Credit Card Transaction Data DQR

Description:

Dataset Name: Card Transactions

Dataset Purpose: Record of credit card transactions of a U.S. government organization over a 12-month period.

Data Source: U.S. Government

Time Period: January 1, 2010 – December 31, 2010

Number of Fields: 10

Number of Records: 96,753

Summary Table:

Numeric Fields:

Column Name	# of Records	% Populated	# Unique Values	# of Zeros	Mean	Std. Dev.	Min.	Max.
Amount	96753	100.00	34909	0	427.89	10006.14	0.01	3102045.53

Categorical Fields:

Column Name	# of Records	% Populated	# Unique Values	Most Common Field
Recnum	96753	100.00	96753	N/A
Cardnum	96753	100.00	1645	5142148452
Date	96753	100.00	365	2/28/10
Merchnum	93378	96.51	13091	930090121224
Merch description	96753	100.00	13126	GSA-FSS-ADV
Merch state	95558	98.76	227	TN
Merch zip	92097	95.19	4567	38118.0
Transtype	96753	100.00	4	P
Fraud	96753	100.00	2	0

Data Field Exploration:

Field 1
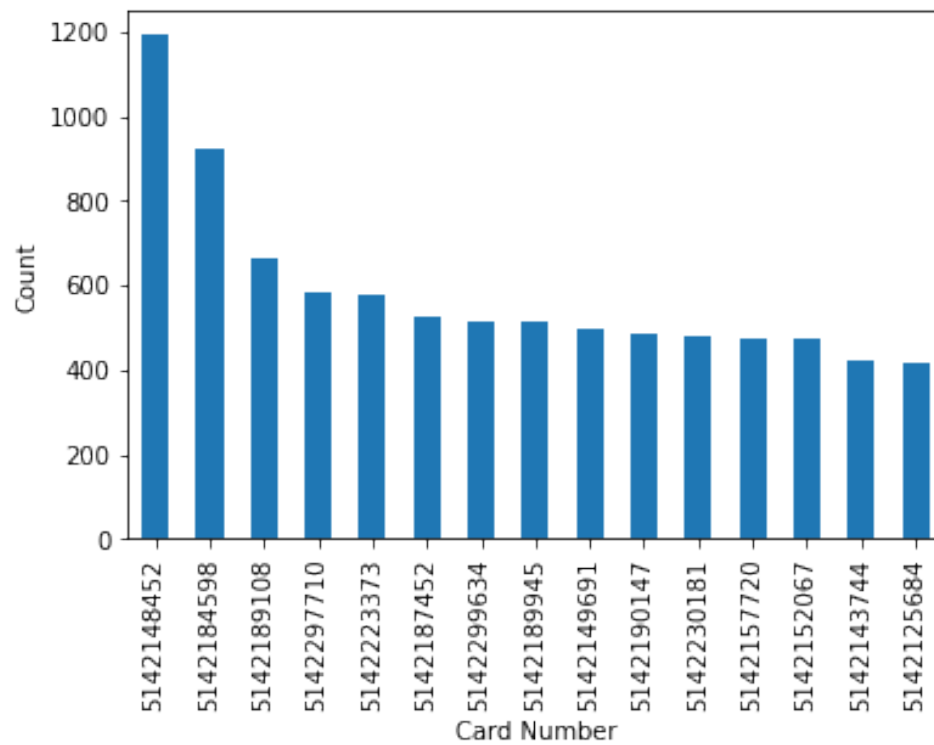
Name: Recnum

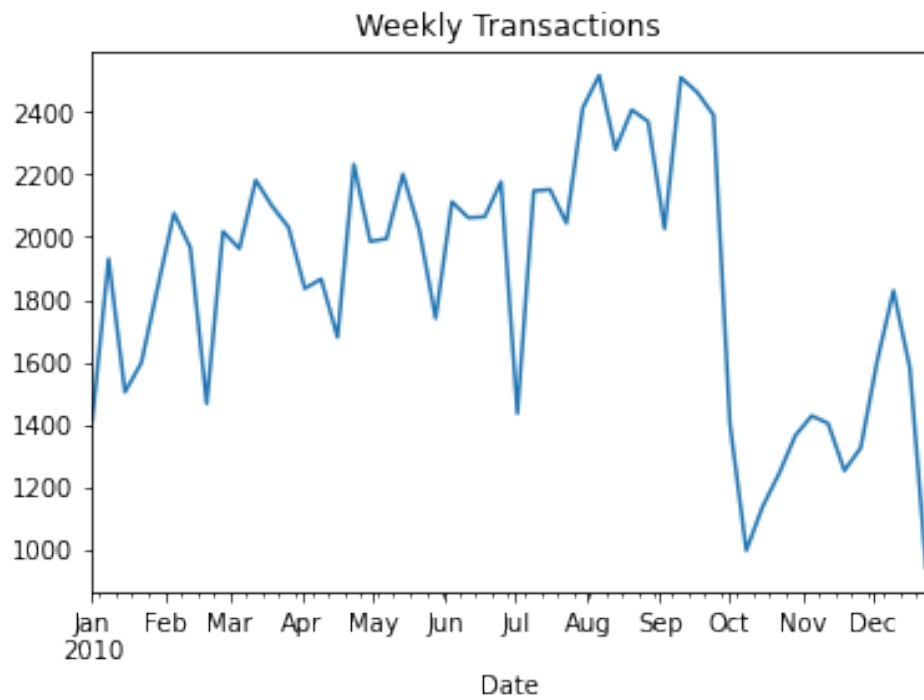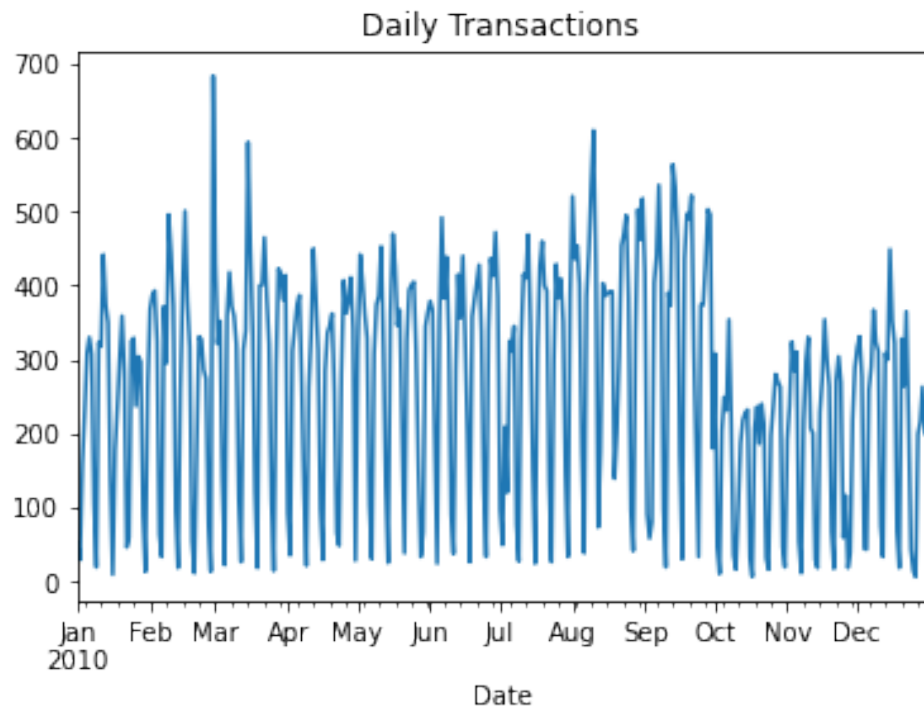Description: Identifier for each data entry.

Field 2

Name: Cardnum

Description: Card number.

<u>Field 3</u>

Name: Date

Description: Date of transaction.

Daily Transactions
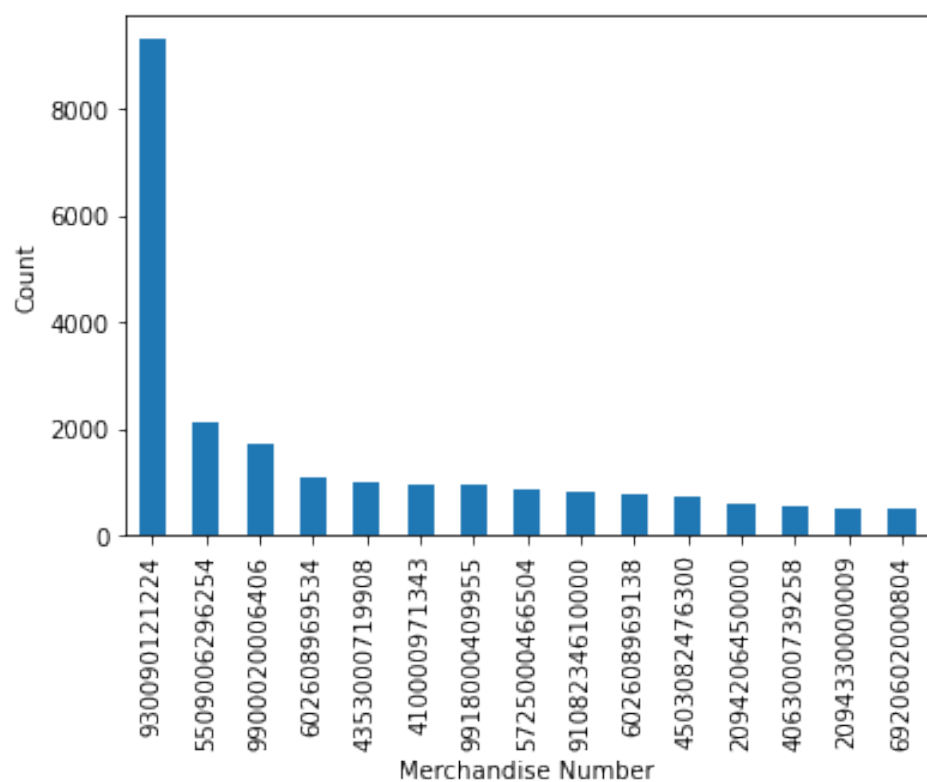

Weekly Transactions

Approximate Monthly Transactions
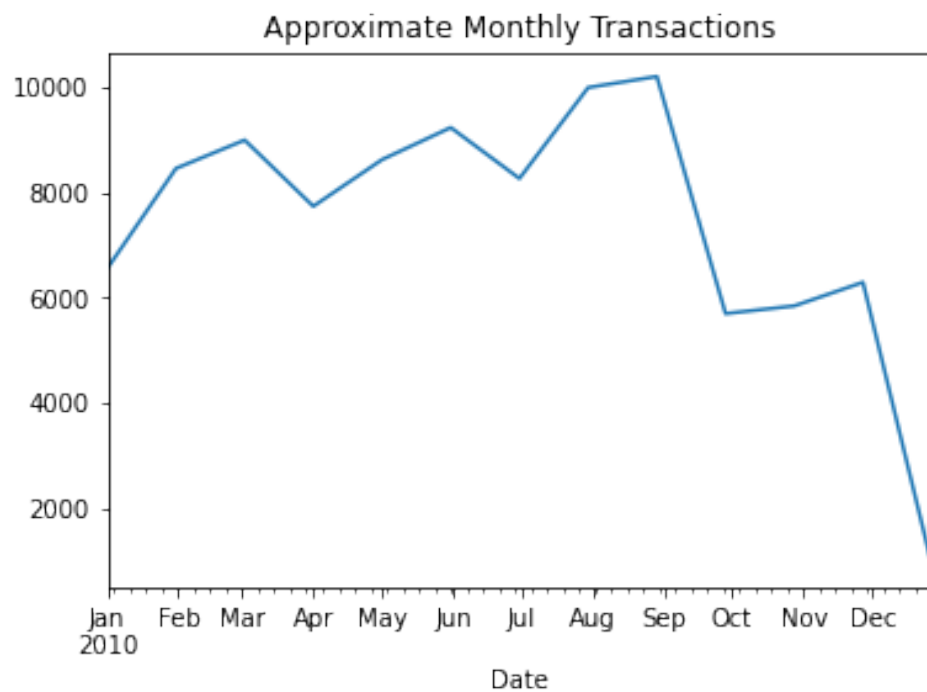
Field 4

Name: Merchnum
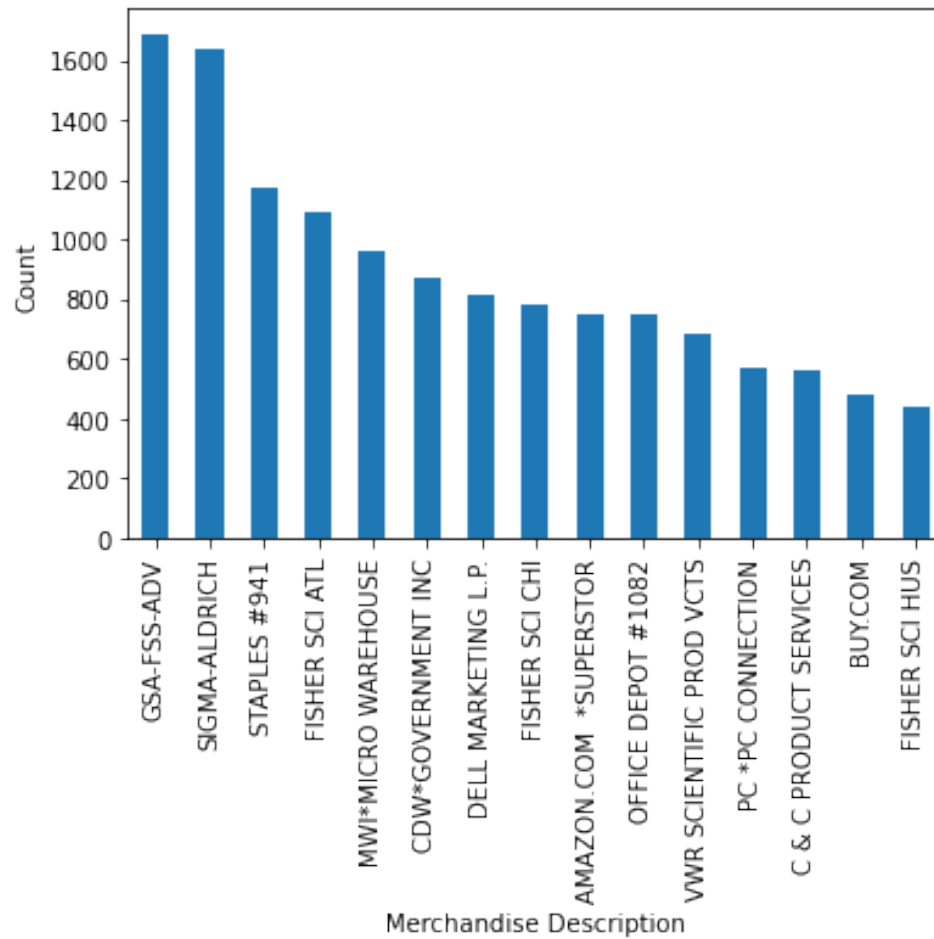
Description: Identifier for merchandise item.

Field 5

Name: Merch description.

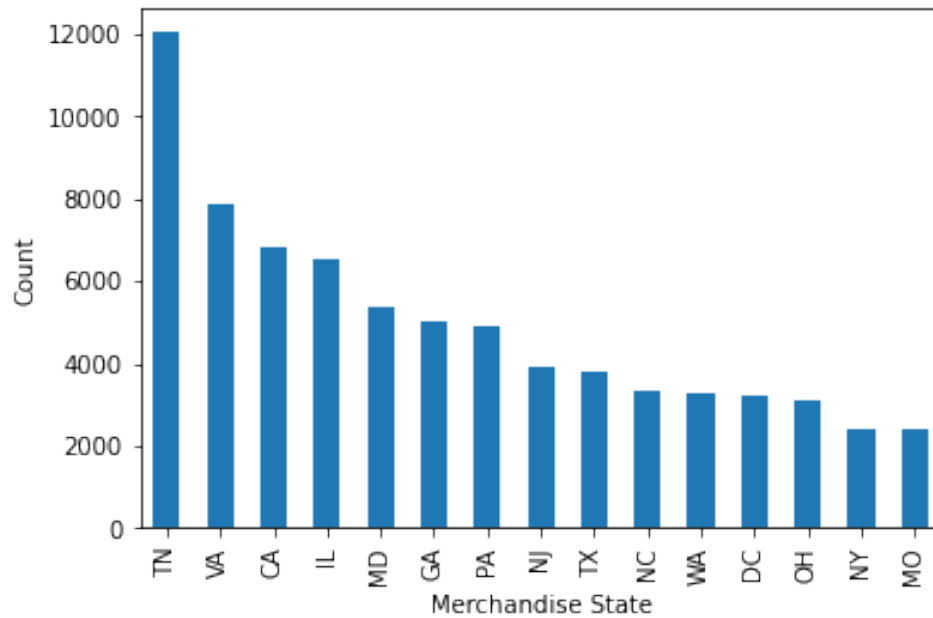Description: Name and location identifier of the business of purchase.
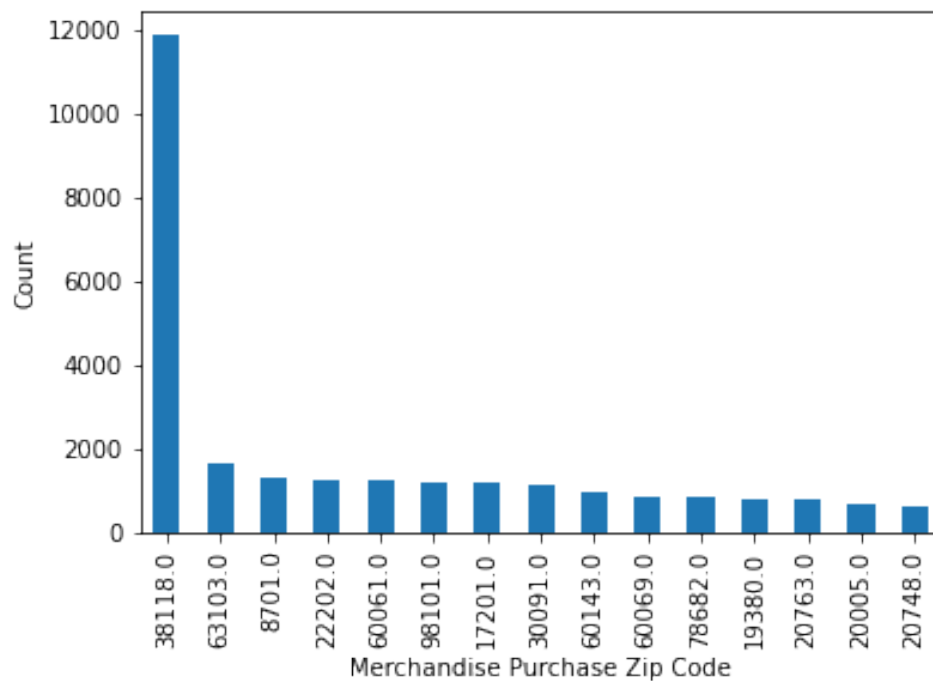
Field 6

Name: Merch state

Description: State or country of transaction.

Field 7

Name: Merch zip
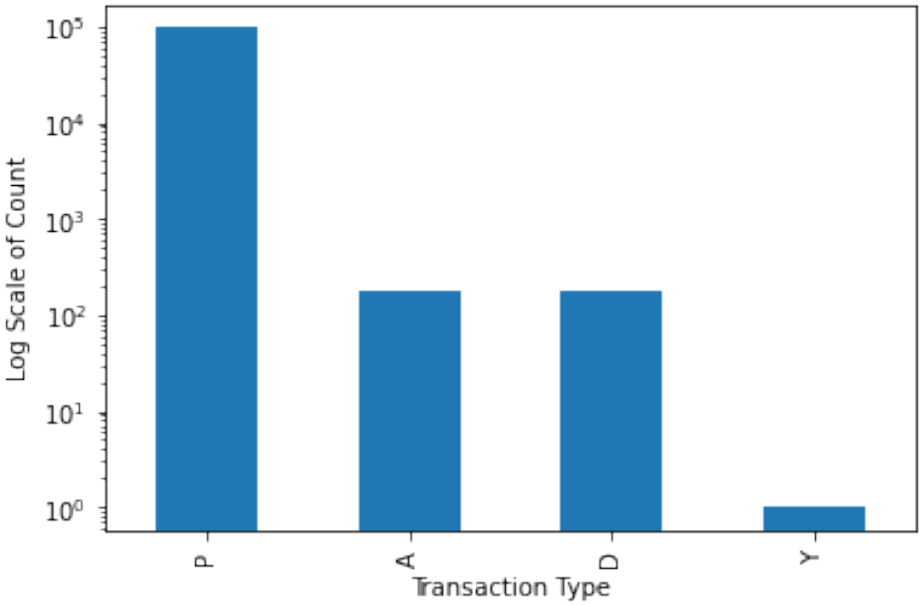
Description: Zip code in which transaction was made.

Field 8

Name: Transtype

Description: Classification of transaction type.

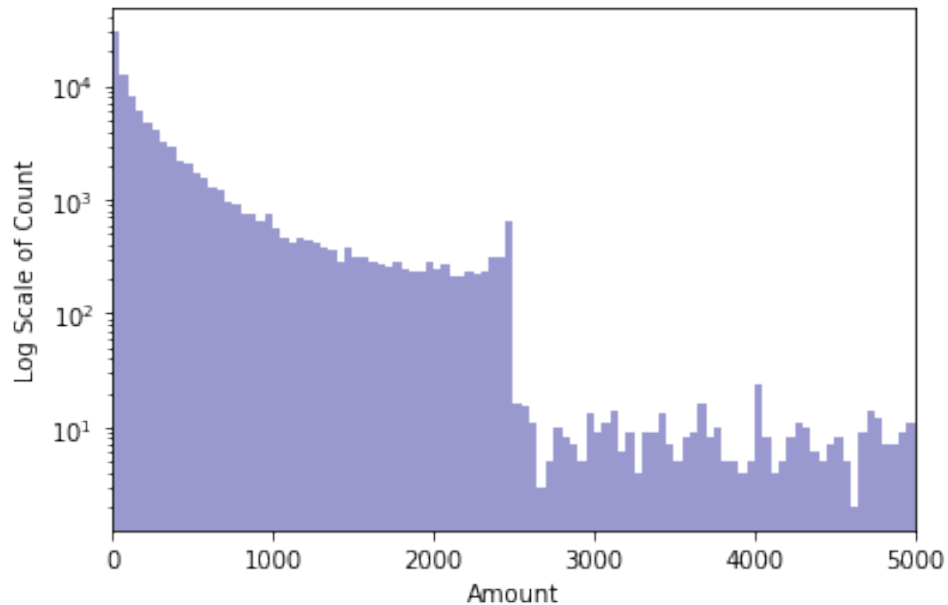| 'P'	'D'	'A'	'Y'
Purchase	Delayed Capture	Unknown	Unknown

<u>Field 9</u>

Name: Amount

Description: Cost of transaction.
Exclude outliers > 5000
Data in histogram is 99.68% populated.

<u>Field 10</u>

Name: Fraud

Description: Binary classification of whether a record is fraudulent: 1 = fraud, 0 = no fraud.