

ASSIGNMENT - 5

Q1. You are an Analyst working for an E-commerce company. Your manager had a discussion with the leadership and asked you to work on the analysis below.

- A. Create a YOY analysis for the count of customers enrolled with the company each month. The output should look like:

| Month | Year_2020 | Year_2021 |
|-------|-----------|-----------|
| 1 | | |
| 2 | | |

- B. Find out the top 3 best-selling products in each of the categories that are currently active on the Website
- C. Find the out the least selling products in each of the categories that are currently active on the website
- D. We are trying to find paired products that are often purchased together by the same user, such as chips and soft drinks, milk and curd etc.. Find the top paired products names.
- E. We want to understand the impact of running a campaign during July'21-Oct'21 what was the total sales generated for the categories "Beauty & Hygiene" and "Bevarages" by
- entire customer base
 - customers who enrolled with the company during the same period
- F. Create a Quarter-wise ranking in terms of revenue generated in each category in Year 2020
- G. Find the top 3 Shipper companies in terms of
- Average delivery time for each category for the latest year
 - Volume for latest year
- H. Find the top 25 customers in terms of
- Total no. of orders placed for Year 2021

b. Total Purchase Amount for the Year 2021

- I. Find out the difference between the last two order dates for each of the customers and categorize the customers in two categories such that if the difference is less than 5 days tag the customer as “Frequent Buyer” else tag it as “Infrequent”.
- J. Find the cumulative average order amount at a monthly level for year 2021
- a. Each category
 - b. Each customer
- K. Find the 3-day rolling average for the total purchase amount by each customer.
- L. Create the below table where values for each Payment method should contain the total order amount by each customer resp.

| CustomerName | Debit Card | POD | PayPal | Credit Card | Wallet | Net banking |
|--------------|------------|-----|--------|-------------|--------|-------------|
| A | | | | | | |
| B | | | | | | |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |

- M. Create a Procedure to filter the orders from Orders table where total purchase amount in a each order id < @x and purchase of year is @Y where @x and @y are the inputs provided by the user.

Q2. What is the difference between Group by and Partition by?

Q3. What is the difference between a Temp Table and a View?

Q4. What is the difference between View and CTE?

Q5. What is the difference between Row Number, Rank and Dense Rank?

Q6. Consider the single-column table below.

| |
|------|
| COL1 |
|------|

| |
|------|
| 1 |
| 1.2 |
| 1.5 |
| 2.2 |
| NA |
| NAN |
| NULL |

Answer the following questions:

1. What is the possible data type of the column 'COL1'?
2. What will the output of the following SQL statements
 - a. 'SELECT COUNT (*) AS ENTRIES FROM TABLE;'
 - b. 'SELECT COUNT(COL1) AS ENTRIES FROM TABLE;'
 - c. 'SELECT COUNT (DISTINCT COL1) AS ENTRIES FROM TABLE;'

---1-a

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains the following SQL code:

```

---1-a
select*from (select month(DateEntered) as month_, year(DateEntered) as year_,
count (distinct customerid) as total_ from Customers1
group by month(dateentered),year(dateentered) )c
pivot (sum(total_) for year_ in ([2020],[2021])) as pivot_
--or
select months,sum([2020]) as year_2020, sum([2021]) as year_2021 from (select
month(dateentered) as months, [2020],[2021] from (select*,year(dateentered)as year_ from Customers1)A
pivot (count(year_) for year_ in ([2020],[2021])) as pvt)b group by B.months;
---1.b
select top 3 categoryname as top_, sum(c.quantity)as total_sold from category A right join Products b on a.CategoryID = b.categoryid
left join OrderDetails c on b.ProductID = c.ProductID
left join orders d on c.OrderID=d.OrderID
where a.Active = '1'
group by CategoryName
order by total_sold desc;
---1.c

```

The Results pane shows the output of the first query (---1-a):

| | month_ | 2020 | 2021 |
|----|--------|------|------|
| 1 | 1 | 10 | 40 |
| 2 | 2 | 23 | 41 |
| 3 | 3 | 30 | 39 |
| 4 | 4 | 30 | 25 |
| 5 | 5 | 21 | 23 |
| 6 | 6 | 20 | 23 |
| 7 | 7 | 10 | 20 |
| 8 | 8 | 15 | 22 |
| 9 | 9 | 22 | 20 |
| 10 | 10 | 18 | 23 |
| 11 | 11 | 10 | 20 |
| 12 | 12 | 10 | 10 |

The status bar at the bottom indicates: "Query executed successfully. LAPTOP-EB69CUGG\NIRAJKR (15... LAPTOP-EB69CUGG\NIRAJKR... ECOMM2 000000 12 rows".

```

select*from (select month(DateEntered) as month_, year(DateEntered) as year_,
count (distinct customerid) as total_ from Customers1
group by month(dateentered),year(dateentered) )c
pivot (sum(total_) for year_ in ([2020],[2021])) as pivot_;

```

--or

```

select months,sum([2020]) as year_2020, sum([2021]) as year_2021 from (select
month(dateentered) as months, [2020],[2021] from (select*,year(dateentered)as year_ from Customers1)A
pivot (count(year_) for year_ in ([2020],[2021])) as pvt)b group by B.months;

```

---1.b

Microsoft SQL Server Management Studio interface showing a query execution window. The query is as follows:

```

---1.a
select*from (select month(DateEntered) as month_, year(DateEntered) as year_,
count (distinct customerId) as total_ from Customers1
group by month_ ,year(dateentered) )c
pivot (sum(total_) for year_ in ([2020],[2021])) as piv1;
--or
select months,sum([2020]) as year_2020, sum([2021]) as year_2021 from (select month(dateentered) as months, [2020],[2021] from (select*,year(dateentered)as year_ from Customers1)A
pivot (count(year_) for year_ in ([2020],[2021])) as piv1b group by 8.months;

---1.b
select top 3 categoryname as top_ , sum(c.quantity)as total_sold from category A right join Products b on a.CategoryID = b.categoryid
left join OrderDetails c on b.ProductID = c.ProductID
left join orders d on c.OrderID=d.OrderID
where a.Active = '1'
group by CategoryName
order by total_sold desc;

---1.c

```

The Results window shows the following data:

| top_ | total_sold |
|--------------------------|------------|
| 1 Beauty & Hygiene | 80213 |
| 2 Gourmet & World Food | 51130 |
| 3 Kitchen, Garden & Pets | 38677 |

Query executed successfully.

```

select top 3 categoryname as top_ , sum(c.quantity)as total_sold from category A right
join Products b on a.CategoryID = b.categoryid
left join OrderDetails c on b.ProductID = c.ProductID
left join orders d on c.OrderID=d.OrderID
where a.Active = '1'
group by CategoryName
order by total_sold desc;

```

---1.c

Microsoft SQL Server Management Studio interface showing a query execution window. The query is as follows:

```

---1.a
select months,sum([2020]) as year_2020, sum([2021]) as year_2021 from (select month(dateentered) as months, [2020],[2021] from (select*,year(dateentered)as year_ from Customers1)A
pivot (count(year_) for year_ in ([2020],[2021])) as piv1b group by 8.months;

---1.b
select top 3 categoryname as top_ , sum(c.quantity)as total_sold from category A right join Products b on a.CategoryID = b.categoryid
left join OrderDetails c on b.ProductID = c.ProductID
left join orders d on c.OrderID=d.OrderID
where a.Active = '1'
group by CategoryName
order by total_sold desc;

---1.c
select top 3 categoryname as top_ , sum(c.quantity)as total_sold from category A right join Products b on a.CategoryID = b.categoryid
left join OrderDetails c on b.ProductID = c.ProductID
left join orders d on c.OrderID=d.OrderID
where a.Active = '1'
group by CategoryName
order by total_sold

---1.d

```

The Results window shows the following data:

| top_ | total_sold |
|-------------------------|------------|
| 1 Fruits & Vegetables | 5410 |
| 2 Beverages | 9109 |
| 3 Bakery, Cakes & Dairy | 9332 |

Query executed successfully.

```

select top 3 categoryname as top_ , sum(c.quantity)as total_sold from category A right
join Products b on a.CategoryID = b.categoryid
left join OrderDetails c on b.ProductID = c.ProductID
left join orders d on c.OrderID=d.OrderID
where a.Active = '1'
group by CategoryName
order by total_sold ;

```

----1.d

---1.e.A

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains the following SQL code:

```

order by total_sold ;

----1.d

---1.e.A
SELECT D.CategoryName ,ROUND(SUM(Total_order_amount),2) AS TOTAL_REVENUE FROM Orders A
LEFT JOIN OrderDetails B ON A.OrderID =B.OrderID LEFT JOIN Products C ON B.ProductID =
C.ProductID
LEFT JOIN Category D ON C.CategoryID = D.CategoryID WHERE A.OrderDate BETWEEN '2021-07-01' AND '2021-10-01'
GROUP BY D.CategoryName HAVING D.CategoryName LIKE '%Beauty & Hygiene%' OR D.CategoryName LIKE '%BevErages%';

--- e.b
SELECT D.CategoryName ,ROUND(SUM(Total_order_amount),3) AS TOTAL_REVENUE FROM Customers1 CT
RIGHT JOIN Orders A ON CT.CustomerID = A.CustomerID LEFT JOIN OrderDetails B ON A.OrderID =B.OrderID
LEFT JOIN Products C ON B.ProductID = C.ProductID LEFT JOIN Category D ON C.CategoryID = D.CategoryID
WHERE A.OrderDate BETWEEN '2021-07-01' AND '2021-10-01' AND CT.DateEntered BETWEEN '2021-07-01' AND '2021-10-01'
GROUP BY D.CategoryName HAVING D.CategoryName LIKE '%Beauty & Hygiene%' OR D.CategoryName LIKE '%BevErages%';

---1.f

```

The Results pane shows the output of the query:

| CategoryName | TOTAL_REVENUE |
|------------------|---------------|
| Beauty & Hygiene | 45984521.88 |
| Beverages | 4666373.56 |

The status bar at the bottom indicates "Query executed successfully." and "2 rows".

```

SELECT D.CategoryName ,ROUND(SUM(Total_order_amount),2) AS TOTAL_REVENUE FROM Orders A
LEFT JOIN OrderDetails B ON A.OrderID =B.OrderID LEFT JOIN Products C ON B.ProductID =
C.ProductID
LEFT JOIN Category D ON C.CategoryID = D.CategoryID WHERE A.OrderDate BETWEEN '2021-07-
01' AND '2021-10-01'
GROUP BY D.CategoryName HAVING D.CategoryName LIKE '%Beauty & Hygiene%' OR
D.CategoryName LIKE '%BevErages%';

```

---1.e.b

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains a T-SQL query with several steps labeled with comments. The results pane at the bottom shows the output of the query, which is a table with two columns: CategoryName and TOTAL_REVENUE. The results are as follows:

| CategoryName | TOTAL_REVENUE |
|------------------|---------------|
| Beauty & Hygiene | 6122317.355 |
| Beverages | 417516.128 |

```
SELECT D.CategoryName ,ROUND(SUM(Total_order_amount),3) AS TOTAL_REVENUE FROM
Customers1 CT
RIGHT JOIN Orders A ON CT.CustomerID = A.CustomerID LEFT JOIN OrderDetails B ON
A.OrderID =B.OrderID
LEFT JOIN Products C ON B.ProductID = C.ProductID LEFT JOIN Category D ON C.CategoryID
= D.CategoryID
WHERE A.OrderDate BETWEEN '2021-07-01' AND '2021-10-01' AND CT.DateEntered BETWEEN
'2021-07-01' AND '2021-10-01'
GROUP BY D.CategoryName HAVING D.CategoryName LIKE '%Beauty & Hygiene%' OR
D.CategoryName LIKE '%BevErages%';
```

---1.f

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains a complex SQL query with several subqueries and joins. The results pane shows a table with 44 rows and 3 columns: CategoryID, revenue, and quarters. The status bar at the bottom indicates 'Query executed successfully' and '44 rows'.

```

SELECT D.CategoryName, ROUND(SUM(Total_order_amount), 3) AS TOTAL_REVENUE FROM Customers1 CT
RIGHT JOIN Orders A ON CT.CustomerID = A.CustomerID LEFT JOIN OrderDetails B ON A.OrderID = B.OrderID
LEFT JOIN Products C ON B.ProductID = C.ProductID LEFT JOIN Category D ON C.CategoryID = D.CategoryID
WHERE A.OrderDate BETWEEN '2021-07-01' AND '2021-10-01' AND CT.DateEntered BETWEEN '2021-07-01' AND '2021-10-01'
GROUP BY D.CategoryName HAVING D.CategoryName LIKE '%Beauty & Hygiene%' OR D.CategoryName LIKE '%Beverages%';

---1.f
select b.CategoryID, sum(a.Total_order_amount) as revenue,
DATEPART(QUARTER, a.OrderDate) as quarters from Orders as a,
(select * from Products) as b where YEAR(a.OrderDate)=2020 group by
b.CategoryID, DATEPART(QUARTER, a.OrderDate);

---1.g.a
select top 3 b.CompanyName as top_, avg(datediff(day, 'orderdate', 'deliverydate')) as avg_ from Orders as A
join Shippers as B on a.ShipperID=b.ShipperID
group by b.companyname
order by avg_;

---1.g.b

```

| CategoryID | revenue | quarters |
|------------|-------------------|----------|
| 5004 | 11756273287.1729 | 2 |
| 5010 | 1889149930.99785 | 4 |
| 5003 | 185555919.50195 | 1 |
| 5008 | 42452692077.6002 | 4 |
| 5011 | 1396214095.854 | 2 |
| 5007 | 3450134434.17877 | 3 |
| 5003 | 14443900915.2864 | 4 |
| 5010 | 210026893.432617 | 1 |
| 5006 | 7953763852.30371 | 2 |
| 5008 | 4720804487.52686 | 1 |
| 5001 | 14438503044.095 | 4 |
| 5009 | 3587977643.06488 | 3 |
| 5005 | 3292518451.16768 | 4 |
| 5008 | 18719897436.4155 | 2 |
| 5004 | 19014254402.2308 | 3 |
| 5007 | 510665389.460449 | 1 |
| 5003 | 47017613683.58177 | 3 |

```

select b.CategoryID, sum(a.Total_order_amount) as revenue,
DATEPART(QUARTER, a.OrderDate) as quarters from Orders as a,
(select * from Products) as b where YEAR(a.OrderDate)=2020 group by
b.CategoryID, DATEPART(QUARTER, a.OrderDate);

```

---1,g.a

```

select top 3 b.CompanyName as top_, avg(datediff(day, 'orderdate', 'deliverydate')) as
avg_ from Orders as A
join Shippers as B on a.ShipperID=b.ShipperID
group by b.companyname
order by avg_;

```

---1 g b

SQLQuery1.sql - L_69CUGG\mk804 (64)*

```

select top 3 b.CompanyName as top_, avg(datediff(day, 'orderdate', 'deliverydate')) as avg_ from Orders as A
join Shippers as B on a.ShipperID=b.ShipperID
group by b.CompanyName
order by avg_;

--1 g b

select top 3 companyname as top_, sum(quantity) as sale_vol_ from orders A right join OrderDetails b on
a.OrderID=b.OrderID left join Shippers c on a.ShipperID=c.ShipperID
group by CompanyName
order by sale_vol_;

--1 h a

select top 25 customerid as top_cust , count(orderid) as order_ from orders
where year(orderdate)='2021'
group by CustomerID
order by order_desc;

```

| top_ | sale_vol_ |
|---------------|-----------|
| 1 FedEx | 34013 |
| 2 Hapag Lloyd | 34712 |
| 3 DTDC | 35117 |

Query executed successfully.

```

select top 3 companyname as top_, sum(quantity) as sale_vol_ from orders A right join
OrderDetails b on
a.OrderID=b.OrderID left join Shippers c on a.ShipperID=c.ShipperID
group by CompanyName
order by sale_vol_;

```

--1 h a

SQLQuery1.sql - L_69CUGG\mk804 (64)*

```

select top 3 b.CompanyName as top_, avg(datediff(day, 'orderdate', 'deliverydate')) as avg_ from Orders as A
join Shippers as B on a.ShipperID=b.ShipperID
group by b.CompanyName
order by avg_;

--1 g b

select top 3 companyname as top_, sum(quantity) as sale_vol_ from orders A right join OrderDetails b on
a.OrderID=b.OrderID left join Shippers c on a.ShipperID=c.ShipperID
group by CompanyName
order by sale_vol_;

--1 h a

select top 25 customerid as top_cust , count(orderid) as order_ from orders
where year(orderdate)='2021'
group by CustomerID
order by order_desc;

```

| top_cust | order_ |
|----------|--------|
| 1 57334 | 19 |
| 2 57643 | 18 |
| 3 57408 | 17 |
| 4 57495 | 17 |
| 5 57383 | 17 |
| 6 57513 | 16 |
| 7 57470 | 16 |
| 8 57396 | 16 |
| 9 57456 | 16 |
| 10 57388 | 16 |
| 11 57392 | 16 |
| 12 57363 | 16 |
| 13 57487 | 15 |
| 14 57490 | 15 |
| 15 57516 | 15 |
| 16 57330 | 15 |
| 17 57393 | 15 |

Query executed successfully.

```

select top 25 customerid as top_cust , count(orderid) as order_ from orders
where year(orderdate)='2021'

```

```
group by CustomerID
order by order_ desc;
```

```
--- 1 h b
```

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains the following SQL code:

```
--1 h a
select top 25 customerid as top_cust , count(orderid) as order_ from orders
where year(orderdate)='2021'
group by CustomerID
order by order_ desc;

--- 1 h b
select top 25 customerid as top_cust , sum(total_order_amount) as total_amnt from orders
where year(orderdate)='2021'
group by customerid
order by total_amnt desc;

--- 1 i
select *, sum(total_order_amount) over ( partition by (select categoryid from category
--- 1 k
select *,
```

The Results pane shows the output of the second query (1 h b):

| | top_cust | total_amnt |
|----|----------|------------------|
| 1 | 57456 | 491256.040504658 |
| 2 | 57281 | 430801.5625 |
| 3 | 57495 | 419785.224121094 |
| 4 | 57315 | 416139.5 |
| 5 | 57396 | 409637.341796675 |
| 6 | 57370 | 392551.680664063 |
| 7 | 57495 | 384733.273925781 |
| 8 | 57392 | 37967.158203125 |
| 9 | 57200 | 374977.411132813 |
| 10 | 57429 | 372805.607421875 |
| 11 | 57102 | 364517.522285625 |
| 12 | 57320 | 364044.593007813 |
| 13 | 57538 | 362462.691601563 |
| 14 | 57477 | 350191.603515625 |
| 15 | 57390 | 346207.681640625 |
| 16 | 57278 | 344834.7265625 |
| 17 | 679460 | 343953.46115875 |

The status bar at the bottom indicates "Query executed successfully." and "25 rows".

```
select top 25 customerid as top_cust , sum(total_order_amount) as total_amnt from orders
where year(orderdate)='2021'
group by customerid
order by total_amnt desc;
```

---1.k

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains the following SQL code:

```
where year(orderdate)='2021'
group by customerId
order by total_amt desc;
---1.i
select *, sum(total_order_amount) over ( partition by (select categoryid from category
---1.j
select *, avg(total_order_amount) OVER(ORDER BY orderDate
ROWS BETWEEN 2 PRECEDING AND CURRENT ROW )
as moving_average
from orders;
---1.k
select*from (select c.Firstname, b.PaymentType, sum(total_order_amount) as total from orders as a join Payments as b on a.PaymentID=b.PaymentID
join customers1 as c on a.customerid=c.customerid
group by c.Firstname, b.PaymentType) as xyz
```

The Results pane displays a table with the following columns: OrderID, CustomerID, PaymentID, OrderDate, ShipperID, ShipDate, DeliveryDate, Total_order_amount, and moving_average. The table contains 17 rows of data.

```
select *,
avg(total_order_amount) OVER(ORDER BY orderDate
ROWS BETWEEN 2 PRECEDING AND CURRENT ROW )
as moving_average
from orders;
```

---1.L.

The screenshot shows the Microsoft SQL Server Management Studio interface. The query editor contains the following SQL code:

```
---1.k
select *,
avg(total_order_amount) OVER(ORDER BY orderDate
ROWS BETWEEN 2 PRECEDING AND CURRENT ROW )
as moving_average
from orders;
---1.L.
select*from (select c.Firstname, b.PaymentType, sum(total_order_amount) as total from orders as a join Payments as b on a.PaymentID=b.PaymentID
join customers1 as c on a.customerid=c.customerid
group by c.Firstname, b.PaymentType) as xyz
select (sum(total) for paymenttype in ([creditcard],[netbanking],[paypal],[pod],[wallet]) ) as prt;
```

The Results pane displays a table with the following columns: Firstname, creditcard, netbanking, paypal, pod, and wallet. The table contains 17 rows of data.

```

select*from (select c.FirstName, b.PaymentType, sum(total_order_amount) as total from
orders as a join Payments as b on a.PaymentID=b.PaymentID
join customers1 as c on a.customerid=c.customerid
group by c.FirstName, b.PaymentType) as xyz
pivot (sum(total) for paymenttype in
([creditcard],[netbanking],[paypal],[pod],[wallet]) )as pvt;

```

Q.no.2 answer:-

A GROUP BY normally reduces the number of rows returned by rolling them up and calculating averages or sums for each row.

PARTITION BY does not affect the number of rows returned, but it changes how a window function's result is calculated.

Q.no.3 answer:-

Temporary tables are just the tables in tempdb. Views are stored queries for existing data in existing tables.

Temporary table needs to be populated first with data, and population is the main performance-concerned issue.

So the data in views already exists and so views are faster than temporary table.

Q.NO 4 answer :-

Views being a physical object on database (but does not store data physically) and can be used on multiple queries, thus provide flexibility and centralized approach.

CTE, on the other hand are temporary and will be created when they are used; that's why they are called as inline view

Q.No 5 answer :-

The row_number gives continuous numbers, while rank and dense_rank give the same rank for duplicates, but the next number in rank is as per continuous order so you will see a jump but in dense_rank doesn't have any gap in rankings.

```

Eg rank      : 1,2,2,4,5,5,5,8,9
Dense_r     : 1,2,2,3,4,4,5,6,7,7,7,8,9
Row         : 1,2,3,4,5,6,7,8,9

```

Q.no.06 answer

1. ans – A column can hold a specific datatype but in col1 it has int and string data type which is not possible in a single column.

So, that this kind of table can't be exist.

2. a. 'SELECT COUNT (*) AS ENTRIES FROM TABLE;' :- count the records for each column present in the table including commons and output of count of col1 as entries

b. 'SELECT COUNT(COL1) AS ENTRIES FROM TABLE;' :- count the specific column col1 including commons from the table and output as entries.

c. 'SELECT COUNT (DISTINCT COL1) AS ENTRIES FROM TABLE;'- will count col1 but will not includes the common records from the table;