

# **COURSE PROJECT**

## **Northwind Traders Sales and Customer Behavior Analysis**

**By: Group\_6**

**Fatima, Naaz (NF1020595)**

**Tirumala Narasimhan, Archana (NF1027045)**

**Prajapati, Charmiben Dineshkumar (NF1014304)**

**Data Analytics Case Study 1 (DAMO-501-5)**

Master in Data Analytics, University of Niagara Falls



**Professor: Omid Isfahanialamdari**

March 16, 2025

## Table of Contents

<b>1</b>	<b>INTRODUCTION .....</b>	<b>4</b>
<b>2</b>	<b>OBJECTIVE .....</b>	<b>4</b>
<b>3</b>	<b>PROJECT PHASES .....</b>	<b>5</b>
<b>4</b>	<b>CHAPTER 1: PROBLEM DEFINITION AND RESEARCH .....</b>	<b>5</b>
4.1	PROBLEM DEFINITION: .....	5
4.2	RESEARCH QUESTION .....	6
4.3	JUSTIFICATION AND RELEVANCE.....	7
<b>5</b>	<b>CHAPTER 2: HYPOTHESES FORMULATION.....</b>	<b>7</b>
5.1	JUSTIFICATION OF HYPOTHESES .....	8
<b>6</b>	<b>CHAPTER 3. DATA COLLECTION AND SQL QUERIES .....</b>	<b>8</b>
6.1	CUSTOMER PURCHASING BEHAVIORS .....	8
6.1.1	SQL Query .....	9
6.2	IMPACT OF SEASONS .....	10
6.2.1	SQL Query .....	11
6.3	FACTORS AFFECTING THE SALES PERFORMANCE.....	12
6.3.1	SQL Query .....	12
6.4	IMPACT OF DIFFERENT TERRITORIES AND SHIPPING PROVIDERS .....	15
6.4.1	SQL Query .....	15
<b>7</b>	<b>CHAPTER 4. DATA UNDERSTANDING .....</b>	<b>17</b>
7.1	THOROUGH EXPLORATION AND ANALYSIS OF THE COLLECTED DATA .....	17
7.1.1	SQL Query-1 Customer purchasing behaviors.....	17
7.1.2	SQL Query-2 Impact of Seasons .....	17
7.1.3	SQL Query-3 Factors affecting the sales performance.....	17
7.1.4	SQL Query-4 Impact of different territories and shipping providers .....	17
7.2	IDENTIFICATION OF KEY METRICS, TRENDS, AND PATTERNS .....	18
7.2.1	SQL Query-1 Customer purchasing behaviors Key-Insights .....	18
7.2.2	SQL Query-2 Impact of Seasons Key-Insights.....	18
7.2.3	SQL Query-3 Factors affecting the sales performance Key-Insights .....	19
7.2.4	SQL Query-4 Impact of different territories and shipping providers Key-Insights.....	19
<b>8</b>	<b>CHAPTER 5. DATA VISUALIZATION .....</b>	<b>20</b>
8.1	CUSTOMER PURCHASING BEHAVIORS .....	20
8.2	IMPACT OF SEASONS .....	22
8.3	FACTORS AFFECTING THE SALES PERFORMANCE.....	23
8.4	IMPACT OF DIFFERENT TERRITORIES AND SHIPPING PROVIDERS .....	25
<b>9</b>	<b>CHAPTER 6. MODEL BUILDING .....</b>	<b>27</b>
9.1	DATA MODELING .....	27
9.2	MODELING TECHNIQUES USED:.....	27
<b>10</b>	<b>CHAPTER 7. MODEL EVALUATION .....</b>	<b>31</b>

10.1	HYPOTHESIS 1:.....	31
10.2	HYPOTHESIS 2:.....	32
10.3	HYPOTHESIS 3:.....	33
10.4	HYPOTHESIS 4:.....	33
<b>11</b>	<b>CONCLUSION: .....</b>	<b>35</b>
<b>12</b>	<b>RECOMMENDATIONS: .....</b>	<b>36</b>
<b>13</b>	<b>REFERENCES .....</b>	<b>36</b>
<b>14</b>	<b>APPENDIX .....</b>	<b>37</b>

# **1 Introduction**

For any organization or a business to be successful, data driven decision making plays an important role. The 4 Ps (Product, Price, Place, and Promotion) of marketing plays an important role in successful business of an organization. The Northwind Trading Company is a global supplier of food and beverage products (teas, syrups, seasonings, seafood, dairy products, baked goods, meats, chocolates, and gourmet delicacies). With the availability of structured database, maintained by the company, which provides an insight into its business operations, including customers, orders, products, employees, suppliers, and shipping logistics.

In order to identify the top-selling products, understanding product demand across different regions, and improving publicity strategies, The knowledge acquired in SQL queries, data visualization, and marketing analytics can be used to analyze the 4Ps from the data set available with Northwind Trading Company.

By analyzing these 4Ps, the valuable recommendations can be outlined which can help in strategic decision-making with respect to marketing strategies, optimizing sales efforts, and boosting overall business performance.

# **2 Objective**

The objective of this case study is to demonstrate the practical application of data analytics techniques in addressing real-world business challenges, enhancing marketing strategies, and improving Northwind Trading Company's overall sales performance, by defining the problem under study, stating the research questions and the hypothesis under investigation and conducting the phases required to investigate data analytics problems including data collection using SQL queries, data understanding, data visualization, model building and evaluation.

### **3 Project Phases**

The various phases of this project case study are:

- **Data Collection:** The relevant data (like information on customers, orders, products, and marketing activities) from the Northwind database is to be extracted by using the SQL queries.
- **Data Understanding:** Customer groups, product sales, order trends, and the efficacy of marketing campaigns can all be perceived by looking at the data that has been gathered.
- **Data Visualization:** It is simpler to identify trends, patterns, and connections quickly when the analyzed data is transformed into visual formats such as pie charts, bar graphs, and pivot tables.
- **Model Building:** Customer buying patterns, sales trends and the effectiveness of marketing campaigns can be anticipated by creating predictive models from past data.
- **Model Evaluation:** The model building can be validated using key metrics like accuracy, precision, recall and F1-score, through which it can be understood how effectively the research questions are answered and hypothesis can also be tested at the same time.

## **4 Chapter 1: Problem Definition and Research**

### **4.1 Problem Definition:**

Northwind Trading Company operates in a constantly evolving market, where sales trends, customer preference, and performance indicators must be tracked to ensure informed decisions are made. Through data analysis, the company can extract valuable information to refine its marketing strategies, increase sales, and develop long-term business growth.

This case study looks into real-world business challenges by exploring customer buying habits, seasonal sales trends, customer groups, and how shipping choices impact order volumes.

By clearly defining the problem and analyzing data using SQL queries, visualizations, and predictive models, Northwind Trading Company can aim to:

- Identify the factors that affect customers' purchasing decisions, such as product categories and order frequency.
- Identify seasonal trends in sales, which can be used to enhance inventory and promotion strategies.
- Understand how consumers are grouped based on their buying habits and how this influences total sales.
- Assess the impact of shipper choice on total sales, evaluating how different regions prefer specific shipping methods.
- Enhance sales forecasting and performance metrics to improve revenue generation.

By following a clear process of gathering, analyzing, and modeling data, this study will uncover practical insights that help Northwind Trading Company make smarter decisions and achieve long-term business growth.

## **4.2 Research Question**

To establish the trends and correlations the following questions to be addressed:

- What are the customer purchasing behaviors in Northwind Trading Company?
- What are the impacts of different seasons on sales performance?
- What factors affect the sales performance among products and regions?
- How do different territories and shipping providers impact on total sales performance?

### 4.3 Justification and Relevance

The research questions are important in understanding business growth and customer retention in an organization. Thus, analyzing purchasing behaviors and sales trends the company can:

- Improve marketing strategies to suit particular customer groups. (for example, female customers can be targeted for beauty products with the proper advertisements).
- Based on demand patterns, it can help in maintaining inventories across different regions.
- It is possible to engage and retain customers, and achieve consistent sales performance of the products by offering promotions and discounts to the customers.
- With proper allocation of the resources towards high-performing products, profitability can be maximized.

## 5 Chapter 2: Hypotheses Formulation

Based on the analysis of the data set the following Null & Alternate hypotheses can be drawn:

- **Null Hypothesis ( $H_0$ ):** Customers who frequently purchase from different product categories do not show consistent ordering patterns.

**Alternate Hypothesis ( $H_1$ ):** Customers who frequently purchase from specific product categories tend to have consistent and repetitive ordering behaviors.

- **Null Hypothesis ( $H_0$ ):** Sales performance remains constant throughout the year, with no significant seasonal impact.

**Alternate Hypothesis ( $H_1$ ):** Sales performance fluctuates significantly across different seasons, with certain periods showing a increase in sales.

- **Null Hypothesis (H<sub>0</sub>):** Some customer segmentation does not significantly impact total sales revenue than others based on their purchasing behavior.

**Alternate Hypothesis (H<sub>1</sub>):** Some customer segments contribute significantly towards total sales revenue than others based on their purchasing behavior.

- **Null Hypothesis (H<sub>0</sub>):** The choice of shipper does not influence total sales, and order volumes remain uniform across regions.

**Alternate Hypothesis (H<sub>1</sub>):** Shipper choice has a significant effect on total sales, with different territories/regions exhibiting varied order volumes based on shipping preferences.

## 5.1 Justification of Hypotheses

These hypotheses are based on past sales trends (extracted from the data set) and marketing principles. These hypotheses can be verified to see if they are true by examining data and visual results. Verification of these hypotheses will assist the company in making strategic choices, enhance client relationships, modify pricing strategies, and increase overall revenue.

## 6 Chapter 3. Data Collection and SQL Queries

Data is extracted from the Northwind database using SQL queries and build a dataset, which is to be used in subsequent chapters. The Northwind database contains information/data about products, suppliers, shippers, orders, employees, customers, and regions & territories, by using appropriate SQL commands ( SELECT, JOIN (INNER JOIN, LEFT JOIN), WHERE, GROUP BY, ORDER BY, HAVING) a dataset can be build.

### 6.1 Customer purchasing behaviors

Customer purchasing behaviors (like, Customer purchase history (Order details), Order frequency per customer, Average Quantity Per Order, Number of repeat orders by customer) in



Northwind Trading Company can be analyzed if the data related to Total Orders, Average Quantity Per Order, Total Repetitive Orders is made available.

### 6.1.1 SQL Query

Use northwindfull;

```
WITH CategoryFrequency AS (
    SELECT
        c.CustomerID,
        p.CategoryID,
        COUNT(o.OrderID) AS Category_Order_Frequency
    FROM
        Customers c
    JOIN
        Orders o ON c.CustomerID = o.CustomerID
    JOIN
        orderdetails od ON o.OrderID = od.OrderID
    JOIN
        Products p ON od.ProductID = p.ProductID
    GROUP BY
        c.CustomerID, p.CategoryID
    HAVING
        COUNT(o.OrderID) >= 3 -- Directly filter consistent category purchases
)
SELECT
    c.CustomerID,
    COUNT(DISTINCT o.OrderID) AS Total_Orders,
    COUNT(DISTINCT p.CategoryID) AS Distinct_Categories,
    ROUND(AVG(od.Quantity), 2) AS Avg_Quantity_Per_Order,
    COUNT(DISTINCT cf.CategoryID) AS Frequent_Categories, -- Directly counts consistent
categories
    SUM(cf.Category_Order_Frequency) AS Total_Repetitive_Orders
FROM
    Customers c
JOIN
    Orders o ON c.CustomerID = o.CustomerID
JOIN
    orderdetails od ON o.OrderID = od.OrderID
JOIN
    Products p ON od.ProductID = p.ProductID
LEFT JOIN
    CategoryFrequency cf
    ON c.CustomerID = cf.CustomerID
GROUP BY
    c.CustomerID
```

## ORDER BY

Frequent\_Categories DESC, Total\_Orders DESC;

The screenshot shows the SQL Server Enterprise Manager interface. The left pane displays the 'Schemas' tree with 'orders' expanded. The main pane shows a SQL query in the 'SQL File 01.sql' editor. The query is as follows:

```
1. [use northwindfull]
2. WITH CategoryFrequency AS (
3.     SELECT
4.         c.CustomerID,
5.         p.CategoryID,
6.         COUNT(o.OrderID) AS Category_Order_Frequency
7.     FROM
8.         Customers c
9.     JOIN
10.        Orders o ON c.CustomerID = o.CustomerID
11.     JOIN
12.        orderdetails od ON o.OrderID = od.OrderID
13.     JOIN
14.        Products p ON od.ProductID = p.ProductID
15. )
16. SELECT
17.     c.CustomerID,
18.     SUM(od.Quantity) AS Total_Orders,
19.     COUNT(DISTINCT p.CategoryID) AS Distinct_Categories,
20.     ROUND(SUM(od.Quantity) / SUM(od.OrderID), 2) AS Avg_Quantity_Per_Order,
21.     COUNT(DISTINCT p.CategoryID) AS Frequent_Categories,
22.     COUNT(o.OrderID) AS Total_Repetitive_Orders
23. FROM Customers c
24. JOIN Orders o ON c.CustomerID = o.CustomerID
25. JOIN orderdetails od ON o.OrderID = od.OrderID
26. JOIN Products p ON od.ProductID = p.ProductID
27. GROUP BY c.CustomerID
28. ORDER BY Frequent_Categories DESC, Total_Orders DESC
```

The bottom pane shows the 'Results' grid with the following data:

CustomerID	Total_Orders	Distinct_Categories	Avg_Quantity_Per_Order	Frequent_Categories	Total_Repetitive_Orders
SAVIA	31	8	42.74	8	13405
BRNGH	30	8	44.54	8	10404
QUICK	28	8	46.06	8	7396
HARGO	19	8	30.62	8	3025
FOLKO	19	8	27.42	8	2025

The bottom pane also shows the 'Output' window with the following messages:

```
34 13:13:44 use northwindfull (0 rows) affected
35 13:13:44 SELECT C.Region AS Territory, S.ShipperName AS ShippingProvider, SUM(OD.Quantity) AS Total_Orders, COUNT(DISTINCT OD.CategoryID) AS Distinct_Categories, ROUND(SUM(OD.Quantity) / SUM(OD.OrderID), 2) AS Avg_Quantity_Per_Order, COUNT(OD.CategoryID) AS Frequent_Categories, COUNT(O.OrderID) AS Total_Repetitive_Orders (53 rows) returned
36 22:26:53 use northwindfull (0 rows) affected
37 22:26:53 WITH CategoryFrequency AS ( SELECT c.CustomerID, p.CategoryID, COUNT(O.OrderID) AS Category_Order_Frequency FROM Customers c JOIN Orders o ON c.CustomerID = o.CustomerID JOIN orderdetails od ON o.OrderID = od.OrderID JOIN Products p ON od.ProductID = p.ProductID GROUP BY c.CustomerID ORDER BY Frequent_Categories DESC, Total_Orders DESC ) (89 rows) returned
```

This SQL query gathers information needed to study customer behavior by looking into Total Orders, Average Quantity Per Order, and Total Repetitive Orders.

- **Total\_Orders:** The total number of distinct orders placed by each customer.
- **Distinct\_Categories:** The number of distinct product categories purchased by each customer.
- **Avg\_Quantity\_Per\_Order:** The average quantity of items ordered per order, rounded to two decimal places.
- **Frequent\_Categories:** The count of product categories where the customer has made at least 3 orders.
- **Total\_Repetitive\_Orders:** The total count of orders that were made for categories.

## 6.2 Impact of Seasons

A study on the effect of the different seasons on sales performance (determining which seasons sell the most, determining which season a particular category of products sells more), would be feasible if information about total sales by seasons and monthly revenue is gathered.

## 6.2.1 SQL Query

Use northwindfull;

SELECT

YEAR(o.OrderDate) AS Year,

MONTH(o.OrderDate) AS Month,

CASE

WHEN MONTH(o.OrderDate) IN (12, 1, 2) THEN 'Winter'

WHEN MONTH(o.OrderDate) IN (3, 4, 5) THEN 'Spring'

WHEN MONTH(o.OrderDate) IN (6, 7, 8) THEN 'Summer'

WHEN MONTH(o.OrderDate) IN (9, 10, 11) THEN 'Fall'

END AS Season,

c.CategoryName,

SUM(od.Quantity \* od.UnitPrice \* (1 - od.Discount)) AS TotalSales

FROM Orders AS o

JOIN OrderDetails AS od

ON o.OrderID = od.OrderID

JOIN Products AS p

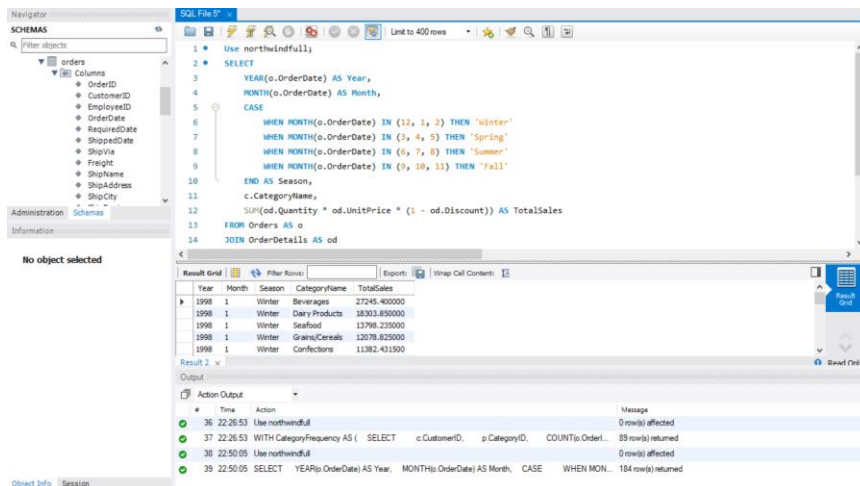
ON od.ProductID = p.ProductID

JOIN Categories AS c

ON p.CategoryID = c.CategoryID

GROUP BY YEAR(o.OrderDate), MONTH(o.OrderDate), Season, c.CategoryName

ORDER BY Year DESC, Month, TotalSales DESC;



Year	Month	Season	CategoryName	TotalSales
1998	1	Winter	Beverages	27245.400000
1998	1	Winter	Dairy Products	18303.850000
1998	1	Winter	Seafood	13796.220000
1998	1	Winter	Grains/Cereals	12078.820000
1998	1	Winter	Confections	11382.410000

This SQL query gathers information needed to study the impact of different seasons on sales by looking into Year, Month, Season, CategoryName, & TotalSales.

- **Year:** Represents the year in which the order was placed.
- **Month:** Indicates the month of the order date (from January to December or 1 to 12).

- **Season:** Categorizes the month into seasons:
  - **Winter** (December, January, February)
  - **Spring** (March, April, May)
  - **Summer** (June, July, August)
  - **Fall** (September, October, November)
- **CategoryName:** The name of the product category (from the Categories table).
- **TotalSales:** Represents the total sales amount calculated for each combination of year, month, season, and category. It is computed using the formula:

$$\text{TotalSales} = \text{SUM}(\text{Quantity} \times \text{UnitPrice} \times (1 - \text{Discount}))$$

## 6.3 Factors affecting the sales performance

To find the factors that affect the sales performance among products and regions, (by finding out which regions contribute the most to sales, Identifying which products category perform best in different regions), This can be achieved if the extracted data provides the Best-selling products and their sales performance, Sales performance by region, and Revenue contribution by different categories.

### 6.3.1 SQL Query

```
WITH SalesData AS (
  SELECT
    c.CustomerID,
    c.CompanyName,
    o.OrderID,
    o.OrderDate,
    p.ProductID,
    p.ProductName,
    p.CategoryID,
    cat.CategoryName,
    od.UnitPrice * od.Quantity AS Total_Sales,
    r.RegionDescription AS Region
  FROM
    Customers c
```

```

JOIN
    Orders o ON c.CustomerID = o.CustomerID
JOIN
    orderdetails od ON o.OrderID = od.OrderID
JOIN
    Products p ON od.ProductID = p.ProductID
JOIN
    Categories cat ON p.CategoryID = cat.CategoryID
JOIN
    Employees e ON o.EmployeeID = e.EmployeeID
JOIN
    EmployeeTerritories et ON e.EmployeeID = et.EmployeeID
JOIN
    Territories t ON et.TerritoryID = t.TerritoryID
JOIN
    Region r ON t.RegionID = r.RegionID
)

SELECT
    sd.CustomerID,
    sd.CompanyName,
    COUNT(DISTINCT sd.OrderID) AS Total_Orders,
    COUNT(DISTINCT sd.CategoryID) AS Distinct_Categories,
    SUM(sd.Total_Sales) AS Total_Sales,
    sd.Region,
    sd.CategoryName,
    CASE
        WHEN SUM(sd.Total_Sales) > 10000 THEN 'High-Value Customer'
        WHEN COUNT(DISTINCT sd.OrderID) >= 10 THEN 'Frequent Buyer'
        ELSE 'Regular Customer'
    END AS Customer_Segment
FROM
    SalesData sd
GROUP BY
    sd.CustomerID, sd.CompanyName, sd.Region, sd.CategoryName
ORDER BY
    Total_Sales DESC, Total_Orders DESC;

```

SQL File P

```

1 Use northwindfull;
2 WITH SalesData AS (
3     SELECT
4         c.CustomerID,
5         c.CompanyName,
6         o.OrderID,
7         o.OrderDate,
8         p.ProductID,
9         p.ProductName,
10        p.CategoryID,
11        cat.CategoryName,
12        od.UnitPrice * od.Quantity AS Total_Sales,
13        r.RegionDescription AS Region
14    FROM

```

CustomerID	CompanyName	Total_Orders	Distinct_Categories	Total_Sales	Region	CategoryName	Customer_Segment
QUICK	QUICK-Stop	7	1	149357.2000	Eastern	Beverages	High-Value Customer
SAVEA	Save-a-lot Markets	5	1	120295.0000	Western	Dairy Products	High-Value Customer
WEREP	W&W Produce	1	1	111932.0000	Western	Beverages	High-Value Customer
PICCO	Piccolo and mehr	2	1	106975.0000	Western	Beverages	High-Value Customer
SAVEA	Save-a-lot Markets	3	1	98791.2000	Western	Meat/Poultry	High-Value Customer

Result 3

Output

#	Time	Action	Message
38	22:50:05	Use northwindfull	0 row(s) affected
39	22:50:05	SELECT YEAR(p.OrderDate) AS Year, MONTH(p.OrderDate) AS Month, CASE WHEN MON...	154 row(s) returned
40	23:08:11	Use northwindfull	0 row(s) affected
41	23:08:11	WITH SalesData AS ( SELECT c.CustomerID, c.CompanyName, o.OrderID, o.Orde...	1148 row(s) returned

This query is used to extract the data to understand the factors affecting the sales performance.

- **CustomerID:** A unique identifier for each customer.
- **CompanyName:** The name of the company associated with each customer.
- **Total\_Orders:** The count of distinct orders placed by each customer.
- **Distinct\_Categories:** The count of distinct product categories purchased by each customer.
- **Total\_Sales:** The sum of all sales amounts for each customer, calculated as the total sales from all products they ordered.
- **Region:** The description of the region in which the customer is located, derived from the territories and regions.
- **CategoryName:** The name of the product category that is associated with the sales data.
- **Customer\_Segment:** A derived categorical field that segments customers based on their purchasing behavior:
  - **'High-Value Customer':** Customers whose total sales exceed \$10,000.
  - **'Frequent Buyer':** Customers who have made 10 or more distinct orders.

- **'Regular Customer'**: Customers who do not meet the criteria for the other segments.

## 6.4 Impact of different territories and shipping providers

To analyze, how do different territories and shipping providers impact on total sales performance (by determining which shippers handle the highest sales volume, Evaluating whether shipping costs affect sales trends, See how different times of the year affect regional sales performance).

### 6.4.1 SQL Query

```
Use northwindfull;
SELECT
    t.TerritoryID,
    t.TerritoryDescription,
    r.RegionDescription,
    sh.ShipperID,
    sh.ShipperName,
    COUNT(o.OrderID) AS TotalOrders,
    SUM(od.Quantity * od.UnitPrice * (1 - od.Discount)) AS TotalSales,
    AVG(TIMESTAMPDIFF(DAY, o.OrderDate, o.ShippedDate)) AS AvgShippingTime
FROM Territories AS t
JOIN Region AS r
    ON t.RegionID = r.RegionID
JOIN EmployeeTerritories AS et
    ON t.TerritoryID = et.TerritoryID
JOIN Employees AS e
    ON et.EmployeeID = e.EmployeeID
JOIN Orders AS o
    ON e.EmployeeID = o.EmployeeID
JOIN orderdetails AS od
    ON o.OrderID = od.OrderID
JOIN Shippers AS sh
    ON o.ShipVia = sh.ShipperID
GROUP BY
    t.TerritoryID,
    t.TerritoryDescription,
    sh.ShipperID,
    sh.ShipperName,
    r.RegionDescription
ORDER BY TotalSales DESC;
```

The screenshot shows a SQL Server Enterprise Manager interface. On the left, the 'Schemas' pane is open, showing a tree view of the database structure. The main window displays a SQL query in the 'SQL File 1' tab. The query is a SELECT statement that joins the 'Territories' table with the 'Region' and 'EmployeeTerritories' tables. It calculates various metrics for each territory, including the total number of orders, total sales, and average shipping time. The results are displayed in a grid below the query. The grid has columns for TerritoryID, TerritoryDescription, RegionDescription, ShipperID, ShipperName, TotalOrders, TotalSales, and AvgShippingTime. The data shows four territories: Rockville, Greensboro, Cary, and Atlanta, all of which are in the Eastern region and use United Package as the shipper. Savannah is in the Southern region and also uses United Package. The results show that Rockville and Greensboro have the highest total sales, while Savannah has the highest average shipping time.

TerritoryID	TerritoryDescription	RegionDescription	ShipperID	ShipperName	TotalOrders	TotalSales	AvgShippingTime
20852	Rockville	Eastern	2	United Package	383	106642.532500	9.0747
27403	Greensboro	Eastern	2	United Package	383	106642.532500	9.0747
27511	Cary	Eastern	2	United Package	383	106642.532500	9.0747
30346	Atlanta	Southern	2	United Package	114	87999.645000	9.5526
31406	Savannah	Southern	2	United Package	114	87999.645000	9.5526

This query is used to extract the data required to understand Impact of different territories and shipping providers.

- **TerritoryID:** A unique identifier for each sales territory. This helps in distinguishing between different geographic sales areas managed by the company.
- **TerritoryDescription:** A description of the territory, providing context and detail about the geographic area associated with the TerritoryID.
- **RegionDescription:** A description of the broader region to which the territory belongs.
- **ShipperID:** A unique identifier for each shipper, indicating the shipping company responsible for delivering orders.
- **ShipperName:** The name of the shipping company associated with the respective ShipperID.
- **TotalOrders:** The count of total orders shipped to customers within the territory, grouped by the shipper.
- **TotalSales:** The total sales amount calculated from all orders, accounting for quantity sold, unit price, and any discounts applied.



- **AvgShippingTime:** The average number of days it takes for shipments to arrive from the order date to the shipping date.

## **7 Chapter 4. Data Understanding**

### **7.1 Thorough exploration and analysis of the collected data**

To establish the trends and correlations related to the research question, four queries executed. The explanation of these queries is as below:

#### **7.1.1 SQL Query-1 Customer purchasing behaviors**

With the help of this data the purchasing patterns of customers can be analyzed, particularly those who frequently buy from specific product categories, thereby allowing for insights into customer segmentation (like engaged customers or casual customers) and preferences based on their purchasing behavior.

#### **7.1.2 SQL Query-2 Impact of Seasons**

With the help of this data seasonal sales trends i.e. the impact of change of seasons on product categories and total sales can be analyzed, based on this input inventories can be planned according to different seasons.

#### **7.1.3 SQL Query-3 Factors affecting the sales performance**

To analyze the different factors contributing to the sales, such as understanding of top selling products in different regions. Based on this company can make the necessary changes in type of products in these regions.

#### **7.1.4 SQL Query-4 Impact of different territories and shipping providers**

To assess the impact of territory and shipping providers on total sales. This helps in understanding if the products are arriving on time to different locations/territories by using the

appropriate shippers. This also helps in understanding of delivery rates of shippers and if any consistent delays are happening by a specific shipper, company can replace this shipper.

## **7.2 Identification of Key Metrics, Trends, and Patterns**

### **7.2.1 SQL Query-1 Customer purchasing behaviors Key-Insights**

The customer ID SAVEA has maximum total order with Average quantity per order as 42.74 and total repetitive orders of 13456.

As the total orders count decreases, the Average quantity per order and total repetitive orders also decreases.

### **7.2.2 SQL Query-2 Impact of Seasons Key-Insights**

#### **Overall Seasonal Performance**

- Spring recorded the highest sales 31% of total sales.
- Winter followed with 32%, showing strong consumer demand.
- Fall accounted for 22%.
- Summer had the lowest sales at 15%.

#### **Category-Wise Trends**

- Spring & Winter: Showed the highest total sales, indicating peak shopping periods.
- Dairy Products & Confections consistently ranked among the top categories across all seasons.
- Seafood & Produce performed better in Winter & Fall compared to Summer.
- Beverages had moderate sales, peaking in Spring.

#### **Recommendations**

- The inventory can be optimized by stocking high-demand products according to seasonal needs.
- Depending upon the season and sales trend, Seasonal promotions can be offered to boost the overall sales.

- Product bundling like grouping medium performing products along with high-performing categories in top seasons can help in additional revenue.

### **7.2.3 SQL Query-3 Factors affecting the sales performance Key-Insights**

#### **Overall Regional Performance**

- Eastern region had the highest total sales (~35%) followed with Western region (~27%).
- Northern region and Southern regions had less sales accounting (~17%) & (~13%) respectively.

#### **Total Orders Distribution**

- Eastern region had the most orders (51% of total orders), followed with Western region which had 27% of total orders.

#### **Business Recommendations**

- As the Southern & Northern regions sales are less, company can focus high on promotions in these areas.
- Concentrate on increasing the sales of high-demand product categories in the Western region, particularly Dairy and Meat.
- Leverage marketing campaigns in the Eastern region to maintain strong sales momentum.

### **7.2.4 SQL Query-4 Impact of different territories and shipping providers Key-Insights**

#### **Total Orders (Territory Wise & Shipper Wise):**

The highest number of total orders comes from specific territories (like Atlanta, Cary, Greensboro), Similarly, for shippers, a few key players handle the majority of orders, with others trailing significantly.

A noticeable pattern shows certain territories and shippers dominating order volumes.

#### **Total Sales (Territory Wise & Shipper Wise):**

A few territories and shippers generate most of the revenue.

The leading territories are significantly ahead, suggesting high-value markets.

### Average Shipping Time:

There is a fluctuating trend in shipping times across different shippers.

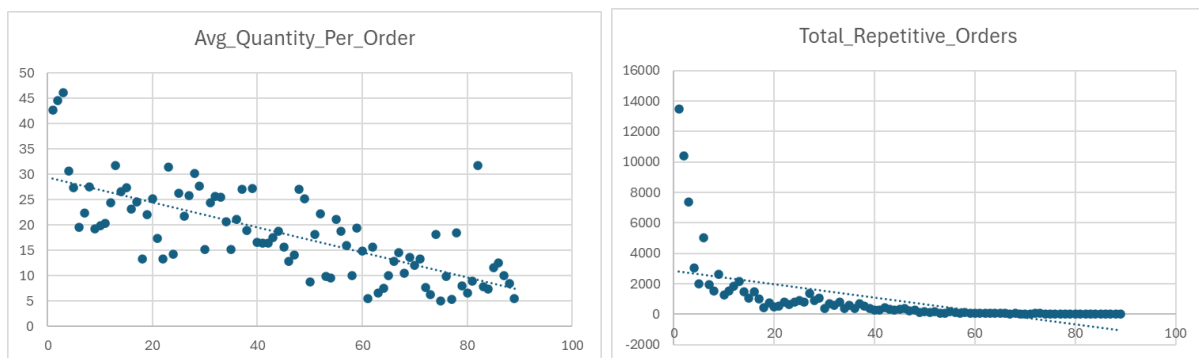
### Key Takeaways:

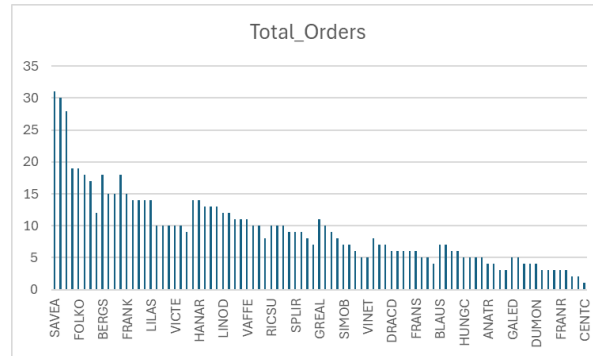
- A few territories and shippers contribute most to orders and sales.
- Some shippers maintain stable shipping times, but others show inconsistencies, which could impact customer satisfaction.
- Identifying high-value territories and optimizing logistics for slow shippers could improve business efficiency.

## 8 Chapter 5. Data Visualization

This chapter provides data visualizations and information related to charts and diagrams created from the data extracted.

### 8.1 Customer purchasing behaviors





### **Avg\_Quantity\_Per\_Order**

There is a clear negative correlation between order frequency and the average quantity per order.

Higher frequency orders tend to have lower quantities per order, while less frequent orders have higher quantities per order.

### **Total\_Repetitive\_Orders**

The trend shows that a few customers or products have extremely high repeat orders, while the majority have significantly fewer.

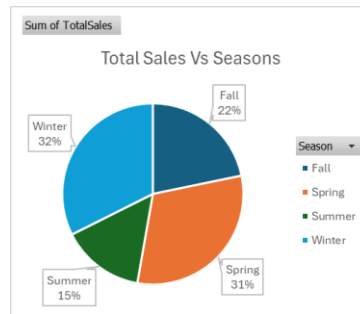
This follows a long-tail distribution, where a small portion of items/customers contribute to most of the repeat orders.

### **Total\_Orders**

Some categories (e.g., SAVA, FOLO, BERS) have significantly higher total orders compared to others. There is a sharp decline in total orders as we move toward the right side of the chart.

## 8.2 Impact of Seasons

Row Labels	Sum of TotalSales
Fall	275409.4475
Spring	392347.9305
Summer	188018.5
Winter	410017.1615
<b>Grand Total</b>	<b>1265793.04</b>



Season	Fall		Season	Spring	
Row Labels	Sum of TotalSales		Row Labels	Sum of TotalSales	
Beverages	48390.42		Beverages	87313.805	
Condiments	26153.91		Condiments	34388.8	
Confections	32622.1115		Confections	56543.6705	
Dairy Products	55709.525		Dairy Products	74933.42	
Grains/Cereals	19976.285		Grains/Cereals	25261.255	
Meat/Poultry	39878.086		Meat/Poultry	39395.5	
Produce	18992.2275		Produce	41129.4925	
Seafood	33686.8825		Seafood	33381.9875	
<b>Grand Total</b>	<b>275409.4475</b>		<b>Grand Total</b>	<b>392347.9305</b>	
Season	Summer		Season	Winter	
Row Labels	Sum of TotalSales		Row Labels	Sum of TotalSales	
Beverages	25260.955		Beverages	106903	
Condiments	15769.935		Condiments	29734.44	
Confections	26525.0375		Confections	51666.4055	
Dairy Products	39269.74		Dairy Products	64594.6	
Grains/Cereals	17746.36		Grains/Cereals	32760.6875	
Meat/Poultry	20775.42		Meat/Poultry	62973.3535	
Produce	16430.82		Produce	23432.04	
Seafood	26240.2325		Seafood	37952.635	
<b>Grand Total</b>	<b>188018.5</b>		<b>Grand Total</b>	<b>410017.1615</b>	

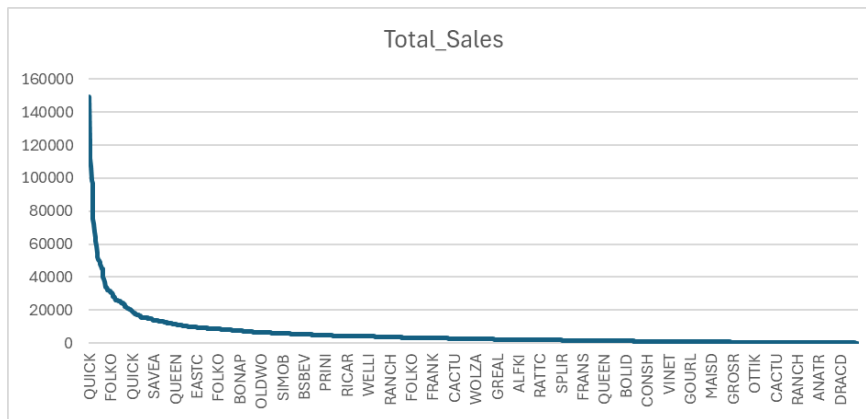
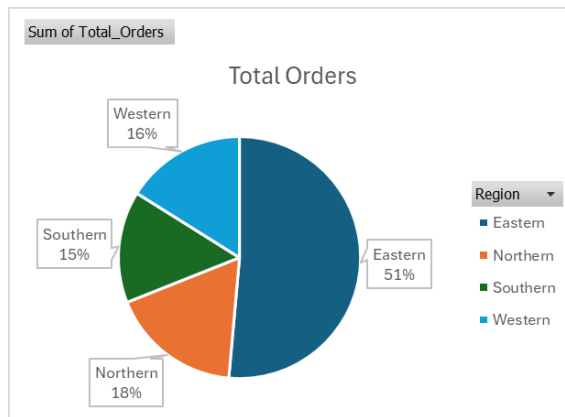
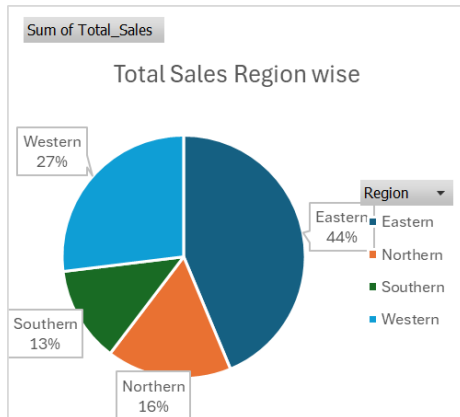


The following points can be inferred from plots:

- Spring recorded the highest sales of 392,347.91 and Winter followed with 410,017.16
- Dairy Products & Confections had the highest total sales across all seasons.
- Spring & Winter had the highest sales in most categories.
- Summer showed the lowest sales, especially in categories like Grains/Cereal and Condiments.
- Meat/Poultry saw peak sales in Winter, indicating demand for comfort and festive foods.

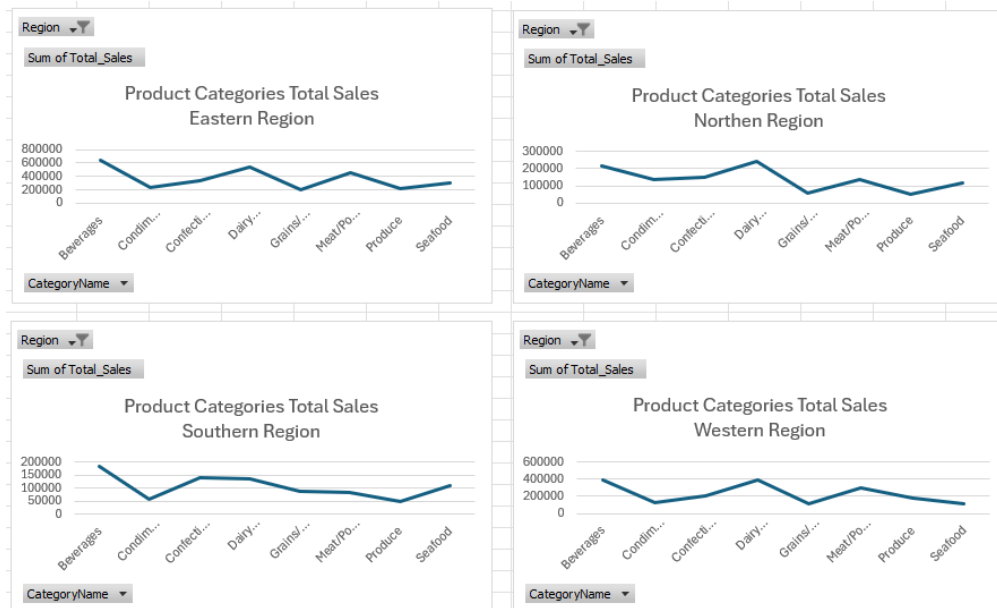
### 8.3 Factors affecting the sales performance

Regions	Sum of Total_Sales	Regions	Sum of Total_Orders
Eastern	2928068.84	Eastern	981
Northern	1113952.12	Northern	335
Southern	852205.2	Southern	285
Western	1803950.4	Western	307
<b>Grand Total</b>	<b>6698176.56</b>	<b>Grand Total</b>	<b>1908</b>



Region	Eastern		Region	Northern	
Row Labels	Sum of Total_Sales		Row Labels	Sum of Total_Sales	
Beverages	645019.35		Beverages	218158.4	
Condiments	237883.9		Condiments	139540.1	
Confections	339618.26		Confections	152038.58	
Dairy Products	533284.1		Dairy Products	240510.1	
Grains/Cereals	203535.4		Grains/Cereals	56034.1	
Meat/Poultry	449130.17		Meat/Poultry	140072.52	
Produce	216242.45		Produce	51416.95	
Seafood	303355.21		Seafood	116181.37	
Grand Total	2928068.84		Grand Total	1113952.12	
Region	Southern		Region	Western	
Row Labels	Sum of Total_Sales		Row Labels	Sum of Total_Sales	
Beverages	186026.2		Beverages	385808.75	
Condiments	56436.6		Condiments	121338	
Confections	139437.24		Confections	202496	
Dairy Products	138185.2		Dairy Products	396230.5	
Grains/Cereals	87592.6		Grains/Cereals	116490.5	
Meat/Poultry	83686.8		Meat/Poultry	296301.4	
Produce	50872.6		Produce	177143.5	
Seafood	109967.96		Seafood	108141.75	
Grand Total	852205.2		Grand Total	1803950.4	

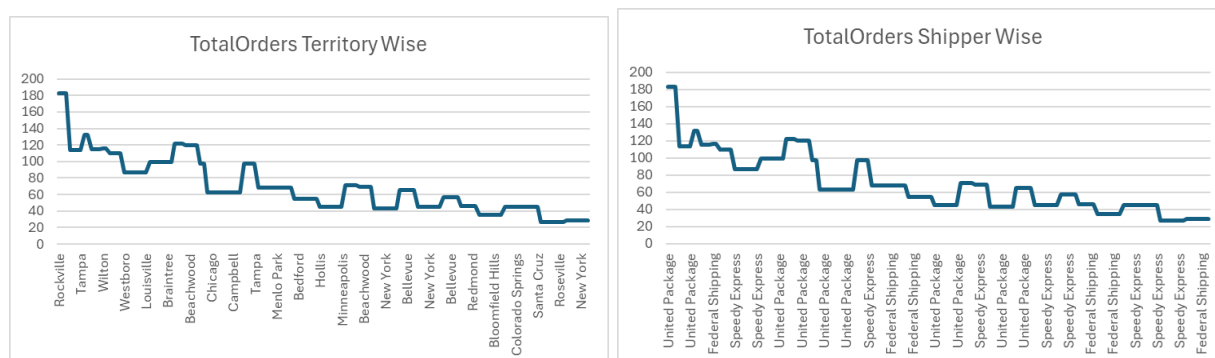


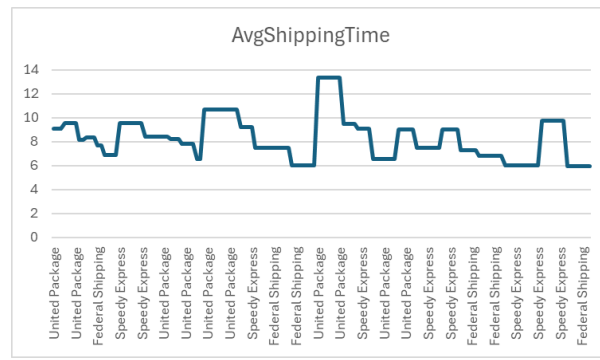
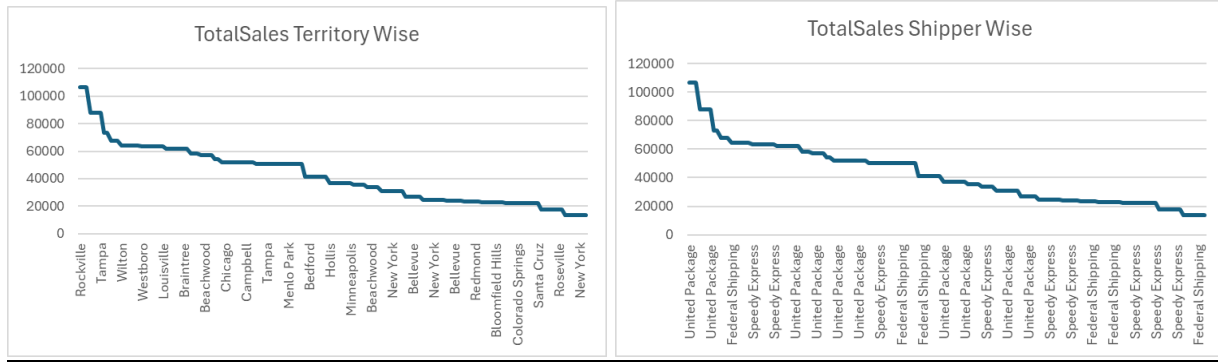


The following points can be inferred from plots:

- Eastern Region Leads in Sales & Orders, with Beverages 645,913.55 was the top-selling category.
- Western Region performed well in Dairy Products & Meat/Poultry.
- Northern region had the lowest sales in Beverages & Condiments.
- Southern region had lower sales across most categories.

## 8.4 Impact of different territories and shipping providers





### Total Orders (Territory Wise & Shipper Wise):

The highest total orders are from specific territories, with a sharp decline as you move down the list. The same with shippers; there are a couple of big players that handle the majority of orders, with the rest following far behind.

### Total Sales (Territory Wise & Shipper Wise):

Sales distribution is a steep curve, meaning that there are limited shippers and territories that contribute to most of the revenue. There is a long tail, meaning many regions have less sales contribution.

### Average Shipping Time:

There is a random trend in shipping time across different shippers. Some shippers have relatively consistent delivery times, while others have variance.

## **9 Chapter 6. Model Building**

### **9.1 Data Modeling**

- The data fetched from the customer\_purchase\_Train, Season\_Sales\_Train, Sales\_perform\_Train, and Territory\_Shipping\_Train sheets, containing key features such as:
  - Total Orders
  - Distinct Categories
  - Avg Quantity per Order
  - Frequent Categories
  - Total Repetitive Orders

### **9.2 Modeling Techniques Used:**

- Linear Regression to predict customer purchasing behavior based on total orders and distinct categories.
- ANOVA (Single Factor Analysis) for analyzing the impact of different seasons on sales performance.
- ANOVA (Single Factor Analysis) for determining the impact of customer segmentation on total sales revenue.
- Multiple Linear Regression for quantifying the impact of shipping vendors and regions on overall sales performance.

## 1) What are the customer purchasing behaviors in Northwind Trading Company?

### Test Performed: Linear Regression

**Rationale:** Linear regression helps to quantify the relationship between different predictors (e.g., Total Orders and Unique Categories) and the response variable for prediction purposes.

**Limitations:** Linear regression relies on the linearity assumption, is sensitive to multicollinearity and outliers, requires homoscedasticity, can be sensitive to overfitting, and conveys correlation not causation.

SUMMARY OUTPUT TRAIN					
Regression Statistics					
Multiple R	0.883583252				
R Square	0.780719363				
Adjusted R Square	0.773286121				
Standard Error	957.7378364				
Observations	62				
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	192681477.6	96340739	105.0308	3.63E-20
Residual	59	94118444.03	917261.8		
Total	61	246799921.7			
	Coefficients	Standard Error	t Stat	P-value	Lower 95%Upper 95%Lower 95.0%Upper 95.0%
Intercept	1267.108471	823.5849072	1.562612	0.123445	-360.881 2935.098 -360.881 2935.098
Total_Orders	393.6863063	29.48835996	13.35057	1.76E-19	334.6802 452.6924 334.6802 452.6924
Distinct_Categories	-603.6167874	138.8119769	-4.34845	5.5E-05	-881.379 -325.855 -881.379 -325.855

SUMMARY OUTPUT TEST									
Regression Statistics									
Multiple R	0.900306005								
R Square	0.810550903								
Adjusted R	0.794763478								
Standard E	927.6544982								
Observations	27								
ANOVA									
	df	SS	MS	F	Significance F				
Regressor	2	88363214	44181607	51.3415529	2.14E-09				
Residual	24	20653029	860542.9						
Total	26	1.09E+08							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	213.026278	646.4836	0.329515	0.744624788	-1121.250207	1547.303	-1121.25	1547.303	
Total_Orders	338.7528091	36.69053	9.232703	2.28249E-09	263.027268	414.4784	263.0273	414.4784	
Distinct_Categories	-357.8358555	127.2065	-2.81303	0.009632907	-620.377096	-95.2946	-620.377	-95.2946	

## 2) What are the impacts of different seasons on sales performance?

### Test Performed: Anova Single Factor

**Rationale:** Used ANOVA to compare the sales performance across different seasons, helping identify whether seasonal changes significantly affect sales

**Limitations:** ANOVA assumes that groups are independent, and may not account for other confounding factors such as promotions or market trends that could influence seasonal sales.

TRAIN						
Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Season_Num	129	333	2.581395349	1.2765262		
TotalSales	129	850622.777	6593.975016	27036942		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2802297306	1	2802297306	207.29394	7.8532E-35	3.8780396
Within Groups	3460728741	256	13518471.64			
Total	6263026047	257				

Anova: Single Factor						
<div></div>						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Season_Num	55	131	2.381818182	1.351515152		
TotalSales	55	415170.2625	7548.550227	45080685.53		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1565978086	1	1565978086	69.47445542	2.71323E-13	3.929011718
Within Groups	2434357092	108	22540343.44			
Total	4000335177	109				

### **3) What factors affect the sales performance among products and regions?**

Test Performed: Anova Single Factor

**Rationale:** ANOVA helps to compare the sales performance across different customer segments, assessing whether variations in customer segments significantly impact total sales.

**Limitations:** ANOVA assumes that the groups are independent and may not account for other confounding factors like product pricing or external economic factors, which could influence sales performance in different regions.

Anova: Single Factor					
SUMMARY					
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>	
Customer_Segment	803	908	1.13076	0.113803	
Total_Sales	803	4811027	5991.316	1.47E+08	
ANOVA					
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>
Between Groups	1.44E+10	1	1.44E+10	196.2216	3.81E-42
Within Groups	1.18E+11	1604	73420831		
Total	1.32E+11	1605			

Anova: Single Factor						
SUMMARY						
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
Customer_Segment_N	345	388	1.1246377	0.1094203		
Total_Sales	345	1887149.6	5469.9989	101375226		
ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	5159231030	1	5.159E+09	101.78485	2.0471E-22	3.85501034
Within Groups	34873077900	688	50687613			
Total	40032308930	689				

#### **4) How do different territories and shipping providers impact on total sales performance?**

Test Performed: Multiple Regression

##### **Rationale:**

Multiple regression was used to evaluate the impact of territories and shipping providers on sales performance, allowing for the analysis of their combined effects on the dependent variable (total sales).

##### **Limitations:**

Multiple regression assumes linear relationships between predictors and the dependent variable and may not account for non-linear effects. Additionally, multicollinearity could impact the interpretation of results if territories or region and shipping providers are highly correlated.

SUMMARY OUTPUT TRAIN								
Regression Statistics								
Multiple R	0.365078484							
R Square	0.1332823							
Adjusted R Square	0.106750125							
Standard Error	16828.51445							
Observations	102							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	4267882431	1422627477	5.023421645	0.002787159			
Residual	98	27753492059	283198898.6					
Total	101	32021374490						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	33526.98668	11668.16902	2.873371702	0.004979718	10371.885	56682.08837	10371.885	56682.08837
AvgShippingTime	2529.661461	1126.164149	2.246263534	0.026931976	294.8253589	4764.497564	294.8253589	4764.497564
RegionDescription_Num	-6191.405253	1601.154849	-3.866837274	0.000198424	-9368.844941	-3013.965565	-9368.844941	-3013.965565
ShipperID	4813.69011	2653.160481	1.81432301	0.07268765	-451.4205168	10078.80074	-451.4205168	10078.80074
SUMMARY OUTPUT Test								
Regression Statistics								
Multiple R	0.576805676							
R Square	0.332704788							
Adjusted R Square	0.283878309							
Standard Error	3184.230384							
Observations	45							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	3	207268766.5	69089588.84	6.814023767	0.000784741			
Residual	41	415712248.6	10139323.14					
Total	44	622981015.1						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	17862.21152	3592.19618	4.972504458	1.2253E-05	10607.62416	25116.79888	10607.62416	25116.79888
AvgShippingTime	-150.6064396	377.9556084	-0.398476531	0.692346601	-913.9032757	612.6903966	-913.9032757	612.6903966
RegionDescription_NUM	1662.551724	419.3073079	3.964995821	0.000267411	815.7434361	2509.360011	815.7434361	2509.360011
ShipperID	-513.0275316	549.6949057	-0.933295045	0.356131682	-1623.158915	597.1038518	-1623.158915	597.1038518

## 10 Chapter 7. Model Evaluation

### 10.1 Hypothesis 1:

**Null Hypothesis (H<sub>0</sub>):** Customers who frequently purchase from different product categories do not show consistent ordering patterns.

**Alternate Hypothesis (H<sub>1</sub>):** Customers who frequently purchase from specific product categories tend to have consistent and repetitive ordering behaviors.

Key Findings: **Customer\_purchase\_Keymetrics**

Metric	Train Dataset	Test Dataset
R-Square	0.7807	0.8106
P value (Model)	3.63E-20	2.1375E-09
Total_Orders	393.686	388.752
Pvalue	1.76E-19	2.28E-09
Distinct_Categories	-603.617	-357.836
Pvalue	5.50E-05	0.0096

## Interpretation

**R-square:** 78-81% of the variability in Total\_Repetitive\_Orders can be explained by Total\_Orders and Distinct\_Categories.

**Coefficient:** For each additional order, Total\_Repetitive\_Orders seems to increase by approximately 393 units

For each additional distinct category, Total\_Repetitive\_Orders seems to decrease by approximately 604 units

**P-Value(Predictors):** Is less than 0.05; hence, it is statistically significant. We reject the null hypothesis and confirm that Total\_Orders and Distinct\_Categories have a strong impact on Total\_Repetitive\_Orders.

**P value (Model):** Is less than 0.05, hence the overall model is statistically significant, confirming the model effectively explains the relationship with Total\_Repetitive\_Orders.

## 10.2 Hypothesis 2:

**Null Hypothesis ( $H_0$ ):** Sales performance remains constant throughout the year, with no significant seasonal impact.

**Alternate Hypothesis ( $H_1$ ):** Sales performance fluctuates significantly across different seasons, with certain periods showing a increase in sales.

## Key Findings: Season\_Sales\_keymetrics

	Train Dataset	Test Dataset
Pvalue	7.85315E-35	2.71323E-13



## Interpretation

**P-value:** The p-values in both the Train and test are less than 0.05. Since the p-value is less than 0.05, we reject the null hypothesis and confirm that seasonality has a statistically significant effect on Total Sales.

### 10.3 Hypothesis 3:

**Null Hypothesis (H<sub>0</sub>):** Some customer segmentation does not significantly impact total sales revenue than others based on their purchasing behavior.

**Alternate Hypothesis (H<sub>1</sub>):** Some customer segments contribute significantly towards total sales revenue than others based on their purchasing behavior.

#### Key Findings: Sales\_perform\_keymetrics

	Train Data	Test Data:
Pvalue	3.81E-42	2.05E-22

## Interpretation

**P-value:** The p-values in both Train and test are less than 0.05, Hence we reject the null and it is statistically significant confirming that Customer Segment has an effect on Total Sales Revenue.

### 10.4 Hypothesis 4:

**Null Hypothesis (H<sub>0</sub>):** The choice of shipper does not influence total sales, and order volumes remain uniform across regions.

**Alternate Hypothesis (H<sub>1</sub>):** Shipper choice has a significant effect on total sales, with different territories/regions exhibiting varied order volumes based on shipping preferences.

## Key Findings: Territory\_Shipping\_keymetrics

Metric	Train Data	Test Data
R-Square	0.133282	0.332705
P value (Model)	0.002787	0.000785
AvgShippingTime	2529.661	0.692347
Pvalue	0.026932	2.28E-09
RegionDescription_Num	-6191.41	1662.552
Pvalue	0.000198	0.072688
ShipperID	4813.69	-513.028
Pvalue	0.072688	0.356132

## Interpretation

**Rsquare:** The model explains 13.3% of the variability in TotalSales for the training set and 33.3% for the test set, showing modest predictive power.

**Coefficient:** RegionDescription\_NUM shows a negative impact in the training set and a positive impact in the test set.

AvgShippingTime and ShipperID are inconsistent and non-significant.

### P-Value(Predictors):

RegionDescription\_NUM is statistically significant in both datasets.

AvgShippingTime and ShipperID are not significant in the test set.

Since the p-value for Region Description is less than 0.05 in the training set, we reject the null hypothesis and confirm that Region Description has a statistically significant effect on Total Sales in the training data. However, in the test data, this relationship is weaker. The Shipping ID p-values indicate that its impact is inconsistent and non-significant.

**P value (Model):** P value < 0.05, confirming the overall model is statistically significant.

## **11 Conclusion:**

The analysis of Northwind Trading Company's sales and customer behavior, supported by both test results, key metrics, and visual insights, has revealed clear trends that inform strategic improvements. The linear regression with an R-square value of 78-81%, indicating that the model effectively explains the variability in repetitive purchases based on total orders and distinct categories results, confirmed that repetitive purchases are closely linked to total order volume, which aligns with visual insights showing a negative correlation between order frequency and average quantity per order. Customers with frequent orders tended to purchase in smaller quantities per transaction. Visualizations also illustrated that a few particular customers such as SAVEA created an outsized share of repeat orders, further validating the significance of focused customer retention.

ANOVA test confirmed seasonal variations in sales, and visualizations showed that Spring and Winter together contributed over 60% of total sales, and Dairy Products and Confections were steady across all seasons. Visual patterns showed that Seafood and Produce experienced a notable increase in Winter and Fall, and Summer sales were the least, particularly for product categories such as Grains/Cereal and Condiments. These visual observations also underscore the necessity of aligning promotions and inventory with seasonal demand.

Customer segmentation report supported that high-value customers and frequent buyers have a significant influence on sales. Visualizations reinforced the same by anticipating the Eastern region as the key driver with 35% of overall sales and 51% of overall orders. In contrast, the Southern and Northern regions had much lower volumes of sales, supporting the need for targeted campaigns in these regions. Visual data on shipping performance revealed that strategic locations like Atlanta and Cary contributed heavily towards total orders, whereas certain

shipping providers exhibited random delivery times, which may have an impact on customer satisfaction.

## **12 Recommendations:**

Targeted marketing efforts in the South and North can induce engagement, while stocking up on high-demand products like Dairy and Confections during Winter and Spring can guarantee top-of-the-line sales.

Packing slower movers with best sellers can drive Summer sales. Starting loyalty programs and targeted promotions can retain quality customers. Concentrating on good shippers and penalizing delay with poor performing providers will improve logistics and customer satisfaction.

These tactics, informed by test data and key metrics, can improve sales performance and fuel long-term growth.

## **13 References**

- Marketing Analytics, A Comprehensive Guide, v1.0 Christina J. Inge,  
[https://scholar.flatworldknowledge.com/books/33141/marketing\\_analytics\\_1-33143-20211021-123616-215205/preview](https://scholar.flatworldknowledge.com/books/33141/marketing_analytics_1-33143-20211021-123616-215205/preview)
- Adobe Experience Cloud Team, 01-14-2025;  
<https://business.adobe.com/blog/basics/learn-about-four-ps-of-marketing>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage Learning.
- Business research methods (2019) (5th ed.), Bryman, A., & Bell, E., Oxford University Press.

- Research methods for business students (2019). (8th ed.), Saunders, M., Lewis, P., & Thornhill, A, Pearson Education.
- The 4Ps of Marketing: Product, Price, Place, and Promotion | Analytics Steps. 2021.  
Pathak R.

## **14 Appendix**

The below data sets and analysis & modeling test results in form of excel files appended to this report are:

- Case Study1\_Course\_Project\_Full\_Customer\_Behavior\_Dataset
- Case Study1\_Course\_Project\_Full\_Impact of Seasons\_Dataset
- Case Study1\_Course\_Project\_Full\_Factors affecting Sales\_Dataset
- Case Study1\_Course\_Project\_Full\_Territories & Shippers affecting Sales\_Dataset
- Case Study1\_Course\_Project\_Full\_Analysis\_Modelling