****

**Enova Data Analysis Case Question**

**Instructions:**

1. Review the materials below. **READ THEM CAREFULLY!**
2. Ensure you understand the timelines and expected output. Clarify any questions you may have with someone from Enova.
3. Decide how you want to divide up the work on your team. Due to the short time constraints, it’s important that you maximize the productivity of everyone on your team.
4. When the time is complete, return the USB drive to the Enova employees per the instructions in this document.

**Scenario**

You are an analytics employee working for a consumer lender. For the past few years, your company has been issuing loans to customers and gathering the resulting historical performance data on this portfolio.

Your company does not lend its own capital. Instead it borrows capital from another source, which it then lends out to the consumers. Your company takes all of the credit risk; that is if the consumer does not repay the loan then you accept the full losses and if the consumer does repay the loan, you collect the full interest.

|  |  |
| --- | --- |
| **Loan Structure** | A one-time loan, with fully amortizing payments occurring at the end of each month |
| **Loan Term** | Either 36 or 60 months |
| **Interest Rate** | 6% - 26% annual |
| **Loan Size (Principal)** | $1000 - $3500 |
| **Discount Rate (WAAC)** | For this exercise, we will take WACC = 0% (no discounting) |

**Goal**

Your boss comes to you with a new list of prospective loans (validation.csv). He asks you to figure out to which of these customers we should lend.

The objective is to determine which customers we should lend to in the future, such that the resulting total realized net present value is maximized.

**Data Set**

You have been provided a training data set (training.csv), which contains all the information about the past loan performance. Note that this includes both current and historical loans. For the current loans, there is no data available for the loan status and realized npv columns.

Previously, your company used to get the loans scored by a third-party vendor. The projected NPV and risk grade columns were obtained from this third-party vendor.

You also have a validation data set (validation.csv) that you are to use to determine which customers in this data set we should lend to in order to maximize the total realized net present value.

Note that your company no longer uses the third-party vendor. Thus there are no projected NPV or risk grade columns in the validation dataset.

The variables are defined in the file data\_dictionary.xlsx. They are categorized as follows:

1. Application data: Data provided by the customer or the lender at the time of application.
2. Credit data: Data obtained from 3rd party credit bureaus that provide insight into a customer’s previous repayment history.
3. Loan Info: Information about the structure of the loan that the customer was provided.
4. Loan Performance: Information about how a particular loan performed. **Notice that this data is only available in the training data set, that too for historical loans. Since we are deciding which of these customers to lend to, we do not yet know what the outcome of the loan will be.**

**Output**

In order to be eligible for scoring, you must provide the items listed below. Each of these files should be stored on the provided USB drive.

1. You are expected to provide the validation data set (validation.csv) with an additional column labeled “decision” coded as 0 or 1. A value of 0 indicates that you DO NOT recommend lending to that customer, 1 indicates that you DO recommend lending to this customer. Please output this file in csv format.
2. A (brief) paragraph explaining the items listed below.
   1. Data preparation
   2. Model building techniques and strategy
   3. Any interesting findings
   4. Conclusions/Business Recommendations
   5. Areas for future research
3. Copies of all code used to complete the analysis. If multiple files are used, create a folder to store them and name it “Code”.