

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS



# 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](http://www.Kaggle.com)

# DATA REFERENCE & ETL

## ❖ Issues :

Country	Item Type	Sales Count	Order Priority	Order Date	Count
Azerbaijan	Αναζήτηση από το παλιότερο προς το νεότερο				5
Panama	Αναζήτηση από το νεότερο προς το παλιότερο				8
Sao Tome	Ταξινόμηση κατά χρώμα				8
Sao Tome	Ταξινόμηση κατά χρώμα				8
Belize	Προβολή φύλλου				1
Denmark	Απαλοιφή φίλτρου από το "Order Date"				4
Germany	Απαλοιφή φίλτρου από το "Order Date"				7
Turkey	Φίλτρο κατά χρώμα				7
United States	Φίλτρα ημερομηνίας				8
Kazakhstan	Αναζήτηση (Όλα)				4
Haiti	Αναζήτηση (Όλα)				5
Italy	Αναζήτηση (Όλα)				2
Malta	Αναζήτηση (Όλα)				2
Jordan	Αναζήτηση (Όλα)				2
Cambodia	Αναζήτηση (Όλα)				5
Saint Kitts and Nevis	Αναζήτηση (Όλα)				5
Cameroon	Αναζήτηση (Όλα)				3
Bahrain	Αναζήτηση (Όλα)				8
Solomon Islands	Αναζήτηση (Όλα)				8
Monaco	Αναζήτηση (Όλα)				5
Comoros	Αναζήτηση (Όλα)				5
Iceland	Αναζήτηση (Όλα)				8

Order Priority	Order Date	Order ID	Ship Date	Units Sold
Αναζήτηση από το παλιότερο προς το νεότερο				934
Αναζήτηση από το νεότερο προς το παλιότερο				4551
Ταξινόμηση κατά χρώμα				9986
Ταξινόμηση κατά χρώμα				9118
Προβολή φύλλου				5858
Προβολή φύλλου				1149
Απαλοιφή φίλτρου από το "Ship Date"				7964
Φίλτρο κατά χρώμα				6307
Φίλτρα ημερομηνίας				8217
Αναζήτηση (Όλα)				2758
Αναζήτηση (Όλα)				1031
Αναζήτηση (Όλα)				1165
Αναζήτηση (Όλα)				3322
Αναζήτηση (Όλα)				4693
Αναζήτηση (Όλα)				4502
Αναζήτηση (Όλα)				9004
Αναζήτηση (Όλα)				6486
Αναζήτηση (Όλα)				2264
Αναζήτηση (Όλα)				3688
Αναζήτηση (Όλα)				5137
Αναζήτηση (Όλα)				3022
Αναζήτηση (Όλα)				786

## ❖ 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 :

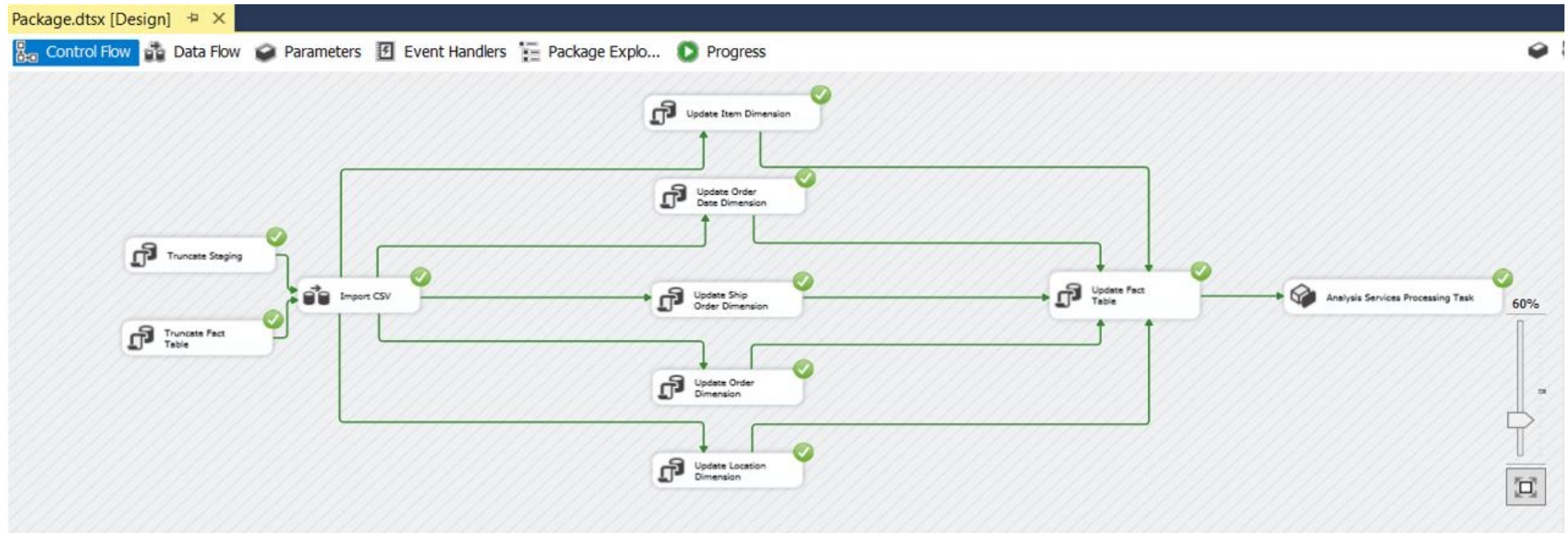
```
Checking for NA's
There are 0 NA's
Checking for NULL's
There are 0 NULL's
```

# DATA REFERENCE & ETL

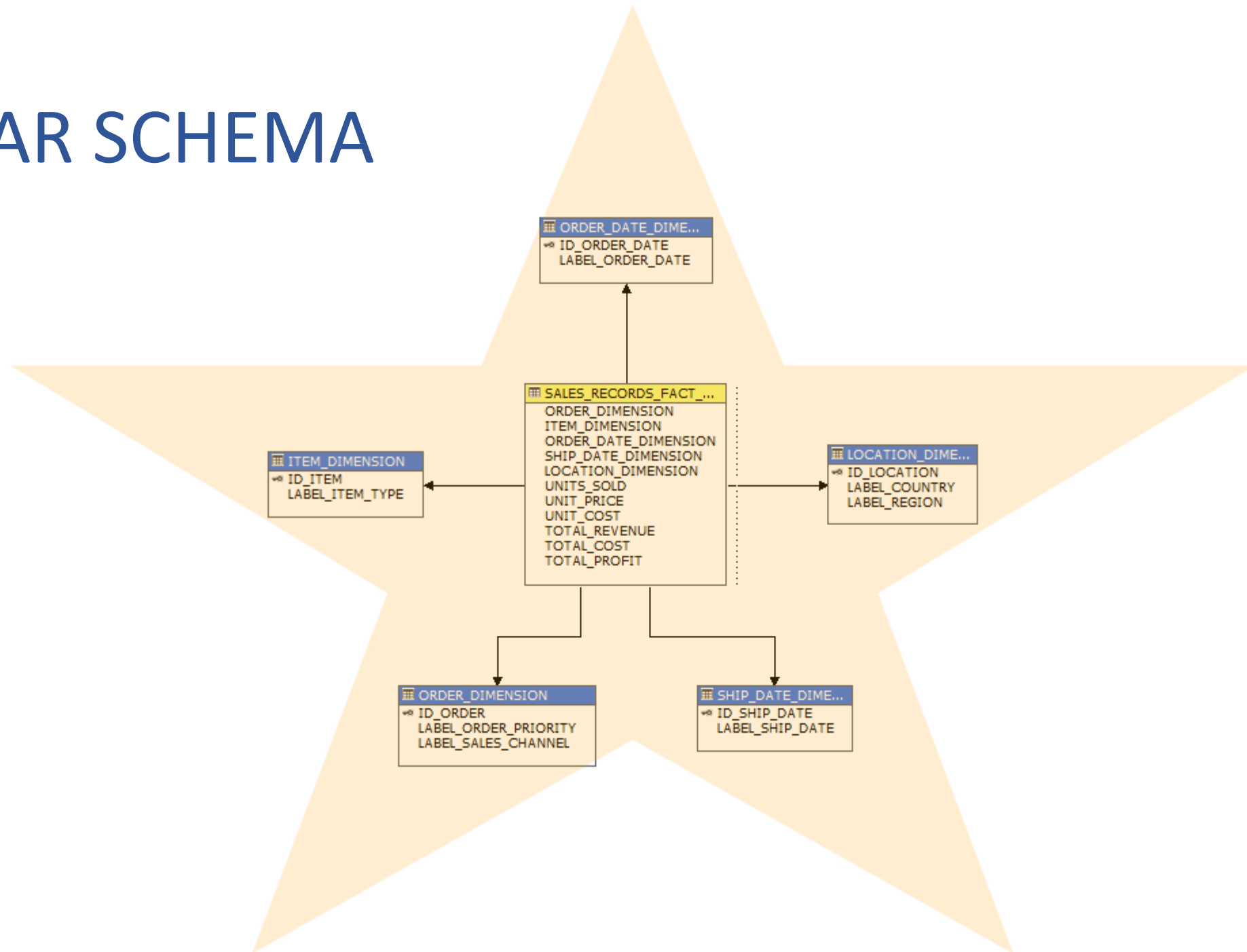
## ❖ Final Format :

```
> 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



# CUBE CREATION

SalesRecordsProject.cube [Design] + X	
Cube Struct... Dimension Usage Calculations KPIs	
Cube Objects	
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

Sales Records Project

Metadata

Search Model

Measure Group:

<All>

Sales Records Project

Measures

SALES RECORDS FACT TABLE

SALES RECORDS FACT TABLE

TOTAL COST

TOTAL PROFIT

TOTAL REVENUE

UNIT COST

UNIT PRICE

UNITS SOLD

AVERAGE\_UNITS\_ORDERED

KPIs

ITEM DIMENSION

LOCATION DIMENSION

ORDER DATE DIMENSION

Dimension

<Select dimension>

AVERAGE\_UNITS\_ORDERED

5001.44617



# EXAMPLE OF OLAP QUERY

## ❖ Slice & Dice :

The screenshot displays the 'SalesRecordsProject.cube [Design]' interface. The top toolbar includes options like 'Cube Struct...', 'Dimension Usage', 'Calculations', 'KPIs', 'Actions', 'Partitions', 'Aggregations', 'Perspectives', 'Translations', and 'Browser'. The main workspace shows a query configuration table with the following data:

Dimension	Hierarchy	Operator	Filter Expression	Parameters
LOCATION DIMENSION	LABEL COUNTRY	Equal	{ Greece }	<input type="checkbox"/> <input type="checkbox"/>
ORDER DATE DIMENSION	LABEL ORDER DATE	Range (Inclusive)	1/1/2010 : 12/31/2010	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				

Below the query configuration, a table displays the results of the query:

LABEL ITEM TYPE	UNITS SOLD
Baby Food	56804
Beverages	77543
Cereal	65908
Clothes	114175
Cosmetics	112962
Fruits	58661
Household	54496
Meat	32281
Office Supplies	67471
Personal Care	55911
Snacks	60219
Vegetables	39259

The left sidebar shows the 'Sales Records Project' tree with folders for 'Metadata', 'Measure Group', 'KPIs', and 'ITEM DIMENSION'. The 'ITEM DIMENSION' folder is expanded, showing 'ID ITEM', 'LABEL ITEM TYPE', 'LOCATION DIMENSION', 'ORDER DATE DIMENSION', and 'ORDER DIMENSION'. The 'Calculated Members' section at the bottom is currently empty.

And yet... we  
are not done



# DIMPET ANALYTICS

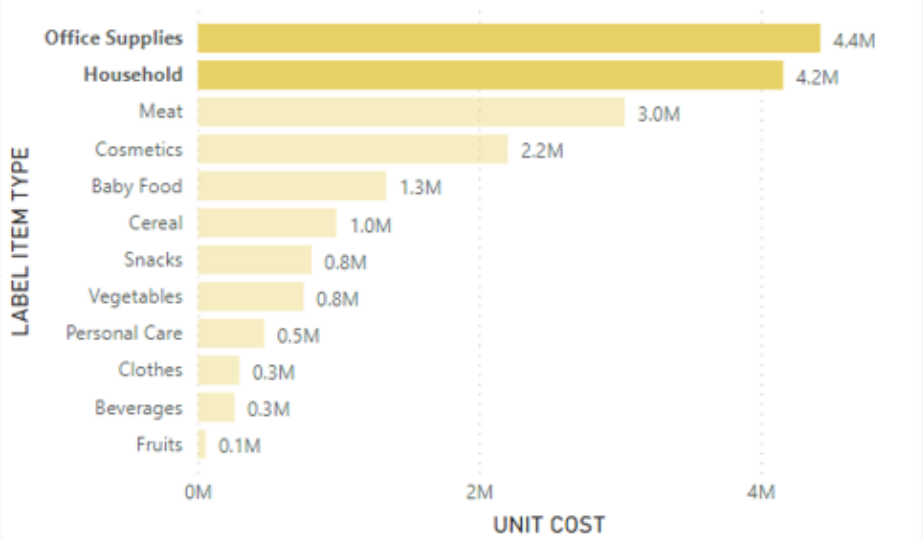
POWER BI

# BUSINESS CASE SCENARIO

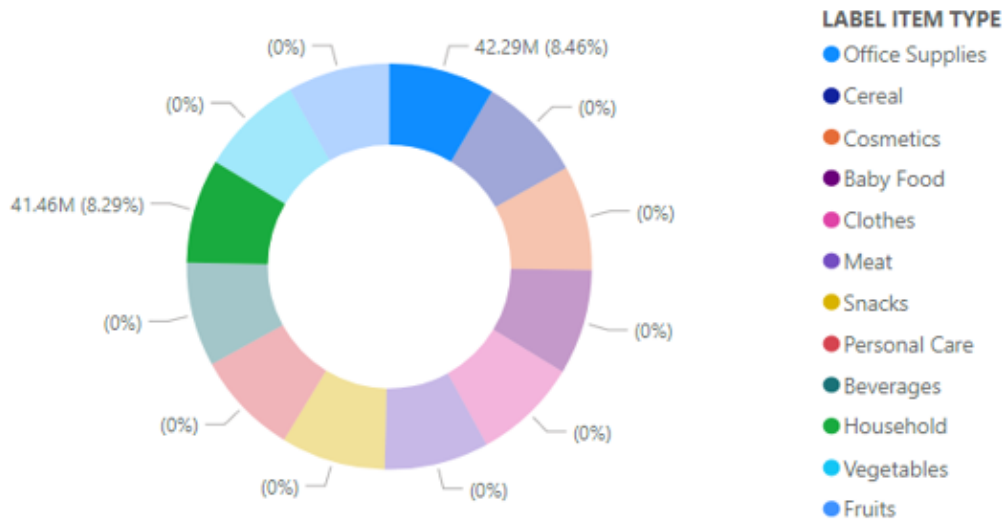
- ❖ Year of analysis : 2017.
- ❖ Customer : Multinational company that wants to invest in two specific product markets (Household items, Office supplies).
- ❖ In which regions should they focus on? In which countries?
- ❖ Should they focus on offline sales channels, on online or both?
- ❖ Is every time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

# Information about Office Supplies and Household Items

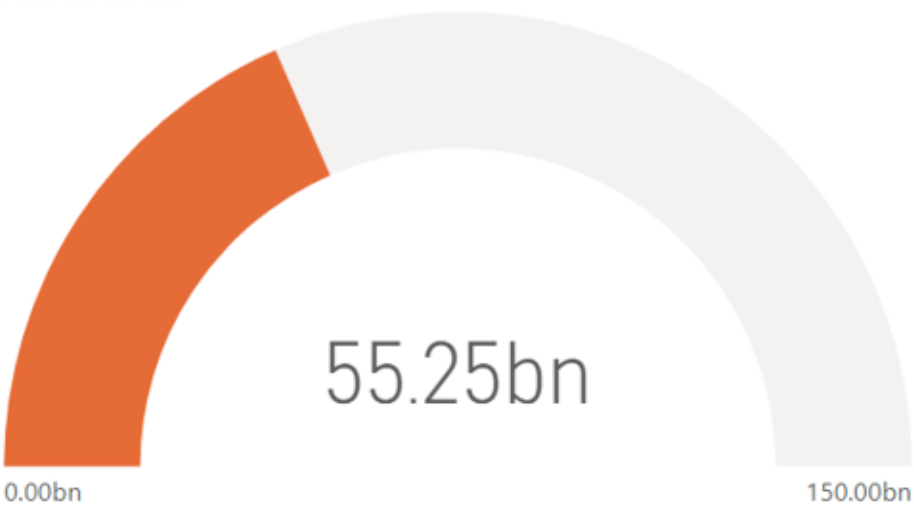
UNIT COST by LABEL ITEM TYPE



UNITS SOLD by LABEL ITEM TYPE

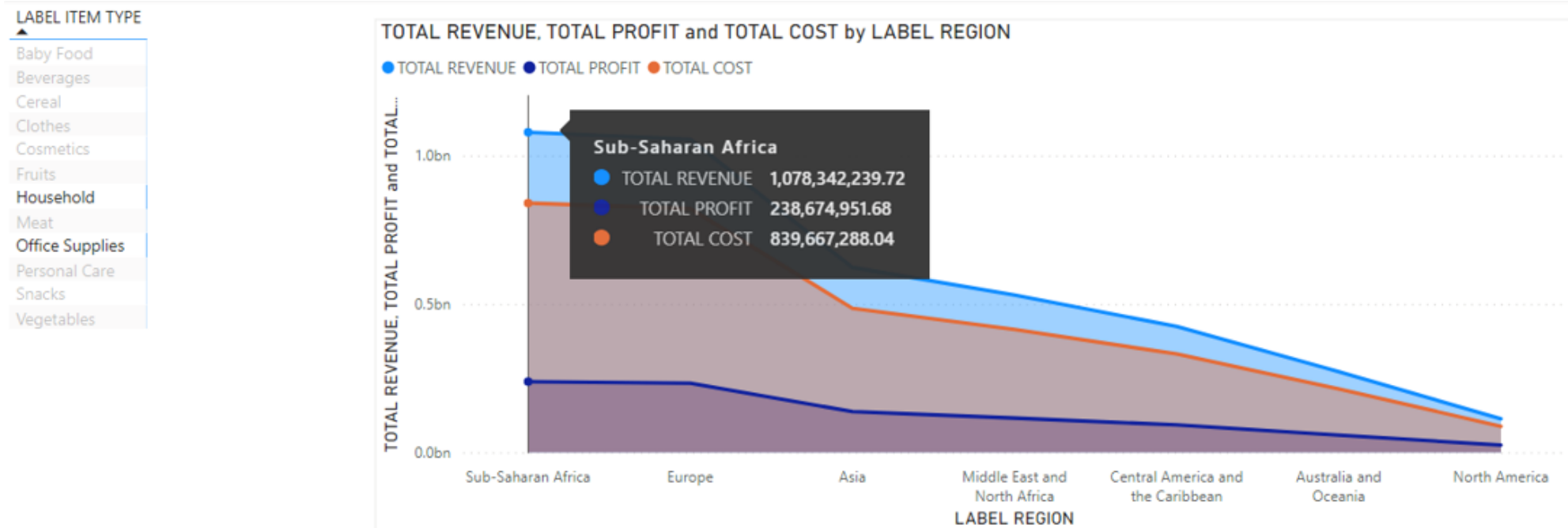


TOTAL REVENUE

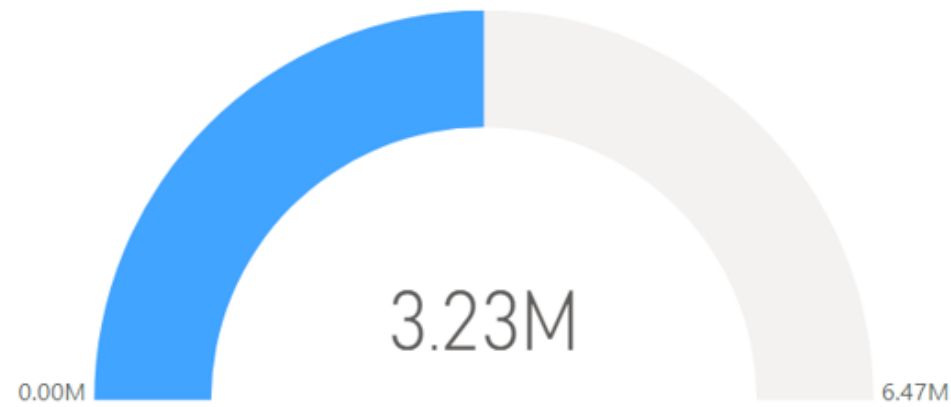




# Information about the Best-Sellers Regions worldwide (Sub-Saharan Africa) in 2017



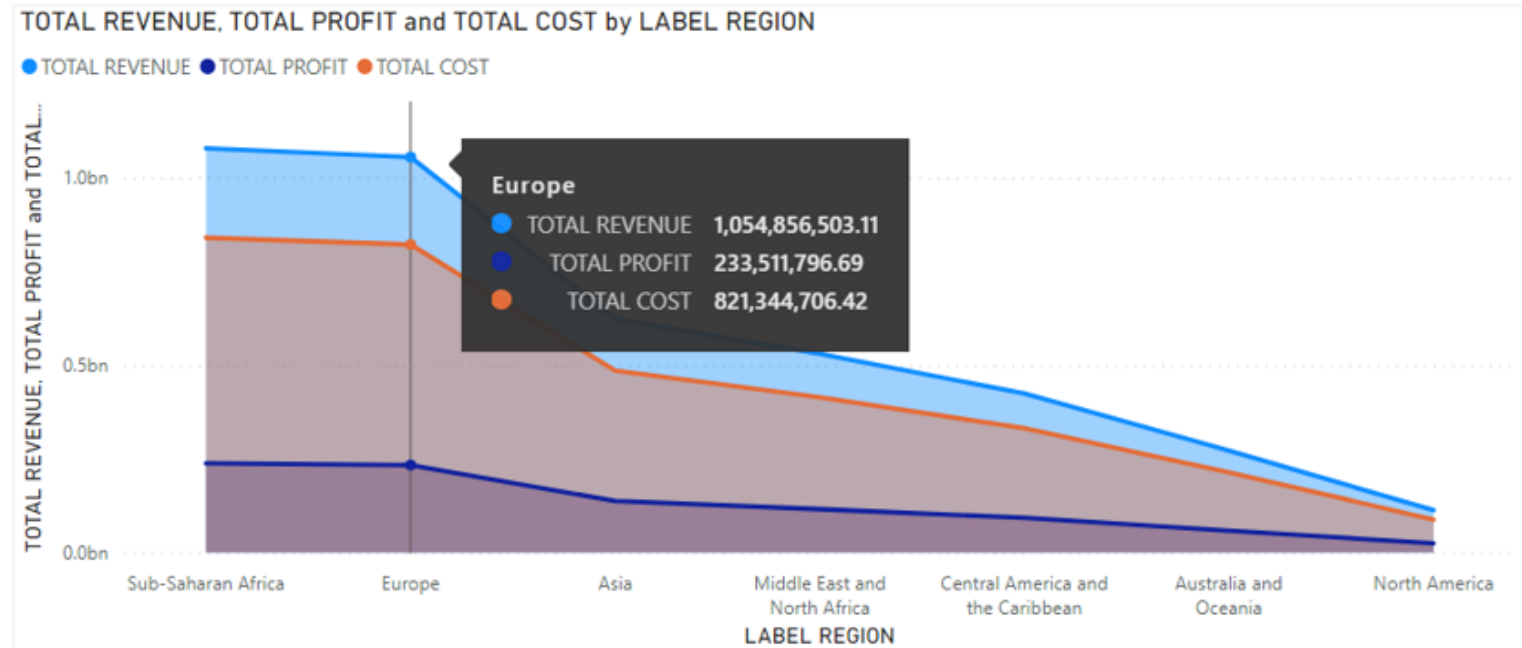
UNITS SOLD



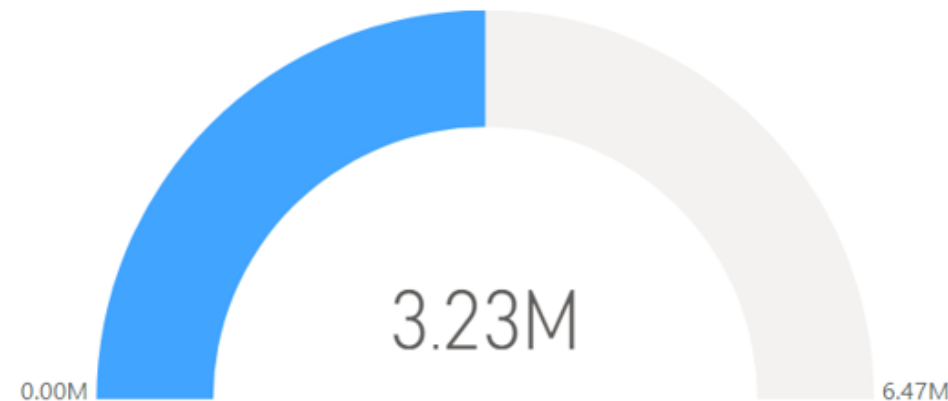
# Information about the Best-Sellers Regions worldwide (Europe) in 2017

LABEL ITEM TYPE

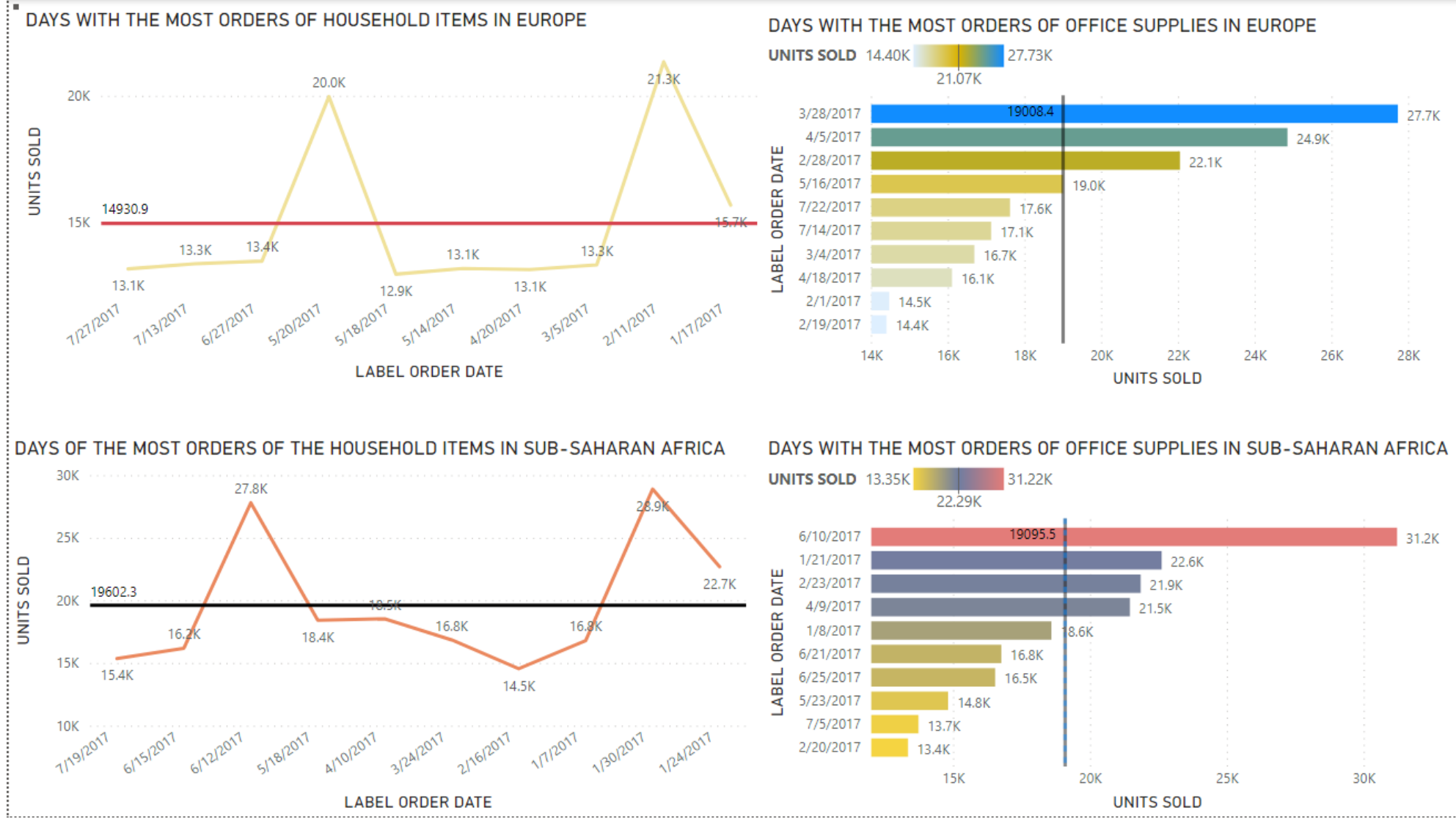
Baby Food  
Beverages  
Cereal  
Clothes  
Cosmetics  
Fruits  
Household  
Meat  
Office Supplies  
Personal Care  
Snacks  
Vegetables



UNITS SOLD



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

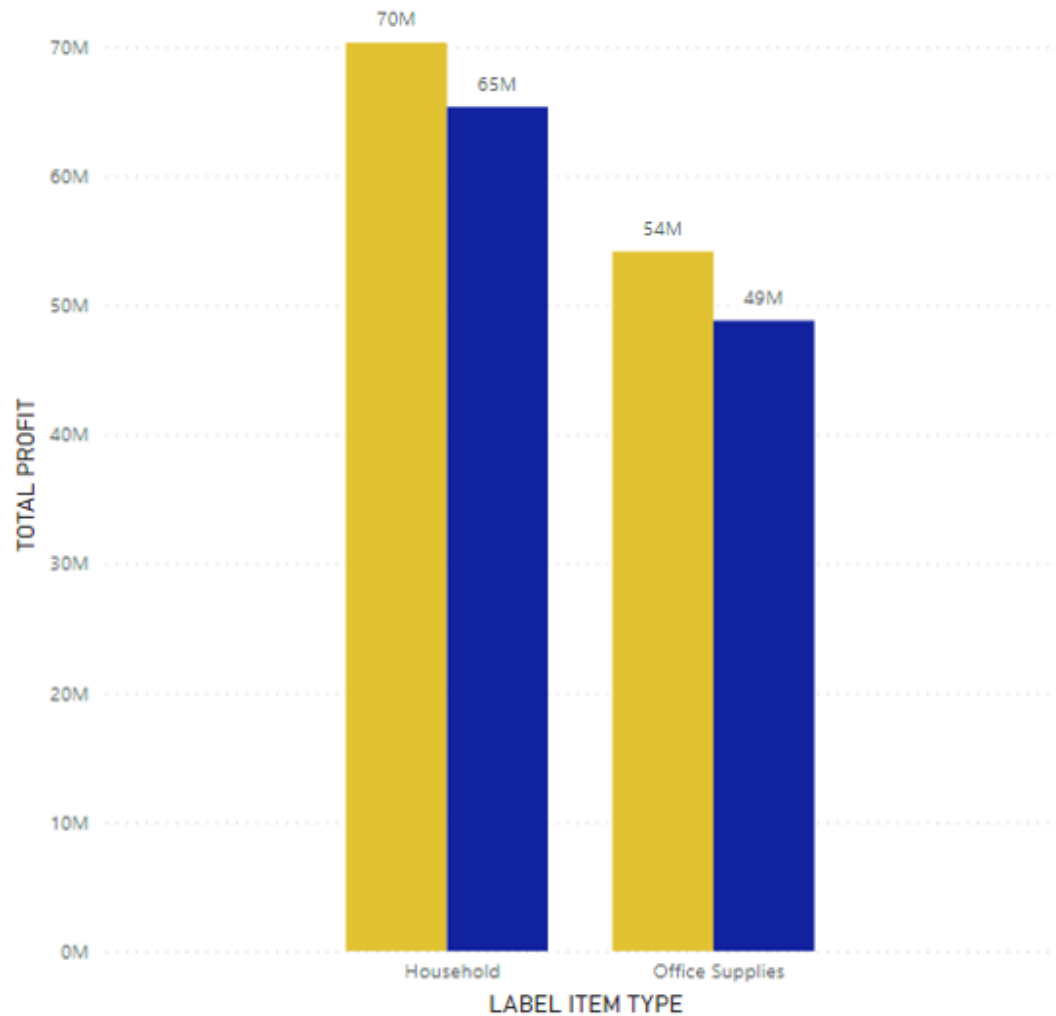




# Information about the sales channels' differences between the two Regions in 2017

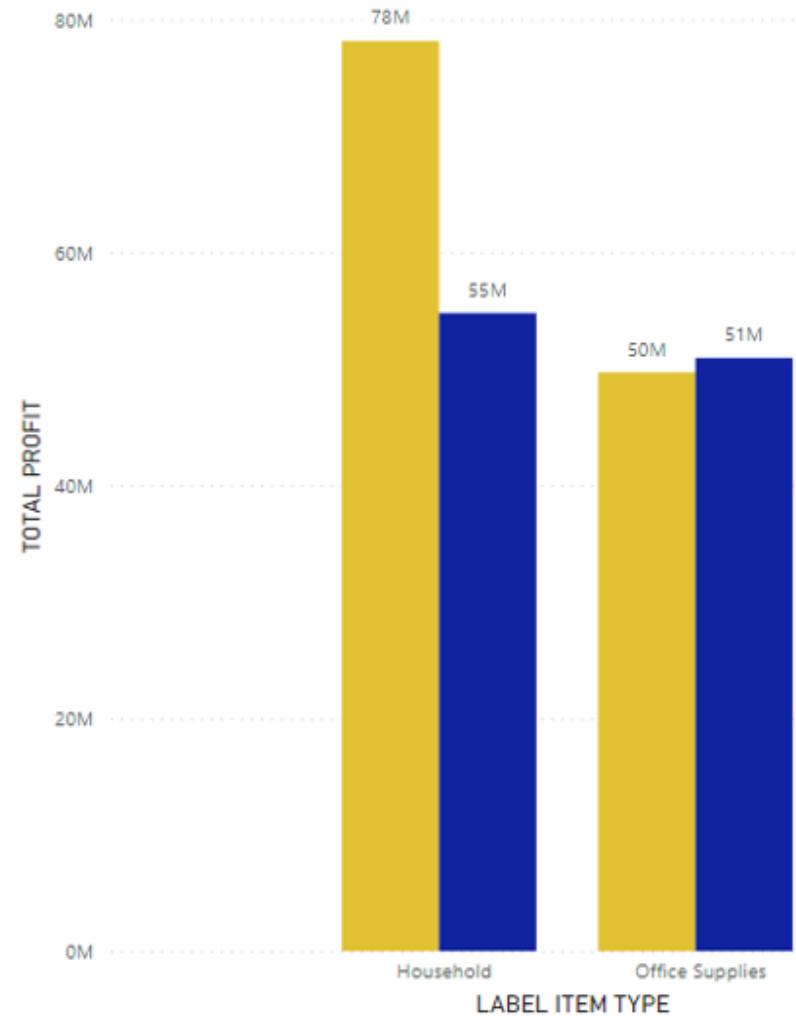
TOTAL PROFIT BY SALES CHANNEL IN SUB-SAHARAN AFRICA IN 2017

LABEL SALES CHANNEL ● Offline ● Online

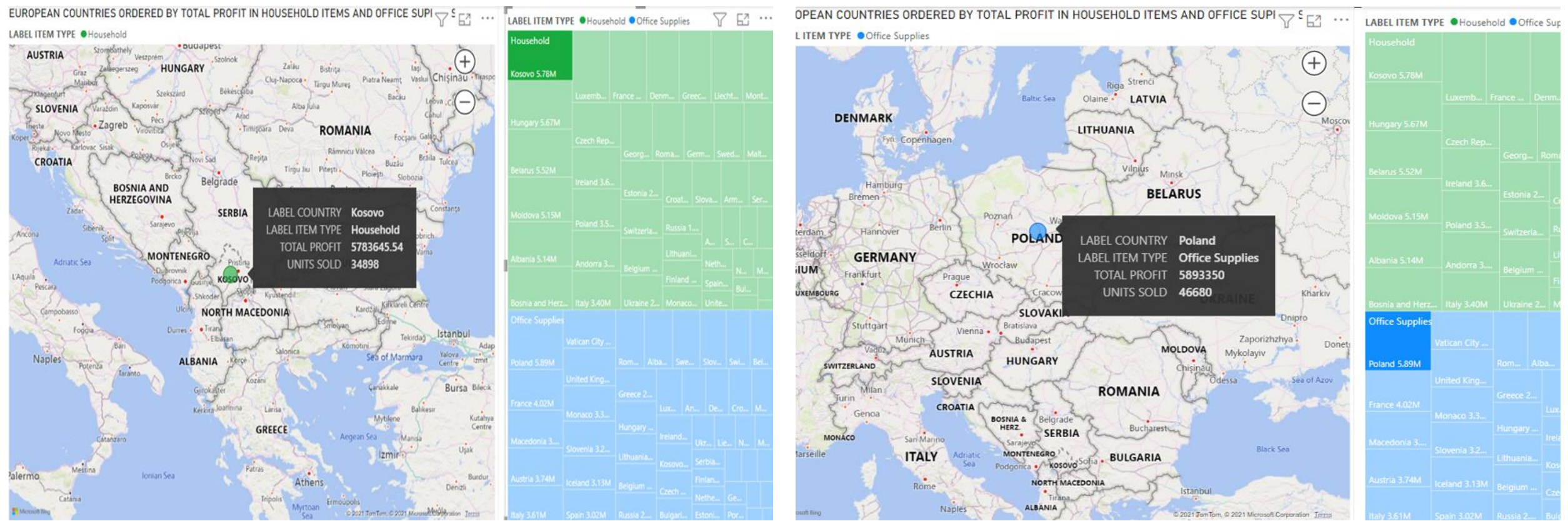


TOTAL PROFIT BY SALES CHANNEL IN EUROPE IN 2017

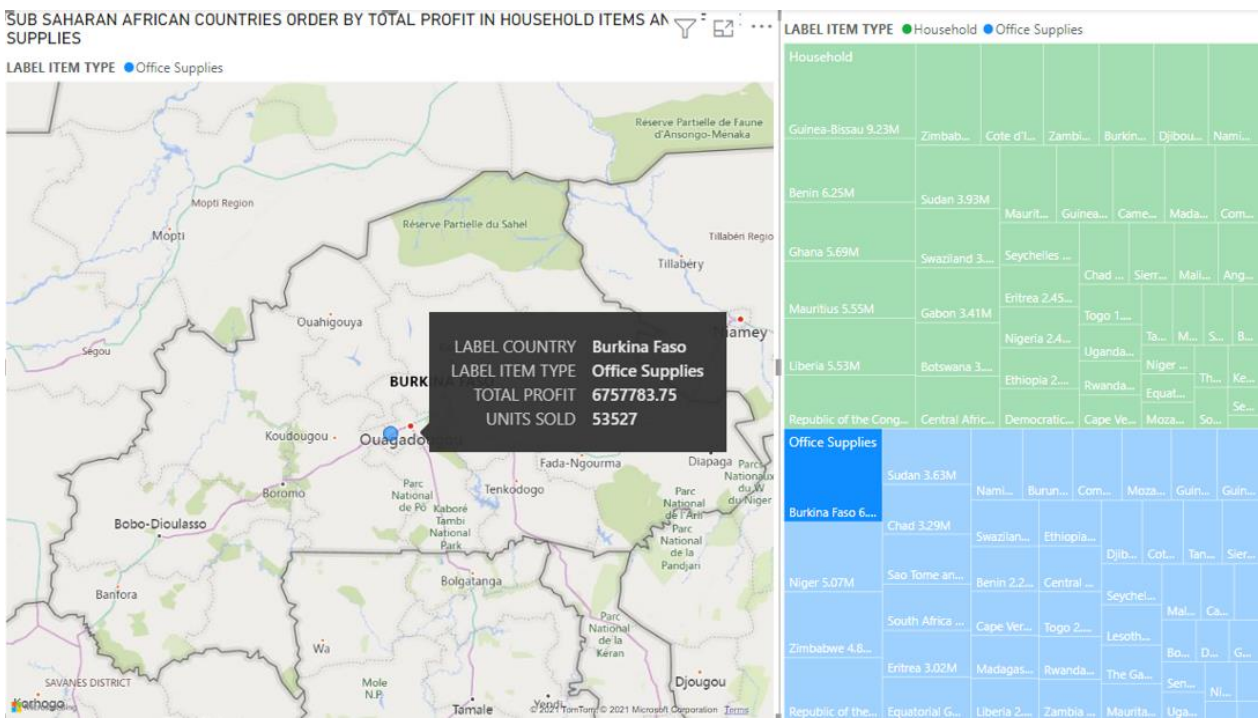
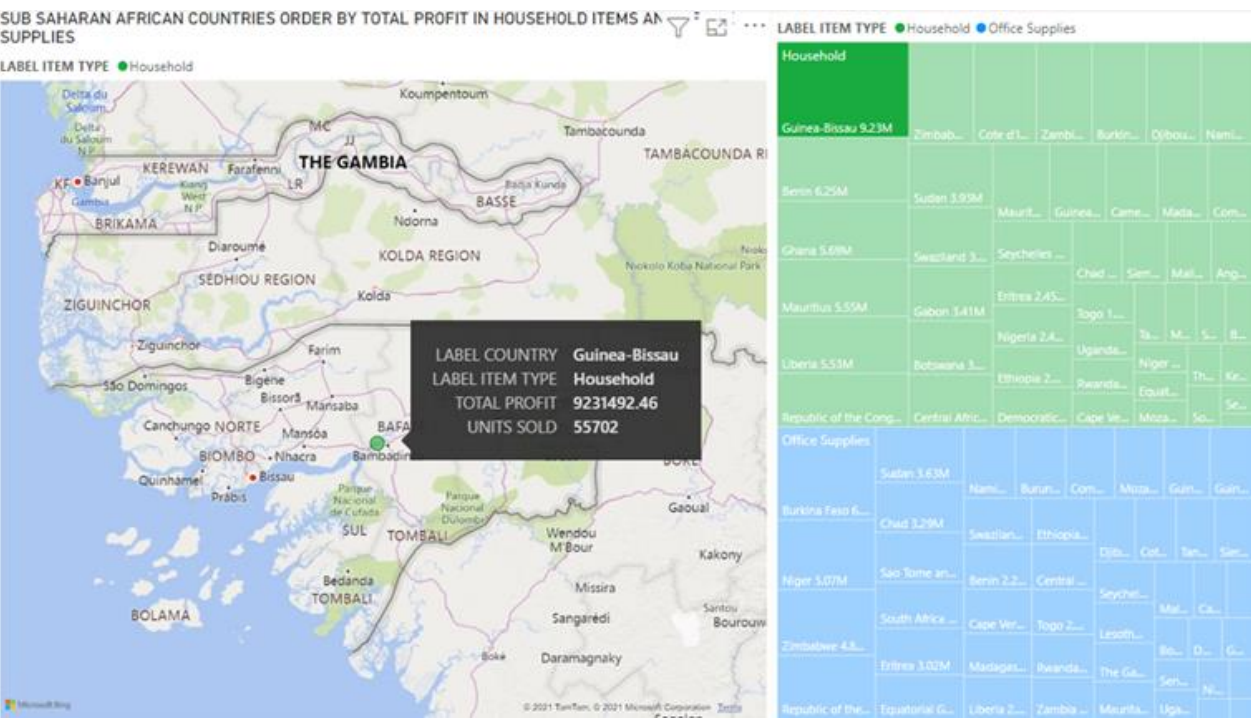
LABEL SALES CHANNEL ● Offline ● Online



# 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**