

A close-up of a colorful background

AI-generated content may be incorrect.  


**Capstone Project – Sales Analytics**

A close-up of a colorful background

AI-generated content may be incorrect. **Northwind Traders – Project Documentation  
  
Overview**

**Data Analytics   
 ~ Priyanshu Joshi**

**Problem Statement**

The objective of this Power BI report is to design an interactive and visually compelling dashboard that delivers key performance insights for **Northwind Traders**.  
The dashboard will consolidate data from multiple sources to provide a **360° view** of the company’s operations, covering **sales performance, customer segmentation, inventory management, and employee productivity**. By leveraging advanced analytics and intuitive visualizations, the report will uncover trends in customer behavior, sales patterns, and workforce efficiency.

This solution will enable stakeholders to:

* **Explore data dynamically** with interactive filters and drill-down capabilities.
* **Identify growth opportunities** through customer and product insights.
* **Monitor performance metrics** across sales, inventory, and operations.
* **Make informed, data-driven decisions** that enhance competitiveness in the wholesale market.

The ultimate impact is to transform how Northwind Traders interacts with its business data, empowering leadership to optimize operations, strengthen customer relationships, and drive sustainable growth.

**Dataset Description**

The dataset is based on the **Northwind database**, which represents a fictional company, *Northwind Traders*, a global importer and exporter of specialty foods. It contains transactional, customer, employee, and product-level data that reflects real-world business scenarios.

The data model comprises multiple interconnected tables:

**1. Customers Table**

Holds details of all customers, including:

* Customer ID (unique identifier)
* Company and contact information
* Address, city, region, postal code, country
* Phone and fax details

**2. Employees Table**

Captures information about employees, including:

* Employee ID (unique identifier)
* Personal details (name, title, birthdate, hire date)
* Contact and address information
* Reporting structure (manager relationships)
* Notes and photo details

**3. Orders Table**

Contains order-level transaction data, including:

* Order ID, Customer ID, Employee ID
* Order, required, and shipped dates
* Freight charges
* Shipping information (ship name, address, city, region, postal code, country, shipper)

**4. Order Details Table**

Provides line-item details of each order:

* Order ID, Product ID
* Unit price, quantity, and discount applied

**5. Products Table**

Stores product-level information, including:

* Product ID, name, supplier, and category
* Quantity per unit
* Pricing and stock details (unit price, stock on hand, units on order, reorder level)
* Product status (discontinued or active)

**6. Suppliers Table**

Contains supplier details, including:

* Supplier ID and company information
* Contact person and title
* Address, city, region, postal code, country
* Communication details (phone, fax, homepage)

**7. Shippers Table**

Holds information about logistics providers:

* Shipper ID
* Company name and phone number

**8. Categories Table**

Defines product categories:

* Category ID, name, and description

A close-up of a colorful background

AI-generated content may be incorrect. **Process**  
  
**How to Access & Connect Northwind Data**

**Step 1: Accessing the Dataset Repository**

1. Open the GitHub link provided for the dataset:  
   👉 [Northwind Dataset Repository](https://github.com/acciojob-data-analytics/NorthWind)
2. In the repository, locate the **dataset files**. These include:
   * .csv files (for Customers, Orders, Products, etc.)
   * .sql file
3. Download the dataset by:
   * Clicking on the **green "Code" button** → "Download ZIP".
   * Or clone it via Git if you are using GitHub Desktop / Git Bash.

**Step 2: Prepare the Dataset for Use**

* CSV files are provided:
  1. Extract the ZIP file to a folder on your computer.
  2. Keep note of the file path.
* SQL script is provided:
  1. Open **SQL Server Management Studio (SSMS)** or **MySQL Workbench**.
  2. Create a new database called Northwind.
  3. Run the SQL script to import tables into the database.

**Step 3: Connect Data in Power BI**

**A: Using CSV files**

1. Open **Power BI Desktop**.
2. Click **Home → Get Data → Text/CSV**.
3. Browse to your extracted folder and load each file (Customers, Orders, Employees, etc.).
4. Repeat for all required files.
5. Once loaded, go to **Model View** and establish relationships:
   * Orders[CustomerID] → Customers[CustomerID]
   * Orders[EmployeeID] → Employees[EmployeeID]
   * Order Details[OrderID] → Orders[OrderID]
   * Order Details[ProductID] → Products[ProductID]
   * Products[SupplierID] → Suppliers[SupplierID]
   * Products[CategoryID] → Categories[CategoryID]
   * Orders[ShipVia] → Shippers[ShipperID]

**B: Using SQL Database**

1. Open **Power BI Desktop**.
2. Click **Home → Get Data → SQL Server**.
3. Enter your server’s name and database name (Northwind).
4. Choose **Import mode** (faster for reports) or **DirectQuery** (real-time connection).
5. Select tables: Customers, Employees, Orders, Order Details, Products, Suppliers, Shippers, Categories.
6. Load and review relationships in **Model View**.

**Step 4: Data Cleaning & Transformation (Power Query)**

Before building dashboards, clean the data:

* Remove unnecessary columns (e.g., fax, notes, photo paths).
* Format dates (e.g., Order Date, Shipped Date).
* Create calculated columns if needed (e.g., Total Sales = Quantity \* UnitPrice \* (1-Discount)).
* Standardized field names

**Step 5: Verifying Data Model**

* Ensure **star schema** design:
  + **Fact Table** → Orders & Order Details
  + **Dimension Tables** → Customers, Employees, Products, Suppliers, Shippers, Categories.
* Check relationships are **1-to-Many** and active.

**Step 6: Building Dashboards**

1. **Sales & Revenue**: Use Orders + Order Details.
2. **Customer Insights**: Use Customers + Orders.
3. **Employee Performance**: Use Employees + Orders.
4. **Product & Inventory**: Use Products + Order Details.

A close-up of a colorful background

AI-generated content may be incorrect.

**Objective**

The objective of this project is to design a comprehensive and visually compelling Power BI dashboard using the Northwind database, which represents a global wholesale food trading company. The report will consolidate sales, customers, products, suppliers, employees, and shipping data to uncover actionable insights.

The analysis aims to provide a **360-degree view of company operations** by focusing on:

* **Sales & Revenue Analysis** → Explore order trends, profitability, and key markets.
* **Customer Insights** → Segment customers based on purchase behavior and geography.
* **Product & Inventory Analysis** → Track top-selling items, product categories, and stock levels.
* **Employee Performance** → Measure sales contribution and efficiency of staff.
* **Supplier Overview** → Evaluate supplier impact on pricing, stock, and distribution.

The goal is to empower **data-driven decision-making**, improve profitability, optimize operations, and strengthen customer and supplier relationships.

A close-up of a colorful background

AI-generated content may be incorrect. **Significance**

Northwind dataset analysis plays a **critical role** for businesses in wholesale and retail trade by offering insights into multiple areas:

1. **For Business Leaders** → Provides a real-time overview of sales, revenue streams, and market trends to support strategic planning and expansion.
2. **For Sales & Marketing Teams** → Identifies top-performing customers, profitable product categories, and seasonal demand trends, enabling targeted campaigns.
3. **For Inventory & Supply Chain Managers** → Monitors stock levels, reorder needs, supplier performance, and delivery timelines, reducing stockouts and inefficiencies.
4. **For HR & Operations Teams** → Evaluates employee contributions, performance KPIs, and workload distribution, leading to effective workforce management.
5. **For Policy & Decision Makers** → Enables benchmarking and forecasting, ensuring the company maintains competitiveness in the global market.

By leveraging this dashboard, Northwind Traders can transform raw data into **actionable insights**, leading to **higher efficiency, improved profitability, and competitive advantage**.

A close-up of a colorful background

AI-generated content may be incorrect. **Northwind Data Dictionary**

**Table: Customers**

* **CustomerID** – Unique identifier for each customer.
* **CompanyName** – Name of the customer’s company.
* **ContactName** – Name of the contact person.
* **ContactTitle** – Designation of the contact person.
* **Address** – Street address of the customer.
* **City** – City of the customer.
* **Region** – Region/state of the customer.
* **PostalCode** – Postal code.
* **Country** – Country of the customer.
* **Phone** – Contact phone number.
* **Fax** – Fax number.
* **Image / ImageThumbnail** – Visual identifier/logo of the customer.

**Table: Orders**

* **OrderID** – Unique identifier for each order.
* **CustomerID** – Foreign key linking to Customers.
* **EmployeeID** – Foreign key linking to Employees.
* **OrderDate** – Date the order was placed.
* **RequiredDate** – Date by which the order is required.
* **ShippedDate** – Date the order was shipped.
* **Freight** – Shipping cost of the order.
* **ShipVia** – Shipper used (links to Shippers).
* **ShipName** – Name under which the order is shipped.
* **ShipAddress / City / Region / PostalCode / Country** – Shipping details.
* **ShippingDuration** – Calculated duration between order and shipped date.

**Table: Order Details**

* **OrderID** – Foreign key linking to Orders.
* **ProductID** – Foreign key linking to Products.
* **UnitPrice** – Price of each product at the time of order.
* **Quantity** – Number of units ordered.
* **Discount** – Discount applied to the order line.
* **LineTotal** – Calculated total for the line item.
* **OrderValue (bins)** – Grouped order value categories.
* **Quantity (bins)** – Grouped quantity categories.
* **UnitPrice (bins)** – Grouped price categories.

**Table: Products**

* **ProductID** – Unique identifier for each product.
* **ProductName** – Name of the product.
* **SupplierID** – Foreign key linking to Suppliers.
* **CategoryID** – Foreign key linking to Categories.
* **QuantityPerUnit** – Packaging details (e.g., 12 units per box).
* **UnitPrice** – Unit price of the product.
* **UnitsInStock** – Available stock.
* **UnitsOnOrder** – Units currently on order.
* **ReorderLevel** – Minimum stock before reorder.
* **Discontinued** – Indicates if product is discontinued.

**Table: Suppliers**

* **SupplierID** – Unique identifier for each supplier.
* **CompanyName** – Name of the supplier’s company.
* **ContactName** – Contact person at the supplier.
* **ContactTitle** – Designation of the contact.
* **Address / City / Region / PostalCode / Country** – Supplier address details.
* **Phone / Fax** – Contact details.
* **HomePage** – Website of the supplier.
* **Average Products per Supplier** – Calculated KPI.
* **Top Supplier** – Flag/measure for best-performing supplier.
* **Total Suppliers** – Aggregated KPI.

**Table: Shippers**

* **ShipperID** – Unique identifier for each shipper.
* **CompanyName** – Name of the shipping company.
* **Phone** – Contact phone number.

**Table: Categories**

* **CategoryID** – Unique identifier for each category.
* **CategoryName** – Name of the product category.
* **Description** – Details about the category.
* **Picture** – Category image/logo.

**Table: Employees**

* **EmployeeID** – Unique identifier for each employee.
* **FirstName / LastName** – Employee’s name.
* **Title** – Job title.
* **BirthDate** – Date of birth.
* **HireDate** – Employment start date.
* **Address / City / Region / PostalCode / Country** – Employee’s address.
* **HomePhone / Extension** – Contact details.
* **Photo / Notes** – Employee profile details.
* **ReportsTo** – Manager’s EmployeeID (hierarchy).
* **TenureYears** – Calculated length of service.
* **TenureYears (bins)** – Grouped tenure categories.

**Table: OrderValueSummary (Calculated / Aggregated Table)**

* **OrderID** – Reference order identifier.
* **OrderValue** – Total value of the order.
* **OrderValue (bins)** – Grouped order value categories.
* **Total Sales** – Aggregated KPI.
* **Sales Volume** – Total number of products sold.
* **TotalCustomers** – Count of unique customers.
* **NumberofCountries** – Count of distinct countries served.
* **Top Product** – Best-performing product (calculated).
* **Top Country** – Best-performing country (calculated).

A close-up of a colorful background

AI-generated content may be incorrect.

**ER diagram**  
  
An **ER Diagram (Entity–Relationship Diagram)** is a type of **data modeling diagram** that visually represents the structure of a database.

It shows:

1. **Entities** – the main objects or concepts in the system (e.g., Customer, Order, Product).
2. **Attributes** – the details or properties of entities (e.g., CustomerName, OrderDate, Price).
3. **Relationships** – how entities are connected to each other (e.g., Customers *place* Orders, Orders *contain* Products).

**Entities and Attributes**

1. **Customers**
   * Attributes: CustomerID, CompanyName, ContactName, Address, City, Country, etc.
   * Represents people or organizations placing orders.
2. **Employees**
   * Attributes: EmployeeID, LastName, FirstName, Title, Address, City, ReportsTo, etc.
   * Represents staff members handling customer orders.
3. **Orders**
   * Attributes: OrderID, CustomerID, EmployeeID, OrderDate, RequiredDate, ShippedDate, ShipVia, Freight, etc.
   * Represents purchase transactions made by customers.
4. **Order\_Details**
   * Attributes: OrderID, ProductID, UnitPrice, Quantity, Discount.
   * Acts as a **bridge (junction table)** between Orders and Products (many-to-many relationship).
5. **Products**
   * Attributes: ProductID, ProductName, SupplierID, CategoryID, UnitPrice, UnitsInStock, etc.
   * Represents items available for sale.
6. **Suppliers**
   * Attributes: SupplierID, CompanyName, ContactName, Address, City, Phone, etc.
   * Represents companies that provide products.
7. **Categories**
   * Attributes: CategoryID, CategoryName, Description, Picture.
   * Groups products into categories (e.g., Beverages, Dairy).
8. **Shippers**
   * Attributes: ShipperID, CompanyName, Phone.
   * Represents delivery companies responsible for shipping orders.
9. **OrderValueSummary (Derived/Calculated Table)**
   * Attributes: OrderID, OrderValue, DistinctCountries, Sales Volume.
   * Likely a summary view/table for analytical purposes.

**Relationships**

* **Customers ↔ Orders**  
  One customer can place many orders (1-to-many).
* **Employees ↔ Orders**  
  One employee can process many orders (1-to-many).
* **Orders ↔ Order\_Details**  
  Each order can include many products, and each product can appear in many orders → many-to-many resolved via **Order\_Details**.
* **Products ↔ Suppliers**  
  Each product is supplied by one supplier, but one supplier can provide many products (1-to-many).
* **Products ↔ Categories**  
  Each product belongs to one category, and each category can include many products (1-to-many).
* **Orders ↔ Shippers**  
  Each order is shipped via one shipper, but one shipper can handle many orders (1-to-many).
* **Order\_Details ↔ Products**  
  Many-to-one relationship (each detail line corresponds to one product).
* **OrderValueSummary ↔ Orders**  
  Analytical relationship summarizing order-level data.

**Key Points**

* The schema resembles the classic **Northwind database model** used for learning relational databases.
* **Normalization:**
  + Products, Suppliers, and Categories are normalized into separate entities to reduce redundancy.
  + Order\_Details acts as a fact table linking Orders and Products.
* **Business Use:**
  + Tracks customers and their orders.
  + Manages employees and their sales activities.
  + Maintains supplier-product relationships.
  + Supports shipment tracking through shippers.
  + Provides summary analytics (OrderValueSummary).

A screenshot of a computer screen

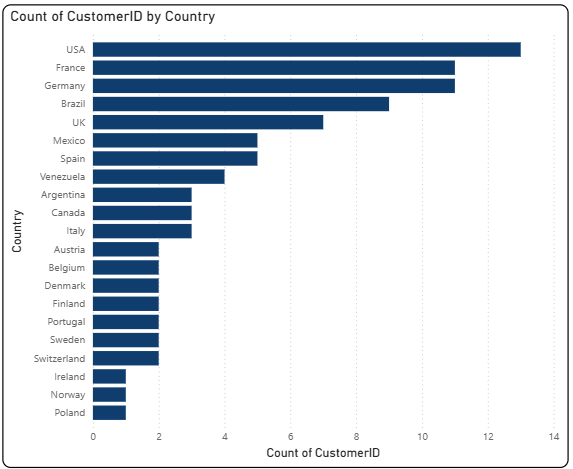
AI-generated content may be incorrect.

A close-up of a colorful background

AI-generated content may be incorrect.  
 **Power BI Problem Statements**

**1. Customer Analysis**

**Q1**. How does customer distribution vary across different countries or cities?

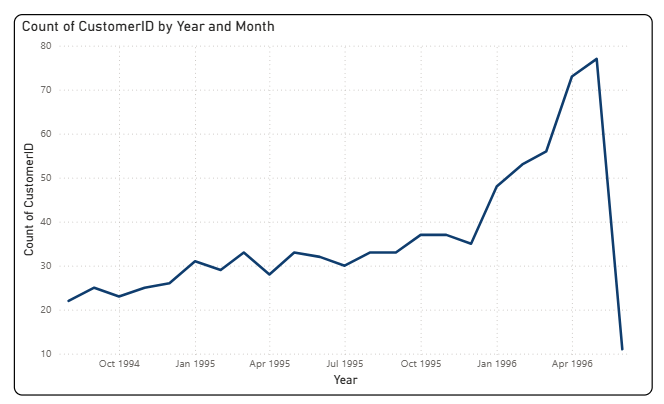


The **United States** emerged as the country with the **highest number of customers**, recording a total of **13 unique CustomerIDs**. This was followed closely by **Germany** and **France**, each contributing **11 CustomerIDs** to the overall dataset. In terms of percentage share, the United States alone accounted for **14.29% of the total customer base**, making it the single largest contributor among all countries.

When the customer distribution is analyzed across all **21 countries**, the number of CustomerIDs varies significantly, ranging from as few as **1 customer** in some regions to as many as **13 customers** in the leading market. This wide range highlights the diversity of the customer base, with certain countries representing strong, concentrated markets while others account for smaller but important portions of the overall distribution.

Such an analysis provides valuable insights into the **geographical spread of customers**, helping businesses identify their strongest markets, recognize potential areas for growth, and tailor region-specific marketing or sales strategies.

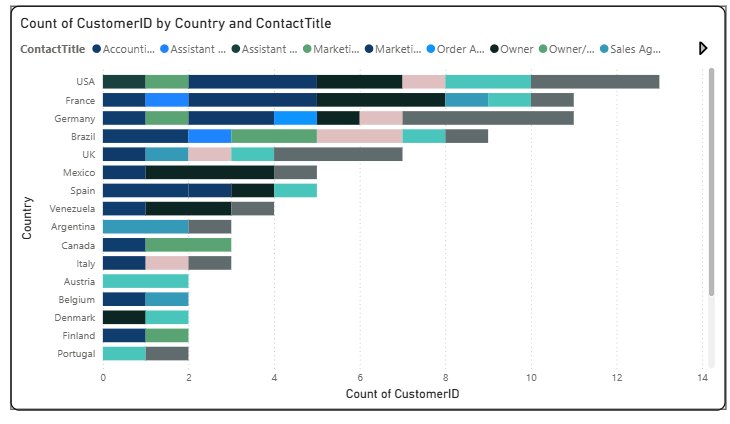
**Q2.** What is the trend in customer orders over time?



The trend analysis of CustomerIDs reveals a significant fluctuation over the observed period. Between **August 1994 and June 1996**, the number of CustomerIDs experienced a sharp **decline of 50%**, signaling a period of reduced customer engagement or loss in market activity. However, this downward trend reversed in **October 1995**, when customer numbers began to rise again at an impressive pace.

Within just **7 months**, the count of CustomerIDs increased by **108.11%**, growing from **37 to 77**. This period, spanning from **October 1995 to May 1996**, marked the steepest incline in customer acquisition, reflecting a strong recovery and potentially the impact of improved business strategies, marketing efforts, or expansion into new markets.

Overall, the data highlights a volatile yet promising trajectory: an initial contraction followed by a rapid rebound, with the latter period showcasing significant growth momentum.

**Q3**. What is the distribution of customers by Contact Title or Region? Use stacked bar chart   
or pie chart to visualize.  
  


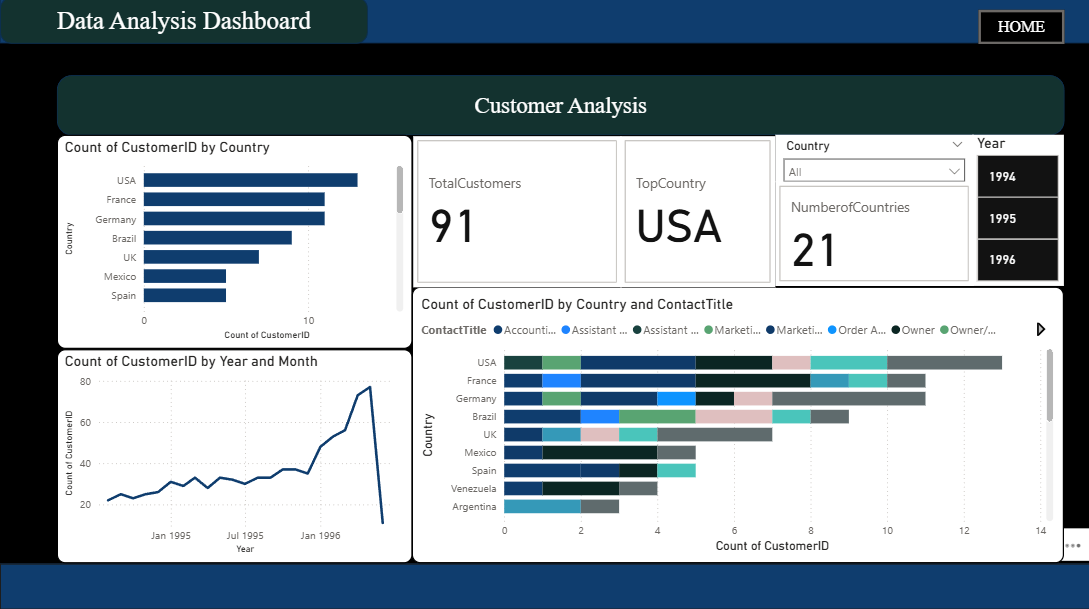
The analysis of customer distribution by **Contact Title** reveals some interesting patterns. Both the **Sales Representative** and **Owner** roles emerged as the leading categories, each contributing the **highest total CustomerID count of 17**. These roles are closely followed by the **Marketing Manager**, which also plays a significant part in the customer base. This suggests that business owners and sales professionals form a core segment of the company’s customer network, while marketing leaders also represent a considerable share.

Looking at the data from a regional perspective, in **Germany**, the **Sales Representative** role alone accounted for **4.40% of the total CustomerID count**, emphasizing the strong presence of sales-driven contacts in that market. This regional insight is particularly valuable as it highlights how specific roles contribute differently depending on the country.

When examining averages, the **Marketing Manager** recorded the **highest average CustomerID count at 1.71**, outperforming both the Sales Representative and Owner categories. This indicates that, although Marketing Managers may not dominate in absolute numbers, they tend to have a consistently higher representation on average compared to other roles.

Overall, the distribution suggests a **balanced mix of decision-makers, sales professionals, and marketing leaders** among customers. This insight can guide targeted engagement strategies:

* **Owners** may respond better to strategic partnership and long-term value propositions.
* **Sales Representatives** may prefer solutions that support efficiency and sales growth.
* **Marketing Managers** could be more inclined toward campaigns, promotions, and brand-driven initiatives.

Such segmentation is essential for tailoring communication, product offerings, and relationship management in order to strengthen customer engagement across diverse roles and regions.  
  
  
  
  


**1. High-Level KPIs (Top Center)**

* **Total Customers: 91**  
  → There are **91 customers** in total.
* **Top Country: USA**  
  → The **USA has the highest number of customers** compared to other countries.
* **Number of Countries: 21**  
  → Customers are spread across **21 different countries**, showing wide global reach.
* **Filters (Right Side):** Country and Year (1994–1996) for segmented analysis.

**2. Customer Distribution by Country (Top Left – Bar Chart)**

* **USA, France, and Germany** are the **top 3 countries** in terms of customer count.
* Other notable countries include **Brazil, UK, Mexico, and Spain**.
* This helps identify key markets and regions where the business is most active.

**3. Customer Growth Over Time (Bottom Left – Line Chart)**

* Customer count shows a **steady upward trend from 1994 to 1996**.
* There is **consistent growth**, indicating expanding customer acquisition.
* A peak is observed in **mid-to-late 1996**, suggesting a **significant boost in new customers** (possibly due to marketing campaigns, product expansion, or new market entry).

**4. Customer Breakdown by Country & Contact Title (Bottom Right – Stacked Bar Chart)**

* Shows how **customers are distributed across job titles** (e.g., Owner, Marketing Manager, Order Administrator, Assistant, etc.) for each country.
* For example:
  + **USA** has a mix of **Owners, Marketing Managers, and Assistants**.
  + **France and Germany** also have diverse roles, indicating strong B2B penetration.
  + Smaller countries like **Venezuela and Argentina** have fewer customers but still diverse titles.
* This insight helps understand the **type of decision-makers and professionals** the company engages with in different markets.

**5. Key Insights**

* The company has a **total of 91 customers across 21 countries**, with the **USA leading in customer count**.
* Customer base has shown **consistent growth over the years**, peaking in **1996**.
* **Job titles are diverse**, meaning the company caters to a **broad professional audience** (business owners, marketing managers, assistants, administrators, etc.).
* The dashboard reveals **geographic strengths (USA, France, Germany)** and helps identify opportunities to expand in **emerging markets (Latin America, Spain, UK, etc.)**.

**In summary:**  
The **Customer Analysis dashboard** highlights customer growth, geographic distribution, and professional segmentation. The USA, France, and Germany are the strongest markets, while customer numbers have steadily increased from 1994 to 1996, reflecting expanding global reach and successful acquisition strategies.

**2. Order Trends**

**Q4**:How does order volume change over time? Use line chart or stacked bar chart to visualize.

A graph showing the growth of the year

AI-generated content may be incorrect.

The analysis of sales performance, measured through the **Sum of Quantity**, shows a period of significant fluctuation. Between **August 1994 and June 1996**, the total quantity of products ordered experienced a sharp **decline of 55.95%**, reflecting a substantial drop in sales during this time. This downward trend may indicate reduced demand, market challenges, or possible supply-side constraints that affected order volumes.

However, beginning in **October 1995**, the trend reversed, and the Sum of Quantity entered a strong growth phase. Over the next **seven months**, the metric increased by an impressive **111.57%**, rising from **2,614 units to 5,497 units**. This sharp recovery highlights a period of renewed market continuation, with customer demand rebounding strongly.

The most pronounced surge occurred between **October 1995 and May 1996**, when the Sum of Quantity nearly doubled in just a few months, climbing from **2,343 to 4,957 units**. This steep growth indicates not just a recovery but a strong expansion phase, suggesting that business conditions improved significantly, possibly due to strategic initiatives, seasonal demand, or the introduction of new products that stimulated customer purchases.

In summary, the data reflects a **two-phase pattern**: an initial contraction in sales volumes followed by a rapid and sustained growth period. While the early decline signals potential challenges that required corrective measures, the subsequent rebound demonstrates the business’s resilience and ability to capture renewed demand in a relatively short timeframe. Such insights are crucial for understanding market dynamics, forecasting demand, and identifying factors that drive sharp recoveries in sales performance.

**Q5**: What is the distribution of order values? Use histogram or box plot to visualize.

A graph of a number of numbers

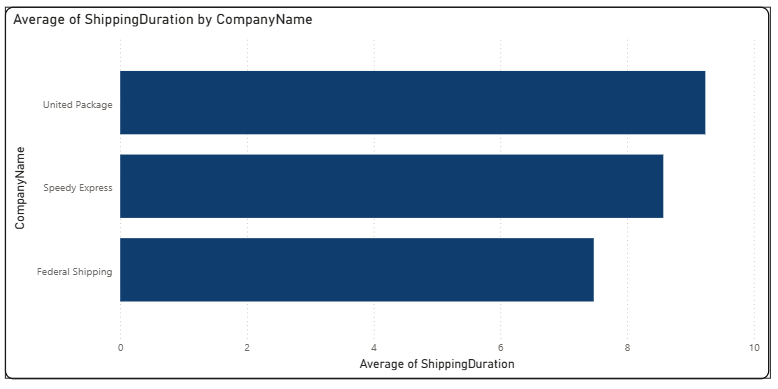
AI-generated content may be incorrect.

The distribution of orders by **OrderValue** highlights some notable trends. The **Count of OrderIDs** reached its highest peak at **234 orders** when the **OrderValue was 0**, making it the single most frequent category in the data set. This was followed by the next highest concentrations at **OrderValue levels of 500 and 1000**, which also recorded strong order counts, though not as dominant as the zero-value orders.

Interestingly, the **zero-value orders** accounted for a substantial portion of activity, representing **28.19% of the total Count of OrderIDs**. This suggests that a significant share of transactions did not generate revenue directly. Such orders could be attributed to **samples, promotional distributions, replacements, or non-billed transactions**, which may play a strategic role in customer acquisition and retention but warrant further analysis to understand their business impact.

Looking across the broader dataset, which includes **25 distinct OrderValue bins**, the **Count of OrderIDs** ranged widely—from as low as **1 order** in the least frequent categories to as high as **234 orders** in the most frequent bin. This wide variation reflects the diversity in purchasing behavior: while many orders cluster around smaller or non-revenue values, there are still orders distributed across mid-to-high value ranges, though with less frequency.

In summary, the analysis reveals a **skewed distribution**, heavily concentrated at the zero-value range, with secondary peaks at moderate values such as 500 and 1000. Understanding why zero-value orders dominate is crucial, as it may influence how the company assesses true revenue performance versus order activity. Additionally, identifying the drivers behind mid-value peaks could provide insights into customer purchasing patterns, promotional effectiveness, and potential opportunities to shift orders toward higher-value categories.

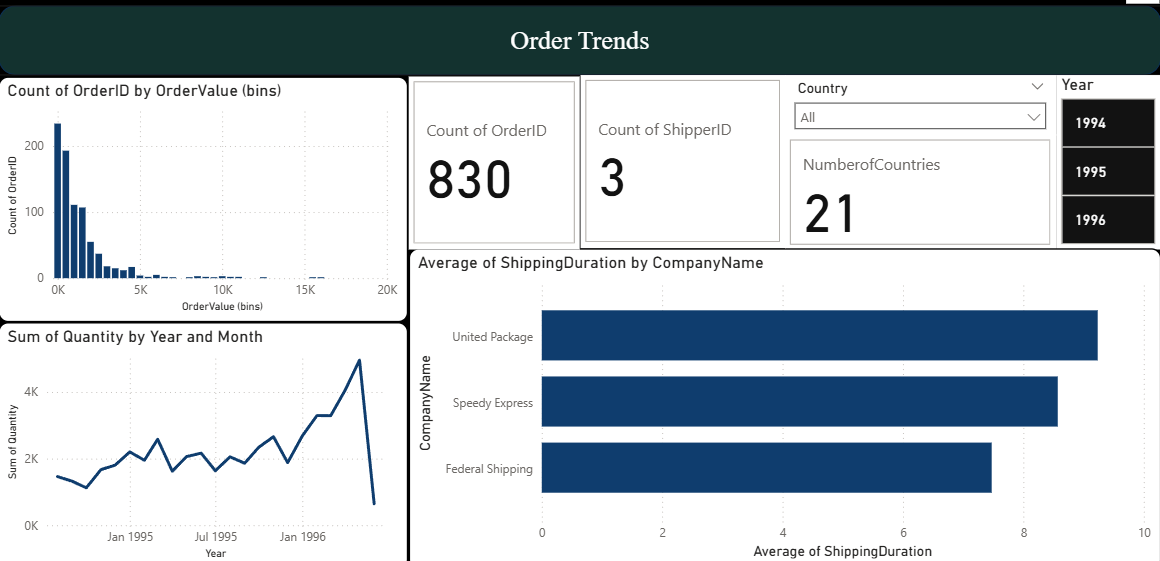
**Q6**: What is the average order shipping duration? Use bar chart or box plot to visualize  
  


The analysis of shipping performance by carrier reveals noticeable differences in delivery efficiency. **United Package** recorded the **highest average shipping duration at 9.23 days**, making it the slowest among the three shipping providers. This average is **23.56% longer than that of Federal Shipping**, which demonstrated the **fastest delivery performance with an average of just 7.47 days**.

**Speedy Express** ranked second, with an average shipping duration of **8.57 days**, positioning it between the fastest and slowest carriers. While it outperformed United Package, it still lagged behind Federal Shipping by more than a day on average.

These results suggest that customers relying on **United Package** may experience noticeably slower deliveries compared to those using Federal Shipping, which could influence customer satisfaction and repeat purchasing behavior. On the other hand, **Federal Shipping’s shorter average duration** highlights its relative efficiency and potential to be positioned as the preferred shipping partner for time-sensitive orders.

Overall, the comparison highlights a **performance gap of nearly two full days** between the fastest and slowest carriers, underscoring the importance of choosing the right shipping provider. Businesses can use this insight to **negotiate service-level improvements, optimize shipping assignments, or offer customers choices** based on delivery speed versus cost trade-offs.



**1. High-Level KPIs (Top Center)**

* **Count of OrderID: 830**  
  → A total of **830 orders** have been placed.
* **Count of ShipperID: 3**  
  → Orders are handled by **3 shipping companies**.
* **Number of Countries: 21**  
  → Orders span across **21 different countries**, showing broad international reach.
* **Filters:** Year (1994, 1995, 1996) and Country allow time-based or geography-based analysis.

**2. Order Value Distribution (Top Left – Histogram)**

* Most orders fall into **lower order value ranges (< $5K)**.
* Only a few orders reach higher order values (above $10K–20K).
* This indicates that the business is driven mainly by **smaller, frequent orders**, with large orders being rare.

**3. Shipping Duration Analysis (Right – Bar Chart)**

* **United Package** has the **longest average shipping duration (~9 days)**.
* **Speedy Express** is slightly faster (~8 days).
* **Federal Shipping** is the **fastest option (~7 days)**.
* This comparison helps identify efficiency gaps between shippers.

**4. Order Trends Over Time (Bottom Left – Line Chart)**

* The **quantity of items ordered** shows a **steady upward trend from 1994 to 1996**.
* There are occasional fluctuations, but overall growth suggests **rising demand and expanding sales**.
* The sharp peak in early 1996 highlights a **surge in order volume**, followed by a dip that may indicate seasonality.

**5. Key Insights**

* The company processed **830 orders across 21 countries**, showing a **global footprint**.
* The majority of orders are **small to medium in value**, but a few high-value orders exist.
* **Federal Shipping is the most efficient shipper**, while **United Package has the slowest delivery times**.
* Demand has been **growing steadily over the years**, with a notable spike in 1996, signaling market expansion.

**In summary:**  
The **Order Trends dashboard** reveals that the company handles a large number of relatively small-value orders across 21 countries, with order volumes growing steadily over time. Shipping performance varies by provider, with **Federal Shipping** being the most reliable and fastest. The steady growth in demand highlights expanding business opportunities and the need for efficient logistics management.

**3. Employee Data**

**Q7:** What is the count of employees by job title or region? Use stacked bar chart or tree map   
to visualize.

A graph with blue lines

AI-generated content may be incorrect.

The analysis of employee distribution by country shows that the **United States** employed the **highest number of staff, with 5 employees**, slightly ahead of the **United Kingdom, which accounted for 4 employees**. Together, these two countries form the core of the workforce, but the United States leads by a narrow margin.

In percentage terms, the **USA represents 55.56% of the total employee base**, making it the majority contributor to overall staffing. The UK, on the other hand, contributes **44.44%**, reflecting a nearly balanced yet slightly smaller share.

This distribution highlights that the organization’s workforce is concentrated across just two countries, with the **USA holding a stronger presence**. Such a concentration may indicate that the company relies heavily on its American operations while maintaining a significant but secondary hub in the UK.

Understanding this geographic split is valuable for assessing workforce allocation, regional capacity, and potential dependencies. For instance, a greater proportion of employees in the USA may suggest that **key operational, managerial, or customer-facing functions** are based there, while the UK presence may support regional markets or specialized roles.

**Q8:** What is the distribution of employee tenure? Use histogram or box plot to visualize.  
  
A graph of a number of blue rectangular objects

AI-generated content may be incorrect.

The analysis of employee tenure shows a clear pattern of balance across the longer service categories. Specifically, each of the tenure groups—31 years, 32 years, and 33 years—recorded an equal EmployeeID count of 3. This uniform distribution suggests that the organization has a consistent number of employees who have remained with the company for three decades or more.

Such stability at the higher end of tenure highlights the presence of a loyal and experienced workforce, with multiple employees sustaining long-term careers within the company. It also reflects positively on employee retention practices, indicating that the organization has been successful in maintaining staff commitment over extended periods.

From a strategic perspective, this trend may imply that the company benefits from a strong pool of institutional knowledge and expertise. However, it could also signal a potential challenge soon if several long-serving employees approach retirement around the same time, creating a need for succession planning and knowledge transfer programs.

In summary, the equal distribution of employees across the 31-, 32-, and 33-year tenure categories not only highlights workforce stability but also underlines the importance of planning ahead to balance long-term employee loyalty with sustainable talent renewal.

**Q9:**What is the reporting structure among employees? Use org chart or hierarchical tree to visualize.  
  
A screenshot of a graph

AI-generated content may be incorrect.

The graph illustrates the **hierarchical distribution of employees** filtered by the **Title = Sales Representative**, with additional breakdowns by **ReportsTo** and **Country**. The analysis begins with the **total number of employees by first name**, which amounts to **9 employees** across all roles included in the visual.

From this total, **6 employees hold the title of Sales Representative**, making it the dominant role within this filtered dataset. Other roles represented include **Inside Sales Coordinator (1 employee)**, **Sales Manager (1 employee)**, and **Vice President of Sales (1 employee)**. This shows that while the Sales Representative role accounts for the majority, there is still representation from other leadership and support positions.

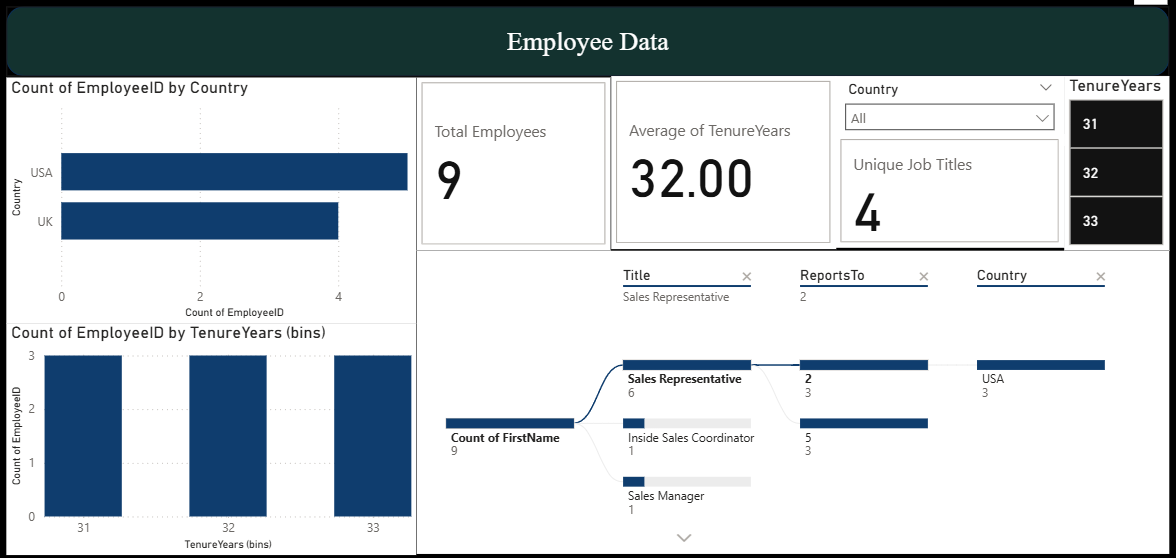
Focusing specifically on the **Sales Representatives**, the breakdown by **ReportsTo** provides insights into reporting structure. Out of the 6 Sales Representatives, **3 report directly to EmployeeID 2**, while another **3 reports to EmployeeID 5**. This balanced split indicates that responsibility for overseeing Sales Representatives is evenly distributed between these two managers.

When adding the **Country dimension**, the analysis shows that all **3 Sales Representatives reporting to EmployeeID 2 are based in the USA**. This highlights the geographical allocation of part of the sales team and points to the United States as a key hub for this role. The employees reporting to EmployeeID 5, while not expanded in this view, account for the other half of the Sales Representatives, suggesting a potential presence in other regions or markets.

Overall, the graph provides a **multi-level view of the workforce structure**, starting from the total employee count and drilling down into **role (Sales Representative), reporting hierarchy, and geographical distribution**. It highlights not only the concentration of employees in certain titles but also reveals insights into managerial oversight and regional workforce allocation.

From a business perspective, this visualization is useful for:

* Understanding **team size and distribution** within the Sales function.
* Analyzing **reporting lines** to identify managerial workload.
* Highlighting **regional presence** of key sales roles, with the USA standing out in this case.
* Supporting decisions related to **resource allocation, leadership balance, and organizational design**.



**1. High-Level KPIs (Top Center)**

* **Total Employees: 9**  
  → The company has **9 employees** in total.
* **Average of Tenure Years: 32.00**  
  → On average, employees have been with the company for **32 years**, showing very high retention and experience levels.
* **Unique Job Titles: 4**  
  → Employees hold **4 different job roles** across the organization.
* **Filters:** Country and Tenure Years (31, 32, 33) are available for drill-down analysis.

**2. Employee Distribution by Country (Top Left)**

* **USA** has the largest number of employees, followed by the **UK**.
* This indicates that the workforce is concentrated mainly in the USA.

**3. Employee Tenure Distribution (Bottom Left)**

* Employees are grouped into tenure ranges (**31, 32, 33 years**).
* The distribution is balanced across these bins, suggesting **long-term loyalty** among staff, with most employees serving over three decades.

**4. Job Roles and Reporting Structure (Right Side – Sankey/Flow Visual)**

* **Sales Representative** is the most common job role, with **6 employees**.
* Other roles include:
  + **Inside Sales Coordinator** → 1 employee
  + **Sales Manager** → 1 employee
* The reporting structure shows connections between employees and supervisors, mainly concentrated in the USA.

**5. Key Insights**

* The company has a **very small but highly experienced workforce** (average tenure of 32 years).
* The majority of employees are in **Sales-focused roles** (Sales Representatives dominate).
* The workforce is **geographically concentrated in the USA and UK**.
* High employee tenure suggests **stability, low turnover, and strong retention**, but may also indicate an **aging workforce** with fewer new hires.

**In summary:**  
This dashboard highlights that the organization is **small (9 employees), sales-oriented, and highly experienced**, with most employees serving over 30 years. The workforce is primarily based in the USA and UK, with Sales Representatives forming the bulk of the team.

**4. Product Insights**

**Q10**: Which products have the highest sales volume? Use bar chart or tree map to visualize.  
  
A graph with blue lines

AI-generated content may be incorrect.

The product-level sales analysis highlights **Camembert Pierrot** as the standout performer, achieving the **highest sales volume of 1,577 units**. This figure places it significantly ahead of the lowest-selling product, **Mishi Kobe Niku**, which recorded just **95 units**, resulting in a staggering **1,560% difference** between the top and bottom performers. Such a gap underscores the wide disparity in demand across the product portfolio.

Following Camembert Pierrot, the **second- and third-highest sales volumes** were recorded for **Raclette Courdavault** and **Gorgonzola Telino**, respectively. These products, while not matching the leader, still demonstrate strong market performance and reflect customer preference for cheese-based offerings, particularly within this product category.

In terms of overall contribution, **Camembert Pierrot accounted for 3.07% of total sales volume across all products**, establishing its importance as a notable driver of sales. Although the percentage may appear modest, in the context of a portfolio containing **77 different products**, it represents a meaningful share of the overall sales mix.

Across the full product range, sales volumes varied significantly, spanning from a minimum of **95 units to a maximum of 1,577 units**. This broad spread illustrates the diverse levels of demand for different products—some being niche items with limited reach, while others, such as Camembert Pierrot, emerge as clear leaders in customer preference.

From a business perspective, these insights are highly valuable. The dominance of products like Camembert Pierrot suggests opportunities to **reinforce marketing and distribution strategies** around top-performing items. Meanwhile, products with consistently lower sales volumes, such as Mishi Kobe Niku, may warrant closer evaluation to determine whether targeted promotions, repositioning, or even product line rationalization could improve their performance.

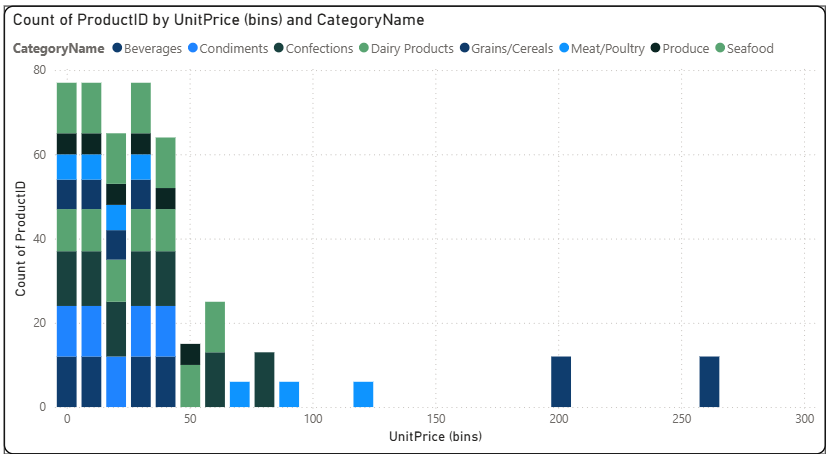
**Q11**:How does the sales volume vary across different product categories? Use stacked bar   
chart or tree map to visualize.  
  
A chart with different colors

AI-generated content may be incorrect.

Within the **Dairy Products** category, **Camembert Pierrot** emerged as a notable contributor, accounting for **3.07% of the total sales volume**. While this percentage may appear modest at first glance, it highlights the product’s steady role in driving sales within a highly competitive category that includes a wide variety of cheeses and other dairy offerings.

The contribution of Camembert Pierrot is particularly significant when considering the breadth of products in the Dairy segment. Its share reflects not only consistent customer demand but also the strong market positioning of the product compared to its peers. The performance indicates that Camembert Pierrot has established a loyal consumer base and continues to play a vital role in sustaining overall category growth.

This level of contribution also suggests potential opportunities for **strategic promotion and expansion**. With focused marketing efforts or enhanced distribution, Camembert Pierrot’s share within the Dairy Products segment could increase further, strengthening both category performance and the company’s overall sales portfolio.

**Q12:**Can we visualize the pricing distribution of products? Use box plot or histogram to visualize.  
  


The **Confections** category recorded the **highest total ProductID count**, reaching **91**, which reflects the wide variety and strong representation of confectionery products in the overall portfolio. This dominance highlights the importance of Confections as a leading category, suggesting both high consumer demand and diverse product offerings.

On the other end of the spectrum, **Grains/Cereals** registered the **lowest total ProductID count at 28**, indicating a relatively smaller product base in this category. This gap between Confections and Grains/Cereals underscores the uneven distribution of product variety across categories, likely driven by differences in consumer preferences, demand cycles, and market focus.

When looking at the **average ProductID count**, the trend remains consistent. **Confections once again leads**, with an average of **13**, showcasing both volume and variety as its key strengths. In contrast, **Produce ranked lowest**, with an average ProductID count of just **5**, reflecting limited product representation compared to other categories.

Overall, these insights reveal how certain categories, such as Confections, are central to the business due to their strong presence, while others, like Grains/Cereals and Produce, may represent niche segments with more targeted but smaller-scale offerings. This distribution suggests opportunities for **further product development and diversification** in underrepresented categories to achieve a more balanced portfolio.

A screenshot of a computer

AI-generated content may be incorrect.

**1. High-Level KPIs (Top Center)**

* **Total Sales: 1.35M**  
  → The combined sales value across all products is 1.35 million.
* **Top Product: Côte de Blaye**  
  → Côte de Blaye generates the highest sales value among all products, making it the standout product.
* **Count of ProductID: 77**  
  → The dataset includes 77 distinct products.
* **Filters (Country & Year: 1994–1996)**  
  → Users can drill down by country and year to analyze performance trends.

**2. Sales by Individual Product (Top Left)**

* **Bar chart of Sales Volume by ProductName**
  + Top-selling products include **Camembert Pierrot, Raclette Courdavault, and Gorgonzola Telino**, each showing significant sales volumes.
  + Côte de Blaye also contributes strongly, reaffirming its place as the top product overall.
  + Other notable products include Pavlova and Guaraná Fantástica.
* This highlights that **a small group of products contributes disproportionately to sales**.

**3. Product Pricing Distribution (Bottom Left)**

* **Count of ProductID by UnitPrice (bins) and CategoryName**
  + Most products fall into the **low-to-moderate pricing ranges (under 50)**.
  + A few products have exceptionally high unit prices (approaching 300), such as Côte de Blaye, which skews the pricing distribution.
  + Categories like **Beverages and Confections** dominate in product count across different price bands.

**4. Sales by Category and Product (Bottom Right)**

* **Stacked bar chart of Sales Volume by CategoryName and ProductName**
  + **Beverages** and **Dairy Products** emerge as the leading sales categories.
  + Other strong contributors: **Confections, Seafood, and Condiments**.
  + Categories like **Grains/Cereals and Produce** contribute less overall sales volume.
* This view helps identify **which categories are driving sales** and which are underperforming.

**5. Overall Insights**

* The product portfolio is **diverse (77 products across multiple categories)**, but sales are **concentrated in a few high-value items**.
* **Côte de Blaye** is the **star performer**, driving significant sales due to its premium price point.
* **Beverages and Dairy Products** are the backbone of category sales.
* Pricing analysis shows that **most products are affordable**, but luxury products (like Côte de Blaye) heavily influence revenue despite being fewer in number.

**In summary:**  
This dashboard provides a **comprehensive view of product performance**, showing both **breadth (77 products across multiple categories)** and **depth (high-value top products like Côte de Blaye)**. It highlights the importance of a few premium items in driving total sales while also showing the steady contribution of everyday products in categories like Beverages and Dairy.

**5. Supplier Overview**

**Q13:** How many products are supplied by each supplier?   
  
A colorful pie chart with text

AI-generated content may be incorrect.

**Pavlova, Ltd.** and **Plutzer Lebensmittelgroßmärkte AG** emerged as the top-performing suppliers in terms of **ProductID count**, with both companies contributing **5 products each** to the portfolio. This tie for the leading position highlights their significant role in expanding product diversity and ensuring a broad range of offerings for customers. Their strong representation suggests not only a consistent supply capacity but also a strategic focus on maintaining a competitive edge in the market.

Following closely behind, **New Orleans Cajun Delights** secured the **second position**, further strengthening the overall variety available to customers. Although slightly behind the leaders, its contribution demonstrates the supplier’s relevance and importance in supporting product diversity within the network.

Supplier that two suppliers tied at the top emphasizes the competitive nature of supplier participation, while the close positioning of New Orleans Cajun Delights underscores how supplier contributions collectively shape the balance of the product portfolio. This distribution suggests that maintaining strong relationships with these key suppliers is vital to sustaining inventory richness, improving customer choice, and driving overall business growth.

**Q14:** How does product pricing vary across different suppliers?   
  
A graph of different colored squares

AI-generated content may be incorrect.

Within the Aux joyeux ecclésiastiques category, Côte de Blaye stood out as a significant contributor, accounting for 11.85% of the total UnitPrice. This makes it one of the most valuable products in the category, reflecting its premium pricing and strong market positioning. Such a high percentage indicates that Côte de Blaye carries considerable weight in driving the overall value of the category, even when compared to other products that may have higher sales volumes but lower individual pricing.

The prominence of Côte de Blaye suggests its role not only as a revenue generator but also as a potential flagship product within the Aux joyeux ecclésiastiques range. Its contribution highlights the importance of premium-priced products in maintaining profitability and balancing categories that might otherwise rely more heavily on volume-driven items.

This insight underscores the need for strategic focus on managing and promoting Côte de Blaye, as it has the capacity to significantly influence category performance. Furthermore, the product’s strong representation in Unit Price distribution suggests that maintaining supply consistency and reinforcing its market appeal could yield long-term advantages for both category growth and overall portfolio profitability.

**Q15:** What is the geographical distribution of suppliers?   
  
A map of the world with blue circles

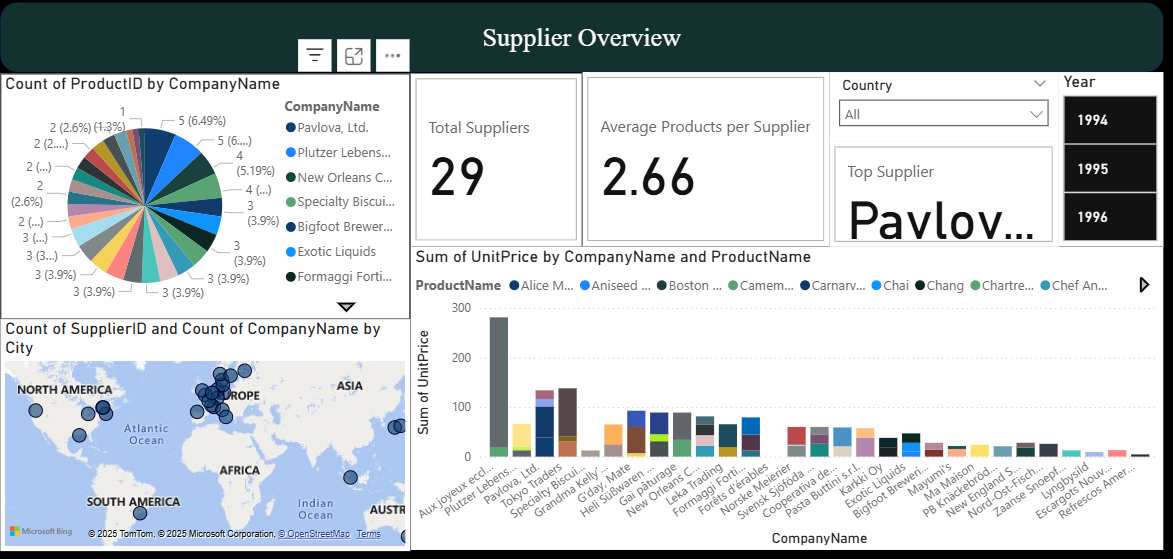
AI-generated content may be incorrect.  
  
This graph is a **geographical map visualization** that shows the **distribution of suppliers by city**, based on the **Count of SupplierID and CompanyName**.

From the visualization, it is clear that:

* **Europe** has the **highest concentration of suppliers**, with dense clustering across multiple cities such as those in Germany, France, and the UK. The close grouping of circles in this region highlights Europe as a major hub for supplier activity.
* **North America**, particularly the United States, also shows a notable presence of suppliers, though fewer compared to Europe.
* **South America** has limited representation, with only one visible city contributing to the supplier network.
* In the **Asia-Pacific region**, there is supplier activity in both **Australia** and parts of **East Asia**, indicating global diversification of supply sources.
* **Africa** and much of **Asia** show little to no supplier presence, suggesting that the network is less dependent on these regions.

The **size of the circles** represents the count of suppliers (SupplierID/CompanyName) in each city. Larger circles indicate cities with a higher number of suppliers, making them critical supply hubs.

Overall, the graph highlights a **supplier base heavily concentrated in Europe**, moderately represented in North America and Asia-Pacific, and sparsely distributed in South America and Africa. This suggests a strong reliance on European suppliers, while other regions contribute to global reach but at a smaller scale.



**1. High-Level KPIs (Top Center)**

* **Total Suppliers: 29**  
  → There are 29 suppliers included in the dataset.
* **Average Products per Supplier: 2.66**  
  → On average, each supplier provides about 3 products.
* **Top Supplier: Pavlova, Ltd.**  
  → Pavlova, Ltd. is identified as the leading supplier based on product count or value contribution.
* **Filters: Country & Year (1994–1996)**  
  → The dashboard can be filtered by country and year for deeper insights.

**2. Supplier Contribution (Pie Chart – Top Left)**

* This shows the **distribution of ProductID counts by CompanyName**.
* Pavlova, Ltd. and Plutzer Lebensmittelgroßmärkte AG hold the highest share, with **5 products each (6.49%)**.
* Other suppliers like New Orleans Cajun Delights, Specialty Biscuits, Ltd., and Bigfoot Breweries follow closely with smaller shares.
* The pie chart highlights that **product contributions are spread across suppliers**, with no extreme dominance by a single company.

**3. Global Supplier Distribution (Map – Bottom Left)**

* A geographical visualization of suppliers by **SupplierID and CompanyName by City**.
* The **largest cluster is in Europe**, especially in Germany, the UK, and France, showing this region as a supplier hub.
* Suppliers are also present in **North America, South America, Australia, and Asia**, indicating a global but uneven distribution.
* Africa has almost no visible representation.

**4. Pricing Insights (Bar Chart – Bottom Right)**

* This chart shows the **Sum of UnitPrice by CompanyName and ProductName**.
* **Aux joyeux ecclésiastiques (Côte de Blaye)** stands out with the highest unit price contribution, followed by suppliers offering premium or specialty items.
* Other companies show more moderate or balanced contributions, suggesting varied pricing strategies across suppliers.

**5. Overall Insights**

* The supplier base is **diverse, with 29 suppliers worldwide**, but **Europe is the central hub**.
* **Top-performing suppliers like Pavlova, Ltd. and Plutzer Lebensmittelgroßmärkte AG** provide the widest product ranges.
* High-value items (like Côte de Blaye) significantly influence the **UnitPrice distribution**, even if product count is lower.
* The dashboard provides both **volume (product count)** and **value (pricing impact)** perspectives for a well-rounded supplier analysis.

**In summary:**  
This dashboard gives a clear picture of supplier diversity, geographic distribution, and product value contribution. It highlights which suppliers dominate in product volume, which contribute premium pricing, and how the global supplier network is structured across regions.  
  
  
  
Top Countries: USA led with 13 customers (14.29% of total), followed by Germany and France (11 each). Customer counts ranged from 1–13 across 21 countries. By Role: Sales Representative and Owner tied for highest customer count (17), with Marketing Manager having the highest average count (1.71). In Germany, Sales Representatives contributed 4.40% of total customers. Sales Trends: Quantity fell 55.95% (Aug 1994–Jun 1996) before rising 111.57% in 7 months, peaking between Oct 1995–May 1996 (+111.59%). Orders: Most common order value was $0 (234 orders, 28.19% of total), followed by $500 and $1,000. Shipping: United Package had the slowest average delivery (9.23 days), 23.56% longer than Federal Shipping’s fastest (7.47 days). Employees: USA employed the most staff (5), followed by the UK (4), representing 55.56% of total. Top Products: Camembert Pierrot had the highest sales volume (1,577), 1,560% higher than the lowest (Mishi Kobe Niku, 95), and contributed 3.07% of total volume. Categories: Confections led in product count (91) and average count (13), while Grains/Cereals were lowest (28, avg. 5). Suppliers: Pavlova, Ltd. and Plutzer Lebensmittelgroßmärkte AG tied for most products (5 each). Côte de Blaye held 11.85% of total unit price in Aux joyeux ecclésiastiques.

**Conclusion**

In conclusion, the dashboards provide clear insights into employees, orders, and customer trends across different regions and years. The analysis highlights strong markets such as the USA, steady customer and order growth over time, and variations in shipping performance and employee tenure. These findings can support data-driven decision-making, improve strategic planning, and help identify new opportunities for growth.

**Prepared by:**  
*Priyanshu*