**ICA**

**BIG DATA AND BUSINESS INTELLIGENCE**

**NAME: [PUT NAME HERE]**

**STUDENT ID: [PUT STUDENT ID HERE]**

**Table of Contents**

**Section 1: Business Intelligence Design**

**1 Dataset Description and Business Intelligence Question**

1.1 Data Source Description

1.2 Reason for choosing the dataset

1.3 Dataset Description

1.4 Loading Data into Power BI

1.5 Business Intelligence Requirements

**2 Data pre-processing and Data Cleansing**

2.1 Loading Data into Query Editor for transformation

2.2 Pre-processing and Cleansing Data for Analysis

**3 BI Data Modelling**

3.1 Adding Calendar Table to the Model

**Section 2: Business Report**

1. Introduction

2. Key Findings

3. Conclusion

4. Recommendations

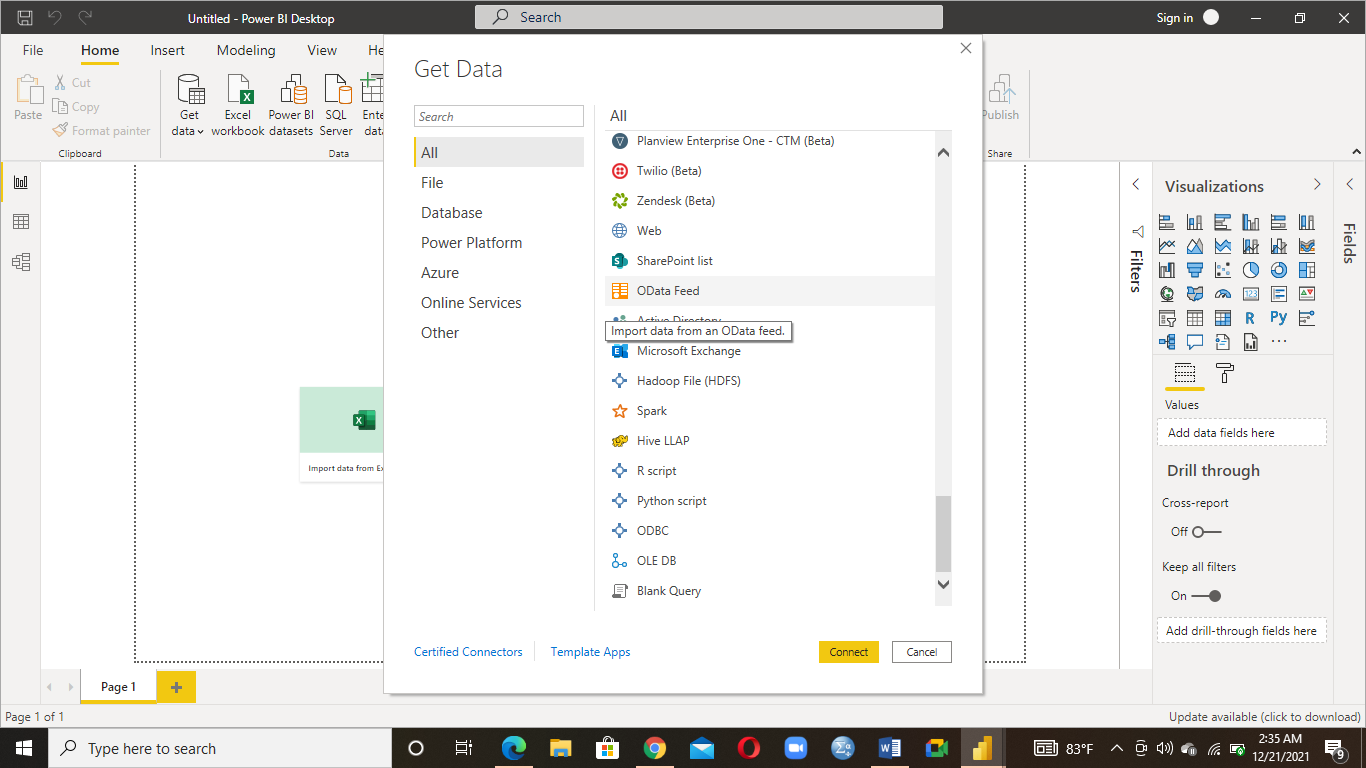
**1.1 Data Source Description**

In this Power BI project, we will make use of the Northwind dataset accessible from the OData Feed source of Power BI. To locate it, we clicked on **Get Data** then **OData Feed**. A url is provided to access the data and is entered in the box displayed after selecting **OData Feed**.

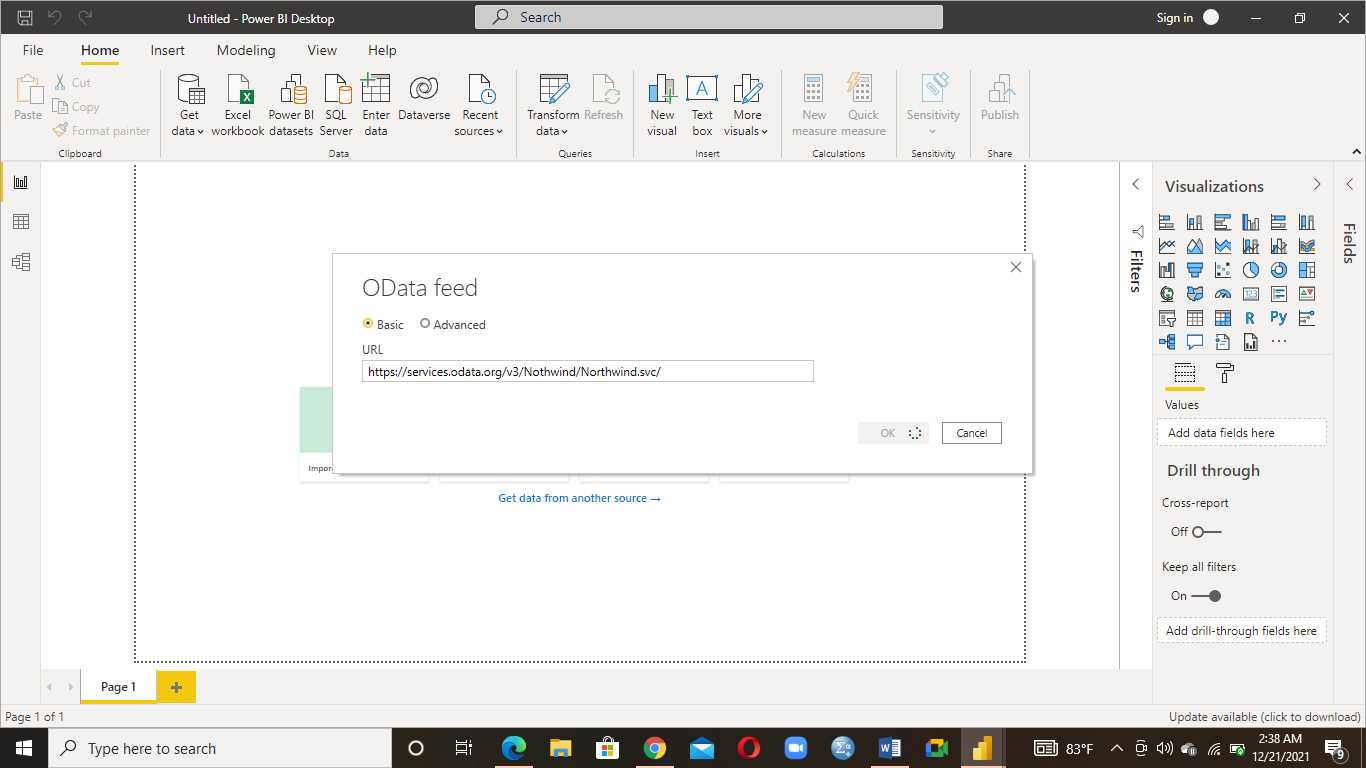
The url is:

<https://services.odata.org/V3/Northwind/Northwind.svc/>

The database is made available by Microsoft Corporation for use as practices in a variety of database products for many years. It contains the data on sales made by a fictitious company called Northwind Traders. The company specialises in imports and exports of specialty food. The database has historically been used in both Microsoft and non-Microsoft database management software.



**Fig 1: Screenshot showing the OData Feed source being selected**

****

**Figure 2: Screenshot of the url used to access the data**

**1.2 Reason for Choosing the Dataset**

The choice of the dataset is informed by fact that it contains the important attributes which can be used to achieve our aims in this project. The business requirements of this assignment will be met by using the attributes obtainable from the data. The database will allow us meet our business intelligence requirements adequately.

**1.3 Dataset Description**

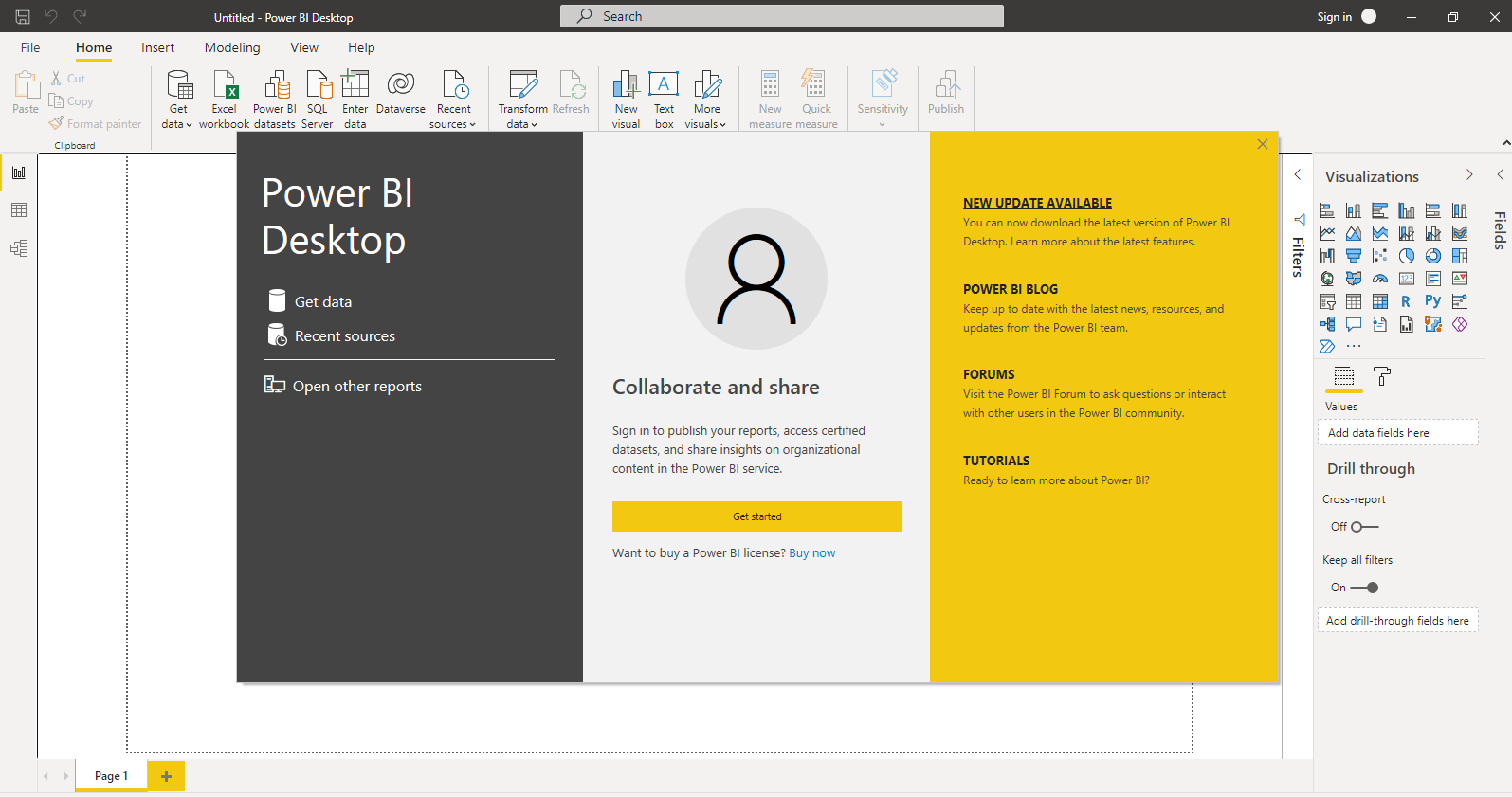
5 tables were selected from the Northwind database for use in this assignment. The tables are order\_details, orders, employees, shippers, customers. They are described in turns in the table below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table Name** | | **Number of Columns** | | | | **Number of Rows** | **Name of columns** | **Description of the columns** |
| Order\_details | | 7 | | | | 2155 | OrderID | Order ID |
| ProductID | Product ID |
| UnitPrice | The unit price at which the product is sold |
| Quantity | Quantity ordered |
| Discount | Amount of discount |
| Order | Order made |
| Product | Product ordered |
| Orders | | 18 | | | | 830 | OrderID | Order ID |
| CustomerID | ID of the customers |
| EmployeeID | ID of the employees |
| OrderDate | Date the orderis made |
| RequiredDate | The date the order is required |
| ShippedDate | The date the order is shipped |
| ShippedVia | Medium through which the order is shipped |
| Freight | The amount of freight |
| ShipName | Name of the ship |
| ShipAddress | Address of the ship |
| ShipCity | City of the ship |
| ShipRegion | Region of the ship |
| ShipPostalCode | Postal code of the ship |
|  |  | | | |  | | ShipCountry | Country of the ship |
| Customer | Name of the customers |
| Employee | Name of the employee in charge of the order |
| Order\_details | The details of the order |
| Shipper | The name of the company that shipped the order |
| Employees | 22 | | 9 | | | | EmployeeID | ID of the employee |
| LastName | The last name of the employees |
| FirstName | The first name of the employees |
| Title | Title of the employees |
| TitleofCourtesy | Title of courtesy |
|  |  | | |  | | | BirthDate | Date of birth of the employees |
| HireDate | The date the employee was hired |
| Address | Address of the employee |
| City | City where the employees reside |
|  | |  | | | |  | Region | Region where the employee resides |
| PostalCode | Postal codes of the employees’ residence |
| Country | Country where the employee comes from |
| HomePhone | Home phone f the employee |
| Photo | Photo of the employee |
| Notes | Notes about the employees |
| ReportsTo | The person the employee reports to |
| PhotoPath | Path of the photo |
| Employees1 | Employees |
| Employee | employee |
| Orders | Orders handled by the emloyees |
| Territories | Territories of the employees |
| Shippers | | 4 | | | | 3 | ShipperID | ID of the Shipper |
| CompanyName | Name of the company |
| Phone | Phone |
| Customers | | 13 | | | | 91 | CustomerID | The unique ID assigned to the customers |
| CompanyName | Name of the company |
| ContactName | Name through which the customers could be contacted |
| ContactTitle | Title of the contact |
| Address | Address of the customers |
|  | |  | | | |  | City | City of the customers |
|  | |  | | | |  | Region | Region of the customers |
|  | |  | | | |  | PostalCode | Postal code |
|  | |  | | | |  | Country | Country of the customers |
|  | |  | | | |  | Phone | Phone of the customer |
|  | |  | | | |  | Fax | Fax of the customer |
|  | |  | | | |  | Orders | Orders made by the customer |
|  | |  | | | |  | CustomerDemographics | Demographics of the customers |

**1.4** **Loading Data into Power BI Environment**

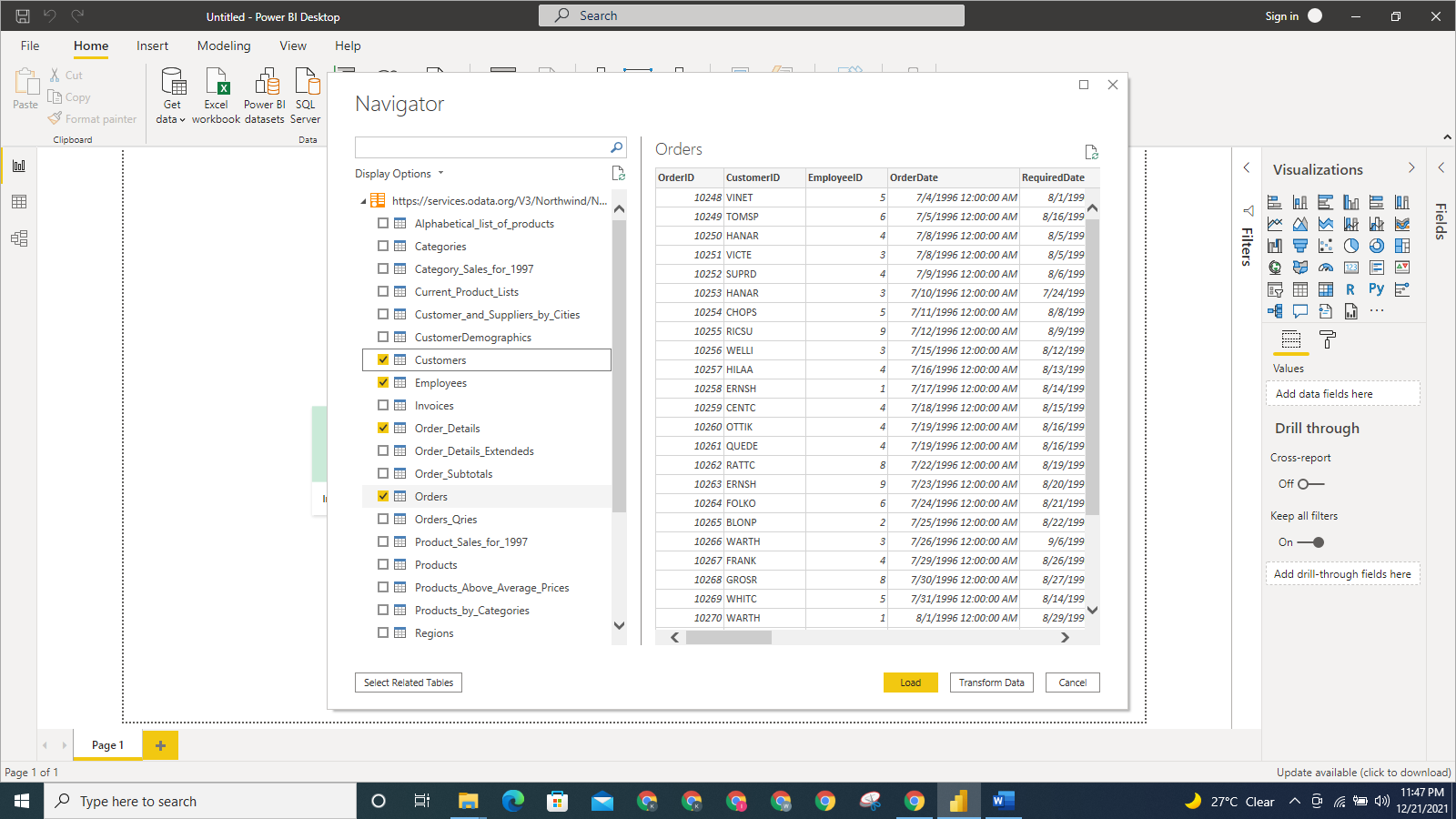
To begin working with the chosen dataset in Power BI, there is a need to load it from its source into the Power BI environment. We can begin working with the data, creating useful visuals and reports, if there is no pre-processing or cleansing to be done. There are many sources from which we can get data into Power BI. These sources include Text/CSV, Web, MS Access, Salesforce, OData Feed, GitHub, Oracle, SQL Server, Excel, Azure SQL Database, BI Connector, Hadoop File (HDFS), Salesforce Reports, PDF, Twilio, etc. For this assignment, we decided to pull data from OData Feed using the url provided in the assignment brief. The data loaded into Power BI from OData Feed is the Northwind database.

The screenshot below shows the Power BI screen when the business intelligence software is opened.

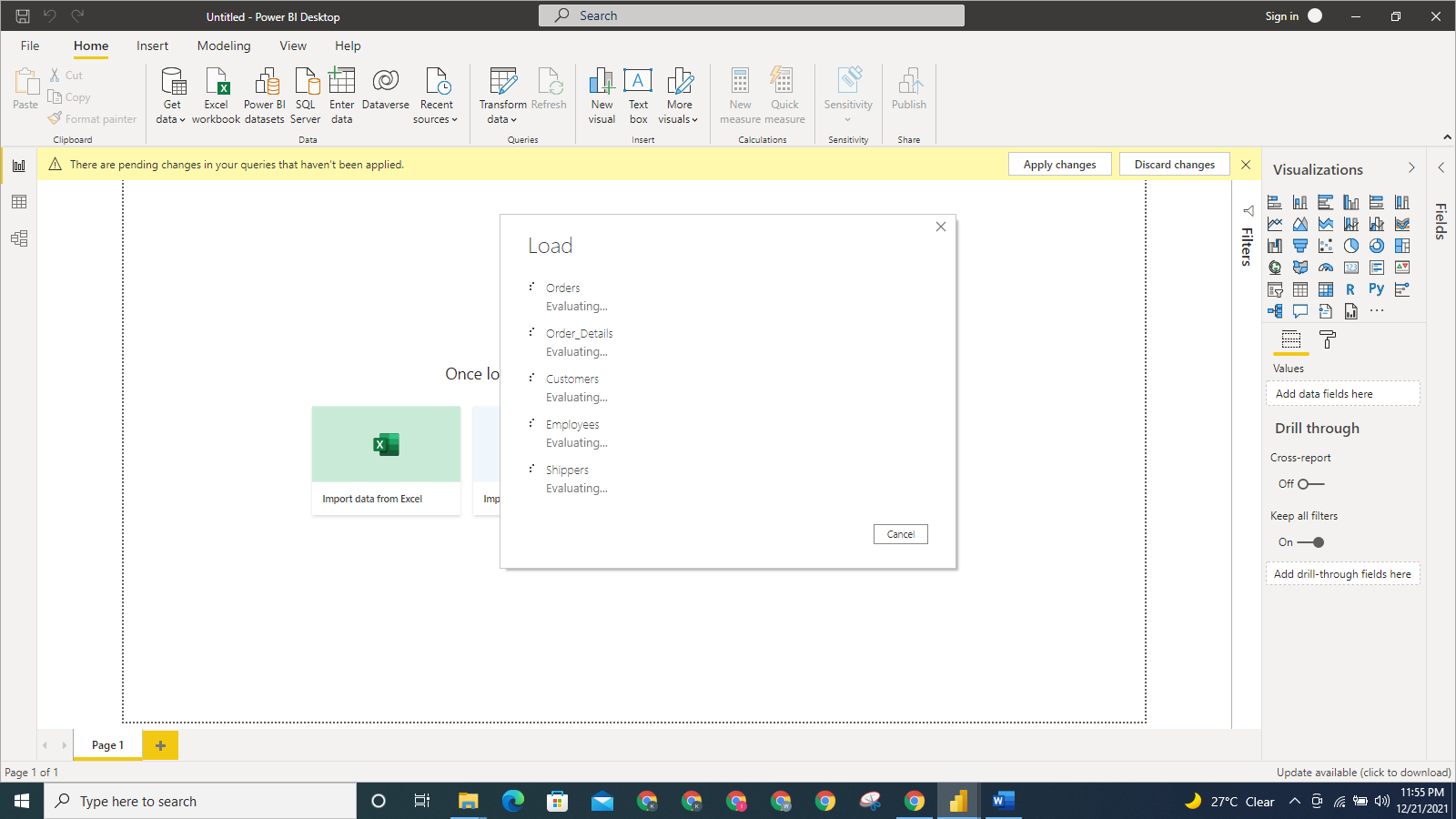


**Fig 3: Screenshot of Power BI start screen**

The tables needed for this project are selected and then loaded into Power BI. Figure 4 below shows a portion of the data from the Orders table in the Northwind database, and figure 5 shows the data being loaded into Power BI after selecting the tables.



**Figure 4: Screenshot of data in the Orders table**



**Figure 5: Screenshot showing the data being loaded into Power BI**

Other tables selected in the OData Feed for use in our projects are order\_details, shippers, employees, customers.

**1.5 Business Intelligence Requirements**

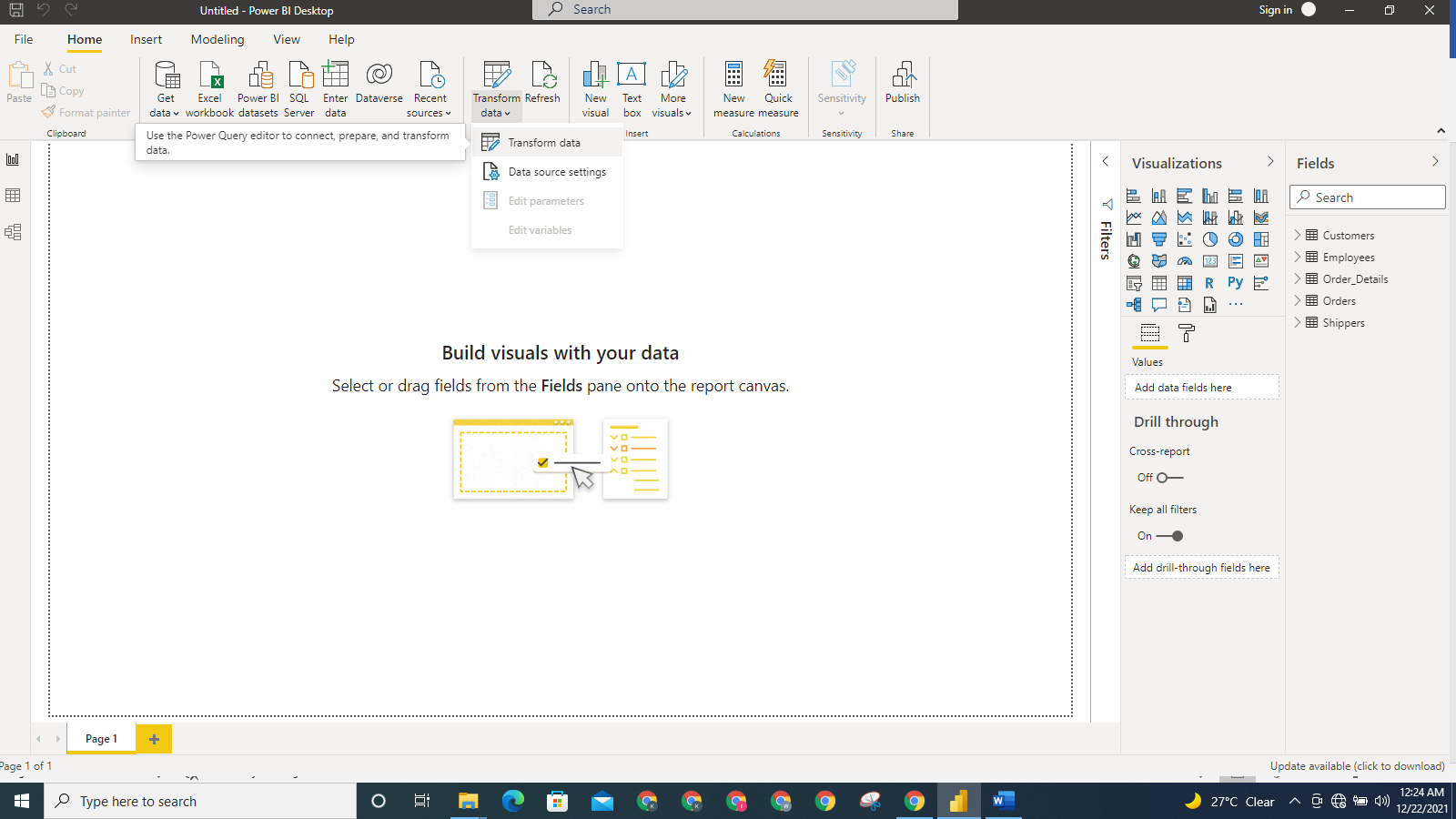
The business requirements in this project include:

* What is the amount spent on freight each year?
* What is the number of orders handled by each employee?
* What is the number of orders by ship country?
* What is the active number of customers each year?
* How many orders were shipped year?
* What is the freight amount to each country?

**2 Data pre-processing and Data Cleansing**

**2.1 Loading Data into Query Editor for transformation**

Data hardly comes in the form they are required for use. Some pre-processing and cleansing are required to put the data in the form and shape they will be needed. The tables in the selected database will be loaded into the Power BI Query Editor, which is a useful tool for getting the data into useful shape and form.



**Figure 6: Screenshot showing the Transform Data being selected**

2.2 Pre-processing and Cleansing Data for Analysis

The cleansing and pre-processing will include dropping unwanted columns, changing data type, renaming columns, etc. We clicked on **Transform Data** option in the **Home** tab of Power BI to load the data into the **Query Editor** for data transformation.

We are going to remove the columns we don’t need from all affected tables. This will help reduce the memory consumption of the data and as well make them easier to work with. To do this, we ensure the Orders table is the active table, then **Remove** **Columns option** was selected from the **HOME** tab. The following columns were removed from Orders table: ShipRegion, ShipPostalCode, Customer, Employee, order\_details, shipper, ShipName, ShipAddress, ShipCity. **Removed Columns** was added to **APPLIED STEPS** in the **Query Settings.**

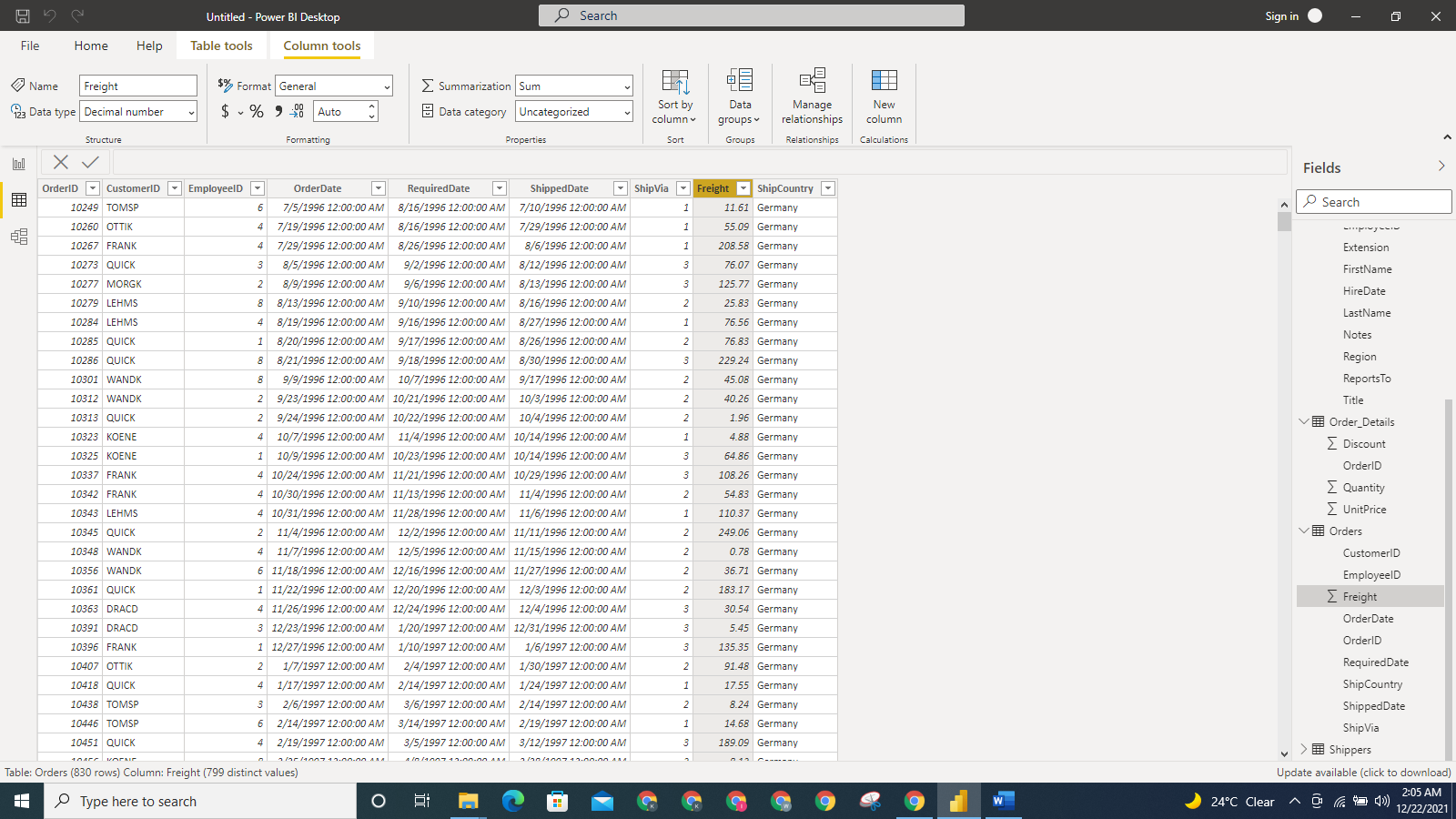


**Figure 7: Screenshot showing Removed Columns step added to the Query Settings**

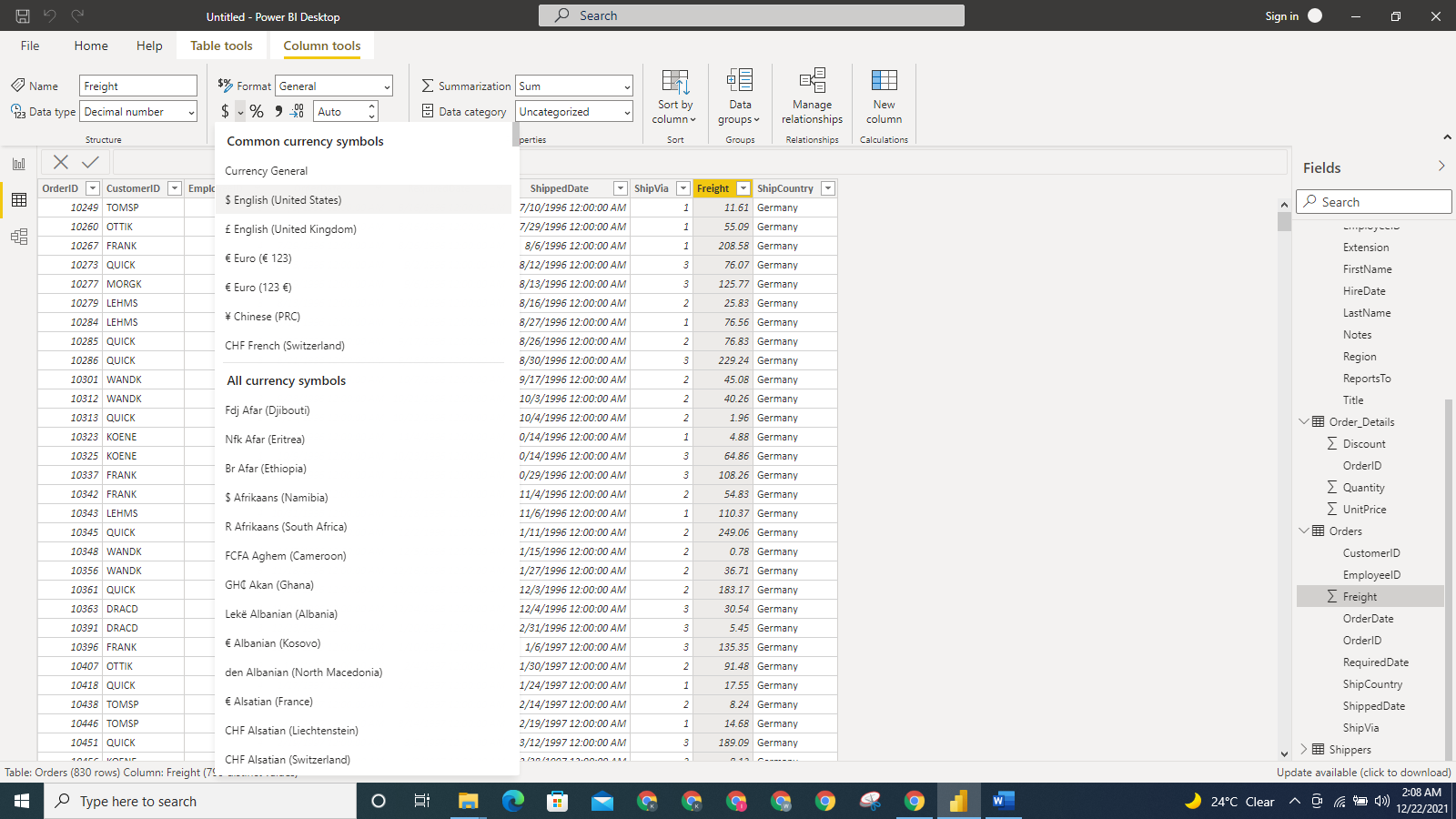
The columns that are not useful for our analysis in the other tables were also removed. This allows us have only the necessary columns useful for the business intelligence requirements.

Now we can load the data and return to the Report view of Power BI by selecting the **Close and Apply** option**.**

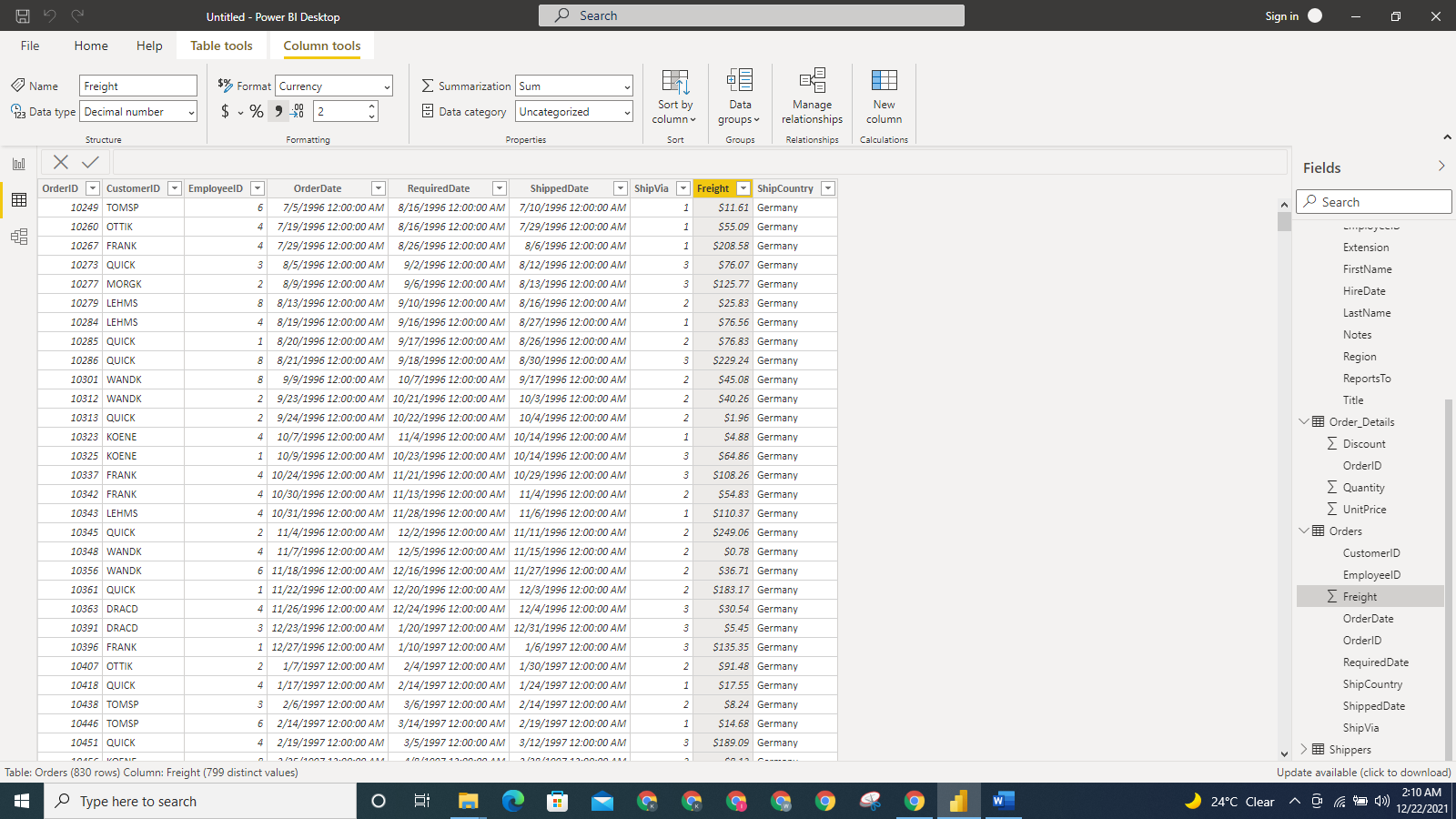
All the columns having currency values are explicitly converted to Currency Format as shown in the screenshots below.



**Figure 8: Screenshot showing the freight column in the Orders table having numeric values yet to be converted to currency**



**Figure 9: Screenshot showing the Currency Format being selected for Freight column**



**Figure 10: Screenshot showing Freight column is now in Currency Format**

In the Order\_details table, the UnitPrice and Discount were converted to Currency Fomat also.

**3 BI DATA MODELLING**

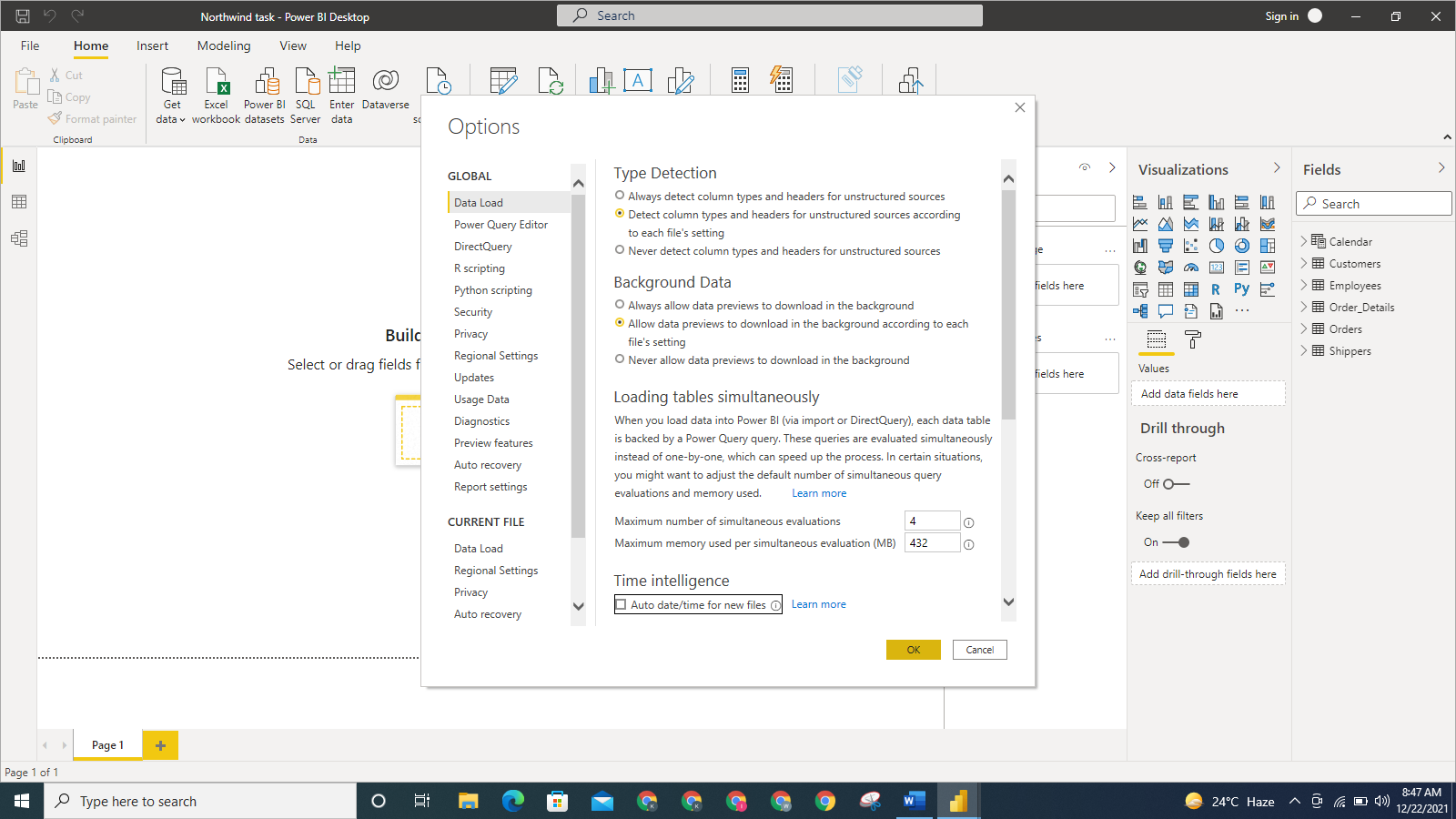
Power BI detected a relationship among the tables in the Northwind dataset automatically. The screenshot below shows the entity-relationship diagram of the tables.



**Figure 11: Screenshot showing the relationship detected automatically**

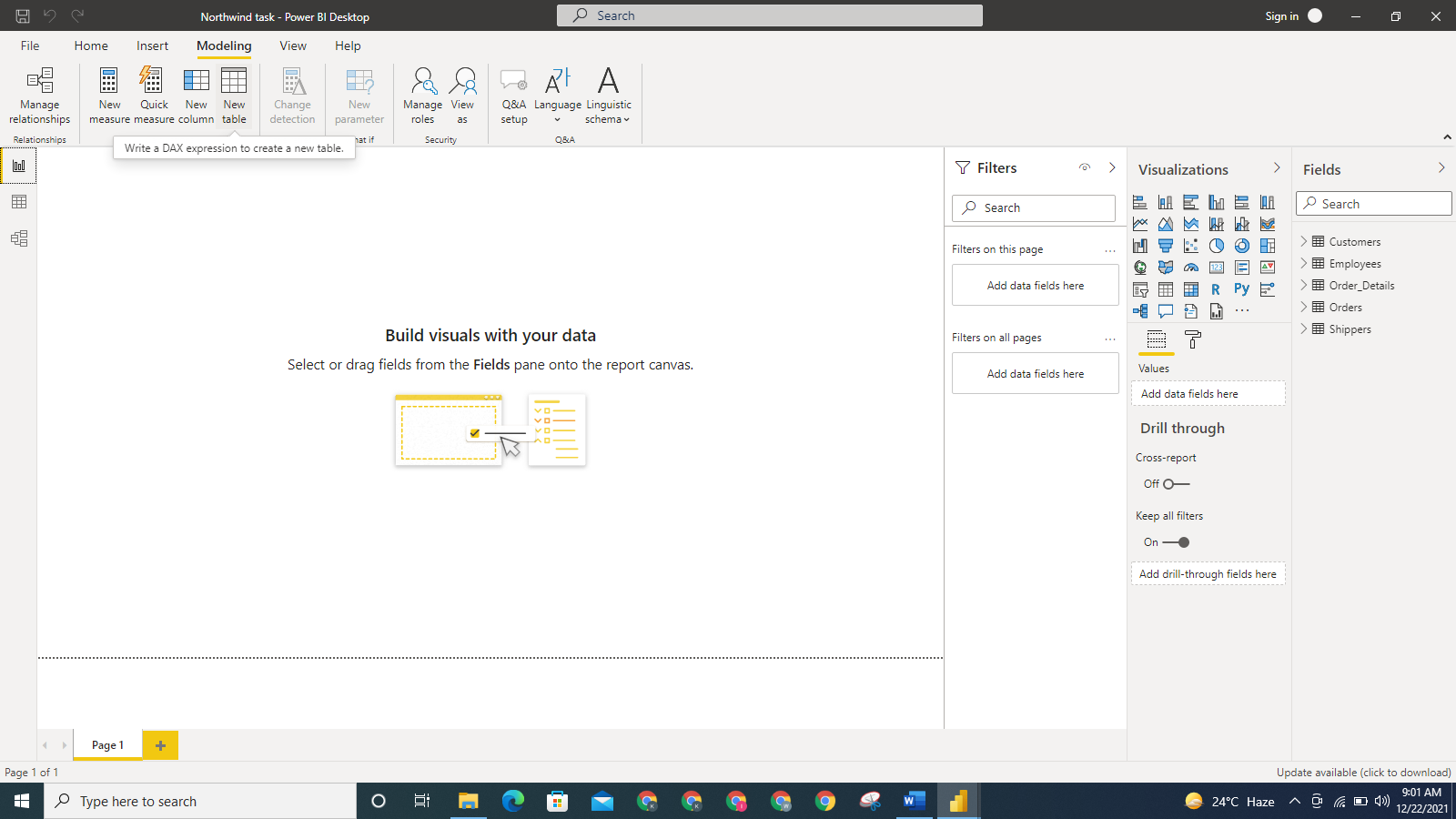
**3.1 Adding Calendar Table to the Model**

We will disable the Time intelligence capability of Power BI under **Options and Settings** so as to create our own time dimension for the Northwind database.

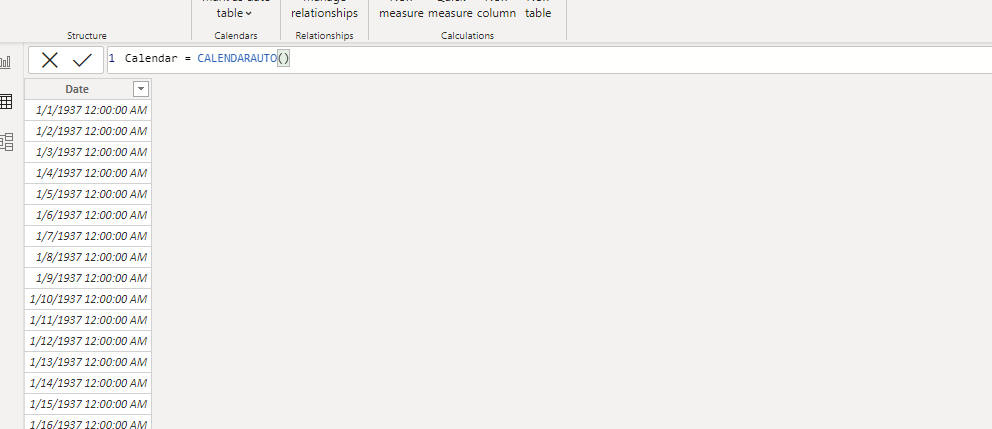


**Figure 12: Screenshot showing the Time intelligence: Auto date/time for new files being disabled**

The Calendar table will be created using the DAX function, CALENDARAUTO(). To create the Calendar table, we selected **Modeling**, then **New table** in the Power Bi environment.

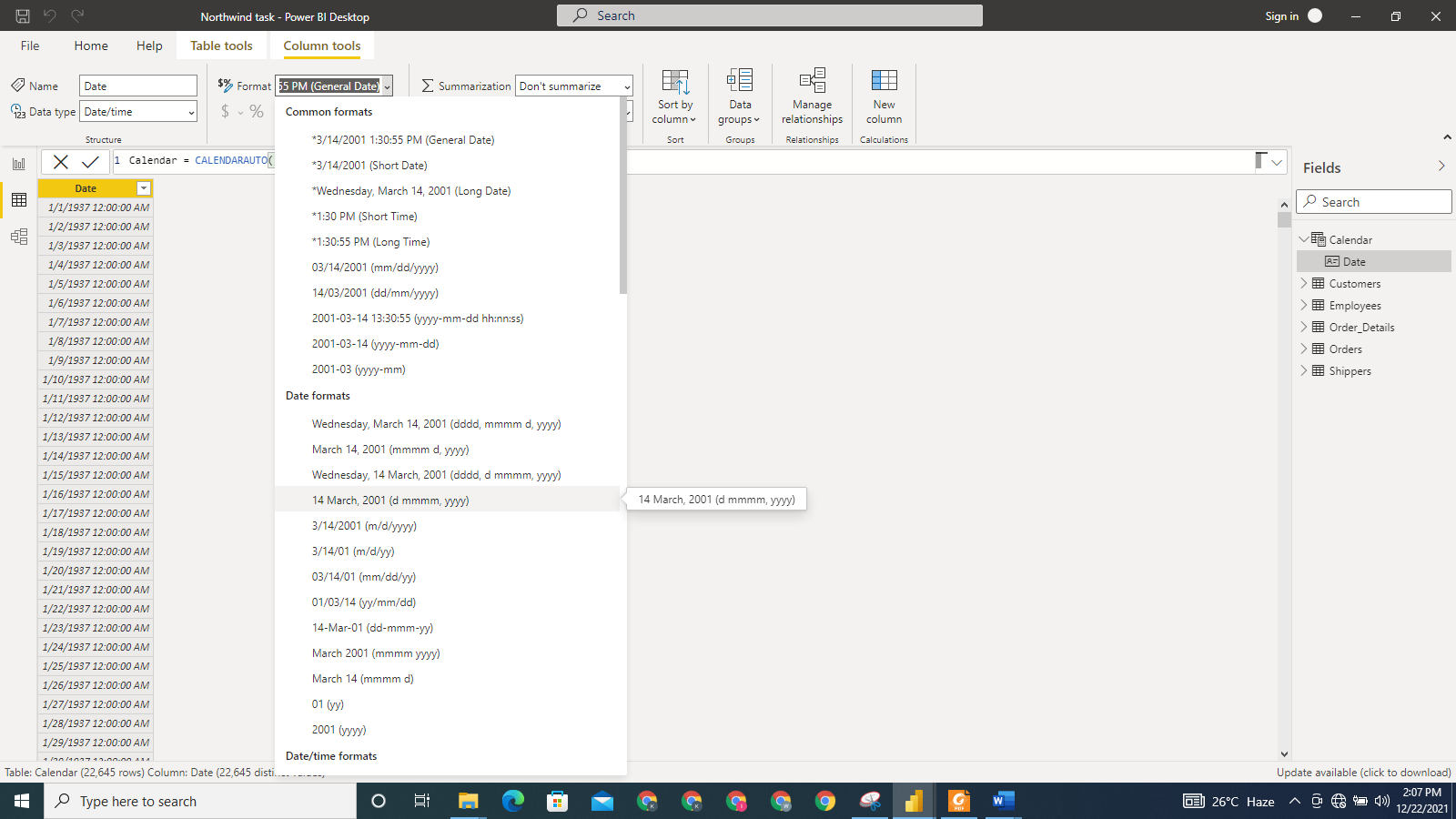


**Figure 13: Screenshot showing the New table being selected**



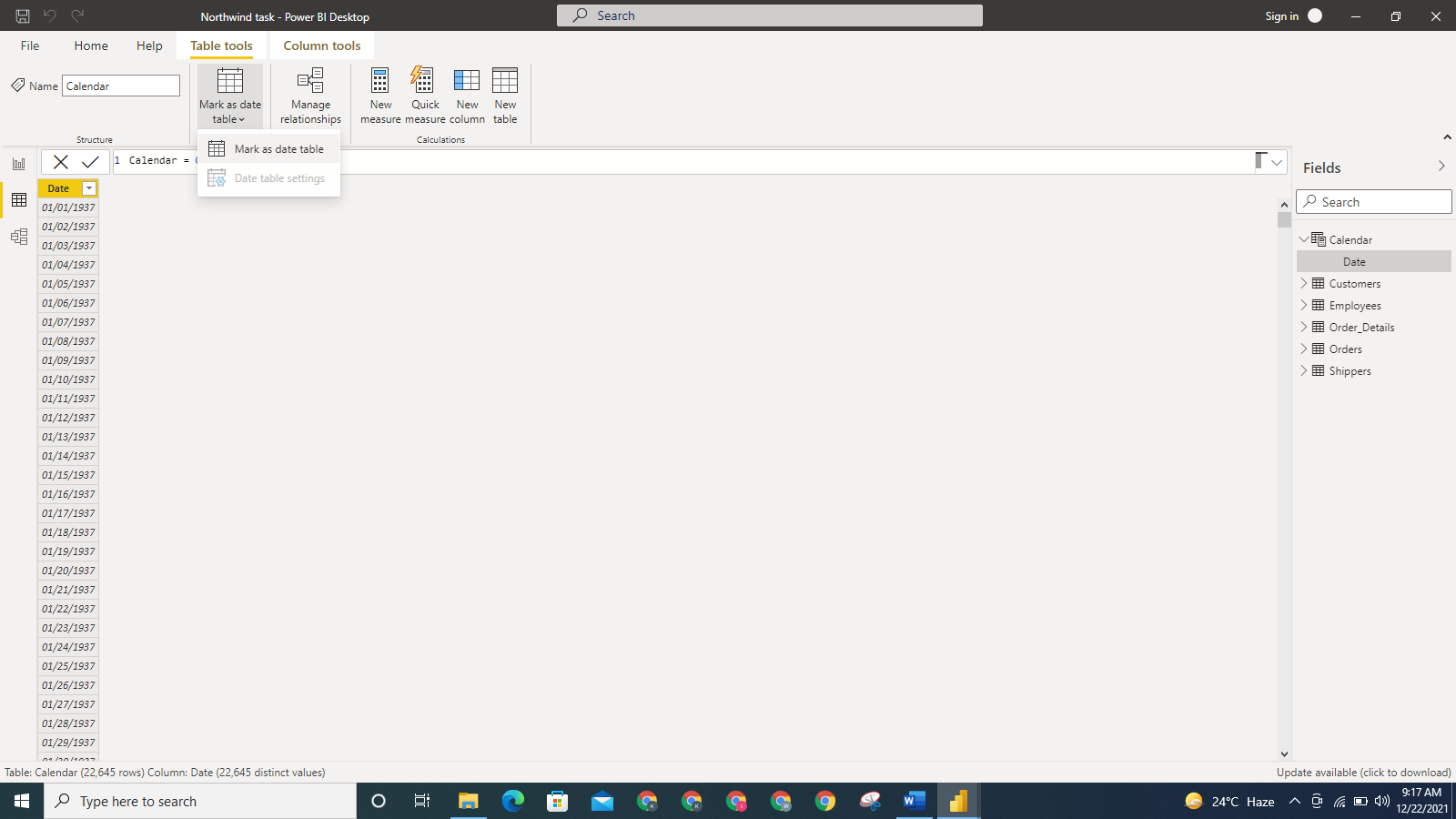
**Figure 14: Screenshot showing the CALENDARAUTO() DAX function**

The date column of the Calendar table is converted properly to d mmmm yyyy Format.



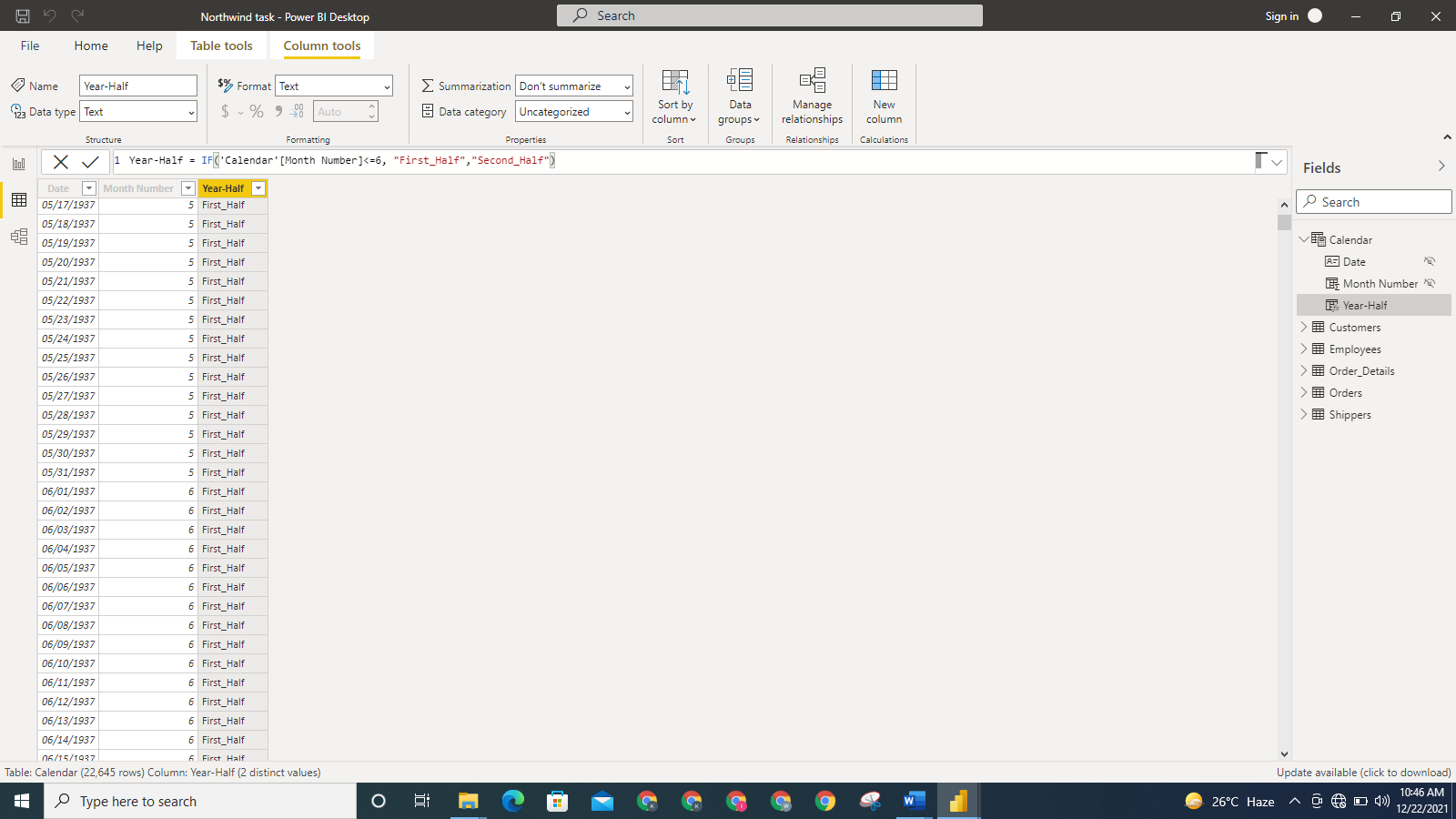
**Figure 15: Screenshot showing the date column being converted to d mmmm yyyy Format**

It is important we mark the Calendar table as a date table. This is a compulsory step that should be taken to ensure Power BI recognises the table as having time dimension. To do this, we select **Mark as date table** in the **Table tools**.



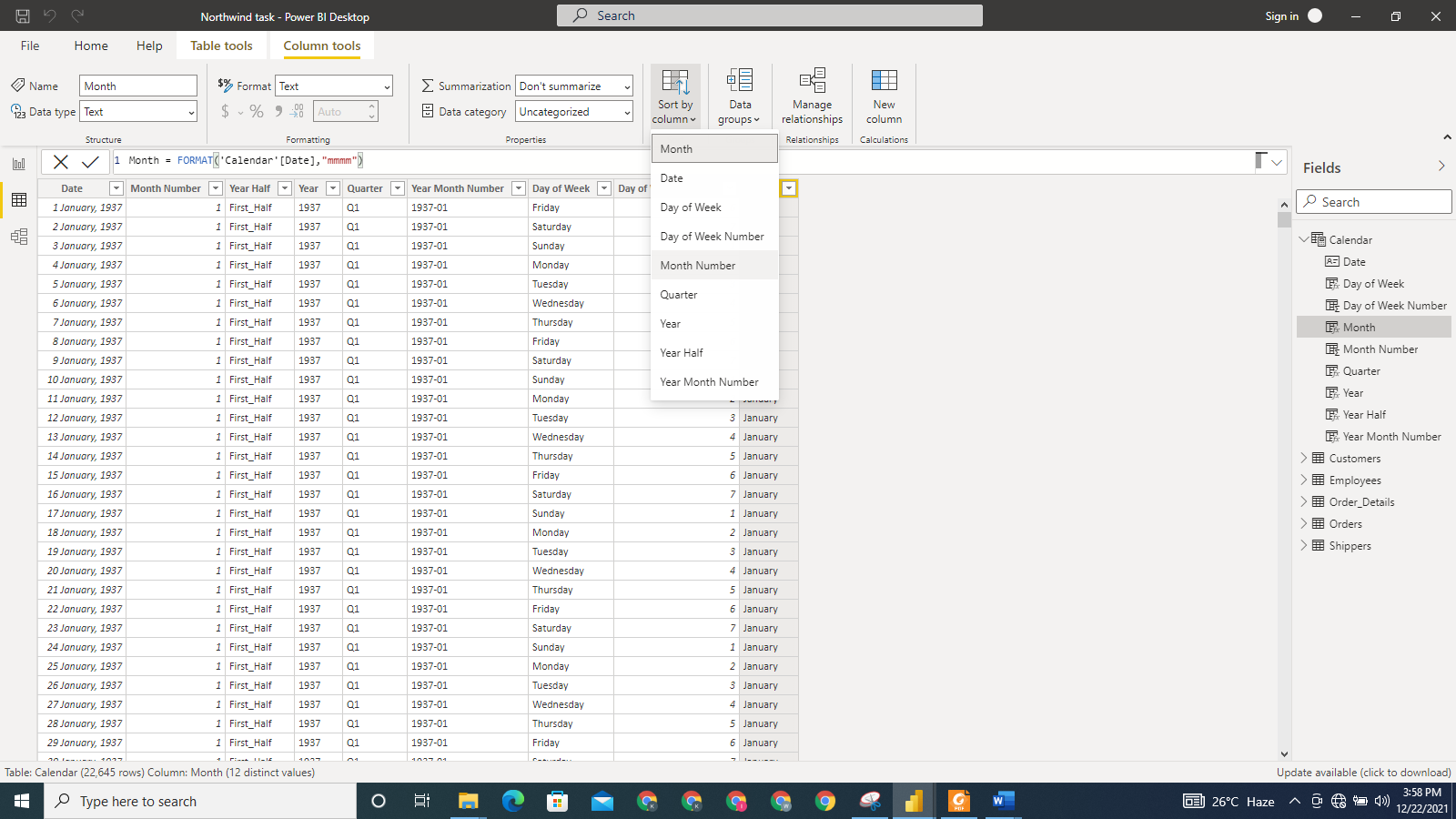
**Figure 16: Screenshot showing we marked the Calendar table as date table**

We will add more calculated columns to the Calendar table just created. For example, to create Month Number and Year-Half columns, the used are Month Number = MONTH(‘calendar’[Date]) and Year Half = IF(‘Calendar’[Month Number]<= 6, “First Half”, “Second Half”) respectively.



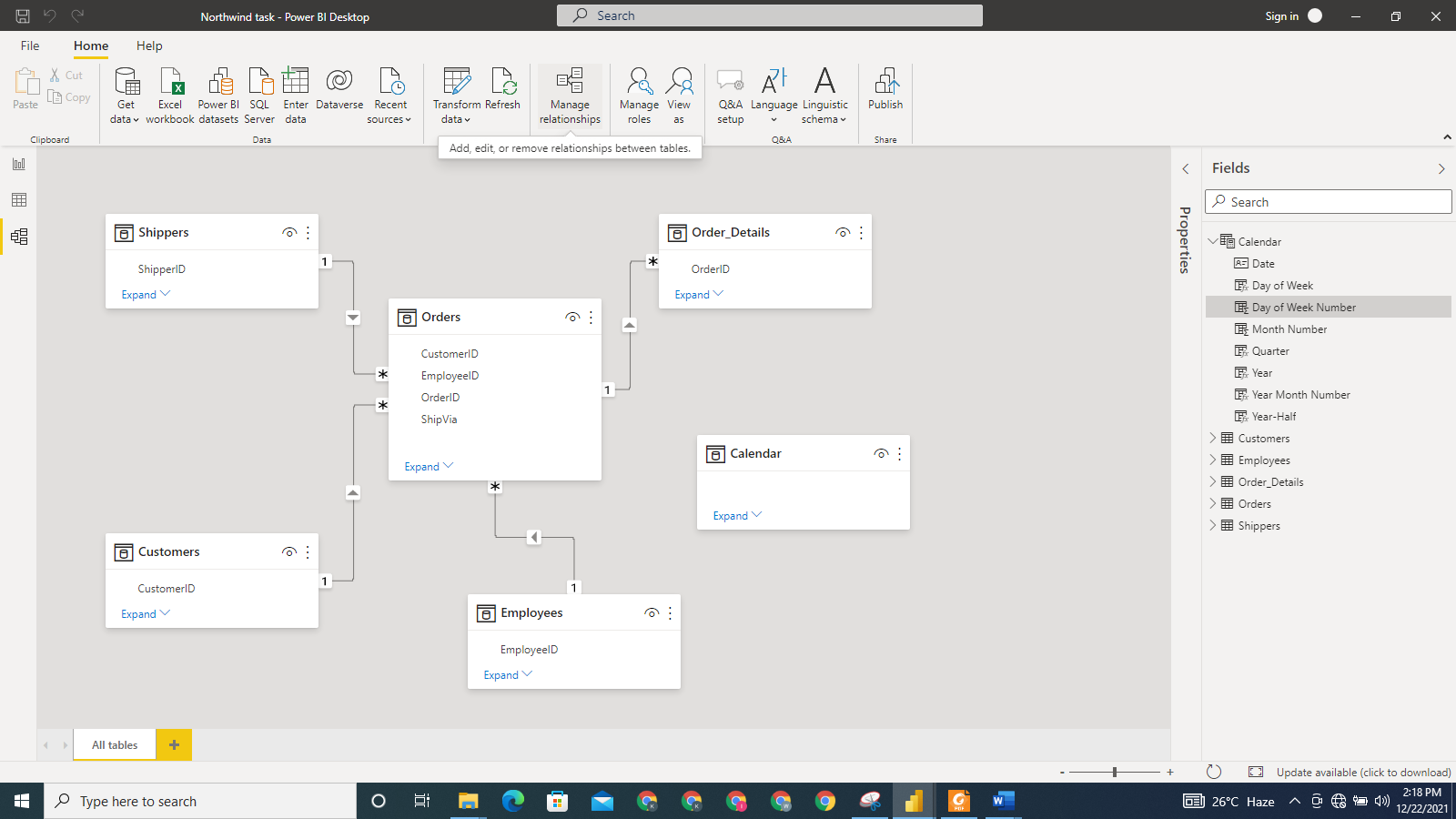
**Figure 17: Screenshot showing the Month Number and Year-Half columns are created and added to the Calendar table**

More columns are added to the Calendar with the use of the FORMAT function. Additional columns included are Year, Month, Quarter, Year Month Number, Day of Week, Day of Week Number. We have to ensure that Calendar table is sorted by calendar months and not in alphabetic order, so we sort the Month column of the table by the Month Number column. To do this, we select Month column, **Column tools** tab, **Sort by Columns** and then pick Month Number column.

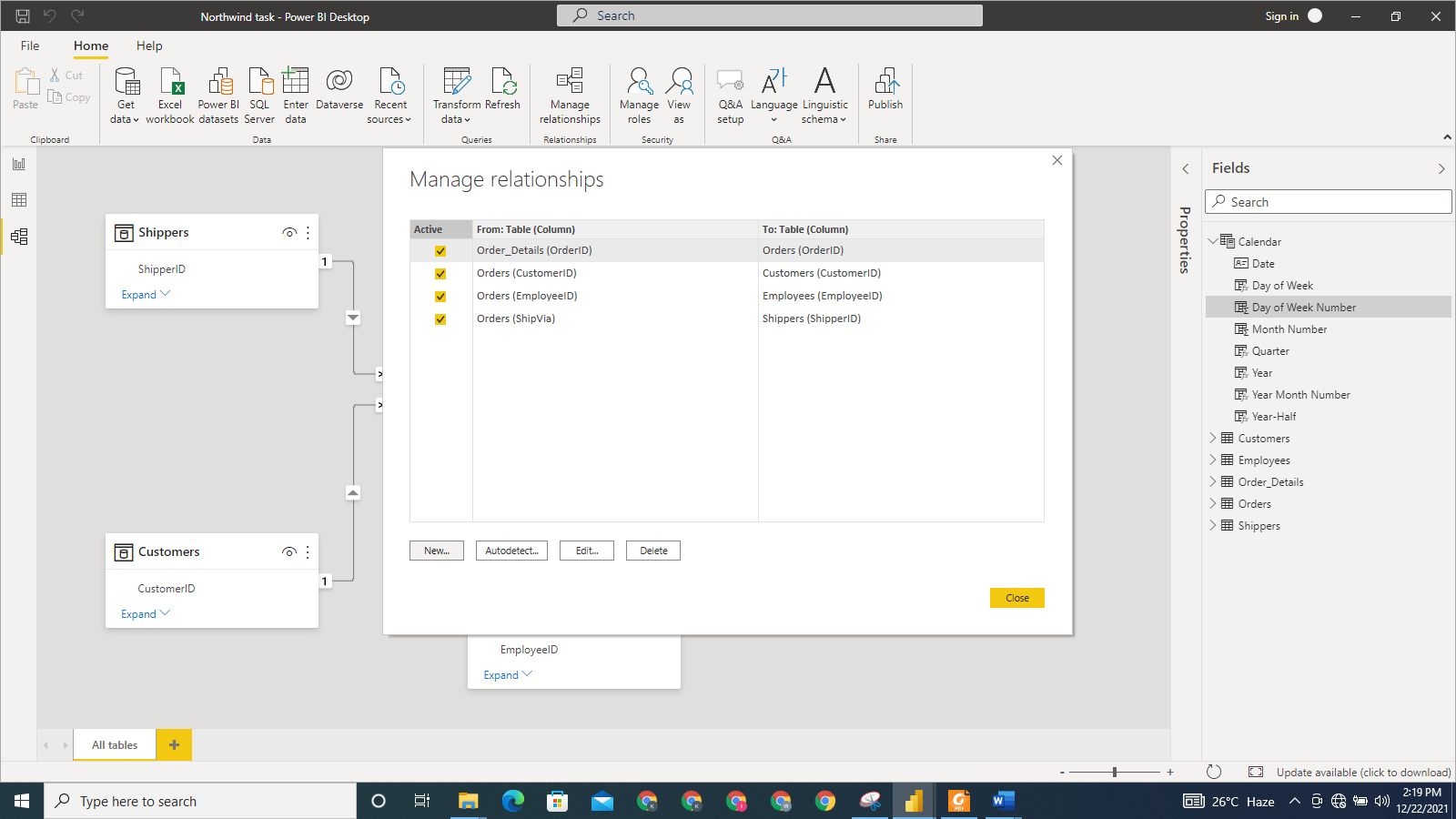


**Figure 18: Screenshot showing that we sort the Calendar table by Month Number column**

We created a relationship between Orders and Calendar tables. To do this, we selected **Manage Relationship** while in the **Diagram view**.

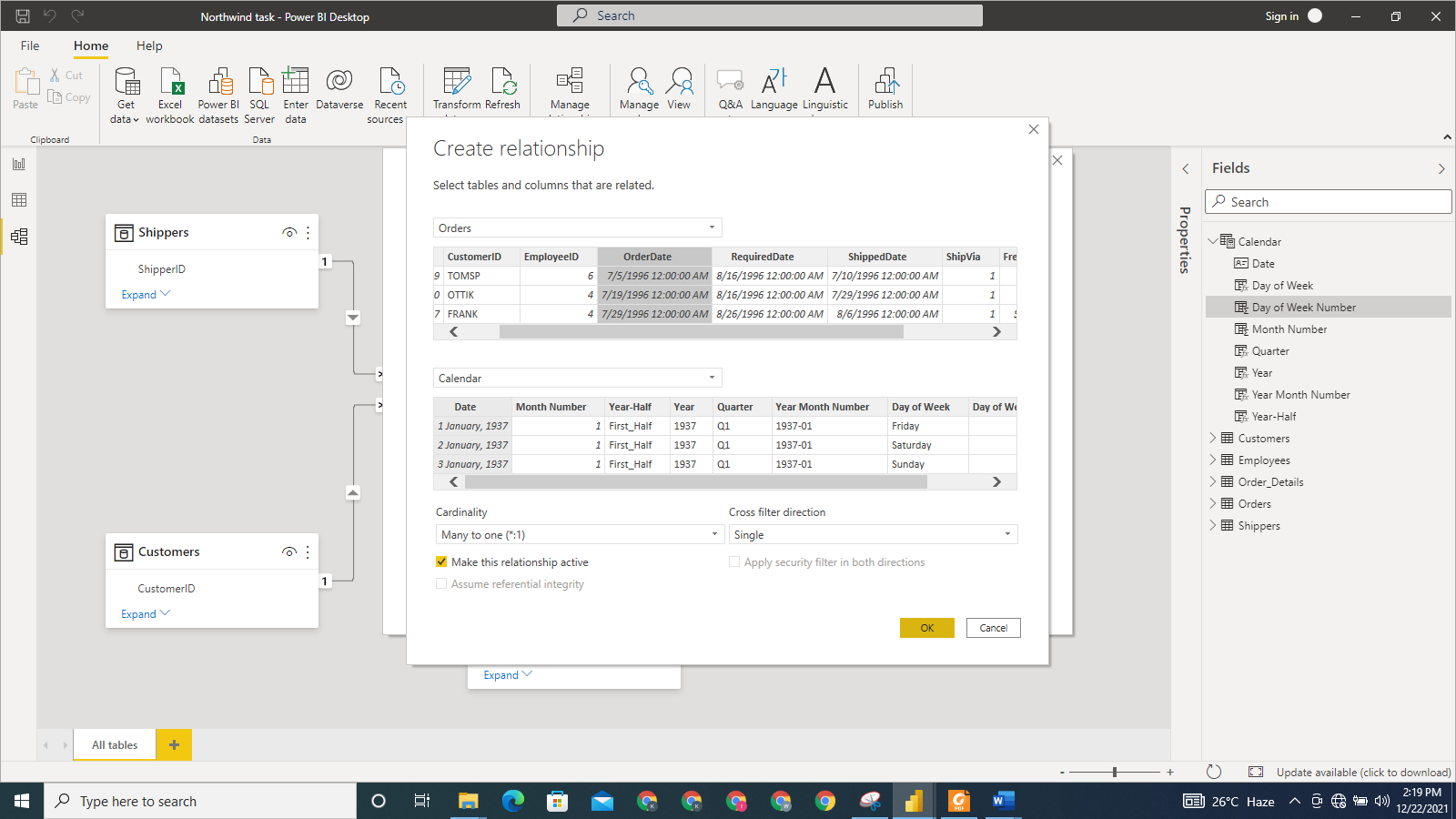


**Figure 19: Screenshot Showing the Manage Relationship being selected**



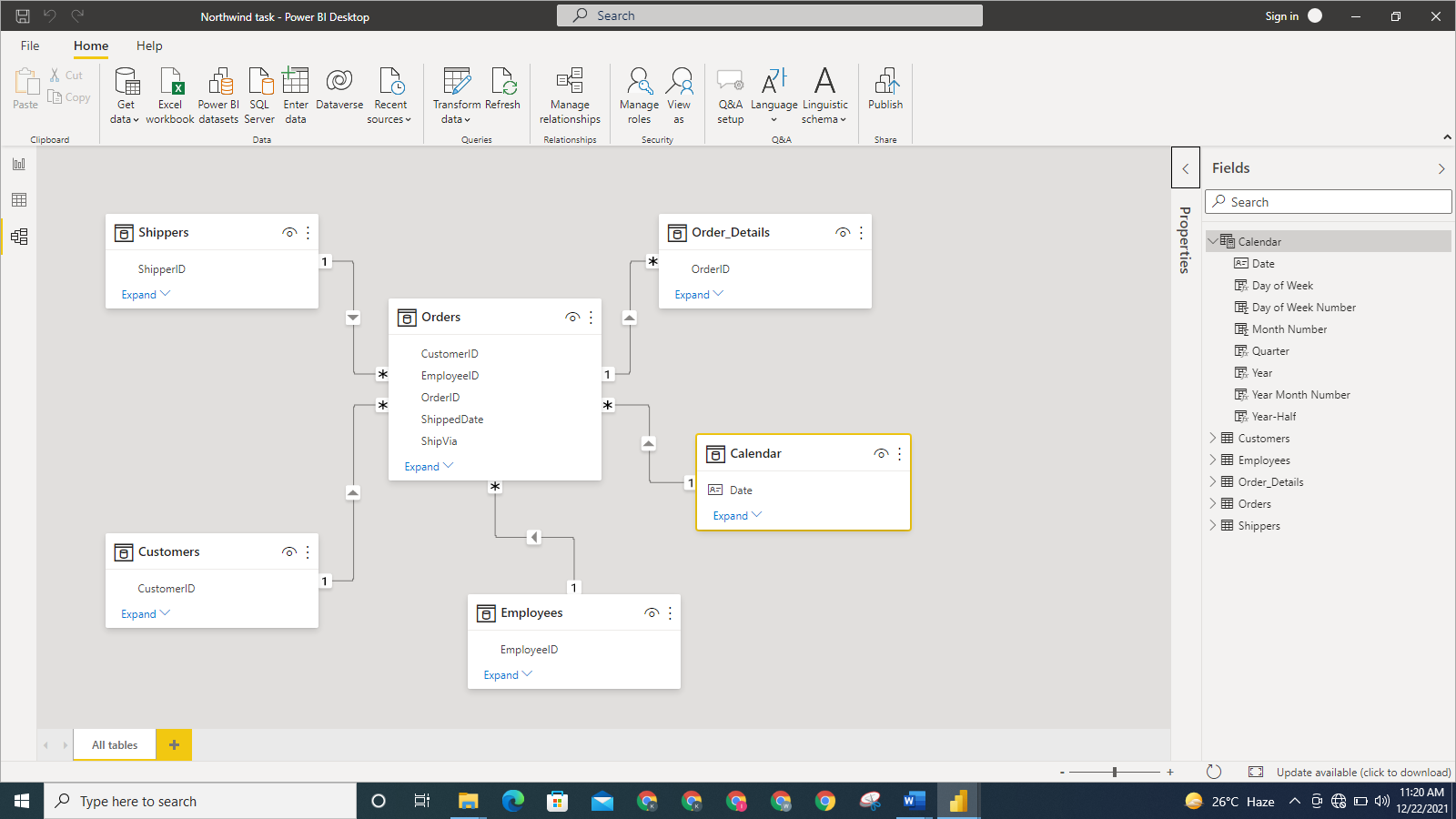
**Figure 20: Screenshot showing the active relationships**

We clicked on **New** to create a new relationship between Orders and Calendar. We used the Date column in the Calendar table and OrderDate column in the Orders table to create the relationship between the two tables.



**Figure 21: Screenshot showing the relationship created between Orders and Calendar tables**

We go ahead to present the model using Star schema in the **Model** view.



**Figure 22: Screenshot showing the model is presented in the star schema**

**SECTION 2: BUSINESS REPORT**

**ANALYSIS OF NORTHWIND FINANCIAL PERFORMANCE**

**NAME: [PLEASE PUT THE NAME HERE]**

**STUDENT ID: [PLEASE PUT THE STUDENT ID HERE]**

**Executive Summary**

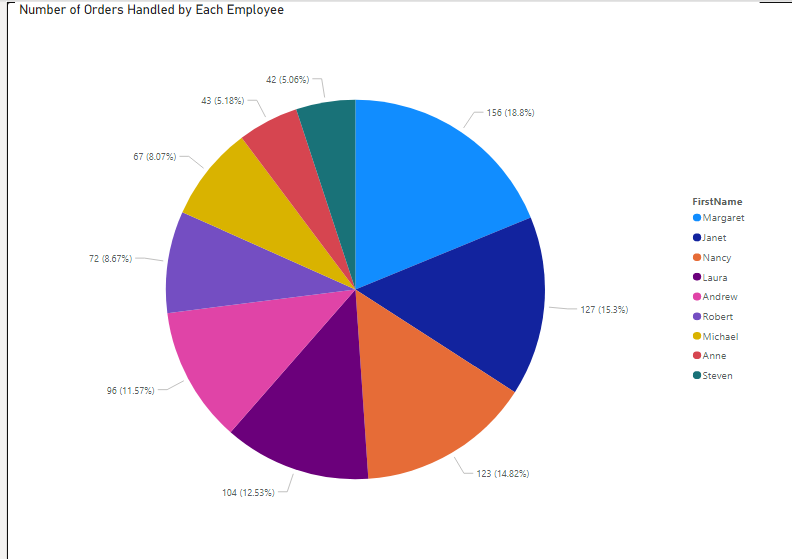
Northwind Traders has been in business for years. With this report, we will like to evaluate how well we have the performed in the previous financial years. Our company will be able to get important answers to key business questions by the virtue of this report.

The business requirements we will be able to address include:

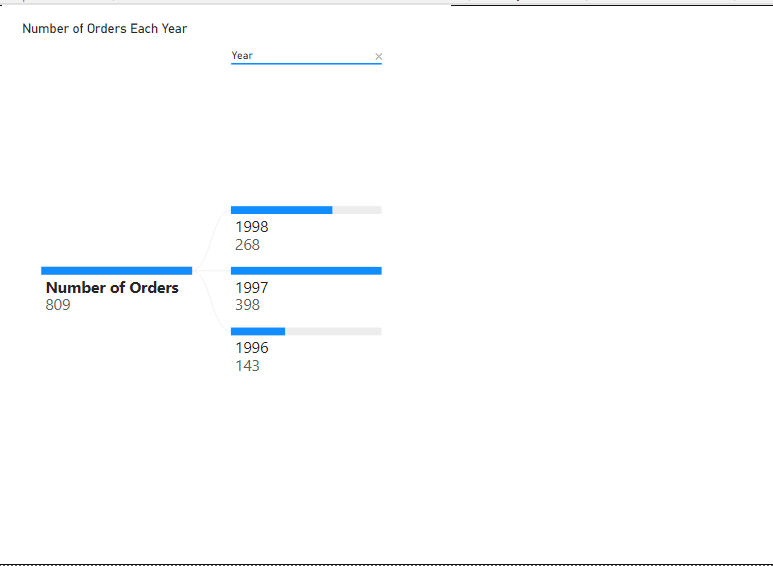
* What is the amount spent on freight each year?
* What is the number of orders handled by each employee?
* What is the number of orders by ship country?
* What is the active number of customers each year?
* How many orders were shipped each year?
* What is the freight amount to each country?

Through our power BI, we were able to come up with the following findings and conclusions.

Findings







Recommendations

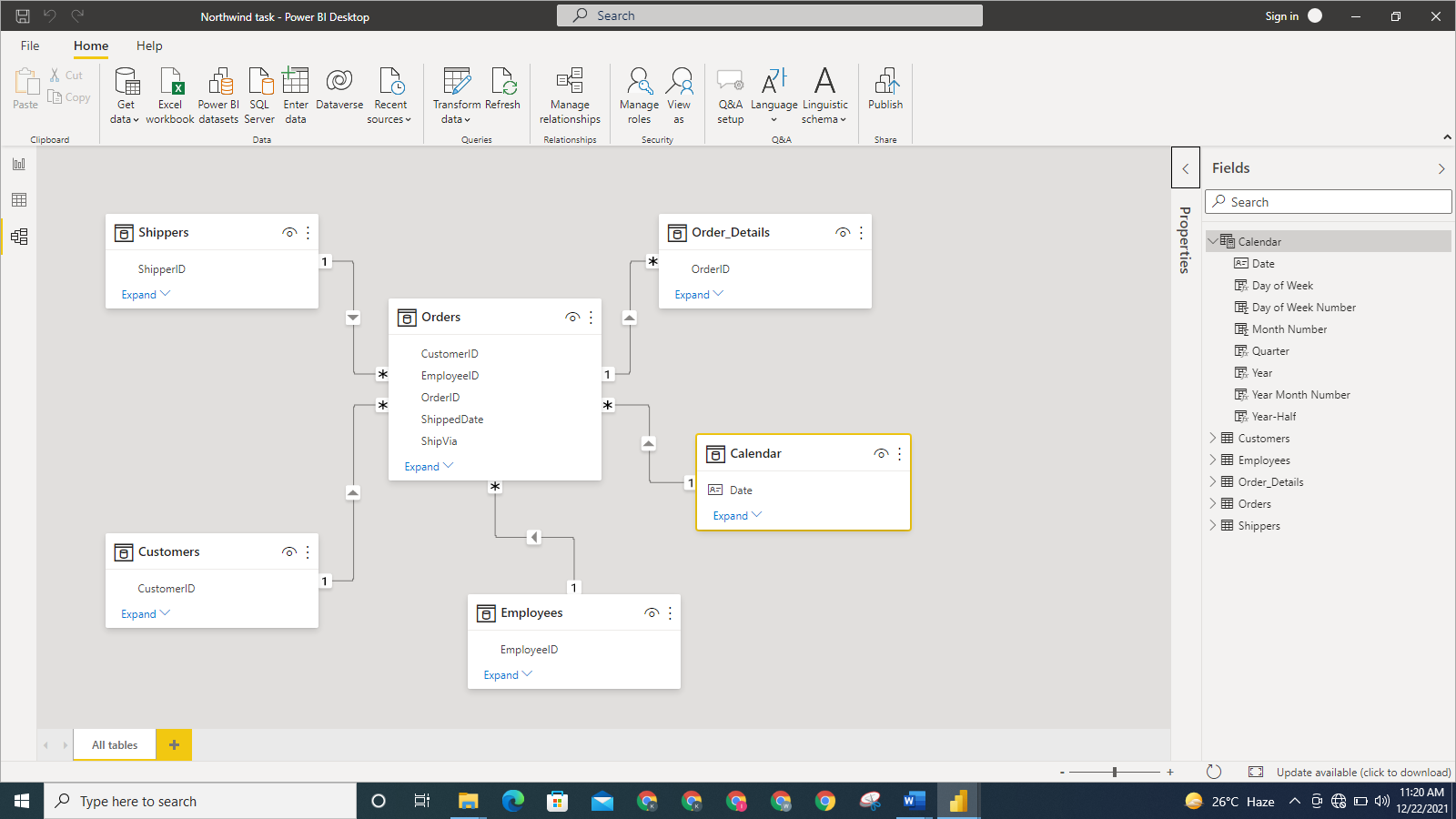
**Introduction**

In this project, we will analyse the data made available by Northwind Traders, a company engaged in in imports and exports of specialty food. The analysis will allow us come up with findings and recommendation that will help Northwind improve its performance.

The data of the company was accessed from the Power BI OData Feed using the url below:

<https://services.odata.org/V3/Northwind/Northwind.svc/>

Using the link, we were able to pull tables necessary for the analysis. We model the relationship among the table by using the Star Schema. The model is presented below:



**Figure 1.: Screenshot of the data model for Northwind dataset presented in a Star Schema**

By using the Star Schema, it means we used a single fact table and several other dimension tables.

We pulled 5 tables from the OData feed. The tables are Shippers, Order\_details, Orders, Customers, and Employees. The relationship among the data was also automatically imported from the OData feed. An additional table- Calendar table- was created, and a relationship was created between it and the fact table which is Orders. These tables will allow us meet our business requirements in this project. The calculated columns earlier created in the Calendar table and the measures that will be created in this section will help facilitate the building of our report.

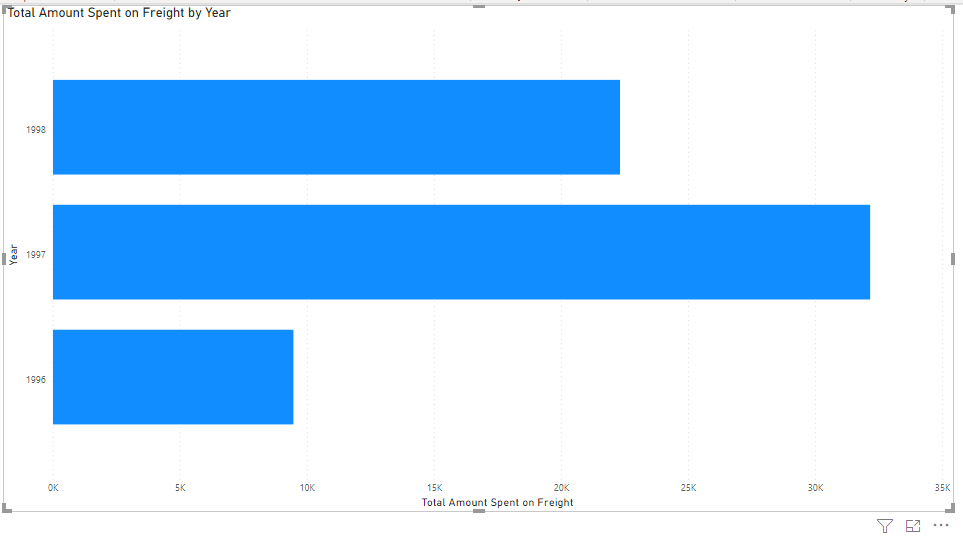
1. **What is the amount spent on freight each year?**



**Figure 1.2: Screenshot showing for total amount spent on Freight measure**

A measure was created to calculate the total amount spent as freight.

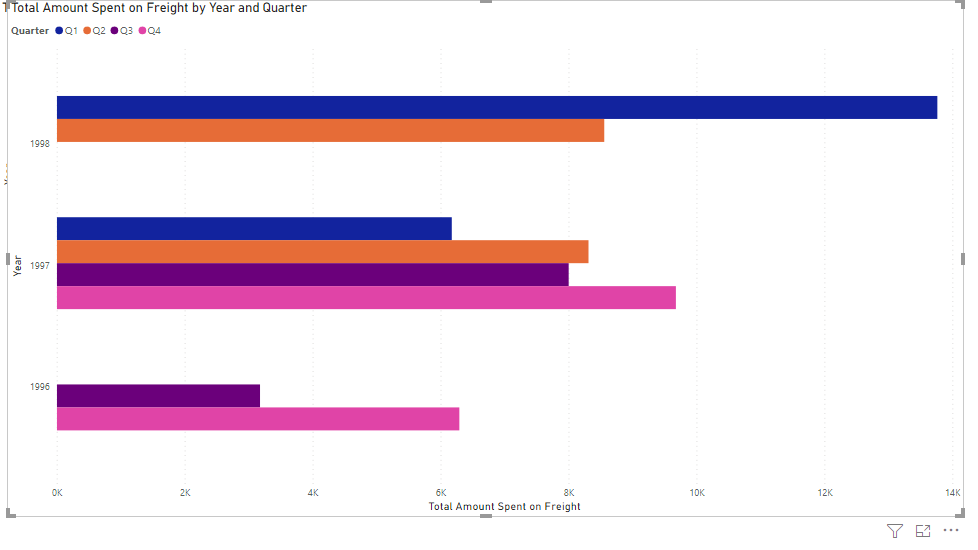
* To create the measure, we right-clicked on the Orders table to select **New Measure**. Then used the function in the above screenshot to create a new measure.



**Fig 1.3 Screenshot of bar chart showing freight spent in each year**

Bar chart is used to show the total amount spent as revenue in each year here. It was found that the total freight was $34,000 in 1997 and reduced to $24,000 in 1998.

* The bar chart was created by selecting the Stacked bar chart; dragging **Total Amount Spent on Freight** measure and dragging **Year** in Calendar tableto **Axis** box.
* The visual was filtered to ensure it included only the years covered for the freight. We did this by using the filter for **Year** under the **Filters** pane to the left of **Visualizations** pane. We selected 1996, 1997 and 1998 as the years we are interested in.



**Figure 1.4: Screenshot of bar chart showing freight spent in each year and quarter.**

Bar chart is used to show the total amount of freight by year and quarter. It was found that the most amount of freight was incurred in the first quarter of 1998.

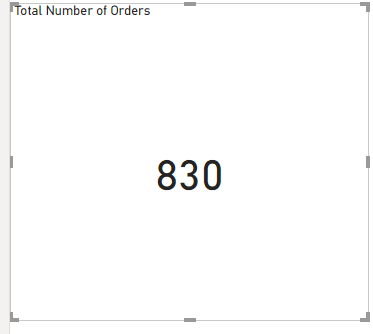
* To create the bar chart, we selected the **Stacked bar chart**. We dragged **Year** in Calendar table to **Axis** box; **Quarter** in Calendar tableto **Legend** box. Lastly, **Total Amount Spent on Freight** was dragged to **Values** box.
* The visual was filtered to ensure it included only the years covered for the freight. We did this by using the filter for **Year** under the **Filters** pane to the left of **Visualizations** pane. We selected 1996, 1997 and 1998 as the years we are interested in.

1. **What is the number of orders handled by each employee?**



**Figure 1.5 Screenshot of measure for the number of orders**

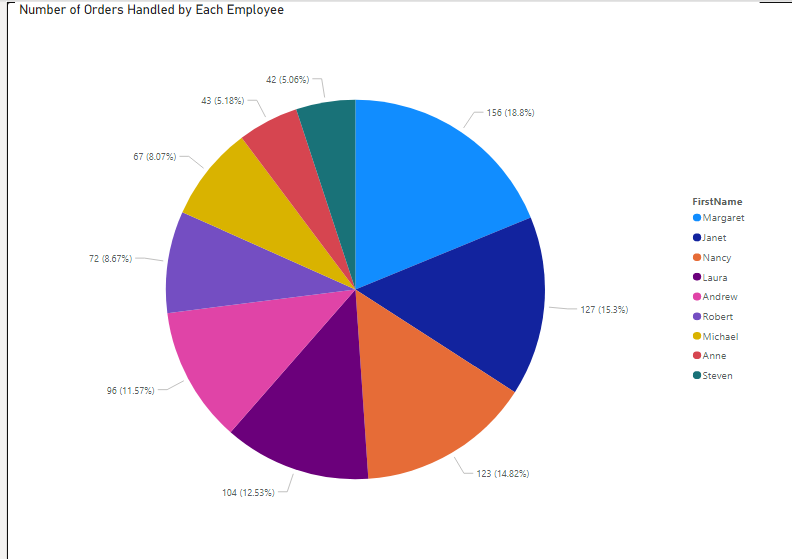
* To create the measure, we clicked on Orders table, select **New Measure** and create a measure using the function in Figure 1.4.



**Figure 1.6: Screenshot of Total Number of Orders measure**

It was found that the total number of orders handled by the company is 830.

* To create the visual in Figure 1.5 above, we selected the Matrix visual, and then drag **Total Number of Orders** measure to **Fields** box.

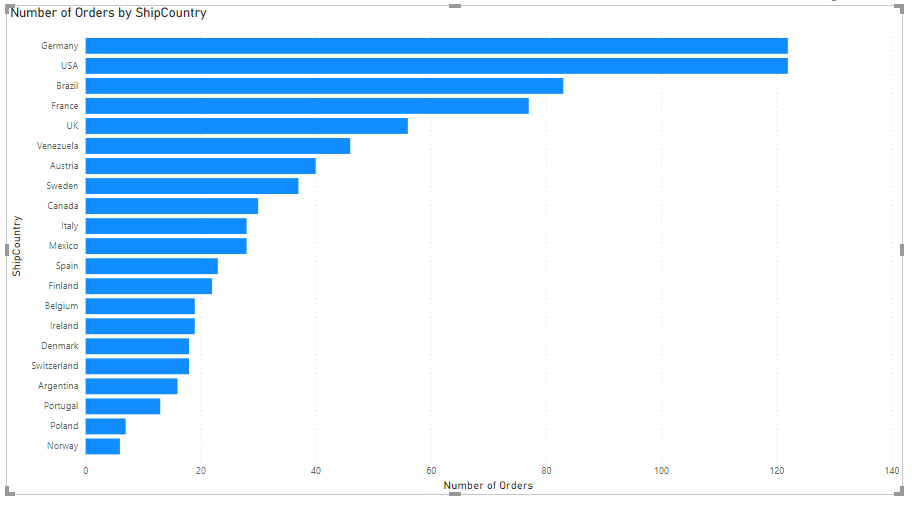


**Figure 1.7: Screenshot of number of orders ordered by each employee**

Of the 830 total orders received by Northwind, Margaret handled the highest number of orders: 154. The least number of orders handled by an employee was 42, and it was handled by Steven.

* To create the Pie chart visual, we select the visual from the **Visualizations** pane. We dragged **FirstName** to **Legend**, and dragged **Numbers of Orders** measure to **Values**.

1. **What is the number of orders by ship country?**



**Figure 1.8: Screenshot of orders by ship country**

To ascertain the country to which most orders are shipped and to which least orders are shipped, we created a Clustered bar chart. It was found that 122 orders were made to both Germany and USA, and this is the highest recorded. The least number of orders, 6 was made to Norway.

* To create the visual, we select the Clustered bar chart from the **Visualizations** pane. We dragged **ShipCountry** to **Axis** box, and **Number of Orders** measure was dragged to **Values** box.

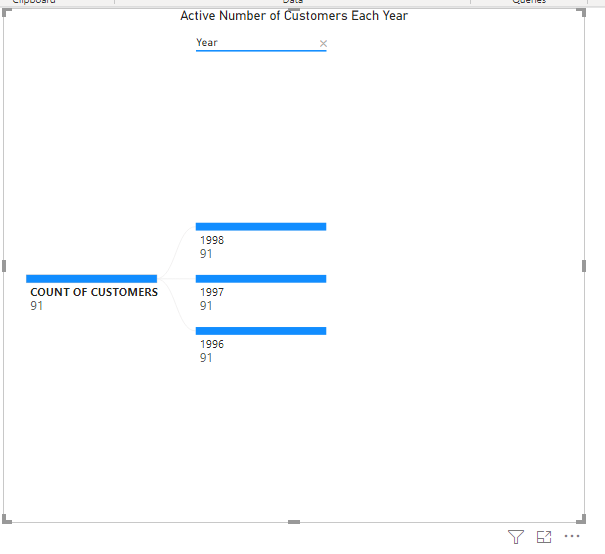


**Figure 1.9: Screenshot of orders by ship country on a map**

With the use of Map visual, we were able to clearly see the country to which we ship the largest order. It also reveals the regions to which we ship most of our orders. Germany and USA take the lead among the countries our orders go to. It was found also that most of our orders were shipped to Europe.

* To make the Map visual, we selected Map visual. We dragged **ShipCountry** to **Location** box and **Number of Orders** measure to **Size.**
* Using **Format** pane under the **Visualizations** pane, we applied different colors to the bubbles representing each country. We did this under the **Data colors** option.

1. **What is the active number of customers each year?**

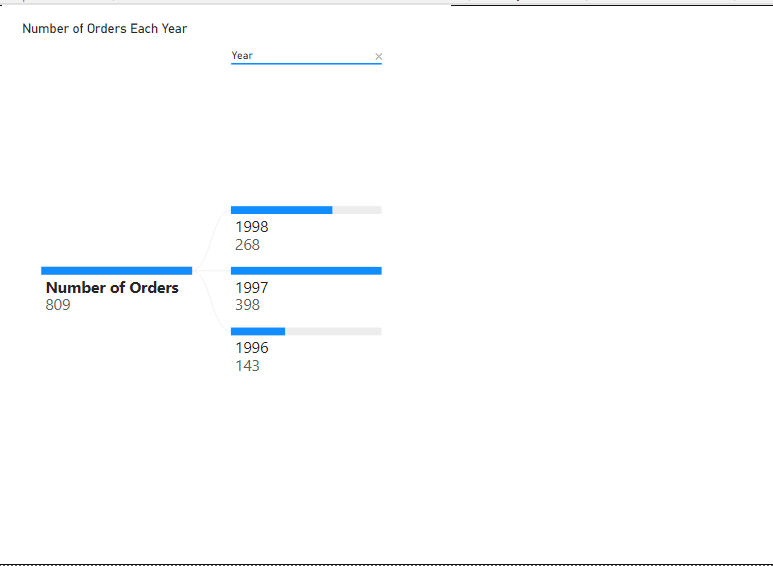


**Figure 1.10: Screenshot of active number of customers each year**

From the Decomposition tree visual, it was found that the active number of customers for each year is 91.

* To create the visual, we selected Decomposition tree, then dragged **Count of Customers** measure to **Analyze** field and dragged **Year** in Calendar table to **Explain by** field.

1. **How many orders were shipped each year?**

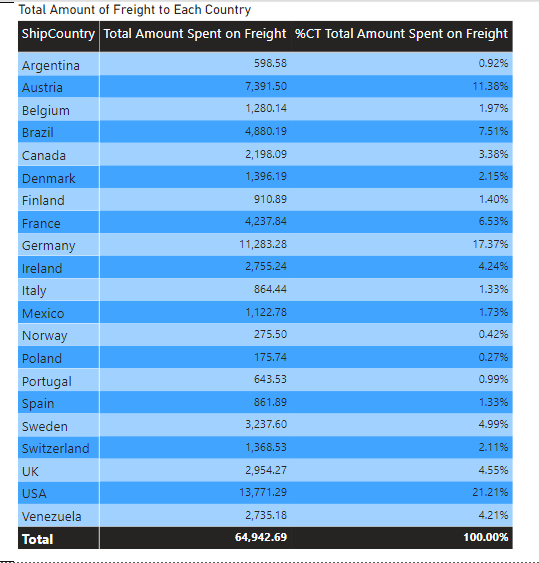


**Figure 1.11: Screenshot of orders by number of years**

Using Decomposition tree, we found we were able to see the number of orders for each year. It was found that there is an increase in the orders between 19996 and 1996, but there is a decline the orders between 1997 and 1998. The total number of orders in 1998 was 268.

* To create the visual, we selected Decomposition tree visual, then dragged **Number of Orders** measure to **Analyze** field and dragged **Year** in Calendar table to **Explain by** field.

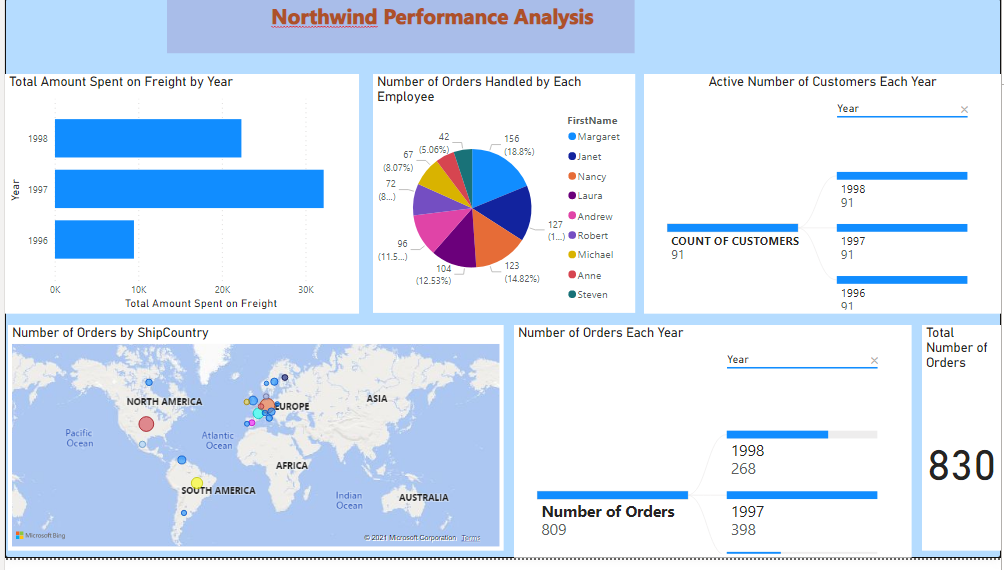
1. **What is the freight amount to each country?**



**Figure 1.12: Screenshot of total amount spent as freight to each country**

The Matrix visual allows us to display total amount spent as freight to each country. It was found that we spent more on freight to the USA more than any other country. The total amount spent as freight to USA alone was $13,771.29, representing 21.22% of the total amount of freight.

* To create the visual, we dragged **ShipCountry** to **Rows** box and **Total Amount Spent on Freight** measure to **Values** box. We dragged **Total Amount Spent on Freight** measure to **Values** box again and converted the value to percent of column total. We did this by clicking on the arrow beside the measure in the **Values** box and selected **Show values as**, and then **Percent of column total.** Doing this enabled us to show the amount spent on freight in both values and percentages.
* The Matrix visual was formatted to make it more presentable. To do this, we clicked on **Format** pane under the **Visualizations** pane. We clicked **Style** to open it and then selected **Bold header flashy rows**.



**Figure 1.13: Screenshot of the dashboard showing Northwind performance analysis**

**Conclusion**

Northwind have been in business for some years. Building a report to know how well we are performing in business is quite essential. With the use of Power BI, we are able to carry out a detailed analysis of our performance, leveraging on the available data. From our analysis, we make the following findings about Northwind Traders:

* The total freight in1998 was $24,000. Meanwhile, it was $34,000 in 1997. This represents a decrease of 29%.
* Drilling down further, we found that the highest amount of freight was incurred in the first quarter of 1998. The amount spent as freight in this quarter alone was $14,000.
* The total number of orders shipped by Northwind for the whole of period was 830. Of the 830 orders, 154 being the highest was handled by an employee named Margaret. Steven handled the lowest orders: 42.
* It was found that the highest orders were shipped to both Germany and USA. This means there is a large market for Northwind’s in both countries. The least order which is 6 was made to Norway.
* Though the largest part of the orders is shipped to USA and Germany, most of the orders are shipped to Europe. This means there is a large market for the product in Europe.
* The active number of customers for the company is 1991.
* There was a decline in order between 1997 and 1998.
* It was found that the amount spent as freight amount to $64942.69.

The following recommendations were made:

* The company should inspect the cause for the rising cost of freight.
* If the cost of freight is not directly proportional to the orders shipped to each control, cost reduction strategy should be adopted by the company. This may be through negotiating with the freight companies for favourable fees.
* The company should concentrate more on Europe since there is a large market for its product there.
* The company should also map out strategies to penetrate some other part of the worlds. This will help the company raise its international recognition and also help it increase the number of its customers worldwide.
* The company should put a solid reward system in place for its employees. This will motivate them to be able to handle more orders.