DATA QUALITY REPORT

For Card Transaction Data

PREPARED BY:
ANKUR OJHA
USCID–3686-2408-29

Data Quality Report

1. Data Description

The dataset consists of 96,753 records and 10 distinct fields, documenting real card payments made between January 1, 2010, and December 31, 2010. Each record in this dataset captures essential transaction details such as card numbers, transaction dates, merchant numbers, descriptions, states, zip codes, payment amounts in US dollars, and fraud scores.

Summary Tables

A. Numerical Table

For this dataset we are considering Date and Amount as Numeric Field.

Table 1

Field Name	% Populated	count	Min	Max	Mean	Stdev	# Unique Values	% Zero
Date	100.00	96753	2010-01-01	2010-12-31	/	/	365	0.00
Amount	100.00	96753	0.01	3102045.53	426.885	10006.14	34909	0.00

B. Categorical Table

There are 8 Categorical Fields in this data set, Table 2 (**generated from python codes**) Showing preliminary statistics for categorical field.

Table 2

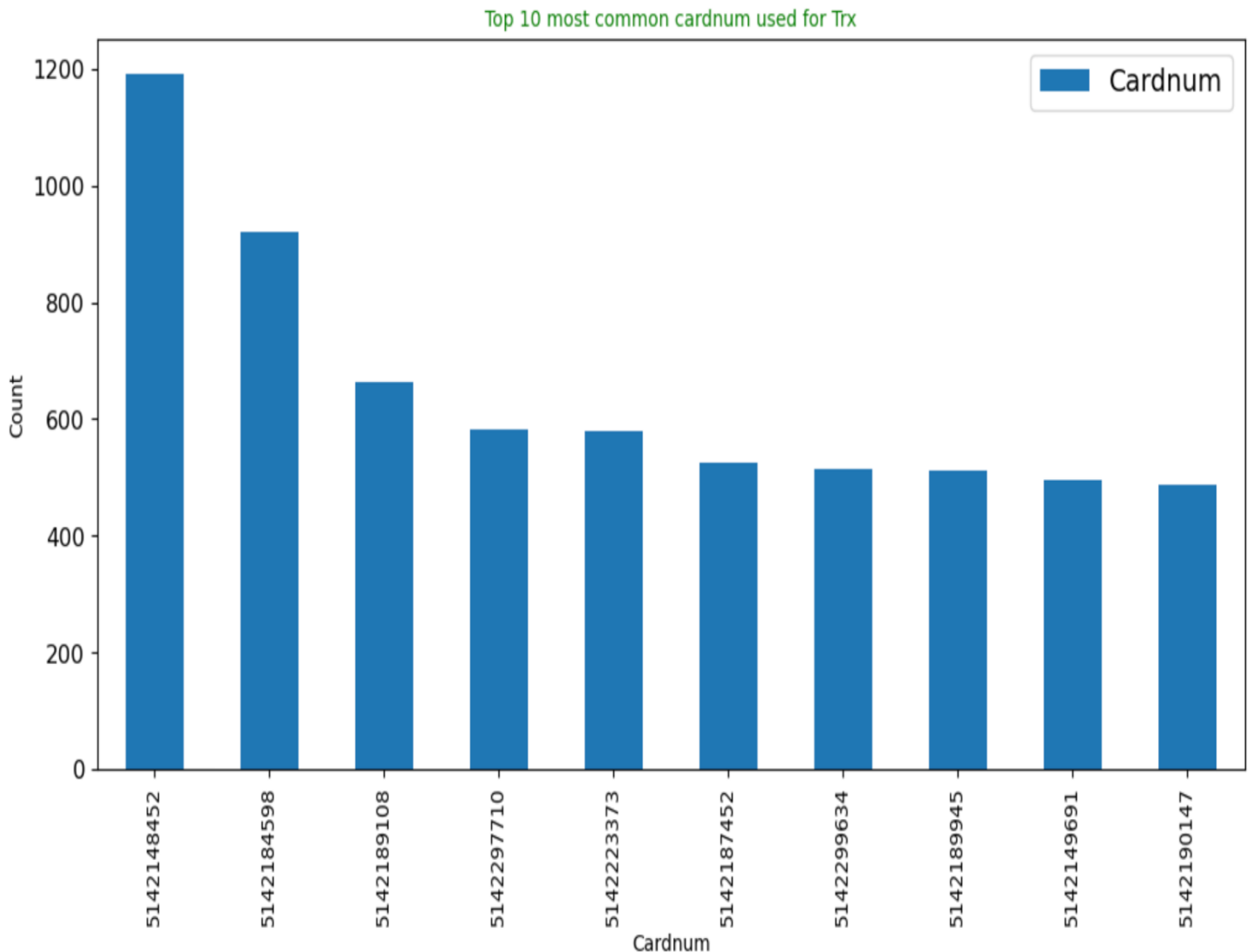| Field Name	% Populated	# blank	# Zeroes	# Unique Values	Count	Most Common Value
Recnum	100	0	0	96753	96753	Unifrom
Cardnum	100	0	0	1645	96753	5142148452
Merchnum	96.51	3375	0	13092	93378	930090121224
Merch description	100	0	0	13126	96753	GSA-FSS-ADV
Merch state	98.76	1195	0	228	95558	TN
Merch zip	95.19	4656	0	4568	92097	38118.0
Transtype	100	0	0	4	96753	P
Fraud	100	0	95694	2	96753	0

2. Visualization of Each Field

The Line charts and bar charts represented for each categorical and numerical field give some high-level information of features of card Transaction data.

1. **Field Name:** record
   Description: A categorical field containing unique positive integers for each record from 1 to 96,753.
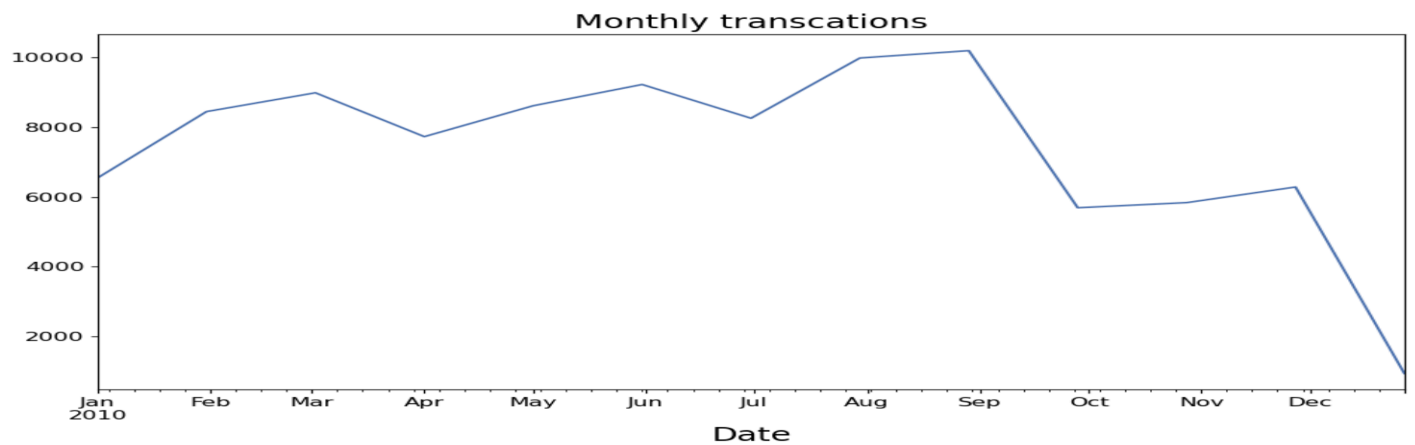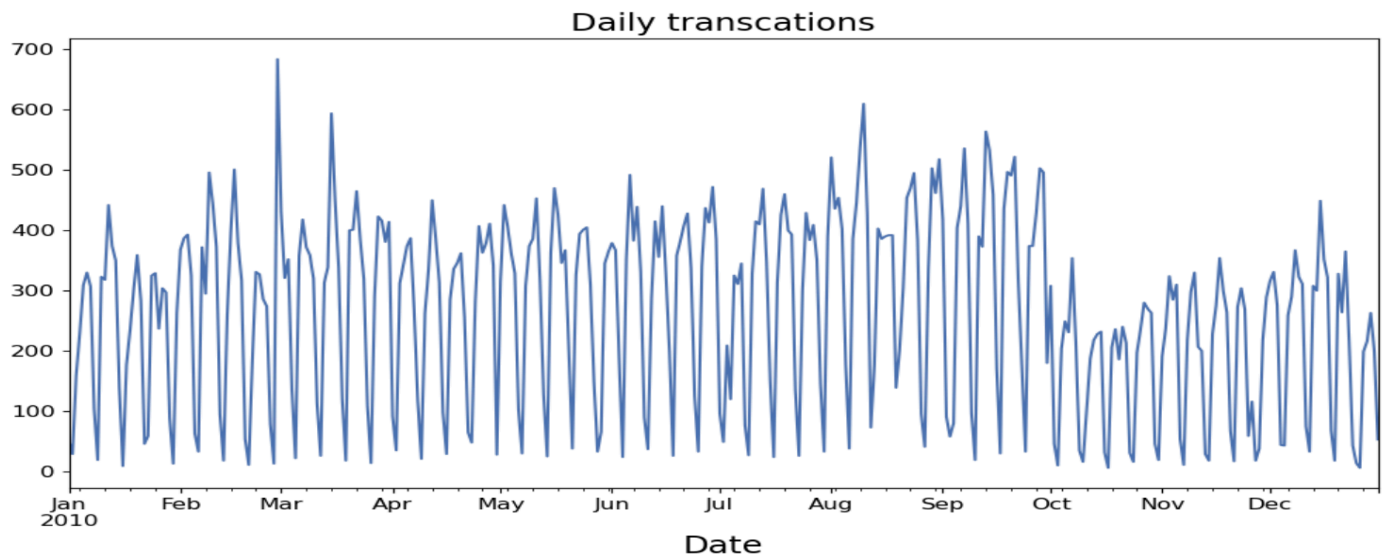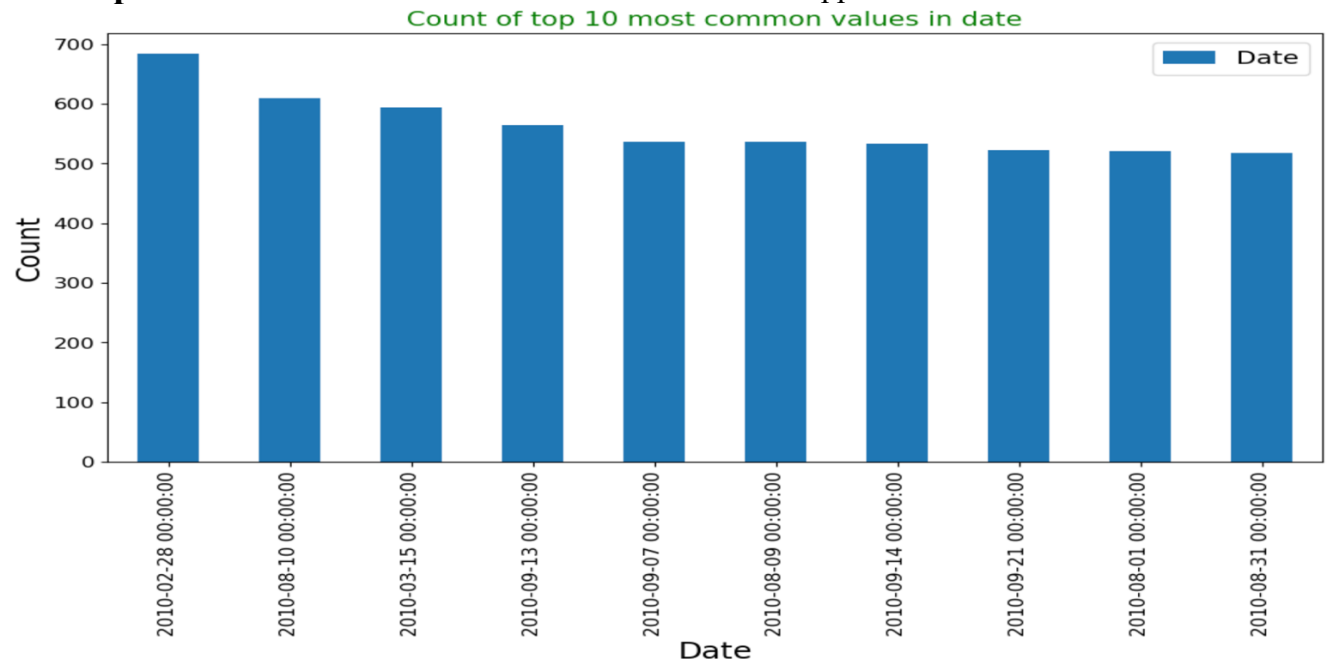
2. **Field Name:** Cardnum
   Description: This field contains the card number of the card, used to pay for transaction.

3. **Field Name:** Date
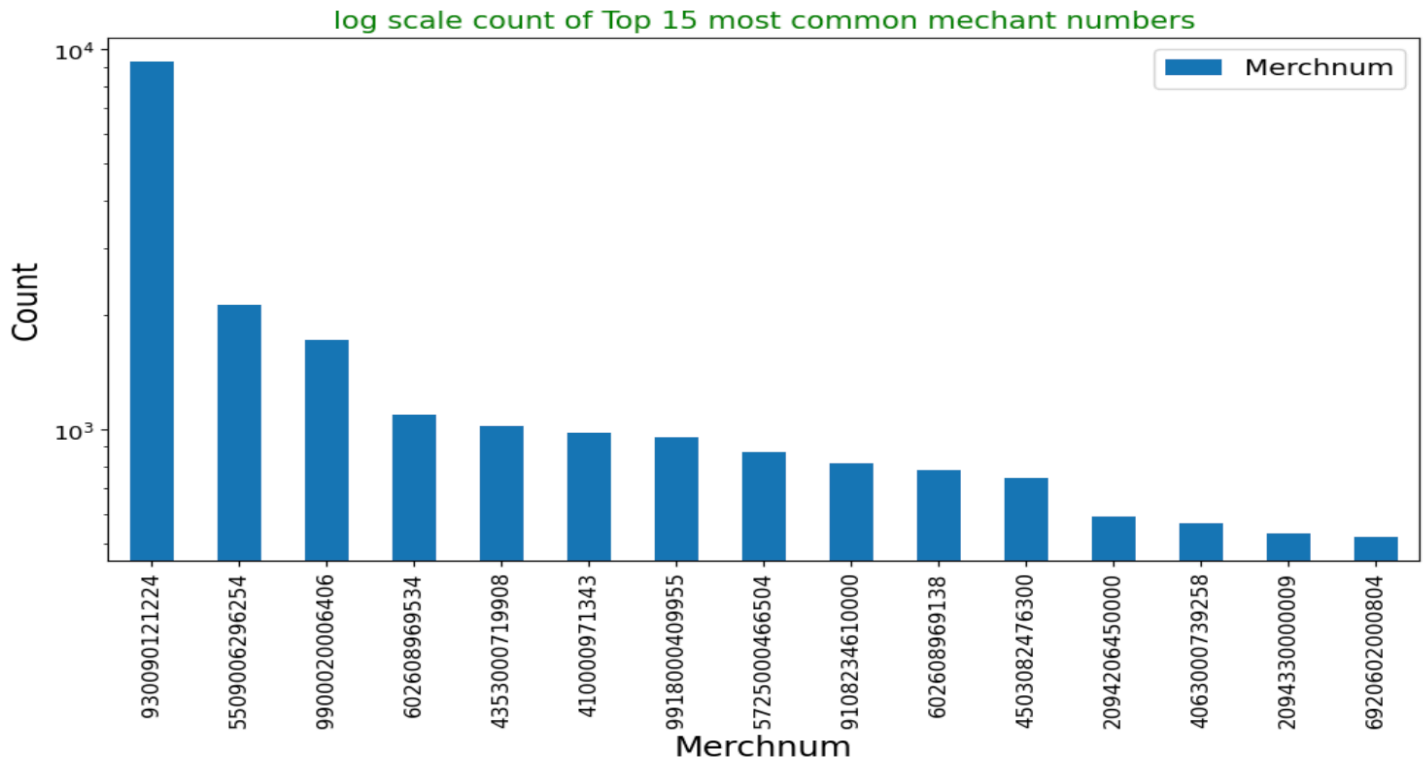   **Description:** Date field records date on which transaction happened.

Count of top 10 most common values in date

Daily transcations
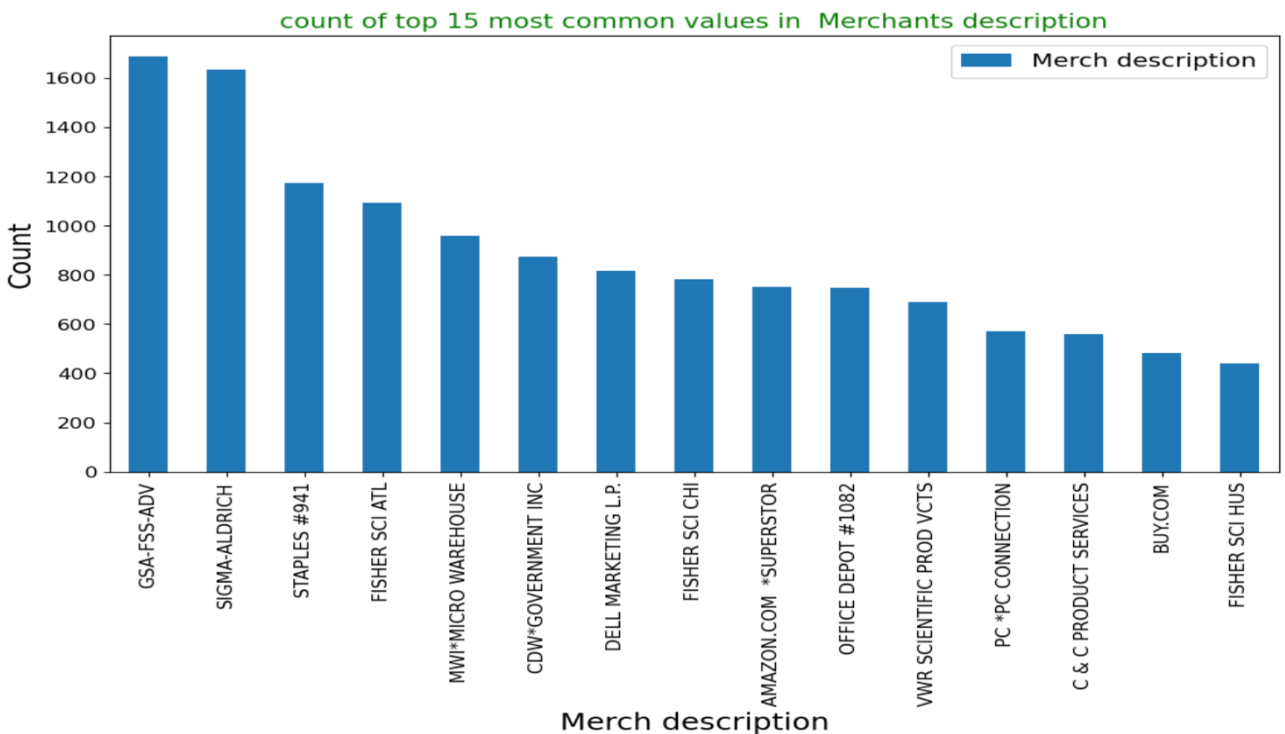

Monthly transcations

3

4. **Field Name:** Merchnum
   Description: This field shows information about the that at which merchant the credit card transaction happened. (Log scale on y axis)
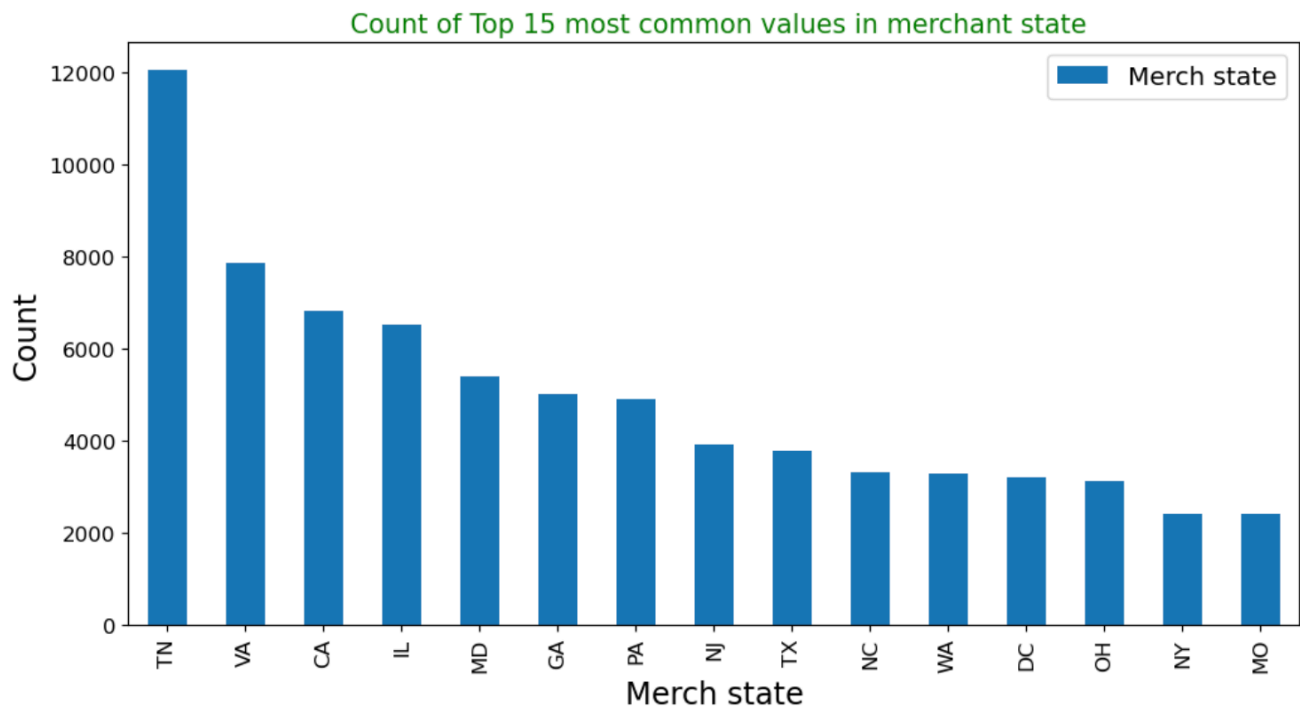

log scale count of Top 15 most common mechant numbers

5. **Field Name:** Merch description
   Description: This field contains brief description of the merchant.
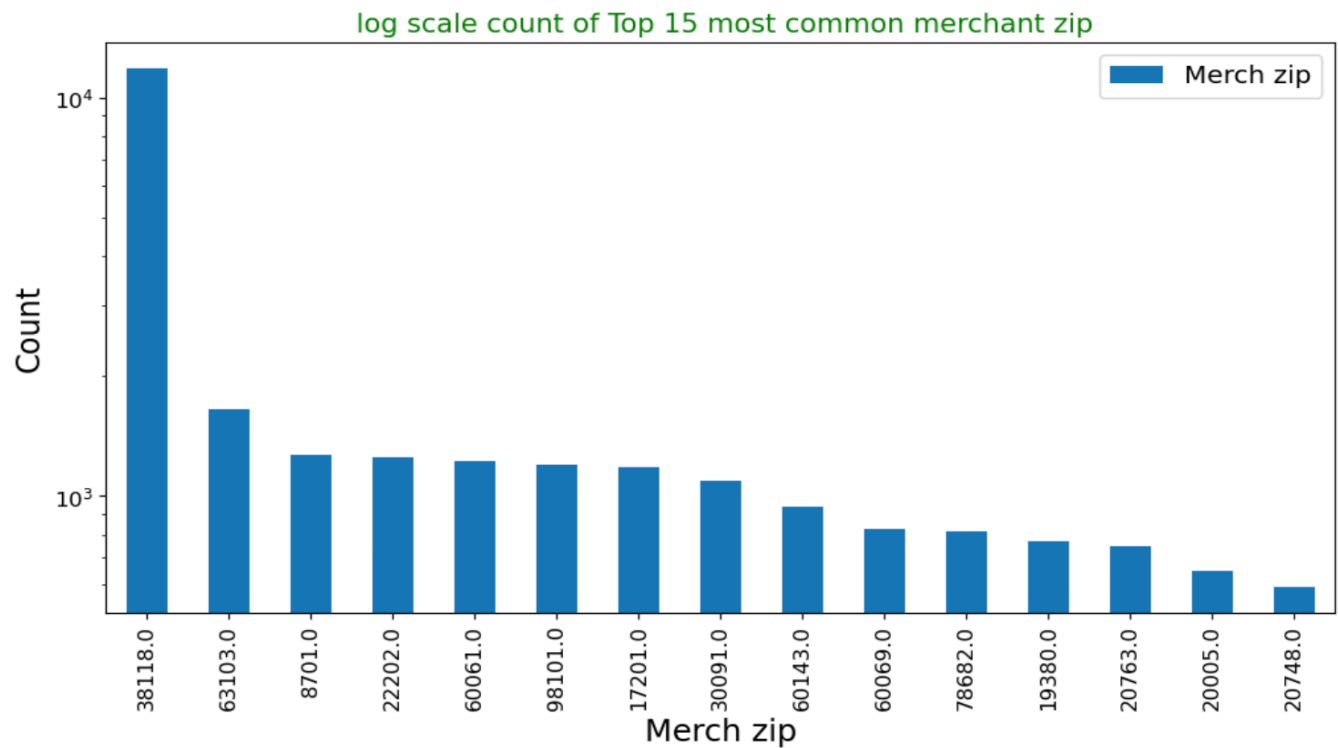

count of top 15 most common values in Merchants description

6. **Field Name:** Merch State
   Description: This Field contains the information of state in which merchant is located.

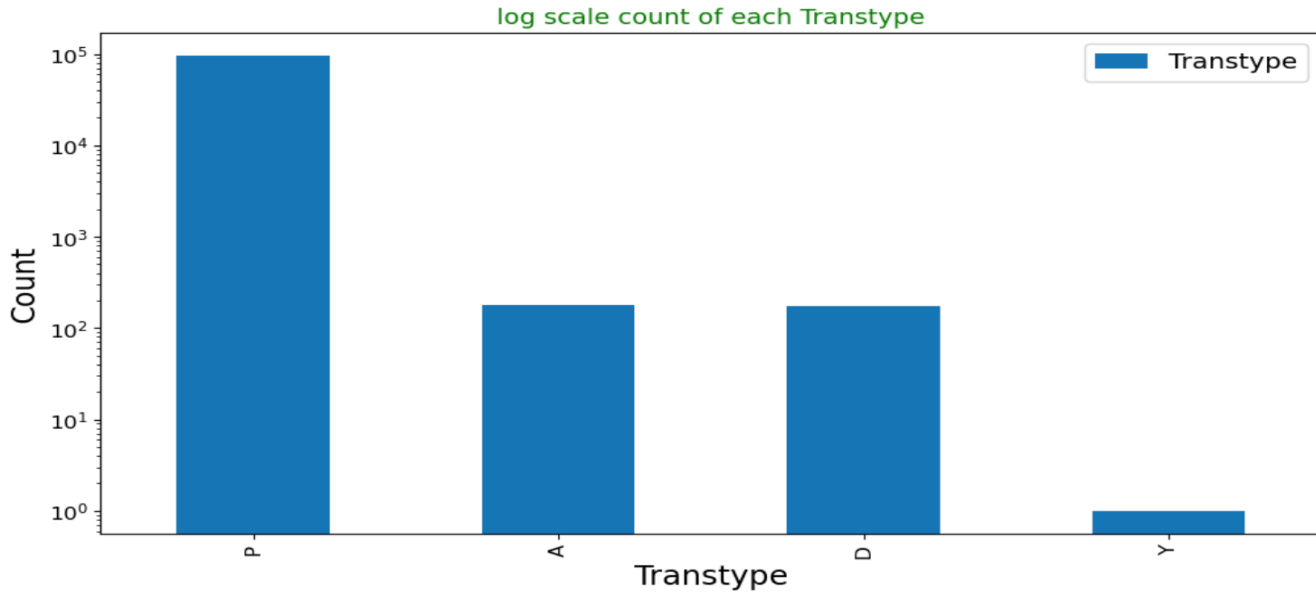Count of Top 15 most common values in merchant state

7. **Field Name:** Merch zip
   Description: Merch Zip represent the zip code of merchant. (Log scale on y axis)

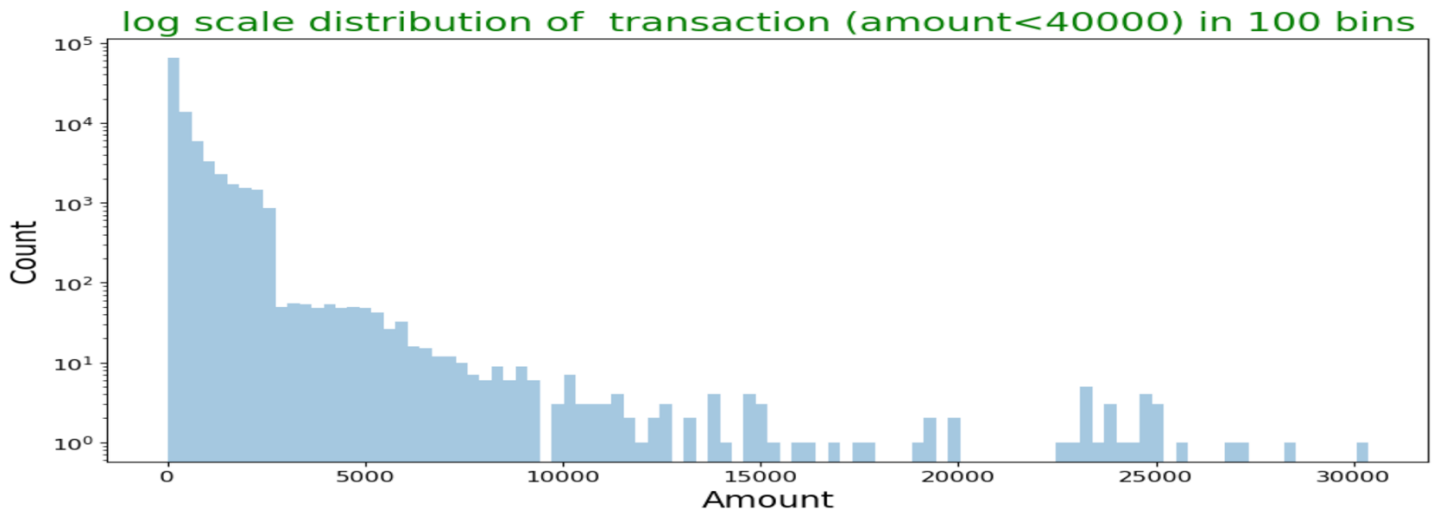log scale count of Top 15 most common merchant zip

8. **Field Name:** Transtype
    Description: This field contains the information about the type of credit card transaction.

9. **Field Name:** Amount
    Description: This field contains the information about amount of money paid in the credit card transaction. (Log scale on y axis)

10. **Field Name:** Fraud:
    Description: This field contains values that indicate whether credit card transaction is fraud or not.

log scale Count of category fraud:1 and not_fraud:0