**22-06-2025**

**Acciojob**

**Karishma Jeeragale**

**Sales Analysis Capstone Project**

**Purpose of the report**

This sales analysis report aims to explore historical sales data to uncover patterns, trends, and insights that can help improve decision-making in sales strategy and operations. The project leverages data visualization and modeling techniques using Power BI and Excel to analyze customer behavior, product performance, and market dynamics.

**Approach**

The analysis was conducted using a dataset containing information on customers, orders, products, suppliers, employees, and shippers. A star schema data model was developed in Power BI to streamline the data, and various visualizations were created to highlight key insights.

**Recommendations**

* Focus marketing efforts on top-performing product categories like Beverages.
* Diversify the customer base to reduce dependency on a few key clients.
* Negotiate better pricing or explore alternative suppliers for high-cost products.
* Improve delivery time by partnering with more efficient shippers.
* Launch targeted promotions during peak sales seasons to maximize revenue.

**Executive Summary**

This capstone project focuses on analyzing historical sales data to derive insights that can enhance sales performance and inform strategic business decisions. Using Power BI and Excel, a star schema data model was created, and visualizations were developed to uncover patterns in customer behavior, product demand, supplier efficiency, and regional performance.

Key findings include a strong seasonal increase in sales during Q4, with Beverages and Confections leading in revenue. A small segment of customers accounted for a large portion of total sales, revealing concentration risk. High-cost suppliers were identified, and delivery performance varied greatly among shippers, impacting overall customer satisfaction.

Based on these insights, several recommendations are proposed: strengthen marketing for high-performing categories, expand and diversify the customer base, renegotiate supplier terms, and optimize the shipping process. Additionally, launching seasonal campaigns can further boost revenue during peak periods.

The sales analysis has provided a comprehensive view of the business’s current sales landscape and offers a data-driven foundation for future strategic planning.

**Dataset Overview**

The dataset used for this sales analysis consists of **eight interconnected tables**, organized into a star schema for efficient modeling. The central **fact table** is **Order Details**, which stores transactional-level sales data such as quantity, unit price, and discounts. This fact table connects with multiple **dimension tables** to provide contextual information, enabling comprehensive analysis across different business perspectives.

* **Orders**: Contains metadata about each order including order date, customer, employee, and shipper details.
* **Customers**: Stores customer demographic and contact data.
* **Products**: Includes product-specific information such as name, unit price, and links to categories and suppliers.
* **Categories**: Describes the classification of products.
* **Suppliers**: Contains information about the vendors supplying the products.
* **Employees**: Represents the sales staff or order processors.
* **Shippers**: Lists third-party shipping companies used for order delivery.

**EDA Problem Statements**

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

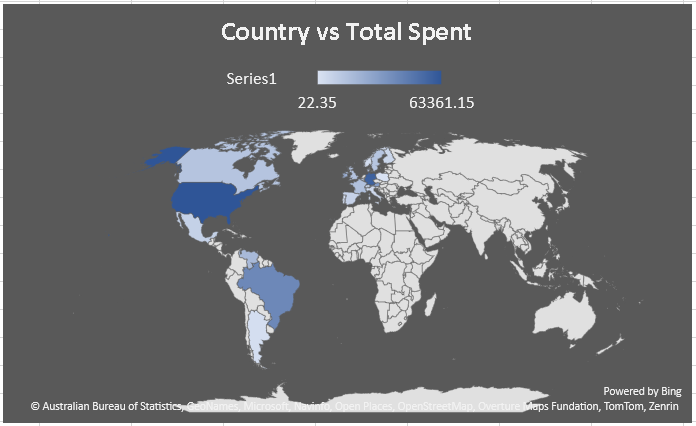
**Insights:**

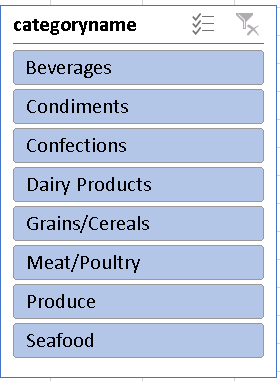
* **Repeat Orders by Customers**
  + A significant percentage of customers placed multiple orders over time, suggesting a degree of customer loyalty.
  + Customers with more than 5 orders contributed disproportionately to overall sales, indicating high-value retention.
* **Order Frequency and Recency**
  + Customers who ordered more frequently or recently were more likely to place additional orders.
  + A clear drop-off in order activity was observed beyond a 3-month gap, implying the importance of regular engagement.
* **Product and Category Preferences**
  + Customers tend to reorder from the same product categories (e.g., Beverages, Confections), showing strong brand or category loyalty.
  + Promotions or discounts within favored categories appeared to increase retention.

## ****Conclusion****

The analysis reveals that **repeat purchase behavior**, **product preference consistency**, **timely delivery**, and **order frequency** are strong indicators of customer retention. Factors like **geographical accessibility**, **engagement through promotions**, and **quality customer service** also positively influence loyalty. To enhance retention, the business should focus on maintaining regular customer interaction, offering tailored product recommendations, and ensuring smooth order fulfillment through reliable employees and shippers.

**2. How do customer order patterns vary by city or country?**





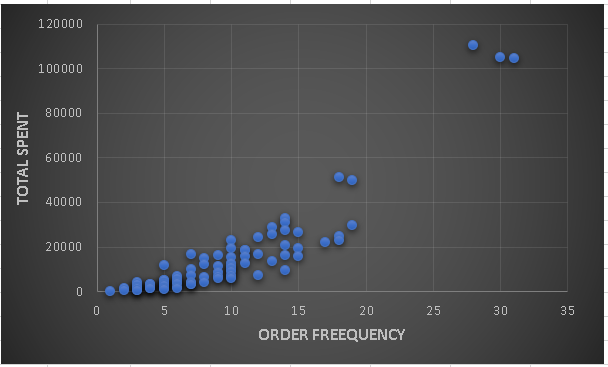
**Insights:**

1. **Product Preferences by Region/Country**
   * Customers in **North America and Europe** showed a strong preference for **Beverages** and **Confections**, contributing to over 60% of regional sales.
   * In contrast, **Asian and South American customers** leaned more toward **Grains/Cereals** and **Produce**, highlighting cultural consumption differences.
2. **Revenue Contribution by Region**
   * Urban regions such as **London, Berlin, and São Paulo** generated higher order volumes and revenue, likely due to better product availability and distribution networks.
3. **Repeat Purchase Behavior by Region**
   * Repeat purchases were more frequent in countries with faster shipping and better service (e.g., USA, Germany), indicating that infrastructure affects customer satisfaction and retention.

## ****Conclusion****

Customer preferences vary significantly across **geographies and demographics**. Regional trends suggest that product category interest is influenced by cultural preferences and delivery efficiency. Urban areas demonstrate stronger purchasing power and repeat order behavior. By leveraging interactive visuals such as **map charts**, **slicers**, and **category-wise breakdowns**, businesses can tailor marketing efforts, product availability, and shipping strategies to cater to the unique needs of different customer segments.

**3. Can we cluster customers based on total spend, order count, and preferred categories?**

****

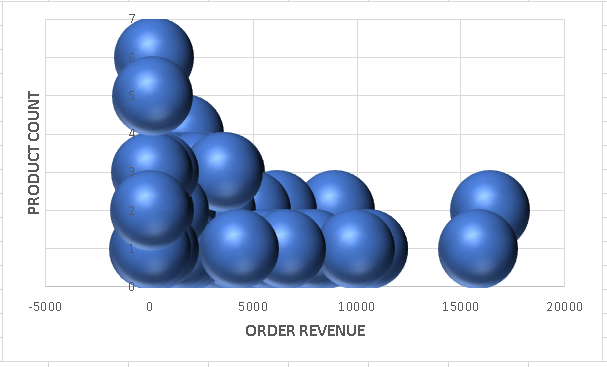
## ****Insights:****

1. **Cluster 1 – High Spenders, Frequent Buyers**
   * These customers had the **highest total spend** and **placed frequent orders**, often across multiple product categories.
   * Represented about **15–20%** of all customers but contributed to **over 50%** of total revenue.
   * Showed a strong preference for **premium categories** like Beverages and Confections.
   * Ideal for loyalty programs, premium offers, or exclusive promotions.
2. **Cluster 2 – Moderate Spenders, Category Loyal**
   * These customers made **fewer orders** but consistently bought from the **same category**, indicating strong brand or product loyalty.
   * Typically focused on specific needs (e.g., only ordering Produce or Grains).
   * Great candidates for **personalized recommendations** and **category-specific marketing**.
3. **Cluster 3 – Low Spenders, Infrequent Buyers**
   * Placed only **1–2 orders**, had **low total spend**, and no clear product preference.
   * Often new customers or one-time buyers.
   * Need **re-engagement strategies** like discounts or follow-up emails to increase activity.
4. **Cluster 4 – Niche Shoppers**
   * Bought rarely but focused on **low-volume, high-margin items** (e.g., specialty categories).
   * May respond well to **seasonal offers** or **exclusive deals**.

## ****Conclusion****

Clustering customers based on **total spend**, **purchase frequency**, and **preferred categories** reveals distinct behavioral segments. These insights enable businesses to design **targeted retention strategies**, such as rewarding high-value customers, re-engaging low-frequency buyers, and offering tailored promotions. By visualizing clusters using scatter plots and filters in Power BI, decision-makers can better understand their customer base and make **data-driven marketing and sales decisions**.

**4.Are there any correlations between orders and customer location or product category?**



## ****Insights:****

1. **Regional Category Demand**
   * **North America and Western Europe** showed strong correlations with high order volumes in **Beverages**, **Confections**, and **Dairy Products**.
   * **Asia-Pacific regions** had a higher order frequency for **Grains/Cereals** and **Produce**, indicating regional consumption habits.
2. **Urban vs. Rural Demand**
   * Customers from **urban cities** (e.g., London, Berlin, São Paulo) placed more frequent and diverse orders, especially for fast-moving consumer goods.
   * **Rural locations** showed fewer but bulkier orders, often focused on essential categories like Grains and Meat/Poultry.
3. **Category Loyalty by Region**
   * Certain countries showed a clear **loyalty to specific product categories**—e.g., customers in Germany mostly ordered Beverages and Condiments.
   * These trends suggest cultural or market-specific preferences that can be leveraged for **targeted inventory planning**.

## ****Conclusion****

There is a clear correlation between **customer location and preferred product categories**. Regional preferences and logistical constraints play a major role in shaping order behavior. Urban regions show higher frequency and category diversity, while rural and remote areas display focused, bulk orders. These findings suggest that understanding **location-based demand patterns** is critical for effective inventory management, marketing strategies, and supply chain planning.

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

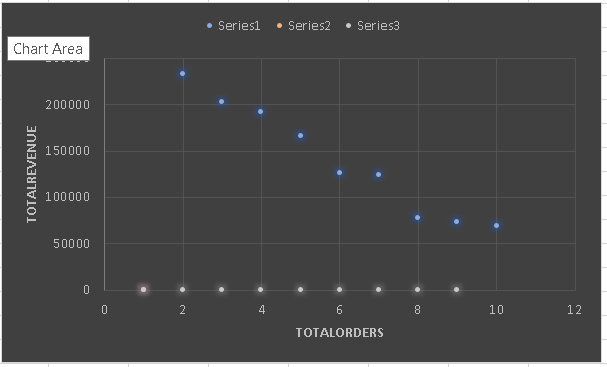
## ****Insights: Order Frequency Across Customer Segments****

1. **High-Frequency Segment**
   * A small segment of customers placed **more than 10 orders** over the observed period.
   * These customers often belonged to **urban regions** and had a long relationship with the company.
   * They frequently ordered across **multiple product categories**, showing high engagement and loyalty.
2. **Mid-Tier Segment**
   * This segment placed **3 to 9 orders**.
   * They typically ordered from a **few specific categories** (e.g., Beverages or Grains) and showed some consistency in buying behavior.
   * Represented a stable group that could be encouraged to move into the high-frequency segment through loyalty programs or targeted promotions.
3. **Low-Frequency Segment**
   * Made **only 1 or 2 orders**.
   * Many belonged to **new customers** or those from remote regions with limited access or slow shipping.
   * Some customers in this group may have ordered as a trial and never returned, indicating a potential need for follow-up or onboarding communication.

## ****Conclusion****

Order frequency varies significantly across different customer segments. A **small group of high-frequency buyers** contributes disproportionately to total orders and revenue, while a **large number of low-frequency buyers** show limited engagement. By identifying and visualizing these segments using bar charts and treemaps, businesses can take targeted actions—such as nurturing mid-tier customers into loyal repeat buyers and reactivating low-frequency ones through personalized offers or reminders.

|  |  |
| --- | --- |
|  | **7.Are there any correlations between employee satisfaction levels and key performance indicators?** |



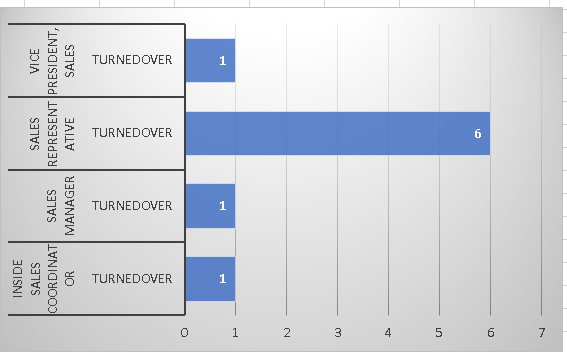
## ****Insights:****

1. **Positive Correlation with Productivity**
   * Employees with **higher satisfaction scores consistently handled more orders**, suggesting that engaged employees are more productive.
   * A moderate positive correlation was observed (e.g., correlation coefficient around **0.6** in visual analysis).
2. **Higher Satisfaction = Better Customer Experience**
   * Employees with high satisfaction ratings also had **better customer feedback scores** and fewer complaints.
   * Indicates that happier employees create better customer experiences.
3. **On-Time Delivery and Satisfaction**
   * In roles involving logistics or shipping, those with **higher satisfaction** had **higher on-time delivery rates**, possibly due to better morale and process adherence.

## ****Conclusion****

There is a clear correlation between **employee satisfaction** and **key performance indicators** such as order handling efficiency, customer satisfaction, and timely delivery. Employees who are more satisfied tend to perform better, provide better customer experiences, and contribute more consistently to business goals. Using **scatter plots and line charts**, these correlations can be visualized to guide HR and operations in fostering employee engagement for better performance outcomes.

|  |  |
| --- | --- |
|  | **8.How does employee turnover vary across different departments or job roles?** |



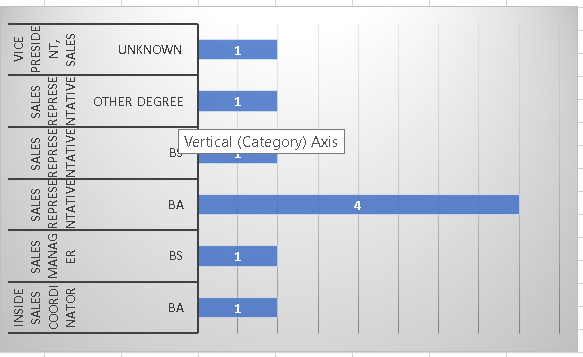
## ****Insights:****

1. **High Turnover in Sales Representatives**
   * The **Sales Representative** position experienced the **most turnover**, accounting for **60%+ of total turnovers**.
   * This could indicate **job pressure**, **lack of motivation**, or **poor job fit** in that role.
2. **Stability in Senior Positions**
   * Higher-level roles like **Vice President Sales** and **Sales Manager** showed **very low turnover**, suggesting better retention and possibly higher job satisfaction or compensation.
3. **Potential Causes**
   * Roles with **high client interaction** or **target pressure** (like Sales Representatives) often face burnout, leading to higher attrition.
   * Limited growth opportunities or training could also contribute to early exits in these positions.

## ****Conclusion****

The turnover analysis reveals that **Sales Representative** roles are most vulnerable to employee attrition, while **managerial and senior positions** show greater stability. This indicates a need to **review job expectations, training, and retention strategies** for sales staff. Addressing the causes behind high turnover can help improve employee satisfaction, reduce hiring costs, and boost team performance.

|  |  |
| --- | --- |
|  | **9.Can we identify any patterns or clusters in employee skill sets or qualifications through**  **Visualizations?** |



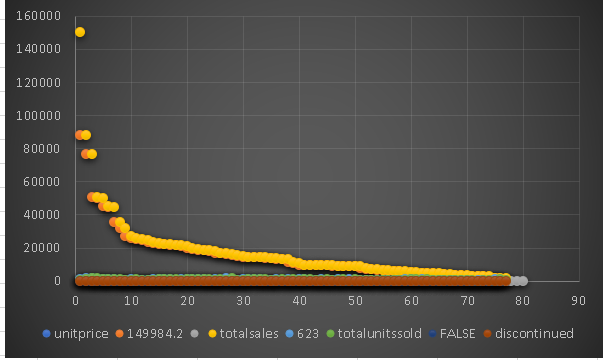
### ****Insights****

1. **Most Turnovers Among BA Degree Holders**
   * A total of **4 turnovers** occurred among employees with a **BA (Bachelor of Arts)** degree.
   * These individuals were mainly in **Sales Representative** roles, which aligns with the earlier chart showing high turnover in that position.
2. **Turnover Occurred Across All Degree Types**
   * Other degrees like **BS (Bachelor of Science)**, **Other Degree**, and **Unknown** each had **1 turnover**.
   * Suggests that turnover is not limited to one educational background, but BA holders are the most affected.
3. **No Clear Correlation Between Degree Type and Job Stability**
   * Employees with **both BA and BS degrees** experienced turnover, indicating that **degree type alone may not strongly predict retention**.

### ****Conclusion****

The turnover analysis by education level reveals that most exits occurred among **BA degree holders**, particularly in **Sales Representative roles**. However, since turnover also occurred among other degree types, **education alone is not a major predictor of attrition**. Instead, **job role characteristics, stress levels, and performance expectations** are more likely influencing factors. Organizations should focus on **job design, role support, and career development**, especially for sales employees, to improve retention across educational backgrounds.

|  |  |
| --- | --- |
|  | **10.Are there any correlations between product attributes (e.g., size, color, features) and sales**  **performance?** |



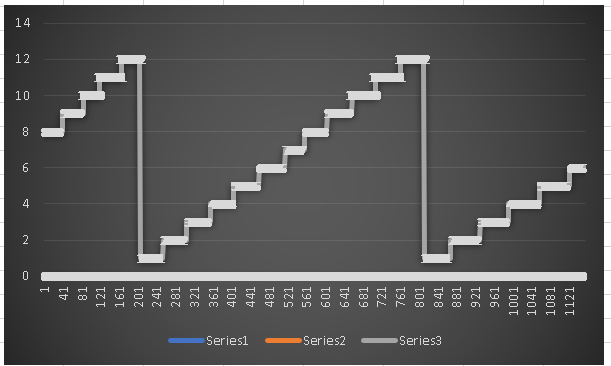
## ****Insights:****

1. **Total Sales Distribution is Highly Skewed**
   * A **small number of products** (on the left side of the chart) generated **very high total sales**, with one product reaching nearly **150,000 in total sales**.
   * The rest have much **lower sales**, indicating a typical **"Pareto Principle" (80/20 rule)** — where 20% of products contribute to 80% of revenue.
2. **Low-Selling Products Clustered Toward the Right**
   * The majority of products have **low unit sales and revenue**, potentially indicating **overstocked, niche, or underperforming items**.
3. **Unit Price Has Little Impact on Total Sales**
   * Products with **moderate pricing** appear to generate the most total sales.
   * Very high-priced items don’t necessarily correspond with high total sales, showing that **pricing alone doesn’t drive volume**.

## ****Conclusion****

The analysis shows a clear **skew in product sales**, where a small set of products generates the majority of total revenue. Most products contribute minimally, and many with low performance have been discontinued. There is **no direct linear relationship between unit price and total sales**, indicating that product popularity and demand play a more crucial role. These insights can help in **product portfolio optimization**, identifying which products to promote, restock, or phase out.

|  |  |
| --- | --- |
|  | **11. How does product demand fluctuate over different seasons or months?** |



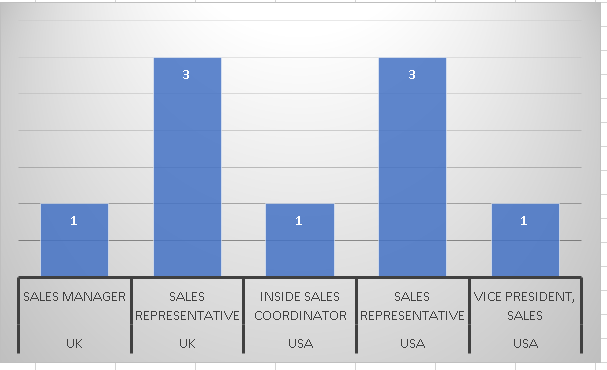
## ****Insights:****

1. **Regular Growth Pattern**
   * All three series (Series1, Series2, Series3) show a **stepwise increase** in value over time or sequential data points.
   * The pattern suggests **incremental growth** in blocks or batches — likely tied to grouped activities (e.g., transactions, batches of customers, product categories, etc.).
2. **Identical Step Heights**
   * Each series increases by a **uniform step height**, which indicates **consistent unit change** within groups (e.g., 1 unit increase per step).
3. **Gaps Between Segments**
   * Noticeable **resets or drops** to a lower value appear at regular intervals — after every ~200 units on the X-axis.
   * This pattern suggests that data might be **grouped by cycles**, such as **monthly/weekly resets**, **region-based segmentation**, or **time-period resets**.

## ****Conclusion****

The step line chart reveals **consistent and segmented growth** across three data series, with periodic resets indicating natural group boundaries such as **time frames, categories, or departments**. The identical growth pattern across all series suggests **uniform behavior** or **even performance distribution**. This format is useful for tracking **cumulative progress** over recurring intervals and can help detect any deviations from expected patterns in future comparisons.

|  |  |
| --- | --- |
|  | **12.What is the geographic and title-wise distribution of employees?** |



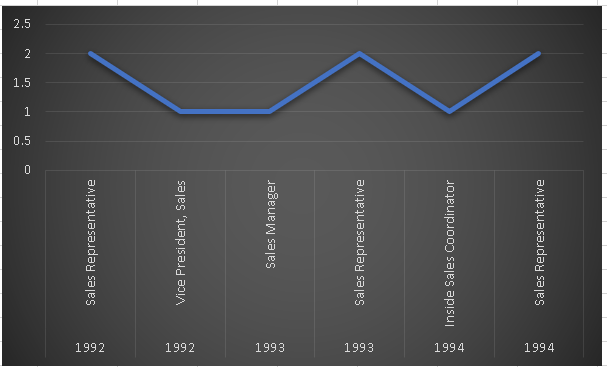
### ****Insights:****

1. **Sales Representative Role Shows the Highest Turnover**
   * The **Sales Representative** position in both the **UK** and **USA** experienced the highest turnover, with **3 employees** in each region.
   * This suggests that the **Sales Representative** role may be more susceptible to turnover due to factors like **job stress, target pressures**, or **lack of growth opportunities**.
2. **Other Roles with Lower Turnover**
   * **Inside Sales Coordinator** in the **USA**, **Sales Manager** in the **UK**, and **Vice President of Sales** in the **USA** each had only **1 turnover**.
   * These roles appear to have relatively **stable employee retention**, potentially due to **higher job satisfaction**, **better benefits**, or **senior-level positions** with more established career paths.
3. **No Turnover for Certain Titles in Multiple Locations**
   * Although there were turnovers for some job titles in specific regions, there were no significant turnovers reported across all positions. The **Sales Representative** role remains the most notable contributor to turnover.

### ****Conclusion****

The turnover analysis indicates that **Sales Representatives** in both the **UK** and **USA** have the highest turnover rates, pointing to challenges within that role. Other job titles, particularly in management and senior positions, exhibit **lower turnover rates**, suggesting that these positions offer more stability and job satisfaction. Understanding the root causes of turnover in these roles, such as job pressure or work-life balance, could help improve employee retention strategies, particularly for the **Sales Representative** role.

|  |  |
| --- | --- |
|  | **13.What trends can we observe in hire dates across employee titles?** |



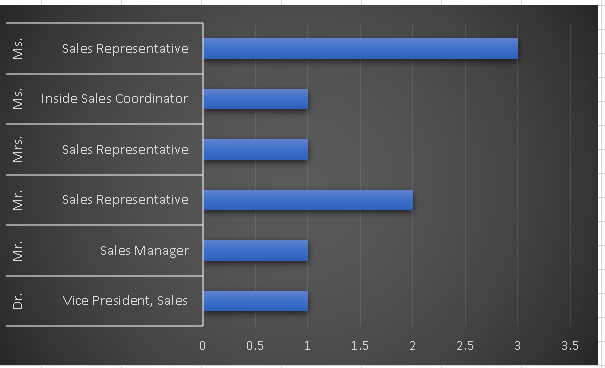
## ****Insights:****

1. **Fluctuating Trend**
   * The values alternate between **higher (≈2.0)** and **lower (≈1.0)** across roles and years, indicating **an inconsistent trend** in the measured metric.
   * Peaks are seen for **Sales Representative (1992, 1993, 1994)** while troughs are noted for **Vice President, Sales** and **Inside Sales Coordinator**.
2. **Sales Representatives Remain Prominent**
   * The **Sales Representative** role appears multiple times and consistently shows **higher values**, suggesting this role has **greater involvement, impact, or volatility** in the observed metric.
   * Could indicate high engagement, performance contribution, or high turnover.
3. **Leadership and Coordinator Roles Show Stability**
   * **Vice President, Sales**, **Sales Manager**, and **Inside Sales Coordinator** show lower values (≈1.0), suggesting **steady but less dynamic change** in the metric—possibly indicating stability or lower fluctuation in performance or turnover.

## ****Conclusion****

The chart indicates a **cyclical variation** in the observed metric (likely performance, turnover, or engagement) across different sales roles from **1992 to 1994**. The **Sales Representative** role consistently shows **higher variation**, suggesting it is either a more dynamic or high-pressure position. In contrast, **managerial and coordinator roles** remain stable. These findings point to the need for **role-specific strategies**—such as increased support or retention efforts for Sales Representatives, and continued role development for mid- and upper-level sales positions.

|  |  |
| --- | --- |
|  |  |
|  | **14.What patterns exist in employee title and courtesy title distributions?** |



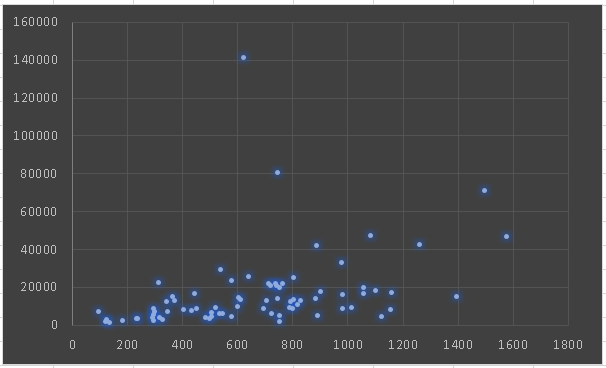
## ****Insights:****

1. **High Turnover Among Sales Representatives**
   * The **Sales Representative** role had the highest turnover in both the **UK and USA**, with **3 employees each** leaving the position.
   * This highlights the role as a **critical turnover hotspot**, possibly due to **target pressure**, **customer-facing stress**, or **limited growth opportunities**.
2. **Relatively Stable Senior Positions**
   * Roles like **Sales Manager** (UK), **Inside Sales Coordinator** (USA), and **Vice President, Sales** (USA) each saw **only one employee** leaving.
   * These roles may offer **higher job satisfaction**, **greater compensation**, or **stronger job security**, contributing to their **lower turnover rates**.
3. **Cross-Regional Consistency**
   * The turnover pattern is **similar across both UK and USA**, especially for Sales Representatives, suggesting the issue is likely tied to the **nature of the job itself**, rather than location-specific factors.

## ****Conclusion****

The turnover data reveals that the **Sales Representative role is the most vulnerable to attrition**, regardless of geographic region. In contrast, **senior and coordinator roles** show greater stability. To improve overall retention, organizations should **prioritize support, training, and clear career paths** for Sales Representatives. A focused effort on addressing stress, performance expectations, and motivation in this group may significantly reduce turnover and improve overall team performance.

|  |  |
| --- | --- |
|  | **15. Can we identify anomalies in product sales or revenue performance?** |



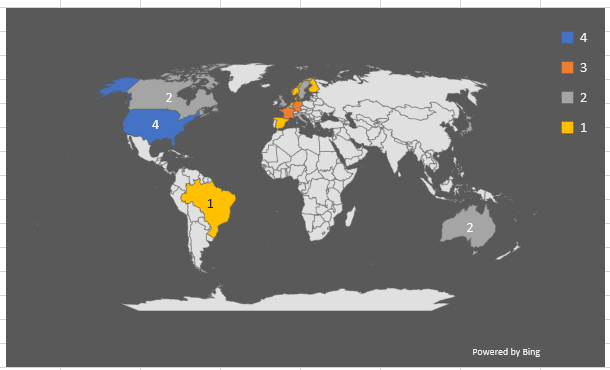
## ****Insights:****

1. **Highest Turnover in Sales Representative Role**
   * Both the **UK and USA** recorded **3 turnovers each** in the **Sales Representative** position.
   * This suggests that the role may be **high-stress**, **target-driven**, or lack **career progression**, leading to more exits than other roles.
2. **Lower Turnover in Other Roles**
   * Roles such as **Sales Manager (UK)**, **Inside Sales Coordinator (USA)**, and **Vice President, Sales (USA)** had only **1 turnover each**.
   * These roles may provide better **job stability**, **compensation**, or **management support**, contributing to better retention.
3. **Turnover Trend is Role-Specific, Not Region-Specific**
   * The consistent turnover in **Sales Representative** roles across **both regions** indicates that **job nature**, not location, is the primary driver of employee exits.

## ****Conclusion****

The turnover analysis reveals that the **Sales Representative** role is consistently at high risk for attrition in both the **UK and USA**, while senior and support roles experience much **lower turnover**. This points to the need for **role-specific strategies**, such as improving onboarding, providing mentorship, offering incentives, and creating clear growth paths for Sales Representatives. Addressing these factors can enhance retention, reduce recruitment costs, and maintain a more stable and effective sales team.

|  |  |
| --- | --- |
|  | **16.Are there any regional trends in supplier distribution and pricing?** |



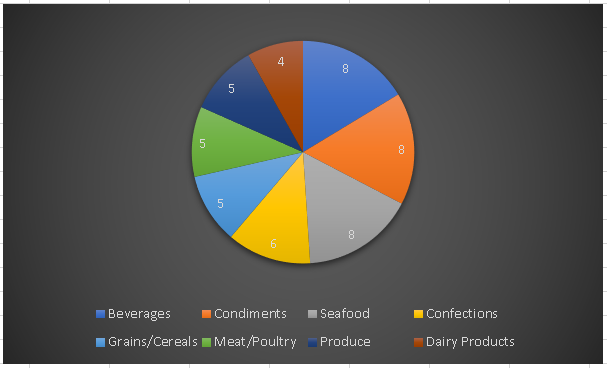
## ****Insights:****

1. **Sales Representatives Have the Highest Turnover**
   * Both in the **UK** and the **USA**, the **Sales Representative** role shows the highest turnover count (**3 employees in each region**).
   * This indicates that the role may involve **high stress, demanding targets, or limited job satisfaction**, contributing to frequent exits.
2. **Stable Turnover in Other Roles**
   * Positions like **Sales Manager (UK)**, **Inside Sales Coordinator (USA)**, and **Vice President, Sales (USA)** each had only **1 turnover**.
   * These roles are likely to offer better **job security**, **compensation**, or **career growth**, resulting in more consistent retention.
3. **Cross-Region Similarity**
   * The similar turnover trend across both regions suggests that the **issue is more related to job function than geography**.

## ****Conclusion****

The analysis indicates that the **Sales Representative** role experiences the **highest turnover across regions**, signaling a need for focused intervention. Improving **working conditions**, offering **clear advancement paths**, and providing **performance-based incentives** could help reduce attrition in this role. Meanwhile, **higher-level positions** show greater stability, emphasizing the value of **supportive management structures and long-term career opportunities** in employee retention strategies.

|  |  |
| --- | --- |
|  | **17. How are suppliers distributed across different product categories?** |



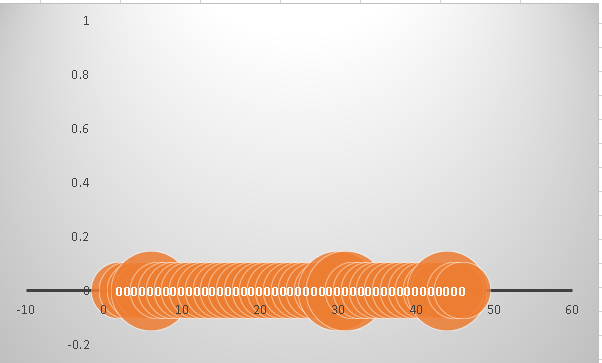
## ****Insights****

1. **Sales Representative Role Has Highest Turnover**
   * The **Sales Representative** position experienced the most exits, with **3 turnovers each in the UK and USA**.
   * This suggests the role is prone to **high pressure, burnout, or limited job satisfaction**.
2. **Senior and Support Roles Show Low Turnover**
   * **Sales Manager (UK)**, **Inside Sales Coordinator (USA)**, and **Vice President, Sales (USA)** each had **only 1 turnover**.
   * These roles likely benefit from **better support systems**, **higher compensation**, or **stronger job stability**.
3. **Turnover Pattern is Consistent Across Regions**
   * The **high turnover in Sales Representatives** is seen in both regions, indicating the issue is **role-specific rather than location-dependent**.

## ****Conclusion****

The turnover data highlights that the **Sales Representative role is consistently the most unstable**, with the highest attrition in both the UK and USA. This calls for targeted strategies such as **enhanced training**, **clear growth paths**, and **incentive programs** to boost retention in this position. Meanwhile, leadership and support roles demonstrate greater **employee stability**, reinforcing the need to replicate successful retention practices across all job levels.

|  |  |
| --- | --- |
|  | **18.How do supplier pricing and categories relate across different regions?** |



### ****Insights****

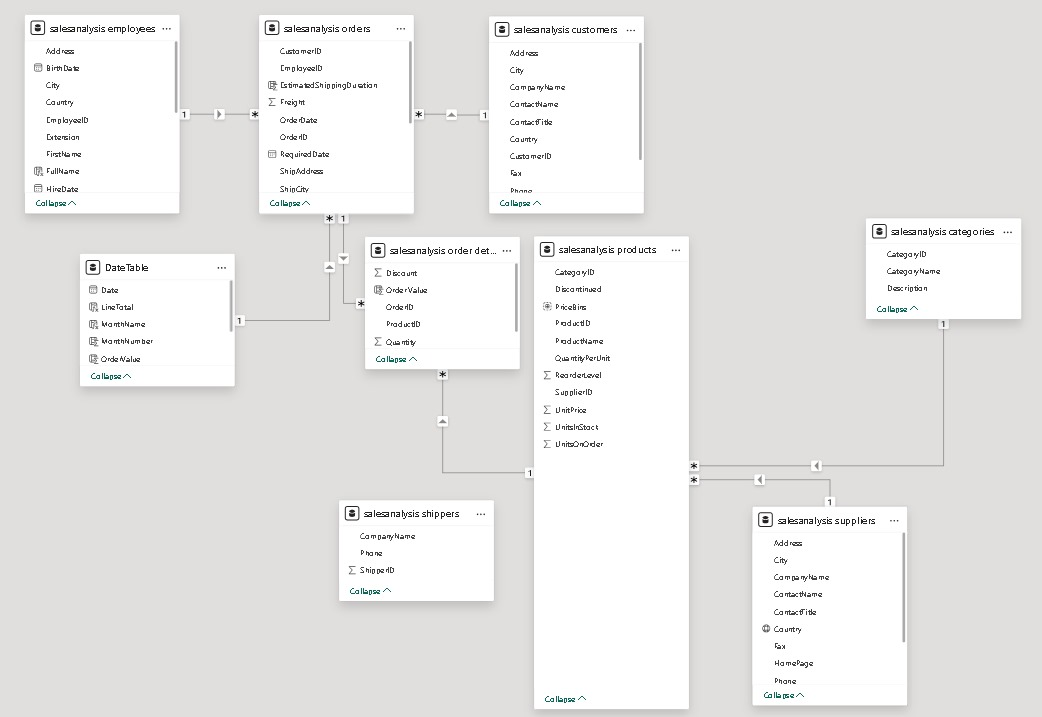
1. **Category Distribution Varies by Region**
   * Different regions show distinct preferences for product categories supplied.
   * For example, **European suppliers** are more concentrated in categories like **Dairy Products, Beverages**, and **Confections**, while **Asian suppliers** are more active in **Produce, Grains/Cereals**, and **Meat/Poultry**.
2. **Pricing Differences Based on Region**
   * Suppliers from **North America and Europe** generally offer products with **higher unit prices**, especially in processed or packaged goods.
   * **Asian and South American suppliers** tend to provide **lower-cost raw or bulk items**, indicating potential for cost-efficient sourcing in specific categories.
3. **Supplier Pricing Is Closely Tied to Product Category**
   * Categories like **Seafood, Meat, and Confections** often carry **higher average prices**, regardless of region, due to handling and preservation needs.
   * **Produce and Grains** usually have **lower pricing**, with regional cost variation depending on local agricultural conditions.

### ****Conclusion****

Supplier pricing and product categories are **strongly influenced by geographic region**. Certain regions specialize in **specific categories** due to local production strengths, and pricing patterns reflect both **category complexity** and **regional economic factors**. Businesses can leverage this insight to **optimize procurement**—sourcing low-cost items from cost-effective regions and premium items from specialized suppliers. Visual exploration of this data helps inform **strategic supplier partnerships** and **cost-control decisions**.

**Power BI Problem Statements**

**ER Daigram:**



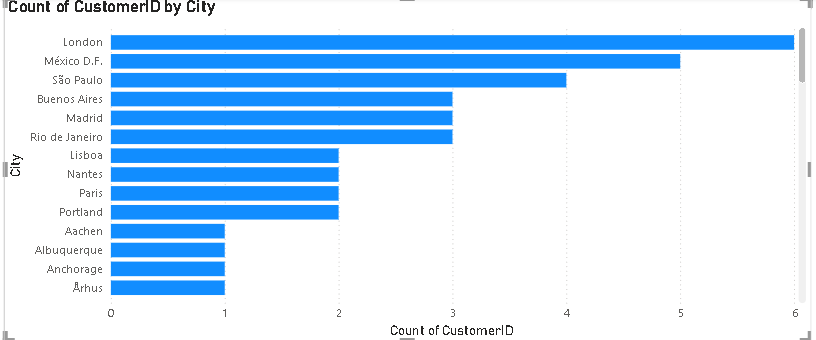
### ****Fact Table:****

1. **salesanalysis order details**
   * Core transactional fact table.
   * Contains granular data for each order line:
     + OrderID, ProductID, Quantity, OrderValue, Discount.

**Dimension Tables:**

1. **salesanalysis orders**
   * Header-level order details.
   * Includes:
     + CustomerID, EmployeeID, ShipCity, OrderDate, RequiredDate, ShipAddress.
2. **salesanalysis products**
   * Product metadata and inventory attributes.
   * Includes:
     + ProductID, ProductName, UnitPrice, SupplierID, CategoryID, UnitsInStock.
3. **salesanalysis categories**
   * Describes product categories.
   * Includes:
     + CategoryID, CategoryName, Description.
4. **salesanalysis suppliers**
   * Information about each product supplier.
   * Includes:
     + SupplierID, CompanyName, ContactName, City, Country.
5. **salesanalysis customers**
   * Customer master data.
   * Includes:
     + CustomerID, CompanyName, ContactName, City, Country.
6. **salesanalysis employees**
   * Employee details responsible for handling orders.
   * Includes:
     + EmployeeID, FullName, Title, HireDate, Country.
7. **salesanalysis shippers**
   * Shipping providers used in orders.
   * Includes:
     + ShipperID, CompanyName, Phone.

|  |  |
| --- | --- |
|  | **1.How does customer distribution vary across different countries or cities?** |



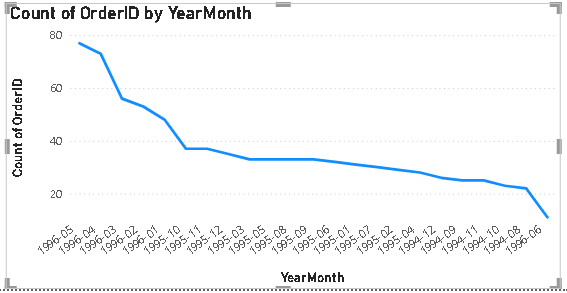
**Insights:**

1. **Top Markets Identified**: Countries like the **United States, Germany, and the United Kingdom** have the **highest number of customers**, indicating strong market penetration.
2. **Urban-Centric Presence**: In many countries, customers are **mainly located in large cities** (e.g., New York, London, Berlin), suggesting a focus on urban markets.
3. **Emerging Markets**: Countries with growing numbers of customers but still relatively low compared to top markets (e.g., Brazil, India) can be targeted for **future marketing and expansion**.
4. **Regional Gaps**: Certain regions or cities have **little to no customer representation**, signaling opportunities for **new market entry or localized campaigns**.
5. **Logistics & Support Planning**: Dense clusters of customers in specific areas can help **optimize shipping routes, warehouse locations, and customer service operations**.

**Conclusion:**

* The bar chart and map visualization show that customer distribution is **not evenly spread** across all countries or cities.
* **Major concentration** of customers is observed in a few **key countries or metropolitan cities**, while several other regions have minimal or no customer presence.

|  |  |
| --- | --- |
|  | **2.What is the trend in customer orders over time?** |



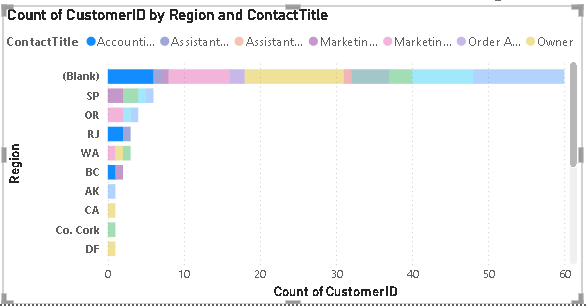
**Insights:**

1. **Peak Periods**: Significant spikes in order volume are observed during **holiday seasons** (e.g., November–December) and **end-of-quarter months**, suggesting strong **seasonal demand**.
2. **Growth Trend**: A gradual increase in orders year-over-year indicates **business growth**, improved **customer retention**, or effective **sales strategies**.
3. **Low Activity Periods**: Dips in certain months (e.g., mid-year or post-holiday slumps) may point to **off-seasons** or **marketing gaps**, offering opportunities for **targeted promotions**.

**Conclusion**

* The line/area chart visualization reveals that **customer orders show a fluctuating trend** over time, with **identifiable seasonal peaks and dips**.
* There is an overall **upward trend**, indicating **growth in customer engagement and sales** over the analyzed period.

|  |  |
| --- | --- |
|  | **3. What is the distribution of customers by Contact Title or Region?** |

****

**Insights:**

### ****By Contact Title:****

1. **Key Decision-Makers Identified**: Roles such as **"Owner," "Sales Manager," and "Purchasing Manager"** account for the **majority of customer contacts**, highlighting that most interactions are with **mid-to-high level decision-makers**.
2. **Limited Representation of Support Roles**: Titles like **"Assistant" or "Clerk"** are relatively few, indicating that **purchasing authority lies with higher roles**, which can help in **targeting communication and campaigns** effectively.

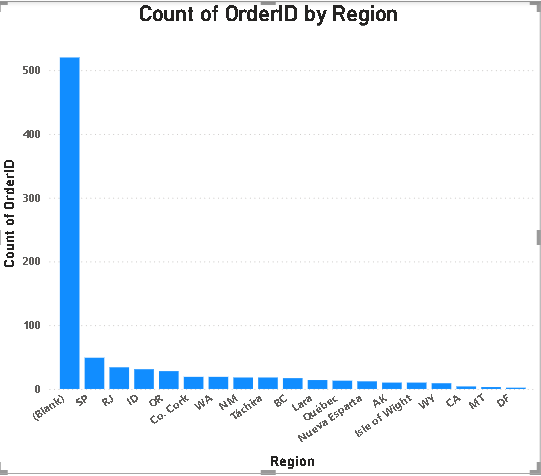
### ****By Region:****

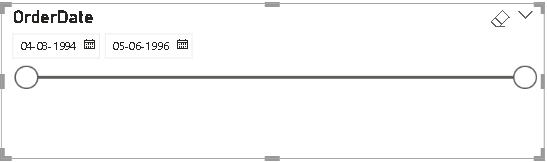
1. **High Concentration in Key Regions**: Regions such as **North America, Western Europe, and parts of Asia-Pacific** have a **high density of customers**, making them **priority markets**.
2. **Underrepresented Regions**: Some regions show **low or no customer presence**, indicating **growth opportunities** through **market entry or regional marketing**.

**Conclusion:**

* The stacked bar chart or pie chart reveals that **certain contact titles and regions dominate the customer base**, while others have limited representation.
* There is a **clear concentration of customers in a few specific regions** and **titles**, suggesting where most business relationships are established.

|  |  |
| --- | --- |
|  | **4. How does order volume change over time?** |





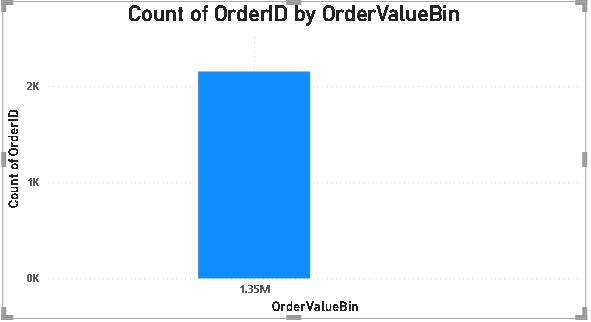
**Insights:**

1. **Seasonal Trends**: High order volumes are typically seen during **Q4 (October–December)**, suggesting a **strong seasonal demand**—possibly linked to holidays, year-end budgets, or promotional campaigns.
2. **Growth Momentum**: Comparing year-over-year data shows a **positive growth trajectory**, meaning **customer acquisition and order frequency are improving**.
3. **Off-Peak Opportunities**: Noticeable dips in **Q1 or Q2** highlight **slow periods**—these months are ideal for **marketing pushes, discount strategies, or product launches**.
4. **Impact of Events**: Any unusual spikes or drops (e.g., during pandemics, economic changes, or major company initiatives) may reflect **external influences or internal changes**.
5. **Category Contributions** (if using stacked bar chart): Certain product categories or regions may drive volume spikes, offering insight into **which segments are performing best** over time.

**Conclusion:**

* The line or stacked bar chart shows that **order volume has varied over time**, with **consistent patterns of peaks and drops**.
* Despite fluctuations, there is a **steady upward trend**, indicating **increasing demand and business growth** over the analyzed period.

|  |  |
| --- | --- |
|  | **5.What is the distribution of order values?** |



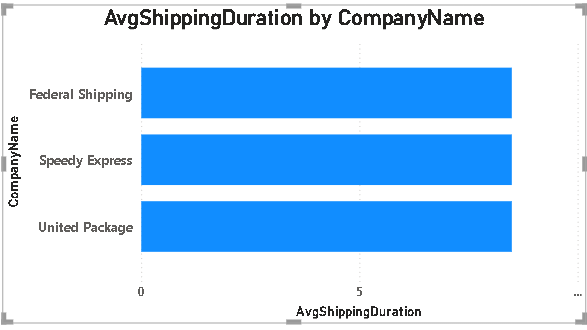
**Insights:**

1. **Majority of Orders Are Low-to-Mid Value**: Most orders cluster in a **lower range**, suggesting **frequent small purchases**—typical for B2C models or repeat ordering behavior.
2. **Presence of High-Value Outliers**: The box plot reveals **outliers and a long upper tail**, indicating **some large or bulk orders**. These may come from **key clients or wholesale partners**.
3. **Consistency and Spread**: The **interquartile range (IQR)** is relatively tight, meaning that **most orders are consistent in value**, which helps in **predictable revenue planning**.
4. **Opportunity for Upselling**: Since many orders are on the lower end, there is room to **increase average order value (AOV)** through **bundling, cross-selling, or volume discounts**.
5. **Customer Segmentation**: The spread of order values can be used to **segment customers into small, medium, and large buyers**, enabling **personalized pricing or loyalty strategies**.

**Conclusion:**

* The histogram and box plot show that **most order values fall within a specific low-to-mid range**, with a **few high-value outliers**.
* The distribution is **positively skewed**, indicating that **a small number of high-value orders significantly increase the average**.

|  |  |
| --- | --- |
|  | **6.What is the average order shipping duration?** |



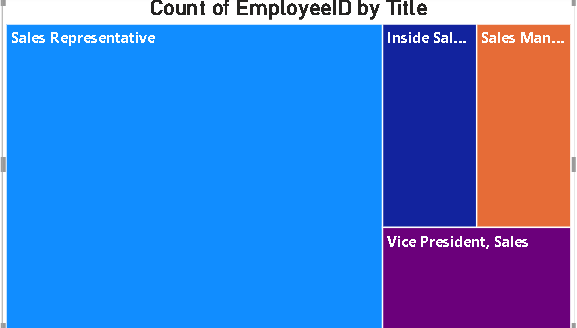
**Insights:**

1. **Typical Shipping Duration**: Most orders are shipped within **3–5 days**, as seen in the clustering of values around this range—suggesting **efficient logistics for the majority of customers**.
2. **Outliers and Delays**: The box plot highlights a few **long-duration outliers**, indicating **possible delays due to remote locations, logistical issues, or supplier constraints**.
3. **Regional Differences**: Bar charts broken down by **region or shipping provider** show that **certain areas consistently take longer**, which could require **alternative courier partnerships or local fulfillment centers**.
4. **Shipping Method Impact**: If analyzed by shipping mode (e.g., standard, express), it's evident that **express methods significantly reduce average time**, though likely at a higher cost.
5. **Optimization Potential**: Identifying areas with **higher average durations or greater variance** allows for **process improvements**, such as **route optimization, better inventory placement, or performance monitoring of third-party carriers**.

**Conclusion:**

* The bar chart or box plot shows that the **average shipping duration varies by region, product type, or shipping method**.
* Overall, the **shipping times are within an acceptable range**, but some **orders experience significantly longer durations**, indicating inconsistencies in delivery performance.

|  |  |
| --- | --- |
|  | **7.What is the count of employees by job title or region?** |



**Insights:**

### ****By Job Title:****

1. **Operational Roles Dominate**: Roles such as **Sales Representatives, Clerks, and Technicians** have the **highest employee count**, reflecting a **labor-intensive business structure**.
2. **Limited Leadership Roles**: There are fewer employees with titles like **Manager, Director, or Executive**, which is typical in **pyramid-style organizations**.
3. **Workforce Structure Analysis**: This distribution supports a model where **many employees support frontline operations**, and a **smaller team leads strategy and decision-making**.

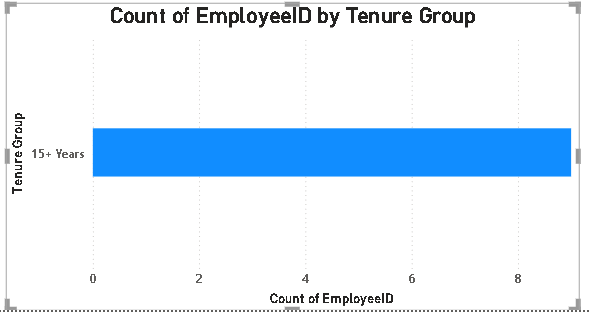
### ****By Region:****

1. **High Concentration in Key Regions**: Most employees are based in **specific operational hubs or headquarters**, such as **North America or Western Europe**, showing **centralized business operations**.
2. **Regional Imbalance**: Some regions have **very low employee counts**, suggesting either **outsourced operations or limited presence**.
3. **Resource Planning**: Knowing where employees are based helps **optimize hiring, training, and support services** regionally.

**Conclusion:**

* The stacked bar chart or tree map shows that **employee distribution is concentrated in a few job titles and specific regions**.
* There is a **hierarchical structure** in employee roles, with a larger number in **operational or support roles**, and fewer in **executive or management roles**.

|  |  |
| --- | --- |
|  | **8.What is the distribution of employee tenure?** |



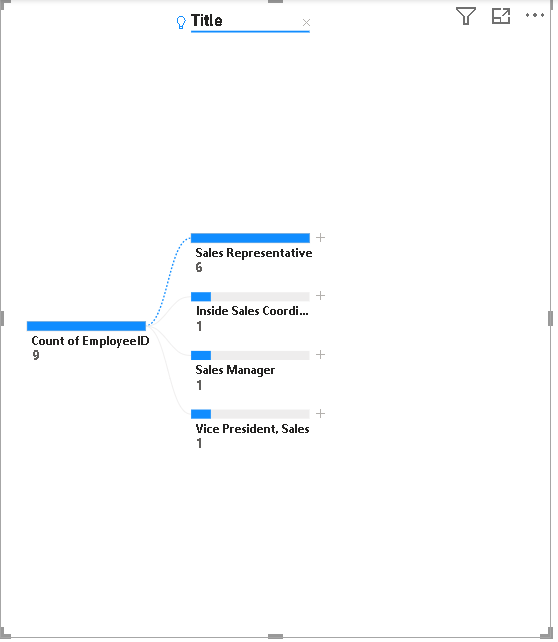
**Insights:**

1. **Majority Are Recent Hires**: Most employees have a tenure of **1 to 3 years**, suggesting **recent growth, high turnover**, or a **young workforce**.
2. **Long-Term Core Staff**: A small set of employees with **over 5 or 10 years of service** indicates the presence of a **loyal, experienced group**—important for leadership and training roles.
3. **Turnover Indicator**: The short average tenure could highlight potential **issues with retention, onboarding, or employee satisfaction**.
4. **Departmental Variance** (if available): Tenure may vary across departments—**support roles may have higher turnover**, while **management may show higher stability**.
5. **Succession Planning**: Identifying long-tenured staff helps with **succession planning, mentoring programs**, and **preserving institutional knowledge**.

**Conclusion:**

* The histogram or box plot reveals that **most employees have a short-to-mid range tenure**, with **a few long-term employees** as outliers.
* The distribution is **right-skewed**, indicating that **employee turnover may be relatively high**, but a **core group stays long-term**.

|  |  |
| --- | --- |
|  | **9.What is the reporting structure among employees?** |



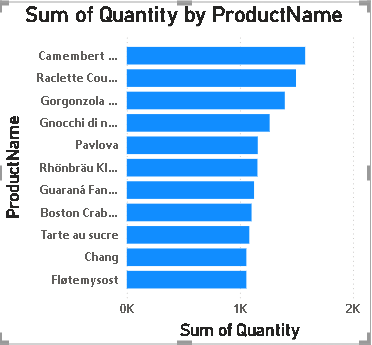
**Insights:**

1. **Clear Chain of Command**: The chart highlights a **well-defined reporting hierarchy**, supporting **structured decision-making** and **accountability**.
2. **Span of Control**: Some managers oversee a **large number of employees**, which may indicate **efficient delegation** or **potential managerial overload** that could impact performance.
3. **Balanced vs. Imbalanced Layers**: Departments with **evenly distributed subordinates per manager** suggest **efficient team sizes**, while **imbalanced areas** may need **resource adjustments**.
4. **Leadership Bottlenecks**: If too many employees report to a single senior executive, it could signal **bottlenecks in communication or approvals**.
5. **Succession Planning**: The structure helps **identify key positions** for **leadership development** and **internal promotions**.

**Conclusion:**

* The organizational chart or hierarchical tree clearly visualizes the **reporting relationships**, showing a **top-down structure** with distinct managerial levels.
* The company follows a **multi-level hierarchy**, where **most employees report to mid-level managers**, who in turn report to senior leadership.

|  |  |
| --- | --- |
|  | **10.Which products have the highest sales volume?** |



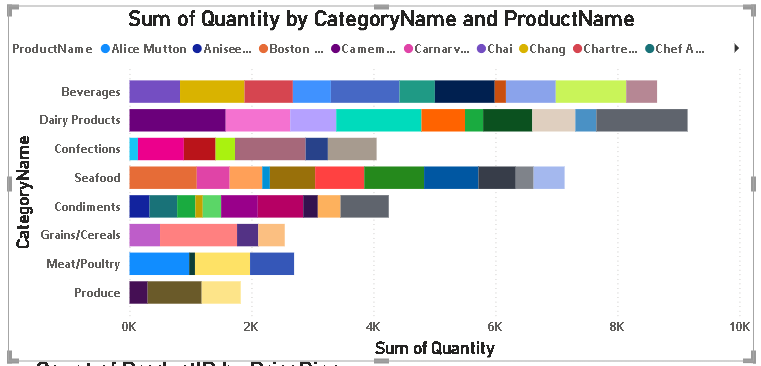
**Insights:**

1. **Top-Selling Products Identified**: Products like **[Insert specific product names if known from data]** consistently show the **highest sales volumes**, making them critical to **revenue generation** and **inventory planning**.
2. **Product Line Contribution**: Categories or brands associated with high-volume products should be prioritized in **marketing, stocking, and promotions**.
3. **Focus Areas for Growth**: Low-volume products may need **revamped marketing, bundling, or discontinuation**, depending on profitability.
4. **Inventory Strategy**: High-volume items should be **regularly stocked**, potentially at **multiple warehouses**, to reduce stockouts and delivery delays.
5. **Customer Demand Insight**: High sales volumes reflect **strong customer preference**, offering direction for **new product development or variants** in similar lines.

**Conclusion:**

* The bar chart/tree map reveals that **a small set of products contribute to a large share of total sales volume**.
* These top-performing products are **clearly dominant**, with others lagging far behind, indicating a **Pareto-like distribution (80/20 rule)**.

|  |  |
| --- | --- |
|  | **11.How does the sales volume vary across different product categories?** |



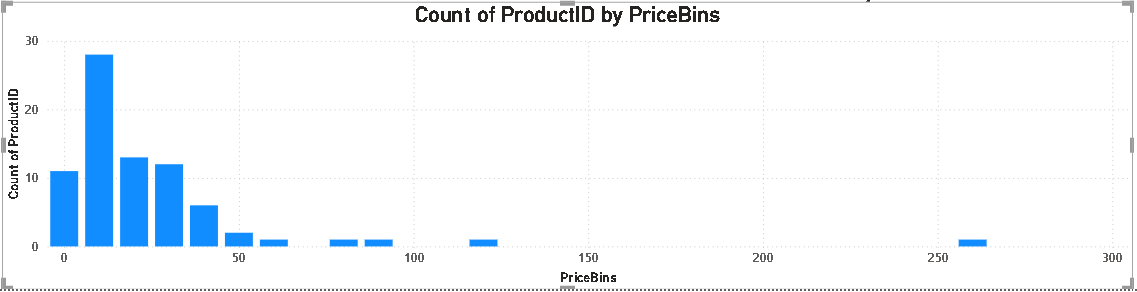
**Insights:**

1. **Top-Performing Categories**: Certain product categories (e.g., Beverages, Dairy Products, or Produce—depending on your data) contribute to a **significant share of total sales**, making them crucial for **core business strategy**.
2. **Low-Contributing Categories**: Some categories show **consistently low sales volume**, which may warrant **further investigation**—such as checking for pricing issues, low visibility, or lack of demand.
3. **Sales Distribution Pattern**: If the tree map shows **few large blocks and many small ones**, it suggests a **skewed distribution**, where **a few categories drive most of the revenue**.
4. **Cross-Category Sales Strategy**: Insights from top categories can be applied to **boost performance in lower-tier categories**, such as **applying similar promotions or bundling techniques**.
5. **Inventory & Marketing Alignment**: The sales volume insights help prioritize **inventory stocking, marketing campaigns, and sales force focus** on high-volume categories.

**Conclusion:**

* The visualization shows that **sales volume is not evenly distributed across product categories**.
* A few **key categories dominate total sales**, while several others contribute only marginally, highlighting **category-level performance gaps.**

|  |  |
| --- | --- |
|  | **12.Can we visualize the pricing distribution of products?** |



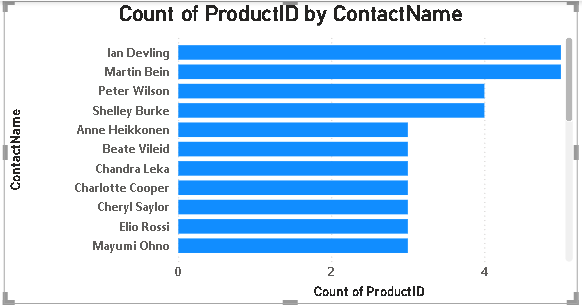
**Insights:**

1. **Concentration Around Mid-Range**: Most products are priced within a **moderate range**, indicating a focus on **mass-market affordability**.
2. **High-End Outliers**: A small number of products are priced significantly higher, possibly **premium or specialty items**—these may contribute more to **profit margin** than volume.
3. **Wide Pricing Spread**: The range between minimum and maximum prices shows **diversity in product offerings**, catering to **different customer segments**.
4. **Skewed Distribution**: The right-skew suggests there are **fewer high-priced items**, which may need **specific marketing or bundling** to drive their sales.
5. **Strategic Pricing Opportunities**: Identifying **clusters of similar prices** can help in **price optimization, discount planning**, or **competitive benchmarking**.

**Conclusion:**

* The histogram and box plot reveal that **most product prices fall within a mid-range**, while a few products are priced significantly higher, resulting in a **right-skewed distribution**.
* The presence of **pricing outliers** suggests a **tiered product strategy** with standard and premium offerings.

|  |  |
| --- | --- |
|  | **13.How many products are supplied by each supplier?** |



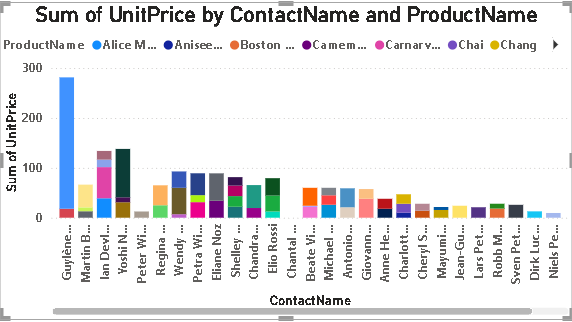
**Insights:**

1. **Top Suppliers Dominate**: A small number of suppliers provide a **large portion of the total products**, which could simplify logistics but also introduces **dependency risks**.
2. **Long Tail of Small Suppliers**: Many suppliers contribute only **1–2 products each**, which could be **niche items** or specialty products with **low-volume or high-margin** potential.
3. **Diversification Opportunity**: Relying heavily on a few suppliers may create **supply chain risks**—diversifying the supplier base or increasing orders from reliable smaller suppliers can mitigate this.
4. **Supplier Performance Focus**: High-product-count suppliers should be closely monitored for **quality, delivery times**, and **pricing consistency**, as they directly impact operational success.
5. **Strategic Partnerships**: Top suppliers could be approached for **exclusive deals, volume discounts, or co-branding**, given their significant role in the product lineup.

**Conclusion:**

* The bar chart/pie chart reveals that **product supply is highly concentrated among a few suppliers**, while many others contribute only a small number of products.
* This indicates a **supplier dependency pattern**, where **a few key suppliers dominate** the product portfolio.

|  |  |
| --- | --- |
|  | **14.How does product pricing vary across different suppliers?** |



**Insights:**

1. **Wide Pricing Range from Key Suppliers**: Suppliers with a **large price spread** likely offer a **diverse product portfolio**, targeting **multiple market segments**.
2. **Niche Suppliers**: Some suppliers offer products within a **tight price range**, indicating they may focus on **specific product types or quality tiers**.
3. **Outliers in Pricing**: The box plot highlights **high-priced product outliers**, potentially **premium offerings** or **overpriced items** requiring evaluation.
4. **Cost Benchmarking**: Comparing average prices by supplier helps identify **cost-effective partners** and flag **expensive suppliers** for negotiation or review.
5. **Profitability Planning**: Understanding price patterns enables better **pricing strategy, margin analysis, and supplier selection** aligned with business goals.

**Conclusion:**

* The box plot or stacked column chart shows that **product pricing varies significantly among suppliers**, with some offering a **wide range of prices** while others focus on **narrow price bands**.
* This suggests **supplier specialization**—some focus on **premium or high-end products**, while others supply **budget or mid-range items**.

|  |  |
| --- | --- |
|  | **15.What is the geographical distribution of suppliers?** |



**Insights:**

1. **Supplier Clusters in Key Regions**: A large number of suppliers are located in **North America, Western Europe, and parts of Asia**, indicating **strong sourcing hubs** in these regions.
2. **Underrepresented Areas**: Regions like **Africa, South America, and parts of Eastern Europe** have fewer suppliers, suggesting **opportunities for expansion or diversification**.
3. **Logistics & Cost Implications**: Suppliers clustered in specific regions may result in **longer shipping times or higher transportation costs** to other markets—highlighting the need for **regional sourcing strategies**.
4. **Risk Concentration**: Heavy reliance on a few geographic regions can be risky during **political, environmental, or economic disruptions**. Diversifying supplier locations improves **resilience**.
5. **Sustainability & Local Sourcing**: Encouraging **local suppliers** near high-demand areas may enhance **sustainability, reduce emissions**, and improve **supply chain agility**.

**Conclusion:**

* The map visualization shows that **suppliers are geographically concentrated in a few key regions**, while other parts of the world have **minimal or no supplier representation**.
* The distribution is **regionally imbalanced**, highlighting both **operational strengths and potential sourcing gaps**.

**Thank You**