**TRANSACTION FRAUD MONITORING DASHBOARD REPORT**

**Dataset:** Simulated Credit Card Transactions generated using Sparkov (Kaggle)

* **Rows:** 555,719
* **Columns:** 23
* **Description:** Synthetic dataset highlighting fraud patterns across days, months, and years. It examines fraud trends by category, city, state, and customer behavior by job, age group, and transaction amount.

**1. Data Cleaning & Preparation**

**SQL Data Cleaning:**

* **Primary Key:** id column used as index.
* **Datatype Conversion:**
  + id → int
  + amt → decimal
  + gender → nchar(2)
  + dob → date (renamed to birth\_date)
* **Standardization:**
  + Columns merchant, job, gender converted to lowercase and trimmed.
* **Null Handling:**
  + lat, long, merch\_lat, merch\_long contain nulls; not replaced to preserve valid 0.0 coordinates.
* **Dropped Columns:** unix\_time (represents 2013, synthetic mismatch).
* **New Columns:** full\_name (nvarchar(70)) added.
* **Outlier/Value Updates:**
  + Replace land, make, sub → unknown
  + Corrected barista → barrister

**Power BI Data Preparation:**

* age column calculated in years.
* age\_group column:
  + 18–30 → Young Adults
  + 31–50 → Middle Age
  + 51–110 → Senior
* amt\_group column:
  + 1–8000 → Low
  + 8001–15000 → Medium
  + 15001–23000 → High

**2. SQL Query Insights (Not Covered in Dashboard)**

**Q1. Top 5 Cities by Fraud Count**

* **Camden:** 27 fraud cases
* **Birmingham:** 25 fraud cases
* Other cities range: 2–19

**Q2. Fraud Rate by Day of Week & Month**

* No consistent pattern; fraud occurs across all weeks and months. Peaks vary, indicating year-round distribution.

**Q3. City vs Population, Transactions & Fraud Rate**

* Higher population or more transactions do not imply higher fraud:
  + Burrton: 503 transactions, 19 fraud cases, population 1,689
  + Clark Mills: 1,474 transactions, 18 fraud cases, population 606
  + Phoenix: 2,222 transactions, 15 fraud cases, population 1,312,922
  + Houston: 1,679 transactions, 0 fraud cases, population 2,906,700

**3. Power BI Dashboard Insights**

**Page 1 – Overview / Summary**

| **Insight** | **Value / Observation** |
| --- | --- |
| Total Transactions | 0.56M |
| Fraud Count | 2,145 in 555,719 transactions |
| Fraud Rate | <1% of total transactions |
| Total Transaction Amount | $38.56M |
| Unique Credit Card Holders | 924 customers |
| Fraud by Date | Highest fraud on **4th October** with 40 cases. Fraud varies daily; higher from **July to November** |
| Fraud by State | **New York:** 175 cases; fraud ranges from 0–175 |
| Fraud by Category | **shopping\_net:** 506 cases (online payments)  **grocery\_pos:** 485 cases (POS cash payments) |
| Fraud by Gender | Almost evenly distributed; females slightly higher |

**Page 2 – Customer Insights**

| **Insight** | **Observation** |
| --- | --- |
| Fraud by Age Group & Gender | **Senior females:** 630 cases  **Senior males:** 533 cases  Seniors more involved than young adults |
| Fraud by Customer | **Many Williams:** 19 cases  **Gina Grimes:** 18 cases |
| Fraud by Category & Gender | Females → grocery\_pos  Males → shopping\_net |
| Fraud by Job | Science writers & unknown jobs highest |
| Fraud by Age Group | Middle Age: 819  Senior: 1,163  Young Adults: 163 |
| Fraud by Amount Group | All 2,145 fraud cases in **Low** amount group; indicates small-amount product frauds |

**Slicers:** Date, Amount Group, Age Group, Gender

**Page 3 – Geographical Distribution**

| **Insight** | **Observation** |
| --- | --- |
| Fraud by Business/Store | Highest fraud in 3 businesses:  - Lemke-Gutmann  - Mosciski, Ziemann and Farrell  - Romaguera, Cruickshank and Greenholt  Each has multiple branches |
| Bubble Map | Fraud by Business/Store visualized by branch locations |
| Fraud by State | NY: 175 cases  PA: 114 cases |
| State Map | Visualizes overall state-level fraud distribution |

**Page 4 – Transaction Insights & Trends**

| **Insight** | **Observation** |
| --- | --- |
| Fraud by Day | Day 25 & Days 4–8 highest fraud |
| Fraud by Month | August: 415 (highest)  October: 384 (second highest)  Monthly trend smooth & stable |
| Transactions by Amount Group | Low: 0.56M transactions  Medium & High: very few transactions |
| Business Matrix | Shows total transactions vs fraud per business; highlights high-risk stores |

**4. Limitations**

* unix\_time and job columns contain inconsistent or incorrectly formatted values.
* Many job titles have duplicates with different wording.
* Dataset does not specify currency type.
* Demographic info may be inaccurate or incomplete.

**5. Recommendations**

* Ensure **data accuracy & consistency** during entry and collection.
* Align transaction timestamps (unix\_time vs trans\_date\_trans\_time) to avoid errors.
* Standardize job column values for accurate analysis.
* Implement targeted fraud monitoring for **low-amount transactions**, senior customers, and high-risk states/stores.

✅ **Summary:**

* This dashboard tells a story from **overview → customer insights → geographic distribution → transaction trends**.
* SQL queries supplement dashboard insights, providing detailed views of cities, day-of-week patterns, and population vs fraud correlations.
* The dashboard uses **interactivity (slicers)** to analyze fraud by date, gender, age group, amount group, and state.