DATA WAREHOUSING AND BUSINESS INTELLIGENCE

PRACTICAL PROJECT

by

Daniel Pereira (10391381@mydbs.ie)
Maja Pomohaczi (10392409@mydbs.ie)
Sylwia Wojciechowska (10392584@mydbs.ie)

HIGHER DIPLOMA IN SCIENCE IN DATA ANALYTICS
Module: B8IT104 Data Warehousing and Business Intelligence
Lecturer: Noel Cosgrave (noel.cosgrave@dbs.ie)

DUBLIN BUSINESS SCHOOL
Dublin, Ireland
2019

ABSTRACT

This report will discuss the implementation of a dimensional data warehouse. The concept of a dimensional data warehouse modelling is a technique of analysis and design. It was developed by Ralph Kimball, (Kimball, 2013), in order to help large corporations, organise data and bring it to business users to facilitate decision making. It has been broadly accepted and has become the primary technique for DW/BI presentation.

We intent to discuss a strategy for the design, development and implementation of this structure using various tools available such as Microsoft, IBM and Tableau. We will also, present steps for several modelling techniques, implementation of schemas, data manipulation and visualization in a user - friendly way.

KEYWORDS

Data Warehousing, Business Intelligence, Business Analytics

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Our Team

The below Data Warehousing and BI Project is an outcome of a group work. Our group consists of three members: Dan, Maja and Sylwia. As we all possess a diverse, individual set of skills we could all equally contribute to the project through sharing our knowledge and exchanging our opinions. We divided the workload between all the members, and we met twice a week to discuss our progress. We shared this document so that we can keep each other updated, comment and share ideas.

The purpose of this project is to develop a business plan for a food company and suggest changes needed to improve the company's performance. Our approach to the project is to act as a consulting company that has been hired to perform an analysis, create a data warehouse and draw conclusions for further improvement.

We will endeavour to create a strategic plan for the company while taking into consideration the main business drivers that are associated with every profit - oriented organization. We would like to demonstrate different techniques for creating a data warehouse which will enable us to filter all the relevant information needed to create reports and visualizations. This in turn, will help us to analyse results and bring improvements to the company.

The project will be analysed using the sample Northwind database that provides sales data for Northwind Traders – a mock company that imports and exports specialty foods from around the world. Northwind is a well know relational database created by Microsoft used for teaching, demonstration and practicing purposes. We will use it to construct our functional data warehouse. (Microsoft, 2019)

We have selected this sample database for our study exercise for the following reasons:

i) Simplicity: it is a well know relational database with an easy and simple to understand data schema;

ii) Awareness: the mock structure provides a classic design of a real - world system used for order processing, customer storage and suppliers that most of us will be familiar with;

iii) Academic resources: there are various academic examples of how to explore and analyse the sample data and its schema. We wanted to have some academic background to support our analysis.

This report aims to provide a guide to anyone who wishes to implement a data warehouse using dimensional modelling, however, is was created from the perspective of a study exercise/project only, which means it has its limitations. Considering the high cost involved to introduce a data warehouse, this report gives examples of tools and techniques to be considered by individuals and small business.

1.2 Alta Rica Analytics

We are a consulting company, Alta Rica Analytics that specializes in analysing companies' performance, identifying their downsides and proposing an action plan that will help them achieve an operational excellence. We have been hired by the Super Cool Foods as a consultancy company to carry out a thorough analysis of their overall performance with a special emphasis on their financial and sales activities and help them achieve their goals.

Nowadays, all organizations produce huge amounts of data, which is stored on a so – called source systems. They use this data and reports as part of their decision - making processes. The data contains important information on customer behaviours, sales performance, operation gaps and many more. Using a dimensional data warehouse allows businesses to access all data quickly and in a user-friendly interface. As a result, we decided that the optimal solution for this company would be implementing a functional data warehouse that will support the analysis and reporting processes as well as any improvements that can be made.

Before we could outline the business plan, the management of the Super Cool Foods had to answer pre-project questions (PPQs) in relation to what they are looking for, where they are now and where they would like to be in the next five years. These questions were the basis of our project proposal and helped us create a strategic business plan. Based on the company's vision, mission and main goals a balanced scorecard and KPI's were created. Then, we built and implemented a data warehouse to obtain a quick and easy access to all the data needed for reporting. Once we had all that information, we could start analysing the data provided. Finally, we drew conclusions and were able to propose any changes to improve the company's position on the market.

2 BUSINESS DRIVES

As the consultants responsible for this report, we will analyse sales as the main business driver considering the effectiveness of sales people, numbers of products sold and identifying the best suppliers.

Broadly defined sales are essential for the business to grow and they are a good indicator of the general company's performance. By analysing the facts and figures we can identify the potential problems as well as the opportunities.

2.1 Company's Mission

"The overarching mission of the organization provides the starting point by defining why the organization exists or how business unit fits within a broader corporate architecture. The mission and the core values that accompany it remain fairly stable over time." (Kaplan & Norton, 2004)

The Super Cool Foods aim to provide excellence and quality in gourmet food to a wide range of customers as well as high quality service and satisfaction to our customers and suppliers.

2.2 Company's Vision

"The organization's vision paints a picture of the future that clarifies the organization's direction and helps individuals understand why and how they should support the organization. In addition, the vision set the organization in motion, from the stability of the mission and core values to the dynamics of strategy, the next step in the continuum." (Kaplan & Norton, 2004)

Our goal is to continually improve and expand our product range and create exceptional customer service.

2.3 Key stakeholders

Key stakeholders are all the individuals or groups of people who have an interest in an organization. Their individual interests can vary and so be conflicting at times. The business decisions of a company will be largely determined by their actions and influences. Stakeholders' support and engagement is essential in creating effective management.

We can divide stakeholders into:

- **Internal:** employees, managers, directors, trade unions, shareholders, etc.
- **External:** suppliers, customers, competitors, general public, etc.

The main key stakeholders in the Super Cool Foods that we'll be focusing on are customers, employees and suppliers.

i. *Customers* are our major stakeholder group. Their main interest in the company is the quality of goods and services available and the process and payment terms. We must constantly work to increase the business knowledge and understanding of this group to enable effective collaboration. We strive to maintain the trust they place in our products and brands and ensure that we offer good products consistently.

ii. *Employees* are another crucial stakeholder group. They are the most valuable resource in an organization. Some of their main interests in the organization would be the security of employment, remuneration, responsibilities, career opportunities,

working conditions and improvement of skills. In order to provide excellent customer service, we have to offer them attractive terms and conditions in order to retain them.

iii. *Suppliers* are carefully selected according to their reliability, quality, range of products and pricing. Some of their main interest in the company are the size and value of contracts and the speed of invoice payment. Maintaining our relationship with our key suppliers is essential to provide a successful business operation.

2.4 SWOT Analysis

One of the strategic, powerful tools for analysing an organization, its performance and position against the external environment is SWOT analysis. SWOT stands for strengths, weaknesses, opportunities and threats. "A basic SWOT analysis considers how an organisation can build on its strengths and take advantage of its opportunities, while minimising its weaknesses and avoiding threats." (Claire, 2004)

By providing a summary of the important issues faced by a company it allows to develop a strategic plan that will help to grow the business.

Strengths usually refer to all the features that make us unique and better than the competitors. We have identified the following strengths of the company:

a) The worldwide presence. Operating in numerous markets around the world the company has a strong global presence that can reduce business risks.
b) Strongly established brand name and reputation for good quality. It sets the company apart from other food companies and enables it to retain the customer base.
c) Provision of a wide range of high - quality products. The company fulfils the demand from customers for high quality products. Those customers are willing to pay a bit more for the extra quality. Being able to offer premium quality products gives the company a competitive advantage over other companies in the industry that don't provide that types of products.

Weaknesses present areas of the organization that need to be improved. The weaknesses identified are:

d) Small range of customers. The number of customers is not proportionate in relation to a large number of suppliers that they currently have.
e) Small product range. The limited choice gives the competitors an advantage in the market which means that Super Cool Foods cannot compete with other leading companies in the industry. It also results in high transport costs.
f) Large number of suppliers which makes it very difficult to build strong and reliable relationships. It also poses a challenge because the company is not in a position to negotiate better prices and the delivery costs are much bigger.

Opportunities are external factors that we can use in our favour and achieve competitive advantage. The opportunities presented are:

g) Social trends. The company needs to explore the current trends and pay attention to the growing consumers' health consciousness and changing food habits. Adapting to those changes and looking for ways to benefit from the new trends can help attract more customers.

h) Exclusivity agreement with supplier. Explore the option of becoming the main supplier for a food producer or become a leader in a certain food type for example a high - quality pâté.
i) New technologies. Respond to the constant changes and development of new technologies in order to fulfil the customers' demands and expectations.
j) New markets. Look for new potential retail stores, restaurants that will be interested in the company's range of products.

Threats often come up from competitors or factors outside of the company's control. They can potentially harm and damage the performance of an organisation in the marketplace. Companies need to look for ways to minimise the negative effects of them to a greater extent than their competitors. The threats faced by the Super Cool Foods are:

k) Increasing competition and lower competitors' prices. In order to improve sales, we have to maintain the quality but also offer more competitive prices and that entails provision of changes to our relatively small product range and large number of suppliers.
l) Changes in legislations and taxation relating to the food industry. Reacting to emerging challenges for example Brexit, use of plastic packaging, etc.
m) Contamination or defects in the products. It can hugely affect the image on the company and the trust of its customers.
n) Social trends. the company needs to keep up to date with the current changes in food habits and consumers' buying behaviour.

Once we perform the SWOT analysis, taking into consideration the internal and external environments we can set the final, realistic and achievable objectives of the company.

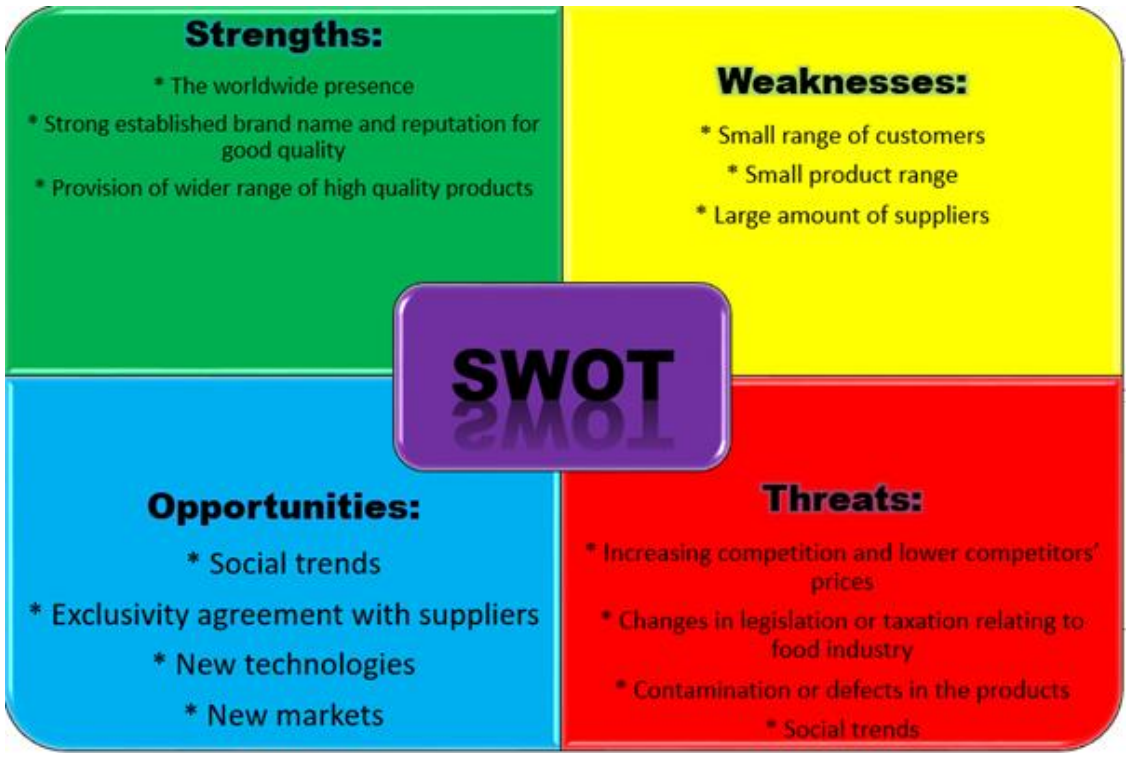

Figure 1 - SWOT Analysis of Super Cool Foods

2.5 Company's Goal

The main aim of Super Cool Foods is to increase revenue by 15% compared to last year. Our recommendation is to increase the customer database and keep 90% of customers active, which means that they will place an order at least once a month. We will also aim to increase the customer pool by 10% and reduce the cost of shipment by 5%.

2.5.1 Strategic Plan

Designing and building the Data Warehouse will help the company to complete their mission and strategy as well as achieve their main goal – an increase in revenue. We will accomplish that by understanding the financial performance of our company. We will be able to take out all information in relation to sales performance based on the product, region, customer or employee, etc. then analyse all the details and draw conclusions. Based on the knowledge gained from the sales performance reports we can determine areas of improvement and implement any necessary changes needed to excel in our business sector. Based on our main goals we will set up our Balanced Scorecard and KPI's.

2.6 Balance Scorecard

Balanced Scorecard is the strategy planning, management and performance measurement tool which was introduced first in 1987 by Arthur Schneiderman. As Robert Kaplan and David Norton broadly disseminated the term of The Balanced Scorecard in 1992 in the article in Harvard Business Review "The Balanced Scorecard: measures that drive performance", they are considered the founders of the present Balance Scorecard. Kaplan and Norton conducted some research study to evaluate and understand new methods for measuring performance which they based on the balance of financial and nonfinancial perspectives. (Kaplan & Norton, 1996) and (Balance Scorecard Institute, 2019).

As a management tool, Balanced Scorecard provides "a formalism, methodology and framework that translates strategy to actionable and measurable objectives" (Nair, 2004). Kaplan and Norton describe Balanced Scorecard as follows: "The Balance Scorecard translates mission and strategy into objectives and measures, organized into four different perspectives: financial, customer, internal business process, and learning and growth. The scorecard provides a framework, a language, to communicate mission and strategy; it uses measurements to inform employees about the drivers of current and future success." (Kaplan & Norton, 1996)

The Balanced Scorecard should be used by companies to communicate what they would like and try to achieve, to align the everyday tasks that people in an organisation do with the main strategy, to prioritize projects and finally, to measure and monitor the progress of accomplishing the goal.

According to Kaplan and Norton's assumptions we should consider our company from four different perspectives and develop objectives and measures (KPI's) as well targets. Then, we can establish the appropriate actions that need to be undertaken.

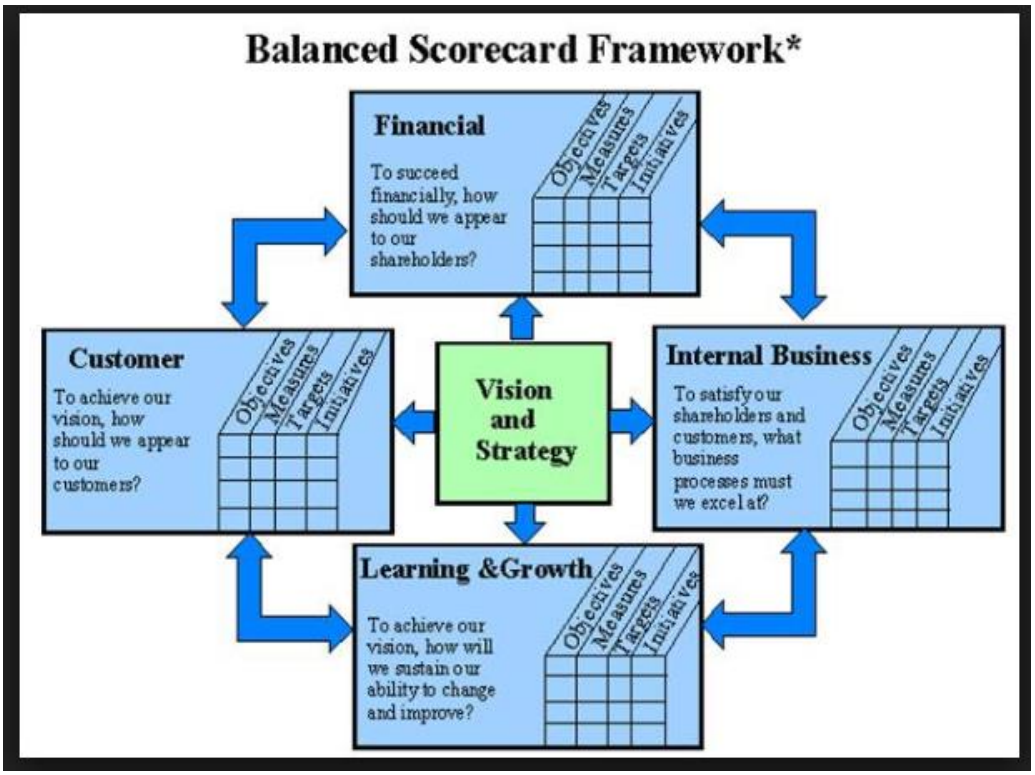

Figure 2 - Balance Scorecard Framework

*The Balance Scorecard Framework[1]

The four perspectives can be either financial or nonfinancial and all of them equally contribute to the main vision, mission and strategy of the organization. Kaplan and Norton identified four perspectives known as the four traditional perspectives: financial, customer, internal and learning & growth perspective.

Financial perspective - in order to properly define that perspective, the organization needs to establish first the financial goals that must be accomplished to achieve the overall strategy. To measure it we need to take into consideration an increase in funding from different sources, reduction of operational costs and improvements in the usage of assets.

Customer Perspective – to outline this perspective we need to specify what objectives relating to customers must be reached as well as what the company needs to do to gain customers' acquisition, acceptance and preservation. The measures, among others, include the capability of the organization to provide quality goods and services, the effectiveness of their delivery, overall customer service and satisfaction from customer perception.

Internal Processes Perspective – to define it we need to know what measures have to be implemented to improve the company's performance in relation to its customers and other stakeholders. In every organization, internal procedures, culture and processes must be in place to ensure that the customer and financial objectives are accomplished. The business needs to identify the key processes at which they need to excel. The measurement includes core competencies, productivity and service delivery.

Learning and Growth Perspective – is the base for all the previous perspectives. It reminds us that without skilled, motivated and satisfied employees as well as high quality information systems the internal, customer and financial perspective goals cannot be achieved. This perspective is about how people absorb new ideas and turn them into actions.

[1] The Balance Scorecard Framework figure was adapted from Robert S. Kaplan and David P. Norton, "Using the Balanced Scorecard as a Strategic Management System", Harvard Business Review (January-February 1996): page 76.

Our recommendations for the Super Cool Foods in relation to the four perspectives outlined above are as follows:

i. Financial Perspective:
   a. Objectives: increase the revenue and profitability;
   b. Goals: increase both of them by 15%;
   c. Indicators / measurements: financial statements;
   d. Initiatives: negotiating better prices with suppliers and reducing shipment costs;

ii. Customer Perspective:
   a. Objectives: maintain a high number of active customers and increase the average customer rating;
   b. Goals: increase the number of customers who place an order at least once in two weeks by 5% and increase the value of minimal order by 10%;
   c. Indicators / measurements: an invoice value per sale and a monthly count of orders per customer;
   d. Initiatives: improve and increase the product range and set up special seasonal promotions.

iii. Internal Processes Perspective:
   a. Objectives: improve delivery efficiency and reduce shipment delays;
   b. Goals: decrease the delivery waiting time and reduce the amount of human error;
   c. Indicators / measurements: monthly count of days per shipment per customer;
   d. Initiatives: automate the order processing by introducing barcode scanners, implement online order service and redesign the customer service processes.

iv. Learning and Growth Perspective:
   a. Objectives: increase the knowledge of customers and employees in relation to products and promotions and maintain a skilled sales team
   b. Goals: train all the sales team in relation to sales techniques and knowledge of the products; regularly inform customers
   c. Indicators: quarterly knowledge test
   d. Initiatives: online customer service courses for the sales team, lunches and learning events

2.7 Strategy Map

Once we outlined the four perspectives and the balanced scorecard, we created the strategy map, which will help the organization to achieve its main goal and vision.

Strategy map is a diagram which represents the relationships and links between the different perspectives and their contribution to achievement of the main financial goal.

According to Kaplan and Norton “The strategy map describes the logic of the strategy, showing clearly the objectives for the critical internal processes that create value and the intangible assets required to support them” (Kaplan & Norton, 2004). The strategy map helps to set up actions and initiatives which will allow the objectives and targets to be accomplished. By illustrating the cause and effect relationship, the business can set up the action plan and implement further changes to the organization.

Presented by Kaplan, the strategy map (see Figure 3) provides a normative checklist of strategy’s components and interrelationships. The strategy map provides the missing link between a strategy formulation and a strategy execution.

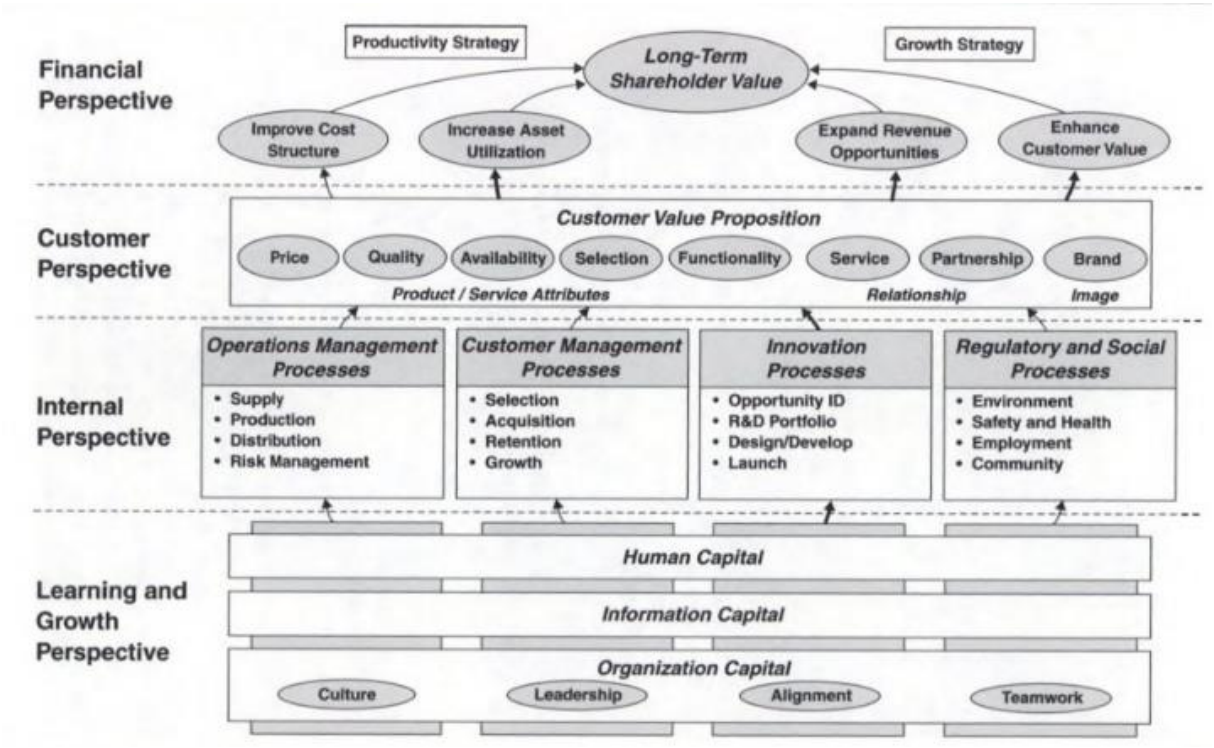

Figure 3 - A Strategy Map Represents How the Organization Creates Value (Kaplan & Norton, 2004)

Based on the above guidelines, we have created a customized Super Cool Foods' strategy map (see Figure 4), that describes how to enhance the performance of internal processes and deliver value to the identified stakeholders.

Figure 4 - Strategy map created for Super Cool Foods

2.8 Key Performance Questions

The management of Super Cool Foods needs to define the Key Performance Questions in relation to the strategic objectives, to identify the information need. The KPQs help to develop better and more important KPIs.

a) How well are we fulfilling our customer requirements?
b) How could we build and develop our supplier relationship?
c) How could we improve our customer relationship?
d) How well are we taking control of our budget?
e) To what extent are we improving our costs in relation to product delivery?
f) How well are retaining loyalty among our most beneficial customers?
g) How well are we using new technology tools?
h) How well are we managing our supply chain?

2.9 Key Performance Indicators

Key Performance Indicators (KPIs) are metrics of the company's progress in relation to achieving the anticipated results as well as the strategic and operational goals. They are measurable values that demonstrate the effectiveness of the company in achieving critical objectives. (KPI.org, 2019)

Based on the information provided by the management of Super Cool Foods in relation to Key Performance Questions we established the below KPIs:

i. Sales/ Financial KPI's:
    a. Target sales per person – the sales target needs to increase by 15% in comparison to the sales value of the same month from the previous year (as well as quarterly, half-yearly and yearly). This needs to be measured on a monthly basis;
    b. Active Customers/Customers Retention – increase the number of customers who place an order at least twice a month by 5%; it will be measured on a fortnightly basis;
    c. Active customers – to extend the pool of active customers by 10% within three months timeframe;
    d. Product Sales Target - based on the season / holiday periods – to increase sales of selected products by 35%, based on the time of the year; it will be measured quarterly;
ii. Customer Satisfaction KPI's:
    a. Shipping efficiency – decrease the delivery waiting time; this will be measured by the number of days needed to complete the order on a monthly basis;
    b. Product availability – decrease the amount of time the product is out of stock by 15%; it will be measured on a monthly basis by counting the cancelled orders;
    c. Product returns – decrease the number of returned products due to the poor quality or package damage – less than 5% per month;
    d. Customer Support / Service – improve the level of communication, making sure the customers receive notifications in relation to any product shortage and delivery delays; it will be measured through customer satisfaction surveys on a monthly basis;
iii. Internal Processes KPI's:
    a. Minimum Order Volume – set the minimum order volume to be at least 15 items; it will be measured on a monthly basis;

b. Shipping costs – to reduce shipment cost by 3% in comparison to last year, taking into consideration petrol prices fluctuations; it will be measured on a monthly basis;
c. Workforce – adjusting the workforce depending on the volume of orders; if the orders decrease, reduce the number of staffs; this will be measured quarterly;

iv. Learning and Growth KPI's:
a. Employees – increase an individual employee's sales performance by 15%; it will be measured on monthly basis;
b. Employee satisfaction - improve the work – life balance and satisfaction; this will be measured by employee satisfaction survey on a quarterly basis;

3 DATA MODELLING

3.1 Design Pattern

We will talk about different design patterns of implementing a warehouse and discuss some aspects that refer to a dimensional warehouse such as: business (process), logical (dimensional) and physical models.

We know that the design phase of a warehouse implementation is extremely important. The data model helps to understand and even visualize relationships among the data. As, Ralph Kimball suggests, “The dimensional model is extremely robust. It can withstand serious changes to the content of the database without requiring existing applications to be rewritten” (Kimball, 1996).

Bill Inmon (Inmon, 2005, p. 29), considered by many to be the father of data warehousing, defines: “A data warehouse is a subject-oriented, integrated, non-volatile, and time-variant collection of data in support of management’s decisions.”

3.2 Business Model

The business model enables us to understand what information is available and the relationship between the attributes. We will aim to answer some of the below KPQs during the business modelling process:

1. Who are our customers?
2. Who are our sales team?
3. What has been sold?
4. What are the details of the sales?

3.2.1 OLTP vs OLAP

OLTP (online transaction processing) system is a term used for a transaction system that is primarily responsible for capturing and storing data related to day-to-day business functions. It addresses a critical business need for a real-time report with no redundancy (*normalized*) of data but does not serve for ad hoc analysis and time variant comparison reports. (Sharda, et al., 2018, p. 184)

OLAP (online analytical processing) system needs the data to be structured for user-friendly queries and reporting. As a result, the data tends to have a lot of redundancy. This is also called the *de-normalization* process.

The existing data model used by Super Cool Foods is an OLTP system. It was mainly built with the business operations in mind rather than reporting and growth analysis. The **Error! Reference source not found.** below demonstrates the entity-relationship diagram (ERD) with main subject areas and relationships.

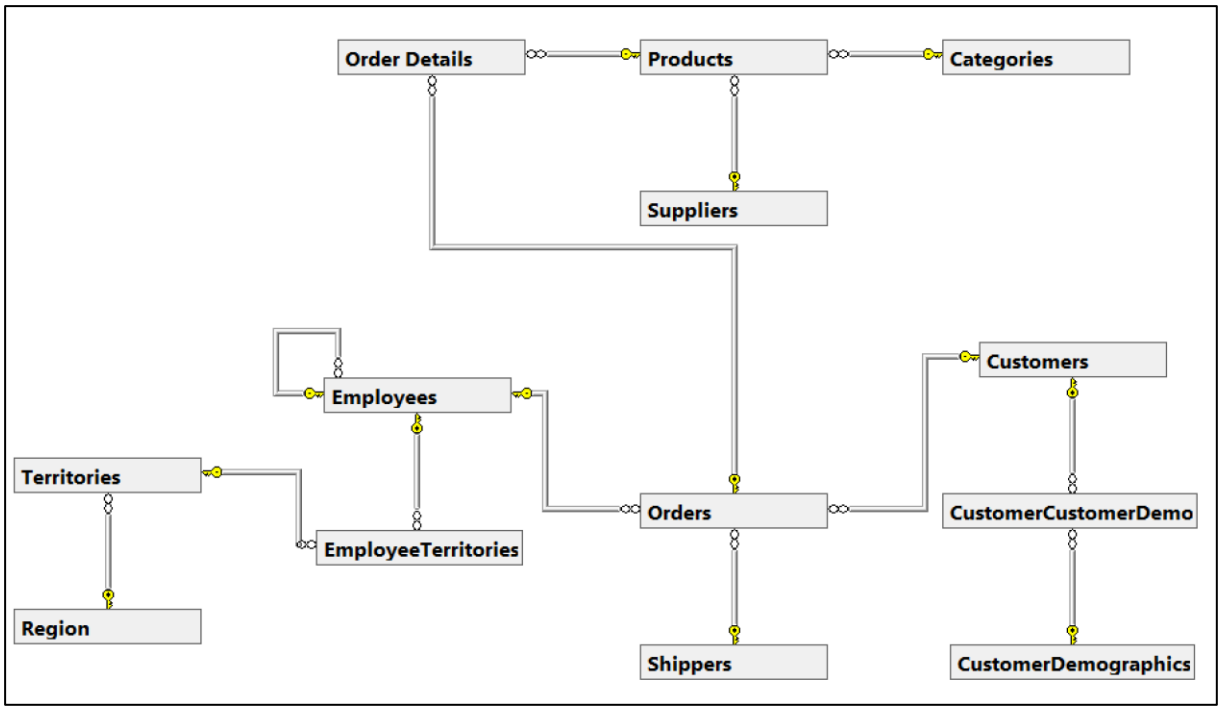

Figure 5 - Northwind Business ERD

The diagram shows the existing relationships found in different tables in the database. It can be noted that the **Orders** table has the most connections. We will use this table as a starting point when creating the dimensional model later.

We have used SQL Server[2] to manipulate the data stored in our source database. Table 1 shows an example of Orders stored and their related IDs.

OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate	ShippedDate	ShipVia	ShipCountry
10248	VINET	5	04/07/1996	01/08/1996	16/07/1996	3	France
10249	TOMSP	6	05/07/1996	16/08/1996	10/07/1996	1	Germany
10250	HANAR	4	08/07/1996	05/08/1996	12/07/1996	2	Brazil
10251	VICTE	3	08/07/1996	05/08/1996	15/07/1996	1	France
10252	SUPRD	4	09/07/1996	06/08/1996	11/07/1996	2	Belgium
10253	HANAR	3	10/07/1996	24/07/1996	16/07/1996	2	Brazil
10254	CHOPS	5	11/07/1996	08/08/1996	23/07/1996	2	Switzerland
10255	RICSU	9	12/07/1996	09/08/1996	15/07/1996	3	Switzerland
10256	WELLI	3	15/07/1996	12/08/1996	17/07/1996	2	Brazil
10257	HILAA	4	16/07/1996	13/08/1996	22/07/1996	3	Venezuela
10258	ERNSH	1	17/07/1996	14/08/1996	23/07/1996	1	Austria
10259	CENTC	4	18/07/1996	15/08/1996	25/07/1996	3	Mexico
10260	OTTIK	4	19/07/1996	16/08/1996	29/07/1996	1	Germany
10261	QUEDE	4	19/07/1996	16/08/1996	30/07/1996	2	Brazil
10262	RATTC	8	22/07/1996	19/08/1996	25/07/1996	3	USA

Table 1 - Sample of Orders table and data

[2] Microsoft SQL Server is a relational database management system developed by Microsoft. We have used the version 2017 of the system with SQL Server Management Studio tool to generate the diagrams and queries.

3.3 Logical Model

Considering the entities and relationships presented in the Business model, we can start analysing the information we would like to extract and finally, find a way to organize all that in a meaningful manner.

The Logical Data Model adds information to the elements, defines the structure and relationships between them.

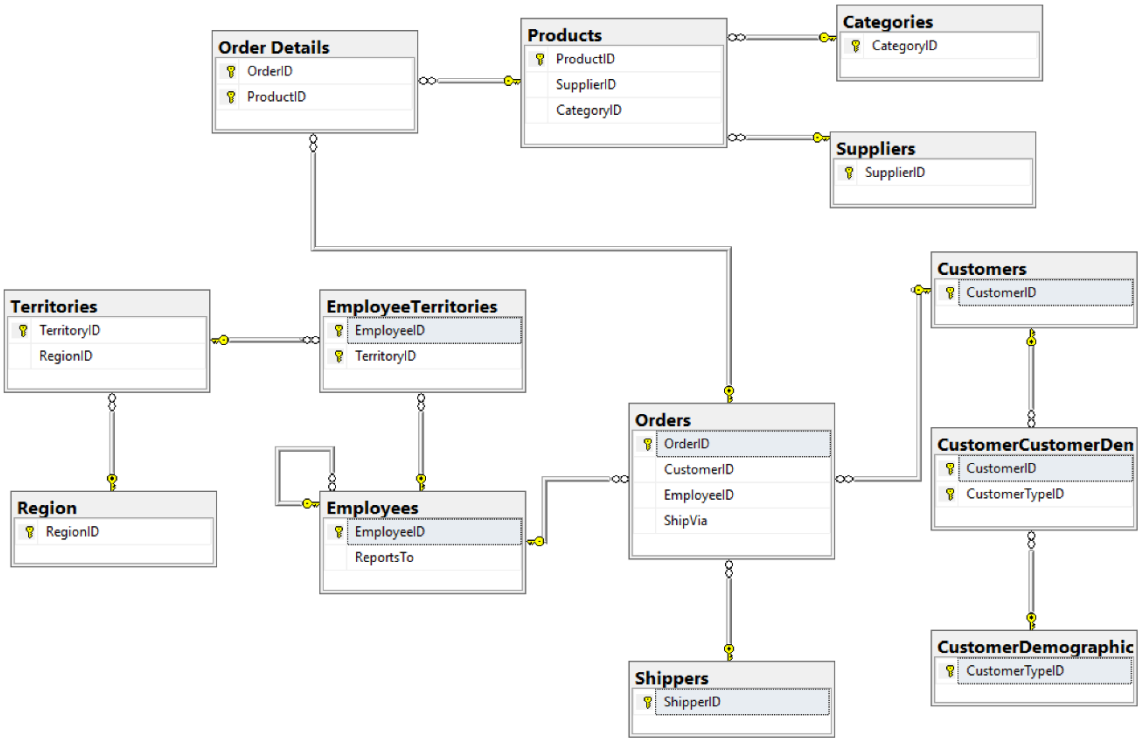

Figure 6 - Northwind Logical ERD

While analysing the logical data model (Figure 6) above, we were able to identify the entity keys and relationships. This step was important in understanding which entities are essential to fulfil the business goals.

3.4 Physical Model

The physical data model is reflected in the entity meta data from the above logical data model. It contains physical properties for each entity including table name, column name, data type and other attribute definitions like primary key, foreign key, null options, etc.

The Figure 7 below demonstrates the physical model of the Northwind database sample. (In the Appendix section, there is the full database schema of the Northwind database used by Super Cool Foods).

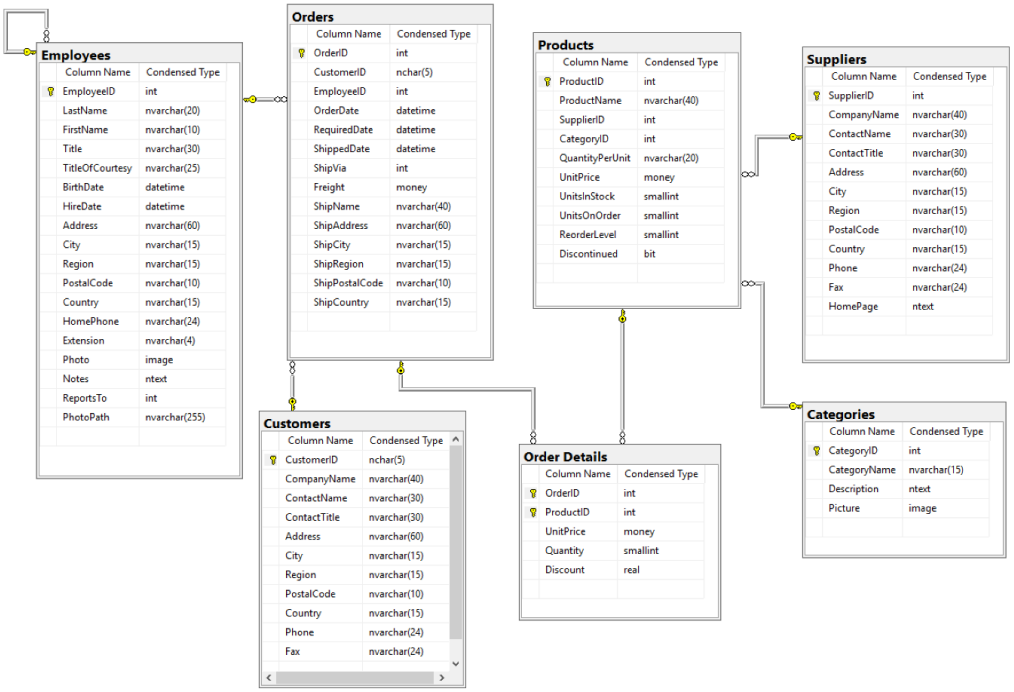

Figure 7 - Northwind Physical ERD

3.5 Dimensional Model

Paularj Ponniah said that the principles of dimensional modelling start with gathering requirements and the definition of those requirements drives the data design for a data warehouse. Data design consists of putting together various data structures. The logical data design includes business elements and establishes the relationship among the data structures, while the physical data design is about how these structures are implemented. The data design process results in a dimensional data model. (Ponniah, 2010, p. 225).

Reviewing the Northwind data model, we noticed three types of data entities:

a) measurements;

b) business dimensions;

c) attributes of each business element;

As a result, when we put together the dimensional model to represent the information for Super Cool Foods, we need to come up with data structures to represent these three types of data entities.

The dimensional data modelling is different than the usual OLTP 3rd normal form. The design has two kinds of tables: fact and dimensional.

3.5.1 Dimensional Table

This table consists of important parameters for business reports that are not measurements. This is also known as a lookup table.

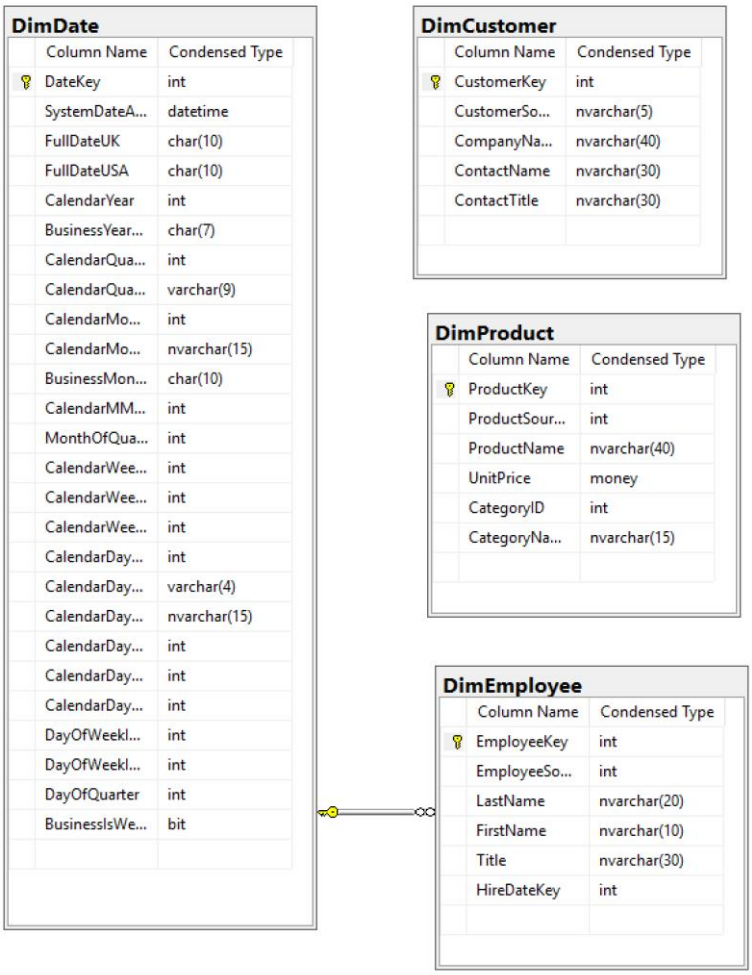

Figure 8 - SernikDW dimension tables

3.5.2 Fact Table

Fact tables contain measures of data or information on facts needed to build an analytical report. They contain a large number of rows that correspond to the observed facts and foreign keys.

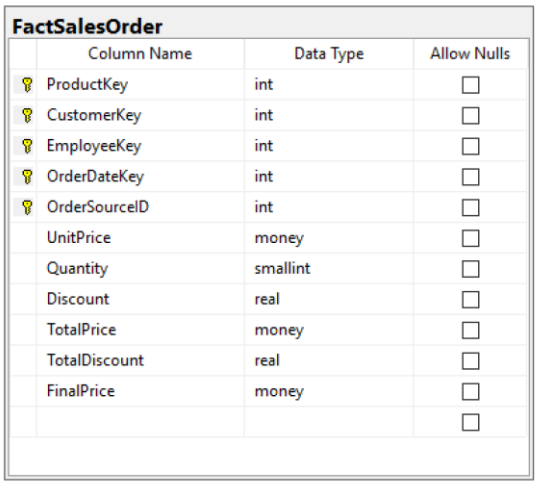

Figure 9 - SernikDW fact table

3.5.3 Star Schema

After defining what the dimension and fact tables are, we started combining them to create relationships. Paulraj (Ponniah, 2010, p. 230), lists the basic criteria for combining the tables into a dimensional model:

- The model should provide the best data access;
- The whole model must be query-centric;
- It must be optimized for queries and analysis;
- The model must show that the dimension tables interact with the fact table;
- It should also be structured in such a way that every dimension can interact equally with the fact table;
- The model should also drill down or roll up along dimension hierarchies;

A star schema contains a central fact table surrounded by and connected to several **dimensional tables** (Adamson, 2006). It is designed to provide fast query-response time, simplicity, and ease of maintenance for read-only database structures.

3.5.4 Snowflake Schema

The **snowflake schema** is a logical extension of the star schema, obtained by normalizing the dimensional tables into multiple tables. The entity-relationship diagram resembles a snowflake in shape, hence the name of the schema. It improves the performance of queries due to small storage requirements, however, it requires more maintenance.

3.5.5 Data Granularity

Granularity represents the level of detail in the fact table. If the fact table is at the lowest grain, then, the users can drill down to the lowest level of detail.

The disadvantage of a fact table at the lowest grain is the amount of storage and high maintenance. Lowest grain means large numbers of fact table rows.

One of the main advantages of granular fact tables, is the fact that they can serve as natural destinations for operational data that are frequently added from operational systems. Further, the more recent data mining applications need details at the lowest grain. (Ponniah, 2010, p. 239).

3.5.6 Time Dimension

Time dimension is an essential part of data warehousing planning, our DimDate dimension is a virtual data to be linked with the OrderDate and the HireDate. The DimDate dimension was created to provide a drilldown reporting feature such as business year, quarter, week number and day of the week. The SQL script for creation the DimDate can be found on Appentix section: Table 3.

3.5.7 The Defined Dimensional Data Model

Considering the earlier identified Business Drivers, we have designed the data dimensional model using the star schema, the order management as our fact table and the granularity as shown in Figure 10. The advantage of that approach is that:

a) we can keep querying simpler as compared to the snowflake schema;
b) there are many examples of the star schema usage on the Internet;
c) the business logic and requirements are simple enough to continue using the star schema.

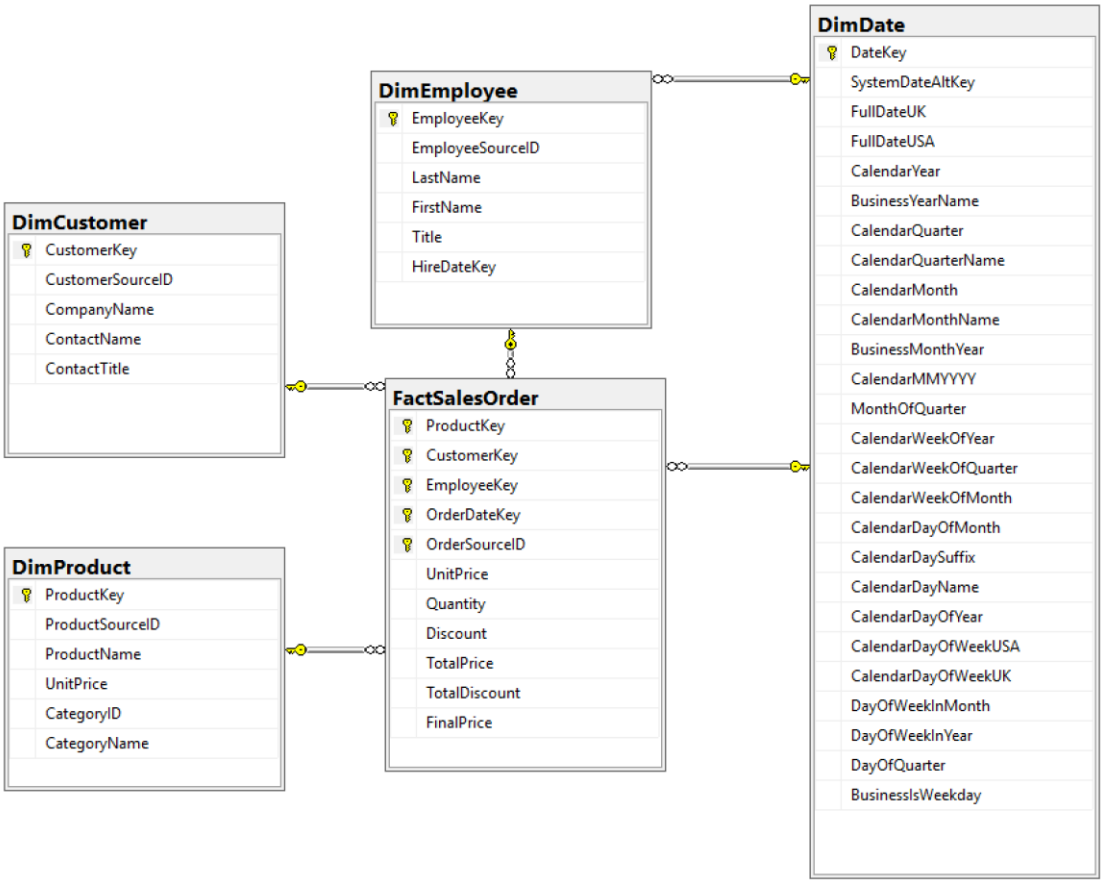

Figure 10 - SernikDW Star schema

3.6 Extraction, Transformation, and Load

"The Extract-Transform-Load (ETL) system is the foundation of the data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions." (Kimball & Caserta, 2004)

Ramesh Sharda described the ETL process as *extraction* (i.e. reading data from one or more databases), *transformation* (i.e. converting the extracted data from its previous form into the form it needs to be. Then, it can be placed into a data warehouse or simply to another database), and *load* (i.e. putting the data into a data warehouse).

The process of loading data into a data warehouse can be performed either through data transformation tools with rich, user-friendly interface or through more traditional methods, such as developing specific programs or SQL scripts. (Sharda, et al., 2018, p. 174)

We have taken into consideration the volume of the data provided by Super Cool Foods, their primary objectives and KPIs and other aspects such as:

a. Specific tools that can be very expensive;
b. Specific tools that may have a long learning curve;
c. Specific tools that may request big infrastructure.

4 IMPLEMENTATION

4.1 Database Management Systems

4.1.1 SQL Server

Microsoft SQL Server is a relational database management system (RDBMS) with the primary function of storing and retrieving data. This software application supports a wide variety of transaction processing, business intelligence and analytics applications in business IT environments. Microsoft SQL Server is created on top of a standardized programming language and combines different data management and data analytics tools into SQL server. The latest version contains features for performance tuning, real-time operational analytics, data visualization and reporting on mobile devices. SQL Server is an application for storing information inside a "table" structure and supports different data type as such as Integer, Float, Decimal, Char and many more. (TechTarget, 2019), (Tutorials Point, 2019)

4.1.2 IBM Cognos

IBM Cognos Business Intelligence is a web-based integrated business intelligence suite that provides a wide range of functionalities. It helps organizations analyse and understand the performance by providing easy to use tool for reporting, analytics and monitoring of events and metrics. The software consists of several components designed to meet the different information requirements in a company. IBM Cognos has a variety of functionalities and components, for instance: IBM Cognos Framework Manager, IBM Cognos Cube Designer, IBM Cognos Transformer, Cognos Connection, Cognos Insight and Cognos Workspace. (IBM, 2019), (Tutorials Point, 2019)

4.1.3 Teradata

Teradata is a market leader in database services and products. It is one of the best and most powerful data warehouse tools that can be used for viewing and managing huge amounts of data. As a result, it is used extensively by many enterprise organizations for insights, analytics and decision making.

Some of the most important features and characteristics are:

- Low management costs with simple and effective solutions
- Designed for organization of any size
- Quick and insightful decision - based analytics
- Hybrid storage, which means the database can be deployed on premises or in the cloud. Support can be obtained either from Teradata directly or through Amazon Web Services
- It allows multiple concurrent users to ask complex questions related to data
- It is entirely built on a concept of parallel architecture which enables users to analyse data easily and more efficiently
- It offers sophisticated workload management and high performance, with diverse queries and more consistent response times
- If enables to segregate data into hot and cold data. The less frequently used cold data is stored on hard disk drives and the frequently used hot data is stored on a solid-state drive that offer higher performance

4.1.4 Oracle

Oracle is a data warehousing platform designed to provide customers with business insights and analytics that enables them to make fast and more informed business decisions. Oracle 12 c is a unique and comprehensive standard solution that offers high performance, scalability and optimization in data warehousing.

Some of the most important features are:

- Low cost of ownership
- On premises, in-cloud and hybrid solutions
- Big Data capability, it can manage and store large volumes of data
- Designed to serve a variety of industries, including government, finance, manufacturing, health, utilities, and high-tech
- Excellent, consistent performance and consolidation
- Delivers faster and more sophisticated analytic queries though Database In-Memory
- Increased and constant innovation that makes it very competitive

4.1.5 Our recommendation

We have chosen the Microsoft SQL Server to implement the data warehouse because the Super Cool Foods's database (Northwind) is considered to be a small size business and Microsoft is well known for their long-time service and consistent products. We had to consider another big factor that is the cost of licensing, implementation and maintenance. Finally, for SQL the learning curve is much smaller than other DBMS[3] solutions.

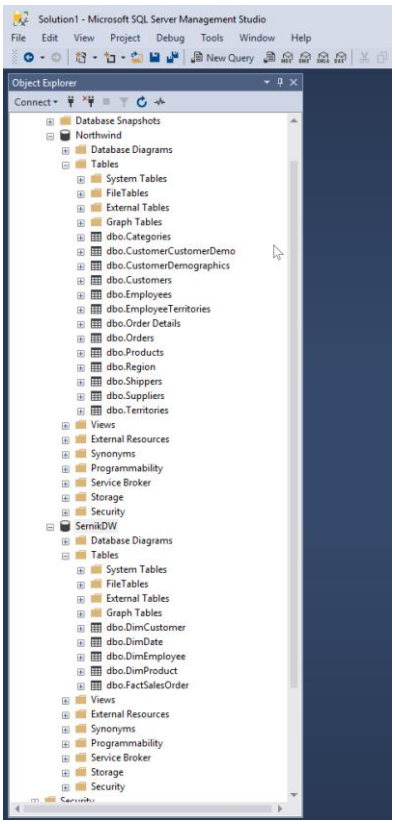

Figure 11 - Northwind and SernikDW databases in SQL Server

[3] A database management system (DBMS) is system software for creating and managing databases. The DBMS provides users and programmers with a systematic way to create, retrieve, update and manage data. The DBMS manages three important things: the data, the database engine and the database schema. Source: (Tech Target, 2019)

4.2 Staging Area

In a typical data warehousing architecture, there are several tasks that are required, those involved pulling data from multiple operational systems and cleansing of data to be imported to the data warehouse dimensions and fact tables.

Despite of in the Super Cool Foods using the Northwind example, there is only one data source to extract and load data, we will use the recommended practice and create a temporary data storage location where data is copied. This is a good solution to support the growth and changes that we will recommend for the company.

The Figure 12 below, shows the implementation of ETL process with the data source tables and temporary staging files. The requirements from Super Cool Foods are small enough that justifies the use of flat files instead of a totally separated database for staging. Once the data is transformed and clean, it is loaded to DW to be consumed by the reporting services to support analysis and actions by all stakeholders.

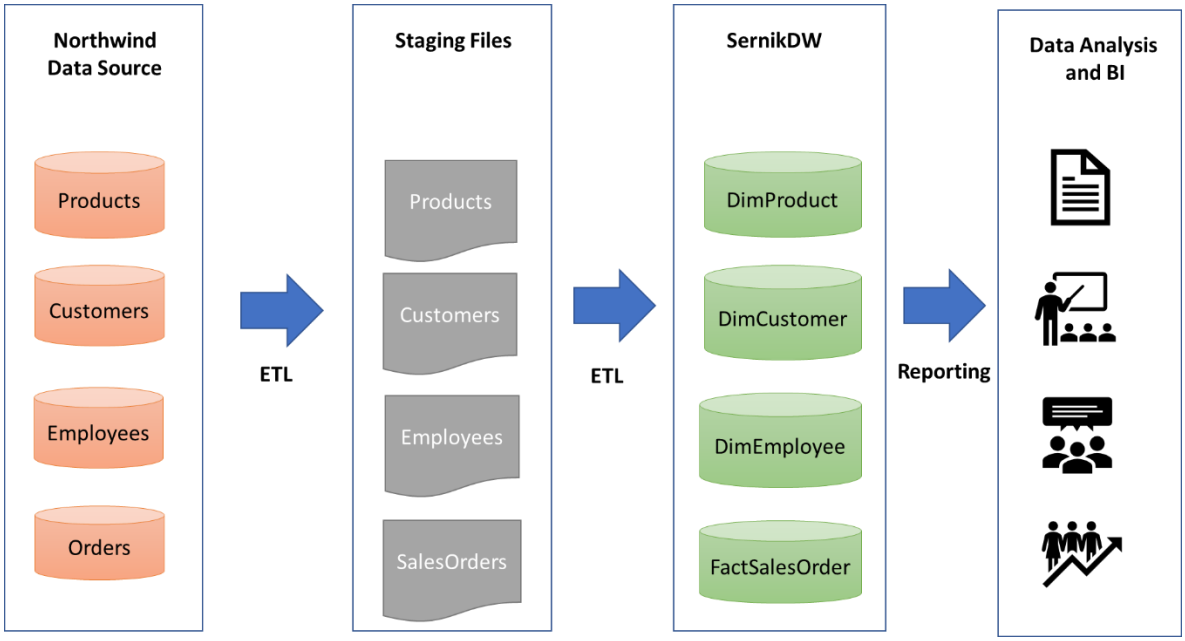

Figure 12 – ETL Implementation

4.2.1 Staging Package overview

To implement the ETL process above described, we have used the MS SQL SSIS[4] application mainly because is the easiest tool to use and manage packages at zero cost for a research purpose. When compared with other commercial solutions, the SQL SSIS provides a quick install and run for a small data warehouse as it is in our case. Also, we preferred to use graphical solution instead of SQL scripts for extraction and loading process. The Figure 13 shows the creation of staging packages into SSIS.

[4] Microsoft SQL Server Integration Services (SSIS) is a platform for building data integration solutions, including ETL packages for data warehousing. SSIS includes graphical tools and wizards for building and debugging packages; executing SQL statements and sending e-mail messages; data source extraction and destination for loading data. (Mcrosoft, 2019)

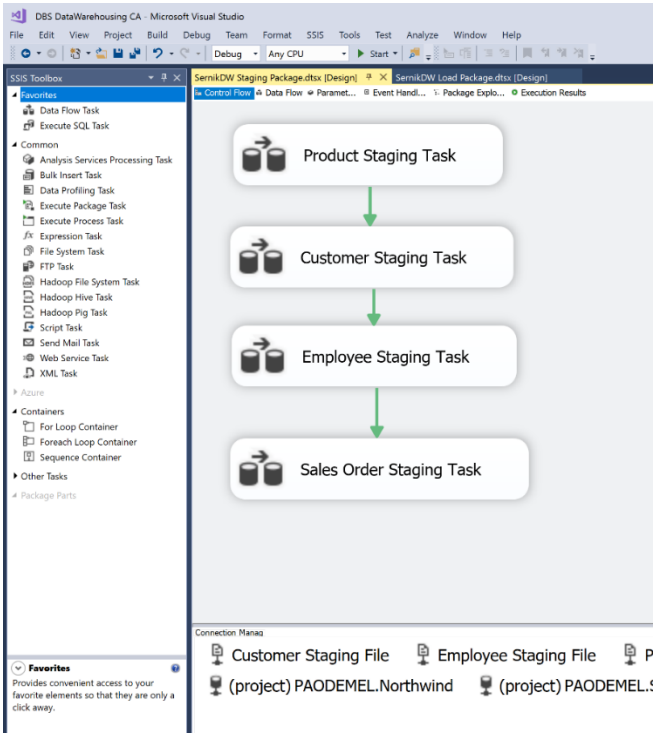

Figure 13 - Stage Package overview of SernikDW

4.2.2 Product staging

To keep very simple analysis of products during the transformation process we are using the category Id to lookup the category name and store that into the staging file.

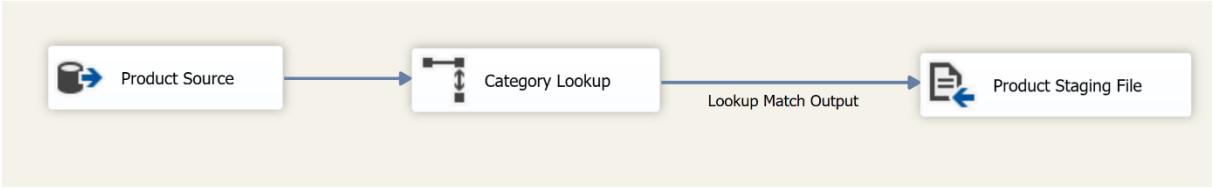

Figure 14 - Product staging task

4.2.3 Customer staging

Figure 15 - Customer staging task

4.2.4 Employee staging

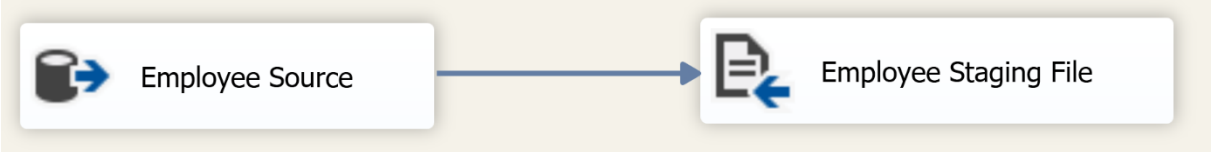

Figure 16 - Employee staging task

4.2.5 Sales Order staging

The SalesOrder Source task use a special View created to combine all information that we need regarding orders and the linkage with product, customer and employee. The View can be found on Appendix section: Table 4 - SQL script to create a sales order view

Figure 17 - SalesOrder staging task

4.3 Our Loading

The SSIS is a flexible set of tools that provides a variety of options for connecting to, and loading data into, SQL Data Warehouse. We have created a package that uses Data Flow task that contains source as the Flat file used for Staging and destination as the SernikDW tables.

The loading process design should be implemented considering one of the following strategies:

1. only load data that is already archived or off-line;
2. only load data that is contained in existing application; and
3. only load incremental changes of data from the operational system since the last time the data was loaded.

4.3.1 Loading Package overview

The Figure 18 shows the loading packages created into SSIS. We have added a SQL script task to clean up the tables before data is loaded. This is a solution proposed for this report only and not ideal for final implementation because data should be loaded in incremental basis, but we didn't have this as a requirement to be implemented.

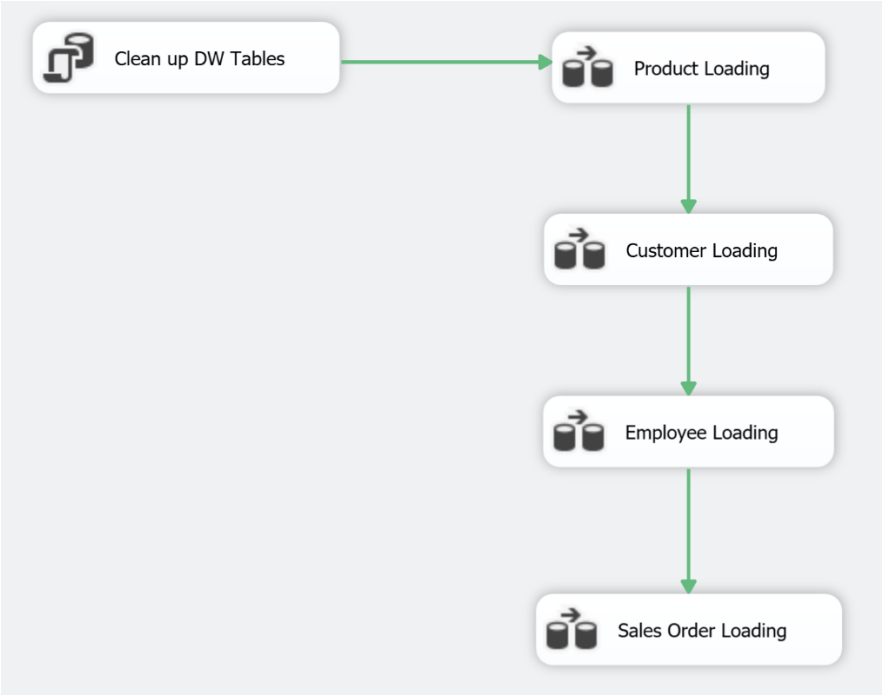

Figure 18 - Loading Package overview of SernikDW

4.3.2 Product loading

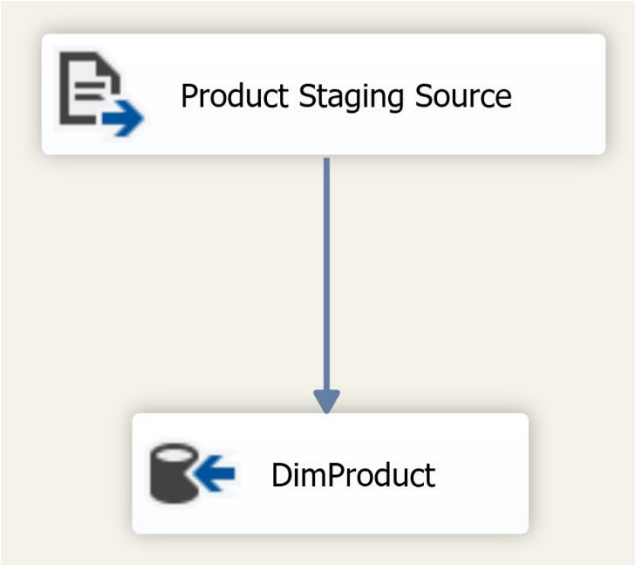

Figure 19 - DimProduct Loading task

4.3.3 Customer loading

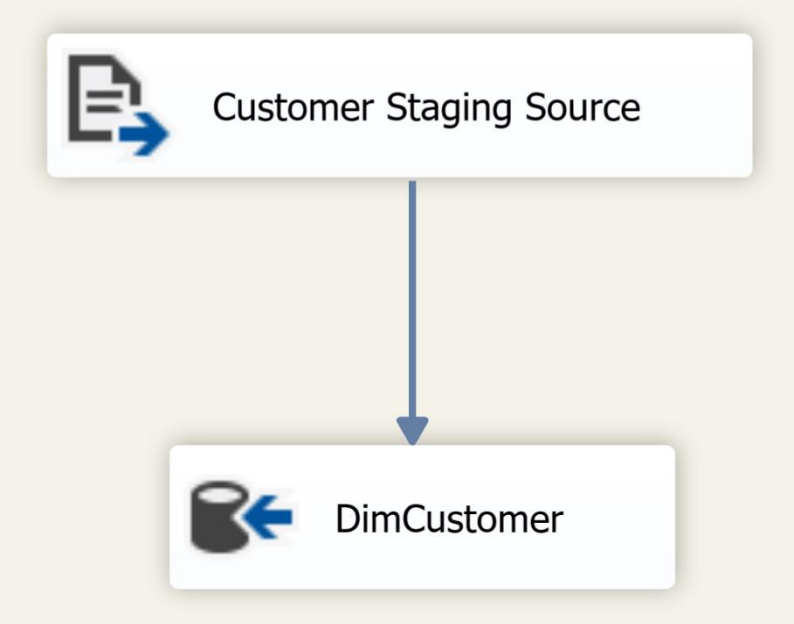

Figure 20 - DimCustomer Loading task

4.3.4 Employee loading

The Employee loading process requires that we use the lookup task to link the HireDate field with the DimDate dimensional table.

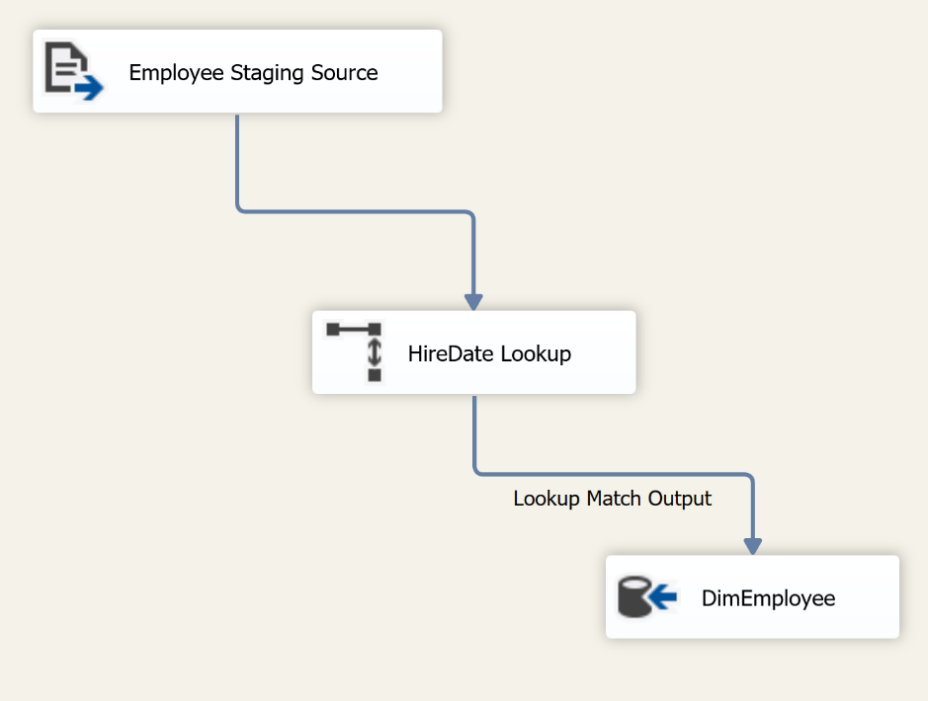

Figure 21 - DimEmployee Loading task

4.3.5 Sales Order loading

As the purpose of the fact table is to store measurement data only, we have to use the lookup task to load the correct primary keys for product, customer and employee. And same as with the employee loading task, the sales order dimension requires to lookup for a OrderDate and associate with the DimDate dimension.

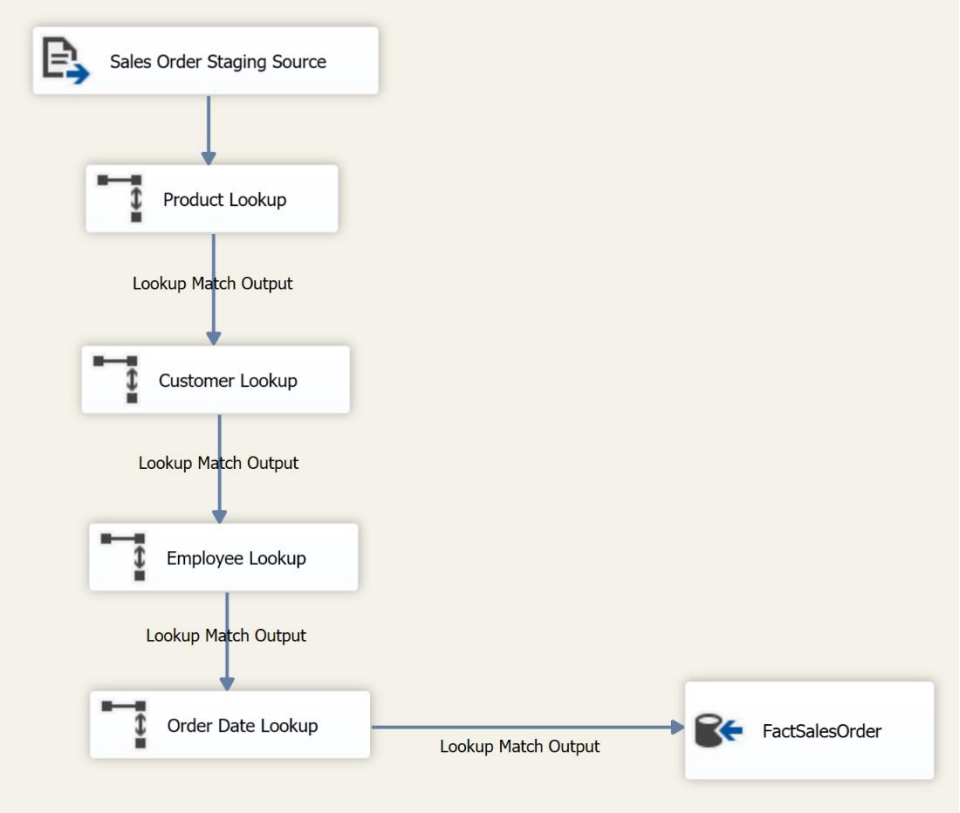

Figure 22 - FactSalesOrder Loading task

5 REPORTING AND ANALYSIS

5.1 Power BI vs Tableau

Power BI is a business analytics collection of software services, apps and connectors that work together to transform unrelated sources of data into clear and interactive insights. Power BI has been developed by Microsoft and is delivered via the Azure Cloud. This tool transforms the data into visuals, customized dashboards and interactive reports. It can also assist in the analytics and decision processes for group projects, divisions or entire corporations. It is easy to use, and it can develop complex analysis from different data sources. This tool is broadly used by public who shares visualizations and video tutorials, however, the Microsoft' s customer support is not of the highest standard.

Power BI is a tool which easily connects to any data source, for instance, excel spreadsheets, local databases or data warehouses, cloud-based or on-premises. We can distinguish 4 main parts.

- Power BI Desktop – Windows desktop application
- Power BI service – online Software as a Service (SaaS)
- Power BI mobile app – for Windows, iOS and Android devices
- Power BI Report Server – to publish reports on an on-premises report server after creating them on Power BI Desktop

Power BI is a relatively cheap option, the license price is 80% lower than other cloud BI products on the market. (Microsoft, 2019) and (Tutorials Point, 2019)

Tableau - is a powerful tool which can produce interactive data visualizations and it is used broadly in the business intelligence industry. It transforms raw data into easy to understand and analyse formats. Tableau is simple and user - friendly tool which does not require any technical or programming skills to create reports, dashboards or simple statistical analytics. This tool can connect to all kinds of data sources, from simple database such an excel or pdf, to a complex database like Oracle, Microsoft SQL database and various other data sources.

The Tableau Product range consists of:

- Tableau Desktop – allows to create customized reports, charts and dashboards;
- Tableau Public – for cost-effective usage; the created workbooks cannot be saved locally;
- Tableau Online – the data is stored and shared in the cloud;
- Tableau Server – to share visualizations and reports across organizations;
- Tableau Reader – it is a free tool to view the visualizations and workbooks created via Tableau Desktop or Public. (Tableau, 2019), (Guru99, 2019) and (Tutorials Point, 2019)

We have used the Power BI as our reporting tool mainly because our team had the application already in place running where the Tableau required install a trial version.

5.2 Strategic and Operations

We will use graphs and dashboards from Power BI to support our analysis of operation of Super Cool Foods.

5.2.1 General Revenue Analysis

5.2.1.1 General Sales per Year

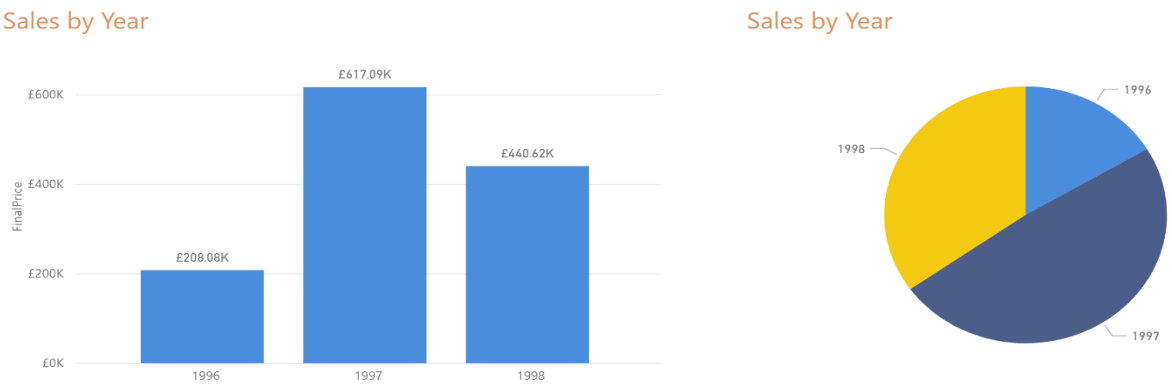

Figure 23 - General sales per Year

The 'Sales by Year' graph shows us that the best sales results were reached in 1997 and the lowest in 1996. However, we cannot fully compare those years because the available data ranges from Q3 of 1996 until Q2 of 1998 only. In the last two quarters of 1996 the sales figure is £208.08k, the year 1997 was closed off with a result of £6170.09k and in the first two quarters of 1998 the company sold products worth of £440.62k.

5.2.1.2 General Sales per Quarter

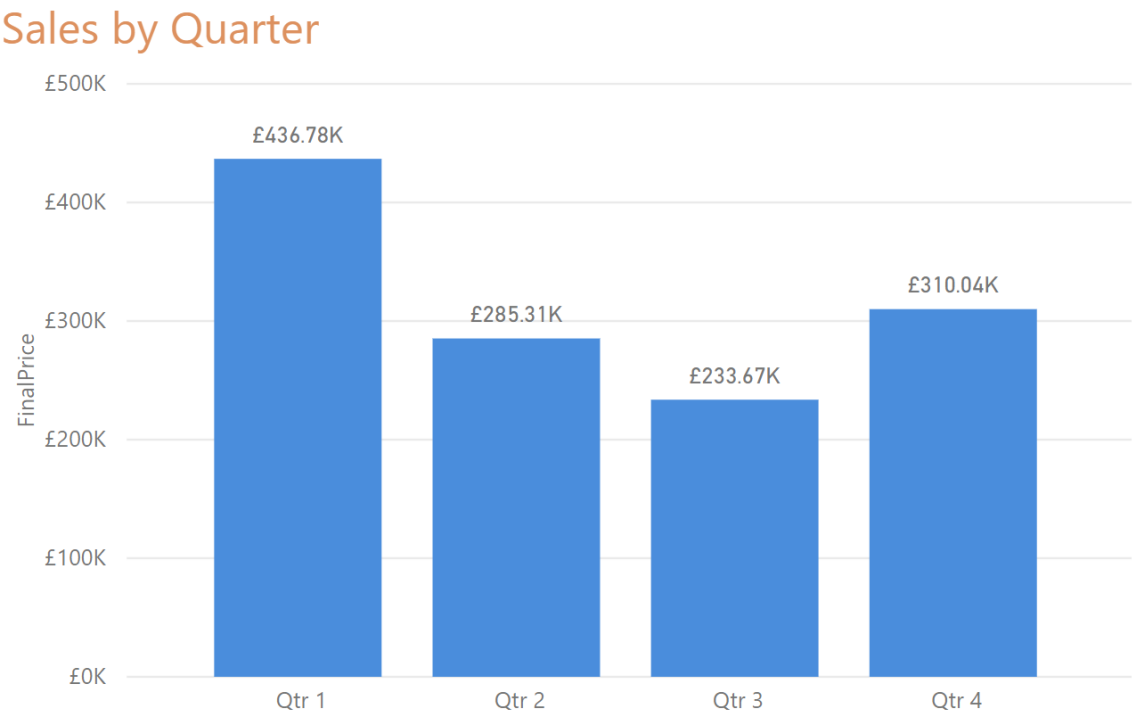

Figure 24 - General sales per Quarter

The 'Sale by Quarter' graph represents the summarized sales of each quarter. From the graph we can read that the best overall results were achieved in Q1 (1997 & 1998) and the lowest in Q3 (1996 & 1997). In Q1, 1997 and 1998 the company reached £436.78k which is the highest and almost doubles the result from Q3, 1996 and 1997, when Super Cool Foods sold goods for £233.67k.

5.2.1.3 General Sales per Month

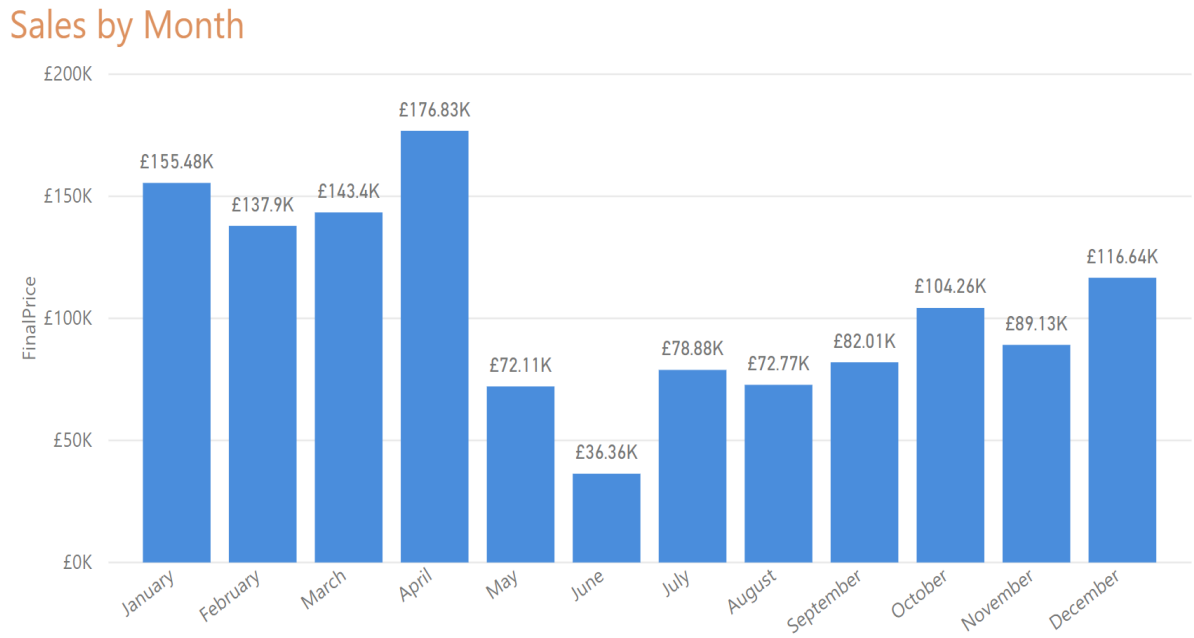

Figure 25 - General Sales per Month

The 'Sales by Month' graph represents the summarized sales figures for each month for the period of Q3, 1996 till Q2, 1998. The best sales result was reached in April 1997 and 1998. The lowest sale was recorded in June with £36.36k which was only 20% of the April sales when the company reached £176.83k. Overall, Super Cool Foods noted the best sales performance in first four months of 1997 and 1998.

5.2.2 Costumer Analysis

5.2.2.1 Top 20 Customers by Sales

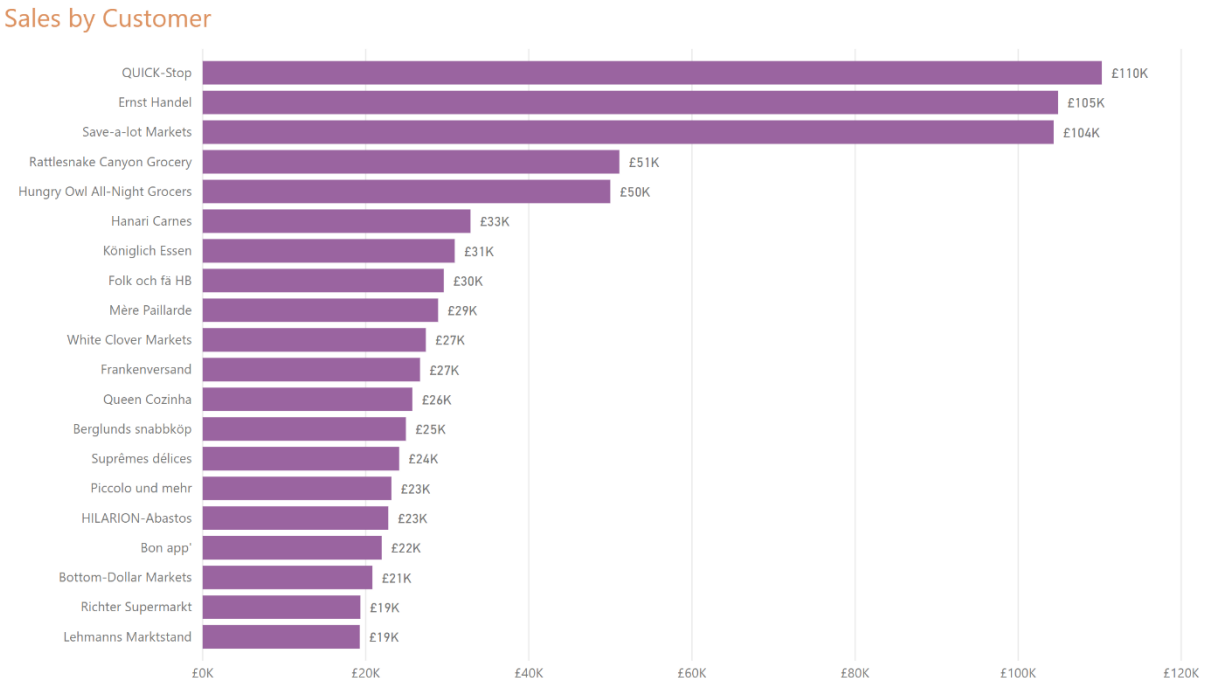

Figure 26 - Top 20 Customers by Sales

From the above graph 'Sales by Customer' we can see that in the list of the best twenty customers, the first five bought products for over £50k for the period of Q3, 1996 until Q2, 1998. Also, the top three bought products for over £100k. The rest of the customers spent between £35k and £19k in that period. The best client of Super Cool Food is Quick Stop who spent £110k, and at the bottom of the list is Lehmanns Martstand and Richter Supermarkt who spent £19k each.

5.2.2.2 Top 20 Sales per Customer per Year

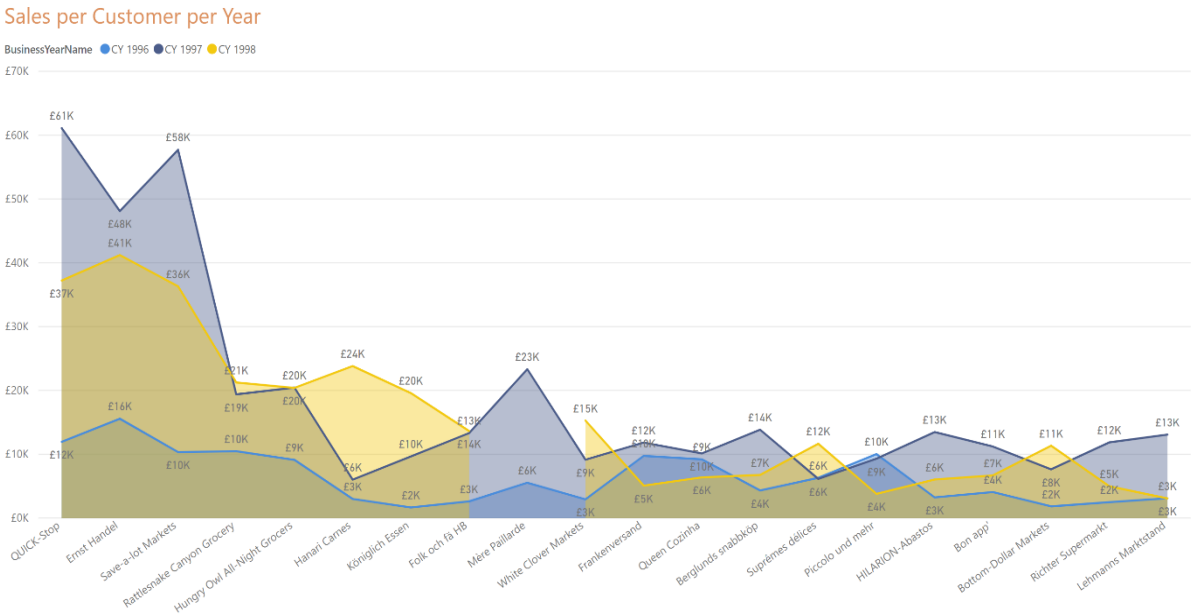

Figure 27 - Top 20 Customers per Year

The graph 'Sales per Customer per Year' shows that the highest purchase was made in 1996 by Ernst Handel who bought goods for £16k. In 1997 the top customer was Quick Stop who bought goods for £61k and the second best was Save-a-Lot Markets, who spent £58k in the same year. In 1998 the best customer was Ernst Handel again as he bought goods worth of £41k. In 1996 we had two customers who

bought goods for only £2k; in 1997 we can observe the lowest spending of £2k and in 1998 we can note the lowest spending of £4k. We can also see that one of our clients - Mere Paillarde did not purchase any goods in the year of 1998.

5.2.3 Product Analysis

5.2.3.1 Top 20 Products sales per Year

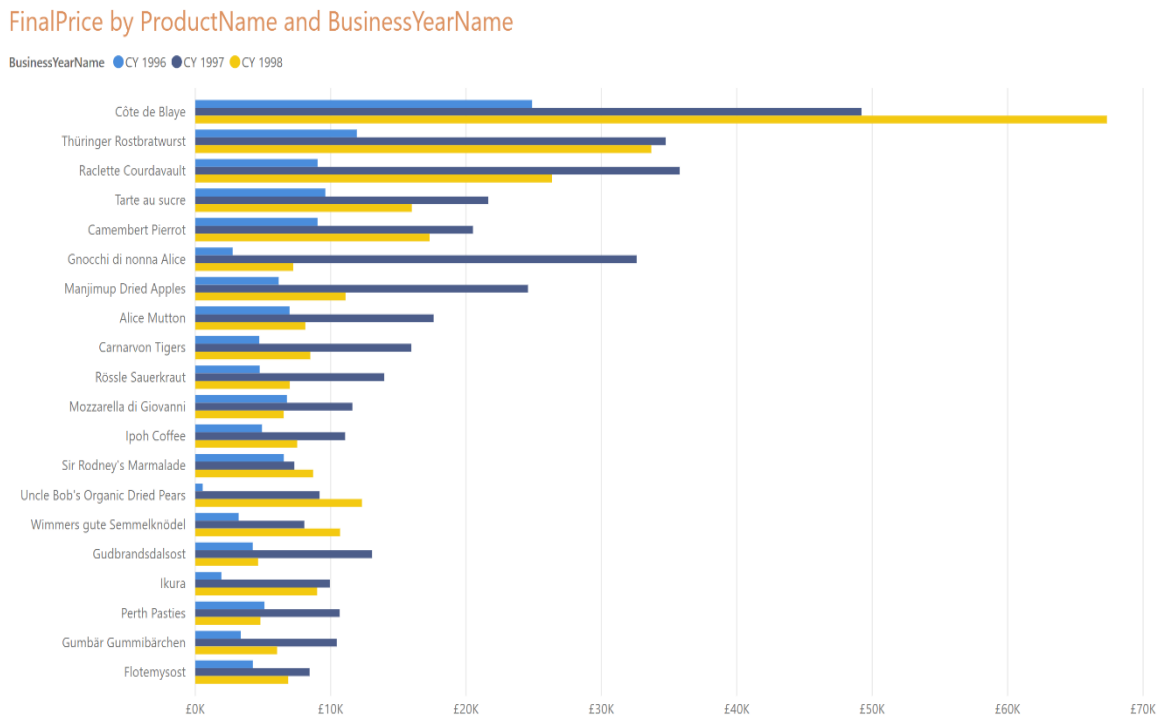

Figure 28 - Top 20 Products Sales per Year

The above graph shows us figures in relation to the top twenty products sales per year. The best-selling product between Q3, 1996 and Q2, 1998 was Cote de Blaye with an increase of sales every single year. The highest number was achieved in the first two quarters of 1998. Cote de Blaye was sold in 1997 for a value of almost £50k and in 1998 for almost £70k. Over 50% of the products from the top twenty list have not reached £20k and the products from the bottom of the list have not reached £10k in any of the recorded years.

5.2.3.2 Total Sales per Product Category

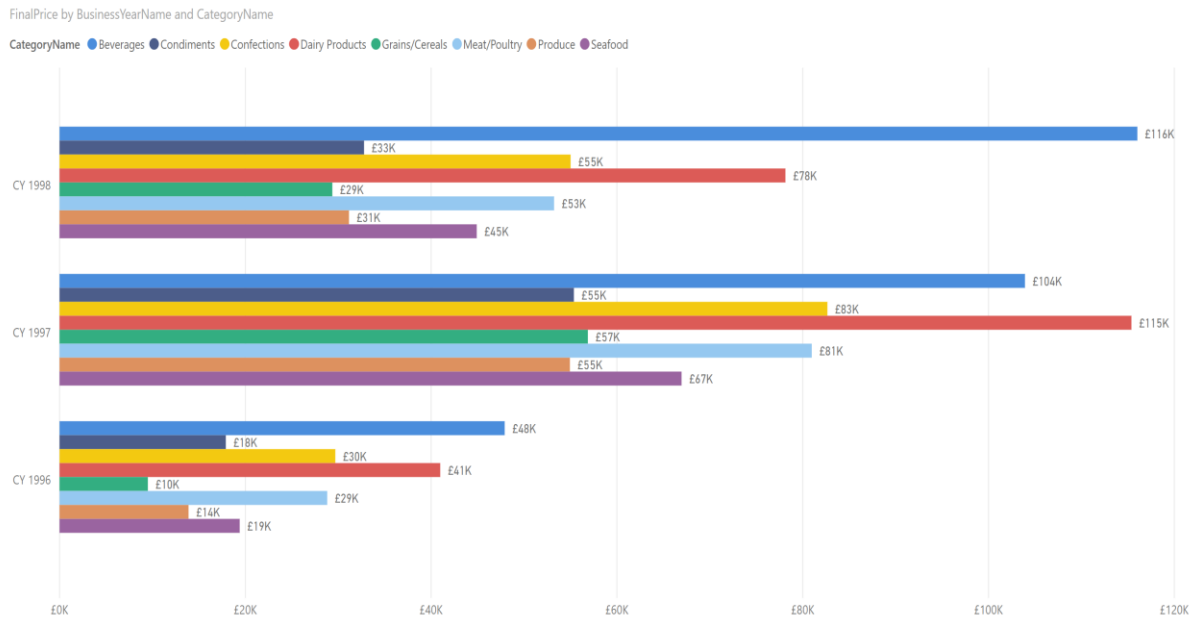

Figure 29 - Total Sales per Product Category

The above graph shows us sales figures per product categories per year. The best-selling product category in 1996 and 1998 was Beverages, and 1997 it was Dairy Products with £115K.

5.2.4 Employee Analysis

5.2.4.1 Overall Sales per Employee

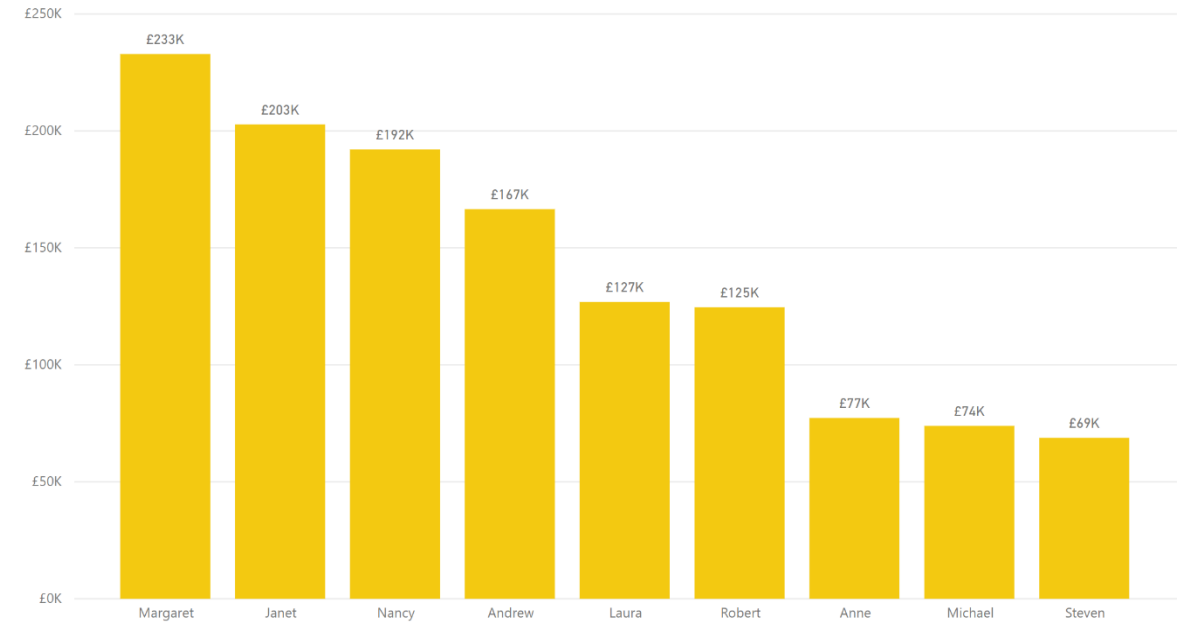

Figure 30 - Overall Sales per Employee

As we can see on the above graph Margaret recorded the best overall sales number as she sold goods for £233k. The second best was Janet with the total sales of £203k. The lowest sales results were

obtained by Steven. His total amount of sale is £69k and this figure is almost 3.5 times worse than the best total of the best sales person – Margaret. Two of the employees sold for over £200k, in the middle we have three members of the team with sales between £125k and £167k and three of the employees recorded a value less than £100k.

5.2.4.2 Overall Sales per Employee per Year

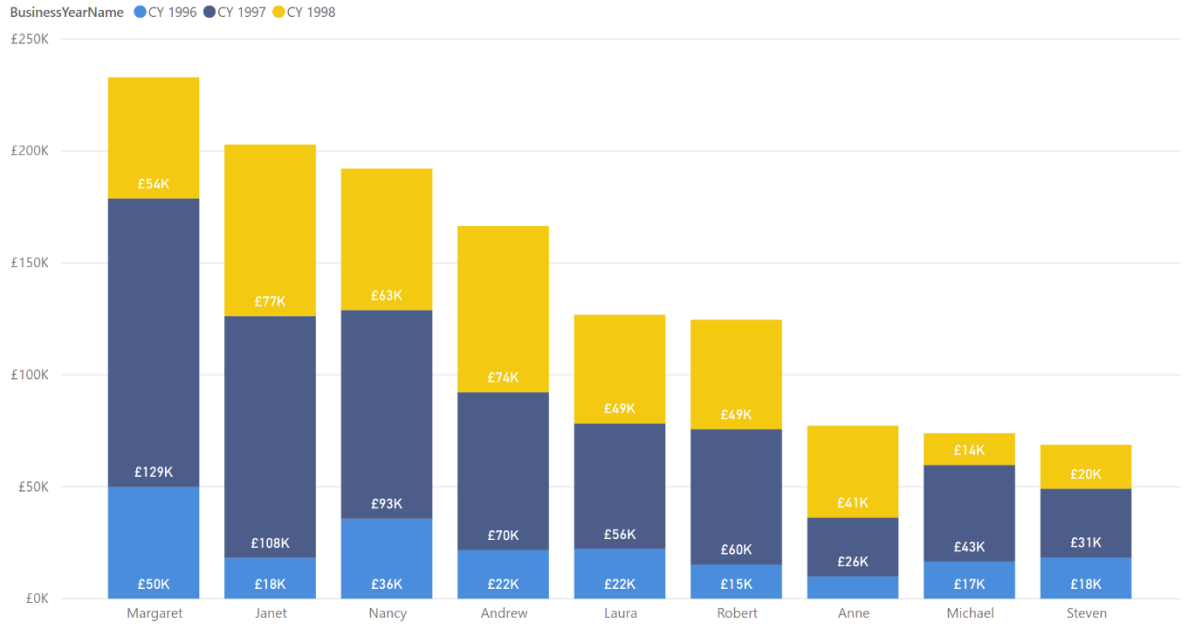

Figure 31 - Sales per Employee per Year

The 'Sales Per Person Per Year' graph shows us who performed the best each year and thus, helps us to find answers in relation to employees' performance. As we can see in the above graph in both 1996 and 1997 Margaret was the best, however, in 1998 the best sales person was Janet. The lowest sales results were obtained by Steven. He only reached a total amount of £69k and this figure is almost 3.5 times worse than the best sales person – Margaret. Margaret got the best overall sales figures for three years. Even if, in 1998 she was the 4th and her year total was 2.5 times worse than year before she still managed to keep her top position of the company's best sales person. Only two employees maintained a steady increase of their sales figures in those three years. Andrew and Anne improved their performance each year.

6 FUTURE WORK

This report provides a quick implementation of data warehouse for Super Cool Foods, we would like to recommend improvements to be made by the company on both business and technical aspects.

6.1.1 Business recommendations

i. They should find ways of controlling the shipping costs more effectively, for instance, introducing a minimum value or quantity orders that would reduce the number of small purchases that aren't generating any real revenue;
ii. Reduce the cost of shipments from individual suppliers by decreasing the number vendors and introducing bulk orders;
iii. They should have a system of measuring the delivery times and record any delays or issues encountered, and look for ways of improvements;
iv. They can introduce surveys measuring the customer's satisfaction which should be recorded and analysed
v. They should work towards increasing the number of their customers by providing appropriate customer care training and improving the employees' sales skills
vi. They can introduce a yearly bonus based on employee's performance and/or commission pay in order to increase their motivation
vii. They should analyse the performance of employees who don't reach their targets and consider more training for them or look for other solutions;

6.1.2 Technical recommendations

viii. The dimensional schema contains some set of data to support the business goals as defined. The company should consider increase the schema with details to allow a deep analysis of aspects such as suppliers, shipment and customer relations.
ix. The company could consider changes on database to support future business expansion for example the online shopping.
x. For performance reasons, we have decided to clean up the DW before each loading process. In the 'real world' we recommend to incrementally load the data into the data warehouse with both start and end points. Also, the validation checks for blank data and duplicate records should be considering before loading the data.
xi. We have used MS Power BI as a reporting solution, but the company could avail other options such as Tableau or even Excel depending of the cost and knowledge of the analysis team.
xii. They should update the ETL process with implementation of a schedule to run it outside the business hours and add audit tasks to send notifications of success or fail.

7 CONCLUSION

This report intent to provide the reader the practical steps to build a data warehouse from the business consideration to technical implementation. We have found data warehousing a very complex topic and through the 'consulting approach' were could use our theoretical knowledge and general understanding of the module content in practice and follow all the steps to create an actual data warehouse. We've created and implemented the ETL, built a star schema for reporting purposes.

Together, we discussed the various aspects of the company and defined the business drivers. We wrote the business plan for the company, conducted SWOT analysis, analysed KPI's, identified and analysed the most important stakeholders of our company. Based on established business drivers and our KPI's we could decide what we want to focus on in our analysis. Using the balanced scorecard, we were able to link the company's vison, mission and strategy with the four perspectives. This enabled us to get a better insight into the company, which in turn helped us to draw conclusions and identify specific recommendations.

Unfortunately, we did notice that the data extracted from the operations system has not been satisfactory to implement a full data warehousing solution. We have used the four characteristics from Bill Inmon's data warehouse definition:

- Subject oriented – the data is organised around the functional applications of the company, we have used this concept to keep our dimensions DimProduct, DimCustomer and DimEmployee.
- Integrated – we can combine data from various systems into a repository that can be used to get a general view of the business; we have used only one data source, the Northwind database, however, we know about this functionality and have learnt the different tools we could use to integrate the different databases and other structured on non-structured files.
- Non-volatile - we learnt that if we try to change data in the source this will generate a new record in a data mart but will not update the existing one
- Time variant - helps to create a unique source of truth where every unit of data is accurate, and many case times stamped.

Different than what may be thinking, the Data Warehouse does not contain all the data in the company. Its structured to support reporting for decision makers to access from a business intelligence tools for strategic and tactical decisions.

We hope the readers will find a good understanding of implementing a data warehouse and those implications with the business. As a summary we could conclude the following:

a) Before creating a DW you must focus on Business Drives and Company Goals;
b) Data modelling is an important process to define the schema to be used, as this result in how data will be stored and complex access later;
c) The implementation of a DW by a company brings a competitive way to understand business and operational gaps and trends;
d) The biggest drawback that we could see is that even a smaller change made in the database structure requires modification on ETL and dimensional model.

8 BIBLIOGRAPHY

i) Adamson, C., 2006. *Mastering Data Warehouse Aggregates.* 1st ed. Indianopolis: John Wiley & Sons.

ii) Alooma, 2019. *Data Warehousing Tools.* [Online]
Available at: https://www.alooma.com/blog/data-warehousing-tools
[Accessed 6 April 2019].

iii) Balance Scorecard Institute, 2019. *Balanced Scorecard Basics.* [Online]
Available at: https://www.balancedscorecard.org/BSC-Basics/About-the-Balanced-Scorecard
[Accessed 16 March 2019].

iv) Business News Daily, 2019. *SWOT Analysis: What It Is and When to Use It.* [Online]
Available at: https://www.businessnewsdaily.com/4245-swot-analysis.html
[Accessed 27 March 2019].

v) Claire, C., 2004. *Understanding organisational context.* 2nd ed. s.l.:Pearson Education Limited.

vi) Guru99, 2019. *20 Best ETL / Data Warehousing Tools in 2019.* [Online]
Available at: https://www.guru99.com/top-20-etl-database-warehousing-tools.html
[Accessed 6 April 2019].

vii) Guru99, 2019. *What is Tableau? Uses and Applications.* [Online]
Available at: https://www.guru99.com/what-is-tableau.html
[Accessed 06 April 2019].

viii) Heflo, 2019. *Business SWOT analysis: What it is and how to use it in your company..* [Online]
Available at: https://www.heflo.com/blog/swot/business-swot-analysis/
[Accessed 27 March 2019].

ix) IBM, 2019. *IBM Cognos Analytics.* [Online]
Available at: https://www.ibm.com/products/cognos-analytics
[Accessed 03 April 2019].

x) Inmon, W. H., 2005. *Building the Data Warehouse.* 4th ed. Indianopolis: Wiley Publishing.

xi) Kaplan, R. S. & Norton, D. P., 1996. *The balance scorecard: translating strategy into action.* Boston, MA: Harvard Business School Press.

xii) Kaplan, R. S. & Norton, D. P., 2004. *Strategy Maps: Converting Intangible Assets Into Tangible Outcomes.* s.l.:Harvard Business School Press.

xiii) Kimball, R., 1996. Dangerous Preconceptions: Discovering the liberating truths that can lead to a successful data warehouse project. *DBMS Magazine*, August.

xiv) Kimball, R., 2013. *The Data Warehouse Toolkit.* 3rd ed. s.l.:Wiley.

xv) Kimball, R. & Caserta, J., 2004. *The data warehouse ETL toolkit.* s.l.:Wiley Publishing Inc. .

xvi) KPI.org, 2019. *WHAT IS A KEY PERFORMANCE INDICATOR (KPI)?.* [Online]
Available at: https://kpi.org/KPI-Basics
[Accessed 20 March 2019].

xvii) Lumen Learning, 2019. *Business Stakeholders.* [Online]
Available at: https://courses.lumenlearning.com/boundless-management/chapter/business-stakeholders/
[Accessed 27 March 2019].

xviii) Mcrosoft, 2019. *SSIS How to Create an ETL Package.* [Online]
Available at: https://docs.microsoft.com/en-us/sql/integration-services/ssis-how-to-create-an-etl-package?view=sql-server-2017
[Accessed 14 April 2019].

xix) Microsoft, 2019. *Northwind and pubs Sample Databases for SQL Server 2000.* [Online]
Available at: https://www.microsoft.com/en-us/download/details.aspx?id=23654
[Accessed 04 04 2019].

xx) Microsoft, 2019. *Power BI.* [Online]
Available at: https://powerbi.microsoft.com/en-us/
[Accessed 03 March 2019].

xxi) Monitis, 2019. *Top 5 data warehouses on the market today.* [Online]
Available at: https://www.monitis.com/blog/top-5-data-warehouses-on-the-market-today/
[Accessed 6 April 2019].

xxii) Nair, M., 2004. *Essentials of Balanced Scorecard.* s.l.: John Wiley & Sons, Inc..

xxiii) Oracle Corporation, 2019. *Oracle Autonomous Data Warehouse.* [Online]
Available at: https://www.oracle.com/ie/database/data-warehouse.html
[Accessed 6 April 2019].

xxiv) Ponniah, P., 2010. *Data warehousing fundamentals for IT professionals.* 2nd ed. New Jersey: John Wiley & Sons.

xxv) Sharda, R., Delen, D. & Turban, E., 2018. *Business Intelligence, Analytics, and Data Science.* 4th ed. s.l.:Pearson Education Limited.

xxvi) Tableau, 2019. *Tableau.* [Online]
Available at: https://www.tableau.com/
[Accessed 06 April 2019].

xxvii) Tech Target, 2019. *database management system (DBMS).* [Online]
Available at: https://searchsqlserver.techtarget.com/definition/database-management-system
[Accessed 10 4 2019].

xxviii) TechTarget, 2019. *Microsoft SQL Server.* [Online]
Available at: https://searchsqlserver.techtarget.com/definition/SQL-Server
[Accessed 03 March 2019].

xxix) Teradata, 2019. *Teradata Vantage.* [Online]
Available at: https://www.teradata.com/Products/Software/Vantage
[Accessed 6 April 2019].

xxx) Tutorials Point, 2019. *Cognos Tutorial.* [Online]
Available at: www.tutorialspoint.com/cognos/
[Accessed 03 April 2019].

xxxi) Tutorials Point, 2019. *MS SQL Server - Overview.* [Online]
Available at: www.tutorialspoint.com/ms_sql_server/ms_sql_server_overview
[Accessed 20 March 2019].

xxxii) Tutorials Point, 2019. *Power BI Tutorial.* [Online]
Available at: www.tutorialspoint.com/power_bi/
[Accessed 03 March 2019].

xxxiii) Tutorials Point, 2019. *Tableau Tutorial.* [Online]
Available at: http://www.tutorialspoint.com/tableau
[Accessed 06 April 2019].

9 APPENDIX

9.1 Northwind database – Table structure

The following figures show the tables defined for the Northwind database. We have assumed this is the relational database used by our mock Super Cool Foods.

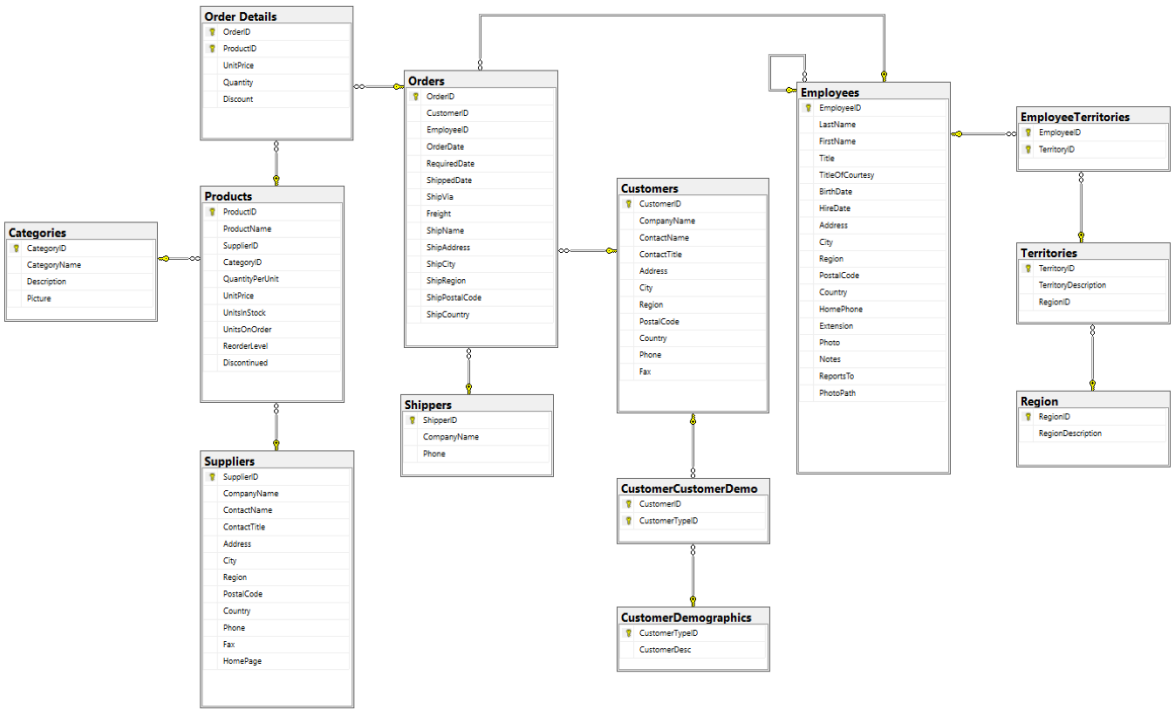

Figure 32 - Northwind database schema

9.2 SernikDW – Create SQL Script

The following script is included together with this report - **SernikDW create tables.sql**.

```
/*
# Continuous Assessment
# Dublin Business School
# B8IT104 Data Warehousing and Business Intelligence
# Lecturer: Noel Cosgrave (noel.cosgrave@dbs.ie)
# ################################################

 Daniel Pereira (10391381@mydbs.ie)
 Maja Pomohaczi (10392409@mydbs.ie)
 Sylwia Wojciechowska (10392584@mydbs.ie)

# ################################################

  This script will create tables for our SernikDW database.

  The SernikDW is the datawarehouse (datamart) solution
*/

USE [SernikDW]
GO
```

```
/* Drop existing tables */
DROP TABLE IF EXISTS [dbo].[FactSalesOrder]
GO

DROP TABLE IF EXISTS [dbo].[DimProduct]
GO

DROP TABLE IF EXISTS [dbo].[DimEmployee]
GO

DROP TABLE IF EXISTS [dbo].[DimCustomer]
GO

DROP TABLE IF EXISTS [dbo].[DimDate]
GO



/* Create tables */
CREATE TABLE DimDate(
        DateKey int PRIMARY KEY,
        SystemDateAltKey datetime NOT NULL,
        [FullDateUK] CHAR(10), -- Date in dd-MM-yyyy format
        [FullDateUSA] CHAR(10),-- Date in MM-dd-yyyy format
        CalendarYear int NOT NULL,
        [BusinessYearName] CHAR(7), --CY 2012,CY 2013
        CalendarQuarter int NOT NULL,
        [CalendarQuarterName] VARCHAR(9),--First,Second..
        CalendarMonth int NOT NULL,
        CalendarMonthName nvarchar(15) NOT NULL,
        [BusinessMonthYear] CHAR(10), --Jan-2013,Feb-2013
        [CalendarMMYYYY] INT,
        [MonthOfQuarter] INT,-- Month Number belongs to Quarter
        [CalendarWeekOfYear] INT,--Week Number of the Year
        [CalendarWeekOfQuarter] INT, --Week Number of the Quarter
        [CalendarWeekOfMonth] INT,-- Week Number of Month
        CalendarDayOfMonth int NOT NULL,
        [CalendarDaySuffix] VARCHAR(4), -- Apply suffix as 1st, 2nd ,3rd etc
        CalendarDayName nvarchar(15) NOT NULL,
        [CalendarDayOfYear] INT,
        [CalendarDayOfWeekUSA] INT,-- First Day Sunday=1 and Saturday=7
        [CalendarDayOfWeekUK] INT,-- First Day Monday=1 and Sunday=7
        [DayOfWeekInMonth] INT, --1st Monday or 2nd Monday in Month
        [DayOfWeekInYear] INT,
        [DayOfQuarter] INT,
        [BusinessIsWeekday] BIT-- 0=Week End ,1=Week Day

)
GO

CREATE TABLE DimProduct(
        ProductKey int identity NOT NULL PRIMARY KEY NONCLUSTERED,
        ProductSourceID int NOT NULL,
        ProductName nvarchar(40) NOT NULL,
        UnitPrice money NULL,
        CategoryID int NULL,
        CategoryName nvarchar(15) NULL
)
GO

CREATE TABLE DimCustomer(
        CustomerKey int identity NOT NULL PRIMARY KEY NONCLUSTERED,
        CustomerSourceID nvarchar(5) NOT NULL,
        CompanyName nvarchar(40) NULL,
        ContactName nvarchar(30) NULL,
        ContactTitle nvarchar(30) NULL
)
GO

CREATE TABLE DimEmployee(
        EmployeeKey int identity NOT NULL PRIMARY KEY NONCLUSTERED,
        EmployeeSourceID int NOT NULL,
        LastName nvarchar(20) NOT NULL,
```

```
        FirstName nvarchar(10) NOT NULL,
        Title nvarchar(30) NOT NULL,
        HireDateKey int NOT NULL
)
GO
ALTER TABLE [dbo].[DimEmployee]  WITH CHECK ADD FOREIGN KEY([HireDateKey])
REFERENCES [dbo].[DimDate] ([DateKey])
GO


CREATE TABLE FactSalesOrder(
        ProductKey int NOT NULL,
        CustomerKey int NOT NULL,
        EmployeeKey int NOT NULL,
        OrderDateKey int NOT NULL,
        OrderSourceID int NOT NULL,
        UnitPrice money NOT NULL,
        Quantity smallint NOT NULL,
        Discount real NOT NULL,
        TotalPrice money NOT NULL,
        TotalDiscount real NOT NULL,
        FinalPrice money NOT NULL

        CONSTRAINT [PK_FactSalesOrder] PRIMARY KEY NONCLUSTERED
        (
                [ProductKey],
                [CustomerKey],
                [EmployeeKey],
                [OrderDateKey],
                [OrderSourceID]
        )
)
GO
ALTER TABLE [dbo].[FactSalesOrder]  WITH CHECK ADD FOREIGN KEY([CustomerKey])
REFERENCES [dbo].[DimCustomer] ([CustomerKey])
GO
ALTER TABLE [dbo].[FactSalesOrder]  WITH CHECK ADD FOREIGN KEY([EmployeeKey])
REFERENCES [dbo].[DimEmployee] ([EmployeeKey])
GO
ALTER TABLE [dbo].[FactSalesOrder]  WITH CHECK ADD FOREIGN KEY([OrderDateKey])
REFERENCES [dbo].[DimDate] ([DateKey])
GO
ALTER TABLE [dbo].[FactSalesOrder]  WITH CHECK ADD FOREIGN KEY([ProductKey])
REFERENCES [dbo].[DimProduct] ([ProductKey])
GO
```

Table 2 - SQL script to create SernikDW

9.3 SernikDW – Insert SQL Script

The following is the script to generate dates into the DimDate dimensional table. This is important because as a principle, a DW is 'time-variant'. The file is included as part of this report. See **DimDate insert.sql**.

```
/*
# Continuous Assessment
# Dublin Business School
# B8IT104 Data Warehousing and Business Intelligence
# Lecturer: Noel Cosgrave (noel.cosgrave@dbs.ie)
# ################################################

 Daniel Pereira (10391381@mydbs.ie)
 Maja Pomohaczi (10392409@mydbs.ie)
 Sylwia Wojciechowska (10392584@mydbs.ie)

# ################################################
```

```
  This script will insert values into DimDate dimensional table.

# ################################################
*/

USE [SernikDW]


--Define the date range to be created
--StartDate must be less than EndDate
--Dates must be in USA format
DECLARE @StartDate DATETIME = '01/01/1990'
DECLARE @EndDate DATETIME = '12/31/2030'

--Temporary Variables To Hold the Values During Processing of Each Date of Year
DECLARE
        @DayOfWeekInMonth INT,
        @DayOfWeekInYear INT,
        @DayOfQuarter INT,
        @WeekOfMonth INT,
        @CurrentYear INT,
        @CurrentMonth INT,
        @CurrentQuarter INT

/*Table Data type to store the day of week count for the month and year*/
DECLARE @DayOfWeek TABLE (DOW INT, MonthCount INT, QuarterCount INT, YearCount INT)

INSERT INTO @DayOfWeek VALUES (1, 0, 0, 0)
INSERT INTO @DayOfWeek VALUES (2, 0, 0, 0)
INSERT INTO @DayOfWeek VALUES (3, 0, 0, 0)
INSERT INTO @DayOfWeek VALUES (4, 0, 0, 0)
INSERT INTO @DayOfWeek VALUES (5, 0, 0, 0)
INSERT INTO @DayOfWeek VALUES (6, 0, 0, 0)
INSERT INTO @DayOfWeek VALUES (7, 0, 0, 0)

--Extract and assign various parts of Values from Current Date to Variable
DECLARE @CurrentDate AS DATETIME = @StartDate
SET @CurrentMonth = DATEPART(MM, @CurrentDate)
SET @CurrentYear = DATEPART(YY, @CurrentDate)
SET @CurrentQuarter = DATEPART(QQ, @CurrentDate)



/********************************************************************************************/
--Proceed only if Start Date(Current date ) is less than End date you specified above
WHILE @CurrentDate <= @EndDate
BEGIN

/*Begin day of week logic*/

    /*Check for Change in Month of the Current date if Month changed then change variable value*/
        IF @CurrentMonth != DATEPART(MM, @CurrentDate)
        BEGIN
                UPDATE @DayOfWeek
                SET MonthCount = 0
                SET @CurrentMonth = DATEPART(MM, @CurrentDate)
        END

    /* Check for Change in Quarter of the Current date if Quarter changed then change variable
value*/
        IF @CurrentQuarter != DATEPART(QQ, @CurrentDate)
        BEGIN
                UPDATE @DayOfWeek
                SET QuarterCount = 0
                SET @CurrentQuarter = DATEPART(QQ, @CurrentDate)
        END

     /* Check for Change in Year of the Current date if Year changed then change variable value*/
        IF @CurrentYear != DATEPART(YY, @CurrentDate)
        BEGIN
                UPDATE @DayOfWeek
```

```
                SET YearCount = 0
                SET @CurrentYear = DATEPART(YY, @CurrentDate)
        END

    -- Set values in table data type created above from variables
        UPDATE @DayOfWeek
        SET
                MonthCount = MonthCount + 1,
                QuarterCount = QuarterCount + 1,
                YearCount = YearCount + 1
        WHERE DOW = DATEPART(DW, @CurrentDate)

        SELECT
                @DayOfWeekInMonth = MonthCount,
                @DayOfQuarter = QuarterCount,
                @DayOfWeekInYear = YearCount
        FROM @DayOfWeek
        WHERE DOW = DATEPART(DW, @CurrentDate)

/*End day of week logic*/


/* Populate Your Dimension Table with values*/

        INSERT INTO [dbo].[DimDate]
        SELECT
                CONVERT (char(8),@CurrentDate,112) as DateKey,
                @CurrentDate AS SystemDateAltKey,
                CONVERT (char(10),@CurrentDate,103) as FullDateUK,
                CONVERT (char(10),@CurrentDate,101) as FullDateUSA,
                DATEPART(YEAR, @CurrentDate) AS CalendarYear,
                'CY ' + CONVERT(VARCHAR, DATEPART(YEAR, @CurrentDate)) AS BusinessYearName,
                DATEPART(QQ, @CurrentDate) AS CalendarQuarter,
                CASE DATEPART(QQ, @CurrentDate)
                        WHEN 1 THEN 'First'
                        WHEN 2 THEN 'Second'
                        WHEN 3 THEN 'Third'
                        WHEN 4 THEN 'Fourth'
                        END AS CalendarQuarterName,
                DATEPART(MM, @CurrentDate) AS CalendarMonth,
                DATENAME(MM, @CurrentDate) AS CalendarMonthName,
                LEFT(DATENAME(MM, @CurrentDate), 3) + '-' + CONVERT(VARCHAR, DATEPART(YY,
@CurrentDate)) AS BusinessMonthYear,
                RIGHT('0' + CONVERT(VARCHAR, DATEPART(MM, @CurrentDate)),2) + CONVERT(VARCHAR,
DATEPART(YY, @CurrentDate)) AS CalendarMMYYYY,
                CASE
                        WHEN DATEPART(MM, @CurrentDate) IN (1, 4, 7, 10) THEN 1
                        WHEN DATEPART(MM, @CurrentDate) IN (2, 5, 8, 11) THEN 2
                        WHEN DATEPART(MM, @CurrentDate) IN (3, 6, 9, 12) THEN 3
                        END AS CalendarMonthOfQuarter,
                DATEPART(WW, @CurrentDate) AS CalendarWeekOfYear,
                (DATEDIFF(DD, DATEADD(QQ, DATEDIFF(QQ, 0, @CurrentDate), 0), @CurrentDate) / 7) + 1
AS CalendarWeekOfQuarter,
                DATEPART(WW, @CurrentDate) + 1 - DATEPART(WW, CONVERT(VARCHAR,
                DATEPART(MM, @CurrentDate)) + '/1/' + CONVERT(VARCHAR,
                DATEPART(YY, @CurrentDate))) AS CalendarWeekOfMonth,
                DATEPART(DD, @CurrentDate) AS CalendarDayOfMonth,

                --Apply Suffix values like 1st, 2nd 3rd etc..
                CASE
                        WHEN DATEPART(DD,@CurrentDate) IN (11,12,13)
                        THEN CAST(DATEPART(DD,@CurrentDate) AS VARCHAR) + 'th'
                        WHEN RIGHT(DATEPART(DD,@CurrentDate),1) = 1
                        THEN CAST(DATEPART(DD,@CurrentDate) AS VARCHAR) + 'st'
                        WHEN RIGHT(DATEPART(DD,@CurrentDate),1) = 2
                        THEN CAST(DATEPART(DD,@CurrentDate) AS VARCHAR) + 'nd'
                        WHEN RIGHT(DATEPART(DD,@CurrentDate),1) = 3
                        THEN CAST(DATEPART(DD,@CurrentDate) AS VARCHAR) + 'rd'
                        ELSE CAST(DATEPART(DD,@CurrentDate) AS VARCHAR) + 'th'
                        END AS CalendarDaySuffix,

                DATENAME(DW, @CurrentDate) AS CalendarDayName,
                DATEPART(DY, @CurrentDate) AS CalendarDayOfYear,
```

```
                DATEPART(DW, @CurrentDate) AS CalendarDayOfWeekUSA,

                -- check for day of week as Per US and change it as per UK format
                CASE DATEPART(DW, @CurrentDate)
                        WHEN 1 THEN 7
                        WHEN 2 THEN 1
                        WHEN 3 THEN 2
                        WHEN 4 THEN 3
                        WHEN 5 THEN 4
                        WHEN 6 THEN 5
                        WHEN 7 THEN 6
                        END
                        AS CalendarDayOfWeekUK,

                @DayOfWeekInMonth AS DayOfWeekInMonth,
                @DayOfWeekInYear AS DayOfWeekInYear,
                @DayOfQuarter AS DayOfQuarter,
                CASE DATEPART(DW, @CurrentDate)
                        WHEN 1 THEN 0
                        WHEN 2 THEN 1
                        WHEN 3 THEN 1
                        WHEN 4 THEN 1
                        WHEN 5 THEN 1
                        WHEN 6 THEN 1
                        WHEN 7 THEN 0
                        END AS BusinessIsWeekday


        SET @CurrentDate = DATEADD(DD, 1, @CurrentDate)
END
```

Table 3 - SQL script to insert data into DimDate dimensional table

9.4 SernikDW – SalesOrder View SQL Script

The following script was created to support a view during the 'staging' phase of ETL process. See **Create View SalesOrder.sql.**

```
/*
Continuous Assessment
Dublin Business School
B8IT104 Data Warehousing and Business Intelligence
Lecturer: Noel Cosgrave (noel.cosgrave@dbs.ie)

# ################################################

 Daniel Pereira (10391381@mydbs.ie)
 Maja Pomohaczi (10392409@mydbs.ie)
 Sylwia Wojciechowska (10392584@mydbs.ie)

# ################################################

  This script will return Orders from Northwind in a way to be stagged
*/

CREATE VIEW [dbo].[VW_Staging_SalesOrder] AS

        SELECT
                o.CustomerID AS [CustomerID],
                o.EmployeeID AS [EmployeeID],
                o.OrderID AS [OrderID],
                o.OrderDate AS [OrderDate],
                prod.ProductID AS [ProductID],
                od.Quantity AS [Quantity],
                od.UnitPrice AS [UnitPrice],
                od.Discount AS [Discount],
                (od.Quantity * od.UnitPrice)  AS [TotalPrice],
                od.Discount*(od.Quantity * od.UnitPrice) AS [TotalDiscount],
```

```
                CONVERT(money,(od.Quantity * od.UnitPrice) - od.Discount*(od.Quantity *
od.UnitPrice)) AS [FinalPrice]

        FROM
                [dbo].[Orders] o
                INNER JOIN [dbo].[Order Details] od ON o.OrderID = od.OrderID
                INNER JOIN [dbo].[Products] prod ON od.ProductID = prod.ProductID
                INNER JOIN [dbo].[Customers] cus ON o.CustomerID = cus.CustomerID
                INNER JOIN [dbo].[Employees] e ON o.EmployeeID = e.EmployeeID
```

Table 4 - SQL script to create a sales order view