BI Practical File

Sr. No.	Practicals	Date
1	**Import the legacy data from different sources such as (Excel, OData Feed, etc.) and load in the target system.**	**5/12/22**
2	**Data Analysis using Time Series Analysis**	**12/12/22**
3	**Prediction using Linear Regression**	**12/12/22**
4	**Perform Extraction, Transformation & Loading ETL Process in Power BI**	**26/12/22**
5	**Data Visualization from ETL Process**	**2/1/23**
6	**What-if Analysis using excel for data visualization**	**9/1/23**
7	**K-means clustering using R**	**16/1/23**
8	**Pivot Table**	**6/2/23**
9	**Height and Weight Regression**	**20/2/23**

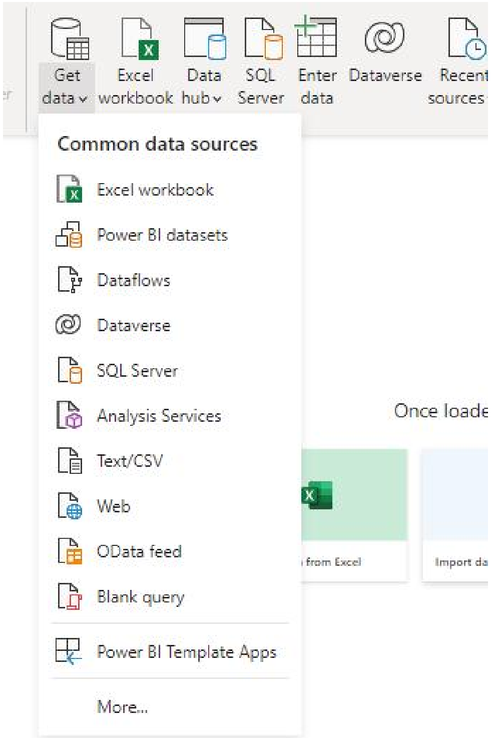
PRACTICAL 1

Aim : Import the legacy data from different sources such as (Excel, OData Feed, etc.) and load in the target system.

Import Excel Data

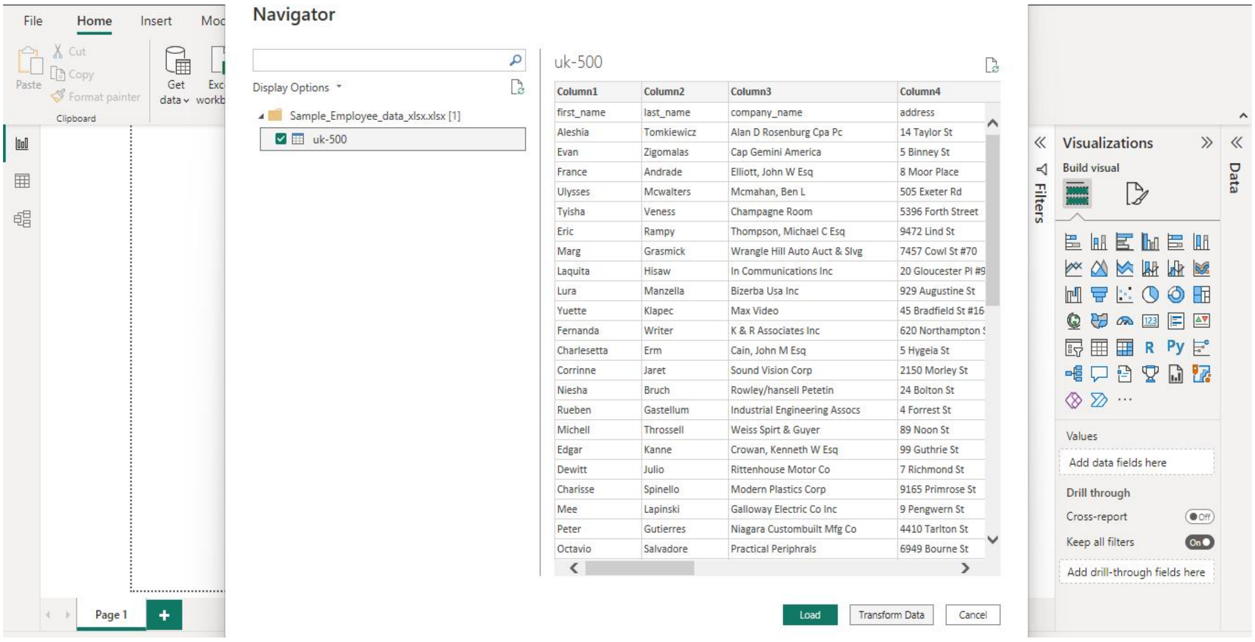
1) Launch Power BI

2) From the Home ribbon, select Get Data. Excel is one of the Most Common data connections, so you can select it directly from the Get Data menu.



1) If you select the Get Data button directly, you can also select Get data > Excel Workbook

2) You will get the data & you can transform it according to your own



uk-500

Column1	Column2	Column3	Column4
first_name	last_name	company_name	address
Aleshia	Tomkiewicz	Alan D Rosenburg Cpa Pc	14 Taylor St
Evan	Zigomalas	Cap Gemini America	5 Binney St
France	Andrade	Elliott, John W Esq	8 Moor Place
Ulysses	Mcwalters	Mcmahan, Ben L	505 Exeter Rd
Tyisha	Veness	Champagne Room	5396 Forth Street
Eric	Rampy	Thompson, Michael C Esq	9472 Lind St
Marg	Grasmick	Wrangle Hill Auto Auct & Slvg	7457 Cowl St #70
Laquita	Hisaw	In Communications Inc	20 Gloucester Pl #9
Lura	Manzella	Bizerba Usa Inc	929 Augustine St
Yuette	Klapec	Max Video	45 Bradfield St #16
Fernanda	Writer	K & R Associates Inc	620 Northampton S
Charlesetta	Erm	Cain, John M Esq	5 Hygeia St
Corrinne	Jaret	Sound Vision Corp	2150 Morley St
Niesha	Bruch	Rowley/hansell Petetin	24 Bolton St
Rueben	Gastellum	Industrial Engineering Assocs	4 Forrest St
Michell	Throssell	Weiss Spirt & Guyer	89 Noon St
Edgar	Kanne	Crowan, Kenneth W Esq	99 Guthrie St
Dewitt	Julio	Rittenhouse Motor Co	7 Richmond St
Charisse	Spinello	Modern Plastics Corp	9165 Primrose St
Mee	Lapinski	Galloway Electric Co Inc	9 Pengwern St
Peter	Gutierres	Niagara Custombuilt Mfg Co	4410 Tarlton St
Octavio	Salvadore	Practical Periphrals	6949 Bourne St

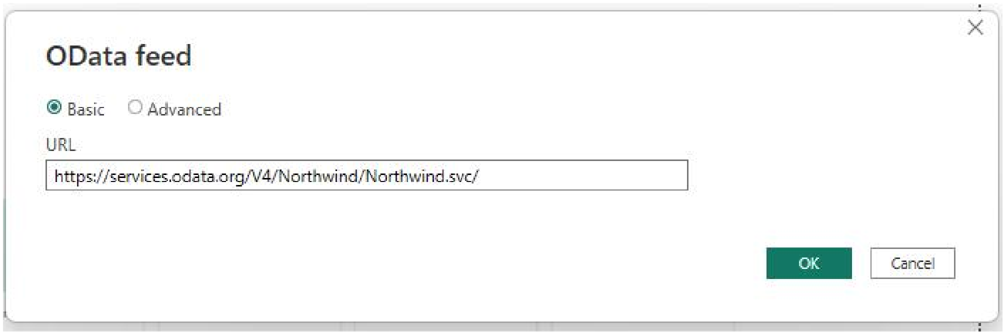
Importing Data from OData Feed -

Connect to an OData feed:

1) From the Home ribbon tab in Query Editor, select Get Data.

2) Browse to the OData Feed data source.

3) In the OData Feed dialog box, paste the URL for the Northwind OData feed.



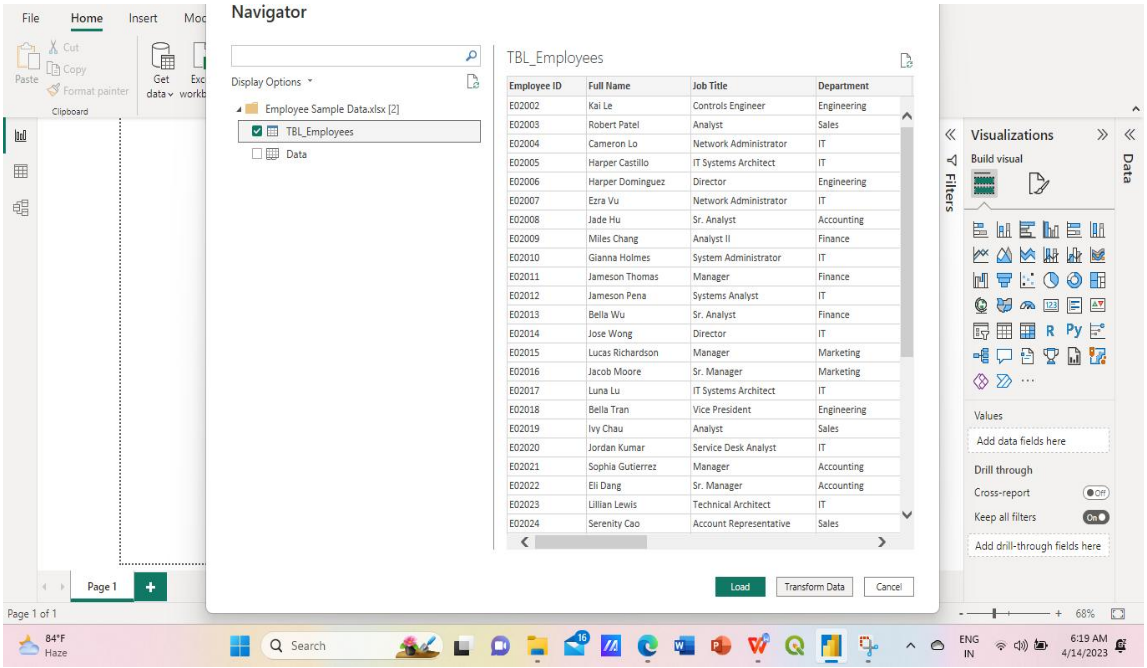
● Basic ○ Advanced

URL

https://services.odata.org/V4/Northwind/Northwind.svc/

4) Select OK.

5) In the Navigator pane, select the Orders table, and then select Edit.



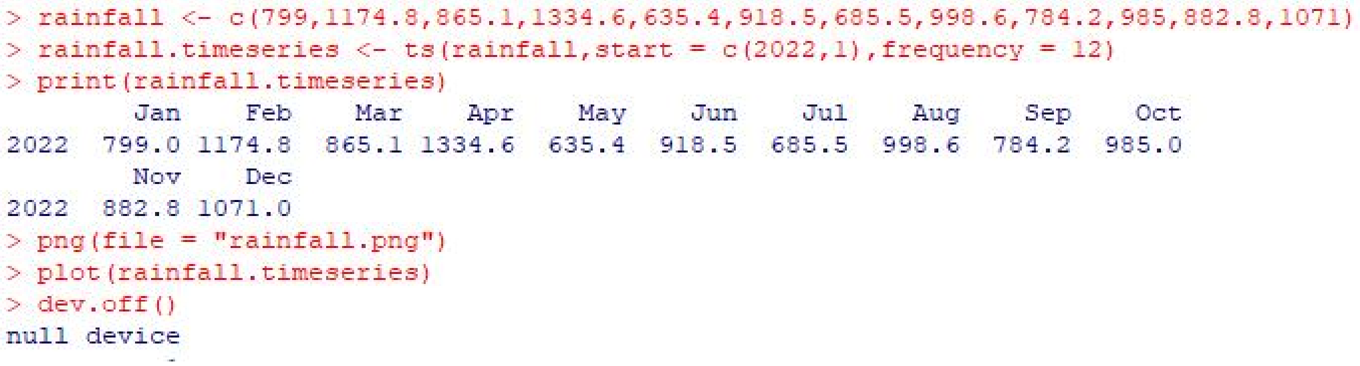
TBL_Employees

Employee ID	Full Name	Job Title	Department
E02002	Kai Le	Controls Engineer	Engineering
E02003	Robert Patel	Analyst	Sales
E02004	Cameron Lo	Network Administrator	IT
E02005	Harper Castillo	IT Systems Architect	IT
E02006	Harper Dominguez	Director	Engineering
E02007	Ezra Vu	Network Administrator	IT
E02008	Jade Hu	Sr. Analyst	Accounting
E02009	Miles Chang	Analyst II	Finance
E02010	Gianna Holmes	System Administrator	IT
E02011	Jameson Thomas	Manager	Finance
E02012	Jameson Pena	Systems Analyst	IT
E02013	Bella Wu	Sr. Analyst	Finance
E02014	Jose Wong	Director	IT
E02015	Lucas Richardson	Manager	Marketing
E02016	Jacob Moore	Sr. Manager	Marketing
E02017	Luna Lu	IT Systems Architect	IT
E02018	Bella Tran	Vice President	Engineering
E02019	Ivy Chau	Analyst	Sales
E02020	Jordan Kumar	Service Desk Analyst	IT
E02021	Sophia Gutierrez	Manager	Accounting
E02022	Eli Dang	Sr. Manager	Accounting
E02023	Lillian Lewis	Technical Architect	IT
E02024	Serenity Cao	Account Representative	Sales

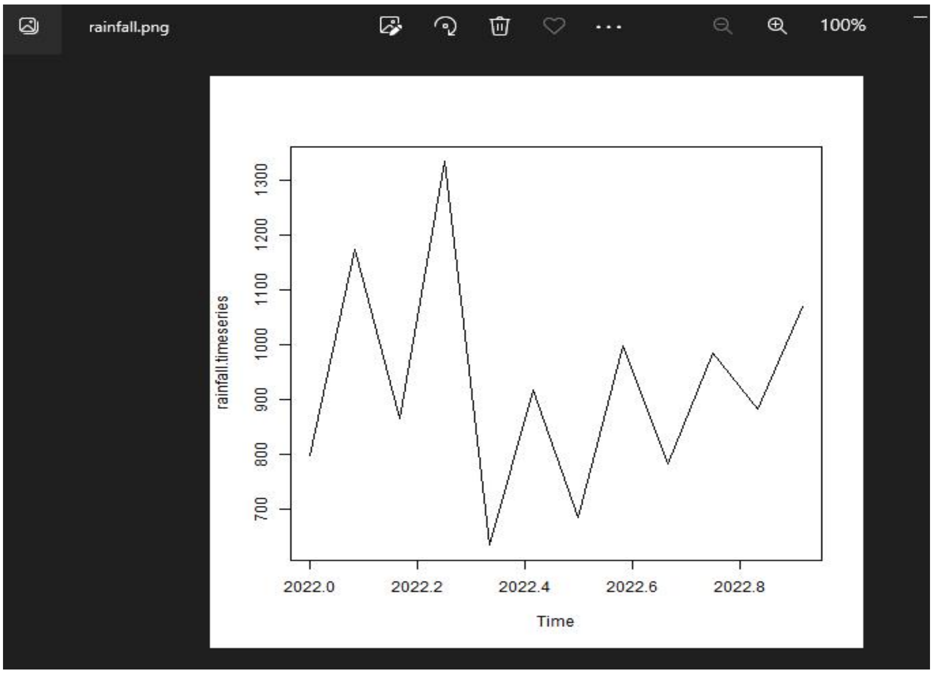
Note - You can click a table name, without selecting the checkbox, to see a preview.

PRACTICAL 2

Aim : Data Analysis using Time Series Analysis



```
> rainfall <- c(799,1174.8,865.1,1334.6,635.4,918.5,685.5,998.6,784.2,985,882.8,1071)
> rainfall.timeseries <- ts(rainfall,start = c(2022,1),frequency = 12)
> print(rainfall.timeseries)
        Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct
2022  799.0 1174.8  865.1 1334.6  635.4  918.5  685.5  998.6  784.2  985.0
        Nov    Dec
2022  882.8 1071.0
> png(file = "rainfall.png")
> plot(rainfall.timeseries)
> dev.off()
null device
```



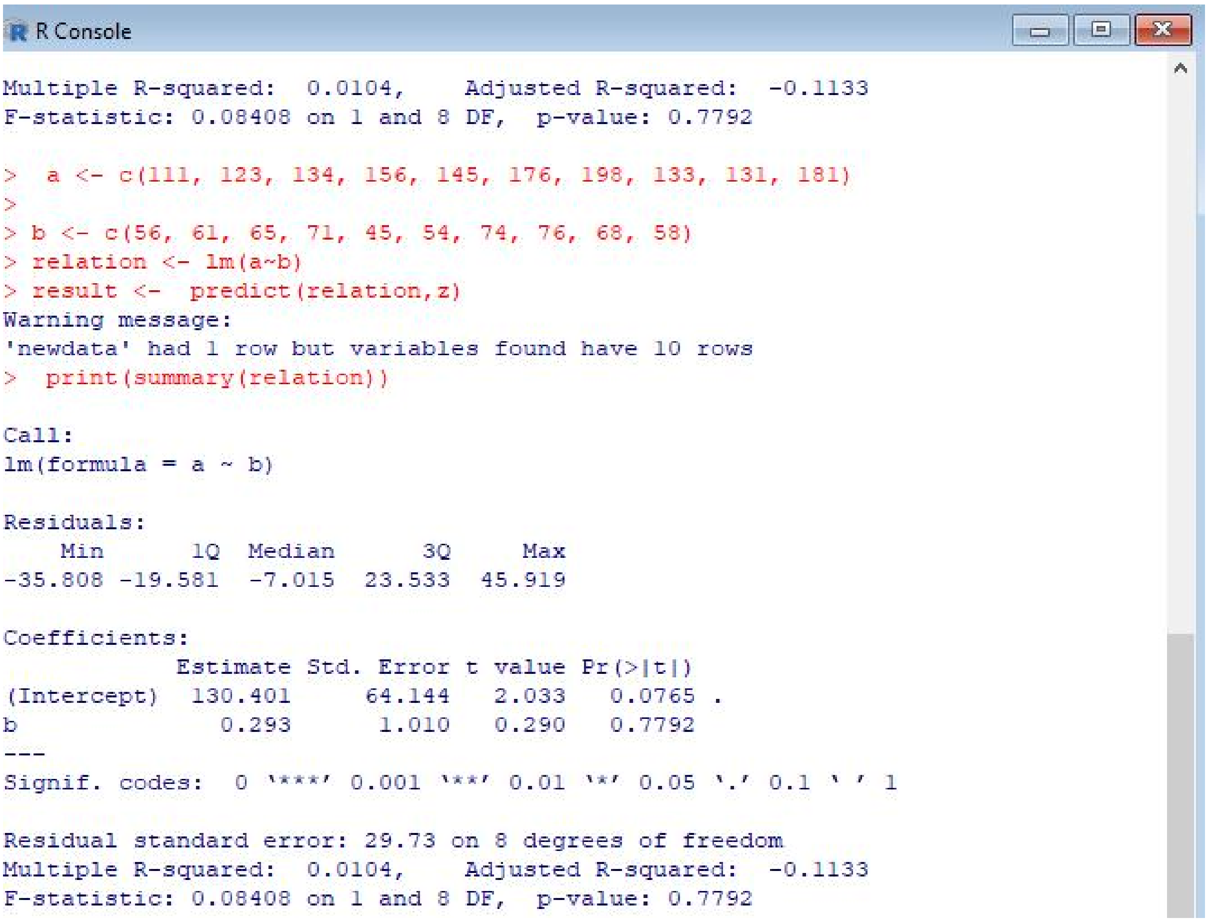
PRACTICAL 3

Aim : Prediction using Linear Regression

Steps :

1) Gather heights and weights like atleast a few observations.
2) Use the lm() function to create the relationship model
3) Use the coefficients from the model and create a mathematical equation.
4) Predict the weight of new observations using the predict() function.

Result <- predict(relation,z)



```
Multiple R-squared:  0.0104,    Adjusted R-squared:  -0.1133
F-statistic: 0.08408 on 1 and 8 DF,  p-value: 0.7792

>  a <- c(111, 123, 134, 156, 145, 176, 198, 133, 131, 181)
>
> b <- c(56, 61, 65, 71, 45, 54, 74, 76, 68, 58)
> relation <- lm(a~b)
> result <-  predict(relation,z)
Warning message:
'newdata' had 1 row but variables found have 10 rows
>  print(summary(relation))

Call:
lm(formula = a ~ b)

Residuals:
    Min      1Q  Median      3Q     Max
-35.808 -19.581  -7.015  23.533  45.919

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  130.401     64.144   2.033   0.0765 .
b              0.293      1.010   0.290   0.7792
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 29.73 on 8 degrees of freedom
Multiple R-squared:  0.0104,    Adjusted R-squared:  -0.1133
F-statistic: 0.08408 on 1 and 8 DF,  p-value: 0.7792
```

PRACTICAL 3

Aim : Perform Extraction, Transformation & Loading ETL Process in Power BI

Step 1 : Data Extraction

There are two types of data extraction

1. Full Extraction : All the data from source systems or operational systems gets extracted tostaging area. (Initial Load)

2. Partial Extraction : Sometimes we get notification from the source system to updatespecific date. It is called as Delta load.

Source System Performance: The Extraction strategies should not affect source system performance.

Step 2 : Data Transformation :

The data transformation is second step.After extracting the data there is big need to do the transformation as per the target system.I would like to give you some bullet points of Data Transformation.

- Data Extracted from source system is in to Raw format. We need to transform itbefore loading in to target server.
- Data has to be cleaned, mapped and transformed
- There are following important steps of Data Transformation :

1.Selection : Select data to load in target
2.Matching : Match the data with target system
3.Data Transforming : We need to change or transform the data as per target table structures

Step 3 : Data Loading

Data loading phase loads the prepared data from staging tables to main tables.

ETL Process in Power BI

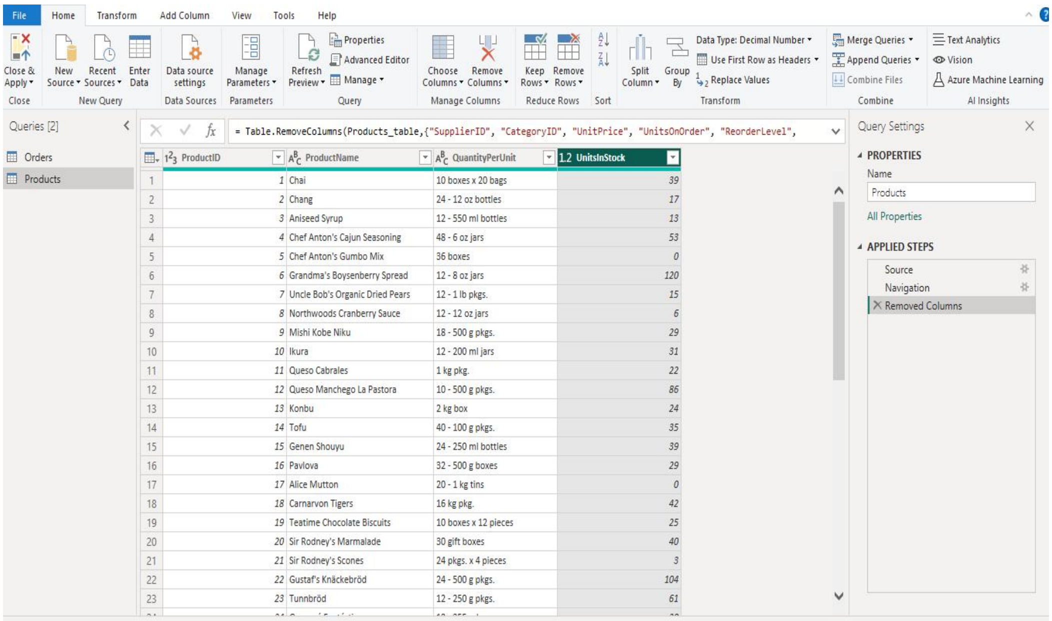
Select orders and products table from northwind.svc

1) **Remove other columns to only display columns of interest**

In this step you remove all columns except **ProductID**, **ProductName**, **UnitsInStock**, and **QuantityPerUnit**

Power BI Desktop includes Query Editor, which is where you shape and transform your data connections. Query Editor opens automatically when you select **Edit** from Navigator. You can also open the Query Editor by selecting Edit Queries from the Home ribbon in Power BI Desktop. The following steps are performed in Query Editor.

1. In **Query Editor**, select the **ProductID, ProductName**, **QuantityPerUnit,** and **UnitsInStock** columns (use **Ctrl+Click** to select more than one column, or **Shift+Click** to select columns that are beside each other).
2. Select **Remove Columns > Remove** Other Columns from the ribbon, or right-clickon a column header and click Remove Other Columns.



```
= Table.RemoveColumns(Products_table,{"SupplierID", "CategoryID", "UnitPrice", "UnitsOnOrder", "ReorderLevel",
```

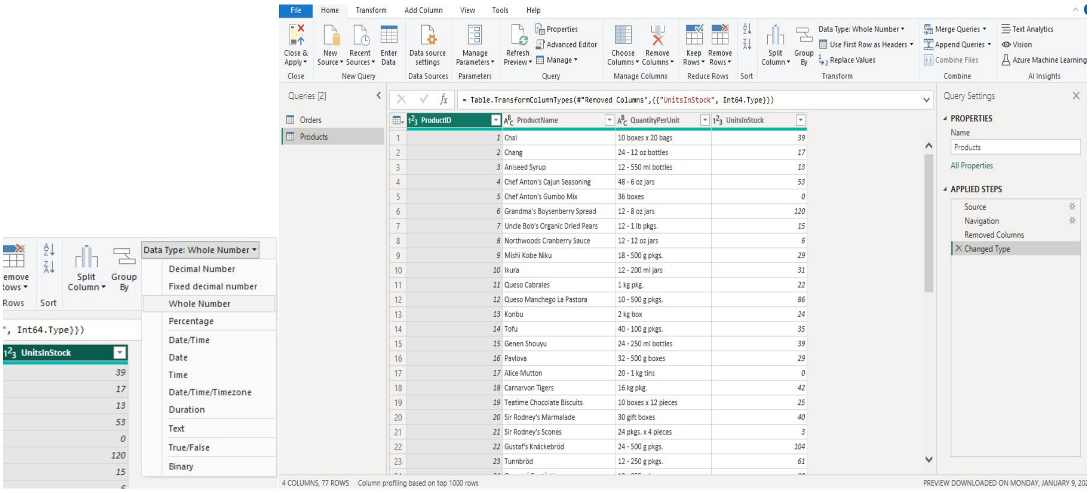
	ProductID	ProductName	QuantityPerUnit	UnitsInStock
1	1	Chai	10 boxes x 20 bags	39
2	2	Chang	24 - 12 oz bottles	17
3	3	Aniseed Syrup	12 - 550 ml bottles	13
4	4	Chef Anton's Cajun Seasoning	48 - 6 oz jars	53
5	5	Chef Anton's Gumbo Mix	36 boxes	0
6	6	Grandma's Boysenberry Spread	12 - 8 oz jars	120
7	7	Uncle Bob's Organic Dried Pears	12 - 1 lb pkgs.	15
8	8	Northwoods Cranberry Sauce	12 - 12 oz jars	6
9	9	Mishi Kobe Niku	18 - 500 g pkgs.	29
10	10	Ikura	12 - 200 ml jars	31
11	11	Queso Cabrales	1 kg pkg.	22
12	12	Queso Manchego La Pastora	10 - 500 g pkgs.	86
13	13	Konbu	2 kg box	24
14	14	Tofu	40 - 100 g pkgs.	35
15	15	Genen Shouyu	24 - 250 ml bottles	39
16	16	Pavlova	32 - 500 g boxes	29
17	17	Alice Mutton	20 - 1 kg tins	0
18	18	Carnarvon Tigers	16 kg pkg.	42
19	19	Teatime Chocolate Biscuits	10 boxes x 12 pieces	25
20	20	Sir Rodney's Marmalade	30 gift boxes	40
21	21	Sir Rodney's Scones	24 pkgs. x 4 pieces	3
22	22	Gustaf's Knäckebröd	24 - 500 g pkgs.	104
23	23	Tunnbröd	12 - 250 g pkgs.	61

2) Change the data type of the UnitsInStock column

When Query Editor connects to data, it reviews each field and to determine the best data type. For the Excel workbook, products in stock will always be a whole number, so in this step you confirm the **UnitsInStock** column's datatype is Whole Number.

1. Select the **UnitsInStock** column.

2. Select the **Data Type drop-down button** in the **Home ribbon**.
   If not already a Whole Number, select **Whole Number** for data type from the drop down(the Data Type: button also displays the data type for the current selection).



```
= Table.TransformColumnTypes(#"Removed Columns",{{"UnitsInStock", Int64.Type}})
```

	ProductID	ProductName	QuantityPerUnit	UnitsInStock
1	1	Chai	10 boxes x 20 bags	39
2	2	Chang	24 - 12 oz bottles	17
3	3	Aniseed Syrup	12 - 550 ml bottles	13
4	4	Chef Anton's Cajun Seasoning	48 - 6 oz jars	53
5	5	Chef Anton's Gumbo Mix	36 boxes	0
6	6	Grandma's Boysenberry Spread	12 - 8 oz jars	120
7	7	Uncle Bob's Organic Dried Pears	12 - 1 lb pkgs.	15
8	8	Northwoods Cranberry Sauce	12 - 12 oz jars	6
9	9	Mishi Kobe Niku	18 - 500 g pkgs.	29
10	10	Ikura	12 - 200 ml jars	31
11	11	Queso Cabrales	1 kg pkg.	22
12	12	Queso Manchego La Pastora	10 - 500 g pkgs.	86
13	13	Konbu	2 kg box	24
14	14	Tofu	40 - 100 g pkgs.	35
15	15	Genen Shouyu	24 - 250 ml bottles	39
16	16	Pavlova	32 - 500 g boxes	29
17	17	Alice Mutton	20 - 1 kg tins	0
18	18	Carnarvon Tigers	16 kg pkg.	42
19	19	Teatime Chocolate Biscuits	10 boxes x 12 pieces	25
20	20	Sir Rodney's Marmalade	30 gift boxes	40
21	21	Sir Rodney's Scones	24 pkgs. x 4 pieces	3
22	22	Gustaf's Knäckebröd	24 - 500 g pkgs.	104
23	23	Tunnbröd	12 - 250 g pkgs.	61

4 COLUMNS, 77 ROWS   Column profiling based on top 1000 rows   PREVIEW DOWNLOADED ON MONDAY, JANUARY 9, 202

3) Expand the Order_Details table

The Orders table contains a reference to a Details table, which contains the individual productsthat were included in each Order. When you connect to data sources with multiples tables (such as a relational database) you can use these references to build up your query

In this step, you expand the **Order_Details** table that is related to the Orders table, to combine the **ProductID**, **UnitPrice**, and **Quantity** columns from **Order_Details** into the **Orders table**. This is a representation of the data in these tables:

The Expand operation combines columns from a related table into a subject table. When the query runs, rows from the related table (**Order_Details**) are combined into rows from the subject table (**Orders**).

After you expand the Order_Details table, three new columns and additional rows are addedto the Orders table, one for each row in the nested or related table.

1. In the Query View, scroll to the Order_Details column.

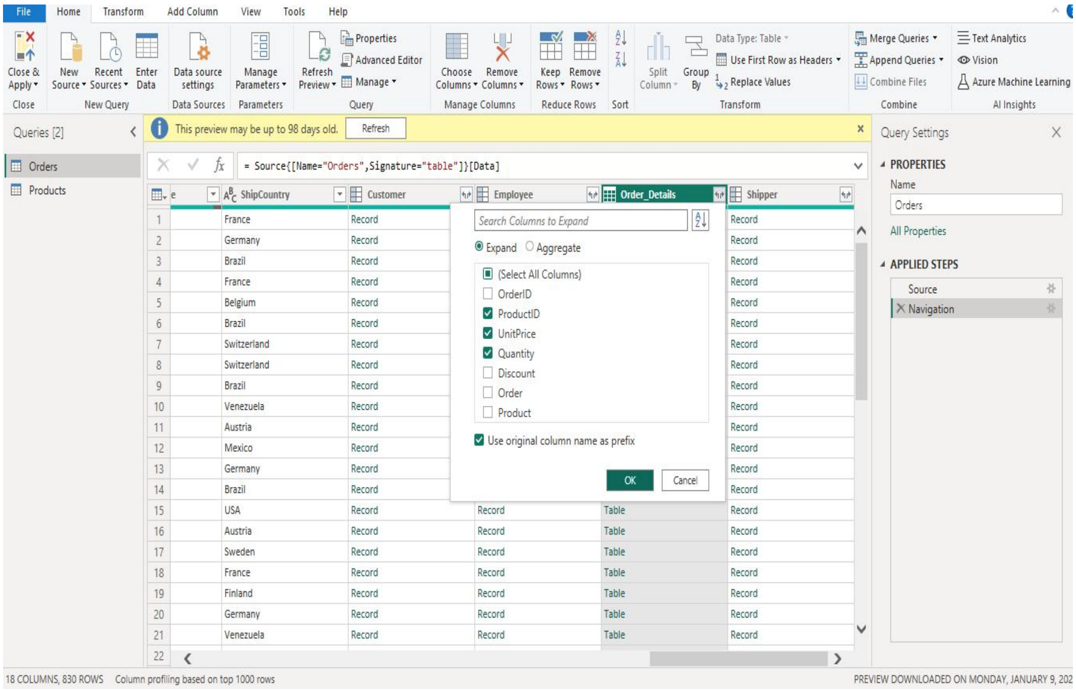
2. In the Order_Details column, select the expand icon ( ).

3. In the Expand drop-down:

   a. Select (Select All Columns) to clear all columns.

   b. Select ProductID, UnitPrice, and Quantity.

   c. Click OK.



This preview may be up to 98 days old.

Name: Orders

```
= Source{[Name="Orders",Signature="table"]}[Data]
```

Search Columns to Expand

● Expand ○ Aggregate

- ■ (Select All Columns)
- □ OrderID
- ☑ ProductID
- ☑ UnitPrice
- ☑ Quantity
- □ Discount
- □ Order
- □ Product

☑ Use original column name as prefix

OK Cancel

	e	ShipCountry	Customer	Employee	Order_Details	Shipper
1		France	Record			Record
2		Germany	Record			Record
3		Brazil	Record			Record
4		France	Record			Record
5		Belgium	Record			Record
6		Brazil	Record			Record
7		Switzerland	Record			Record
8		Switzerland	Record			Record
9		Brazil	Record			Record
10		Venezuela	Record			Record
11		Austria	Record			Record
12		Mexico	Record			Record
13		Germany	Record			Record
14		Brazil	Record			Record
15		USA	Record	Record	Table	Record
16		Austria	Record	Record	Table	Record
17		Sweden	Record	Record	Table	Record
18		France	Record	Record	Table	Record
19		Finland	Record	Record	Table	Record
20		Germany	Record	Record	Table	Record
21		Venezuela	Record	Record	Table	Record
22						

18 COLUMNS, 830 ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED ON MONDAY, JANUARY 9, 202

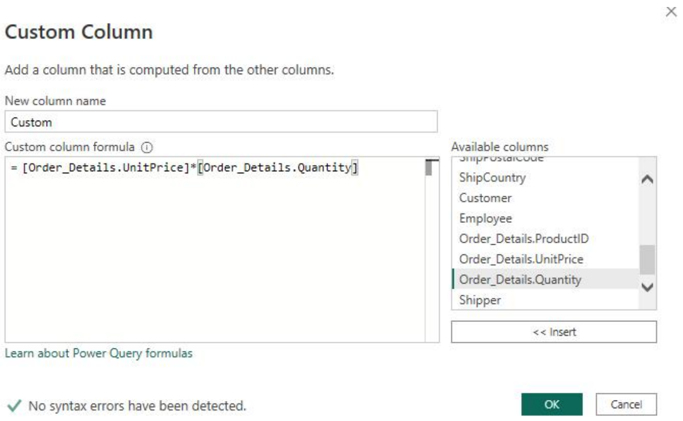
4. Calculate the line total for each Order_Details row

Power BI Desktop lets you to create calculations based on the columns you are importing, so you can enrich the data that you connect to. In this step, you create a Custom Column to calculate the line total for each Order_Details row.

Calculate the line total for each Order_Details row:

1. In the Add Column ribbon tab, click Add Custom Column.
2. In the Add Custom Column dialog box, in the Custom Column Formula textbox, enter [Order_Details.UnitPrice] * [Order_Details.Quantity].
3. In the New column name textbox, enter LineTotal.

   Click OK.



Custom Column

Add a column that is computed from the other columns.

New column name: Custom

Custom column formula:

```
= [Order_Details.UnitPrice]*[Order_Details.Quantity]
```

Available columns: ShipPostalCode, ShipCountry, Customer, Employee, Order_Details.ProductID, Order_Details.UnitPrice, Order_Details.Quantity, Shipper

Learn about Power Query formulas

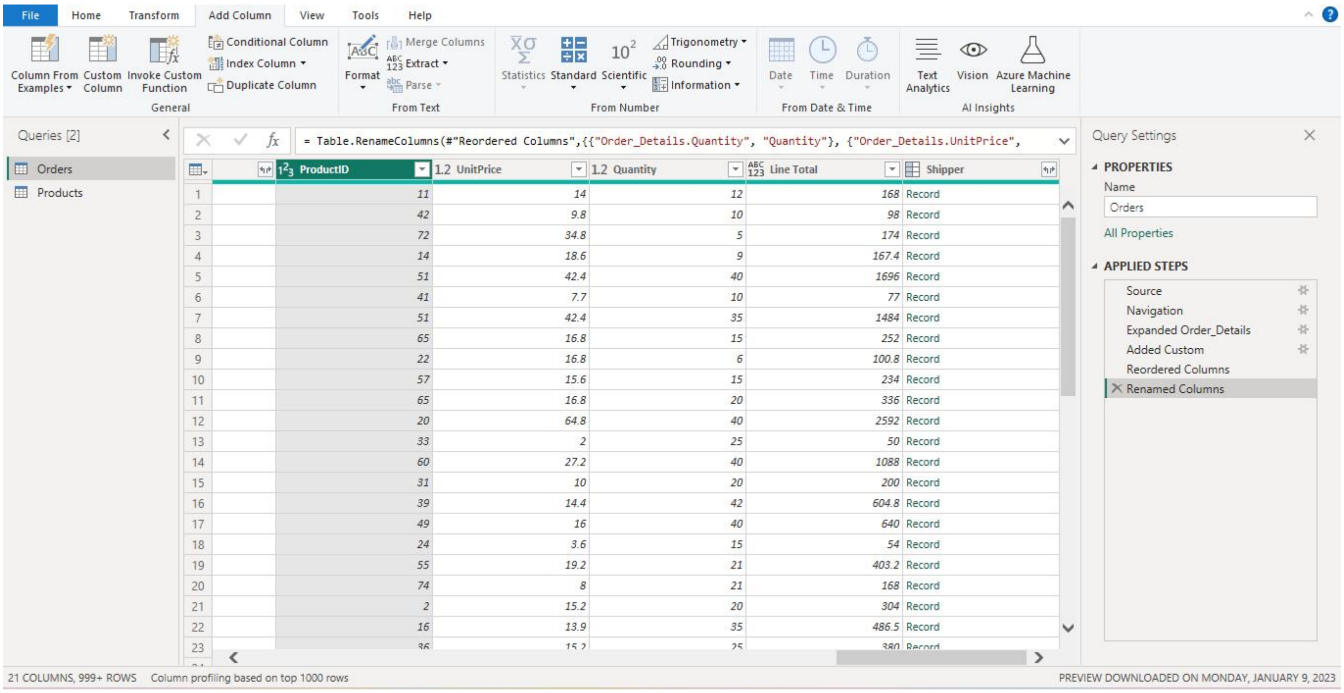
✓ No syntax errors have been detected.

5. Rename and reorder columns in the query

In this step you finish making the model easy to work with when creating reports, byrenaming the final columns and changing their order.

1) In Query Editor, drag the LineTotal column to the left, after ShipCountry.

2) Remove the Order_Details. prefix from the Order_Details.ProductID, Order_Details.UnitPrice and Order_Details.Quantity columns, by double-clicking on each column header, and then deleting that text from the column name.



```
= Table.RenameColumns(#"Reordered Columns",{{"Order_Details.Quantity", "Quantity"}, {"Order_Details.UnitPrice",
```

		ProductID	UnitPrice	Quantity	Line Total	Shipper
1		11	14	12	168	Record
2		42	9.8	10	98	Record
3		72	34.8	5	174	Record
4		14	18.6	9	167.4	Record
5		51	42.4	40	1696	Record
6		41	7.7	10	77	Record
7		51	42.4	35	1484	Record
8		65	16.8	15	252	Record
9		22	16.8	6	100.8	Record
10		57	15.6	15	234	Record
11		65	16.8	20	336	Record
12		20	64.8	40	2592	Record
13		33	2	25	50	Record
14		60	27.2	40	1088	Record
15		31	10	20	200	Record
16		39	14.4	42	604.8	Record
17		49	16	40	640	Record
18		24	3.6	15	54	Record
19		55	19.2	21	403.2	Record
20		74	8	21	168	Record
21		2	15.2	20	304	Record
22		16	13.9	35	486.5	Record
23		36	15.2	25	380	Record

Name: Orders

Applied steps:
- Source
- Navigation
- Expanded Order_Details
- Added Custom
- Reordered Columns
- Renamed Columns

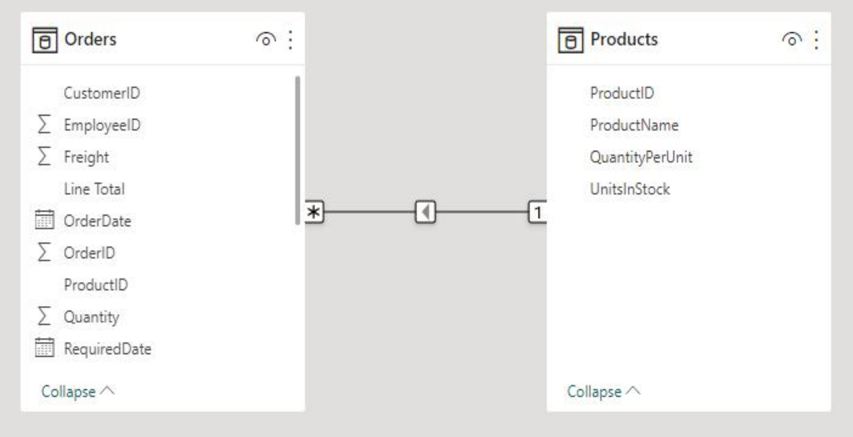
21 COLUMNS, 999+ ROWS   Column profiling based on top 1000 rows   PREVIEW DOWNLOADED ON MONDAY, JANUARY 9, 2023

6) Create Relationship among both the tables

After applying all the changes select close and apply ,go to modeling and then manage relationship.

Select products, orders table will be applied automatically.

Go to the model view section, the relationship has been successful created.



Orders

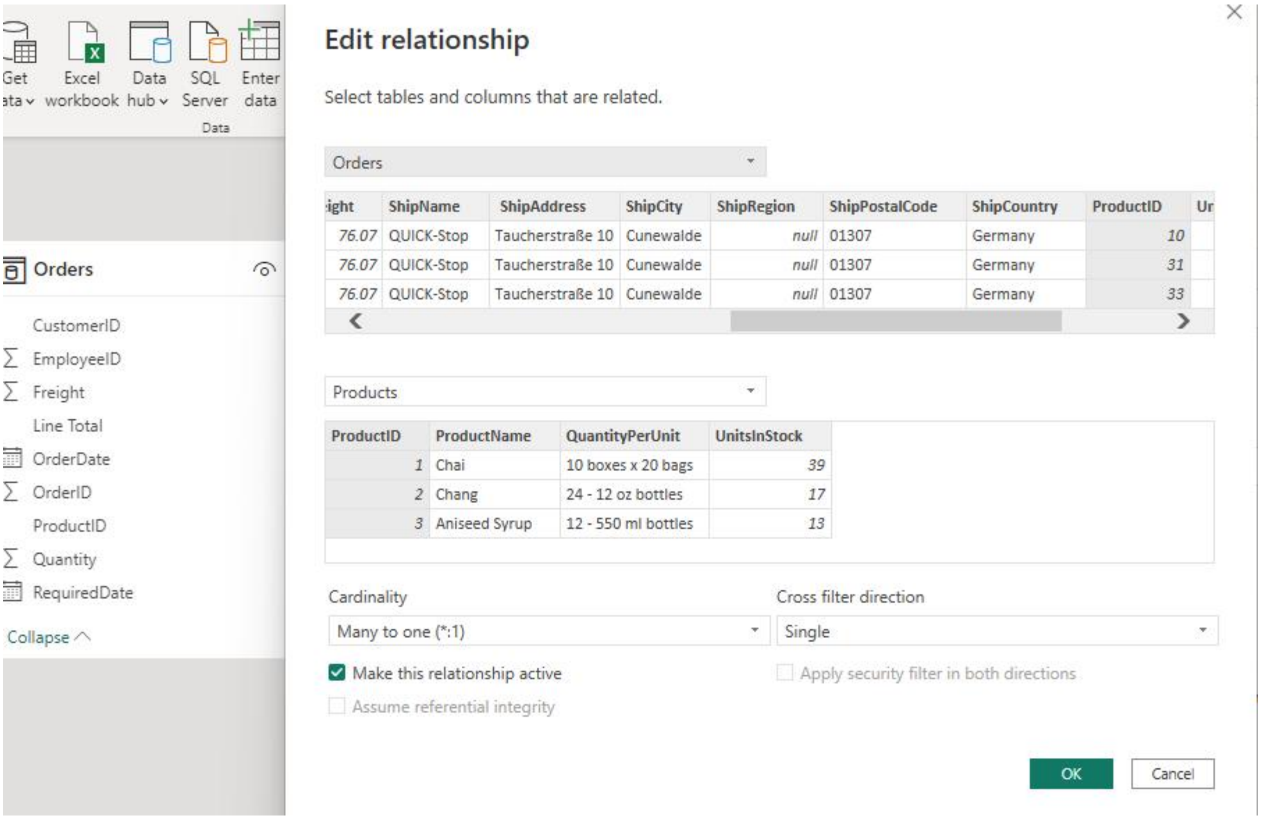
CustomerID
EmployeeID
Freight
Line Total
OrderDate
OrderID
ProductID
Quantity
RequiredDate

Collapse

Products

ProductID
ProductName
QuantityPerUnit
UnitsInStock

Collapse



Orders

CustomerID
EmployeeID
Freight
Line Total
OrderDate
OrderID
ProductID
Quantity
RequiredDate

Collapse

Edit relationship

Select tables and columns that are related.

Orders

ight	ShipName	ShipAddress	ShipCity	ShipRegion	ShipPostalCode	ShipCountry	ProductID	Ur
76.07	QUICK-Stop	Taucherstraße 10	Cunewalde	null	01307	Germany	10	
76.07	QUICK-Stop	Taucherstraße 10	Cunewalde	null	01307	Germany	31	
76.07	QUICK-Stop	Taucherstraße 10	Cunewalde	null	01307	Germany	33	

Products

ProductID	ProductName	QuantityPerUnit	UnitsInStock
1	Chai	10 boxes x 20 bags	39
2	Chang	24 - 12 oz bottles	17
3	Aniseed Syrup	12 - 550 ml bottles	13

Cardinality: Many to one (*:1)

Cross filter direction: Single

☑ Make this relationship active

□ Apply security filter in both directions

□ Assume referential integrity

OK　Cancel

PRACTICAL 5

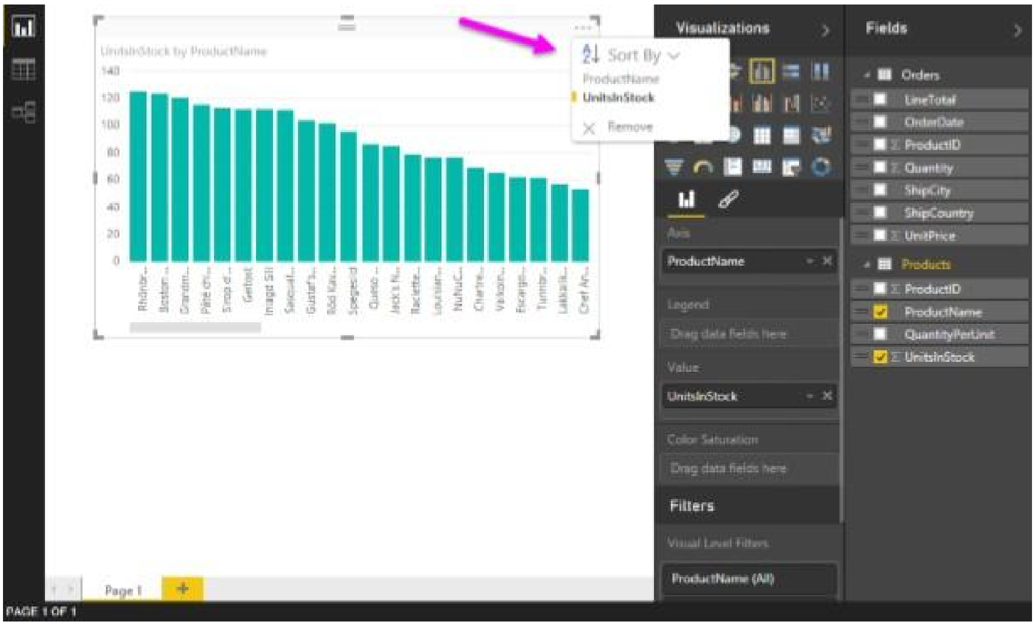
Aim : Data Visualization from ETL Process

Power BI Desktop lets you create a variety of visualizations to gain insights from your data.You can build reports with multiple pages and each page can have multiple visuals. You caninteract with your visualizations to help analyze and understand your data

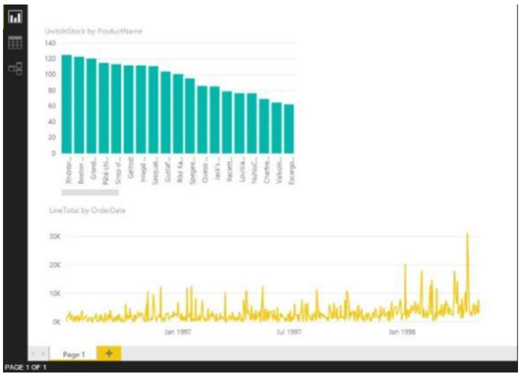
In this task, you create a report based on the data previously loaded. You use the Fields paneto select the columns from which you create the visualizations.

Step 1: Create charts showing Units in Stock by Product and Total Sales by Year

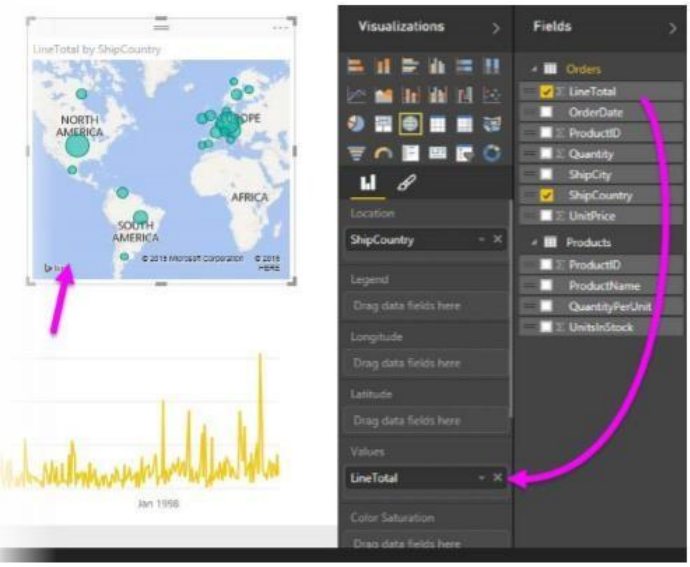
1. Drag UnitsInStock from the Field pane (the Fields pane is along the right of the screen) onto a blank space on the canvas. A Table visualization is created. Next, dragProductName to the Axis box, found in the bottom half of the Visualizations pane. Then we then select Sort By > UnitsInStock using the skittles in the top right corer ofthe visualization.



2. Drag OrderDate to the canvas beneath the first chart, then drag LineTotal (again, from the Fields pane) onto the visual, then select Line Chart. The following visualization iscreated.



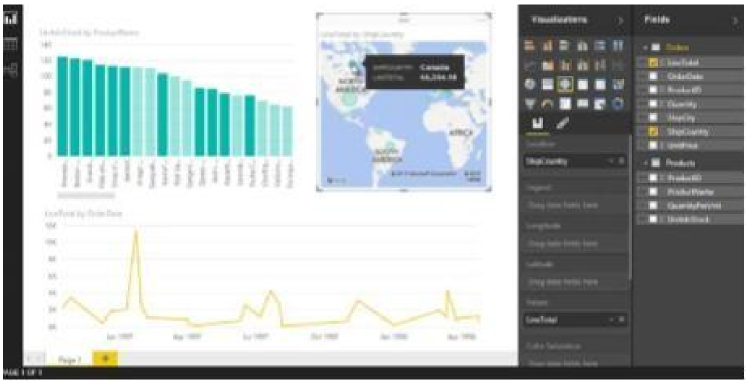
3. Next, drag ShipCountry to a space on the canvas in the top right. Because you selected a geographic field, a map was created automatically. Now drag LineTotal to the Values field; the circles on the map for each country are now relative in size to the LineTotal for orders shipped to that country.



Step 2: Interact with your report visuals to analyze further

Power BI Desktop lets you interact with visuals that cross-highlight and filter each other touncover further trends.

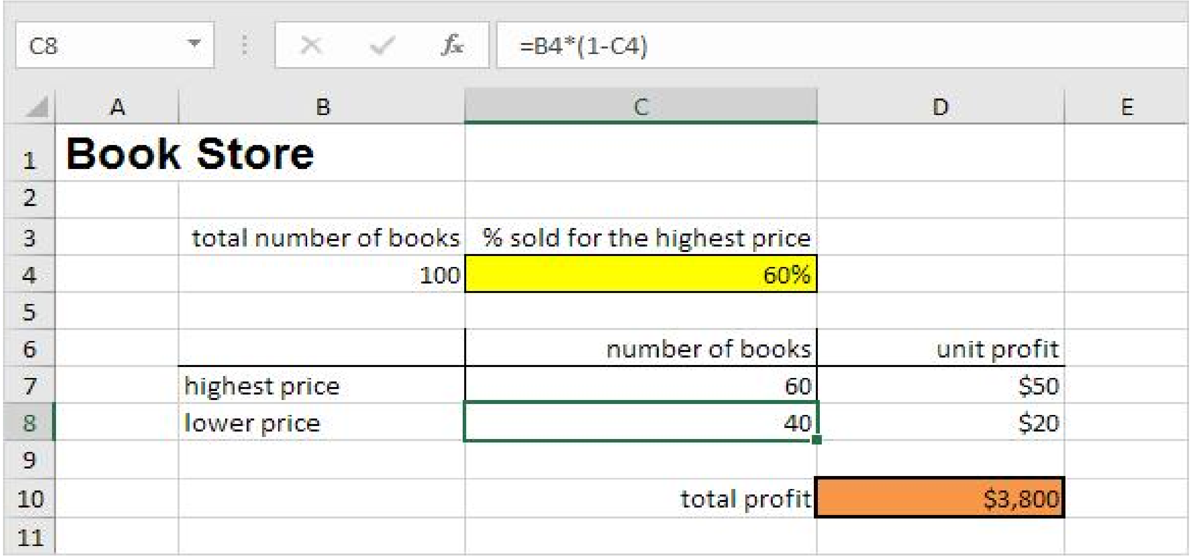
1. Click on the light blue circle centered in Canada. Note how the other visuals are filtered to show Stock (ShipCountry) and Total Orders (LineTotal) just for Canada.



PRACTICAL 6

Aim : What-if analysis using excel for data visualization

A book store and have 100 books in storage. You sell a certain % for the highest price of $50 and a certain % for the lower price of $20.



C8 =B4*(1-C4)

	A	B	C	D	E
1	Book Store				
2					
3		total number of books	% sold for the highest price		
4		100	60%		
5					
6			number of books	unit profit	
7		highest price	60	$50	
8		lower price	40	$20	
9					
10			total profit	$3,800	
11					

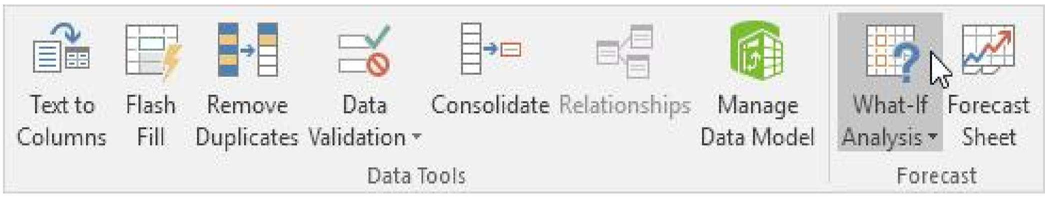
If you sell 60% for the highest price, cell D10 calculates a total profit of 60 * $50 + 40 * $20 = $3800.

Create Different Scenarios

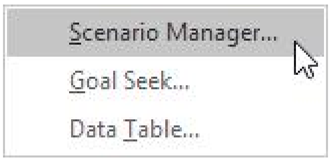
But what if you sell 70% for the highest price? And what if you sell 80% for the highest price? Or 90%, or even 100%? Each different percentage is a different **scenario**. You can use the Scenario Manager to create these scenarios.

Note: You can simply type in a different percentage into cell C4 to see the corresponding result of a scenario in cell D10. However, what-if analysis enables you to easily compare the results of different scenarios. Read on.

1. On the Data tab, in the Forecast group, click What-If Analysis.

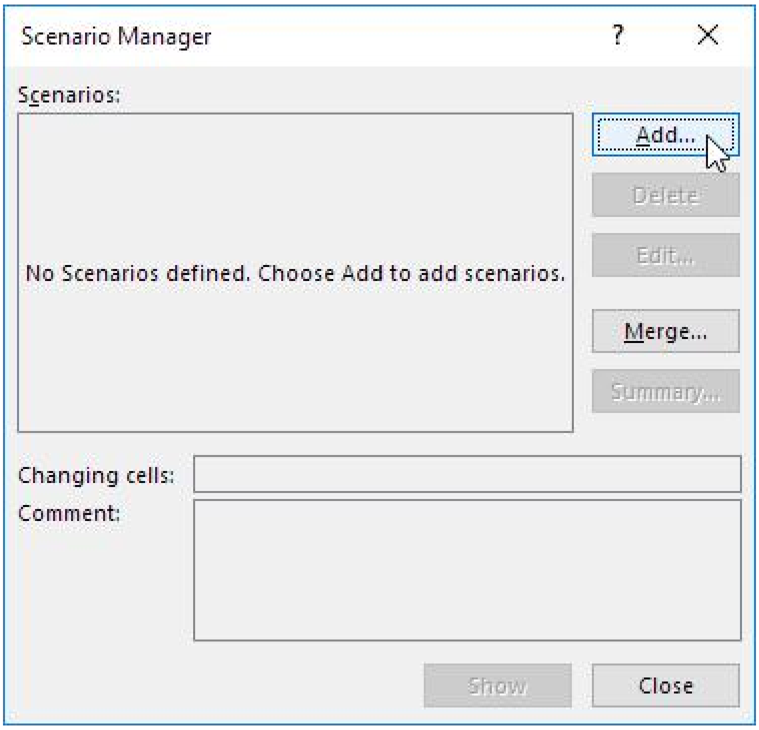


2. Click Scenario Manager.



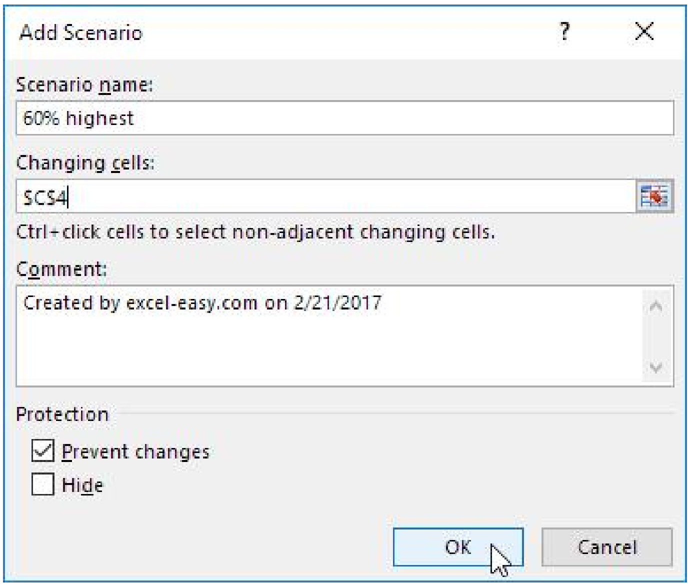
The Scenario Manager dialog box appears.

3. Add a scenario by clicking on Add.



No Scenarios defined. Choose Add to add scenarios.

4. Type a name (60% highest), select cell C4 (% sold for the highest price) for the Changing cells and click on OK.



Scenario name: 60% highest

Changing cells: C4

Ctrl+click cells to select non-adjacent changing cells.

Comment: Created by excel-easy.com on 2/21/2017

Protection

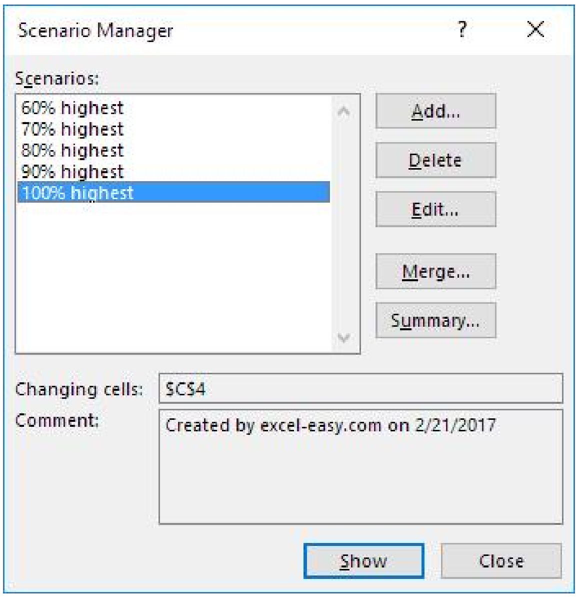
☑ Prevent changes

□ Hide

5. Enter the corresponding value 0.6 and click on OK again.

6. Next, add 4 other scenarios (70%, 80%, 90% and 100%).

Finally, your Scenario Manager should be consistent with the picture below:



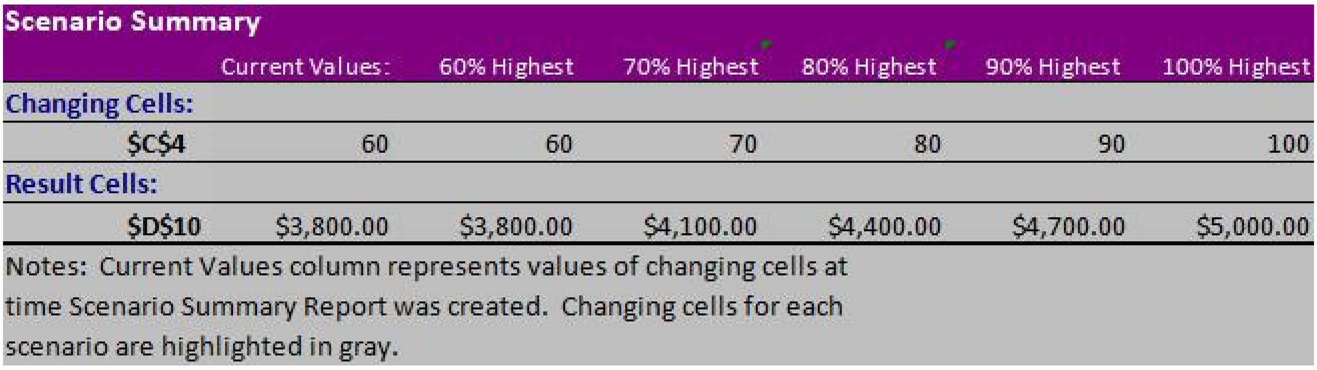
Scenarios:

60% highest
70% highest
80% highest
90% highest
100% highest

Changing cells: C4

Comment: Created by excel-easy.com on 2/21/2017

7. The scenario summary will look like this



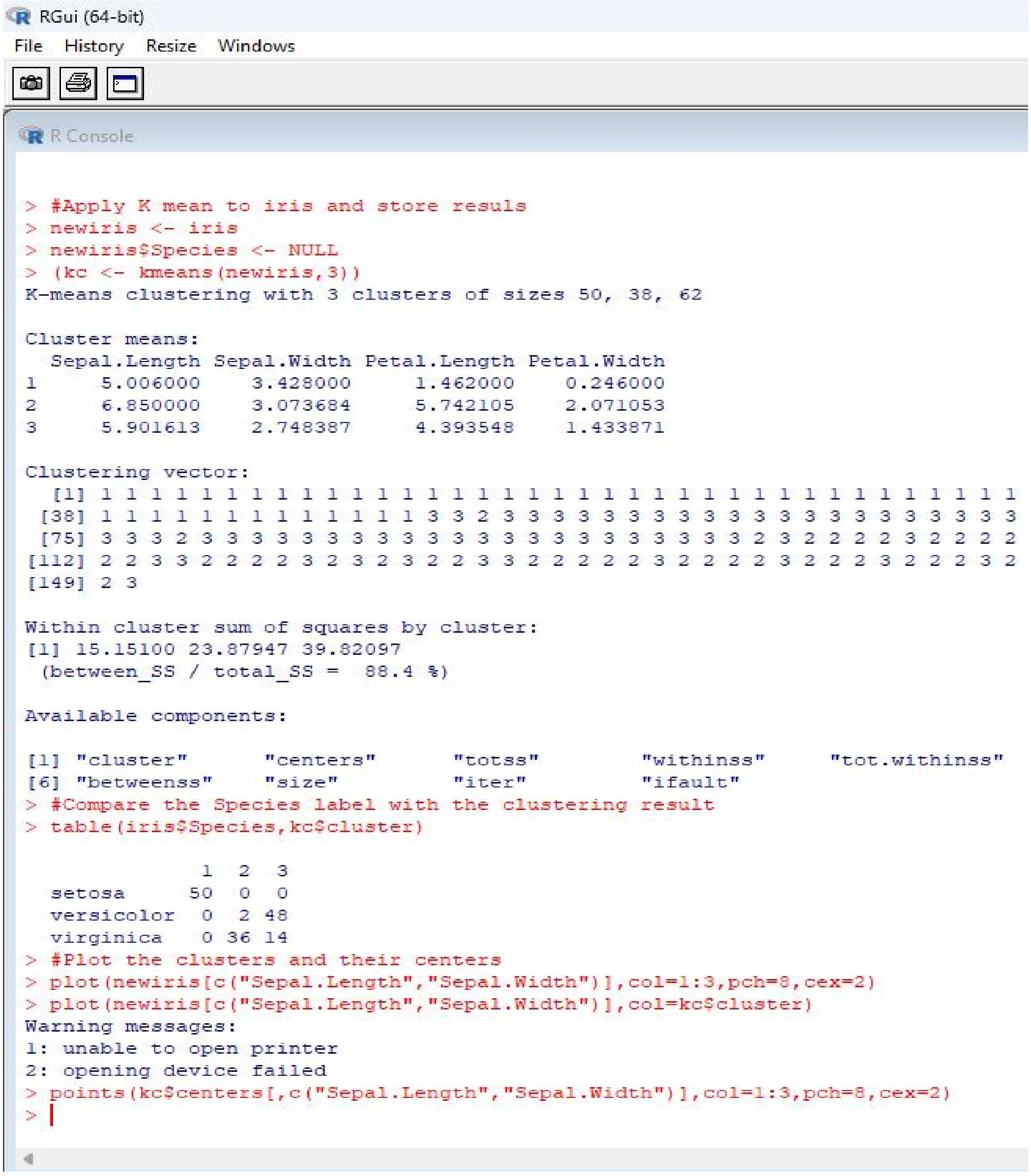
Scenario Summary

	Current Values:	60% Highest	70% Highest	80% Highest	90% Highest	100% Highest
Changing Cells:						
C4	60	60	70	80	90	100
Result Cells:						
D10	$3,800.00	$3,800.00	$4,100.00	$4,400.00	$4,700.00	$5,000.00

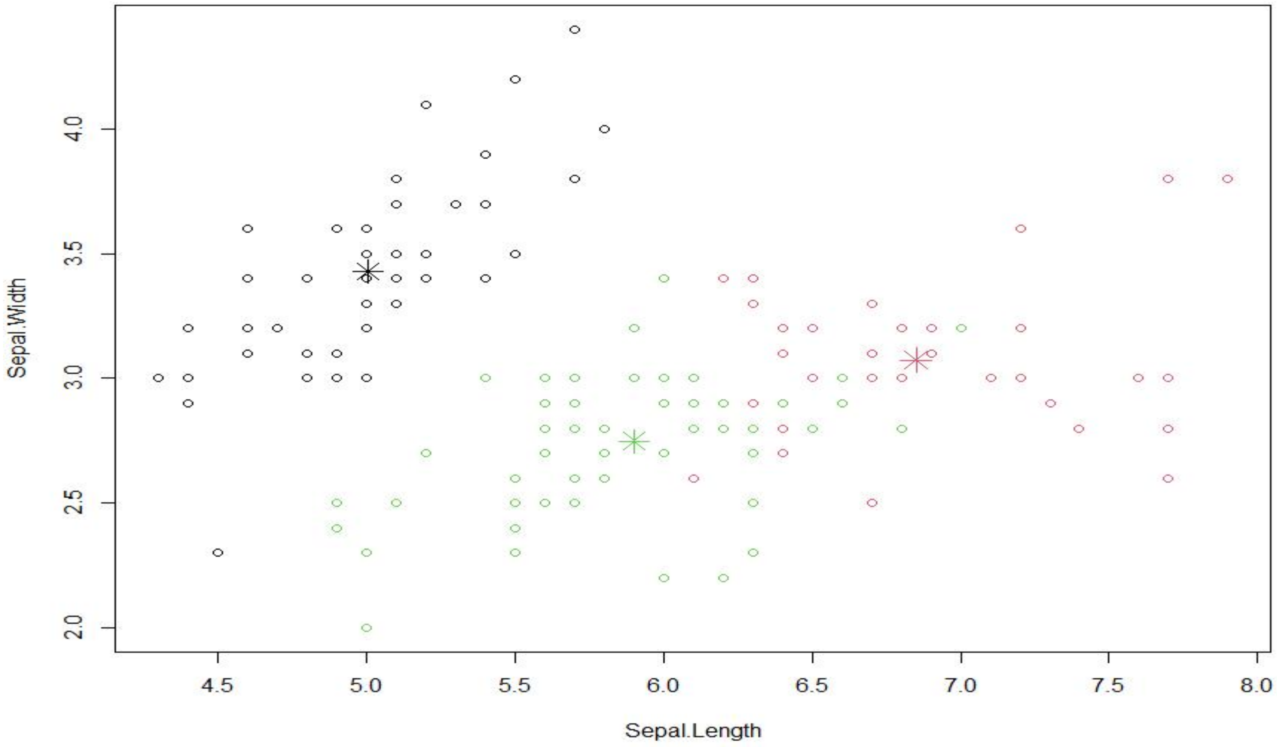
Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.

PRACTICAL 7

Aim : K-means clustering using R



```
> #Apply K mean to iris and store resuls
> newiris <- iris
> newiris$Species <- NULL
> (kc <- kmeans(newiris,3))
K-means clustering with 3 clusters of sizes 50, 38, 62

Cluster means:
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1     5.006000    3.428000     1.462000    0.246000
2     6.850000    3.073684     5.742105    2.071053
3     5.901613    2.748387     4.393548    1.433871

Clustering vector:
  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
 [75] 3 3 3 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 2 2 2 2 3 2 2 2 2
[112] 2 2 3 3 2 2 2 2 3 2 3 2 3 2 2 3 3 2 2 2 2 2 3 2 2 2 2 3 2 2 2 3 2 2 2 3 2
[149] 2 3

Within cluster sum of squares by cluster:
[1] 15.15100 23.87947 39.82097
 (between_SS / total_SS =  88.4 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
> #Compare the Species label with the clustering result
> table(iris$Species,kc$cluster)

             1  2  3
  setosa    50  0  0
  versicolor 0  2 48
  virginica  0 36 14
> #Plot the clusters and their centers
> plot(newiris[c("Sepal.Length","Sepal.Width")],col=1:3,pch=8,cex=2)
> plot(newiris[c("Sepal.Length","Sepal.Width")],col=kc$cluster)
Warning messages:
1: unable to open printer
2: opening device failed
> points(kc$centers[,c("Sepal.Length","Sepal.Width")],col=1:3,pch=8,cex=2)
> 
```

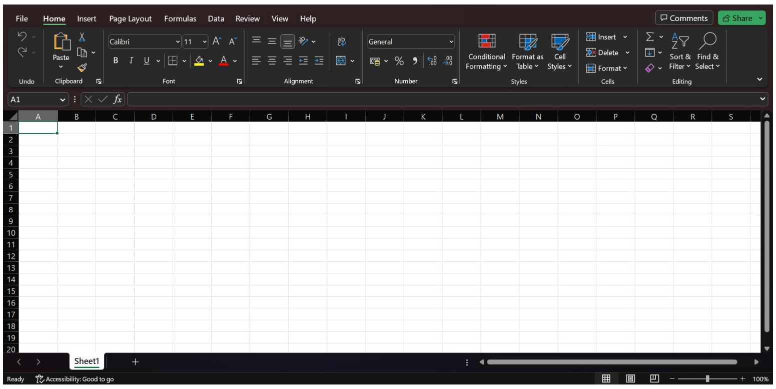


Sepal.Width
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
Sepal.Length

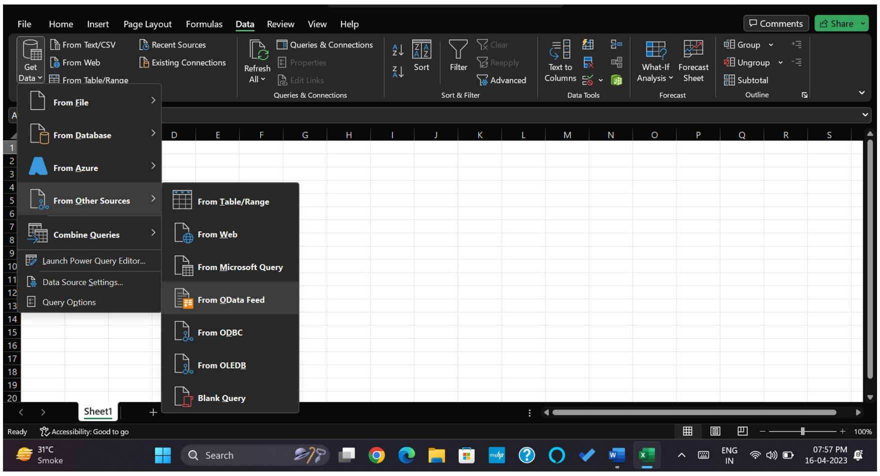
PRACTICAL 8

Aim : Pivot table using excel

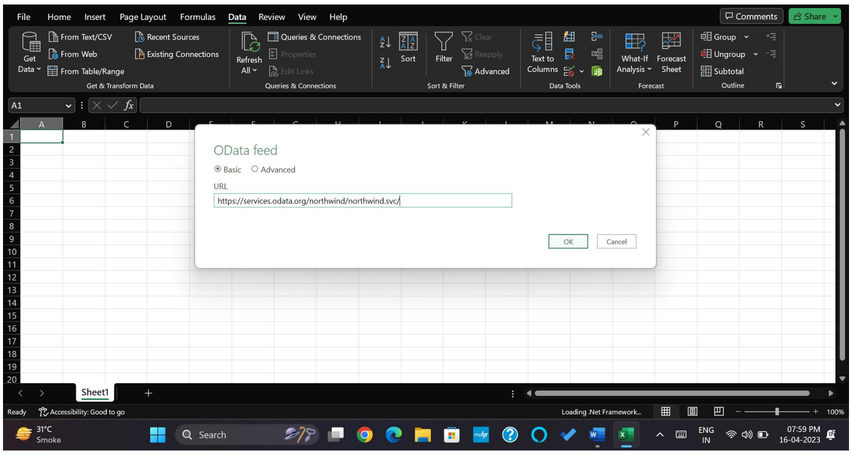
Step 1: Open excel SpreadSheet.



Step 2: Go to Data Tab > Get Data> From Other Sources > From OData Feed



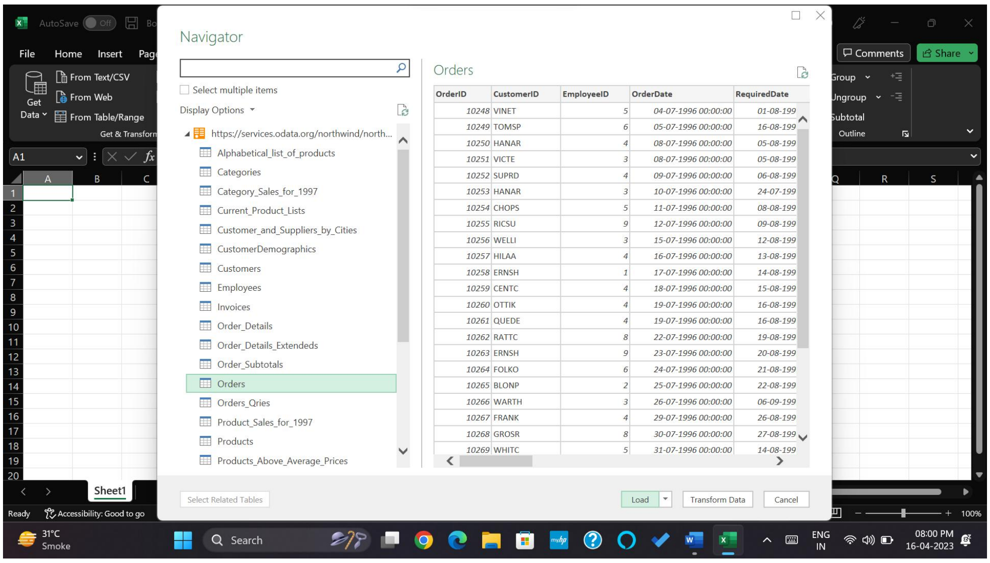
Step 3: Past the following link *https://services.odata.org/northwind/northwind.svc/* and hit OK


OData feed

● Basic ○ Advanced

URL: https://services.odata.org/northwind/northwind.svc/

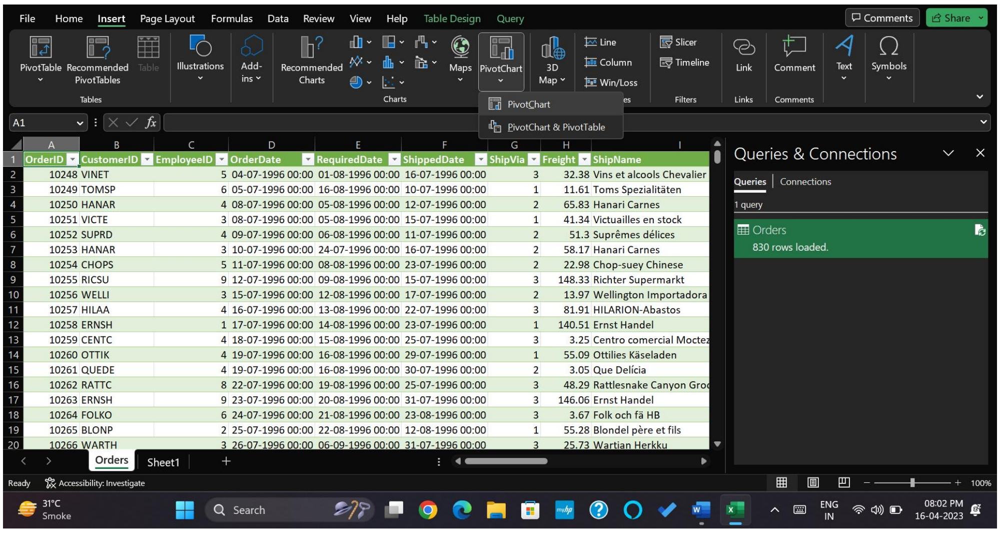
Step 4: Select Any table from left side of the window and hit Load



Orders

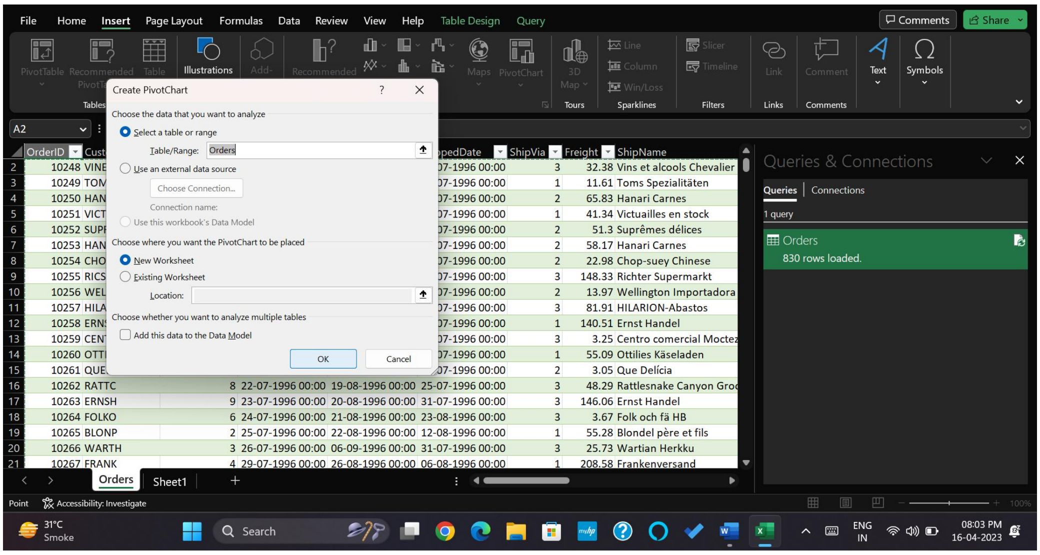
OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate
10248	VINET	5	04-07-1996 00:00:00	01-08-199
10249	TOMSP	6	05-07-1996 00:00:00	16-08-199
10250	HANAR	4	08-07-1996 00:00:00	05-08-199
10251	VICTE	3	08-07-1996 00:00:00	05-08-199
10252	SUPRD	4	09-07-1996 00:00:00	06-08-199
10253	HANAR	3	10-07-1996 00:00:00	24-07-199
10254	CHOPS	5	11-07-1996 00:00:00	08-08-199
10255	RICSU	9	12-07-1996 00:00:00	09-08-199
10256	WELLI	3	15-07-1996 00:00:00	12-08-199
10257	HILAA	4	16-07-1996 00:00:00	13-08-199
10258	ERNSH	1	17-07-1996 00:00:00	14-08-199
10259	CENTC	4	18-07-1996 00:00:00	15-08-199
10260	OTTIK	4	19-07-1996 00:00:00	16-08-199
10261	QUEDE	4	19-07-1996 00:00:00	16-08-199
10262	RATTC	8	22-07-1996 00:00:00	19-08-199
10263	ERNSH	9	23-07-1996 00:00:00	20-08-199
10264	FOLKO	6	24-07-1996 00:00:00	21-08-199
10265	BLONP	2	25-07-1996 00:00:00	22-08-199
10266	WARTH	3	26-07-1996 00:00:00	06-09-199
10267	FRANK	4	29-07-1996 00:00:00	26-08-199
10268	GROSR	8	30-07-1996 00:00:00	27-08-199
10269	WHITC	5	31-07-1996 00:00:00	14-08-199

Step 5: Go to Insert Tab and select PivotChart



OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate	ShippedDate	ShipVia	Freight	ShipName
10248	VINET	5	04-07-1996 00:00	01-08-1996 00:00	16-07-1996 00:00	3	32.38	Vins et alcools Chevalier
10249	TOMSP	6	05-07-1996 00:00	16-08-1996 00:00	10-07-1996 00:00	1	11.61	Toms Spezialitäten
10250	HANAR	4	08-07-1996 00:00	05-08-1996 00:00	12-07-1996 00:00	2	65.83	Hanari Carnes
10251	VICTE	3	08-07-1996 00:00	05-08-1996 00:00	15-07-1996 00:00	1	41.34	Victuailles en stock
10252	SUPRD	4	09-07-1996 00:00	06-08-1996 00:00	11-07-1996 00:00	2	51.3	Suprêmes délices
10253	HANAR	3	10-07-1996 00:00	24-07-1996 00:00	16-07-1996 00:00	2	58.17	Hanari Carnes
10254	CHOPS	5	11-07-1996 00:00	08-08-1996 00:00	23-07-1996 00:00	2	22.98	Chop-suey Chinese
10255	RICSU	9	12-07-1996 00:00	09-08-1996 00:00	15-07-1996 00:00	3	148.33	Richter Supermarkt
10256	WELLI	3	15-07-1996 00:00	12-08-1996 00:00	17-07-1996 00:00	2	13.97	Wellington Importadora
10257	HILAA	4	16-07-1996 00:00	13-08-1996 00:00	22-07-1996 00:00	3	81.91	HILARION-Abastos
10258	ERNSH	1	17-07-1996 00:00	14-08-1996 00:00	23-07-1996 00:00	1	140.51	Ernst Handel
10259	CENTC	4	18-07-1996 00:00	15-08-1996 00:00	25-07-1996 00:00	3	3.25	Centro comercial Moctez
10260	OTTIK	4	19-07-1996 00:00	16-08-1996 00:00	29-07-1996 00:00	1	55.09	Ottilies Käseladen
10261	QUEDE	4	19-07-1996 00:00	16-08-1996 00:00	30-07-1996 00:00	2	3.05	Que Delícia
10262	RATTC	8	22-07-1996 00:00	19-08-1996 00:00	25-07-1996 00:00	3	48.29	Rattlesnake Canyon Groc
10263	ERNSH	9	23-07-1996 00:00	20-08-1996 00:00	31-07-1996 00:00	3	146.06	Ernst Handel
10264	FOLKO	6	24-07-1996 00:00	21-08-1996 00:00	23-08-1996 00:00	3	3.67	Folk och fä HB
10265	BLONP	2	25-07-1996 00:00	22-08-1996 00:00	12-08-1996 00:00	1	55.28	Blondel père et fils
10266	WARTH	3	26-07-1996 00:00	06-09-1996 00:00	31-07-1996 00:00	3	25.73	Wartian Herkku

Step 6: New Dialog Box will appear, hit ok.



Create PivotChart

Choose the data that you want to analyze

● Select a table or range

Table/Range: Orders

○ Use an external data source

Choose Connection...

Connection name:

○ Use this workbook's Data Model

Choose where you want the PivotChart to be placed

● New Worksheet

○ Existing Worksheet

Location:

Choose whether you want to analyze multiple tables

□ Add this data to the Data Model

OK Cancel

Queries & Connections

Queries | Connections

1 query

Orders

830 rows loaded.

Step 7: A new interface will appear.

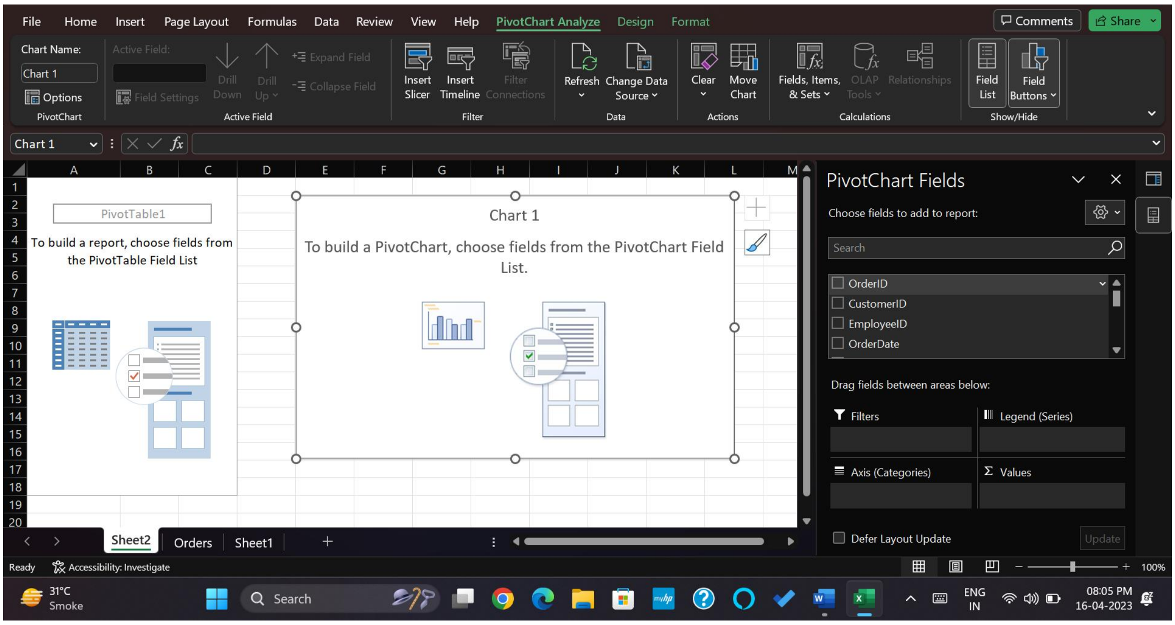


Chart 1

To build a PivotChart, choose fields from the PivotChart Field List.

PivotTable1

To build a report, choose fields from the PivotTable Field List

PivotChart Fields

Choose fields to add to report:

□ OrderID

□ CustomerID

□ EmployeeID

□ OrderDate

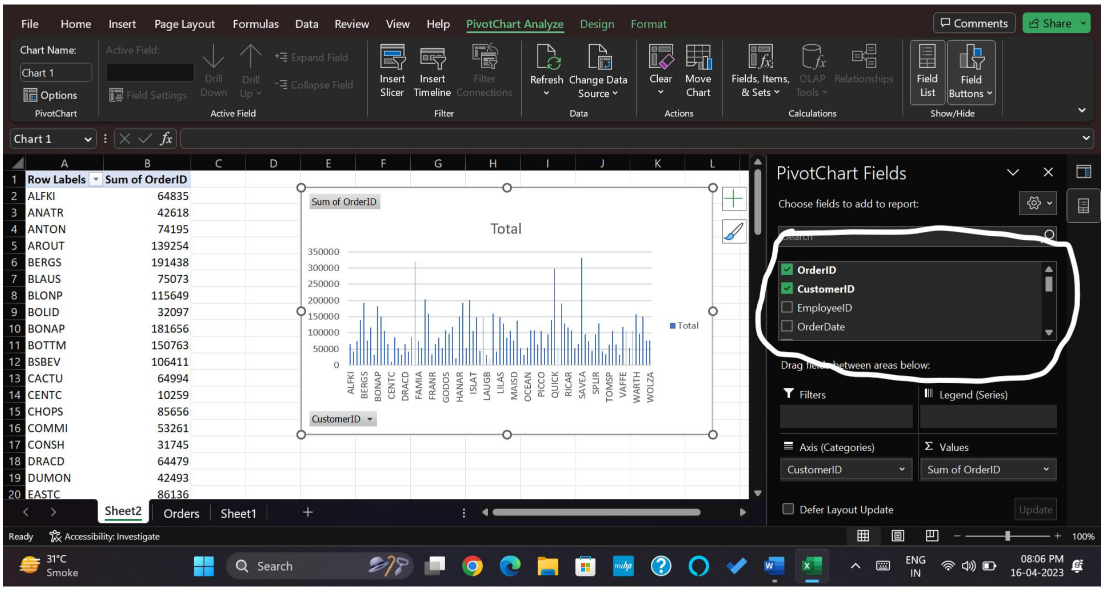
Drag fields between areas below:

Filters | Legend (Series)

Axis (Categories) | Values

□ Defer Layout Update

Step 8: Select required Columns from PivotChart Fields from right side of the window and your PivotChart is ready.



Row Labels	Sum of OrderID
ALFKI	64835
ANATR	42618
ANTON	74195
AROUT	139254
BERGS	191438
BLAUS	75073
BLONP	115649
BOLID	32097
BONAP	181656
BOTTM	150763
BSBEV	106411
CACTU	64994
CENTC	10259
CHOPS	85656
COMMI	53261
CONSH	31745
DRACD	64479
DUMON	42493
EASTC	86136

PivotChart Fields

Choose fields to add to report:

Search

☑ OrderID
☑ CustomerID
☐ EmployeeID
☐ OrderDate

Drag fields between areas below:

Filters

Legend (Series)

Axis (Categories): CustomerID

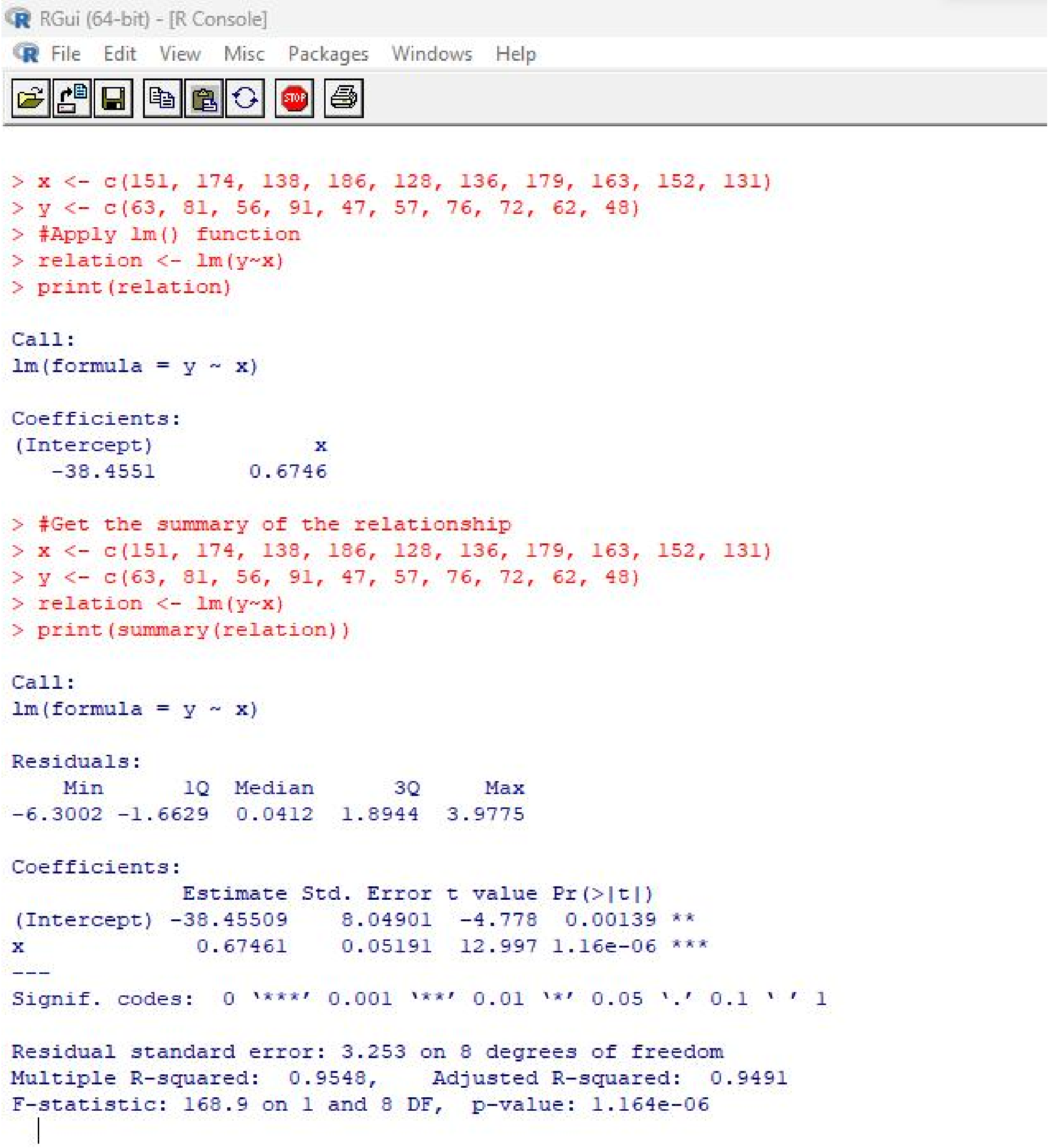
Values: Sum of OrderID

☐ Defer Layout Update

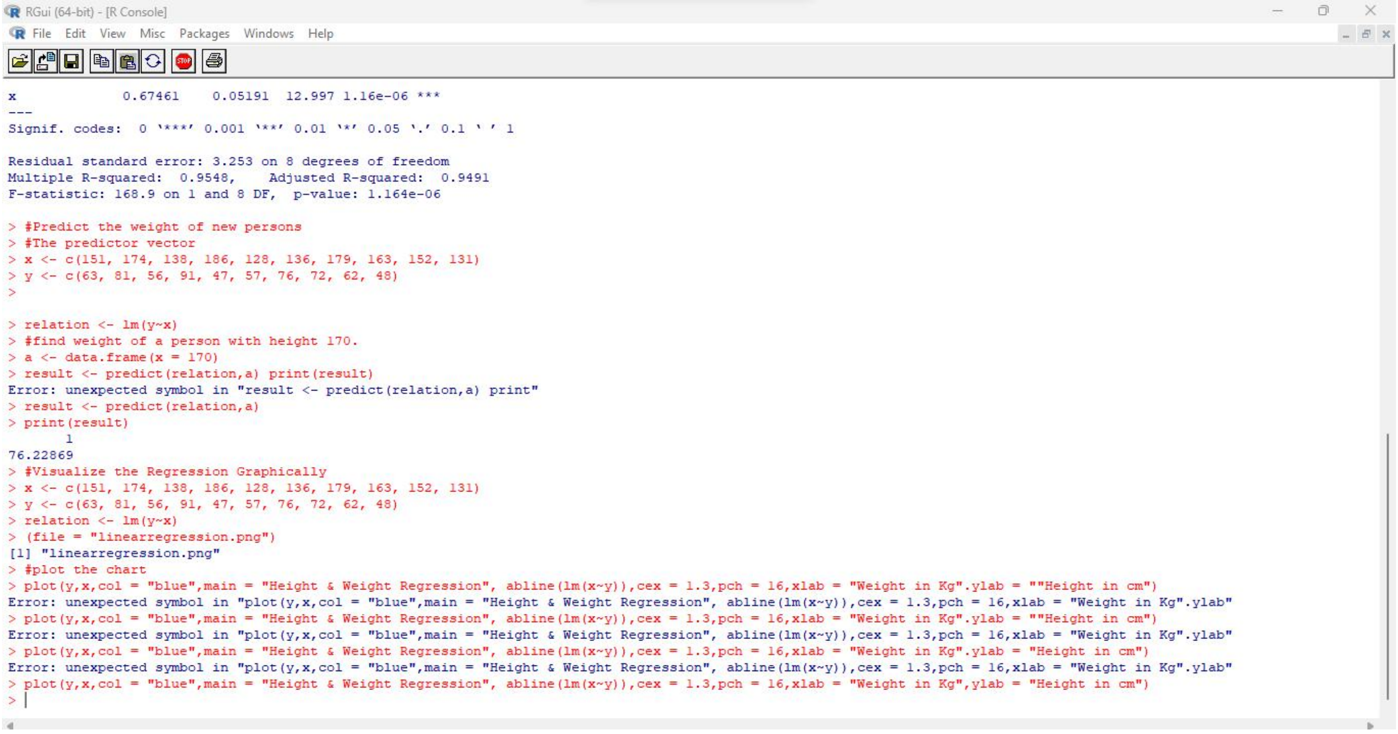
Update

PRACTICAL 9

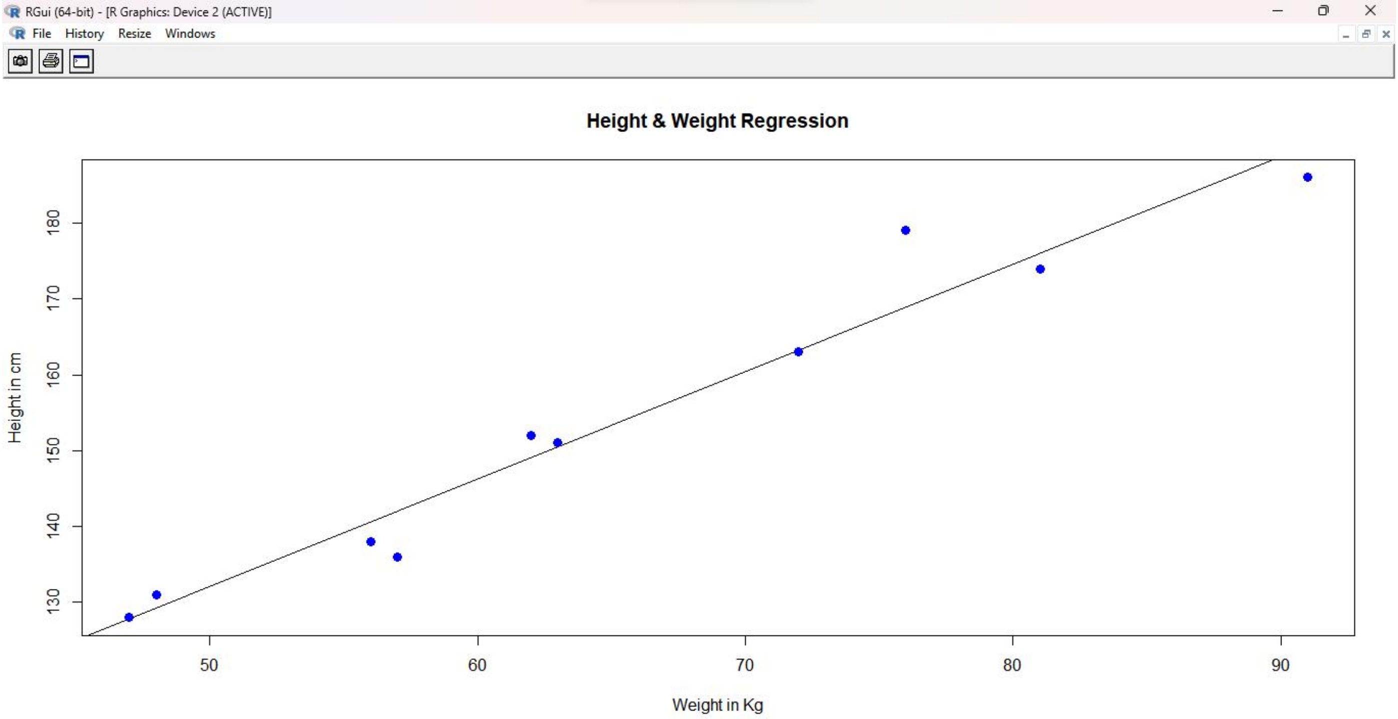
Aim : Height & Weight Regression using R


RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
STOP

```
> x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)
> y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
> #Apply lm() function
> relation <- lm(y~x)
> print(relation)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
   -38.4551       0.6746

> #Get the summary of the relationship
> x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)
> y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
> relation <- lm(y~x)
> print(summary(relation))

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.3002 -1.6629  0.0412  1.8944  3.9775 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -38.45509    8.04901  -4.778  0.00139 ** 
x             0.67461    0.05191  12.997 1.16e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.253 on 8 degrees of freedom
Multiple R-squared:  0.9548,    Adjusted R-squared:  0.9491 
F-statistic: 168.9 on 1 and 8 DF,  p-value: 1.164e-06
```


RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

```
x              0.67461    0.05191  12.997 1.16e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.253 on 8 degrees of freedom
Multiple R-squared:  0.9548,    Adjusted R-squared:  0.9491
F-statistic: 168.9 on 1 and 8 DF,  p-value: 1.164e-06

> #Predict the weight of new persons
> #The predictor vector
> x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)
> y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
>

> relation <- lm(y~x)
> #find weight of a person with height 170.
> a <- data.frame(x = 170)
> result <- predict(relation,a) print(result)
Error: unexpected symbol in "result <- predict(relation,a) print"
> result <- predict(relation,a)
> print(result)
       1
76.22869
> #Visualize the Regression Graphically
> x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)
> y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
> relation <- lm(y~x)
> (file = "linearregression.png")
[1] "linearregression.png"
> #plot the chart
> plot(y,x,col = "blue",main = "Height & Weight Regression", abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg".ylab = ""Height in cm")
Error: unexpected symbol in "plot(y,x,col = "blue",main = "Height & Weight Regression", abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg".ylab"
> plot(y,x,col = "blue",main = "Height & Weight Regression", abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg".ylab = ""Height in cm")
Error: unexpected symbol in "plot(y,x,col = "blue",main = "Height & Weight Regression", abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg".ylab"
> plot(y,x,col = "blue",main = "Height & Weight Regression", abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg".ylab = "Height in cm")
Error: unexpected symbol in "plot(y,x,col = "blue",main = "Height & Weight Regression", abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg".ylab"
> plot(y,x,col = "blue",main = "Height & Weight Regression", abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg",ylab = "Height in cm")
> |
```


RGui (64-bit) - [R Graphics: Device 2 (ACTIVE)]
File History Resize Windows
Height & Weight Regression
Height in cm
Weight in Kg