Ecommerce store data analysis

1. Data Importing:

All files were checked and the type of the files were comma-separated values files (CSV)

The data was imported from (Get Data—Text/CSV--- select the file --- open)

1. Data Transformation:

Each file was opened on power query from (transform data).

I deleted the useless columns with no values or values that cannot be used in the analysis and data with different languages. (remove columns from tab home)

I changed some columns' data types (standard cost, list price) to decimal numbers to be able to use them in calculations.

I replaced the values in the gender and marital status columns with full names (Female, Male), (Single, Married) to be more familiar in the report.

After all cleaning, I select (close and apply)

1. Data Exploring:

As the data is about an e-commerce store, the fact table is (Internet sale)

1. Data Modelling:

I checked the table joining from the model view and created a relationship between:

* Date table (date key) and Internetsale (order date key)
* Product category column (product category key) and Internetsale (Product key)

1. Data analysis and visualization

I created some columns and measures from the available data using DAX

Columns created:

1. SALES

SALES = InternetSale[UnitPrice]+InternetSale[TaxAmt]+InternetSale[Freight]

1. Profit

Gross Profit = InternetSale[SALES]-InternetSale[TotalProductCost]

1. Profit margin ratio

Profit Margin Ratio = DIVIDE(InternetSale[Profit],InternetSale[UnitPrice])

1. Year

Year = YEAR(InternetSale[OrderDate])

1. Quarter

Quarter = QUARTER(InternetSale[OrderDate])

1. Month

Month = MONTH(InternetSale[OrderDate])

1. Day

Day = DAY(InternetSale[OrderDate])

Then I created

1. Year hierarchy that contains (Year, Quarter, Month and day)
2. Geographic hierarchy (territory group, territory country, territory region)

My table

That contain few measures as

1. Total sale

TotalSales = SUM(InternetSale[UnitPrice])

1. Total sales YOY %

From quick measure

1. Growth rate in each year (I applied the measure over years)

Growth rate in 2012 =

VAR sales2011 = CALCULATE(SUM(InternetSale[UnitPrice]),InternetSale[Year]=2011)

VAR sales2012 = CALCULATE(SUM(InternetSale[UnitPrice]),InternetSale[Year]=2012)

VAR sales2013 = CALCULATE(SUM(InternetSale[UnitPrice]),InternetSale[Year]=2013)

VAR sales2014 = CALCULATE(SUM(InternetSale[UnitPrice]),InternetSale[Year]=2014)

RETURN DIVIDE(sales2012-sales2011,sales2011)

1. Number of customers

Number of customers = DISTINCTCOUNT(Customer[CustomerKey])

1. Number of orders

Number of orders = DISTINCTCOUNT(InternetSale[SalesOrderNumber])

1. Customer age calculation:

From Power query- add column tab – date section- select age. The age column is inserted as a duration. From transform the duration is changed to total years to turn it to age with years then the data was rounded down to get the whole number without friction.

I made the report of 4 pages that visualize

1. Sales over years
2. Product analysis
3. Geographic analysis
4. Customer analysis
5. Some insights

From the analysis I can conclude that:

1. Although the number of orders increased over the years from 2011 to 2014 the profit decreased in 2014
2. The highest profit and largest number of orders came from the USA.
3. The total product costs were higher in Australia and Germany.
4. We purchased larger number of orders for Professionals.
5. Our customers from both genders are nearly the same.
6. Bikes are most purchased in the store.