



BUSINESS INTELLIGENCE

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

Project Report

Group 11

Gabriel Cardoso, number: m20201027

João Lucas, number: m20200758

João Chaves, number: m20200627

Luís Almeida, number: m20200666

April 2021

INDEX

1. INTRODUCTION	1
2. BUSINESS UNDERSTANDING.....	2
2.1. Background.....	2
2.2. Data understanding	2
2.3. Business needs and problems	3
3. Dimensional Modeling.....	4
4. Perspectives Description & Assessments	6
5. Conclusions.....	12

1. INTRODUCTION

Understanding data whilst being able to express it quickly and accurately is a necessity in this modern age of business decision making.

This project is meant to reinforce the conceptual knowledge that we will be acquiring throughout the course, where it can be applied 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 chose to address and future projects.

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

It is noteworthy to state that the perspective we have selected on how to approach the data was by incorporating part of an internal Information Management/ Business Intelligence team.

2. 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.

SuperStore is a multi-national level retailer and has its business tendrils reaching far and wide. As such, integration locally and compliance with regional and political policies is a fundamental part of making business and ensuring success in the long run. As this company sells pretty much everything from office related materials (and even art supplies) to telecommunications, these are not markets in which differing political regimes or orientation can be troublesome to the business, except for the telecommunications market, in which there is a major problem in what regards information flow and data protection. With the rise of data circulation, some regimes with more authoritarian approaches can deter our operations on that front.

The economy is still recovering from the impact of the 2008 recession. However, across time we see that the growth is a normal trajectory which coincides with the cyclical nature of the macroeconomic trends. From our data, we found that the market is struggling, but our growth does not follow suit. As such, even though the value of the market is decreasing, the regenerative nature of the economy can explain the fact that we are thriving.

In this way, we can take the conclusion that with the crisis, the survival of the fittest got accentuated, and our growth (and the growth of other thriving competitors) was not able to offset the bigger reduction in market value. Despite this, there is a position of power to take in the market, as other players leave. This is a major opportunity to strengthen our brand and positioning.

2.2.DATA UNDERSTANDING

The Super Store dataset contains data on order details of customers for orders of a superstore point of sale from different locations across the world. This includes the state, region, order date, shipping date, product ordered etc. Throughout this project, we'll define some use cases for this dataset in order to provide a trustful and meaningful analysis with the objective of creating business values and opportunities for SuperStore. Examples of those could be questions related to product sales and category trends, sales and demand forecasting, operational costs analysis and optimization and financial sustainability and performance (profit analysis).

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 Performance 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.

2.3.BUSINESS NEEDS AND PROBLEMS

Although SuperStore is not on any red-light alarm in terms of financial situation, their management required us in order to keep their growth in a positive and steady state. Since it has tons of different point of sales across the world, the main objectives of our work will focus in analysing our dataset to find new strategies and insights that can help us take better strategy decisions for the future of SuperStore at three different perspectives, which are the following:

- Operational, since due to the huge size of our company, SuperStore is dealing with enormous quantities of inventory and their respective mobility can be seen as a bigger slice of their costs. So, it is in our best interest to understand how we can optimize our costs in terms of warehouse location and store's location (p.e reducing the number of warehouses and stores that don't sell a lot in several states, could be use to decrease the shipping costs), identify the main characteristics of orders with the lowest shipping costs and how can it be replicated to the rest of the orders;
- Market wise, since our main goal is to reinforce our position in the United States, is in our best interest to analyse what products are more popular in the US, how can we improve our presence, and how can we replicate our successes in the rest of the world in this specific country. It is important to compare, prices, quantities and unit and shipping costs with other countries where we are present.

3. DIMENSIONAL MODELING

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 SuperStore.

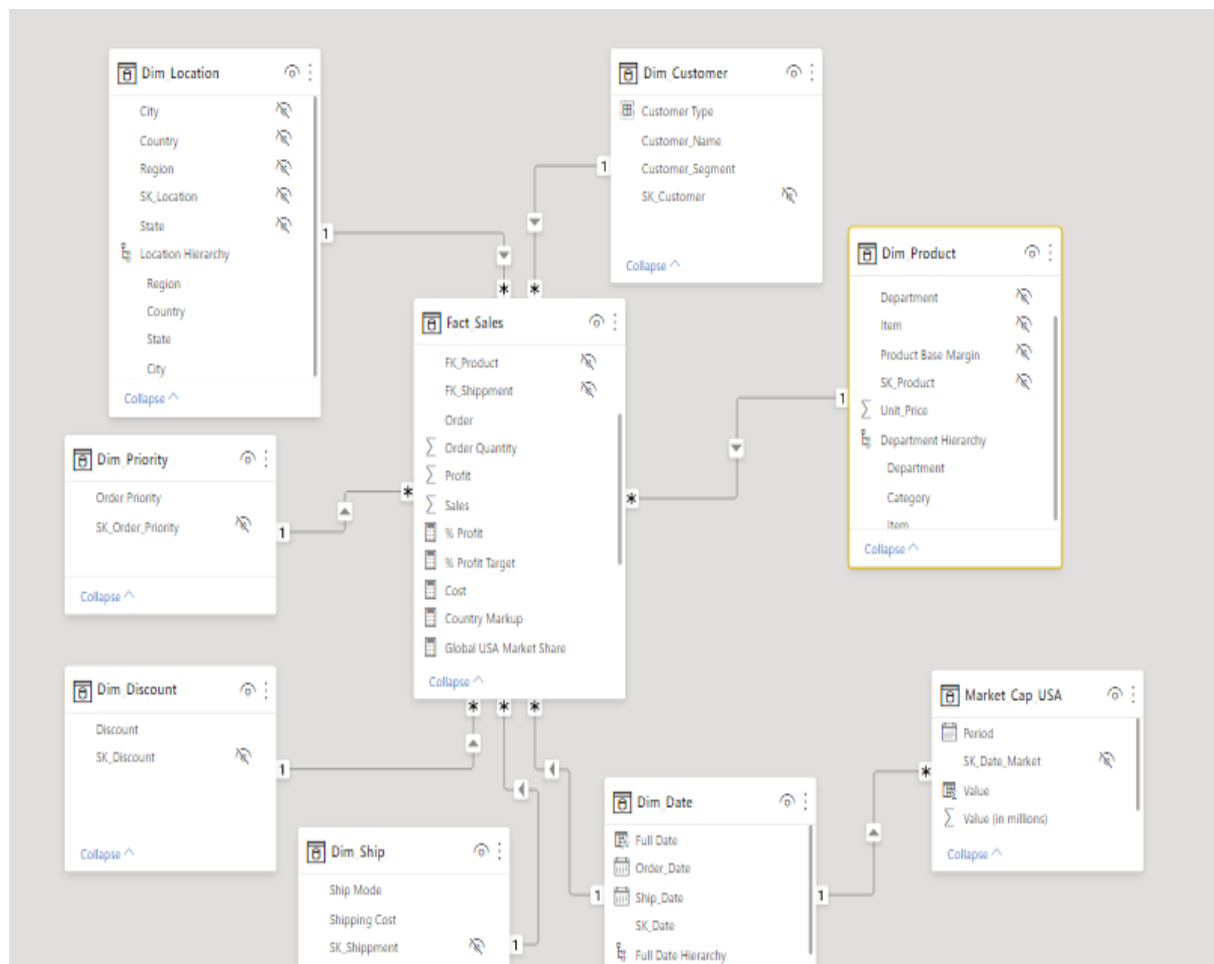
Before trying to implement, it is important to understand what the processes are and where is located their respective information. Since SuperStore business sector is relative to Retail, the dimensions chosen were the following below.

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 is also important to notice that this project contemplates five hierarchies. Those can be found on Location Dimension (Region, City, State, Country, four levels of depth), Market Cap USA and in Full Date (Year, Quarter, Month, Day, four levels of depth) from Date Dimension and in Product Dimension (Department, Category and Item, three levels of depth).

Finally, this modeling is finalized with two Fact tables. The Fact_Sales has information relative to metrics relative to the business activity of SuperStore and the Market_Cap_USA has Information regarding the American Retail Sector. In the Fact_Sales 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. Then in the Market_Cap_USA we have the additive fact of the Value (in millions) and is connected with the Fact_Sales by the date for comparison and performance purposes.

To conclude, here is our Dimensional Model:



4. PERSPECTIVES DESCRIPTION & ASSESSMENTS

In total we approached two different perspectives to which there were some measures and calculated columns created in order to properly go through the available information and draw our conclusions.

Below we're going to dissect those measures and see what were their uses and consequent conclusions.

4.1. MEASURES AND CALCULATED COLUMN

The different measure and calculated columns were used in both operational and market perspectives. Regarding the measures we've calculated 15 different ones that can be divided in three different categories, Profit related, Sales related, Market related:

- **Profit related:** % Profit, % Profit target, LY USA Profit, Total Profit, USA % Profit, USA Profit – Those profit measures as the names indicate give mostly the same information but divided between the global analysis and the USA analysis;
- **Sales related:** Sales Amount, LY USA Sales, USA Sales Amount – Follows the same logic above;
- **Market Analysis:** Global USA Market Share, LY USA Sector Mkts, USA Internal MKT Share, Correlation Coeff (SalesVS Sector Volume), Sector Volume;
- **Cost:** This measure is specific only for the operational analysis whereas the profit and sales related are diluted within the operational and the market perspectives.

As for the calculated columns it was created one, the Value on the Market_Cap_USA Dimension, which is the total value of the sector since the original information was the Value in million.

4.2. OPERATIONAL ANALYSIS

For the operations we analysed how our sales around the world were distributed and which departments and categories were more profitable and had the biggest volume. The USA market is our biggest market and the one where we want to focus our analysis, so we will try to compare worldwide insights with the USA market. To accomplish that we created three report pages with information mainly regarding this perspective, the Global Overview, USA Market and USA Product cost.

For future consideration, the **% Profit is one of the most important kpi** to evaluate our performance, where from < 45% the performance is poor and improvements are due, 45 – 55%

is an attention level, which is above our minimum established objective but is within a close range and lastly 55% > where the performance is deemed as excellent.

In the **Global Overview** report page it is obvious the company chose to give more attention to the USA market, since it is the most profitable one worldwide. We also can observe that the global % profit is 48,9%, representing a good enough performance but still out of the targeted range. Also the top 4 profitable product categories represent around 50% of all the retained profit although 'Office Machines' have a below minimum % of profit which indicates that a revision is due to happen. Selecting USA in the map and looking into the product performance we observe that 'Telephones and Communication' has the highest sales amount and a % of profit very close to our excellence objective, similarly 'Chairs & Chairmats' category is above excellence objective but a little lower on the sales amount. Once again the 'Office Machines' has a behaviour close to its global performance. From the whole categories there are 7 underperforming and the top 3 represent around 29% of the total sales amount, with this we can conclude that 'Office Machine', 'Copiers & Fax' and 'Binders & Binder Accessories' are the main products to look into and increase profitability within our goals.

The second report page dedicated to this perspective is the **USA Market**, that as the main objective of see the company's performance on the american market. It has two kpis to control performance, the USA % Profit Performance and the USA Retail Sector Market Share, a reference of the correlation between the USA Sales amount and the American Retail sector total volume to have in mind the effect of the variations present in the Area Chart, which displays the USA Sales amount (millions) and the Sector Volume (billions) per Year, with the option to drill the year down and look into that year's months behaviour. Also there is a line chart with the display of the information regarding USA Sales Amount in the previous years and with a sales forecast for the year to come with a confidence interval of 95%.

Looking into the overall profit margin we see a global average of about 47,76%, which although is not underperforming is very close to it, which indicates a need for taking contingencies with a medium priority. On other kpi the Market Share has increased steadily in the last years, starting at 0,0029% in 2010 and increasing to 0,0049% in 2013, although this market share is especially low because of the huge volume of the USA retail sector, this represents a relatively big increase in the SuperStore market share. The Forecast made by our analytical model looks conservative at first view but since we applied a confidence interval of 95% we have some optimistic numbers close to the upper bound, which indicates us that maintaining our operations and improving on some specific and relevant points we are able to achieve those optimistic values.

The USA global retail market has been unstable in the last years, experiencing huge declines in the last 4 years, however our company's total sales in that market have been increasing steadily since 2010 with a huge increase in the last year, which indicates that the company has been working well and the opportunity of grabbing more stake in this market is achievable. On another note, we decided to display the correlation between our company's USA sales amount and the Total USA Retail Sector Volume, to have a constant notion of the real impact of the sector's fluctuations.

The last report we created for this perspective is the **USA Product Costs** where we aim to understanding the discounts and costs impact on performance. This report has a table organized by discount, sales amount order quantity, cost and consequent % profit, where we can understand their interactions, a ribbon chart with the visual representation of the costs associated per department and a multi-row card with all the product categories per state with associated costs above 50k.

Overall, the costs increased in all departments through the studied years, especially from 2012 to 2013. Worth to mention that looking into the Office Supplies, costs have been decreasing since 2010 and then a big increase in the last year. SuperStore also applies different discounts to its products which can have an impact on the profit margin. Taking in special focus the USA market we looked into the values associated to the different discounts per department.

It is interesting to observe that sales with a 1% discount were the only ones with a % profit above our excellence level whereas sales with 0% are on our attention zone, this tells us that the discount/sales ratio is not correctly set and needs our attention on further process improvements.

Furniture is the most profitable department, with a very low cost when compared to the sales amount, which allows to have discounts up to 10% and still remain very profitable. Above the 50k cost mark there is only one category with a 54K costs amount but is not a concern since the % profit is set at a value of 61% which is well above the excellence lower limit.

In Office Supplies, we observe that the amount of products sold at discounts is evenly distributed between discounts that go from 0 to 10%, so taking into account that 1% and 6% are the dicounts with better % profit margins and even so performing lower than our goal, it could be a good strategy to study further into this department's finances and try to cut some costs, starting by the 'Binders and Binders Accessories' as in the District of Columbia and in New York their performances are critically low, displaying high costs and low % profit.

Technology, has the odd behaviour where the sales without any discount have a performance below expectations and there isn't one category in this department with a % profit at excellence levels. Observing the categories per state with the cost above 50K we can confirm that the Office Machines have extremely high costs and poor % profit associated, inclusive the Technology category with the highest cost is precisely the Office Machines in Illinois, displaying 70k in cost and a 25% profit margin. We already know that the Office Machines has a high sales amount in the USA so actions to decrease the costs and cut the discount sales on this category must be taken promptly.

4.3. MARKET ANALYSIS

For the Market perspective we focused on accessing information of customer behaviours and geographic patterns that could come up and give useful insights. To do that we created three report pages, **Client Type Profile**, **Product Analysis** and **USA Product Distribution**.

Superstore is in different types of market across the world, being the most profitable USA, China and India respectively where our clients are segmented into:

- Corporate
- Home Office
- Consumer
- Small Business

We decided to also aggregate our clients into four different clusters according to their orders amount and the profit margins they generate, as so we have:

- Cash Cows, clients that buy in high quantities at high margins of profit;
- High Profit, clients that buy in less quantity but also at high margins of profit.
- Medium Profit, clients that buy in less quantity and at medium to low margins of profit.
- No Profit, clients that but small quantities and at low to negative margins of profit.

Observing in further detail the profile of our USA clients there is a predominance of high and medium profit clients, with few cash cows and No Profits.

Regarding on how our profit unfolds and what kind of clients contribute to it, **Corporate** clients represent the most profitable ones (1.343.127€) being that the department where they spent the most is Technology (580.637€), followed by Furniture (505.551€) and Office Supplies (256.937€).

The second most profitable clients are the **Home Office** with 849.489€ of profit, they also have a preference for Technology products (336.042€), folowed by Furniture (302.937€) and Office Supplies (210.509€).

Next we have the **Consumers** that represent 748.717€ of the profit and have a preference for Furniture products (292.289€) followed closely by Technology (292.033€) and Office Supplies (164.394€).

Finally, our least profitable clients are the **Small Businesses** as they represent 676.953€ of the profit and have a preference for Technology (259.820€), Furniture (233.772€) and Office Supplies (183.360€).

On a different prespective and analysing how our different categories and items contribute to our profit worldwide, the Technology department has 45,5% of profit, Furniture with 58,1% and Office Supplies with 43%.

Going in further depth inside each department;

The most profitable items in the Technology department are: Canon PC1080F Personal Copier (119.490€ of profit), Canon PC940 Copier (113.219€ of profit), Adesso Programmable 142-Key Keyboard (112.516€ of profit); while the less profitable are the Polycom VoiceStation 100 (-600€ of profit), Polycom ViaVideo desktop unit (-28€ of profit), Sony IBM Color Diskettes (30€ of profit).

The most profitable items in the Furniture department are: Riverside Palais Bookcase (249.585 of profit), Global Troy Executive Tilter (214.555€ of profit), Bretford CR8500 Furniture (162.990€ of profit); while the less profitables are the Hon Non-Folding Unit Table (-3.351€ of profit), 3M Polarizing Light Filter Sleeves (-38€ of profit), 3M Polarizing Task Lamp (-27€ of profit).

The most profitable items in the Office Supplies department are: Sauder Facets Cabinet (78.870€ of profit), Tennsco Shelving Units (76.523€ of profit), Office Impressions Heavy Duty Drawers (73.071 of profit); while the less profitables are Ibico EPK-21 (-526€ of profit), Ibico EB-19 (-472€ of profit), Hoover Commercial Vacuum (-340€ of profit).

Also on the Product Analysis report page, looking into the Less Profitable Products list we observe that the bottom 3 products come from different departments, being the worst performer a product from a the most % profitable department which comes as a surprise not only for that fact but also for being 3x bigger than the next unprofitable product. To note, that although the negative value is close to even, the 3rd unprofitable product is from the Office Machines category, which not only further reinforces this category's previous assessments but also gives us a priority within the category. Unsuprisingly the top 3 most profitable products of the USA market are all from the Furniture department.

Regarding orders amount and profit through these years, is worthy of note that the Technology department, that represents 38.38% of all the orders, experienced the biggest development, increasing almost 2x the profit from 2010 to 2013 and increasing 50% the orders amount, this indicates that the rise of this department was one of the main reasons for our development, especially with the uncertainty and huge competition on the USA Retail Market. The Furniture department increased overall in profit and in ordered quantity, suffering from a decrease in profit in 2011 but responded effectively in 2012 and mantaining the level in 2013. The Office Supplies department, represents 40,65% of the total orders amount had a similar behaviour as the Furniture department in regard to the profitability with the particularity of the orders amount being more steady.

Drilling down the year's into the overall months, it's observable that the sales amount from January until April are low and steady, then experience a mini spike in May, only to pull back in June, from July the orders amount rises steadily to peak in November, out best month of the year, and pull back in December.

On the USA Product Distribution report page we get the perception of the geographical distribution of the categories sales aligned with the monthly seasonality variation based on the Order Quantity and Sales Amount values for the movement and the size representing the % profit. Generally we can state that the bulk of our clients are from the East side of the country,

far from it, the West Side has a fair amount of orders whereas the North, Centre and South are underperforming and lack sales.

Looking into the Office Machines, the geographical distribution is very sparse, demonstrating to be more dense over the East Coast and a fair performance over the West Coast. Regarding the seasonal behaviour, to note the size of the bubble along the months, specially on the most profitable ones for the company. From September to December, we can understand why this category is underperforming on the % profit, in September the sales amount and the ordered items are at a absolute peak but the size of the bubble indicates that % profit level is very low, onto October where that size barely changes but the sales amount remain high and the ordered amount decreases by almost 50%, November the % profit has a discrete improvement and the sales amount remain stable has the orders amount increases and finally in December is where the sales amount increases and the ordered amount decreases which reflects in a huge improvement on the % profit.

Overall, all categories have increased compared to the previous year (+1,15M€) which is a positive outcome and a obvious statement from our company in this market and a very positive result of all the changes implemented.

5. CONCLUSIONS

Globally the business is growing and there is an opportunity to improve internally in order to monetize our current growth on the market. On the global perspective, it's interesting to identify the product category of Office Machines as a potential improvement since both globally and in the USA has a huge market share on the company's sales and underperforms all the same.

Our focus on the USA market wasn't random and although the company already has a comfortable position there we understood that the global performance is highly affected by the USA performance and since it's our main market, the focus on optimizing operations and identifying new market opportunities is only natural and completely justified. It is important to state that the correlation of our sales and the global USA retail market performance aren't highly correlated, thus the team understood that the USA performance could be improved despite the USA Retail market fluctuations.

The clear priority for improvement of our position in the USA market is the product category Office Machines, looking into it we can easily understand that the costs are high and the % profit is low which led us to conclude that the **first counter measure** to surpass this issue is to review the discounts given and to **adjust the costs with the sale price**, being this regarding manufacturing or transportation, so we suggest a Continuous Improvement Lean/Kaizen Project to solve this issue.

Focusing only on the Office Machines improvement and looking into its geographical distribution knowing that the worst costs associated for this category are in Illinois (70k – 25,2%), New York (67,5k – 41,8%) and California (51,7k – 44,4%) respectively, we concluded we understood that California is a big state with a lot of sales potential and its surrounding states are low sales states, so having a more impactful presence in California could be key to expand our business in the West Coast, so with this in mind we **suggest expand our Operations there in order to decrease the costs associated to the products** and work on **targeted marketing campaigns** to better understand its particularities and understand the more appropriate timings to make discounts.

Looking into the seasonality on Office Machines in California, we observe that the bulk of the demand is between September and December but September is the month with the highest demand but with a low % profit coming from it. Having that in mind we would advise to cut the discounts ratio on that particular month or associate these high demanded products with other profitable ones with a joint balanced discount. In October due to decrease of the orders we would advise to work on those marketing campaigns and sell product with a decreased profit margin, that way we'll work on maintaining September's orders amount and given the right discount the sales will naturally also increase. November and December are highly profitable, with a high % profit associated to large orders amount so we suggest either not do anything or associate low performing products within this category to the high performing ones.