

# Data Quality Report

Alexander Tan

## 1. Dataset Description

The dataset contains **company credit card transaction data** from a **U.S. government organization** in **2010**. The data holds **18 fields** with **96,753 records**. The following credit card transaction fields are card number, date, transaction type, and amount. The following fields concerning the merchant are merchant number, merchant description, merchant state, and merchant zip code. A fraud field is present to label whether the transaction is legitimate.

## 2. Summary Tables

### 2.1 Numerical Table

Field Name	% Populated	Min	Max	Mean	Stdev	% Zero
Date	100.00	2010-01-01	2010-12-31	NA	NA	0.00
Amount	100.00	0.01	3,102,045.53	427.89	10,006.14	0.00

### 2.2 Categorical Table

Field Name	% Populated	# Of Unique Values	Most Common Value
Recnum	100.00	96,753	NA
Cardnum	100.00	1,645	5142148452
Merchnum	96.51	13,091	930090121224
Merch description	100.00	13,126	GSA-FSS-ADV
Merch state	98.76	227	TN
Merch zip	95.19	4,567	38118
Transtype	100.00	4	P
Fraud	100.00	2	0

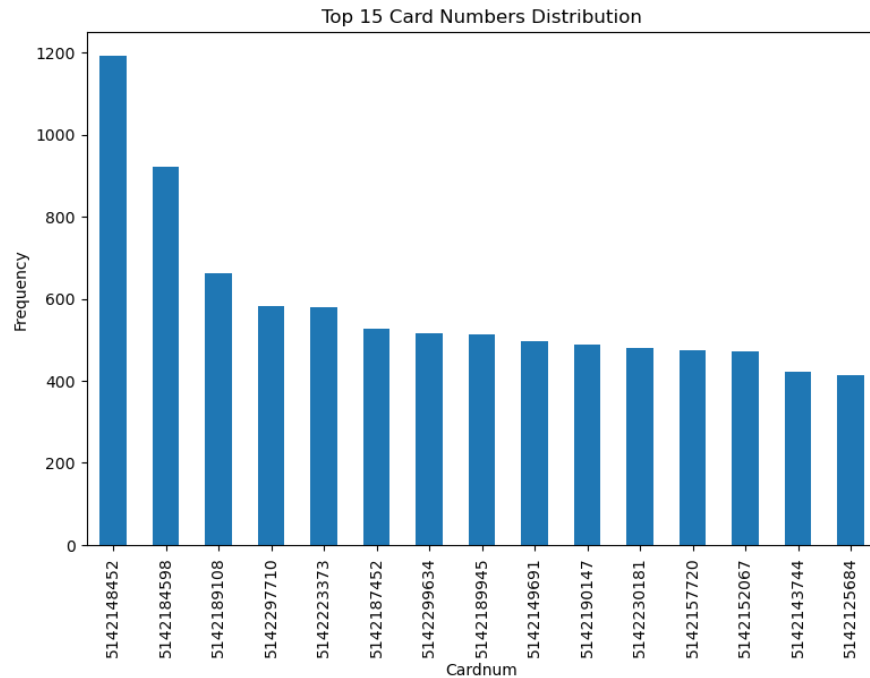
## 3. Field Visualization

### 3.1 Field Name: Recnum

**Description:** Ordinal unique positive integer representing each transaction from 1 to 96,753.

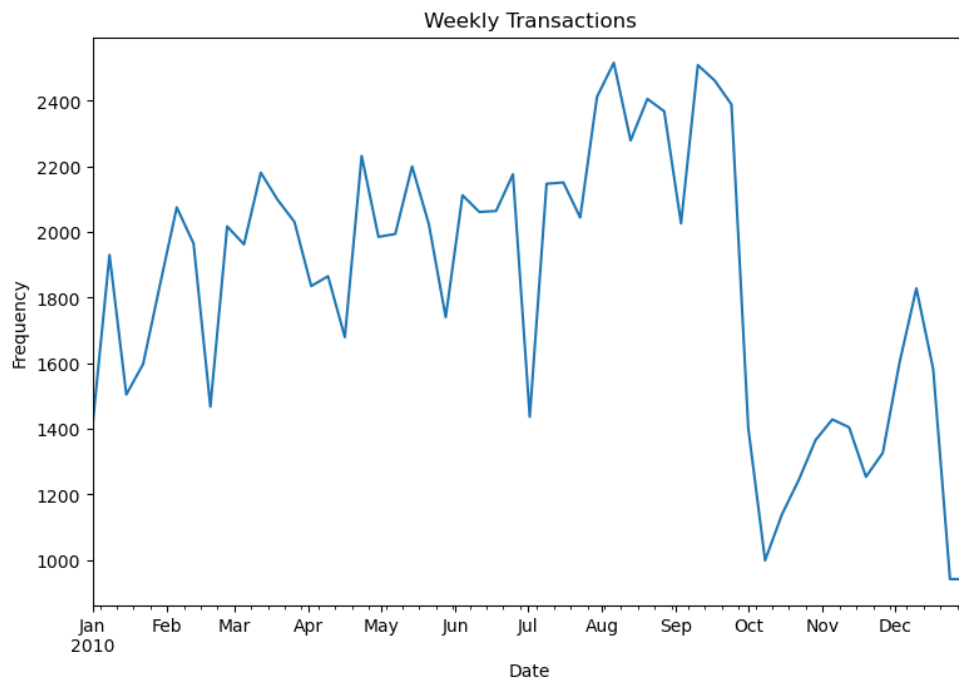
### 3.2 Field Name: Cardnum

**Description:** Categorical field of the credit card number involved in a transaction. The most common card number is 5142148452 with 1,192 occurrences. The distribution is the top 15 reoccurring card numbers.



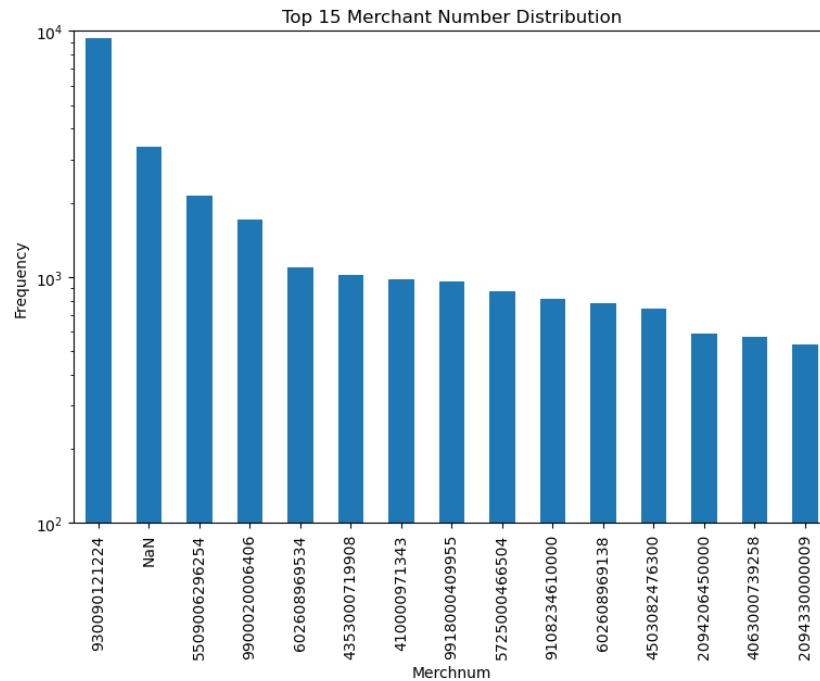
### 3.3 Field Name: Date

**Description:** Datetime field containing the date of each credit card transaction. The day with the most transactions was 2010-02-28 with 684 submissions. Weekly transactions are distributed across time (the year 2010).



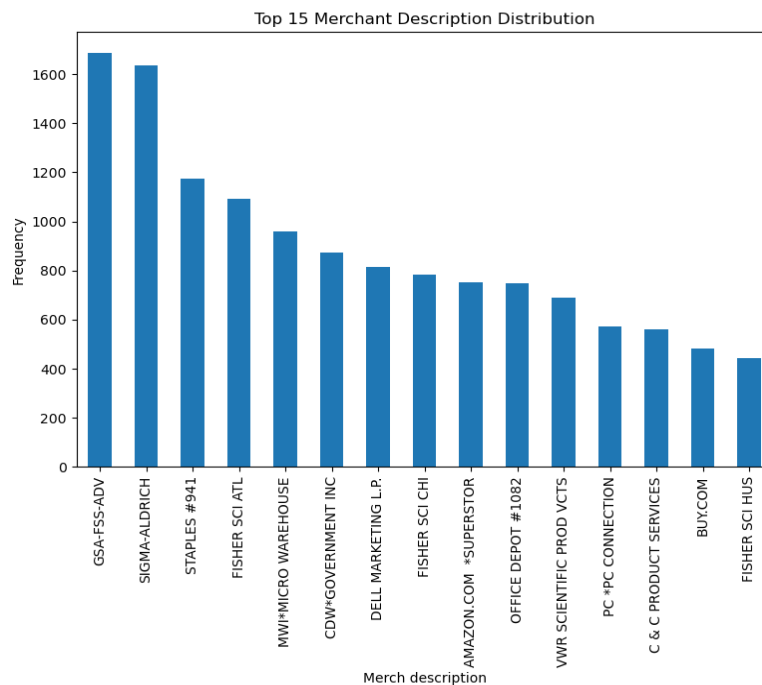
### 3.4 Field Name: Merchnum

**Description:** Categorical field of merchant identifying number for every credit card transaction. The most common merchant number is 930090121224 with 9,310 occurrences. This field also contains 3,375 null values. The distribution is the top 15 reoccurring merchant numbers.



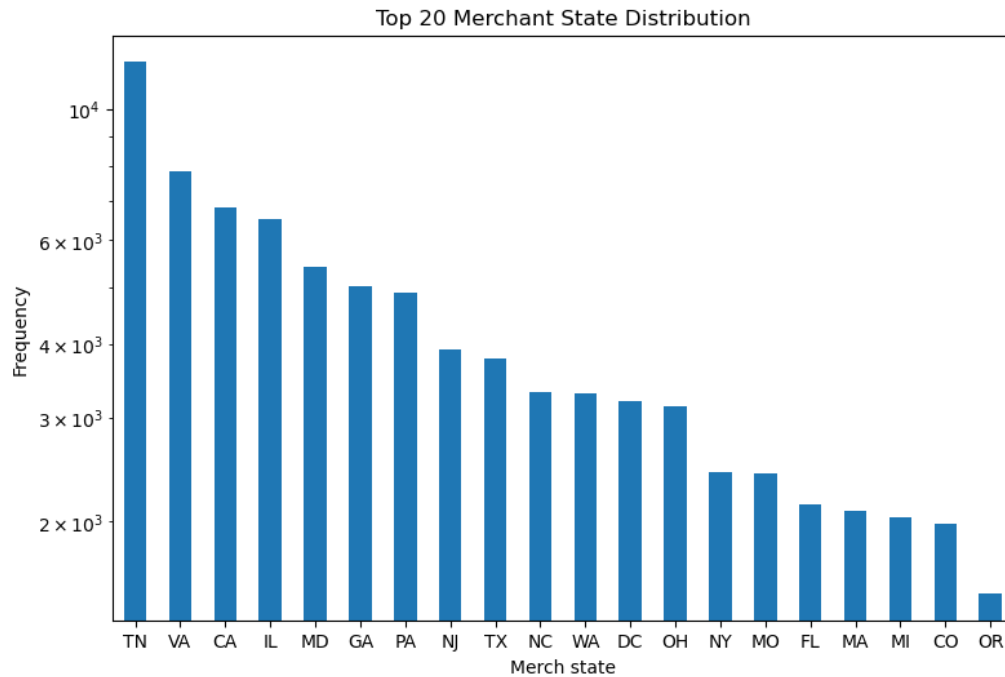
### 3.5 Field Name: Merch description

**Description:** Categorical field of the merchant's description for each credit card transaction. The most merchant description is GSA-FSS-ADV with 1,688 occurrences. The distribution is the top reoccurring 15 merchant descriptions.



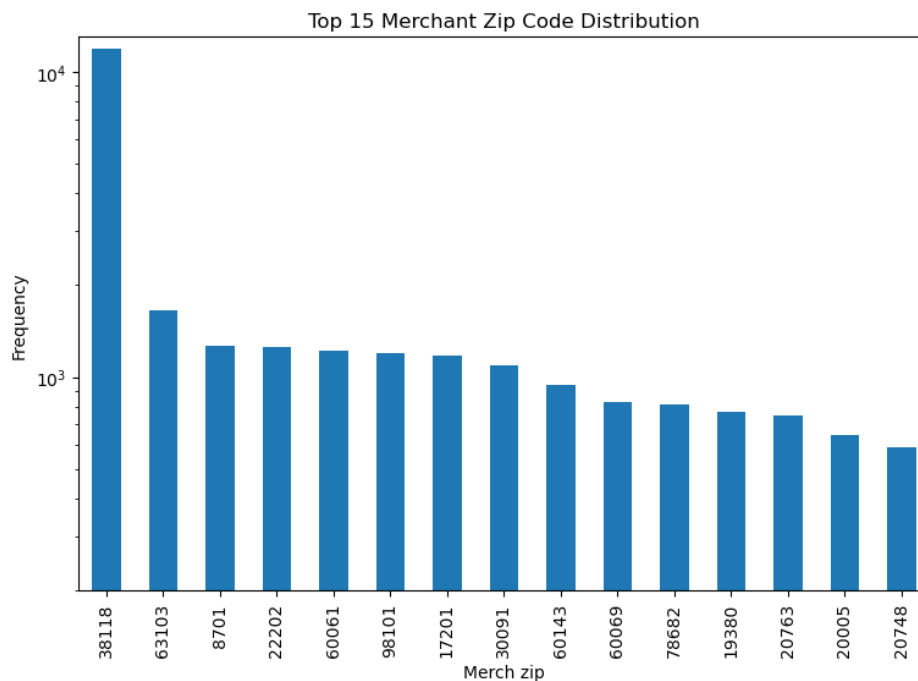
### 3.6 Field Name: Merch state

**Description:** Categorical field of the state where each merchant from a transaction is located. The most common state among applicants is TN with 12,035 occurrences. Additionally, this field contains 1,195 null values. The distribution is the top 20 merchant state occurrences.



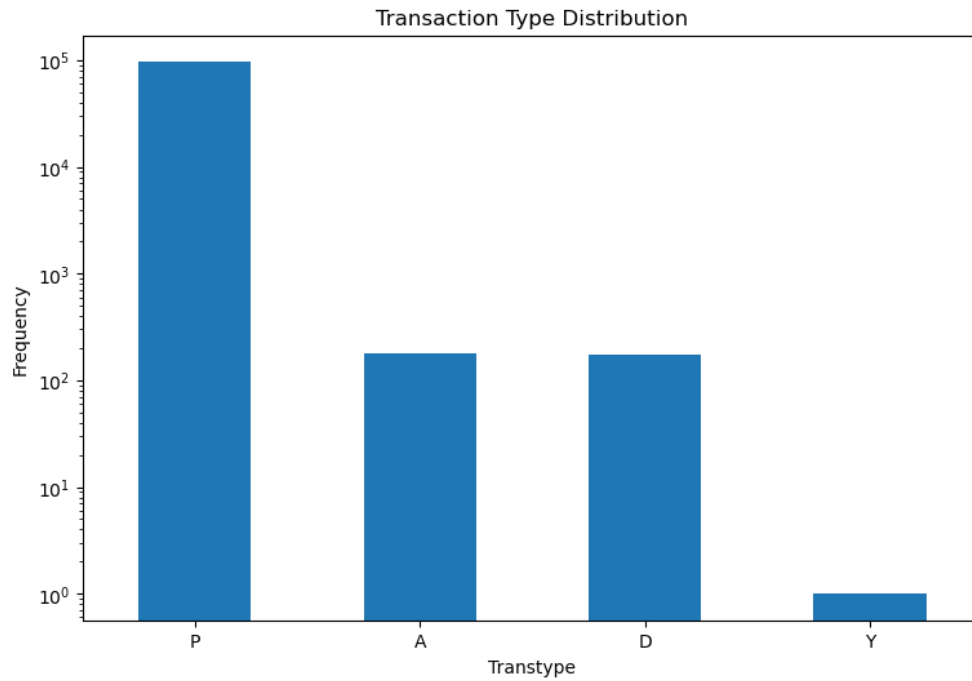
### 3.7 Field Name: Merch zip

**Description:** Categorical field of each merchant's zip code from a transaction. The most common zip code among merchants is 38118 with 11,868 occurrences. Additionally, this field contains 4,656 null values. The distribution is the top 15 reoccurring merchant zip codes.



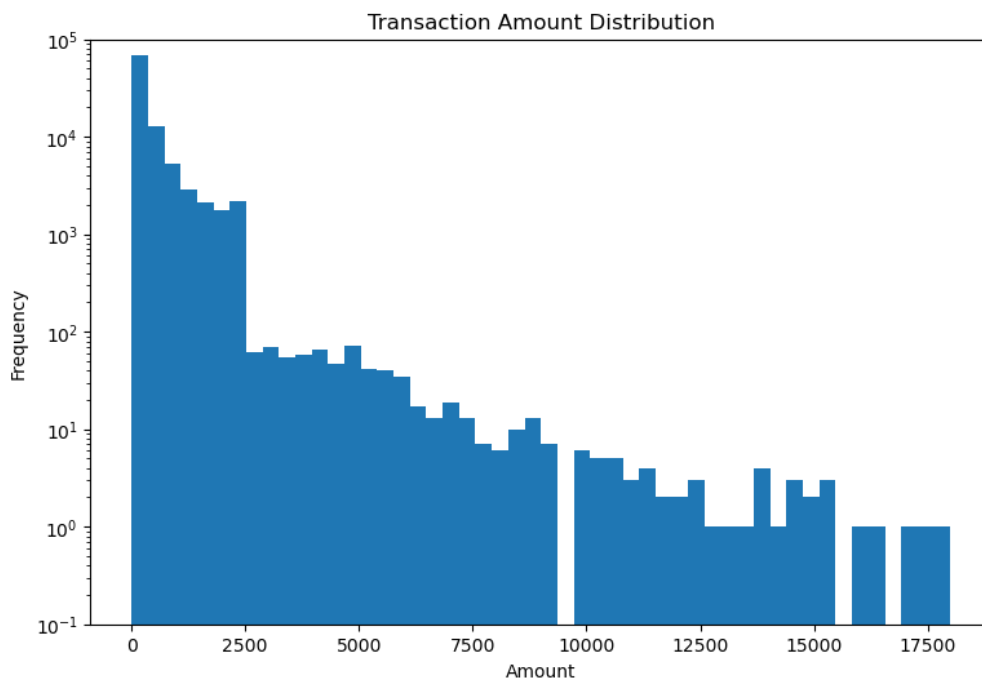
### 3.8 Field Name: Transtype

**Description:** Categorical field of each credit card transaction type. The most common transaction type is P, purchases, with 96,398 occurrences.



### 3.9 Field Name: Amount

**Description:** Numerical field of the amount spent for each credit card transaction. The average amount spent is 427.89 with a standard deviation of 10,006.14. The distribution is the reoccurring transaction amounts.



### 3.10 Field Name: Fraud

**Description:** Categorical field determining whether a credit card transaction is legitimate (0) or fraudulent (1). Most transactions are labelled as 0 with 95,694 occurrences. Transactions with fraud as 1 have 1,059 occurrences.

