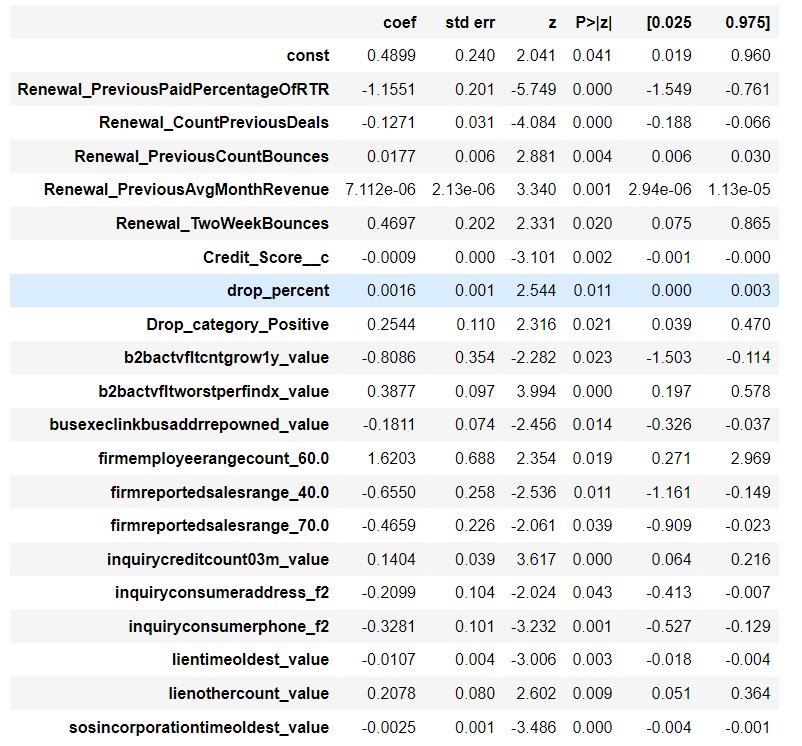
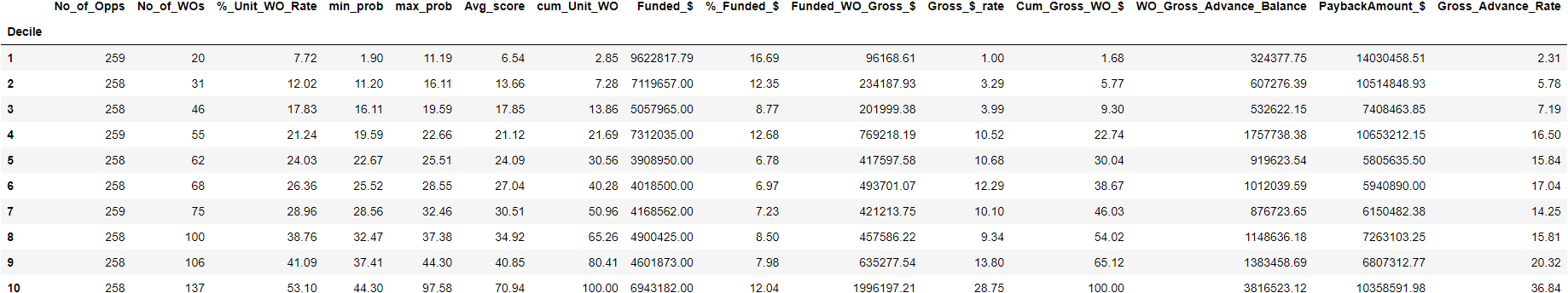
Summary 

Train set



Test set



Validation set



Adding new attribute in renewal only

drop\_percent\_recent\_revenue = (Ren\_previousmonthlyrevenue-last3statementsdeposits)/Ren\_previousmonthlyrevenue

**Summary:**

****

**Train set:**



**Test Set:**



**Validation Set:**



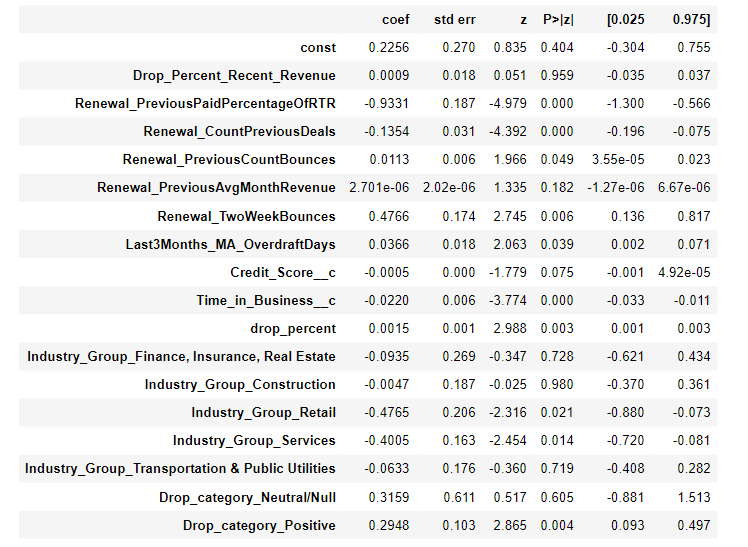
Observation (In Training Deciles) and Upsampling

* Till 9 th decile (where max\_probabilty - 0.42) - model has predicted = 0
* In 10 deciles (min\_prob - 0.42 till max probability) - model has predicted = 1.
  + False\_positive = where Write\_OFF = 0 and Model Predictions = 1
  + All those points are upsampled - so as to move them to upper deciles
* There were 43 such points which were added to the original train set of 2708 points

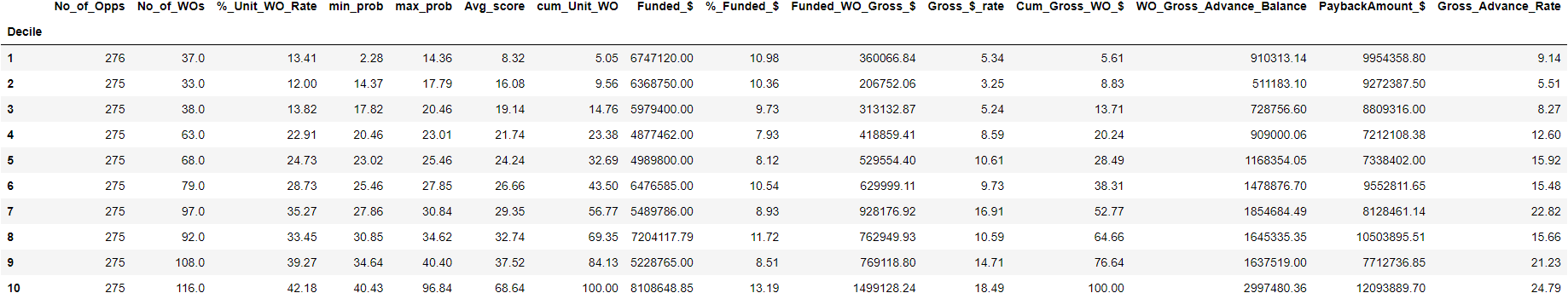
Upsampling



Model Summary



Train Decile



Test Decile



‘

Validation Decile



Advanced Feature Engineering

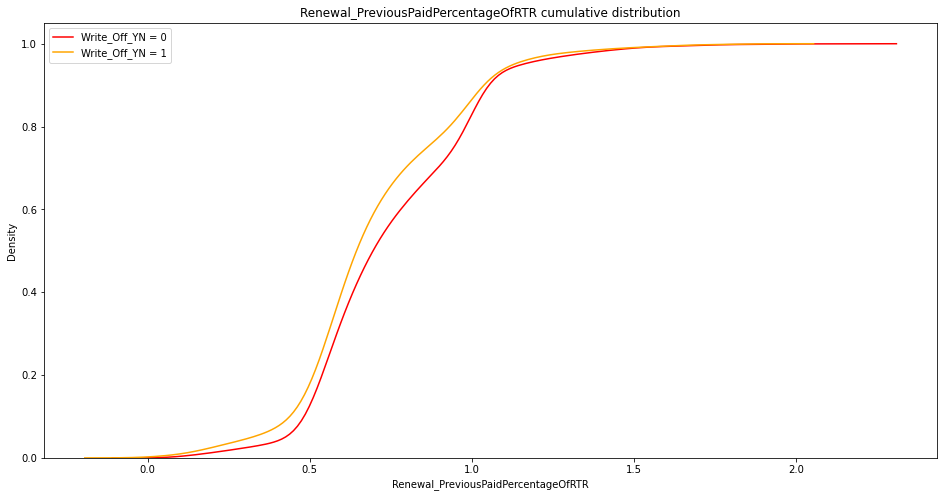
CDF, Distribution, Description, Bins

We need to paste model summary

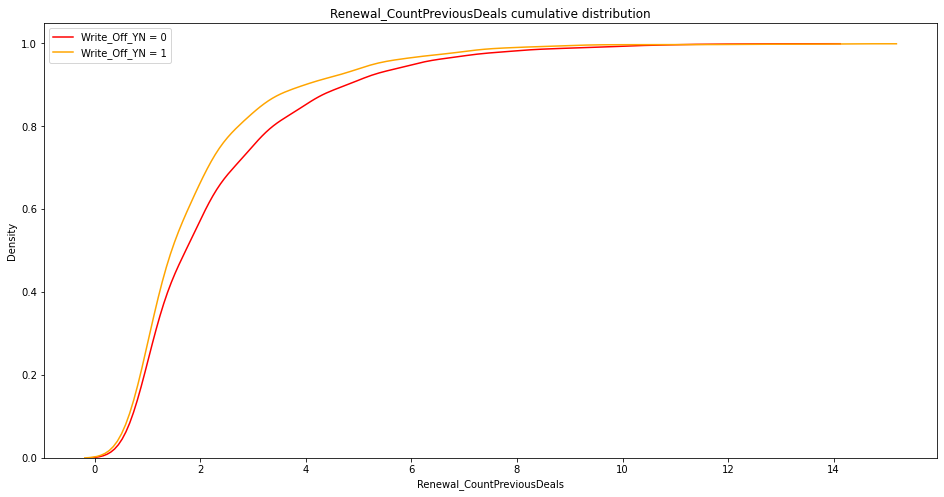
Explain why p value is 1.000 and why it is nan

Paste the results

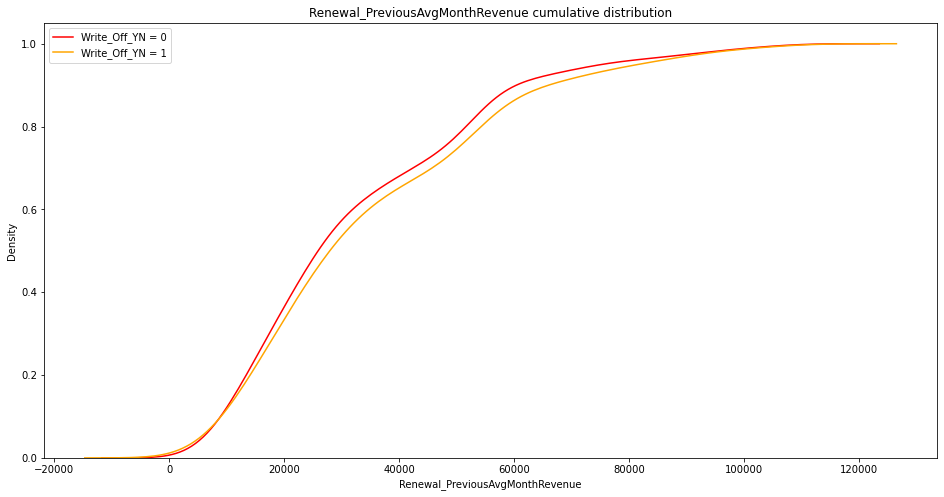
CDFs (These are build to get see the distinguish between output class in order to make bins):



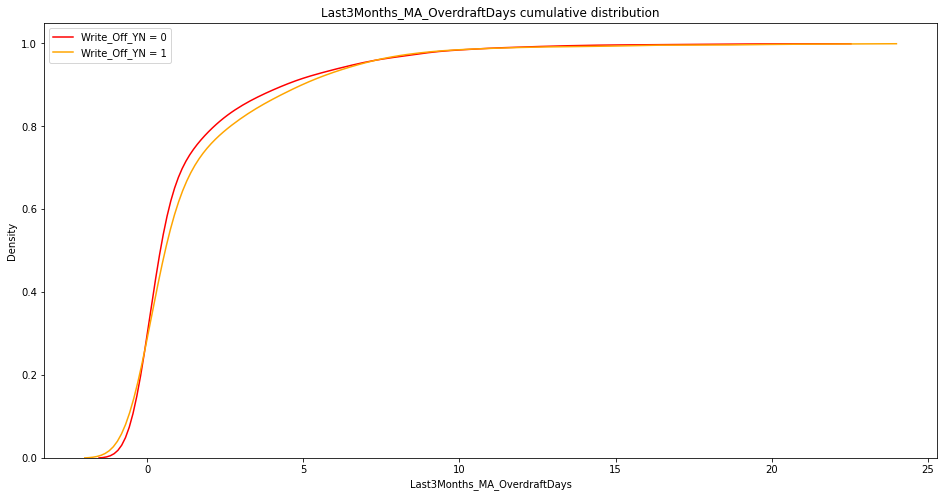
* Bins Distribution (0-0.5 (f1), 0.5-1(f2), Rest (f3))



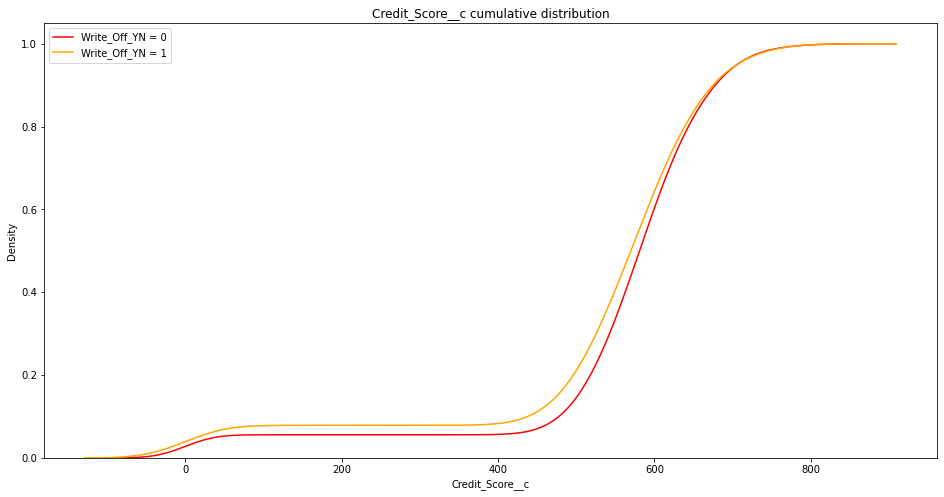
* Bins Distribution (2-4, 4-6, Rest)



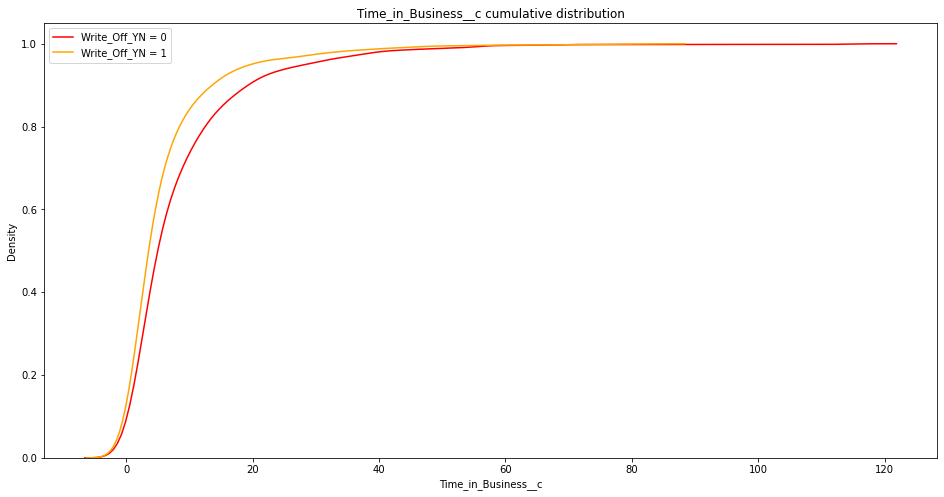
* Bins Distribution (20K-50K, 50K-80K, Rest)



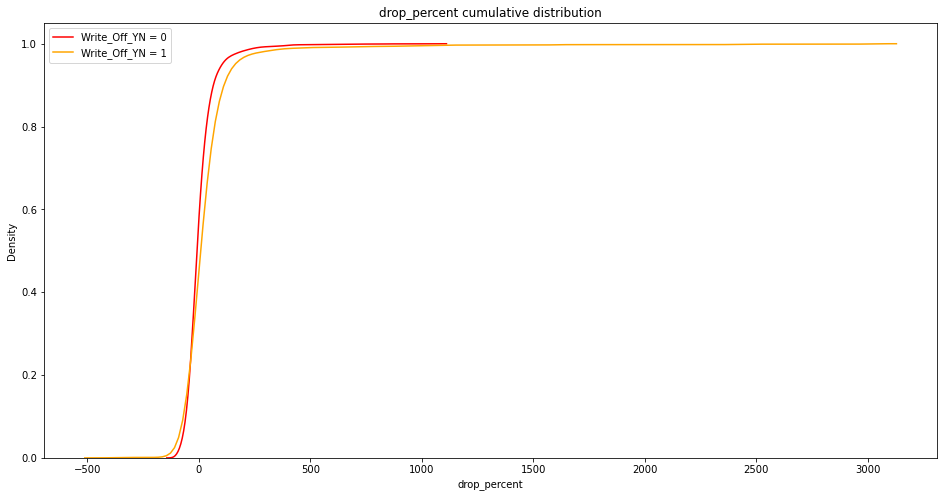
* Bins Distribution (0-5, Rest)



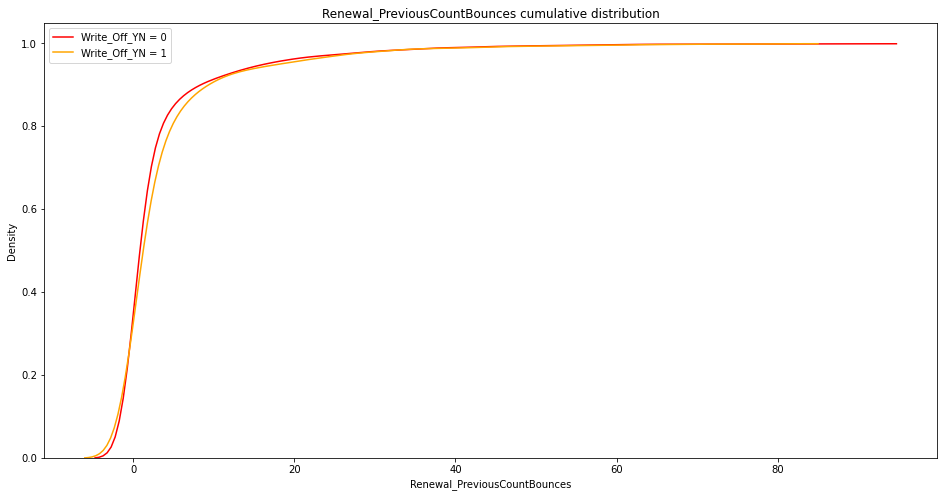
* Bins Distribution (400-500, 500-600, Rest)



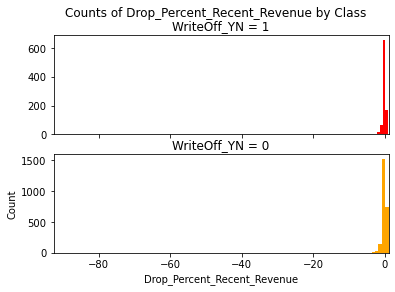
* Bins Distribution (0-20, 20-40, Rest)

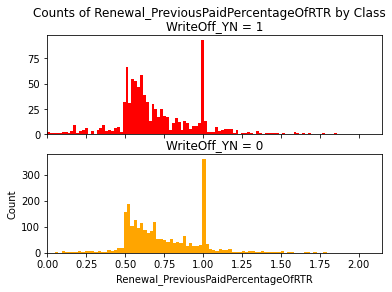
Some attributes could not be broken into bins n Others were binary features

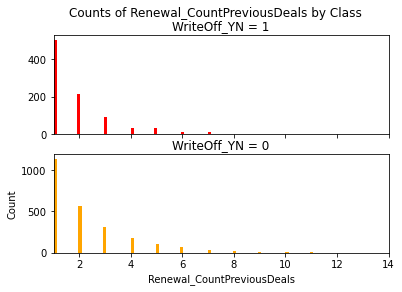




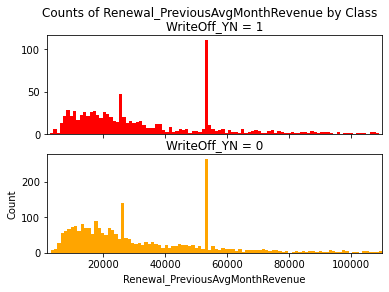
Distributions:-

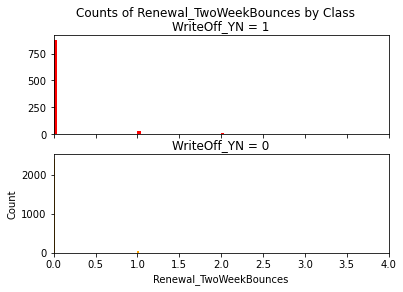


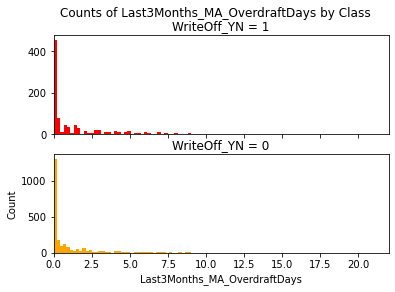


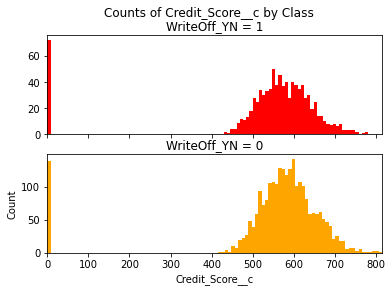


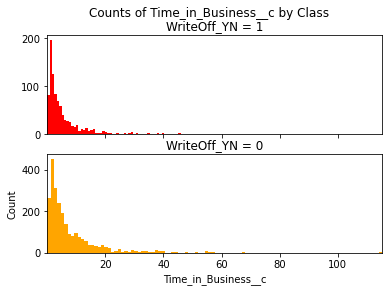


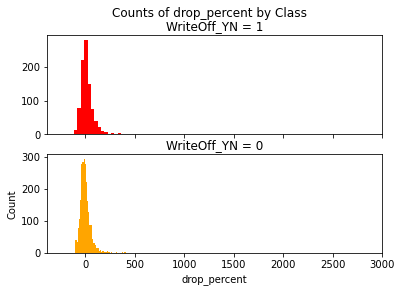


















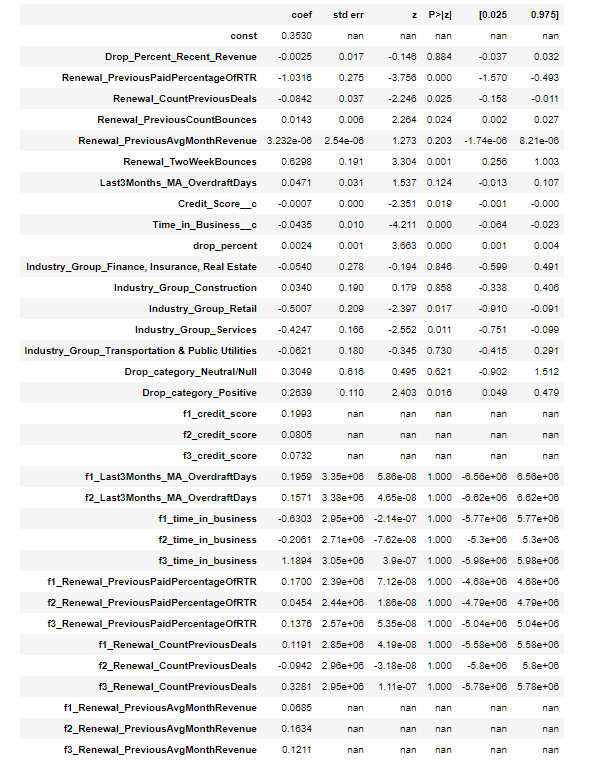








Summary:-



Train set:



Test set:



Validation Set:



Observations:

1. Validation and pentiles are not descent
2. P values of new attributes is either 1 or nan meaning they have no predictive value
3. It could be possible due to use of one hot encoding to make bins
4. One can not assign different weights (label encoding) to every sub-attribute. This can make our model biased for bigger values of weights

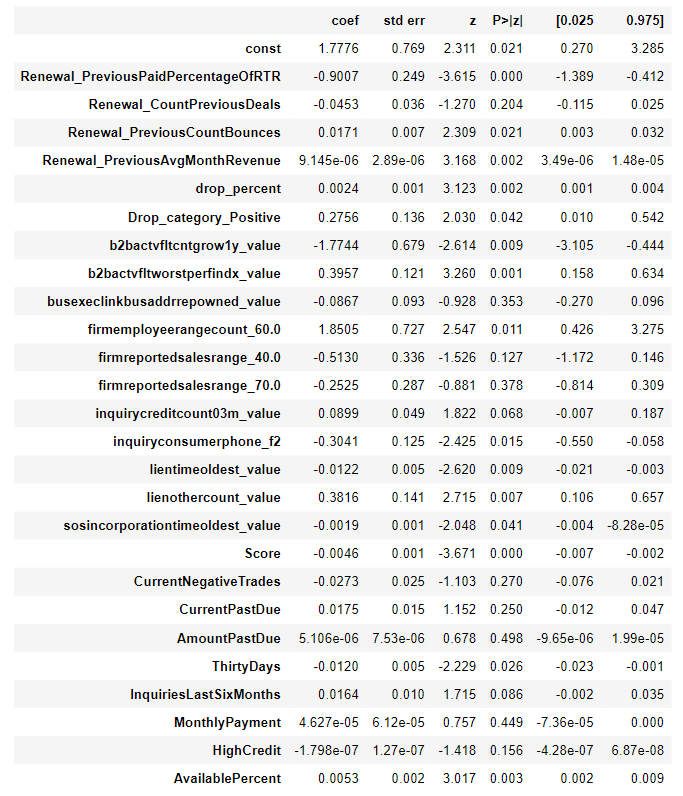
Renewal + LN + CBC:-

Training set - (1851,26)

Test set - (463,26)

Validation set - (326,26)

Summary:



Training Decile:-



Test Decile:-



Validation Decile:-

