Data Quality Report

1. Data Description

The dataset contains credit card transactions obtained from **Mark Nigrini** website of a Tennessee located **US Government organization** for the business purposes in the **year 2010**, including information such as the card number, merchant number, transaction amount, and whether it was flagged as fraud. The data will likely be used to identify patterns and anomalies to improve fraud detection in credit card transactions. Dataset consists of **10 fields** populated with **96753 records**.

1. Summary Tables

**Table 1** : Numerical Fields Table

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Field names** | **# Populated Records** | **% populated** | **# zeros** | **Min** | **Max** | **Mean** | **Most Common** | **Std. dev** |
| Date | 96,753 | 100.00% | 0 | 1/1/2010 | 12/31/2010 | 6/25/2010 | 2/28/2010 | ~99 days |
| Amount | 96,753 | 100.00% | 0 | $0.01 | $3,102,045.53 | $427.88 | $3.62 | $10,006.14 |

**Table 2** : Categorical Fields Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **# Records With Values** | **% Populated** | **# Zeros** | **# Unique Values** | **Most Common** |
| Recnum | 96,753 | 100.00% | 0 | 96,753 | 1 |
| Cardnum | 96,753 | 100.00% | 0 | 1,645 | 5142148452 |
| Merchnum | 93,378 | 96.51% | 0 | 13,091 | 930090121224 |
| Merch description | 96,753 | 100.00% | 0 | 13,126 | GSA-FSS-ADV |
| Merch state | 95,558 | 98.76% | 0 | 227 | TN |
| Merch zip | 92,097 | 95.19% | 0 | 4,567 | 38,118 |
| Transtype | 96,753 | 100.00% | 0 | 4 | P |
| Fraud | 96,753 | 100.00% | 95,694 | 2 | 0 |

1. Visualization of each Field
2. Field Name: *Record Number*

Description: Recnum is a field that represents a **unique identifier for each transaction record**. It is an incremental integer that starts from 1 and increments by 1 for each new transaction based on the date field. This field can be used as an alternative to identify the sequence of transactions performed for the same date.

1. Field Name: *Card Number*

Description: Cardnum is a field that represents the **credit card number used for the transaction**. The Cardnum field can be useful for identifying patterns of fraud or suspicious activity associated with a particular credit card, such as multiple transactions occurring in a short time-period or transactions occurring in geographically distant locations. This field contains sensitive information and proper measures should be taken to protect this data from unauthorized access.

Visualization:

Chart, bar chart

Description automatically generated

1. Field Name: *Date*

Description: Date is a field that represents the **date on which the transaction occurred** and is a critical piece of information for analyzing transaction trends and patterns over time. The date field can be useful for identifying seasonality or other trends in transaction volume, as well as identifying periods of increased or decreased fraud activity. Additionally, the date field can be used to filter transactions based on a specific date range or to aggregate transactions by week, month, or year.

Visualization:

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

1. Field Name: *Merchant Number*

Description: Merchnum is a field that represents a **unique identifier for the merchant** where the transaction occurred. The merchnum field can be useful for identifying patterns of fraud or suspicious activity associated with specific merchants or groups of merchants, such as unusually high transaction volumes. Additionally, the merchnum field can be used to group transactions by merchants for reporting and analysis purposes.

Visualization:

Chart, waterfall chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. Field Name: *Merchant Description*

Description: Merch description is a field that provides a **brief description of the goods or services purchased from the merchant**. This field may contain text descriptions or codes that provide additional information about the transaction. The field can be useful for identifying patterns in transaction activity, such as the types of products or services that are frequently purchased, or for identifying potentially fraudulent transactions based on unusual or unexpected merchandise descriptions. This field can also be used for reporting and analysis purposes, such as grouping transactions by product category or identifying trends in purchasing behavior over time.

Visualization:

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. Field Name: *Merchant state*

Description: Merch state is a field representing the **state where the merchant is located**. The merch state field can be useful for identifying patterns in transaction activity by geographic region, such as identifying areas where a particular merchant is more popular or where fraud is more prevalent.

Visualization:

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. Field Name: *Merchant Zip*

Description: Merchant Zip is a field that represents the **zip code of the merchant's location** where the credit card transaction was made. The zip code can help identify where most of the transactions are taking place and provide insights into the merchant's location. This information can be used to identify unusual or suspicious transactions that occur outside the expected geographic area.

Visualization:

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

1. Field Name: *Transaction Type*

Description: Transtype is a field that represents the **type of transaction made with the credit card**. It provides information on whether the transaction was a Purchase(P), Authorization(A), or some other type of transaction. Transtype is a useful field for detecting fraud, as certain transaction types may be more likely to be fraudulent than others.

Visualization:

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. Field Name: *Amount*

Description: Amount is a field that represents the **dollar amount of the transaction** made with the credit card. Analyzing the Amount field can provide insights into customer behavior, such as what types of products or services are being purchased and at what price, but it is important to consider the context of the transaction. For Ex - A $500 transaction at a high-end department store may be perfectly legitimate, while a $500 transaction at a discount store may be more suspicious.

Visualization:

Chart, histogram

Description automatically generated

Chart, line chart

Description automatically generated

1. Field Name: *Fraud*

Description: Fraud is a field that **indicates whether a transaction is fraudulent or not**. A value of 0 typically indicates that the transaction is legitimate, while a value of 1 indicates that the transaction is fraudulent. Count of Fraud Transactions – 1,059.

Visualization:

Chart, waterfall chart

Description automatically generated

Chart, bar chart

Description automatically generated