Exploratory Data Analysis

BUSINESS PROBLEM

- Amaze Payment Solution (APS) has launched its product so that users can transact online on various merchants with the APS provided credit amount.
- APS has bimonthly credit cycles, On these dates bill has been generated for the user of the transaction amount they have transacted in the cycle.
- Users can pay APS back after the bill generation in order to avail service further.
- Users can also do 'onetime_settlements', which means to pay APS before the bill generation of the cycle.
- Company have found that some of the users were unable to pay back the transaction amount which leads to loss of company.
- Company’s goal to analyze or find the variables which are important for user repayments.

DATA DESCRIPTION

- **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.
- **Cycles:** Bi-monthly cycles information with start and end date.
- **Transaction Data:** Daily transactional data by users for each merchant
- **User Data:** In contains each user information like, name, email id, city, referrer etc.
- **Settlement Data:** This data has user's repayments related variables like settlement amount, date, which cycle, amount settled or not etc.
- **Failure Event Data:** Stores the details about the transactions which got failed and type of error.

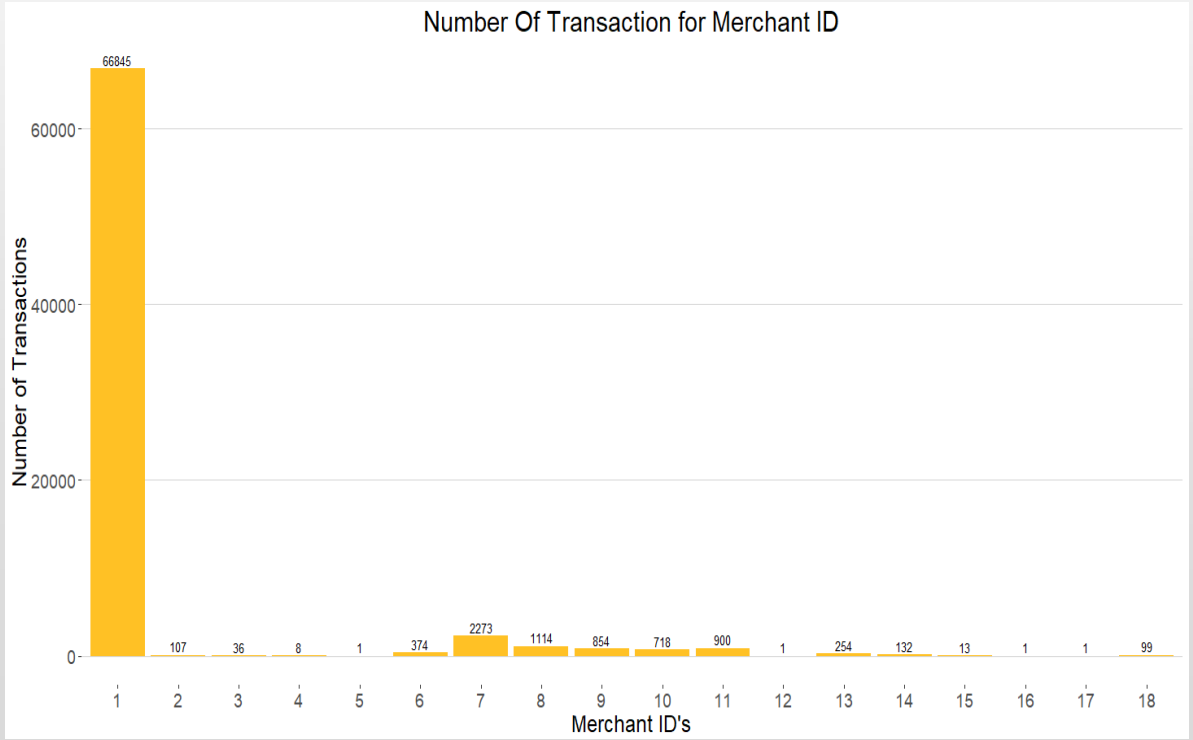
DATA PREPARATION

- **Data Cleaning**
  - Missing or NA values check.
  - NA values substitution
  - Extracted Date and Hour from the Date-Time field.
  - Merging city_id to corresponding user_id in failure event data.
- **Assumptions**
  - In credit data file if any consecutive cycle_id is missing for a user then I've assumed that the credit limit for that cycle (for that user) was same as the previous cycle_id.
  - Settlement data: substituting the days_delay to 0 where settlement_status ='onetime_settlement' and -1 where settlement_status='bill_pending'.
  - For settlement status = 'bill_pending', Date/Month/Hour is from the created date (or bill created date).

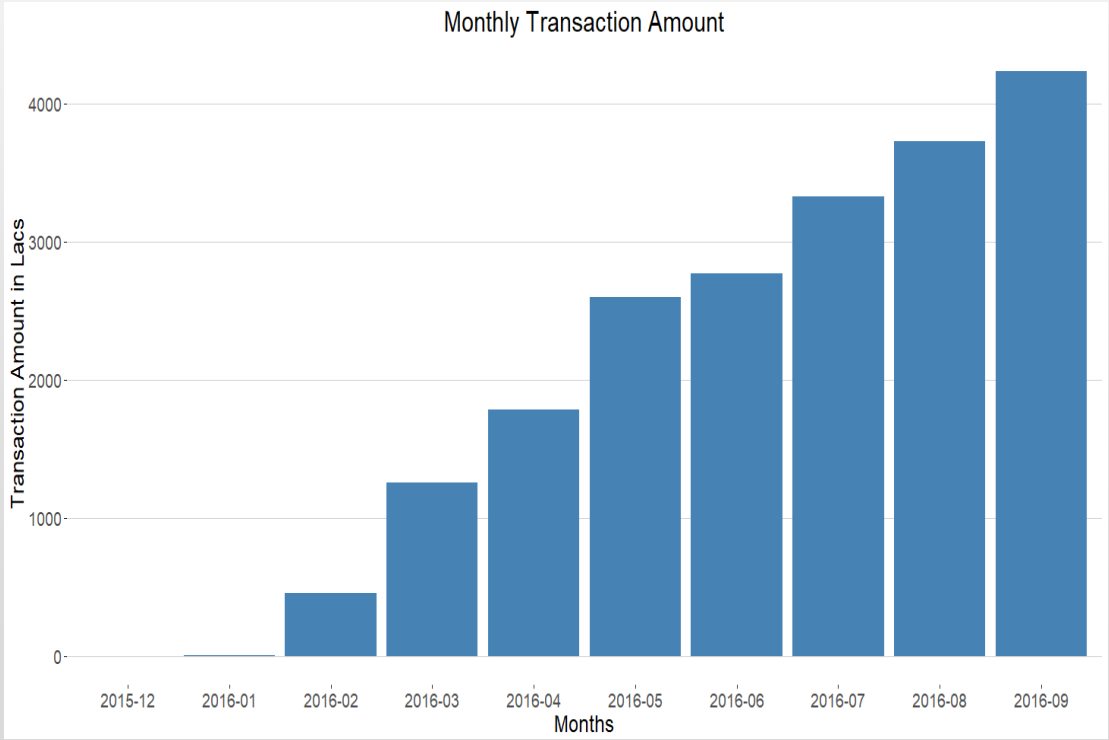
TRANSACTION DATA ANALYSIS

ANALYSIS 1

Users had done majority of transaction on Merchant_Id ‘1’

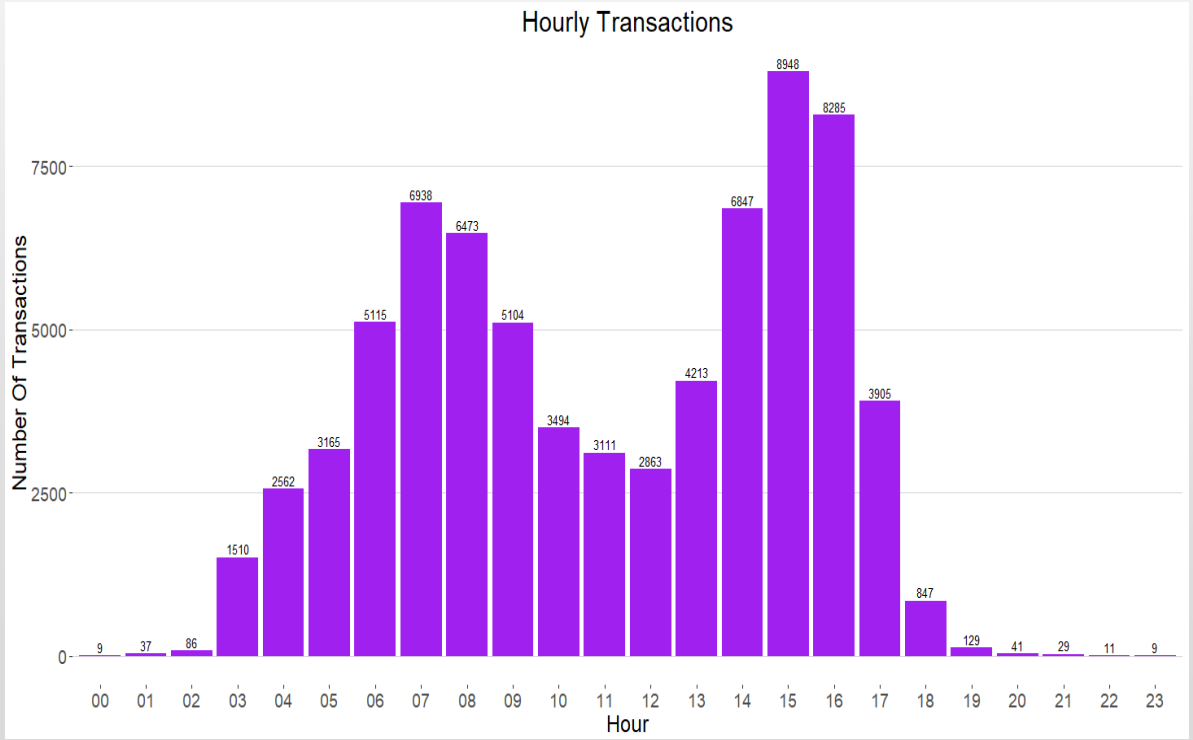


Clearly showing the increase in transactions using the application

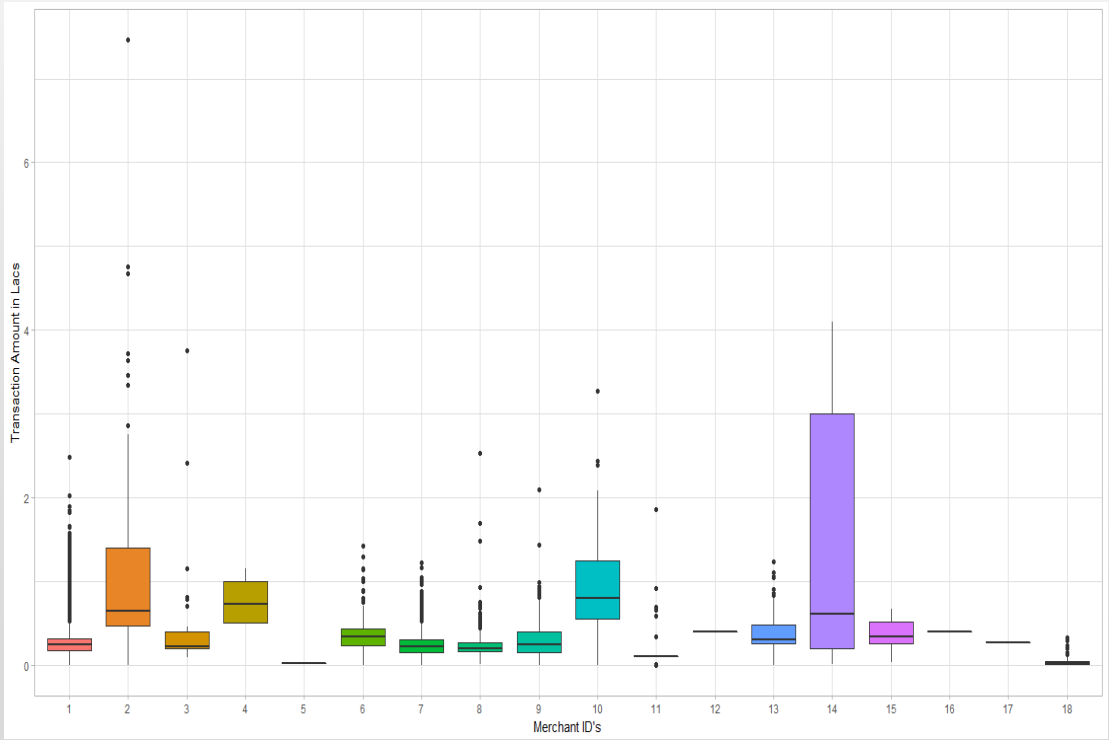


ANALYSIS 2

Majority of transactions happen in morning and late afternoon/early evening



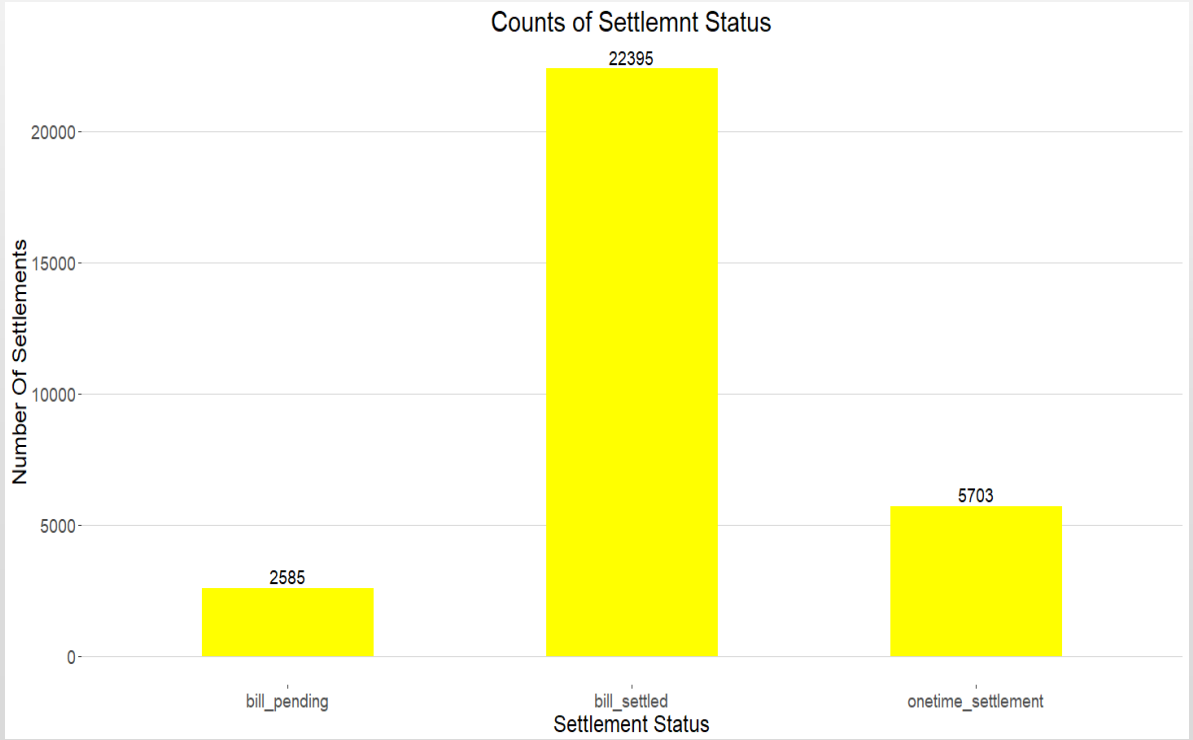
Transaction amount range on different merchants, users done big transactions on merchant 14



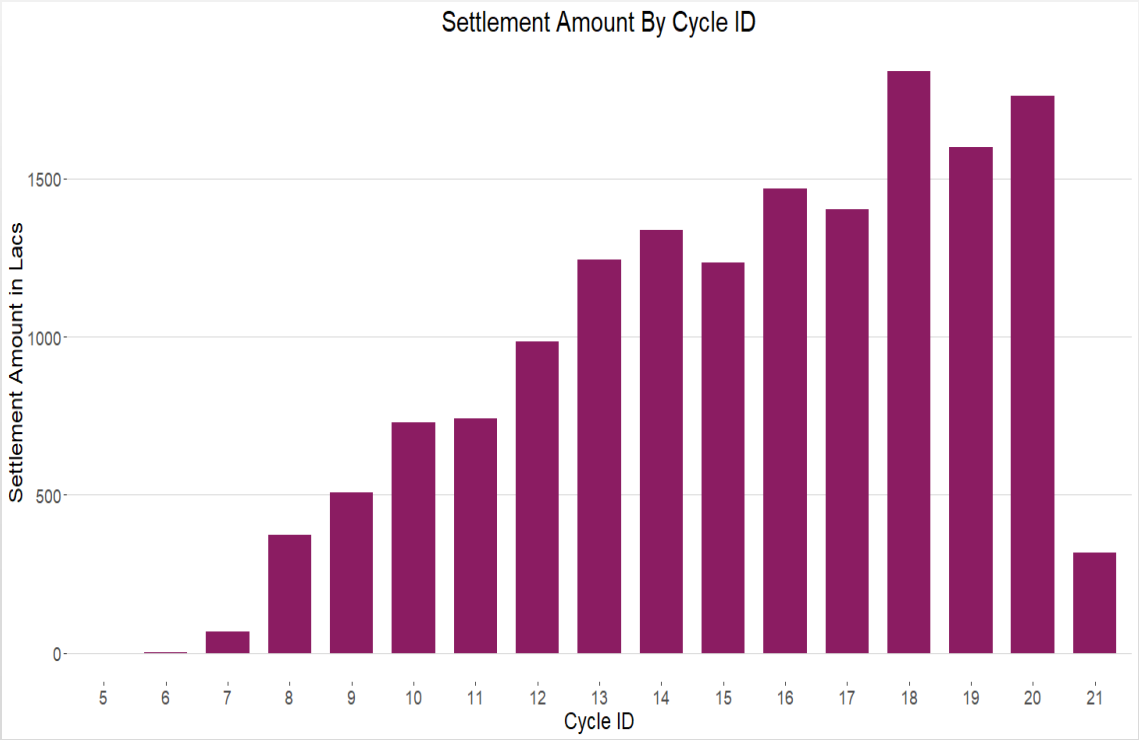
SETTLEMENT DATA ANALYSIS

ANALYSIS 1

People usually preferred to pay bill once the bill is generated

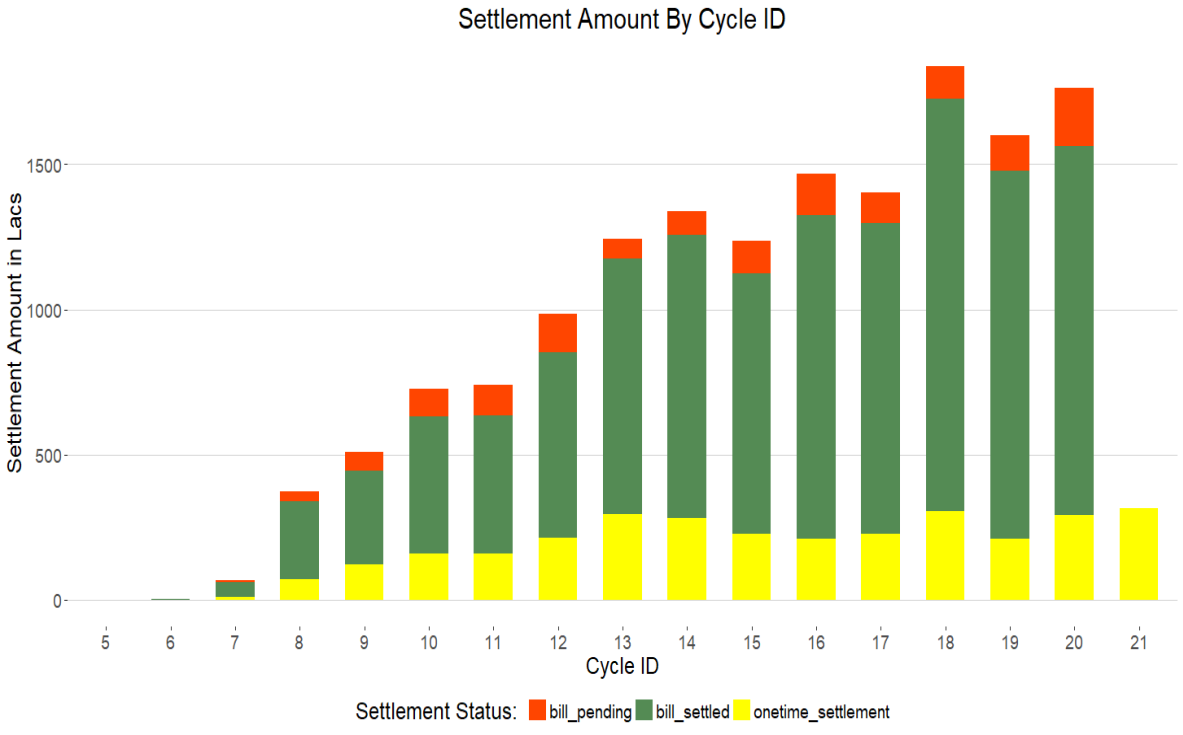


Increasing trend of settlement amount by each cycle (cycle 21 amount is low because it contains onetime_settlement only)

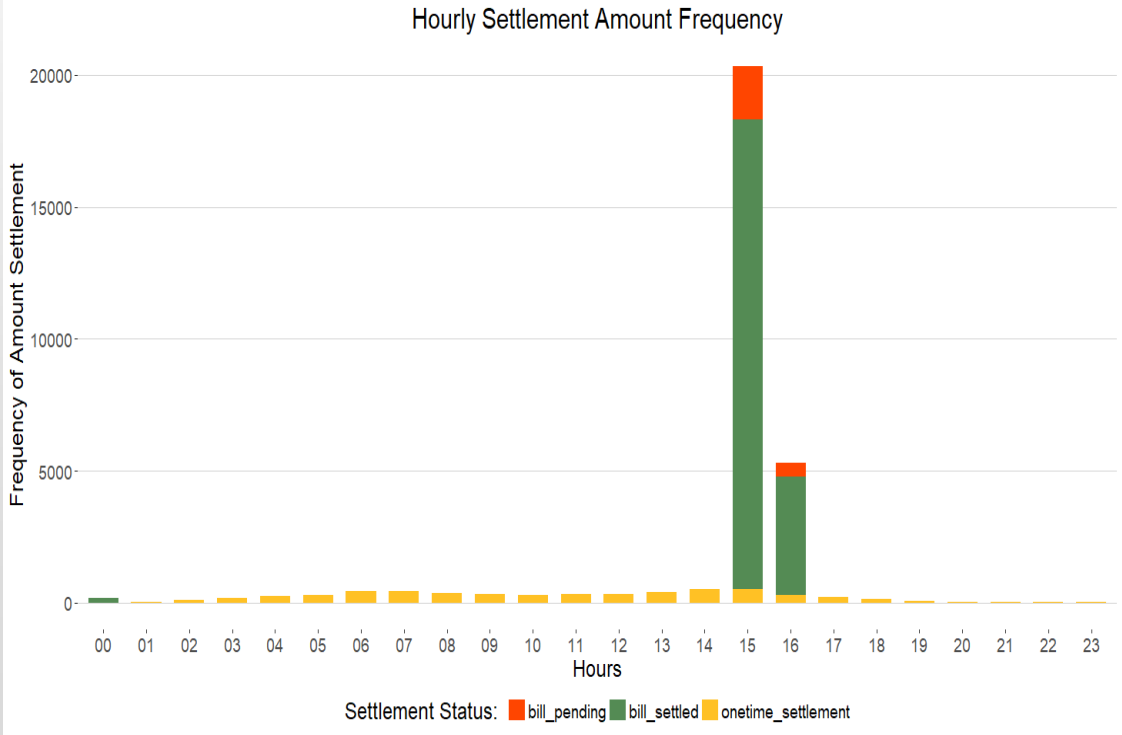


ANALYSIS 2

Each cycle shows increasing trend of bill_settled and onetime_settlement, bill_pending is more or less same

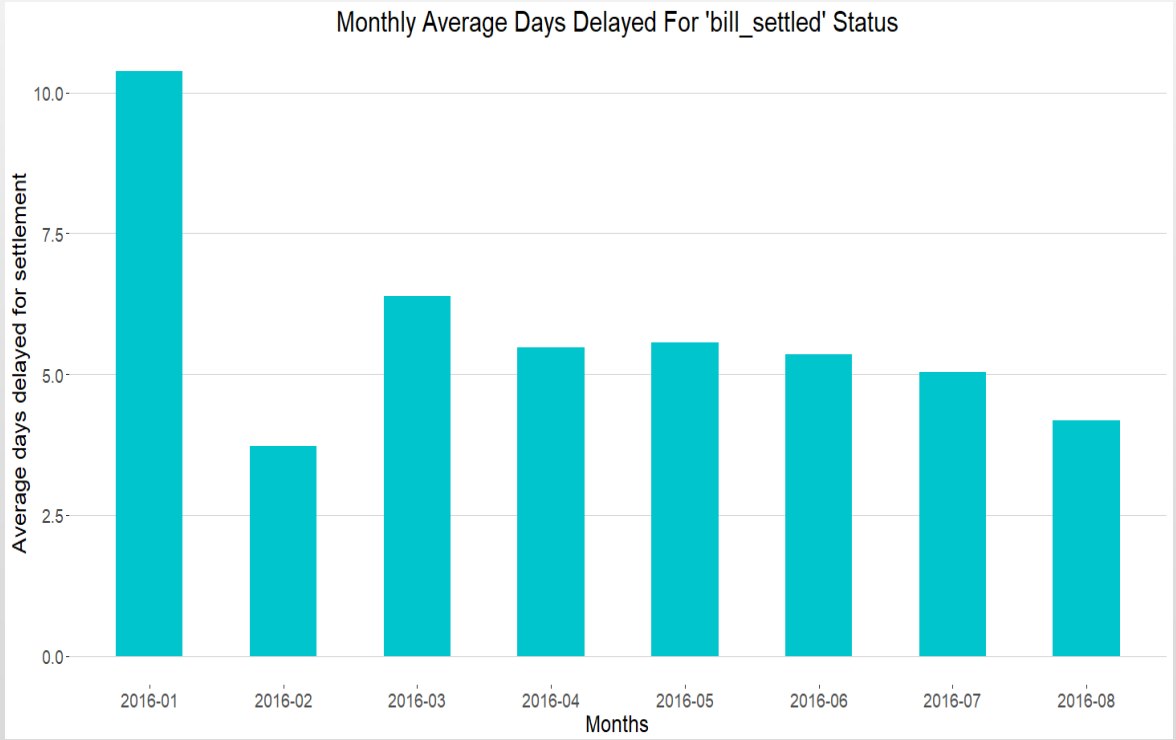


Majority of transactions were settled between 3PM and 4PM

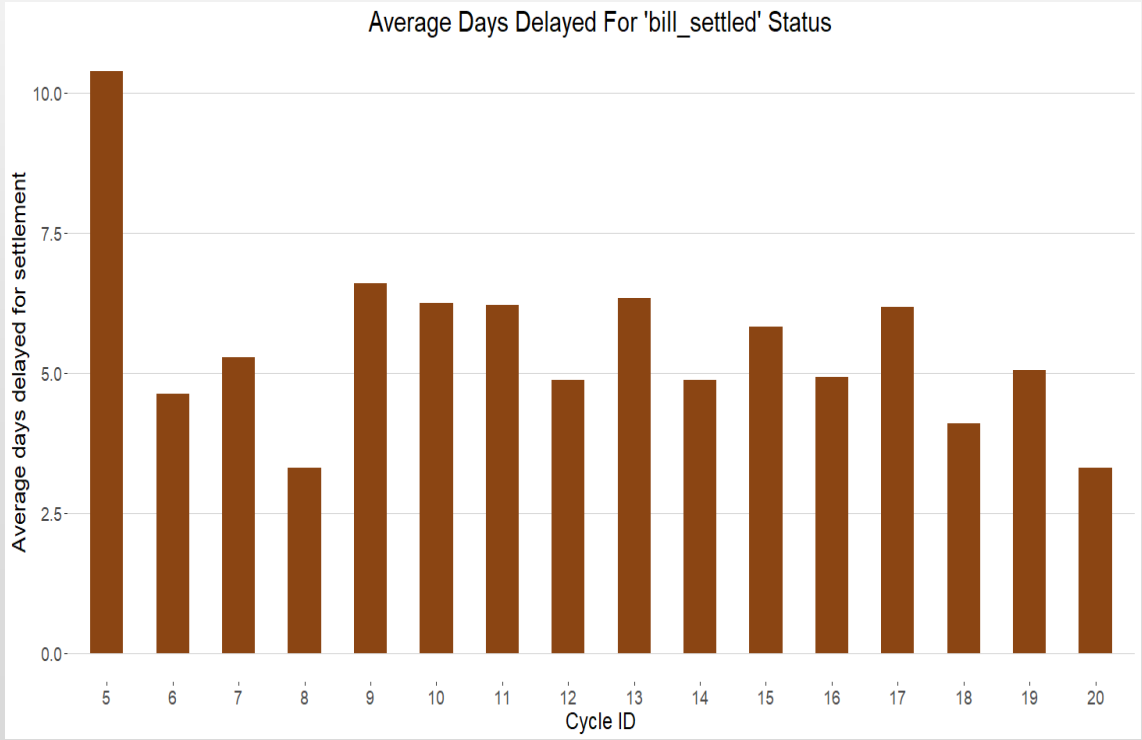


ANALYSIS 3

Average time delayed for bill payment by months. Feb-16 was having less delay in payment



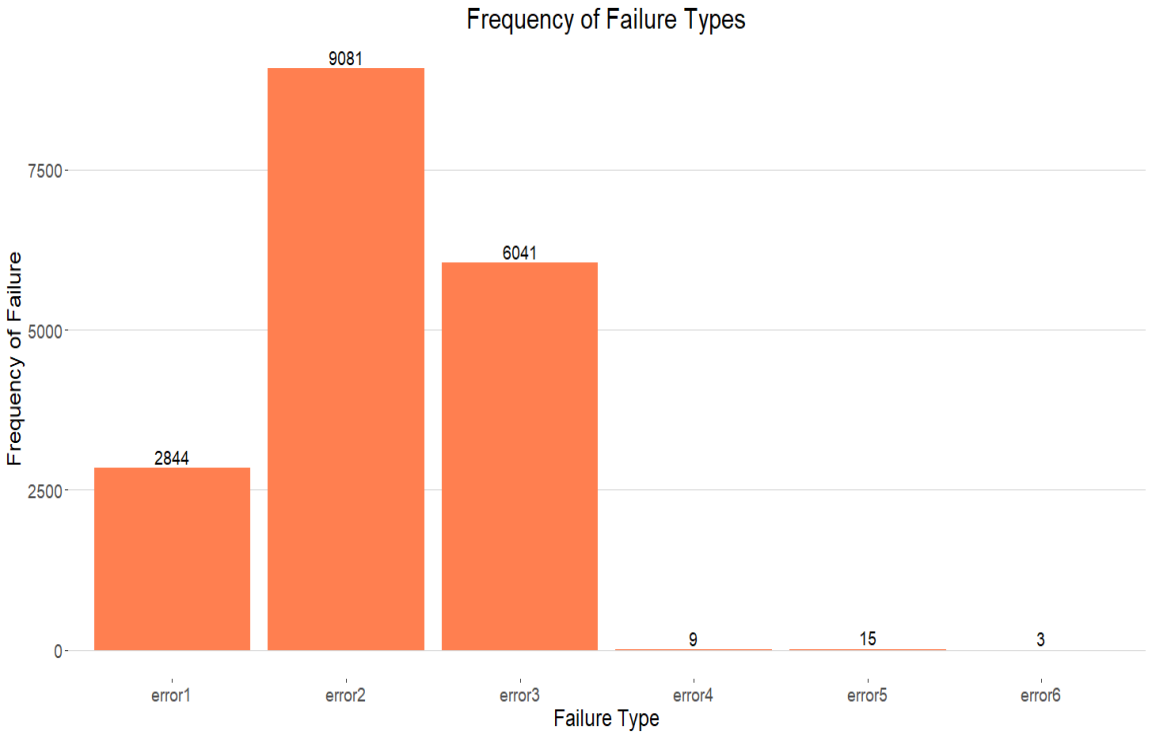
Average time delayed for bill payment by cycles. Cycle 8 was having less delay in payment



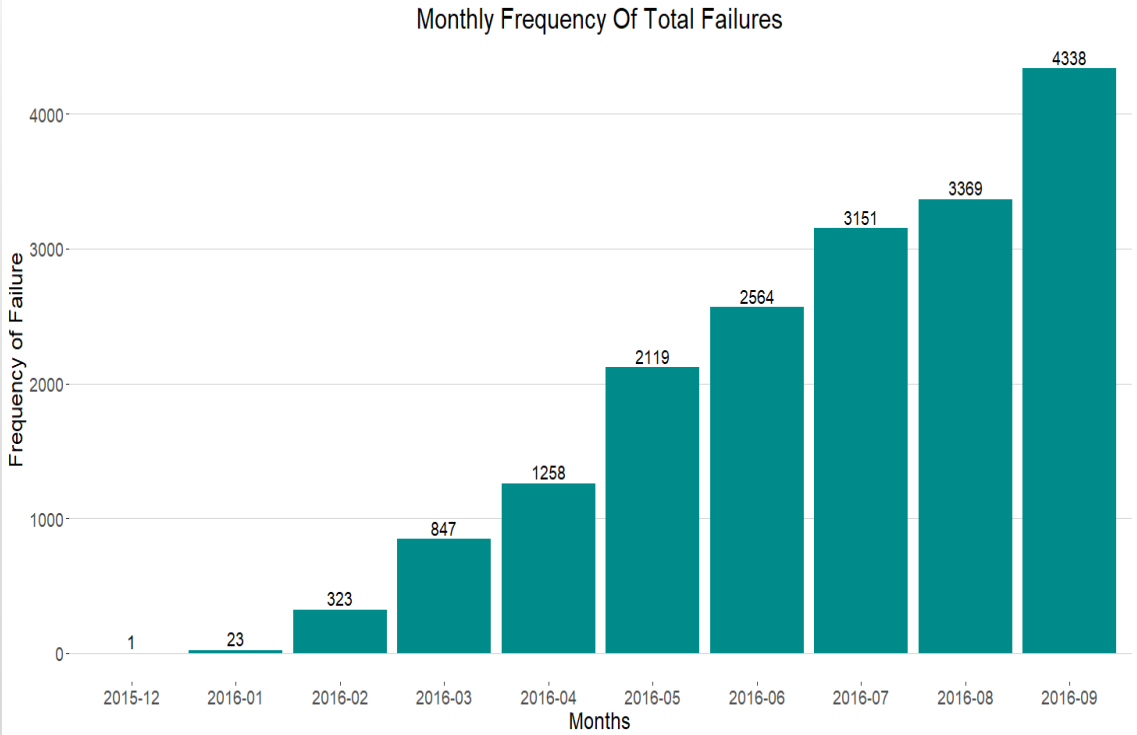
FAILURE EVENT DATA ANALYSIS

ANALYSIS 1

Distribution of failures. Most of the transactions failed because of error type 2

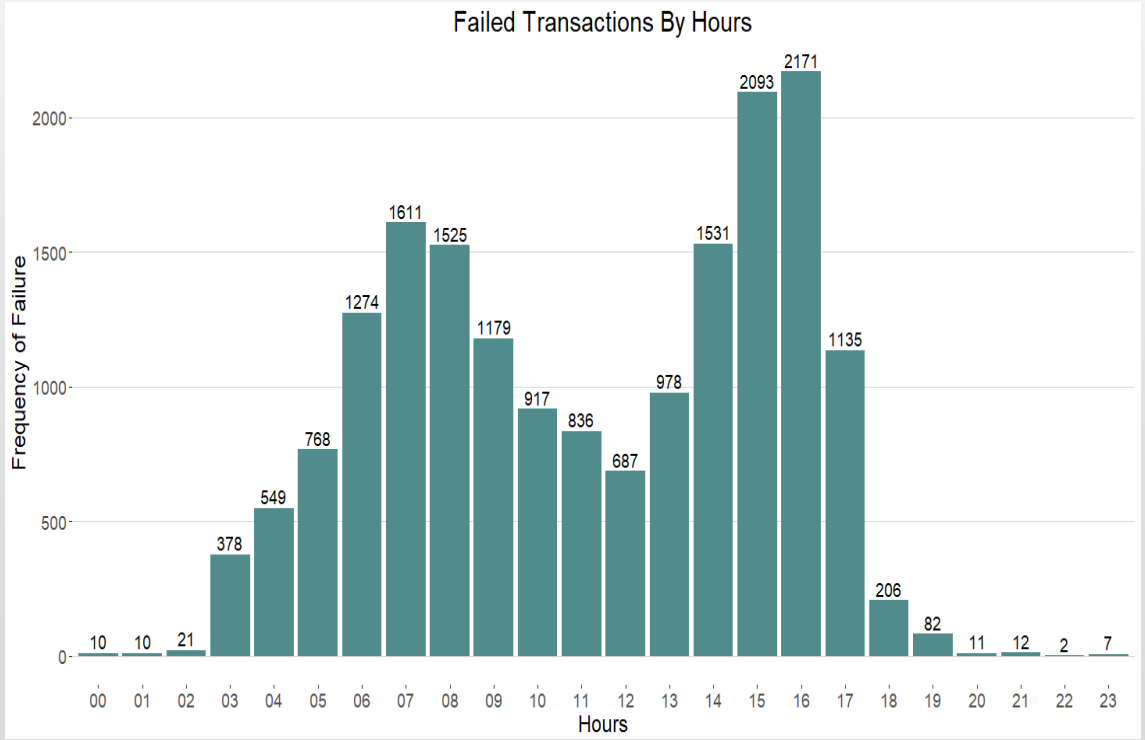


Increasing trend of transaction failure by month (or we can say more the transactions more the chances of failure

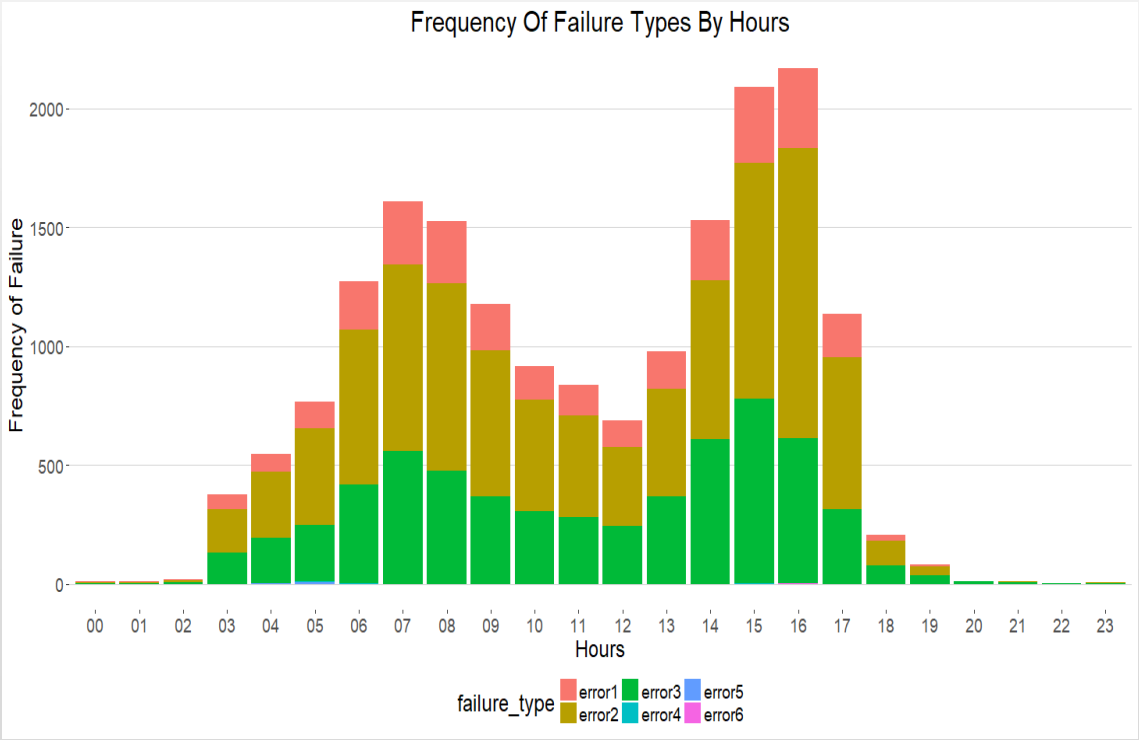


ANALYSIS 2

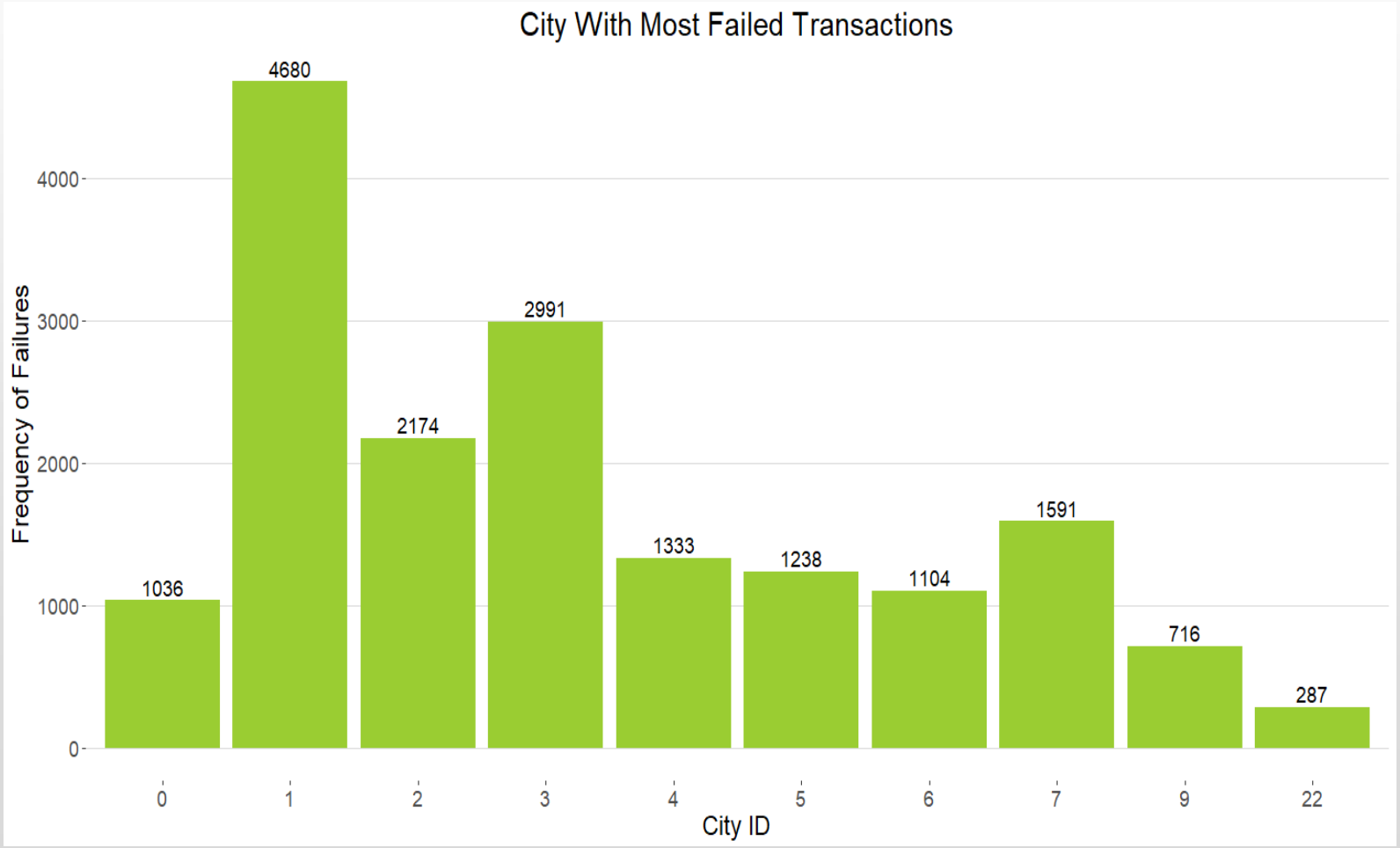
Most of the Transaction failure occurred in morning ,afternoon or early evening (this time most of the people transacts)



error2 and error3 accounts for the majority of failures



ANALYSIS 3



- Graph shows the top 10 cities from where most of the failed transactions happened.
- City 1 hold the most number of transaction failure followed by the city 3.

SUMMARY

- Merchant ID ‘1’ is the preferred/famous platform for transaction.
- Most of the transactions happened in Morning or late afternoon, this is also the time when most of the failures occurred.
- Large amount transactions was performed on merchant ‘14’.
- People usually preferred to pay when the transactions bill is generated.
- Majority of transactions were settled between 3PM and 4PM.
- Usually average delayed time for bill payment is 5-7 days.
- Majority of transactions failed because of error2 and error3, needs to focus on to correct those errors.
- Most of the failed transactions happened from city 1 followed by city 3.