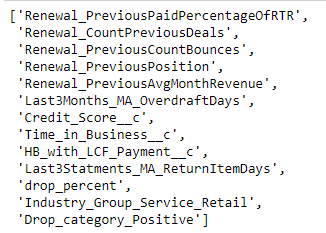
**Data - Distribution:**

* Training Data - (2708)
* Test Data (train-test split) - (678,\_)
* Validation Data (unseen data) - (582,\_)

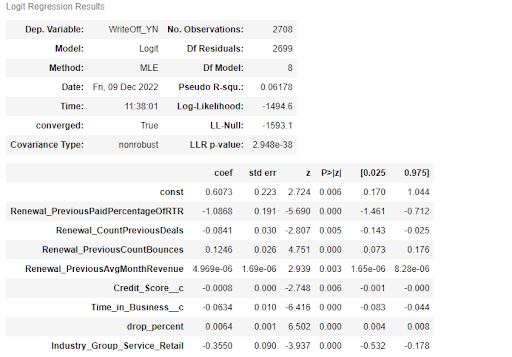
**Only Renewal Data:**

Attributes used for model training:

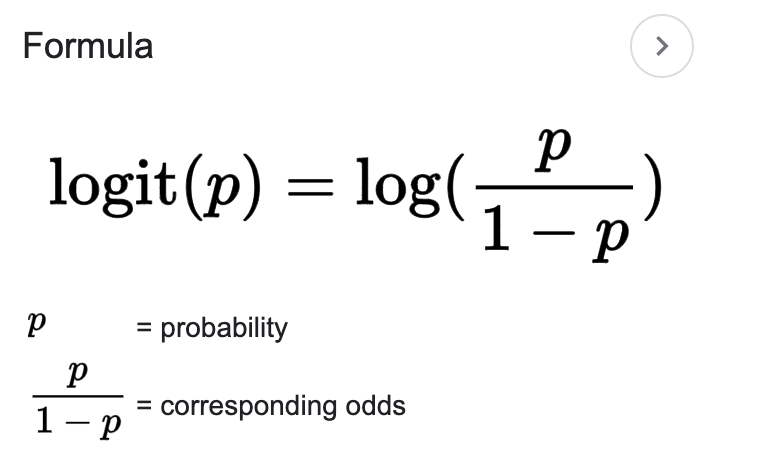


* These are the best attributes given to the model
* Other attributes are dropped due to high missing values, due to very high correlations, and categorical variables.

Model Summary:-



* Ideally
  + P < 0.05 (Statistically significant attribute)
  + P > 0.05 (Statistically insignificant attribute)
* Ideally
  + log\_likelihood - should be close to 0 (0 is the ideal model)
  + try to maximize it
* Ideally
  + Coefficients (b0, b1, b2 . . ) and Log\_Odds



= b\_0 + b1\*x1 + b2\*x2 + . . .

* + +ve coefficient - directly proportional to write\_off
    - If +ve ren\_countbounces - more chances of write\_off
  + -ve coefficient - inversely proportional to write\_off
    - If more credit\_score - less chances of write\_off

Train set:-



Test set (Split):-



Validation set:-



WHat terms should I study