Paytm Epurchase Data

Name : KOULURU NANDA KISHORE REDDY
Email : nandakouluruemail@gmail.com
Linkedin: https://www.linkedin.com/in/kouluru-nanda-kishore-reddy-b42972192/

Task-3

Aim: To analyse the given dataset ‘Paytm Epurchase Data’ and perform the following queries in MS SQL.

Used Databases Names: Purchase_data$

1. What does the "Category_Grouped" column represent, and how many unique categories are there?

ANS

The Category_Grouped column represents which category of item is present in the column.

```
SELECT COUNT(DISTINCT Category_Grouped) FROM Purchase_data$;
```

2. Can you list the top 5 shipping cities in terms of the number of orders?

ANS

```
SELECT TOP 5 Shipping_city,COUNT(*) AS OrdersCount FROM Purchase_data$
GROUP BY Shipping_city ORDER BY OrdersCount DESC;
```

3. Show me a table with all the data for products that belong to the "Electronics" category.

ANS

```
SELECT * FROM Purchase_data$ WHERE Category_Grouped='Electronics';
```

4. Filter the data to show only rows with a "Sale_Flag" of 'Yes'.

ANS

```
SELECT * FROM Purchase_data$ WHERE Category_Grouped = 'Electronics' AND
Sale_Flag = 'Yes';
```

5. Sort the data by "Item_Price" in descending order. What is the most expensive item?

ANS

```
SELECT TOP 1 * FROM Purchase_data$ ORDER BY Item_Price DESC;
```

Name : KOULURU NANDA KISHORE REDDY
Email : nandakouluruemail@gmail.com
Linkedin: https://www.linkedin.com/in/kouluru-nanda-kishore-reddy-b42972192/

6. Apply conditional formatting to highlight all products with a "Special_Price_effective" value below $50 in red.
ANS

```
SELECT *,
 CASE
        WHEN Special_Price_effective < 50 THEN 'Below $50'
 ELSE 'Above $50'
 END AS Price_Category FROM Purchase_data$;
```

7. Create a pivot table to find the total sales value for each category.
ANS

```
SELECT Category_Grouped, SUM(Item_Price) AS TotalSalesValue FROM
Purchase_data$ GROUP BY Category_Grouped;
```

8. Create a bar chart to visualize the total sales for each category.
ANS

```
SELECT SUM(item_price) AS total_sales,Category FROM Purchase_data$
Group By Category ORDER BY total_sales DESC;
```

9. Create a pie chart to show the distribution of products in the "Family" category.
ANS

```
SELECT Family,COUNT(*) AS total_prODUCTS FROM Purchase_data$ GROUP
BY Family;
```

10. Ensure that the "Payment_Method" column only contains valid payment methods (e.g., Visa, MasterCard).
ANS

```
UPDATE Purchase_data$
SET Payment_Method =
 CASE
        WHEN Payment_Method NOT IN ('Visa', 'MasterCard') THEN NULL
ELSE Payment_Method
END;
```

11. Calculate the average "Quantity" sold for products in the "Clothing" category, grouped by "Product_Gender."
ANS

```
SELECT Product_Gender, AVG(Quantity) AS AverageQuantitySold FROM
Purchase_data$ WHERE Category_Grouped = 'Clothing' GROUP BY
Product_Gender;
```

Name : KOULURU NANDA KISHORE REDDY
Email : nandakouluruemail@gmail.com
Linkedin: https://www.linkedin.com/in/kouluru-nanda-kishore-reddy-b42972192/

12. Find the top 5 products with the highest "Value_CM1" and "Value_CM2" ratios. Create a chart to visualize this data.

ANS

```
SELECT TOP 5 *, Value_CM1 / NULLIF(Value_CM2, 0) AS
Ratio_Value_CM1_to_CM2 FROM Purchase_data$ WHERE Value_CM2 <> 0
ORDER BY Ratio_Value_CM1_to_CM2 DESC;
```

13. Identify the top 3 "Class" categories with the highest total sales. Create a stacked bar chart to represent this data.

ANS

```
SELECT TOP 3 Class, SUM(Item_Price) AS TotalSales FROM Purchase_data$
GROUP BY Class ORDER BY TotalSales DESC;
```

14. Use VLOOKUP or INDEX-MATCH to retrieve the "Color" of a product with a specific "Item_NM."

ANS

```
SELECT Color FROM Purchase_data$ WHERE Item_NM =
'your_specific_Item_NM';
```

15. Calculate the total "coupon_money_effective" and "Coupon_Percentage" for products in the "Electronics" category.

ANS

```
SELECT SUM(coupon_money_effective) AS
TotalCouponMoney,SUM(Coupon_Percentage) AS TotalCouponPercentage FROM
Purchase_data$ WHERE Category_Grouped = 'Electronics';
```

16. Perform a time series analysis to identify the month with the highest total sales.

ANS

```
SELECT top 1 DATEPART(month, YourDateColumn) AS SalesMonth,
SUM(Item_Price) AS TotalSales FROM Purchase_data$ GROUP BY
DATEPART(month, YourDateColumn)ORDER BY TotalSales DESC;
```

17. Calculate the total sales for each "Segment" and create a scatter plot to visualize the relationship between "Item_Price" and "Quantity" in this data.

ANS

```
SELECT Segment,SUM(item_price) AS total_sales FROM Purchase_data$ GROUP
BY segment;
```

Name : KOULURU NANDA KISHORE REDDY
Email : nandakouluruemail@gmail.com
Linkedin: https://www.linkedin.com/in/kouluru-nanda-kishore-reddy-b42972192/

18. Use the AVERAGEIFS function to find the average "Item_Price" for products that have a "Sale_Flag" of 'Yes.'

ANS

```
SELECT AVG(Item_Price) AS AverageItemPrice FROM Purchase_data$ WHERE
Sale_Flag = 'Yes';
```

19. Identify products with a "Paid_pr" higher than the average in their respective "Family" and "Brand" groups.

ANS

```
SELECT * FROM Purchase_data$ P1 WHERE Paid_pr > (SELECT
AVG(Paid_pr)FROM Purchase_data$ P2
WHERE P1.Family = P2.Family AND P1.Brand = P2.Brand);
```

20. Create a pivot table to show the total sales for each "Color" within the "Clothing" category and use conditional formatting to highlight the highest sales

ANS

```
SELECT Color,SUM(Item_Price) AS TotalSales FROM Purchase_data$
WHERE Category_Grouped = 'Clothing'GROUP BY Color;
```