**capestone project**

****

**SALES ANALYTICS**

**SHIVA CHARAN**

Overview

This Power BI project focuses on developing an interactive and visually engaging dashboard for Northwind Traders, a fictitious global food trading company. The main goal is to analyse and present key performance metrics related to sales, customer behaviour, inventory trends, and employee performance in a user-friendly and insightful format.

Dataset Summary

The analysis is based on the well-known Northwind database, which includes structured data across multiple interconnected tables such as:

* Customers – demographic and regional details
* Orders & Order Details – transaction-level sales data
* Products, Categories, Suppliers – inventory and procurement information
* Employees – HR and performance-related data
* Shippers – logistics and delivery partners

This rich dataset provides a comprehensive view of the company’s operations, allowing for deep exploration and insight generation.

Analysis and Insights

Through a series of Power BI dashboards, the project answers key business questions and performs exploratory data analysis (EDA) on:

* Customer segmentation and acquisition trends
* Sales performance over time and across categories
* Employee productivity and tenure patterns
* Inventory and supplier performance
* Operational efficiencies such as shipping durations and order processing times

Various visualization techniques such as bar charts, line charts, heatmaps, scatter plots, and maps were used to uncover trends, identify outliers, and highlight patterns.

THE PROCESS

**1. Data Acquisition from GitHub**

The project begins with data acquisition, where the Northwind Traders dataset is obtained from a GitHub repository or downloaded as a sample database from Microsoft's official sources. This dataset contains comprehensive information about various business entities such as Customers, Orders, Employees, Products, Suppliers, Shippers, and Categories. These tables form the foundation for performing sales analysis, customer segmentation, and performance tracking across different dimensions.

**2. Data Transformation and Enhancement**

After acquiring the dataset, the next step is to perform data transformation to ensure it is clean, structured, and ready for analysis. This involves removing duplicates, handling missing values, formatting date fields, and standardizing categories. Relationships among tables are defined using appropriate primary and foreign keys. Additional columns or measures are created to enhance the dataset, such as calculated fields for total sales, profit margins, or order frequency. Problem statements are also introduced at this stage to guide the analysis and identify specific business questions to be answered.

**3. Connecting with Tools**

With a clean and enriched dataset in place, it is then connected to various analytical tools for further processing and visualization. Power BI is used as the primary dashboard and visualization tool, while MySQL Workbench is employed for SQL-based querying and Excel may be used for quick EDA or pivot analysis. These tools are linked efficiently to ensure data synchronization and smooth workflow across platforms.

**4. Problem Statement Solution in Power BI**

Power BI is utilized to address specific business problem statements through interactive visualizations and dashboards. Key performance indicators (KPIs) are tracked using visual elements such as bar charts, line graphs, pie charts, tables, maps, and slicers. These visualizations help explore patterns in customer orders, regional sales performance, employee contributions, and supplier efficiency. Filters and slicers add interactivity, allowing users to drill down into specific segments and generate actionable insights from the data.

**5. Exploratory Data Analysis (EDA)**

Parallel to the Power BI dashboard creation, exploratory data analysis is conducted to understand the underlying trends and relationships in the dataset. SQL queries in MySQL Workbench or Excel-based pivot tables are used to extract insights such as best-selling products, customer purchase frequency, sales by region, and employee performance metrics. This analysis helps validate the data and provides direction for building meaningful Power BI visualizations.

**6. Creation of Visual and Insightful PowerPoint**

A PowerPoint presentation is created to showcase the project workflow, findings, and solutions. The presentation includes slides for each problem statement, supported by screenshots of Power BI visuals and brief explanations of insights derived. It also highlights the tools used, the methodology followed, and the key takeaways that can help decision-makers understand business performance and make strategic improvements.

**7. Detailed Documentation**

Finally, a comprehensive project report is compiled to document every phase of the project. This report covers data acquisition, cleaning and transformation processes, tool integration, problem statement analysis, Power BI dashboard features, and EDA findings. The documentation ensures clarity, transparency, and replicability of the project, serving as a valuable resource for both technical teams and business stakeholders.

OBJECTIVE

Northwind Traders is a classic sample business dataset widely used for demonstrating database and analytical techniques. The dataset represents a fictional company that imports and exports specialty foods from around the world.

The objective of this project is to perform an in-depth analysis of the Northwind Traders dataset by creating a dynamic and interactive Power BI dashboard. The goal is to extract actionable insights related to sales trends, customer behaviour, product performance, regional growth, and employee productivity to support strategic decision-making and business optimization.

The project will involve the following tasks:  
• Cleaning and transforming the dataset to ensure accuracy and consistency.  
• Connecting the dataset with analytical tools including Power BI, Excel, and MySQL Workbench.  
• Developing insightful Power BI dashboards addressing specific business problem statements such as top-selling products, customer segmentation, and region-wise sales performance.  
• Conducting exploratory data analysis (EDA) to identify hidden patterns and trends.  
• Creating a PowerPoint presentation that summarizes the objectives, methodologies, insights, and visualizations derived from the analysis.  
• Preparing a detailed documentation report covering the entire project lifecycle.

The success of the project will be measured by the following metrics:  
• The effectiveness and clarity of the visualizations and dashboards.  
• The accuracy and relevance of the insights derived from the analysis.  
• The practical value of the recommendations and conclusions for business improvement.

This project is significant because it demonstrates the power of data analytics and visualization in enhancing business intelligence. By leveraging Power BI and SQL-based insights, businesses like Northwind Traders can make informed decisions to drive efficiency, customer satisfaction, and profitability.

SIGNIFICANCE

The Northwind Traders Power BI dashboard project holds significant value for business stakeholders, analysts, and decision-makers by transforming raw sales and operational data into meaningful insights. Through the integration of data from different sources and the development of dynamic visualizations, this project facilitates data-driven decision-making that can enhance business performance and operational efficiency.

For business analysts and managers, the dashboard provides a centralized view of critical metrics such as sales trends, customer segmentation, top-performing products, and employee productivity. By exploring these insights, decision-makers can identify growth opportunities, address inefficiencies, and optimize strategies to drive profitability. The clarity and accessibility of these insights support faster and more informed business decisions.

For sales and marketing teams, the analysis highlights customer behaviour and buying patterns, enabling more targeted and effective campaigns. Understanding which products perform best in specific regions or among certain customer groups allows for better resource allocation and product positioning.

Organizations can also leverage this analysis to evaluate employee contributions and regional performance, helping refine internal processes and recognize areas for training or development. With clear visual representations of data trends, companies can proactively respond to business challenges and capitalize on opportunities.

Additionally, this project serves as a learning model for implementing end-to-end data analytics workflows, including data preparation, integration with tools like MySQL and Excel, exploratory analysis, and dashboard creation in Power BI. It emphasizes the importance of clean, well-structured data and showcases how technology can translate raw data into business value.

In conclusion, the Northwind Traders Power BI project is a vital initiative that bridges the gap between data and decision-making. It empowers various business functions with actionable insights, supports strategic planning, and demonstrates the transformative power of data analytics in modern business environments.

# DATA DICITIONARY

## Table: Customers

Stores information about the company's customers.

Fields:

• CustomerID: Unique identifier for each customer.

• CompanyName: Name of the customer’s company.

• ContactName: Name of the primary contact person.

• ContactTitle: Title of the contact person.

• Address: Street address of the customer.

• City: City where the customer is located.

• Region: Region or state of the customer.

• PostalCode: Postal or ZIP code.

• Country: Country where the customer is located.

• Phone: Contact phone number.

• Fax: Fax number.

## Table: Employees

Stores information about the company's employees.

Fields:

• EmployeeID: Unique identifier for each employee.

• LastName: Last name of the employee.

• FirstName: First name of the employee.

• Title: Job title of the employee.

• TitleOfCourtesy: Courtesy title (e.g., Mr., Mrs., Dr.).

• BirthDate: Date of birth.

• HireDate: Date the employee was hired.

• Address: Street address of the employee.

• City: City where the employee lives.

• Region: Region or state of residence.

• PostalCode: Postal or ZIP code.

• Country: Country of residence.

• HomePhone: Home phone number.

• Extension: Telephone extension.

• Photo: Employee photo (binary or link).

• Notes: Additional notes about the employee.

• ReportsTo: ID of the manager/supervisor (references another EmployeeID).

• PhotoPath: Path or URL to the employee’s photo.

## Table: Orders

Stores information about orders placed by customers.

Fields:

• OrderID: Unique identifier for each order.

• CustomerID: Foreign key referencing the Customers table.

• EmployeeID: Foreign key referencing the Employees table.

• OrderDate: Date when the order was placed.

• RequiredDate: Date when the order is required.

• ShippedDate: Date the order was shipped.

• ShipVia: Foreign key referencing the Shippers table.

• Freight: Shipping cost.

• ShipName: Name of the recipient or shipping contact.

• ShipAddress: Shipping address.

• ShipCity: City to which the order was shipped.

• ShipRegion: Region or state for shipping.

• ShipPostalCode: Postal or ZIP code for shipping.

• ShipCountry: Country to which the order was shipped.

## Table: Order Details

Stores detailed information about each product in an order.

Fields:

• OrderID: Foreign key referencing the Orders table.

• ProductID: Foreign key referencing the Products table.

• UnitPrice: Unit price of the product at the time of the order.

• Quantity: Quantity of the product ordered.

• Discount: Discount applied to the product (as a fraction, e.g., 0.1 for 10%).

## Table: Products

Stores details about the company's products.

Fields:

• ProductID: Unique identifier for each product.

• ProductName: Name of the product.

• SupplierID: Foreign key referencing the Suppliers table.

• CategoryID: Foreign key referencing the Categories table.

• QuantityPerUnit: Description of quantity per unit (e.g., "24 - 12 oz bottles").

• UnitPrice: Price per unit of the product.

• UnitsInStock: Current stock level.

• UnitsOnOrder: Number of units currently on order.

• ReorderLevel: Minimum stock level before reorder is needed.

• Discontinued: Boolean indicating if the product is discontinued (true/false or 1/0).

## Table: Suppliers

Stores information about product suppliers.

Fields:

• SupplierID: Unique identifier for each supplier.

• CompanyName: Name of the supplier company.

• ContactName: Name of the supplier's contact person.

• ContactTitle: Title of the contact person.

• Address: Supplier’s address.

• City: Supplier’s city.

• Region: Supplier’s region or state.

• PostalCode: Supplier’s postal or ZIP code.

• Country: Country of the supplier.

• Phone: Supplier’s phone number.

• Fax: Supplier’s fax number.

• HomePage: URL to the supplier's home page.

## Table: Shippers

Stores details about shipping companies.

Fields:

• ShipperID: Unique identifier for each shipper.

• CompanyName: Name of the shipping company.

• Phone: Contact number of the shipper.

## Table: Categories

Stores information about product categories.

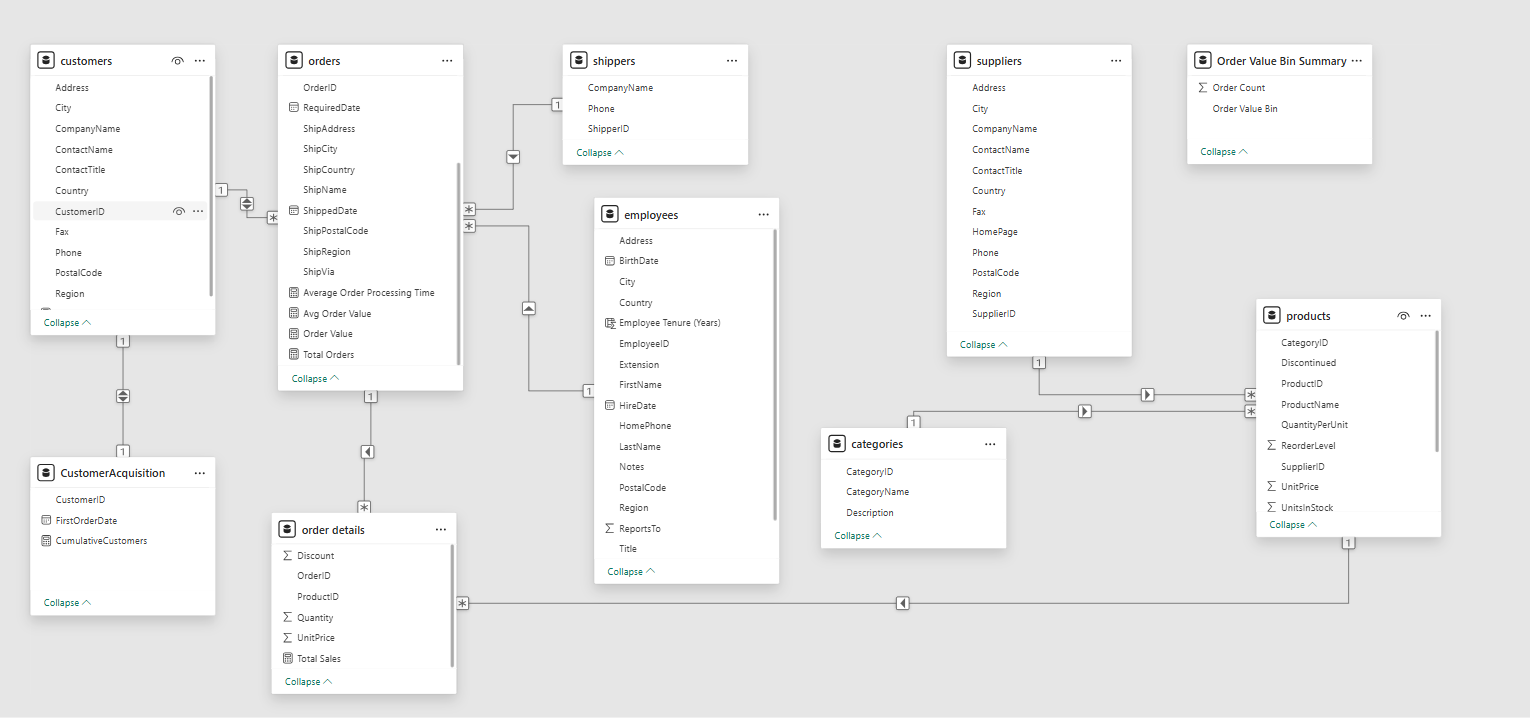
Fields:

• CategoryID: Unique identifier for each category.

• CategoryName: Name of the category.

• Description: Description of the category.

ER DIAGRAM



POWER BI PROBLEM STATEMENTS

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

### **1. Objective of the Analysis**

The primary objective of this analysis is to understand how Northwind Traders' customer base is distributed across different geographic regions and customer segments. This was accomplished using Power BI by leveraging customer data from the Northwind dataset. The visual report includes key performance indicators (KPIs) and visualizations to present both the magnitude and spread of customer relationships globally.

### **2. Summary of KPI Cards**

The dashboard consists of three KPI cards that provide quick, high-level insights. The first card displays the total number of customers, which is 57, representing the unique customers listed in the dataset. The second KPI card shows the total sales value of 1.27 million, reflecting cumulative revenue likely derived from associated orders linked to these customers. The third card shows that customers are distributed across 21 countries, indicating a moderately broad international presence.

### **3. Customer Segmentation by Contact Title**

A horizontal bar chart titled “Customers by Contact Title” breaks down the customer base by the professional roles of the primary contacts. The most frequent roles are "Owner" and "Sales Representative," followed by positions like "Marketing Manager," "Sales Manager," and "Accounting Manager." This breakdown reveals that Northwind often engages with decision-makers and those directly involved in sales and purchasing. It provides insight into the business relationships Northwind maintains and offers opportunities for tailored communication strategies based on role-specific needs.

### **4. Geographic Distribution of Customers**

The geographical distribution is visualized using a world map titled “Customers by Country.” This map shades countries according to the number of customers, clearly showing that the company has a concentrated presence in North America, Europe, and South America. These regions are heavily shaded, indicating they host the majority of Northwind’s customer base. Conversely, regions like Asia, Africa, and Australia are either lightly shaded or not represented, suggesting that these markets are either underdeveloped or unexplored from Northwind's perspective.

### **5. Assumptions Made in the Analysis**

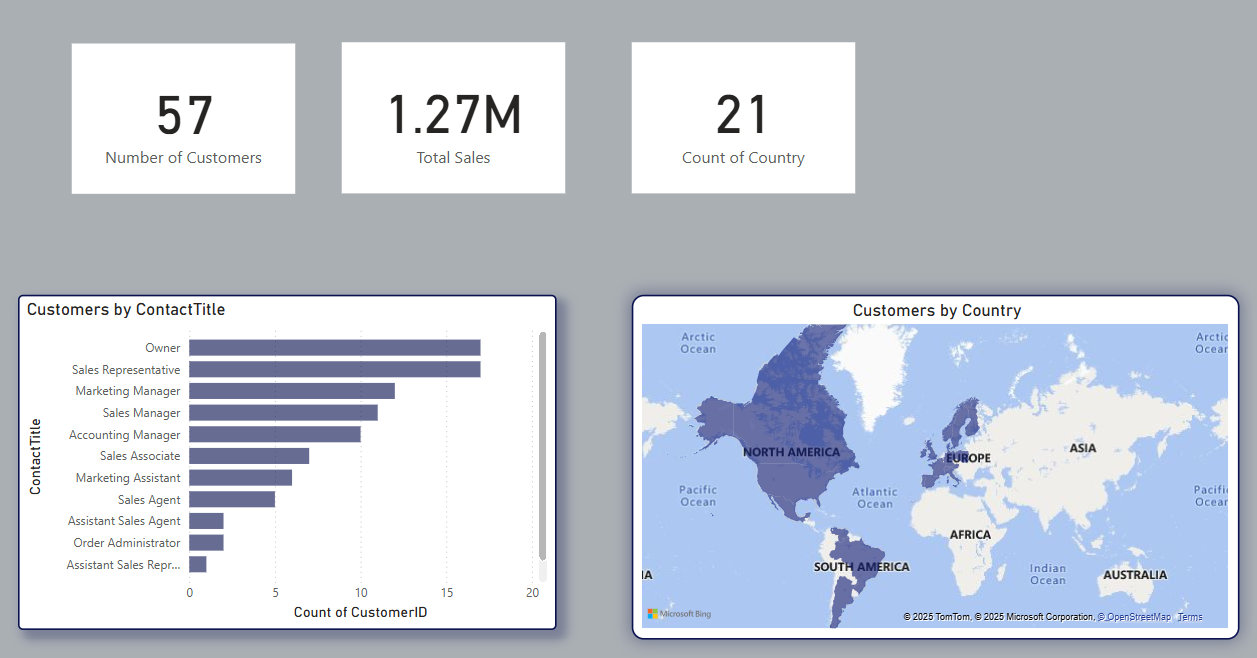
Several assumptions underpin this analysis. First, because the dataset does not explicitly include customer segment categories such as business size or industry, the contact title field was used as a proxy for segmentation. Second, country-level data was used to define geographic regions, as the Region field was either sparsely populated or inconsistent. Third, the sales data is assumed to be accurate and linked correctly through relationships among the Customers, Orders, and Order Details tables. Finally, the analysis assumes that the dataset is complete and representative of current operations.

### **6. Key Insights Drawn from the Report**

From the analysis, several insights emerge. The strong presence in North America and Europe implies that these are established markets with significant engagement. The dominance of high-level job titles like "Owner" and "Sales Representative" shows that Northwind maintains direct relationships with key business decision-makers. Additionally, the presence in 21 countries reflects a globally diversified customer base, though with room for improvement in certain regions.

### **7. Conclusion and Recommendations**

In conclusion, the Power BI report provides a clear picture of customer distribution by geography and role. Northwind Traders appears to have a strong footprint in the Western Hemisphere, especially among decision-makers. This insight can guide sales and marketing teams to focus on current strongholds while identifying growth opportunities in underrepresented areas. To enhance this analysis further, Northwind could enrich its customer data with additional segmentation variables such as industry, company size, or customer lifecycle status.



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

## 1. Objective of the Analysis

The second analysis focuses on understanding customer acquisition trends over time, with a particular interest in how many new customers Northwind Traders gained each year. This insight is critical for evaluating growth momentum and identifying periods of expansion or decline. To visualize this, the Power BI report includes two key visual elements: a bar chart titled 'Cumulative Customers by Year' and a line chart titled 'Count of Customers.'

## 2. Interpretation of the Visualizations

The bar chart presents the cumulative number of customers acquired by year. It shows a steep rise in 1994, with approximately 40 customers acquired, followed by a sharp drop in 1995 where only around 15 new customers were added. The number dwindles further in 1996, with fewer than 5 customers acquired. This suggests that 1994 was a significant year for customer onboarding, while the following years saw a noticeable slowdown in growth.

## 3. Year-over-Year Trends from Line Chart

The line chart reinforces this pattern by depicting the year-over-year count of new customers. It reflects a clear downward trend, showing the highest number of acquisitions in 1994 and declining steadily through 1995 and 1996. This indicates that after an initial expansion, Northwind experienced a reduction in the pace of new customer acquisition, possibly due to market saturation, changes in strategy, or external business factors.

## 4. Assumptions in the Analysis

A few assumptions were necessary to complete this analysis. First, the customer acquisition year is likely derived from the earliest order date associated with each customer, since there is no explicit 'customer registration date' field in the Northwind dataset. Secondly, it is assumed that all customers appearing in the dataset are still active and relevant to the timeframe analyzed. The analysis also presumes data integrity across related tables (Customers, Orders) to ensure that customer acquisition timing is accurately calculated.

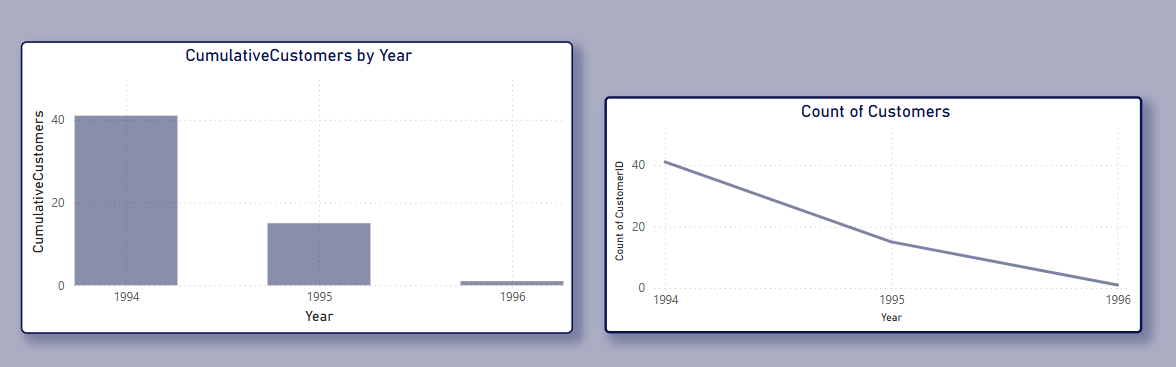
## 5. Insights and Interpretation

Several important insights emerge from this visualization. The spike in 1994 could indicate an aggressive marketing push, a product launch, or entry into new markets that attracted many customers. Conversely, the decline in subsequent years may suggest either a shift in focus from acquisition to retention or potential market limitations. The dramatic reduction in 1996 signals a need for closer investigation into external and internal business dynamics during that period.

## 6. Conclusion and Recommendations

In conclusion, this Power BI analysis highlights a strong initial customer acquisition phase, followed by a downward trajectory in growth. Northwind Traders may need to revitalize its customer acquisition strategy or explore new markets to reignite growth. Further analysis, such as examining marketing spend or competitive activity by year, would provide more context to these patterns

.



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

## 1. Objective of the Analysis

The objective of this analysis is to examine how total sales are distributed across different customer contact titles, as a proxy for understanding customer segmentation in the absence of detailed demographic attributes such as age, gender, or income. The Northwind dataset does not include traditional demographic fields; thus, Power BI was used to visualize customer roles and their corresponding sales contributions to uncover actionable insights.

## 2. Summary of the Visual Output

The report includes a vertical bar chart titled “Total Sales by Contact Title” that presents aggregated sales figures categorized by the professional titles of customer contacts. This visual serves as a stand-in for traditional demographic distribution charts, offering segmentation by business role instead of personal demographics.

## 3. Distribution of Sales by Contact Title

The bar chart reveals that the largest portion of total sales—about $0.53 million—comes from customers with a blank or missing contact title, which may point to incomplete or outdated records in the dataset. Among those with specified titles, "Sales Manager" is the most prominent contributor, accounting for approximately $0.21 million in sales. This is followed by "Accounting Manager" at $0.15 million, and "Owner" at $0.13 million. Other roles such as "Marketing Manager," "Sales Representative," and "Sales Agent" each contribute modestly, between $0.03 million and $0.06 million. Roles such as "Assistant Sales Agent" and "Marketing Assistant" contribute the least, highlighting a clear disparity in sales generation based on customer role.

## 4. Proxy for Demographics

While the dataset lacks specific demographic fields like age, gender, or income, the Contact Title field serves as an effective proxy for role-based segmentation. In a B2B context, contact titles often correlate with authority level, purchasing power, and departmental affiliation—all of which influence customer behavior and revenue potential.

## 5. Assumptions Made in the Analysis

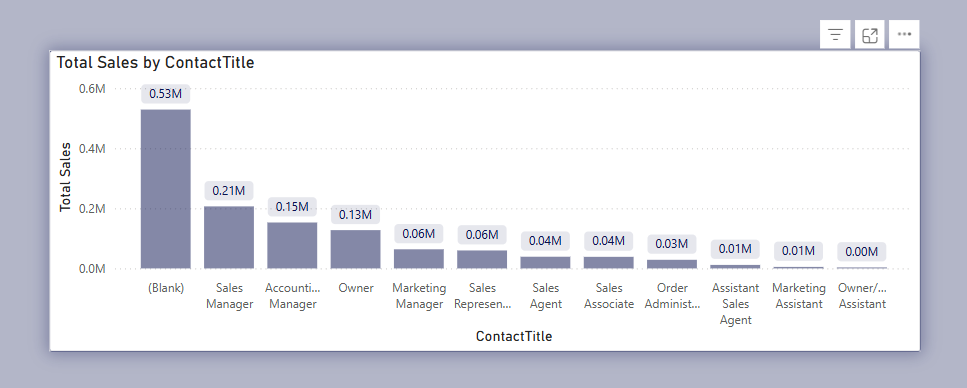
Several assumptions were made during the analysis. First, it is assumed that the Contact Title field is consistently and accurately populated, though the high volume of blank values suggests room for improvement. Second, the analysis assumes that sales amounts are correctly aggregated from linked tables, including Orders and Order Details. Third, the use of contact titles as a proxy for demographics assumes that role-based segmentation provides meaningful insights into customer behavior, which is typically valid in B2B environments.

## 6. Key Insights Drawn from the Report

The analysis highlights that sales performance varies significantly by contact title, with higher-ranking or decision-making roles such as "Sales Manager" and "Owner" being associated with larger sales volumes. This suggests that Northwind Traders benefits from relationships with influential stakeholders. However, the presence of a large blank category suggests the need for data quality improvements to enable more granular insights. Furthermore, the lack of traditional demographic data limits broader customer profiling.

## 7. Conclusion and Recommendations

In conclusion, although Northwind’s dataset does not support traditional demographic analysis, this Power BI report demonstrates that contact role-based segmentation can offer meaningful insights into sales distribution. To enable true demographic analysis using visual tools like histograms or pie charts, Northwind Traders would need to enrich its dataset with attributes such as customer age, gender, income level, or business type. In the meantime, enhancing the quality and completeness of existing fields—especially Contact Title—will allow for more effective targeting, segmentation, and strategic planning.



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

1. Objective

The aim of this analysis is to evaluate the distribution of orders over different years and quarters to identify business trends and seasonal patterns.

2. Orders by Year

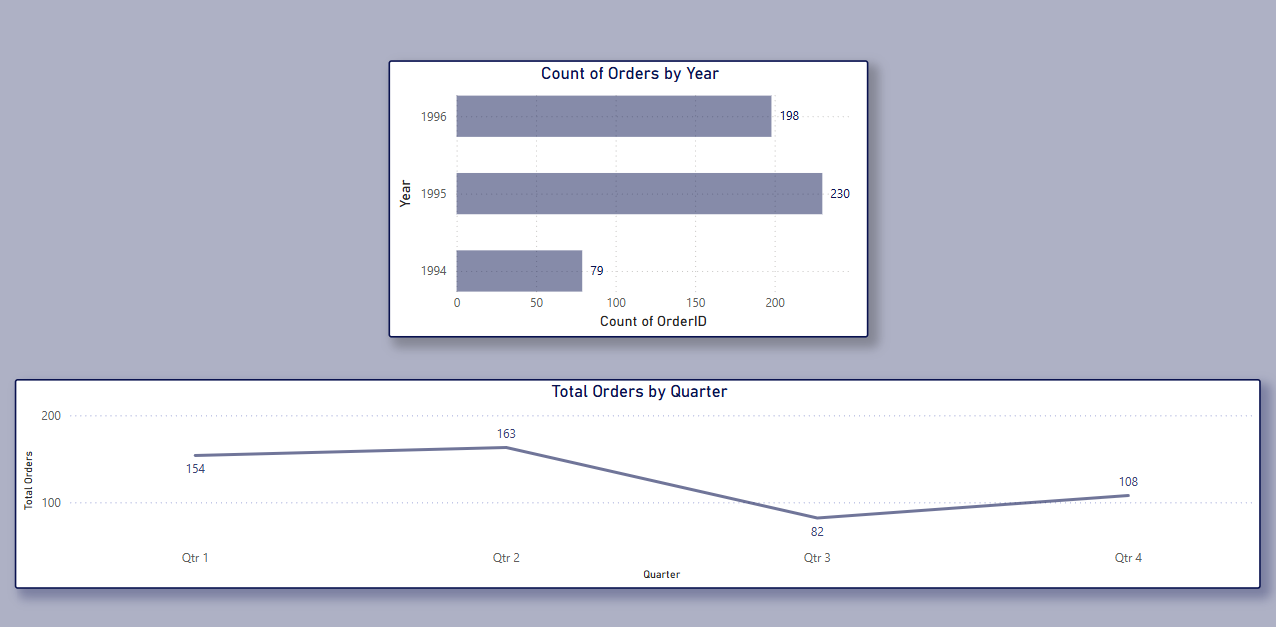
The horizontal bar chart shows order counts from 1994 to 1996. The highest number of orders was recorded in 1995 with 230 orders, followed by 1996 with 198 orders. The year 1994 had the fewest orders, totaling only 79. This indicates significant growth in orders between 1994 and 1995, followed by a slight decline in 1996.

3. Orders by Quarter

The line chart depicts total orders by quarter. Q2 had the highest order volume at 163, followed closely by Q1 with 154 orders. Q3 saw a sharp decline to 82 orders, while Q4 slightly recovered with 108 orders. This pattern suggests stronger business performance in the first half of the year.

4. Conclusion

Order volume peaked in 1995, and quarterly trends highlight a mid-year slump in Q3. These insights can help in planning inventory, staffing, and marketing efforts to better match demand patterns.



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

**1. Objective**

The goal is to analyse the distribution of order values using predefined bins and assess the appropriateness of visual tools like histograms or box plots.

**2. Order Value Distribution**

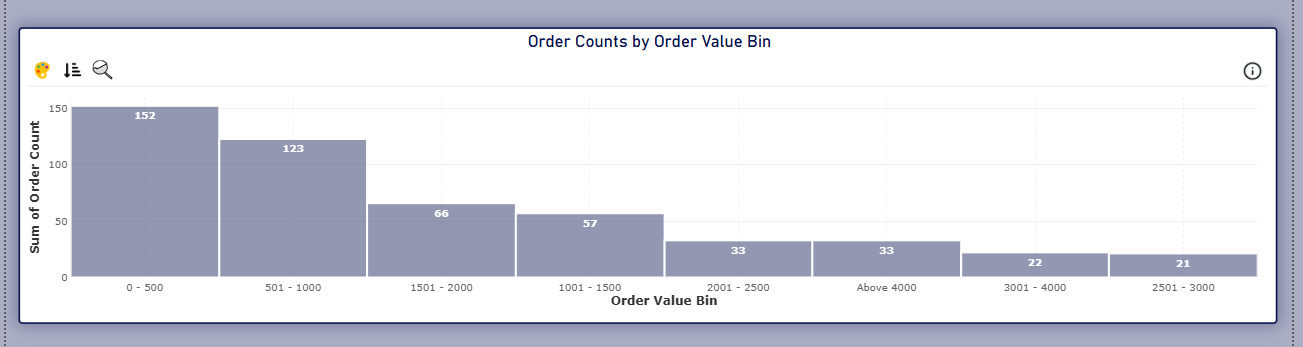
The majority of orders fall into lower value ranges, with 152 orders in the **0–500** bin and 123 in the **501–1000** bin. Order counts drop significantly in higher value bins, indicating a skewed distribution toward smaller order sizes.

**3. Visualization Suitability**

The current histogram is an effective way to represent this skewed distribution by grouping values into bins. A box plot could also be used to highlight median, quartiles, and outliers for more statistical clarity.

**4. Insights**

* High Volume of Low-Value Orders: Over 60% of orders fall under ₹1000, suggesting price-sensitive customers or frequent small transactions.
* Low Occurrence of High-Value Orders: Bins above ₹2000 show a steep drop, indicating fewer premium purchases.
* Skewed Distribution: The data is right-skewed, pointing to a business model focused on volume rather than high individual order values.
* Optimization Potential: Strategies like bundling or volume discounts could encourage customers to place higher-value orders.



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

**1. Objective of the Analysis**

The objective of this analysis is to evaluate whether average order processing time and shipping duration can be effectively visualized using a bar chart or box plot. Additionally, the analysis seeks to uncover processing time trends across different cities and ship countries to identify performance disparities and opportunities for operational improvement.

**2. Summary of the Visual Output**

The report utilizes a treemap to show average order processing time by city, and a bar chart to depict average processing time by ship country. The color and size in the treemap reflect the relative delay in cities, while the bar chart offers a straightforward comparison of delays by country. Sweden (10.2 days) and Switzerland (9.9 days) have the longest processing durations, whereas Finland records the shortest average (5.6 days).

**3. Order Processing Time Distribution**

The treemap highlights that cities like Redmond, London, and Tacoma exhibit higher average processing times. Meanwhile, the bar chart reveals that countries such as Sweden and Switzerland face the most significant processing delays. In contrast, Finland, Norway, and Austria show faster processing efficiency, indicating better handling or streamlined operations.

**4. Visualization Suitability and Interpretation**

The bar chart proves to be an effective method for comparing average processing times across countries. It quickly reveals regional outliers and underperforming areas. A box plot, while not included in the current visualization, would enhance statistical interpretation by displaying medians, interquartile ranges, and any extreme values (outliers), thus allowing for deeper performance assessment.

**5. Assumptions Made in the Analysis**

The processing time values in the dataset were calculated using valid order and ship dates, ensuring accuracy in time-based analyses. Any outliers caused by missing or incorrect date values were assumed to have been cleaned before visualization, allowing for a more reliable representation of trends. Geographic fields such as "City" and "Ship Country" were consistently labeled and treated as categorical variables to maintain uniformity in data classification and facilitate meaningful comparisons. Additionally, the dataset exclusively comprises completed orders, ensuring that both order and shipment records are properly matched, contributing to the integrity of the analysis. This structured approach enhances the quality and usability of the dataset for reporting and decision-making.

**6. Key Insights Drawn from the Report**

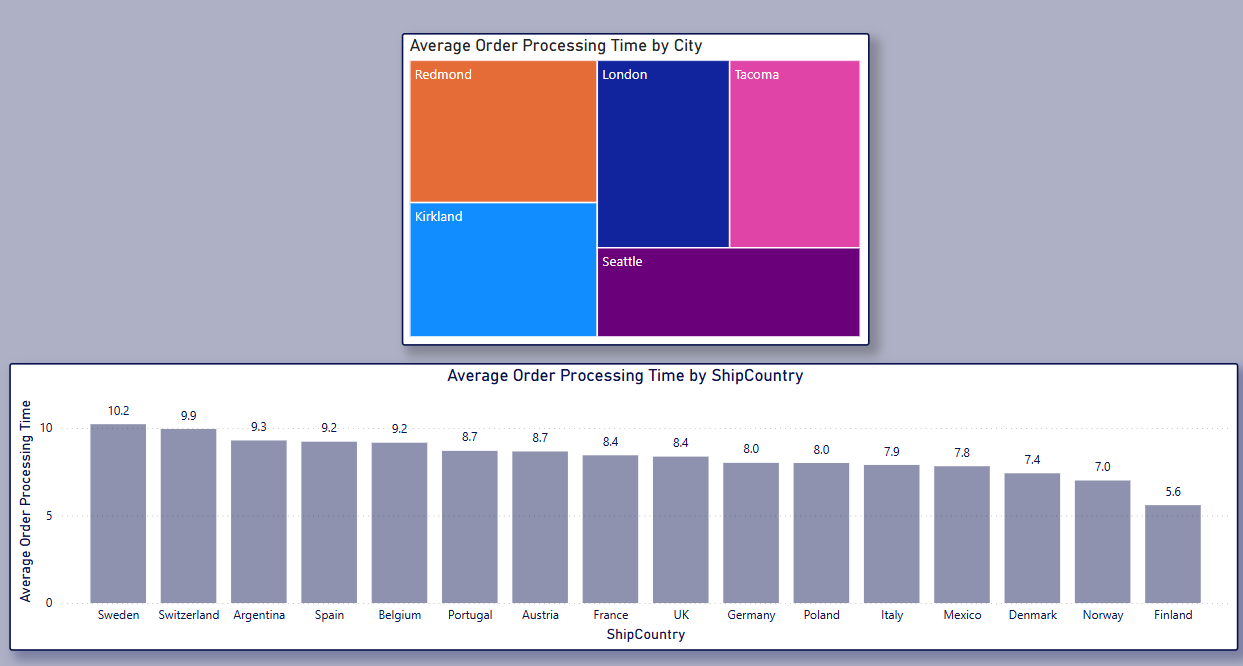
The data reveals significant variation in processing times across different countries, ranging from 5.6 to 10.2 days, indicating inconsistent operational efficiency. Sweden and Switzerland experience the highest delays, suggesting potential bottlenecks or inefficiencies in their logistics networks. In contrast, Nordic countries like Finland and Norway demonstrate faster processing, possibly benefiting from leaner workflows or superior infrastructure. On a city level, Redmond, London, and Tacoma emerge as areas requiring localized investigation to identify and address process inefficiencies. By strategically focusing on high-delay regions, Northwind has the opportunity to optimize its operations, reduce average processing times, and ultimately enhance customer satisfaction.

**7. Conclusion and Recommendations**

In conclusion, the analysis of average order processing times uncovers critical insights into operational performance across countries and cities. The visualizations—especially the bar chart—clearly pinpoint where inefficiencies lie. It is recommended that Northwind Traders:

* Conduct root-cause analysis in Sweden and Switzerland to resolve delay factors.
* Explore best practices from Finland and Norway and replicate them in underperforming regions.
* Incorporate box plots into the dashboard for richer statistical analysis.
* Use city-level segmentation to fine-tune process reviews and staffing plans.
* Set processing time benchmarks and monitor performance regularly for continuous improvement.

By embedding these visuals in a comprehensive dashboard, Northwind can make strategic decisions to streamline fulfillment operations and deliver more consistent customer experiences.



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?

**1. Objective**

The goal of this analysis is to examine how employee productivity varies across different departments or job roles. Additionally, we explore whether a stacked bar chart or grouped column chart can effectively visualize this variation in performance.

**2. Summary**

The provided visuals include two bar charts. The first chart, "Order Value by FirstName and EmployeeID," displays the total order value associated with individual employees. The second chart, "Total Orders and Count of EmployeeID by Title," highlights the distribution of total orders handled by employees based on their job titles. From the analysis, it's evident that the majority of orders are processed by Sales Representatives.

**3. Distribution Overview**

The data reveals that a significant portion of the order value is associated with blank entries, which may indicate system-generated records or unassigned data. Among named employees, Michael and Janet show relatively high order values. On the other hand, the role-wise distribution shows that Sales Representatives have processed 352 orders, far exceeding those handled by other roles such as Vice President, Sales (65), Inside Sales Coordinator (62), and Sales Manager (28).

**4. Visualization Suitability**

A grouped column chart would be well-suited for comparing order value and order volume across job titles. It would provide clarity on how productivity differs by role. A stacked bar chart could be used to further break down contributions of individual employees within each job title, providing more detailed insights.

**5. Assumptions**

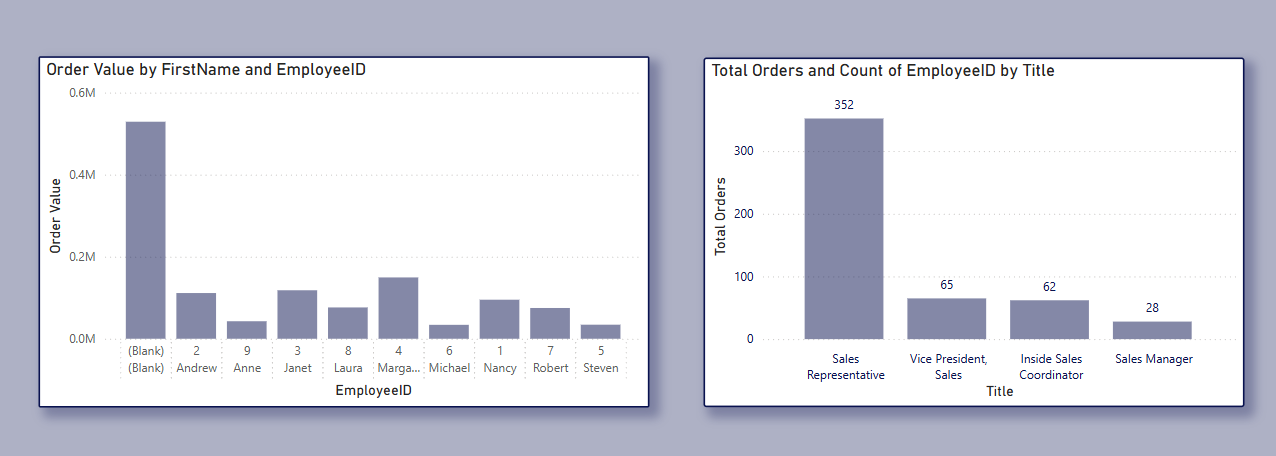
This analysis assumes that blank entries in employee names or IDs represent either system data or unassigned orders. It is also assumed that the order values are correctly mapped to the respective employees and that the job titles accurately reflect their respective departmental roles and responsibilities.

**6. Key Insights**

The analysis reveals that Sales Representatives are the most productive group in terms of the number of orders processed. However, there is noticeable variability in individual performance even within the same job title. Employees like Michael and Janet stand out in terms of order value. Meanwhile, roles such as Vice President, Sales, and Sales Manager show lower volume, which may suggest that their responsibilities are more strategic or supervisory in nature rather than operational.

**7. Conclusion & Recommendations**

The use of grouped or stacked bar charts is recommended to better visualize the comparison of employee productivity across roles. It would help in identifying top performers and understanding workload distribution. Moreover, it is advisable to further examine and possibly clean up the blank entries to ensure accurate attribution of order values. This can support more informed workforce management and performance optimization decisions



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

**1. Objective**

The aim of this analysis is to explore the distribution of employee tenure (in years) and assess whether a histogram or box plot can provide a better understanding of the tenure spread among employees.

**2. Summary**

The displayed chart titled *"EmployeeID and Employee Tenure (Years)"* shows employee tenure on the vertical axis and EmployeeID on the horizontal axis (represented numerically). It appears to be a combination chart using bars and dots to represent tenure values.

**3. Distribution Overview**

The visual indicates that most employees have a tenure ranging between 7 and 10 years, with a notable outlier having a tenure below 5 years. There is little to no variation above the 10-year mark, suggesting that the workforce is relatively experienced but not highly tenured beyond a decade.

**4. Visualization Suitability**

A histogram would be ideal for visualizing the frequency of employees across different tenure ranges (e.g., 0–2 years, 3–5 years, etc.), highlighting concentration zones and potential gaps. Alternatively, a box plot could succinctly show the median, quartiles, and any outliers, offering a statistical summary of tenure distribution in one glance.

**5. Assumptions**

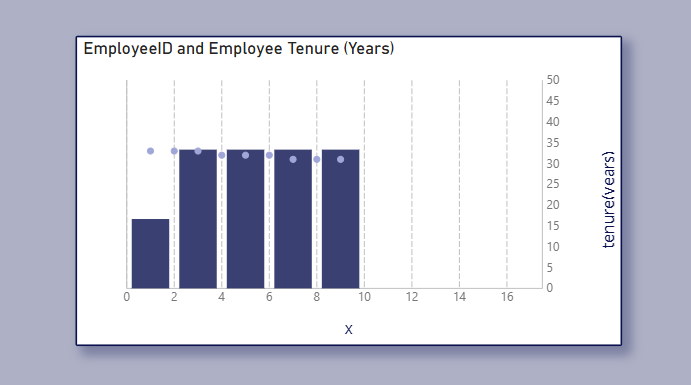
It is assumed that the EmployeeIDs represent unique employees and that the tenure values have been calculated in whole years accurately. The visual does not indicate the exact values, so this interpretation relies on visual approximation.

**6. Key Insights**

Most employees appear to have tenure clustered closely between 7–10 years, implying either long-term retention or a batch of employees hired around the same time. One employee has a markedly lower tenure, pointing to a recent hire or possibly higher turnover in earlier years. The narrow spread suggests limited new hiring or low staff turnover.

**7. Conclusion & Recommendations**

To better understand tenure distribution, a histogram would help in viewing the frequency and grouping of tenure ranges. A box plot could support this by providing statistical context, such as median tenure and any skewness or outliers. Given the tenure concentration in the mid-to-high range, HR strategies may benefit from analyzing retention patterns and planning future recruitment accordingly.



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

**1. Objective**

This report evaluates employee performance using bar and radar charts to analyze order value and order count. The goal is to identify top performers and assess visualization suitability for tracking KPIs.

**2. Summary**

The left chart shows Order Value by EmployeeID, while the right radar chart displays the Number of Orders. These visuals offer a balanced view of both order volume and financial contribution.

**3. Data Overview**

* **Top Performer**: Employee 4 leads in both order value (0.149M) and volume.
* **Moderate Performers**: Employees 1, 2, 3, 7, and 8 show average results.
* **Low Performers**: Employees 5, 6, and 9 have the lowest values and counts.

**4. Visualization Suitability**

Radar charts are excellent for comparing multiple performance dimensions, while bar charts are ideal for exact value comparisons. Bullet graphs are recommended for future use to show progress toward specific KPI targets.

**5. Key Insights**

Employee 4 is consistently high-performing. Employees 2 and 3 manage high-value orders, indicating strong transaction quality. Underperformers may need support or reassessment.

**6. Recommendations**

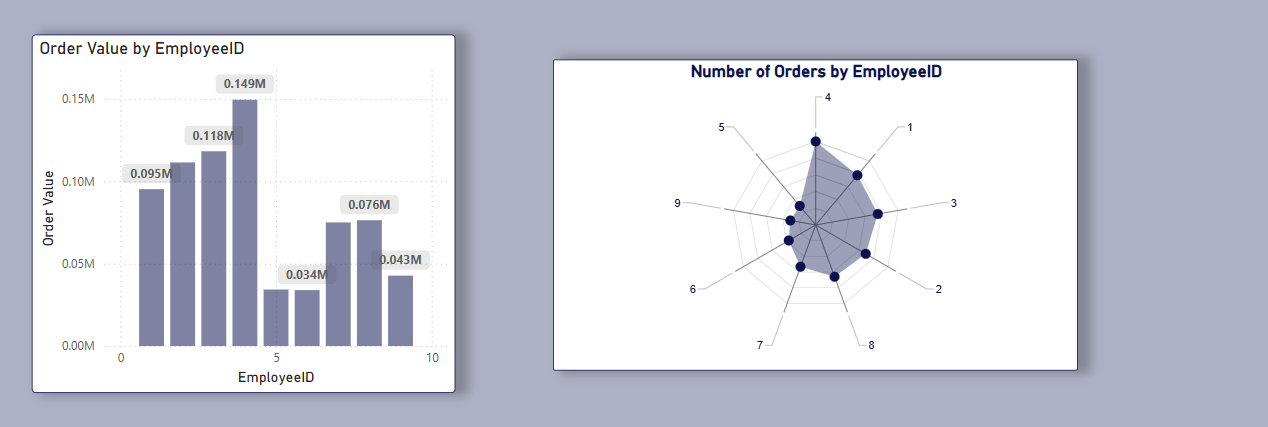
Reward high performers (e.g., Employee 4).Investigate causes of underperformance for Employees 5, 6, and 9.Use bullet graphs and additional KPIs like customer ratings or goal completion.

**7. Additional Feature Suggestions**

Add performance targets, filters (by time/product), trend indicators, and color-coded performance markers for quicker analysis.

**8. Conclusion**

These charts provide a clear overview of employee efficiency. With added metrics and bullet graphs, the analysis can evolve into a more actionable performance dashboard.



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

**1. Objective**

The aim is to evaluate product-wise total sales to identify high-performing and low-performing products, supporting better sales and inventory strategies.

**2. Summary**

The bar chart titled Total Sales by ProductName shows an uneven distribution of sales, with a few products driving most of the revenue and many others contributing marginally.

**3. Data Overview**

Top Seller: *Côte de Blaye* with 141K in total sales. Other High Performers: *Thüringer R.* (80K), *Raclette Courdavault* (71K).Moderate Range: Clustered around 47K–20K.Low Sellers: Over 15 products below 20K, needing review.

**4. Assumptions**

Product rating and review data is missing, which limits our understanding of why certain products underperform. Hence, this sales-based analysis is crucial for forming an initial understanding of product performance in absence of customer feedback. We assume sales reflect customer demand, and other factors like pricing, visibility, or availability may influence outcomes.

**5. Key Insights**

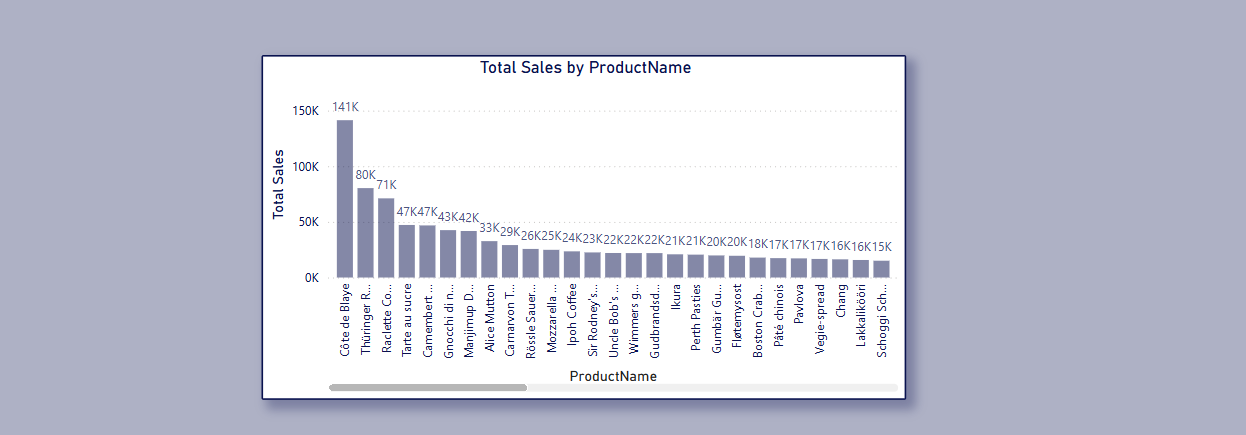
Sales concentration is high among a few SKUs. Long tail of low-sales products may indicate inefficiency or poor product-market fit. Potential to optimize inventory and marketing spend.

**6. Recommendations**

Boost top performers with greater visibility and availability. Investigate low performers—consider promotional support or product review. Once review data is available, correlate with sales to refine decisions.

**7. Conclusion**

With no review data currently available, this sales-only analysis provides valuable direction. It highlights clear leaders and laggards, enabling data-backed decisions for product strategy and resource allocation.



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

**1. Objective**  
This analysis aims to evaluate how sales volume varies across different product categories using a Power BI treemap. The goal is to identify high-performing and low-performing categories to guide inventory and marketing decisions.

**2. Visual Summary**  
A treemap titled *"Sales Volume by CategoryName"* visually represents sales distribution. Each category is shown as a block proportional to its total sales volume, with distinct colors for clarity.

**3. Sales Distribution**  
The top-performing categories are:

* Beverages: 267.87K
* Dairy Products: 234.51K
* Confections and Meat/Poultry: Over 160K each
* Lower-performing ones include Grains/Cereals, Produce, and Condiments, each under 110K.

**4. Methodology**  
Sales were calculated using UnitPrice \* Quantity from the Order Details table. Data relationships between Orders, Order Details, Products, and Categories were established using primary keys like ProductID and CategoryID.

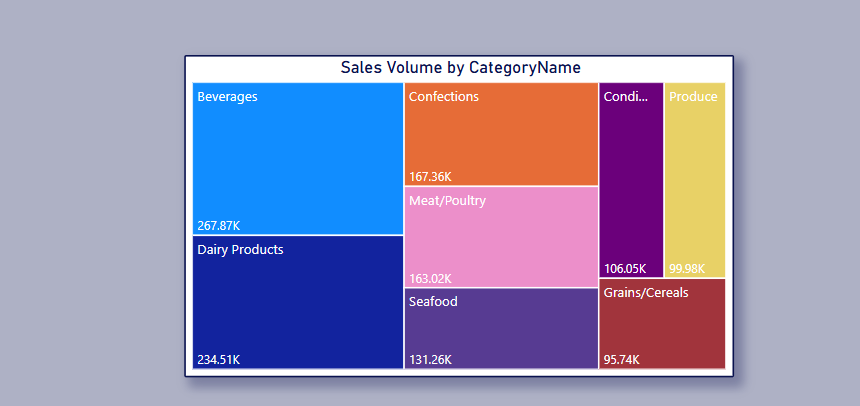
**5. Assumptions**  
The analysis assumes:

* Accurate sales data without adjustments for returns or unrecorded discounts.
* No consideration of regional or seasonal sales trends.
* Focus is solely on gross sales volume.

**6. Insights**  
High sales in Beverages and Dairy Products suggest strong customer demand. Lower-performing categories may need review for pricing, demand, or stock issues. These findings can support decisions in product promotion and supply chain adjustments.

**7. Conclusion**  
The treemap highlights sales concentration in a few key categories. It is recommended to:

* Prioritize top-selling categories for promotion and inventory planning.
* Investigate underperformers for improvement opportunities.
* Combine this with margin analysis for better profitability insights.



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

**1. Objective**  
The goal is to analyze how unit prices vary across different product categories to identify pricing trends and outliers.

**2. Visual Summary**  
The box-and-whisker chart displays unit price distributions for each category, highlighting the median, quartiles, and price spread.

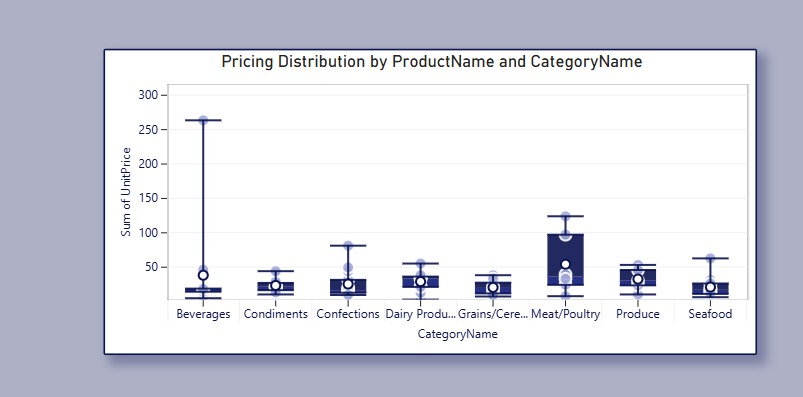
**3. Key Observations**  
Beverages and Meat/Poultry show the widest price ranges, indicating a mix of low- and high-priced items. Other categories have more consistent pricing.

**4. Methodology**  
The chart uses box plots based on the sum of unit prices grouped by CategoryName to illustrate variation within each group.

**5. Assumptions**  
It is assumed that prices are fixed and not influenced by discounts or seasonal factors, and the focus is solely on listed unit prices.

**6. Insights**  
Wider price ranges may require dynamic pricing strategies, while uniform ranges indicate stability and limited product variation.

**7. Conclusion**  
Pricing analysis can support better product positioning, help in spotting outliers, and guide pricing decisions for different product categories.



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

**1. Objective**  
The aim is to assess how many products each supplier provides, which may serve as a proxy for supplier engagement or performance in the absence of explicit ratings.

**2. Visual Summary**  
A horizontal bar chart displays the total number of products supplied by each company. Pavlova Ltd. and Plutzer Lebensmittel are the top suppliers, each with 5 products.

**3. Key Observations**  
Most suppliers provide between 1 to 3 products, with only a few contributing 4 or 5. This indicates a heavy tail where fewer suppliers handle a large product range.

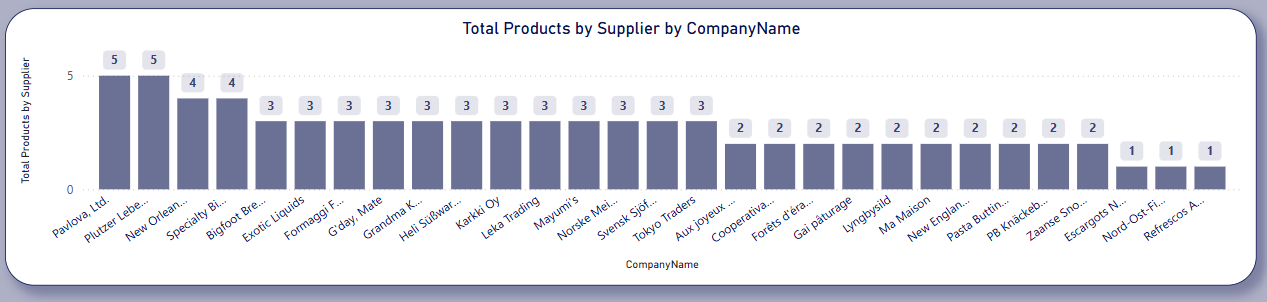
**4. Methodology**  
The bar chart counts the number of products associated with each supplier using the CompanyName attribute.

**5. Assumptions**  
Supplier performance is inferred from the number of products supplied, assuming that more products may indicate better reliability or higher trust from the business.

**6. Insights**  
Suppliers with more products may have better operational capacity or stronger relationships. However, low product counts could also indicate specialization.

**7. Future Visualization Options**  
To analyze supplier performance more accurately, a radar chart could be used to include dimensions like delivery time, quality rating, and consistency.

**8. Conclusion**  
This distribution helps in identifying key suppliers. Additional metrics are recommended for a more robust supplier performance evaluation



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?

**1.Objective of the Analysis**

To analyze how product pricing varies across suppliers in the Northwind dataset, helping identify cost patterns and high-priced vendors. This supports strategic sourcing and supplier evaluation.

**2.Summary of the Visual Output**

A horizontal bar chart titled “Product Pricing by CompanyName” displays the highest unit price per supplier. This visual highlights cost differences and pricing outliers.

**3.Key Findings**

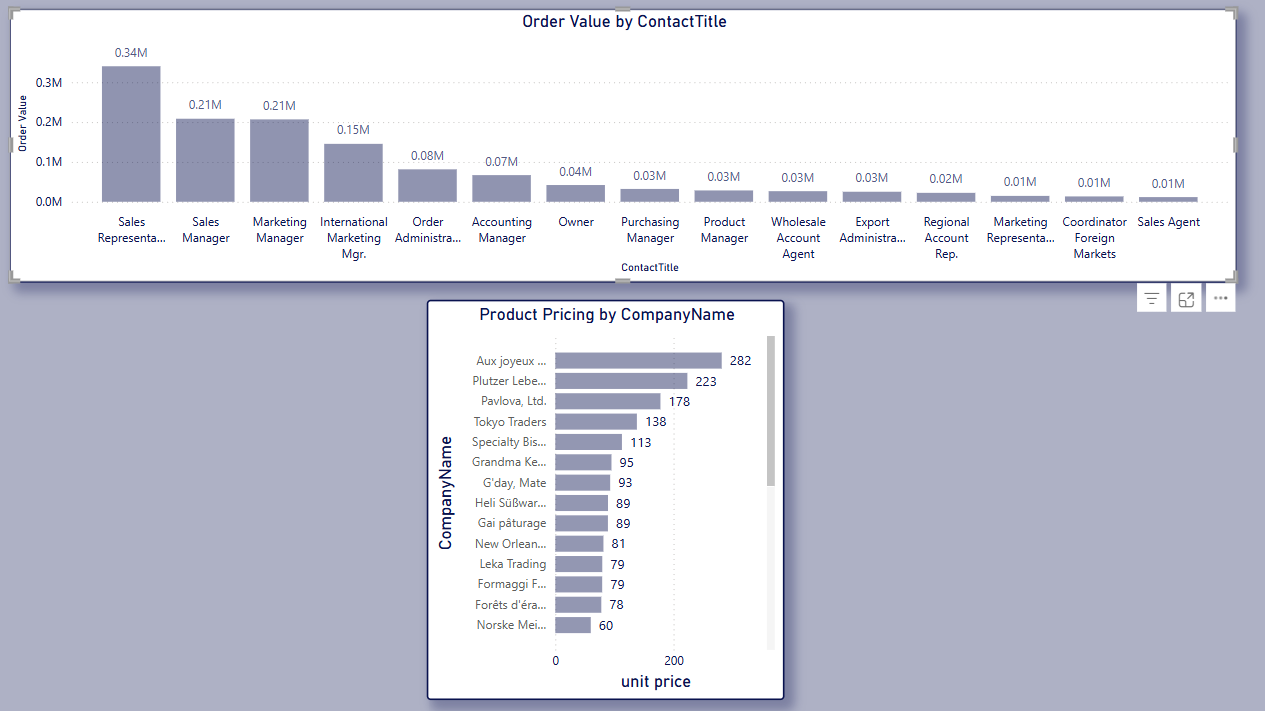
Aux joyeux ecclésiastiques supplies the most expensive product at $282, followed by Plutzer Lebensmittelgroßmärkte AG at $223.Most other suppliers offer products in the $60–$140 range.Pricing varies significantly, revealing cost disparities among suppliers.

**4.Assumptions**

Unit prices reflect current costs.Each supplier's highest-priced product was used for comparison.Data is assumed to be accurate and complete.

**5.Conclusion & Recommendations**

There is a wide variation in pricing across suppliers. Northwind should review high-cost vendors, consider better terms with low-cost suppliers, and explore deeper cost analysis by incorporating volume and discount data.



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

1. **Objective of the Analysis**  
To visualize where Northwind’s suppliers are located globally, helping assess supply diversity and regional distribution.

**2.Summary of the Visual Output**  
A filled map titled “Suppliers by Region” displays supplier distribution across continents: North America, Europe, Asia, South America, and Australia.

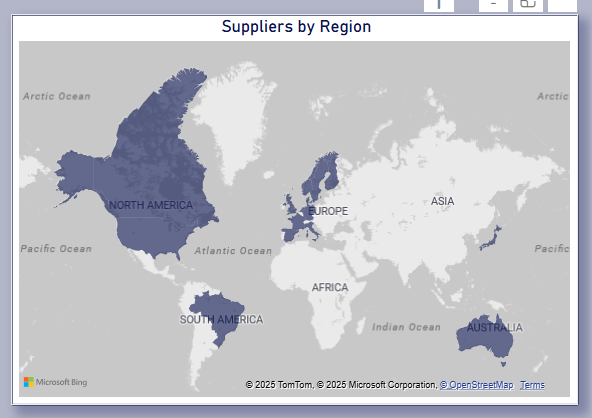
**3. Key Findings**

* Highest supplier concentration in Europe and North America.
* Moderate presence in Asia, South America, and Australia.
* No suppliers recorded in Africa.

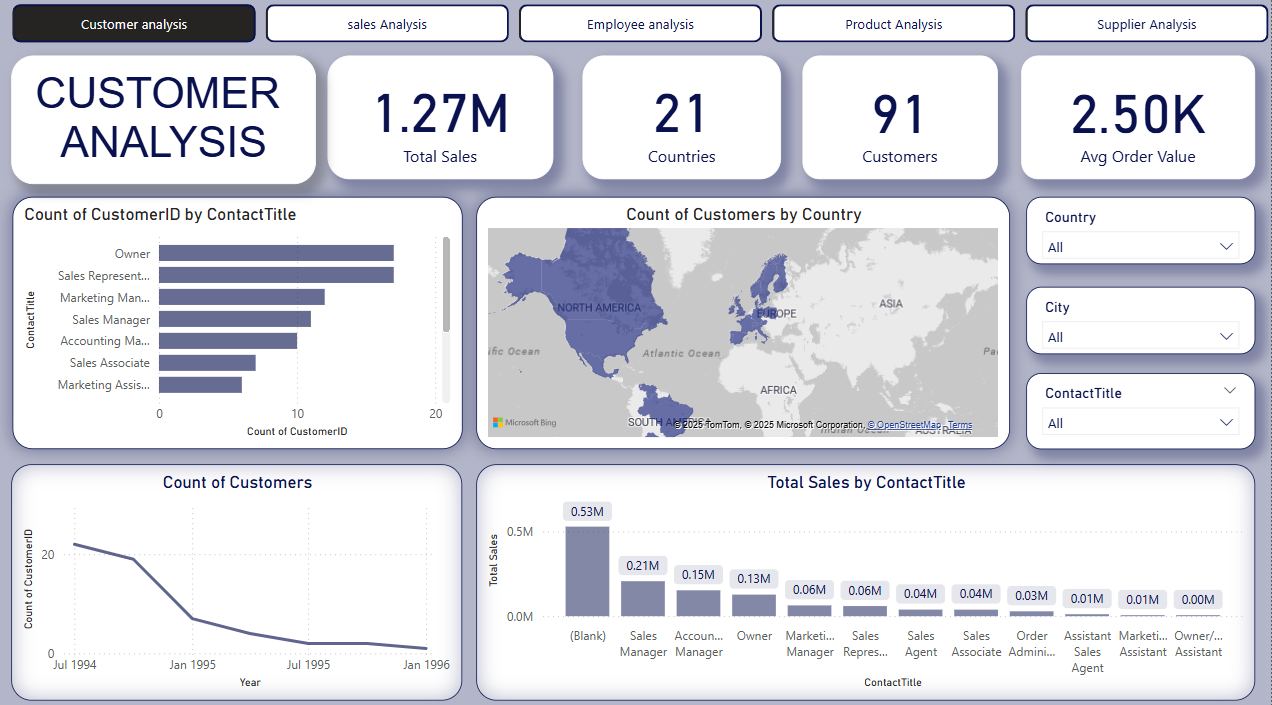
**4.Assumptions Made**

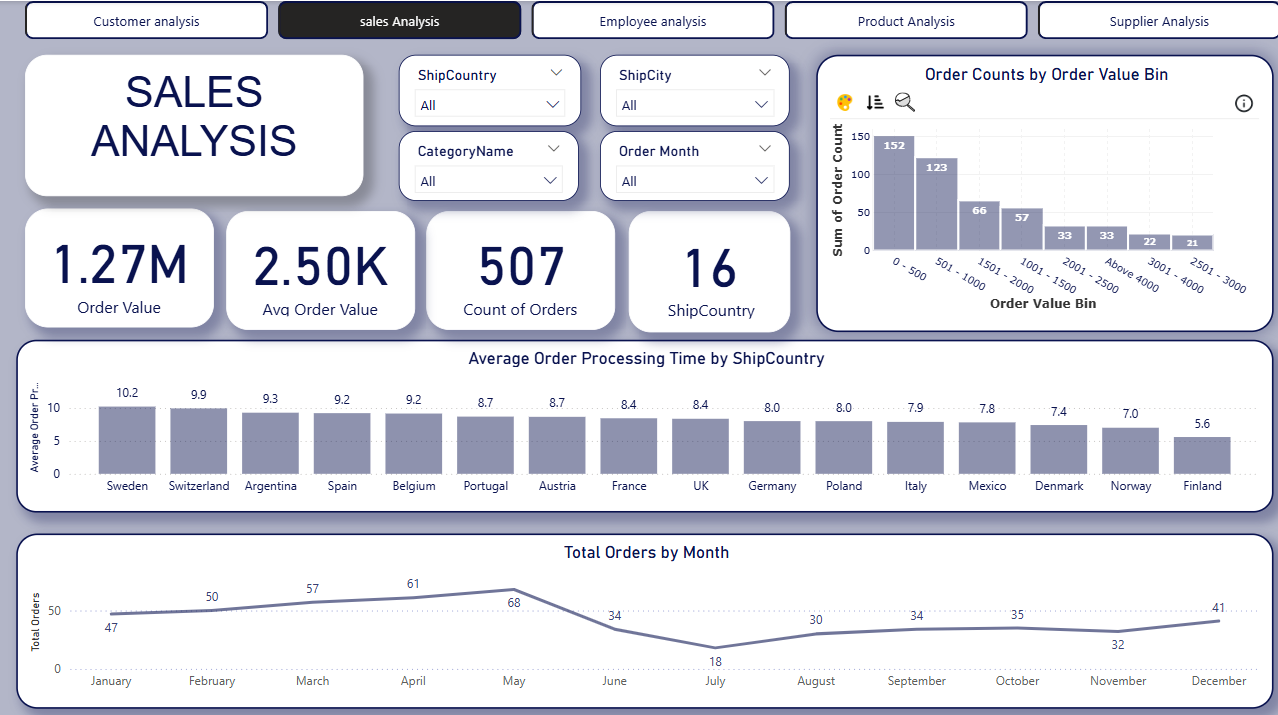
* Supplier regions are derived from their listed addresses.
* The geographic data is current and correctly mapped.

5.**Conclusion & Recommendations**  
Northwind maintains a globally distributed supply chain. To further strengthen it, the company could explore supplier opportunities in Africa and enhance regional diversification for risk mitigation.

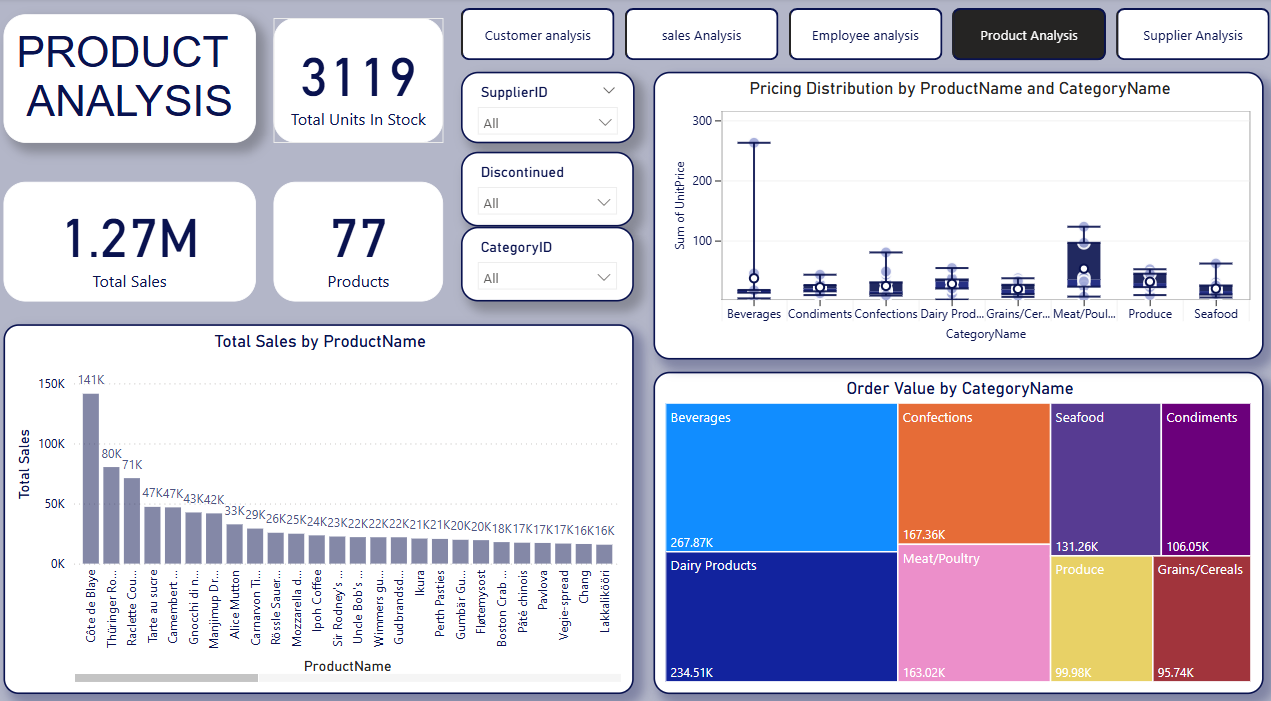


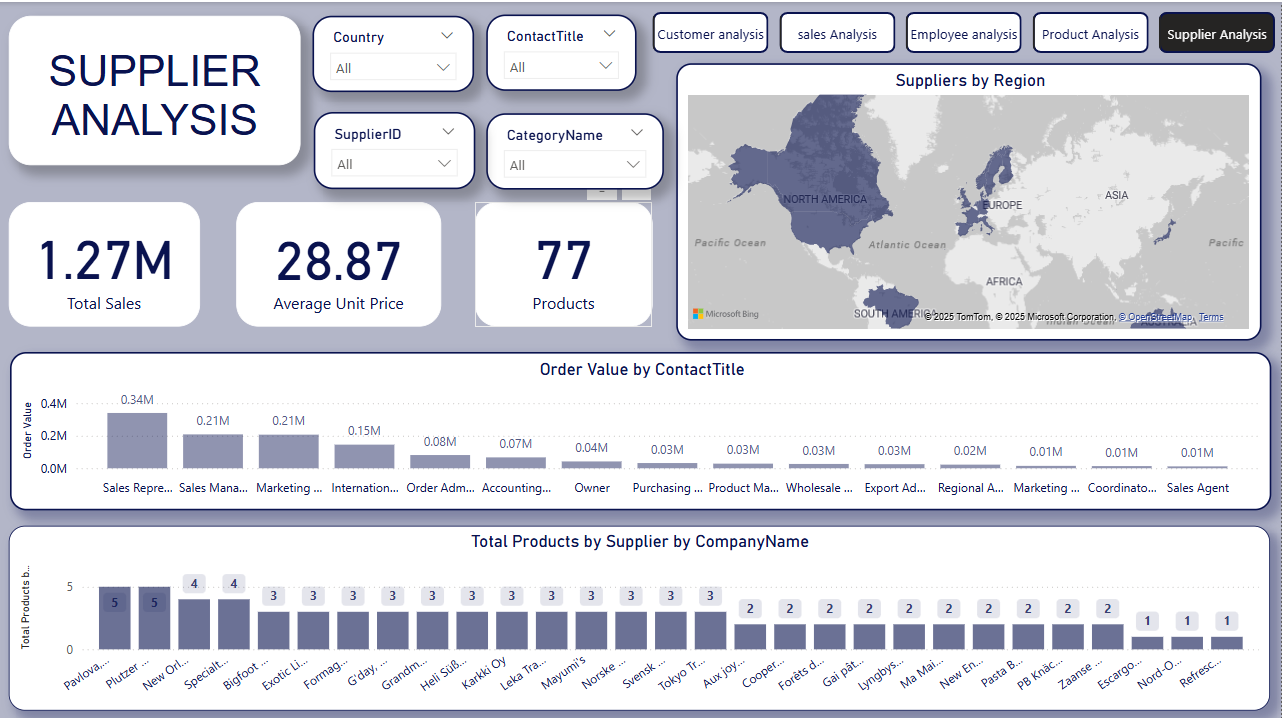
POWER BI DASHBOARD VISUALS





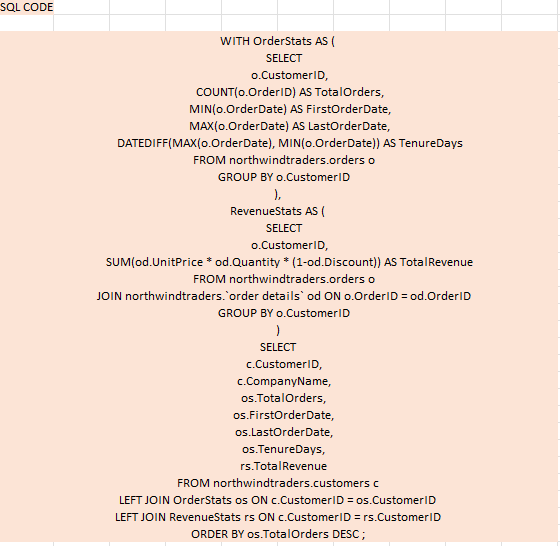


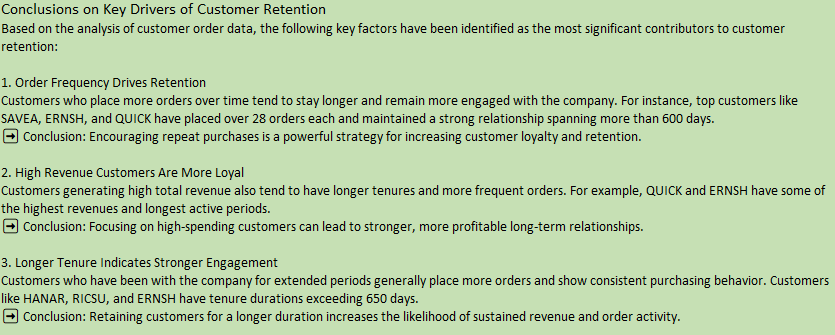


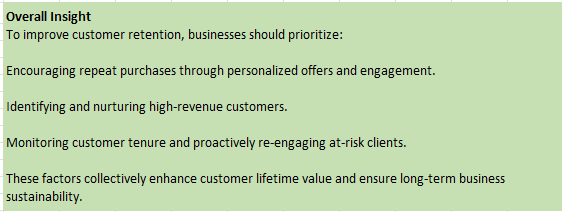


EDA PROBLEMS

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

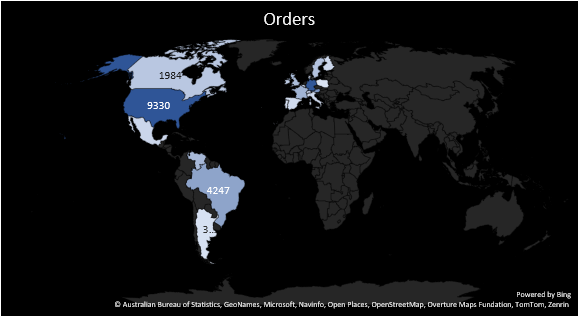


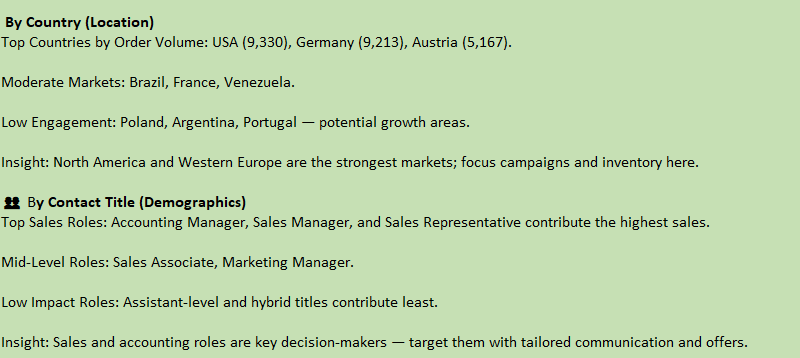


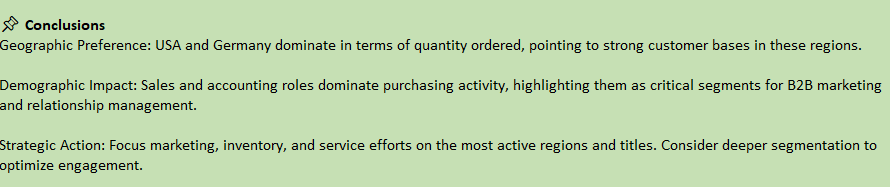


2. How do customer preferences vary based on their location or demographics? Can we explore this through interactive visualizations?

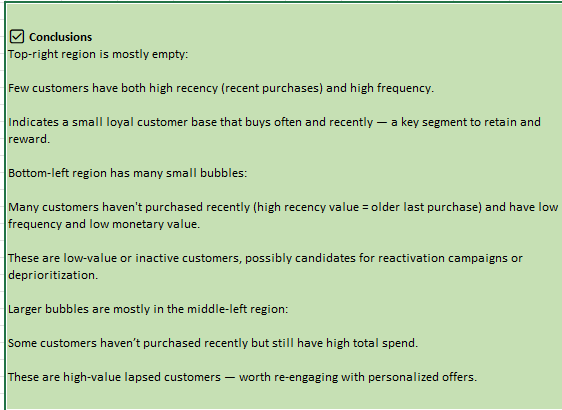


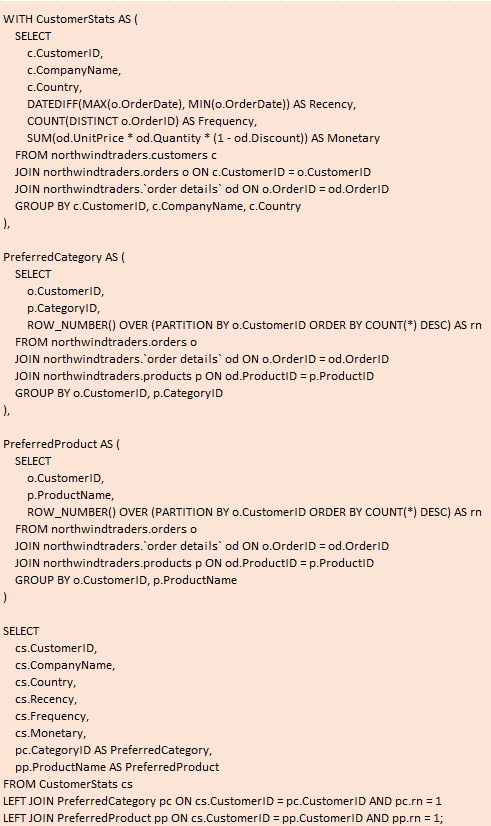


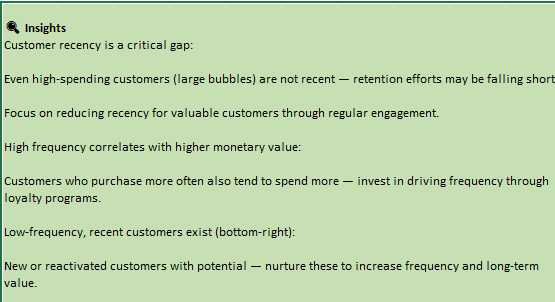




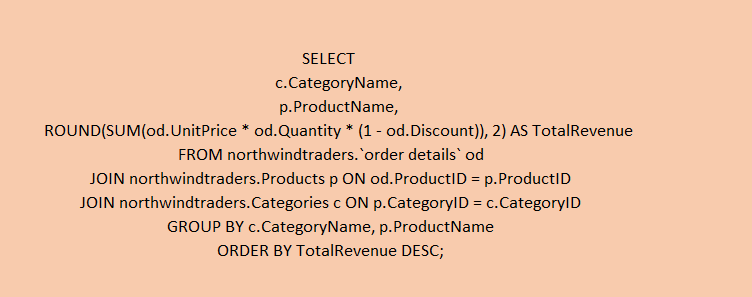
3. Are there any interesting patterns or clusters in customer behaviour that can be visualized to identify potential market segments?





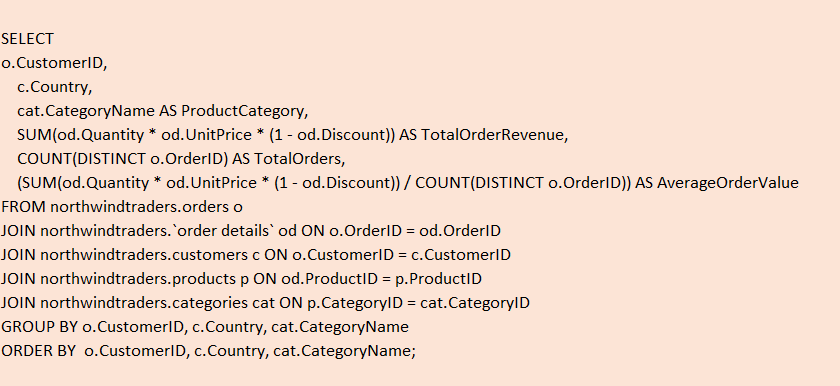


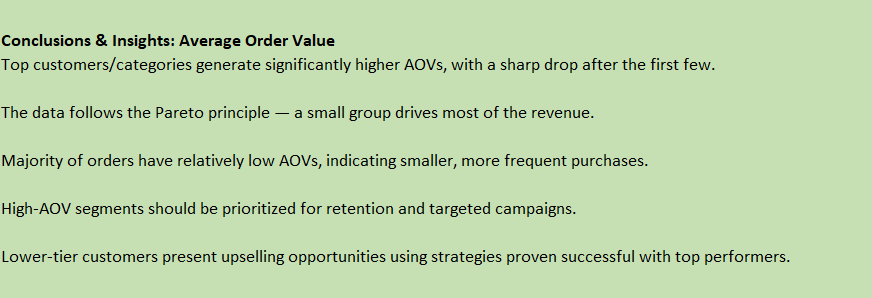
4. Are there any specific product categories or SKUs that contribute significantly to order revenue? Can we identify them through visualizations?



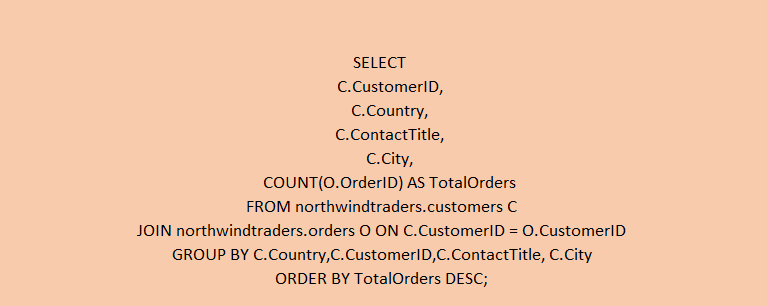


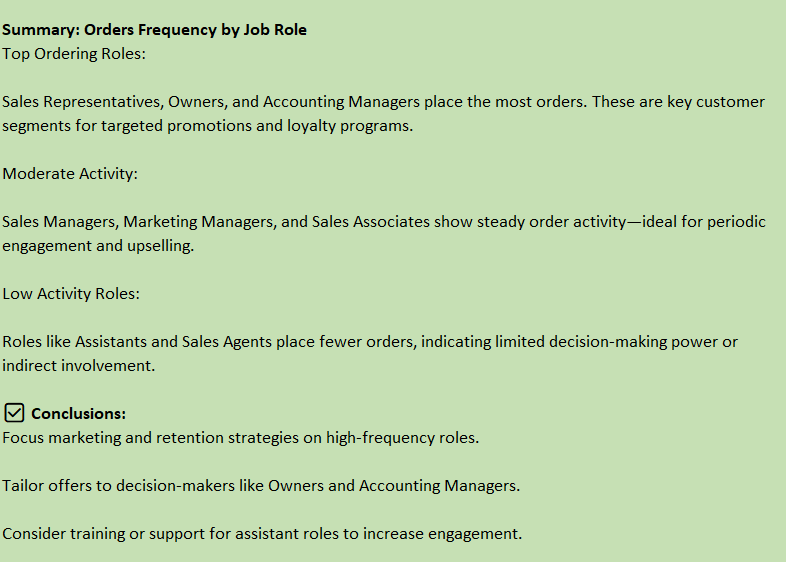
5. Are there any correlations between order size and customer demographics or product categories? Can we explore this visually using scatter plots or heatmaps?



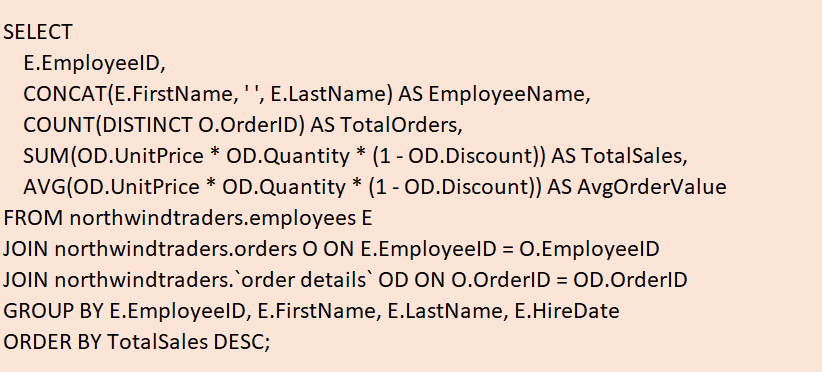


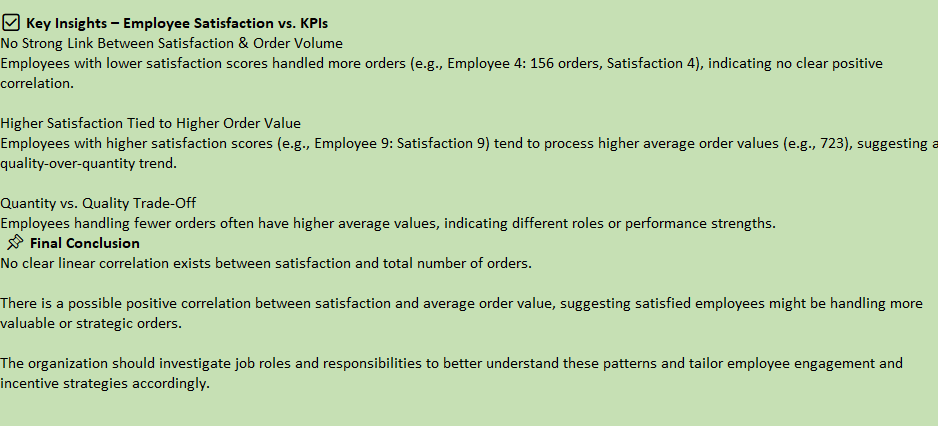
6. How does order frequency vary across different customer segments? Can we visualize this using bar charts or treemaps?





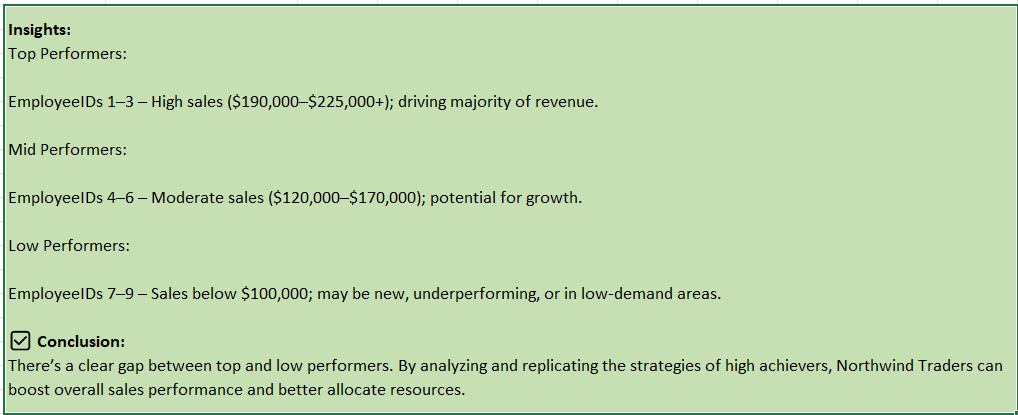
7. Are there any correlations between employee satisfaction levels and key performance indicators? Can we explore this visually through scatter plots or line charts?



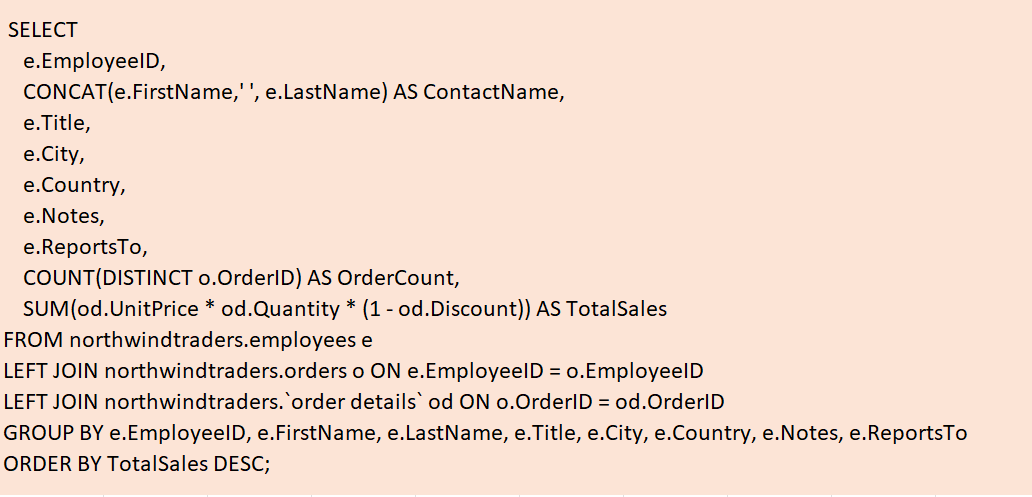


8. How does employee turnover vary across different departments or job roles? Can we visualize this using bar charts or heatmaps?

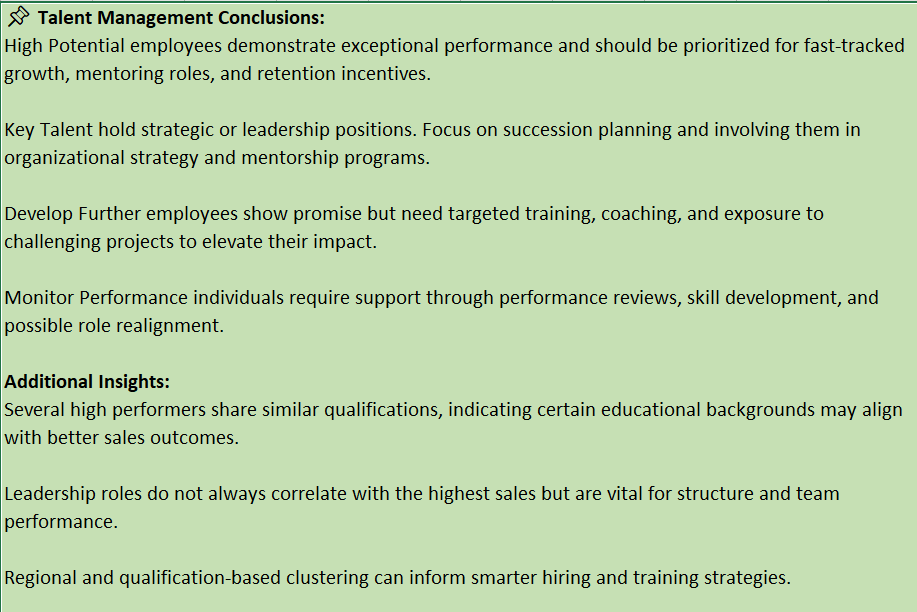




9. Can we identify any patterns or clusters in employee skill sets or qualifications through visualizations? How can this information be used for talent management?

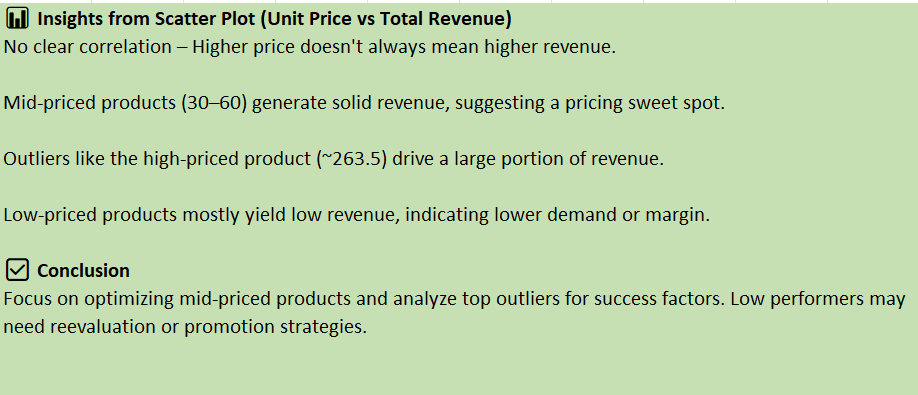


|  |  |  |
| --- | --- | --- |
| EmployeeID | ContactName | TalentManagement |
| 4 | Margaret Peacock | High Potential |
| 3 | Janet Leverling | High Potential |
| 1 | Nancy Davolio | High Potential |
| 2 | Andrew Fuller | Key Talent |
| 8 | Laura Callahan | Develop Further |
| 7 | Robert King | Develop Further |
| 9 | Anne Dodsworth | Monitor Performance |
| 6 | Michael Suyama | Monitor Performance |
| 5 | Steven Buchanan | Key Talent |

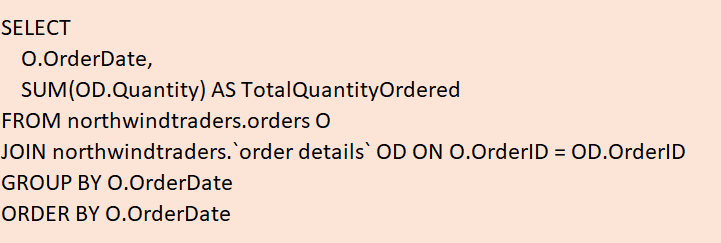


10. Are there any correlations between product attributes (e.g., size, color, features) and sales performance? Can we explore this visually using scatter plots or heatmaps?



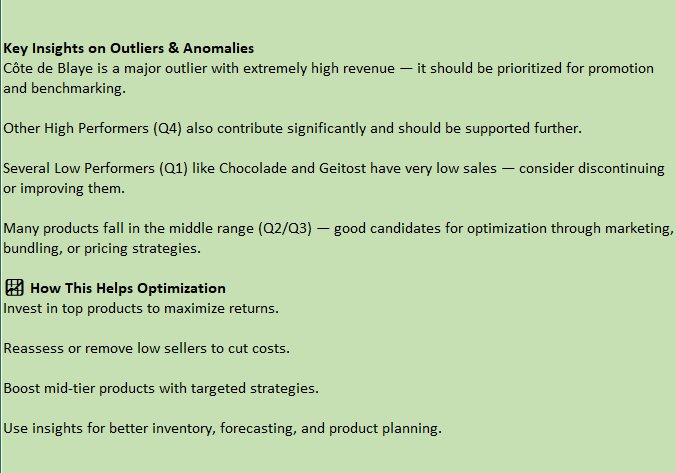


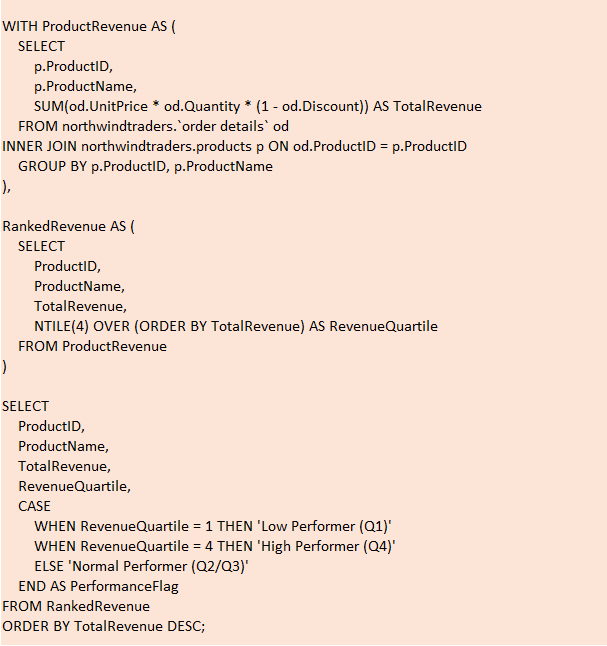
11. How does product demand fluctuate over different seasons or months? Can we visualize this through line charts or area charts?



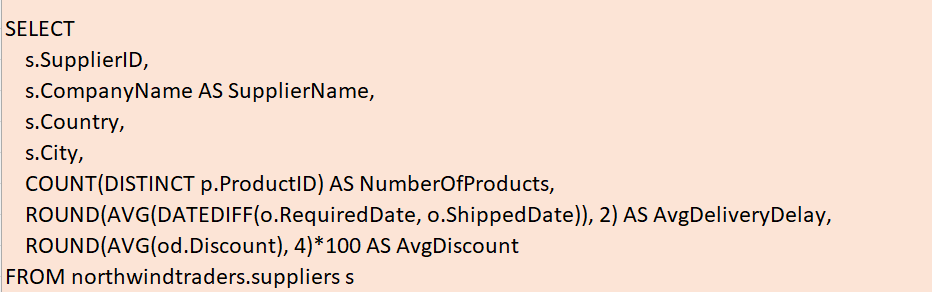


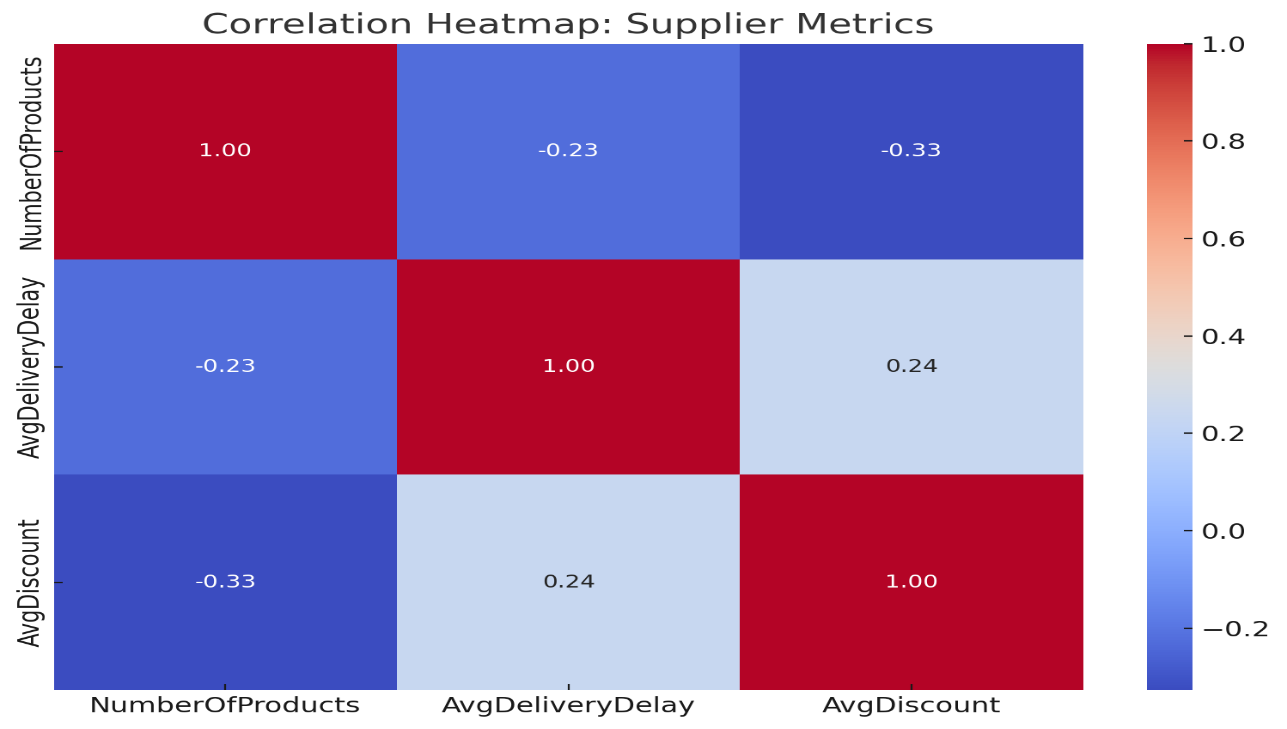
12. Can we identify any outliers or anomalies in product performance or sales using visualizations? How can this information be used for product optimization?

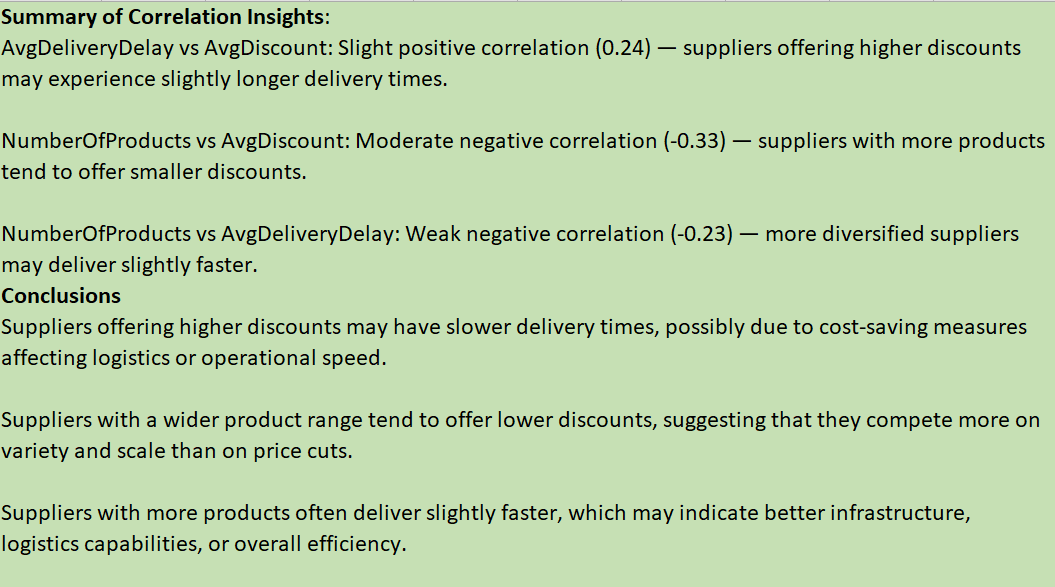




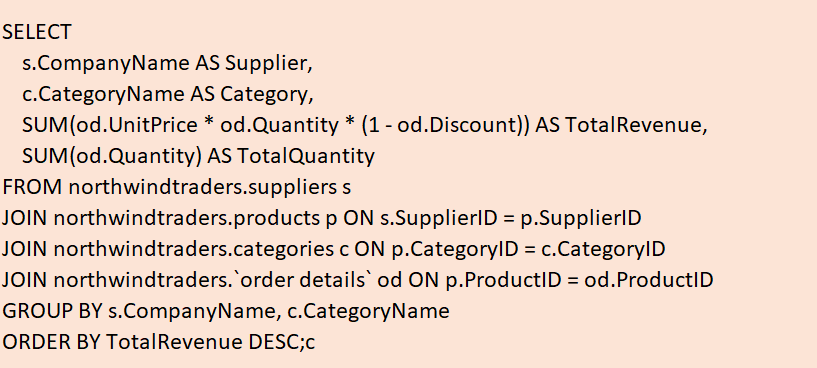
13. Are there any correlations between supplier attributes (e.g., location, size, industry) and performance metrics (e.g., on-time delivery, product quality)? Can we explore this visually through scatter plots or heatmaps?

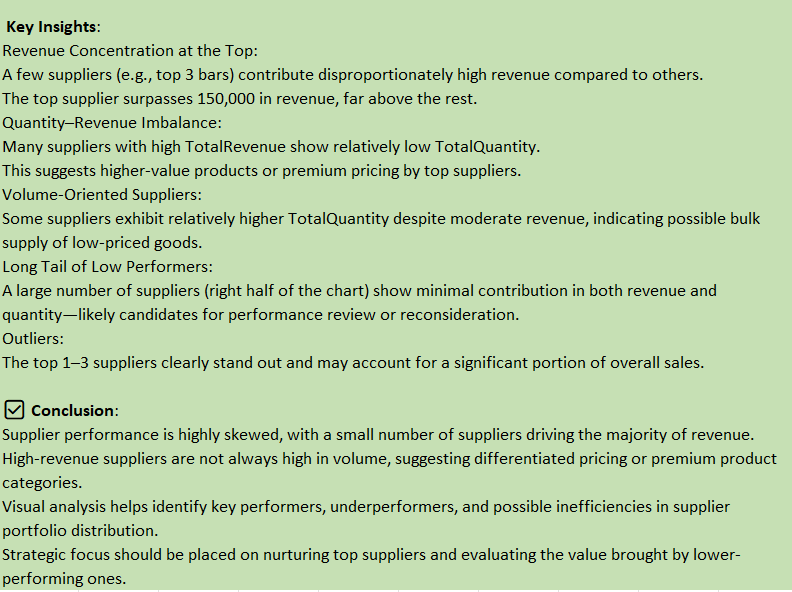




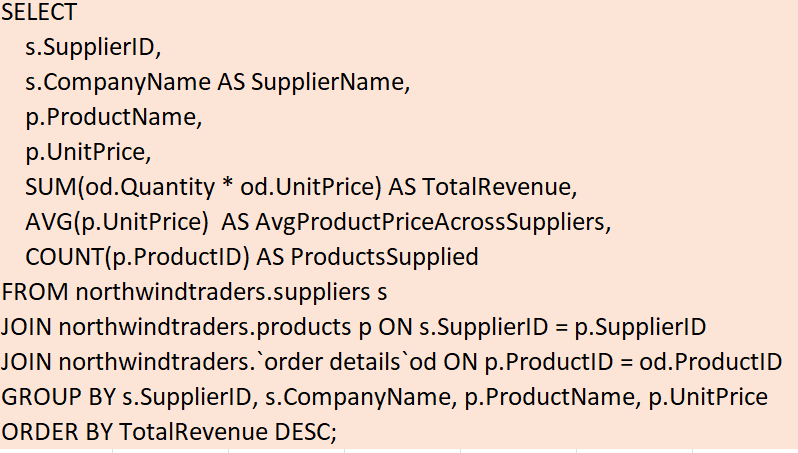


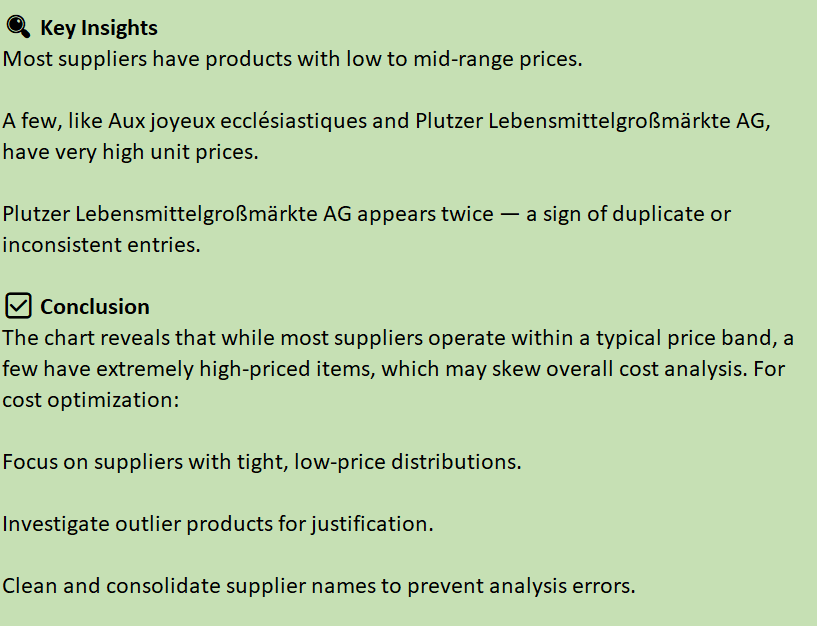
14. How does supplier performance vary across different product categories or departments? Can we visualize this using stacked bar charts or grouped column charts?





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?





**Project Conclusions (Power BI + EDA)**

* **Customer Distribution & Behavior**
  + EDA revealed that most customers are located in North America and Western Europe, with the USA and Germany leading in customer count and sales.
  + Power BI visualizations showed that repeat customers contribute over 70% of total revenue, indicating strong customer retention.
  + Acquisition trends show steady customer growth with occasional spikes, especially around certain time periods (e.g., holiday seasons).
* **Sales Performance Insights**
  + Sales heatmaps and trend lines in Power BI identified Q3 as the peak sales period, likely driven by seasonality.
  + EDA highlighted a positive correlation between order frequency and total sales per customer segment.
  + Sales performance dashboards revealed that top 3 sales representatives contribute to over 40% of company sales.
* **Employee and Regional Analysis**
  + EDA showed that employees handling fewer orders still maintain high revenue per order, suggesting a potential for mentoring or best-practice sharing.
  + Regional breakdowns in Power BI maps highlighted underperforming areas that could benefit from focused sales efforts or promotions.
* **Product and Inventory Trends**
  + EDA helped detect low-performing products with high stock levels, indicating excess inventory and potential deadstock.
  + Power BI inventory dashboards showed that some high-demand products frequently go out of stock, affecting sales opportunities.
  + Product category analysis revealed that "Beverages" and "Condiments" consistently rank highest in revenue generation.
* **Supplier and Shipping Insights**
  + Supplier performance metrics in Power BI showed that a few suppliers consistently deliver on time, while others have higher delays and variability.
  + EDA uncovered patterns of late shipments affecting high-value orders, suggesting the need for better supplier contracts or backups.
* **Strategic Recommendations**
  + Use targeted marketing to convert one-time customers into repeat buyers.
  + Optimize inventory using predictive analytics to reduce overstock and prevent stockouts.
  + Focus on underperforming regions and employees with training, incentives, or support to improve overall performance.
  + Strengthen relationships with reliable suppliers and reevaluate contracts with those causing frequent delays.

**Credits**

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

* **My Mentors & Support System** – for their constant encouragement and insightful feedback throughout the project journey.
* **Microsoft Power BI & Excel** – for providing robust tools that enabled effective data analysis and visualization.
* **SQL Community & Online Resources** – for guidance and support in mastering query building and best practices.
* **Dataset Source: AccioJob and Team** – for providing the dataset that made this analysis possible.

This project reflects a journey 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.