SQL ONLINE SALES ANALYSIS

- SAIPAVAN MEKALA

Introduction

This project involves analyzing a dataset of online sales transactions to uncover insights into sales trends, regional preferences, product popularity, and the impact of different payment methods. The goal is to optimize inventory, marketing strategies, and understand consumer behavior across various regions and product categories.

Description

This dataset contains detailed records of online sales transactions, including order ID, date, product category, product name, quantity sold, unit price, total price, region, and payment method. It provides a comprehensive view of sales activities across different regions and categories

Data Structure of Table

Column Name	Data Type	Description
Order ID	String/Integer	Unique identifier for each sales order
Date	Date/DateTime	Date of the sales transaction
Category	String	Broad category of the product sold
Product Name	String	Specific name or model of the product sold
Quantity	Integer	Number of units of the product sold in the transaction
Unit Price	Float/Decimal	Price of one unit of the product
Total Price	Float/Decimal	Total revenue generated from the sales transaction
Region	String	Geographic region where the transaction occurred
Payment Method	String	Method used for payment

Type of Analysis can perform

- **Sales Trends:** Identify seasonal patterns and peak periods to optimize inventory and marketing strategies.
- **Category and Regional Analysis**: Discover popular product categories in different regions to tailor offerings and maximize sales.
- **Top Products and Payment Methods:** Determine best-selling products and the impact of payment methods on sales to refine promotions and payment options.

Type of Analysis can perform

- **Revenue Analysis:** Evaluate total revenue across different categories and regions to identify key revenue drivers and areas for improvement.
- **Product Performance Analysis:** Compare the performance of specific products in various regions to tailor offerings and marketing efforts to regional preferences and demands.

PROCESS INVOLVED

- **Data Wrangling:** This includes importing data, handling missing values, removing duplicates, standardizing data types, correcting inconsistencies, and calculating derived fields. The goal is to ensure data integrity and usability, enabling accurate and insightful analysis.

- **Exploratory Data Analysis (EDA):** The process of analyzing datasets to uncover key insights. It involves using statistical summaries and visualizations to identify trends, patterns, anomalies, and relationships within the data,

1. Find the total revenue (Total Price) for each Category.

QUERY

```
SELECT
PRODUCT_CATEGORY,
SUM(TOTAL_REVENUE) AS TOTAL_REVENUE
FROM SALES_DATA_2023
GROUP BY PRODUCT_CATEGORY
ORDER BY TOTAL_REVENUE DESC;
```

OUTPUT

PRODUCT_CATEGORY	TOTAL_REVENUE
Electronics	34983.00
Home Appliances	18646.16
Sports	14328.00
Clothing	8131.00
Beauty Products	2621.90
Books	1861.00

2. Retrieve the Region with the most sales and its revenue.

QUERY

```
SELECT Region,SUM(units_sold) Total_Sales,
SUM(total_revenue) Total_Revenue
FROM sales_data_2023
GROUP BY region
ORDER BY Total_Sales DESC ;
```

OUTPUT

Region	Total_Sales	Total_Revenue
Asia	233	22459.00
North America	180	36844.00
Europe	105	21268.06

3. Find the top 5 costliest product.

QUERY

```
SELECT PRODUCT_NAME, MAX(unit_price) AS Price
FROM sales_data_2023
GROUP BY PRODUCT_NAME
ORDER BY PRICE DESC
LIMIT 5;
```

OUTPUT

PRODUCT_NAME	Price
Canon EOS R5 Camera	3900
MacBook Pro 16-inch	2500
Apple MacBook Pro 16-inch	2399
Peloton Bike	1895
HP Spectre x360 Laptop	1600

4. Retrieve the Product Name and Quantity where the Quantity is greater than or equal to 10.

QUERY

```
SELECT PRODUCT_NAME,
SUM(UNITS_SOLD) AS TOTAL_QUANTITY
FROM SALES_DATA_2023
GROUP BY PRODUCT_NAME
HAVING TOTAL_QUANTITY >=10;
```

OUTPUT

PRODUCT_NAME	TOTAL_QUANTITY
Hanes ComfortSoft T-Shirt	10

5.List all unique Regions where sales occurred.

QUERY

```
SELECT
DISTINCT REGION
FROM SALES_DATA_2023;
```

OUTPUT

REGION
North America
Europe
Asia

6. Count the number of transactions for each Payment Method

QUERY

```
SELECT PAYMENT_METHOD ,
COUNT(PAYMENT_METHOD) AS TOTAL_TRANSACTIONS
FROM SALES_DATA_2023
GROUP BY PAYMENT_METHOD;
```

OUTPUT

PAYMENT_METHOD	TOTAL_TRANSACTIONS
Credit Card	120
PayPal	80
Debit Card	40

7.Retrieve the Transaction_ID, Product Name, and Total Price for transactions on a specific date (e.g., '2023-02-01').

QUERY

```
SELECT TRANSACTION_ID,PRODUCT_NAME,
(UNITS_SOLD * UNIT_PRICE) AS TOTAL_PRICE
FROM SALES_DATA_2023
WHERE DATE = '2023-02-01';
```

OUTPUT

TRANSACTION_ID	PRODUCT_NAME	TOTAL_PRICE
10032	Instant Pot Duo	270

8. Find the date with the highest total revenue and list all transactions on that date.

QUERY

```
WITH TOTAL_REVENUE AS (
SELECT `DATE`,MAX(TOTAL_REVENUE) AS TOTAL_REVENUE,TRANSACTION_ID
FROM SALES_DATA_2023
GROUP BY `DATE`,TRANSACTION_ID
ORDER BY TOTAL_REVENUE DESC
LIMIT 1
),
ALL_TRANSACTION AS (
SELECT TRANSACTION_ID,`DATE` FROM SALES_DATA_2023
WHERE DATE = (SELECT `DATE` FROM TOTAL_REVENUE)
)
SELECT TR.`DATE`,TR.TOTAL_REVENUE, AT.TRANSACTION_ID
FROM TOTAL_REVENUE AS TR
JOIN ALL_TRANSACTION AS AT
ON TR.`DATE`=AT.`DATE`;
```

OUTPUT

DATE	TOTAL_REVENUE	TRANSACTION_ID
2023-04-12	3900.00	10103

9.Find the Top 3 Transaction_ID’s and Total Price for orders where the Total Price is higher than the average Total Price of all orders.

QUERY

```
SELECT TRANSACTION_ID, TOTAL_REVENUE
FROM SALES_DATA_2023
WHERE TOTAL_REVENUE > (SELECT AVG(TOTAL_REVENUE) FROM SALES_DATA_2023)
ORDER BY TOTAL_REVENUE
DESC LIMIT 3;
```

OUTPUT

TRANSACTION_ID	TOTAL_REVENUE
10103	3900.00
10086	2599.98
10007	2500.00

10. Calculate the percentage contribution of each Category to the total revenue.

QUERY

```
SELECT PRODUCT_Category,
SUM(TOTAL_REVENUE) AS category_total_revenue,
ROUND((SUM(TOTAL_REVENUE)/(SELECT SUM(TOTAL_REVENUE) FROM sales_data_2023) * 100),1)
AS percentage_contribution
FROM sales_data_2023
GROUP BY PRODUCT_Category;
```

OUTPUT

PRODUCT_Category	category_total_revenue	percentage_contribution
Electronics	34983.00	43.4
Home Appliances	18646.16	23.1
Clothing	8131.00	10.1
Books	1861.00	2.3
Beauty Products	2621.90	3.3
Sports	14328.00	17.8

11. Get the Region with the highest average Total Price per transaction.

QUERY

```
SELECT REGION,
AVG(UNIT_PRICE) AS HIGHEST_AVG_TRANSACTION
FROM SALES_DATA_2023
GROUP BY REGION
ORDER BY HIGHEST_AVG_TRANSACTION DESC
LIMIT 1;
```

OUTPUT

REGION	HIGHEST_AVG_TRANSACTION
North America	353.8750

12. Find the Product Name with the highest total revenue (Total Price) in each Region.

QUERY

```
WITH REVENUE AS (
SELECT PRODUCT_NAME,REGION,(UNITS_SOLD * UNIT_PRICE) AS TOTAL_REVENUE
FROM SALES_DATA_2023
),
REGION AS (
SELECT REGION,MAX(TOTAL_REVENUE) AS MAX_REVENUE
FROM REVENUE GROUP BY REGION
)
SELECT R.PRODUCT_NAME,R.REGION,R.TOTAL_REVENUE FROM REVENUE R JOIN REGION RE ON
 R.REGION = RE.REGION AND R.TOTAL_REVENUE = RE.MAX_REVENUE;
```

OUTPUT

PRODUCT_NAME	REGION	TOTAL_REVENUE
Peloton Bike	Asia	1895
LG OLED TV	Europe	2600
Canon EOS R5 Camera	North America	3900

Conclusion

- Total revenue generated is $80,571.06.
- Electronic accessories generate the most revenue, accounting for 43% of total sales, followed by home appliances.
- North America generates the highest revenue, totaling $36,844.
- Most sales came from Asia.
- Most transactions are conducted using credit cards.
- The costliest product was the Canon EOS R5 Camera, priced at $3,900.

Thank You

- SAIPAVAN MEKALA