The data is a structure dataset pull from the Kowope Mart’s database with few important features that can be used to predict defaulters in retail network. The data sample has been split into train and test with a volume of 56,000 and 24,000.

The files for download are:

* **train.csv** - you will use this to train your model.
* **test.csv** - you will use this to test your model.
* **SampleSubmission.csv** - is an example of what your submission file should look like. The order of the rows does not matter, but the names of the UIDs must be correct. You must predict growth\_stage for the corresponding images.

**Variables**

* Application\_ID: Unique Customer Application Identification number

***Risk Scoring features:***

* form\_field1: Customer Creditworthiness score based on historical data
* form\_field2: A score that measures the number and riskiness of credit enquiries made by a borrower.

***Severity features: Severity is a function of amount, time since default, and number of defaults***

* form\_field3: Severity of default by the borrower on any loan(s).
* form\_field4: Severity of default by the borrower on auto loan(s).
* form\_field5: Severity of default by the borrower on education loan(s).

***Credit features***

* form\_field6: Minimum of credit available on all credit cards that is automatically renewed as debts are paid off on the customer’s cards (in NGN)
* form\_field7: Maximum of credit available on customer’s active credit lines (in NGN)
* form\_field8: Maximum of credit available on all active credit cards that is automatically renewed as debts are paid off on the customer’s cards (in NGN)
* form\_field9: Sum of available credit on credit cards that the borrower has missed 1 payment (in NGN)
* form\_field10: Total amount of credit available on accepted credit lines (in NGN)
* form\_field11: The amount of dues collected post-default where the due amount was more than 500 (in NGN)
* form\_field12: Sum of the amount due on active credit cards (in NGN)
* form\_field13: Annual amount paid towards all credit cards during the previous year (in NGN)
* form\_field14: Annual income (in NGN)
* form\_field15: The estimated market value of a property owned/used by the borrower (in NGN)

***Credit card features:***

* form\_field16: Number of active credit card that is automatically renewed as debts are paid off on which full credit limit is utilized by the borrower
* form\_field17: Number of active credit cards on which full credit limit is utilized by the borrower
* form\_field18: Number of active credit lines on which full credit limit is utilized by the borrower
* form\_field19: Number of active credit cards on which at least 75% credit limit is utilized by the borrower
* form\_field20: Number of active credit lines on which at least 75% credit limit is utilized by the borrower
* form\_field21: Average utilization of active revolving credit card loans (%)
* form\_field22: Average utilization of line on all active credit lines activated in last 2 years (%)
* form\_field23: Average utilization of line on all active credit cards activated in last 1 year (%)
* form\_field24: Average utilization of line on credit cards on which the borrower has missed 1 payment during the last 6 months (%)
* form\_field25: Average tenure of active revolving credit cards (in days)
* form\_field26: Tenure of oldest credit card among all active credit cards (in days)
* form\_field27: Tenure of oldest revolving credit card among all active revolving credit cards (in days)

***Time-related features***

* form\_field28: Number of days since last missed payment on any credit line
* form\_field29: Tenure of the oldest credit line (in days)
* form\_field30: Maximum tenure on all auto loans (in days)
* form\_field31: Maximum tenure on all education loans (in days)
* form\_field32: Sum of tenures (in months) of active credit cards
* form\_field33: Sum of tenures (in months) of active credit cards
* form\_field34: Number of active credit lines over the last 6 months on which the borrower has missed 1 payment
* form\_field35: Number of revolving credit cards over the last 2 years on which the borrower has missed 1 payment
* form\_field36: Number of active credit lines
* form\_field37: Number of credit cards with an active tenure of at least 2 years
* form\_field38: Number of credit lines activated in the last 2 years
* form\_field39: Number of credit lines on which the borrower has current delinquency

***Percentage related features***

* form\_field40: Utilization of line on active education loans (%)
* form\_field41: Utilization of line on active auto loans (%)

***Other features***

* form\_field42: Financial stress index of the borrower. This index is a function of collection trades, bankruptcies files, tax liens invoked, etc.
* form\_field43: Number of credit lines on which the borrower has never missed a payment in the last 2 years, yet considered as high-risk loans based on the market prediction of the economic scenario
* form\_field44: Ratio of the maximum amount due on all active credit lines and the sum of amounts due on all active credit lines
* form\_field45: Number of mortgage loans on which the borrower has missed 2 payments
* form\_field46: Number of auto loans on which the borrower has missed 2 payments
* form\_field47: Type of product that the applicant applied for. (C = Charge; L = Lending)
* form\_field48: Undefined Variable
* form\_field49: Undefined Variable
* form\_field50: Ratio of the minimum amount due on all active credit lines and the sum of amounts due on all active credit lines

***Target features***

* default\_status: defaulted or not. (yes:1, no: 0)