

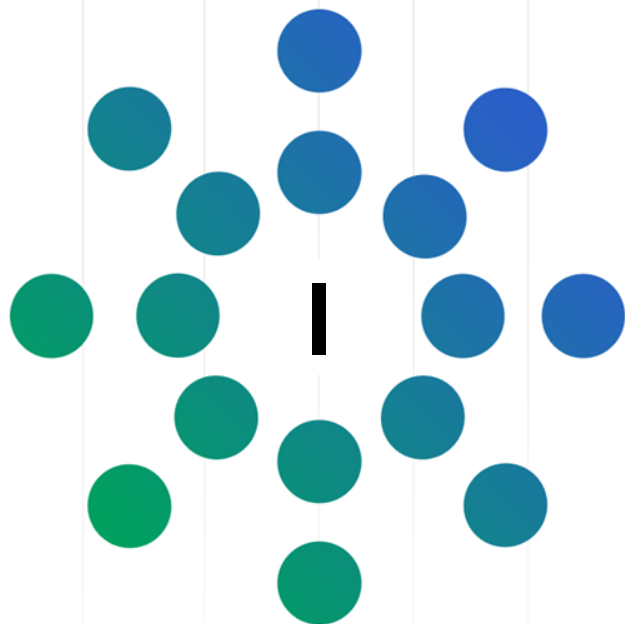
Mini Project Data Engineering

Study Case: Northwind Traders Database

by Adrian Maulana Muhammad

Table of Contents

- I. Business Understanding
- II. Analysis Objectives
- III. Data Processing Flowcharts
- IV. Data Processing Queries
- V. Data Analysis
- VI. Summaries & Recommendations
- VII. References



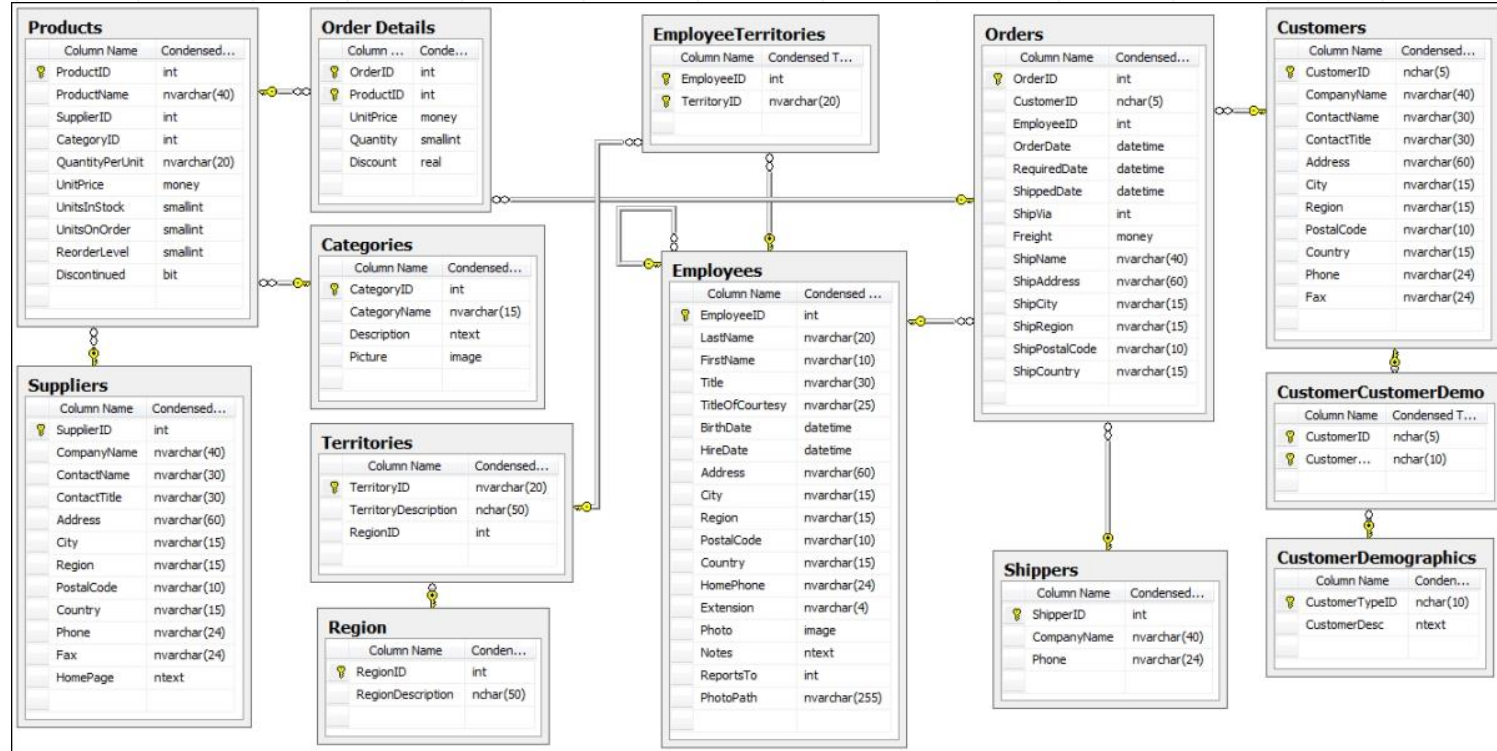
Business Understanding

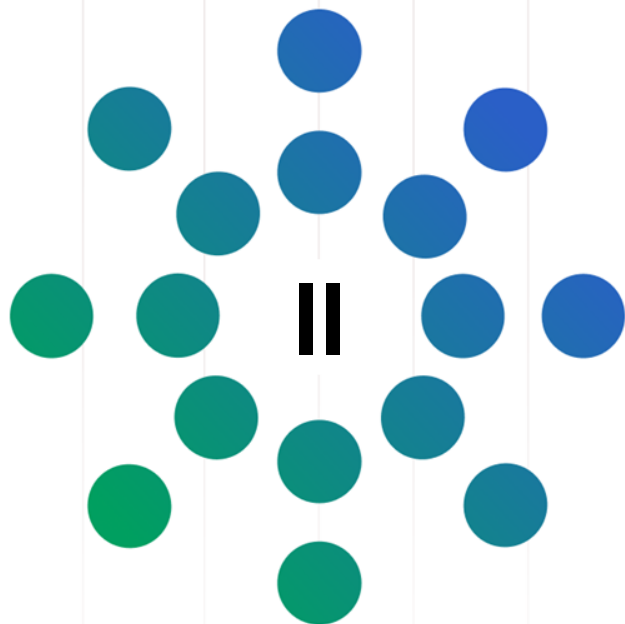
Northwind Traders

Northwind Traders is a fictional company that operates as a wholesale supplier of food and non-food products. The database represents a typical business operating in the retail industry, which includes a wide range of products, customers, orders, and employees.



Northwind Traders Database Schema





Analysis Objectives

Analysis Scope & Goals

Scope	Goals
Product Analysis	Category of Items Sold The Most
	The Best-Selling Products
Customer Analysis	Customer Spending by Country
	The Most Frequent Customers by Number of Orders
Shipper Analysis	Average Delivery Time per Shipper
	The Busiest Shippers

Product Analysis

- Analysis of Category of Items Sold The Most
 - a) Use join to combine the data from the "Products", "OrderDetails", and "Categories" tables.
 - b) Use aggregation to group the joined data by category and calculate the sum of quantity sold and total revenue for each category.
 - c) Use filtering to display only categories that have sold units above the average.
- Analysis of The Best-Selling Products
 - a) Join the Orders and Order Details tables to retrieve order details information.
 - b) Filter the order details information to retrieve only the needed information.
 - c) Group the data by product ID and aggregate the data to find the total quantity of each product sold.
 - d) Sort the data in descending order to find the best-selling product.
 - e) Join the Products table to retrieve the product name.
 - f) Select the product name and total quantity to display the best-selling product information.

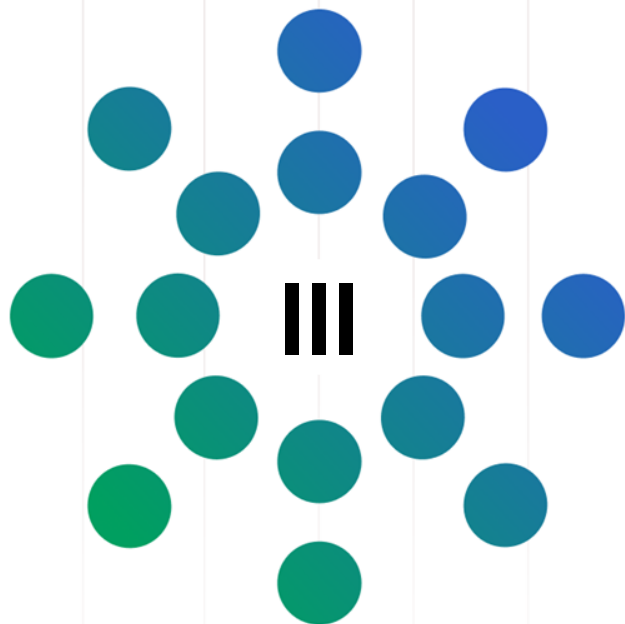
Customer Analysis

- Analysis of Customer Spending by Country
 - a) Use join to combine the data from the "Customers" and "Orders Details" tables.
 - b) Use aggregation to group the joined data by country and calculate the sum of total revenue for each country.
 - c) Use filtering to display only countries where customer spending is above the average.
- Analysis of The Most Frequent Customers by Number of Orders
 - a) Use CTE to create a subquery that counts the number of orders per customer.
 - b) Use filtering to display only customers with more than average orders.
 - c) Use sorting to order the data by the number of orders per customer and display the top 10 most frequent customers.

Shipper Analysis

- Analysis of Average Delivery Time per Shipper
 - a) Use join to combine the data from the "Orders" and "Shippers" tables.
 - b) Use condition to calculate the difference between the "ShippedDate" and "OrderDate" columns.
 - c) Use aggregation to group the joined data by shipper and calculate the average delivery time for each shipper.

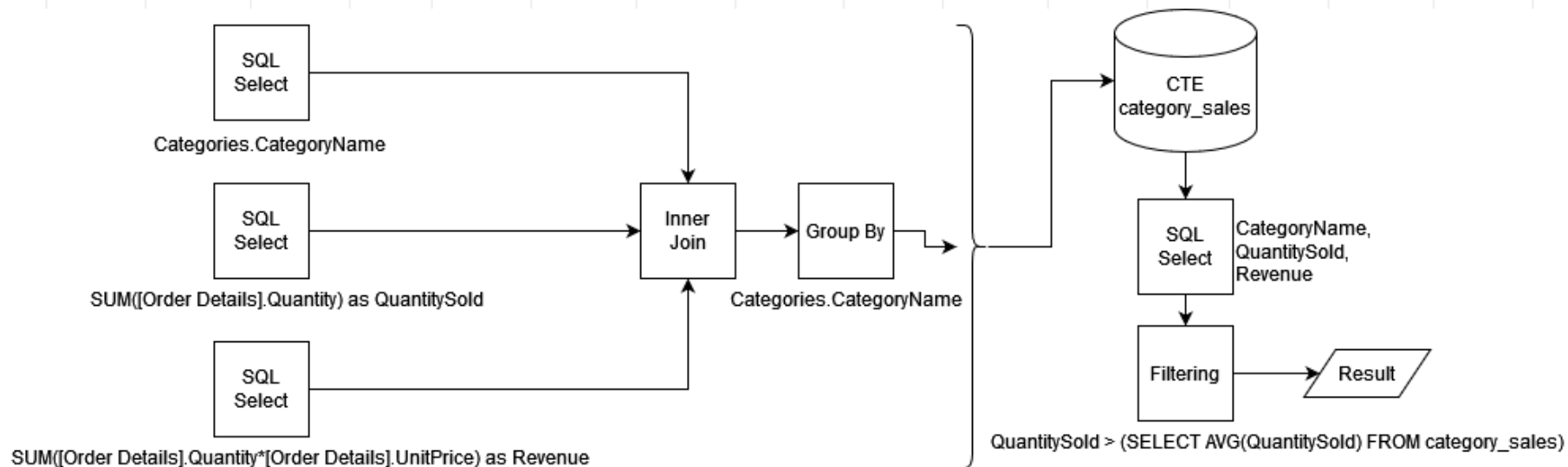
- Analysis of The Busiest Shippers
 - a) Use join to combine the data from the "Orders" and "Shippers" tables.
 - b) Use aggregation to count the number of orders shipped by each shipper.
 - c) Use sorting to order the data by the number of shipped orders and display the top 10 busiest shippers.



Data Processing Flowcharts

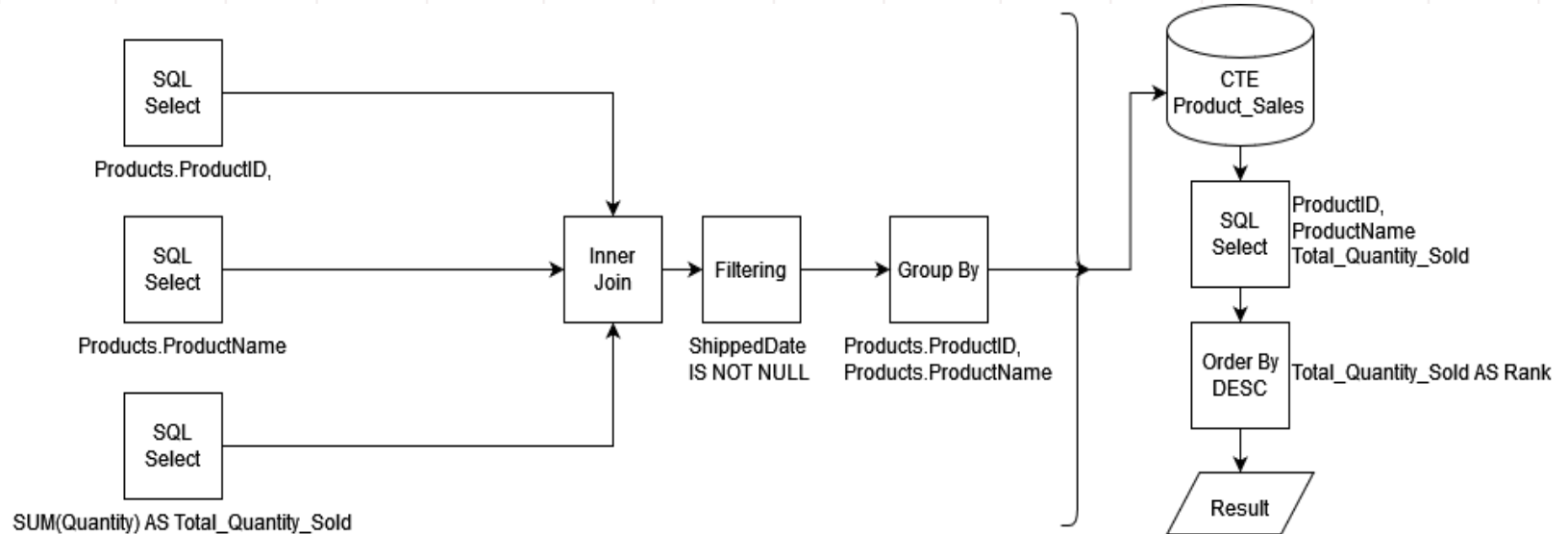
Product Analysis Flowcharts

Flowchart of Analysis of Category of Items Sold The Most



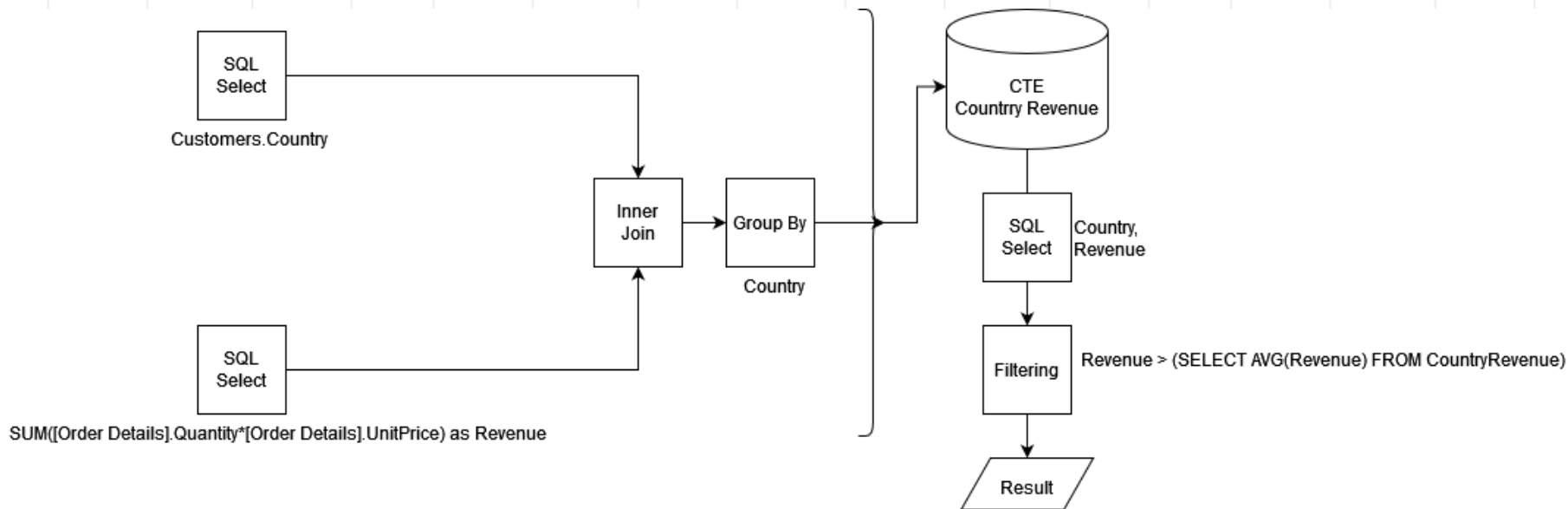
Product Analysis Flowcharts

Flowchart of Analysis of The Best-Selling Products



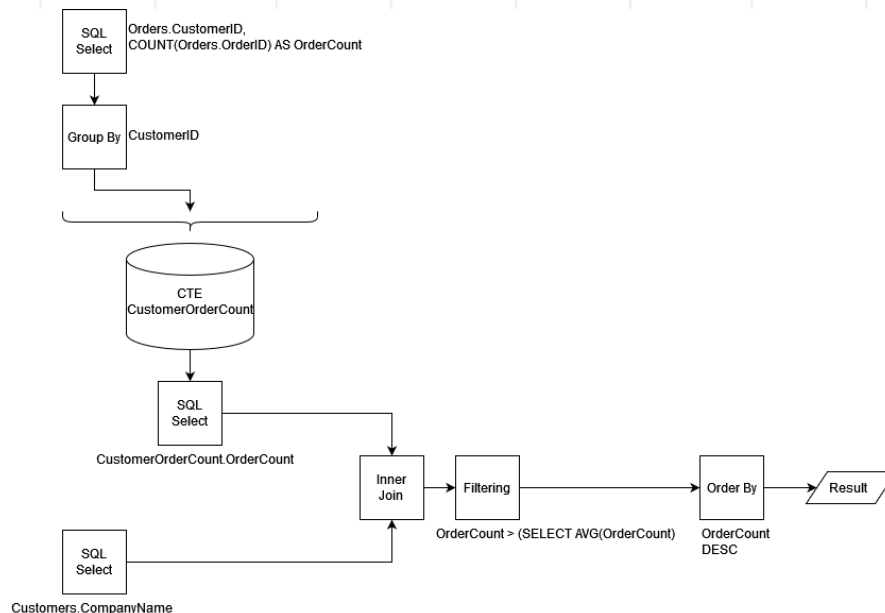
Customer Analysis Flowcharts

Flowchart of Analysis of Customer Spending by Country



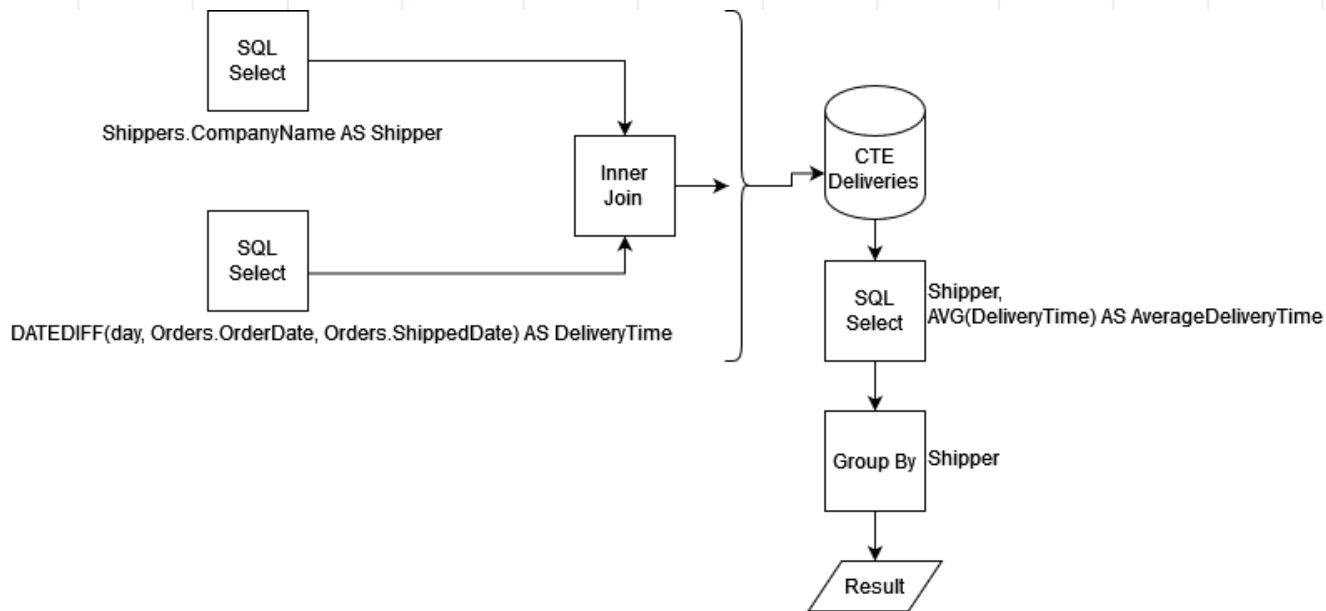
Customer Analysis Flowcharts

Flowchart of Analysis of The Most Frequent Customers by Number of Orders



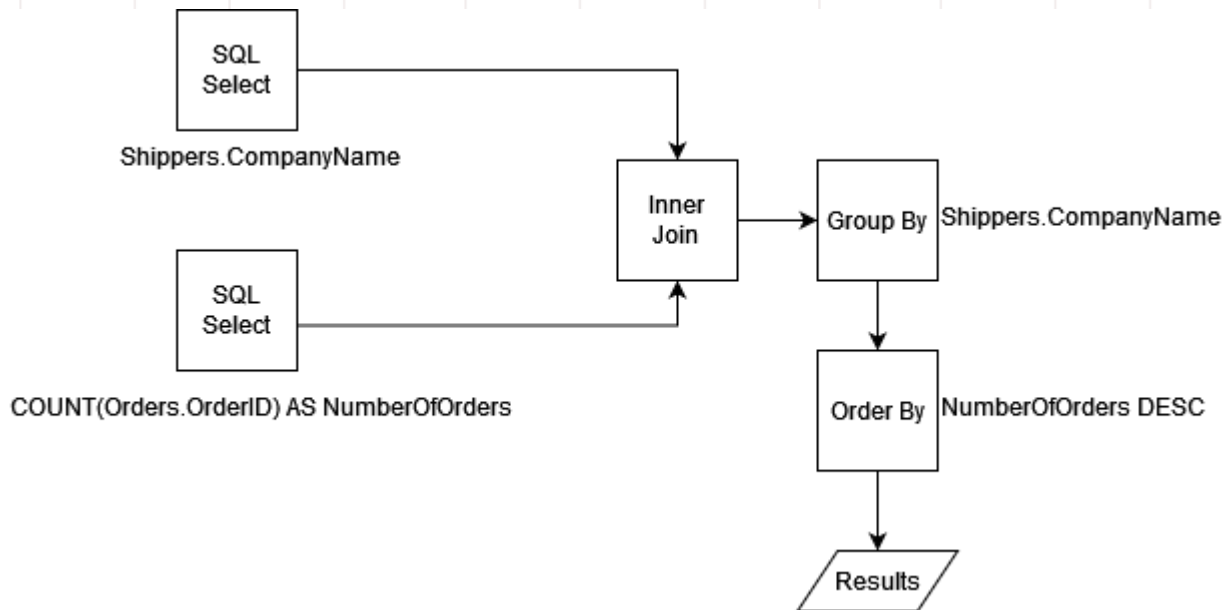
Shipper Analysis Flowcharts

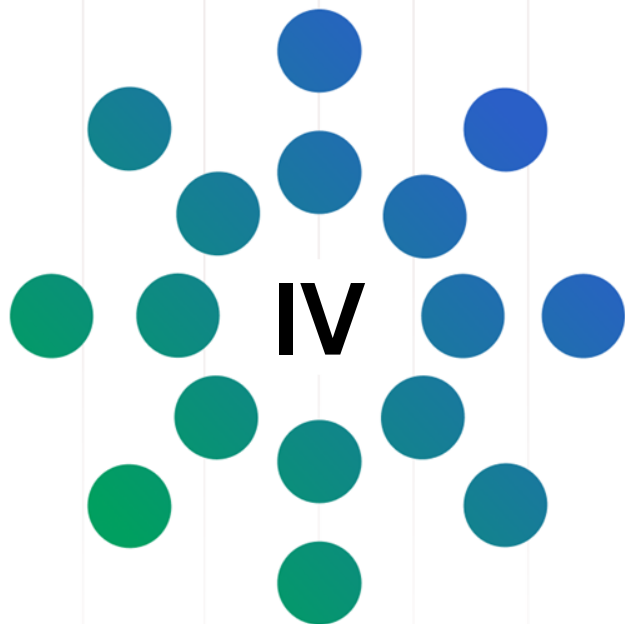
Flowchart of Analysis of Average Delivery Time per Shipper



Shipper Analysis Flowcharts

Flowchart of Analysis of The Busiest Shippers





Data Processing Queries

Product Analysis Queries

Query of Analysis of Category of Items Sold The Most

```
WITH category_sales AS (  
    SELECT  
        Categories.CategoryName,  
        SUM([Order Details].Quantity) as QuantitySold,  
        SUM([Order Details].Quantity*[Order Details].UnitPrice) as Revenue  
    FROM  
        Products  
    JOIN  
        [Order Details] ON Products.ProductID = [Order Details].ProductID  
    JOIN  
        Categories ON Products.CategoryID = Categories.CategoryID  
    GROUP BY  
        Categories.CategoryName  
)  
SELECT  
    CategoryName,  
    QuantitySold,  
    Revenue  
FROM  
    category_sales  
WHERE  
    QuantitySold > (SELECT AVG(QuantitySold) FROM category_sales)
```

	CategoryName	QuantitySold	Revenue
1	Beverages	9532	286526,95
2	Confections	7906	177099,10
3	Dairy Products	9149	251330,50
4	Seafood	7681	141623,09

Product Analysis Queries

Query of Analysis of The Best-Selling Products

```
WITH Product_Sales AS (  
    SELECT  
        Products.ProductID, Products.ProductName,  
        SUM(Quantity) AS Total_Quantity_Sold  
    FROM  
        Orders  
    INNER JOIN  
        [Order Details] ON Orders.OrderID = [Order Details].OrderID  
    INNER JOIN  
        Products ON [Order Details].ProductID = Products.ProductID  
    WHERE  
        ShippedDate IS NOT NULL  
    GROUP BY  
        Products.ProductID, Products.ProductName  
)  
SELECT TOP 10  
    ProductID,  
    ProductName,  
    Total_Quantity_Sold,  
    ROW_NUMBER() OVER (ORDER BY Total_Quantity_Sold DESC) AS Rank  
FROM  
    Product_Sales
```

	ProductID	ProductName	Total_Quantity_Sold	Rank
1	60	Camembert Pierrot	1504	1
2	59	Raclette Courdavault	1496	2
3	31	Gorgonzola Telino	1377	3
4	56	Gnocchi di nonna Alice	1263	4
5	75	Rhönbräu Klosterbier	1151	5
6	16	Pavlova	1112	6
7	40	Boston Crab Meat	1103	7
8	24	Guaraná Fantástica	1095	8
9	62	Tarte au sucre	1083	9
10	71	Flotemysost	1036	10

Customer Analysis Queries

Query of Analysis of Customer Spending by Country

```
WITH CountryRevenue AS(  
  SELECT  
    Customers.Country,  
    SUM([Order Details].Quantity*[Order Details].UnitPrice)  
  as Revenue  
  FROM Customers  
  JOIN Orders ON Customers.CustomerID = Orders.CustomerID  
  JOIN [Order Details] ON [Order Details].OrderID =  
    Orders.OrderID  
  GROUP BY Country)  
SELECT * FROM CountryRevenue  
WHERE Revenue > (SELECT AVG(Revenue) FROM CountryRevenue)
```

	Country	Revenue
1	USA	263566,98
2	Brazil	114968,48
3	Germany	244640,63
4	Austria	139496,63
5	France	85498,76

Customer Analysis Queries

Query of Analysis of The Most Frequent Customers by Number of Orders

```
WITH CustomerOrderCount AS (  
  SELECT  
    CustomerID,  
    COUNT(OrderID) AS OrderCount  
  FROM Orders  
  GROUP BY CustomerID  
)  
SELECT TOP 10  
  Customers.CompanyName,  
  CustomerOrderCount.OrderCount  
FROM Customers  
JOIN CustomerOrderCount ON Customers.CustomerID =  
  CustomerOrderCount.CustomerID  
WHERE  
  CustomerOrderCount.OrderCount > (SELECT AVG(OrderCount) FROM  
  CustomerOrderCount)  
ORDER BY  
  CustomerOrderCount.OrderCount DESC
```

Results Messages		
	CompanyName	OrderCount
1	Save-a-lot Markets	31
2	Ernst Handel	30
3	QUICK-Stop	28
4	Folk och få HB	19
5	Hungry Owl All-Night Grocers	19
6	Berglunds snabbköp	18
7	HILARION-Abastos	18
8	Rattlesnake Canyon Grocery	18
9	Bon app'	17
10	Frankenversand	15

Shipper Analysis Queries

Query of Analysis of Average Delivery Time per Shipper

```
WITH deliveries AS (  
    SELECT  
        Shippers.CompanyName AS Shipper,  
        DATEDIFF(day, Orders.OrderDate, Orders.ShippedDate)  
    AS DeliveryTime  
    FROM Orders  
    JOIN Shippers ON Orders.ShipVia = Shippers.ShipperID  
)  
SELECT  
    Shipper,  
    AVG(DeliveryTime) AS AverageDeliveryTime  
FROM deliveries  
GROUP BY Shipper
```

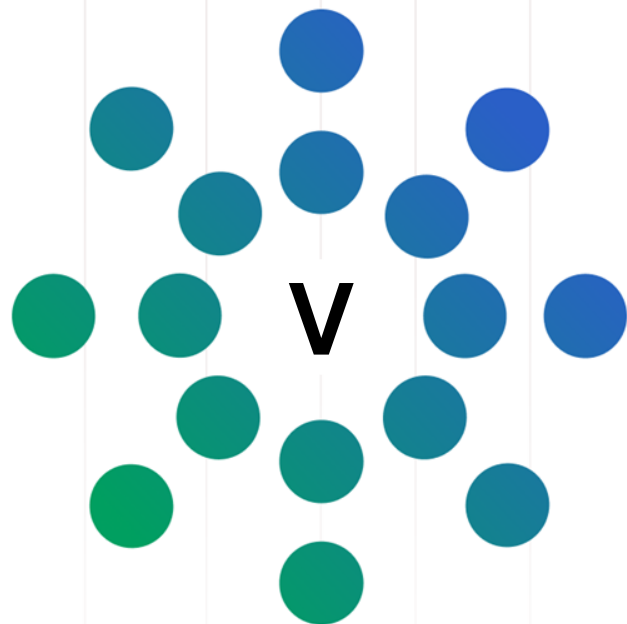
	Shipper	AverageDeliveryTime
1	Federal Shipping	7
2	Speedy Express	8
3	United Package	9

Shipper Analysis Queries

Query of Analysis of The Busiest Shippers

```
SELECT
    Shippers.CompanyName,
    COUNT(Orders.OrderID) AS NumberOfOrders
FROM
    Orders
    INNER JOIN Shippers
    ON Orders.ShipVia = Shippers.ShipperID
GROUP BY
    Shippers.CompanyName
ORDER BY
    NumberOfOrders DESC
```

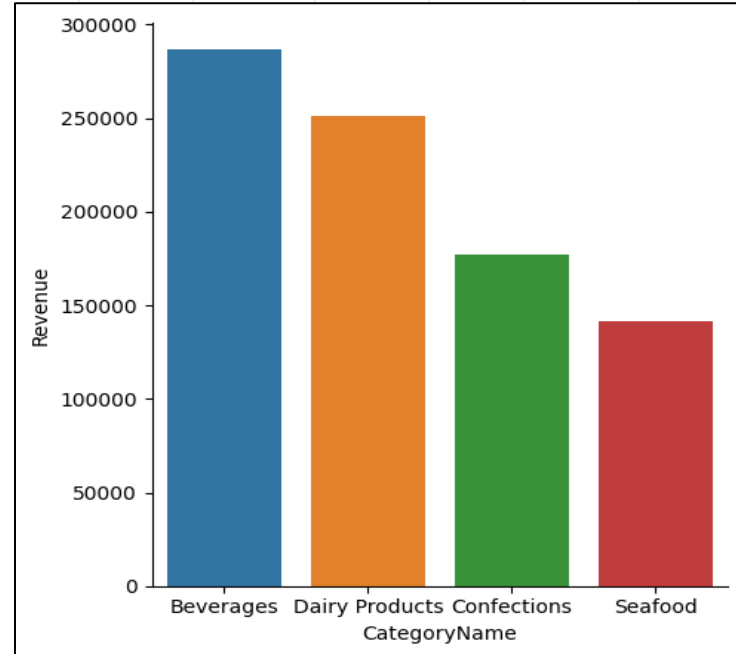
Results			Messages	
	CompanyName	NumberOfOrders		
1	United Package	326		
2	Federal Shipping	255		
3	Speedy Express	249		



Data Analysis

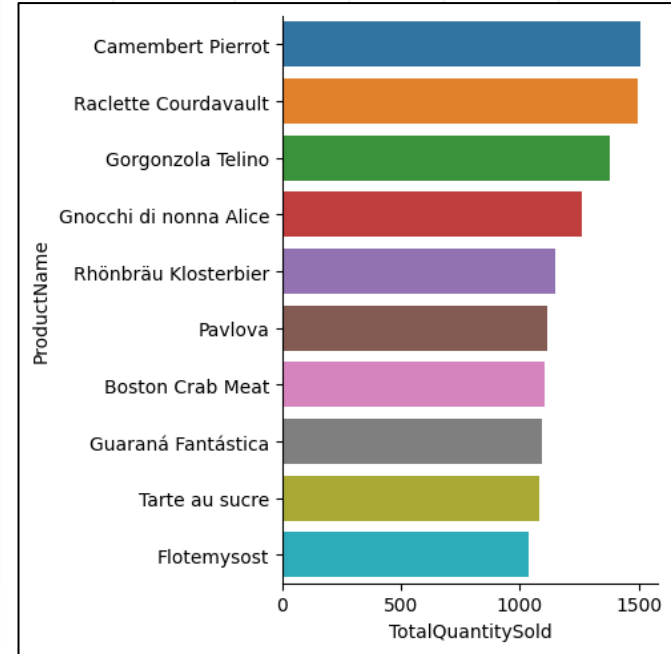
What Category of Items Sold The Most?

Based on the data visualization, it is evident that the product category that generated the highest revenue and sold the most units is Beverages, while the category with the lowest revenue is Seafood. The findings are aligned with the analysis of the Northwind database and provide valuable insights into the market performance of different product categories.



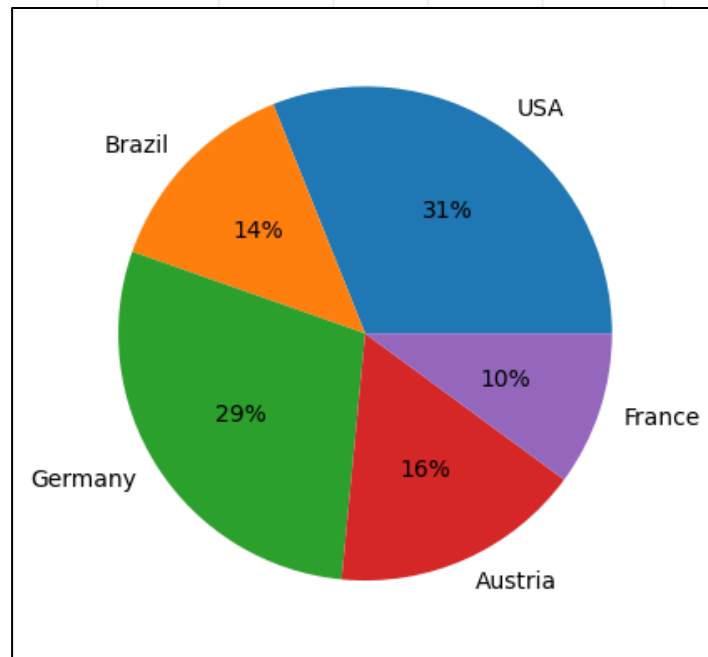
What is The Best-Selling Product?

Based on the data visualization displayed alongside, it is evident that the most highly sold product is Camembert Pierrot. Further analysis of the sales figures for this product indicates that it has consistently been one of the top-performing products within the Northwind inventory.



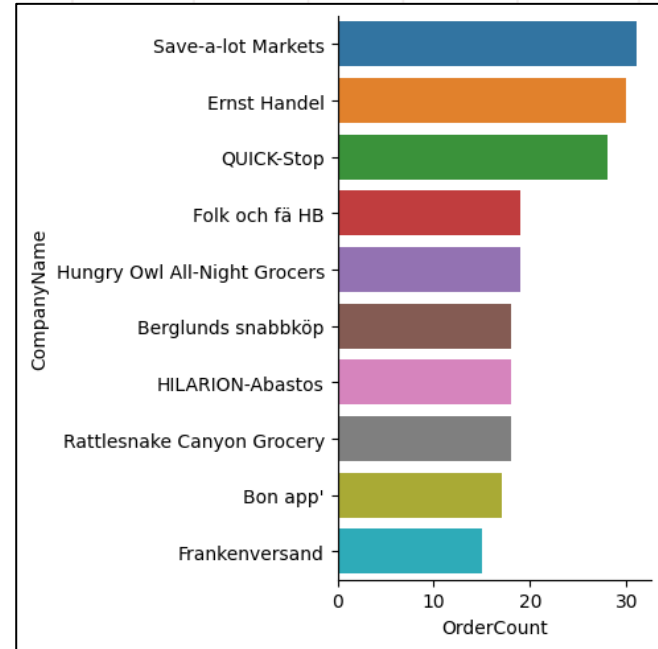
From Which Country Do The Customers Originate?

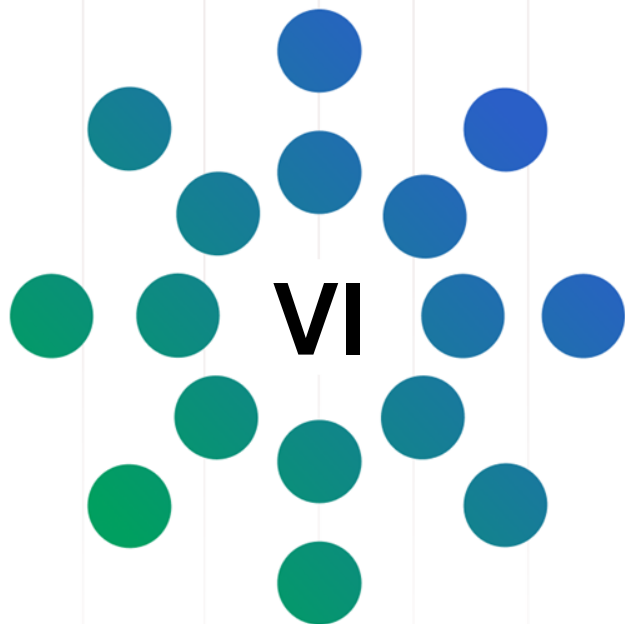
Based on the visualization data displayed, it is evident that most customers are located in the United States, while the few are from France. This can provide valuable insights into the geographical distribution of the customer base and can be used to inform future business strategies and marketing efforts.



Which Company Places The Most Orders?

Based on the visualization data shown, it can be seen that the company that places the most orders is Save-a-lot Markets, and the company that places the least orders is Frankenversand. The insights gleaned from this analysis indicate the purchasing patterns of different companies and could provide valuable information for businesses looking to optimize their sales and marketing strategies.





Summaries & Recommendations

Based on the above data analysis, the following conclusions can be drawn

- From the above facts, it can be concluded that the Beverages product category is the most profitable and highly sold in the Northwind database. This information can inform product development and inventory management strategies by prioritizing the production of similar products or expanding the Beverages category.
- Furthermore, the high sales performance of Camembert Pierrot highlights the potential for similar products to perform well in the market. This could inform product development and marketing efforts for similar products.
- The majority of customers are located in the United States, which suggests that there may be potential for further growth in this market. Marketing and sales efforts could be focused on this region to target potential customers and increase sales.
- Save-a-lot Markets place the most orders, suggesting they may be a valuable target for business-to-business sales and marketing efforts. On the other hand, the low number of orders placed by Frankenversand highlights the potential for further growth in this market and the opportunity to target this company for increased sales.



VII

References

- <https://github.com/yugabyte/yugabyte-db/wiki/Northwind-Sample-Database>
- <https://medium.com/analytics-and-data/the-path-to-learning-sql-and-mastering-it-to-become-a-data-engineer-256ea0fef4e7>
- <https://www.mssqltips.com/sqlservertip/6921/data-engineering-solution-python-sql-server/>

Thank You!

