Financial Programming

Group Project – Financial Data Set

The group project is a Python programming assignment where the students are required to construct a data science basetable from the provided financial data set.

Purpose

The main purpose of this assignment is to assess the students' skills on the these following criteria:

1. The ability to process complex raw data to create a data science basetable correctly and creatively.
2. The ability to explore, extract and report insightful information from the created basetable.
3. The ability to organize a data science project, including data, scripts, documents, and other deliverables.
4. Collaborating and teamworking.

Description

To create the basetable, the students need to follow the below guidance:

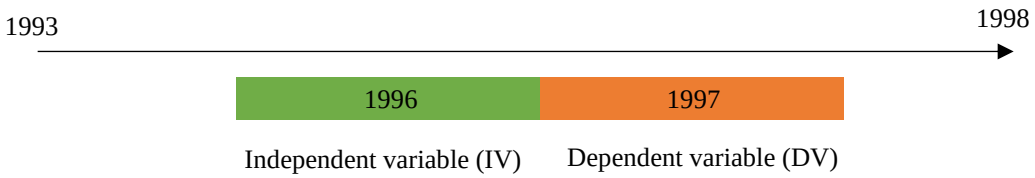

Figure 1: Illustration of the independent and dependent variables time window

1. Extract information from the financial data set and create a data science basetable with the following criteria:
    - The granularity (lowest level of observation): clients who are account owners.
    - The time window of the independent variables: 1996 (1 year, Fig. 1).
    - The time window of the dependent variables: 1997 (1 year, Fig. 1).
2. Create the independent variables (e.g. gender, age, age group, etc.):
    - Should be calculated only for the independent variables time window (i.e. 1996).
    - Should be calculated only for clients having full information in the independent variables time window (i.e. 1996).
3. Create the dependent variables (or target variables):
    - Target variable #1: Client had granted loan in the dependent variables time window (i.e. 1997), binary value (0 = did not have granted loan, 1 = had granted loan).
    - Target variable #2: Client had credit card issued (for account owner or disponent) in the dependent variables time window (i.e. 1997), binary value (0 = did not have credit card issued, 1 = had credit card issued).

After creating the basetable, the students should also report and analyze the basetable on the following points:

1. Describe to the variables of the basetable (e.g. variable name, description, data type, value, etc.).
2. Describe the data correction and transformation (if any) and the applied methods.
3. Analyze and visualize the independent variables and dependent variables.

Submission

- A project folder should include:
    - The Python Jupyter Notebook(s) (*.ipynb)
    - The data science basetable (*.csv)
    - The report describes and analyzes the basetable (*.pdf)
- The organization of the project folder can follow the suggestion in the README.txt file.

Evaluation

The project will be evaluated on:

- The quality of the Jupyter Notebook(s) and the basetable [50%]
- The quality of the report and the analyses of the basetable [40%]
- The well-organized project folder [10%]
- Peer-2-peer evaluation

Importantly, using the work (text or programming scripts) of other people without citing the source is considered as plagiarism, and is strictly prohibited.

Timeline information

- Deadline: 30 November 2021.