Assignment instructions

A year ago, Bohemia Bank a.s. started to offer a new loan product to retail customers: 30 000 CZK loan payable in one year. The special feature of this loan includes no monthly installments. The loan is simply repaid after one year altogether with interest. As a bonus to customers, there are no fees associated with the loan underwriting or administration.

One year after the product introduction has elapsed, it seems that the product wasn’t as profitable as expected. Moreover, there are some rumors that the bank has lost a lot of money due to wrong calculation of expected costs.

You have historical dataset of 10 000 loans in following structure:

|  |  |  |
| --- | --- | --- |
| Variable | Character | Description |
| Client | N/A | Identification number of the applicant. |
| Age | discrete | Age of applicant. |
| Gender | categorical | 0 – Male  1 – Female |
| MarStat | categorical | Marital Status of the applicant.  1 – divorced  2 – married  3 – single  4 – window/male widow |
| Housing | categorical | Type of applicants’ accommodation.  1 – lives with with parents  2 – registered at the town hall  3 – in an own house/apartment  4 – in a rented house/apartment |
| Educ | categorical | Type of applicants’ education.  1 – elementary  2 – secondary school  3 – university  4 – PhD. |
| Inc\_Gr | categorical | Volume of applicants’ income.  1 – receives social security  2 – receives salary less than 15 000 CZK  3 – receives salary between 15 000 CZK and 30 000 CZK  4 – receives salary more than 30 000 CZK |
| Children | discrete | Number of applicant’s children. |
| IQ | discrete | Applicant’s IQ |
| Default\_flg | discrete | Flag suggesting whether the loan was paid back (=0) or not (default=1) |
| Rate | continuous | Annual interest rate of the applicant’s loan |
| Cost\_type | character | Type of distributional channel. |

Additionally, you have historically observed, that you can collect larger share of defaulted debt (debt of customers who failed to pay back) from loans of applicants who are owners of their houses and flats. You know that in these cases, you usually collect 80% of the debt. In other words, you usually loose only 20% of the financed amount. (Share of lost part of the loan which defaulted and is not collected back is called Loss Given Default – LGD). In case that the applicant isn’t an owner of his/her apartment/flat, the lost amounts are on average much higher: 70% of the financed amount.

There are three distributional channels: www, Facebook and Brokers. WWW stands for online, Facebook stands for online offering via Facebook and Broker stands for external third party which offers relevant clients directly to the bank. The most cost-efficient is the WWW which costs on average   
200 CZK for each loan. Facebook costs 300 CZK per loan and the most cost intensive are the Brokers who charge 1 000 CZK for each loan.

Analyze the problem setting using the provided data and follow the list of the tasks:

## Tasks:

### Dataset description

1. Do variables IQ and age have a direct influence on the selection of the distributional channel? Can you describe and interpret this relationship? (if any)

2. Is there any relationship between IQ and age?

3. What is the default rate (share of customers who defaulted to the total population in the sample) in all classes of variable Inc\_Gr? Is there any relationship between income and default rate? Does it have an economic interpretation?

4. Describe the whole portfolio using the following metrics (ignore the distribution channels segmentation):

* Number of observations and defaults
* Default rate, Average LGD (Loss Given Default)
* Average IQ , Age;t
* The most frequent characteristic for following variables: Housing, Educ, Inc\_Gr.

### Profitability analysis

5. Calculate the historical total profit of the loan portfolio

* What is the profit for each used distribution channels?
* Which of the distributional channels proved to be more profitable?
* Why do you think it is more profitable compared to the others?

6. What are the typical characteristics of the top and the least profitable customers?

### Recommendations to the management

7. Summarize your findings in a form of presentation. The presentation should include the solutions to the tasks described above as well as the summary of the current situation and your recommendation regarding the future of this portfolio. Also, the presentation should include description of your methodology.