**DATA QUALITY REPORT**

**PROJECT 3:**

**CARD TRANSACTION FRAUD MODEL**

**Prepared for:**

Professor Coggeshall

USC Marshall DSO 562 Fraud Analytics

Homework 7

**Prepared by:**

Nanchun Shi

March 18, 2020

**TABLE OF CONTENTS**

[**PART I - INTRODUCTION**](#_jlgkpo7oov1a) **------------------------------------------------------------------------------ 2**

[**PART II - SUMMARY OF DATASET**](#_f5z4j27jbf5k) **----------------------------------------------------------------- 2**

[**PART III - FIELD DESCRIPTION**](#_8lvwnqf88br7) **--------------------------------------------------------------------- 3**

# 

# 

# PART I - INTRODUCTION

The dataset of interest “card transactions” contains information on 96,753 credit card transaction records received from January 1st, 2010 to December 31th, 2010. The dataset contains 10 fields including record number, card number, transaction date, merchant information, transaction type and fraud label, which indicates if a transaction is considered to be a card transaction fraud.

## **PART II - SUMMARY OF DATASET**

**Below are the summary tables for the 10 fields in the dataset:**

**Categorical fields:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field Name** | **# of Records** | **% Populated** | **# Unique Values** | **Most Common Value** |
| **Recnum** | 96,753 | 100% | 96,753 | - |
| **Cardnum** | 96,753 | 100% | 1,645 | 5142148452 |
| **Date** | 96,753 | 100% | 365 | 2010-02-28 |
| **Merchnum** | 93,378 | 96.51% | 13,092 | 930090121224 |
| **Merch description** | 96,753 | 100% | 13,126 | GSA-FSS-ADV |
| **Merch state** | 95,558 | 98.76% | 228 | TN |
| **Merch zip** | 92,097 | 95.19% | 4,568 | 38118 |
| **Transtype** | 96,753 | 100% | 4 | P |
| **Fraud** | 96,753 | 100% | 2 | 0 |

**Numerical field:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Field Name** | **# of Records** | **% Populated** | **# Unique Values** | **# Records w/ value 0** | **Mean** | **Standard Deviation** | **Max** | **Min** |
| **Amount** | 96,753 | 100% | 34,909 | 0 | 427.89 | 10,006.14 | 3,102,046 | 0.01 |

## **PART III - FIELD DESCRIPTION**

**1.**

**Field name:** Recnum

**Field type:** int64

**Description:** unique index number for each record, ascending from 1 to 96753.

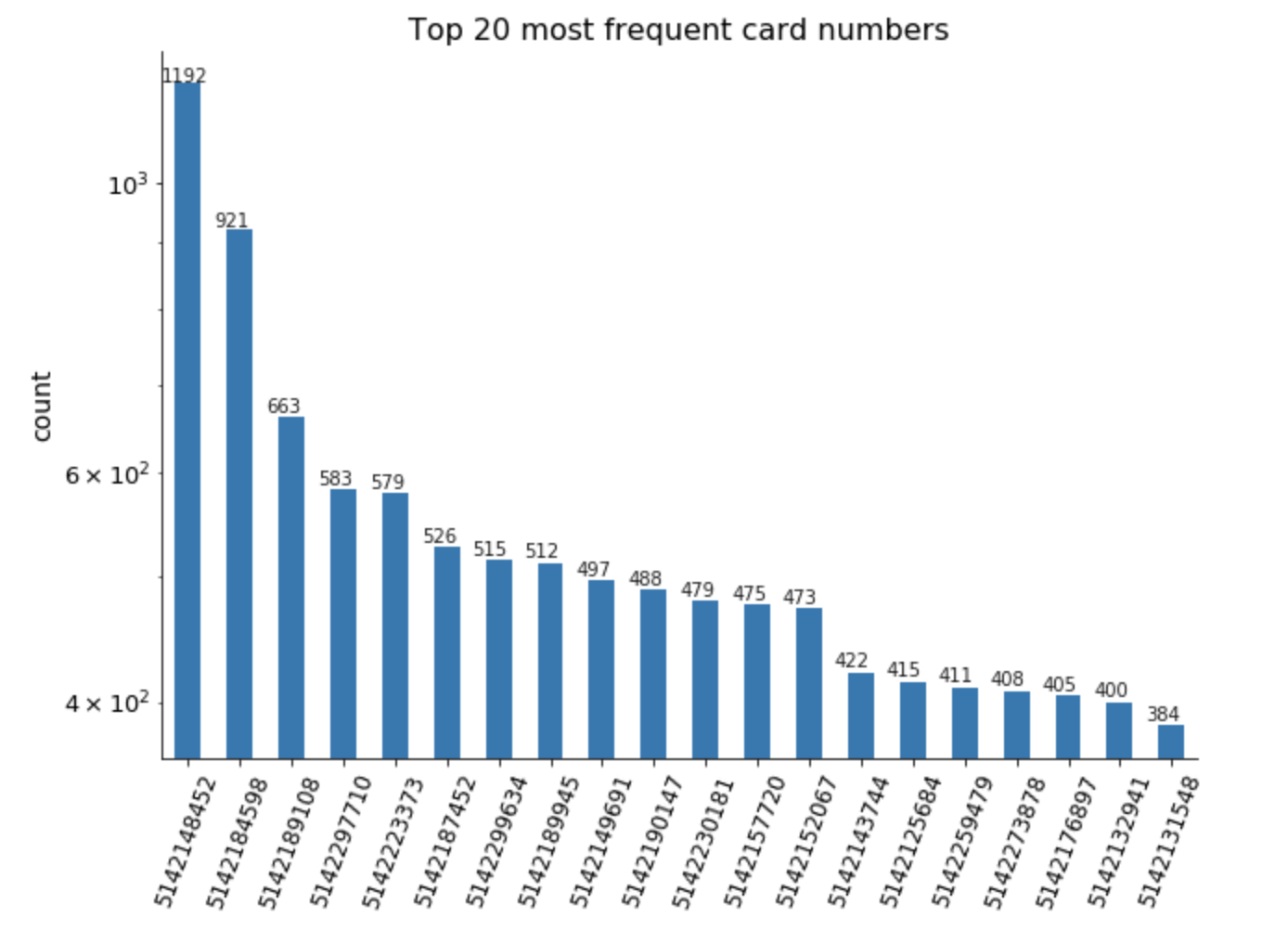
**2.**

**Field name:** Cardnum

**Field type:** int64

**Description:** credit card number used in a transaction.

Top 20 most frequent card numbers:



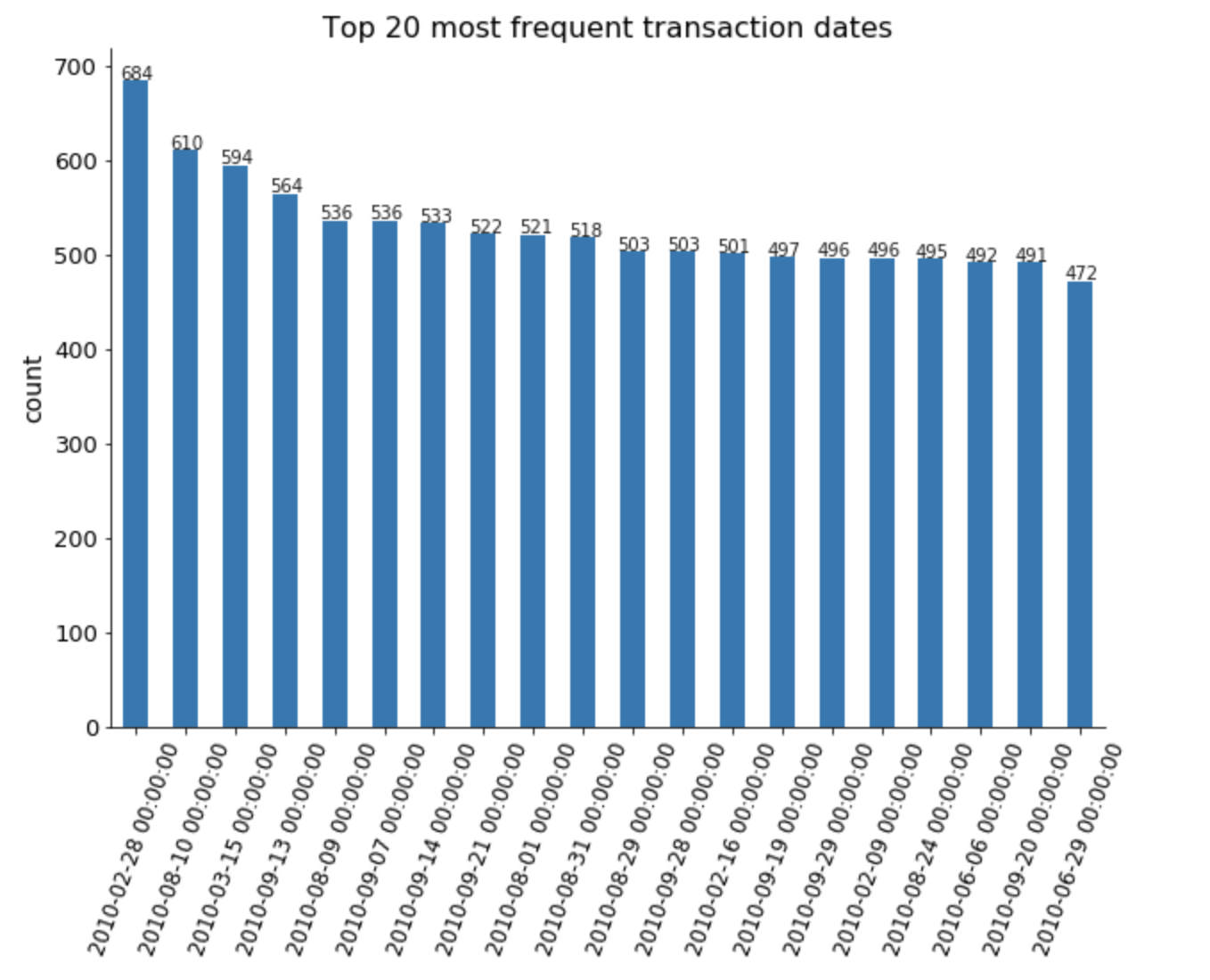
**3.**

**Field name:** Date

**Field type:** datetime64[ns]

**Description:** date of a transaction.

Top 20 most frequent transaction dates:



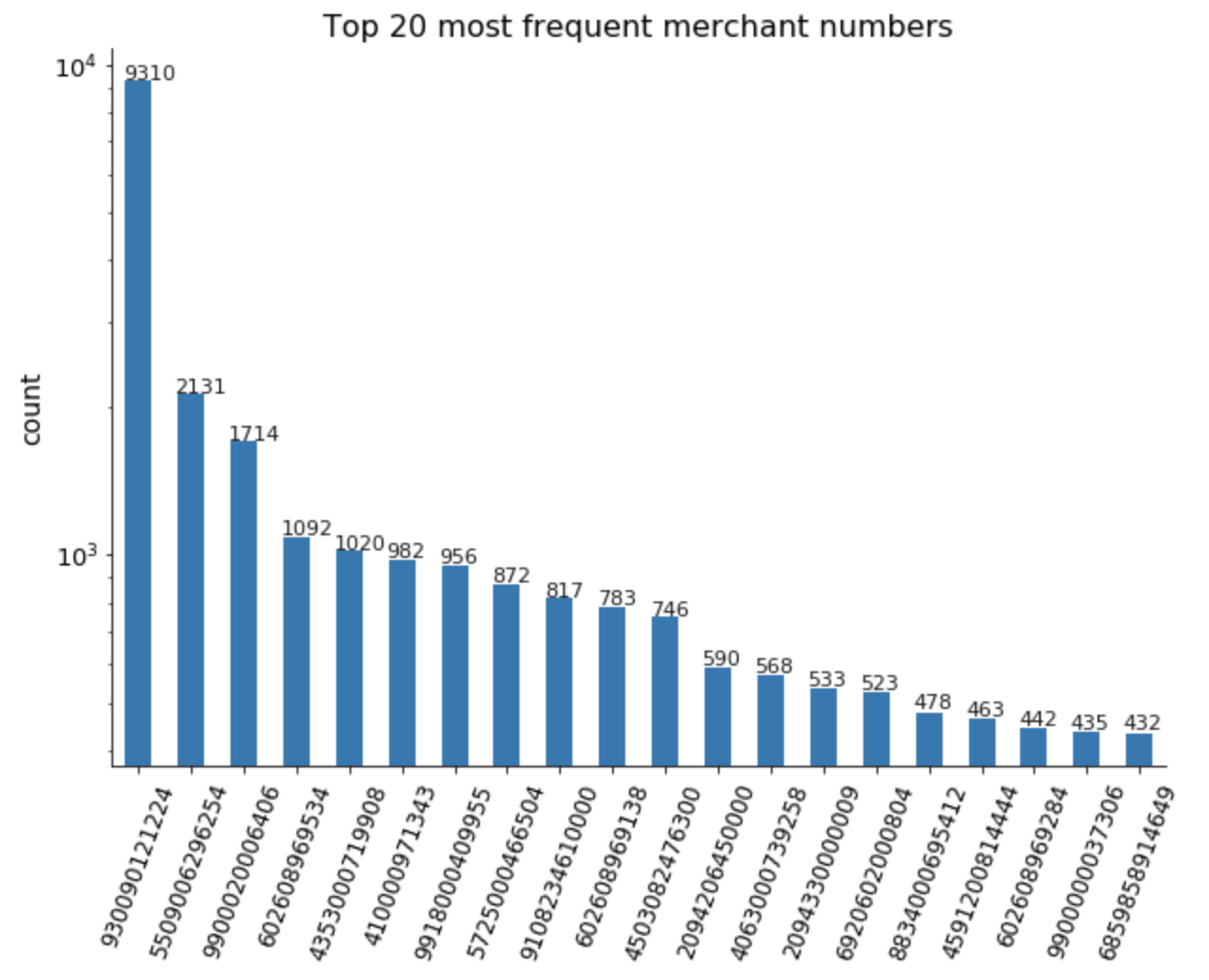
**4.**

**Field name:** Merchnum

**Field type:** object

**Description:** identity code of each merchant.

Top 20 most frequent merchant numbers:

****

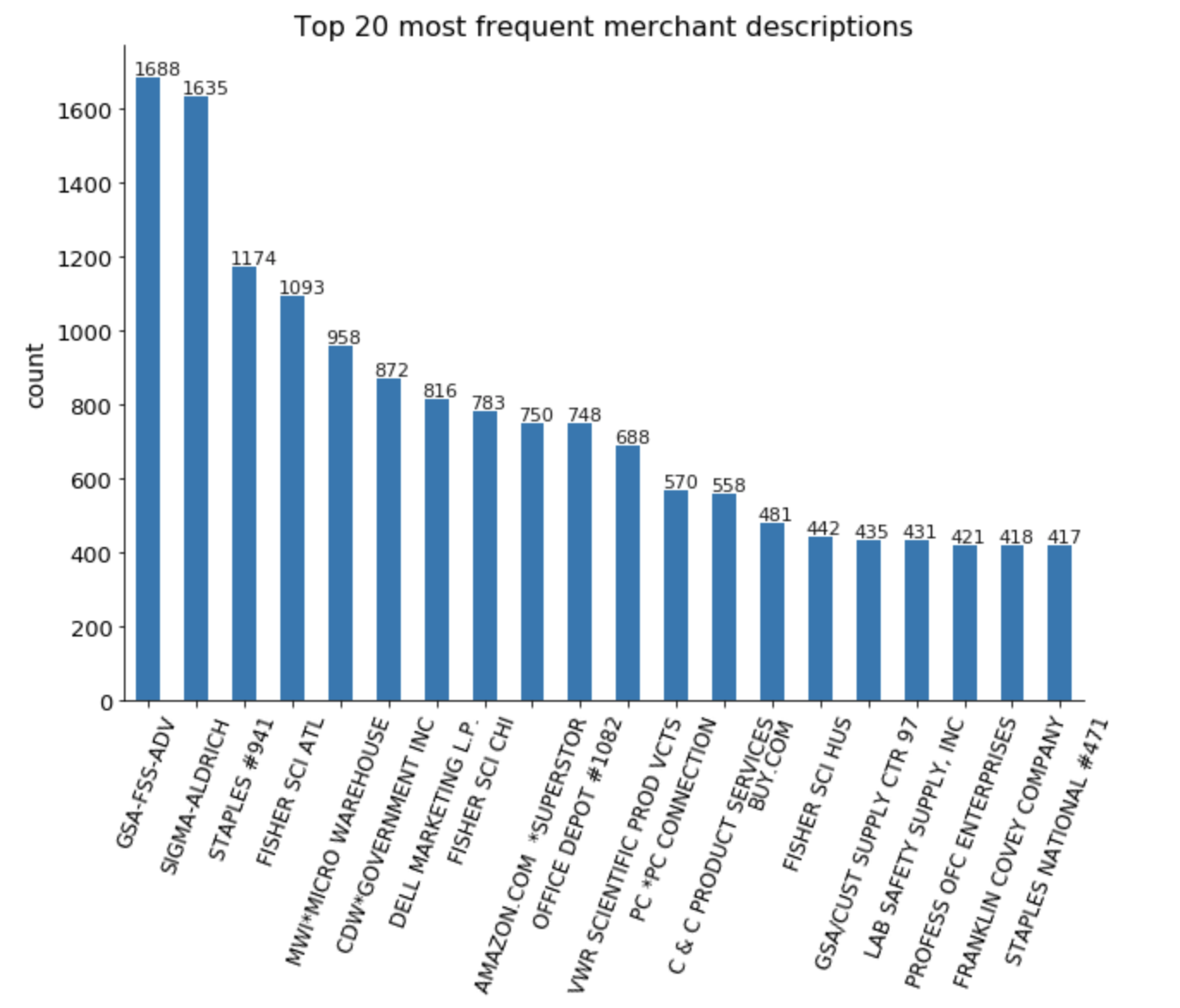
**5.**

**Field name:** Merch description

**Field type:** object

**Description:** text description of each merchant.

Top 20 most frequent mechant descriptions:



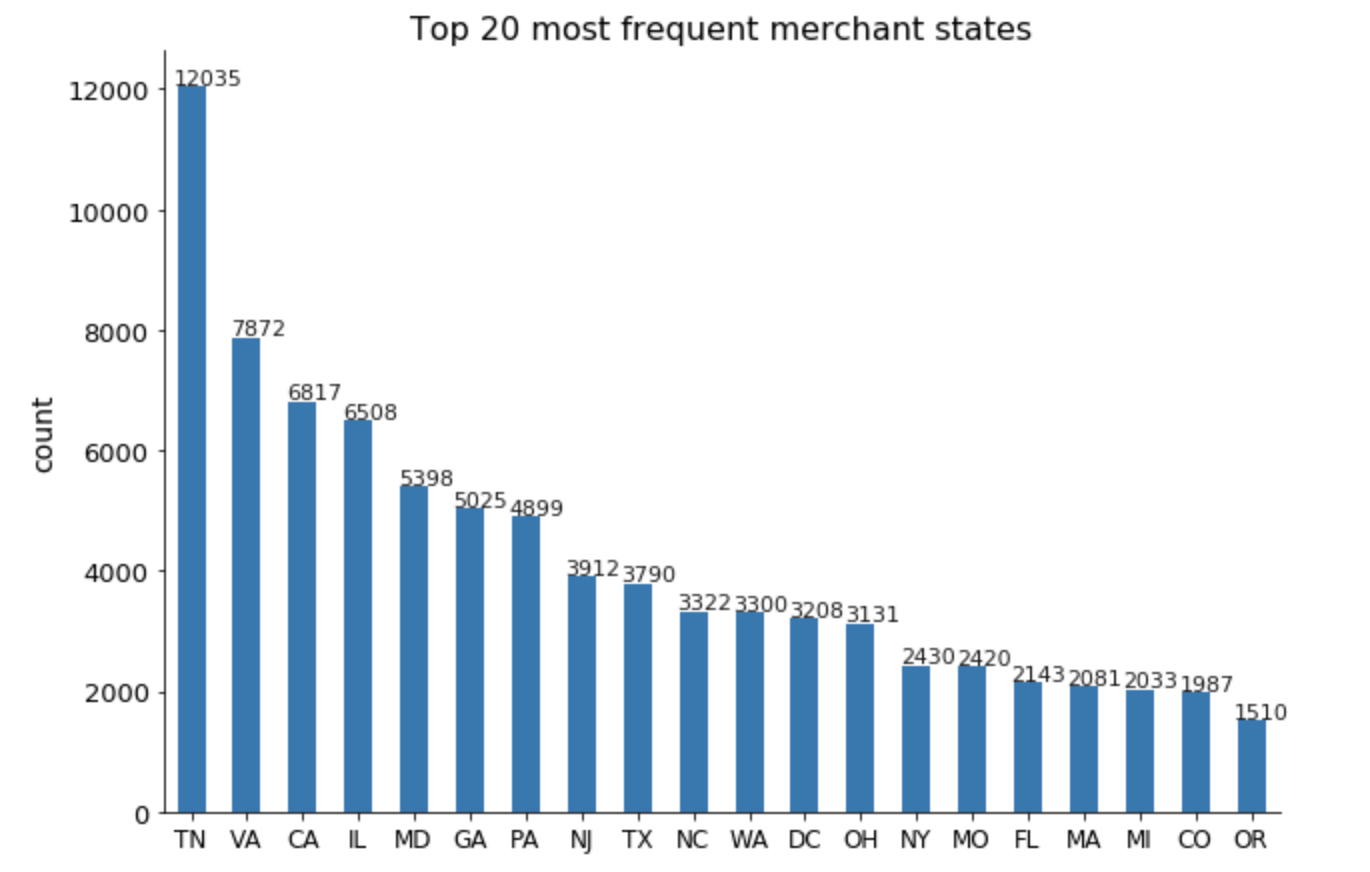
**6.**

**Field name:** Merch state

**Field type:** object

**Description:** state code of each merchant’s location.

Top 20 most frequent merchant states:



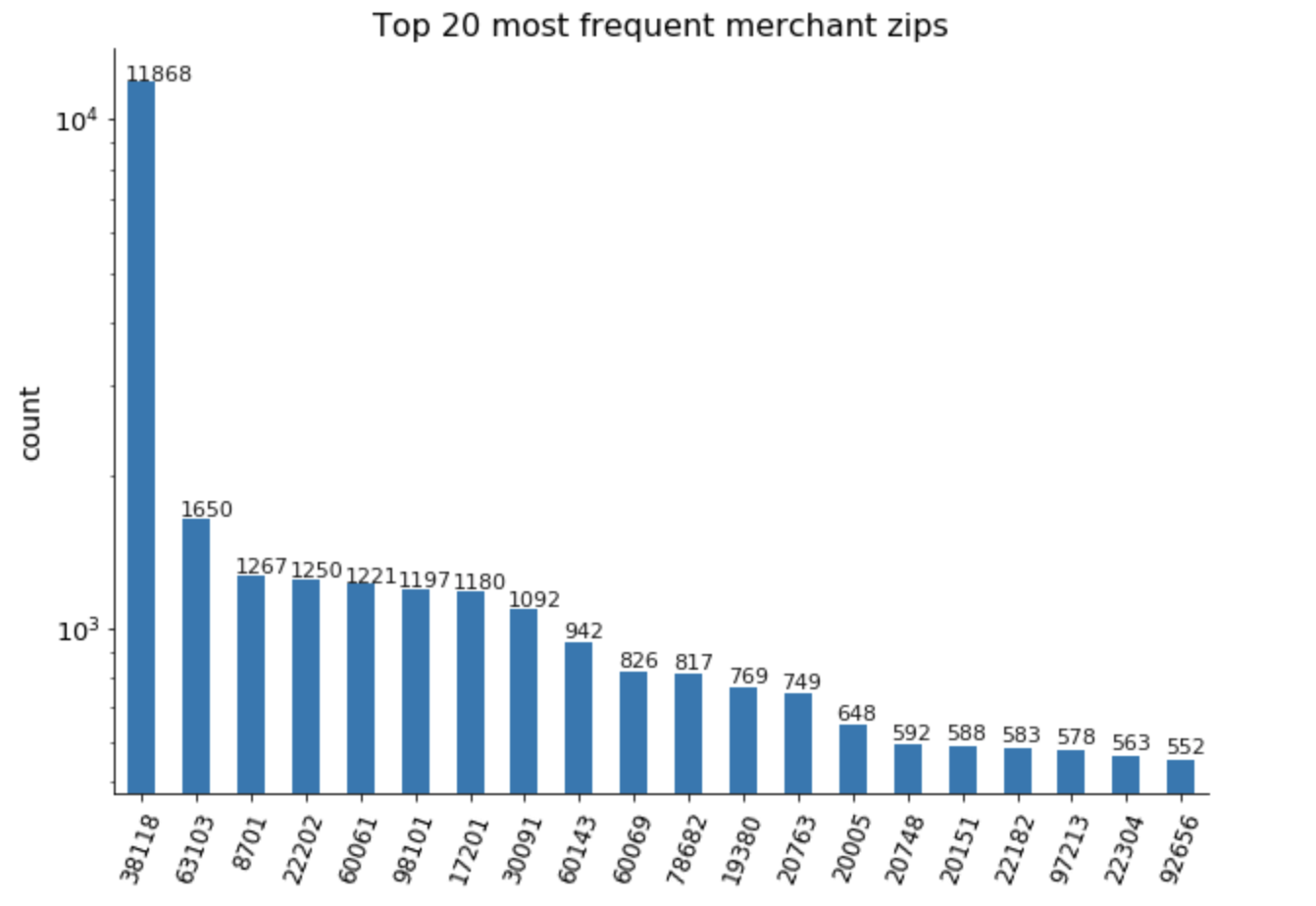
**7.**

**Field name:** Merch zip

**Field type:** float64

**Description:** zip code of a merchant.

Top 20 most frequent merchant zip codes:

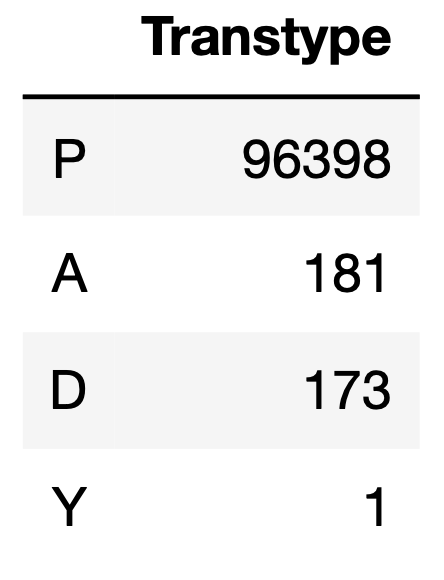


**8.**

**Field name:** Transtype

**Field type:** object

**Description:** type of a transaction.

Counts for four types:

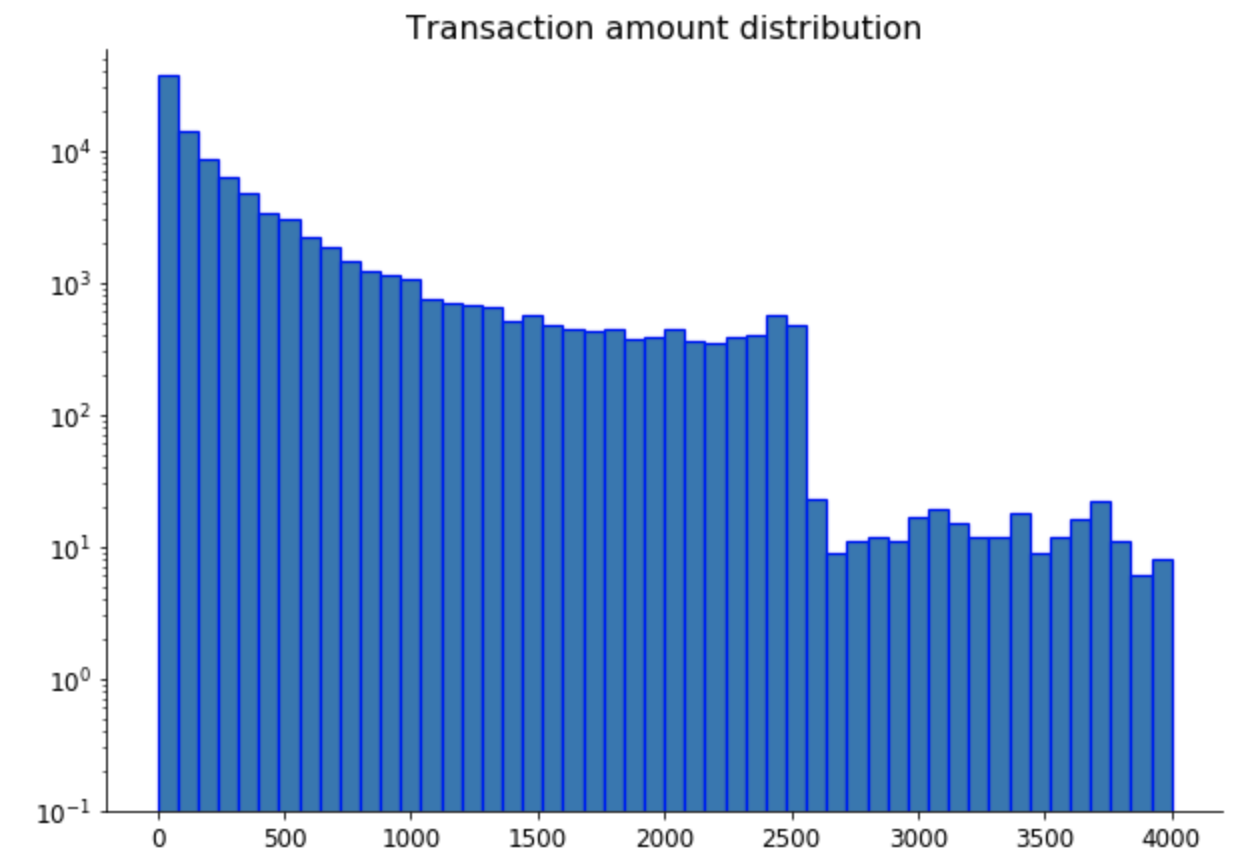
**9.**

**Field name:** Amount

**Field type:** float64

**Description:** transaction amount.

Distribution between $0 - $4000:



**10.**

**Field name:** Fraud

**Field type:** int64

**Description:** binary indicator; 1 means the record is considered as a fraud, 0 otherwise.

Counts for two labels:

