Customer Segmentation (RFM Analysis)

- Can you provide a query to segment customers based on their recency, frequency, and monetary value of purchases?

```
CREATE VIEW RFM AS ( SELECT DISTINCT customer_id, TransactionID,
COUNT(transactionID) OVER(PARTITION BY customer_id) AS "Frequency"
FROM transaction
ORDER BY Frequency DESC);

SELECT RFM.customer_id, PurchaseDate, Frequency, DATEDIFF(NOW(), purchaseDate) AS "Recency",
SUM(PurchaseAmount) AS "Monetary" FROM transaction t JOIN
RFM ON t.TransactionID = RFM.transactionID
GROUP BY RFM.customer_id, PurchaseDate, Frequency, Recency;
```

Output-

customer_id	PurchaseDate	Frequency	Recency	Monetary
CUSTID3523	2023-06-20 00:00:00	16	295	294
CUSTID7215	2019-09-03 00:00:00	12	1681	125
CUSTID9019	2017-08-26 00:00:00	13	2419	828
CUSTID1380	2023-01-25 00:00:00	12	441	566
CUSTID9195	2023-04-19 00:00:00	11	357	790
CUSTID2473	2017-10-17 00:00:00	7	2367	991
CUSTID4078	2020-05-17 00:00:00	12	1424	471
CUSTID9794	2021-06-16 00:00:00	10	1029	866

- Can you segment customers based on their RFM scores? How would you determine the appropriate thresholds for segmentation?

```
CREATE VIEW RFM AS (SELECT
    DISTINCT customer_id,
    MAX(PurchaseDate) AS PurchaseDate,
    SUM(PurchaseAmount) AS TotalPurchase,
    COUNT(TransactionID) AS Frequency,
    DATEDIFF(NOW(), MAX(purchaseDate)) AS Recency,
    SUM(PurchaseAmount) AS Monetary
FROM
    transaction
GROUP BY
    customer_id);
SELECT
    customer_id, CASE
        WHEN Freq_seg + Rec_seg + Mon_seg > 10 THEN 'High Value'
        WHEN Freq_seg + Rec_seg + Mon_seg > 5 THEN 'Medium Value'
        ELSE 'Low Value' END AS Segmentation FROM (
        SELECT customer_id,
            NTILE(5) OVER (ORDER BY Frequency) AS Freq_seg,
            NTILE(5) OVER (ORDER BY Recency) AS Rec_seg,
            NTILE(5) OVER (ORDER BY Monetary) AS Mon_seg
        FROM RFM) AS quartiles GROUP BY customer_id
ORDER BY customer_id;
```

Output-

customer_id	Segmentation
CUSTID122	Medium Value
CUSTID1220	Medium Value
CUSTID1221	Low Value
CUSTID1222	Medium Value
CUSTID1223	Medium Value
CUSTID1224	Medium Value
CUSTID1225	High Value
CUSTID1226	Medium Value
CUSTID1227	Medium Value

- How would you analyse purchasing behaviour differences between different demographic groups, such as age, profession, family size or gender?

```
SELECT CASE
WHEN age BETWEEN 18 AND 30 THEN "Adults"
WHEN age BETWEEN 31 AND 45 THEN "Old Adults"
WHEN age BETWEEN 46 and 60 THEN "Old"
WHEN age > 60 THEN "Senior" ELSE ""
END AS "age_group",
gender, proffession, location,
sum(purchaseAmount) AS "total_purchase" FROM transaction t JOIN
customers c ON t.customer_id = c.customer_id
GROUP BY 1,2,3,4
ORDER BY total_purchase DESC, Proffession,total_purchase, location;
```

Output-

age_group	gender	proffession	location	total_purchase
Adults	M	Information Technology	South Jonathanstad	534237
Old Adults	F	Information Technology	Lake Sarahburgh	506349
Old	F	Healthcare	Smithburgh	506322
Adults	M	Information Technology	Sherylmouth	479690
Old Adults	F	Information Technology	North Michael	471275
Adults	M	Information Technology	Michaelburgh	464623
Old	M	Information Technology	Williamsfort	459246
Old	F	Information Technology	South Bryanmouth	434426

proffession	total_purchase
Information Technology	60004123
Finance and Banking	38886496
Healthcare	38558845
Education	19136991
Manufacturing and Engineering	17979395

location	total_purchase
Sherylmouth	6398036
South Jonathanstad	6322821
Johnstad	6271409
Benjaminside	6216330
Williamsfort	6189383

Age_group	total_purchase
45+	72358273
31-45	55370720
18-30	43046097

gender	total_purchase
M	86576732
F	85566632
X	2422486

- What are the characteristics of customers who are likely to churn, and how can we segment them?

```
SELECT t.customer_id, segment, DATEDIFF(NOW(), purchasedate) AS "recency",
ROUND(AVG(customer_service_rating),1) AS "avg_rating"
FROM transaction t
JOIN feedbacks f ON t.customer_id = f.customer_id
JOIN segments s ON t.customer_id = s.customer_id
GROUP BY 1,2,3
HAVING recency > 365 AND avg_rating < 3.5 AND segment = "low_value"
;
```

Output-

customer_id	segment	recency	avg_rating
CUSTID9034	low_value	2108	2.9
CUSTID9034	low_value	2089	2.9
CUSTID9034	low_value	1819	2.9
CUSTID9034	low_value	1799	2.9
CUSTID9034	low_value	1756	2.9
CUSTID9034	low_value	855	2.9

- Can you identify any seasonal trends or patterns in customer purchasing behaviour?

```
CREATE TEMPORARY TABLE segments (SELECT customer_id,
CASE WHEN freq_seg + rec_seg + mon_seg > 10 THEN "high_value"
WHEN freq_seg + rec_seg + mon_seg > 5 THEN "medium_value"
ELSE "low_value"
END AS "segment"
FROM (SELECT customer_id,
NTILE(5) OVER(ORDER BY frequency) AS "freq_Seg",
NTILE(5) OVER(ORDER BY recency) AS "rec_seg",
NTILE(5) OVER(ORDER BY monetary) AS "mon_seg"
FROM rfm) AS quartile
GROUP BY customer_id ORDER BY customer_id);

SELECT segment, MONTHNAME(purchasedate) AS "month",
SUM(purchaseamount) AS "total_purchaSE"
FROM transaction t JOIN segments s
ON t.customer_id = s.customer_id
GROUP BY segment, month, MONTH(PurchaseDate)
ORDER BY MONTH(PurchaseDate),MONTHNAME(PurchaseDate) ;
```

Output-

	segment	month	total_purchaSE
▶	low_value	January	926937
	medium_value	January	7299642
	high_value	January	6546107
	low_value	February	1075612
	high_value	February	5619785
	medium_value	February	6762654
	low_value	March	1206843
	medium_value	March	7489774
	high_value	March	6194775

```sql
SELECT MONTHNAME(purchasedate) AS month,
SUM(PurchaseAmount) AS revenue
FROM transaction
GROUP BY MONTH(purchasedate), MONTHNAME(purchasedate)
ORDER BY MONTH(purchasedate);

SELECT QUARTER(purchasedate) AS "quarter",
sum(PurchaseAmount) AS "revenue" FROM transaction
GROUP BY 1
ORDER BY quarter DESC;

SELECT YEAR(purchasedate) AS "year",
sum(PurchaseAmount) AS "revenue" FROM transaction
GROUP BY 1
ORDER BY year DESC;
```

	month	revenue
▶	January	14772686
	February	13458051
	March	14891392
	April	14492979
	May	14671450
	June	14265659
	July	14715054
	August	14879060
	September	14398852
	October	14849506
	November	14274382
	December	14896779

	quarter	revenue
▶	4	44020667
	3	43992966
	2	43430088
	1	43122129

	year	revenue
▶	2024	5833870
	2023	25033217
	2022	24781808
	2021	24822830
	2020	25183161
	2019	24898277
	2018	24880771
	2017	19131916