Assignment 1

Due Week 4 Beginning of Class. Please submit the following documents in a single zip file through Canvas:

- *Your 'ipynb' source file. (Please make sure that it is runnable to replicate your results.)*
- *A brief report of your answers, take-aways, and thoughts.*

Please name your zip file in the format of "A@_G#.zip", where "@" is the assignment number and "#" is your group number on Canvas. For example, if you are submitting Assignment 2 for Group 1, please name your zip file as "A2_G1.zip". Only one submission per group is needed.

References to resources that are not in the textbook or class handouts should be explicitly mentioned in the write-ups and source codes.

In this assignment, we will revisit the credit lending problem, but on a more realistic data set. We will extend the model by considering more features such as loan amount, term, interest rate, and grade. Our final goal is to explore the performance of logistic regression and its variants.

- Since this is the first assignment, we will walk through the whole process using a step-by-step approach: We will start from the ETL (Extract, Transform, Load) and data cleaning, followed by building logistic models, and finally, how to produce a credit decision rule. We will also explore regularized logistic regression.
  - The step-by-step guidance is contained in the file "*DBA3713_assignment_1.ipynb*." Please complete the tasks and fill in the blanks/empty cells. (45 points in total)
- Please also write a brief report of your answers and take-aways. In the report, also think of the following question: "Based on your model, how might LendingClub use the results to set interest rates?" (5 points in total)
- A dataset containing the full loan data (as well as its explanations) has been uploaded on Canvas.
  - The loan data file is "*lendingclub_full_data_set_no_id.csv*"
  - There is also a file "*lendingclub_data_dictionary.xlsx*" that explains the meaning of each field.