Digital Egypt Pioneers

وزارة الاتصـــــــالات
وتكنولوجيا المعلومات

Ministry of Communications and Information Technology

Digital Egytion Pioneers Initiative

Online Payment Fraud Detection Analysis

A Graduation Project

Submitted in Partial Fulfillment of

Data Analysis track at

AST-Data Analyst Specialist-GIZ3_DAT1_S2

Prepared By

Kareem Mahmoud

Rawan Osama

Sama Saleh

Alaa Ahmed Taha

Mohamed Allam

Omar Ashraf

Supervised by

Dr.Mohamed Farouk

وزراة الإتصالات وتكنولوجيا المعلومات

مبادرة رواد مصر الرقمية

Online Payment Fraud Detection Analysis

مشروع تخرج

مقدم لاستيفاء جزء من متطلبات

مسار تحليل البيانات للفرقة

AST-Data Analyst Specialist-GIZ3_DAT1_S2

إعداد

آلاء أحمد طه

محمد علاّم

عمر أشرف

كريم محمود

روان أسامة

سما صالح

تحت إشراف

د/ محمد فاروق

Table of Contents

1. Introduction & Problem Definition

1.1 Background and Story of the Problem

Digital payments in India have grown massively in recent years, with credit and debit cards being used for everything from small grocery purchases to high-value ecommerce orders and travel bookings. This rapid growth brings convenience—but also a major vulnerability: **credit card fraud**.

Imagine a typical customer in India using their card across Zomato, Amazon, Uber, Big Bazaar, and local merchants. Every transaction travels through multiple systems: issuing bank, payment gateway, card network, and merchant. At each step, there is a risk of:

- Stolen card details being used online (Card Not Present fraud)
- Skimming devices capturing card data at POS terminals
- Social engineering and phishing leading to Account Takeover or Identity Theft

For banks and financial institutions, this means:

- **Financial loss** (refunds, chargebacks, operational investigation costs)
- **Customer trust erosion** (“my bank can’t protect my money”)
- **Regulatory and compliance pressure** to detect and mitigate fraud early

The dataset in this project simulates **credit card transactions in India**, featuring attributes such as card type, issuing bank, state, merchant, fraud type, and a fraud score. Out of 1,000 transactions, **286 are flagged as fraud**, giving a fraud rate of about **28.6%** in the sample. Total spend is about ₹12.2M, and fraudulent transactions account for roughly **27% of total value**.

The business challenge is therefore:

> How can we **monitor and understand fraud patterns** across customers, merchants, locations, banks, and card types, so that risk teams can act proactively instead of reacting only after fraud occurs?

1.2 Why an India-focused Dataset?

India is a particularly relevant context for several reasons:

- **High volume and growth** in card usage and e-commerce.
- **Diverse customer and merchant base** across states, income levels, and digital maturity.
- **Multiple payment players** (banks, card networks, wallets, gateways, apps) are creating complex risk surfaces.
- **Regulatory attention** on customer protection, dispute resolution, and secure digital payments.

Using an India-specific dataset allows the project to:

- Reflect realistic **merchant names** (Zomato, Swiggy, Amazon India, Flipkart, Tata Cliq, etc.).
- Capture **regional variation** in risk (states like Karnataka, Maharashtra, etc.).
- Make the insights more **transferable to real Indian banking/fintech environments**.

1.3 Business Questions

The dashboard and analysis are driven by practical questions a fraud risk team or business stakeholder would ask, such as:

1. **Overall risk & impact**
    - What is the overall **fraud rate** (count-based and amount-based)?
    - How much **money is at risk** due to fraudulent transactions?

2. **Patterns by product & channel**
   - Which **fraud types** (Card Not Present, Identity Theft, Phishing, etc.) are most common and most costly?
   - Which **transaction categories** (Food Delivery, Groceries, Electronics, Transportation, Apparel, E-commerce) show higher fraud rates?
3. **Risk by customer / card / bank**
   - Which **card types** (Visa, Mastercard, Amex, Rupay) are more exposed to fraud?
   - Which **banks** show higher fraud rate or fraud amount?
4. **Geographical & merchant risk**
   - Which **states** exhibit higher fraud risk (e.g., Karnataka, Maharashtra, West Bengal, etc.)?
   - Which **merchants and merchant locations** contribute the most to fraud volume and value?
5. **Risk scoring**
   - How does the **Fraud Score** relate to actual fraudulent cases?
   - Can we identify thresholds where risk is significantly higher?

1.4 Project Objectives and Scope

Overall Objective

To design and implement a **Power BI fraud risk dashboard** for Indian credit card transactions that provides clear, actionable insights for fraud monitoring, pattern detection, and decision-making.

Specific Objectives

- Build a **clean, well-structured data model** based on the provided transaction dataset.

- Develop **DAX measures** to quantify fraud volume, value, and rates across multiple dimensions.

- Design **professional visualizations** that answer the business questions above.

- Summarize **key insights and risk hotspots** (by card type, bank, state, merchant, category, fraud type).

- Provide a foundation that could be extended to **real-time monitoring or predictive modeling** in the future.

Scope

- Single main dataset: **Credit card transaction records in India** (1,000 rows, 14 columns).

- Time horizon: ~1 year of transactions (2023-12-16 to 2024-12-15).

- Tools: **Excel (optional) + Power BI**.

- No machine learning model—focus is on **descriptive & diagnostic analytics**, not prediction.

- Target users: **Fraud risk analysts, product managers, operations leadership**.

2. Project Plan

2.1 Timeline (Gantt-style Overview)

Phase	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Project initiation & requirements	■					
Data understanding & exploration	■	■				
Data cleaning & transformation		■	■			
Data modelling & DAX measures			■	■		
Dashboard design & visualization				■	■	
Validation, storytelling, report					■	■
Final presentation & handover						■

2.2 Milestones & Deliverables

Milestone 1 – Project Kick-off

- Stakeholder requirements documented
- List of business questions and KPIs defined
- Initial understanding of dataset structure

Milestone 2 – Data Prepared

- Data imported into Power BI
- Data types fixed (dates, numerics, categories)
- Derived columns for date (Year, Month, etc.) and risk buckets if needed

Milestone 3 – Model & Measures Ready

- Star / single-table model confirmed
- Core DAX measures created: Total Transactions, Total Fraud, Fraud Rate %, Total Amount, Fraud Amount %, Avg Fraud Score, etc.

Milestone 4 – Dashboard Prototype

- At least 2–3 report pages: Overview, Fraud Patterns, Merchant/Geography
- Slicers and interactions working
- Initial visual design and layout completed

Milestone 5 – Final Dashboard & Documentation

- Fully formatted Power BI report

- Slide deck or written report summarizing methodology, challenges, and insights
- Handover notes (data sources, assumptions, and future enhancement ideas)

2.3 Resource Allocation & Roles (6-person team)

1. **Kareem Mahmoud – Business Analyst**
    - Gather requirements
    - Define business questions & KPIs
    - Manage timeline and communication
    - Lead final story & presentation
2. **Sama Saleh – Data Engineer**
    - Handle data import from CSV/Excel
    - Ensure correct data types, encoding, and consistency
    - Create Power Query transformations (date columns, cleaning, etc.)
3. **Alaa Ahmed – Data Analyst**
    - Perform exploratory data analysis (EDA)
    - Calculate descriptive statistics (fraud rates, amount distributions)
    - Identify key patterns and anomalies

4. **Mohamed Allam – Power BI Developer / BI Specialist**
    - Build DAX measures and calculated columns (where needed)
    - Design report pages and visuals
    - Implement interactions, drill-downs, and slicers
5. **Omar Ashraf – Domain Expert (Fraud Risk / Banking)**
    - Validate if patterns make sense from a risk perspective
    - Help interpret fraud types, severity, and thresholds
    - Suggest additional useful views (e.g., fraud score cutoffs)
6. **Rawan Osama – QA & Documentation / Presentation Lead**
    - Test dashboard for correctness and usability
    - Check filters, tooltips, and calculations
    - Prepare documentation, user guide, and final slides

---

3. Methodology

3.1 Data Understanding

- Loaded the **Credit Card Fraud Risk Analysis** dataset (1,000 rows, 14 columns).
- Columns include: Transaction ID, Customer Name, Merchant Name, Transaction Date, Transaction Amount (INR), Fraud Risk, Fraud Type, State, Card Type, Bank, IsFraud, Fraud Score, Transaction Category, Merchant Location.

- Time range: about **one year** of transactions.
- Fraud flag (`IsFraud`) is binary (0/1).
- `Fraud Score` is a risk indicator from 10 to 95.

Initial checks showed:

- No missing values in key columns.
- Reasonable ranges for amounts and scores.
- Diverse categories and states.

3.2 Data Cleaning

Steps taken (in Power Query and/or Excel):

- **Type conversion**
    - Converted `Transaction Date` to proper Date type.
    - Ensured `Transaction Amount (INR)`, `Fraud Score`, and `IsFraud` are numeric.
- **Consistency checks**
    - Verified there are **no duplicate Transaction IDs**.
    - Checked that each categorical field (Card Type, Bank, Fraud Type, State, Merchant Location, Transaction Category) has valid, expected values.
- **Optional derived date fields** (if done in Power Query):
    - `Year` = YEAR([Transaction Date])

    - `Month Number` = MONTH([Transaction Date])
    - `Month Name` = FORMAT([Transaction Date], "MMM")
    - `Year-Month` = combination for trend charts

3.3 Data Transformation & Feature Engineering

To make the analysis more powerful:

- Created **Fraud vs Non-Fraud classification** using `IsFraud`:
    - 1 → Fraud
    - 0 → Legit
- Considered **Fraud Score buckets** (if implemented):
    - 0–30 → Low
    - 31–60 → Medium
    - 61–80 → High
    - 80 → Critical
- Enabled **hierarchies** for analysis:
    - Geography: State → Merchant Location
    - Business: Merchant Name → Transaction Category

These transformations make it easier to build slicers and drill-downs in Power BI.

3.4 Modelling & DAX Calculations

A simple but effective **single-table model** was used (one fact table with all attributes). On top of that, core **DAX measures** were defined, such as:

- **Volume Measures**
  - `Total Transactions = COUNTROWS(Transactions)`
  - `Total Fraud Transactions = CALCULATE([Total Transactions], IsFraud = 1)`
- **Value Measures**
  - `Total Amount = SUM(Transaction Amount (INR))`
  - `Total Fraud Amount = CALCULATE([Total Amount], IsFraud = 1)`
- **Rates**
  - `Fraud Rate % = DIVIDE([Total Fraud Transactions], [Total Transactions])`
  - `Fraud Amount % = DIVIDE([Total Fraud Amount], [Total Amount])`
- **Averages**
  - `Avg Transaction Amount = AVERAGE(Transaction Amount (INR))`
  - `Avg Fraud Transaction Amount = CALCULATE([Avg Transaction Amount], IsFraud = 1)`
  - `Avg Fraud Score = AVERAGE(Fraud Score)`

- **Context-aware measures**
  These measures are reused across visualizations with different dimensions (Card Type, State, Merchant Name, etc.), making the model highly flexible.

3.5 Dashboard Design & Visualization

Report pages were structured as:

1. **Overview Page**
   - KPI cards for total transactions, fraud transactions, fraud rate %, total amount, fraud amount, fraud amount %.
   - Line chart of fraud trend over time.
   - Bar/column charts showing fraud by Card Type, Bank, and Transaction Category.
2. **Fraud Patterns Page**
   - Bar charts for fraud by Fraud Type and Fraud Score bucket.
   - Matrix showing Fraud Type × Transaction Category by fraud amount and fraud rate.
3. **Merchant & Geography Page**
   - Map of fraud rate by State.
   - Bar charts for top merchants and merchant locations by fraud amount.
   - Detailed table for drill-in investigation.

---

4. Findings & Insights from the Analysis

These are example insights derived from the dataset you provided:

4.1 Overall Fraud Burden

- Total transactions: **1,000**
- Fraudulent transactions: **286**
- **Fraud rate (by count): ~28.6%**
- Total transaction amount: **≈ ₹12.23 million**
- Fraudulent amount: **≈ ₹3.31 million**
- **Fraud amount share: ~27.0%** of total spend

In this sample, fraud is **significant** both in volume and value.

4.2 Fraud by Card Type

From the analysis:

- **Visa**
  - 261 transactions, 88 fraud → **~33.7% fraud rate**
  - Fraud amount ~31% of Visa spend
- **Mastercard & Amex**
  - Fraud rate ~27–28%
- **Rupay**
  - Lowest fraud rate (~24.7%) and lowest fraud amount share.

Insight: In this dataset, **Visa cards are relatively more exposed to fraud**, suggesting a need for closer monitoring or stricter controls on this network.

4.3 Fraud by Fraud Type

Fraud types show different risk levels:

- **Highest fraud rates**:
  - **Identity Theft**: ~32.8% fraud rate, ~33% of amount fraudulent
  - **Card Not Present**: ~31.2% fraud rate, ~30% fraud amount share
- **Moderate:**
  - Phishing, Card Skimming, Account Takeover around 24–28%

Insight:
Digital and identity-driven frauds (**Identity Theft** and **Card Not Present**) appear to be **major drivers of risk**, consistent with online and remote commerce threats.

4.4 Fraud by State

Ranking states by fraud rate:

- **Highest fraud rates**:
  - Karnataka (~35.4%)
  - Maharashtra (~33.6%)
  - West Bengal (~31.7%)
- **Lower fraud rates**:
  - Kerala (~22.3%)

    - Delhi (~22.8%)
    - Gujarat (~24.5%)

Insight:
States like **Karnataka and Maharashtra** are risk hotspots in this dataset and may reflect higher digital activity or concentration of high-risk merchants/customers.

4.5 Fraud by Merchant & Category

- **Top merchants by fraud amount** include:
    - Zomato, Big Bazaar, Tata Cliq, Myntra, Reliance Digital, Swiggy, Uber, Flipkart.
- **By Transaction Category:**
    - Highest fraud rates in:
        - **Food Delivery** (~32.4%)
        - **E-commerce** (~31.0%)
        - **Electronics** (~30.2%)

Insight:
Categories associated with **online or app-based purchases (Food Delivery, E-commerce, Electronics)** show higher fraud risk, aligning with typical vulnerabilities in card-not-present transactions.

4.6 Fraud Score Behavior

- Overall average fraud score: ~53.2
- Average fraud score for fraud transactions: ~53.9

Insight:
Fraudulent transactions generally have **slightly higher scores**, indicating the score has signal, but future work could focus on **optimizing thresholds** and turning this into a predictive model.

---

5. Conclusion and Future Work

5.1 Project Conclusion

This project developed a **Power BI–based fraud risk analytics dashboard** using an India-focused credit card transactions dataset. Through structured cleaning, transformation, modeling, and visualization, the team:

- Quantified the **scale of fraud** in the sample (≈28.6% by count, ≈27% by amount).
- Identified **high-risk combinations** of card types, fraud types, states, merchants, and categories.
- Delivered a **multi-page dashboard** suited for both high-level overviews and detailed investigations.

5.2 Future extensions can include:

- Integrating **real-time or near-real-time** transaction feeds.
- Building a **machine learning model** for fraud prediction using the Fraud Score and other features.
- Adding **customer segmentation** (new vs existing, high vs low spenders) and **merchant risk profiling**.
- Implementing **alerting logic** based on dynamic thresholds and patterns uncovered in this analysis.