Data Quality Report

1. Data Description

The dataset is **Credit Card Transaction Data**, which contains real credit card transaction records from a US government organization for business purposes. The data came from the website of author Mark Nigrini over a **one-year period** from January 1, 2010, to December 31, 2010. There are **10 fields** and a total of **96,753 records**.

2. Summary Tables

Numeric Fields Table

Field Name	# Records Have Values	% Populated	# Zeros	Min	Max	Mean	Most Common	Stdev
Date	96,753	100.00%	0	1/1/10	12/31/10	6/25/10	2/28/10	98 days 21:38
Amount	96,753	100.00%	0	0.01	3102045.53	427.88	3.62	10006.14

Categorical Fields Table

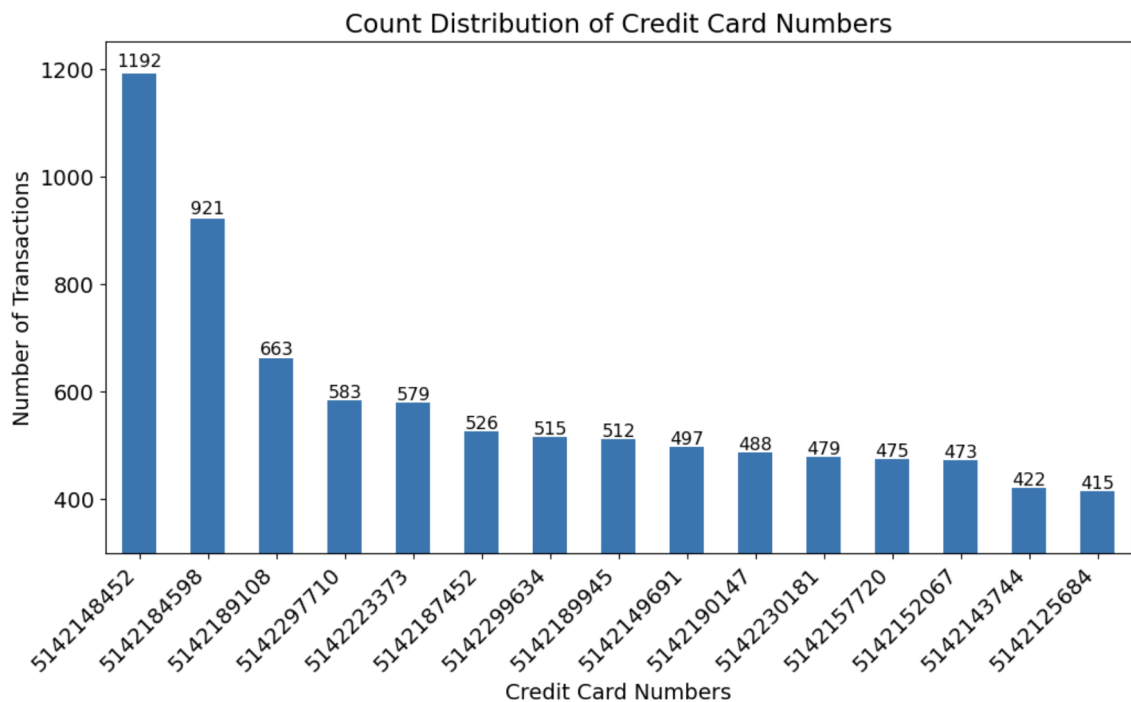| Field Name	# Records Have Values	% Populated	# Zeros	# Unique Values	Most Common
Recnum	96,753	100.00%	0	96,753	1
Cardnum	96,753	100.00%	0	1,645	5142148452
Merchnum	93,378	96.51%	0	13,091	930090121224
Merch description	96,753	100.00%	0	13,126	GSA-FSS-ADV
Merch state	95,558	98.76%	0	227	TN
Merch zip	92,097	95.19%	0	4,567	38118
Transtype	96,753	100.00%	0	4	P
Fraud	96,753	100.00%	95,694	2	0

3. Visualization of Each Field

1) **Field Name: Recnum**

   Description: Record Number. A unique identifier for each transaction record, sequentially from 0 to 96,752.
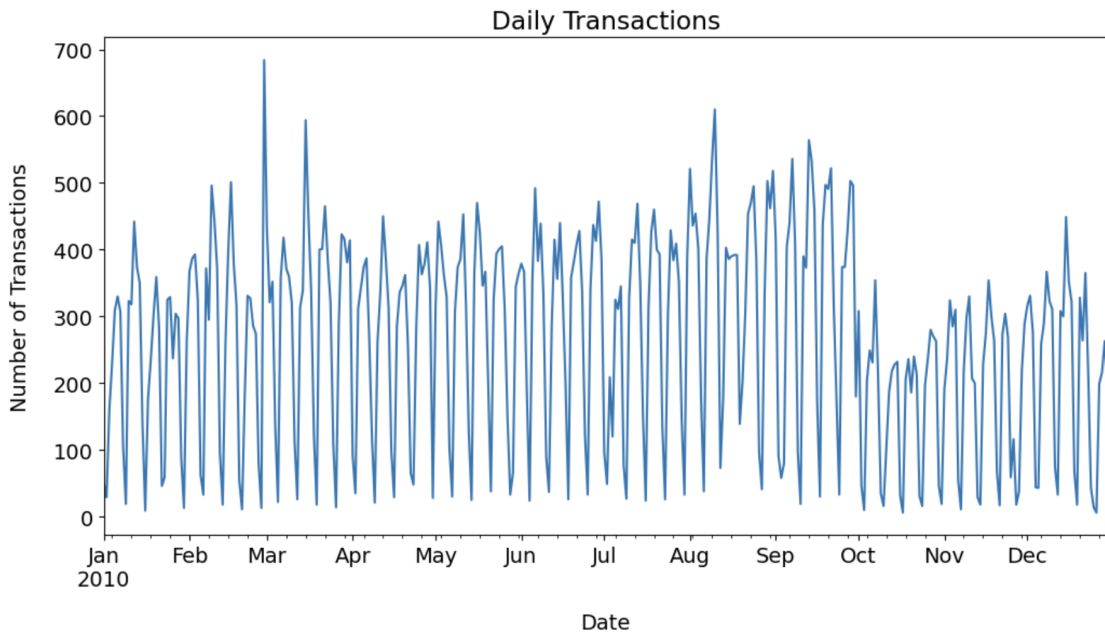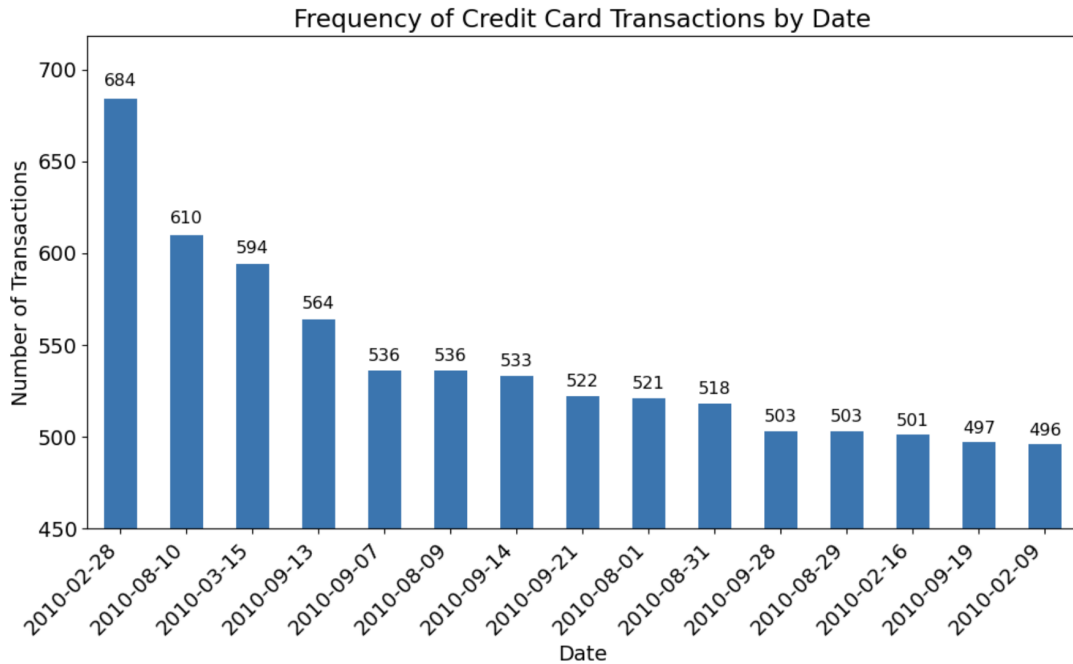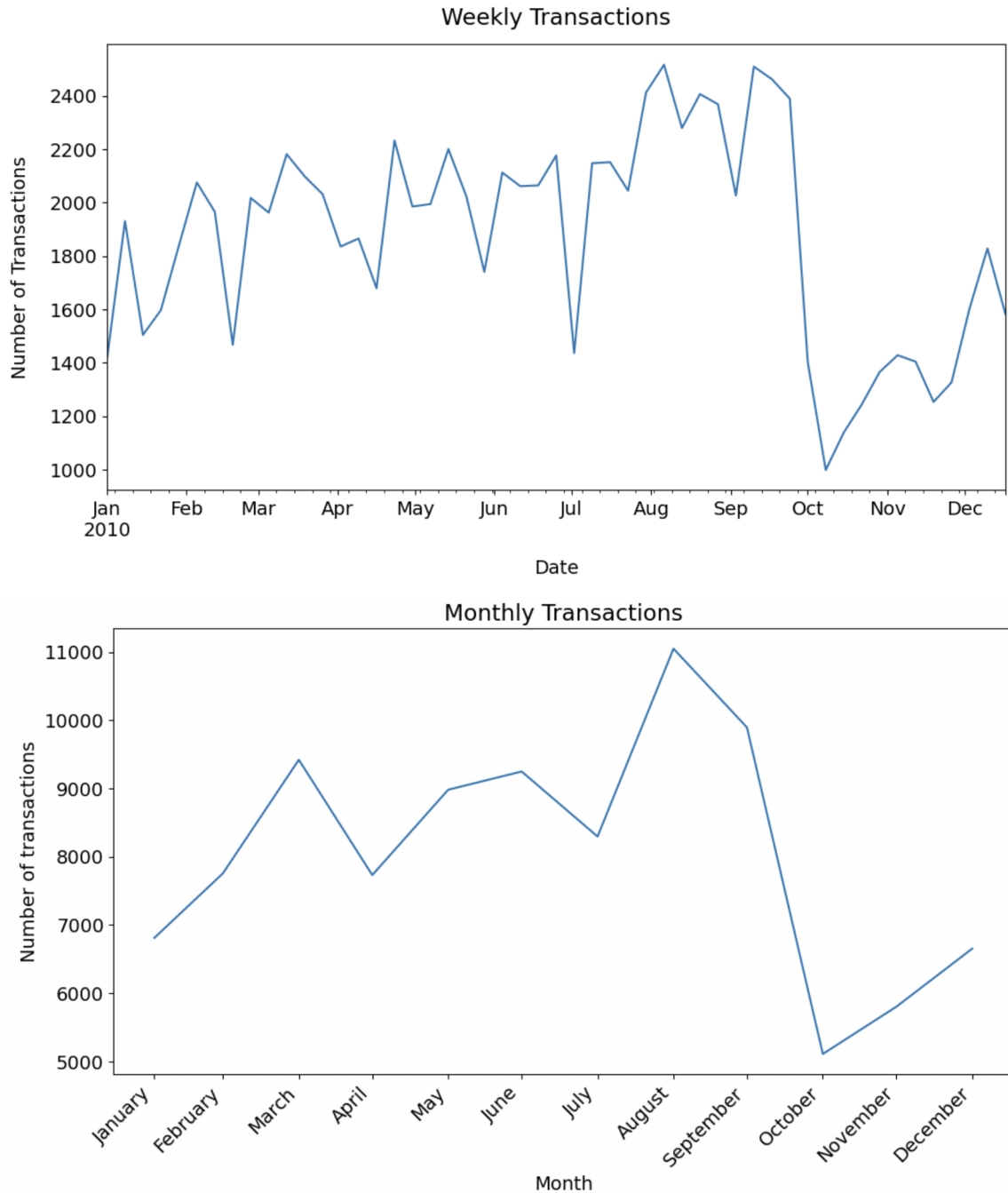
2) **Field Name: Cardnum**

   Description: Credit Card Number. 1,645 distinct values. The distribution shows the top 15 field values of Cardnum. The most common card number is 5142148452, with a total count of 1,192.
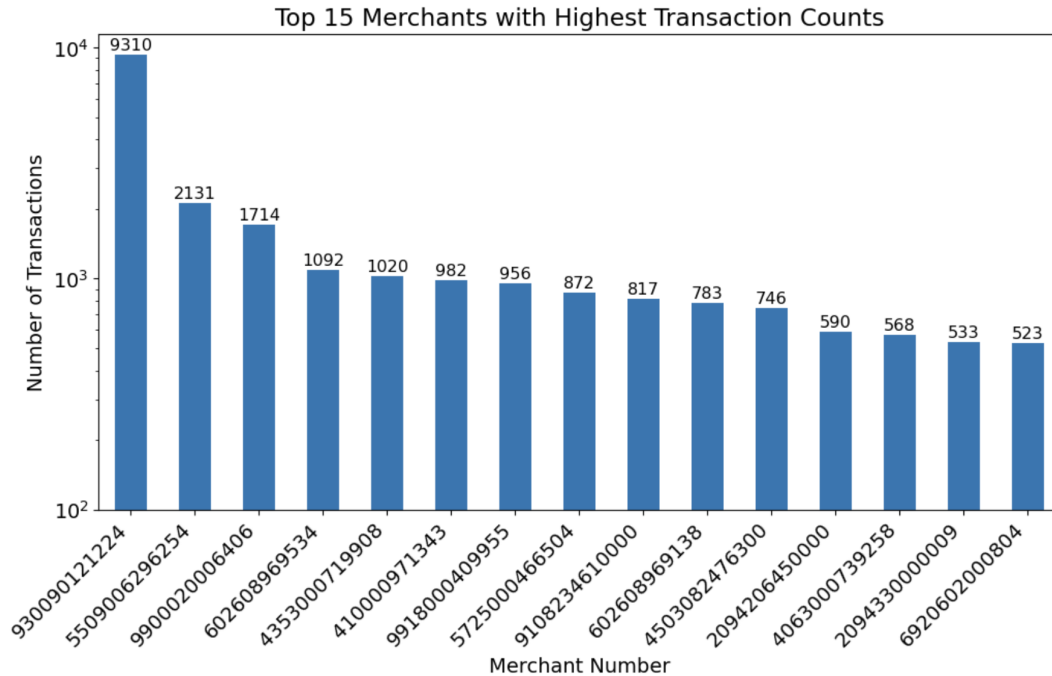
3) Field Name: Date

Description: Date on which transaction took place. The first distribution shows the maximum number of transactions (684) took place on February 28, 2010. The subsequent distributions show the trend of daily transactions, weekly transactions and monthly transactions across time.
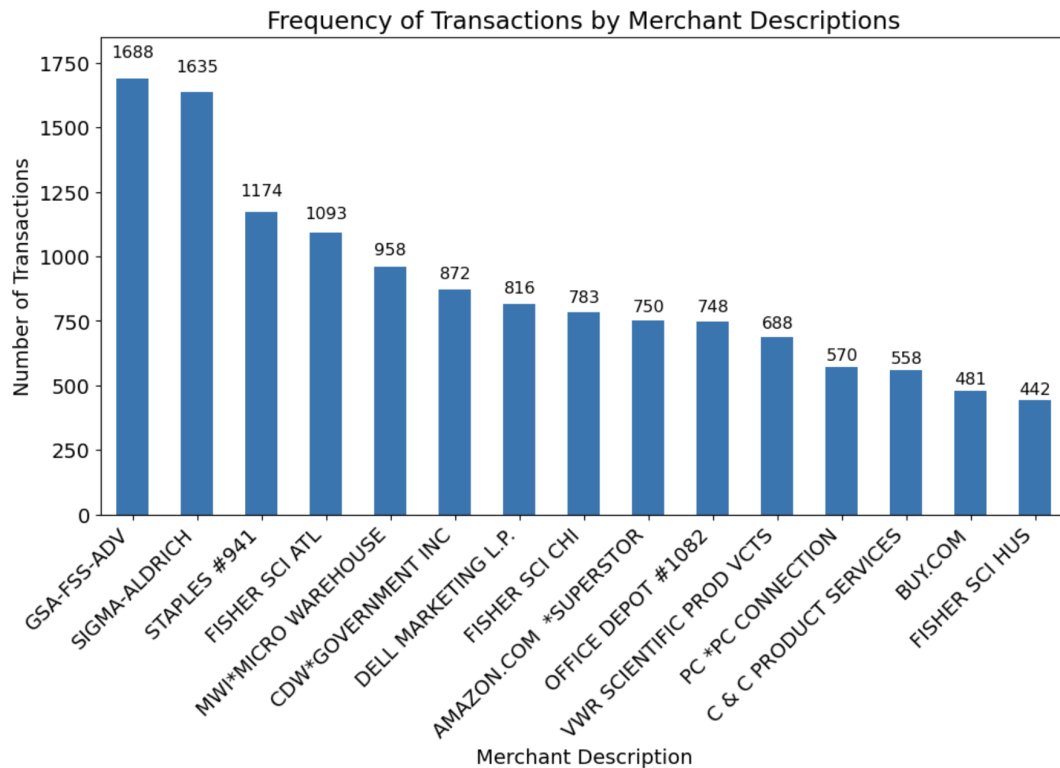
Weekly Transactions

Monthly Transactions

4) Field Name: Merchnum

Description: Merchant Number. The distribution shows the top 15 field values of Merchnum with highest transaction count. The most common merchant number is 930090121224, with a total count of 9,310.
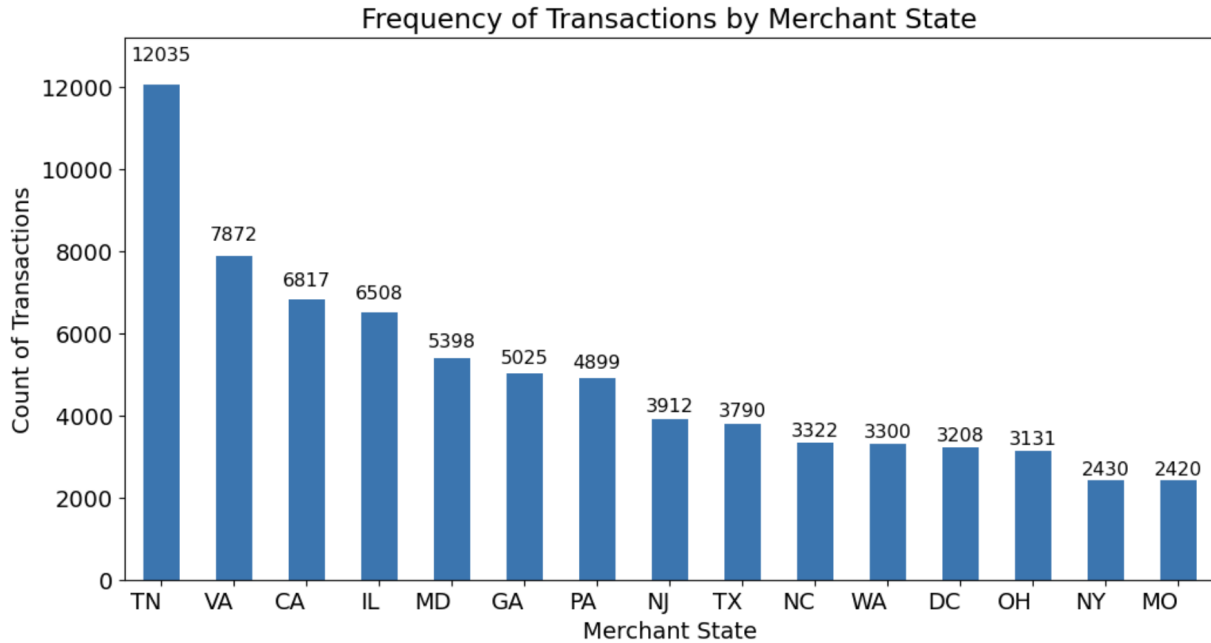
Top 15 Merchants with Highest Transaction Counts

5) Field Name: Merch description

Description: Merchant's Description. The distribution shows the top 15 field values of Merchant's Description. The most common last name is 'GSA-FSS-ADV', with a total count of 1,688.

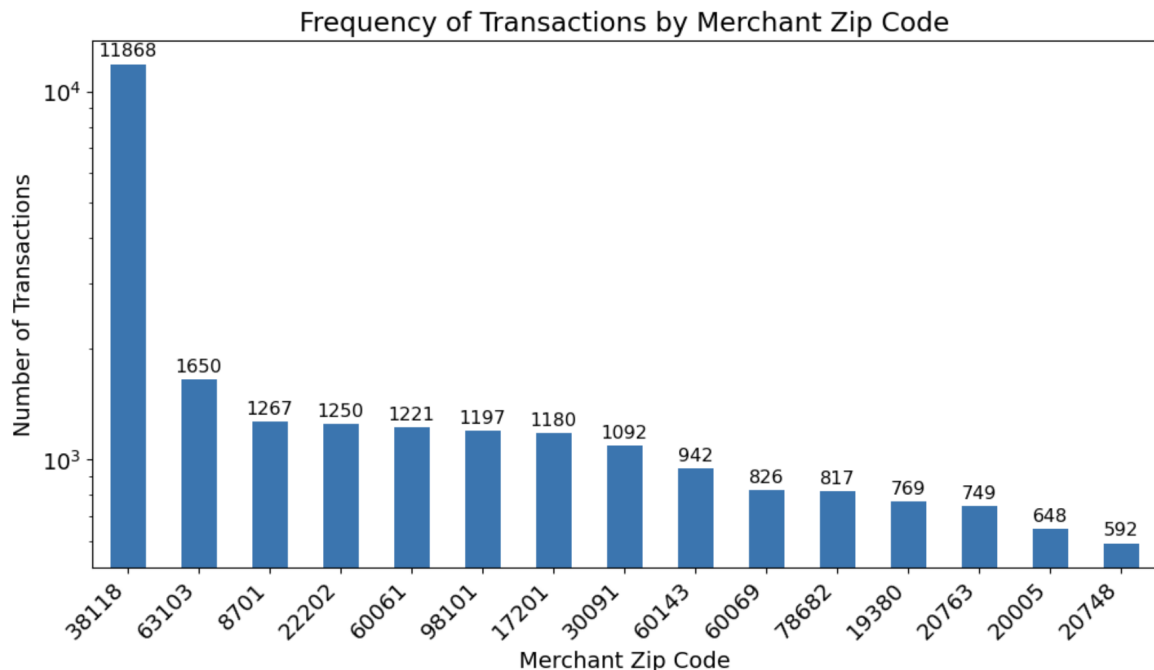Frequency of Transactions by Merchant Descriptions

6) Field Name: Merch state

Description: Merchant's State. The distribution shows the top 15 field values of Merchant's State. The most common address is 'TN', with a total count of 12,035.

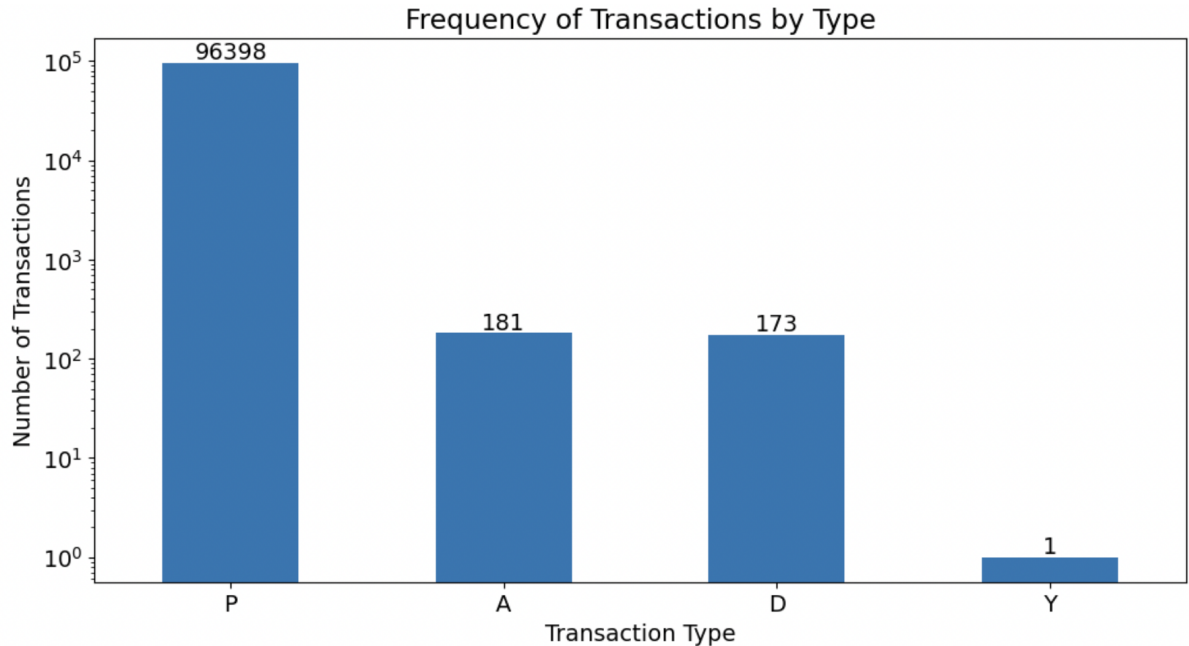Frequency of Transactions by Merchant State

7) Field Name: Merch zip

Description: Merchant's zip code. The distribution shows the top 15 field values of a merchant's zip code. The most common zip code is 38118, with a total count of 11,868.

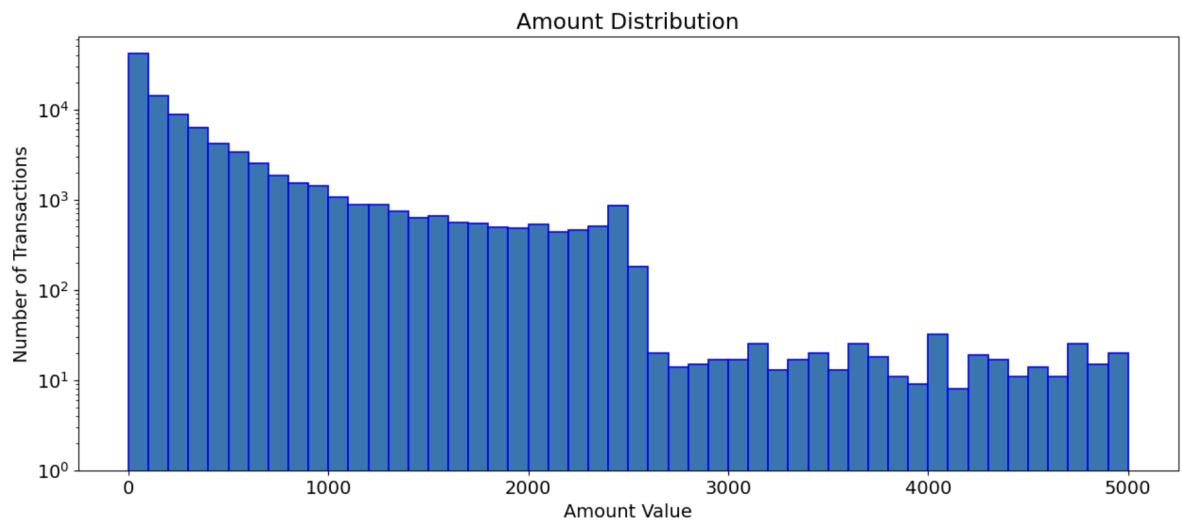Frequency of Transactions by Merchant Zip Code
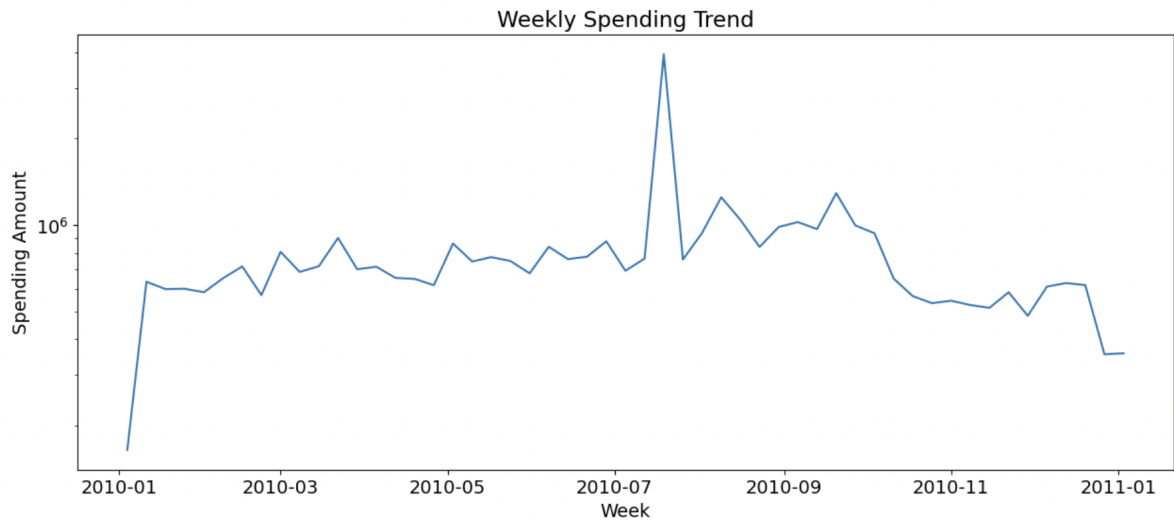
8) Field Name: Transtype

Description: Transaction Type. The distribution shows the top 15 field values of transaction types. The most common transaction type is P, with a total count of 96,398.

9) Field Name: Amount

Description: Amount of Transaction Value in dollars. The first distribution shows the number of transactions vs amount. The second distribution shows the weekly spending trend.

Weekly Spending Trend

10) Field Name: Fraud

Description: Fraud identification label. Fraud = 0 (Not fraudulent), Fraud =1 (Fraud identified). The total count fraud = 0 is 95,694. The total count of fraud = 1 is 1059. The other two distributions show fraudulent transactions trend across days and card numbers.

Fraudulent Transactions by Date

Fraudulent Transactions by Card