Business Problem Statement

A rapidly scaling digital payments fintech company has observed a significant increase in fraudulent UPI-based transactions across multiple consumer payment channels such as UPI Transfer, QR-based payments, Merchant Pay, and Wallet-to-UPI conversions. These fraudulent activities are leading to major financial losses, poor customer trust, disputes, operational escalations, regulatory compliance pressure, and increased overhead in manual fraud review cycles.

The Fraud Risk, Payments, and Data Intelligence teams want to understand:

- which payment categories are most vulnerable
- what transaction behaviors indicate fraudulent intent
- which user accounts repeatedly attempt or receive fraudulent transactions
- the financial exposure driven by specific fraud patterns
- how fraud evolves over time and which time windows show attack surges

The company aims to use cloud-based analytics to detect fraud patterns proactively, reduce financial leakage, strengthen fraud rules, and guide future ML-based risk scoring.

You are tasked with analyzing UPI transaction data stored on AWS to answer the core business question:

“How can the organization leverage cloud data analytics to identify fraud exposure patterns, detect risky accounts, and strengthen fraud prevention strategies to reduce financial loss in UPI payments?”

Deliverables

1. **Cloud Data Preparation & Integration (AWS)**
    - Store raw UPI transactions in S3.
    - Use AWS Glue for schema detection and table cataloging.
    - Build serverless SQL layer on Athena.
2. **Fraud Analytics (SQL – Athena)**
    - Derive KPIs and risk signals from transactional data.
    - Identify top fraud categories, high-risk accounts and fraud-prone channels.
    - Perform investigative fraud behavior segmentation.
3. **Visualization & Risk Insights (Power BI)**
    - Build a Fraud Intelligence Dashboard including KPIs such as Fraud Rate %, Avg Ticket Exposure, Fraud by Type, Payment Channel Vulnerabilities and High-Risk Sender Accounts.

- Create executive-level visuals for Fraud Risk Teams.

4. **Report & Presentation**
    - Document business problem, architecture, methodology, KPI outcomes, findings and prioritized recommendations for leadership.
    - Provide strong actionable insights to define fraud response & next phase modeling roadmap.
5. **GitHub Repository**
    - Maintain clean structured folder with SQL queries, architecture PNG, dataset reference, Power BI PBIX file, and project report PDF.
    - Clearly include README.md describing objectives, workflow, tech stack, insights & value.

Expected Outcomes

- Cloud-driven scalable fraud analytics foundation for future ML scoring.
- Clear identification of high-risk payment pathways and fraud typologies.
- Prioritization of top sender / receiver accounts for monitoring and blocking strategies.
- Executive insights to reduce fraud losses, optimize rule tuning and improve customer trust.