ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
SCHOOL OF BUSINESS

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
DEPARTMENT OF MANAGEMENT SCIENCE & TECHNOLOGY

Data Management and Business Intelligence

ASSIGNMENT 2

STUDENTS : Charilaos Petrakogiannis ,Ilias Dimos

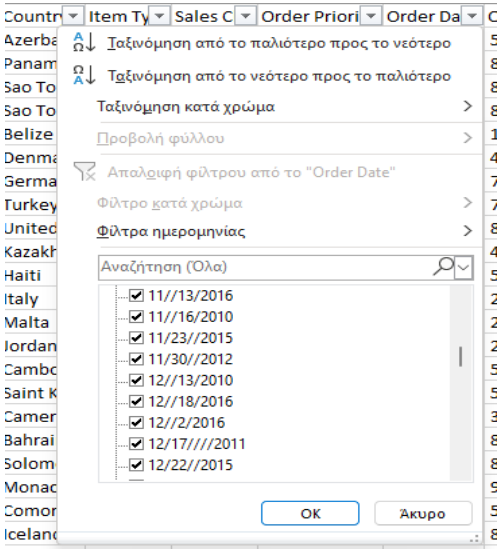
PROFESSORS : Damianos Chatziantoniou , Spiros Safras

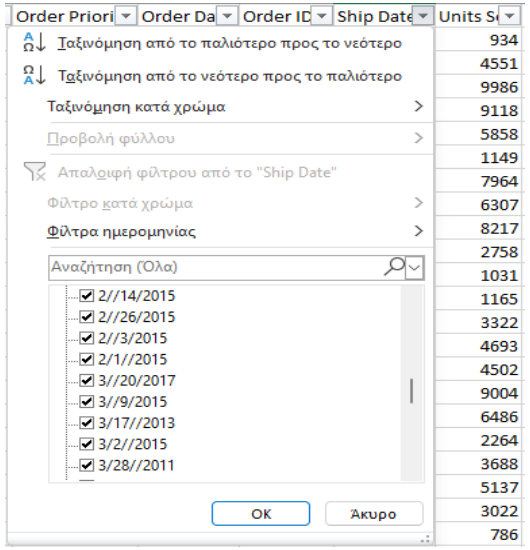
DATA REFERENCE & ETL

- Sales of 12 different item types Globally
- Time Period : 2010-2017
- Total Amount of Observations : 100.000.
- Source : www.Kaggle.com

DATA REFERENCE & ETL

❖ Issues :





❖ Solution :

```
getwd()

dataset<- read.csv("C:\\Users\\elgr9\\OneDrive\\Desktop\\Sales.csv")

View(dataset)
str(dataset)


dataset$Order.Date<-gsub("//","/",dataset$Order.Date) # where the script finds "//" it replaces them with "/"

dataset$Ship.Date <-gsub("//","/",dataset$Ship.Date)

nrow(dataset) # In order to see if any line has been deleted


write.csv(dataset,'Sales Records.csv') # We save the new file as Sales Records to continue with the ETL process
```

DATA REFERENCE & ETL

❖ Checking for NAs and NULL values with R programming:

```
for(i in nrow(Sales_Records))
{
  cat("Checking for NA's",sep="\n")
  cat("There are",sum(is.na(i)),"NA's",sep =" ")

  cat("\n")

  cat("Checking for NULL's",sep="\n")
  cat("There are",sum(is.null(i)),"NULL's",sep =" ")

}
```

❖ Output :

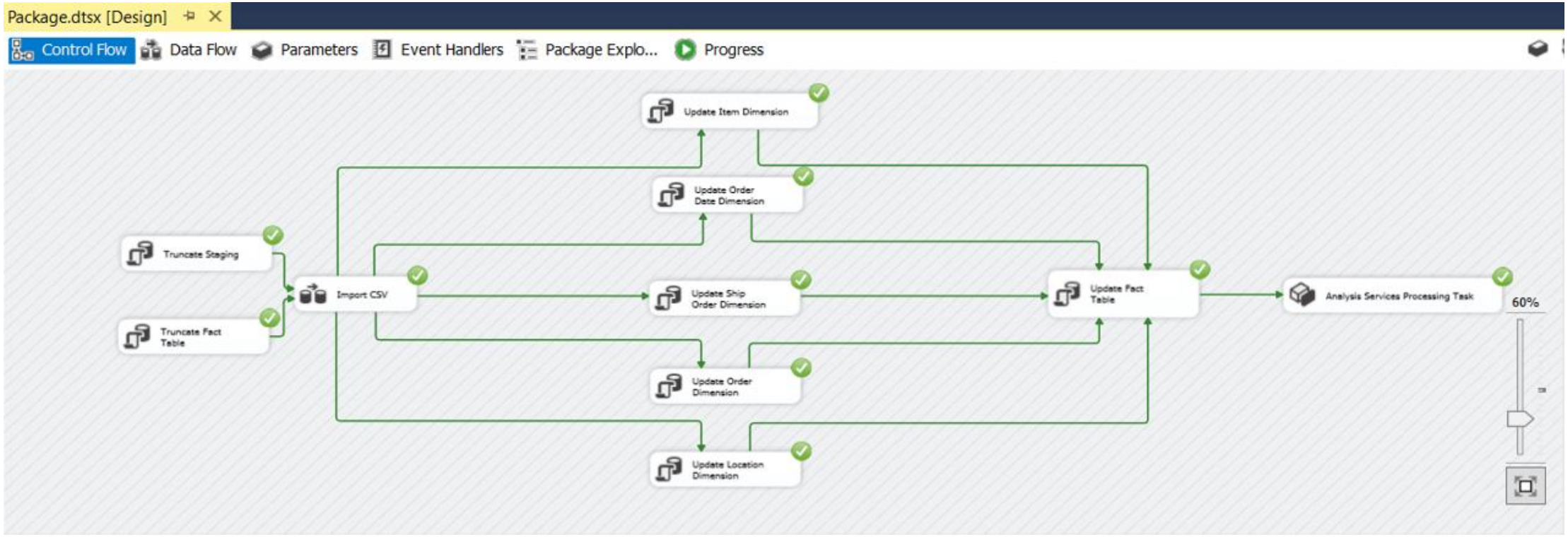
```
Checking for NA's
There are 0 NA's
Checking for NULL's
There are 0 NULL's
```

DATA REFERENCE & ETL

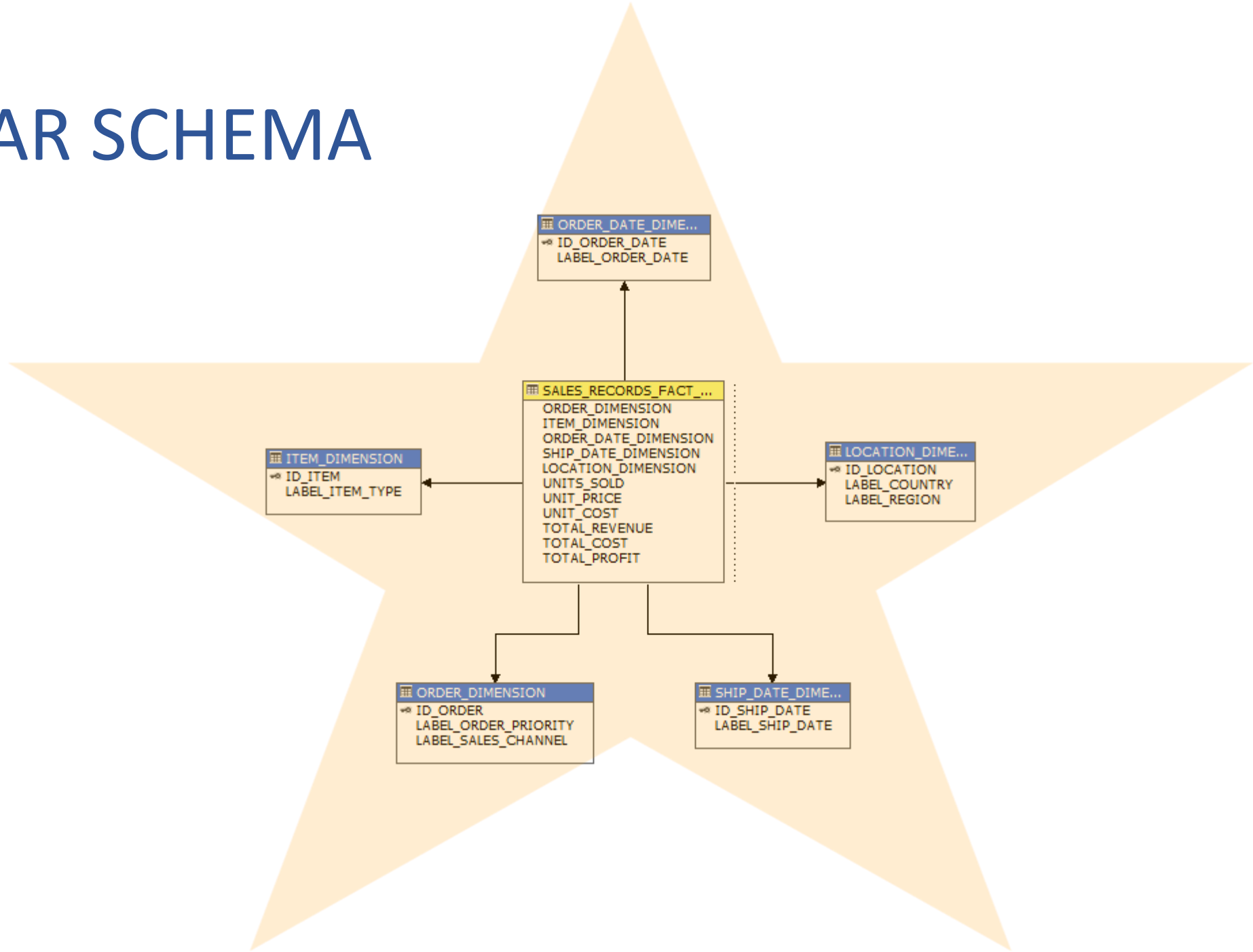
❖ Final Format :

```
> Sales_Records <- read.csv("C:\\Users\\elgr9\\OneDrive\\Desktop\\Sales Records.csv")
> str(Sales_Records)
'data.frame':   100000 obs. of  14 variables:
 $ Region        : chr  "Middle East and North Africa" "Central America and the Caribbean" "Sub-Saharan Africa" "Sub-Saharan Africa" ...
 $ Country       : chr  "Azerbaijan" "Panama" "Sao Tome and Principe" "Sao Tome and Principe" ...
 $ Item.Type     : chr  "Snacks" "Cosmetics" "Fruits" "Personal Care" ...
 $ Sales.Channel : chr  "Online" "Offline" "Offline" "Online" ...
 $ Order.Priority: chr  "C" "L" "M" "M" ...
 $ Order.Date    : chr  "10/8/2014" "2/22/2015" "12/9/2015" "9/17/2014" ...
 $ Order.ID      : int  535113847 874708545 854349935 892836844 129280602 473105037 754046475 772153747 847788178 471623599 ...
 $ Ship.Date     : chr  "10/23/2014" "2/27/2015" "1/18/2016" "10/12/2014" ...
 $ Units.Sold    : int  934 4551 9986 9118 5858 1149 7964 6307 8217 2758 ...
 $ Unit.Price    : num  152.58 437.2 9.33 81.73 668.27 ...
 $ Unit.Cost     : num  97.44 263.33 6.92 56.67 502.54 ...
 $ Total.Revenue : num  142510 1989697 93169 745214 3914726 ...
 $ Total.Cost    : num  91009 1198415 69103 516717 2943879 ...
 $ Total.Profit  : num  51501 791282 24066 228497 970846 ...
```

DATA REFERENCE & ETL



STAR SCHEMA

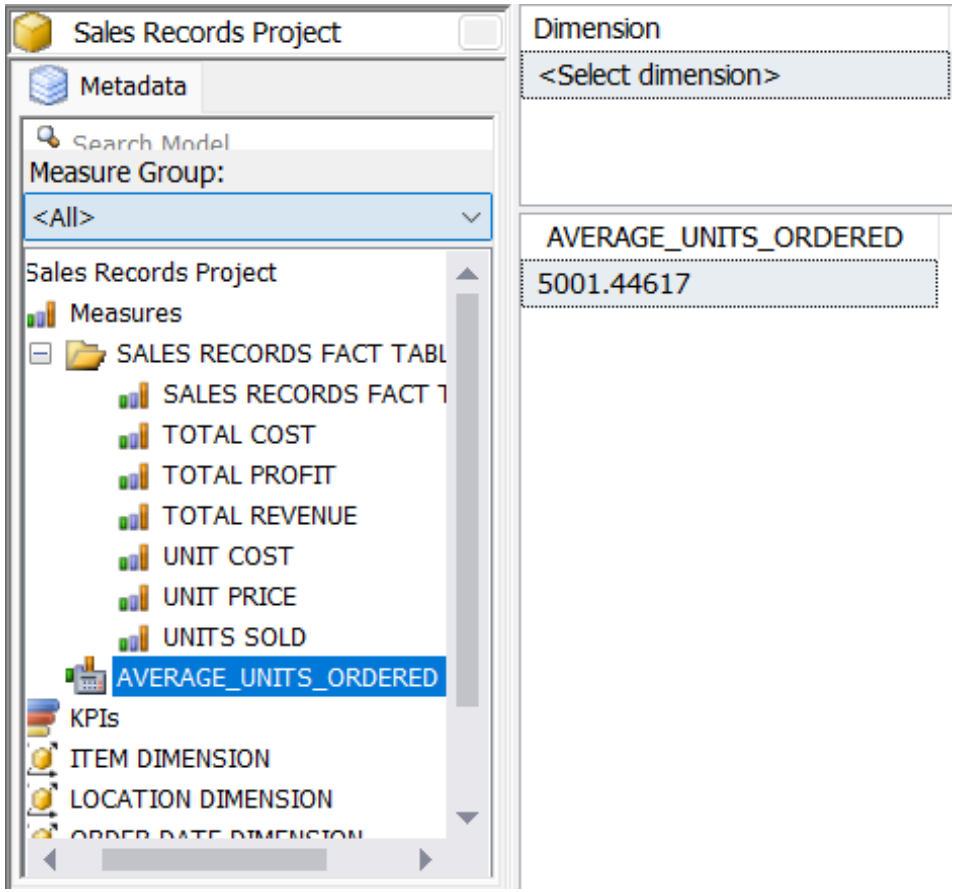

ORDER_DATE_DIME...
ID_ORDER_DATE
LABEL_ORDER_DATE
SALES_RECORDS_FACT_...
ORDER_DIMENSION
ITEM_DIMENSION
ORDER_DATE_DIMENSION
SHIP_DATE_DIMENSION
LOCATION_DIMENSION
UNITS_SOLD
UNIT_PRICE
UNIT_COST
TOTAL_REVENUE
TOTAL_COST
TOTAL_PROFIT
ITEM_DIMENSION
ID_ITEM
LABEL_ITEM_TYPE
LOCATION_DIME...
ID_LOCATION
LABEL_COUNTRY
LABEL_REGION
ORDER_DIMENSION
ID_ORDER
LABEL_ORDER_PRIORITY
LABEL_SALES_CHANNEL
SHIP_DATE_DIME...
ID_SHIP_DATE
LABEL_SHIP_DATE

CUBE CREATION

SalesRecordsProject.cube [Design]

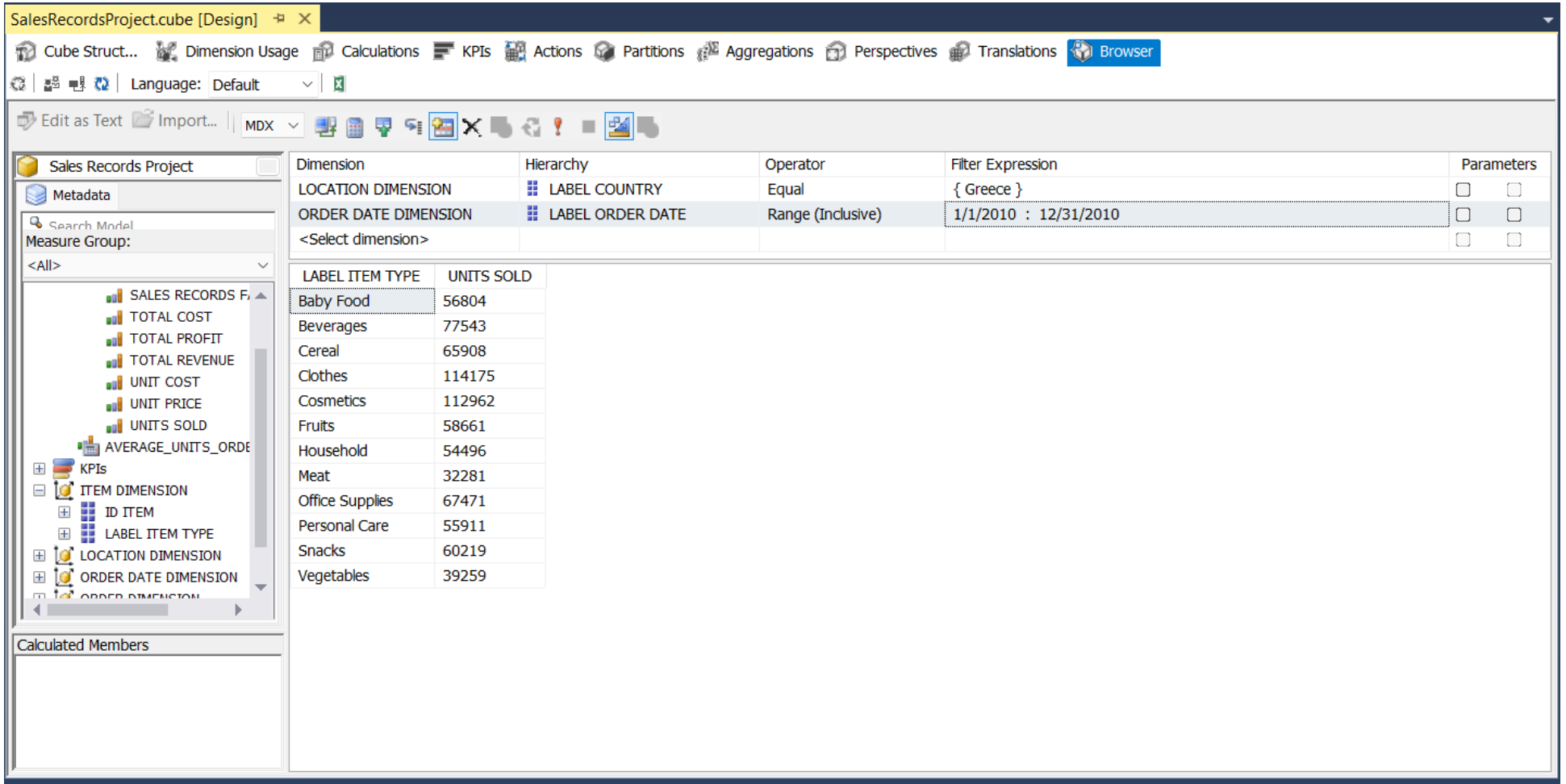
Cube Struct... | Dimension Usage | Calculations | KPIs

Cube Objects	Object Type
Sales Records Project	Name
	DefaultMeasure
− Measure Groups	
− SALES RECORDS FACT TABLE	MeasureGroup
UNITS SOLD	Measure
UNIT PRICE	Measure
UNIT COST	Measure
TOTAL REVENUE	Measure
TOTAL COST	Measure
TOTAL PROFIT	Measure
SALES RECORDS FACT TABLE Count	Measure
− Dimensions	
+ ORDER DIMENSION	CubeDimension
+ ITEM DIMENSION	CubeDimension
+ LOCATION DIMENSION	CubeDimension
+ ORDER DATE DIMENSION	CubeDimension
+ SHIP DATE DIMENSION	CubeDimension
− Calculations	
AVERAGE_UNITS_ORDERED	CalculatedMember



EXAMPLE OF OLAP QUERY

❖ Slice & Dice :



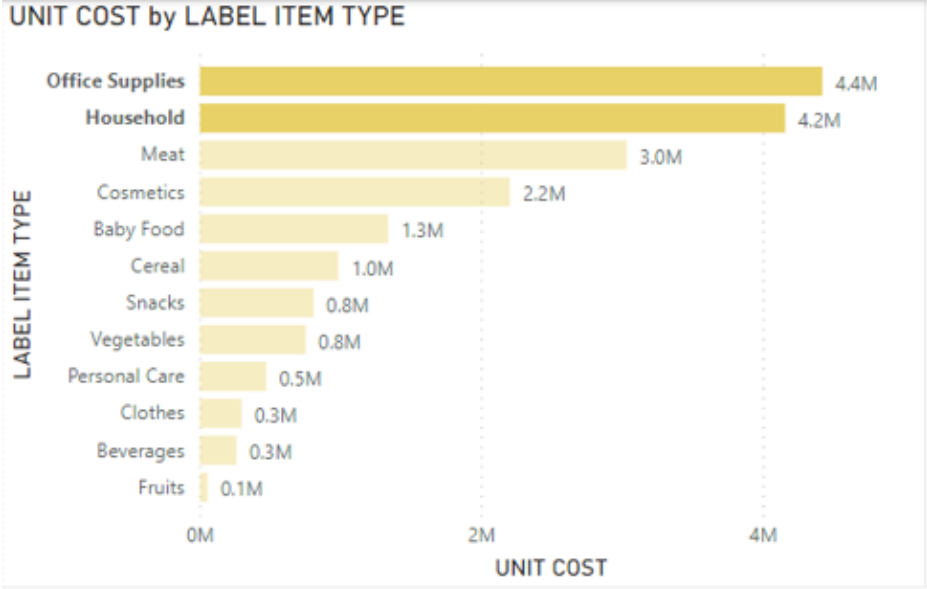
And yet… we are not done

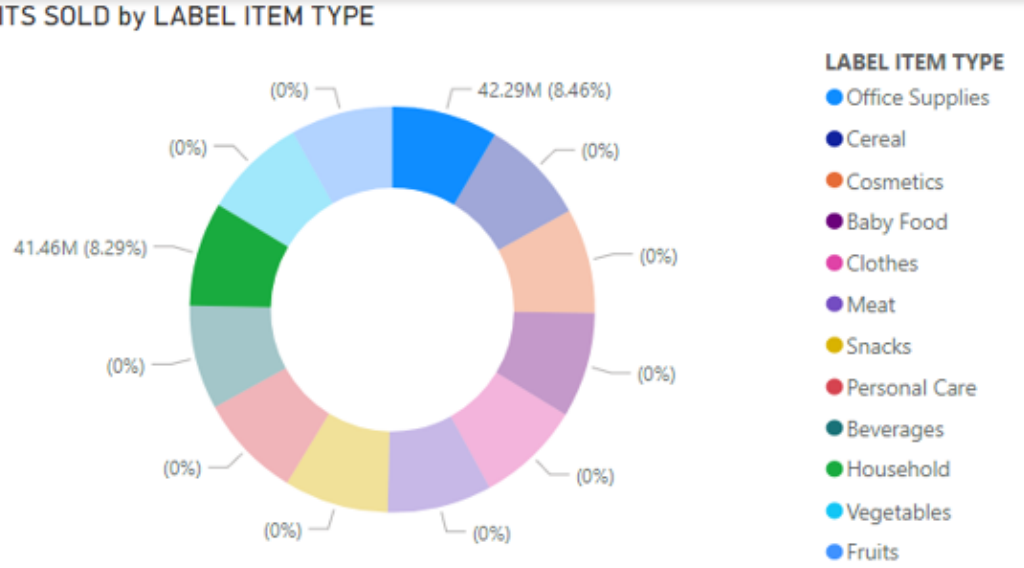
POWER BI

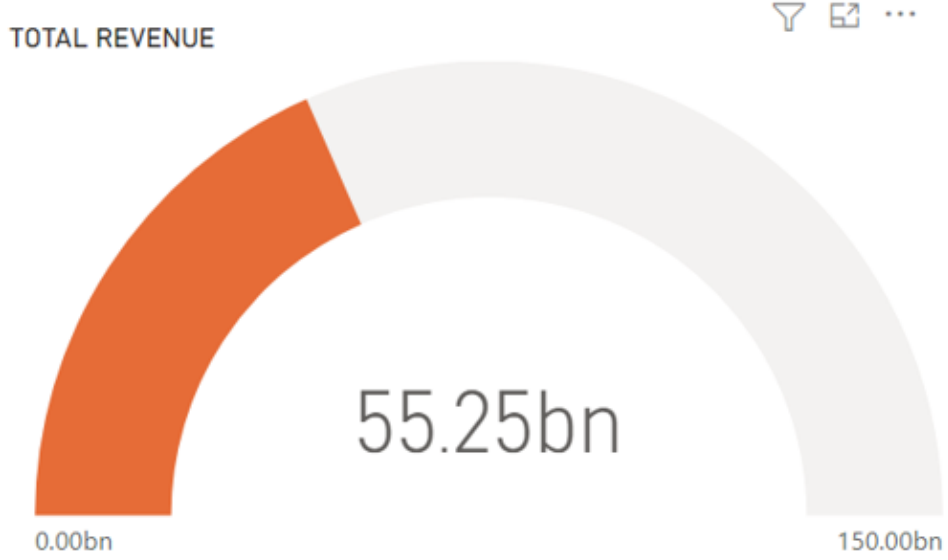
BUSINESS CASE SCENARIO

- Year of analysis : 2017.
- Customer : Multinational company that wants to invest in two specific product markets (Household items, Office supplies).
- In which regions should they focus on? In which countries?
- Should they focus on offline sales channels, on online or both?
- Is every time of the year the same in terms of load and if not what should they expect?

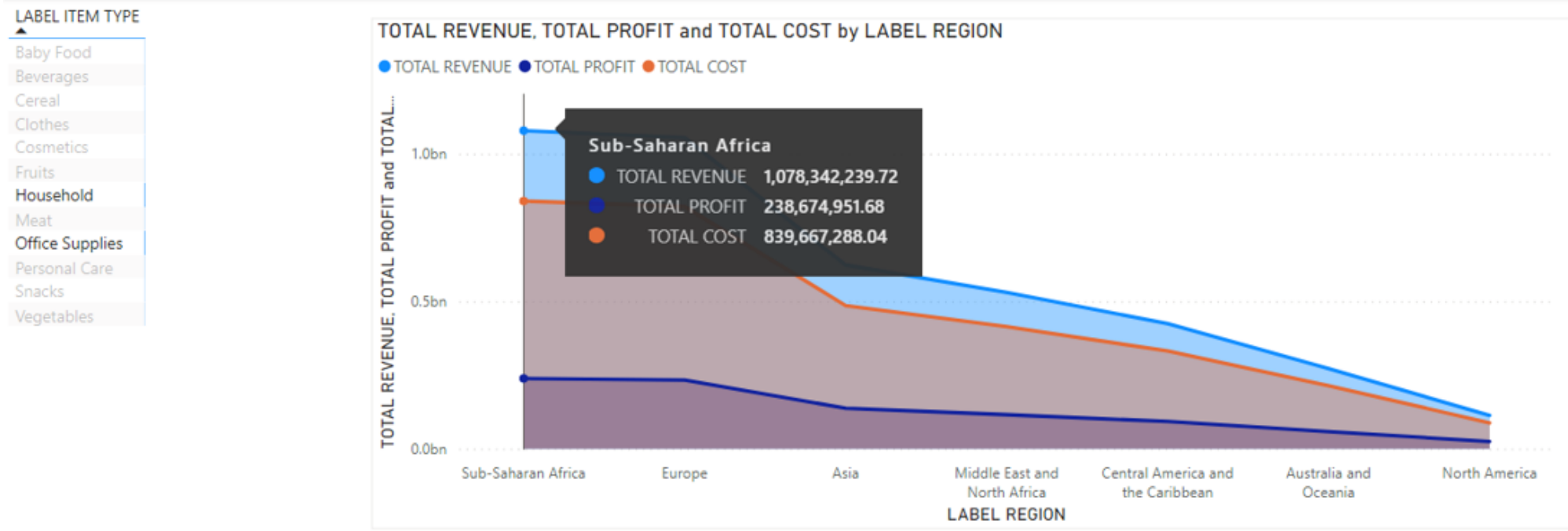
Information about Office Supplies and Household Items



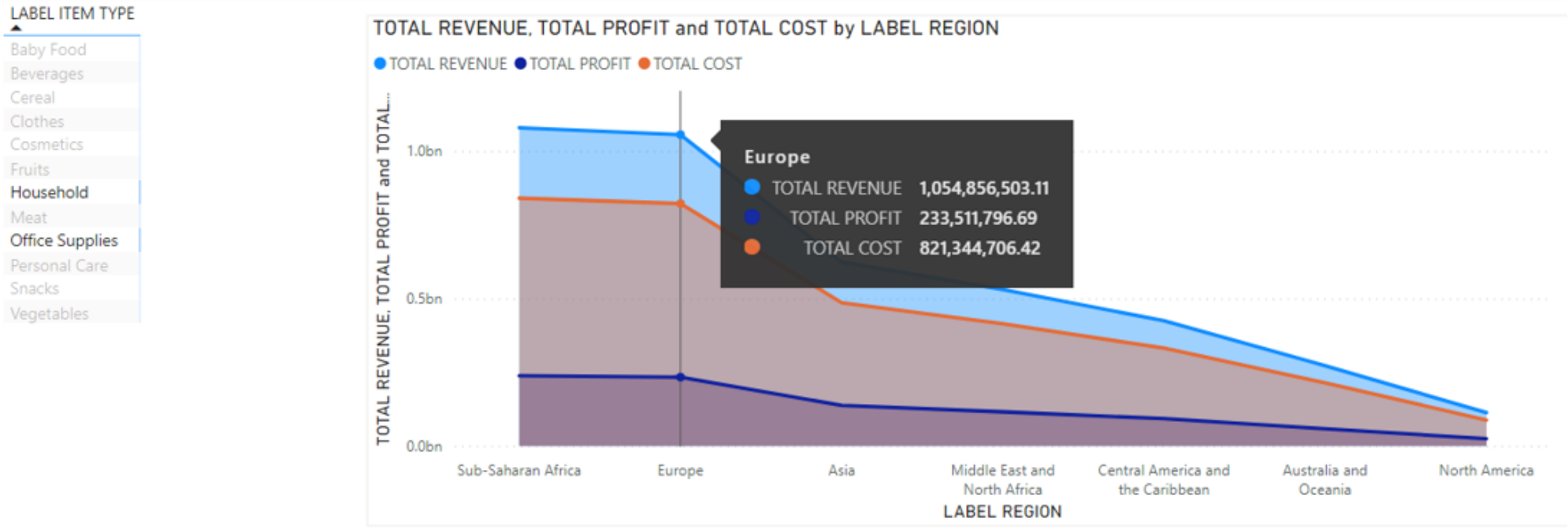


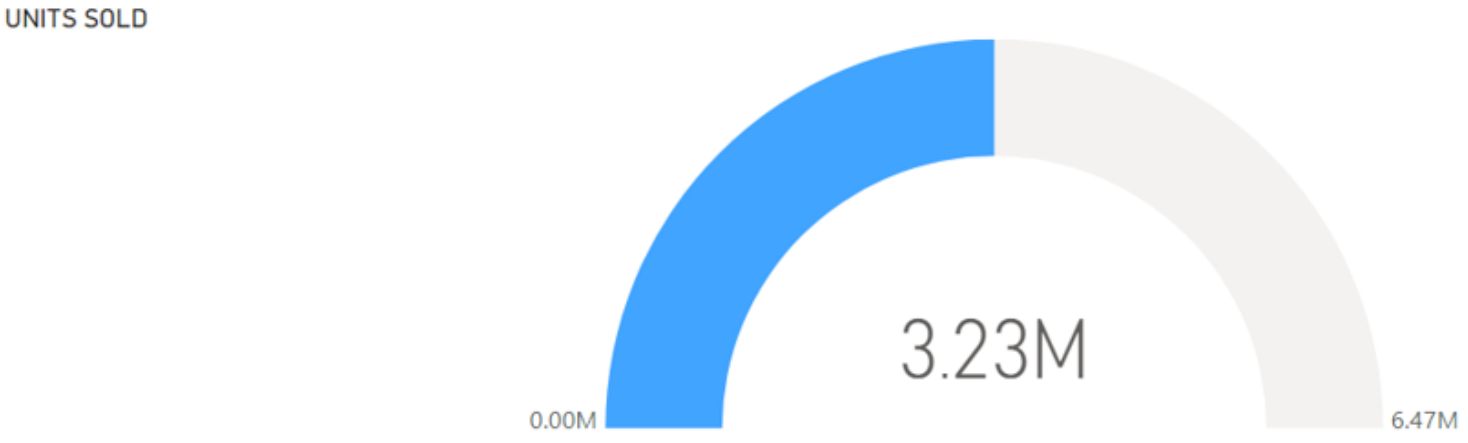


Information about the Best-Sellers Regions worldwide (Sub-Saharan Africa) in 2017

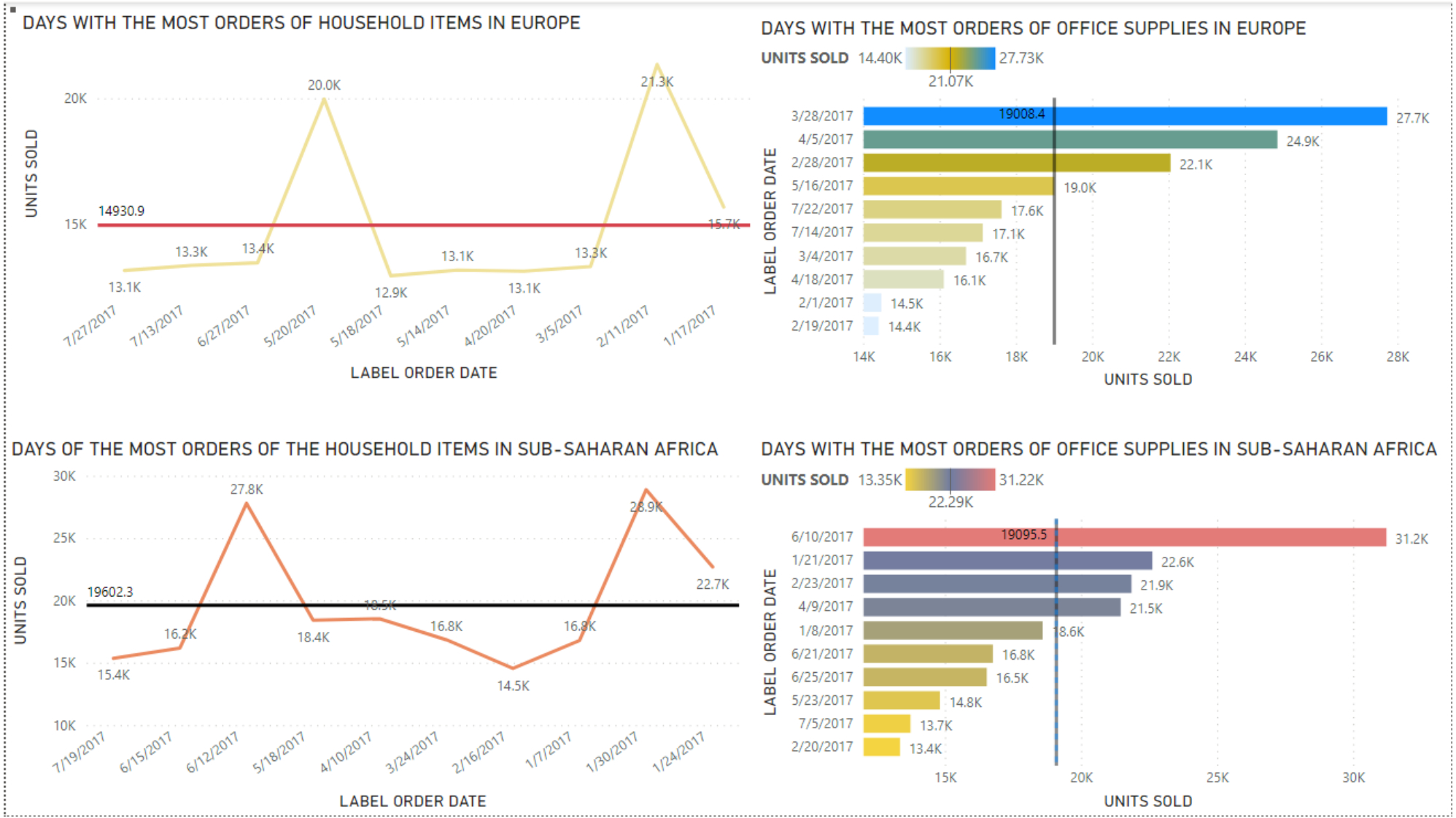


Information about the Best-Sellers Regions worldwide (Europe) in 2017

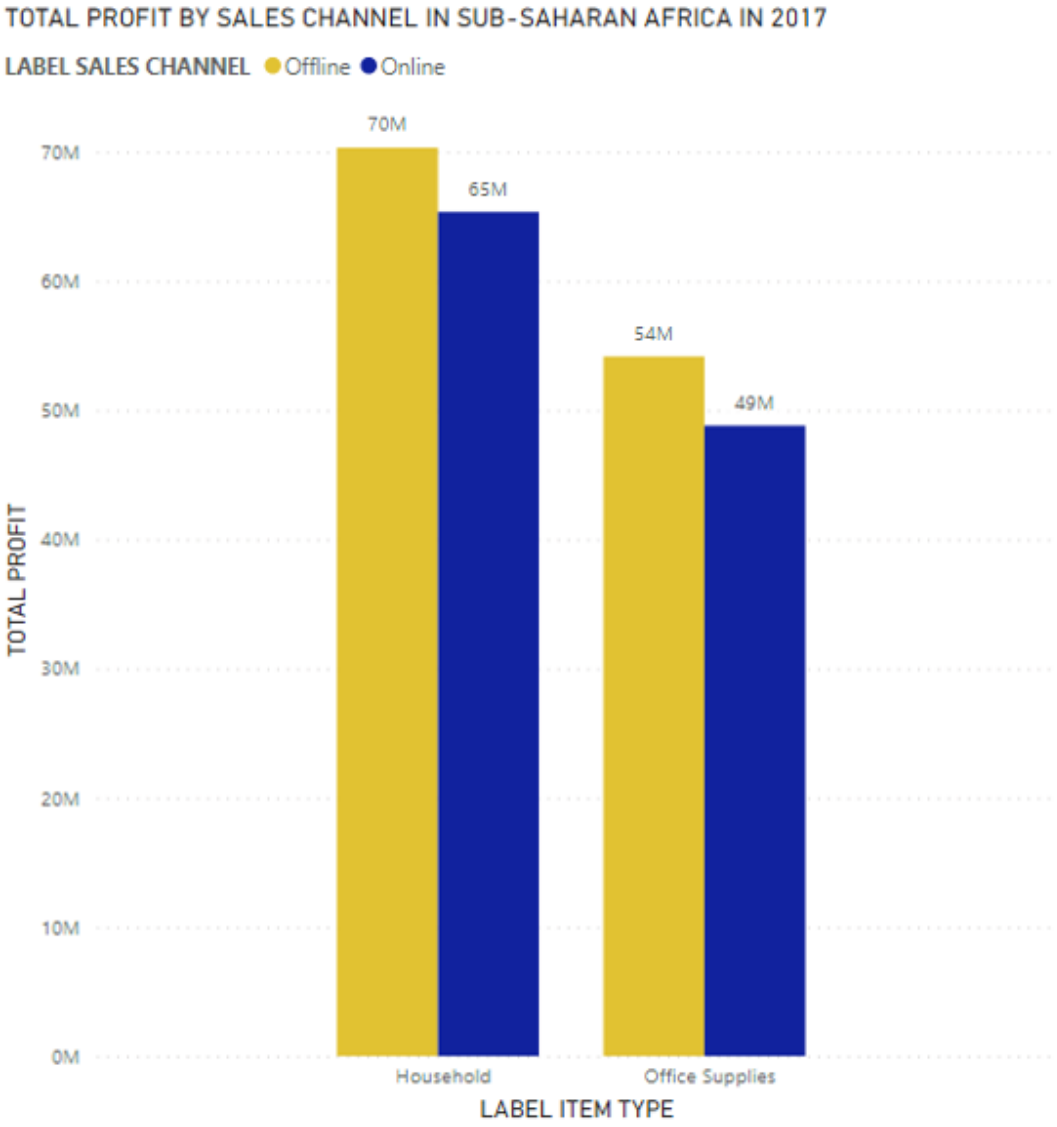


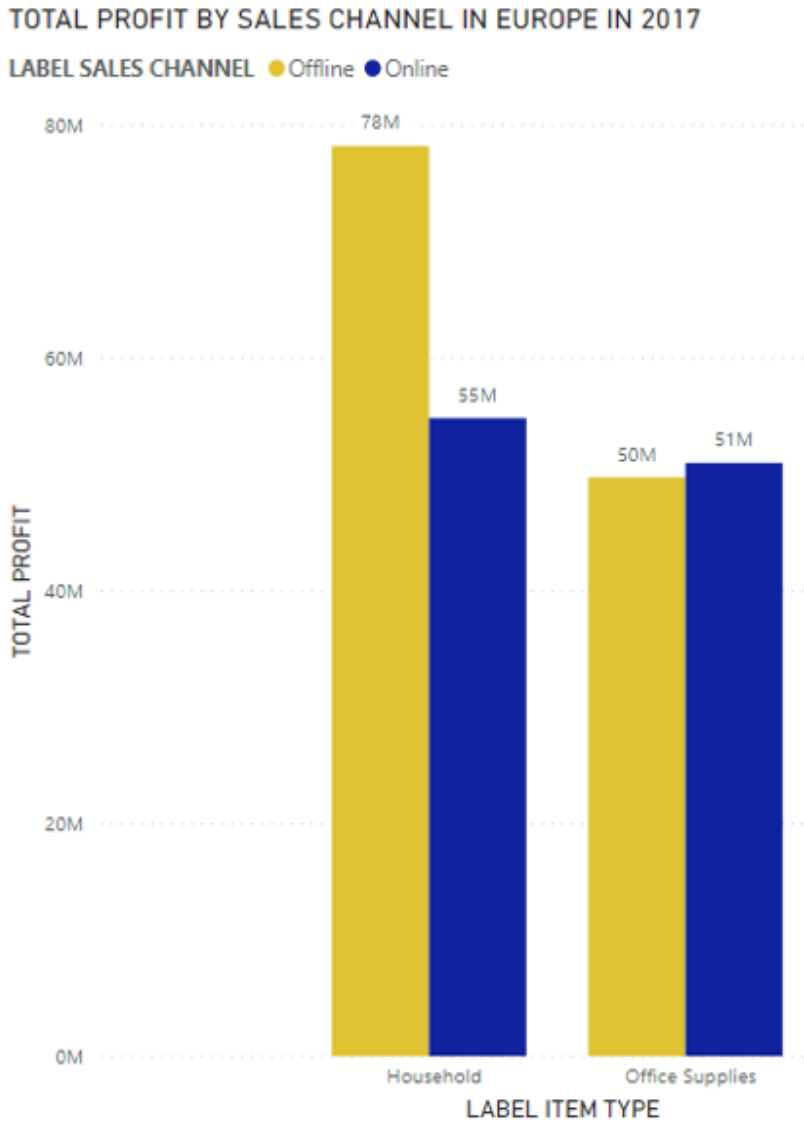


Information about the top 10 days with the largest number of orders per Region in 2017

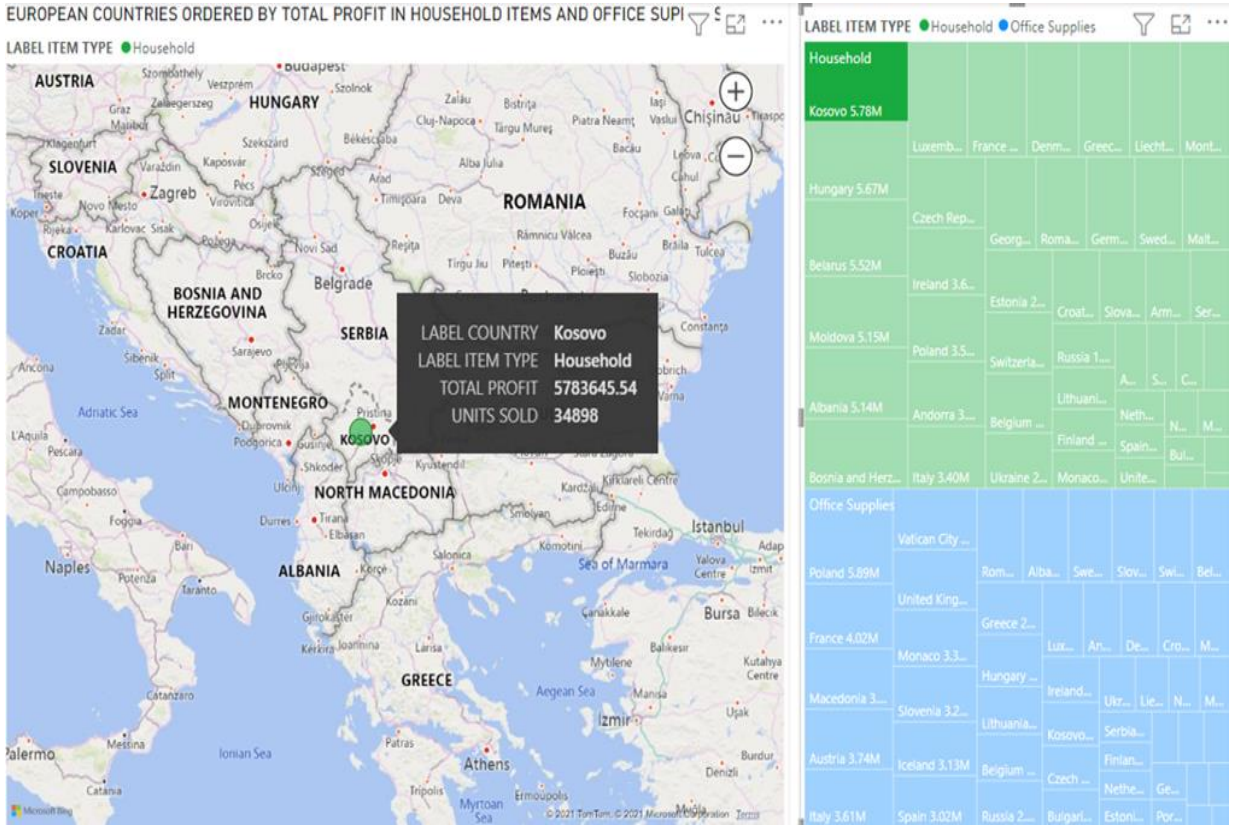


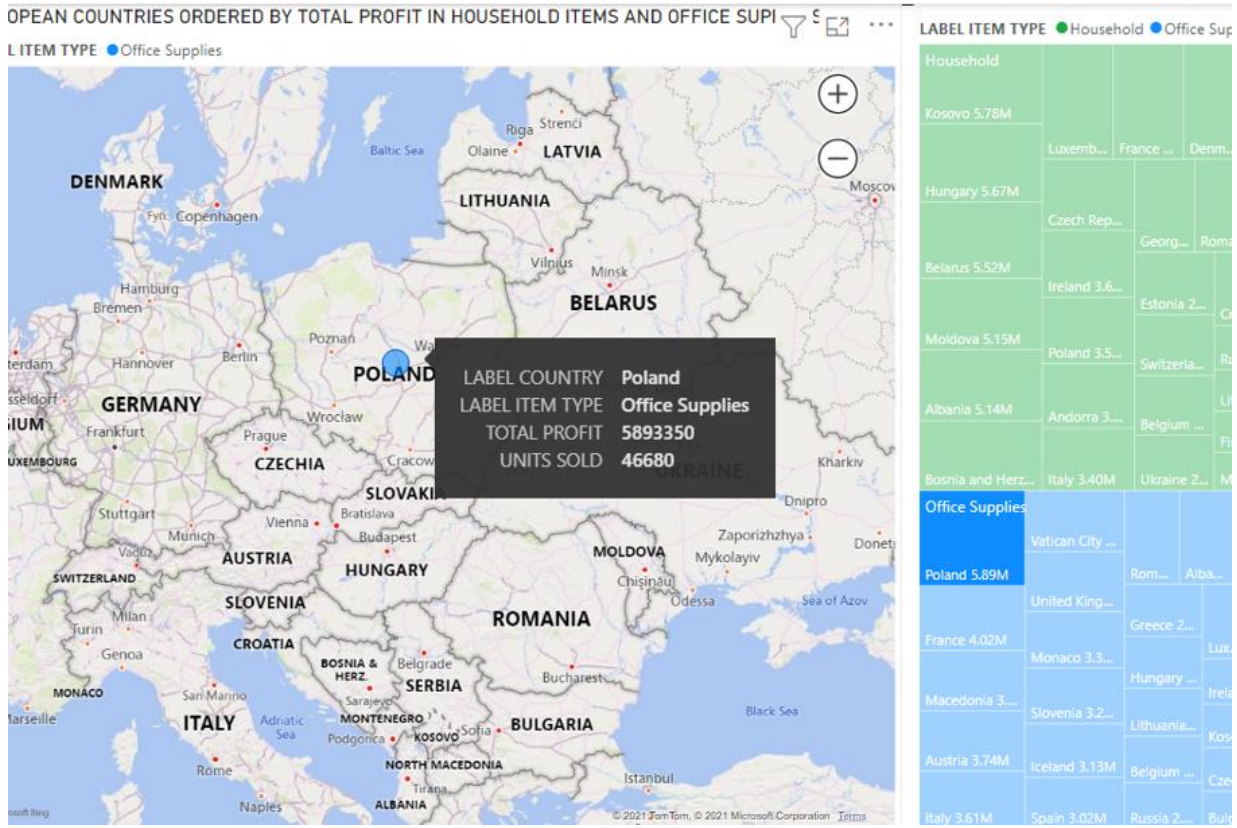
Information about the sales channels' differences between the two Regions in 2017



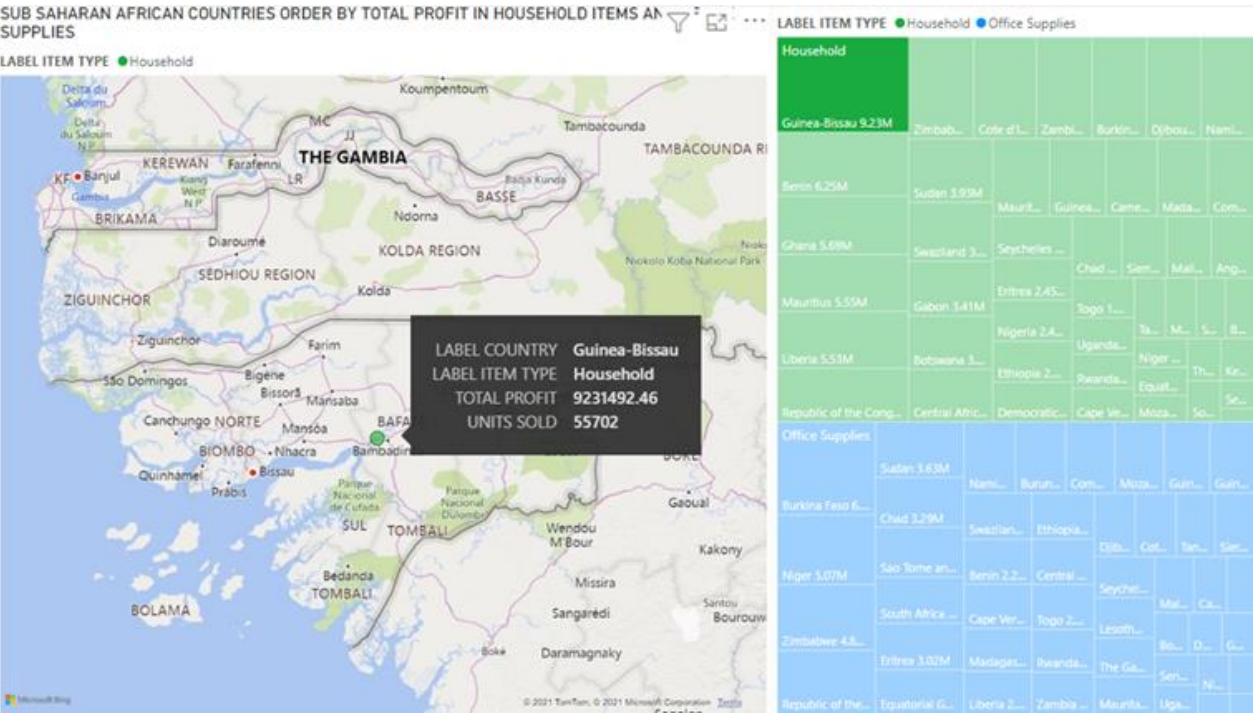


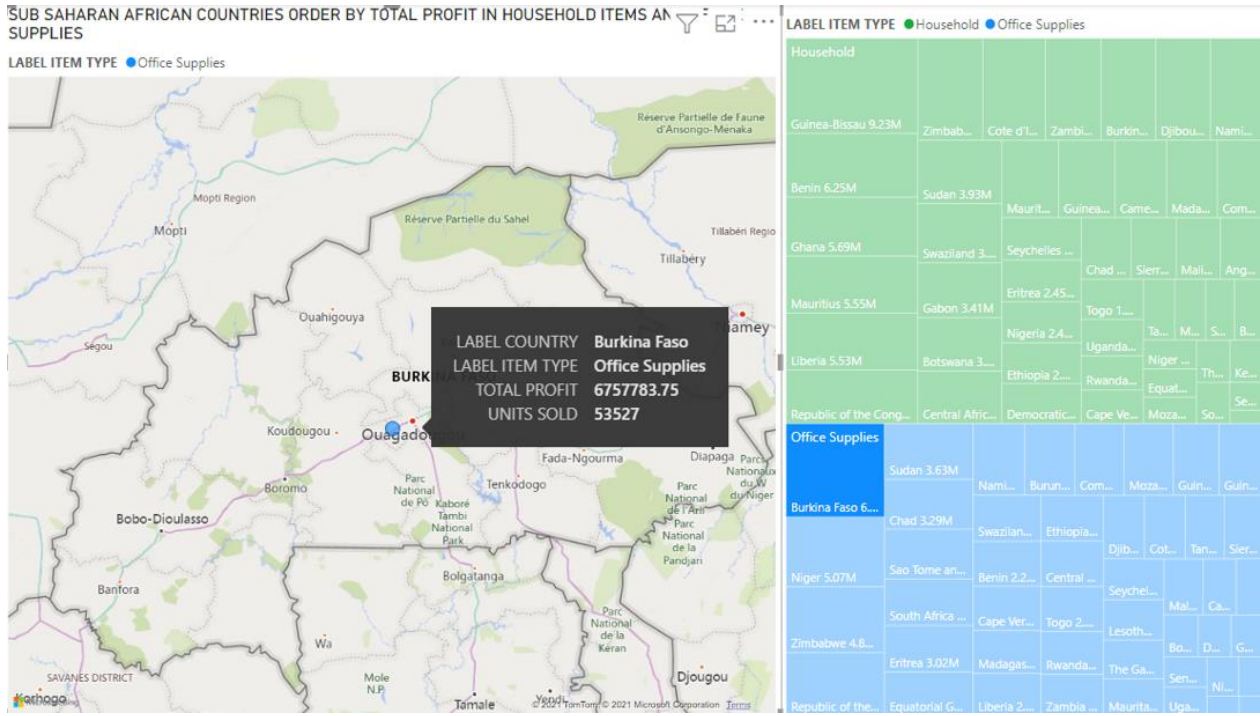
Information about the European Countries order by the total profit of Household items and Office Supplies in 2017





Information about the Sub-Saharan Africa Countries order by the total profit of Household items and Office Supplies in 2017





CONCLUSION

- ✓ Transition from a traditional data management into a multidimensional model.
- ✓ Business Goals Achieved


THANK YOU FOR YOUR ATTENTION
PLEASE CLAP AND DON'T ASK TOUGH QUESTIONS