

# DATA QUALITY REPORT

<b>DQR Prepared By</b>	Chinmayi Bengaluru Prakash
<b>Initials</b>	CBP
<b>Date Created</b>	March 30, 2022
<b>Version No.</b>	1.0

## 1. File Description

This section provides a high-level description of the dataset file and the dataset.

The table below provides general information on the dataset file:

<b>Data File Name</b>	“card transactions.csv”
<b>Data File Description</b>	<p>The dataset represents credit card transactions.</p> <p>The dataset contains record number, card number, date of transaction, merchant number, merchant description, merchant’s state, merchant’s zipcode, transaction type, transaction amount and fraud label.</p> <p>Fraud Label categorizes a transaction as good (not fraud: 0) and bad (fraud: 1).</p>
<b>Granularity</b>	A record for each card transaction.

The table below provides general information on the dataset:

<b>Time Period of Data</b>	2006-01-01 to 2006-12-31
<b>Number of Fields (Columns)</b>	10
<b>Number of Records (Rows)</b>	96753
<b>Number of Numeric Fields</b>	2
<b>Number of Categorical Fields</b>	8

## 2. Fields Summary Tables

This section provides information on two field summary tables listing all fields with summary information.

### 2.1 Numeric Fields Table

The following table lists all the numeric fields along with their summary statistics:

Field Name	% Populated	Min	Max	Mean	Standard Deviation	% Zero
Date	100.00%	2006-01-01	2006-12-31	NA	NA	0.00%
Amount	100.00%	0.01	3,102,045.53	428.20	10,006.14	0.00%

Therefore, this dataset has two numeric fields.

### 2.2 Categorical Fields Table

The following table lists all the categorical fields along with their summary statistics:

Field Name	% Populated	# Unique Values	Most Common Value
Recnum	100.00%	96,753	-
Cardnum	100.00%	1,645	5142148452
Merchnum	96.51%	13,091	930090121224
Merch description	100.00%	13,126	GSA-FSS-ADV
Merch state	98.76%	227	TN
Merch zip	95.19%	4,567	38118.0
Transtype	100.00%	4	P
Fraud	100.00%	2	0

Therefore, this dataset has eight categorical fields.

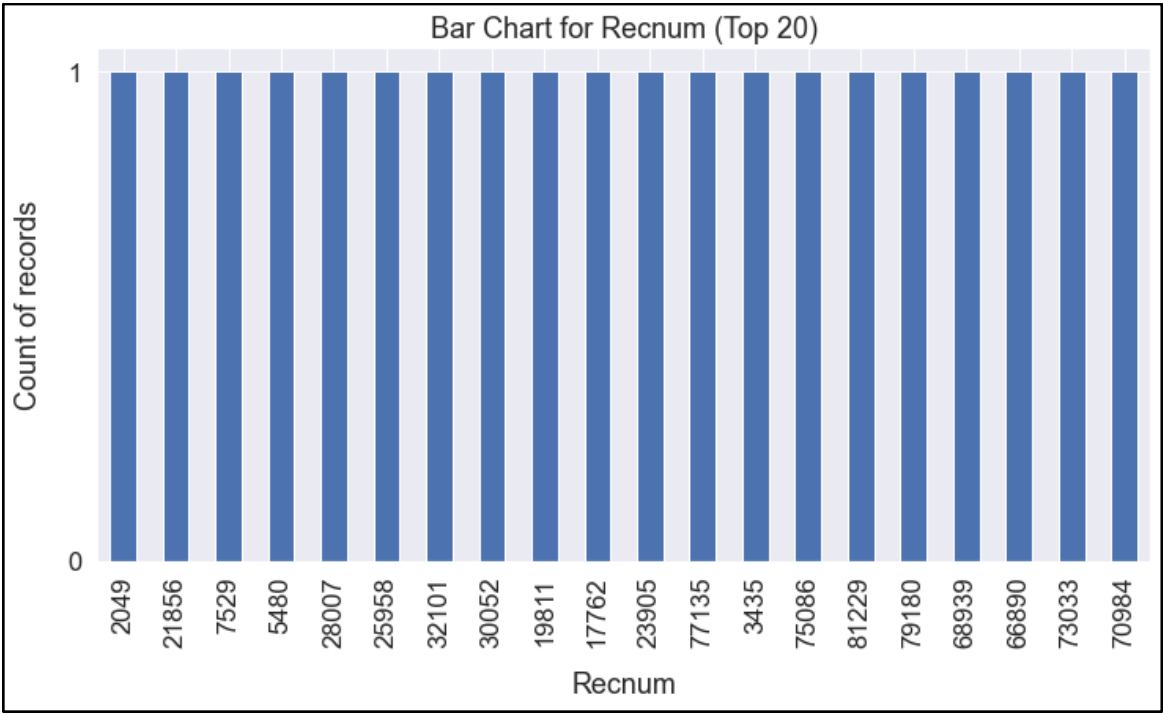
3. Description of the fields

This section provides a brief description for each of the 10 fields:

3.1 Recnum

Field Name	Recnum
Field Description	<b>Categorical field</b> with no missing values.
	This field denotes a unique number for each transaction.
	This field has a unique value for each row.

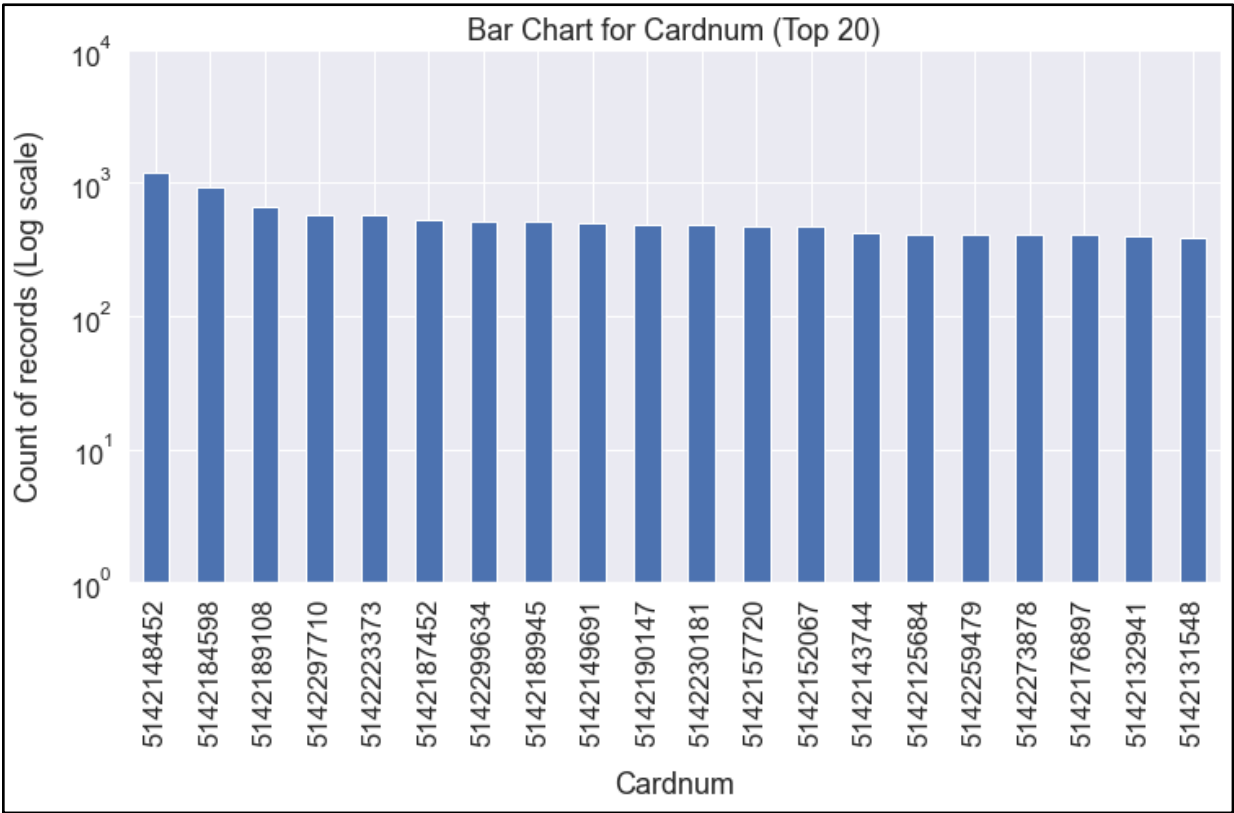
The bar chart below displays top 20 values of “Recnum” field:



3.2 Cardnum

Field Name	Cardnum
Field Description	<b>Categorical field</b> with no missing values.
	This field denotes the credit card number.
	This field has 1,645 unique values.
	The most common value is 5142148452.

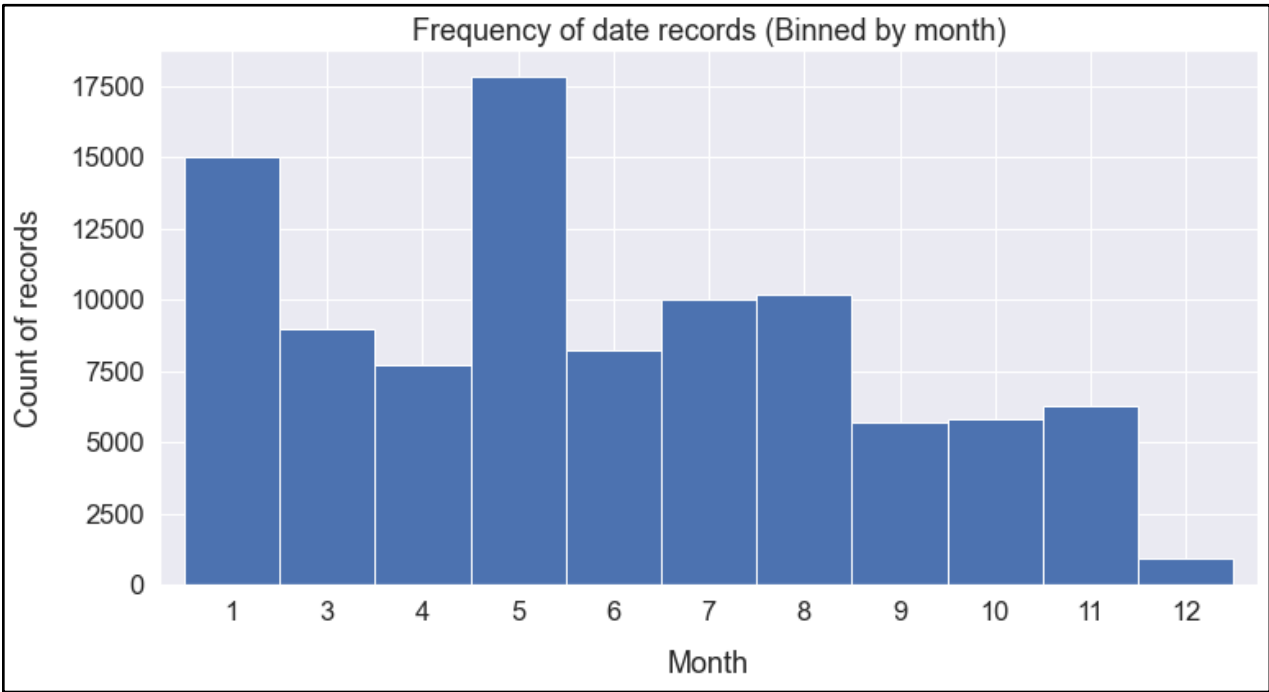
The bar chart below displays top 20 values of “Cardnum” field:



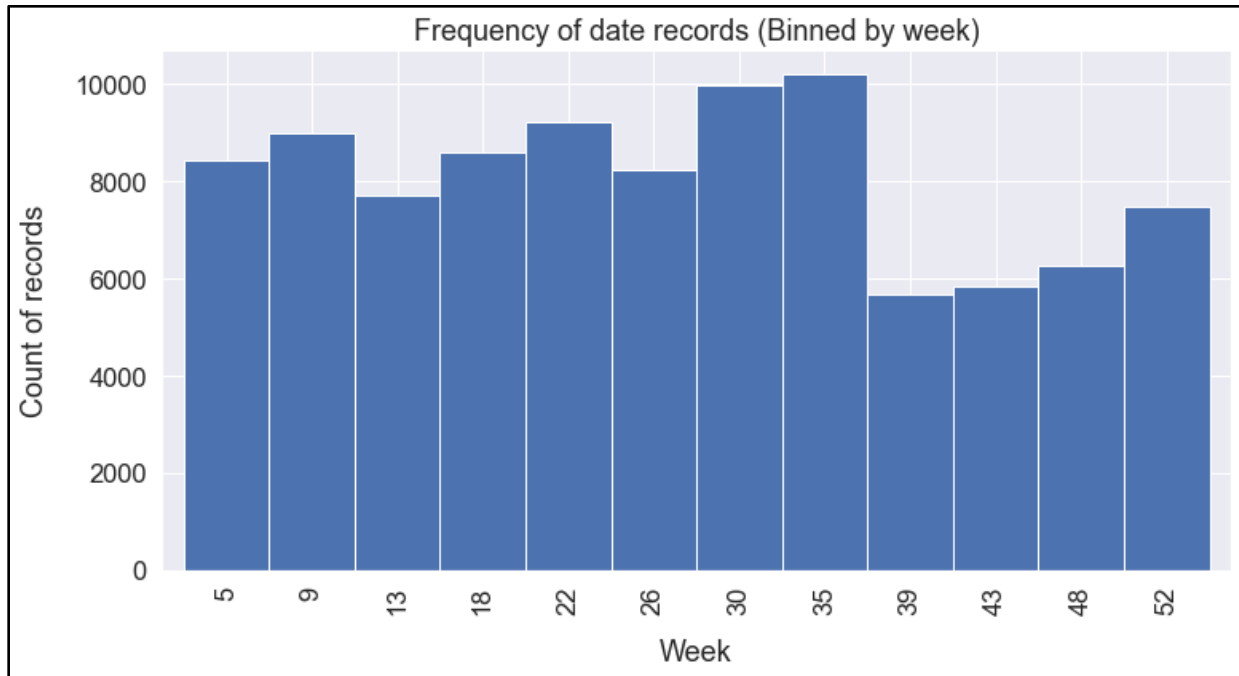
3.3 Date

Field Name	Date
Field Description	<b>Numerical field</b> with no missing values.
	This field denotes the date of credit card transaction.
	The Date field ranges from 2006-01-01 to 2006-12-31

The following is the distribution of values of “Date” field binned by 12 months of 2006:



The following is the distribution of values of “Date” field binned by 52 weeks of 2006:



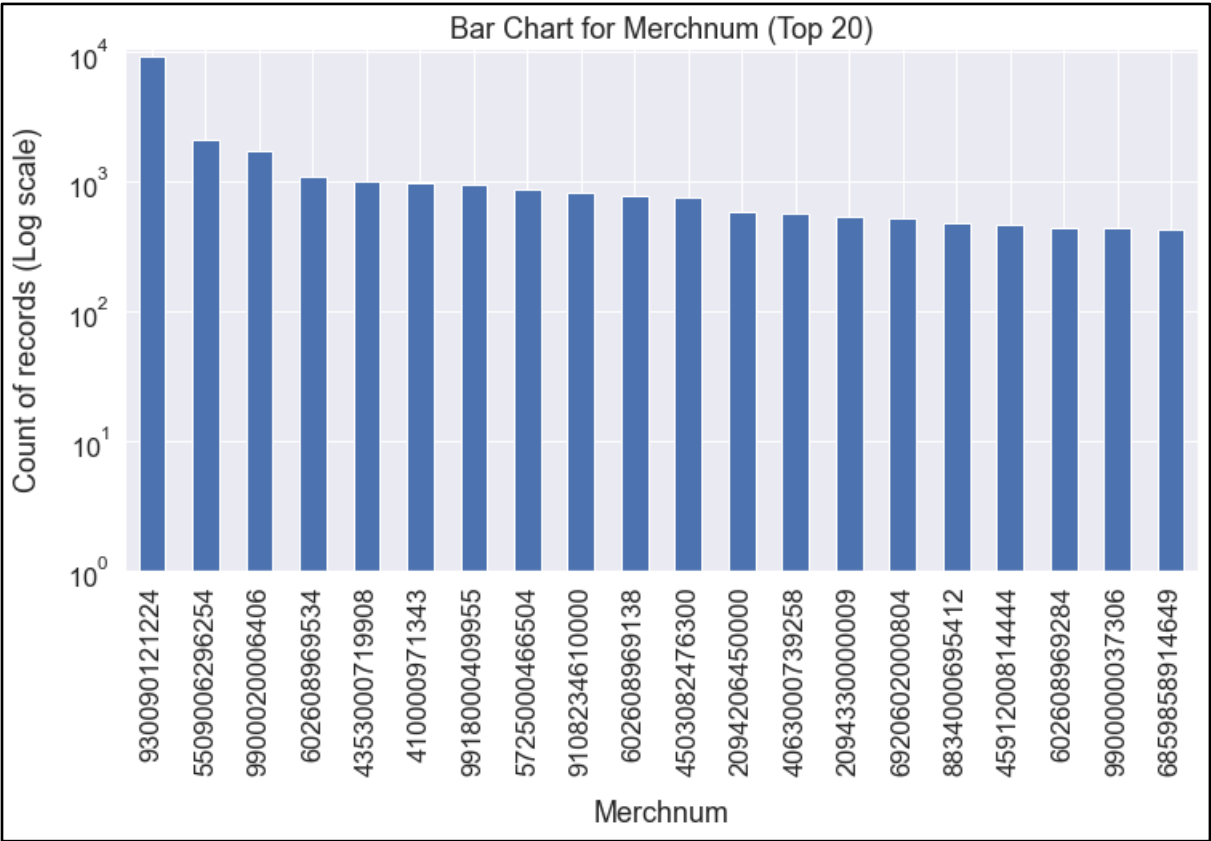
The following table shows the top 15 highest transactions made on respective dates:

Date	# Transactions
2006-02-28	684
2006-08-10	610
2006-03-15	594
2006-09-13	564
2006-08-09	536
2006-09-07	536
2006-09-14	533
2006-09-21	522
2006-08-01	521
2006-08-31	518
2006-09-28	503
2006-08-29	503
2006-02-16	501
2006-09-19	497
2006-02-09	496

3.4 Merchnum

Field Name	Merchnum
Field Description	<b>Categorical field</b> with 3,375 missing values.
	This field denotes the name of the merchant.
	This field has 13,091 unique values.
	The most common value is 930090121224.

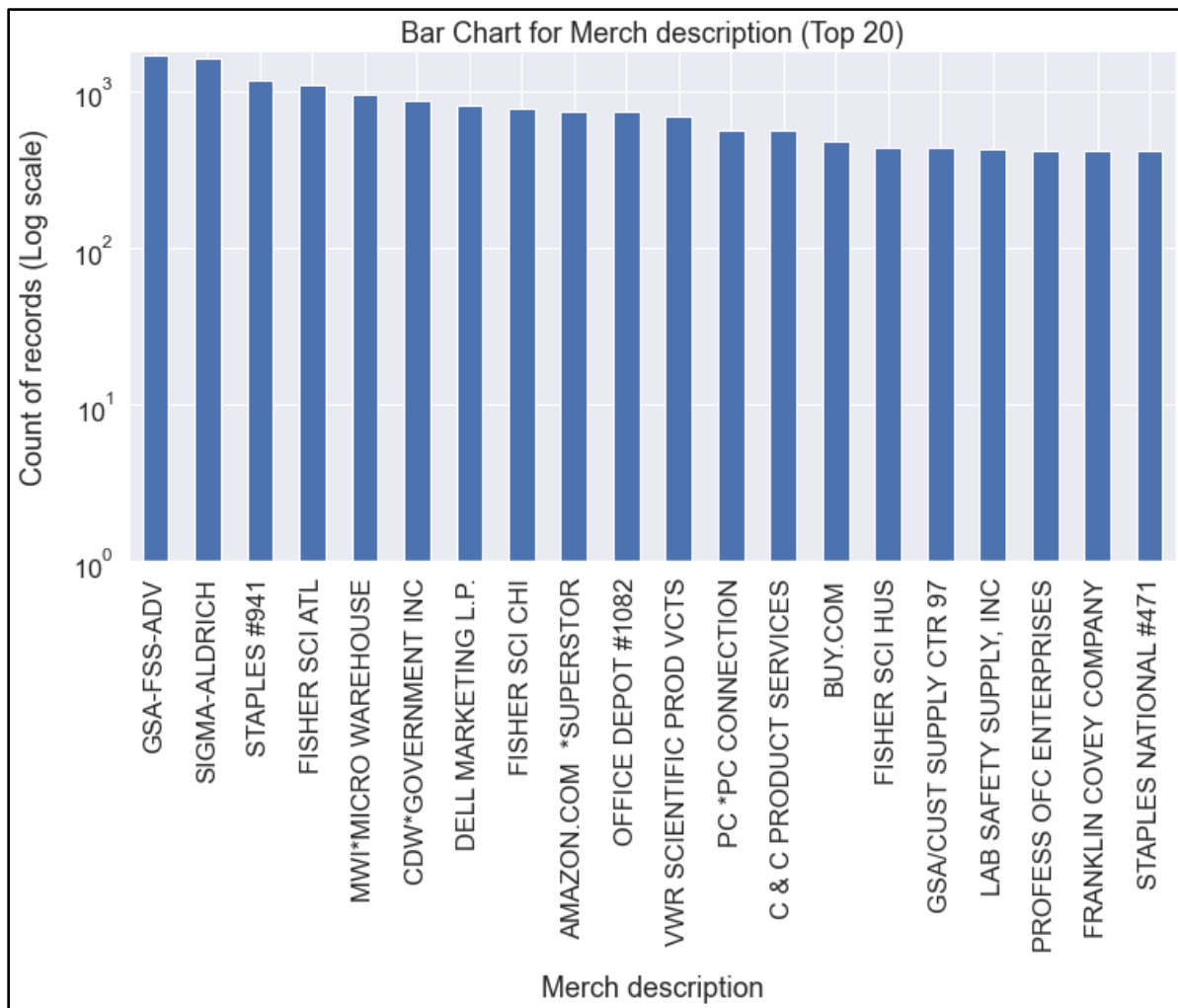
The bar chart below displays the top 20 unique values of “Merchnum” field:



### 3.5 Merch description

Field Name	Merch description
Field Description	<b>Categorical field</b> with no missing values.
	This field denotes the description of Merchant.
	This field has 13,126 unique values.
	The most common value is GSA-FSS-ADV.

The bar chart below displays the top 20 unique values of “Merch description” field:

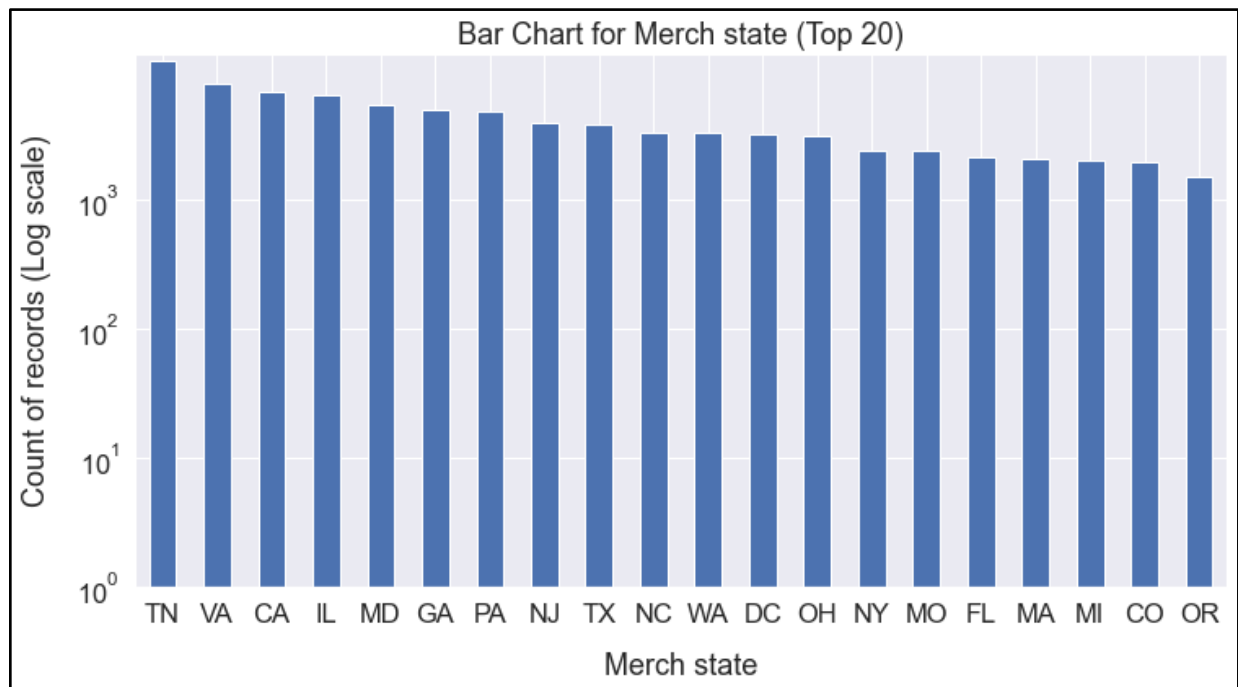




### 3.6 Merch state

<b>Field Name</b>	Merch state
<b>Field Description</b>	<b>Categorical field</b> with 1,195 missing values.
	This field denotes the state where the Merchant is registered to be located.
	This field has 227 unique values.
	The most common value is TN.

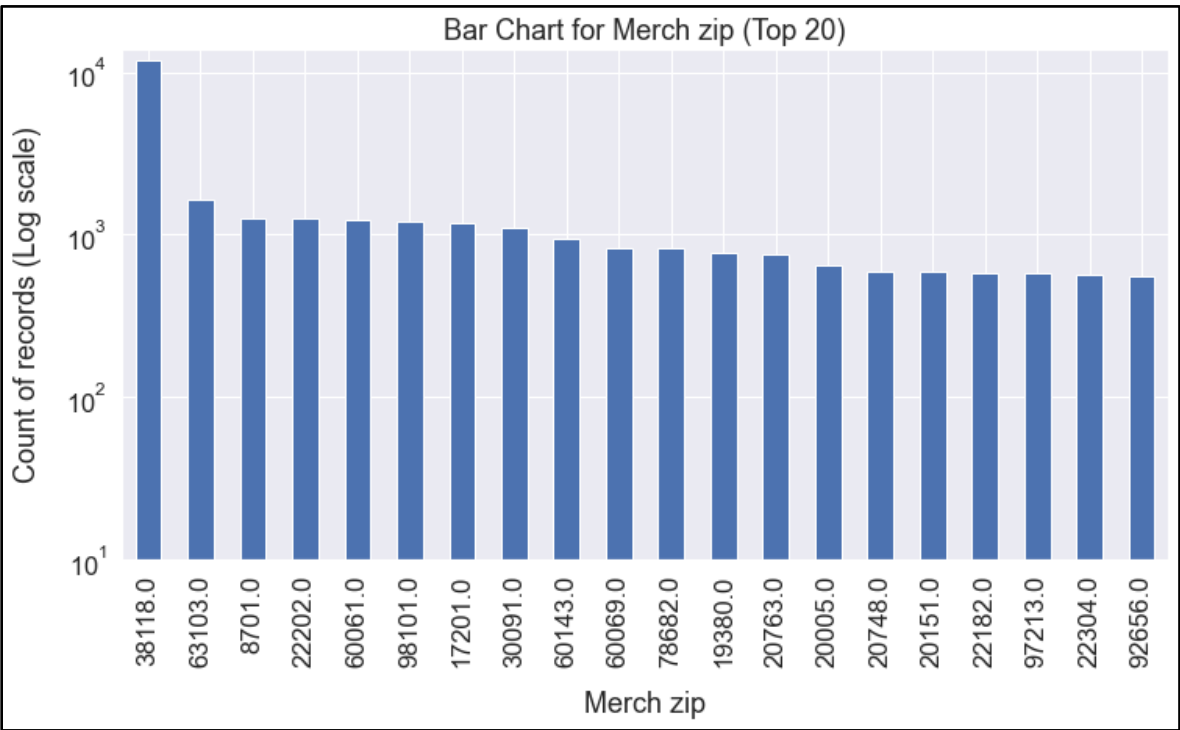
The bar chart below displays the top 20 unique values of “Merch state” field:



3.7 Merch zip

Field Name	Merch zip
Field Description	<b>Categorical field</b> with 4,656 missing values.
	This field denotes the 4-digit or 5-digit zip code where the Merchant is registered to be located.
	This field has 4,567 unique values.
	The most common value is 38118.0

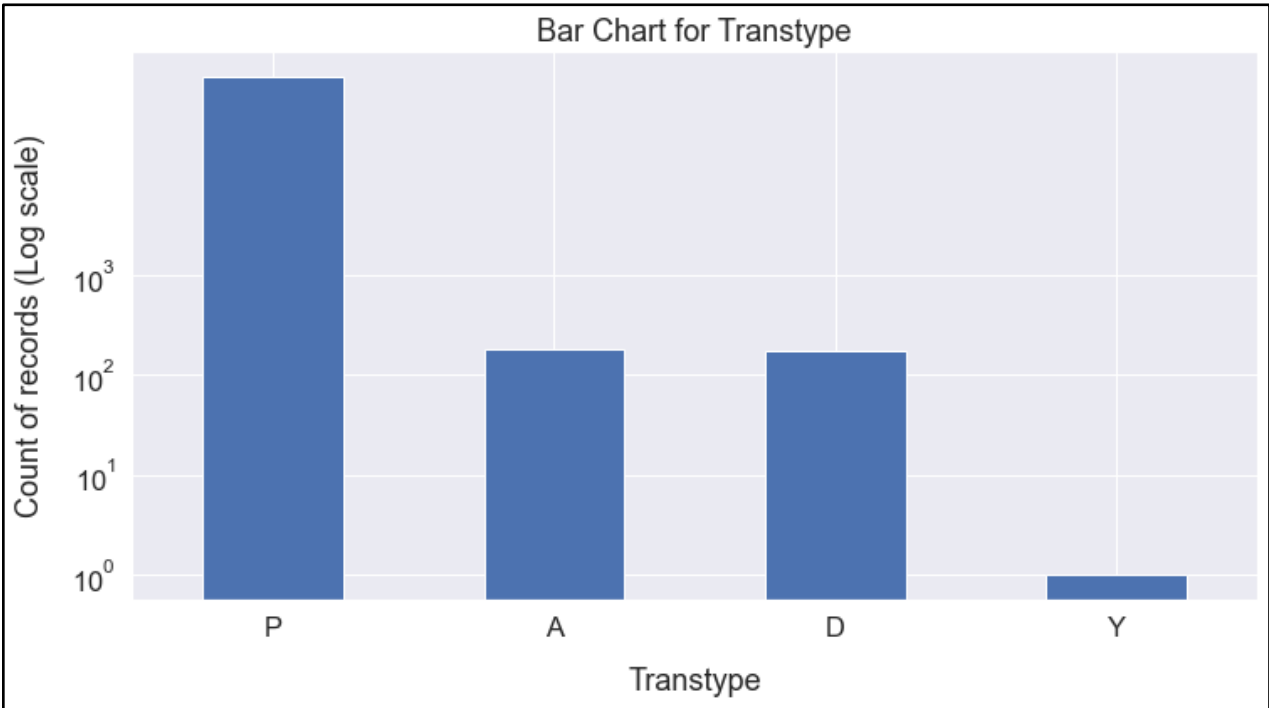
The bar chart below displays the unique values of “Merch zip” field:



3.8 Transtype

Field Name	Transtype
Field Description	<b>Categorical field</b> with no missing values.
	This field denotes the type of transaction made.
	This field has 4 unique values. P – Purchase A – No info provided D – No info provided Y – No info provided
	The most common value is P.

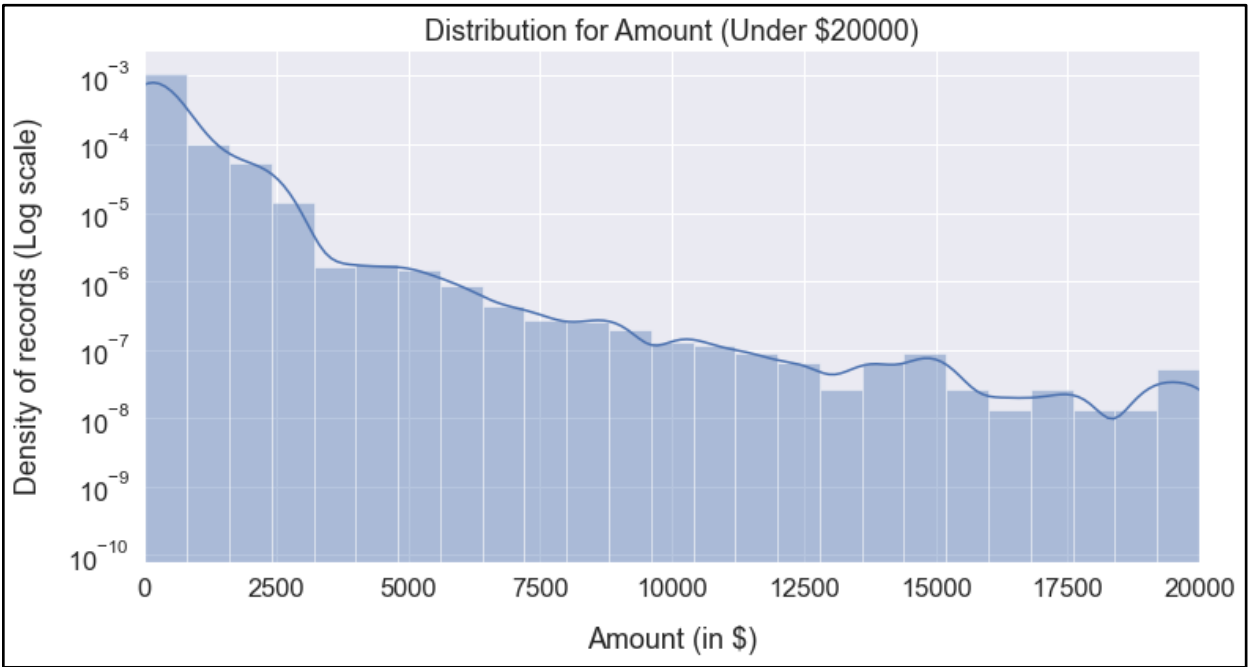
The bar chart below displays the 4 unique values of “Transtype” field:



3.9 Amount

Field Name	Amount
Field Description	<b>Numerical field</b> with no missing values.
	This field denotes the amount that is transacted in each transaction.
	The Amount field ranges from: <b>\$ 0.01 to \$ 3102045.53</b>

The following is the distribution of values of “Amount” field within \$20000 which accounts for 99.97% of the total data (all “Amount” fields):



3.10 Fraud

Field Name	Fraud
Field Description	<b>Categorical field</b> with no missing values.
	This field denotes whether a transaction is fraudulent or not.
	0 - Not fraud
	1 - Fraud
	This field has 2 unique values.
	The most common value is 0.

The bar chart below displays the unique values of “Fraud” field:

