

Capstone Project

North Wind Traders

Sales Analytics

By: - Divyansh Mishra

ABSTRACT

The Power BI report for Northwind Traders offers a comprehensive overview of the company's operations, leveraging data from multiple tables to provide valuable insights into sales performance, customer behaviour, inventory trends, and employee effectiveness. By integrating information from the Customers, Employees, Orders, Order Details, Products, Suppliers, Shippers, and Categories tables, the report delivers a visually appealing and user-friendly dashboard. Key performance metrics such as sales analysis, customer segmentation, and employee performance are presented through interactive visualizations and dynamic filters, enabling stakeholders to make informed, data-driven decisions. The report aims to revolutionize how Northwind Traders interacts with its data, empowering the company to remain competitive and drive business growth in the wholesale market landscape.

INDEX

Project Overview.....	04
Key Components of the Report.....	05
Information about Dataset Table.....	06-07
Data Cleaning.....	08-10
Primary Key and Foreign Key.....	11
ER Diagram.....	12
MECE breakdown.....	13
Understanding and solving PowerBi ques.....	14-30
Dashboard.....	31-34
Solving EDA questions.....	35-49
Conclusion.....	50-51

PROJECT OVERVIEW

Objective:

The objective of this Power BI report is to develop a comprehensive dashboard for Northwind Traders, a fictitious company specializing in importing and exporting specialty foods. The report aims to provide valuable insights into various aspects of the company's operations, including sales analysis, customer segmentation, inventory trends, and employee performance. The primary goal is to facilitate data-driven decision-making processes for stakeholders by presenting key performance metrics in an intuitive and visually appealing manner.

Dataset Description:

The dataset used for this project is the Northwind database, containing sales data for Northwind Traders. It comprises several interconnected tables, including Customers, Employees, Orders, Order Details, Products, Suppliers, Shippers, and Categories. Each table contains specific information related to customers, employees, orders, products, suppliers, and shipping details, enabling comprehensive analysis of the company's operations.

Key Components of the Report:

1. Sales Analysis:

1. Visualizations depict sales trends over time, including total sales, sales by product category, and geographical distribution of sales.
2. Analysis of top-selling products, customer segments, and sales performance by region or customer demographics.

2. Customer Segmentation:

1. Segmentation of customers based on various attributes such as geographic location, purchase history, and customer type.
2. Customer retention analysis and identification of high-value customers for targeted marketing strategies.

3. Inventory Trends:

1. Visualizations illustrating inventory levels, including stock availability, products on order, and reorder levels.
2. Analysis of inventory turnover rates, identifying slow-moving or obsolete products.

4. Employee Performance:

1. Evaluation of employee performance metrics such as sales contribution, order processing time, and customer satisfaction ratings.
2. Comparison of performance across different sales territories or teams.

Information about Dataset Tables

- What is Dataset?

A dataset is a collection of data.

- In this Project, we have got dataset from a company called “Northwind Traders”. A company that deals with importing and exporting food items across different countries.
- In this dataset, we have 8 different tables.

1. Customers Table
2. Employees Table
3. Orders Table
4. Order Details Table
5. Products Table
6. Suppliers Table
7. Shippers Table
8. Categories Table

Tables and their columns: -

1. **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.
2. **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.

3. **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.
4. **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.
5. **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.
6. **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.
7. **Shippers 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.
8. **Categories Table:** - This table stores information about the company's categories. 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.

Data Cleaning

After getting dataset we must understand each column in each table and perform data cleaning process on our data for further use.

- **What is Data Cleaning?**

The process of identifying, correcting, and removing errors, inconsistencies, and inaccuracies from a dataset to improve its quality and reliability for analysis. It is a crucial step in the data analysis workflow as it ensures that the data used for analysis is accurate, consistent, and complete. Data cleaning involves various tasks, including:

- 1. Handling missing values:** Identifying and dealing with missing data points, which may involve imputation (replacing missing values with estimated values), deletion of incomplete records, or other strategies depending on the nature of the data and the analysis.
- 2. Removing duplicates:** Identifying and removing duplicate entries in the dataset to avoid redundancy and ensure that each record is unique.
- 3. Standardizing formats:** Ensuring consistency in data formats, such as date formats, units of measurement, and categorical variables, to facilitate analysis and comparison across different data points.
- 4. Correcting errors:** Identifying and correcting errors in data entry, such as typos, outliers, or inconsistencies in data values, to improve the accuracy of the dataset.
- 5. Handling outliers:** Identifying and addressing outliers, which are data points that significantly deviate from the rest of the dataset, by either correcting them if they are genuine data points or removing them if they are errors or anomalies.
- 6. Dealing with inconsistencies:** Resolving inconsistencies in data values, such as conflicting information or discrepancies between different sources, to ensure data integrity and reliability.

- 7. Validating data:** Verifying the accuracy and integrity of the data by cross-referencing it with external sources or validating it against predefined criteria or business rules.

Overall, data cleaning aims to prepare the dataset for analysis by ensuring that it is accurate, consistent, and reliable, thereby enabling meaningful insights to be derived from the data analysis process.

Data cleaning, I have done in this dataset.

1. Replacing null value with N/A from each tables having region column.

A	B	C	Region
			null
BC			null
			BC
			null
			null
			null
SP			null
			SP
			null
			N/A
SP			N/A
			SP
			N/A

- Delete unnecessary columns having null values in each row from each table such as report to, image, imagethumbnail, etc.

Image	ImageThumbnail
null	null

- Formatting data columns in correct format such as DATE, DECIMAL, WHOLE NUMBER, TEXT, etc.
- Created new columns based on questions requirements.

After cleaning all the table, dataset is ready for the visualization.

To visualise the dataset, first we must identify Keys in each table such as Primary Keys, Foreign Keys.

- **What is Primary Key and Foreign Key?**

A **primary key** is a column or set of columns in a table that uniquely identifies each row (or record) in that table. It must have a unique value for each row, and it cannot contain null values.

A **foreign key** is a column or set of columns in one table that refers to the primary key in another table. It establishes a link or relationship between two tables by referencing the primary key of one table as a foreign key in another table.

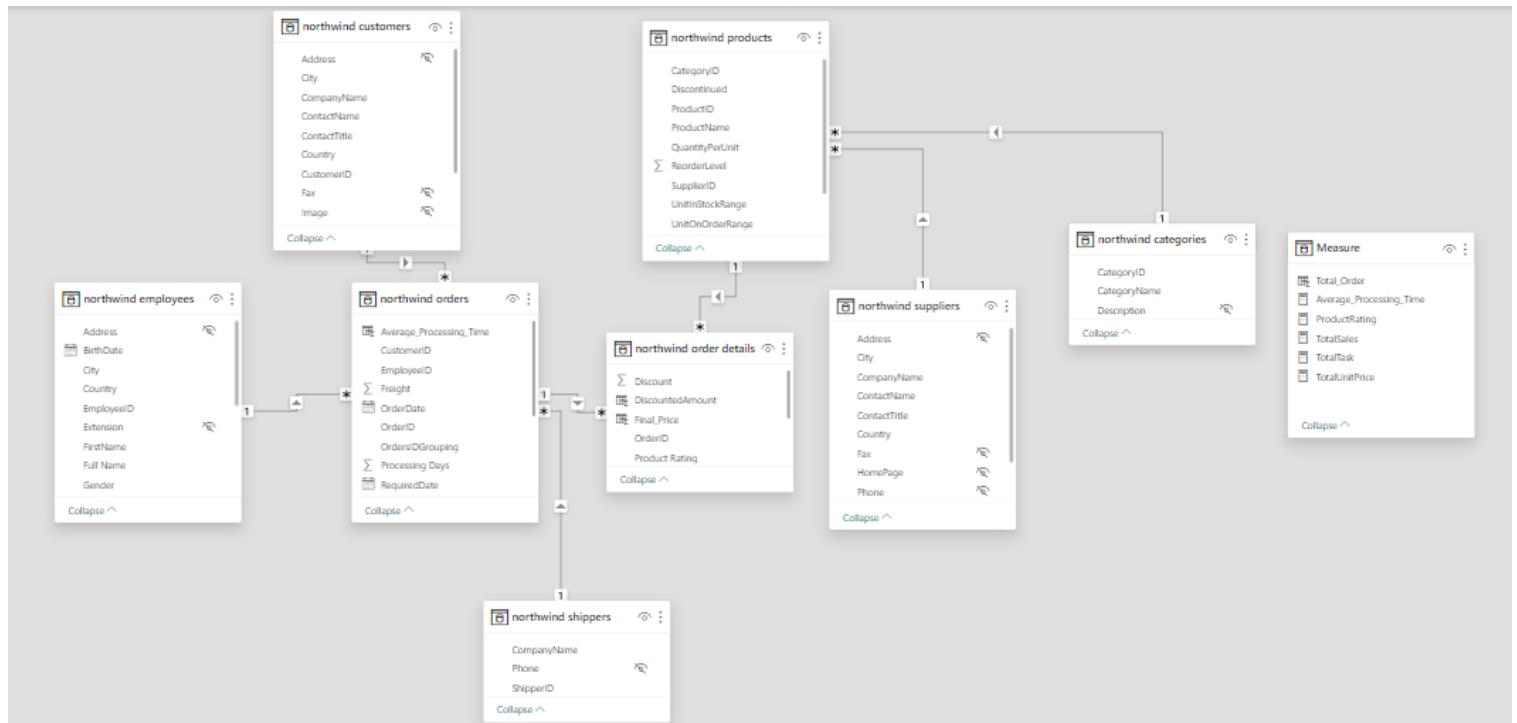
Identify all the keys in each table shown below:

Customers			Order Details			Employees			Orders		
CustomerID	PK	Numerical	OrderID	PK	Numerical	EmployeeID	PK	Numerical	OrderID	PK	Numerical
City		Text	ProductID	FK	Numerical	FirstName		Text	OrderDate		DateTime
CompanyName		Text	Quantity		Numerical	LastName		Text	RequiredDate		DateTime
ContactName		Text	Unit Price		Numerical	BirthDate		DateTime	Freight		Numerical
ContactTitle		Text	Discount		Numerical	City		Text	CustomerID	FK	Numerical
Country		Text				Country		Text	EmployeeID	FK	Numerical
Address		Text				Extension		Numerical	ShipAddress		Text
Phone		Numerical				HireDate		DateTime	ShipCity		Text
PostalCode		Text				HomePhone		Numerical	ShipCountry		Text
Region		Text				Address		Text	ShipName		Text
Image		Text				Notes		Text	ShippedDate		DateTime
ImageThumbnail		Text				Photos		Text	ShipPostalCode		Text
						PostalCode		Text	ShipRegion		Text
						Region		Text	ShipVia		Text
						ReportsTo		Text			
						Title		Text			
						TitleOfCourtesy		Text			
Products			Suppliers			Categories					
ProductID	PK	Numerical	SupplierID	PK	Numerical	CategoryID	PK	Numerical			
ProductName		Text	CompanyName		Text	CategoryName		Text			
QuantityPerUnit		Numerical	ContactName		Text	Description		Text			
CategoryID	FK	Numerical	Address	FK	Text	Picture		Text			
Discontinued		Numerical	City		Text						
ReorderLevel		Numerical	ContactTitle		Text						
SupplierID	FK	Numerical	Country	FK	Text						
UnitPrice		Numerical	Fax		Text						
UnitInStock		Numerical	PostalCode		Text						
UnitsOnOrder		Numerical	Phone		Numerical						
			Region		Text						
			HomePage		Text						

PK - Primary Key
FK - Foreign Key

After all this task we have to create ER diagram for our dataset.

ER Diagram:



Above ER diagram shows all the tables and their connections with each other. It will help use to understand the flow of connections and type of connection such as One-to-one or One-to-many. In this ER Diagram we have use One-to-one connection.

Tables connections:

1. Employees Table connected with Orders Table using EmployeeID.
2. Customers Table connected with Orders Table using CustomerID.
3. Orders Details Table connected with Orders Table using OrderID.
4. Shipper Table connected with Orders Table using ShipperID.
5. Products Table connected with Order Details Table using ProductID.
6. Products Table connected with Suppliers Table using SupplierID.
7. Products Table connected with Categories Table using CategoryID.

After all this steps we can start with our questions.

MECE Breakdown

MECE stands for "Mutually Exclusive, Collectively Exhaustive". MECE breakdown refers to the principle of breaking down a problem or a set of information into mutually exclusive and collectively exhaustive (MECE) categories. This approach is commonly used in problem-solving, analysis, and communication to ensure clarity, completeness, and effectiveness. MECE breakdowns are essential because they provide a structured framework for organizing and understanding complex information, facilitating better decision-making and communication.



Understanding & Solving Power Bi Questions

In power bi we can divide questions into 5 different categories.

- 1. Customers Analysis:** - It contains all the record of Customers such as Customer ID, Company name, contact name, contact title, City, Country, etc. which help us to find customer distribution, acquisition, and demographic information.
- 2. Employee Analysis:** - It contains all the records of employees such as Emp ID, first name, last name, Date of birth, hire date, city, country, etc. it helps us to solve questions related to employee tenure, employee productivity and employee performance.
- 3. Order Analysis:** - It contains all the records related to orders such as Order ID, order date, required date, shipped date, ship name, ship via, ship address, etc. it will answer the following questions such as average order processing time, order volume and order value.
- 4. Product Analysis:** - It stores all the products record such as Product ID, Product name, quantity, unit price, unit in stock, unit on order, etc. it help us to show pricing of product, sales volume across different categories, product ratings.
- 5. Supplier Analysis:** - It stores records such as Supplier ID, supplier company name, supplier contact name, contact title, city, country. It will help us to find geographical distribution of supplier, pricing structure of supplier and supplier performance.

- To solve questions, we have to make few columns & measures based on the requirements.

I have created few columns & measures for dataset based on my requirements.

The screenshot shows the Power BI Data View interface. On the left, under the 'Measures' section, there is a list of six measures: Average_Processing_Time, ProductRating, Total_Order, TotalSales, TotalTask, and TotalUnitPrice. On the right, under the 'northwind employees' table, there is a list of columns: BirthDate, City, Country, EmployeeID, FirstName, Full Name, Gender, HireDate, LastName, Region, Tenure, and Title. The 'Full Name' and 'Tenure' columns are highlighted with yellow backgrounds.

Measures	Columns
Average_Processing_Time	BirthDate
ProductRating	City
Total_Order	Country
TotalSales	EmployeeID
TotalTask	FirstName
TotalUnitPrice	Full Name
	Gender
	HireDate
	LastName
	Region
	Tenure
	Title

Columns

- Different between Measures and Columns

Measures:

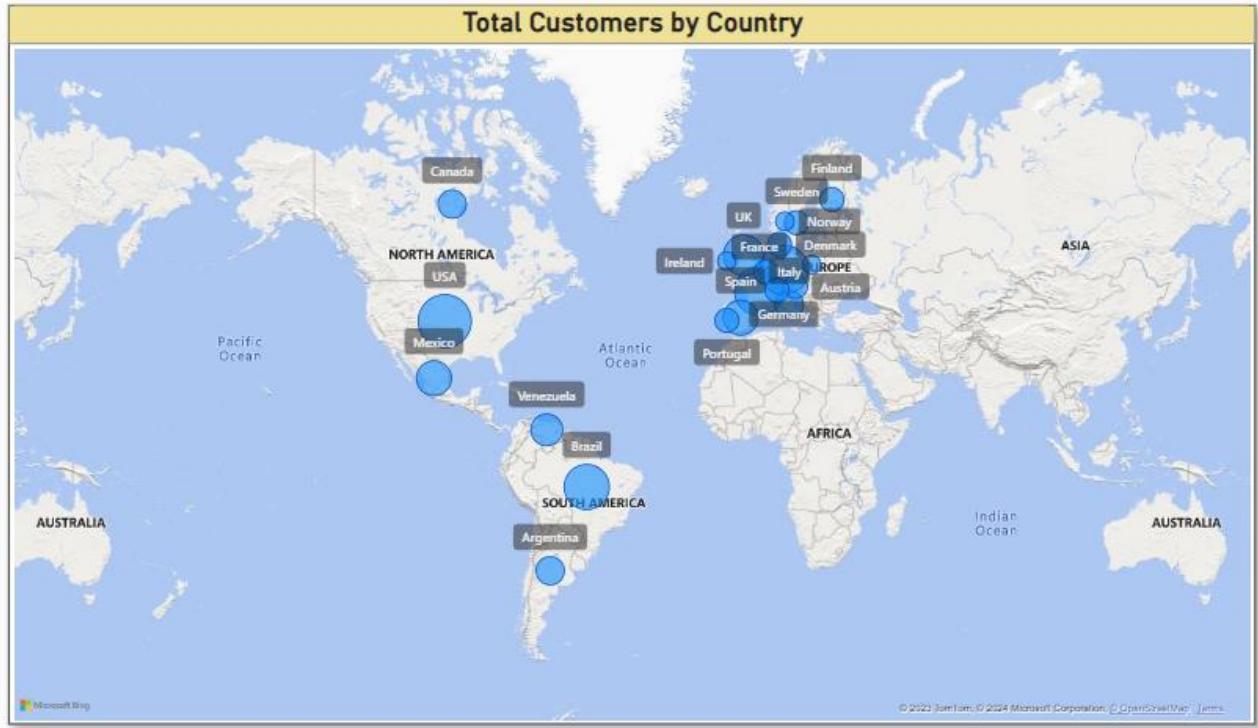
1. Measures are dynamic calculations based on the data in the model.
2. They are created using DAX (Data Analysis Expressions), a formula language in Power BI, and can perform calculations such as sums, averages, counts, or more complex aggregations.

Columns:

1. Columns are fields in a table that contain static data values for each row.
2. They are derived from the underlying data source or calculated during data loading using Power Query (M language).
3. Columns represent attributes or characteristics of the data, such as customer names, product categories, or order dates.

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

Answer:



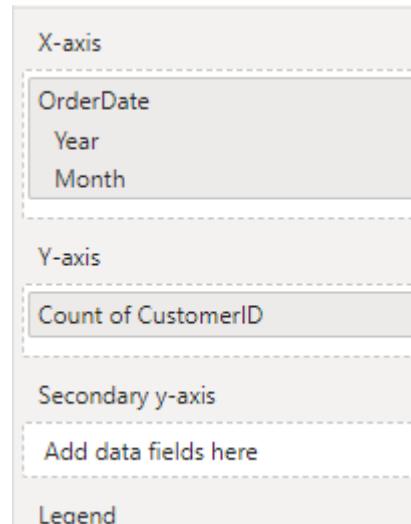
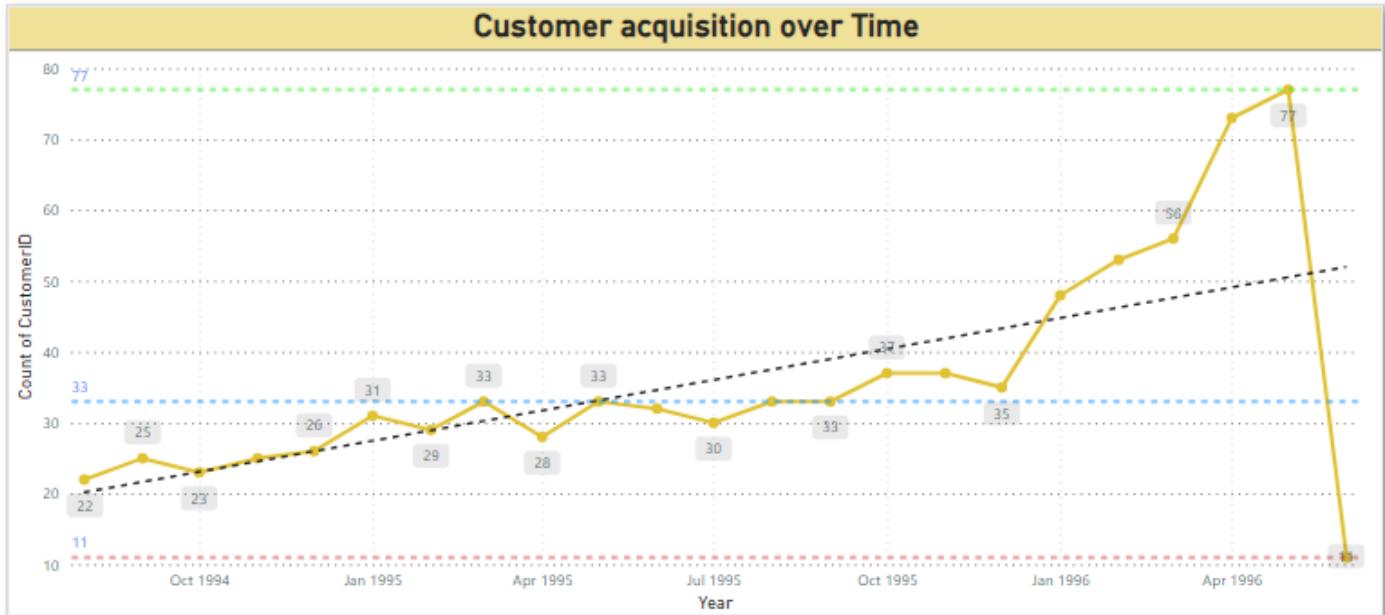
In this I have used map chart to show customers distribution across different countries, it will help us to identify which countries have maximum customer distributions.

Location
Country
Legend
Add data fields here
Latitude
Add data fields here
Longitude
Add data fields here
Bubble size
Count of CustomerID

Yes, we can create map chart to visualize this question. In this question I have used "Country" in location section and "Count of CustomerID" in bubble size. It will count the customers using count function and display the total number of customers in each country.

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

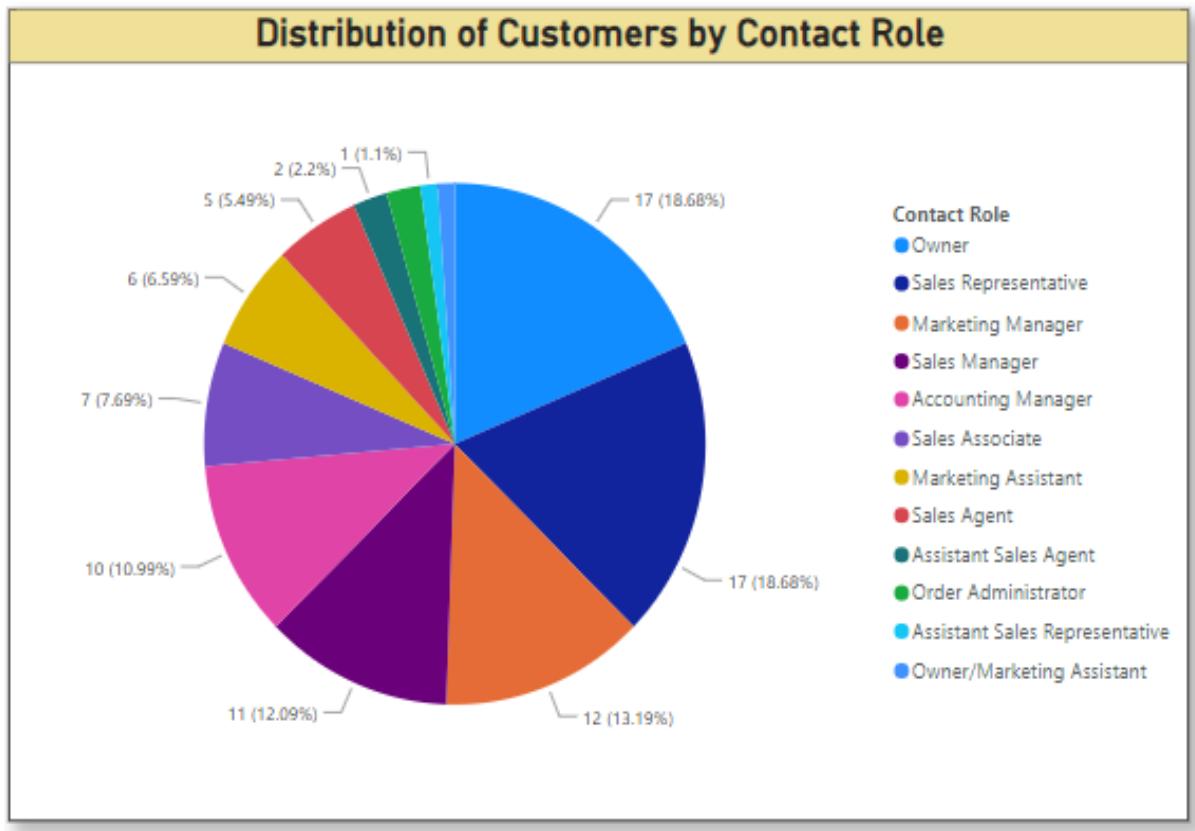
Answer:



Yes, we can create line chart to analyze trend over time. To create this visualization, I have used “OrderDate” from order table in X-axis and “Count of CustomerID” from order table in Y-axis. It will count the total number of customers in months and years, which will help us to find maximum customers in which year. Its shows **maximum number of customers in April 1996 having 77 customers and oct 1994 having minimum 22 customers.**

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

Answer:



Legend

Contact Role

Values

Count of CustomerID

Details

Yes, we can create pie chart to find customers demographic, to calculate distribution of customers I have used contact role from demographic. It will provide us the number of customers in each contact role which will help us to analyse which role is having maximum distribution. I have used “ContactTitle as Contact Role” from customer table in Legend and “Count of Customer ID” from customer table in Values. It shows **“Sales Representative” have highest number of customers having 18.68%.**

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

Answer: Yes, we can create time series charts to visualise changes in order volume over time.

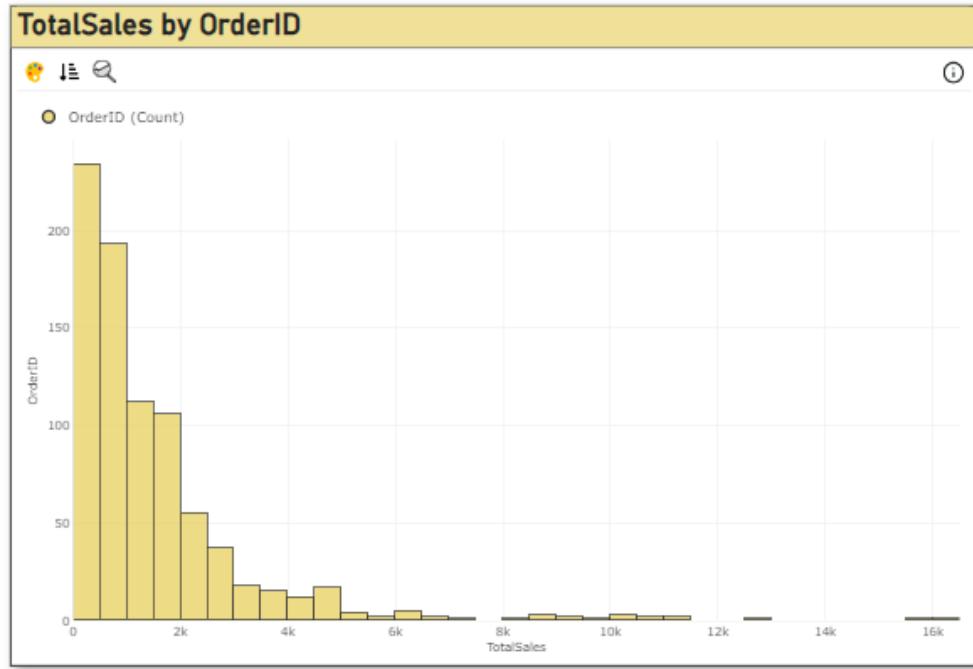
X-axis
OrderDate
Year
Quarter
Month
Y-axis
Count of OrderID

I have use “OrderDate” from Order table in X-axis and “Count of OrderID” from Order table in Y-axis. To find the trend of number of orders in different months, quarters & years. It shows **April 1996 is having maximum number of orders 161 and July 1994 is having minimum number of orders 47**. Which help us to understand trend.



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

Answer:



Value 1
TotalSales
Value 2
Add data fields here
Value 1 weight
OrderID
Value 2 weight

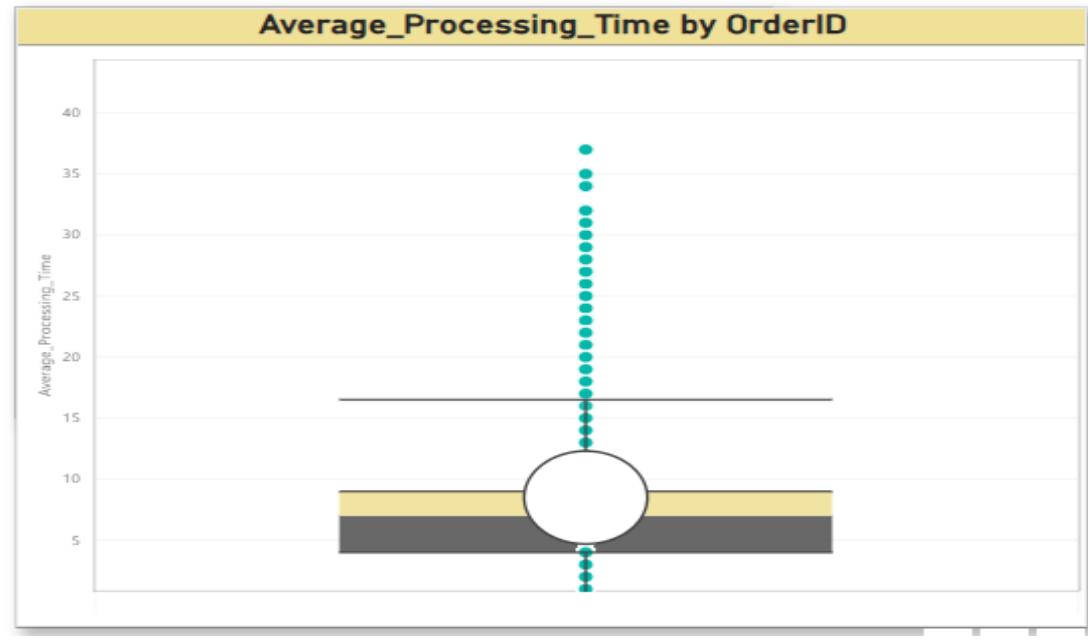
Yes, we can create a histogram chart to understand the distribution range of order volume. I have used “TotalSales” from measures table in Value 1 and “OrderID” from order table in Value 1 weight. It will create distribution of total sales in difference of 500 and provide us the number of orders in each distribution range. It displays range between 00-500 having maximum number of order count 234.

Measure created to calculate Total Sales given below.

```
1 TotalSales = SUMX('northwind order details', ('northwind order details'[UnitPrice]-'northwind order details'[DiscountedAmount])*'northwind order details'[Quantity])
```

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

Answer:



Median Type	Inclusive
Whisker Type	= 1.5 IQR
Mean	8.49
Quartile 1	4.00
Median	7.00
Quartile 3	9.00
Maximum	37.00
Minimum	1.00
IQR	5.00
Upper Whisker	16.50
Lower Whisker	-3.50

Yes, we can create box plot chart to calculate average processing time for shipping duration for this I have used “OrderID” from order table in Axis and

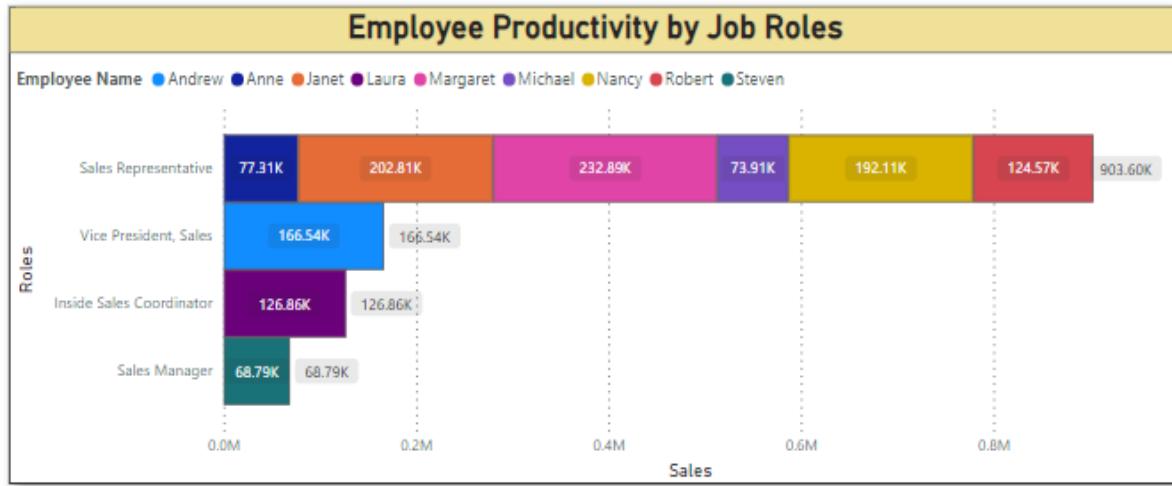
“Average_Processing_Time” from measures table to find average processing time for delivery. **It also shows outliers, mean 8.49 and median 7.00.**

Measure created to calculate Average Processing Time given below.

```
Average_Processing_Time = DIVIDE(SUM('northwind orders'[Processing Days]),DISTINCTCOUNT('northwind orders'[OrderID]))
```

Question 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?

Answer:



Y-axis
Roles
X-axis
Sales
Legend
Employee Name

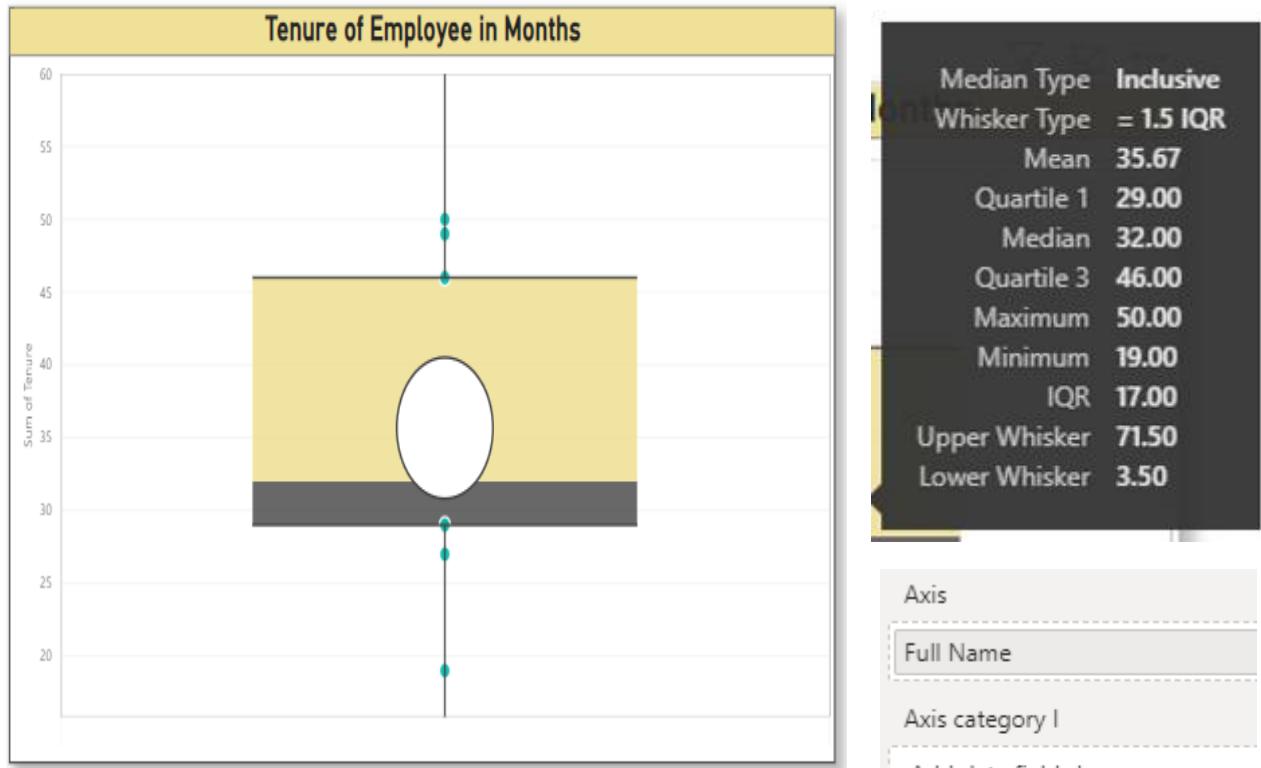
Yes, we can create stacked bar chart to find employee productivity across different job roles. I have used “Title as Roles” from employee table in Y-Axis, “TotalSales as Sales” from measure table in X-Axis and “FirstName as Employee Name” from employee table in Legend. It will display “Sales Representative” role having highest sales around 903k and employee name “Margaret” have done maximum sales 232.8k

Measure created to calculate Total Sales given below.

```
1 TotalSales = SUMX('northwind order details', ('northwind order details'[UnitPrice]-'northwind order details'[DiscountedAmount])*'northwind order details'[Quantity])
```

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

Answer:



Yes, we can create a box plot chart to identify Employee tenure in months. I have created new column in employee table as Full Name using concat function between First name and Last name and new column as tenure to calculate number of months using Datediff function between hire date and order date. To calculate employee tenure, **it shows maximum tenure of employee is 50 months and minimum is 19.**

Measure created to calculate Tenure given below.

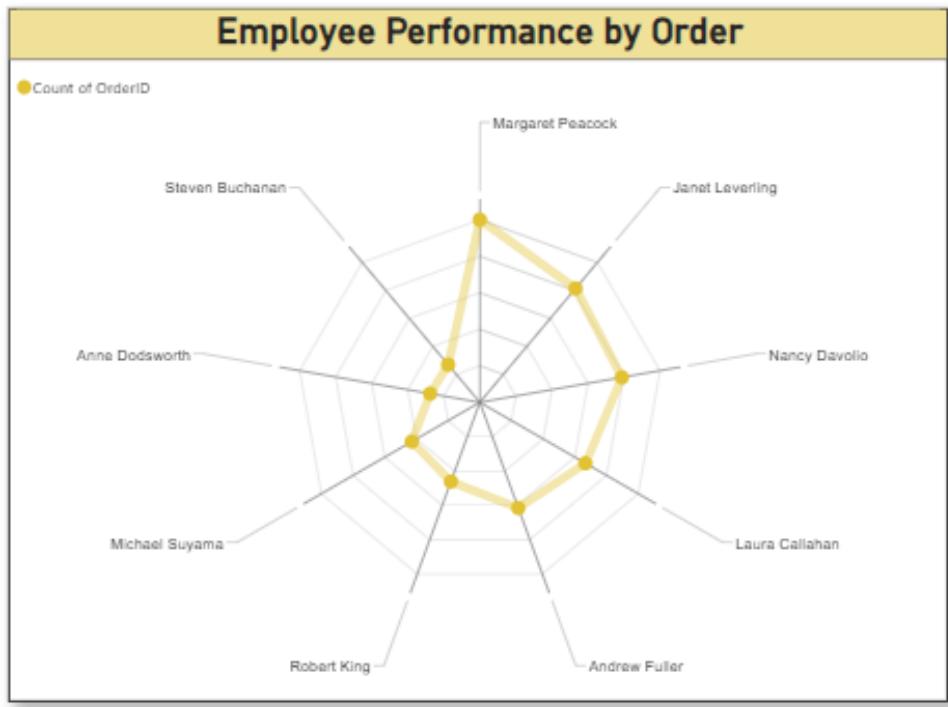
```
Tenure = DATEDIFF('northwind employees'[HireDate],MAX('northwind orders')[OrderDate]),MONTH)
```

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

Answer:

Yes, we can calculate employee performance using radar chart. I have used “Full Name” from Employee table in Category and “Count of Order ID” from order table in Y-Axis which will provide us the number of orders by each employee.

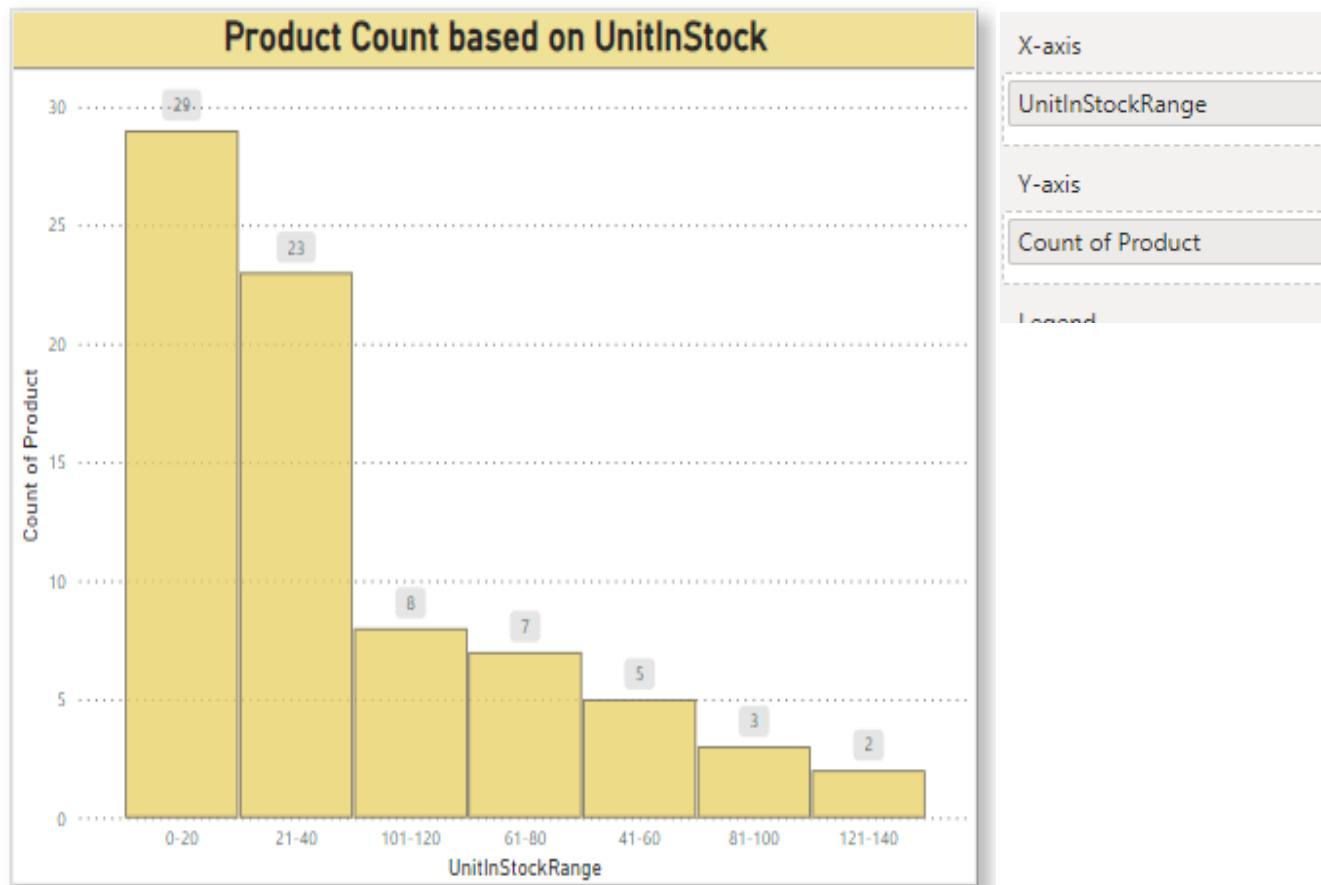
Which will help us to recognize best performance Employee maximum the number of order maximum the employee ratings. **It shows Margaret Peacock is the best performing employee having 156 orders.**



Category
Full Name
Y Axis
Count of OrderID
Drill through

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

Answer:

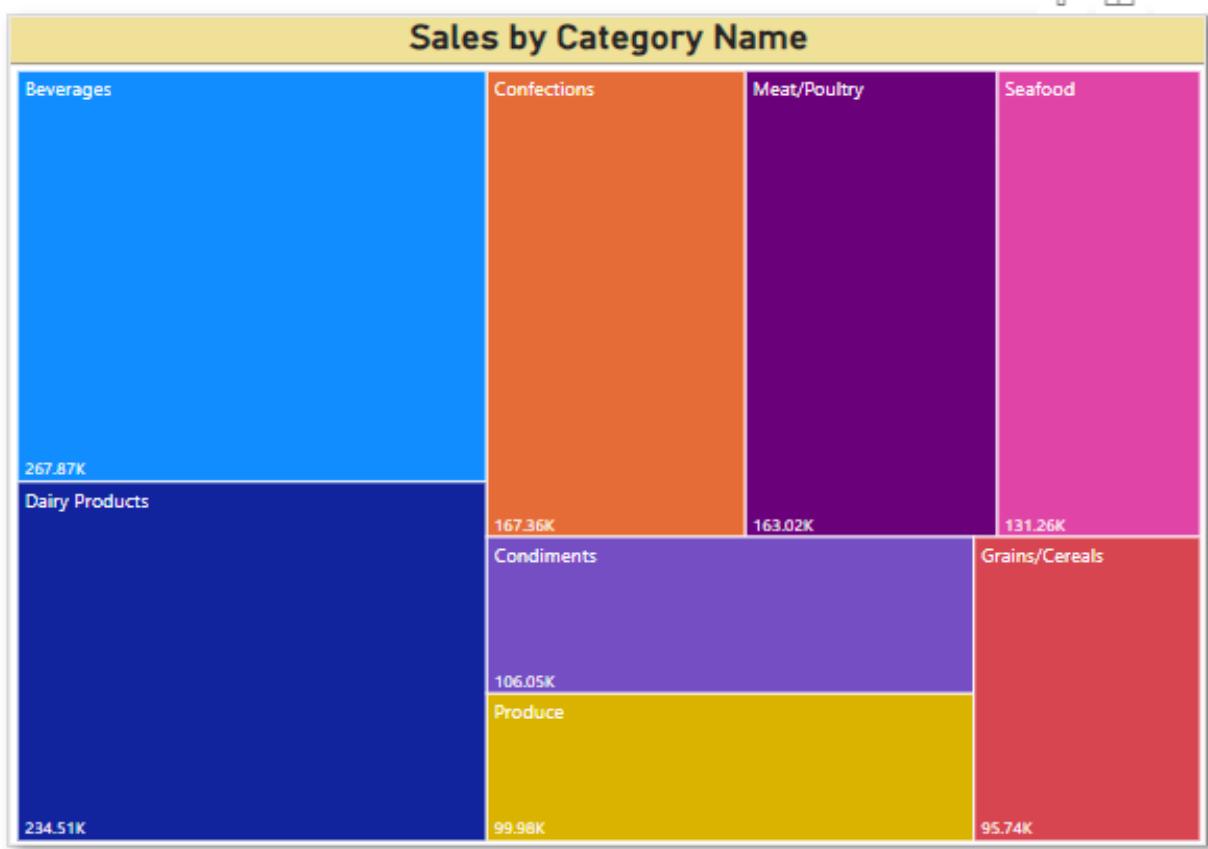


Yes, we can create a histogram to calculate distribution of product rating.

I have created a new conditional column as UnitInStockRange to which shows range of UnitInStock. Highest the unit in stock shows the highest demand of product which means if unit in stock is more demand of that product is more which will help us to identify the best rating product. In this I have used "UnitInStockRange" in X-Axis and "Count of Product" in Y-Axis. **There is total 29 product in lies between 0-20 range which shows 29 product can have most demand.**

Question 11: - How does the sales volume vary across different product categories? Can we create a bar chart or treemap to display it?

Answer:



Category

Category Name

Details

Add data fields here

Values

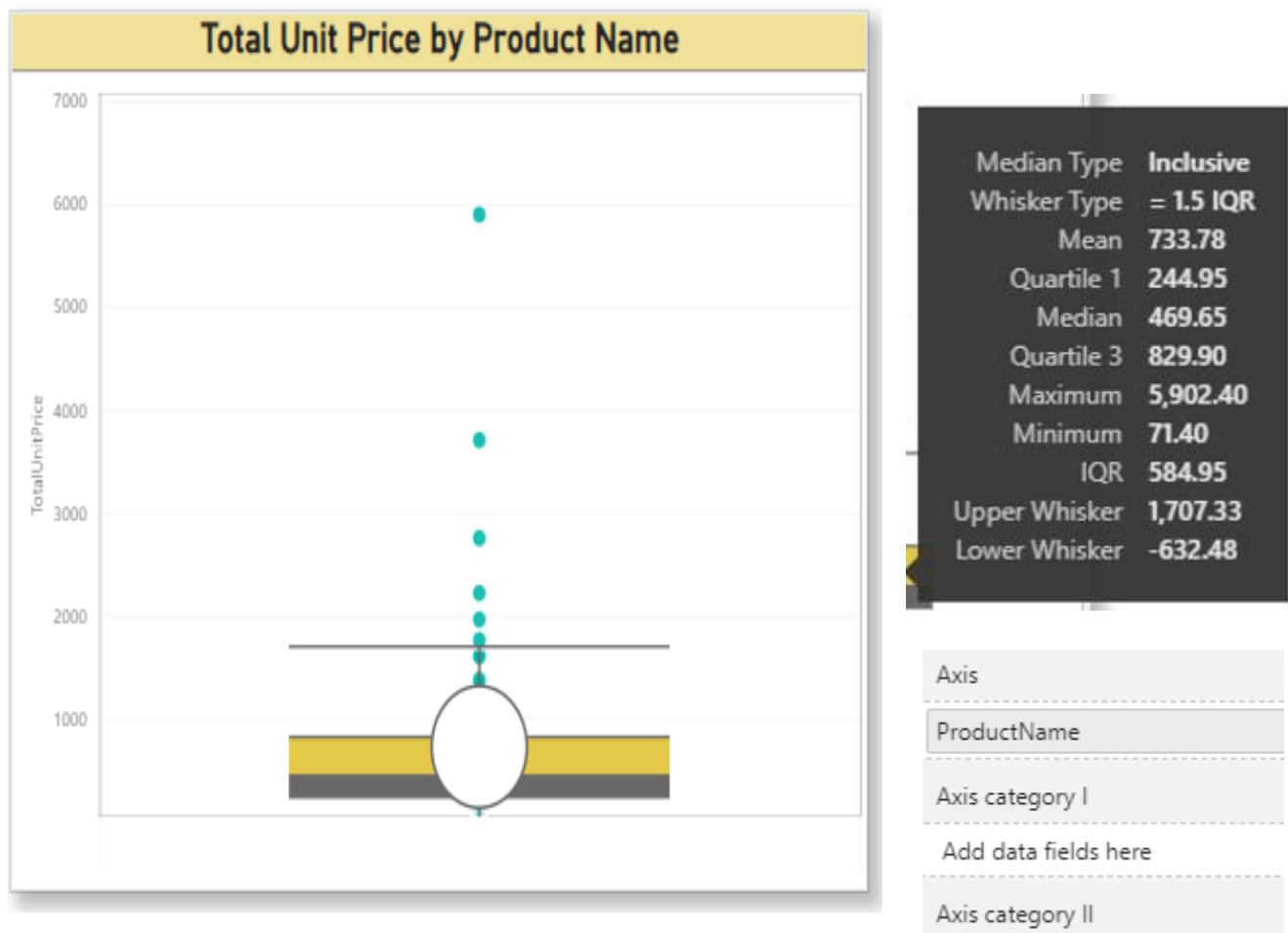
Sales

Tooltips

Yes, we can identify Total Sales across different product categories with the help of "Category Name" from Categories table in Category and "TotalSales as Sales" from measures in Values. It shows category name "Beverages" having maximum sales around 267.87K.

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

Answer:



Yes, we can create box plot for pricing distribution of product it will shows us the product price. To display this, I have used ProductName from product table and created measure name as Total Unit price which will calculate the total unit price of each product and put it in Values and put ProductName in Axis. It will show us the range between which maximum pricing lies and outlier pricing.

Measure created to calculate Total Unit Price given below.

```
1 TotalUnitPrice = SUM('northwind order details'[UnitPrice])
```

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

Answer:

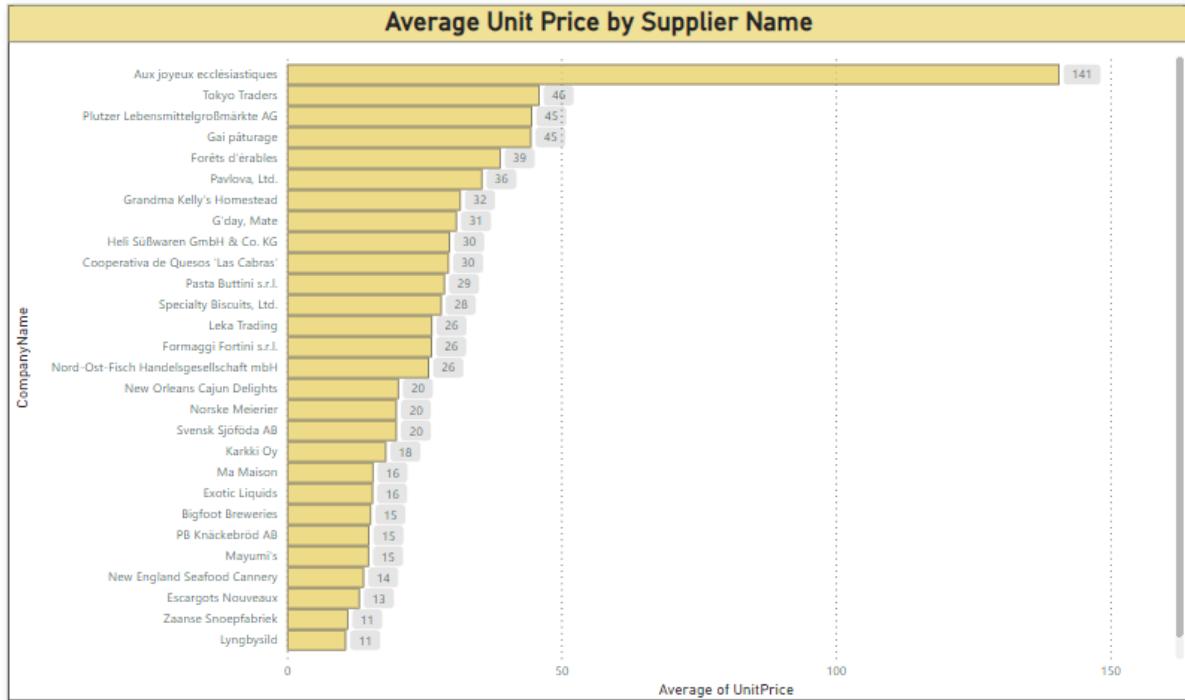


Yes, we can create a bar chart to visualize supplier performance, we can calculate supplier performance based on total quantity of product company have supplied to customers maximum se Total Quantity maximum se performance metrics. I have used “CompanyName” from supplier table in Y-Axis and “Sum of Quantity” from orders details in X-Axis. **It will display supplier company name Plutzer is having maximum quantity 4.1k.**



Question 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?

Answer:



Y-axis
CompanyName

X-axis
Average of UnitPrice

Legend

Yes, we can create stacked bar charts to identify pricing structure of each supplier. To calculate pricing structure, I have used “CompanyName” from supplier in Y-Axis and “AverageUnitPrice” from product in X-Axis which will shows us the average pricing of each supplier company. **It shows supplier company name Aux Joyeux is having maximum average price 141.**

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

Answer:



Location

Country

Legend

Add data fields here

Latitude

Add data fields here

Longitude

Add data fields here

Bubble size

Count of SupplierID

Yes, we can create map chart to identify suppliers in different countries. I have used "Country" from supplier table in Location and "Count of supplierID" from supplier table in Bubble size. Which will display us the number of suppliers in each country.

Dashboard

Home Page Dashboard

INDEX

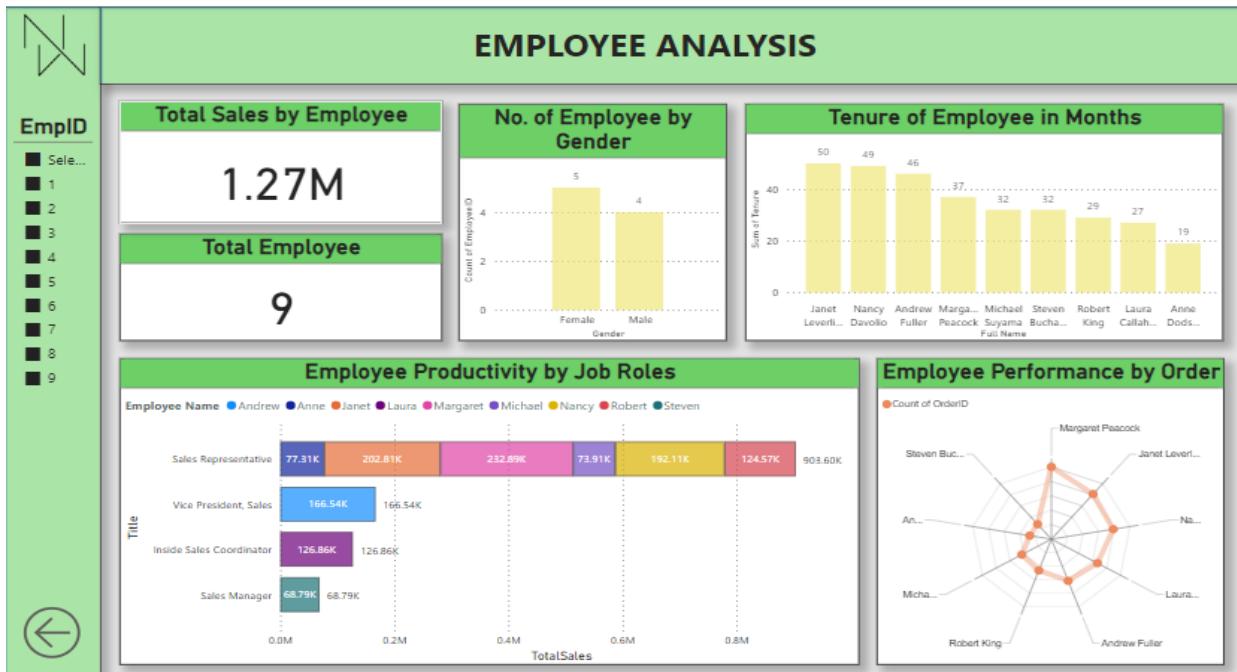
NORTHWIND TRADERS DASHBOARD

The dashboard features a vertical sidebar on the left with a stylized 'N' logo at the top and the word 'INDEX' below it. The main area has a light green header bar. Below it is a grid of five categories, each with a small icon and a title. To the right of the grid is a large text area containing detailed descriptions for each category.

Category	Description
EMPLOYEES	Employee Performance: - Evaluation of employee performance metrics such as sales contribution, order processing time, and customer satisfaction ratings. - Comparison of performance across different sales territories or teams.
CUSTOMERS	Customer Segmentation: - Segmentation of customers based on various attributes such as geographic location, purchase history. - Customer retention analysis and identification of high-value customers for targeted marketing.
ORDERS	Sales Analysis: - Visualizations depicting sales trends over time, including total sales, sales by product category, and geographical distribution of sales. - Analysis of top-selling products, customer segments, and sales performance by region or customer demographics.
PRODUCTS	Inventory Trends: - Visualizations illustrating inventory levels, including stock availability, products on order, and reorder levels. - Analysis of inventory turnover rates, identifying slow-moving or obsolete products.
SUPPLIERS	

This is the home page of my Northwind Traders Dashboard. It displays all the 5 categories: - Employees, Customers, Orders, Products, Suppliers after click on any of the category this will navigate us to that category dashboard.

Employee Dashboard



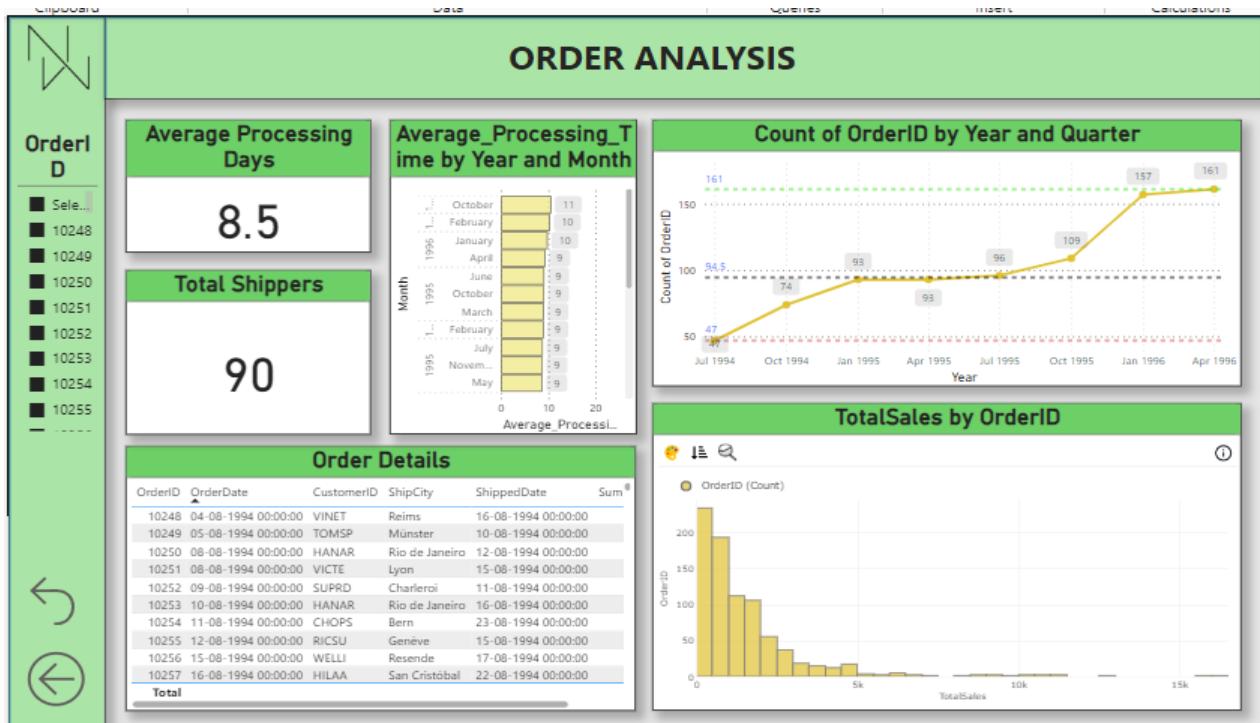
This is the Employee Dashboard which will help use to analyse Employee performance. It displays total employees and total sales by employees.

Customer Dashboard



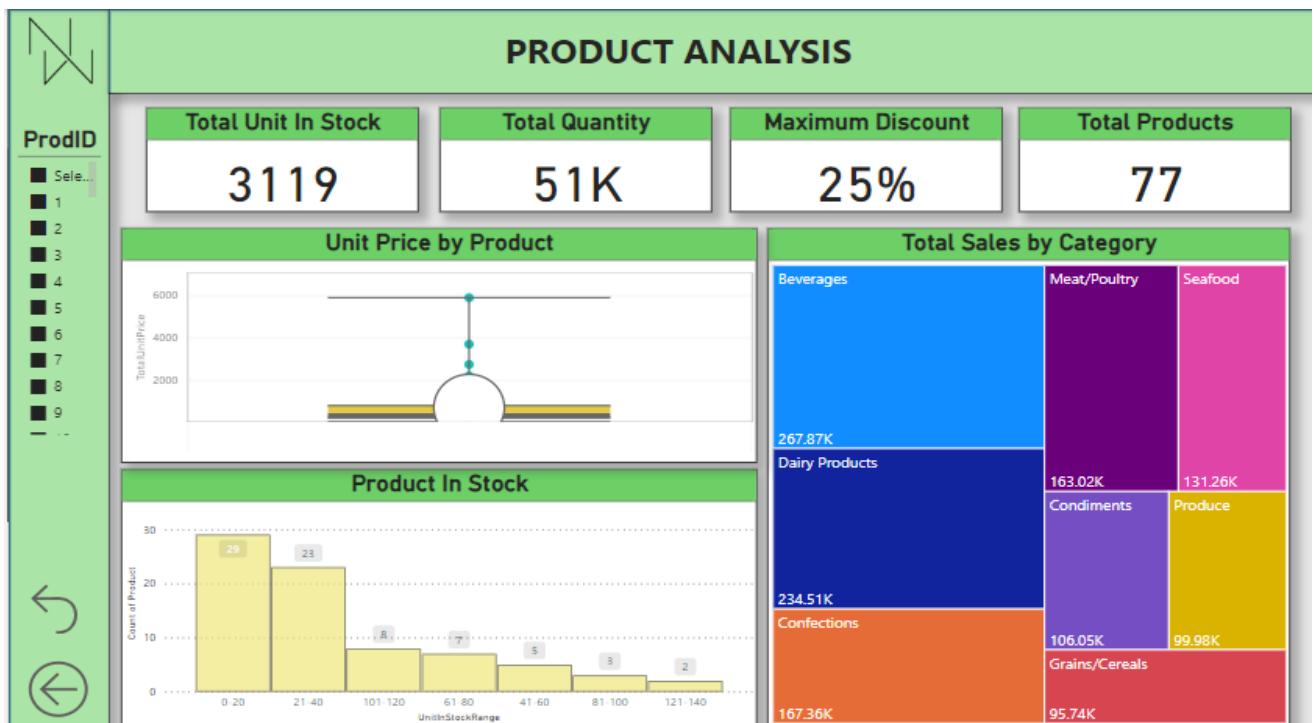
It displays all the customers performance and help use to understand customers trend over time.

Order Dashboard



It displays all the important factors of order and order details. It helps us to track order processing days and total sales by each order ids.

Product Dashboard



It displays all the product details which will help us traders to maintain their stocks based on demand.

Supplier Dashboard



This dashboard will show suppliers in different countries, pricing structures and quantity shipped by each supplier.

Solving EDA Questions

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

Answer:

```
select
o.CustomerID,c.Country, round(AVG(Monetary),2) as AvgMonetary,
count(o.OrderID) Frequency,
min(datediff((select max(date(orderDate)) from orders),date(o.OrderDate))) Recency
from orders o
join customers c
on c.CustomerID = o.CustomerID
join northwind.`order details` od
on o.OrderID = od.OrderID
join
(
    select od.OrderID,
    round(sum( UnitPrice * Quantity * ( 1 - Discount )),2) as Monetary
    from northwind.`order details` od
    group by 1
) mon on od.OrderID = mon.OrderID
group by 1
order by 1
```

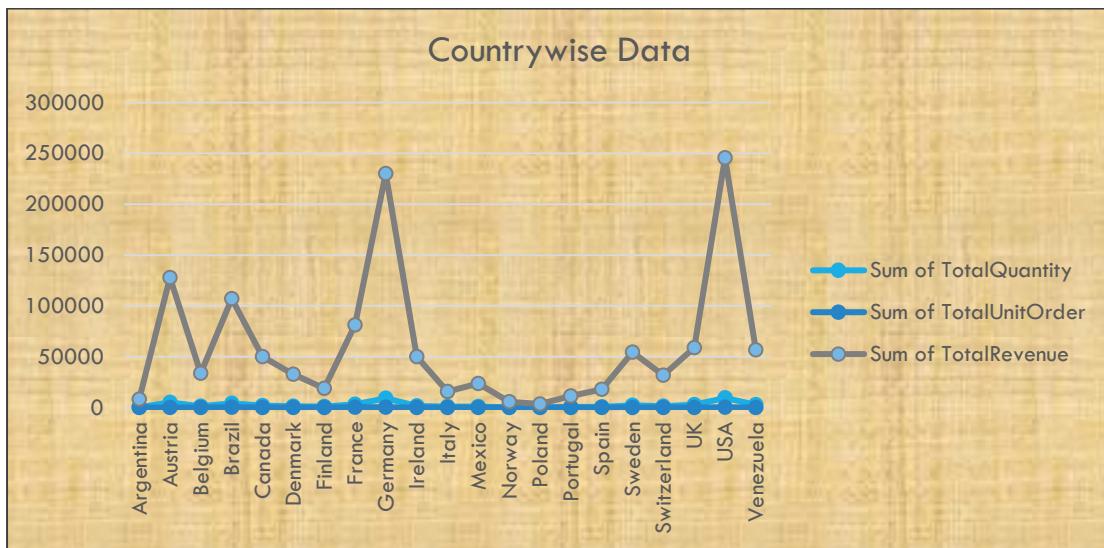
RFM by Customers			
CustomerID	Sum of AvgMonetary	Sum of Frequency	Sum of Recency
ALFKI	706.88	12	27
ANATR	399.45	10	63
ANTON	1175.53	17	98
AROUT	1258.89	30	26
BERGS	1455.98	52	63
BLAUS	540.56	14	7
BLONP	2504.74	26	114
BOLID	1537.28	6	43
BONAP	1465.81	44	0
BOTTM	1731.15	35	12
BSBEV	603.41	22	22
CACTU	335.69	11	8
CENTC	100.8	2	657
CHOPS	1564	22	14
COMMI	1093.95	10	14
CONSH	581.4	7	103
DRACD	738.04	10	2
DUMON	494.77	9	79
EASTC	2049.08	21	8
ERNSH	3908.1	102	1
FAMIA	665.87	19	187
FOLIG	2682.32	16	135
GOLIK	1001.11	15	0

Key factors to check customers loyalty is based on RFM. R stands for Recency, F stands for Frequency, M stands for Monetary. It helps you to understand loyalty of customers. Higher the frequency and lower the Recency gives you the loyal customers.

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

Answer:

```
select c.CustomerID, c.Country, SUM(od.Quantity) as TotalQuantity , COUNT(od.OrderID) as TotalUnitOrder,
round(sum( UnitPrice * Quantity * ( 1 - Discount )),2) as TotalRevenue
from Customers c
join Orders o on c.CustomerID = o.CustomerID
join `order details` od on o.OrderID = od.OrderID
GROUP BY 1;
```



We can find customer preferences vary across different location using this query which will provide us the Total Quantity, Total Unit in Order, and Total Revenue by customers in each country, which will help us to identify customers sales trend across different countries.

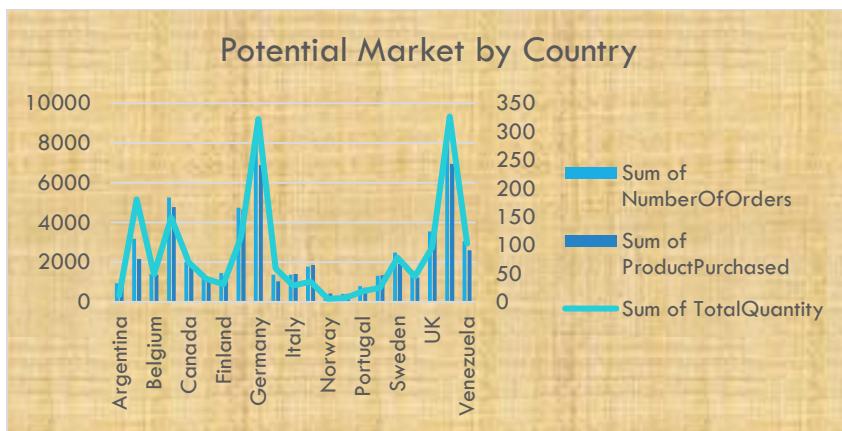
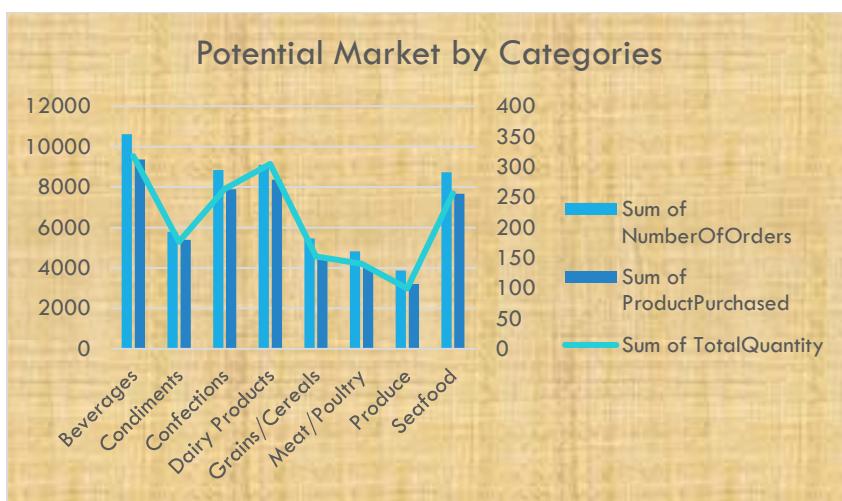
CustomerID
Countrywise data value

CustomerID	Country	Sum of TotalQuantity	Sum of TotalUnitOrder	Sum of TotalRevenue
FOLIG	Argentina	339	34	8119.1
FOLKO	Austria	5167	125	128003.84
FRANK	Belgium	1392	56	33824.85
FRANR	Brazil	4247	203	106925.77
FRANS	Canada	1984	75	50196.29
FURIB	Denmark	1170	46	32661.02
GALED	Finland	885	54	18810.05
GODOS	France	3254	184	81358.32
	Germany	9213	328	230284.62
	Ireland	1684	55	49979.9
	Italy	822	53	15770.15
	Mexico	1025	72	23582.08
	Norway	161	16	5735.15
	Poland	205	16	3531.95
	Portugal	533	30	11472.36
	Spain	718	54	17983.2

Question 3: - Are there any interesting patterns or clusters in customer behavior that can be visualized to identify potential market segments?

Answer:

```
select cu.CustomerID, cu.Country, c.CategoryName, count(Distinct p.ProductID) ProductPurchased,  
count(Distinct o.OrderID) NumberOfOrders, sum(od.Quantity) TotalQuantity  
from Customers cu  
join Orders o on cu.CustomerID = o.CustomerID  
join `order details` od on o.OrderID = od.OrderID  
join products p on p.ProductID = od.ProductID  
join categories c on c.CategoryID = p.CategoryID  
Group by 1,2,3
```

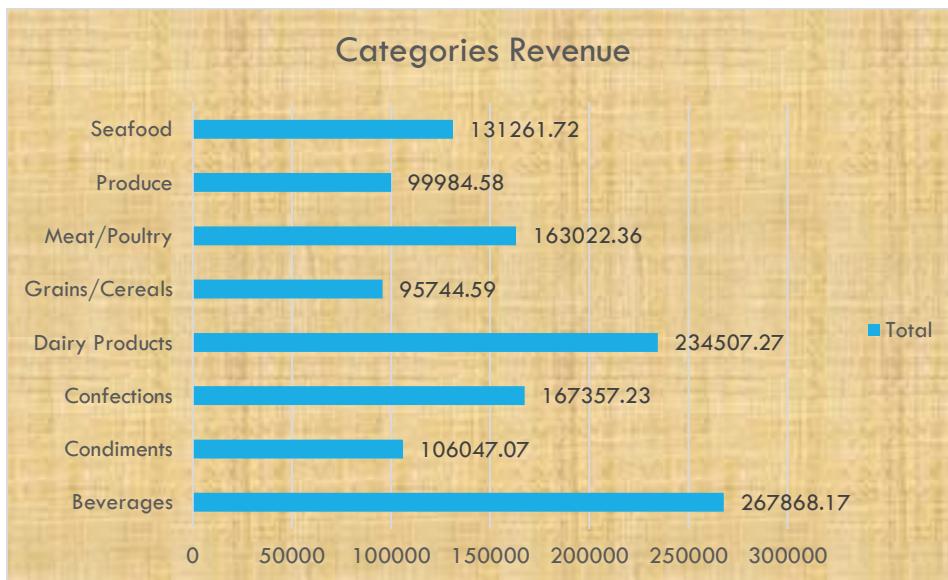


This provides us the market pattern of customer across different country and different categories, it tell us the total quantity purchased by customer in each categories which will provide us the best market segment to focus on.

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

Answer:

```
Select c.CategoryName, p.ProductName, round(SUM(od.UnitPrice*od.Quantity*(1-od.Discount)),2) as TotalRevenue
from products p
join categories c on p.CategoryID = c.CategoryID
join `order details` od on od.ProductID = p.ProductID
Group by 1,2
order by 3 desc;
```



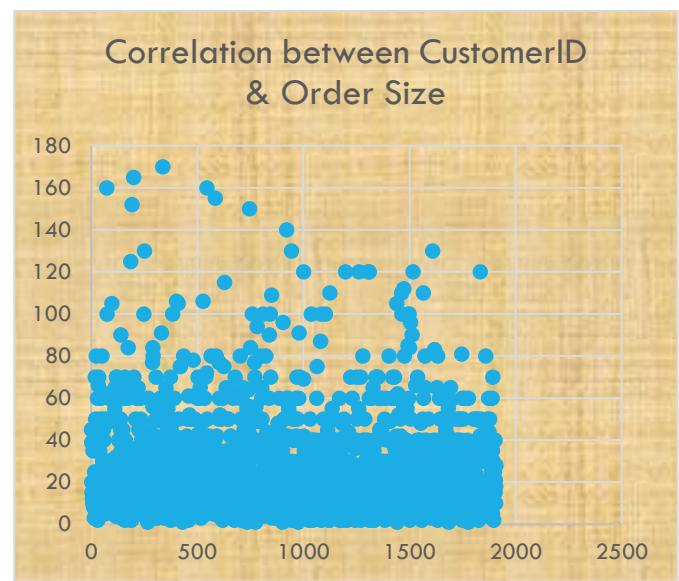
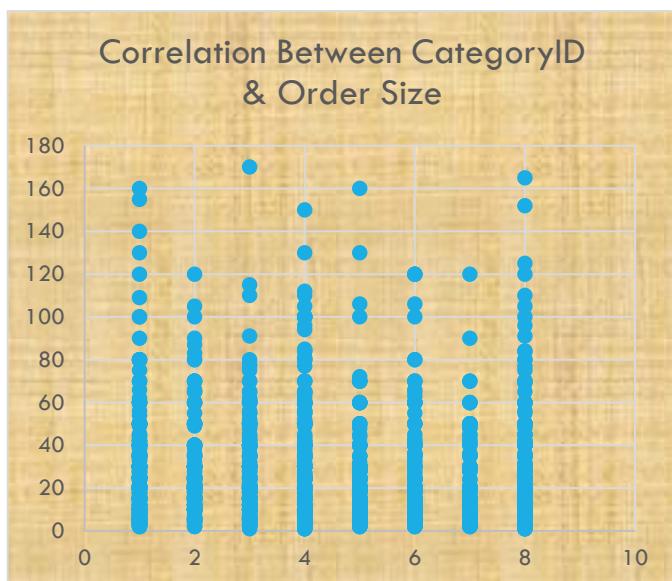
Category Revenue	
Product Category	Sum of TotalRevenue
Beverages	267868.17
Condiments	106047.07
Confections	167357.23
Dairy Products	234507.27
Grains/Cereals	95744.59
Meat/Poultry	163022.36
Produce	99984.58
Seafood	131261.72
Grand Total	1265792.99

This query will give order revenue generated by each category. It helps use to identify highest order revenue by categories. It shows Beverages have highest revenue of \$267868.

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

Answer:

```
select cu.Country, cu.CustomerID, ca.CategoryID, ca.CategoryName, od.OrderID, SUM(od.Quantity) as TotalOrderSize
from customers cu
join orders o
on cu.CustomerID = o.CustomerID
join `order details` od
on od.OrderID = o.OrderID
join products p
on od.ProductID = p.ProductID
join categories ca
on ca.CategoryID = p.CategoryID
group by 1,2,3,4,5
```



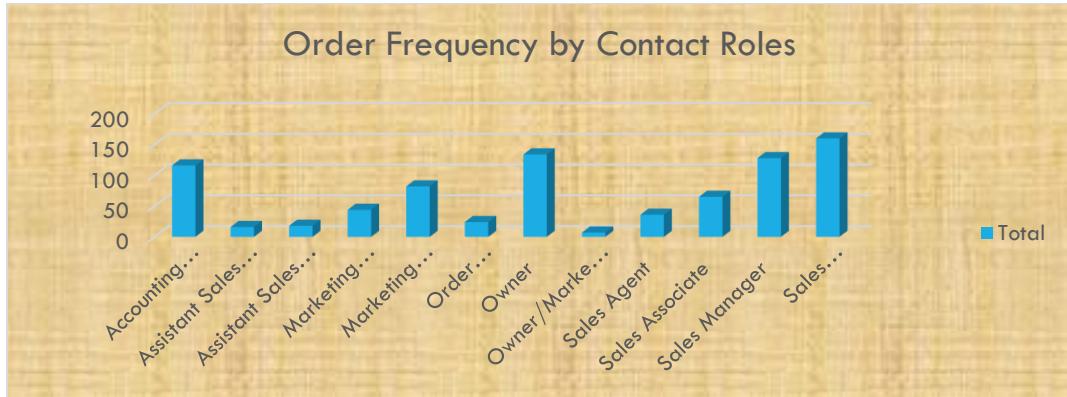
Category Order Size	
CategoryName	Sum of TotalOrderSize
Beverages	9532
Condiments	5298
Confections	7906
Dairy Products	9149
Grains/Cereals	4562
Meat/Poultry	4199
Produce	2990
Seafood	7681

This query helps us to create correlation between customers and order size. It also creates correlation between category and order size. Which display categoryID and amount of order done by it. And display most ordered Category name.

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

Answer:

```
select cu.CustomerID, cu.ContactTitle, cu.Country, COUNT(o.OrderID) as OrderFrequency
from customers cu
join orders o
on cu.CustomerID = o.CustomerID
group by 1,2,3
order by 4 desc;
```



This SQL command will provide us the order frequency across different customer segments. we have use 2 customer segments.

1. Customers Country

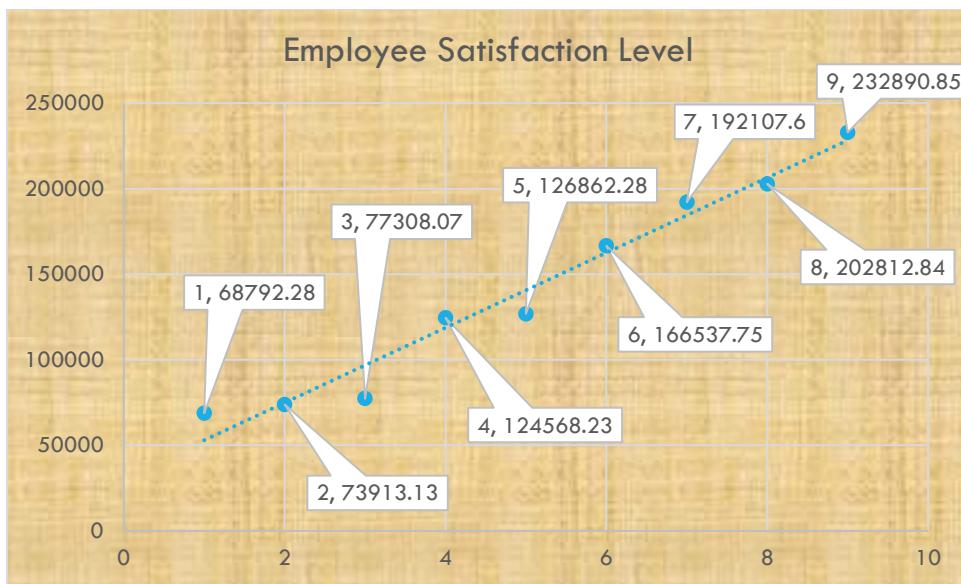
2. Customer Contact Role

It will show us the order frequency of customers across different country and different contact roles.

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

Answer:

```
select e.EmployeeID, concat(e.FirstName, " ", e.LastName) as Full_Name, count(distinct o.CustomerID) as NumOfCustomers,
round(sum(UnitPrice * Quantity * (1 - Discount)),2) as TotalSales
from employees e
join orders o
on e.EmployeeID = o.EmployeeID
join `order details` od
on od.OrderID = o.OrderID
group by 1
order by 4 desc;
```



Employee by TotalSales & Satisfaction Level			
Employee Full Name	Sum of NumOfCustomers	Sum of TotalSales	Sum of Satisfaction Level
Andrew Fuller	59	166537.75	6
Anne Dodsworth	29	77308.07	3
Janet Leverling	63	202812.84	8
Laura Callahan	56	126862.28	5
Margaret Peacock	75	232890.85	9
Michael Suyama	43	73913.13	2
Nancy Davolio	65	192107.6	7
Robert King	45	124568.23	4
Steven Buchanan	29	68792.28	1
Grand Total	464	1265793.03	

This command will provide us the Employee full name, total customers and total sales based on total sales we have given levels to each employee higher the sales higher the level based on that we have created scatter

chart to identify correlation patterns. It shows employee Margaret Peacock have highest. Total Sales \$232890 with satisfaction level 9.

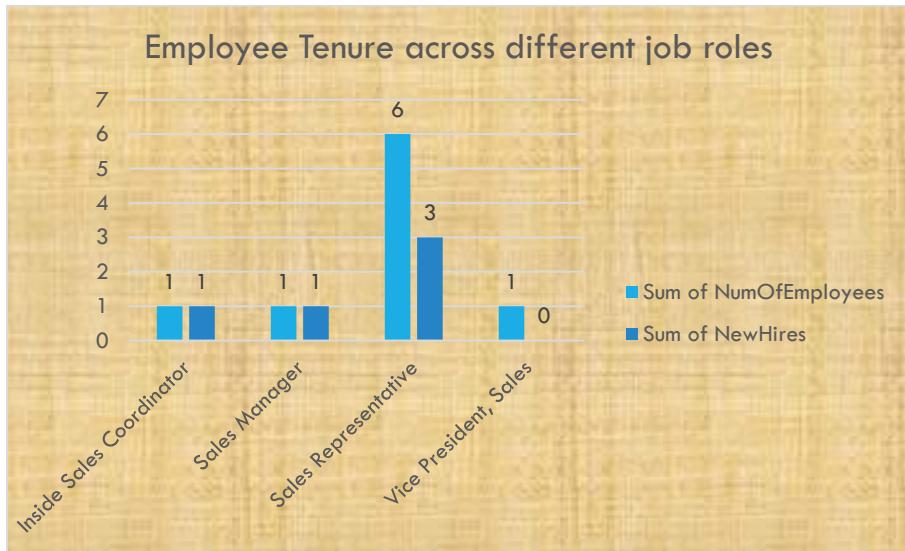
Question 8: - How does employee turnover vary across different departments or job roles? Can we visualize this using bar charts or heatmaps?

Answer:

```

SELECT
    e.Title as JobTitle,
    COUNT(e.EmployeeID) as NumOfEmployees,
    sum(CASE
        WHEN floor(timestampdiff(year,hiredate,MaxDate))<3
        THEN 1 ELSE 0 END) as NewHires
FROM Employees e
left join (select employeeId,max(orderdate) as MaxDate from orders
group by 1) md on md.EmployeeID = e.EmployeeID
GROUP BY 1;

```



Employee tenure across Job Role		
Job Role	Sum of NumOfEmployees	Sum of NewHires
Inside Sales Coordinator	1	1
Sales Manager	1	1
Sales Representative	6	3
Vice President, Sales	1	0
Grand Total	9	5

This SQL query will provide the last 3 years employee hire date. It will help us to analyse employee turnover across different roles. It shows Sales Representative have highest employee new hires and total employee number of employees.

Question 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?

Answer:

```
Select e.EmployeeID, concat(e.FirstName, " ", e.LastName, " ", e.Title) as Full_Name_Title,
count(distinct o.OrderID) as NumOfTasks, timestampdiff(year,hiredate,MaxDate) as Tenure,
round(SUM(od.UnitPrice*od.Quantity*(1-od.Discount)),2) as Total_Sales
from employees e
left join (
    select employeeID, max(orderdate) as MaxDate from orders
    group by 1) md on e.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
order by 5 desc;
```

EmployeeID	Full_Name_Title	NumOf	Tenure	Total_Sales
1	Nancy Davolio, Sales Representative	123	4	1728968.44
2	Andrew Fuller, Vice President, Sales	96	3	1498839.79
3	Janet Leverling, Sales Representative	127	4	1825315.59
4	Margaret Peacock, Sales Representative	156	3	2096017.61
5	Steven Buchanan, Sales Manager	42	2	619130.54
6	Michael Suyama, Sales Representative	67	2	665218.16
7	Robert King, Sales Representative	72	2	1121114.11
8	Laura Callahan, Inside Sales Coordinator	104	2	1141760.5
9	Anne Dodsworth, Sales Representative	43	1	695772.6

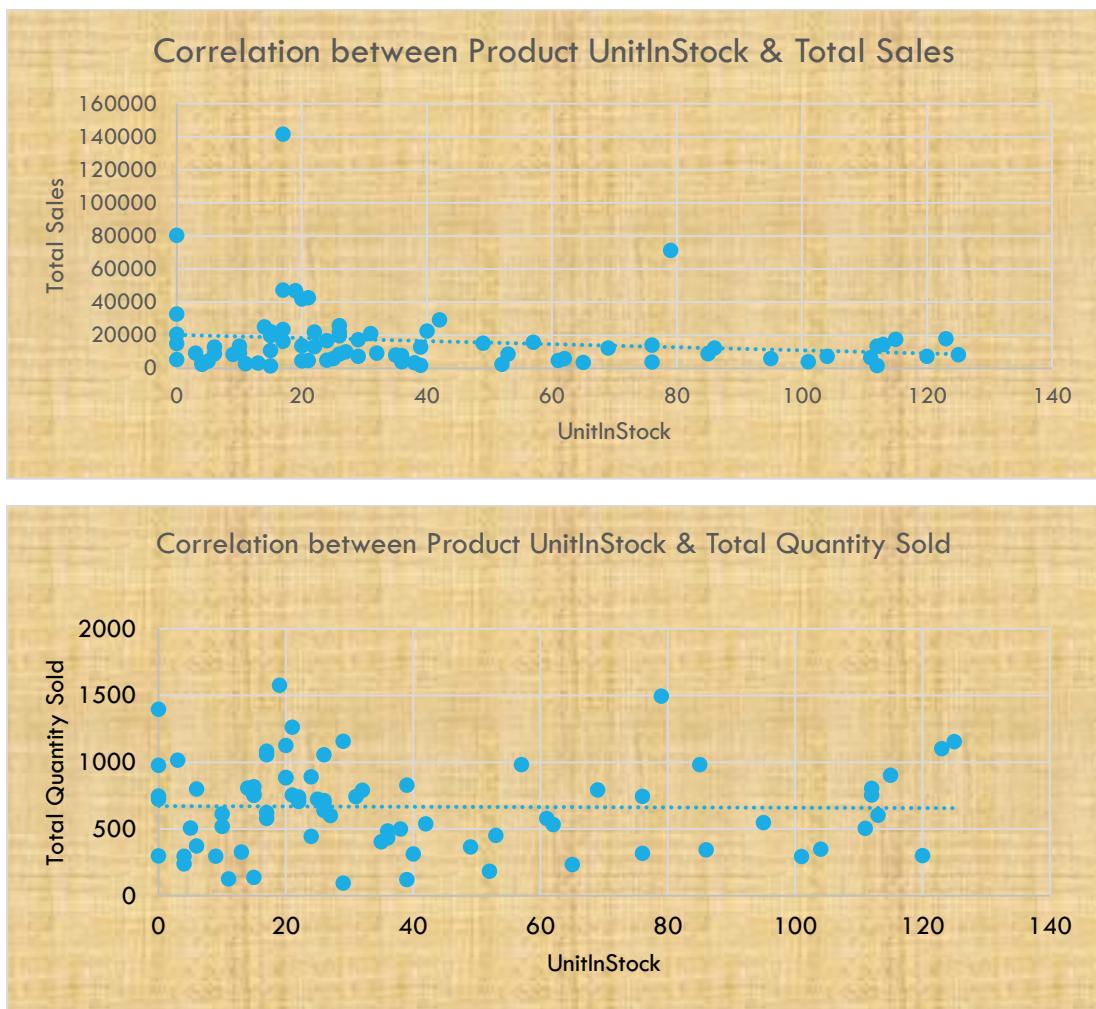


This shows skill set of employees with the help of total tasks and total sales. EmployeeID 4 – Margaret Peacock, Sales Representative have the highest number of tasks with total sales \$2096017.

Question 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?

Answer:

```
select p.ProductID, p.ProductName, p.UnitsInStock, sum(od.Quantity) as TotalQuantitySold,  
round(sum( od.UnitPrice * od.Quantity * ( 1 - od.Discount )),2) as TotalSales  
from products p join `order details` od  
on p.ProductID = od.ProductID  
group by 1  
order by 5 desc;
```

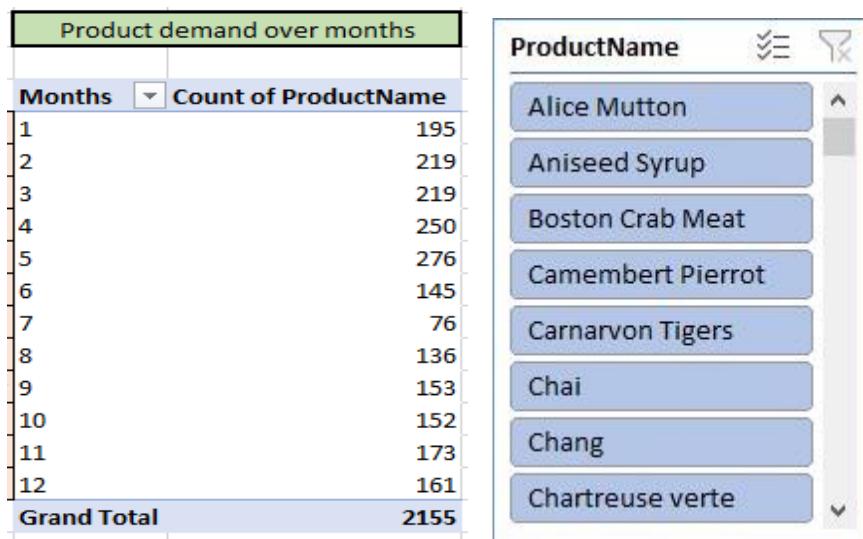


In this question we have to find any correlation between products and sales performance. It shows weak relation between product unit in stock and total quantity sold this is because the less promote or product is not useful.

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

Answer:

```
select p.ProductID, p.ProductName, month(o.OrderDate) as Months  
from products p  
left join `order details` od  
on p.ProductID = od.ProductID  
left join orders o  
on od.OrderID = o.OrderID
```



This query will provide the product demands in each month. It shows 5th month is having highest product demand with 276 product count.

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

Answer:

```
SELECT p.ProductName, SUM(od.Quantity) as TotalQuantity, round(sum( od.UnitPrice * od.Quantity * ( 1 - od.Discount )),2) as TotalSales
FROM Products p
JOIN `order details` od ON p.ProductID = od.ProductID
JOIN Orders o ON od.OrderID = o.OrderID
GROUP BY p.ProductName
ORDER BY p.ProductName;
```



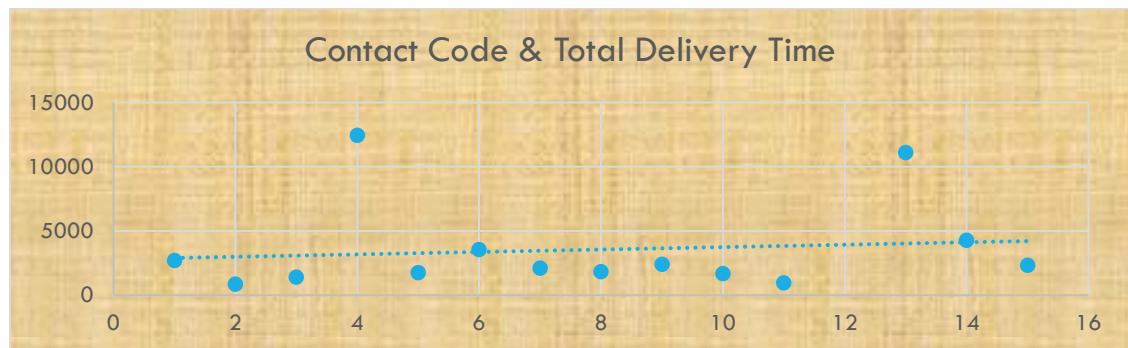
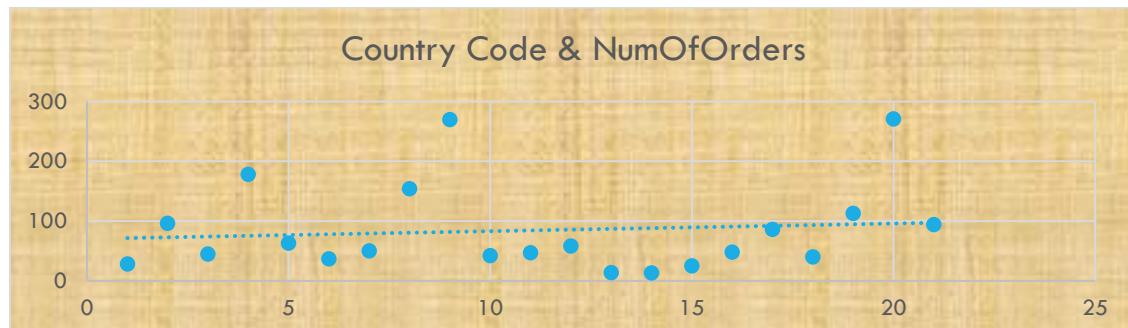
This query will provide the outliers in Total sales and Total quantity in product. Which will help us to analyse the product performance ranges based on total sales and total quantity.

Question 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?

Answer:

```

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='Sales 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;
  
```



This SQL command will show weak correlation between supplier attributes and its performance metrics. We have made scatter chart to identify supplier contact title and delivery time.

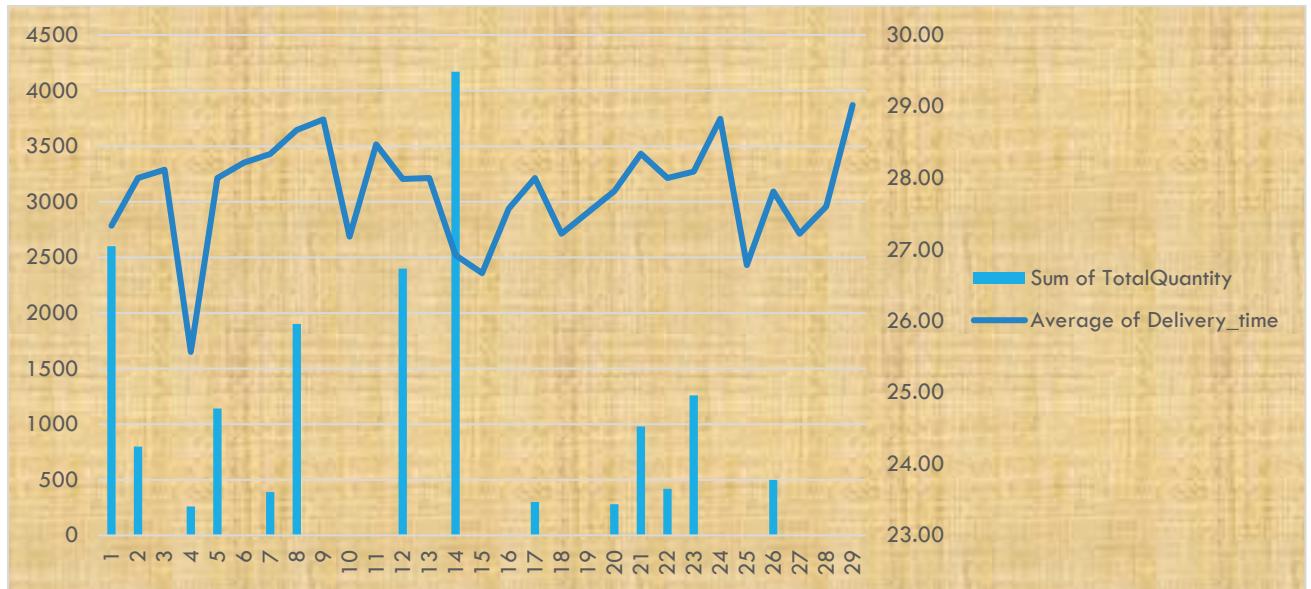
Question 14: - How does supplier performance vary across different product categories or departments? Can we visualize this using stacked bar charts or grouped column charts?

Answer:

```

SELECT
  distinct s.supplierID, s.CompanyName, s.ContactName, s.ContactTitle, c.CategoryName, SUM(p.UnitsOnOrder) 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;

```



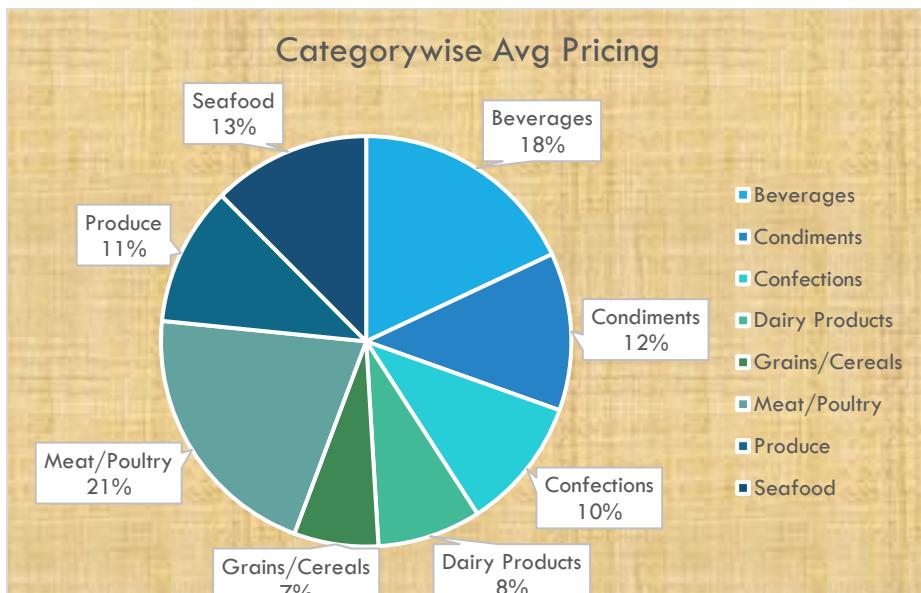
CategoryName	Filter
Beverages	
Condiments	
Confections	
Dairy Products	
Grains/Cereals	
Meat/Poultry	
Produce	
Seafood	

This SQL command will help us to identify supplier performance across different categories.
we have made combo chart to display supplier Ids and their average time and total quantity supplied by them in different product categories.

Question 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?

Answer:

```
select s.CompanyName as SupplierName, p.CategoryID, c.CategoryName, AVG(p.UnitPrice) as AvgPrice
from products p
join suppliers s
on p.supplierID = s.SupplierID
join categories c
on c.CategoryID = p.CategoryID
group by 1,2;
```



SupplierName	Supplier Avg Unit Price
Aux joyeux ecclésiastiques	265.8333333
Bigfoot Breweries	183.2
Cooperativa de Quesos 'L...	153.885
Escargots Nouveaux	120.4333333
Exotic Liquids	98
Forêts d'érables	308.315
Formaggi Fortini s.r.l.	161.85
Gai pâturage	183.415
Grand Total	1474.931667

This command will provide us the average pricing of each supplier in different categories. Which help us to compare pricing between different suppliers in different categories.

Conclusion

Customer Segmentation: By analyzing customer purchasing behavior and demographics (such as region, industry, etc.), Northwind Traders can segment their customers effectively. This segmentation could help in targeting specific customer groups with tailored marketing strategies or product offerings.

Sales Performance: Examining sales data, including total sales, trends over time, and sales by product category or region, can provide insights into which products are performing well and which may need improvement. Identifying top-selling products and regions can inform inventory management and sales strategies.

Employee Performance: Analyzing employee sales performance, such as total sales generated, number of orders processed, or average order value, can help identify top-performing employees and areas where additional training or support may be needed.

Inventory Trends: Monitoring inventory levels, including units in stock, units on order, and reorder levels, can help optimize inventory management processes. Identifying products that frequently run out of stock or have excess inventory can inform purchasing and stocking decisions.

Supplier Relationships: Evaluating supplier performance, such as on-time delivery rates or quality of products, can help Northwind Traders maintain strong supplier relationships and ensure a reliable supply chain.

Shipping Efficiency: Analyzing shipping data, including average shipping times and costs by shipper, can help optimize shipping processes and reduce delivery times and costs.

Overall Business Performance: By combining insights from all areas of analysis, Northwind Traders can gain a holistic view of their business performance. This can help identify areas of strength and opportunities for improvement, ultimately leading to better decision-making and strategic planning.

In conclusion, leveraging the data available in the Northwind database through Power BI can provide valuable insights into customer behaviour, sales patterns, employee performance, inventory management, supplier relationships, shipping efficiency, and overall business performance. These insights can empower Northwind Traders to make data-driven decisions that drive business growth and competitiveness in the wholesale market landscape.