**Submitted By**: Mehtab Alam

**Email**: mehtabalam9155@gmail.com

Project Report on

**SALES ANALYSIS**

NORTHWIND TRADERS

**1. Project Overview**

The Northwind Traders Dashboard Project is designed to provide a comprehensive data analysis solution using Microsoft Excel, Power BI, and SQL. The aim of this project is to consolidate and analyze sales, customer behavior, inventory, and employee performance data from the Northwind Traders database, which stores essential business information for a fictitious company dealing with international food imports and exports. This dashboard will enable key stakeholders, such as managers and decision-makers, to easily understand and visualize critical metrics related to their business operations.

The project leverages SQL for data extraction and transformation, Microsoft Excel for preliminary data analysis and manipulation, and Power BI for interactive data visualization and reporting. The ultimate goal is to create a user-friendly dashboard that can drive informed decision-making and optimize business strategies across various aspects of Northwind Traders' operations**.**

**2. Problem Statement**

The objective of this project is to design a dynamic and interactive Power BI dashboard that communicates key performance indicators (KPIs) to stakeholders at Northwind Traders. With growing business complexity, Northwind Traders needs a centralized platform that can effectively showcase customer behavior, sales performance, employee productivity, and inventory trends in a visually appealing manner. Currently, accessing and analyzing data from multiple sources is time-consuming and complex, hindering timely decision-making.

By developing this Power BI dashboard, the project seeks to:

* Consolidate data from various tables in the Northwind database.
* Analyze business performance across various dimensions, such as sales, customer segments, inventory levels, and employee contributions.
* Empower stakeholders with a tool that enables them to interact with the data and draw actionable insights.

The dashboard aims to streamline the decision-making process and enhance the operational efficiency of Northwind Traders.

**3. Objective**

The primary objectives of this project are as follows:

* **Data Consolidation and Integration:** Combine data from multiple tables, such as Customers, Orders, Products, Employees, and Order Details, to provide a unified view of the company’s operations. This integration is facilitated using SQL queries that join and filter relevant data for analysis.
* **Exploratory Data Analysis (EDA):** Conduct an in-depth analysis of the sales data using Microsoft Excel, focusing on identifying trends, patterns, and potential outliers in areas such as sales growth, customer behavior, and inventory management.
* **Dashboard Design and Development**: Use Power BI to create an intuitive, interactive dashboard that displays key metrics like sales performance, customer segmentation, inventory levels, and employee performance. The dashboard will include dynamic filters to allow users to explore the data from different perspectives.
* **User Interactivity**: Ensure the dashboard provides features like drill-downs, slicers, and filters to facilitate detailed analysis. This interactivity will allow stakeholders to perform customized analysis and gain insights into specific segments or time periods.

**4. Dataset Description**

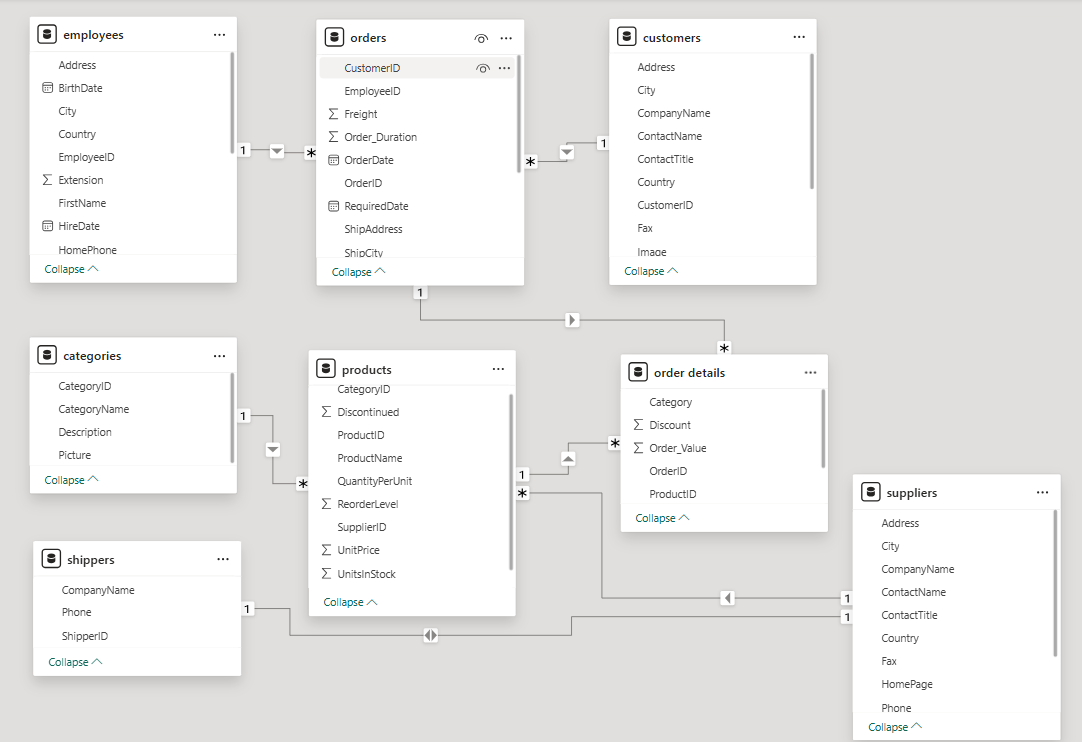
The Northwind database serves as the foundation for this project. It contains essential business data for the fictitious Northwind Traders company. The dataset includes tables with information on customers, employees, orders, products, suppliers, categories, and shippers, among other things. Below is a brief description of the key tables used in this project:

* **Customers Table**: Contains information about the company’s customers, including company name, contact details, address, and geographic information.
* **Employees Table**: Stores employee data, such as names, titles, contact details, hire date, and reporting relationships.
* **Orders Table**: Captures order-related data, including order IDs, customer IDs, employee IDs, order dates, shipping details, and freight costs.
* **Order Details Table**: Provides detailed information about the products within each order, including quantities, unit prices, and discounts.
* **Products Table**: Stores data on products, including product IDs, names, suppliers, unit prices, quantities in stock, and reorder levels.
* **Suppliers Table**: Contains information about suppliers, such as company names, contact details, and shipping information.
* **Shippers Table**: Includes data on shipping companies, including their names and contact details.
* **Categories Table**: Defines product categories and includes category names and descriptions.

**5. Methodology**

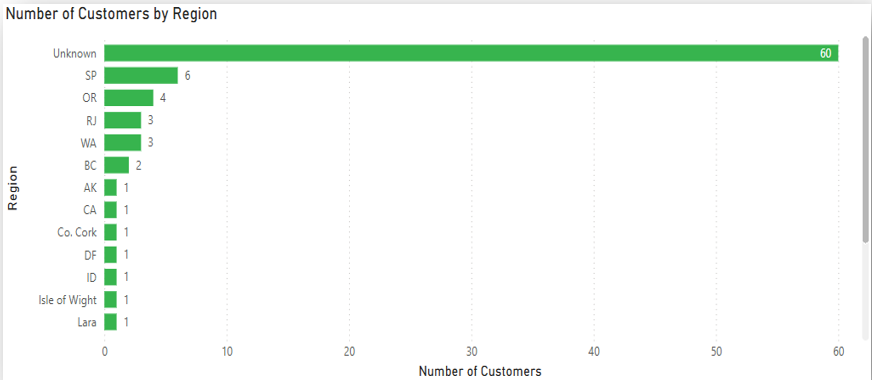
* **Data Extraction and Transformation (SQL**): The data was initially extracted from the Northwind database using SQL queries. Key transformations were applied to merge data from various tables, filter relevant records, and prepare it for analysis.
* **Data Analysis (Excel):** The raw data was imported into Microsoft Excel for preliminary analysis. This included calculating summary statistics, identifying trends, and preparing data for visualization. Pivot tables and charts were used to explore sales patterns, customer segments, and inventory trends.
* **Data Visualization (Power BI):** After preparing the data, the final step involved creating the dashboard in Power BI. Interactive visualizations, such as bar charts, line graphs, and pie charts, were used to display sales trends, inventory levels, and employee performance. Slicers and filters were incorporated to allow users to interact with the data.

**ER Diagram:**

****

**POWER BI**

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

****

**Visualization Description:**

The bar chart titled **“Number of Customers by Region”** displays the count of customers grouped by their respective regions. The x-axis represents the **number of customers**, while the y-axis lists different **region names**.

* A total of **16 regions** are included in the chart.
* The data is sorted in descending order of the number of customers per region.
* A prominent observation is the presence of a significant number of customers from the **"Unknown"** category, which dominates the chart.

**Insights and Interpretation:**

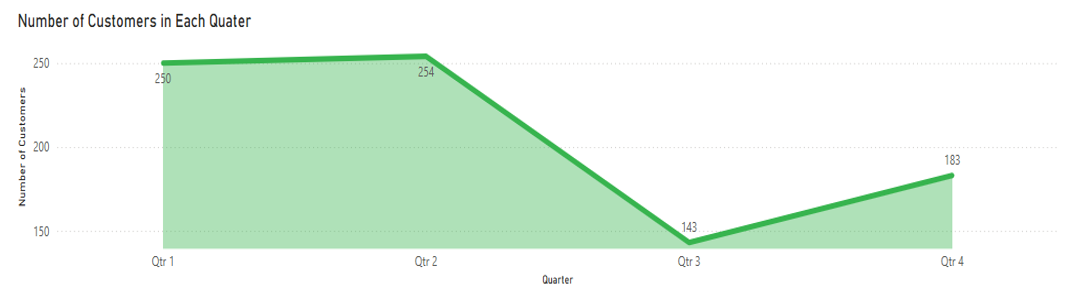
1. **High Count of "Unknown" Region:**
   * The largest segment, comprising **60 customers**, is labeled under the "Unknown" category.
   * This suggests either missing data or customers who **did not disclose** their region.
   * It is a **data quality issue** that may hinder region-based analytics and targeting strategies.
2. **Regions with Noticeable Customer Presence:**
   * **SP (São Paulo)** follows with **6 customers**, showing a modest concentration.
   * **OR (Oregon)** and **RJ (Rio de Janeiro)** have **4 and 3 customers** respectively.
   * These are the **only regions with customer counts greater than 2**, other than "Unknown".
3. **Long-Tail Distribution:**
   * Several regions have only **1 customer** each:
     + AK, CA, Co. Cork, DF, ID, Isle of Wight, Lara
   * This indicates a **sparse distribution** across a wide range of regions.

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

**2: Trend in Customer Acquisition Over Time – Time Series Analysis**

**Objective:**

To examine the quarterly trend in customer acquisition and understand how customer signups change across different time periods. The focus is on identifying growth, decline, and recovery phases.



**Visualization Description:**

The **area chart** titled **“Number of Customers in Each Quarter”** shows customer acquisition over four quarters.

* **X-axis (horizontal)**: Represents the **four quarters** (Qtr 1 to Qtr 4).
* **Y-axis (vertical)**: Shows the **number of customers acquired** in each quarter.
* A filled area beneath the line gives a clear visual cue for volume across time.
* Data values:
  + Qtr 1: **250 customers**
  + Qtr 2: **254 customers**
  + Qtr 3: **143 customers**
  + Qtr 4: **183 customers**

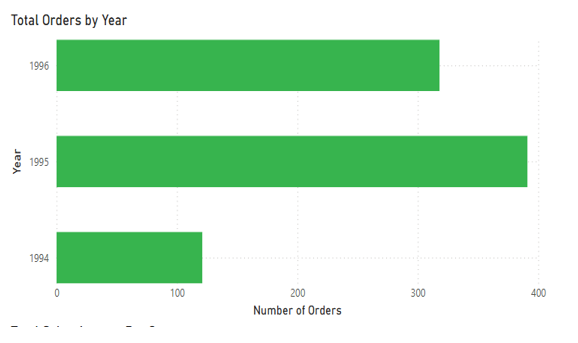
**Insights and Interpretation:**

1. **Initial Growth (Qtr 1 to Qtr 2):**
   * Slight increase in customer acquisition from **250 to 254**.
   * Indicates stable or slightly improving performance.
2. **Sharp Decline (Qtr 2 to Qtr 3):**
   * Dramatic drop to **143 customers**, a **43.7% decrease**.
   * This dip could be due to:
     + Seasonal trends
     + Marketing budget cuts
     + Product issues
     + External factors like economic slowdown
3. **Recovery Phase (Qtr 3 to Qtr 4):**
   * Moderate increase to **183 customers**.
   * Shows potential **recovery or response to corrective actions**, such as new campaigns or product improvements.

**3: Order Volume Trends Over the Years – Time Series Summary**

**Objective:**

To analyze how the number of orders changed annually and determine which year had the highest and lowest order volumes.



**Visualization Description:**

The **horizontal bar chart** titled **“Total Orders by Year”** displays order volume across three years.

* **Y-axis (vertical)**: Represents the **year** – 1994, 1995, and 1996.
* **X-axis (horizontal)**: Indicates the **number of orders** placed each year.
* **Data values:**
  + **1994**: ~120 orders (lowest)
  + **1995**: ~390 orders (highest)
  + **1996**: ~330 orders

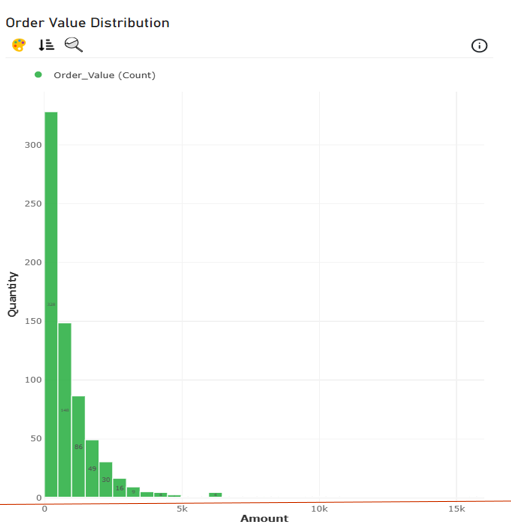
**Insights and Interpretation:**

1. **Growth Spike (1994 → 1995):**
   * Orders more than tripled from ~120 to ~390.
   * Suggests a major growth phase – possibly due to:
     + Expansion in customer base
     + New product launches
2. **Slight Decline (1995 → 1996):**
   * Orders dropped slightly to ~330.
   * While lower than 1995, 1996 still maintained strong performance compared to 1994.
   * Could indicate market saturation or shift in demand.

**4: Order Value Distribution – Histogram Analysis**

**Objective:**

To analyze the frequency distribution of order values and understand how product orders are spread across different value ranges.



**Visualization Description:**

The **vertical histogram** titled **“Order Value Distribution”** represents how many products were ordered within specific monetary ranges.

* **X-axis (horizontal):** Order value ranges (e.g., 0–500, 500–1000, ..., up to 15,000).
* **Y-axis (vertical):** Quantity of products sold (frequency).
* **Green bars** represent the number of products sold within each order value bin.

**Key Observations:**

1. **Highly Skewed Distribution (Right-Skewed):**
   * Majority of orders are **concentrated in the lower value ranges**.
   * The **0–500** range has the **highest count (328 products)**.
   * The counts drop off steeply as the order value increases.
2. **Sparse High-Value Orders:**
   * Very few orders were placed in high-value ranges (above ₹6000).
   * Example: Only **4 orders** in the **₹6000–₹6500** range.
   * One bar appears around ₹10,000–₹15,000 with a count of **1**, showing outliers.

**Insights and Interpretation:**

* **Typical Order Size:** Most customers prefer low-to-mid value purchases.
* **Revenue Strategy:** Even though high-value orders are rare, they may contribute significantly to revenue per order.
* **Product Affordability:** Majority of products fall in accessible price bands, indicating affordability drives sales.

**Implications:**

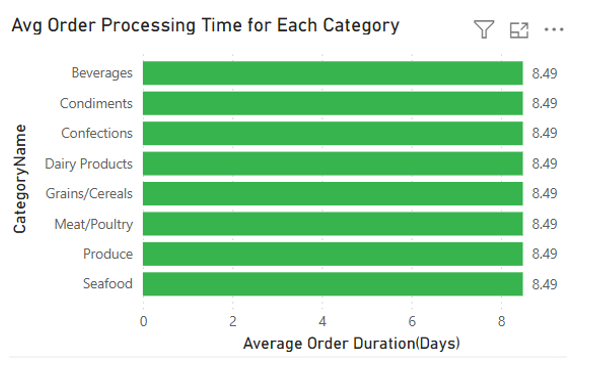
* **Inventory Planning:** Focus on keeping low-value items well-stocked due to high demand.
* **Marketing Strategy:**
  + Bundle lower-value products to increase average order value.
  + Promote high-value items with discounts or financing options to increase adoption.

**5: Average Order Processing Time – Bar Chart Analysis**

**Question:**

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

**Answer:** Yes. The bar chart below is a clear example of visualizing the average order processing time by product category.



**Visualization Description:**

The **horizontal bar chart** titled **“Avg Order Processing Time for Each Category”** shows:

* **Y-axis (Category Name):** Lists all product categories like Beverages, Condiments, Produce, etc.
* **X-axis (Average Order Duration in Days):** Fixed value across all categories.
* **Bar Values:** Every category shows an average of **8.49 days**.

**Key Observations:**

* **Uniform Processing Time:**
  + Each category has the **exact same average order duration** of **8.49 days**.
  + No variation exists among the different categories.
* **No Outliers or Spread:**  
  This visualization doesn't reveal any spread or range in values; it only shows a **mean**.

**Insights and Interpretation:**

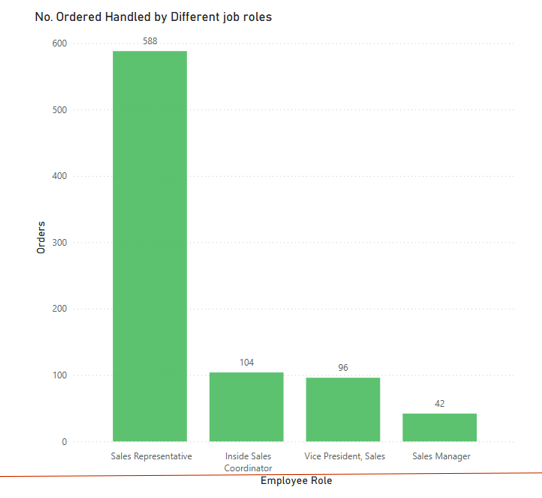
* **Standardized Order Workflow:**  
  It appears that the order processing system applies a uniform process regardless of category.

**Implications:**

* **No Category Requires Special Handling Timewise:**
  + From a logistics perspective, all items are treated equally.

**6: Employee Productivity by Job Role**

To evaluate employee productivity across different job roles, the number of orders handled by each role was analyzed and visualized using a column chart. This chart provides a clear comparison of performance levels among key sales-related positions.

****

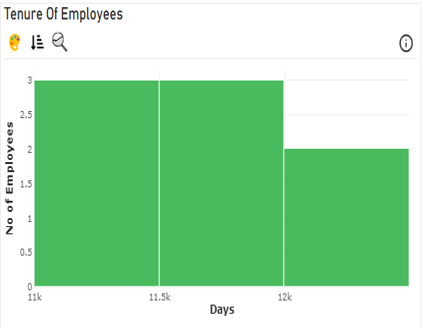
**Key Insights:**

* The Sales Representative role handled the highest volume of orders, with a total of 588 orders, indicating a strong contribution to day-to-day operational activities.
* The Inside Sales Coordinator and Vice President, Sales roles followed with 104 and 96 orders, respectively.
* The Sales Manager role handled the fewest orders, totaling 42, suggesting a possible focus on management and oversight responsibilities rather than direct order handling.

**Interpretation:**

The distribution of order volume highlights that operational roles like Sales Representatives are primarily engaged in order processing tasks. In contrast, managerial and executive roles appear to be more focused on strategic functions, which is reflected in their lower order counts.

**7: Analysis of Employee Tenure Distribution**

****

To understand the tenure distribution of employees, a histogram was generated using the number of days each employee has been with the organization. The x-axis represents the tenure in days, divided into intervals, while the y-axis indicates the number of employees within each interval.

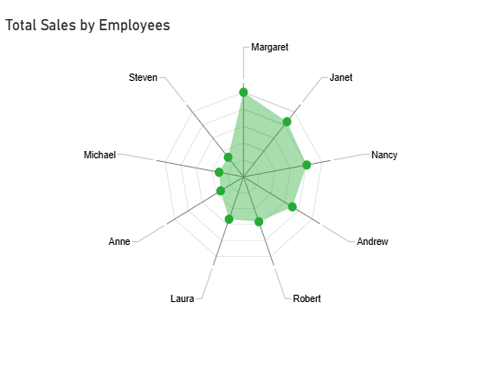
The data reveals the following:

* Three employees have tenures between 11,000 and 11,500 days.
* Two employees fall within the 11,500 to 12,000 days range.
* Two employees have worked for more than 12,000 days.

This distribution shows a concentration of employee tenure between 11,000 and 12,000 days, with a small number extending beyond that range. Overall, five out of seven employees have tenure below 12,000 days, while the remaining two have slightly longer tenure.

**8: Employee Performance by Total Sales**

To assess individual employee performance, a radar chart was used to visualize the total sales (order amounts) handled by each employee. This approach offers a clear comparative view of performance levels across a group of employees, making it easier to identify both top and underperformers.



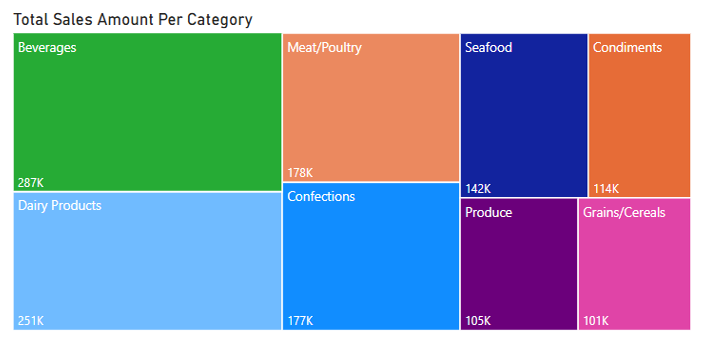
**Key Insights**:

* Margaret recorded the highest total order amount, indicating outstanding individual sales performance.
* Janet and Nancy also performed strongly, handling substantial volumes of sales.
* On the other hand, Steven handled the lowest order amount, suggesting a need for performance review or potential role reassessment.
* Employees such as Anne, Laura, Michael, and Robert also exhibited relatively lower performance compared to top contributors.

**Interpretation:**  
The radar chart reveals a wide variation in performance among employees. High-performing individuals like Margaret play a critical role in driving sales, while those on the lower end may require additional support, training, or revised targets. This visualization helps pinpoint where to focus improvement efforts and how to balance team productivity.

**9: Sales Volume by Product Category**

To evaluate sales performance across various product categories, a treemap chart was utilized to visualize the total sales amount contributed by each category. The size of each block within the treemap corresponds to the total sales value, offering an intuitive and proportional comparison.

****

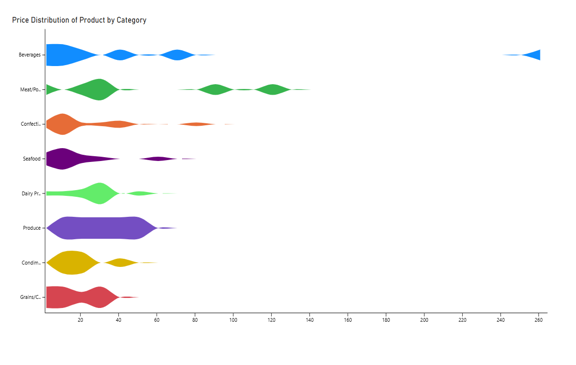
**Key Insights:**

* The Beverages category recorded the highest total sales, amounting to 287K, making it the top-performing product segment.
* Dairy Products followed closely with 251K in sales, also contributing significantly to overall revenue.
* Other notable categories include Meat/Poultry (178K) and Confections (177K).
* The Grains/Cereals category accounted for the lowest sales, totaling 101K, indicating relatively lower consumer demand or limited product offerings in this segment**.**

**Interpretation:**The treemap visualization effectively highlights disparities in sales distribution among product categories. Categories like Beverages and Dairy Products dominate in terms of revenue generation, suggesting strong market preference or product turnover. In contrast, categories such as Grains/Cereals and Produce show comparatively lower sales, which may require deeper analysis into product placement, pricing strategy, or customer preferences.

**10: Product Pricing Distribution by Category**

**To analyze how prices vary across different product categories, a violin plot was used. This type of plot effectively displays both the distribution and density of product prices, providing insights into the range and variability within each category.**

****

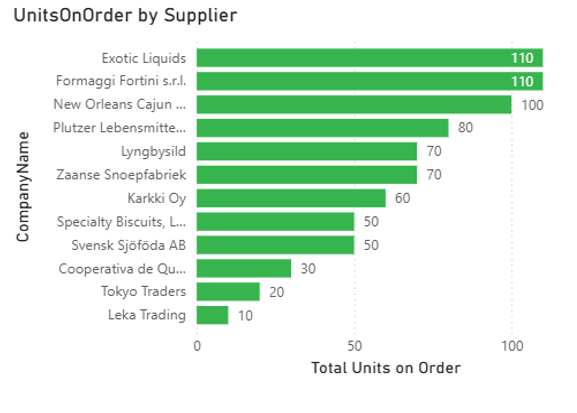
**Key Insights:**

* The "Produce" category exhibits the widest price distribution, as shown by the broader and more uneven violin shape. This suggests significant variation in product pricing within this category.
* In contrast, the "Confectionery" category has a narrower and more concentrated price range, reflected by a thinner violin shape. This implies more consistent pricing across products in this group.
* Other categories, such as Seafood and Condiments, show moderate variability, indicating a mix of standard and premium-priced products.

**Interpretation:**The violin plot reveals important pricing dynamics across product categories. Broad distributions like that of Produce may indicate a diverse product range or fluctuating market prices. Meanwhile, narrow distributions, such as in Confectionery, may reflect standardized pricing or limited product variation. Understanding these patterns can help in pricing strategy, inventory planning, and market positioning.

**11: Supplier Performance Metrics (Units on Order)**

**To evaluate supplier performance, the number of units on order from each supplier was used as a performance metric. This bar chart visualization helps identify the suppliers contributing most to inventory inflow based on current order volumes.**

****

**Key Insights:**

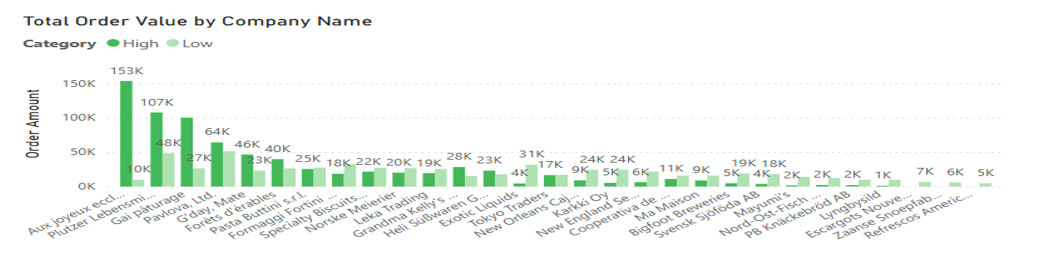
* Exotic Liquids and Formaggi Fortini s.r.l. lead with the highest units on order (110 each), indicating strong supplier engagement or high product demand.
* Suppliers such as New Orleans Cajun, Plutzer Lebensmittelgroßmärkte, and Lyngbysild follow with substantial order volumes, showcasing consistent supply activity.
* On the lower end, suppliers like Leka Trading and Tokyo Traders have minimal units on order, potentially suggesting limited product range, niche demand, or recent onboarding**.**

**Interpretation:**This bar chart clearly highlights disparities in supplier activity levels. High order volumes may reflect top-performing suppliers or high-turnover products, while low numbers could suggest opportunities for growth or underutilization. Tracking such metrics aids in supplier performance management and procurement planning.

**12: Pricing Structure Across Suppliers**

**To understand how pricing varies across suppliers, this stacked bar chart breaks down total order value into two pricing categories:**

* **High: Products priced above 1000**
* **Low: Products priced below 1000**

****

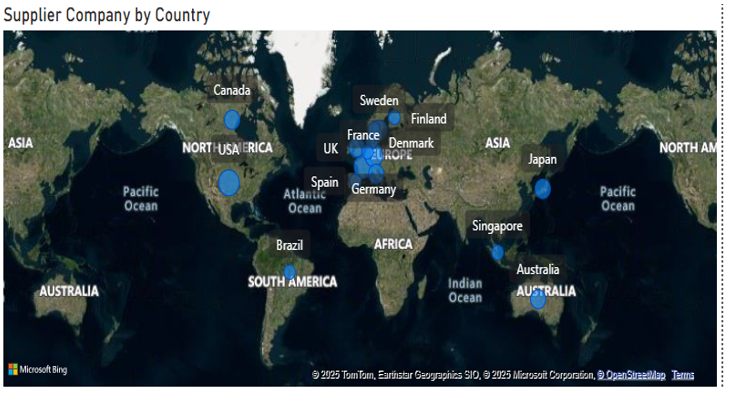
**Key Observations:**

* Aux joyeux ecclésiastiques leads with a total order value of 153K, predominantly contributed by high-priced products, indicating a premium supplier.
* Plutzer Lebensmittelgroßmärkte AG and Gai pâturage follow with 107K and 64K in total order value, respectively.
* A clear disparity is visible between suppliers whose value is mostly in high-price tiers versus those focused on lower-cost items.

**Interpretation:**This visualization helps differentiate suppliers based on pricing strategy. Suppliers with a high share of expensive product orders (like Aux joyeux ecclésiastiques) may specialize in luxury or niche products, while others (like Formaggi Fortini or Tokyo Traders) may cater to more economical or high-volume segments. Such insights support procurement decisions and vendor segmentation.

**13: Geographical Distribution of Suppliers**

**To understand where suppliers are located globally, this map chart visualizes the supplier distribution using bubbles sized by frequency or concentration.**

****

**Key Observations:**

* A significant cluster of suppliers is located in Europe, including countries like Germany, France, UK, Denmark, Sweden, and Finland.
* Other suppliers are spread across North America, Asia, and Australia, but with smaller representation.
* Countries like Brazil, Singapore, Japan, and Australia host fewer suppliers, suggesting more regional diversification in those cases.

**Interpretation:**The map illustrates a European dominance in the supplier base, potentially due to historical trade ties, supply chain integration, or proximity advantages. This insight can guide logistics planning and risk assessment, especially in scenarios involving geopolitical or economic disruptions in specific regions.

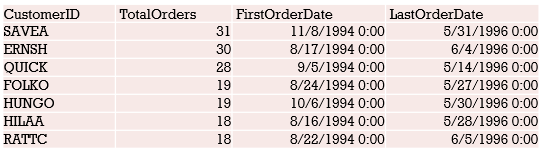
**Microsoft Excel**

**1: Customer Retention & Loyalty Insights**

**To identify customer loyalty, we analyzed the number of orders placed over time using the following SQL query:**

SELECT

CustomerID,



COUNT(OrderID) AS TotalOrders,

MIN(OrderDate) AS FirstOrderDate,

MAX(OrderDate) AS LastOrderDate

FROM

Orders

GROUP BY

CustomerID

ORDER BY

TotalOrders DESC;

**Key Observations:**

* CustomerID "SAVEA" has the highest number of orders (31), with activity spanning from November 1994 to May 1996.
* Other loyal customers include ERNSH and QUICK, with 30 and 28 orders respectively.
* Consistent order frequency over multiple years suggests strong customer engagement and potential satisfaction.

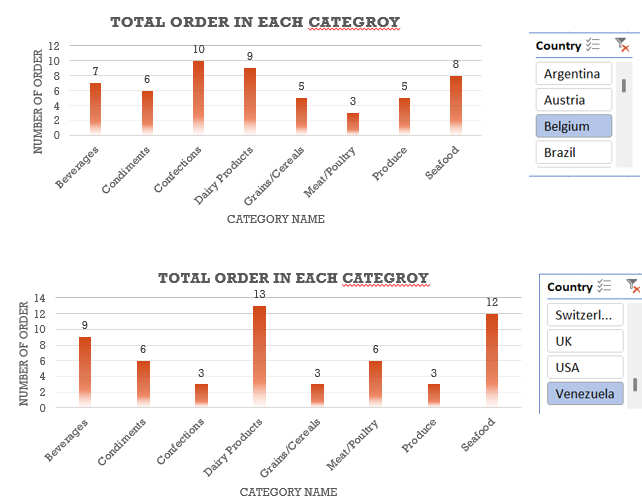
**Interpretation:**This analysis highlights frequency and duration of purchases as critical indicators of loyalty. Customers with high order counts and long engagement periods (like SAVEA) are valuable targets for retention strategies and personalized marketing.

**2**: **Customer Preferences by Location – Insights from Interactive Visualizations**

To understand how product demand differs by geography, we used the following SQL query:

WITH Data AS (

SELECT



C.Country,

CC.CategoryName,

COUNT(O.OrderID) AS Total

FROM customers C

INNER JOIN orders O ON C.CustomerID = O.CustomerID

INNER JOIN order\_details D ON D.OrderID = O.OrderID

INNER JOIN products P ON D.ProductID = P.ProductID

INNER JOIN categories CC ON P.CategoryID = CC.CategoryID

GROUP BY C.Country, CC.CategoryName

)

SELECT Country, CategoryName, Total FROM Data;

**Key Observations:**

* **Belgium** shows the highest demand in the **Confections** category, indicating a strong preference for sweets or desserts.
* **Venezuela**, on the other hand, has a significantly higher demand for **Dairy Products**, followed closely by **Seafood**.

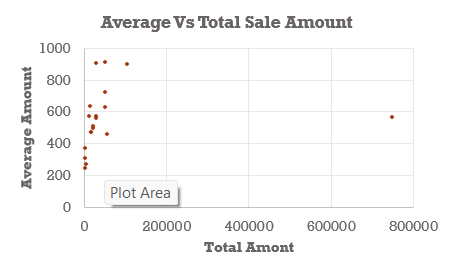
**Interpretation:**  
These visualizations clearly show how regional tastes and demographics influence product preferences. This insight can guide:

* Localized marketing strategies
* Inventory planning
* Tailored product recommendations for each market

**3**: **Identifying Customer Behavior Patterns Across Regions**

To analyze regional sales behavior, we used the following SQL query:

SELECT



COALESCE(C.Region, 'Unknown') AS Region,

COUNT(DISTINCT O.OrderID) AS Total\_Customers,

ROUND(SUM(D.UnitPrice \* D.Quantity \* (1 - D.Discount)), 0) AS Total\_Amount,

ROUND(AVG(D.UnitPrice \* D.Quantity \* (1 - D.Discount)), 0) AS Average\_Amount

FROM customers C

INNER JOIN orders O ON C.CustomerID = O.CustomerID

INNER JOIN order\_details D ON O.OrderID = D.OrderID

GROUP BY Region

ORDER BY Total\_Customers DESC, Total\_Amount DESC;

**Insights from the Scatter Plot (Average vs. Total Sales Amount):**

* The distribution is widespread, with no clear clustering of regions.
* Regions with the highest total sales don't necessarily exhibit high average order values—they tend to have moderate averages.
* This suggests that volume of transactions rather than order size is driving total sales in top-performing regions.

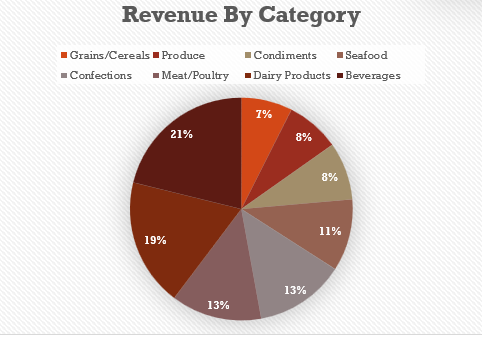
**Interpretation & Business Implications:**

* No distinct clusters indicate the market is fragmented in behavior.
* Marketing strategies should focus on customer engagement and transaction frequency rather than simply increasing order size.

**4: Which Product Categories Drive the Most Revenue?**

To understand revenue contribution by product category, we used the following query:

SELECT



C.CategoryName,

SUM(D.UnitPrice \* D.Quantity) AS Revenue

FROM categories C

INNER JOIN products P ON P.CategoryID = C.CategoryID

INNER JOIN order\_details D ON P.ProductID = D.ProductID

GROUP BY C.CategoryName;

**Insights from the Pie Chart ("Revenue by Category"):**

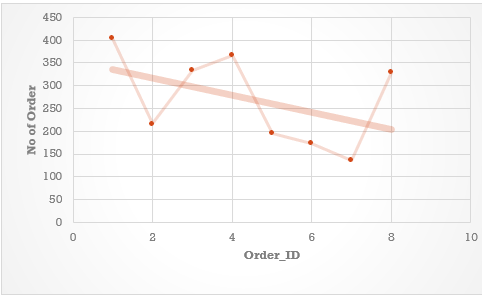
* Beverages is the top revenue-generating category, contributing 21% of total revenue.
* Other strong performers include Dairy Products (19%) and Meat/Poultry (13%).
* The category with the lowest revenue contribution is Grains/Cereals (7%).

**Interpretation & Business Strategy:**

* The business may benefit from targeted promotions for high-margin categories like Beverages to sustain revenue.
* Low-performing categories like Grains/Cereals may require a reassessment of pricing, placement, or promotion strategies.

**5:Do Order Sizes Correlate with Product Categories?**

To explore whether product categories influence order volumes, the following SQL query was used:



SELECT

C.CategoryID,

COUNT(D.OrderID) AS Total\_Orders

FROM categories C

INNER JOIN products P ON C.CategoryID = P.CategoryID

INNER JOIN order\_details D ON P.ProductID = D.ProductID

GROUP BY C.CategoryID;

**Insights from the Chart:**

* The line plot visualizes the number of orders for each product category (identified by CategoryID).
* There’s a general downward trend, suggesting that higher CategoryIDs are associated with fewer orders.
* A negative correlation is visible, although the relationship isn't strictly linear.

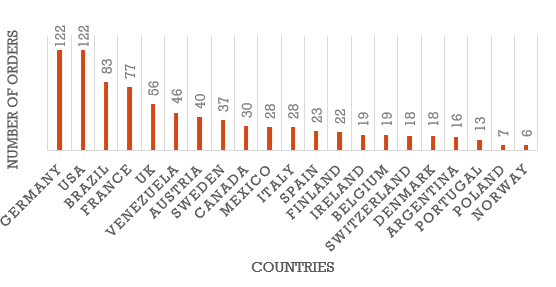
**Interpretation & Considerations:**

* While CategoryID itself is a surrogate key, this trend may still reflect category popularity if IDs were assigned in a chronological or logical order.

**6**:**How Does Order Frequency Vary Across Customer Segments?**

To explore geographic trends in order frequency, the following SQL query was used:

SELECT



C.Country,

COUNT(O.OrderID) AS Total\_Orders

FROM Customers C

INNER JOIN Orders O ON C.CustomerID = O.CustomerID

GROUP BY C.Country

ORDER BY Total\_Orders DESC;

**Insights from the Bar Chart:**

* Germany and the USA lead with the highest number of orders (122 each).
* Brazil, France, and the UK follow with substantial order counts.
* Countries like Norway, Poland, and Portugal show relatively low order activity.

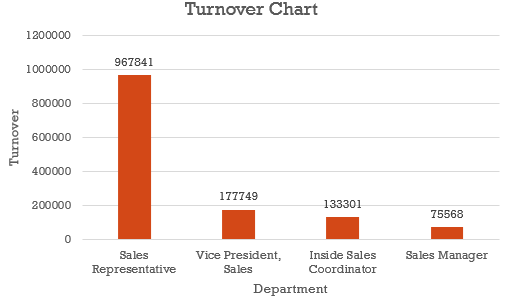
**Interpretation:**

* This distribution can highlight key markets and potential regions for growth.
* Countries with high order frequencies may reflect larger customer bases, better market penetration, or higher engagement.

**7: How Does Order Frequency Vary Across Customer Segments?**

**To explore geographic trends in order frequency, the following SQL query was used:**

SELECT



C.Country,

COUNT(O.OrderID) AS Total\_Orders

FROM Customers C

INNER JOIN Orders O ON C.CustomerID = O.CustomerID

GROUP BY C.Country

ORDER BY Total\_Orders DESC;

**Insights from the Bar Chart:**

* Germany and the USA lead with the highest number of orders (122 each).
* Brazil, France, and the UK follow with substantial order volumes.
* Countries like Norway, Poland, and Portugal show relatively low order activity.

**Interpretation:**

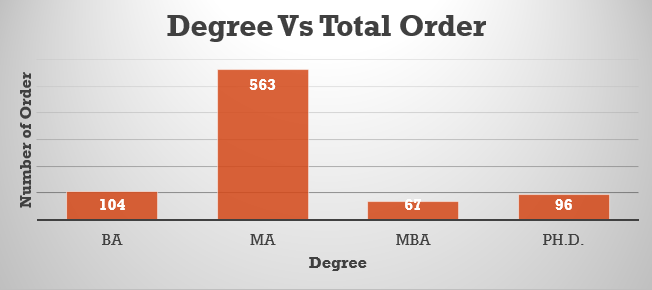
* This distribution can help identify key markets and uncover opportunities for expansion.
* Countries with high order frequencies likely reflect larger customer bases, better market penetration, or stronger customer engagement.

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

**To analyze the impact of educational qualifications on order processing, the following SQL query was used:**

SELECT

E.FirstName,



CASE

WHEN E.Notes LIKE '%Ph.D.%' THEN 'Ph.D.'

WHEN E.Notes LIKE '%MBA%' THEN 'MBA'

WHEN E.Notes LIKE '%MA%' THEN 'MA'

WHEN E.Notes LIKE '%BA%' THEN 'BA'

WHEN E.Notes LIKE '%BS%' THEN 'BS'

WHEN E.Notes LIKE '%BSC%' THEN 'BSC'

ELSE 'Unknown'

END AS Degree,

COUNT(O.OrderID) AS total\_orders

FROM employees E

INNER JOIN orders O ON E.EmployeeID = O.EmployeeID

GROUP BY E.FirstName, 2

ORDER BY total\_orders DESC;

**Insights from the Bar Chart:**

* Employees with an MA degree completed the highest number of orders (563).
* Employees holding a BA degree followed with 104 orders.
* Ph.D. and MBA holders recorded lower order counts—96 and 67 respectively.

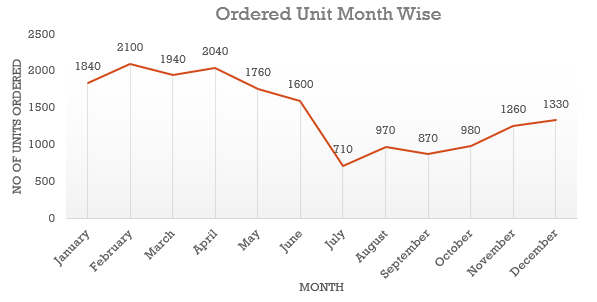
**Interpretation:**

* Employees with MA degrees appear to contribute the most to order fulfillment, possibly indicating better alignment of their qualifications with job responsibilities.
* This insight can guide talent acquisition and training, suggesting that certain degrees may correlate with higher performance in order handling roles.
* It also helps identify skill clusters that may be prioritized in hiring or internal development programs.

**9: How does product demand fluctuate over different seasons or months? Can we visualize this through line charts or area charts?**

**To examine monthly trends in product orders, the following SQL query was used:**

SELECT



MONTHNAME(O.OrderDate) AS Months,

SUM(P.UnitsOnOrder) AS OrderedUnits

FROM orders O

INNER JOIN order\_details D ON O.OrderID = D.OrderID

INNER JOIN products P ON D.ProductID = P.ProductID

GROUP BY 1

ORDER BY 2 DESC;

**Insights from the Line Chart:**

* February shows the highest number of ordered units (2100), followed closely by April (2040) and March (1940).
* A decline begins after April, reaching the lowest point in July (710).
* From August onward, order volumes begin to recover gradually, ending at 1330 units in December.

**Interpretation:**

* Seasonality plays a significant role in product demand, with peaks early in the year and dips during mid-year (summer months).
* Businesses can plan inventory, promotions, and staffing around these trends to optimize operations.

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

**To detect outliers in total product sales, the following SQL query was used:**

SELECT

P.ProductName,

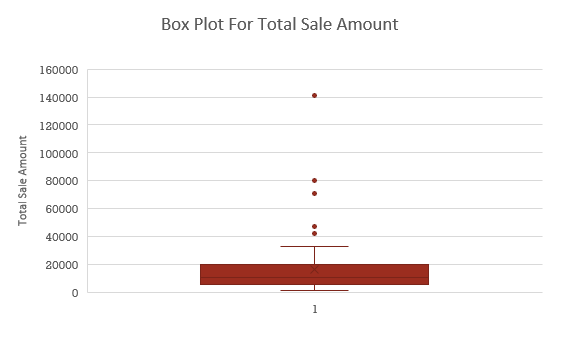
COUNT(DISTINCT D.OrderID) AS Total\_Count,

ROUND(SUM(D.UnitPrice \* D.Quantity \* (1 - D.Discount)), 0) AS Total\_Sales\_Amount

FROM products P

INNER JOIN order\_details D ON P.ProductID = D.ProductID

GROUP BY 1;



**Insights from the Box Plot:**

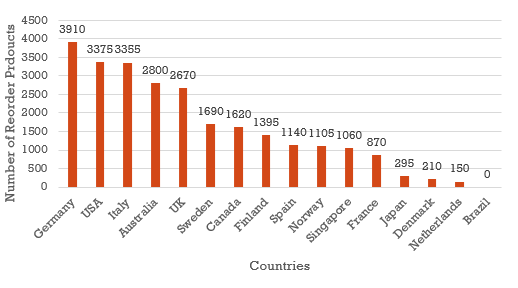
* Outlier sales values include 41,820, 46,825, 71,156, 80,369, and the highest at 141,347.
* These values lie significantly above the interquartile range, marking them as statistical outliers.

**Interpretation:**

* The products corresponding to these high sales figures are top performers and warrant deeper analysis.
* Product optimization opportunities include:
  + Scaling production of these high-demand items.
  + Enhancing visibility through targeted marketing campaigns.
  + Improving distribution channels to meet demand effectively.

**11: Are there any correlations between supplier attributes and performance metrics?**

SELECT



S.Country,

SUM(P.ReorderLevel) AS Total\_Order\_Products

FROM suppliers S

INNER JOIN products P ON S.SupplierID = P.SupplierID

INNER JOIN order\_details D ON P.ProductID = D.ProductID

INNER JOIN orders O ON O.OrderID = D.OrderID

GROUP BY 1;

**Insights from the Bar Chart:**

* Top 5 Countries by Reorder Volume:
  + Germany (3,910)
  + USA (3,375)
  + Italy (3,355)
  + Australia (2,800)
  + UK (2,670)
* Notable Observation: Brazil had 0 reorder products, making it an outlier with minimal to no replenishment demand.

**Interpretation:**

* Suppliers from Germany, USA, Italy, Australia, and UK are key contributors to the reorder volume, possibly due to:
  + Strong distribution networks
  + Reliable delivery timelines
  + Product popularity or high consumption rates
* Brazil's zero value could indicate:
  + Ineffective supply
  + Poor product demand
  + Data anomalies or inactive suppliers

**12: How does supplier performance vary across product categories?**

SELECT

C.CategoryName,

ROUND(AVG(DATEDIFF(O.ShippedDate, O.OrderDate)), 0) AS average\_Ship\_Time

FROM categories C

INNER JOIN Products P ON C.CategoryID = P.CategoryID

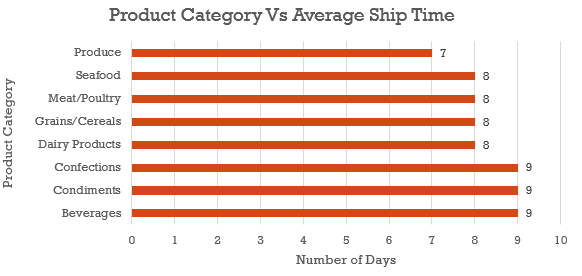
INNER JOIN order\_details D ON P.ProductID = D.ProductID

INNER JOIN orders O ON O.OrderID = D.OrderID

GROUP BY 1

ORDER BY 2 DESC;

**Chart Analysis – Average Ship Time by Product Category:**



| **Product Category** | **Avg Ship Time (days)** |
| --- | --- |
| Beverages | 9 |
| Condiments | 9 |
| Confections | 9 |
| Dairy Products | 8 |
| Grains/Cereals | 8 |
| Meat/Poultry | 8 |
| Seafood | 8 |
| **Produce** | **7** |

**Key Observations:**

* Longest shipping delays (9 days) were seen in Beverages, Condiments, and Confections.
* Fastest category: Produce, which took 7 days on average — indicating higher efficiency or priority.

**Business Insights:**

* Optimize slow categories:
  + Investigate supply chain delays for Beverages and Confections.
  + Consider localizing suppliers or improving packaging/distribution workflows.
* Use Produce as a benchmark:
  + Identify best practices in that category and replicate across others.

**13: Can we identify trends in supplier costs or pricing through visualizations?**

SELECT

S.SupplierID,

ROUND(SUM(D.UnitPrice \* D.Quantity \* (1 - Discount)), 0) AS Total\_Sales\_Amount,

COUNT(DISTINCT O.OrderID) AS Total\_Orders,

ROUND(SUM(O.Freight), 0) AS Transport\_Cost,

ROUND((SUM(O.Freight) / SUM(D.UnitPrice \* D.Quantity \* (1 - Discount))) \* 100, 2) AS percent

FROM suppliers S

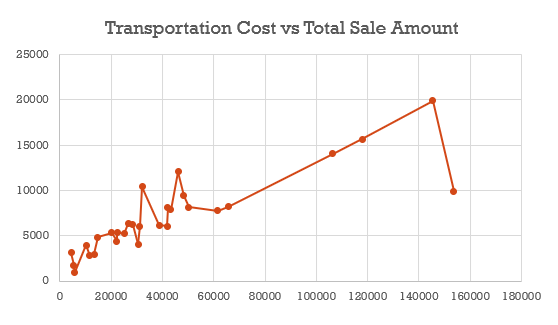
INNER JOIN products P ON S.SupplierID = P.SupplierID

INNER JOIN order\_details D ON D.ProductID = P.ProductID

INNER JOIN orders O ON D.OrderID = O.OrderID

GROUP BY 1

ORDER BY 2 DESC;



**Chart Analysis – Transportation Cost vs Total Sales Amount:**

* A positive correlation is observed: as total sales increase, the transportation cost (absolute) also rises.
* SupplierID 18 is an outlier:
  + Has the highest total sales.
  + But maintains a very low transportation cost percentage (~6%).

**Procurement Optimization Insights:**

1. Benchmark Efficient Suppliers:
   * Suppliers like SupplierID 18 can serve as efficiency benchmarks.
   * Investigate how they manage logistics, possibly using bulk shipping, localized hubs, or better freight contracts.
2. Cost Control Measures:
   * For suppliers with high transport cost percentages, explore opportunities to bundle shipments, use alternate carriers, or consolidate orders.

**Thank You!**