Analysis of Defaulter

From the perspective of risk control, estimating the probability of default will be more meaningful than classifying customers into the binary results – risky and non-risky

Table of Contents

What is Outstanding Amount

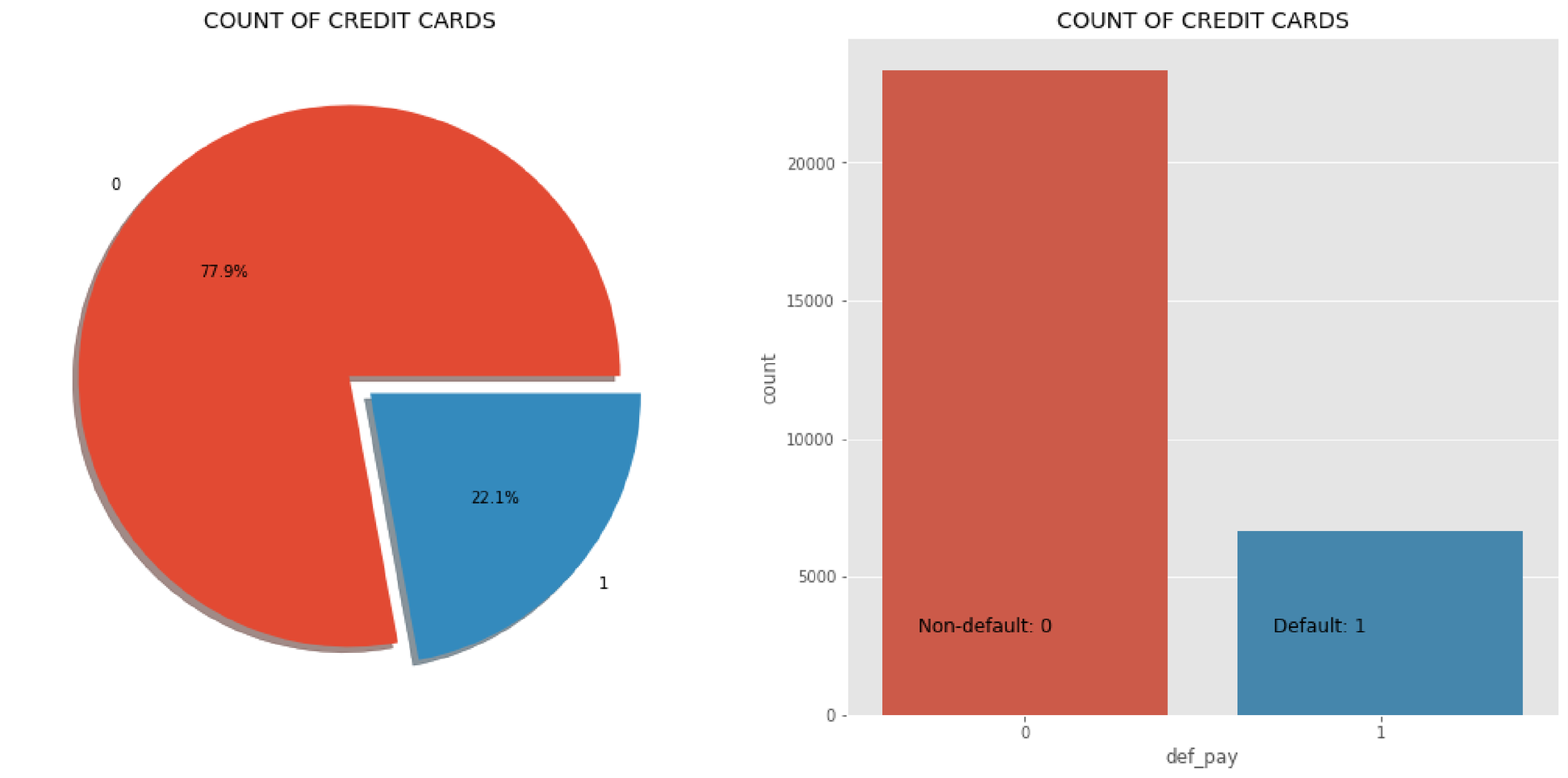
Outstanding Amount
Card issuers assign specific credit limits (spending limits) on your cards.
Your outstanding balance helps determine how much credit (money left to spend) you have available.
To find out how much you have left, simply subtract your outstanding balance from your credit limit.
In simple words Money utilized

Outstanding Amount for Bank will be Sum of limit Balance

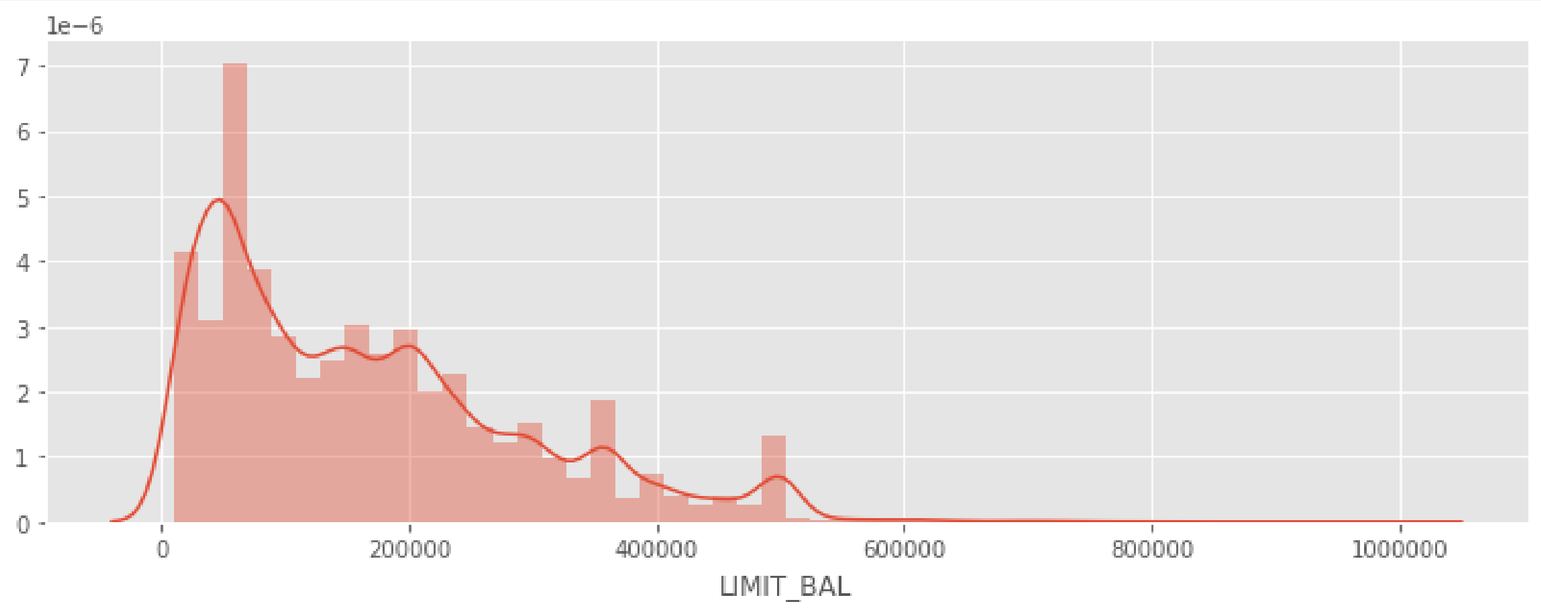
General Trends in our data

- The average value for the amount credit card limit is 167,484 NT dollars.
- The standard deviation is 129,747 NT dollars, ranging from 10,000 to 1M NT dollars.
- Education level is mostly graduate school (1) and university (2).
- Most of the clients are either married or single (less frequent than the other status).
- The average age is 35.5 years, with a standard deviation of 9.2 years.

'COUNT OF CREDIT CARDS'

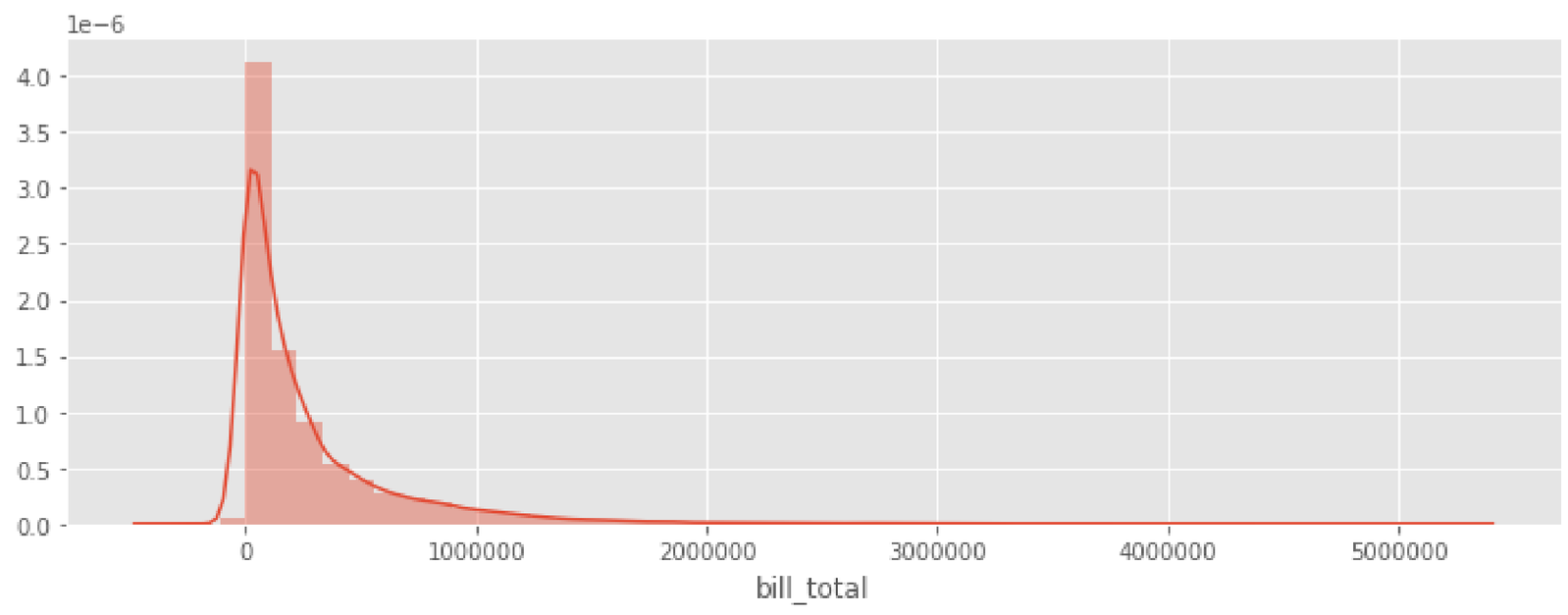


Mapping the target: categorizing From this sample of 30,000 credit card holders, there were 6,636 default credit cards; that is, the proportion of default in the data is 22,1%



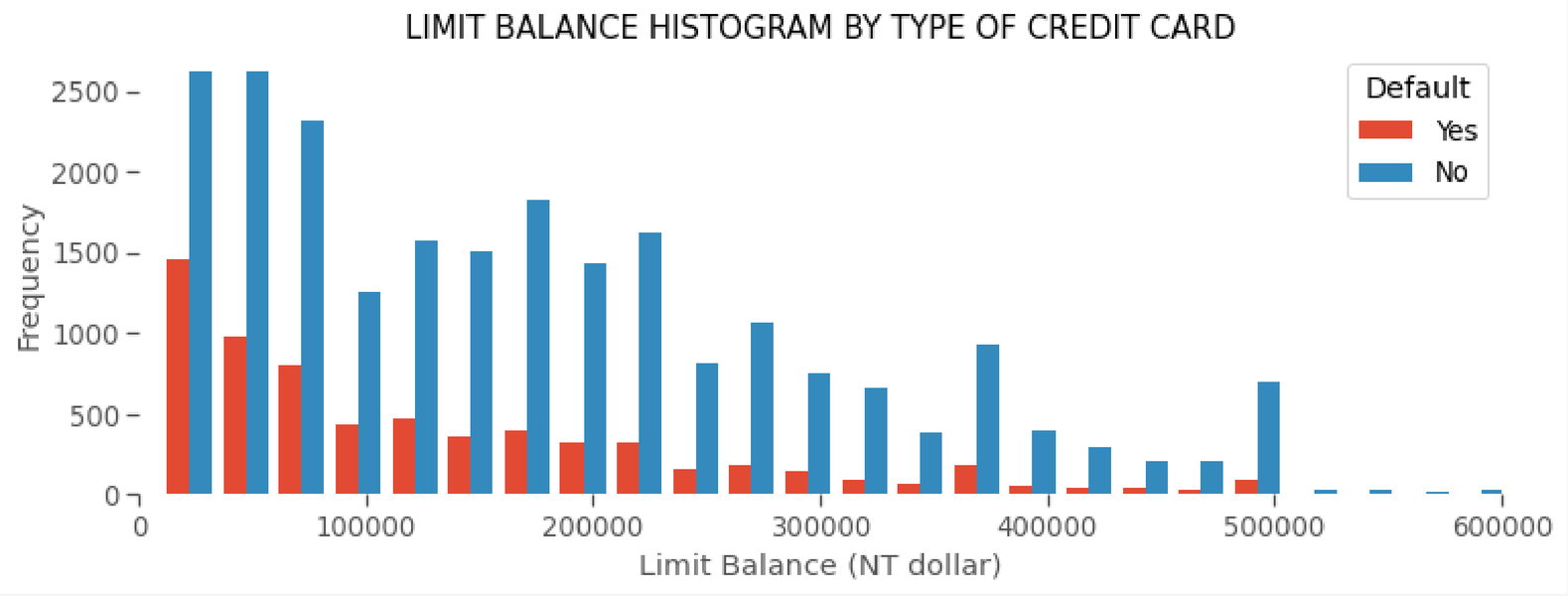
Outstanding Amount for Bank will be Sum of limit Balance

- The trend on the outstanding amount for the bank



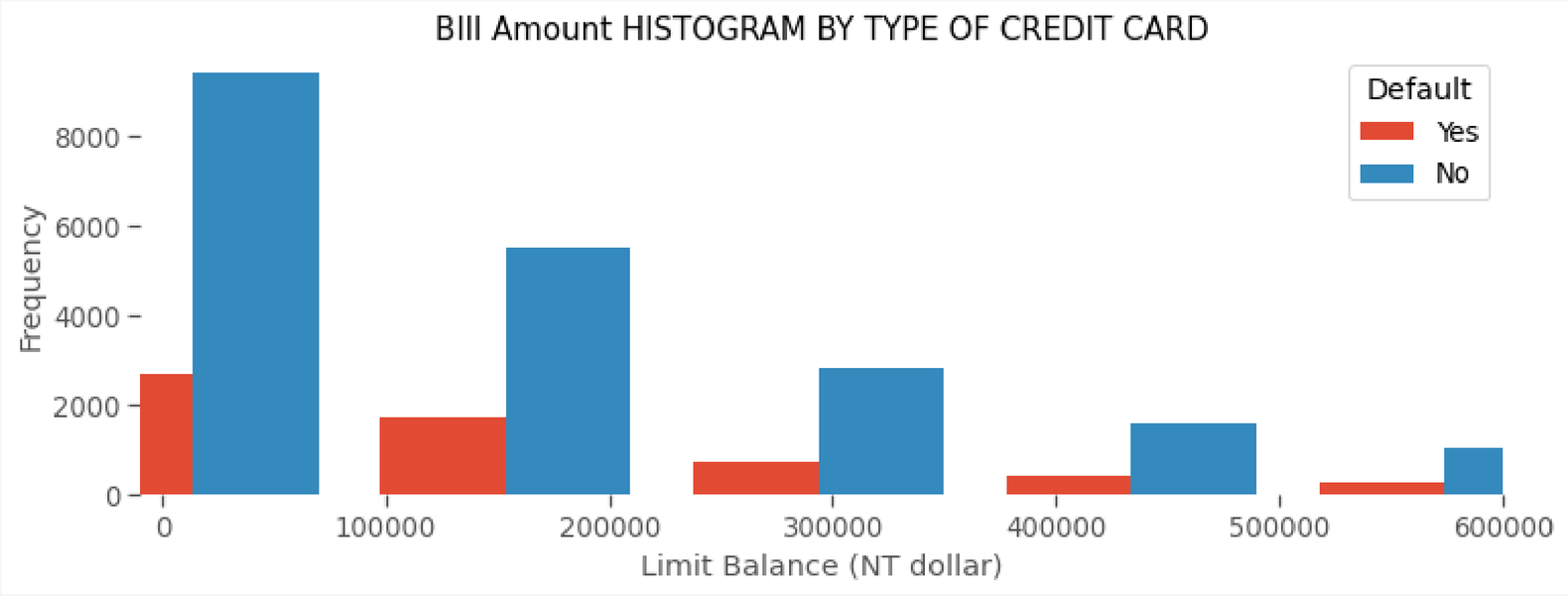
The trend on the outstanding amount for the Customer

The negative sign indicates an overpayment of the bill and you may be eligible for a refund.

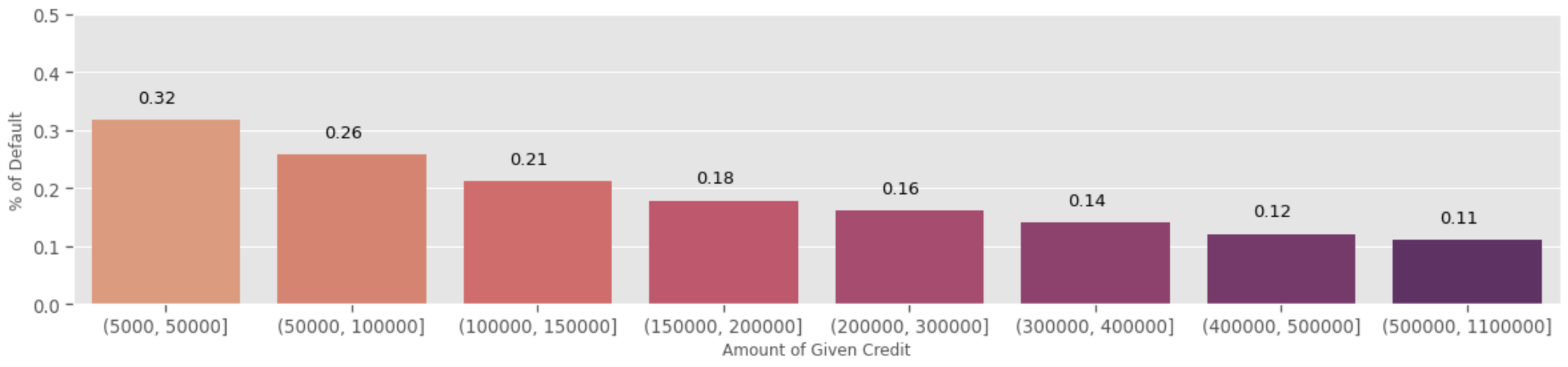


LIMIT BALANCE HISTOGRAM BY TYPE OF CREDIT CARD

BILL AMOUNT HISTOGRAM BY TYPE OF CREDIT CARD

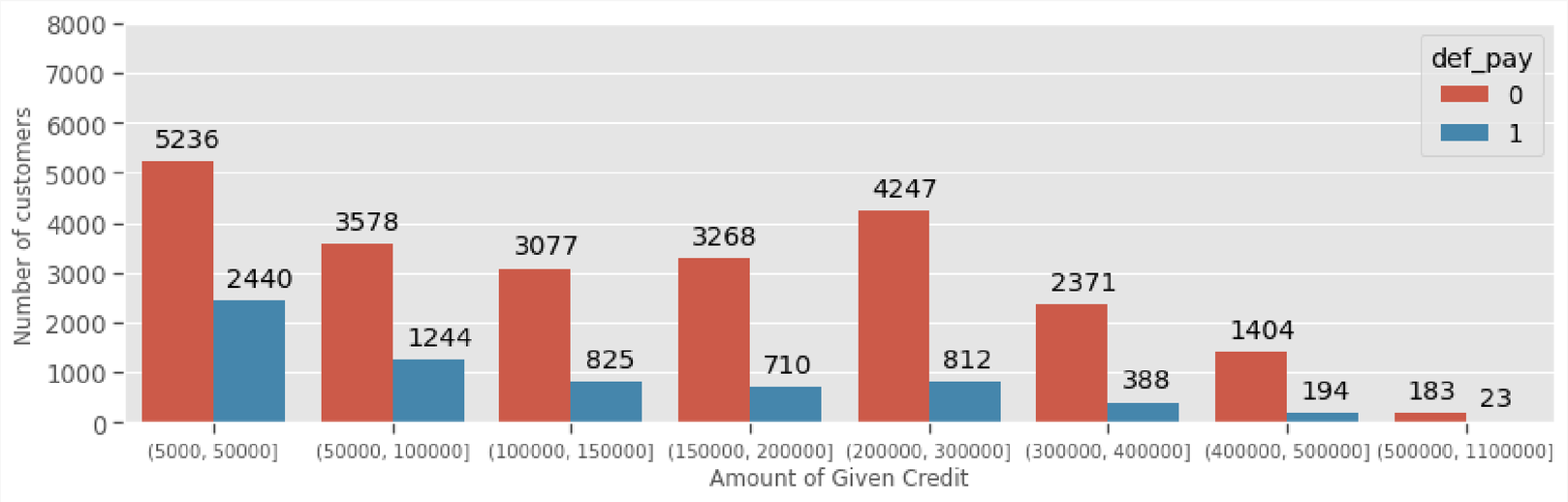


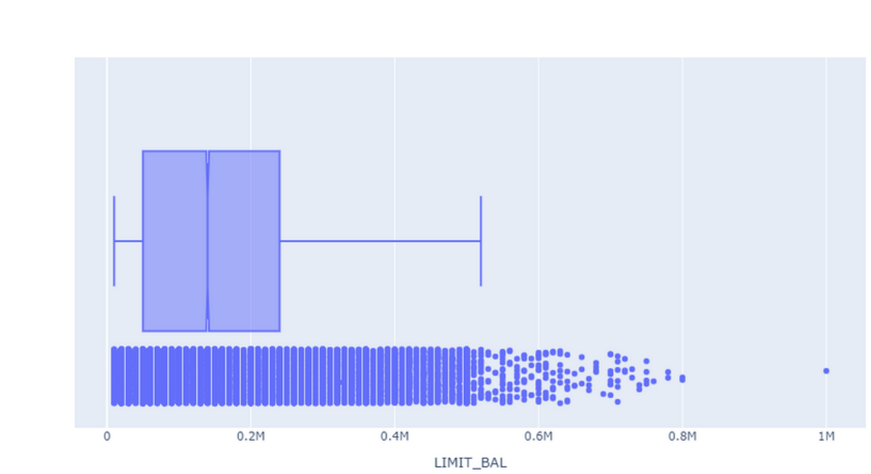
Number of customers with the max outstanding amount (in different outstanding amount buckets)



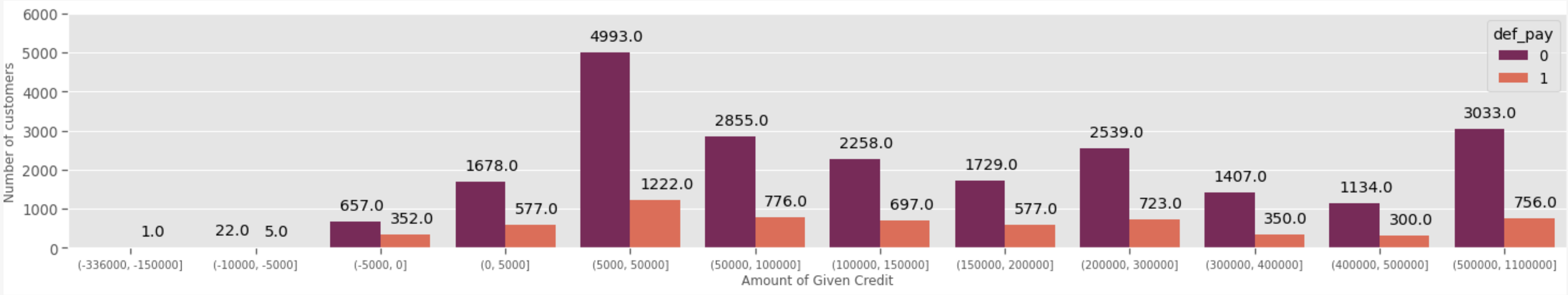
Outliers are clearly accounted for in both plots.

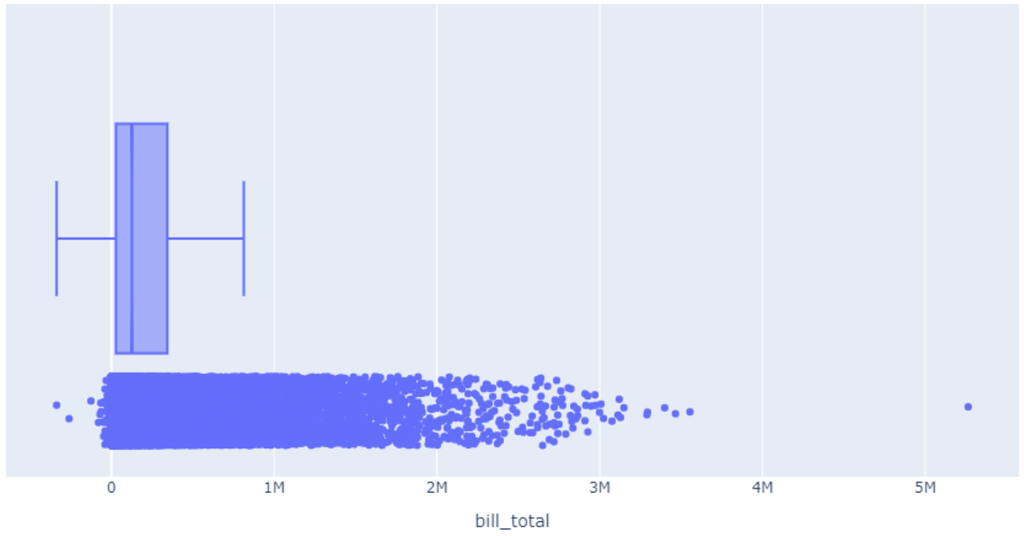
The analysis shows a significant rate of default (over 30%) from customers with 50k or less of the credit limit. We also can see that the higher the limit, the lower is the chance of defaulting.





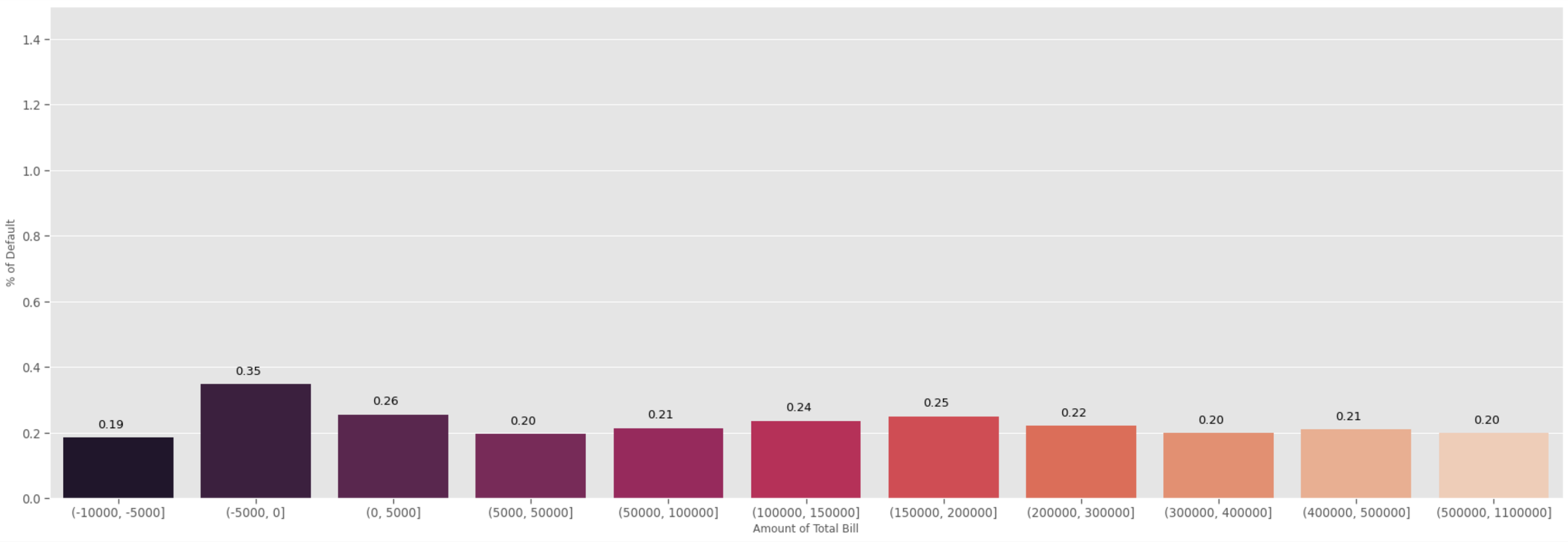
Analysis of outstanding Amount of customer





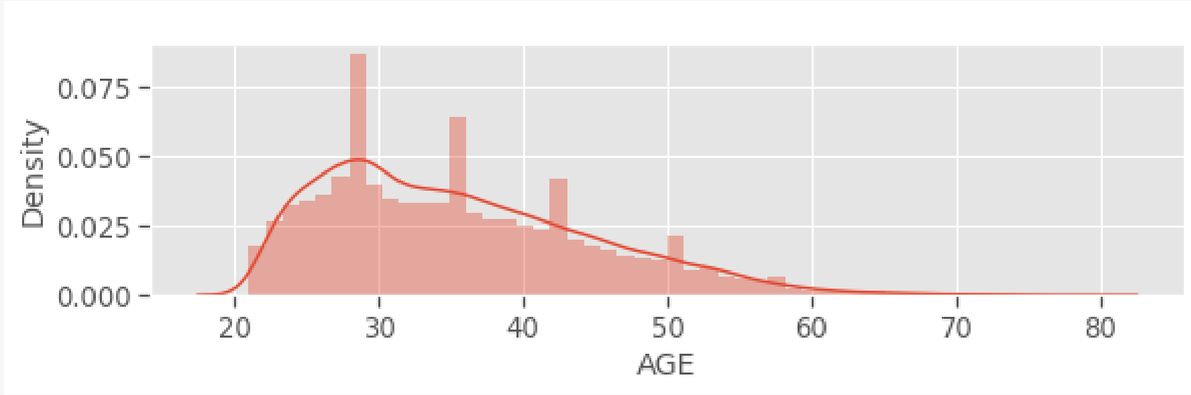
Outliers are there which need to be removed as they will be affecting our analysis negatively

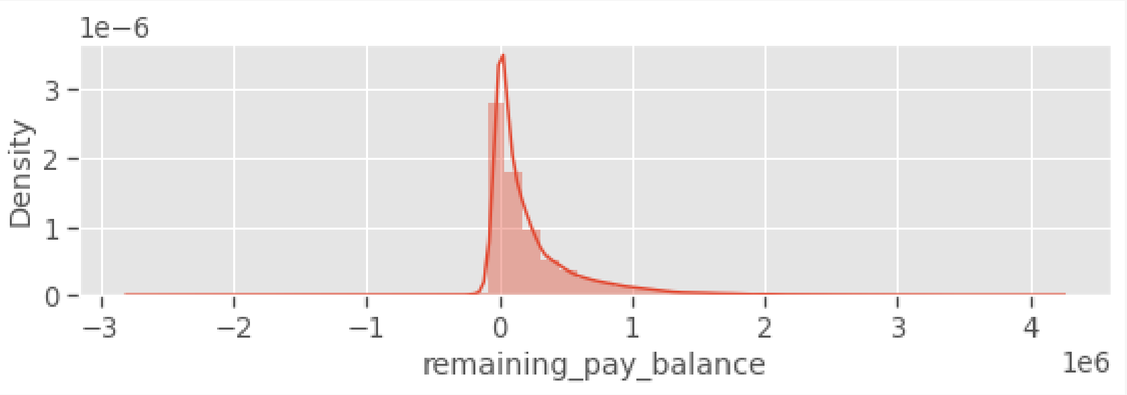
The extreme negative value should be neglected as it is the outliers



As expected, those who have a negative bill statement have a lower chance of default than the rest. What stands out is that there is a little higher chance of default for those who didn't have a bill in the previous months.

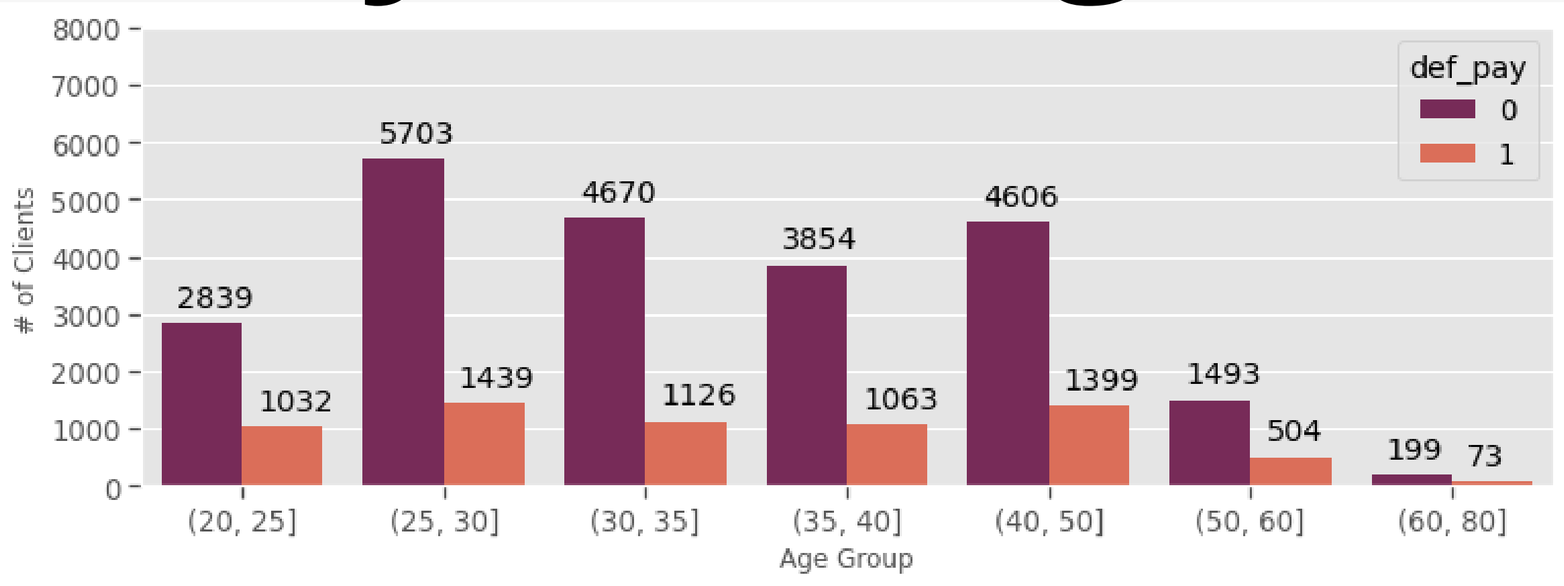
Distribution of Age

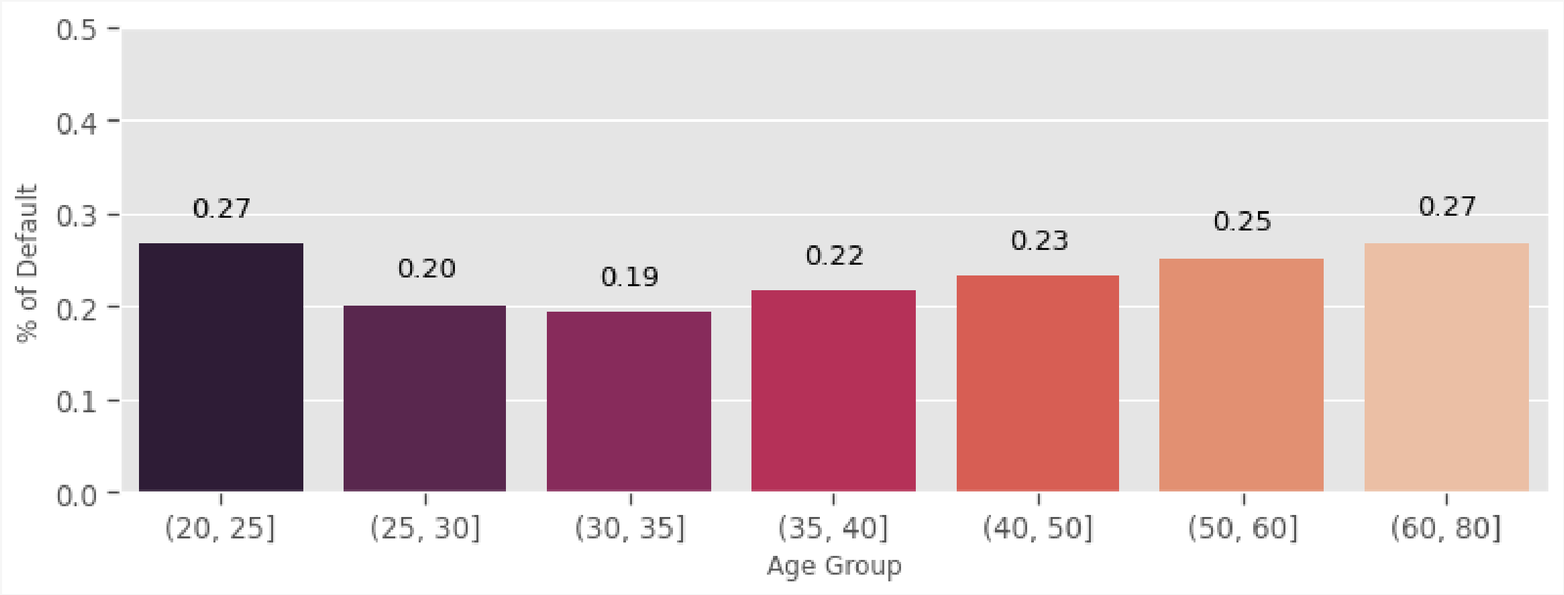




Distribution of Remaining Balance

Analysis of Age

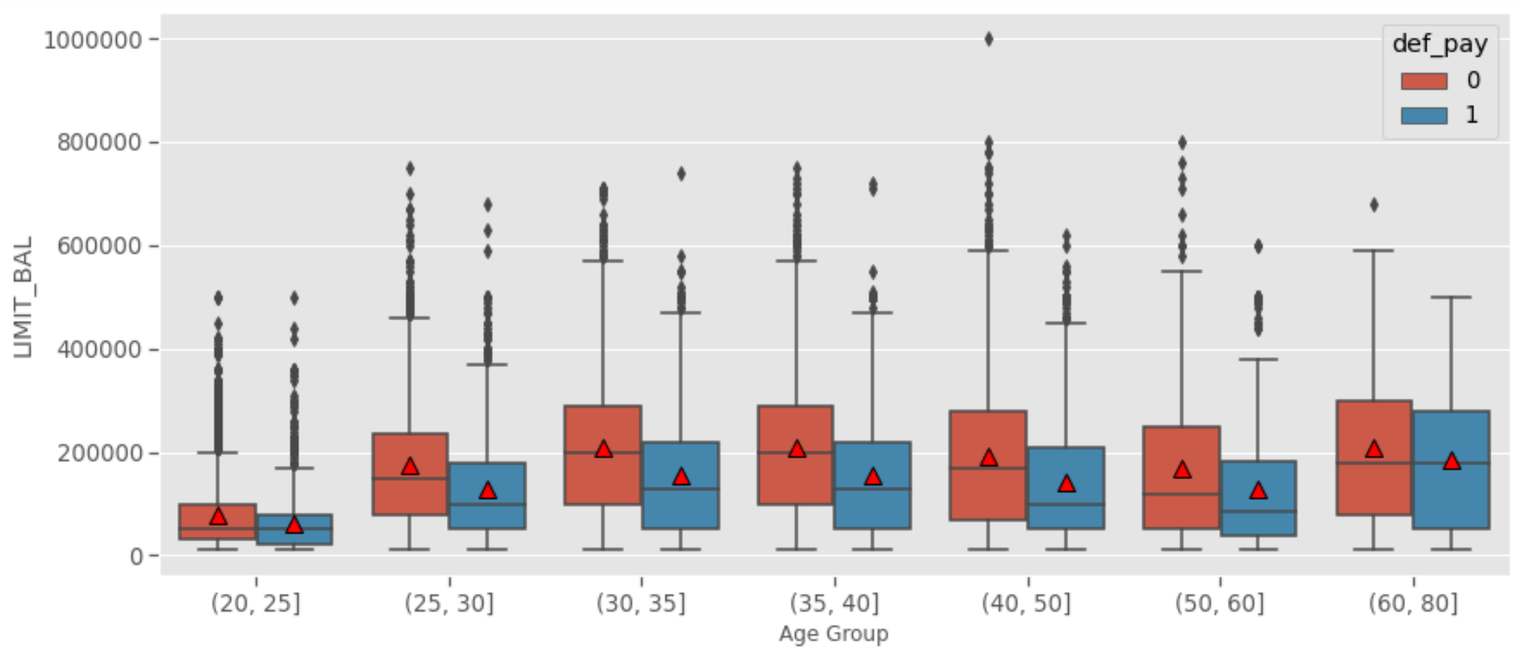




As expected, the lowest chances of default are registered around 30-35 years old, while the highest occur at the extremes (20-25 and 60+).

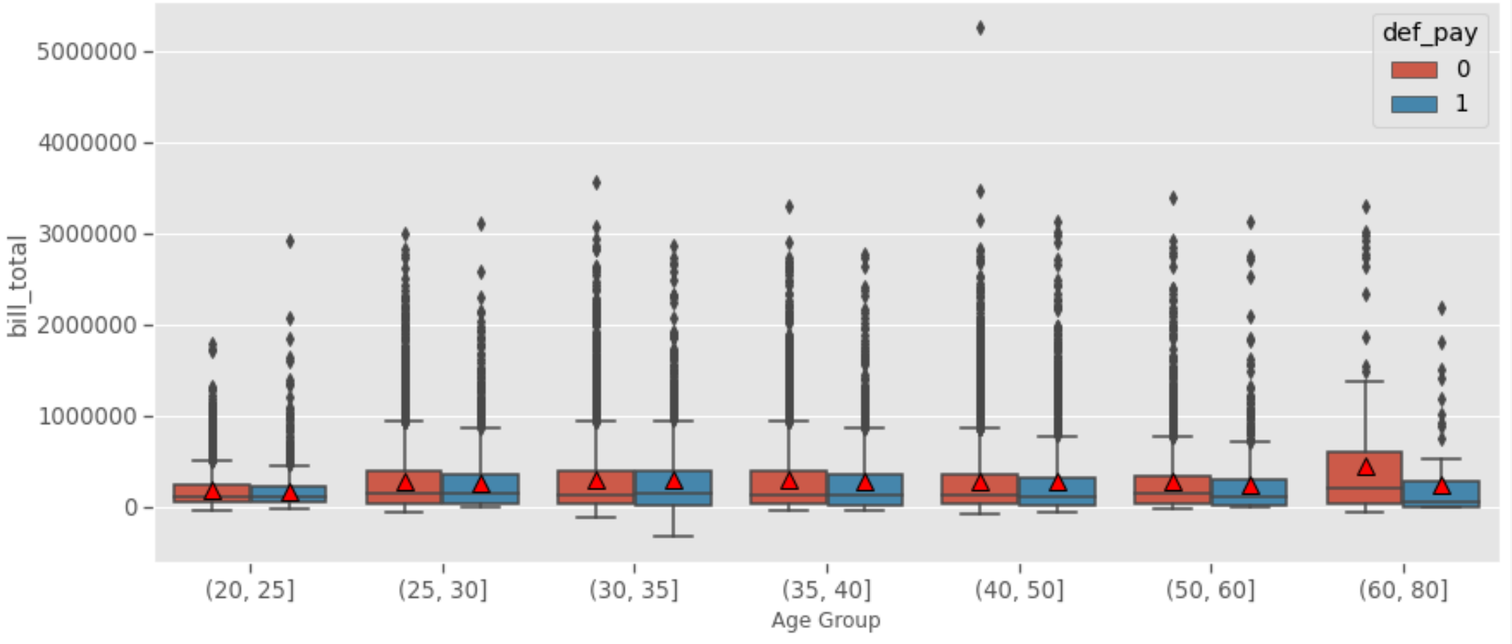
(Though keeping the numbers in Mind)

Analysis of age with Limit and Bill

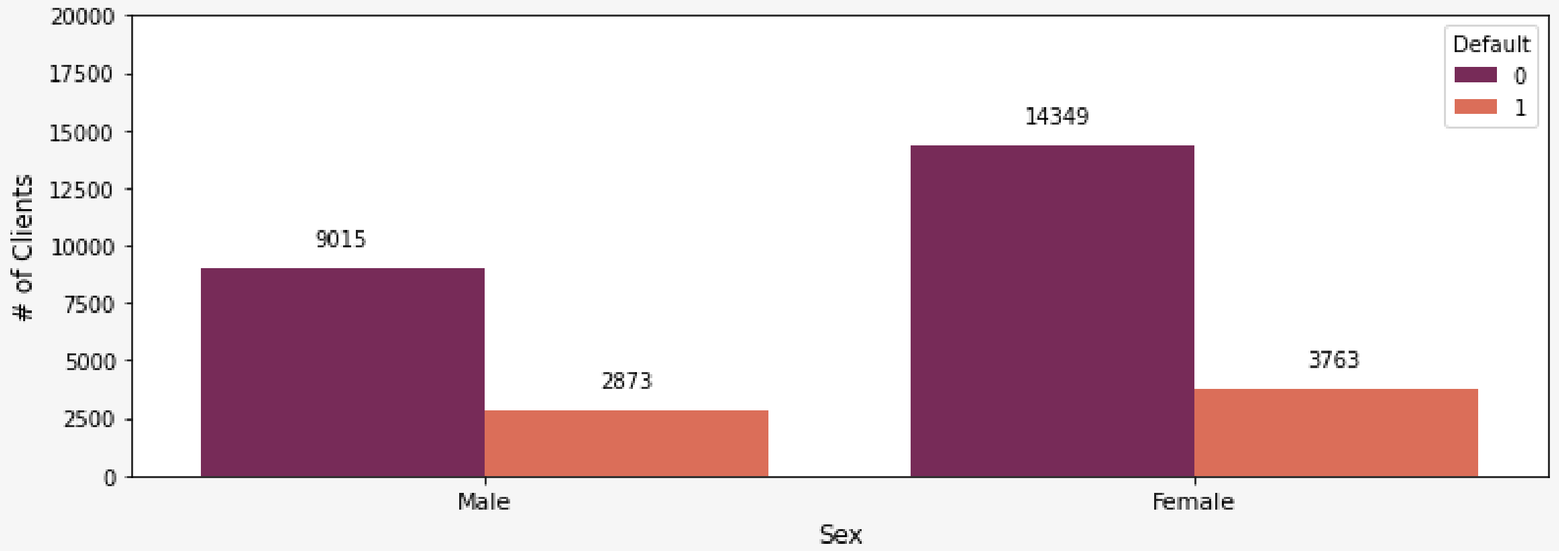


The 60+ people are less but are playing a role in our analysis

We can evidently see that Age is a vital component as with it the mean varies

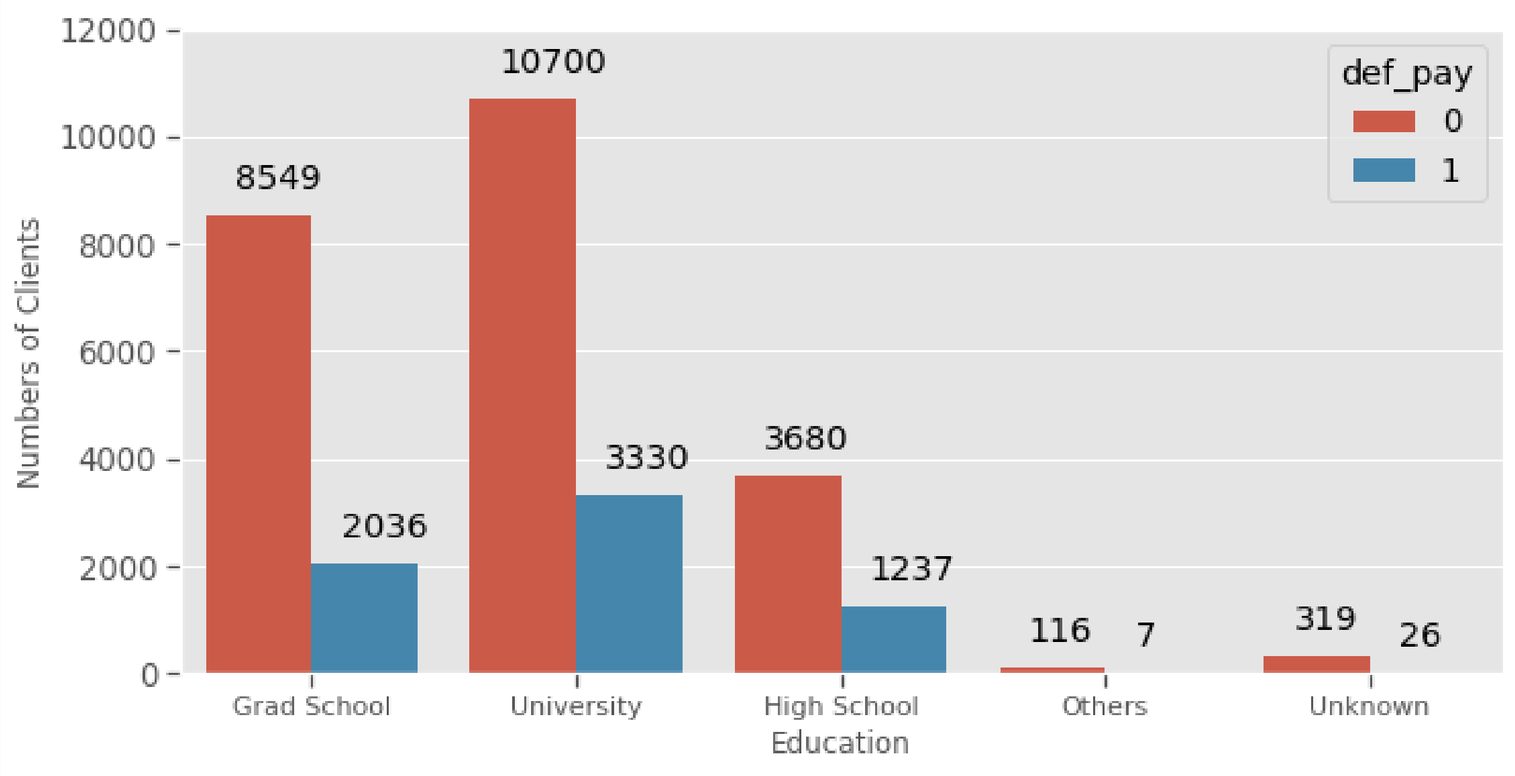


Analysis With Gender

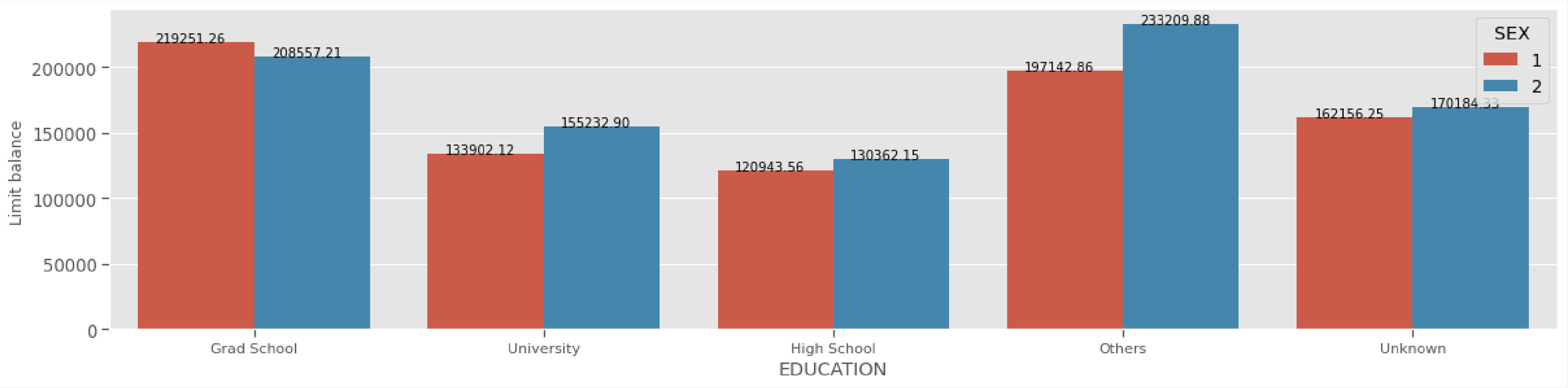


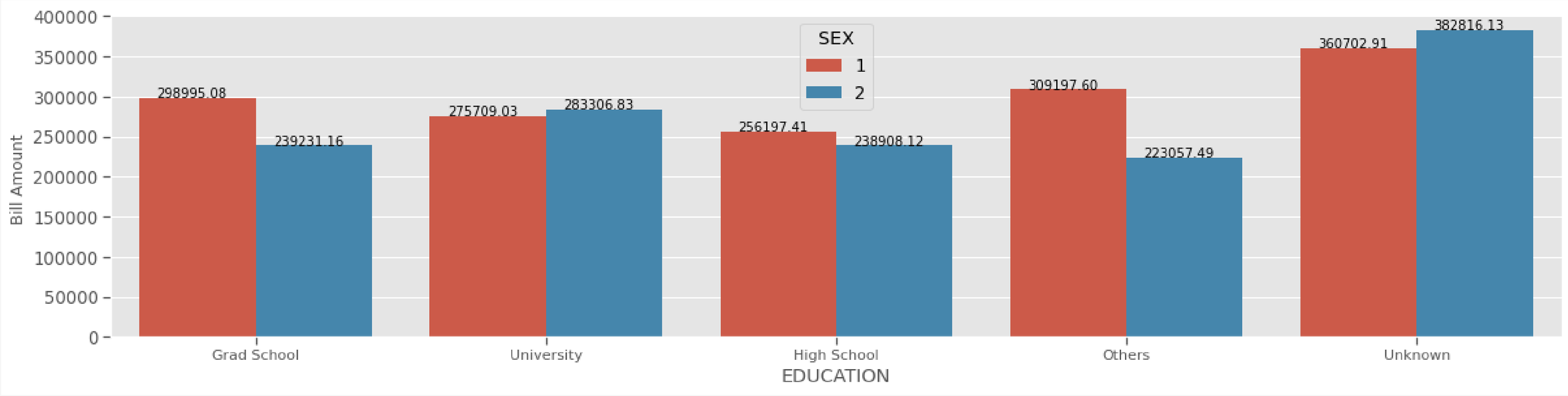
There are more women than men in our dataset and, apparently, men have a slightly higher chance of default. We will try to confirm it a little later by comparing men and women with the same education and marital status.

Analysis With Education

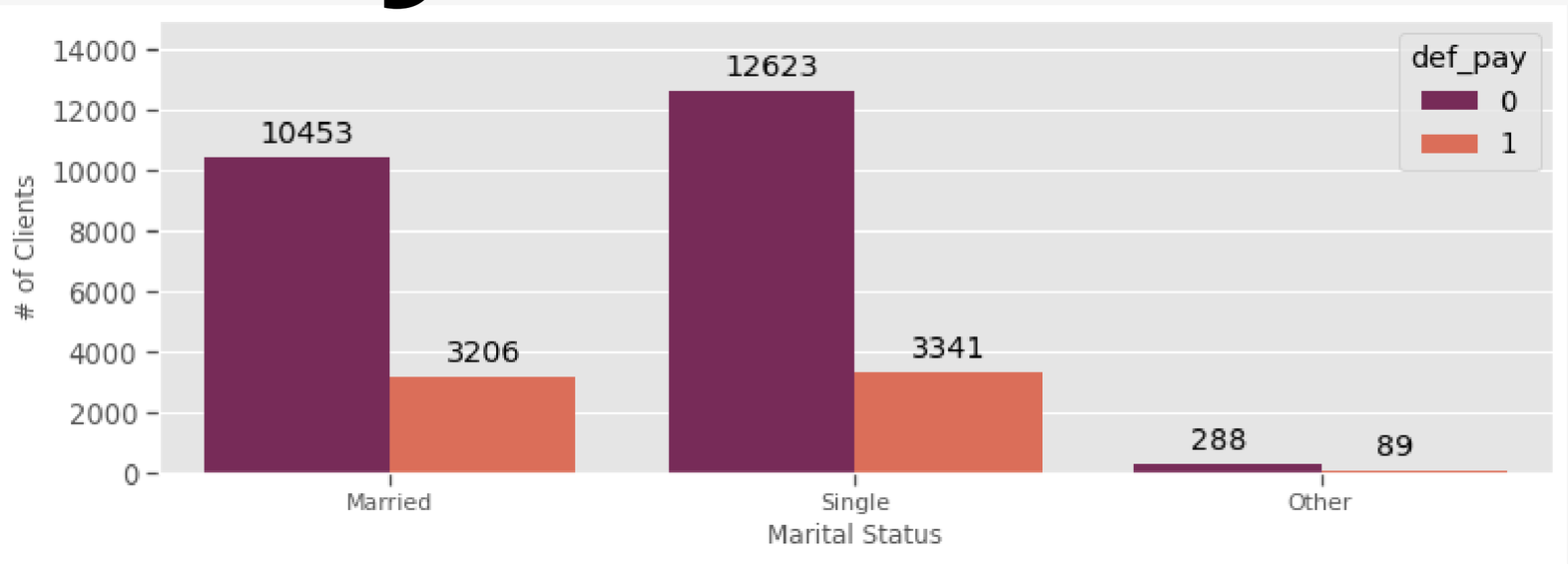


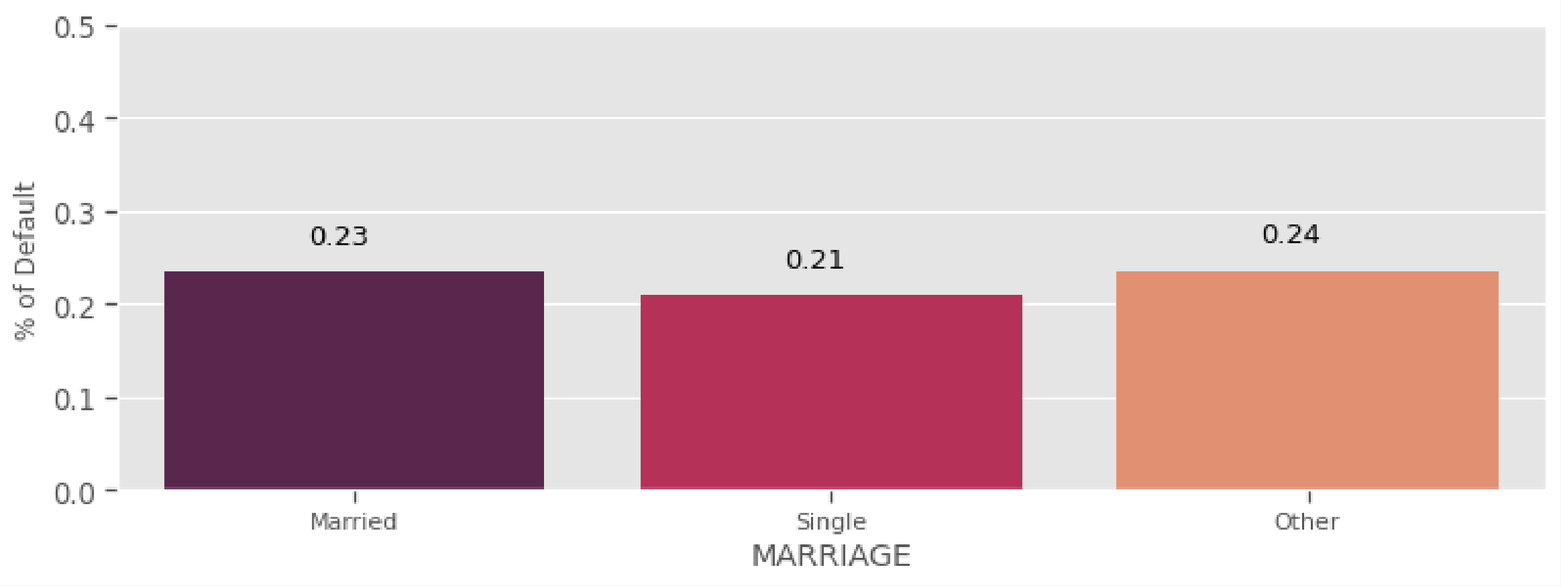
Analysis of Education with Limit and Bill(Sex)





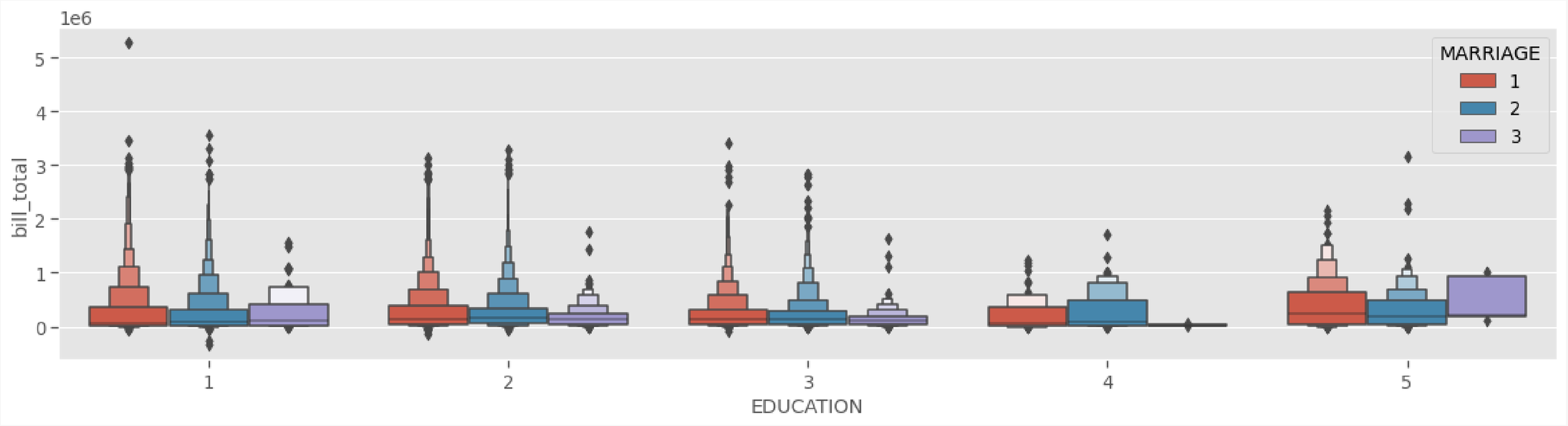
Analysis with Marital Status

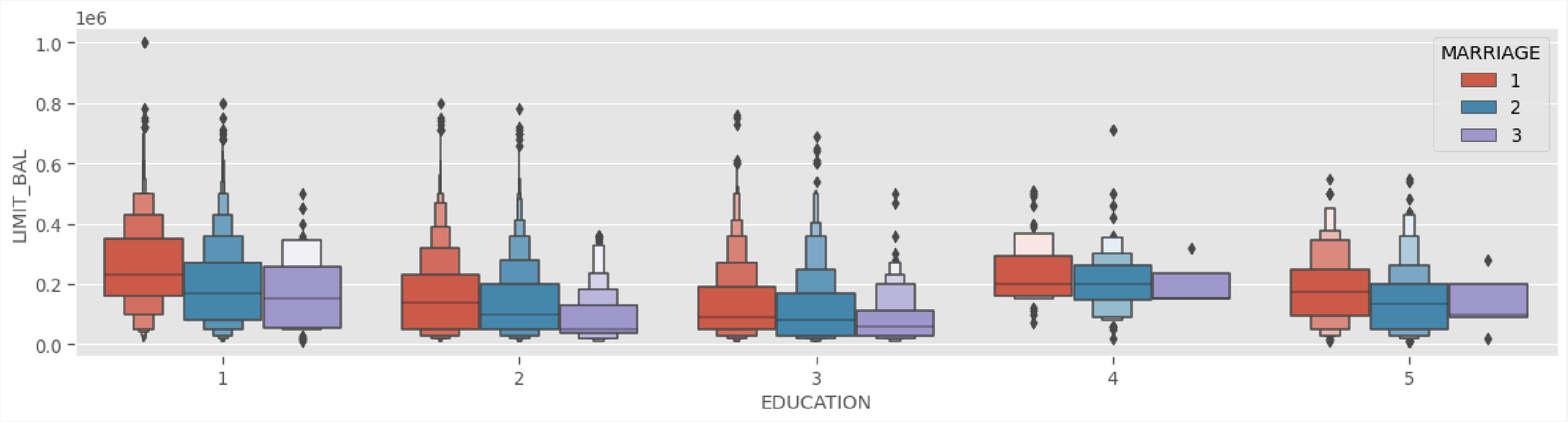




Here it seems that married people are most likely to default as well as the "Other" (which is again numerically less relevant than the others)

This analysis made it more concrete that this feature has less weight in our analysis

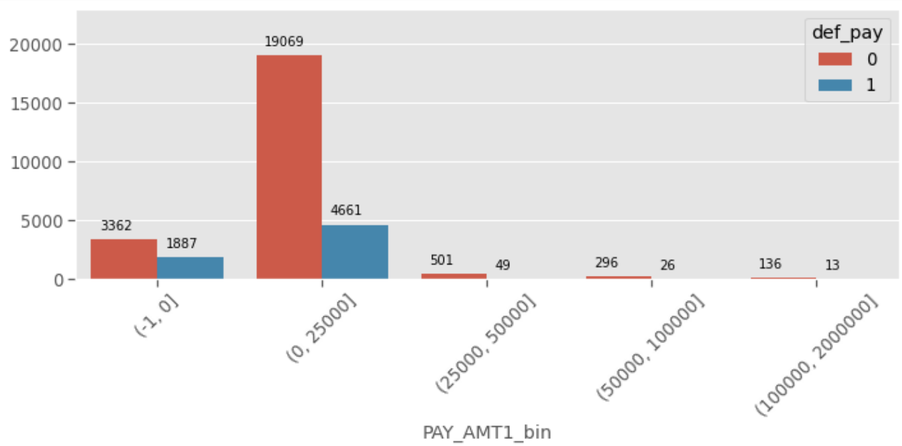


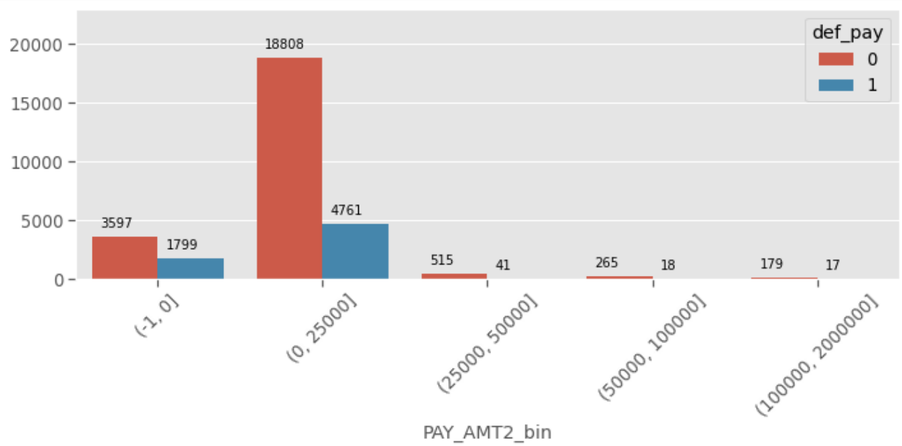


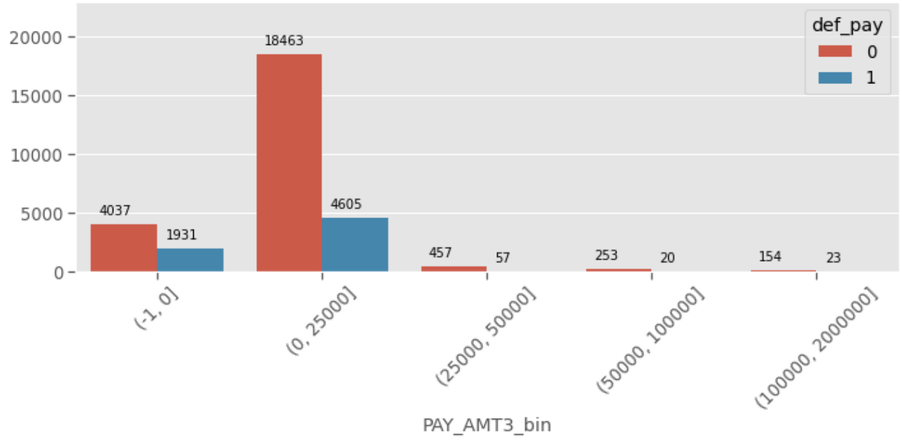
- General Trend states that As people ask for more months they are more likely to not be a defaulter after observing the 6 plots
- Most customers are duly paying their credit card bills. And it's pretty clear that their likelihood of default is much lower than the rest.

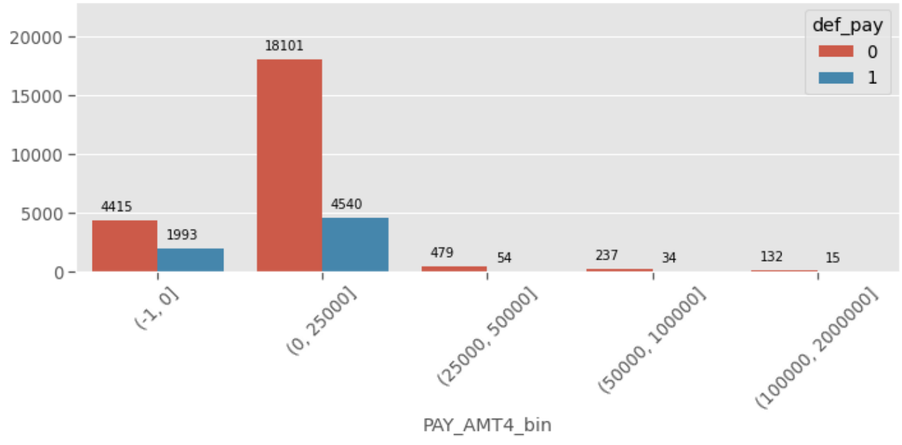
Since there are only a few customers with 4+ months of delay in all PAY_X features

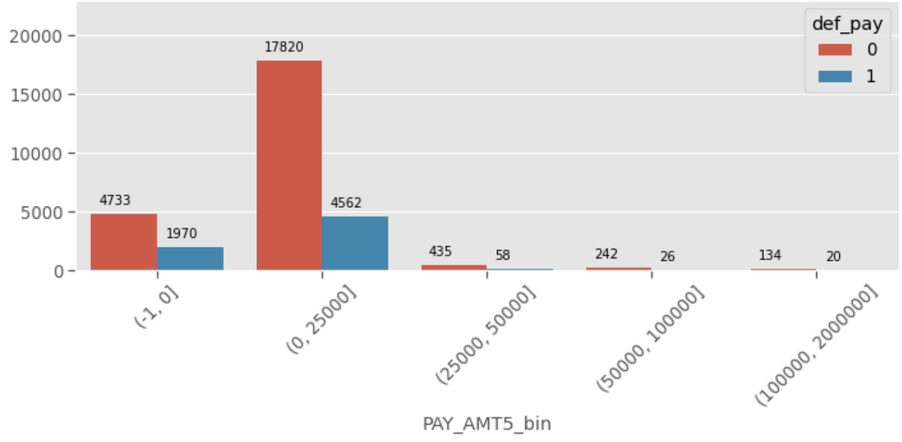
Amount of previous payment

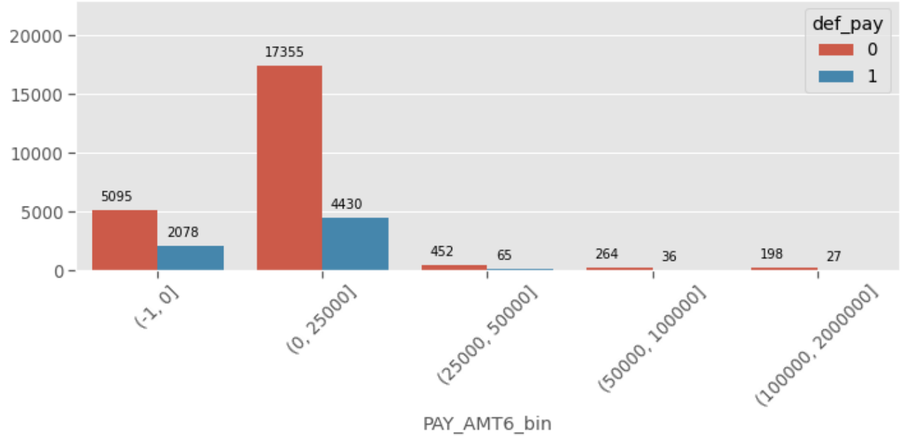


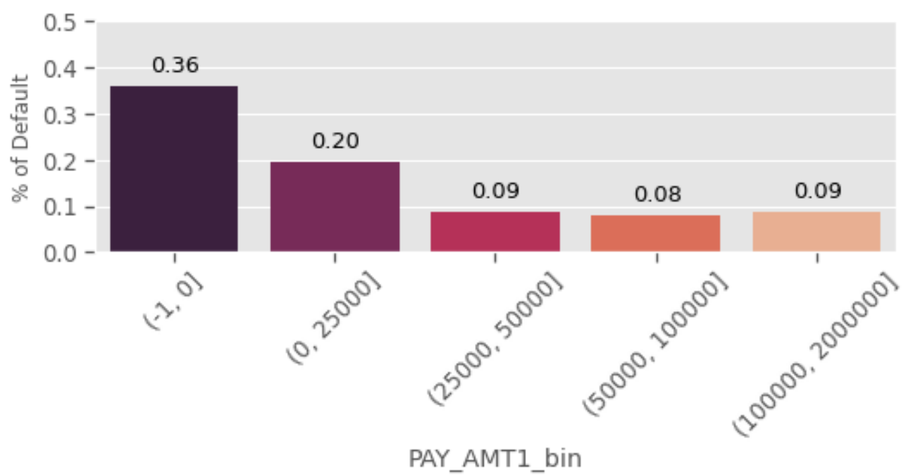


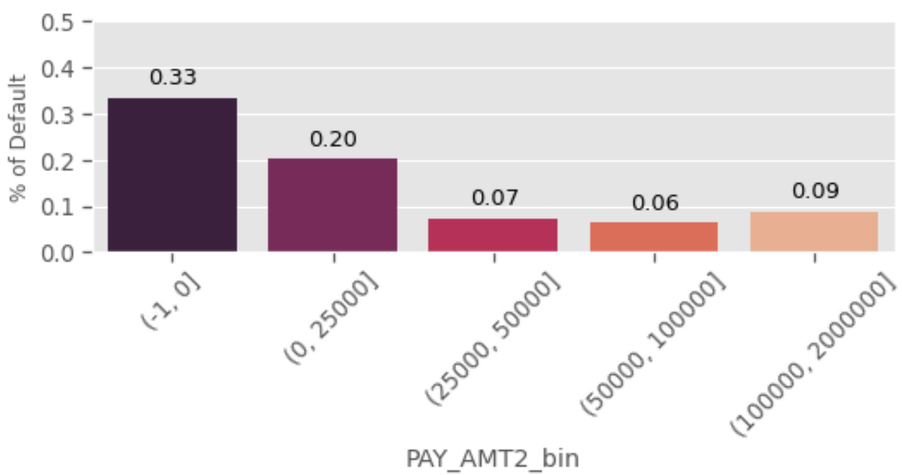


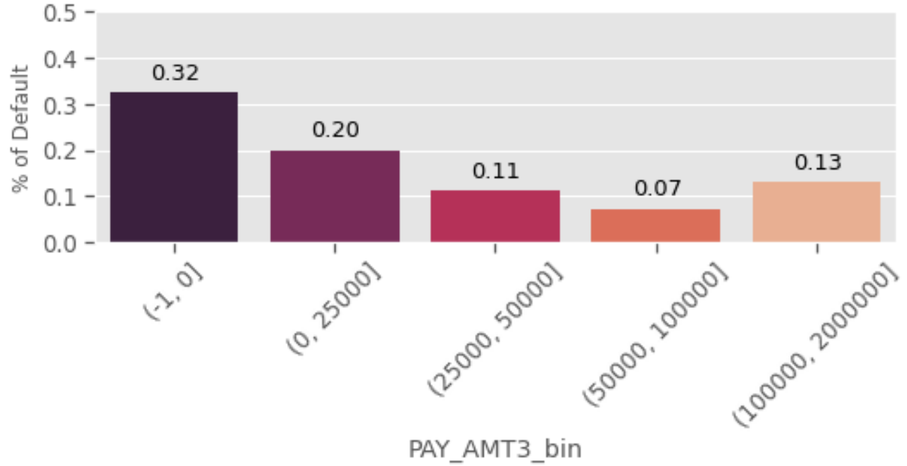


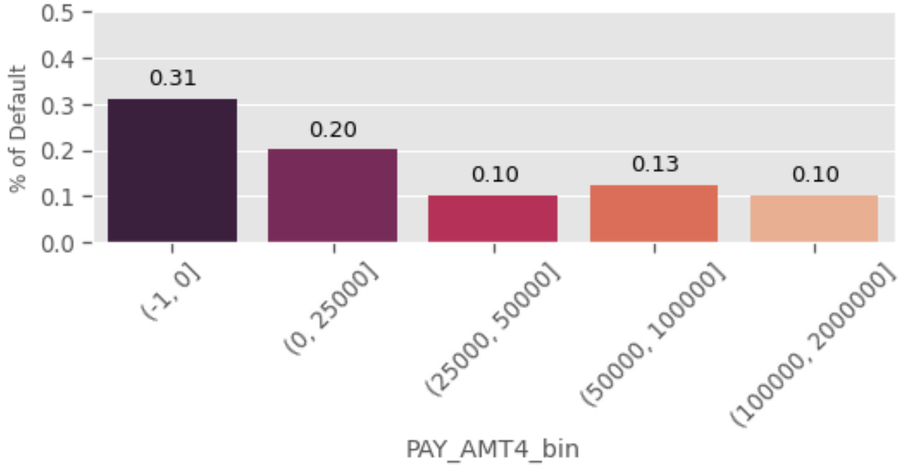


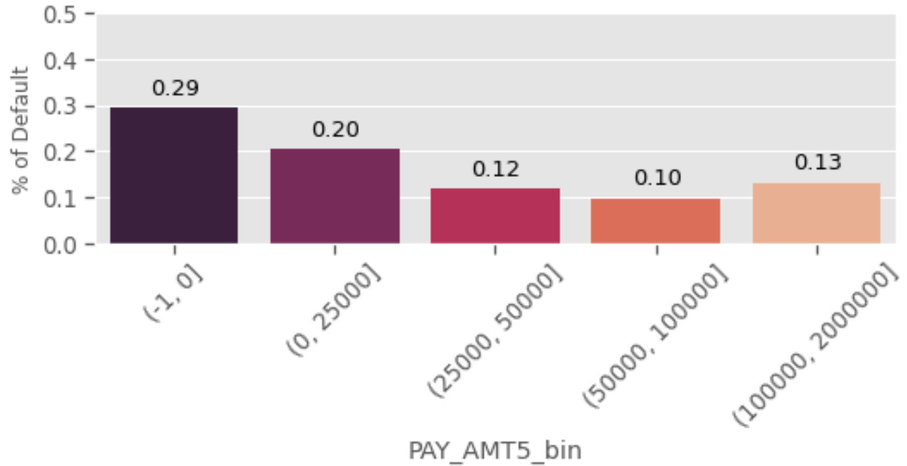


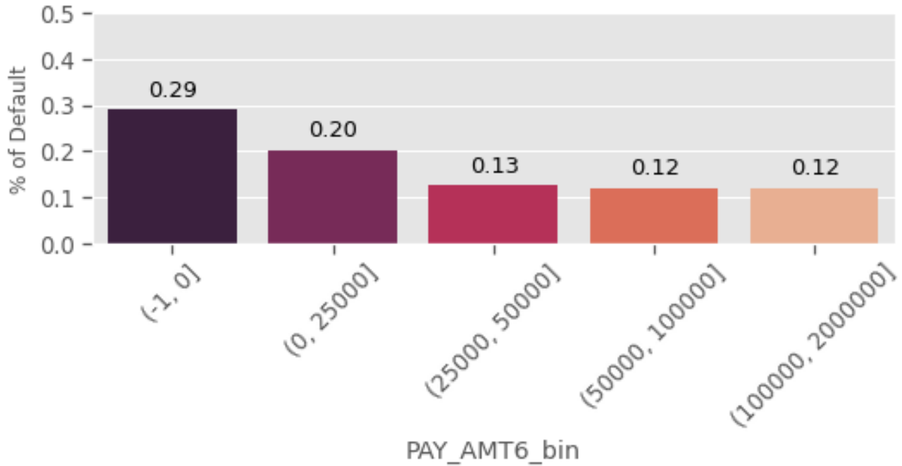




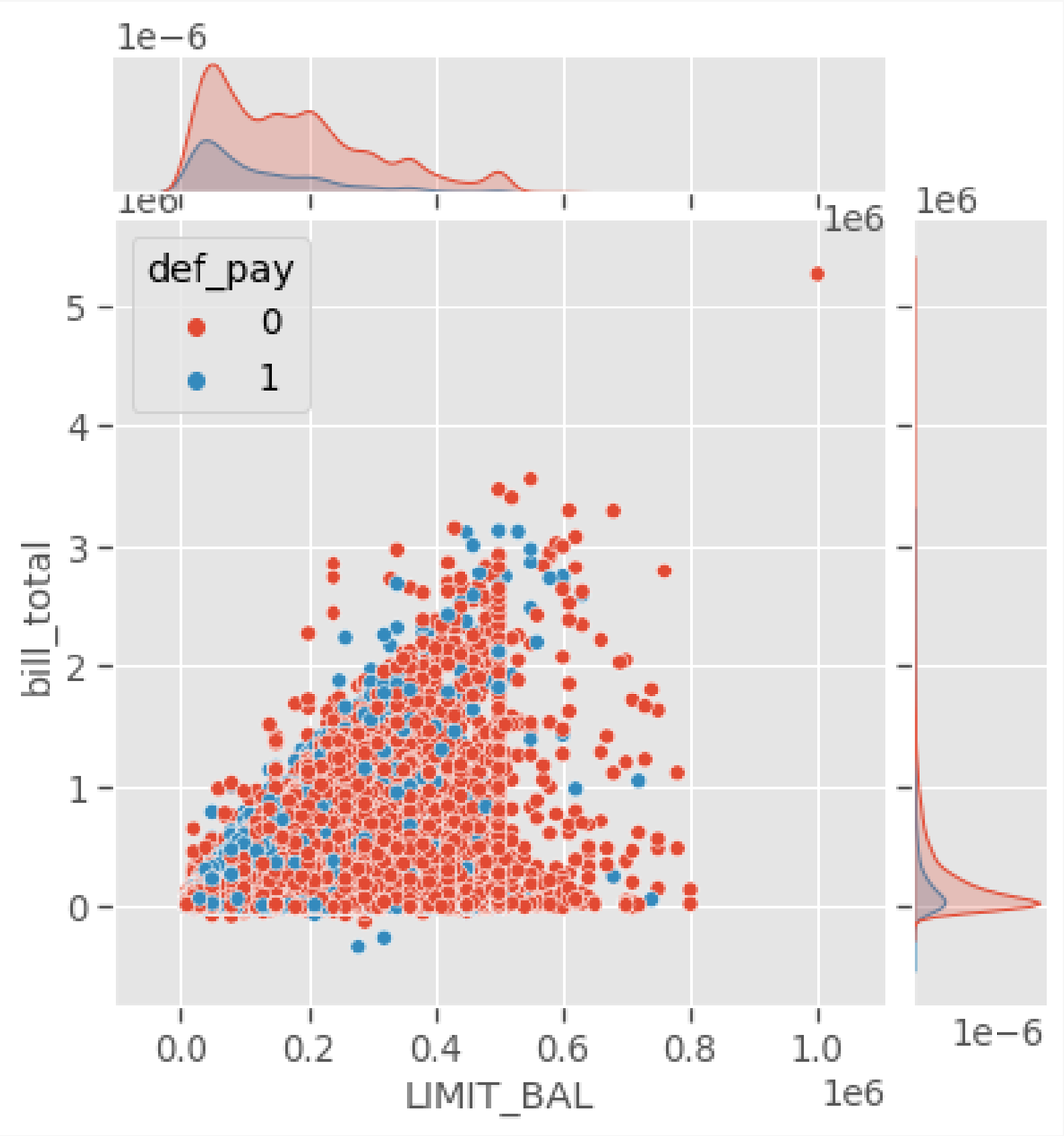






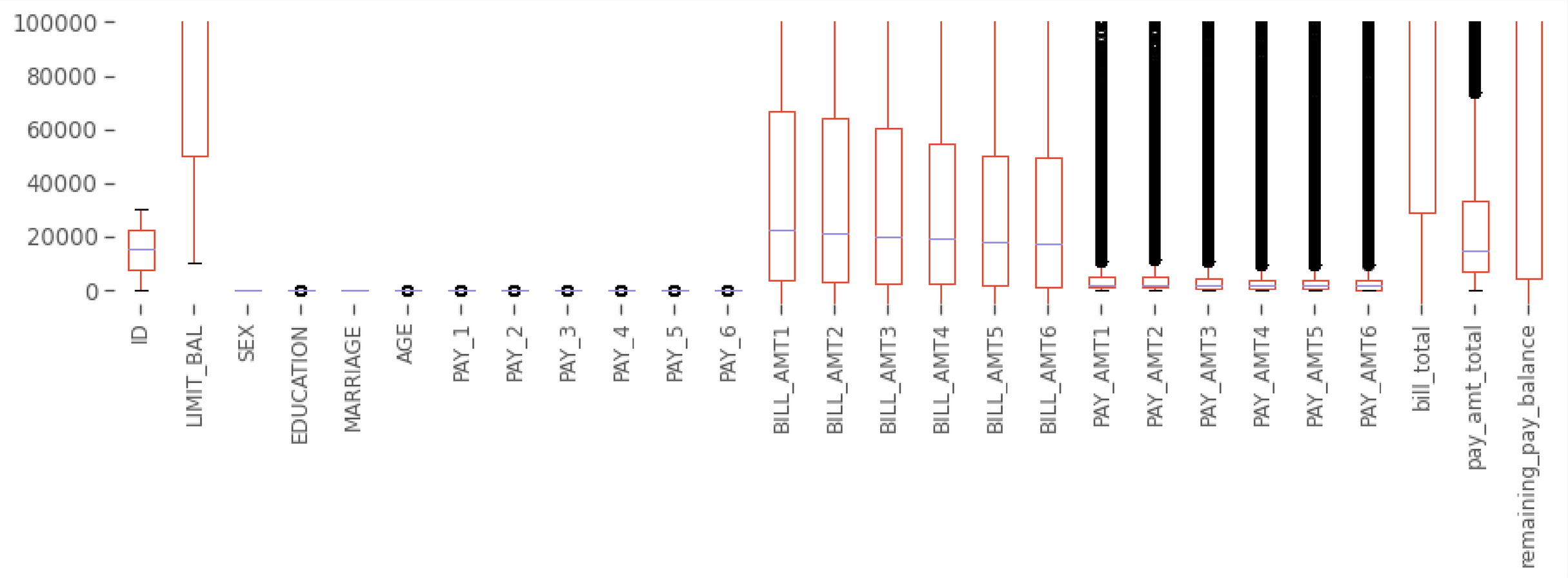


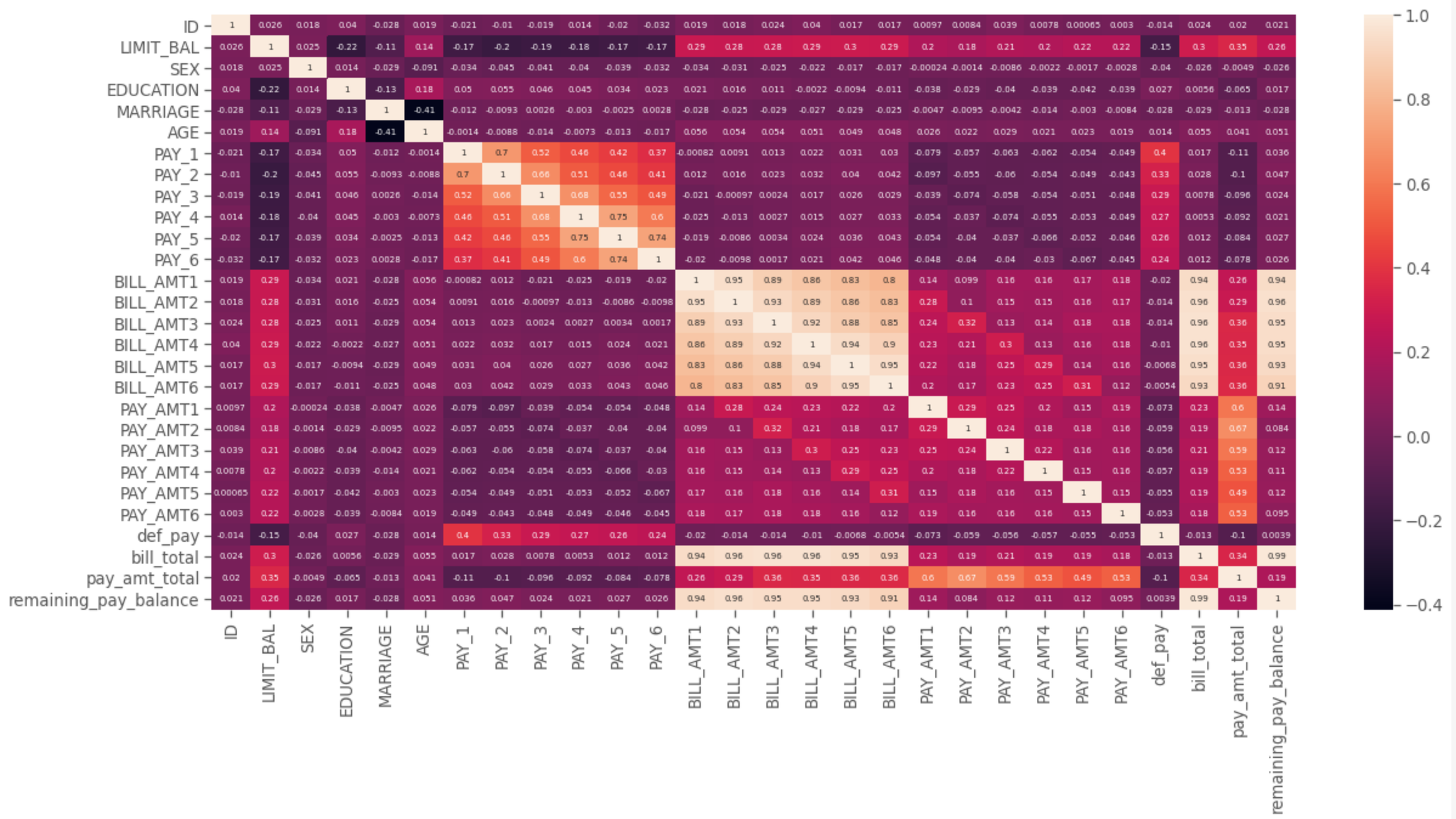
The analysis shows that there is a higher default rate among those who paid nothing in previous months and lower rates among those who paid over 25k of NT dollars.



- Limit Amount and Bill Amount both are right Skewed
- Bill Amount Suffers from Kurtosis
- Both have outliers that should be treated
- They should be Normalized so as to make our model more accurate for prediction

The boxplot below reveals that features are in different scales and units.
Hence they need to standerdized





Till Now we have done with Univarient and bivarient analysis Now looking at the corelation will give us a faire say On Features

Summary

- Overall, the offered credit limit seems to match the ability to pay in time within each category.
- The age group of 60+, which, despite having one of the highest chances of default, receives, on average, the highest credit value. by age group. (**Risky Group**)
- The analysis shows that there is a higher default rate among those who paid nothing in previous months and lower rates among those who paid over 25k of NT dollars.
- Higher the Limit Amount lesser are the chances of being a Defaulter And same goes with Our Total bill amount. (Keeping outliers in mind)

Analyzing the trends with age, education, marriage, credit limit

- The credit limit is the most important factor followed by the age
- The order of importance of other features is as followed Education, Sex, Marriage.
- With Bivariate analysis and Correlation also we can say that the History of past payment(PAY_N) Out of which also September plays a major role in our analysis as it gives a clear picture of Delay in the months for payment.