Linear & Logistic Regression Case Study

The Objective: One of the leading banks in Germany would like to solve the following business problems.

1. The drivers for the bad loans and predict the bad loans given the drivers (perform **logistic regression** on "**Badloan**" using the Other variables as the independent variables)
2. The drivers for the loan amount and would like to predict the loan amount for given drivers(perform **Linear regression** on "**Credit Amount**" using the other variables as the independent variables)

Expectations from the Trainees:

1. Understand the data & perform the data preparation before the model building
2. Understand the output from the software and explain the model fit
3. How would you determine what is the "best" linear model?
4. Apply transformations to the given variables and find out the possible best model after transformations.
5. Generate the final equation
6. Validate the model and present the results in Excel or PPT.

Data Dictionary:

Number of Variables in the data set: 21

Description of the Variables:

1. Badloan:
   1:Bad loan
   0:Good Loan
2. Account Balance: Status of existing checking account
   1 : ... < 0 USD
   2 : 0 <= ... < 200 USD
   3 : ... >= USD / salary assignments for at least 1 year
   4 : no checking account
3. Duration of Credit (month): Duration in months
4. Payment Status of Previous Credit: Credit history
   0 : no credits taken/ all credits paid back duly
   1 : all credits at this bank paid back duly
   2 : existing credits paid back duly till now
   3 : delay in paying off in the past
   4 : critical account/ other credits existing (not at this bank)
5. Purpose
   0 : car (new)
   1 : car (used)
   2 : furniture/equipment
   3 : radio/television
   4 : domestic appliances
   5 : repairs

6 : education
7 : (vacation - does not exist?)
8 : retraining
9 : business
10 : others

6. Credit amount: Loan amount

7. Value Savings/Stocks: Savings account/bonds

1 : ... < 100 DM
2 : 100 <= ... < 500 DM
3 : 500 <= ... < 1000 DM
4 : .. >= 1000 DM
5 : unknown/ no savings account

8. Length of current employment: Present employment since

1 : unemployed
2 : ... < 1 year
3 : 1 <= ... < 4 years
4 : 4 <= ... < 7 years
5 : .. >= 7 years

9. Instalment per cent: Installment rate in percentage of disposable income

10. Sex & Marital Status: Personal status and sex

1 : male : divorced/separated
2 : female : divorced/separated/married
3 : male : single
4 : male : married/widowed
5 : female : single

11. Guarantors: Other debtors

1 : none
2 : co-applicant
3 : guarantor

12. Duration in Current address: Present residence since

13. Most valuable available asset

1 : real estate
2 : if not 1 : building society savings agreement/ life insurance
3 : if not 1/2 : car or other, not in attribute 6
4 : unknown / no property

14. Age (years): Age in years

15. Concurrent Credits: Other installment plans

1 : bank
2 : stores
3 : none

16. Type of Apartment

1 : rent
2 : own
3 : for free

17. No of Credits at this Bank: Number of existing credits at this bank

18. Occupation

1 : unemployed/ unskilled - non-resident
2 : unskilled - resident
3 : skilled employee / official
4 : management/ self-employed/highly qualified employee/ officer

19. Number of Dependents: Number of people being liable to provide maintenance for
20. Telephone
    1 : none
    2 : yes, registered under the customer's name
21. foreign worker
    1 : yes
    2 : no