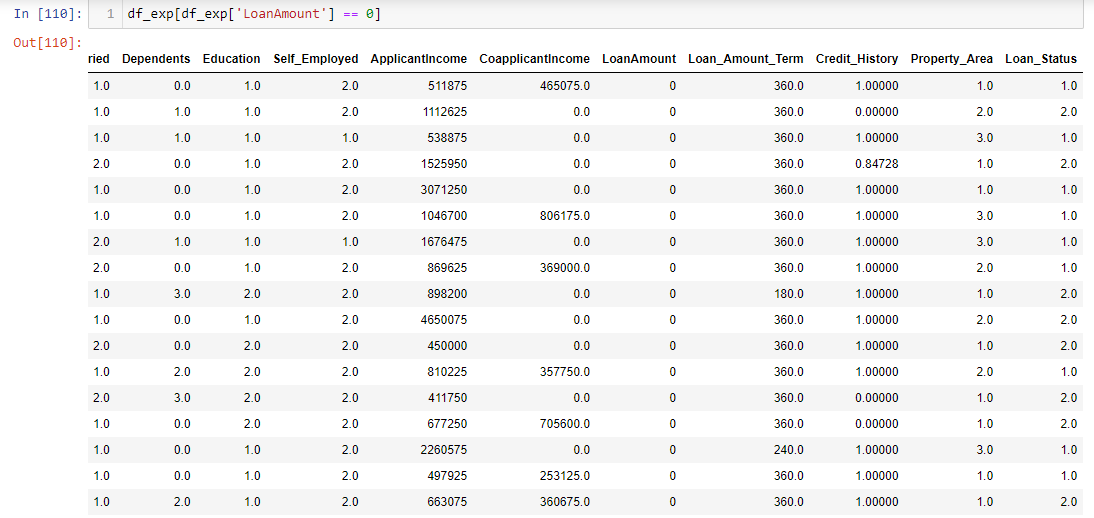
Milestone 3 (Customer Creditworthiness):

We’re starting with observing more on to the data and found that LoanAmount minimum value is 0.

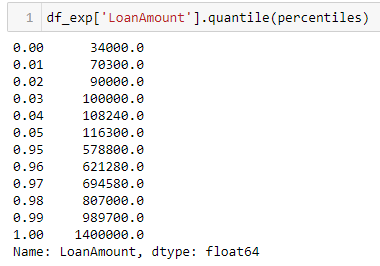


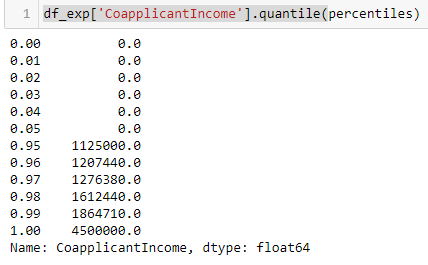
Which is not making sense that 1.0 (Yes) as Loan status where loan amount is zero. Deleting these rows as it seems faulty.

**Outliers check:**

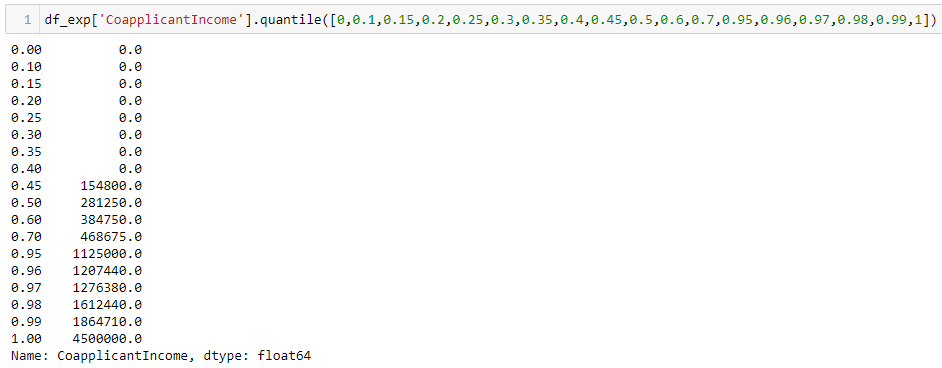
* **Percentile method.**

There are no outliers in any of data. Since now we know about the data whether particular number can be right or not as below.

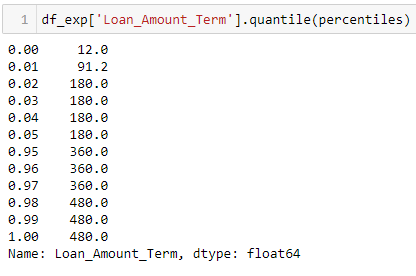


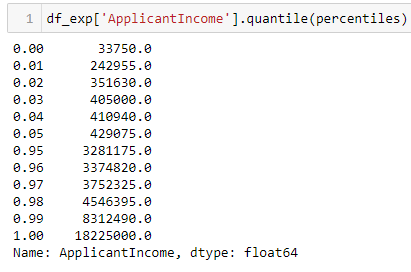


Here it was not clear, so we added more percentile.



Clearly visible that 45% of coapplicant aren’t earning.





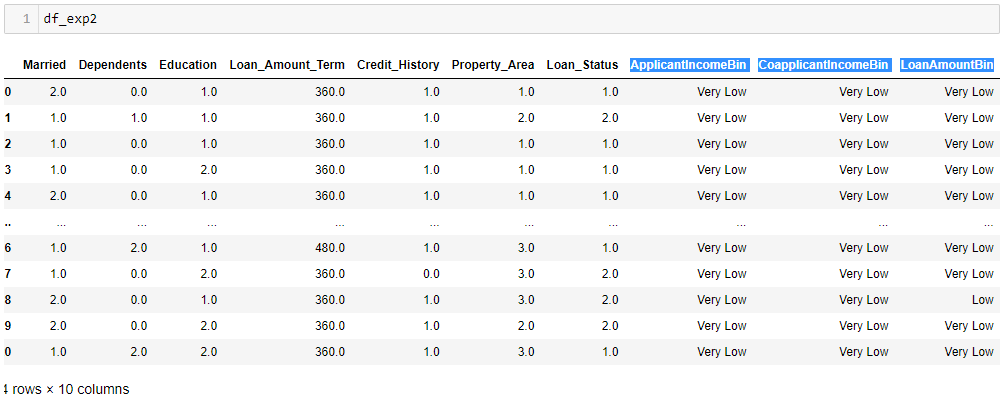
* **Finding outlier via IQR method**

Checked via IQR method as well as Z-score but having a domain knowledge makes us believe, they’re not outliers.

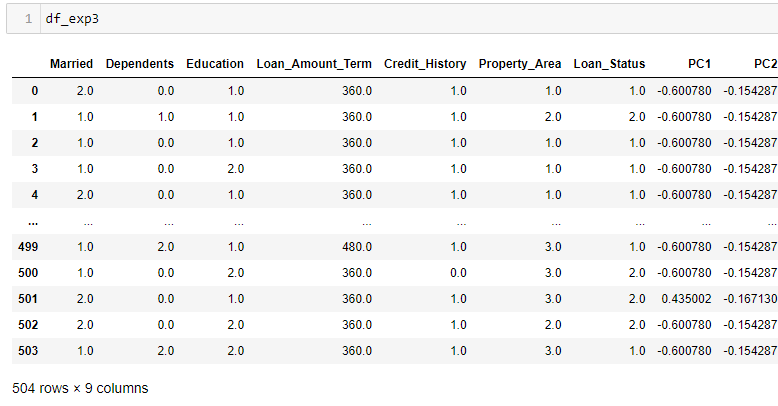
**Dropping feature**

As concluding in M2, we’re dropping Loan\_ID, Gender, Self\_Employed because they’re not valuable for our output.

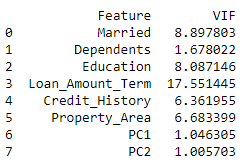
**Binning**



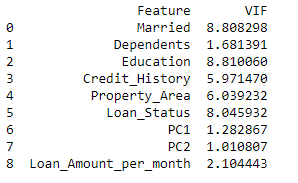
**Used PCA for dimension reduction and Scaling too. We’ve applied Principal Component Analysis for Applicant Income, Coapplicant Income and Loan Amount. We got the result as below.**



**VIF checking**

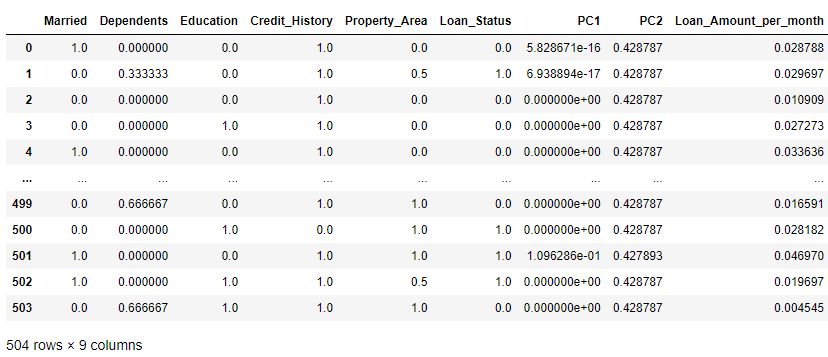


Since VIF more than 10 is big concern for us. Either we should drop to treat it. So we didn’t reduce more feature but created new as Loan\_amount\_per\_month (by dividing loan\_amount to Loan\_amount\_term) and replaced Loan\_amont\_term. Which helped us controlling VIF.



**Normalization**

We’ve applied this to make values between 0 and 1 so our model become more stable and gives accurate results.



**Balancing data**

We are not going to do Undersampling/ Oversampling/ SMOTE as our class distribution is 68-32. So, it is not highly imbalance. It is better that we make model on original data.