

Northwind Sales Analysis

SQL application for data analysis

December 2024

Trainee :Dai Ahmed Mohamed El-Ashry

Supervised by : Instructor : Asmaa El-Taher

CARERHA – Tech Her Up Program

Data Analysis Track – Group 3

Introduction :

In sales field, data analysis is no longer an option, it's a necessity for staying competitive and achieving sustainable growth, empowering sales teams to gain a deeper understanding of sales patterns, customer behavior and enhance their sales strategies .

SQL (Structured Query Language) is a powerful tool for analyzing sales data as it allows businesses to efficiently store, retrieve, and manipulate large volumes of data , additionally , its versatility and compatibility with various database systems make it a go-to solution for managing and analyzing sales data effectively.

The **Northwind database** included in SQLite simulates the operations of a fictional company, which imports and exports specialty foods from around the world. It is widely used for learning and demonstrating relational database concepts, SQL querying, and business data analysis.

Objectives :

By writing sophisticated SQL queries , the objective of this project is To gain a deeper understanding of sales patterns and customer behavior within the Northwind database , which will help enhancing sales and customer relationships .

This report contains 5 main sections :

1. Customer behavior : Segmenting customers based on their purchase patterns .
2. Product Analysis : Determining best products .
3. Order Analysis : Understanding orders' patterns .
4. Employee performance : Determining employees' performance in terms of sales and orders purchased .
5. Conclusion and Recommendations : an overall conclusion of the analysis process and suggest how can Northwind company enhance the sales strategies and strengthen customer loyalty .

SECTION 1 :Customers behavior :

1.1 RFM (Recency, Frequency, and Monetary) Analysis :

To segment customers to 3 main categories :

- **Champions** : Highest frequency , highest monetary and lowest recency.
- **Potential loyalist** : High frequency or high monetary .
- **At risk** : Any customers else .

Based on RFM analysis as follows :

- **Recency**: How recently a customer made a purchase (Days since last order).
- **Frequency** : How often a customer makes purchases (Total number of orders made) .
- **Monetary** : How much a customer spends (Total amount spent) .

First of all , recency , frequency and monetary were calculated for each customers and their ranges were found as follows :

- Recency : (426: 596) days .
- Frequency : (154: 210) orders .
- Monetary: (3,965,464.95 : 6,154,115.34) dollars

So based on the information , Champions were decided to be those who spent at least 5,500,000 \$ for their total orders , ordered at least 190 times and their last order was purchased at most 490 days ago .

While potential loyalists are those who spent at least 4,500,000 \$ or at least their total number of orders is 164 and customers at risk are anyone else except them.

Second, recursive queries were required to calculate the RFM matrix for customers, segment based on it, and determine which customers do the company have as shown in the following SQL query.

```

--§ 1. RFM ( Recency , Frequency , Monetary ) Analysis:
-- Step 1: Calculate RFM metrics for each customer :
WITH RFM AS (
    SELECT
        C.CustomerID,
        round(JULIANDAY('now') - JULIANDAY(MAX(O.OrderDate)),0) AS Recency,
        COUNT(DISTINCT(O.OrderID)) AS Frequency,
        SUM(OD.UnitPrice * OD.Quantity*(1-OD.discount)) AS Monetary
    FROM Customers C
    INNER JOIN Orders O ON C.CustomerID = O.CustomerID
    INNER JOIN 'Order Details' OD ON O.OrderID = OD.OrderID
    GROUP BY C.CustomerID
),
--Segment Customers to Champions , Potential Loyalist and At risk based on RFM metrics :
Customers_Segmentation AS (
    SELECT
        customerid
        ,Recency
        ,Frequency
        ,Monetary
        ,CASE
            WHEN Recency <= 490 AND Frequency >= 190 AND Monetary >= 5500000 THEN 'Champion'
            WHEN Frequency >= 164 OR Monetary >= 4500000 THEN 'Potential Loyalist'
            ELSE 'At Risk'
        END AS Segment
    FROM
        RFM )
--Define Customers distribution based on their categories.
SELECT *
FROM Customers_Segmentation
GROUP BY 1
HAVING segment = 'Champion'

```

Running the query , the customer where segmented as follows:

Segment	Num. of Customers
Champion	7
Potential Loyalist	75
At Risk	11

Table_1 : Customers' RFM segments .

The majority of customers are potential loyalist, there are 7 special customers who made highest number of orders and spend highest amount .

Customer	Recency	Frequency	Monetary
ANATR	435	195	5534356.65
BSBEV	440	210	6154115.34
FOLIG	443	195	5505502.85
GOURL	445	202	5552309.805
HUNGC	433	198	5698023.67
RANCH	457	194	5559110.08
RICAR	466	203	5524517.31

Table_2 : Champions .

Also there are 11 customers the company may lose since their frequency and monetary value are at lowest , that is a problem that needs to be handled .

1.2 Average Revenue Value :

On average , customers 'order values range from 713 \$ to 755 \$.

Customer were segmented based on average order value the pay to company so that :

- High-Value: Those who pay at least 740\$ for one order on average .
- Medium-Value : Those who pay at least 725\$ for an order on average.
- Low-Value : anyone else .

```
--§ 2. Average Order Value :  
WITH AvgOrderValue AS (  
  SELECT  
    O.CustomerID  
  , avg(OD.UnitPrice * OD.Quantity*(1-OD.discount)) AS AVG_Revenue  
  FROM Orders O  LEFT JOIN 'Order Details' OD ON O.OrderID = OD.OrderID  
  GROUP BY 1  
  ORDER BY 2 DESC ) ,  
AvgRevenue_Segmentation AS (  
  SELECT  
    CustomerID  
  , CASE  
    WHEN AVG_Revenue >= 740 THEN 'High-Value'  
    WHEN AVG_Revenue >= 725 THEN 'Medium-Value'  
    ELSE 'Low-Value'  
    END AS 'AverageRevenue'  
  FROM AvgOrderValue )  
  
  SELECT AverageRevenue , COUNT(customerid) N_Customers  
  FROM AvgRevenue_Segmentation  
  GROUP BY 1
```

Running the query , customers were segmented to :

Average order value	Num. of Customers
High- Value	29
Medium-Value	53
Low-Value	12

Table_3: Average Revenue Segments

On average , the majority of customers' order values are either high or medium values , only 12 customers pay low order values.

SECTION 2 :Product Analysis :

The aim of this section is to understand the sales process of the company's products.

2.1 Revenue Generator Products :

To identify the products that generates the highest revenue values to the company , the sql query below calculates total revenue for each products and obtain the highest 10 as the top revenue generator products .

```
2 $ High Revenue Value: Identify the top 10 revenue generator products.*/
3 SELECT
4     P.ProductName
5     , sum(OD.unitprice* OD.quantity*(1-OD.discount)) REVENUE
6 FROM Products P LEFT JOIN 'Order Details' OD
7 ON P.ProductID = OD.ProductID
8 GROUP BY 1
9 ORDER BY 2 DESC
10 LIMIT 10
```

Product Name	Revenue
Coast of Blaye	53,265,895.235
Thuringian Grilled Sausage	24,623,469.232
Mishi Kobe Beef	19,423,037.5
Sir Rodney's Marmalade	16,653,807.36
Carnarvon Tigers	12,604,671.875
Courdavault Raclette	11,216,410.7
Manjimup Dried Apples	10,664,768.65
Sugar Pie	9,952,936.07
Ipoh Coffee	9,333,374.7
Little Horse Sauerkraut	9,252,765.44

Table_4:Top 10 revenue generator products

Coat of Bayle - Probably wine - is the best revenue generating product with **53 million dollars** generated for the company , followed by the **Thuringian Grilled Sausage** that generated **about 24 million dollars** .

2.2 Top 10 Sales Volume :

This part aims to identify the highest frequently ordered products so that the Northwind traders can identify the most popular products among customers .

By Counting how many times each product was ordered with SQL query, the top 10 highly ordered products were generated as follows :

```
/*§ High Sales Volume: Determine the top 10 most frequently ordered products.*/  
SELECT  
    P.ProductName  
    , COUNT(orderid) Num_Orders  
FROM Products P LEFT JOIN 'Order Details' OD  
ON P.ProductID = OD.ProductID  
GROUP BY 1  
ORDER BY 2 DESC  
LIMIT 10
```

Product Name	Order Frequency
Louisiana Hot Spiced Okra	8040
Teatime Chocolate Biscuits	8024
Outback Lager	8020
Sir Rodney's Marmalade	7999
Gumbär Gummibärchen	7999
Gudbrandsdal Cheese	7991
Courdavault Raclette	7982
Angelo Ravioli	7969
Konbu (Kelp)	7968
Telino Gorgonzola (Italian blue cheese)	7964

Table_5:Top 10 popular products

The most popular product among customers is the Louisiana Hot Spiced Okra that it's ordered 8040 times , followed by chocolate biscuits (Teatime) that is ordered 8024 times in whole time interval .

2.3 Top 5 Slow Movers :

As It's important for traders to identify the most popular products , it is also very important to highlight the lowest popular products among customers that are not highly ordered .

The following query was to gain the lowest 5 products in terms of how many times they were ordered and those products were found as follows:

```
SELECT
    P.ProductName
    , COUNT(OD.orderid) Num_Orders
FROM Products P LEFT JOIN 'Order Details' OD
ON P.ProductID = OD.ProductID
GROUP BY 1
ORDER BY 2 ASC
LIMIT 5
```

ProductName	N. of Orders
Zaan Cookies	7785
Fabioli Mascarpone	7817
Perth Pasties	7817
Scottish Longbreads	7818
Tofu	7834

Table_6:Top 5 Slow Mover Products

From table(6): Zaan Cookies is worst products in terms of total number of orders made .

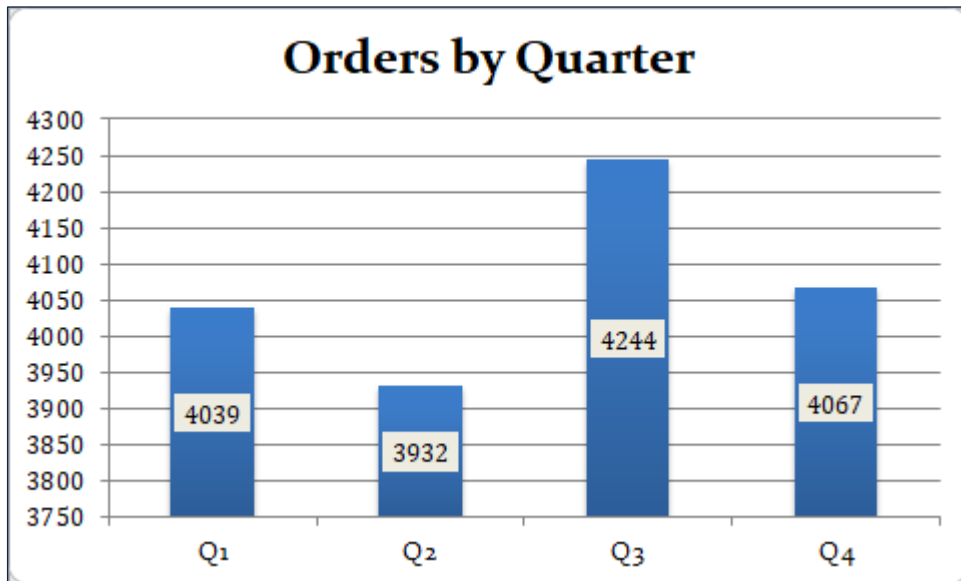
SECTION 3 : Order Analysis :

This section focuses on analyzing order trends and determining when orders are at highest and lowest times .

3.1 Seasonality:

First , a sql query was written to obtain the quarters of the year , then for each quarter , the total number of orders was calculated so that the quarter effect on order trend can be shown .

```
$ Seasonality: Identify any seasonal fluctuations in order volume.*/  
SELECT  
    CASE  
        WHEN 0 + strftime('%m', O.OrderDate) BETWEEN 1 AND 3 THEN 'Q1'  
        WHEN 0 + strftime('%m', O.OrderDate) BETWEEN 4 AND 6 THEN 'Q2'  
        WHEN 0 + strftime('%m', O.OrderDate) BETWEEN 7 AND 9 THEN 'Q3'  
        WHEN 0 + strftime('%m', O.OrderDate) BETWEEN 10 AND 12 THEN 'Q4'  
    END AS quarter  
    , COUNT (O.OrderID) No_Orders  
FROM Orders O  
GROUP BY 1
```



Figure_1:Orders by Quarter

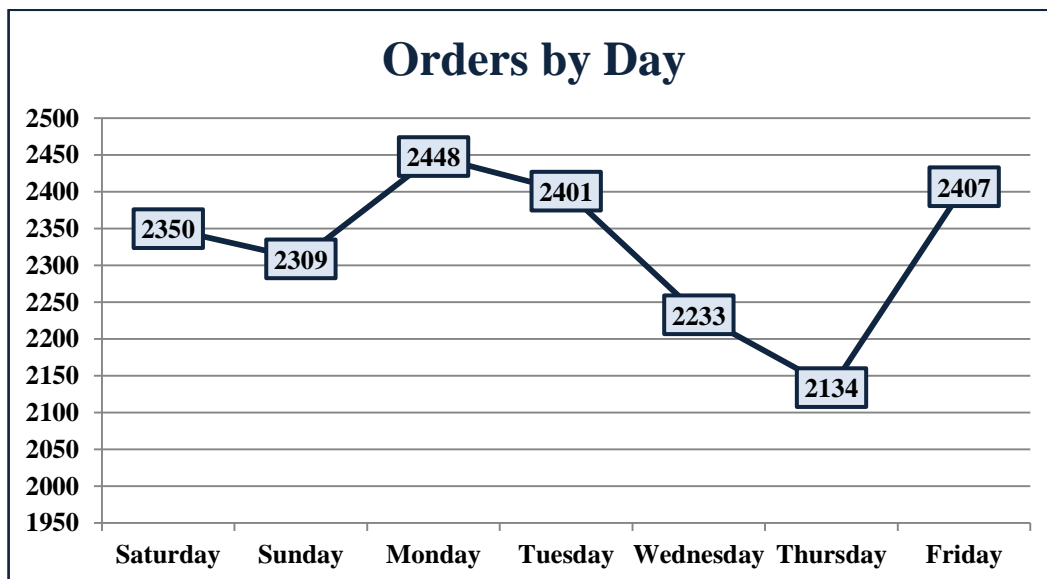
As figure_1 shows , The orders gets higher in the first quarter, then decreases in the second quarters in the third quarter , but it increases to the highest possible value in the third quarter but decreases again in the fourth quarter.

Generally , the orders increases in the first (January , February , may) and third quarter (July , August , September) , but it decreases in the second quarter (April , May ,June) and fourth quarter (October , November , December)

3.2 Day-of-the-Week Analysis:

Determine the most popular order days for customers, by calculating number of orders by day

```
/*§ Day-of-the-Week Analysis: Determine the most popular order days.*/  
SELECT  
    strftime('%w',O.OrderDate) Day  
    , COUNT (O.OrderID) No_Orders  
FROM Orders O  
GROUP BY 1  
--from 0 : sunday to 6 : saturday
```



Figure_2 : Orders by day

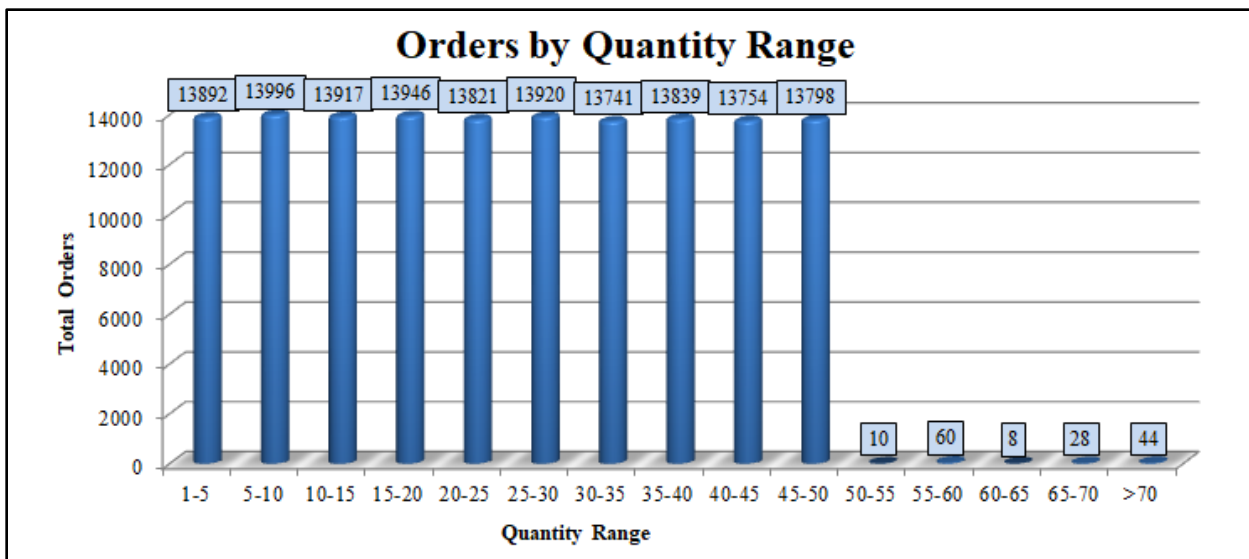
From figure_2 : Monday is the day when orders reach the highest limit , while orders decrease after Monday until they reach 2134 orders in Thursday , then increases to 2407 in Friday .

3.3 Order Size Analysis:

A SQL query was used to analyze the distribution of order quantities by creating quantity ranges. The query calculated the number of orders that fall within each range to identify the most frequently ordered quantities.

```
/*$ Order Size Analysis: Analyze the distribution of order quantities.
*/
SELECT
  CASE
    WHEN quantity BETWEEN 1 AND 5 THEN '1-5'
    WHEN quantity BETWEEN 5 AND 10 THEN '5-10'
    WHEN quantity BETWEEN 10 AND 15 THEN '10-15'
    WHEN quantity BETWEEN 15 AND 20 THEN '15-20'
    WHEN quantity BETWEEN 20 AND 25 THEN '20-25'
    WHEN quantity BETWEEN 25 AND 30 THEN '25-30'
    WHEN quantity BETWEEN 30 AND 35 THEN '30-35'
    WHEN quantity BETWEEN 35 AND 40 THEN '35-40'
    WHEN quantity BETWEEN 40 AND 45 THEN '40-45'
    WHEN quantity BETWEEN 45 AND 50 THEN '45-50'
    WHEN quantity BETWEEN 50 AND 55 THEN '50-55'
    WHEN quantity BETWEEN 55 AND 60 THEN '55-60'
    WHEN quantity BETWEEN 60 AND 65 THEN '60-65'
    WHEN quantity BETWEEN 65 AND 70 THEN '65-70'
    WHEN quantity > 70 THEN '>70'
    ELSE 'nulll'
  END AS 'QuantityRange'
, COUNT (orderid) Orders
FROM 'Order Details'
GROUP BY 1
```

Visualizing the results in the following graph :



Figure_3 :Quantity Distribution

As figure_3 shows , The majority of orders lay in the range (5-10) items followed by range (1-5) items , Also there are orders than have a very high quantity .

Section4 : Employee Performance:

4.1 Revenue generator Employees :

Determining employees performance in terms of total revenue generated :

```
--$ Total sales volume generated:
SELECT
    CONCAT(E.firstname, ' ', E.lastname) Employee
    , sum(OD.unitprice* OD.quantity*(1-OD.discount)) REVENUE
FROM Employees E INNER JOIN Orders O
ON E.EmployeeID = O.EmployeeID
INNER JOIN 'Order Details' OD
ON O.OrderID = OD.OrderID
GROUP BY 1
ORDER BY 2 DESC
```

Employee	Revenue
Margaret Peacock	51,488,395
Steven Buchanan	51,386,459
Janet Leverling	50,445,574
Nancy Davolio	49,659,423
Robert King	49,651,899
Laura Callahan	49,281,137
Michael Suyama	49,139,967
Anne Dodsworth	49,019,678
Andrew Fuller	48,314,101

Table_7: revenue by employees

Margret Peacock and Steven Buchannan are highest revenue generators , while Andrew Fuller generated the lowest total revenue value .

4.2 Order Achieving Employees :

The aim of this part is to evaluate employee performance based on total number of orders achieved , the following query is to calculate the total orders purchased by each employee .

```
--$ Number of orders processed:
SELECT
    CONCAT(E.firstname, ' ', E.lastname) Employee
    , COUNT (O.OrderID) No_Orders
FROM Employees E INNER JOIN Orders O
ON E.EmployeeID = O.EmployeeID
GROUP BY 1
ORDER BY 2 DESC
```

Employee	Orders
Margaret Peacock	1908
Nancy Davolio	1846
Janet Leverling	1846
Steven Buchanan	1804
Laura Callahan	1798
Robert King	1789
Andrew Fuller	1771
Anne Dodsworth	1766
Michael Suyama	1754

Table_8 : orders by each employee

Margret Peacock is still the number one employee in terms of total orders purchased, followed by Nancy Davoilio , Steven Buchanan is the second employee in terms of revenue generating but he is the fourth employee in terms of orders purchased .

4.3 Average Order Value :

In this part , the performance of employees will be evaluated by the average order value that each employee can generate , the SQL query obtained each employee's average order value.

```
--$ Average order value :  
SELECT  
    CONCAT(E.firstname, ' ', E.lastname) Employee  
    , avg(OD.unitprice* OD.quantity*(1-OD.discount)) AvgOrderValue  
FROM Employees E INNER JOIN Orders O  
ON E.EmployeeID = O.EmployeeID  
INNER JOIN 'Order Details' OD  
ON O.OrderID = OD.OrderID  
GROUP BY 1  
ORDER BY 2 DESC
```

Employee	Avg_Value per order
Michael Suyama	742.41
Janet Leverling	739.17
Anne Dodsworth	738.67
Robert King	737.10
Margaret Peacock	736.90
Steven Buchanan	735.48
Nancy Davolio	734.40
Laura Callahan	731.16
Andrew Fuller	728.00

Table_9: Average Order value generated by employee

Michael Suyama has the highest average order value among all employees, while Margaret Peacock ranks fifth in average order value compared to her colleagues.

Section 5 : Conclusions and recommendations :

5.1 Conclusions :

The Analysis showed that :

1. Each Northwind 's customer at least ordered 150 times, which gives a good indicator for a the customers' trust and loyalty .
2. With Customers segmentation , Majority are either champion or potential loyalist and pay for one order either high or medium values.
3. There are 11 customers Northwind traders are at risk to lose due to low frequency or monetary value .
4. Customers are more likely to buy spiced okra , chocolate biscuits and Outback Lager since each product was about 8000 times ordered , while zaan cookies and mascarpone are the most unpopular compared to them since they are 7700 times ordered.
5. Comparing Order frequency of highest and lowest ordered products, the gap is not wide , which means that all products of Northwind are highly popular among customers .
6. **The best seller :** (Coat of bale) generates significantly higher revenue than other products, despite not being among the top 10 ordered selected .
7. The third quarter is the most busiest period for customers' order, followed by the first quarter .
8. Monday is the highest in terms of order frequency, suggesting peak activity at th start of the week .
9. On average , all employees generate the same order value – all around 730 \$ - but Margret Peacock is the best employee in terms of both revenue and order frequency.

5.2 Recommendations :

To improve sales strategies and customers' loyalty :

- 1. For 'Champion' customers :** It's important to offer them an exclusive special offers and personalized gifts for strengthen long-term relationships and reinforce their value to the company.
- 2. For 'At Risk' customers :** The company should establish effective and proactive communications – specially with them - to understand the reasons for their reduced engagement or low orders .
This will Identify and address any potential issues related to company policies, product quality, or service to rebuild trust and improve satisfaction.
- 3. For slow mover products:** It's better to create feedback forms to customers to highlight the concerns about them and offer special offers so that we increase their order frequency .
- 4.** When Launching marketing campaigns for low popular products , highlight the top 10 popular products in offers to attract higher number of customers.
- 5.** Given its high revenue generation, focus marketing efforts on **the Coat of Bale** to increase its order frequency.
- 6. For seasonal and weekly trends :**
 - Launch seasonal campaigns specially in the third quarter to maximize sales volume .
 - Schedule weekly offers specially on Monday and Friday to encourage customers to increase their order volume .
- 7. For improving employees performance :**
 - Reward Margret Peacock for her outstanding performance to that she will keep her level of performance , also this will encourage other employees to enhance their strategies .
 - Provide effective training programs from high-performed employees like Margret to ensure consistence of the team's improvement.