Data Quality Report

1. Data Description

The Dataset is **Credit Card Transaction Data,** comprising transaction details (*amount, card number, merchant number and description, state and zip, and date of transaction*) between **1645 different credit card users** with **13091 different merchants** across **227 states** in the **year 2010**. There are a total of **97,852 recorded transactions** (records) and **10 different fields**.

2. Summary Tables

Numeric Fields Table :

	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	Min	Max	Mean	Standard Deviation	Most Common
0	Date	Numeric	97,852.00	100.00	0.00	2010-01-01 00:00:00	2010-12-31 00:00:00	NaN	NaN	2010-02-28 00:00:00
1	Amount	Numeric	97,852.00	100.00	0.00	0.01	3102045.53	425.47	9,949.85	3.62

Categorical Fields Table :

	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	# Unique Values	Most Common
0	Recnum	Categorical	97,852.00	100.00	0.00	97,852.00	1
1	Cardnum	Categorical	97,852.00	100.00	0.00	1,645.00	5142148452
2	Merchnum	Categorical	94,455.00	96.53	0.00	13,091.00	930090121224
3	Merch description	Categorical	97,852.00	100.00	0.00	13,126.00	GSA-FSS-ADV
4	Merch state	Categorical	96,649.00	98.77	0.00	227.00	TN
5	Merch zip	Categorical	93,149.00	95.19	0.00	4,567.00	38118.0
6	Transtype	Categorical	97,852.00	100.00	0.00	4.00	P
7	Fraud	Categorical	97,852.00	100.00	95,805.00	2.00	0

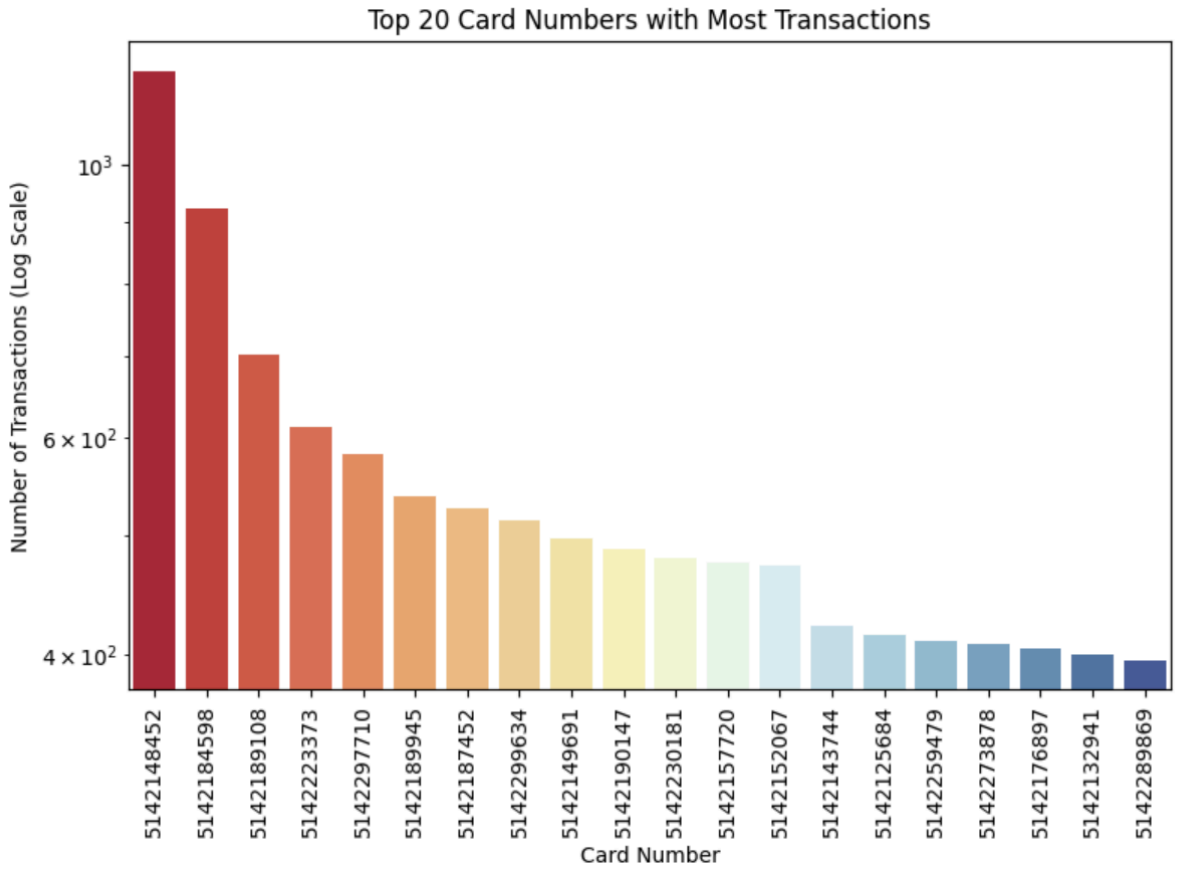
3. Visualization of Each Field

1. **Field Name : Recnum**

   Description: Unique, positive record numbers from 1 to 97,852, depicting all the unique transactions made by the card users. (It does not make sense to visualize all 97,852 record numbers.)

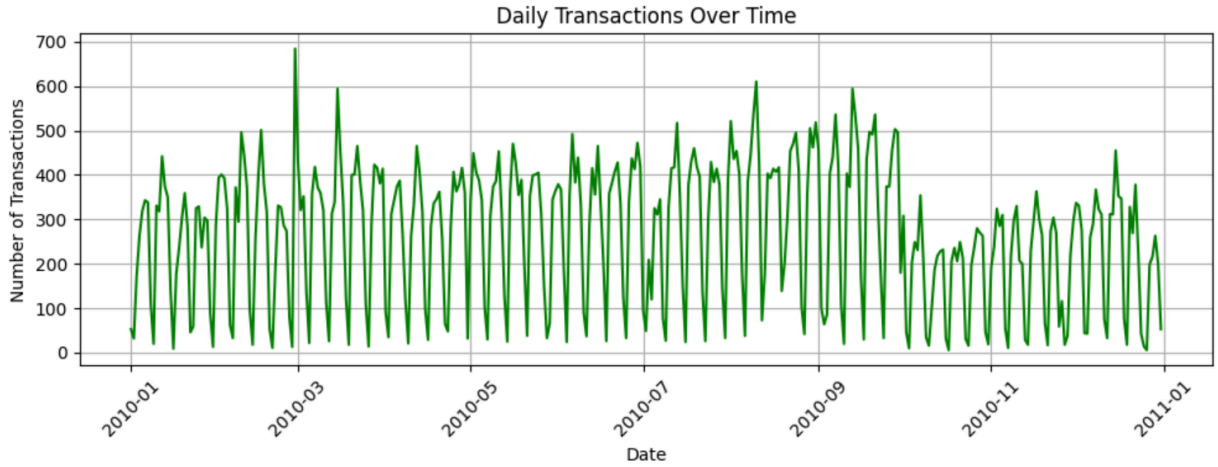
2. **Field Name : Cardnum**

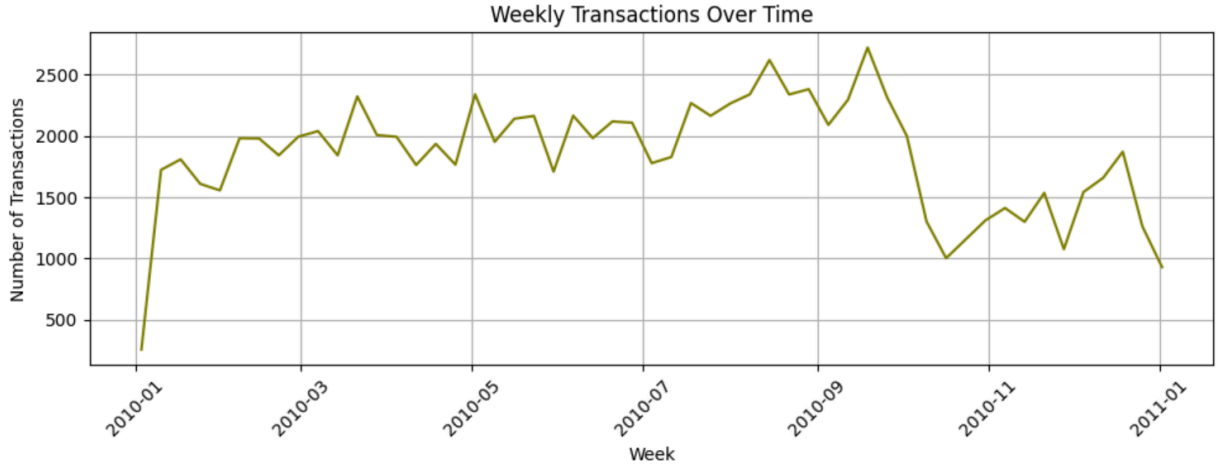
   Description: The Credit Card Number. A 10 digit number identifying the card. There are a total of 1645 unique cards carrying out transactions in this dataset. Plotting a barchart of the top 20 cards with most transactions.

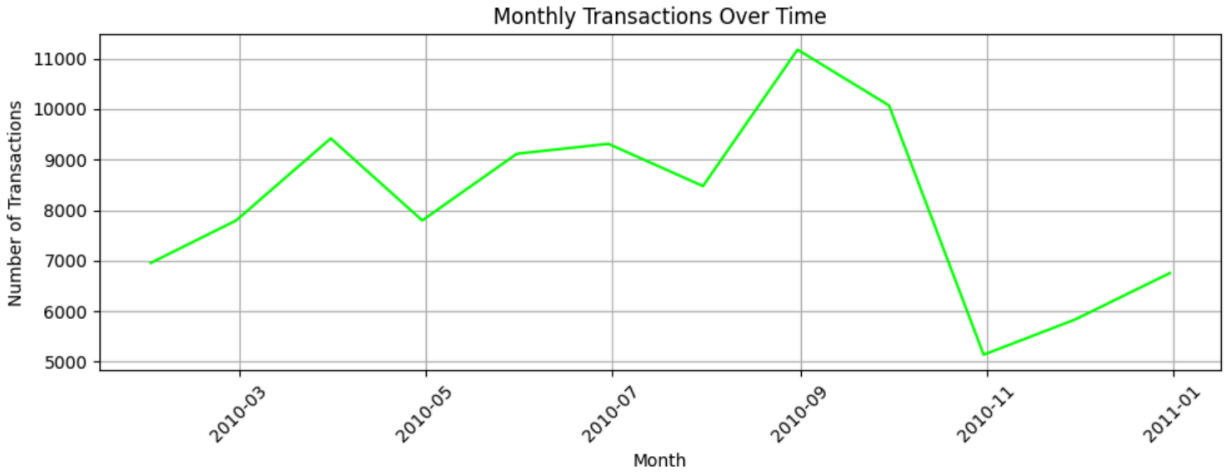


3. **Field Name : Date**

Description: Date of the Transaction. Plotting three graphs to show (a) Daily transactions, (b) Weekly Transactions, and (c) Monthly Transactions.

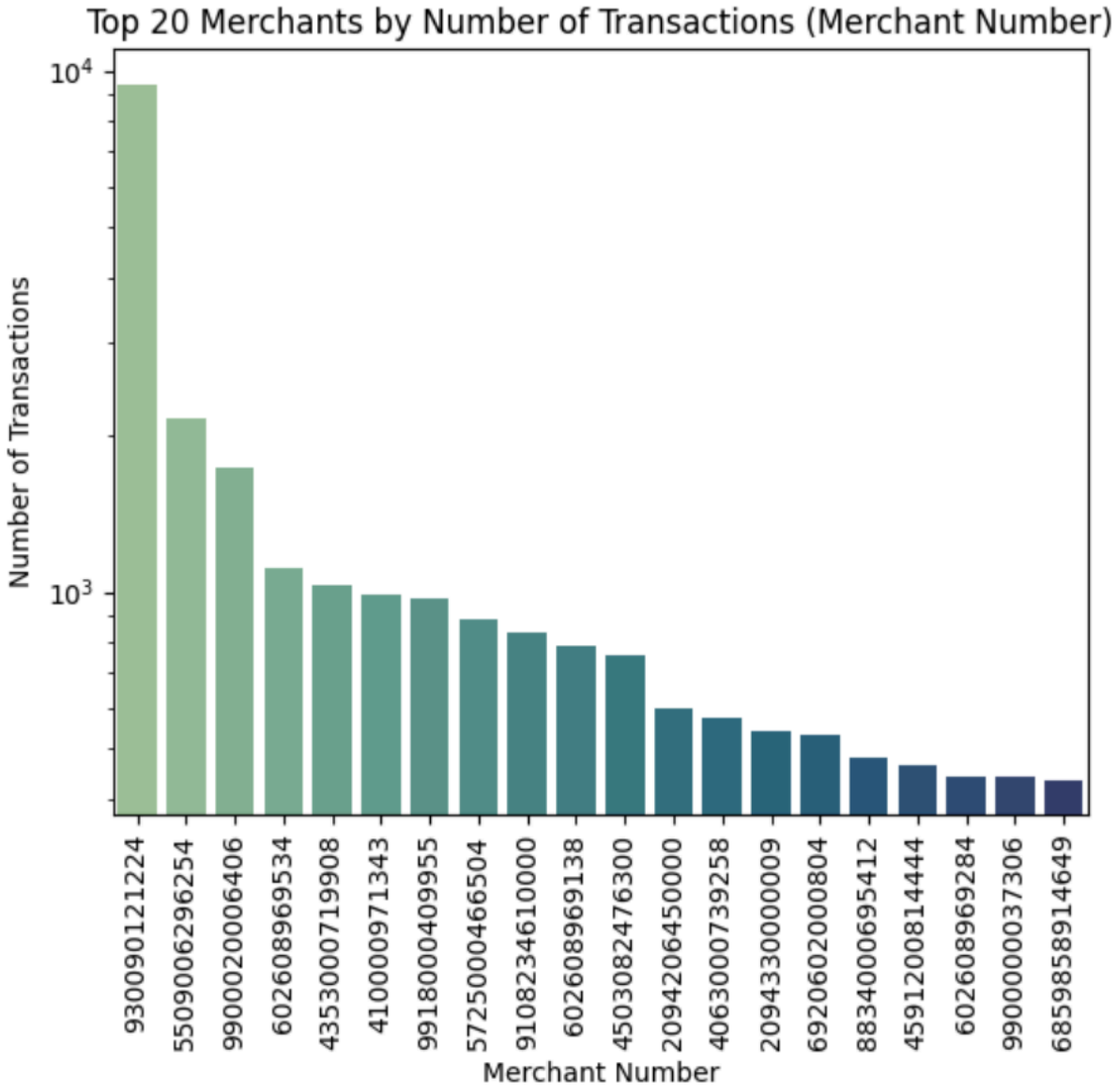






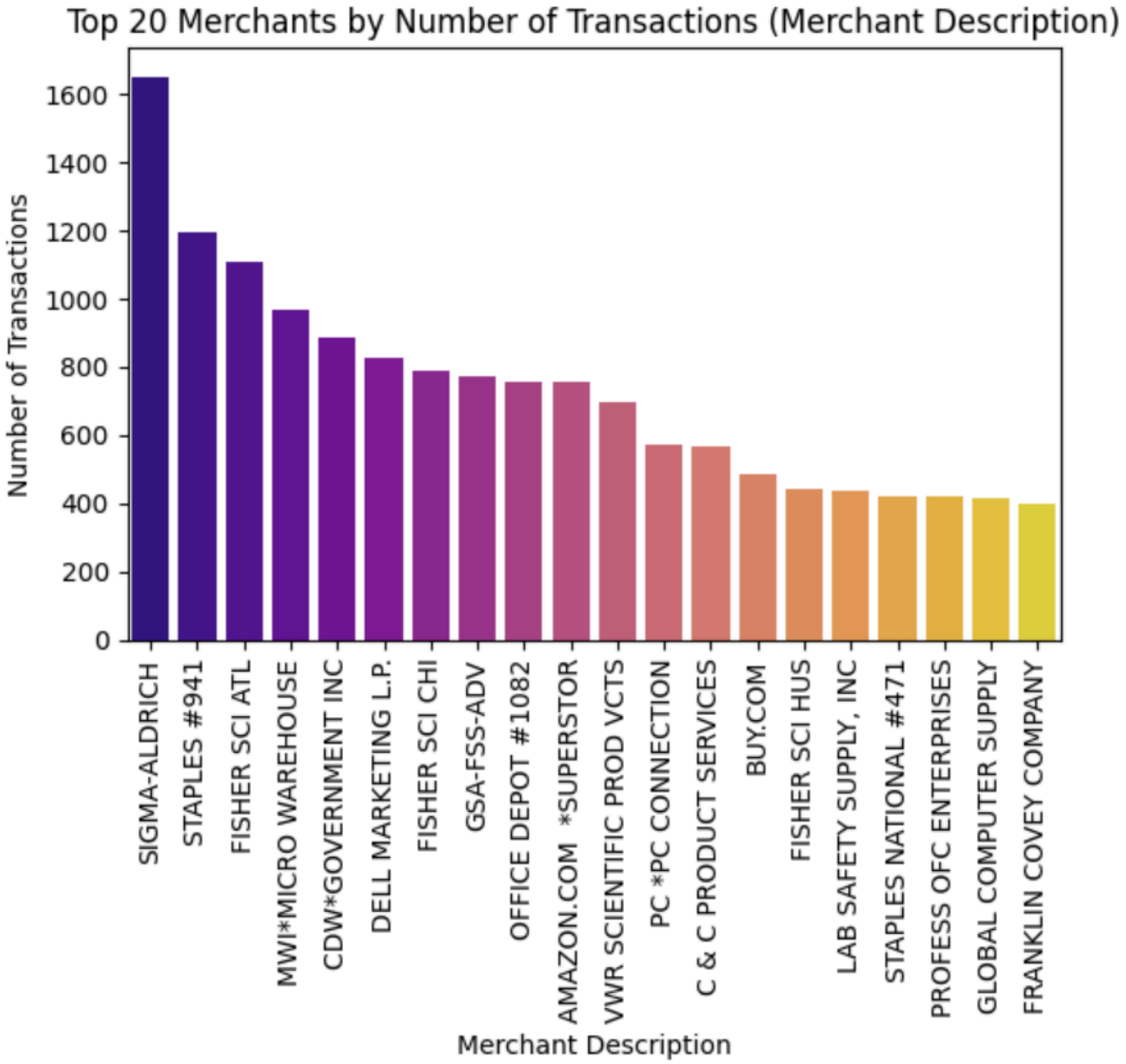
4. Field Name : Merchnum

Description: Merchant Number. Unique ID/number of the merchant with whom the customer is carrying out the transaction. 13091 unique merchant number values are present in the dataset. Depicting the top 20 merchants with most transactions carried out.



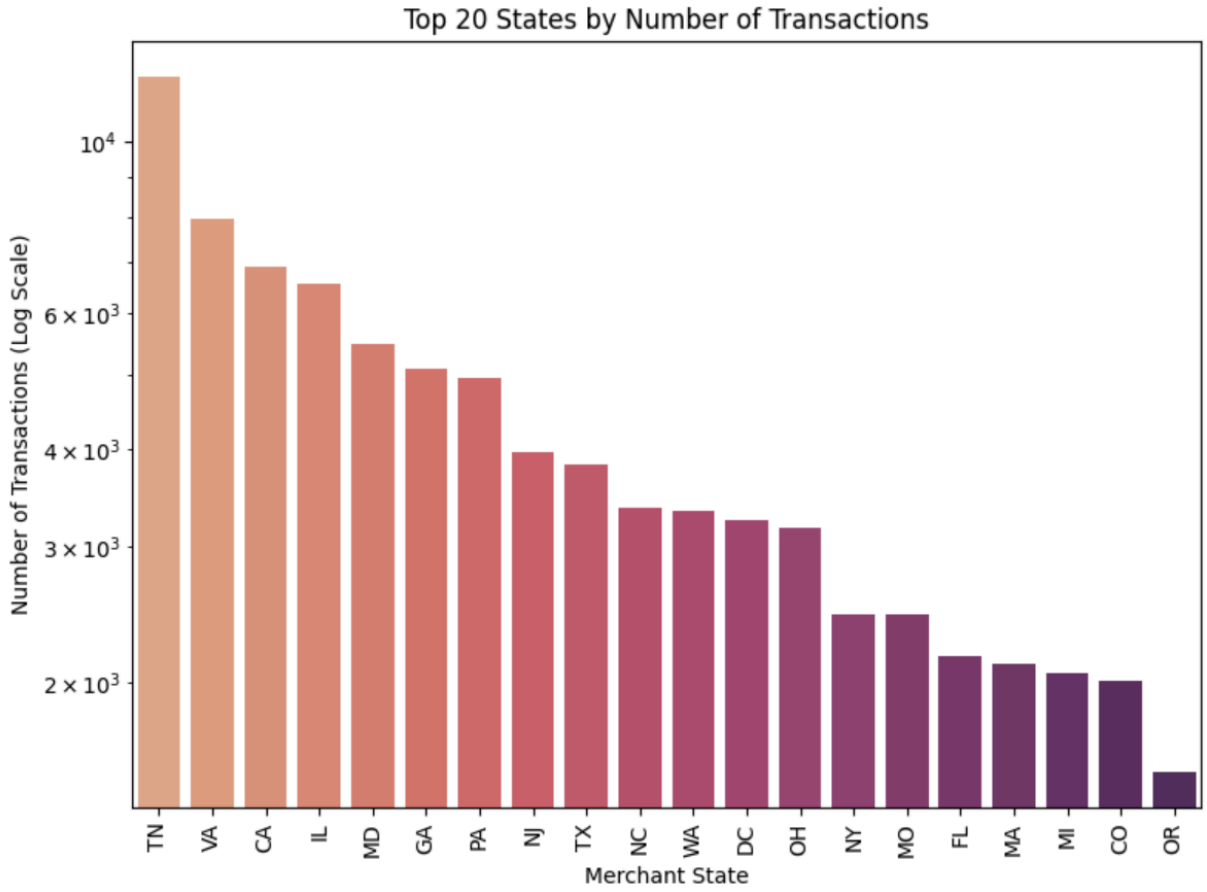
5. **Field Name : Merch description**

Description: Merchant Description. Name of the merchant with a slight description. 13,126 different values.



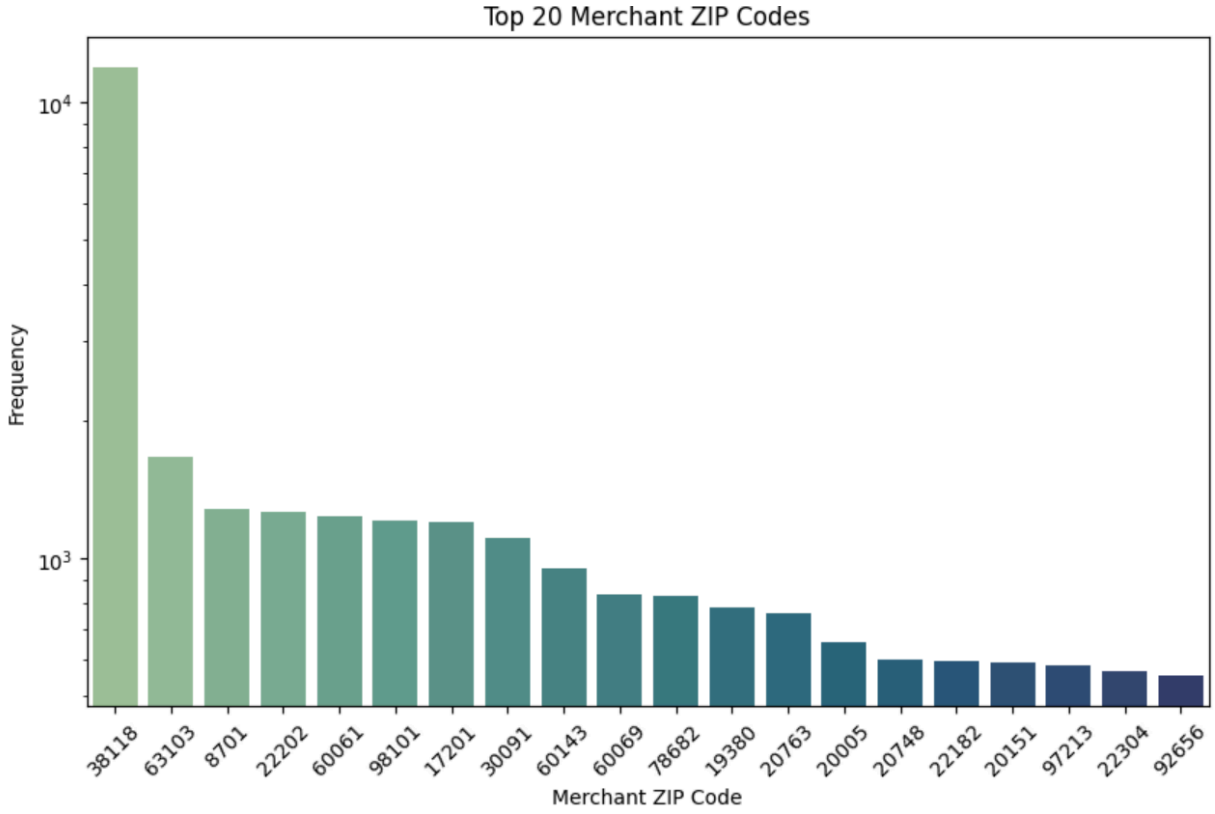
6. **Field Name : Merch state**

Description: Merchant State. The state of the merchant from where the transaction was carried out.



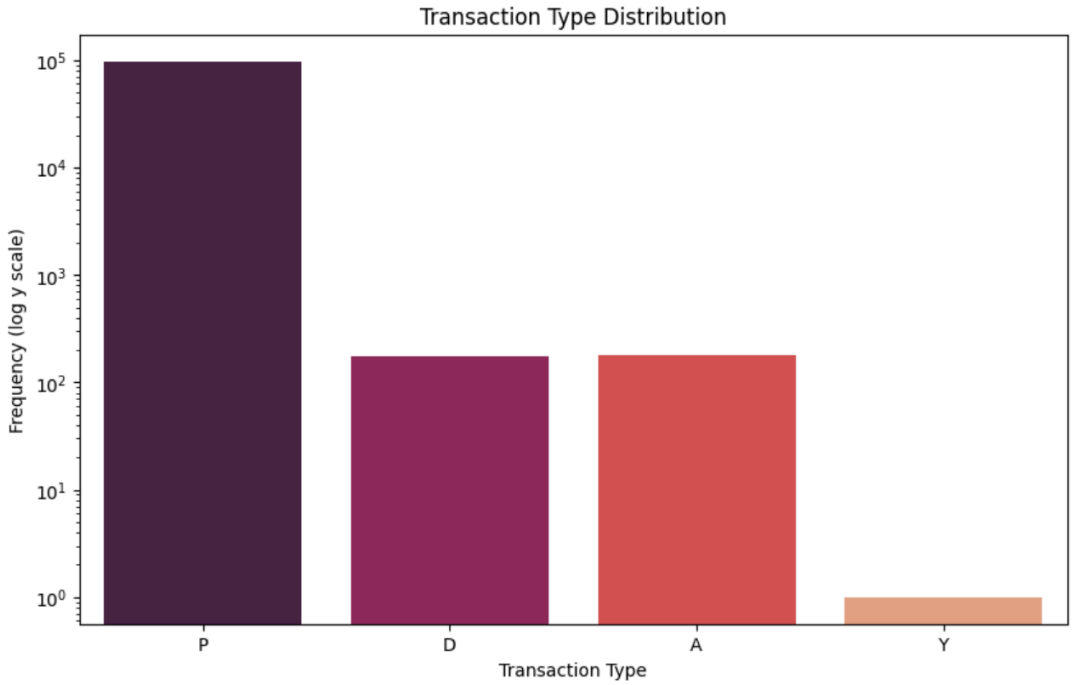
7. **Field Name : Merch zip**

Description: Merchant zip code. The 5 digit zip code of the merchant from where the transaction was carried out. Ignoring the empty records, the plot I get is -



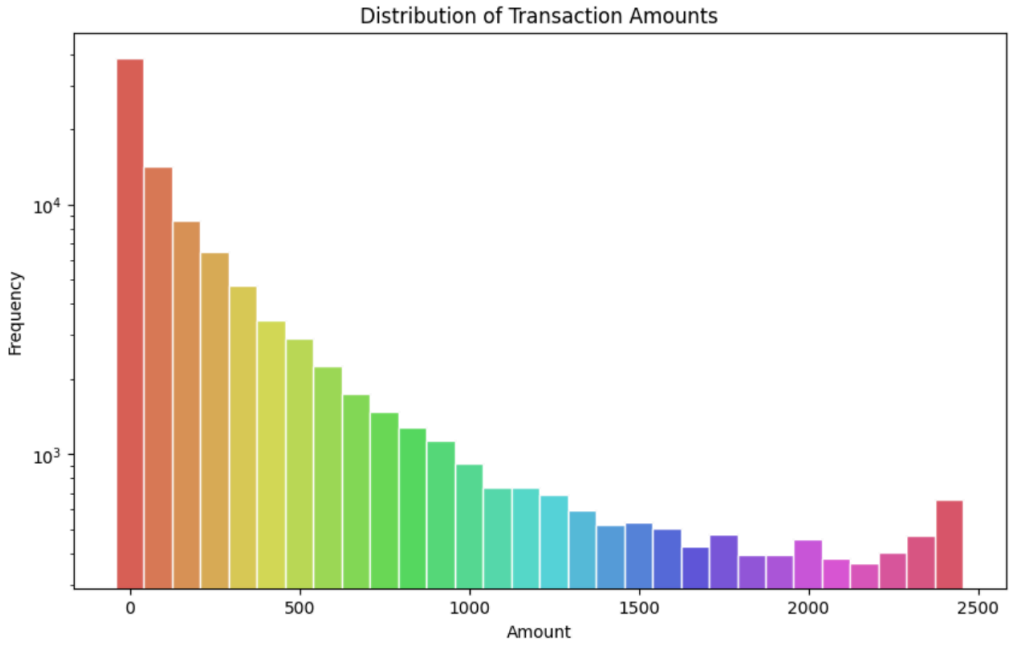
8. Field Name : Transtype

Description: Transaction Type. This field depicts the type of transaction and has four categories- P, D, A, and Y. These could correspond to Authorized purchases, Pre-authorized transactions, ATM Withdrawals, Denied Transactions etc. Clearly the transaction type “P” is the most common type of transaction in this dataset with 97,497 occurrences, followed by "A" with 181 occurrences, next is "D" with 173 occurrences and “Y” has only 1 occurrence.

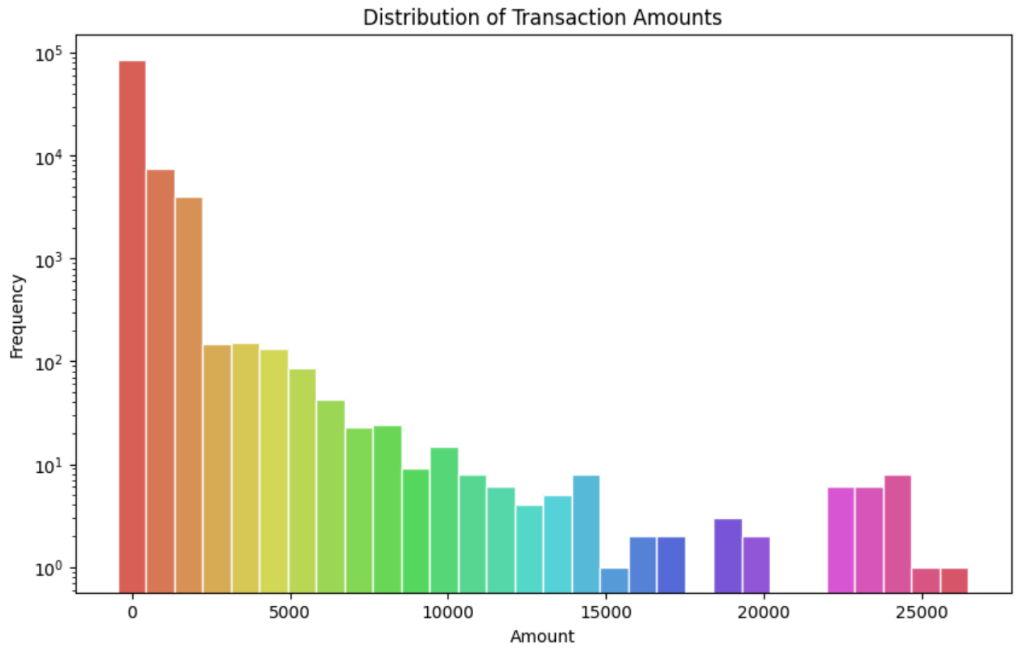


9. **Field Name : Amount**

Description: The actual Amount transacted. The lowest transaction amount is 0.01 and the highest amount is 3102045.53 in this dataset. There are a few outliers in the amount field with very large values, causing the plot to be very complicated and out of range. Hence I am plotting the amount distribution on 99% of the data (amounts ~ 2500) -



On tweaking the plot to show amounts ~25000, I start seeing the outliers i.e Bigger transactions-



10. Field Name : Fraud

Description: Indicates whether the transaction is fraudulent or not. 0 means no fraud, while 1 indicates a fraudulent transaction.