DATA ANALYSIS FOR RETAIL SHOP IN SQL

Import Retail Sales dataset

select \* from [SQL - Retail Sales Analysis\_utf ]

select count(\*) from [SQL - Retail Sales Analysis\_utf ]

select \* from [SQL - Retail Sales Analysis\_utf ]

where transactions\_id is null

select \* from [SQL - Retail Sales Analysis\_utf ]

where sale\_date is null

select \* from [SQL - Retail Sales Analysis\_utf ]

where sale\_time is null

Data Cleaning

Find Null values in all columns

select \* from [SQL - Retail Sales Analysis\_utf ]

where sale\_time is null

or

sale\_date is null

or

gender is null

or

category is null

or

quantiy is null

or

price\_per\_unit is null

or

cogs is null

or

total\_sale is null

Delete Null values from all columns

delete from [SQL - Retail Sales Analysis\_utf ]

where sale\_time is null

or

sale\_date is null

or

gender is null

or

category is null

or

quantiy is null

or

price\_per\_unit is null

or

cogs is null

or

total\_sale is null

To Verify

select count(\*) from [SQL - Retail Sales Analysis\_utf ]

-------------------------------------------------------------------

select count(DISTINCT customer\_id) as total\_sales from [SQL - Retail Sales Analysis\_utf ]

Distinct category

select count(DISTINCT category) as total\_sales from [SQL - Retail Sales Analysis\_utf ]

select DISTINCT category from [SQL - Retail Sales Analysis\_utf ]

Data Analysis & Business Key problem & Answer

1-Q write an SQL query to retrieve all sales made on '2022-11-05'

select \* from [SQL - Retail Sales Analysis\_utf ]

where sale\_date='2022-11-05';

2-Q write an SQL query to retrieve all transactions where the category is 'clothing' and the quantity sold is more than 10 in the

month of Nov-2022--

SELECT \*

FROM [SQL - Retail Sales Analysis\_utf ]

WHERE category = 'Clothing'

AND FORMAT(sale\_date, 'yyyy-MM') = '2022-11'

AND quantiy >=4;

3-Q write an SQL query to calculate the total sales (total sales) for each category

SELECT

category,

SUM(total\_sale) AS net\_sale,

COUNT(\*) AS total\_orders

FROM [SQL - Retail Sales Analysis\_utf]

GROUP BY category;

4-Q write a SQL query to find average age of customers who purchased items from the 'Beauty' Category.

SELECT

AVG(age) AS avg\_age

FROM [SQL - Retail Sales Analysis\_utf]

WHERE category = 'Beauty';

5-Q write an SQL query to find all transaction where the total sale is greater than 1000

Select \* from [SQL - Retail Sales Analysis\_utf ]

where total\_sale >=1000

6-Q write an sql query to find the total number of transactions (transaction\_id) made by each gender in each category.

SELECT

category,

gender,

COUNT(\*) AS total\_transaction

FROM [SQL - Retail Sales Analysis\_utf]

GROUP BY category, gender;

7-Q write an SQL query to calculate the average sale for each month. Find out best-selling month in each year-

WITH MonthlySales AS (

SELECT

YEAR(sale\_date) AS sale\_year,

MONTH(sale\_date) AS sale\_month,

AVG(total\_sale) AS avg\_sale

FROM [SQL - Retail Sales Analysis\_utf]

GROUP BY YEAR(sale\_date), MONTH(sale\_date)

),

RankedSales AS (

SELECT

sale\_year,

sale\_month,

avg\_sale,

RANK() OVER (PARTITION BY sale\_year ORDER BY avg\_sale DESC) AS rank\_sales

FROM MonthlySales

)

SELECT

sale\_year,

sale\_month,

avg\_sale

FROM RankedSales

WHERE rank\_sales = 1

ORDER BY sale\_year, sale\_month;

8-Q write an sql query to find the top 5 customers based on the higest top sales

select \* from [SQL - Retail Sales Analysis\_utf ]

SELECT top 5

customer\_id,

SUM(total\_sale) AS total\_sales

FROM [SQL - Retail Sales Analysis\_utf]

GROUP BY customer\_id

ORDER BY total\_sales DESC;

9-Q write an sql query to find the number of unique customers who purchased items from each category

SELECT

category,

COUNT(DISTINCT customer\_id) AS unique\_customers

FROM [SQL - Retail Sales Analysis\_utf]

GROUP BY category;

10-Q write an sql query to create each shift and number of orders (Example Morning <=12,afternoon between 12 & 17, evening >17)

---select \* FROM [SQL - Retail Sales Analysis\_utf]

SELECT

CASE

WHEN DATEPART(HOUR, sale\_time) <= 12 THEN 'Morning'

WHEN DATEPART(HOUR, sale\_time) BETWEEN 12 AND 17 THEN 'Afternoon'

ELSE 'Evening'

END AS shift,

COUNT(transactions\_id) AS number\_of\_orders

FROM [SQL - Retail Sales Analysis\_utf]

GROUP BY

CASE

WHEN DATEPART(HOUR, sale\_time) <= 12 THEN 'Morning'

WHEN DATEPART(HOUR, sale\_time) BETWEEN 12 AND 17 THEN 'Afternoon'

ELSE 'Evening'

END

ORDER BY shift;