SQL PROJECT ON RETAIL SALES

Hello!

My Name is **Waleed Ilyas**. In this project, I utilized SQL Queries to solve questions related to Retail Sales.

PROJECT Introduction

This project is designed to demonstrate SQL skills and techniques typically used by data analysts to explore, clean, and analyze retail sales data. The project involves setting up a retail sales database, performing exploratory data analysis (EDA), and answering specific business questions through SQL queries.

OBJECTIVES

01. **Set up a retail sales database:** Create and populate a retail sales database with the provided sales data.

02. **Data Cleaning:** Identify and remove any records with missing or null values.

03. **Exploratory Data Analysis (EDA):** Perform basic exploratory data analysis to understand the dataset.

04. **Business Analysis:** Use SQL to answer specific business questions and derive insights from the sales data.

Set up a retail sales database

```sql
-- Create Database
CREATE DATABASE Retail_Sale;

-- 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
);
```

Data Cleaning

```sql
-- Data Cleaning
Select * from retail_sales
where
transactions_id is null
OR
sale_date is null
OR
sale_time is null
OR
customer_id is null
OR
gender is null
OR
age 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;
```

After Finding out the null values, Run this Query

```sql
Delete from retail_sales
where
transactions_id is null
OR
sale_date is null
OR
sale_time is null
OR
customer_id is null
OR
gender is null
OR
age 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;
```

```sql
-- Data Exploration

-- How many Sales we have?

Select COUNT(*) AS Total_Sale FROM Retail_Sales;
```

	total_sale bigint 🔒
1	1987

```sql
-- How many Unique Customers we have?

Select COUNT(DISTINCT customer_id) AS Total_Customers FROM Retail_Sales;
```

total_customers bigint 🔒
155

```sql
-- How many Categories we have?

Select DISTINCT category FROM retail_sales;
```

	category character varying (15) 🔒
1	Electronics
2	Clothing
3	Beauty

Business Analysis

Q # 1: Write a SQL query to retrieve all columns for sales made on '2022-11-05'.

```
Select * from retail_sales
where sale_date = '2022-11-05';
```

transactions_id [PK] integer	sale_date date	sale_time time without time zone	customer_id integer	gender character varying (15)	age integer	category character varying (15)	quantiy integer
180	2022-11-05	10:47:00	117	Male	41	Clothing	3
240	2022-11-05	11:49:00	95	Female	23	Beauty	1
1256	2022-11-05	09:58:00	29	Male	23	Clothing	2
1587	2022-11-05	20:06:00	140	Female	40	Beauty	4
1819	2022-11-05	20:44:00	83	Female	35	Beauty	2
943	2022-11-05	19:29:00	90	Female	57	Clothing	4
1896	2022-11-05	20:19:00	87	Female	30	Electronics	2
1137	2022-11-05	22:34:00	104	Male	46	Beauty	2
856	2022-11-05	17:43:00	102	Male	54	Electronics	4
214	2022-11-05	16:31:00	53	Male	20	Beauty	2
1265	2022-11-05	14:35:00	86	Male	55	Clothing	3

Q # 2: Write a SQL query to retrieve all transactions where the category is 'Clothing' and the quantity sold is more than 4 in the month of Nov-2022

```sql
Select * from retail_sales
where
category = 'Clothing'
AND
TO_CHAR(sale_date, 'YYYY-MM') = '2022-11'
AND
quantiy > 3;
```

transactions_id [PK] integer	sale_date date	sale_time time without time zone	customer_id integer	gender character varying (15)	age integer	category character varying (15)	quantiy integer
1484	2022-11-23	09:29:00	22	Female	19	Clothing	4
64	2022-11-15	06:34:00	7	Male	49	Clothing	4
284	2022-11-12	09:17:00	129	Male	43	Clothing	4
1885	2022-11-09	07:32:00	148	Female	52	Clothing	4
547	2022-11-14	07:36:00	3	Male	63	Clothing	4
159	2022-11-10	21:30:00	42	Male	26	Clothing	4
699	2022-11-21	22:21:00	129	Female	37	Clothing	4
1259	2022-11-03	17:31:00	105	Female	45	Clothing	4
146	2022-11-10	22:01:00	74	Male	38	Clothing	4
1476	2022-11-11	22:27:00	130	Female	27	Clothing	4
1296	2022-11-26	20:42:00	45	Female	22	Clothing	4
1696	2022-11-21	17:59:00	24	Female	50	Clothing	4
1497	2022-11-19	21:44:00	109	Male	41	Clothing	4

Q # 3: Write a SQL query to calculate the total sales (total_sale) for each category

```sql
Select category,
SUM(total_sale) as Net_Sale,
Count(*) as Total_Orders
From retail_sales
Group by 1;
```

	category character varying (15) 🔒	net_sale double precision 🔒	total_orders bigint 🔒
1	Electronics	311445	678
2	Clothing	309995	698
3	Beauty	286790	611

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

```sql
Select
Round(AVG(age),2) as AVG_Age
from retail_sales
where category = 'Beauty';
```

	avg_age numeric
1	40.42

Q # 5: Write a SQL query to find all transactions where the total_sale is greater than 1000.

```sql
Select * from retail_sales
where total_sale > 1000;
```

customer_id integer	gender character varying (15)	age integer	category character varying (15)	quantiy integer	price_per_unit double precision	cogs double precision	total_sale double precision
52	Male	46	Beauty	3	500	145	1500
5	Female	40	Clothing	4	300	84	1200
48	Male	46	Beauty	3	500	235	1500
49	Female	40	Clothing	4	300	144	1200
66	Female	37	Clothing	3	500	235	1500
59	Female	37	Clothing	3	500	185	1500
135	Female	19	Clothing	4	300	75	1200

Q # 6: Write a SQL query to find the total number of transactions (transaction_id) made by each gender in each category.:

```sql
Select category, gender,
Count(*) as Total_Trans
From retail_sales
Group By
category, gender
Order by 1
```

category character varying (15)	gender character varying (15)	total_trans bigint
Beauty	Female	330
Beauty	Male	281
Clothing	Female	347
Clothing	Male	351
Electronics	Male	343
Electronics	Female	335

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

```sql
Select year, month, avg_sale
FROM(
Select
EXTRACT(YEAR from sale_date) as year,
EXTRACT(MONTH from sale_date) as month,
AVG(total_sale) as avg_sale,
RANK() OVER(PARTITION BY EXTRACT(YEAR from sale_date) Order BY AVG(total_sale) DESC ) as Rank
from retail_sales
Group By 1,2
) as t1
where rank = 1
```

year numeric	month numeric	avg_sale double precision
2022	7	541.3414634146342
2023	2	535.531914893617

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

```sql
Select customer_id,
SUM (total_sale) as Total_Sale
from retail_sales
group by customer_id
Order by total_sale DESC
Limit 5
```

customer_id integer	total_sale double precision
3	38440
1	30750
5	30405
2	25295
4	23580

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

```sql
Select category,
COUNT(DISTINCT customer_id) as cnt_unique_cs
From retail_sales
Group BY category;
```

category character varying (15) 🔒	cnt_unique_cs bigint 🔒
Beauty	141
Clothing	149
Electronics	144

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

```sql
With hourly_sale
As
(Select *,
case
    WHEN EXTRACT (HOUR FROM sale_time) < 12 THEN 'MORNING'
    WHEN EXTRACT (HOUR From sale_time) BETWEEN 12 AND 17 THEN 'AFTERNOON'
    ELSE 'Evening'
    END as Shift
From retail_sales
)
Select Shift,
COUNT(*) as total_orders
From hourly_sale
Group By Shift
```

shift text	total_orders bigint
Evening	1062
AFTERNOON	377
MORNING	548

Thank You

📞 +92 3176063654

in linkedin.com/in/waleed-ilyas-664839213

✉️ waleedilyas99@gmail.com