

|  |
| --- |
| Capstone project – Sales Analysis |
|  |
| November 7  Data analytics  Authored by: Sanket meshram |

# 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. |
| “It is a capital mistake to theorize before one has data” |
|  |

1. DATA ACQUISITION FROM GITHUB:

*THE PROCESS*

Obtain the requisite dataset from a designated GitHub repository,

Containing essential information on sales analysis,

Encompassing various countries and their performance across distinct ranking systems.

1. DATA TRANSFORMATION AND ENHANCEMENT:

If necessary, execute data transformation procedures to ensure data

Quality and consistency. Additionally, consider augmenting the

Dataset with new problem statement to enrich the analysis potential.

3. CONNECTING WITH TOOLS:

Establish connections between the dataset and various analytical tools. interface the dataset with POWER BI, EXCEL and MySQL Workbench, facilitating seamless data integration and processing.

4.PROBLEM STATEMENT SOLUTION IN POWER BI:

Utilize power bi to delve into the specifies problem statement. employ its robust features for data visualization, exploration and analysis, effectively deriving insights and solutions.

5. EXPLORATORY DATA ANALYSIS (EDA):

Perform exploratory data analysis using either excel or SQL workbench, depending on the complexity of the analysis. Extract meaningful patterns, relationship, and trends from the data to inform subsequent decision making

6. CREATION OF VISUAL AND INSIGHTFUL POWER POINT:

Develop a comprehensive power point presentation that encapsulates the project’s objective, methodologies, problem statement solutions, and key visualization. Each problem statement should be accompanied by a dedicated section with pertinent conclusions and insights.

7. DETAILED DOCUMENTATION:

Compile a detailed report that meticulously document the entire project lifecycle. Include sections on data collection, transformation, problem statements formulation, tools integration, power BI solutions, EDA insights and power point solutions, EDA insights, and Power Point visualization.

OBJECTIVE

Performance Evaluation: Assess the overall sales performance to determine if the company is meeting its revenue targets and growth objectives.

Trend Identification: Identify sales trends over time to understand the direction of the business and anticipate future sales patterns.

Product Analysis: Determine which products are top sellers, underperforming, or have growth potential to optimize inventory and marketing efforts.

Customer Segmentation: Segment customers based on demographics, behavior, or purchase history to tailor marketing strategies and enhance customer experiences.

Pricing Strategy: Analyze the impact of pricing changes on sales and profitability to optimize pricing strategies.

Geographic Analysis: Understand regional variations in sales to allocate resources effectively and identify opportunities in specific markets.

Seasonal Trends: Recognize and plan for seasonal fluctuations in sales to ensure adequate inventory and marketing efforts.

Sales Forecasting: Create accurate sales forecasts to improve inventory management and resource allocation.

Profit Margin Analysis: Assess the profitability of different products, customer segments, or sales channels to make informed decisions about pricing and resource allocation.

Inventory Management: Optimize inventory levels to meet demand without overstocking or running out of products.

Conversion Rate Analysis: Evaluate the effectiveness of the sales funnel and identify areas for improvement in the conversion process.

Operational Efficiency: Analyze the efficiency of sales and distribution operations to reduce costs and improve overall effectiveness.

These objectives will guide your data analysis efforts and help you derive actionable insights from your sales data to make informed business decisions. The specific objectives should align with your company's strategic goals and the challenges it faces in the market.

SIGNIFICANCE

[Sales analysis is a structured review of a company’s sales data that considers all aspects of the sales process, including the underlying factors driving sales, the efficacy of key performance indicators (KPIs), and whether or not resources are being used effectively](https://www.pipefy.com/blog/sales-analysis/) .

[It plays a critical role in business as it allows teams to uncover new opportunities and identify and solve problems in their sales process 1](https://www.pipefy.com/blog/sales-analysis/). [A thorough sales analysis can reveal what’s working, what isn’t, and bring opportunities for improvement into clearer focus 1](https://www.pipefy.com/blog/sales-analysis/).

[It provides a wide range of metrics and data that can help your sales team improve their sales pipeline model and make timely and informed decisions that drive revenue and reduce waste 1](https://www.pipefy.com/blog/sales-analysis/).

[By tracking all aspects of the buyer journey, from points of contact and messaging to engagement with content and more, you can give your sales team a deeper understanding of how to target potential customers and retain existing ones 1](https://www.pipefy.com/blog/sales-analysis/).

[An in-depth sales analysis can reveal patterns in behavior that can help your sales team create more effective buyer journeys and increase pipeline conversion rates 1](https://www.pipefy.com/blog/sales-analysis/).

DATA DICTIONARY

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

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

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

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

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

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

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

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

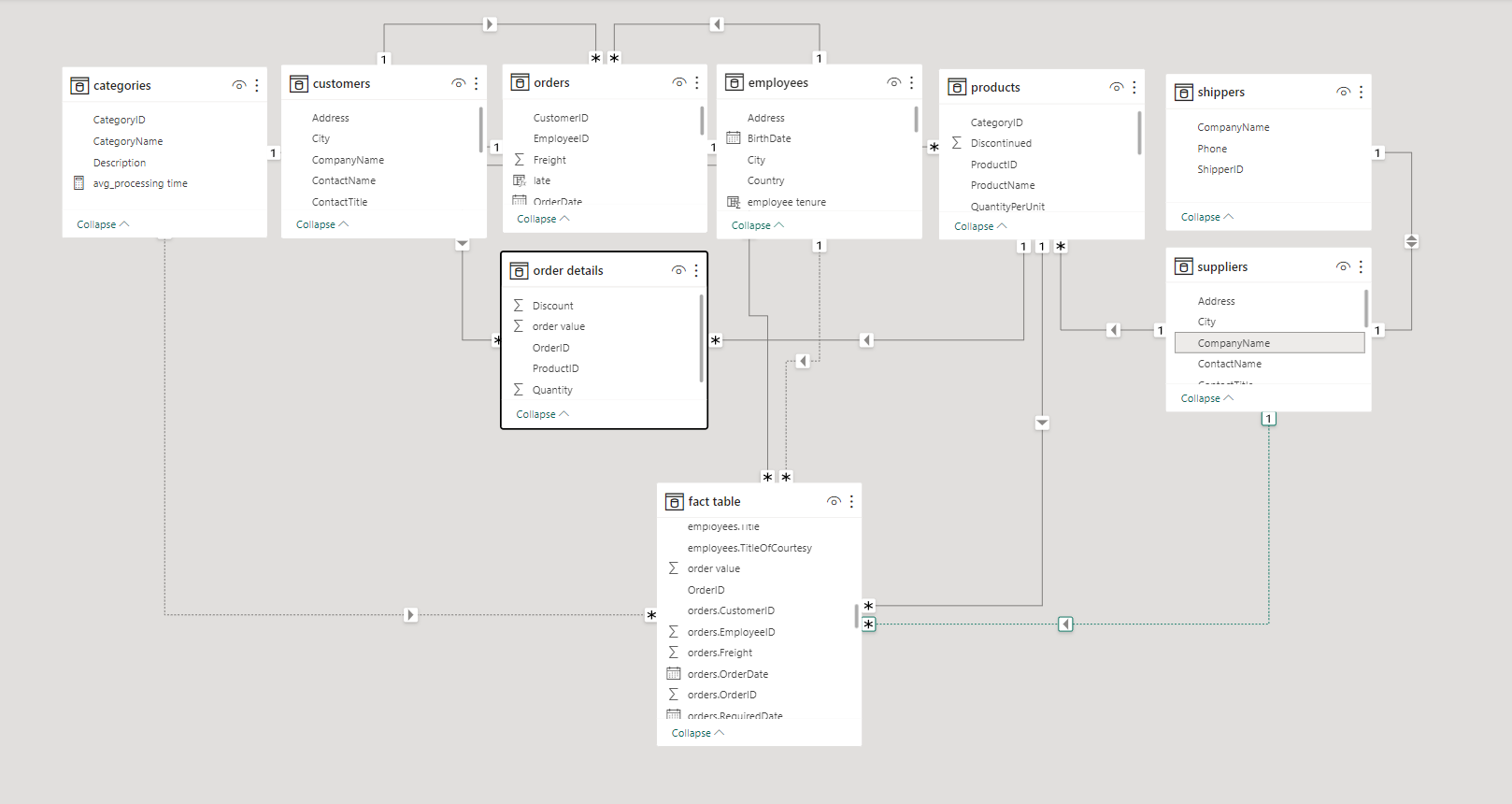
Phone: phone nuber of the supplier

PostalCode: postal codeoof the supplier

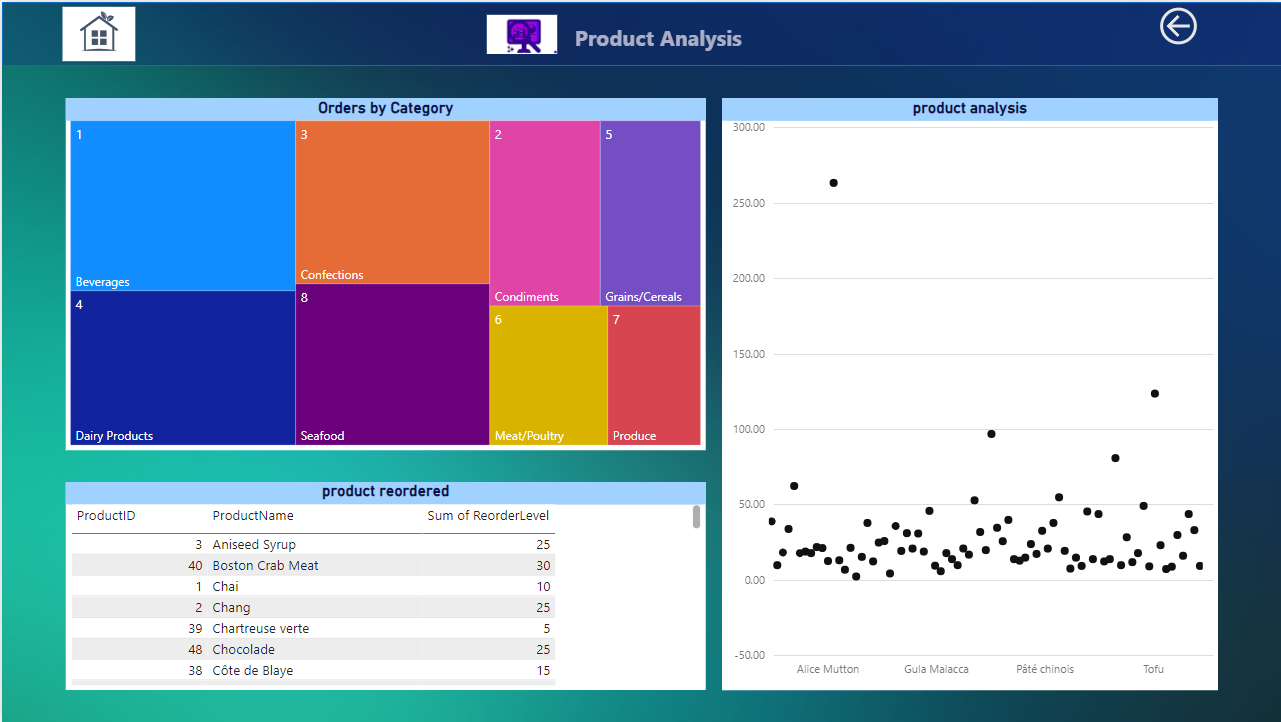
Region: location region of the supplier

SupplierID:unique supplier id of the supplier

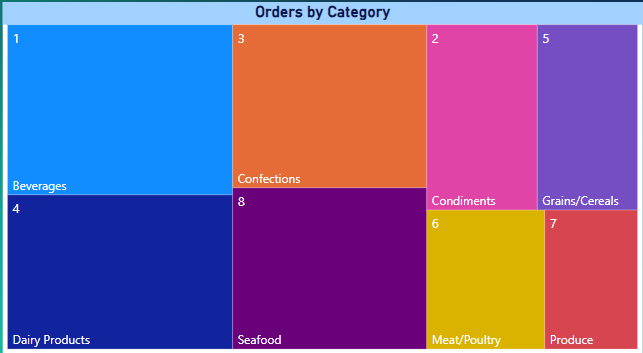
ER DAIGRAM:







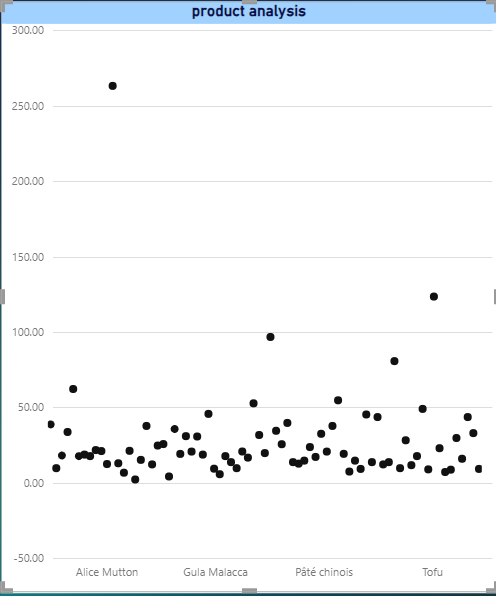
How does the sales volume vary across different product categories? Can we create a bar chart or tree map to display it?



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

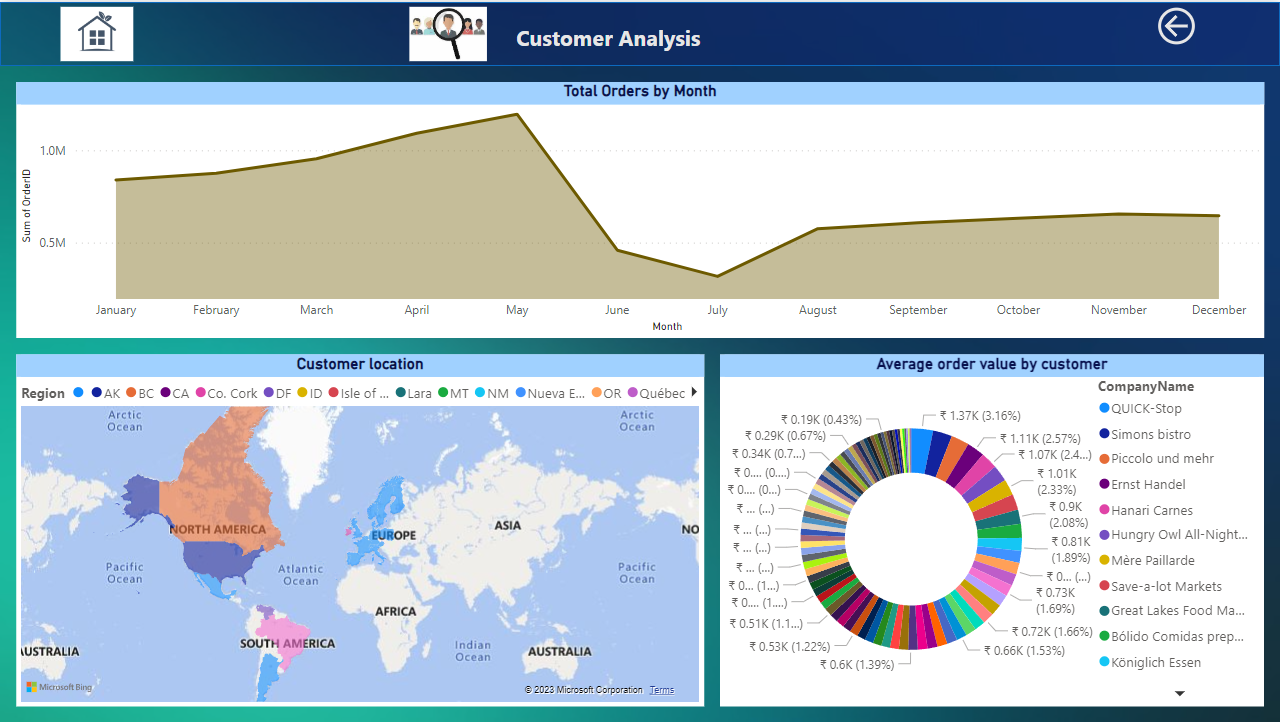
Yes, you can create a tree map to display the distribution of sales volume across different product categories. A tree map is an excellent choice for visualizing the distribution of sales volume or the count of orders by categories, especially when you want to show how each category contributes to the overall total. Each category will be represented as a rectangle within the tree map, and the size of the rectangle will correspond to the sales volume or order count for that category. This provides a visual representation of how different categories compare in terms of their contribution to sales or order counts.

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

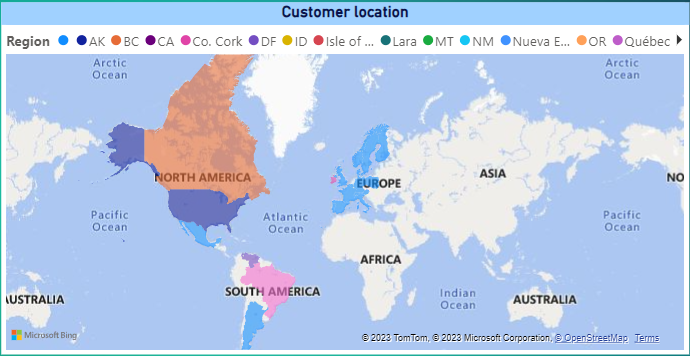


Here is the price distribution of products ,

and we can see that Alice Mutton is at the highest



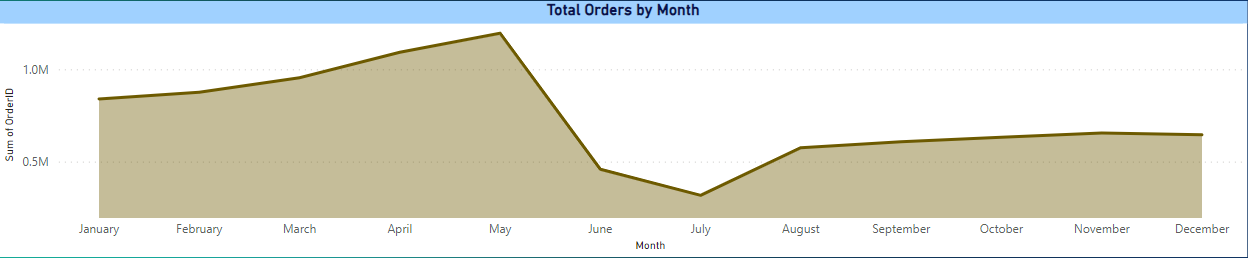
How does customer distribution vary across different regions or customer segments? Can we visualize it on a map or bar chart?



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

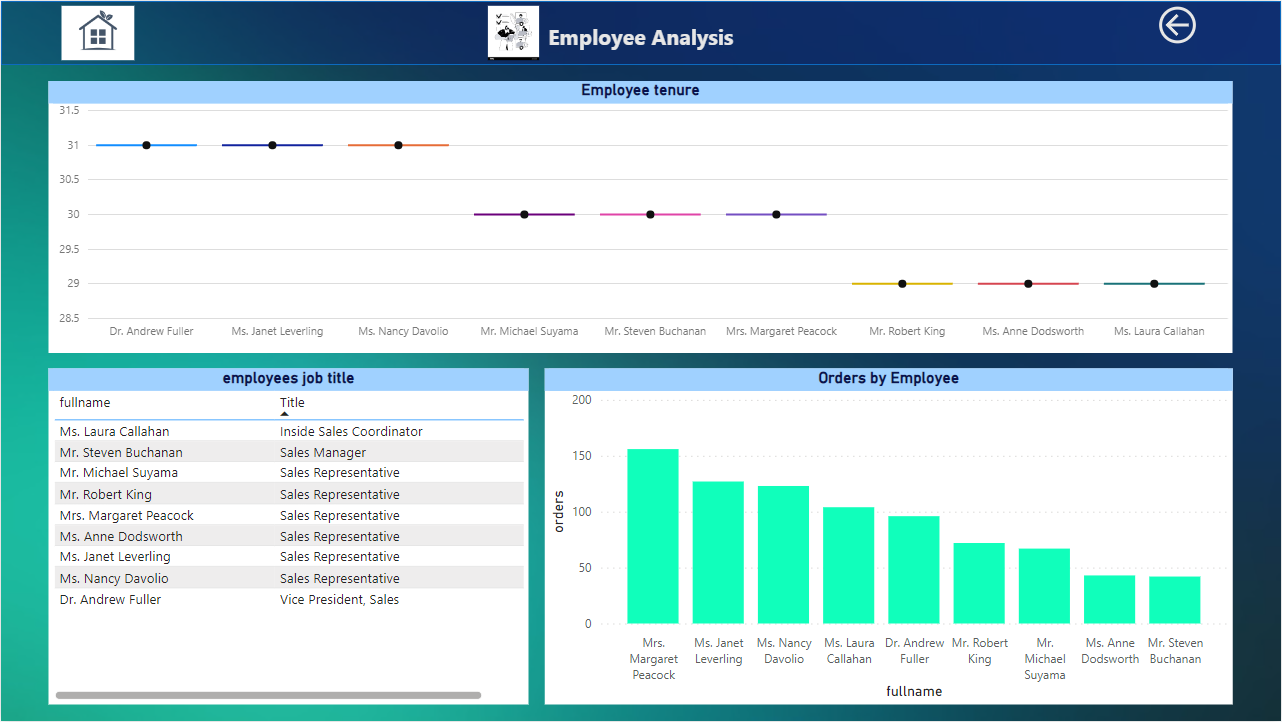
What is the trend in customer acquisition over time? Can we create a line chart or area chart to display it?



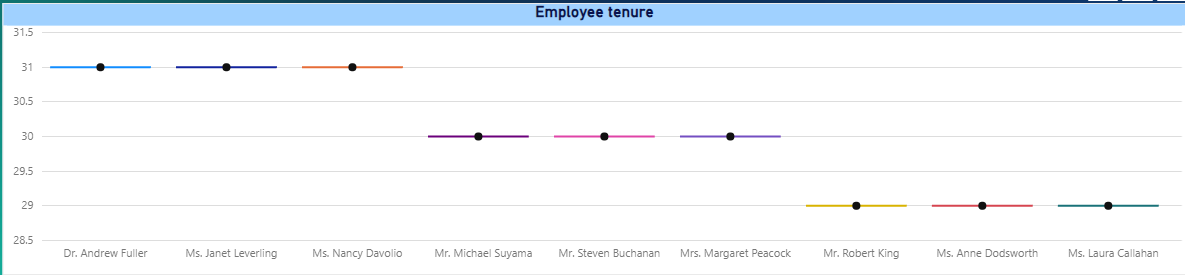
YES, We can distribute numbers of orders we get corresponding to

month

Here May month has maximum sales followed by April and July month has lowest sales



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



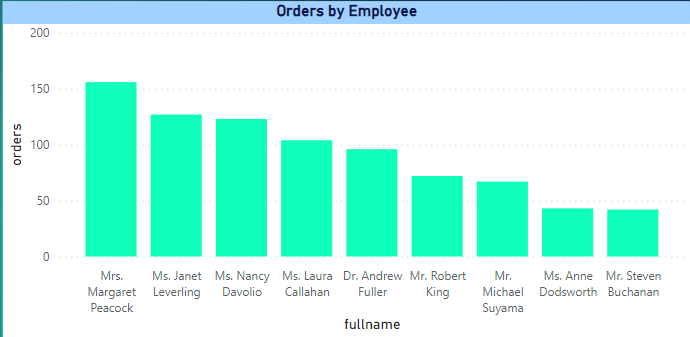
The distribution of employee tenure can be displayed using a histogram or a box plot, which provides insights into the spread and central tendency of employee tenure lengths.

Here the average tenure can be taken as 30 years

And maximum tenure is 31 years

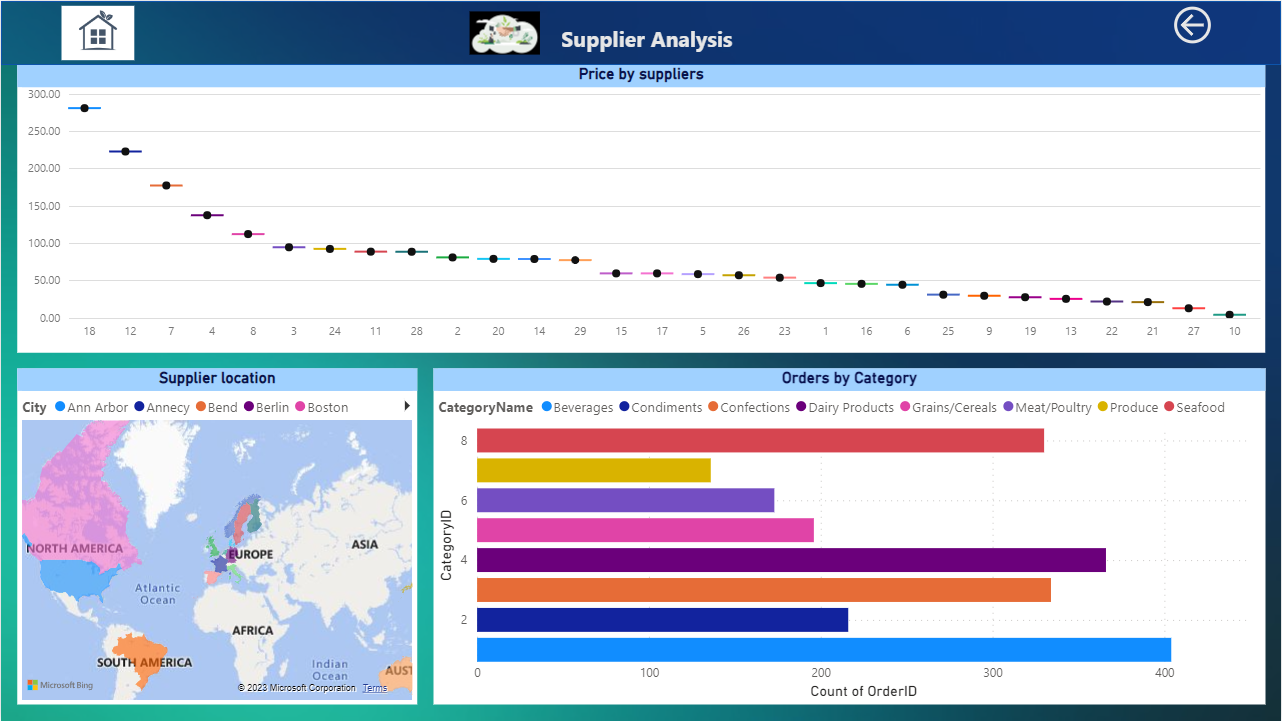
And minimum tenure is 29 years

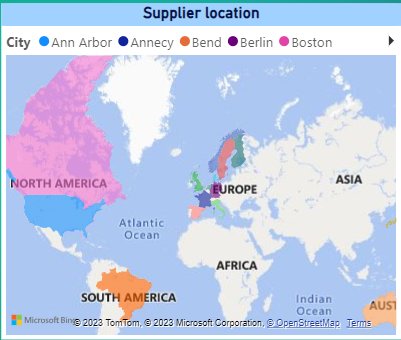
How does employee productivity vary across different departments or job roles? Can we create a stacked bar chart or grouped column chart to visualize it?



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

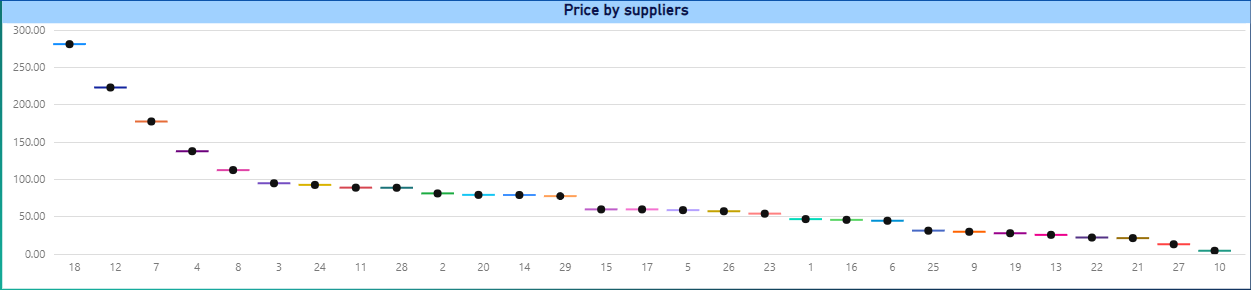




Yes, you can visualize the geographical distribution of suppliers using a map or a bubble chart, both of which are effective for displaying supplier locations.

Here USA country region has the maximum suppliers

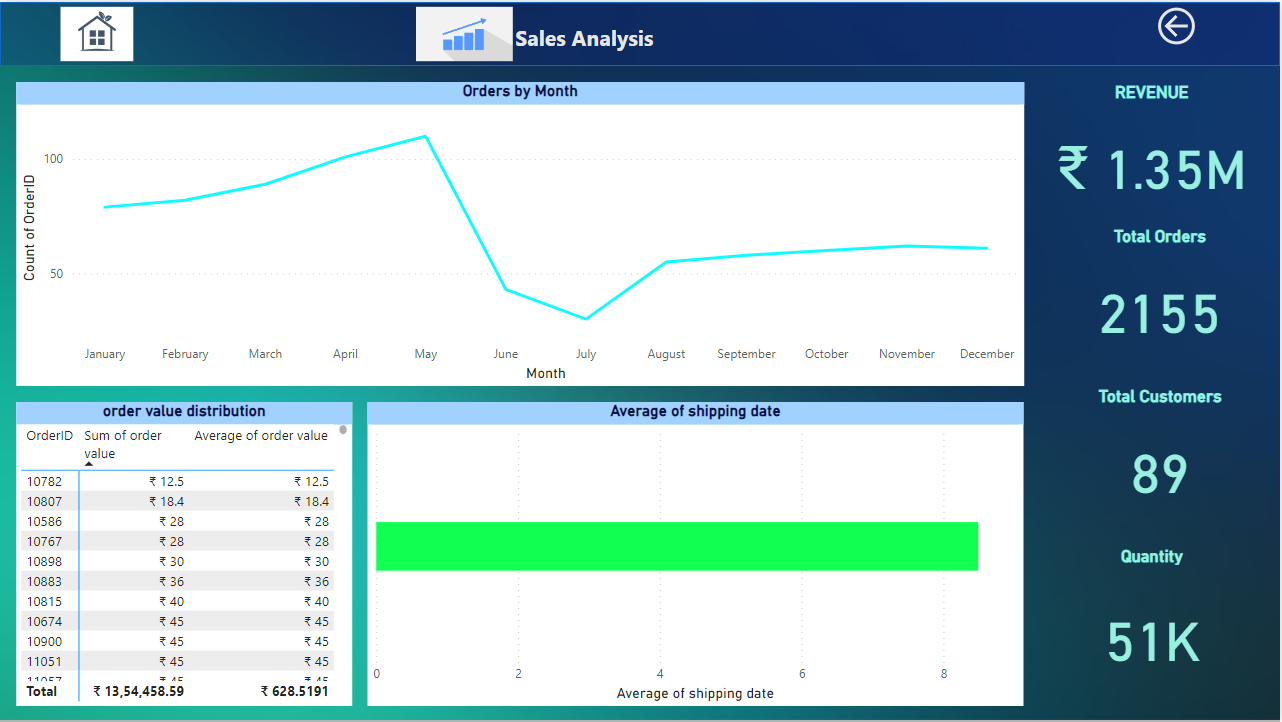
Can we visualize the geographical distribution of suppliers using a map or bubble chart?



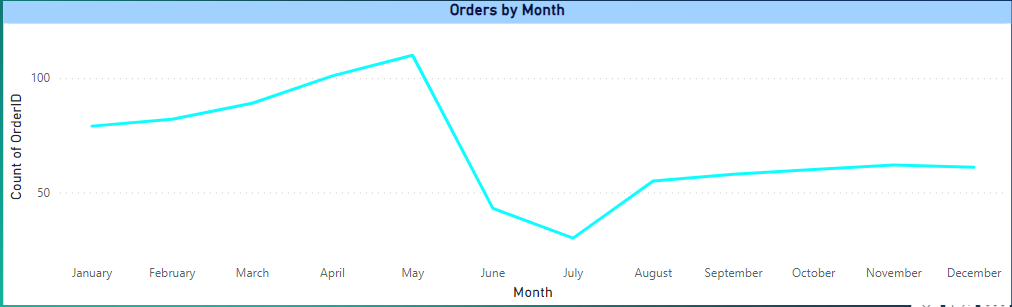
To visualize the variation in cost or pricing structures across different suppliers, you can create a box plot or a stacked bar chart, allowing for comparisons of pricing information among suppliers.

Here Aux Joyeux ecclesiastiques (company name) 18. supplier has the maximum price

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

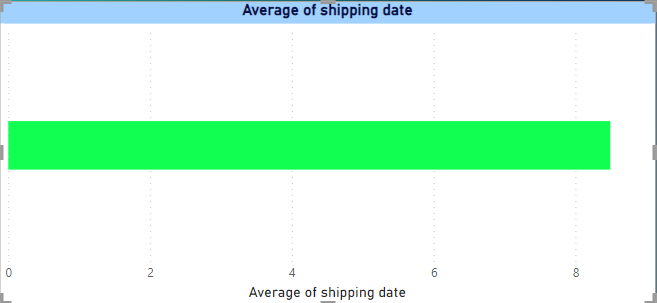


How does order volume change over time? Can we create a time series chart or stacked bar chart to visualize it?



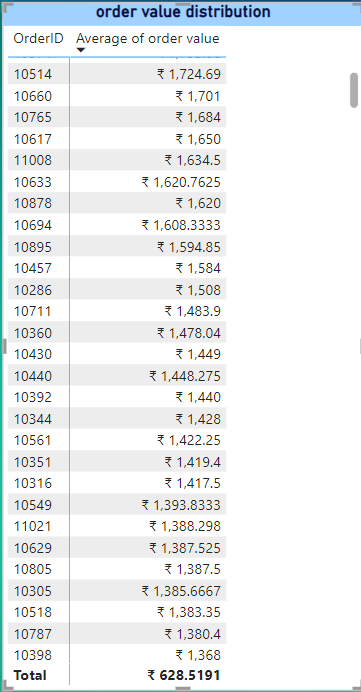
To visualize changes in order volume over time, you can create a time series chart, which is ideal for showing trends in order volume. Stacked bar charts are typically used for different purposes, like showing composition within a single time period.

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



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

Average shipping date can be between 9 to 10



Yes, we can create a box plot chart but it will be very difficult to understand.

I think table or matrix will be more suitable.

orderID 10865 10981has the maximum ordervalue which is 15,810 and the lowest orderID 10462 has the lowest sales order value 4.8

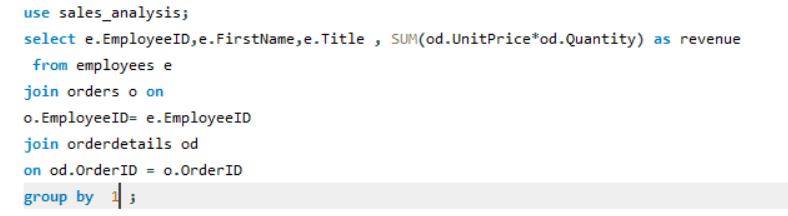
Here, we have taken the average order value.

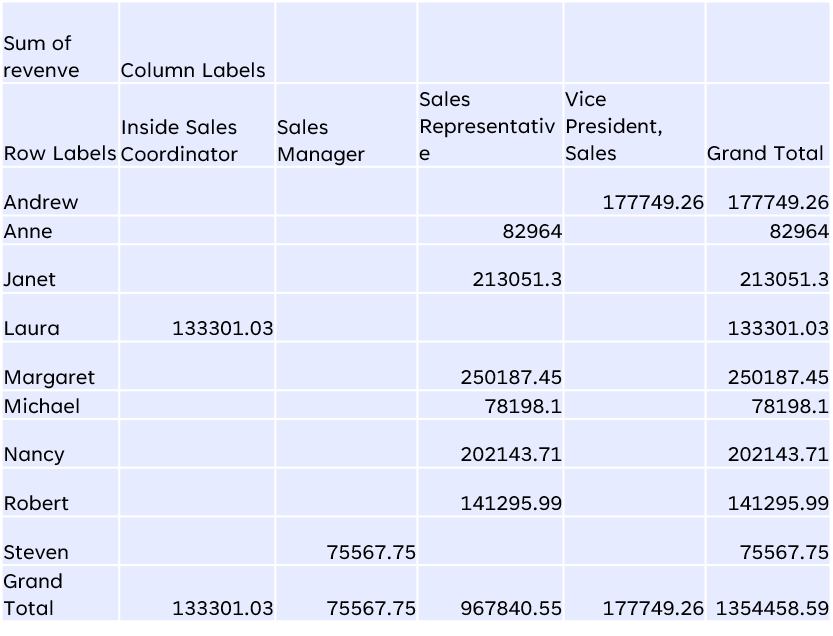
What is the distribution of order values? Can we create a histogram or box plot to display it?

EDA PROBLEM STATEMENTS

EDA:

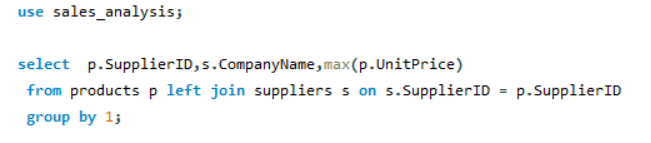
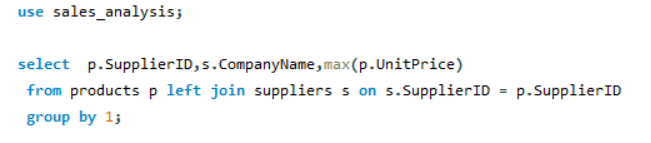
8. How does employee turnover vary across different departments or job roles? Can we visualize this using bar charts or heatmaps?





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?

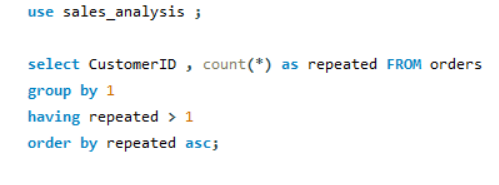
 

Here, Aux Joyeux ecclesiastiques (company name) supplier has the maximum price and Refrescos americans LTDA has lowest price

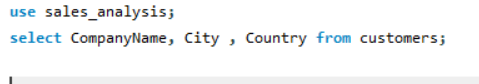
Here, Aux Joyeux ecclesiastiques (company name) supplier has the maximum price and Refrescos americans LTDA has lowest price

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 | | | | | |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

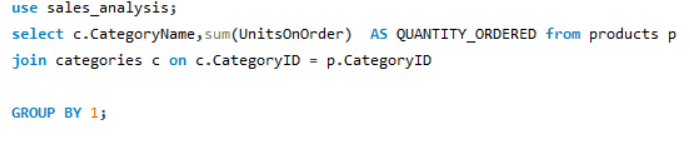
2. How do customer preferences vary based on their location or demographics? Can we explore this through interactive visualizations?

|  |  |
| --- | --- |
| 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 |

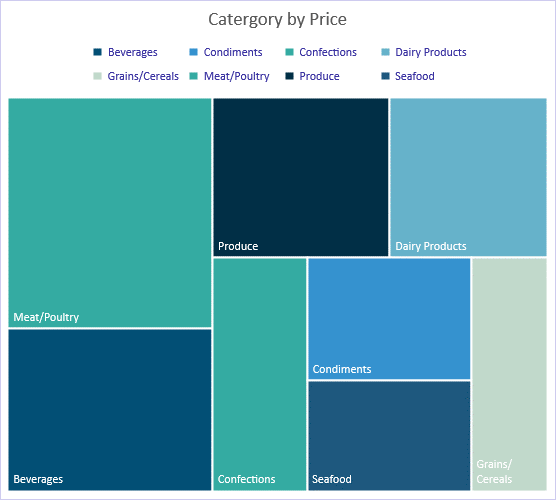
3. Are there any interesting patterns or clusters in customer behavior that can be visualized to identify potential market segments?

|  |  |
| --- | --- |
| 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 |

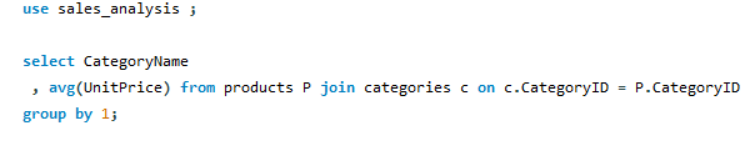
|  |  |
| --- | --- |
| CATEGORYID | QUANTITY ORDERED |
| Beverages | 60 |
| Condiments | 170 |
| Confections | 180 |
| Dairy Products | 140 |
| Grains/Cereals | 90 |
| Meat/Poultry | 0 |
| Produce | 20 |
| Seafood | 120 |

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

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

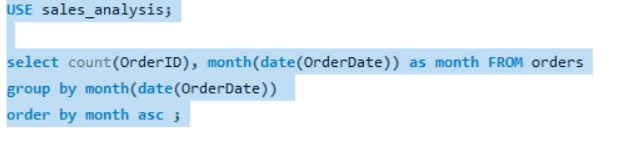


|  |  |
| --- | --- |
| categoryid | average unit price |
| Beverages | 37.97916667 |
| Condiments | 23.0625 |
| Confections | 25.16 |
| Dairy Products | 28.73 |
| Grains/Cereals | 20.25 |
| Meat/Poultry | 54.00666667 |
| Produce | 32.37 |
| Seafood | 20.6825 |
|  |  |

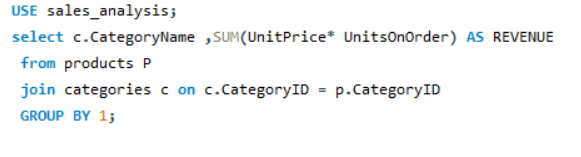
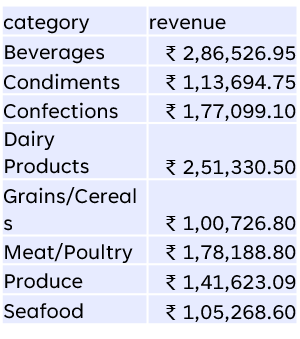


11. How does product demand fluctuate over different seasons or months? Can we visualize this through line charts or area charts?

|  |  |  |
| --- | --- | --- |
| 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 |



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





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

CONCLUSION

* The company total revenue is 1.35 m
* Total placed orders are 2155
* There are 89 customer total
* Maximum sales month or season is may followed by April
* Minimum sales month or season is July.
* Beverage category has the maximum sales units
* Average tenure of the employee is 30 years and there are total 9 employees out of which 6 are from sales representative

Total revenue of products sold by these 6 sales representative is 967840.55 which more than half of the total revenue combined.

* Mrs. Margert peacock has highest sales orders from these 9 employee.
* Average shipping date takes around 9 -10 days to ship the order
* Average order value can be taken around 1000.
* USA and Germany has the maximum customer and orders 172 and 156 respectively.
* Brazil , Spain ,UK ,France are the growing customer base countries
* , Aux Joyeux ecclesiastiques (company name) supplier has the maximum Product price and Refrescos americans LTDA has lowest product price