

Data Quality Report

1. Data Description

The dataset is **Credit Card Transaction Data**, which contains **Records of Purchase Information via credit cards**. It includes 97852 transactions in 2010, nine fields (two Numerical and eight Categorical), and one label - Fraud.

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	9949.80	3.62
1	Fraud	numeric	97852	100.0%	95805	0.00	1.00	0.020919	0.14	0.00

Temporal Field Table

	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	Most Early	Most Late	Most Common
0	Date	datetime	97852	100.0%	0	2010-01-01	2010-12-31	2010-02-28

Categorical Fields Table

	Field Name	Field Type	# Records Have Values	% Populated	# Zeros	# Unique Values	Most Common
0	Recnum	categorical	97852	100.0%	0	97852	1
1	Cardnum	categorical	97852	100.0%	0	1645	5142148452
2	Merchnum	categorical	94455	96.5%	0	13091	930090121224
3	Merch description	categorical	97852	100.0%	0	13126	GSA-FSS-ADV
4	Merch state	categorical	96649	98.8%	0	227	TN
5	Merch zip	categorical	93149	95.2%	0	4567	38118.0
6	Transtype	categorical	97852	100.0%	0	4	P

3. Visualization of Each Field

Numeric:

1) **Field Name: Amount** (Positive Continuous Numeric Data)

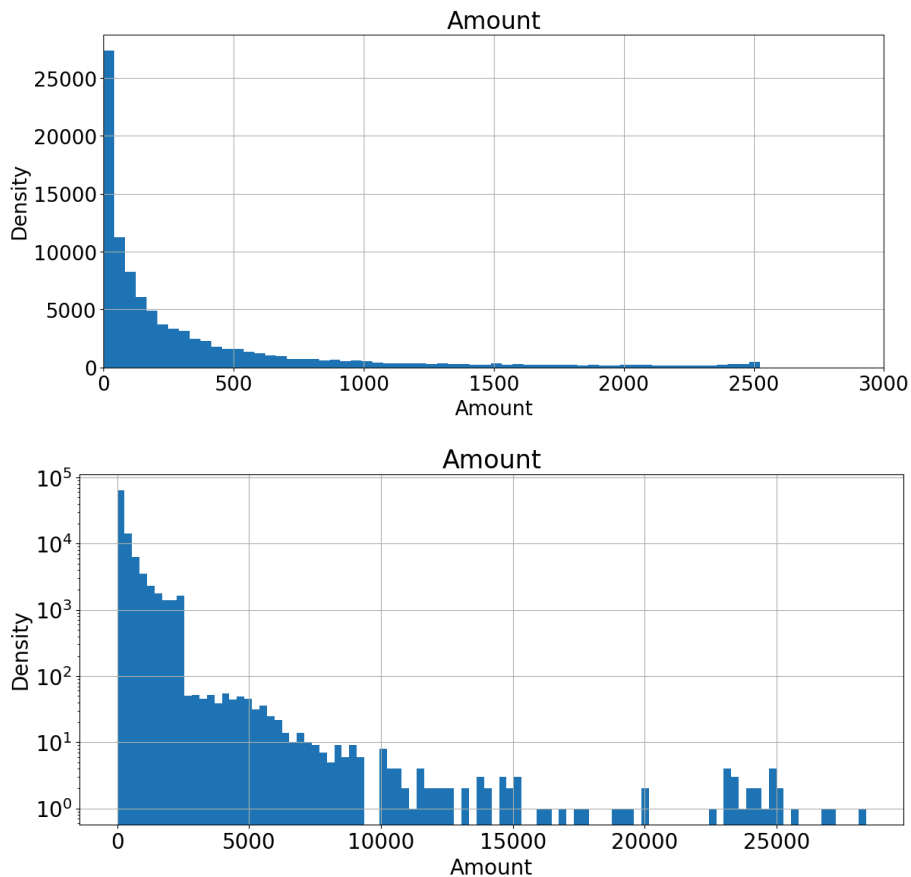
Description: Total Dollars spent in this transaction.

Outliers: 3 records (Threshold \$30275)

The distribution was highly right-skewed, implying most transactions have low spending.

The first distribution has an x-range of 0 to 3000, including 0.99% of the data.

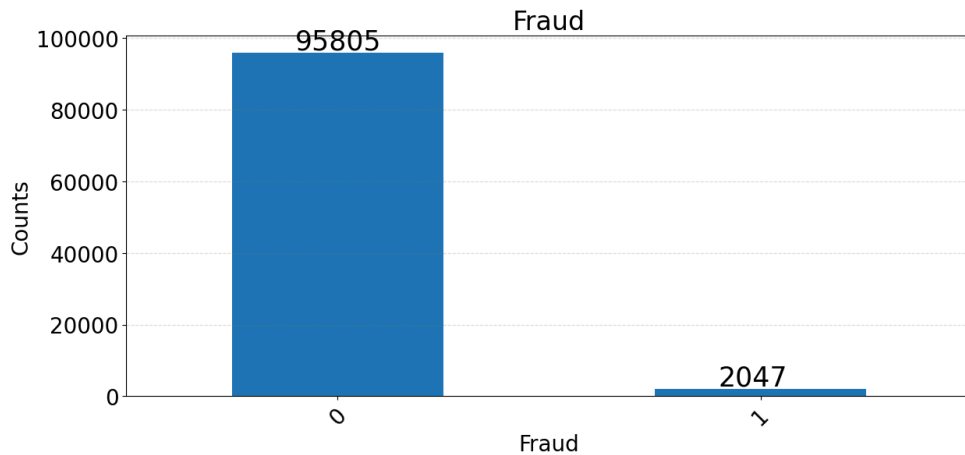
The second distribution log-transformed the y-axis, including all data but the outliers.



2) Field Name: Fraud (Label)

Description: 1 = Fraud Transaction, 0 = Not Fraud Transaction

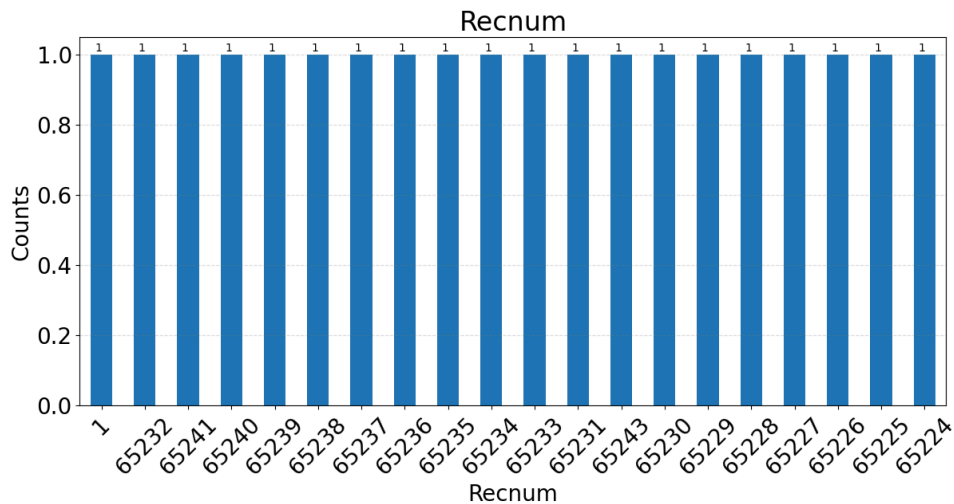
This is the label/dependent value for the dataset, with 95805 non-fraud transactions and 2047 fraud transactions.



Categorical:

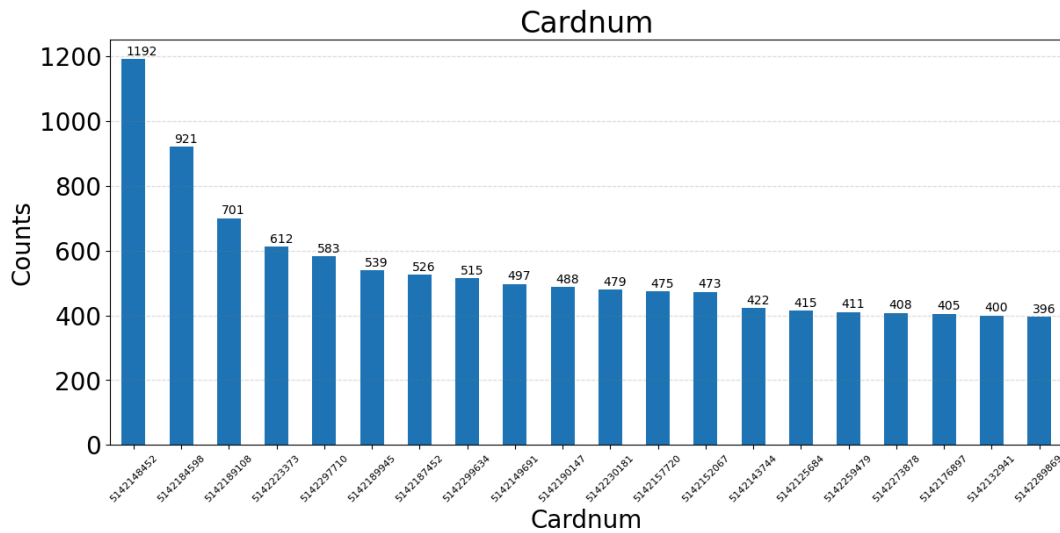
3) Field Name: Recnum

Description: The record number of each transaction. One unique value for each record.



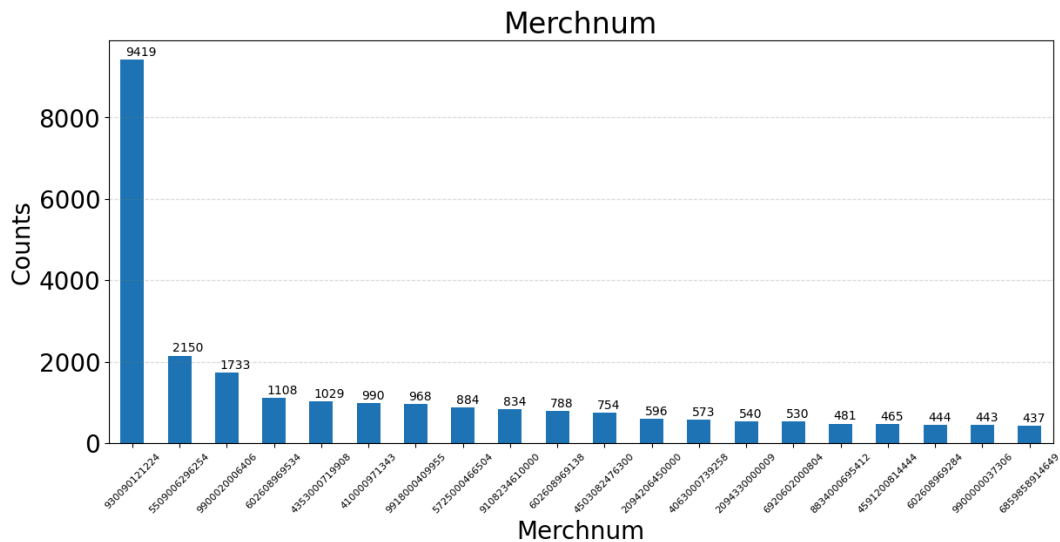
4) Field Name: Cardnum

Description: The credit card numbers used in the transactions. The distribution shows the top 20 field values of Cardnum. The most common Cardnum used is 5142148452, with a total count of 1,192.



5) Field Name: Merchnum

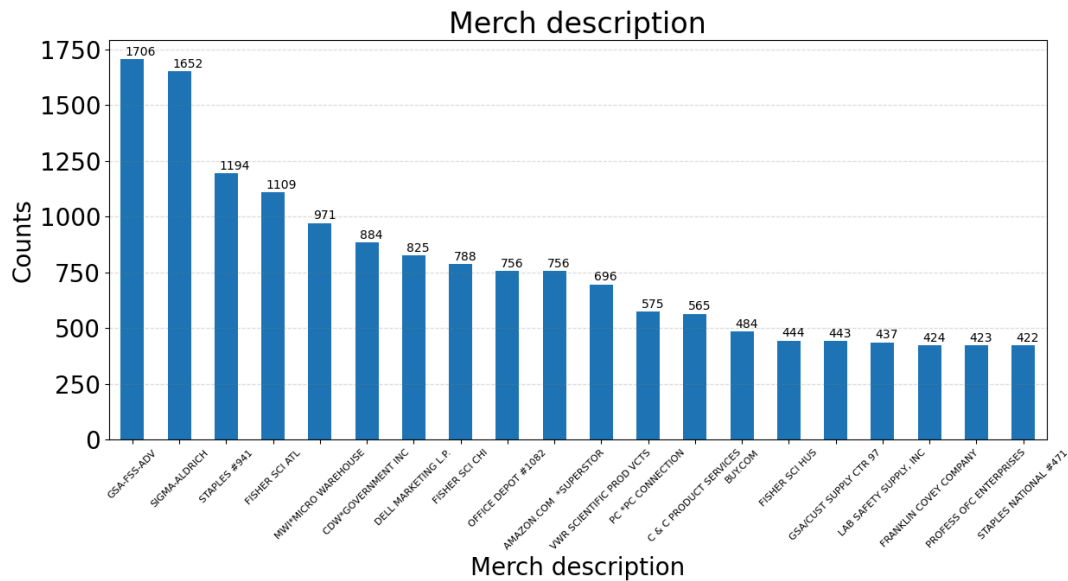
Description: The merchandise number purchased in the transaction. The distribution shows the top 20 field values of Merchnum. The most common Merchnum used is 930090121224, with a total count of 9,419.



6) Field Name: Merch description

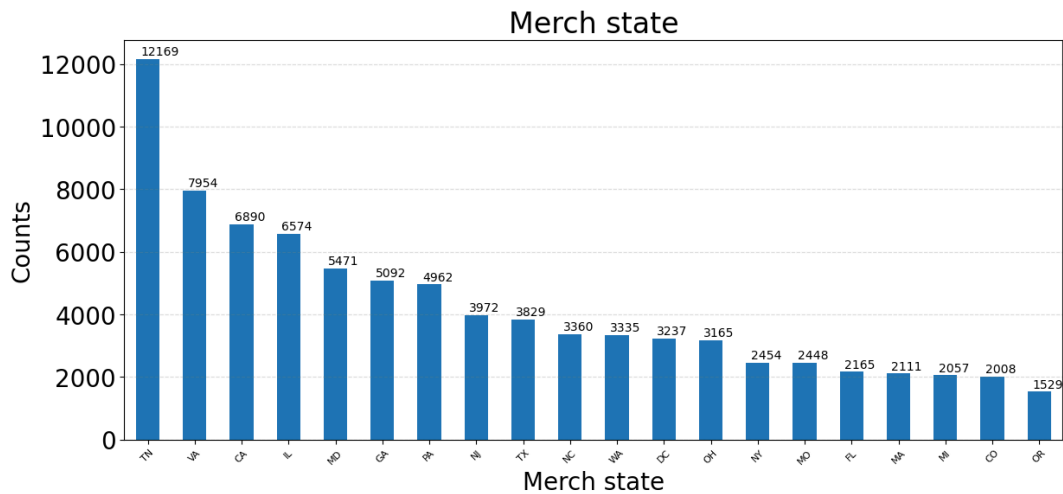
Description: The description of the merchandise in the transaction. The distribution shows the top 20 field values of Merch description. The most common Merch description used is GSA-FSS-ADV, with a total count of 1,706.

Observation: The highest Merchnum 930090121224 doesn't align with the highest Merch description since the highest Merchnum description includes shipping dates that alter through time. On the other hand, the highest Merch description GSA-FSS-ADV's Merchnum is 9900020006406, which is the third highest Merch num.



7) Field Name: Merch state

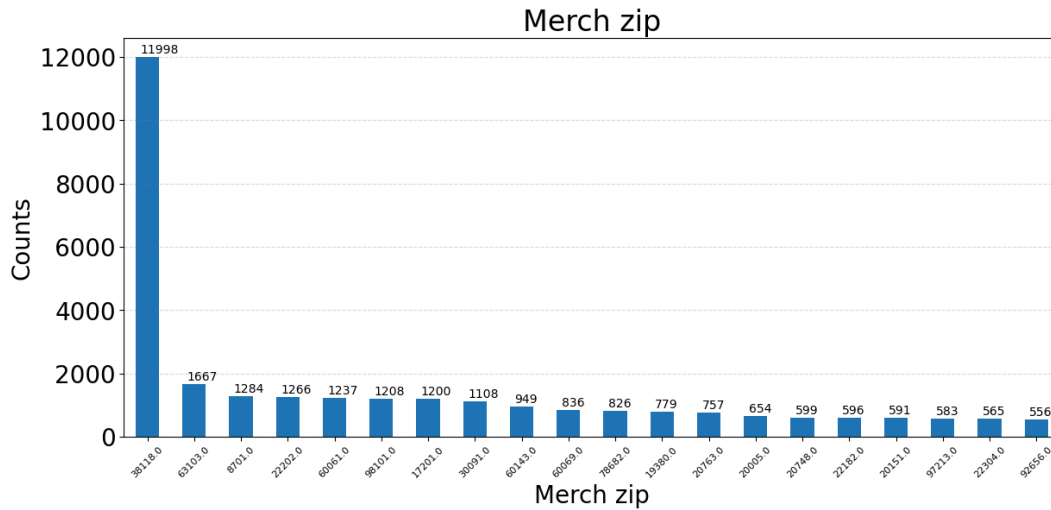
Description: The Merch state indicates where the transaction was made. The distribution shows the top 20 field values of Merch state. The most common Merch state is TN, with a total count of 12,169.



8) Field Name: Merch zip

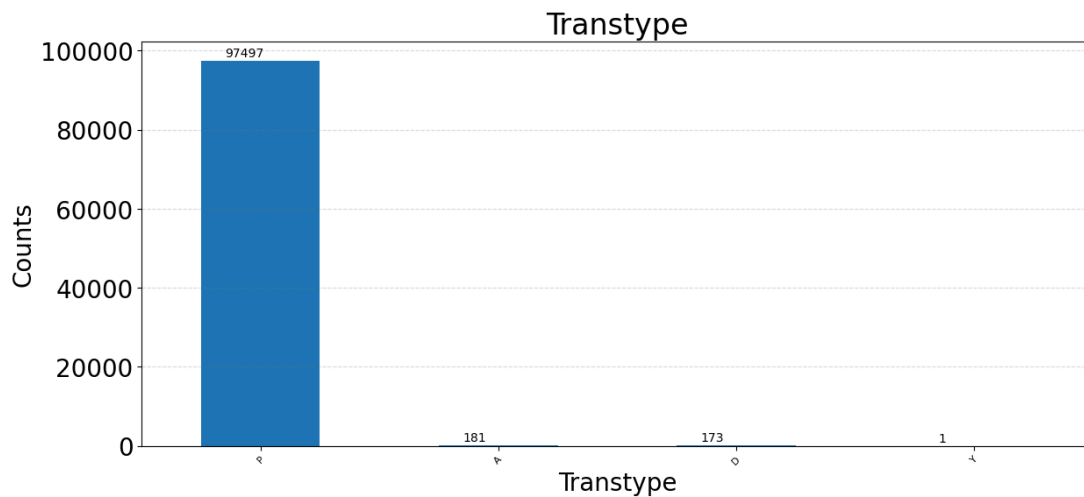
Description The Merch zip indicates the zip code where the transaction was made. The distribution shows the top 20 field values of Merch zip. The most common Merch zip is 38118, with a total count of 11,998.

Observation: The zip code 38118 is in TN, aligning with Merch state.



9) Field Name: Transtype

Description The **Transtype** indicates the transaction type. The distribution shows the most transactions are type P with 97,497, 181 for type A, 173 for type D, and 1 for type Y.



10) Field Name: Date

Description: Transaction Date. The first distribution shows the number of daily transactions across time. The second distribution shows the number of weekly transactions across time.

