Data Quality Report

1. Data Description

The data is about **Transaction Fraud**. The data is about fraud during account usage process like Credit Card transactions, Tax returns, Money transfer from a comprised account and Insurance claims from January 1st, 2010, to December 31st, 2010. It has 96,753 rows and 10 fields.

2. Summary Table

I. 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	3102045.53	427.88	10006.140	0.00

II. Categorical Table

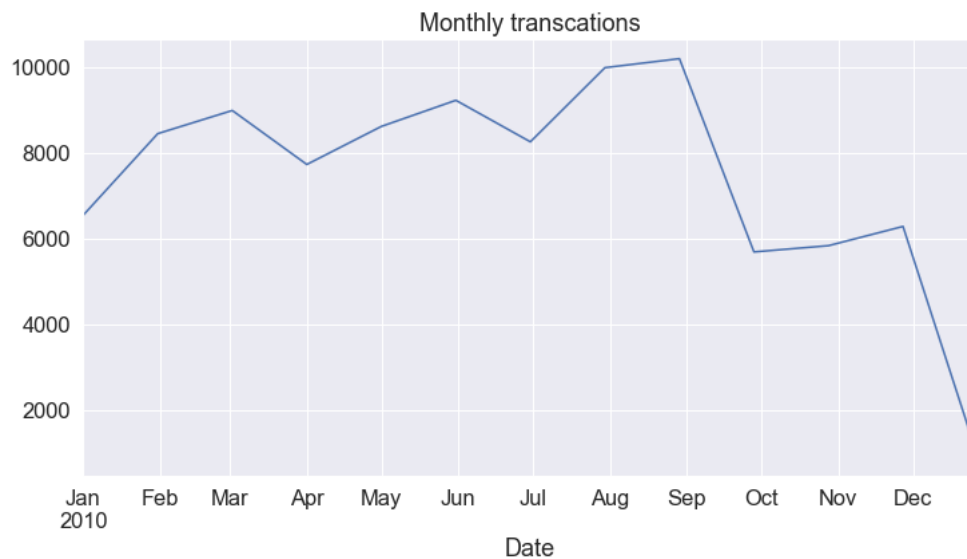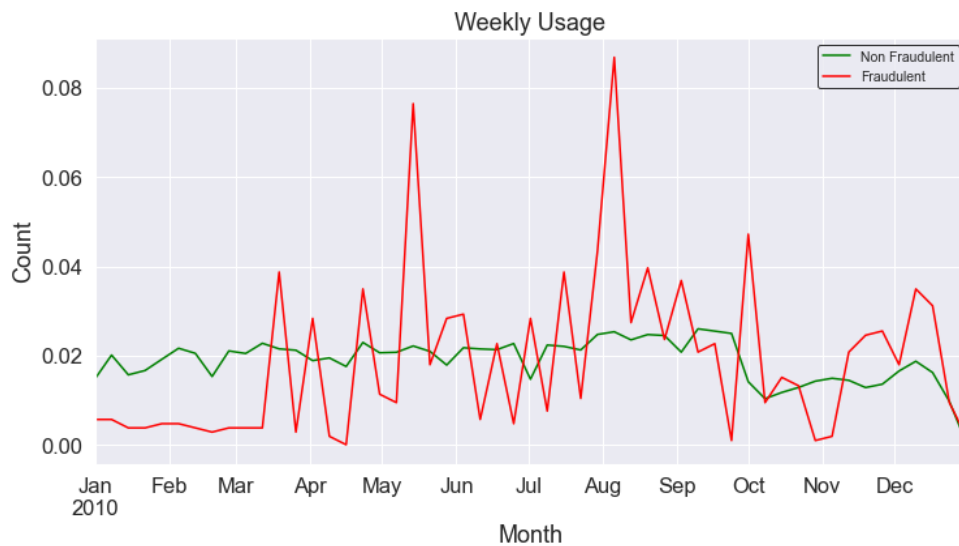| Field Name	% Populated	Blank	Zeroes	#Unique Value	Most Common Value
Recnum	100.00	0	0	96,753	1
Cardnum	100.00	0	0	1,645	5142148452.0
Merchnum	96.51	3375	231	13,091	930090121224
Merch Description	100.00	0	0	13,126	GSA-FSS-ADV
Merch state	98.76	1195	0	227	TN
Merch Zip	95.19	4656	0	4567	38118
Transtype	100.00	0	0	4	P
Fraud	100.00	0	0	2	0.0

Visualization of Each Field:

(1) Field Name: Recnum

   Description: Recnum: Ordinal Unique positive integer for each record from 1 to 96753.
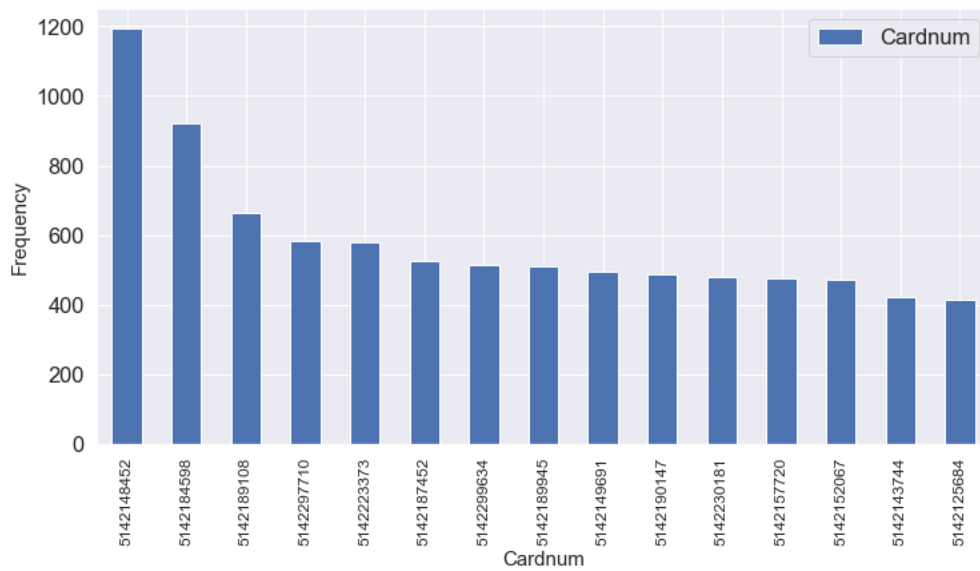
(2) Field Name: Date

   Description: Date: The "date" field indicates the date on which the individual has raised a request for an application. The Weekly use distribution across time.

(3) Field Name: Cardnum

Description: SSN: SSN refers to the Card number of the user. The distribution indicates the top 20 most common card users.
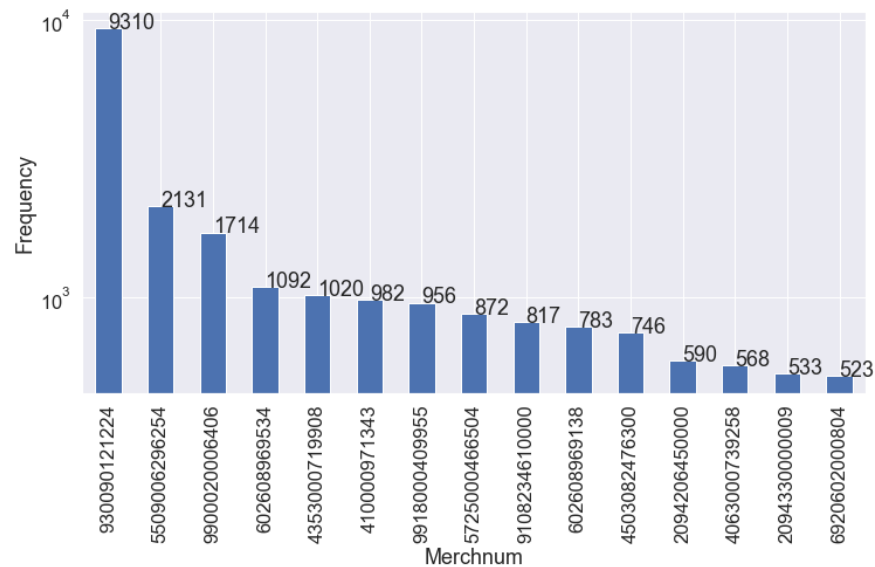
The most common value of the 5142148452, the count is 1192.
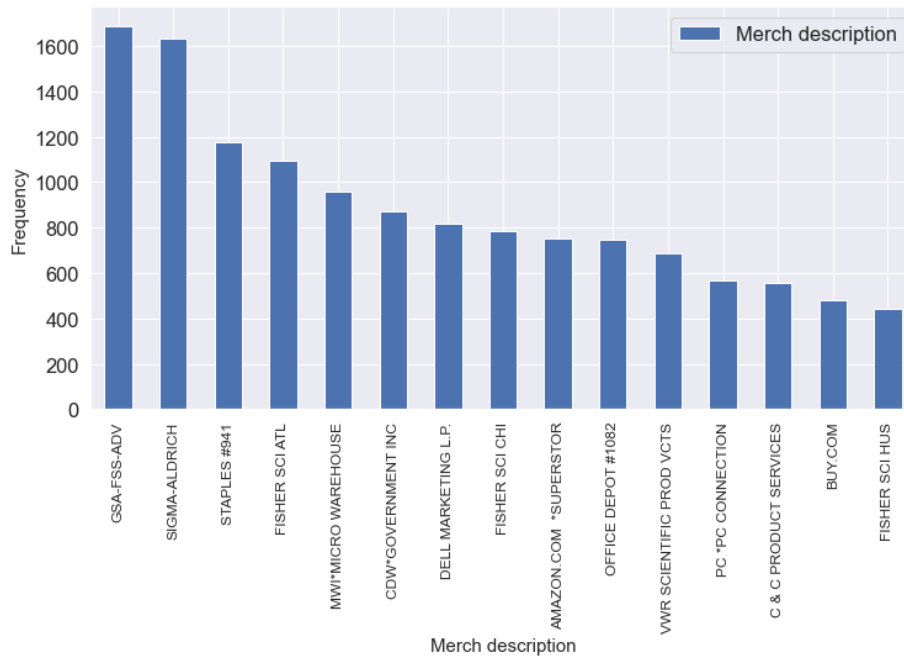

(4) Field Name: Merchnum

Description: Merchnum: The field "Merchnum" refers to the number of the Merchant. The distribution indicates the top 15 most common Merchant number used by the user.

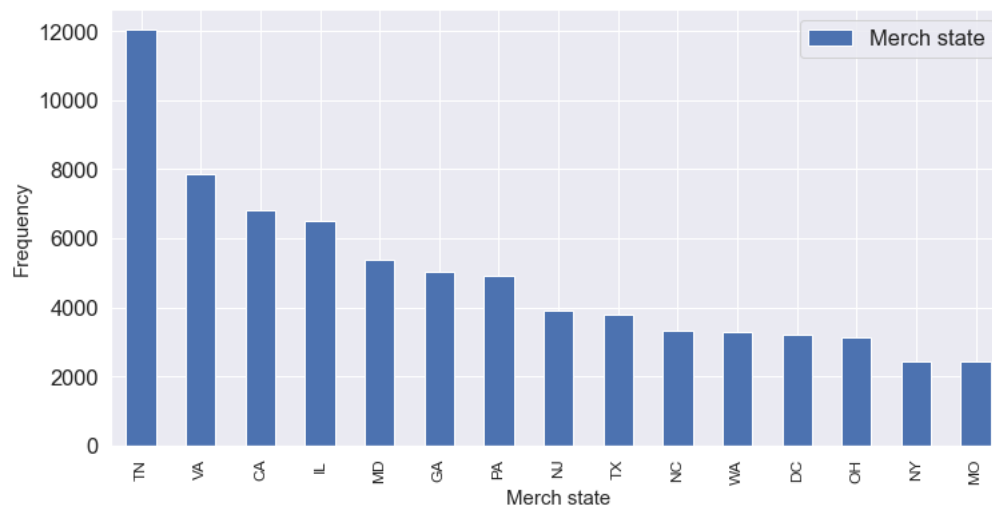The most common value of the Merch num is 930090121224, the count is 9310.

(5) Field Name: Merch description

    Description: Merch description: The field "Merch description" refers to description of the merchant. The distribution indicates the top 15 most common Merchant used by the applicant. The most common value of the Merch Description is GSA-FSS-ADV, the count is 1688.
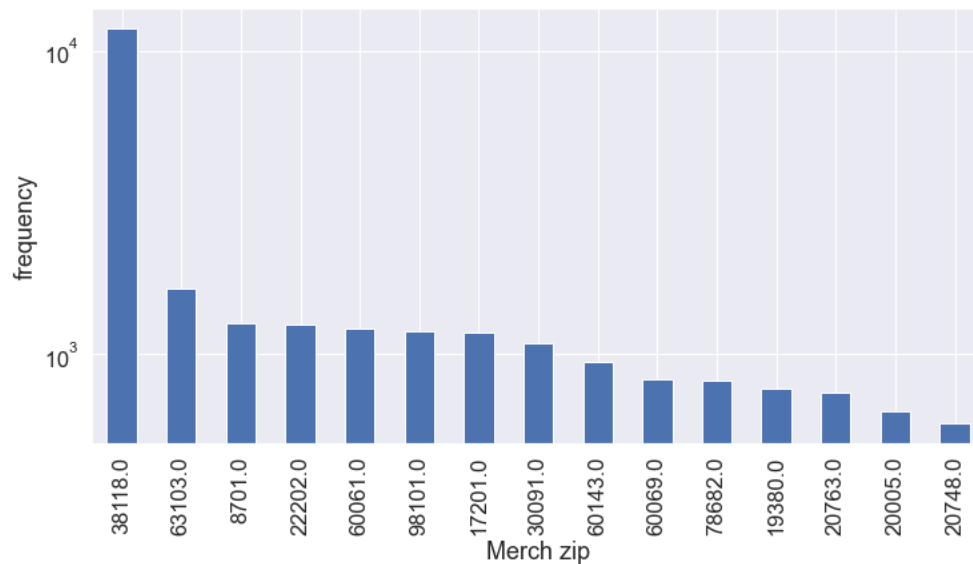

(6) Field name: Merch state

    Description: Address: The field "Merch state" refers to the state of the merchant. The distribution indicates the top 15 most common merchant state used.

The most common value of the Merch state is TN, the count is 12,35.

(7) Field Name: Merch Zip

Description: Merch Zip: The field "Merch Zip" refers to the Zip code of the user. The distribution indicates the top 15 most common Zip code used.
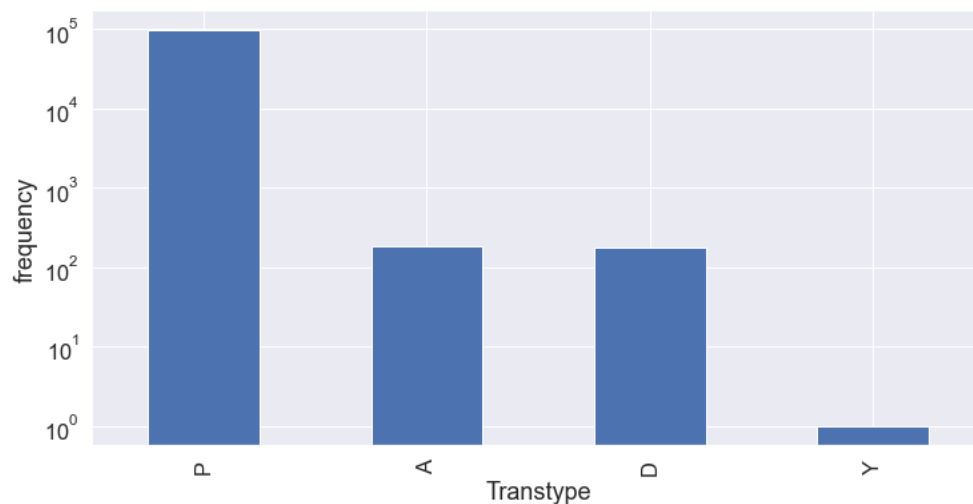
The most common value of the Zip code is 38,118, the count is 11868.

(8) Field Name: Transtype

Description: Transtype: The field "Transtype" refers to the type of transaction. The distribution indicates the 4 types of transactions where P means Purchase, A stands for Approval, D stands for Debit and Y stands for Year-end.
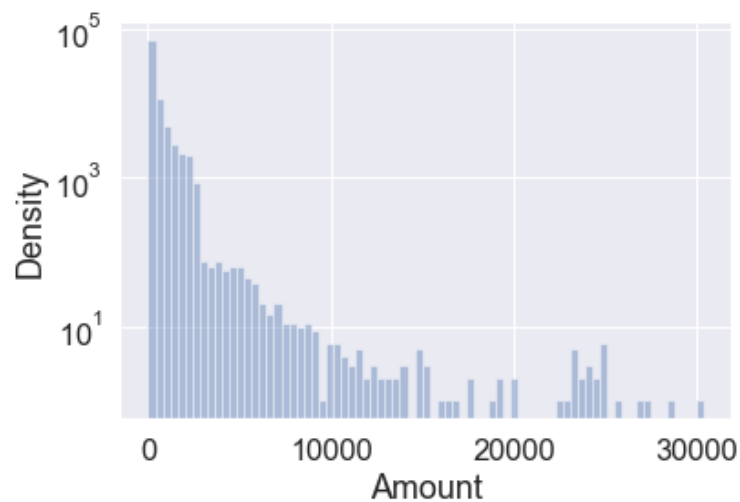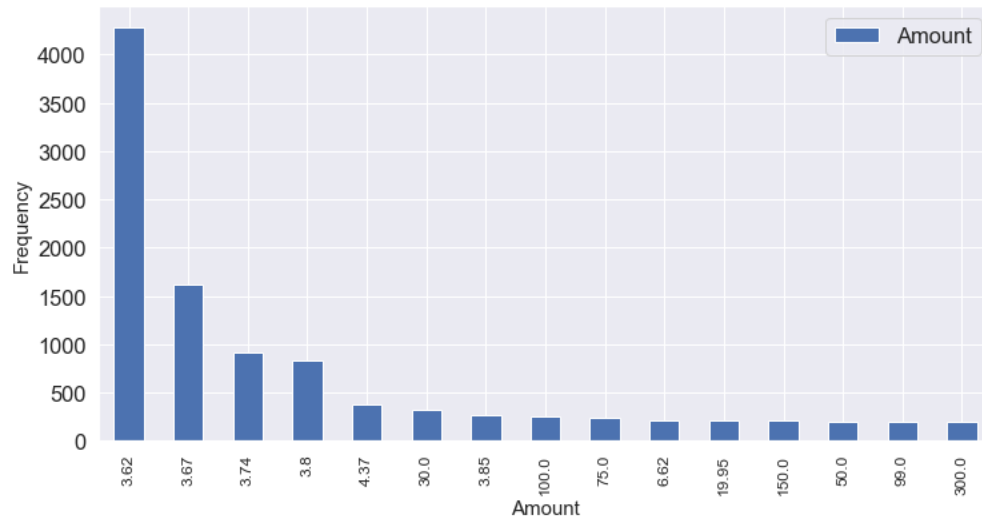
The most type of value of the transaction is P, the count is 96,398.

(9) Field Name: Amount

Description: Amount: The field "Amount" refers to the Transaction Amount. The distribution indicates the top 10 most common transaction amount.

The most common value of the Amount is 3.62, the count is 4,283.

(10) Field Name: Fraud

Description: Fraud = 0 (no fraud label), Fraud=1 (fraud label)

The count of Fraud=0 (95,694) and Fraud =1 (1,059)