**Fraud Detection and Transaction Trends in Credit Card Data: An Analytical Approach**

**Introduction**

This report analyzes credit card transaction data from 2019 and 2020, focusing on fraud detection, customer spending behavior, and merchant performance. Utilizing SQL Server for data processing and Power BI for visualization, we examined over one million records across 24 columns. The study aims to identify fraudulent transactions, understand spending trends, and improve fraud detection methodologies.

## ****Objectives of the Project****

The objective of this project is to analyze credit card transactions to detect fraudulent activities, understand customer spending behavior, and assess merchant performance. This involves identifying patterns in fraudulent transactions, detecting anomalies using statistical techniques like Z-scores, and evaluating revenue trends across different merchants. Additionally, the project aims to provide actionable insights that help mitigate fraud risks, improve transaction monitoring, and enhance financial security through data-driven decision-making.

**Dataset Overview**

The dataset consists of 24 columns, including:

* **Transaction Information:** id, trans\_date\_trans\_time, amt
* **Customer Information:** cc\_num, first, last, gender, dob
* **Merchant Information:** merchant, category, merch\_zipcode, merch\_lat, merch\_long
* **Location Information:** city, state, zip, lat, long
* **Fraud Indicator:** is\_fraud

**Data Preprocessing and Outlier Detection**

**Handling Missing Values**

* **Numerical Columns:** Missing values were filled with the median or mean.
* **Categorical Columns:** Mode imputation was used to fill missing values.
* **Excessive Missing Data:** Transactions with significant missing values were removed to maintain data integrity.

**Outlier Detection Using Z-Scores**

* **Method:** The Z-score method was applied to transaction amounts.
* **Threshold:** Transactions with Z-scores greater than 3 standard deviations were flagged as potential fraud.
* **Impact:** This method helped in detecting unusually high transaction amounts that deviate from normal spending behavior.

**Key Questions Answered**

1. **What percentage of transactions were fraudulent?**
   * Fraudulent transactions accounted for a small percentage but had significantly higher average amounts.
2. **Which merchants had the highest fraud rates?**
   * "Fraud\_Padberg\_Welch" had the highest fraud rate at **33.9%**.
3. **Did fraudulent transactions have a pattern over time?**
   * Fraudulent transactions occurred in clusters, with certain months showing spikes.
4. **How does customer spending behavior vary by gender?**
   * Male customers contributed to **45.41%** of transactions, while females accounted for **54.59%**.
5. **Which categories recorded the highest spending?**
   * "Grocery\_POS" recorded the highest total spending at **$11.69M**.

**Analysis and Insights**

**Customer Analysis**

* **Spending Behavior:** The top 10% of customers contributed significantly to total revenue.
* **Spending Trends:** A decline from **$419,210.75** in 2019 to **$62,102.18** in 2020 was observed.
* **High-Value Transactions:** Identified customers with the highest average spending per transaction.

**Merchant Analysis**

* **Top Performing Categories:** "Grocery\_POS" recorded the highest total spending at **$11.69M**.
* **Revenue Growth:** Analyzed year-over-year changes in merchant revenue.
* **Fraudulent Merchants:** "Fraud\_Padberg\_Welch" had the highest fraud rate (**33.9%**).

**Fraud Detection**

* **Fraud vs. Non-Fraudulent Transactions:** Fraudulent transactions had significantly higher amounts.
* **Time Between Frauds:** Average time between fraudulent activities was analyzed per merchant.
* **Z-Score Outlier Detection:** High-value fraudulent transactions were identified.

**Power BI Dashboards**

**1. Merchant Performance and Insights Dashboard**

**Key Metrics:**

* **Total Revenue:** $73.69M
* **Total Transactions:** 1M
* **Unique Customers:** 943
* **Total Merchants:** 693

**Visualizations:**

* **Slicers:** Merchant-based filtering.
* **Area Chart:** Yearly and monthly transaction trends, with **December 2019** recording the highest amount at **$9.92M**.
* **Donut Chart:** Top 3 merchants by fraudulent transactions, led by "Fraud\_Padberg\_Welch" at **33.9%**.
* **Bar Chart:** Total amount by merchant, with "Fraud\_Kilback\_LLC" leading at **$315.13K**.

**2. Customer Spending Insights Dashboard**

**Key Metrics:**

* **Male Customers:** **45.41%** of total spending.
* **Female Customers:** **54.59%** of total spending.

**Visualizations:**

* **Line Chart:** Spending trends over **2019-2020**, showing a sharp decline in 2020.
* **Pie Chart:** Gender-wise spending distribution.
* **Bar Charts:**
  + **Top 10 spending cities**, with **"Meridian"** leading at **$0.36M**.
  + **Top 10 spending categories**, with **"Grocery\_POS"** at **$11.69M**.

**Recommendations**

1. **Enhance Fraud Detection Models:**
   * Implement **machine learning algorithms** for real-time fraud detection.
2. **Transaction Monitoring:**
   * Use **dynamic thresholds** to flag unusual spending patterns.
3. **Merchant Risk Assessment:**
   * Establish **fraud risk scores** for merchants based on historical data.
4. **Customer Education:**
   * Raise awareness about fraud prevention and secure transactions.
5. **Policy Improvements:**
   * Collaborate with **financial institutions** to enhance fraud response mechanisms.

**Conclusion**

This analysis provides valuable insights into credit card transactions, focusing on fraud detection, spending behavior, and merchant performance. By leveraging SQL Server and Power BI, businesses can strengthen fraud prevention measures, identify high-risk transactions, and enhance financial security.