

Northwind Data Analysis Project

Comprehensive Business Insights

PROJECT BY:

ANFAL MAHMOUD

ESRAA EBRAHIM

MOHAMED ALQMASH

1. Introduction to the Project:

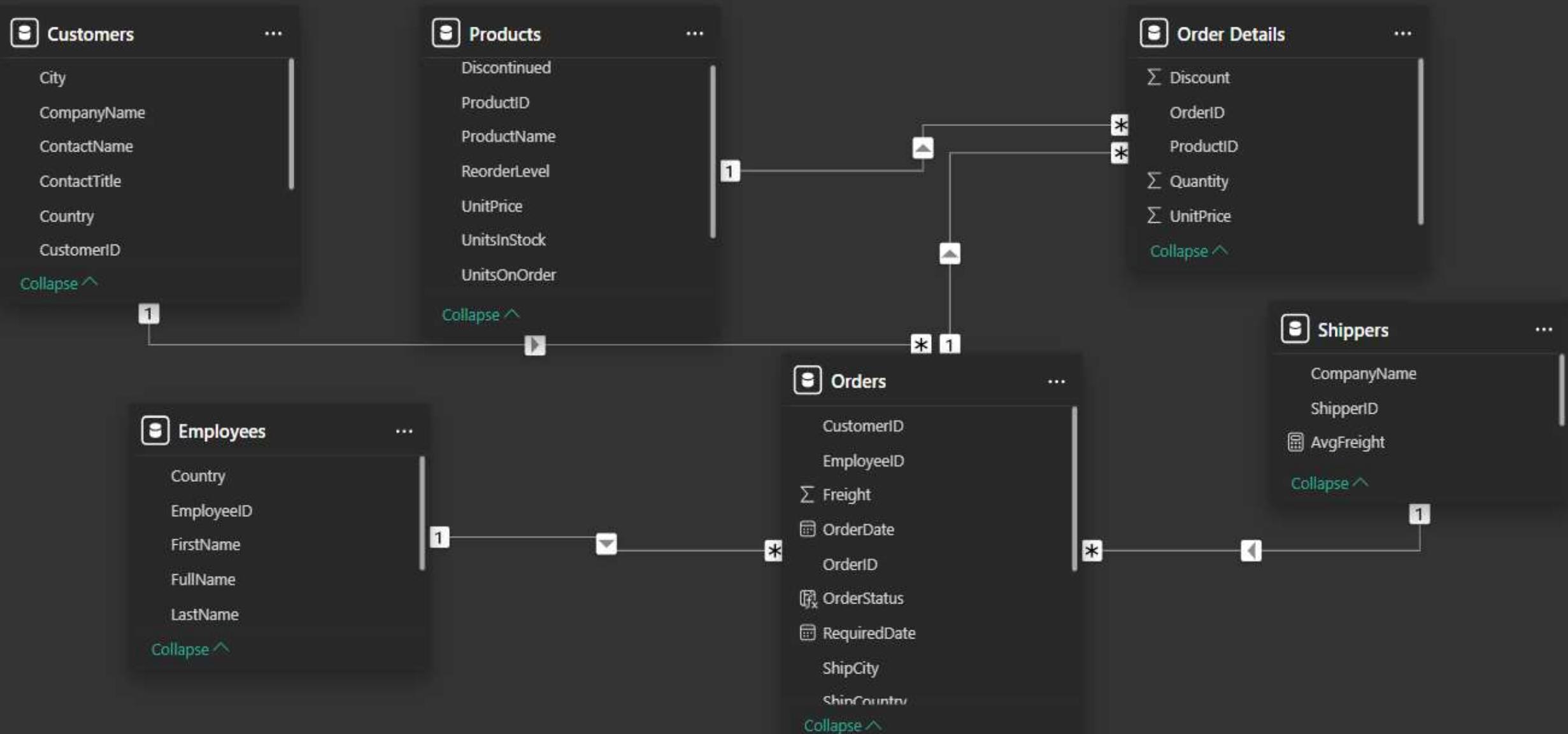
- **Objective:** The purpose of this analysis is to extract valuable business insights from the Northwind dataset, a classic example of sales, product, customer, and employee data. Our goal is to analyze sales performance, customer trends, shipping efficiency, and employee productivity to provide actionable insights for decision-making.
- Throughout this presentation, I will showcase how we used SQL for querying the data, Power Query for Pre-processing Data, and Power BI for visualizing the results.

2. Data Overview Slide:

- Let's start by looking at the data we have. The Northwind dataset consists of several tables representing different aspects of the business:
 - **Orders:** This contains detailed information about the sales transactions.
 - **Customers:** This table lists all the customers who have made purchases.
 - **Products:** Includes information about the products that were sold.
 - **Employees:** Data about the employees who handled the orders.
 - **Shippers:** Information about the companies responsible for shipping the orders.

- **Schema and Relationships:**
 - The dataset is well-structured with clear relationships between tables. For example:
 - Each order is linked to a customer and an employee.
 - Products are linked to orders and categorized by product type.
 - Shippers handle the delivery of each order.
 - These relationships help us extract insights from multiple angles.

- Model View:



3. Key Analyses and Findings Slides:

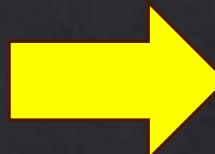


3.1. Product Revenue Analysis:

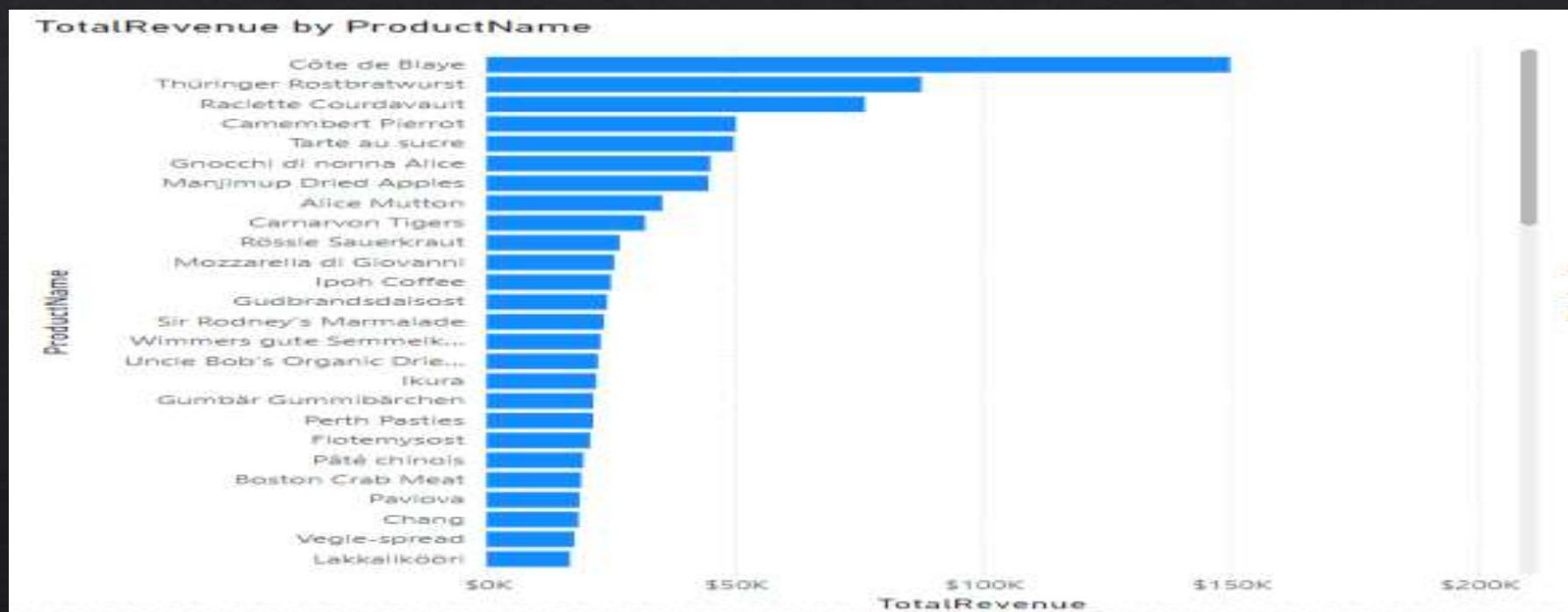
- ❖ This query calculates the **total revenue** generated by each product by multiplying the **quantity sold** by the **unit price**. It ranks products based on their total revenue. This analysis provides key insights such as:
 1. **Top Revenue-Generating Products:** Identifying which products contribute the most to the company's revenue helps prioritize marketing and sales efforts.
 2. **Sales Trends:** Understanding which products drive the most revenue over time can guide product development and promotions.
 3. **Inventory Optimization:** By focusing on high-revenue products, the business can manage stock levels more effectively, ensuring availability of the best-selling items.

•Here's a Queries and chart showing Product Revenue Analysis:

```
SELECT P.ProductName,  
       SUM(OD.Quantity * OD.UnitPrice) AS TotalRevenue  
  FROM dbo.[Order Details] AS OD  
 JOIN dbo.Products AS P ON OD.ProductID = P.ProductID  
 GROUP BY P.ProductName  
 ORDER BY TotalRevenue DESC;
```



	ProductName	TotalRevenue
1	Côte de Blaye	149984.20
2	Thüringer Rostbratwurst	87736.40
3	Raclette Courdavault	76296.00
4	Camembert Pierrot	50286.00
5	Tarte au sucre	49827.90
6	Gnocchi di nonna Alice	45121.20
7	Manjimup Dried Apples	44742.60
8	Alice Mutton	35482.20



3.2. Customer Spending Analysis:

- ❖ This query calculates the **total amount spent** by each customer by summing the **quantity** and **unit price** for all their orders. The customers are ranked in descending order based on their total spend. This analysis provides the following key insights:
1. **Identify High-Value Customers:** Recognizing which customers contribute the most to sales allows the business to focus on maintaining and nurturing relationships with them.
 2. **Targeted Marketing:** The company can design personalized marketing campaigns and loyalty programs for top spenders to improve retention.
 3. **Revenue Growth:** By understanding which customers generate the most revenue, the company can allocate resources effectively to maximize future sales.

- Here's a Queries and chart showing Customer Spending Analysis:

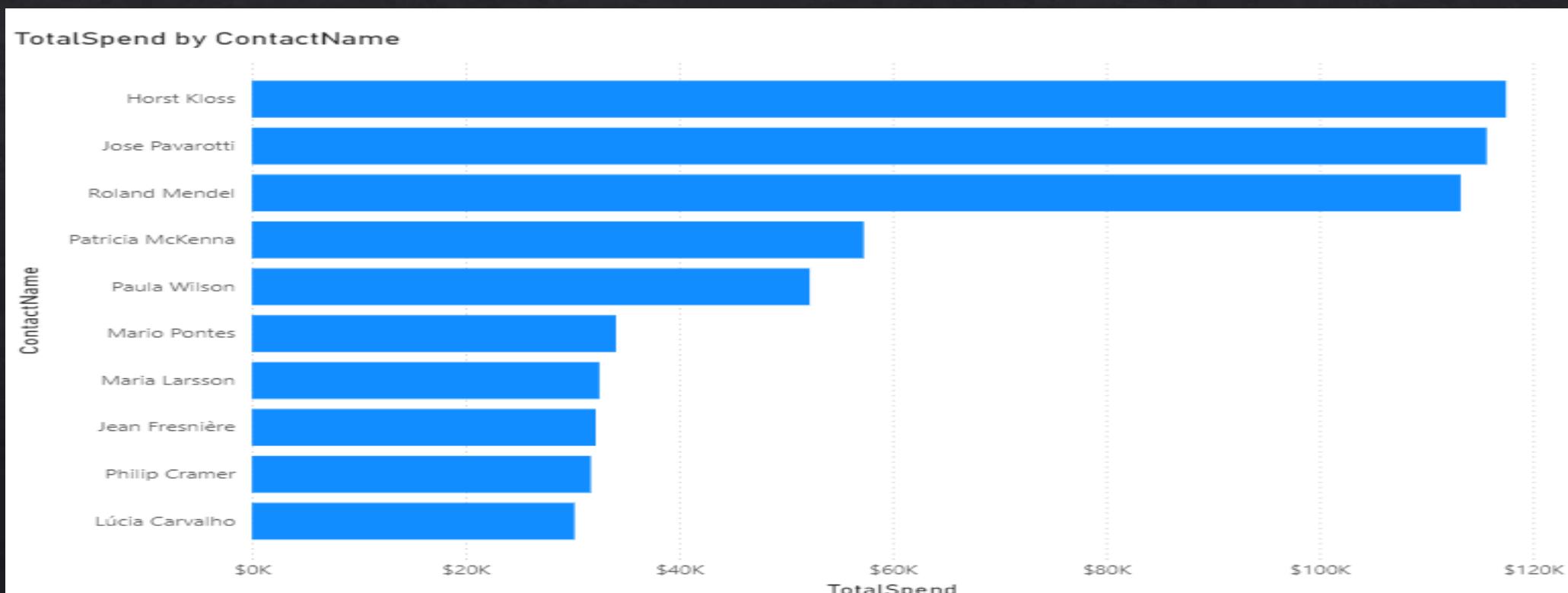
```

SELECT C.CustomerID, C.CompanyName,
       SUM(OD.Quantity * OD.UnitPrice) AS TotalSpend
  FROM dbo.Orders AS O
 JOIN dbo.[Order Details] AS OD ON O.OrderID = OD.OrderID
 JOIN dbo.Customers AS C ON O.CustomerID = C.CustomerID
 GROUP BY C.CustomerID, C.CompanyName
 ORDER BY TotalSpend DESC;
    
```



Results Messages

	CustomerID	CompanyName	TotalSpend
1	QUICK	QUICK-Stop	117483.39
2	SAVEA	Save-a-lot Markets	115673.39
3	ERNSH	Ernst Handel	113236.68
4	HUNGO	Hungry Owl All-Night Grocers	57317.39
5	RATTC	Rattlesnake Canyon Grocery	52245.90
6	HANAR	Hanari Carnes	34101.15
7	FOLKO	Folk och fä HB	32555.55
8	MEREP	Mère Paillardé	32203.90



3.2.1. Top Products Purchased by Each Customer:

- ❖ In this query, we identified the **most purchased product** by each customer, calculating both the **total amount spent** and the **total quantity** of that product. This helps in understanding customer preferences and spending patterns, allowing us to:
 1. **Enhance Product Recommendations:** Knowing which product a customer buys the most can help personalize future offers.
 2. **Optimize Inventory:** Focus on stocking high-demand products for specific customers to meet their needs better.
 3. **Target High-Value Customers:** Customers spending the most on particular products can be targeted for loyalty programs or special discounts.

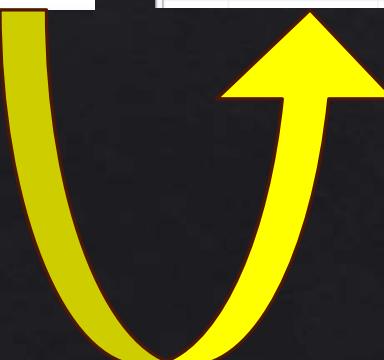
```

WITH CustomerProductSales AS (
    SELECT C.CustomerID, C.CompanyName, P.ProductName,
           SUM(OD.Quantity * OD.UnitPrice) AS TotalSpend,
           SUM(OD.Quantity) AS TotalQuantity,
           ROW_NUMBER() OVER (PARTITION BY C.CustomerID ORDER BY SUM(OD.Quantity) DESC) AS RowNum
    FROM dbo.Orders AS O
   JOIN dbo.[Order Details] AS OD ON O.OrderID = OD.OrderID
   JOIN dbo.Customers AS C ON O.CustomerID = C.CustomerID
   JOIN dbo.Products AS P ON OD.ProductID = P.ProductID
  GROUP BY C.CustomerID, C.CompanyName, P.ProductName
)
SELECT CustomerID, CompanyName, ProductName, TotalSpend, TotalQuantity
FROM CustomerProductSales
WHERE RowNum = 1
ORDER BY TotalSpend DESC;

```

Results Messages

	CustomerID	CompanyName	ProductName	TotalSpend	TotalQuantity
1	HUNGO	Hungry Owl All-Night Grocers	Thüringer Rostbratwurst	18568.50	150
2	GREAL	Great Lakes Food Market	Côte de Blaye	11857.50	45
3	RATTC	Rattlesnake Canyon Grocery	Tarte au sucre	8180.00	185
4	QUICK	QUICK-Stop	Camembert Pierrot	7888.00	243
5	FOLIG	Folies gourmandes	Carnarvon Tigers	5000.00	80
6	BOTTM	Bottom-Dollar Markets	Tarte au sucre	4781.30	101
7	FOLKO	Folk och fä HB	Manjimup Dried Apples	4664.00	88
8	KOENE	Königlich Essen	Raclette Courdavault	4279.00	85



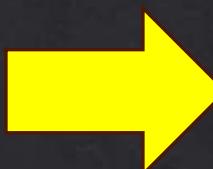
ContactName	TotalSpend	TotalQuantity
Alejandra Camino	\$1,467.29	91
Alexander Feuer	\$5,042.2	172
Ana Trujillo	\$1,402.95	63
Anabela Domingues	\$7,310.62	251
André Fonseca	\$8,702.23	315
Ann Devon	\$15,033.66	569
Annette Roulet	\$10,272.35	442
Antonio Moreno	\$7,515.35	359
Aria Cruz	\$4,438.9	357
Art Braunschweiger	\$12,489.7	327
Bernardo Batista	\$6,973.63	394
Carine Schmitt	\$3,172.16	69
Carlos González	\$17,825.06	836
Carlos Hernández	\$23,611.58	1096
Catherine Dewey	\$10,430.58	320
Christina Berglund	\$26,968.15	1001
Daniel Tonini	\$1,992.05	83
Dominique Perrier	\$2,423.35	48
Eduardo Saavedra	\$836.7	42
Elizabeth Brown	\$1,719.1	87
Elizabeth Lincoln	\$22,607.7	956
Felipe Izquierdo	\$17,889.55	970
Fran Wilson	\$4,258.6	134
Francisco Chang	\$100.8	11
Frédérique Citeaux	\$19,088	666
Georg Pips	\$26,259.95	624
Giovanni Rovelli	\$7,603.85	433
Guillermo Fernández	\$4,242.2	208
Hanna Moos	\$3,239.8	140
Hari Kumar	\$17,172.05	818
Helen Bennett	\$6,146.3	295
Total	\$1,354,458.59	51317

3.3. Employee Order Management Analysis:

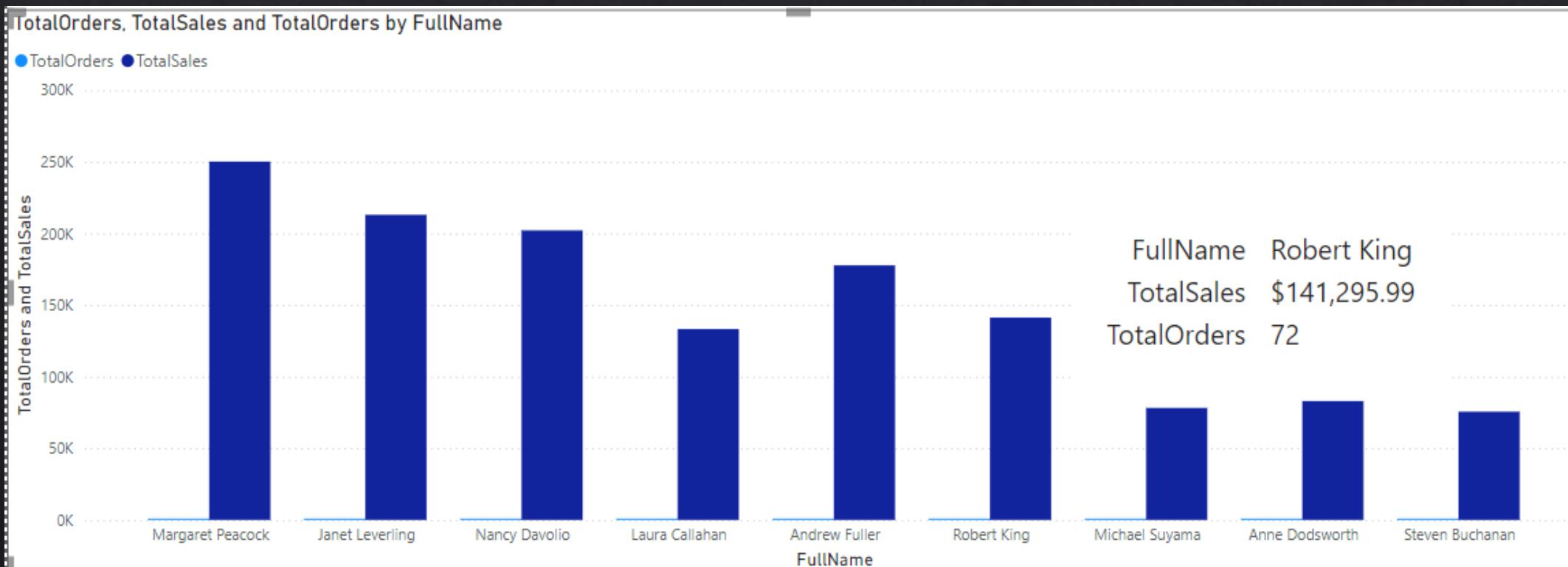
- ❖ This query counts the **total number of orders** managed by each employee, displaying the **employee's full name** and the number of orders they handled. The employees are ranked in descending order based on the total number of orders. This analysis provides:
 1. **Identify High-Performing Employees:** Recognizing which employees handle the most orders allows the company to assess workload and efficiency.
 2. **Resource Allocation:** Insights from this query help in distributing tasks more evenly among employees or identifying areas where additional support may be needed.
 3. **Employee Performance Review:** Understanding order management on an employee basis can be used for performance appraisals and to design incentives for top performers.

- Here's a Queries and chart ranking employees based on their sales performance:

```
SELECT E.EmployeeID, E.FirstName + ' ' + E.LastName AS [Full Name],
       COUNT(O.OrderID) AS TotalOrders
  FROM dbo.Orders AS O
 JOIN dbo.Employees AS E ON O.EmployeeID = E.EmployeeID
 GROUP BY E.EmployeeID, E.LastName, E.FirstName
 ORDER BY TotalOrders DESC;
```



	EmployeeID	Full Name	TotalOrders
1	4	Margaret Peacock	156
2	3	Janet Leverling	127
3	1	Nancy Davolio	123
4	8	Laura Callahan	104
5	2	Andrew Fuller	96
6	7	Robert King	72
7	6	Michael Suyama	67
8	9	Anne Dodsworth	43



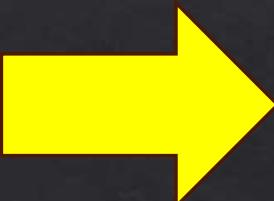
3.3.1.Top and Total sales value per employee :

- ❖ This query calculates the **total sales** made by each employee, showing the **Employee ID, Last Name, and First Name**, and sorting them by their total sales in descending order. This analysis provides several key insights:
 1. **Identify Top Performers:** Highlighting which employees are generating the highest sales helps in recognizing top-performing staff.
 2. **Targeted Training:** It allows for focused training programs to improve the performance of employees with lower sales figures.
 3. **Sales Incentives:** Insights from this query can be used to establish incentive plans for high-performing employees to boost overall company performance.

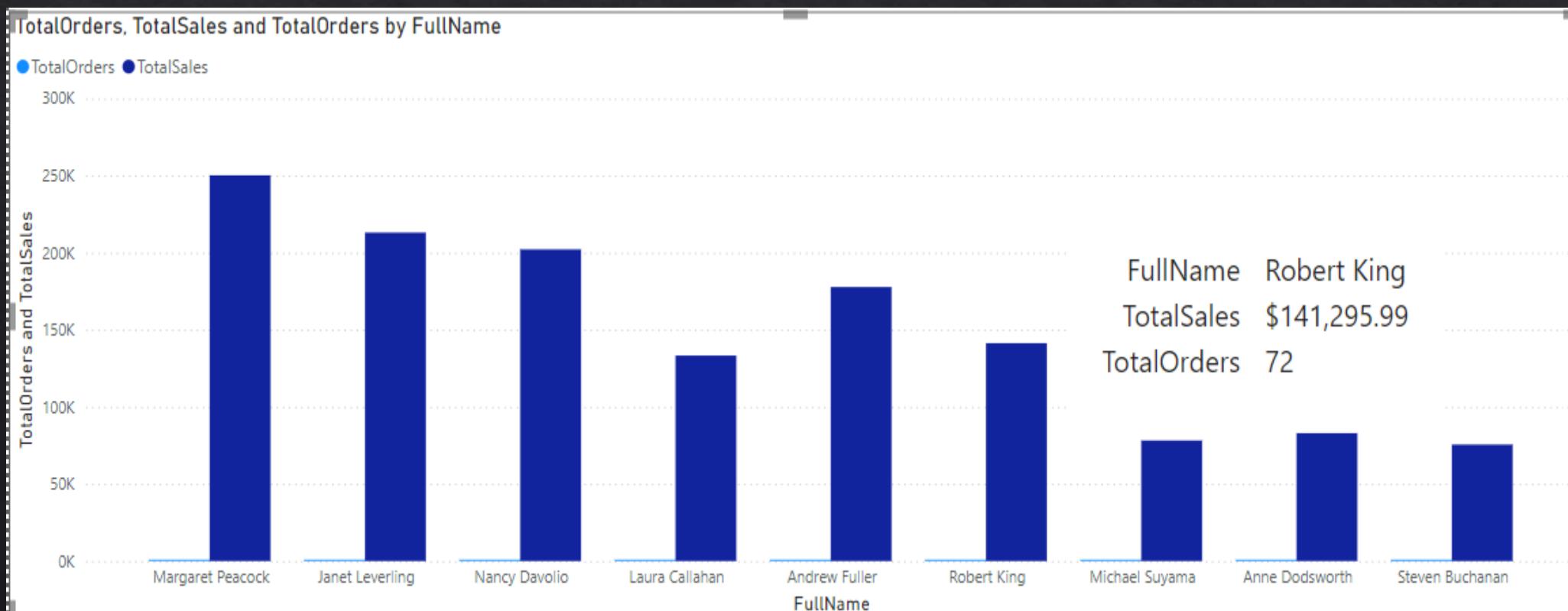
- Here's a Queries and chart showing Employee Order Management Analysis:

```

SELECT E.EmployeeID, E.FirstName + ' ' + E.LastName AS [Full Name],
       SUM(OD.Quantity * OD.UnitPrice) AS TotalSales
  FROM dbo.Orders AS O
 JOIN dbo.[Order Details] AS OD ON O.OrderID = OD.OrderID
 JOIN dbo.Employees AS E ON O.EmployeeID = E.EmployeeID
 GROUP BY E.EmployeeID, E.LastName, E.FirstName
 ORDER BY TotalSales DESC;
    
```



	EmployeeID	Full Name	TotalSales
1	4	Margaret Peacock	250187.45
2	3	Janet Leverling	213051.30
3	1	Nancy Davolio	202143.71
4	2	Andrew Fuller	177749.26
5	7	Robert King	141295.99
6	8	Laura Callahan	133301.03
7	9	Anne Dodsworth	82964.00
8	6	Michael Suyama	78198.10



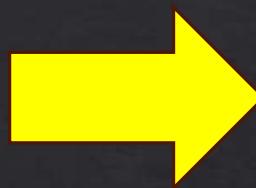
3.4. Sales Trends Over Time:

- ❖ This query calculates **monthly sales totals** by aggregating the **quantity** and **unit price** of orders for each **year** and **month**. The output is ordered chronologically, providing insights into sales trends over time. Key benefits include:
 1. **Seasonality Insights:** Identifying months or periods where sales peak or decline helps businesses adjust marketing and inventory strategies.
 2. **Performance Monitoring:** Comparing sales performance across different months and years provides a clear picture of growth and any market fluctuations.
 3. **Forecasting:** Recognizing historical trends enables more accurate sales forecasting, helping in better planning and decision-making.

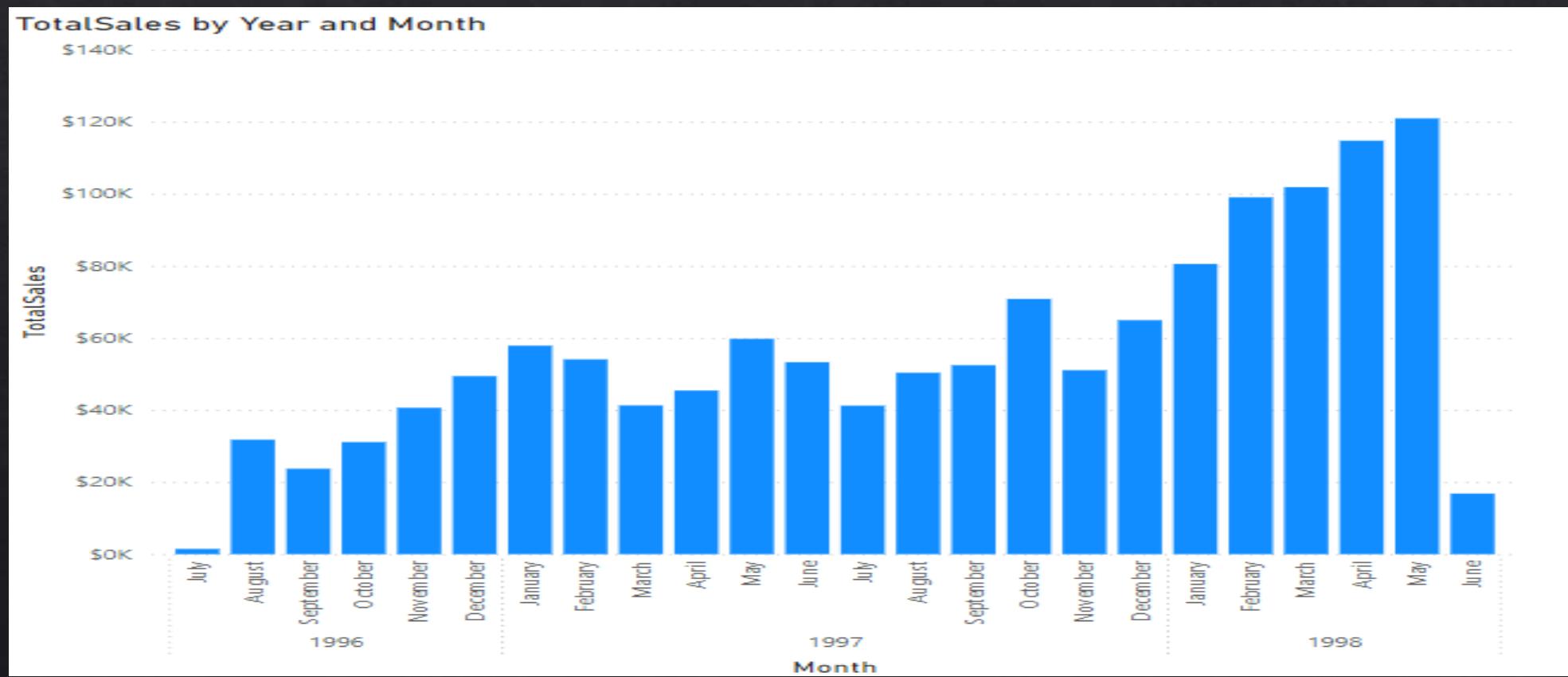
- Here's a Queries and chart showing Sales Trends Over Time :

```

SELECT YEAR(0.OrderDate) AS OrderYear,
       MONTH(0.OrderDate) AS OrderMonth,
       SUM(OD.Quantity * OD.UnitPrice) AS TotalSales
  FROM dbo.Orders AS O
 JOIN dbo.[Order Details] AS OD ON O.OrderID = OD.OrderID
 GROUP BY YEAR(0.OrderDate), MONTH(0.OrderDate)
 ORDER BY OrderYear, OrderMonth
    
```



	Results	Messages	
	OrderYear	OrderMonth	TotalSales
1	1996	7	30192.10
2	1996	8	26609.40
3	1996	9	27636.00
4	1996	10	41203.60
5	1996	11	49704.00
6	1996	12	50953.40
7	1997	1	66692.80
8	1997	2	41207.20



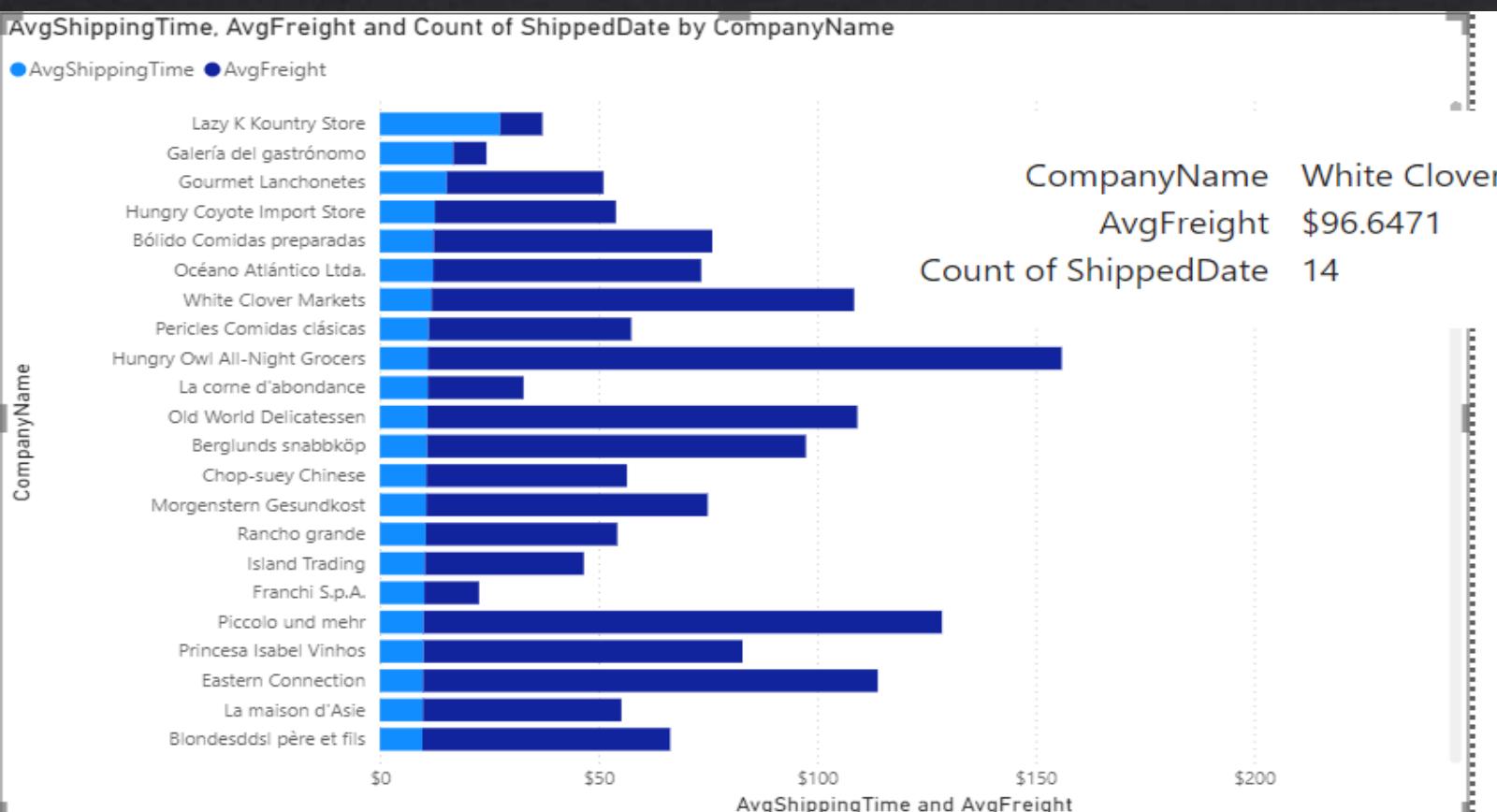
3.5. Shipping Performance by Shipper and Country:

- ❖ This query calculates the **average shipping time** and the **average freight cost** for each shipper, broken down by the destination country. It excludes orders that have not been shipped. The insights gained include:
 1. **Evaluate Shipper Efficiency:** By measuring the **average shipping time** for each shipper and country, businesses can assess which shippers are the most efficient.
 2. **Cost Management:** The **average freight cost** helps in determining the shipping costs and optimizing logistics expenses.
 3. **Improve Customer Satisfaction:** Understanding the shipping performance allows the company to make informed decisions on which shippers to prioritize for specific regions, potentially improving customer experience through faster delivery times.

- Here's a Queries and chart showing Shipping Performance by Shipper and Country:

```

SELECT S.CompanyName, O.ShipCountry,
       AVG(DATEDIFF(DAY, O.OrderDate, O.ShippedDate)) AS AvgShippingTime,
       Avg(O.Freight) AS AvgFreight
FROM dbo.Orders AS O
JOIN dbo.Shippers AS S ON O.ShipVia = S.ShipperID
WHERE O.ShippedDate IS NOT NULL
GROUP BY S.CompanyName, O.ShipCountry
ORDER BY S.CompanyName, AvgShippingTime ASC;
    
```



Results

	CompanyName	ShipCountry	AvgShippingTime	AvgFreight
1	Federal Shipping	Italy	4	40.672
2	Federal Shipping	Switzerland	4	104.2033
3	Federal Shipping	Argentina	5	17.4566
4	Federal Shipping	Belgium	5	34.9662
5	Federal Shipping	Canada	5	53.0181
22	Speedy Express	Norway	2	13.37
23	Speedy Express	Finland	4	34.7512
24	Speedy Express	Austria	5	184.8808
25	Speedy Express	Brazil	7	49.7716
26	Speedy Express	Denmark	7	66.8066
43	United Package	Finland	5	79.916
44	United Package	Canada	5	110.5688
45	United Package	Denmark	5	146.6175
46	United Package	Mexico	6	55.9088
47	United Package	Portugal	6	79.9425

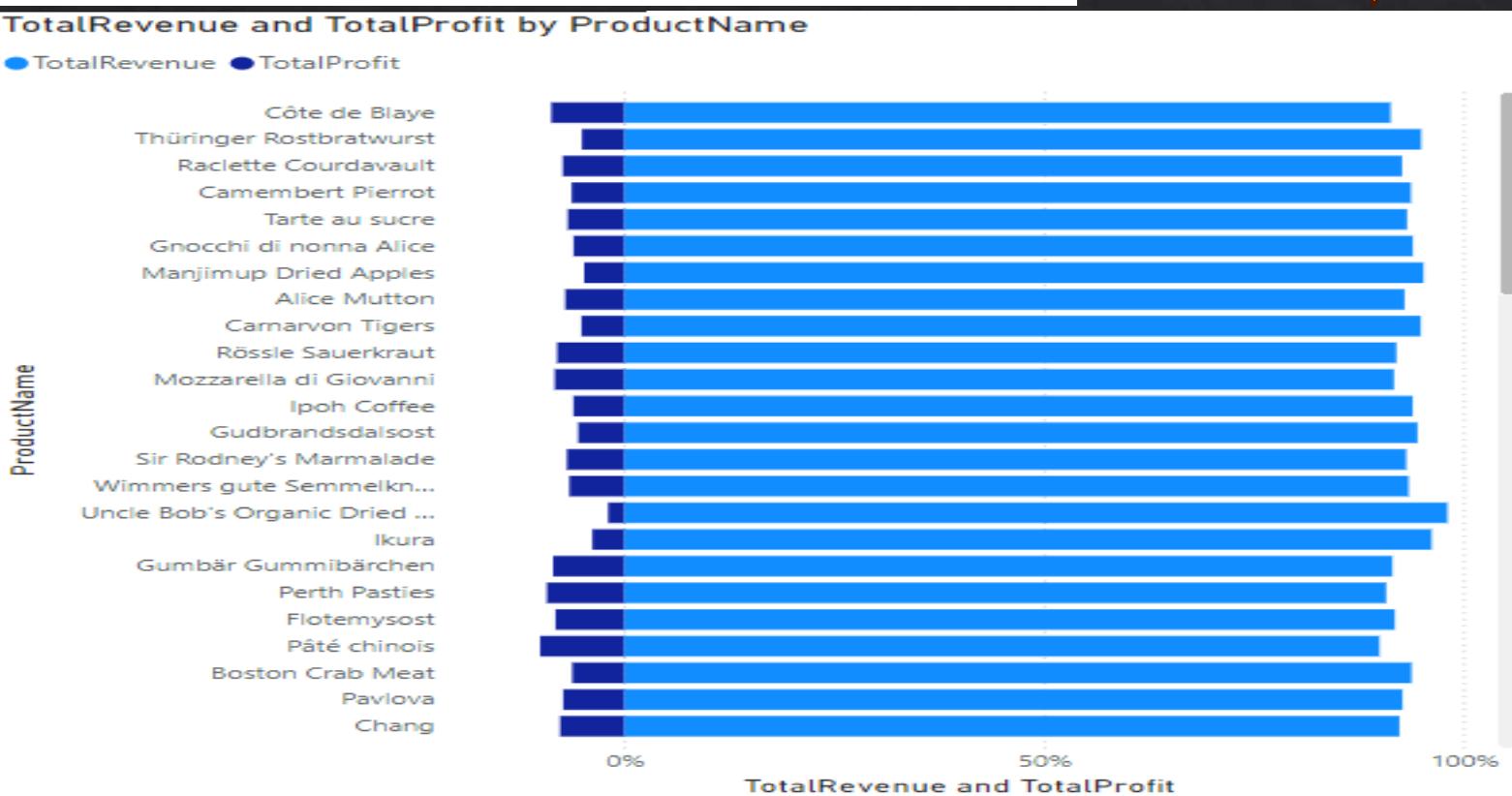
3.6. Product Profitability Analysis:

- ❖ This query calculates the **total profit** generated by each product by subtracting the **product cost** (P.UnitPrice) from the **sales price** (OD.UnitPrice), then multiplying it by the **quantity sold**. It helps rank the products based on profitability. Key benefits include:
 1. **Identify High-Profit Products:** Pinpointing the most profitable products allows the company to focus on promoting and selling those items.
 2. **Optimize Pricing:** Insights from product profitability can help fine-tune pricing strategies, ensuring that the business maximizes profit margins.
 3. **Inventory Decisions:** Understanding which products yield the highest profits can guide stocking and inventory decisions, helping ensure the most profitable items are always available.

- Here's a Queries and chart showing Product Profitability Analysis:

```

SELECT P.ProductName,
       SUM((OD.UnitPrice - P.UnitPrice) * OD.Quantity) AS TotalProfit
  FROM dbo.[Order Details] AS OD
 JOIN dbo.Products AS P ON OD.ProductID = P.ProductID
 GROUP BY P.ProductName
 ORDER BY TotalProfit DESC;
    
```



Results

ProductName	TotalProfit
Laughing Lumberjack Lager	-14.00
Gustaf's Knäckebröd	-75.60
Genen Shouyu	-77.50
Rogede sild	-85.50
Konbu	-111.60
Filo Mix	-116.20
Geitost	-174.00
Grandma's Boysenberry Spread	-180.00

Messages

?

Results

ProductName	TotalProfit
Rössle Sauerkraut	-2318.40
Alice Mutton	-2659.80
Gnocchi di nonna Alice	-2872.80
Camembert Pierrot	-3332.00
Tarte au sucre	-3564.00
Thüringer Rostbratwurst	-4610.94
Raclette Courdavault	-5984.00
Côte de Blaye	-14176.30

Messages

3.7. Order Delivery Status Analysis:

- ❖ This query determines the **status of each order** based on its **shipping date** and **required date**. It classifies orders into three categories:
 - **Late**: Orders shipped after the required date.
 - **Cancelled**: Orders that were never shipped (ShippedDate is NULL).
 - **On Time**: Orders shipped on or before the required date.
- ❖ This analysis helps in:
 1. **Assessing Delivery Performance**: By identifying late or cancelled orders, the company can evaluate its ability to meet customer expectations.
 2. **Improving Logistics**: Insights into delivery delays enable the company to optimize supply chain and shipping processes.
 3. **Customer Satisfaction**: Understanding delivery performance can help address service issues, improving overall customer experience.

- Here's a Queries and chart showing Order Delivery Status Analysis:

SELECT

```
o.OrderID,
o.OrderDate,
o.RequiredDate,
o.ShippedDate,
```

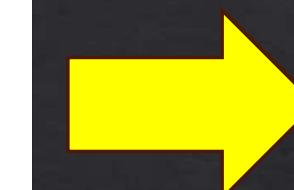
CASE

```
WHEN o.ShippedDate > o.RequiredDate THEN 'Late'
WHEN o.ShippedDate IS NULL THEN 'Cancelled'
ELSE 'On Time'
```

END AS OrderStatus

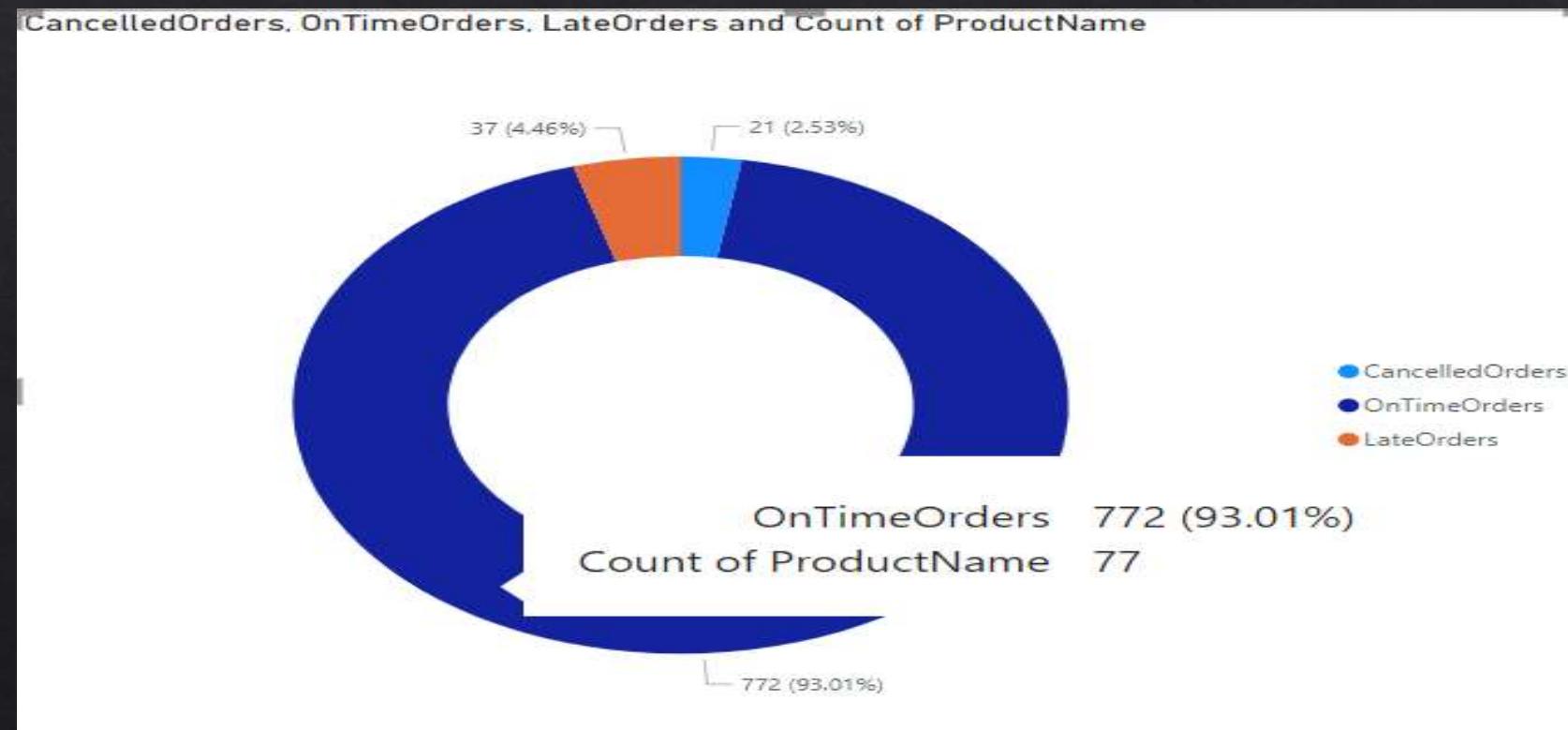
FROM

```
Orders o;
```



Results Messages

	OrderID	OrderDate	RequiredDate	ShippedDate	OrderStatus
29	10276	1996-08-08 00:00:00.000	1996-08-22 00:00:00.000	1996-08-14 00:00:00.000	On Time
30	10277	1996-08-09 00:00:00.000	1996-09-06 00:00:00.000	1996-08-13 00:00:00.000	On Time
31	10278	1996-08-12 00:00:00.000	1996-09-09 00:00:00.000	1996-08-16 00:00:00.000	On Time
32	10279	1996-08-13 00:00:00.000	1996-09-10 00:00:00.000	1996-08-16 00:00:00.000	On Time
33	10280	1996-08-14 00:00:00.000	1996-09-11 00:00:00.000	1996-09-12 00:00:00.000	Late
809	11056	1998-04-28 00:00:00.000	1998-05-12 00:00:00.000	1998-05-01 00:00:00.000	On Time
810	11057	1998-04-29 00:00:00.000	1998-05-27 00:00:00.000	1998-05-01 00:00:00.000	On Time
811	11058	1998-04-29 00:00:00.000	1998-05-27 00:00:00.000	NULL	Cancelled
812	11059	1998-04-29 00:00:00.000	1998-06-10 00:00:00.000	NULL	Cancelled



4.1. Preprocessing through Power Query:

Advanced Editor



Customers

Display Options ▾



```
let
    Source = Sql.Database("ALQMASH\SQLEXPRESS", "northwind", [CreateNavigationProperties=false]),
    dbo_Customers = Source{[Schema="dbo", Item="Customers"]}[Data],
    #"Removed Columns" = Table.RemoveColumns(dbo_Customers, {"Address", "Region", "PostalCode", "Phone", "Fax"})
in
    #"Removed Columns"
```

✓ No syntax errors have been detected.

Done

Cancel

For the Customers table, here's what you did in Power Query:

1.Disabled Navigation Properties: You added [CreateNavigationProperties=false] to the Sql.Database function, which prevents Power Query from automatically creating relationships between the Customers table and other related tables. This gives you more control over managing these relationships manually.

2.Loaded the Customers Table: You accessed the Customers table from the dbo schema using:

```
dbo_Customers = Source{[Schema="dbo", Item="Customers"]}[Data],
```

3. Removed Unnecessary Columns: You removed columns that are not essential for your analysis.

Specifically, you removed: "Address": Customer's street address, which may not be needed in your visualizations. "Region": The region column, possibly not useful for the scope of your analysis. "postal Code": The postal code, which might not contribute directly to your insights. "Phone": Customer phone numbers, which might not be critical for your dashboard. "Fax": Fax numbers, likely outdated or irrelevant.

Summary of Transformations: The goal of these changes is to clean up the Customers table by focusing only on key columns like CustomerID, Company Name, Contact Name, and other critical information. This makes the data more streamlined for your visualizations in Power BI, improving performance and clarity.

4.2. Preprocessing through Power Query:

Advanced Editor



Employees

Display Options ▾



✓ No syntax errors have been detected.

Done

[Cancel](#)

4.2. Preprocessing through Power Query:

For the Employees table, here's what you've done in Power Query:

1.Disabled Navigation Properties: Similar to the Customers table, you added

[CreateNavigationProperties=false] to the Sql.Database function. This disables automatic relationship creation between the Employees table and other tables.

2.Loaded the Employees Table: You accessed the Employees table from the dbo schema

using:

- ◆ dbo_Employees = Source{[Schema="dbo", Item="Employees"]}[Data],

3.Removed Unnecessary Columns: You removed some columns that aren't essential for your dashboard, including:"TitleOfCourtesy": Titles like Mr., Ms., etc., which may not be necessary for the analysis."BirthDate", "HireDate": These date fields might not be needed for the current visualization goals."Address", "City", "Region", "PostalCode": Geographical and postal details, which might not add value in this context.

4.Added a Custom Column (Full Name): You created a new column "FullName" by concatenating the FirstName and LastName columns:

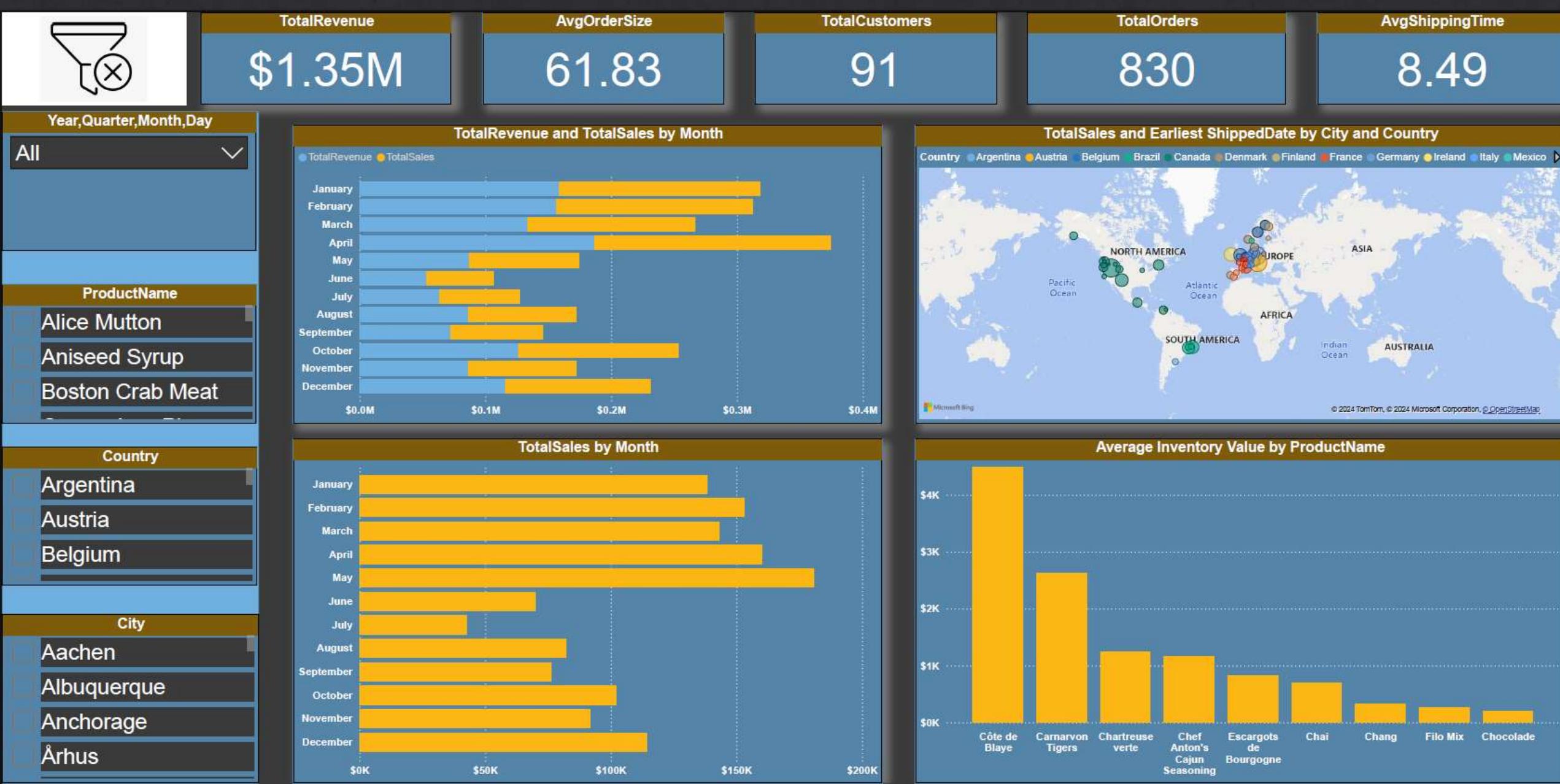
```
#"Added Custom" = Table.AddColumn(#"Removed Columns", "FullName", each [FirstName] & " "  
& [LastName]),
```

5. Changed Data Type: You ensured that the new "FullName" column has the correct data type, which is text:

```
#"Changed Type" = Table.TransformColumnTypes(#"Added Custom", {{"FullName", type text}})
```

Summary of Transformations: In the Employees table, you streamlined the data by removing unnecessary columns, created a new column for the employee's full name to simplify visualizations, and ensured the correct data type for the new column. This will make it easier to work with employee-related data in your Power BI dashboard, especially when displaying employee names in charts or tables.

5. Visualizations through Power BI:



❖ 1. Key Metrics Cards:

- **Total Revenue:** Displays the overall sales revenue (\$1.35M).
- **Avg Order Size:** Shows the average size of orders placed (61.83).
- **Total Customers:** The total number of customers (91).
- **Total Orders:** The total count of orders made (830).
- **Avg Shipping Time:** The average time it takes for an order to be shipped (8.49 days).

◆ 2. Filters:

- **Year/Quarter/Day Filter:** Enables filtering the data based on time periods.
- **Product Name Filter:** Allows filtering the data by specific products (e.g., Alice Mutton, Aniseed Syrup).
- **Country and City Filters:** Enables data exploration by geographical location.

◆ 3. Charts & Visuals:

- **Total Revenue and Total Sales by Month:** A bar chart comparing revenue and sales across each month of the year.
- **Total Sales by Month:** Another chart focused purely on sales figures over the months.
- **Map Visual:** Depicts the total sales and the earliest shipping dates by city and country, showing global distribution of orders.
- **Average Inventory Value by Product Name:** A bar chart comparing the average inventory value for different products.

❖ **4. Insights Provided:**

- The dashboard allows quick analysis of sales performance by month, product, and location.
- Key metrics such as revenue, orders, and shipping time are clearly highlighted for decision-making.
- Filters provide flexibility in analyzing different dimensions like time, product, country, and city.

6. Recommendations for Improvement:

- ❖ This section focuses on recommendations aimed at improving business performance based on the analysis conducted. Here are potential areas for improvement:
- **Inventory Management Optimization:** Based on the analysis of high-performing and low-performing products, the company can improve inventory management by focusing on stocking high-demand products and better managing low-performing items. This can reduce stockouts and excess inventory.
- **Enhancing Customer Experience:** By identifying top-spending customers, the company can develop targeted marketing campaigns and special offers for these high-value customers. This will help increase customer loyalty and drive repeat business.
- **Improving Shipping Efficiency:** Through the analysis of shipping performance by shipper and country, the company can enhance logistical efficiency by selecting the most reliable and fastest shippers based on the destination. This can reduce shipping times and potentially lower shipping costs.
- **Boosting Employee Performance:** Based on the employee performance analysis, the company can create focused training programs for underperforming employees and introduce incentive programs for top-performing staff to motivate and improve overall productivity.