BUSINESS INTELLIGENCE REPORT – GROUP 006

Spring Semester 2021

MASTER'S IN DATA SCIENCE AND ADVANCED ANALYTICS

Course Coordinator: Professor Miguel de Castro Neto

Teaching assistant: Bruno Jardim

MENDELEY – SMART SUPPLY CHAIN

Project team:

HENRIQUE COSTA // [20200652]

LILIT TONOYAN // [20190930]

NGUYEN HUY PHUC // [20200566]

PEDRO MENDES // [20200648]

TABLE OF CONTENTS

List of figures

1) INTRODUCTION

This project was proposed as part of the Business Intelligence course for the master's in Data Science and Advanced Analytics at Nova IMS for the Spring Semester 2020/2021. This project is meant to reinforce the conceptual knowledge acquired throughout the course and deliver an end-to-end self-service BI solution to support the different levels of business decision-makers.

This report will describe, in a detailed fashion, the core concepts of business intelligence, tools, and methodologies used on the development of a dimensional model and related Power BI dashboard.

We will start with a short introduction of the chosen company by explaining the type of business it does and how it operates around the world. In addition, we will do a description of the data source used to build our dimensional model. Moreover, we will identify the business needs and the perspectives of analysis chosen to build our Power BI dashboard. Finally, an analysis and critical assessment of the results will also be shown.

2) DESCRIPTION OF THE DATA SOURCE

This project uses a fictitious company that we have found on the Mendeley data website [4]. Three CSV files were available: the first one with 180 519 observations and 53 features of which 29 are numeric and 24 categorical variables and has data on customers, orders, products, and shipping details; the second dataset has 469 977 observations and 8 features including details about the products available; the third dataset had the description of the variables (shown in Annex 1). Some assumptions regarding the company business model and name had to be made to make the company credible and presentable, always taking into consideration the data available and the BI project proposed.

As suggested in the project guidelines, our dataset includes transactional records that can represent quantifiable facts and will be used in our dimensional model. In terms of Business Intelligence, transactional records are "data that is used to count, sum or otherwise keep track of quantities of something that the organization deeply cares about" [5]. For this purpose, we have data on shipping days (real and scheduled), profit per order, sales per customer, quantities per order, product discounts, and product prices.

In terms of dates, our database runs from January 2015 to February 2018, fulfilling the requirement of the guidelines of more than one year of transactional history. We have however only considered the data until October 2017 as the data after that is very erratic, with no clear trends or patterns, and so we have decided to disregard those last 3 months as this would not affect the project guidelines and would make our conclusions more robust.

3) PRESENTATION OF THE BUSINESS AND DISCOVERY PROCESS

The company has chosen (hereby called Mendeley) is a multinational company that ships products all over the world in a timely and cost-efficient manner.

Our dataset has data from January 2015 to September 2017 (inclusive) where Mendeley had sales of $31 million and profited $3.7 million while selling around 375,000 products. Most of the sales, almost 50%, were from the Fan Shop department which includes more sports-related categories from hunting & shooting to fishing and water sports. The apparel, golf, and footwear departments are also very significant and, together with the Fan Shop, represent more than 95% of the sales.

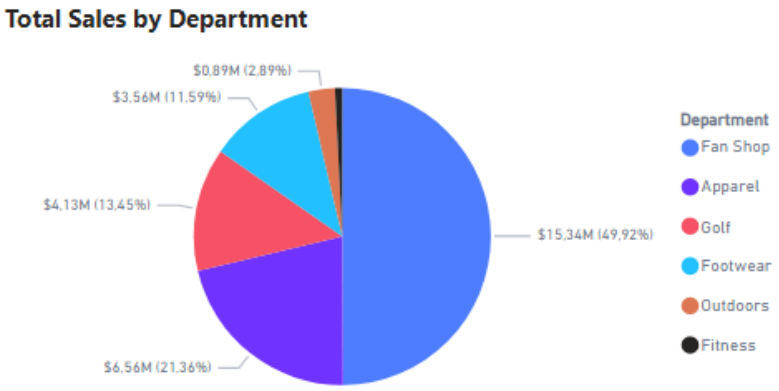

Figure 1 - Total Sales by department

Mendeley has warehouses all around the United States (US) and Puerto Rico (PR) and ships its products all over the world. The US warehouses ship 60% of the total number of products, while PR's ship the remaining 38,5%. It is worth noticing that we have around 1,5% of the products with an unknown origin which we have decided not to delete as they represent legitimate sales, but this is an opportunity for improvement for Mendeley as knowing from which warehouse each product is shipped is business-relevant information.

Mendeley's biggest markets are Latin America and Europe which, together, represent about $17 million in sales. Pacific Asia, USCA (the United States and Canada), and Africa are also important markets that must be taken into consideration. represent about $17 million in sales. Pacific Asia, USCA (the United States and Canada), and Africa are also important markets that must be taken into consideration.

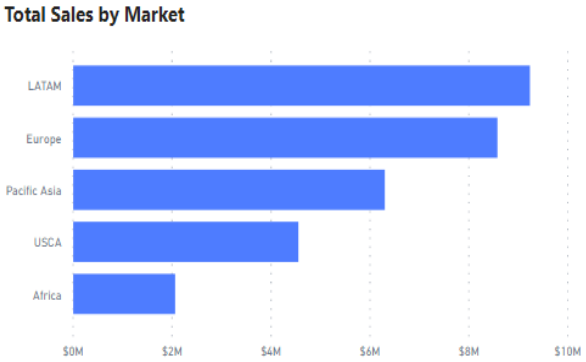

Figure 2 - Total Sales by Market

Regarding Mendeley's biggest client segments, private consumers represent around 50% of the total sales, followed by corporate clients and home offices.

From the supply point of view, Mendeley has registered a total of 57 thousand orders with an average of 3 units per order and average order.

It is also important noticing that Mendeley's sales and profits have remained relatively constant throughout the analyzed period as observed on the graph below (please take into consideration that the decline in both the sales and profit amount in 2017 is largely explained by the fact that only 9 months are being considered versus the 12 months in both 2015 and 2016).

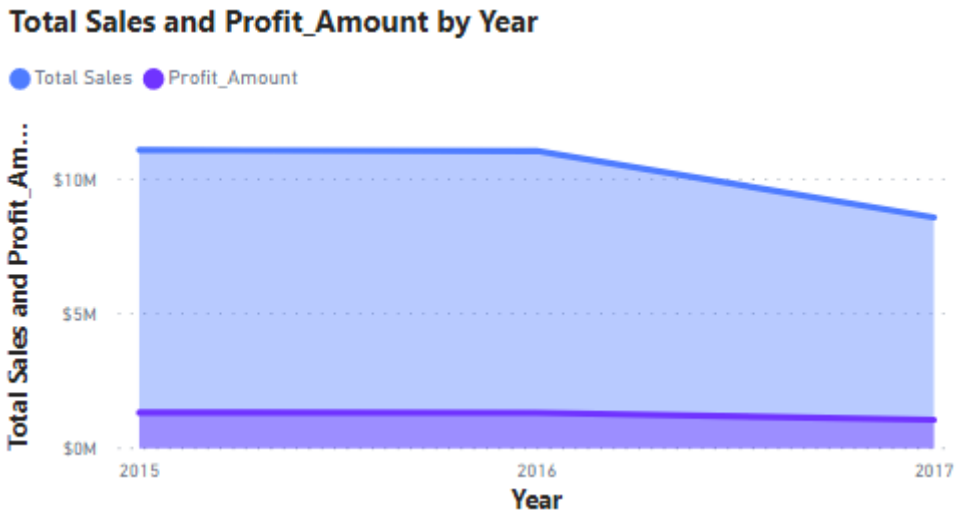

Figure 3 - Total Sales and Profit Amount by year

4) IDENTIFICATION OF BUSINESS NEEDS AND PROBLEM

The initial analysis made on Mendeley's dataset helped us to identify its main trends and business needs. Based on this analysis, our group has highlighted two issues that need to be addressed and will be explored in this report:

1) We have noticed on our initial analysis that although overall there is a profit of 3.7M, some products in specific countries are losing money. It is important to analyze this problem and create visualizations that make it easier to identify these products so further actions can be taken.
2) Mendeley's projections on shipping times are consistently underestimating their real shipping times: in fact, only 43% of the orders arrive on time (i.e., before the initially scheduled date). This issue can greatly affect customers' expectations and satisfaction, in fact, 69% of consumers "are much less or less likely to shop with a retailer in the future if an item they purchased is not delivered within two days of the date promised." [1] Some research even suggests that "17% of respondents will stop shopping with a retailer after receiving a late delivery one time" [2]. Moreover, a Harvard Business Review article [3] suggests that acquiring new customers can cost anywhere from five to 25 times more than retaining an existing customer, thus making it imperative that this topic is analyzed.

5) DEFINITION OF PERSPECTIVE OF ANALYSIS

Having in mind the issues previously identified, we have decided to analyze our dataset from two different perspectives:

5.1. Sales

In this perspective, the sales and orders data will be recorded and summarized so meaningful business insights can be drawn. The report will allow sales managers and executives to monitor sales revenues, number of orders in different regions, product categories and stores. After our analysis and the final report are done, we aim to explain not only how Mendeley is currently performing but also to understand its weaknesses and how they can be improved. There are several questions and metrics to analyze in this perspective, for instance:

- What are the key departments in terms of sales?
- Where is(are) the key market(s) of the company?

- Is there any seasonality in the products purchasing patterns?
- KPIs on sales, quantities, and profit margins – Annually, quarterly, monthly sales
- Whih country shipped most products?
- How did each store perform compared to the prior year?
- What are the most and least profitable products?

5.2. Supply Chain

Mendeley dispatches its orders all over the world and it is important to understand the shipping times of those orders, where the orders are shipped from, where they arrive and which dispatching option is used by the customer. In this perspective, the model will provide detailed monitoring across the delivery operations of the company, including metrics such as on-time delivery, shipping mode, and the status of each order to identify delivery issues at any point of the process. The operations manager and executives will be able to drill down into shipment operations history to understand the strengths and weaknesses of their current operations, thus having sufficient information to carry out effective decisions and continuous improvements. The key areas to analyze in this perspective can be described below:

- Measure on-time delivery KPI
- Quantify orders delivery status
- Average shipment time to each region in the world
- Quantify preferred shipment mode

6) DESIGN AND DESCRIPTION OF THE DIMENSIONAL MODEL

6.1. Identify the business process

As clearly described in the previous section, the business objective of this data warehouse will be to address the metrics and performance of Sales and Supply Chain activities of the company. Fortunately, the columns available in the source file is sufficient to provide information needed for calculation and aggregation of the sales metrics and delivery metrics such as sales quantity, sales revenue, profit ratio, delivery status, etc.

6.2. Identify the granularity

The lowest level of detail in the data source can be described as an invoice transaction line which contain the information of **each product purchased per each order per day**, along with information such as purchasing price, purchasing quantity, order date, customer information, store information, order information such as location, status, market, and more.

6.3. Identify the dimensions

The information available for each row in the data source file can be used for creating dimension tables as described below:

Dimension tables	Description	Columns' name
Dim Product	*Basic information about each product item*	*Name, Price, Category and Department*
Dim Store	*Data about the location of the store and its segment (e.g., Home-office or Retail customer)*	*Segment, Street, City, State, Country, Latitude, Longitude*
Dim Payment	*Data about the payment of each order such as payment type (e.g., cash or card) and the status of the order*	*Payment Status, Payment Type*

	(e.g., On hold, Canceled, Complete, etc.)	
Dim Shipment	*Data about the shipment status of each order such as ship mode (e.g., Standard or Premium) and the status of the delivery (e.g., On time, Late, etc.)*	*Delivery Status, Shipping Mode*
Dim Order Location	*Data about the location of the order*	*Market, City, State, Country, Region*
Dim Date	*Dimension of date for analyzing order date*	*Order Date*

Figure 4 - Dimension Tables description

6.4. Identify the facts

There are several transaction facts already available in the original data source to create a transaction fact table: Total sale amount, Product quantity, Profit ratio, Discount rate, Actual shipping days, Scheduled shipping days.

1.2 DiscountRate	1.2 ProfitRatio	1²3 Quantity	1.2 TotalSales	1²3 ProductStatus	1²3 ActualShipDay	1²3 ScheduledShipDay
0.039999999	-0.270000011	2	115.1800003	0	5	2
0.150000006	0.310000002	2	67.98000336	0	4	4
0.150000006	0.479999989	1	110.4899979	0	6	4
0.039999999	0.209999993	1	287.980011	0	3	4
0.039999999	0.349999994	2	96	0	6	2
0.100000001	0.330000013	1	116.9899979	0	2	4
0.01	0.25999999	5	494.9500122	0	6	2

Figure 5 - Facts for transaction fact table of sales

Beside these available facts, there are several measures that can be further calculated and aggregated when developing the detailed report for each perspective. Below are some suggested approaches for the measures supporting the analysis.

Analysis on the sales perspectives will help indicate the business' ability to generate revenue, measuring the success of each product offered, each market and online activity. Some of the critical sales metrics for monitoring are: Total revenue, divided by products, stores and regions; Total profit; Number of Orders; Average order value; Top performing products, etc.

Concerning the continuous growth, development, and success of Mendeley's delivery efforts, supply chain performance metrics are the most invaluable tools. Some of the significant metrics for Mendeley are: Perfect Order Rate, On-time Shipping; Delivery time, Number of shipments, etc.

After identifying all the possible design patterns for the data warehouse, in the next section, we will describe the actual processes to create the data warehouse for this report.

7) DATA INTEGRATION, TRANSFORMATION, AND MODELLING

7.1. Import the data

The data is imported under "csv" format. There are 2 flat files to be imported: "DataCoSupplyChainDataset.csv" which contains the sales data and "tokenized_access_logs.csv" which contains the web access data.

To optimize the data reading process when working on multiple computers, we create an automation procedure to edit the data source path for all the tables. Whenever the file is opened and edited on a different computer, the user just needs to adjust the path variable showed in queries with the path that contains the project folder, and all the tables' source will be automatically updated accordingly.

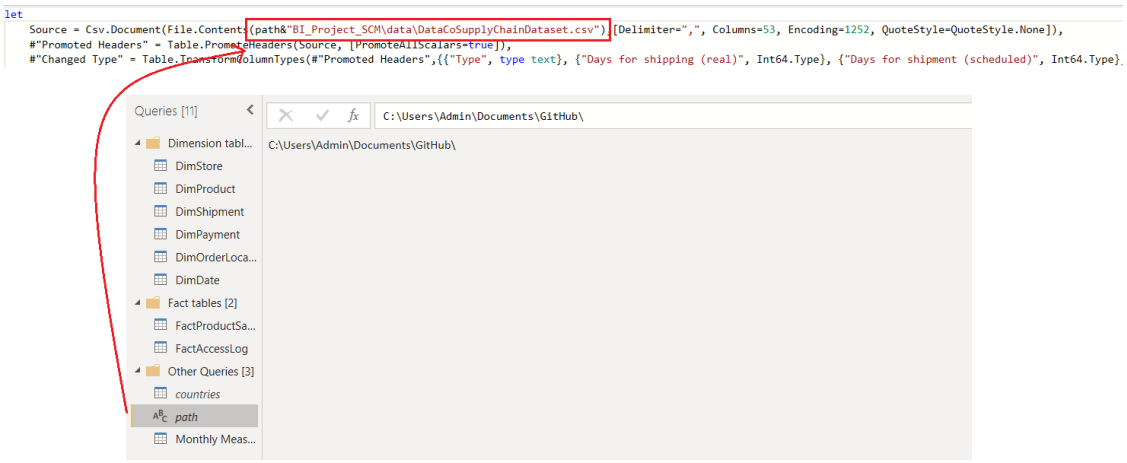

Figure 6 - Importing the data

7.2. Normalize the data source

The original data is a single flattened file that contains all the columns in one file. Thus, the first process that we do is to normalize the file and convert it into different dimension tables and fact tables using Power Query Editor.

7.2.1. Generate dimension tables

The process of normalizing the file is the same for each dimension table. Firstly, the imported flat table is duplicated into a new table. And then we removed all the columns that are not selected for the target dimension table. Then all the duplicated rows in the table are removed. If the table does not have a primary key as index yet, a new index column will be created and treated as the surrogate key column. And finally, to create a relationship with the fact table, unless there is already a foreign key in the fact table to create the relationship directly, the columns that can be used as **composite primary key** will be concatenated together and later be used to join with the fact table in which there will be the same concatenated columns.

SK_Shipment	DeliveryStatus	ShippingMode	FK_formerge
1	Advance shipping	Standard Class	Standard ClassAdvance shipping
2	Late delivery	Standard Class	Standard ClassLate delivery
3	Shipping on time	Standard Class	Standard ClassShipping on time
4	Shipping canceled	Standard Class	Standard ClassShipping canceled
5	Late delivery	First Class	First ClassLate delivery
6	Late delivery	Second Class	Second ClassLate delivery
7	Shipping canceled	Second Class	Second ClassShipping canceled
8	Shipping on time	Same Day	Same DayShipping on time
9	Late delivery	Same Day	Same DayLate delivery
10	Shipping canceled	Same Day	Same DayShipping canceled
11	Shipping on time	Second Class	Second ClassShipping on time
12	Shipping canceled	First Class	First ClassShipping canceled

```
let
    Source = Csv.Document(File.Contents(path&"BI_Project_SCM\data\DataCoSupplyChainDataset.csv"),[Delimiter=",", Columns=53, Encoding=
    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Type", type text}, {"Days for shipping (real)", Int64.Type}, {
    #"Removed Other Columns" = Table.SelectColumns(#"Changed Type",{"Delivery Status", "Shipping Mode"}),
    #"Removed Duplicates" = Table.Distinct(#"Removed Other Columns"),
    #"Added Index" = Table.AddIndexColumn(#"Removed Duplicates", "Index", 1, 1, Int64.Type),
    #"Reordered Columns" = Table.ReorderColumns(#"Added Index",{"Index", "Delivery Status", "Shipping Mode"}),
    #"Renamed Columns" = Table.RenameColumns(#"Reordered Columns",{{"Index", "SK_Shipment"}, {"Delivery Status", "DeliveryStatus"},{"S
    #"Added Custom" = Table.AddColumn(#"Renamed Columns", "FK_formerge", each [ShippingMode]&[DeliveryStatus])
in
    #"Added Custom"
```

Figure 7 - Example of the steps in creating a dimension table from the flat file

The above processes would be applied the same for the creations of all 5 dimension-tables DimStore, DimProduct, DimPayment, DimShipment, and DimOrderLocation. Below are the concatenated columns to generate the surrogate key for dimension tables.

Dimension table	Store	Product	Shipment	Payment	Order Location
Composite key / Surrogate key	Store ID	Product ID	Delivery Status + Shipping Mode	Type + Status	City + Country

The date dimension is generated manually without importing any data from another source. The date table is then limited with the time range of the original dataset. From the perspective of analysis, it is decided to only generate the date hierarchy up to a month as the lowest level.

```
let
    Source = List.Dates(#date(2015,1,1),Number.From(#date(2018,5,1)) - Number.From(#date(2015,1,1)), #duration(1,0,0,0)
    #"Converted to Table" = Table.FromList(Source, Splitter.SplitByNothing(), null, null, ExtraValues.Error),
    #"Renamed Columns" = Table.RenameColumns(#"Converted to Table",{{"Column1", "Date"}}),
    #"Changed Type" = Table.TransformColumnTypes(#"Renamed Columns",{{"Date", type date}}),
    #"Inserted Year" = Table.AddColumn(#"Changed Type", "Year", each Date.Year([Date]), Int64.Type),
    #"Inserted Day" = Table.AddColumn(#"Inserted Year", "Day", each Date.Day([Date]), Int64.Type),
    #"Inserted Month" = Table.AddColumn(#"Inserted Day", "Month", each Date.Month([Date]), Int64.Type),
    #"Reordered Columns" = Table.ReorderColumns(#"Inserted Month",{"Date", "Day", "Month", "Year"}),
    #"Inserted Month Name" = Table.AddColumn(#"Reordered Columns", "Month Name", each Date.MonthName([Date]), type text
    #"Reordered Columns1" = Table.ReorderColumns(#"Inserted Month Name",{"Date", "Day", "Month", "Month Name", "Year"})
    #"Renamed Columns1" = Table.RenameColumns(#"Reordered Columns1",{{"Date", "SK_Date"}}),
    #"Added Custom" = Table.AddColumn(#"Renamed Columns1", "Quarter", each Date.QuarterOfYear([SK_Date])),
    #"Filtered Rows" = Table.SelectRows(#"Added Custom", each [SK_Date] < #date(2017, 9, 30))
in
    #"Filtered Rows"
```

Figure 8 - Example of the steps in creating the date dimension table

7.2.2. Transforming the fact tables:

As mentioned previously, the original flat files have all the categorical and numerical columns to re-produce the fact tables supporting the analysis of the report.

The fact table of Product Sales [FactProductSales] is generated from the file “DataCoSupplyChainDataset.csv”. It is noticed that the type of the data columns from the file is in string and with the format of “yy/dd/mm-hh/mm/ss”. Thus, firstly, it is necessary to split the columns and only keep the date values, then convert them into European Date format. Secondly, to create the relationship with the dimension tables previously created the columns that are used to create the **composite primary key** in the dimension tables are kept and concatenated to generate the same key. The next step is to merge the dimension tables and fact table together using the generated composite key, then expanded the merged queries to get the index or the primary key from the dimension table as the foreign key in the fact table. After all merged and foreign keys are created, the concatenated column will be deleted from the fact table. Finally, the numerical data and metrics that are needed for the analysis shall be kept while all the irrelevant columns shall be removed.

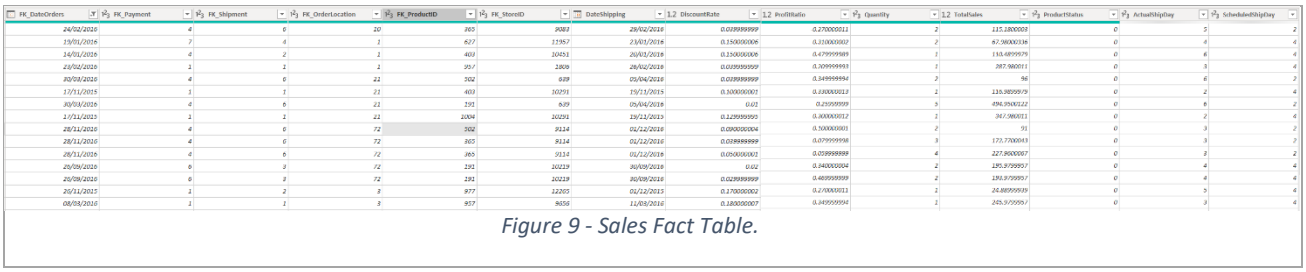

Figure 9 - Sales Fact Table.

7.3. Creating model relationships

The model will be created under star schema with 2 fact tables and 6 dimension tables. After creating all dimensions and fact table, the next step is to create the relationship amongst them using the foreign keys and surrogate keys created in the previous steps.

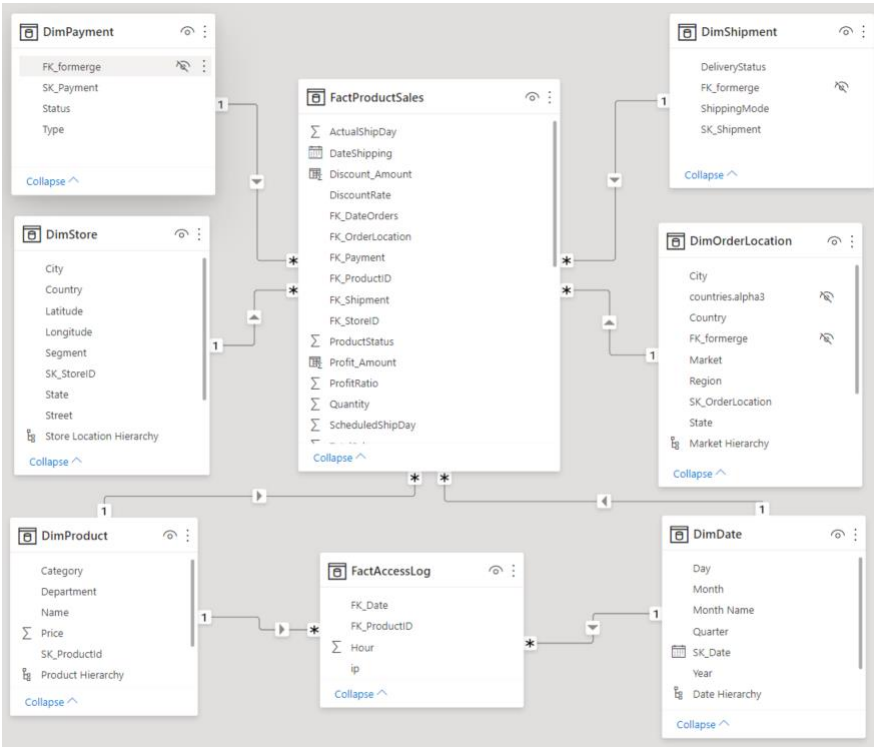

Figure 10 - The final data model

Then, finally, we need to assign the hierarchy for each dimension table that have hierarchy available. There are 4 hierarchies defined in the model across 4 dimensions of product, store location, customer and date.

Dimension table	Hierarchy
DimProduct	Product < Category < Department
DimStore	City < State < Country
DimOrderLocation	City < State < Country < Region < Market
DimDate	Date < Month < Quarter < Year

Figure 11 - Hierarchies

8) Description of Power BI Report

We have an interactive decision-making dashboard with the goal of informing corporate management about the current and future state of the firm. Our dashboard's primary purpose is to change current sales techniques to increase profitability and efficiency by enabling the decision-maker to reach judgments about how to update the business plan.

When creating the dashboard, we wanted to accomplish three main goals from the sales perspective: we wanted the dashboard to present the state of the business (profit, total sales, total quantity, etc.) at any point in time; we wanted data from an individual store point of view to be easily accessible; and finally, we also wanted to quickly understand with which products we were losing money.

From the Supply Chain point of view, we planned to have an overall understanding of each market performance (LATAM, Europe, USCA, etc.), orders dispatched and shipping modes. To achieve these objectives, we calculated several columns and key measures from the current data model, complying with the project requirements, as can be observed in Annex 2:

Let us now discuss each sheet on the sales part individually, while explaining why and how each graph was built:

8.1. Summary monthly progress page with major metrics and KPIs

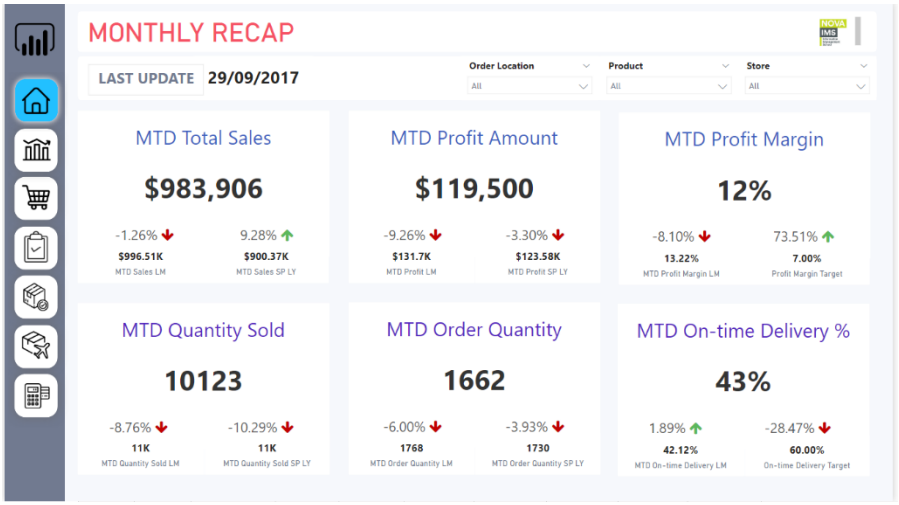

Figure 12 - Summary monthly progress page with major metrics and KPIs

We realize that the report pages are for careful investigation, would require a lot of time to accomplish. As a result, on the first page, we present a monthly summary for any shop manager or executive who wants to get a brief overview of their current performance. This page displays the month to date (MTD) value of the primary metrics and KPIs, as well as a comparison of the last month's performance to the same time last year. We assume that by using this page, you will be able to save a considerable amount of time, and this will give a slight increase to the report's efficiency.

In order to calculate the MTD value of the measures, we used the TOTALMTD function in DAX for that measure, and for the value of last month and last year, we add the filter DATEADD to the function. Moreover, for easiness in navigating between the pages, we also created a navigation pane with buttons that allow the user to navigate between the pages by clicking at the icons.

In this page, we use a new KPI visual that can be add freely from the AppSource called *Dynamic KPI Card by Sereviso*. With this visual, we will able to show not only the KPI but also 2 other measures along with the upward/ downward trend highlighted.

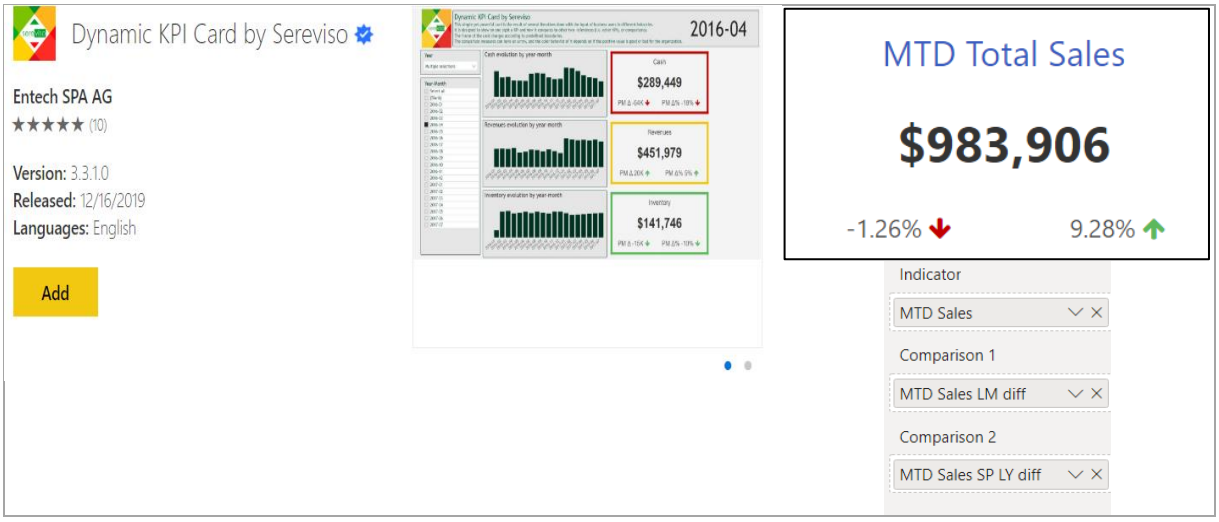

Figure 13 - Dynamic KPI Card example

8.2. Sales perspective pages:

8.2.1. Sales Summary

On top of that, we have what we considered to be the most important filters to analyze the overall state of the business at any point in time. The reader could choose which year to analyze, which quarter, or even which specific month; moreover, he could choose a specific market, region, or country in Order Location to study the specifics of a specific location; on the other hand, if the reader wanted to analyze this dashboard from a different perspective (say he wanted to see how a specific category, department or even product is performing) this would also be possible using the product filter; last but not least, we have the store perspective if one wants to examine how each store is operating (if one wants a deeper analysis from the store perspective, one should interact on the second available sheet “store” about which we will discuss next).

Continuing the “Sales” sheet, we have three KPI’s: total sales, quantity, and profit amount.

From our point of view, these are the most important measures and are convenient for quick analysis, moreover, the measures are observable on various pages.

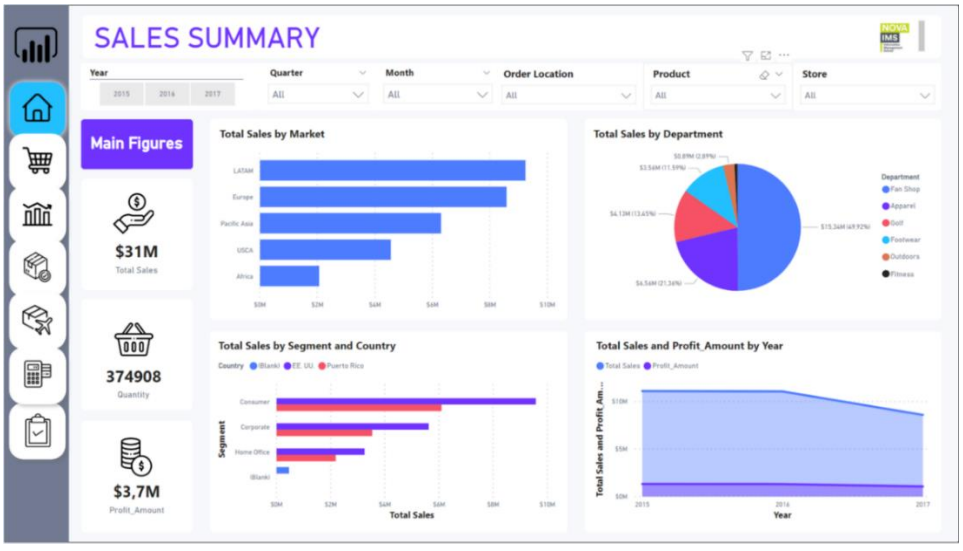

Figure 14 - Sales Summary sheet

Let us now analyze each graph individually, how it was built and why:

- On the top left, the Total Sales by Market clustered bar chart shows the value in dollars of the sales on each market (which can be drilled down by region, country, state, and city). To build this graph, we have created a DAX measure called Total Sales which results from the sum of the column TotalSales on the FactProductSales table.
- On the top right, the Total Sales by Department pie chart shows the value in dollars (and the relative percentage) of the sales by product hierarchy (department, category, and name). To build this graph, the Total Sales measure was again used.
- On the bottom left, the Total Sales by segment and country clustered bar chart shows the value in dollars that each type of client consumed and separated by the country of origin of that order. Once again, the Total Sales measure was used.
- On the bottom right, the Total Sales and Profit Amount by year area chart. It shows the values in dollars of both the total sales and the profit amount by date hierarchy (year, quarter, or month). To build this graph, the Total Sales measure was again used and a calculated column using DAX named Profit_Amount (see Annex 2 for definition) and calculates the total value in dollars of the profit.

8.2.2. Store Review

This page intends to further explore how each store is performing from the sales, quantity, profit, and comparison to the prior year's point of view.

On top of the page, we have again the same filters as on the sales summary as was explained before. Also, the main figures on the left side of the page work the same way: Total Sales use the measure Total Sales; Quantity uses the column with the same name and Profit Amount uses the calculated column with the same name described above.

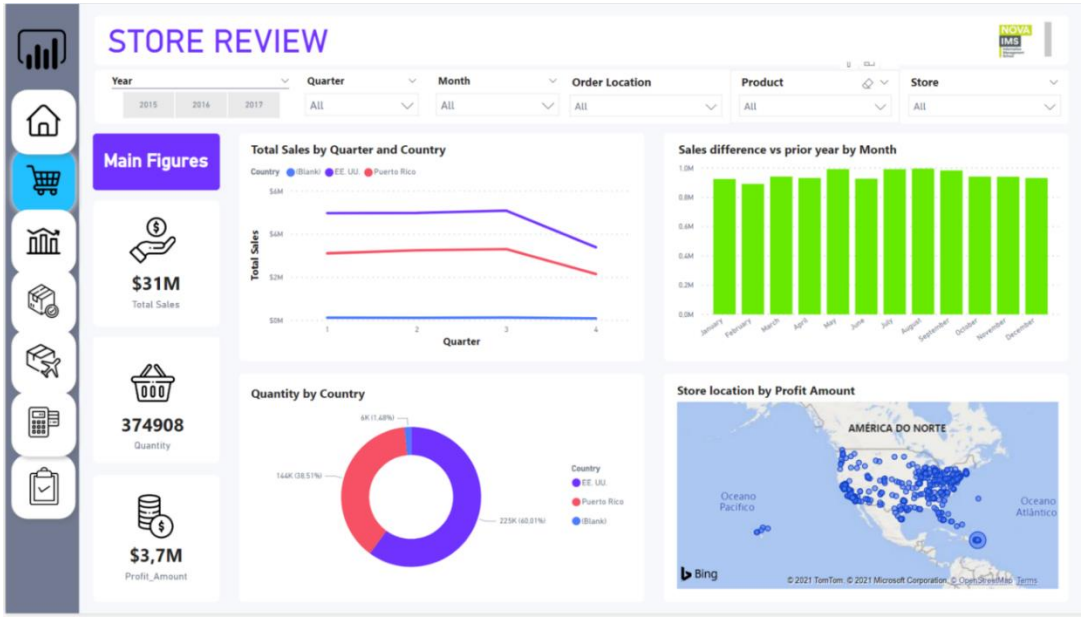

Figure 15 - Store review sheet

Let us then analyze each graph individually, how it was built and what information it shows:

- On the top left, the total sales by quarter and country line chart show how much each country with warehouses (PR and US) sold by date hierarchy (year, quarter, and month). To build this graph,

we used the Total Sales measure and the country of the store location to know the origin of each order.

- On the top right, the sales difference vs prior year by Month clustered column chart shows how the selected stores performed in comparison to the same period the year before (see image below an example for the year 2016 where you can see the comparison vs 2015). To build this graph, a DAX Time Intelligence measure was created called *DIFF.BEETWEEN_TY_AND_LY* (see Annex 2 for definition) and measures the difference in sales between years. It is also important to notice that to build the DIFF.BEETWEEN_TY_AND_LY measure, another measure called *Total Sales LY* (see Annex 2 for definition) was also built.

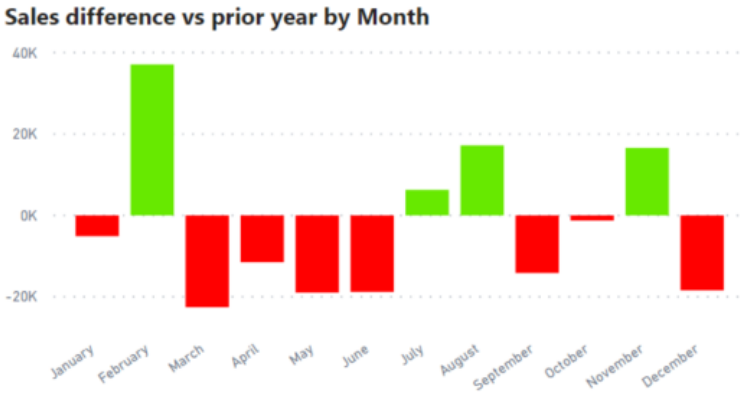

Figure 16 - Sales difference vs prior year

- On the bottom left, the quantity by country pie chart, shows how many products each location produced and uses the store location hierarchy (country, state, and city) and quantity to show such data.
- Finally, on the bottom right, the store location by profit amount shows where each of Mendeley's stores is located on the map. When you hover your mouse over a certain location a tooltip with data about that certain location (name of the city, total sales, quantity, and total profits) is shown. To build this map besides the profit amount calculated column and the city from the store dimension table, we also used the latitude and longitude information on the store dimension table to pinpoint the location of each store without incurring in the mistake of pinpointing a store in a different location but with the same city name. Finally, to enhance the information provided by the map, we also create a tooltip to add information about Total Sales, Quantity and Profit Amount by each location pointed at on the map.

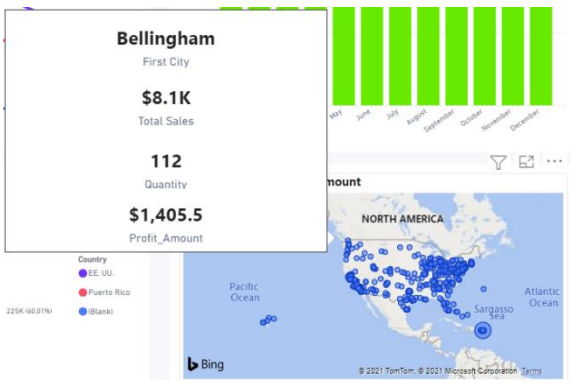

Figure 17 - Tooltip

8.2.3. Profitability review

This page was designed to solve one of the main business needs identified: early on, we discovered that there were some products with which Mendeley was losing money. This sheet intends to easily perform an analysis on product profitability.

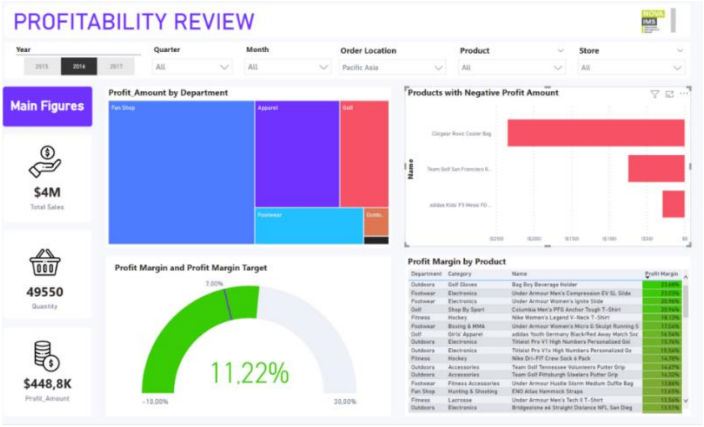

Figure 18 - Profitability review sheet

Let us then analyze each graph individually, how it was built and what information it shows:

- On the top left, the profit amout by department treemap, shows the most profitable products either by department, category, or name. To build this treemap, we used both product hierarchy from the product dimension table and the calculated column of profit amount.
- On the top right, the products with negative profit amount on the clustered bar chart, show the name and amount of loss each product has according to the filters chosen. Please be aware that this graph was built only to show products with profit_amount below zero and so, if all the products filtered are profitable, this graph will appear blank.
- On the bottom left, the profit margin and profit margin gauge show the total profit margin for all the filtered products and the target profit margin of 7%. This targeted profit margin is merely representative of this KPI and can be changed at will by the business owner. We do believe that this KPI visualization can be extremely useful for the users of this dashboard to quickly access if they are meeting the required targeted profits. To build this visual, new measures were created: *Profit Margin (see definition in Annex 2)* and *Profit Margin Target = 0.07*. To change the target profit, one simply changes this last measure.
- On the bottom right is the profit margin by product table, which shows in detail the profit margin for each filtered product. Here, we can observe the respective department and category. It is a remarkably interesting insight to understand which are the most and least profitable products in any given country, reorder market, etc.

8.3. Supply chain perspective pages

In order to calculate the number of orders, we have to use the DAX measure **GROUPBY** to group the 5 foreign keys FK_OrderDate, FK_OrderLocation, FK_Payment, FK_Shipment, FK_StoreID to create a composite key for Order_ID. The number of on-time delivery is calculated by sum the number of orders with filter of order completed and order advanced divided by total number of orders

8.3.1. Orders Review

This specific page was created to be able to provide a review of orders quantity. As it was in the Sales section here as well, we have three KPI's that are displayed for quick analysis: Quantity (by pointing at, we can also see the details of all the order status like number the of orders completed, order canceled), number

of On-time Delivery, and Average units per order. We have the possibility to compare the current quantity with the quantity of last year in the same period.

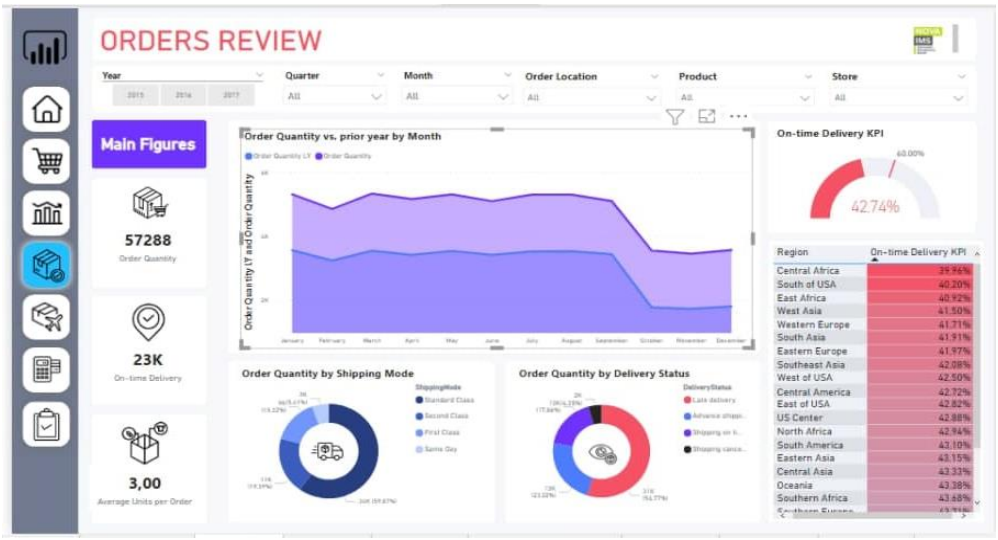

Figure 19 - Orders review sheet

Let us look at each graph individually to see what data it displays.

- On the top left, we have Order Quantity vs. Prior year by Month Area chart. It shows order quantity by month, which helps to understand and analyze the months with better and worse performances.
- On the bottom part of our dashboard page, we have two doughnut charts. The charts are for Order Quantity by Shipping Mode and Order Quantity by Delivery Status. From chart it can be concluded that Mendeley company is severely suffered from late deliveries which might significantly affect the customer satisfaction.
- On the right side the on-time delivery KPI is located. Here we have the target of 60%, we also set a rule for the color of the bar and the text, if the score is over the target, the color will change from red to green.
- On the bottom right, the One-time delivery by region table shows the on-time delivery for each filtered region in detail. On this table, it is also possible to take a further look to observe the region performance by using the tooltips.

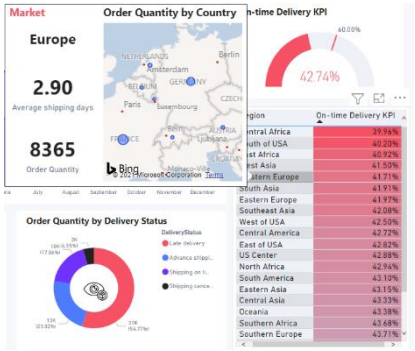

Figure 20 - Tooltip

8.3.2. Market

This page will delve deeper into how each market is performing through the years of our dataset. Here as it was in previous pages, we have the most important filters such as Years, Products, Order Location, etc. The same Key Performance Indicators are present on this page which give a good overall picture of the current market state.

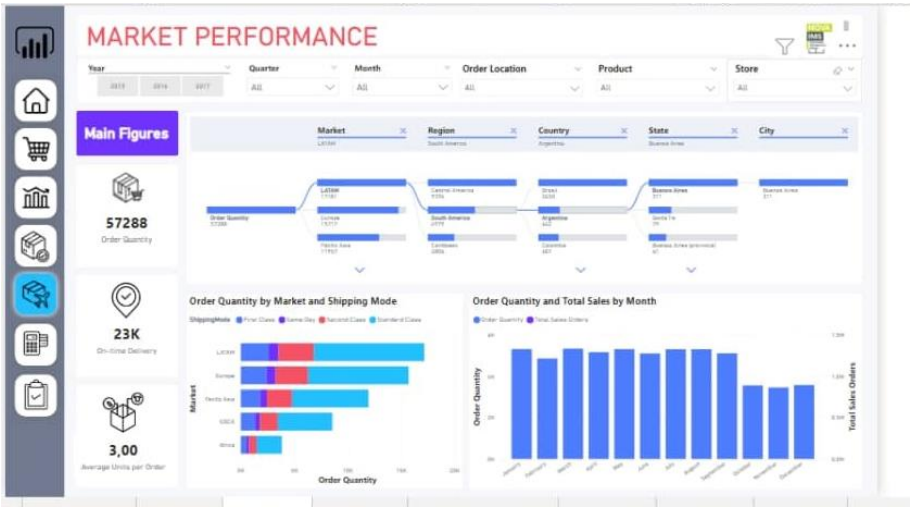

Figure 21 - Market performance sheet

To better understand Market Performance let us give a brief introduction to our graphs.

- On the top of our dashboard page, we have a decomposition tree visual, which lets visualize data across multiple dimensions. It collects data automatically and allows you to drill down into your dimensions in any sequence. The figure is showing order quantity with location hierarchy from the market up to the city. We also utilize the tooltip to visualize the location of the selected location on map along with average shipping day and number of orders.

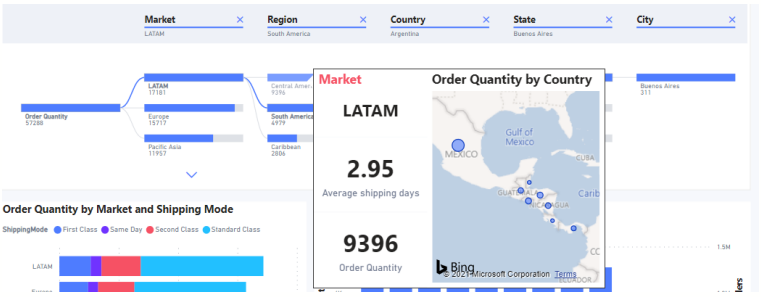

Figure 22 - Tooltip

- The left bar chart is for performing Order Quantity by Market and Shipping mode. With year filters here we can compare the Market Performance in 2015 and 2017 and notice a remarkable decrease in the Africa and US markets.
- Right side of the page is for clustered column chart that compares Order Quantity and Total sales by Month. Having this, we can analyze each month individually and understand market performance each year. To be able to build this chart we used Total Sales and Order Quantity measures and Month Name.

8.4. Advanced Analytics with Forecasting tool

This page is particularly useful to perform forecasts on the total sales, quantity, and orders. A line graph is shown for each perspective, and we have chosen to perform forecasts for the next 3 months, with a confidence level of 90% and a seasonality of 90 points (quarterly seasonality, by season). This page can be used by management to predict and manage the stocks needed.

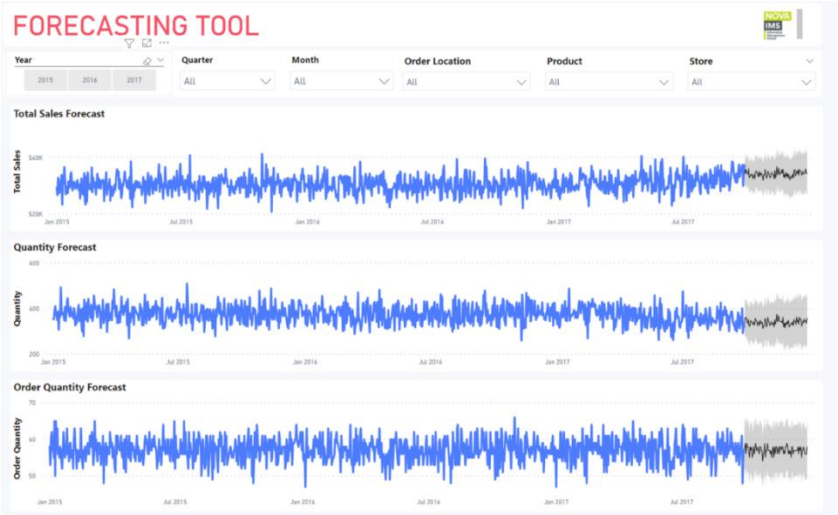

Figure 23 - Forecasting tool sheet

9) ANALYSIS AND DISCUSSION

To analyze the data, we will answer the proposed questions when we defined the two perspectives of analysis:

a. What are the key departments in terms of sales?

In any given year, Fan Shop is the department with the highest sales followed by apparel and golf. These three departments, together, represent around 85% of Mendeley's total sales. If we deep dive on the categories we can also observe that fishing, cleats, camping & hiking and cardio equipment are on the top of the most sold (representing more than 50% of the total sales), meaning that the company's sales are heavily based on outdoor sports.

Figure 24 - Sales review

b. Where is(are) the key market(s) of the company?

As it was already mentioned, Mendeley's sells its products to five different markets, being the biggest one Latin America. It is interesting to notice that the key markets change from year to year: for example, in 2015, products were sold only in Latin America, Europe and Pacific Asia; in 2016, products were sold to four markets (which could mean that the Mendeley was expanding its customer base) but Latin America (overall the biggest market) did not sell anything in 2016. There are two plausible explanations for this: either Mendeley did a weird experience in 2016 a stopped shipping products to its biggest market or our dataset has some defects. Let us discuss this further on the critical assessment.

c. Is there any seasonality in the products purchasing patterns?

Almost no seasonality was found. In February, a small decrease in the total sales is observed but it is not significant. This may be explained by the fact that the after-Christmas period is known as a sales period when most stores want to liquidate their stocks and so it is normal that most people rather do their purchases in December and January with the demand decreasing in February.

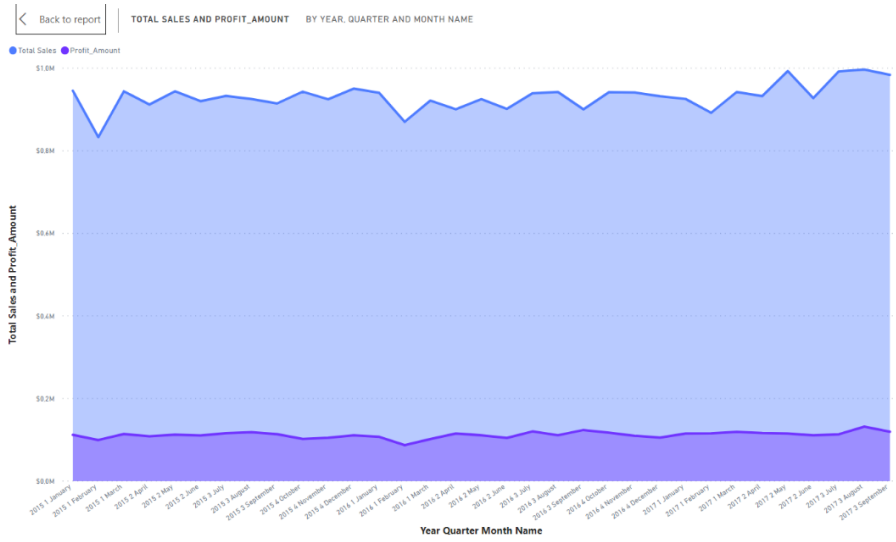

Figure 25 - Seasonality

d. KPIs on sales, quantities, and profit margins – Annually, quarterly, monthly

We did not observe any large fluctuations in sales, quantities, and profit margins. In 2015, Mendeley's had revenues of $11 million, selling 138 thousand units and profiting $1.3 million. In 2016, Mendeley's had the same revenues and profits although it sold 1000 units less. Finally, in 2017, all these figures diminished mostly because only nine months are being considered when, in the other years, we had data for the full year. In 2017, Mendeley had revenues of $9 million, 99 thousand units sold, and profits of $1.1 million.

As discussed previously regarding seasonality, no specific patterns are noticeable and apart from the decrease in sales in February, we could not highlight any specific trend in any given month. However, when looking at the data from a quarterly perspective it is interesting to notice that Q4 in each year is slightly higher than the other quarters and that in 2017 there is a clear upward trend regarding the total sales.

Figure 26 - Total profit amount

e. Which country shipped most products?

As it was already discussed in this paper, all our products are shipped from our stores located in the United States and Puerto Rico. When analyzing the products shipped, we can easily conclude that most products (both from the revenue and quantity point of view) are shipped from the United States with 60% of the products shipped from there. In the limitations chapter, we will also discuss all the orders that had the countries of origin as "blank" (1.48%) as these are real orders, that contribute to the total revenue, but the missing country of origin could be an area of improvement for the business.

f. How did each store perform compared to the prior year?

Our dashboard allows us to study how each specific store performed in comparison to the prior year. It would be impossible to study the performance of each store in detail and share that information in this report, but we can give an example: in 2016, the store in the stet of California in the city of Anaheim, performed worse in February, April, May, June, July, September, and December in comparison to the prior year. In the other months, it sold more in comparison to the prior year. The reader should however be aware of the magnitude of these differences as many times they are less than 1000 dollars per month and, as such, may not be relevant from a business perspective.

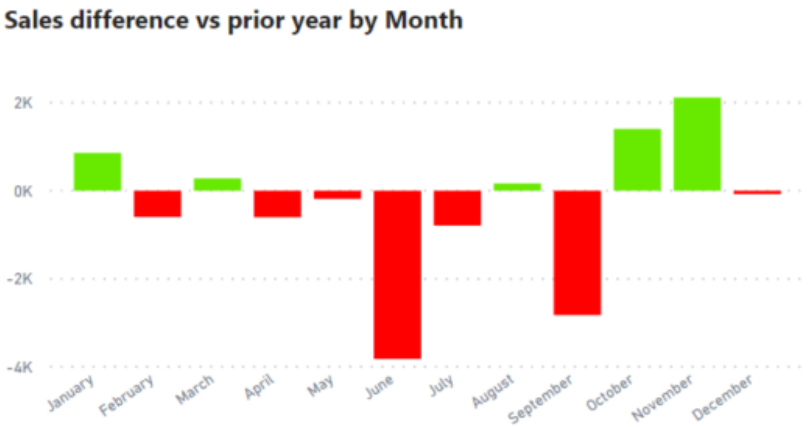

Figure 27 - Sales difference vs prior year

g. What are the most and least profitable products?

To answer this question, we have created a page on the dashboard where the reader can understand which are the most and the least profitable products (with which, sometimes, Mendeley is even losing money) from a region, country, or even store location. It is important to have both the profit margin (in percentage) and the profit amount (in dollars) so informed decisions can be made taking into consideration also the magnitude of the profit or loss. When considering the entire dataset, the most profitable item in terms of profit margin is

the Bowflex dumbbells with a profit margin of 27.30% (total sales of $2000 and a profit of $551). The least profitable product is the GoPro Hero3+ Black edition camera with a profit margin of only 2.13%.

It is important to mention that, overall, Mendeley is not losing money with any of the products sold. However, when deep diving on a country or region level, we observe that some of the products sold are creating a loss to the company and thus should be easily identifiable so actions can be taken.

For instance, if you look at Angola sales in the entire dataset you discover that Mendeley lost almost $1500 with a Field & Stream gun fire safe (-13.06% profit margin).

Overall, however, Mendeley is making a profit and so the average profit margin is 12% (11.91% in 2015, 11.87% in 2016, and 12.29% in 2017).

h. Measure on-time delivery KPI

One of the main problems that we found early on with this dataset is that many orders were arriving after the estimated date. In fact, out of the 57288 orders only around 23000 arrive on time (42.74% of the total). This is a systematic problem as almost every region has on-time delivery quantities below 50% every year. The only exceptions are in 2017 when five regions have this KPI above 50%. It is quite hard to explain the reasons behind this phenomenon due to the lack of data regarding suppliers and no clear patterns shown. For instance, the number of regions to where the products are exported fluctuates greatly from year to year (10 regions in 2015, 20 regions in 2016 and 17 regions in 2020) which makes it hard to understand which regions we could use as benchmark. Moreover, there some very strange conclusions that could be drawn but are senseless: if one looks at the overall picture, it can be concluded that to West Africa 46.68% of the products arrive on time; however, to the US Center, only 42.88% arrive on time (similar numbers are noticeable for Est US and West US). This does not make any sense as the warehouses are all located in either the US or PR and geographically closer to these regions. We believe this could be a very interesting topic to further explore but, unfortunately, cannot be due to the lack of data. It would be interesting to compare, for example, the shipping times of different shipping companies and the means of transport used, so better recommendations could be made but, unfortunately, that data is not available.

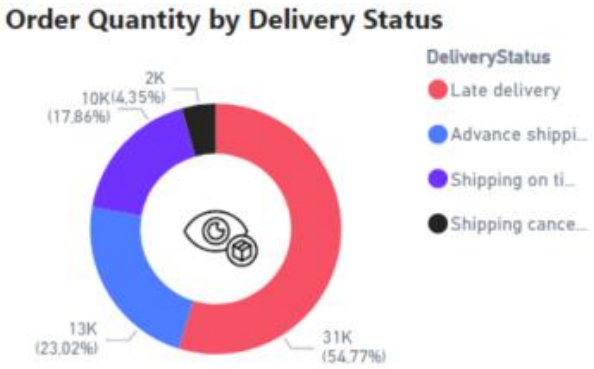

Figure 28 - Order quantity by delivery status

i. Quantify orders delivery, status

In total, Mendeley delivered 57288 orders: 20904 that15, 20859 in 2016 and 15525 in 2017. Several delivery statuses could be very useful to have a better understanding of how supply chain is working: 54.77% of the orders arrived later than it was initially scheduled; 23.02% arrived before the scheduled time and canceled % of the orders arrived on time (the rest of the orders, 4.35%, were cancelled). This is a quite worrying statistic, meaning that the predictive system for arrival time is completely flawed. Our deliveries must arrive

on time so we could meet the client expectations and if an early delivery is not that troublesome (although, in some cases, the fact that the client is not expecting the package might mean he/she will not be able to receive it), late deliveries cause dissatisfied customers who are less likely to come back.

j. Average shipment time to each region in the world

To understand the average shipment time to each region in the world, we have created a tooltip on the on-time delivery KPI table. Once again, the data does not make sense sometimes and it even is a bit randomized: to Canada, the average shipping time is 2.94 days; to Southern Europe, however, is just 2.91 days! Although this difference might seem small, it can make a significant difference when trying to predict the shipment time to a certain region of the globe. Moreover, it is quite unusual that Canada (which is very close to the location of our warehouses) has a longer average shipment time than a region on the other side of the Atlantic Ocean.

k. What is the preferred shipment mode?

There are four different modes that Mendeley uses to ship products: standard class (59.87% of the orders), second class (19.39%), first class (15.32%) and same day (5.41%). We did not have any description on the definition of each of the modes but there are some conclusions we can derive from analyzing the dashboard: when using the standard class, a whopping 60.14% of the deliveries arrive on time, with the lowest on-time delivery KPI being higher than 55%!

However, when analyzing second class shipments only 20.06% arrive on time. One of our recommendations would be to stop immediately this mode or, at least, to change the suppliers as soon as possible as they are doing a terrible service to Mendeley. It would be also interesting to understand what the price difference for both the client and Mendeley is of a standard versus a second-class shipping mode, but there was no available data.

Unfortunately, when analyzing both the first class and same-day orders, we faced data inconsistencies. Regarding the first-class mode, no information on the delivery times is available on the dataset so it is impossible to know how our KPI would improve/worsen this mode. Regarding the same-day deliveries, we observe that only 52.24% arrive on time. Is there a chance that Mendeley's is labeling same-day delivery orders that arrive late? That is possible but extremely unlikely, or, at least, it shows how inconsistent Mendeley's current delivery predictions are.

l. Forecast

It is mentioned that we have created a page, to forecast sales, quantities ordered and orders. It can predict, how the customers will behave, based on time and past behavior, and help Mendeley to manage its stocks, reduce costs, delivery times and fulfill most of the clients' orders. No peaks or bottoms in demand are observed on the next 3 months of our forecast in none of the studied metrics and so Mendeley should be able to keep the same level of sales and profits in the short term.

10) CRITICAL ASSESSMENT

When first choosing our dataset, we were mainly concerned with checking all the requirements asked from the teaching team to perform in this project. Those minimum requirements were fulfilled with this dataset but, unfortunately, as we started working with it, we realized that this dataset was inconsistent. Besides that, we have managed to complete the requested task and report successfully as not only the dataset checks all the required requirements but also because we were able to build a dashboard that has all the mandatory hierarchies, calculated columns, DAX measures, KPI visuals, advanced visuals and storytelling techniques. Moreover, we feel that our dashboard is not only very insightful but is also well designed in terms of structure and user-friendliness.

Regarding the data source, the data used in this project is a flat file containing all the data from all operational processes of the business, which is not a common thing in real situation. We acknowledge that data source in real world business might come from a lot of different sources, data marts or data silos from different department of the company, which will require much more effort in the ETL process. However, we believe that the knowledge and experience gained from the process of migration and transformation data both in this project and in practical classes has equipped us a strong foundation to overcome any future challenge in the ETL process of building a data warehouse.

Overall, we have done a good job but must highlight some of the aspects that should be taken into consideration in future works or with other groups using the same dataset:

- LATAM (Latin America market) is our biggest market overall with almost a third of the sales, $9 million. However, in 2016, and for no apparent reason, LATAM sold 0 dollars. Not a single sale was registered in 2016. We believe that this should be a mistake (why would you not sell to your biggest market for an entire year?) and should be addressed by Mendeley.
- Unfortunately, no seasonality is observed. It seems strange how linear the entire dataset without any big variations in terms of sales, quantity, and orders (as observed on the Forecast page).
- There are about 6000 products without a known origin. We believe that this might be an error in the data input and should also be addressed.
- The small variations in terms of seasonality and on a store level (year on year variations, for instance) are very small and it is quite hard to make any kind of recommendations or drive any conclusions.
- The first-class delivery option did not have provided information on delivery time, so it was impossible to understand how our KPI would perform.

Due to the time limit of the project, we only deliver the metrics on a monthly basis. However, in real world situation, especially supply chain management, it is very crucial to monitor the operational performance on a weekly basis in order to carry out in-time decision whenever any abnormal happen. Thus, for future work, the report should also develop the date hierarchy up to week for some metrics in the supply chain management perspective.

Advanced services such as Q&A also can be considered to add to the report to improve the efficiency in using the dashboard and looking for information from the data that have not been showed on the report.

11) CONCLUSION

Based on what has been covered so far, the company's management must grasp the current global order to design a successful plan. Due to the consistency of total sales, the company needs to change and update its operating model to remain competitive in the market. Although it is true that the company is profitable and has been profitable every year, it is somewhat concerning that both sales and profits remain somewhat constant, and that no growth is observed.

We believe that if Mendeley focuses its strategy on both sales and supply chain, it will see its profit margins grow. First, on the sales side, Mendeley should reduce its portfolio of products sold to certain countries/regions as they proved to be unprofitable. Also, it is odd that for long periods of time sometimes (like it happened with Latin America in 2016) entire regions are closed and no sales are registered. We believe this might be due to data inconsistencies and should be investigated by the company. Also on the

sales side, all the products that have no stores of origin data should be reviewed as this could give insightful information on the associated costs and profit margin of said product.

On the supply chain side, Mendeley should review its entire distribution strategy as orders arrive consistently later than originally planned. It is hard to understand why these constant delays happen, but we have concluded that, for instance, second class shipping is very damaging for the company's image as only 20% of the orders arrive on time and should be terminated immediately. It is also peculiar that sometimes orders sent farther away on average arrive earlier, than to geographically closer places. This could be due data inconsistencies or a poorly designed supply system.

Although we wouldn't advise the use of the same dataset to future groups, we have delivered (and learned) a great amount from it. As this report shows, we have built a dashboard that, could prove very insightful to the sales managers and executives to drive quick insights on the data. We believe we have used all (and more) of the required tools and that we were able to make a streamlined and comprehensible dashboard and consequent report from somewhat odd and complicated dataset.

12) REFERENCES

[1] Howen, A. (2014). *The Impact of Late and Inaccurate Deliveries on Customer Loyalty*. Website Magazine. https://www.websitemagazine.com/blog/the-impact-of-late-and-inaccurate-deliveries-on-customer-loyalty

[2] *Impact of Late or Inaccurate Deliveries Can Be Disastrous, Study Shows*. (2012). Supply Chain Brain. https://www.supplychainbrain.com/articles/14912-impact-of-late-or-inaccurate-deliveries-can-be-disastrous-study-shows

[3] *The Value of Keeping the Right Customers*. (2014, November 5). Harvard Business Review. https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

[4] Constante, F. (2019, March 12). DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS. Mendeley. https://data.mendeley.com/datasets/8gx2fvg2k6/

[5] Mckinsey Global Institute. (2016, December). THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20and%20Social%20Sector/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.pdf

13) ANNEX

Annex 1

FIELD	DESCRIPTION
Type	Type of transaction made (which can be DEBIT, CASH, TRANSFER, etc.)
Days for shipping (real)	Actual shipping days of the purchased product
Days for shipment (scheduled)	Days of scheduled delivery of the purchased
Benefit per order	Earnings per order placed
Sales per customer	Total sales per customer made per customer
Delivery status	Delivery status of orders (Advanced shipping, Late delivery, Shipping on time, etc)
Late_delivery_risk	Categorical variable that indicates if send...
Category Id	Product category code
Category Name	Description of the product category
Customer City	City where the customer made the purchase
Customer Country	Country where the customer made the purchase
Customer Email	Customer's Email

Customer Fname	Customer name
Customer Id	Customer Id
Customer Lname	Customer Last name
Customer Password	Customer Password
Customer Segment	Types of Customers (Consumer , Corporate, Home office)
Customer State	State to which the store where the purchase has been made
Customer Street	Street to which the store where the purchas has been made
Customer Zipcode	Zipcode of the customer
Department Id	Department code of store
Department Name	Department name of store
Market	Market to where the order is delivered
Order City	Destination city of the order
Order Country	Destination country of the order
Order Customer Id	Customer order code
order date (DateOrders)	Date on which the order is made
Order Id	Order code
Order Item Cardprod Id	Product code generated through the RFID reader
Order Item Discount	Order item discount value
Order Item Id	Order item code
Order Item Quantity	Number of products per order
Sales	Value in sales
Order Item Total	Total amount per order
Order Profit Per Order	Order Profit Per Order
Order Region	Region of the world where the order is delivered
Order State	State of the region where the order is delivered
Order Status	Order Status : COMPLETE , PENDING , CLOSED)
Product Card Id	Product code
Product Description	Product Description
Product Name	Product Name
Product Price	Product Price
Product Status	Product Status
Shipping date (DateOrders)	Exact date and time of shipment
Shiping Mode	The following shipping modes are presented (First lass, Second Class, Standard Class

Annex 2

Columns	Formula
Discount_Amount	(FactProductSales[TotalSales]*FactProductSales[DiscountRate])/(1-FactProductSales[DiscountRate])
Profit_Amount	(FactProductSales[TotalSales]*FactProductSales[ProfitRatio])

Measures	Formula
SALES PERSPECTIVE	
Total Sales	SUM(FactProductSales[TotalSales])
Total Sales LY	CALCULATE([Total Sales], SAMEPERIODLASTYEAR(DimDate[SK_Date]))
Profit Margin	DIVIDE(SUM(FactProductSales[Profit_Amount]),[Total Sales],0)
Profit Amount Measures	SUMX(FactProductSales,FactProductSales[ProfitRatio]*FactProductSales[TotalSales])
DIFF.BEETWEEN_TY_AND _LY	IF(ISBLANK([Total Sales LY]),BLANK(), [Total Sales]-[Total Sales LY])

SUPPLY CHAIN PERSPECTIVE	
Order Quantity	COUNTROWS(GROUPBY(FactProductSales, FactProductSales[FK_DateOrders], FactProductSales[FK_OrderLocation], FactProductSales[FK_Payment],FactProductSales[FK_Shipment],FactProductSales[FK_StoreID]))
Order Quantity LY	CALCULATE([Order Quantity],DATEADD(DimDate[SK_Date],-1,YEAR))
Units per Order	SUMX(GROUPBY(FactProductSales,FactProductSales[FK_DateOrders],FactProductSales[FK_OrderLocation],FactProductSales[FK_Payment],FactProductSales[FK_Shipment],FactProductSales[FK_StoreID],"Total Units", COUNTX(CURRENTGROUP(), FactProductSales[TotalSales])), [Total Units])
Average Units per Order	DIVIDE([Units per Order],[Order Quantity])
Average shipping days	AVERAGEX(GROUPBY(FactProductSales,FactProductSales[FK_DateOrders], FactProductSales[FK_OrderLocation],FactProductSales[FK_Payment],FactProductSales[FK_Shipment],FactProductSales[FK_StoreID]), AVERAGE(FactProductSales[ScheduledShipDay]))
On-time Delivery	CALCULATE([Order Quantity],FILTER(FactProductSales,RELATED(DimShipment[DeliveryStatus]) <> "Late delivery"), FILTER(FactProductSales,RELATED(DimShipment[DeliveryStatus]) <> "Shipping canceled"))
Delivered Delivery	CALCULATE([Order Quantity],FILTER(FactProductSales,RELATED(DimShipment[DeliveryStatus]) <> "Shipping canceled"))
On-time Delivery KPI	DIVIDE([On-time Delivery],[Delivered Delivery],0)
MONTHLY RECAP	
Current Date	MAX(FactProductSales[FK_DateOrders])
MTD Sales	TOTALMTD([Total Sales],DimDate[SK_Date])
MTD Sales LM	TOTALMTD([Total Sales],DATEADD(DimDate[SK_Date],-1,MONTH))
MTD Sales LM diff	([MTD Sales]-[MTD Sales LM])/[MTD Sales LM]
MTD Sales SP LY	TOTALMTD([Total Sales],DATEADD(DimDate[SK_Date],-1,YEAR))
MTD Sales SP LY diff	([MTD Sales]-[MTD Sales SP LY])/[MTD Sales SP LY]
… more MTD measures	