Data Quality Report (DQR)

1. Basic Information for Dataset:
   This is a card transaction fraud dataset in 2010. It is provided to build a supervised model which can identify the propensity of a card transaction to be a fraud. There are 96708 records and 10 fields with 1 numeric variable and 9 categorical variables in the dataset. Each row is one record of card transaction with basic card transaction information and fraud label.

2. Summary Statistics:
   a. basic statistics
   Here is the summary table that includes field number, field name, count of records, number of unique value, and percent of populated records.

No.	Field Name	Count	# of Unique Values	Percent of populated
1	Recordnum	96708	96708	100
2	Cardnum	96708	1644	100
3	Date	96708	365	100
4	Merchantnum	93333	13091	96.51
5	Merch Description	96708	13125	100
6	Merchant State	95513	228	98.76
7	Merchant Zip	92052	4568	95.19
8	Transtype	96708	4	100
9	Amount	96708	34876	100
10	Fraud	96708	2	100

   b. additional statistics for numeric data
   Here is an additional summary table that includes field number, field name, mean, standard deviation, minimum, first quantile, median, third quantile, and maximum.

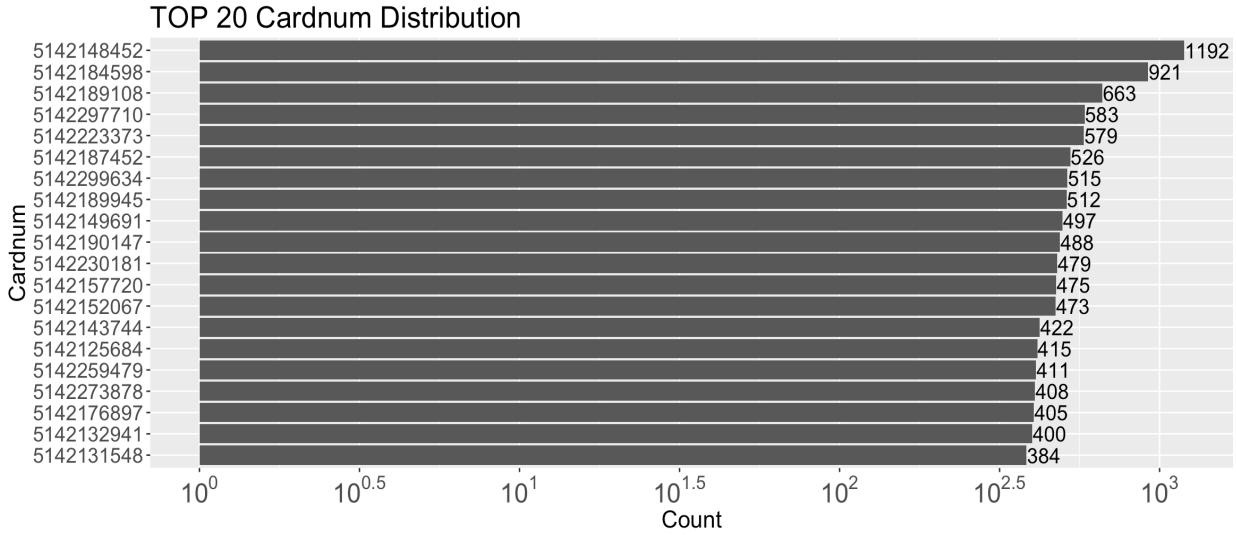
   These are the statistics that are calculated after removing the unusual value in the dataset. The amount value of this outlier is 3102045.53 with money measuring unit in Mexico. Our client aims to remove it.

No.	Field Name	Min	1st Q	Median	3rd Q	Max	Mean	SD
9	Amount	0.01	33.45	137.90	427.66	47900.00	395.79	832.05

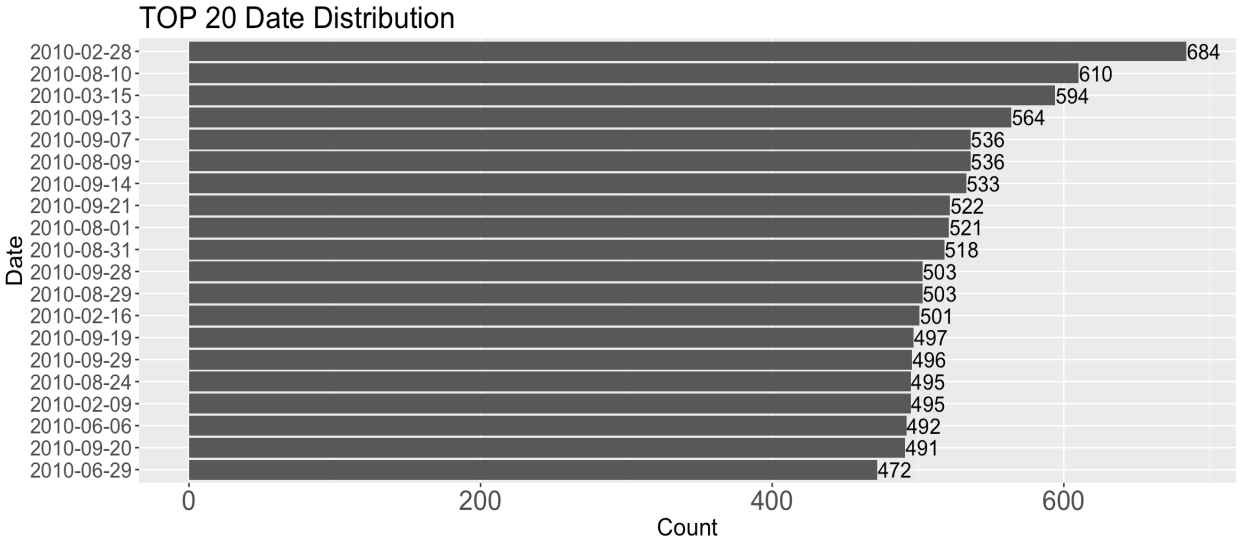
3. Detailed Information for Each Field:

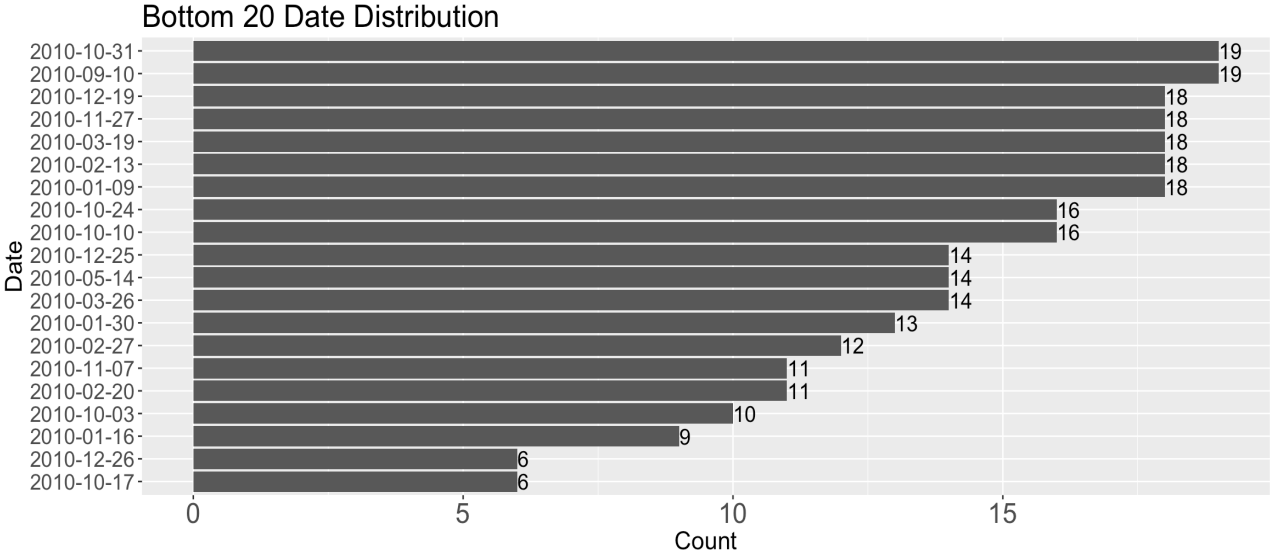
No.	Field Name	# of Unique Values	Percent of populated	Description
1	Recordnum	96708	100	Describe order of transactions

No.	Field Name	# of Unique Values	Percent of populated	Description
2	Cardnum	1644	100	The card number of transactions

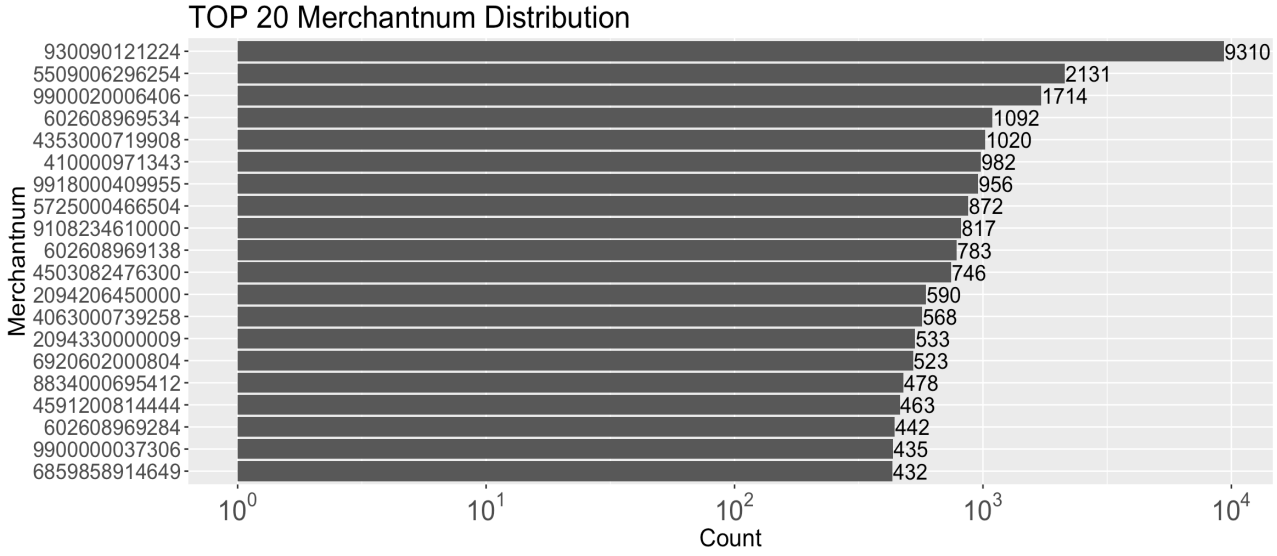


No.	Field Name	# of Unique Values	Percent of populated	Description
3	Date	365	100	The date of the transactions

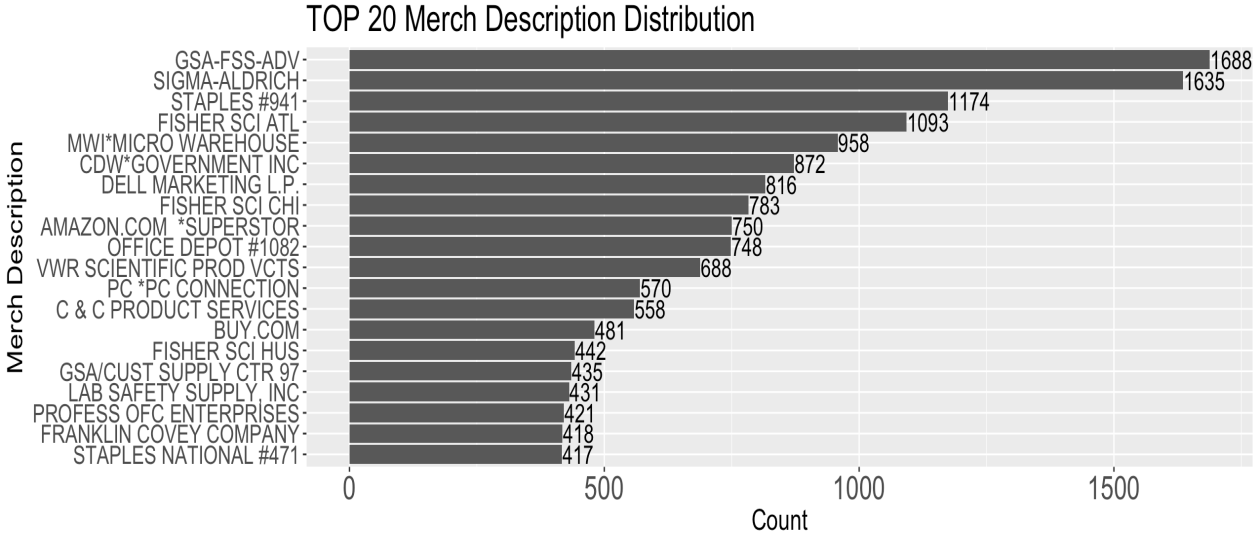




No.	Field Name	# of Unique Values	Percent of populated	Description
4	Merchantnum	13091	96.51	The identification number of merchant

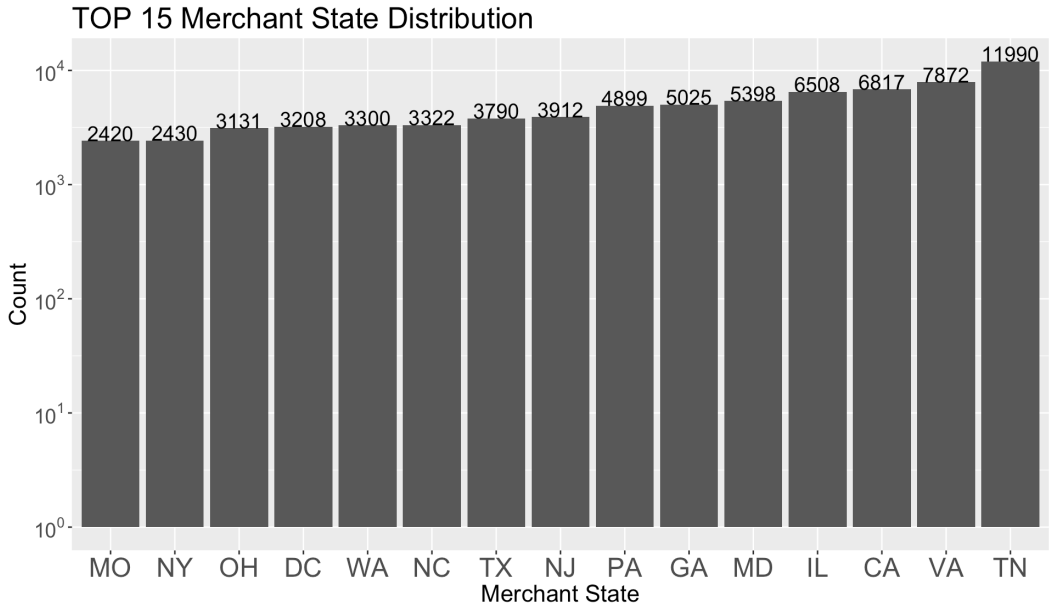


No.	Field Name	# of Unique Values	Percent of populated	Description
5	Merch Description	13125	100	The description of merches

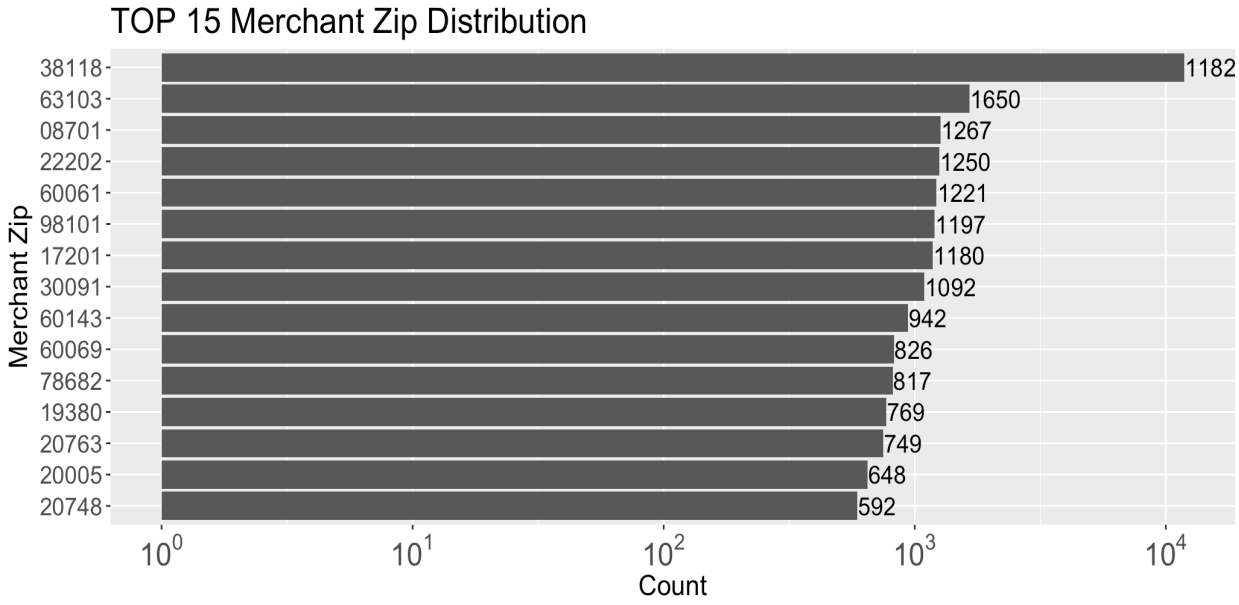


No.	Field Name	# of Unique Values	Percent of populated	Description
6	Merchant State	228	98.76	The state of the merchant

There are 228 merchant states in the dataset which is not consistent with the reality. Typos or other geographical regions in other countries may be involved. Caution and further analysis is needed for this field.



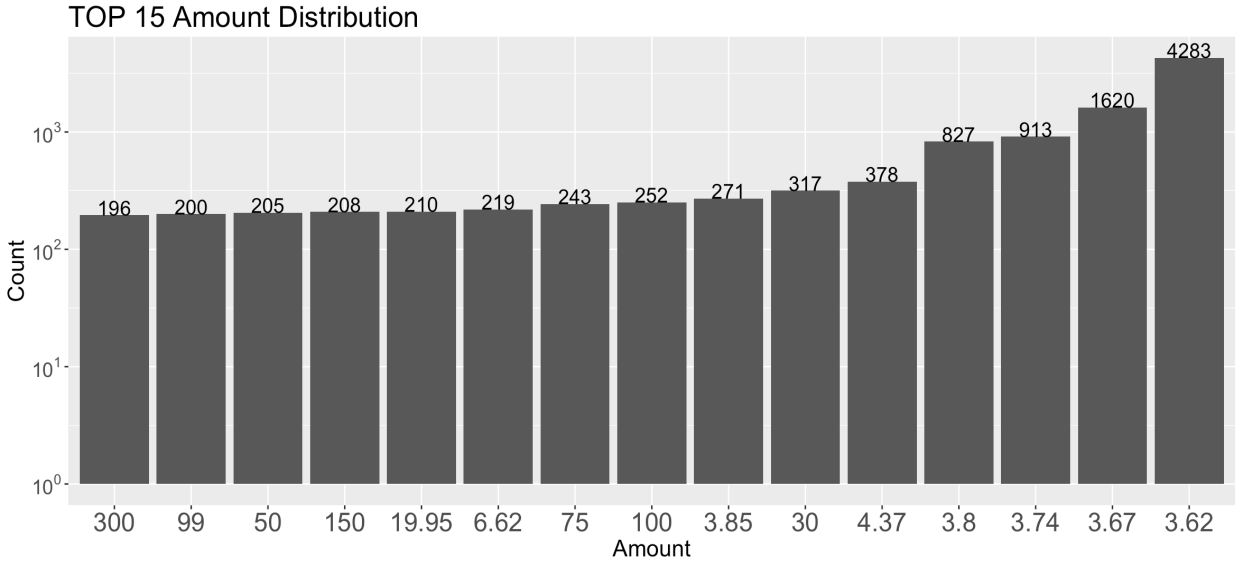
No.	Field Name	# of Unique Values	Percent of populated	Description
7	Merchant Zip	4568	95.19	The zip code of merchant



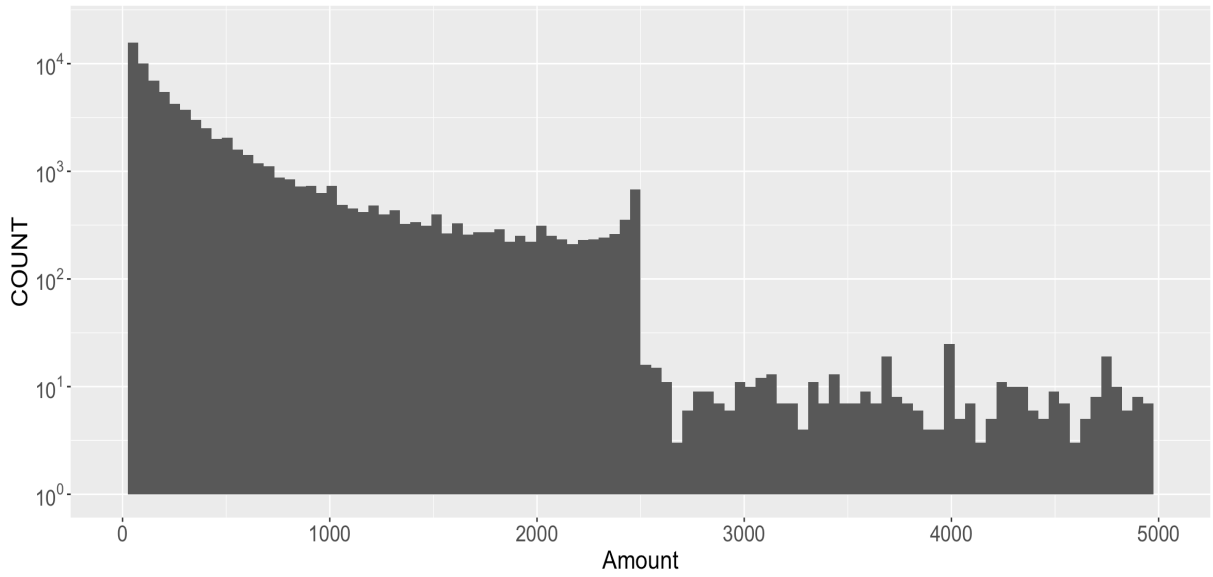
No.	Field Name	# of Unique Values	Percent of populated	Description
8	Transtype	4	100	The transaction type

Transtype	Count
P	96353
A	181
D	173
Y	1

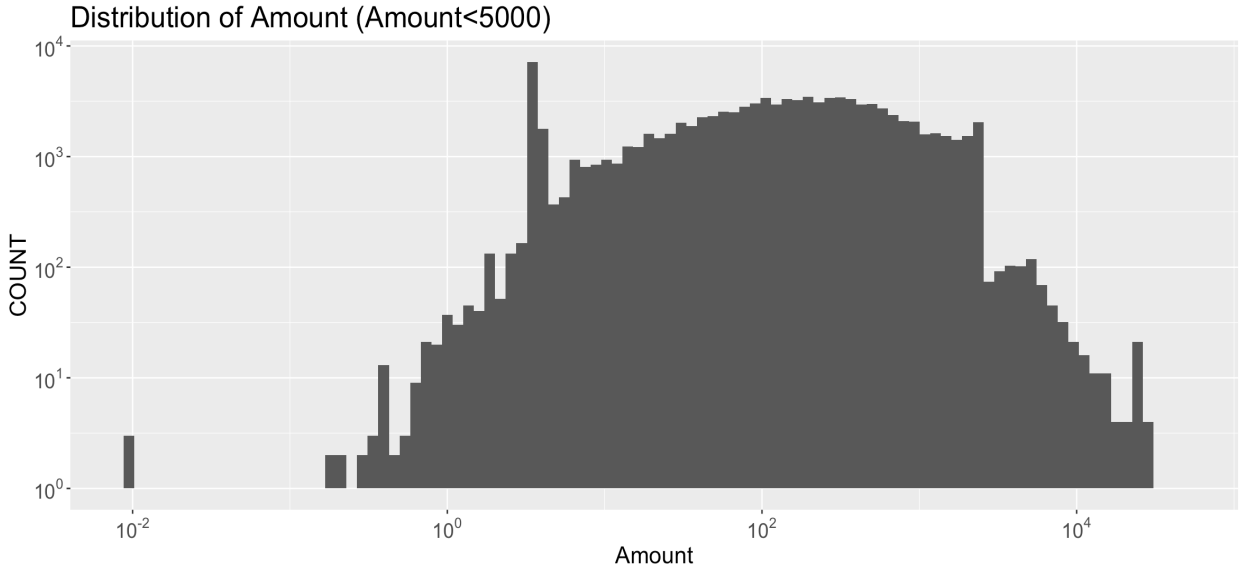
No.	Field Name	# of Unique Values	Percent of populated	Description
9	Amount	34876	100	Transaction amount



a. normal scale



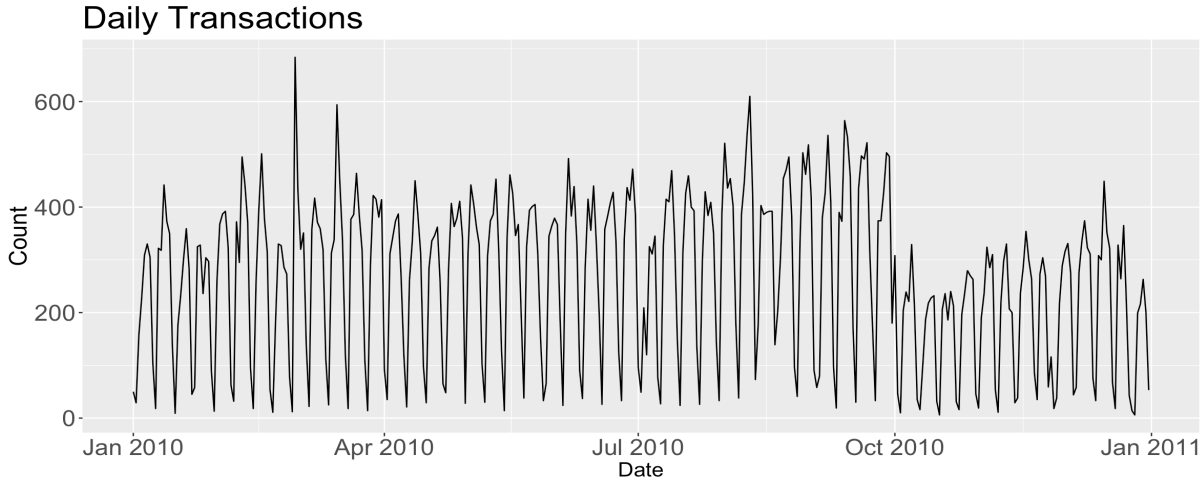
b. log scale



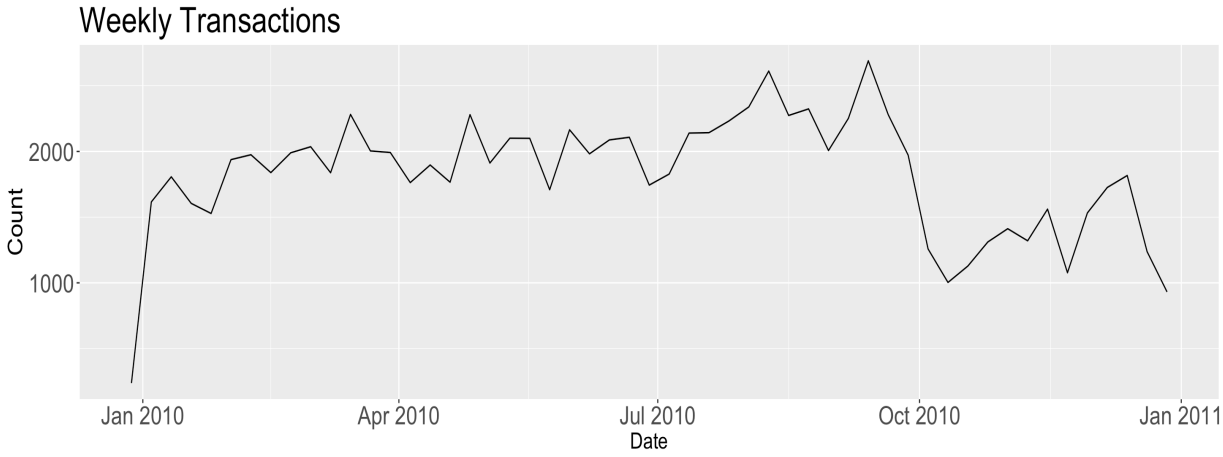
No.	Field Name	# of Unique Values	Percent of populated	Description
10	Fraud	2	100	1: the transaction is fraud 0: the transaction is not fraud

Fraud	Count
1	1014
0	95694

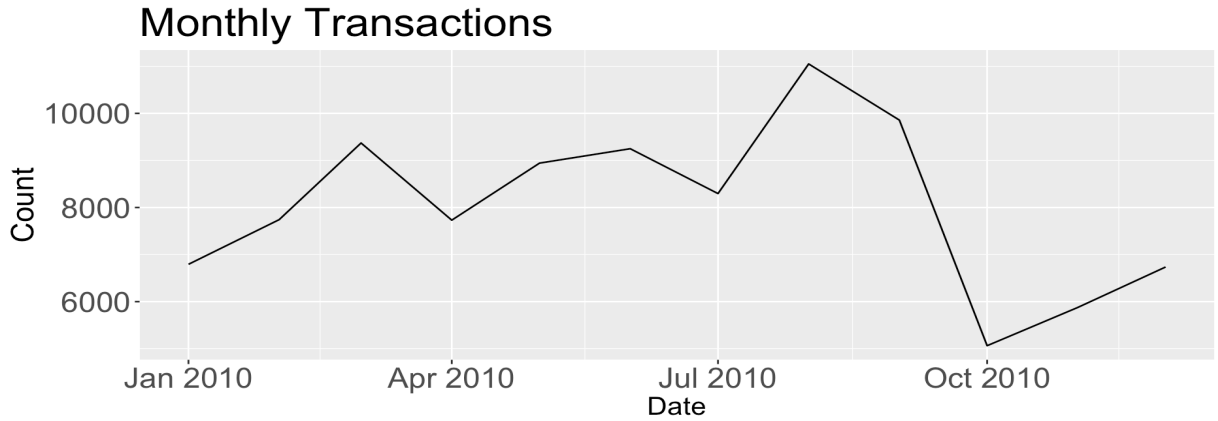
4. Time Series for Number of Transactions:
    a. number of transactions per day



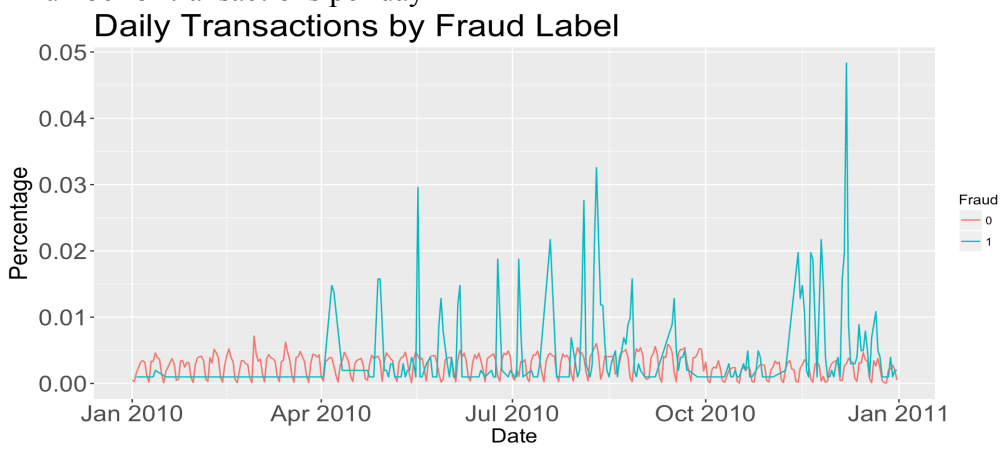
    b. number of transactions per week



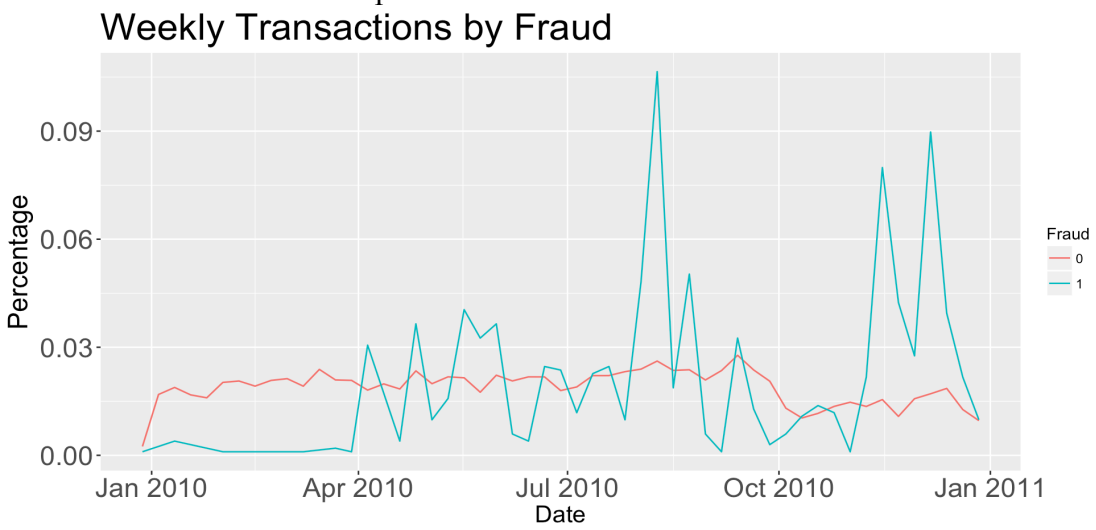
    c. number of transactions per month



5. Fraud Analysis
    a. number of transactions per day



    b. number of transactions per week



    c. number of transactions per month