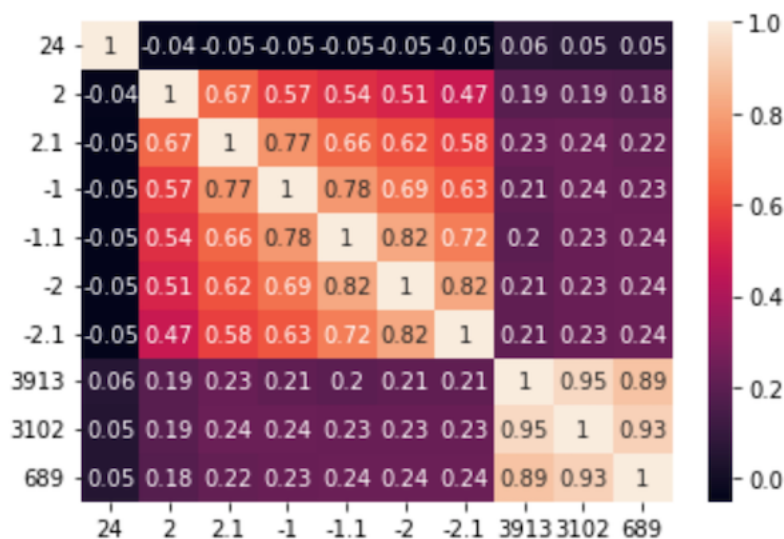


I think that the best way to ensure customers can and will pay their loans is to be very cautious when issuing them credit to begin with. It might be in Credit One's best interest to do further research into finding predictors that will provide them with more accurate results than the ones provided in this data set. Based on the data provided to us, I do not believe that it is possible to approve customers with very high certainty. I think that it could be possible to get a fairly decent certainty at some point, but the data provided to us is simply not going to get us there. Maybe if we include more suitable predictors, this might increase the certainty.

In data pre-processing, finding out if there's a high correlation between the predictor variables are important to avoid the multicollinearity. Multicollinearity is observed when two or more independent predictors are highly correlated to each other in a multiple regression model. Multicollinearity is a problem in regression model since it undermines the statistical significance of an independent predictor.



In the plot, we can see strong correlations with limit balance between several attributes, such as between PAY_4 and PAY_5. Highly correlated variables (>0.7) should be removed as predictors in the regression models.

As the credit limit is a continuous variable, so the appropriate method is regression and not classification. We can try some different regression methods to see if prediction accuracy is improved.