**Power BI**

*Q. What is the difference between Data Analyst and Business Analyst?*

|  |  |
| --- | --- |
| Data Analyst | Business Analyst |
| * It analyses data and find insights. | Work on insights from data analyst and form strategies |
| * They generally work with data and has good technical skills (SQL, Statistics, Power Bi, etc) | They generally have good problem solving, critical thinking, communication skills, analytical thinking skill. |
| * Work flow – gather data, clean, analyse, and insights | Work flow – understand the problem, data collection, decision based on insights |

Power Bi Components:

1. Power BI Desktop

2. Power BI Services

3. Power BI Mobile

*What we can Do?*

**Power BI Desktop: *Note: Only works on Windows.***

1. Dashboard
2. Load, Clean
3. Graphs
4. Visualization.

**Power BI Services:**

* Online cloud and cloud services
* All data insights can be accessed by all teams and key members in the project.
* Report Sharing
* Self Service BI, Data Refresh
* Security.

**Power Bi Mobile:**

* View and interactions between DA and BA and all C level people.
* Notifications.
* Only viewing.

*Q. Why do people jump from Excel to power BI?*

**Ans:**

Initially data size was manageable by Excel but as data sizes got exceeding excel was not able to handle it.

So, people start using Data bases to store such huge data and other options also which were available in the market. And when we want to apply data analysis for such huge data, we need tool like Power BI.

Hence many people start jumping from excel to Power BI.

**Power BI Desktop UI:**

* Report View**:** All reporting slicing, visualization happens in this area,
* Tabel View: To see how data look like
* Model View: To know and make the relationship between datasets
* DAX**:** M language

Power BI Data Flow:

**Data Collection** **Power Bi Desktop** **Power BI Services** **Client**

Create visuals Data Refresh Mobile

Data Formatting Row Level Security

Data Cleaning Dashboards

DAX, M Language Reports

Power Query Self-service BI

**Data Collection:**

**3 Form of data:**

* Structured data – The data which is available in row and column format or tabular format.
* Unstructured data – The data which is not in any form, like text data, video, image etc,
* Semi-structured data – The data which has both structured and unstructured data. (xml, Json)

**Data collected from following sources:**

* Database, Datawarehouse, DataMart’s, API’s using ETL and web scraping processes.
* For data processing and management two methods used are **OLAP** and **OLTP**.

**Databases:**

* Structured collection of data generally in tabular format.
* It only stores current data.
* It has transactional data (day to day data) or frequent data with basic query like CRUD
* CRUD -
* **C -** Create
* **R -** Read
* **U -** Update
* **D** - Delete
* So, data is stored in structured format with basic query in databases.

**Data Warehouses:**

It is a centralised repository for storing data from multiple sources. It has current and historical data.

These data warehouses are usually used for analysis, business intelligence and strategy’s purpose.

It is designed to handle complex queries like aggregation, joints, and mergers.

**Data Matrix’s:**

It is a subset of data warehouse that store the information about particular subject.

After filtering bigdata we form the data marts and re-arranging as per particular subject.

**Data Collection Methods:**

1. ETL – **E** – Extract, **T** – Transform, **L** – Load
2. Web Scrapping
3. IOT – Internet of the things like data available from sensors devices smart watches etc.

**Data Processing and Management Processes:**

**OLTP** –

**O**nline **T**ransactional **P**rocessing -

It stores transactional data with simple queries like CRUD.

Prioritizes low-latency response which is very quick.

Stores current and operational data only.

**OLAP** –

**O**nline **A**pplication **P**rocessing

This method used for complex queries and multi-dimensional analysis which is very useful for business intelligence and data analysis.

This analysis is performed very easily with complex queries like aggregation, logic, calculation, etc.

It holds current and historical data as well.

Since it performs complex queries, its response is very slow.

Power BI Dashboard for Global Superstore retail shop

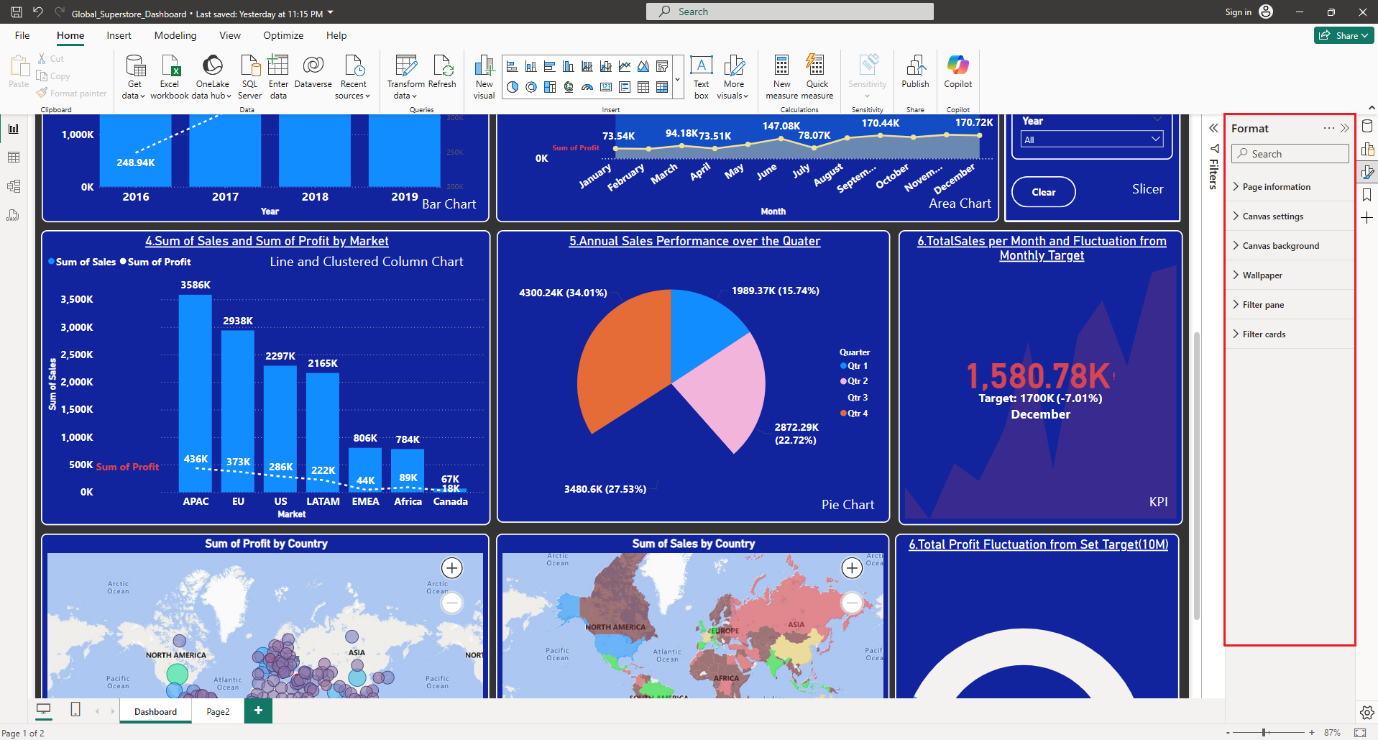
**Power BI UI**

In Visualization we have two Options

* Page Information.
* Format Report Page

In Format Report page we have

* Page Information
* Canvas Setting
* Canvas Background
* Wallpaper
* Filter pane
* Filter cards



Selection of plot is based on how data being showed and how much visual is convincing the data.

In dashboard there should be 8 visuals and should be only one table if needed.

**Visuals Used so far:**

Bar chart, Area chart, line and clustered column chart, pie chart

KPI – For two quantities measured with target value or fluctuations from target value.

New Variable – To set target value [Monthly Sale Target] = 1.7 million, etc

Bubble Chart – For Global Distribution

Filled Map Chart – For Global Distribution

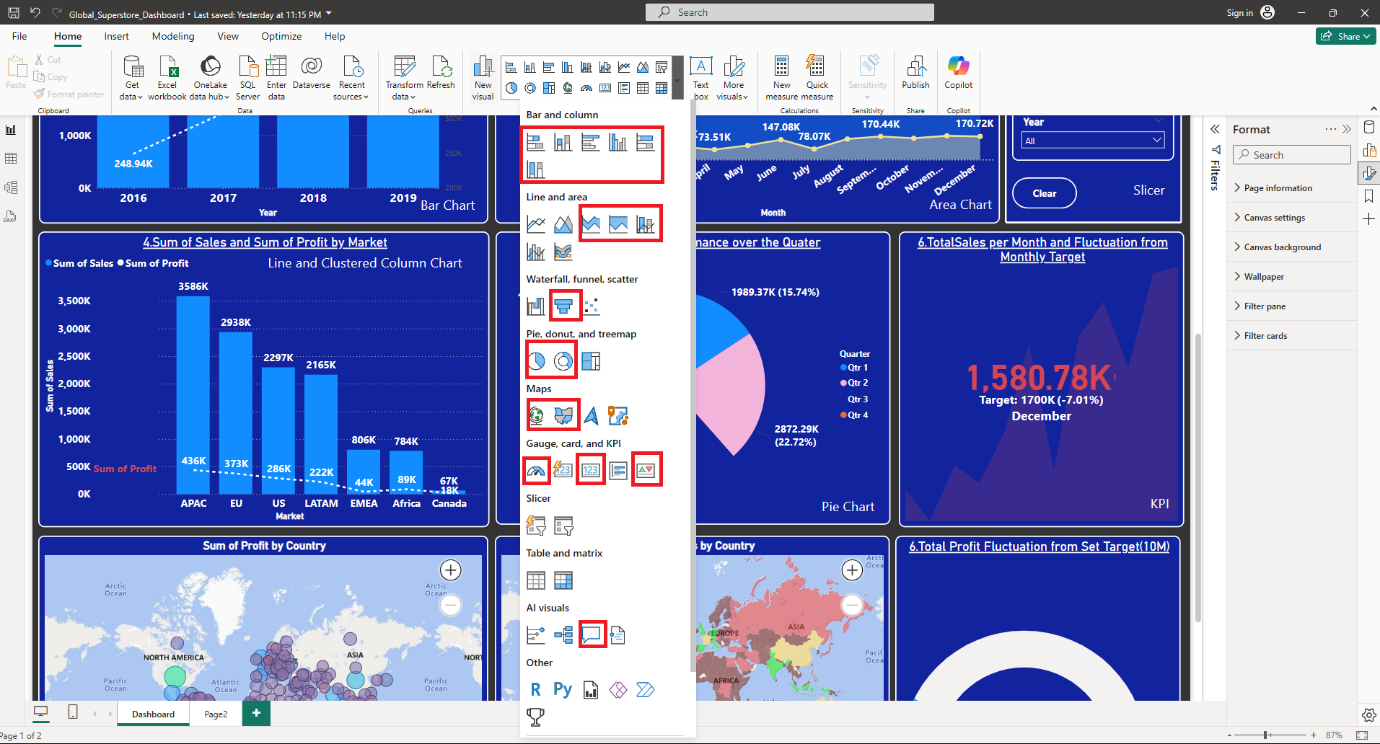
Slicers – For Selecting different segments or parameters from data.

Guage – When any parameter is compared with target value without any other axis then gauge is used.

Bookmark – Used to store screenshot on any particular parameter selection and its visual.

Q&A Button – to ask any questions to AI.

Clear all section button – to clear all selections.



Funnel Chart – when any category is need to be compared with previous and next category then funnel chart is used.

Tooltip or Small Multiples - We can measure four parameters on two axes, only by using Legend or small multiple option.

Grid Visuals – These visuals are used whenever one parameter is needs to be sub-divided based on different parameters like Sum of Sale needs to be sub-divided based on region or category etc.

There are two type of grid visuals

1. Table
2. Matrix

1.Table – When one parameter needs to be subdivided then table is used. In table we have to provide only one parameter that is column.

2. Matrix – Matrix is used when we need pivot form of table. In matrix we have to provide three parameters row. column, and values.

**Interview Question**

*Q. When on main graphs filter has applied, same filter will get applied to tooltip also?*

Ans- yes

**Filters**

In Power BI there are 3 types of filters

1. Filter on this page – Current page will get filtered
2. Filter on all pages – All pages will get filtered
3. Filter on this visual – Only particular visual will get filtered.

***Note: All type has 2 options 1. Advance Filtering, 2. Basic Filtering.***

*Q. Why there is need of slicer and filter? When we have slicer then why do we need filter and vice-versa?*

Ans -

**Slicer** -

* Slicer is type of visual which is available through insert option.
* It is usually can be seen on working pane.
* It allows us to do basic selection and deselection which is not based on condition or couple of parameters.
* Slicers have interaction with all visuals initially but it can be disconnected.
* It has lower applicability and performance.

**Filters** -

* Filter is type of functionality it is applied on visuals and available through view pane.
* It allows us to logical selection which is based on couple of conditions or parameters.
* Filters can be applied on one page, all pages, or selected visual but its interaction with other visuals cannot be disconnected.
* It has higher applicability and performance.

***Note: Filter gets applied first rather than slicer.***

**Data Modelling:**

It is basically creating relationship between two data tables so that they can communicate among themselves or we will have required information in one table.

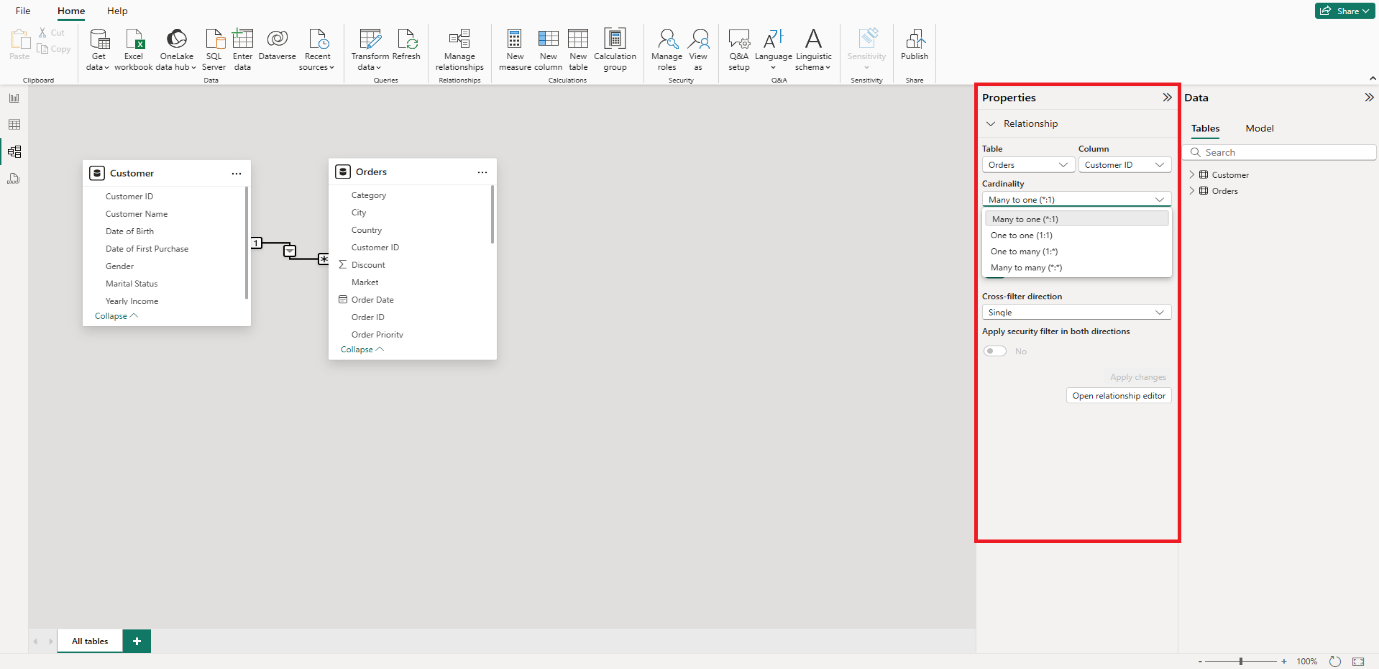
It is used whenever given information is not incomplete but it is scattered if different data tables.

So, in this case we have to create the relationship between two tables to get required information in one table. This relation is always based on common factor between two table. (common column)

**Cardinality:**

While connecting two tables we have one option that is cardinality. Cardinality is nothing but how row of one table is connected to another table. There are four options of cardinality

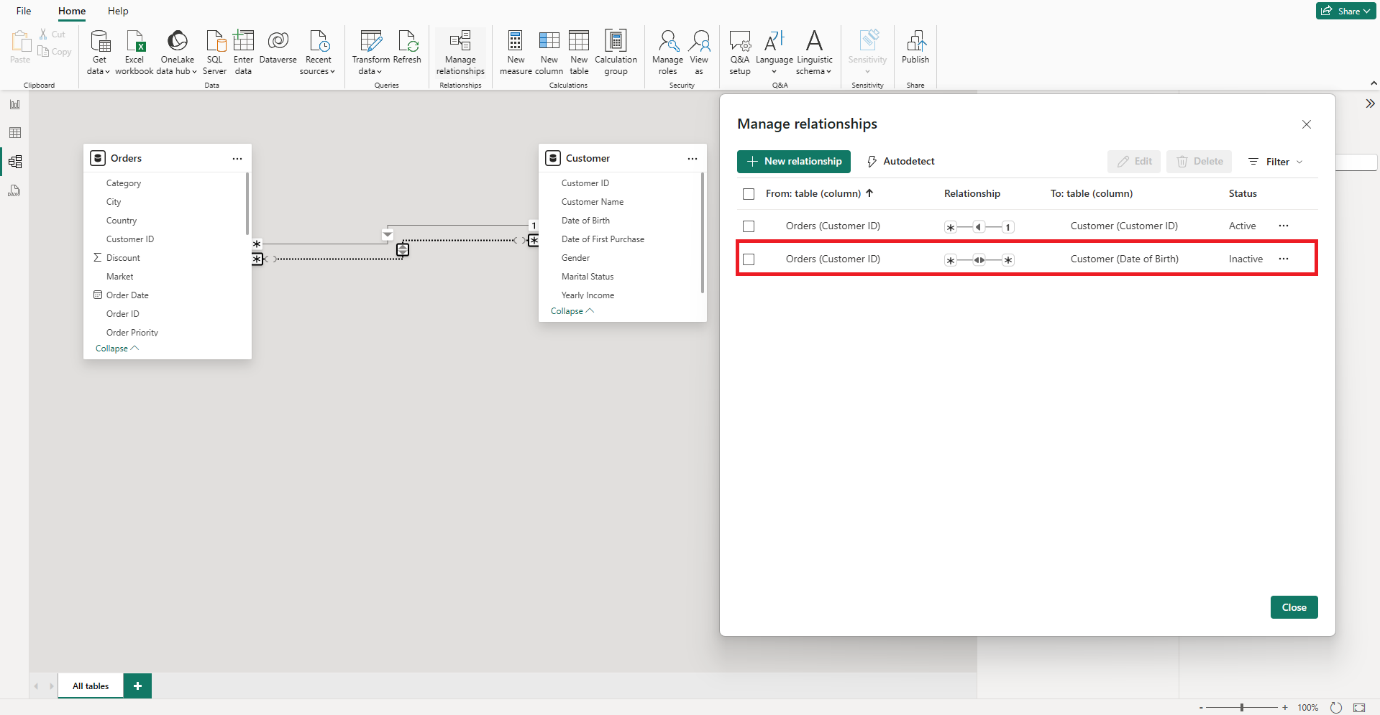
1. Many to one (\*:1) – one row(table-1) to any one row(table-2) [Adhar card – Pan card]
2. One to one (1:1) - one row(table-1) to many rows (table-2) [student - subject]
3. One to many (1: \*) – many rows(table-1) to one row(table-2) [child - father]
4. Many to many (\*: \*) - many rows(table-1) to 1st one row(table-2) and vice versa. [students and subject chosen]



*Q. Can we have multiple relations between two tables?*

Ans

Yes, but only one relationship will be active at a time. The relationship with dotted line is inactive relationship as shown in below snap. Solid one is active.



Most of the time Power BI detects the cardinality by its own.

**Cross-Filter Direction:**

Cross Filter direction allows us to decide the direction of filter, i.e. selection from one table will affect both table or one table will be decided by Cross -Filter Direction.

If you select one option from table-1 then accordingly 2nd table will get affected and vice versa.

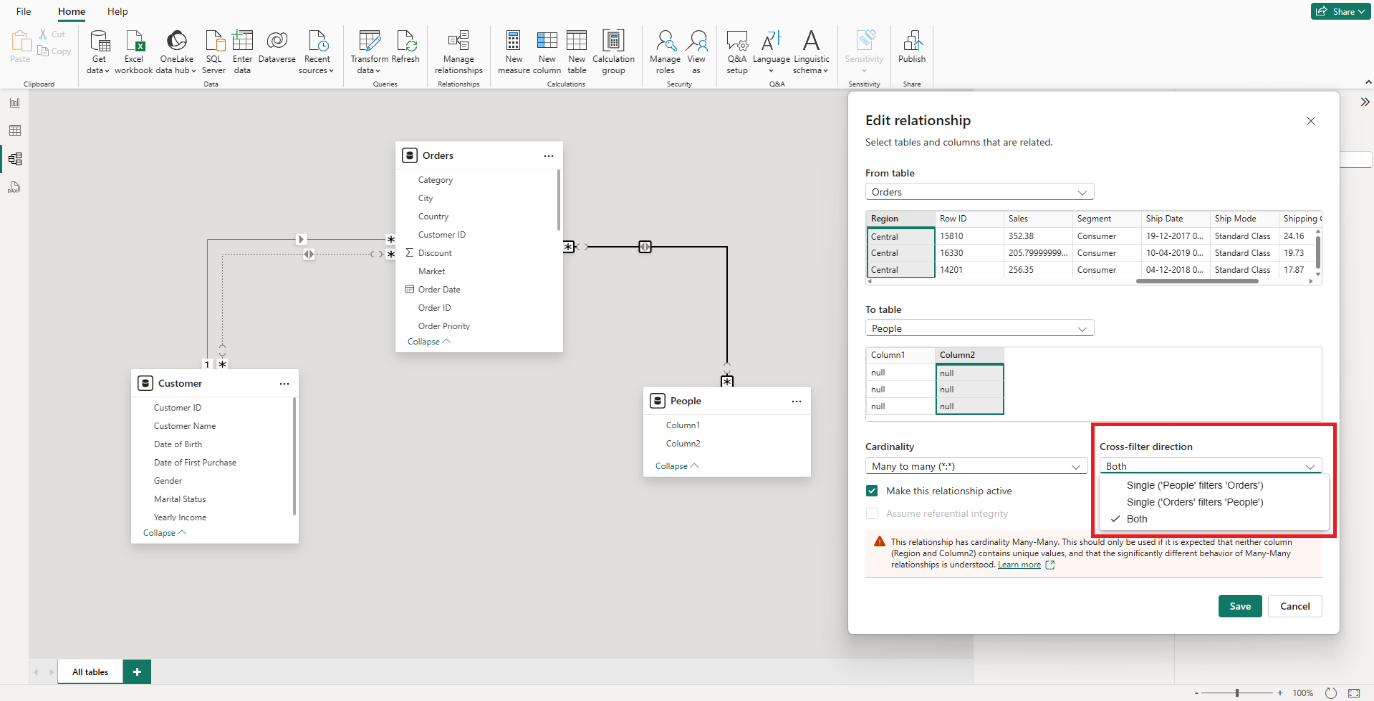
There are three options in cross filter direction, viz table-1

To table-2, table-2 to table1, and both.

Table-1 to Table-2 – based on selection in Table-1 only, Table-2 will get affected.

Table-2 to Table-1 – based on selection in Table-2 only, Table-1 will get affected.

Both – Both tables get affected by any one selection from both tables.



Introduction to DAX (Data Analysis Expression):

It is coding language of Power Bi

**Link -** [**DAX Documentation**](https://learn.microsoft.com/en-us/dax/)

**Mail Link -** [**Data Science\Projects\Power BI Visualization Dashboards\01\_Global Superstore**](file:///C:\Users\tarwa\AppData\Roaming\Microsoft\Word\Data%20Science\Projects\Power%20BI%20Visualization%20Dashboards\01_Global%20Superstore)

**Basic DAX**

**Extract Year/Month/Day/Day of week:**

Month = MONTH (Orders [Order Date]) It will give us the number (1 to 12) as month (January to December)

Year = YEAR (Orders [Order Date]) To extract year

Date = DAY (Orders [Order Date]) To extract day of month (1 to 31)

Quarter = QUARTER (Orders [Order Date]) It will give us the Quarter (1 to 4)

Day of Week = WEEKDAY (Orders [Order Date], 2) It will give us the day of week (1 to 7)

Sunday = 1 and Saturday = 7 by default, but if we put 2 as <return type> then Monday = 1 and Sunday = 7.

***Note: [Order Date] column should be in Date/time type data format.***



**Format:**

It is basically used to convert particular value in desired format. Like we need day in the format of sun, Mon, Tue or month in January, February March, format.

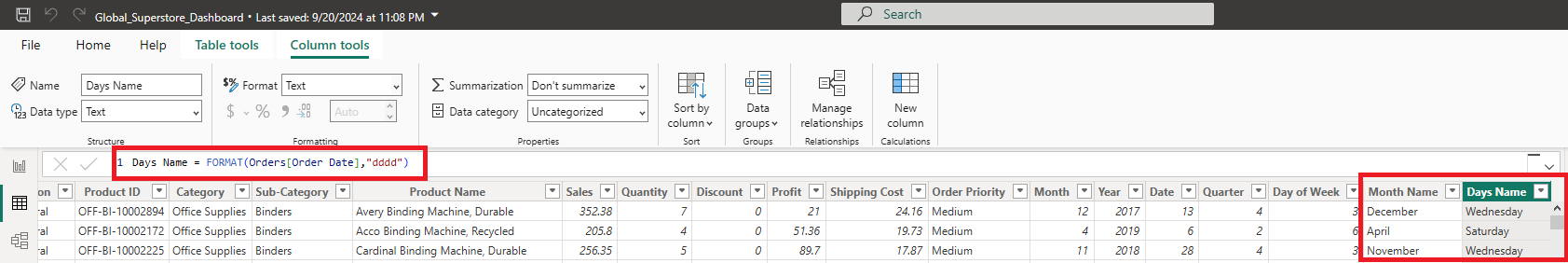
We can convert three values using Format DAX and those are Numeric, Text and Date.

Month Name = FORMAT (Orders [Order Date],”mmmm”)

It will give us month name as January to December. “mmm” will give us month name in short form Jan to Dec.

Days Name = FORMAT (Orders [Order Date],”dddd”)

It will give us days name as Sunday to Saturday. “ddd” will give us the days name in short form Sun to Sat.



CONTRY = UPPER (Orders [Country]) it converts each alphabet to upper case.

Left = LEFT (Orders [Region],3) It will give us 3 characters from left of word ‘Central’

Right = RIGHT (Orders [Region],3) It will give us 3 characters from right of word ‘Central’

Mid = MID (Orders [Region],1,3) It will give us characters starting from 1st position till 3rd characters from 1st position of word ‘Central’. 1- starting character, 3 – no. of characters including 1st one.

Business Problems based on basic DAX (Mail 5)

***Q. Client wants Sale performance on weekdays vs Weekend so we will use switch DAX***

**Switch:**

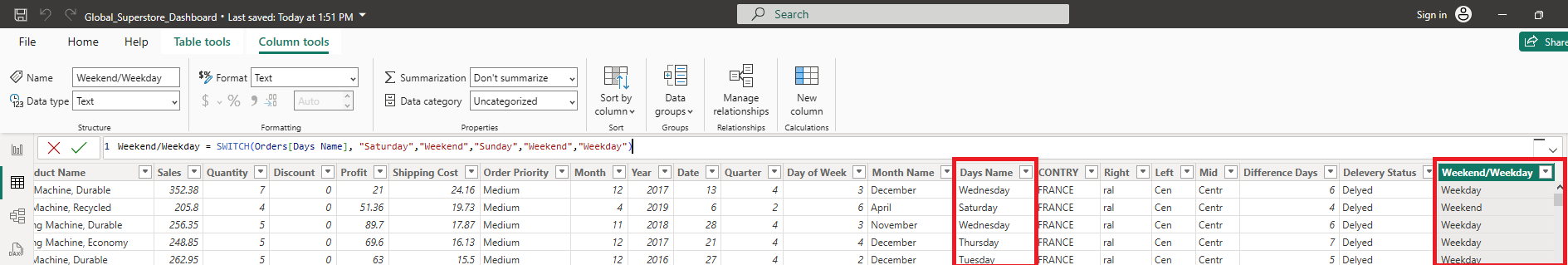
Weekend/Weekday = SWITCH (Orders [Day Name],

“Saturday”, “Weekend”,

“Sunday”, “Weekend”,

“Weekday” )

So basic syntax of switch is SWITCH (Expression/test, Value1, Result1, Value2, Result2, Else if test fails)



**Another Approach:**

Weekend/Weekday

var days = FORMAT (Orders [Order Date], ”dddd”)

return

if (days in {“Saturday”, “Sunday”},” Weekend”, “Weekday”)

Whenever you use var you must have to use return, otherwise value won’t get displayed/print.

***Q. Client wants sales analysis based on financial month and quarter.***

So, Financial years start with April and ends at March (1st April to 31st March)

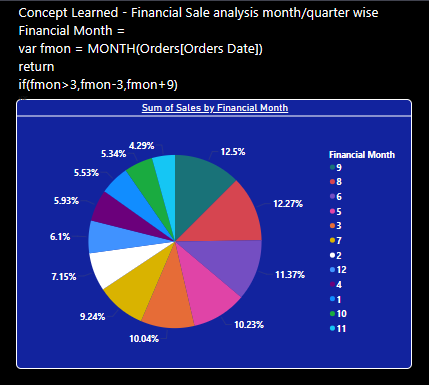
Logic will month>3, month-3 else month+9

**Financial Month =**

var fmon = MONTH (Orders [Order Date])

return

if(fmon>3, fmon-3, fmon+9)

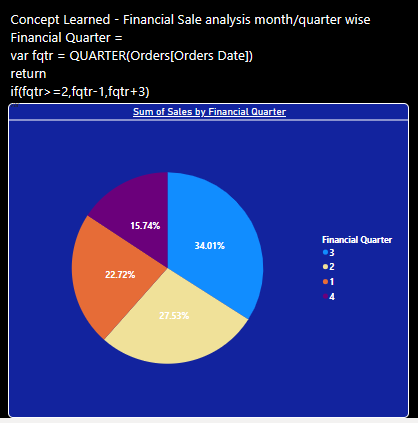


**Financial Quarter =**

var fqtr = QUARTER (Orders [Order Date])

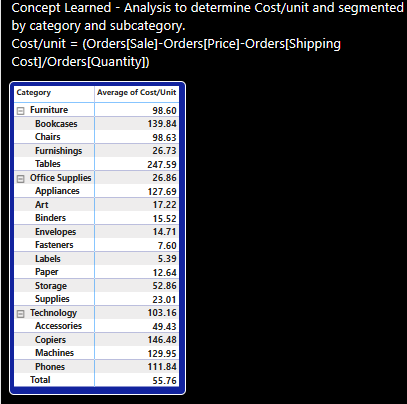
return

if(fqtr>=2, fqtr-1, fqtr+3)



***Q. Concept Learned - Analysis to determine Cost/unit and segmented by category and subcategory.***

Cost/unit = (Orders [Sale]-Orders [Price]-Orders [Shipping Cost]/Orders [Quantity])



***Q. Concept Learned - client wants cluster of consumers in different age group then total consumers category wise and contribution of each category in total consumers.***

Category =

var Age = DATEDIFF (Customer [Date of Birth], DATE (2019,12,31), YEAR)

return

    SWITCH (TRUE (),

            Age<=19,"Teen",

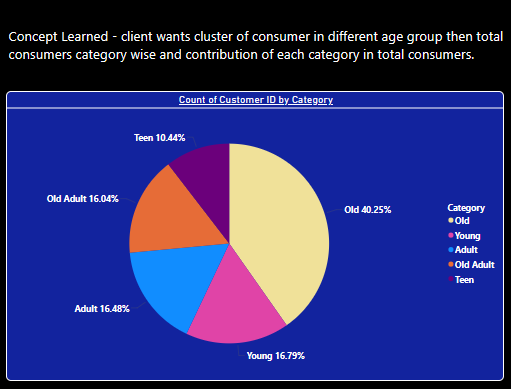
            Age>20 && Age<=30,"Young",

            Age>30 && Age< 40,"Adult",

            Age>40 && Age< 50,"Old Adult",

            "Old"

    )



***Q. Concept Learned - Client wants cluster of consumer income group wise also number of orders place across different regions, age category and income category.***

Age Category =

var Age = DATEDIFF (Customer [Date of Birth], DATE (2019,12,31), YEAR)

return

    SWITCH (TRUE (),

            Age<=19,"Teen",

            Age>20 && Age<=30,"Young",

            Age>30 && Age< 40,"Adult",

            Age>40 && Age< 50,"Old Adult",

            "Old")

Income Group =

var grp = Customer [Yearly Income]

return

SWITCH (TRUE (),

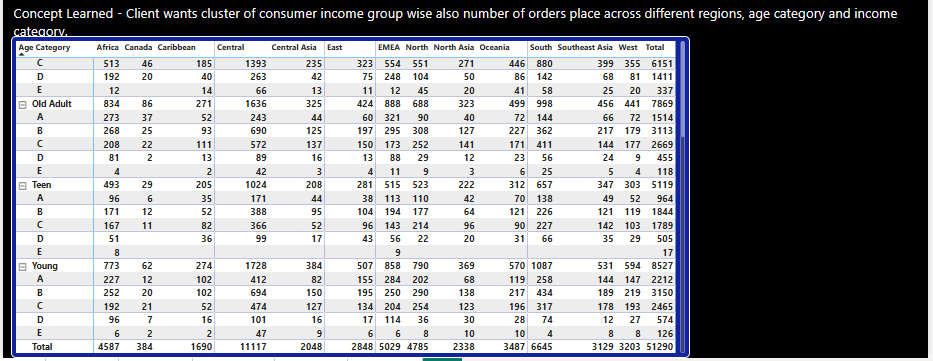
        grp<=30000,"A",

        grp>30000 && grp<=60000,"B",

        grp>60000 && grp<=100000,"C",

        grp>100000 && grp<=150000,"D","E"

)



***Note: Whenever you don’t have any expression in SWITCH () DAX, then you can give TRUE () as default expression****.*

SWITCH (TRUE (), test1, result1, test2, result2)

Whenever you have to evaluate multiple conditions then go for SWITCH ()

**Q. Client wants sale between 15th April 2019 to 15th November 2019.**

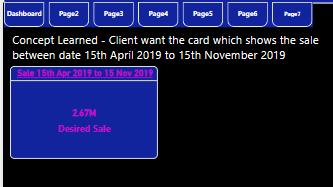
Here we are using DAX CALCULATE.

CALCULATE = CALCULATE (Expression, FILTER ())

FILTER = FILTER (Table, Filter expression)

E.g.

Desired Sale = CALCULATE (sum(Orders[Sales]), FILTER(Orders,Orders[Order Date]>DATE(2019, 4, 15)), FILTER(Orders,Orders[Order Date]<DATE(2019, 11, 15)))



**Q. Client wants to value their top 10 customers based on CLV. CLV is Customer Loyalty Value.**

**Formula for CLV**

CLV = (Average Order Value \* Average Profit Margin)

Purchase Frequency

Concept Learned –

Formula for CLV

CLV  =      (Average Order Value \* Average Profit Margin ) / Purchase Frequency

Profit Margin =

var mp = (Orders[Sales])/(Orders[Quantity]\*(1-Orders[Discount]))

return

mp - Orders[Cost/Unit]

CLV =

var  cust = Customer[Customer ID]

var  sale = CALCULATE(sum(Orders[Sales]), FILTER(Orders, Orders[Customer ID] = cust ))

var  quantity = CALCULATE(sum(Orders[Quantity]), FILTER(Orders, Orders[Customer ID] = cust ))

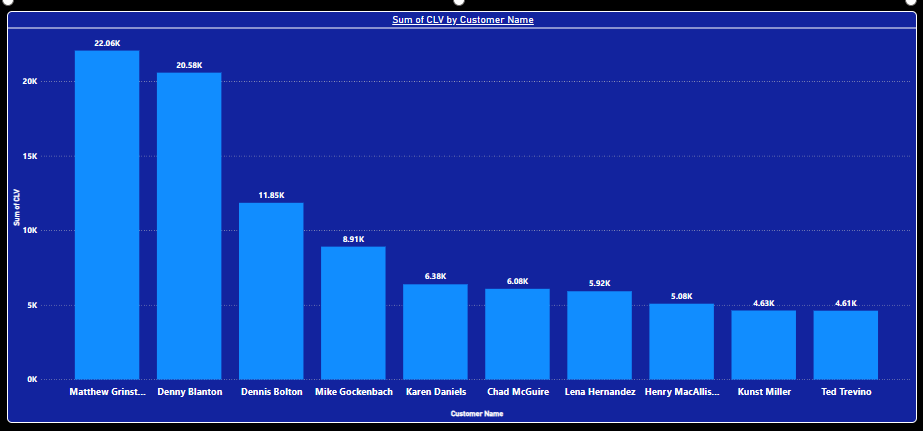
var  profit = CALCULATE(AVERAGE(Orders[Profit Margin]), FILTER(Orders, Orders[Customer ID] = cust ))

var  purchase = CALCULATE(DISTINCTCOUNT(Orders[Order ID]), FILTER(Orders, Orders[Customer ID] = cust ))

var  avg\_order\_value = sale/quantity

return

(avg\_order\_value \* profit)/purchase



***Q. Client wants individual gender which is above and below avg income, and their preference for sub category across different regions***

Concept Learned –

Client wants individual gender which is above and below avg income, and their preference for sub category across different regions

Average =

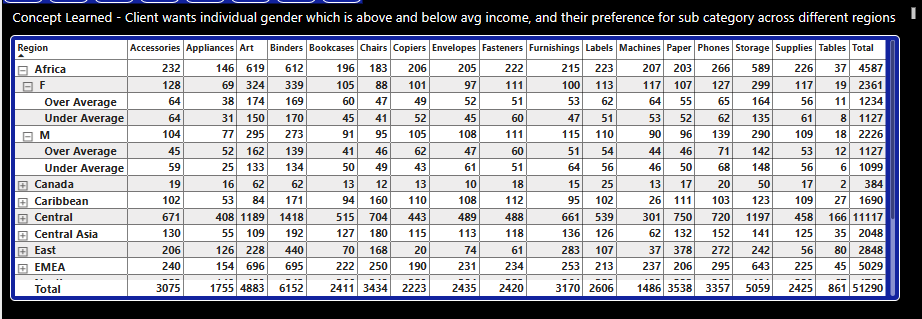
 var male = CALCULATE(AVERAGE(Customer[Yearly Income]), FILTER(Customer,Customer[Gender] = "M"))

 var female = CALCULATE(AVERAGE(Customer[Yearly Income]), FILTER(Customer,Customer[Gender] = "F"))

 return

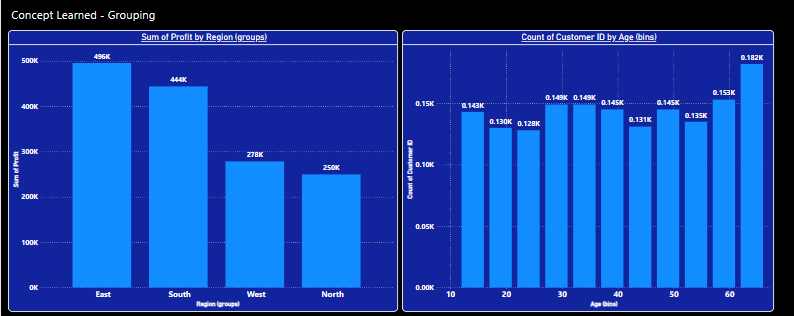
 if(male>Customer[Yearly Income] && Customer[Gender] = "M", "Under Average",

    if(female>Customer[Yearly Income] && Customer[Gender] = "F", "Under Average","Over Average"))



**Grouping:**

Consolidating different categories into minimum possible groups. Grouping the Age as teen, young, adult etc bast on age range like 14 to 20 falls into teen group.



Numeric grouping is not preferred as its loose readability. We can create histogram instead, by connecting to python.

**Measure:**

Measure is nothing but aggregations or calculations done either to build logic or to solve business problems. Most use case DAX are measures.

There are two types of Measures.

1. Implicit Measures: The aggregations which are perform by power bi itself are known as implicit measures. Like count, count (Distinct), max, min, mod etc.
2. Explicit Measures: The measures which are defined by user are known as explicit measures.

***Note: It is usual practice that you have to keep all measures at one place using table***.

So, Measures are aggregations. Also Measures aggregates dynamically based on context of visual.

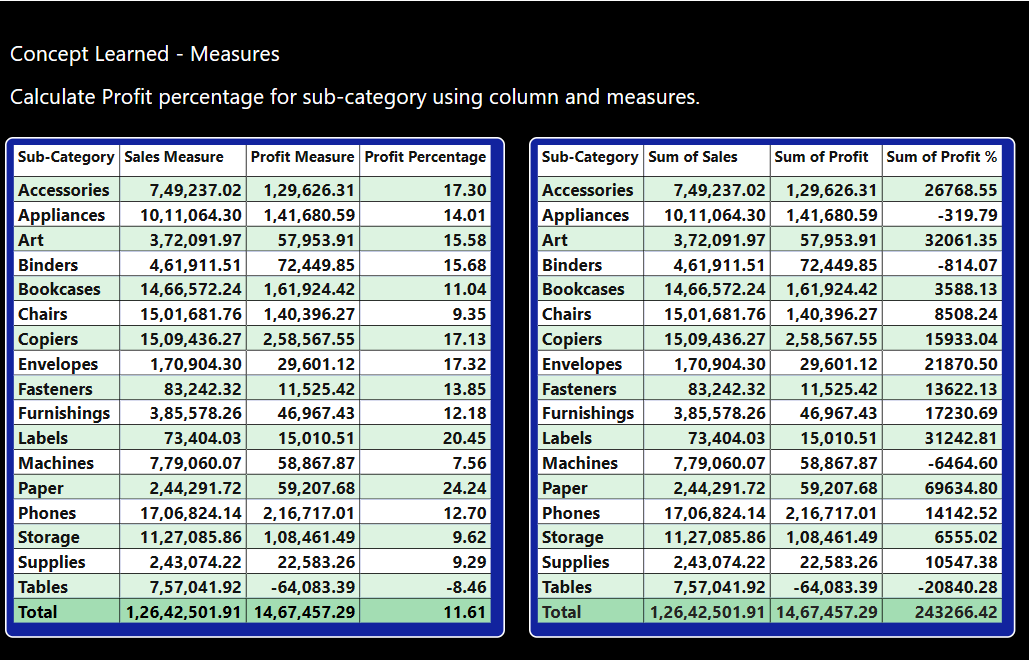
* Columns and measures are created by using DAX so question is which one to use and when?
* Columns are not space saving, means whenever columns are created, they are created permanently.
* Whereas measures are space saving. until and unless measures are called, they are dead they don’t occupy any space.
* Whenever we need row by row values then we have to use columns and when we need aggregations like sum, avgas, mean, mode, etc then we have to use measures.
* Let’s take one problem and that is calculate profit percentage per sub-category.

Formula will be

Profit percentage = (sum of profit/ sum of sale)\*100

So, result is shown in below snap. But same problem will be solved by new columns then result will be different. First table is calculated using measures and second table is calculated using New Column.

So, when new columns are used it will sum the profit % from each category (row by row value) and hence there is difference in results.

****

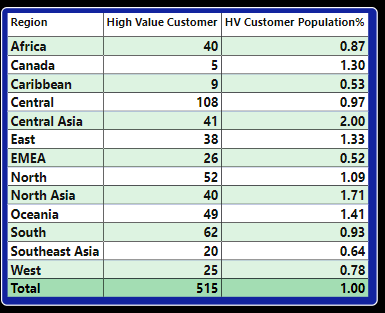
***Q. High Value Customer/region ? and out of total orders of that region how much population is high value customer?***

﻿

Ans

High Value Customer = CALCULATE(COUNT(Orders[Order ID]), FILTER(Orders, Orders[Sales]>1500), FILTER(Orders,Orders[Profit %]>25))

HV Customer Population% = ([High Value Customer]/COUNT(Orders[Order ID]))\*100



***Q Variance between the desired sale of 2019 and desired sale of 2018 for the same date values.***

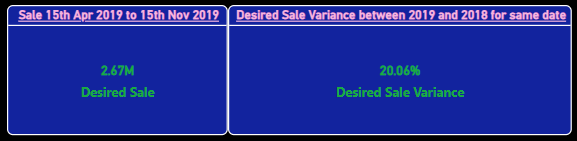
**Ans**

Desired Sale Variance =

var Twenty\_ninteen = CALCULATE(sum(Orders[Sales]), FILTER(Orders,Orders[Order Date]>DATE(2019, 4, 15)), FILTER(Orders,Orders[Order Date]<DATE(2019, 11, 15)))

var Twenty\_eighteen = CALCULATE(sum(Orders[Sales]), FILTER(Orders,Orders[Order Date]>DATE(2018, 4, 15)), FILTER(Orders,Orders[Order Date]<DATE(2018, 11, 15)))

return ((Twenty\_ninteen-Twenty\_eighteen)/(Twenty\_ninteen))\*100



This is the variance of desired sale between 2019 and 2018 for same date value 15Apr to 15Nov

**Concept Learned - Measures**

***Q. Client ask for Avg. Order Value/customer and identification of top 15 customers based on highest order value.***

**Ans**

Avg Order Value = ([Sales Measure]/CALCULATE(DISTINCTCOUNT(Orders[Order ID])))



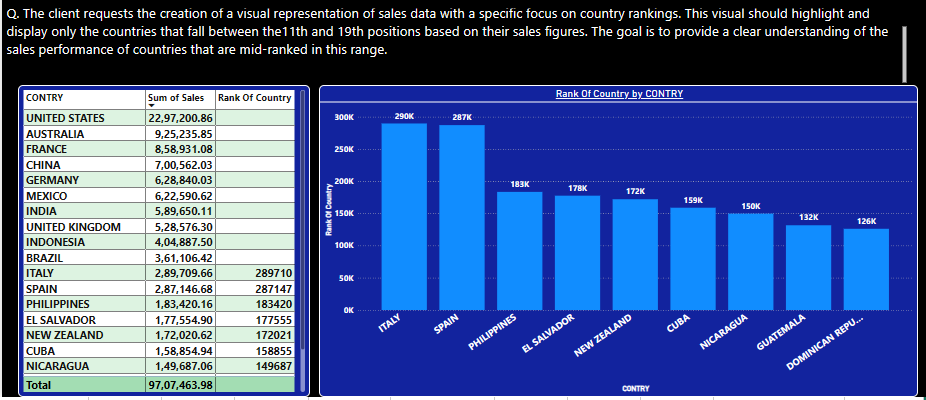
**Concept Learned - Measures**

***Q. The client requests the creation of a visual representation of sales data with a specific focus on country rankings. This visual should highlight and display only the countries that fall between the11th and 19th positions based on their sales figures. The goal is to provide a clear understanding of the sales performance of countries that are mid-ranked in this range*.**

**Ans**

[Sale Measure] = SUM(Orders[Sale])

Rank Of Country = IF(RANKX(all (Orders[CONTRY]),[Sales Measure]) > 10 && RANKX(all (Orders[CONTRY]),[Sales Measure])<20 , [Sales Measure], BLANK())



***Q. The client desires a visual representation of the cumulative sales sum for each year. This visual should clearly illustrate the progressive accumulation of sales over time within each year.***

**Ans**

Cumulative Sum = CALCULATE(sum(Orders[Sales]), Orders[Order Date].[Year]<= SELECTEDVALUE(Orders[Order Date].[Year]))

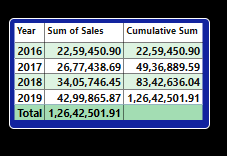
Cumulative sum =

2016 -🡪 2016

2017 -🡪 2016+2017

2018 -🡪 2016+2017+2018

2019 -🡪2016+2017+2018+2019



**Concept Learned – Measure**

***Q. The CEO has requested a report showing the average daily sales amount across all years (2016, 2017,2018, and 2019).***

**Ans**

Average Sales =

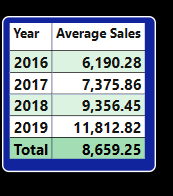
var lowest = min(Orders[Order Date])

var highest = max(Orders[Order Date])

var numdays = DATEDIFF(lowest,highest,DAY)

return

[Sales Measure]/numdays



**Filtered Context:**

Data being displayed is filtered out based on which factors is nothing but Filtered context.

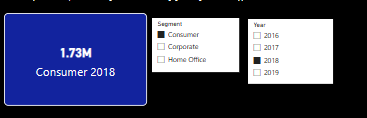
There are three types of Filtered Context

1. Clear Context
2. Fix Context
3. Relative Context

**Fix Context:** The context in which some parameters are fixed, are known as Fixed Context. E.g.

If we want sale of consumer for the year of 2018. Then consumer from segment and 2018 from year are fixed, and based on that we are calculating sale.

Consumer 2018 = CALCULATE(sum(Orders[Sales]),Filter(Orders, Orders[Segment]="Consumer"), Filter(Orders, Orders[Order Date].[Year]= 2018))

****

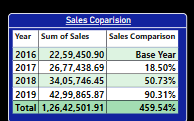
**Relative Context:** When one value is compared with reference to other value then that is relative context. Like client wants to know increase in sales % for each year compared to 2016

**Sales Comparison =**

var Sale\_2016 = CALCULATE([Sales Measure], Orders[Order Date].[Year]= 2016)

return

if(SELECTEDVALUE(Orders[Order Date].[Year])=2016, "Base Year", ([Sales Measure]- Sale\_2016)/Sale\_2016)

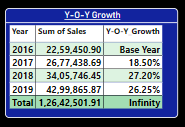
****

**Now Year-on-Year growth**

Y-O-Y\_growth = var previous\_year\_sale = CALCULATE([Sales Measure], Orders[Order Date].[Year] = SELECTEDVALUE(Orders[Order Date].[Year])-1)

return

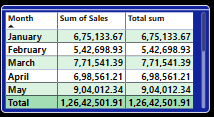
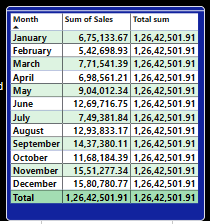
if(SELECTEDVALUE(Orders[Order Date].[Year])=2016, "Base Year",([Sales Measure]-previous\_year\_sale)/previous\_year\_sale)

****

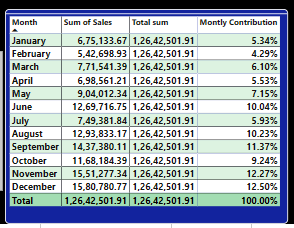
**Clear Context:**

When any value is filtered out based on context, but we want that filter should not get applied and value should get calculated then that is called clear context

So, if client wants month contribution in % to total sale then we want sum of sale and total of sum of sale (1,26,42,501.91) but whenever we put sale in table the sum of sale gets filtered based on month as shown in below snap. Hence, we are going to use ALL DAX to avoid this effect.

 Total sum = CALCULATE(sum(Orders[Sales]), all(Orders)) 

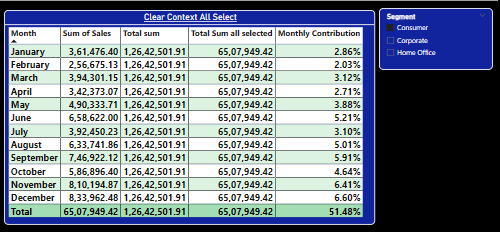
So now we can calculate the Monthly contribution = (Sum of sale/Total Sum)\*100



**All Select:**

All select nullify the filters of the visuals but gets affected by outside slicers (outside the visual).

As we can see consumer from segment is selected and Total Sum all selected column gets affected by it.

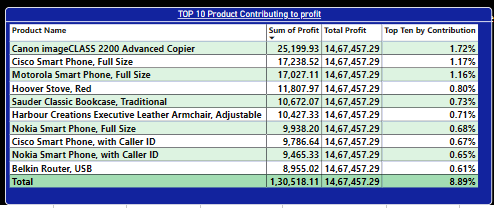


***Q. Analise the company's product portfolio to determine the top 10 products that contribute the most to overall profitability.***

**Ans :**

Total Profit = CALCULATE(sum(Orders[Profit]), all(Orders))

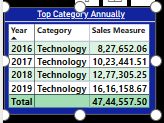
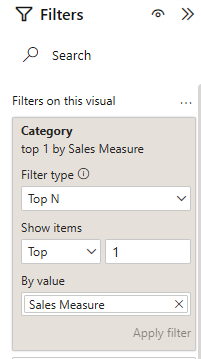
Top Ten by Contribution = [Profit Measure] / [Total Profit] press %



***Q. Determine the leading product category for each year by analyzing annual sales data. This analysis will help identify trends and shifts in consumer preferences, enabling the company to make***

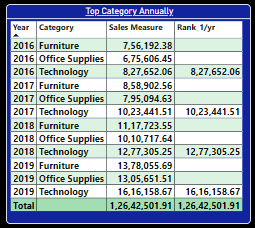
**Ans:**

**1st on by just applying filter on category.**

** **

**Another Approach**

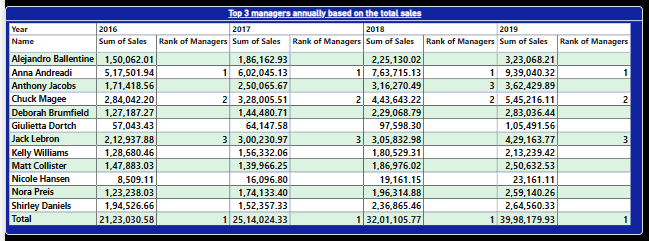
Rank\_1/yr = if (RANKX(all(Orders[Category]),[Sales Measure]) = 1, [Sales Measure], BLANK())



But not sure about first approach….! Because Total are different.

***Q. For each year, identify the top 3 managers annually based on the total sales they generated. This analysis will provide insights into the most effective leaders in driving sales performance, allowing the company to recognize and reward top talent, and replicate successful strategies across different regions***

**Ans** Rank of Managers = if(RANKX(all(People[Name]), [Sales Measure])<4, RANKX(all(People[Name]), [Sales Measure]), BLANK())



***Q. Assess how much the sales value of the top-ranked country deviates from the sales values of other countries within the top 10. This analysis will highlight the performance gap between the leading country and its peers, offering insights into potential areas for growth and improvement in underperforming regions.***

**Ans** Rank of Ctry = RANKX(all(Orders[CONTRY]), [Sales Measure])

**Deviation =**

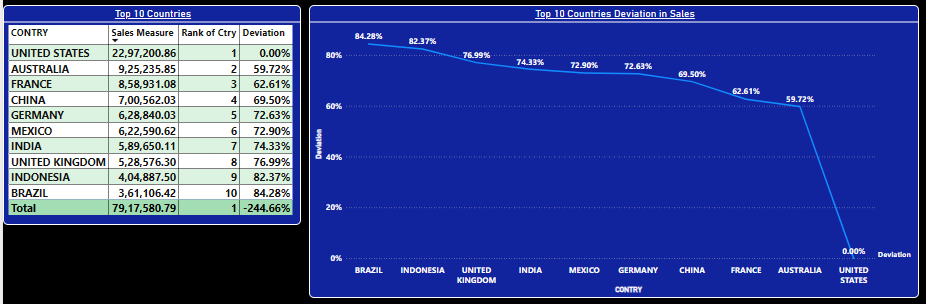
var top1 =CALCULATE(MAXX(TOPN(1, all(Orders[CONTRY]),[Sales Measure],DESC),[Sales Measure]))

var Sale = [Sales Measure]

var SaleDifference = top1-Sale

return

SaleDifference/top1

****

***Q. Evaluate the year-over-year growth in new customer acquisition and the percentage growth in the total customer base. This analysis will help the company understand the effectiveness of its customer acquisition strategies and track the expansion of its customer base over time, providing insights for future marketing and retention efforts.***

**Ans: Customer Base =**

var year = SELECTEDVALUE(Orders[Order Date].[Year])

*-- for current year*

var churned = if(year=2019,

            CALCULATE(COUNTROWS('Retention Table'),FILTER('Retention Table','Retention Table'[Last Purchase Date]<year)),

            CALCULATE(COUNTROWS('Retention Table'),FILTER('Retention Table','Retention Table'[Last Purchase Date]<=year))

            )

var new = CALCULATE(COUNTROWS('Retention Table'), FILTER('Retention Table','Retention Table'[First Purchase Date]<=year))

var current\_base = new - churned

*-- for previous year*

var previous\_year = year-1

var previous\_churned = if(previous\_year=2019,

            CALCULATE(COUNTROWS('Retention Table'),FILTER('Retention Table','Retention Table'[Last Purchase Date]<previous\_year)),

            CALCULATE(COUNTROWS('Retention Table'),FILTER('Retention Table','Retention Table'[Last Purchase Date]<=previous\_year))

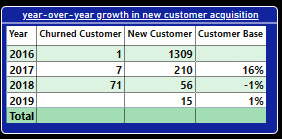
            )

 var previous\_new = CALCULATE(COUNTROWS('Retention Table'), FILTER('Retention Table','Retention Table'[First Purchase Date]<=previous year))

 var previous\_base = previous\_new - previous\_churned

return

if(year = 2016, " ",(current\_base - previous\_base)/ previous\_base)



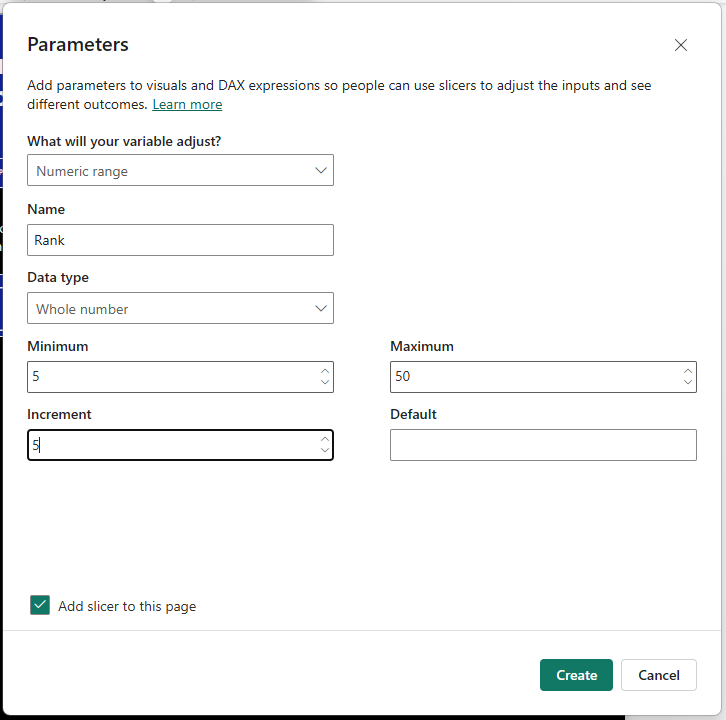
**Concept Learned –**

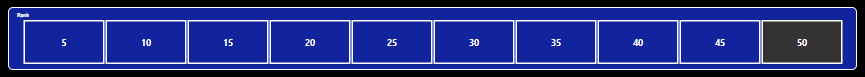
**Parameterized Chart:**

Parameterized charts are nothing but slicers in which parameters are selected by human or users. In Modelling we have option to create parameters.

There are two options under New Parameters Numeric Range and Fields.

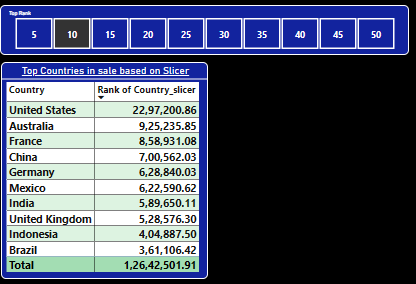
Numeric Range

****

****

***Note: Add slicer must be checked.***

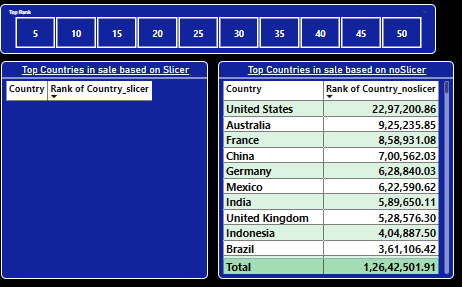
Rank of Country\_slicer = if(RANKX(all(Orders[Country]),[SalesMeasure])<SELECTEDVALUE('Rank'[Rank]), [Sales Measure],BLANK())

****

**If any value is not selected in slicer**

Rank of Country\_noslicer =

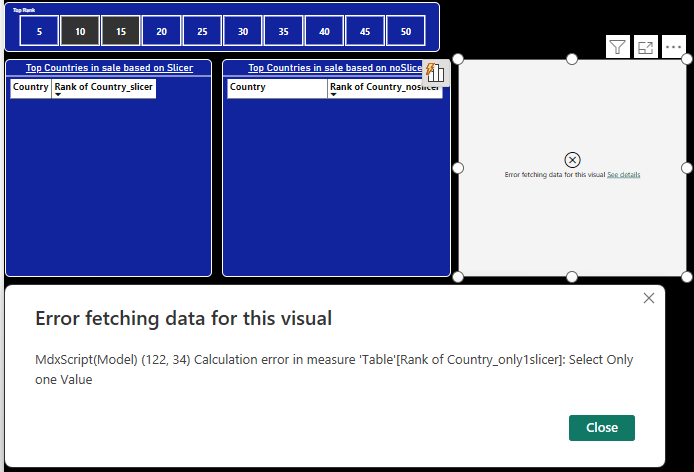
if(ISFILTERED('Rank'[Rank]),if(RANKX(all(Orders[Country]), [Sales Measure])<=SELECTEDVALUE('Rank'[Rank]), [Sales Measure],BLANK()),[Sales Measure])



So, if any value is not selected in slicer, then it’s expected that visual must show all countries ranking base on sale as shown in above visual. Because empty visual doesn’t seem appropriate. So instade of empty visual it must show all countries with sale.

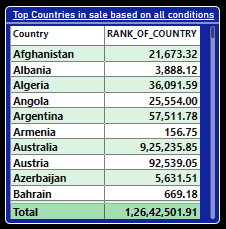
If more than one value is selected from slicer then it should give use the error message and for that we have HASONEFILTER() DAX.

**Rank of Country\_only1slicer =** if(HASONEFILTER('Rank'[Rank]),if(RANKX(all(Orders[Country]), [Sales Measure])<=SELECTEDVALUE('Rank'[Rank]), [Sales Measure],BLANK()),ERROR("Select Only one Value"))

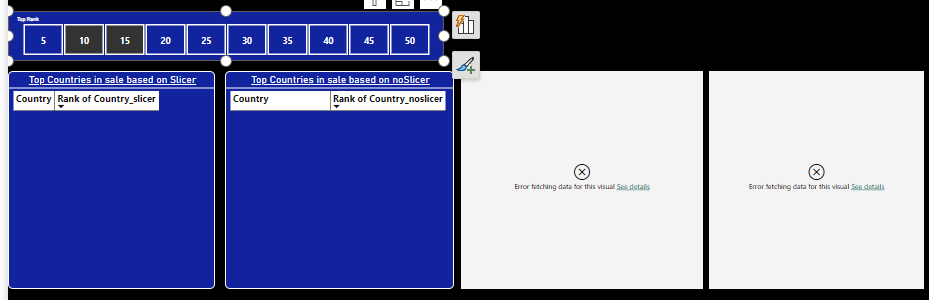


This the combine Dax considering all above conditions**.**

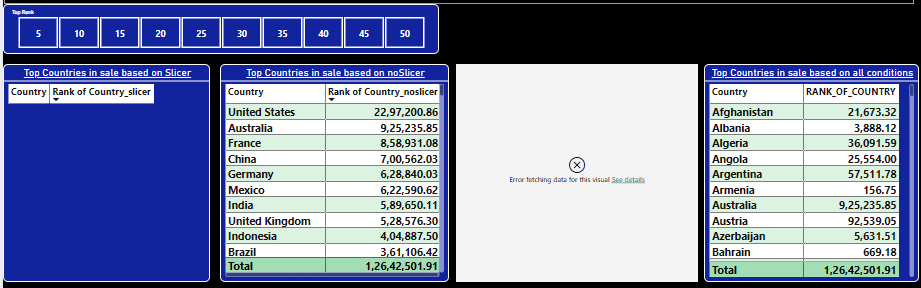
RANK\_OF\_COUNTRY = if(ISFILTERED('Rank'[Rank]), if(HASONEVALUE('Rank'[Rank]), if(RANKX(all(Orders[Country]), [Sales Measure])<=SELECTEDVALUE('Rank'[Rank]), [Sales Measure], BLANK()), ERROR("Select Only one Value.")), [Sales Measure])



When two values selected it showing error message as shown in below snap.



And when no value is selected it must show all countries.

****

***Q. Client wants a visual where he will have authority to select the profit % and only the country that lies in that profit % range should be displayed.***

**Ans**

Profit Percentage = ([Profit Measure]/[Sales Measure])\*100

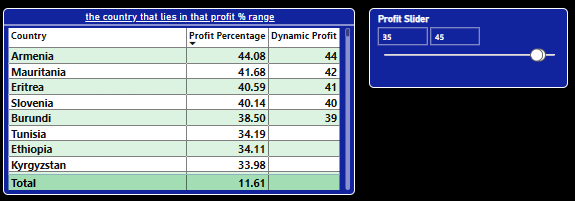
Dynamic Profit =

var minimum = min('Profit Slider'[Profit Slider])

var maximum = max('Profit Slider'[Profit Slider])

return

if ([Profit Percentage]>=minimum&& [Profit Percentage]<=maximum, [Profit Percentage], BLANK())



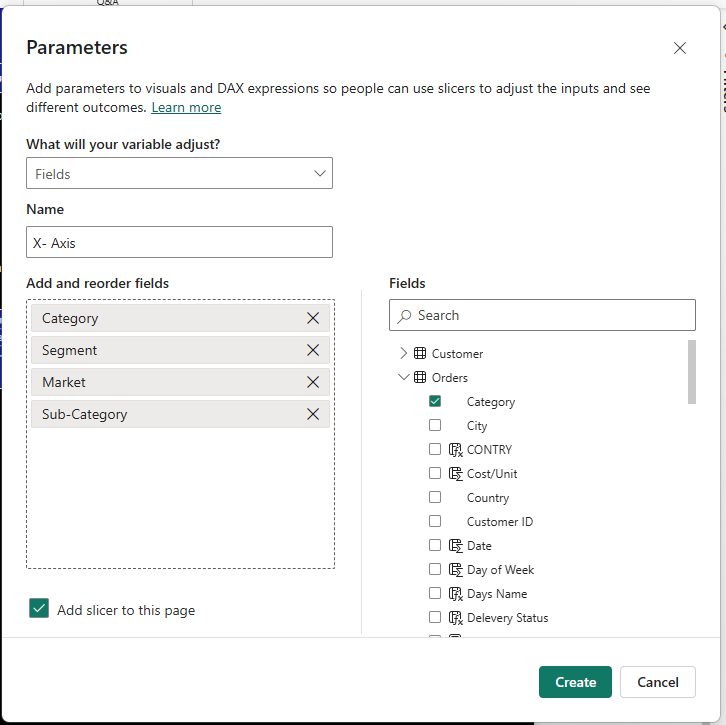
So only those countries are displayed who are fall in profit range of 35 to 45 % profit.

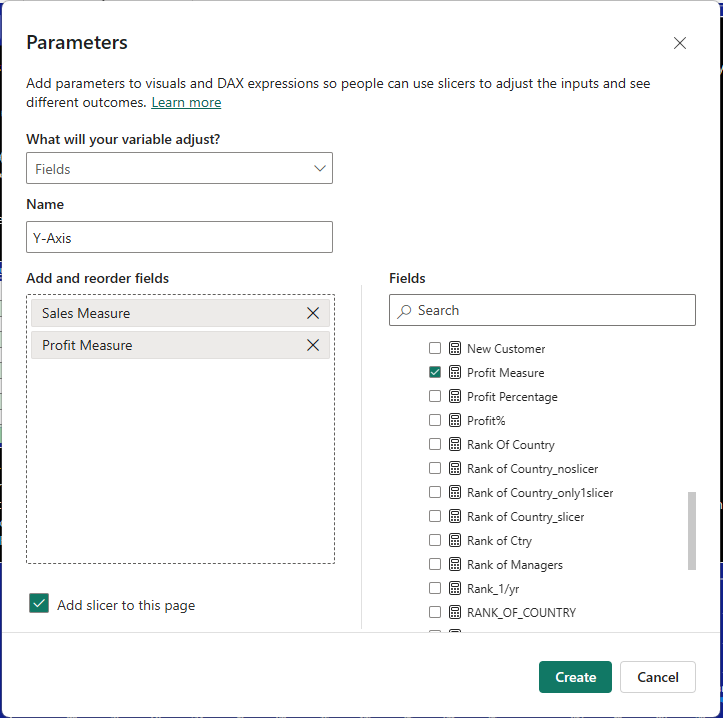
**Field Parameters:**

The parameters which are created based on Column are nothing but Field Parameters

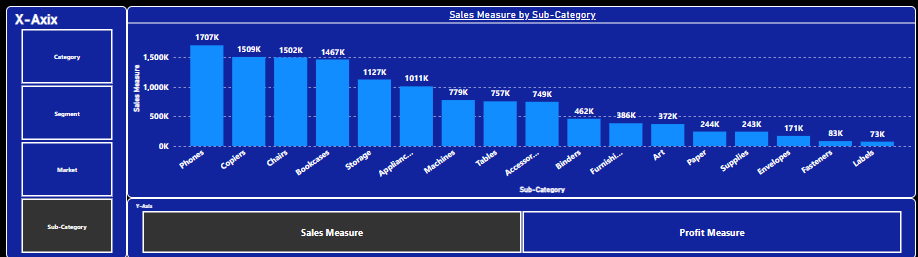
E.g. clients want the control over selection of Category, segment, market, sub-category and client want Sales and profit against this selection so this can be archived by field parametsrs.

We have to select Field from Modelling tab and following dialog-box will appears.

****

****

Select the parameters as we want on X and Y-Axis and then we have blow visual as per selection.

****

***Note: All functions with X like RANKX, SUMX, MAXX etc are iterative function they perform on each row.***

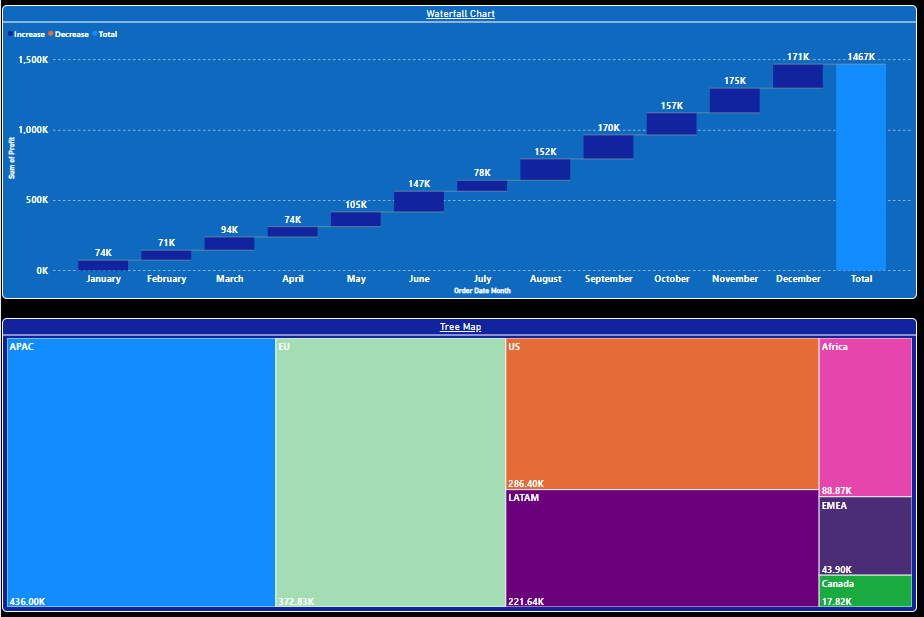
**Waterfall Chart:**

Used to show the contribution of different parameters in one particular Variable. Like Months contribution in total sale of that particular year.

**Tree Map:**

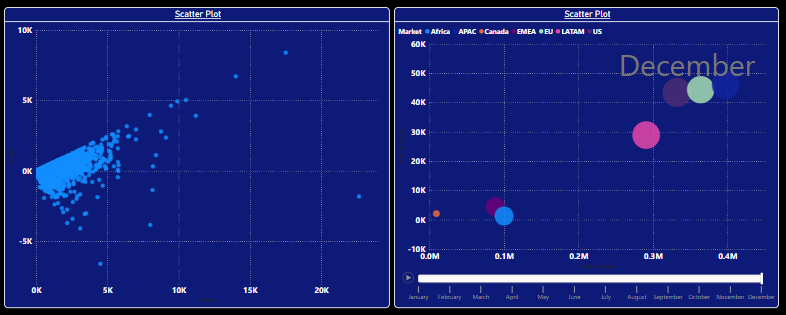
Used to show the contribution of different parameters in one particular Variable. Like Months contribution in total sale of that particular year.

Tree Map never show the contribution of negative values as Waterfall Chart



**Scatter Plot:**

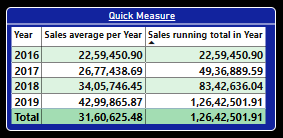
Basically, it used to compare two numerical quantities. Best for time-series analysis

****

**Concept Learned - Quick Measures**

Quick Measures are those measures which are created by power bi itself.

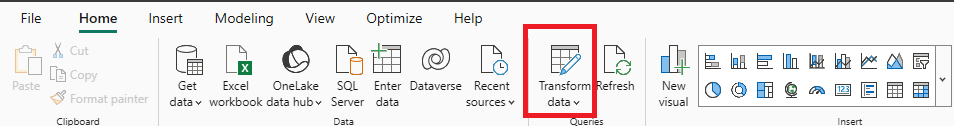
Under Modelling Tab, we have option of Quick Measure.

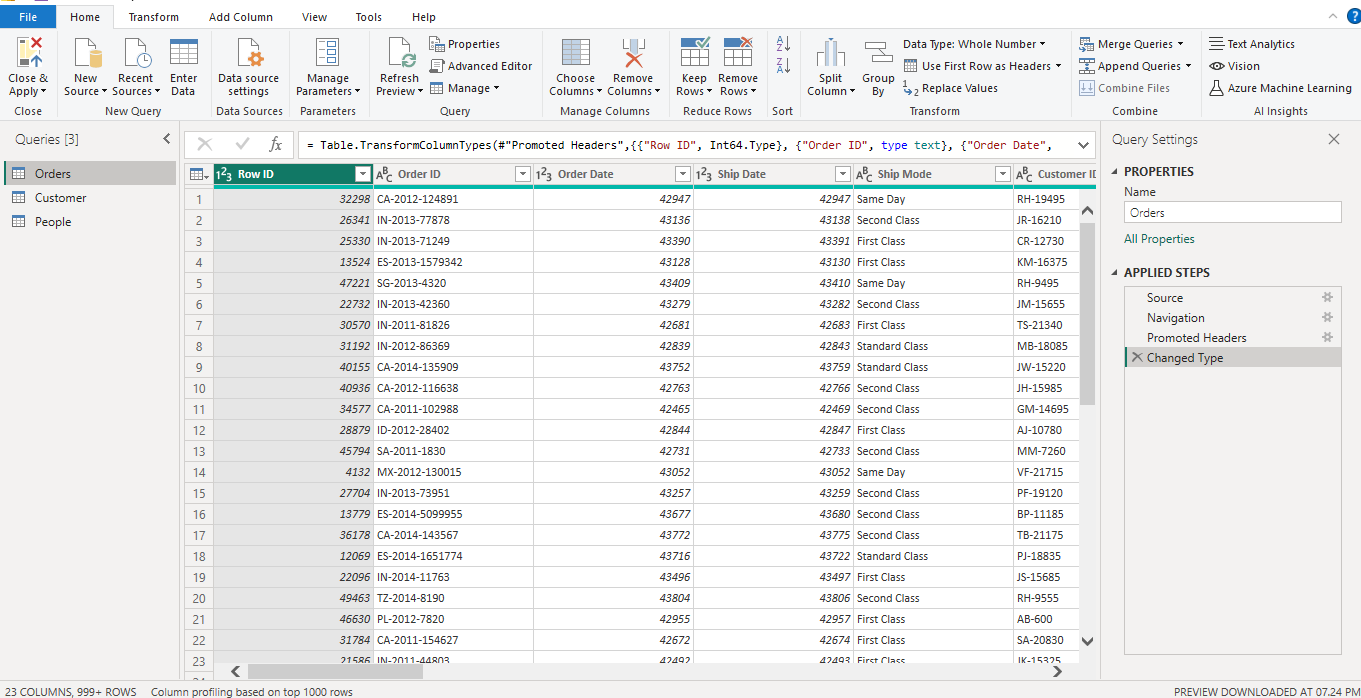


Power Query Introduction:

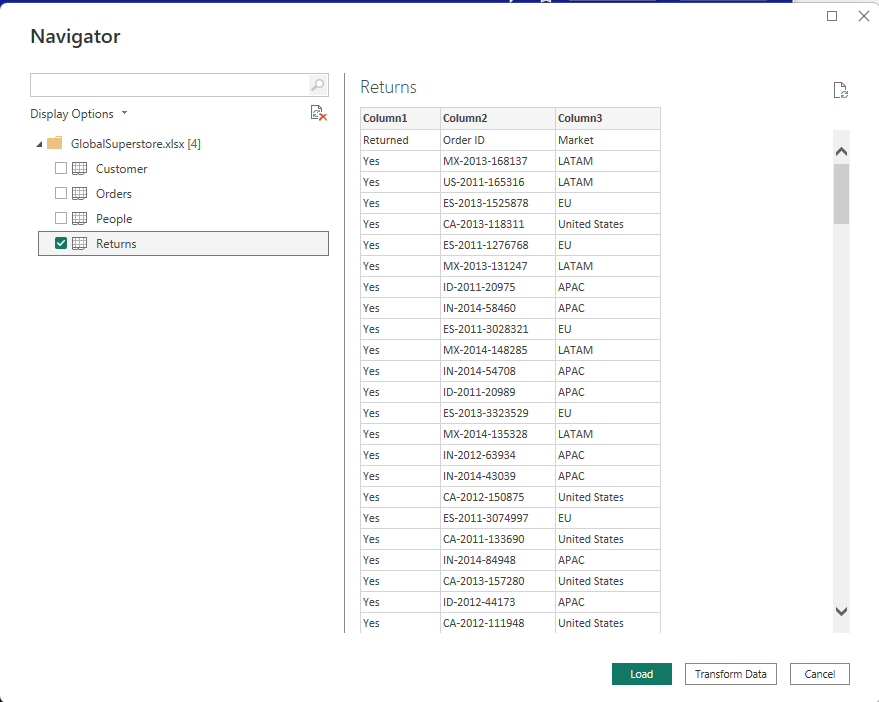
It is dedicated tool for data cleaning, data processing, data transformation etc.

We can see that in Return table columns name become row, so to correct this we use power query transform data option. It opens the Data transform editor

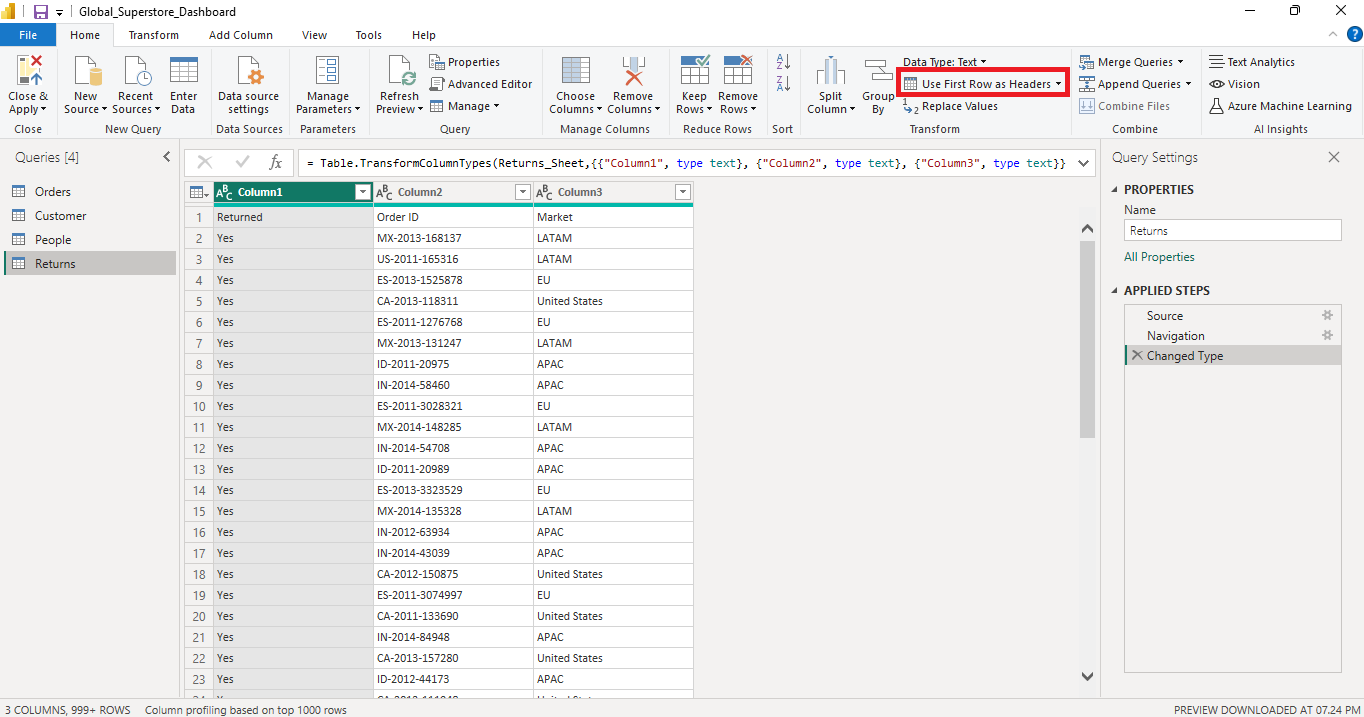


****

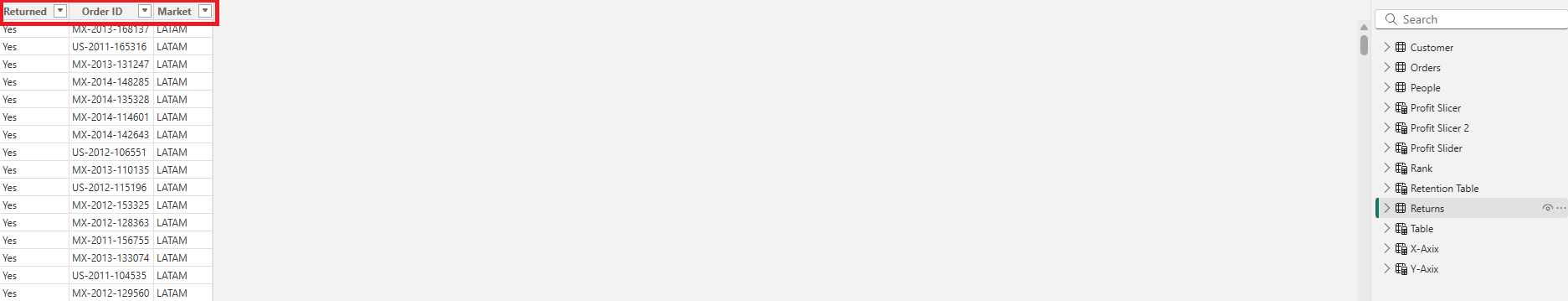
We can see that in Return table columns name become row

****

So instead of load now we have to click on Transform Data button. There are two option to load data into power bi. One you load the data then transform and second in first transform the data and then load.

****

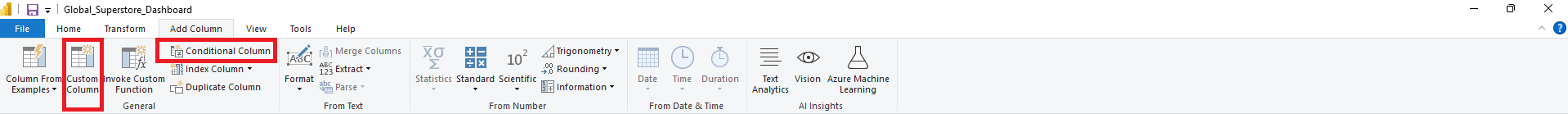
To resolve this problem, we have to click on “Use First Row as header” option as shown in red rectangle**.** And don’t forget to click on Close and Apply option. So that all changes we do in transform data editor will get load in to power bi. And it affects permanently.

****

* Flow of power bi is
* Source -🡪 power query --🡪 power bi desktop --🡪 power bi services --🡪 power bi mobile. Hence the changes perform in power bi desktop will not reflect in to power query.
* **Note: No DAX are used in Power Query. DAX are used in Power BI Desktop only. All codes generated in power query are written in M language.**
* Query Dependency show the source on which the query is depend. Or in other words we say it show the source from which the query is coming from.

**Conditional Columns**

We have two options to create conditional columns that are Conditional Columns and Custom Columns in Add Columns option. Conditional column allows us to create the new column based on some condition. But it only allows condition in 'if else' format. Also, you can’t use conditional operators



**Conditional Columns**

We have two options to create conditional columns that are Conditional Columns and Custom Columns in Add Columns option. Conditional column allows us to create the new column based on some condition. But it only allows condition in 'if else' format. Also, you can't use conditional operators.

**Custom Columns**

In custom columns we can write our own code using M Language and conditions like AND and OR can be used in M language. when multiple conditions need to be evaluated then custom columns are pref**er**

***Q. We can create structural changes like month, year, week, quarter, Days of week etc using DAX and power query so which one should we select?***

***(interview question)***

Ans:

The flow of power Bi is like source to power query to power bi desktop. So whatever structural changes need to be done, do it in the power query. It is the best practice. Because when these changes are done in power query the got created before loading it into power bi desktop.

So, because of this whenever these columns are used to create any visual in power bi desktop and if those visuals are get changed due to some selection like slicer, the formula behind these columns will not be recalculated since it has been run in power query.

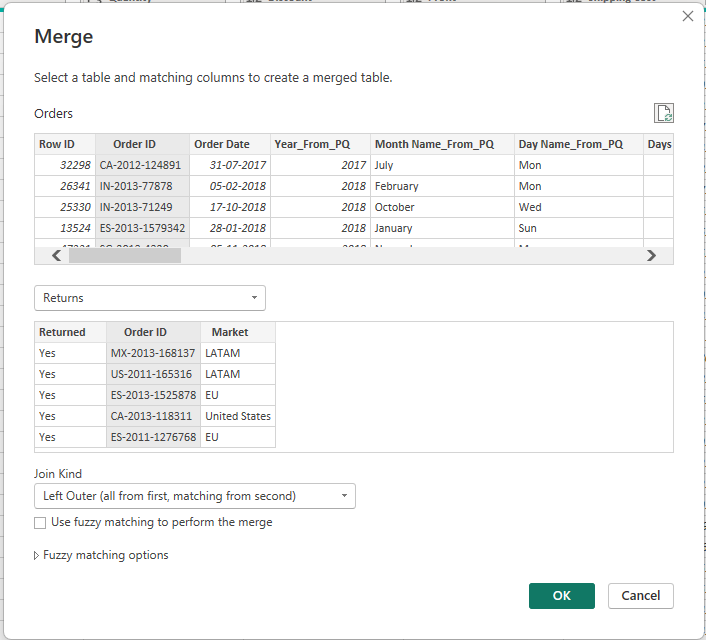
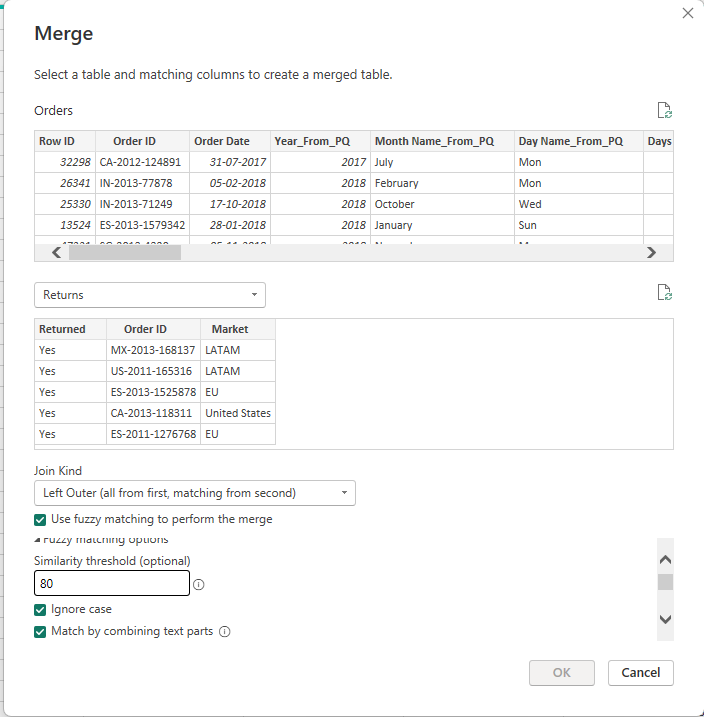
On the other hand, if we create these columns in DAX the **logic behind column gets executed every time** **(because power bi is dynamic)** even if there is refresh **which will consume working space unnecessarily.**

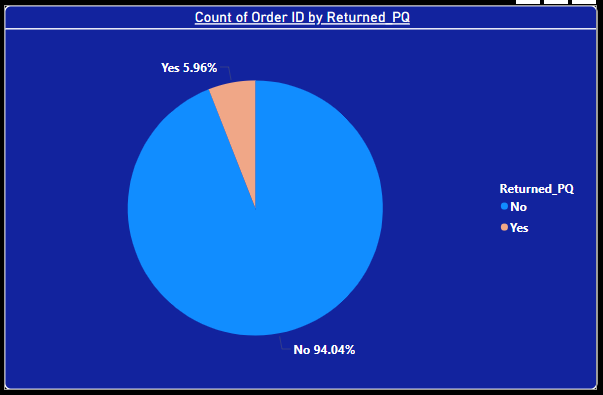
**Q. Find the number of orders returned.**

Ans

Now we have Order ID in orders table and Returned Order ID in Returned Table so have to combine these two tables. So have to use Merge Queries option from Home Menu. In power bi join are referred as Merge.

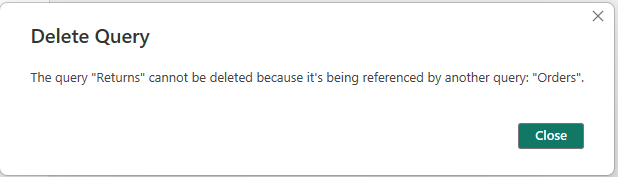
Also, one more option is there in Merge that is Fuzzy Matching which will match the two columns base on similarities. Ans we can provide percen9tage of similarity.

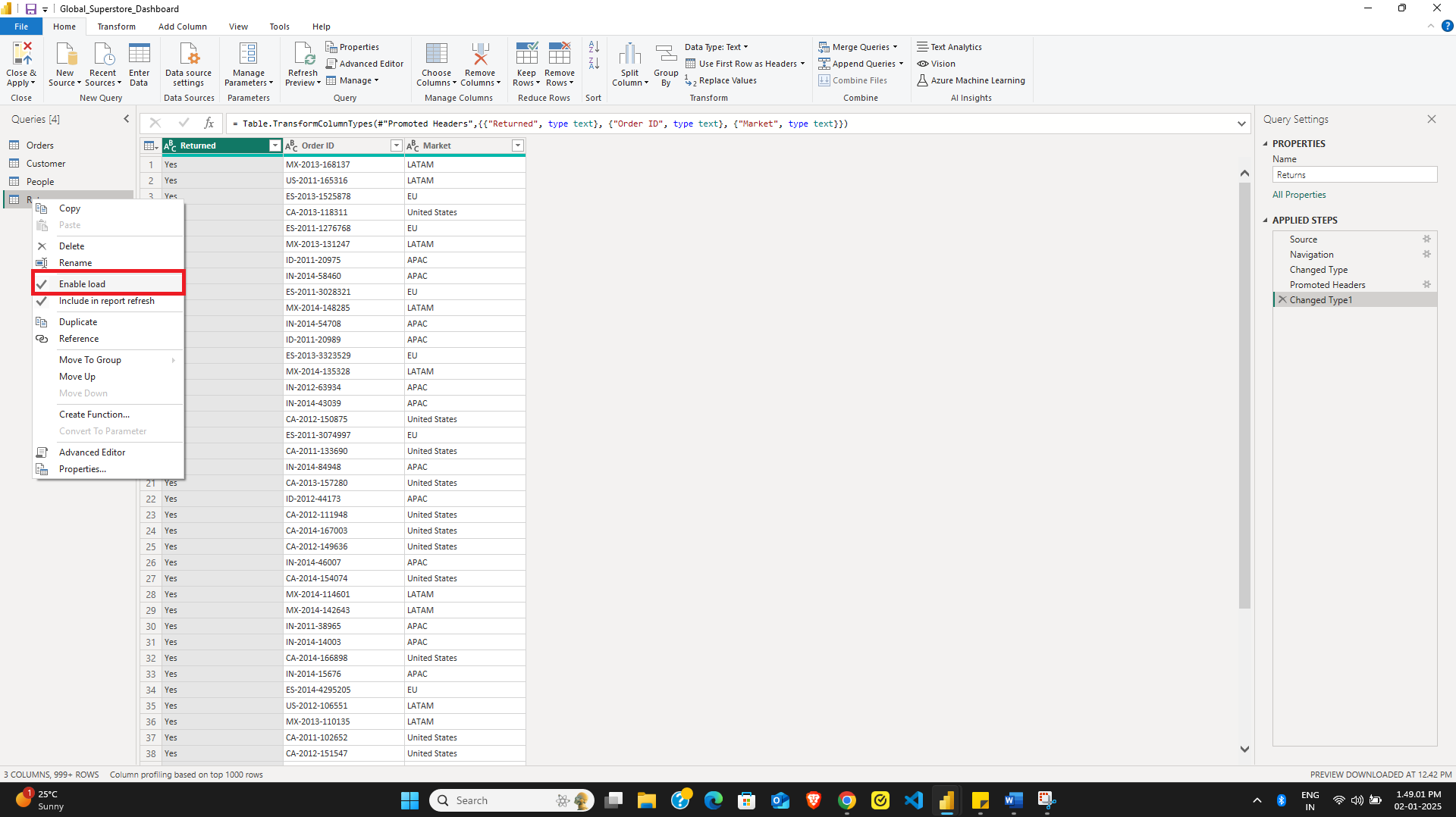
 



Now in power bi Return Table is of no use now. So, we can delete it to save the space. But Bi won’t allow us to delete Return Table because Returned column is coming from that table.

So, we one option of Enable Load which will not load that table into power bi desktop.

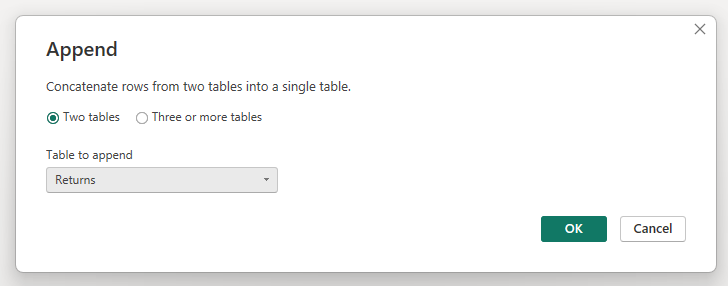




**Append Queries:**

Append Queries are used whenever999 we want to join two excel sheets vertically. But they will join only when the name of column and data type of each column in both sheets must be same.

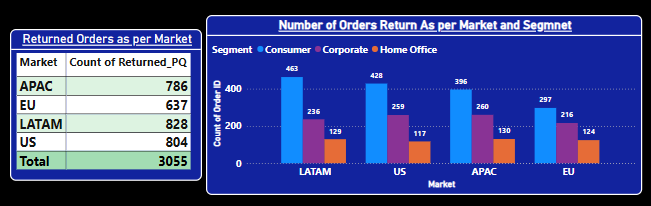
*(If I want to join People and Return Table)*



**Mail -8**

**Q. Find the total numbers of orders returned per market, and also analyse the returns for each market on the basis segment.**

**Ans.**

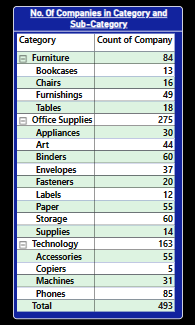
just drag and drop.

**Mail-8**

**Q. Show the number of companies, dealing in each category and subcategory, so that client can have an overview of the diversities of the company.**

**Ans.**

We don’t have any column as Company but if you look at Product Name column then you will easily come to know that 1st Text before space is company name hence, we can go to Transform data and we can add new column with Text before delimiter option.



**Mail-8**

**Q. What is the percentage of quantity sold of routers, keyboards, headsets to total quantity sold in Accessories sub-category.**

**Ans.**

Router % =

var total\_routers = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("router", Orders[Product Name], 1, 0)>0))

var Accessories\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Accessories"))

return

(total\_routers/Accessories\_total)\*100

Keyboard % =

var total\_routers = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("keyboard", Orders[Product Name], 1, 0)>0))

var Accessories\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Accessories"))

return

(total\_routers/Accessories\_total)\*100

Headset % =

var total\_routers = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("Headset", Orders[Product Name], 1, 0)>0))

var Accessories\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Accessories"))

return

(total\_routers/Accessories\_total)\*100



**Mail-8**

**Q. What is the percentage of quantity sold of leather Arm chair, rocking chair, adjustable chairs in Chairs sub-category.**

**Ans.**

Leather Arm Chair =

var total\_larm\_chair = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("leather Armchair", Orders[Product Name], 1, 0)>0))

var Chairs\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Chairs"))

return

Rocking Chair =

var total\_rokng\_chair = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("rocking chair", Orders[Product Name], 1, 0)>0))

var Chairs\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Chairs"))

return

(total\_rokng\_chair/Chairs\_total)\*100



**Q. What is the percentage of iPhone, Samsung, Motorola phones sold to total phones sold.**

**Ans.**

iPhone =

var total\_iphone = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("Apple Smart Phone", Orders[Product Name], 1, 0)>0))

var Phone\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Phones"))

return

(total\_iphone/Phone\_total)\*100

Samsung =

var total\_samsung = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("Samsung Smart Phone", Orders[Product Name], 1, 0)>0))

var Phone\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Phones"))

return

(total\_samsung/Phone\_total)\*100

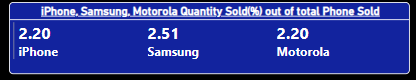
Motorola =

var total\_motorola = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, SEARCH("Motorola Smart Phone", Orders[Product Name], 1, 0)>0))

var Phone\_total = CALCULATE(SUM(Orders[Quantity]),FILTER(Orders, Orders[Sub-Category]="Phones"))

return

(total\_motorola/Phone\_total)\*100



**Q&A Session:**

**What is cross filter direction?**

**Ans.**

Whenever we connect two table, to have some relationship between them. We connect them based on direction of filtration that is nothing but cross-filter direction. Usually there are two options single and both.

Now let say we two table Orders and Peoples. We connect them based on single cross filtration direction; common column is Region.

Orders

Peoples

See the direction of arrow, it means that orders table will get filtered based on people table selection and not vies versa. But if we want filtration in both directions then simply, we have to select cross-filtration direction as “Both”.

**When to Use SUMMARIZE Function?**

**Ans.**

Now client want the analysis of Avg. income based on gender, marital status and age category. So, client actually wants the Avg. income by grouping three things gender, marital status, and age.

So, in such cases SUMMARAIZE function should be used

**What is Static and dynamic RLS?**

**Ans.**

In Static **R**ow **L**evel **S**ecurity – roles are defined by user and in

Dynamic **R**ow **L**evel **S**ecurity – roles are defined by Power BI automatically.

**When to use related and relatedtable?**

**Ans.**

If we have two tables and somehow one column from 1st table is related or required in 2nd table, then we have to bring that column into 2nd table and that operation is done by related DAX.

Related return single value but related-table return multiple values.

**What is the difference between datediff and subtracting two dates manually?**

**Ans.**

If we subtract two dates manually then we get number of days only. We can’t find out months or quarters etc. if we want it then we have to use datediff() function

**\*Imp –**

In Power BI services if you want to share your report to user who don’t have Microsoft account, then in your report there shouldn’t be RLS. Otherwise, you are not able to share.

Data Flow

It is created to avoid the server being get overloaded by refresh requests from reports. So, Data flow is created and connected to source and then refreshed data is supplied to reports.

Now we have to connect the reports to Data flow which is done in power bi desktop. Refer below fig.

Without data flow

Report-1

Semantic Model

Source

Report-2

Semantic Model

Source

With data flow

Data Flow

Semantic Model

Source

Report-1

Semantic Model

Report-2

In power bi go to transform data and then Data flow’s advanced editor & copy entire code and past to that data’s **(*original*)** source (***which is available in Applied Steps***) from which reports has been created.

**Power BI Interview Questions**

(Most obvious asked questions are marked in )

1. **Is it possible to establish multiple functional relationships between two tables in Power BI data model?**

Ans:

Yes, we can establish more than one relationship but not functional or active relationships. For single query we cannot create multiple relationships.

1. **While in Power Query Editor, how effective does the DAX works?**

Ans:

DAX never works in Power Query, in power query M Language used.

1. **What are the different types of Filters available in Power BI and what is their order of applicability?**

Ans:

We have three types of filters Order of applicability

- Filter on this visual 3

- Filter on this page 2

- Filter on all pages 1

*\* Slicers will get applied after filters*

1. **What is the difference between a matrix and a table visual in power bi?**

**(Grid Visuals-either Matrix or Table)**

Ans:

-In table only, we can put any value as column, where as in Matrix you can decide which value should be row and which value should be column.

-Table are unpivot(non-expanded) form and Matrix are pivot(expanded) form of Table

1. **Can you explain the concept of data lineage in Power BI?**

Ans:

Data Lineage is nothing bur link between data source, semantic model and reports created from that data. See below fig.

Data Source

Report

Semantic Model

1. **What is the difference between search and find?**

Ans:

Search is case insensitive New, new, and NEW are same for search. But find is case sensitive.

We can use wildcard characters (\*) in search but in find we can’t.

1. **How All is different from Allselect?**

Ans:

All ignores all filter’s [Visual Filter] applied or in future filter being applied. All will ignore slicers also.

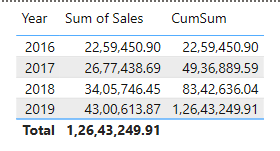
Whereas Allselect ignore contextual filter but filter on this page and filter on all pages are accepted. Filter on this visual will be ignored by Allselect. Slicers will not be ignored in Allselect.

1. **Provide a measure that gives the running/cumulative total using the data table.**

**Ans:**

**Cum Sum =**

CALCULATE(SUM(Orders[Sales]), Orders[Order Date].[year]<= SELECTEDVALUE(Orders[Order Date].[year]))



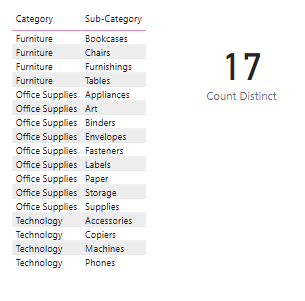
1. **How to get distinct count of two columns ?**

Ans:

To find the unique combination between two columns we can use Summarize and count rows functions as shown below.

Count Distinct = COUNTROWS(SUMMARIZE(Orders,Orders[Category],Orders[Sub-Category]))

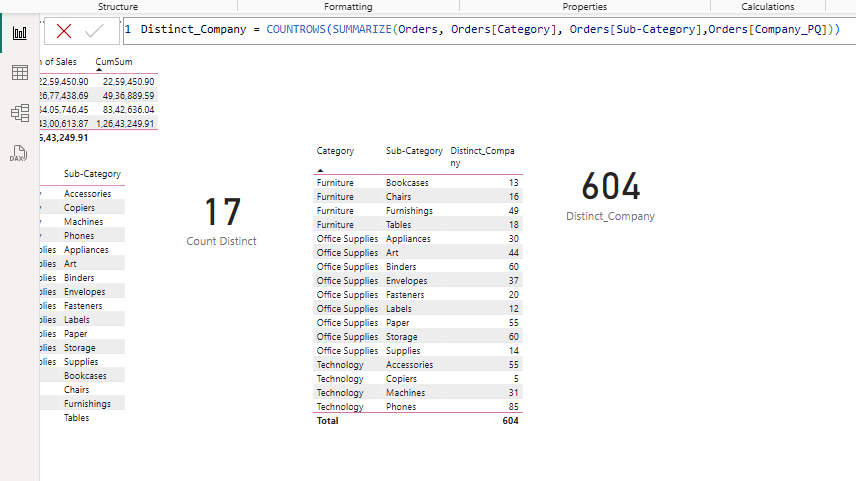
So, summarize function will create the table based on category and subcategory and count row will count the row of that table. Which is 17 I in this case.



1. **Global super store wants to know the number of companies for each category and sub-category?**

Ans:

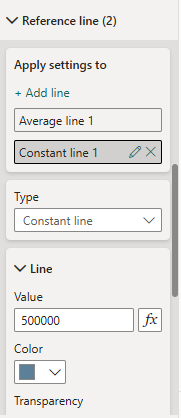
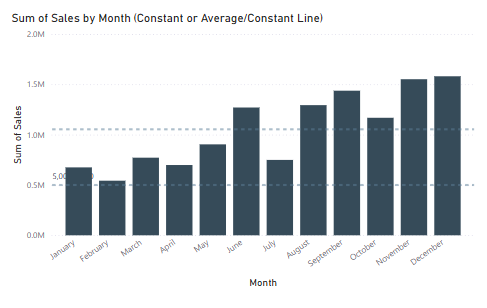
Distinct\_Company = COUNTROWS(SUMMARIZE(Orders, Orders[Category], Orders[Sub-Category],Orders[Company\_PQ]))



1. **How can I add constant or average/constant line in my visual.**

**Ans:**

We can use Reference line option in Format pane.

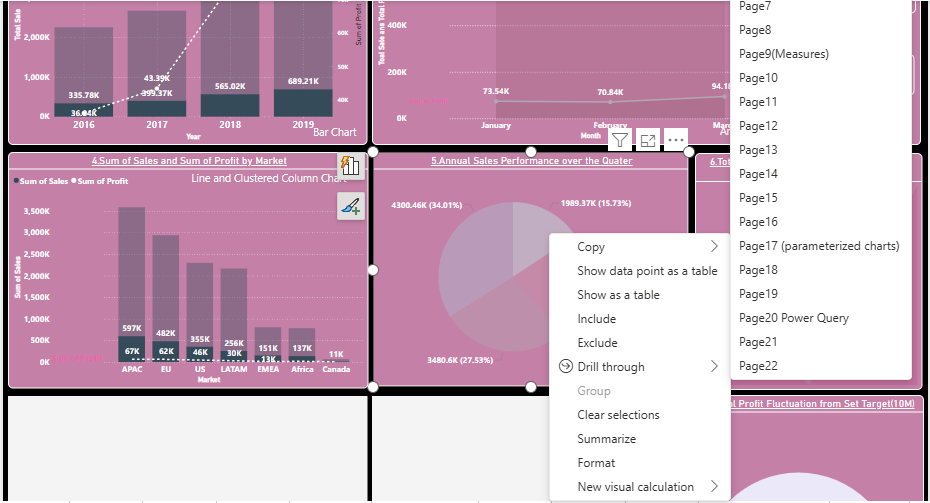


1. **What is drill through in power bi? Have drill through any connection with page level filter.?**

Ans:

When we filter the second page(entire) based on visual which on first page that could be any visual, then that is drill through. It is also in multi-level digging or multi-level analysis like year-month-week-day etc.

So yes, drill through have connection with page level filter.



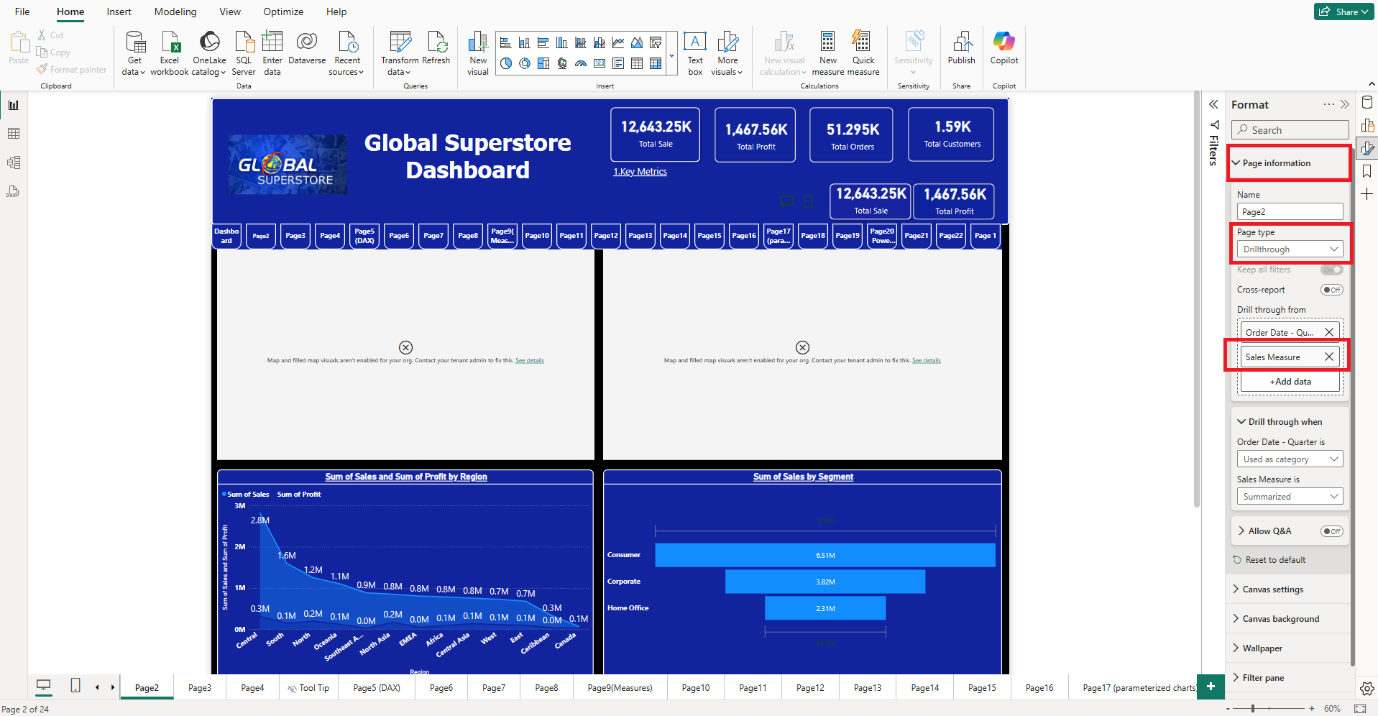
So, we have selected Page 7 and Page 7 get filtered as shown below.

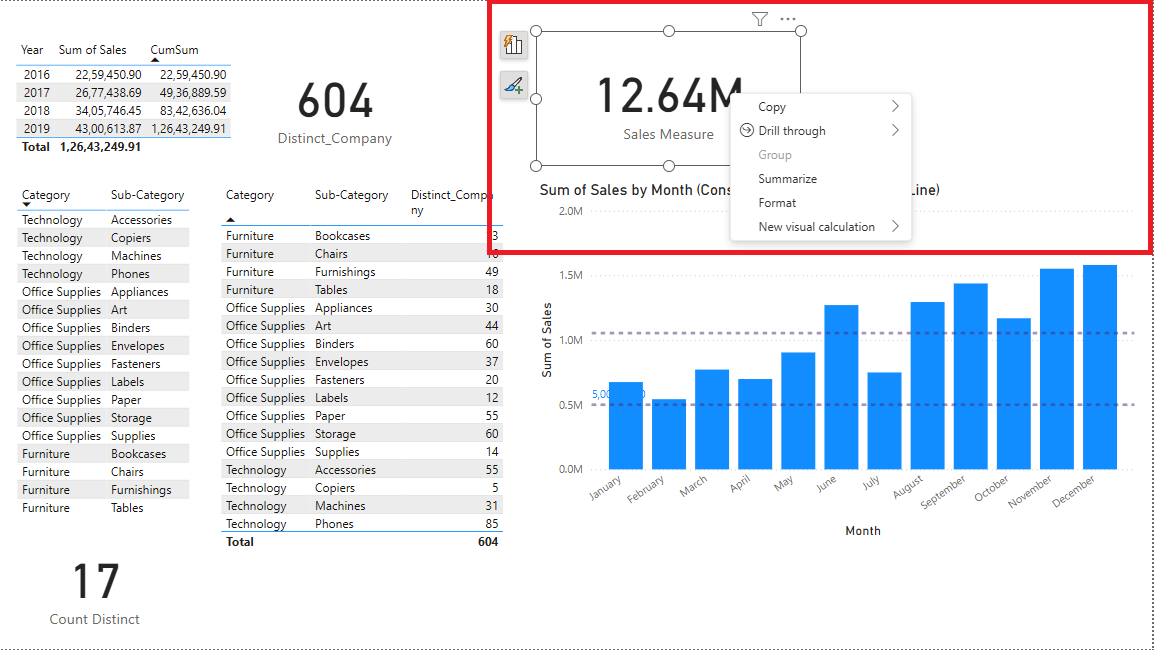


1. **Can we drill through using Card Visual?**

Ans:

Yes, we can drill through card visual also provided that the page to which we are drilling must be drill through and same measure should be selected. And Card must have measure not any aggregate.



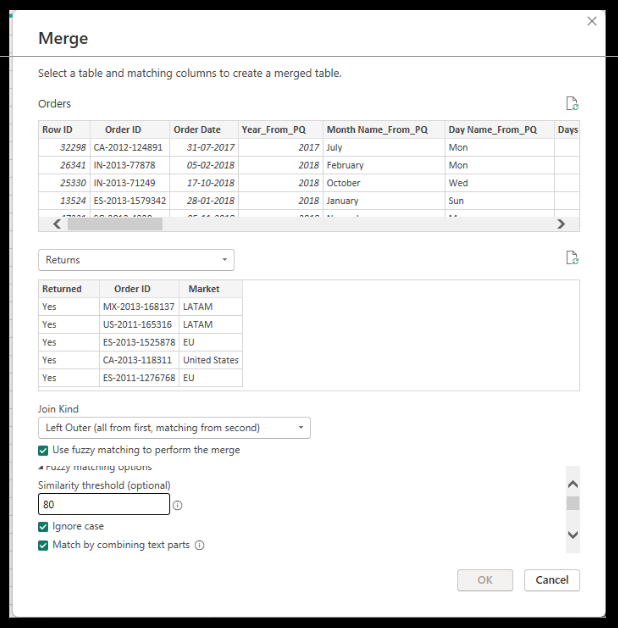


1. **What is Fuzzy Matching in power bi?**

Ans:

Now we have Order ID in orders table and Returned Order ID in Returned Table so have to combine these two tables. So have to use Merge Queries option from Home Menu. In power bi join are referred as Merge.

Also, one more option is there in Merge that is Fuzzy Matching which will match the two columns base on similarities. Ans we can provide percen9tage of similarity.



*In power bi whenever we are joining two tables which has similar columns then merge them, at merging we can provide the percentage of similarity between two columns and that is nothing but fuzzy matching.*

1. **What is the difference between schedule refresh and incremental refresh?**

Ans:

Schedule refresh is refresh at particular time. Like 8:00pm. So at 8:00pm entire data get refreshed and all visuals, measures and all calculations get updated every time and that is the problem with schedule refresh.

Where as in incremental refresh we divide data into two parts static data (historic data) and refresh data, so static part will not get refresh every time and only changes in refresh part will get updated not entire refresh data part.

Refresh Data

Only changes will get updated from this part

Static Data

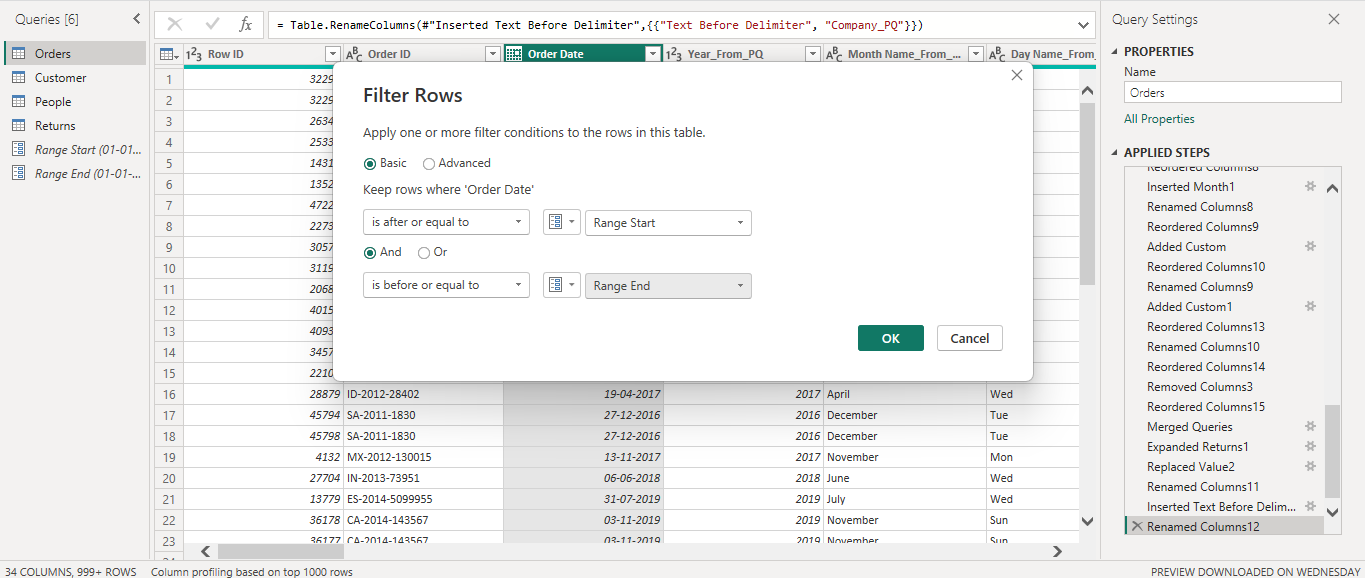
This part will not get refreshed.

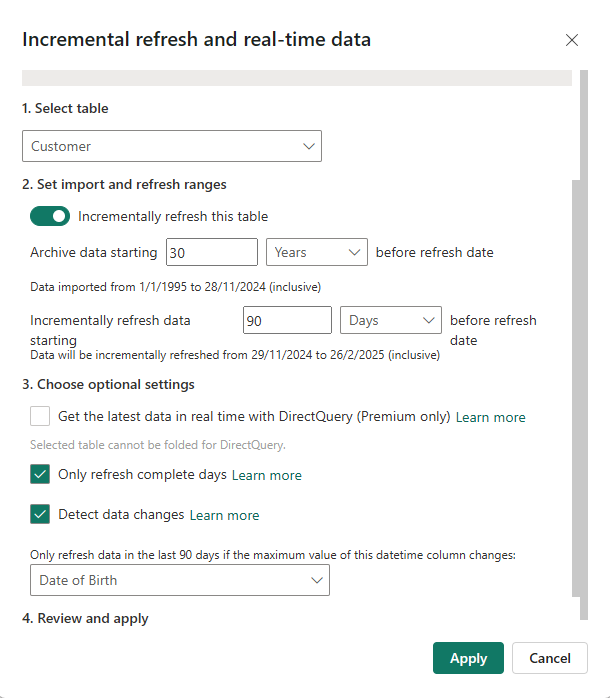
**Pre-requisite to incremental refresh:**

* Incremental refresh only applicable at query folding**.**
* Incremental refresh only applied to Import Mode.
* Date/Date-time column must be there where we are applying incremental refresh.

First, we have set start and end parameter which is done in power query editor and then we have to apply increment which done in power bi desktop.

To set parameter we have option Manage Parameter in Transform data. And then select the column on which you want to apply incremental refresh. Right click on that filter arrow and select Date Filter option, then dialog box as shown below will open, then select parameters as shown.

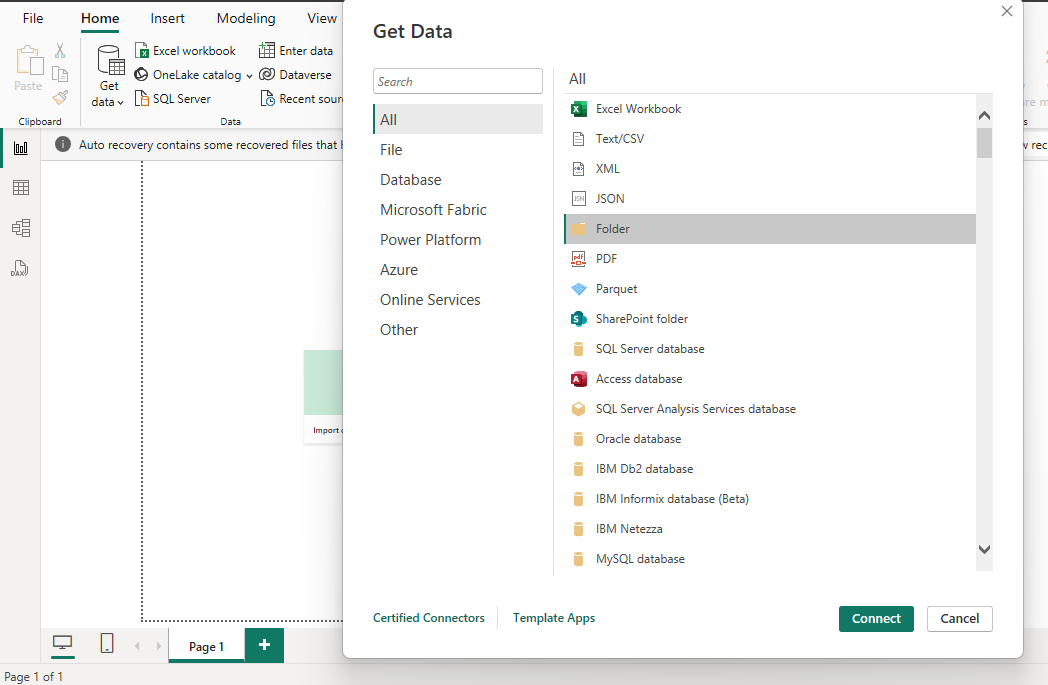




1. **Client has multiple files of same type stored in a folder, he wants to combine them to one, how he/she can do it?**

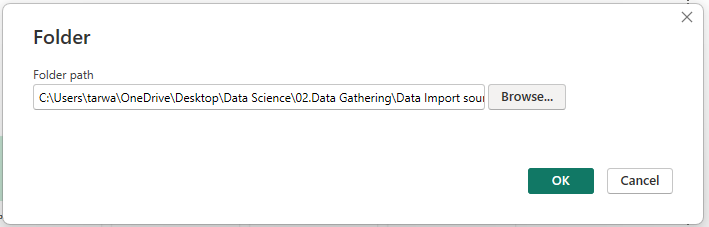
Ans:

We have to go to Get Data option and in that we have to select Folder option, then click on connect.

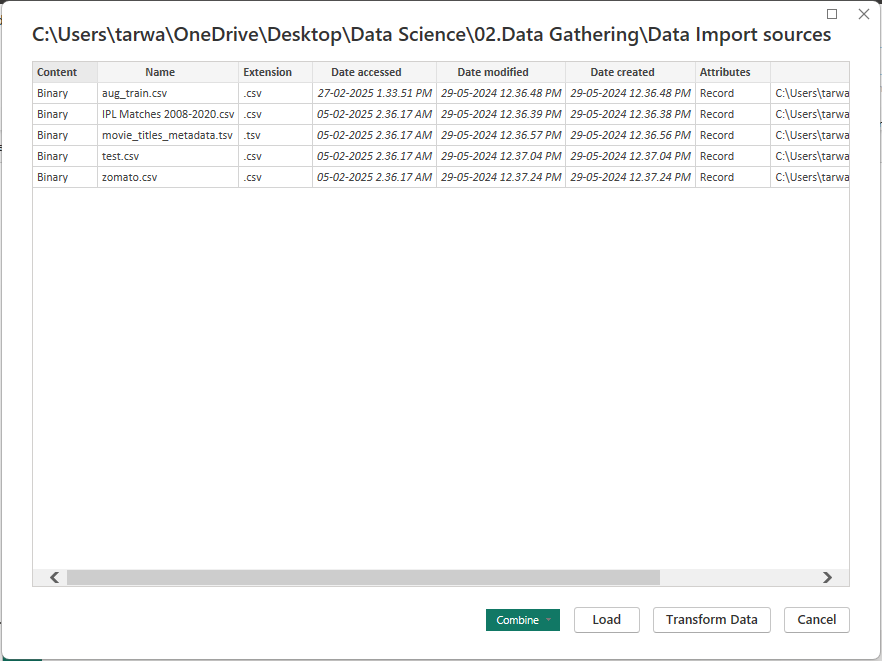


*Note – files in this folder must be of same type.*

Then we have to provide the folder path in which files are there.



After that we have option combine and transform where all those files will get combine to one.

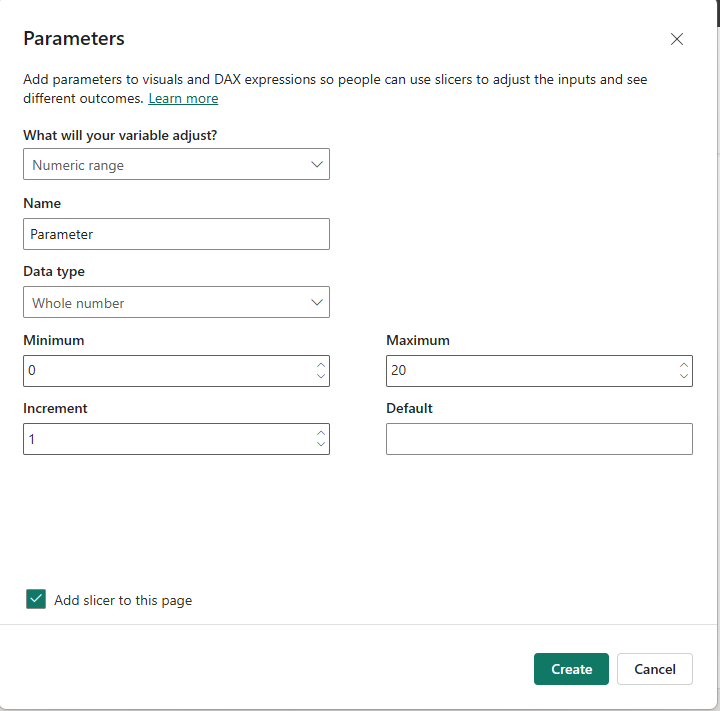


In this way all the files will get combine to one file.

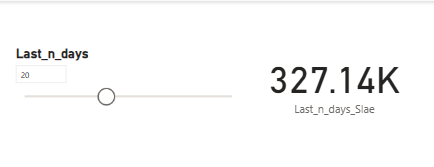
1. **Create a visual that will show the last n days sales and those last n days will be selected by client as per his/her need.**

**Ans:**

First, we will use parameter to chose number of days. Then we have to create measure of last n days sales.



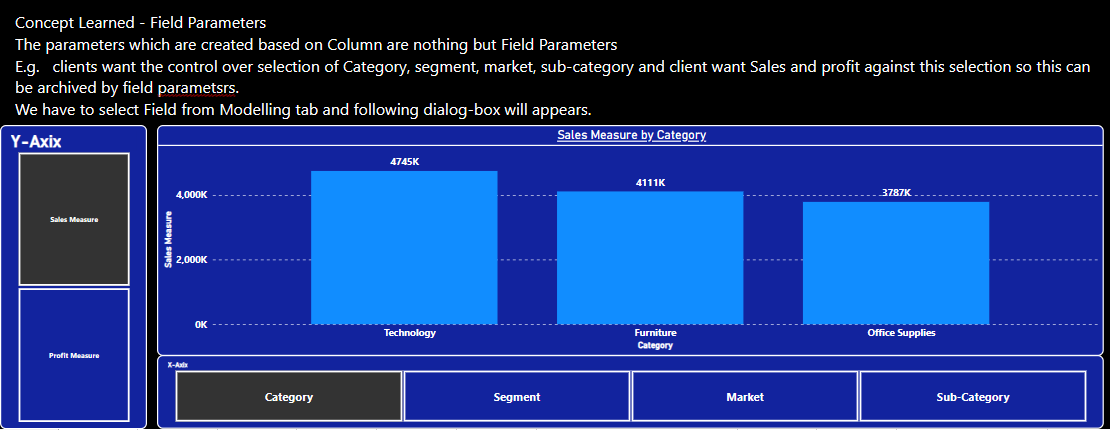
Last\_n\_days\_Sale = CALCULATE(SUM(Orders[Sales]), FILTER(all(Orders[Order Date]), Orders[Order Date]>= MAX(Orders[Order Date])-SELECTEDVALUE(Last\_n\_days[Last\_n\_days])&&Orders[Order Date]<= MAX(Orders[Order Date])))



1. **In a chart all values of axis are fixed. If I want to make it dynamic, how can I do it?**

Ans:

We have to create the Field parameter and then we can make dynamic axis as shown below.



1. **What is bi-directional cross Filtering?**

Ans:

If we two tables, then one table get filter based on selection in other table and vice-verse then it is known as bi-directional cross filtering.

Filtered

Selection

Table-B

Table-A

Selection

Filtered

**DAX Has been Used:**

[**https://learn.microsoft.com/en-us/dax/rankx-function-dax**](https://learn.microsoft.com/en-us/dax/rankx-function-dax)

**SWITCH ()** –

Evaluates an expression against a list of values and returns one of multiple possible result expressions. This function can be used to avoid having multiple nested [IF](https://learn.microsoft.com/en-us/dax/if-function-dax) statements.

SWITCH(<expression>, <value>, <result>[, <value>, <result>]…[, <else>])

**MONTH/DAY/QUARTER/YEAR ()** –

Returns the month as a number from 1 (January) to 12 (December).

MONTH(<datetime>)

**FORMAT () –**

Converts a value to text according to the specified format.

FORMAT(<value>, <format\_string>[, <locale\_name>])

**DATEDIFF () –**

Returns the number of interval boundaries between two dates.

DATEDIFF(<Date1>, <Date2>, <Interval>)

**CALCULATE (SUM/AVG/COUNT/DISTINCTCOUNT, FILTER) –**

Evaluates an expression in a modified filter context.

CALCULATE(<expression>[, <filter1> [, <filter2> [, …]]])

The expression used as the first parameter is essentially the same as a measure.

Filters can be:

* Boolean filter expressions
* Table filter expressions
* Filter modification functions

When there are multiple filters, they can be evaluated by using the AND (&&) [logical operator](https://learn.microsoft.com/en-us/dax/dax-operator-reference#logical-operators), meaning all conditions must be TRUE, or by the OR (||) logical operator, meaning either condition can be true.

**RANKX () –**

Returns the ranking of a number in a list of numbers for each row in the *table* argument.

RANKX(<table>, <expression>[, <value> [, <order>[, <ties>]]])

**SELECTEDVALUE () –**

Returns the value when the context for columnName has been filtered down to one distinct value only. Otherwise returns alternateResult

SELECTEDVALUE(<columnName>[, <alternateResult>])

**DATEDIFF () – Working?**

**SUMMARIZE () – Working?**

**COUNTROWS () – Working?**

**COUNTROWS (FILTER ()) – Working?**

**MAXX () – Working?**

**TOPN () – Working?**

**ISFILTERD () – working?**

**HASONEFILTER () – working?**

**All selected except All etc. – working?**

**Important Interview Questions**

SQL  
1. What is the difference between a clustered index and a non-clustered index?  
2. How do you handle transactions in SQL Server, and what are the ACID properties?  
3. What are table-valued functions, and how do they differ from scalar functions?  
  
Excel  
1. How do you create and use PivotTables for data analysis in Excel?  
2. What are Excel’s data validation rules, and how can they be used to maintain data integrity?  
3. How do you use Excel’s Solver add-in for optimization problems?  
  
Power BI  
1. How do you use DAX (Data Analysis Expressions) to create calculated columns and measures in Power BI?  
2. What is the difference between Import mode, DirectQuery, and Live Connection in Power BI?  
3. How do you handle and visualize time-series data in Power BI?  
  
Python  
1. How do you handle exceptions in Python, and why is it important in data processing?  
2. What is the Pandas merge function, and how is it used for combining DataFrames?  
3. How do you deploy a machine learning model built with Python into production?  
  
Data Visualization  
1. How do you ensure data integrity and avoid misleading visualizations?  
2. What are the best practices for choosing colors and fonts in data visualizations?  
3. How do you use GIS (Geographic Information System) data for spatial visualizations in tools like Tableau or Power BI?