CAPSTONE PROJECT NORTH WIND TRADERS

BY SAHIL BHARDWAJ

TOPICS

- Project Overview

- Information about Dataset tables

- Data cleaning and MECE-Breakdown

- Solving Power Bi and EDA Questions

- Dashboard Analysis

- Conclusion

Project overview

The North wind database contains the sales data for a fictitious company called "North wind traders" which imports and exports specialty foods from around the world. The report aims to generate insights into customer behavior, sales patterns, and employee performance to aid decision-making processes. It will cover sales analysis, customer segmentation, inventory trends, and employee performance, consolidating data from multiple tables for a comprehensive view of the company's operations. The objective of this report to empower stakeholders to make data-driven decisions by offering valuable insights and facilitating data exploration through interactive visualizations and dynamic filters.
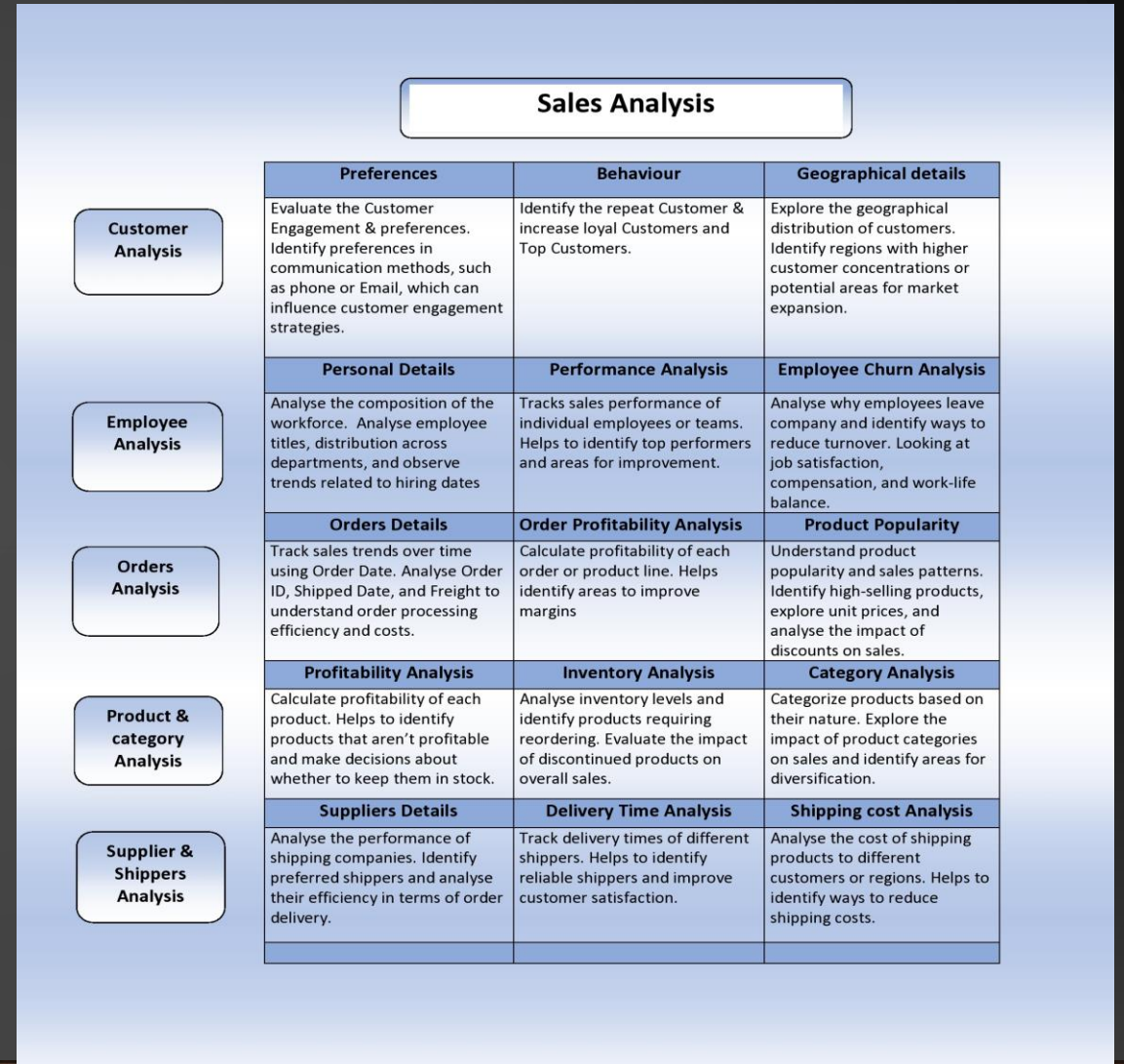
Information about dataset table

➤ **Customers**: This table contains information about customers, including their company details, contact information, and location.

➤ **Employees:** This table stores employee details like name, title, contact information, and hire date. Orders: this table captures order details such as order ID, customer ID, date, items ordered, shipping information, and freight costs.

➤ **Order details:** This table provides a breakdown of each order, including specific products ordered, their quantity, and unit price.

➤ Products: This table showcases product information like name, category, unit price, stock levels, and supplier details. Suppliers: this table contains information about the company's suppliers, including their contact details and location.

➤ Suppliers: This table contains information about the company's suppliers, including their contact details and location.

➤ Shippers: This table lists the shipping companies utilized by north wind traders, along with their contact information.

➤ Categories: This table categorizes the products offered by north wind traders.

Data cleaning and MECE-breakdown

Data cleaning:

The Table contains many null value almost in each attributes ,Replaced each with N/A and remove those column that are not needed for visualization such as Fax ,Phone Number , Image , Image thumbnail etc. After transformation I have created a MECE-breakdown for better understanding of dataset.

Sales Analysis

| Analysis			
Customer Analysis	**Preferences**	**Behaviour**	**Geographical details**
	Evaluate the Customer Engagement & preferences. Identify preferences in communication methods, such as phone or Email, which can influence customer engagement strategies.	Identify the repeat Customer & increase loyal Customers and Top Customers.	Explore the geographical distribution of customers. Identify regions with higher customer concentrations or potential areas for market expansion.
Employee Analysis	**Personal Details**	**Performance Analysis**	**Employee Churn Analysis**
	Analyse the composition of the workforce. Analyse employee titles, distribution across departments, and observe trends related to hiring dates	Tracks sales performance of individual employees or teams. Helps to identify top performers and areas for improvement.	Analyse why employees leave company and identify ways to reduce turnover. Looking at job satisfaction, compensation, and work-life balance.
Orders Analysis	**Orders Details**	**Order Profitability Analysis**	**Product Popularity**
	Track sales trends over time using Order Date. Analyse Order ID, Shipped Date, and Freight to understand order processing efficiency and costs.	Calculate profitability of each order or product line. Helps identify areas to improve margins	Understand product popularity and sales patterns. Identify high-selling products, explore unit prices, and analyse the impact of discounts on sales.
Product & category Analysis	**Profitability Analysis**	**Inventory Analysis**	**Category Analysis**
	Calculate profitability of each product. Helps to identify products that aren't profitable and make decisions about whether to keep them in stock.	Analyse inventory levels and identify products requiring reordering. Evaluate the impact of discontinued products on overall sales.	Categorize products based on their nature. Explore the impact of product categories on sales and identify areas for diversification.
Supplier & Shippers Analysis	**Suppliers Details**	**Delivery Time Analysis**	**Shipping cost Analysis**
	Analyse the performance of shipping companies. Identify preferred shippers and analyse their efficiency in terms of order delivery.	Track delivery times of different shippers. Helps to identify reliable shippers and improve customer satisfaction.	Analyse the cost of shipping products to different customers or regions. Helps to identify ways to reduce shipping costs.

Power BI Questions

1. HOW DOES CUSTOMER DISTRIBUTION VARY ACROSS DIFFERENT REGIONS OR CUSTOMER SEGMENTS? CAN WE VISUALIZE IT ON A MAP OR BAR CHART?

2. WHAT IS THE TREND IN CUSTOMER ACQUISITION OVER TIME? CAN WE CREATE A LINE CHART OR AREA CHART TO DISPLAY IT?

3. CAN WE VISUALIZE THE DISTRIBUTION OF CUSTOMER DEMOGRAPHICS SUCH AS AGE, GENDER, OR INCOME USING HISTOGRAMS OR PIE CHARTS?

4. HOW DOES ORDER VOLUME CHANGE OVER TIME? CAN WE CREATE A TIME SERIES CHART OR STACKED BAR CHART TO VISUALIZE IT?

5. WHAT IS THE DISTRIBUTION OF ORDER VALUES? CAN WE CREATE A HISTOGRAM OR BOX PLOT TO DISPLAY IT?

6. CAN WE VISUALIZE THE AVERAGE ORDER PROCESSING TIME OR SHIPPING DURATION USING A BAR CHART OR BOX PLOT?

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?

8. WHAT IS THE DISTRIBUTION OF EMPLOYEE TENURE? CAN WE CREATE A HISTOGRAM OR BOX PLOT TO DISPLAY IT?

9. CAN WE VISUALIZE EMPLOYEE PERFORMANCE RATINGS OR KPIS USING A RADAR CHART OR BULLET GRAPH?

10. WHAT IS THE DISTRIBUTION OF PRODUCT RATINGS OR REVIEWS? CAN WE CREATE A HISTOGRAM OR STACKED BAR CHART TO VISUALIZE IT?

11. HOW DOES THE SALES VOLUME VARY ACROSS DIFFERENT PRODUCT CATEGORIES? CAN WE CREATE A BAR CHART OR TREE MAP TO DISPLAY IT?

12. CAN WE VISUALIZE THE PRICING DISTRIBUTION OF PRODUCTS USING A BOX PLOT OR VIOLIN PLOT?

13. WHAT IS THE DISTRIBUTION OF SUPPLIER RATINGS OR PERFORMANCE METRICS? CAN WE CREATE A BAR CHART OR RADAR CHART TO VISUALIZE IT?

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?

15. CAN WE VISUALIZE THE GEOGRAPHICAL DISTRIBUTION OF SUPPLIERS USING A MAP OR BUBBLE CHART?

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

Yes , we can create a map for visualization of customer distribution across different countries . and also we can create a bar chart for it.

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

The customer acquisition trend is increasing over time, but with some fluctuations. There seems to be a steeper increase from October 1994 to April 1995, followed by a period of slower growth until June 1996.
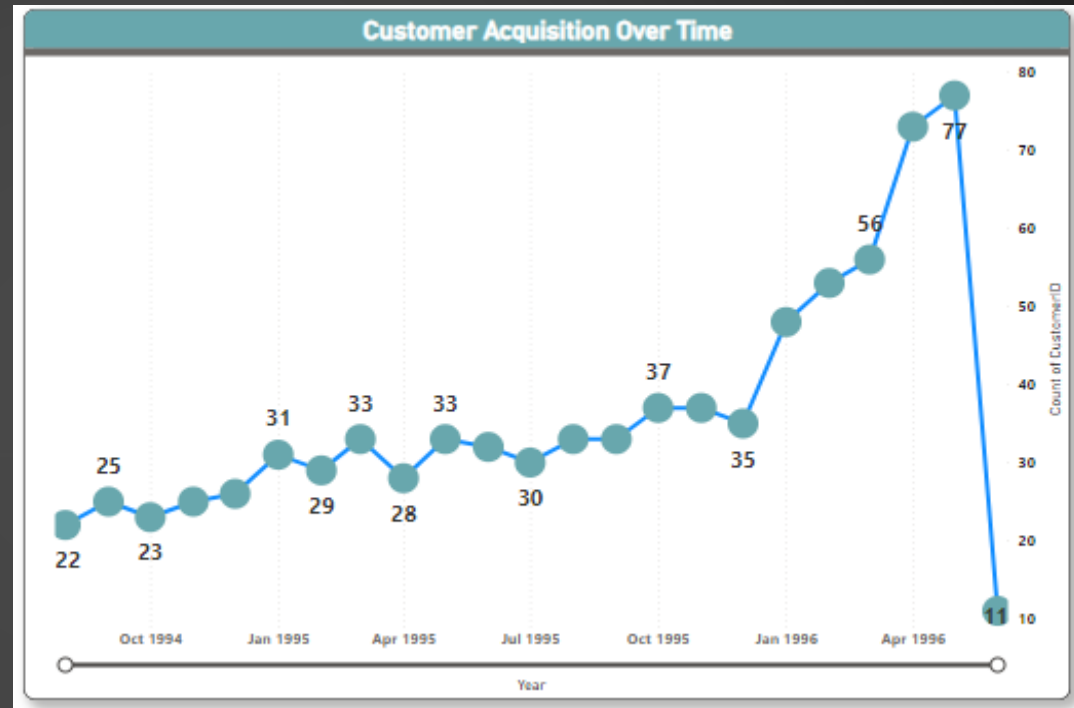
3.Can we visualize the distribution of customer demographics such as age, gender, or income using histograms or pie charts?

➢ **Distribution of Customers by Country:**

A majority of customers around 60% are from Germany and the United States. The remaining customers are spread across various countries, with no other country exceeding 10%.

➢ **Customer Distribution by Contact Role:**

•The largest group of customers around 40% have the contact role of Sales Representative.

•Other significant groups include Sales Manager around 14% and Accounting Manager around 10%.

•Smaller groups include Sales Associate, Marketing Manager, Order Administrator, Assistant Sales, and Owner/Marketing.
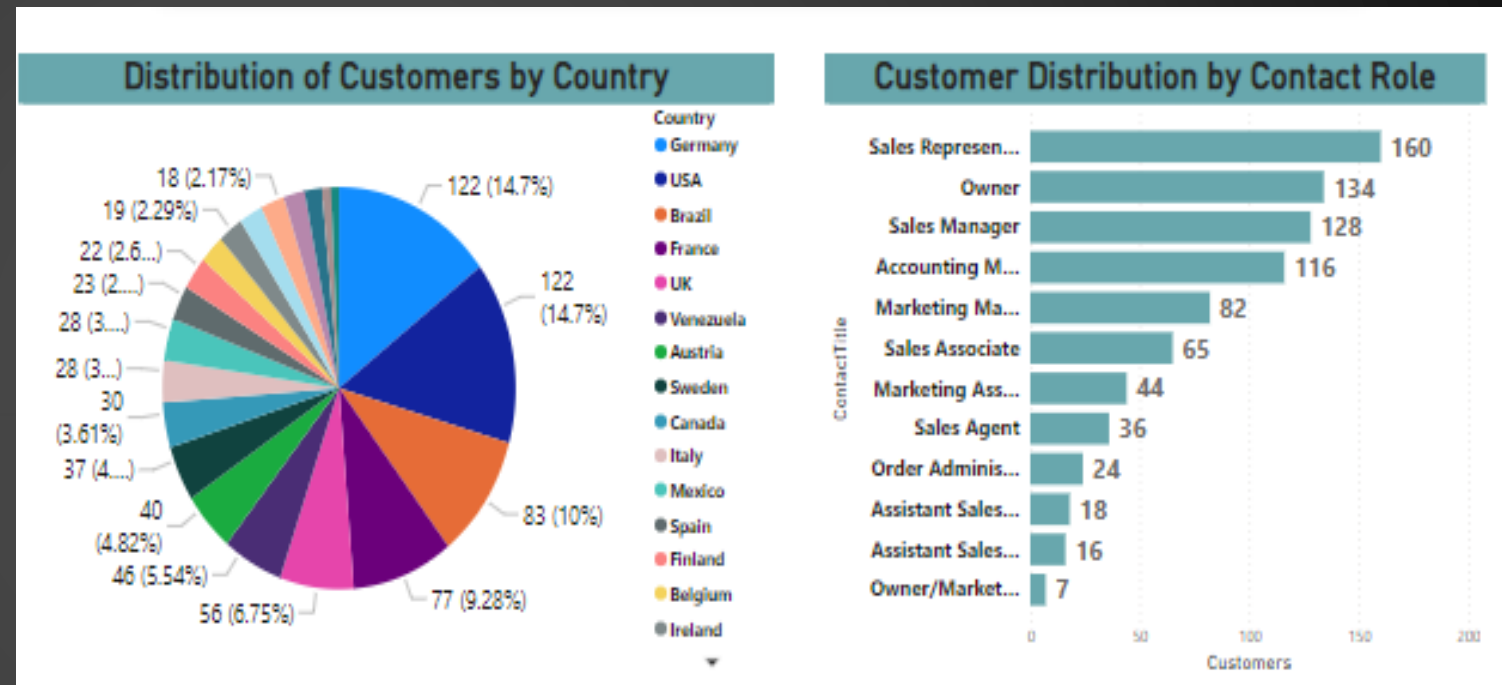
4.How does order volume change over time? Can we create a time series chart or stacked bar chart to visualize it?

The order volume shows a gradual increase over time, with some fluctuations. There's a notable jump in volume between July 1995 and April 1996
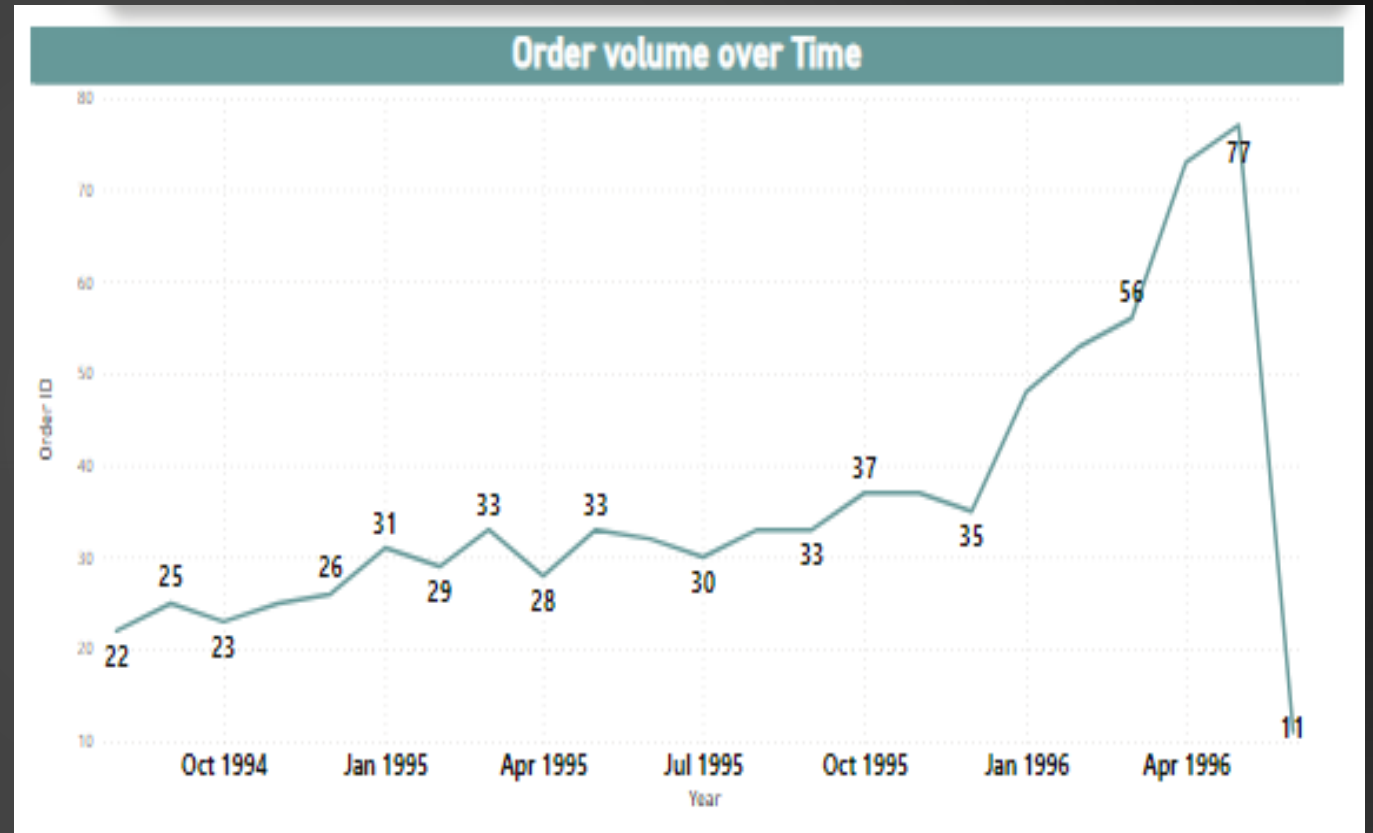
Order volume over Time

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

Majority of orders have a low to moderate value. The peak of the distribution is concentrated around the $2,000-$4,000 range, suggesting a high number of orders fall within this price. While most orders are concentrated in the lower to mid-range, there's a noticeable tail extending towards higher order values, indicating a presence of some large orders potentially impacting overall revenue
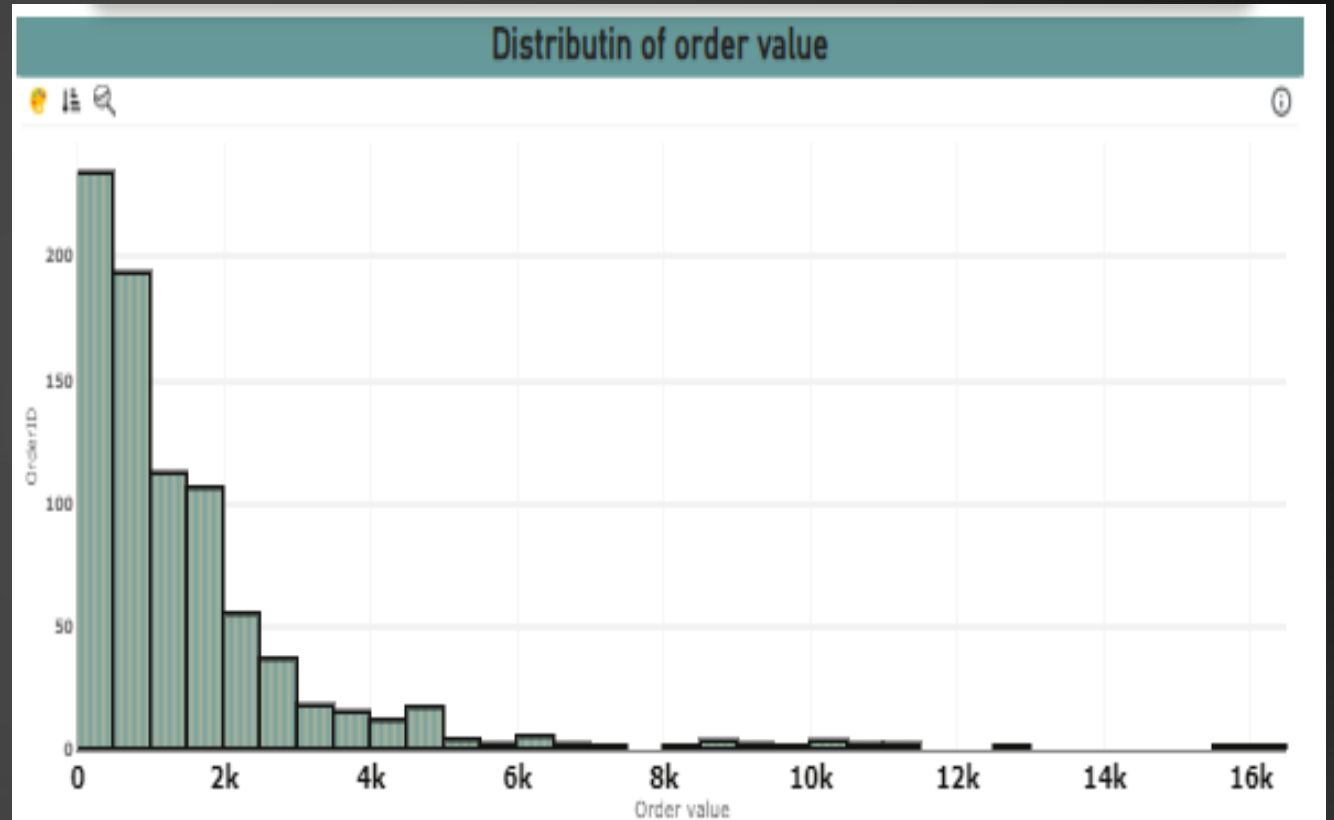
6.Can we visualize the average order processing time or shipping duration using a bar chart or box plot?

The x-axis shows whisker type, labeled 1.5 IQR. The y-axis shows average processing time in days. There are four horizontal lines representing quartiles and outliers. The median processing time is 7 days. The average processing time is 8 days. The whisker range is from -4 days to 17 days. There are two outliers, one at 37 days and one at 1 day.
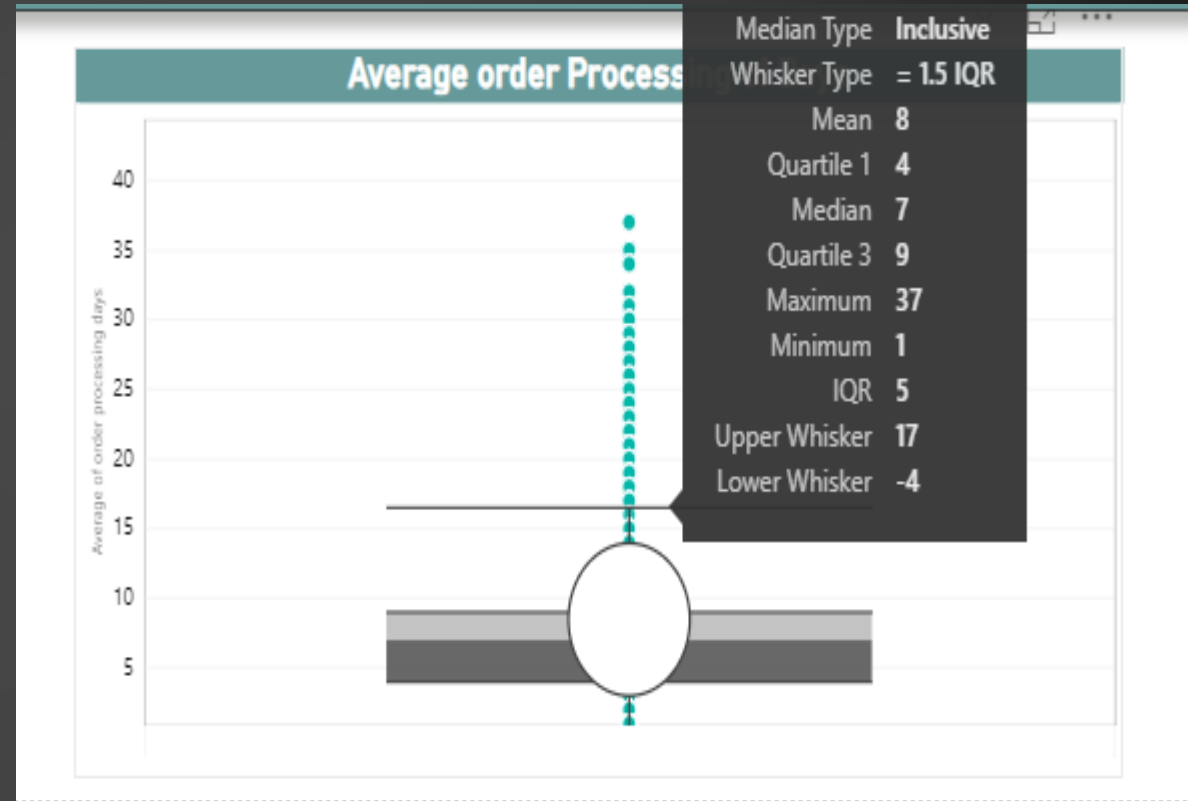
7.How does employee productivity vary across different departments or job roles? Can we create a stacked bar chart or grouped column chart to visualize it?

Sales representatives have the highest average productivity. Most of orders are done by the sale representatives. Insider sales coordinator and vice president have done almost equal numbers of orders and sales manager have the lowest Productivity. There's a significant difference in productivity between the top and bottom roles
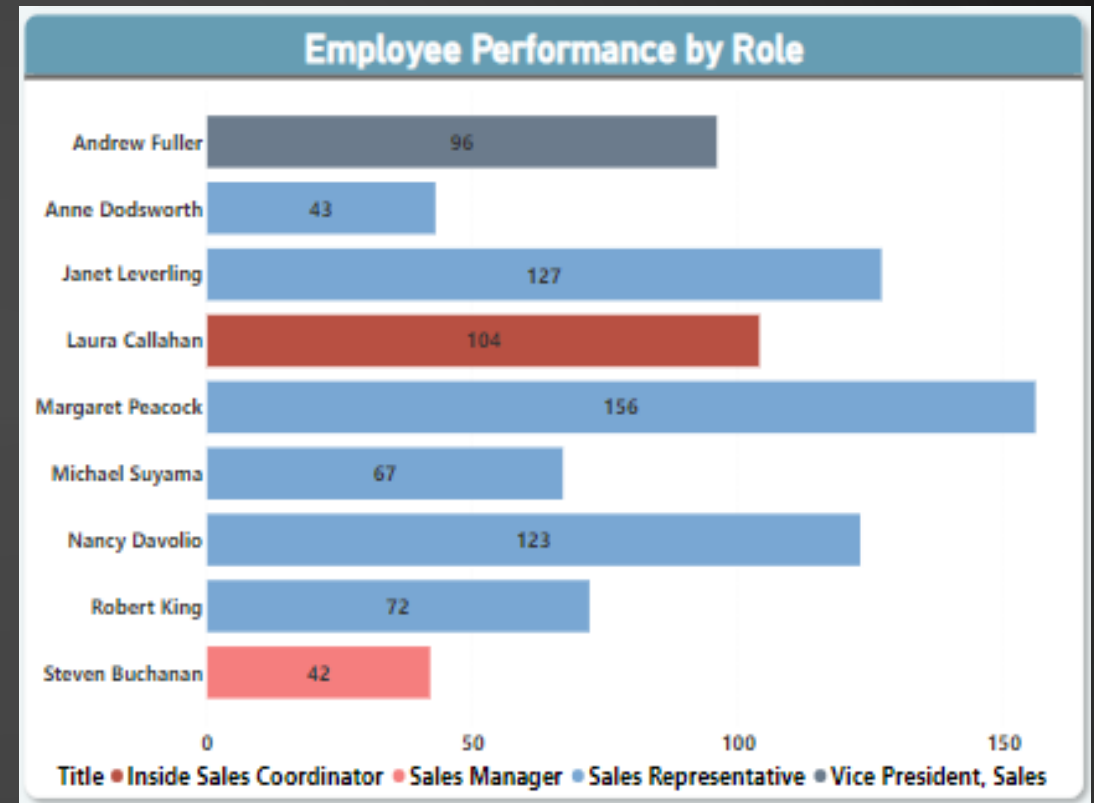
8. What is the distribution of employee tenure? Can we create a Histogram or box plot chart to display it?

Yes we can create a box plot for it .
Majority tenure of employees have
been with the company
around between 29 and 46 Months.
Average Tenure of employees is 35.6.
Highest tenure of the employee is 50
month and lowest is 19 months .This
indicate a presence of employees
who have been with the company
for a specific period of times.
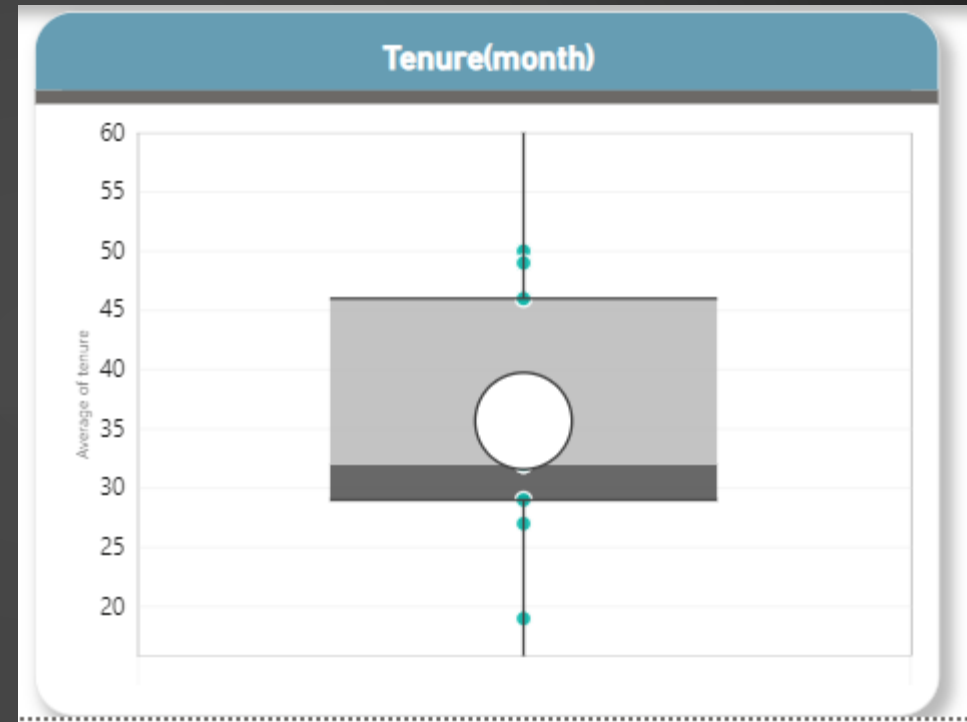
Tenure(month)

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

Yes , we can visualize it to check employee performance . Here employe ID 4 have done the highest task and lowest tasks done by Emp Id 5 (42 tasks) This help us to identify the performance of the employees.

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

Yes , we can create a stacked bar chart to check products performance by distributing them into a product rating . While the majority of products fall within the average and below average range, only a small percentage achieve "good" and "excellent" ratings. This signifies a potential opportunity to improve product quality or perception to increase customer satisfaction and positive feedback.
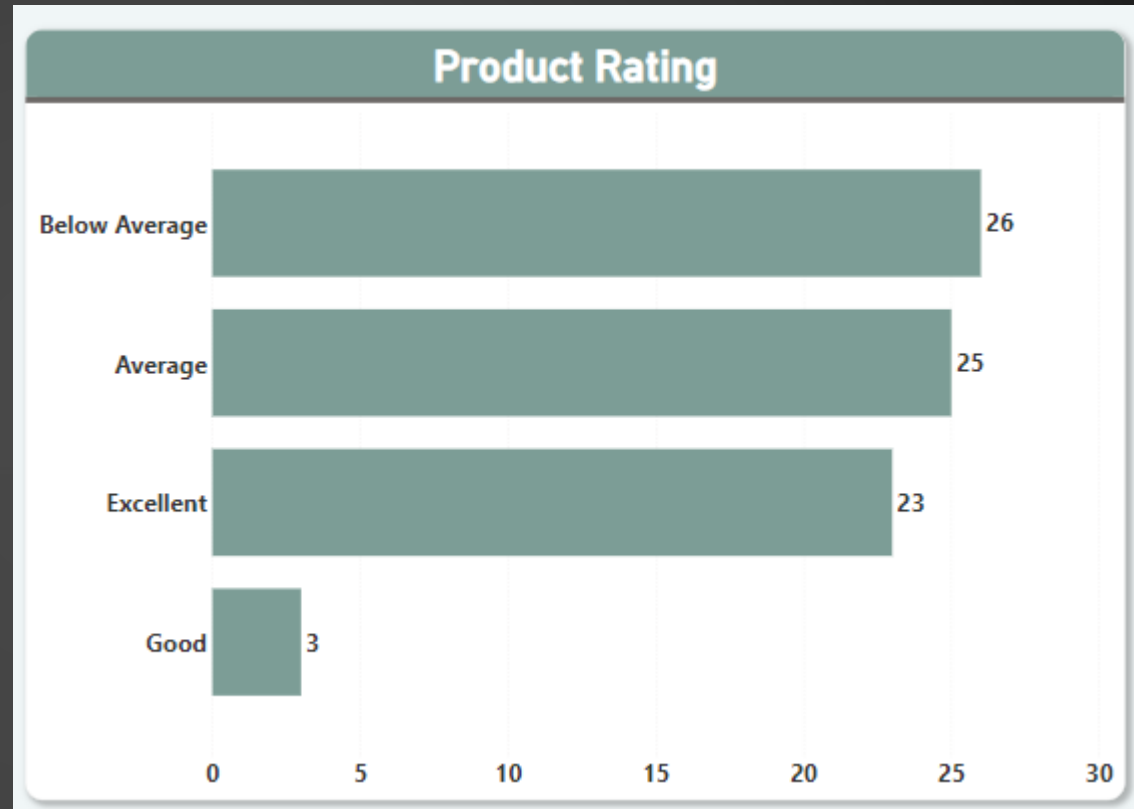
11. How does the sales volume vary across different product categories? Can we create a bar chart or tree map to display it?

Yes , we can create a tree map to check sales volume across the different Product categories.

- Beverages and dairy products are the top sellers, generating significantly higher sales volume compared to other categories.
- Confections, seafood, and meat/poultry show moderate sales volume, forming a middle tier and other categories show the lowest sales volume
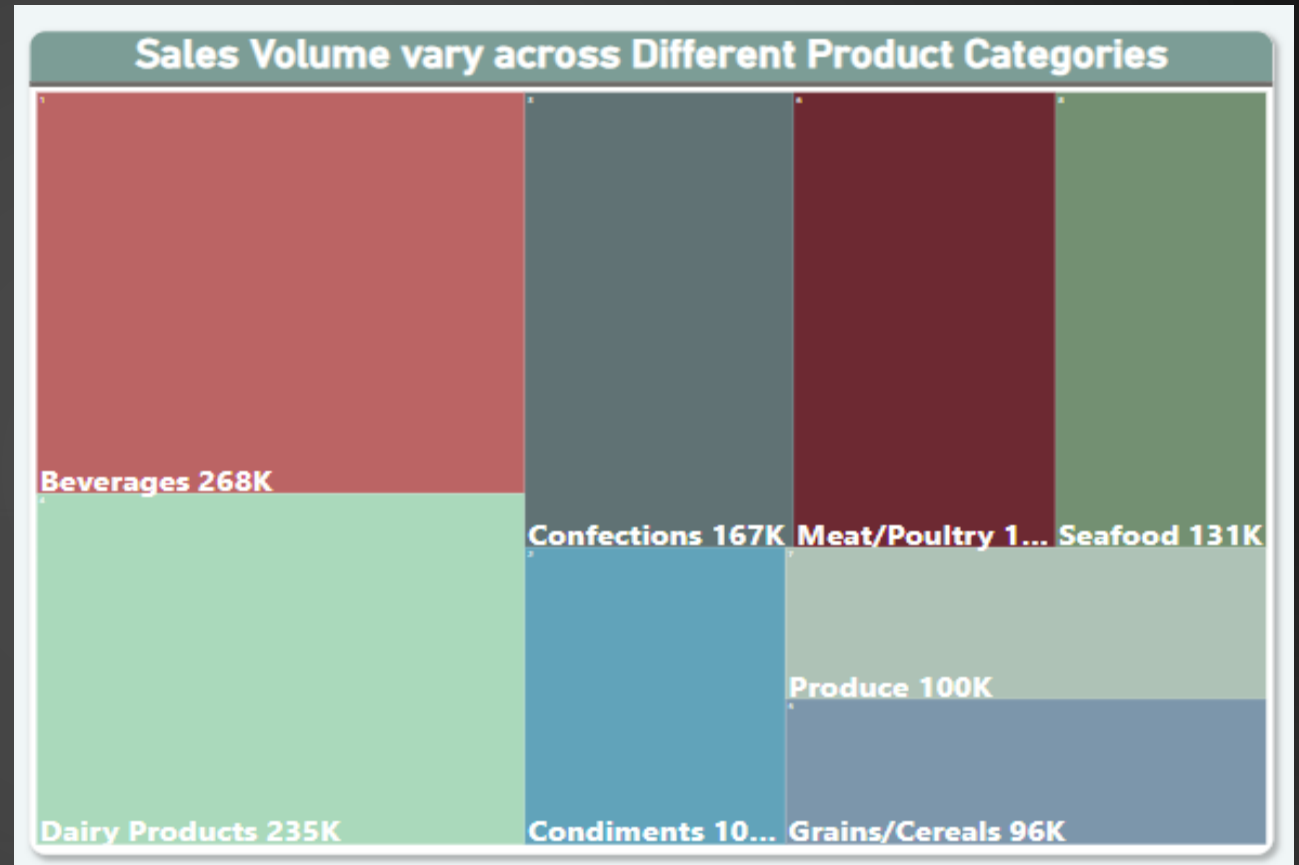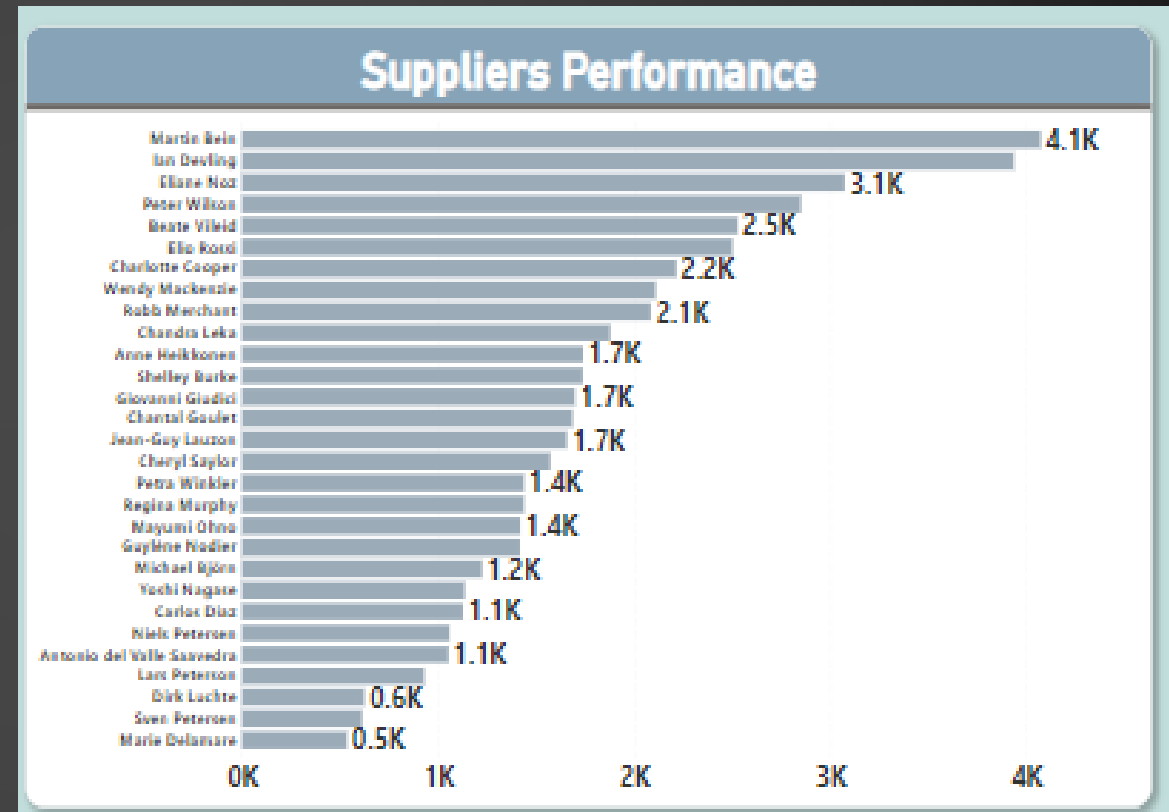
12. Can we visualize the pricing distribution of products using a box plot or violin plot?

Yes, we can visualize the price distribution of products using box plot .It help us to identify the outlier in the product price means which product have the highest price and lowest different from the average price of all the products . The most of product price lie b/w 13 and 33 .The max price of the product is 263 and min price is 2.5 and the average of price of the product is 28.
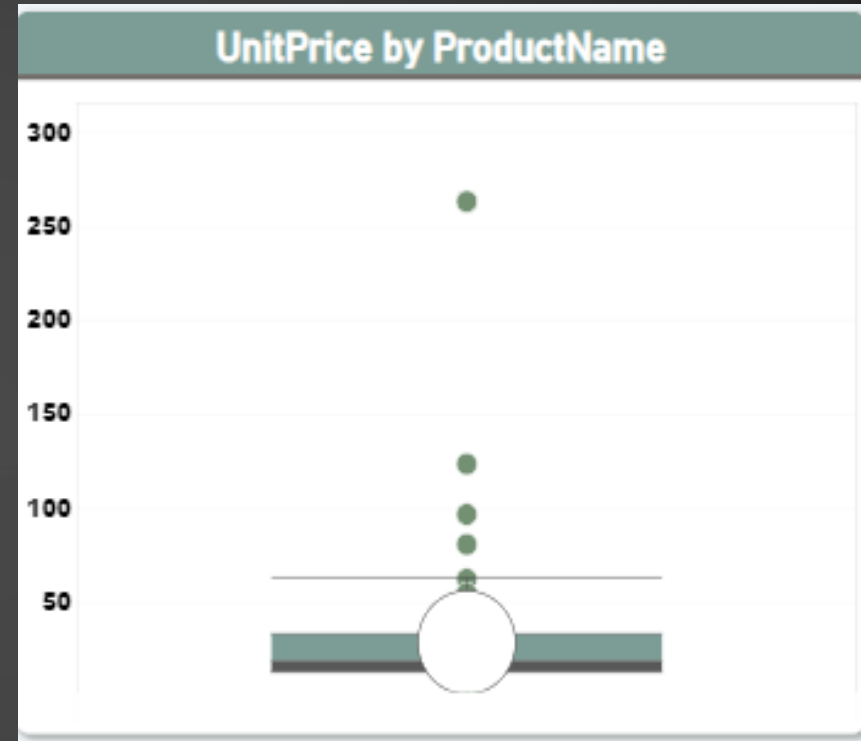
UnitPrice by ProductName

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

Yes we can create a bar chart to visualize the distribution of supplier rating by check which supplier have highest no. of order quantity delivery. if the company buying product continuously then we can compare the supplier with there order quantity .
From the chart we can see that supplier martin Bei
 have the highest order quantity.

Suppliers Performance

Supplier	Value
Martin Bein	4.1K
Ian Devling	
Eliane Noz	3.1K
Peter Wilson	
Beate Vileid	2.5K
Elio Rossi	
Charlotte Cooper	2.2K
Wendy Mackenzie	
Robb Merchant	2.1K
Chandra Leka	
Anne Heikkonen	1.7K
Shelley Burke	
Giovanni Giudici	1.7K
Chantal Goulet	
Jean-Guy Lauzon	1.7K
Cheryl Saylor	
Petra Winkler	1.4K
Regina Murphy	
Mayumi Ohno	1.4K
Guylène Nadier	
Michael Björn	1.2K
Yoshi Nagase	
Carlos Diaz	1.1K
Niels Petersen	
Antonio del Valle Saavedra	1.1K
Lars Peterson	
Dirk Luchte	0.6K
Sven Petersen	
Marie Delamare	0.5K

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?

Yes ,we can create a bar char to visualize the cost price structure vary across different suppliers by compare the average unit price provided by the suppliers .
This help us to check and compare the unit price of product by different Supplier to identity the which supplier is best for which product.
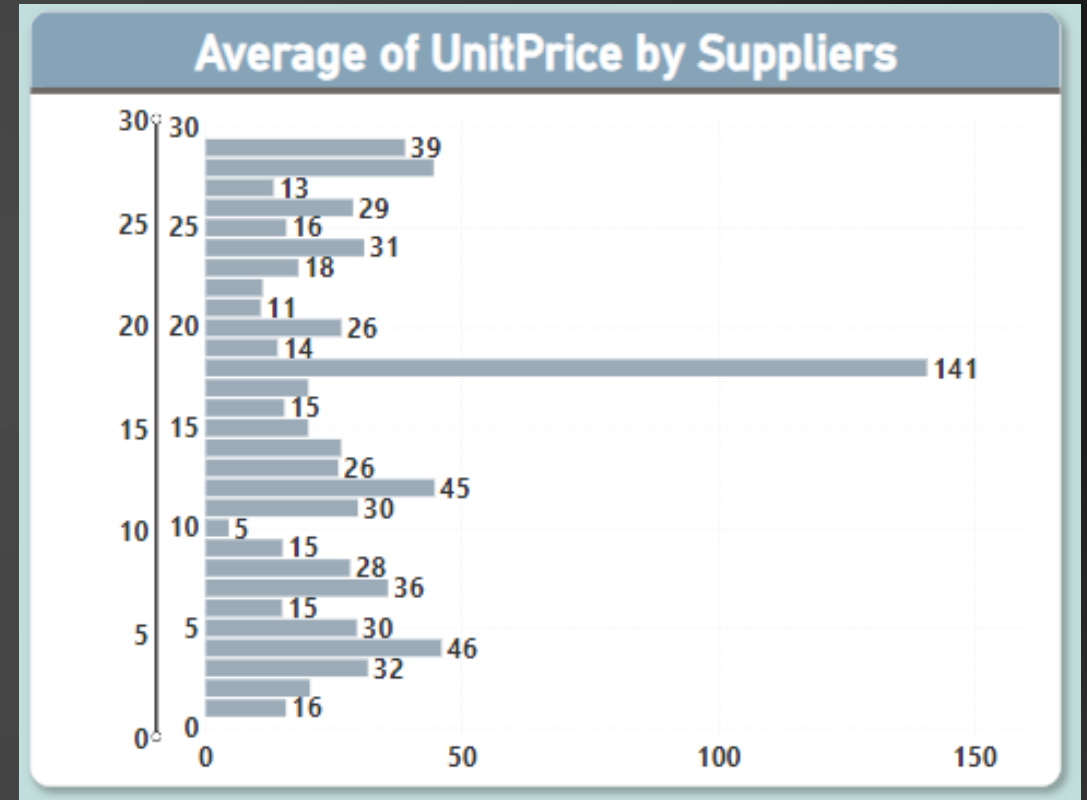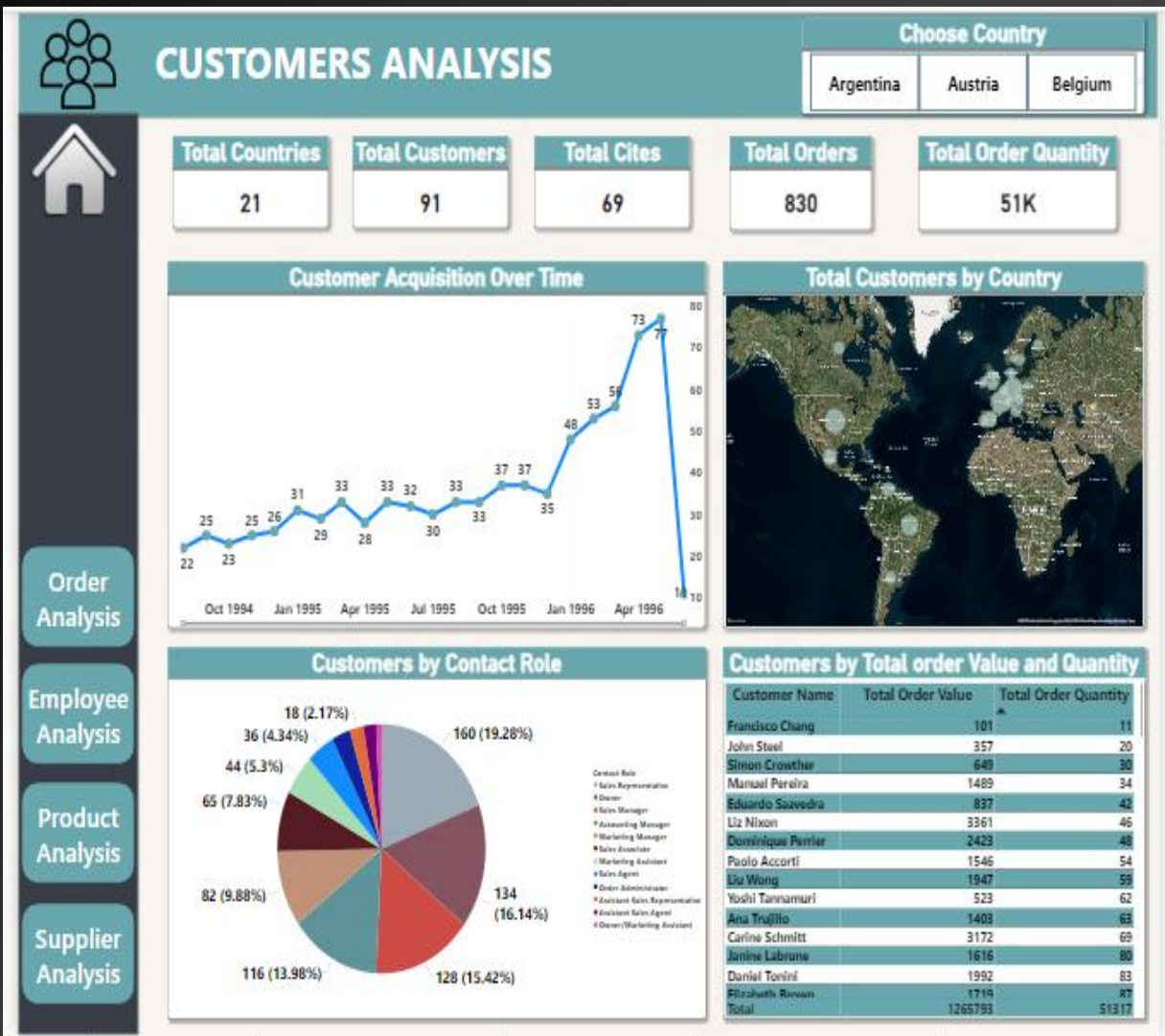
15.Can we visualize the geographical distribution of suppliers using a map or bubble chart?
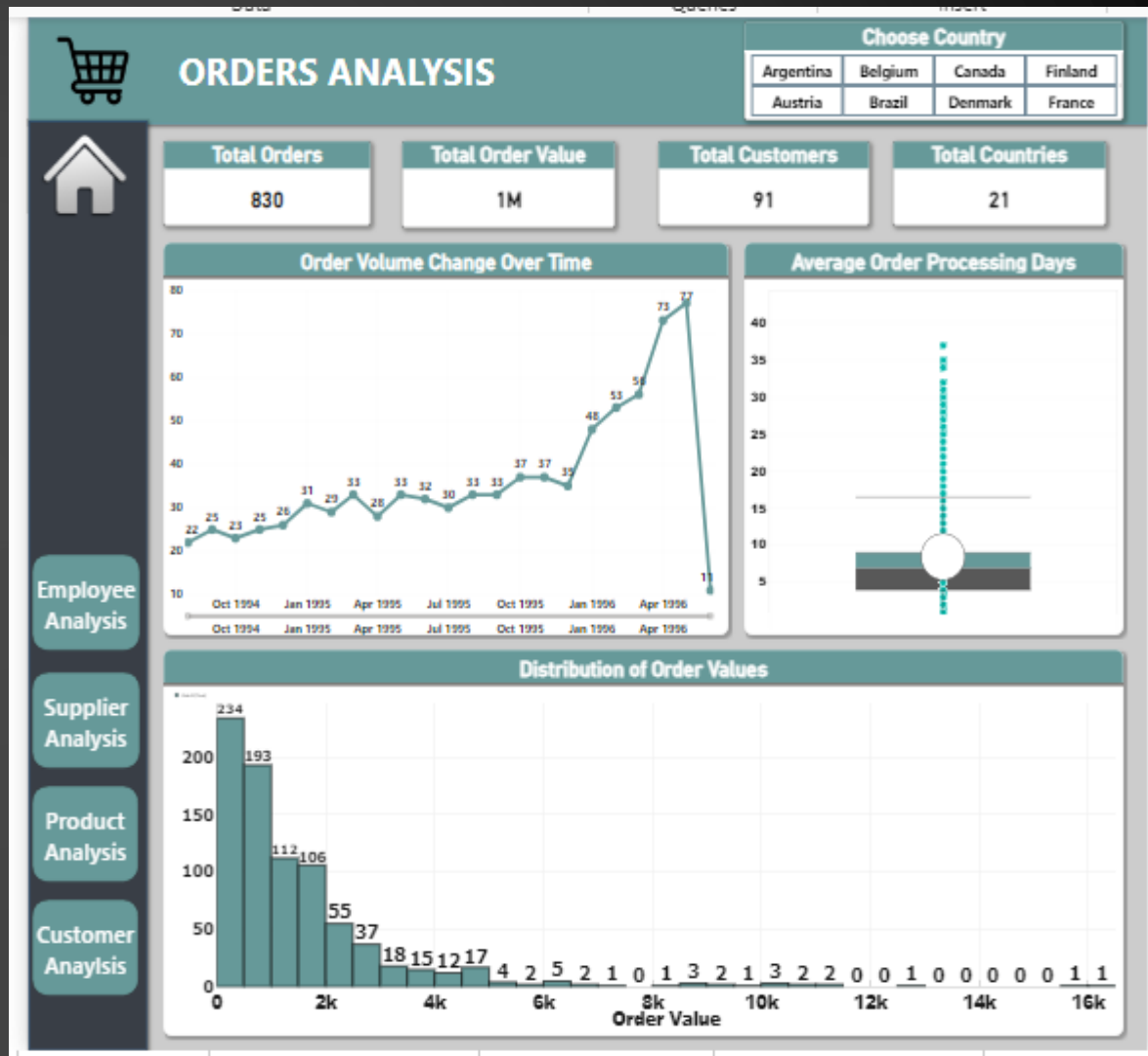
Yes , we can visualize the geographical distribution of supplier by map chart .
This help us to identify the country and region of supplier and how many supplier in a each country
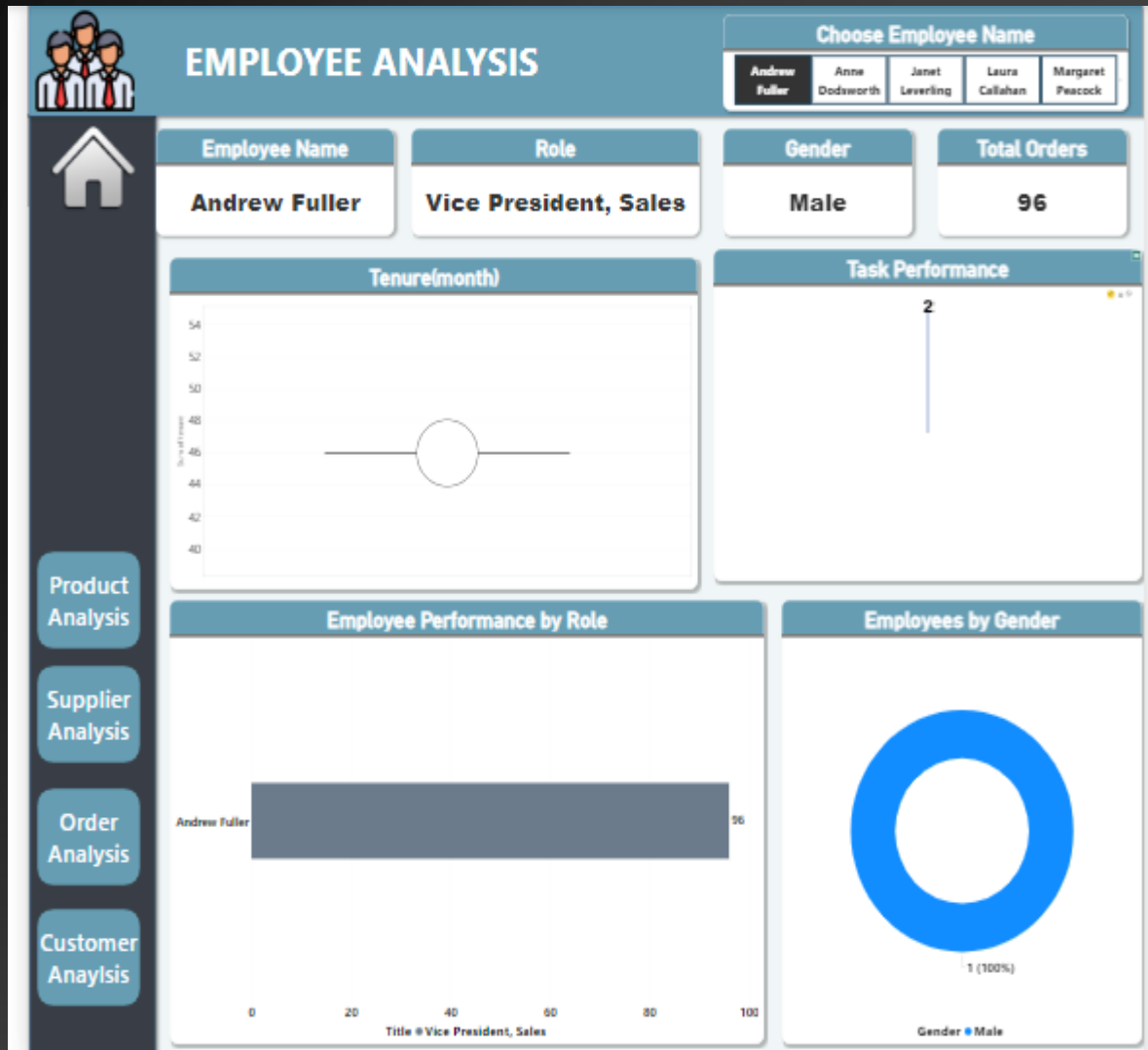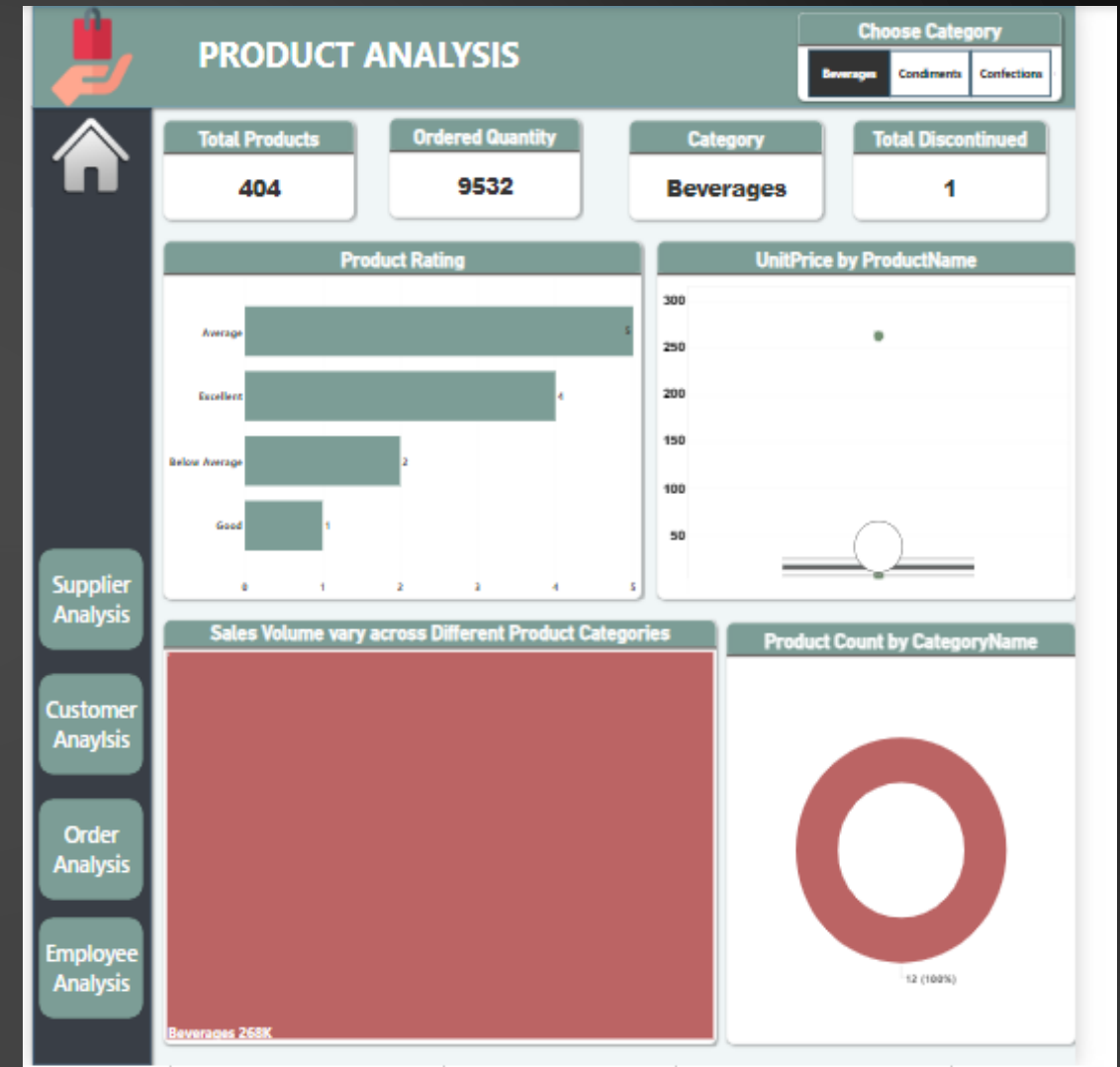
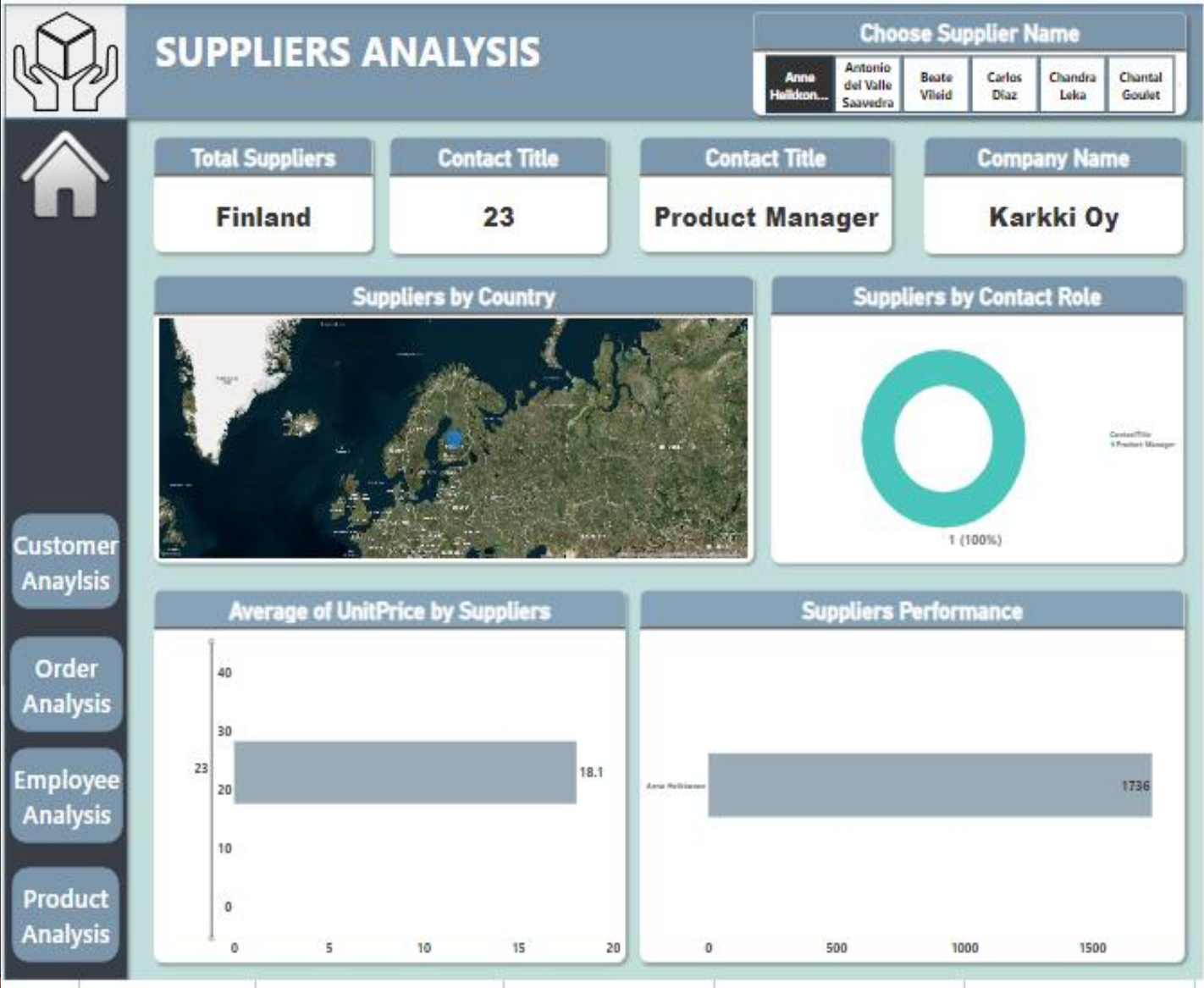# Customers Analysis

Orders Analysis

Employee Analysis

Product Analysis

Supplier Analysis
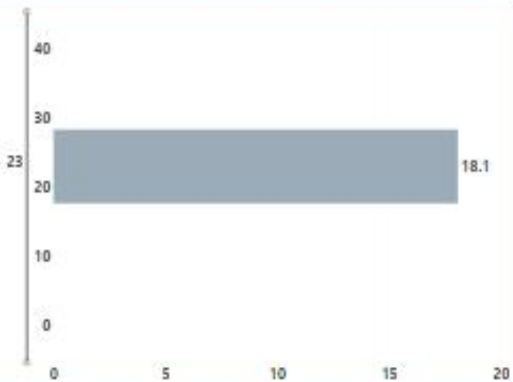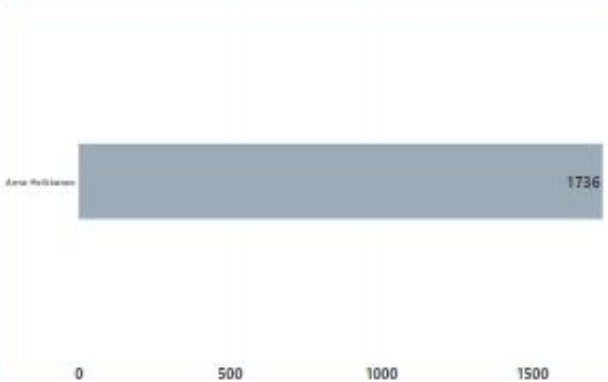
EDA QUESTIONS

1. WHAT ARE THE KEY FACTORS INFLUENCING CUSTOMER RETENTION OR LOYALTY BASED ON THE DATASET?

2. HOW DO CUSTOMER PREFERENCES VARY BASED ON THEIR LOCATION OR DEMOGRAPHICS? CAN WE EXPLORE THIS THROUGH INTERACTIVE VISUALIZATIONS?

3. ARE THERE ANY INTERESTING PATTERNS OR CLUSTERS IN CUSTOMER BEHAVIOR THAT CAN BE VISUALIZED TO IDENTIFY POTENTIAL MARKET SEGMENTS?

4. ARE THERE ANY SPECIFIC PRODUCT CATEGORIES OR SKUS THAT CONTRIBUTE SIGNIFICANTLY TO ORDER REVENUE? CAN WE IDENTIFY THEM THROUGH VISUALIZATIONS?

5. ARE THERE ANY CORRELATIONS BETWEEN ORDER SIZE AND CUSTOMER DEMOGRAPHICS OR PRODUCT CATEGORIES? CAN WE EXPLORE THIS VISUALLY USING SCATTER PLOTS OR HEATMAPS?

6. HOW DOES ORDER FREQUENCY VARY ACROSS DIFFERENT CUSTOMER SEGMENTS? CAN WE VISUALIZE THIS USING BAR CHARTS OR TREE MAPS ?

7. ARE THERE ANY CORRELATIONS BETWEEN EMPLOYEE SATISFACTION LEVELS AND KEY PERFORMANCE INDICATORS? CAN WE EXPLORE THIS VISUALLY THROUGH SCATTER PLOTS OR LINE CHARTS?

8. HOW DOES EMPLOYEE TURNOVER VARY ACROSS DIFFERENT DEPARTMENTS OR JOB ROLES? CAN WE VISUALIZE THIS USING BAR CHARTS OR HEATMAPS?

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?

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?

11. HOW DOES PRODUCT DEMAND FLUCTUATE OVER DIFFERENT SEASONS OR MONTHS? CAN WE VISUALIZE THIS THROUGH LINE CHARTS OR AREA CHARTS?

12. CAN WE IDENTIFY ANY OUTLIERS OR ANOMALIES IN PRODUCT PERFORMANCE OR SALES USING VISUALIZATIONS? HOW CAN THIS INFORMATION BE USED FOR PRODUCT OPTIMIZATION?

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?

14. HOW DOES SUPPLIER PERFORMANCE VARY ACROSS DIFFERENT PRODUCT CATEGORIES OR DEPARTMENTS? CAN WE VISUALIZE THIS USING STACKED BAR CHARTS OR GROUPED COLUMN CHARTS?

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?

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

conclusion : This query calculates RFM values (Recency, Frequency, Monetary Value) for each customer, enabling segmentation based on their purchasing behavior . Recency show the number of days between the customer's latest purchase and a reference date, often the current date. It provides a straightforward understanding of how recently a customer has engaged in buying activity.

```
create view Loyalty as select
o.CustomerID,
count(o.OrderID) Frequency,
min(datediff((select max(date(OrderDate)) from orders),date(o.OrderDate))) Recency,
round(sum( UnitPrice * Quantity * ( 1 - Discount )),2) Monetary
from orders o
join northwind.`order details` od
on o.OrderID = od.OrderID
group by 1
order by 1
```
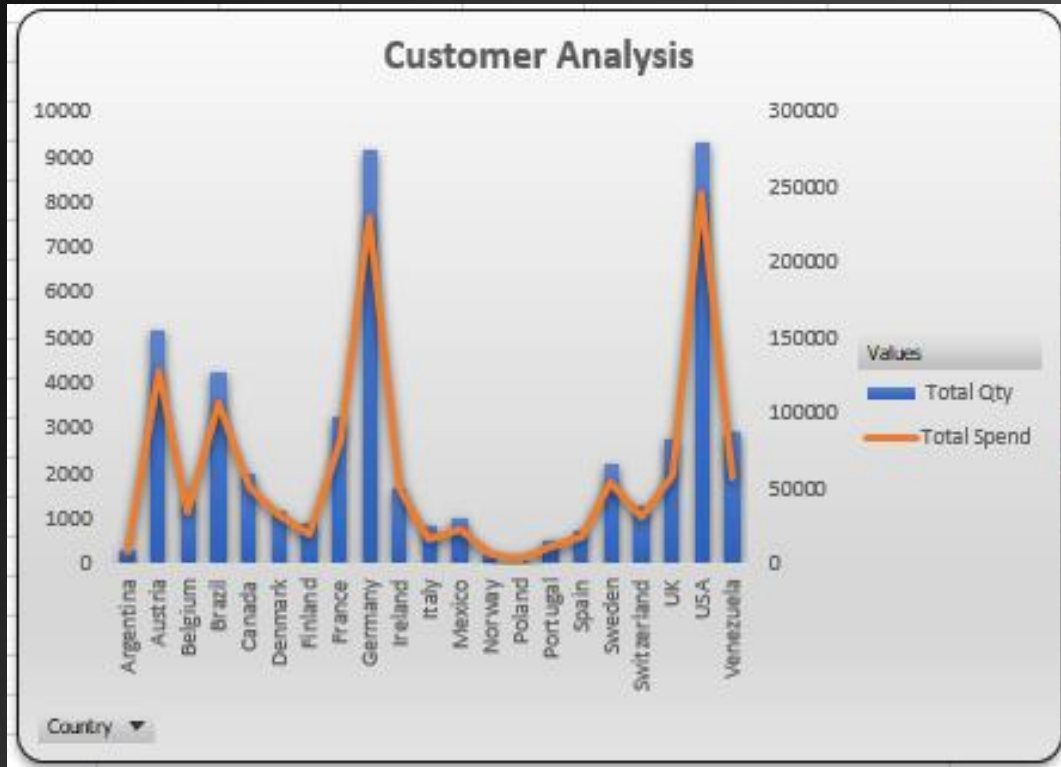
Row Labels	order Frequency	Recency_	Monetary value
ALFKI	12	27	4273
ANATR	10	63	1402.95
ANTON	17	98	7023.98
AROUT	30	26	13390.65
BERGS	52	63	24927.58
BLAUS	14	7	3239.8
BLONP	26	114	18534.08
BOLID	6	43	4232.85
BONAP	44	0	21963.25
BOTTM	35	12	20801.6
BSBEV	22	22	6089.9
CACTU	11	8	1814.8
CENTC	2	657	100.8
CHOPS	22	14	12348.88
COMMI	10	14	3810.75
CONSH	7	103	1719.1
DRACD	10	2	3763.21
DUMON	9	79	1615.9
EASTC	21	8	14761.04
ERNSH	102	1	104874.98
FAMIA	19	187	4107.55
FOLIG	16	135	11666.9
FOLKO	45	9	29567.56
FRANK	48	27	26656.56
FRANR	6	43	3172.16
FRANS	10	6	1545.7

The most loyal customers:
1. SAVEA :This customer has the highest number of orders 116 and has spent the most money in total 104361.95 . They also placed an order recently within 5 days.
2. ERNSH: This customer has the highest number of orders 102 and has spent the most money in total 104874.98 . They also placed an order recently within 1 days.
3.RATTC: This customer has the highest number of orders 77 and has spent the most money in total 51097.8 . They also placed an order recently within 0 days.
4.QUICK: This customer has the highest number of orders 86 and has spent the most money in total 110277 . They also placed an order recently within 22 days.
5.BONAP: This customer has the highest number of orders 44 and has spent the most money in total 21963 . They also placed an order recently within 0days.

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

➢ **Customer Total spend and total Quantity in Each country**


```
create view Analyzing_Product_Preferences as
SELECT  distinct c.customerID, c.Country, SUM(od.Quantity) AS TotalQty
, SUM(od.Quantity * od.UnitPrice * (1-od.Discount) )AS TotalSpend
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY c.customerID, c.Country, p.ProductName
ORDER BY c.Country, TotalQty DESC;
```

Row Labels	Total Qty	Total Spend
Argentina	319	7839.1
Austria	5167	128003.838
Belgium	1392	33824.855
Brazil	4237	106813.776
Canada	1984	50196.2899
Denmark	1170	32661.0225
Finland	885	18810.0525
France	3244	81246.3224
Germany	9153	229684.633
Ireland	1684	49979.9049
Italy	822	15770.155
Mexico	1005	23332.0775
Norway	161	5735.15
Poland	205	3531.95
Portugal	533	11472.3625
Spain	718	17983.2
Sweden	2235	54495.14
Switzerland	1275	31692.659
UK	2742	58971.31
USA	9330	245584.61
Venezuela	2936	56810.6289
Grand Total	**51197**	**1264439.04**

This query helps to check the total spend and Total Quantity purchased in each country .
By this we can identify the top country that contribute most in increasing the revenue.
USA have highest Total spend around $245584.

Product Preference by Country
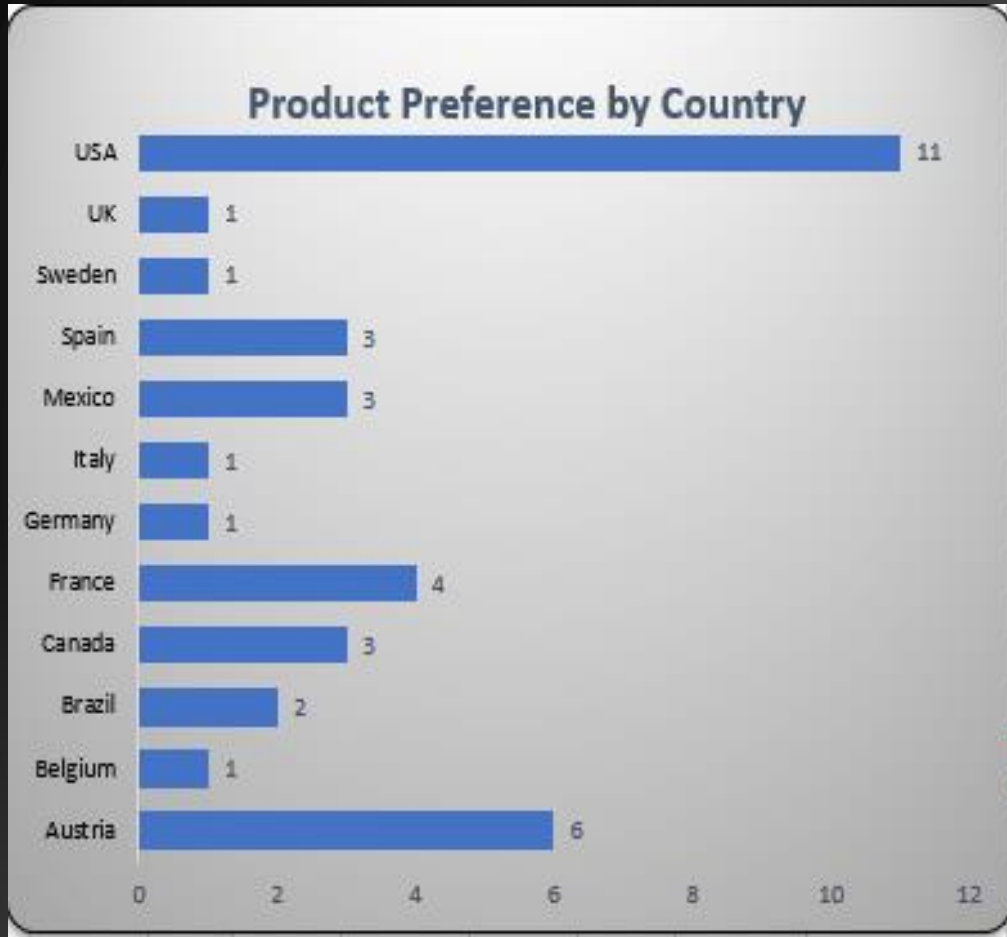
ProductName	Alice Mutton 🔽
Row Labels 🔽	**Customer Preference**
Austria	6
Belgium	1
Brazil	2
Canada	3
France	4
Germany	1
Italy	1
Mexico	3
Spain	3
Sweden	1
UK	1
USA	11
Grand Total	**37**

```sql
create view customer_Preference as
SELECT c.Country, p.ProductName, COUNT(*) AS CustomerCount
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY c.Country, p.ProductName
ORDER BY  CustomerCount DESC;
```

This query provides a list of countries and the top-selling product based on the number of unique customers who purchased it in each country. It focuses on customer preference rather than total quantity.

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

➢ RFM Segmentation
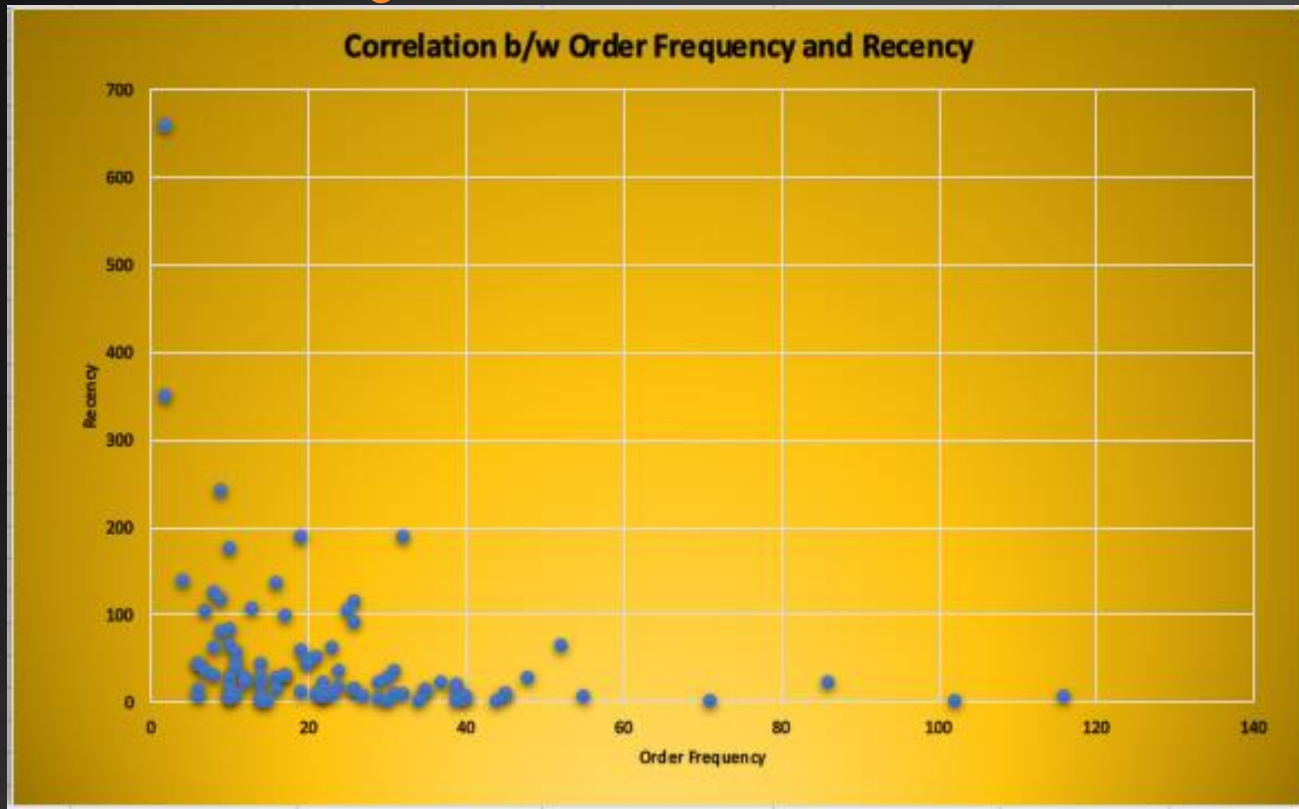
Row Labels	Sum of Frequency	Sum of Recency	Sum of Monetary
ALFKI	12	27	4273
ANATR	10	63	1402.95
ANTON	17	98	7023.98
AROUT	30	26	13390.65
BERGS	52	63	24927.58
BLAUS	14	7	3239.8
BLONP	26	114	18534.08
BOLID	6	43	4232.85
BONAP	44	0	21963.25
BOTTM	35	12	20801.6
BSBEV	22	22	6089.9
CACTU	11	8	1814.8
CENTC	2	657	100.8
CHOPS	22	14	12348.88
COMMI	10	14	3810.75
CONSH	7	103	1719.1
DRACD	10	2	3763.21
DUMON	9	79	1615.9
EASTC	21	8	14761.04
ERNSH	102	1	104874.98
FAMIA	19	187	4107.55
FOLIG	16	135	11555.0

Question 1   question 2   **Question 3**   question 4   Quest

This provide the Correlation b/w Recency and order Frequency of customer from this we can identify the customer behavior and purchasing pattern. Recency show the number of days between the customer's latest purchase . It provides a straightforward understanding of how recently a customer has engaged in buying activity.

Product Preferences by customers
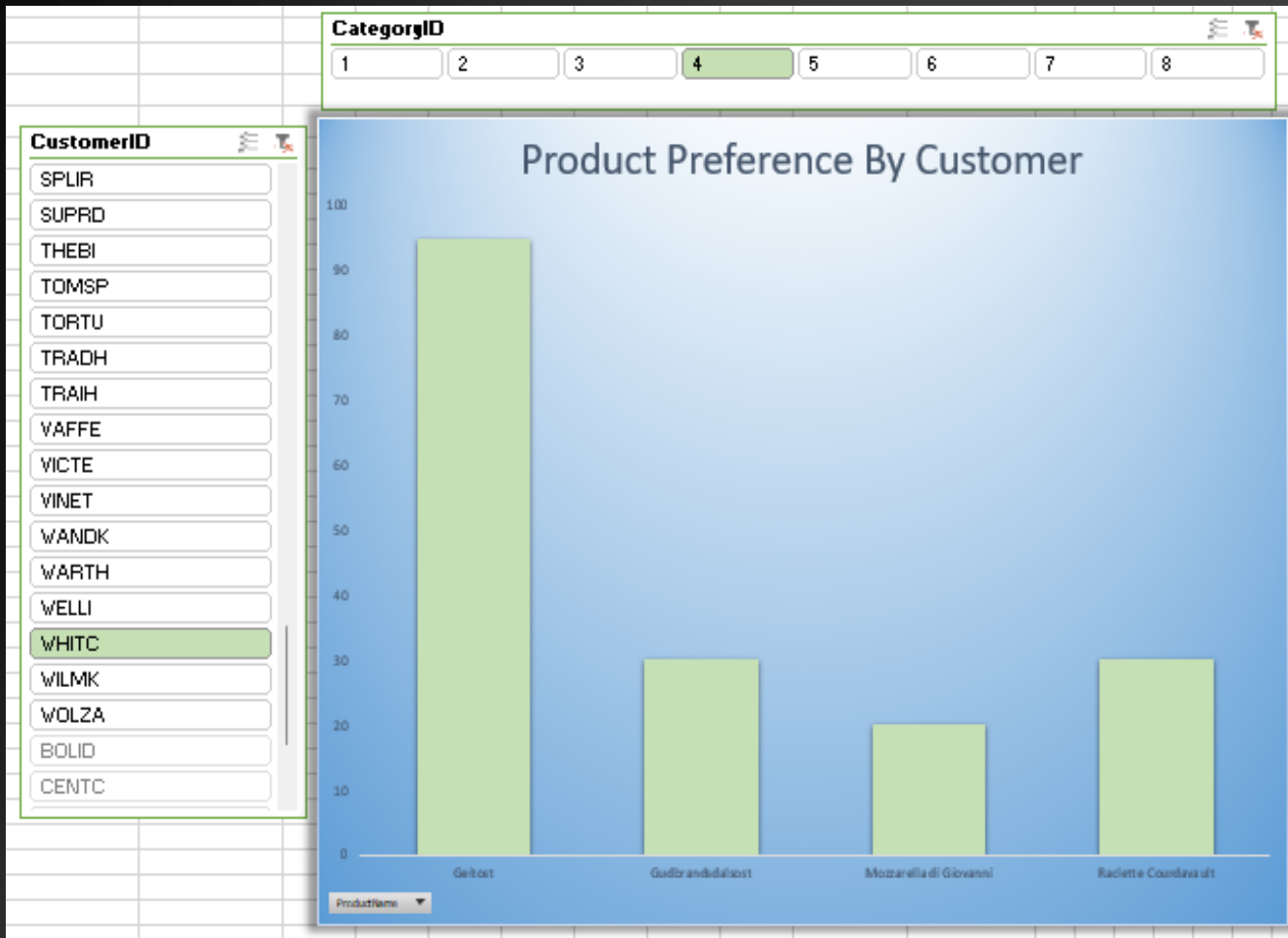

```
create view Product_category_preference as
SELECT c.CustomerID, c.CompanyName, p.CategoryID,p.ProductName, SUM(od.Quantity) AS TotalQty
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY c.CustomerID, c.CompanyName, p.CategoryID, p.ProductName
ORDER BY c.CustomerID, TotalQty DESC;
```

Row Labels	Sum of TotalQty
Geitost	95
Gudbrandsdalsost	30
Mozzarella di Giovanni	20
Raclette Courdavault	30
Grand Total	**175**

This query provide the preference of customer for different Category. This show that the each Customer Total Quantity From the each Product.

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


Specific Product Category

Row Labels	Total Revenue	Total Quantity
Chai	12788.10	828
Chang	16355.96	1057
Chartreuse verte	12294.54	793
Côte de Blaye	141396.73	623
Guaraná Fantástica	4504.36	1125
Ipoh Coffee	23526.70	580
Lakkalikööri	15760.44	981
Laughing Lumberjack Lager	2396.80	184
Outback Lager	10672.65	817
Rhönbräu Klosterbier	8177.49	1155
Sasquatch Ale	6350.40	506
Steeleye Stout	13644.00	883
Grand Total	**267868.18**	**9532**

```
create view specific_product_categories as
SELECT p.CategoryID, p.ProductName,
    SUM(od.Quantity * od.UnitPrice) AS TotalRevenue,
    SUM(od.Quantity) AS TotalQuantity
FROM Products p
INNER JOIN northwind.`order details` od ON p.ProductID = od.ProductID
INNER JOIN Orders o ON od.OrderID = o.OrderID
GROUP BY p.CategoryID, p.ProductName
ORDER BY TotalRevenue DESC;
```

This provide total revenue and Total Quantity of each product for a specific category and in the chart secondary axis show the total Quantity and bar show the total revenue from each product. This help us to compare the total sale of each product of different category.

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


Correlation b/w Country Code and OrderSize
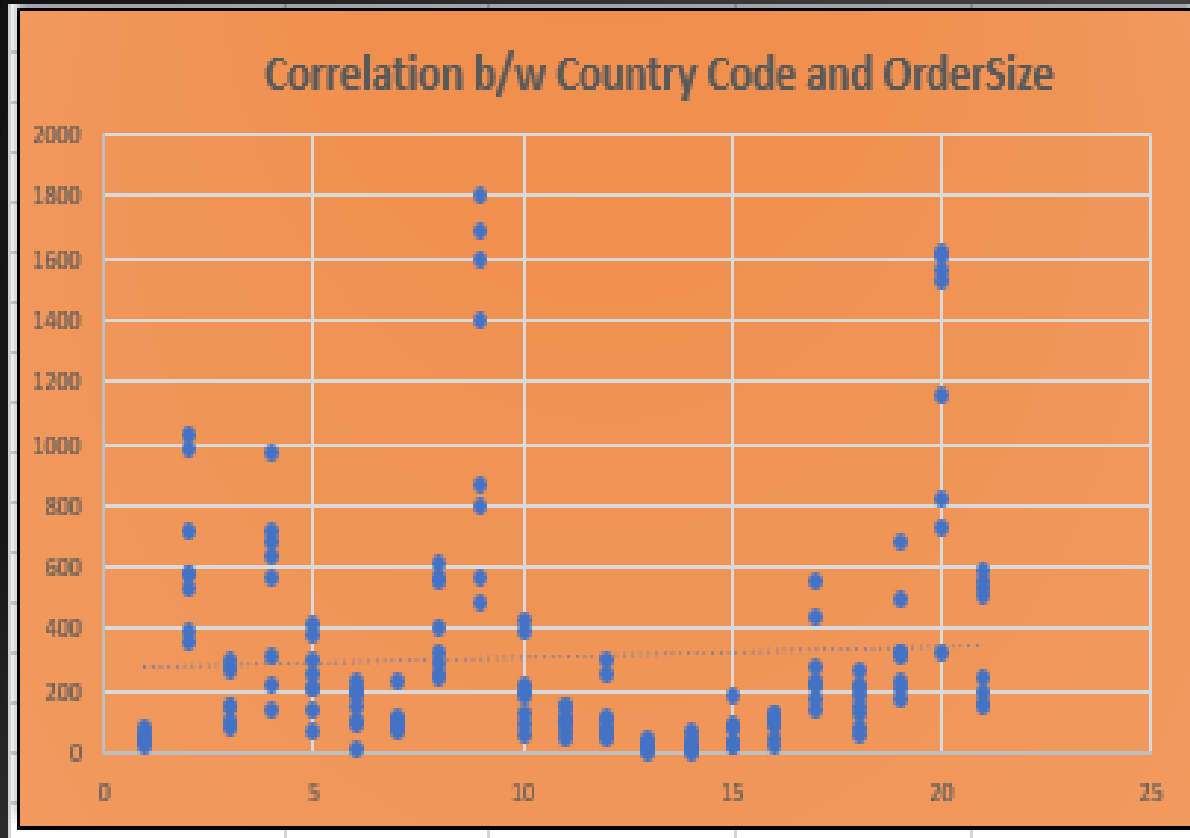
```sql
create view correlations_order_size_customer_demographics as
SELECT c.Country, p.CategoryID,
case when c.country='Argentina' then 1  when c.country='Austria' then 2   when c.country="Belgium" then 3
when c.country="Brazil" then 4 when c.country="Canada" then 5 when c.country="Denmark" then 6
when c.country="Finland" then 7 when c.country="France" then 8  when c.country="Germany" then 9
when c.country="Ireland" then 10 when c.country="Italy" then 11 when c.country="Mexico" then 12
when c.country="Norway" then 13 when c.country="Poland" then 14 when c.country="Portugal" then 15
when c.country="Spain" then 16  when c.country="Sweden" then 17 when c.country="Switzerland" then 18
when c.country="UK" then 19  when c.country="USA" then 20 when c.country="Venezuela" then 21 end as countrycode,
SUM(od.Quantity) AS OrderSize
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY 1,2,3
```

The correlation coefficient between country code and order size is 0.0586.This indicates a very weak positive correlation. Based on this correlation, we cannot conclude that there's a significant relationship between country code and order size. Other factors are likely more influential in determining order size.

6. How does order frequency vary across different customer segments? Can we visualize this using bar charts or tree maps ?
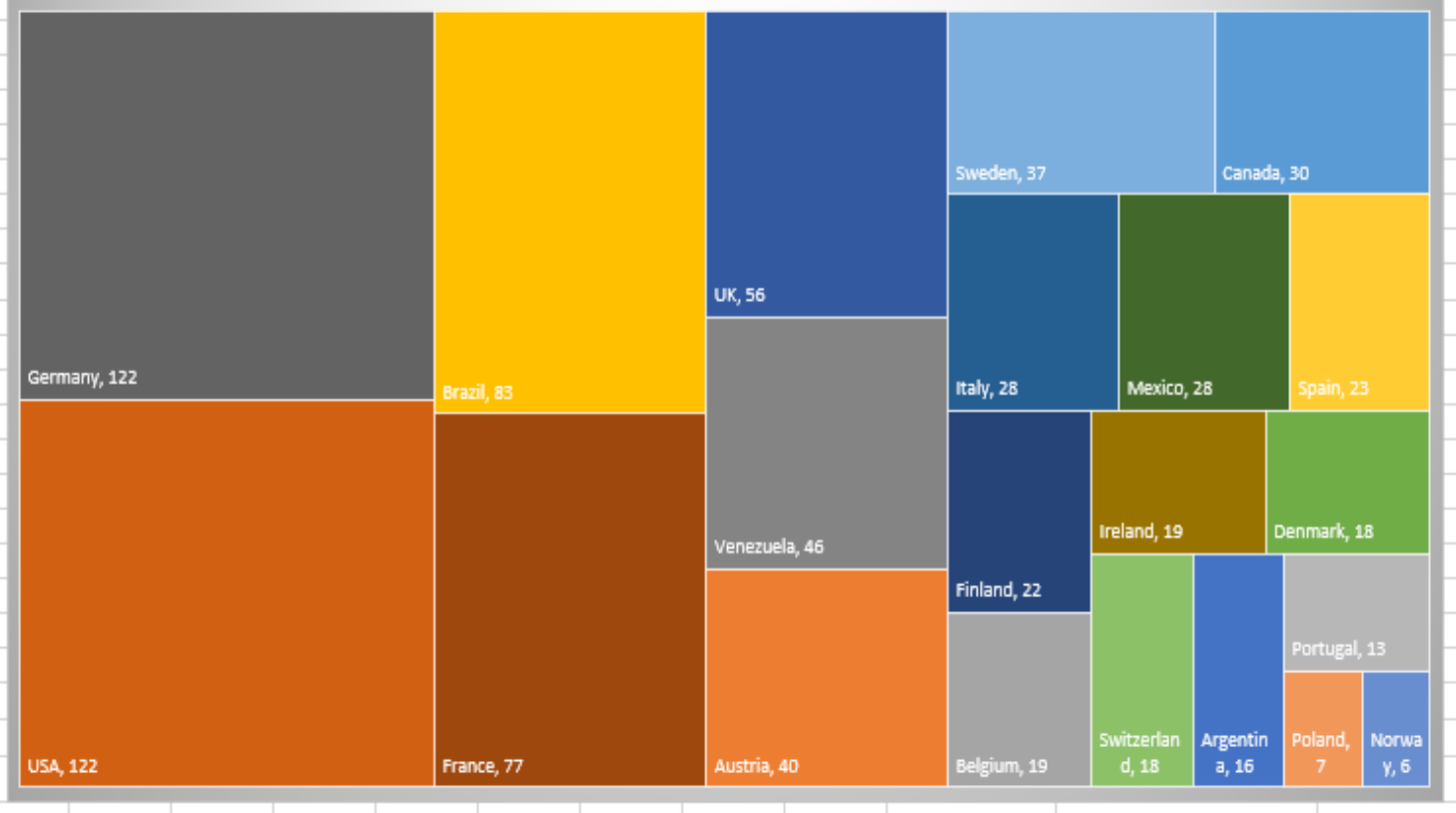
```sql
create view customer_segments as SELECT
    c.Country,
    c.City,
    COUNT(o.OrderID) AS NumberOfOrders
FROM Customers c
LEFT JOIN Orders o ON c.CustomerID = o.CustomerID
GROUP BY c.Country, c.City
ORDER BY c.Country, c.City
```


Order Frequency vary Across Differnent Country

Country	Orders
Argentina	16
Austria	40
Belgium	19
Brazil	83
Canada	30
Denmark	18
Finland	22
France	77
Germany	122
Ireland	19
Italy	28
Mexico	28
Norway	6
Poland	7
Portugal	13
Spain	23
Sweden	37
Switzerland	18
UK	56
USA	122
Venezuela	46

The United States and Germany are the two countries with the most orders. This could be due to a number of factors, such as the size of the population, the level of economic development, or the popularity of the product or service being sold.
Some countries have relatively few orders. These include Portugal, Ireland, and Denmark.

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

Correlation b/w Total Sale and Satisfaction Level

```
create view Emp_performance as
select e.EmployeeID,concat(e.FirstName," ",e.LastName) As EmployeeName,
count(distinct o.CustomerID) as Total_Customers ,
round(sum(od.unitprice * od.Quantity*(1-od.discount)),2) as Total_sale from employees e
inner join orders o on o.EmployeeID=e.EmployeeID
inner join `order details` od on od.OrderId=o.OrderID
group by 1,2
```

| Row Labels	Total Customers	Total sale	Satisfaction Level.
Andrew Fuller	59	166537.75	5
Anne Dodsworth	29	77308.07	1
Janet Leverling	63	202812.84	6
Laura Callahan	56	126862.28	4
Margaret Peacock	75	232890.85	8
Michael Suyama	43	73913.13	2
Nancy Davolio	65	192107.6	7
Robert King	45	124568.23	3
Steven Buchanan	29	68792.28	1

The correlation b/w Total sale and satisfaction level is 0.0979 and this show the weak correlation .
Margaret Peacok (Employee ID 4) have most satisfaction level 8 and Total sale of 132890.85.

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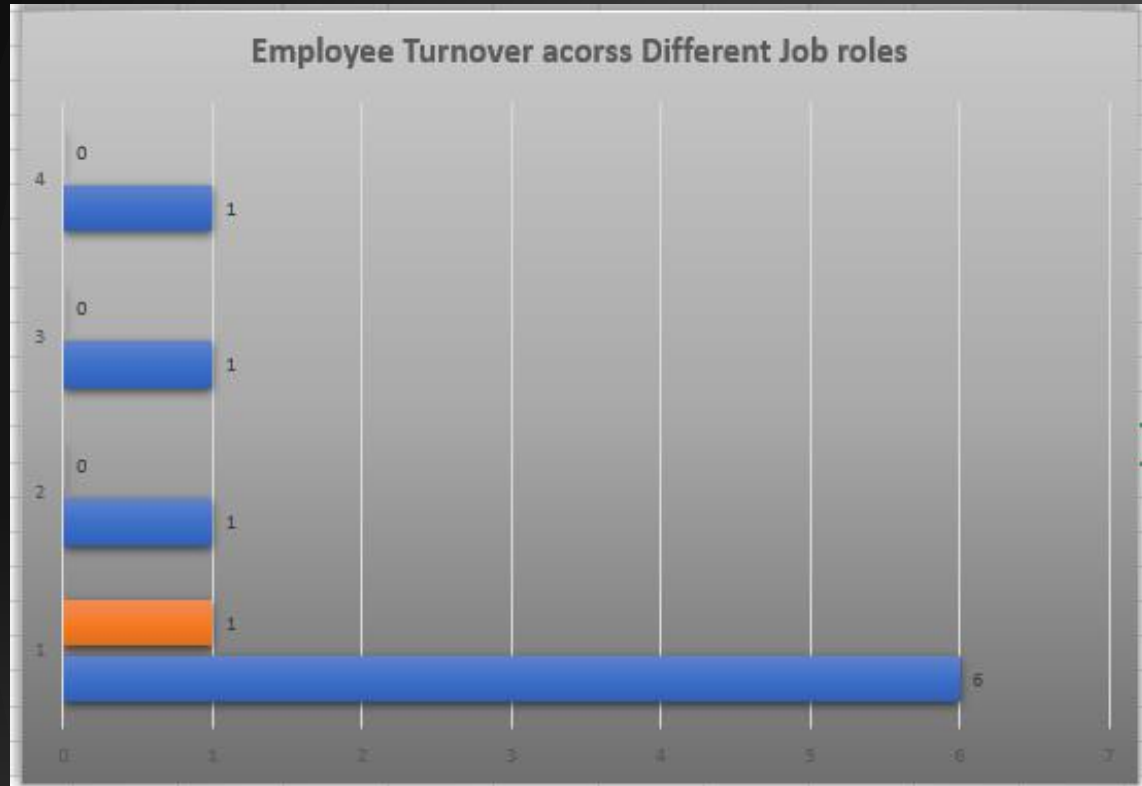

```
create view turnover as SELECT
e.Title as JobRole,
COUNT(e.EmployeeID) as TotalEmployees,
sum( CASE WHEN (timestampdiff(year,hiredate,maxD))<2
THEN 1 ELSE 0 END)
as NewHires
FROM Employees e
left join (select employeeId,max(orderdate) as maxD from orders
group by 1)m on m.EmployeeID=e.EmployeeID
GROUP BY 1|
```

JobRole	TotalEmployees	NewHires
Sales Representative	6	1
Vice President, Sales	1	0
Sales Manager	1	0
Inside Sales Coordinator	1	0

This query provide last 2 year hire data and visualization provide a simple yet effective way to analyze employee turnover across different role. This data indicates that the Sales Representative role likely involves high employee turnover and ongoing hiring given the disproportionately large headcount compared to management positions.

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?

Employee Performance

(Bubble chart — Tenure vs. Total Task)

Nancy Davolio, Sales Representative, 1
Janet Leverling, Sales Representative, 3
Robert King, Sales Representative, 7
Andrew Fuller, Vice President, Sales, 2
Steven Buchanan, Sales Manager, 5
Margaret Peacock, Sales Representative, 4
Laura Callahan, Inside Sales Coordinator, 8
Michael Suyama, Sales Representative, 6
Anne Dodsworth, Sales Representative, 9
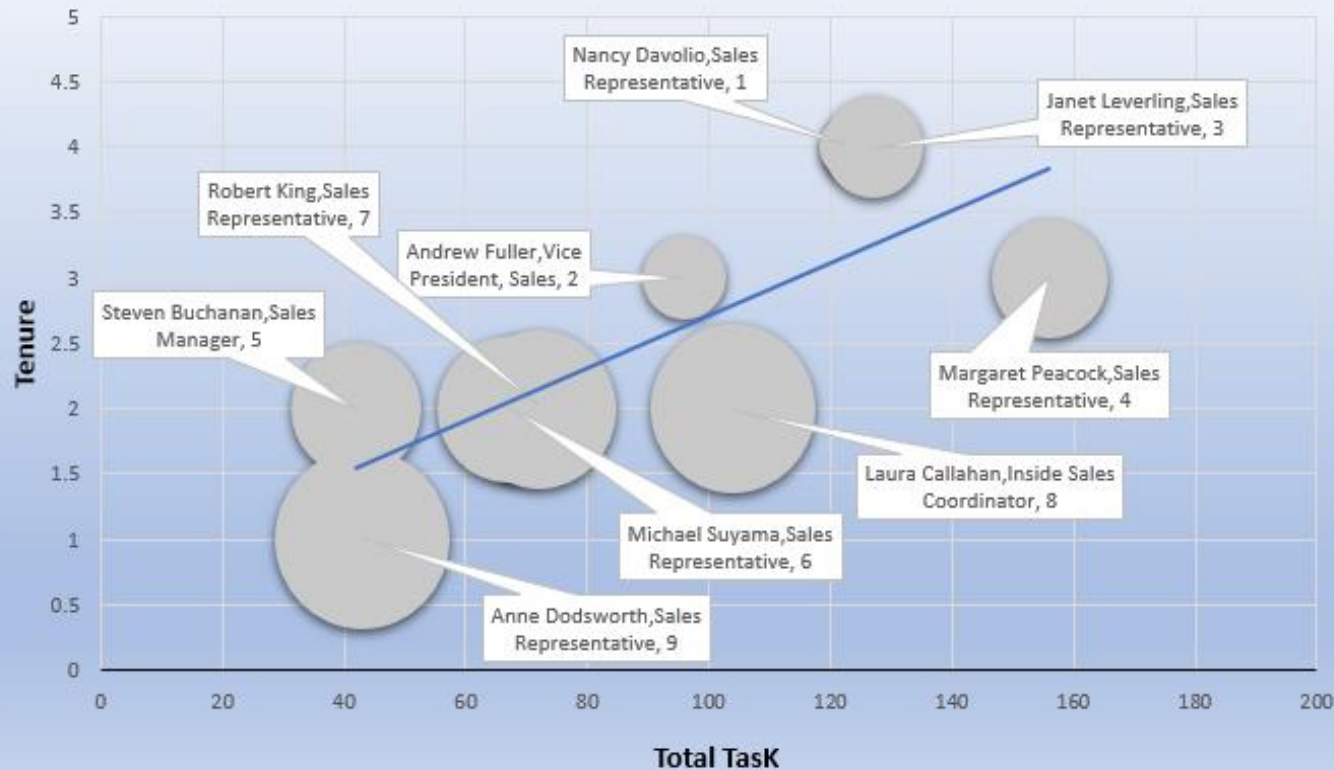
```
create view Emp_performance as
SELECT
e.employeeid,concat(e.firstName," ",e.LastName,",",e.title) as Full_Name_with_Title,
count(distinct o.OrderID) as total_tasks,timestampdiff(year,hiredate,maxD)as tenure,
round(sum(od.unitprice * od.Quantity*(1-od.discount)),2) as total_sale
FROM Employees e
left join (select employeeId,max(orderdate) as maxD from orders
group by 1)m on m.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
```

employeeid	Full_Name_with_Title	total_tasks	tenure	total_sale
1	Nancy Davolio, Sales Representative	123	4	192107.6
2	Andrew Fuller, Vice President, Sales	96	3	166537.75
3	Janet Leverling, Sales Representative	127	4	202812.84
4	Margaret Peacock, Sales Representativ	156	3	232890.85
5	Steven Buchanan, Sales Manager	42	2	68792.28
6	Michael Suyama, Sales Representative	67	2	73913.13
7	Robert King, Sales Representative	72	2	124568.23
8	Laura Callahan, Inside Sales Coordinato	104	2	126862.28
9	Anne Dodsworth, Sales Representative	43	1	77308.07

The sales representatives seem to be performing well, with an average of 99.8 total tasks completed and an average of $160,153.06 in total sales.
Sales representative number 4 is the highest performing salesperson, with 156 tasks completed and $232,890.85 in total sales. Sales representatives number 1 and 3 are also high performers, with 123 and 127 tasks completed, and $192,107.60 and $202,812.84 in total sales, respectively.
However, there is some variability in performance, with sales representative number 9 completing only 43 tasks and generating $77,308.07 in total sales

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?

Correlation b/w Units In Stock and Total Sold

ProductName	UnitsInStock	TotalSold
Camembert Pierrot	19	1577
Raclette Courdavault	79	1496
Gorgonzola Telino	0	1397
Gnocchi di nonna Alice	21	1263
Pavlova	29	1158
Rhönbräu Klosterbier	125	1155
Guaraná Fantástica	20	1125
Boston Crab Meat	123	1103
Tarte au sucre	17	1083
Chang	17	1057
Fløtemysost	26	1057
Sir Rodney's Scones	3	1016
Jack's New England Cla	85	981
Lakkalikööri	57	981
Alice Mutton	0	978
Pâté chinois	115	903
Konbu	24	891
Manjimup Dried Apple	20	886
Steeleye Stout	20	883
Chai	39	828
Outback Lager	15	817
Mozzarella di Giovanni	14	806
Inlagd Sill	112	805
Scottish Longbreads	6	799
Chartreuse verte	69	793
Original Frankfurter gri	32	791
Uncle Bob's Organic Dr	15	763
Geitost	112	755

```
create view prouduct_attribute as

SELECT p.ProductName, p.UnitsInStock, SUM(od.Quantity) AS TotalSold

FROM Products p

JOIN `Order Details` od ON p.ProductID = od.ProductID

GROUP BY 1,2

ORDER BY TotalSold DESC
```

The data suggests that there is a Weak correlation between the number of units in stock and the total number of units sold. This means that more product units in stock tend to sell more units overall. This could be because these are more popular with customers, or because the store is not able to promote them more heavily.

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

Products demand Flucuate over Different Months

ProductName list: Alice Mutton, Aniseed Syrup, Boston Crab Meat, Camembert Pierrot, Carnarvon Tigers, Chai, Chang, Chartreuse verte, Chef Anton's Cajun S..., Chef Anton's Gumbo..., Chocolade, Côte de Blaye, Escargots de Bourgo..., Filo Mix, Fløtemysost, Geitost, Genen Shouyu
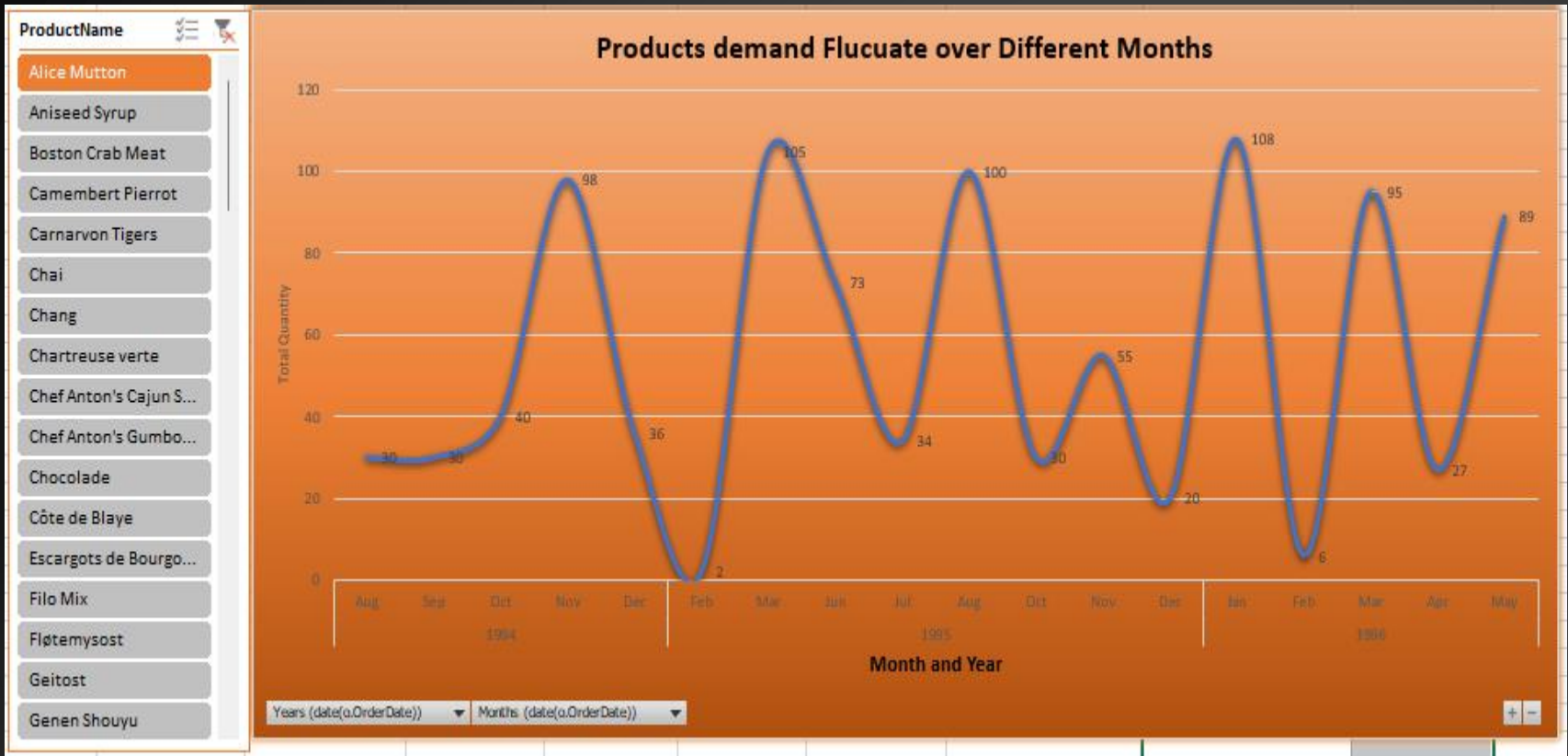
```
create view product_demand as
SELECT
    date(o.OrderDate) ,
    p.ProductID,p.ProductName,
    SUM(Quantity) AS TotalQuantity
FROM Orders o
left JOIN `Order Details`od ON o.OrderID=od.OrderID
left join products as p on p.ProductID=od.ProductID
left join categories c on c.CategoryID=p.CategoryID
GROUP BY 1,2,3
ORDER BY 1,2,3
```

Row Labels	Sum of TotalQuantity
⊟1994	234
Aug	30
Sep	30
Oct	40
Nov	98
Dec	36
⊟1995	419
Feb	2
Mar	105
Jun	73
Jul	34
Aug	100
Oct	30
Nov	55
Dec	20
⊟1996	325
Jan	108
Feb	6
Mar	95
Apr	27
May	89
Grand Total	**978**

Overall demand increased in 1995: The total quantity of the product sold in 1995 (25,007) was almost three and a half times higher than the total quantity sold in 1994 (7,381). This suggests a significant increase in demand for the product over that period.
Demand fluctuated throughout the year: Even within years, demand fluctuated considerably. In 1994, December saw the highest demand (1,804 units), while August saw the lowest (1,462 units). This pattern is repeated in 1995, with November being the strongest month (2,657 units) and July being the weakest (1,635 units).

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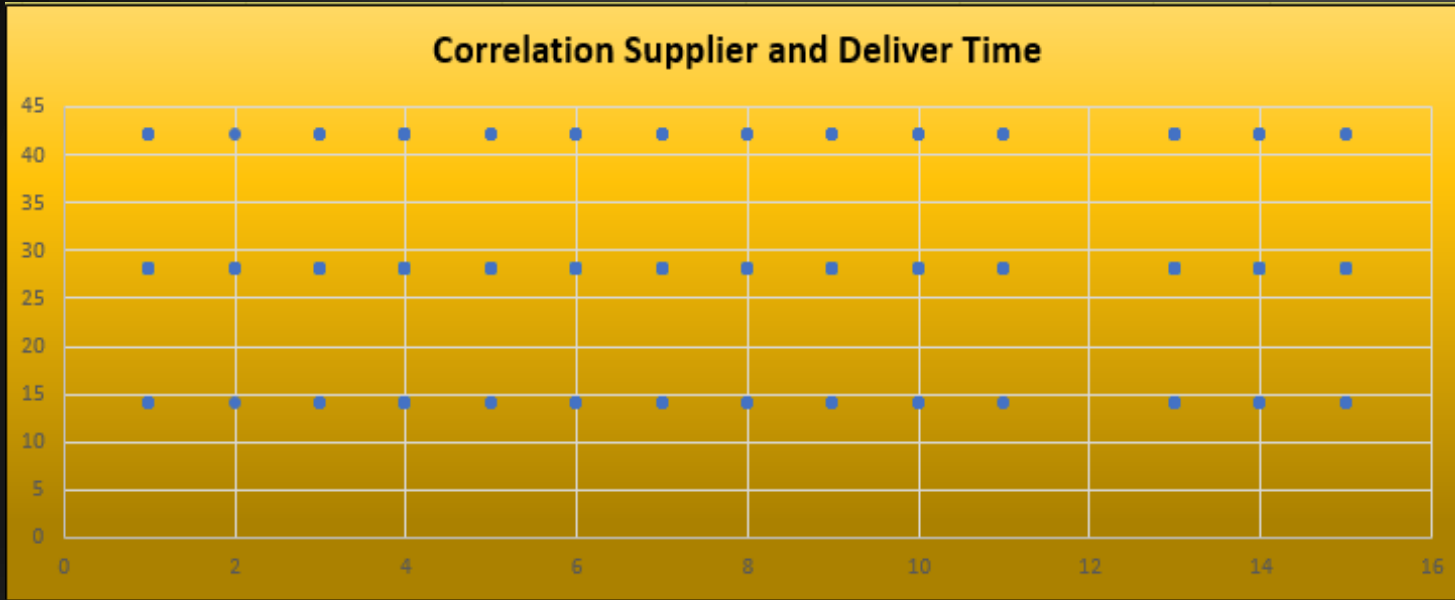

This give us visualization to identify outliers in sale performance of the products and we can also analyze product with Highest sales and lowest sales. From this we can compare the average price of the product with other for each category

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

Row Labels	Total Quantity	Average UnitPrice
Beverages	9532	35.58333333
Chai	828	17
Chang	1057	18
Chartreuse verte	793	17
Côte de Blaye	623	246
Guaraná Fantástica	1125	4
Ipoh Coffee	580	43
Lakkalikööri	981	17
Laughing Lumberjack Lager	184	14
Outback Lager	817	14
Rhönbräu Klosterbier	1155	7
Sasquatch Ale	506	13
Steeleye Stout	883	17
Grand Total	9532	35.58333333

```sql
create view product_performance as
SELECT
    p.ProductName,c.CategoryName,
    SUM(od.Quantity) AS TotalQuantity,
    round(avg(od.Unitprice),0) as avg_UnitPrice FROM products p
left JOIN `Order Details`od ON p.ProductID=od.ProductID
left join categories c on c.CategoryID=p.CategoryID
GROUP BY 1,2
ORDER BY 1,2,3
```

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?

Correlation Supplier and Deliver Time


```
create view supplier_performace as
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
```

ContactCod	countrycode	OrderID	Delivery_Time
3	8	10248	28
7	8	10248	28
13	8	10248	28
5	9	10249	42
13	9	10249	42
15	4	10250	28
13	4	10250	28
6	4	10250	28
11	8	10251	28
6	8	10251	28
13	3	10252	28
4	3	10252	28
13	4	10253	14
14	4	10253	14
8	4	10253	14
4	18	10254	28
9	18	10255	28
4	18	10255	28
13	18	10255	28
13	4	10256	28
4	4	10256	28
14	21	10257	28
4	21	10257	28

Correlation Coefficient (Contact Code vs. delivery Time):
The off-diagonal value of 0.001354 is the correlation coefficient between Contact Code and delivery Time . A correlation coefficient close to zero indicates a weak or negligible linear relationship between the two variables. In this case, the correlation coefficient is negative, but the magnitude is very small.

# Correlation Between Ship Country/Country code and Delivery Time

Correlation b/w Country and Delivery Time

```
create view supplier_performace  as
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
```

Correlation Coefficient (country code vs delivery time) is 0.01319.
The positive sign indicates a positive correlation between country code and delivery Time.
However, the correlation coefficient is relatively small , suggesting a weak positive linear relationship

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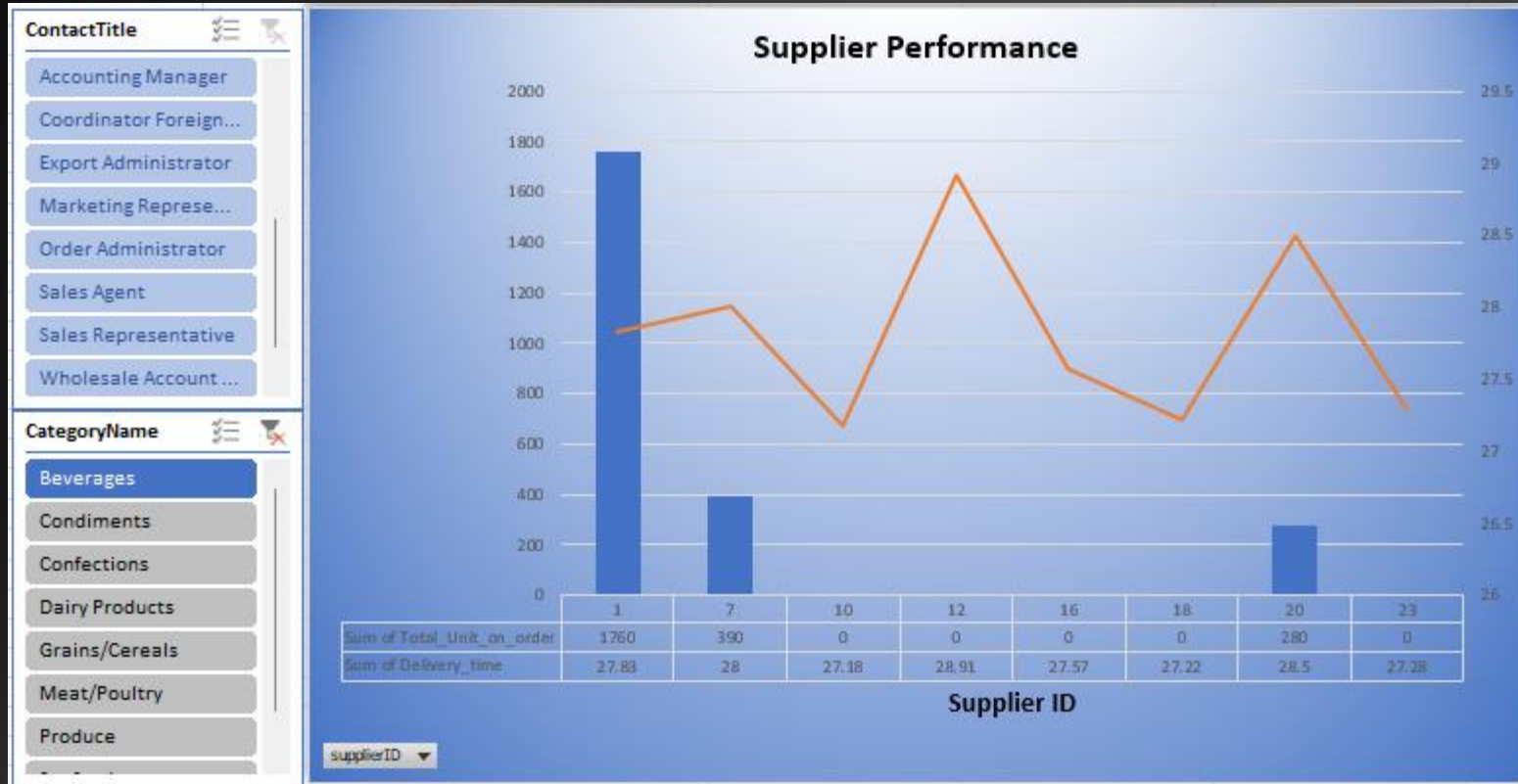

```sql
create view supply_performance as
SELECT
distinct s.supplierID,s.ContactTitle ,c.CategoryName, SUM(od.Quantity) AS TotalQuantity,
round(AVG(timestampdiff(day, OrderDate, Requireddate)),2) AS Delivery_time FROM suppliers s
LEFT JOIN products AS p ON p.SupplierID = s.SupplierID
LEFT JOIN `order 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
```

Supplier Performance chart with Supplier ID on x-axis:

	1	7	10	12	16	18	20	23
Sum of Total_Unit_on_order	1760	390	0	0	0	0	280	0
Sum of Delivery_time	27.83	28	27.18	28.91	27.57	27.22	28.5	27.28

Row Labels	Sum of Total_Unit_on_order	Average of Delivery_time
1	1760	27.83
7	390	28
10	0	27.18
12	0	28.91
16	0	27.57
18	0	27.22
20	280	28.5
23	0	27.28
Grand Total	2430	27.81125

This query provide the supplier ID ,contact tile, category name ,total Unit on order, average delivery time by suppliers . For Category Beverages total unit on order is 1760 and average delivery time by the supplier id 1 is 27 days and supplier id 2 provide average deliver time of 28 for 390 total unit on order and supplier id 20 provide average deliver time of 28.5 days for 280 units . By this we can compare the average delivery time provide by the suppliers to check the supplier performance for different category and Job Roles.

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?

CategoryID
1
2
3
4
5
6
7
8

Supplier Cost Structure

```
Refrescos Americanas LTDA    4.5
Plutzer Lebensmittelgroßmärkte AG    7.75
Pavlova, Ltd.    15
Leka Trading    46
Karkki Oy    18
Exotic Liquids    18.5
Bigfoot Breweries 15.33333333
Aux joyeux ecclésiastiques    140.75
```

Average Unit Price in Each Category

```sql
create view cost_pattern as

SELECT s.CompanyName, p.CategoryID, AVG(p.UnitPrice) AS AvgUnitPrice

FROM Products p

JOIN Suppliers s ON p.SupplierID = s.SupplierID

GROUP BY s.CompanyName, p.CategoryID
```

Row Labels	UnitPrice
Aux joyeux ecclésiastiques	140.75
Bigfoot Breweries	15.333333
Exotic Liquids	18.5
Karkki Oy	18
Leka Trading	46
Pavlova, Ltd.	15
Plutzer Lebensmittelgroßmärkte AG	7.75
Refrescos Americanas LTDA	4.5
Grand Total	33.229167

Suppliers with higher vs lower average pricing overall.
Categories where certain suppliers have exceptionally high prices.
Comparing pricing between suppliers for the same category .
Based on Analysis: Negotiate lower pricing in categories where a supplier's prices

are consistently high Switch suppliers for certain categories to reduce costs.

Conclusion:

By providing a visually compelling and user-friendly dashboard with interactive features, this Power BI report empowers stakeholders at Northwind Traders to make data-driven decisions. This comprehensive analysis of customers, sales, inventory, and employees offers valuable insights into business operations and facilitates data exploration. The expected impact is a transformation in how Northwind interacts with its data, ultimately enhancing competitiveness and driving the company forward in the wholesale market.