DATA QUALITY REPORT

Credit Card Transaction Data

By Shiyi (Letty) Hua

Section 1 – Introduction of Data

1.1 **Name of Dataset:** Card Transactions

1.2 **Description:** This is credit card transaction data. To be more specific, each row in the dataset is a record of a credit card transaction on a specific date with a specific merchant, including the transaction amount. The data also indicates whether each transaction is fraud or not.

1.3 **Time Period:** 01/01/2010 – 12/31/2010

1.4 **Number of Fields:** 10

1.5 **Number of Records:** 96,753

Section 2 – Summary of Fields

Table 2.1 Information of All Fields:

Field Name	Field Type	# Non-null Records	% Populated	# Zero Records
Recnum	int64	96,753	100	0
Cardnum	int64	96,753	100	0
Date	datetime64[ns]	96,753	100	0
Merchnum	object	93,378	96.51	231
Merch description	object	96,753	100	0
Merch state	object	95,558	98.76	0
Merch zip	float64	92,097	95.19	0
Transtype	object	96,753	100	0
Amount	float64	96,753	100	0
Fraud	int64	96,753	100	95,694

Table 2.2 Summary Statistics of Numeric Field:

Field Name	Mean	Standard Deviation	Min	Max
Amount	427.89	10,006.14	0.01	3,102,046

Table 2.3 Summary Information of Categorical Fields:

Field Name	# Unique Values	Most Common Field Value
Recnum	96,753	N/A
Cardnum	1,645	5142148452
Date	365	2010-02-28
Merchnum	13,092	930090121224
Merch description	13,126	GSA-FSS-ADV
Merch state	228	TN
Merch zip	4568	38118.0
Transtype	4	P
Fraud	2	0

Section 3 – Description of Each Field

3.1 **Recnum:** Unique integer identifier to label each record of the dataset

3.2 **Cardnum:** Credit card number for each transaction. Some credit cards transact for multiple times. Below is a bar chart of top 15 credit card numbers with their frequencies.
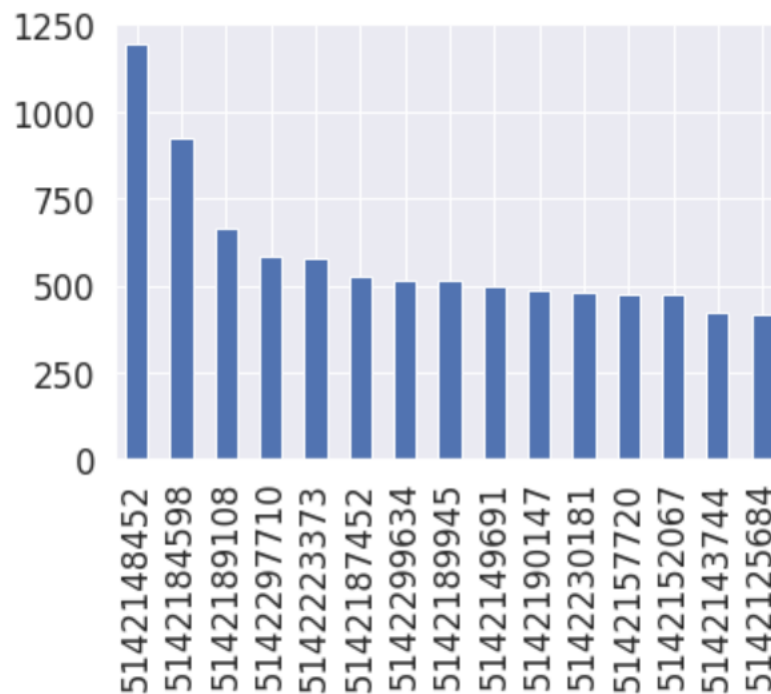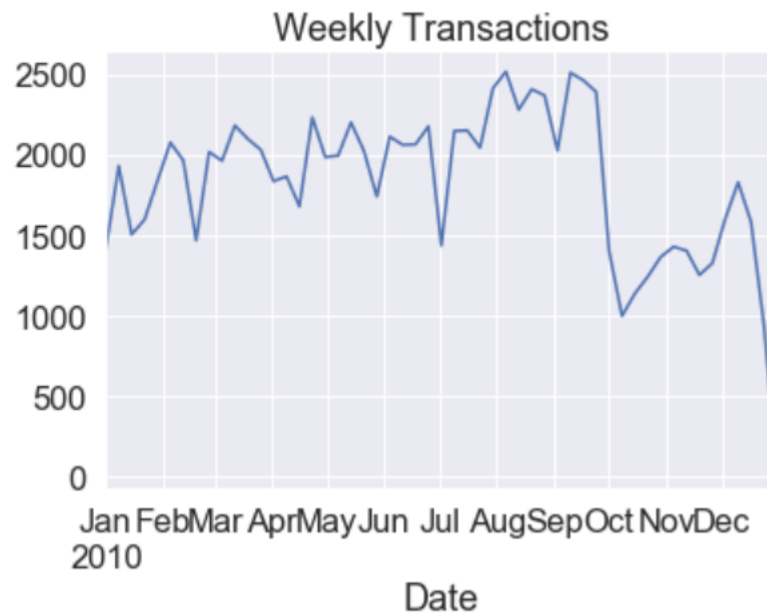
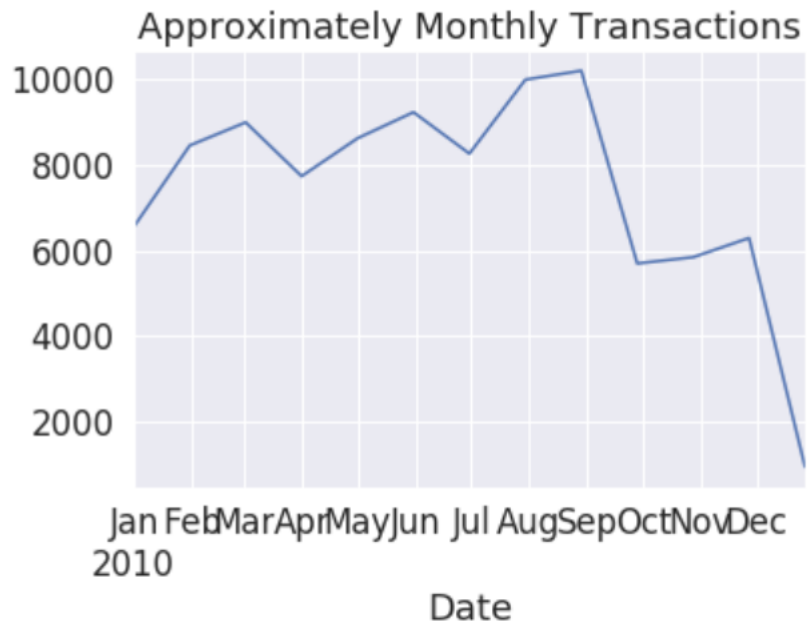3.3 **Date:** Transaction date. Below is a bar char of top 15 dates with their frequencies.
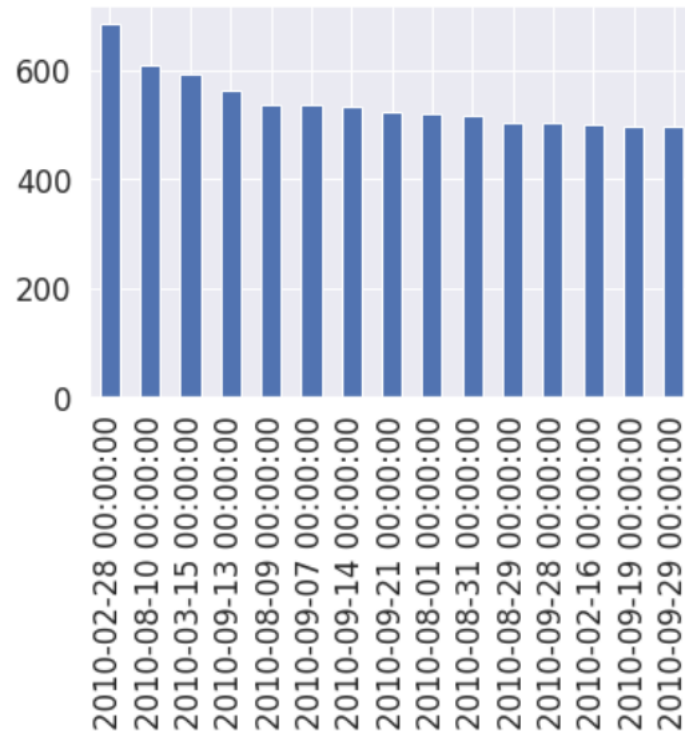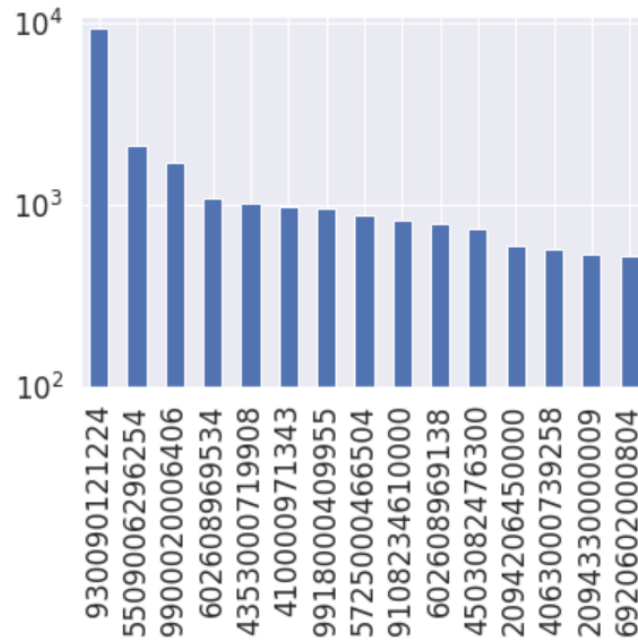

Below is a line plot of daily transactions.
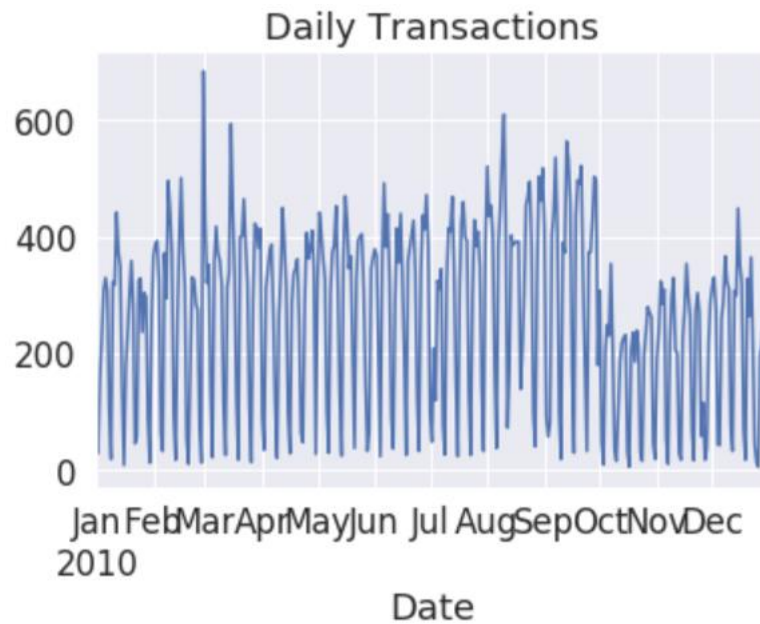
Below is a line plot of weekly transactions.

Weekly Transactions

Below is a line plot of approximately monthly transactions.

**Approximately Monthly Transactions**

3.4 **Merchnum:** Merchant bank account number for each transaction. Below is a bar chart of top 15 merchant bank account numbers with their frequencies.

3.5 **Merch description:** Merchant name/(online) address for each transaction. Below is a bar chart of top 15 merchant names with their frequencies.
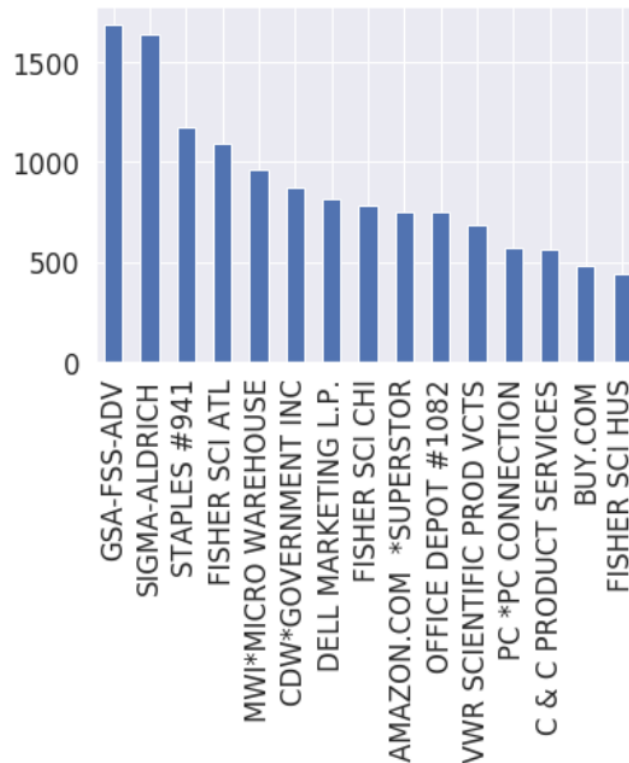
3.6 **Merch state:** State where the merchant locates for each transaction record. Below is a bar chart of top 15 states with their frequencies.
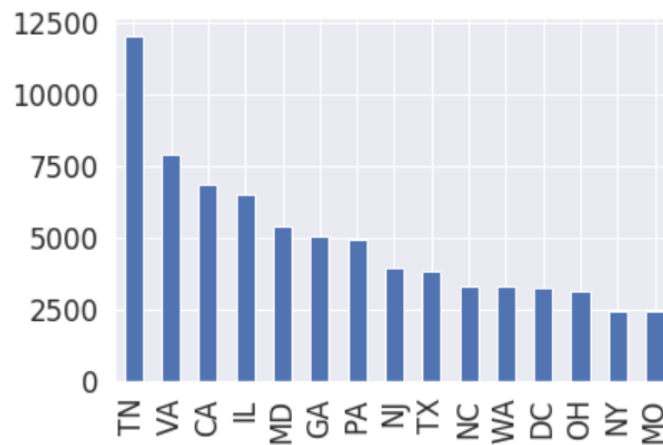

3.7 **Merch zip:** Zip code for the location of the merchant for each transaction. Below is a table of top 15 zip codes with their frequencies.

38118.0	11868
63103.0	1650
8701.0	1267
22202.0	1250
60061.0	1221
98101.0	1197
17201.0	1180
30091.0	1092
60143.0	942
60069.0	826
78682.0	817
19380.0	769
20763.0	749
20005.0	648
20748.0	592

3.8 **Transtype:** Transaction type for each record. There are 4 types. Below is a table of each transaction type with its frequency.

P	96398
A	181
D	173
Y	1

3.9 **Amount:** Transaction amount spent for each record. Below is a distribution plot of transaction spending amount. Selected values are under 4000.
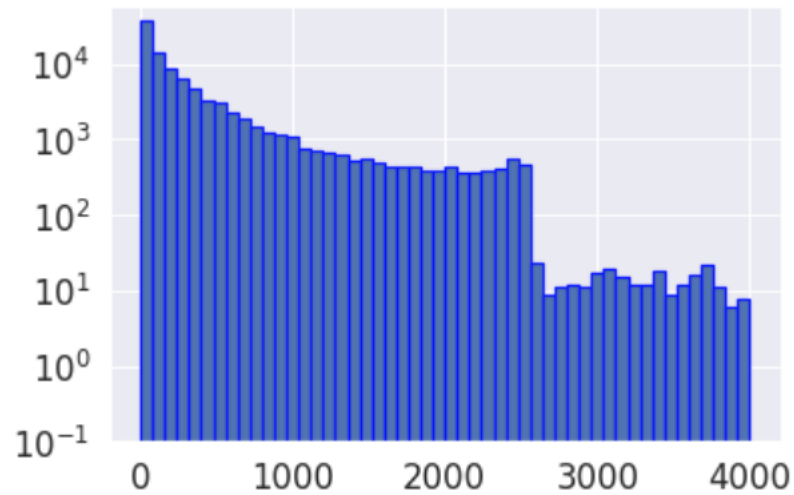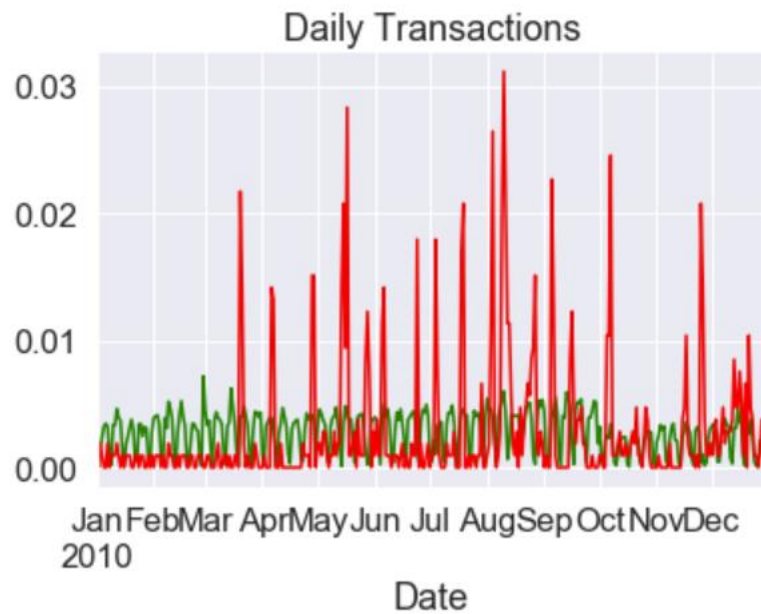

3.10 **Fraud:** Dummy label that indicates whether each transaction record is fraud or not. Below is a table of fraud transaction and non-fraud transaction with their frequencies.
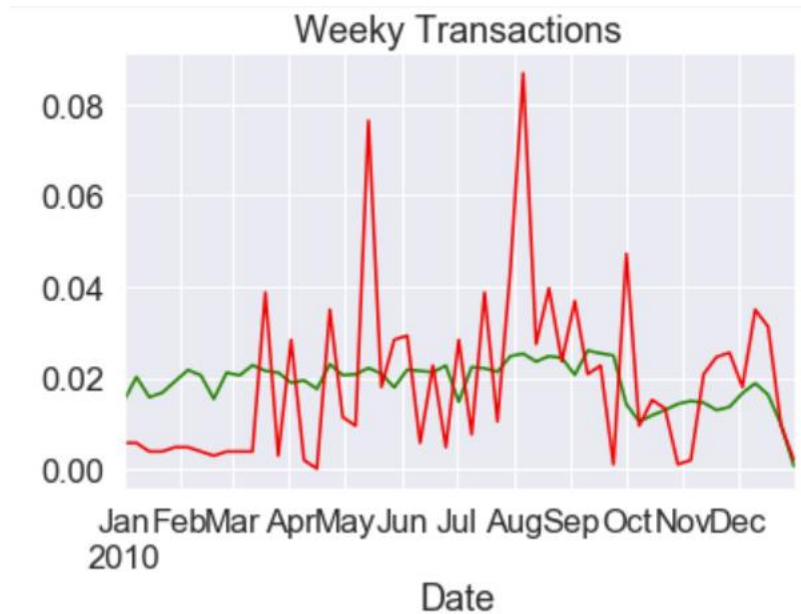
0	95694
1	1059

Below is a line plot of daily transactions with proportions of fraud transaction (red) and non-fraud transaction (green).

Below is a line plot of weekly transactions with proportions of fraud transaction (red) and non-fraud transaction (green).

Weeky Transactions

Below is a line plot of approximately monthly transactions with proportions of fraud transaction (red) and non-fraud transaction (green).

Approximately Monthly Transactions