

Lessons Learned Report

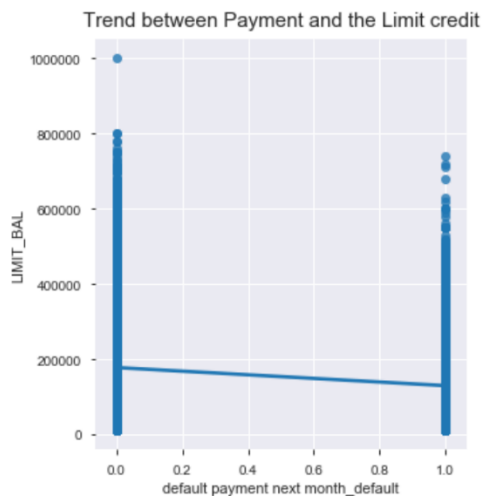
Credit one is having issues, increasing in customer default rates which could end losing revenue and customers for its clients and of course for Credit One.

One of the functions I have used a lot to check for data details is called **GroupBy** in Pandas which together with **matplotlib** and **seaborn** for visualization are very helpful to slice, dice and find different insights from data.

Furthermore, in order to analyze the data in a better way, discretization is a great technique, using the function **cut** in Pandas helps with this goal.

Finally, correlation and covariance give a good guideline about relationship between the variables which help to identify some strong relationships between features that could cause issues in the modeling.

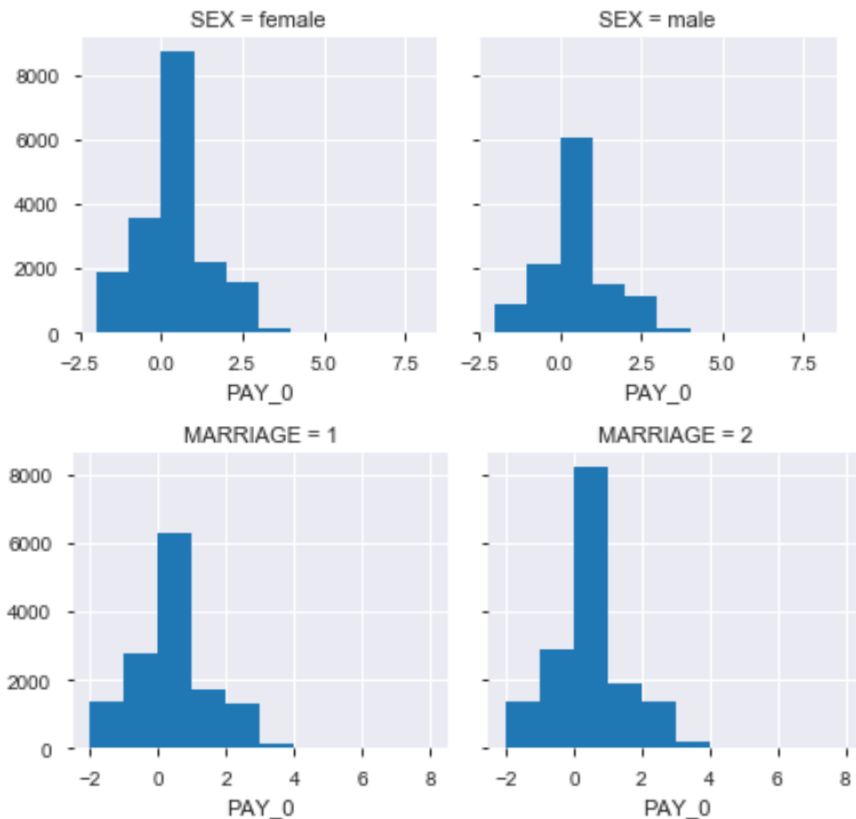
- ✓ 70% of the customer's credit limit goes from \$10,000 to \$200,000, and from the 6,636 customers (22% of the total) who will default 78.6% will fall in this range. Therefore, a lot of attention need to be done in this range.
- ✓ There seems to be a general linear trend between Payment and the Limit credit. The visualization shows that the higher the limit is, the less chance he/she would default.



- ✓ 92% of the customer's age goes from 21 to 50 years old, and 72% are below 41 years old. From the 6,636 customers who will default 70% will fall in this range (below 40 years old) There is a trend in Age level, the older the customer, the less chance he/she would default.
- ✓ 47% of the customer's education falls at University level, and only 16% in high school. However, from the 16% in high school, 25% will default with the highest percentage, follow by university level with 24% and graduate school with 19%. There is a trend in Education

level similar to the Age, the higher the level of education, the less chance he/she would default.

- ✓ Other insights from data are: 60% of the customers are female and the difference between married (1) and single (2) is not very relevant, 45% and 53% respectively. Furthermore, single and female customers tend to delayed more in their payments.



- ✓ Using correlation and covariance analysis, there are strong relationship between some of the variables. Therefore, there is a possibility of collinearity problems.

In addition, using the same functions mentioned at the beginning of the document, I have checked the payment status were -1 means “Paid in Full”. After checking this option in all the data, there are still 284 customers who have “Paid in Full” all the time, and they are classified as “default”. These data could be taken out for further analysis.

In short, a lot of attention and more analysis to be required for younger people with lower level of education. Probably, a reduction in the credit limit could help they do not default.

On a side note, I have found the following error: *KeyError: "['default payment next month_default'] not in index"* which means the variable is not within the name of the data columns. However, I realized that the name of the column/variable had been change after running the function: “profile_report” from “*default payment next month_default*” to “*default payment_next_month_default*”.