Customer Segmentation using SQL (MySQL)

Problem Statement:

- To perform RFM Analysis for a retail store.
- 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.

R: Recency (More points for customer who made purchase recently

F: Frequency (More points for customer who purchase many times

M: Monetary (More points for customer whose purchase value is larger

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 have not 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 have not returned for a long time.	Win them back via renewals or newer products, do not 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.

Dataset:

Query:

```
create database Customer_Seg_Proj;

use Customer_Seg_Proj;

create table Txn_table(
Invoice_no int(7),
Stock_code text,
Description_Item text,
Quantity int(7),
Invoice_date timestamp,
Unit_price decimal,
Cust_ID int,
Country text);

show tables;
```

Load the local dataset in the Txn_table.

Step1: Calculating Bill Amount for each invoice

We have details for all invoices, for our analysis we need for each invoice what is bill amount

Query:

```
create table invoice_bill as (
SELECT Invoice_no, sum(Quantity*Unit_price) as bill FROM txn_table group by Invoice
```

Step2: Getting details for customers

Query1:

```
create table cust_invoices as (
SELECT Cust_ID, Invoice_no, min(date(Invoice_date)) as txn_date from txn_table group by
Cust_ID,Invoice_no);
```

Query2: Joining first two tables to get all details of customer and invoices

```
create table RFM_data as (
select cust_invoices.*, invoice_bill.bill from cust_invoices right join invoice_bill on
cust_invoices.Invoice_no=invoice_bill.Invoice_no)
```

Step3: Getting Recency, Frequency and Monetary for each customer

Recency is calculated as difference between ref date and purchase date.
Frequency is calculated as no of purchases/month.
Monetary is calculated as sum of all bills for customer.

Query:

```
with cte1 as (
select Cust_ID, date(min(txn_date)) as first_purchase, date(max(txn_date)) as last_purchase,
count(Invoice_no) as total_purchases, sum(bill) as monetary from RFM_data group by Cust_ID),

cte2 as (
select date(max(txn_date)) as ref_date from RFM_data),

cte3 as (
select * from cte1, cte2)
select * from cte3;

create table RFM_final as (
select Cust_ID, datediff(ref_date,last_purchase)+1 as Recency,
total_purchases/(datediff(last_purchase,first_purchase)+1) as Frequency, monetary from cte3
order by Cust_ID);
```

Step4: Dividing R, F and M into 5 quantiles

- We need to divide Recency, frequency and Monetary into 5 parts for segmentation.
- Initially these columns are divided into 100 parts (percentile) and then 5 percentiles (20^{th}, 40^{th}, 60^{th}, 80^{th} and 100^{th}) are calculated for segmentation

Query:

```
select @r20 := Recency from RFM_final order by Recency limit 1 offset 863;
select @r40 := Recency from RFM_final order by Recency limit 1 offset 1727;
select @r60 := Recency from RFM_final order by Recency limit 1 offset 2591;
select @r80 := Recency from RFM_final order by Recency limit 1 offset 3455;
select @r100 := Recency from RFM_final order by Recency limit 1 offset 4319;

select @f20 := Frequency from RFM_final order by Frequency limit 1 offset 863;
select @f40 := Frequency from RFM_final order by Frequency limit 1 offset 1727;
select @f60 := Frequency from RFM_final order by Frequency limit 1 offset 2591;
select @f80 := Frequency from RFM_final order by Frequency limit 1 offset 3455;
select @f100 := Frequency from RFM_final order by Frequency limit 1 offset 4319;

select @m20 := monetary from RFM_final order by monetary limit 1 offset 863;
select @m40 := monetary from RFM_final order by monetary limit 1 offset 1727;
select @m60 := monetary from RFM_final order by monetary limit 1 offset 2591;
```

```
select @m80 := monetary from RFM_final order by monetary limit 1 offset 3455;
select @m100 := monetary from RFM_final order by monetary limit 1 offset 4319;
```

Step5: Calculating RFM score

- Based on quantiles, we need to give RFM score to each customer from 1 to 5.
- Most recent customers (recency percentile less than 20) will be given 5 points.
- Most frequent buyers (frequency percentile between 80-100) will be given 5 points.
- Most expensive buyers (Monetary percentile between 80-100) will be given 5 points.
- F and M scores are combined to reduce large no of segments.

Query:

```
create table rfm_score as ( SELECT Cust_ID, Recency, Frequency, monetary,
case when Recency<=@r20 then 5
when Recency>@r20 and Recency<=@r40 then 4
when Recency>@r40 and Recency<=@r60 then 3
when Recency>@r60 and Recency<=@r80 then 2
when Recency>@r80 and Recency<=@r100 then 1 end as RScore,

case when Frequency<=@f20 then 1
when Frequency>@f20 and Frequency<=@f40 then 2
when Frequency>@f40 and Frequency<=@f60 then 3
when Frequency>@f60 and Frequency<=@f80 then 4
when Frequency>@f80 and Frequency<=@f100 then 5 end as FScore,

case when monetary<=@m20 then 1
when monetary>@m20 and monetary<=@m40 then 2
when monetary>@m40 and monetary<=@m60 then 3
when monetary>@m60 and monetary<=@m80 then 4
when monetary>@m80 and monetary<=@m100 then 5 end as MScore FROM RFM_final);
select * from rfm_score;

create table final_scores_data as (
select Cust_ID, Recency, Frequency, monetary, RScore, round((FScore+MScore)/2) as FM_score
from rfm_score);
```

Step6: Segmentation based on RFM scores

- Based on RFM score, customers are divided into 11 segments as follows

Sr No	Segments
1	Champions
2	Loyal Customers
3	Potential Loyalists
4	Recent Customers
5	Promising
6	Customers Needing Attention
7	About to Sleep
8	At Risk
9	Cant Lose Them
10	Hibernating
11	Lost

- Based on score these segments will be assigned as follows

		FM				
		1	2	3	4	5
R	1	Lost	Hibernating	At Risk	Cant Lose Them	Cant Lose Them
	2	About to Sleep	Customers Needing Attention	Customers Needing Attention	At Risk	At Risk
	3	Promising	Customers Needing Attention	Potential Loyalists	Loyal Customers	Loyal Customers
	4	Promising	Potential Loyalists	Potential Loyalists	Loyal Customers	Champions
	5	Recent Customers	Potential Loyalists	Loyal Customers	Champions	Champions

Query:

```
create table output_score_segmentation as (SELECT Cust_ID, Recency, Frequency, monetary,
RScore, FM_score,
CASE WHEN (RScore = 5 AND FM_score = 5) OR (RScore = 5 AND FM_score = 4) OR (RScore = 4
AND FM_score = 5) THEN 'Champions'
WHEN (RScore = 5 AND FM_score =3) OR (RScore = 4 AND FM_score = 4) OR (RScore = 3 AND
FM_score = 5) OR (RScore = 3 AND FM_score = 4) THEN 'Loyal Customers'
WHEN (RScore = 5 AND FM_score = 2) OR (RScore = 4 AND FM_score = 2) OR (RScore = 3 AND
FM_score = 3) OR (RScore = 4 AND FM_score = 3) THEN 'Potential Loyalists'
WHEN RScore = 5 AND FM_score = 1 THEN 'Recent Customers' WHEN (RScore = 4 AND FM_score
= 1) OR (RScore = 3 AND FM_score = 1) THEN 'Promising'
WHEN (RScore = 3 AND FM_score = 2) OR (RScore = 2 AND FM_score = 3) OR (RScore = 2 AND
FM_score = 2) THEN 'Customers Needing Attention'
WHEN RScore = 2 AND FM_score = 1 THEN 'About to Sleep'
WHEN (RScore = 2 AND FM_score = 5) OR (RScore = 2 AND FM_score = 4) OR (RScore = 1 AND
FM_score = 3) THEN 'At Risk'
WHEN (RScore = 1 AND FM_score = 5) OR (RScore = 1 AND FM_score = 4) THEN 'Cant Lose Them'
WHEN RScore = 1 AND FM_score = 2 THEN 'Hibernating'
WHEN RScore = 1 AND FM_score = 1 THEN 'Lost' END AS rfm_segment FROM final_scores_data);

select * from output_score_segmentation;
```