Capstone project-

Sales Analysis

"Revolutionizing Data-Driven Decision Making at Northwind Traders"

-Balam Gowthami

Overview

**The Northwind Traders project involves the development of a comprehensive Power BI dashboard to analyse and visualize key performance metrics within the context of a fictitious wholesale company named Northwind Traders. The overarching goal is to provide stakeholders with valuable insights into various aspects of the company's operations, facilitating informed decision-making processes. The project focuses on leveraging the Northwind database, a dataset containing sales data, to extract meaningful information related to customer behaviour, sales patterns, inventory management, and employee performance**.

**Expected Impact:**

**The implementation of the Power BI dashboard is anticipated to revolutionize how Northwind Traders interacts with its data. The visualizations and interactive features will empower stakeholders to explore and understand complex datasets easily. This, in turn, will enable the company to remain competitive and make strategic decisions based on a comprehensive understanding of its performance metrics.**

**The Process**

**1. Introduction:**

Data Exploration and Analysis (EDA) is a crucial step in understanding and deriving insights from data. This document outlines a comprehensive process of extracting data from a GitHub repository, specifically focusing on SQL and Excel files. The extracted data will be transformed and modeled using Power BI, culminating in the creation of insightful visualizations.

**2. Data Source:GitHub Repository:**

**The GitHub repository chosen for this analysis is[https://github.com/acciojob-data-analytics/NorthWind]. It contains the type of data, e.g., SQL and Excel files that we aim to explore and analyze.**

**3.Connecting with tools:**

**Examine the structure and content of the SQL and Excel files present in the repository. Understand the relationships between different tables in SQL files and the sheets in Excel files. Interface the dataset with Excel, Power bi files and Mysql Workbench.**

**4.Data Transformation in Power BI:**

**Utilize Power Query Editor within Power BI for data cleaning and preprocessing. Document each transformation step, including filtering, renaming columns, handling missing values, and any other necessary data cleansing operations.**

**5. Data Modelling in Power BI**

**Design a robust data model that aligns with the business logic. Establish relationships between different tables, define calculated columns, and consider the hierarchy of data for optimal visualization.**

**6. Visualizations in Power BI:**

**Create a variety of visuals using Power BI's visualization options. Choose appropriate chart types for different insights, customize visuals for clarity, and ensure that the visuals align with the analysis goals.**

**7. Testing and Validation in EDA:**

**Implement strategies to test and validate the accuracy and reliability of the data. Cross-verify visuals with the source data to ensure that the insights derived are accurate. Highlight any limitations encountered during the process, such as data quality issues or constraints imposed by the data source. Provide considerations for future improvements or modifications.**

**8. Documentation:**

**Document each step of data transformation and modelling to provide a clear understanding of the decisions made during the process. Include explanations for data shaping choices and any compromises or assumptions. Summarize key findings, insights, and the impact of the EDA process on decision-making. Offer recommendations for next steps based on the analysis.**

**Data Description:**

**The dataset used for this project is the Northwind database, containing comprehensive sales data for Northwind Traders. The dataset includes tables such as Customers, Employees, Orders, Order Details, Products, Suppliers, Shippers, and Categories, each providing crucial information about different aspects of the company's operations.**

**Table Contents:**

**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.**

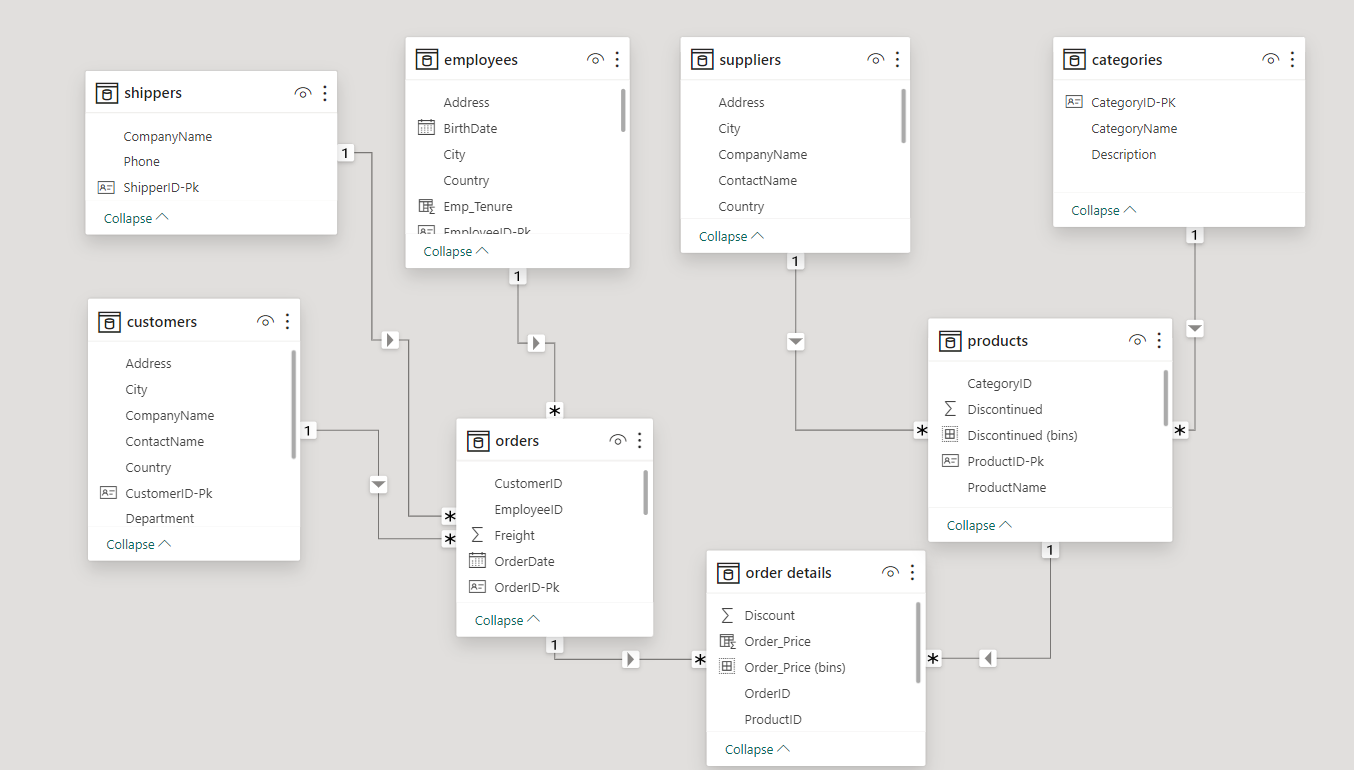
**Shippers Table:**

**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.**

ER Diagram:

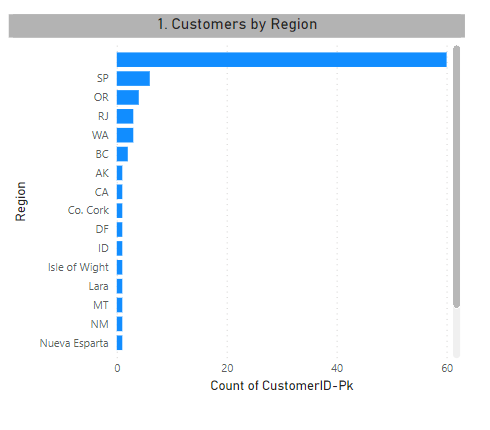


Power BI Visuals:

1.How does customer distribution vary across different regions or customer segments?

**We present a customer distribution to gain insights into regional and segment-based variations. we** initiate the analysis with an overview of the overall customer distribution, summarizing the total number of customers and identifying any discernible trends or patterns. This initial assessment provides a foundation for more detailed investigations.

Based on our analysis, we propose actionable recommendations, such as tailoring marketing campaigns to specific regions or segments and implementing customer retention strategies in areas with declining engagement. These recommendations are designed to enhance customer satisfaction and foster sustained business growth.

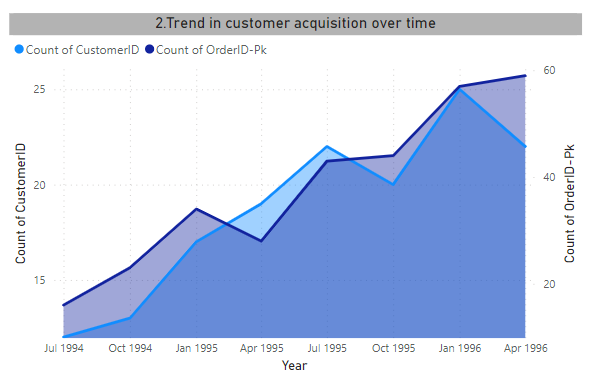


**Future Considerations:**

Recognizing the dynamic nature of markets, we highlight potential changes in customer distribution that may occur in the future.

2. What is the trend in customer acquisition over time?

Assess customer retention by examining whether existing customers are consistently placing orders or if there's a high rate of one-time orders.High customer retention suggests a loyal customer base, while low retention may indicate a need for improved customer engagement strategies.



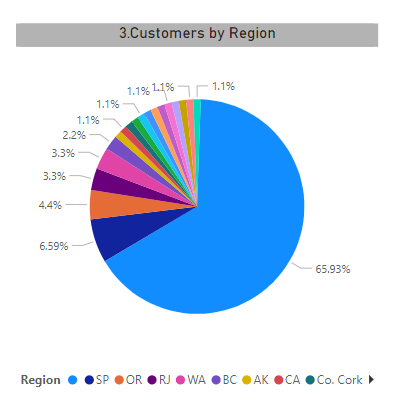
**Future Considerations:**

Consider what changes or events might impact customer acquisition in the future. Outline potential areas of focus for ongoing analysis and adaptation of strategies.

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

Pie charts work best when illustrating the parts of a whole. Identify which regions have the highest percentage of customers. This can help in focusing marketing efforts or tailoring products/services to the needs of the dominant regions. Understanding these disparities can be crucial for regional targeting or expansion strategies.

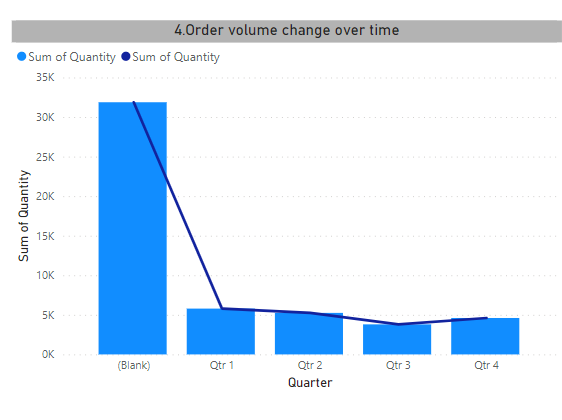
Assess whether certain regions are approaching market saturation in terms of customer acquisition. This information can be essential for strategic planning and resource allocation.



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

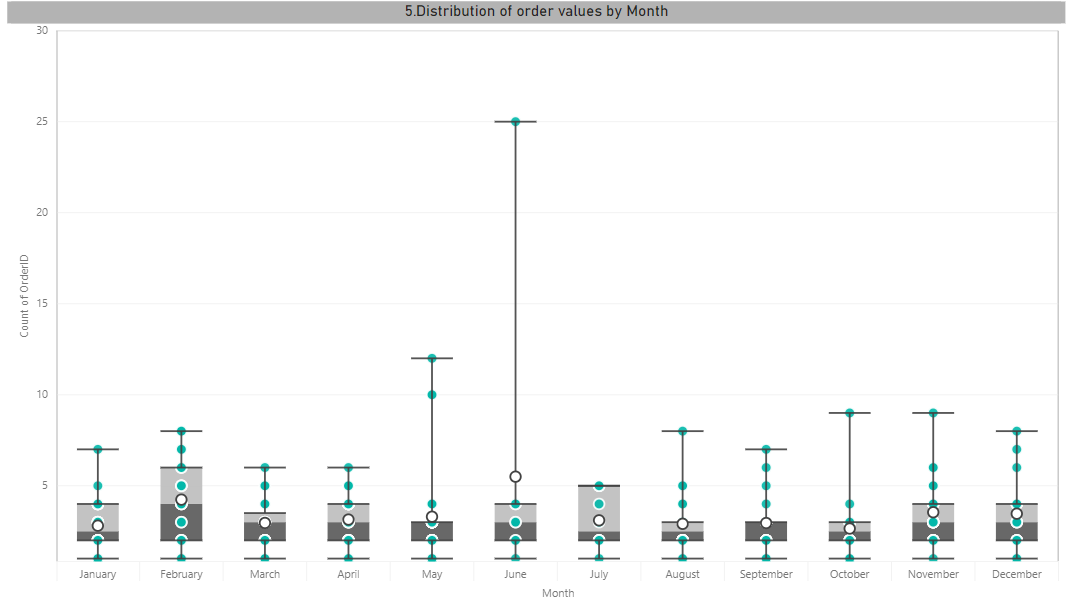
**changes in order volume over time using a time series chart is a valuable way to analyze trends and patterns in your business. Look for recurring patterns or seasonality in the data, such as higher order volume during certain months or days of the week.**

Identify the periods of highest and lowest order volumes.Assess the impact of promotions, marketing campaigns, or other events on order volume.Use historical trends to make informed predictions about future order volumes.



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

A box plot typically displays the distribution of a dataset by showing the median, quartiles, and potential outliers.



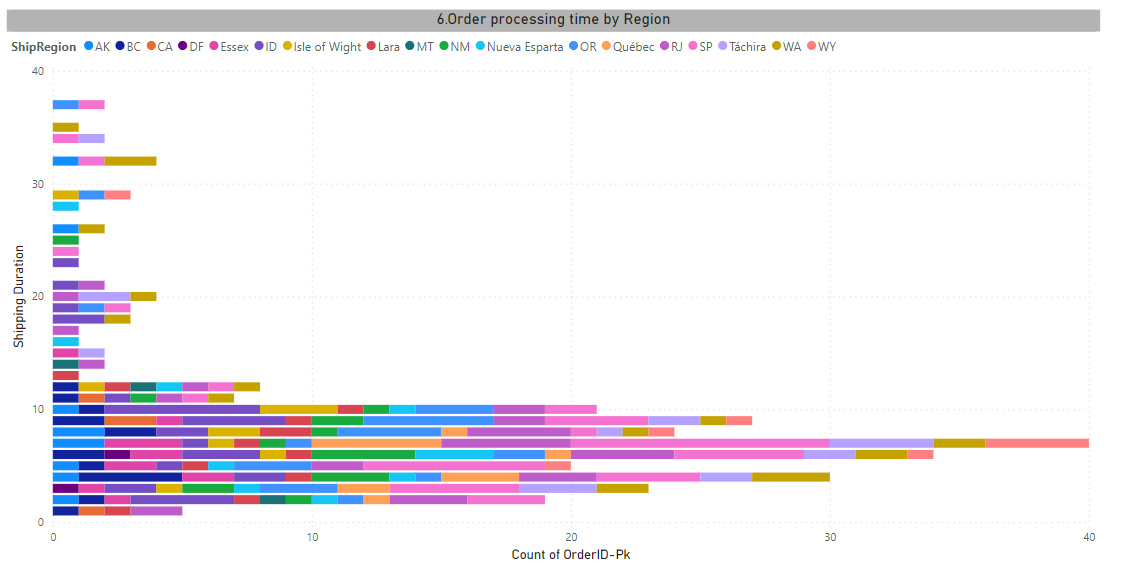
**Median and Interquartile Range (IQR):**

If the box is skewed to one side, it indicates that a significant portion of the data lies in that direction. If the median is towards the lower quartile, it suggests that the majority of order values are on the lower side. If the median is towards the upper quartile, it suggests that the majority of order values are on the higher side.

**Outliers:**

Outliers, points beyond the "whiskers" of the box plot, may indicate extreme values in the order data. If there are outliers on the higher side, it could suggest some unusually large orders. If there are outliers on the lower side, it could indicate some unusually small orders.

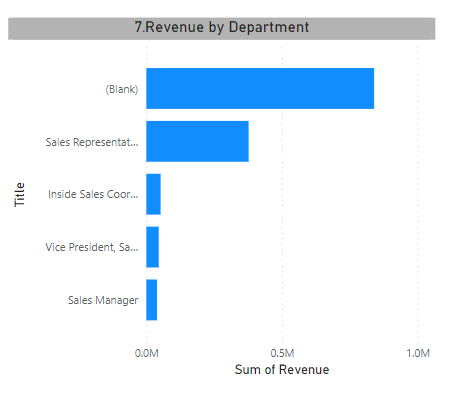
6. Can we visualize the average order processing time or shipping duration using a bar chart or box plot?



We will explore the use of a bar chart to visually represent the average order processing time and shipping duration for a given set of data.We provide recommendations or actions based on the insights gained from the visualization. For example, if certain products consistently have longer shipping durations, consider optimizing the logistics for those items.

The key findings and insights derived from the bar chart, emphasizing the importance of monitoring and improving order processing and shipping efficiency for enhanced customer satisfaction.

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?

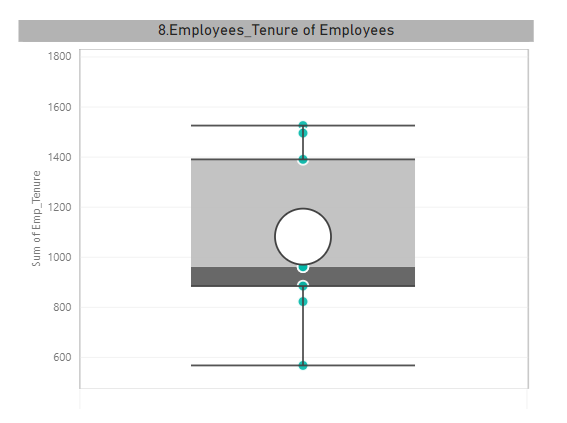


Identify and categorize the data based on departments or job roles. Analyze the chart to identify patterns, trends, and variations in productivity across different departments or job roles.Any key observations, such as high-performing or underperforming areas are noticed.

Discuss potential implications of the productivity variations and suggest recommendations for improvement. This could include resource reallocation, training programs, or process optimizations.

Emphasize the importance of regularly updating the data and the chart to reflect the latest information. This ensures that decision-making is based on current and relevant data.

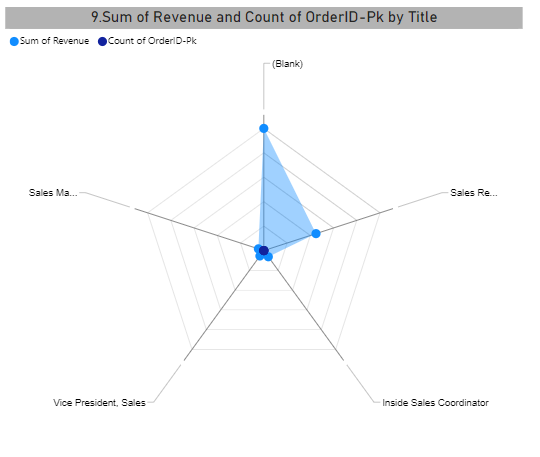
8. What is the distribution of employee tenure? Can we create a histogram or box plot to display it?



Employee tenure is a crucial metric in understanding the stability and longevity of the workforce within an organization. A box plot is an effective visualization tool that displays the distribution of employee tenure, highlighting key statistical measures such as the median, quartiles, and potential outliers.

Consider updating the analysis regularly to track changes in employee tenure distribution over time. Additionally, explore correlations with other organizational metrics to gain a holistic understanding of workforce dynamics.

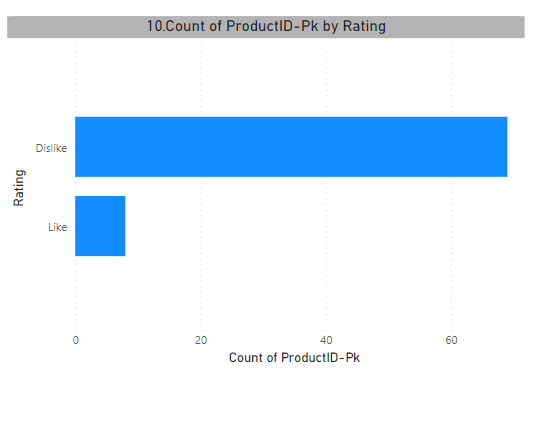
9. Can we visualize employee performance ratings or KPIs using a radar chart or bullet graph?



Visualizing employee performance ratings or Key Performance Indicators (KPIs) using a radar chart can be an effective way to showcase multiple performance metrics for each employee. A radar chart, also known as a spider chart or star chart, is a two-dimensional chart that displays multivariate data in the form of a spider web or radar.

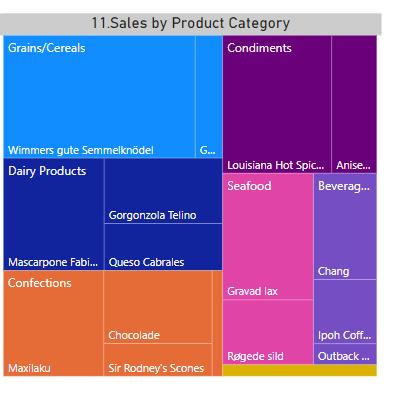
Explain how to interpret the radar chart. The distance from the center to each point on the chart represents the performance score for a specific KPI. A larger area enclosed by the shape indicates better overall performance.

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



Visualizing the distribution of product ratings or reviews using a stacked bar chart is a great way to understand how customers perceive a product across different rating categories. Focus on lower-rated segments to understand areas where the product can be improved.

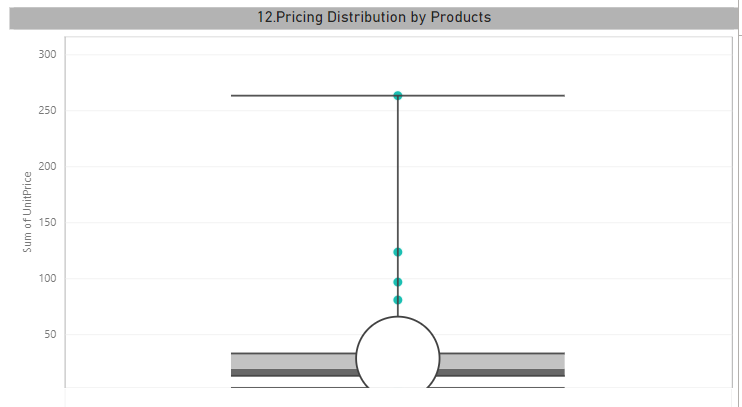
11. How does the sales volume vary across different product categories? Can we create a bar chart or tree map to display it?



A Tree map is a visual representation of hierarchical data that displays quantities in nested rectangles. Each branch of the hierarchy is given a coloured rectangle, and its size corresponds to a certain metric, such as sales volume in your case.

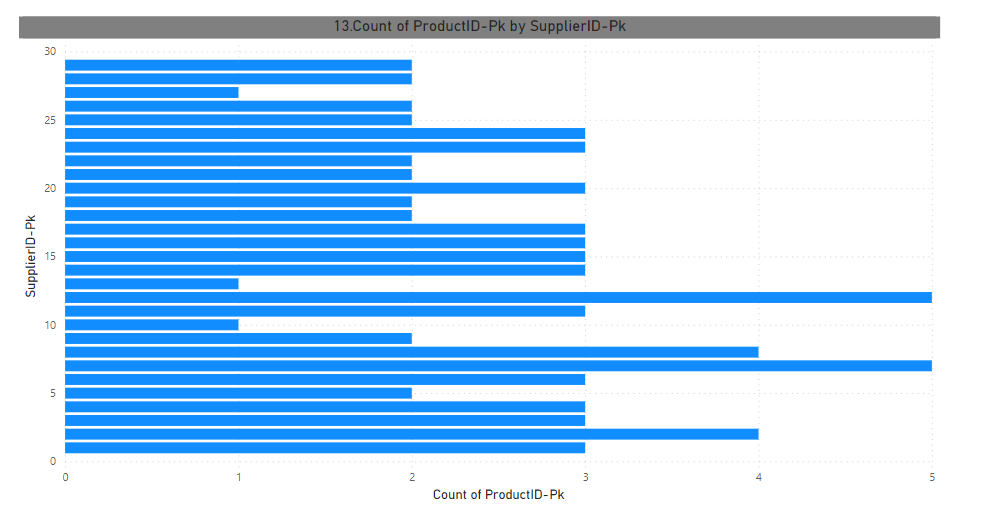
Larger rectangles represent higher sales volumes within their respective categories. Colors help differentiate between categories, making it visually intuitive. Patterns or clusters of rectangles may reveal insights, such as which categories contribute the most to overall sale

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



Visualizing the pricing distribution of products using a box plot is an effective way to understand the central tendency, spread, and potential outliers in the data. A box plot may reveal that most products have prices concentrated within a certain range (the interquartile range), with a few higher-priced outliers.

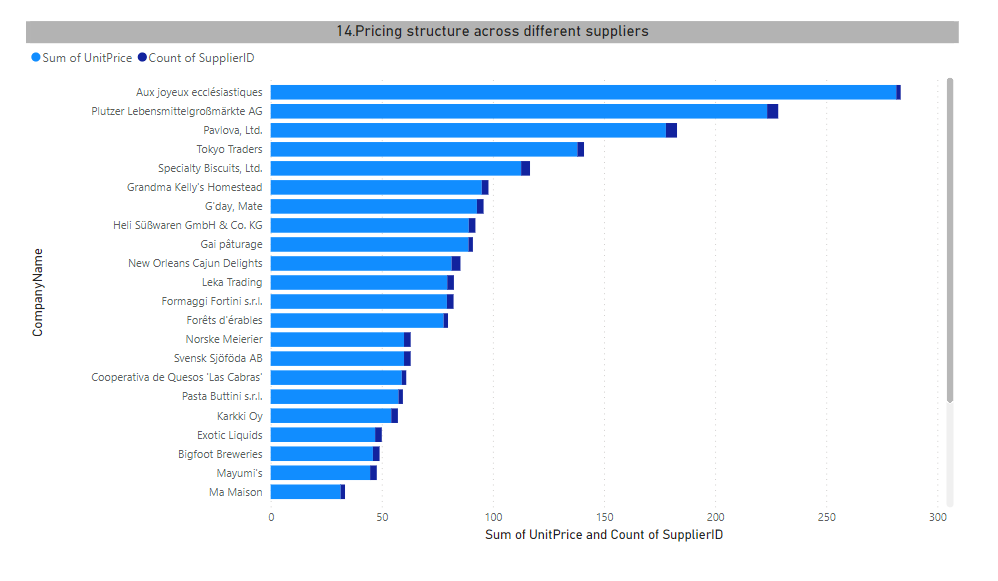
13. What is the distribution of supplier ratings or performance metrics? Can we create a bar chart or radar chart to visualize it?



A bar chart is a visual representation of data where individual bars represent different categories or groups. The length of each bar corresponds to the value it represents. The chart allows for quick identification of trends, such as which categories have the most suppliers or the highest average performance.

In summary, a bar chart is an effective and straightforward way to visualize the distribution of supplier ratings or performance metrics, providing a quick overview of how suppliers are performing across different categories or criteria.

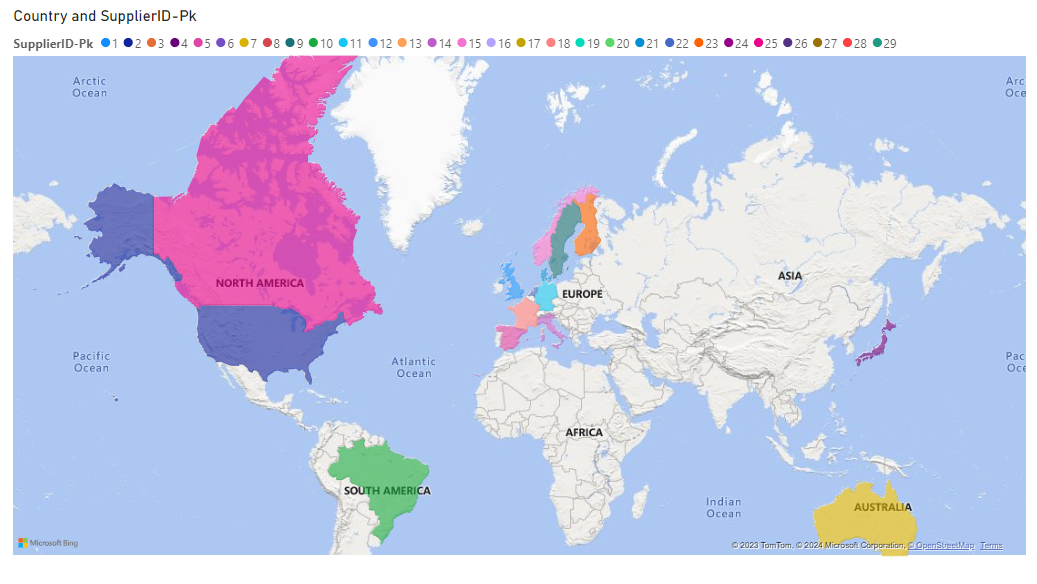
14. How does the cost or pricing structure vary across different suppliers? Can we create a box plot or stacked bar chart to display it?



The stacked bar chart provides a visual breakdown of the cost or pricing structure across different suppliers. Each segment of the stacked bar represents a specific cost component, and the overall height of the bar represents the total cost or price.It allows for quick identification of which cost components contribute the most to the total cost for each supplier.

In the stacked bar chart, you might observe that for one supplier, the majority of the cost comes from raw materials, while for another supplier, labor costs are a significant portion. This visual representation helps in understanding the cost composition for each supplier.

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

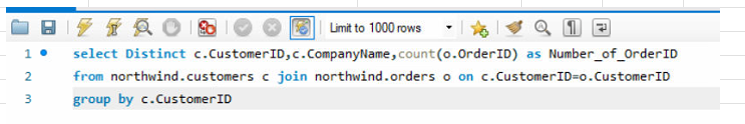


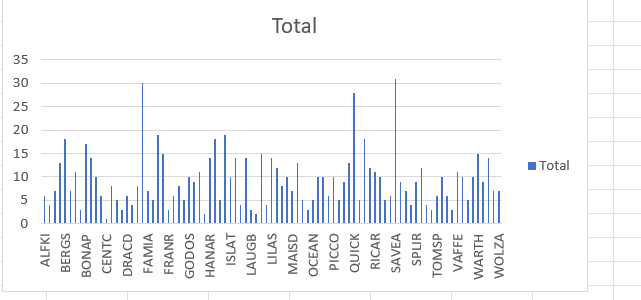
Absolutely! Visualizing the geographical distribution of suppliers using a map is a powerful way to gain insights into the global or regional spread of your suppliers.

The map visually displays the geographical locations of your suppliers. Patterns or clusters on the map may indicate concentrations of suppliers in specific regions. A quick glance at the map allows for insights into the global or regional distribution of your suppliers.

EDA Analysis:

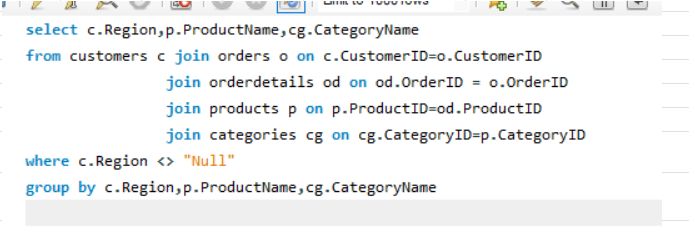
1.What are the key factors influencing customer retention or loyalty based on the dataset?

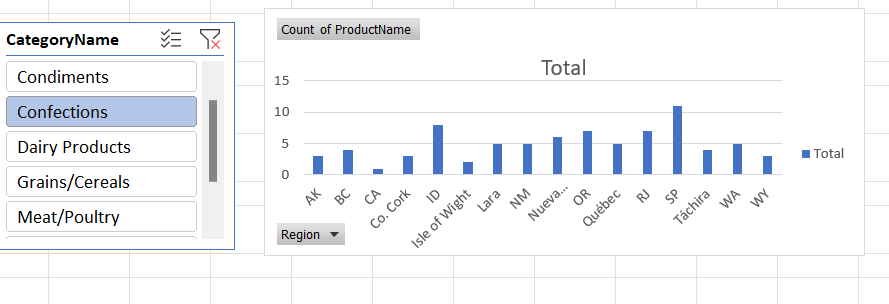




We can measure the loyalty of Customers by Observing Maximum Number of Orders.Most Of the Orders done from Sava-a-lot Markets and Ernst Handel.

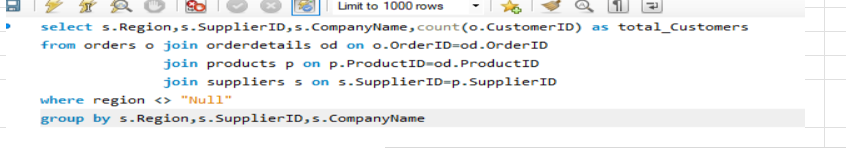
2.How do customer preferences vary based on their location or demographics? Can we explore this through interactive visualizations?

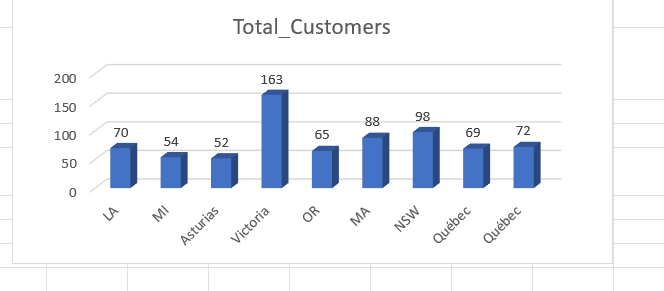




Through the Analyses we can conclude the customer Preferences region wise Product and categories.

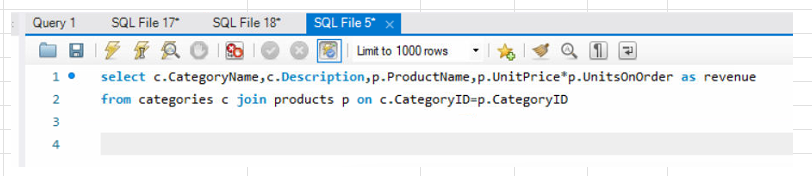
3.Are there any interesting patterns or clusters in customer behaviour that can be visualized to identify potential market segments?

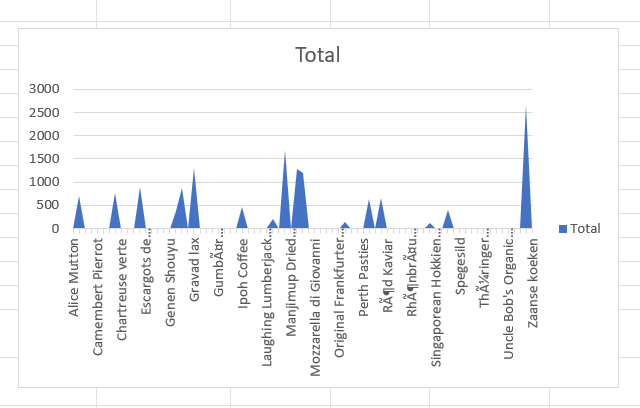




We can analyze the customer behaviour through potential markets i.e,suppliers.

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

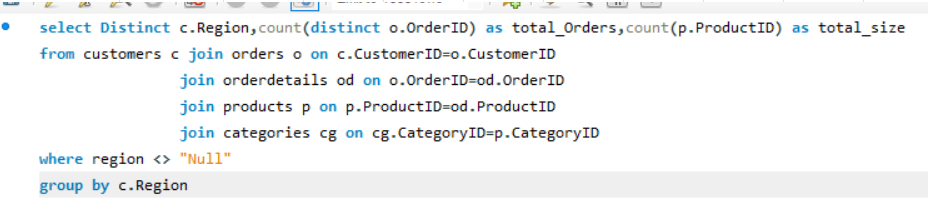


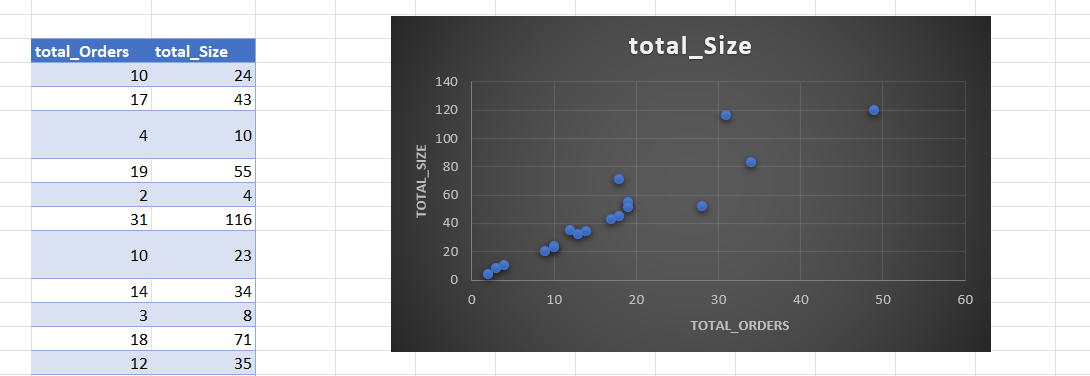


In which category product is giving max and min Revenue. In which category we can enhance or exchange the product quality.

We can also conclude which product is having better revenue as compare to other products of same Category.

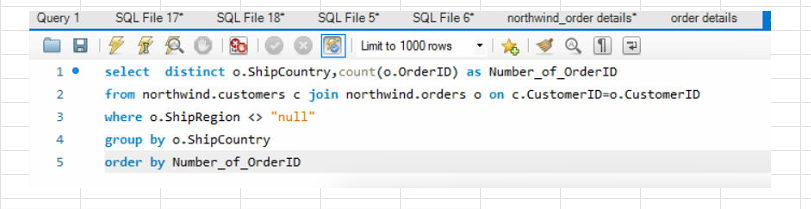
5.Are there any correlations between order size and customer demographics or product categories? Can we explore this visually using scatter plots or heatmaps?

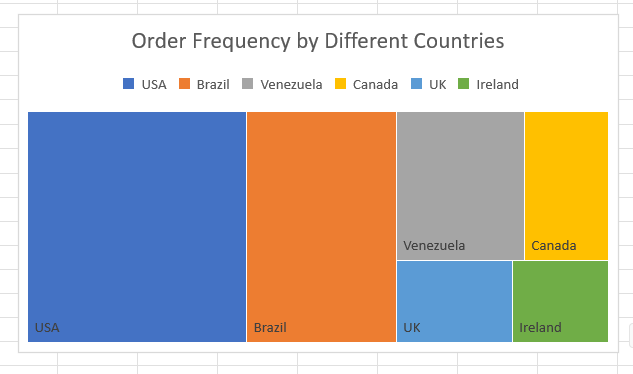




The above Graph represents the correlation between Customer Region and Product categories i.e,Size and total Orders.

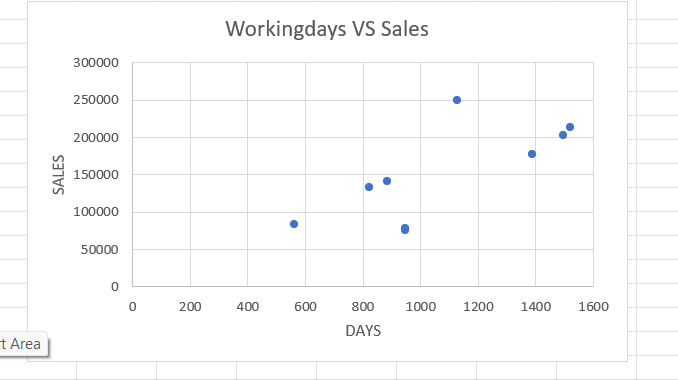
6.How does order frequency vary across different customer segments? Can we visualize this using bar charts or treemaps?

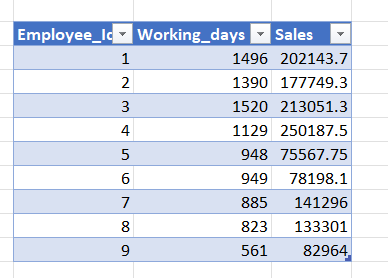




We can observe the pattern of number of orders across different countries.we get to know which country is having highest and lowest number of orders.

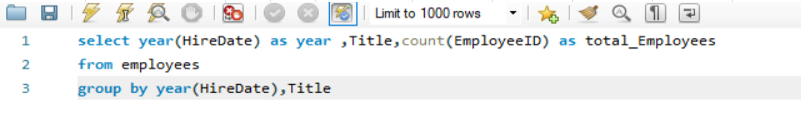
7.Are there any correlations between employee satisfaction levels and key performance indicators? Can we explore this visually through scatter plots or line charts?

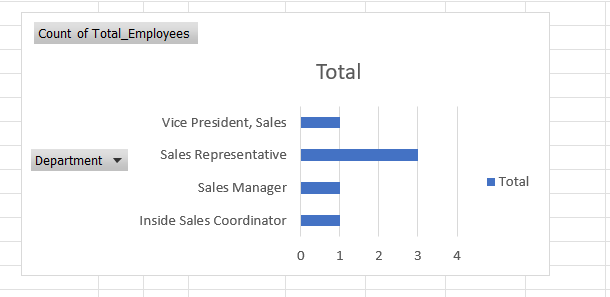


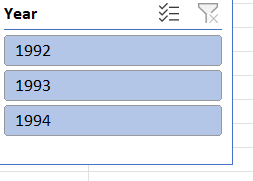


The above Graph shows the correlation between Employee working days and Sales of Employees.

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

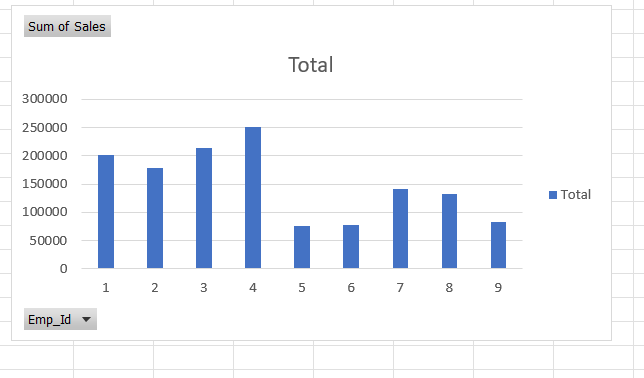


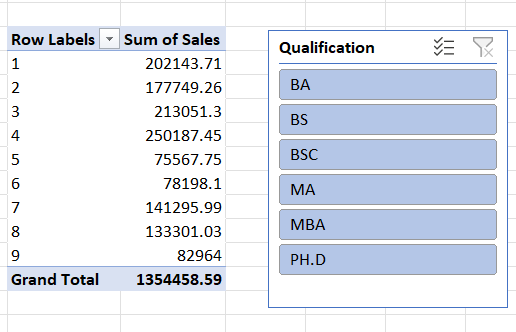




We can Analyse the Employee details regarding each department year wise.

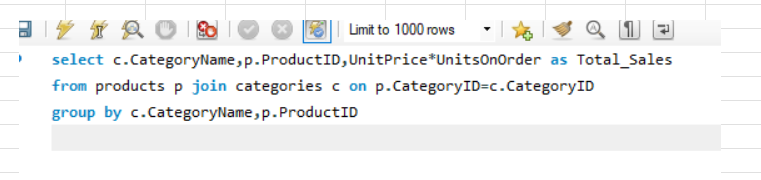
9.Can we identify any patterns or clusters in employee skill sets or qualifications through visualizations? How can this information be used for talent management?

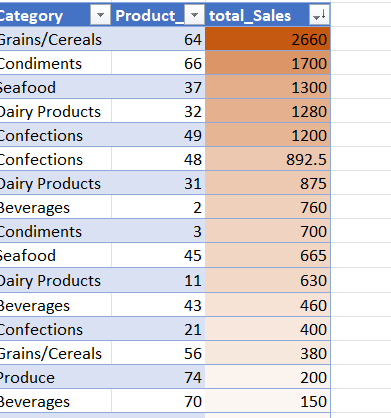




Graph Shows the total Sales of each employee regarding the Qualification.

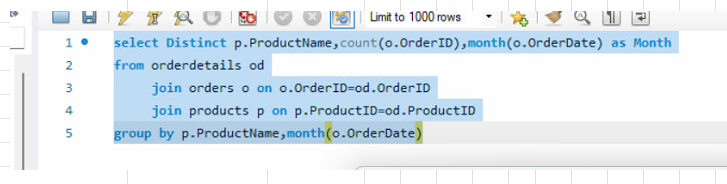
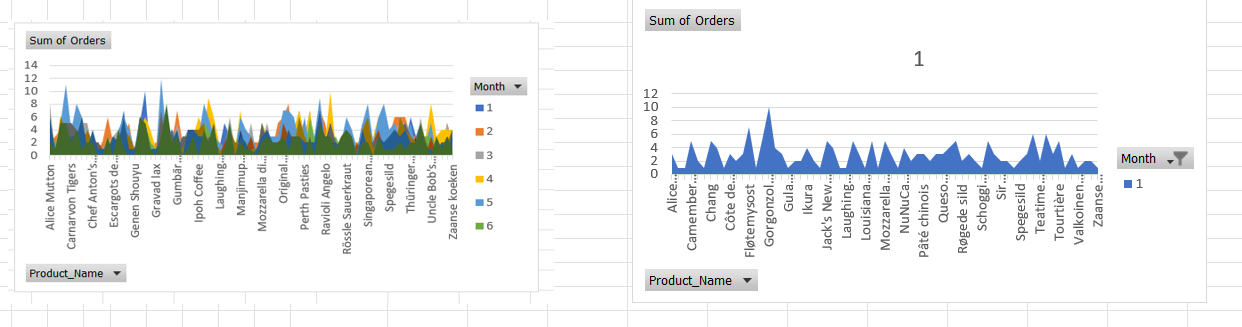
10.Are there any correlations between product attributes (e.g., size, color, features) and sales performance? Can we explore this visually using scatter plots or heatmaps?





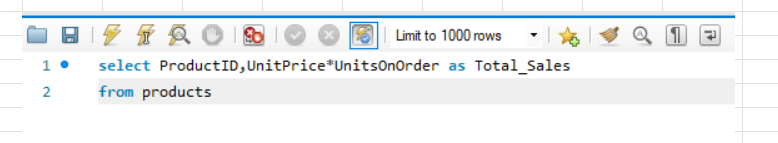
We can see the correlation of product category and sales and In Grains category we are having more number of Sales.

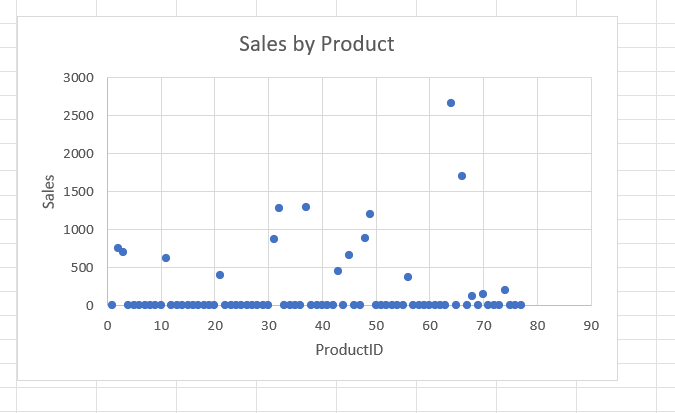
11.How does product demand fluctuate over different seasons or months? Can we visualize this through line charts or area charts?



By taking monthly product sales we can analyse which product is in more demand and in which season.

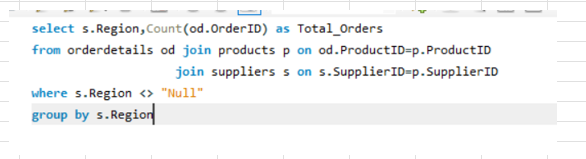
12.Can we identify any outliers or anomalies in product performance or sales using visualizations? How can this information be used for product optimization?

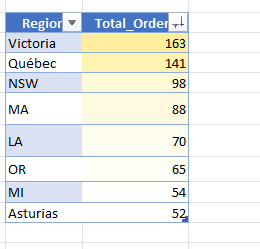




Product Id with 64 is an outlier with having highest sales 2660

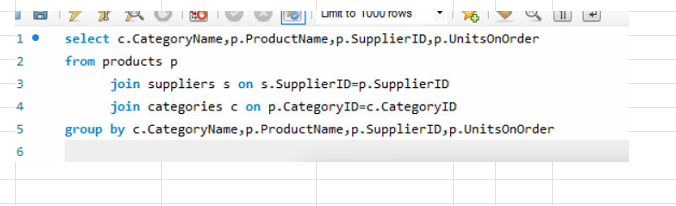
13.Are there any correlations between supplier attributes (e.g., location, size, industry) and performance metrics (e.g., on-time delivery, product quality)? Can we explore this visually through scatter plots or heatmaps?

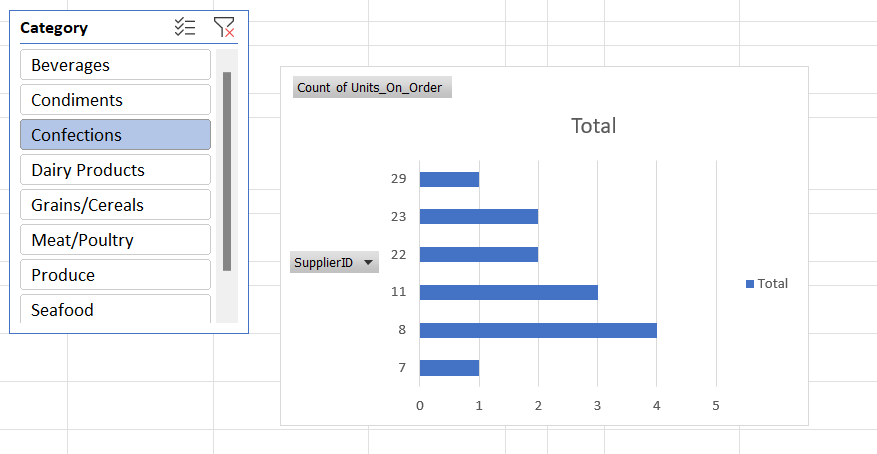




Analysing total Orders with respect to Region we can get to know correlation between them and the Region Victoria is having more number of Orders.

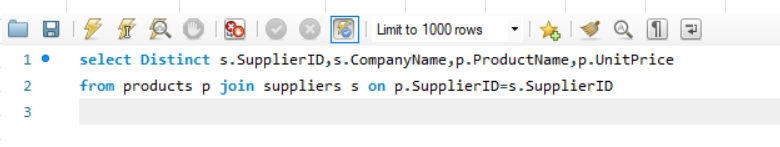
14.How does supplier performance vary across different product categories or departments? Can we visualize this using stacked bar charts or grouped column charts?

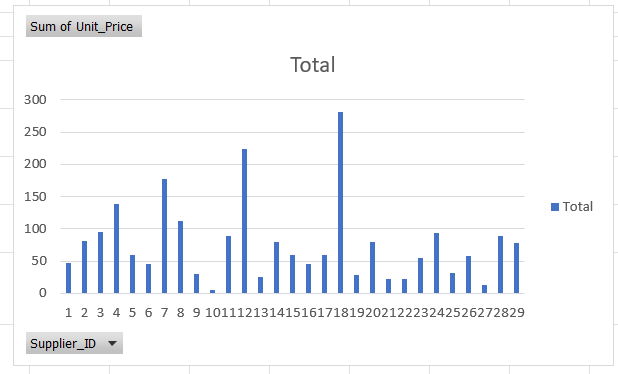




Analysing total sales with respect to product Categories along with its supplier details will help us in identifying performance of supplier.

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?





We can see the different trends occurred in the pricing Structure of Suppliers.Analysing these helps in order to obtain increase in order value which is used for procurement optimization.