CAPSTONE PROJECT NORTH WIND TRADERS

By sahil Bhardwaj

Abstract

The abstract for the "Capstone Project North Wind Traders" encompasses a comprehensive analysis of the sales data for the fictitious company "North Wind Traders." The project aims to provide insights into customer behavior, sales patterns, and employee performance to facilitate data-driven decision-making. The documentation covers various aspects, including the project overview, information about dataset tables, data cleaning, MECE-breakdown, Power BI and EDA questions, and dashboard analysis.

The project utilizes interactive visualizations and dynamic filters to explore customer distribution, customer acquisition trends, customer demographics, order volume, order values, employee productivity, employee tenure, product performance, supplier metrics, and geographical distribution of suppliers. The visualizations and analyses offer valuable insights into customer preferences, market segments, product performance, and supplier relationships, enabling stakeholders to make informed decisions and identify opportunities for improvement.

The abstract provides a glimpse into the detailed analysis and insights derived from the project, showcasing the use of Power BI and EDA to extract meaningful information from the dataset. The project's objective is to empower stakeholders with valuable insights and facilitate data exploration through interactive visualizations and dynamic filters, ultimately aiding in informed decision-making and operational improvements.

CONTENTS

INTRODUCTION

➢ **Project Overview:**

The North Wind Traders Capstone Project is a comprehensive analysis of the sales data for a fictitious company called "North Wind Traders," which imports and exports specialty foods from around the world. The project aims to generate insights into customer behaviour, sales patterns, and employee performance to aid decision-making processes. It covers sales analysis, customer segmentation, inventory trends, and employee performance, consolidating data from multiple tables for a comprehensive view of the company's operations. The objective of the project is to empower stakeholders to make data-driven decisions by offering valuable insights and facilitating data exploration through interactive visualizations and dynamic filters. The project utilizes Power BI and EDA questions to extract meaningful information from the dataset and offers sample visualizations and analyses to showcase the use of interactive visualizations and dynamic filters. The project's key functionality includes sales analysis, customer segmentation, inventory trends, employee performance analysis, product performance evaluation, supplier metrics visualization, and interactive visualizations and dynamic filters. The project's documentation includes a project overview, information about dataset tables, data cleaning and MECE-breakdown, Power BI and EDA questions, dashboard analysis, and conclusion.

➤ Introduction of data set Table:

The North Wind Traders Capstone Project includes several tables that store information about the company's operations.

The tables included in the project are:

Customers Table

This table stores information about the company's customers. It includes fields for customer ID, company name, contact name, contact title, address, city, region, postal code, country, phone, and fax.

Employees Table

This table stores information about the company's employees. It includes fields for employee ID, last name, first name, title, title of courtesy, birth date, hire date, address, city, region, postal code, country, home phone, extension, photo, notes, reports to, and photo path.

Orders Table

This table stores information about the company's orders. It includes fields for order ID, customer ID, employee ID, order date, required date, shipped date, ship via, freight, ship name, ship address, ship city, ship region, ship postal code, and ship country.

Order Details Table

This table stores detailed information about the items within each order. It includes fields for order ID, product ID, unit price, quantity, and discount.

Products Table

This table stores information about the company's products. It includes fields for product ID, product name, supplier ID, category ID, quantity per unit, unit price, units in stock, units on order, reorder level, and whether the product is discontinued.

Suppliers Table

This table stores information about the company's suppliers. It includes fields for supplier ID, company name, contact name, contact title, address, city, region, postal code, country, phone, fax, and home page.

Shippers Table

This table stores information about the company's shipping companies. It includes fields for shipper ID, company name, and phone.

Categories Table

This table stores information about the product categories. It includes fields for category ID, category name, and description.

These tables are consolidated to provide a comprehensive view of the company's operations and generate insights into customer behavior, sales patterns, and employee performance. The project utilizes interactive visualizations and dynamic filters to facilitate data exploration and offer valuable insights to stakeholders

➢ Reason For the Project:

The North Wind Traders Capstone Project aims to create a visually appealing and user-friendly dashboard that communicates key performance metrics effectively. The project's objective is to generate

insights into customer behavior, sales patterns, and employee performance to aid decision-making processes. It covers sales analysis, customer segmentation, inventory trends, and employee performance, consolidating data from multiple tables for a comprehensive view of the company's operations. The report is designed to empower stakeholders to make data-driven decisions by offering valuable insights and facilitating data exploration through interactive visualizations and dynamic filters. The expected impact of the project is to revolutionize how North Wind Traders interacts with its data, enabling the company to remain competitive and drive its business forward in the wholesale market landscape. The dataset includes tables for customers, employees, orders, order details, products, suppliers, shippers, and categories, providing a rich source of information for the analysis. The project involves data cleaning and MECE-breakdown to ensure the quality and structure of the data. Additionally, it includes a set of Power BI and EDA (Exploratory Data Analysis) questions to guide the analysis and visualization process, covering various aspects such as customer distribution, order volume, employee productivity, product performance, supplier metrics, and customer retention. The project's comprehensive approach and the use of interactive visualizations and dynamic filters are aimed at providing a deep understanding of the company's operations and facilitating informed decision-making.

➢ Key Functionality of the Project:

The key functionalities of this Power BI project for Northwind Traders include:

Data Integration and Transformation:

Importing and integrating data from multiple tables (Customers, Employees, Orders, etc.) into Power BI.

Performing data transformations such as cleaning, shaping, and merging to prepare the data for analysis.

Visual Data Exploration:

Creating interactive visualizations to explore key performance metrics such as sales trends, customer segmentation, inventory levels, and employee performance.

Utilizing various chart types (line charts, bar charts, pie charts, etc.) to present data in a meaningful and visually appealing manner.

Implementing slicers, filters, and drill-down capabilities to allow users to dynamically interact with the data and gain deeper insights.

Sales Analysis:

Analyzing sales data to identify trends, patterns, and opportunities for growth. Calculating key sales metrics such as total revenue, average order value, and sales by product category or region.

Visualizing sales performance over time and comparing it with historical data to track progress.

Customer Segmentation:

Segmenting customers based on demographic data (location, industry, etc.) to better understand their behavior and preferences.

Creating visualizations to show the distribution of customers across segments and analyze their purchasing patterns.

Identifying high-value customers and developing targeted marketing

strategies to enhance customer engagement and retention.

Inventory Management:

Monitoring inventory levels and trends to optimize stock levels and minimize stockouts.

Calculating inventory metrics such as turnover rate, stock levels, and reorder frequency.

Visualizing inventory data to identify slow-moving items, excess stock, and potential supply chain issues.

Employee Performance Analysis:

Evaluating employee sales performance based on metrics such as total sales, number of orders processed, and average order value.

Ranking employees by performance and identifying top performers for recognition and reward.

Providing insights into employee productivity and effectiveness to inform decision-making and performance management.

Interactive Dashboards and Reports: Designing interactive dashboards with a user-friendly layout and intuitive navigation. Incorporating storytelling elements to guide users through the data and highlight key insights.

Enabling users to customize views, apply filters, and drill down into specific data points for deeper analysis.

Report Sharing and Collaboration:

Publishing the Power BI report to the Power BI Service for online access. Sharing the report with relevant stakeholders and teams within the organization. Collaborating on data analysis and decision-making by providing access to the report and facilitating discussions around the insights generate

➢ Aim and Objective:

Consolidate Data: Integrate and consolidate data from multiple tables within the Northwind database to provide a comprehensive view of the company's operations.

Analyze Sales Patterns: Analyze sales data to identify trends, patterns, and fluctuations in sales performance over time, by product, region, and customer segment.

Understand Customer Behavior: Analyze customer data to understand buying behavior, preferences, and demographics, enabling targeted marketing strategies and improved customer engagement.

Optimize Inventory Management: Monitor inventory levels, analyze stock movement, and identify inventory trends to optimize inventory management processes and reduce carrying costs.

Evaluate Employee Performance: Assess employee sales performance, productivity, and effectiveness to identify top performers, training needs, and opportunities for improvement.

Provide Actionable Insights: Present insights derived from data analysis in a visually appealing and intuitive dashboard format, enabling stakeholders to make informed decisions and take proactive actions.

Foster Data-Driven Culture: Promote a culture of data-driven decision-making within the organization by providing access to timely and relevant data, empowering stakeholders at all levels to leverage data for strategic planning and operational improvements.

Enhance Business Competitiveness: Use data insights to identify competitive advantages, market opportunities, and areas for innovation, enabling Northwind Traders to stay ahead in the wholesale market landscape.

Facilitate Continuous Improvement: Continuously monitor and analyze key performance metrics, gather feedback from stakeholders, and iterate on the Power BI dashboard to ensure it remains relevant, insightful, and aligned with business objectives.

By achieving these objectives, the Power BI project aims to empower Northwind Traders with actionable insights, enabling them to make informed decisions, drive business growth, and maintain a competitive edge in the marketplace.

DATA CLEANING

Data cleaning is the process of identifying and correcting or removing errors, inconsistencies, and inaccuracies in data. It is an important step in data analysis as it ensures that the data is accurate, complete, and consistent. Data cleaning is used to improve the quality of data and to make it suitable for analysis. It involves identifying missing values, removing duplicates, correcting spelling errors, and standardizing data formats. Data cleaning is essential because it helps to ensure that the insights generated from the data are reliable and accurate.

Some common challenges faced during data cleaning include:

Missing Data: Dealing with missing values in the dataset, which may require imputation or removal of incomplete records.

Inconsistent

Data: Addressing inconsistencies in data formats, such as date formats, currency symbols, or units of measurement.

Duplicate Records: Identifying and removing duplicate entries, which can skew analysis and lead to inaccurate results.

Outliers: Handling outliers that can significantly impact statistical analysis and visualization.

Data Standardization: Ensuring that data is consistent and standardized across different sources or systems.

Data Validation: Verifying the accuracy and integrity of the data, which may involve cross-referencing with external sources or known benchmarks.

Data Transformation: Converting data into a suitable format for analysis, such as aggregating, pivoting, or normalizing data.

Data Quality: Ensuring the overall quality of the data, including accuracy, completeness, and reliability.

These challenges are common in the data cleaning process and require careful attention to detail to ensure the accuracy and integrity of the data for analysis.

Process of Cleaning in North Wind dataset tables:

➢ Customers table:

➢ In Customer table region column contain many null values that are replaced with N/A.

➢ After this removed these columns that are not necessary for the visualization. These columns are Phone, Fax, Image and Image Thumbnail

➢ **Employee table:**
From the employee table these columns are removed that are not providing and information for visualization. These columns are Home Phone, Photo, Report To.

➢ **Supplier table:**
Supplier table also contain these columns that are not necessary for visualization. These columns are Phone, Fax, Home Page.

➢ **Categories table:** categories table contain one column that are not necessary for visualization.

After cleaning all the table are used for the visualization in power BI and build a relationship between the tables :

➢ **Employees table connected to the orders table.**
➢ **Customer table connected to the orders table.**
➢ **Supplier table connected to the Product table.**
➢ **Product table connected to the order details and categories table.**
➢ **Order details table connected to the orders table.**
➢ **Shipper table also connected to the orders table.**

All the tables are connected with each in a one-to-many relationship.

➢ **Overview of Transformed dataset (North wind**

traders)
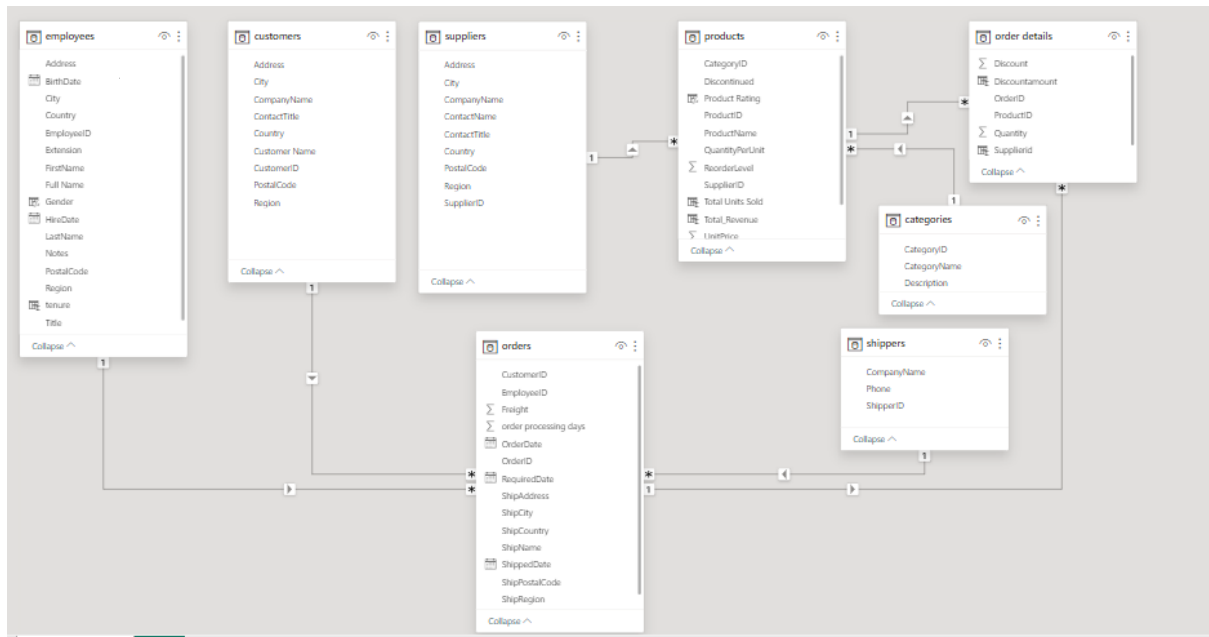
Entity Relationship Diagram

MECE-Breakdown:

MECE stands for "Mutually Exclusive, Collectively Exhaustive." It's a principle often used in problem-solving and structuring information to ensure that all elements being considered are distinct from one another (mutually exclusive) and that together they cover all possible options (collectively exhaustive).

Sales Analysis

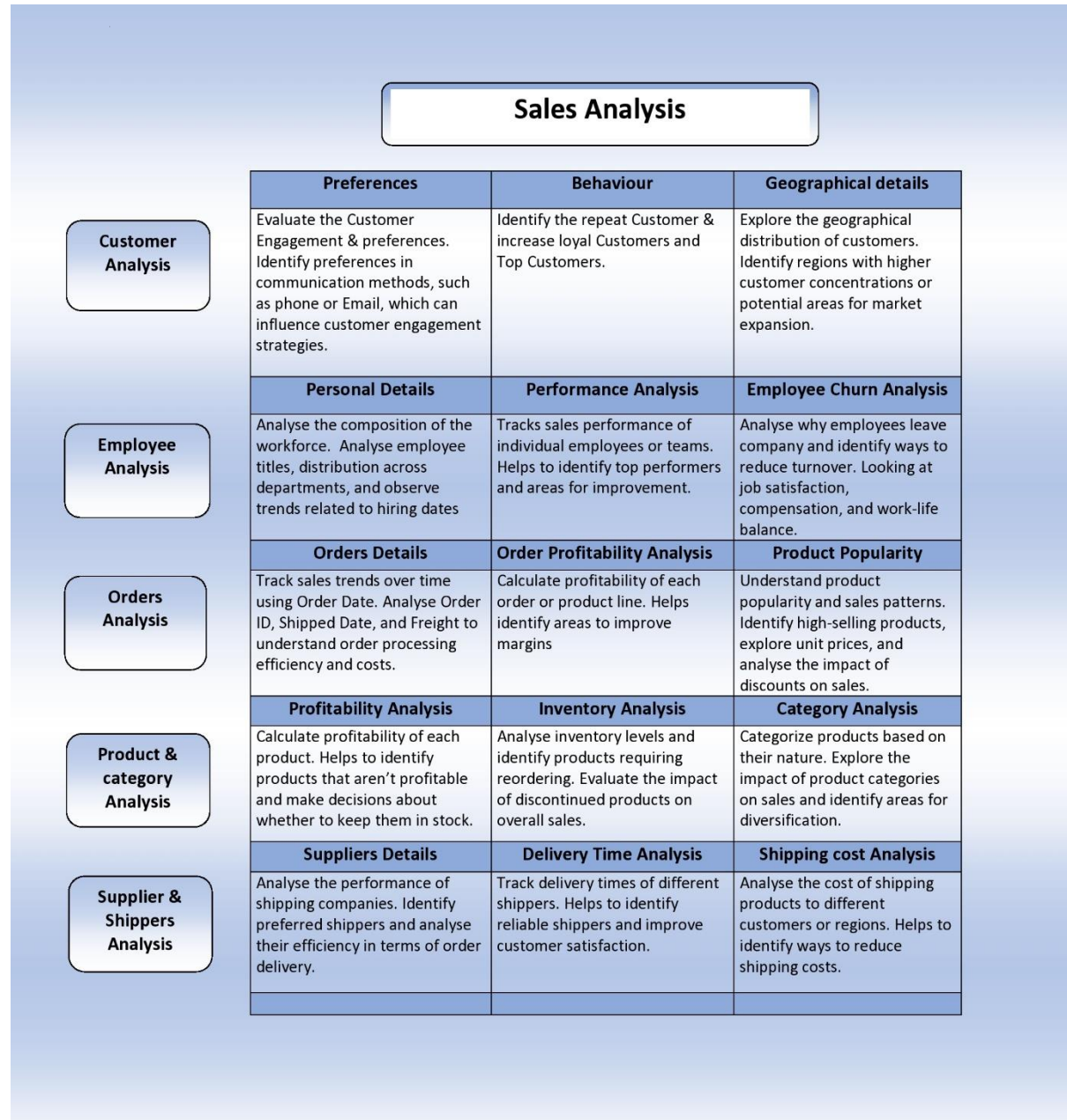|	Preferences	Behaviour	Geographical details
Customer Analysis	Evaluate the Customer Engagement & preferences. Identify preferences in communication methods, such as phone or Email, which can influence customer engagement strategies.	Identify the repeat Customer & increase loyal Customers and Top Customers.	Explore the geographical distribution of customers. Identify regions with higher customer concentrations or potential areas for market expansion.
	Personal Details	**Performance Analysis**	**Employee Churn Analysis**
Employee Analysis	Analyse the composition of the workforce. Analyse employee titles, distribution across departments, and observe trends related to hiring dates	Tracks sales performance of individual employees or teams. Helps to identify top performers and areas for improvement.	Analyse why employees leave company and identify ways to reduce turnover. Looking at job satisfaction, compensation, and work-life balance.
	Orders Details	**Order Profitability Analysis**	**Product Popularity**
Orders Analysis	Track sales trends over time using Order Date. Analyse Order ID, Shipped Date, and Freight to understand order processing efficiency and costs.	Calculate profitability of each order or product line. Helps identify areas to improve margins	Understand product popularity and sales patterns. Identify high-selling products, explore unit prices, and analyse the impact of discounts on sales.
	Profitability Analysis	**Inventory Analysis**	**Category Analysis**
Product & category Analysis	Calculate profitability of each product. Helps to identify products that aren't profitable and make decisions about whether to keep them in stock.	Analyse inventory levels and identify products requiring reordering. Evaluate the impact of discontinued products on overall sales.	Categorize products based on their nature. Explore the impact of product categories on sales and identify areas for diversification.
	Suppliers Details	**Delivery Time Analysis**	**Shipping cost Analysis**
Supplier & Shippers Analysis	Analyse the performance of shipping companies. Identify preferred shippers and analyse their efficiency in terms of order delivery.	Track delivery times of different shippers. Helps to identify reliable shippers and improve customer satisfaction.	Analyse the cost of shipping products to different customers or regions. Helps to identify ways to reduce shipping costs.

UNDERSTANDING & SOLVING POWER BI QUESTION

There 15 Question of Power BI and all these Questions are divided into 5 parts:

➢ **Customer Analysis:**
- Distribution: Visualize regional/segment spread (map/bar chart) to understand customer base diversity.
- Acquisition: Track customer growth over time (line/area chart) to identify trends and potential strategies.
- Demographics: Explore age, gender, income distribution (histograms/pie charts) for targeted marketing.

➢ **Order Analysis:**
- Volume: Track order volume changes over time (time series/stacked bar chart) to assess demand fluctuations.
- Value: Analyze order value distribution (histogram/box plot) to understand purchase behavior and potential upselling opportunities.
- Processing/Shipping: Visualize processing/shipping times (bar/box plot) to identify areas for improvement.

➢ **Employee Analysis:**
- Productivity: Compare productivity across departments/roles (stacked bar/grouped column chart) to optimize resource allocation.
- Tenure: Understand employee retention with tenure distribution (histogram/box plot) to identify potential churn risks.
- Performance: Visualize performance ratings/KPIs (radar/bullet chart) for individual and team evaluation.

➢ **Product Analysis:**
- Sales: Track sales volume by category (bar/tree map) to identify best-sellers and potential growth areas.
- Pricing: Analyze product pricing distribution (box/violin plot) to
- understand competitiveness and profitability.

> ➢ **Supplier Analysis:**
> - **Ratings:** Assess supplier performance with rating distribution (bar/radar chart) to inform sourcing decisions.
> - **Cost/Pricing:** Compare supplier costs/pricing structures (box plot/stacked bar chart) to optimize procurement.
> - **Distribution:** Visualize supplier locations (map/bubble chart) to understand logistical considerations.
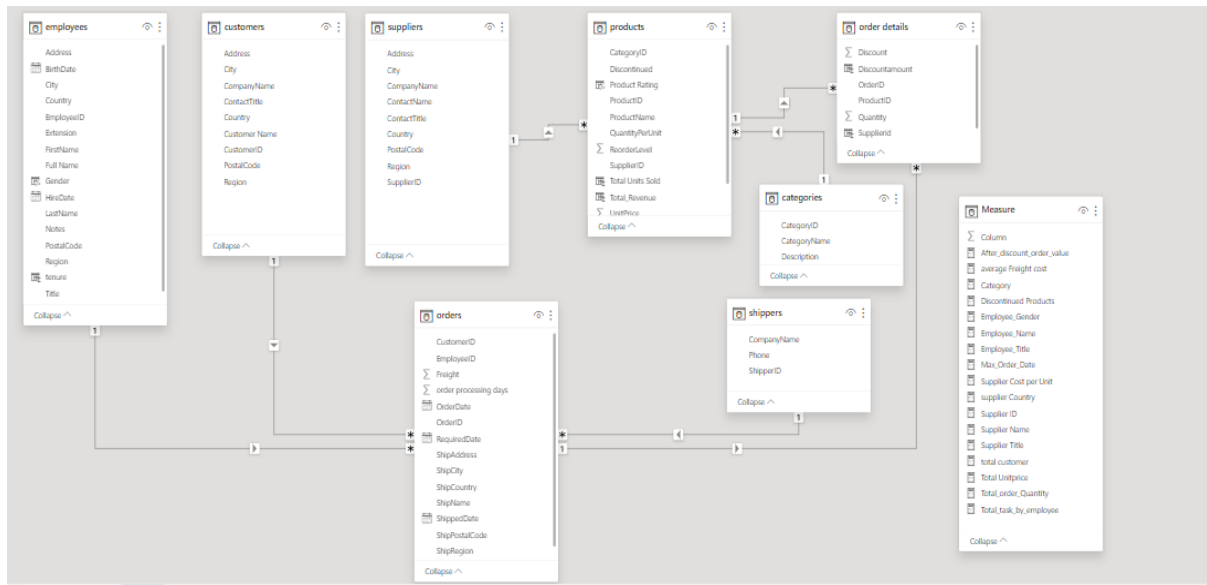

ER-Diagram

> ➢ **Before doing any analysis, I have created a measure table contain all measure that are used in in Analysis in power bi**
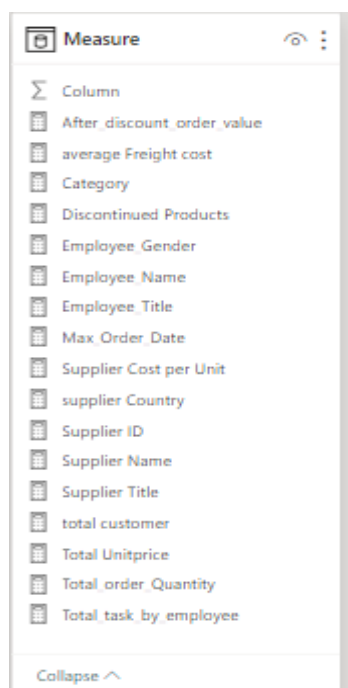
1. How does customer distribution vary across different regions or customer segments? Can we visualize it on a map or bar chart?

➢ **Yes, we can create a map for visualization of customer distribution across different countries and also, we can create a bar chart for it.**

➢ **In Power BI, a map visualization displays geographic data using interactive maps, allowing users to visualize and analyze spatial relationships, trends, and patterns within their data**.
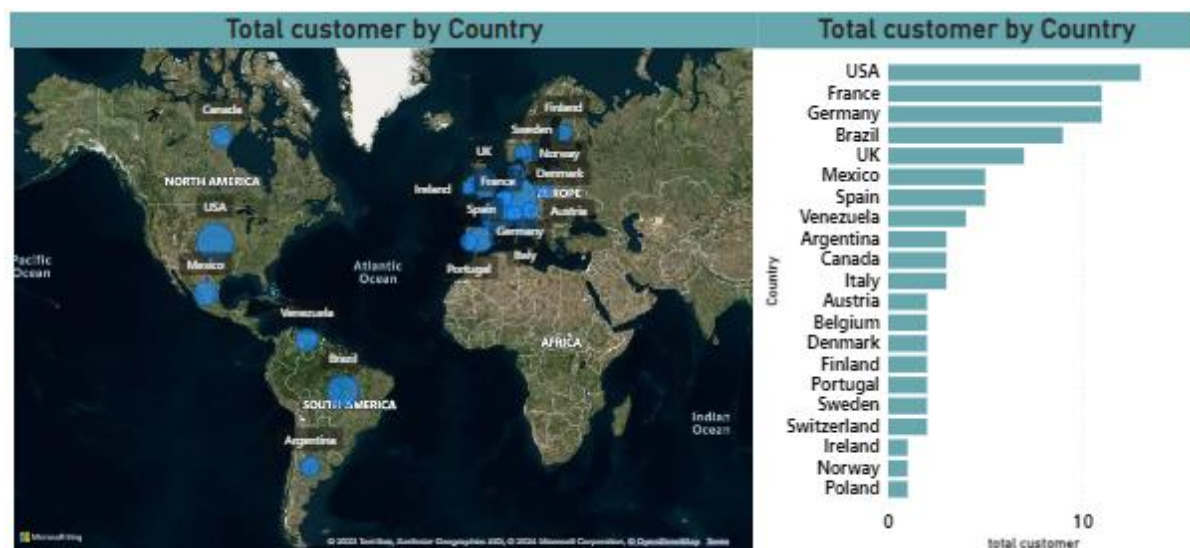

For this analysis Country Column is used from orders table and also to find the total number of customers from each country measure is used that count the distinct customer id from the customer table.

This shows the Distribution of the customers across country


```
total customer = count(customers[CustomerID])
```

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

**Customer Acquisition Over Time**
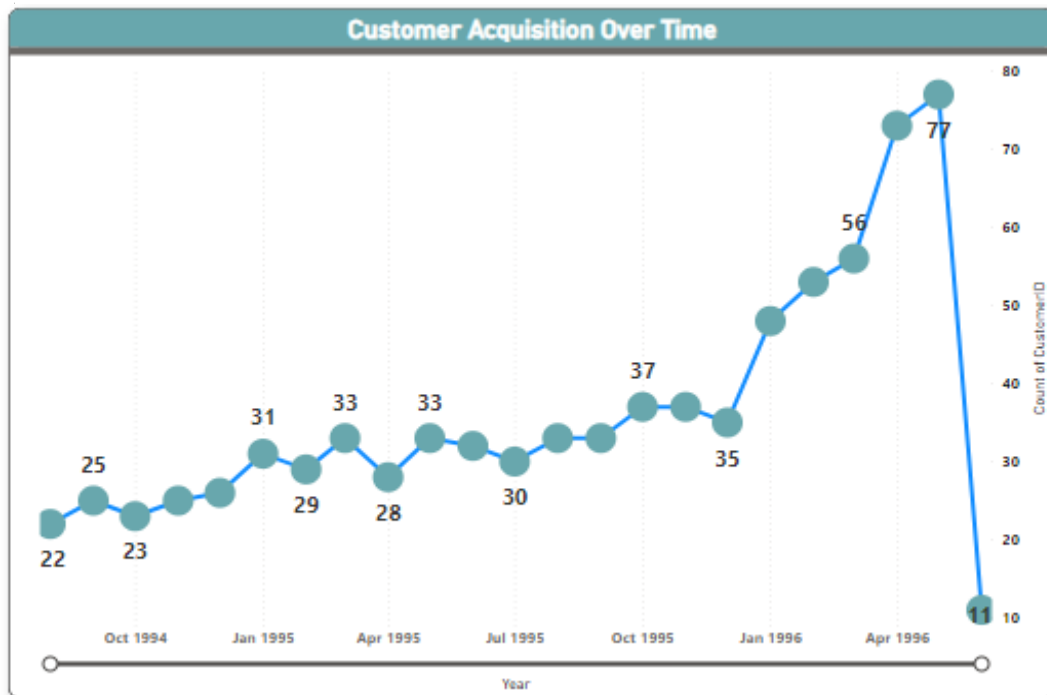
For this analysis, order date from the orders table and count of customers from the orders table are used.

The customer acquisition trend is increasing over time, but with some fluctuations. There seems to be a steeper increase from October 1994 to April 1995, followed by a period of slower growth until June 1996.

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

> ➢ For this analysis, choose the country column from the customers table and customer ID from the orders table this give the distribution of customers across country and also use the contact Tile column to get the distribution of customers by contact Title in bar chart.
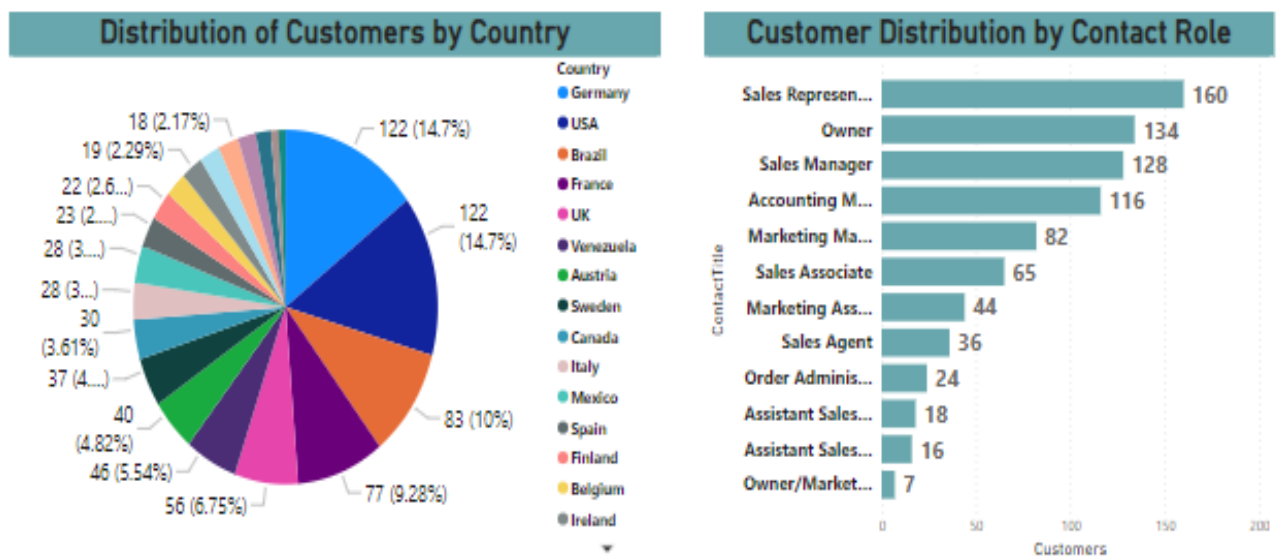
> ➢ **Distribution of Customers by Country:**
>   A majority of customers around 60% are from Germany and the United States. The remaining customers are spread across various countries, with no other country exceeding 10%.
> ➢ **Customer Distribution by Contact Role:**
> • The largest group of customers around 40% have the contact role of Sales Representative.
> • Other significant groups include Sales Manager around 14% and Accounting Manager around 10%.
> • Smaller groups include Sales Associate, Marketing Manager, Order Administrator, Assistant Sales, and Owner/Marketing.

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

For this analysis, choose the column order date from the orders table and order Id from the orders table and this give the order volume over time.
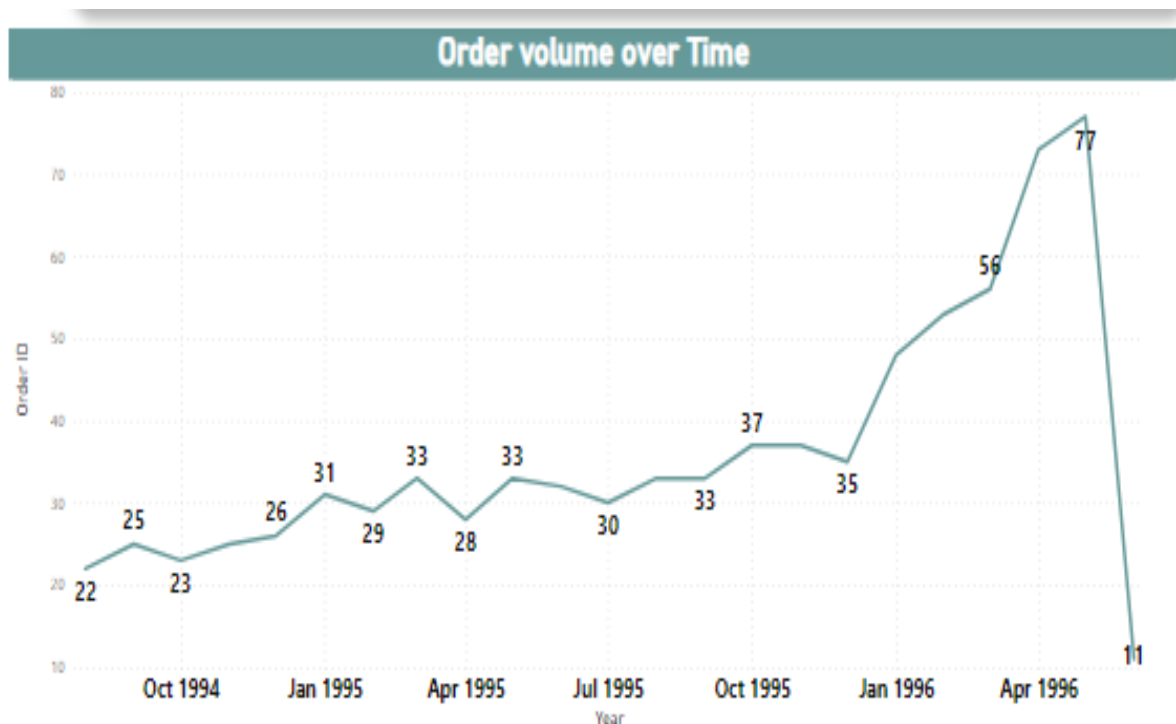

The order volume shows a gradual increase over time, with some fluctuations. There's a notable jump in volume between July 1995 and April 1996.

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

For this analysis, choose the after-discount order value from the measure and order Id from the orders table, this give the distribution of order value.

```
After_discount_order_value = sumx('order details',('order details'[UnitPrice]-'order details'[Discountamount])*'order
details'[Quantity])
```


Distributin of order value
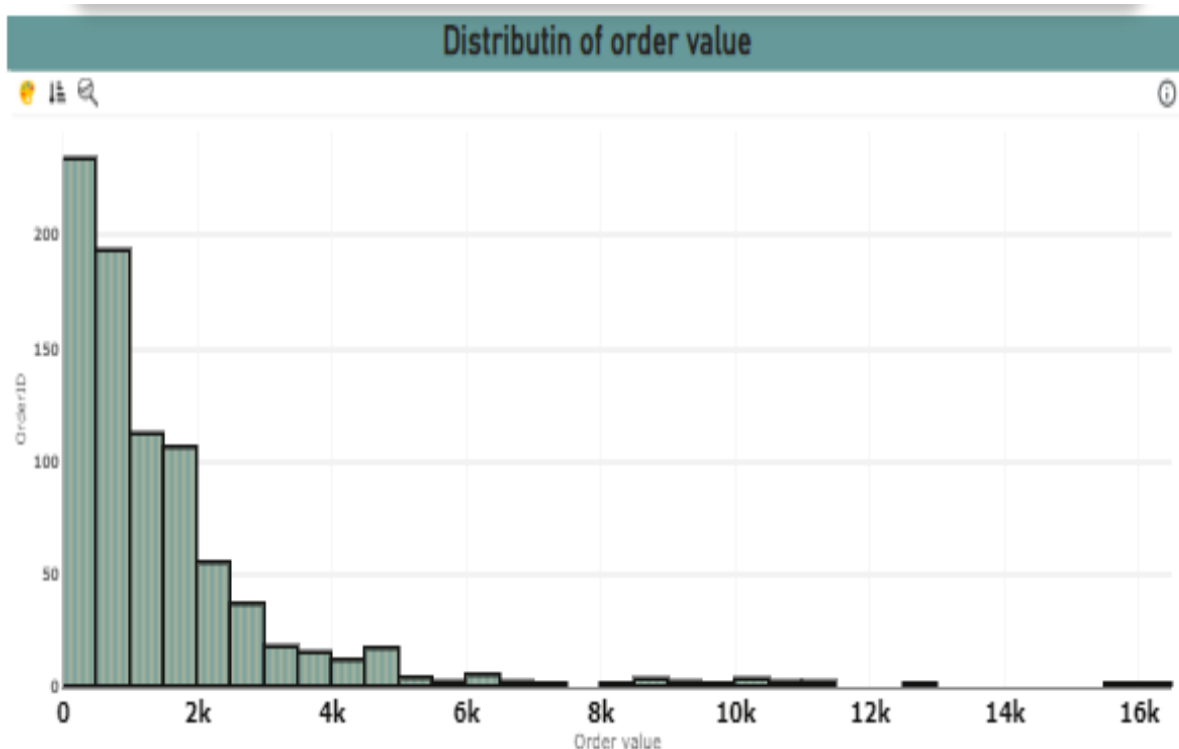
Majority of orders have a low to moderate value. The peak of the distribution is concentrated around the $2,000-$4,000 range, suggesting a high number of orders fall within this price. While most orders are concentrated in the lower to mid-range, there's a noticeable tail extending towards higher order values, indicating a presence of some large orders potentially impacting overall revenue

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

For this analysis, choose the average delivery time column form orders table (order date-ship date) and Order Id from the orders table. And visualize it on the box plot chart.
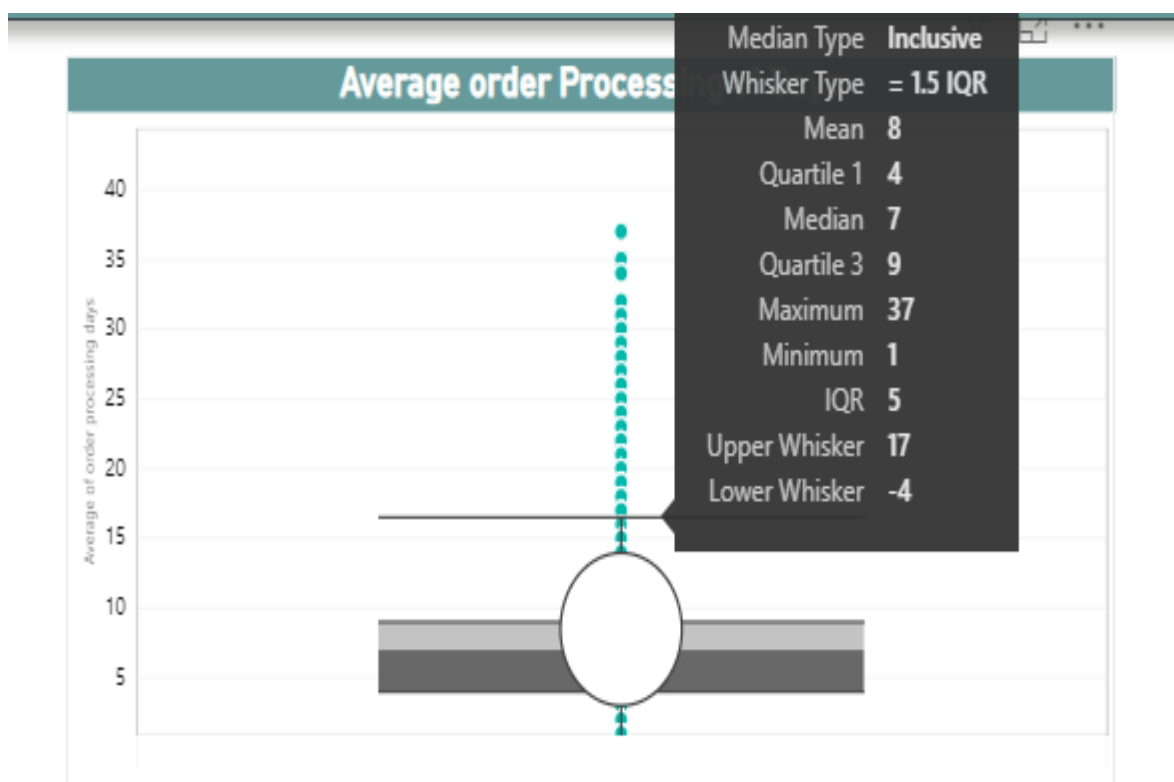

The x-axis shows whisker type, labeled 1.5 IQR. The y-axis shows average processing time in days. There are four horizontal lines representing quartiles and outliers. The median processing time is 7 days. The average processing time is 8 days. The whisker range is from -4 days to 17 days. There are two outliers, one at 37 days and one at 1 day.

7.How does employee productivity vary across different departments or job roles? Can we create a stacked bar chart or grouped column chart to visualize it?

For this analysis , choose the Employee Full name from the employees table and used the measure to check the total task of the each employee in different job roles.

```
Total_task_by_employee = count(orders[EmployeeID])
```
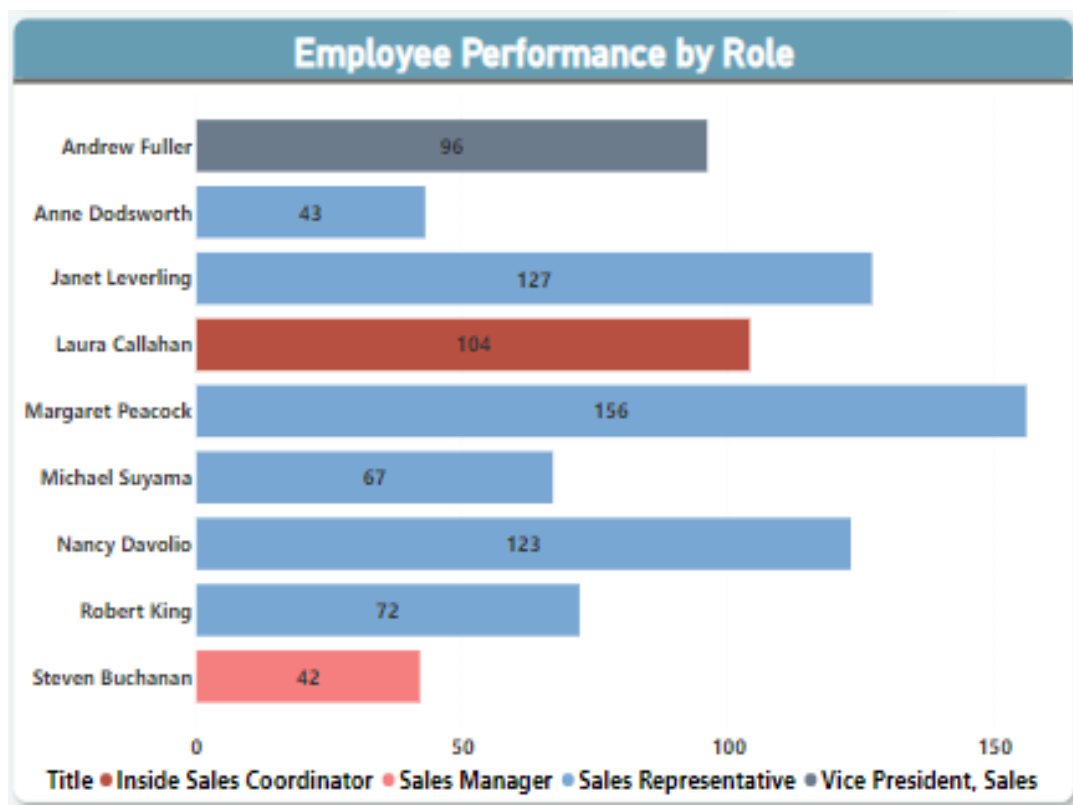

Sales representatives have the highest average productivity. Most of orders are done by the sale representatives. Insider sales coordinator and vice president have done almost equal numbers of orders and sales manager have the lowest Productivity. There's a significant difference in productivity between the top and bottom roles

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

For this analysis, First created a new column Tenure (month) in the employees table this give the Total tenure of the each employee and then visualize it on the box plot chart to check the outlier in tenure of employee.
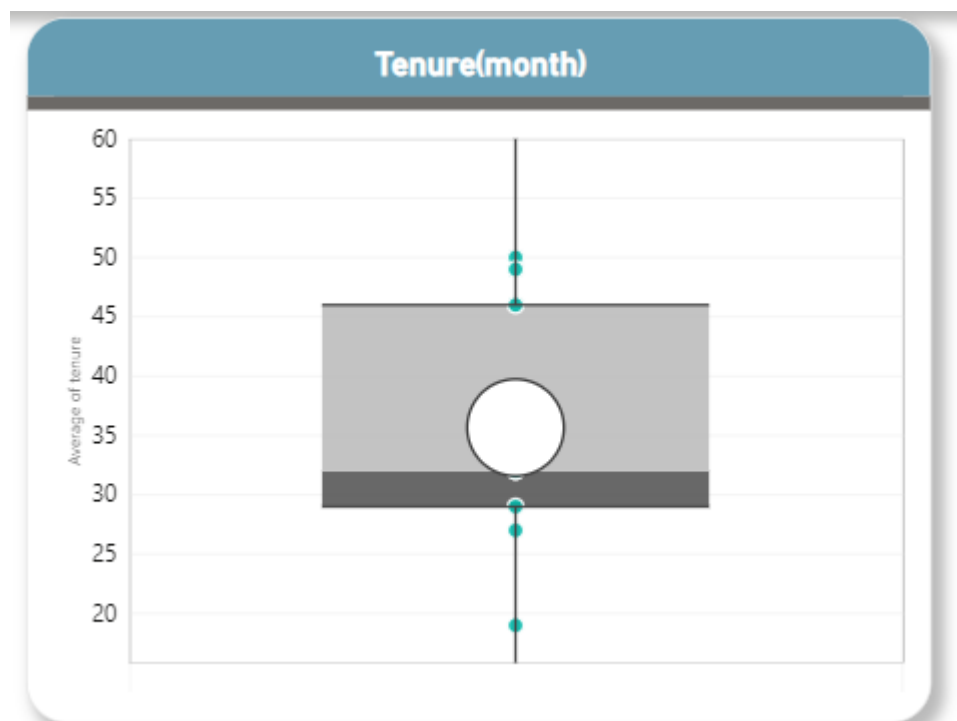

Majority tenure of employees have been with the company around between 29 and 46 Months. Average Tenure of employees is 35.6.Highest tenure of the employee is 50 month and lowest is 19 months .This indicate a presence of employees who have been with the company for a specific period of times.

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

For this analysis, check that how many order I'd done by employees.

This will give a criterion to compare the performance of the employees.
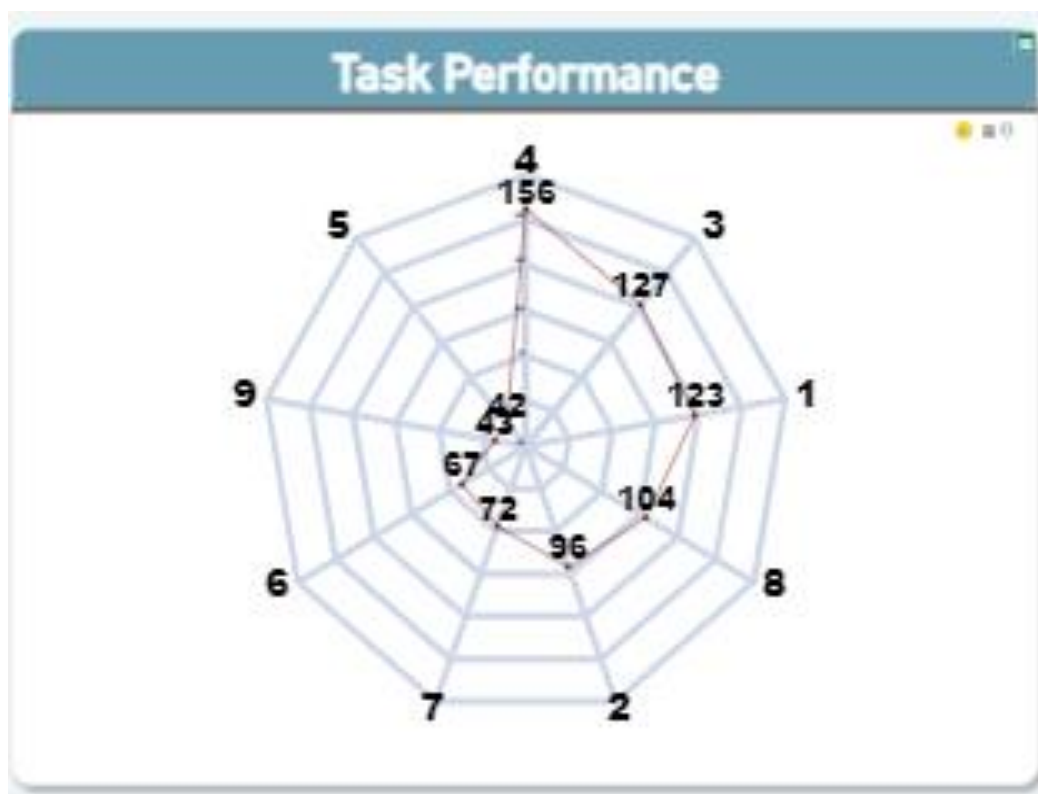

Yes, we can visualize it by radar chart to check employee performance. Here employe ID 4 have done the highest task and lowest tasks done by Emp Id 5 (42 tasks) This help us to identify the performance of the employees.

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

For this analysis, First Created a column product Rating in product table by this Dax function:

```
Product Rating =
IF(AND(products[Total_Revenue] < 5000,products[Total Units Sold] < 100), "Low",IF( AND( products[Total Units Sold] < 600,
        products[Total_Revenue]< 10000 ), "Below Average",IF( AND(  products[Total Units Sold] < 1000,
            products[Total_Revenue] < 20000 ), "Average", IF(   AND(  products[Total Units Sold] < 1500,
                products[Total_Revenue] < 80000 ), "Excellent",  "Good")) ))
```

After this selected bar chart and drag the product rating column in the y-axis and product id from the orders table in the X-axis .This will give a count of products lie in different criteria .
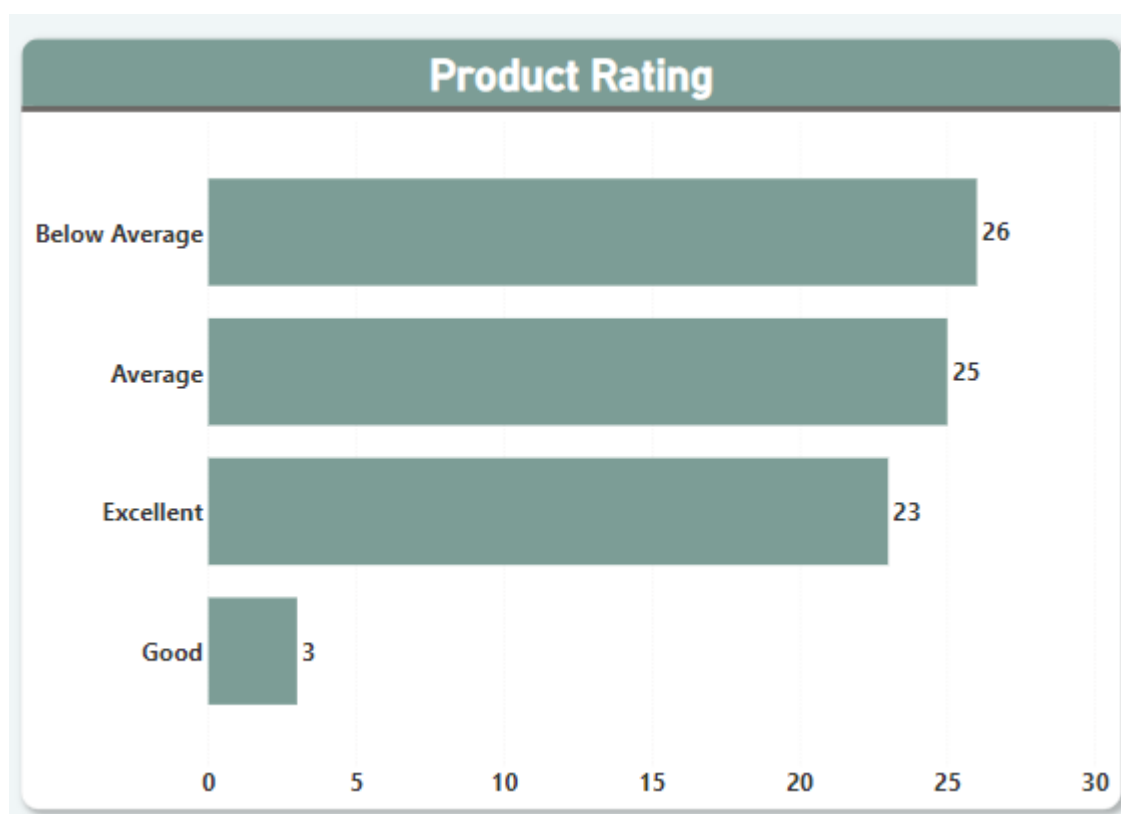
Product Rating

we can create a stacked bar chart to check products performance by distributing them into a product rating. **While the majority of products fall within the average and below average range, only a small percentage achieve "good" and "excellent" ratings. This signifies a potential opportunity to improve product quality or perception to increase customer satisfaction and positivefeedback.**

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

For this analysis, choose the category name from the categories table and choose After discount order value from the measure table. This will give the order volume vary across diffenent categories.

```
After_discount_order_value = sumx('order details',('order details'[UnitPrice]-'order details'[Discountamount])*'order details'[Quantity])
```
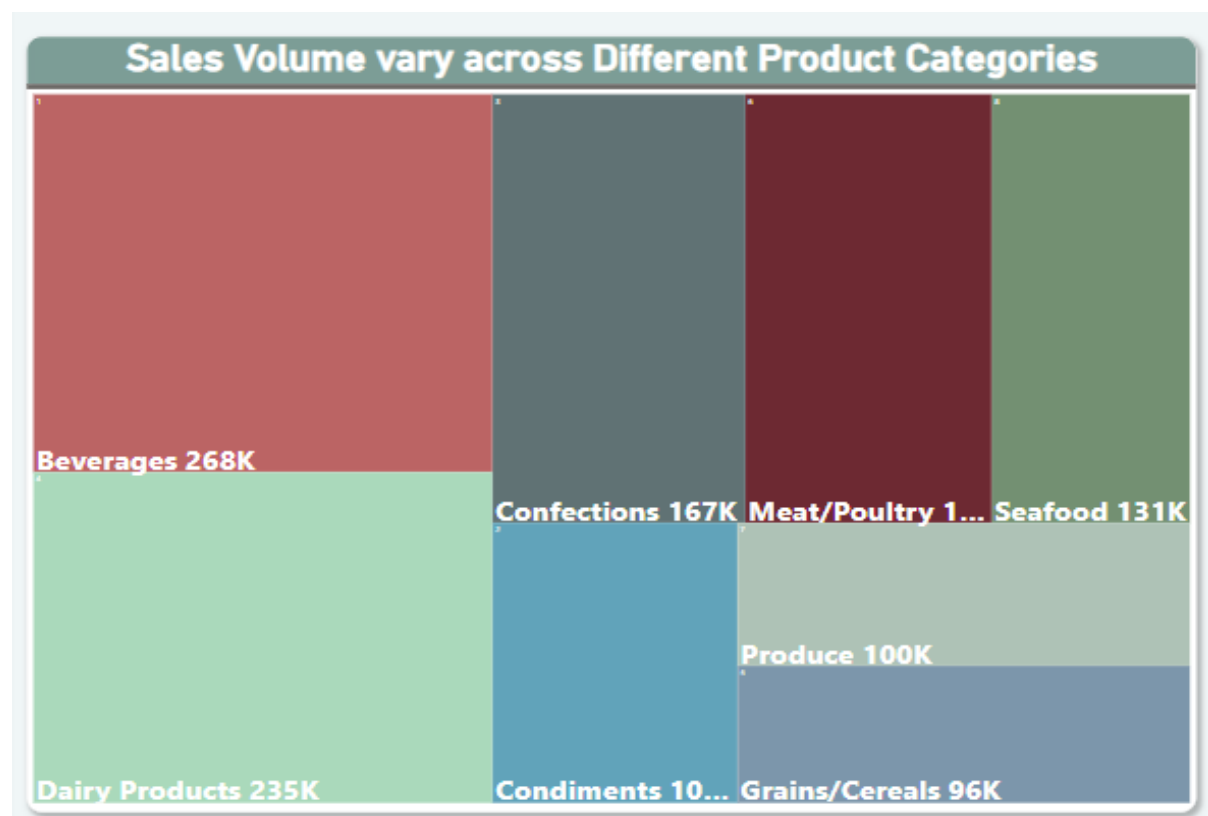

Yes, we can create a tree map to check sales volume across the different Product categories.

- Beverages and dairy products are the top sellers, generating significantly higher sales volume compared to other categories.

- Confections, seafood, and meat/poultry show moderate sales volume, forming a middle tier and other categories show the lowest sales volume.

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

For this analysis, First created a measure total unit price.

```
1  Total Unitprice = AVERAGE('order details'[UnitPrice])
```

After this drag the product from the products table and also drag the total unit price measure into a box plot chart. This shows the outliers in product price.
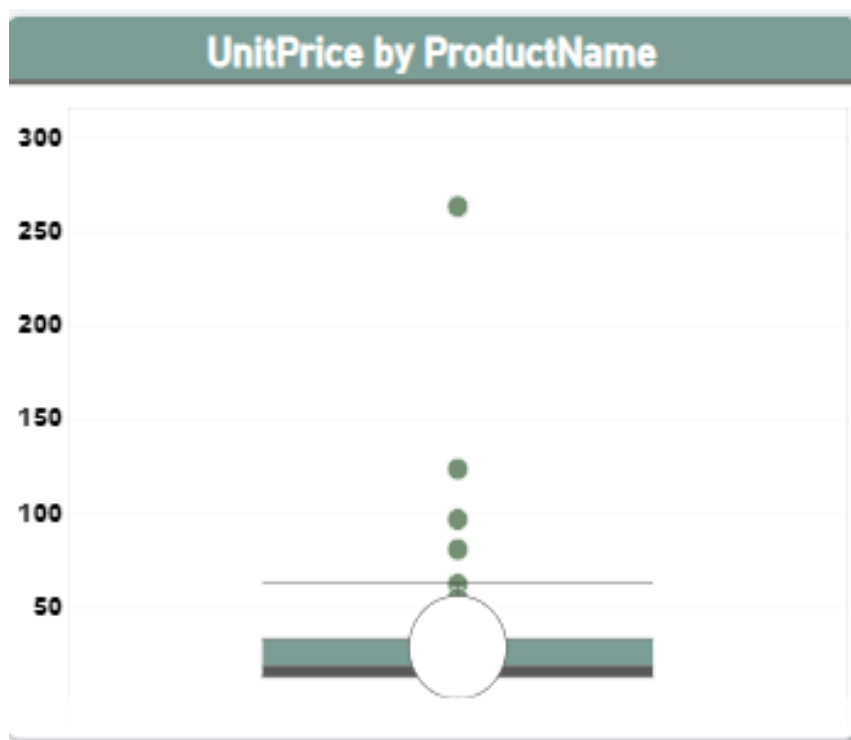
we can visualize the price distribution of products using box plot. It help us to identify the outlier in the product price means which product have the highest price and lowest different from the average price of all the products. The most of product price lie b/w 13 and 33. The max price of the product is 263 and min price is 2.5 and the average of price of the product is 28.

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

For this analysis, choose the Supplier Name from the suppliers table and choose the Quantity column from the order details table by comparing the total quantity deliver by each supplier.
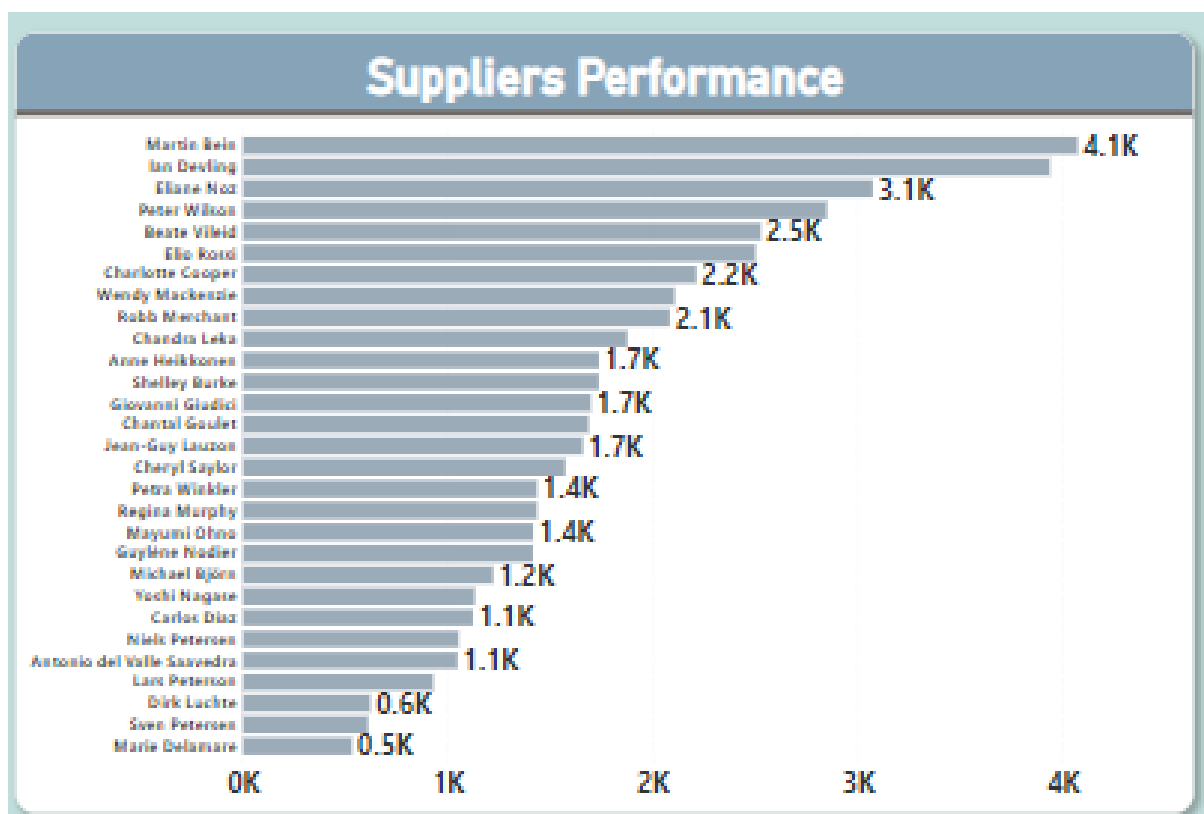
we can create a bar chart to visualize the distribution of supplier rating by check which supplier have highest no. of order quantity delivery. if the company buying product continuously then we can compare the supplier with there order quantity. From the chart we can see that supplier martin Bein have the highest order quantity.

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

For this analysis, choose the contact's name column from the Supplier table and unit price column from the product to get the average unit price provided by each supplier

Average of UnitPrice by Suppliers

we can create a bar char to visualize the cost price structure vary across different suppliers by compare the average unit price provided by the suppliers. This help us to check and compare the unit price of product by different Supplier to identity the which supplier is best for which product.

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

For this analysis, choose the country column from the Supplier from Supplier table and also supplier id to get the count of supplier in each country.

we can visualize the geographical distribution of supplier by map chart. This helps us to identify the country and region of supplier and how many suppliers in each country

DASHBOARD

NORTHWIND TRADERS DASHBOARD

NORTHWIND TRADERS

Customer Anaylsis	This Dashboard help us to Identify the identify and target high valuable customers.By understanding the customer demographics,Preferences and buying habits.
Order Analysis	This Dashboard help us to identify the top selling product and Categories, Analyze the sales trend by region, country or over period and evaluate the effectiveness of promotion/marketing.
Employee Analysis	This Dashboard help us to identify top performing employees and understand the employee Productivity across differnent job roles which helps us to track engagment and satifaction.
Product Analysis	This Dashboard help us to identify the popular and unpopular products by analyzing the sales data and reviews. Analyze the customer purchase pattern and also show the product price and over all sales.
Supplier Analysis	This Dashboard help us to analyze the Supplier Performance ,cost ,Geographical Distribution of the Suppliers and market trend to nagotiate the Supplier cost.

Customer Analysis Dashboard:
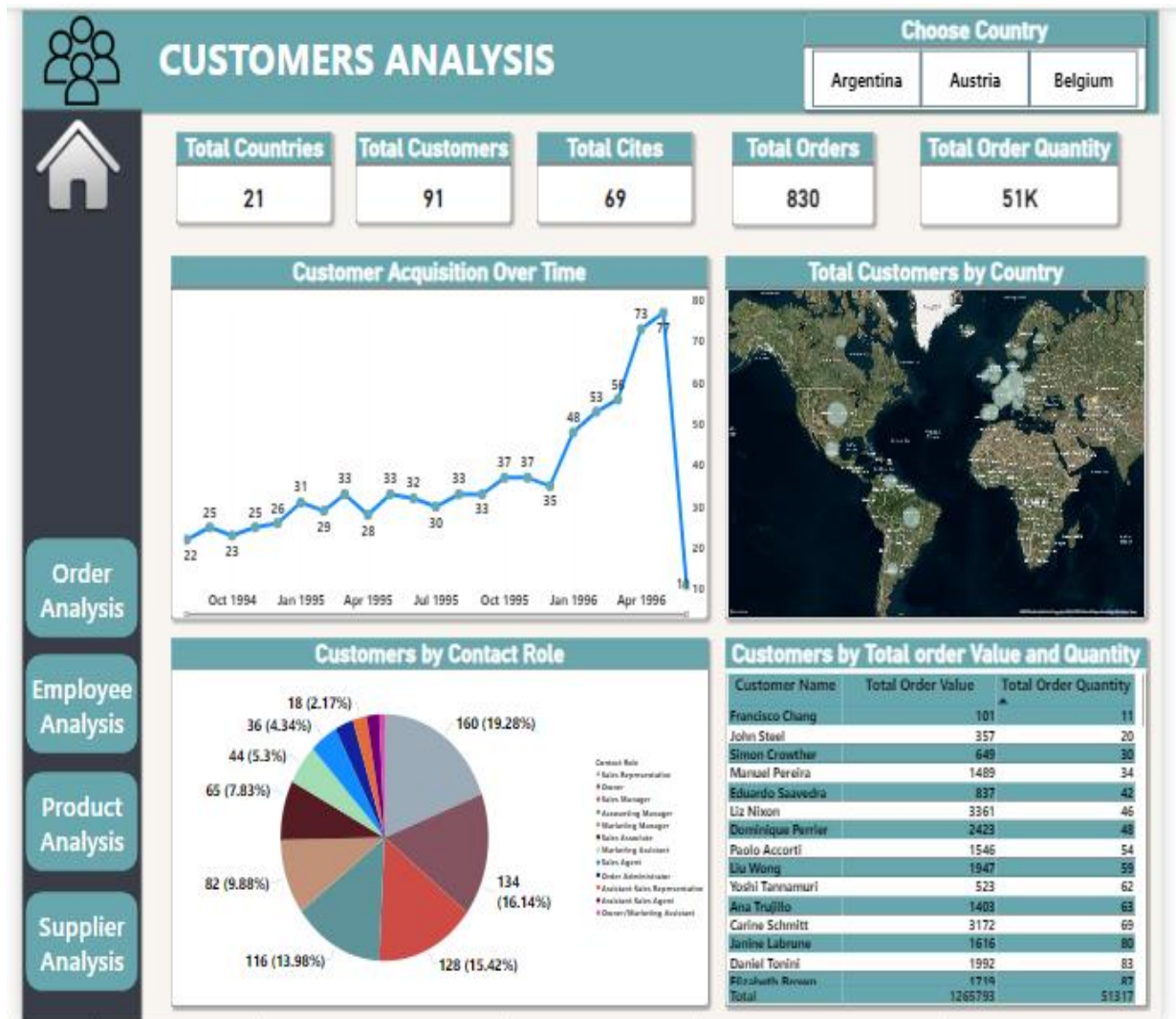
Orders Analysis Dashboard:

Employee Analysis Dashboard:

Product Analysis Dashboard:

Product Analysis Dashboard:

UNDERSTANDING & SOLVING EDA QUESTIONS

There are 15 Question of EDA and these questions are divided into 4 parts:

> **Customer Analysis:**

Retention/Loyalty: Identify key factors driving customer retention and brand loyalty.

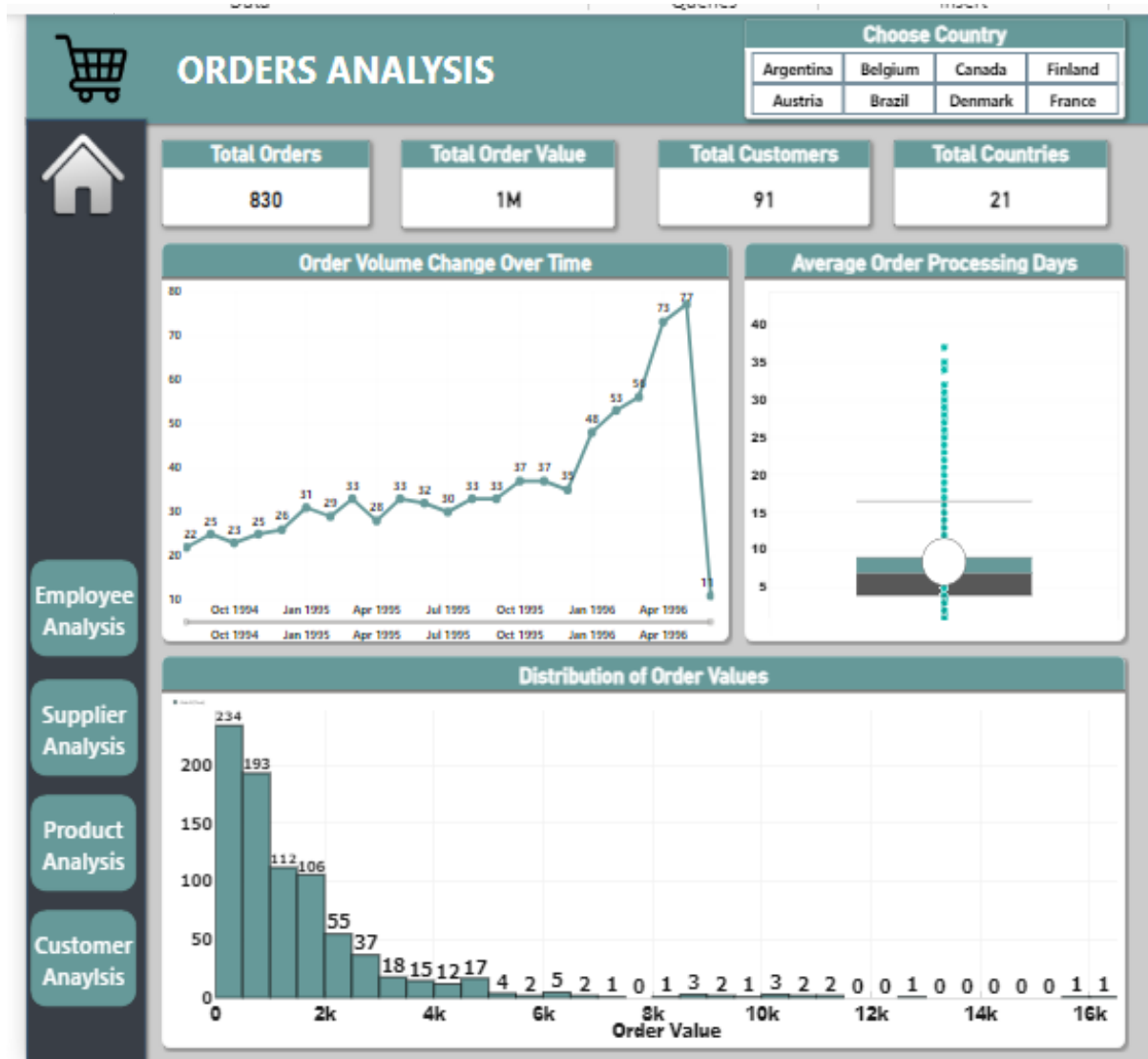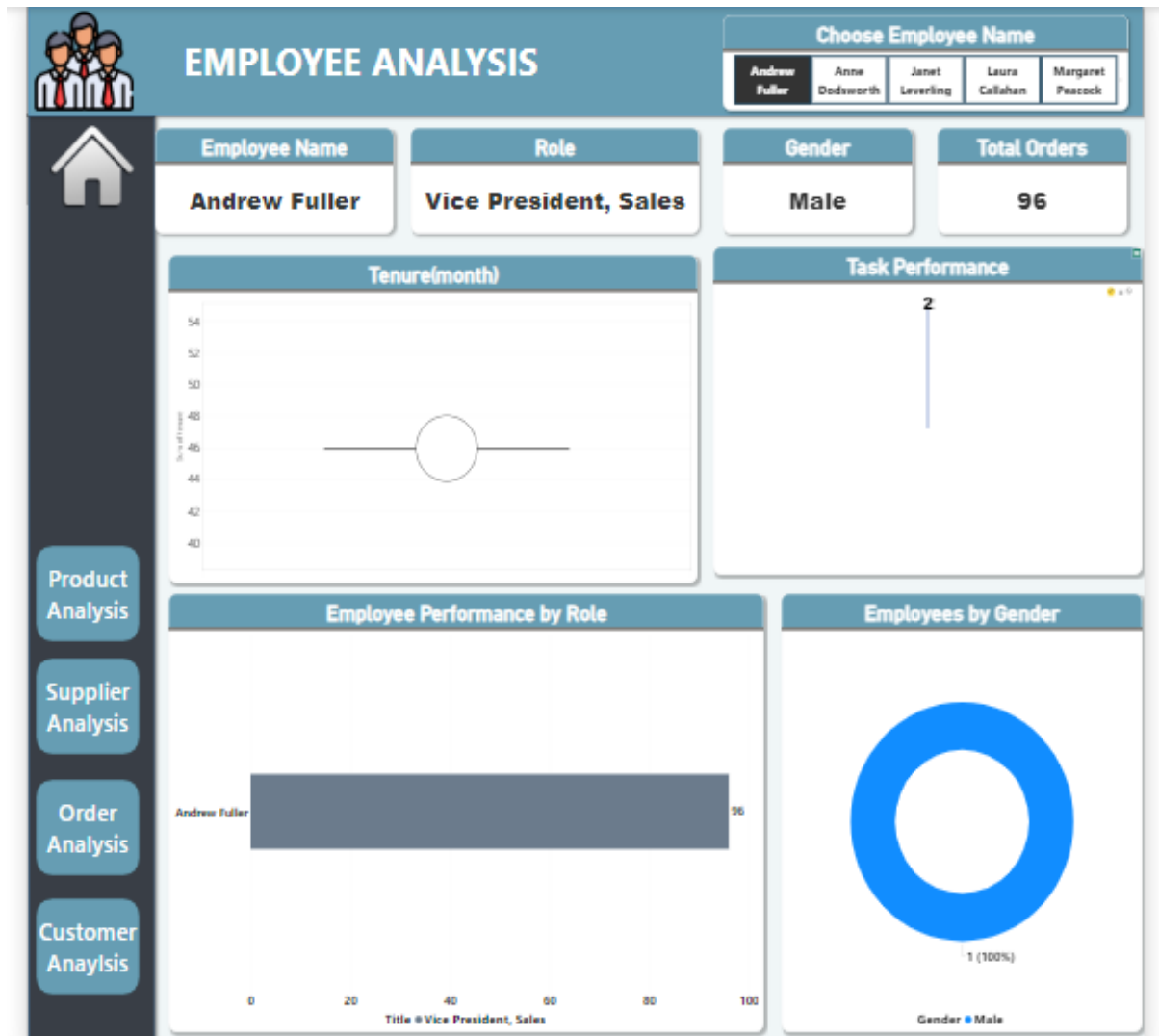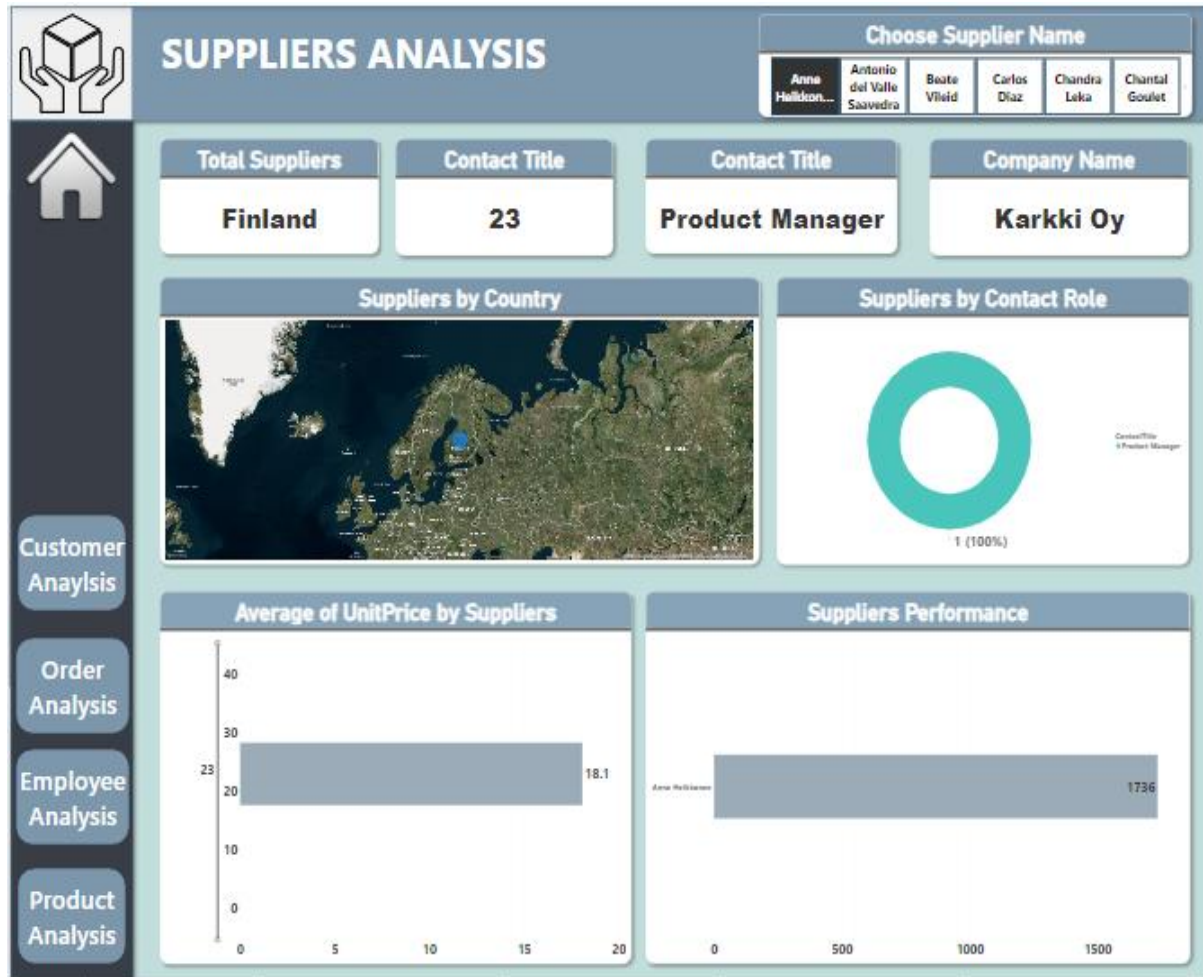**Preferences:** Explore variations in preferences based on location and demographics (interactive visualizations encouraged).

Segmentation: Uncover patterns and clusters in customer behavior for potential market segmentation (visualizations recommended).

> **Product Analysis:**

Revenue Drivers: Identify product categories and SKUs with the highest impact on order revenue (visualizations suggested).

Order Size Correlations: Explore correlations between order size and customer demographics or product categories (scatter plots/heatmaps).

Order Frequency Distribution: Visualize order frequency across different customer segments (bar charts/tree maps).

> **Employee Analysis:**

Satisfaction & Performance: Investigate correlations between employee satisfaction and key performance indicators (scatter plots/line charts).

Turnover Variations: Visualize employee turnover variations across departments or job roles (bar charts/heatmaps).

Skill & Qualification Clusters: Identify patterns or clusters in employee skill sets for talent management (visualizations helpful).

➤ **Supplier Analysis:**
  Product Attributes & Sales: Explore correlations between product attributes and sales performance (scatter plots/heatmaps).
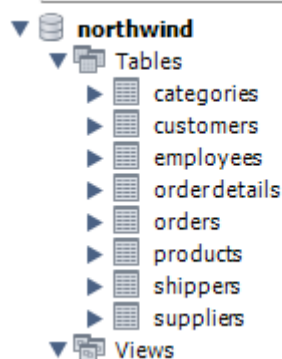 Demand Fluctuations: Visualize product demand fluctuations over time (line/area charts).
  Performance Variations: Analyze supplier performance variations across categories/departments (stacked bar/grouped column charts).
  Cost & Pricing Trends: Identify trends in supplier costs and pricing structures for procurement optimization (visualizations beneficial).

Before starting solving the EDA questions :

➤ **Load the Dataset SQL file into the SQL server.**

▼ 🗄 **northwind**
  ▼ 🗂 Tables
    ▶ 🎛 categories
    ▶ 🎛 customers
    ▶ 🎛 employees
    ▶ 🎛 orderdetails
    ▶ 🎛 orders
    ▶ 🎛 products
    ▶ 🎛 shippers
    ▶ 🎛 suppliers
  ▼ 🗂 Views

➤ **After loading created a view in SQL server of each table that are necessary for solving the EDA Questions.**

▼ 🗂 Views
  ▶ ⊞ analysing_purchase_history
  ▶ ⊞ analyzing_product_preferences
  ▶ ⊞ analyzing_product_preferences_by_country
  ▶ ⊞ correlations_order_size_customer_demographics
  ▶ ⊞ cost_pattern
  ▶ ⊞ customer_behaviour
  ▶ ⊞ customer_preference
  ▶ ⊞ customer_segmentation_based_on_rfm
  ▶ ⊞ customer_segments
  ▶ ⊞ emp_performance
  ▶ ⊞ identifying_churning_customers
  ▶ ⊞ loyalty
  ▶ ⊞ product_category_preference
  ▶ ⊞ product_demand
  ▶ ⊞ product_demand_fluctuate
  ▶ ⊞ product_performance
  ▶ ⊞ prouduct_attribute

Iministration   Schemas

northwind [32]
- analysing_purchase_history
- analyzing_product_preferences
- analyzing_product_preferences_by_coun...
- correlations_order_size_customer_demo...
- cost_pattern
- customer_behaviour
- customer_preference
- customer_segmentation_based_on_rfm
- customer_segments
- emp_performance
- identifying_churning_customers
- loyalty
- product_category_preference
- product_demand
- product_demand_fluctuate
- product_performance

➤ **After these steps connected the Excel with SQL server on local host to access the created view in SQL server work bench. After All these steps we can start solving EDA Questions.**

X

From ODBC

Data source name (DSN)

MySql-ODBC ▾

▷ Advanced options

OK   Cancel

analysing_purchase_history
Preview downloaded on 30 January 2024

CustomerID	CompanyName	ProductName
ALFKI	Alfreds Futterkiste	Escargots de Bourgogne
ALFKI	Alfreds Futterkiste	Chartreuse verte
ALFKI	Alfreds Futterkiste	Vegie-spread
ALFKI	Alfreds Futterkiste	Fløtemysost
ALFKI	Alfreds Futterkiste	Rössle Sauerkraut
ALFKI	Alfreds Futterkiste	Grandma's Boysenberry Spread
ALFKI	Alfreds Futterkiste	Lakkalikööri
ALFKI	Alfreds Futterkiste	Raclette Courdavault
ALFKI	Alfreds Futterkiste	Aniseed Syrup
ALFKI	Alfreds Futterkiste	Spegesild
ALFKI	Alfreds Futterkiste	Original Frankfurter grüne Soße
ANATR	Ana Trujillo Emparedados y helados	Camembert Pierrot
ANATR	Ana Trujillo Emparedados y helados	Mascarpone Fabioli
ANATR	Ana Trujillo Emparedados y helados	Konbu
ANATR	Ana Trujillo Emparedados y helados	Mozzarella di Giovanni
ANATR	Ana Trujillo Emparedados y helados	Teatime Chocolate Biscuits
ANATR	Ana Trujillo Emparedados y helados	Outback Lager
ANATR	Ana Trujillo Emparedados y helados	Singaporean Hokkien Fried Mee
ANATR	Ana Trujillo Emparedados y helados	Tofu
ANATR	Ana Trujillo Emparedados y helados	Queso Cabrales
ANATR	Ana Trujillo Emparedados y helados	Gudbrandsdalsost
ANTON	Antonio Moreno Taquería	Queso Cabrales

Load ▾   Transform Data   Cancel

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

Row Labels	order Frequency	Recency_	Monetary value
ALFKI	12	27	4273
ANATR	10	63	1402.95
ANTON	17	98	7023.98
AROUT	30	26	13390.65
BERGS	52	63	24927.58
BLAUS	14	7	3239.8
BLONP	26	114	18534.08
BOLID	6	43	4232.85
BONAP	44	0	21963.25
BOTTM	35	12	20801.6
BSBEV	22	22	6089.9
CACTU	11	8	1814.8
CENTC	2	657	100.8
CHOPS	22	14	12348.88
COMMI	10	14	3810.75
CONSH	7	103	1719.1
DRACD	10	2	3763.21
DUMON	9	79	1615.9
EASTC	21	8	14761.04
ERNSH	102	1	104874.98
FAMIA	19	187	4107.55
FOLIG	16	135	11666.9
FOLKO	45	9	29567.56
FRANK	48	27	26656.56
FRANR	6	43	3172.16
FRANS	10	6	1545.7

```
create view Loyalty as select
o.CustomerID,
count(o.OrderID) Frequency,
min(datediff((select max(date(OrderDate)) from orders),date(o.OrderDate))) Recency,
round(sum( UnitPrice * Quantity * ( 1 - Discount )),2) Monetary
from orders o
join northwind.`order details` od
on o.OrderID = od.OrderID
group by 1
order by 1
```

conclusion :This query calculates RFM values (Recency, Frequency, Monetary Value) for customer, enabling segmentation based on their purchasing behavior . Recency show the number of days between the customer's latest purchase and a reference date,often the current date. It provides a straightforward understanding of how recently a customer has engaged inbuying activity.

The most loyal customers:

SAVEA: This customer has the highest number of orders 116 and has spent the most money in total 104361.95. They also placed an order recently within 5 days.
ERNSH: This customer has the highest number of orders 102 and has spent the most money in total 104874.98. They also placed an order recently within 1 day.
**RATTC:** This customer has the highest number of orders 77 and has spent the most money in total 51097.8. They also placed an order recently within 0 days.
QUICK: This customer has the highest number of orders 86 and has spent the most money in total 110277. They also placed an order recently within 22 days.
BONAP: This customer has the highest number of orders 44 and has spent the most money in total 21963. They also placed an order recently within 0 day.

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

> **Customer Total spend and total Quantity in Each country**
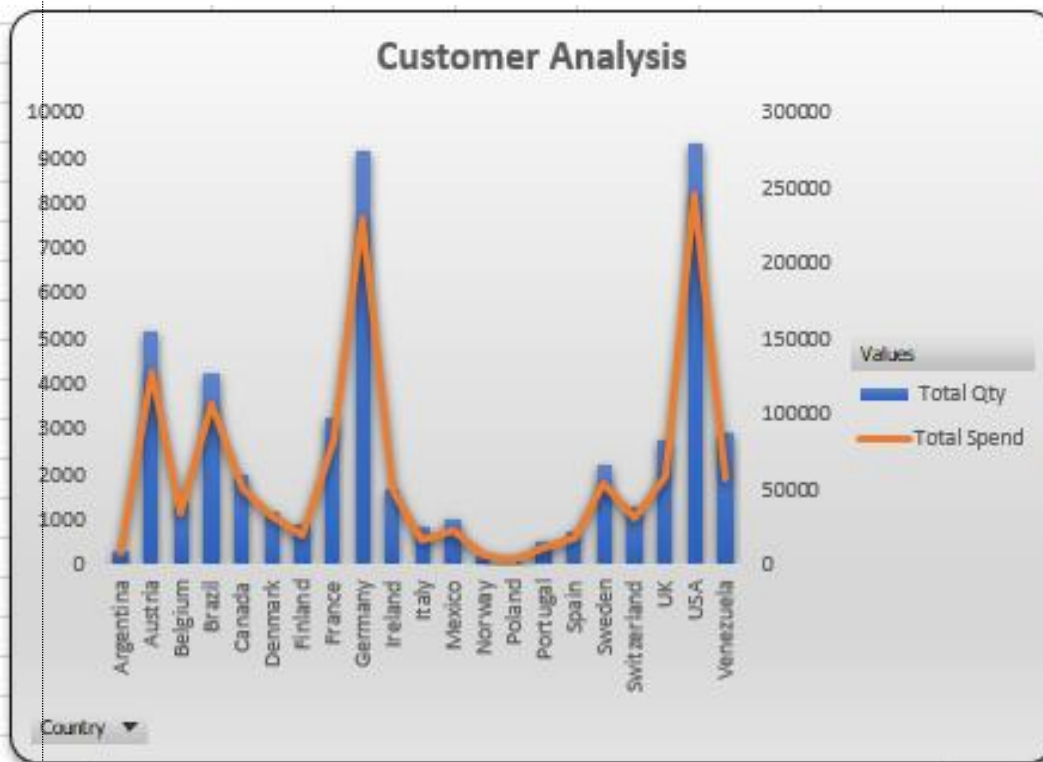
Row Labels	Total Qty	Total Spend
Argentina	319	7839.1
Austria	5167	128003.838
Belgium	1392	33824.855
Brazil	4237	106813.776
Canada	1984	50196.2899
Denmark	1170	32661.0225
Finland	885	18810.0525
France	3244	81246.3224
Germany	9153	229684.633
Ireland	1684	49979.9049
Italy	822	15770.155
Mexico	1005	23332.0775
Norway	161	5735.15
Poland	205	3531.95
Portugal	533	11472.3625
Spain	718	17983.2
Sweden	2235	54495.14
Switzerland	1275	31692.659
UK	2742	58971.31
USA	9330	245584.61
Venezuela	2936	56810.6289
Grand Total	**51197**	**1264439.04**

```
create view Analyzing_Product_Preferences as
SELECT  distinct c.customerID, c.Country, SUM(od.Quantity) AS TotalQty
, SUM(od.Quantity * od.UnitPrice * (1-od.Discount) )AS TotalSpend
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY c.customerID, c.Country, p.ProductName
ORDER BY c.Country, TotalQty DESC;
```

This query helps to check the total spend and Total Quantity purchased in each country. By this we can identify the top country that contribute most in increasing the revenue. USA have highest Total spend around $245584.

# ➢ Total Customers for each Product in Different Countries


```
create view customer_Preference as
SELECT c.Country, p.ProductName, COUNT(*) AS CustomerCount
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY c.Country, p.ProductName
ORDER BY  CustomerCount DESC;
```

ProductName Alice Mutton	
Row Labels	Customer Preference
Austria	6
Belgium	1
Brazil	2
Canada	3
France	4
Germany	1
Italy	1
Mexico	3
Spain	3
Sweden	1
UK	1
USA	11
Grand Total	37

This query provides a list of countries and the top-selling product based on the number of unique customers who purchased it in each country. It focuses on customer preference rather than total quantity.
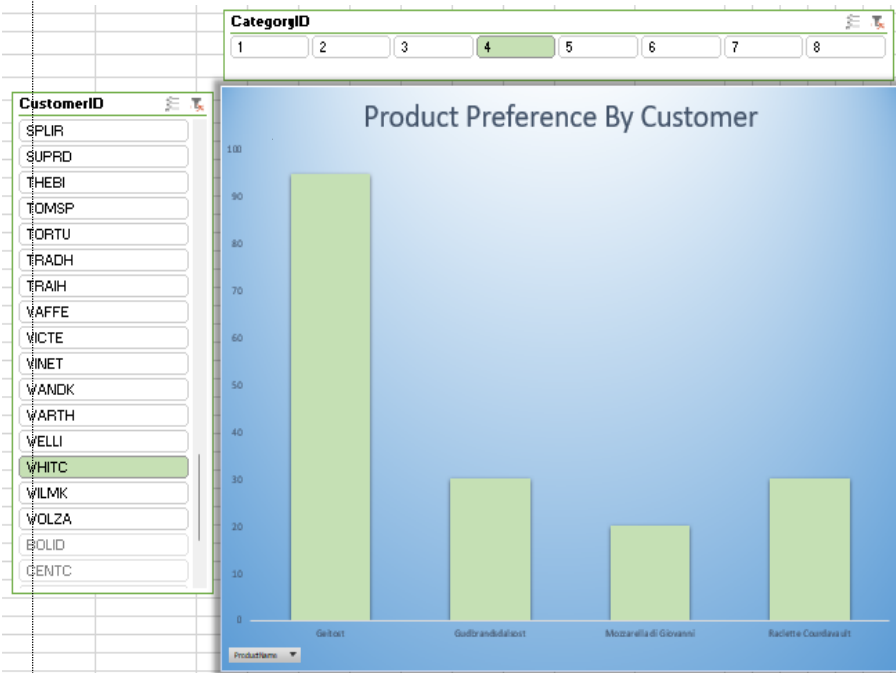
3. Are there any interesting patterns or clusters in customer behavior that can be visualized to identify potential market segments?
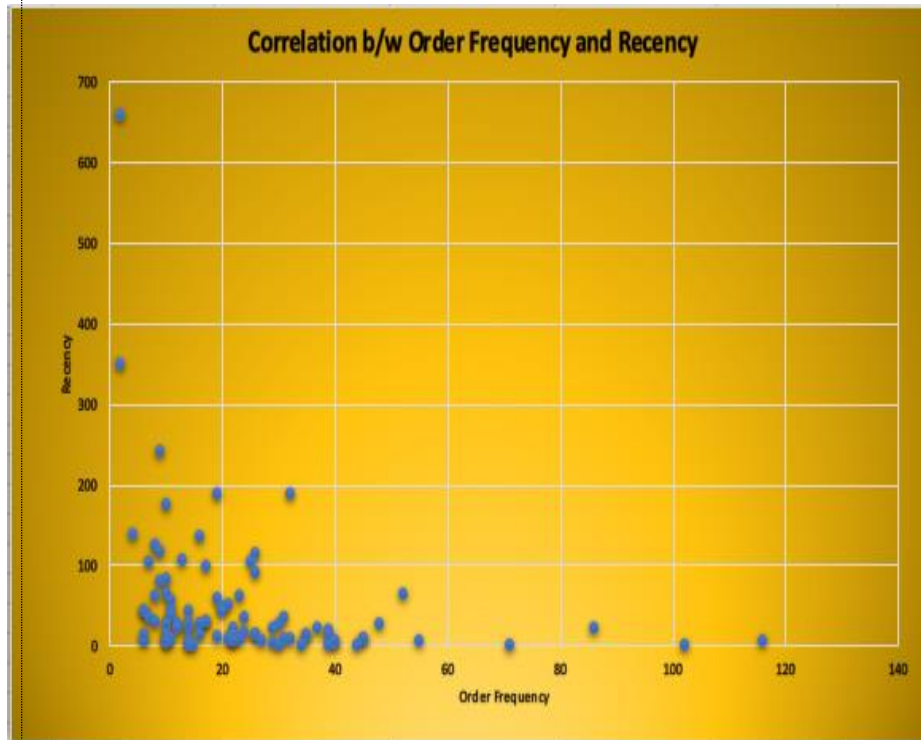
Row Labels	Sum of Frequency	Sum of Recency	Sum of Monetary
ALFKI	12	27	4273
ANATR	10	63	1402.95
ANTON	17	98	7023.98
AROUT	30	26	13390.65
BERGS	52	63	24927.58
BLAUS	14	7	3239.8
BLONP	26	114	18534.08
BOLID	6	43	4232.85
BONAP	44	0	21963.25
BOTTM	35	12	20801.6
BSBEV	22	22	6089.9
CACTU	11	8	1814.8
CENTC	2	657	100.8
CHOPS	22	14	12348.88
COMMI	10	14	3810.75
CONSH	7	103	1719.1
DRACD	10	2	3763.21
DUMON	9	79	1615.9
EASTC	21	8	14761.04
ERNSH	102	1	104874.98
FAMIA	19	187	4107.55
FOLIC	16	125	11655.0

Question 1   question 2   **Question 3**   question 4   Questi

This provide the Correlation b/w Recency and order Frequency of customer from this we can identify the customer behavior and purchasing pattern. Recency show the number of days between the customer's latest purchase . It provides a straightforward understanding of how recently a customer has engaged in buying activity.

➢ Product Preferences by customers

CategoryID

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

CustomerID

| SPLIR |
| SUPRD |
| THEBI |
| TOMSP |
| TORTU |
| TRADH |
| TRAIH |
| VAFFE |
| VICTE |
| VINET |
| WANDK |
| WARTH |
| WELLI |
| WHITC |
| WILMK |
| WOLZA |
| BOLID |
| CENTC |

Product Preference By Customer

Row Labels	Sum of TotalQty
Geitost	95
Gudbrandsdalsost	30
Mozzarella di Giovanni	20
Raclette Courdavault	30
Grand Total	**175**

```
create view Product_category_preference as
SELECT c.CustomerID, c.CompanyName, p.CategoryID,p.ProductName, SUM(od.Quantity) AS TotalQty
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY c.CustomerID, c.CompanyName, p.CategoryID, p.ProductName
ORDER BY c.CustomerID, TotalQty DESC;
```

This query provides the preference of customer for different Category. This show that each Customer Total Quantity From each Product.

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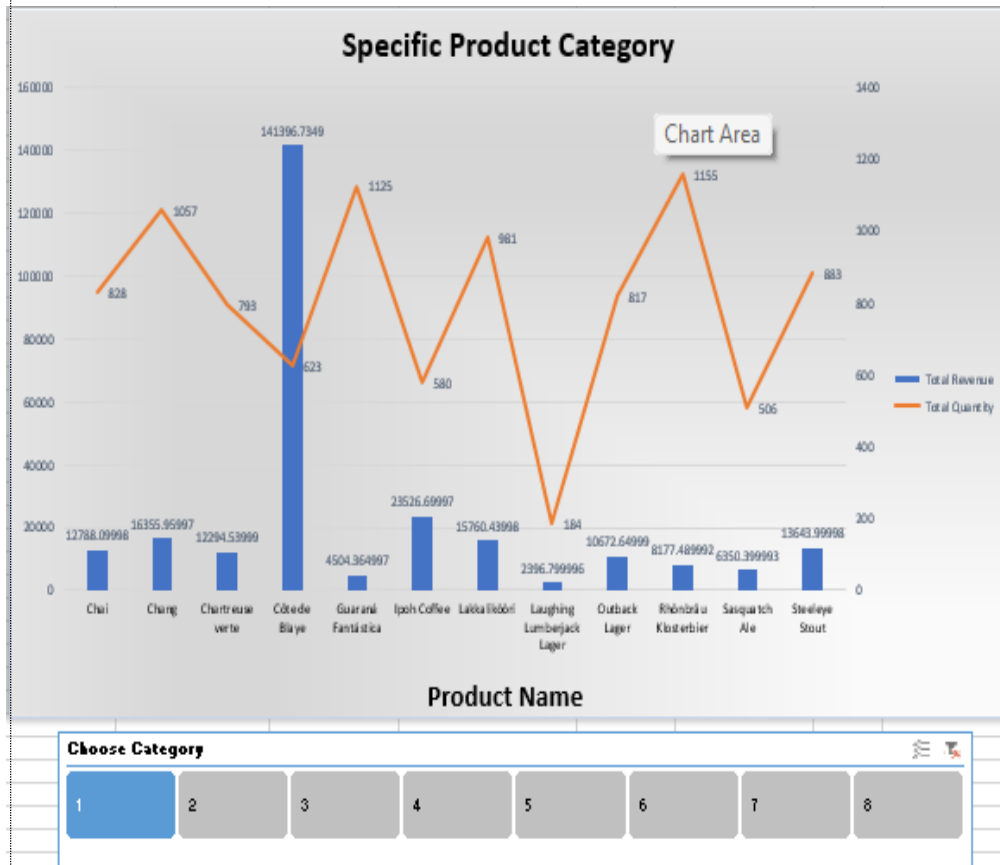

Row Labels	▼	Total Revenue	Total Quantity
Chai		12788.10	828
Chang		16355.96	1057
Chartreuse verte		12294.54	793
Côte de Blaye		141396.73	623
Guaraná Fantástica		4504.36	1125
Ipoh Coffee		23526.70	580
Lakkalikööri		15760.44	981
Laughing Lumberjack Lager		2396.80	184
Outback Lager		10672.65	817
Rhönbräu Klosterbier		8177.49	1155
Sasquatch Ale		6350.40	506
Steeleye Stout		13644.00	883
Grand Total		**267868.18**	**9532**

```
create view specific_product_categories as
SELECT p.CategoryID, p.ProductName,
       SUM(od.Quantity * od.UnitPrice) AS TotalRevenue,
       SUM(od.Quantity) AS TotalQuantity
FROM Products p
INNER JOIN northwind.`order details` od ON p.ProductID = od.ProductID
INNER JOIN Orders o ON od.OrderID = o.OrderID
GROUP BY p.CategoryID, p.ProductName
ORDER BY TotalRevenue DESC;
```

This provides total revenue and Total Quantity of each product for a specific category and in the chart secondary axis show the total Quantity and bar show the total revenue from each product. This help us to compare the total sale of each product of different category.

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

```
create view correlations_order_size_customer_demographics as
SELECT c.Country, p.CategoryID,
case when c.country='Argentina' then 1  when c.country='Austria' then 2  when c.country="Belgium" then 3
when c.country="Brazil" then 4 when c.country="Canada" then 5 when c.country="Denmark" then 6
when c.country="Finland" then 7 when c.country="France" then 8  when c.country="Germany" then 9
when c.country="Ireland" then 10 when c.country="Italy" then 11 when c.country="Mexico" then 12
when c.country="Norway" then 13 when c.country="Poland" then 14 when c.country="Portugal" then 15
when c.country="Spain" then 16  when c.country="Sweden" then 17 when c.country="Switzerland" then 18
when c.country="UK" then 19  when c.country="USA" then 20 when c.country="Venezuela" then 21 end as countrycode,
SUM(od.Quantity) AS OrderSize
FROM Customers c
INNER JOIN Orders o ON c.CustomerID = o.CustomerID
INNER JOIN northwind.`order details` od ON o.OrderID = od.OrderID
INNER JOIN Products p ON od.ProductID = p.ProductID
GROUP BY 1,2,3
```


Correlation b/w Country Code and OrderSize
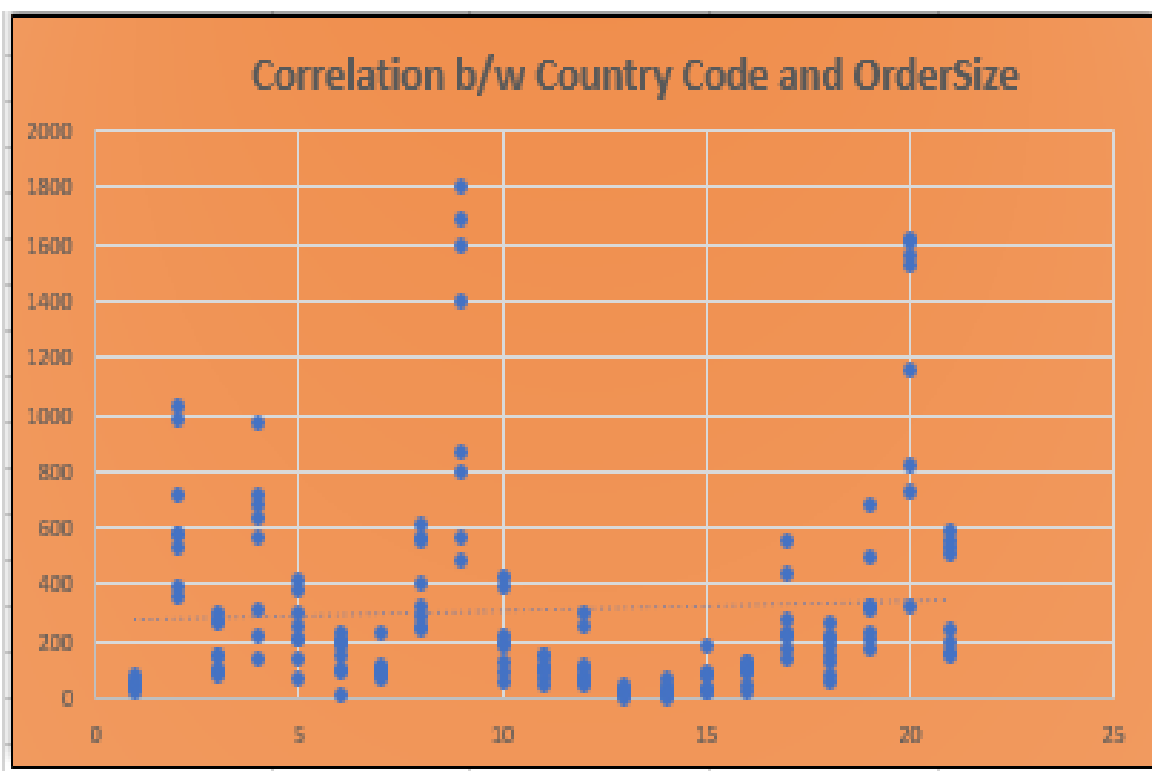
The correlation coefficient between country code and order size is 0. 0586.This indicates a very weak positive correlation. Based on this correlation, we cannot conclude that there's a significant relationship between country code and order size. Other factors are likely more influential in determining order size.

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

```
create view customer_segments as SELECT
   c.Country,
   c.City,
   COUNT(o.OrderID) AS NumberOfOrders
FROM Customers c
LEFT JOIN Orders o ON c.CustomerID = o.CustomerID
GROUP BY c.Country, c.City
ORDER BY c.Country, c.City
```

The United States and Germany are the two countries with the most orders. This could be due to a number of factors, such as the size of the population, the level of economic development, or the popularity of the product or service being sold. Some countries have relatively few orders. These include Portugal, Ireland, and Denmark.

Country	Orders
Argentina	16
Austria	40
Belgium	19
Brazil	83
Canada	30
Denmark	18
Finland	22
France	77
Germany	122
Ireland	19
Italy	28
Mexico	28
Norway	6
Poland	7
Portugal	13
Spain	23
Sweden	37
Switzerland	18
UK	56
USA	122
Venezuela	46

Order Frequency vary Across Differnent Country
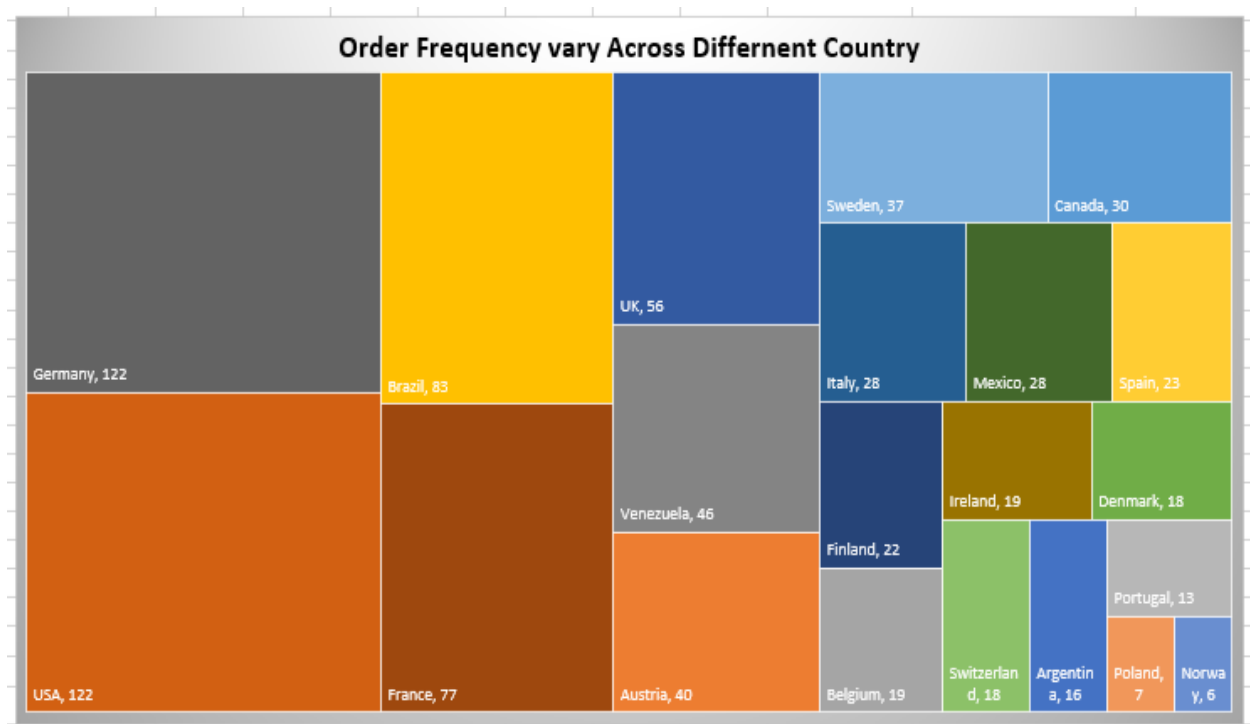
7. **There any correlations between employee satisfaction levels and key performance indicators? Can we explore this visually through scatter plots or line charts?**

Row Labels	Total Customers	Total sale	Satisfaction Level.
Andrew Fuller	59	166537.75	5
Anne Dodsworth	29	77308.07	1
Janet Leverling	63	202812.84	6
Laura Callahan	56	126862.28	4
Margaret Peacock	75	232890.85	8
Michael Suyama	43	73913.13	2
Nancy Davolio	65	192107.6	7
Robert King	45	124568.23	3
Steven Buchanan	29	68792.28	1

Correlation b/w Total Sale and Satisfaction Level

```
create view Emp_performance as

select e.EmployeeID,concat(e.FirstName," ",e.LastName) As EmployeeName,

count(distinct o.CustomerID) as Total_Customers ,

round(sum(od.unitprice * od.Quantity*(1-od.discount)),2) as Total_sale from employees e

inner join orders o on o.EmployeeID=e.EmployeeID

inner join `order details` od on od.OrderId=o.OrderID

group by 1,2
```

The correlation b/w Total sale and satisfaction level is 0.0979 and this show the weak correlation .Margaret Peacok (Employee ID 4) have most satisfaction level 8 and Total sale of 132890.85.

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

```
create view turnover as SELECT
e.Title as JobRole,
COUNT(e.EmployeeID) as TotalEmployees,
sum( CASE WHEN (timestampdiff(year,hiredate,maxD))<2
THEN 1 ELSE 0 END)
as NewHires
FROM Employees e
left join (select employeeId,max(orderdate) as maxD from orders
group by 1)m on m.EmployeeID=e.EmployeeID
GROUP BY 1
```

| JobRole	TotalEmployees	NewHires
Sales Representative	6	1
Vice President, Sales	1	0
Sales Manager	1	0
Inside Sales Coordinator	1	0

Employee Turnover acorss Different Job roles
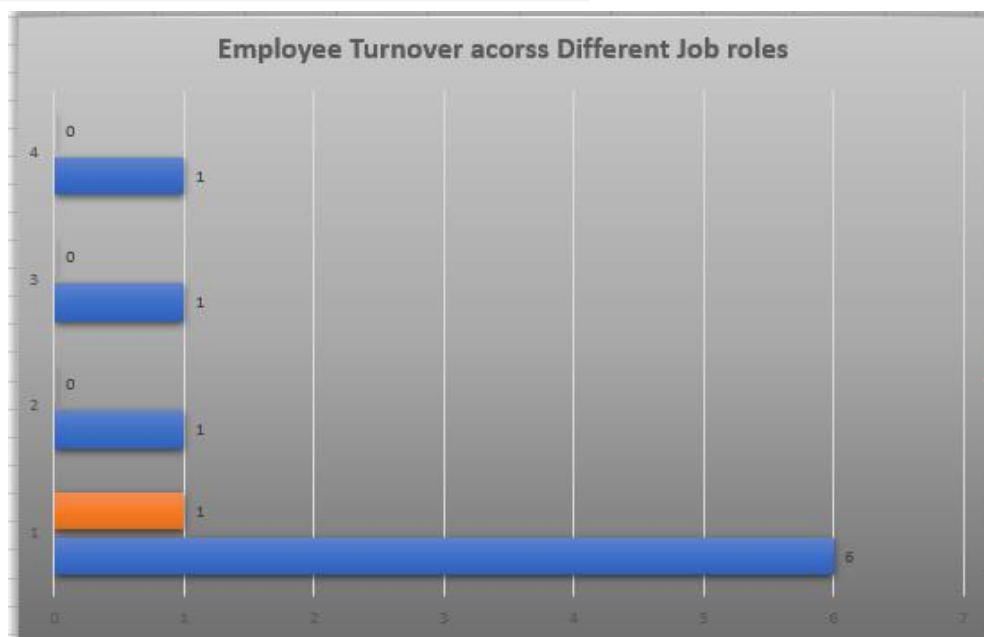
This query provides last 2-year hire data and visualization provide a simple yet effective way to analyze employee turnover across different role. This data indicates that the Sales Representative role likely involves high employee turnover and ongoing hiring given the disproportionately large headcount compared to management positions.

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

```
create view Emp_performance as
SELECT
e.employeeid,concat(e.firstName," ",e.LastName,",",e.title) as Full_Name_with_Title,
count(distinct o.OrderID) as total_tasks,timestampdiff(year,hiredate,maxD)as tenure,
round(sum(od.unitprice * od.Quantity*(1-od.discount)),2) as total_sale
FROM Employees e
left join (select employeeId,max(orderdate) as maxD from orders
group by 1)m on m.EmployeeID=e.EmployeeID
left join orders o on o.EmployeeID=e.EmployeeID
left join `order details` od on od.OrderId=o.OrderID
group by 1,2,4
```

employeeid	Full_Name_with_Title	total_tasks	tenure	total_sale
1	Nancy Davolio,Sales Representative	123	4	192107.6
2	Andrew Fuller,Vice President, Sales	96	3	166537.75
3	Janet Leverling,Sales Representative	127	4	202812.84
4	Margaret Peacock,Sales Representativ	156	3	232890.85
5	Steven Buchanan,Sales Manager	42	2	68792.28
6	Michael Suyama,Sales Representative	67	2	73913.13
7	Robert King,Sales Representative	72	2	124568.23
8	Laura Callahan,Inside Sales Coordinato	104	2	126862.28
9	Anne Dodsworth,Sales Representative	43	1	77308.07
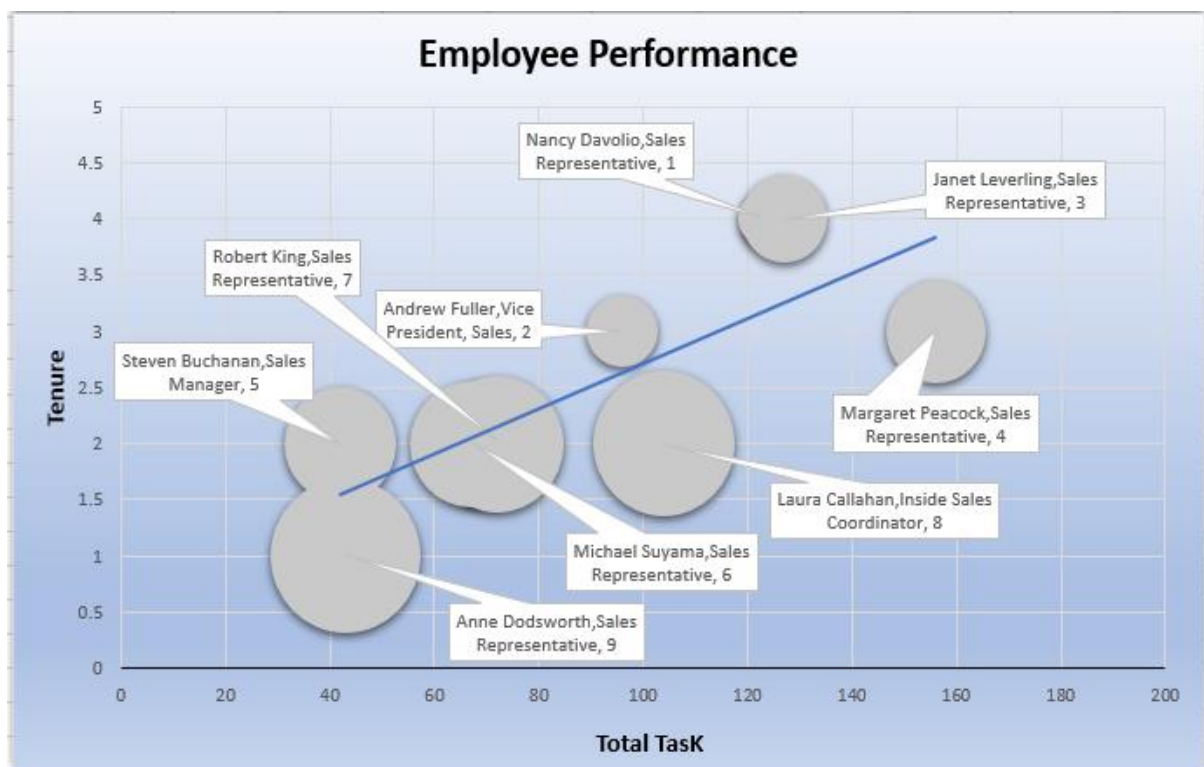
The sales representatives seem to be performing well, with an average of 99.8 total tasks completed and an average of $160,153.06 in total sales.

Sales representative number 4 is the highest performing salesperson, with 156 tasks completed and $232,890.85 in total sales. Sales representatives number 1 and 3 are also high performers, with 123 and 127 tasks completed, and $192,107.60 and $202,812.84 in total sales, respectively.
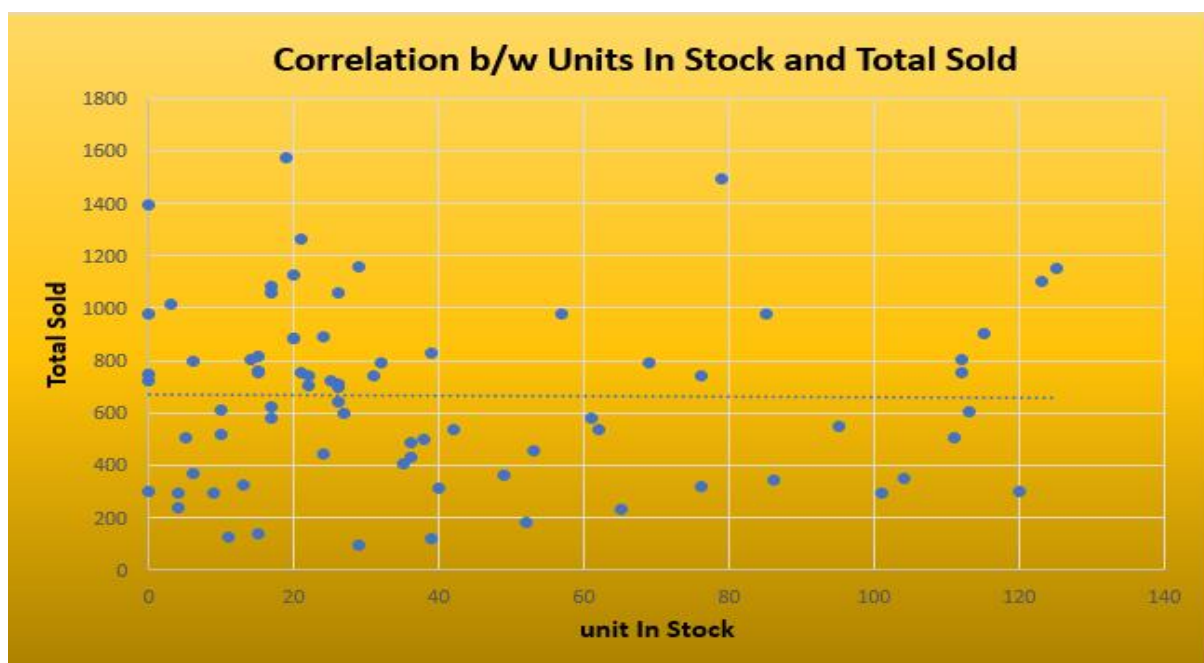
However, there is some variability in performance, with sales representative number 9 completing only 43 tasks and generating $77,308.07 in total sales.

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

```sql
create view prouduct_attribute as

SELECT p.ProductName, p.UnitsInStock, SUM(od.Quantity) AS TotalSold

FROM Products p

JOIN `Order Details` od ON p.ProductID = od.ProductID

GROUP BY 1,2

ORDER BY TotalSold DESC
```

ProductName	UnitsInStoc	TotalSold
Camembert Pierrot	19	1577
Raclette Courdavault	79	1496
Gorgonzola Telino	0	1397
Gnocchi di nonna Alice	21	1263
Pavlova	29	1158
Rhönbräu Klosterbier	125	1155
Guaraná Fantástica	20	1125
Boston Crab Meat	123	1103
Tarte au sucre	17	1083
Chang	17	1057
Fløtemysost	26	1057
Sir Rodney's Scones	3	1016
Jack's New England Cla	85	981
Lakkalikööri	57	981
Alice Mutton	0	978
Pâté chinois	115	903
Konbu	24	891
Manjimup Dried Apple	20	886
Steeleye Stout	20	883
Chai	39	828
Outback Lager	15	817
Mozzarella di Giovanni	14	806
Inlagd Sill	112	805
Scottish Longbreads	6	799
Chartreuse verte	69	793
Original Frankfurter gri	32	791
Uncle Bob's Organic Dr	15	763
Geitost	112	755

The data suggests that there is a Weak correlation between the number of units in stock and the total number of units sold. This means that more product units in stock tend to sell more units overall. This could be because these are more popular with customers, or because the store is not able to promote them more heavily.

Correlation b/w Units In Stock and Total Sold

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

```sql
create view product_demand as
SELECT
    date(o.OrderDate) ,
    p.ProductID,p.ProductName,
    SUM(Quantity) AS TotalQuantity
FROM Orders o
left JOIN `Order Details`od ON o.OrderID=od.OrderID
left join products as p on p.ProductID=od.ProductID
left join categories c on c.CategoryID=p.CategoryID
GROUP BY 1,2,3
ORDER BY 1,2,3
```

Row Labels	Sum of TotalQuantity
⊟ 1994	234
Aug	30
Sep	30
Oct	40
Nov	98
Dec	36
⊟ 1995	419
Feb	2
Mar	105
Jun	73
Jul	34
Aug	100
Oct	30
Nov	55
Dec	20
⊟ 1996	325
Jan	108
Feb	6
Mar	95
Apr	27
May	89
Grand Total	**978**

Overall demand increased in 1995: The total quantity of the product sold in 1995 (25,007) was almost three and a half times higher than the total quantity sold in 1994 (7,381). This suggests a significant increase in demand for the product over that period.

Demand fluctuated throughout the year: Even within years, demand fluctuated considerably. In 1994, December saw the highest demand (1,804 units), while August saw the lowest (1,462 units). This pattern is repeated in 1995, with November being the strongest month (2,657 units) and July being the weakest (1,635 units).
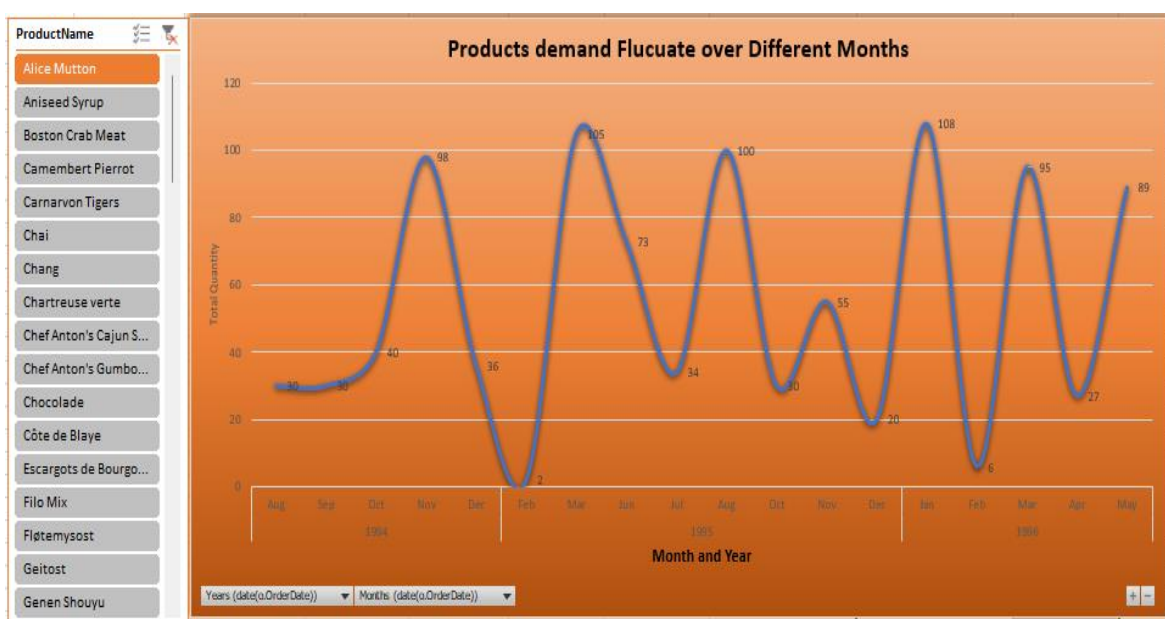
12.Are there any correlations between supplier attributes (e.g., location, size, industry) and performance metrics (e.g., on-time delivery, product quality)? Can we explore this visually through scatter plots or heatmaps?

```sql
create view supplier_performace as
select distinct
case when ContactTitle="Accounting Manager" then 1 when ContactTitle='Coordinator Foreign Markets' then 2
when ContactTitle='Export Administrator' then 3 when ContactTitle='International Marketing Mgr.' then 4
when ContactTitle='Marketing Manager' then 4 when ContactTitle='Marketing Representative'then 5
when ContactTitle='Order Administrator' then 6 when ContactTitle='Owner' then 7
when ContactTitle='Product Manager' then 8 when ContactTitle='Purchasing Manager' then 9
when ContactTitle='Regional Account Rep.' then 10  when ContactTitle='Sales Agent' then 11
when ContactTitle='Purchasing Manager'then 12  when ContactTitle='Sales Representative'then 13
when ContactTitle='Sales Manager'then 14 when ContactTitle='Wholesale Account Agent'then 15 end as ContactCode,
case when o.ShipCountry='Argentina' then 1  when o.ShipCountry='Austria' then 2  when o.ShipCountry="Belgium" then 3
when o.ShipCountry="Brazil" then 4 when o.ShipCountry="Canada" then 5 when o.ShipCountry="Denmark" then 6
when o.ShipCountry="Finland" then 7 when o.ShipCountry="France" then 8  when o.ShipCountry="Germany" then 9
when o.ShipCountry="Ireland" then 10 when o.ShipCountry="Italy" then 11 when o.ShipCountry="Mexico" then 12
when o.ShipCountry="Norway" then 13 when o.ShipCountry="Poland" then 14 when o.ShipCountry="Portugal" then 15
when o.ShipCountry="Spain" then 16  when o.ShipCountry="Sweden" then 17 when o.ShipCountry="Switzerland" then 18
when o.ShipCountry="UK" then 19  when o.ShipCountry="USA" then 20 when o.ShipCountry="Venezuela" then 21 end as countrycode ,
os.OrderID,Delivery_Time from suppliers s
left join products as p on p.SupplierID=s.SupplierID
left join 'order details' od on od.ProductID=p.ProductID
left join (select OrderID, avg(timestampdiff(day,OrderDate,Requireddate)) as Delivery_Time from orders
where orderDate is not null and ShippedDate is not null
group by 1) os on od.OrderID=os.OrderID
join orders o on o.OrderID=od.OrderID
where Delivery_Time is not null
```

ContactCode	countrycode	OrderID	Delivery_Time
3	8	10248	28
7	8	10248	28
13	8	10248	28
5	9	10249	42
13	9	10249	42
15	4	10250	28
13	4	10250	28
6	4	10250	28
11	8	10251	28
6	8	10251	28
13	3	10252	28
4	3	10252	28
13	4	10253	14
14	4	10253	14
8	4	10253	14
4	18	10254	28
9	18	10255	28
4	18	10255	28
13	18	10255	28
13	4	10256	28
4	4	10256	28
14	21	10257	28
4	21	10257	28

Correlation Supplier and Deliver Time
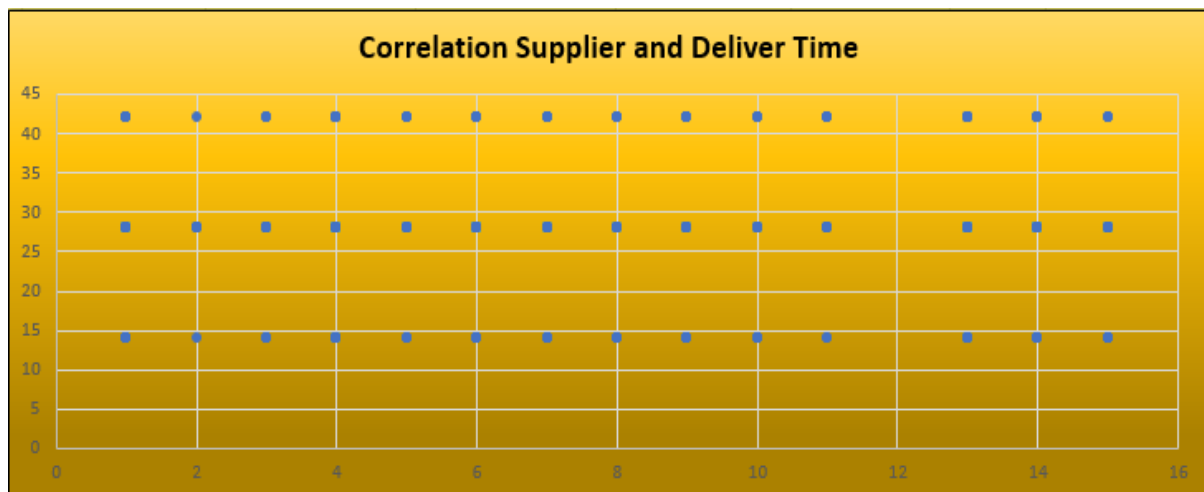
Correlation Coefficient (Contact Code vs. delivery Time):

The off-diagonal value of 0.001354 is the correlation coefficient between Contact Code and delivery Time. A correlation coefficient close to zero indicates a weak or negligible linear relationship between the two variables. In this case, the correlation coefficient is negative, but the magnitude is very small.

> Correlation Between Ship Country/Country code and Delivery Time

Correlation Coefficient (country code vs delivery time) is 0.01319.

The positive sign indicates a positive correlation between country code and delivery Time. However, the correlation coefficient is relatively small, suggesting a weak positive linear relationship.

14 . How does supplier performance vary across different product categories or departments? Can we visualize this using stacked bar charts or grouped column charts?
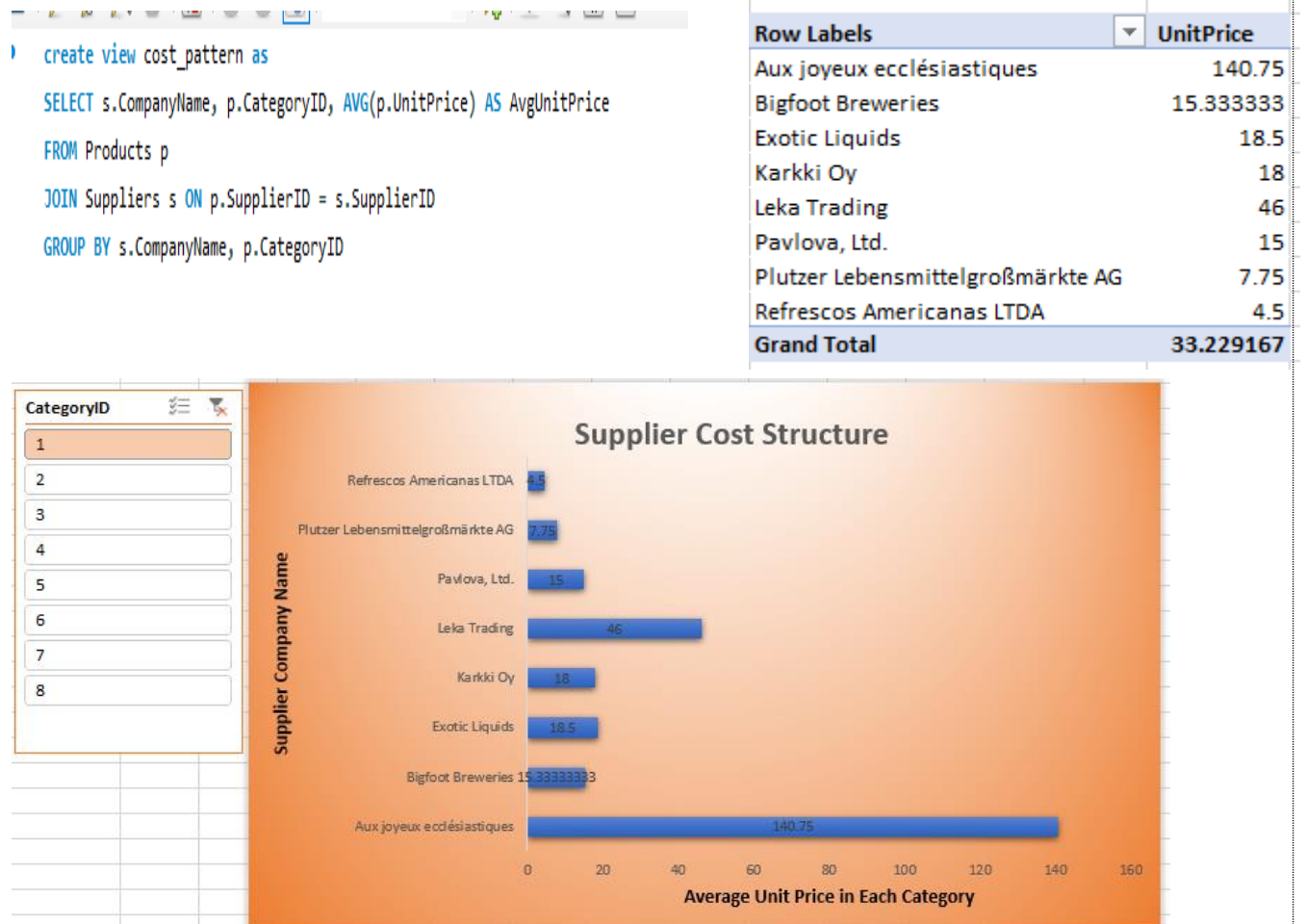
```sql
create view supply_performance as
SELECT
distinct s.supplierID,s.ContactTitle ,c.CategoryName, SUM(od.Quantity) AS TotalQuantity,
round(AVG(timestampdiff(day, OrderDate, Requireddate)),2) AS Delivery_time FROM suppliers s
LEFT JOIN products AS p ON p.SupplierID = s.SupplierID
LEFT JOIN `order details` od ON od.ProductID = p.ProductID
LEFT JOIN orders o ON o.OrderID = od.OrderID
LEFT JOIN categories c ON c.CategoryID = p.CategoryID
GROUP BY s.supplierID, s.ContactName, c.categoryID, c.CategoryName
HAVING Delivery_time IS NOT NULL
```

Row Labels ▼	Sum of Total_Unit_on_order	Average of Delivery_time
1	1760	27.83
7	390	28
10	0	27.18
12	0	28.91
16	0	27.57
18	0	27.22
20	280	28.5
23	0	27.28
Grand Total	**2430**	**27.81125**
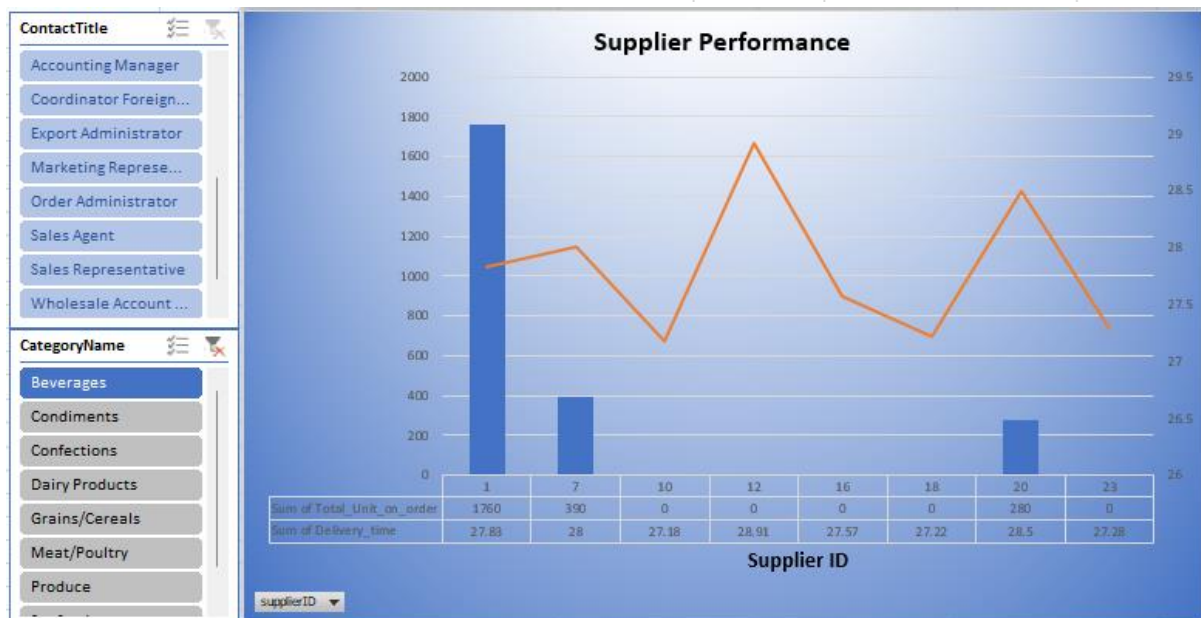
This query provide the supplier ID ,contact tile, category name ,total Unit on order, average delivery time by suppliers . For Category Beverages total unit on order is 1760 and average delivery time by the supplier id 1 is 27 days and supplier id 2 provide average deliver time of 28 for 390 total unit on order and supplier id 20 provide average deliver time of 28.5 days for 280 units . By this we can compare the average delivery time provide by the suppliers to check the supplier performance for different category and Job Roles.

15. Can we identify any trends or patterns in supplier costs or pricing structures through visualizations? How can this information be used for procurement optimization?

```
create view cost_pattern as
SELECT s.CompanyName, p.CategoryID, AVG(p.UnitPrice) AS AvgUnitPrice
FROM Products p
JOIN Suppliers s ON p.SupplierID = s.SupplierID
GROUP BY s.CompanyName, p.CategoryID
```

Row Labels	UnitPrice
Aux joyeux ecclésiastiques	140.75
Bigfoot Breweries	15.333333
Exotic Liquids	18.5
Karkki Oy	18
Leka Trading	46
Pavlova, Ltd.	15
Plutzer Lebensmittelgroßmärkte AG	7.75
Refrescos Americanas LTDA	4.5
Grand Total	33.229167

Suppliers with higher vs lower average pricing overall. Categories where certain suppliers have exceptionally high prices. Comparing pricing between suppliers for the same category .
Based on Analysis: Negotiate lower pricing in categories where a supplier's prices are consistently high Switch suppliers for certain categories to reduce costs.

Conclusion:

By providing a visually compelling and user-friendly dashboard with interactive features, this Power BI report empowers stakeholders at Northwind Traders to make data-driven decisions. This comprehensive analysis of customers, sales, inventory, and employees offers valuable insights into business operations and facilitates data exploration. The expected impact is a transformation in how Northwind interacts with its data, ultimately enhancing competitiveness and driving the company forward in the wholesale market.