**FRAUDULENT TRANSACTION ANALYSIS USING SQL**

**1.Total number of fraud transactions**

SELECT COUNT(\*) as Fraud\_count

FROM fraud\_detection.fraudtrain

WHERE is\_fraud=1;

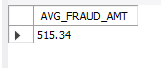


**2. Average fraud transaction amount**

SELECT ROUND(AVG(amt),2) as AVG\_FRAUD\_AMT

FROM fraud\_detection.fraudtrain

WHERE is\_fraud=1;



**3. Top 5 merchants by total fraud\_amount**

SELECT merchant, SUM(amt) as Tot\_fraud\_amt

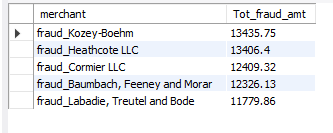
FROM fraud\_detection.fraudtrain

WHERE is\_fraud=1

GROUP BY merchant

ORDER BY Tot\_fraud\_amt DESC

LIMIT 5;



**4.Percentage of notable fradulent transactions per merchant**

SELECT\*FROM(

SELECT merchant,(COUNT(CASE WHEN is\_fraud=1 THEN 1 END)\*100/count(\*) )AS fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY merchant

ORDER BY fraud\_percent DESC) AS T

WHERE fraud\_percent>1;



**5. Number of fraud transactions per month**

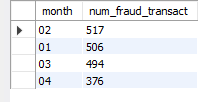
SELECT DATE\_FORMAT(trans\_date\_trans\_time, '%m') as month,COUNT(\*) AS num\_fraud\_transact

FROM fraud\_detection.fraudtrain

WHERE is\_fraud=1

GROUP BY month

ORDER BY num\_fraud\_transact DESC;



**6. TOP 5 cities with number of fraud transactions**

SELECT city,COUNT(\*) as num\_fraud\_transact

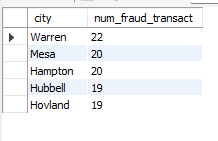
FROM fraud\_detection.fraudtrain

WHERE is\_fraud=1

GROUP BY city

ORDER BY num\_fraud\_transact DESC

LIMIT 5;



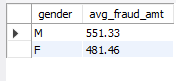
**7. Average fradulent transaction amount per card holder by gender**

SELECT gender,ROUND(AVG(amt),2)as avg\_fraud\_amt

FROM fraud\_detection.fraudtrain

WHERE is\_fraud=1

group by gender;



**8. highest risk merchant based on fraud percentage**

SELECT merchant,COUNT(\*) as total\_transactions,

SUM(CASE WHEN is\_fraud=1 THEN 1 ELSE 0 END) AS fraud\_transact,

(SUM(CASE WHEN is\_fraud=1 THEN 1 ELSE 0 END)/COUNT(\*))\*100 AS fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY merchant

ORDER BY fraud\_percent DESC

LIMIT 1;



**9. Fraud patterns by time of day(top 6)**

SELECT HOUR(trans\_date\_trans\_time) AS hour\_of\_day,COUNT(\*) as total\_transact,

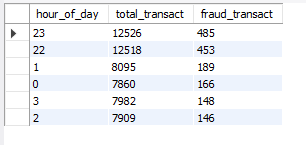
COUNT(CASE WHEN is\_fraud=1 THEN 1 END) as fraud\_transact

FROM fraud\_detection.fraudtrain

GROUP BY hour\_of\_day

ORDER BY fraud\_transact DESC

LIMIT 6;



**10. Fraud numbers and percentage based on gender**

SELECT gender,COUNT(\*) AS total\_transactions,

COUNT(CASE WHEN is\_fraud=1 THEN 1 END) AS num\_fraud\_transact,

(COUNT(CASE WHEN is\_fraud=1 THEN 1 END)/COUNT(\*))\*100 AS fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY gender;



**11. TOP fraud per category**

SELECT category,COUNT(\*) as total\_transact,

COUNT(CASE WHEN is\_fraud=1 THEN 1 END) AS fraud\_transact,

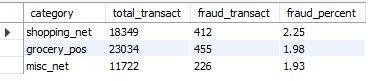
ROUND(COUNT(CASE WHEN is\_fraud=1 THEN 1 END)\*100/COUNT(\*),2) as fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY category

HAVING fraud\_percent>1

ORDER BY fraud\_percent DESC;



**12. Least fraud per category**

SELECT category,COUNT(\*) as total\_transact,

COUNT(CASE WHEN is\_fraud=1 THEN 1 END) AS fraud\_transact,

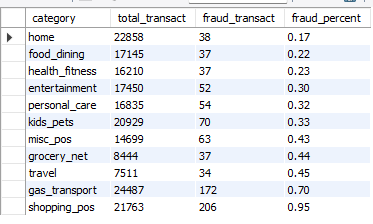
ROUND(COUNT(CASE WHEN is\_fraud=1 THEN 1 END)\*100/COUNT(\*),2) as fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY category

HAVING fraud\_percent<1

ORDER BY fraud\_percent;



**12. TOP FRAUD percentage by state**

SELECT state,COUNT(\*) AS total\_transact,

COUNT(CASE WHEN is\_fraud=1 THEN 1 END) as total\_fraud\_transact,

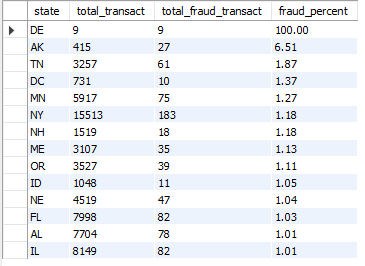
ROUND(COUNT(CASE WHEN is\_fraud=1 THEN 1 END)\*100/COUNT(\*),2) AS fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY state

HAVING fraud\_percent>1

ORDER BY fraud\_percent DESC;



**13. NO FRAUD percentage by state**

SELECT state,COUNT(\*) AS total\_transact,

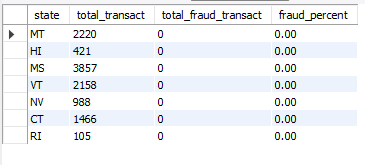
COUNT(CASE WHEN is\_fraud=1 THEN 1 END) as total\_fraud\_transact,

ROUND(COUNT(CASE WHEN is\_fraud=1 THEN 1 END)\*100/COUNT(\*),2) AS fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY state

HAVING fraud\_percent=0;



**14. least fraud percentage by state**

SELECT state,COUNT(\*) AS total\_transact,

COUNT(CASE WHEN is\_fraud=1 THEN 1 END) as total\_fraud\_transact,

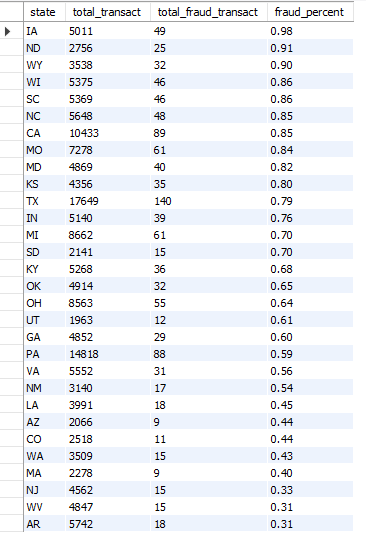
ROUND(COUNT(CASE WHEN is\_fraud=1 THEN 1 END)\*100/COUNT(\*),2) AS fraud\_percent

FROM fraud\_detection.fraudtrain

GROUP BY state

HAVING fraud\_percent>0 AND fraud\_percent<1

ORDER BY fraud\_percent DESC;



**CONCLUSION:**

We can deduce the following:

* Herman Terutel and Dickens have the highest fraud risk
* Warren is the city with most frauds
* Shopping\_net,mic\_net and grocery\_pos has the highest risk of fraud.
* Delaware(DE) has the highest fraud rate with a 100% chance
* MT, MI, HS, VT, NV, CT, and RI states have no fraudulent transactions whatsoever.
* The home category has the least chance of fraud
* Males are more prone to fraud than women.
* Most of the fraud transactions occur at 22:00 and 23:00 hrs
* Kozey-Bohem and Heatchcote LLC have the highest fraudulent amount
* Most of the frauds occurred in January and February
* IA have the highest fraud rate

**Data Source**: [Kaggle fraud transactions data](https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTrain.csv)

**By Saswata Maity**