2025-2026 Fall Group Project

Due Date: Monday December 15th
*Submit your group project report as **a single pdf** file by email*

Loan Appraisal for FNB Bank

You are a credit analyst working for FNB Bank. Your responsibilities include analysing the loan applications and making recommendations to management, based upon your findings, to help them make data-driven decisions on lending.

"**Loan.csv**" file includes data on 40,000 FNB customers that were granted a loan in the past and the respective outcome, i.e. whether they were identified as "write-offs" or "not write-offs".

Using this data, analyse the new loan applications whose data are provided in "**NewApplications.csv**" and try to predict whether each new applicant will repay the requested loan if it is approved.

Data available:

1. Gender → M=male; F=female
2. Age → an integer parameter
3. marital_status → widowed; married; single; divorced
4. education → basic; highsch; univ; postgrad
5. nb_depend_child → number of dependent children (0,1,2,3)
6. employ_status → employment status (full_time; part_time; unemployed; self_employ; retired)
7. yrs_current_job → years at the current employment
8. yrs_employed → total number of years employed so far
9. net_income → an integer parameter
10. spouse_work → yes; no
11. spouse_income → if the spouse works, what is his/her income?
12. residential_status → home owner (owner); tenant; home owner with a mortgage (owner_morg); living with parents (w_parents)
13. yrs_current_address → years at the current address
14. loan_amount → an integer parameter
15. loan_purpose → debt consolidation (debt_consol); wedding; home improvement (home_improv); vehicle; holidays; other
16. loan_length → the duration of the loan
17. collateral → yes; no
18. writeoff → yes; no

Your report should contain:

1. an investigation of the data and a summary of your descriptive analyses; (15 pts)
2. a discussion on the pros and cons of the prediction methods that can be used to address FNB's loan appraisal problem; (5 pts)
3. a brief description (and the assumptions made, if any) of how selected prediction methods are applied; (5 pts)
4. *R* codes developed; (10 pts)
5. an evaluation of the results obtained by each prediction method tried on the data; (20 pts)
6. a comparative analysis of the results; (20 pts)
7. your final recommendation to FNB on which customers should be granted loan; (20 pts)
8. a discussion on any additional data that you think would be useful, if collected, to make better predictions in the future. (5 pts)