TASK DESCRIPTION

1. The task

You are faced with the following task:

The Retail Customer Relationship managers heard about big data and start to think about all the data they had to collect and are available in the bank. Since they are not data analysts themselves and don't know what big data really means, they ask you to "analyze the data".
Their goal is to identify new target customers for mortgages. It is your task to analyze the data, find promising new customers and give a short presentation of your results.

2. The data

You are provided with two files, *Retail data.csv* and *Potential Customers.csv*.
Information about Age, Sex, Age at origination of loan (only for *Retail data.csv),* Years with bank, marital status, education level, employment status, customer income, current balance and if the customer has a mortgage or not, among other columns is provided in these data tables.

3. The modelling

For accomplishing this task please use either **R or Python**. You can use any statistical method you see fit (Cluster analysis, neural networks, Linear/ Logistic Regression…).

For training the model, please use the file *Retail data.csv.* This model should then be used on *Potential Customers.csv*. Its outcome should provide some guidance on which customers should be approached (for example by calculating a probability of taking a mortgage).

4. The presentation

You can choose to present your results live in the Programming-environment or in a PowerPoint presentation. This presentation is aimed at Customer Relationship managers (who have heard only a little bit about big data) with a time frame of 15 min.

Please provide your outputs in a zip file which should contain presentation and all the relevant codes.

5. The evaluation

The presentation, the modelling (use of data/transformations) as well as the structure of the code are evaluated. Additionally, of course, your answer to our topic related questions.