# Business Case: Evaluating Fraud Risk Through Data Science

## Context & Business Problem

Online payment fraud is a **major risk** for fintech companies. To combat fraudulent activity, fraud analytics teams require **clean, well-structured, and insightful data** to detect suspicious transactions efficiently.

This case study will assess **your ability to build a structured dataset and derive meaningful fraud insights**, ensuring data is prepared for effective fraud detection.

### Problem Statement

You have just joined a **fintech company's fraud analytics team**. Your first task is to **design a small-scale fraud analytics pipeline** and **generate key insights** from online transaction data.

**Key Challenges:**

1. **Clean and structure transaction data** for fraud detection.
2. **Generate fraud insights**, such as fraud rates by time of day, transaction amount, and device type.
3. **Ensure that fraud analysts can efficiently query and use the data**.

Data and metadata can be found here.

### Evaluation Criteria

- **Feature Engineering & Data Processing:** How well do you construct meaningful features from raw transaction data?
- **Risk Modeling & Anomaly Detection:** How effectively does your approach classify suspicious vs. non-suspicious transactions?
- **Explainability & Transparency:** Can your model's predictions be interpreted by risk officers and regulators?
- **Data Pipeline Design:** Can your approach be integrated into an AML monitoring system?

**1. Data Preparation & Pipeline Design**

- Effective handling of missing values, duplicates, and outliers.
- Well-structured fraud dataset optimized for easy querying and analysis.
- Efficient and scalable data processing approach.

**2. Fraud Analytics & Insights**

- Identification of key fraud trends (e.g., by time of day, transaction type, device).
- Use of basic analytics to generate meaningful fraud risk metrics.
- Clear and insightful presentation of trends using simple charts.
- Integration of external fraud intelligence sources (e.g., an internet-based search for emerging fraud patterns).

**3. Clarity & Simplicity**

- Readability and organization of code and documentation.
- Usability of the fraud dataset for fraud analysts.
- Logical and concise presentation of insights.

**4. Alerting Mechanism & Real-time Monitoring**

- Implementation of an alert system (e.g., email, SMS, Slack, or webhooks) to flag suspicious transactions or anomalies.
- Customization of alerts based on risk levels and fraud patterns.

**5. Time Management & Prioritization**

- Focus on high-value tasks rather than unnecessary complexity.
- Justification of trade-offs made in simplifying or optimizing work.

# Submission Guidelines

Candidates should submit:

**1. Fraud Analytics Data Pipeline (Python or SQL Script)**

- Loads, cleans, and structures the dataset
- Handles missing values, duplicates, and outliers
- Creates a structured fraud analytics table with useful fraud risk features

**2. Fraud Analytics Summary Report (Short Report or Notebook)**

- 1-2 pages or a Jupyter Notebook summary
- Summary of key fraud insights (percentages, patterns, trends)
- 2-3 simple visualizations (bar charts, histograms, line plots)
- High-level explanation of fraud risk factors found

- What would you do to improve fraud rates?

**3. Submission format:** PDF Report, Jupyter (Colab) notebook. Github repository a huge plus.