

Capstone Project North Wind Traders Sales Analytics

By - Divyansh Mishra





Topics

- ❖ Project Overview
- ❖ Key Components of Report
- ❖ Information about Dataset Tables
- ❖ Data Cleaning and MECE
- ❖ Power BI Questions Discussions
- ❖ EDA Questions Discussions
- ❖ Dashboard Analysis

Project Overview

Objective:

The objective of this Power BI report is to develop a comprehensive dashboard for Northwind Traders, a fictitious company specializing in importing and exporting specialty foods. The report aims to provide valuable insights into various aspects of the company's operations, including sales analysis, customer segmentation, inventory trends, and employee performance. The primary goal is to facilitate data-driven decision-making processes for stakeholders by presenting key performance metrics in an intuitive and visually appealing manner.

Dataset Description:

The dataset used for this project is the Northwind database, containing sales data for Northwind Traders. It comprises several interconnected tables, including Customers, Employees, Orders, Order Details, Products, Suppliers, Shippers, and Categories. Each table contains specific information related to customers, employees, orders, products, suppliers, and shipping details, enabling comprehensive analysis of the company's operations.

Key Components of the Report:

- 1. Sales Analysis:**
 1. Visualizations depicting sales trends over time, including total sales, sales by product category, and geographical distribution of sales.
 2. Analysis of top-selling products, customer segments, and sales performance by region or customer demographics.
- 2. Customer Segmentation:**
 1. Segmentation of customers based on various attributes such as geographic location, purchase history, and customer type.
 2. Customer retention analysis and identification of high-value customers for targeted marketing strategies.
- 3. Inventory Trends:**
 1. Visualizations illustrating inventory levels, including stock availability, products on order, and reorder levels.
 2. Analysis of inventory turnover rates, identifying slow-moving or obsolete products.
- 4. Employee Performance:**
 1. Evaluation of employee performance metrics such as sales contribution, order processing time, and customer satisfaction ratings.
 2. Comparison of performance across different sales territories or teams.

Information about Dataset Tables

- ❖ What is Dataset ?

- A dataset is a collection of data.

- ❖ In this Project, we have got dataset from a company called “Northwind Traders”. A company that deals with importing and exporting food items across different countries.

- ❖ In this dataset, we have 8 different tables.

1. Customers Table
2. Employees Table
3. Orders Table
4. Order Details Table
5. Products Table
6. Suppliers Table
7. Shippers Table
8. Categories Table

Information about Dataset Tables Continue

1. **Customers Table:** - This table stores information about the company's customers. It includes fields for customer ID, company name, contact name, contact title, address, city, region, postal code, country, phone and fax.
2. **Employees Table:** - This table stores information about the company's employees. It includes fields for employee ID, last name, first name, title, title of courtesy, birth date, hire date, address, city, region, postal code, country, home phone, extension, photo, notes, reports to, and photo path.
3. **Orders Table:** - This table stores information about the company's orders. It includes fields for order ID, customer ID, employee ID, order date, required date, shipped date, ship via, freight, ship name, ship address, ship city, ship region, ship postal code, and ship country.
4. **Order Details Table:** - This table stores detailed information about the items within each order. It includes fields for order ID, product ID, unit price, quantity, and discount.

Information about Dataset Tables Continue

5. **Products Table:** - This table stores information about the company's products. It includes fields for product ID, product name, supplier ID, category ID, quantity per unit, unit price, units in stock, units on order, reorder level, and whether the product is discontinued.
6. **Suppliers Table:** - This table stores information about the company's suppliers. It includes fields for supplier ID, company name, contact name, contact title, address, city, region, postal code, country, phone, fax, and home page.
7. **Shippers Table:** - This table stores information about the company's employees. It includes fields for employee ID, last name, first name, title, title of courtesy, birth date, hire date, address, city, region, postal code, country, home phone, extension, photo, notes, reports to, and photo path.
8. **Categories Table:** - This table stores information about the company's orders. It includes fields for order ID, customer ID, employee ID, order date, required date, shipped date, ship via, freight, ship name, ship address, ship city, ship region, ship postal code, and ship country.

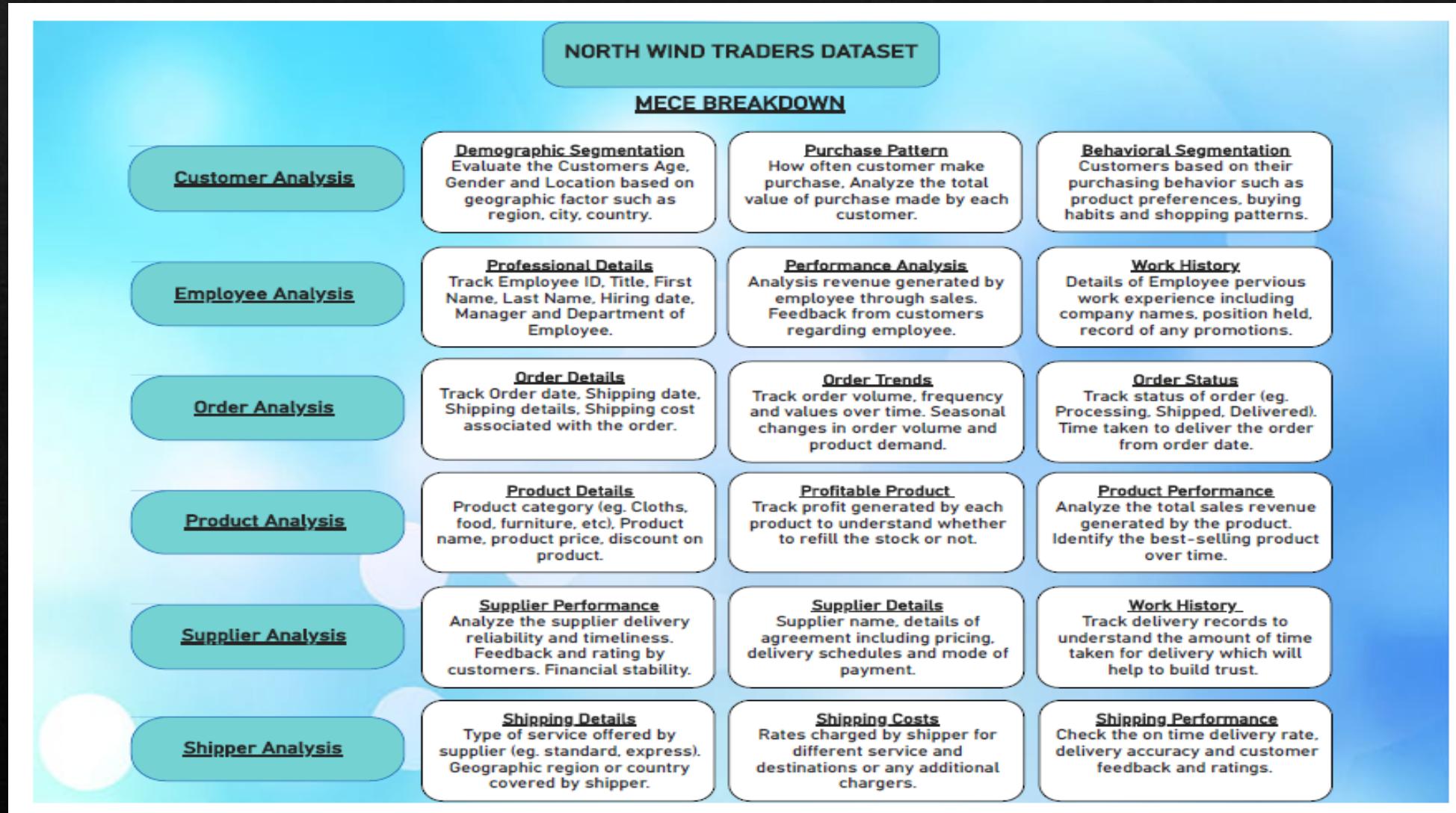
Clean Dataset Tables

1. **Customers Table:** - After cleaning the unnecessary columns. It includes fields for customer ID, company name, contact name, contact title, city, country.
2. **Employees Table:** - After cleaning the unnecessary columns. It includes fields for employee ID, last name, first name, title, title of courtesy, birth date, hire date, city, country
3. **Orders Table:** - After cleaning the unnecessary columns in the company's orders table. It includes fields for order ID, customer ID, employee ID, order date, required date, shipped date, ship via, freight, ship name, ship city, and ship country.
4. **Order Details Table:** - After cleaning the unnecessary columns in the company's order details table. It includes fields. It includes fields for order ID, product ID, unit price, quantity, and discount.

Clean Dataset Tables

5. **Products Table:** - After cleaning unnecessary data. It includes fields for product ID, product name, supplier ID, category ID, quantity per unit, unit price, units in stock, units on order, reorder level, and whether the product is discontinued.
6. **Suppliers Table:** - After cleaning unnecessary data. It includes fields for supplier ID, company name, contact name, contact title, city, region, postal code, country,
7. **Shippers Table:** - After cleaning unnecessary data. It includes fields for employee ID, last name, first name, title, title of courtesy, birth date, hire date, city, region, postal code, country.
8. **Categories Table:** - After cleaning unnecessary data. It includes fields for order ID, customer ID, employee ID, order date, required date, shipped date, ship via, freight, ship name, ship city, and ship country.

MECE Breakdown



Power Bi Questions

Power Bi Question

1. How does customer distribution vary across different regions or customer segments? Can we visualize it on a map or bar chart?

Answer:

Yes, we can visualize using Map Chart to understand Total Customers across different countries.

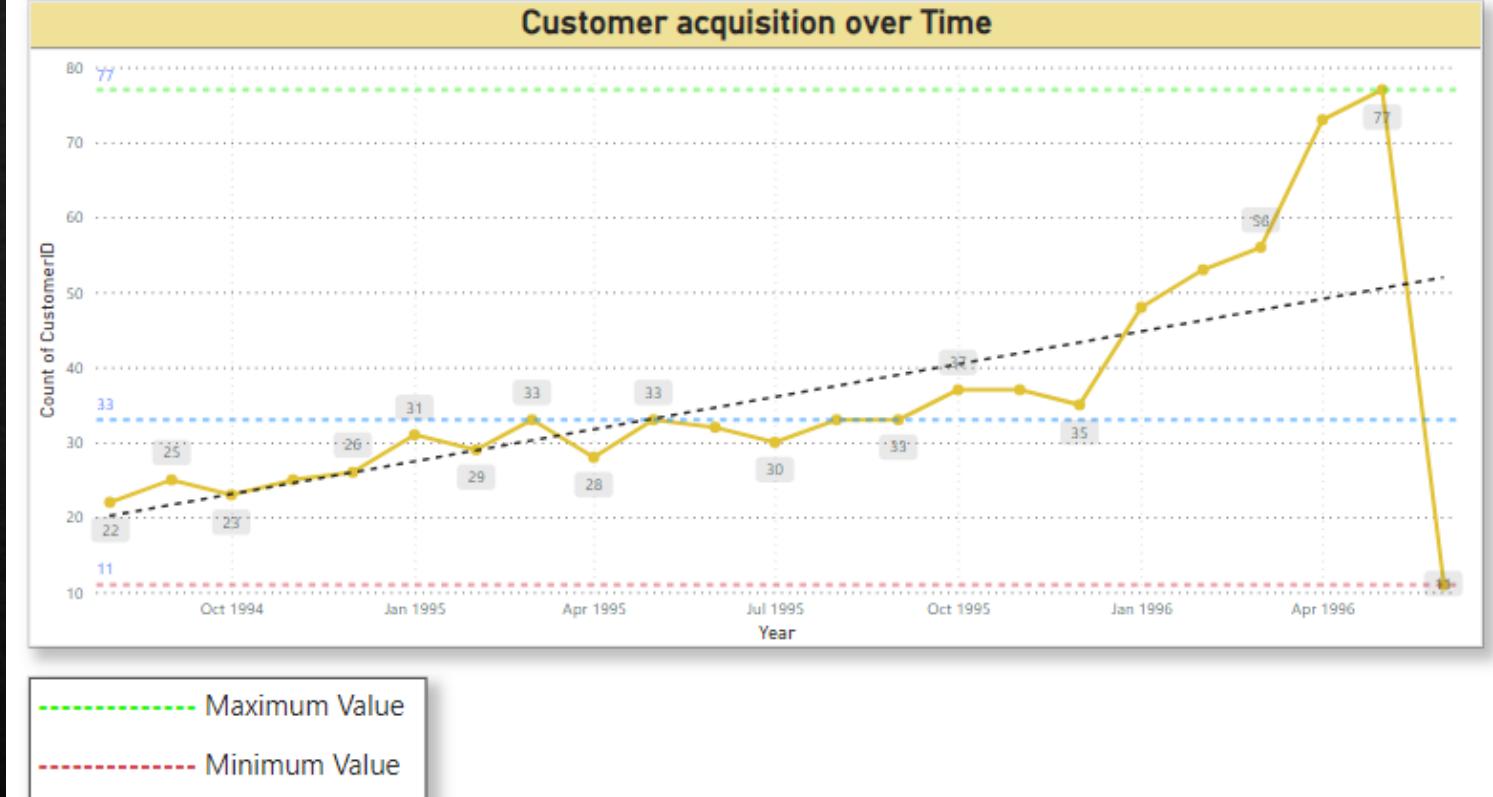


Power Bi Question

2. What is the trend in customer acquisition over time? Can we create a line chart or area chart to display it ?

Answer:

Yes, we can create line chart to visualize the trend in customer acquisition over time. It help us to understand customers behavior changing over the years.

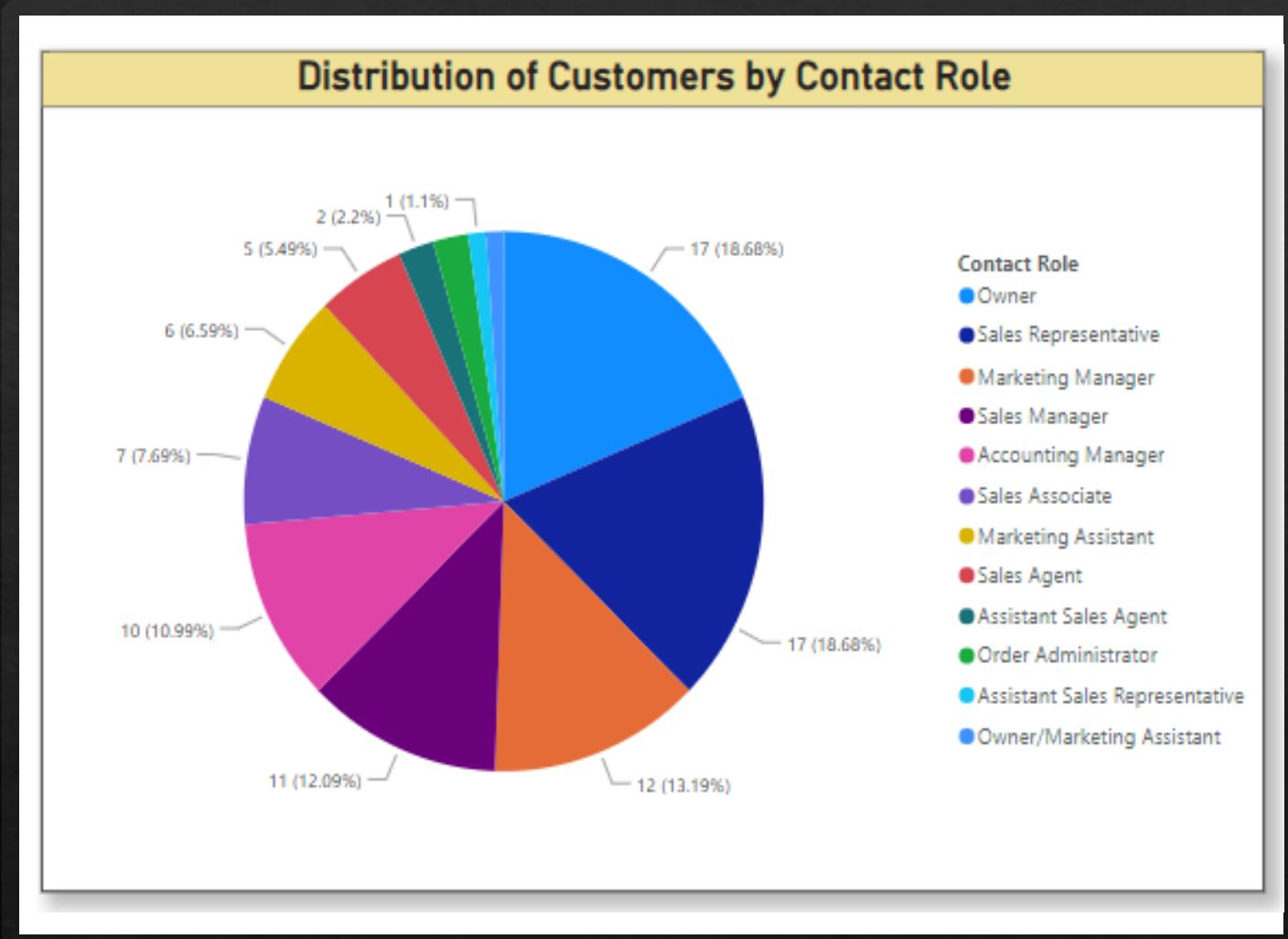


Power Bi Question

3. Can we visualize the distribution of customer demographics such as age, gender, or income using histograms or pie charts ?

Answer:

Yes, we can create pie chart to visualize customer distribution. It shows 17 customers having contact role as Owner which is 18.68% of customers.

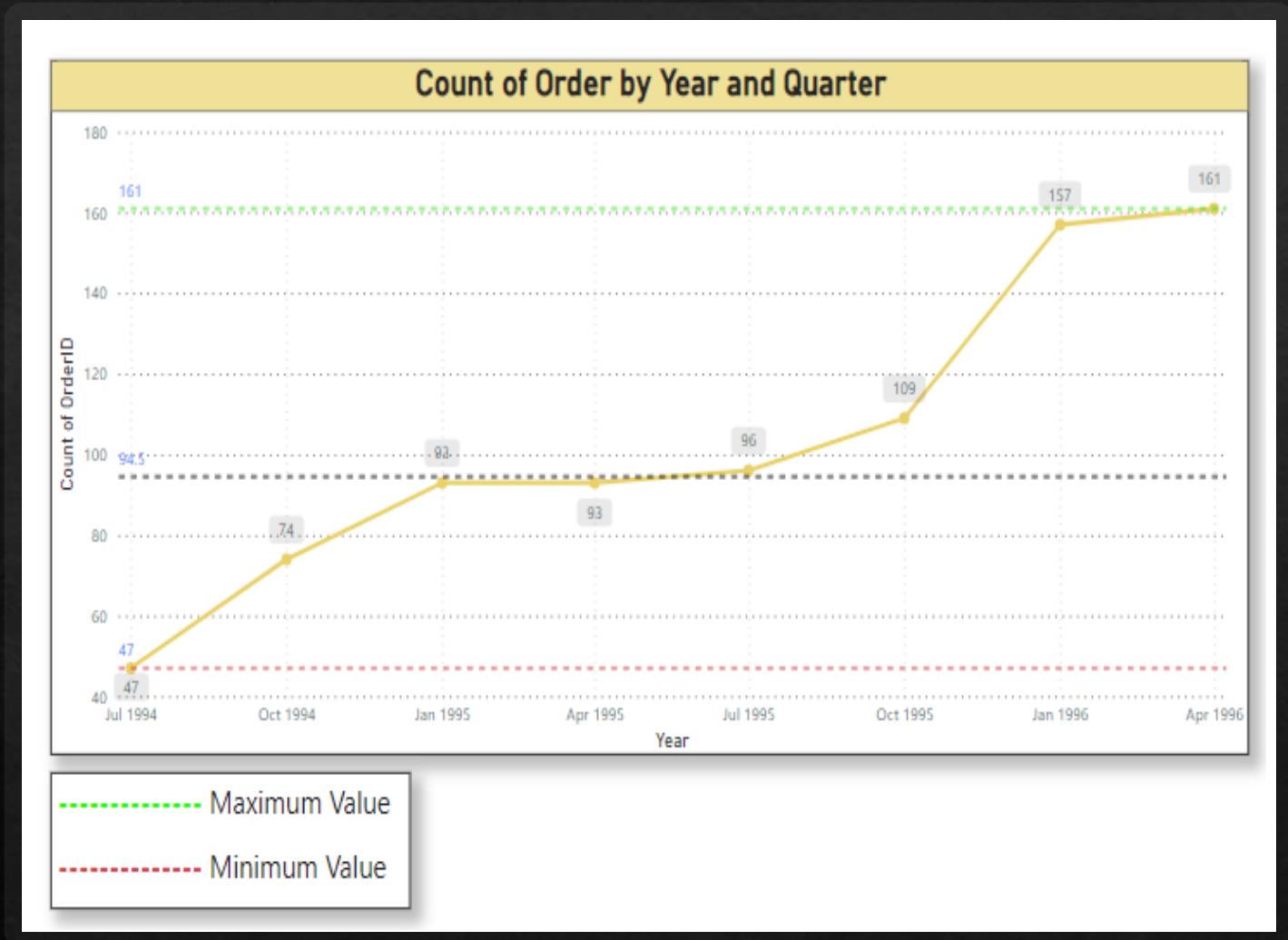


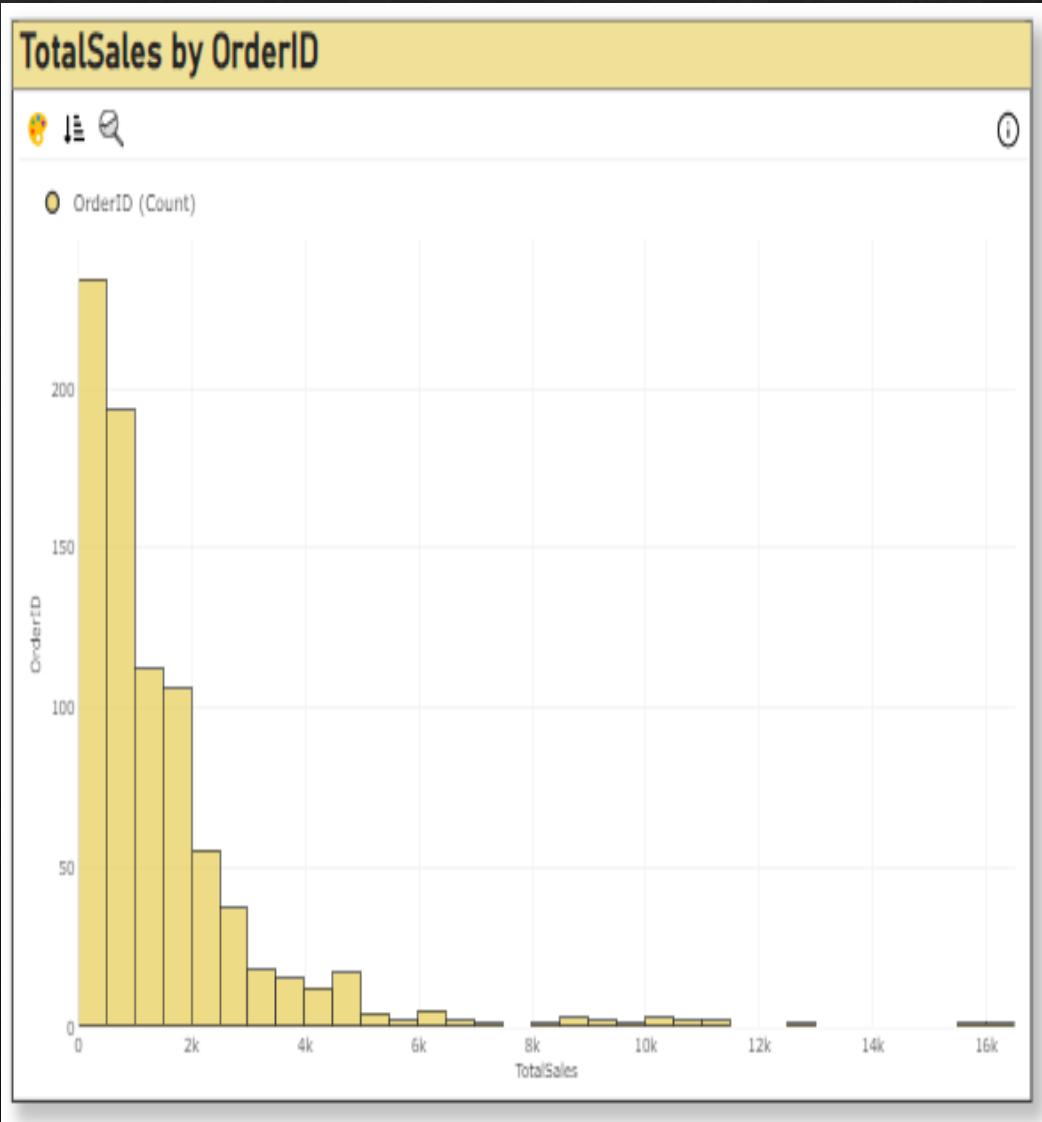
Power Bi Question

4. How does order volume change over time ? Can we create a time series chart or stacked bar chart to visualize it ?

Answer:

Yes, we can create line chart for time visualization. It display no. of orders over different quarter and year. Which will help us to identify orders trend over time.





Power Bi Question

5. What is the distribution of order values ? Can we create a histogram or box plot to display it ?

Answer:

Yes, we can create histogram chart to understand order values as Total Sales using no. of orders. Which will help us to gain information about how much orders vary in different total sales range. Maximum orders lies between 0 to 2000 having 234 orders.

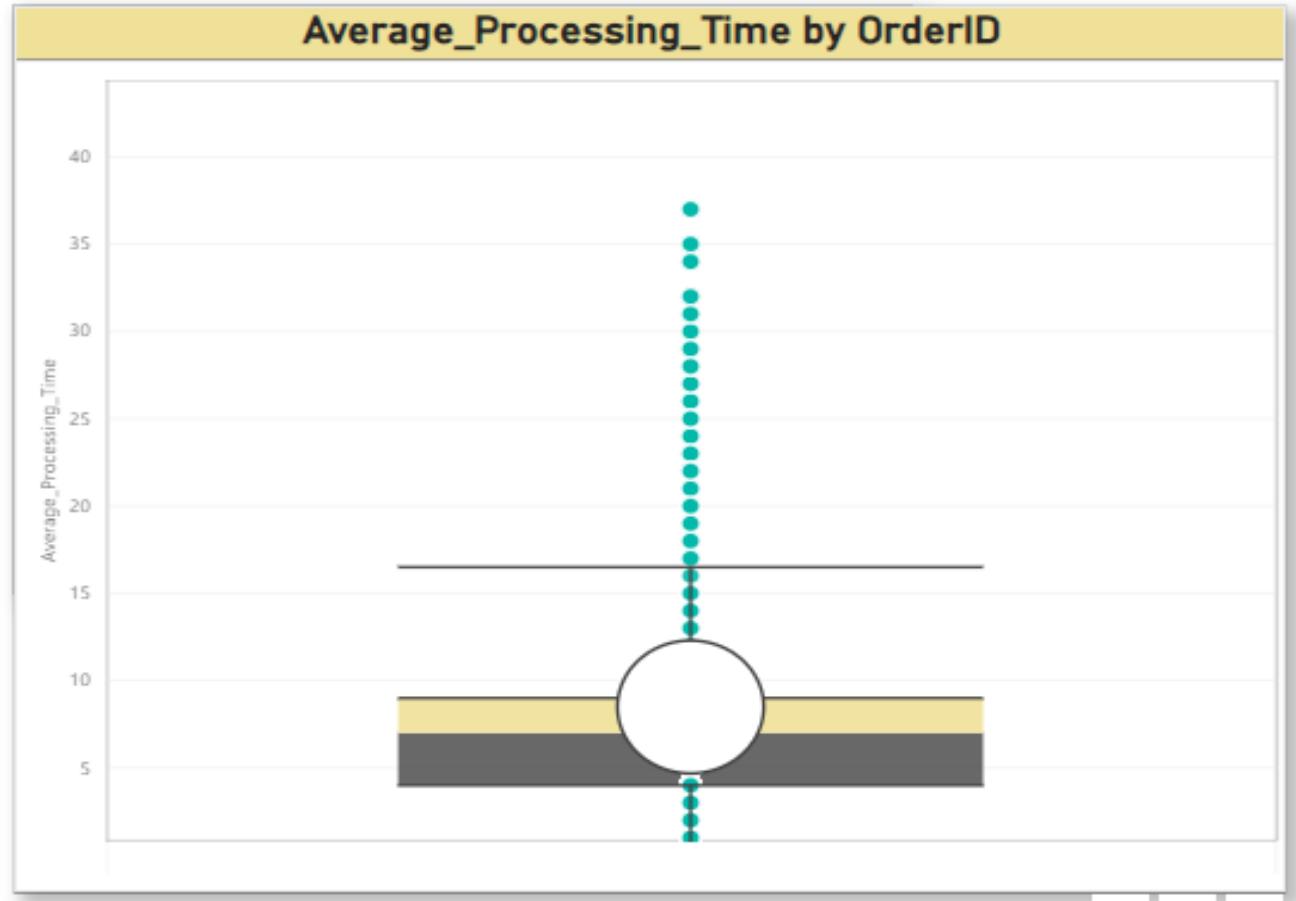
Power Bi Question

6. Can we visualize the average order processing time or shipping duration using a bar chart or box plot ?

Answer:

Yes, we can create box plot chart. It help us to identify outliers. The median processing time (the middle of the box) is around 7 days and mean is 8.49 days.

The interquartile range (IQR), represented by the box itself, is around 5 days, indicating that the middle 50% of orders take between 4 and 9 days to process.

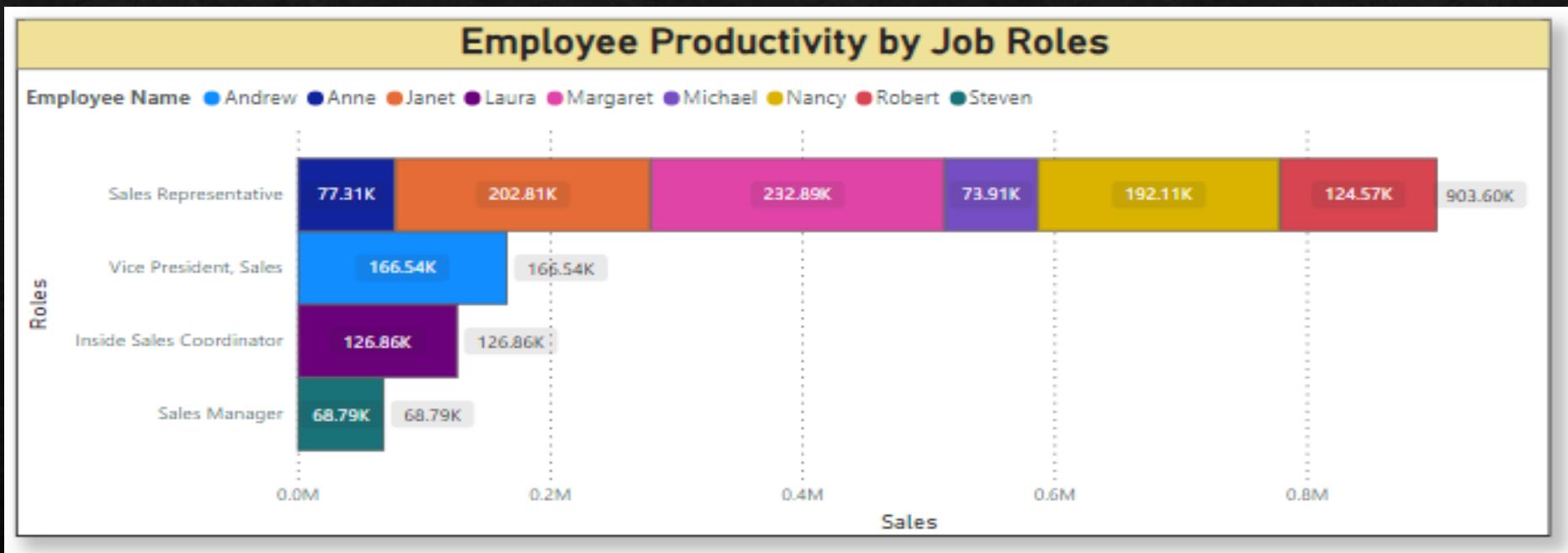


Power Bi Question

7. How does employee productivity vary across different departments or job roles? Can we create a stacked bar chart or grouped column chart to visualize it?

Answer:

Yes, we can create stacked bar chart, which shows sales representative have highest employee productivity around 903k and Employee Name Margaret have 232.89k sales.



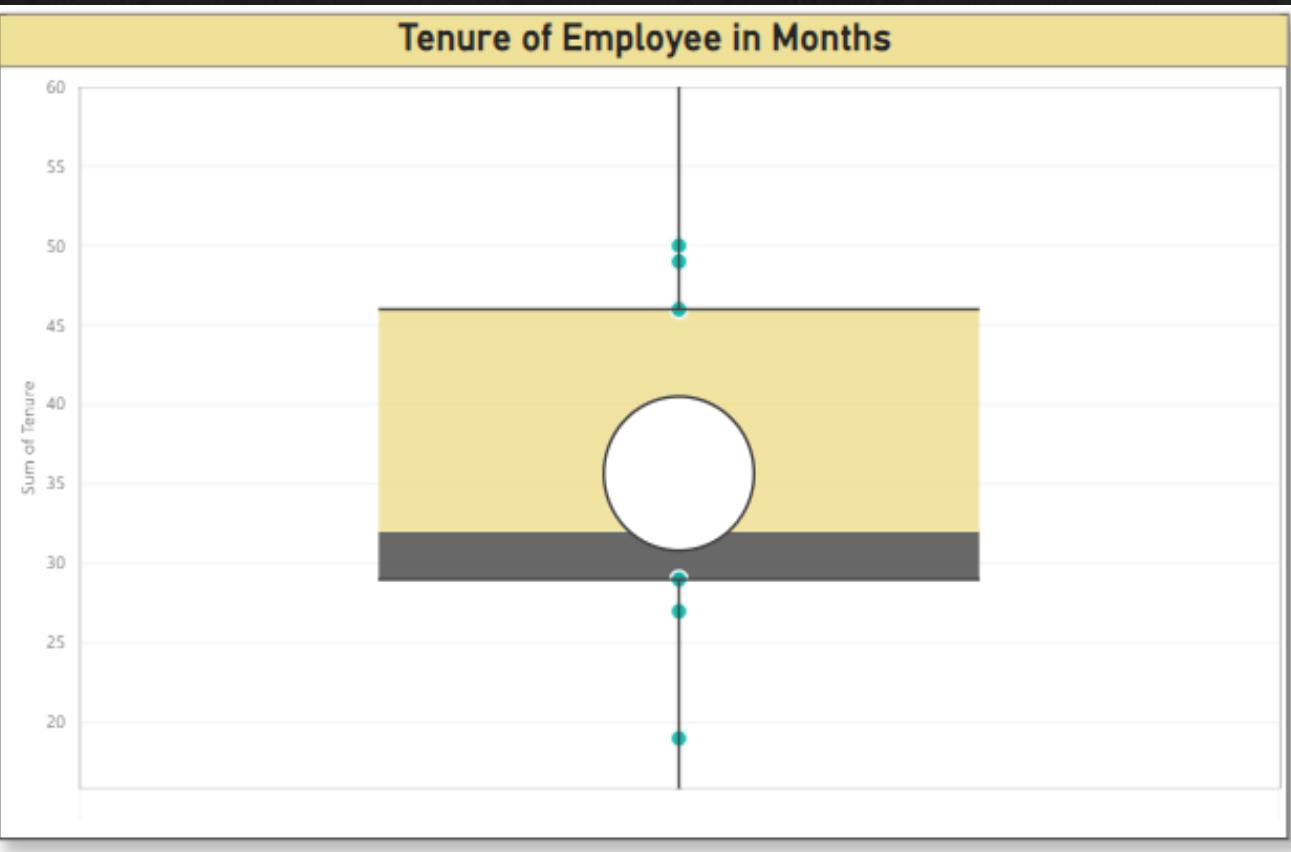
Power Bi Question

8. What is the distribution of employee tenure ? Can we create a histogram or box plot to display it ?

Answer:

Yes, we can create box plot chart, It appears to show employee tenure data from a company. The median tenure is 32 months and mean is 35.67.

The interquartile range (IQR) is around 17 months, most of employees have tenures between 29 and 46 months.

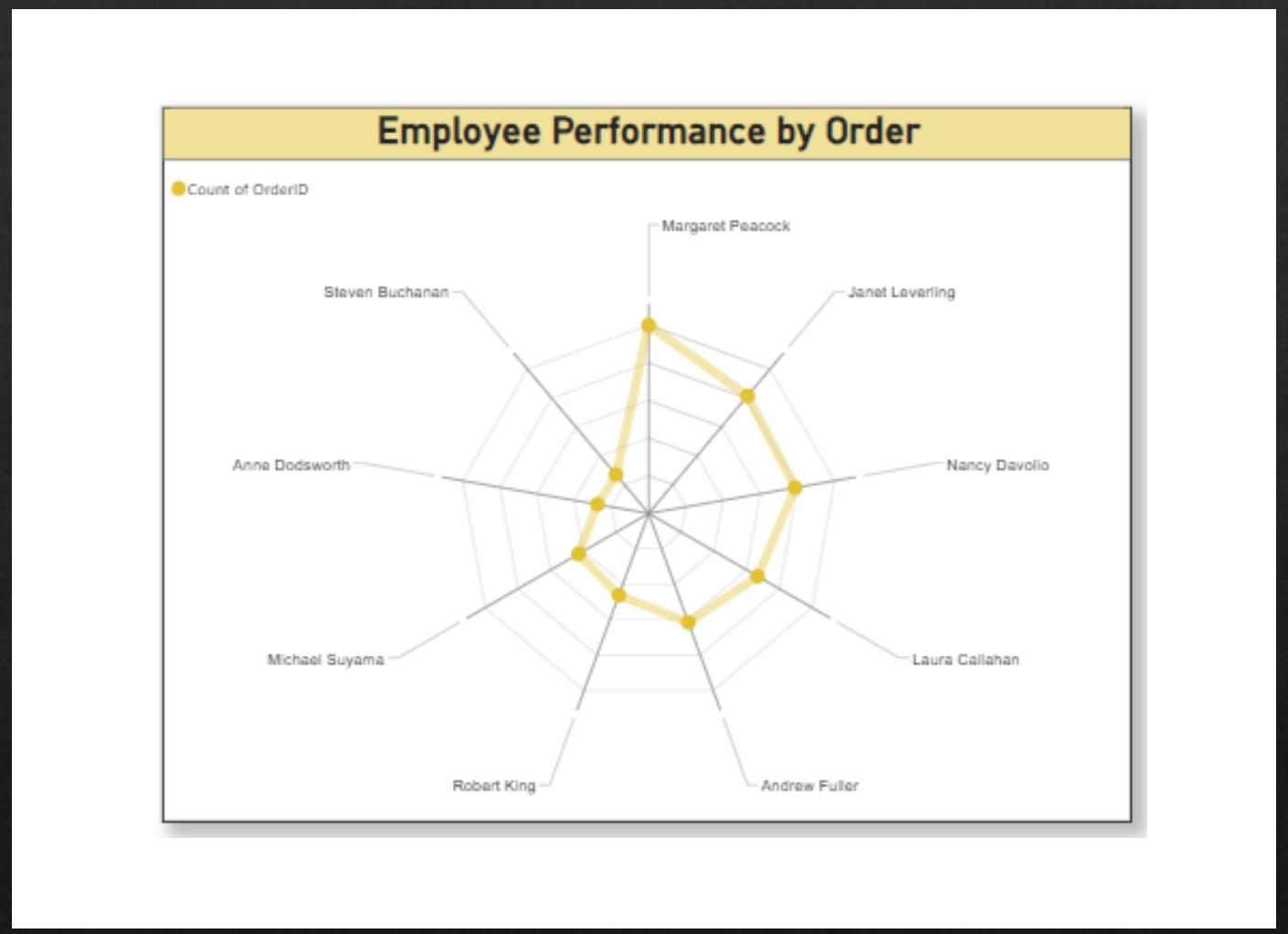


Power Bi Question

9. Can we visualize employee performance ratings or KPIs using a radar chart or bullet graph?

Answer:

Yes, we can create radar chart to check employee performance by number of orders. It shows Margaret Peacock have sold maximum 156 orders help us to identify employee performance.



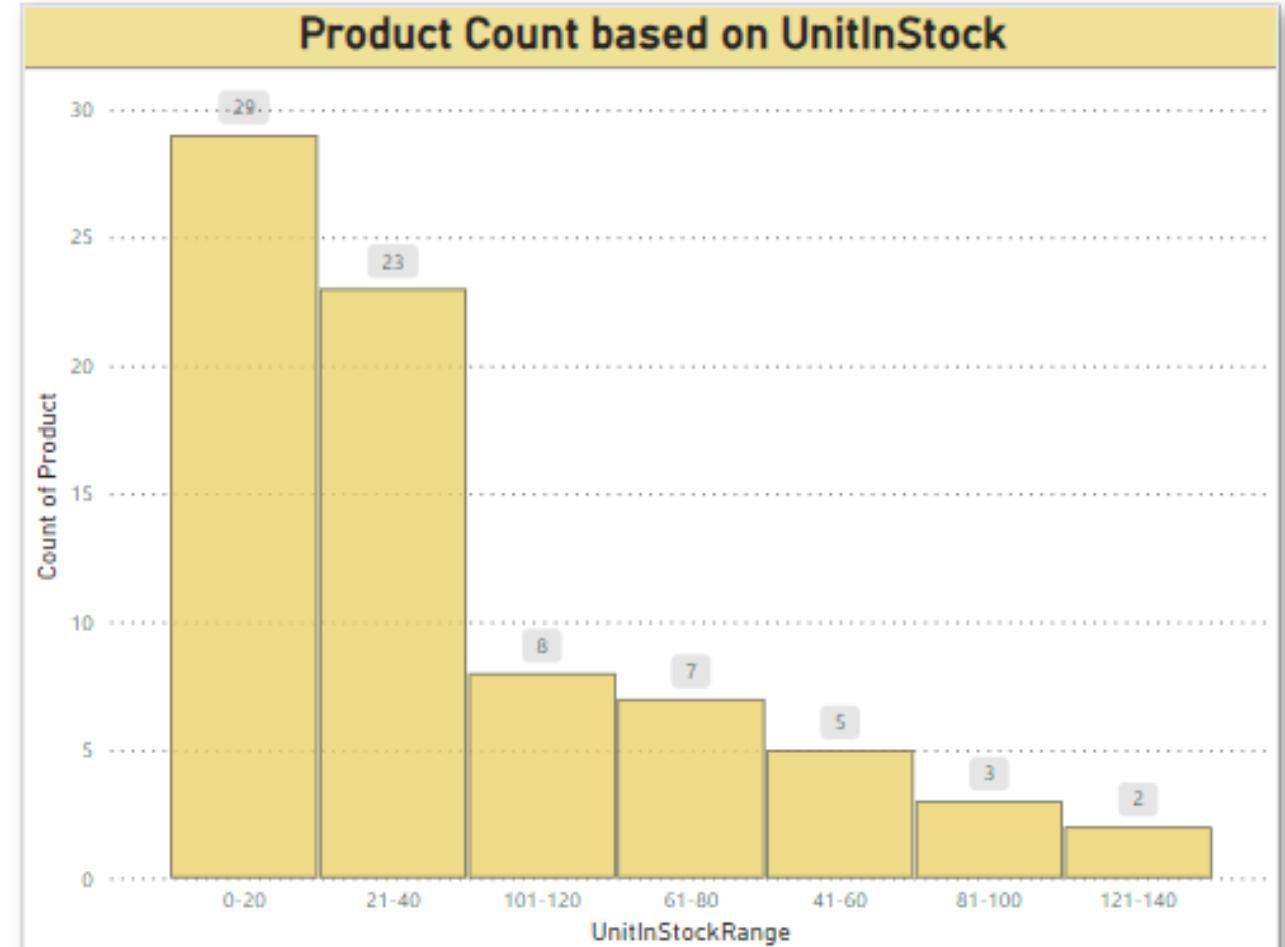
Power Bi Question

10. What is the distribution of product ratings or reviews? Can we create a histogram or stacked bar chart to visualize it?

Answer:

Yes, we can create histogram chart to find product rating based on product count. Maximum the product in stock, maximum demand of that product.

It help us to identify the best rating product.

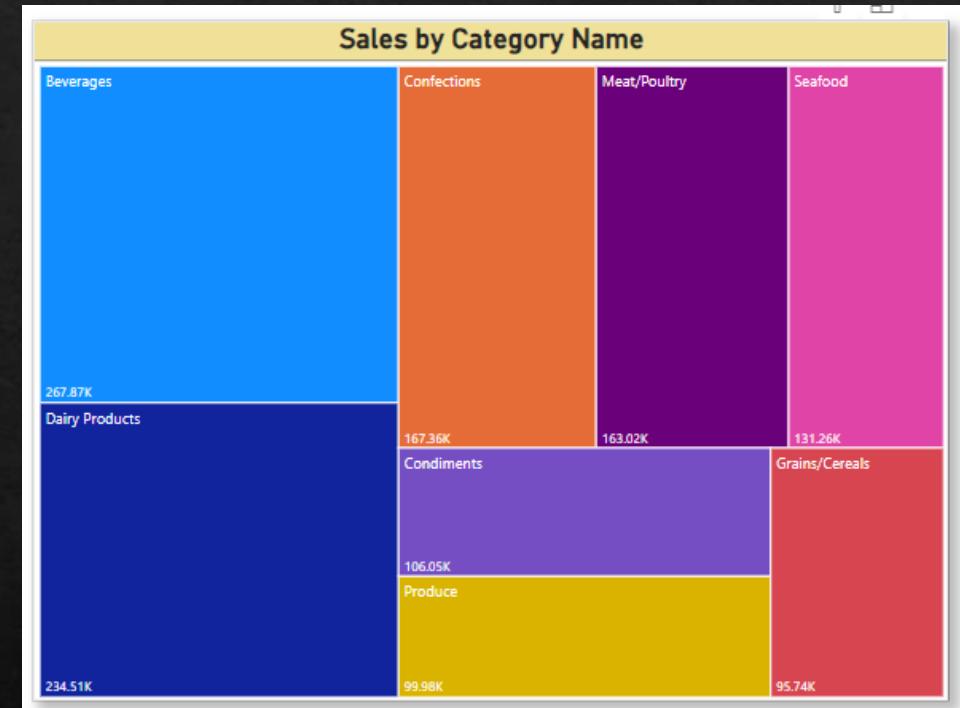


Power Bi Question

11. How does the sales volume vary across different product categories? Can we create a bar chart or treemap to display it ?

Answer:

Yes, we can create treemap to identify sales volume across different categories which help us to focus on that category to gain more profits. According to this chart, Beverages have the maximum sales volume.

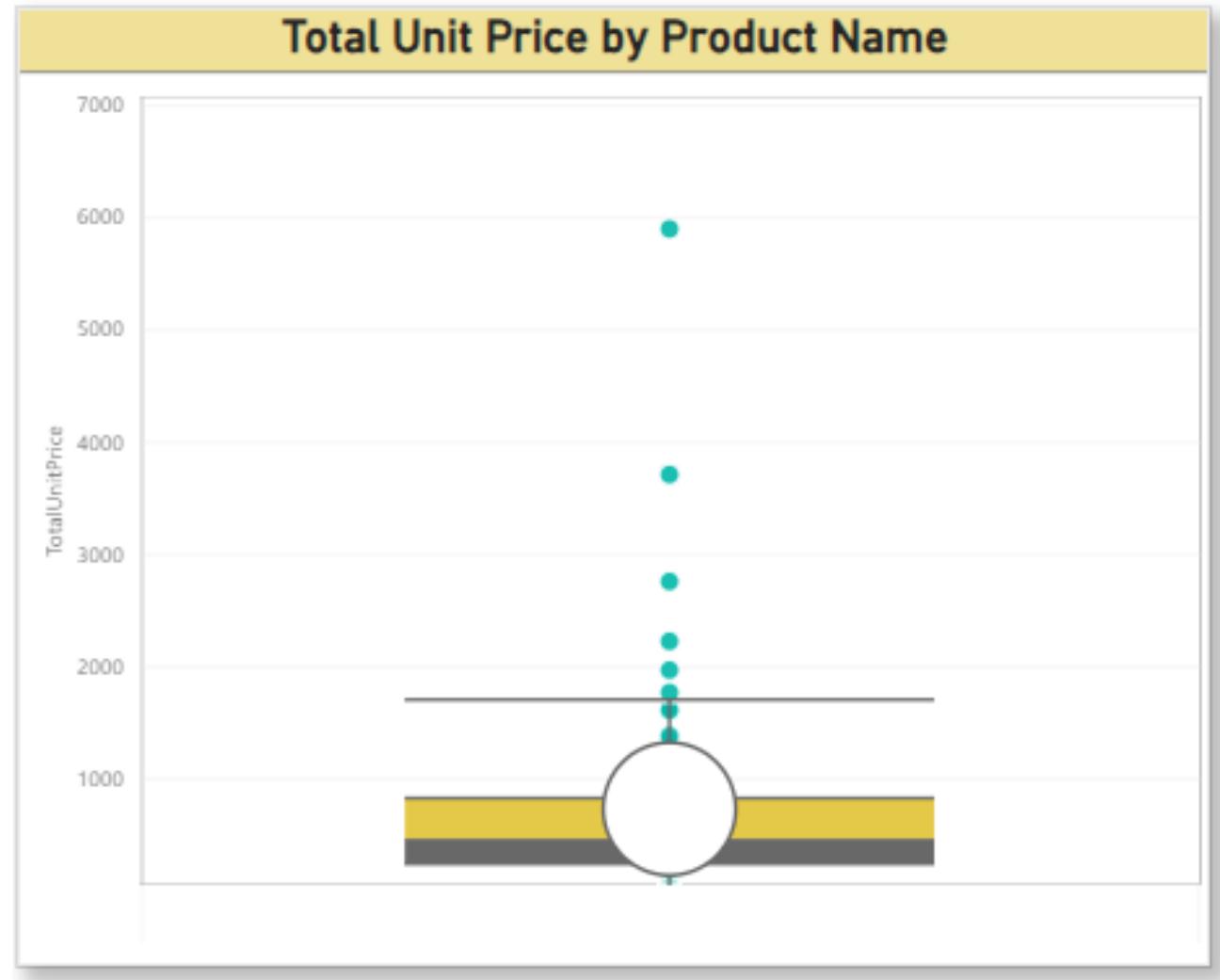


Power Bi Question

12. Can we visualize the pricing distribution of products using a box plot or violin plot ?

Answer:

Yes, we can create box plot chart to measure pricing distribution of product. It help us to identify outlier price. IQR is 584.95 and mean is 733.78.



Power Bi Question

13. What is the distribution of supplier ratings or performance metrics? Can we create a bar chart or radar chart to visualize it

Answer:

Yes, we can create bar chart to identify supplier performance, it shows the total quantity shipped by each supplier company, which will help us to identify most trustworthy supplier.

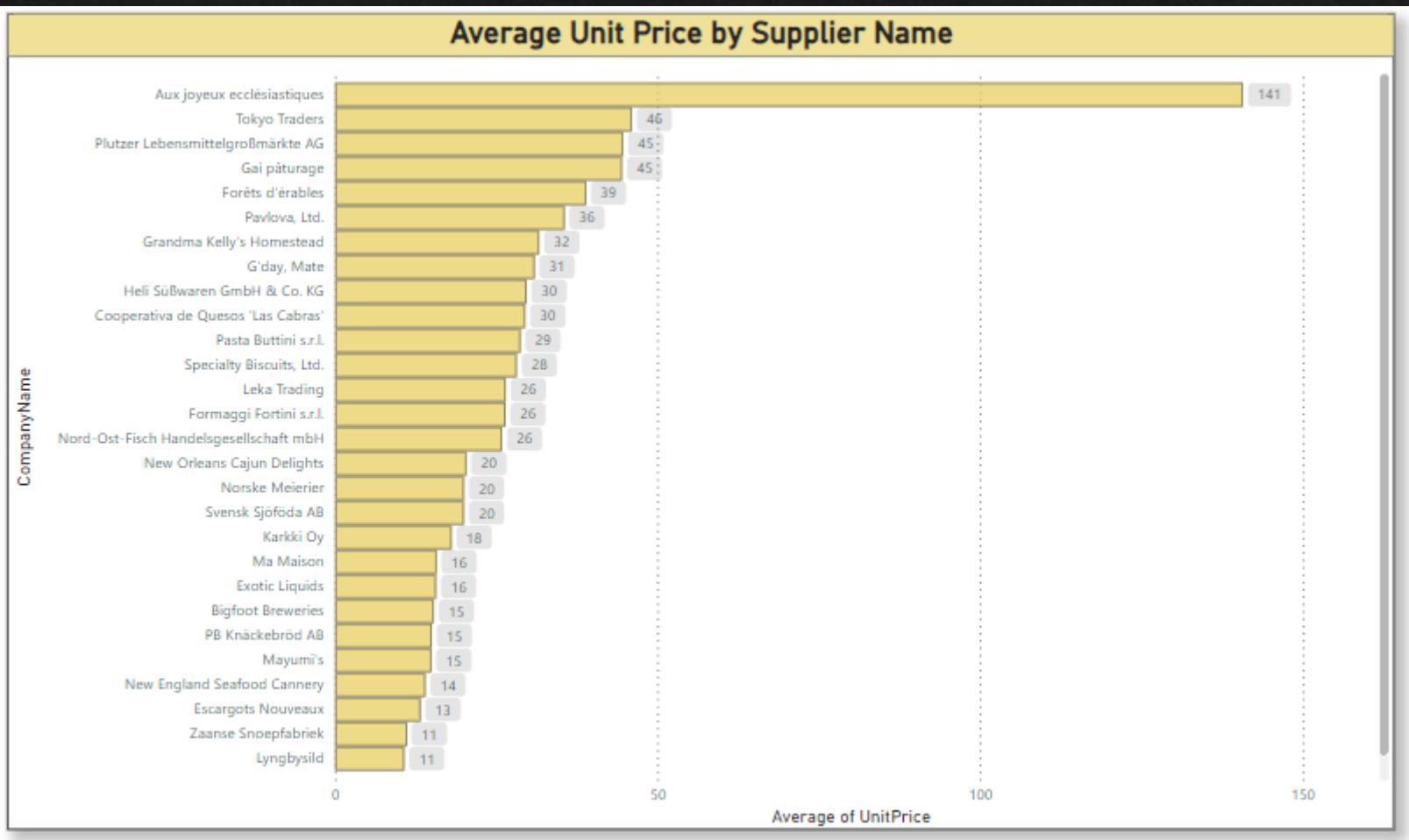


Power Bi Question

14. How does the cost or pricing structure vary across different suppliers? Can we create a box plot or stacked bar chart to display it?

Answer:

Yes, we can create bar chart to identify average pricing structure of each supplier company.



Distribution of Supplier over Country



Power Bi Question

15. Can we visualize the geographical distribution of suppliers using a map or bubble chart?

Answer:

Yes, we can create map to understand suppliers' distribution across different country.

Dashboard Analysis

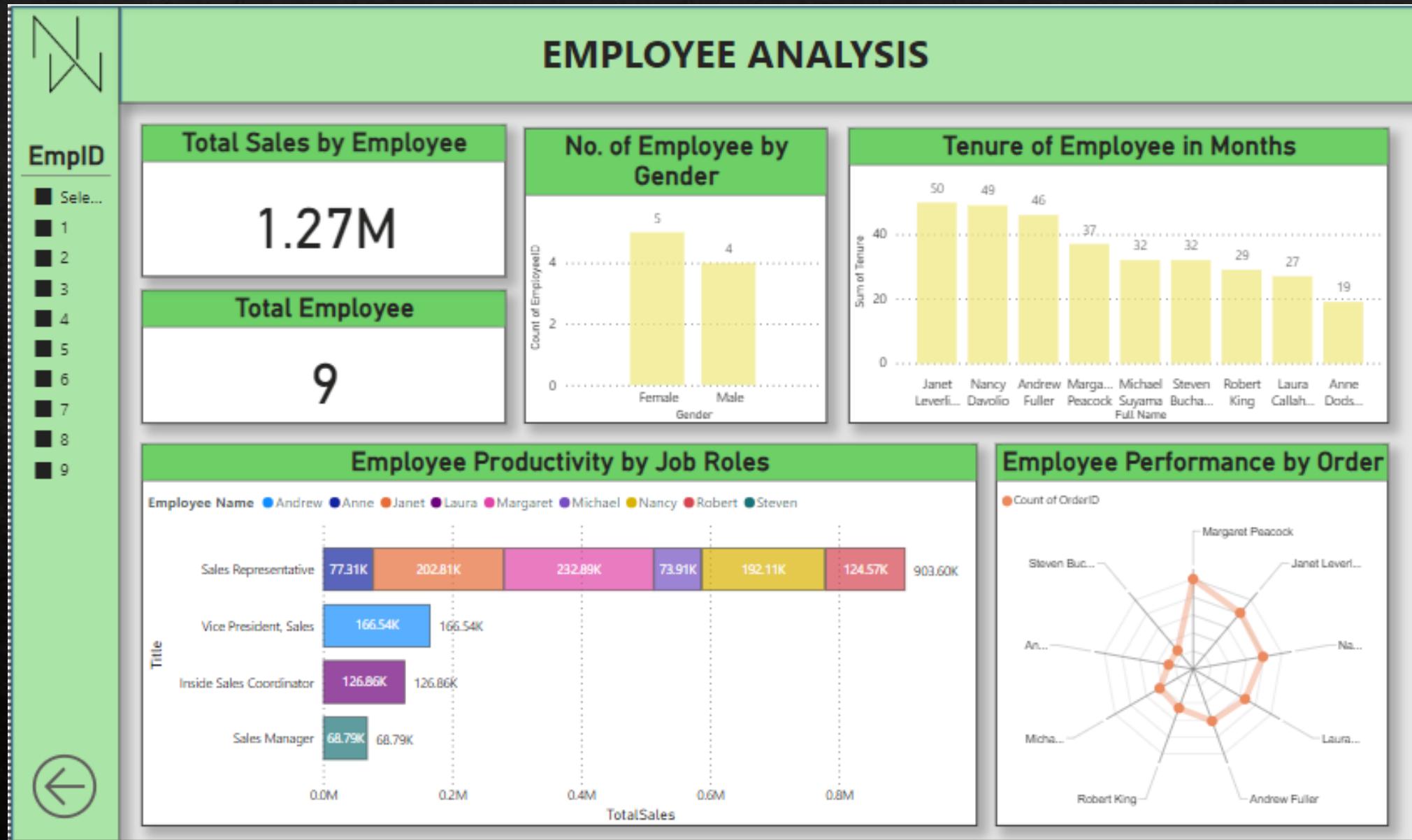
Interactive Features:

The report will feature interactive elements such as dynamic filters, slicers, and drill-down capabilities to enable users to explore the data according to their specific needs. Users can filter data by date range, product category, customer segment, or employee, allowing for deeper insights into various aspects of the company's operations.

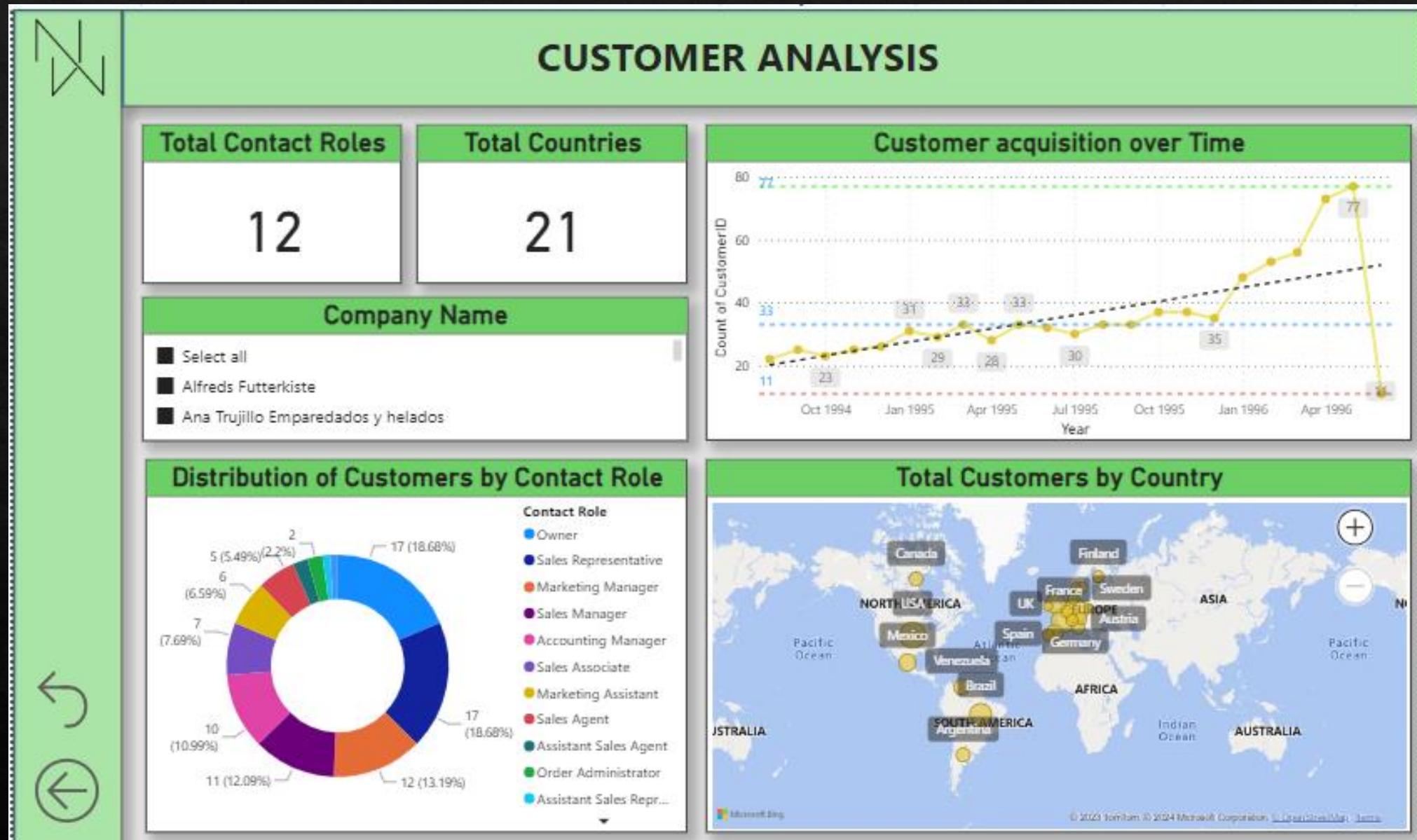
Expected Impact:

The implementation of this Power BI report is expected to revolutionize how Northwind Traders interacts with its data, providing stakeholders with actionable insights that drive business growth and competitiveness in the wholesale market landscape. By leveraging data-driven decision-making processes, the company can optimize sales strategies, improve customer satisfaction, and streamline inventory management, ultimately enhancing overall operational efficiency and profitability.

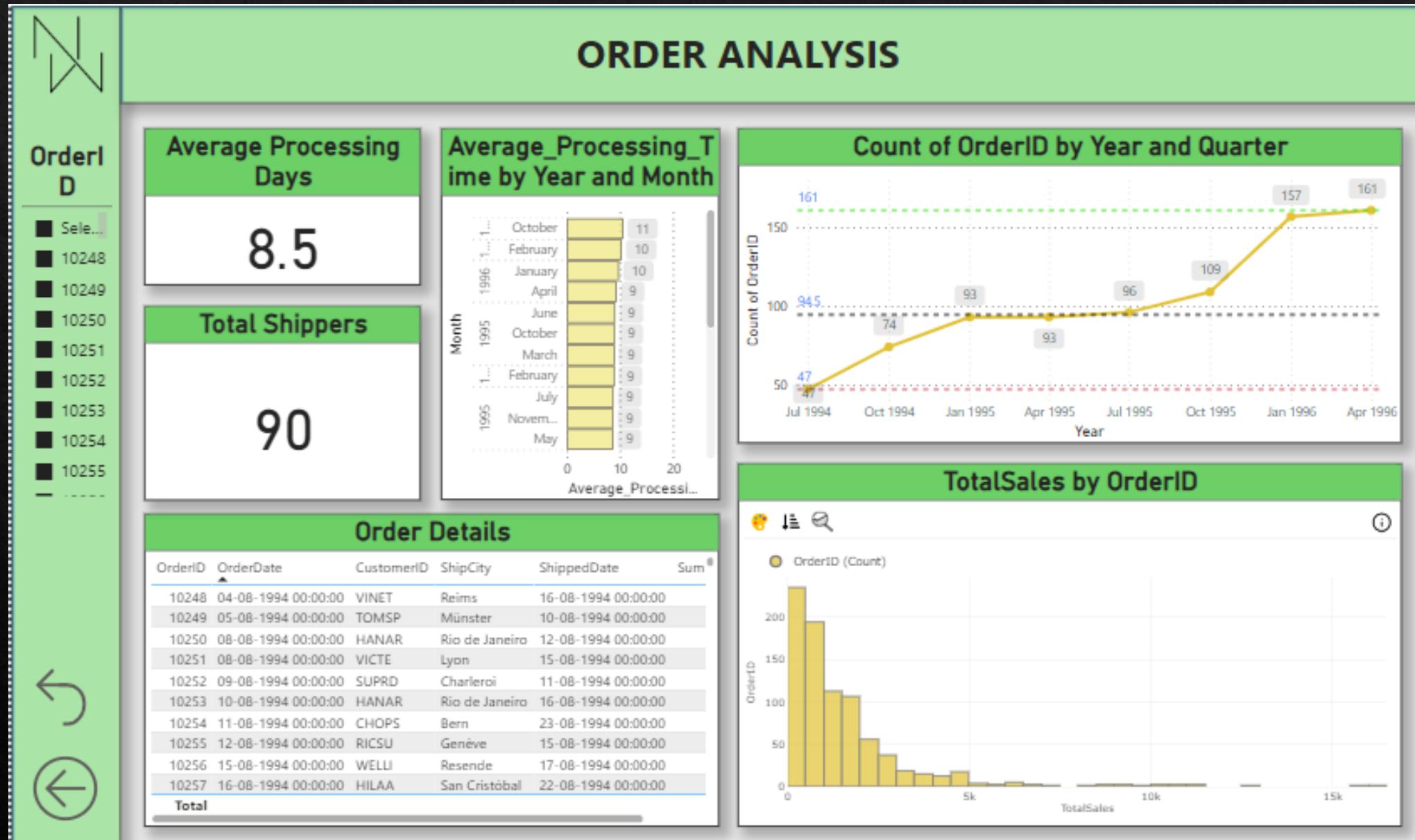
EMPLOYEE DASHBOARD



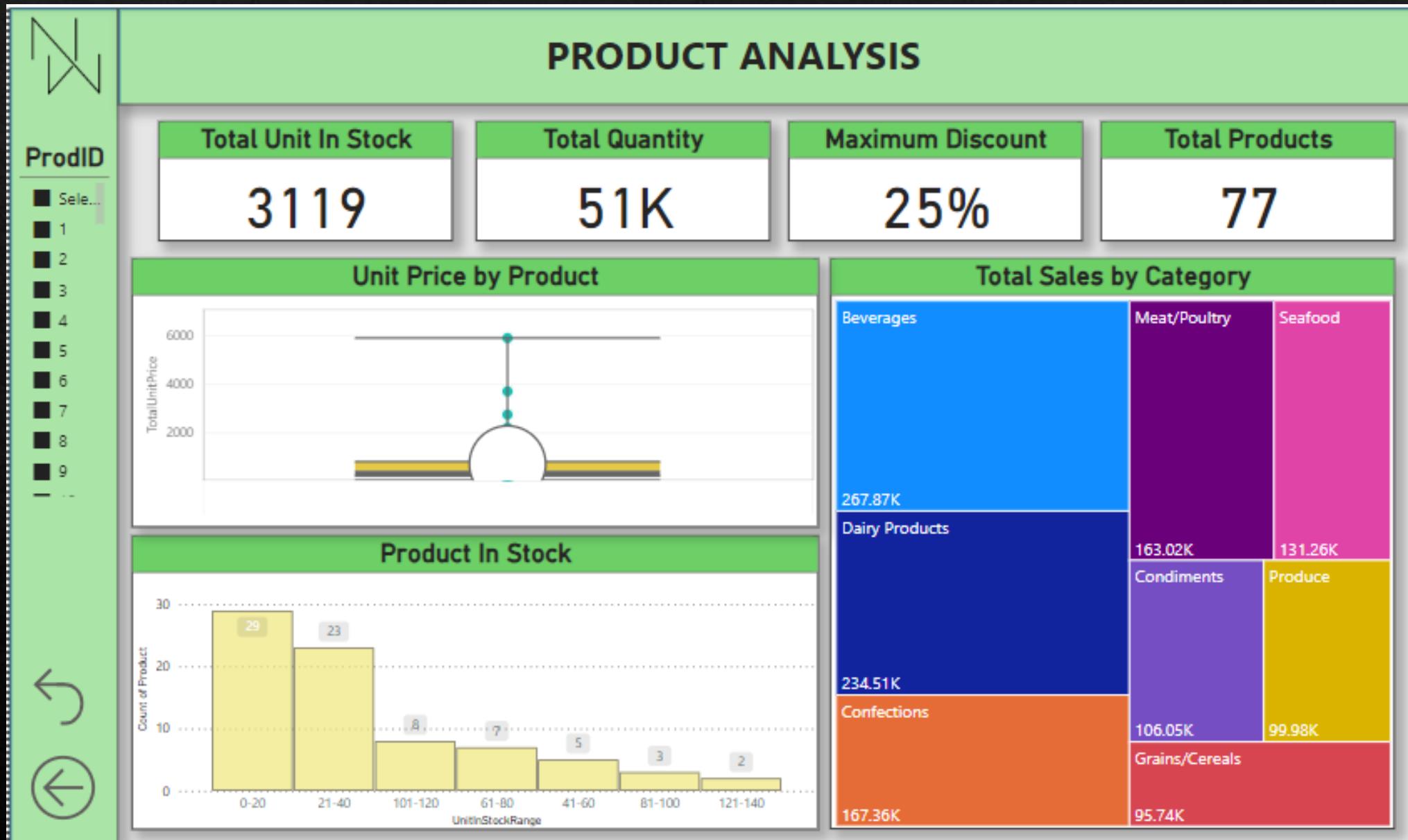
CUSTOMER DASHBOARD



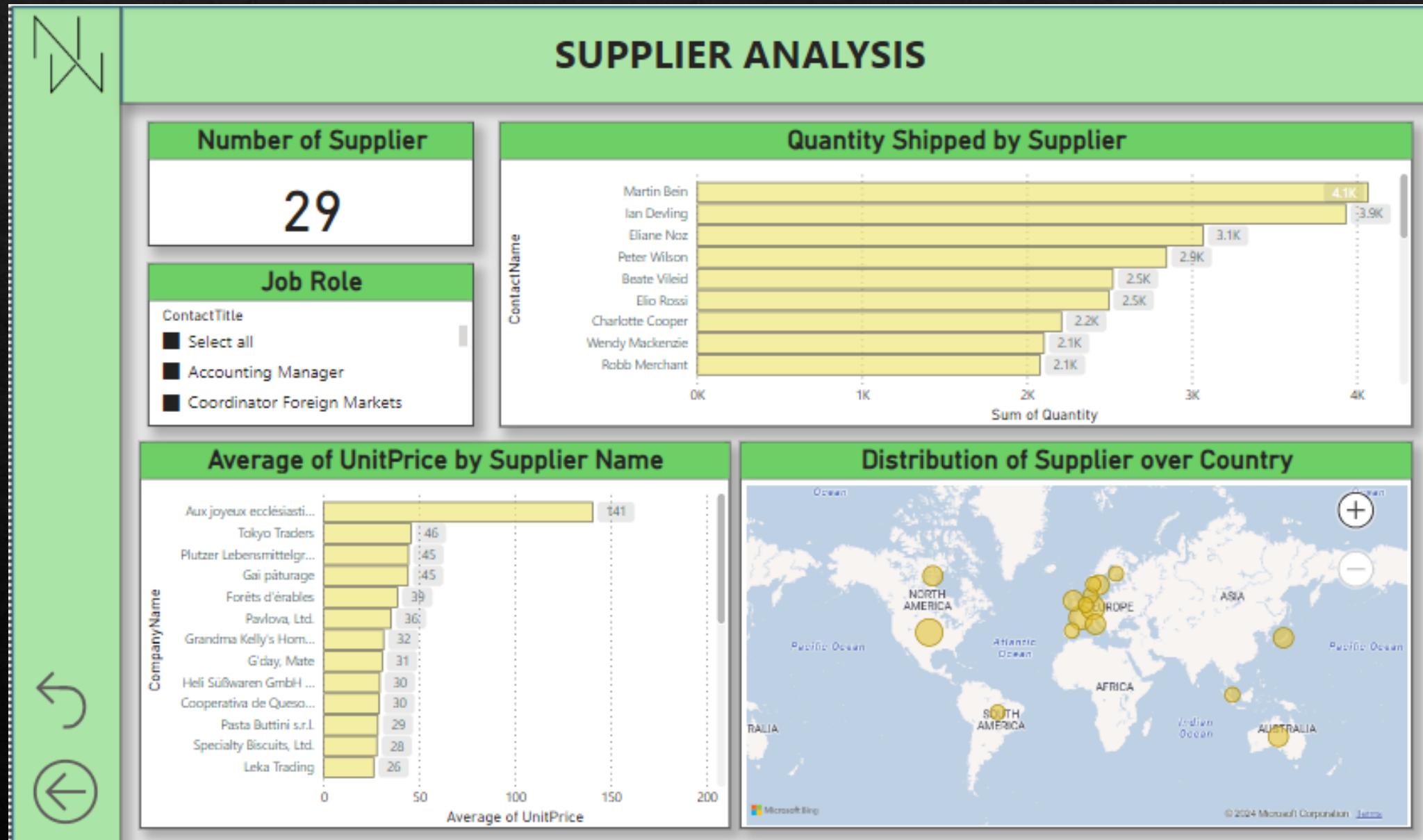
ORDER DASHBOARD



PRODUCT DASHBOARD



SUPPLIER DASHBOARD



EDA Questions

EDA Question

1. What are the key factors influencing customer retention or loyalty based on the dataset ?

```
select
o.CustomerID,c.Country, round(AVG(Monetary),2) as AvgMonetary,
count(o.OrderID) Frequency,
min(datediff((select max(date(OrderDate)) from orders),date(o.OrderDate))) Recency
from orders o
join customers c
on c.CustomerID = o.CustomerID
join northwind.`order details` od
on o.OrderID = od.OrderID
join (
    select od.OrderID,
    round(sum( UnitPrice * Quantity * ( 1 - Discount )),2) as Monetary
    from northwind.`order details` od
    group by 1
) mon on od.OrderID = mon.OrderID
group by 1
order by 1
```

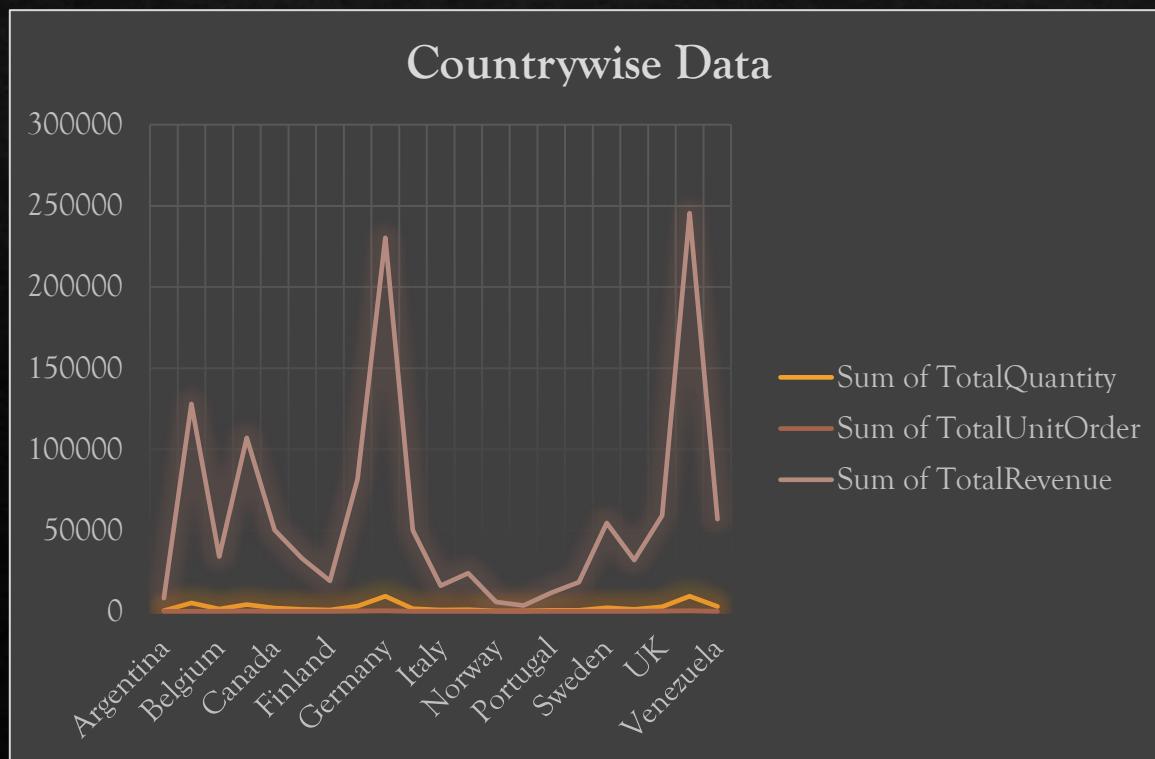
RFM by Customers			
CustomerID	Sum of AvgMonetary	Sum of Frequency	Sum of Recency
ALFKI	706.88	12	27
ANATR	399.45	10	63
ANTON	1175.53	17	98
AROUT	1258.89	30	26
BERGS	1455.98	52	63
BLAUS	540.56	14	7
BLONP	2504.74	26	114
BOLID	1537.28	6	43
BONAP	1465.81	44	0
BOTTM	1731.15	35	12
BSBEV	603.41	22	22
CACTU	335.69	11	8
CENTC	100.8	2	657
CHOPS	1564	22	14
COMMI	1093.95	10	14
CONSH	581.4	7	103
DRACD	738.04	10	2
DUMON	494.77	9	79
EASTC	2049.08	21	8
ERNSH	3908.1	102	1
FAMIA	665.87	19	187
FOLIG	2682.32	16	135
FOUG	1021.11	15	0

Conclusion: - Key factors to check customers loyalty is based on RFM. R stands for Recency, F stands for Frequency, M stands for Monetary. It helps you to understand loyalty of customers. Higher the frequency and lower the Recency gives you the loyal customers.

EDA Question

2. How do customer preferences vary based on their location or demographics? Can we explore this through interactive visualizations?

```
select c.CustomerID, c.Country, SUM(od.Quantity) as TotalQuantity , COUNT(od.OrderID) as TotalUnitOrder,
round(sum( UnitPrice * Quantity * ( 1 - Discount )),2) as TotalRevenue
from Customers c
join Orders o on c.CustomerID = o.CustomerID
join `order details` od on o.OrderID = od.OrderID
GROUP BY 1;
```



Countrywise data value			
Country	Sum of TotalQuantity	Sum of TotalUnitOrder	Sum of TotalRevenue
Argentina	339	34	8119.1
Austria	5167	125	128003.84
Belgium	1392	56	33824.85
Brazil	4247	203	106925.77
Canada	1984	75	50196.29
Denmark	1170	46	32661.02
Finland	885	54	18810.05
France	3254	184	81358.32
Germany	9213	328	230284.62
Ireland	1684	55	49979.9
Italy	822	53	15770.15
Mexico	1025	72	23582.08
Norway	161	16	5735.15
Poland	205	16	3531.95
Portugal	533	30	11472.36
Spain	718	54	17983.2

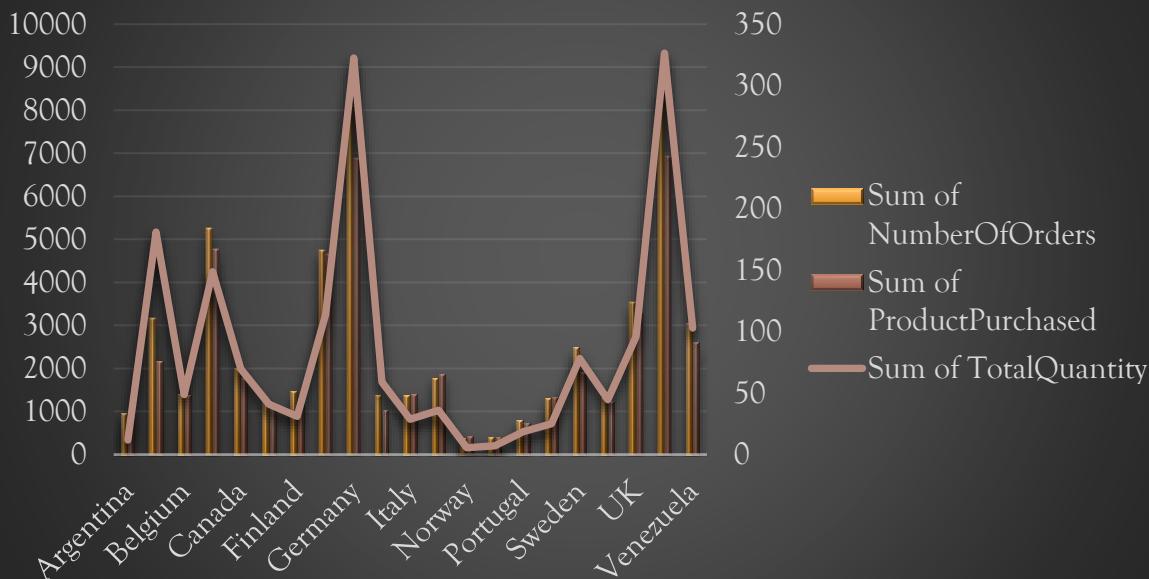
Conclusion: - This query will provide us the Total Quantity, Total Unit in Order, and Total Revenue by customers in each country, which will help us to identify customers sales trend across different countries.

EDA Question

3. Are there any interesting patterns or clusters in customer behavior that can be visualized to identify potential market segments ?

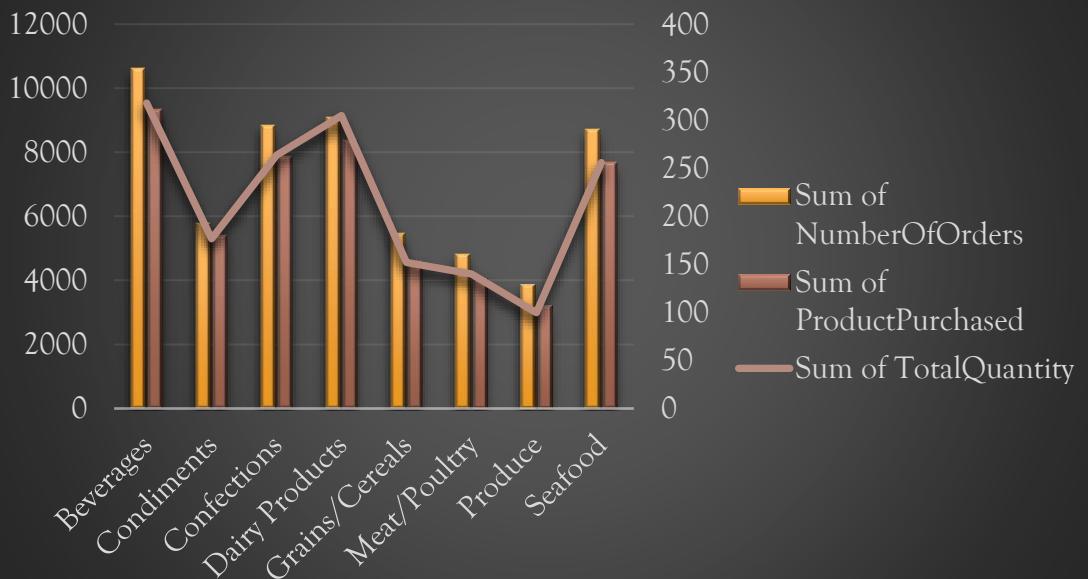
```
select cu.CustomerID, cu.Country, c.CategoryName, count(Distinct p.ProductID) ProductPurchased,  
count(Distinct o.OrderID) NumberOfOrders, sum(od.Quantity) TotalQuantity  
from Customers cu  
join Orders o on cu.CustomerID = o.CustomerID  
join `order details` od on o.OrderID = od.OrderID  
join products p on p.ProductID = od.ProductID  
join categories c on c.CategoryID = p.CategoryID  
Group by 1,2,3
```

Potential Market by Country



Conclusion: - This provide us the market pattern of customer across different country and different categories, it tell us the total quantity purchased by customer in each categories which will provide us the best market segment to focus on.

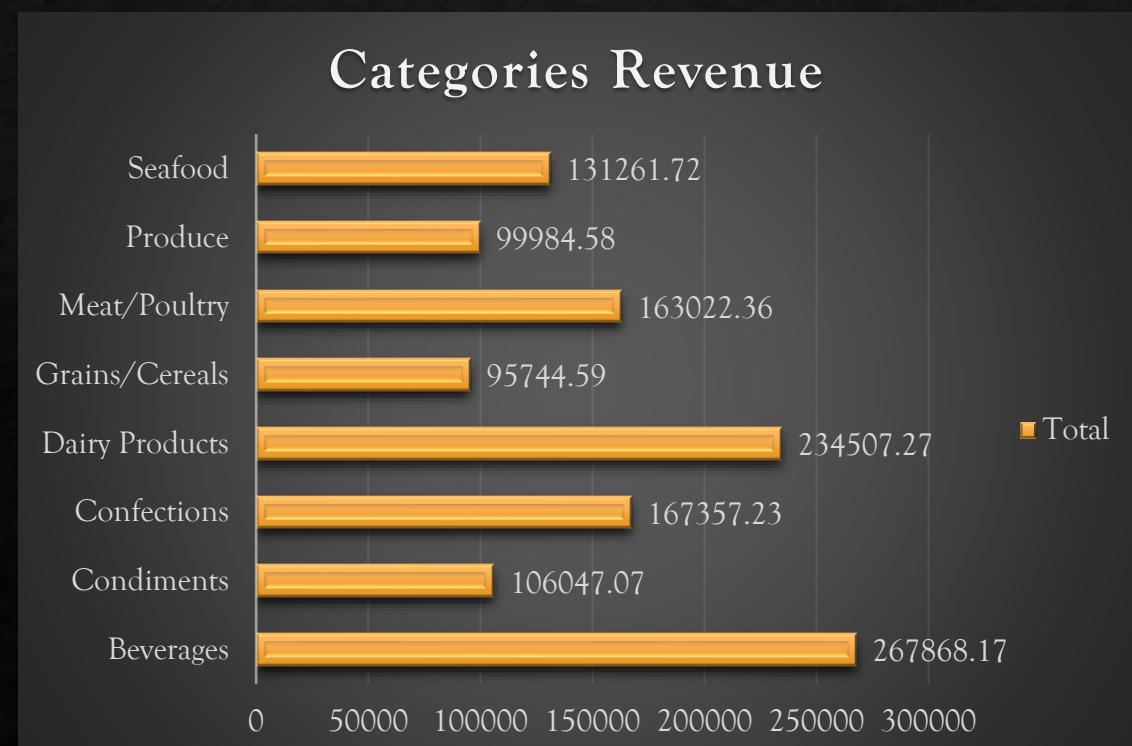
Potential Market by Categories



EDA Question

4. Are there any specific product categories or SKUs that contribute significantly to order revenue? Can we identify them through visualizations?

```
Select c.CategoryName, p.ProductName, round(SUM(od.UnitPrice*od.Quantity*(1-od.Discount)),2) as TotalRevenue  
from products p  
join categories c on p.CategoryID = c.CategoryID  
join `order details` od on od.ProductID = p.ProductID  
Group by 1,2  
order by 3 desc;
```



Category Revenue	
Product Category	Sum of TotalRevenue
Beverages	267868.17
Condiments	106047.07
Confections	167357.23
Dairy Products	234507.27
Grains/Cereals	95744.59
Meat/Poultry	163022.36
Produce	99984.58
Seafood	131261.72
Grand Total	
	1265792.99

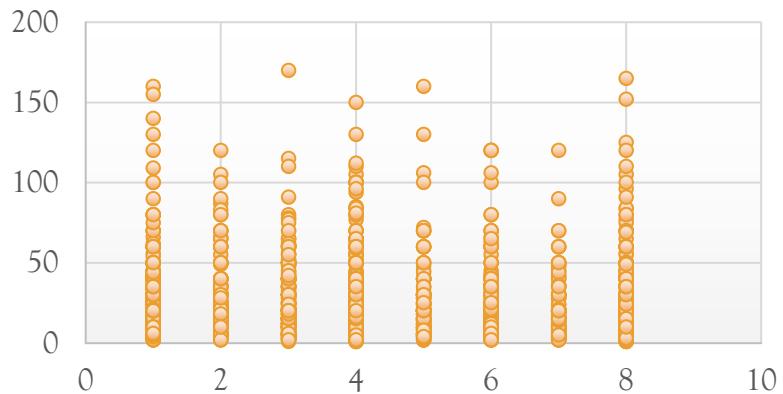
Conclusion: - This query will give order revenue generated by each category. It help use to identify highest order revenue by categories. It show Beverages have highest revenue of \$267868.

EDA Question

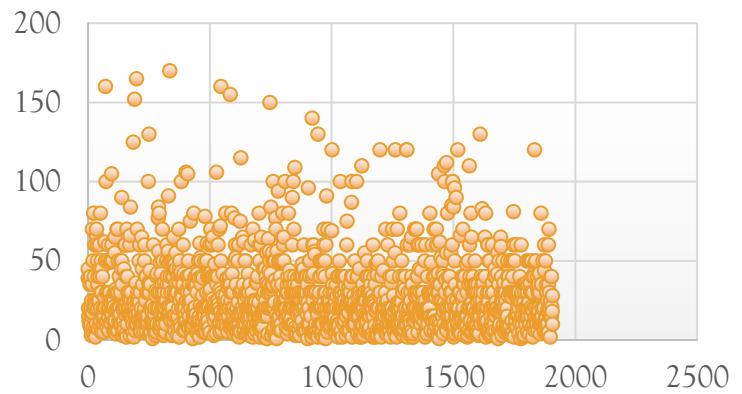
5. Are there any correlations between order size and customer demographics or product categories? Can we explore this visually using scatter plots or heatmaps ?

```
select cu.Country, cu.CustomerID, ca.CategoryID, ca.CategoryName, od.OrderID, SUM(od.Quantity) as TotalOrderSize  
from customers cu  
join orders o  
on cu.CustomerID = o.CustomerID  
join `order details` od  
on od.OrderID = o.OrderID  
join products p  
on od.ProductID = p.ProductID  
join categories ca  
on ca.CategoryID = p.CategoryID  
group by 1,2,3,4,5
```

Correlation Between CategoryID & Order Size



Correlation between CustomerID & Order Size



Conclusion: - This query help us to create correlation between customers and order size. It also create correlation between category and order size. Which display categoryID and amount of order done by it. And also display most ordered Category name.

Category Order Size

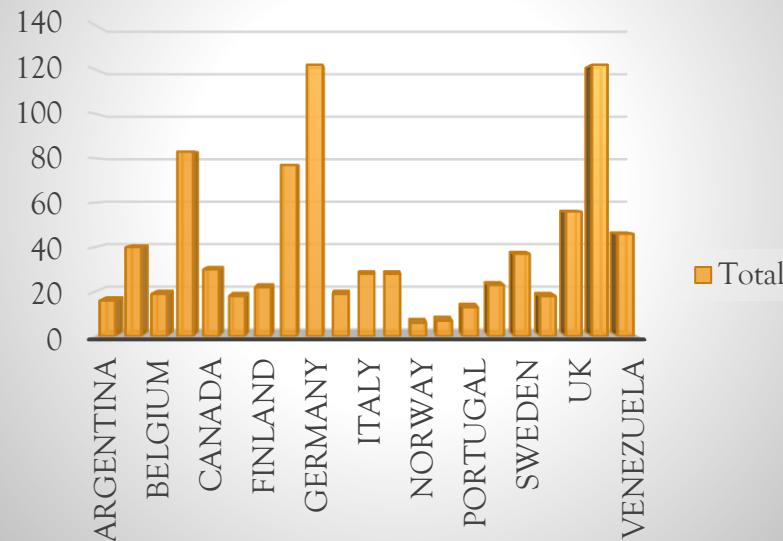
CategoryName	Sum of TotalOrderSize
Beverages	9532
Condiments	5298
Confections	7906
Dairy Products	9149
Grains/Cereals	4562
Meat/Poultry	4199
Produce	2990
Seafood	7681

EDA Question

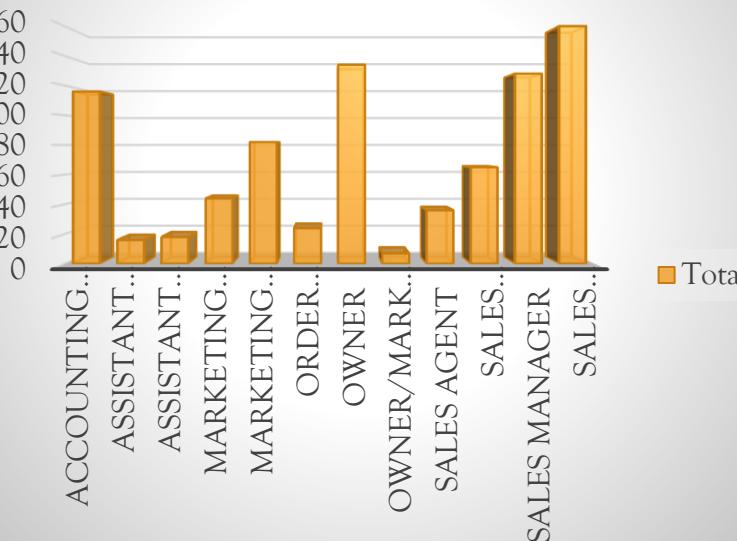
6. How does order frequency vary across different customer segments? Can we visualize this using bar charts or treemaps?

```
select cu.CustomerID, cu.ContactTitle, cu.Country, COUNT(o.OrderID) as OrderFrequency  
from customers cu  
join orders o  
on cu.CustomerID = o.CustomerID  
group by 1,2,3  
order by 4 desc;
```

Order Frequency by Country



Order Frequency by Contact Roles



Conclusion: - This sql command will provide us the order frequency across different customer segments. we have use 2 customer segments

1. Customers Country
 2. Customer Contact Role
- It will show us the order frequency of customers across different country and different contact roles.

EDA Question

7. Are there any correlations between employee satisfaction levels and key performance indicators? Can we explore this visually through scatter plots or line charts ?

Conclusion: - This command will provide us the Employee full name, total customers and total sales based on total sales we have given levels to each employee higher the sales higher the level based on that we have created scatter chart to identify correlation patterns. It shows employee Margaret Peacock have highest Total Sales \$232890 with satisfaction level 9.

```
select e.EmployeeID, concat(e.FirstName, " ", e.LastName) as Full_Name, count(distinct o.CustomerID) as NumOfCustomers,
round(sum( UnitPrice * Quantity * ( 1 - Discount )),2) as TotalSales
from employees e
join orders o
on e.EmployeeID = o.EmployeeID
join `order details` od
on od.OrderID = o.OrderID
group by 1
order by 4 desc;
```



Employee by TotalSales & Satisfaction Level			
Employee Full Name	Sum of NumOfCustomers	Sum of TotalSales	Sum of Satisfaction Level
Andrew Fuller	59	166537.75	6
Anne Dodsworth	29	77308.07	3
Janet Leverling	63	202812.84	8
Laura Callahan	56	126862.28	5
Margaret Peacock	75	232890.85	9
Michael Suyama	43	73913.13	2
Nancy Davolio	65	192107.6	7
Robert King	45	124568.23	4
Steven Buchanan	29	68792.28	1
Grand Total	464	1265793.03	

EDA Question

8. How does employee turnover vary across different departments or job roles? Can we visualize this using bar charts or heatmaps?

```
SELECT
    e.Title as JobTitle,
    COUNT(e.EmployeeID) as NumOfEmployees,
    sum(CASE
        WHEN floor(timestampdiff(year,hiredate,MaxDate))<3
        THEN 1 ELSE 0 END) as NewHires
FROM Employees e
left join (select employeeId,max(orderdate) as MaxDate from orders
group by 1) md on md.EmployeeID = e.EmployeeID
GROUP BY 1;
```

Conclusion: - This sql query will provide the last 3 years employee hire date. It will help us to analyze employee turnover across different roles. It shows Sales Representative have highest employee new hires and total employee number of employees.

Employee Tenure across different job roles



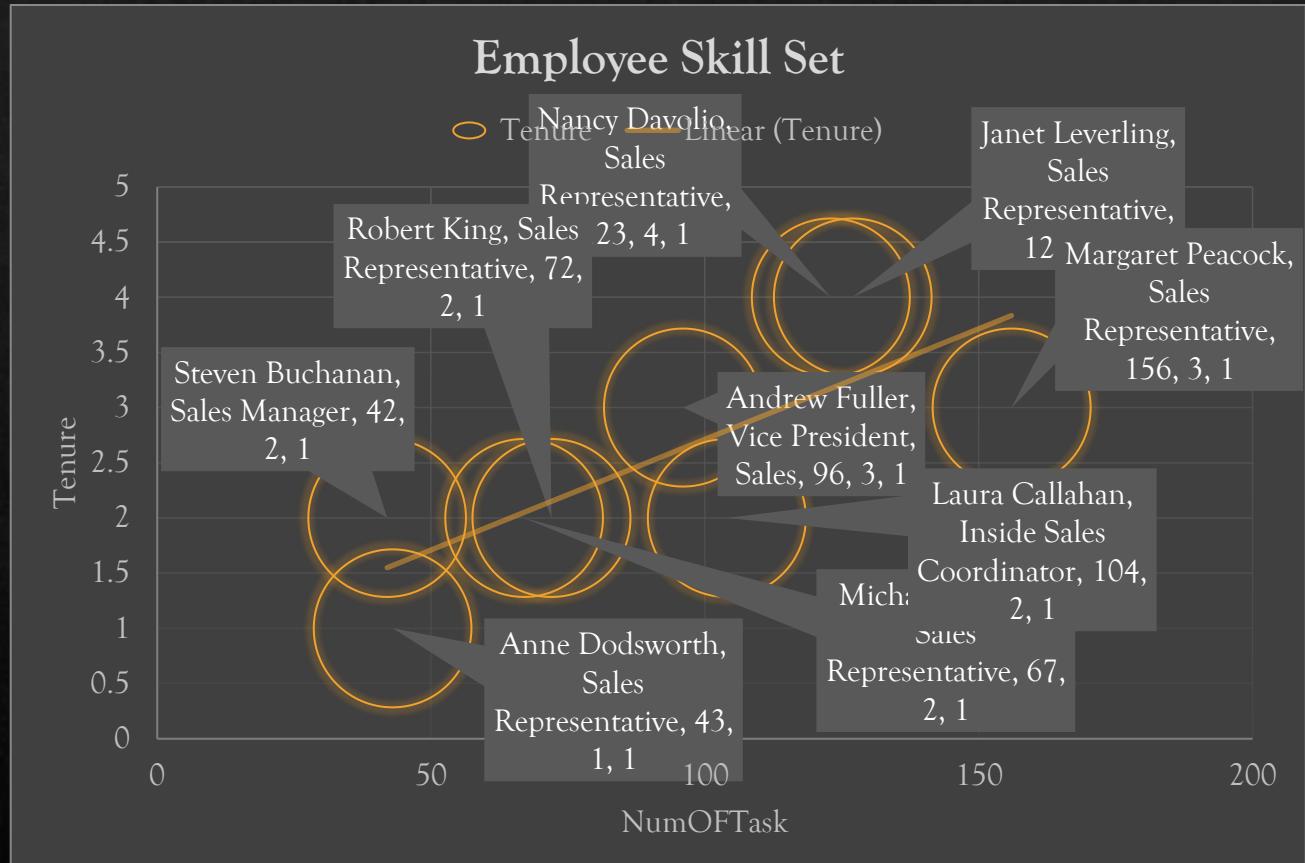
Employee tenure across Job Role		
Job Role	Sum of NumOfEmployees	Sum of NewHires
Inside Sales Coordinator	1	1
Sales Manager	1	1
Sales Representative	6	3
Vice President, Sales	1	0
Grand Total	9	5

EDA Question

9. Can we identify any patterns or clusters in employee skill sets or qualifications through visualizations? How can this information be used for talent management?

```
Select e.EmployeeID, concat(e.FirstName, " ",e.LastName, " ",e.Title) as Full_Name_Title,
count(distinct o.OrderID) as NumOfTasks, timestampdiff(year,hiredate,MaxDate) as Tenure,
round(SUM(od.UnitPrice*od.Quantity*(1-od.Discount)),2) as Total_Sales
from employees e
left join (
    select employeeID, max(orderdate) as MaxDate from orders
    group by 1) md on e.EmployeeID = e.EmployeeID
left join orders o on o.EmployeeID = e.EmployeeID
left join `order details` od on od.OrderID = o.OrderID
group by 1,2,4
order by 5 desc;
```

EmployeeID	Full_Name_Title	NumOf	Tenure	Total_Sales
1	Nancy Davolio, Sales Representative	123	4	1728968.44
2	Andrew Fuller, Vice President, Sales	96	3	1498839.79
3	Janet Leverling, Sales Representative	127	4	1825315.59
4	Margaret Peacock, Sales Representative	156	3	2096017.61
5	Steven Buchanan, Sales Manager	42	2	619130.54
6	Michael Suyama, Sales Representative	67	2	665218.16
7	Robert King, Sales Representative	72	2	1121114.11
8	Laura Callahan, Inside Sales Coordinator	104	2	1141760.5
9	Anne Dodsworth, Sales Representative	43	1	695772.6



Conclusion: - This shows skill set of employees with the help of total tasks and total sales. EmployeeID 4 - Margaret Peacock, Sales Representative have the highest number of task with total sales \$2096017.

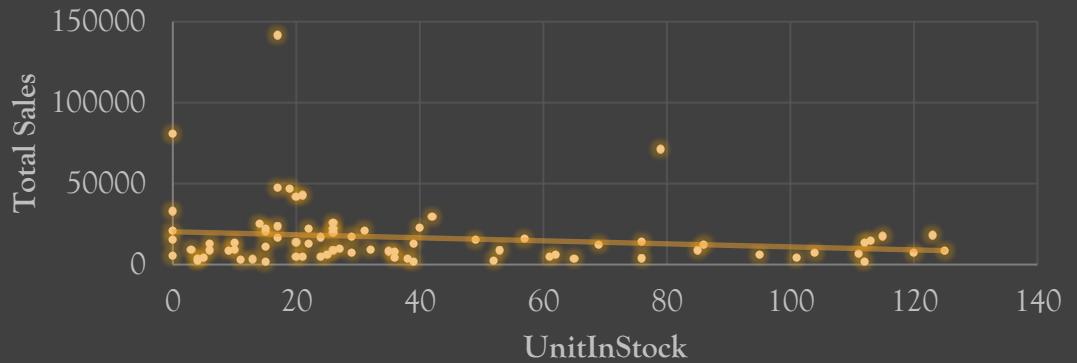
EDA Question

10. Are there any correlations between product attributes (e.g., size, color, features) and sales performance? Can we explore this visually using scatter plots or heatmaps ?

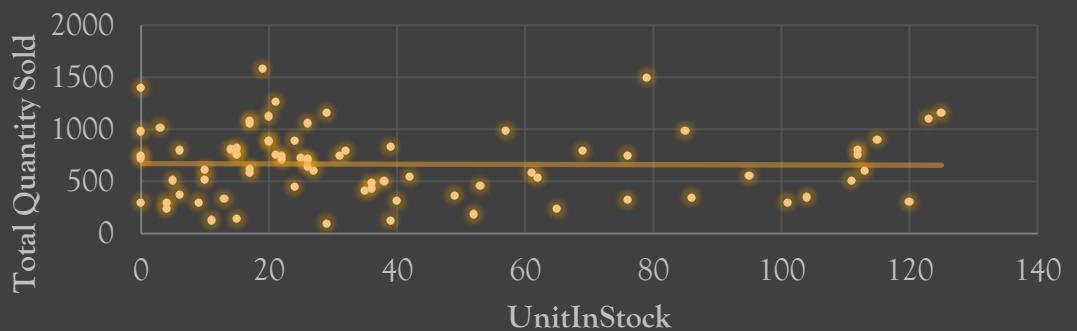
```
select p.ProductID, p.ProductName, p.UnitsInStock, sum(od.Quantity) as TotalQuantitySold,  
round(sum( od.UnitPrice * od.Quantity * ( 1 - od.Discount )),2) as TotalSales  
from products p join `order details` od  
on p.ProductID = od.ProductID  
group by 1  
order by 5 desc;
```

Conclusion: - In this question we have to find any correlation between products and sales performance. It shows weak relation between product unit in stock and total quantity sold this is because the less promote or product is not useful.

Correlation between Product UnitInStock & Total Sales



Correlation between Product UnitInStock & Total Quantity Sold



EDA Question

11. How does product demand fluctuate over different seasons or months? Can we visualize this through line charts or area charts ?

```
select p.ProductID, p.ProductName, month(o.OrderDate) as Months  
from products p  
left join `order details` od  
on p.ProductID = od.ProductID  
left join orders o  
on od.OrderID = o.OrderID
```

Product demand over months	
Months	Count of ProductName
1	195
2	219
3	219
4	250
5	276
6	145
7	76
8	136
9	153
10	152
11	173
12	161
Grand Total	
	2155

Product Name Filter:

- Alice Mutton
- Aniseed Syrup
- Boston Crab Meat
- Camembert Pierrot
- Carnarvon Tigers
- Chai
- Chang
- Chartreuse verte



Conclusion: - This query will provide the product demands in each months. It shows 5th month is having highest product demand with 276 product count.

EDA Question

12. Can we identify any outliers or anomalies in product performance or sales using visualizations? How can this information be used for product optimization ?

```
SELECT p.ProductName, SUM(od.Quantity) as TotalQuantity, round(sum( od.UnitPrice * od.Quantity * ( 1 - od.Discount )),2) as TotalSales  
FROM Products p  
JOIN `order details` od ON p.ProductID = od.ProductID  
JOIN Orders o ON od.OrderID = o.OrderID  
GROUP BY p.ProductName  
ORDER BY p.ProductName;
```



Conclusion: - This query will provide the outliers in Total sales and Total quantity in product. Which will help us to analyze the product performance ranges based on total sales and total quantity.

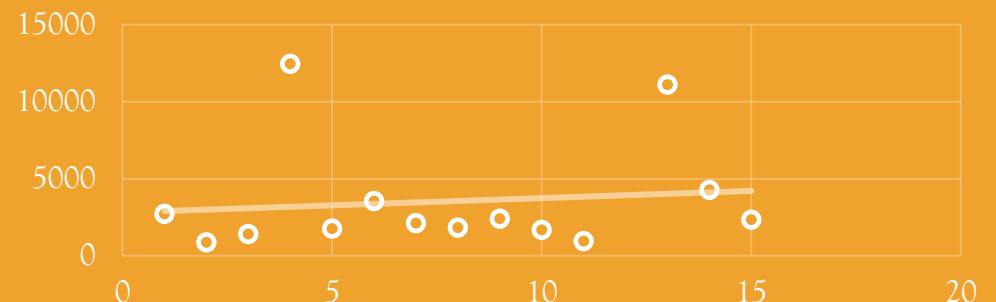
EDA Question

13. Are there any correlations between supplier attributes (e.g., location, size, industry) and performance metrics (e.g., on-time delivery, product quality) ? Can we explore this visually through scatter plots or heatmaps ?

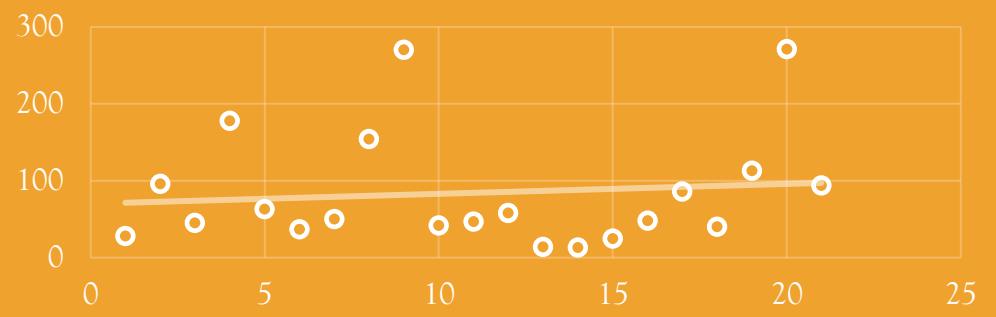
```
select distinct
case when ContactTitle="Accounting Manager" then 1 when ContactTitle='Coordinator Foreign Markets' then 2
when ContactTitle='Export Administrator' then 3 when ContactTitle='International Marketing Mgr.' then 4
when ContactTitle='Marketing Manager' then 4 when ContactTitle='Marketing Representative' then 5
when ContactTitle='Order Administrator' then 6 when ContactTitle='Owner' then 7
when ContactTitle='Product Manager' then 8 when ContactTitle='Purchasing Manager' then 9
when ContactTitle='Regional Account Rep.' then 10 when ContactTitle='Sales Agent' then 11
when ContactTitle='Purchasing Manager' then 12 when ContactTitle='Sales Representative' then 13
when ContactTitle='Sales Manager' then 14 when ContactTitle='Wholesale Account Agent' then 15 end as ContactCode,
case when o.ShipCountry='Argentina' then 1 when o.ShipCountry='Austria' then 2 when o.ShipCountry="Belgium" then 3
when o.ShipCountry="Brazil" then 4 when o.ShipCountry="Canada" then 5 when o.ShipCountry="Denmark" then 6
when o.ShipCountry="Finland" then 7 when o.ShipCountry="France" then 8 when o.ShipCountry="Germany" then 9
when o.ShipCountry="Ireland" then 10 when o.ShipCountry="Italy" then 11 when o.ShipCountry="Mexico" then 12
when o.ShipCountry="Norway" then 13 when o.ShipCountry="Poland" then 14 when o.ShipCountry="Portugal" then 15
when o.ShipCountry="Spain" then 16 when o.ShipCountry="Sweden" then 17 when o.ShipCountry="Switzerland" then 18
when o.ShipCountry="UK" then 19 when o.ShipCountry="USA" then 20 when o.ShipCountry="Venezuela" then 21 end as countrycode,
os.OrderID,Delivery_Time from suppliers s
left join products as p on p.SupplierID=s.SupplierID
left join `order details` od on od.ProductID=p.ProductID
left join (select OrderID, avg(timestampdiff(day,OrderDate,Requireddate)) as Delivery_Time from orders
where orderDate is not null and ShippedDate is not null
group by 1) os on od.OrderID=os.OrderID
join orders o on o.OrderID=od.OrderID
where Delivery_Time is not null;
```

Conclusion: - This sql command will show weak correlation between supplier attributes and its performance metrics. We have made scatter chart to identify supplier contact title and delivery time.

CONTACT CODE & TOTAL DELIVERY TIME



COUNTRY CODE & NUMOFORDERS



EDA Question

14. How does supplier performance vary across different product categories or departments ? Can we visualize this using stacked bar charts or grouped column charts ?

```
SELECT
distinct s.supplierID,s.CompanyName, s.ContactName, s.ContactTitle, c.CategoryName, SUM(p.UnitsOnOrder) AS TotalQuantity,
round(AVG(timestampdiff(day, OrderDate, Requireddate)),2) AS Delivery_time FROM suppliers s
LEFT JOIN products AS p ON p.SupplierID = s.SupplierID
LEFT JOIN `order details` od ON od.ProductID = p.ProductID
LEFT JOIN orders o ON o.OrderID = od.OrderID
LEFT JOIN categories c ON c.CategoryID = p.CategoryID
GROUP BY s.supplierID, s.ContactName, c.categoryID, c.CategoryName
HAVING Delivery_time IS NOT NULL;
```



Conclusion: - This sql command will help us to identify supplier performance across different categories.

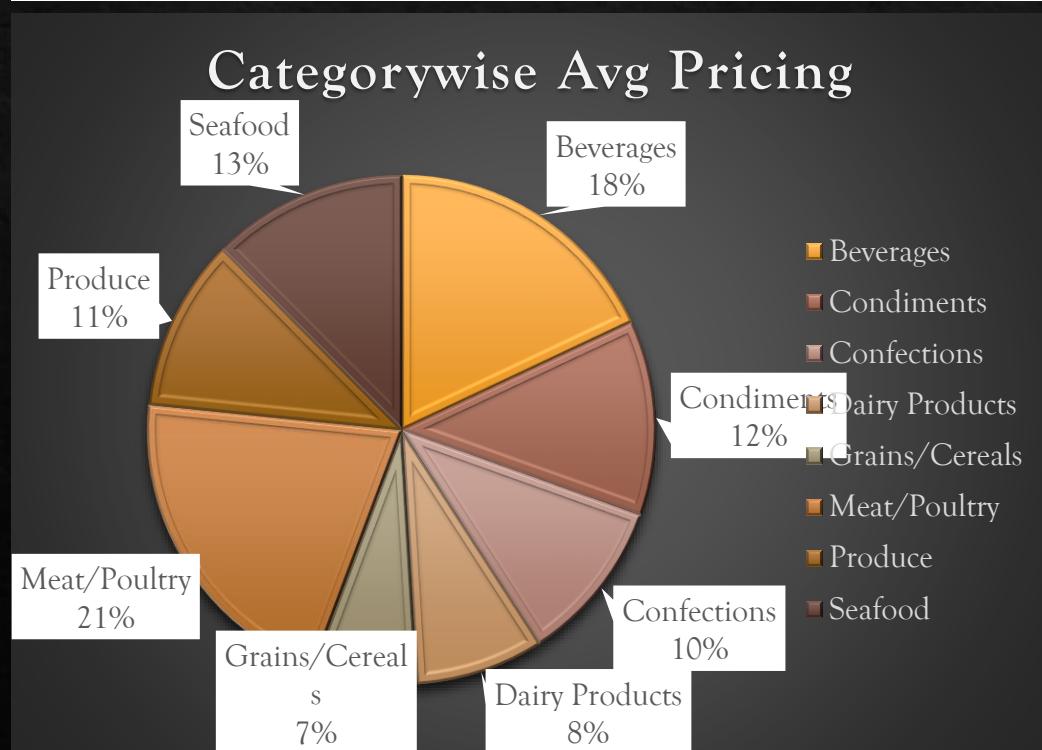
we have made combo chart to display supplier Ids and their average time and total quantity supplied by them in different product categories.

CategoryName	⋮	✖
Beverages		
Condiments		
Confections		
Dairy Products		
Grains/Cereals		
Meat/Poultry		
Produce		
Seafood		

EDA Question

15. Can we identify any trends or patterns in supplier costs or pricing structures through visualizations? How can this information be used for procurement optimization ?

```
select s.CompanyName as SupplierName, p.CategoryID, c.CategoryName, AVG(p.UnitPrice) as AvgPrice  
from products p  
join suppliers s  
on p.supplierID = s.SupplierID  
join categories c  
on c.CategoryID = p.CategoryID  
group by 1,2;
```



SupplierName	Supplier Avg Unit Price
Category Name	Sum of AvgPrice
Aux joyeux ecclésiastiques	265.8333333
Bigfoot Breweries	183.2
Cooperativa de Quesos 'L...	153.885
Escargots Nouveaux	120.4333333
Exotic Liquids	98
Forêts d'érables	308.315
Formaggi Fortini s.r.l.	161.85
Gai pâturage	183.415
Grand Total	1474.931667

Conclusion: - This command will provide us the average pricing of each suppliers in different categories. Which help us to compare pricing between different suppliers in different categories.

Conclusion

In conclusion, leveraging the data available in the Northwind database through Power BI can provide valuable insights into customer behavior, sales patterns, employee performance, inventory management, supplier relationships, shipping efficiency, and overall business performance. These insights can empower Northwind Traders to make data-driven decisions that drive business growth and competitiveness in the wholesale market landscape.