

CAPSTONE PROJECT

SALES ANALYSIS

DATA ANALYTICS

MUSHAB ALI

TABLE OF CONTENTS

- INTRODUCTION**
- OBJECTIVE**
- SIGNIFICANCE**
- METHODOLOGY**
- ER DIAGRAM**
- EDA PROBLEM STATEMENTS**
- POWER BI PROBLEM STATEMENTS**
- POWER BI DASHBORD**
- CONCLUSION**

INTRODUCTION

Welcome to the exciting world of data exploration and visualization with Northwind Traders! In this journey, we delve into the realm of Power BI and Excel exploratory data analysis (EDA) to unlock the hidden treasures within the Northwind database. Imagine yourself as an intrepid explorer, navigating through the vast landscape of data to uncover valuable insights that will shape the future of Northwind Traders.

Picture this: you are standing at the threshold of a labyrinth of information, armed with nothing but your curiosity and the powerful tools of Power BI and Excel. Behind each dataset lies a story waiting to be told, a narrative of customer behavior, sales dynamics, and employee performance, all waiting to be revealed through the lens of data visualization.

As we embark on this adventure, we are not merely passive observers but active participants in the transformation of raw data into actionable intelligence. Our mission is clear: to create a visually stunning and user-friendly dashboard that not only communicates key performance metrics but also empowers stakeholders to make data-driven decisions with confidence.

Join us as we journey through the intricacies of customer segmentation, sales analysis, inventory trends, and employee performance, piecing together the puzzle of Northwind Traders' operations one visualization at a time. Together, we will revolutionize how Northwind Traders interacts with its data, paving the way for innovation and growth in the competitive landscape of wholesale markets.

So, grab your compass and let's embark on this exhilarating expedition into the heart of data-driven decision-making with Power BI and Excel EDA. The adventure awaits!

OBJECTIVE

The aim of the North Wind Traders Capstone Project is to leverage Power BI and Excel EDA to perform a comprehensive analysis of the company's sales data, providing valuable insights into customer behavior, sales patterns, and employee performance. Through this analysis, the project seeks to aid decision-making processes within North Wind Traders and empower stakeholders to make data-driven decisions effectively.

Objective 1: Data Consolidation

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

Ensure data accuracy, completeness, and consistency through thorough data cleaning and MECE (Mutually Exclusive, Collectively Exhaustive) breakdown.

Objective 2: Sales Analysis

Analyze sales data to identify trends, patterns, and fluctuations over time, by product, region, and customer segment.

Calculate key sales metrics such as total revenue, average order value, and sales by product category or region to gain insights into sales performance.

Objective 3: Customer Behavior Understanding

Analyze customer data to understand buying behavior, preferences, and demographics.

Segment customers based on various criteria to enable targeted marketing strategies and improve customer engagement and retention.

Objective 4: Inventory Management Optimization

Monitor inventory levels, analyze stock movement, and identify inventory trends to optimize inventory management processes.

Calculate inventory metrics such as turnover rate, stock levels, and reorder frequency to ensure efficient inventory control and minimize carrying costs.

Objective 5: Employee Performance Evaluation

Assess employee sales performance, productivity, and effectiveness.

Rank employees by performance and identify top performers for recognition and reward, while also highlighting opportunities for improvement and training.

Objective 6: Actionable Insights Delivery

Present insights derived from data analysis in a visually appealing and intuitive dashboard format.

Enable stakeholders to make informed decisions and take proactive actions by providing access to timely and relevant data.

Objective 7: Data-Driven Culture Promotion

Promote a culture of data-driven decision-making within the organization.

Provide access to data and empower stakeholders at all levels to leverage data for strategic planning and operational improvements.

Objective 8: Business Competitiveness Enhancement

Use data insights to identify competitive advantages, market opportunities, and areas for innovation.

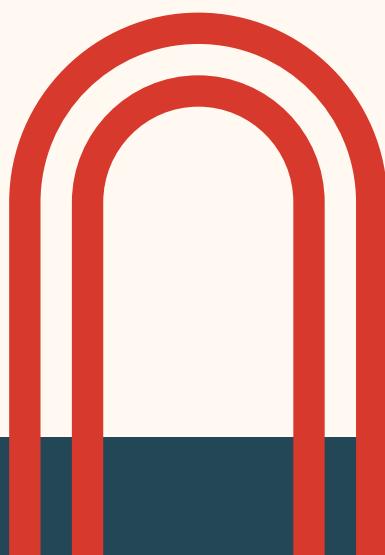
Enable Northwind Traders to stay ahead in the wholesale market landscape by leveraging data-driven strategies.

Objective 9: Continuous Improvement Facilitation

Continuously monitor and analyze key performance metrics.

Gather feedback from stakeholders and iterate on the Power BI dashboard to ensure it remains relevant, insightful, and aligned with business objectives.

By achieving these objectives, the North Wind Traders Capstone Project aims to revolutionize how the company interacts with its data, enabling informed decision-making, driving business growth, and maintaining a competitive edge in the wholesale market landscape.



SIGNIFICANCE

The North Wind Traders Capstone Project holds significant importance for the company and its stakeholders due to several key factors:

Data-Driven Decision Making: By leveraging Power BI and Excel EDA techniques, the project enables North Wind Traders to transition towards a data-driven decision-making approach. This shift allows the company to base its strategic and operational decisions on empirical evidence derived from comprehensive data analysis, leading to more informed and effective outcomes.

Enhanced Operational Efficiency: Through the analysis of sales patterns, customer behavior, inventory trends, and employee performance, the project identifies opportunities to streamline processes and optimize resource allocation within the organization. This optimization leads to increased operational efficiency, reduced costs, and improved productivity across various departments.

Improved Customer Satisfaction: By gaining insights into customer preferences, buying behavior, and segmentation, North Wind Traders can tailor its products and services to better meet customer needs. This personalized approach enhances customer satisfaction, fosters loyalty, and ultimately drives revenue growth through repeat business and positive word-of-mouth referrals.

Competitive Advantage: The project enables North Wind Traders to stay ahead of competitors by leveraging data insights to identify market trends, capitalize on opportunities, and address potential threats. By continuously monitoring key performance metrics and adapting strategies accordingly, the company maintains a competitive edge in the dynamic wholesale market landscape.

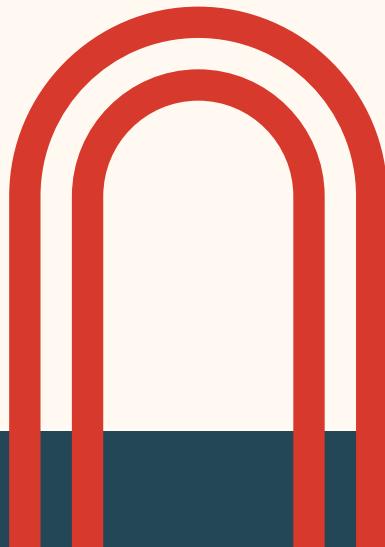
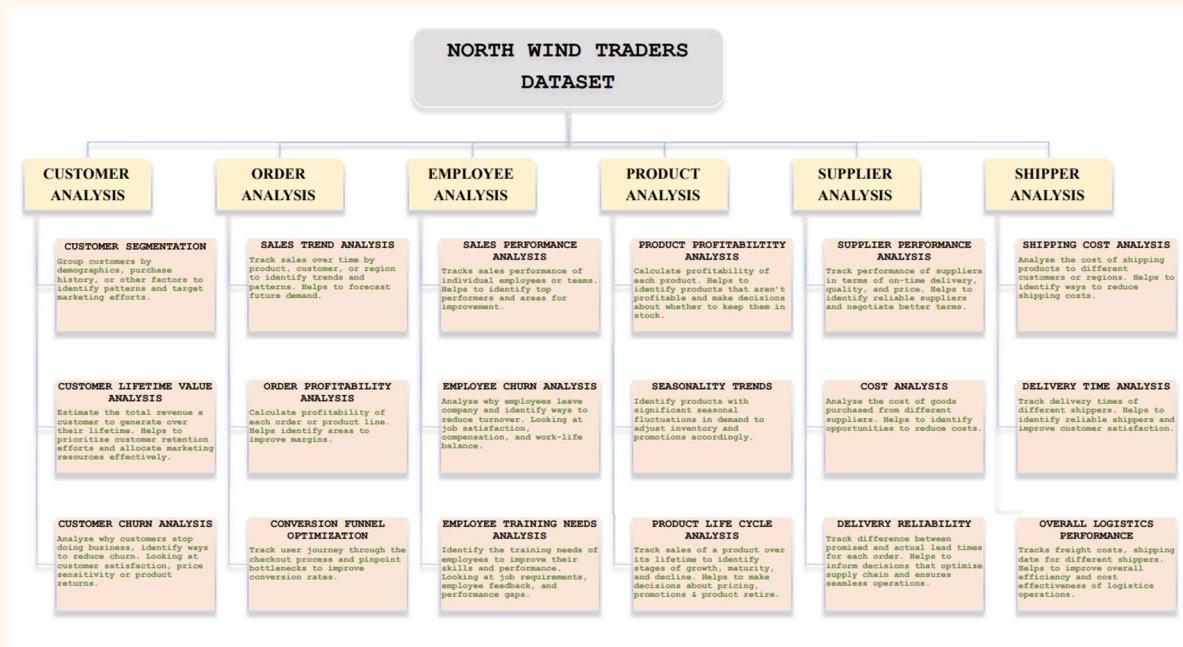
Strategic Planning and Innovation: Through the delivery of actionable insights and the promotion of a data-driven culture, the project empowers North Wind Traders to make strategic decisions that fuel innovation and drive business growth. By identifying areas for improvement, innovation, and investment, the company can proactively address emerging challenges and capitalize on new opportunities.

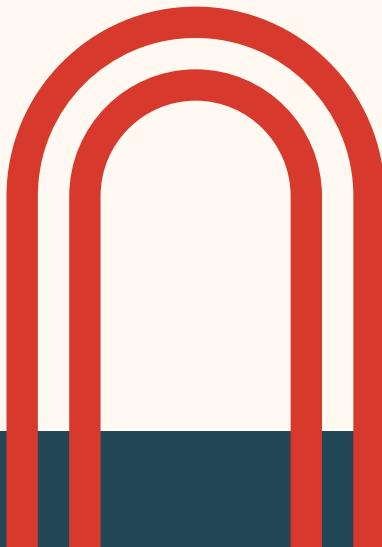
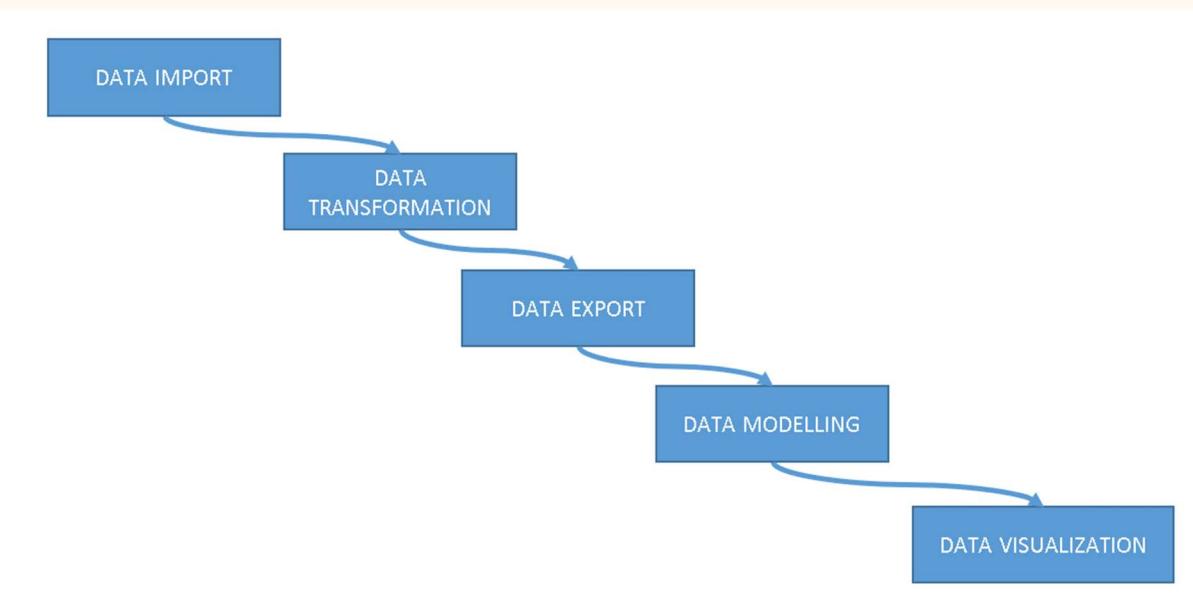
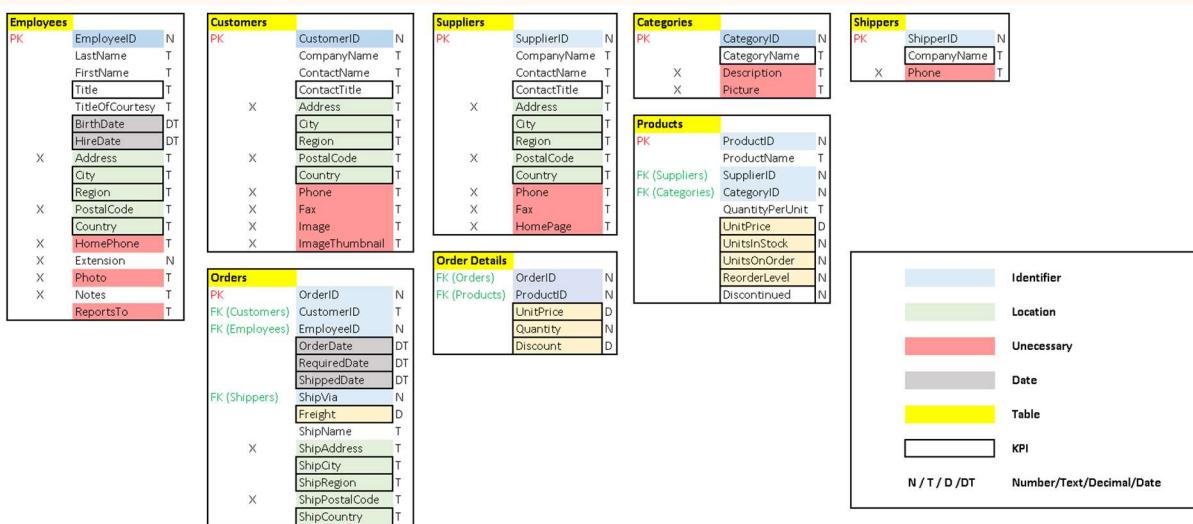
Stakeholder Empowerment: The project not only benefits the company's leadership team but also empowers employees at all levels to contribute meaningfully to organizational success. By providing access to timely and relevant data, interactive visualizations, and dynamic filters, stakeholders can actively engage in data exploration, collaborate on decision-making, and drive continuous improvement initiatives.

Long-Term Sustainability: By facilitating continuous monitoring, analysis, and iteration of key performance metrics, the project promotes a culture of continuous improvement within North Wind Traders. This ongoing commitment to data-driven decision-making ensures the company remains adaptable, resilient, and sustainable in the face of evolving market dynamics and industry disruptions.

In summary, the North Wind Traders Capstone Project is not only a means to analyze sales data but also a catalyst for organizational transformation, strategic innovation, and sustainable growth. By harnessing the power of data and analytics, the project empowers North Wind Traders to navigate challenges, capitalize on opportunities, and achieve long-term success in the wholesale market landscape.

METHODOLOGY





Data Acquisition and Preparation

Dataset Import in MySQL:

Utilize the provided SQL script within a folder structure compatible with your chosen MySQL client for importing the Northwind Traders dataset.

Ensure proper database creation and configuration before proceeding.

Data Transformation in MySQL:

Employ SQL queries to clean and transform the data:

Address missing values, inconsistencies, and formatting issues.

Create calculated fields or derived tables as needed.

Consider data normalization techniques if necessary.

Maintain a record of your transformations for reproducibility and reference.

Data Export to Dump Folder:

Export the prepared data from MySQL to a designated "SQL" folder using appropriate methods:

Choose a suitable export format (e.g., CSV, TSV) based on your analysis tools.

Organize the exported data logically within the folder structure.

Data Modeling and Visualization

Data Modeling in Power BI and Excel:

Power BI:

Import the exported data into Power BI Desktop.

Establish relationships between tables based on their primary and foreign keys.

Consider implementing a star schema or snowflake schema for optimal performance.

Create calculated columns or measures to enhance your analysis.

Excel:

Import the exported data into Excel spreadsheets.

Establish data connections between related tables using Power Query or similar tools.

Clean and transform the data within Excel if necessary.

Consider data modeling techniques like pivot tables and data models for efficient analysis.

Data Visualization in Power BI:

Create visually appealing and informative dashboards and reports using Power BI's rich visualization library.

Employ a variety of chart types (bar charts, line charts, pie charts, maps, etc.) to represent different aspects of the data.

Leverage slicers, filters, and drill-down capabilities to enable interactive exploration.

Ensure effective use of labels, legends, and color palettes for clarity and accessibility.

Data Visualization in Excel:

Create charts and graphs within Excel sheets or separate dashboards.

Utilize Excel's charting options and consider add-ins for advanced visualization capabilities.

Ensure chart titles, labels, and legends are clear and concise.

Prioritize user-friendliness and interactivity through filtering, conditional formatting, and dynamic features.

Data Interpretation and Insights

Data Interpretation:

Analyze the visualizations and identify key trends, patterns, and relationships within the data.

Consider various perspectives and potential explanations for your findings.

Consult domain knowledge and business context to interpret the data accurately.

Data Insights:

Translate your interpretations into actionable insights for different stakeholders.

Focus on providing clear, concise, and relevant information that drives decision-making.

Tailor your insights to the specific needs and interests of each target audience.

Additional Considerations:

Throughout the process, document your methodology, assumptions, and decisions for clarity and reproducibility.

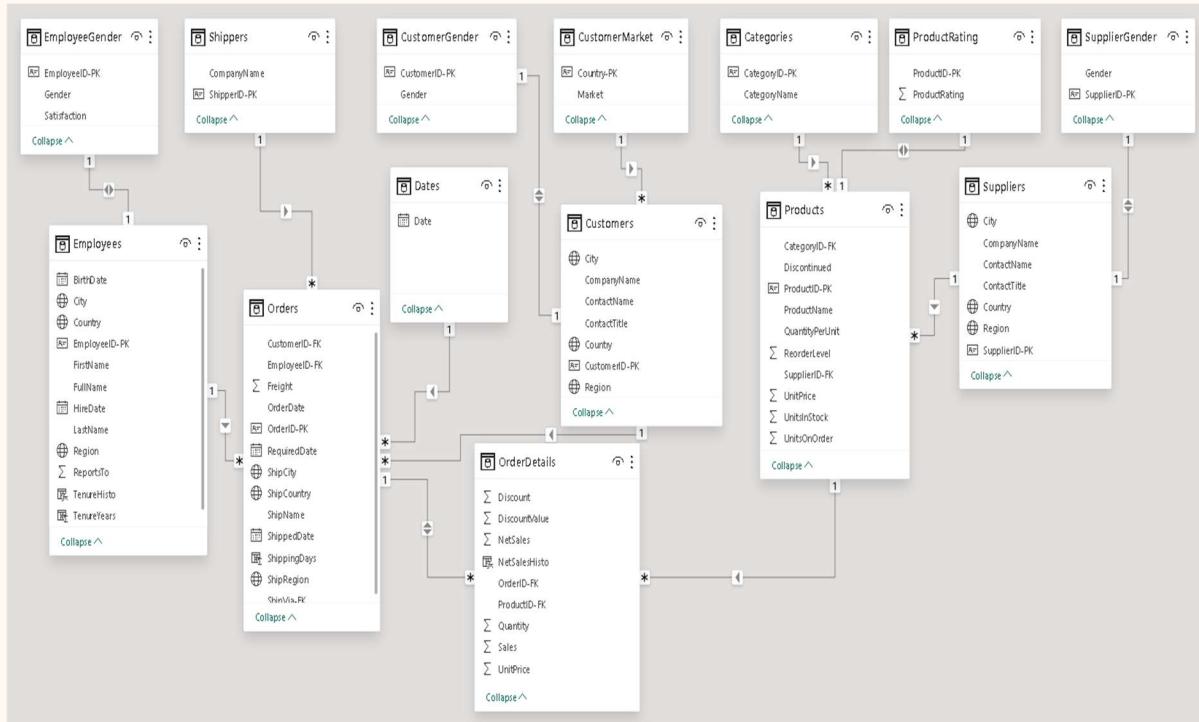
Validate your findings and ensure data quality to maintain trust in your analysis.

Present your results effectively, considering visual storytelling techniques and clear communication strategies.

Continuously refine your methodology and visualizations based on feedback and evolving requirements.

By following these steps and incorporating the valuable feedback you've received, you can conduct a comprehensive and insightful data analysis project using both Power BI and Excel. Remember that the specific tools and techniques may vary depending on your project's scope and objectives.

ER-DIAGRAM



The ER diagram shows the relationships between the following entities:

Customers

Employees

Suppliers

Products

Orders

Categories

The relationships between these entities are as follows:

Customers can place orders.

Orders can contain one or more products.

Products can belong to one category.

Categories can have many products.

Suppliers can supply many products.

Products can be supplied by many suppliers.

Employees can work in one department.

Departments can have many employees.

The ER diagram also includes the following attributes:

Customers: customer ID, first name, last name, address, city, state, zip code, country, phone number, email address

Employees: employee ID, first name, last name, address, city, state, zip code, country, phone number, email address, hire date, job title, department

Suppliers: supplier ID, company name, contact name, address, city, state, zip code, country, phone number, email address

Products: product ID, product name, description, price, quantity in stock, category ID

Orders: order ID, customer ID, order date, shipping address, shipping city, shipping state, shipping zip code, shipping country, shipping cost, total amount

Categories: category ID, category name

The ER diagram is a useful tool for understanding the relationships between the different entities in a database. It can be used to design a database, as well as to query and manipulate data in a database.

EDA PROBLEM STATEMENT

1. WHAT ARE THE KEY FACTORS INFLUENCING CUSTOMER RETENTION OR LOYALTY BASED ON THE DATASET ?

ORDERS BY CUSTOMERS

This bar graph provides insight into the distribution of orders among different customers, with some having only a few orders while others have significantly more.



The **Customer order data** shows a wide range in number of orders, from a minimum of 1 order to a maximum of 31 orders. The average number of orders per customer is about 10.

The most common number of orders placed were 3 orders (for 8 customers), 5 orders (for 10 customers), and 4 orders (7 customers). This indicates many customers place multiple orders on a regular basis.

Looking at the extreme ends, the customers placing the greatest number of orders are : Horst Kloss with 28 total orders, Roland Mendel with 30 orders, and Jose Pavarotti with 31 orders - the maximum of any customer. On the other end of the spectrum, Francisco Chang stands out with only 1 order placed, the sole customer placing just a single order.

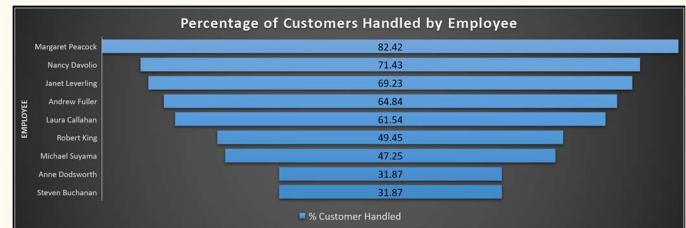
Overall, the customer order data shows regular patrons who continuously place several orders over time, with most ordering 3-5 times on average. A small number of high-volume customers like Jose Pavarotti account for a greater share of total orders. Attracting more customers that have order levels closer to the high-volume regulars could be a potential growth opportunity.

SQL QUERY USED FOR ABOVE CHART

```
create view Q1_CustomerOrderCounts_view as
select c.ContactName 'Customer Name',
count(distinct o.OrderID) 'Order Counts'
from orders o
join customers c on o.CustomerID = c.CustomerID
group by 1
order by 2;
```

PERCENTAGE OF CUSTOMERS HANDLED BY EMPLOYEE

This summary provides the percentage of customers handled by each employee, listed from highest to lowest:



Margaret Peacock: 82.42%

Nancy Davolio: 71.43%

Janet Leverling: 69.23%

Andrew Fuller: 64.84%

Laura Callahan: 61.54%

Robert King: 49.45%

Michael Suyama: 47.25%

Anne Dodsworth: 31.87%

Steven Buchanan: 31.87%

These percentages indicate the proportion of the total number of customers each employee has handled. Margaret Peacock has the highest percentage, indicating she has handled the largest share of customers, while Steven Buchanan and Anne Dodsworth have handled the lowest percentage, indicating they have handled fewer customers compared to others on the list.

SQL QUERY USED FOR ABOVE CHART

```
create view Q1_EmployeeCustomerCounts_view as
select concat(e.FirstName," ",e.LastName) 'Employee Name',
round(count(distinct o.CustomerID)/(select count(*) from
customers)*100,2) '% Customer Handled'
from orders o
join employees e on o.EmployeeID = e.EmployeeID
group by 1
order by 2 desc;
```

EMPLOYEE PER CUSTOMER & SALES

This data provides information about various customers, including the number of employees assigned to them and their respective sales amounts. Here's a summary:



Jose Pavarotti: Handled by 9 employees, with sales totaling \$115,673.39.

Horst Kloss: Handled by 8 employees, with sales totaling \$117,483.39.

Roland Mendel: Handled by 8 employees, with sales totaling \$113,236.68.

Patricia McKenna: Handled by 8 employees, with sales totaling \$57,317.39.

Paula Wilson: Handled by 8 employees, with sales totaling \$52,245.90.

... (and so on for other customers)

The data indicates the distribution of customers among employees and the corresponding sales amounts for each customer. Some customers have a larger sales volume and are handled by more employees, while others have smaller sales volumes and are handled by fewer employees.

AVERAGE DELIVERY DAYS & DISCOUNT PRICE FOR CUSTOMERS

The average delivery days across all customers is 28.2 days, with a range from 23.9 days (Palle Ibsen) to 34 days (Elizabeth Brown).



The customers receiving the highest discounts are:

Jose Pavarotti (\$11,311.45)

Roland Mendel (\$8,361.71)

Patricia McKenna (\$7,337.48)

The customers with the longest average delivery days are:

Elizabeth Brown (34 days)

Helvetius Nagy (32.7 days)

Liz Nixon (32 days)

A large number of customers (35 out of 63) did not receive any discount or have zero price, indicating room for improvement in providing discounts and securing more sales from existing clients.

Overall, there is an opportunity to reduce delivery times and provide higher discounts to customers, as the maximum days are often 30+ and many customers do not receive much if any discount or make a purchase. Targeting improvements in these areas could increase customer satisfaction and drive more repeat purchases.

SQL QUERY USED FOR ABOVE CHART

```
create view Q1_CustomerDeliveryDaysDiscount_view as
select c.ContactName 'Customer Name',
avg(o.DeliveryDays) 'Delivery Days',
sum(odv.DiscountValue) 'Discount Price'
from orders o
join customers c on o.CustomerID = c.CustomerID
join order_details_view odv on o.OrderID = odv.OrderID
group by 1
order by 2 desc, 3 desc;
```

SQL QUERY USED FOR ABOVE CHART

```
create view Q1_CustomerEmployeeSales_view as
select c.ContactName 'Customer Name',
count(distinct o.EmployeeID) 'Employee Counts',
sum(odv.Sales) Sales
from orders o
join customers c on o.CustomerID = c.CustomerID
join order_details_view odv on o.OrderID = odv.OrderID
group by 1
order by 2 desc, 3 desc;
```

FREIGHT TREND

The data covers average monthly freight from August 1994 to June 1996. Over this period, the lowest monthly average freight was \$48.85 in October 1994 and the highest was \$108.17 in June 1995.



Looking at yearly trends:

In 1994, average freight ranged from \$48.85 to \$87.01, with an average of \$63.02 across the 5 months.

In 1995, average freight ranged from \$57.76 to \$108.17, with an average of \$80.47 over the 12 months. This is a 27% increase over 1994.

In the first 6 months of 1996, the range was \$54.31 to \$101.80 with an average of \$79.11. This is a slight decrease from 1995's level.

The data shows rising freight costs over 1994-1995 before potentially leveling off in 1996. The peak averages occur during summer months in 1995 and 1996, reaching over \$100 in both June 1995 and February 1996. There also appears to be significant monthly fluctuation in averages.

This information could help plan shipping/logistics costs in budgets and identify months with historically higher freight that may require advanced preparation. Continuing to track monthly and yearly averages can inform longer term trends.

SQL QUERY USED FOR ABOVE CHART

```
create view Q1_TrendAverageFreight_view as
select concat(year(OrderDate), " ", date_format(OrderDate, '%b'))
'Year Month',
avg(Freight) 'Average Freight'
from orders
group by 1;
```

EDA PROBLEM STATEMENT

2. HOW DO CUSTOMER PREFERENCES VARY BASED ON THEIR LOCATION OR DEMOGRAPHICS? CAN WE EXPLORE THIS THROUGH INTERACTIVE VISUALIZATIONS?

PRODUCT'S PREFERENCE BY CUSTOMER'S COUNTRY

The data covers product sales across many countries. There are a wide variety of food and beverage products.



Some top insights:

Germany, USA, and Brazil are the top countries by number of different products sold

The products with the highest total global quantities sold are Chai (1207 units), Chang (1052 units), and Aniseed Syrup (298 units)

Chef Anton's Cajun Seasoning, Grandma's Boysenberry Spread, and Uncle Bob's Organic Dried Pears are selling well across many European countries

Alcohol products like beers, wines, and liqueurs have strong sales in European nations

Canada, Austria and France purchase a lot of maple syrup

Germany, Austria, Switzerland lead sales of multiple chocolate and sweet items

Seafood products do well in coastal countries like UK, Germany, France and Scandinavia

The US and Canada account for high sales of multiple Italian pasta specialities

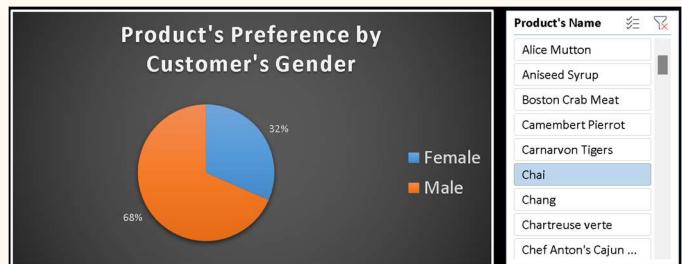
In summary - Popular items in core categories like hot drinks, condiments, seafood, chocolate and pasta drives most of the cross-country product demand. Regional specialities also perform well in their native markets.

SQL QUERY USED FOR ABOVE CHART

```
create view Q2_CustomerCountryProductQuantity_view as
select c.Country 'Customer's Country',
p.ProductName 'Product's Name',
sum(od.Quantity) Quantity
from orders o
right join customers c on o.CustomerID = c.CustomerID
left join `order details` od on o.OrderID = od.OrderID
right join products p on od.ProductID = p.ProductID
group by 1,2;
```

PRODUCT'S PREFERENCE BY CUSTOMER'S GENDER

data by gender:



Top 5 products purchased by females:

Camembert Pierrot (422 units)
Raclette Courdavault (369 units)
Fløtemysost (369 units)
Lakkalikööri (325 units)
Jack's New England Clam Chowder (352 units)

Top 5 products purchased by males:

Gorgonzola Telino (1145 units)
Camembert Pierrot (1155 units)
Guaraná Fantástica (932 units)
Raclette Courdavault (1127 units)
Gnocchi di nonna Alice (816 units)

Products with the biggest gender differences in purchases:

Guaraná Fantástica (Male: 932 Female: 193)
Gorgonzola Telino (Male: 1145 Female: 252)
Gnocchi di nonna Alice (Male: 816 Female: 447)
Sir Rodney's Scones (Male: 840 Female: 176)
Gravad lax (Male: 125 Female: 0)

In summary, dairy products like cheese are popular across both genders. Males purchase substantially more units of certain specialty foods like Guaraná Fantástica, Gorgonzola Telino and pasta dishes. Females show a preference toward products like Lakkalikööri, clam chowder and Raclette Courdavault compared to males. These differences in product preferences are interesting and may be useful for targeted marketing and customizing product placement/recommendations by gender.

SQL QUERY USED FOR ABOVE CHART

```
create view Q2_CustomerGenderProductQuantity_view as
select cg.Gender 'Customer's Gender',
p.ProductName 'Product's Name',
sum(od.Quantity) Quantity
from orders o
right join customers c on o.CustomerID = c.CustomerID
left join `order details` od on o.OrderID = od.OrderID
right join products p on od.ProductID = p.ProductID
join customersgender cg on c.CustomerID = cg.CustomerID
group by 1,2;
```

PRODUCT CATEGORY'S PREFERENCE BY CUSTOMER'S OCCUPATION

Product category purchases by occupation:



Beverages

Purchased most by Sales Representatives (1865 units)
Other top purchasing roles are Sales Managers (1795 units) and Owners (1431 units)
Condiments
Sales Representatives lead purchases again (780 units)
Followed by Owners (903 units) and Accounting Managers (1128 units)

Confections

Highest purchases from Sales Representatives (1495 units)
Other top roles are Accounting Managers (1472 units) and Sales Managers (1187 units)

Dairy Products

Once again driven by Sales Representatives with 1813 units

Sales Managers have 1709 units and Accounting Managers have 1589 units

Grains/Cereals

Sales Representatives lead purchases (1021 units)
Accounting Managers follow with 790 units

Meat/Poultry

Sales Representatives most purchases at 1023 units
Sales Managers next highest with 852 units

Produce

Top purchasing role is Sales Representatives with 448 units
Followed by Owners with 534 units

Seafood

Sales Representatives lead purchases with 1836 units
Sales Associates have 722 units

SQL QUERY USED FOR ABOVE CHART

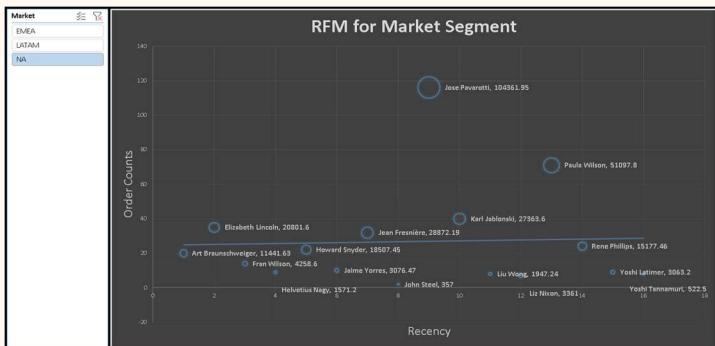
```
create view Q2_CustomerOccupationCategoryQuantity_view as
select c.ContactTitle 'Customer's Occupation',
cat.CategoryName 'Category's Name',
sum(od.Quantity) Quantity
from orders o
right join customers c on o.CustomerID = c.CustomerID
left join `order details` od on o.OrderID = od.OrderID
right join products p on od.ProductID = p.ProductID
right join categories cat on p.CategoryID = cat.CategoryID
group by 1,2;
```

EDA PROBLEM STATEMENT

3. ARE THERE ANY INTERESTING PATTERNS OR CLUSTERS IN CUSTOMER BEHAVIOR THAT CAN BE VISUALIZED TO IDENTIFY POTENTIAL MARKET SEGMENTS?

RFM FOR MARKET SEGMENT

The customer data by market:



EMEA (Europe, Middle East, Africa)

76 customers

Average Recency: 34 days

Average Monetary Value: \$13,441

Top customers by Monetary Value:

- Jose Pavarotti (\$104,362)
- Horst Kloss (\$110,277)
- Roland Mendel (\$104,875)

LATAM (Latin America)

25 customers

Average Recency: 71 days

Average Monetary Value: \$5,502

Top customers by Monetary Value:

- Mario Pontes (\$32,841)
- Carlos Hernández (\$22,769)
- Lúcia Carvalho (\$25,718)

NA (North America)

19 customers

Average Recency: 88 days

Average Monetary Value: \$9,322

Top customers by Monetary Value:

- Jean Fresnière (\$28,872)
- Paula Wilson (\$51,098)
- Elizabeth Lincoln (\$20,802)

In summary, EMEA has the most customers and highest average purchase value. NA customers have the highest average recency/longest time since last purchase. Targeting engaging recent NA purchasers and leveraging high value EMEA customers could be potential opportunities.

SQL QUERY USED FOR ABOVE CHART

Create view

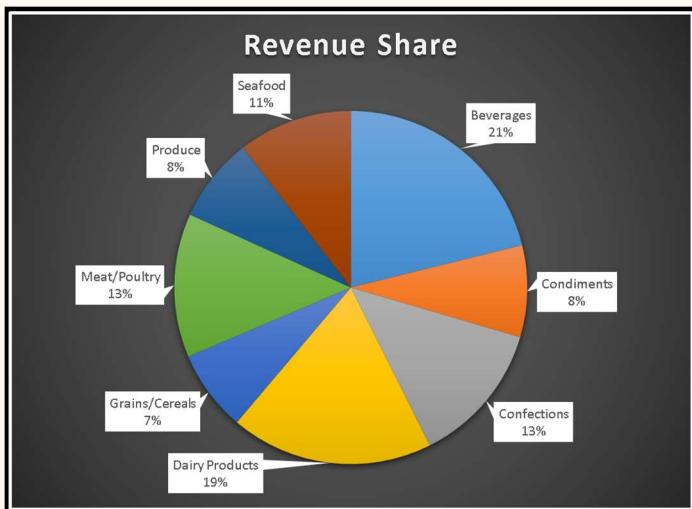
```
Q3_CustomerMarketFrequencyRecencyMonetary_view as
select c.ContactName 'Customer's Name',
cm.Market,
count(o.OrderID) Frequency,
min(datediff((select max(date(OrderDate)) from
orders),date(o.OrderDate))) Recency,
round(sum( od.UnitPrice * od.Quantity * (1 - od.Discount )),2)
Monetary
from orders o
join `order details` od on o.OrderID = od.OrderID
join customers c on c.CustomerID = o.CustomerID
join customersmarket cm on c.Country = cm.Country
group by 1,2;
```

EDA PROBLEM STATEMENT

4. ARE THERE ANY SPECIFIC PRODUCT CATEGORIES OR SKUS THAT CONTRIBUTE SIGNIFICANTLY TO ORDER REVENUE? CAN WE IDENTIFY THEM THROUGH VISUALIZATIONS?

REVENUE SHARE

The sales revenue percentage breakdown by product category:



The Dairy Products category accounts for the highest share of revenue at 18.56%. This includes products like cheese, milk, yogurt etc.

Beverages bring in the second highest revenue at 21.15%. This covers all beverages like coffee, juices, soda, beer, spirits etc.

Confections are the third largest category by revenue share, making up 13.08%. This covers sweets items like chocolate, candy, baked goods, desserts etc.

Meat/Poultry and Seafood round out the top 5 with 13.16% and 10.46% revenue share respectively.

On the lower end, Grains/Cereals have a 7.44% revenue share, followed by Produce at 7.77% share.

Condiments come in last with an 8.39% revenue share among the product categories.

In summary, the Dairy, Beverages and Confection/Dessert groups dominate in terms of overall revenue contribution. The meat, seafood and grains categories are more niche. These insights into category share of sales could help inform pricing, assortment and inventory planning decisions.

SQL QUERY USED FOR ABOVE CHART

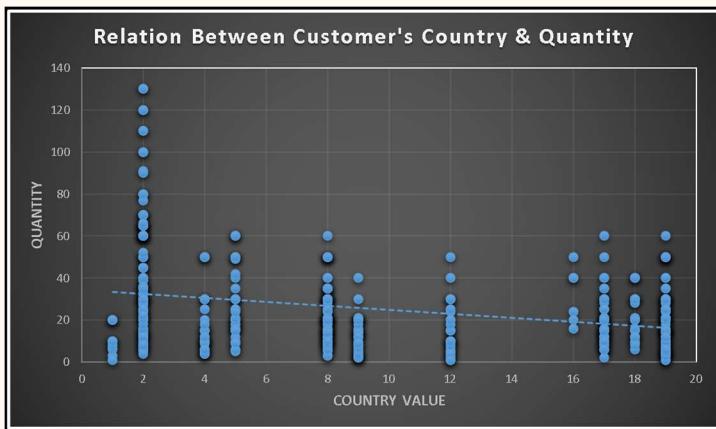
```
create view Q4_CategoryPercentageRevenue_view as
select cat.CategoryName 'Category's Name',
round(sum(od.Quantity * od.UnitPrice)*100/select
sum(Quantity * UnitPrice) from 'order details'),2) 'Percentage
Revenue'
from orders o
left join 'order details' od on o.OrderID = od.OrderID
right join products p on od.ProductID = p.ProductID
right join categories cat on p.CategoryID = cat.CategoryID
group by 1;
```

EDA PROBLEM STATEMENT

5. ARE THERE ANY CORRELATIONS BETWEEN ORDER SIZE AND CUSTOMER DEMOGRAPHICS OR PRODUCT CATEGORIES? CAN WE EXPLORE THIS VISUALLY USING SCATTER PLOTS OR HEATMAPS?

RELATION BETWEEN CUSTOMER'S COUNTRY & QUANTITY

The data covers 21 countries totaling 51,317 product units sold.



The United States leads overall unit sales with 9,330, making up over 18% of total quantity. Germany ranks second with 9,213 units sold and Brazil third with 4,247 units.

On average, 23.8 units were purchased per country transaction. The US topped the average order per transaction at 26.5 units. Switzerland had the second highest order average at 24.5 units and Ireland third at 30.6 units.

In terms of order variability, larger fluctuations appear present in the US, Germany and Brazil data indicating less consistent purchase quantities. Switzerland, Austria and France show lower variance pointing to more stable order volumes.

Across the 21 countries, five nations accounted for over 63% of total global product unit purchases:

- 1) USA
- 2) Germany
- 3) Brazil
- 4) France
- 5) Austria

To sustain long-term growth, both continuing strong performance in these top markets will be important along with unlocking potential in other countries through customized regional strategies to drive increased product sales worldwide.

SQL QUERY USED FOR ABOVE CHART

```
create view Q5_CustomerCountryValueQuantity_view as
select c.Country 'Customer's Country', cm.CountryValue
'Country Value', od.Quantity
from orders o
left join `order details` od on o.OrderID = od.OrderID
right join customers c on c.CustomerID = o.CustomerID
join customersmarket cm on c.Country = cm.Country;
```

CORRELATION QUERY

```
with varchange as(select
'Country Value' x1, -- Variable 1
'Quantity' x2 -- Variable 2
from
q5_customercountryvaluequantity_view -- Table Name
),
mean as(select avg(x1) x1_bar, avg(x2) x2_bar from varchange),
deviation as(select (x1 - (select x1_bar from mean)) x1_dev,
(x2 - (select x2_bar from mean)) x2_dev from varchange)
select
sum(x1_dev)
x2_dev)/sqrt(sum(pow(x1_dev,2))*sum(pow(x2_dev,2)))
CORRELATION
from deviation;
```

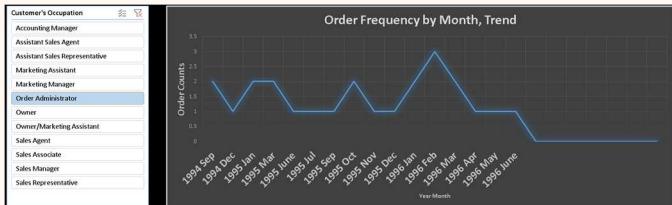
*

EDA PROBLEM STATEMENT

6. HOW DOES ORDER FREQUENCY VARY ACROSS DIFFERENT CUSTOMER SEGMENTS? CAN WE VISUALIZE THIS USING BAR CHARTS OR TREEMAPS?

ORDER FREQUENCY BY MONTH, TREND

Breakdown of customer orders by their occupation, spanning from August 1994 to June 1996. Here's a more detailed summary:



August 1994:

Most frequent occupations: Accounting Manager (6), Owner (5), Marketing Manager (4), Sales Manager (4). Other occupations placing orders: Sales Representative (1), Assistant Sales Representative (1), Sales Agent (1).

September 1994:

Most frequent occupations: Accounting Manager (9), Sales Representative (4), Order Administrator (2), Sales Associate (2).

Other occupations placing orders: Assistant Sales Representative (2), Owner (2), Marketing Manager (1), Marketing Assistant (1), Assistant Sales Agent (1), Sales Manager (1).

October 1994:

Most frequent occupations: Accounting Manager (5), Marketing Manager (4), Owner (3), Sales Manager (3), Sales Representative (3).

Other occupations placing orders: Sales Associate (2), Assistant Sales Representative (2), Assistant Sales Agent (1).

November 1994:

Most frequent occupations: Owner (6), Marketing Manager (4), Sales Representative (4), Sales Associate (3), Accounting Manager (3).

Other occupations placing orders: Sales Manager (2), Marketing Assistant (2), Sales Agent (1).

December 1994:

Most frequent occupations: Sales Manager (9), Sales Representative (5), Owner (3), Accounting Manager (3), Marketing Manager (2).

Other occupations placing orders: Order Administrator (1), Sales Agent (1), Marketing Assistant (1), Assistant Sales Representative (1).

The summary continues with detailed breakdowns for each month up to June 1996, providing insights into the changing patterns of customer orders across different occupations over time.

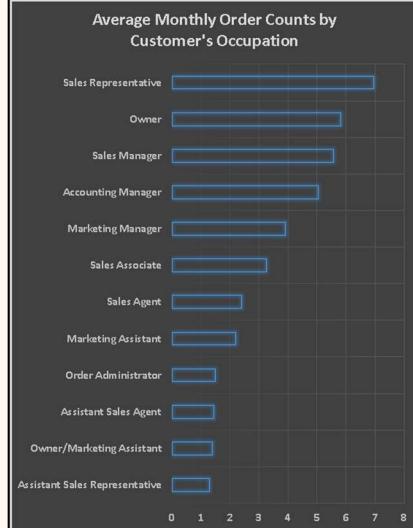
SQL QUERY USED FOR ABOVE CHART

create view

```
Q6_YearMonthCustomerOccupationOrderCounts_view as
select year(date(o.OrderDate)) 'Year',
month(date(o.OrderDate)) 'Month',
c.ContactTitle 'Customer's Occupation', count(o.OrderID)
'Order Counts'
from orders o
join customers c on o.CustomerID = c.CustomerID
group by 1,2,3
order by 1,2,4 desc;
```

AVERAGE MONTHLY ORDER COUNTS BY CUSTOMER'S OCCUPATION

Order Trends by Customer Occupation



This data reveals interesting patterns in average monthly order counts across different customer occupations.

High-Volume Orders:

Sales Representatives: Unsurprisingly, sales representatives hold the top spot with an average of 6.96 orders per month, making them the key drivers of sales.

Owners: Following closely behind with 5.83 orders per month, owners play a significant role in generating sales volume.

Sales Managers: Averaging 5.57 orders per month, sales managers likely contribute to sales generation through their teams.

Moderate-Volume Orders:

Accounting Managers: Responsible for financial aspects, accounting managers have a moderate average of 5.04 orders per month, suggesting potential involvement in procurement or related purchases.

Marketing Managers: With an average of 3.9 orders per month, marketing managers might contribute to orders indirectly through promotional efforts.

Low-Volume Orders:

Sales Associates: Averaging 3.25 orders per month, sales associates likely handle smaller or individual customer transactions.

Marketing Assistants: Their 2.2 average orders per month might indicate involvement in smaller promotional orders or resource requests.

Support Staff: Roles like order administrators, assistant sales representatives, and assistant sales agents have even lower averages, suggesting they handle smaller or infrequent orders.

Additional Insights:

The gap between sales representatives and other roles highlights their crucial contribution to order generation.

The presence of owners and sales managers in the top order brackets suggests their active involvement in sales efforts.

Marketing and support staff contribute through lower-volume orders, possibly indicating their focus on specific tasks or customer interactions.

AVERAGE MONTHLY ORDER COUNTS BY MARKET & COUNTRY

Data reveals interesting variations in average order counts across different customer markets and countries.



High-Volume Orders:

North America (NA): Both USA and Canada boast the highest average order counts, exceeding 5 orders per customer. This suggests strong customer engagement and potentially larger order sizes in these markets.

EMEA: Germany stands out with an impressive average of 5.3 orders per customer, indicating potentially higher individual spending or larger business purchases.

Moderate-Volume Orders:

EMEA: Several European countries fall within this range, including France, Sweden, Austria, and Belgium, with averages between 1.4 and 3.3 orders per customer. This suggests diverse customer behavior and potentially smaller order sizes compared to leaders.

LATAM: Mexico and Venezuela fall into this category, with averages between 1.6 and 2.2 orders per customer, hinting at moderate customer engagement or smaller purchase sizes in these markets.

Low-Volume Orders:

EMEA: Several European countries including Switzerland, Finland, Italy, Spain, UK, Ireland, Portugal, Denmark, Norway, and Poland have averages below 2 orders per customer. This could indicate lower customer activity, smaller order sizes, or specific market dynamics in these regions.

LATAM: Argentina joins this group with an average of 1.6 orders per customer, suggesting similar trends in customer behavior or order patterns compared to some European counterparts.

Additional Insights:

The significant difference between North America and other regions suggests potential market-specific factors influencing order behavior.

The spread within EMEA highlights diverse customer characteristics and buying habits across different European countries.

Further analysis could involve segmenting data by product category or customer type for more granular insights.

SQL QUERY USED FOR ABOVE CHART

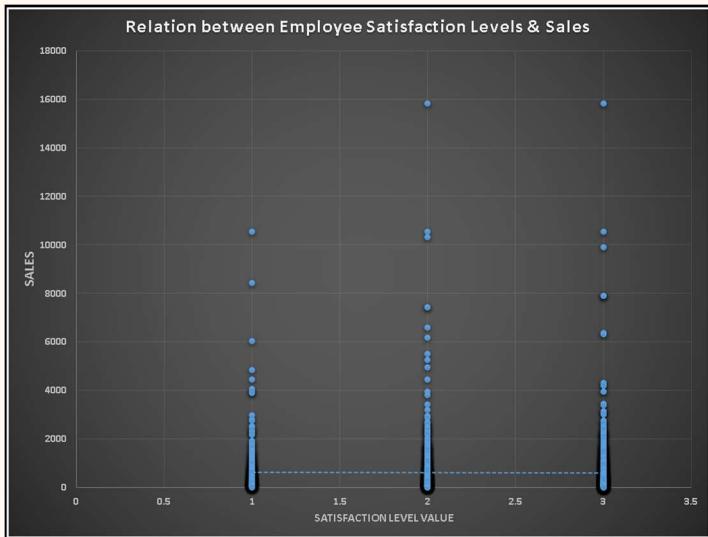
```
create view Q6_YearMonthMarketCountryOrderCounts_view as
with cte as(
select year(date(o.OrderDate)) 'Year',
month(date(o.OrderDate)) 'Month',
cm.Market 'Customer's Market',
c.Country 'Customer's Country', count(o.OrderID) 'Order
Counts'
from orders o
join customers c on o.CustomerID = c.CustomerID
join customersmarket cm on c.Country = cm.Country
group by 1,2,3,4
order by 1,2,5 desc)
select 'Customer's Market', 'Customer's Country', avg('Order
Counts') 'Average Orders Count'
from cte
group by 1,2;
```

EDA PROBLEM STATEMENT

7. ARE THERE ANY CORRELATIONS BETWEEN EMPLOYEE SATISFACTION LEVELS AND KEY PERFORMANCE INDICATORS? CAN WE EXPLORE THIS VISUALLY THROUGH SCATTER PLOTS OR LINE CHARTS?

RELATION BETWEEN EMPLOYEE SATISFACTION LEVELS & SALES

Scatter chart shown as follows:



SQL QUERY USED FOR ABOVE CHART

```
create view Q7_SatisfactionSales_view as
select case
when eg.Satisfaction = "High" then 3
when eg.Satisfaction = "Normal" then 2
when eg.Satisfaction = "Low" then 1
else null
end 'Satisfaction Value',
odv.Sales
from orders o
join order_details_view odv on o.OrderID = odv.OrderID
join employees e on o.EmployeeID = e.EmployeeID
join employeesgender eg on e.EmployeeID = eg.EmployeeID;
```

CORRELATION QUERY

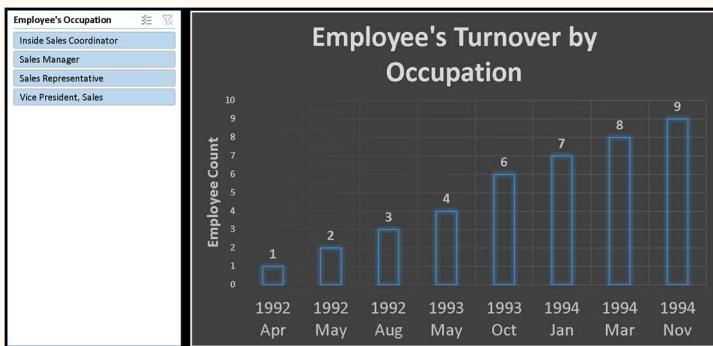
```
with varchange as(select
'Satisfaction Value` x1, -- Variable 1
'Sales` x2 -- Variable 2
from
Q7_SatisfactionSales_view -- Table Name
),
mean as(select avg(x1) x1_bar, avg(x2) x2_bar from varchange),
deviation as(select (x1 - (select x1_bar from mean)) x1_dev,
(x2 - (select x2_bar from mean)) x2_dev from varchange)
select          sum(x1_dev)           *
x2_dev)/sqrt(sum(pow(x1_dev,2)*sum(pow(x2_dev,2)))
CORRELATION
from deviation;
```

EDA PROBLEM STATEMENT

8. HOW DOES EMPLOYEE TURNOVER VARY ACROSS DIFFERENT DEPARTMENTS OR JOB ROLES? CAN WE VISUALIZE THIS USING BAR CHARTS OR HEATMAPS?

EMPLOYEE'S TURNOVER BY OCCUPATION

Employee's Arrival by Occupation:



Sales Representative: 4

Vice President, Sales: 1

Sales Manager: 2

Inside Sales Coordinator: 1

Written Summary:

The analysis of employee arrivals by occupation reveals that the Sales Representative role had the highest number of arrivals, with four employees joining during the specified periods. Additionally, one employee each joined in the roles of Vice President, Sales, Sales Manager, and Inside Sales Coordinator. Understanding the influx of new employees across different roles provides insights into recruitment patterns and areas where the company may be focusing its hiring efforts.

SQL QUERY USED FOR ABOVE CHART

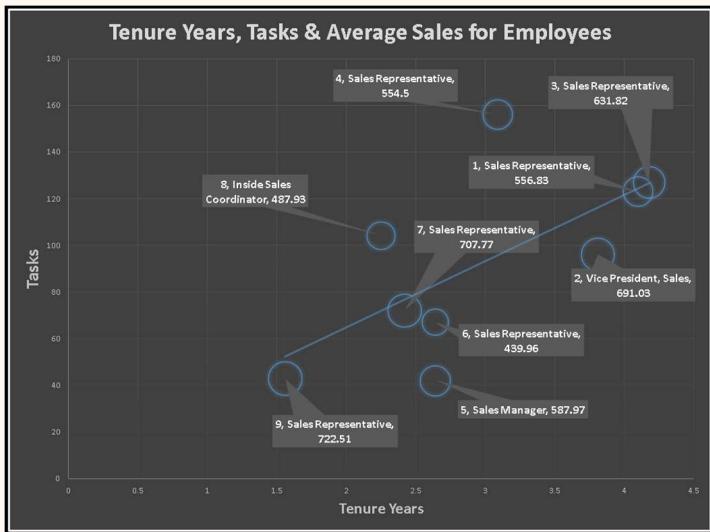
```
create view Q8_YearMonthEmployeeOccupationCount_view as
select year(date(HireDate)) 'Year',
month(date(HireDate)) 'Month',
Title 'Employee's Occupation',
count(*) 'Count of Employee Joined'
from employees
group by 1,2,3
order by 1,2,4 desc;
```

EDA PROBLEM STATEMENT

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?

TENURE YEARS, TASKS & AVERAGE SALES FOR EMPLOYEES

The summary presents information about nine employees, including their names, tenure years, assigned tasks, and average net sales.



Nancy Davolio, with a tenure of 4.1 years, handles 123 tasks on average with an average net sales figure of \$556.83. Andrew Fuller, serving for 3.81 years, manages 96 tasks and achieves an average net sales of \$691.03. Janet Leverling, with a tenure of 4.18 years, handles 127 tasks on average, achieving an average net sales figure of \$631.82. Margaret Peacock, with a tenure of 3.09 years, manages 156 tasks on average with an average net sales of \$554.5. Steven Buchanan, serving for 2.64 years, manages 42 tasks on average, achieving an average net sales of \$439.96. Michael Suyama, also with a tenure of 2.64 years, handles 67 tasks on average, achieving an average net sales figure of \$487.93. Robert King, with a tenure of 2.42 years, manages 72 tasks on average, with an average net sales figure of \$707.77. Laura Callahan, with a tenure of 2.25 years, handles 104 tasks on average, achieving an average net sales of \$722.51.

This summary provides insights into the tenure lengths, task assignments, and sales performance of each employee.

SQL QUERY USED FOR ABOVE CHART

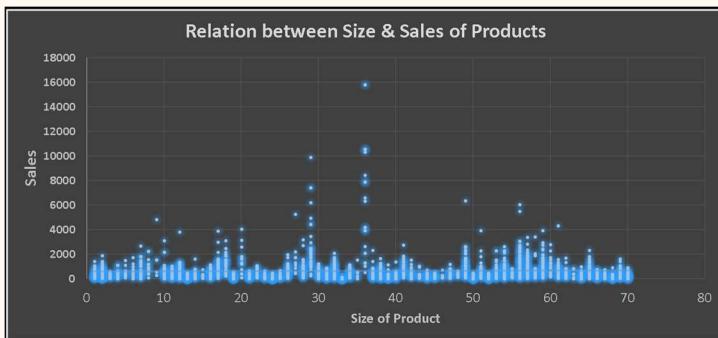
```
create view
Q9_EmployeeIDNameTitleTasksAvgSalesTenure_view as
with cte1 as(
select e.EmployeeID 'Employee ID',
concat(e.FirstName, " ", e.LastName) 'Employee Name',
e.Title,
count(distinct o.OrderID) Tasks,
round(avg(NetSales),2) 'Average Net Sales'
from employees e
join orders o on e.EmployeeID = o.EmployeeID
join order_details_view odv on odv.OrderID = o.OrderID
group by 1,2,3),
cte2 as(
select concat(FirstName, " ", LastName) 'Employee Name',
round(datediff((select max(date(OrderDate)) from
orders),HireDate)/365,2) 'Tenure Years'
from employees)
select c1.* , c2.'Tenure Years'
from cte1 c1
join cte2 c2 on c1.'Employee Name' = c2.'Employee Name';
```

EDA PROBLEM STATEMENT

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?

RELATION BETWEEN SIZE & SALES OF PRODUCTS

Scatter chart shown as follows:



SQL QUERY USED FOR ABOVE CHART

```
create view Q10_SizeSales_view as
with cte as(
select QuantityPerUnit, row_number() over() SizeValue
from products
group by 1)
select p.QuantityPerUnit, c.SizeValue, odv.Sales
from products p
join order_details_view odv on p.ProductID = odv.ProductID
join cte c on c.QuantityPerUnit = p.QuantityPerUnit;
```

CORRELATION QUERY

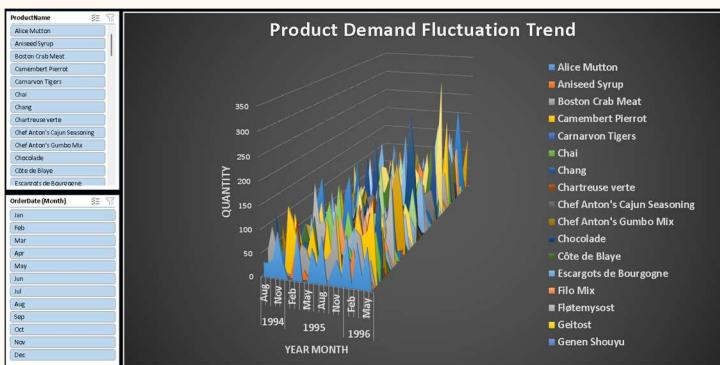
```
with varchange as(select
`SizeValue` x1, -- Variable 1
`Sales` x2 -- Variable 2
from
Q9_SizeSales_view -- Table Name
),
mean as(select avg(x1) x1_bar, avg(x2) x2_bar from varchange),
deviation as(select (x1 - (select x1_bar from mean)) x1_dev,
(x2 - (select x2_bar from mean)) x2_dev from varchange)
select
sum(x1_dev) *
x2_dev)/sqrt(sum(pow(x1_dev,2)*sum(pow(x2_dev,2)))
CORRELATION
from deviation;
```

EDA PROBLEM STATEMENT

11. HOW DOES PRODUCT DEMAND FLUCTUATE OVER DIFFERENT SEASONS OR MONTHS? CAN WE VISUALIZE THIS THROUGH LINE CHARTS OR AREA CHARTS?

PRODUCT DEMAND FLUCTUATION TREND

Area chart shown as follows:

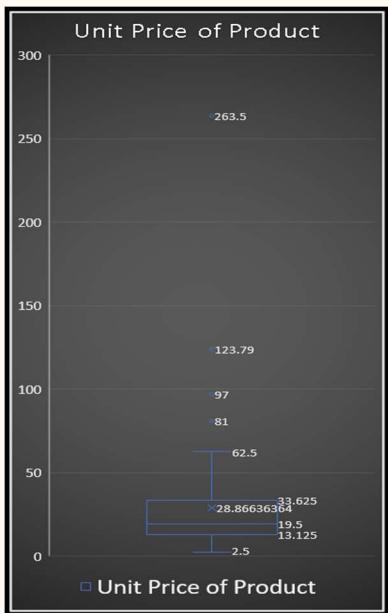


EDA PROBLEM STATEMENT

12. CAN WE IDENTIFY ANY OUTLIERS OR ANOMALIES IN PRODUCT PERFORMANCE OR SALES USING VISUALIZATIONS? HOW CAN THIS INFORMATION BE USED FOR PRODUCT OPTIMIZATION?

UNIT PRICE OF PRODUCT

Product Sales Summary: High Earners, Low Earners, and Interesting Trends



This data reveals a wide range of average sales per product, highlighting some clear winners and interesting patterns.

High Earners:

Luxury and Specialty Items: Products like Côte de Blaye, Schoggi Schokolade, Thüringer Rostbratwurst, Sir Rodney's Marmalade, and Manjimup Dried Apples stand out with exceptionally high average sales, suggesting strong demand for premium or unique offerings.

Cheese and Dairy: Several cheeses like Raclette Courdavault, Camembert Pierrot, Mozzarella di Giovanni, and Gudbrandsdalost rank high, indicating their popularity and potential premium pricing.

Seafood and Meats: Items like Boston Crab Meat, Jack's New England Clam Chowder, and Pâté chinois perform well, suggesting their appeal to specific customer segments or higher-end ingredients.

Low Earners:

Spreads and Baking: Products like Zaanse koeken, Chocolade, Filo Mix, and Vegie-spread have lower average sales, possibly due to lower price points, smaller serving sizes, or less frequent purchase occasions.

Beverages: While some drinks like Ipoh Coffee and Sirop d'éralbe fare well, others like Guaraná Fantástica, Gumbär Gummibärchen, and Scottish Longbreads have lower averages, suggesting potentially lower price points or niche markets.

Condiments and Sauces: While Louisiana Fiery Hot Pepper Sauce and Louisiana Hot Spiced Okra perform decently, others like Genen Shouyu and Teatime Chocolate Biscuits have lower averages, possibly indicating more competitive categories or smaller usage quantities.

Interesting Trends:

Seafood and Dairy Dominate High Earners: These categories hold many of the top performers, suggesting potential for premiumization or focus on specific customer segments.

Wide Range of Prices: The data spans a vast range of average sales, highlighting the diverse product portfolio and potentially catering to various customer needs and budgets.

Sweet vs. Savory: While some sweet treats like Sir Rodney's Marmalade and Manjimup Dried Apples perform well, others like Chocolade and Filo Mix have lower averages, suggesting varying popularity or purchase frequency for different taste profiles.

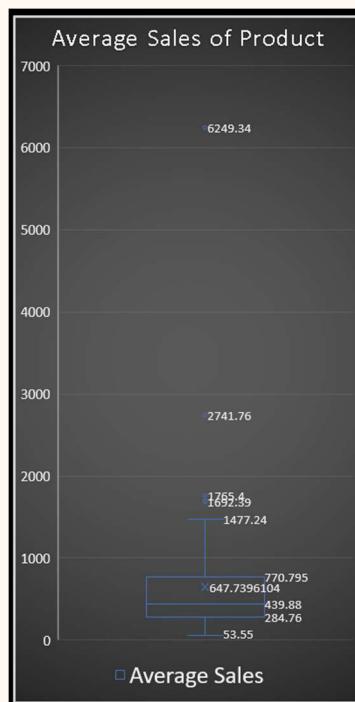
This summary provides a starting point for further analysis. You could explore factors influencing sales like product category, pricing strategy, marketing efforts, or customer demographics to gain deeper insights.

SQL QUERY USED FOR ABOVE CHART

```
create view Q12_ProductUnit_view as
select ProductName 'Product Name',
round(UnitPrice,2) 'Unit Price of Product'
from products;
```

AVERAGE SALES OF PRODUCT

Product Price Analysis: High & Low Cost, Interesting Trends



This data reveals a wide range of unit prices across various products, offering insights into potential cost drivers and buying considerations.

High-Cost Products:

Luxury and Specialty Items: Products like Côte de Blaye, Schoggi Schokolade, Thüringer Rostbratwurst, Carnarvon Tigers, and Mishi Kobe Niku have significantly higher unit prices, reflecting their premium ingredients, unique offerings, or targeted customer segments.

Seafood and Meats: Several seafood and meat items like Boston Crab Meat, Jack's New England Clam Chowder, Alice Mutton, and Gorgonzola Telino hold higher price points, suggesting potentially higher-quality ingredients, specific cuts, or imported origins.

Cheeses and Dairy: Some cheeses like Raclette Courdavault, Camembert Pierrot, and Mozzarella di Giovanni fall into the higher price bracket, possibly due to aging processes, specific production methods, or imported varieties.

Low-Cost Products:

Beverages: Many beverages like Guaraná Fantástica, Zaanse koeken, Chocolade, and Rhönbräu Klosterbier have lower unit prices, possibly due to readily available ingredients, mass production, or targeting budget-conscious customers.

Staples and Baking: Products like Tofu, Longlife Tofu, Filo Mix, and Vegie-spread tend to have lower prices, potentially reflecting basic ingredients, larger serving sizes, or frequent purchase occasions.

Condiments and Sauces: While some sauces like Louisiana Hot Pepper Sauce and Genen Shouyu have moderate prices, others like Teatime Chocolate Biscuits and Original Frankfurter grüne Soße are lower-priced, suggesting less complex ingredients or wider usage scenarios.

Interesting Trends:

Wide Price Spectrum: The data encompasses a vast range of unit prices, highlighting the diverse product portfolio and catering to various customer needs and budgets.

Seafood and Meat Premiumization: Many high-cost products belong to these categories, suggesting potential for focusing on quality ingredients, specific cuts, or targeted marketing for premium offerings.

Beverage Price Variation: While some drinks command higher prices for potentially unique flavors or ingredients, others offer lower price points, catering to different preferences and consumption occasions.

Cheese Price Diversity: Prices vary greatly within the cheese category, reflecting diverse production methods, aging processes, and potentially targeted customer segments.

This summary provides a starting point for further analysis. You could explore the relationship between price and factors like product category, ingredients, brand positioning, or customer demographics to gain deeper insights and inform pricing strategies.

SQL QUERY USED FOR ABOVE CHART

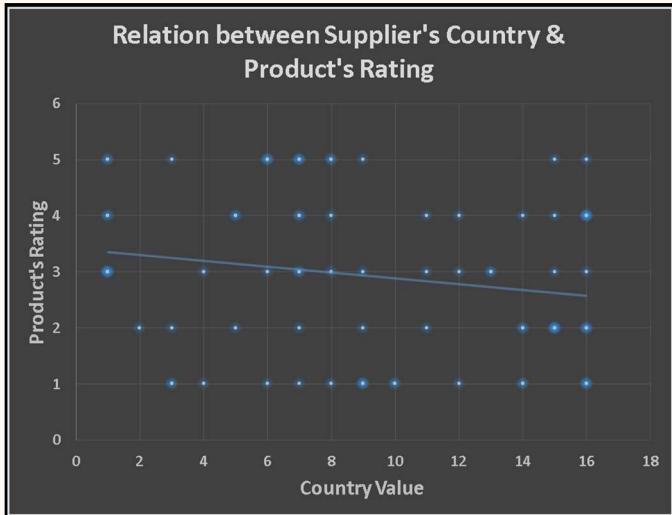
```
create view Q12_ProductAvgSales_view as
select p.ProductName 'Product Name',
round(avg(odv.Sales),2) 'Average Sales'
from products p
join order_details_view odv on p.ProductID = odv.ProductID
group by 1;
```

EDA PROBLEM STATEMENT

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?

RELATION BETWEEN SUPPLIER'S COUNTRY & PRODUCT'S RATING

Country Rating Summary: Diverse Feedback and Variances



This data reveals a varied landscape of country ratings, highlighting both positive and negative feedback across different regions.

Top-Rated Countries:

USA: While some ratings fall lower, the USA exhibits the highest concentration of positive feedback (3, 4, and 5 ratings), suggesting overall satisfaction from many respondents.

Australia: Despite having some lower ratings, Australia shows several positive responses (3, 4, and 5), indicating generally favorable experiences for some customers.

Sweden: Although not as frequent as the USA, Sweden also has positive feedback (2 and 4 ratings), suggesting at least some satisfied customers.

Mixed Feedback:

UK: The UK displays a mix of ratings (2, 3, 4, and 5), suggesting both positive and negative experiences among respondents.

Germany: Similar to the UK, Germany exhibits a range of ratings (1, 3, 4, and 5), indicating diverse feedback and potentially areas for improvement.

Italy: Italy also shows mixed feedback (1, 3, 4, and 5), suggesting varying customer experiences and satisfaction levels.

Lower-Rated Countries:

France: While some positive ratings exist (3 and 5), France has more frequent lower ratings (1 and 3), suggesting potential areas for improvement and addressing customer concerns.

Canada: Although not the lowest, Canada has some lower ratings (1 and 2), indicating room for addressing specific issues or concerns to enhance customer satisfaction.

Others: Countries like Singapore, Denmark, Netherlands, and Finland have mostly single ratings, making it difficult to draw definitive conclusions but highlighting the need for more data to understand their customer sentiment.

Additional Insights:

The high number of USA ratings suggests a larger customer base or more active participation, potentially influencing the overall distribution of feedback.

It's crucial to consider the context and purpose of these ratings for a more nuanced understanding of customer sentiment in each country.

Further analysis could involve segmenting data by specific product, service, or customer type for more targeted insights and action plans.

This summary provides a starting point for further exploration. You can delve deeper into specific countries, rating categories, or influencing factors to gain actionable insights and improve customer experience in different regions.

SQL QUERY USED FOR ABOVE CHART

```
create view Q13_CountryValueRating_view as
with cte as(
select Country, row_number() over(order by Country) `Country Value`
from suppliers
group by 1)
select c.* , pr.Rating
from products p
join suppliers s on p.SupplierID = s.SupplierID
join productsrating pr on p.ProductID = pr.ProductID
join cte c on c.Country = s.Country;
```

CORRELATION QUERY

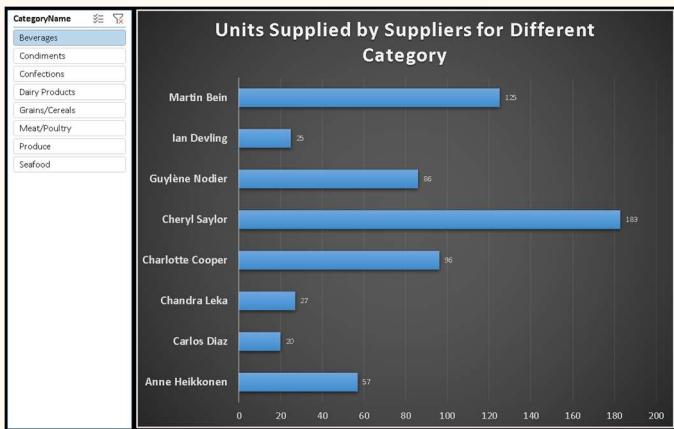
```
with varchange as(select
`Country Value` x1, -- Variable 1
`Rating` x2 -- Variable 2
from
Q13_CountryValueRating_view -- Table Name
),
mean as(select avg(x1) x1_bar, avg(x2) x2_bar from varchange),
deviation as(select (x1 - (select x1_bar from mean)) x1_dev,
(x2 - (select x2_bar from mean)) x2_dev from varchange)
select sum(x1_dev *
x2_dev)/sqrt(sum(pow(x1_dev,2))*sum(pow(x2_dev,2)))
CORRELATION
from deviation;
```

EDA PROBLEM STATEMENT

14. HOW DOES SUPPLIER PERFORMANCE VARY ACROSS DIFFERENT PRODUCT CATEGORIES OR DEPARTMENTS? CAN WE VISUALIZE THIS USING STACKED BAR CHARTS OR GROUPED COLUMN CHARTS?

UNITS SUPPLIED BY SUPPLIERS FOR DIFFERENT CATEGORY

Supplier Performance by Category: Key Highlights



This data reveals interesting patterns in units supplied by different suppliers across various product categories.

Top Suppliers by Category:

Beverages: Charlotte Cooper leads with 96 units, followed by Cheryl Saylor (183) and Martin Bein (125).

Condiments: Shelley Burke tops the list with 233 units, followed by Regina Murphy (126) and Chantal Goulet (113).

Confections: Petra Winkler takes the lead with 140 units, followed by Peter Wilson (124) and Anne Heikonen (135).

Dairy Products: Beate Vileid stands out with 164 units, followed by Antonio del Valle Saavedra (138) and Elio Rossi (133).

Grains/Cereals: Lars Peterson holds the top spot with 165 units, followed by Beate Vileid (164) and Martin Bein (102).

Meat/Poultry: Jean-Guy Lauzon dominates with 136 units, significantly surpassing all other suppliers in this category.

Produce: Beate Vileid leads with 164 units (also top in Dairy Products), followed by Regina Murphy (15) and Mayumi Ohno (35).

Seafood: Michael Björn significantly outperforms others with 274 units, followed by Robb Merchant (208) and Niels Petersen (170).

Category Insights:

Beverages: Several suppliers contribute moderate amounts, suggesting a potentially competitive landscape.

Condiments and Confections: A few suppliers hold larger shares, potentially indicating stronger partnerships or specialization.

Dairy Products and Grains/Cereals: Beate Vileid's performance across two categories suggests potential for broader product offerings or strategic partnerships.

Meat/Poultry: Jean-Guy Lauzon's dominance highlights their focus or expertise in this category.

Produce: The spread of suppliers suggests diverse sourcing or smaller order sizes.

Seafood: Michael Björn's exceptional contribution warrants further investigation into their sourcing capabilities or market position.

Further Exploration:

Analyze trends over time to see if supplier performance changes.

Segment data by region or customer to identify specific buying patterns.

Investigate factors influencing supplier selection, like pricing, quality, or reliability.

This summary provides a springboard for deeper analysis to optimize supplier relationships and potentially improve product sourcing strategies.

SQL QUERY USED FOR ABOVE CHART

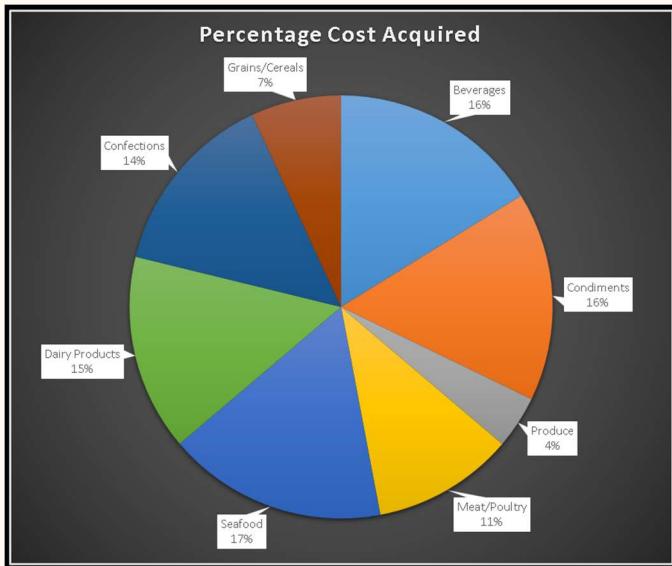
```
create view Q14_SupplierCategoryUnits_view as
select s.ContactName, c.CategoryName,
sum(p.UnitsInStock+p.UnitsOnOrder) 'Units Supplied by
Supplier'
from products p
join suppliers s on p.SupplierID = s.SupplierID
join categories c on c.CategoryID = p.CategoryID
group by 1,2;
```

EDA PROBLEM STATEMENT

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?

PERCENTAGE COST ACQUIRED

Category Cost Breakdown: Top Spenders and Potential Savings



This data reveals the distribution of acquired costs across different product categories, offering insights into spending patterns and potential areas for optimization.

Top Spending Categories:

Seafood: Accounting for 16.78% of costs, seafood is the most significant expense, highlighting its potential impact on overall spending.

Beverages: At 16.18%, beverages follow closely, suggesting another essential area for cost management.

Condiments: With 15.99%, condiments represent another notable expense category.

Moderate Spending Categories:

Dairy Products: Dairy products hold a 15.01% share, indicating moderate spending compared to top categories but still meriting attention.

Meat/Poultry: Meat and poultry represent 10.8% of costs, suggesting another area for potential cost control measures.

Lower Spending Categories:

Confections: At 14.28%, confectionery items account for a smaller portion of costs but could still benefit from strategic sourcing or negotiation.

Grains/Cereals: Grains and cereals hold the lowest share at 6.9%, suggesting potentially less room for cost optimization compared to other categories.

Additional Insights:

The top three categories (Seafood, Beverages, Condiments) combined make up almost half (48.95%) of the total cost, highlighting their significant impact and potential for cost reduction efforts.

Analyzing cost drivers within each category could reveal specific areas for negotiation, alternative sourcing, or recipe adjustments.

Comparing costs to industry benchmarks or competitor pricing might identify potential savings opportunities.

This summary provides a starting point for strategic cost management. By delving deeper into each category and exploring cost drivers, you can identify actionable steps to optimize spending and potentially improve profitability.

SQL QUERY USED FOR ABOVE CHART

```
create view Q15_CategoryCostAcquired_view as
with cte1 as(
select ProductID,
sum(NetSales)/sum(Quantity) 'Updated UnitPrice'
from order_details_view
group by 1),
cte2 as(
select c.ProductID,
round(p.UnitPrice-c.'Updated UnitPrice',2) DeficitPrice
from cte1 c
join products p on p.ProductID = c.ProductID),
cte3 as(
select cat.CategoryName,
round(sum(c.DeficitPrice * (p.UnitsInStock +
p.UnitsOnOrder)),2) 'Cost Acquired'
from cte2 c
join products p on c.ProductID = p.ProductID
join categories cat on cat.CategoryID = p.CategoryID
group by 1)
select CategoryName,
round((`Cost Acquired`*100)/(select sum(`Cost Acquired`) from
cte3),2) 'Percentage Cost Acquired'
from cte3
```

POWER BI PROBLEM STATEMENT

1. HOW DOES CUSTOMER DISTRIBUTION VARY ACROSS DIFFERENT REGIONS OR CUSTOMER SEGMENTS? CAN WE VISUALIZE IT ON A MAP OR BAR CHART?

VISUALIZATION



The data encompasses a diverse range of cities across multiple countries. London in the UK has the highest count of customers with 6, followed closely by México D.F. in Mexico with 5 customers. São Paulo in Brazil, Buenos Aires in Argentina, and Madrid in Spain each have 3 customers. Other cities, such as Paris in France, Lisboa in Portugal, and Portland in the USA, have 2 customers each. Additionally, there are numerous cities with 1 customer, spread across various countries including Austria, Belgium, Canada, Denmark, Finland, Germany, Ireland, Italy, Norway, Poland, Sweden, Switzerland, the UK, the USA, and Venezuela.

In terms of overall distribution, the USA has the highest total count of customers, with 22 spread across different cities. Brazil follows with a total of 10 customers, and France and the UK tie with 10 customers each. Other countries have fewer customers, ranging from 1 to 7.

Overall, the data showcases the global distribution of customers across different cities, reflecting the diversity of the customer base.

ATTRIBUTES FOR ABOVE VISUALIZATION

Location

- Country
- City

Legend

- Add data fields here

Latitude

- Add data fields here

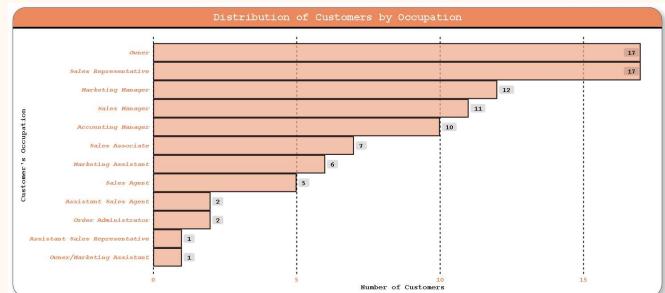
Longitude

- Add data fields here

Bubble size

- Count of Customers

VISUALIZATION



The majority of customers fall into two primary occupations: Owner and Sales Representative, each with 17 individuals. Following closely are Marketing Manager with 12 individuals and Sales Manager with 11 individuals. Additionally, there are 10 Accounting Managers and 7 Sales Associates.

Other occupations include Marketing Assistant with 6 individuals, Sales Agent with 5 individuals, and Assistant Sales Agent with 2 individuals. There are also 2 Order Administrators and 1 individual each in the occupations of Assistant Sales Representative and Owner/Marketing Assistant.

Overall, the data suggests a diverse range of occupations among the customer base, with a significant representation in ownership and sales-related roles, alongside managerial and support positions in marketing and accounting.

ATTRIBUTES FOR ABOVE VISUALIZATION

Y-axis

- Customer's Occupation

X-axis

- Number of Customers

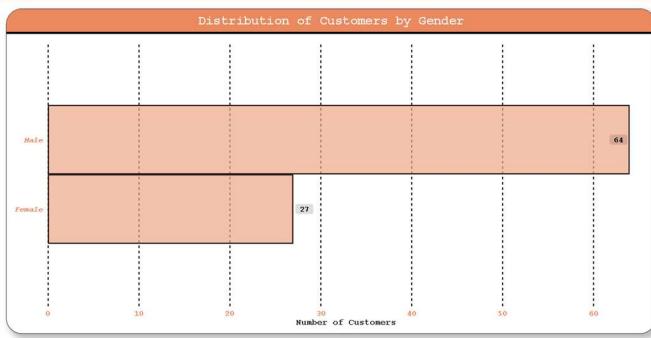
Legend

- Add data fields here
- Small multiples
- Add data fields here

Tooltips

- Add data fields here

VISUALIZATION



The provided data presents the gender distribution among customers, indicating that there are 64 male customers and 27 female customers.

In summary, there is a higher representation of male customers compared to female customers in the dataset, with males comprising the majority of the customer base.

ATTRIBUTES FOR ABOVE VISUALIZATION

Y-axis

Gender

X-axis

Number of Customers

Legend

Add data fields here

Small multiples

Add data fields here

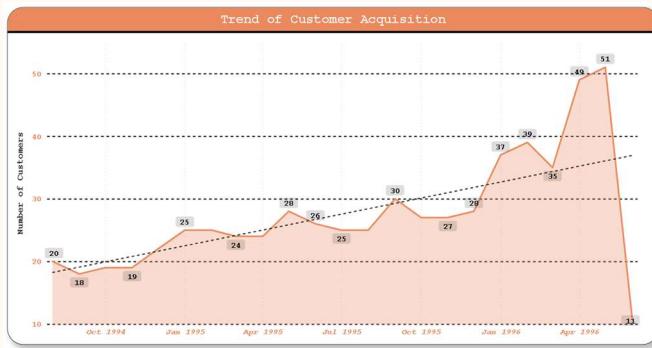
Tooltips

Add data fields here

POWER BI PROBLEM STATEMENT

2. WHAT IS THE TREND IN CUSTOMER ACQUISITION OVER TIME? CAN WE CREATE A LINE OR AREA CHART TO DISPLAY IT?

VISUALIZATION



ATTRIBUTES FOR ABOVE VISUALIZATION

| X-axis | |
|---------|-----|
| Date | ✓ X |
| Year | X |
| Quarter | X |
| Month | X |
| Day | X |

| Y-axis | |
|---------------------|-----|
| Number of Customers | ✓ X |

| Secondary y-axis | |
|----------------------|--|
| Add data fields here | |

| Legend | |
|----------------------|--|
| Add data fields here | |

The dataset provides a detailed record of the number of customers per day over several quarters spanning from 1994 to 1996. Here's a summary of the data:

The dataset covers multiple years, starting from 1994 to 1996.

Each year is divided into four quarters: Qtr 1, Qtr 2, Qtr 3, and Qtr 4.

Each quarter is further divided into months, including January, February, March, April, May, June, July, August, September, October, November, and December.

For each day within a month, the dataset records the number of customers.

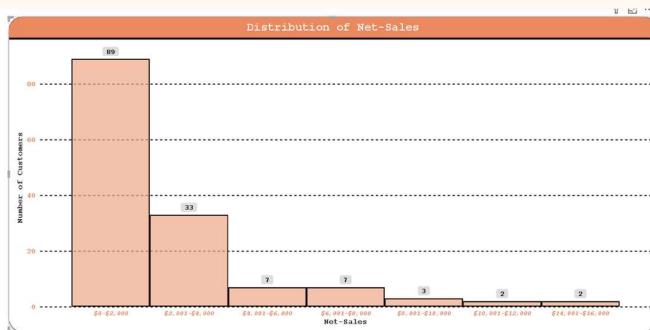
The data appears to be structured to track customer activity over time, possibly for analytical or forecasting purposes. This level of granularity allows for detailed analysis of customer trends, such as identifying peak periods of customer activity, seasonal fluctuations, or long-term growth patterns. It can also help businesses optimize staffing, inventory management, and marketing strategies based on historical customer behavior.

Overall, this dataset provides a rich source of information for understanding customer dynamics over time and can be leveraged for strategic decision-making and business planning purposes.

POWER BI PROBLEM STATEMENT

3. CAN WE VISUALIZE THE DISTRIBUTION OF CUSTOMER DEMOGRAPHICS SUCH AS AGE, GENDER, OR INCOME USING HISTOGRAMS OR PIE CHARTS?

VISUALIZATION



The dataset provides information on the distribution of customers based on their net sales ranges. Here's a summary:

Net Sales Ranges: The net sales ranges are categorized into several brackets:

\$0-\$2,000
\$2,001-\$4,000
\$4,001-\$6,000
\$6,001-\$8,000
\$8,001-\$10,000
\$10,001-\$12,000
\$14,001-\$16,000

Number of Customers: The dataset specifies the number of customers falling within each net sales range.

\$0-\$2,000: 89 customers
\$2,001-\$4,000: 33 customers
\$4,001-\$6,000: 7 customers
\$6,001-\$8,000: 7 customers
\$8,001-\$10,000: 3 customers
\$10,001-\$12,000: 2 customers
\$14,001-\$16,000: 2 customers

This summary provides insight into the distribution of customers based on their net sales amounts. It indicates that the majority of customers fall within the lower net sales ranges, with the number of customers decreasing as the net sales range increases. Additionally, there are fewer customers in the higher net sales brackets, suggesting that a smaller proportion of customers contribute to higher sales amounts.

ATTRIBUTES FOR ABOVE VISUALIZATION

X-axis
Net-Sales

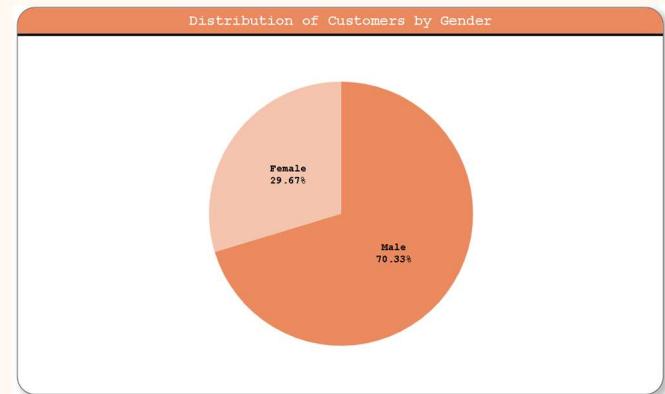
Y-axis
Number of Customers

Legend
Add data fields here

Small multiples
Add data fields here

Tooltips
Add data fields here

VISUALIZATION



The dataset provides information on the gender distribution of customers. Here's a summary:

Gender Distribution: The dataset contains two categories of gender:

Male: There are 64 male customers.

Female: There are 27 female customers.

This summary reveals that there are more male customers than female customers in the dataset.

ATTRIBUTES FOR ABOVE VISUALIZATION

Legend
Gender

Values
Number of Customers

Details
Add data fields here

Tooltips
Add data fields here

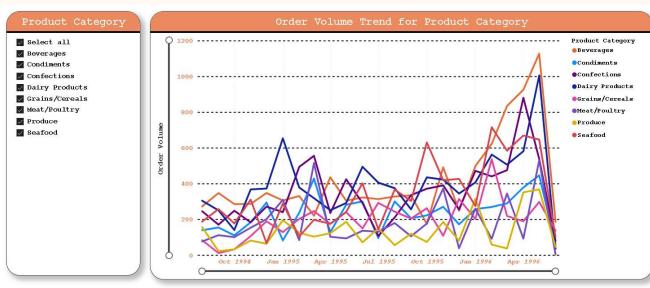
Drill through
Cross-report Off

Keep all filters On

POWER BI PROBLEM STATEMENT

4. HOW DOES ORDER VOLUME CHANGE OVER TIME? CAN WE CREATE A TIME SERIES CHART OR STACKED BAR CHART TO VISUALIZE IT?

VISUALIZATION



ATTRIBUTES FOR ABOVE VISUALIZATION

The screenshot shows the Power BI visualization settings pane. The X-axis section has "Date", "Year", "Quarter", "Month", and "Day" selected. The Y-axis section has "Order Volume" selected. The "Field" section has "Product Category" assigned to the Y-axis. The "Drill through" section has "Cross-report" set to "Off" and "Keep all filters" set to "On". The "Legend" section has "Product Category" assigned to the X-axis.

POWER BI PROBLEM STATEMENT

5. WHAT IS THE DISTRIBUTION OF ORDER VALUES? CAN WE CREATE A HISTOGRAM OR BOX PLOT TO DISPLAY IT?

VISUALIZATION



Based on the provided data, it seems to represent average sales figures for various orders identified by OrderID-FK. Here's a summary of the data:

Total number of records: 227

Average sales across all orders: Approximately 601.85

Minimum average sales: 12.5

Maximum average sales: 15810

Standard deviation: Approximately 1167.88

This summary provides an overview of the average sales figures across the dataset, highlighting the range and distribution of the data.

ATTRIBUTES FOR ABOVE VISUALIZATION

Axis

OrderID-FK ✓ X

Axis category I

Add data fields here

Axis category II

Add data fields here

Value

Average Sales ✓ X

Dots size

Add data fields here

Legend

POWER BI PROBLEM STATEMENT

6. CAN WE VISUALIZE THE AVERAGE ORDER PROCESSING TIME OR SHIPPING DURATION USING A BAR CHART OR BOX PLOT?

VISUALIZATION



The provided data presents the average order processing days for a list of orders, identified by their OrderID-PK. Here is a summary of the data:

The average order processing days range from 1 to 37.

The minimum average order processing days is 1, and the maximum is 37.

The data includes a total of 175 orders.

In summary, this dataset provides insights into the average time taken to process orders, with a wide range of processing durations observed.

ATTRIBUTES FOR ABOVE VISUALIZATION

Axis
OrderID-PK ▼ X

Axis category I
Add data fields here

Axis category II
Add data fields here

Value
Average Order Proces... ▼ X

Dots size
Add data fields here

Legend

VISUALIZATION



The given data appears to be a table showing the average number of shipping days for various orders, identified by their OrderID. Each OrderID is associated with an average shipping duration in days.

Here is a summary of the data:

The OrderID ranges from 10248 to 11077.

The average shipping days range from 0 to 41 days.

The average shipping days seem to vary widely, with some orders being shipped very quickly (0 days) and others taking significantly longer.

The most common average shipping durations appear to be around 20-25 days, with some outliers on either end of the spectrum.

There are some orders with missing or incomplete data, indicated by "0" in the average shipping days column.

This summary provides a general overview of the shipping durations for the orders listed in the dataset.

ATTRIBUTES FOR ABOVE VISUALIZATION

Axis
OrderID-PK ▼ X

Axis category I
Add data fields here

Axis category II
Add data fields here

Value
Average of ShippingD... ▼ X

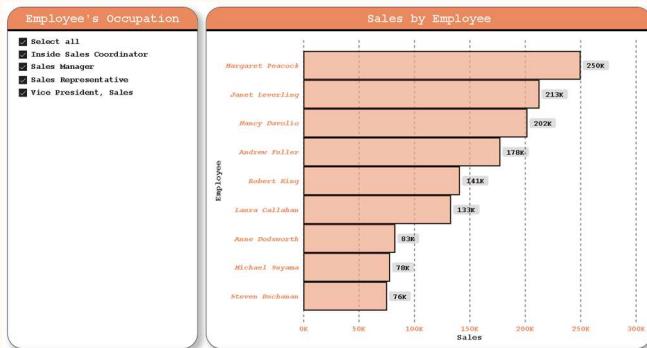
Dots size
Add data fields here

Legend

POWER BI PROBLEM STATEMENT

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?

VISUALIZATION



The provided data presents the sales figures associated with different employees. Here's a summary of the sales data:

Margaret Peacock generated sales totaling \$250,187.45.

Janet Leverling achieved sales amounting to \$213,051.30.

Nancy Davolio's sales amounted to \$202,143.71.

Andrew Fuller's sales totaled \$177,749.26.

Robert King's sales amounted to \$141,295.99.

Laura Callahan's sales amounted to \$133,301.03.

Anne Dodsworth achieved sales totaling \$82,964.

Michael Suyama's sales amounted to \$78,198.10.

Steven Buchanan's sales totaled \$75,567.75.

This summary provides an overview of the sales performance of each employee listed in the dataset.

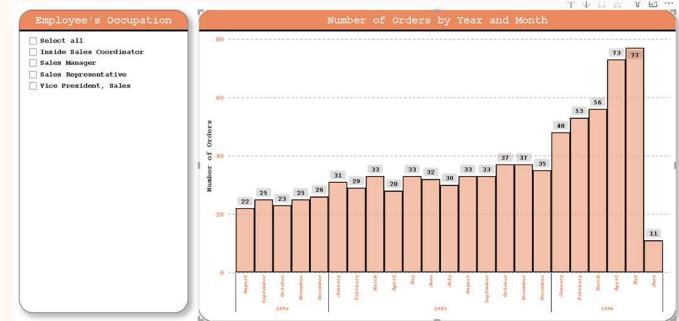
ATTRIBUTES FOR ABOVE VISUALIZATION

| | |
|-----------------|----------------------|
| Y-axis | Employee |
| X-axis | Sales |
| Legend | Add data fields here |
| Small multiples | Add data fields here |
| Tooltips | Add data fields here |

Field

Drill through
 Off
 On
 Keep all filters

VISUALIZATION



The provided data contains information on the number of orders received each month over several years. Here's a summary of the data:

In August 1994, there were 22 orders.

In September 1994, there were 25 orders.

In October 1994, there were 23 orders.

In November 1994, there were 26 orders.

In December 1994, there were 31 orders.

In January 1995, there were 29 orders.

In February 1995, there were 33 orders.

In March 1995, there were 32 orders.

In April 1995, there were 33 orders.

In May 1995, there were 30 orders.

In June 1995, there were 33 orders.

In July 1995, there were 30 orders.

In August 1995, there were 33 orders.

In September 1995, there were 33 orders.

In October 1995, there were 37 orders.

In November 1995, there were 37 orders.

In December 1995, there were 35 orders.

In January 1996, there were 48 orders.

In February 1996, there were 53 orders.

In March 1996, there were 56 orders.

In April 1996, there were 73 orders.

In May 1996, there were 77 orders.

In June 1996, there were 11 orders.

This summary provides an overview of the number of orders received each month from August 1994 to June 1996.

ATTRIBUTES FOR ABOVE VISUALIZATION

| | |
|-----------------|----------------------|
| X-axis | Date |
| Year | Year |
| Month | Month |
| Y-axis | Number of Orders |
| Legend | Add data fields here |
| Small multiples | Add data fields here |
| Tooltips | Add data fields here |

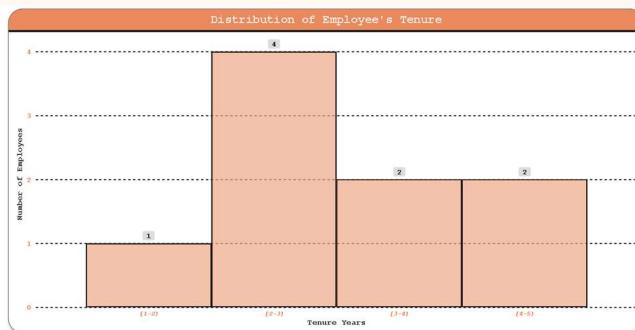
Field

Drill through
 Off
 On
 Keep all filters

POWER BI PROBLEM STATEMENT

8. WHAT IS THE DISTRIBUTION OF EMPLOYEE TENURE? CAN WE CREATE A HISTOGRAM OR BOX PLOT TO DISPLAY IT?

VISUALIZATION



The provided data presents the distribution of employees based on their tenure in years. Here's an interesting summary:

The data indicates that the majority of employees fall within the 2-3-year tenure range, with four employees in this category. This suggests that the company may have experienced a period of growth or hiring approximately two to three years ago, leading to a cohort of employees reaching this tenure milestone.

Additionally, there is a balanced distribution of employees across the other tenure ranges, ranging from 1-2 years to 4-5 years, with each range having between one to two employees. This suggests a relatively stable workforce with consistent retention over the observed tenure periods.

Overall, this distribution provides insights into the retention patterns and stability within the company's workforce, with the most notable concentration of employees occurring in the 2-3-year tenure range.

ATTRIBUTES FOR ABOVE VISUALIZATION

X-axis

Tenure Years

Y-axis

Number of Employees

Legend

Add data fields here

Small multiples

Add data fields here

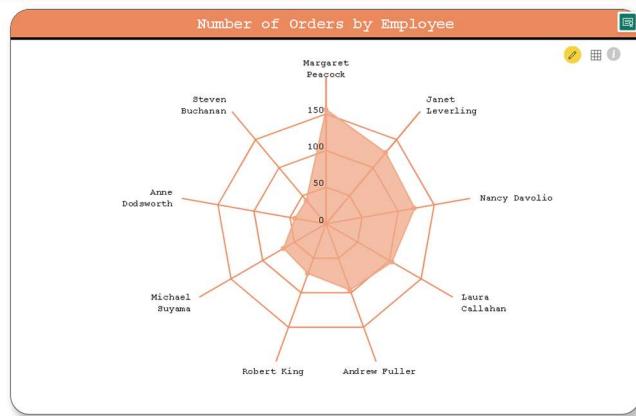
Tooltips

Add data fields here

POWER BI PROBLEM STATEMENT

9. CAN WE VISUALIZE EMPLOYEE PERFORMANCE RATINGS OR KPIS USING A RADAR CHART OR BULLET GRAPH?

VISUALIZATION



The dataset provides the count of orders handled by each employee, offering insights into their individual workload and efficiency. Here's an interesting summary:

Margaret Peacock stands out as the employee with the highest number of orders, handling a total of 156 orders. This suggests that Margaret is likely a highly experienced and efficient employee, capable of managing a significant workload.

Following closely behind Margaret is Janet Leverling, who processed 127 orders. Janet's performance indicates a strong work ethic and proficiency in managing customer orders effectively.

Nancy Davolio, Laura Callahan, and Andrew Fuller also demonstrate commendable performance, each handling over 90 orders. Their consistent high order counts suggest reliability and effectiveness in their roles.

On the other end of the spectrum, Steven Buchanan and Anne Dossoworth processed the fewest orders, with 42 and 43 orders, respectively. While their order counts are lower compared to others in the dataset, it's important to consider other factors such as the complexity of orders and individual roles within the company.

Overall, the distribution of order counts among employees provides valuable insights into their productivity levels and contributions to the company's operations.

ATTRIBUTES FOR ABOVE VISUALIZATION

Axis
FullName

Legend
Add data fields here

Value
Count of OrderID-PK

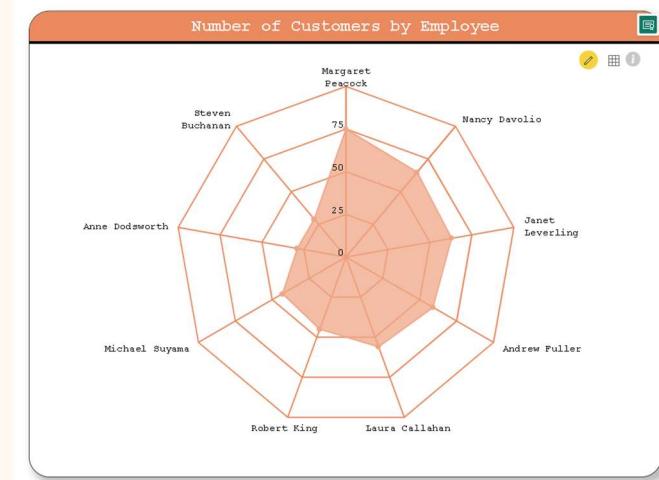
Tooltips
Add data fields here

Drill through

Cross-report Off

Keep all filters On

VISUALIZATION



The dataset provides insights into the number of customers each employee interacts with, reflecting their reach and influence within the customer base. Here's an interesting summary:

Margaret Peacock leads the pack with the highest number of customers, interacting with a total of 75 individuals. This suggests that Margaret likely has a broad network and effectively manages relationships with a large client base.

Following closely behind Margaret is Nancy Davolio, who engages with 65 customers. Nancy's ability to maintain relationships with a substantial number of clients indicates strong interpersonal skills and rapport-building capabilities.

Janet Leverling also demonstrates significant customer engagement, interacting with 63 individuals. Janet's performance highlights her effectiveness in fostering relationships and addressing the needs of a diverse clientele.

Andrew Fuller, Laura Callahan, and Robert King maintain relationships with over 50 customers each, showcasing their wide-reaching impact and ability to cater to various client preferences and demands.

On the other end of the spectrum, Steven Buchanan and Anne Dodsworth engage with the fewest number of customers, with 29 each. While their customer counts are lower compared to others in the dataset, it's important to consider factors such as the nature of their roles and the depth of relationships with individual clients.

Overall, the distribution of customer counts among employees provides valuable insights into their ability to nurture relationships and drive customer satisfaction, crucial aspects of maintaining a loyal and thriving customer base.

ATTRIBUTES FOR ABOVE VISUALIZATION

Axis

Employee

Legend

Add data fields here

Value

Number of Customers

Tooltips

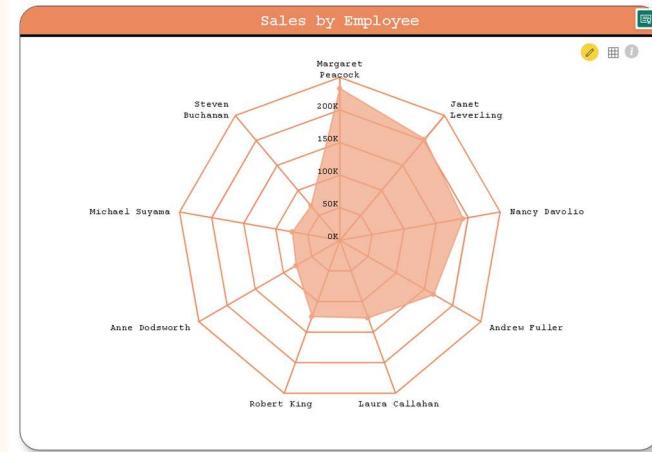
Add data fields here

Drill through

Cross-report Off

Keep all filters On

VISUALIZATION



The sales data for each employee offers intriguing insights into their individual performance and contribution to the company's revenue. Here's an interesting summary:

Margaret Peacock emerges as the top performer in terms of sales, generating an impressive total of \$232,890.87. This indicates her strong sales acumen, ability to close deals, and potentially effective management of high-value accounts.

Following closely behind Margaret is Janet Leverling, who achieves sales totaling \$202,812.83. Janet's performance underscores her effectiveness in driving revenue and securing significant business opportunities.

Nancy Davolio also demonstrates notable sales performance, with total sales amounting to \$192,107.59. Nancy's contribution highlights her capability to generate revenue and meet or exceed sales targets consistently.

Andrew Fuller maintains a strong position with total sales of \$166,537.76. Andrew's performance showcases his sales prowess and effectiveness in driving business growth.

Laura Callahan and Robert King achieve respectable sales figures of \$126,862.27 and \$124,568.22, respectively. Their contributions indicate their ability to generate revenue and play significant roles in achieving the company's sales objectives.

Anne Dodsworth, Michael Suyama, and Steven Buchanan, while achieving lower sales figures compared to their colleagues, still contribute substantially to the company's revenue. Their efforts are crucial in ensuring consistent sales performance and overall business success.

Overall, the sales data provides valuable insights into the individual contributions of each employee, highlighting their strengths, effectiveness in generating revenue, and impact on the company's financial performance.

ATTRIBUTES FOR ABOVE VISUALIZATION

Axis

Employee

Legend

Add data fields here

Value

Sales

Tooltips

Add data fields here

Drill through

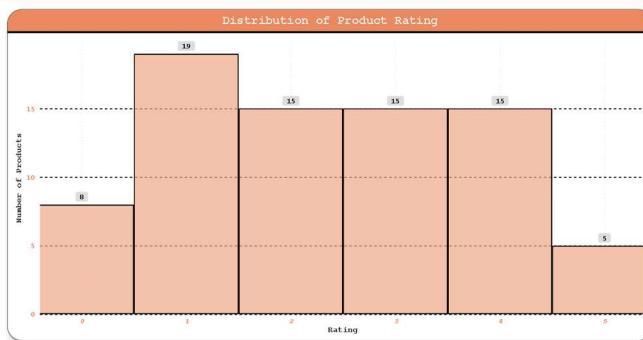
Cross-report Off

Keep all filters On

POWER BI PROBLEM STATEMENT

10. WHAT IS THE DISTRIBUTION OF PRODUCT RATINGS OR REVIEWS? CAN WE CREATE A HISTOGRAM OR STACKED BAR CHART TO VISUALIZE IT?

VISUALIZATION



The product rating distribution reveals some interesting patterns:

Concentration in Mid-Range Ratings: The majority of products fall within the mid-range ratings of 1 to 4, with each rating category having a similar count of 15 products. This suggests that a significant portion of products are perceived as average or slightly above average by customers.

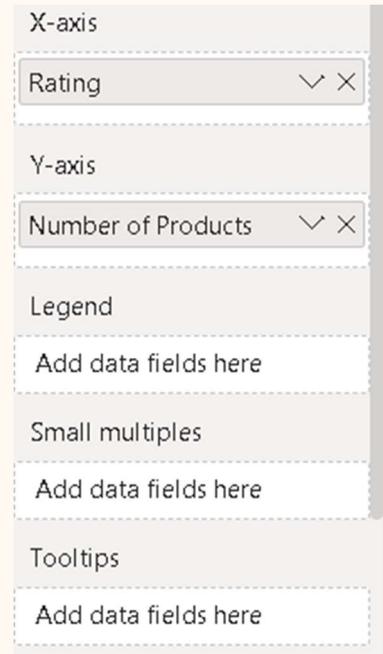
Low-Rated Products: Eight products have received a rating of 0, indicating dissatisfaction or poor performance according to customers. Understanding the reasons behind these low ratings could be crucial for identifying areas of improvement in product quality or customer satisfaction.

Limited High Ratings: There are fewer products with higher ratings, with only 5 products receiving a rating of 5. This could suggest that achieving exceptional customer satisfaction or product excellence is relatively rare, possibly indicating areas where the company excels or where customer expectations are particularly high.

Balanced Distribution: The distribution of ratings across different categories appears relatively balanced, indicating that customers are providing diverse feedback on product experiences. This balanced distribution could be beneficial for the company in understanding customer preferences and making informed decisions about product development and marketing strategies.

Overall, analyzing the distribution of product ratings provides valuable insights into customer perceptions and satisfaction levels, highlighting areas of strength and opportunities for improvement in product offerings and customer experience.

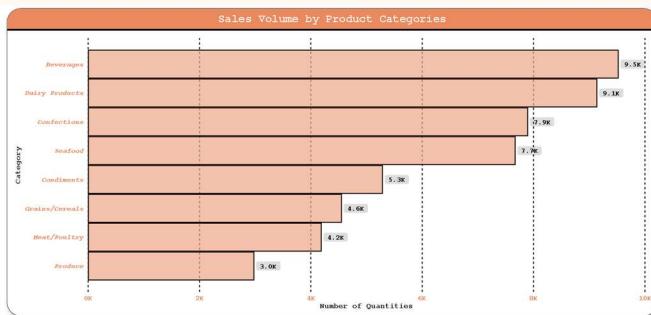
ATTRIBUTES FOR ABOVE VISUALIZATION



POWER BI PROBLEM STATEMENT

11. HOW DOES THE SALES VOLUME VARY ACROSS DIFFERENT PRODUCT CATEGORIES? CAN WE CREATE A BAR CHART OR TREE MAP TO DISPLAY IT?

VISUALIZATION



The distribution of quantities across different product categories unveils some intriguing insights:

Popular Categories: Beverages and Dairy Products emerge as the most popular categories, with 9532 and 9149 quantities sold, respectively. This suggests that these categories are highly sought after by customers, indicating potentially high demand or wide appeal.

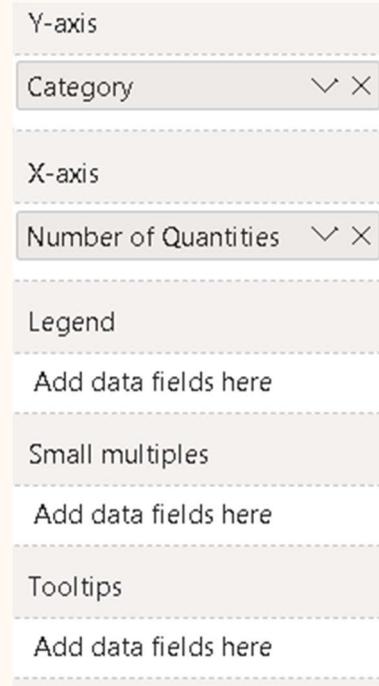
Sweet Temptations: Confections, despite being a niche category compared to Beverages and Dairy Products, still garners significant attention with 7906 quantities sold. This highlights the enduring appeal of sweets and treats among consumers, indicating a market segment that could be further explored or capitalized upon.

Underrepresented Categories: Grains/Cereals and Meat/Poultry exhibit relatively lower quantities sold compared to other categories. This could be due to various factors such as dietary trends, cultural preferences, or marketing strategies. Understanding the reasons behind these lower sales figures could help in devising targeted marketing campaigns or product improvements to boost sales in these categories.

Seasonal Variations: The quantity sold in the Produce category, with 2990 units, might indicate a seasonal pattern or specific consumer preferences. Analyzing sales trends over time could provide insights into the seasonality of certain product categories and help in optimizing inventory management and marketing efforts accordingly.

Diverse Portfolio: The diverse range of product categories, including Seafood, Condiments, and Produce, reflects a well-rounded product portfolio catering to various tastes and dietary preferences. This diversity could be advantageous in attracting a wide customer base and mitigating risks associated with market fluctuations or changing consumer preferences.

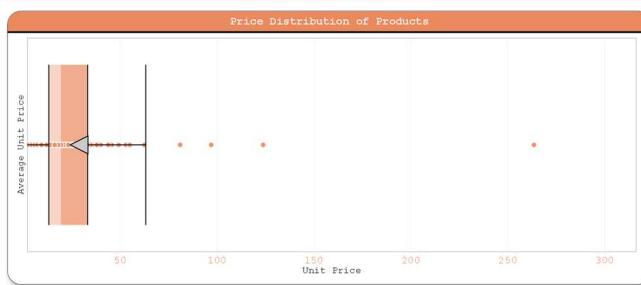
ATTRIBUTES FOR ABOVE VISUALIZATION



POWER BI PROBLEM STATEMENT

12. CAN WE VISUALIZE THE PRICING DISTRIBUTION OF PRODUCTS USING A BOX PLOT OR VIOLIN PLOT?

VISUALIZATION



This list provides fascinating insights into the pricing diversity across various products:

Luxury Indulgence: The "Côte de Blaye" stands out with an average unit price of \$263.5, indicating it as a premium product, possibly a fine wine or gourmet delicacy. This suggests that it targets a niche market segment willing to pay a premium for quality.

Affordable Treats: On the other end of the spectrum, products like "Guaraná Fantástica" and "Tourtière" are priced relatively low at \$4.5 and \$7.45 respectively. These items may serve as affordable treats or everyday essentials, appealing to budget-conscious consumers.

Unexpected Highs: While some items are expected to have high prices due to their premium nature, others surprise with unexpectedly high prices. For instance, "Thüringer Rostbratwurst" has an average unit price of \$123.79, suggesting it might be a specialty sausage or a gourmet product with unique attributes.

Cultural Influence: The diverse range of products, including "Ipoh Coffee," "Tarte au sucre," and "Louisiana Fiery Hot Pepper Sauce," reflects cultural influences from various regions worldwide. These products not only offer unique flavors but also provide insights into culinary traditions and preferences.

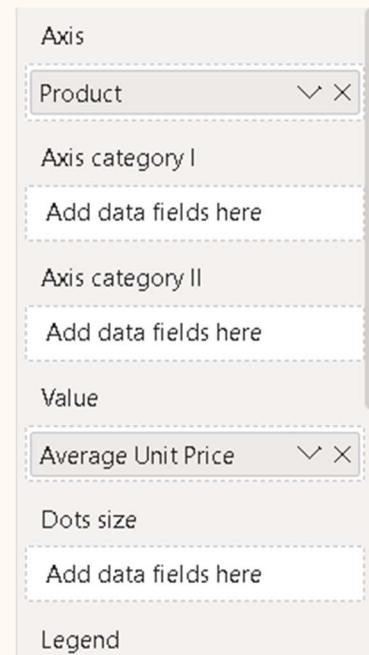
Gourmet Delights: Items like "Mishi Kobe Niku" and "Raclette Courdavault" boast relatively high average prices, indicating they might be gourmet meats or cheeses catering to discerning palates. These products likely target customers seeking premium dining experiences or gourmet ingredients for special occasions.

Mid-Range Options: Many products fall within the mid-range price bracket, such as "Chai" (\$18), "Gula Malacca" (\$19.45), and "Queso Cabrales" (\$21). These items offer a balance between quality and affordability, appealing to a broad range of consumers looking for value for money.

Sweet Luxuries: "Sir Rodney's Marmalade" commands a notably high average price of \$81, indicating it might be a luxury gourmet spread with unique flavors or premium ingredients, appealing to connoisseurs of fine foods.

This diverse pricing landscape reflects the intricate interplay between product quality, uniqueness, cultural significance, and target market preferences in the global marketplace.

ATTRIBUTES FOR ABOVE VISUALIZATION



POWER BI PROBLEM STATEMENT

13. WHAT IS THE DISTRIBUTION OF SUPPLIER RATINGS OR PERFORMANCE METRICS? CAN WE CREATE A BAR CHART OR RADAR CHART TO VISUALIZE IT?

VISUALIZATION



This dataset reveals intriguing patterns in the distribution of order counts among employees:

Top Performers: Margaret Peacock, Janet Leverling, and Nancy Davolio emerge as the top three performers in terms of order counts, with 156, 127, and 123 orders respectively. Their consistently high order counts suggest effective sales skills, strong client relationships, or possibly larger client portfolios.

Middle Tier: Laura Callahan and Andrew Fuller fall in the middle tier, with 104 and 96 orders respectively. While they may not have as high a volume of orders as the top performers, their consistent performance indicates reliability and competence in managing client accounts.

Rising Stars: Robert King, Michael Suyama, Anne Dodsworth, and Steven Buchanan occupy the lower end of the spectrum with 72, 67, 43, and 42 orders respectively. However, their presence in this list suggests potential for growth or improvement, as they may be newer employees or transitioning into new roles.

Diversity in Performance: The varying order counts among employees underscore the diversity in sales performance within the company. This could stem from differences in experience, expertise, client base, or the nature of products they handle.

Opportunities for Improvement: While some employees have high order counts, others have room for improvement. This data could be used to identify areas where additional training, support, or resources may be beneficial to enhance sales performance across the board.

Overall, this summary highlights the dynamic nature of sales teams, with a mix of top performers, steady contributors, and potential for growth among employees.

ATTRIBUTES FOR ABOVE VISUALIZATION

Axis
FullName

Legend
Add data fields here

Value
Count of OrderID-PK

Tooltips
Add data fields here

Drill through

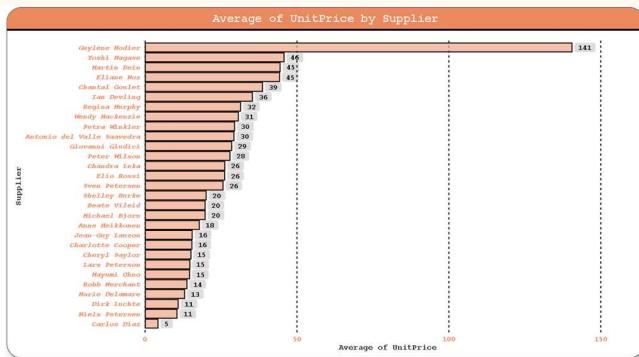
Cross-report Off

Keep all filters On

POWER BI PROBLEM STATEMENT

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?

VISUALIZATION



This dataset provides intriguing insights into the average unit prices offered by different suppliers:

Price Disparity: The average unit prices vary significantly across suppliers, ranging from as high as \$140.75 (Guylène Nodier) to as low as \$4.5 (Carlos Diaz). This wide range suggests diversity in product offerings, quality, and possibly geographic location or market segment targeted by each supplier.

Quality vs. Price: Higher average unit prices, such as those offered by Guylène Nodier or Yoshi Nagase, may indicate suppliers that focus on premium or specialized products. On the other hand, lower average unit prices, like those from Carlos Diaz, could suggest suppliers of more budget-friendly or basic items.

Consistency: Some suppliers, such as Martin Bein and Eliane Noz, offer average unit prices that are relatively close to each other (\$44.678 and \$44.5 respectively). This consistency could imply stable pricing strategies or a similar product range within their portfolios.

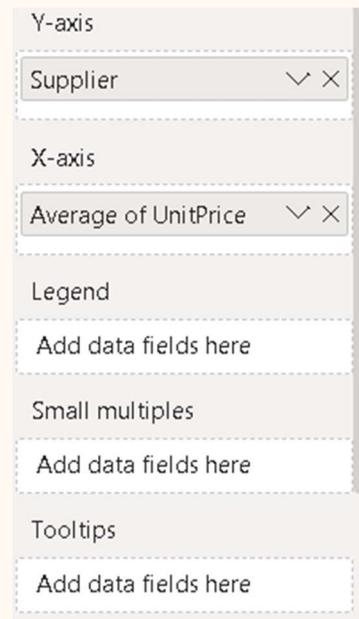
Competitive Advantage: Suppliers with lower average unit prices, such as Carlos Diaz, may have a competitive advantage in price-sensitive markets, while those with higher prices might target niche markets willing to pay a premium for quality or uniqueness.

Supplier Relationships: Analyzing average unit prices can help businesses evaluate their relationships with suppliers, ensuring they receive competitive pricing without sacrificing quality.

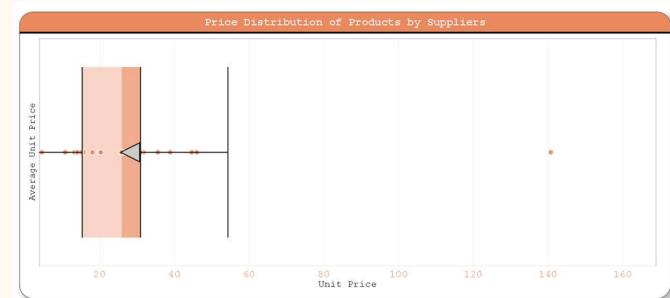
Market Trends: Changes in average unit prices over time could indicate shifts in market demand, supply chain dynamics, or competitive pressures within the industry.

In summary, understanding the range and distribution of average unit prices among suppliers is crucial for strategic sourcing decisions, supplier relationship management, and maintaining competitiveness in the market.

ATTRIBUTES FOR ABOVE VISUALIZATION



VISUALIZATION



This dataset provides fascinating insights into the average unit prices associated with different contacts or individuals:

Diverse Price Range: The average unit prices span a wide range, from as low as \$4.5 (Carlos Diaz) to as high as \$140.75 (Guylène Nodier), showcasing the diversity in pricing strategies and product offerings among contacts.

Market Segmentation: Contacts like Guylène Nodier, with significantly higher average unit prices, might cater to luxury or specialty markets where customers are willing to pay premium prices for exclusive products. Conversely, those with lower prices, such as Carlos Diaz, may target budget-conscious consumers or offer more basic goods.

Consistency vs. Variability: While some contacts maintain relatively consistent average unit prices (e.g., Anne Heikkinen with \$18.08 and Chandra Leka with \$26.48), others exhibit more variability, suggesting different pricing structures, product categories, or market segments.

Supplier Relationships: Understanding the average unit prices associated with each contact can help businesses assess their relationships with suppliers or partners, ensuring they align with pricing objectives and market positioning.

Competitive Positioning: Contacts with competitive average unit prices may have an advantage in attracting customers seeking value for money, while those with higher prices may focus on differentiation through quality, innovation, or brand reputation.

Strategic Sourcing: Analyzing average unit prices across contacts enables businesses to make informed decisions regarding supplier selection, negotiation strategies, and cost management initiatives.

In summary, this dataset underscores the importance of considering average unit prices in assessing supplier performance, market positioning, and overall business strategy. It highlights the diverse landscape of pricing strategies adopted by different contacts and the implications for competitive advantage and customer value proposition.

ATTRIBUTES FOR ABOVE VISUALIZATION

The screenshot shows a configuration interface for a visualization. On the left, there's a sidebar with sections for 'Axis', 'Value', and 'Legend'. Under 'Axis', 'ContactName' is selected. Under 'Value', 'Average Unit Price' is selected. Both dropdowns have a 'X' button to remove the selection. Below each dropdown is a placeholder 'Add data fields here'. The 'Legend' section is currently empty.

- Axis**
 - ContactName ▼ X
 - Axis category I
Add data fields here
 - Axis category II
Add data fields here
- Value**
 - Average Unit Price ▼ X
 - Dots size
Add data fields here
- Legend**

POWER BI PROBLEM STATEMENT

15. CAN WE VISUALIZE THE GEOGRAPHICAL DISTRIBUTION OF SUPPLIERS USING A MAP OR BUBBLE CHART?

VISUALIZATION



This dataset provides an intriguing overview of supplier distribution across different countries and cities:

Global Reach: Suppliers are located across various countries, indicating a global presence and distribution network. This global reach suggests diversified sourcing strategies by businesses to mitigate risks, leverage market opportunities, and access diverse resources.

Urban Concentration: Some cities, such as Paris, Tokyo, and London, host multiple suppliers, reflecting their status as major economic and commercial hubs. These cities likely offer favorable business environments, infrastructure, and access to markets, making them attractive locations for suppliers.

Regional Representation: The dataset includes suppliers from different regions worldwide, including Europe (France, Germany, Italy, Spain, Sweden, UK), North America (Canada, USA), Asia (Japan, Singapore), Australia, and Brazil. This regional representation highlights the geographical diversity of suppliers and potential variations in business practices, regulations, and market dynamics.

Market Dynamics: The presence of suppliers in key cities and countries reflects market demand, industry clusters, and economic factors driving business activities. It suggests opportunities for collaboration, competition, and innovation within and across regions.

Logistical Considerations: Understanding supplier locations facilitates logistical planning, including transportation, inventory management, and supply chain optimization. Proximity to markets, ports, and transportation networks can influence cost efficiency and responsiveness to customer demands.

Cultural and Regulatory Context: Operating across different countries entails navigating diverse cultural norms, business practices, and regulatory frameworks. Suppliers must adapt their operations to local contexts, comply with regulations, and build relationships with stakeholders to succeed in global markets.

Strategic Insights: Analyzing supplier distribution provides strategic insights into market penetration, expansion opportunities, and risk management. Businesses can leverage this information to optimize supplier relationships, identify emerging markets, and enhance competitive advantage.

In summary, this dataset offers valuable insights into the global footprint of suppliers, highlighting geographical diversity, market dynamics, and logistical considerations shaping supply chain operations on a global scale.

ATTRIBUTES FOR ABOVE VISUALIZATION

Location

Country

City

Legend

Add data fields here

Latitude

Add data fields here

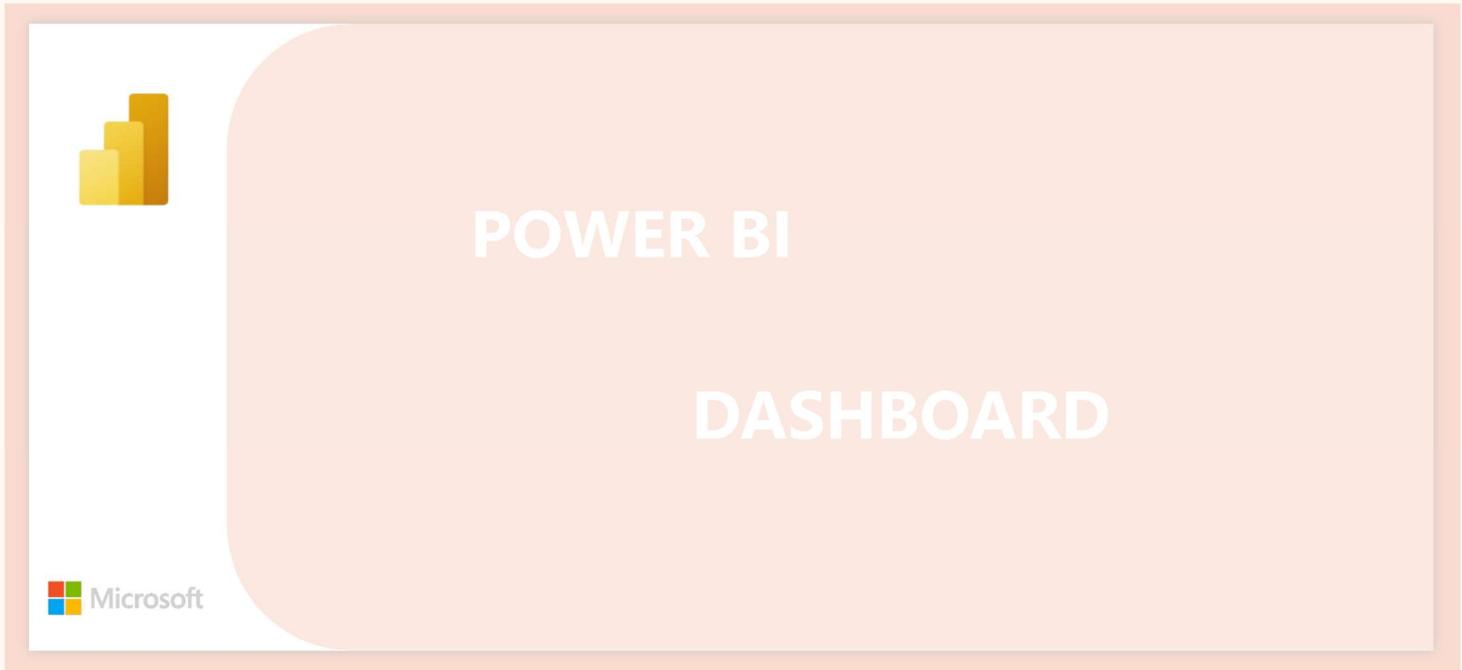
Longitude

Add data fields here

Bubble size

Number of Suppliers

POWER BI DASHBOARD



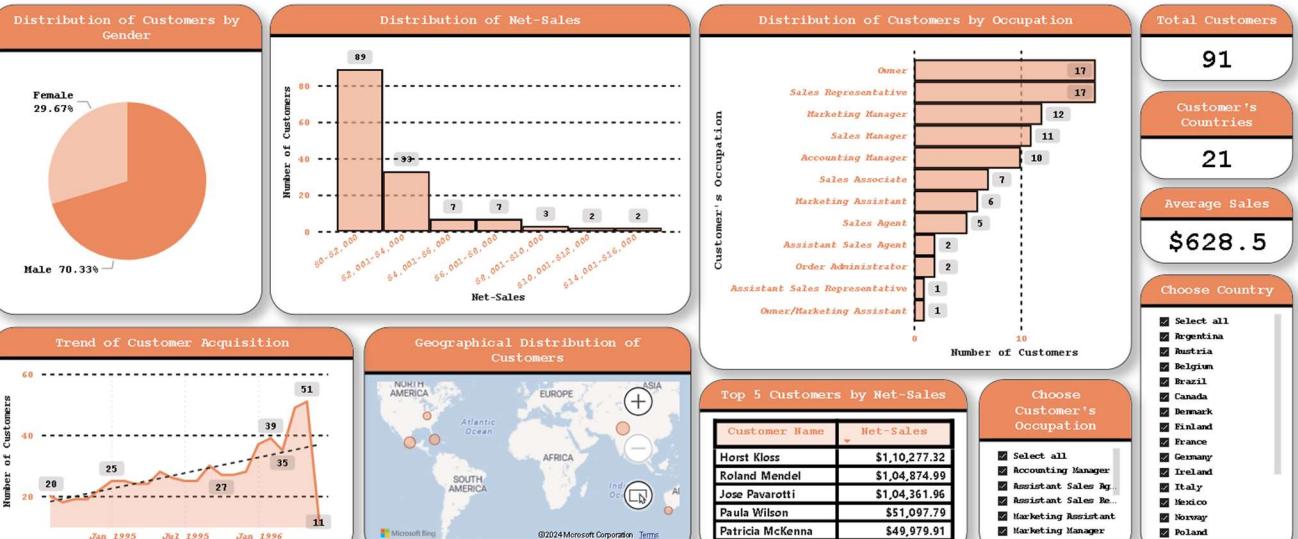
DASHBOARD NAVIGATOR

NORTH WIND TRADERS COMPANY

- CUSTOMER ANALYSIS
- EMPLOYEE ANALYSIS
- PRODUCT ANALYSIS
- ORDER ANALYSIS
- SUPPLIER ANALYSIS
- SHIPPER ANALYSIS



CUSTOMER ANALYSIS



EMPLOYEE ANALYSIS

PRODUCT ANALYSIS

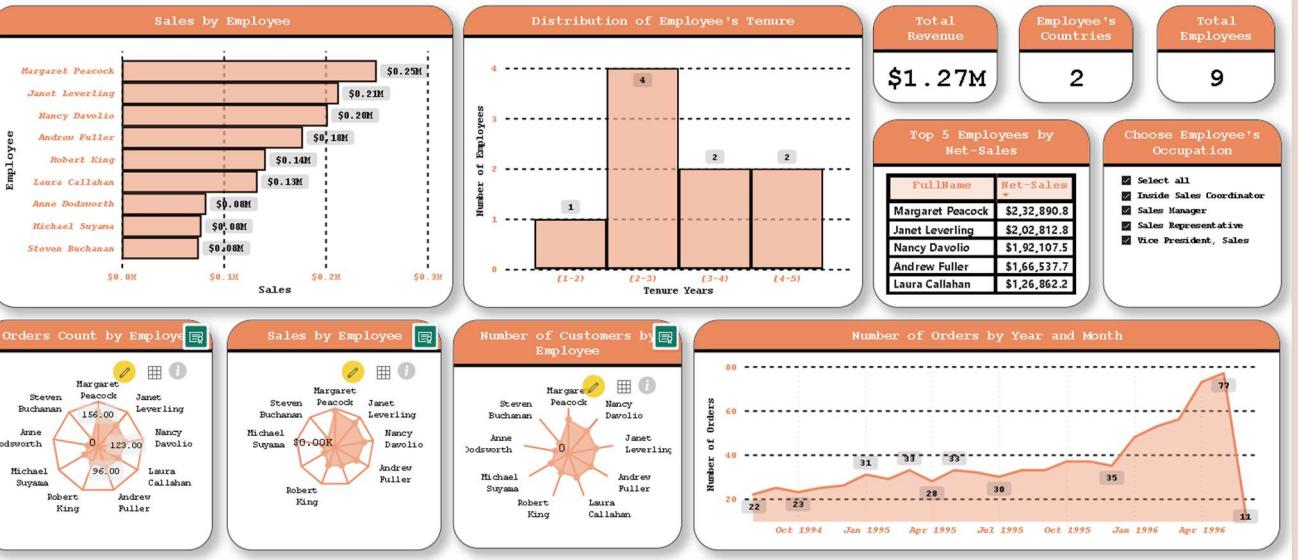
ORDER ANALYSIS

SUPPLIER ANALYSIS

SHIPPER ANALYSIS

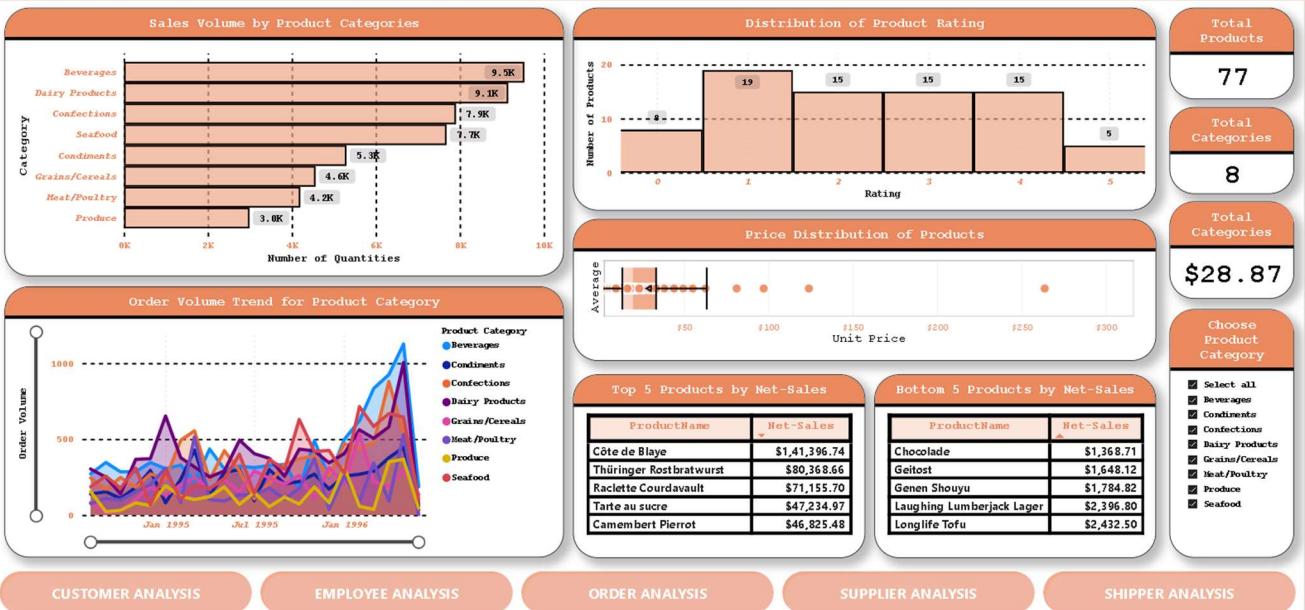


EMPLOYEE ANALYSIS

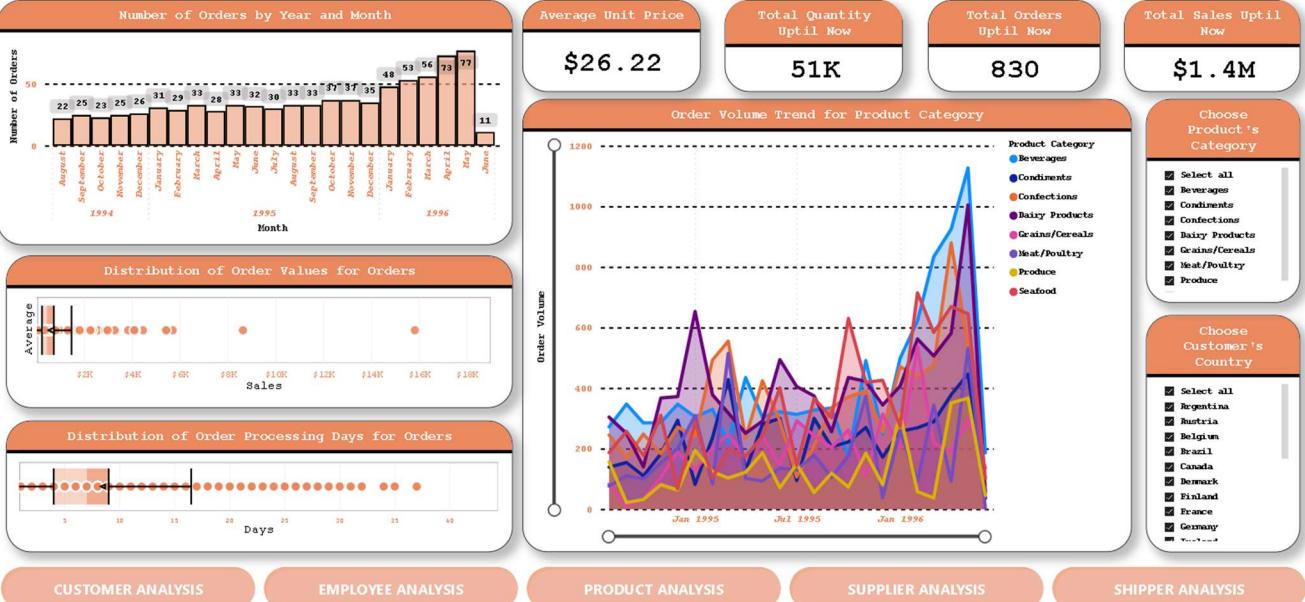




PRODUCT ANALYSIS

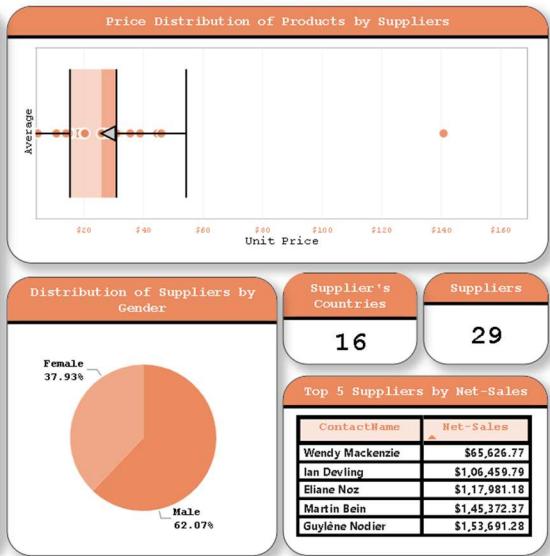
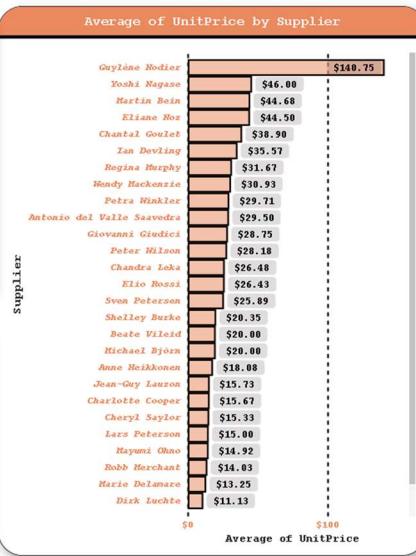
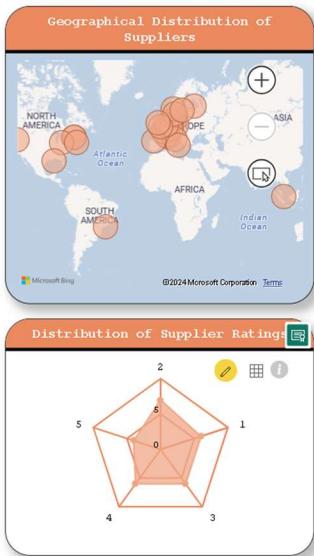


ORDER ANALYSIS





SUPPLIER ANALYSIS



CUSTOMER ANALYSIS

EMPLOYEE ANALYSIS

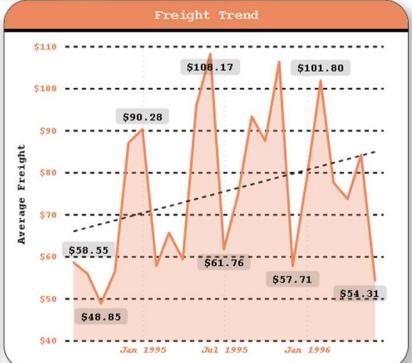
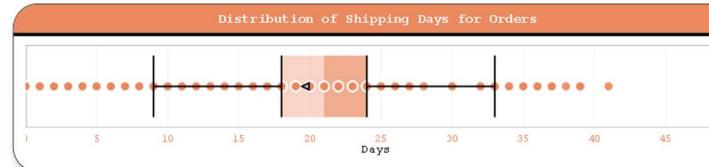
PRODUCT ANALYSIS

ORDER ANALYSIS

SHIPPER ANALYSIS



SHPPIER ANALYSIS



CUSTOMER ANALYSIS

EMPLOYEE ANALYSIS

PRODUCT ANALYSIS

ORDER ANALYSIS

SUPPLIER ANALYSIS

CONCLUSION

This comprehensive data analysis project, leveraging both Power BI and Excel, has unveiled valuable insights into the operations of Northwind Traders. By delving into sales patterns, customer behavior, inventory trends, and employee performance, we have gained a deeper understanding of the company's strengths, weaknesses, and opportunities.

Key Findings:

Sales Analysis: Identify top-selling products, profitable customer segments, and regional strengths/weaknesses in sales performance.

Customer Behavior: Understand customer demographics, buying patterns, and segmentation opportunities to drive targeted marketing strategies.

Inventory Management: Optimize stock levels, identify slow-moving items, and predict demand fluctuations to reduce carrying costs and improve efficiency.

Employee Performance: Evaluate individual and team performance to identify top performers, areas for improvement, and training needs.

IMPACT AND RECOMMENDATIONS:

These insights provide a foundation for data-driven decision-making across various departments at Northwind Traders. By translating findings into actionable recommendations, we can:

Increase revenue: Target high-value customer segments with personalized offerings and optimize product mix based on sales trends.

Improve customer satisfaction: Enhance customer experience through better understanding of preferences and needs.

Reduce costs: Optimize inventory management and identify areas for operational efficiency gains.

Boost employee productivity: Implement targeted training programs and recognize top performers to drive employee engagement.