```sql
CREATE DATABASE SELLER_ABUSE_PREVENTION;
USE SELLER_ABUSE_PREVENTION;
--data preview
SELECT  TOP 5 *  FROM LISTING_CLEANED;

SELECT TOP 5 * FROM SUSPICIOUS_CLEANED;

--Percentage of sellers flagged for suspicious activity
SELECT
(COUNT(DISTINCT SELLER_ID) *100)/(SELECT COUNT(*) FROM SELLERS_CLEANED)
AS percentage_of_sellers_flagged
FROM
SUSPICIOUS_CLEANED;
-- 61% of sellers flagged, lets dive deeper

SELECT
ACTIVITY_TYPE
,TYPE_PERCENTAGE
,SUM(TYPE_PERCENTAGE) OVER(ORDER BY TYPE_PERCENTAGE DESC ROWS BETWEEN
UNBOUNDED PRECEDING AND CURRENT ROW) AS ROLLING_PERCENTAGE
FROM
(
SELECT
ACTIVITY_TYPE
,COUNT(*) AS TOTAL_SELLERS
,COUNT(DISTINCT SELLER_ID)*100/(SELECT COUNT(*) FROM SELLERS_CLEANED)
AS TYPE_PERCENTAGE
FROM
SUSPICIOUS_CLEANED
GROUP BY
ACTIVITY_TYPE) AS A
ORDER BY TYPE_PERCENTAGE DESC

/*61% of sellers have been flagged for suspicious activity at least
once, showing widespread risk.
The 90% cumulative flag rate across activities means many sellers are
flagged for multiple abuses, indicating complex, overlapping
problems.*/
SELECT
COUNT(DISTINCT SELLER_NAME) AS
TOTAL_SELLERS_FLAGGED_FOR_MULTIPLE_ABUSES
FROM
(
SELECT
A.  SELLER_NAME AS SELLER_NAME
,COUNT(DISTINCT B.ACTIVITY_TYPE) AS NUMBER_OF_TYPE
FROM
SELLERS_CLEANED A
LEFT JOIN SUSPICIOUS_CLEANED B ON
A.  SELLER_ID=B.SELLER_ID
GROUP BY
A.  SELLER_NAME
HAVING COUNT(DISTINCT B.ACTIVITY_TYPE) >1) AS SUBQUERY
/*27 sellers are flagged for multiple types of suspicious activities,
showing they pose a higher risk.
Focusing on these sellers can help reduce abuse more effectively by
addressing complex, repeated behaviors.*/
```

```sql
SELECT
 A.  SELLER_NAME
,A.SELLER_ID
,MIN(B.LISTED_DATE) AS FIRST_LISTED_DATE
,MAX(C.DETECTED_ON) AS LAST_DETECTED_ON
,DATEDIFF(DAY,MIN(B.LISTED_DATE),MAX(C.DETECTED_ON)) AS
NUMBER_OF_DAYS_ACTIVE
,ROUND(SUM(B.UNITS_SOLD *B.PRICE),2) AS TOTAL_SALE
FROM
SELLERS_CLEANED A
JOIN LISTING_CLEANED B ON
B.  SELLER_ID=A.SELLER_ID
JOIN SUSPICIOUS_CLEANED C ON
C.  SELLER_ID=B.SELLER_ID
WHERE A.SELLER_ID IN (SELECT SELLER_ID FROM SUSPICIOUS_CLEANED)
GROUP BY
A.  SELLER_NAME
,A.SELLER_ID
ORDER BY TOTAL_SALE DESC
```

/*Sellers flagged for suspicious activity generated over $1.7 million
in sales before being detected, with some active for nearly a year.
While not all sales are confirmed abuse, this volume represents
significant potential risk to revenue and marketplace trust.
Strengthening
early detection and data quality will help minimize losses and protect
genuine sellers and customers.*/

```sql
SELECT
C.SEVERITY
,COUNT(*) AS NUMBER_OF_INCIDENTS
,COUNT(DISTINCT C.SELLER_ID) AS NUMBER_OF_SELLERS
FROM SUSPICIOUS_CLEANED C
GROUP BY
C.SEVERITY ORDER BY
NUMBER_OF_INCIDENTS DESC;
```

/*The majority (~95%) of abuse incidents are categorized within Low,
Medium, and High severity, showing a spread across different impact
levels.*/

```sql
WITH CTE AS
(
SELECT
YEAR
,MONTH
,SELLER_ID
,SELLER_NAME
,CATEGORY
,ROUND(TOTAL_SALE,2) AS TOTAL_SALE
FROM
```

```sql
(
SELECT
YEAR(A.LISTED_DATE) AS YEAR
,MONTH(A.LISTED_DATE) AS MONTH
,B.SELLER_ID AS SELLER_ID
,B.SELLER_NAME AS SELLER_NAME
,A.CATEGORY AS CATEGORY
,SUM(A.PRICE*A.UNITS_SOLD) AS TOTAL_SALE
, DENSE_RANK() OVER(PARTITION BY YEAR(A.LISTED_DATE),
MONTH(A.LISTED_DATE) ORDER BY SUM(A.PRICE*A.UNITS_SOLD) DESC) AS DR
FROM
LISTING_CLEANED A
JOIN SELLERS_CLEANED B ON
A.  SELLER_ID=B.SELLER_ID
WHERE B.SELLER_ID IN (SELECT SELLER_ID FROM SUSPICIOUS_CLEANED)
GROUP BY
YEAR(A.LISTED_DATE)
,MONTH(A.LISTED_DATE)
,B.SELLER_ID
,B.SELLER_NAME
,A.CATEGORY) AS SUBQUERY
WHERE DR=1)
SELECT
YEAR
,MONTH
,SELLER_NAME
,CATEGORY
,TOTAL_SALE
,CONCAT(ROUND((TOTAL_SALE *100)/(SELECT SUM(TOTAL_SALE) FROM
CTE),2),'%') AS PERCE
FROM
CTE
ORDER BY ROUND((TOTAL_SALE *100)/(SELECT SUM(TOTAL_SALE) FROM CTE),2)
DESC;


/*A small set of suspicious sellers are responsible for over 90% of
total flagged sales, indicating high concentration of abuse revenue.

Top 5 entries alone contribute ~45% of all suspicious revenue — key
players like:

Davis-Owens (Books) – □382K (9.66%)

Lee-Watkins (Clothing) – □358K (9.04%)

Reyes-Campbell (Electronics) – □356K (9.00%)

Clothing, Electronics, and Books are high-risk categories, frequently
dominated by top suspicious sellers.*/




#Final insight
Key Insights:
Top 5 suspicious sellers alone contributed over 44% of the total
suspicious revenue:
```

Davis-Owens (Oct 2024) – □382.6K (9.66%)
Lee-Watkins (Sep 2024) – □358.1K (9.04%)
Reyes-Campbell (Jun 2024) – □356.6K (9.00%)
Miller, Richardson and Parker (Aug 2024) – □345.1K (8.71%)
Lee-Watkins appeared multiple times (Books & Clothing), indicating
recurring high-revenue abuse across categories — likely a high-risk
profile.
Unidentified sellers (SELLER_NAME = 0) account for significant revenue
(□700K+), signaling either data masking or deliberate evasion — needs
urgent review.
Clothing and Electronics are the most abused categories, dominating the
top 10 in revenue — this could signal vulnerabilities in pricing or
listing controls.


```
SELECT
*  FROM
LISTING_CLEANED L
JOIN SELLERS_CLEANED S ON
S.SELLER_ID=L.SELLER_ID
JOIN SUSPICIOUS_CLEANED A ON
A.  SELLER_ID=S.SELLER_ID
```