with Barclays |
| capital-loss | Capital loss of customer with Barclays |
| hours-per-week | Working hours per week of customer |
| native-country | Nationality of customer |
| customer | Whether or not is customer of ABC product  (1 is yes, 0 is no) |

In this exercise, we expect you could help:

1. Analyse existing customers of ABC products to understand their characteristics.
2. Create a model to evaluate who will be the potential customer for ABC product among millions of Barclays’ customers.

This exercise is testing your approach to solving problems and is open to any data analysis techniques.

We expect a technical analytics report to describe the findings from this exercise including how you create and validate the model to predict potential customers.

Any technology or visualisation tools can be used.

During the interview, you will have five minutes to present the results, and then 20 minutes will be spent with questions about the report, including background to any of the methods that were used