

DQR – PROJECT 2

1. Data Description

The dataset comprises of Credit Card Transaction data.

It has 96753 records and 10 fields – ‘Recnum’, ‘Cardnum’, ‘Date’, ‘Merchnum’, ‘Merch d
escription’, ‘Merch state’, ‘Merch zip’, ‘Transtype’, ‘Amount’, ‘Fraud’.

The data spans across the year 2010.

2. Summary Tables

Numerical Table

Field Name	% Populated	Min	Max	Mean	Stdev	% Zero
Date	100	2010-01-01	2010-12-31	N/A	N/A	0
Amount	100	0.01	3,102,045.53	427.88	10,006.14	0

Categorical Table

Field Name	% Populated	# Unique Values	Most Common Value
Recnum	100	96,573	N/A
Cardnum	100	1,645	5142148452
Merchnum	96.512	13,091	930090121224
Merch Description	100	13,126	GSA-FSS-ADV
Merch state	98.765	227	TN
Merch zip	95.188	4,567	38118
Transtype	100	4	P
Fraud	100	2	0

3. Visualization of Each Field

a. Field Name: Recnum

This is an ordinal unique positive integer which is incremental. This is assigned to each record entry from 1 to 96,753.

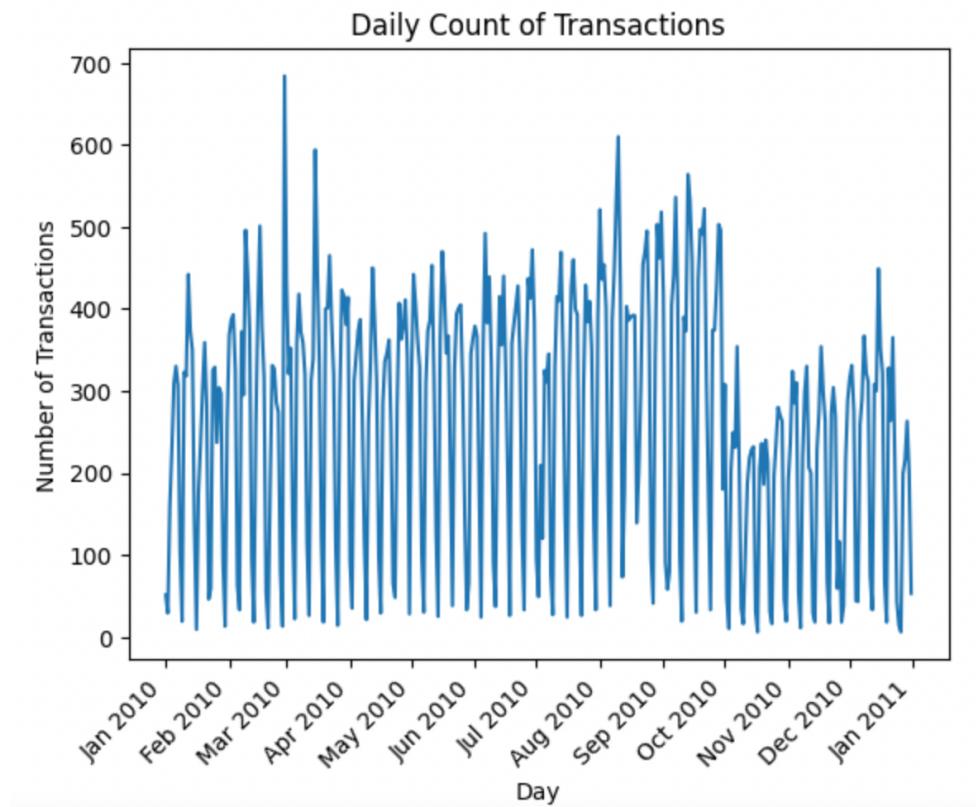
b. Field Name: Date

This is a numerical field which captures the date on which the transactions occurred. The transactions are captured between 2010-01-01 and 2010-12-31. The count of transactions has been captured on a daily, weekly, and monthly basis:

Daily Count of Transactions –

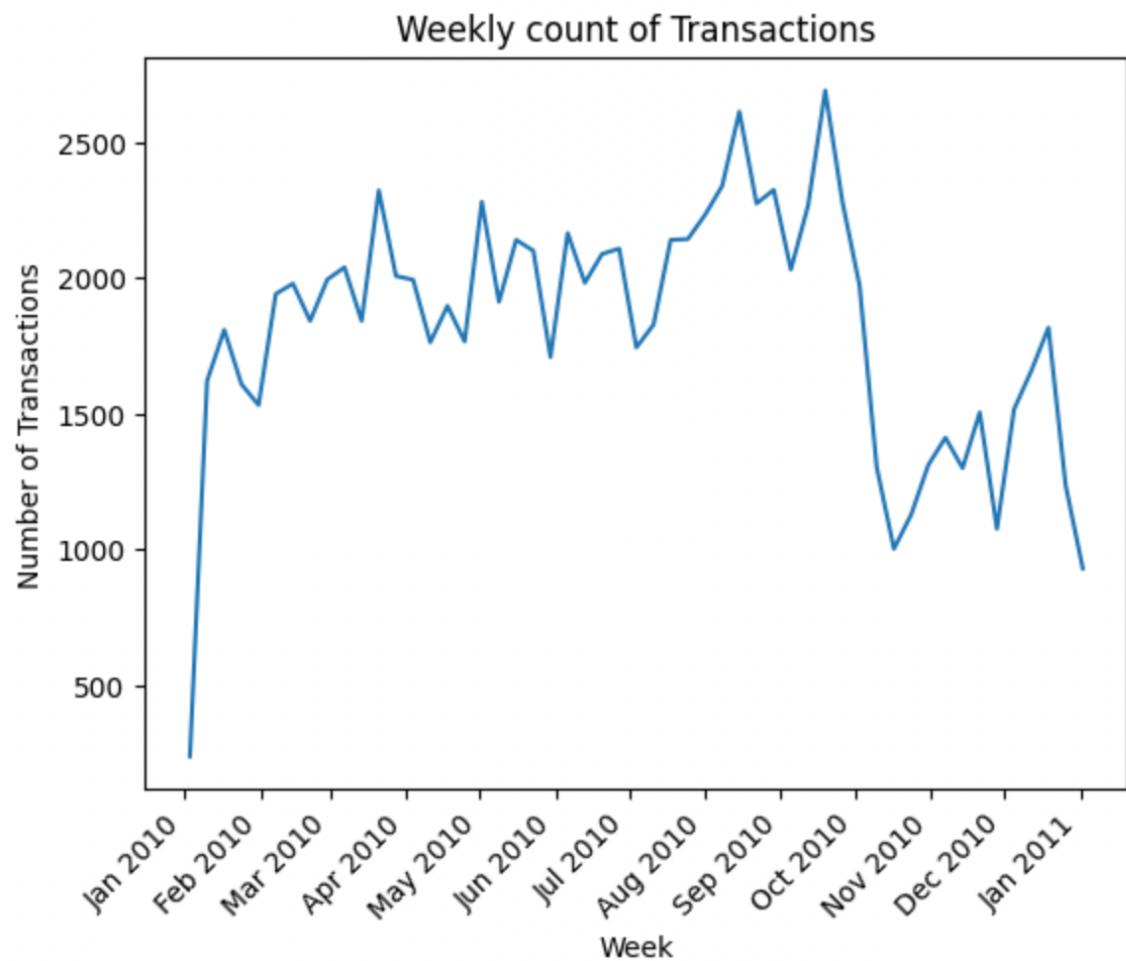
It is seen that on January 1st, 2010, there were 51 transactions and on December 31st, 2010, there were 53 transactions.

The highest number of transactions occurred on 28th February 2010, capturing 684 transactions on that day.



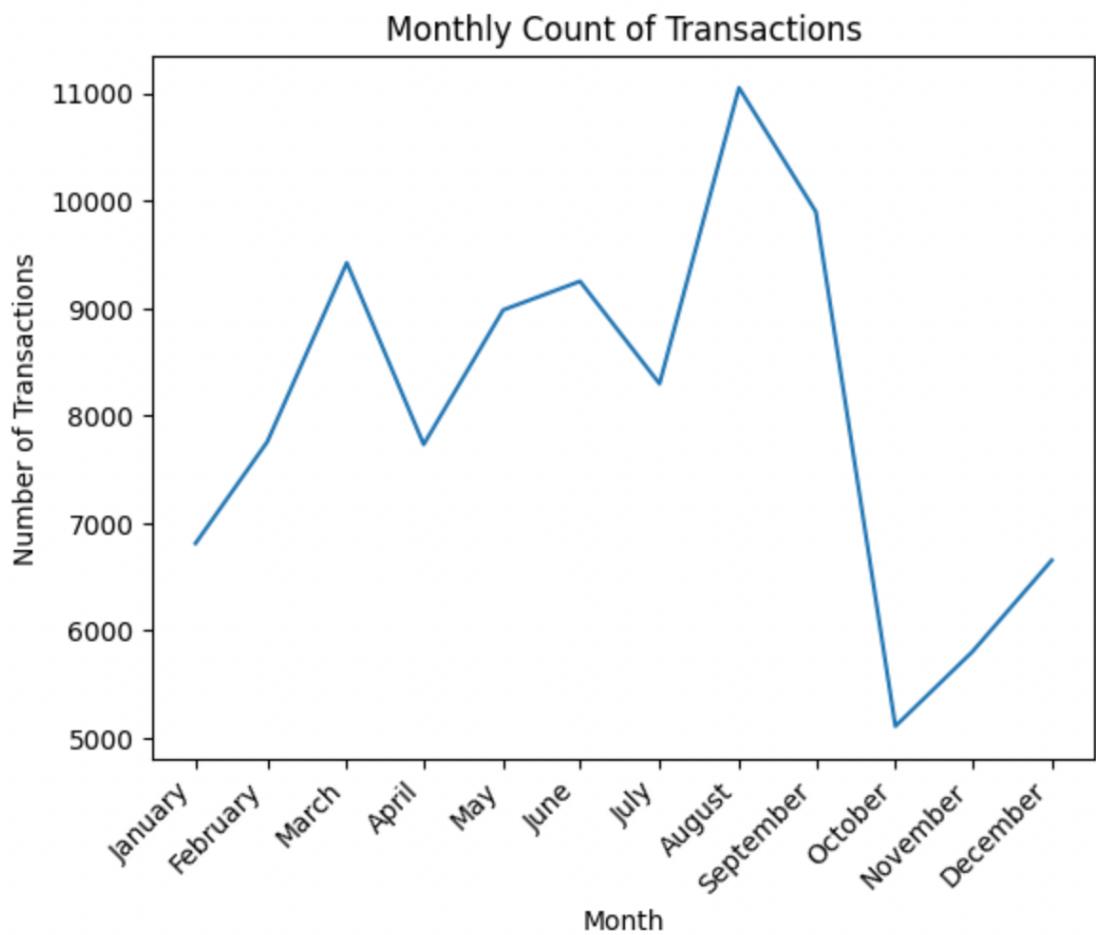
Weekly Count of Transactions –

Week 01(week ending January 3rd, 2010) captures 239 transactions and the last week (week ending February 2nd 2011) captures 931 transactions as seen below.



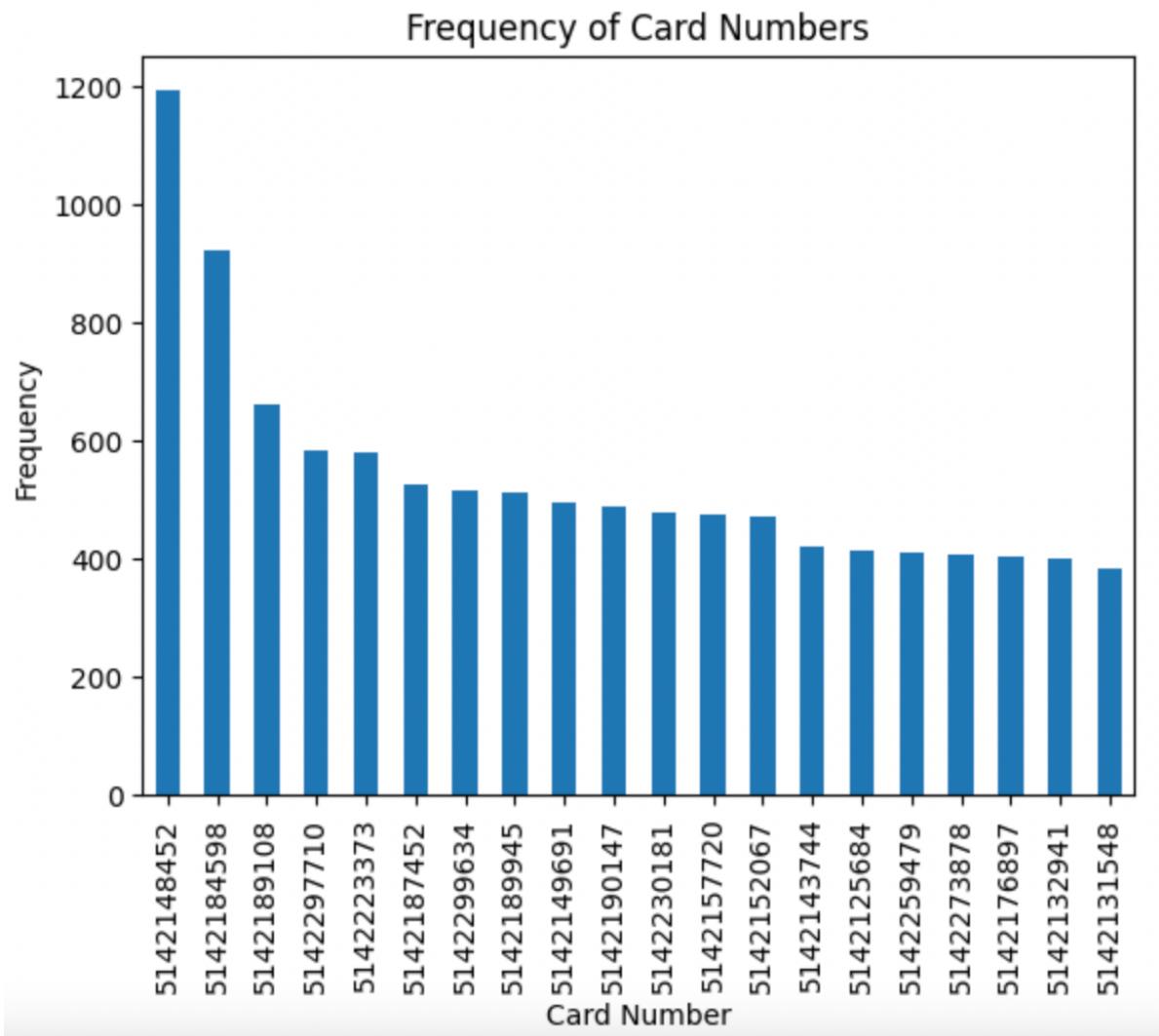
Monthly Count of Transactions –

August is the month where the highest number of transactions (11,050) have been captured. October is the month with the least number of transactions (5,109).



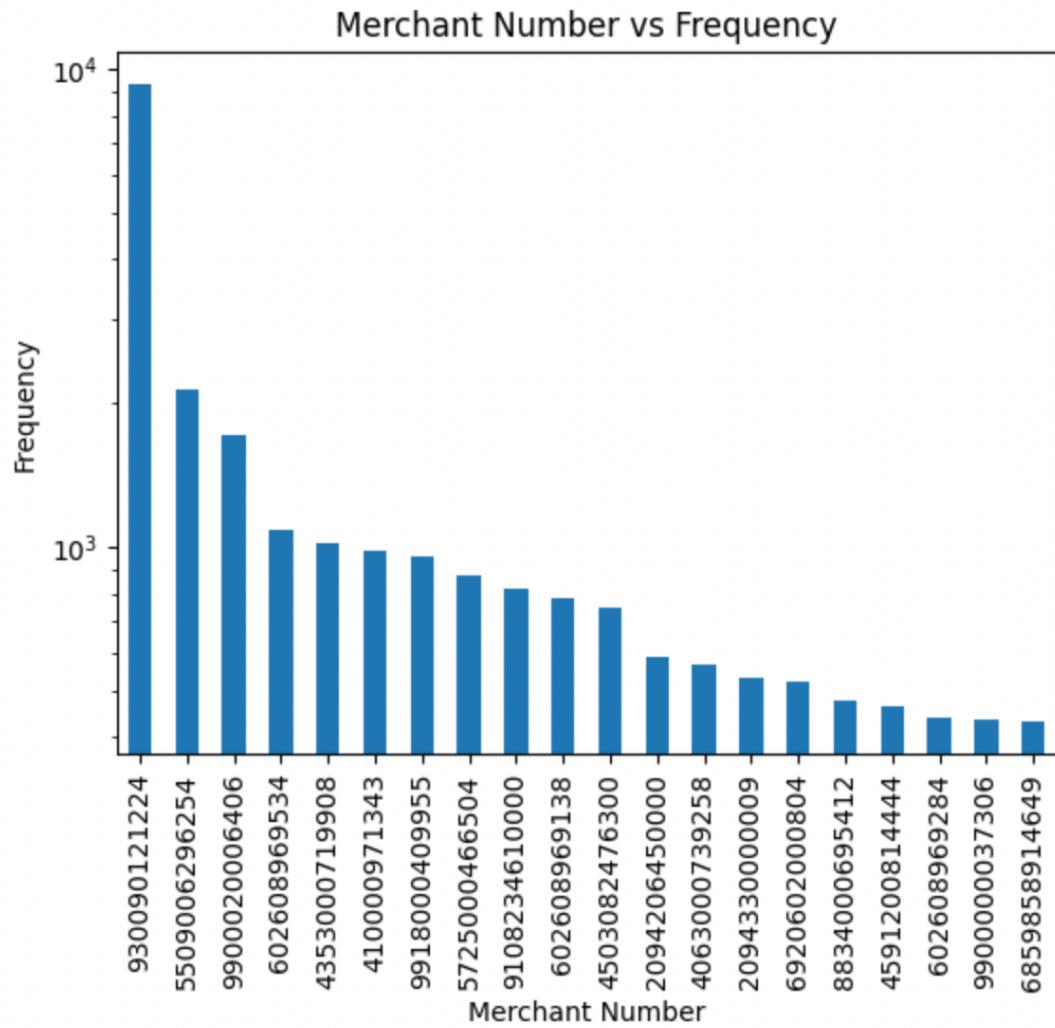
c. Field Name: Cardnum

This is a categorical field which displays the card number used to make each transaction. The card number 5142148452 was used 1192 number of times, making this card number the most used, in the dataset.



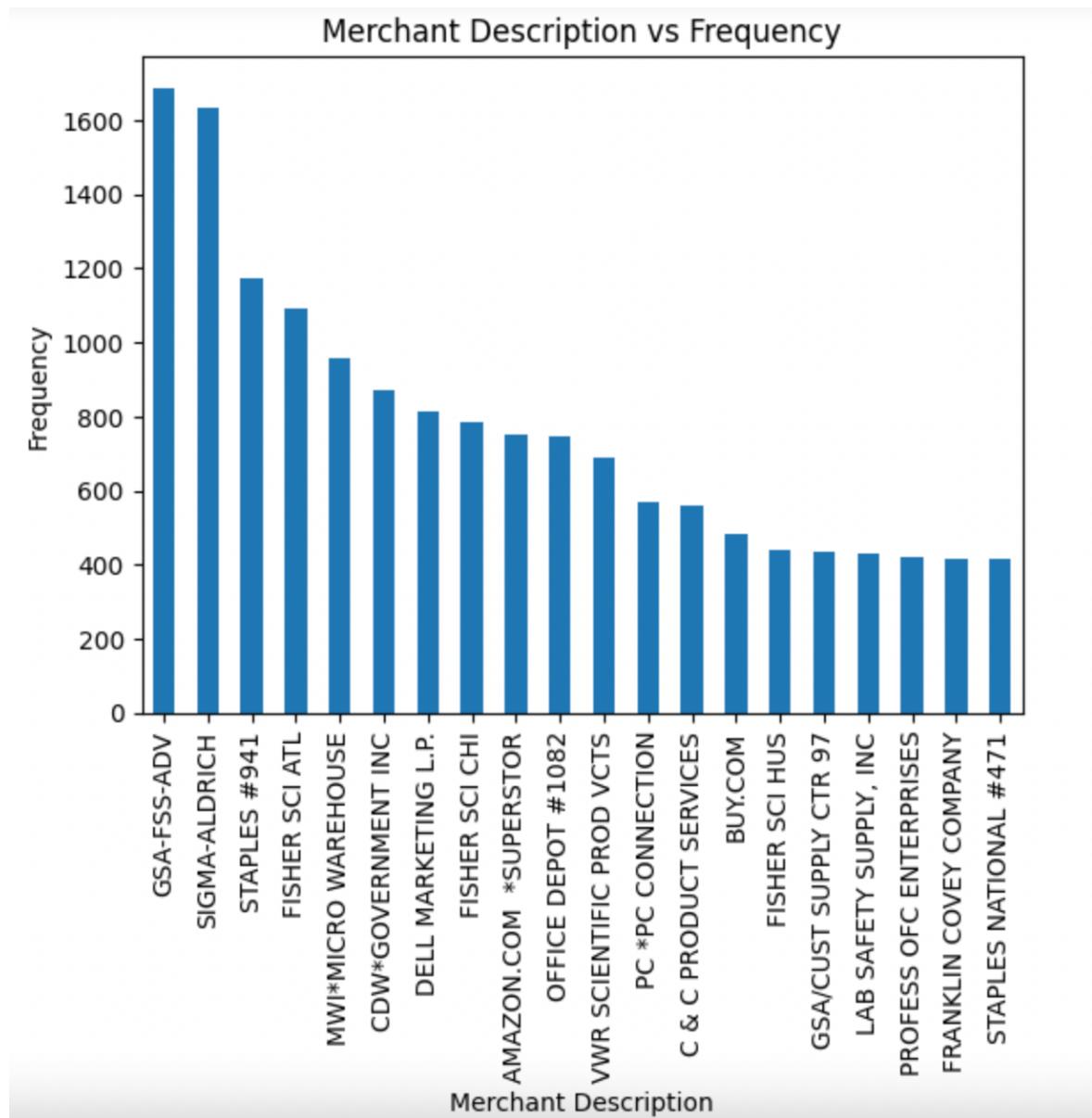
d. Field Name: Merchnum

A merchant number is a code that is issued by credit card processors that tells the banks and credit card processors where the money is going. 930090121224 is the merchnum with the highest number of transactions (9310).



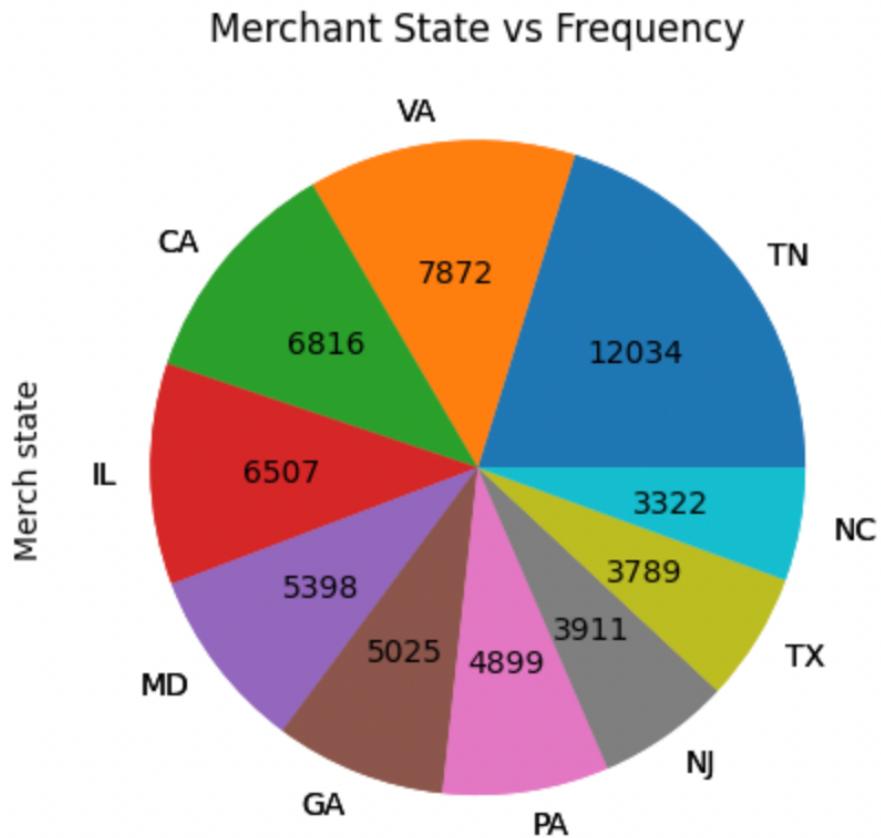
e. Field name: Merchant description

This field describes the merchants, with GSA-FSS-ADV being the merchant description with the highest occurrence (1688).



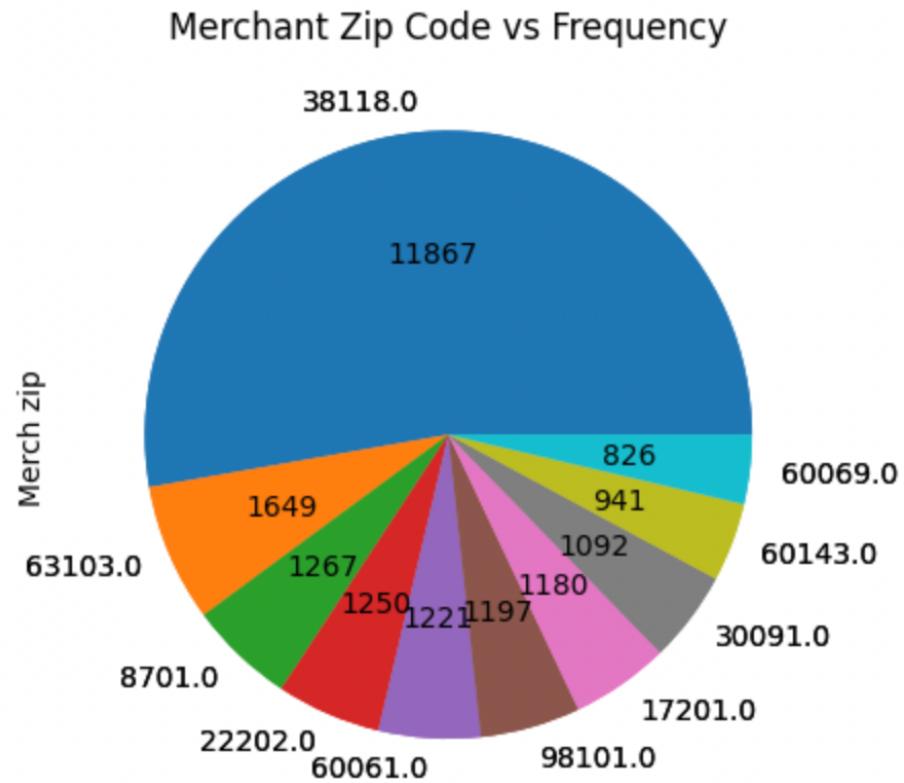
f. Field name: Merch state

Merchant state describes the state where the transactions take place. As seen below, most of the transactions originate in Tennessee (12,034).



g. Field name: Merch zip

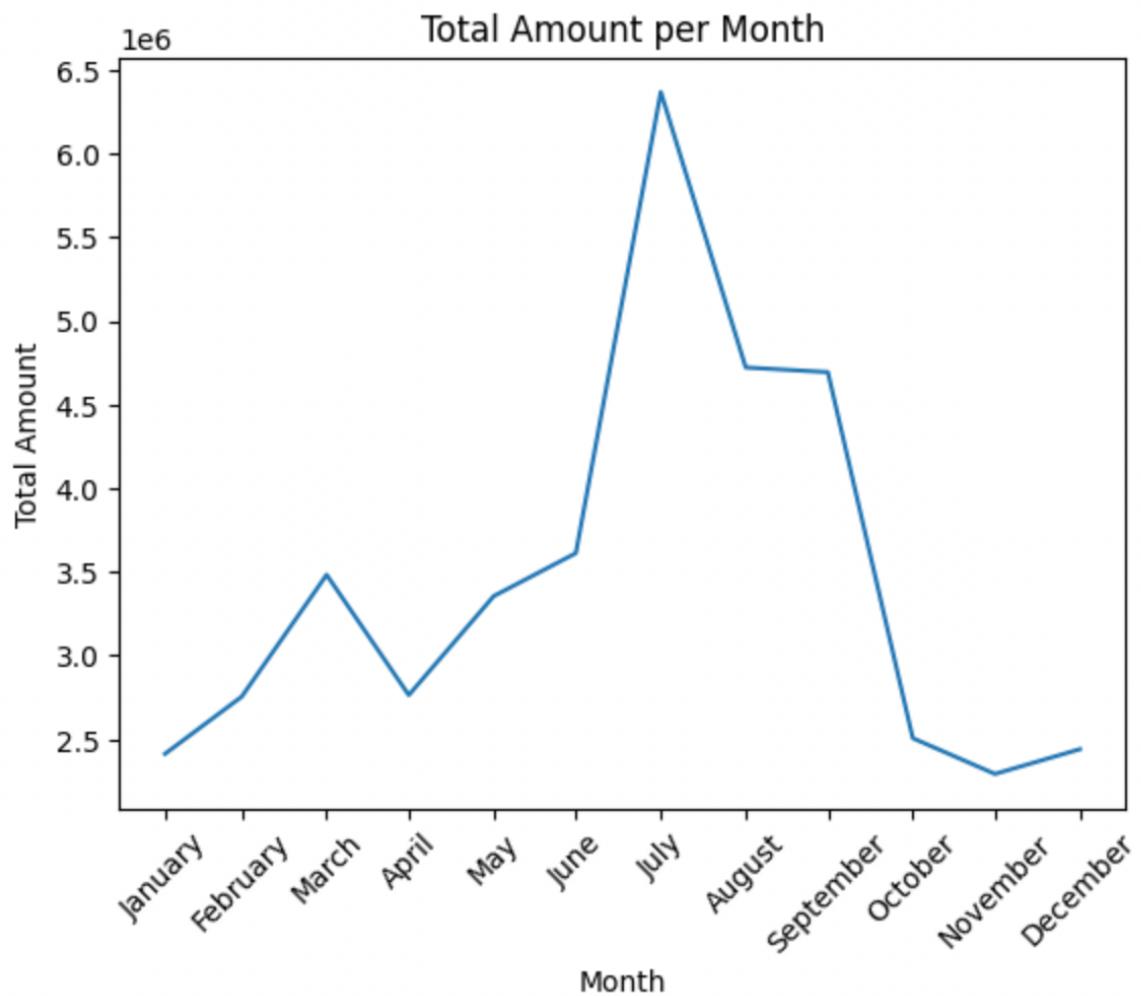
Captures the merchant zip codes. The zip code 38118 has the highest number of transactions (11867).



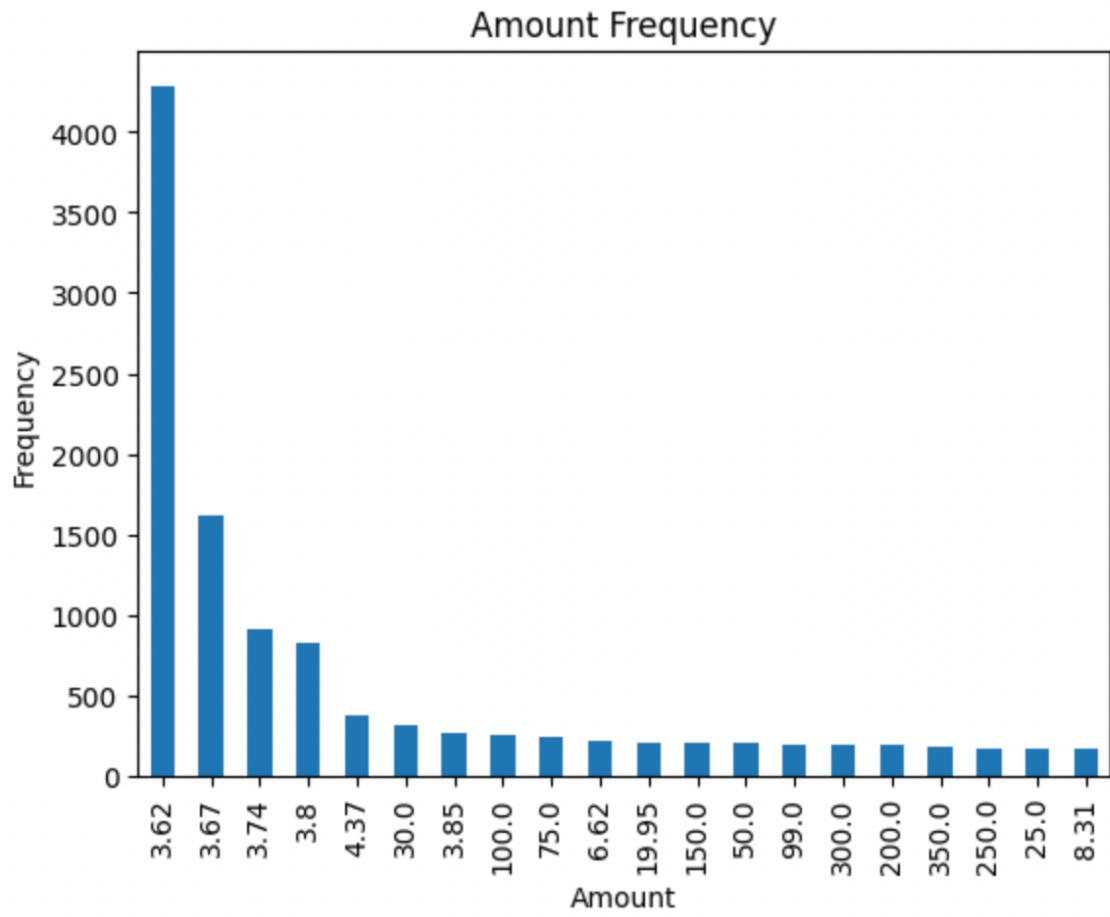
h. Field Name: Amount

This field displays the amount transacted in each of the transactions.

The below plot shows the total amount transacted in each month. July is the month where the largest amount has been transacted (63,67,486).

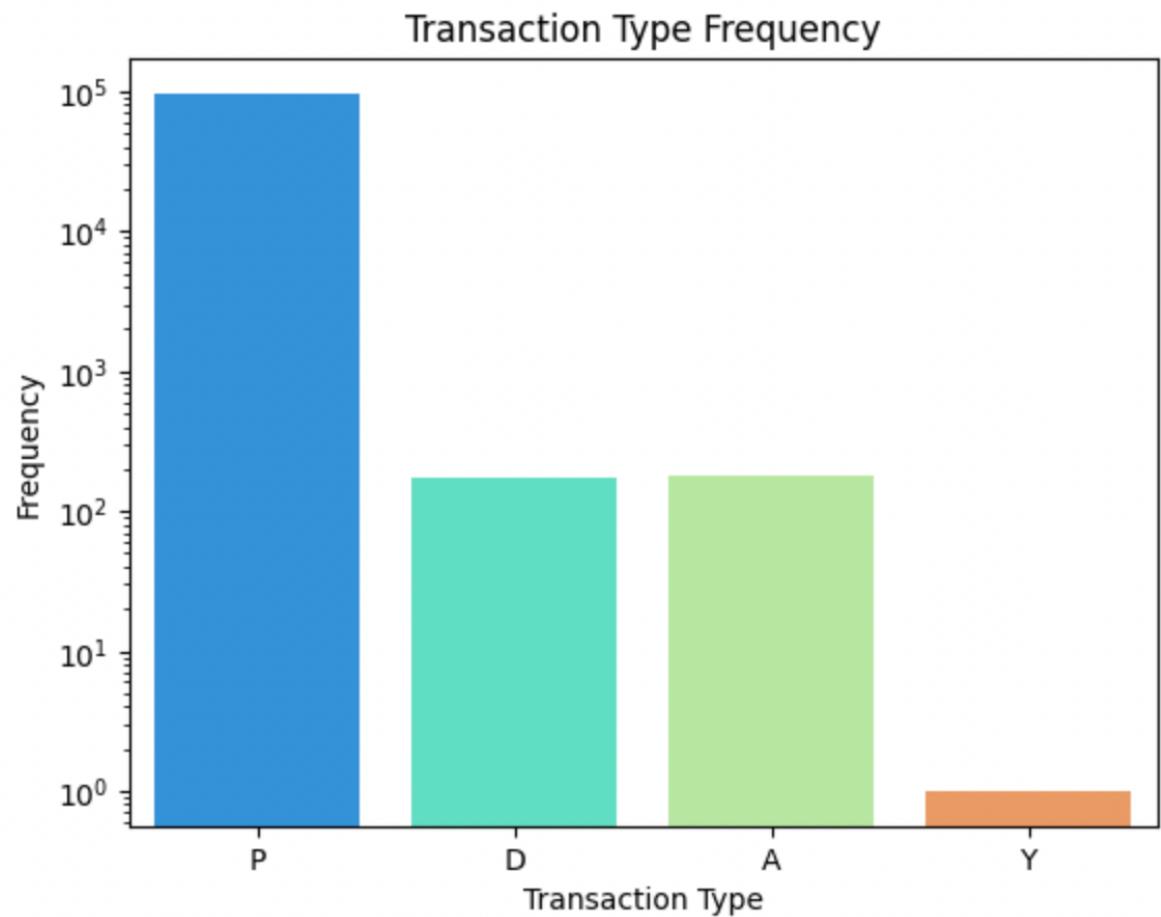


The below plot shows the most common dollar amounts that were transacted. The frequency is depicted in the log scale. The amount which is transacted the most is 3.62, which has occurred 4,283 times.



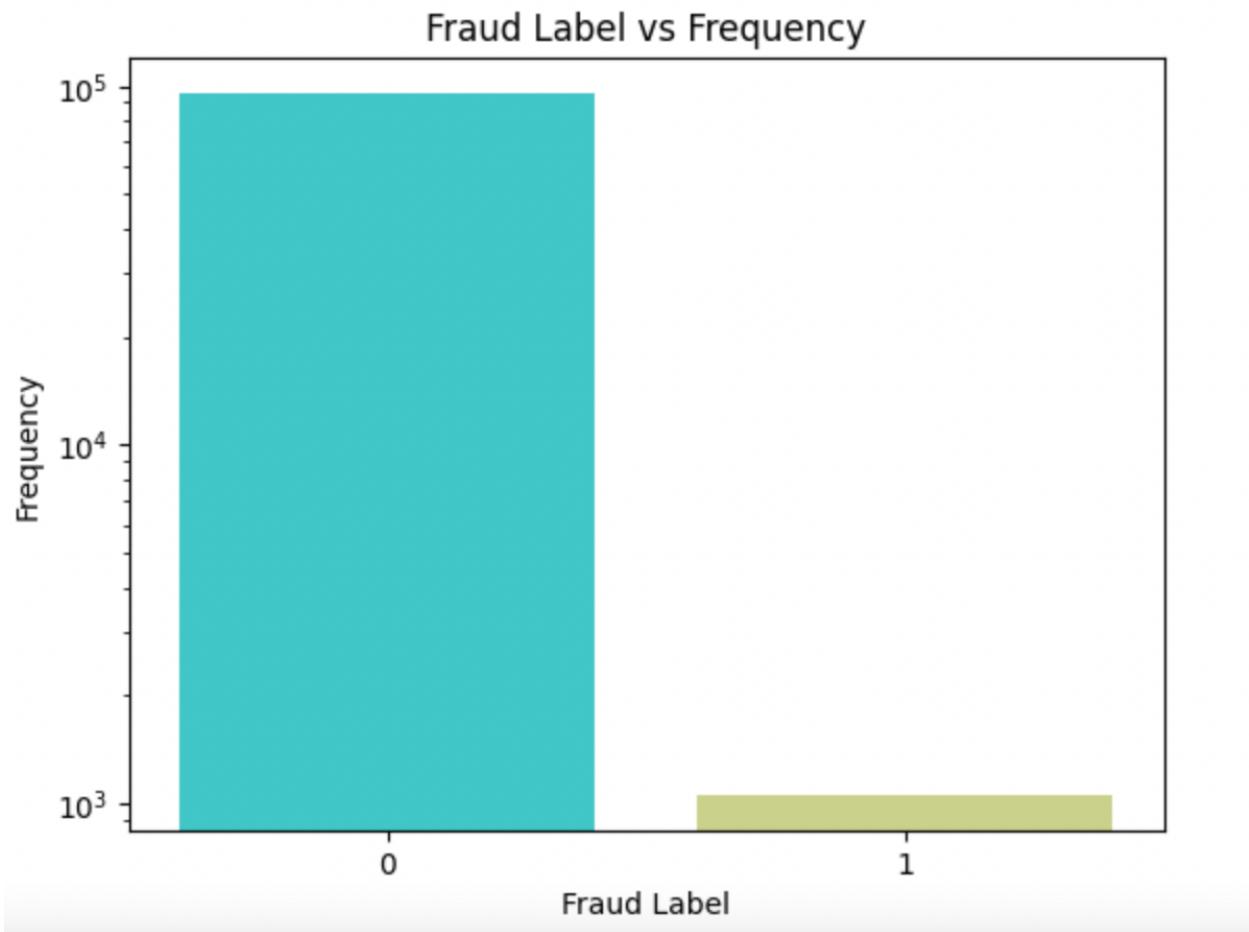
i. Transaction Type

Categorizes the transactions into 4 categories – P, D, A, Y. Most of the transactions fall under the “P” category.



j. **Field Name: Fraud**

This field depicts whether the transaction is categorized as “Fraud” or not. All the transactions with Fraud = 1 are Fraud transactions and the ones with Fraud = 0 are not fraud.



Fraud Label	Count	Percentage %
0	95,694	99%
1	1059	1%