Online Retile System

This document contains a study and analysis of online sales data and the creation of a customer segmentation system. It also includes some recommendations.

Asmaa Ali Saber Ali

Online Retile System

The **Online Retail** system is a system that allows companies to sell their products and services online. This system usually includes a company website where products and services are displayed, and customers can browse and add items to their shopping cart.
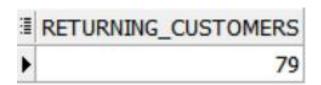
 The **Objective** of analyzing the online retail system is to evaluate the system's performance and understand how to improve it to achieve desired goals. This involves analyzing a range of different data and variables related to the online retail system

Using this data and information, the performance of the online retail system is analyzed, and its strengths and areas for improvement are identified to achieve desired goals.

Let's explore together the data we have and brainstorm some ideas that may help the business.

Q1.1 Customer retention rate:

```sql
WITH CUSTOMER_COUNTS AS (
    SELECT
        CUSTOMER_ID,
        COUNT(DISTINCT INVOICE) AS TOTAL_PURCHASES
    FROM
        ONLINERETAIL
    GROUP BY
        CUSTOMER_ID
)
SELECT
    COUNT(*) AS RETURNING_CUSTOMERS
FROM
    CUSTOMER_COUNTS
WHERE
    TOTAL_PURCHASES > 1;
```

RETURNING_CUSTOMERS
79

By calculating the customer retention rate, we can evaluate the company's success in maintaining its customers and ensure that they do not lose to competitors. This rate can also be used to determine the extent of customer satisfaction with the services or products provided by the company, and to determine the factors that affect customer retention in the company.

In addition, the customer retention rate can be used as a measure of a company's sustainable growth, as maintaining existing customers can reduce the costs associated with attracting new customers, and increase continued revenue through existing customers.

Q1.2 Monthly Active Users:

```sql
SELECT
    TO_CHAR(TO_DATE(INVOICEDATE, 'MM/DD/YYYY
HH24:MI'), 'YYYY-Mon') AS MONTH,
    COUNT(DISTINCT CUSTOMER_ID) AS
MONTHLY_ACTIVE_USERS
FROM
    ONLINERETAIL
GROUP BY
    TO_CHAR(TO_DATE(INVOICEDATE, 'MM/DD/YYYY
HH24:MI'), 'YYYY-Mon')
ORDER BY
    MONTHLY_ACTIVE_USERS DESC ;
```

MONTH	MONTHLY_ACTIVE_USERS
2011-Nov	44
2011-Sep	37
2011-May	32
2011-Oct	29
2011-Mar	29
2011-Jul	24
2011-Jun	24
2011-Dec	23
2010-Dec	22
2011-Jan	20
2011-Feb	20
2011-Aug	19
2011-Apr	14

This query calculates the number of active users per month, by counting the distinct customers who made a purchase in each month. The TO_CHAR function is used to format the invoice date as a month-year string

Q1.3 What if we want to look at a specific day

```sql
SELECT
    TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI') AS
INVOICE_DATE,
    SUM(PRICE*QUANTITY) AS DAILY_REVENUE
FROM
    ONLINERETAIL
WHERE
    TO_CHAR(TO_DATE(INVOICEDATE, 'MM/DD/YYYY
HH24:MI'), 'YYYY-MM') = '2011-12'
GROUP BY
    TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI')
ORDER BY
    TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI');
```

INVOICE_DATE	DAILY_REVENUE
12/1/2011 10:07:00 AM	780
12/1/2011 12:14:00 PM	312.29
12/1/2011 12:53:00 PM	139.03
12/1/2011 12:55:00 PM	264.65
12/1/2011 1:03:00 PM	616.32
12/1/2011 1:18:00 PM	460.3
12/1/2011 1:47:00 PM	287.9
12/1/2011 3:55:00 PM	1171.4
12/2/2011 10:15:00 AM	5.45
12/2/2011 10:52:00 AM	213.1
12/2/2011 11:12:00 AM	273.78
12/2/2011 2:26:00 PM	385.56
12/2/2011 2:30:00 PM	42.88

This query calculates the daily revenue for December 2011, by filtering for invoices that occurred in that month and aggregating the total revenue for each day

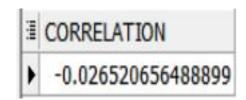## Q1.4 Monthly revenue and total number of orders per month

```sql
SELECT
    TO_CHAR(TO_DATE(INVOICEDATE, 'MM/DD/YYYY
HH24:MI'), 'YYYY-MM') AS MONTH,
    COUNT(DISTINCT INVOICE) AS TOTAL_ORDERS,
    SUM(PRICE*QUANTITY) AS MONTHLY_REVENUE
FROM
    ONLINERETAIL
GROUP BY
    TO_CHAR(TO_DATE(INVOICEDATE, 'MM/DD/YYYY
HH24:MI'), 'YYYY-MM')
ORDER BY
    TOTAL_ORDERS DESC, MONTHLY_REVENUE DESC;
```

MONTH	TOTAL_ORDERS	MONTHLY_REVENUE
2011-11	117	45320.65
2011-09	59	27853.82
2011-10	56	19059.69
2011-05	49	16490.37
2011-07	48	15664.54
2011-03	46	15547.86
2011-08	39	38072.94
2011-06	38	11370.61
2011-12	38	10685.63
2011-02	31	12947.2
2011-01	31	8819.48
2010-12	31	8488.56

We also see here that the **Highes**t month in sales in terms of the number of orders and in terms of returns is the month of **November 2011**

Q1.5 Is there a relationship between the number of items sold and the price of the item?

```sql
SELECT
    CORR(PRICE, QUANTITY) as
CORRELATION
FROM
    ONLINERETAIL
```

CORRELATION
-0.026520656488899

I was trying to find out if there is a relationship between the number of orders and the price of one item, but based on the correlation number, which is equal to **-0.02**, this means that there is **no clear relationship** between them

Q1.6 Study sales during each quarter of the year

```sql
select  distinct Year_, quarter,
 count(QUANTITY) as Quantity,
SUM(PRICE*QUANTITY) AS Quarter_REVENUE
from (
SELECT OnlineRetail.*,
TO_CHAR(TO_DATE(INVOICEDATE, 'MM/DD/YYYY
HH24:MI'), 'YYYY') AS Year_,
 NTILE(4) OVER(PARTITION BY EXTRACT(YEAR
FROM TO_DATE(INVOICEDATE, 'MM/DD/YYYY
HH24:MI')), ((EXTRACT(MONTH FROM
TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI')) -
1) / 3) + 1 ORDER BY TO_DATE(INVOICEDATE,
'MM/DD/YYYY HH24:MI')) AS quarter
FROM OnlineRetail)
Group By quarter,Year_
order By Quarter_REVENUE DESC , Quantity DESC
```

YEAR_	QUARTER	QUANTITY	QUARTER_REVENUE
2011	2	2678	65567.65
2011	3	2676	56678.56
2011	4	2672	56025.59
2011	1	2681	53441.13
2010	2	114	3137.48
2010	4	113	2108.43
2010	3	114	1856.17
2010	1	114	1386.48

When the sales distribution indicates that the majority of sales occur in the **second** quarter of the year, it usually suggests that there may be a specific season for selling these products and that customers purchase them in large quantities during this period. However, if there is no particular season, studying sales during this period can help identify the factors that affect sales. Therefore, this knowledge can be used to improve sales performance throughout the year.

Studying sales during this period can help identify the best-selling products that match customers' needs during this period and identify differences in demand for products between the other quarters of the year. Consequently, this information can be used to improve marketing and advertising strategies to increase sales throughout the year, especially in the **first quarter**, which is considered the least selling period.

Customer Segmentation

What is The RFM System ?

The RFM system is a method used to analyze customer behavior and evaluate their value to the company. Customers are analyzed according to three main criteria:

1. **Recency:** Measures how recently the customer made a purchase.

2. **Frequency:** Measures how often the customer made a purchase over a specific time period.

3. **Monetary:** Measures the amount of money the customer spent on products or services over a specific time period.

Customers are categorized into different groups based on their evaluation for each of these criteria, which helps companies understand customer behavior and identify customers who represent significant value to the company. Therefore, companies can direct their marketing and sales efforts more effectively to retain current customers and attract new ones.

RFM Model

```sql
WITH Data_ AS (
  SELECT DISTINCT CUSTOMER_ID,
    TRUNC(LAST_VALUE(TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI')) OVER (ORDER BY
TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI') ROWS BETWEEN UNBOUNDED PRECEDING AND
UNBOUNDED FOLLOWING) - LAST_VALUE(TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI')) OVER
(PARTITION BY CUSTOMER_ID ORDER BY TO_DATE(INVOICEDATE, 'MM/DD/YYYY HH24:MI') RANGE
BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING)) AS recency,
    COUNT(INVOICE) OVER (PARTITION BY CUSTOMER_ID ORDER BY INVOICE ROWS BETWEEN UNBOUNDED
PRECEDING AND UNBOUNDED FOLLOWING) AS frequency,
    SUM(PRICE * QUANTITY) OVER (PARTITION BY CUSTOMER_ID) AS monetary
  FROM ONLINERETAIL
),
rfm_score As (
    select Data_.* ,
        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 Data_
                    ),
fm_score As (
    select customer_id,recency,frequency,monetary,r_score,( f_score+m_score/2) as fm_score
    from rfm_score)

select customer_id,recency,frequency,monetary,r_score, fm_score,

CASE
    WHEN (r_score = 5 AND fm_score IN (4, 5)) OR
        (r_score = 4 AND fm_score = 5) THEN 'Champions'
    WHEN (r_score = 5 AND fm_score = 2) OR
        (r_score = 4 AND fm_score in(2,3)) OR
        (r_score = 3 AND fm_score =3) THEN 'Potential Loyalists'
    WHEN (r_score = 5 AND fm_score = 3) OR
        (r_score = 4 AND fm_score = 4) OR
        (r_score = 3 AND fm_score in(5,4)) THEN 'Loyal Customers'
    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 IN (4, 5)) OR
        (r_score = 1 AND fm_score = 3) THEN 'At Risk'
    WHEN (r_score = 1 AND fm_score IN (4, 5)) THEN 'Cant Lose Them'
    WHEN r_score = 1 AND fm_score = 2 THEN 'Hibernating'
    WHEN r_score = 1 AND fm_score = 1 THEN 'Lost'
    Else 'About to sleep'
END AS cust_segment
from fm_score
```

OutPut

CUSTOMER_ID	RECENCY	FREQUENCY	MONETARY	R_SCORE	FM_SCORE	CUST_SEGMENT
12823	74	5	1759.5	2	3	Customers Needing Attention
12822	70	46	948.88	3	4.5	About to sleep
12842	69	34	1118.99	3	4.5	About to sleep
12886	67	82	1378.4	3	6	About to sleep
12843	65	103	1702.26	3	6	About to sleep
12923	63	29	176.97	3	3.5	About to sleep
12939	63	47	11581.8	3	5.5	About to sleep
12885	62	71	1175.22	3	6	About to sleep
12824	58	25	397.12	3	3	Potential Loyalists
12836	58	175	2612.86	3	7.5	About to sleep
12940	54	99	913.54	3	5.5	About to sleep
12863	51	5	521.2	3	2	Customers Needing Attention
12879	43	10	573.22	3	2	Customers Needing Attention
▶12895	41	22	372.8	3	3	Potential Loyalists

Recommendation

1. **Offer exceptional customer service:** Providing exceptional customer service can create a positive impression on customers and increase the likelihood of their return. Customer service can be improved by training employees to be responsive, helpful, and knowledgeable.

2. **Conduct customer surveys:** Conducting customer surveys can help identify areas for improvement and provide insights into what customers value most. Feedback from surveys can be used to make necessary changes to products, services, and the overall customer experience.

3. **Personalize customer interactions:** Personalizing customer interactions can make customers feel valued and increase their loyalty. Personalization can be achieved by using customer data to tailor marketing messages, offers, and product recommendations to individual customers.

4. **Inventory preparation: The company can prepare for November by ensuring the appropriate and sufficient inventory is available to meet the increased demand during this month.**

5. **Improve shipping and delivery processes: Efforts should be made to improve shipping and delivery processes to meet the demands in a timely manner during November.**

6. **Identify factors influencing returns: Efforts should be made to identify the factors leading to increased returns and to improve processes and products to reduce returns and increase sales.**

7. **Enhance customer experience: The company can enhance the customer experience by providing additional payment and delivery options, improving customer service, and making the customer experience more convenient and comfortable to achieve high customer satisfaction.**