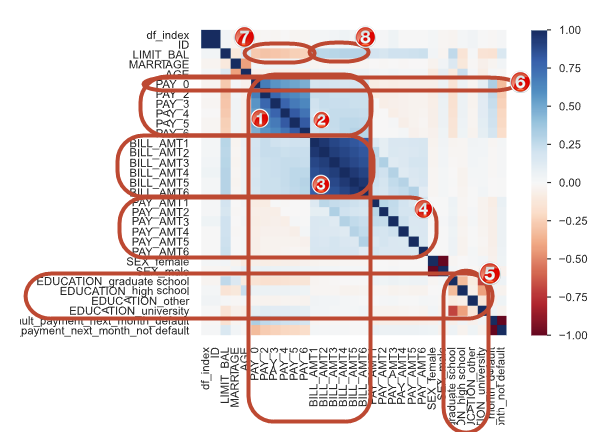
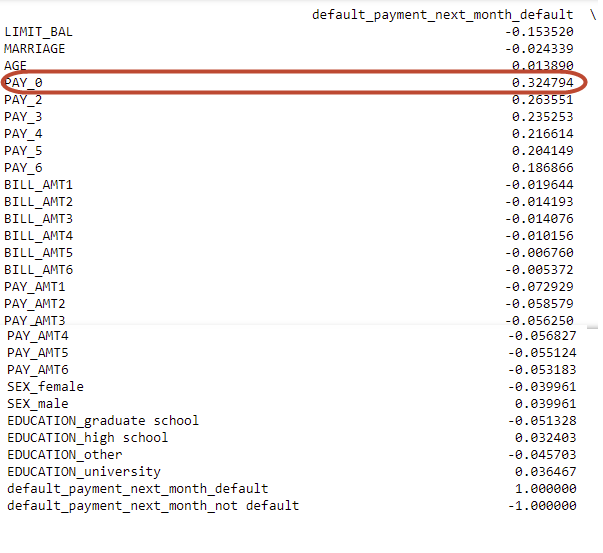
Guido,

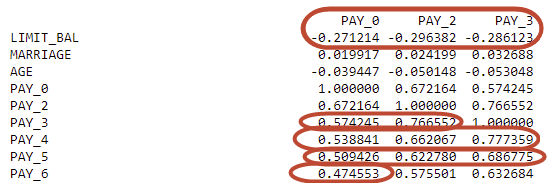
From reviewing the data, there are several items of interest I’ve found from exploring the data. We cannot with 100% certainty ensure that customers pay their loads. However, we can use correlation of several factors to help determine if customers are likely to pay their loans.

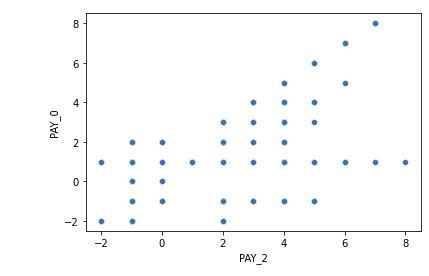
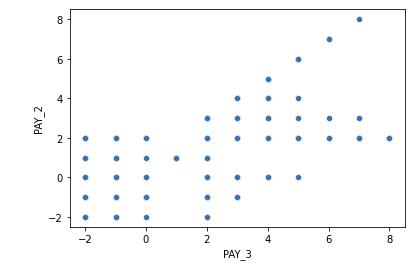
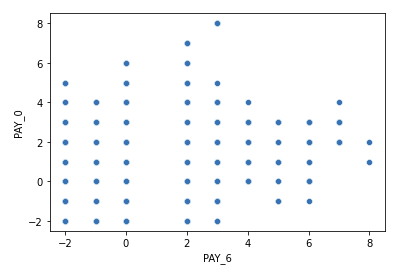
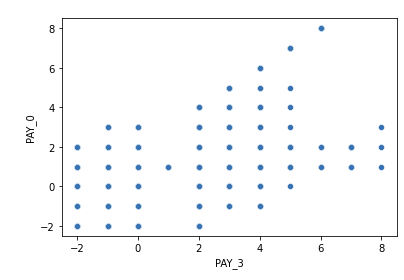
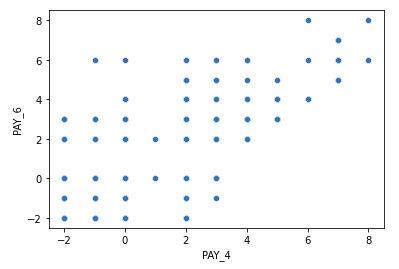
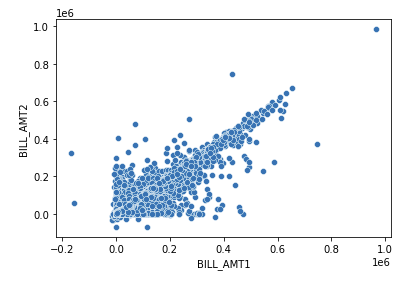
There was no correlation found between any values and age, limit balance, and marriage.

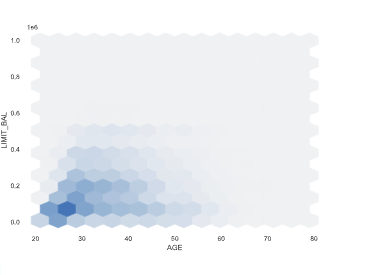
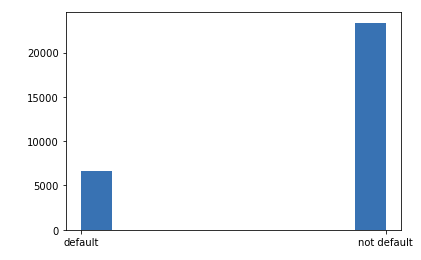
**Correlation Heatmap:**

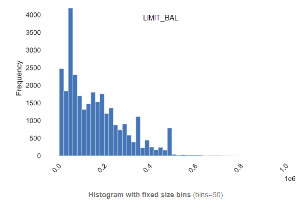
* 1. Strong positive correlation between all Pay variables (Pay\_0, Pay\_2, Pay\_3, Pay\_4, Pay\_5, and Pay\_6. Pay\_6 and Pay\_0 had the weakest
  2. A slightly positive correlation between all Pay and Bill\_Amt variables.
  3. Strong positive correlation between all BILL\_AMT variables; the strongest correlation was Bill\_AMT1 and BILL\_AMT2.
  4. Negligible-extremely small positive correlation between Pay\_AMT variables
  5. Strong negative correlation between university education and graduate school
     + \*Note: All Education variables have a bias because it’s counting each level as distinct rather than inclusive. Example: a person cannot attend university until they complete high school, etc.
  6. There was a small positive correlation between all Pay variables and the default payment variable. However, Pay\_0 had the highest correlation with default payment of all pay variables.
     + This means, if someone defaults on a payment, it’s most likely to be the first payment.
  7. Moderately negative correlation between Limit\_Bal and Pay\_0, Pay\_2, Pay\_3, Pay\_4,Pay\_5
  8. Moderately positive correlation between Limit\_Bal and all Bill\_Amt variables

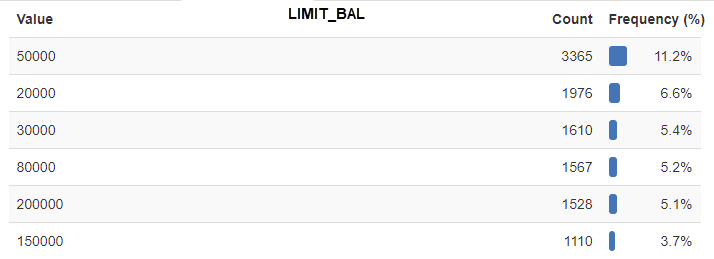






Younger customers receive a lower limit balance. Most customer behavior is to not default.

The limit balance with the most frequency is 50,000.



**Summary:**

We cannot with 100% certainty ensure that customers pay their loads. Overall, more people don’t default on their loans. There’s no correlation between gender or marital status as to if someone will default on a loan. There’s a stronger correlation to Pay\_0 and default payment than any other pay value. This means as pay progresses, it’s less likely to default.

There is a slightly positive or negative relationship between default or not default next month and Pay, Bill\_Amt, Pay\_Amt, the relationship gets a smaller correlation as the months progress. Example: if initially Bill\_Amt1 has a 0.019 correlation with default next month, the correlation gets smaller each month until it reaches 0.00537 by BILL\_AMT6.

In the future, we can use machine learning to use multiple variables to uncover if a customer is more or less likely to default because that will take into account multiple variables with the dataset.

Thanks, Kristen