Case Study (data related + modelling)

Problem Statement

These files contain complete loan data for all loans issued through 2007-2015, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information. The file containing loan data through the "present" contains complete loan data for all loans issued through the previous completed calendar quarter. Additional features include credit scores, number of finance inquiries, address including zip codes, and state, and collections, among others. The file is a matrix of about 890 thousand observations and 75 variables. A data dictionary is provided in a separate file.

The objective of the exercise is

- **Role** - You are a lending portfolio Analyst

- Task 1 - Provide your insights on this portfolio and develop a lending scorecard
- Task 2 - Suggest a portfolio management strategy, e.g. Underwriting, Pricing, Provisions, ticket size, moratorium etc

> Portfolio Objective - We want to grow portfolio cautiously in Post COVID world. The lending scorecard is a broad area, but we will like you to think and articulate the objective of the scorecard - *decide the outcome, outcome period, the dependent and independent variables.*

Note - you need not cover all aspects of portfolio strategy. Feel free to choose a couple and explain the rationale for doing so.

- 3 to 5 slides to present your views
- Analyse data and develop a strategy while you set up the context
- Please share your analysis notebook (preferable an ipython notebook else an R-Studio Notebook and share the same.

Files

- LCDataDictionary - A dictionary describing all the columns in the data.
- loan.csv - Raw data.

https://drive.google.com/file/d/19bHzZHS0XJsWhz0q0C7O9gimOP8fK1Lf/view?usp=sharing