

Data science learning studio

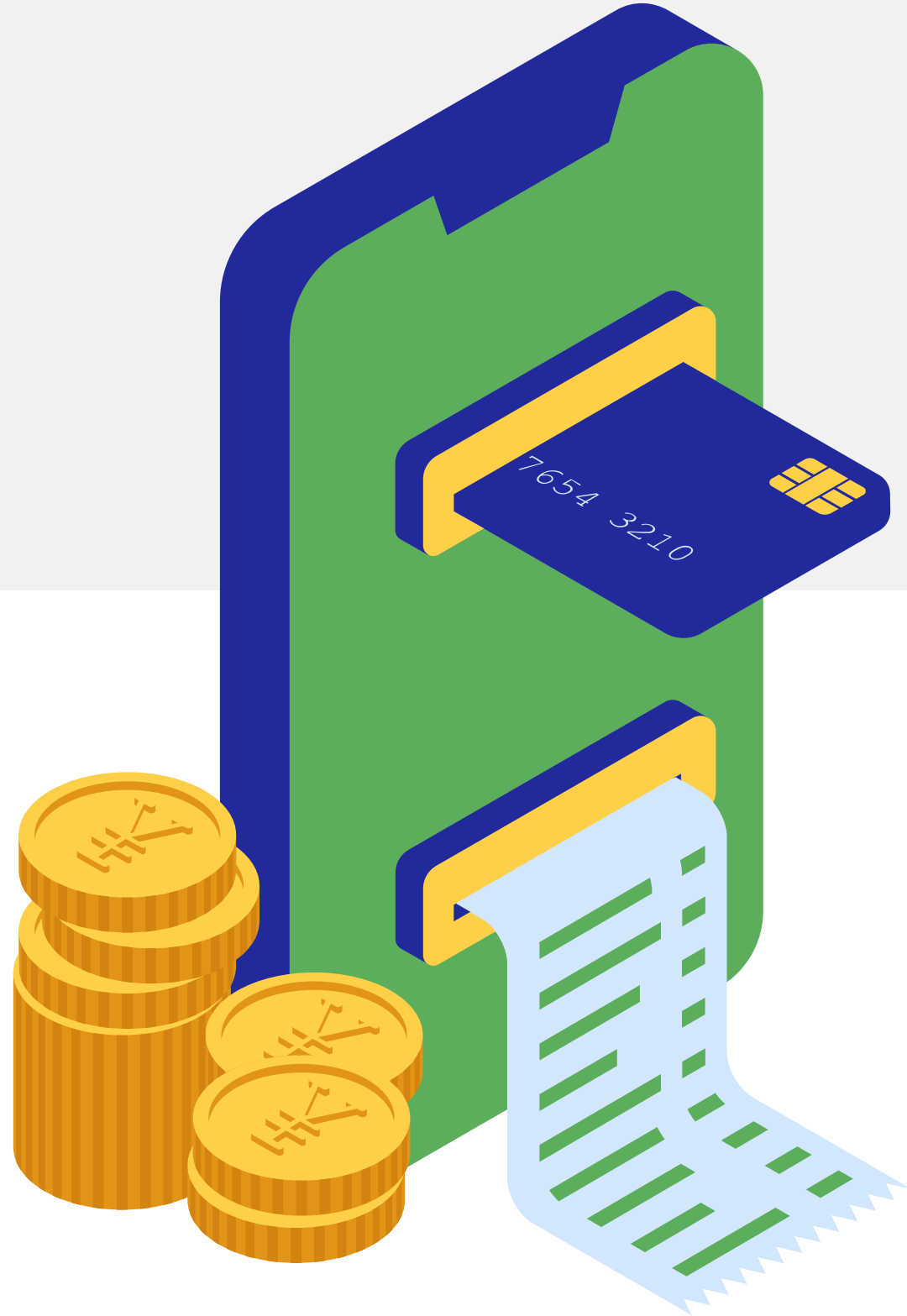
Data Engineering Project

Muhammad Farras Rizki

Table Of Contents



- 01 Introduction
Business Understanding & Tools
- 02 Extract Data
Extract Data & Check ERD
- 03 Transform Data
Querying & Flowchart
- 04 Load Data
Analysis & Report



Step 1

Introduction

Business Understanding

Background & Objective

Northwind Traders is a company engaged in the export and import of food. From the transaction history, some insights will be obtained that can be used to improve the company's performance.

Tools



Microsoft SQL Server

Database and
Querying



Google Data Studio

Visualization Data

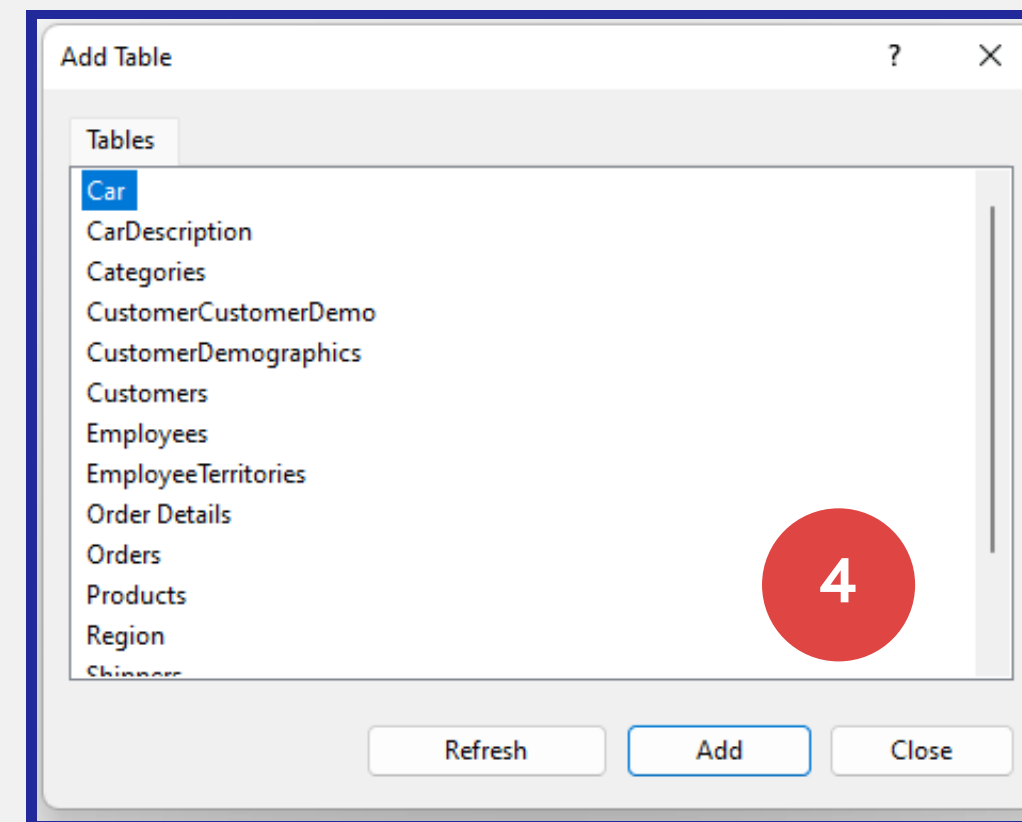
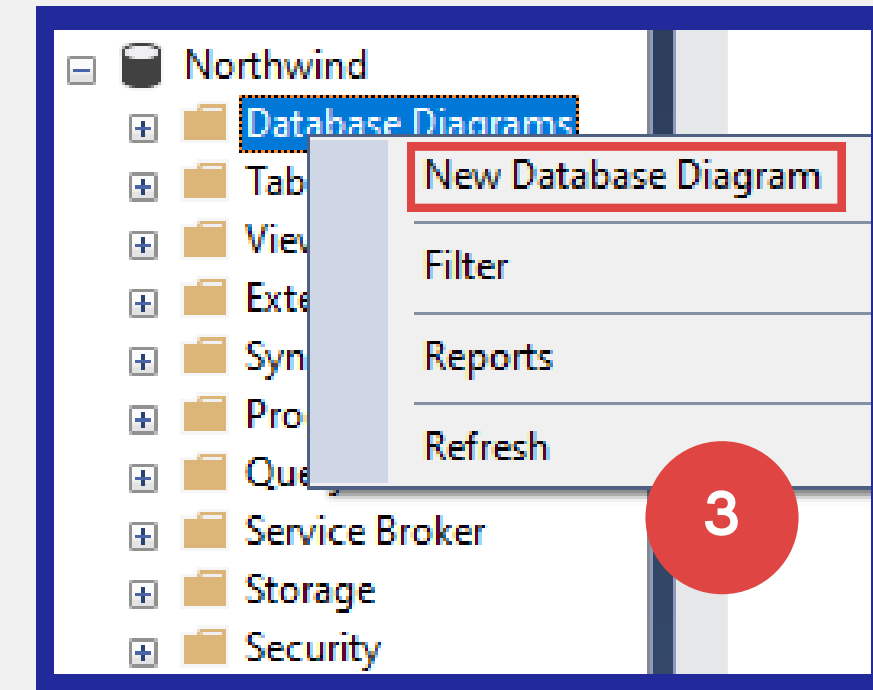
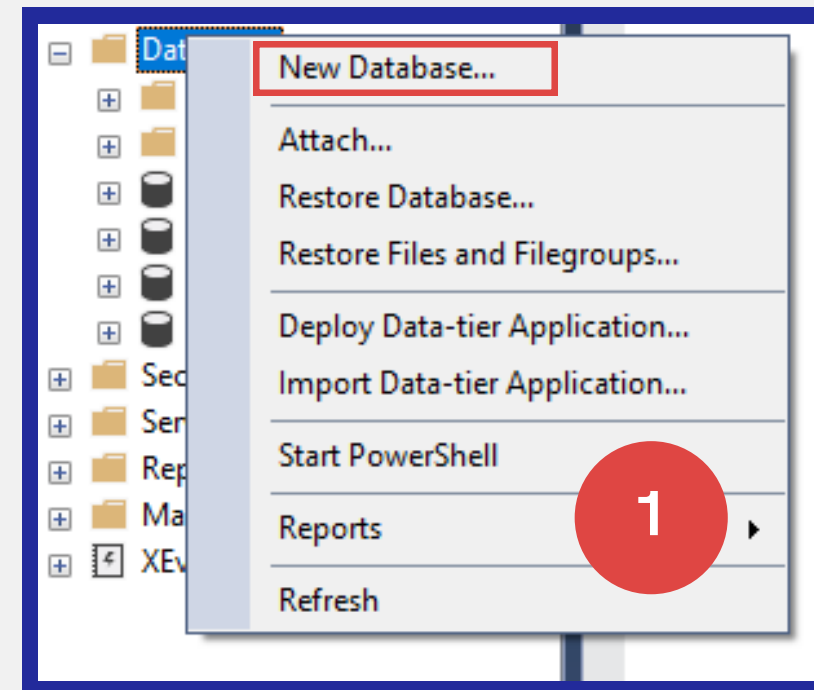


Step 2

Extract Data

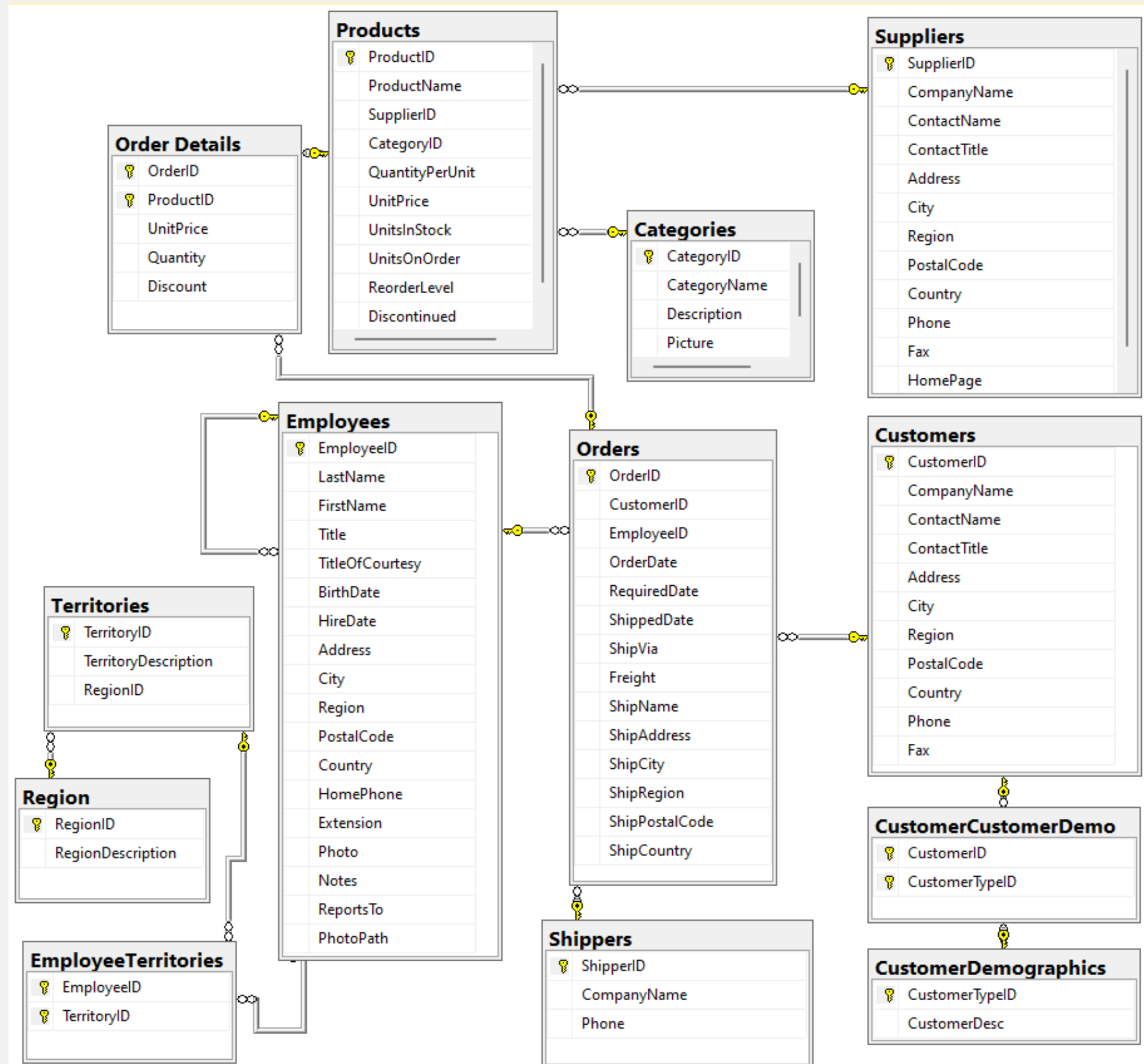
Extract Data

1. Create New Database
2. Open the query file and execute it
(data source format is query)
3. Create New Database Diagram
4. Select the table that will be used



ERD

(Entity Relationship Diagram)





Step 3

Transform Data

1. Intermediate Queries
2. Study Case

A. Intermediate Queries

1. Number of customers for each month who placed orders in 1997

Query :

```
## 1.  
SELECT  
MONTH(OrderDate) AS Month,  
COUNT(CustomerID) AS TotalCustomer  
FROM Orders  
WHERE YEAR(OrderDate) = 1997  
GROUP BY MONTH(OrderDate);
```



Output :

	Month	TotalCustomer
1	1	33
2	2	29
3	3	30
4	4	31
5	5	32
6	6	30
7	7	33
8	8	33
9	9	37
10	10	38
11	11	34
12	12	48

Conclusion :

The highest number of customers who placed orders was in December. From the table, quarter 4 is the quarter with the highest number of customers placing orders.



A. Intermediate Queries

2. Show names of employees who are Sales Representatives

Query :

```
## 2.  
SELECT  
FirstName + ' ' + Lastname AS Name  
FROM Employees  
WHERE Title = 'Sales Representative';
```



Output :

	Name
1	Nancy Davolio
2	Janet Leverling
3	Margaret Peacock
4	Michael Suyama
5	Robert King
6	Anne Dodsworth

Conclusion :

Northwind Traders has 6 employees with the the job title Sales Representative.



A. Intermediate Queries


3. Top 5 products with the highest quantity ordered in January 1997

Query :

```
WITH Table1 as
(SELECT
TOP 5 SUM(od.Quantity) as TotalOrder,
ProductName
FROM [Order Details] od
    LEFT JOIN Products p ON od.ProductID = p.ProductID
    LEFT JOIN Orders o ON od.OrderID = o.OrderID
WHERE YEAR(OrderDate) = 1997
AND MONTH(OrderDate) = 1
GROUP BY ProductName
ORDER BY 1 DESC)

SELECT ProductName, TotalOrder
FROM Table1
```


Output :



	ProductName	TotalOrder
1	Geitost	119
2	Raclette Courdavault	115
3	Gnocchi di nonna Alice	105
4	Vegie-spread	100
5	Côte de Blaye	99

Conclusion :

The best-selling product in January of 1997 was 'Geitost'.




A. Intermediate Queries

4. Companies that ordered 'Chai' in June 1997

Query :

```
SELECT CompanyName
FROM [Order Details] od
    LEFT JOIN Products p ON od.ProductID = p.ProductID
    LEFT JOIN Orders o ON od.OrderID = o.OrderID
    LEFT JOIN Customers c ON o.CustomerID = c.CustomerID
WHERE YEAR(OrderDate) = 1997
AND MONTH(OrderDate) = 6
AND ProductName = 'Chai';
```


Output :



	CompanyName
1	Tortuga Restaurante

Conclusion :

The company that ordered Chai products in June 1997 was Tortuga Restaurante.



A. Intermediate Queries

5. Show the number of OrderIDs that have made purchases by categories (≤ 100), ($100 < x \leq 250$), ($250 < x \leq 500$), and (> 500)

Query :

```
WITH table1 as
(SELECT o.OrderID, SUM(UnitPrice * Quantity) as TotalSales
FROM Orders o
    LEFT JOIN Customers c ON o.CustomerID = c.CustomerID
    LEFT JOIN [Order Details] od ON o.OrderID = od.OrderID
GROUP BY o.OrderID)

SELECT
COUNT(CASE WHEN TotalSales <= 100 then OrderID END) AS '<=100',
COUNT(CASE WHEN TotalSales >100 AND TotalSales <= 250 then OrderID END) AS '101-250',
COUNT(CASE WHEN TotalSales >250 AND TotalSales <= 500 then OrderID END) AS '251-500',
COUNT(CASE WHEN TotalSales >500 then OrderID END) AS '>500'
FROM table1
```



Output :

	<=100	101-250	251-500	>500
1	37	64	118	611



Conclution :

The majority of transactions have a purchase amount of more than 500 dollars

A. Intermediate Queries

6. Show the company name of customers who made purchases above 500 dollar in 1997

Query :

```
WITH table1 as
(SELECT SUM(UnitPrice * Quantity) as TotalSales, CompanyName
FROM Orders o
    LEFT JOIN Customers c ON o.CustomerID = c.CustomerID
    LEFT JOIN [Order Details] od ON o.OrderID = od.OrderID
WHERE YEAR(OrderDate) = 1997
GROUP BY CompanyName)

SELECT CompanyName
FROM table1
WHERE TotalSales > 500
```

Output :

	CompanyName
1	Vaffeljemet
2	Ottilies Käseladen
3	Santé Gourmet
4	Let's Stop N Shop
5	Hungry Coyote Import Store
6	Morgenstem Gesundkost
7	HILARION-Abastos
8	Ricardo Adocicados
9	Pericles Comidas clásicas
10	Wartian Herkku
67	Victuailles en stock
68	Berglunds snabbköp
69	Suprêmes délices
70	Blondesddd père et fils
71	The Cracker Box
72	Trail's Head Gourmet Pro...
73	Gourmet Lanchonetes
74	Chop-suey Chinese
75	Hungry Owl All-Night Groc...

Conclusion :

There were 75 companies that made transactions with purchases of more than 500 dollars

A. Intermediate Queries

7. Show the Top 5 highest sales product names for each month in 1997.

Query :

```
WITH table1 as
(SELECT
MONTH(OrderDate) as Month,
ProductName,
row_number() over
(partition by MONTH(OrderDate) order by SUM(od.UnitPrice * Quantity) desc) as Ranking
FROM Orders o
LEFT JOIN [Order Details] od ON o.OrderID = od.OrderID
LEFT JOIN Products p ON od.ProductID = p.ProductID
WHERE YEAR(OrderDate) = 1997
GROUP BY MONTH(OrderDate), ProductName)

SELECT *
FROM table1
WHERE Ranking <= 5
```



Output :

	Month	ProductName	Ranking
1	1	Côte de Blaye	1
2	1	Raclette Courdavault	2
3	1	Vegie-spread	3
4	1	Gnocchi di nonna Alice	4
5	1	Alice Mutton	5
6	2	Pâté chinois	1
7	2	Sirop d'érable	2
8	2	Gnocchi di nonna Alice	3
9	2	Thüringer Rostbratwurst	4
10	2	Rössle Sauerkraut	5

Conclusion :

The product with the highest sales in January was 'Côte de Blaye', while in February it was 'Pâté chinois'.



A. Intermediate Queries


8. Create a view to show Order Details which contains OrderID, ProductID, ProductName, UnitPrice, Quantity, Discount, Price after discount.

Query :

```
CREATE VIEW [NewTable] AS
SELECT OrderID, od.ProductID, ProductName, od.UnitPrice, Quantity, Discount,
od.UnitPrice * (1-Discount) as PriceAfterDisc
FROM [Order Details] od
LEFT JOIN Products p ON od.ProductID = p.ProductID;

SELECT *
FROM NewTable
```

Output :



	OrderID	ProductID	ProductName	UnitPrice	Quantity	Discount	PriceAfterDisc
1	10248	11	Queso Cabrales	14,00	12	0	14
2	10248	42	Singaporean Hokkien Fried Mee	9,80	10	0	9,8
3	10248	72	Mozzarella di Giovanni	34,80	5	0	34,8
4	10249	14	Tofu	18,60	9	0	18,6
5	10249	51	Manjimup Dried Apples	42,40	40	0	42,4
6	10250	41	Jack's New England Clam Chowder	7,70	10	0	7,7

A. Intermediate Queries


9. Create an 'Invoice Procedure' to call CustomerID, CustomerName/company name, OrderID, OrderDate, RequiredDate, ShippedDate if there is a specific CustomerID input.

Query :

```
CREATE PROCEDURE Invoice1 (@CustomerID NCHAR(5)) AS
BEGIN
SET NOCOUNT ON
SELECT O.CustomerID, CompanyName, OrderID, OrderDate, RequiredDate, ShippedDate
FROM Orders O
INNER JOIN Customers C ON C.CustomerID=O.CustomerID
WHERE O.CustomerID = @CustomerID
END

EXEC Invoice1 ANTON;
```

Output :



	CustomerID	CompanyName	OrderID	OrderDate	RequiredDate	ShippedDate
1	ANTON	Antonio Moreno Taquería	10365	1996-11-27 00:00:00.000	1996-12-25 00:00:00.000	1996-12-02 00:00:00.000
2	ANTON	Antonio Moreno Taquería	10507	1997-04-15 00:00:00.000	1997-05-13 00:00:00.000	1997-04-22 00:00:00.000
3	ANTON	Antonio Moreno Taquería	10535	1997-05-13 00:00:00.000	1997-06-10 00:00:00.000	1997-05-21 00:00:00.000
4	ANTON	Antonio Moreno Taquería	10573	1997-06-19 00:00:00.000	1997-07-17 00:00:00.000	1997-06-20 00:00:00.000
5	ANTON	Antonio Moreno Taquería	10677	1997-09-22 00:00:00.000	1997-10-20 00:00:00.000	1997-09-26 00:00:00.000
6	ANTON	Antonio Moreno Taquería	10682	1997-09-25 00:00:00.000	1997-10-23 00:00:00.000	1997-10-01 00:00:00.000
7	ANTON	Antonio Moreno Taquería	10856	1998-01-28 00:00:00.000	1998-02-25 00:00:00.000	1998-02-10 00:00:00.000


B. Case Study

1. How quantity trend of 'Chartreuse verte' products in every year ?

Query :

```
SELECT
ProductName,
SUM(Quantity) AS TotalQuantity,
YEAR(OrderDate) AS Year
FROM [Order Details] AS od
    LEFT JOIN Products AS p ON od.ProductID = p.ProductID
    LEFT JOIN Orders AS o ON o.OrderID = od.OrderID
WHERE ProductName = 'Chartreuse verte'
GROUP BY YEAR(OrderDate), ProductName
```

Output :



	ProductName	TotalQuantity	Year
1	Chartreuse verte	266	1996
2	Chartreuse verte	283	1997
3	Chartreuse verte	244	1998

Conclution :

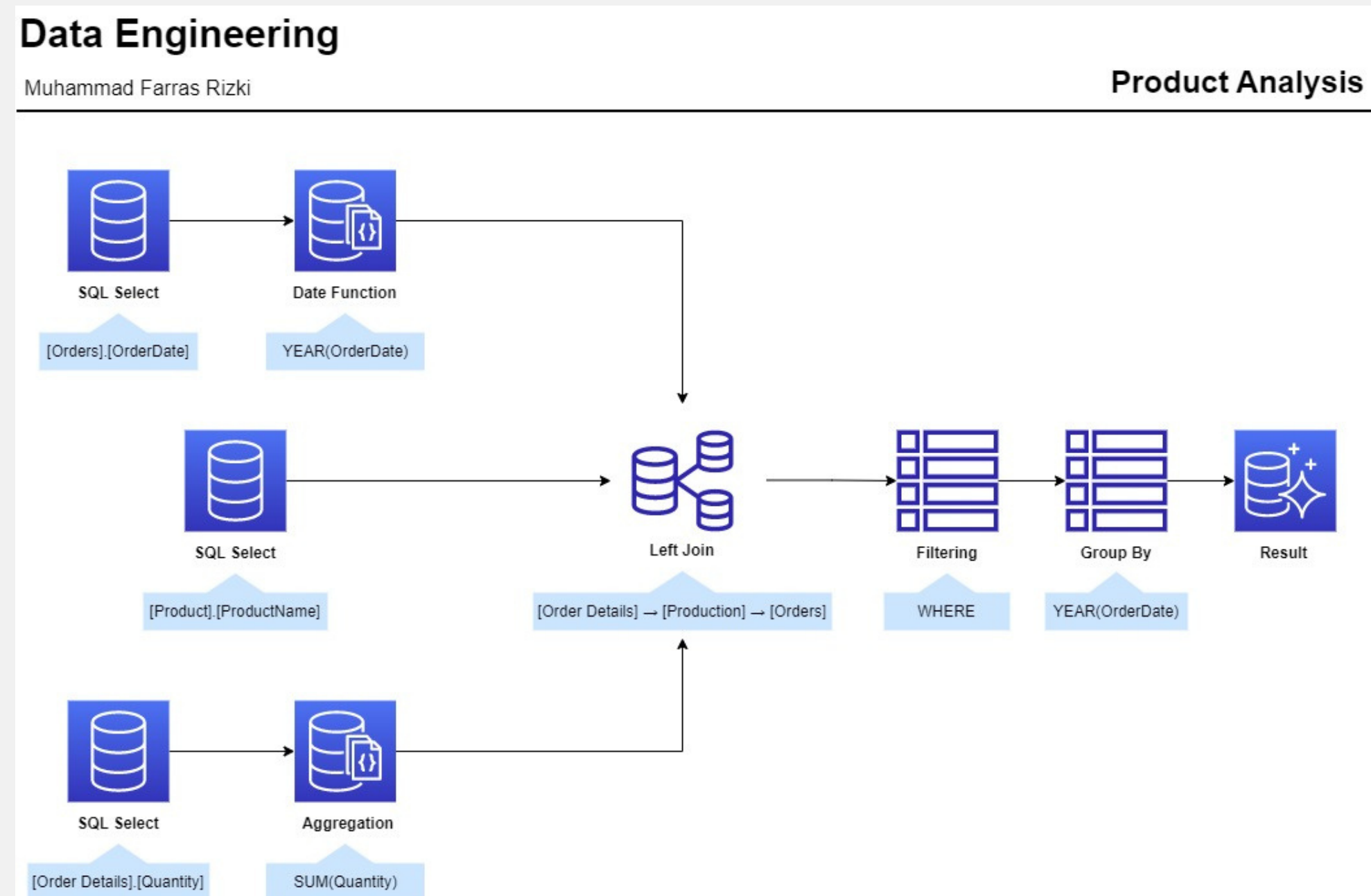
Sales of 'Chartreuse verte' product were highest in 1997.



B. Case Study

1. How quantity trend of 'Chartreuse verte' products in every year ?

Flowchart :




B. Case Study

2. Show 5 most loyal companies customers by highest number of transactions

Query :

```
SELECT TOP 5 CompanyName, SUM(UnitPrice * Quantity) AS Revenue
FROM [Order Details] AS od
    LEFT JOIN Orders AS o ON o.OrderID = od.OrderID
    LEFT JOIN Customers AS c ON o.CustomerID = c.CustomerID
GROUP BY CompanyName
ORDER BY SUM(UnitPrice * Quantity) DESC
```


Output :



	CompanyName	Revenue
1	QUICK-Stop	117483,39
2	Save-a-lot Markets	115673,39
3	Ernst Handel	113236,68
4	Hungry Owl All-Night Grocers	57317,39
5	Rattlesnake Canyon Grocery	52245,90

Conclusion :

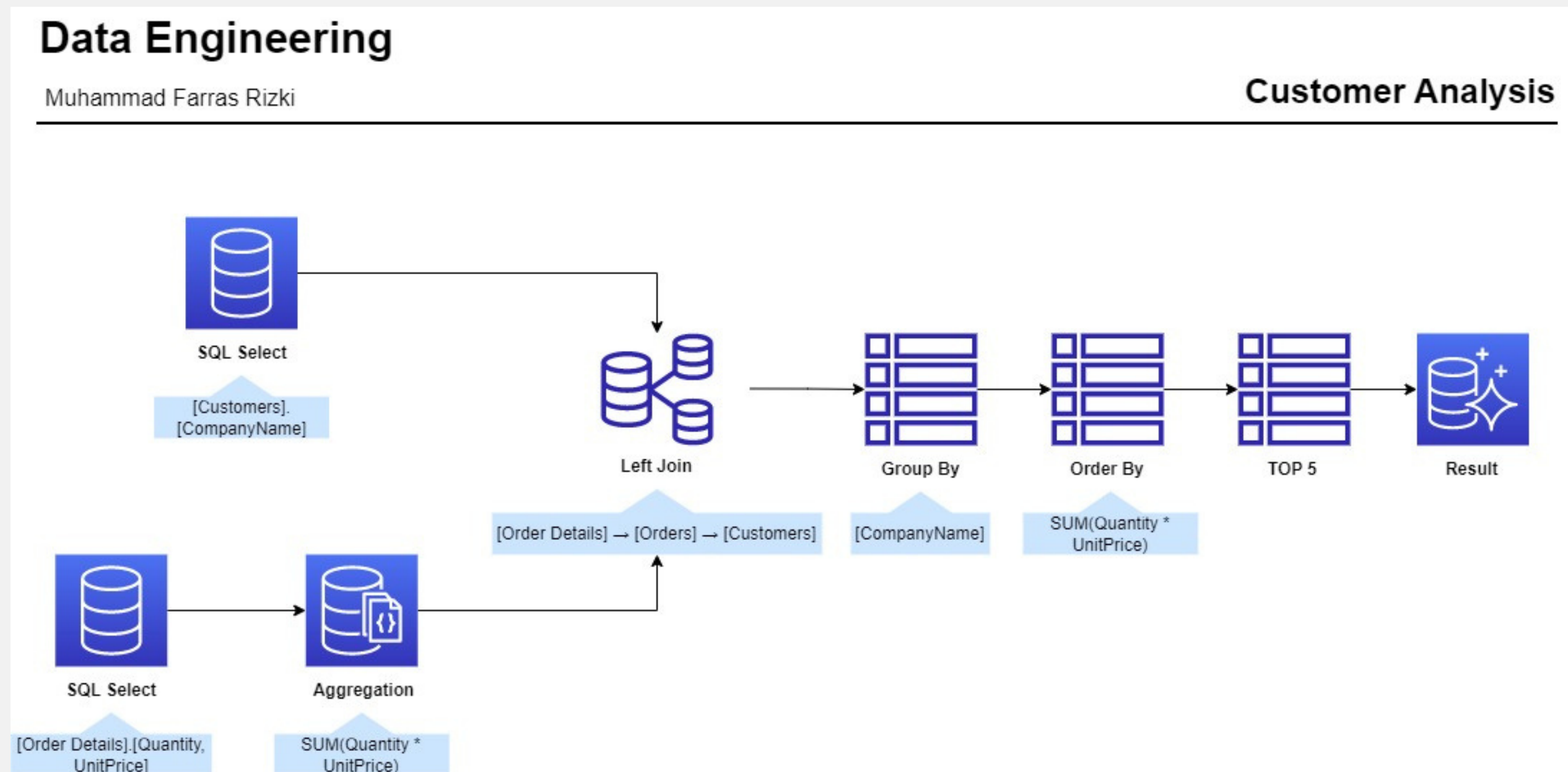
The most loyal company / has the highest number of transactions is
QUICK-Stop



B. Case Study

2. Show 5 most loyal companies customers by highest revenue

Flowchart :




B. Case Study

3. Top 5 cities with the highest number of shipments

Query :

```
SELECT TOP 5 ShipCity, COUNT(OrderID) AS TotalShipping
FROM Orders
GROUP BY ShipCity
ORDER BY COUNT(OrderID) DESC
```




Output :

	ShipCity	TotalShipping
1	Rio de Janeiro	34
2	London	33
3	Boise	31
4	Sao Paulo	31
5	Graz	30

Conclution :

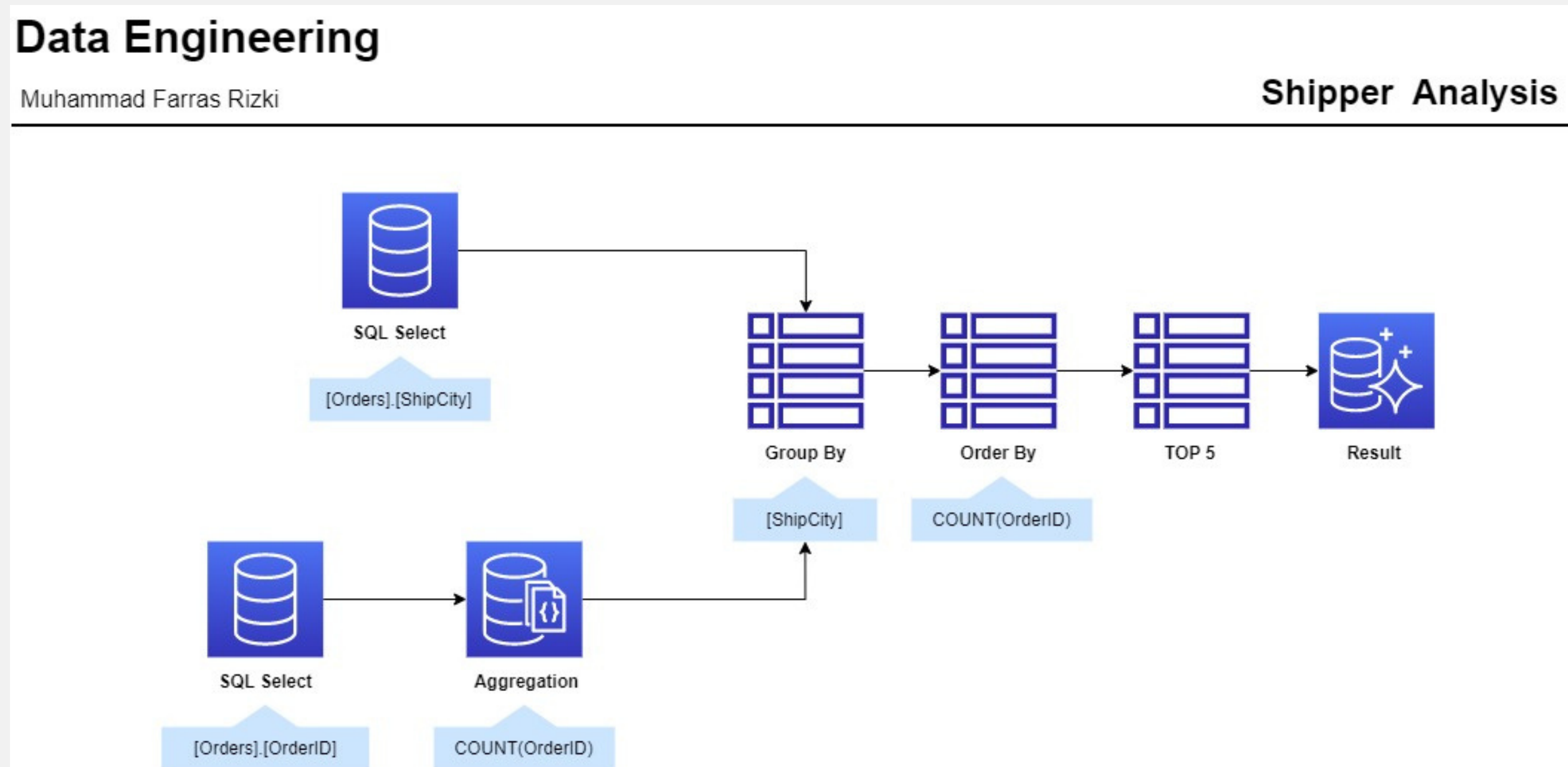
The top 5 shipping destinations are Rio de Janeiro, London, Boise, Sao Paulo, and Graz.



B. Case Study

3. Top 5 cities with the highest number of shipments

Query :




B. Case Study

4. What are the 3 most purchased products with 25% discount ?

Query :

```
SELECT TOP 3 ProductName, Quantity, Discount
FROM [Order Details] AS od
    LEFT JOIN Products AS p ON od.ProductID = p.ProductID
WHERE Discount = 0.25
ORDER BY Quantity DESC
```

Output :



	ProductName	Quantity	Discount
1	Sirop d'érable	120	0,25
2	Chang	100	0,25
3	Raclette Courdavault	100	0,25

Conclusion :

The most purchased products with 25% discount are 'Sirop d'érable', 'Chang' and 'Raclette Courdavault'.




B. Case Study

5. Top 3 product categories are most purchased

Query :

```
SELECT TOP 3 CategoryName, SUM(Quantity) AS Quantity
FROM [Order Details] AS od
    LEFT JOIN Products AS p ON od.ProductID = p.ProductID
    LEFT JOIN Categories AS c ON c.CategoryID = p.CategoryID
GROUP BY CategoryName
ORDER BY Quantity DESC
```


Output :



	CategoryName	Quantity
1	Beverages	9532
2	Dairy Products	9149
3	Confections	7906

Conclusion :

The most sold product categories are Beverages, Dairy Products, and Confections.





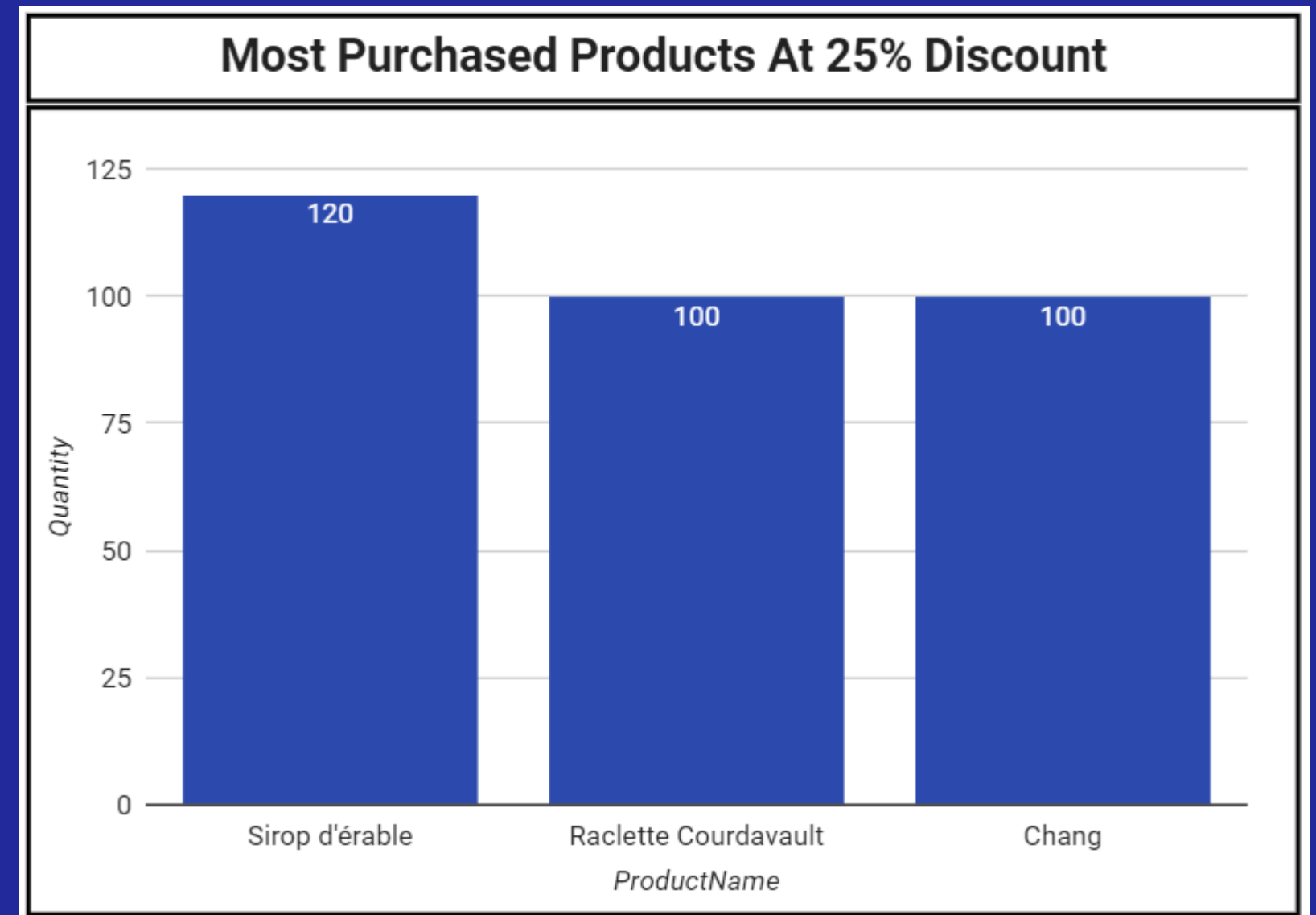
Step 4

Load Data

Product Analysis

Top 3 Purchased Product with 25% Discount

The most purchased product with a 25% discount was 'Sirop d'érable', with a total of 120 items sold.

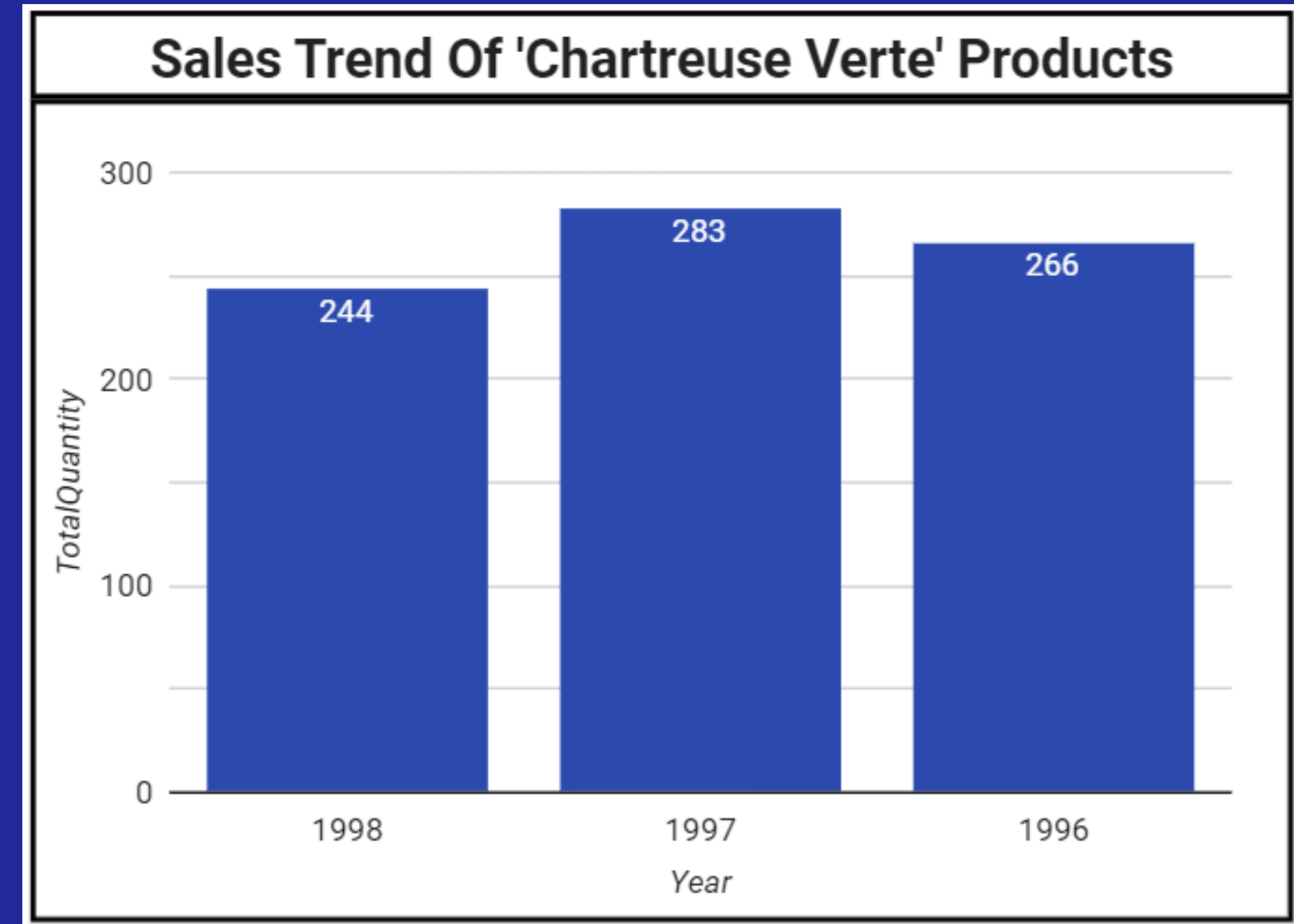


Product Analysis

Trend of 'Chartreuse verte' products

Sales of 'Chartreuse verte' were highest in 1997, with a total of 283 items sold.

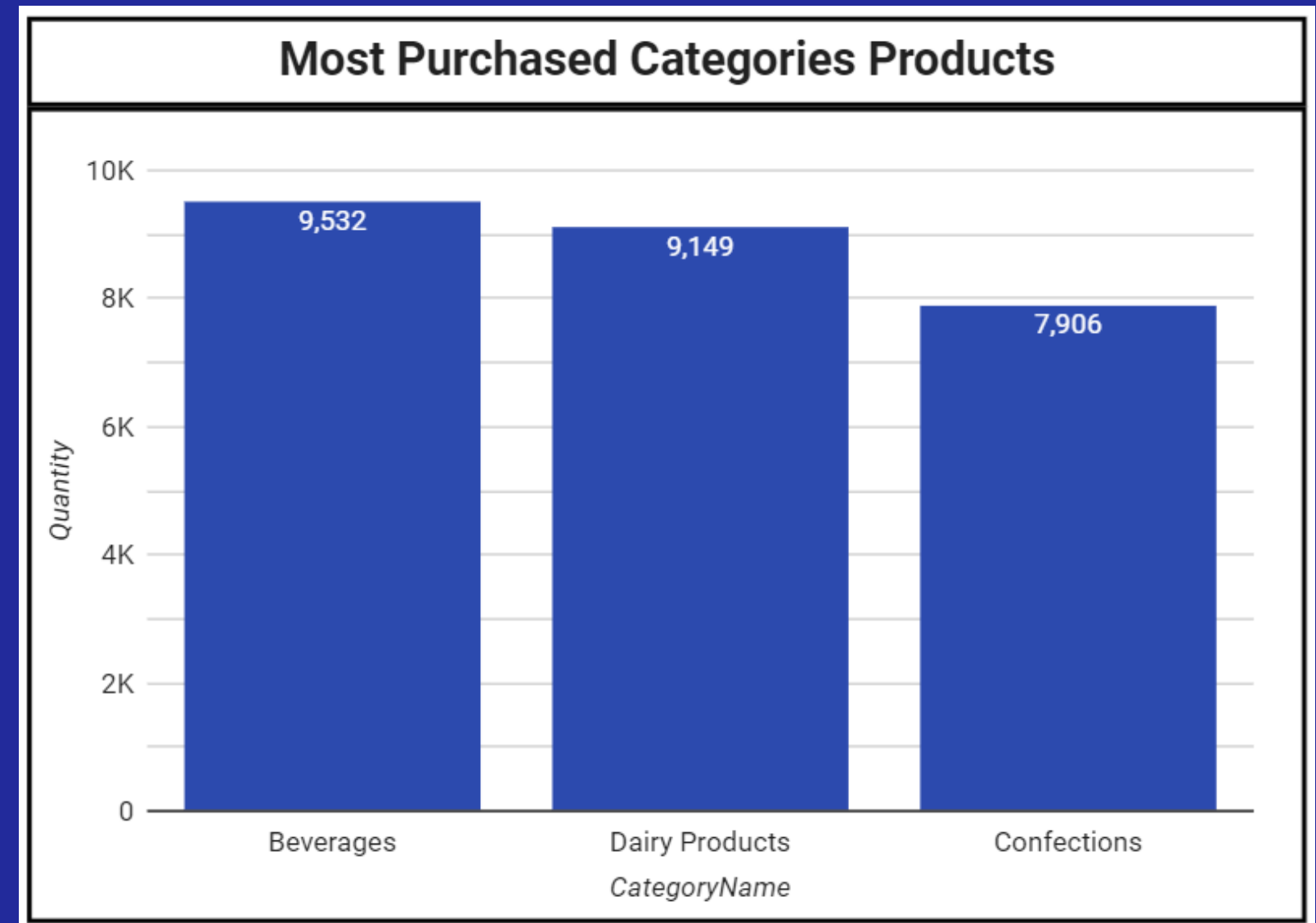
When viewed from the bar chart, there is no significant change in the sales of 'Chartreuse verte' products.



Product Analysis

Top 3 Purchased Categories Product

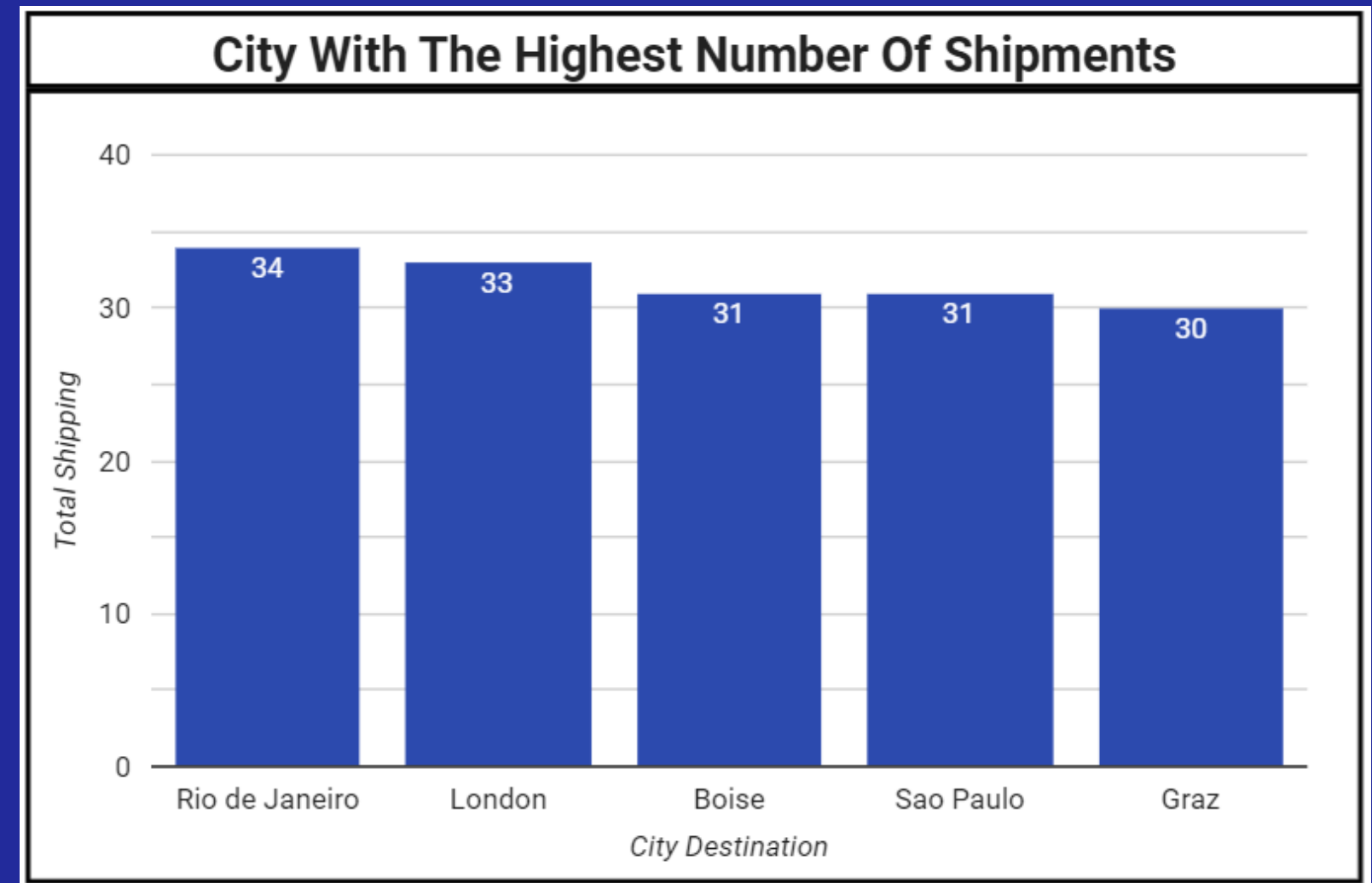
The most sold product category is Beverages with total sales of 9532 pieces, followed by Dairy Products and Confections.



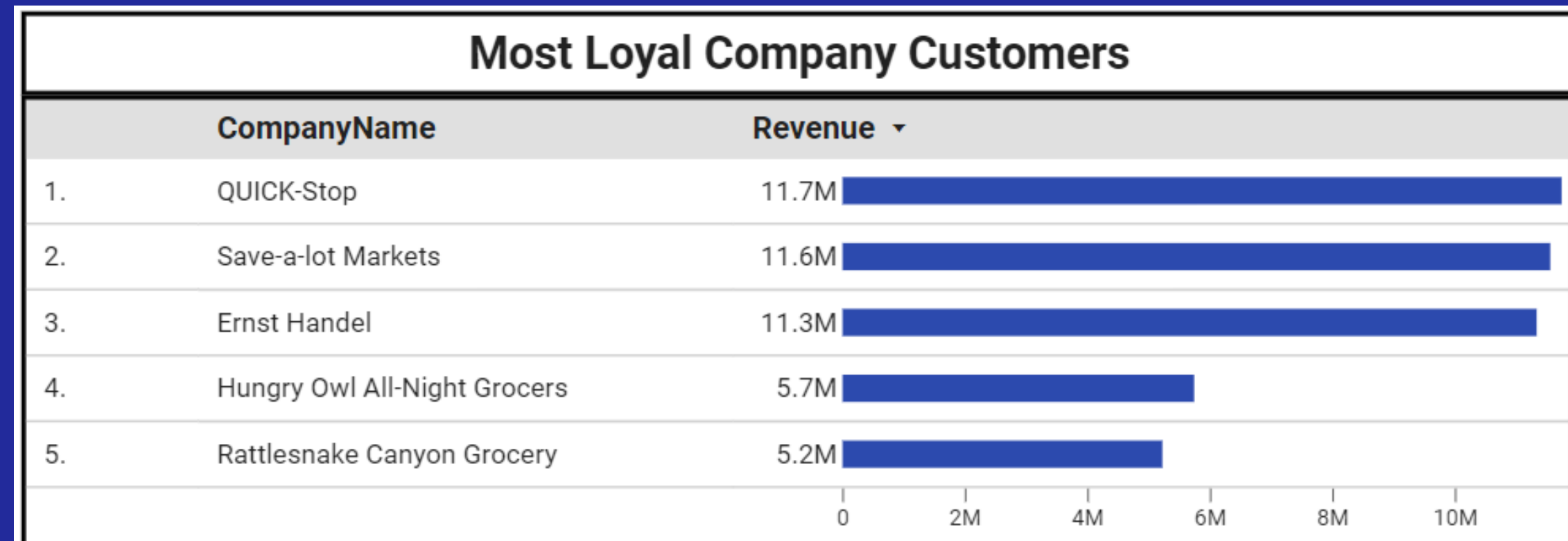
Shipper Analysis

Top 5 Shipping Destination City

Rio De Janeiro was the city with the highest total shipments, with 34 shipments



Customer Analysis



Most Loyal Company Customers

There are 3 customer companies that contributed the highest revenue of more than 11 million dollars, namely 'QUICK-Stop', 'Save-a-lot Markets' and 'Ernst Handel'. 'QUICK-Stop' is in the highest position with a revenue of 11.7 million dollars.

Recommendation

- If company wanted to do a trial campaign, do it during the 4th quarter with the Beverage product category, because it has the highest sales rate.
- Expand expeditions to Rio De Janeiro and London, so that the performance of shipping remains good despite high orders.
- Give rewards or discounts to loyal Customer Company, to keep them subscribed



Thank You

Very open to constructive criticism and suggestions

Contact Me

mfarrasrizki@gmail.com

<http://www.linkedin.com/in/mfarrasrizki/>

<http://www.github.com/FarrasRizki>

