“Banking Fraud Detection”

Batch Number: DA | BN001

Name: ALUGU PRANUSHA

Project Title: Fraud Detection in Banking Using Power BI

ENTRI
elevate

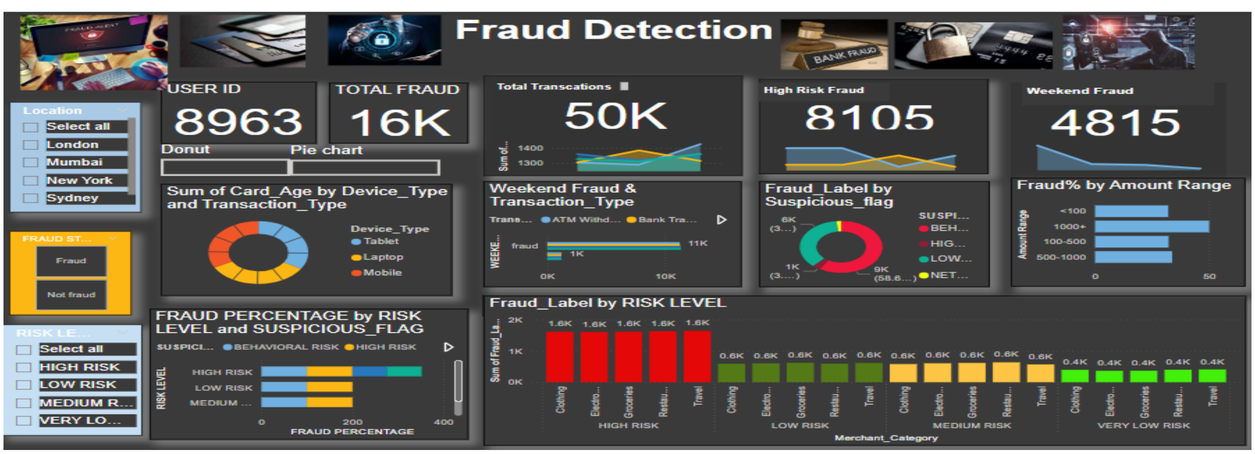
TELUGU REGION

Banking Fraud Detection

Contents

Objectives:

"This project aims to analyze transaction data to identify key indicators of fraudulent behavior using Excel, Power Query, and Power BI. The goal is to support the banking industry in detecting high-risk patterns, understanding user behaviors, and building data-driven fraud prevention strategies."



A. Pranusha

Data Collection: (Kaggle)

For this mini-project on **Fraud Detection in Banking**, a **synthetic dataset** with approximately **50,000 transaction records** was used. The dataset was designed to simulate real-world banking scenarios and includes detailed attributes such as:

- Transaction ID, User ID, Transaction Amount, Transaction Type, Timestamp
- Device Type, Location, Merchant Category, Card Type, Risk Score
- Failed_Transaction_Count_7d, Previous fraudulent Activity, Fraud Label, and more.

The dataset was downloaded from a public repository (Kaggle) titled **"Synthetic Financial Datasets for Fraud Detection"**, which is commonly used for analytical practice.

The data was provided in **CSV format** and imported using:

- Microsoft Excel – for initial review, sorting, and column checking
- Extracted Date Parts with timestamping**:** (Using Text function)

  Date: TEXT(E2,"DD-MM-YYYY")

  Time: =TEXT(E2,"HH:MM")

  Day of week: =TEXT(E2,"DDDD")

  Month: =TEXT(E2,"MMMM")

  Year: =TEXT(E2,"YYYY")

  And VLOOKUP value based on card type and Authentication_ Method

  =VLOOKUP([@[Card_Type]],fraud[[#All],[Card_Type]:[VLOOKUP]],4,FALSE)

- *Pivot Table Analysis*:

  Created Pivot Tables to analyze fraud patterns:
  1. Fraud Count by Day of Week
  2. Fraud Count by Month
  3. Avg Transaction Amount by Day
  4. Count of High-Risk Score by Month

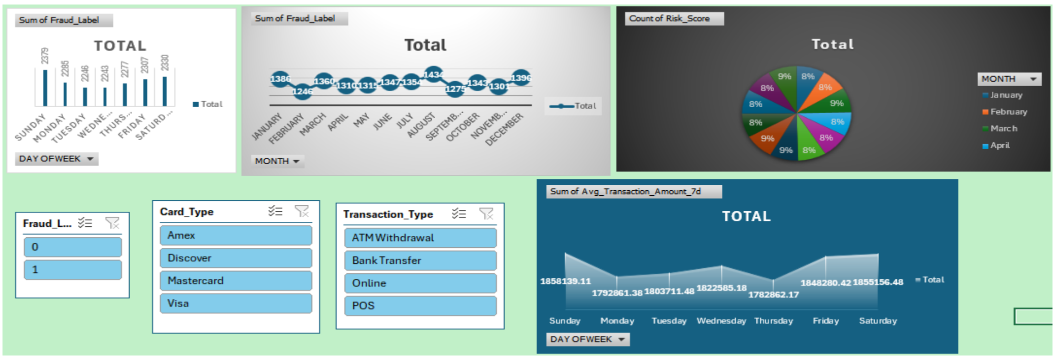
Row Labels	Sum of Fraud_Label
Sunday	2379
Monday	2285
Tuesday	2246
Wednesday	2243
Thursday	2277
Friday	2307
Saturday	2330
Grand Total	**16067**

Row La	Sum of Fraud_Label
January	1386
February	1246
March	1360
April	1310
May	1315
June	1347
July	1354
August	1434
Septembe	1275
October	1343
Novembei	1301
December	1396
Grand Tot	**16067**

Row La	Sum of Avg_Transaction_Amount_7d
Sunday	1858139.11
Monday	1792861.38
Tuesday	1803711.48
Wednesdı	1822585.18
Thursday	1782862.17
Friday	1848280.42
Saturday	1855156.48
Grand Tot	**12763596.22**

Row La	Count of Risk_Score
January	4188
February	3903
March	4259
April	4106
May	4162
June	4160
July	4213
August	4384
Septembe	4087
October	4164
Novembei	4085
December	4289
Grand Tot	**50000**

- Charts in Excel**:**



- Power BI (Power Query) – for data cleaning, transformation, column creation, and model structuring

This dataset serves as the foundation for fraud analysis using **descriptive and diagnostic analytics** techniques—focusing on identifying high-risk transactions, frequent failures, and fraudulent behavior patterns through visuals and calculated metrics.

This analysis is entirely driven by **DAX measures, calculate, and interactive dashboards** using **Excel and Power BI tools**.

Data Preparation:

Open Power BI → Import dataset → Go to Power Query Editor inside Power BI

Power Query in Power BI

After importing the dataset into Power BI, the **Power Query Editor** was used for structured and efficient data cleaning and transformation.

Handling Missing Values

- Checked for nulls in key columns such as Transaction Amount, Risk Score, and Authentication method.
- Replaced missing values:

- Numerical fields → filled with **median** or **0** depending on context.

  (For numerical columns, use =IF(ISBLANK(A2), MEDIAN (A: A), A2).

- Categorical fields → filled with **“Unknown”**.

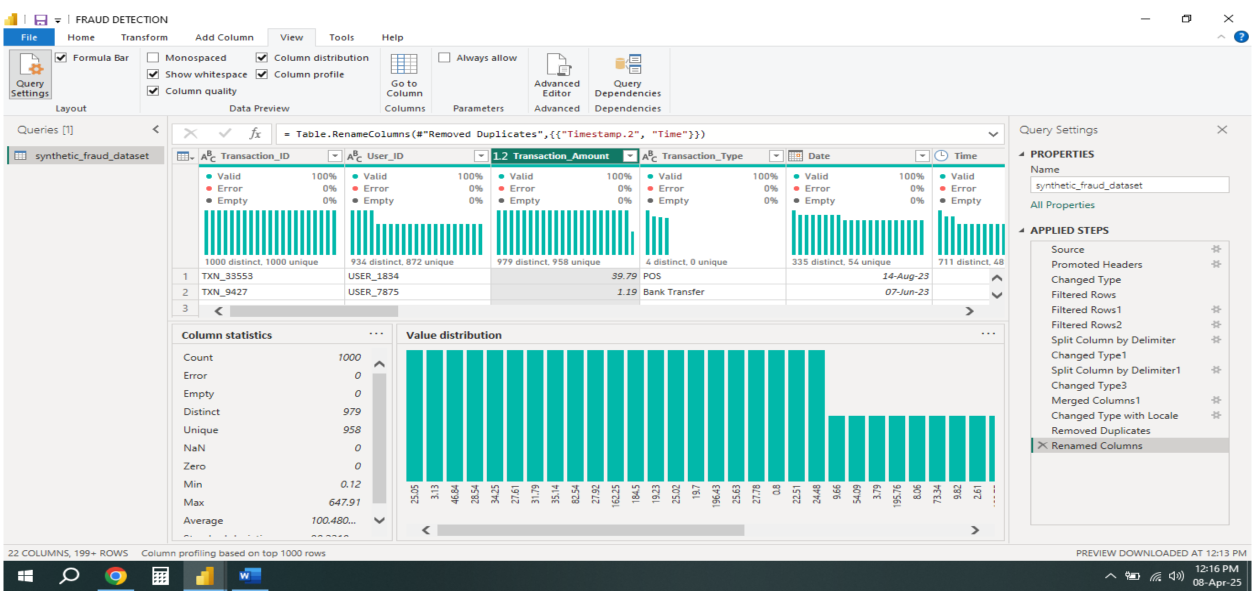
  (use =IF(ISBLANK(A2), "Unknown", A2).

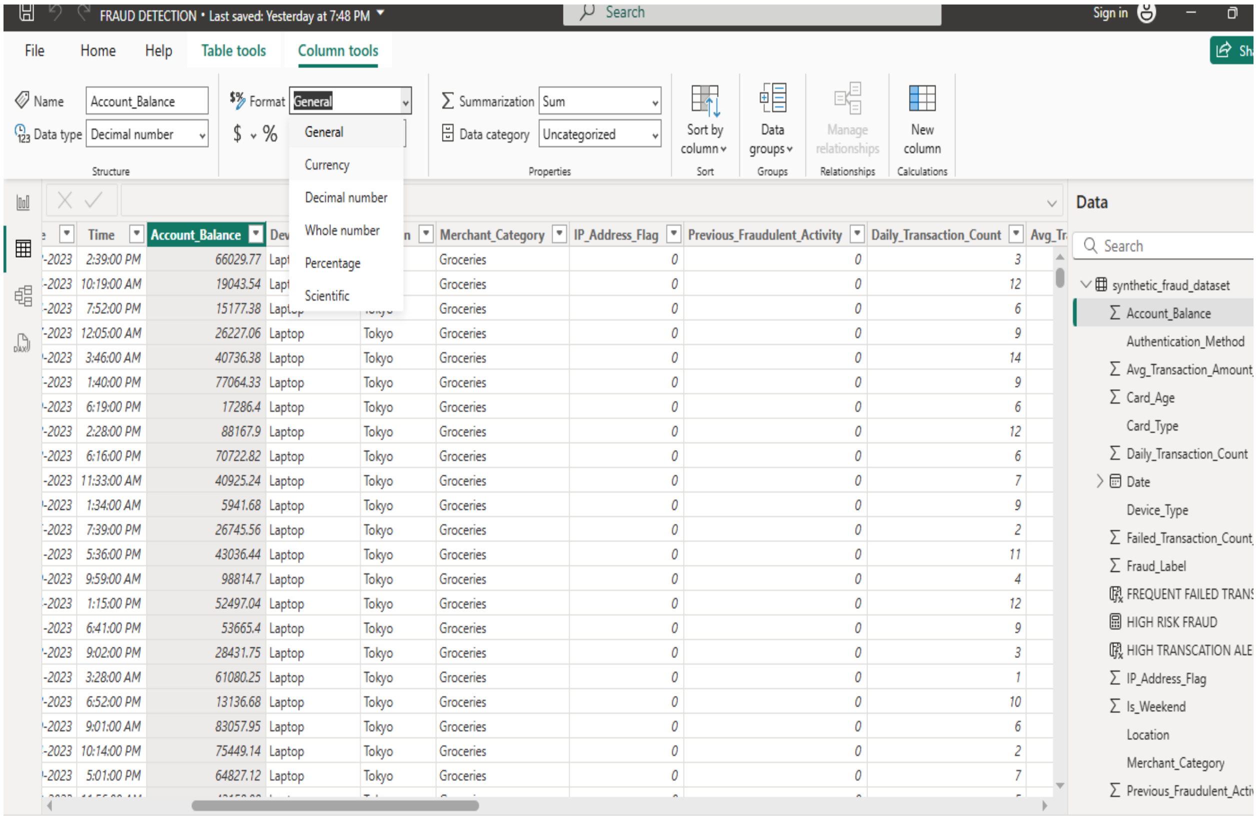
Data Type Conversion

- Converted:
  - Timestamp to **Date/Time**
  - Risks Score to **Decimal Number**
  - Transaction amount and Account Balance to **Currency**

Remove Irrelevant Columns

- Removed columns like Merchant Category, IP Address Flag (optional), if not helpful for fraud detection logic. [But I did not removed I want to do based on all columns]





Here are some key analytical questions based on your dataset:

1.What types of transactions are most frequently associated with fraud?

2.Are there patterns in fraud based on time (e.g., weekends, night transactions)?

3.Which users or accounts have the highest number of fraudulent activities?

4.Is there a relationship between transaction amount and fraud likelihood?

5.How do device type, location, and authentication method influence fraud risk?

6.What role does a highrisk score or failed transaction history play in identifying fraud?

Detect patterns and behaviors that indicate potential fraud.

- Segment transactions and users into risk categories (High, Medium, Low).
- Visualize fraud distribution across multiple factors like transaction type, location, and time.
- Identify high-risk users for closer monitoring.
- Provide actionable insights to support fraud prevention strategies.

DAX columns: (Creating Calculated Columns Fraud Flags)

Calculated Columns.

---

1. High Transaction Alert (Flag if amount > 5)

DAX Calculated Column:

High Transaction Alert =

IF (' synthetic fraud dataset '[Transaction Amount] > 5, "High", "Normal")

---

2. Frequent Failed Transactions Alert (Flag if more than 3 failed transcations in 7 days)

DAX Column:

Frequent Failed Transactions =

IF (synthetic fraud dataset [Failed_Transaction_Count_7d] > 3, "Suspicious", "Okay")

---

3. Risk Score Label (Categorize Risk Score as Risk Level)

DAX Column:

RISK LEVEL = SWITCH (TRUE (), synthetic fraud dataset [Risk Score]>0.8,"HIGH RISK", synthetic fraud dataset[Risk Score]>0.5,"MEDIUM RISK", synthetic fraud dataset[Risk Score]>0.2,"LOW RISK","VERY LOW RISK")

4.Weekend Fraud Flag(Based on Is weekend & Fraud Label)

If you want to analyze behavior on weekends:

DAX Column:

WEEKEND FRAUD = IF(synthetic fraud dataset[Is Weekend]=1 && synthetic fraud dataset[Fraud Label]=1,"weekend fraud" ,"non fraud")

Explanation:

- Is Weekend = 1 → means Yes, it's weekend

- Fraud Label = 1 → means It's a fraud
- So together, this marks transactions that are both weekend and fraud.

---

5. WEEKEND_VS_WEEKDAY(Based on Fraud Label & IS weekend)

	Fraud	Is weekend
Weekend fraud →	1	1
Weekday fraud→	1	0
Not fraud →	0	0 or 1

Fraud label(1=Fraud, 0=Not fraud)

Is-weekend(1=weekend, 0= weekday)

DAX Column:

```
WEEKEND_VS_WEEKDAY =
SWITCH(
   TRUE(),
   'synthetic fraud dataset'[Fraud Label] = 1 && 'synthetic fraud dataset'[Is Weekend] = 1,
"Weekend Fraud"
   'synthetic fraud dataset'[Fraud Label] = 1 && 'synthetic fraud dataset'[Is Weekend] = 0,
"Weekday Fraud",
   "Not Fraud"
)
```

---

6. **SUSPICIOUS_FLAG Report:**

According to by data to calculate separate columns on both IP address flag and Failed Transcation-count-7d I have used switch DAX function to get Suspicious-falg

Example: A hacker could succed in the first try. But use a suspicious IP, you'd miss it if you only looked at failed attempts!

DAX Column:

SUSPICIOUS_FLAG: (Based on Failed transcation count & IP address flag)

SUSPICIOUS_FLAG =
SWITCH(TRUE(),synthetic_fraud_dataset[Failed_Transaction_Count_7d]>3&&synthetic_fraud_dataset[IP_Address_Flag]=1,"HIGH RISK",synthetic_fraud_dataset[Failed_Transaction_Count_7d]>3,"BEHAVIORAL RISK", synthetic fraud dataset[IP Address Flag]=1,"NETWORK RISK","LOW RISK")

Create DAX Measures:

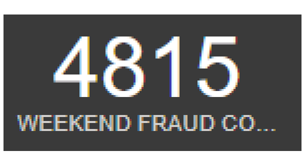
1:Total Fraud Count: (Based on Fraud Label)

TOTAL FRAUD COUNT =
CALCULATE(COUNTROWS(synthetic_fraud_dataset),synthetic_fraud_dataset[Fraud_Label]=1)

2.High Risk Fraud: (Based on risk level & fraud label)

LHIGH RISK FRAUD =
CALCULATE(COUNTROWS(synthetic_fraud_dataset),synthetic_fraud_dataset[RISK LEVEL]="HIGH RISK"&&synthetic_fraud dataset[Fraud_Label]=1)

3. WEEKEND FRAUD COUNT: (Based on Weekend vs Weekday) WEEKEND FRAUD COUNT =
CALCULATE(COUNTROWS(synthetic_fraud_dataset),synthetic_fraud_dataset[WEEKEND_VS_WEEKDAY]="WEEKEND FRAUD")



Here are some key analytical questions based on your dataset:

1.Types of transactions are most frequently associated with fraud?

we can analyze the relationship between Transaction Type and Fraud Label to find the most risky transaction types.

2. **Are there patterns in fraud based on time (e.g., weekends, night transactions)?**

A. Findings from week-Based Analysis:

1. Weekend vs Weekday Fraud
   - Fraud incidents are more frequent during weekends or weekday, when manual monitoring may be lower.
2. Weekend fraud count
   Based on weekend vs weekday fraud analysis
3. Fraud Status.

Conclusion: Fraudulent **activity tends to increase during weekends and outside regular business** hours. This indicates the need for stronger monitoring during off peak times.

3. Which users or accounts have the highest number of fraudulent activities?

A. Used the User ID column along with the Fraud Label to identify users with repeated frauds.

By finding total fraud count then calculate new table to for TOP5 Fraudulent users by using DAX functions

Top5 fraudulent users = TOPN(5,SUMMARIZE(synthetic_fraud_dataset,synthetic_fraud_dataset[User_ID],"Fraud count",[TOTAL FRAUD COUNT]),[TOTAL FRAUD COUNT],DESC)

“Top 5 users were involved in multiple fraudulent transactions, suggesting either compromised accounts or high-risk behavior. Continuous monitoring and alert mechanisms should be prioritized for these accounts.”

4. Is there a relationship between transaction amount and fraud likelihood?

A. To understand whether high or low transaction amounts are more likely to be fraudulent.

Step 1: Create a new calculated column to bin transaction amounts

Amount Range = SWITCH(TRUE(),synthetic_fraud_dataset[Transaction_Amount]<100,"<100"

,synthetic_fraud_dataset[Transaction_Amount]<500,"100-500",synthetic_fraud_dataset[Transaction_Amount]<1000,"500-1000","1000+")

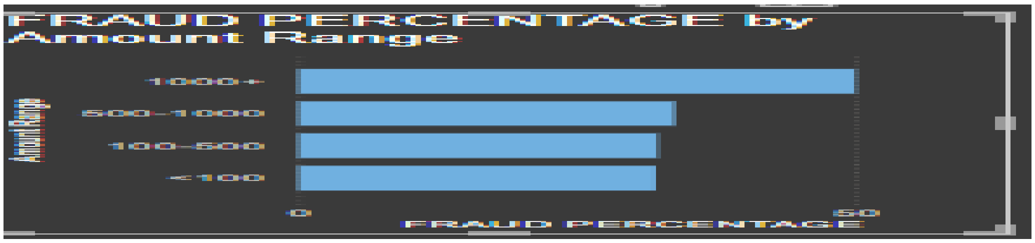
Step 2: Create a Fraud Percentage Measure

FRAUD PERCENTAGE = DIVIDE([TOTAL FRAUD COUNT],[TOTAL TRANSCATIONS],0)*100

Bar Chart:

- Amount Range
- Fraud Percentage

Now you'll see which amount ranges have higher fraud rates.



You can conclude whether higher transactions have more fraud or if fraudsters target smaller values to avoid detection.

5. How do device type, location, and authentication method influence fraud risk?

Goal:

Understand which devices, locations, or authentication methods are more prone to fraud, and uncover hidden risk patterns.

Step 1: Use Existing Columns

Ensure these columns are cleaned and available:

- Device Type
- Location
- Authentication Method
- Fraud Label (0/1)

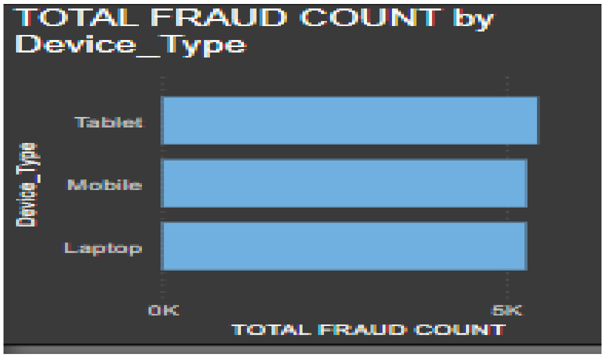
Step 2: Use already Create Measures

Total Fraud count

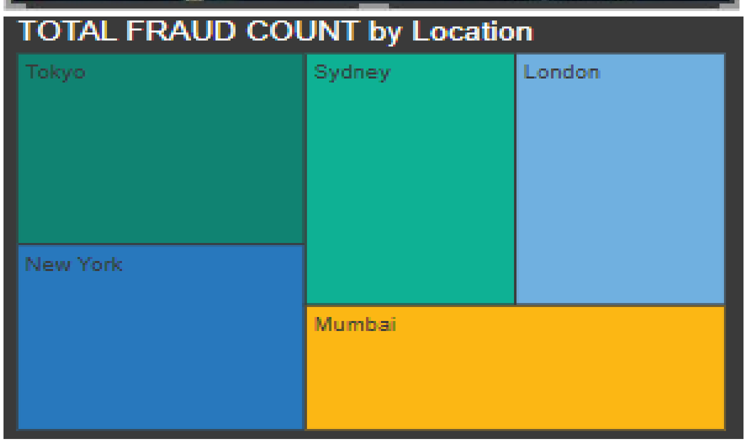
Fraud Percentage

Step 3: Create Visuals

1. Device Type vs Fraud
   Visual: Bar Chart
   Axis: Device type
   Values: Total fraud count

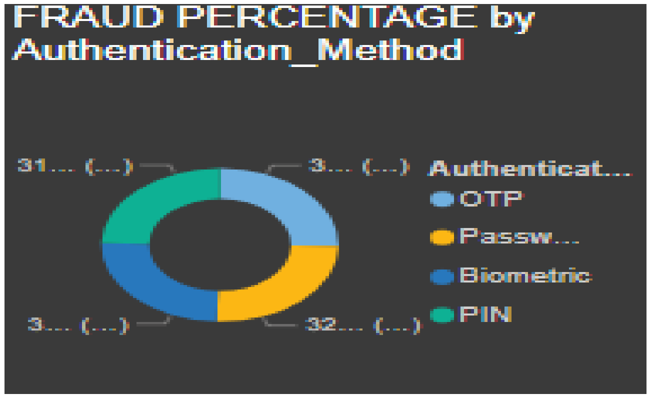


2. Location vs Fraud:
   Visuals: Map and Tree
   Location: Location column
   Values: Fraud count



A. Pranusha

3. Authentication Method vs Fraud:
   Visual: Donut
   Axis: Authentication Method
   Values: Fraud percentage



"Mobile devices and public Wi-Fi locations showed a 2x higher fraud rate than desktop usage at home, indicating a correlation between device & location with fraud."

6.What role does a high risk score or failed transaction history play in identifying fraud?

1. High Risk Score

The Risk Score is typically a numeric value (e.g., between 0 and 1) that estimates the likelihood of a transaction being fraudulent.

Insight:

"Higher risk scores are strongly correlated with fraud, and most fraudulent transactions in our dataset had scores above 0.8."

2. Failed Transaction History

This refers to the number of failed attempts made by a user in a short time frame (e.g., 7 days).

- **Role:**
- A user with frequent failed transactions may be:
  - Attempting to guess credentials
  - Facing authentication issues
  - Using stolen cards or credentials
- Repeated failed attempts are strong indicators of potential fraud attempts.

➢ **Insight:**

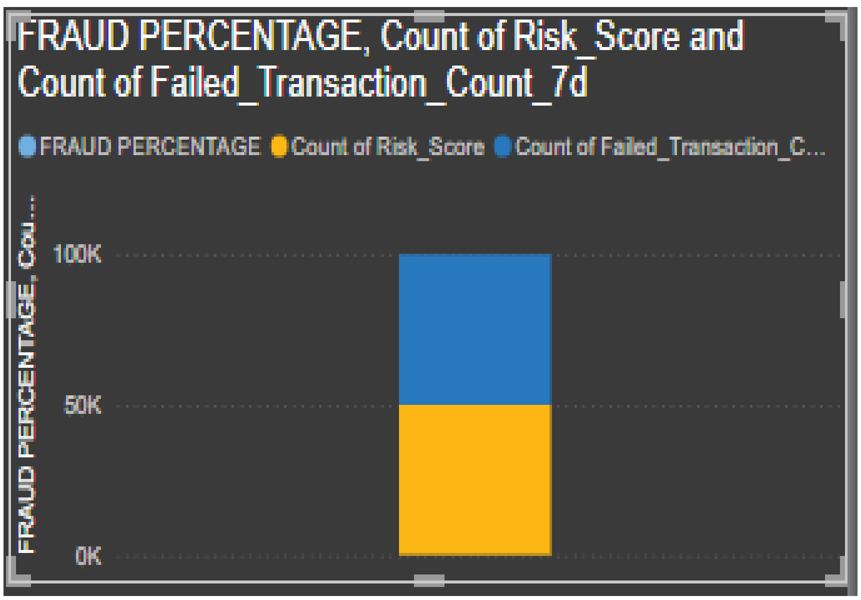
"Users with more than 3 failed transactions in a 7-day window were 4x more likely to be flagged as fraudulent."

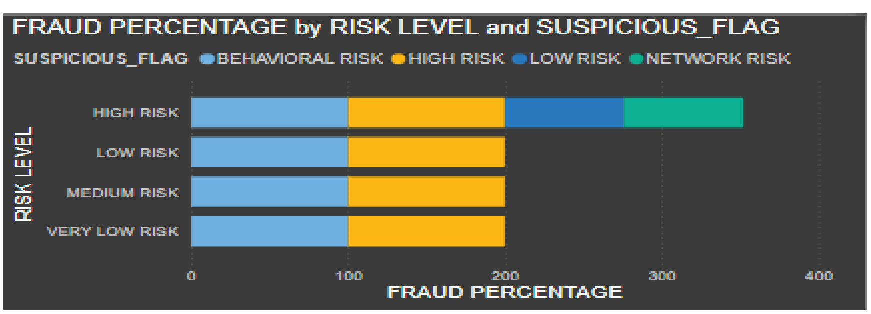
DAX column:

FREQUENT FAILED TRANSCATION = IF(synthetic_fraud_dataset[Failed_Transaction_Count_7d]>3,"SUSPICIOUS","OKAY")

Use in visualizations:

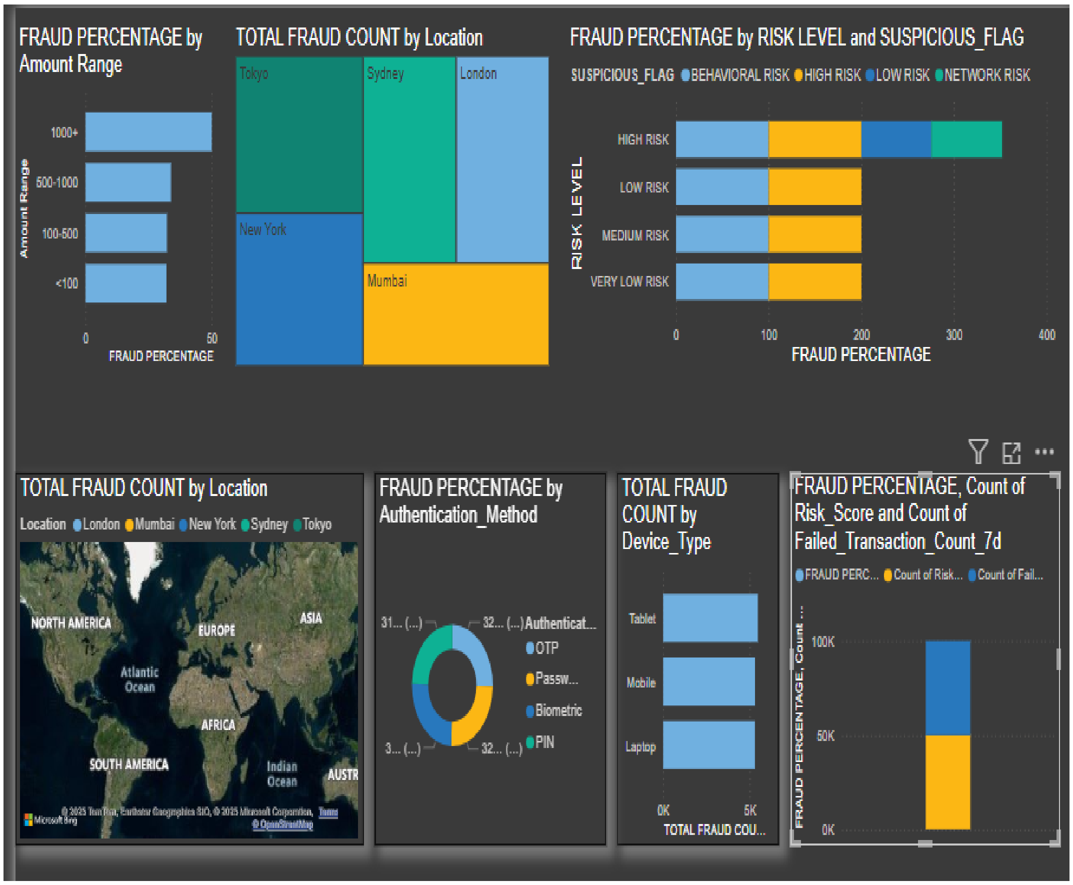
- Fraud % by Risk Score (Line/Scatter Chart)
- Fraud % by Risk Level and Suspicious Falg(Bar Chart)
- Combine both in Stacked Columns or Matrix





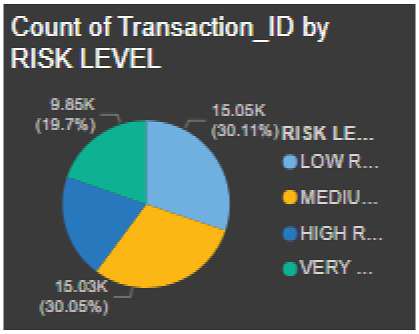
Bar Chart:

High risk scores and frequent failed transactions are strong behavioral indicators of potential fraud. Monitoring these factors can significantly improve fraud detection and help prioritize cases for review.



Create the Visual:

1. Go to **Report View** in Power BI.
2. From the **Visualizations pane**, click on **Donut chart** or **Pie chart**.
3. Drag and drop:
    - **Legend**: Risk_Level
    - **Values**: Transaction ID (or any column with unique transactions; it will count them automatically)



Now you'll see the distribution of risk levels!

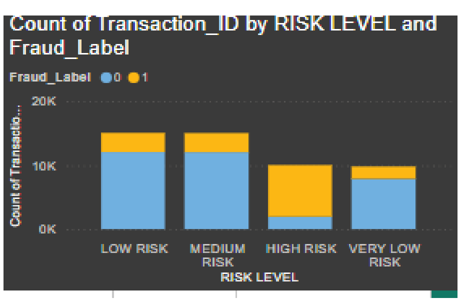
Use Slicers

- Add a **Slicer** visual.
- Drag in Risk_Level.
- Now users can filter the entire report based on Risk Category.

Combine with Fraud Label:

Create a **stacked column chart** to compare how many transactions in each Risk_Level are actually labeled fraud:

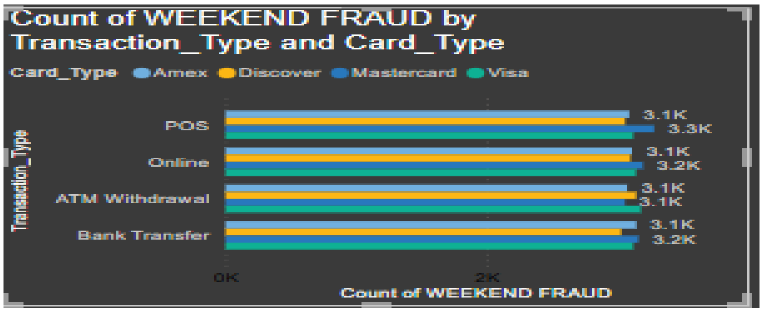
1. Axis: Risk_Level
2. Value: Count of Transaction ID
3. Legend: Fraud Label (0 = Not Fraud, 1 = Fraud)



This shows how frauds correlate with different risk categories.

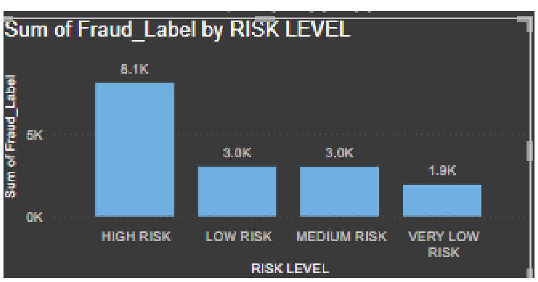
Weekend vs Weekday Fraud Count (Bar Chart):

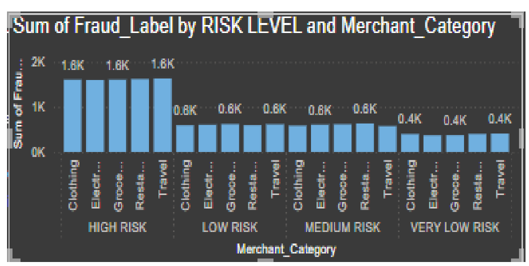
- Go to **Visualizations pane** → click on **Clustered Bar Chart**
- On the **Fields pane**, drag:
  - Weekend Fraud → **Axis**
  - Transaction ID (or any unique field like User ID) → **Values** → it auto-aggregates as **Count**
  - *(Optional)* Card Type or (Location) → **Legend** (to break it down by category)



1.Stacked Column Chart:

:(Risk vs fraud)

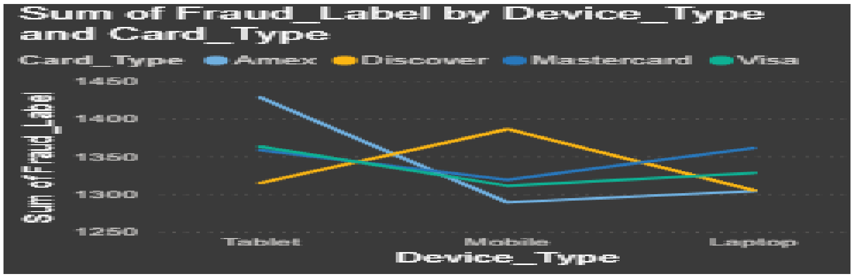




x-axis→Risk level
→Merchant Category
y-axis→Sum of Fraud label
we can do Drill up an dDrill Down and also hieracy

2.Line chart:

x-axis: Device type
y-axis: sum of fraud label
Legend: Card type



3.Table:

Detailed data table of: transaction ID, Amount, Risk_Level, Fraud Label

Transaction_ID	Sum of Account_Balance	RISK LEVEL	Sum of Fraud_Label
TXN_0	$93,915.02	LOW RISK	0
TXN_1	$91,495.28	MEDIUM RISK	0
TXN_10	$48,484.2	MEDIUM RISK	0
TXN_100	$89,219.38	VERY LOW RISK	0
TXN_1000	$40,090.04	VERY LOW RISK	0
TXN_10000	$30,837.25	HIGH RISK	0
TXN_10001	$18,273.83	MEDIUM RISK	1
TXN_10002	$4,380.72	HIGH RISK	1
TXN_10003	$87,541.23	VERY LOW RISK	0
TXN_10004	$50,643.87	VERY LOW RISK	1
TXN_10005	$7,095.38	LOW RISK	0
TXN_10006	$34,200.02	VERY LOW RISK	0
TXN_10007	$43,966.37	VERY LOW RISK	1
Total	$2,514,703,299.039994		16067

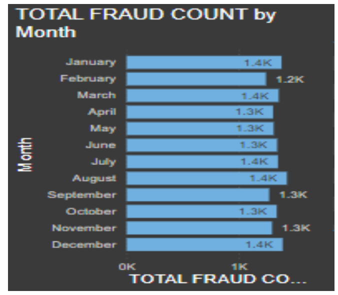
Add slicers for:

- Location
- Merchant Category
- Authentication Method

So we can drill down and interact with the visuals.

4.Clustered Bar Chart – Fraud Count by Month:

Clustered Bar chart:



y Axis: Month

x axis: Total fraud Count

Insight: Spot fraud spikes month

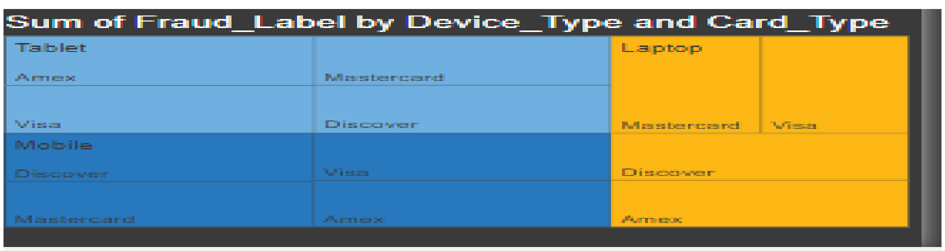
5.Tree Map:

Category: Device type

: Merchant category

Details: card-type

Values: sum of fraud label

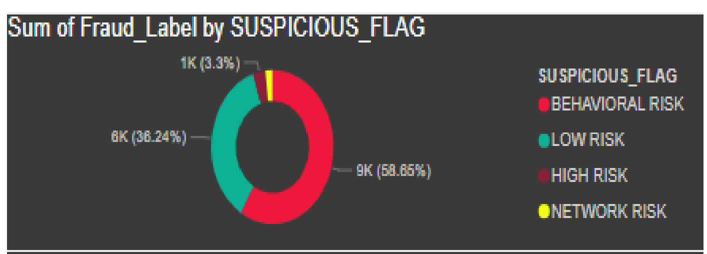


6.Used Donut chart:

Legend: SUSPICIOUS FLAG

Values: sum of fraud label

In Visualization pane→ Format your visuals→ Slices→colors→change color as Red→High Risk



Light Red→Behavioral Risk

Yellow→Network Risk

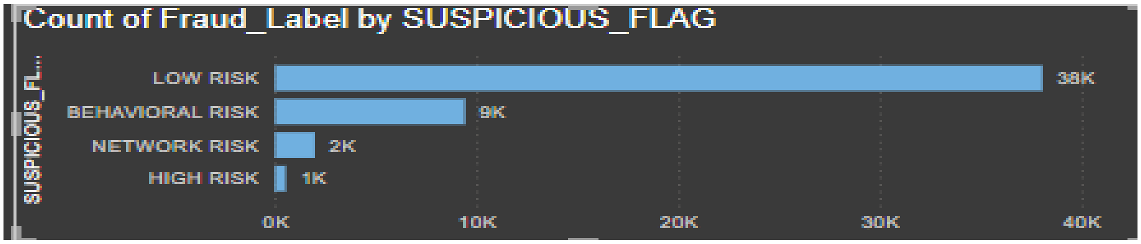
Green→Low Risk

Insight:

A. Pranusha

To show which type of risk is most common in fraud

7.Stacked Bar Chart:

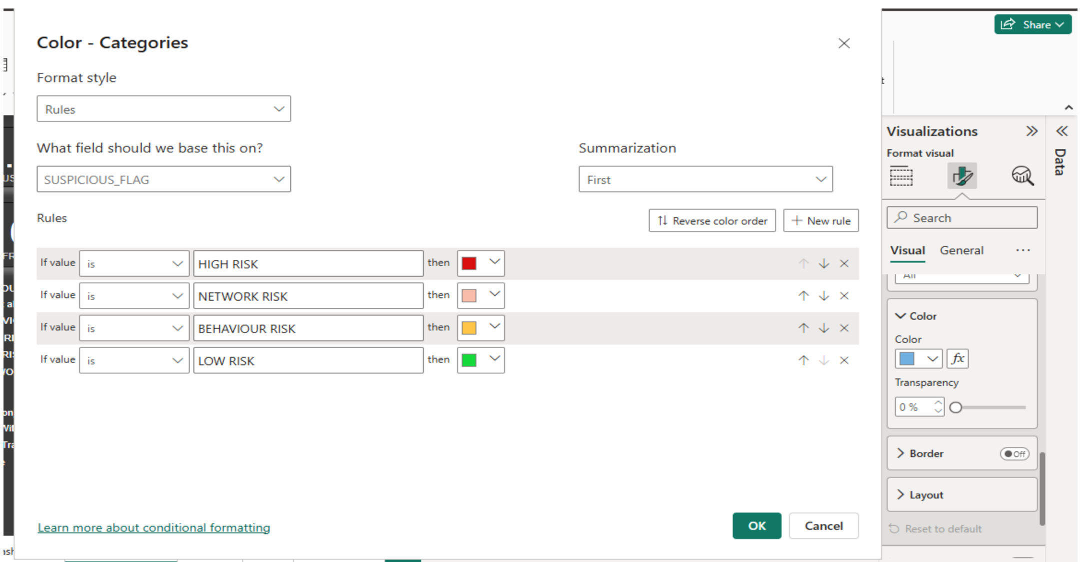


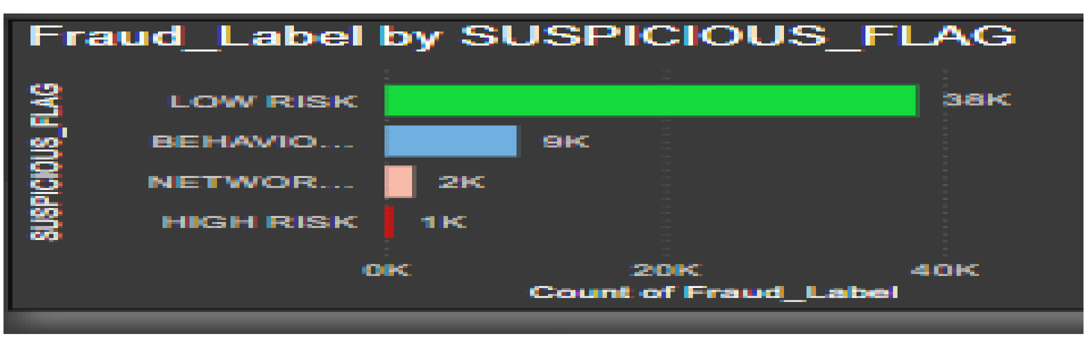
: Suspicious Flag,

Values: Count of Fraud Label = 1

Insight: Compare how many frauds fall under each risk type

Used conditional formatting for colors to identify easily risk level





A. Pranusha

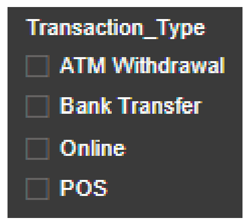
8.Matrix Table:

Card_Type	BEHAVIORAL RISK	HIGH RISK	LOW RISK	NETWORK RISK	Total
Amex	2320	124	9501	474	12419
Discover	2363	145	9344	476	12328
Mastercard	2353	120	9719	501	12693
Visa	2387	142	9503	528	12560
Total	9423	531	38067	1979	50000

Rows: Card Type,
Columns: Suspicious Flag,
Values: Count of Fraud Label
Insight: Analyze how different card types behave under risk flags

9. Added Slicer for Suspicious Flag:

Want to see only "High Risk" frauds then select that option if not any other option in the slicer according to our need. (Interactive and clear for viewers.)
And



Added Slicer to Transcation type:
ATM Withdrawal
Bank transfer
Online
Pos

1. Card Visuals: count of distinct Risk Score
   Sum of daily transcation count

Fraud Percentage:

Step 1: Create a Measure for Total Transactions

TOTAL TRANSCATIONS = COUNT (synthetic fraud dataset [Transaction_ ID])

Step 2: Create a Measure for Fraud Transactions

TOTAL FRAUD TRANSCATIONS = CALCULATE (COUNTROWS (synthetic_ fraud_ dataset), synthetic fraud dataset[Fraud Label]=1)

Step 3: Create the Fraud Percentage Measure

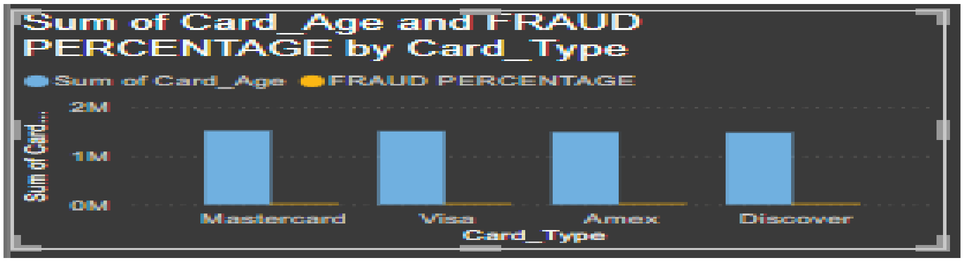
FRAUD PERCENTAGE = DIVIDE ([TOTAL FRAUD TRANSCATIONS], [TOTAL TRANSCATIONS],0) * 100

Fraud % by Card Type:

Clustered Column Chart

Axis (X): Card Type

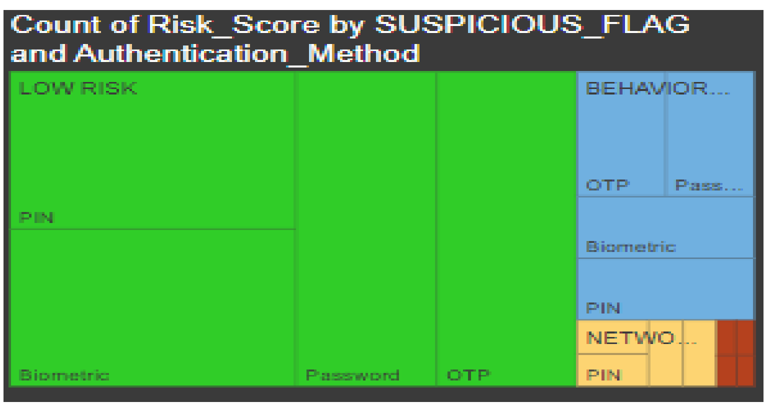
Values (Y): Fraud Percentage By Card Type measure (DAX) and sum of card age



Fraud % by Weekend/Weekday:

Tree Map

- Group by: Risk Score bucket or your created suspicious flag
- Catergory : suspicious flag
- Details: authentication method
- Values: count of risk score

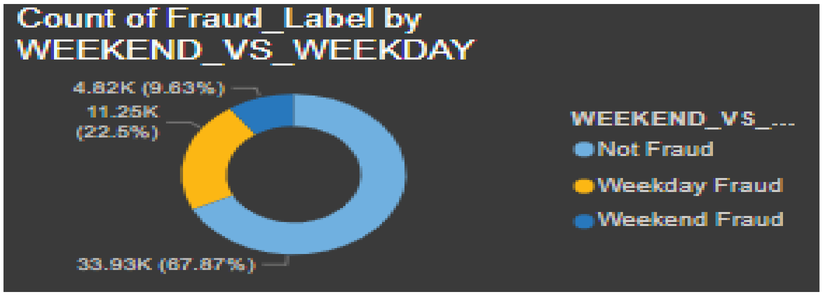


Sum of Fraud by Weekend/Weekday:

Donut Chart

- Legend / Axis: "Weekend vs Weekday" (I already have a column like Weekend vs weekday Fraud)

- Values: Fraud Count

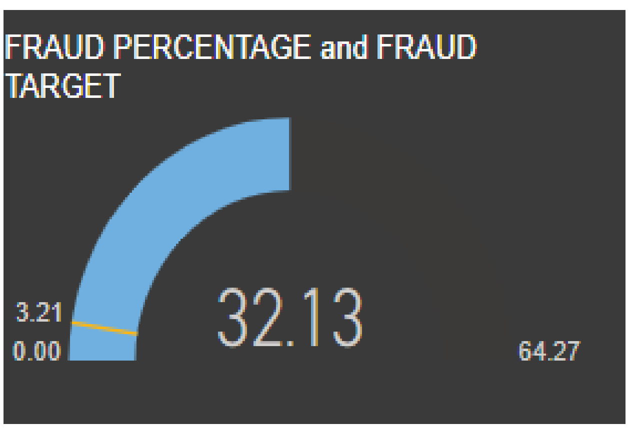


Target Value Measure:

FRAUD TARGET = synthetic fraud dataset [FRAUD PERCENTAGE] *0.10

Insert the Gauge Visual:

1. Go to Visualizations Pane → Click Gauge
2. Drag the fields:
   - Value → Fraud Percentage
   - Target Value → Fraud Target
   - (Optional) Set Min Value = 0, Max Value = 0.10 (or 10% as upper range)

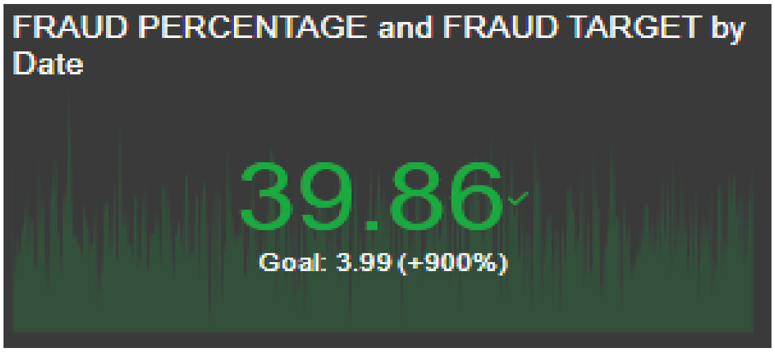


Insert KPI Visual:

Go to Visualizations pane → Select KPI

Drag:

- Indicator → Fraud Percentage
- Target Goal → Fraud Target
- (Optional) Trend Axis → Transaction Date (must be in date format)



Fraud Percentage vs Target:

Step-by-Step Guide to Create Toggle Buttons:

Gauge Visual and KPI Visual on your report page.

Keep both in the same position (overlapping).

Select one visual (Gauge and KPI), then go to **View** → Enable **Selection Pane** and **Bookmarks Pane and also create buttons for that**

Create Bookmarks:

1. **Turn off KPI visual** in **Selection Pane**, only **Gauge visible**.
2. Go to **Bookmarks Pane**, click **Add**, rename to Show Gauge.
3. Now **hide Gauge**, **show KPI**.
4. Add another Bookmark, rename to Show KPI.

Create Buttons

1. Go to **Insert** → **Buttons** → **Blank** (or use built-in icons).
2. Rename one as Gauge Button, the other as KPI Button.
3. Place both on the report.

Assign Bookmarks to Buttons:

1. Select Gauge Button → **Format** → **Action: On**
2. Set **Type: Bookmark** → Select Show Gauge
3. Select KPI Button → **Action: On**
4. Set **Type: Bookmark** → Select Show KPI
5. Now clicking buttons will toggle between visuals!

Slicer:

To add slicer for fraud I have created a new column in order to know Fraud status using DAX function if

FRAUD STATUS = IF (synthetic fraud dataset [Fraud Label] =1,"Fraud","Not fraud")

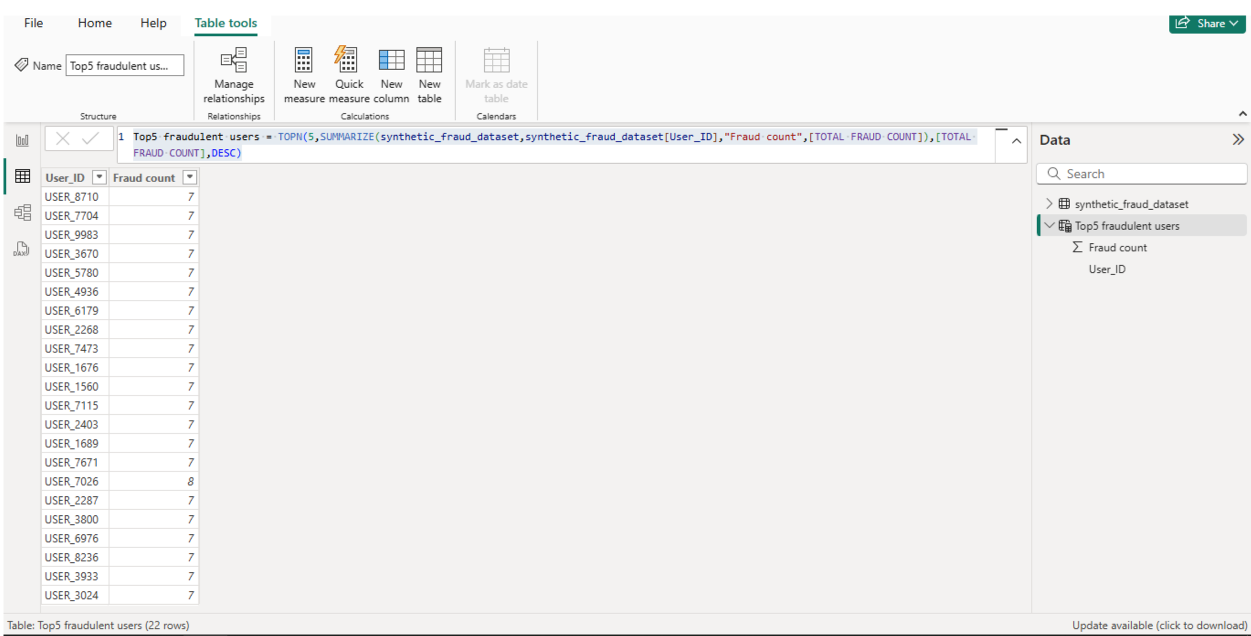
Then added slicer to it to know the status of fraud.

Created a NEW TABLE:

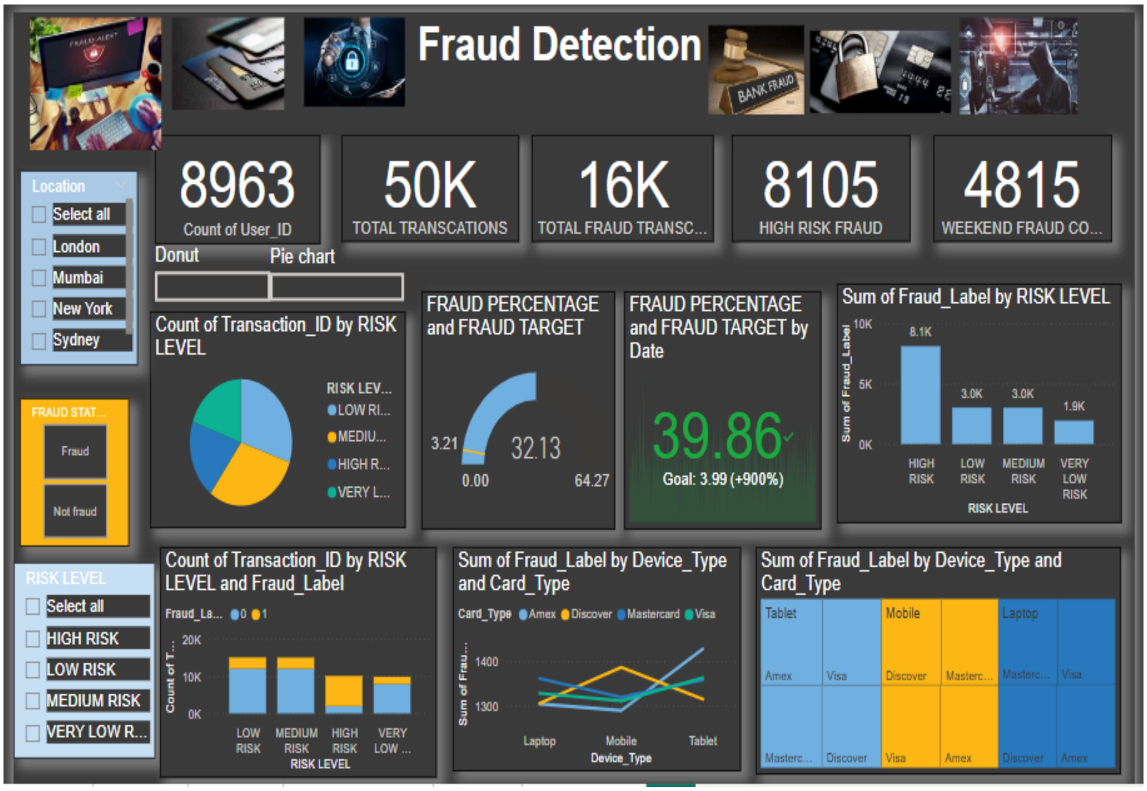
In Modeling → New Table,

Use the User ID column along with the Fraud Label to identify **users** with repeated frauds.

Top5 fraudulent users = TOPN (5, SUMMARIZE (synthetic fraud dataset,synthetic fraud dataset [User ID],"Fraud count",[TOTAL FRAUD COUNT]),[TOTAL FRAUD COUNT],DESC)

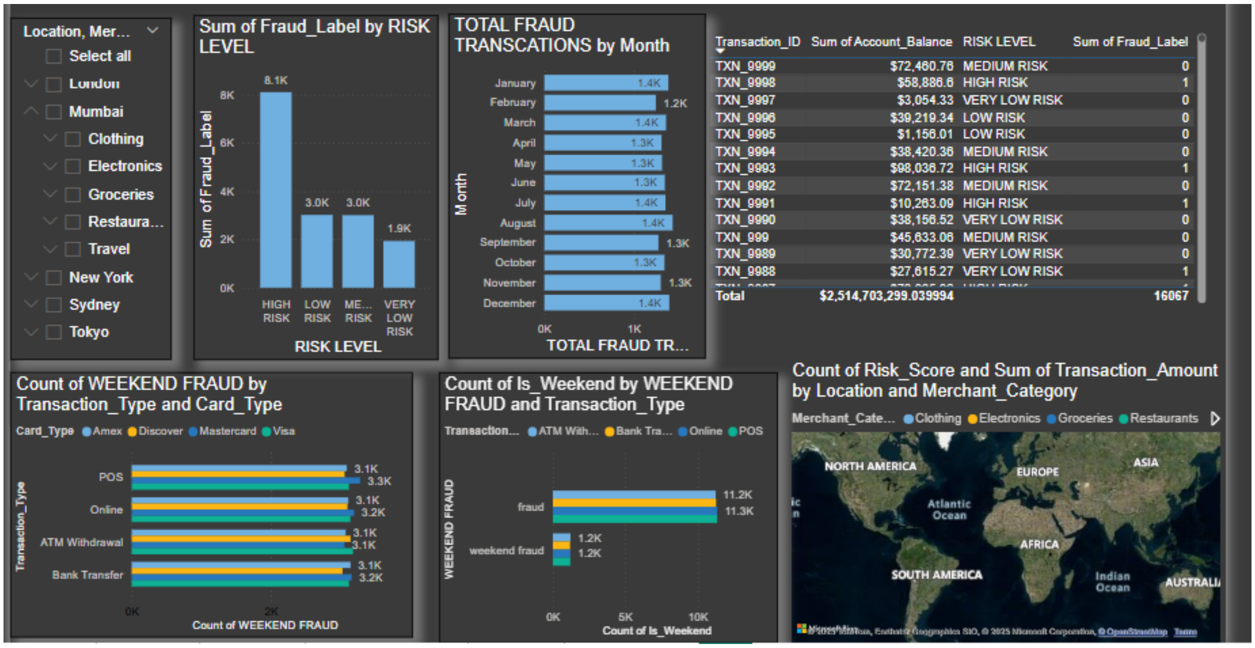


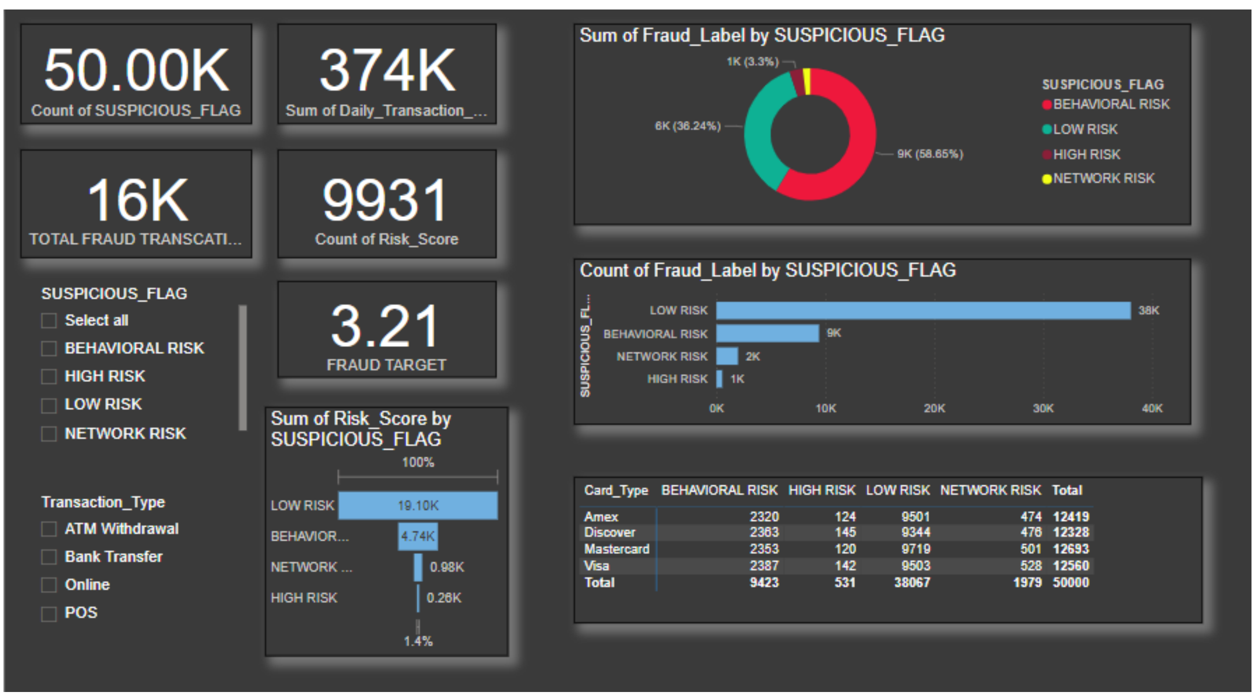
“Top 5 users were involved in multiple fraudulent transactions, suggesting either compromised accounts or high-risk behavior. Continuous monitoring and alert mechanisms should be prioritized for these accounts.”



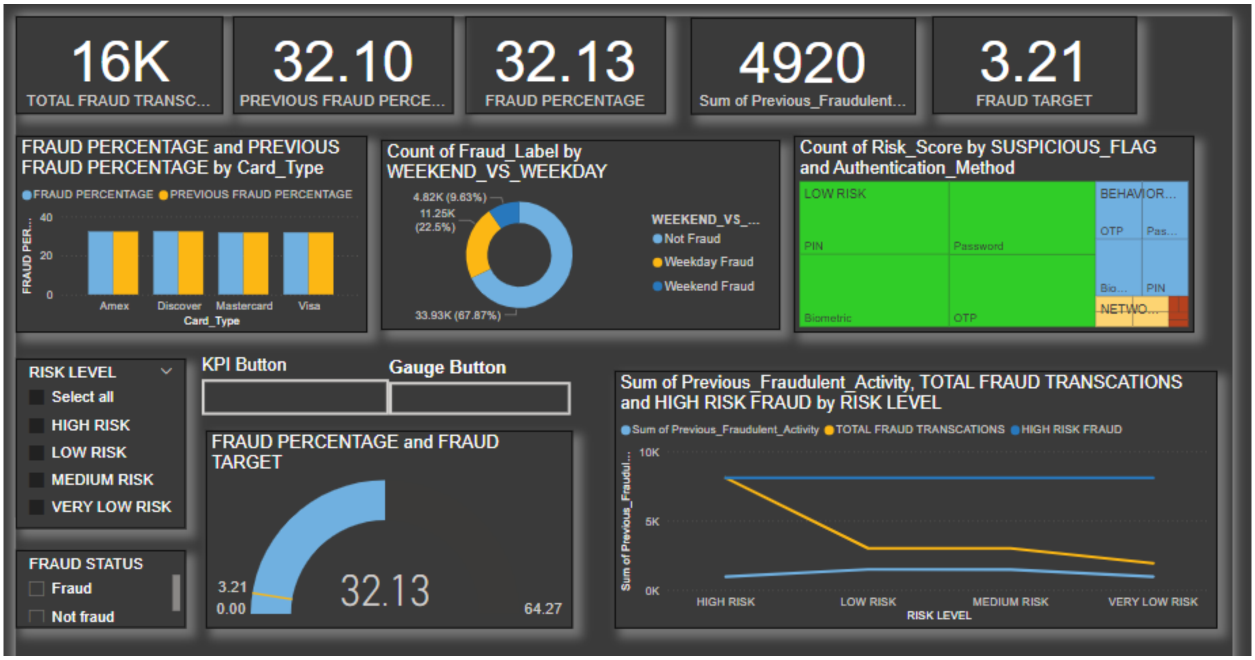
A. Pranusha

Reports:





A. Pranusha



Conclusion:

In this project, a comprehensive fraud analysis was conducted using a synthetic banking dataset consisting of 50,000 transactions. The analysis was performed entirely using **Excel, Power Query, and Power BI**, focusing on identifying patterns and trends related to fraudulent activities.

Key accomplishments include:

- **Data Cleaning and Transformation** using Power Query to prepare the dataset for analysis.
- Creation of multiple **custom flags and DAX measures** such as:
    - High Transaction Alert
    - Frequent Failed Transactions
    - Weekend Fraud Indicator
    - Risk Level and Suspicious Flags
- Use of **Power BI visuals** to analyze fraud across different dimensions such as:
    - Transaction Type, Card Type, Device Type, Location, and Risk Score
    - Time-based trends using Transaction Date and Weekend vs Weekday
- **Insightful dashboards** displaying:

  - Total Fraud Cases
  - Fraud % by Card Type, Risk Level, and Weekend/Weekday

This project demonstrates how real-world fraud detection logic can be applied using standard business intelligence tools without the need for machine learning or programming. The analysis helps in better **understanding suspicious patterns**, improving fraud detection logic, and enabling data-driven decision-making for financial institutions.