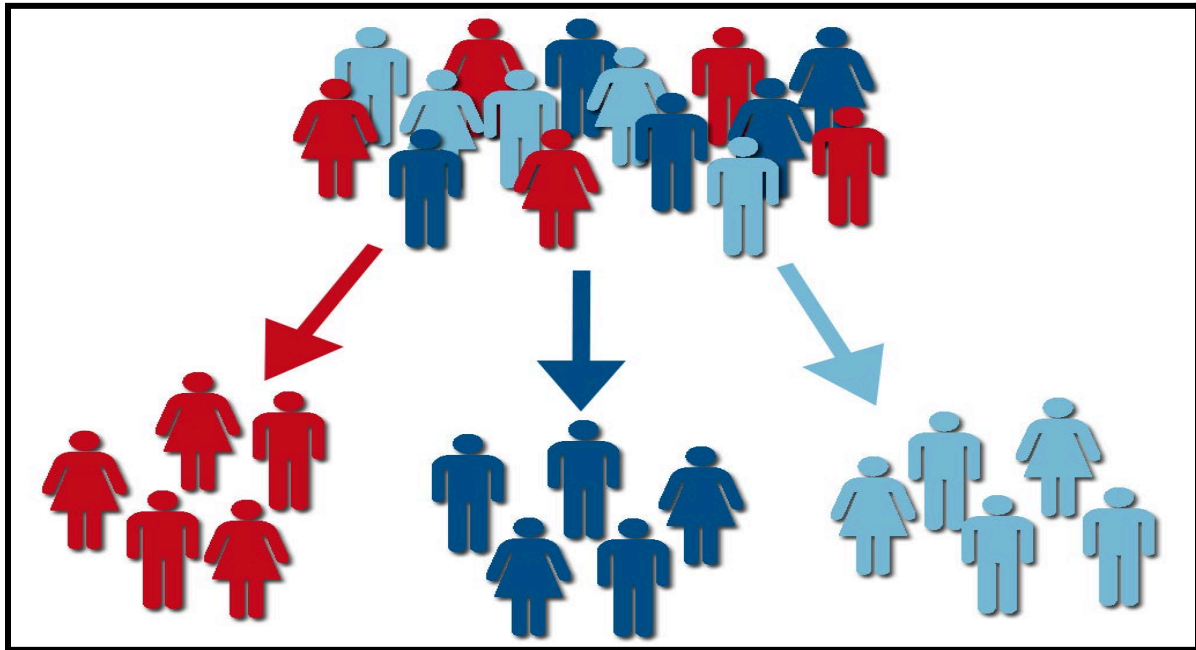


Customer segmentation using SQL & Tableau



Problem Statement:

We will be performing a RFM analysis for a chain of retail stores that sells a lot of different items and categories.

The stores need to adjust their marketing budget and have better targeting of customers so they need to know which customers to focus on and how important they are for the business.

RFM segmentation will answer key questions for your business:

- Who are my best customers?
- Which customers are at the verge of churning?
- Who has the potential to be converted into more profitable customers?
- Who can you view as lost customers?
- Which customers must you retain?
- Who are your loyal customers?
- Which group of customers is most likely to respond to your current campaign?

How to do RFM Analysis?

- RFM values - to calculate the RFM score you will first need to identify each customer's RFM values.
- Recency (R) – period since last purchase.
Frequency (F) – number of transactions.
Monetary (M) – total money spent, often called Customer Lifetime Value (CLV)
- 5 bands for each of the R, F and M-values.
- Final RFM score is calculated by combining individual RFM scores.

Based on customer segment, targeted action can be taken, Eg: Marketing

Customer Segment	Activity	Actionable Tip
Champions	Bought recently, buy often and spend the most!	Reward them. Can be early adopters for new products. Will promote your brand.
Loyal Customers	Spend good money with us often. Responsive to promotions.	Upsell higher value products. Ask for reviews. Engage them.
Potential Loyalist	Recent customers, but spent a good amount and bought more than once.	Offer membership / loyalty program, recommend other products.
Recent Customers	Bought most recently, but not often.	Provide on-boarding support, give them early success, start building relationship.
Promising	Recent shoppers, but haven't spent much.	Create brand awareness, offer free trials
Customers Needing Attention	Above average recency, frequency and monetary values. May not have bought very recently though.	Make limited time offers, Recommend based on past purchases. Reactivate them.
About To Sleep	Below average recency, frequency and monetary values. Will lose them if not reactivated.	Share valuable resources, recommend popular products / renewals at discount, reconnect with them.
At Risk	Spent big money and purchased often. But long time ago. Need to bring them back!	Send personalized emails to reconnect, offer renewals, provide helpful resources.
Can't Lose Them	Made biggest purchases, and often. But haven't returned for a long time.	Win them back via renewals or newer products, don't lose them to competition, talk to them.
Hibernating	Last purchase was long back, low spenders and low number of orders.	Offer other relevant products and special discounts. Recreate brand value.
Lost	Lowest recency, frequency and monetary scores.	Revive interest with reach out campaign, ignore otherwise.

[Data & Marketing association](#)

Data

- **InvoiceNo:** Invoice number. Unique 6-digit to each transaction. If this code starts with letter 'c', it indicates a cancellation.
- **StockCode:** Product (item) code. unique 5-digit for distinct product.
- **Description:** Product (item) name.
- **Quantity:** The quantities of each product (item) per transaction. Numeric.
- **InvoiceDate:** Invoice Date and time. Numeric, the day and time when each transaction was generated.
- **UnitPrice:** Unit price. Numeric, Product price per unit.
- **CustomerID:** Customer number. unique 5-digit assigned to each customer.
- **Country:** Name of the country where each customer resides.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HE	6	2010-12-01 8:26:00	2.55	17850	United Kingdom
536365	71053	WHITE METAL LAN	6	2010-12-01 8:26:00	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEA	8	2010-12-01 8:26:00	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FL	6	2010-12-01 8:26:00	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTT	6	2010-12-01 8:26:00	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA N	2	2010-12-01 8:26:00	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROS	6	2010-12-01 8:26:00	4.25	17850	United Kingdom
536366	22633	HAND WARMER UNI	6	2010-12-01 8:28:00	1.85	17850	United Kingdom
536366	22632	HAND WARMER REC	6	2010-12-01 8:28:00	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUR	32	2010-12-01 8:34:00	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOU	6	2010-12-01 8:34:00	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOU	6	2010-12-01 8:34:00	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCE	8	2010-12-01 8:34:00	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED ML	6	2010-12-01 8:34:00	1.65	13047	United Kingdom

What we need?

- How to set up BigQuery: [link](#)
 - Dataset link: [sales.csv](#)
-

Work flow:

The RFM Segmentation can be executed using these five steps:

1. Data processing
2. Compute for recency, frequency, and monetary values per customer
3. Determine quantiles for each RFM metric
4. Assign scores for each RFM metric
5. Define the RFM segments using the scores in step 4

Data Processing

Adding the data to BigQuery:

Project name: Customer_segmentation, Dataset name: Retail, table name: Sales.

- find the total cost for that product i.e., quantity * unit price.

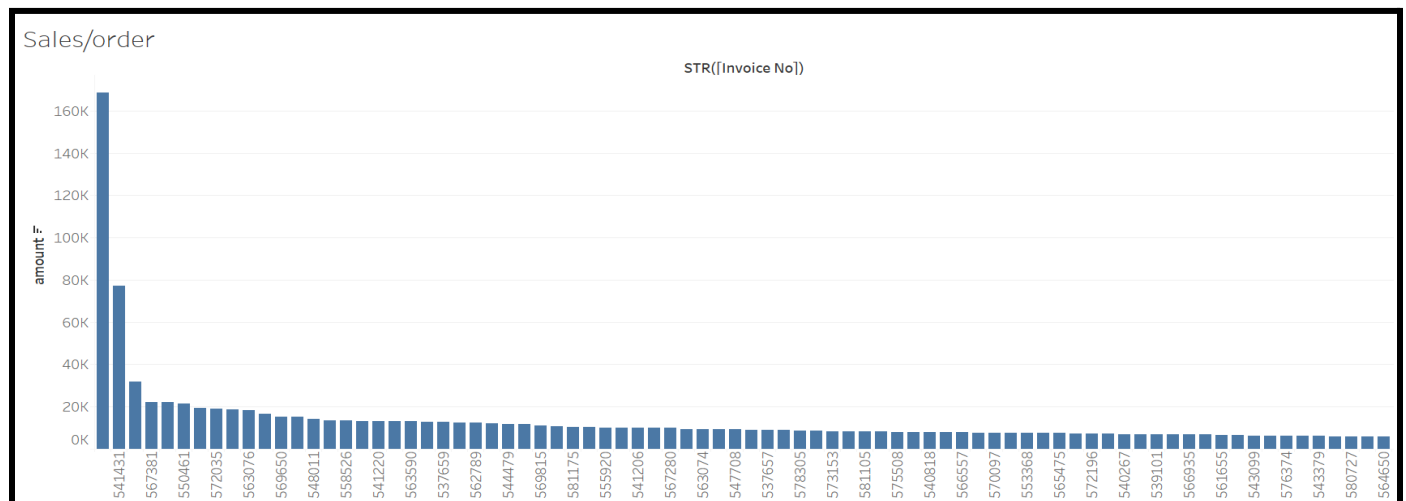
```
select InvoiceNo, (Quantity*UnitPrice)amount from `retail.sales`
```

Row	InvoiceNo	amount
1	571035	6.8
2	571035	13.52
3	571035	24.900000000000...
4	571035	16.5
5	571035	20.8
6	571035	8.5
7	571035	14.75

- Find total amount spent per order

```
with bills as(  
select InvoiceNo, (Quantity*UnitPrice)amount  
from `retail.sales`)  
select InvoiceNo, sum(amount) total  
from bills  
group by InvoiceNo  
Order by total desc
```

InvoiceNo	total
581483	168469.6
541431	77183.6
567423	31698.15999999...
567381	22104.8
556917	22007.33000000...
550461	21535.9
563614	19487.23999999...
572035	19023.11999999...



Save this data as a `bill` table

Compute for recency, frequency and monetary values per customer :

We will join the `bill` table that we saved with the `sales` table and add the total cost on the customer level for monetary value. - total sale per customer

```
select s.CustomerID, date(max(s.InvoiceDate)) as last_purchase_date,
date(min(s.InvoiceDate)) as first_purchase_date, count(distinct
s.InvoiceNo)no_of_purchases, sum(b.total) as monetary
from `retail.sales` s
join `retail.bill` b
on s.InvoiceNo = b.InvoiceNo
group by s.CustomerID
Order by monetary desc
```

CustomerID	last_purchase_date	first_purchase_date	no_of_purchases	monetary
14096.0	2011-12-05	2011-08-30	17	16830944.49999...
14646.0	2011-12-08	2010-12-20	72	16349866.53999...
12415.0	2011-11-15	2011-01-06	20	9119035.620000...
14911.0	2011-12-08	2010-12-01	197	4356322.590000...
14298.0	2011-12-01	2010-12-14	44	4285059.370000...
13081.0	2011-11-28	2010-12-03	11	3742499.509999...
17511.0	2011-12-07	2010-12-01	31	3446318.640000...
14156.0	2011-11-30	2010-12-03	52	3218471.699999...
17841.0	2011-12-08	2010-12-01	124	2954547.169999...
14088.0	2011-11-29	2011-01-21	13	2723180.459999...

We can save this table as 'monetary'

Recency & Frequency

We chose a reference date, which is the *most recent purchase* in the dataset. Find the *date difference between the reference date and the last purchase date of each customer*. This is the recency value for that particular customer.

For frequency we calculate the months the person has been a customer by difference in first and last purchase +1 (for when first and last month are same and the customer should be considered a customer for at least 1 month)

```
select *, date_diff(reference_date ,last_purchase_date,day)recency, no_of_purchases/
(months_cust) frequency
from ( select *, max(last_purchase_date) over()+1 as reference_date,
date_diff(last_purchase_date,first_purchase_date,month)+1 months_cust
from `retail.monetary`)
```

order by CustomerID

CustomerID	last_purchase_date	first_purchase_date	no_of_purchases	monetary	reference_date	months_cust	recency	frequency
12346.0	2011-01-18	2011-01-18	1	77183.6	2011-12-10	1	326	1.0
12347.0	2011-12-07	2010-12-07	7	122083....	2011-12-10	13	3	0.538...
12348.0	2011-09-25	2010-12-16	4	13230.0...	2011-12-10	10	76	0.4
12349.0	2011-11-21	2011-11-21	1	84951.8...	2011-12-10	1	19	1.0
12350.0	2011-02-02	2011-02-02	1	4710.4	2011-12-10	1	311	1.0
12352.0	2011-11-03	2011-02-16	7	12528.2...	2011-12-10	10	37	0.7
12353.0	2011-05-19	2011-05-19	1	58.6000...	2011-12-10	1	205	1.0
12354.0	2011-04-21	2011-04-21	1	45371.5...	2011-12-10	1	233	1.0

save it as another table named `RFM`.

Determine quintiles for each RFM metric

- Quintiles are like percentiles, but instead of dividing the data into 100 parts, you divide it in 5 equal parts.
- The `approximate_quantiles` will return an array for each percentile and for creating quintiles out of it we will need values at 20, 40 and so on. We save those values as m20, m40 for monetary and f, r for frequency and recency respectively.

select

a.*,

#all percentiles for monetary

b.percentiles[offset(20)] m20,

b.percentiles[offset(40)] m40,

b.percentiles[offset(60)] m60,

b.percentiles[offset(80)] m80,

b.percentiles[offset(100)] m100,

#all percentiles for frequency

c.percentiles[offset(20)] f20,

c.percentiles[offset(40)] f40,

c.percentiles[offset(60)] f60,

c.percentiles[offset(80)] f80,

c.percentiles[offset(100)] f100,

#all percentiles for recency

d.percentiles[offset(20)] r20,

d.percentiles[offset(40)] r40,

d.percentiles[offset(60)] r60,

d.percentiles[offset(80)] r80,

d.percentiles[offset(100)] r100

```

From `retail.RFM` a,
(select approx_quantiles(monetary, 100) percentiles from `retail.RFM` )b,
(select approx_quantiles(frequency, 100) percentiles from `retail.RFM` )c,
(select approx_quantiles(recency, 100)percentiles from `retail.RFM` )d
order by CustomerID

```

CustomerID	last_purchase	first_purchase	no_of_pur	monetary	reference_date	months_cust	recency	frequency	m20	m40	m60	m80	m100
12346.0	2011-01-18	2011-01-18	1	77183.6	2011-12-10	1	326	1.0	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...
12347.0	2011-12-07	2010-12-07	7	12208...	2011-12-10	13	3	0...	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...
12348.0	2011-09-25	2010-12-16	4	13230...	2011-12-10	10	76	0.4	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...
12349.0	2011-11-21	2011-11-21	1	84951...	2011-12-10	1	19	1.0	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...
12350.0	2011-02-02	2011-02-02	1	4710.4	2011-12-10	1	311	1.0	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...
12352.0	2011-11-03	2011-02-16	7	12528...	2011-12-10	10	37	0.7	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...
12353.0	2011-05-19	2011-05-19	1	58.600...	2011-12-10	1	205	1.0	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...
12354.0	2011-04-21	2011-04-21	1	45371...	2011-12-10	1	233	1.0	2394.1...	7374.0...	17904.5...	49565.77...	16830944.4...

save these as a new table named `quantile`.

Assign scores for each RFM metric :

While with F and M, we give higher scores for higher quintiles, R should be reversed as more recent customers should be scored higher in this metric.

Frequency and Monetary value are combined (as both of them are indicative to purchase volume anyway). to reduce the possible options from 125 to 50.

Use CASE to get values and assign scores accordingly, so we just get the data from the `quintiles` table that we stored assign scores.

```

select  CustomerID,    m_score,    f_score,r_score,    recency,    frequency,    monetary,
cast(round((f_score+ m_score)/2,0)as int64) as fm_score
from(select *,
case when monetary <= m20 then 1
      when monetary <= m40 and monetary > m20 then 2
      when monetary <= m60 and monetary > m40 then 3
      when monetary <= m80 and monetary > m60 then 4
      when monetary <= m100 and monetary > m80 then 5
end as m_score,
case when frequency <= f20 then 1
      when frequency <=f40 and frequency > f20 then 2
      when frequency <= f60 and frequency > f40 then 3
      when frequency <= f80 and frequency > f60 then 4
      when frequency <= f100 and frequency > f80 then 5
end as f_score,

```

```
#recency scoring s reverse
case when recency <= r20 then 5
      when recency <= r40 and recency > r20 then 4
      when recency <= r60 and recency > r40 then 3
      when recency <= r80 and recency > r60 then 2
      when recency <= r100 and recency > r80 then 1
end as r_score
from `retail.quintiles`)
```

CustomerID	m_score	f_score	r_score	recency	frequency	monetary	fm_score
12346.0	5	3	1	326	1.0	77183.6	4
12347.0	5	2	5	3	0.538...	122083.2...	4
12348.0	3	1	2	76	0.4	13230.00...	2
12349.0	5	3	4	19	1.0	84951.89...	4
12350.0	2	3	1	311	1.0	4710.4	3
12352.0	3	2	3	37	0.7	12528.20...	3
12353.0	1	3	1	205	1.0	58.60000...	2
12354.0	4	3	1	233	1.0	45371.54...	4
12355.0	2	3	1	215	1.0	4554.0	3
12356.0	5	1	1	246	0.5	54613.48...	3

Save table as 'score'

Define the RFM segments using these scores :

- For example, in the **Champions** segment, customers should have bought recently, bought often, and spent the most. Therefore, their R score should be 5 and their combined FM score should be 4 or 5.
- On the other hand, **Can't Lose Them** customers made the biggest purchases, and often, but haven't returned for a long time. Hence their R score should be 1, and FM score should be 4 or 5.

SELECT

```
CustomerID,
recency,frequency,monetary,
r_score, f_score, m_score,
```



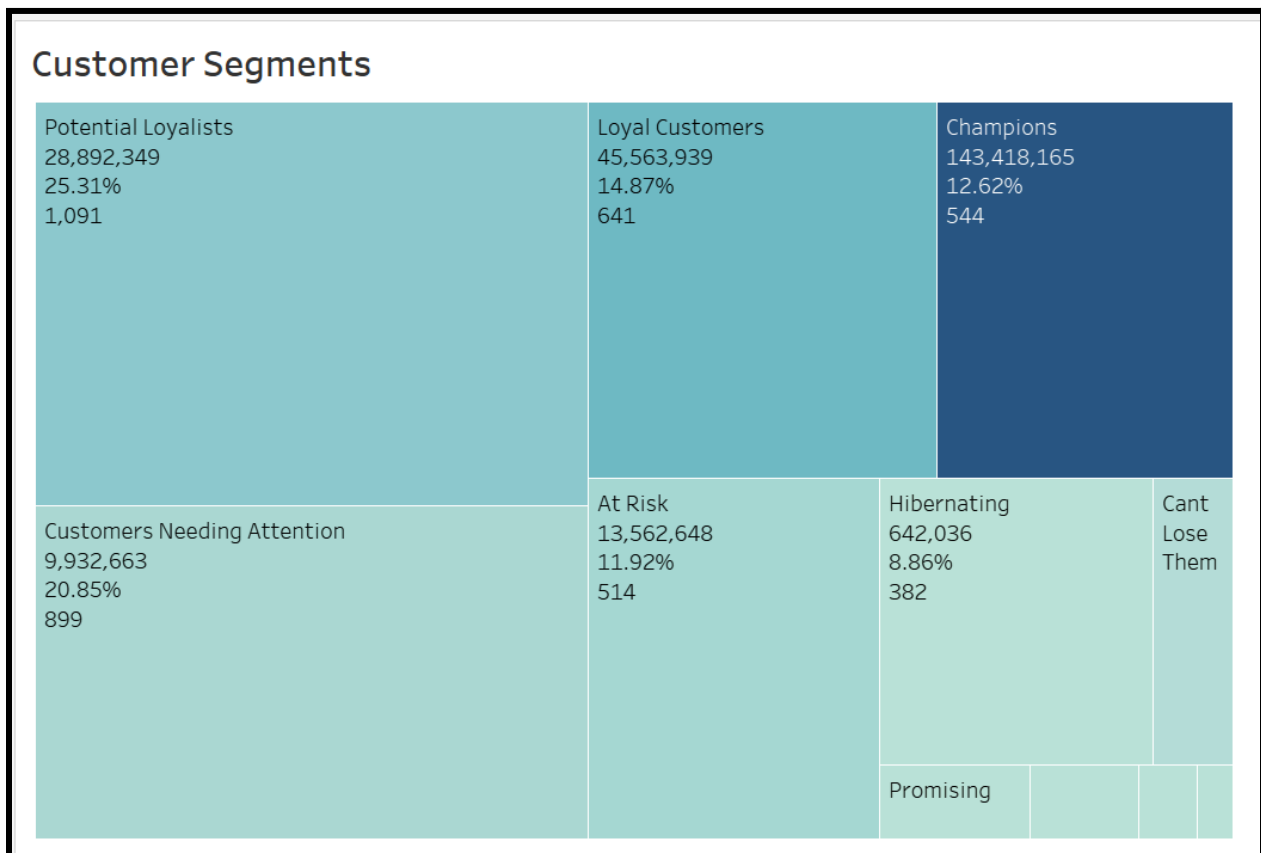
```

fm_score,
CASE WHEN (r_score = 5 AND fm_score = 5)
    OR (r_score = 5 AND fm_score = 4)
    OR (r_score = 4 AND fm_score = 5)
THEN 'Champions'
WHEN (r_score = 5 AND fm_score = 3)
    OR (r_score = 4 AND fm_score = 4)
    OR (r_score = 3 AND fm_score = 5)
    OR (r_score = 3 AND fm_score = 4)
THEN 'Loyal Customers'
WHEN (r_score = 5 AND fm_score = 2)
    OR (r_score = 4 AND fm_score = 2)
    OR (r_score = 3 AND fm_score = 3)
    OR (r_score = 4 AND fm_score = 3)
THEN 'Potential Loyalists'
WHEN r_score = 5 AND fm_score = 1 THEN 'Recent Customers'
WHEN (r_score = 4 AND fm_score = 1)
    OR (r_score = 3 AND fm_score = 1)
THEN 'Promising'
WHEN (r_score = 3 AND fm_score = 2)
    OR (r_score = 2 AND fm_score = 3)
    OR (r_score = 2 AND fm_score = 2)
THEN 'Customers Needing Attention'
WHEN r_score = 2 AND fm_score = 1 THEN 'About to Sleep'
WHEN (r_score = 2 AND fm_score = 5)
    OR (r_score = 2 AND fm_score = 4)
    OR (r_score = 1 AND fm_score = 3)
THEN 'At Risk'
WHEN (r_score = 1 AND fm_score = 5)
    OR (r_score = 1 AND fm_score = 4)
THEN 'Cant Lose Them'
WHEN r_score = 1 AND fm_score = 2 THEN 'Hibernating'
WHEN r_score = 1 AND fm_score = 1 THEN 'Lost'
END AS rfm_segment
FROM `retail.score`
ORDER BY CustomerID

```

CustomerID	recency	frequency	monetary	r_score	f_score	m_score	fm_score	rfm_segment
12346.0	326	1.0	77183.6	1	3	5	4	Cant Lose Them
12347.0	3	0.5384...	122083.2...	5	2	5	4	Champions
12348.0	76	0.4	13230.00...	2	1	3	2	Customers Needing A...
12349.0	19	1.0	84951.89...	4	3	5	4	Loyal Customers
12350.0	311	1.0	4710.4	1	3	2	3	At Risk
12352.0	37	0.7	12528.20...	3	2	3	3	Potential Loyalists
12353.0	205	1.0	58.60000...	1	3	1	2	Hibernating
12354.0	233	1.0	45371.54...	1	3	4	4	Cant Lose Them
12355.0	215	1.0	4554.0	1	3	2	3	At Risk
12356.0	246	0.5	54613.48...	1	1	5	3	At Risk

Each customer should have an RFM segment assignment like this.



Insights:

- Total Customers Analyzed: 4,300
- Customers Falling into 6 Major Segments: 94%

- Champions
Percentage of Customers: 13%
Revenue Contribution: ₹14.4 Crores
Average Revenue per Customer: ₹270,000
Features: These customers are highest revenue contributors, with frequent purchases and high spending.
- Loyal Customers
Number of Customers: 641
Revenue Contribution: ₹4.5 Crores
Characteristics: Regular purchasers who contribute significantly to overall revenue.
- Potential Loyalists
Percentage of Customers: 25%
Characteristics: Recent buyers who have made multiple purchases and spent a good amount. This is the largest segment, indicating potential for conversion to loyal customers.
- Customers Needing Attention
Number of Customers: 900
Revenue Contribution: ₹1 Crore
Average Revenue per Customer: ₹11,000
Characteristics: These customers have above-average Recency, Frequency & Monetary (RFM) values but have not made recent purchases.
- At Risk
Percentage of Customers: 12%
Characteristics: These customers used to spend a lot and purchased frequently but have not made purchases in a long time.
- Hibernating and Lost Customers
Combined Percentage: Less than 1%
Characteristics: With low customer attrition, indicating positive business health
- Additional Insights
New Customers Added: 21 (0.5% of total customers)
Indicates a need to focus on new customer acquisition strategies.

Analysis Summary:

- Top Performers: Champions and Loyal Customers are key revenue drivers.
- Growth Opportunities: Potential Loyalists and Customers Needing Attention present significant opportunities for increasing customer loyalty and revenue.
- Business Health: Low percentage of Hibernating and Lost Customers demonstrates a strong customer retention rate.
- These insights highlight the importance of focusing on high-value segments while developing strategies to engage and retain customers in the 'Potential Loyalists' and 'Customers Needing Attention' segments. Additionally, efforts should be made to attract new customers to ensure sustained growth.

Recommendations:

- Strategies for Champions, Loyal Customers, and Potential Loyalists
Cross-Selling and Up-Selling:
Implement strategies to recommend complementary or higher-end products to these segments to increase the average order value.
Loyalty Programs and Exclusive Benefits, extra convenient features
These initiatives will keep high-value customers engaged and promote brand.

Customer Feedback:
Actively seek feedback from these segments to understand their needs and preferences better, enabling improved service and product offerings.

Enhanced Customer Experience:
Ensure a seamless buying process, responsive customer service, and robust after-sales support to enhance overall customer satisfaction.
- Strategies for Potential Loyalists

Targeted Marketing Campaigns, Recommend products based on purchase history.
Use personalized offers and product reminders to encourage repeat purchases and increase customer loyalty.
- Strategies for Customers Needing Attention and At Risk

Reactivation Campaigns:
Use limited-time offers & discounts for next purchases.

Provide personalized recommendations highlighting the benefits and features of new or popular products they haven't purchased yet.

Customer Engagement:

Highlight the benefits of recent arrivals and popular products to attract their interest and encourage re-engagement.

- Strategies for New Customer Acquisition(Recent customers)

Increase Footfall:

Use blogs, social media, referrals, and promo codes to attract new customers.

Offer a wide range of products to cater to diverse customer needs.

Digital Engagement:

Enhance digital engagement with customers through various social media platforms to build a strong online presence.

Summary

- Implementing these recommendations will help maintain and enhance the engagement of high-value customers, reactivate less active segments, and attract new customers. Focusing on personalized marketing, loyalty programs, and customer feedback will drive customer satisfaction and revenue growth, while churn analysis will ensure continuous improvement in addressing customer needs.

Explanation Video link: [Detailed explanation](#)

Tableau visualization link: [Customer_Segmentation 1](#)

Medium article link : [RFM analysis](#)

Contact Prakyath Shetty

7026169082

pprakyathshetty@gmail.com