

# CT130-3-M-ODL-DAP– DATA ANALYTICAL PROGRAMMING INDIVIDUAL ASSIGNMENT

Intake : YOUR INTAKE

Lecturer : Mr. DHASON PADMAKUMAR

Date Assigned :

Date Due & Location :

**LEARNING OUTCOMES**

At the end of this assignment, you should be able to:

1. Assess various forms of datasets by reading, combining and categorizing using data analytical programming techniques

2. Produce analytical data models by creating summary reports and enhanced listing

3. Formulate visualization and Discovery Strategies

**INSTRUCTIONS**

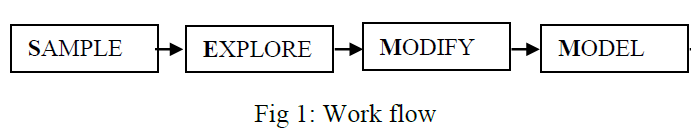
Lasiandra Finance Inc. (LFI) New York, USA is a leading private financing company which caters the funding needs of Small and Medium enterprises (SME). LFI clearly understood that some business dreams need that extra push to see them accelerated. Hence it allows its loaning process as tailor-made and customer centric. In the past few years, it has tremendously expanded its wings and to speed up the process, it needs to automate the loan eligibility process based on customer portfolio entered online.

**PROBLEM**

The main problem faced by the LFI is the approval process of the loans. Because it is a complicated procedure of verification and validation but still there is no guarantee whether the chosen applicant is the deserving one out of all applicants. Hence, it needs a model which can predict the loan approval.

**TASK**

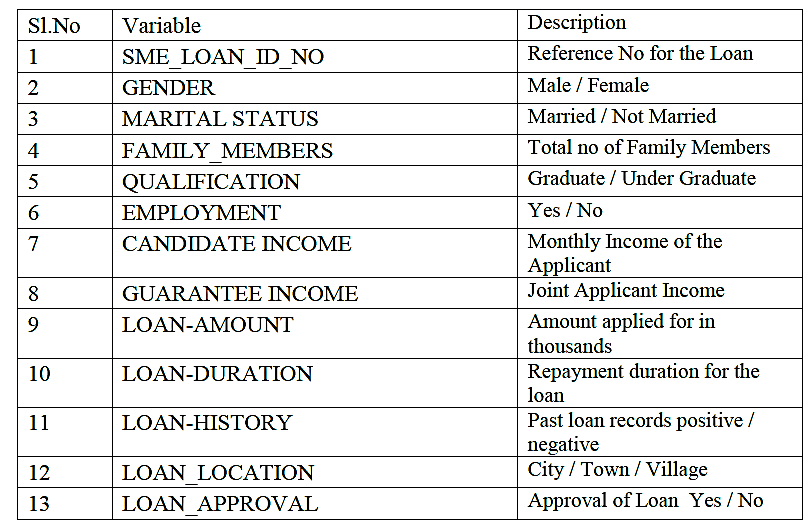
You are the recent hire as one of the Data Scientist at the Headquarters of LFI, Washington, D.C., United States. You have been assigned to analyze the data set obtained from the past customers, and build a most accurate model to predict the approval process as approved or rejected. Your department chief hands over this assignment to you with the work flow as shown in Figure 1.



You are required to follow the work flow and to build the model proposed. Report the output produced by the model you found, out of your analytical study to the management of LFI.

**DATASETS**

*Note*: Do **NOT** edit the data in any way. The variables on the dataset are as follows:



**DELIVERABLES & FULFILLMENT**

1) You are required to conduct a thorough analysis on the dataset by writing programs with **SAS**.

2) The report requires you to introduce the data, methods and coding problems that precipitated your interest in this study. You should also discuss your ultimate goal and objective when undertaking this analysis.

3) Produce output results, interpret them and write a report and/or presentation using the outputs.

**PROGRAM DEMONSTRATION & CODING**

Be sure to break down complicated procedures into their essential steps, and to demonstrate the steps with sub sets of PROC, SQL code, macro code or supplementary code. At each step, explain whether the code worked the way you anticipated. Identify areas of improvement and next steps. Discuss obstacles and don't worry about including mistakes. It’s also perfectly reasonable to come up against a seemingly intractable programming problem, or one that threatens to overtake much of your effort on the project, and make some

modifications to the project. Often these modifications involve a change in your goal, possibly by simplifying code and output. Be straightforward and honest about these obstacles in your write-up.

At the end of the process, you should have a complete set of code that you demonstrate on your full data set, along with the product of the code, data sets or tables, ODS (Output Delivery System) output, etc. Be sure to write a conclusion summarizing your project and reflecting on what you learned. Include suggestions for improvement or enhancements to your final product, and discuss their feasibility. Useful comments also help describing the intent of each step in the program.

Your report should include the following:

**Abstract** – A self-contained, short, and powerful statement/brief that describes your work. It may contain the scope, purpose, results, and contents of the work.

**Introduction** - The purpose of your report. Background information about the topic. You also have to place some brief details of your methods applied for the study. Include an outline of the structure of the report.

**Meta data**

**Literature review** - Carefully structure your findings.

**Initial Data Exploration**

**Data Pre-processing**

**Exploratory Data Analysis (EDA)**

**Model Creation & Prediction**

**Report Generation using SAS ODS & SAS MACRO**

**Conclusion** – In this section, you need to state your position about what you gained in this assignment that can contribute to other readers.

**Documentation Format:**

* + Typeface: Times New Roman. Boldface, italic & lines can be used for emphasizing and to enhance readability.
  + Font size: 12 (except titles and headings).
  + Margins: 1” from the left, right, top & bottom of the edges of the A4 paper.
  + Spacing: 1.5 lines between texts of a paragraph.
  + Alignment: Justify.
  + Headers and footers can be used all pages must be numbered accordingly.
  + Standard cover page as available in the learning management system

**Word Count:**

3000 ~ 5000 Words

**Submission date & time: Before 11:59 PM ON DATE TO MOODLE**

**ASSESSMENT CRITERIA**

|  |  |
| --- | --- |
| **Individual - (100%)** | **Marks** |
| **PART 1** | **10%** |
| **Introduction** | **5** |
| **Metadata** | **5** |
| **PART 2** | **50%** |
| **Literature review** | **10** |
| **Data Pre-processing**  Correctness of identification of the attribute types, the pre-processing procedures, results and explanation of the steps undertaken. | **15** |
| **Exploratory Data Analysis (EDA) – graph**  Depth of data understanding - how comprehensive are the explanations of your explorative results, appropriateness of illustrations. | **15** |
| **Model creation and prediction** | **10** |
| **PART 3** | **40%** |
| **Formulate visualization and report generation** | **30** |
| **Presentation** | **10** |
|  | **100** |

**Grading**

|  |  |
| --- | --- |
|  | **Range** |
| Distinction | >=75% |
| Merit | 65 -74% |
| Pass | 50 - 64% |
| Fail | 0 - 49% |

**APPENDIX 1**

|  |  |  |
| --- | --- | --- |
| **Marking Band** | | **Band Description** |
| **Characteristics of Student Achievement by Marking Band** | 75-100 % | Exceptional problem solving with supporting analysis and application development skills with extraordinary originality and adoption of the tool, platforms, methods & technologies. Work may be considered for further expansion. |
| 65-74 % | Satisfactory knowledge base that supports some analysis, evaluation and problem- solving and application development. |
| 50-64 % | Limited knowledge base. Limited understanding of discipline. Difficulty with problem solving and application development. Little evidence of understanding of discipline.  Significant difficulty with problem solving and application development. |
| 0-49 % | Inadequate knowledge base. Inadequate understanding of discipline. Major difficulty with problem solving and application development.  Awarded for: (i) non-submission; (ii) dishonesty; (iii) in situations where the student fails to address the assignment brief (e.g.: answers the wrong question) and/or related learning outcomes |