CAPSTONE PROJECT – SALES ANALYSIS

Northwind Traders Sales Data Analysis

ABSTRACT

This project employs data analysis and visualization on Northwind Traders' dataset to reveal insights into customer behavior, sales patterns, and employee performance, promoting data-driven decision-making for business growth.

Aditi Arya

Overview

A sales analysis project involves examining sales data to gain insights and make data-driven decisions. It includes data collection, cleaning, exploration, defining metrics, comparative analysis, forecasting, customer and competitor analysis, and ends with recommendations and action plans to improve sales performance. It's an ongoing process that helps a company optimize strategies and increase revenue.

This project focuses on conducting exploratory data analysis (EDA) and visualization to gain insights from a dataset containing information about customers, orders, employees, products, suppliers, and categories from the Northwind Traders company. The objectives of the analysis are to understand customer behavior, sales patterns, employee performance, and supplier metrics to inform decision-making processes and drive business growth. Through the exploration of various EDA questions, such as customer retention factors, customer preferences, employee turnover, product performance, and supplier attributes, valuable insights are obtained. Interactive visualizations, including charts, graphs, and maps, are utilized to effectively communicate the findings and facilitate data exploration. The project highlights the importance of data-driven decision-making and demonstrates the power of EDA and visualization techniques in uncovering actionable insights for business success.

DATA DICTIONARY

1. **Table 1: categories**

Fields: CategoryID: every category name has unique category number.

CategoryName : name of the category

Description : It consist of products name according to categoryname

2. **Table 2: customers**

Fields: Address : location of customer where they live.

City: city location of the customer

CompanyName : company name of the customer

ContactName : name of company owner or representative

ContactTitle: job post title

Country: country location of the customer

CustomerID: unique no. of every customer

Fax: fax details of the customer

Phone: phone number of customer

PostalCode : postal code of customer

Region : region location of the customer

3. **Table 3: employees**

Fields: : Address

City : location city of the employee

Country :location country of employee

Region: location region of employee

BirthDate: birth date of the employee

EmployeeID: unique employee number for every employee

Extension :unique number

FirstName: first name of the employee

HireDate: hire date of the employee

HomePhone :home phone number of the employee

LastName ;last name of the employee

PostalCode : postal code of the employee

Region: location of the employee

Reportsto : reporting number of the employee

Title :job title of the employee

TitleOfcourtesy: courtesy title of the employee

4. **Table 4: order details**

Fields: Discount : discount percentage in the employee

OrderID: unique order id for orders

ProductID :unique number id for every product

Quantity :quantity number of every product

UnitPrice: price of every product

5. **Table 5: orders**

Fields: CustomerID: unique no. of the customer

EmployerID:unique no. of the employeer

Freight: goods carried by ship

OrderDate: order placed date

RequiredDate: expecting date of order

ShipCity: customer city

ShipAddress: address of the customer

ShipCountry :country of the customer

ShipName: name of the ship where goods are boarded

ShipPostalCode: postal code of ship

Shipregions: regionof the ship where they have to board

6. **Table 6 : products**

Fields: CategoryID unique id of the categories

Discontinued: goods which are not available

ProductID: unique product id no of the product

ProductName :name of the product

QuantityPerUnit quantity of product order

ReorderLevel :reorder quantity of the product

SupplierID: unique no. of the suppliers

UnitPrice: price of the product per items

UnitStock; number of product quantity we have

UnitOnOrder

7. **Table 7: shippers**

Fields: CompanyName : company name of the shipper id

Phone :phone number of the shipper company

ShipperID: unique id for the shipper company name

8. **Table 8: suppliers**

Fields: Address: address of the supplier

City: Location city of the supplier

CompanyName: company namr of the supplier

ContactName: Companys representative person

Country:location country of supplier

Fax :fax no. of supplier

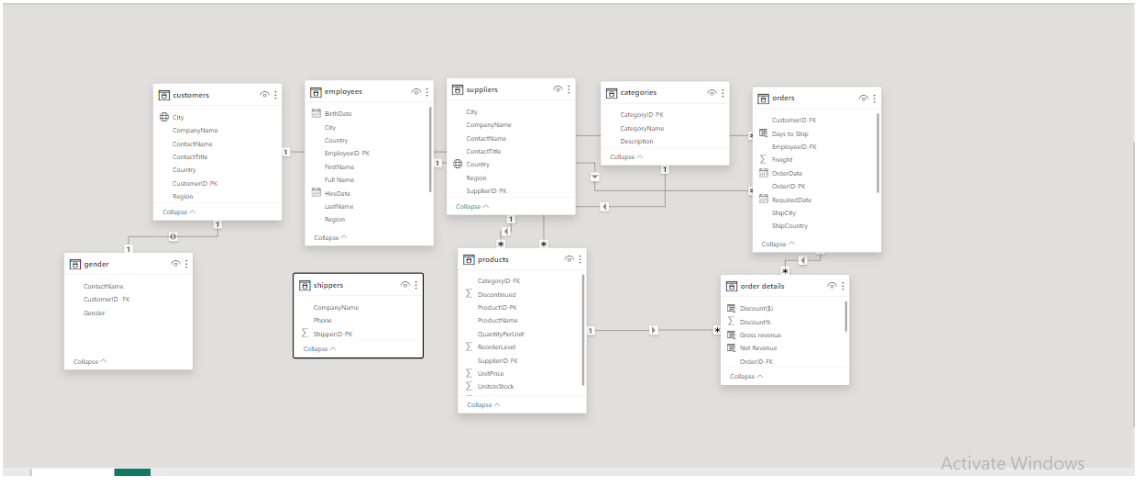
Phone: phone nuber of the supplier

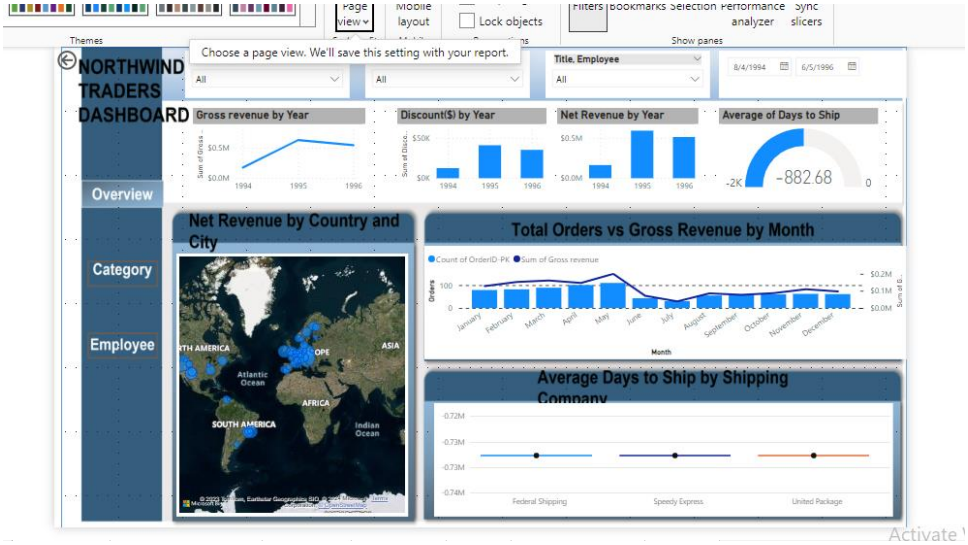
PostalCode: postal codeoof the supplier

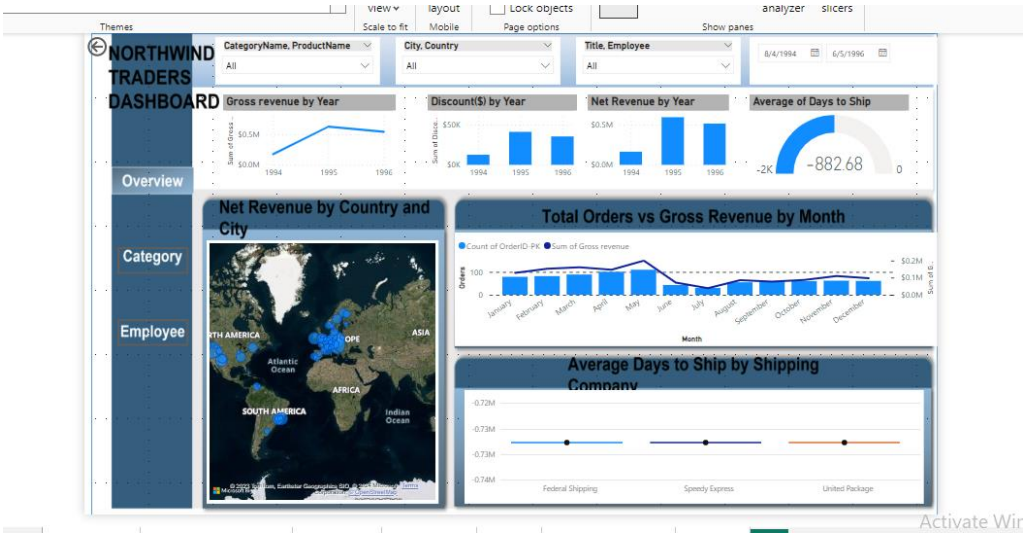
Region: location region of the supplier

SupplierID:unique supplier id of the supplier

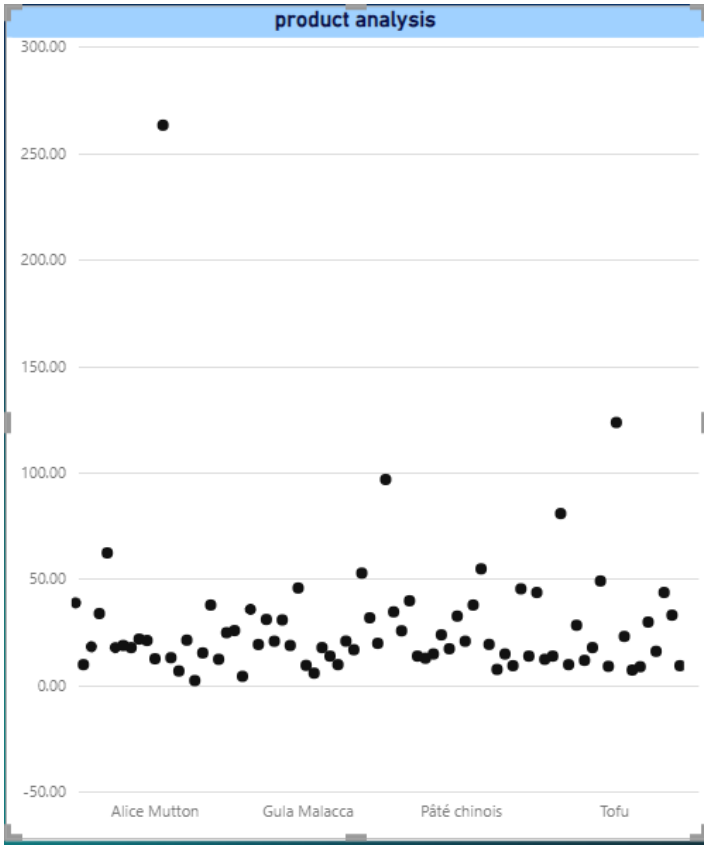
ER DAIGRAM:


customers
City
CompanyName
ContactName
ContactTitle
Country
CustomerID PK
Region
Collapse
employees
BirthDate
City
Country
EmployeeID PK
FirstName
Full Name
HireDate
LastName
Region
Collapse
suppliers
City
CompanyName
ContactName
ContactTitle
Country
Region
SupplierID PK
Collapse
categories
CategoryID PK
CategoryName
Description
Collapse
orders
CustomerID FK
Days to Ship
EmployeeID FK
Freight
OrderDate
OrderID PK
RequiredDate
ShipCity
ShipCountry
Collapse
gender
ContactName
CustomerID FK
Gender
Collapse
shippers
CompanyName
Phone
ShipperID PK
Collapse
products
CategoryID FK
Discontinued
ProductID PK
ProductName
QuantityPerUnit
ReorderLevel
SupplierID FK
UnitPrice
UnitsInStock
Collapse
order details
Discount($)
Discount%
Gross revenue
Net Revenue
OrderID FK
Collapse
Activate Windows


Themes
Page view
Mobile layout
Lock objects
Filters
Bookmarks
Selection
Performance analyzer
Sync slicers
Show panes
Choose a page view. We'll save this setting with your report.
NORTHWIND TRADERS DASHBOARD
Title, Employee
All
8/4/1994
6/5/1996
Gross revenue by Year
Discount($) by Year
Net Revenue by Year
Average of Days to Ship
-882.68
Overview
Category
Employee
Net Revenue by Country and City
Total Orders vs Gross Revenue by Month
Count of OrderID-PK
Sum of Gross revenue
Month
Average Days to Ship by Shipping Company
Federal Shipping
Speedy Express
United Package
Atlantic Ocean
Indian Ocean
AFRICA
SOUTH AMERICA
ASIA



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

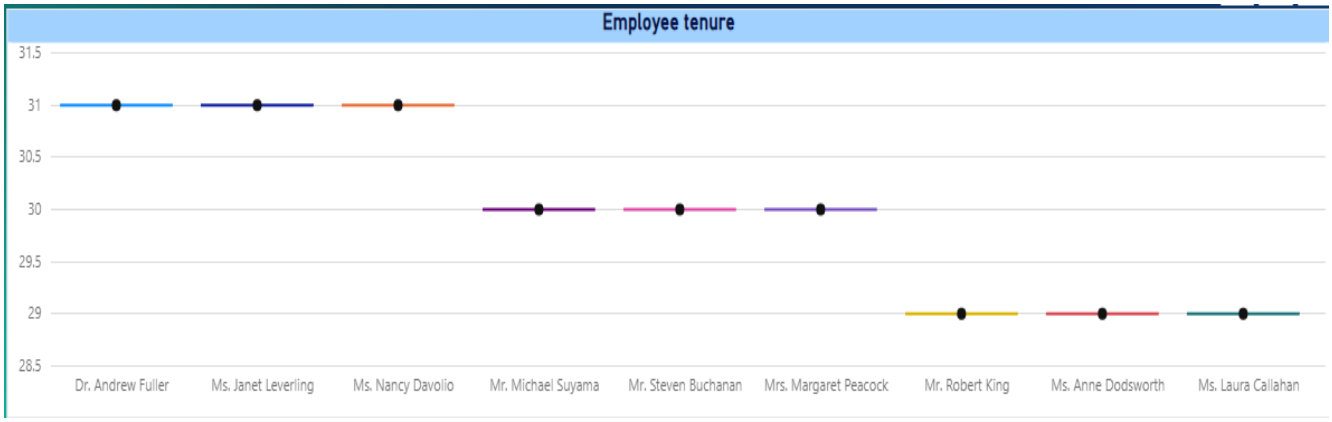


Here is the price distribution of products ,

and we can see that Alice Mutton is at the highest

You can visualize customer distribution across different regions on a map and across customer segments using a bar chart. Use a map to show the geographical distribution by region and a bar chart to compare customer counts across segments or non-geographical categories.

Here, North America region has the maximum customers

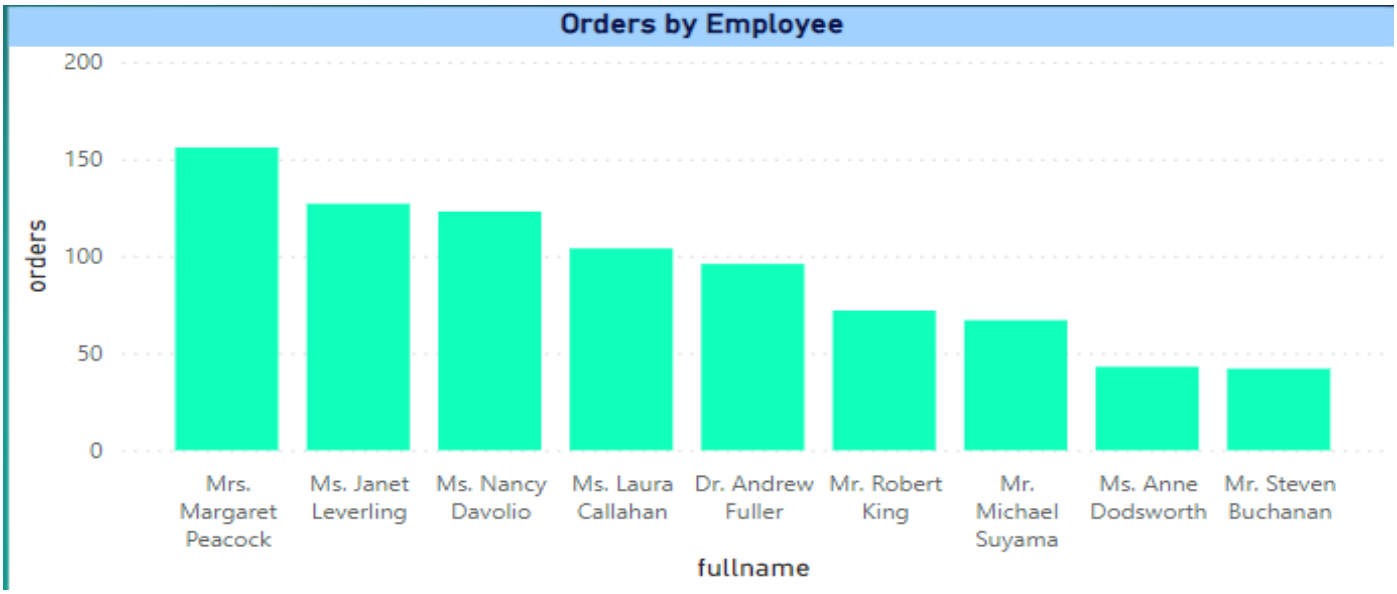


The distribution of employee tenure can be displayed using a histogram or a box plot, which provides insights into the spread and central tendency of employee tenure lengths.

Here the average tenure can be taken as 30 years

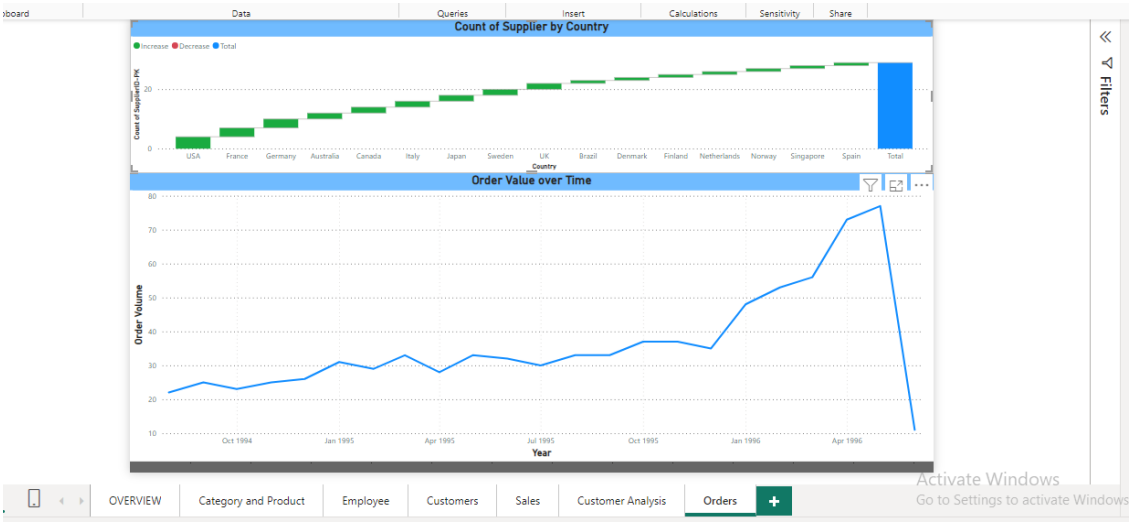
And maximum tenure is 31 years

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?

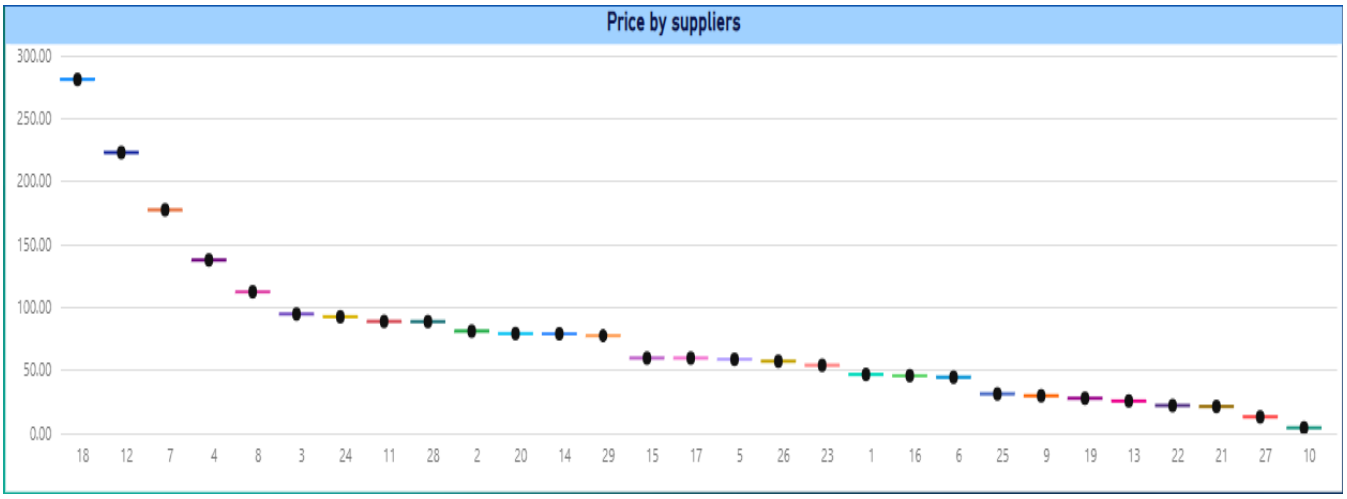


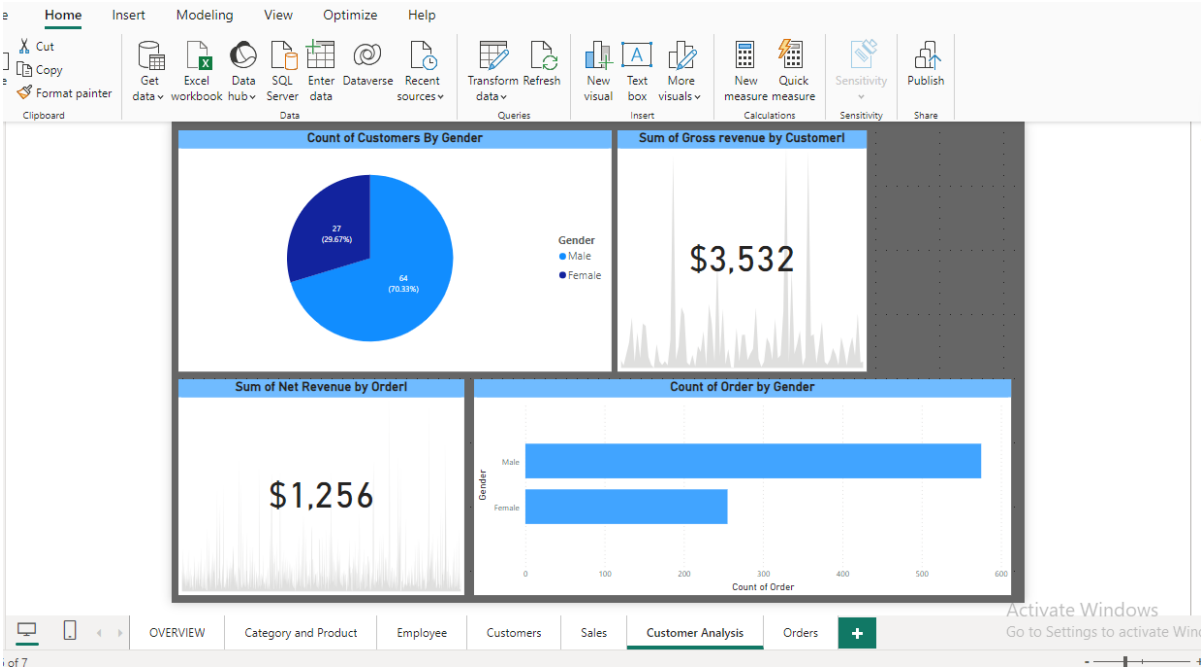
To visualize how employee productivity varies across different departments or job roles, you can create a stacked bar chart or grouped column chart, which allows for easy comparison of productivity levels within these categories

Here Mrs. Margaret Peacock has sold maximum product and she is from Sales Representative

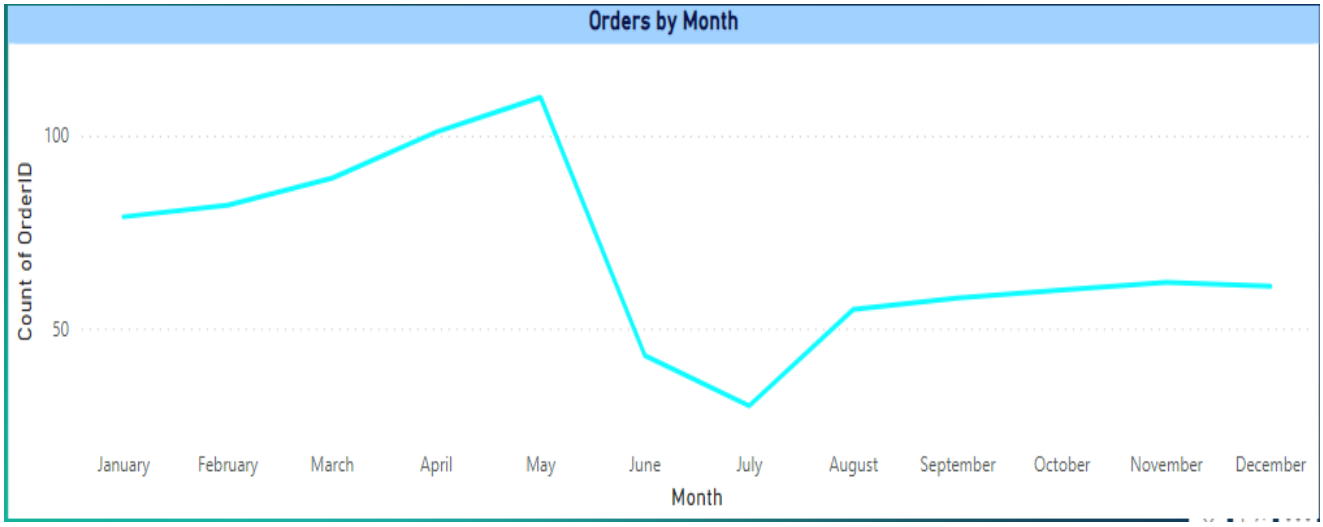

Data
Queries
Insert
Calculations
Sensitivity
Share
Count of Supplier by Country
Increase
Decrease
Total
Count of SupplierID-PK
20
0
USA
France
Germany
Australia
Canada
Italy
Japan
Sweden
UK
Brazil
Denmark
Finland
Netherlands
Norway
Singapore
Spain
Total
Country
Order Value over Time
Order Volume
80
70
60
50
40
30
20
10
Oct 1994
Jan 1995
Apr 1995
Jul 1995
Oct 1995
Jan 1996
Apr 1996
Year
Filters
OVERVIEW
Category and Product
Employee
Customers
Sales
Customer Analysis
Orders
Activate Windows
Go to Settings to activate Windows

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




Home
Insert
Modeling
View
Optimize
Help
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
Data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New measure
Quick measure
Calculations
Sensitivity
Publish
Share
Count of Customers By Gender
27
(29.67%)
64
(70.33%)
Gender
Male
Female
Sum of Gross revenue by CustomerI
$3,532
Sum of Net Revenue by OrderI
$1,256
Count of Order by Gender
Male
Female
Gender
0
100
200
300
400
500
600
Count of Order
OVERVIEW
Category and Product
Employee
Customers
Sales
Customer Analysis
Orders
Activate Windows
Go to Settings to activate Win
of 7

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

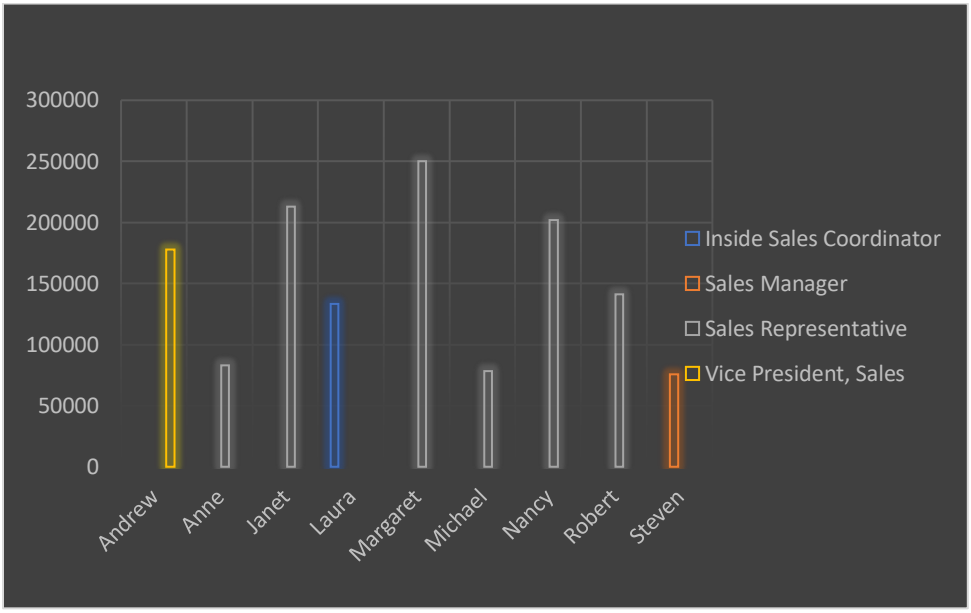


To visualize changes in order volume over time, you can create a time series chart, which is ideal for showing trends in order volume. Stacked bar charts are typically used for different purposes, like showing composition within a single time period.

Here May month has maximum order volume and July has minimum order volume

EDA PROBLEM STATEMENTS

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




```
use sales_analysis;
select e.EmployeeID,e.FirstName,e.Title , SUM(od.UnitPrice*od.Quantity) as revenue
 from employees e
join orders o on
o.EmployeeID= e.EmployeeID
join orderdetails od
on od.OrderID = o.OrderID
group by  1 ;
```

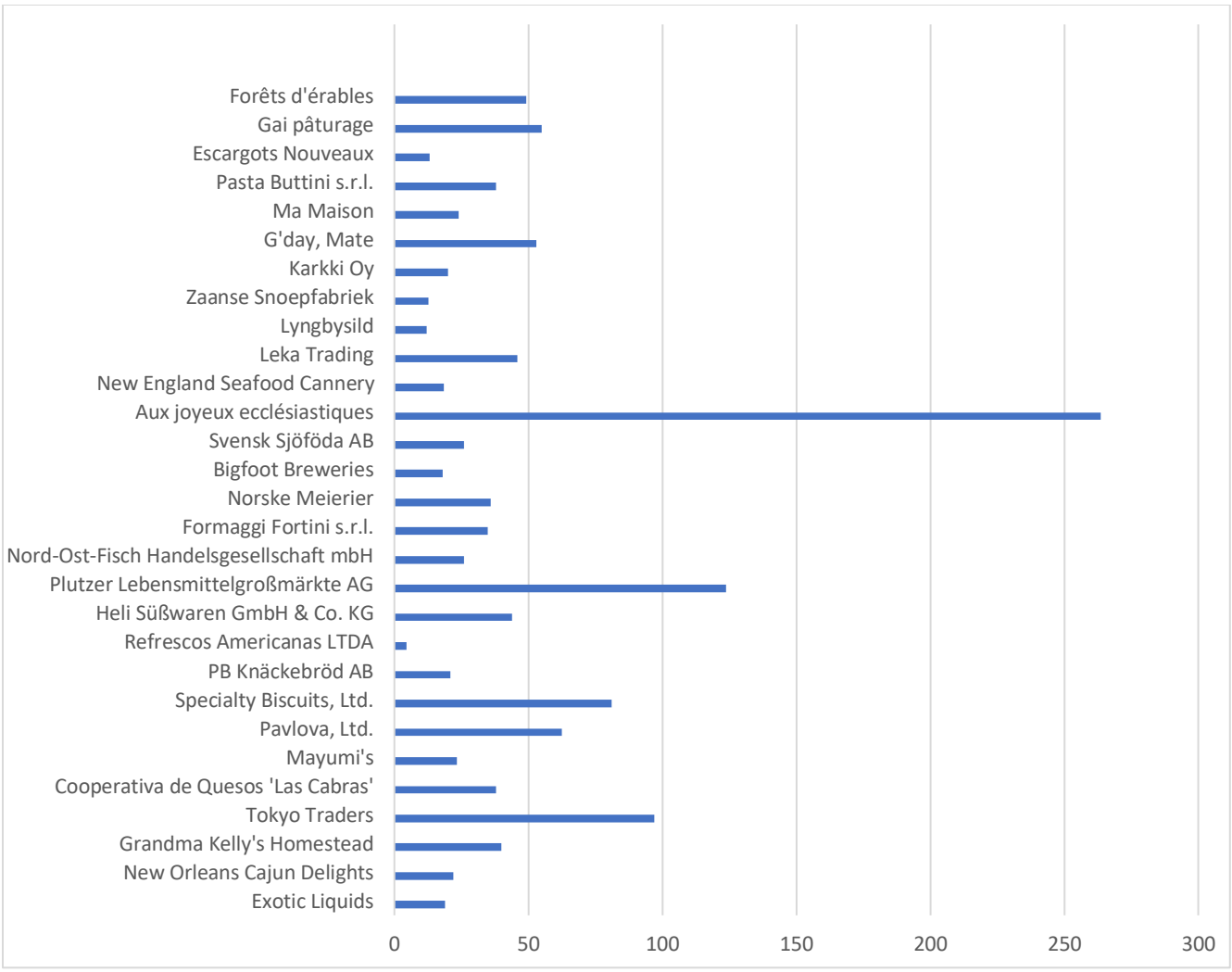
Sum of revenve	Column Labels				
Row Labels	Inside Sales Coordinator	Sales Manager	Sales Representative	Vice President, Sales	Grand Total
Andrew				177749.26	177749.26
Anne			82964		82964
Janet			213051.3		213051.3
Laura	133301.03				133301.03
Margaret			250187.45		250187.45
Michael			78198.1		78198.1
Nancy			202143.71		202143.71
Robert			141295.99		141295.99
Steven		75567.75			75567.75
Grand Total	133301.03	75567.75	967840.55	177749.26	1354458.59

Here, there are 6 employees in sales representative and make 967840.55 revenue which more than half of total revenvue

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?

```
use sales_analysis;

select  p.SupplierID,s.CompanyName,max(p.UnitPrice)
 from products p left join suppliers s on s.SupplierID = p.SupplierID
 group by 1;
```

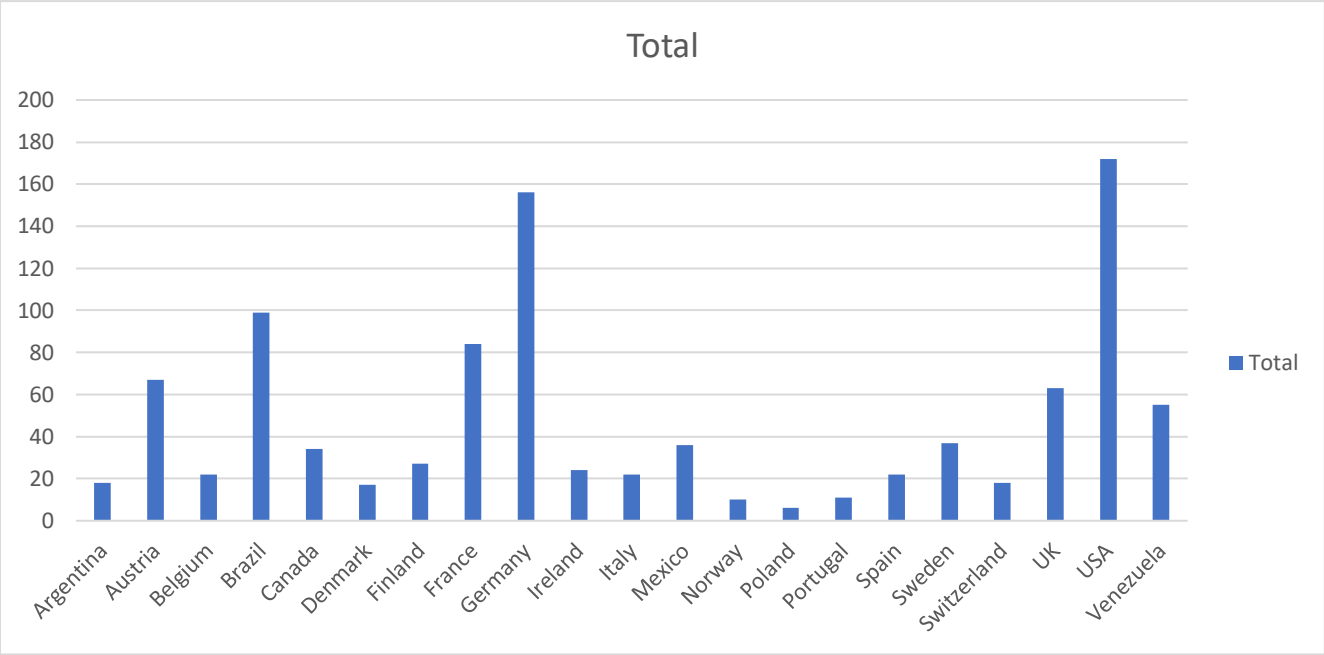

Forêts d'érables
Gai pâturage
Escargots Nouveaux
Pasta Buttini s.r.l.
Ma Maison
G'day, Mate
Karkki Oy
Zaanse Snoepfabriek
Lyngbysild
Leka Trading
New England Seafood Cannery
Aux joyeux ecclésiastiques
Svensk Sjöföda AB
Bigfoot Breweries
Norske Meierier
Formaggi Fortini s.r.l.
Nord-Ost-Fisch Handelsgesellschaft mbH
Plutzer Lebensmittelgroßmärkte AG
Heli Süßwaren GmbH & Co. KG
Refrescos Americanas LTDA
PB Knäckebröd AB
Specialty Biscuits, Ltd.
Pavlova, Ltd.
Mayumi's
Cooperativa de Quesos 'Las Cabras'
Tokyo Traders
Grandma Kelly's Homestead
New Orleans Cajun Delights
Exotic Liquids
0
50
100
150
200
250
300

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

these are the customers with repeating order and factors can be

1. order is shipped after 5 6 average days of order date
2. giving discount in low sales regions to promote products
3. customer repeating the order it means the product quality is good

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




```
use sales_analysis ;

select CustomerID , count(*) as repeated FROM orders
group by 1
having repeated > 1
order by repeated asc;
```

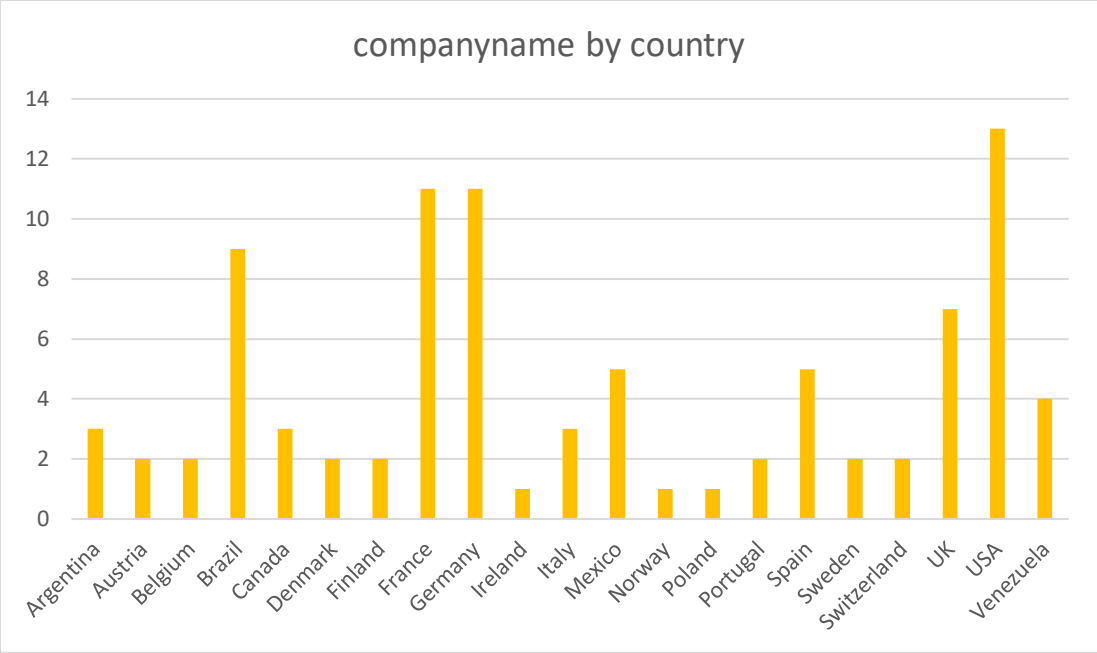
product name	(All)

Row Labels	Count of product name
Argentina	18
Austria	67
Belgium	22
Brazil	99
Canada	34
Denmark	17
Finland	27
France	84
Germany	156
Ireland	24
Italy	22
Mexico	36
Norway	10
Poland	6
Portugal	11
Spain	22
Sweden	37
Switzerland	18
UK	63
USA	172
Venezuela	55
Grand Total	1000

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

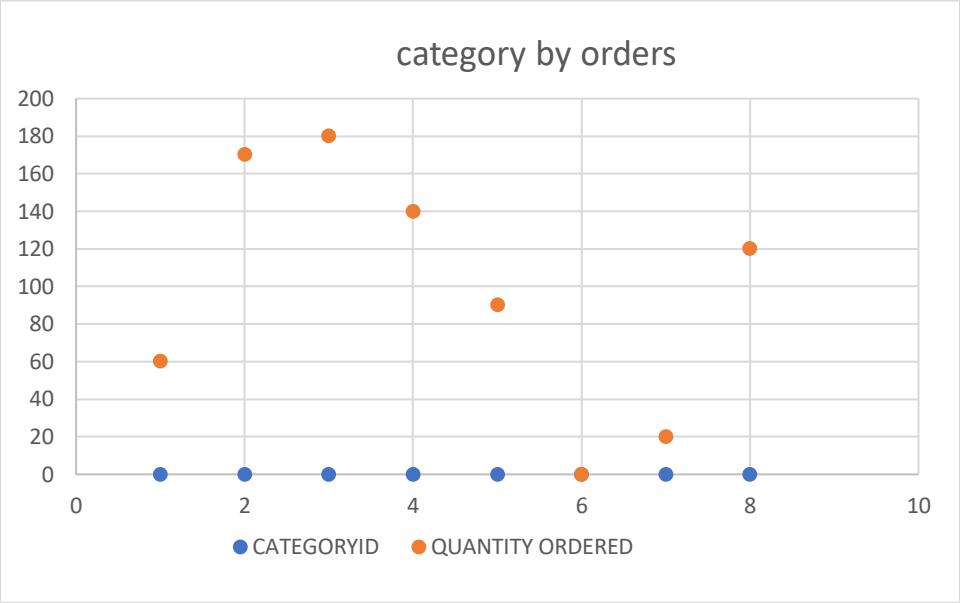
country	Count of companyname
Argentina	3
Austria	2
Belgium	2
Brazil	9
Canada	3
Denmark	2
Finland	2
France	11
Germany	11
Ireland	1
Italy	3
Mexico	5
Norway	1
Poland	1
Portugal	2
Spain	5
Sweden	2
Switzerland	2
UK	7
USA	13
Venezuela	4

```
use sales_analysis;
select CompanyName, City , Country from customers;
```


companyname by country
14
12
10
8
6
4
2
0
Argentina
Austria
Belgium
Brazil
Canada
Denmark
Finland
France
Germany
Ireland
Italy
Mexico
Norway
Poland
Portugal
Spain
Sweden
Switzerland
UK
USA
Venezuela

CATEGORYID	QUANTITY ORDERED
Beverages	60
Condiments	170
Confections	180
Dairy Products	140
Grains/Cereals	90
Meat/Poultry	0

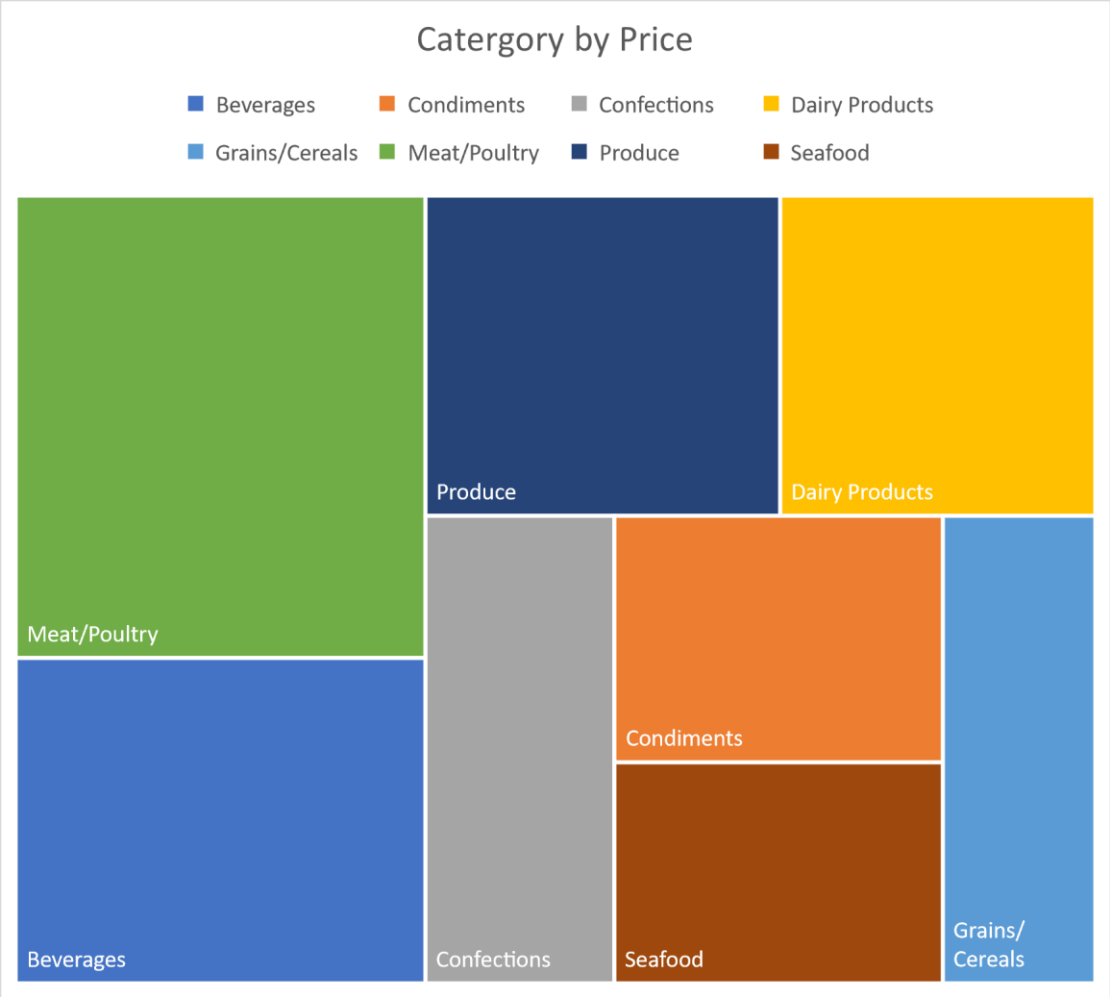
5.	Produce	20	Are there any correlations between order size and customer demographics or product categories? Can we explore this visually using scatter plots or heatmaps?
	Seafood	120	




```
use sales_analysis;
select c.CategoryName,sum(UnitsOnOrder)  AS QUANTITY_ORDERED from products p
join categories c on c.CategoryID = p.CategoryID

GROUP BY 1;
```

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



categoryid	average unit price
Beverages	37.97916667
Condiments	23.0625
Confections	25.16
Dairy Products	28.73
Grains/Cereals	20.25
Meat/Poultry	54.00666667

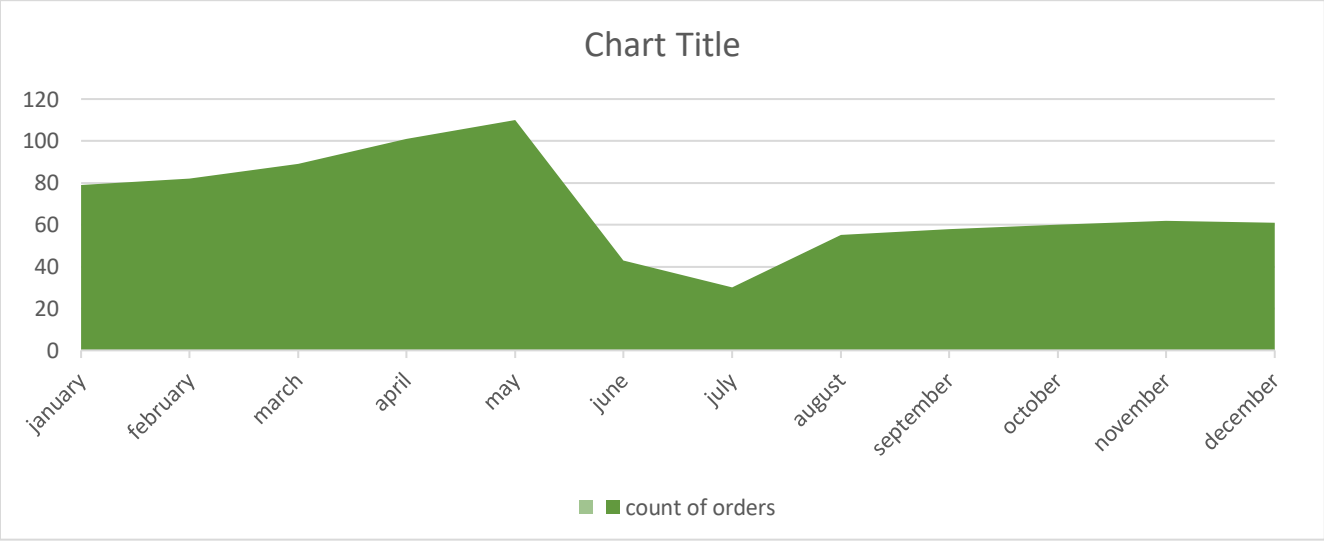
Produce	32.37
Seafood	20.6825

```
use sales_analysis ;

select CategoryName
 , avg(UnitPrice) from products P join categories c on c.CategoryID = P.CategoryID
group by 1;
```

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

month name	count of orders
january	79
february	82
march	89
april	101
may	110
june	43
july	30
august	55
september	58
october	60
november	62
december	61



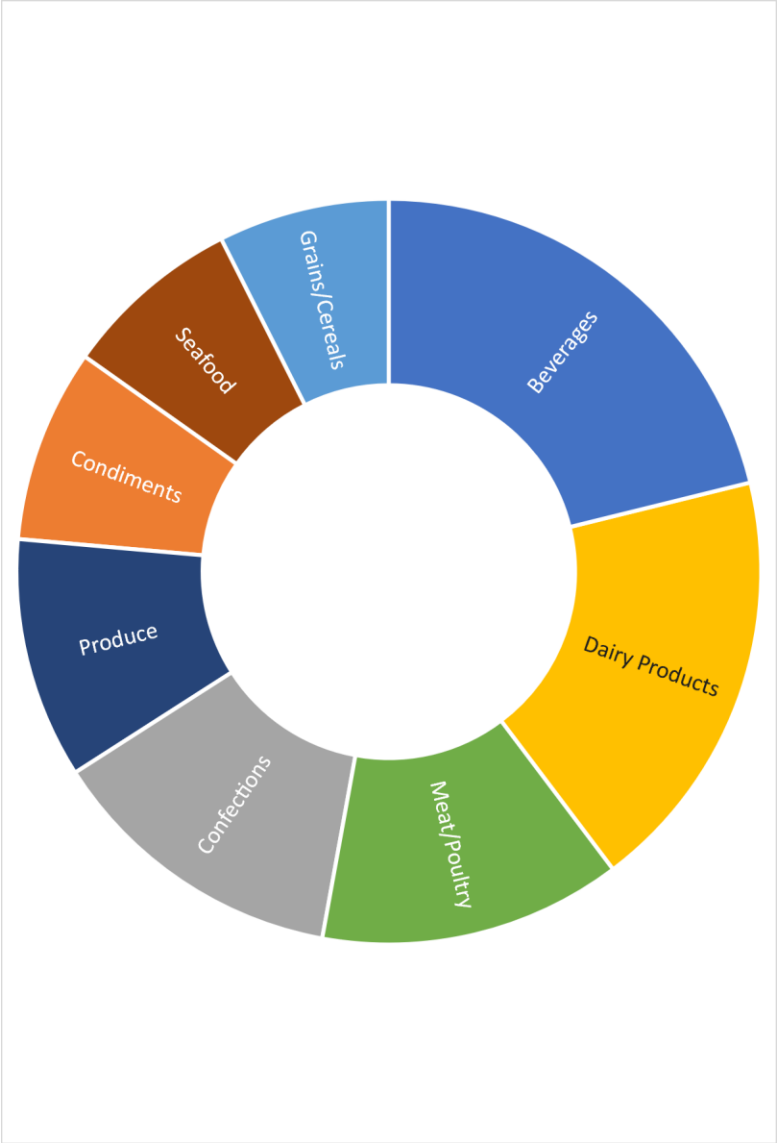

```
USE sales_analysis;

select count(OrderID), month(date(OrderDate)) as month FROM orders
group by month(date(OrderDate))
order by month asc ;
```

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

```
USE sales_analysis;
select c.CategoryName ,SUM(UnitPrice* UnitsOnOrder) AS REVENUE
 from products P
 join categories c on c.CategoryID = p.CategoryID
 GROUP BY 1;
```

category	revenue
Beverages	₹ 2,86,526.95
Condiments	₹ 1,13,694.75
Confections	₹ 1,77,099.10
Dairy Products	₹ 2,51,330.50
Grains/Cereals	₹ 1,00,726.80
Meat/Poultry	₹ 1,78,188.80
Produce	₹ 1,41,623.09
Seafood	₹ 1,05,268.60



Here, beverages have highest sales revenue and produce have lowest sales revenue