

Customer Segmentation

IMPLEMENTING RFM STRATEGIES USING SQL





CUSTOMER SEGMENTATION

- Customer segmentation involves categorizing customers into specific groups based on factors such as age, shopping behaviors, preferences, and geographic location.
- This strategic method empowers businesses to gain a profound insight into their customer demographics and tailor products or services to cater to the distinct preferences of each segment.
- By implementing personalized strategies for individual segments, businesses can enhance customer satisfaction, cultivate loyalty, and boost revenue streams.

RFM (RECENCY , FREQUENCY, MONETARY)

Recency (R):

- The timing of the customer's last purchase plays a vital role. Customers who have recently made a purchase are more inclined to make another purchase as the product is still fresh in their minds. Recency is usually gauged in days, but for some products, it could be measured in years, weeks, or even hours.

Frequency (F):

- How frequently did the customer make purchases within a specific timeframe? Customers who have previously made purchases are more likely to make repeat purchases. First-time customers can also be targeted for follow-up marketing to encourage them to become regular customers.

Monetary Value (M):

- The total amount a customer spent within a specific period is crucial. Customers who spend more are likely to continue spending in the future and are of high value to a business.

WHY RFM ANALYSIS?



Simple
Segmentation



Identify
Valuable
Customers



Personalized
Marketing



Improve ROI



Strategic
Decision
Making

RFM ANALYSIS PROCESS

Data Collection: Gather customer transaction data

RFM Calculation: Calculate Recency, Frequency, and Monetary values for each customer based on their transaction history.

RFM Segmentation: Segment customers into RFM groups (e.g., Champion, Loyal, Potential Loyalist) using predefined criteria or percentiles.

Analysis and Insights: Analyze each RFM segment's characteristics, behavior, and value to the business.



TABLE OVERVIEW

InvoiceNo ▾ //	StockCode ▾ //	Description ▾ //	Quantity ▾ //	InvoiceDate ▾ //	UnitPrice ▾ //	CustomerID ▾ //	Country ▾
571035	21238	RED RETROSPOT CUP	8	2011-10-13 12:50:00 UTC	0.85	12446.0	RSA
571035	21243	PINK POLKA DOT PLATE	8	2011-10-13 12:50:00 UTC	1.69	12446.0	RSA
571035	23240	SET OF 4 KNICK KNACK TINS ...	6	2011-10-13 12:50:00 UTC	4.15	12446.0	RSA
571035	23209	LUNCH BAG VINTAGE DOILY	10	2011-10-13 12:50:00 UTC	1.65	12446.0	RSA
571035	23201	JUMBO BAG ALPHABET	10	2011-10-13 12:50:00 UTC	2.08	12446.0	RSA
571035	23205	CHARLOTTE BAG VINTAGE AL...	10	2011-10-13 12:50:00 UTC	0.85	12446.0	RSA
571035	21936	RED RETROSPOT PICNIC BAG	5	2011-10-13 12:50:00 UTC	2.95	12446.0	RSA
571035	22620	4 TRADITIONAL SPINNING TO...	12	2011-10-13 12:50:00 UTC	1.45	12446.0	RSA
571035	22619	SET OF 6 SOLDIER SKITTLES	4	2011-10-13 12:50:00 UTC	3.75	12446.0	RSA
571035	21889	WOODEN BOX OF DOMINOES	12	2011-10-13 12:50:00 UTC	1.25	12446.0	RSA

CREATING RFM METRICS



```
WITH rfm AS (
    SELECT
        CustomerID,
        DATE(MIN(InvoiceDate)) AS first_order_date,
        DATE(MAX(InvoiceDate)) AS last_order_date,
        COUNT(DISTINCT InvoiceNo) AS Total_orders,
        ROUND(SUM(Quantity * UnitPrice), 2) AS Total_amount
    FROM `sales.sales`
    GROUP BY CustomerID
)
```

This part of the query calculates RFM metrics for each customer in the dataset.

First_order_date: The earliest order date for each customer

Last_order_date: The latest order date for each customer

Total_orders: The total number of orders placed by each customer

Total_amount: Total monetary amount spent by each customer

CREATING RFM METRICS



```
SELECT  
    distinct CustomerID,  
    DATE_DIFF(MAX_date, tbl1.last_order_date, DAY) AS Recency,  
    ROUND((tbl1.Total_orders / tbl1.no_of_months), 2) AS Frequency,  
    tbl1.Total_amount AS Monetary  
FROM (  
    SELECT  
        *,  
        MAX(rfm.last_order_date) OVER() + 1 AS max_date,  
        DATE_DIFF(rfm.last_order_date, rfm.first_order_date, MONTH) + 1 AS no_of_months  
    FROM rfm  
    JOIN `sales.sales` AS s USING(CustomerID)
```

Recency: Calculates the number of days between the last order date and the maximum order date in the dataset.

Frequency: Calculates the average number of order per month by dividing ‘Total_order’ by the number of months between the first and last order dates

Monetary: Represents the total amount spent by each customer

CREATING RFM SCORES

```
SELECT  
CustomerID,  
Recency,  
Frequency,  
Monetary,  
NTILE(5) OVER (ORDER BY Recency DESC) AS R_score,  
NTILE(5) OVER (ORDER BY Frequency) AS F_score,  
NTILE(5) OVER (ORDER BY Monetary DESC) AS M_score  
FROM rfm_analysis;
```

Dividing customers into 5 equal-sized groups (quintiles) based on their Recency, Frequency, and Monetary Values using NTILE(5)

- **NTILE(5):** Splits customers into 5 equal-sized groups based on their Recency, Frequency, and Monetary Values.
- **Order by:** Defines the sequence of customers within each quintile, for example, arranging by recency in descending order (placing the most recent customers at the top of the list).

CONCATENATING RFM SCORES



```
SELECT  
    CustomerID,  
    Recency,  
    Frequency,  
    Monetary,  
    CONCAT(R_score, '-', F_score, '-', M_score) AS RFM_STRING  
FROM (  
    -- Previous Query Here  
) AS rfm_segmented;
```

This part of the query concatenates the individual RFM scores (R_score, F_score, M_score) into a single string format (RFM_STRING).

CONCAT: combines the scores with hyphens (-) as separators to create the RFM_STRING.

ASSIGNING RFM SEGMENTS

```
SELECT
    distinct CustomerID,
    Recency,
    Frequency,
    Monetary,
    f.RFM_STRING,
    CASE
        WHEN RFM_STRING IN ('5-5-5', '5-5-4', '5-4-4', '5-4-5', '4-5-4', '4-5-5', '4-4-5') THEN
            'Champion'
        WHEN RFM_STRING IN ('5-4-3', '4-4-4', '4-3-5', '3-5-5', '3-5-4', '3-4-5', '3-4-4', '3-3-5')
        THEN 'Loyal'
        WHEN RFM_STRING IN ('5-5-3', '5-5-1', '5-5-2', '5-4-1', '5-4-2', '5-3-3', '5-3-2', '5-3-1',
                            '4-5-2', '4-5-1', '4-4-2', '4-4-1', '4-3-1', '4-5-3', '4-3-3', '4-3-2',
                            '4-2-3', '3-5-3', '3-5-2', '3-5-1', '3-4-2', '3-4-1', '3-3-3', '3-2-3')
        THEN 'Potential Loyalist'
        WHEN RFM_STRING IN ('5-1-2', '5-1-1', '4-2-2', '4-2-1', '4-1-2', '4-1-1', '3-1-1')
        THEN 'New Customer'
        WHEN RFM_STRING IN ('5-2-5', '5-2-4', '5-2-3', '5-2-2', '5-2-1', '5-1-5', '5-1-4', '5-1-3',
                            '4-2-5', '4-2-4', '4-1-3', '4-1-4', '4-1-5', '3-1-5', '3-1-4', '3-1-3')
        THEN 'Promising'
        WHEN RFM_STRING IN ('5-3-5', '5-3-4', '4-4-3', '4-3-4', '3-4-3', '3-3-4', '3-2-5', '3-2-4')
        THEN 'Needs Attention'
        WHEN RFM_STRING IN ('3-3-1', '3-2-1', '3-1-2', '2-2-1', '2-1-3', '2-3-1', '2-4-1', '2-5-1')
        THEN 'About To Sleep'
        WHEN RFM_STRING IN ('2-5-5', '2-5-4', '2-4-5', '2-4-4', '2-5-3', '2-5-2', '2-4-3', '2-4-2',
                            '2-3-5', '2-3-4', '2-2-5', '2-2-4', '1-5-3', '1-5-2', '1-4-5', '1-4-3',
                            '1-4-2', '1-3-5', '1-3-4', '1-3-3', '1-2-5', '1-2-4') THEN 'At Risk'
        WHEN RFM_STRING IN ('1-5-5', '1-5-4', '1-4-4', '2-1-4', '2-1-5', '1-1-5', '1-1-4', '1-1-3')
        THEN 'Cannot Lose Them'
        WHEN RFM_STRING IN ('3-3-2', '3-2-2', '2-3-3', '2-3-2', '2-2-3', '2-2-2', '1-3-2', '1-2-3',
                            '1-2-2', '2-1-2', '2-1-1') THEN 'Hibernating Customer'
        WHEN RFM_STRING IN ('1-1-1', '1-1-2', '1-2-1', '1-3-1', '1-4-1', '1-5-1') THEN 'Lost Customer'
    END AS RFM_SEGMENT
FROM (
    -- Previous Query Here
) AS rfm_segmented;
```

This final part of the query assigns RFM segments to customers based on their RFM_STRING.

The CASE statement checks the RFM_STRING values and assigns corresponding segment names ('Champion', 'Loyal', etc.).

You can add more WHEN conditions for other RFM segments as needed.

The ELSE part specifies a default segment ('Other Segment') for RFM_STRING values that do not match any specified conditions.

COMPREHENSIVE SQL QUERY FOR RFM SEGMENTATION AND CUSTOMER ANALYSIS

```
WITH rfm AS (
    SELECT
        CustomerID,
        DATE(MIN(InvoiceDate)) AS first_order_date,
        DATE(MAX(InvoiceDate)) AS last_order_date,
        COUNT(DISTINCT InvoiceNo) AS Total_orders,
        ROUND(SUM(Quantity * UnitPrice), 2) AS Total_amount
    FROM `sales.sales`
    GROUP BY CustomerID
),
rfm_segment AS (
    SELECT
        distinct CustomerID,
        DATE_DIFF(MAX_date, tbll.last_order_date, DAY) AS Recency,
        ROUND((tbl.Total_orders / tbll.no_of_months), 2) AS Frequency,
        tbll.Total_amount AS Monetary
    FROM (
        SELECT
            *,
            MAX(rfm.last_order_date) OVER() + 1 AS max_date,
            DATE_DIFF(rfm.last_order_date, rfm.first_order_date, MONTH) + 1 AS no_of_months
        FROM rfm
        JOIN `sales.sales` AS s USING(CustomerID)
    ) AS tbll
),
final_query AS (
    SELECT
        distinct CustomerID,
        CONCAT(R_score,'-', F_score,'-', M_score) AS RFM_STRING
    FROM (
        SELECT
            CustomerID,
            NTILE(5) OVER (ORDER BY Recency DESC) AS R_score,
            NTILE(5) OVER (ORDER BY Frequency) AS F_score,
            NTILE(5) OVER (ORDER BY Monetary) AS M_score
        FROM rfm_segment
    ) AS subquery
)
```

```
SELECT
    distinct CustomerID,
    Recency,
    Frequency,
    Monetary,
    f.RFM_STRING,
    CASE
        WHEN RFM_STRING IN ('5-5-5', '5-5-4', '5-4-4', '5-4-5', '4-5-4', '4-5-5', '4-4-5') THEN
            'Champion'
        WHEN RFM_STRING IN ('5-4-3', '4-4-4', '4-3-5', '3-5-5', '3-5-4', '3-4-5', '3-4-4', '3-3-5') THEN
            'Loyal'
        WHEN RFM_STRING IN ('5-5-3', '5-5-1', '5-5-2', '5-4-1', '5-4-2', '5-3-3', '5-3-2', '5-3-1',
                            '4-5-2', '4-5-1', '4-4-2', '4-4-1', '4-3-1', '4-5-3', '4-3-3', '4-3-2',
                            '4-2-3', '3-5-3', '3-5-2', '3-5-1', '3-4-2', '3-4-1', '3-3-3', '3-2-3') THEN
            'Potential Loyalist'
        WHEN RFM_STRING IN ('5-1-2', '5-1-1', '4-2-2', '4-2-1', '4-1-2', '4-1-1', '3-1-1') THEN
            'New Customer'
        WHEN RFM_STRING IN ('5-2-5', '5-2-4', '5-2-3', '5-2-2', '5-2-1', '5-1-5', '5-1-4', '5-1-3',
                            '4-2-5', '4-2-4', '4-1-3', '4-1-4', '4-1-5', '3-1-5', '3-1-4', '3-1-3') THEN
            'Promising'
        WHEN RFM_STRING IN ('5-3-5', '5-3-4', '4-4-3', '4-3-4', '3-4-3', '3-3-4', '3-2-5', '3-2-4') THEN
            'Needs Attention'
        WHEN RFM_STRING IN ('3-3-1', '3-2-1', '3-1-2', '2-2-1', '2-1-3', '2-3-1', '2-4-1', '2-5-1') THEN
            'About To Sleep'
        WHEN RFM_STRING IN ('2-5-5', '2-5-4', '2-4-5', '2-4-4', '2-5-3', '2-5-2', '2-4-3', '2-4-2',
                            '2-3-5', '2-3-4', '2-2-5', '2-2-4', '1-5-3', '1-5-2', '1-4-5', '1-4-3',
                            '1-4-2', '1-3-5', '1-3-4', '1-3-3', '1-2-5', '1-2-4') THEN
            'At Risk'
        WHEN RFM_STRING IN ('1-5-5', '1-5-4', '1-4-4', '2-1-4', '2-1-5', '1-1-5', '1-1-4', '1-1-3') THEN
            'Cannot Lose Them'
        WHEN RFM_STRING IN ('3-3-2', '3-2-2', '2-3-3', '2-3-2', '2-2-3', '2-2-2', '1-3-2', '1-2-3',
                            '1-2-2', '2-1-2', '2-1-1') THEN
            'Hibernating Customer'
        WHEN RFM_STRING IN ('1-1-1', '1-1-2', '1-2-1', '1-3-1', '1-4-1', '1-5-1') THEN
            'Lost Customer'
    END AS RFM_SEGMENT
)
FROM final_query AS f
JOIN rfm_segment USING (CustomerID)
ORDER BY CustomerID
```

QUERY OUTPUT

Row	CustomerID	Recency	Frequency	Monetary	RFM_STRING	RFM_SEGMENT
1	12346.0	326	1.0	77183.6	1-4-5	At Risk
2	12347.0	3	0.54	4078.95	5-2-5	Promising
3	12348.0	76	0.4	1437.24	2-1-4	Cannot Lose Them
4	12349.0	19	1.0	1287.15	4-4-4	Loyal
5	12350.0	311	1.0	294.4	1-3-2	Hibernating Customer
6	12352.0	37	0.7	1147.44	3-2-4	Needs Attention
7	12353.0	205	1.0	29.3	1-4-1	Lost Customer
8	12354.0	233	1.0	925.95	1-3-4	At Risk
9	12355.0	215	1.0	414.0	1-4-2	At Risk
10	12356.0	246	0.5	1911.4	1-2-5	At Risk
11	12357.0	34	1.0	5291.27	3-5-5	Loyal
12	12358.0	2	0.33	908.16	5-1-4	Promising
13	12359.0	58	0.4	3642.33	3-1-5	Promising
14	12360.0	53	0.5	1914.26	3-2-5	Needs Attention
15	12361.0	288	1.0	174.9	1-4-1	Lost Customer

STRATEGIC INSIGHTS FROM RFM ANALYSIS

- **Customer Engagement Trends**

- Active and engaged customers are often classified as "Champion" or "Loyal," with higher Recency and Frequency scores.

- **Revenue Contribution**

- RFM Monetary values assist in identifying segments that contribute the most revenue, guiding strategies for retaining high-value customers.

- **Identifying New Opportunities**

- Growth opportunities are present within the "New Customer" segment, which can be cultivated into loyal and high-value customers.

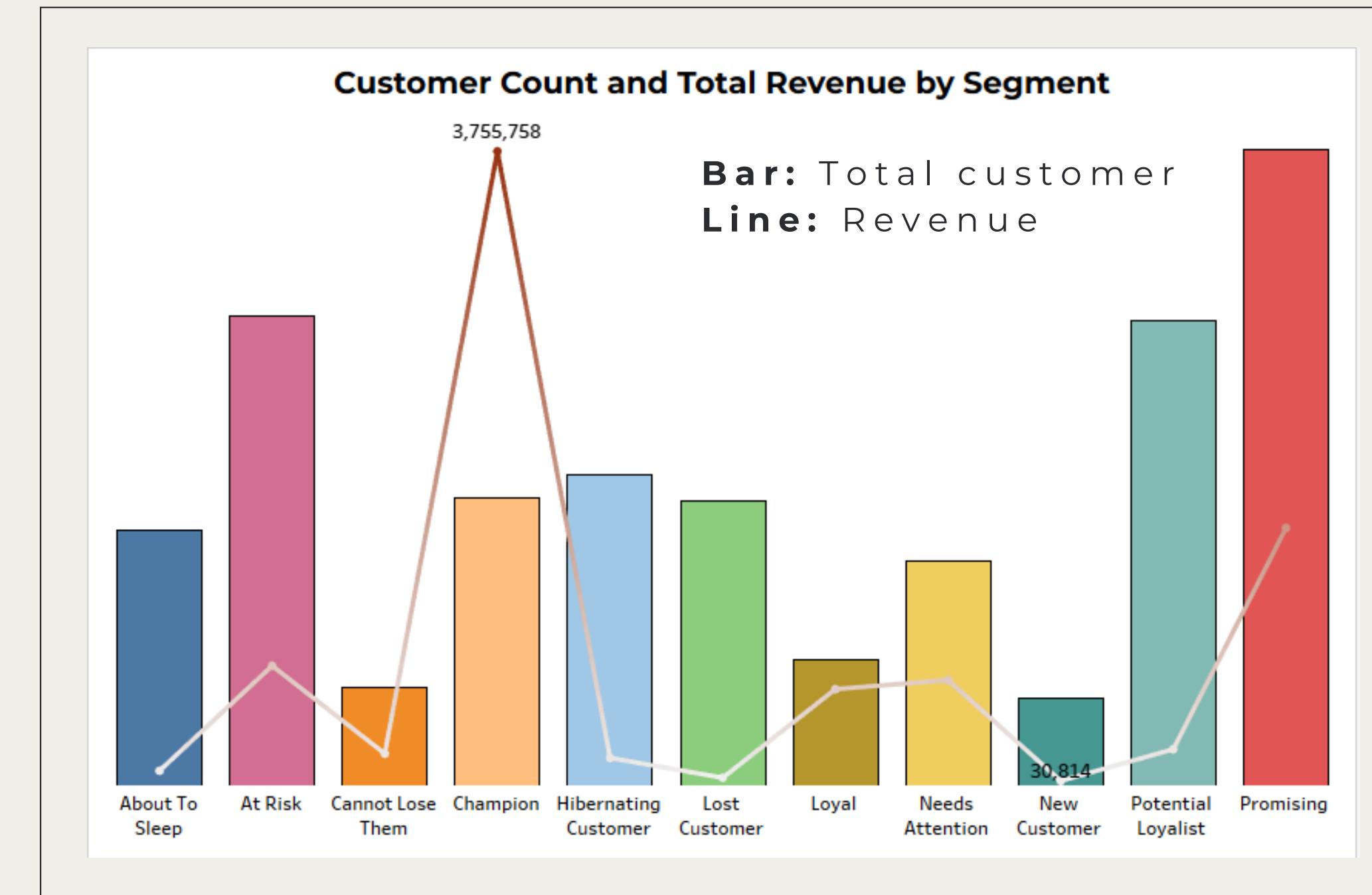
- **Risk Mitigation**

- Implement proactive measures for the "At Risk" and "Lost Customers" segments to reduce churn and enhance retention rates.



CUSTOMER COUNT AND REVENUE INSIGHTS

- **"At Risk and Potential"**: New customers with potential, but they might leave. Engage them to build loyalty.
- **"Champion" and "Promising"**: Lots of customers who spend well. Offer personalized services to keep them loyal and spending more.
- **"New Customer" and "Lost Customer"**: Work on keeping new customers engaged, and try to win back lost customers.



Thank you!



FOLLOW [WWW.LINKEDIN.COM/IN/PRADEEPARAJA-R](https://www.linkedin.com/in/pradeeparaja-r)

