RETAIL SALES SQL PROJECT

--SQL SALES ANALYSIS PROJECT --

CREATE DATABASE Retail Sales DB;

--CREATE TABLE--

CREATE TABLE retail\_sales (transactions\_id INT PRIMARY KEY,

sale\_date DATE,

sale\_time TIME,

customer\_id INT,

gender VARCHAR(15),

age INT,

category VARCHAR(15),

quantiy INT,

price\_per\_unit FLOAT,

cogs FLOAT,

total\_sale FLOAT

) ;

-- VIEW COMPLETE DATA --

select \* from retail\_sales;

-- CONVERT TIME TO hh:mm:ss --

ALTER TABLE retail\_sales

ALTER COLUMN sale\_time TIME(0);

--VIEW THE DATA --

SELECT TOP 10 \* FROM retail\_sales;

SELECT count( \* ) FROM retail\_sales;

--DATA CLEANING--

--VIEW NULL VALUES--

select \* from retail\_sales

WHERE transactions\_id IS NULL

OR

sale\_date IS NULL

OR

sale\_time IS NULL

OR

gender IS NULL

OR

category IS NULL

OR

quantiy IS NULL

OR

cogs IS NULL

OR

total\_sale IS NULL;

-- DELETE NULL VALUES --

DELETE FROM retail\_sales

WHERE transactions\_id IS NULL

OR

sale\_date IS NULL

OR

sale\_time IS NULL

OR

gender IS NULL

OR

category IS NULL

OR

quantiy IS NULL

OR

cogs IS NULL

OR

total\_sale IS NULL;

--CHECK THE NUMBER OF ROWS LEFT --

select count(\*) from retail\_sales;

--DATA EXPLORATION

--HOW MUCH SALES DO WE HAVE?

SELECT COUNT(\*) as Total\_Sales FROM retail\_sales;

--HOW MANY CUSTOMERS WE HAVE?

SELECT

COUNT(DISTINCT customer\_id) AS Total\_Customers

FROM retail\_sales;

--HOW MANY CATEGORIES WE HAVE?

SELECT

COUNT(DISTINCT category) as Categories

from retail\_sales;

--DATA ANALYSIS (BUSINESS KEY PROBLEMS AND ANSWERS)

--Q1.sql query to retrieve information of the sales made on 2022-11-05

SELECT \* FROM retail\_sales

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

--Q2. retreive the transactions mase in the mont of Novemeber 2022 where the category is clothing and the quantity sold is more than or equal to 4

SELECT

category ,

SUM(quantiy)

FROM retail\_sales

GROUP BY category;

SELECT \*

FROM retail\_sales

WHERE category='Clothing'

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

AND quantiy >= 4;

-- Q.3 Write a SQL query to calculate the total sales (total\_sale) for each category.

SELECT

category,

sum(total\_sale) as total\_sales,

count(\*) as total\_orders

FROM retail\_sales

GROUP BY category;

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

SELECT

ROUND(AVG(age \*1.0),2) as average\_age

FROM retail\_sales

WHERE category='Beauty';

-- Q.5 Write a SQL query to find all transactions where the total\_sale is greater than 1000.

SELECT \*

FROM retail\_sales

WHERE total\_sale > 1000;

-- Q.6 Write a SQL query to find the total number of transactions (transaction\_id) made by each gender in each category.

SELECT

category,

gender,

count(transactions\_id) AS transactions

FROM retail\_sales

GROUP BY gender,category

ORDER BY 1;

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

SELECT \* FROM

(

SELECT

DATEPART(year,sale\_date) AS yr,

DATENAME(month, sale\_date) AS mnth,

ROUND(AVG(total\_sale),2) AS avg\_sales,

RANK() OVER (PARTITION BY DATEPART(year,sale\_date) ORDER BY AVG(total\_sale) DESC) AS rank

FROM retail\_sales

GROUP BY DATENAME(month, sale\_date),DATEPART(year,sale\_date)

) AS t1

WHERE rank=1;

-- Q.8 Write a SQL query to find the top 5 customers based on the highest total sales

SELECT

TOP 5 customer\_id ,

sum(total\_sale) AS total\_sale

FROM retail\_sales

GROUP BY customer\_id

ORDER BY 2 DESC;

-- Q.9 Write a SQL query to find the number of unique customers who purchased items from each category.

SELECT

category AS categories,

COUNT(DISTINCT customer\_id) AS cnt\_unique\_customers

FROM retail\_sales

GROUP BY category;

-- Q.10 Write a SQL query to create each shift and number of orders (Example Morning <12, Afternoon Between 12 & 17, Evening >17)

WITH Shift\_sales

AS

(

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

FROM retail\_sales

)

SELECT

shift,

COUNT(transactions\_id) AS no\_of\_orders

FROM Shift\_sales

GROUP BY shift;

--Q11. the age group where maximum orders were placed

WITH age\_sales

AS

(

SELECT \*,

CASE

WHEN age <20 THEN 'teenager'

WHEN age between 20 AND 40 THEN 'middle-age'

ELSE 'elderly'

END AS age\_group

FROM retail\_sales

)

SELECT

age\_group,

COUNT(\*) as Customers

FROM age\_sales

GROUP BY age\_group

ORDER BY Customers DESC;

--Q12. Which category of goods has the highest COGS and the corresponding sales

SELECT

category,

ROUND(SUM(cogs),2) AS Cost\_of\_goods\_sold,

SUM(total\_sale) AS sales

FROM retail\_sales

GROUP BY category

ORDER BY 2 DESC;

--Q13.Is there a strong correlation between quantity bought and profit per transaction?

SELECT

quantiy,

ROUND(AVG(total\_sale-cogs),2) AS Profit

FROM retail\_sales

GROUP BY quantiy

ORDER BY quantiy;

--Q14.Which category has the highest profit margin?

SELECT

category,

ROUND(AVG(total\_sale - cogs), 2) AS avg\_profit,

ROUND(AVG((total\_sale - cogs) \* 1.0 / NULLIF(cogs, 0)), 2) AS avg\_margin\_ratio

FROM retail\_sales

GROUP BY category

ORDER BY avg\_margin\_ratio DESC;

--Q15. Which customers are repeat buyers and how much do they spend on average?

SELECT

customer\_id,

COUNT(\*) AS visits,

ROUND(AVG(total\_sale),2) AS avg\_amount\_spend

FROM retail\_sales

GROUP BY customer\_id

HAVING 2>1

ORDER BY 2 DESC;