

CREDIT CARD TRANSACTION DATA DATA QUALITY REPORT

SUBMITTED BY – SIDDHARTH JAIN

File Description

File Name: Card transactions.xlsx

Description: File contains records of credit card transaction information and a fraud label for each record

Number of Records: 96,708

Number of Features: 10 (9 features, 1 fraud label)

Data Description

VARIABLE NAME	VARIABLE TYPE	UNIQUE VALUE COUNT	POPULATED %
1.Recordnum	Numeric	96708	100%
2.Cardnum	Numeric	1644	100%
3.Date	Date	365	100%
4. Merchantnum	Categorical/String	13091	96.51%
5. Merch Description	Text	13125	100%
6. Merchant State	Categorical/String	228	98.76%
7. Merchant Zip	Categorical/String	4568	95.19%
8. Transtype	Categorical/String	4	100%
9. Amount	Numeric	34876	100%
10.Fraud	Categorical	2	100%

Feature Description:

1. Record

Description: Manually added field to uniquely identify each observation.

Percent Populated: 100%

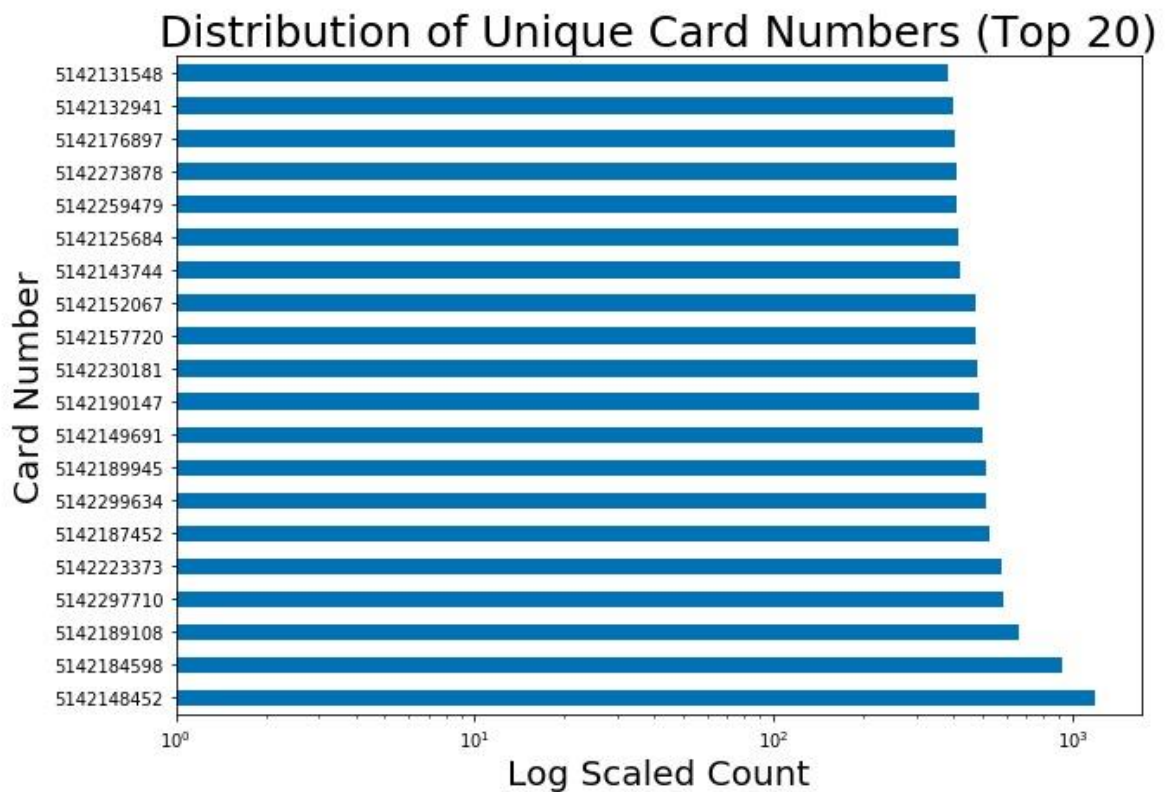
Unique Values: 96,708

2. Cardnum

Description: Credit card number used for transaction.

Percent Populated: 100%

Unique Values: 1644

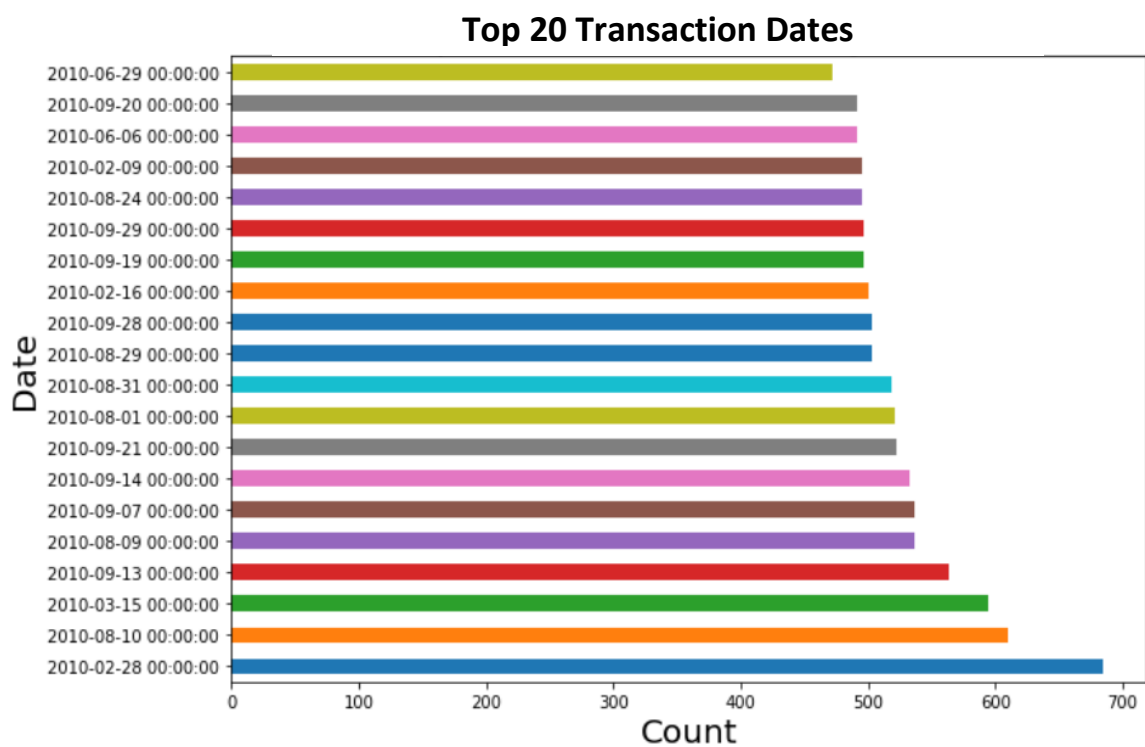
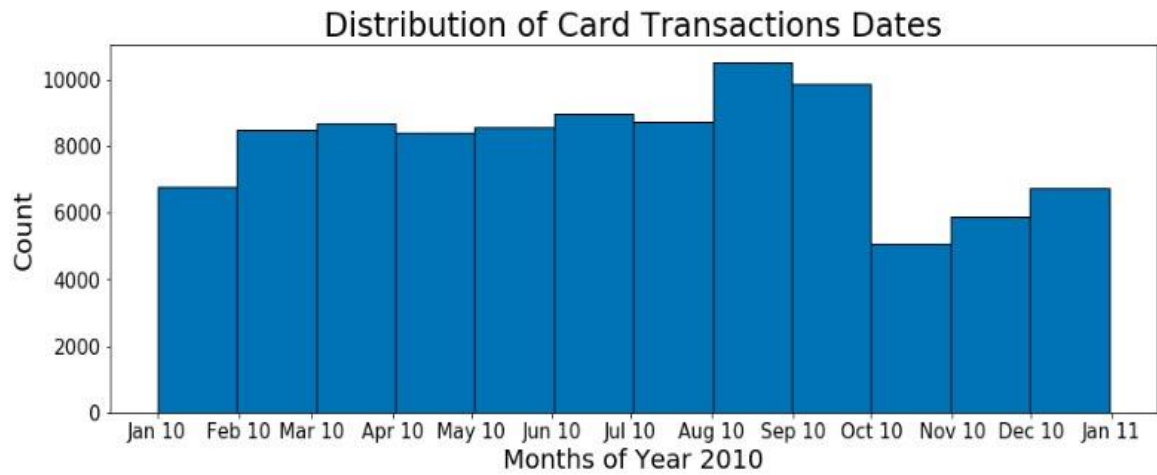


3. Date

Description: Date on which the transaction occurred.

Percent Populated: 100%

Unique Values: 365

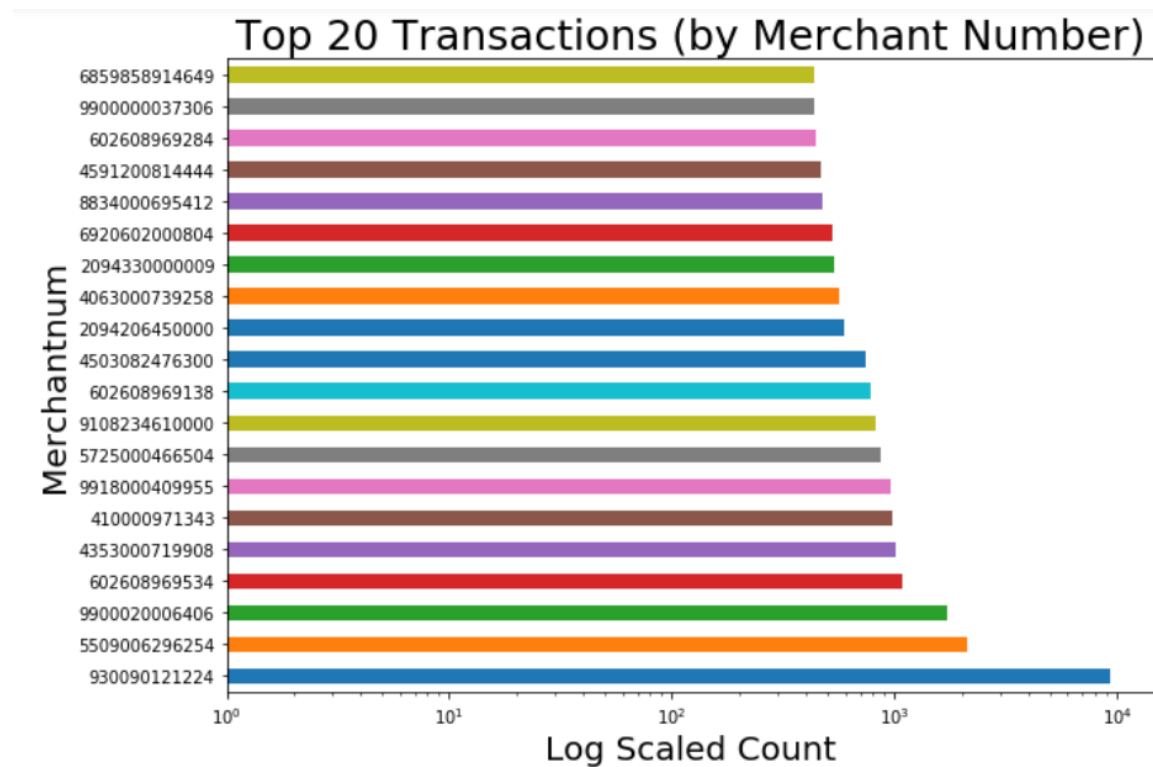


4. Merchantnum

Description: Unique number to identify merchants.

Percent Populated: 96.51%

Unique Values: 13,091

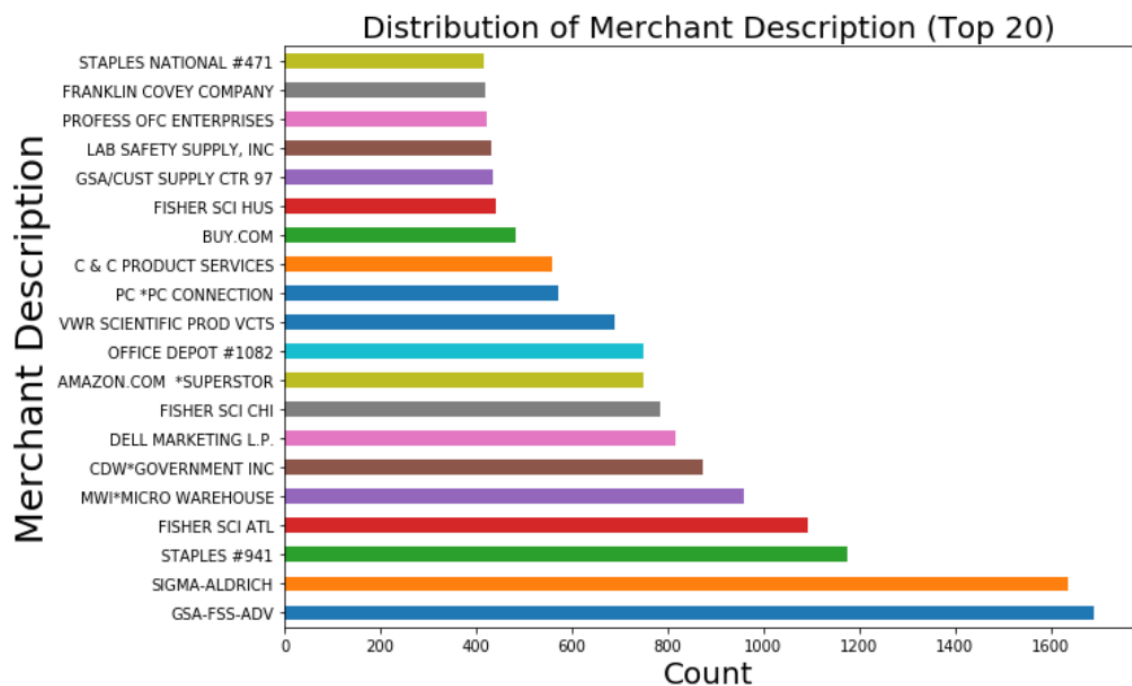


5. Merch Description

Description: Details about merchant

Percent Populated: 100%

Unique Values: 13,125

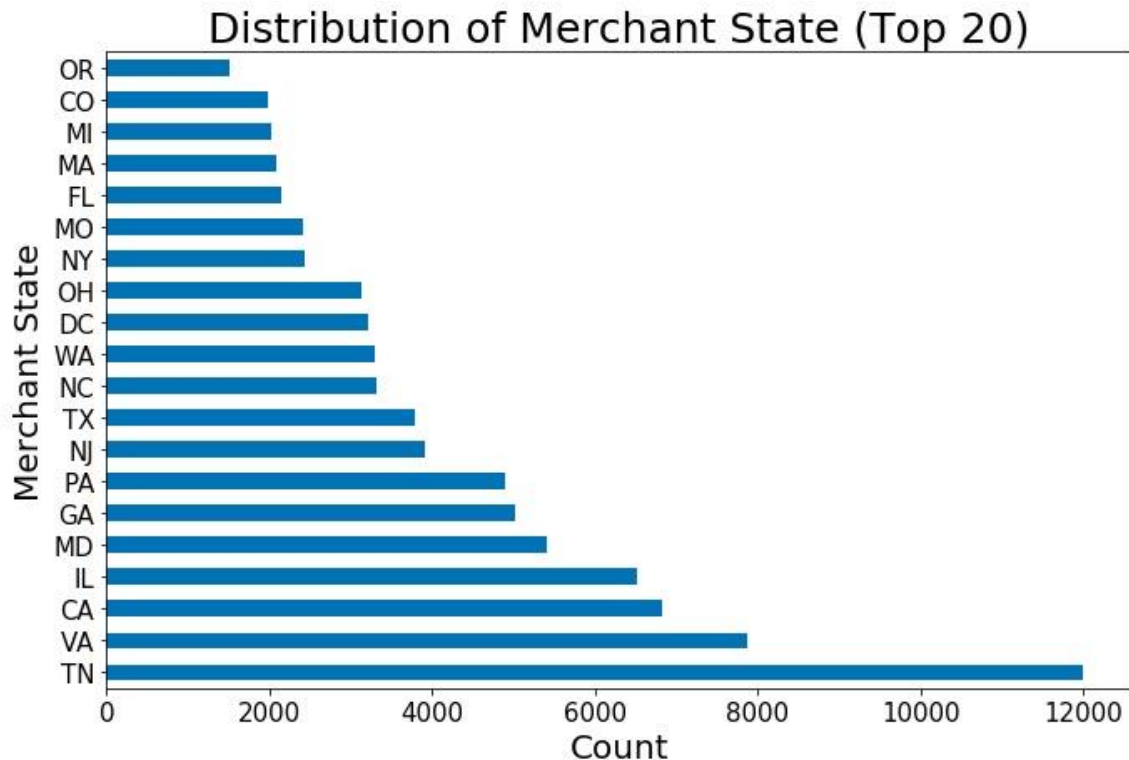


6. Merchant State

Description: State to which the merchant belongs

Percent Populated: 98.76%

Unique Values: 228



An interesting thing to notice is that for this field, we have numerical entries as well.

```
In [11]: data['Merchant State'].value_counts().tail(10)
```

```
Out[11]: 450      1
          085      1
          108      1
          799      1
          737      1
          443      1
          650      1
          180      1
          718      1
          329      1
          Name: Merchant State, dtype: int64
```

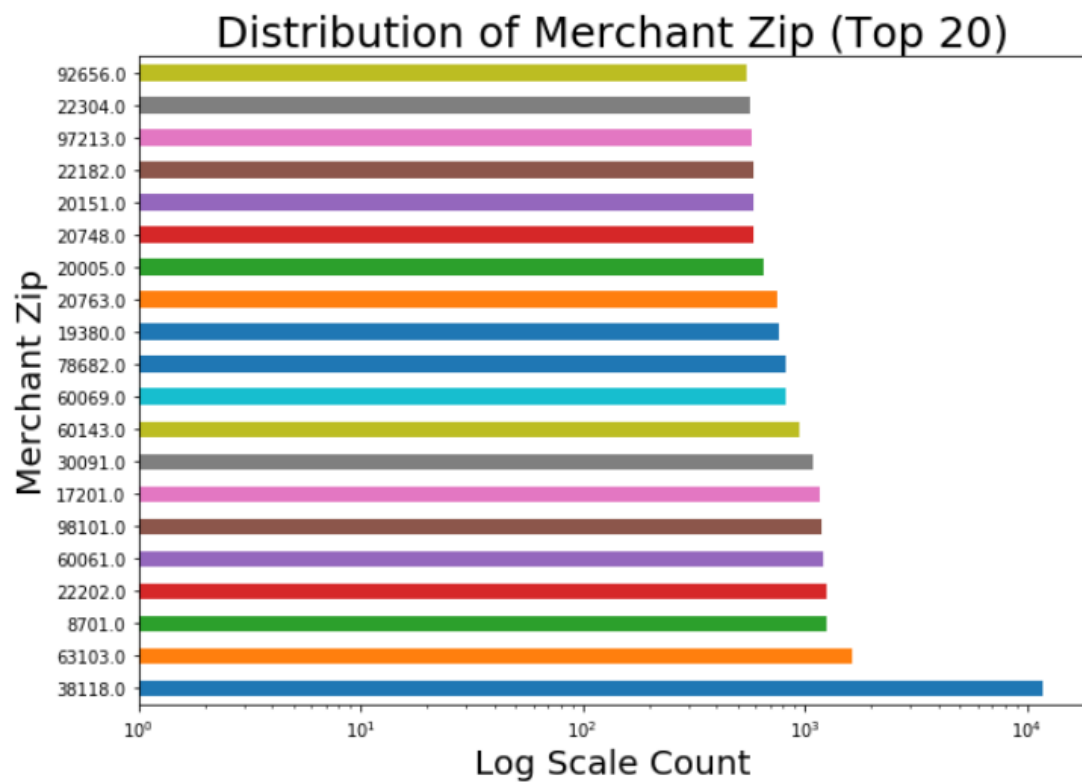
For all these records, the field **merchantnum** is 0.

7. MerchantZip

Description: Zip of the merchant

Percent Populated: 95.19%

Unique Values: 4,568

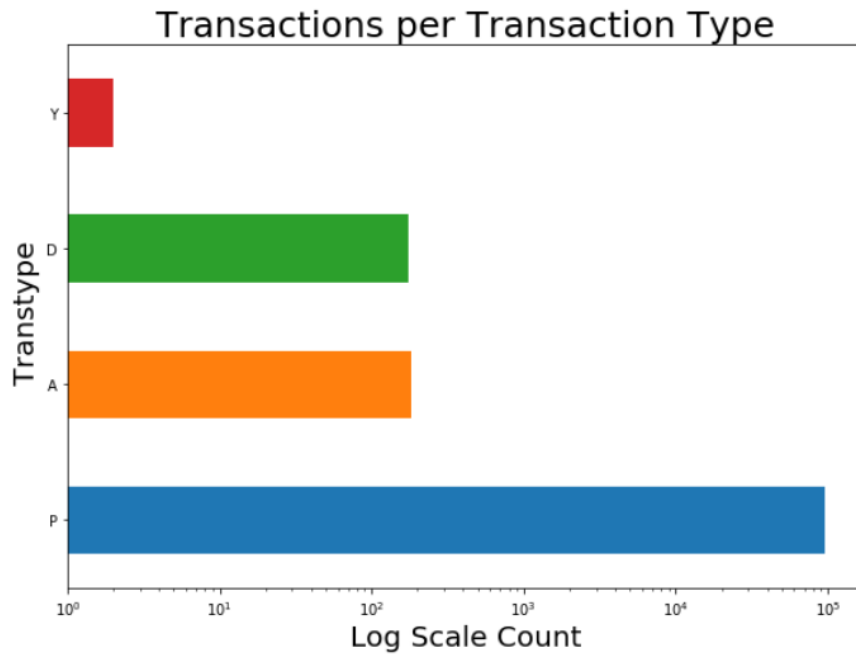


8. Transtype

Description: Classification of transaction.

Percent Populated: 100%

Unique Values: 4



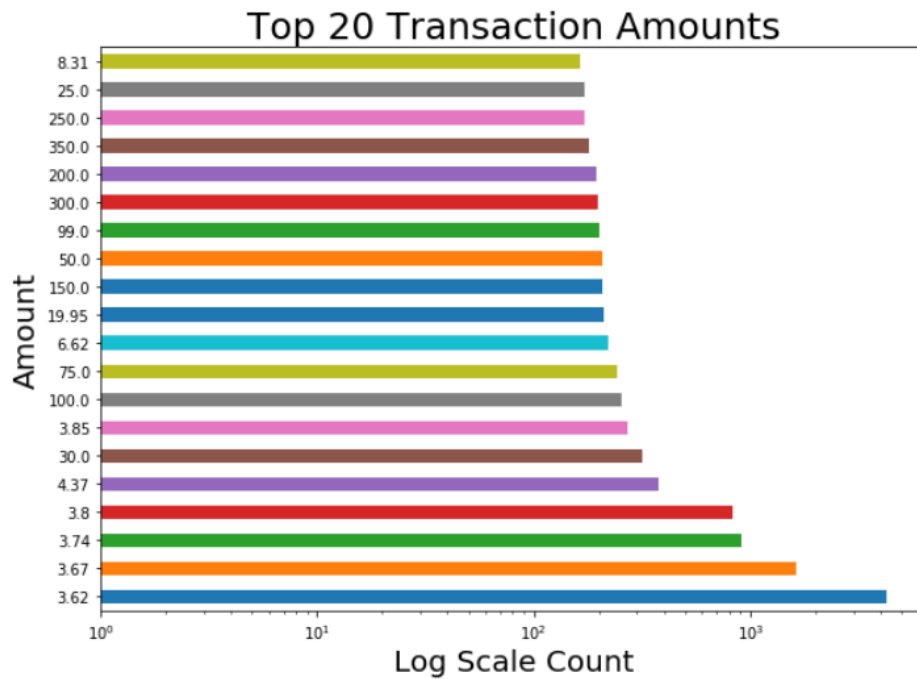
For our analysis, we will only consider the transactions that were processed (**P**) as the other type of transactions have not gone through.

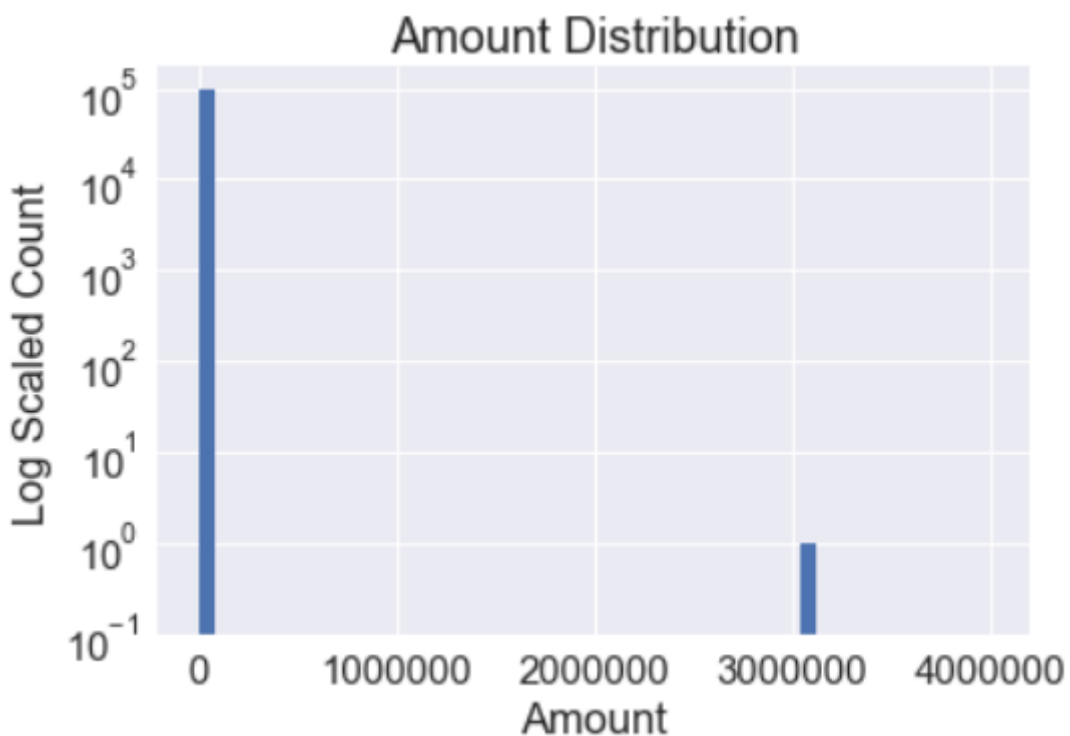
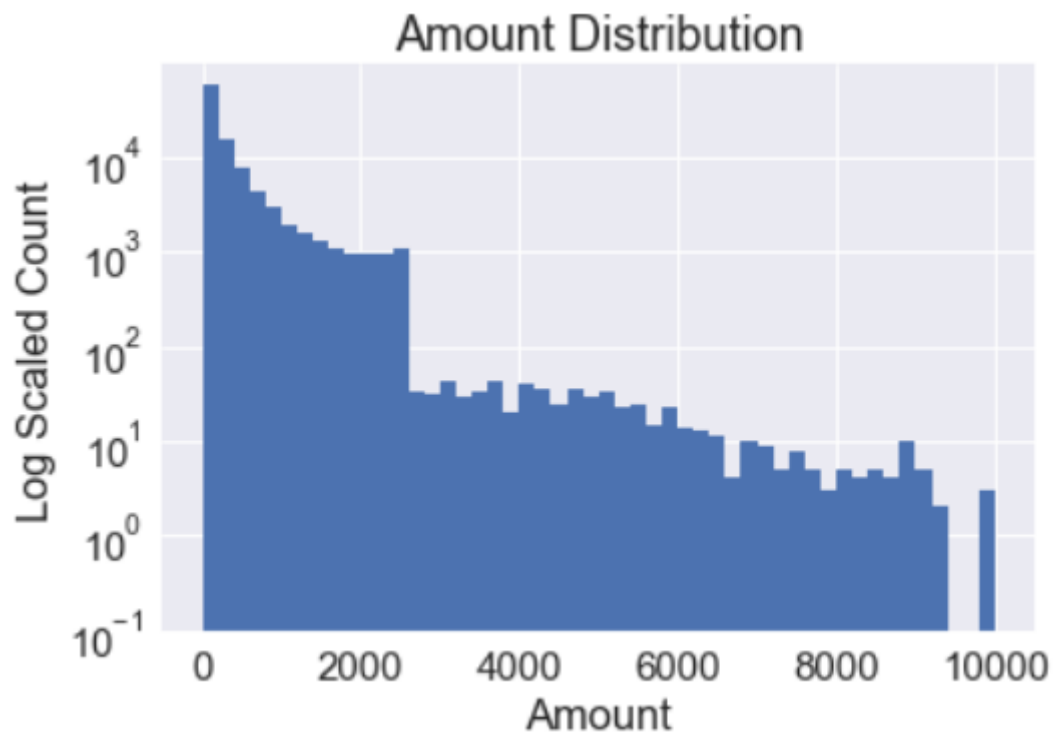
9. Amount

Description: Transaction amount.

Percent Populated: 100%

Unique Values: 34,876





The above graph shows us that one transaction has an abnormally large amount associated with it. Further analysis shows that this value is 3,102,046. It is noticed that this transaction was done with a Mexican business and hence the payment was made in Pesos and not in Dollars,

explaining the high value. We will not consider this transaction for further analysis.

10.Fraud

Description: Dependent variable (Telling whether the record is fraud or not)

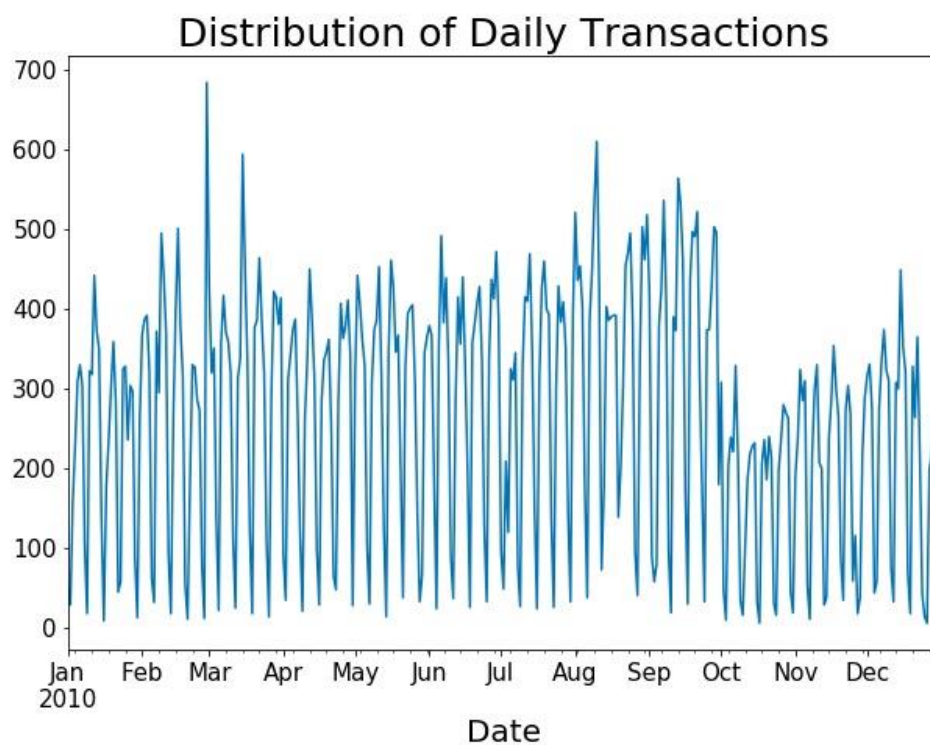
Percent Populated: 100%

Unique Values: 2 (0 – No fraud, 1- Fraud)

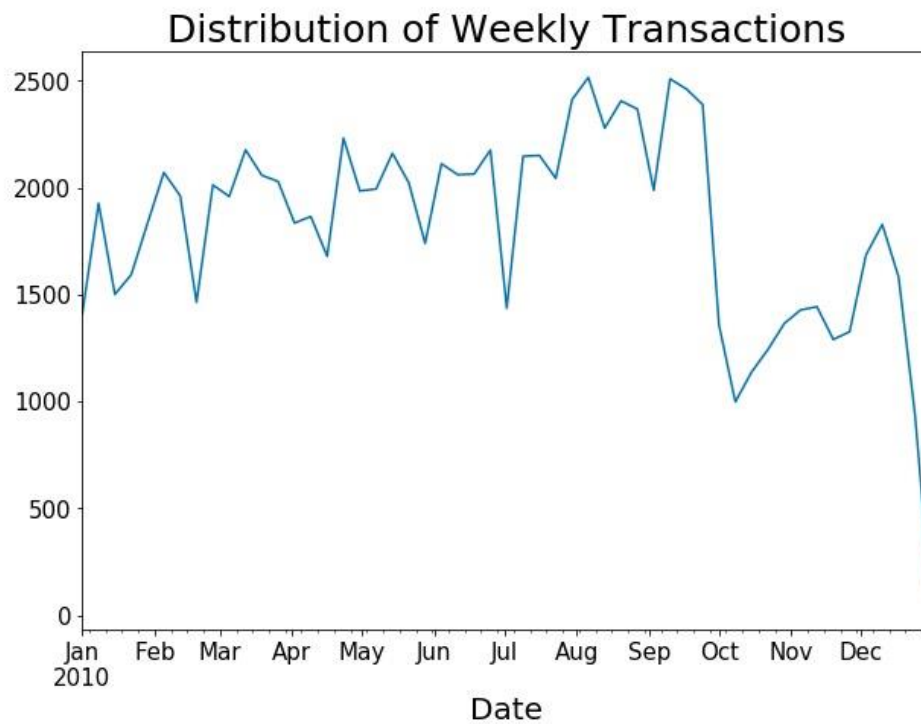
Specific Count: 0 – 95,694; 1 – 1,014

Time Based Analysis:

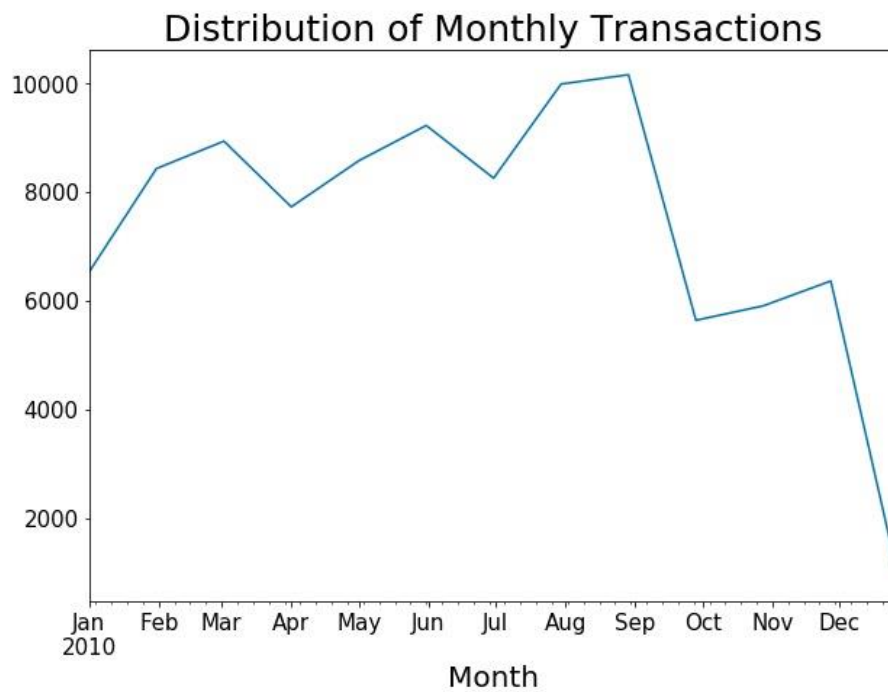
- 1) Number of card transactions received per day



2) Number of card transactions received per week

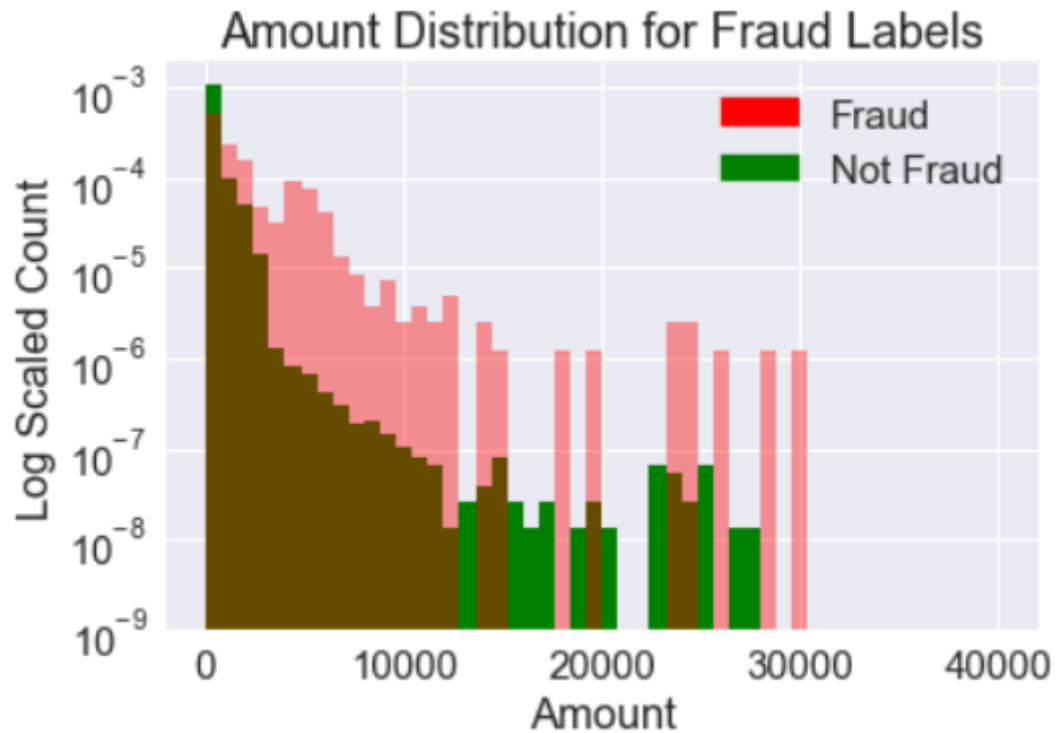


3) Number of card transactions received per month

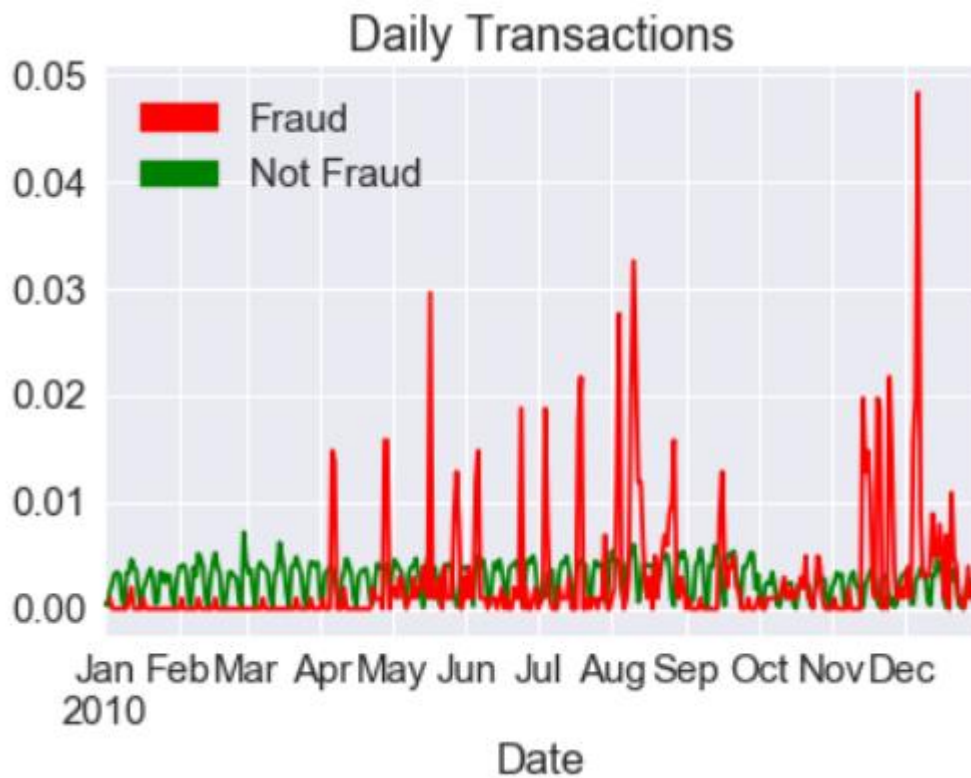


Fraud Label Distribution

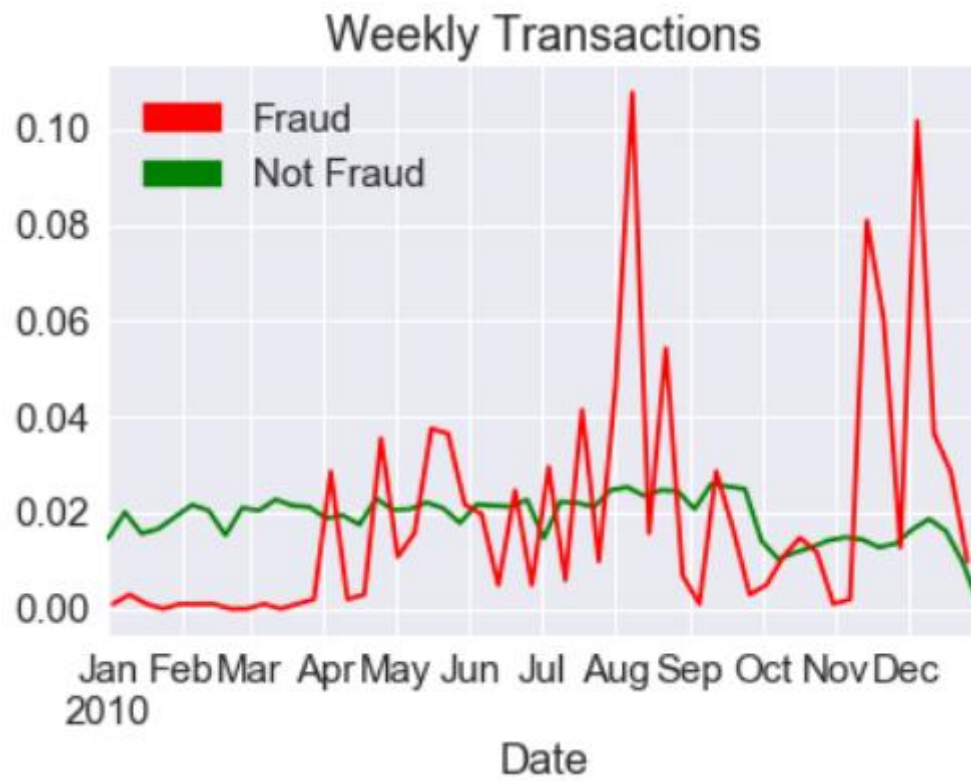
1. Difference based on amount distributions



2. Different based on daily transactions



3. Different based on weekly transactions



4. Different based on monthly transactions

