Problem Statement:

Amaze Payment Solution (APS) has launched its product in Feb'16. Using this product, users can transact online on various merchants with the credit amount which APS provides. APS has bimonthly credit cycle and dates being 1st and 16th of every month. On these dates bill will be generated for the user of the transaction amount they have transacted in the cycle. Users can pay APS back after the bill generation and continue using the service. Although, users can also do 'onetime_settlements', which means to pay APS before the bill generation of the cycle. If user doesn't pays APS after the bill generation, the bill remains in pending state and user cannot transact until he/she clears his/her dues. APS has data from the launch. Now, they want to predict the repayment behaviour of the users who transacted in the latest cycle (id = 22).

Note:

Transaction data(transactions_data) is given till 2016-09-30 15:30:00 +0000 +0000 (cycle_id=22) Repayments data(settlements_data) is given till settled_at < 2016-09-15 15:30:00 +0000 +0000 (cycle_id=21)

Goal:

You are expected to do EDA on the data and generate a report on variables which are important for user repayments. Also, you have to share us findings, insights and recommendations on the basis of EDA.

Data:

Datasets:

1. credit_data
2. cycles
3. transactions_data
4. users_data
5. settlements_data
6. failure_events_data

1. credit_data:

This dataset has user's credit limit data at cycle level. Users's credit limit gets updated depending upon his transactional and settlements behavior. This dataset gives information of user's credit for a cycle.

Variables: Definition
user_id: Unique indentifier of a user
cycle_id: Unique indentifier of a cycle
global_credit_limit: Credit limit of a user

2. cycles:

Variables: Definition
cycle_id: Unique indentifier of a cycle
start_date: Timestamp when cycle started
end_date: Timestamp when cycle ended

3. transactions_data:

Variables: Definition
transaction_id: Unique indentifier of a transaction
user_id: Unique indentifier of a user
merchant_id: Unique indentifier of a merchant trasnsacting with
transaction_amount: amount of transaction
created_at: transaction timestamp

4. users_data:

Variables: Definition
user_id: Unique indentifier of a user
referrer: User approval source
email: email id of user
name: user name
city_id: city from which user transacts

5. settlements_data:

This data is has user's repayments related variables

Variables: Definition
settlement_id: Unique indentifier of a settlement
user_id: Unique indentifier of a user
cycle_id: Unique indentifier of a cycle
settlement_amount: repayment amount paid by user
settlement_status: type of settlement (bill_pending: if bill is due
bill_settled: if bill is settled
onetime_settlement: paid before bill generation)
days_delayed: number of days repayment delayed
settlment_created_at: bill generation timestamp or onetime_settlement creation timestamp
settlement_updated_at: bill paid timestamp or onetime_settlement paid timestamp

6. failure_events_data:

This data has errors happened while transaction. Due to these errors transaction was interrupted

Variables: Definition
failure_event_id: Unique indentifier of failure event
transaction_id: Unique identifier of transaction
user_id: Unique indentifier of a user
error_type: type of error occured while transacting
amount_in_paise: amount of transaction in while error occured
created_at: timestamp when error occured