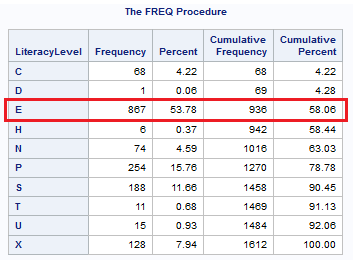
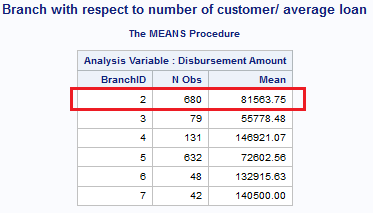
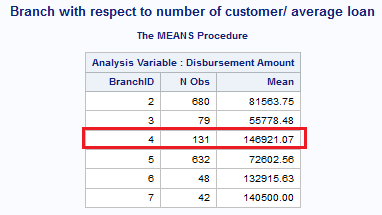
1. What Literacy Level occurs most frequently?



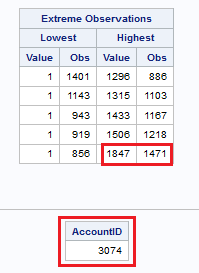
2. Which Branch is largest in terms of the number of customers?



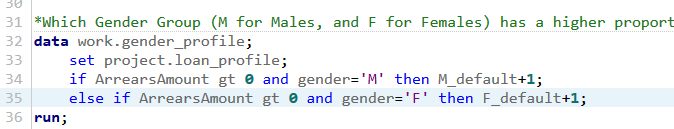
3. Which Branch disburses the largest average loan?



4. Which Customer AccountID has the largest number of days in Arrears?

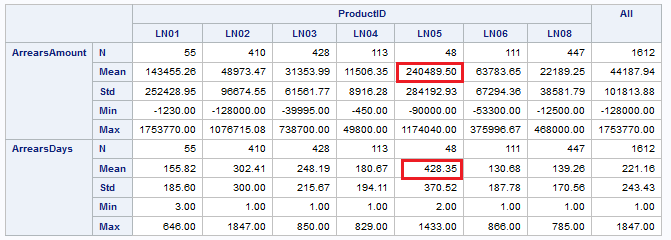


5. Which Gender Group (M for Males, and F for Females) has a higher proportion of “Bad” loans?



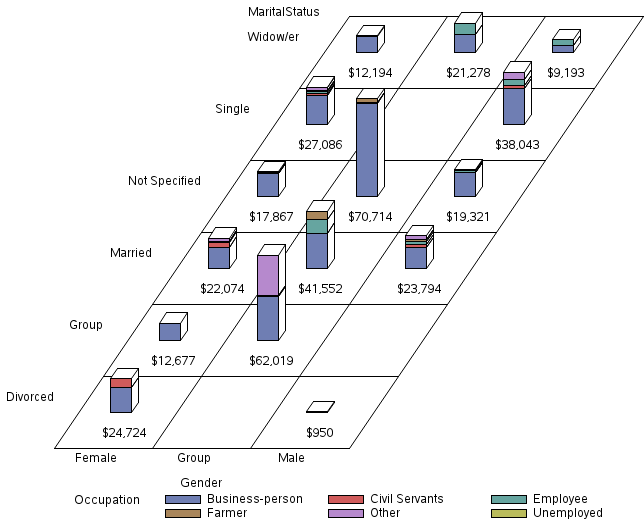
Based on the output of this code, male had higher proportion of arrearsamount.

Proc tabulate



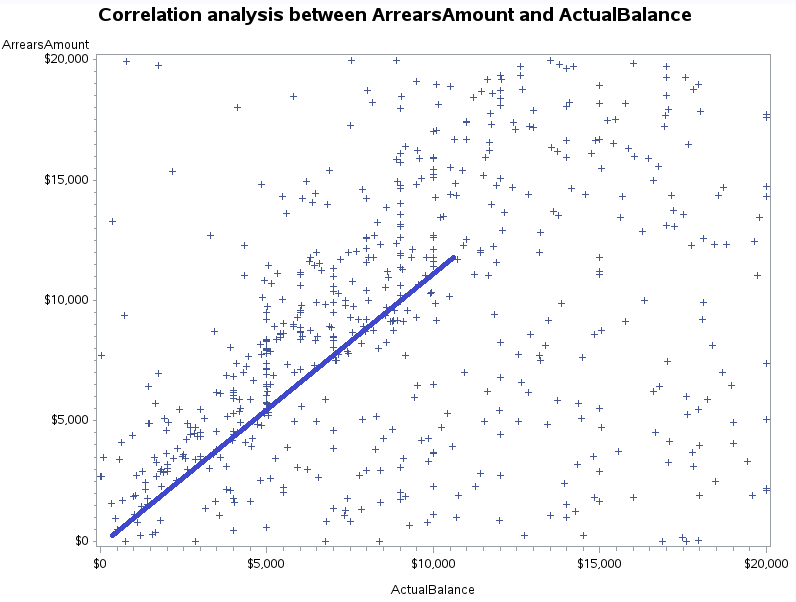
Key Finding: Product LN05 has the largest average arrear amount and the longest arrears days. This shows that LN05 is more likely to end up being a “bad loan”

Proc gchart



Key Finding: Loan taken by a “Group”, of where marital status is not specified, and where the occupation is business-person results in largest arrear amount, and hence such loans are highly risky.

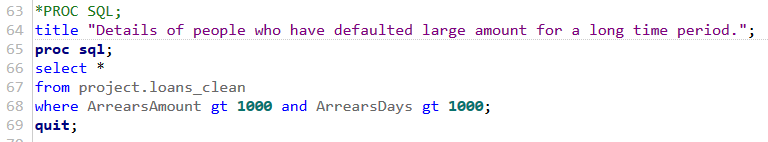
Proc gplot



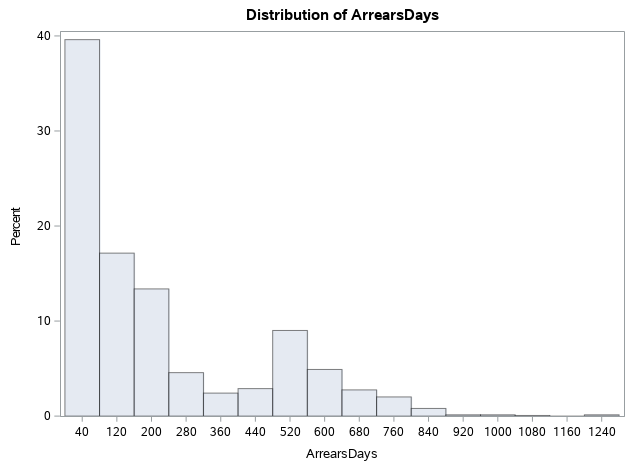
Key Finding: ArrearsAmount and ActualBalance are positively correlated, hence the bank should minimize the amount of loan which it disburses to reduce the total number of bad loan.

Proc sql

This can be used to view part of the table, where the rows satisifies certain conditions for further analysis.

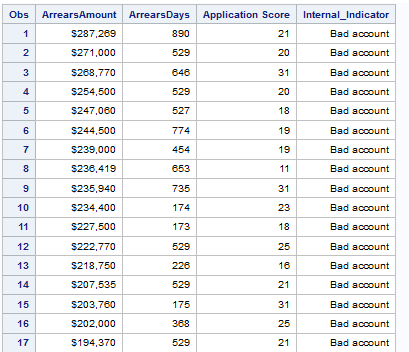


Proc univariate



Key Findings: The histogram shows that the ArrearsDays probability distribution can be modelled suing a exponential probability distribution function.

Proc sort



Key Finding: By using proc sort, it was determined that accounts with large arrears amount and long arrears time duration were correctly classified as “Bad Account”

Proc contents

This proc can be used to detect whether anything had gone amiss during the data cleaning process.

