

# Data Quality Report

## 1. Data Description

The dataset is **Card Transaction Data**, which contains **Each Transaction's Identifying Information** of credit cards and it's fraud results. The data came from a hundred thousand real U.S. transaction record **over the year of 2010**. There are **10 fields** and **97,852 records**.

## 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	Recnum	Categorical	97852	100.0%	0	97852	1
1	Date	Categorical	97852	100.0%	0	365	2010-02-28 00:00:00
2	Cardnum	Categorical	97852	100.0%	0	1645	5142148452
3	Merchnum	Categorical	94455	96.5%	0	13091	930090121224
4	Merch description	Categorical	97852	100.0%	0	13126	GSA-FSS-ADV
5	Merch state	Categorical	96649	98.8%	0	227	TN
6	Merch zip	Categorical	93149	95.2%	0	4567	38118.0
7	Transtype	Categorical	97852	100.0%	0	4	P
8	Fraud	Categorical	97852	100.0%	95805	2	0

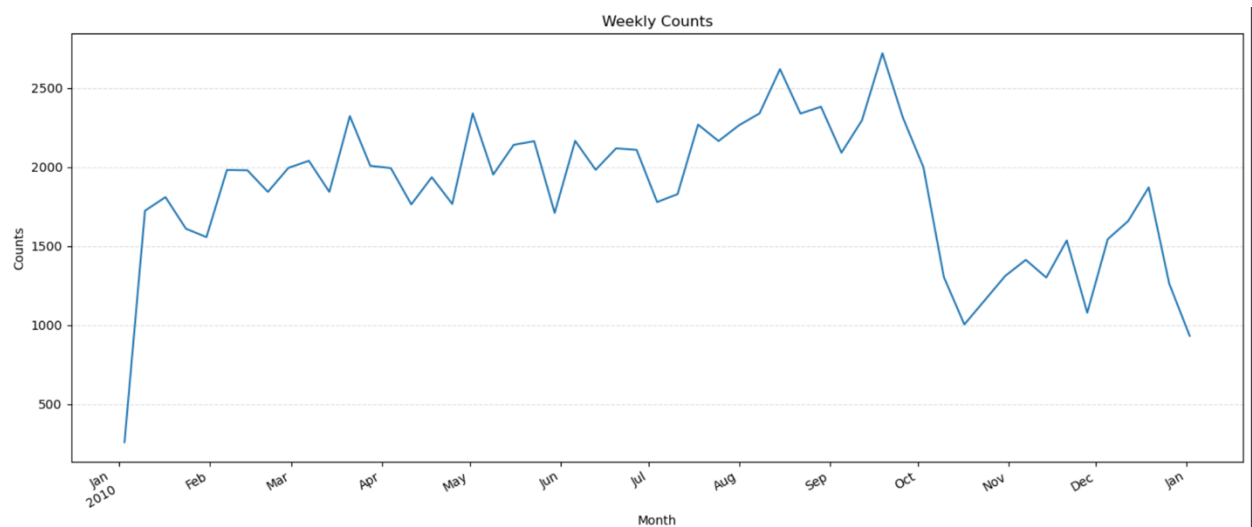
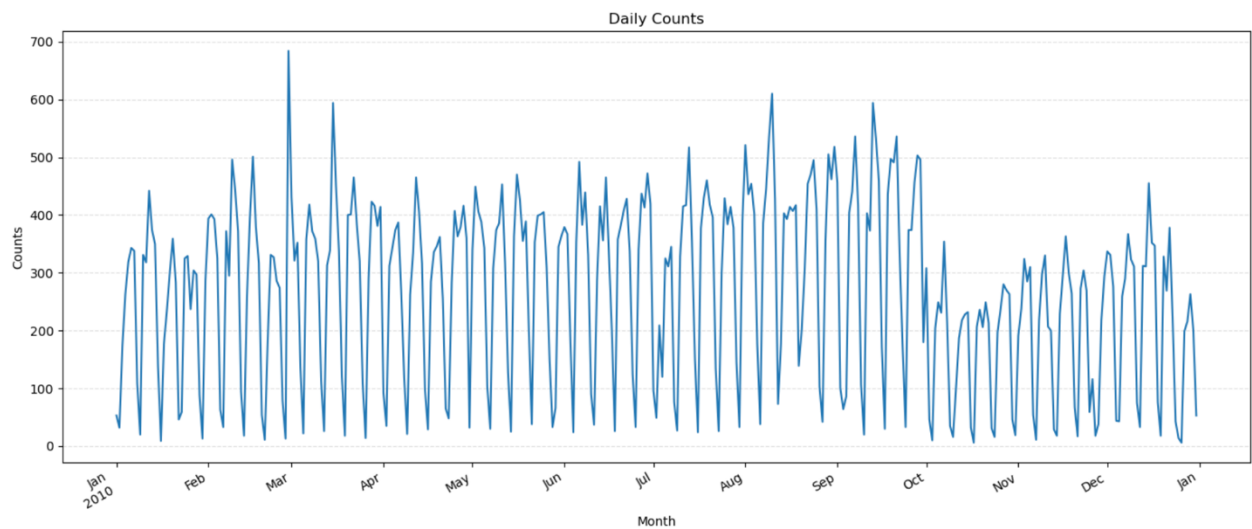
### 3. Visualization of Each Field

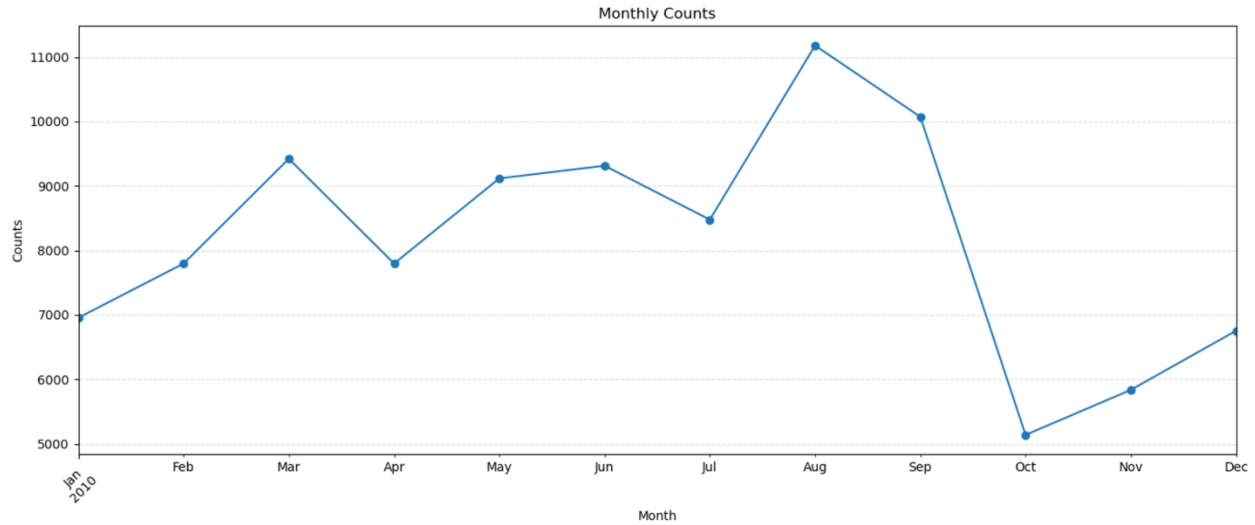
#### 1) Field Name: Recnum

Description: Ordinal unique positive integer for each transaction record, from 1 to 97,852.

#### 2) Field Name: Date

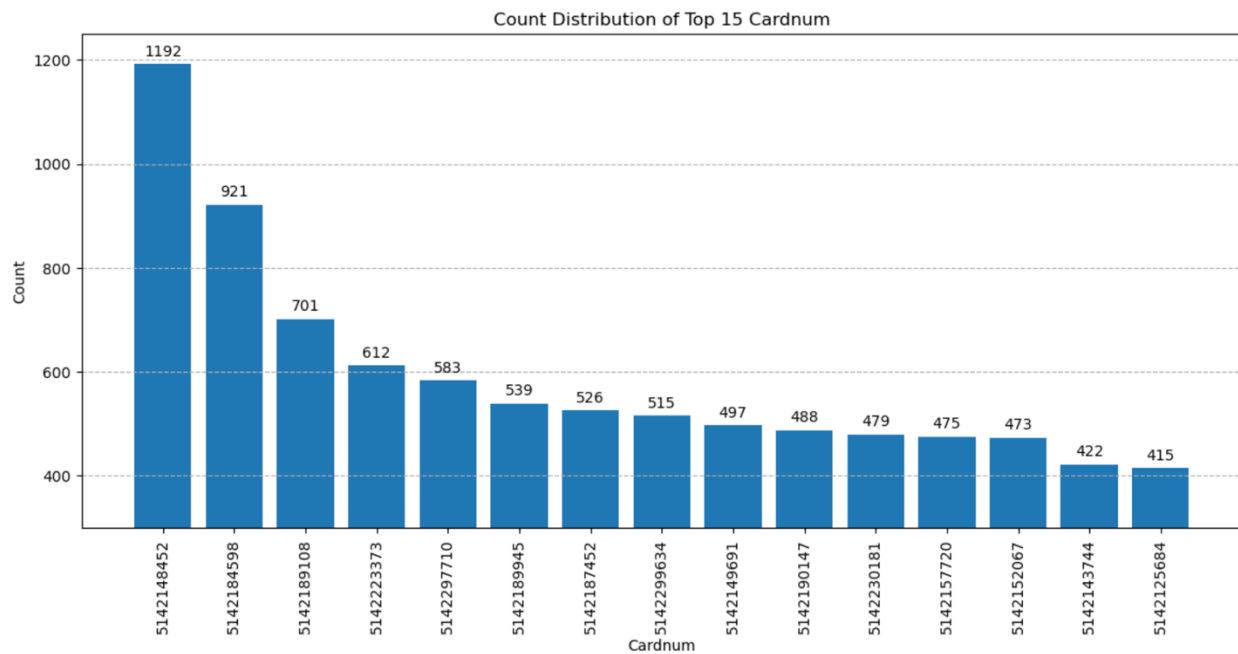
Description: Transaction date (Date). The first distribution shows the number of daily applications across time. The second distribution shows the number of weekly applications across time. The third distribution shows the number of month applications across time





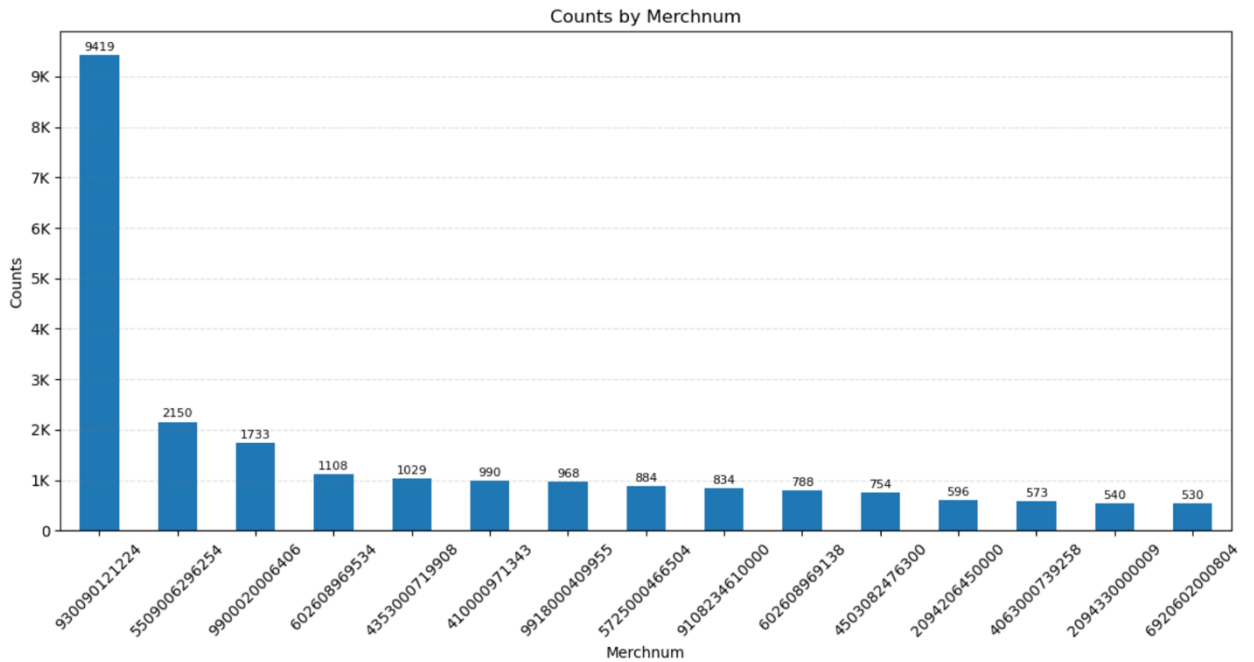
### 3) Field Name: Cardnum

Description: Each transaction's card number (Cardnum). The distribution displays the top 15 occurrences of card numbers used. The most common card number is 5142148452, showing a total count of 1,192, indicating a high frequency of use which could signify a card favored for regular transactions or by a heavy user.



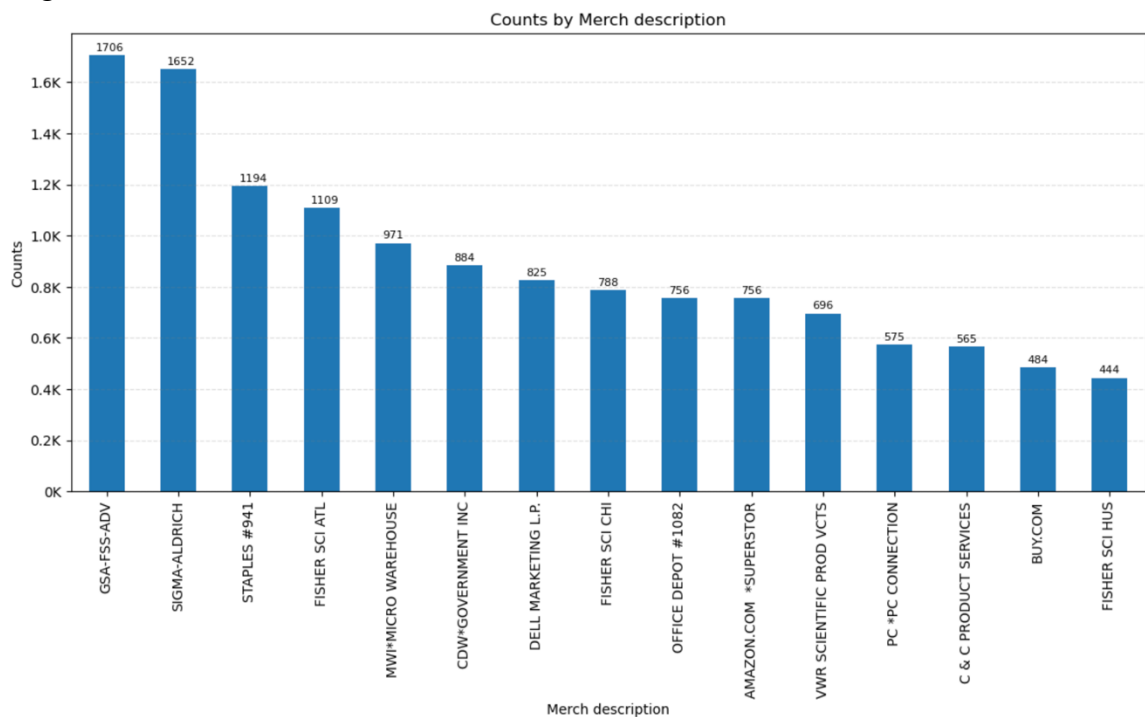
### 4) Field Name: Merchnum

Description: Merchant number (Merchnum). The highest occurring Merchnum stands out markedly at 9,419 counts which is 930090121224, suggesting it might be a focal point for transactions. Other Merchnum counts span from just over 2,000 down to around 500, highlighting a substantial variation in transaction frequency among merchants.



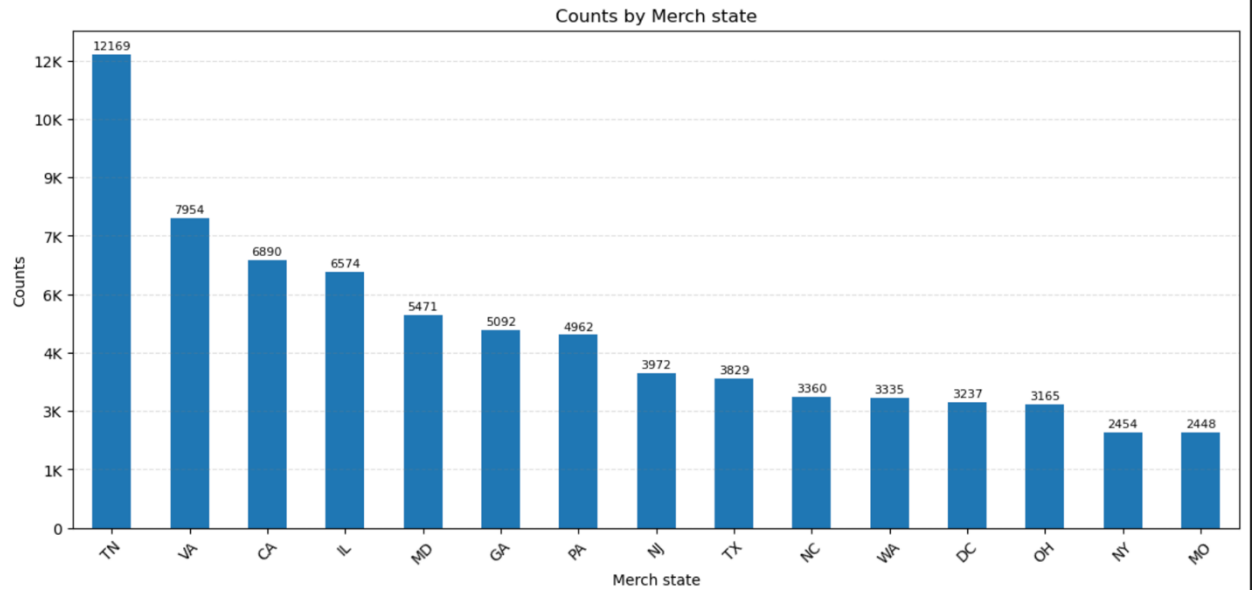
##### 5) Field Name: Merch description

Description: The description of Merchant (Merch description). The bar chart details transaction counts for various merchant descriptions, with 'GSA-FSS ADV' leading at 1,706 transactions. The data reflects a descending order of activity, illustrating a wide range in transaction volumes across merchants.



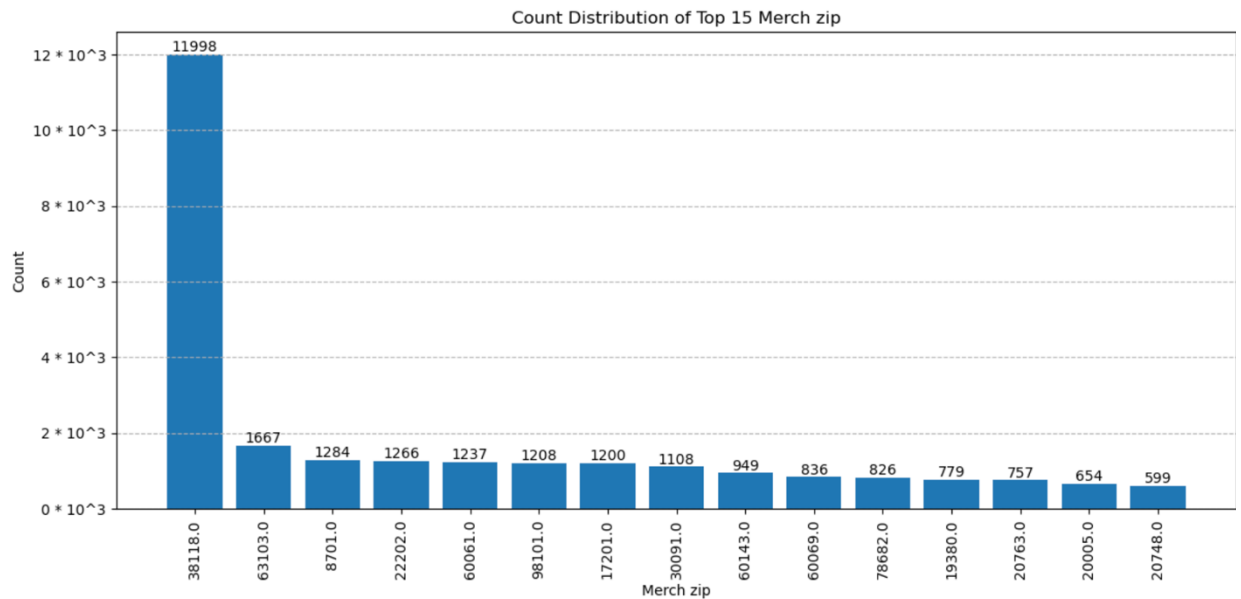
## 6) Field Name: Merch State

Description: Transaction's State (Merch State). The bar chart presents a count of transactions by merchant state, showing Tennessee (TN) with the highest at over 12,000. A clear trend of decreasing transaction counts is visible across the states displayed.



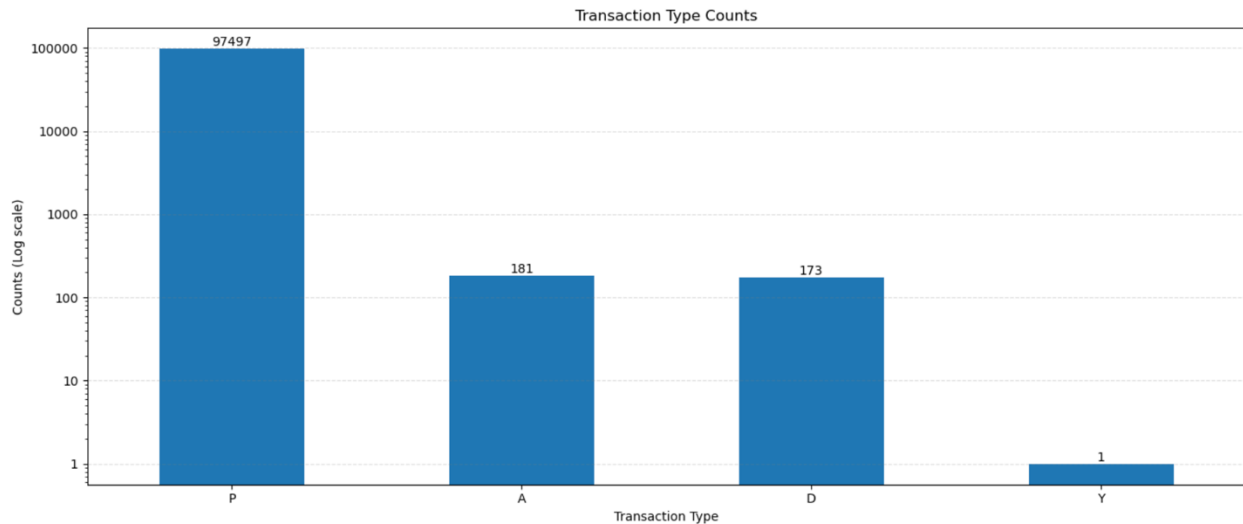
## 7) Field Name: Merch zip

Description: Transaction's zip code. This bar chart outlines the count distribution for the top 15 merchant ZIP codes. The ZIP code 38118.0 leads significantly, with nearly 12,000 counts, indicating a possible hotspot of commercial activity.



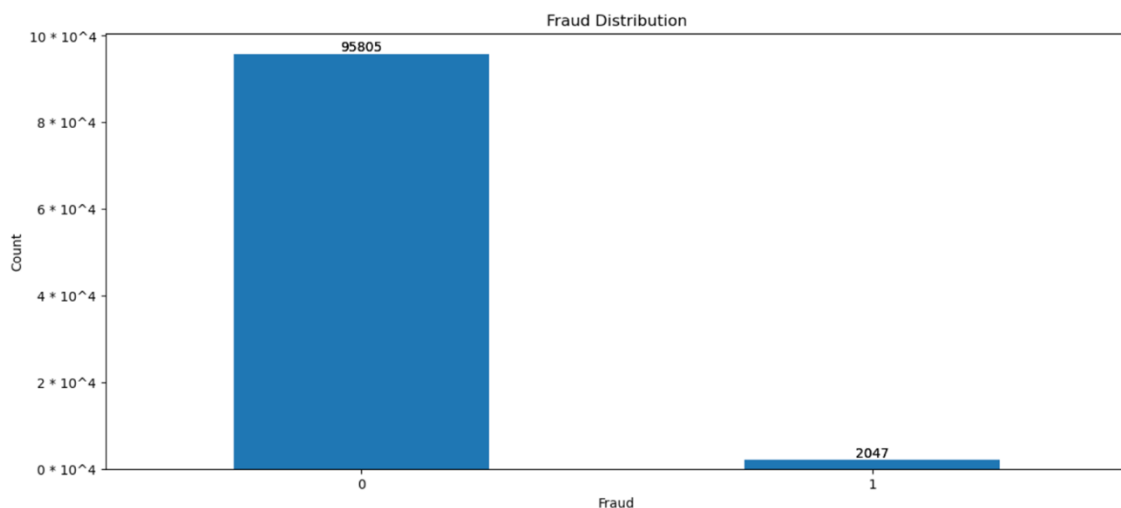
### 8) Field Name: Transtype

Description: Transaction's Type (Transtype). The bar chart illustrates the counts of different transaction types on a logarithmic scale. Type 'P' transactions dominate the chart with 97,497 occurrences, substantially outnumbering the others. Types 'A' and 'D' have a comparable presence, while 'Y' is scarcely represented with a single transaction.



### 9) Field Name: Fraud

Description: The bar chart displays a fraud distribution where non-fraudulent transactions, labeled '0', vastly outnumber the fraudulent ones, labeled '1', with counts of 95,805 and 2,047 respectively. This visual disparity underscores the relative infrequency of fraud in the dataset.



### 10) Field Name: Amount

Description: Each transaction's amount (Amount). This histogram displays the distribution of the "Amount" variable using a logarithmic scale on the y-axis to accommodate the wide range of values. The majority of the data is heavily concentrated at the lower end, near zero, indicating a highly skewed distribution with a sharp peak at the first bin. Additionally, there is a smaller but noticeable increase at the highest end of the range, suggesting the presence of outliers or a long tail to the right of the distribution.

