## **Assignment Tasks:**

Please prepare a presentation (and csv for the second task) that includes the answers to the tasks described below. You’ll ﬁnd more information for the tasks in the data set folder shared with you on email.

1. Using the attached datasets (train\_set) analyse the data and visualize the most important aspects using your preferred method. Python libraries are preferred. Furthermore, share three ideas on how to reduce the % of default. **Default loans are unpaid loans that have exceeded 45 days since disbursement date**. Note this data was extracted on 7th January 2022, document your steps where needed.

2. Predict whether a loan is going to be defaulted or repaid on time. Use train\_set to train your model and test\_set to predict the target. Please document your steps and method used. The csv “SubmissionExample” will help with the format.

# **About the Data sets provided**

The data sets provided include loans data(loans\_data.csv), merchant transaction data(transaction\_data.csv) and merchant KYC data(kyc\_data.xlsx). The id column in transaction\_data.csv corresponds to merchant\_score\_reason\_id in the loans\_data.csv as should be used in merging the two datasets. The challenge is to predict the outcome of a loan as either default(bad) or paid(good) using the respective datasets to train and test your models.

**NB**: candidates will have to engineer the target variable(default non default) from the other variables and information provided.

**1 = Defaulted customers**

**0 = Non default customers**

DATA DICTIONARY

|  |  |
| --- | --- |
| **Loan\_data attributes** | |
| **attribute** | **description** |
| Id | unique identifier of loan data |
| merchant\_score\_reason\_id | Unique identifier of merchant score reasons |
| merchant\_id | unique identifier of merchant for the partner |
| partner | The company that partners with Zion so that Zion can offer loan facilities through the partner company e.g. Jumia, Twiga Foods |
| loan\_id | Unique loan identifier |
| app\_amount | Principal loan offer amount |
| app\_interest\_rate | loan interest rate |
| mpesa\_time | loan disbursement time through Mpesa |
| duration | loan tenor or number of days it will take to repay a loan |
| loan\_status | loan progress status marker e.g funding, funded, paid, processing, cancelled |
| created\_at | Date when loan was created |
| customer\_id | Zion customer unique identifer |
| loan\_count | Total number of loans a customer has borrowed from Zion |
| npa\_closed\_accounts | Count of closed non-performing accounts |
| pa\_open\_accounts\_with\_dh | Count of open performing accounts with default history |
| pa\_closed\_mobi\_loan\_accounts\_with\_dh | Count of closed performing mobile loan accounts with default history |
| npa\_open\_accounts | Count of open non-performing accounts |
| pa\_closed\_accounts | Count of closed performing accounts |
| pa\_accounts\_with\_dh | Count of performing accounts with default history |
| npa\_open\_mobi\_loan\_accounts | Count of open non-performing mobile loan accounts |
| last\_mobi\_loan\_listing\_date\_my\_sector | Date of last mobile loan listing |
| pa\_accounts | Count of performing accounts |
| pa\_closed\_mobi\_loan\_accounts | Count of closed performing mobile loan accounts |
| last\_mobi\_loan\_principal\_amount | last mobile loan principal amount |
| mobi\_loan\_accounts | Count mobile loan accounts |
| credit\_history | Length of credit history in months |
| min\_mobi\_loan\_principal\_amount | minimum mobile loan principal amount |
| avg\_mobi\_loan\_principal\_amount | average mobile loan principal amount |
| last\_mobi\_loan\_listing\_date\_other\_sectors | Date of last mobile loan listing |
| max\_mobi\_loan\_principal\_amount | maximum mobile loan principal amount |
| npa\_closed\_mobi\_loan\_accounts | Count of closed non-performing mobile loan accounts |
| pa\_open\_accounts | Count of open performing accounts |
| npa\_accounts | Count of non-performing accounts |
| pa\_open\_mobi\_loan\_accounts\_with\_dh | Count of open performing mobile loan accounts with default history |
| pa\_closed\_accounts\_with\_dh | Count of closed performing accounts with default history |
| updated\_at | Last loan update date |

|  |  |
| --- | --- |
| **Transaction\_data attributes** | |
| **Attribute** | **description** |
| Id | Unique identifier of transaction data |
| master\_record\_id | Unique identifier of master record |
| merchant\_id | Unique identifier of merchant for partner |
| customer\_id | Unique identifier of merchant for partner. Similar to merchant\_id |
| transaction\_days | Count of days a merchant has been transacting with the partner |
| num\_orders | Count of orders or transactions |
| most\_recent\_transaction | Last transaction date |
| earliest\_transaction | First transaction date |
| active\_days | Count of days a merchant has been transacting with the partner. Has to be at least 1 day. Equivalent to (most\_recent\_transaction - earliest\_transaction)+1 days |
| total\_sales | Total amount that a merchant has transacted with the partner over a period of time. |
| unique\_days\_sales | Number of unique days that a merchant actually transacted |
| sales\_activity\_ratio | proportion of unique days a merchant has transacted out of total days a merchant has been with the partner. Equivalent to unique\_days\_sales/active\_days |
| expected\_orders(fundable\_merchants) | Count of orders that a merchant who qualifies for a loan is expected to make. |
| avg\_order\_size | Average amount of order. Equals total\_sales/num\_orders |
| depot\_name | depot or location name of the merchant |
| hypothetical\_weekly\_order | Average number of orders per week |
| txn\_weeks | number of weeks a customer has transacted with the partner |
| avg\_del/week | average delivery amount or average order value per week |
| funding\_label | defines whether a customer is qualified for a loan or not. Has 2 categories: 1. Fundable, 2. Not\_fundable |
| average\_daily\_sales | Average amount of transaction per day |
| reason\_type | reason why a customer is not offered a loan |
| min\_loan | minimum loan offer amount |
| max\_loan | maximum loan offer amount |
| graduation\_difference | difference between minimum and maximum loan offer |
| recency | Number of days since merchant’s last transaction. Difference between today and last transaction date. |
| frequency | Average number of days a customer takes to transact again. |
| monetary\_value | Value of a customer based on order value. Equivalent to total sales amount |
| r\_quartile | recency score on a scale of 1 to 4. 1 means most recent and 4 means worst recency. |
| m\_quartile | monetary score on a scale of 1 to 4. 4 means best and 1 means worst monetary value |
| f\_quartile | frequency score on a scale of 1 to 4. 4 means most frequent and 1 means least frequent |
| RMFScore | Recency Monetary Frequency score |
| merchant\_score | merchant credit score |
| **KYC DATA** | |
| **Attribute** | **Description** |
| customer\_id | Unique identifier of merchant for partner. Similar to merchant\_id |
| DOB | Merchant Date of birth |
| customer\_date\_created | Date merchant was registered on the platform |
| channel | Merchant general business classification by type of business operated |
| sub\_channel | Merchant specific business classification by type of business operated |
| depot\_name | Merchant general classification as per area of operation |
| route\_name | Merchant specific classification as per area of operation |
| ekyc\_status | Merchant loan registration status  YES : Means merchant has consented to taking loans  No : Means merchant ha not consented to taking loans |