

**BUSINESS INTELLIGENCE**

**MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS – MAJOR IN BUSINESS ANALYTICS**

**Intermediate Delivery**

Group 11

Gabriel Cardoso, number: m20201027

João Lucas, number: m20200758

João Chaves, number: m20200627

Luís Almeida, number: m20200666

April 2021

INDEX

[1. INTRODUCTION 1](#_Toc67926338)

[2. BUSINESS UNDERSTANDING 2](#_Toc67926339)

[2.1. Background 2](#_Toc67926340)

[2.2. Project Objectives 2](#_Toc67926341)

[2.4. Situation assessment 2](#_Toc67926342)

[3. Dimensional Model 3](#_Toc67926343)

# INTRODUCTION

The project for this class is meant to reinforce the conceptual knowledge that we’ll be acquiring throughout the course, where we can apply core concepts of business intelligence, tools and methodologies to deliver an end-to-end Self Service BI solution to support strategic, tactical and operational business decisions, for the information problem that we choose to address.

This intermediate report briefs on the data selection process, the understanding of what we initially had to work and the conceptual direction we’ve taken.

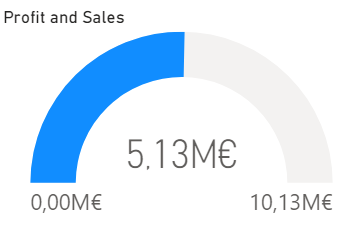
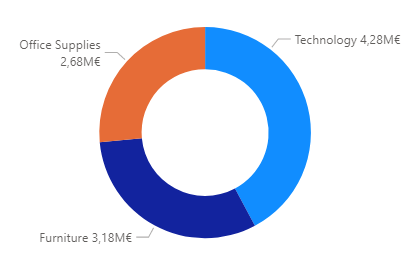
To note that the perspective we selected to approach this data was by incorporating the part of an internal Information Management/ Business Intelligence team.

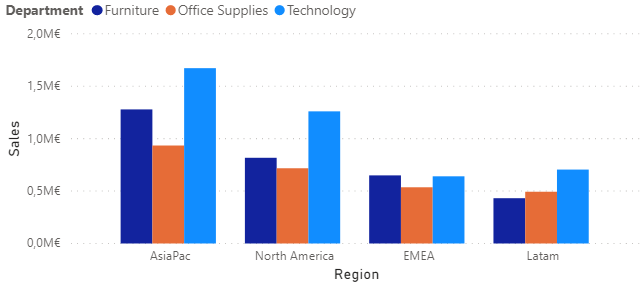
# BUSINESS UNDERSTANDING

## 2.1. Background

Superstore is a multimillion-dollar retailer with operations all over the globe. It is a recognized player, dedicated to quality in every area of business and respected for the outstanding business ethics and despite the large size and explosive international expansion, it continues to provide a family atmosphere in which employees thrive and succeed.

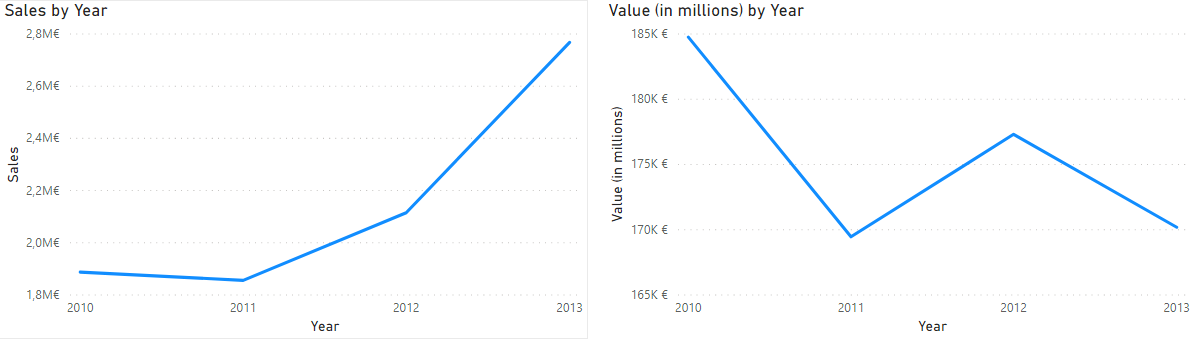
Full year results’13



These results were exceptional, representing a 37% sales raise compared to 2012. This improvement happened naturally, without any extra investment nor innovation which was intriguing for the top management.

Looking at the USA market and crossing the behaviour of our business and the whole market for the past three years there was something obvious happening.



On the left the company’s sales volume on USA market and on the right the total retail sales volume.

The conclusion drawn, was that although the total retail volume in the USA dropped, our sales volume still increased.

## 2.2. Project Objectives

Expand the business and operations of the company on USA retail market.

## 2.3. Situation assessment

Previously it was shown that indicators of SuperStore performance throught the timeline studied, however it is also important to analyse the data before implementing any dimensional modeling.

Both datasets come from the website data.world. The first one is related to a SuperStore where, after some statistical analysis, is possible to reach out to the following statements:

* Orders relative to 1313 unique products between three departments (Office Supplies, technology, Furniture) which involves 17 distinct categories ;
* Orders Location are relative to North America, Pacific Asia, Latin America, Europe, Middle East and Africa from 28 Countries and 1382 cities.
* There are 17 discount rates;
* In relation to the order priority, it has five different types of urgency (Low, Medium, High, Critical and Not Specified);
* It has three shipping modes (Regular Air, Express Air and Delivery Truck) and about 49 unique shipping costs;
* Finally, this dataset is relative to 11270 distinct orders where 7125 are unique.

For the second dataset, it contains USA Retail Peformance from 1992 until 2016. However, for a more significative and trustful analysis, it was decided to use this dataset, to compare the SuperStore performance in comparison to the USA Retail Sector, only throughout 2010 until 2013.

* Perspective of Analysis

**PORTER’S FIVE FORCES**

**PESTEL ANALYSIS**

# Dimensional Model

After an extended analysis on how the database was being created and which features were being registered, it was decided to implement a star schema modeling in order to optimize the data warehouse of SuperShop.

Before trying to implement, it’s important to understand what the processes are and where is located their respective information. Since SuperStore business sector is relative to Retail, the dimensions chosen were the following below.

Firstly, the Customer Dimension is to answer the question who is buying the product, since the profile of each customer is one of the most important assets of any business. Secondly, the place where the order is being made and sent is perfect to establish different types of policies according to the culture of the location of the customer. So, Location Dimension is to reply this issue. Thirdly, Product Dimension let us realize what is being ordered. For the following dimensions, they represent efficient ways of understanding the information better and guide the company policies in a more direct and easy direction related to shipping cost (Shipment dimension), promotions and discounts that are being done and influence the order that is being made by each customer (Discount dimension) and the impact of the urgency in terms of the sales and the respective product ordered (Priority dimension). Finally, the Date Dimension refers to the question when the order is being made and shipped to perceive if there are seasons when the product is being bought more often.

It’s also important to notice that this project contemplates four hierarchies on what all of them have four levels of depth. Those can be found on Location Dimension (Region, City, State, Country), Market Cap USA and in Order Date and Ship Date (Year, Quarter, Month, Day) from Date Dimension.

Finally, this modeling is finalized with a Fact table that has information relative to metrics relative to the business activity of SuperStore. In this specific case, it tracks all the orders being made, the respective amount of revenue and profit of each sale/order. It has three additive facts (Order Quantity, Sales and Profit) and one non-additive fact related to the order identification number. In order to measure this type of an event, this fact table contains foreign keys for each of its associated dimensions.

