**Business Problem:**

Silver Sterling is a banking company also into the credit card services. The company was facing various problems cause of payment defaults by the credit card holders. They want to assess the risk of payment default for its various credit card holders by identifying defaulters and non- defaulters

**Variable Information:**

Default payment (Yes = 1, No = 0)  
X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.   
X2: Gender (1 = male; 2 = female).   
X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).   
X4: Marital status (1 = married; 2 = single; 3 = others).   
X5: Age (year).   
X6 - X11: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows: X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .;X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.   
X12-X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September, 2005; X13 = amount of bill statement in August, 2005; . . .; X17 = amount of bill statement in April, 2005.   
X18-X23: Amount of previous payment (NT dollar). X18 = amount paid in September, 2005; X19 = amount paid in August, 2005; . . .;X23 = amount paid in April, 2005.

Question:

1. What would be your dependent and independent variables?

2. How would you check for any outlier?

3. How would you check for multicollinearity amongst the variables?

4. What statistical method would you use here and why?

5. Explain the statistical output in business terms

6. From the perspective of risk management, what is more important? The overall classification result credible and non-credible results or the result of predictive accuracy of the estimated probability of default?

7. Are these variables good enough for a sound model or any some other attributes could have added value?