Data Quality Report

Description

Dataset Name: Card Transactions

Dataset Purpose: Credit card purchases from a US government organization.

Data Source: Unnamed US government organization.

Time Period: 1/1/2010 – 12/31/2010

Number of Fields: 10

Number of Records: 96753

Summary Tables

Numeric Fields

Field Name	# Of Records	Percent Populated	Unique Values	Mean	Standard Deviation	Minimum Value	Maximum Value	# Of Zeros
Amount	96753	100%	34909	427.89	10006.14	0.01	3102045.53	0

Categorical Fields

Field Name	# Of Records	% Populated	Unique Values	Most Common Field Value	Datetime Field
Recnum	96753	100%	96753	N/A	No
Cardnum	96753	100%	1645	5142148452	No
Date	96753	100%	365	2/28/2010	Yes
Merchnum	93378	96.51%	6627	930090000000	No
Merch description	96753	100%	13126	GSA-FSS-ADV	No
Merch state	95558	98.76%	228	TN	No
Merch zip	92097	95.19%	4568	38118.0	No
Transtype	96753	100%	4	P	No
Fraud	96753	100%	2	0	No

Data Field Exploration
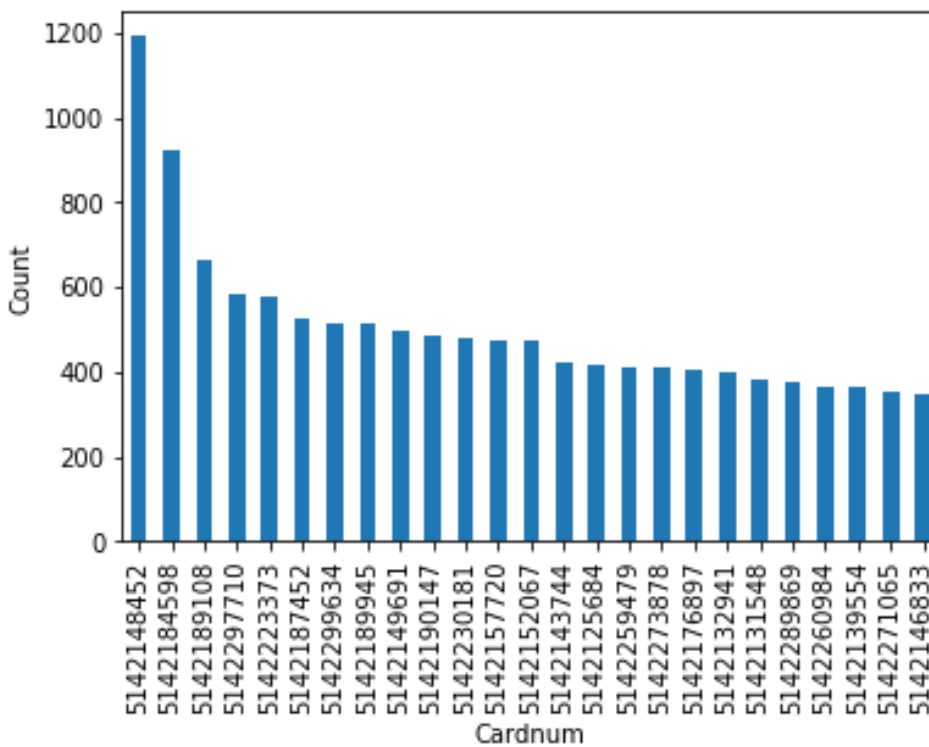
Field 1:

Name: Recnum

Description: Record number of each entry in the dataset.

Field 2:

Name: Cardnum

Description: Card Number associated with the transaction.
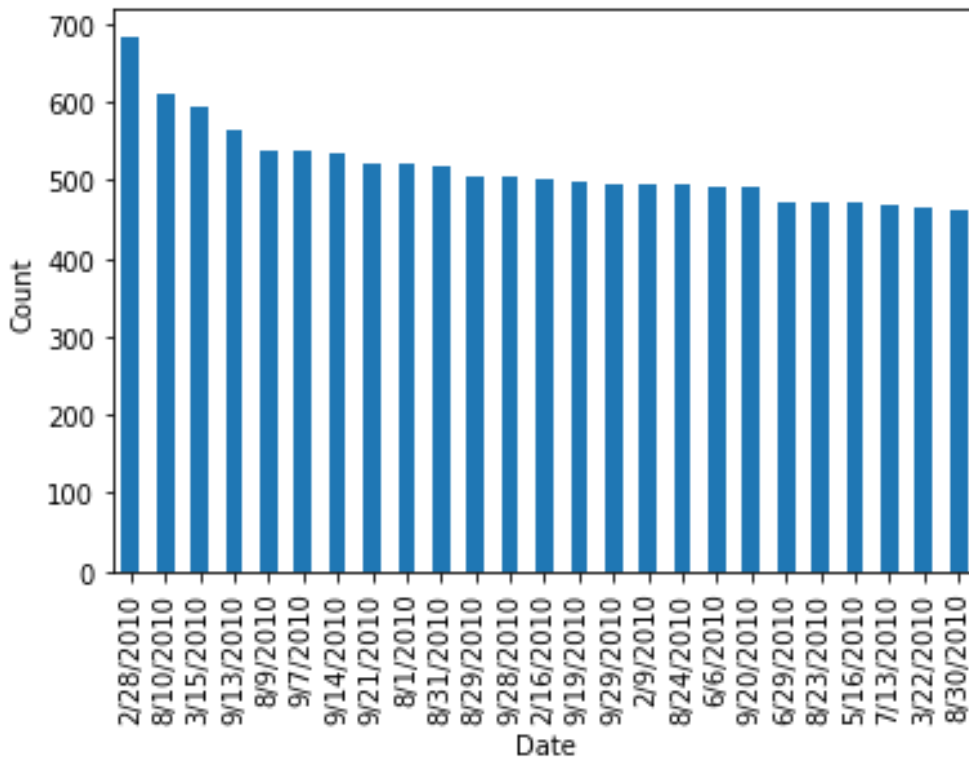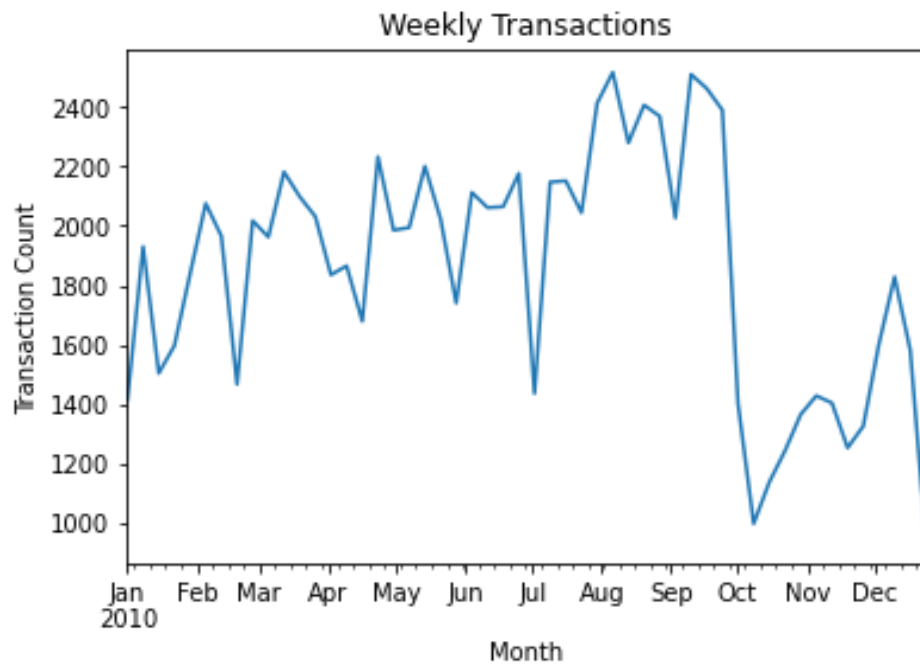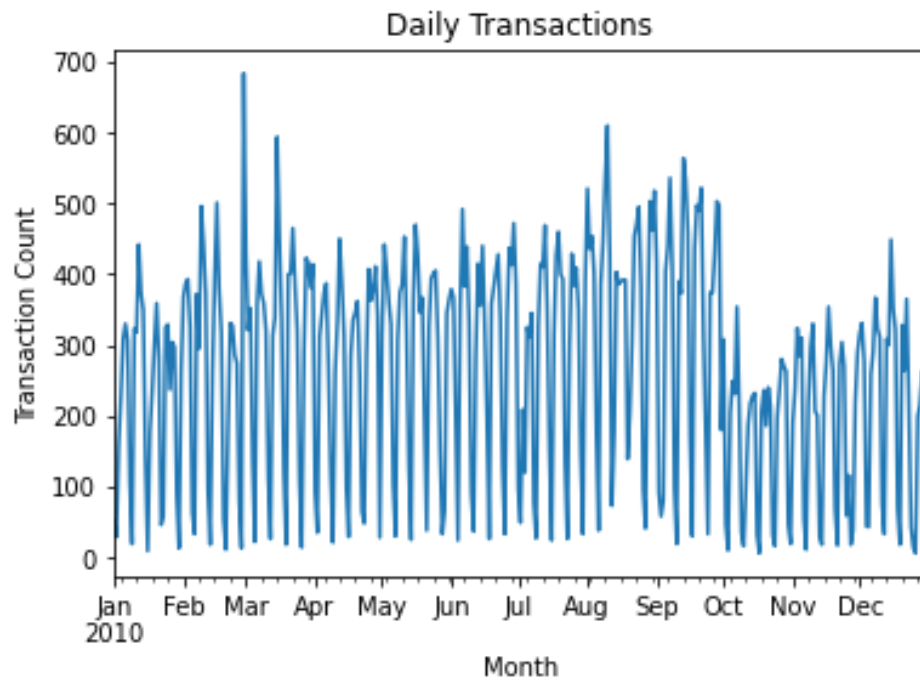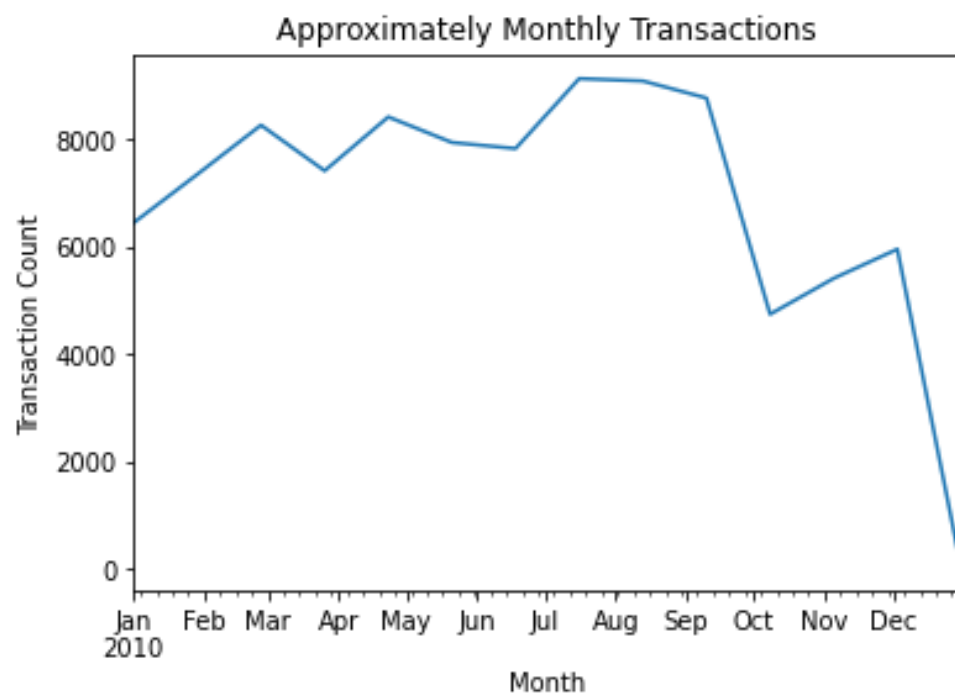
Top 25 records shown.

Field 3:

Name: Date

Description: Transaction date.

Showing top 25 records for the following plot.

Daily Transactions

Weekly Transactions

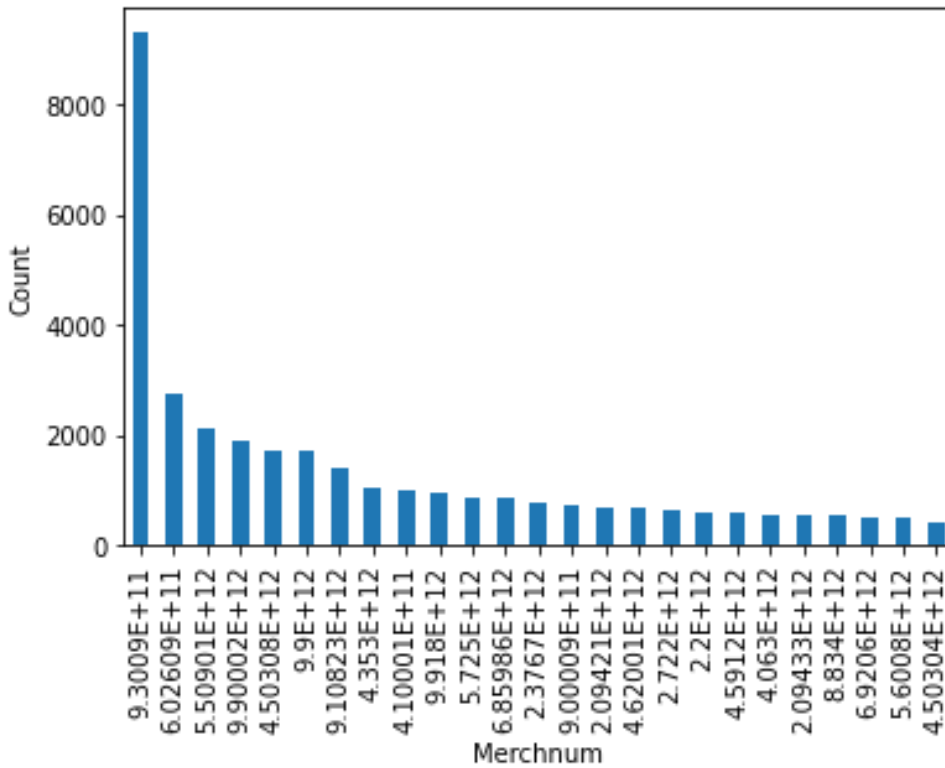Approximately Monthly Transactions

Field 4:

Name: Merchnum

Description: Merchant number.
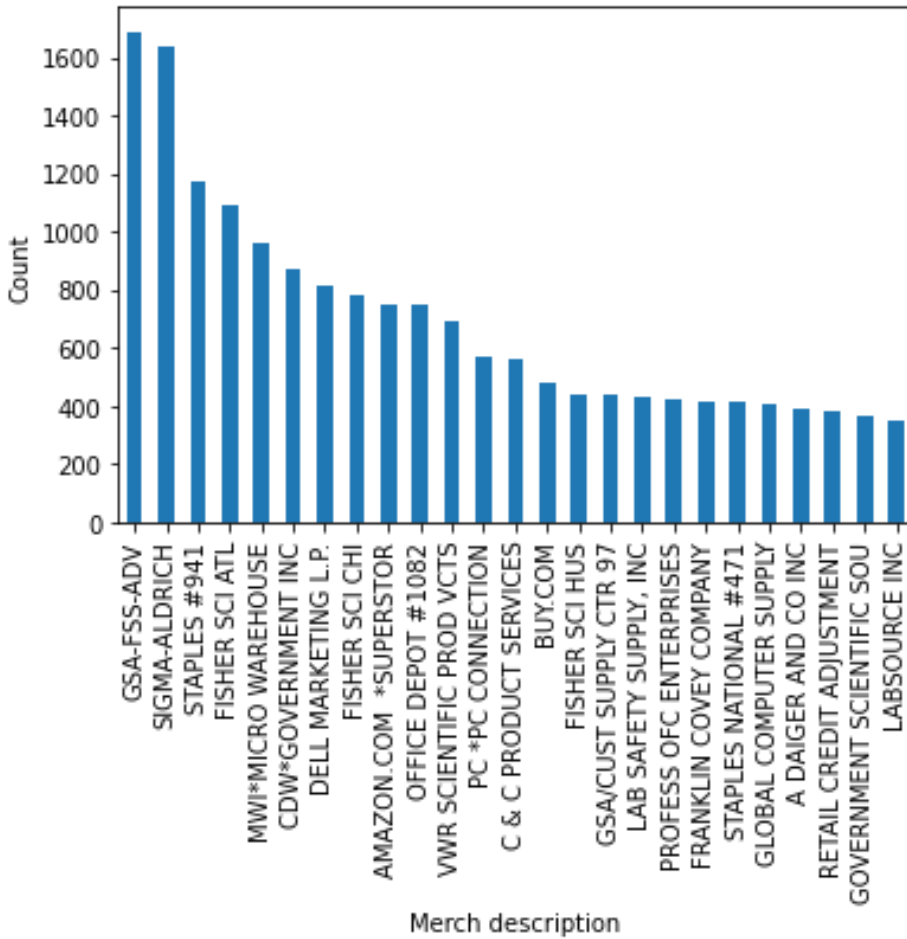
Showing top 25 records.

Field 5:

Name: Merch description.

Description: Merchant description.
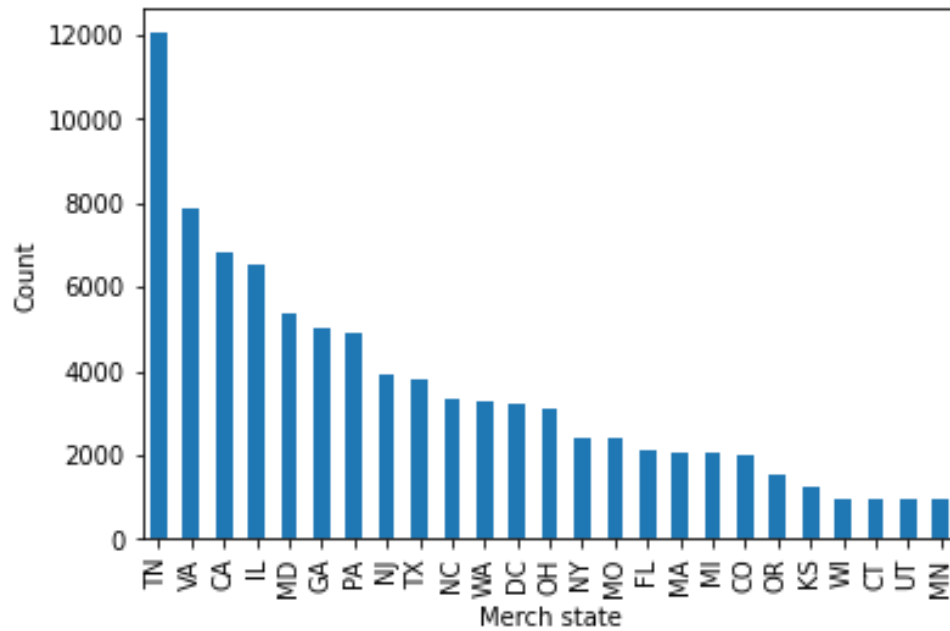
Showing top 25 records

Field 6:

Name: Merch state

Description: US state where the merchant is based from (can also contain Mexican states and Canadian provinces).
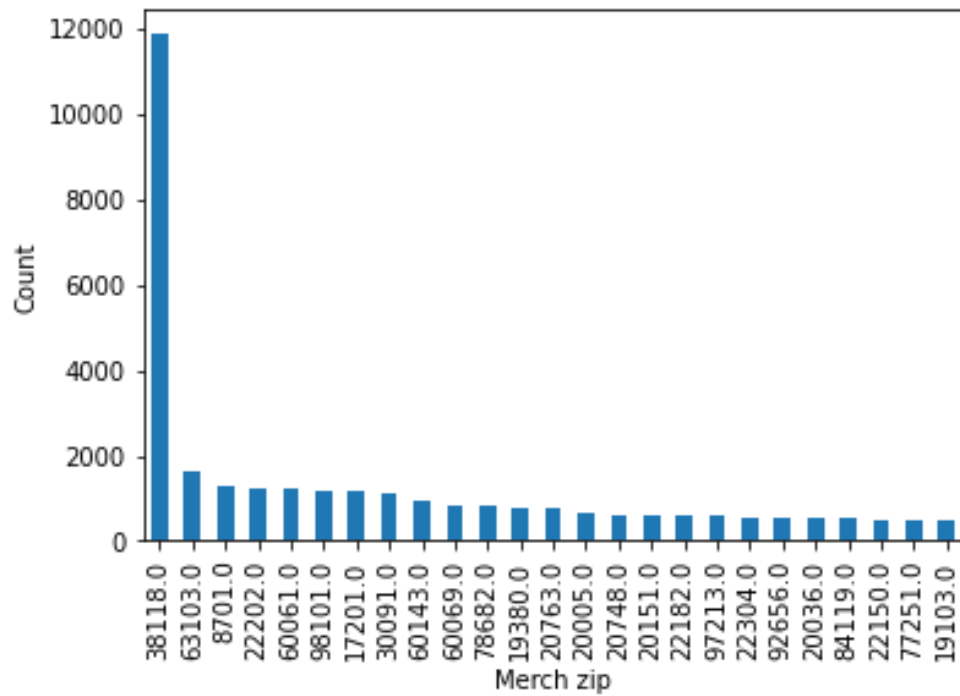
Showing top 25 records

Field 7:

Name: Merch zip

Description: Zip Code corresponding to the area where the merchant is based from.
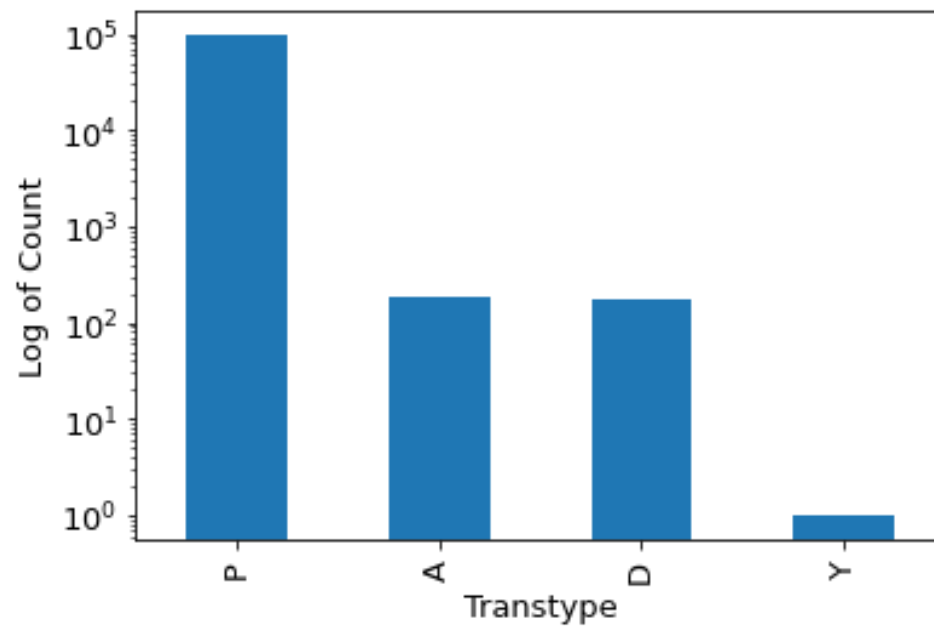
Showing top 25 records

Field 8:

Name: Transtype

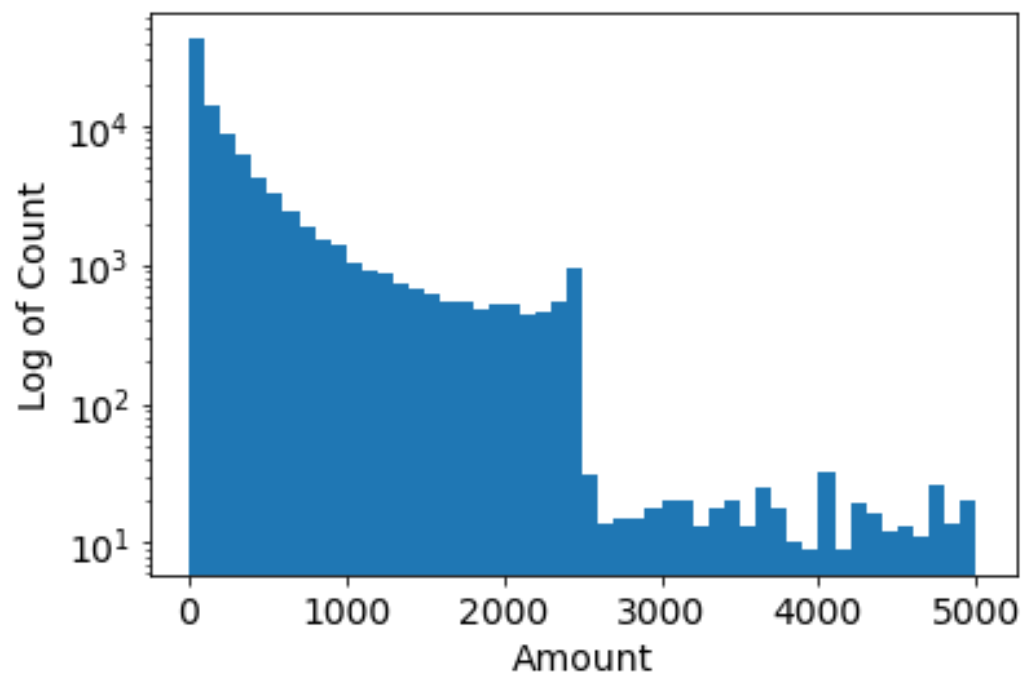Description: Transaction type, four possible values (P, A, D, Y).

Field 9:

Name: Amount

Description: Transaction amount.

Data was filtered to show only transactions with an amount less than or equal to 5,000.

99.67% of the data is represented in this plot.

Field 10:

Name: Fraud

Description: Dummy variable signaling if the record is flagged as fraud or not.

1 = Fraud

0 = Not Fraud