Power BI Dataset Assessment Form

# Your Name : Farsan K

# Dataset's Link ( Drive link): [Transactions](https://www.kaggle.com/datasets/skullagos5246/upi-transactions-2024-dataset)

# Project Goal :

# This project aims to analyze UPI transaction data to identify spending trends, detect potential fraud, and provide actionable insights for data driven financial decisions

# Data Quality

How will you handle missing values in important fields (if any)?

I will handle missing values logically dropping only when essential fields are null, imputing or labeling others with domain relevant defaults, and validating results through SQL queries and Power BI visuals.

Provide the source of the dataset:

kaggle.com

# Data Structure

What is the file format of the dataset (e.g., CSV, Excel)?

CSV

How many records (rows) does the dataset contain?

255000

How many columns does the dataset have?

17

# 5. Data Content

List the columns in the dataset and type of data in them

transaction id (Text)

timestamp(Date/Time)

transaction type(Text)

merchant\_category(Text)

amount (INR) (Whole number)

transaction\_status (Text)

sender\_age\_group (Text)

receiver\_age\_group (Text)

sender\_state (Text)

sender\_bank (Text)

receiver\_bank (Text)

device\_type (Text)

network\_type (Text)

fraud\_flag (Whole number)

hour\_of\_day (Whole number)

day\_of\_week (Text)

is\_weekend (Whole number)

Do the columns contain consistent data (e.g., no mixed data types within the same column)?

Yes

Are there any calculated fields in the dataset (e.g., profit margin, growth rate)? If so, list them:

No

Does the dataset include any categorical variables (e.g., product category, region)? If yes, list them:

Yes, the dataset includes several categorical variables: transaction type, merchant\_category, transaction\_status, sender\_age\_group, receiver\_age\_group, sender\_state, sender\_bank, receiver\_bank, device\_type, network\_type, day\_of\_week, is\_weekend

Does the dataset include any time-based data (e.g., dates, timestamps)? If yes, explain how it is formatted:

Yes, the dataset includes time-based data: timestamp (YYYY-MM-DD HH:MM:SS), hour\_of\_day, day\_of\_week, and is\_weekend.

Are there any columns with geographical data (e.g., country, city)? If yes, list them:

Yes, the dataset includes geographical data: sender\_state.

Are there any numerical variables with large ranges or outliers? If so, how will you handle them?

Yes, the dataset includes numerical variables with large ranges or potential outliers: amount (INR).

Does the dataset include any hierarchical or multi-level categories (e.g., subcategories, regions within countries)?

No, the dataset does not include any hierarchical or multi-level categories.

Does the dataset include any columns that could be used for filtering or grouping data (e.g., product ID, customer segment)?

Yes, the dataset includes columns that can be used for filtering or grouping data: transaction type, merchant\_category, sender\_age\_group, receiver\_age\_group, sender\_state, sender\_bank, and receiver\_bank.

Explain whether the dataset allows you to explore trends, comparisons, or correlations:

Yes, the dataset allows exploration of trends, comparisons, and correlations.

# 6. Size & Performance

What is the size of the dataset in MB?

30.51 MB