EXPLORATORY DATA ANALYSIS ASSIGNMENT

AGENDA:

To perform Exploratory Data Analysis of a given dataset.

UNDERSTANDING THE DATA SET:

The given dataset is of a lending institution which has given loan to various customers/debtors based on certain parameters like credit rating, debt to income ratio, type of account. The dataset is in xlsx format having 51 entries. I have used Tableau and Excel as tools for this assignment. I have re-named some of the column headings based on my Assumption. The data set is given below.

Status	Credit Grading	Amount of Borrowing	Duration of Borrowing	Rate of Borrowing	Debt.To.Income.Ratio
Current	C	5000	4	0.15	0.04
Current	HR	1900	6	0.265	0.02
Current	HR	1000	3	0.15	0.02
Current	HR	1000	5	0.29	0.02
Current	AA	2550	8	0.0795	0.033
Current	NC	1500	2	0.26	0.03
Current	HR	3001	6	0.2875	0.02
Current	E	2000	6	0.25	0.02
Current	AA	2550	1	0.0875	0.04
Current	A	1500	4	0.085	0.04
Current	AA	2500	5	0.075	0.01
Current	HR	2000	8	0.2375	0.02
Current	AA	7500	8	0.08	0.023
Current	D	3000	7	0.242	0.03
Current	D	1000	8	0.2195	0.013
Current	B	1000	10	0.15	0.02
Default	HR	1000	8	0.2375	0.01
Current	HR	1000	10	0.149	0.015
Current	AA	2500	8	0.08	0.016
Current	HR	1170	5	0.27	0.01
Current	AA	2550	6	0.05	0
Current	AA	3000	6	0.0849	0.01
Current	HR	2000	5	0.25	0.03
Current	A	5000	5	0.1019	0.03
Current	HR	1000	1	0.29	0.02
Current	NC	1000	1	0.05	0.01
Current	E	2000	1	0.15	0.02
Current	A	1500	4	0.0975	0.02
Current	HR	1100	8	0.23	0.01

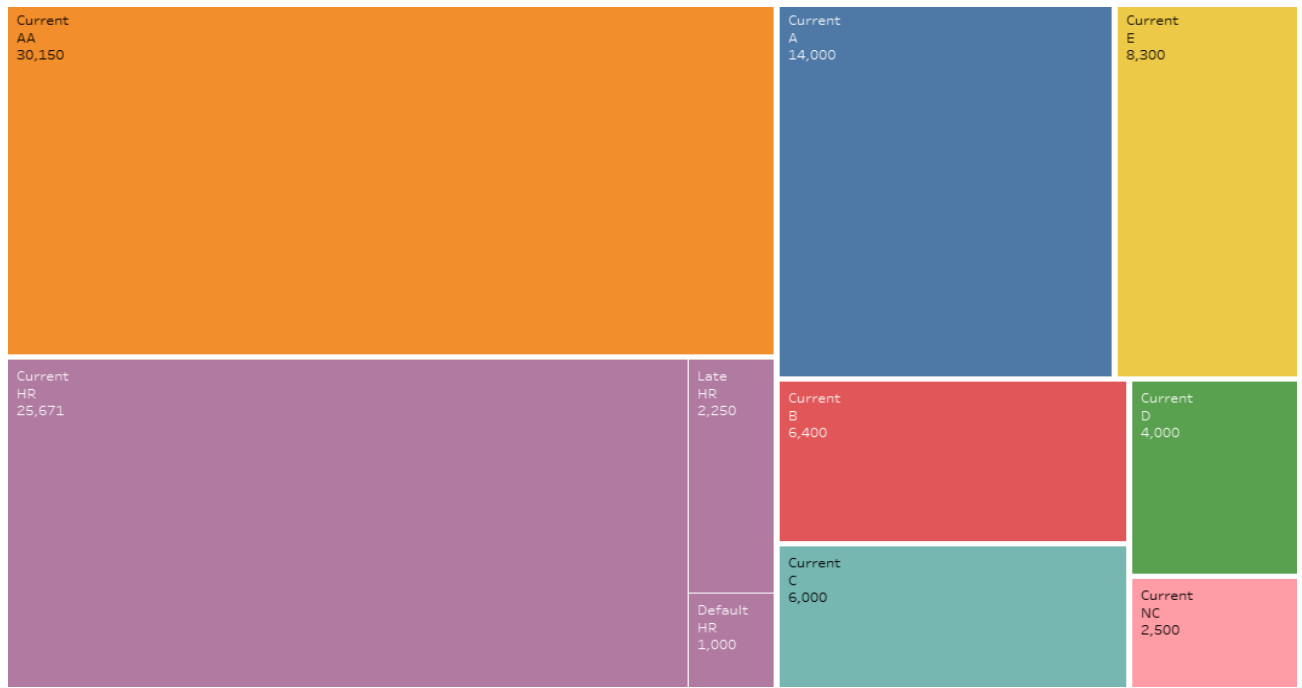| Current	HR	1800	7	0.2075	0.03
Current	HR	1000	8	0.21	0.013
Current	A	1000	3	0.0725	0.01
Current	E	2300	7	0.2025	0.02
Current	E	1000	3	0.23	0.01
Current	HR	1700	1	0.19	0.03
Current	C	1000	9	0.2	0.018
Current	HR	1000	7	0.2	0.01
Late	HR	1200	6	0.24	0.01
Current	HR	1000	10	0.1775	0.031
Current	AA	1000	6	0.074	0.03
Current	HR	3000	10	0.22	0.023
Current	B	3000	1	0.09	0.01
Current	AA	2000	1	0.0774	0.03
Current	HR	1000	3	0.29	0.03
Current	A	5000	9	0.11	0.031
Current	AA	3000	3	0.08	0.02
Current	AA	1000	3	0.07	0.04
Late	HR	1050	7	0.2875	0.03
Current	E	1000	1	0.205	0.03
Current	B	2400	3	0.085	0.04

ANALYSIS:

1)
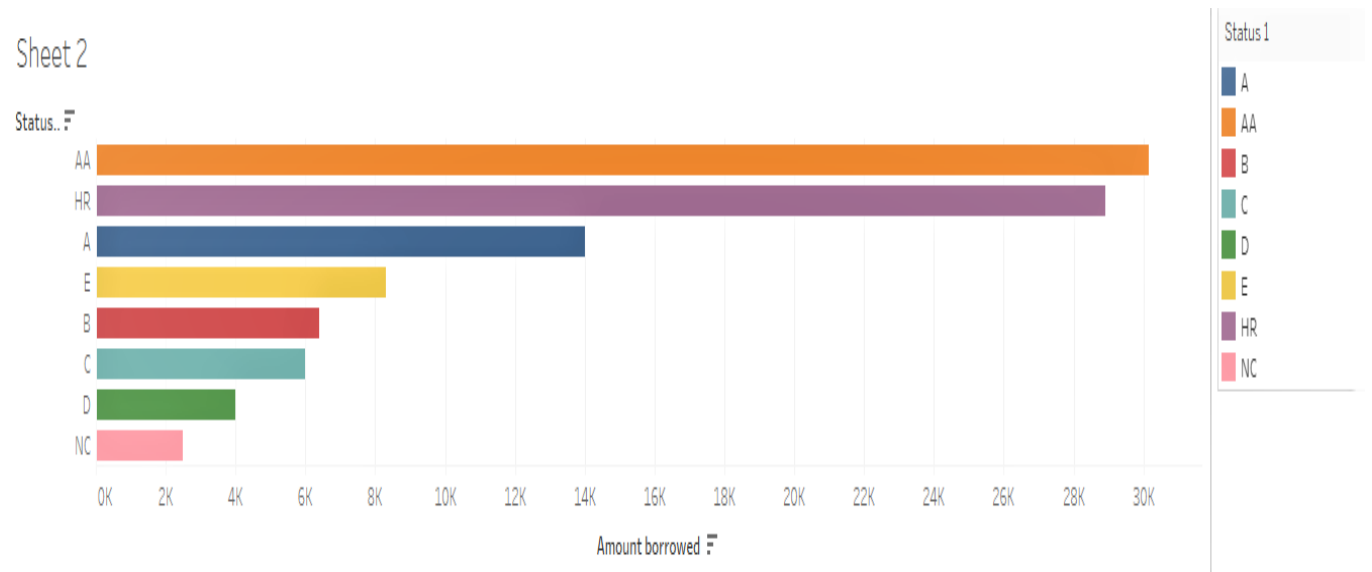
In The above graph the Status 1 (Credit grading) is taken on the Y- Axis and Amount Borrowed has been taken on the X- Axis. The above graph shows the fact that the Top 3 borrowers are of the grading AA, HR and A.

2) The tree maps given below show the fact that some borrowers graded HR being one of the large borrowers have late payments and some have even defaulted. This needs to be investigated. The organization should properly classify such loan in its balance sheet. The late payment (Rs.2250) and default (Rs.1000) comprises of 11.24% of the total loans granted to HR.

Current AA 30,150		Current A 14,000	Current E 8,300
Current HR 25,671	Late HR 2,250	Current B 6,400	Current D 4,000
	Default HR 1,000	Current C 6,000	Current NC 2,500

3) The boxplot below represents summary of the distribution with regards to amounts borrowed and Credit grading. The average amount borrowed is Rs 12533.88. The median is Rs 7500. The upper quartile is Rs 21460.50 the upper whisker is Rs 30150 the lower quartile is Rs 5000 the lower whisker is Rs 2500 and the Inter quartile range is Rs 16460.50. The values outside the box are AA, HR NC and D. The points AA lie at 30150 which is the upper whisker and HR lies at the point 28921. The bigger values towards the right of the box are AA and HR towards the higher side which pull the mean towards the higher side. The distribution is hence right skewed. NC and D lie below the box.
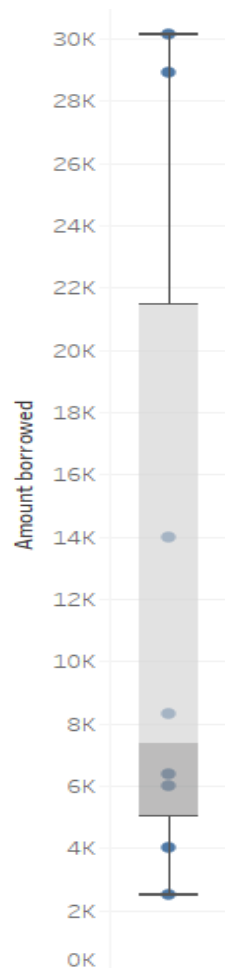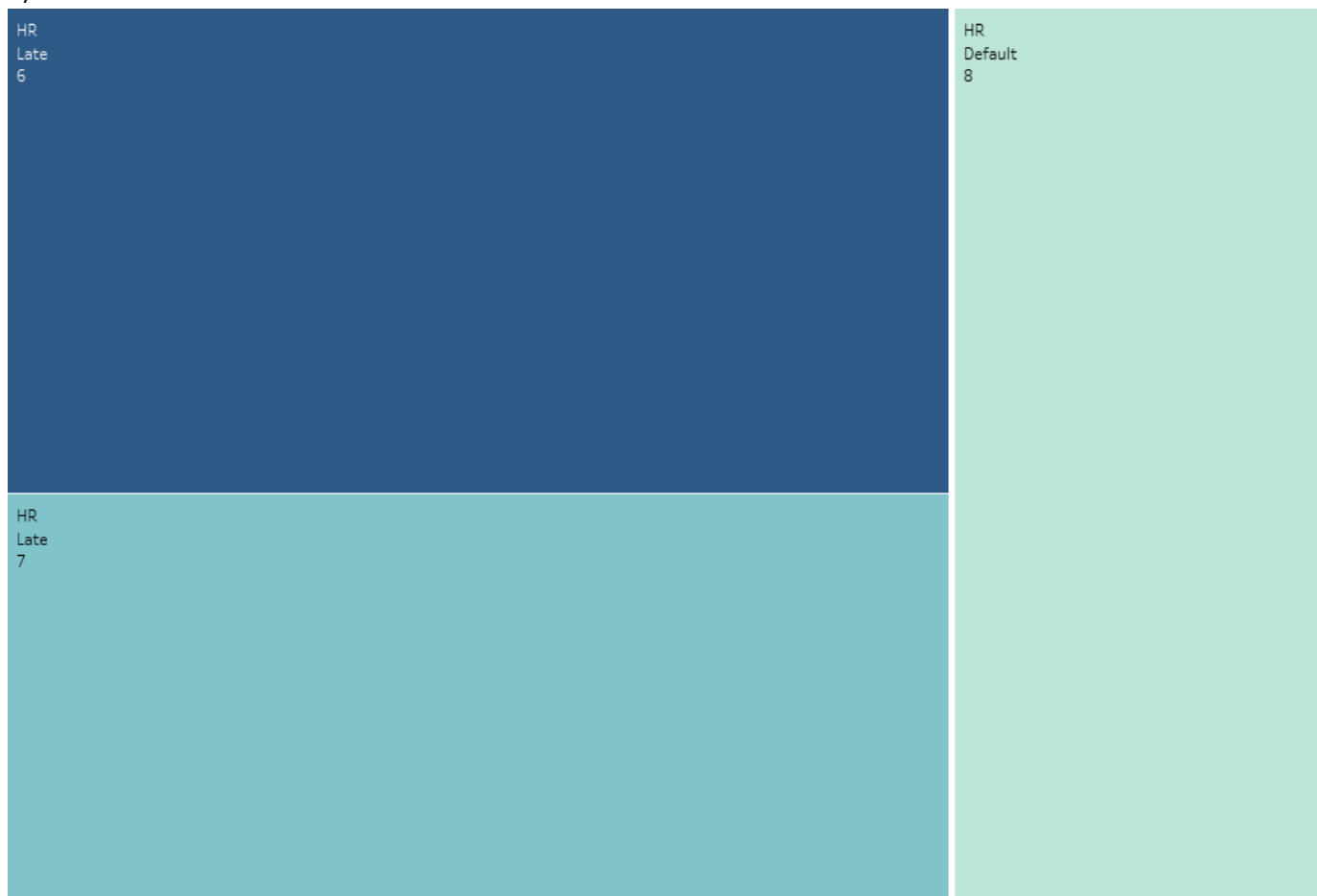
4)
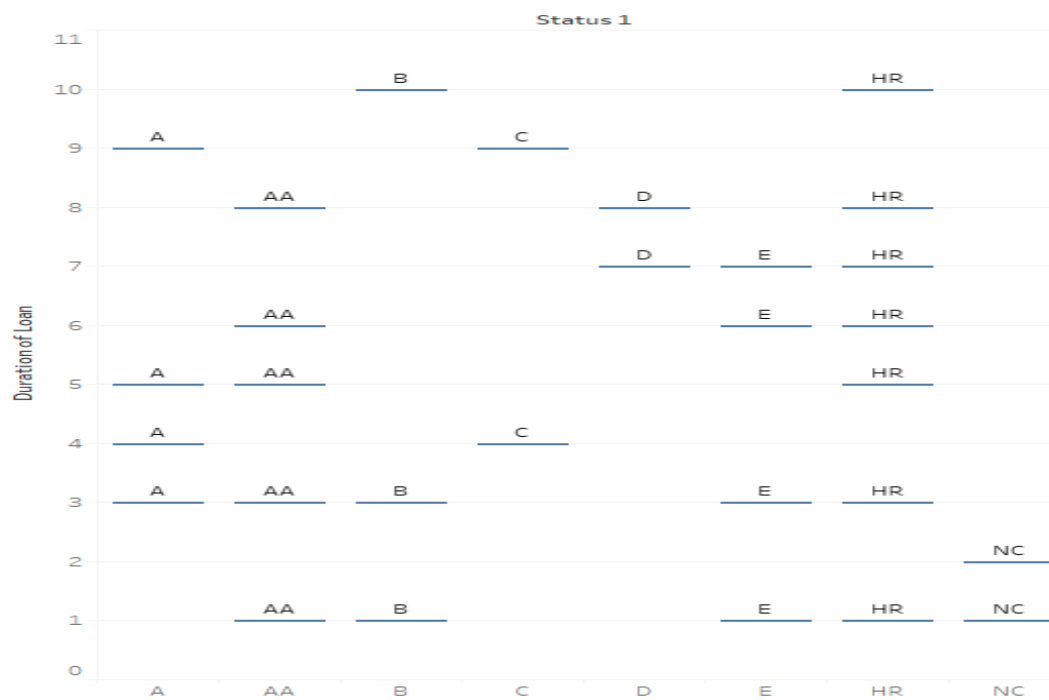
The Gantt chart given above represents the duration of the loan for each rating. We observe that HR and B have the highest duration of Loan followed by A and C then by AA. AA which is the second highest borrower has comparatively lower duration.

5)

The Tree-map given above shows us that the defaulters and late payers of HR have a high duration borrowing. This could also be due to these balances appearing in the account and no amount is recoverable from them. This fact needs to be investigated. If that is the case these balances should be written off from the books of account after obtaining necessary approvals.

6) The bar graph below represents Average debt to Income ratio on the Y axis and Credit rating on the X Axis.It reveals the Borrowers C,A,B have highest average debt to income ratio. This also reveals the fact that HR being the highest borrower has the lowest Average Debt to Income Ratio. This could be due to high Income levels of the HR Borrowers. AA being the second highest borrower has the 3rd lowest debt to income ratio which is due to high income. NC being the lowest borrower has the Average Debt to income ratio same as HR. The lenders should be careful while lending high amounts to low income group

Status 1
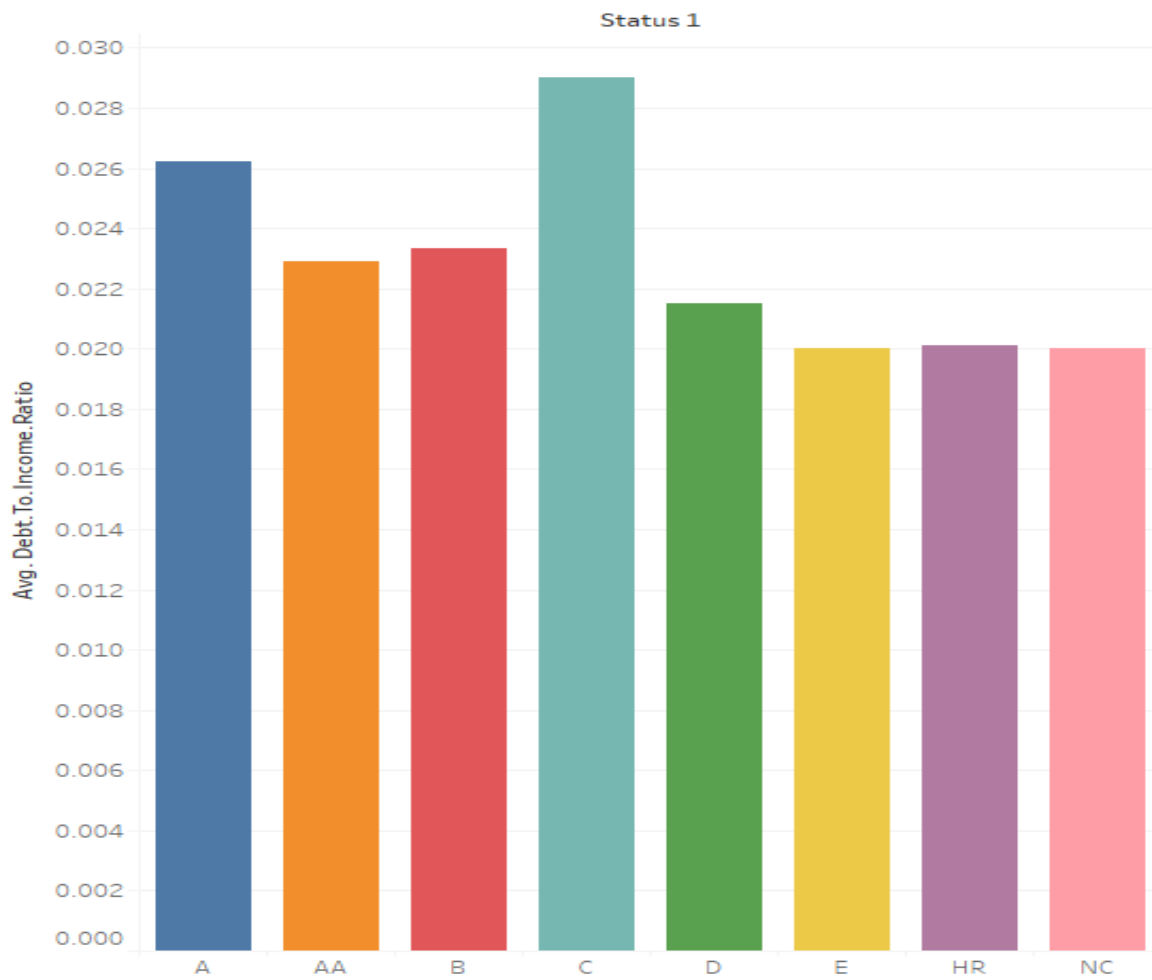
7) The boxplot given below is with respect to Average debt to Income Ratio and Credit grading. The median is 0.022205. The upper and lower quartiles are 0.024767 and 0.020050 respectively. The Interquartile range is 0.004717. The upper and Lower Whiskers are 0.029 and 0.020. The distribution is more or less uniform.
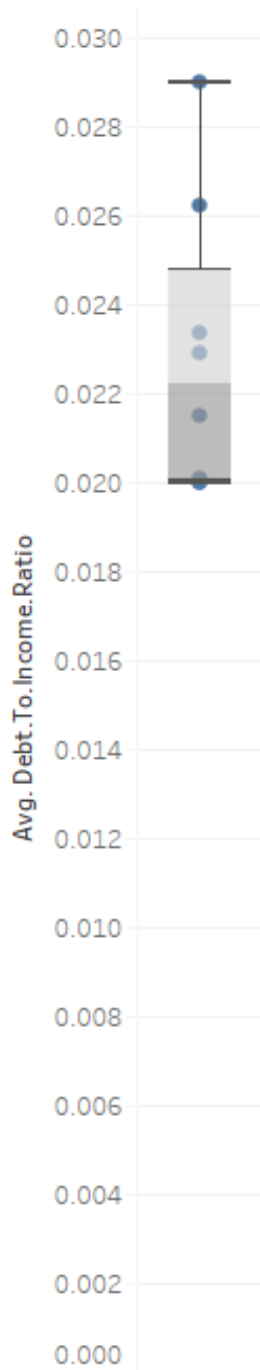
8) The bar chart below shows the fact that HR D and E grading have borrowed an amount bearing the highest Average rate of Interest. It is interesting to Note that AA having highest borrowed amount has comparatively lower Average Interest Rate. NC having lowest borrowed amount has a comparatively higher Average rate of Interest.

Status 1
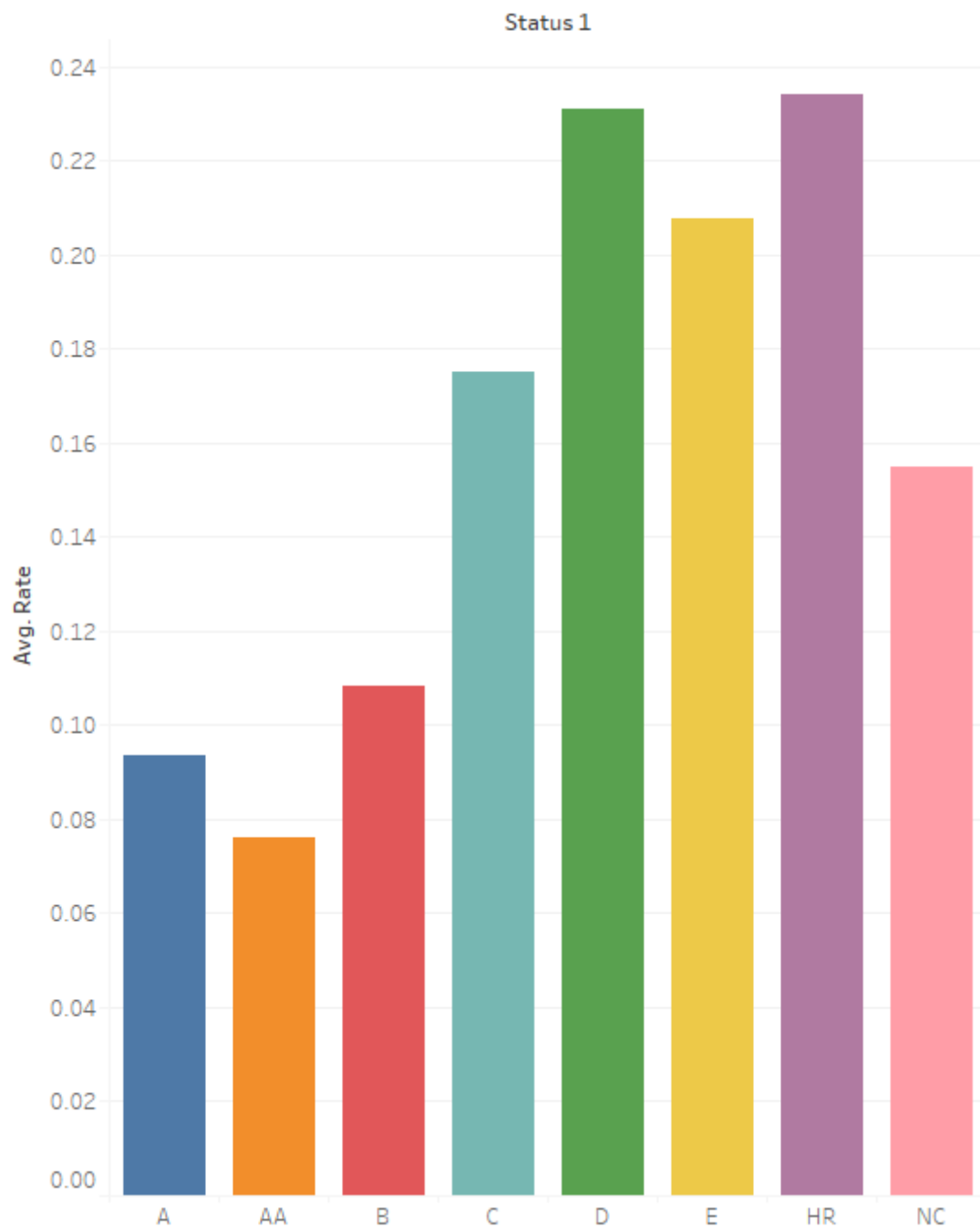
9) The stacked bar chart shows the fact that HR having highest Average Interest Rate has a history of Late Payment and second highest Average rate of Interest has defaulted. This needs to be analysed.
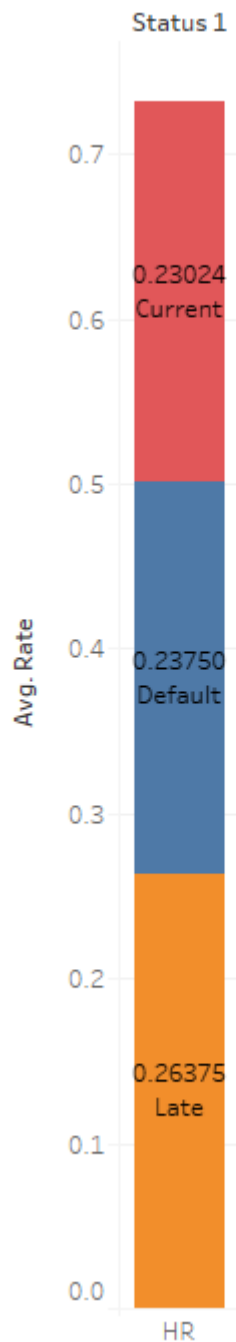
CONCLUSION:

It can be seen that Exploratory Data Analysis is necessary to find hidden information in the dataset so that we can take proper decision in a business. We have taken help of various graphs such as Boxplots, Bar graphs, tree-maps, gantt chart.