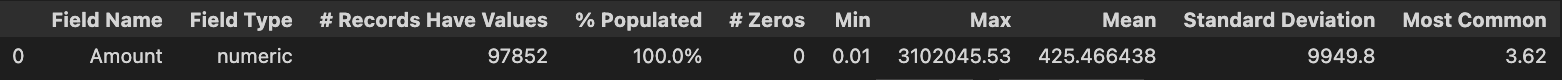
**Data Quality Report**

**1. Data Description**

The dataset is **Card Transaction** **Data**, which contains **Each Transaction’s Identifying Information** of credit cards and it’s fraud results. The data came from a hundred thousand real U.S. transaction record **over the year of 2010**. There are **10 fields** and **97,852 records**.

**2. Summary Tables**

**Numeric Fields Table**

****

**Categorical Fields Table**

**A black and white chart with numbers

Description automatically generated**

**3. Visualization of Each Field**

1. **Field Name: Recnum**

Description: Ordinal unique positive integer for each transaction record, from 1 to 97,852.

1. **Field Name: Date**

Description: Transaction date (Date). The first distribution shows the number of daily applications across time. The second distribution shows the number of weekly applications across time. The third distribution shows the number of month applications across time

**A graph showing a number of blue lines

Description automatically generated**

A graph showing a line

Description automatically generated

A graph with a line and a blue line

Description automatically generated

1. **Field Name: Cardnum**

Description: Each transaction’s card number (Cardnum). The distribution displays the top 15 occurrences of card numbers used. The most common card number is 5142148452, showing a total count of 1,192, indicating a high frequency of use which could signify a card favored for regular transactions or by a heavy user.

**A graph of a number of different sizes and colors

Description automatically generated with medium confidence**

1. **Field Name: Merchnum**

Description: Merchant number (Merchnum). The highest occurring Merchnum stands out markedly at 9,419 counts which is 930090121224, suggesting it might be a focal point for transactions. Other Merchnum counts span from just over 2,000 down to around 500, highlighting a substantial variation in transaction frequency among merchants.

A graph with numbers and a bar

Description automatically generated

1. **Field Name: Merch description**

A graph of blue bars

Description automatically generatedDescription: The description of Merchant (Merch description). The bar chart details transaction counts for various merchant descriptions, with 'GSA-FSS ADV' leading at 1,706 transactions. The data reflects a descending order of activity, illustrating a wide range in transaction volumes across merchants.

1. **Field Name: Merch State**

Description: Transaction’s State (Merch State). The bar chart presents a count of transactions by merchant state, showing Tennessee (TN) with the highest at over 12,000. A clear trend of decreasing transaction counts is visible across the states displayed.A graph of blue bars

Description automatically generated

1. **Field Name: Merch zip**

Description: Transaction’s zip code. This bar chart outlines the count distribution for the top 15 merchant ZIP codes. The ZIP code 38130 leads significantly, with nearly 12,000 counts, indicating a possible hotspot of commercial activity.A graph with numbers and a line of blue squares

Description automatically generated

1. **Field Name: Transtype**

Description: Transaction’s Type (Transtype). The bar chart illustrates the counts of different transaction types on a logarithmic scale. Type 'P' transactions dominate the chart with 97,497 occurrences, substantially outnumbering the others. Types 'A' and 'D' have a comparable presence, while 'Y' is scarcely represented with a single transaction.

A graph with blue squares

Description automatically generated

1. **Field Name: Fraud**

Description: The bar chart displays a fraud distribution where non-fraudulent transactions, labeled '0', vastly outnumber the fraudulent ones, labeled '1', with counts of 95,805 and 2,047 respectively. This visual disparity underscores the relative infrequency of fraud in the datasetA blue and white bar graph

Description automatically generated

1. **Field Name: Amount**

Description: Each transaction’s amount (Amount). This histogram displays the distribution of the "Amount" variable using a logarithmic scale on the y-axis to accommodate the wide range of values. The majority of the data is heavily concentrated at the lower end, near zero, indicating a highly skewed distribution with a sharp peak at the first bin. Additionally, there is a smaller but noticeable increase at the highest end of the range, suggesting the presence of outliers or a long tail to the right of the distribution.

A graph with numbers and a number

Description automatically generated