Global Banking Fraud Detection & Risk Analysis

An end-to-end data analytics case study using Excel, SQL, and Power BI

Garv Dudy
Data Analyst Project | December 2025

Executive Summary

Modern banking systems process millions of transactions daily, making it increasingly difficult to detect fraud early without data-driven monitoring. This project analyzes a large-scale banking transaction dataset to uncover transaction behavior, risk exposure, and fraud patterns using an end-to-end analytics pipeline.

Key Highlights:

- Analyzed 153,000 + banking Transactions
- Built a full pipeline from Excel → SQL → Power BI
- Identified fraud concentration, risk signals, and spending behavior
- Delivered executive-level and analyst-level dashboards for decision making.

Business Problem

Banks face a critical challenge: high Transactions volume increases fraud risk, but manual monitoring cannot scale. Fraud often represents a small percentage of total transactions, yet even a minor increase can lead to significant financial losses and reputational damage.

Key Challenges Addressed:

- Detecting Fraud Hidden within large transaction volumes
- Understanding risk beyond confirm fraud
- Identifying high-risk merchants, regions, and channels
- Supporting early intervention, not just reactive controls

Project Objectives

This project aims to move beyond simple fraud counts and instead analyze transaction behavior holistically to support smarter banking decisions.

Objectives:

- Analyze transaction volume and monetary flow across the system
- Measure Fraud rate and transaction risk exposure
- Identify high-risk merchant categories and transaction types
- Build Dashboards that support executives, risk teams and analysts

Data Overview

The Dataset represents realistic banking transaction behavior over multiple years and channels, enabling deep behavioral analysis.

Dataset Characteristics:

- 153000+ Transactions Records
- Time-based data (year, quarter, month)
- Merchant categories and transaction channels
- Geographic indicators (countries/regions)
- Risk Score and Fraud Flag indicators

Methodology (end-to-end Pipeline)

To Ensure scalability and realism, the project follows a structured analytics pipeline like real banking environments.

Workflow:

- Excel: Data cleaning, formatting, and preparation
- SQL: Querying, Aggregations, fraud analysis, and KPI calculations
- Power BI: Interactive dashboards and storytelling visuals

This Layered approach ensures accuracy, performance, and clarity.

Data Cleaning & Preparation (Excel)

Raw Transactional data often contains inconsistencies that can distort analysis. Significant effort was invested in preparing the dataset for reliable analytics.

Cleaning steps performed:

- Standardized column formats and naming conventions
- Handled missing and inconsistent values
- Split and normalized columns for SQL compatibility
- Optimized structure for Power BI modeling

Result: A clean, analysis-read dataset suitable for scalable querying and visualizations.

SQL Analysis Layer

SQL acted as the analytical engine of this project, transforming cleaned transactional data into structured insights before visualization. Rather than pushing raw data directly into Power BI, SQL was used to validate trends, calculate risk metrics, and isolate fraud behavior, ensuring analytical accuracy and scalability.

Through SQL, this project replicated how banks typically analyze transaction data at the backend level before exposing results to dashboards.

Key SQL analysis performed.

- Aggregation Transaction volume and Total monetary flow
- Calculated Fraud rate and avg risk score
- Identified high-risk Merchant category and Transaction type
- Time-series trend extraction

This layer ensure accurate KPIs and reusable logic before visualization.

Key Performance Indicators (KPIs)

These KIPs act as an early-warning system for banking risk. Rather than isolating fraud alone, they measure transaction intensity, monetary exposure, and behavioral risk simultaneously, enabling precative intervention.

Together, these indicators answer four critical questions:

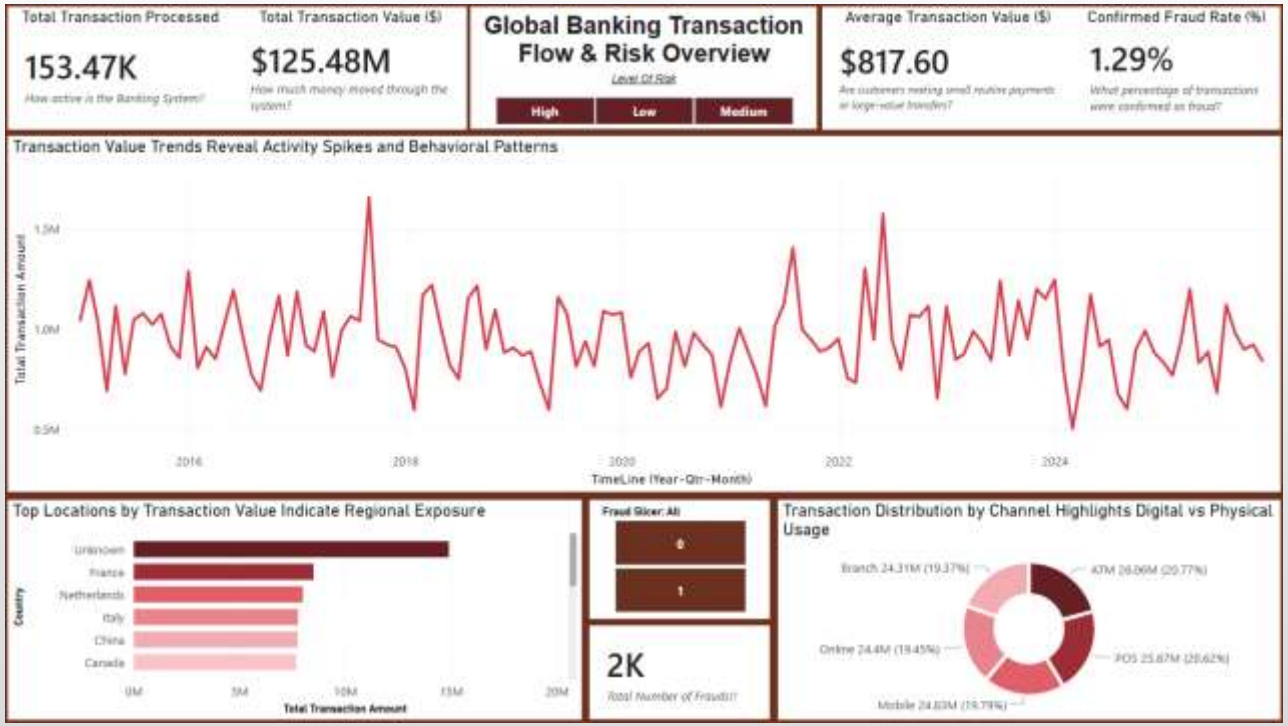
- Total Transaction flow reflects system activity & scale
- Fraud rate highlights hidden risk within normal transactions
- Average fraud transaction value signals financial impact severity

Total Transaction Processed	Total Transaction Value ($)	Average Transaction Value ($)	Confirmed Fraud Rate (%)
153.47K	$125.48M	$817.60	1.29%
How active is the Banking System?	*How much money moved through the system?*	*Are customers making small routine payments or large-value transfers?*	*What percentage of transactions were confirmed as fraud?*

Global Banking Transaction Flow & Risk Overview

What this dashboard answers:

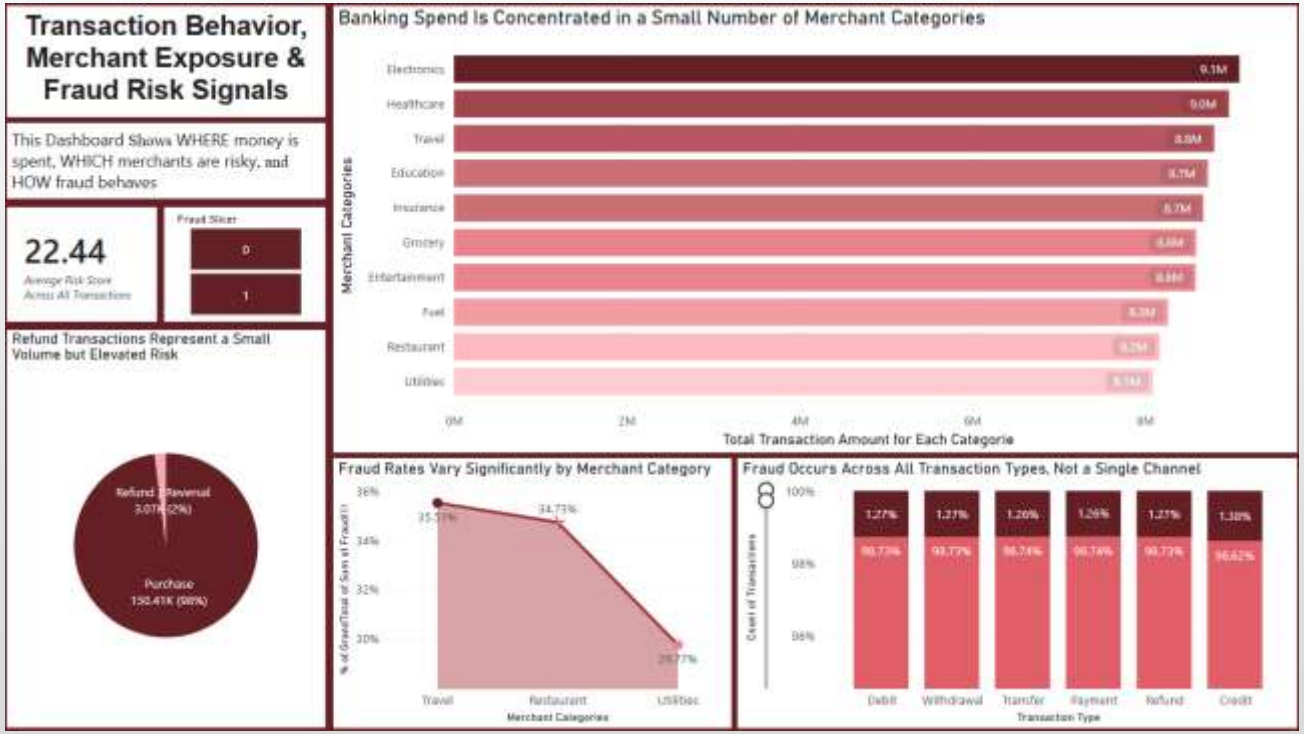
- How active is the global banking system overtime?
  - System was active with an average of 1M Transaction Each Month, except some outliers.
- Where is money concentrated geographically?
  - Most amount is transferred from unknow country with 14M+ and France holding 2nd with 8.5M+, then most of the countries having 7M in avg spending. India as last in list with 6.5M Transaction
- How are customers transacting (digital vs physical)?
  - Well customers preferring Digital transactions (Mobile + POS + Online) which is 75M+ in Total more then Physical with only 50M+ Transfers (ATM + Branch Visit).
- What is the overall fraud exposure?
  - There is in total of 2K Frauds with 18.65M of Money and most important think to note is the Average spending which comes out to be $9.45K each Transaction in case of fraud, whereas the normal Avg spending is around $800
  - Canada is the 2nd Know country for fraud transaction after Unknow with 1.6M+ total in money.



Transaction Behavior, Merchant Exposure & Fraud Risk Signals

What this dashboard reveals:

- While merchant categories dominate spending
  - Money spent on electronics is the highest among all with 9.1M+, Healthcare has a total of 9.0M+ spend. One major think to notice is they all have similar spending amount.
- Where fraud rates vary significantly by category
  - Approx 1.65M+ was detected as fraud in restaurants, Education holds a total of 1.55M+ in fraud whereas Online Retail is the lest spent in fraud.
  - Transaction detected as fraud in Entertainment, Healthcare and ATM withdrawal are nearly equivalent to each other.
- How refunds and reversals differ in risk behavior
  - In total There was 2% of Refund/Reversal which makes 3.07K in numbers. Whereas confirmed purchase was 150.41K in total.
  - For Fraud Refund there was only 0.05K whereas for Fraud purchase it's 1.93K
- Whether fraud is concentrated or distributed across transaction types
  - In shot all Transaction Type Faced fraud equally, but zooming in gives us that credit was used slightly more with approx 0.11% in frauds then any other means.



Key Findings and Strategic Recommendations

Key Findings:

The analysis reveals that banking fraud is not driven by transaction volume alone but emerges from behavioral patterns across merchant categories, transaction intent, and refund activity. While the overall fraud rate remains low, its financial impact and distribution highlight the need for proactive risk monitoring rather than reactive controls.

- However, Fraud Rate remains low overall (~1.29%) but is persistent and makes a total of 18.65M in Transaction Amount.
- Risk is destituted across all transaction types
- Certain merchant categories exhibit elevated fraud rates
- Refund transactions show higher relative risk despite low volume
- High Transaction Value does not always correlate with fraud

Strategic Recommendations:

- Implement behavior-based risk scoring, not value-based flags
- Increase monitoring on refund and reversal patterns
- Prioritize merchant category risk profiling
- Shift fraud detection from reactive alerts → predictive signals

Conclusion

This project demonstrates how an end-to-end analytics pipeline can convert large-scale transactional data into **actionable banking intelligence**. By integrating Excel, SQL, and Power BI, the analysis moves from raw data to executive insight in a structured, scalable way.

Final Outcomes achieved:

- Studied and analyze the data using analytical thinking.
- Cleaned Data with best practice in Excel
- Using SQL queried the data before moving to visualize it
- Created pipelines between three top software
- Built a realistic banking analytics pipeline
- Delivered executive and analyst-focused dashboards
- Identified fraud patterns and behavioral risk signals
- Created a framework applicable to real world banks