Data Quality Report

1. Data Description

The dataset contains **transaction records from card payments**, capturing a wide array of attributes including transaction amounts, merchant details, and fraud indicators. The data comes from real-world financial transactions over **1 year** and includes both numerical and categorical fields. It contains **10 fields** and **97,852 records** and is designed for analytical exploration and fraud detection model development.

2. Summary Tables

Numeric Fields Table

	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	Min	Max	Mean	Standard Deviation	Most Common
0	Amount	numeric	97852	100.0%	0	0.01	3102045.53	425.466438	9949.8	3.62

Categorical Fields Table

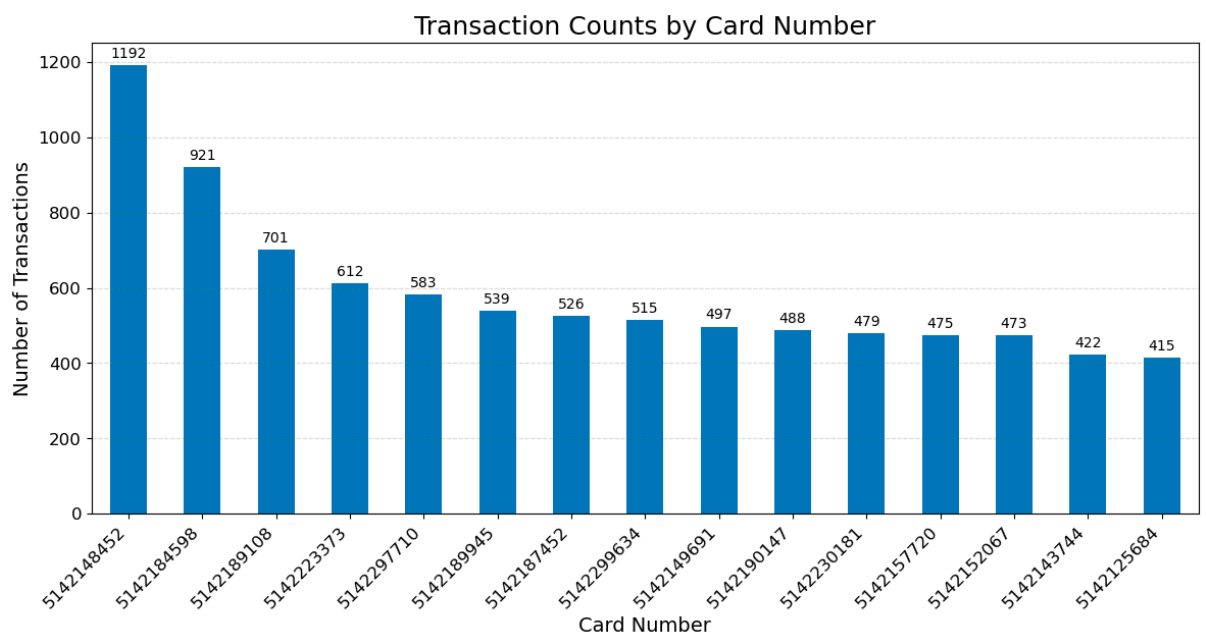|	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	# Unique Values	Most Common
0	Date	categorical	97852	100.0%	0	365	2/28/10
1	Merchnum	categorical	94455	96.5%	0	13091	930090121224
2	Merch description	categorical	97852	100.0%	0	13126	GSA-FSS-ADV
3	Merch state	categorical	96649	98.8%	0	227	TN
4	Transtype	categorical	97852	100.0%	0	4	P
5	Recnum	categorical	97852	100.0%	0	97852	1
6	Fraud	categorical	97852	100.0%	95805	2	0
7	Cardnum	categorical	97852	100.0%	0	1645	5142148452
8	Merch zip	categorical	93149	95.2%	0	4567	38118

3. Visualization of Each Field

1) Field Name: Recnum
Description: Ordinal unique positive integer for each transaction record, from 1 to 97,852.
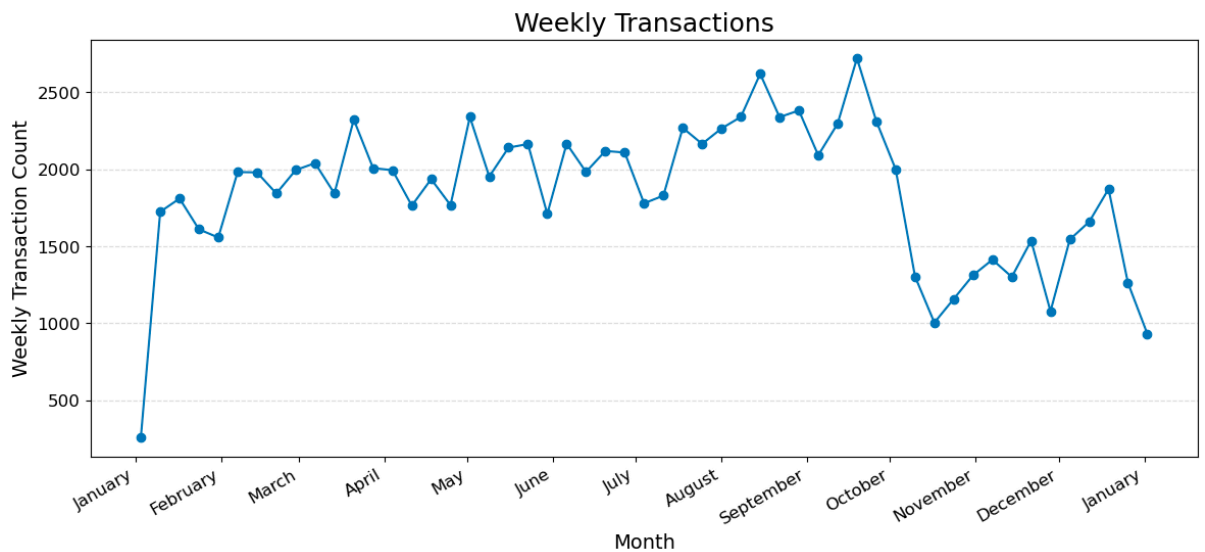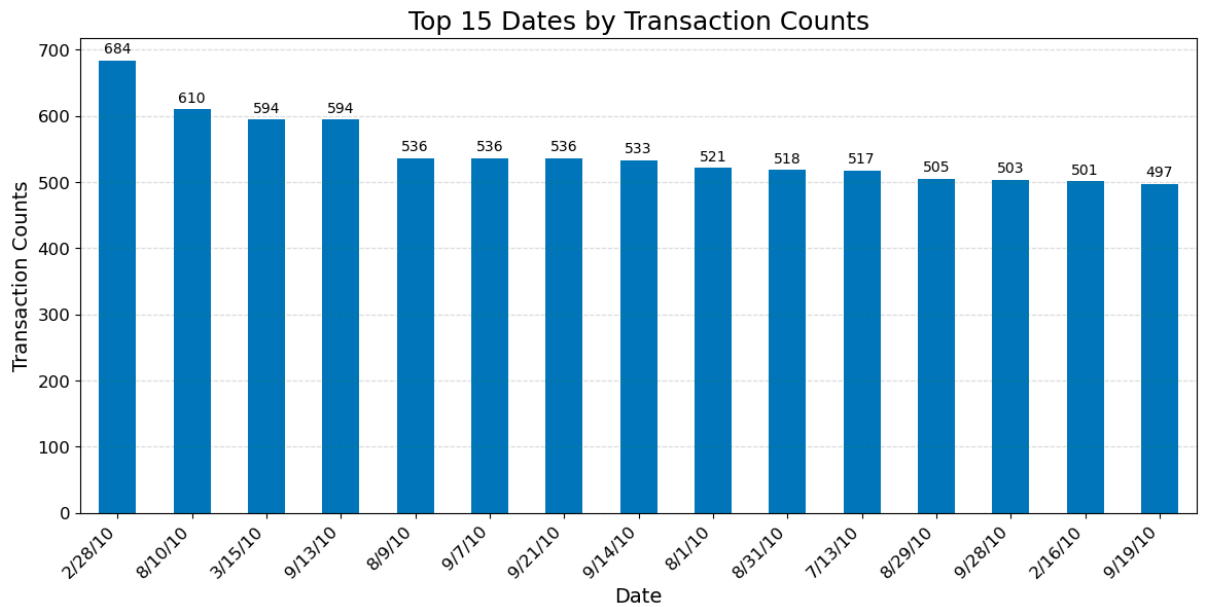
2) Field Name: Cardnum
Description: Applicant's card number. The distribution shows the top 15 field values of card numbers. The most used card number is 5142148452, with a total count of 1,192.

Transaction Counts by Card Number

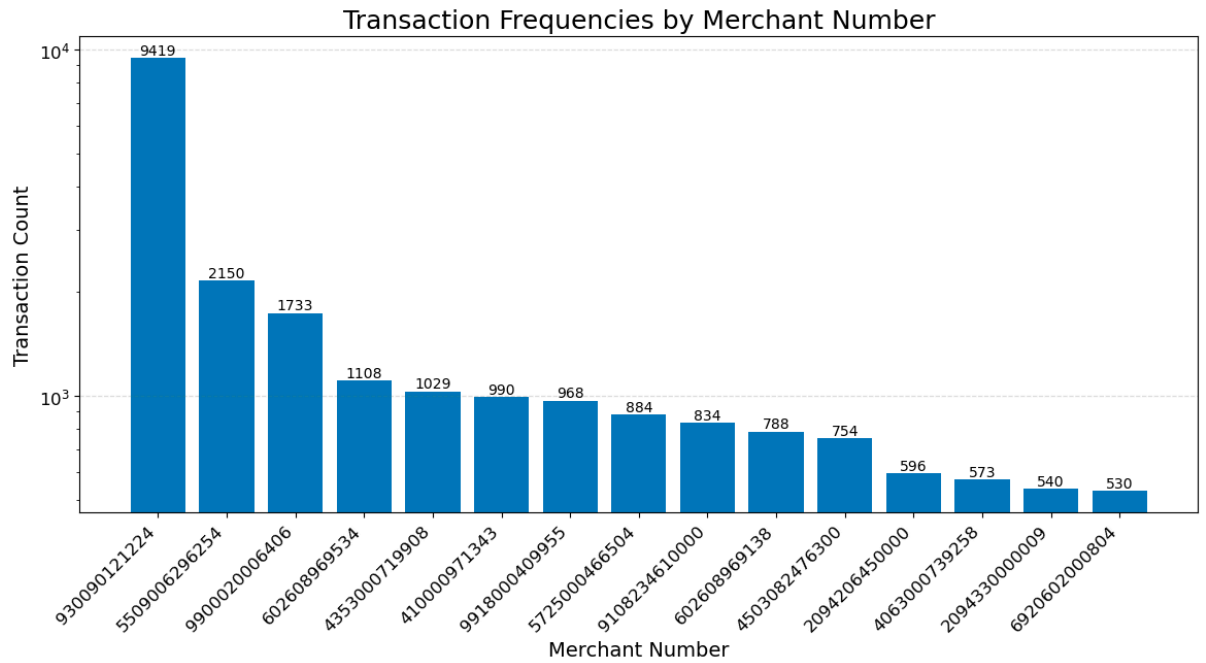| Card Number	Number of Transactions
5142148452	1192
5142184598	921
5142189108	701
5142223373	612
5142297710	583
5142189945	539
5142187452	526
5142299634	515
5142149691	497
5142190147	488
5142230181	479
5142157720	475
5142152067	473
5142143744	422
5142125684	415

3) **Field Name: Date**
   Description: Date of the transaction, spanning a period from 1st January 2010 to 31st December 2010, with the most transactions recorded on 28th February.
   The first graph shows the top 15 dates when transactions occurred. The second graph visualizes the number of weekly transactions over time.
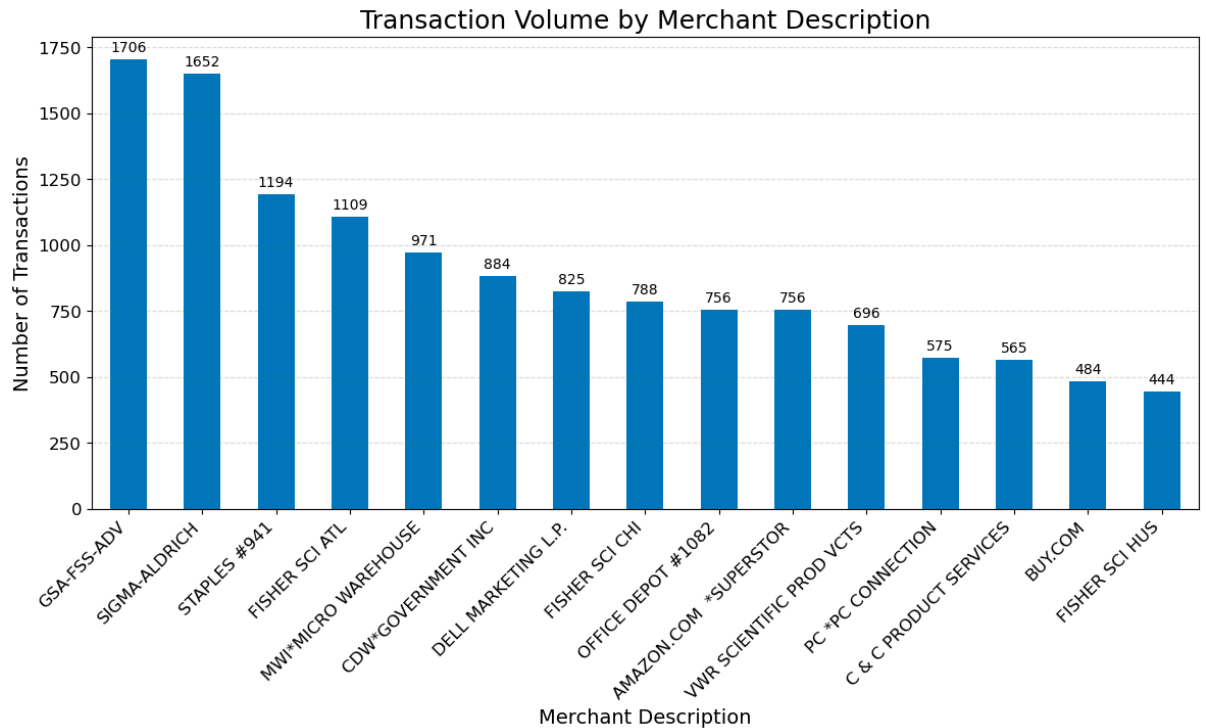

Top 15 Dates by Transaction Counts

Weekly Transactions

4) Field Name: Merchnum

Description: Merchant number. The distribution shows the top 15 merchants that received the most transactions by merchant number, and the most frequent merchant number having 9,419 transactions.
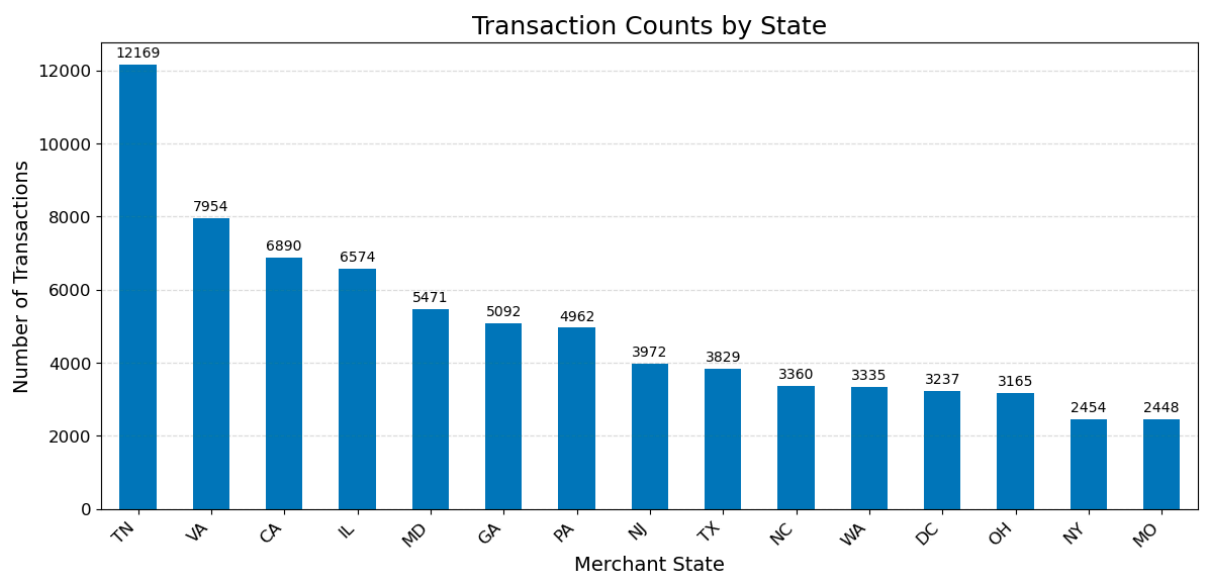
Transaction Frequencies by Merchant Number

A bar chart titled "Transaction Frequencies by Merchant Number" with a logarithmic y-axis labeled "Transaction Count" and x-axis labeled "Merchant Number". The bars show the following values:

Merchant Number	Transaction Count
930090121224	9419
550906296254	2150
990020006406	1733
602608969534	1108
435300071908	1029
41000971343	990
991800040955	968
572500466504	884
910823461000	834
602608969138	788
450308247630	754
209420645000	596
406300073925	573
209433000009	540
692060200804	530

5) Field Name: Merch description

Description: Description of the merchant. The distribution shows the top 15 merchants that received the most transactions by merchant description, with 'GSA-FSS-ADV' appearing most frequently at 1,706 times.
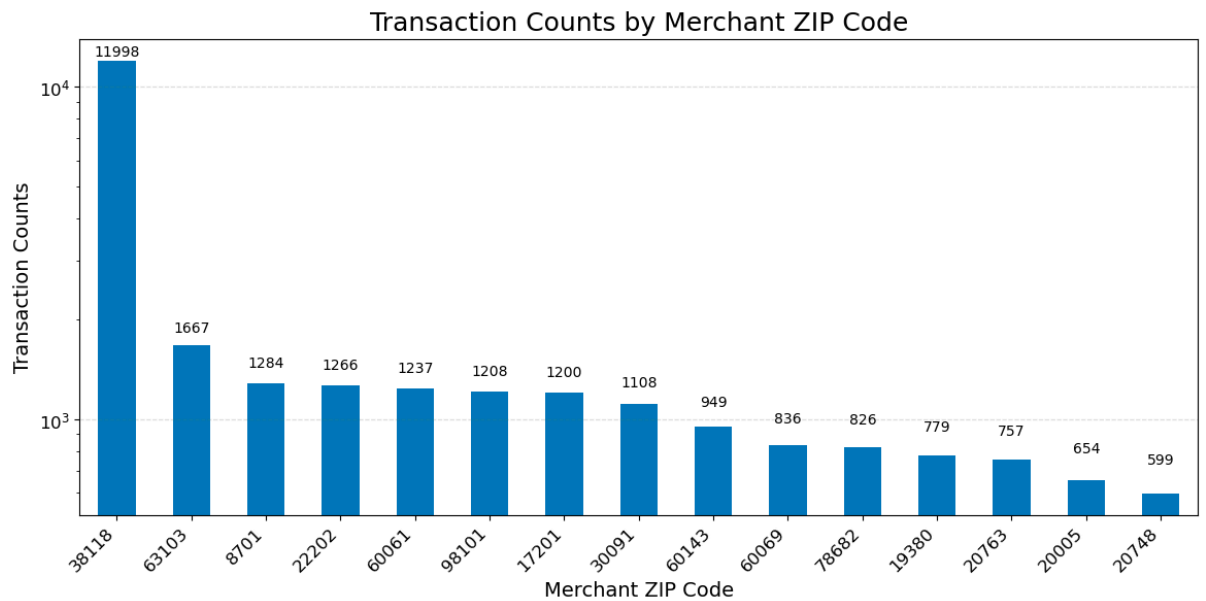
Transaction Volume by Merchant Description

6) Field Name: Merch state

Description: State where the merchant is located, with transactions occurring across 227 different states, most commonly in Tennessee.
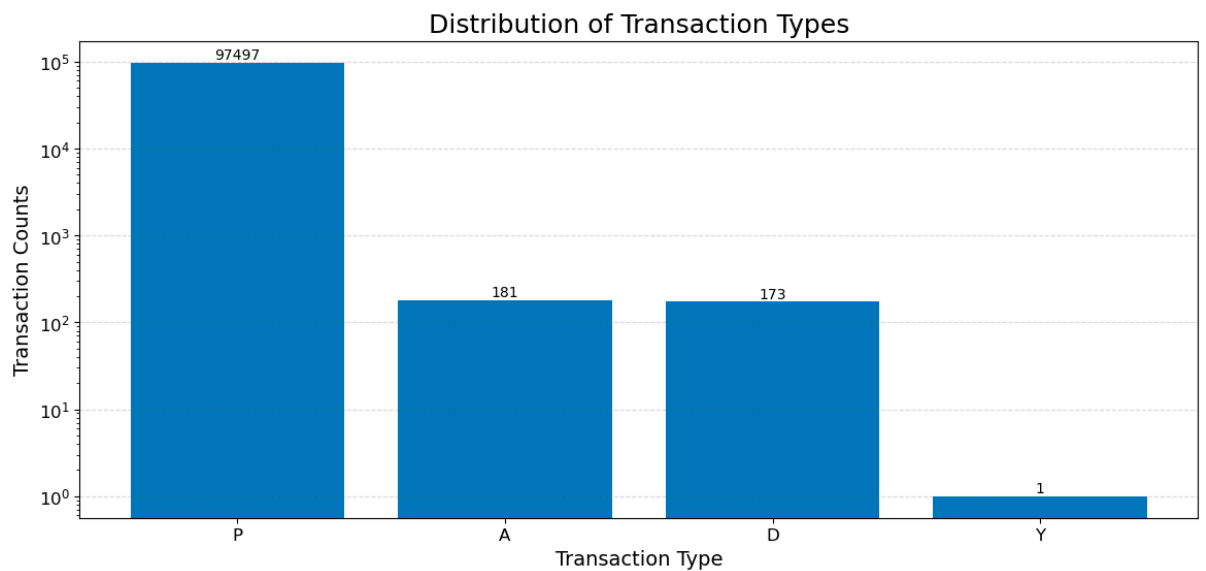
Transaction Counts by State

7) Field Name: Merch zip

Description: ZIP code of the merchant, with 4,567 unique ZIP codes in the dataset, indicating a wide geographical spread.

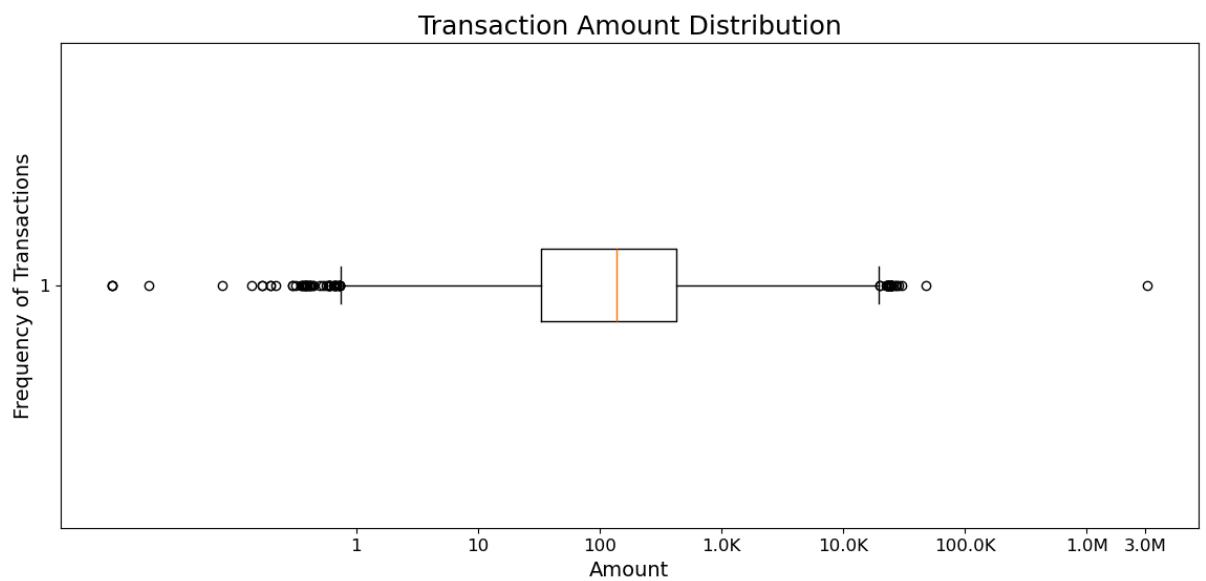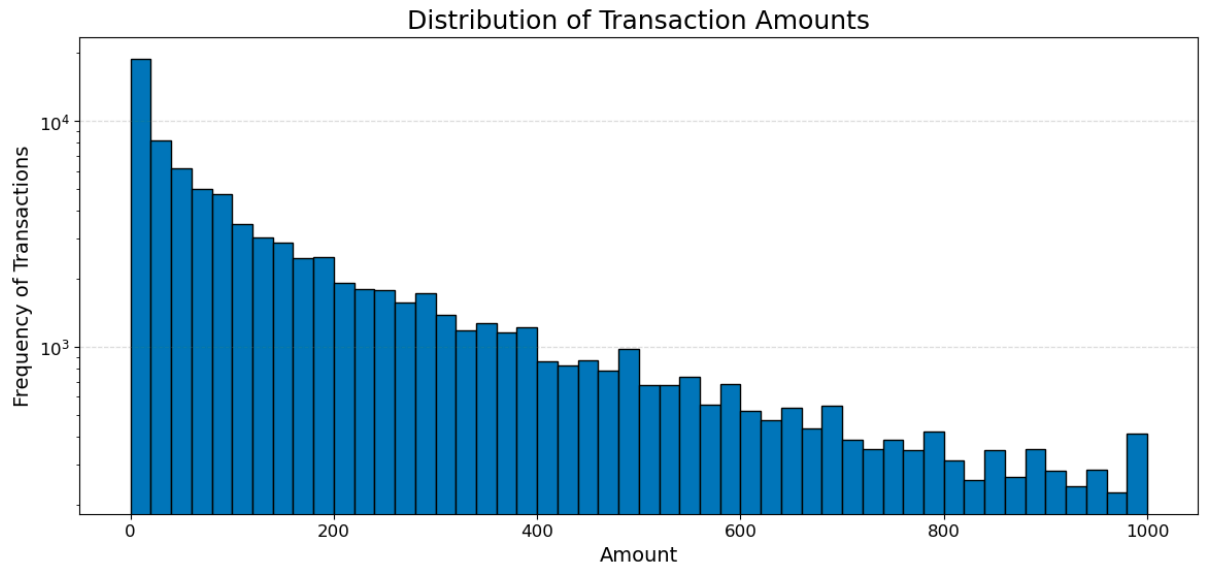

Transaction Counts by Merchant ZIP Code

8) Field Name: Transtype

Description: Type of transactions with four unique types. The 'P' or purchase type is by far the most common, with over 97,000 transaction records.

Distribution of Transaction Types

9) Field Name: Amount

Description: The dollar amount of the transaction, varying widely from as little as $0.01 to over $3 million, showcasing a vast range of transaction values. The first graph shows the distribution of transaction amounts ranging from $0 to $1000 since the maximum number of transactions occurred in this range. The second graph shows a box plot with several high-value outliers.

10) Field Name: Fraud

Description: Fraud identification label. Fraud = 0 (Not fraudulent), Fraud = 1 (Fraudulent). The total count of fraud = 0 is 95,805. The total count of fraud = 1 is 2,047.