Data Analysis with MS Power BI

BUSINESS INTELLIGENCE AND DECISION SUPPORT, 2022

Ricardo Pinto (A14622)

Introduction

The current project has the main goal of analyzing a relational database, throughout the different phases from data preparation and modeling to processing and finally data visualization.

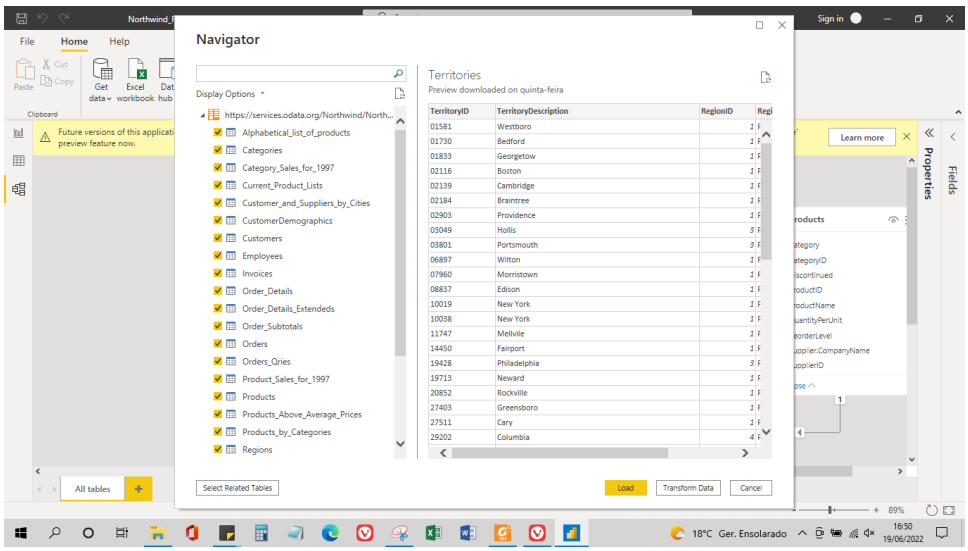
The adopted database is Northwind, which is a database used by Microsoft to demonstrate the features of some of its products, including SQL Server and Microsoft Access. The database contains the sales data for Northwind Traders, a fictitious specialty foods exportimport company.

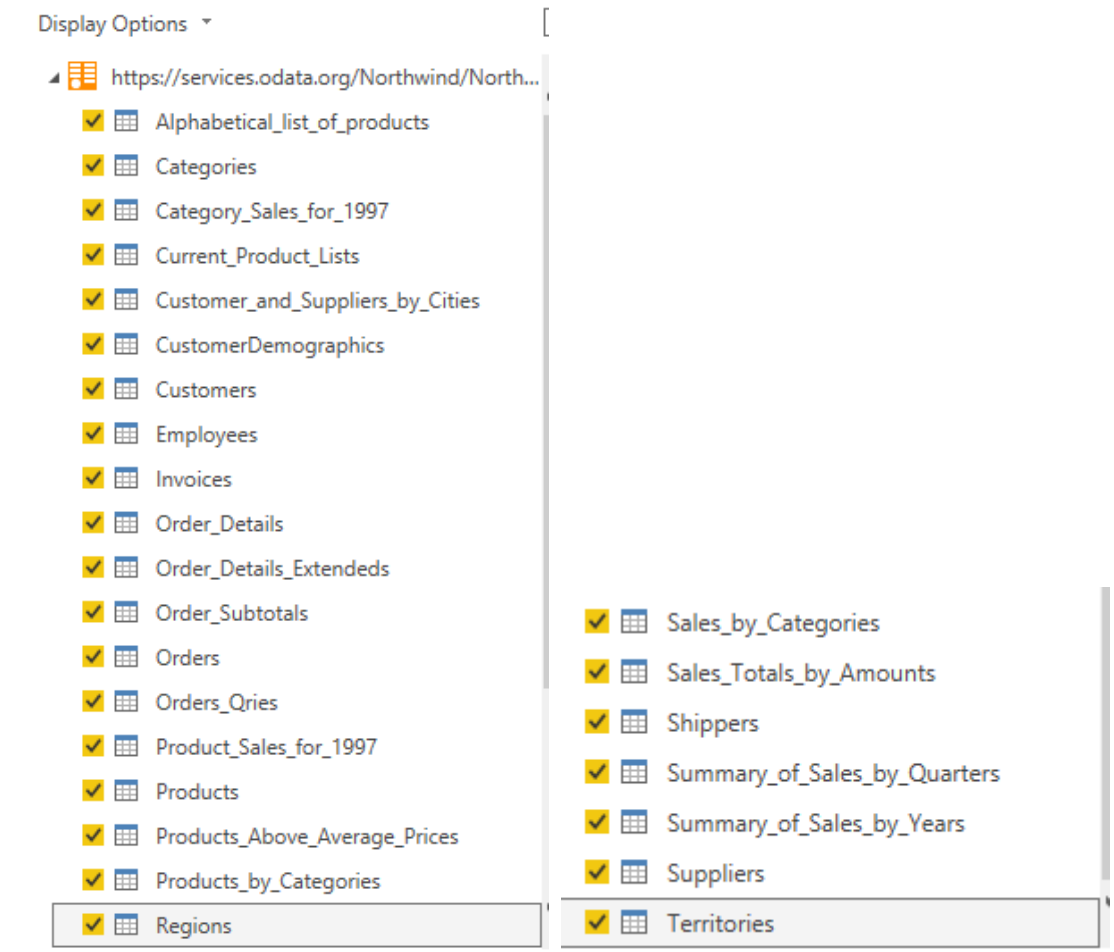
1. Obtaining and Preparing data

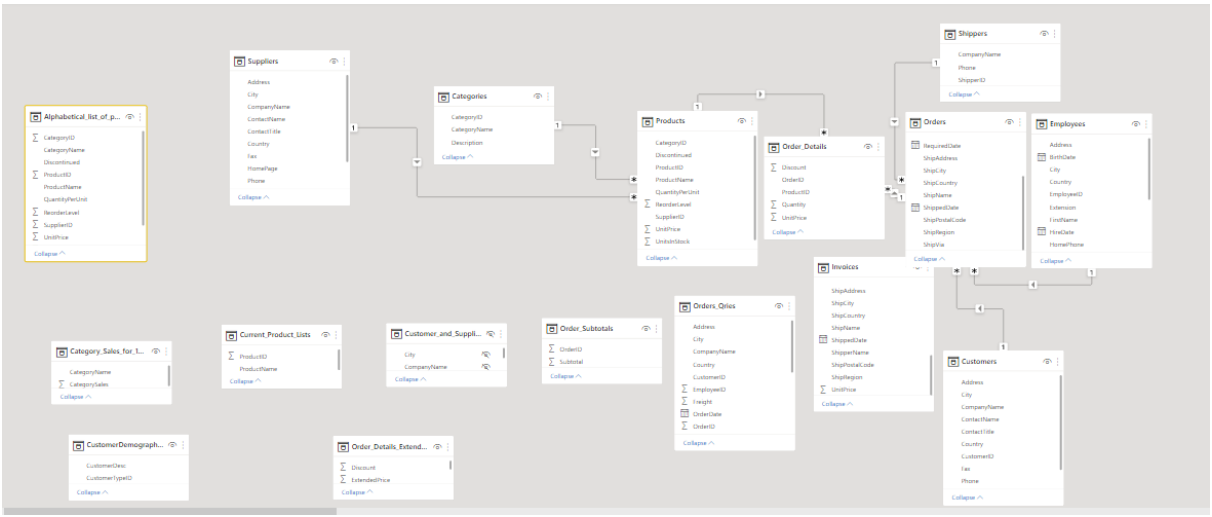
The Northwind database was uploaded through the Odata feed:

https://services.odata.org/Northwind/Northwind.svc/

All tables have been loaded:




Display Options
https://services.odata.org/Northwind/North...
Alphabetical_list_of_products
Categories
Category_Sales_for_1997
Current_Product_Lists
Customer_and_Suppliers_by_Cities
CustomerDemographics
Customers
Employees
Invoices
Order_Details
Order_Details_Extendeds
Order_Subtotals
Orders
Orders_Qries
Product_Sales_for_1997
Products
Products_Above_Average_Prices
Products_by_Categories
Regions
Sales_by_Categories
Sales_Totals_by_Amounts
Shippers
Summary_of_Sales_by_Quarters
Summary_of_Sales_by_Years
Suppliers
Territories



2. Data modeling and processing

2.1. MODELING

The following expansions were made in the Products Table:

Keys	Table	Merged	New Columns
CategoryID	Products	Categories	Category Name
Supplier ID	Products	Supplier Name	Supplier Company Name

Redundant or not relevant tables were eliminated for the analysis purpose of this project.

The Fact Table is the Invoices Table

As such, InnerJoin was made between the Orders table and the Invoices Table (Fact Table) through the keys of the table below and an active dependency was created between the Customers / Invoices and Employees / Invoices tables

Foreign Keys	Source Table	Fact Table
EmployeeID	Orders	Invoices
CustomerID	Orders	Invoices

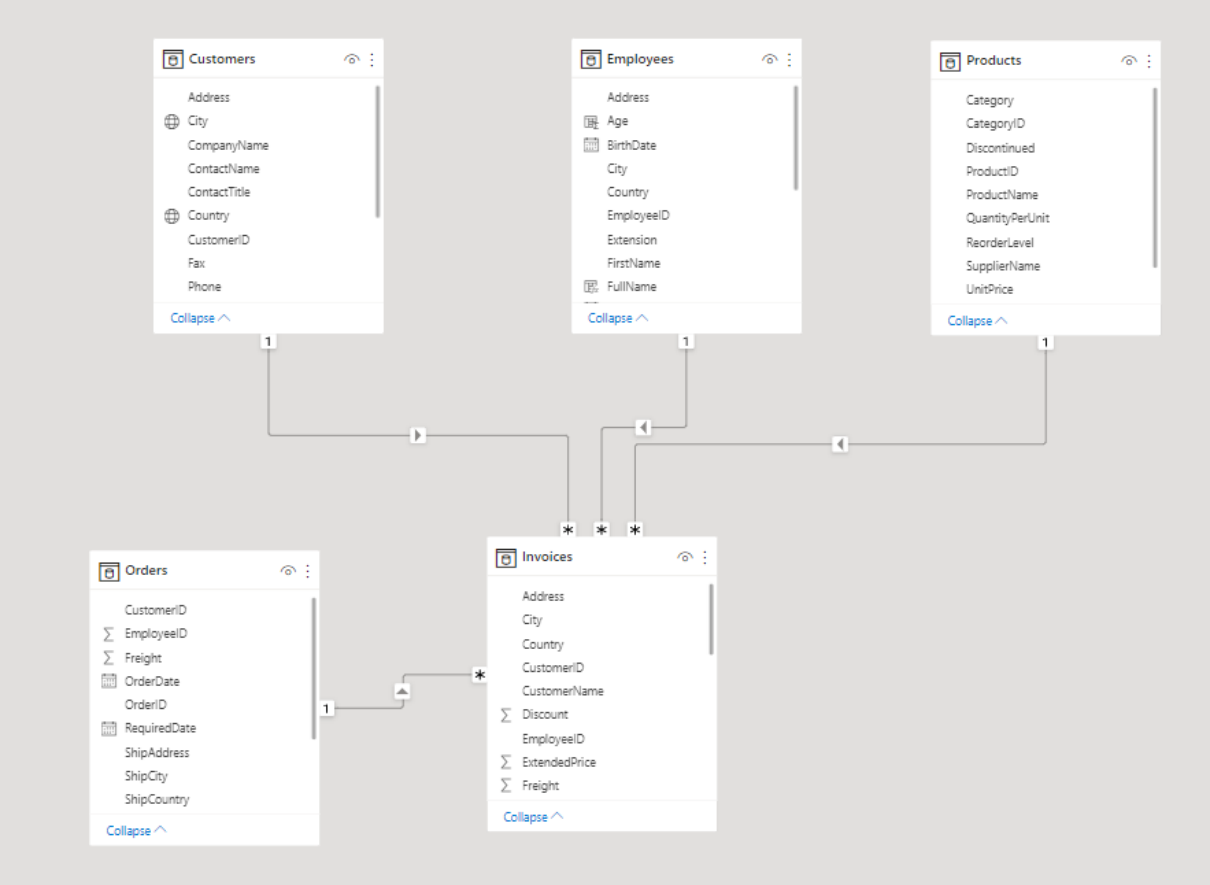
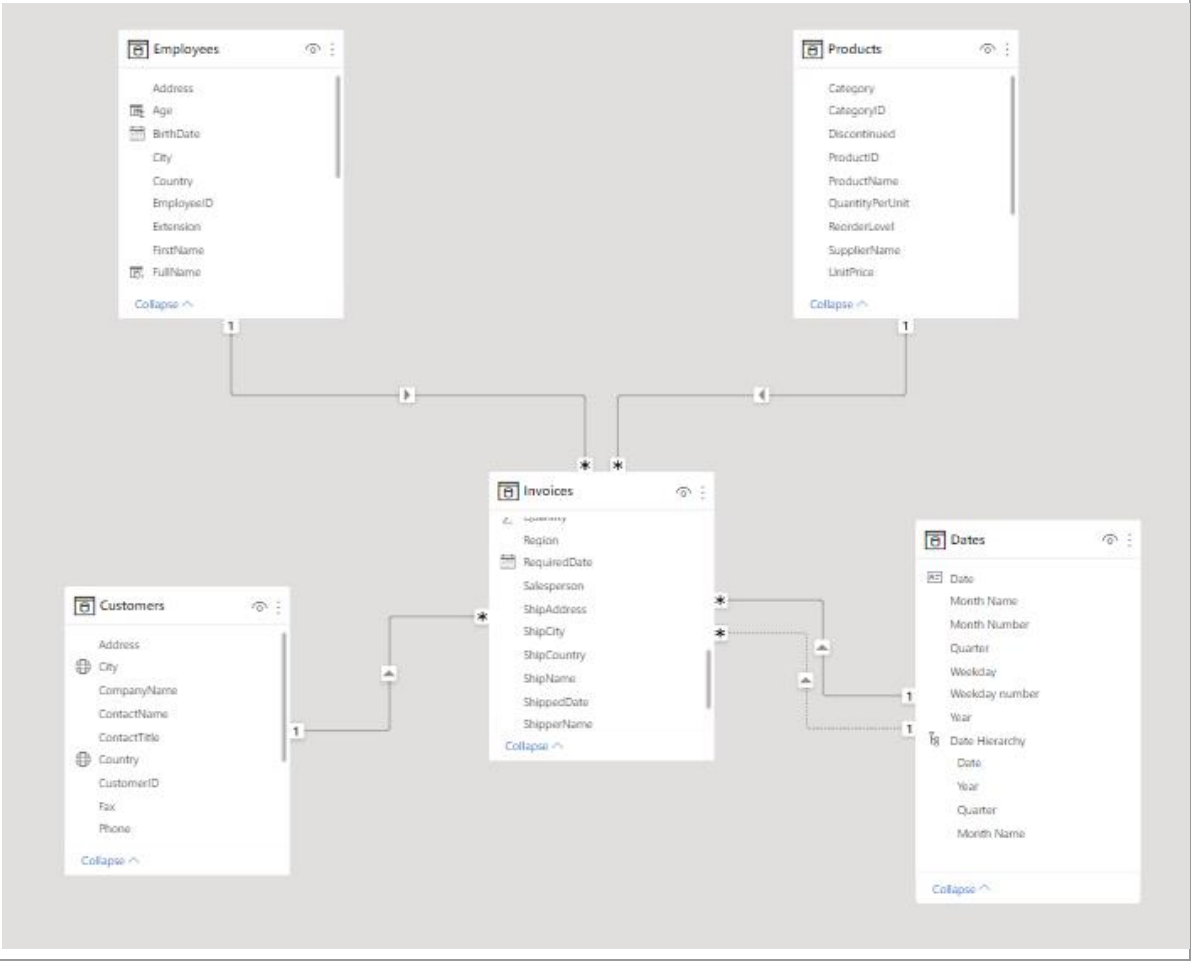


Table "Orders" has replicated attributes and keys from the "Invoices" Table. Therefore we decided to delete it.

The "Dates" Table was added as described in the following section 2.2. We've reach then the final final scheme below:

Adopted final Star Schema



Active Relationships

Active	From: Table (Column)	To: Table (Column)
☑	Invoices (CustomerID)	Customers (CustomerID)
☑	Invoices (EmployeeID)	Employees (EmployeeID)
☑	Invoices (OrderDate)	Dates (Date)
☑	Invoices (ProductID)	Products (ProductID)
☐	Invoices (ShippedDate)	Dates (Date)

2.2. NEW TABLES

Dates

The "Dates" table with date history (from oldest to most recent) was added as a dimensional table to connect into the OrderDate and ShippedDate column of the Invoices table.

High Value Sales

We intended here to divide the product sales amount by classes (low, medium, high and very high)

GDP

We've focus on integrating different data sources, and so an external table with GDP per capita for all countries was integrated with the main database Northwind.

https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD&country=

Task: to Upload the Countries GDP table in PowerBI and GDP expansion by country in 1997 in the "Invoices" table.

2.3. NEW COLUMNS

Invoices

Gross Sales = Invoices[UnitPrice] * Invoices[Quantity]

```
Net Sales = Invoices[Gross Sales] - Invoices[Discount$]

Discount$ = Invoices[Gross Renevue] * Invoices[Discount%]

Days to Ship = Invoices[ShippedDate]-Invoices[OrderDate]

End of month Date =
EOMONTH (Invoices[OrderDate], 0)
```

Employees

```
FullName = Employees[FirstName] & " & Employees[LastName]

Age = int(YEARFRAC(Employees[BirthDate], TODAY(), 1))
```

2.4. NEW MEASURES

MainMeasures

```
CountOrders = DISTINCTCOUNT(Invoices[OrderID])

CustomerRank = RANKX(ALL(Customers), [TotalAmount])

Discount$ YTD =
TOTALYTD(SUM('Invoices'[Discount$]), 'Dates'[Date])

Net Sales = YTD
TOTALYTD(SUM('Invoices'[Net Sales]), 'Dates'[Date])

Quantity = YTD
TOTALYTD(SUM('Invoices'[Quantity]), 'Dates'[Date])

TotalAmount = ROUND(SUMX(Invoices, Invoices[UnitPrice]*Invoices[Quantity]), 2)

TotalAmount = YTD
```

```
TOTALYTD([TotalAmount], 'Dates'[Date])
```

```
TotalDiscount = SUMX(Invoices, Invoices[Quantity] * Invoices[UnitPrice] *  Invoices[Discount%])
```

```
TotalQty = SUM(Invoices[Quantity])
```

Time Intelligence

PrevM_Amount

TotalAmount MoM%

TotalAmount LY

ChangeYoY

Shippersmeasures

CountShippedOrders

2.5. REMAINING TRANSFORMATIONS

Customers City and Country geographically categorized

Countries grouped by Continent

Month Name ordered as Month Number

Year Hierarchy created

Country Hierarchy created

3. Data visualization

The general purpose is to characterize the main KPIs of the business to consulted by the Company Management and divided into 4 main categories:

- Main (Financial);

- Products;

- Orders;

- Customers;

- TotalAmount Influencers (supplementary information).

On the Products, Orders, and Customers pages it is possible through drop-down slicers to filter by:

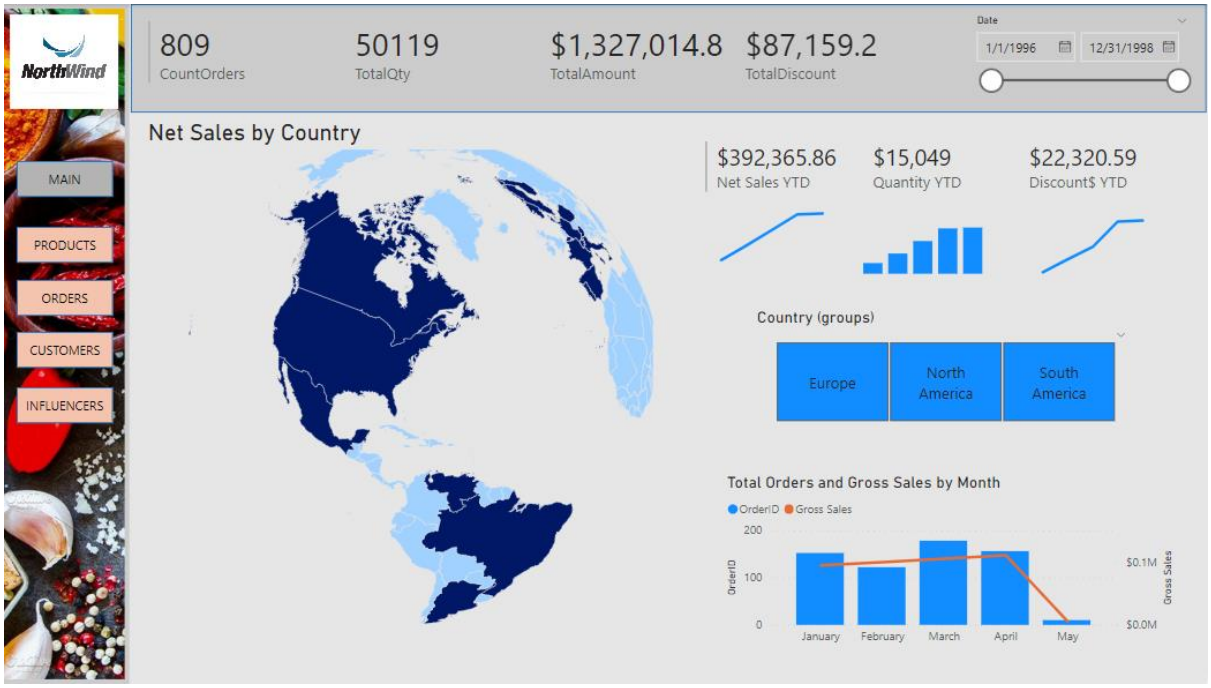
- Product & Category;

- City, Country & Continent;

- Customer Name;

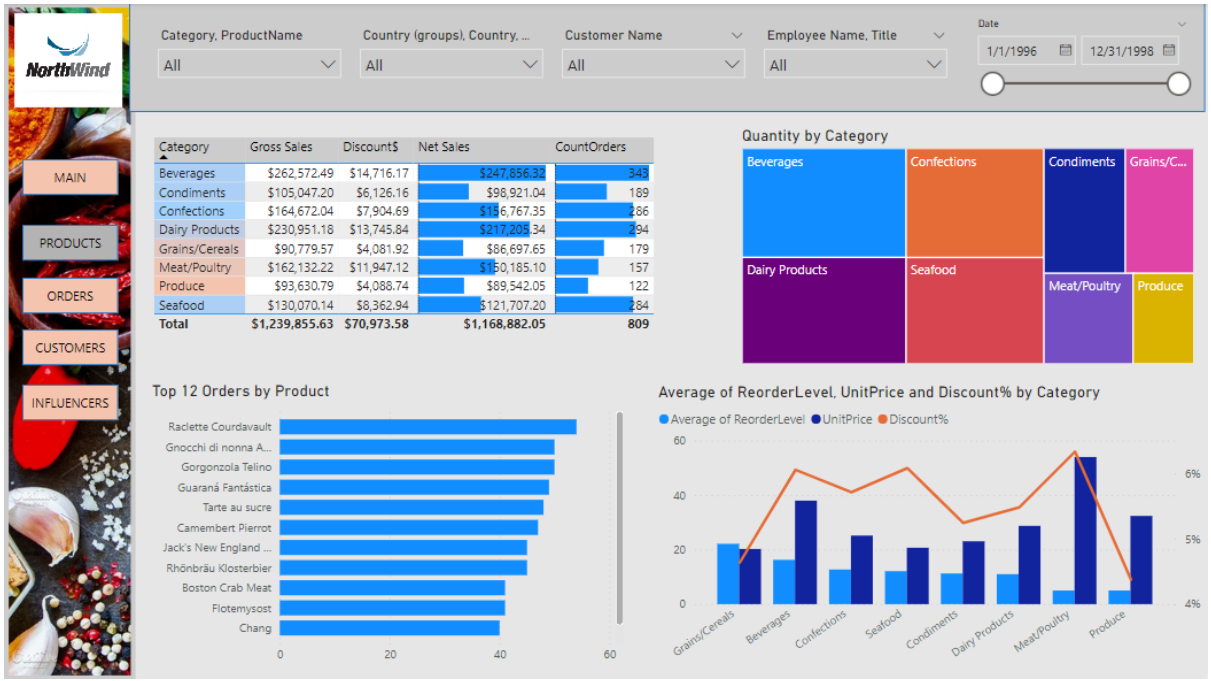
- Employee Name & Title.

3.1. MAIN



The main objective is to provide a general reading of KPIs (especially financial), focusing on the current year (YTD) and by geolocation.

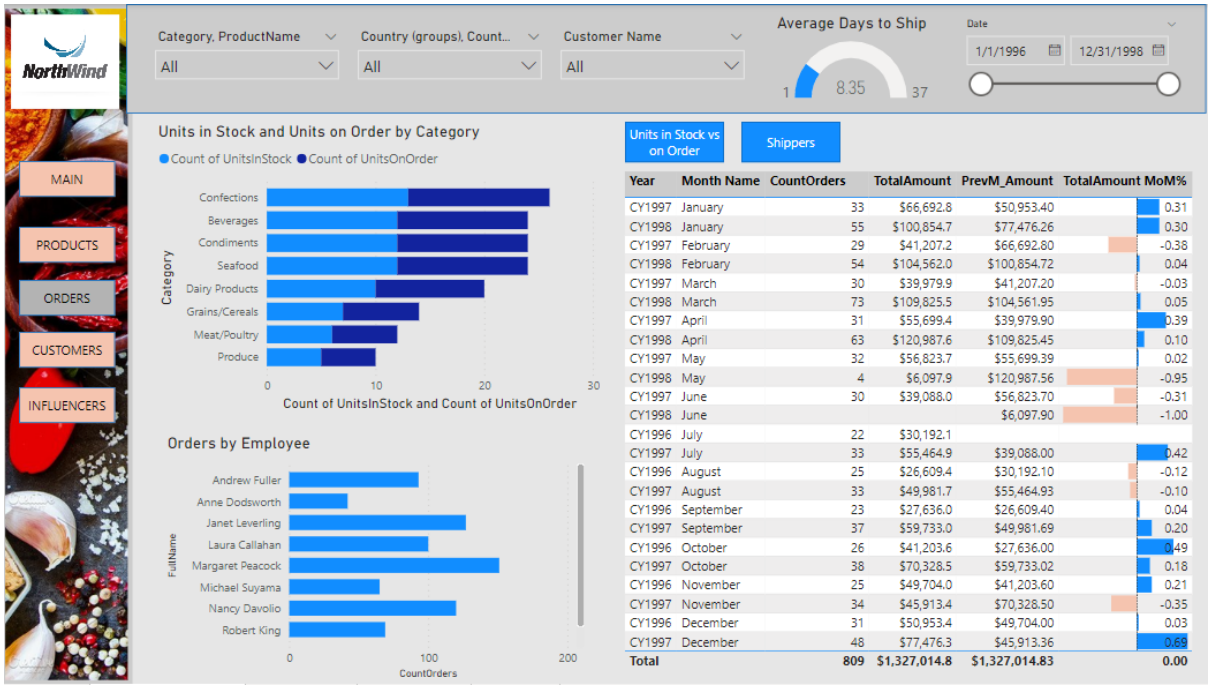
3.2. PRODUCTS



In this page, one can query the products with higher volume by name and sales by category.

Also is visible Information of the reorder level by categories and their relationship with unit price and discount.

3.3. ORDERS



Year	Month Name	CountOrders	TotalAmount	PrevM_Amount	TotalAmount MoM%
CY1997	January	33	$66,692.8	$50,953.40	0.31
CY1998	January	55	$100,854.7	$77,476.26	0.30
CY1997	February	29	$41,207.2	$66,692.80	-0.38
CY1998	February	54	$104,562.0	$100,854.72	0.04
CY1997	March	30	$39,979.9	$41,207.20	-0.03
CY1998	March	73	$109,825.5	$104,561.95	0.05
CY1997	April	31	$55,699.4	$39,979.90	0.39
CY1998	April	63	$120,987.6	$109,825.45	0.10
CY1997	May	32	$56,823.7	$55,699.39	0.02
CY1998	May	4	$6,097.9	$120,987.56	-0.95
CY1997	June	30	$39,088.0	$56,823.70	-0.31
CY1998	June			$6,097.90	-1.00
CY1996	July	22	$30,192.1		
CY1997	July	33	$55,464.9	$39,088.00	0.42
CY1996	August	25	$26,609.4	$30,192.10	-0.12
CY1997	August	33	$49,981.7	$55,464.93	-0.10
CY1996	September	23	$27,636.0	$26,609.40	0.04
CY1997	September	37	$59,733.0	$49,981.69	0.20
CY1996	October	26	$41,203.6	$27,636.00	0.49
CY1997	October	38	$70,328.5	$59,733.02	0.18
CY1996	November	25	$49,704.0	$41,203.60	0.21
CY1997	November	34	$45,913.4	$70,328.50	-0.35
CY1996	December	31	$50,953.4	$49,704.00	0.03
CY1997	December	48	$77,476.3	$45,913.36	0.69
Total		**809**	**$1,327,014.8**	**$1,327,014.83**	**0.00**

This page concerns:

Bookmarks created for 2 charts:

- Units in Stock and Units on Order;

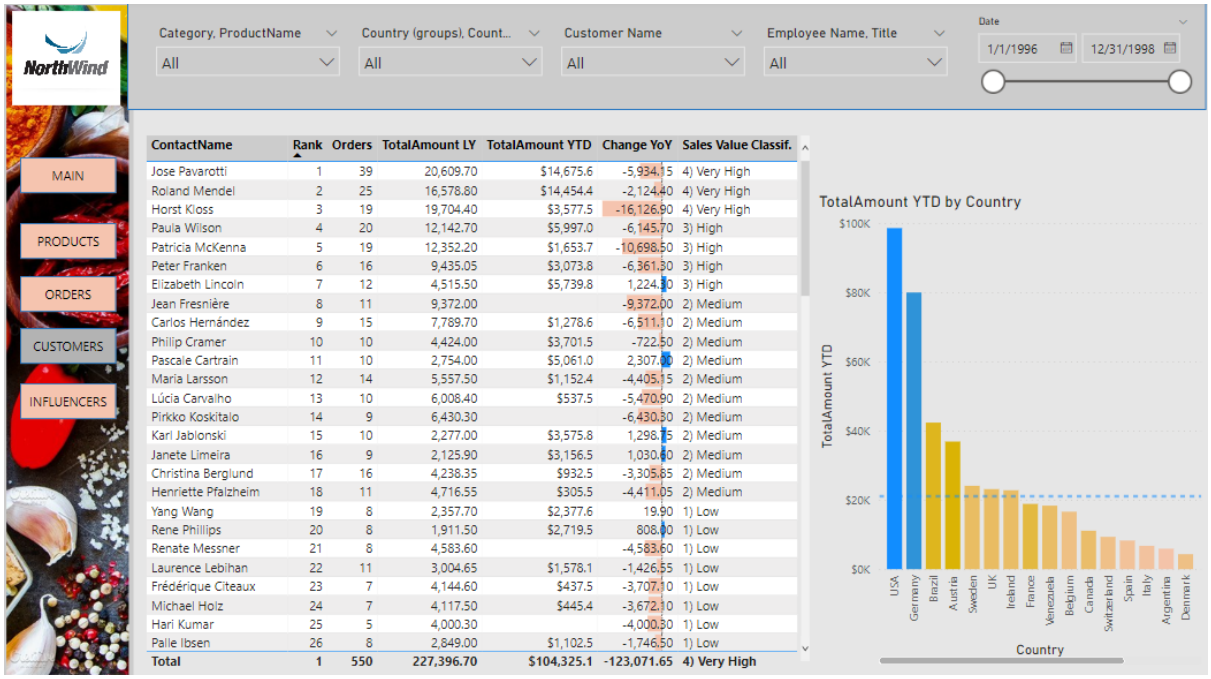
- Shippers.

Comparison between units on stock and on order by category.

Comparison between shippers: average delivery time, freight and shipped orders.

Table of orders and sales values over time and variation over the previous month.

3.4. CUSTOMERS



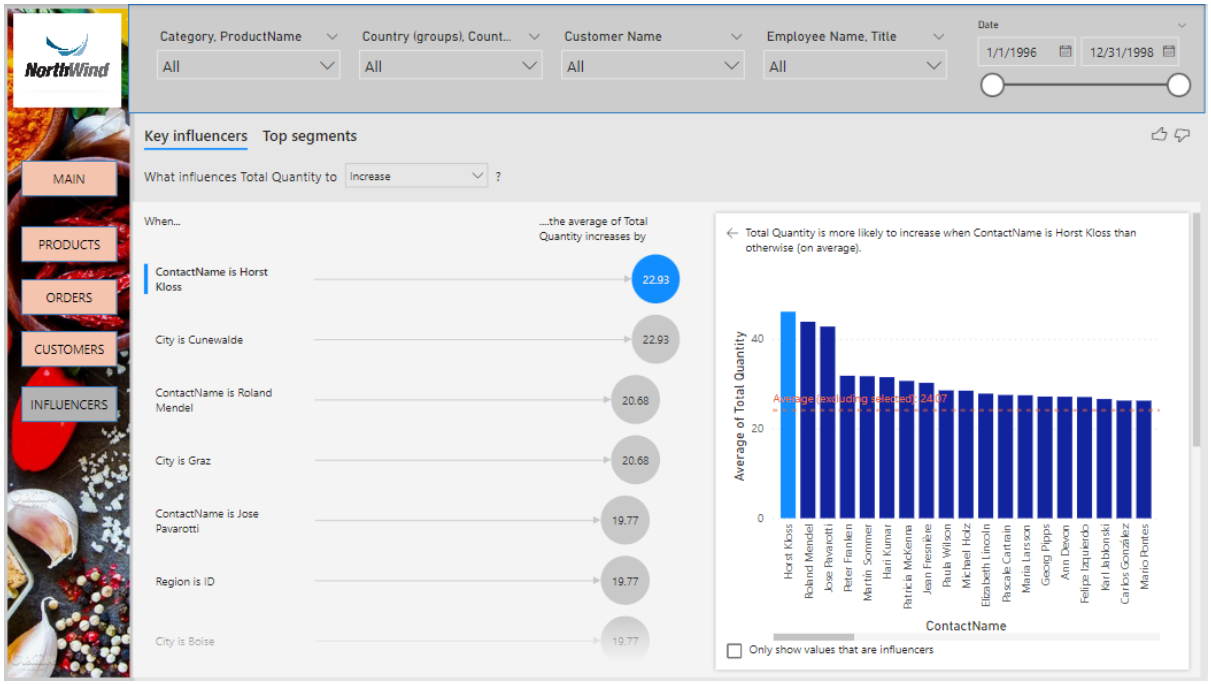
ContactName	Rank	Orders	TotalAmount LY	TotalAmount YTD	Change YoY	Sales Value Classif.
Jose Pavarotti	1	39	20,609.70	$14,675.6	-5,934.15	4) Very High
Roland Mendel	2	25	16,578.80	$14,454.4	-2,124.40	4) Very High
Horst Kloss	3	19	19,704.40	$3,577.5	-16,126.90	4) Very High
Paula Wilson	4	20	12,142.70	$5,997.0	-6,145.70	3) High
Patricia McKenna	5	19	12,352.20	$1,653.7	-10,698.50	3) High
Peter Franken	6	16	9,435.05	$3,073.8	-6,361.30	3) High
Elizabeth Lincoln	7	12	4,515.50	$5,739.8	1,224.30	3) High
Jean Fresnière	8	11	9,372.00		-9,372.00	2) Medium
Carlos Hernández	9	15	7,789.70	$1,278.6	-6,511.10	2) Medium
Philip Cramer	10	10	4,424.00	$3,701.5	-722.50	2) Medium
Pascale Cartrain	11	10	2,754.00	$5,061.0	2,307.00	2) Medium
Maria Larsson	12	14	5,557.50	$1,152.4	-4,405.15	2) Medium
Lúcia Carvalho	13	10	6,008.40	$537.5	-5,470.90	2) Medium
Pirkko Koskitalo	14	9	6,430.30		-6,430.30	2) Medium
Karl Jablonski	15	10	2,277.00	$3,575.8	1,298.75	2) Medium
Janete Limeira	16	9	2,125.90	$3,156.5	1,030.60	2) Medium
Christina Berglund	17	16	4,238.35	$932.5	-3,305.85	2) Medium
Henriette Pfalzheim	18	11	4,716.55	$305.5	-4,411.05	2) Medium
Yang Wang	19	8	2,357.70	$2,377.6	19.90	1) Low
Rene Phillips	20	8	1,911.50	$2,719.5	808.00	1) Low
Renate Messner	21	8	4,583.60		-4,583.60	1) Low
Laurence Lebihan	22	11	3,004.65	$1,578.1	-1,426.55	1) Low
Frédérique Citeaux	23	7	4,144.60	$437.5	-3,707.10	1) Low
Michael Holz	24	7	4,117.50	$445.4	-3,672.10	1) Low
Hari Kumar	25	5	4,000.30		-4,000.30	1) Low
Palle Ibsen	26	8	2,849.00	$1,102.5	-1,746.50	1) Low
Total	**1**	**550**	**227,396.70**	**$104,325.1**	**-123,071.65**	**4) Very High**

Information regarding customers ordered by raking and comparison between sales last year vs current year.

Linking of customers sales to countries of origin through the country sales chart.

Possibility to compare sales by country with their respective GDP per capita (Tooltip with GDP countries activated in the chart).

3.5. INFLUENCERS



Influencers provide additional information on

- the variables that most impact the total quantity of sold products ("TotalQty");
- the relationships between independent and dependent variables;
- data related to trends and patterns.

For example, Total Quantity increases by 0.98 when GDP grows $9551.15

Conclusion

With the current database it was possible to make an analysis with some level of detail of sales over time, through its different attributes such as products and category, destination countries and employees.

The visualizations also provided several insights related to other important variables such as Orders, Shipping and Customers.

Bibliography

UC Business Intelligence classes and Decision Support Systems, PGDAB, 2022.

https://www.powerbitraining.com.au/advanced-scenarios-with-dax-studio-power-bi-advanced-dax-scenarios-with-dax-studio/

https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD&country=