Data Engineering Mini Project

By Hendra Kuswantoro

✉ hendrakuswantoro20@gmail.com

in www.linkedin.com/in/hendrakuswantoro/

Northwind Report

Data Consultant Bootcamp 2023

Microsoft   DATA SCIENCE INDONESIA   Data Science Learning Studio

Outline

1 Introduction

2 Study Case Phase

3 Process Phase

4 Analysis & Visualization Phase

5 Act Phase

Introduction

Northwind
- Database Diagrams
- Tables
  - System Tables
  - FileTables
  - External Tables
  - Graph Tables
  - dbo.Categories
  - dbo.CustomerCustomerDemo
  - dbo.CustomerDemographics
  - dbo.Customers
  - dbo.Employees
  - dbo.EmployeeTerritories
  - dbo.Order Details
  - dbo.Orders
  - dbo.Products
  - dbo.Region
  - dbo.Shippers
  - dbo.Suppliers
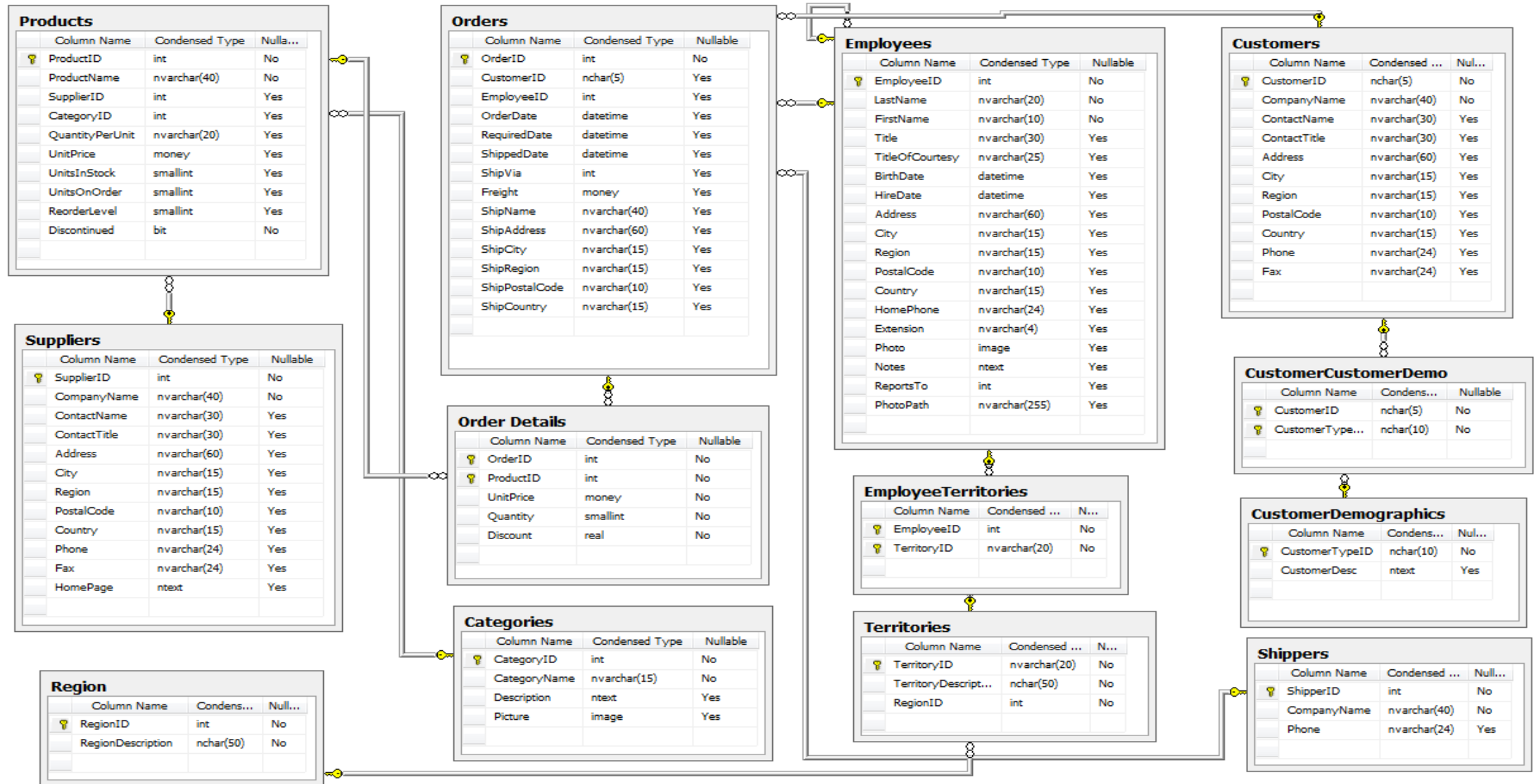 - dbo.Territories
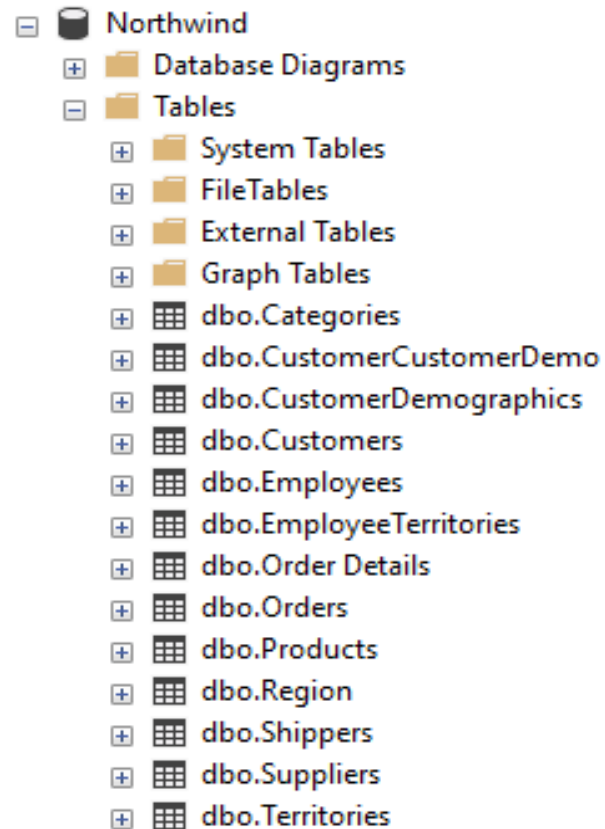
The **Northwind** database describes a database belonging to a fictitious company called **Northwind** Traders where this company is engaged in the export and import of food. In this database there are tables like in an import-export company such as tables of **Orders, Products, Customers, Shippers** and so on.

Schema Diagram of Northwind Database

Products

	Column Name	Condensed Type	Nulla...
🔑	ProductID	int	No
	ProductName	nvarchar(40)	No
	SupplierID	int	Yes
	CategoryID	int	Yes
	QuantityPerUnit	nvarchar(20)	Yes
	UnitPrice	money	Yes
	UnitsInStock	smallint	Yes
	UnitsOnOrder	smallint	Yes
	ReorderLevel	smallint	Yes
	Discontinued	bit	No

Orders

	Column Name	Condensed Type	Nullable
🔑	OrderID	int	No
	CustomerID	nchar(5)	Yes
	EmployeeID	int	Yes
	OrderDate	datetime	Yes
	RequiredDate	datetime	Yes
	ShippedDate	datetime	Yes
	ShipVia	int	Yes
	Freight	money	Yes
	ShipName	nvarchar(40)	Yes
	ShipAddress	nvarchar(60)	Yes
	ShipCity	nvarchar(15)	Yes
	ShipRegion	nvarchar(15)	Yes
	ShipPostalCode	nvarchar(10)	Yes
	ShipCountry	nvarchar(15)	Yes

Employees

	Column Name	Condensed Type	Nullable
🔑	EmployeeID	int	No
	LastName	nvarchar(20)	No
	FirstName	nvarchar(10)	No
	Title	nvarchar(30)	Yes
	TitleOfCourtesy	nvarchar(25)	Yes
	BirthDate	datetime	Yes
	HireDate	datetime	Yes
	Address	nvarchar(60)	Yes
	City	nvarchar(15)	Yes
	Region	nvarchar(15)	Yes
	PostalCode	nvarchar(10)	Yes
	Country	nvarchar(15)	Yes
	HomePhone	nvarchar(24)	Yes
	Extension	nvarchar(4)	Yes
	Photo	image	Yes
	Notes	ntext	Yes
	ReportsTo	int	Yes
	PhotoPath	nvarchar(255)	Yes

Customers

	Column Name	Condensed ...	Nul...
🔑	CustomerID	nchar(5)	No
	CompanyName	nvarchar(40)	No
	ContactName	nvarchar(30)	Yes
	ContactTitle	nvarchar(30)	Yes
	Address	nvarchar(60)	Yes
	City	nvarchar(15)	Yes
	Region	nvarchar(15)	Yes
	PostalCode	nvarchar(10)	Yes
	Country	nvarchar(15)	Yes
	Phone	nvarchar(24)	Yes
	Fax	nvarchar(24)	Yes

Suppliers

	Column Name	Condensed Type	Nullable
🔑	SupplierID	int	No
	CompanyName	nvarchar(40)	No
	ContactName	nvarchar(30)	Yes
	ContactTitle	nvarchar(30)	Yes
	Address	nvarchar(60)	Yes
	City	nvarchar(15)	Yes
	Region	nvarchar(15)	Yes
	PostalCode	nvarchar(10)	Yes
	Country	nvarchar(15)	Yes
	Phone	nvarchar(24)	Yes
	Fax	nvarchar(24)	Yes
	HomePage	ntext	Yes

Order Details

	Column Name	Condensed Type	Nullable
🔑	OrderID	int	No
🔑	ProductID	int	No
	UnitPrice	money	No
	Quantity	smallint	No
	Discount	real	No

CustomerCustomerDemo

	Column Name	Condens...	Nullable
🔑	CustomerID	nchar(5)	No
🔑	CustomerType...	nchar(10)	No

EmployeeTerritories

	Column Name	Condensed ...	N...
🔑	EmployeeID	int	No
🔑	TerritoryID	nvarchar(20)	No

CustomerDemographics

	Column Name	Condens...	Nul...
🔑	CustomerTypeID	nchar(10)	No
	CustomerDesc	ntext	Yes

Categories

	Column Name	Condensed Type	Nullable
🔑	CategoryID	int	No
	CategoryName	nvarchar(15)	No
	Description	ntext	Yes
	Picture	image	Yes

Territories

	Column Name	Condensed ...	N...
🔑	TerritoryID	nvarchar(20)	No
	TerritoryDescript...	nchar(50)	No
	RegionID	int	No

Shippers

	Column Name	Condensed ...	Null...
🔑	ShipperID	int	No
	CompanyName	nvarchar(40)	No
	Phone	nvarchar(24)	Yes

Region

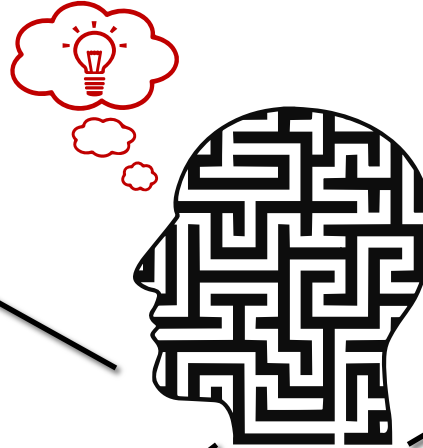|	Column Name	Condens...	Null...
🔑	RegionID	int	No
	RegionDescription	nchar(50)	No

Study Case Phase

Product Analysis

❖ What is the product sales trend in 3 years?
❖ What is the total product sales by category?
❖ What products are the top sales?

Customer Analysis

❖ What is the contact title of the customer based on the number of product sales?
❖ What is the number of customers who place orders in various countries and cities?
❖ What is three categories of customers based on the number of product sales? (Top, Middle, Low)

Shipper Analysis

❖ Which Shipper Company has the most orders?
❖ What is the number of product sales based on the three shipper categories? (Top, Middle, Low)
❖ What country has the most orders using the shipper?

Process Phase

Data Engineering Process of Product Analysis

dbo.Orders

dbo.[Order Details]

dbo.Products

dbo.Categories

INNER JOIN → GROUP BY → SELECT
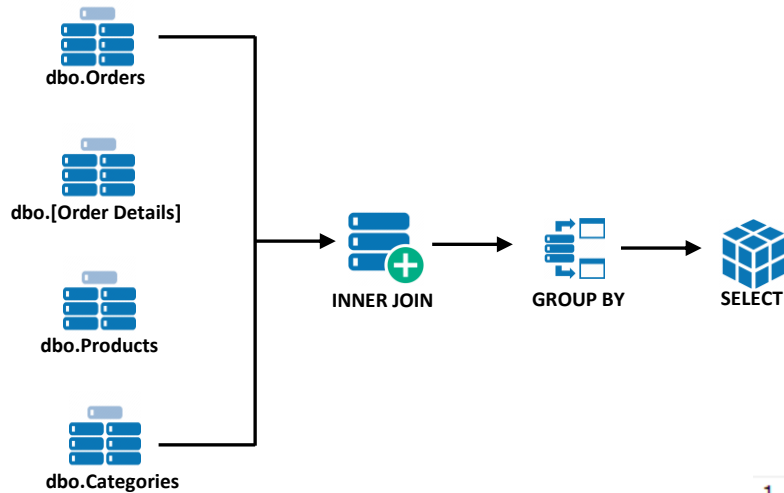
Output

	OrderDate	Product Name	Category Name	Order	Order Price	Product Price	Gap Price	Gross Product Sales	Nett Product Sales
1	1996-07-04 00:00:00.000	Mozzarella di Giovanni	Dairy Products	1	34,80	34,80	0,00	174,00	174
2	1996-07-04 00:00:00.000	Queso Cabrales	Dairy Products	1	14,00	21,00	7,00	168,00	168
3	1996-07-04 00:00:00.000	Singaporean Hokkien Fried Mee	Grains/Cereals	1	9,80	14,00	4,20	98,00	98
4	1996-07-05 00:00:00.000	Manjimup Dried Apples	Produce	1	42,40	53,00	10,60	1696,00	1696
5	1996-07-05 00:00:00.000	Tofu	Produce	1	18,60	23,25	4,65	167,40	167,4
6	1996-07-08 00:00:00.000	Gustaf's Knäckebröd	Grains/Cereals	1	16,80	21,00	4,20	100,80	100,5
7	1996-07-08 00:00:00.000	Jack's New England Clam Chowder	Seafood	1	7,70	9,65	1,95	77,00	77
8	1996-07-08 00:00:00.000	Louisiana Fiery Hot Pepper Sauce	Condiments	2	33,60	42,10	8,50	588,00	585,75
9	1996-07-08 00:00:00.000	Manjimup Dried Apples	Produce	1	42,40	53,00	10,60	1484,00	1478,75
10	1996-07-08 00:00:00.000	Ravioli Angelo	Grains/Cereals	1	15,60	19,50	3,90	234,00	233,25
11	1996-07-09 00:00:00.000	Camembert Pierrot	Dairy Products	1	27,20	34,00	6,80	1088,00	1088
12	1996-07-09 00:00:00.000	Geitost	Dairy Products	1	2,00	2,50	0,50	50,00	48,75
13	1996-07-09 00:00:00.000	Sir Rodney's Marmalade	Confections	1	64,80	81,00	16,20	2592,00	2590
14	1996-07-10 00:00:00.000	Chartreuse verte	Beverages	1	14,40	18,00	3,60	604,80	604,8

Process Phase

Data Engineering Process of Customer Analysis

dbo.Customers

dbo.[Order Details]

dbo.Orders

INNER JOIN → CASE WHEN → GROUP BY → SELECT
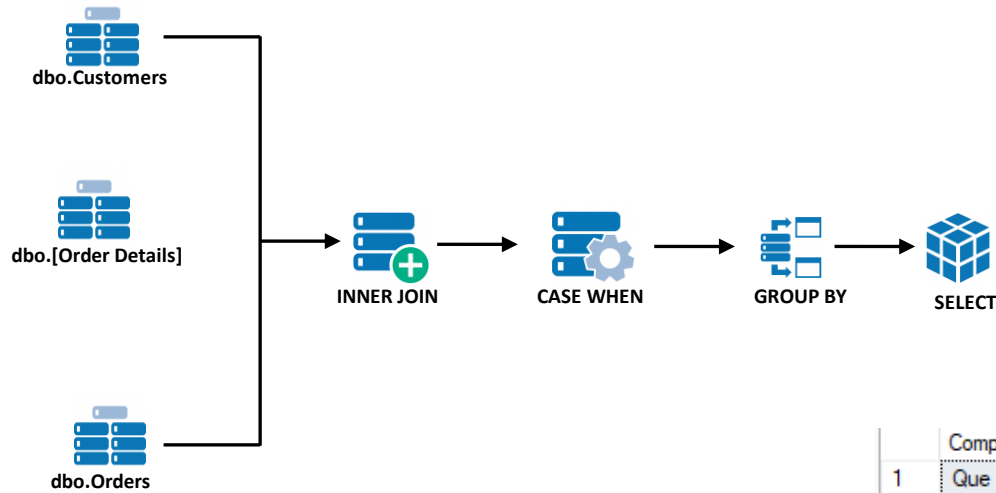
Output

	Company Name	Contact Title	Country	City	Total Order	Product Sales	Customer Category
1	Que Delícia	Accounting Manager	Brazil	Rio de Janeiro	24	6973,63	Low
2	Blondesddsl père et fils	Marketing Manager	France	Strasbourg	26	19088,00	Middle
3	Eastern Connection	Sales Agent	UK	London	21	15033,66	Middle
4	Berglunds snabbköp	Order Administrator	Sweden	Luleå	52	26968,15	Middle
5	Let's Stop N Shop	Owner	USA	San Francisco	10	3490,02	Low
6	Tortuga Restaurante	Owner	Mexico	México D.F.	29	10812,15	Middle
7	Morgenstern Gesundkost	Marketing Assistant	Germany	Leipzig	11	5042,20	Low
8	France restauration	Marketing Manager	France	Nantes	6	3172,16	Low
9	Ana Trujillo Emparedados y helados	Owner	Mexico	México D.F.	10	1402,95	Low
10	Lazy K Kountry Store	Marketing Manager	USA	Walla Walla	2	357,00	Low
11	Centro comercial Moctezuma	Marketing Manager	Mexico	México D.F.	2	100,80	Low
12	La corne d'abondance	Sales Representative	France	Versailles	11	1992,05	Low
13	Lonesome Pine Restaurant	Sales Manager	USA	Portland	14	4258,60	Low
14	Seven Seas Imports	Sales Manager	UK	London	26	17172,05	Middle

Process Phase

Data Engineering Process of Shipper Analysis
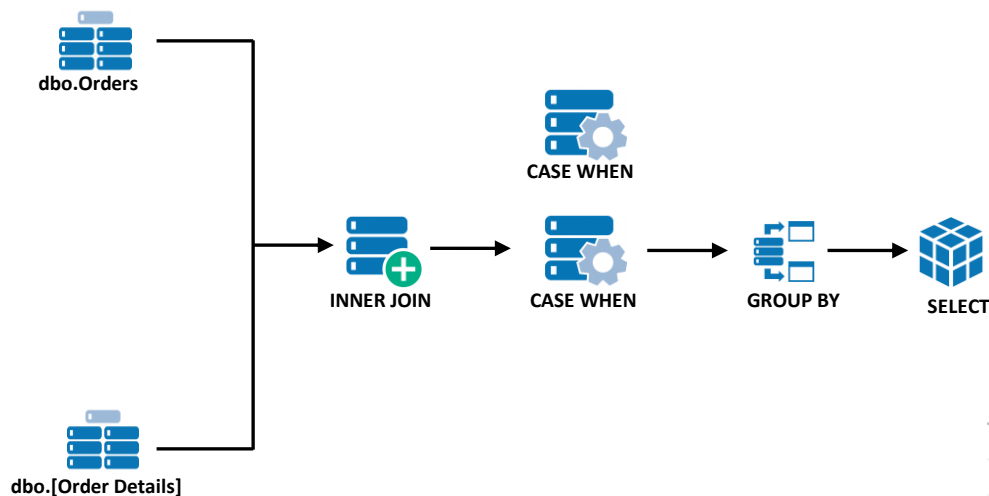
Output

	OrderID	Ship Via	Ship City	Ship Country	Product Sales	Shipper Company	Shipper Category
1	10248	3	Reims	France	27	Federal Shipping	Low
2	10249	1	Münster	Germany	49	Speedy Express	Low
3	10250	2	Rio de Janeiro	Brazil	60	United Package	Middle
4	10251	1	Lyon	France	41	Speedy Express	Low
5	10252	2	Charleroi	Belgium	105	United Package	Top
6	10253	2	Rio de Janeiro	Brazil	102	United Package	Top
7	10254	2	Bern	Switzerland	57	United Package	Middle
8	10255	3	Genève	Switzerland	110	Federal Shipping	Top
9	10256	2	Resende	Brazil	27	United Package	Low
10	10257	3	San Cristóbal	Venezuela	46	Federal Shipping	Low
11	10258	1	Graz	Austria	121	Speedy Express	Top
12	10259	3	México D.F.	Mexico	11	Federal Shipping	Low
13	10260	1	Köln	Germany	102	Speedy Express	Top
14	10261	2	Rio de Janeiro	Brazil	40	United Package	Low

Analysis & Visualization Phase

1. What is the product sales trend in three years?

Trend of Product Sales by Order Date
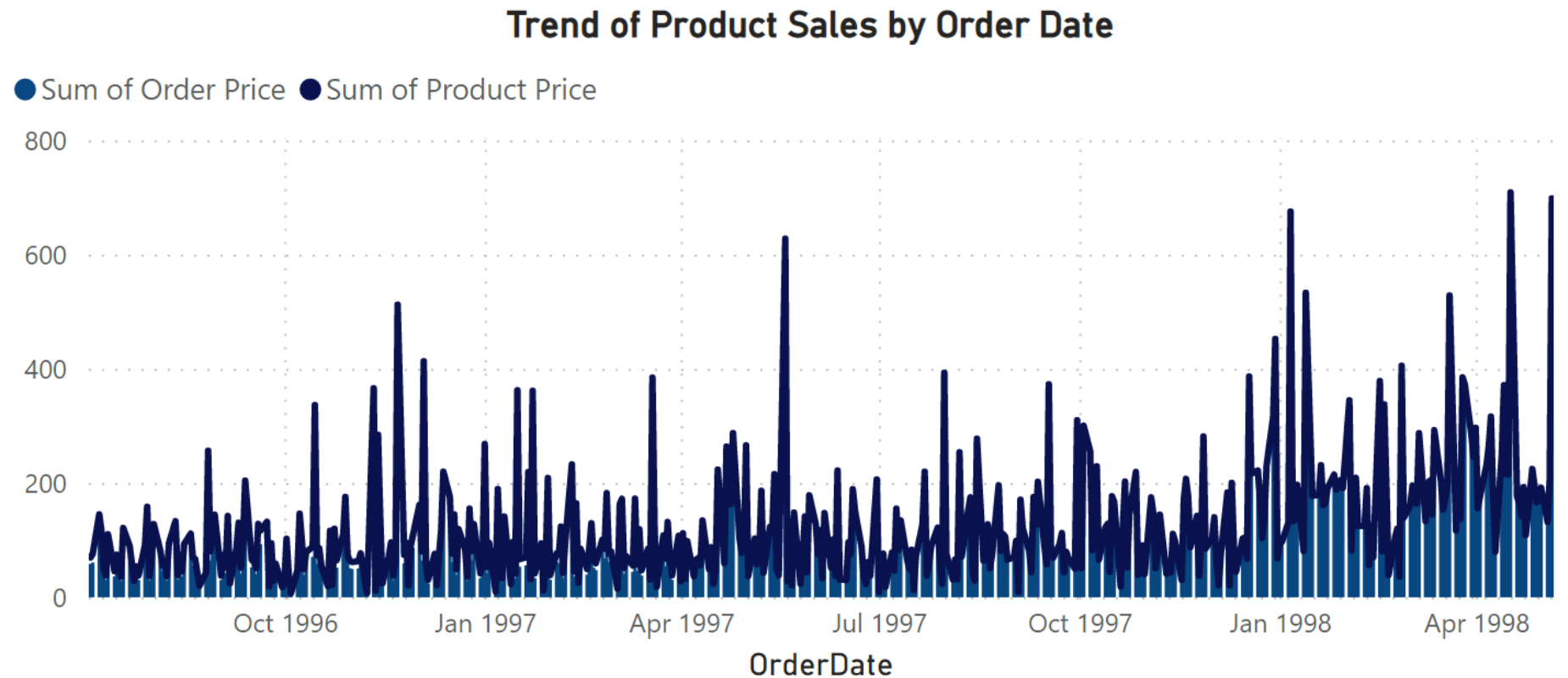
Analysis & Visualization Phase

2. What is the total product sales by category?

Total of Gross Product Sales by Category Names
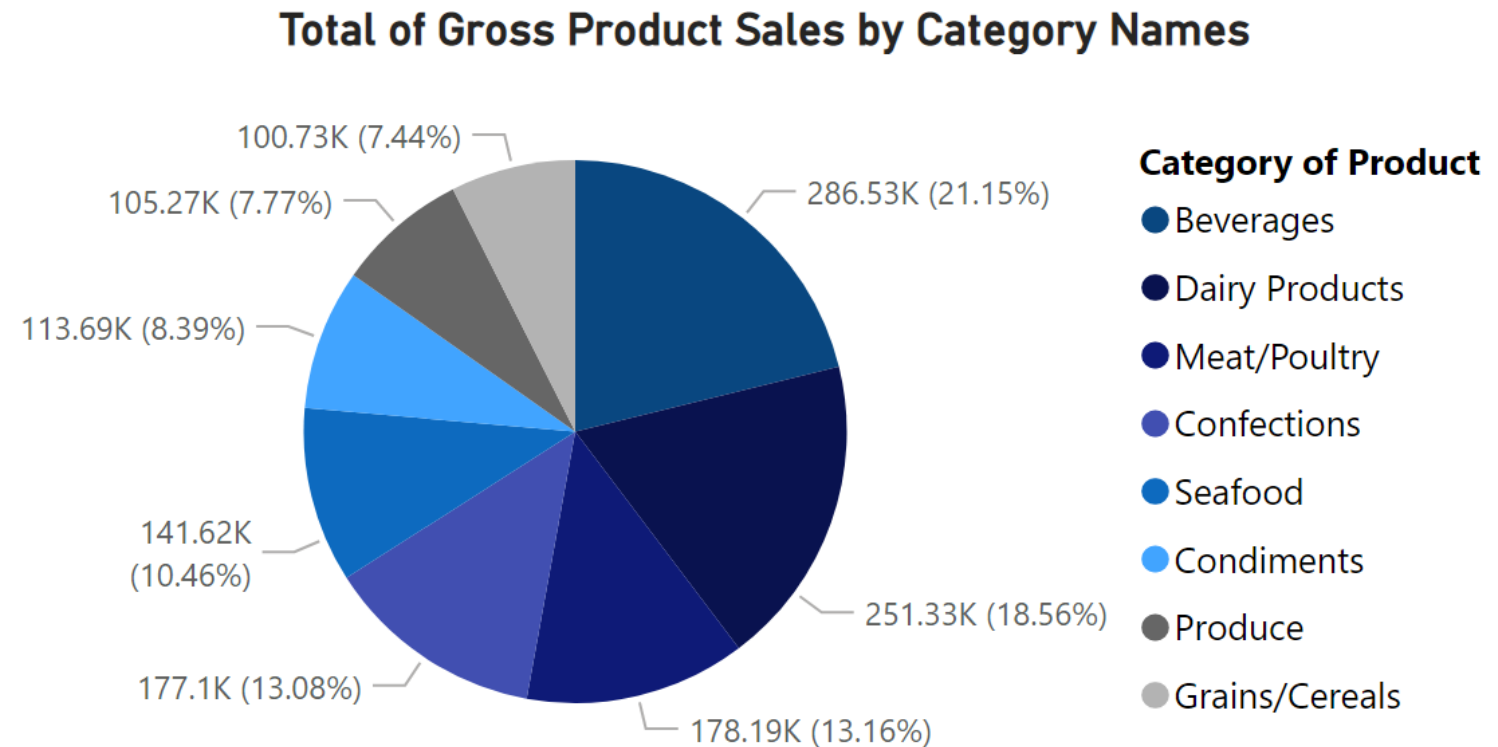
Analysis & Visualization Phase

3. What products are the top sales?

Total Order by Top 10 Product Name
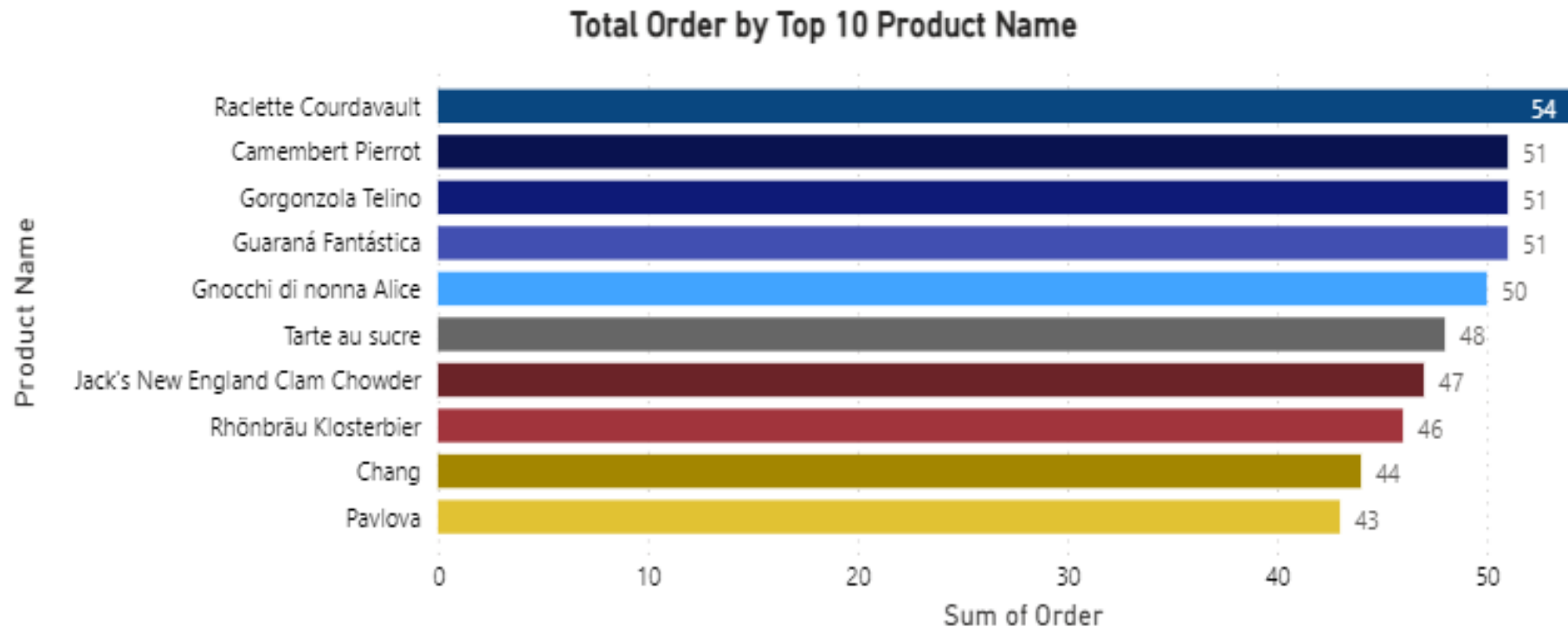
Analysis & Visualization Phase

4. What is the contact title of the customer based on the number of product sales?

Total of Product Sales by Contact Title

Analysis & Visualization Phase

5. How is distribution of customers who place orders in various countries and cities?

Total Order by Country

Analysis & Visualization Phase

6. What is three categories of customers based on the number of product sales? (Top, Middle, Low)

Total Sales by Customer Category

Customer Category ●Low ●Middle ●Top
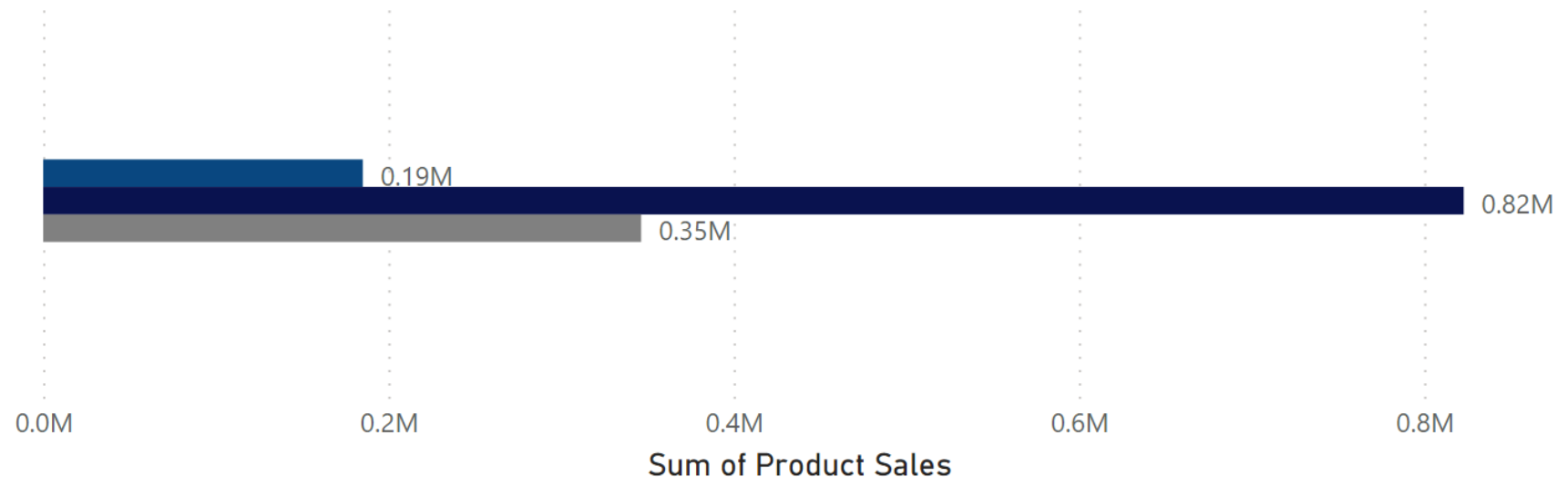
Analysis & Visualization Phase

7. Which Shipper Company has the most orders?

Total Sales by Shipper Company
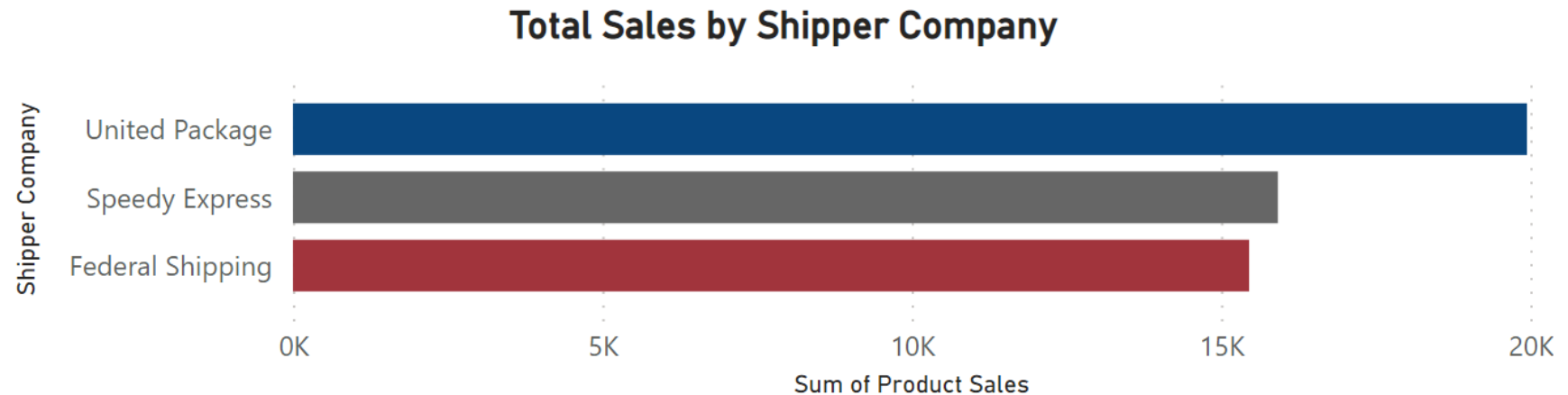
Analysis & Visualization Phase

8. What is the number of product sales based on the three shipper categories? (Top, Middle, Low)

Total Sales by Shipper Category
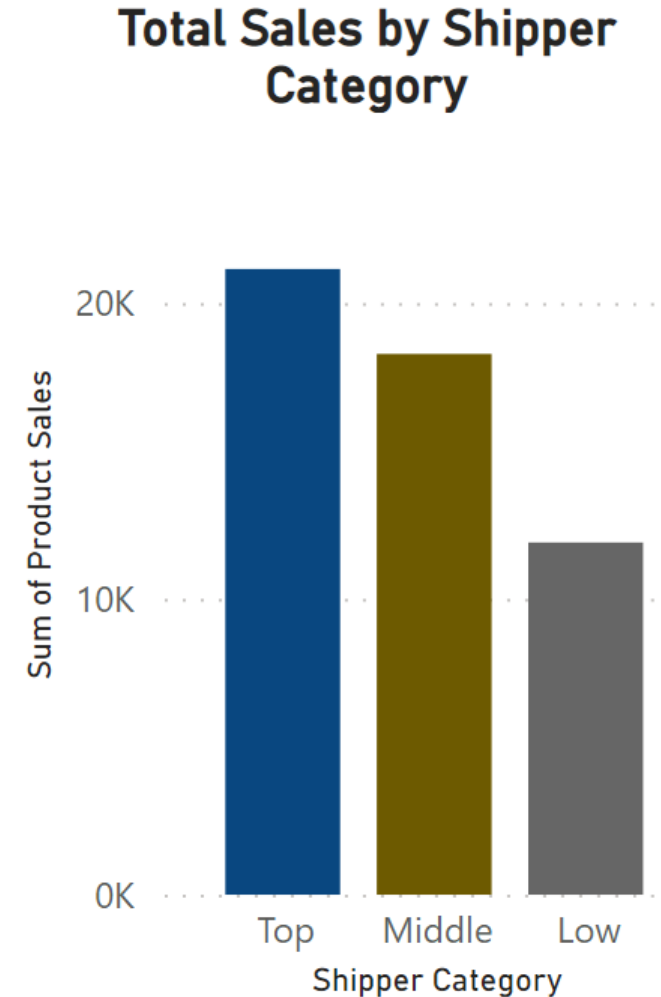
Analysis & Visualization Phase

9. What country has the most orders using the shipper?

Total Sales by Ship City and Ship Country

Act Phase

A. Product Analysis

- Monitor the movement of sales trends every month so they can make the right decision
- Increase products stock for top products and develop and improve promotions for products that are rarely ordered

B. Customer Analysis

- Give discounts for top sales products to customers who have the title of top customer
- Held a big promotion in a country lacking orders

C. Shipper Analysis

- Give award to shipper company with top predicate
- Adding to the shipper company's future work to expand customer reach

Your best quote that reflects your approach… "It's one small step for man, one giant leap for mankind."

- NEIL ARMSTRONG