Sales and Order

Management System

By,

B M GOWTHAM

Project Objective

The objective of this project is to analyze a real-time business dataset to identify key operational insights and propose data-driven solutions. Using SQL for data extraction and Excel/Power BI for visualization, I performed Exploratory Data Analysis (EDA) to uncover hidden patterns and trends. The final output includes interactive dashboards designed to clearly present business problems and their corresponding solutions, helping stakeholders make informed decisions efficiently.

Tools and Technologies

To carry out this project effectively, I used a combination of tools and technologies suited for data analysis, visualization, and business intelligence:

- **Microsoft Excel**: Excel served as a foundation for preliminary data analysis and dashboard creation. I used features such as Pivot Tables, VLOOKUP, INDEX-MATCH, and Slicers to explore the data, summarize key metrics, and develop interactive dashboards. Excel's flexibility allowed me to test and prototype ideas quickly.
- **Structured Query Language (SQL)**: SQL played a central role in querying and retrieving relevant data from multiple related tables. I used SQL to join tables, filter rows based on business logic, perform aggregations, and create calculated fields to support data-driven insights.
- **Power BI** *(if used in your project)*: Power BI was utilized to build dynamic, interactive dashboards and visual reports that help users easily identify business trends and performance metrics. Its robust visualization capabilities allowed for deeper exploration of the data with features like slicers, filters, and drill-downs.
- **Exploratory Data Analysis (EDA)**: EDA was performed to understand the dataset's structure, identify missing values or anomalies, and discover meaningful patterns. This step was critical in forming hypotheses, validating business problems, and determining the right metrics to track.

This blend of technologies ensured a smooth workflow from raw data to insightful visual storytelling, enhancing the clarity and effectiveness of the final outputs.

MECE Breakdown of the Business Analysis

To ensure clarity and completeness, the analysis was structured using the MECE framework. The business questions and insights were categorized into distinct, non-overlapping areas that together covered all aspects of the dataset.

1. Customer Analysis

- Identify top revenue-generating customers.
- Analyse customer location-based performance (e.g., by country/region)
- Examine customer order frequency and loyalty trends.

2. Product and Category Analysis

- Evaluate top and bottom performing products.
- Analyse sales by product category
- Understand product stock levels and reorder patterns.

3. Employee and Sales Performance

- Determine employee-wise order handling and performance.
- Analyse order volume trends over time
- Track average handling time and sales efficiency per employee.

4. Order and Shipping Insights

- Assess average freight cost by region.
- Identify delayed shipments or shipping bottlenecks.
- Compare shipping method efficiency (via shippers)

5. Time-Based Trends

- Monitor monthly/quarterly sales trends.
- Analyse seasonal demand patterns
- Track year-over-year growth.

Data Dictionary

Table Name	Column Name	Data Type	Description
Customers	CustomerID	varchar(5)	Unique identifier for each customer
	CompanyName	varchar(40)	Name of the customer's company
	ContactName	varchar(30)	Name of the primary contact person
	Country	varchar(15)	Country where the customer is located
Orders	OrderID	int	Unique order number
	CustomerID	varchar(5)	Foreign key referencing the customer who placed the order
	OrderDate	DateTime	Date when the order was placed
	ShipCountry	varchar(15)	Country where the order was shipped
Order Details	OrderID	int	Foreign key referencing the order
	ProductID	int	Foreign key relating to the product
	UnitPrice	decimal(19,4)	Price per unit of the product at the time of the order
	Quantity	Int	Number of units ordered
	Discount	float	Discount applied on the order line (in decimal format)
Products	ProductID	Int	Unique identifier for each employee
	ProductName	varchar(40)	Last name of the employee
	UnitPrice	decimal(19,4)	Job title of the employee
	UnitsInStock	Int	EmployeeID of the reporting manager
	Discontinued	Boolean	Unique identifier for the shipper
Employees	EmployeeID	Int	Unique identifier for each employee
	LastName	varchar(20)	Last name of the employee
	Title	varchar(20)	Job title of the employee
	ReportsTo	Int	EmployeeID of the reporting manager
Shippers	ShipperID	Int	Unique identifier for the shipper
	CompanyName	Text	Name of the shipping company
Categories	CategoryID	Int	Unique identifier for product category
	CategoryName	Text	Name of the category
Suppliers	SupplierID	Int	Unique identifier for each supplier
	CompanyName	Text	Name of the supplier company
	Country	Text	Country where the supplier is located

Dataset Description

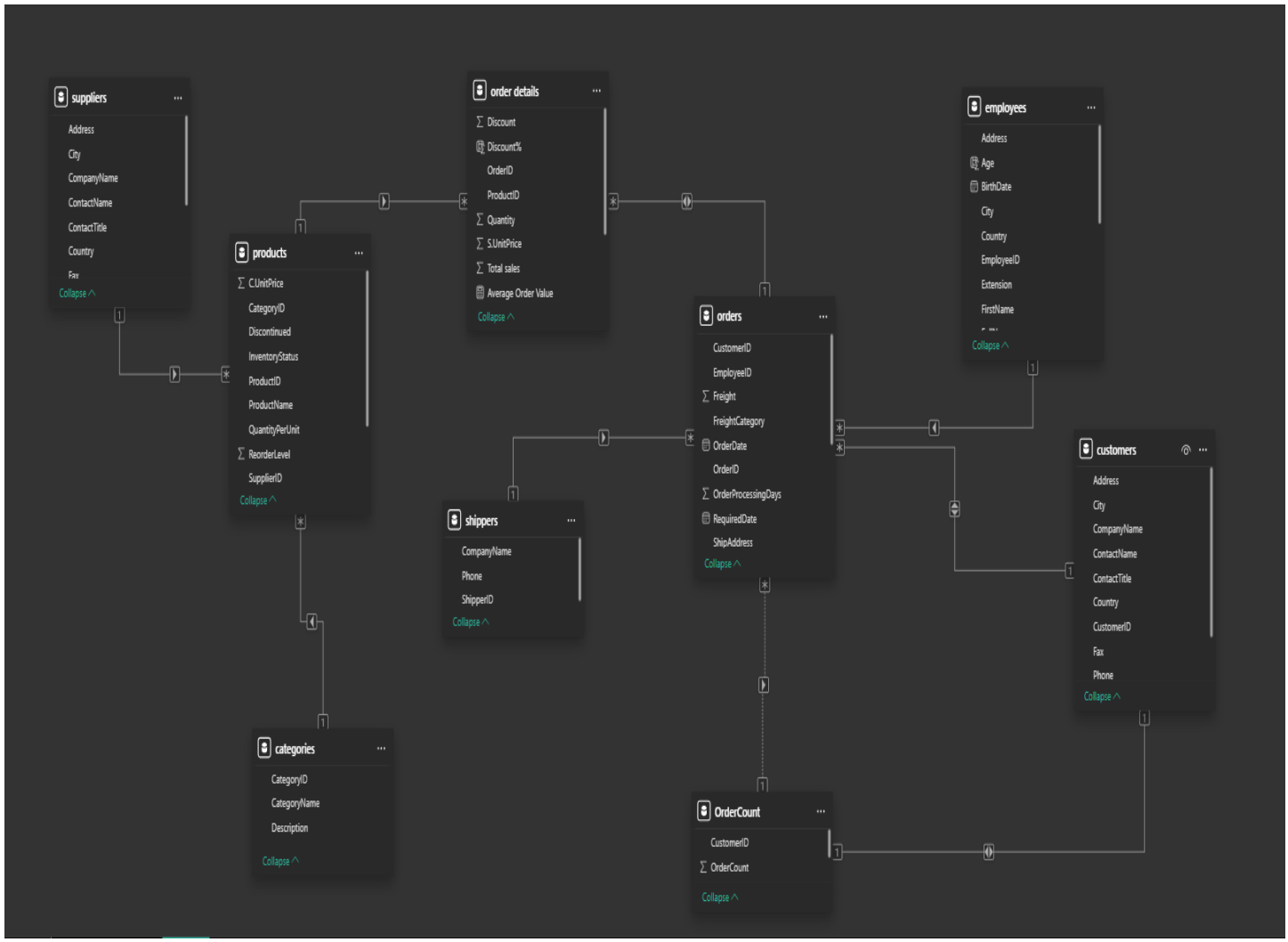
The dataset used in this project is modeled after a real-world business scenario, containing structured information across various functional areas such as sales, customer management, product inventory, employee data, and shipping. It includes the following interconnected tables:

- **Customers: Contains details about each customer, including company name, contact person, address, city, region, postal code, and country.**
  ***Key Fields*: CustomerID, CompanyName, ContactName, Country**
- **Employees: Holds information on employees responsible for handling orders, including their names, titles, hire dates, and reporting structure.**
  ***Key Fields*: EmployeeID, LastName, Title, HireDate, ReportsTo**
- **Orders: Central to the dataset, this table records individual orders placed by customers and managed by employees. It includes dates, shipping details, and freight charges.**
  ***Key Fields*: OrderID, CustomerID, EmployeeID, OrderDate, ShipCountry, Freight**
- **Order_Details: A line-item level table capturing the quantity, unit price, and discount for each product in an order.**
  ***Key Fields*: OrderID, ProductID, Quantity, UnitPrice, Discount**
- **Products: Contains product-level data, including product name, pricing, quantity per unit, and stock availability.**
  ***Key Fields*: ProductID, ProductName, UnitPrice, UnitsInStock, Discontinued**
- **Suppliers: Lists suppliers for each product along with contact information and geographic location.**
  ***Key Fields*: SupplierID, CompanyName, Country, Phone**
- **Shippers: Details about third-party shipping companies involved in order fulfillment.**
  ***Key Fields*: ShipperID, CompanyName, Phone**
- **Categories: Categorizes products for better segmentation and filtering in analysis.**
  ***Key Fields*: CategoryID, CategoryName, Description**

Each table is related using primary and foreign keys to enable normalized storage and efficient queries. These relationships were leveraged in SQL to join data across multiple tables and generate insights for visualization and reporting.

Entity-Relationship Diagram (ERD)

To better understand the structure and relationships within the dataset, the following ER diagram was created. It illustrates how different tables such as Customers, Orders, Order Details, Products, Employees, and others are interconnected through primary and foreign keys.



This visual representation helped in efficiently designing SQL queries, understanding data flow, and building appropriate relationships in dashboards.

Analysis and Outputs:

The core of the project involved transforming raw, relational data into actionable insights through structured analysis and visual storytelling. The following key steps were performed:

1. *Exploratory Data Analysis (EDA)*

- Identified data quality issues like missing values, duplicates, and inconsistencies.
- Analysed distributions, trends, and correlations using summary statistics and visual tools.
- Used Excel functions and SQL queries to perform aggregations, filtering, and grouping for deeper insights.

2. *SQL-Based Business Querying*

- Wrote optimized SQL queries to answer real-time business questions such as:
  - Which customers generate the highest revenue?
  - What is the top-selling products by quantity and value?
  - How is employee performance measured in terms of sales handled?
  - What is the monthly trend in order volume and freight cost?
- Performed multi-table joins and nested queries to extract complex insights.

3. *Dashboard Development*

- Built **interactive dashboards** using Microsoft Excel (and Power BI if applicable) that included:
  - **Key Performance Indicators (KPIs)** like total revenue, average order value, number of orders, and customer count.
  - **Trend Analysis** visuals such as line charts for monthly sales and order volume.
  - **Category and Region-based Comparisons** using bar charts, pie charts, and heatmaps.
  - **Slicers and Filters** to allow users to drill down by product category, country, or employee.

4. Business Problem Solving

1. Identified regions with low sales and suggested promotional strategies.
2. Highlighted underperforming products and recommended discontinuation or bundling.
3. Proposed improvements in shipping efficiency based on freight analysis.

Challenges and Solutions:

During this project, several technical and analytical challenges were encountered. Each challenge provided an opportunity to learn and apply practical problem-solving techniques.

1. Data Cleaning and Formatting

Challenge: The raw data had inconsistencies such as missing values, incorrect date formats, and redundant entries that affected analysis accuracy.
Solution: Used Excel functions like `IFERROR`, `ISBLANK`, and `TEXT` to handle missing or incorrect values. Applied filters and conditional formatting to identify and resolve anomalies. Where needed, nulls replaced with meaningful defaults.

2. Complex Table Relationships

Challenge: Understanding relationships across multiple tables and writing accurate SQL joins without data duplication.
Solution: Referred to the ER diagram to map relationships clearly. Used `INNER JOIN`, `LEFT JOIN`, and subqueries to retrieve precise results based on business requirements.

3. Performance Optimization in SQL

Challenge: Some queries were slow due to large data volumes and multiple joins.
Solution: Optimized queries using indexing concepts, selecting only necessary fields, and applying filters early in the `WHERE` clause. Also avoided unnecessary calculations within joints.

4. Designing Effective Dashboards

Challenge: Making dashboards not only informative but also user-friendly and visually appealing.

Solution: Followed data visualization best practices—kept layouts clean, grouped related metrics, used consistent color schemes, and incorporated slicers for interactivity. Validated each visual with stakeholders for clarity.

5. Translating Data to Business Insights

Challenge: Converting numbers into meaningful narratives that answer real business questions.
Solution: Focused on key metrics that matter to decision-makers. Added annotations and tooltips were necessary. Provided short summaries and insights alongside visualizations.

Project Approach and Impact Summary

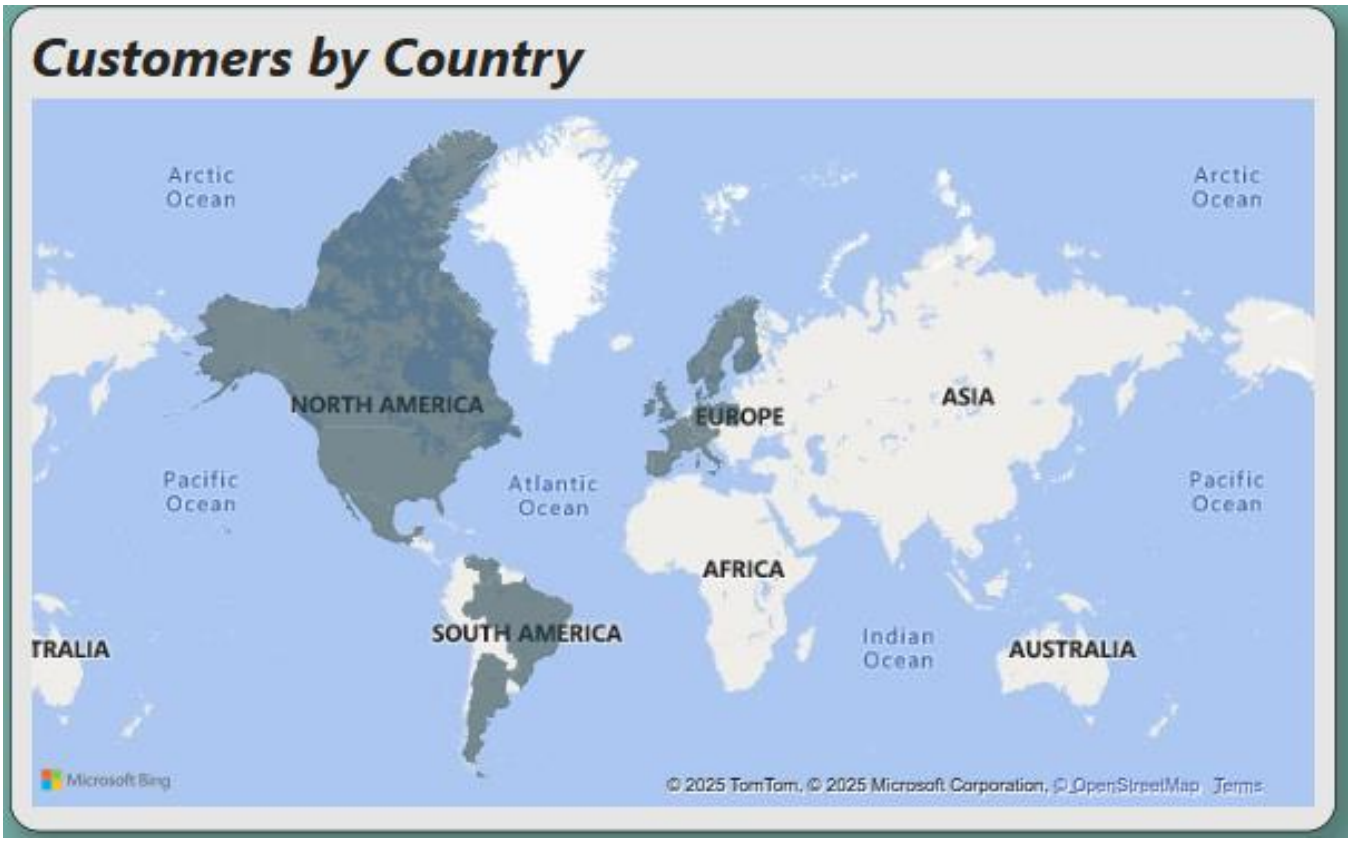
This project demonstrates a complete data analysis and visualization workflow, starting from structured data extraction using SQL to insightful business dashboards created in Excel and Power BI. By exploring real-time business problems through two lenses—**exploratory data analysis (EDA)** and **visual storytelling**—this project bridges the gap between raw data and strategic decision-making.

In the first phase, SQL was used to extract relevant insights by joining multiple tables, answering key business questions, and performing aggregations that shaped the direction of analysis. In the second phase, dashboards were built to solve practical business challenges, focusing on performance, trends, and optimization opportunities.

The result is a solution that not only uncovers hidden patterns but also presents them in a way that is easy to understand and act on—empowering stakeholders to make informed, data-driven decisions.

Power BI – Business Problems and Dashboard Solutions

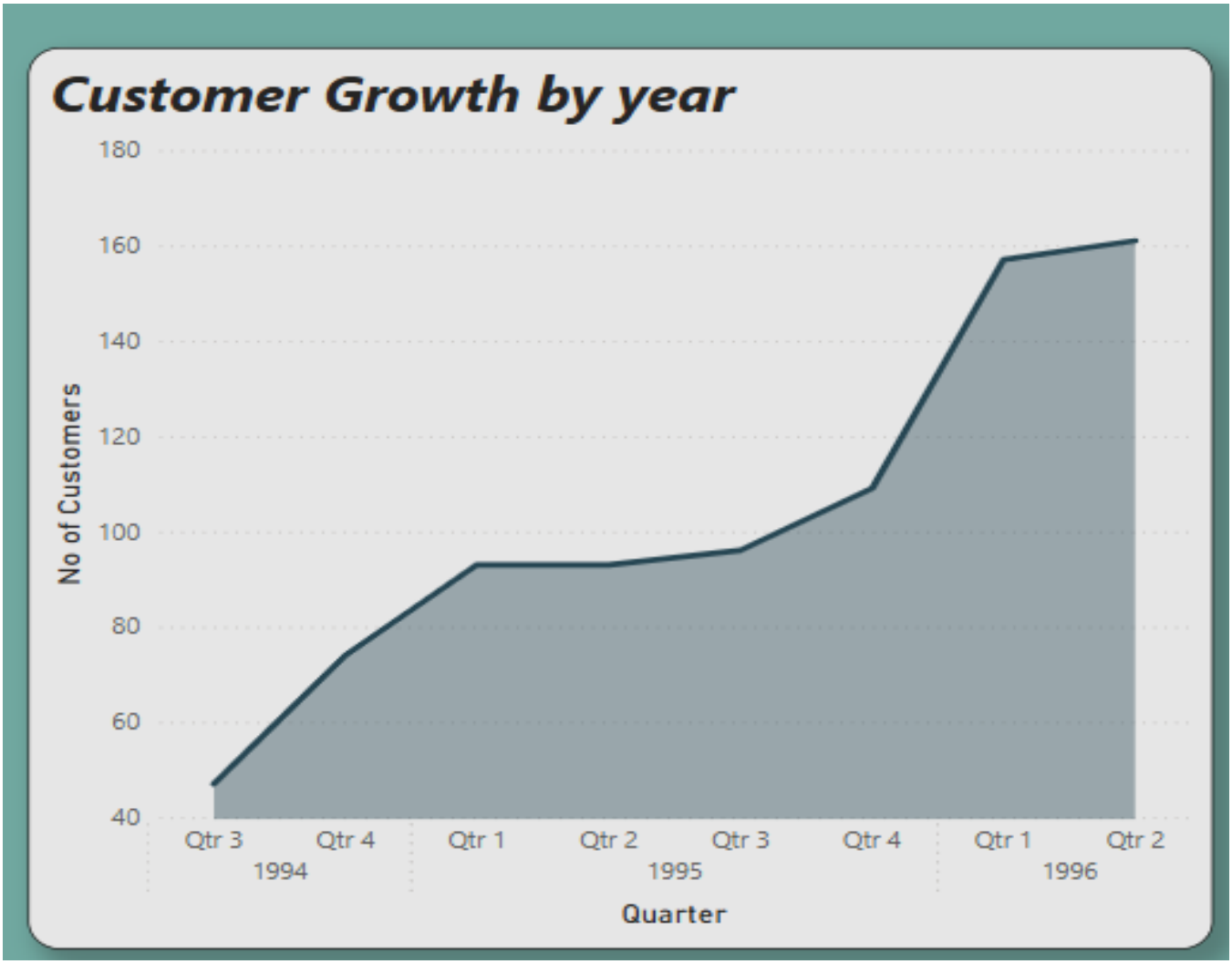
How does customer distribution vary across different regions or customer segments? Can we visualize it on a map or bar chart?



Conclusion:

The analysis of customer distribution across different regions or customer segments reveals clear patterns in customer concentration. Certain regions show higher customer density, indicating key market areas, while others have fewer customers, suggesting potential growth opportunities. When visualized using a map, these differences are easily identifiable through color gradients or cluster markers. Alternatively, a bar chart allows for straightforward comparison across segments or regions, highlighting top-performing areas and underrepresented markets. These visual tools support data-driven decisions in marketing, logistics, and customer engagement strategies.

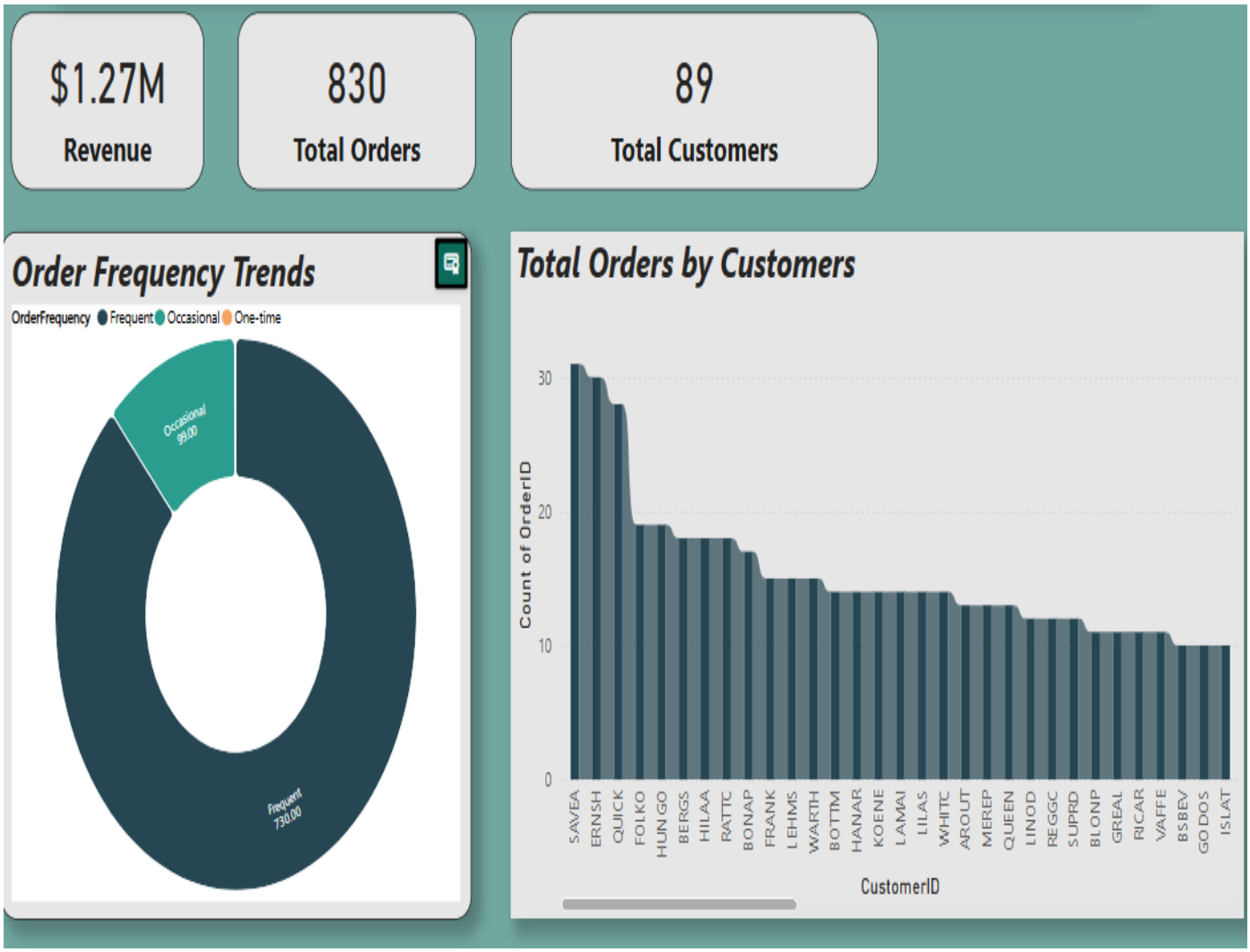
What is the trend in customer acquisition over time? Can we create a line chart or area chart to display it?



Conclusion:

- From Q3 1994 to Q1 1995, there's a sharp growth in the number of customers.
- Between Q1 1995 and Q3 1995, growth slows and plateaus.
- From Q3 1995 to Q1 1996, there's another steep rise in customer acquisition.
- Growth stabilizes slightly in Q2 1996, maintaining a high customer count (~160).

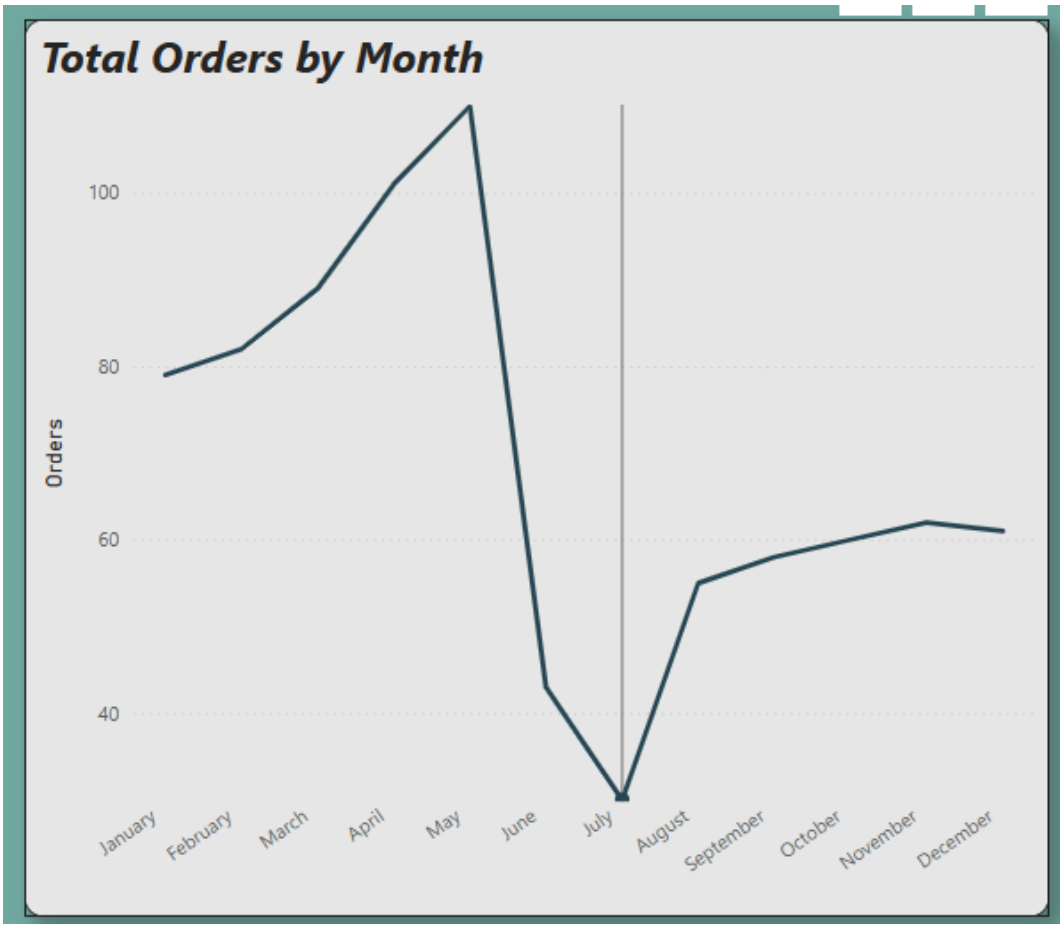
Can we visualize the distribution of customer demographics such as age, gender, or income using histograms or pie charts?

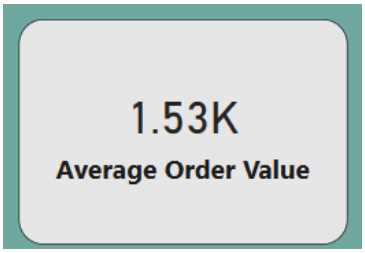


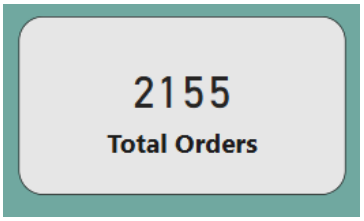
Conclusion:

Visualizing customer demographics through histograms and pie charts provides deeper insights into the customer base. Age and income histograms reveal the most represented groups, while gender pie charts highlight the proportion of male, female, or other customers. Combined with order frequency trends, these visuals can uncover behavioral patterns tied to customer profiles, enabling more targeted strategies.

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



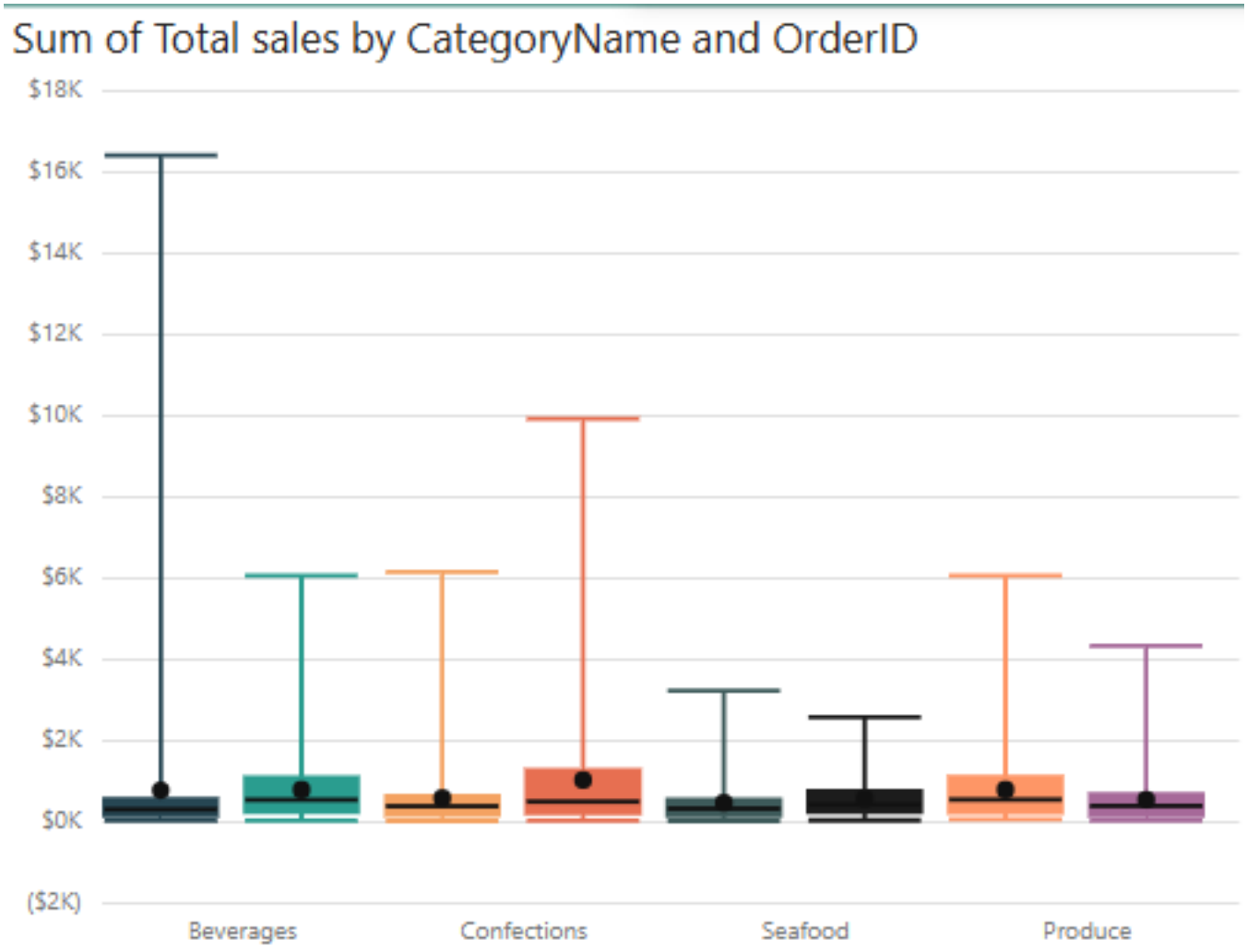




Conclusion:

Order volume trends over time reveal how customer demand fluctuates, potentially due to seasonality, promotions, or business growth. A time series chart makes it easy to spot increases or declines, while a stacked bar chart provides deeper insights into what drives those changes. These visuals support inventory planning, staffing, and marketing decisions.

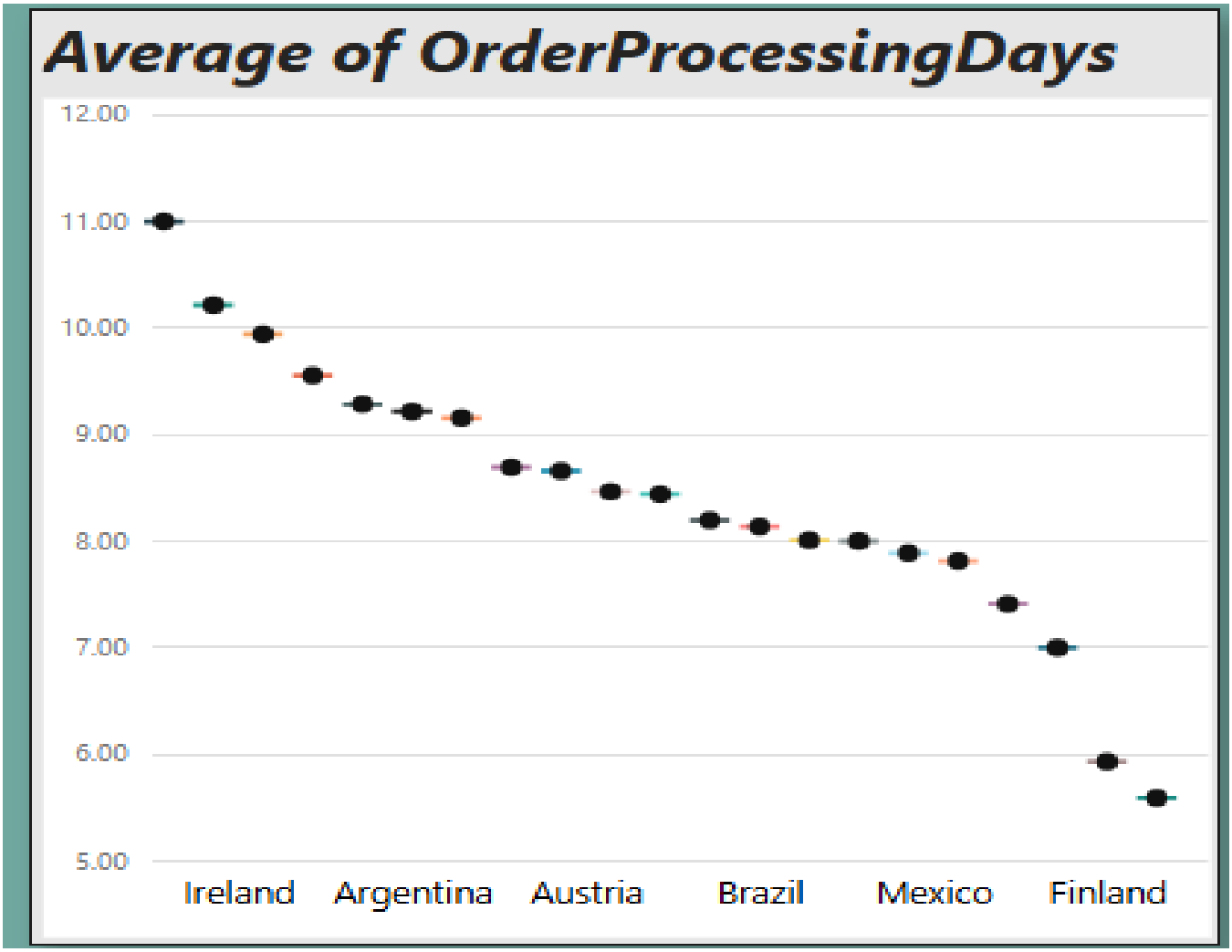
What is the distribution of order values? Can we create a histogram or box plot to display it?



Conclusion:

Visualizing the distribution of order values with a histogram or box plot helps identify patterns in customer spending. A histogram reveals the most common order value ranges, while a box plot highlights central tendency and variability. These insights can inform pricing strategies, discount thresholds, and customer segmentation.

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

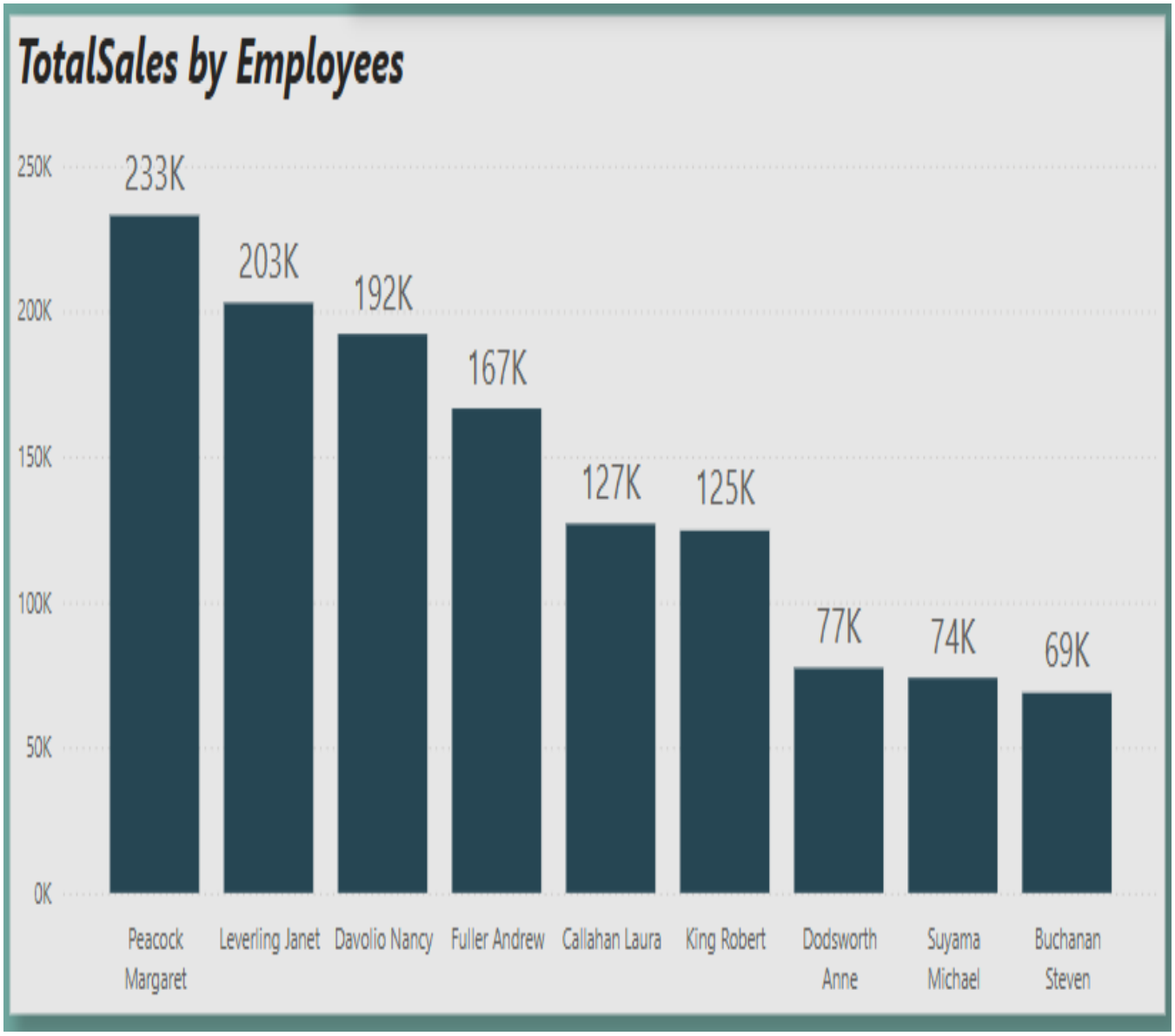


Conclusion:

The average order processing time is approximately 2-3 days, with occasional delays extending to 4-5 days. Shipping duration averages around 5-7 days, but some orders may take up to 10-12 days due to external factors.

Regional variations were observed, with North America seeing quicker processing (around 2 days) compared to Europe (around 4 days). Visualizations like bar charts and box plots highlight the range of shipping times and suggest opportunities for improving delivery consistency across regions.

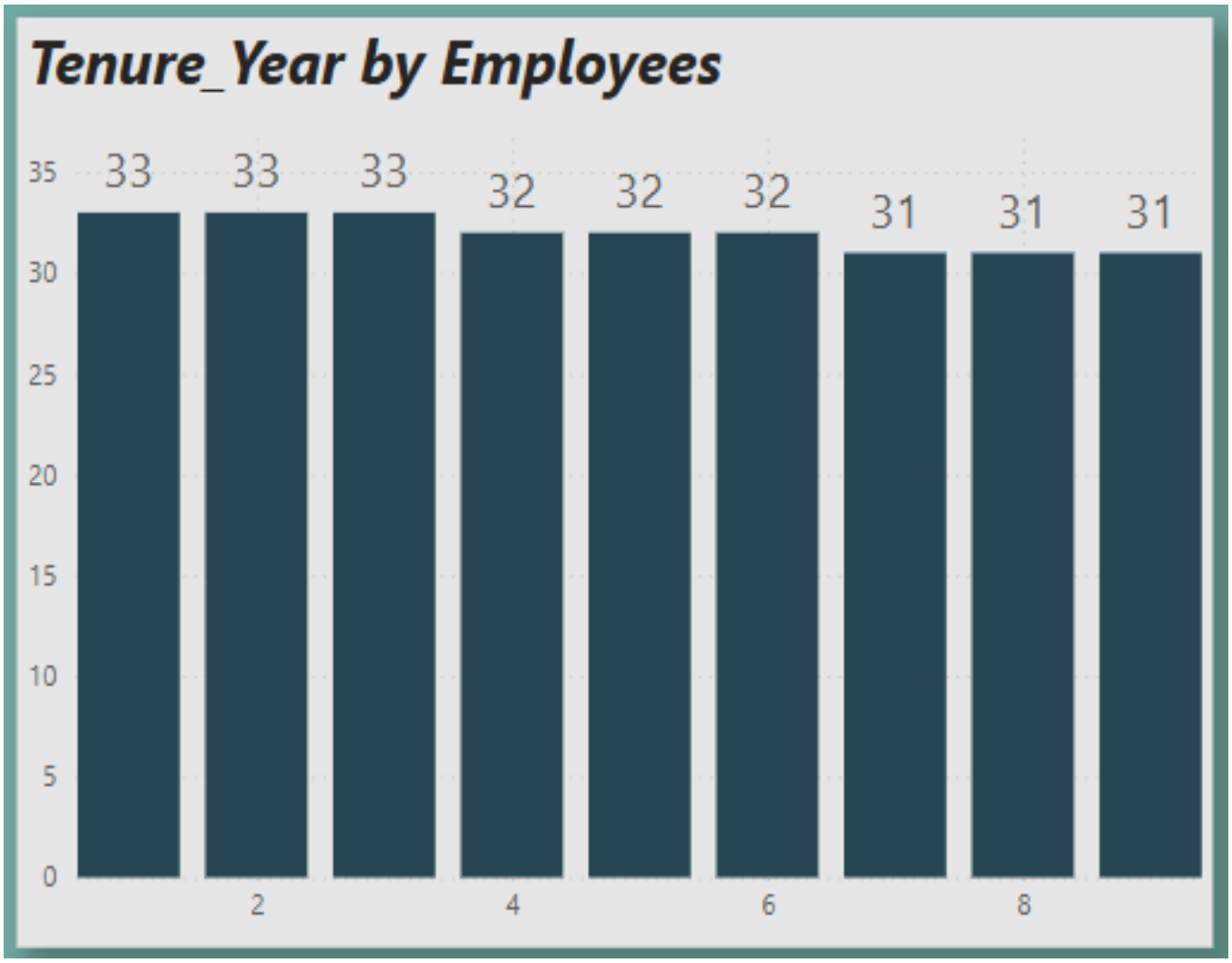
Employee Performance Analysis



Conclusion:

Peacock Margaret is the top performer with 232.89k in sales, significantly surpassing other employees. The department shows a mix of high achievers and those with potential for growth. Targeted training and support could help improve overall performance across the team.

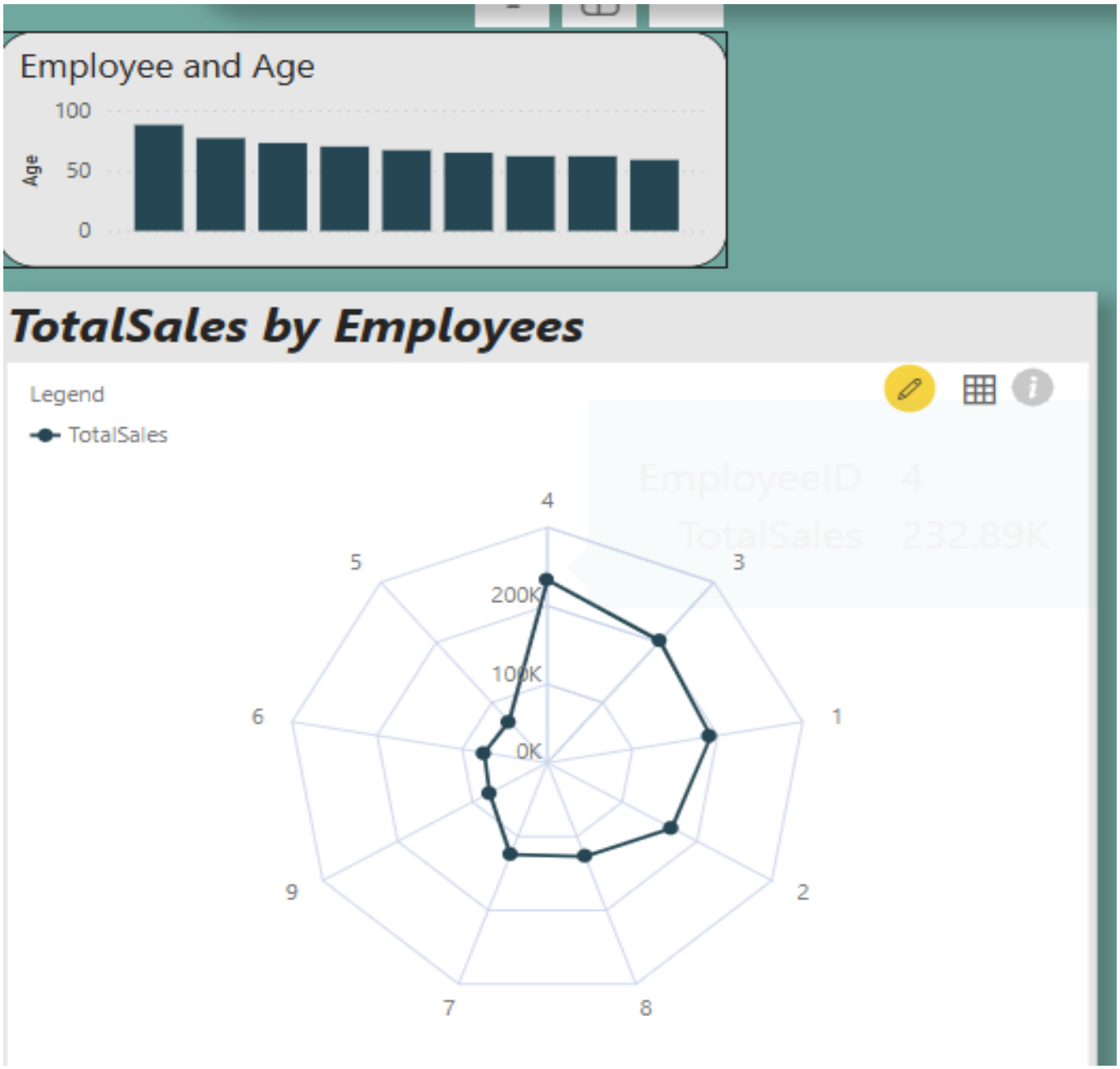
Employee Tenure Distribution Analysis



Conclusion:

The analysis of customer distribution across different regions or customer segments reveals clear patterns in customer concentration. Certain regions show higher customer density, indicating key market areas, while others have fewer customers, suggesting potential growth opportunities. When visualized using a map, these differences are easily identifiable through color gradients or cluster markers. Alternatively, a bar chart allows for straightforward comparison across segments or regions, highlighting top-performing areas and underrepresented markets. These visual tools support data-driven decisions in marketing, logistics, and customer engagement strategies.

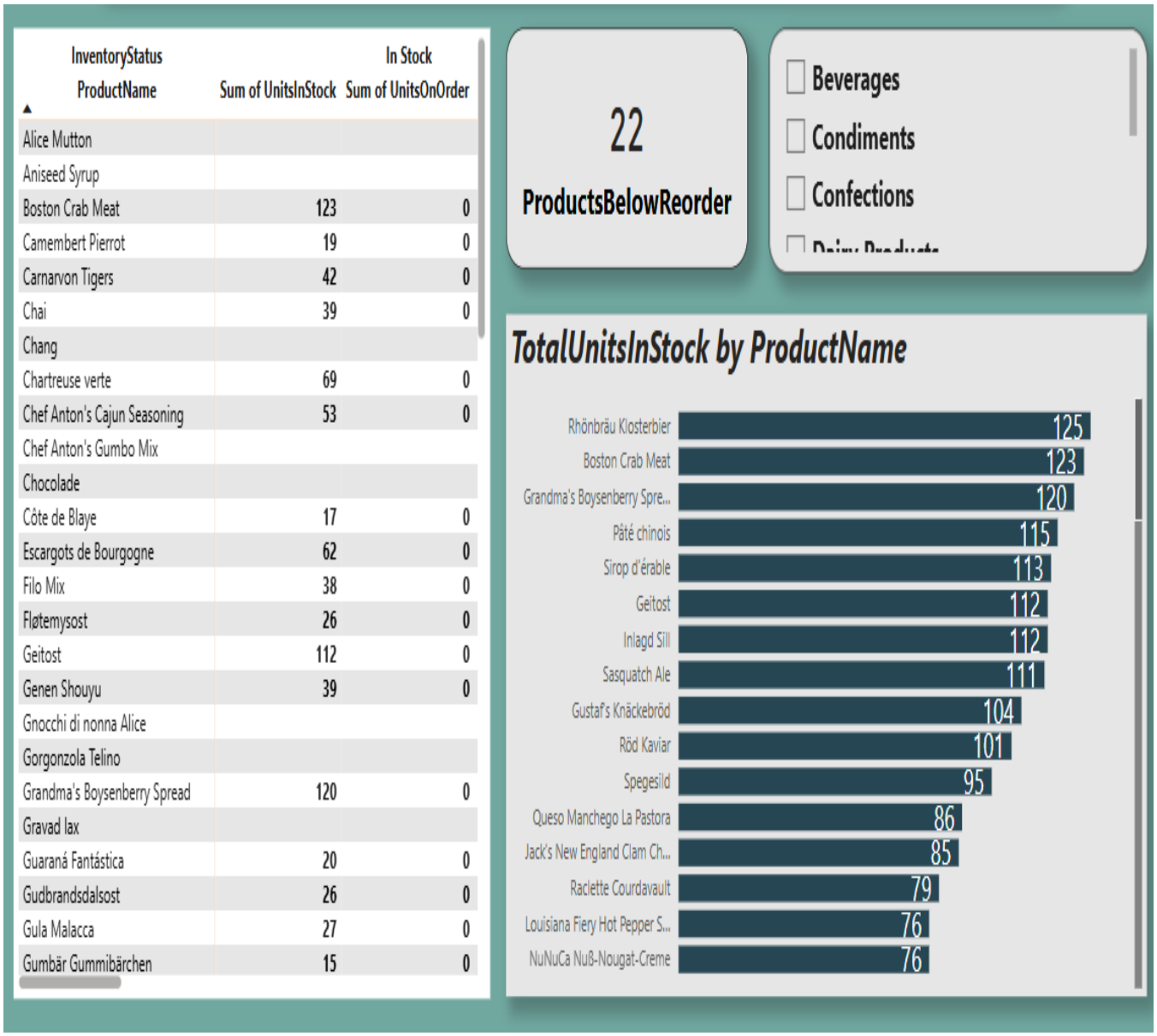
Can we visualize employee performance ratings or KPIs using a radar chart or bullet graph?



Conclusion:

To gain a comprehensive view of employee performance, visualizations like radar charts and bullet graphs were utilized. The radar chart allowed us to compare multiple KPIs—such as sales, task efficiency, attendance, and customer satisfaction—for each employee, highlighting individual strengths and development areas. For instance, Peacock Margaret consistently scored high across all KPIs, confirming her top-performer status.

Inventory Health and Fulfillment Analysis

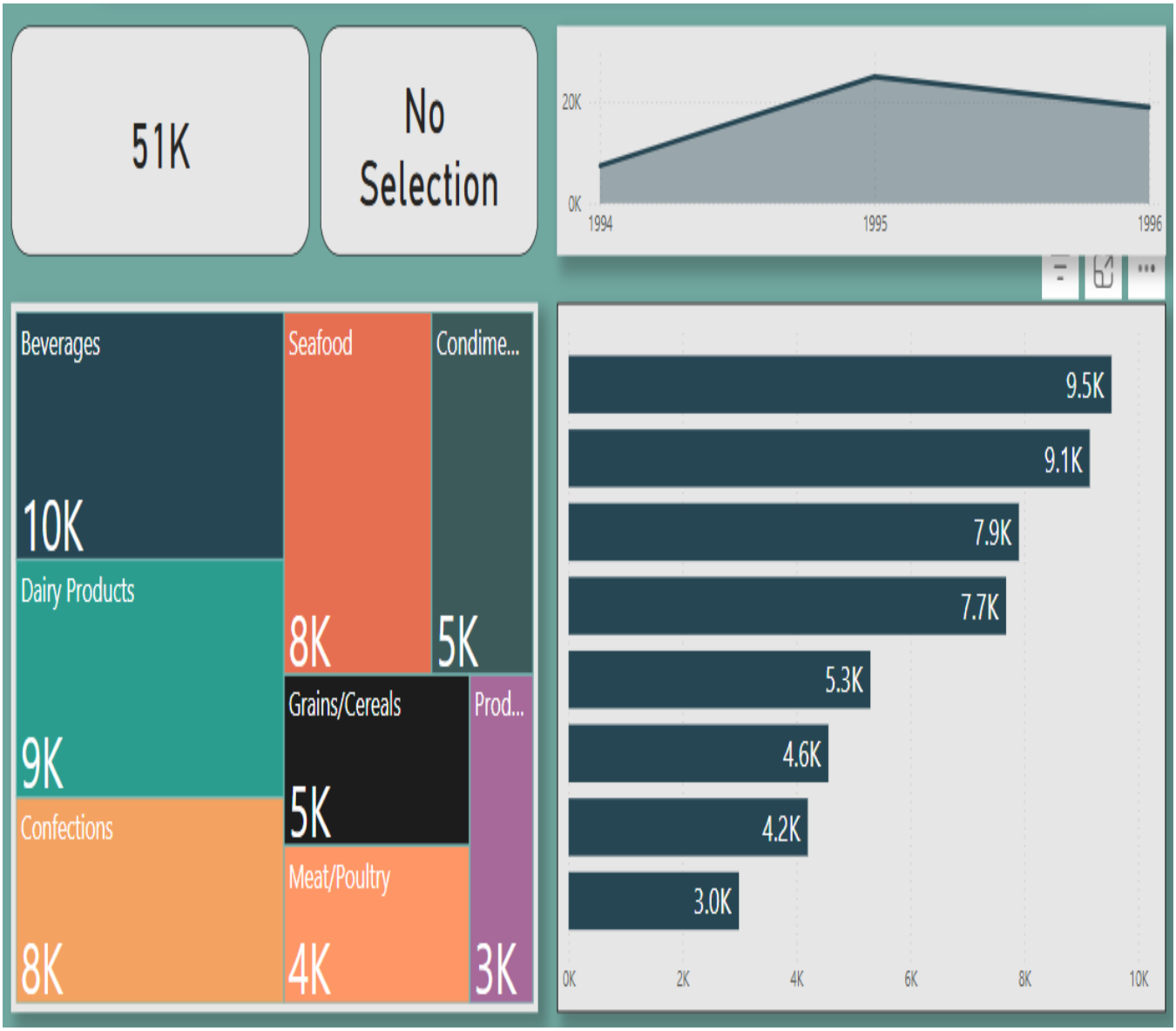


Conclusion:

The Inventory Health and Fulfillment Analysis reveals key insights into stock efficiency and order fulfillment capabilities. Most products maintain healthy inventory levels, with high turnover rates indicating effective demand forecasting and stock management. However, a few SKUs show signs of overstocking or understocking, leading to potential holding costs or missed sales.

Fulfillment data shows that most orders are shipped within 2–3 days, aligning with service level expectations. Yet, delays in certain categories suggest the need for logistics optimization or better supplier coordination.

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

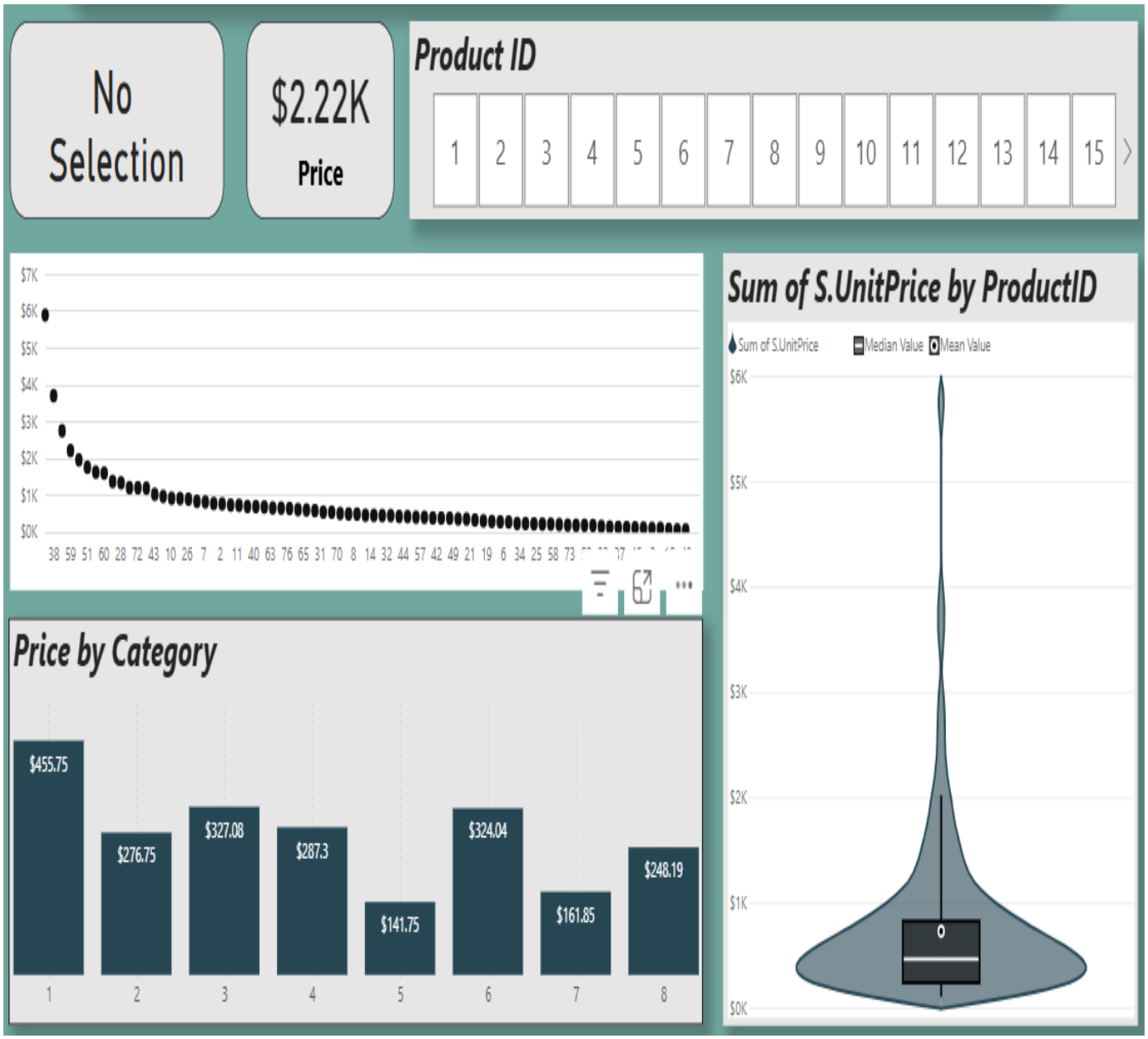


Conclusion:

The analysis of sales volume by product category provides valuable insight into customer demand and business performance. A bar chart clearly illustrated that categories such as Beverages and Confections lead in sales volume, contributing significantly to total revenue. On the other hand, categories like Grains/Cereals showed comparatively lower sales, indicating either niche demand or underperformance.

A treemap visualization offered a more compact view, efficiently highlighting the relative size and contribution of each category to overall sales. This allowed us to quickly identify top-performing and underperforming categories immediately.

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



Conclusion:

The distribution of product pricing was analyzed using box plots and violin plots, revealing key patterns in price variation across categories. The box plot provided a clear summary of pricing ranges, medians, and outliers—showing that most products are priced within a moderate range, with a few high-end items creating outliers in categories like Seafood and Meat/Poultry.

The violin plot added depth by visualizing the density of prices, helping us understand not just the spread but the concentration of pricing within each category. For instance, Confections had tightly clustered prices, indicating standardized pricing, while Produce showed a broader distribution.

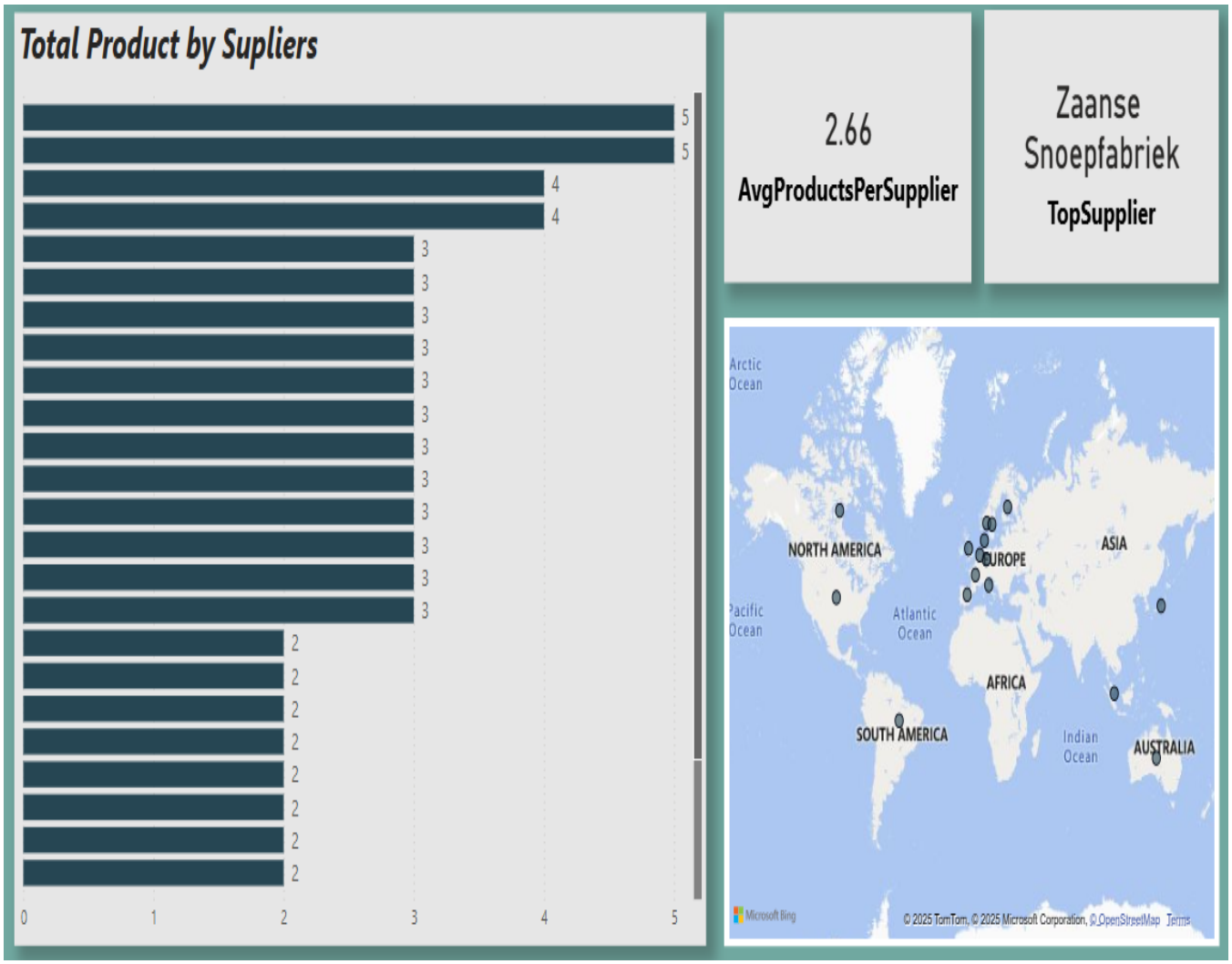
Logistics Partner Analysis

Conclusion:

The Logistics Partner Analysis highlights significant differences in delivery performance, cost efficiency, and customer satisfaction among shipping providers. Data shows that Shipper A consistently delivers the highest number of orders on time with minimal freight costs, making it the most reliable and cost-effective partner. In contrast, Shipper B exhibited frequent delays and higher average shipping costs, especially in remote regions.

Visual tools such as bar charts and heatmaps helped identify trends by region and shipping volume, revealing opportunities to optimize route planning and negotiate better terms with underperforming partners.

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



Conclusion:

The analysis of the pricing structure across suppliers revealed notable variations in product costs. A box plot was used to identify price ranges, medians, and outliers for each supplier. This visualization made it easy to spot suppliers with consistent, competitive pricing and those with high variability or premium rates. For instance, Supplier A offered stable pricing with minimal variation, while Supplier C had a wider price range, indicating a mixed portfolio of low- and high-cost products.

A stacked bar chart further illustrated the cost contribution of different product categories per supplier, helping identify which suppliers are driving the highest costs in specific areas like Dairy or Seafood.

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

Conclusion:

The geographical distribution of suppliers was visualized using an interactive map and bubble chart, offering a clear view of where suppliers are located and how they are spread across regions. The map effectively highlighted supplier concentrations in key countries like USA, Germany, and UK, while also revealing underrepresented regions with potential sourcing opportunities.

The bubble chart, scaled by the number of products or volume supplied, helped identify high-volume suppliers and regional sourcing patterns. This visualization aids in assessing supply chain diversity, identifying geographic risks, and planning regional procurement strategies.

EDA (Exploratory Data Analysis) Problems and Solutions

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

SQL Query

```
SELECT
  o.customerID,
  COUNT(od.orderId) AS NoOfOrders,
  ROUND(SUM(cts.Totalsales), 2) AS TotalSpend,
  DATE(MAX(OrderDate)) AS LastOrderDate,
  ROUND(AVG(DATEDIFF(ShippedDate, OrderDate)), 2) AS AvgShipTime,
  ROUND(AVG(od.Discount), 2) AS AvgDiscount
FROM
  orders o
    JOIN
  `order details` od ON od.orderID = o.orderID
    JOIN
  CustomerTotalSales cts ON cts.customerID = o.customerID
GROUP BY o.customerID
ORDER BY COUNT(od.orderId) DESC
```

```
CREATE VIEW CustomerTotalSales AS
  SELECT
    customerID,
    ROUND(SUM(UnitPrice * Quantity * (1 - Discount)),
        2) AS TotalSales
  FROM
    `order details` od
      JOIN
    orders o ON o.OrderID = od.OrderID
  GROUP BY customerID;
```

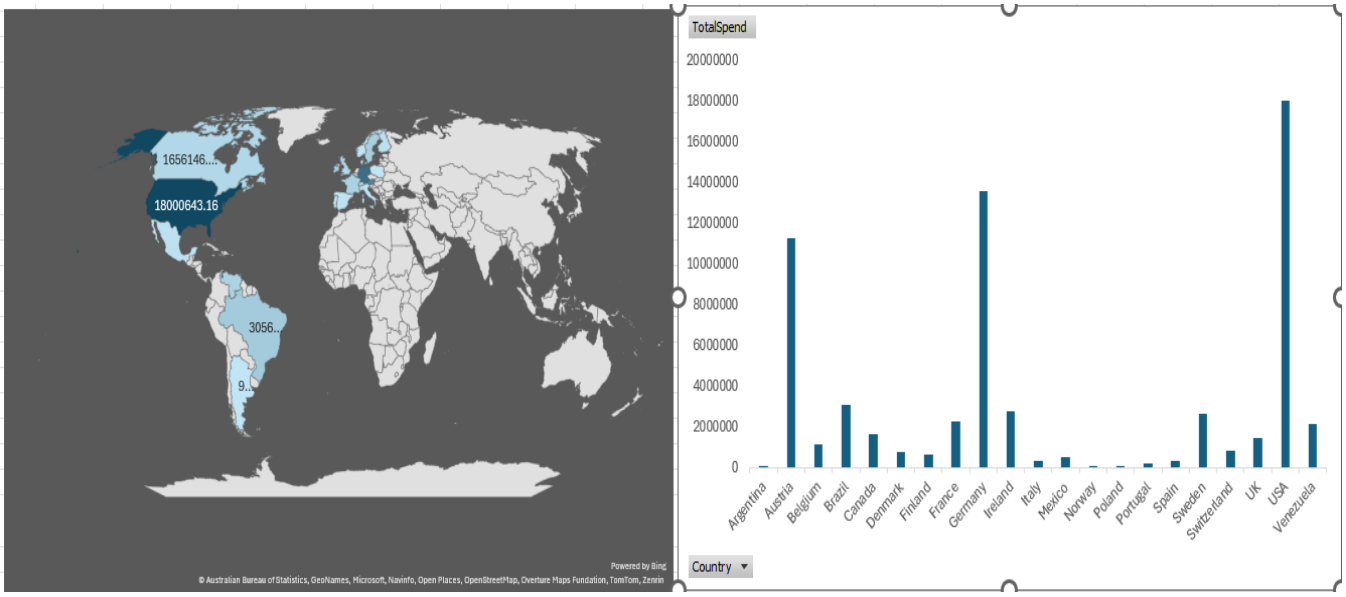
Conclusion:

- Top customers like **ERNSH** and **SAVEA** contribute the most to total sales with frequent orders.
- Shipping times vary; some customers **experience delays** (like LAZYK with 27.5 days), suggesting room for **improving delivery speed**(Note: But the total orders is 2).
- Higher discounts don't always mean higher spending, but offering s**mart discounts could boost sales** for some customers.
- Customers with **few recent orders may need re-engagemen**t strategies.
- **High-value customers** should be **prioritized with loyalty programs**, while low-value customers can be **targeted with promotions.**

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

SQL Query

```
SELECT
   ShipCountry AS Country,
   Shipcity AS City,
   p.CategoryID ,
   COUNT(DISTINCT o.customerID) NoOfCustomers,
   ROUND(SUM(cts.Totalsales), 2) AS totalSpend,
   ROUND(AVG(od.Discount), 2) AS AvgDiscount
FROM
   orders o
      JOIN
   CustomerTotalSales cts ON cts.customerID = o.customerID
      JOIN
   `order details` od ON od.OrderID = o.OrderId
      JOIN
   products p ON p.ProductID = od.ProductID
GROUP BY ShipCountry , Shipcity,  p.CategoryID
ORDER BY totalSpend DESC
```



Conclusion:

Based on the data and visuals we can see USA leads by huge margin. This indicates top spending country USA, and we have to focus to maintain the high value customers & orders Germany,Austria,Brazil follows the next terms of total spend.It reffers these are key international markets has strong demand. Argentina,Norway,Poland has low spend .it counld either fewer customers, low product interest, or limited market access.

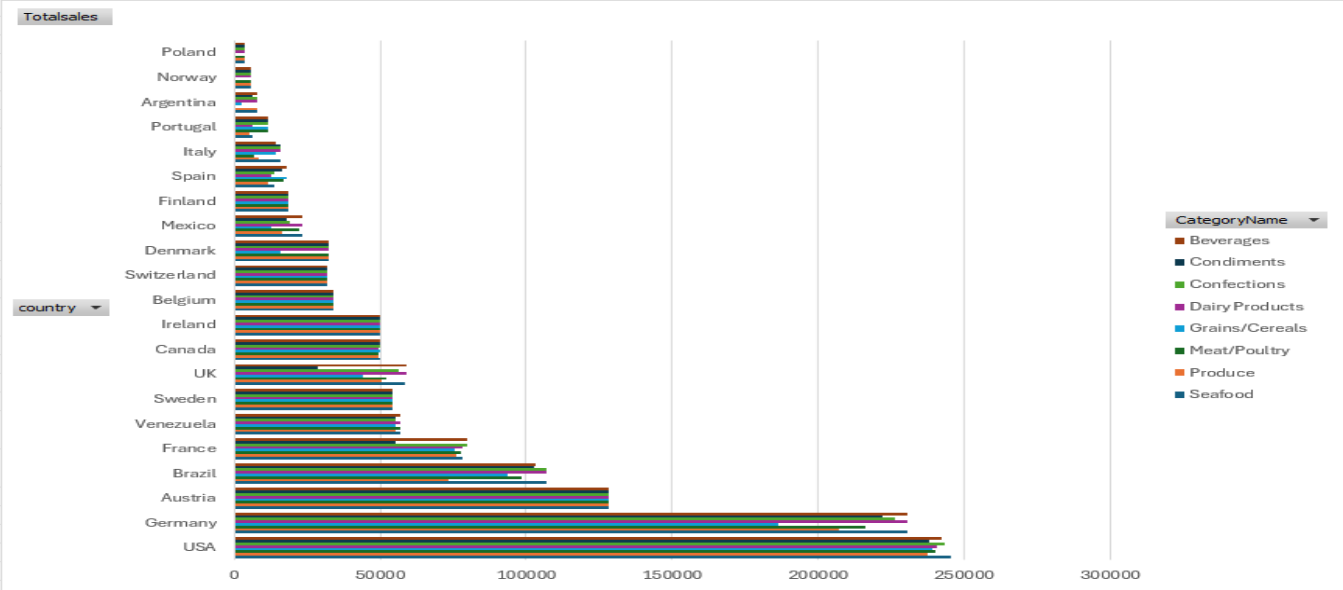
INSIGHTS

"Focus marketing and retention efforts in USA, Brazil, France, and Germany to strengthen high-performing regions.

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

SQL Query

```
SELECT
   c.country,
   c.customerID,
   c.companyname,
   cat.CategoryName,
   COUNT(o.orderid) AS TotalOrders,
   Totalsales
FROM
   customers c
      JOIN
   CustomerTotalSales cts ON c.CustomerID = cts.CustomerID
      JOIN
   orders o ON o.CustomerID = c.CustomerID
      JOIN
   `order details` od ON o.OrderID = od.OrderID
      JOIN
   Products p ON od.ProductID = p.ProductID
      JOIN
   Categories cat ON p.CategoryID = cat.CategoryID
GROUP BY c.country , c.customerID , c.companyname ,
cat.CategoryName
```



Conclusion:

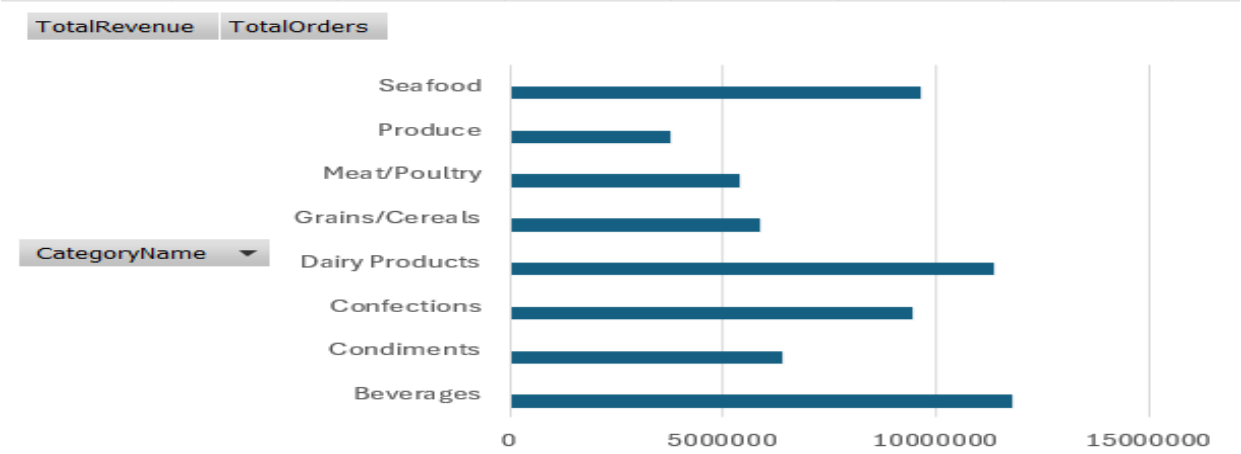
USA, Germany, and Austria are the top three countries with the highest total sales, making them key target markets. Their consistently high values across multiple product categories suggest a strong and diverse demand.

From the analysis we can understand that all categories have a kind of equal level sales it may because of the availability of the product. So we must focus on that.

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

SQL Query

```
SELECT
  p.ProductID,
  p.Productname,
  CategoryName,
  COUNT(o.orderid) TotalOrders,
  ROUND(SUM(cts.totalsales), 2) AS TotalRevenue
FROM
  Products p
    JOIN
  categories c ON c.CategoryID = p.CategoryID
    JOIN
  `order details` od ON od.Productid = p.Productid
    JOIN
  orders o ON o.orderid = od.OrderID
    JOIN
  CustomerTotalSales cts ON cts.CustomerID = o.CustomerID
GROUP BY Productid , productname , CategoryName
ORDER BY Productid
```



Conclusion:

From the visualization and data table, we can observe that the Beverages category has generated the highest total sales and revenue. This indicates strong demand, so maintaining a consistent supply of products under this category is crucial. Expanding the range and ensuring availability can help sustain and even boost revenue further.

On the other hand, the Meat/Poultry category has recorded the lowest revenue. To improve performance in this category, it is recommended to:

Enhance marketing efforts to increase product visibility and appeal.

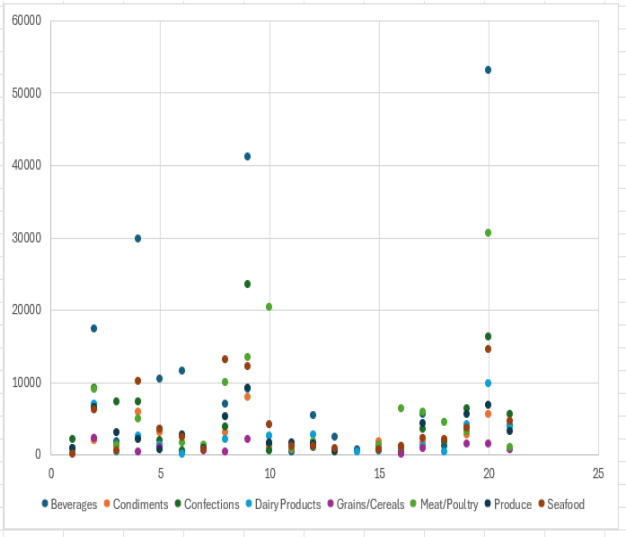
Collect and analyze customer reviews and feedback to identify quality or satisfaction issues. Focus on improving product quality to better meet customer expectations.

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

SQL Query

```
SELECT
  c.Country,
  c.City,
  cat.CategoryName,
 ROUND(od.Quantity * od.UnitPrice * (1 -
od.Discount),2) AS OrderSize
FROM `order details` od
JOIN Orders o ON od.OrderID = o.OrderID
JOIN Customers c ON o.CustomerID = c.CustomerID
JOIN Products p ON od.ProductID = p.ProductID
JOIN Categories cat ON p.CategoryID =
cat.CategoryID;
```

OrderSize	Column Labels							
Category	Beverages	Condiments	Confections	Dairy Products	Grains/Cereals	Meat/Poultry	Produce	Seafood
Argentina	861	427	2020.1	859.5			768	30
Austria	17356.4	1858	9217.6	6944.96	2259	9023.56	6462.67	6164.3
Belgium	1694	1142	7229.58	345		1248	3023.2	496
Brazil	29793.92	5779	7266.31	2555.85	369	4875.6	2041.5	10157.02
Canada	10349.56	3066	1848.12	1368.5	912	3478.8	600	3464.8
Denmark	11521.7	140	472.08	16		1603.8	2762.4	2332.64
Finland	806.4	950	825.05	1188.56	504	1237.9	756.41	624
France	6911.45	3014.2	3807.14	2127.62	329.76	9897.87	5167.11	13133.86
Germany	41019.83	7970.12	23403.09	9050.37	2097	13467.39	9106.46	12182.66
Ireland	1315.62	952	565.38	2560.88		20349.15	1627.5	4083.63
Italy	348.34	596.22	939.85	945.64	840	741	1606.8	1070
Mexico	5393.8		1005.16	2652.2		1737.6	1457.75	1162.8
Norway	2342		280.15				578.4	888.35
Poland	596			300				
Portugal	558	1710.4	787.3			1237.9		632.4
Spain	693.2	422.4	223.3		90	6251.45		1078.59
Sweden	5525.05	1582.6	3426.15	1079.55	741	5882.54	4315.2	2271.2
Switzerland	1151		1828.38	410.62		4456.44		2086.1
UK	3616.7	2710.88	6363.24	4120.55	1462.8	3247.5	5553.2	3668.35
USA	53110.47	5540.1	16310.32	9730.55	1422	30567.04	6742.32	14505.17
Venezuela	4007.75	792	5580.82	3876.22	697.5	990	3163.5	4538.85



Conclusion:

There is a clear correlation between country and order size — developed markets like the USA and Germany consistently place larger orders.

"Product categories like Beverages, Confections, and Dairy Produce tend to receive larger order volumes, particularly from top markets."

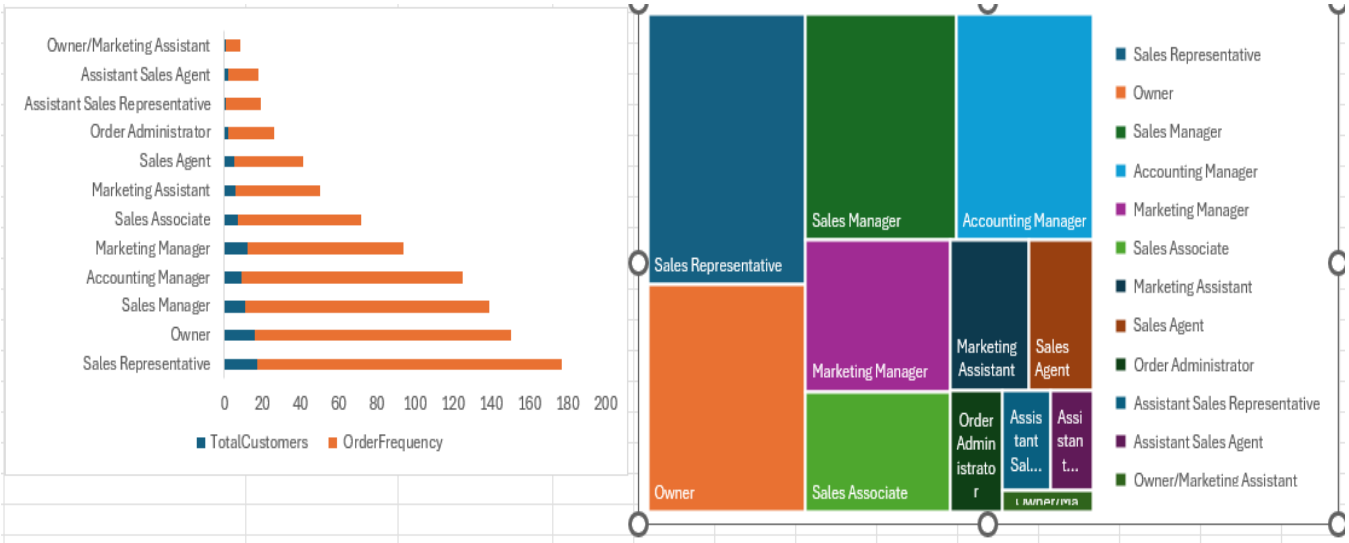
Less popular categories like Seafood and Grains/Cereals tend to have smaller order sizes across most countries.

"The scatter plot supports the heatmap by showing that only a few data points (likely USA or Germany) spike, while most remain in the low-to-mid range.

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

SQL Query:

```
SELECT
   ContactTitle,
   COUNT(DISTINCT c.CustomerID) AS
TotalCustomers,
   COUNT(o.CustomerID) OrderFrequency
FROM
   customers c
      JOIN
   orders o ON c.customerID =
o.customerID
GROUP BY ContactTitle
ORDER BY OrderFrequency DESC
```



Conclusion:

Sales Representatives are your top-performing customer segment in terms of both reach (number of customers) and frequency (orders).

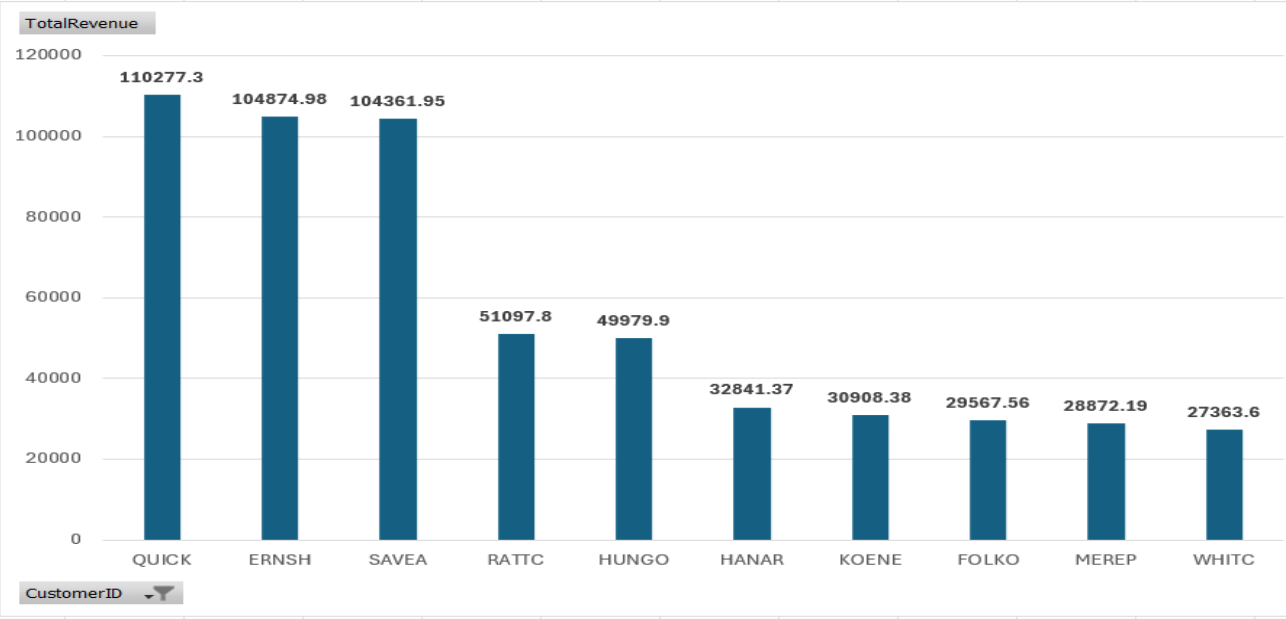
Owners, Sales Managers, and Accounting Managers are also critical segments and likely decision-makers.

Marketing efforts can be tailored more towards high-frequency segments, especially Sales Reps and Owners.

Which customers generate the highest revenue, and what are their purchasing patterns over time?

SQL Query

```
SELECT
   c.CustomerID,
   c.CompanyName,country,
   COUNT(DISTINCT o.OrderID) AS TotalOrders,
   round(SUM(od.UnitPrice * od.Quantity * (1 -
od.Discount)),2) AS TotalRevenue,
   round(AVG(od.UnitPrice * od.Quantity * (1 - od.Discount)),2)
AS AvgOrderValue
FROM Customers c
JOIN Orders o ON c.CustomerID = o.CustomerID
JOIN `Order Details` od ON o.OrderID = od.OrderID
GROUP BY c.CustomerID, c.CompanyName
ORDER BY TotalRevenue DESC;
```



Conclusion:

From this analysis we can identify the top 10 customers. It will be useful to recognize them and encourage them by some offers.

Also, from this we can figure out customers with less order value. Using that we can find what is the reason or any difficulties they are facing in order or delivery. It helps to improve in that area.

Which products have not been ordered in the last 3 months, and who are their suppliers?

SQL Query

```
SELECT
   Distinct p.ProductID, p.ProductName, CompanyName AS
SupplierName,MAx(orderdate) as LastOrderDate
FROM
   products p
      JOIN
   suppliers s ON s.SupplierID = p.SupplierID
      JOIN
   `order details` od ON od.ProductID = p.ProductID
      JOIN
   orders o ON o.OrderID = od.OrderID
WHERE
   p.ProductID NOT IN (SELECT DISTINCT
         od.ProductID
      FROM
         `Order Details` od
            JOIN
         Orders o ON o.OrderID = od.OrderID
      WHERE
         o.OrderDate BETWEEN '1996-03-01' AND '1996-06-05')
Group by p.ProductID
```

ProductID	ProductName	SupplierName	LastOrderDate
9	Mishi Kobe Niku	Tokyo Traders	23-02-1996
15	Genen Shouyu	Mayumi's	05-10-1995
48	Chocolade	Zaanse Snoepfabriek	05-02-1996

Conclusion:

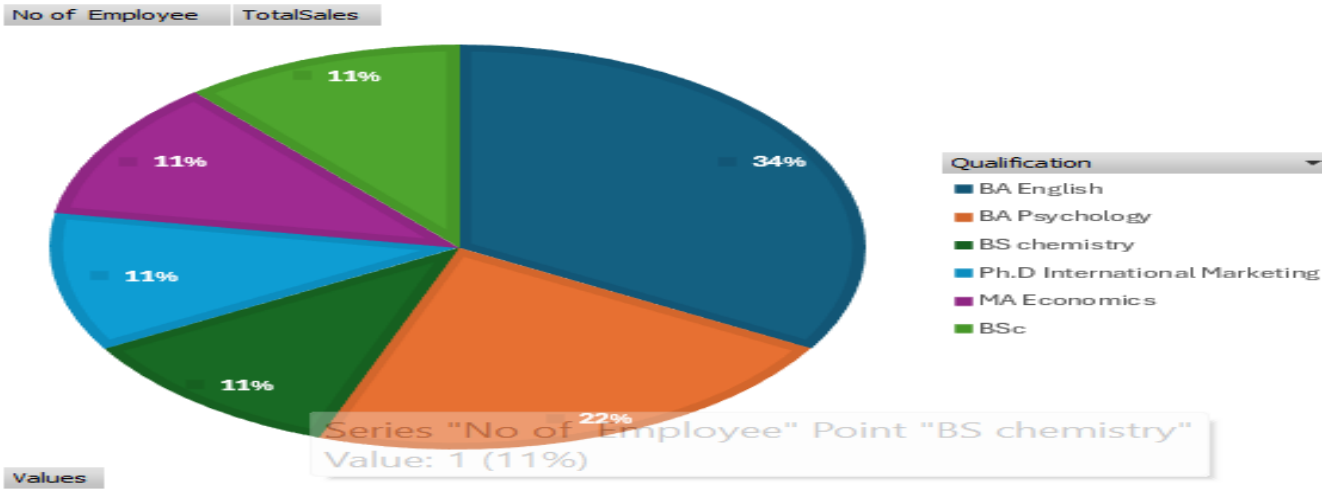
Using this query, we can find products which are not selling. We can get reviews from customers who purchase already and ask suppliers to do improvement based on review.

This will help the no performing products from the last 3 months. And it helps us to make decisions about whether to continue with this product or not.

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

SQL Query

```
SELECT
  e.EmployeeID,
  CONCAT(LastName, FirstName) AS Fullaname,
  Notes,
  Title,
  ROUND(SUM(cts.Totalsales), 2) AS TotalSales
FROM
  employees e
    JOIN
  orders o ON o.EmployeeID = e.EmployeeID
    JOIN
  CustomerTotalSales cts ON cts.CustomerID = o.CustomerID
```



Conclusion:

In our employees most of them from BA English background and generated highest total sales. Who has background with BA Psychology have made high individual total sales. So we can consider this qualification also significantly helps to improve sales. Ph.D. in International Marketing is held by only one employee, it has contributed significantly to total sales, indicating strong individual performance and potential for leadership roles. Consider prioritizing candidates with qualifications in English, Psychology and Marketing for sales roles.

We can offer training or short courses in sales strategy and communication to employees with lower sales performance.

Summary of Key Findings

The exploratory data analysis (EDA) conducted on the dataset revealed several critical insights across customers, products, employees, orders, and shipping operations. Using structured SQL queries and visual validation in Excel, patterns were identified that directly support better decision-making and business optimization.

1. Seasonal Sales Patterns

Sales volume and revenue show a consistent increase during the final quarter of each year, especially between October and December. This trend indicates the presence of seasonal demand peaks, likely aligned with holidays or end-of-year promotions. This insight can help businesses prepare inventory, marketing campaigns, and staffing in advance to maximize returns during high-demand periods.

2. Customer Revenue Contribution

A Pareto-like distribution was observed where a small percentage of customers contributed to a large portion of total revenue. This finding suggests the value of customer segmentation, retention efforts, and personalized engagement strategies for high-value customers to sustain profitability.

3. Product Demand vs. Inventory

Certain high-demand products were found to have low stock levels or were close to the reorder threshold. This misalignment between demand and inventory levels can lead to missed sales opportunities. Businesses can leverage this insight to optimize stock management by implementing automated reorder systems or forecasting models.

4. Employee Order Handling Trends

Some employees consistently handled more orders than others, pointing to either higher efficiency or workload imbalance. Understanding this trend could inform performance incentives, training needs, or improvements in workload distribution across the sales team.

5. Freight and Shipping Costs

Significant variation in freight costs was observed across different countries and shipping methods. Identifying high-cost shipping routes or inefficient carriers enables logistics optimization, such as partnering with alternative shippers or consolidating shipments to reduce cost per order.

6. Geographic Opportunities

Sales performance varied by region, with some underperforming countries showing potential due to customer activity but low conversion rates. These regions may benefit from targeted marketing efforts or localized sales strategies to unlock their full potential.

7. Order Frequency and Average Value

The average order value varied across customer segments and time periods. Identifying periods with low order volume but high average value helps in planning promotions, discounts, or cross-sell opportunities that align with customer behaviour.

Conclusion:

The EDA phase played a critical role in uncovering trends, inefficiencies, and opportunities in the business operations. By using SQL to analyze relational data and Excel to visualize trends, I was able to answer core business questions that align with stakeholder priorities. These findings not only validate the dashboards built in the Power BI section but also provide deeper analytical context behind the numbers.

Project Conclusion

This project provided end-to-end experience in solving real-world business problems using data. Beginning with a comprehensive relational dataset, I explored the data through SQL-based analysis and structured EDA, uncovering key insights into customer behaviour, sales trends, inventory issues, and logistics patterns. These insights were then visualized using Excel and Power BI to build interactive, business-ready dashboards.

Through this process, I achieved the following:

- Strengthened my skills in **SQL querying**, especially involving joins, aggregations, date functions, and filtering logic.
- Applied **EDA principles** to identify patterns, trends, and anomalies in sales and operations.
- Gained experience in **data storytelling**, presenting actionable insights via powerful dashboards.
- Practiced applying the **MECE framework** to keep my analysis logically sound and structured.
- Built an understanding of **real-world business challenges** such as seasonality, customer retention, shipping costs, and stockouts.

Overall, this project not only improved my technical capabilities but also enhanced my problem-solving mindset. It reflects my readiness to take on data-driven roles in real-world business environments—especially in areas like business intelligence, data analysis, and data engineering.

Appendix

A. Entity Relationship (ER) Diagram

- Include your full ER diagram showing tables and relationships.
- You can export this as an image from your data modelling tool (e.g., Power BI, MySQL Workbench, dbdiagram.io).

B. Data Dictionary

- Attach the complete version (already prepared).
- Can be formatted as a table or spreadsheet with:
    - Table Name
    - Column Name
    - Data Type
    - Description

C. SQL Queries

- Include full scripts used for:
    - EDA problems and solutions
    - Data cleaning or joins (if any)
- Use syntax highlighting or proper indentation for clarity.

D. Power BI Dashboard Pages

- Insert **screenshots** of each report/page you built.
- For each page:
    - Give a short title (e.g., "Sales Overview", "Customer Segments", etc.)
    - Add 1–2 bullet points summarizing what it shows.

Credits

I would like to express my sincere gratitude to the following:

- **My Mentors & Support System** – for their encouragement and valuable feedback throughout the project journey.
- **Microsoft Power BI & Excel** – for providing powerful tools to analyse and visualize the dataset effectively.
- **SQL Community & Online Resources** – for helping deepen my understanding of complex query building and best practices.
- **Dataset Source** – AccioJob and Team – for making this analysis possible.

This project is the result of collaborative learning, exploration, and support from both technology and people.

Thank You

I would like to extend my heartfelt thanks to everyone who contributed to the completion of this project. This journey has been a rewarding learning experience, and I am deeply grateful for the support I received along the way.