Student: Fionnuala Murphy

ID: 10361683

Module: Data Warehousing and Business Intelligence.

CA: Northwind Data Warehouse Report

**Introduction:**

There are two types of databases that companies tend to have. The one most people think about is the day-to-day transactional database that records all transactions that take place (purchases, sales, invoices, salaries) for the business to function. This style of database is highly normalised with a focus on efficiency in processing these day-to-day transactions. Unfortunately, this does not lend itself to time-series analysis, or indeed, many other complex queries that a company may need to make strategic decisions about the future.

As noted in the Data Warehouse Toolkit, “one of the most important assets of any organisation is its information”. The generic Database is for operational record keeping, but a Data Warehouse is used for analytical decision making.

Data Warehouses can take information from various sources, including multiple databases. The data is extracted from its source location, cleaned and transformed into a standardised format and then loaded into the warehouse. From there the information can be loaded into BI dashboards for analysis.

The BI dashboard is used to transform the information from the warehouse into a visual story. It can allow the end user to see how the company is performing week to week or quarter to quarter.

For this assignment I took the Northwind Trader company and created a Data Warehouse for it. I extracted data from the underlying database, transformed it and then loaded what I needed into the warehouse. I then loaded the data from the Warehouse into Power BI to create dashboards for the areas I wanted to analyse.

**Key Performance Questions:**

In advance of creating a warehouse, the management of a company need to determine what aspects of the business they want to measure. Setting these out at the beginning is key to a well-structured database. The business sets out its requirements and the development team will use these to model the dimension tables of the warehouse. Dimensions are the entry points to the data from a BI analysis perspective.

In addition to the dimension tables, the business also needs to know what to measure the dimensions against. This is where KPIs come in.

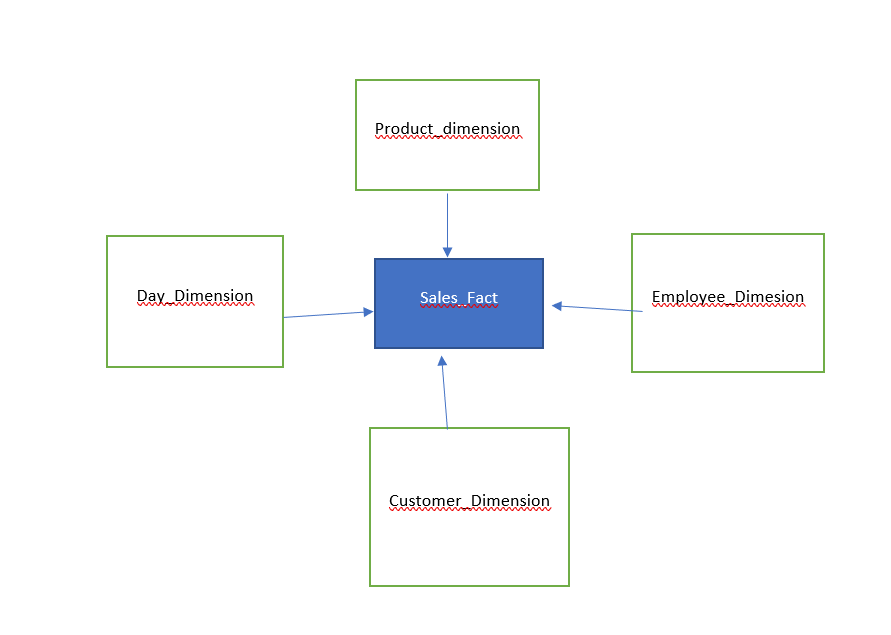
For the Northwind Trading company I wanted to analyse the performance of parts of the company using Sales Revenue. I chose this measure as a trading company typically measures its success against it’s revenue.

From a management perspective I would like to look at the following:

* How are my sales year on year?
* Which products are bringing in the most sales revenue?
* Which products are increasing their revenue year on year?
* Who are my best customers?
* Where are my customers based?
* Which customers have increased their purchases?
* Which of my employees is the most effective sales person?

These questions lead to the design of the warehouse. It is important to ensure that the data is captured without much, if any, duplication.

Using a Star Schema approach, I set out the dimensions: Day, Product, Customer, Employee. As noted above, I used sales to be the measure against which these dimensions would be analysed.



**Extract, Transform, Load:**

Once I had decided on the dimensions I wanted to analyse I had to determine what information I needed to extract from the Northwind Database.

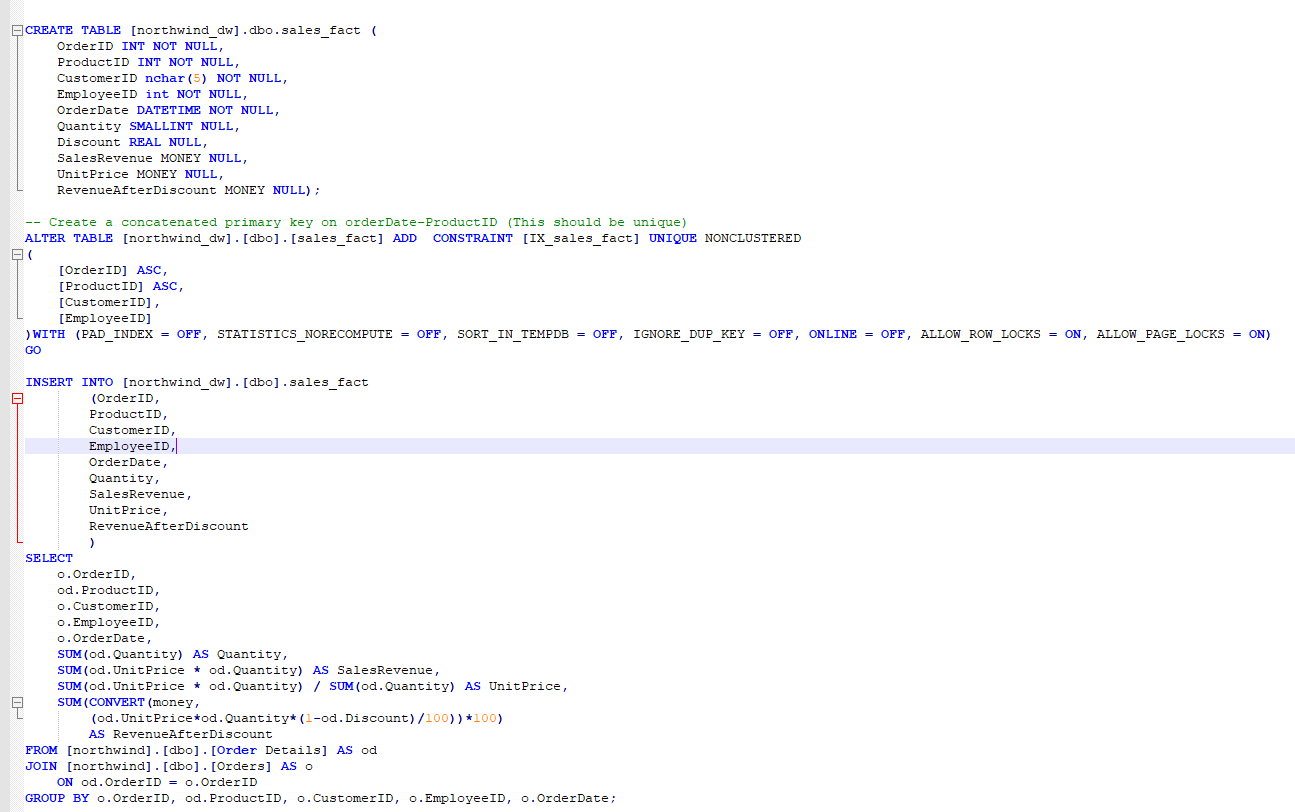
Day\_Dimension – The code for this table created dates for the period of the company’s existence. These dates would be used to compare revenue across time periods.

Product Dimension – For this dimension I determined I needed the following from the Products table on the Northwind Database: ProductID and ProductName. From the Suppliers table I needed the Company name and from the categories table I wanted the CategoryName.

Customer Dimension: I took the following from the Customers table - CustomerID, CompanyName, City, Region, Country.

Employee Dimension: From the Employees table in Northwind.DBO - EmployeeID, FirstName, LastName, HireDate, BirthDate, City, Region, Country.

The sales.fact table was set up using the following code:



The data was extracted from a number of different tables on the Northwind Database and re-labelled before loading into the Warehouse dimension and fact tables.

**BI Dashboards:**

Once the data is loaded into the warehouse the end business user can load that information into a BI tool. There are several BI tools currently on the market. Gartner has assessed the market leaders against the following criteria:

* Agile Centralized BI Provisioning.
* Decentralized Analytics.
* Governed Data Discovery.
* OEM or Embedded BI.
* Extranet Deployment.

Based on their assessment, Tableau and Microsoft are the leaders with companies such as Salesforce and IBM seen as visionaries.

As per Gartner’s latest report Tableau offers a “highly interactive and intuitive visual-based exploration experience for business users to easily access, prepare and analyse their data without the need for coding”. The primary weakness for Tableau is its cost. In comparison with its rivals, it is expensive.

For this assignment I used Microsoft Power BI. According to Gartner, Microsoft “offers a broad range of BI and analytics capabilities with its Power BI suite, delivered via the Azure cloud. Power BI Desktop can be used as a stand-alone, on premises option for individual users, or when power users are authoring complex data mashups involving on-premises data sources. Power BI offers data preparation, data discovery and interactive dashboards via a single design tool.”

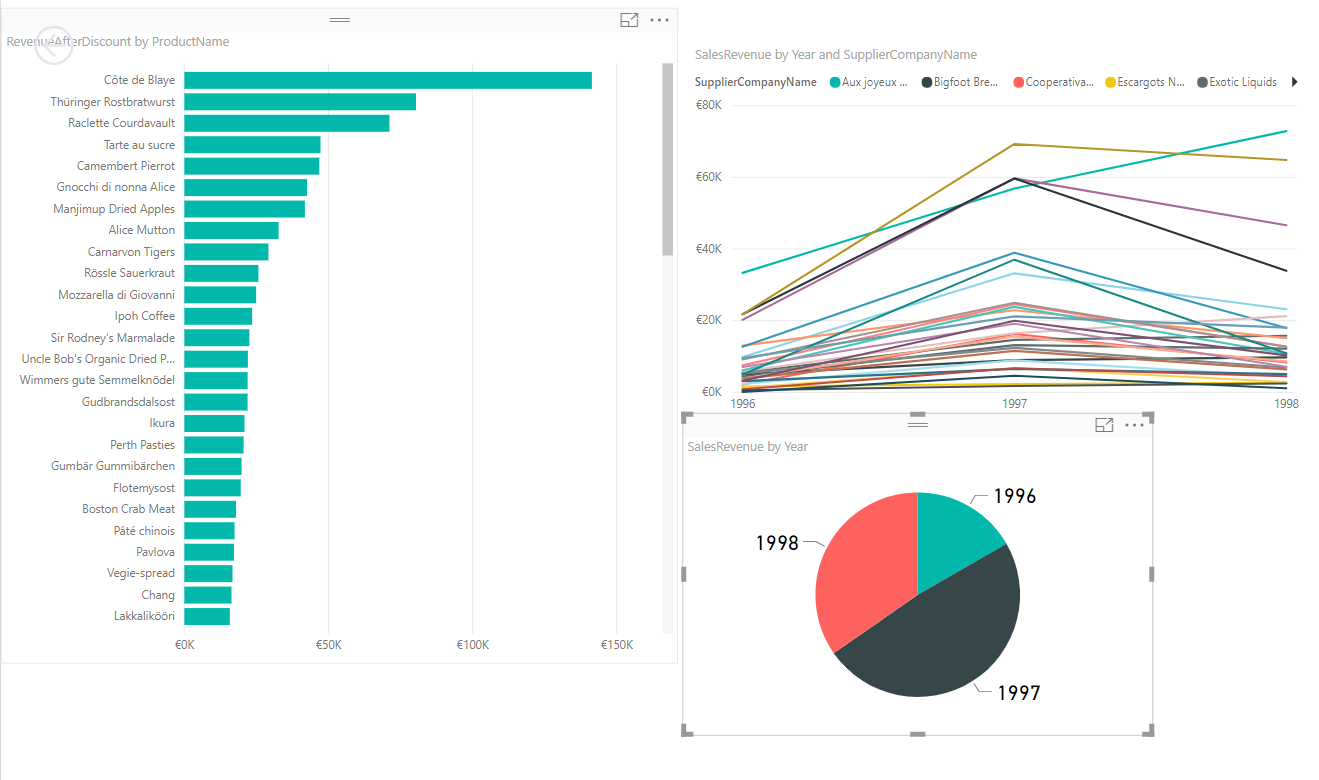
Power BI is free to download and, if you are familiar with other Microsoft products, relatively easy to use. The main weakness with this tool is that is currently only cloud based, and it can take some time to crunch the data.

When creating dashboards, it is worthwhile experimenting with layout and design to ensure the data you are analysing is presented in the clearest way. If possible, design dashboards with multiple devices in mind. A CEO may be looking at the dashboard on their iPhone or tablet, and not on a PC. The data should be clear across all formats.

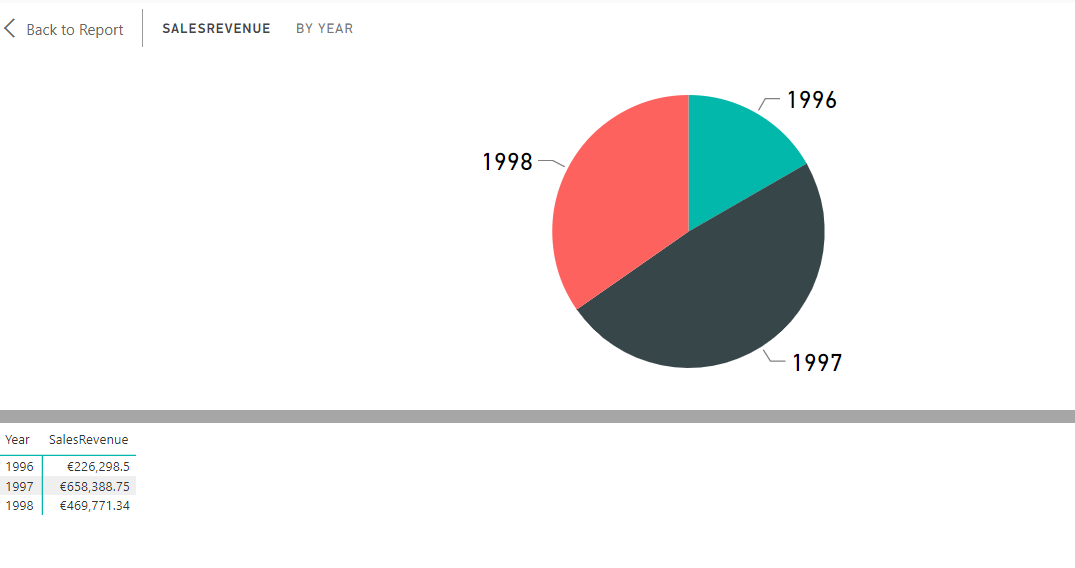
When analysing the information in the Data Warehouse, it’s important to look back to the KPQs and KPIs. For this assignment I picked the following:

* How are my sales year on year?
* Which products are bringing in the most sales revenue?
* Which products are increasing their revenue year on year?
* Who are my best customers?
* Where are my customers based?
* Which customers have increased their purchases?
* Which of my employees is the most effective sales person?

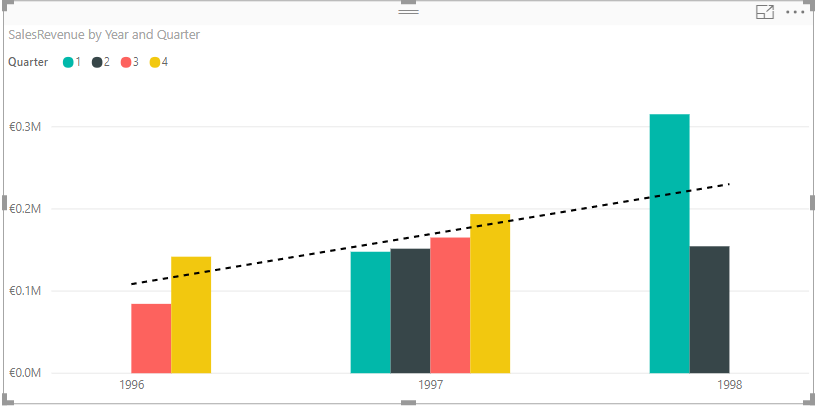
For my first dashboard I wanted to look at the sales by year and product.



With Power BI you can drill into each element of the dashboard to get more information. For example, the Sales Revenue by year pie chart lets me know that 1997 was the best year sales-wise, however I can’t see the actual figures. By Drilling down I get more information:



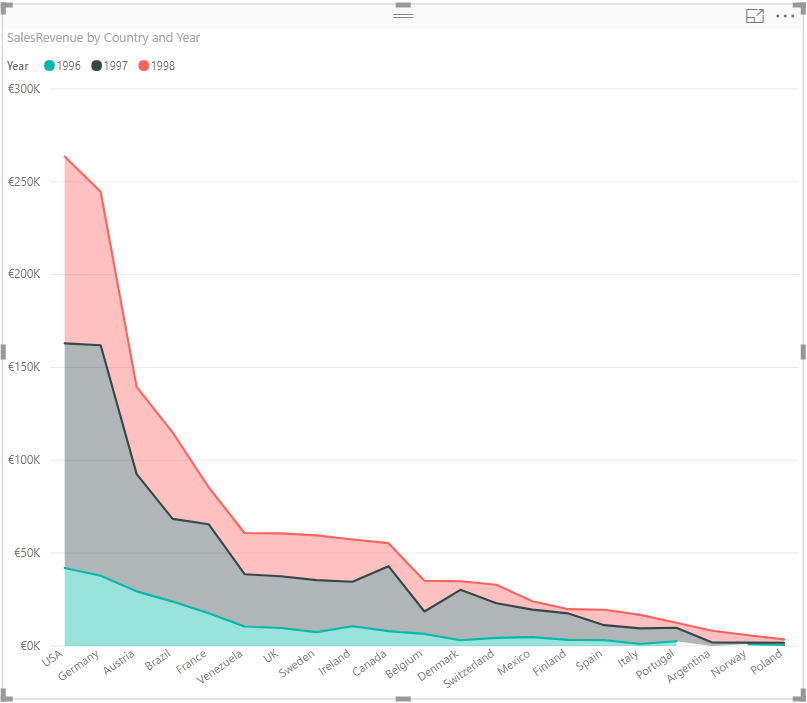
**Analysis by Sales Revenue:**



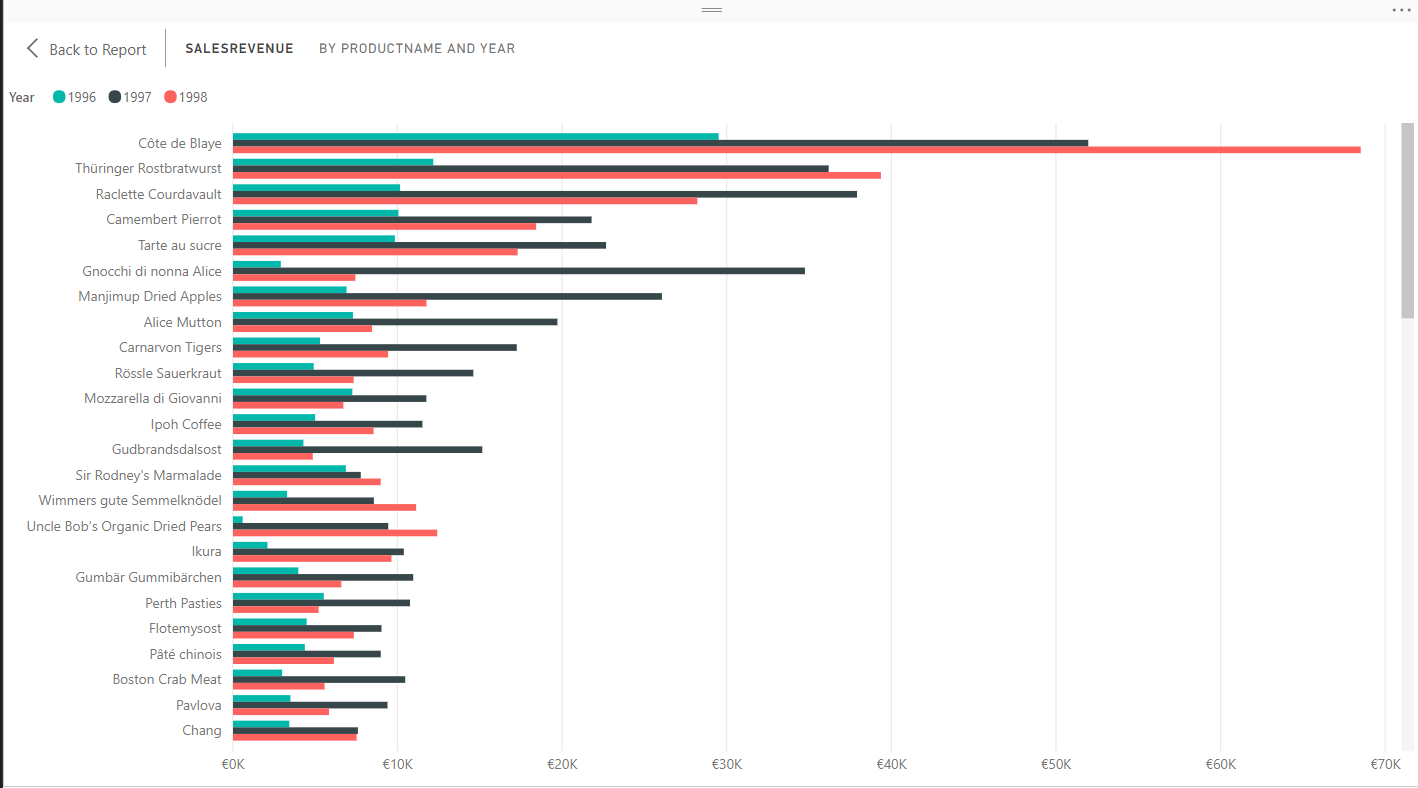


Sales revenue has increased quarter on quarter up until Q2 1998. Q2 ’98 is in line with Q2 ’97. Further data is required to confirm if this Q1 performance is predictive of future revenue, or if it is an outlier.

Based on the Revenue dashboard, the US & Germany are the main markets for Northwind, with revenue consistently high in these countries year-on-year:

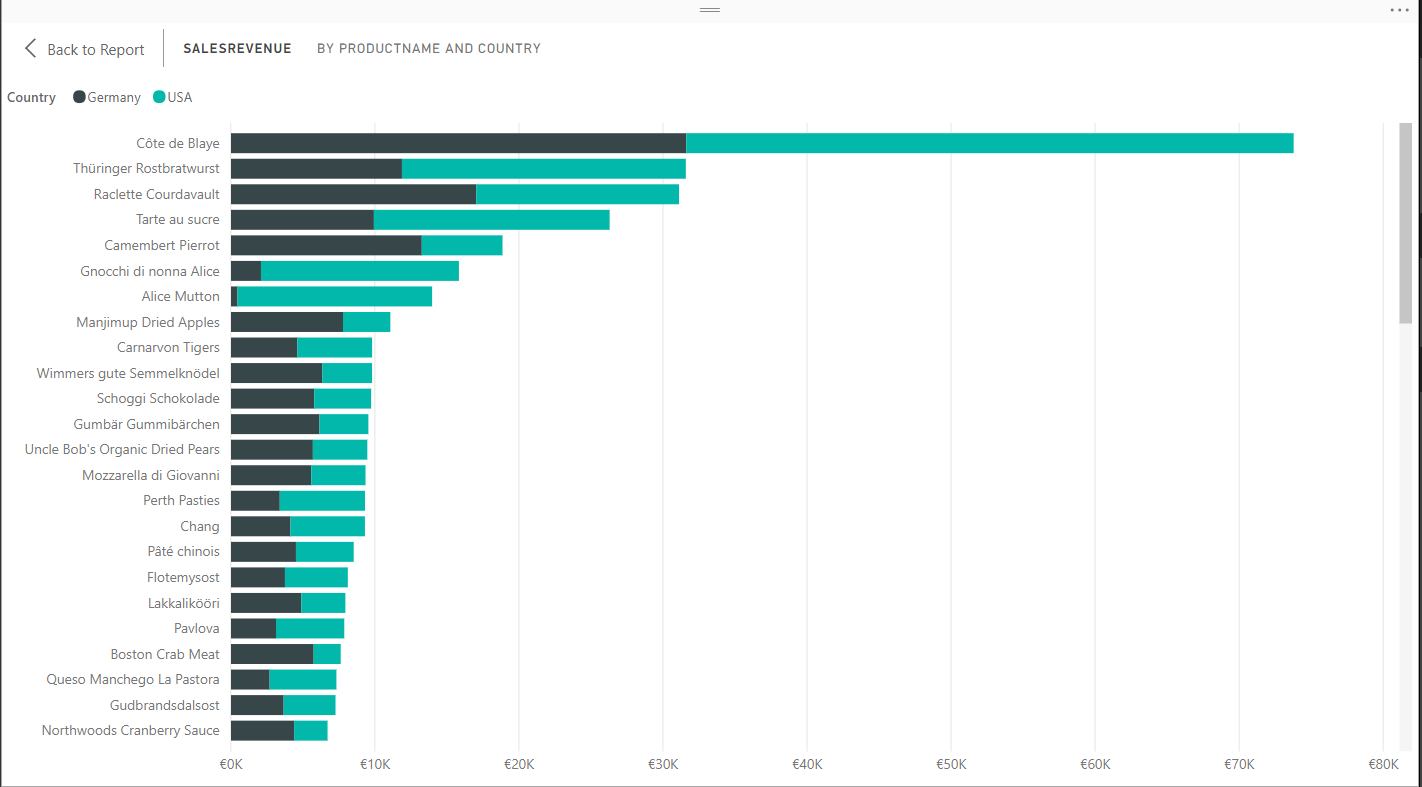


**Analysis by Product:**

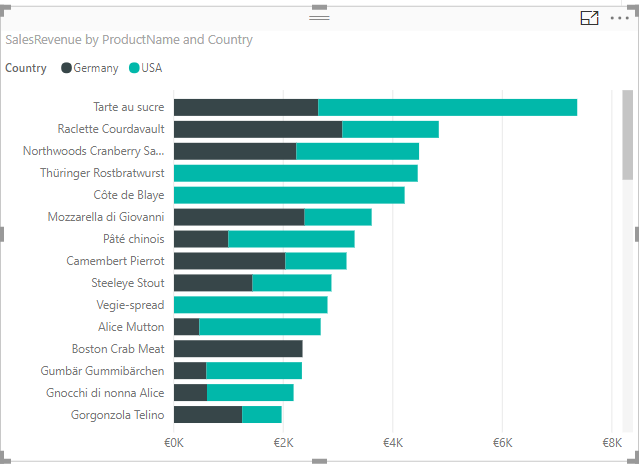


Cote de Blaye and Thuringer Rostgratwurst are the two best selling products with Cote de Blaye selling significant amounts in 1998.

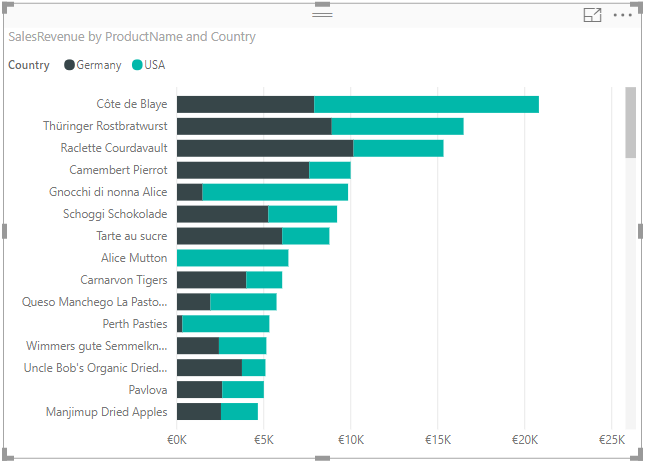
Using the data gathered from the Sales Dashboard, I refined the Product Dashboard to show product sales in the top two markets Northwind sold to.



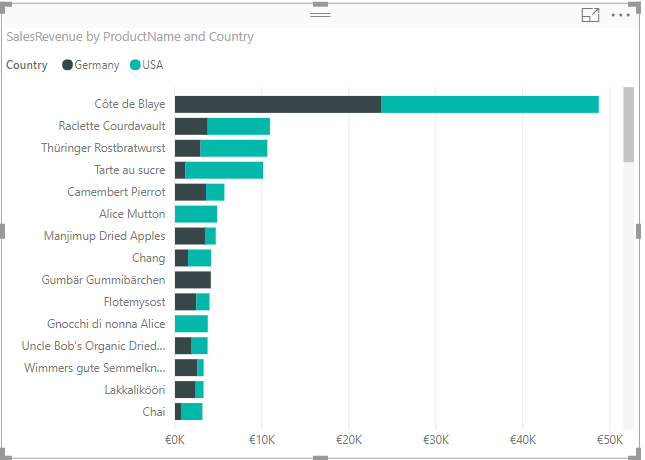
1996 – Top products sold in Germany & the USA:



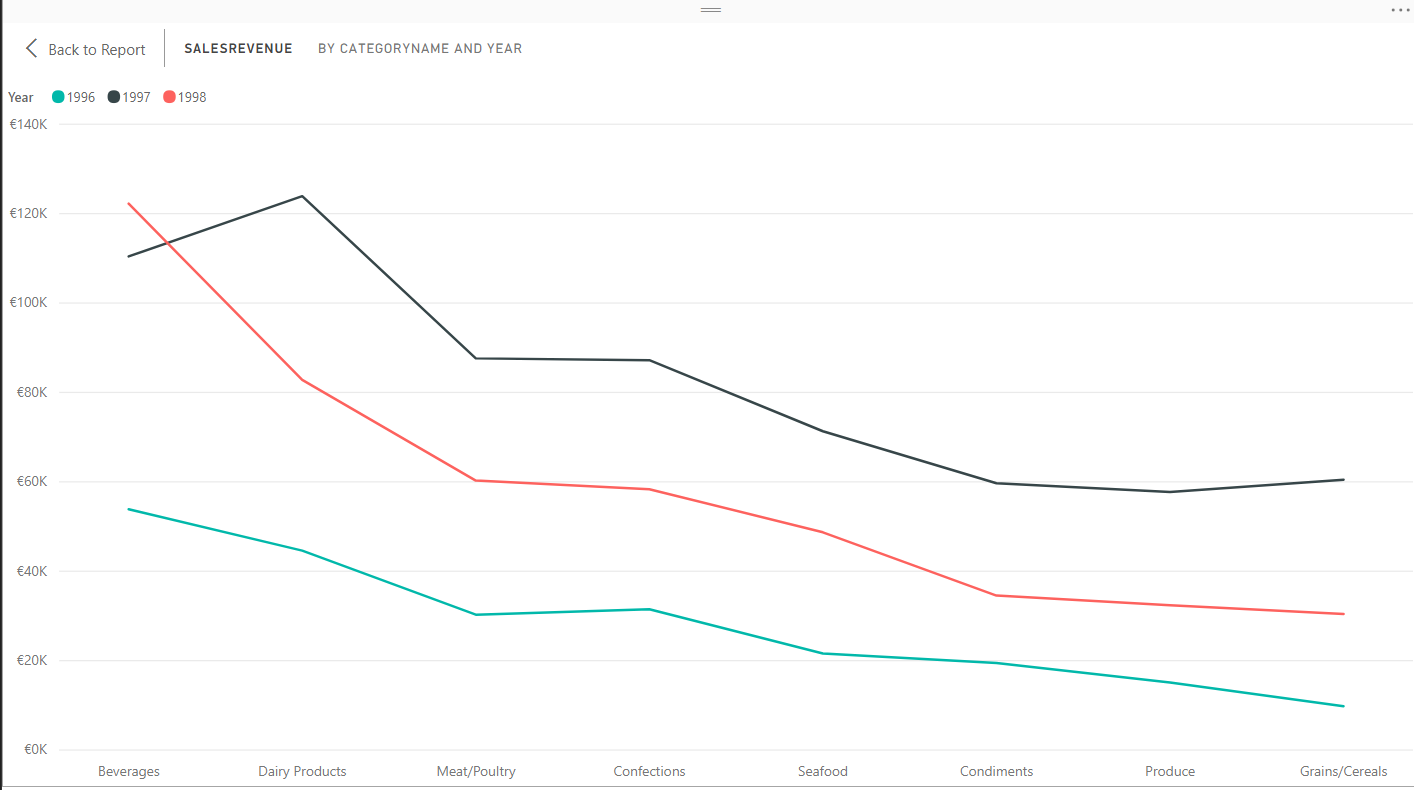
1997 – Top products sold in Germany & the USA:



1998 – Top products sold in Germany & the USA:



Northwind’s sales in Dairy and beverages are the main source of revenue, with beverages seeing a tick up in 1998.

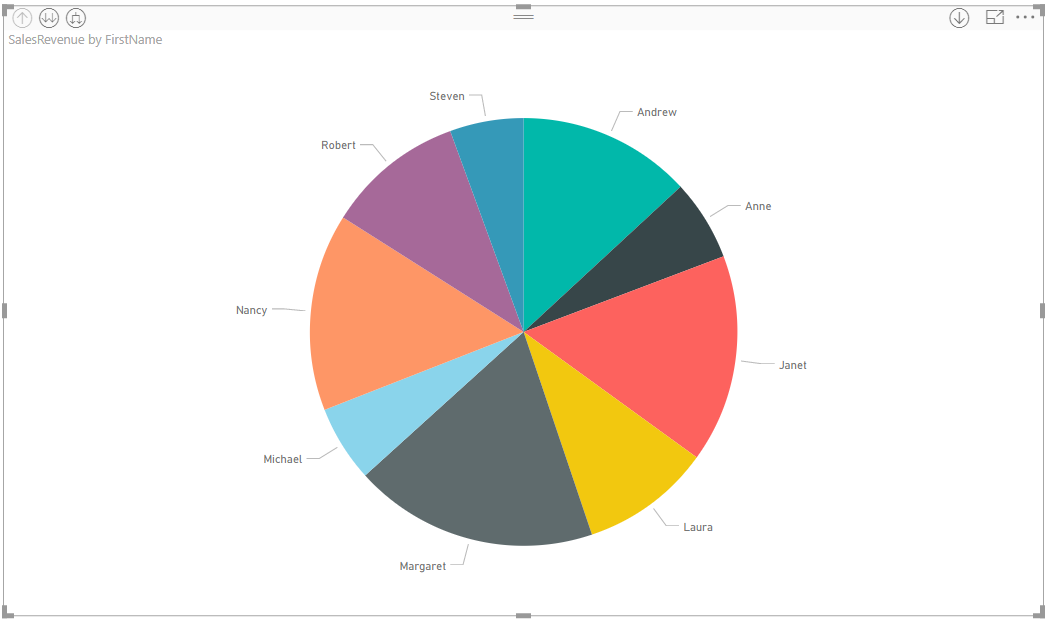


For future quarters, the business should look to increase their business in these areas outside of the core markets of the US/Germany.

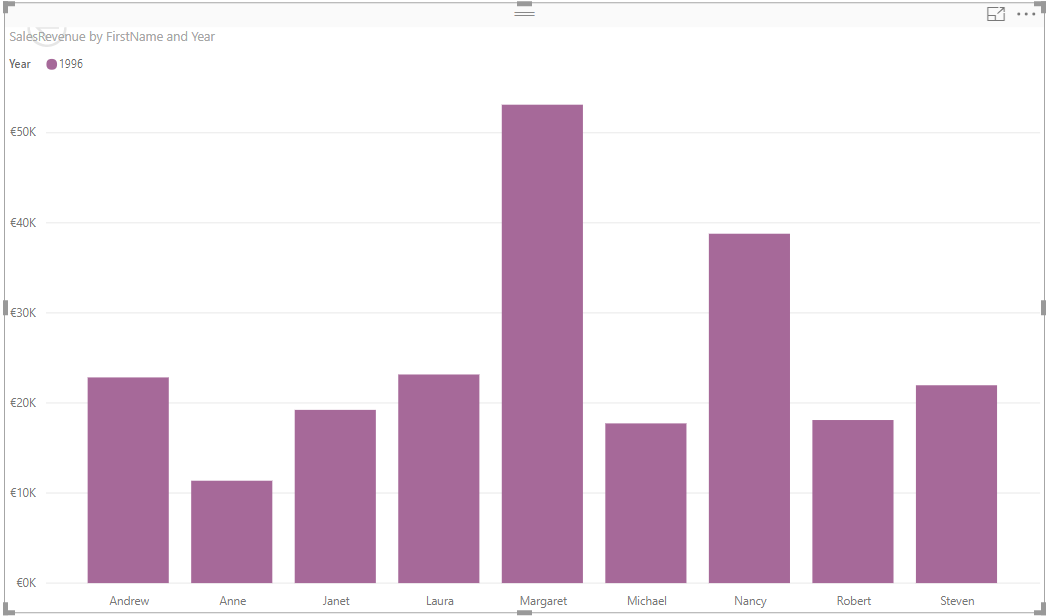
**Analysis by Employee:**

In order to ascertain which employee is deserving of a raise or bonus, an Employee dashboard was created to track sales across quarters:

Sales Revenue by Employee, Inception to Q2 1998:

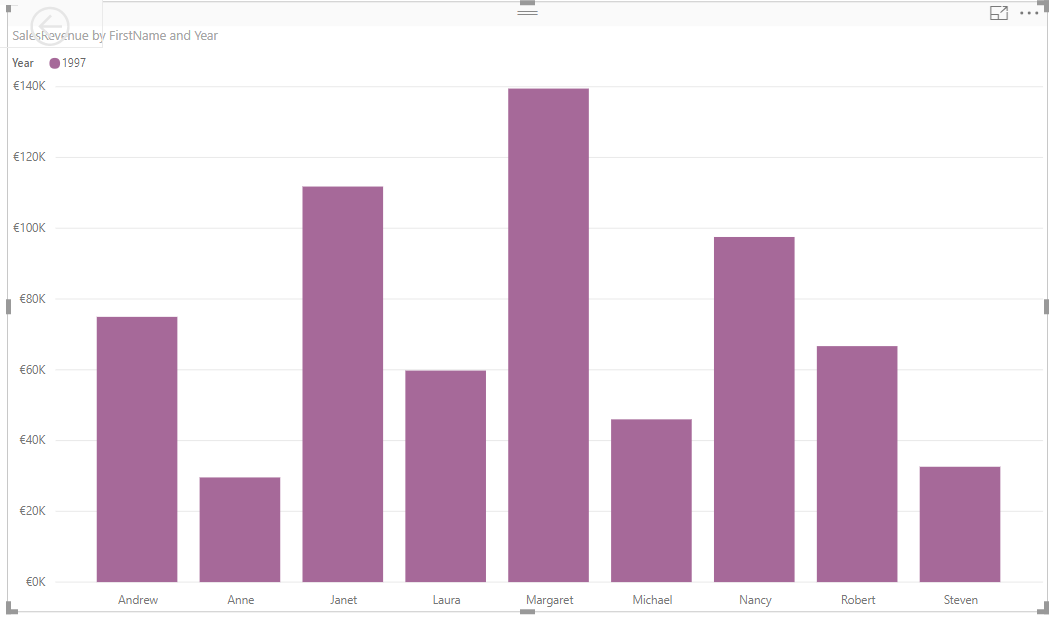


Sales Revenue by Employee, 1996:



From the above data, Margaret is the best salesperson for 1996, bringing in €53,114.80 for the two operating quarters.

Sales Revenue by Employee, 1997:



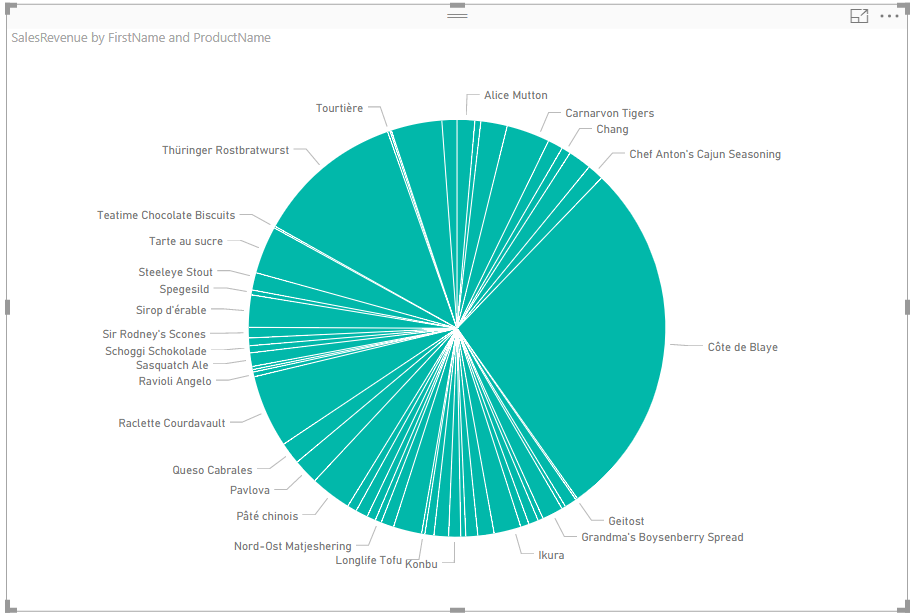
In 1997, Margaret once again brought in the most sales revenue (€139,477.70). Given this consistency, Margaret should be considered for promotion or a sales bonus in 1998.

Looking at the first two quarters of 1998 it is notable that Margaret is no longer the top sales person:

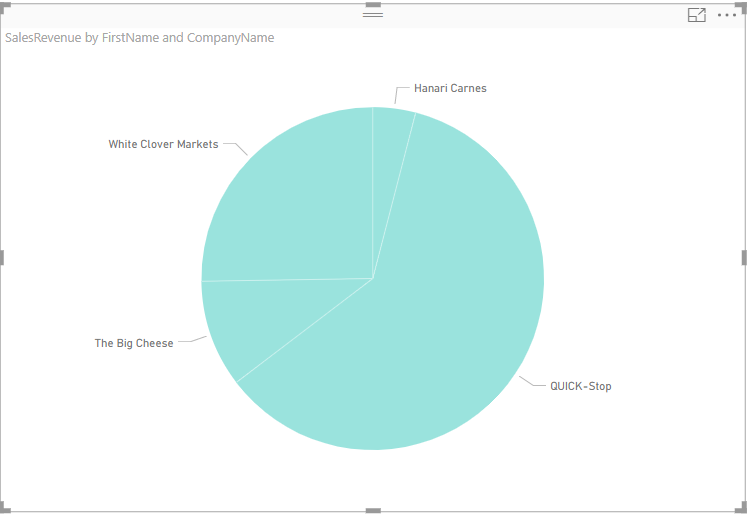
Sales revenue by Employee, 1998 (Q1 & Q2):



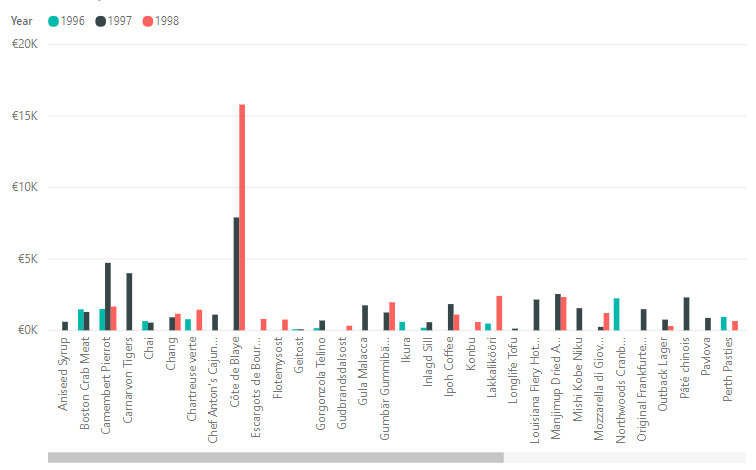
It is possible to cross reference the sales revenue brought in by employees against the products they sold. In doing so for 1998, we can see that Andrew increased his sales in Cote se Blaye:



Andrew sells to the following customers, the largest being Quick-Stop:



If we drill into Quick-Stop’s purchases over the lifetime of Northwind it’s obvious that the increase in sales of Cote de Blaye in Q1 1998 is due to Andrew developing the customer relationship and increasing the sales of this product:



**Conclusion:**

Using the Power BI dashboards, it’s possible to look at the data from the warehouse in a number of different ways and find connections that may not be obvious looking at the day-to-day numbers.

For instance, the increase in sales in Q1 1998 might not have been linked to Andrew without the ability to drill into both the revenue generated by an individual employee and the purchases made by customers.

It is through data warehouses and the complex queries they allow that this information is available for analysis by the business.