Data Processing and Visualization (P02)

DECISION SUPPORT SYSTEMS

Students:

18825 – João Fernandes

18836 – Carlos Martins

Teacher:

Joaquim Gonçalves

Braga, January, 2021

Index

Illustration Index

1. Introduction

The goal of this project is the development of a data processing and visualization software solution that provides information about the data from Adventure works, such as: sales information, customers, products, etc.

In this project we have done data preparation, measures and columns that we thought necessary to perform the data presentation. These measures and columns were created so that we could create several dashboards in Power BI, since this allows us to show data in a better way.

2. Data acquisition and preparation

In order to optimize the information that exists in the tables, we have cleaned some unnecessary data. Examples of such **data are those that are *"null"***, **products that are unavailable**, photos that are **invalid**, **change values** in order to fetch external data and **add columns** that can help us later. Here are some examples of what has been done:

Remove products with category *"null":*

```
Table.RemoveColumns(#"Filtered Rows",{"ProductSubcategoryKey"})
```

Remove from the table "DimProduct", the "StartDate" and "EndDate":

```
Table.RemoveColumns(dbo_DimProduct,{"StartDate", "EndDate"})
```

Remove from the column "LargePhoto" from the table "DimProduct":

```
Table.RemoveColumns(#"Filtered Rows",{"LargePhoto"})
```

A column "Country" was added in the table "DimCustomer", in order to relate to the table "DimGeography":

```
Country = RELATED(DimGeography[EnglishCountryRegionName])
```

We added a column "FullName" in the table "DimCustomer", in order to combine the first name and the last name of each customer.

```
FullName = DimCustomer[FirstName] & " " & DimCustomer[MiddleName] & " " &
DimCustomer[LastName]
```

To have a region with all the data that the tables present, we created a **hierarchy** with the following columns: "EnglishCountryRegion", "SateProvinceName" and "City". **Data categories**, *State or Province* and *City*, have been assigned for the columns "StateProvinceName" and the "City", respectively.

In the fields that show values, we assign the **currency format** and **two decimal cases**.

3. Data modelling and processing

3.1. Measures

To relate tables and be able to show more useful information, measures and columns have been created. Now, we will talk about each of them:

3.1.1. Most ordered product

This measure is used to search for the most purchased product and shows the name of that product.

```
Most_orderedProduct =
CONCATENATEX(
    TOPN(
        1
        , VALUES(DimProduct[ModelName])
        , CALCULATE(SUM(FactInternetSales[OrderQuantity]))
    )
    , [ModelName]
    , ", "
)
```

3.1.2. Number of customers and products

"DISTINCTCOUNT" was used to count the number of customers and of products without repeating them. This will tell us the total number of customers and products.

```
num_customers = DISTINCTCOUNT(FactInternetSales[CustomerKey])
```

```
num_products = DISTINCTCOUNT(DimProduct[ProductKey])
```

3.1.3. Order average

This measure calculates the average cost of the orders. It uses the AVERAGEX function with the first argument being the fact table and the other argument is the total order cost.

```
order_avg = AVERAGEX(FactInternetSales,
(FactInternetSales[OrderQuantity]*FactInternetSales[UnitPrice])*(1-
FactInternetSales[UnitPriceDiscountPct]) + FactInternetSales[Freight] +
FactInternetSales[TaxAmt])
```

3.1.4. Product price average

We repeat the process that we used in the previous measure to calculate the average product price.

```
product_price_avg = AVERAGEX(DimProduct, DimProduct[ListPrice])
```

3.1.5. Last year's sales

This measure calculates last year's total sales given a certain day. For example, if the date is 19/01/2022 it calculates the total amount of sales in 19/01/2021.

```
prev_year_total_amountINT = CALCULATE([total_amountINT],
DATEADD('DimDate'[FullDateAlternateKey], -1, YEAR))
```

3.1.6. Most sold product per region

This measure shows the most sold product and the total sales of that same product. We use it to know this in each region.

```
Top_1_ProductPerRegion =
VAR
    RankingContext = VALUES(DimProduct[EnglishProductName])
    Return
    CALCULATE([total_amountINT],
        TOPN(1, ALL(DimProduct[EnglishProductName]), [total_amountINT]),
        RankingContext)
```

3.1.7. Total internet sales revenue

This measure calculates the sum of every full price of each order. This full price includes the product price, freight, taxes and a discount if there is one.

```
total_amountINT = SUMX(FactInternetSales,
(FactInternetSales[OrderQuantity]*FactInternetSales[UnitPrice])*(1-
FactInternetSales[UnitPriceDiscountPct]) + FactInternetSales[Freight] +
FactInternetSales[TaxAmt])
```

3.1.8. Month over month revenue variation

Here, we used a quick measure to calculate the month-over-month change in the total amount of orders.

```
total_amountINT MoM% =
VAR __PREV_MONTH =
    CALCULATE(
        [total_amountINT],
        DATEADD('DimDate'[FullDateAlternateKey], -1, MONTH)
    )
RETURN
    DIVIDE([total_amountINT] - __PREV_MONTH, __PREV_MONTH)
```

4. Data visualization

4.1. Internet sales dashboard

As we can see in the image below, we have a general dashboard for Internet sales. This dashboard has some **cards** with values and information that we think are important in these sales, such as the total value of sales, the bestselling product, the total number of customers, and an average of the amount spent on each order. Some of the measures mentioned above were used for these cards.

The **pie chart** shows the total sales in each country and a **tooltip** that has information about the total sales of the previous and current year selected and an **area chart** that shows the total sales per month. On the **clustered column chart**, we have the sales per month, we also added a **tooltip** on that chart to know which category was the most sold in that month.

In one **table** we show the total amount of sales per category, in the other we show the variance in percentage per month of sales to know if the following month had better sales than the previous one.

To filter by year and category, the information given in the dashboard, we added two **slicers**.
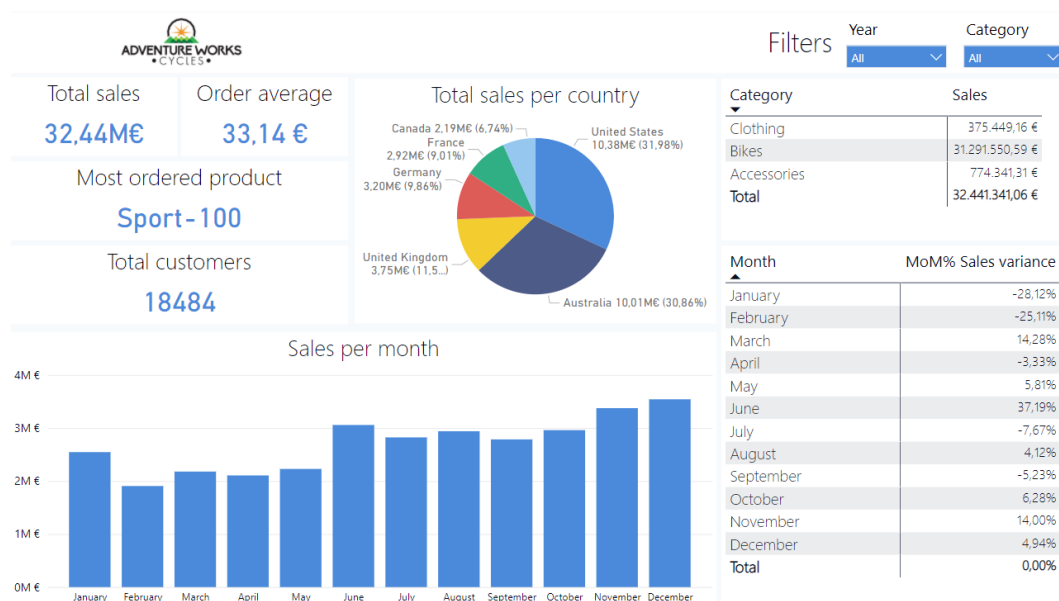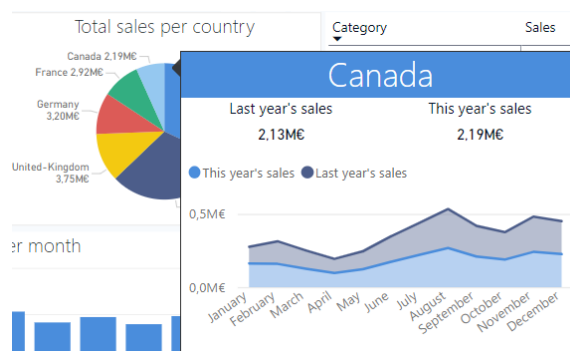

Figure 1 - Internet sales dashboard
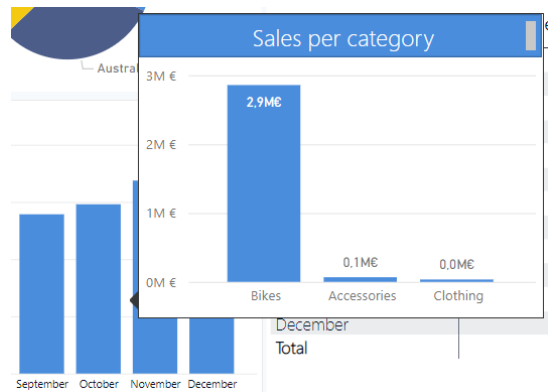

Figure 2 - Previous year sales tooltip

Figure 3 - Sales per category tooltip

4.2. Customer sales report dashboard

This dashboard is the customer sales report, where we can see information about sales by customers.

We have a **card** that indicates the total number of customers. A **table** that shows the list of customers and the amount already spent by each one. A **funnel** that indicates the top 10 customers, the first customer being the one who spent the most, and so on. In the table and funnel, a **tooltip** was added, where we can see the data of the selected customer.

We added four **slicers** that filter by year, customer name, product category and country.
To show a customer purchase history we used a **multi-row card**, to see the orders of each customer.
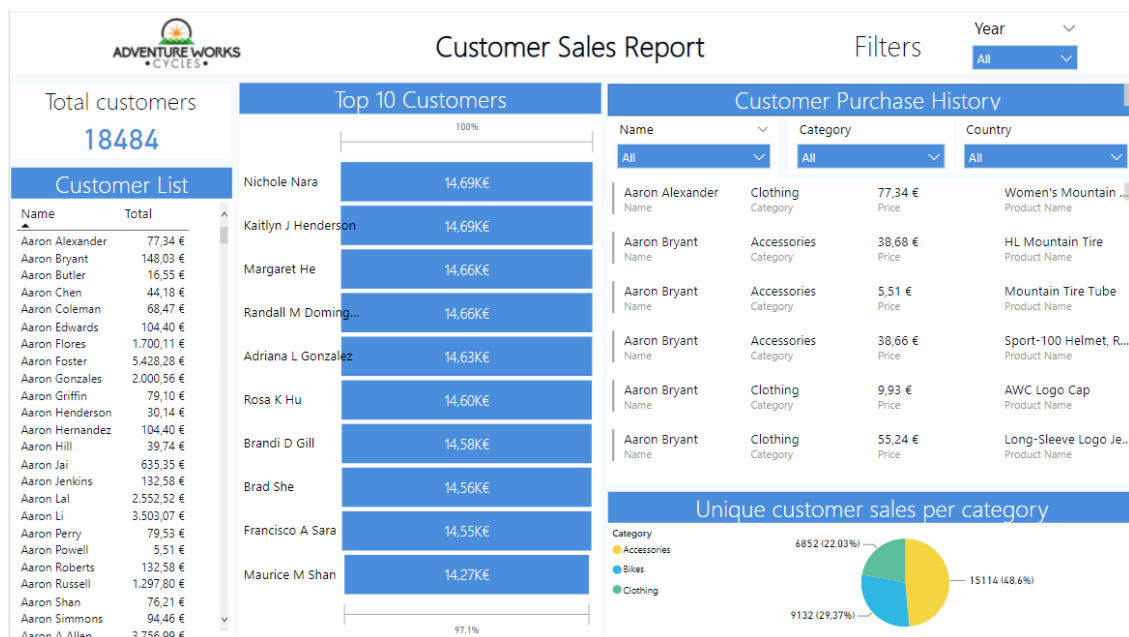Finally, a **pie chart** that shows the number of customers for each category.

Figure 4 - Customer sales dashboard

Figure 5 - Customer information tooltip

4.3. Product sales report dashboard

This dashboard shows a product sales report, where we can see information about sales by products.

We have two **cards** that indicates the total number of products and the price average of all the products. Two **tables** that shows the list of products, their named and prices. A **map** showing the most sold product for each country. In the same visualization we added a **tooltip**, that shows which product is the most sold, its price and the total amount raised from that product.

To show orders per category, we add a **pie chart** that shows us the total of orders for each category. Two **slicers** were added to filter by category: Accessories, Bikes and Clothing, and filter by year.
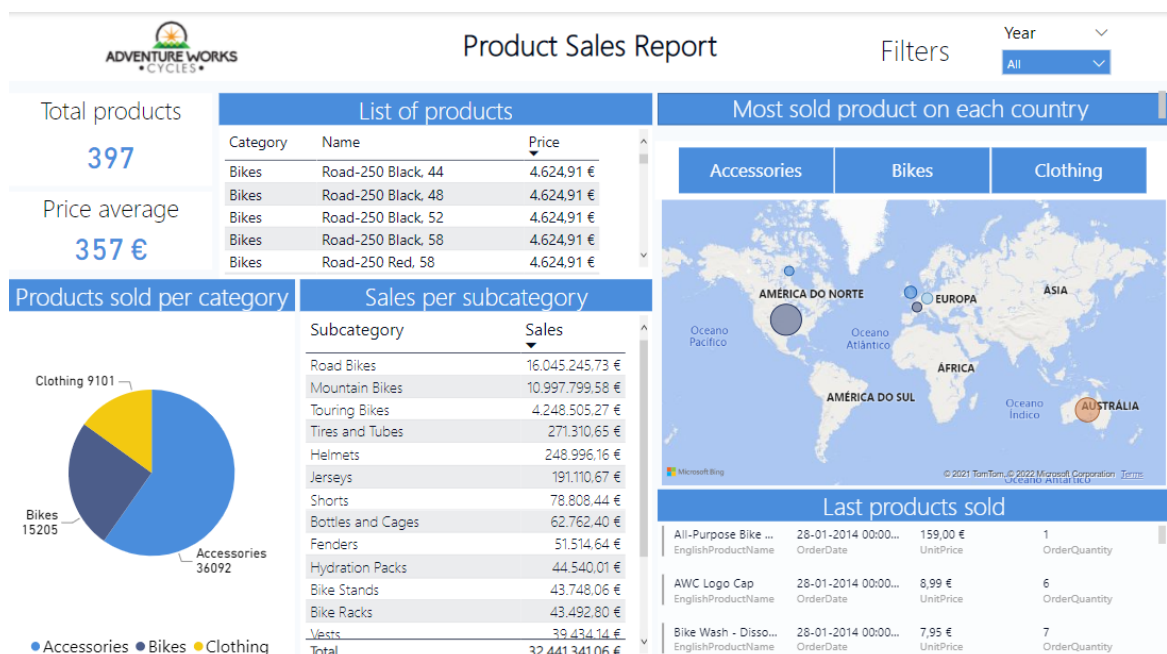

Figure 6 - Product sales dashboard

Figure 7 - Map tooltip

5. Conclusion

With the completion of this work, we have gained a better understanding of what can be done when analytical processing and data visualisation is required and how this will be useful in future projects. When put into practice we have a much greater notion of what can be done at a professional level.