Data Quality Report

Basic Information on the Dataset

This is a dataset on card transactions for the year of 2010 with a target variable 'fraud'. There are 96,708 records and 10 columns.

Summary statistics for each field

	Recordnum	Cardnum	Date	Merchantnum	Merch Description
count	96708	9.67E+04	96708	93333	96708
top	NaN	NaN	28/2/2010 0:00	9.3009E+11	GSA-FSS-ADV
freq	NaN	NaN	684	9310	1688
first	NaN	NaN	1/1/2010 0:00	NaN	NaN
last	NaN	NaN	31/12/2010 0:00	NaN	NaN
mean	48354.5	5.14E+09	NaN	NaN	NaN
std	27917.33925	5.39E+04	NaN	NaN	NaN
min	1	5.14E+09	NaN	NaN	NaN
25%	24177.75	5.14E+09	NaN	NaN	NaN
50%	48354.5	5.14E+09	NaN	NaN	NaN
75%	72531.25	5.14E+09	NaN	NaN	NaN
max	96708	5.14E+09	NaN	NaN	NaN

	Merchant State	Merchant Zip	Transtype	Amount	Fraud
count	95513	92052	96708	9.67E+04	96708
top	TN	NaN	P	NaN	NaN
freq	11990	NaN	96353	NaN	NaN
first	NaN	NaN	NaN	NaN	NaN
last	NaN	NaN	NaN	NaN	NaN
mean	NaN	44709.8176	NaN	4.28E+02	0.010485
std	NaN	28376.09735	NaN	1.00E+04	0.101859
min	NaN	1	NaN	1.00E-02	0
25%	NaN	20855	NaN	3.35E+01	0
50%	NaN	38118	NaN	1.38E+02	0
75%	NaN	63103	NaN	4.28E+02	0
max	NaN	99999	NaN	3.10E+06	1

Percentage populated and Description

Variables	Percentage populated %	# of Unique Values	# of Missing Values	Description
Recordnum	100	96,708	0	Record (row) number of card transaction

Cardnum	100	1,644	0	Credit card number
Date	100	365	0	Date of the transaction
Merchantnum	96.51	13,091	3,375	Merchant number of the place where the card transaction was made
Merch Description	100	13,125	0	Merchant's name/description
Merchant State	98.76	228	1,195	Merchant's state of address
Merchant Zip	95.19	4,568	4,656	Merchant's zip code
Transtype	100	4	0	Type of card transaction (A, D, P, Y)
Amount	100	34,876	0	Transaction amount in dollars
Fraud	100	2	0	Fraudulent or Not Fraudulent

Distributions

1. **Cardnum**

Top 20 card numbers of transaction count
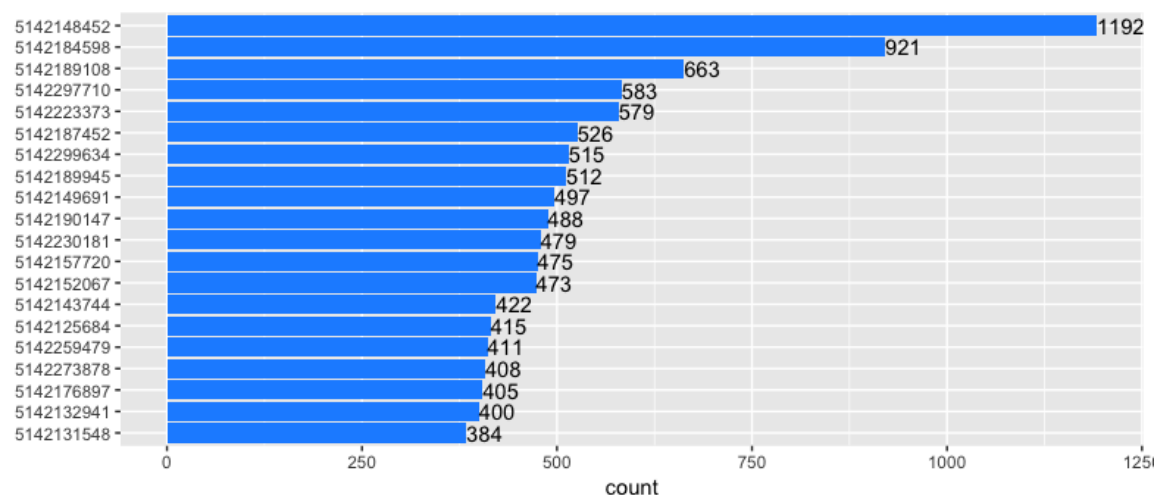
**Top 20 card numbers with highest number of fraudulent transactions**
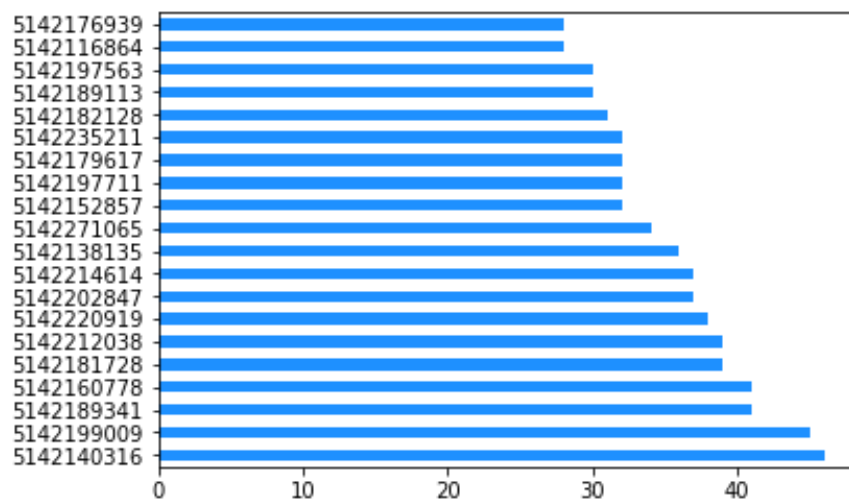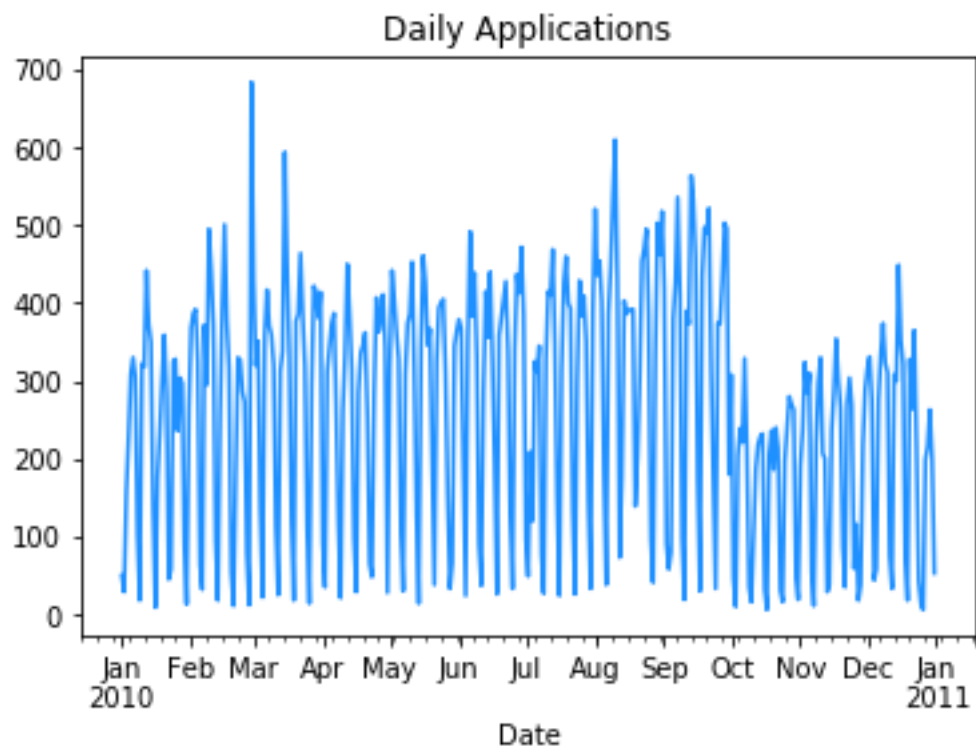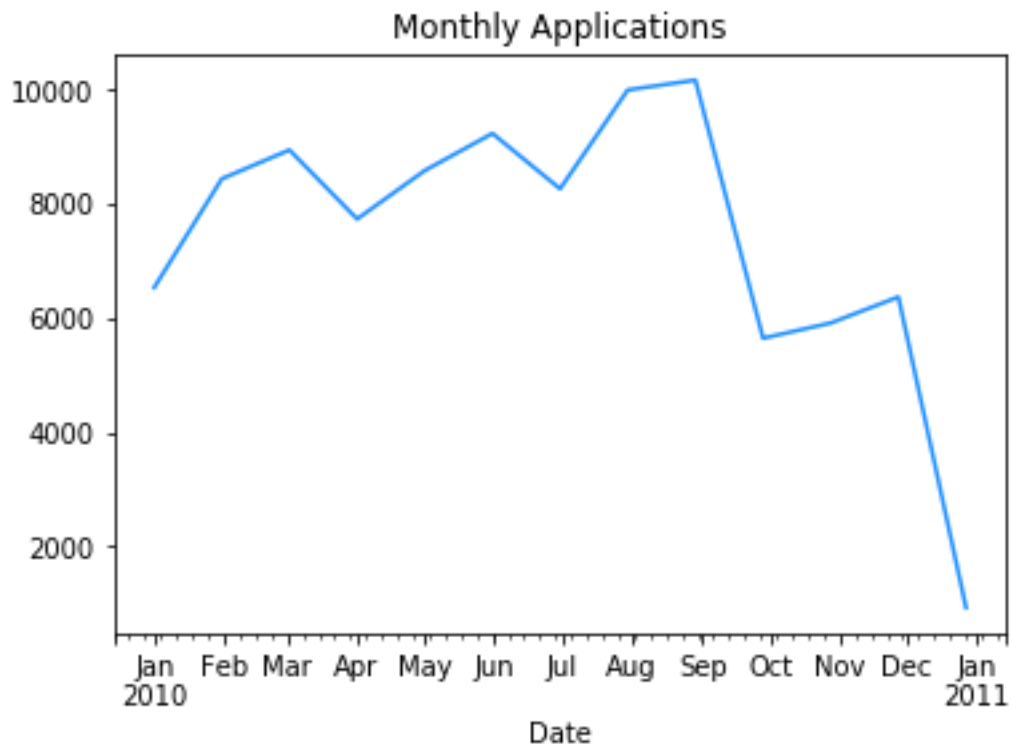
Top 10 card numbers with highest fraud rate

Card Number	Fraud Rate
5142189113	100.00%
5142110434	100.00%
5142117814	100.00%
5142134652	100.00%
5142212038	72.22%
5142199009	62.50%
5142179617	54.24%
5142112858	50.00%
5142228988	50.00%
5142273038	50.00%

2. Date

Top 20 dates of transaction count

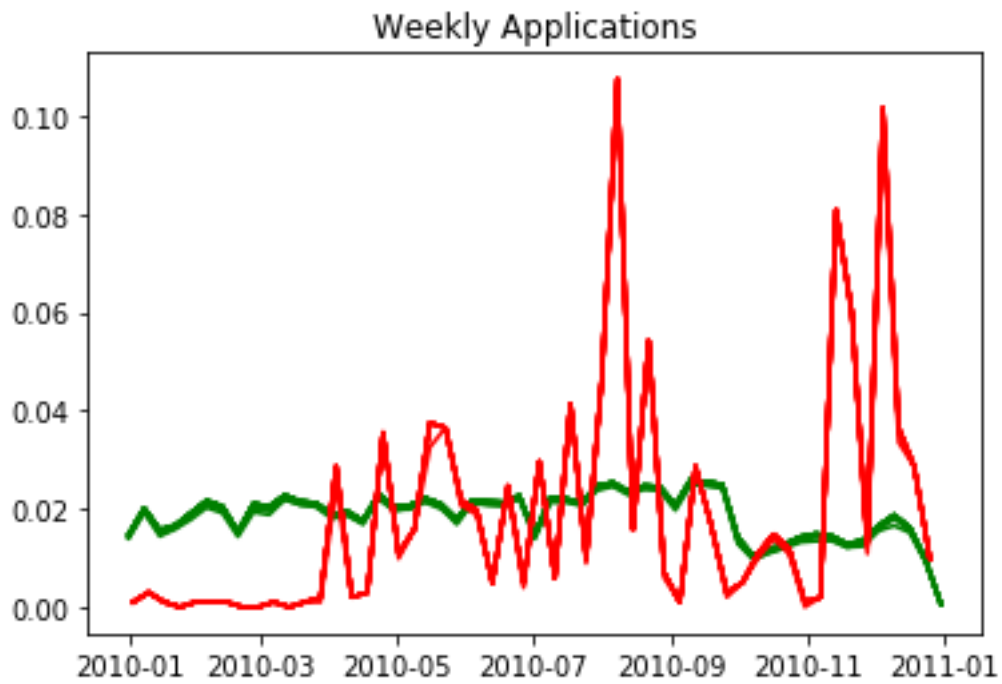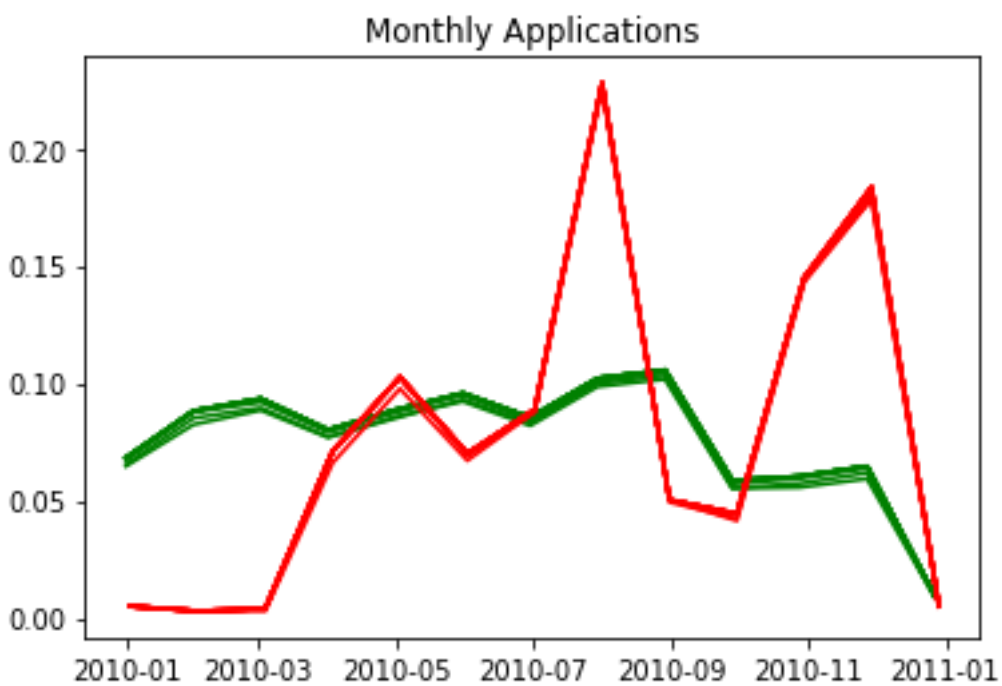Daily Applications

Weekly Applications
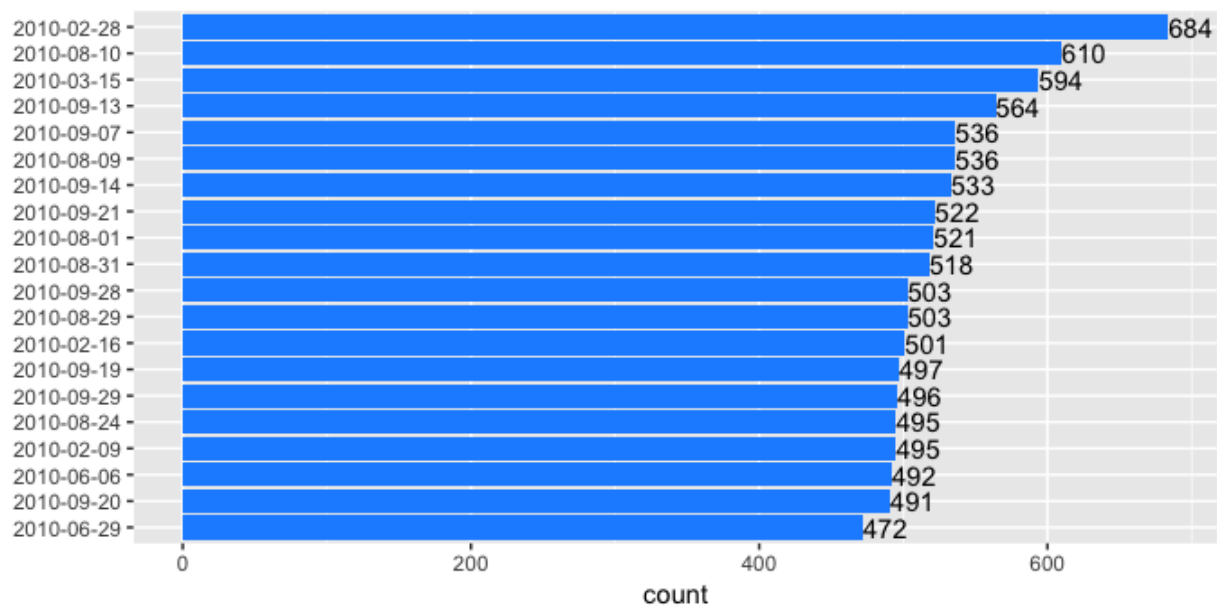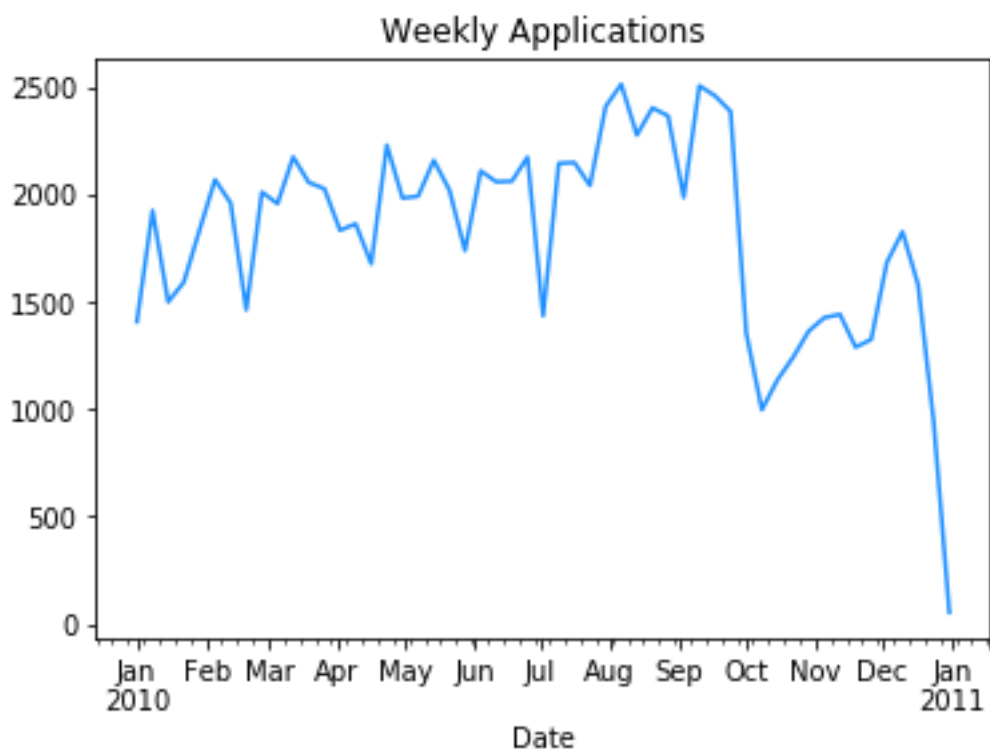
Normalized Daily Applications colored with 'Fraud' (Fraud=1, red; Fraud=0, green)
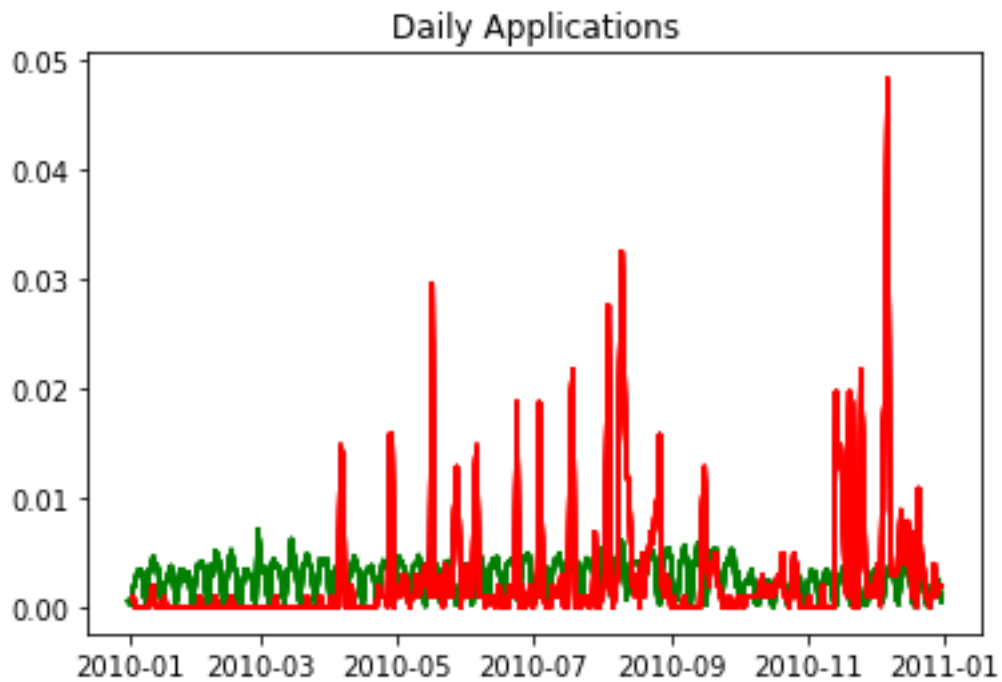
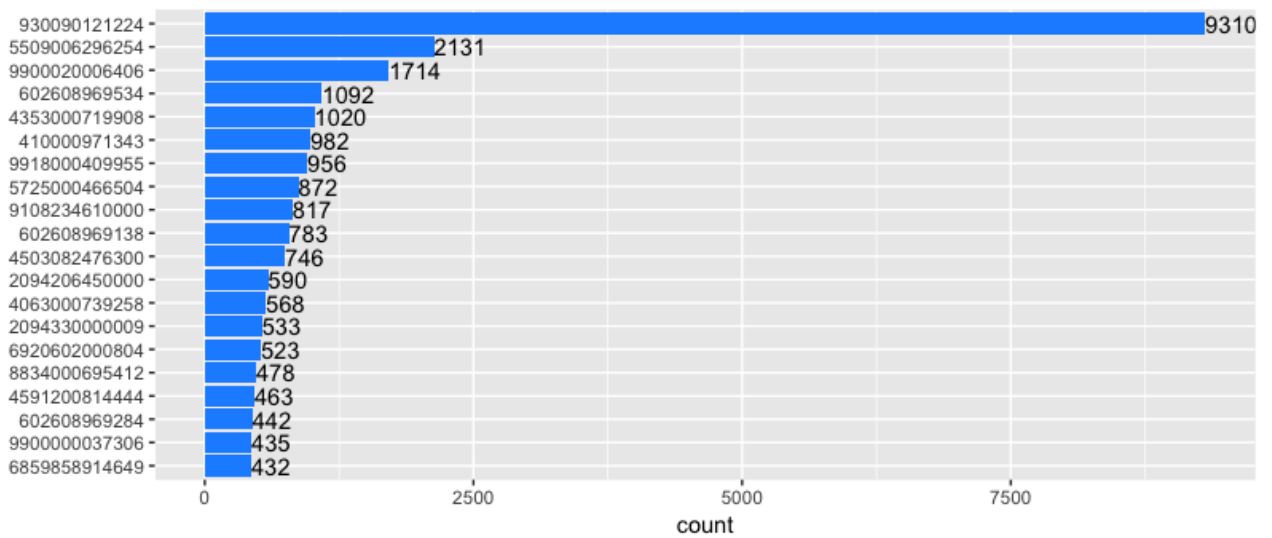**Normalized Weekly Applications colored with 'Fraud' (Fraud=1, red; Fraud=0, green)**

Weekly Applications

Normalized Monthly Applications colored with 'Fraud' (Fraud=1, red; Fraud=0, green)

Monthly Applications

3. **Merchantnum**

   Top 20 merchant numbers of transaction count

	count
930090121224	9310
5509006296254	2131
9900020006406	1714
602608969534	1092
4353000719908	1020
410000971343	982
9918000409955	956
5725000466504	872
9108234610000	817
602608969138	783
4503082476300	746
2094206450000	590
4063000739258	568
2094330000009	533
6920602000804	523
8834000695412	478
4591200814444	463
602608969284	442
9900000037306	435
6859858914649	432

Top 10 merchant numbers with highest fraud rate

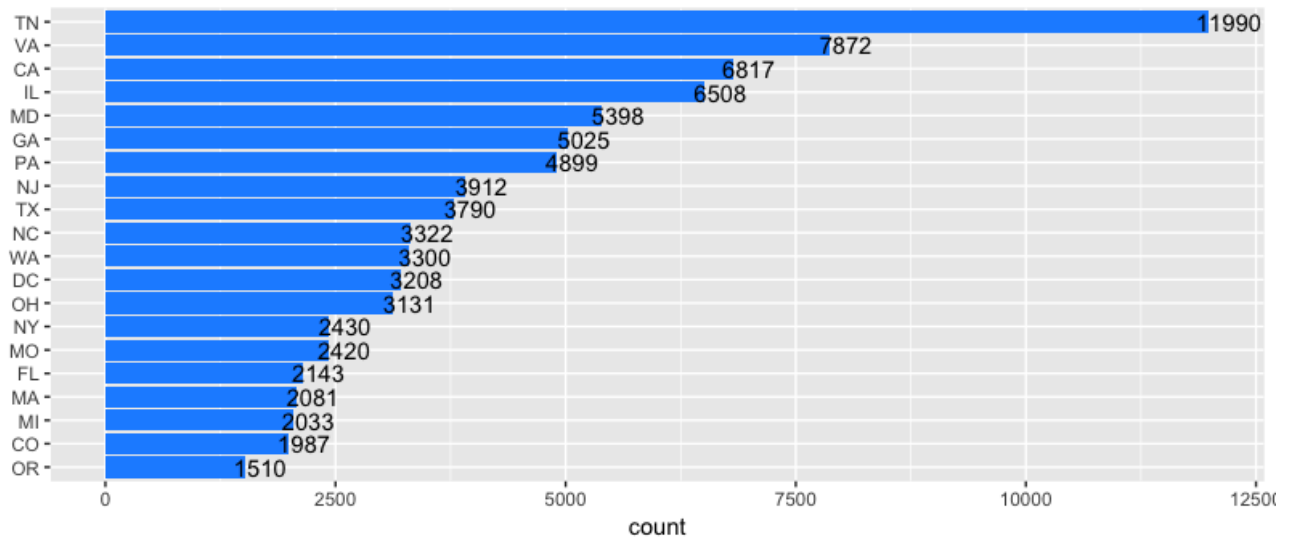| Merchant Number	*Fraud Rate*
8834000695423	***100.00%***
6006333528866	***100.00%***
92891948003	***100.00%***
6070095870009	***93.10%***
938909877224	***76.19%***
8292309000040	***68.97%***
6929	***66.67%***
6005030600003	***55.81%***
08-0091000339	***50.00%***
08-3508354542	***50.00%***

4. Merch Description

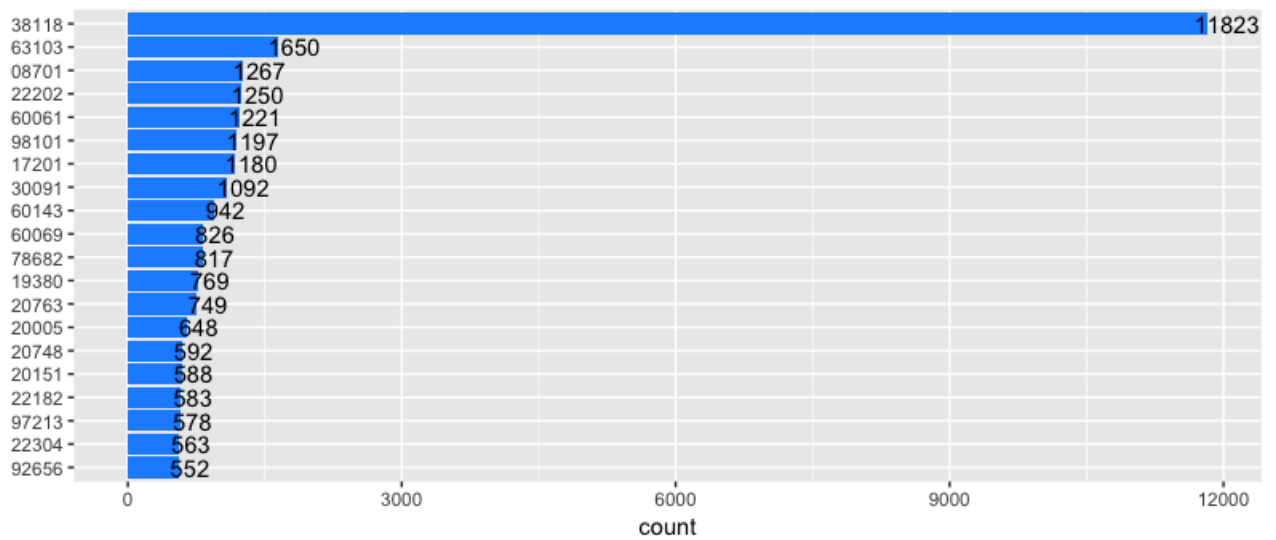**Top 20 merchant descriptions of transaction count**
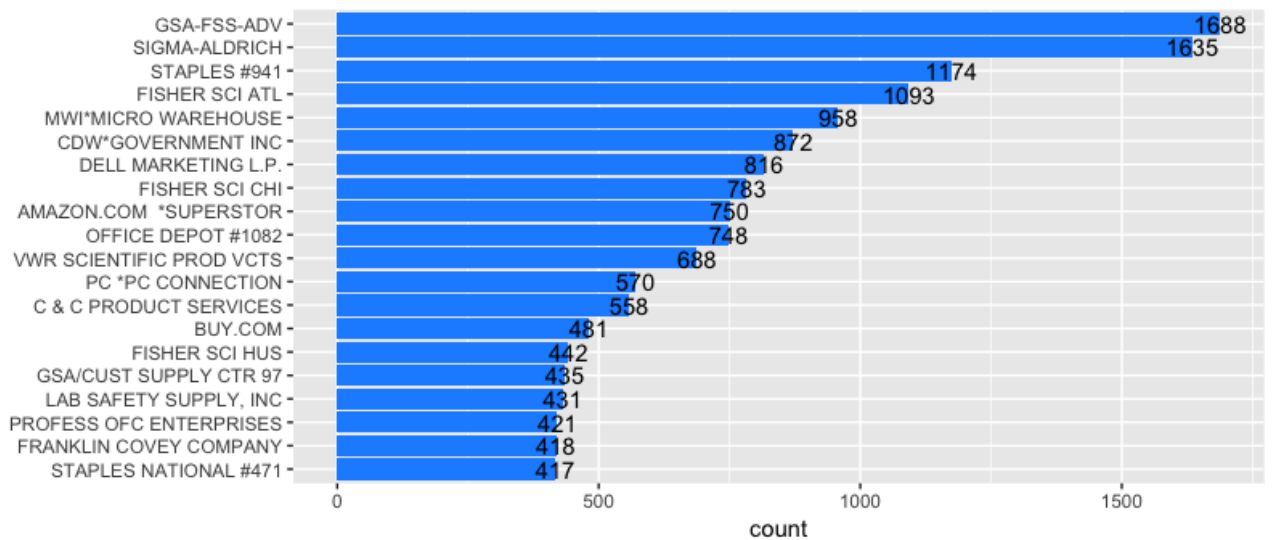
5. Merchant State
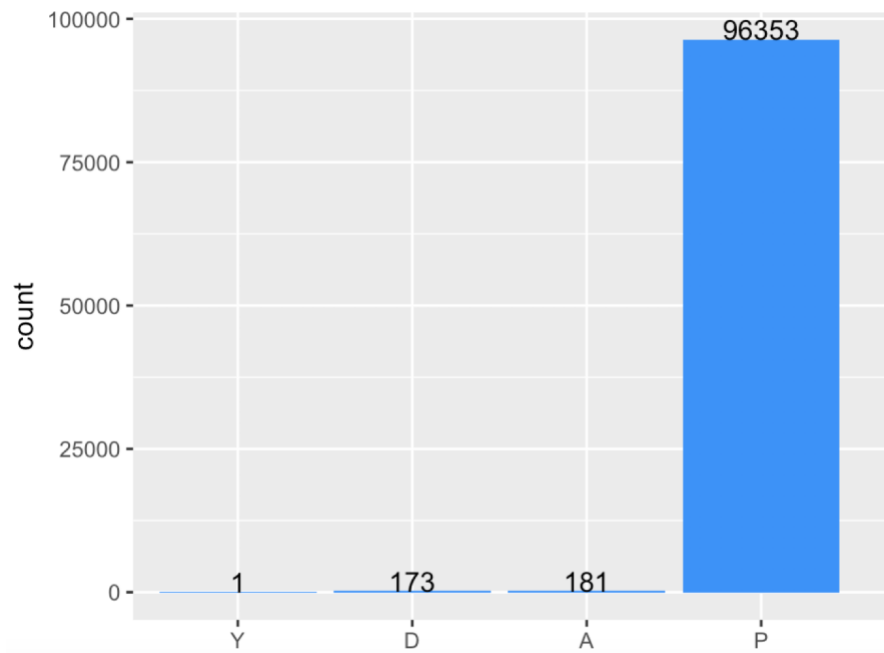
Top 20 merchant states of transaction count

6. Merchant Zip
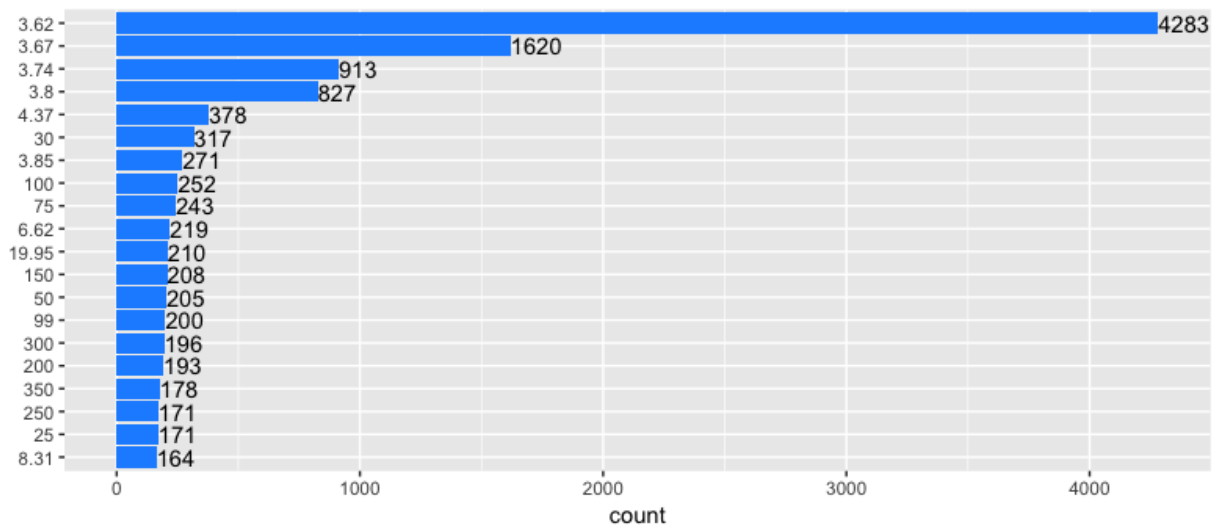
Top 20 merchant zips of transaction count

7. Transtype

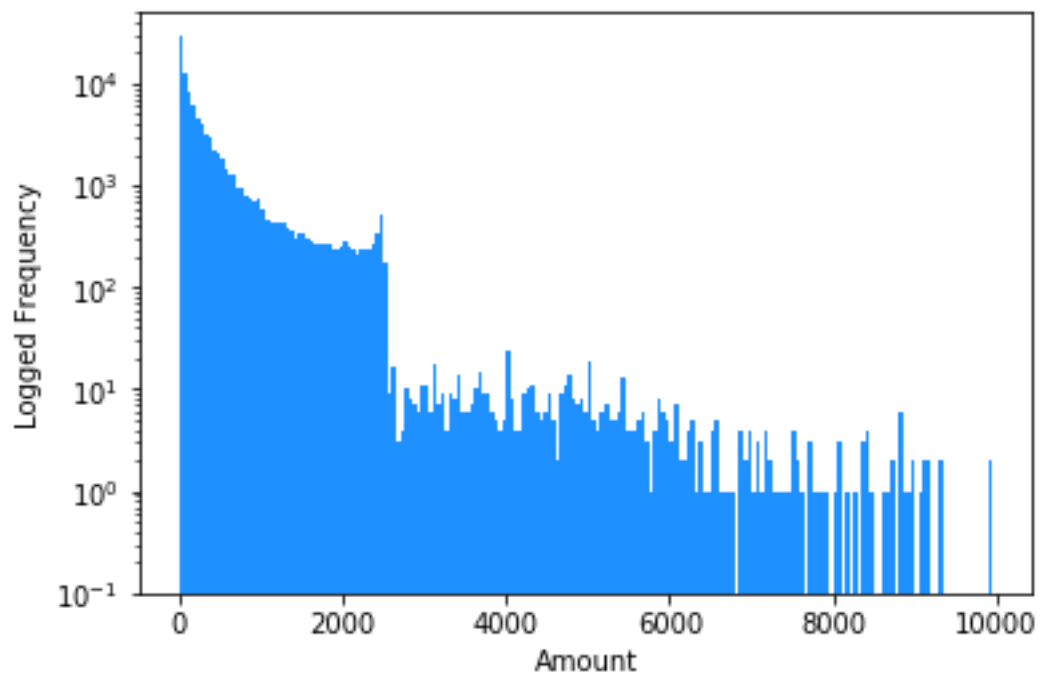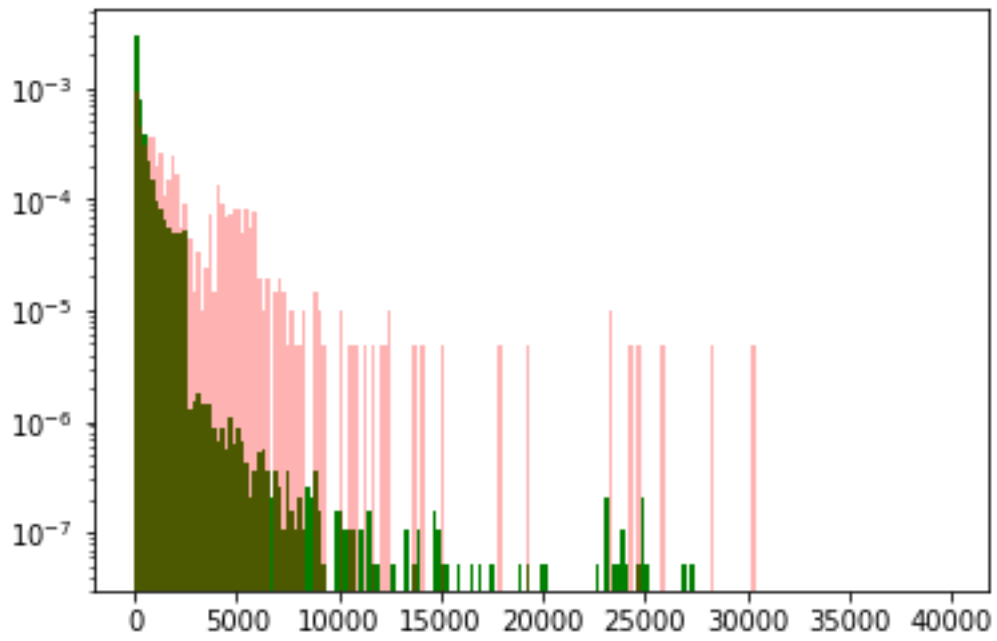**P (purchase) takes the most count.**

8. Amount

Top 20 amounts of transaction count

Logged frequency distribution

Logged frequency distribution colored with 'Fraud' (Fraud=1, red; Fraud=0, green)

9. Fraud