Exploratory Analysis of Customer Segmentation (RFM Analysis) in SQL

Analyst: Olanrewaju Olatunji

Business Request/Problem Statement:

Hi, Olanrewaju

The organization is looking to lunch a new campaign on all products. We need you to segment our customers by finding the best and the worse group of customers for the campaign.

Objectives: -Rank and group customers

-Identify good and bad customers

-Who are our most valuable customers.

Technique: Rank and group customers based on Recency, Frequency and Monetary total of their transactions to identify the top customers and perform targeted marketing campaign.

Analyst: Olanrewaju Olatunji

What is RFM Analysis?

RFM Analysis stands for Recency, Frequency and Monetary.

It is an indexing technique that uses past purchase behaviour to segment customers. An RFM report is a way of segmenting customers using three metrics;

-Recency (How long ago was the customer's last purchase).

-Frequency (How often the customers purchase from us).

-Monetary Value (How much a customer spends).

Analyst: Olanrewaju Olatunji

Step 1:

Data Filtering: I filtered the dataset to what is needed to avoid ambiguity. The columns needed were; order id, order date, customer name and sales

Raw Data:

	order_id	order_date	customers_name	country	state	city	region	segment	ship_mode	category	sub_category	product_name
1	AZ-2011-9050313	2011-01-03	Summer Hayward	United Kingdom	England	Southport	North	Consumer	Economy	Furniture	Bookcases	Dania Corner Shelving, Tradit
2	AZ-2011-6674300	2011-01-04	Devin Huddleston	France	Auver...	Valence	Cen...	Consumer	Economy	Office ...	Art	Binney & Smith Sketch Pad, I
3	AZ-2011-617423	2011-01-05	Daniel Burke	France	Auver...	Echirolles	Cen...	Home O...	Priority	Office ...	Art	Binney & Smith Pencil Sharpe
4	AZ-2011-617423	2011-01-05	Daniel Burke	France	Auver...	Echirolles	Cen...	Home O...	Priority	Office ...	Art	Sanford Canvas, Fluorescent
5	AZ-2011-2918397	2011-01-07	Fredrick Beverid...	France	Prove...	La Seyn...	Cen...	Corporate	Priority	Furniture	Bookcases	Bush Floating Shelf Set, Pine
6	AZ-2011-2918397	2011-01-07	Fredrick Beverid...	France	Prove...	La Seyn...	Cen...	Corporate	Priority	Office ...	Fasteners	Accos Thumb Tacks, Assorte
7	AZ-2011-2918397	2011-01-07	Fredrick Beverid...	France	Prove...	La Seyn...	Cen...	Corporate	Priority	Office ...	Storage	Smead Lockers, Industrial
8	AZ-2011-6712797	2011-01-11	Evie Flockhart	Italy	Liguria	Genoa	South	Consumer	Economy	Office ...	Binders	Ibico Hole Reinforcements, F
9	AZ-2011-4827146	2011-01-11	Faith Greenwood	Austria	Vienna	Vienna	Cen...	Consumer	Economy	Office ...	Art	Boston Canvas, Fluorescent
10	AZ-2011-4827146	2011-01-11	Faith Greenwood	Austria	Vienna	Vienna	Cen...	Consumer	Economy	Office ...	Storage	Smead Trays, Single Width
11	AZ-2011-6439906	2011-01-11	Summer Hayward	Spain	Murcia	Murcia	South	Consumer	Economy	Office ...	Labels	Novimex File Folder Labels, A
12	AZ-2011-7053593	2011-01-11	Gracie Powell	United Kingdom	England	Woking	North	Consumer	Immediate	Furniture	Chairs	SAFCO Executive Leather Ar
13	AZ-2011-7053593	2011-01-11	Gracie Powell	United Kingdom	England	Woking	North	Consumer	Immediate	Office ...	Art	Binney & Smith Canvas, Blue

Query executed successfully. | DESKTOP-170U4IG (16.0 RTM) | sa (63) | AdventureWorks2019 | 00:00:00 | 6,311 rows

Analyst: Olanrewaju Olatunji

Filtered Dataset:

```sql
-- Step 1: Data Filtering

SELECT
    order_id,
    order_date,
    customers_name,
    sales
FROM [dbo].[sample]
```

100 %

Results | Messages | Client Statistics

	order_id	order_date	customers_name	sales
1	AZ-2011-9050313	2011-01-03	Summer Hayward	854
2	AZ-2011-6674300	2011-01-04	Devin Huddleston	140
3	AZ-2011-617423	2011-01-05	Daniel Burke	90
4	AZ-2011-617423	2011-01-05	Daniel Burke	207
5	AZ-2011-2918397	2011-01-07	Fredrick Beveridge	155
6	AZ-2011-2918397	2011-01-07	Fredrick Beveridge	33
7	AZ-2011-2918397	2011-01-07	Fredrick Beveridge	716
8	AZ-2011-6712797	2011-01-11	Evie Flockhart	22
9	AZ-2011-4827146	2011-01-11	Faith Greenwood	55
10	AZ-2011-4827146	2011-01-11	Faith Greenwood	97
11	AZ-2011-6439906	2011-01-11	Summer Hayward	40
12	AZ-2011-7053593	2011-01-11	Gracie Powell	1384
13	AZ-2011-7053593	2011-01-11	Gracie Powell	103
14	AZ-2011-5702370	2011-01-12	Hershel Snyder	552

Query executed successfully.

Step 2:

Calculating RFM Values: I used some aggregate functions to determine the RFM Values. Summed up the sales to get Monetary value (It shows the amount spent by each customer on every purchase made). I also did a count of the Order-ID to get my Frequency value (it gives the total order count for each customer). To be able to get the Recency Value, I did a max of the Order-Date to get the last Order-Date for each customer purchase. Then I used a select statement of the max order-Date to get the maximum date in the whole Dataset. And lastly, I used a DateDiff function to subtract the last order date from the maximum order date to be able to get my Recency Value. (This gave the number of days a customer last purchase).

```
WITH RFM_Value AS (
    SELECT
        customers_name,
        COUNT(order_id) AS Frequency,
        SUM(sales) AS Monetary,
        MAX(order_date) AS Last_order_date,
        (SELECT MAX(order_date) FROM [dbo].[sample]) AS Max_ordr_date,
        DATEDIFF(DD, MAX(order_date), (SELECT MAX(order_date) FROM [dbo].[sample])) AS Recency

    FROM [dbo].[sample]
    GROUP BY customers_name
),
```

Step 3:

Customer Ratings based on Recency, Frequency and Monetary: I assigned a rating to each value (Recency, Frequency and Monetary), the value will be based on the volume of my customers and transactions, I used rating of 1-5 and used NTILE function to assign the ratings based on my RFM Values.

```
--Customers Ratings based on Recency,Frequency and Monetary
    RFM_Score as (
SELECT
    customers_name,
    Recency,
    Frequency,
    Monetary,

    NTILE(5) OVER (ORDER BY Recency DESC) AS R,
    NTILE(5) OVER (ORDER BY Frequency) AS F,
    NTILE(5) OVER (ORDER BY Monetary) AS M


FROM RFM_Value
```

Step 4:

RFM Score: To get RFM Score, I concatenated RFM Ratings as string using CAST function. These scores each customer based on the above RFM ratings

```
SELECT
customers_name,
Recency,
Frequency,
Monetary,
R,F,M,

      Cast (R as varchar)+ Cast(F as varchar)+ Cast(M as varchar)as rfm_score


FROM RFM_Score
```

Step 5:

Customer Segmentation: The last step is segmenting customers based on the RFM Scores assigned to each customer by using a CASE Statement. This helped to identify top customers, loyal customers, customers that needs attention, lost customers etc.

```
--Customers RFM segmentation
    SELECT
    customers_name,
    Recency,
    Frequency,
    Monetary,
    R,F,M,
    rfm_score,
        case
        when rfm_score in (213, 221, 231, 241, 251, 312, 321, 331) then 'About to Sleep'
        when rfm_score in (124,125,133,134,135,142,143,145,152,153,224,225,234,235,242,243,244,245,252,253,254,255  ) then
        when rfm_score in (113,114,115,144,154,155,214,215) then 'Can Not Lose Them'
        when rfm_score in (445,454,455,544,545,554,555) then 'Champions'
        when rfm_score in (122,123,132,211,212,222,223,231,232,233,241,251,322,332) then 'Hibernating Customers'
        when rfm_score in (111,112,121,131,141,151) then 'Lost Customers'
        when rfm_score in (335,344,345,354,355,435,444,543) then 'Loyal Customers'
        when rfm_score in (324,325,334,343,434,443,534,535) then 'Need Attention'
        when rfm_score in (311,411,412,421,422,511,512) then 'New Customers'
        when rfm_score in (323,333,341,342,351,352,353,423,431,432,433,441,442,451,452,453,531,532,533,541,542,551,552,553
        when rfm_score in (313,314,315,413,414,415,424,425,513,514,515,521,522,523,524,525) then 'Promising '

    end as rfm_segment
    FROM RFM_Segment
```

Analyst: Olanrewaju Olatunji

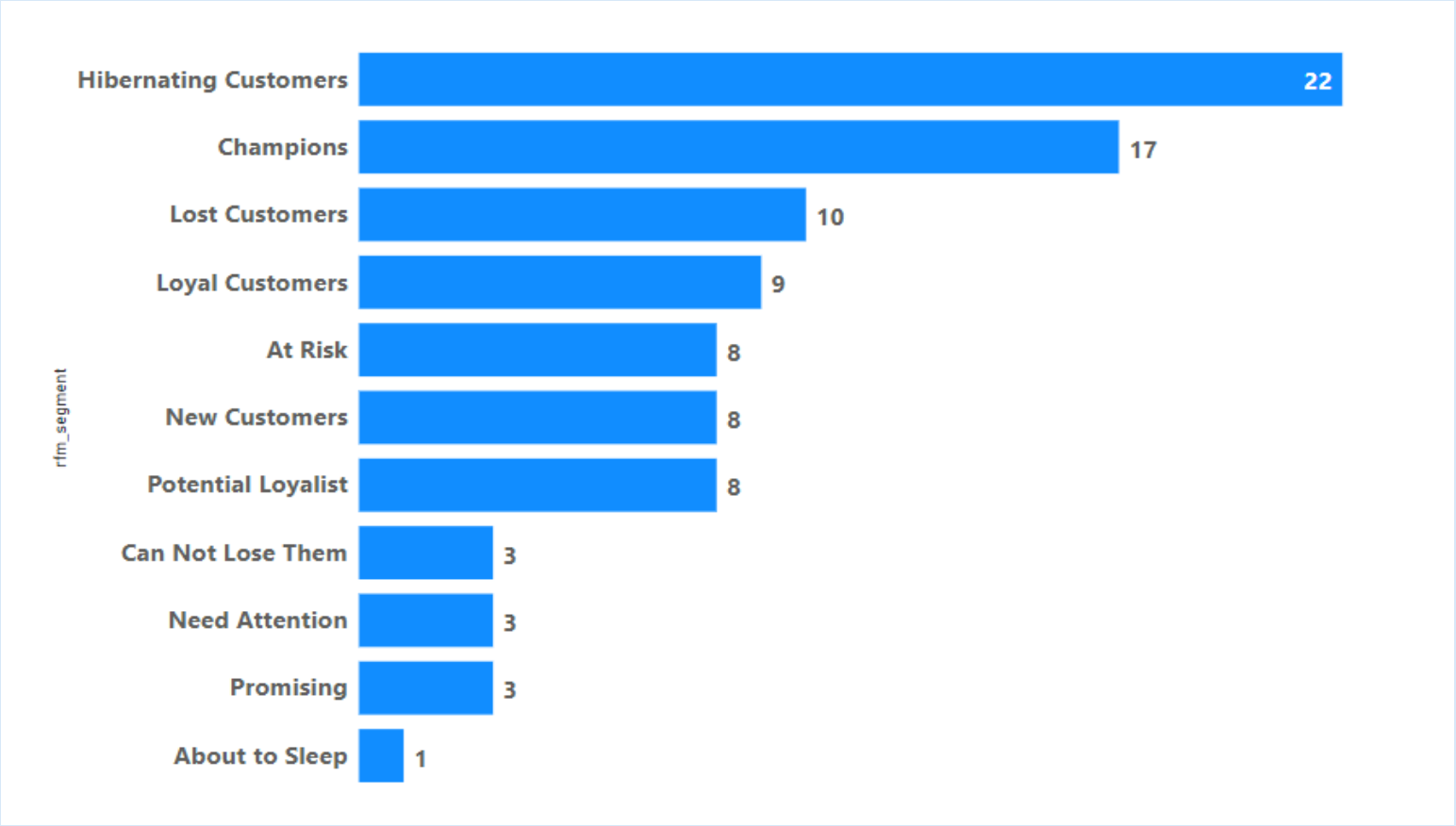
RFM Table:

Results | Messages | Client Statistics

	customers_name	Recency	Frequency	Monetary	R	F	M	rfm_score	rfm_segment
1	James Brown	742	1	14	1	1	1	111	Lost Customers
2	Jodie Garner	1343	1	16	1	1	1	111	Lost Customers
3	Molly Hawkins	721	1	19	1	1	1	111	Lost Customers
4	Grace Brown	497	1	23	1	1	1	111	Lost Customers
5	Paige Bibi	376	2	38	2	1	1	211	Hibernating C...
6	Marie Lawson	387	1	40	2	1	1	211	Hibernating C...
7	Max Baker	1238	1	44	1	1	1	111	Lost Customers
8	Alexandra Mah...	118	1	47	3	1	1	311	New Customers
9	James Fleming	1237	1	50	1	1	1	111	Lost Customers
10	Imogen Daly	1147	3	56	1	1	1	111	Lost Customers
11	Eleanor Pollard	358	3	56	2	1	1	211	Hibernating C...
12	Herbert Harrell	676	1	57	1	1	1	111	Lost Customers
13	Clifford Defreitas	509	2	59	1	1	1	111	Lost Customers
14	Kian Thomson	134	2	65	3	1	1	311	New Customers

Analyst: Olanrewaju Olatunji

Customer Segmentation: By scoring our customers by Recency, Frequency and Monetary, I am able to Identify 11 major segments for targeting.



Analyst: Olanrewaju Olatunji

Conclusion:

Champion and loyal customers are our top customers that have purchased an average of 100 times during the evaluation period, spending an average of $156,569 and have purchased in recent days.
While potential and promising customer have purchased an average of 50 times and have spent an average of $100,000 in the last 34days.

New Customers, Need Attention, About to Sleep, At Risk, Can not Lose, and Hibernating customers accounted for 48.8% of the total customers.
While 10% of the our customers are lost.

Recommendation:
We will create a target campaign to re-engage our Lost, At risk, Can not lose, Hibernating, Need attention and About to Sleep Customers with incentives/discounts.

Further analyse our top customers, loyal and new customers to maximise revenue.

Analyst: Olanrewaju Olatunji

THANK YOU!

Don’t forget to hit the like button, share, comment and follow for more.

linkedin.com/in/olatunji-olanrewaju-5b910480

Analyst: Olanrewaju Olatunji