Credit Card Fraud Risk Analysis – Project Documentation

---

1. Project Overview

This project focuses on analysing **credit card transaction data** to identify fraudulent patterns, risk levels, and transaction behaviour using **SQL and Power BI**.

The objective was to:

- Clean and analyze large transactional data using SQL
- Identify fraud trends and high-risk transactions
- Visualize insights using Power BI dashboards

---

2. Tools & Technologies Used

Tool	Purpose
MySQL	Data storage, querying, analysis
Power BI Desktop	Data visualization & dashboard creation
SQL	Data cleaning, aggregation, risk analysis
CSV Dataset	Source data

---

3. Dataset Description

- **Total Records:** ~284,807
- **Fraud Transactions:** 492
- **Fraud Rate:** ~0.17%
- **Features:**
    - Time
    - V1 – V28 (anonymized transaction attributes)
    - Amount
    - Class (0 = Legit, 1 = Fraud)

---

4. Data Loading & Preparation (SQL)

Step 1: Database Creation

```
CREATE DATABASE sql_risk_analysis_queries;

USE sql_risk_analysis_queries;
```

Step 2: Table Creation

```
CREATE TABLE credit_card_data (
    Time INT,
    V1 FLOAT, V2 FLOAT, V3 FLOAT, V4 FLOAT, V5 FLOAT,
    V6 FLOAT, V7 FLOAT, V8 FLOAT, V9 FLOAT, V10 FLOAT,
    V11 FLOAT, V12 FLOAT, V13 FLOAT, V14 FLOAT, V15 FLOAT,
    V16 FLOAT, V17 FLOAT, V18 FLOAT, V19 FLOAT, V20 FLOAT,
    V21 FLOAT, V22 FLOAT, V23 FLOAT, V24 FLOAT, V25 FLOAT,
    V26 FLOAT, V27 FLOAT, V28 FLOAT,
    Amount FLOAT,
    Class INT
);
```

Step 3: Data Import

```
LOAD DATA INFILE 'creditcard.csv'
INTO TABLE credit_card_data
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
```

---

5. Data Validation & Cleaning

✓ **Record Count Validation**

```
SELECT COUNT(*) FROM credit_card_data;
```

✓ **Missing Value Check**

```
SELECT
    SUM(CASE WHEN Time IS NULL THEN 1 ELSE 0 END) AS missing_time,
    SUM(CASE WHEN Amount IS NULL THEN 1 ELSE 0 END) AS missing_amount,
```

```
    SUM(CASE WHEN Class IS NULL THEN 1 ELSE 0 END) AS missing_class
FROM credit_card_data;
```

---

6. Exploratory Data Analysis (EDA)

6.1 Fraud vs Legit Transactions

```
SELECT Class, COUNT(*) AS transaction_count
FROM credit_card_data
GROUP BY Class;
```

6.2 Transaction Amount Analysis

```
SELECT
    AVG(Amount) AS avg_amount,
    MIN(Amount) AS min_amount,
    MAX(Amount) AS max_amount
FROM credit_card_data;
```

---

7. Risk-Based Analysis (Core Part)

7.1 Risk Segmentation Logic

```
CASE
    WHEN Amount >= 2000 THEN 'High Risk'
    WHEN Amount BETWEEN 500 AND 1999 THEN 'Medium Risk'
    ELSE 'Low Risk'
END
```

7.2 Risk Distribution

```
SELECT
    CASE
        WHEN Amount >= 2000 THEN 'High Risk'
        WHEN Amount BETWEEN 500 AND 1999 THEN 'Medium Risk'
        ELSE 'Low Risk'
    END AS risk_segment,
    COUNT(*) AS total_txns
FROM credit_card_data
```

```
GROUP BY risk_segment;
```

7.3 Fraud Rate by Risk Segment

```
SELECT
   risk_segment,
   COUNT(*) AS total_txn,
   SUM(CASE WHEN Class = 1 THEN 1 ELSE 0 END) AS fraud_cases,
   ROUND(SUM(CASE WHEN Class = 1 THEN 1 ELSE 0 END) * 100.0 / COUNT(*), 2) AS fraud_rate
FROM (
   SELECT *,
        CASE
          WHEN Amount >= 2000 THEN 'High Risk'
          WHEN Amount BETWEEN 500 AND 1999 THEN 'Medium Risk'
          ELSE 'Low Risk'
        END AS risk_segment
   FROM credit_card_data
) t
GROUP BY risk_segment;
```

---

8. Time-Based Fraud Analysis

```
SELECT
   FLOOR(Time / 3600) AS hour_bucket,
   COUNT(*) AS total_txn,
   SUM(CASE WHEN Class = 1 THEN 1 ELSE 0 END) AS fraud_txn
FROM credit_card_data
GROUP BY hour_bucket
ORDER BY fraud_txn DESC;
```

---

9. Power BI Dashboard Overview

Dashboard Sections:

- **Executive Summary**
  - Total Transactions
  - Fraud Count
  - Fraud Percentage
- **Risk Distribution**
  - Low / Medium / High Risk Donut Chart
- **Transaction Amount Analysis**
  - Histogram of amounts
- **Fraud Trend Analysis**
  - Hour-wise fraud patterns

---

10. Key Insights

- Fraud rate is extremely low (~0.17%) but high-value transactions pose greater risk.
- Majority of fraud occurs in **medium and high risk segments**.
- Certain time windows show higher fraud concentration.
- Risk-based segmentation helps in targeted fraud monitoring.

---

11. Business Value

- Enables early fraud detection
- Helps banks prioritize investigations
- Reduces financial loss
- Supports data-driven decision-making

Conclusion This project demonstrates effective use of SQL and Power BI to analyze, visualize, and interpret financial transaction data for fraud detection.