

Name: Dharmit Shah.

Roll no: 31011020018.

Business Intelligence

[JOURNAL]

Name: Dharmit Shah.

Roll no: 31011020018.

Index

1.	Import the legacy data from different sources such as (Excel, SQLServer, Oracle etc.) and load in the target system.
2.	Perform the Extraction Transformation and Loading (ETL) process to construct the database in the SQL Server / Power BI.
3.	Create the cube with suitable dimension and fact tables based on OLAP
4.	Apply the what-if Analysis for data visualization
5.	Perform the data classification using classification algorithm
6.	K-Means clustering using R
7.	Prediction Using Linear Regression
8.	Perform the logistic regression on the given data warehouse data

Name: Dharmit Shah.

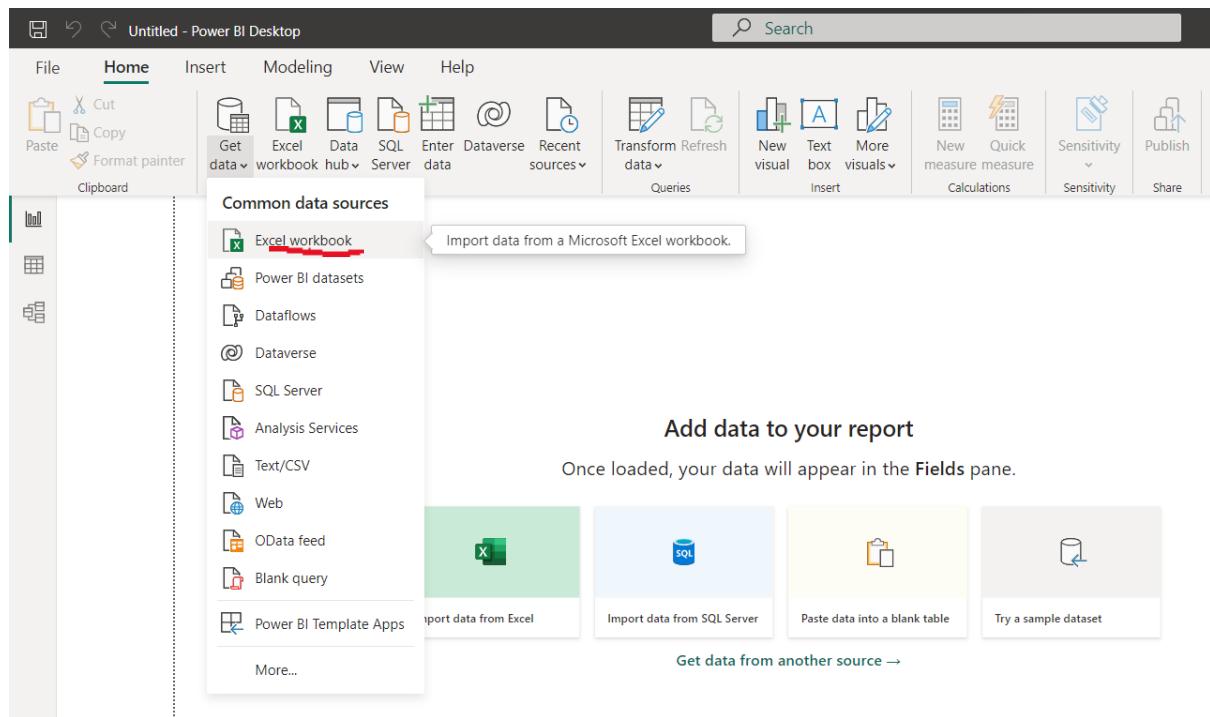
Roll no: 31011020018.

Practical 1

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

Step 1: Importing Excel Data to Power BI

**Open power BI ->Click on Get Data following list will be displayed
->select Excel**



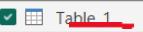
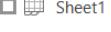
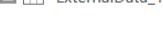
Select required file and click on Open, Navigator screen appears -> Select file and click on Load ->

Name: Dharmit Shah.

Roll no: 31011020018.

Navigator

Display Options  

Products.xlsx [3]
  Table_1
  Sheet1
  ExternalData_1

Table_1

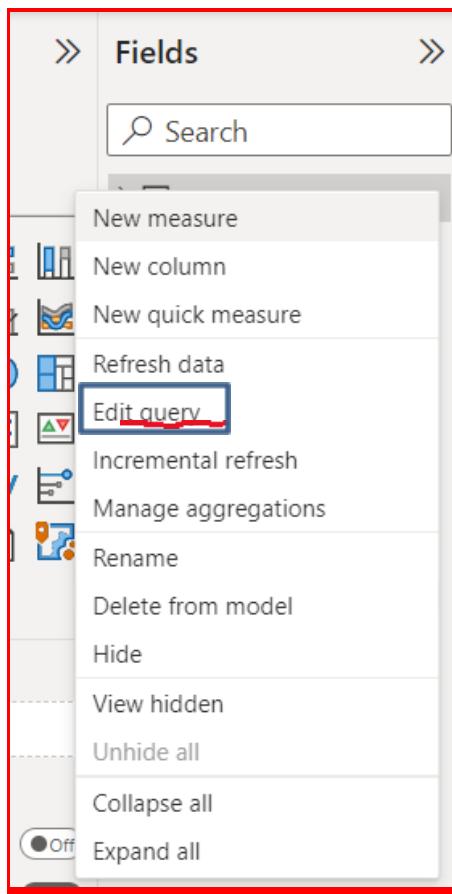
ProductID	ProductName	SupplierID	CategoryID	Quan
1	Chai	1	1	10
2	Chang	1	1	24
3	Aniseed Syrup	1	2	12
4	Chef Anton's Cajun Seasoning	2	2	48
5	Chef Anton's Gumbo Mix	2	2	36
6	Grandma's Boysenberry Spread	3	2	12
7	Uncle Bob's Organic Dried Pears	3	7	12
8	Northwoods Cranberry Sauce	3	2	12
9	Mishi Kobe Niku	4	6	18
10	Ikura	4	8	12
11	Queso Cabrales	5	4	1
12	Queso Manchego La Pastora	5	4	10
13	Konbu	6	8	21
14	Tofu	6	7	40
15	Genen Shouyu	6	2	24
16	Pavlova	7	3	32
17	Alice Mutton	7	6	20
18	Carnarvon Tigers	7	8	16
19	Teatime Chocolate Biscuits	8	3	10
20	Sir Rodney's Marmalade	8	3	30
21	Sir Rodney's Scones	8	3	24
22	Gustaf's Knäckebröd	9	5	24
23	Tunnbröd	9	5	12

Name: Dharmit Shah.

Roll no: 31011020018.



Power query editor appears

The screenshot shows the Power Query Editor window with the following details:

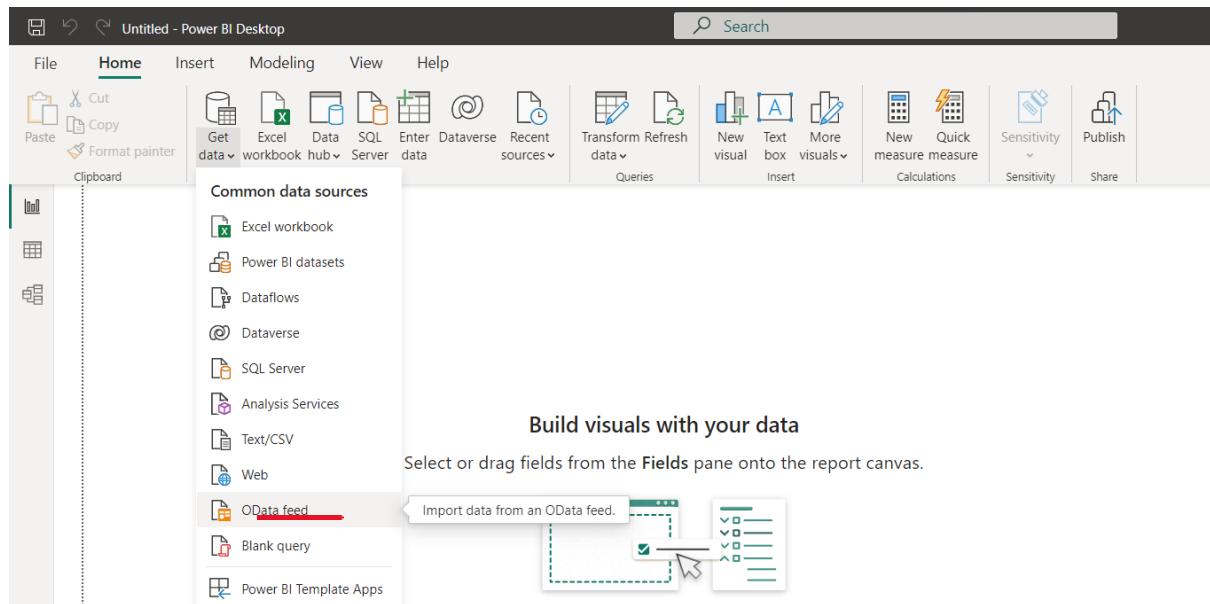
- File Bar:** Untitled - Power Query Editor, File, Home, Transform, Add Column, View, Tools, Help.
- Toolbars:** Close & Apply, New Source, Recent Sources, Enter Data, Data source settings, Manage Parameters, Refresh, Advanced Editor, Properties, Choose Columns, Remove Columns, Keep Rows, Remove Rows, Split Column, Group By, Use First Row as Headers, Data Type: Whole Number, Merge Queries, Append Queries, Combine Files.
- Queries:** Queries [1] - Table_1.
- Table Preview:** Shows the structure of Table_1 with columns: ProductID, ProductName, SupplierID, CategoryID, QuantityPerUnit, and UnitPrice.
- Properties Panel:** Name: Table_1, All Properties.
- Applied Steps Panel:** Source, Navigation, Changed Type.
- Bottom Status:** 10 COLUMNS, 77 ROWS, Column profiling based on top 1000 rows, PREVIEW DOWNLOADED AT 13:03.

Step 2: Importing ODATA to Power BI

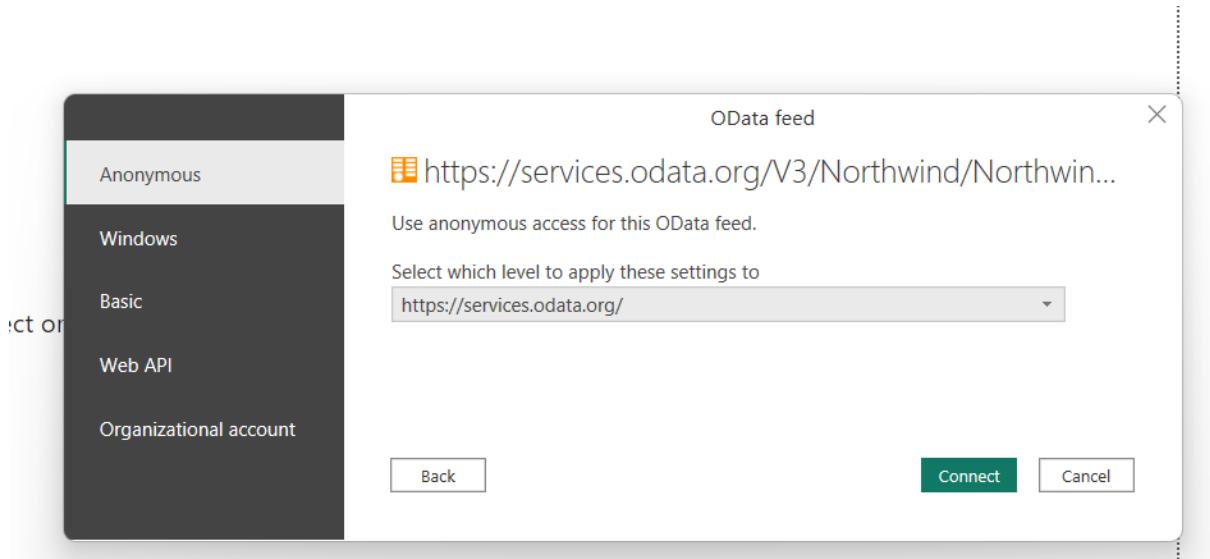
Again, go to Power BI and Get Data and select OData feed

Name: Dharmit Shah.

Roll no: 31011020018.



Paste URL as (<https://services.odata.org/V3/Northwind/Northwind.svc/>) ->Click OK ->connect



Name: Dharmit Shah.

Roll no: 31011020018.

Select Orders table ->Click on Load

Navigator

Display Options ▾

- https://services.odata.org/V3/Northwind/Northwind.svc/
- Alphabetical_list_of_products
- Categories
- Category_Sales_for_1997
- Current_Product_Lists
- Customer_and_Suppliers_by_Cities
- CustomerDemographics
- Customers
- Employees
- Invoices
- Order_Details
- Order_Details_Extendeds
- Order_Subtotals
- Orders**
- Orders_Qries
- Product_Sales_for_1997
- Products
- Products_Above_Average_Prices
- Products_by_Categories
- Regions

Orders

OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate
10248	VINET	5	04-07-1996 00:00:00	01-08-1996 00:00:00
10249	TOMSP	6	05-07-1996 00:00:00	16-08-1996 00:00:00
10250	HANAR	4	08-07-1996 00:00:00	05-08-1996 00:00:00
10251	VICTE	3	08-07-1996 00:00:00	05-08-1996 00:00:00
10252	SUPRD	4	09-07-1996 00:00:00	06-08-1996 00:00:00
10253	HANAR	3	10-07-1996 00:00:00	24-07-1996 00:00:00
10254	CHOPS	5	11-07-1996 00:00:00	08-08-1996 00:00:00
10255	RICSU	9	12-07-1996 00:00:00	09-08-1996 00:00:00
10256	WELLI	3	15-07-1996 00:00:00	12-08-1996 00:00:00
10257	HILAA	4	16-07-1996 00:00:00	13-08-1996 00:00:00
10258	ERNSH	1	17-07-1996 00:00:00	14-08-1996 00:00:00
10259	CENTC	4	18-07-1996 00:00:00	15-08-1996 00:00:00
10260	OTTIK	4	19-07-1996 00:00:00	16-08-1996 00:00:00
10261	QUED	4	19-07-1996 00:00:00	16-08-1996 00:00:00
10262	RATTC	8	22-07-1996 00:00:00	19-08-1996 00:00:00
10263	ERNSH	9	23-07-1996 00:00:00	20-08-1996 00:00:00
10264	FOLKO	6	24-07-1996 00:00:00	21-08-1996 00:00:00
10265	BLONP	2	25-07-1996 00:00:00	22-08-1996 00:00:00
10266	WARTH	3	26-07-1996 00:00:00	06-09-1996 00:00:00
10267	FRANK	4	29-07-1996 00:00:00	26-08-1996 00:00:00
10268	GROSR	8	30-07-1996 00:00:00	27-08-1996 00:00:00
10269	WHITC	5	31-07-1996 00:00:00	14-08-1996 00:00:00
10270	WARTH	1	01-08-1996 00:00:00	29-08-1996 00:00:00

Select Related Tables

OUTPUT

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Queries [2]

Table_1

Orders

OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate	ShippedDate	ShipVia
1	10248	VINET	5	04-07-1996 00:00:00	01-08-1996 00:00:00	16-07-1996 00:00:00
2	10249	TOMSP	6	05-07-1996 00:00:00	16-08-1996 00:00:00	10-07-1996 00:00:00
3	10250	HANAR	4	08-07-1996 00:00:00	05-08-1996 00:00:00	12-07-1996 00:00:00
4	10251	VICTE	3	08-07-1996 00:00:00	05-08-1996 00:00:00	15-07-1996 00:00:00
5	10252	SUPRD	4	09-07-1996 00:00:00	06-08-1996 00:00:00	11-07-1996 00:00:00
6	10253	HANAR	3	10-07-1996 00:00:00	24-07-1996 00:00:00	16-07-1996 00:00:00
7	10254	CHOPS	5	11-07-1996 00:00:00	08-08-1996 00:00:00	23-07-1996 00:00:00
8	10255	RICSU	9	12-07-1996 00:00:00	09-08-1996 00:00:00	15-07-1996 00:00:00
9	10256	WELLI	3	15-07-1996 00:00:00	12-08-1996 00:00:00	17-07-1996 00:00:00
10	10257	HILAA	4	16-07-1996 00:00:00	13-08-1996 00:00:00	22-07-1996 00:00:00
11	10258	ERNSH	1	17-07-1996 00:00:00	14-08-1996 00:00:00	23-07-1996 00:00:00
12	10259	CENTC	4	18-07-1996 00:00:00	15-08-1996 00:00:00	25-07-1996 00:00:00
13	10260	OTTIK	4	19-07-1996 00:00:00	16-08-1996 00:00:00	29-07-1996 00:00:00
14	10261	QUED	4	19-07-1996 00:00:00	16-08-1996 00:00:00	30-07-1996 00:00:00
15	10262	RATTC	8	22-07-1996 00:00:00	19-08-1996 00:00:00	25-07-1996 00:00:00
16	10263	ERNSH	9	23-07-1996 00:00:00	20-08-1996 00:00:00	31-07-1996 00:00:00
17	10264	FOLKO	6	24-07-1996 00:00:00	21-08-1996 00:00:00	23-08-1996 00:00:00
18	10265	BLONP	2	25-07-1996 00:00:00	22-08-1996 00:00:00	12-08-1996 00:00:00
19	10266	WARTH	3	26-07-1996 00:00:00	06-09-1996 00:00:00	31-07-1996 00:00:00
20	10267	FRANK	4	29-07-1996 00:00:00	26-08-1996 00:00:00	06-08-1996 00:00:00
21	10268	GROSR	8	30-07-1996 00:00:00	27-08-1996 00:00:00	02-08-1996 00:00:00
22	10269	WHITC	5	31-07-1996 00:00:00	14-08-1996 00:00:00	09-08-1996 00:00:00
23	10270	WARTH	1	01-08-1996 00:00:00	29-08-1996 00:00:00	02-08-1996 00:00:00
24	10271	SPUR	6	01-08-1996 00:00:00	29-08-1996 00:00:00	30-08-1996 00:00:00
25	10272	RATTC	6	02-08-1996 00:00:00	30-08-1996 00:00:00	06-08-1996 00:00:00
26	10273	QUICK	3	05-08-1996 00:00:00	02-09-1996 00:00:00	12-08-1996 00:00:00
27	10274	VINET	6	06-08-1996 00:00:00	03-09-1996 00:00:00	16-08-1996 00:00:00

18 COLUMNS, 199+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 13:28

File Home Transform Add Column View Tools Help

Close & Apply Close New Recent Enter Data Data Sources Manage Parameters Refresh Preview Advanced Editor Properties Choose Columns Remove Columns Keep Rows Remove Rows Reduce Rows Sort Data Type: Whole Number Use First Row as Headers Merge Queries Append Queries Text Analytics Vision Azure Machine Learning Combine Files AI Insights

Queries [2]

Table_1

Orders

Source {{Name="Orders",Signature="table"}},Data

Properties Name Orders All Properties

Applied Steps Source Navigation

ENG IN 13:30 29-12-2022

Name: Dharmit Shah.

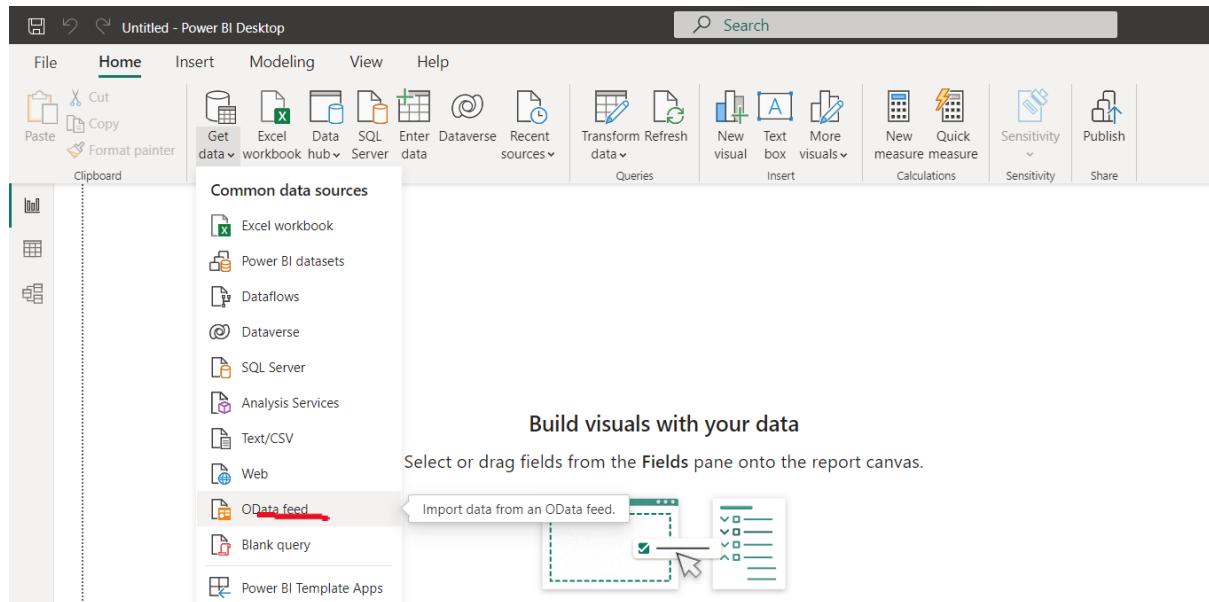
Roll no: 31011020018.

Practical 2

Aim:Perform the Extraction and Transformation and Loading (ETL) process to construct the database in the Power BI.

Step 1:ETL Process in Power BI

Open Power BI ->Click on Get Data -> OData Feed



Paste URL as

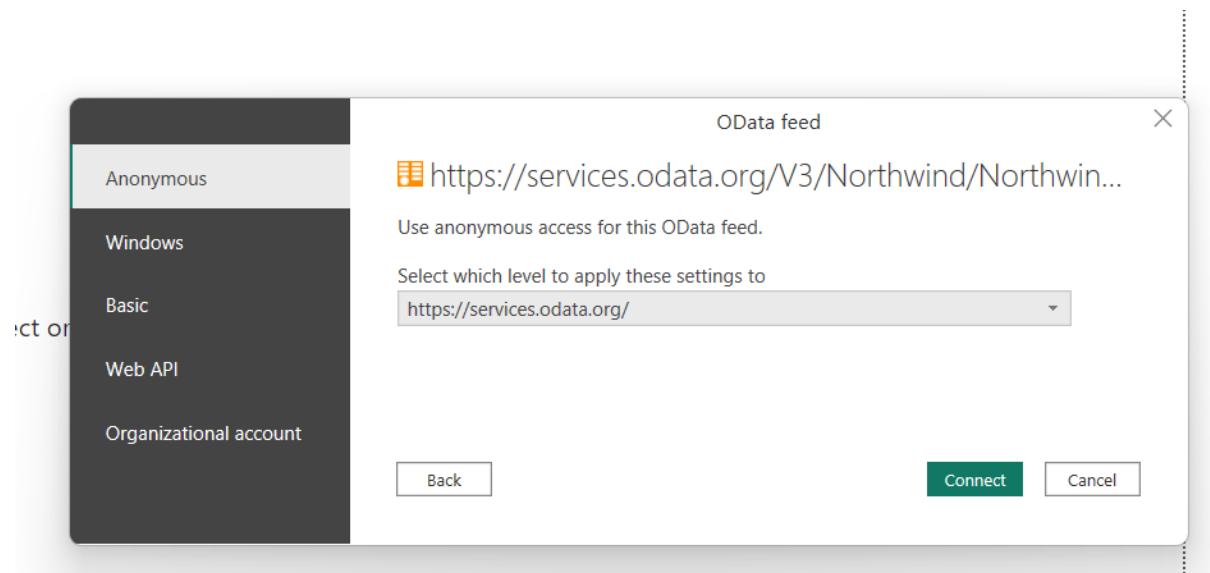
(<https://services.odata.org/v3/Northwind/Northwind.svc/>

) ->Click OK ->connect



Name: Dharmit Shah.

Roll no: 31011020018.



Click on Check Box of Products table and then Click on Load

The screenshot shows the Power BI Navigator interface. On the left, a tree view of tables is displayed, with the 'Products' table being selected, indicated by a red underline. The main pane shows a table titled 'Products' with columns: ProductID, ProductName, SupplierID, CategoryID, and QuantityPerUnit. The table contains 23 rows of product data. At the bottom, there are buttons for 'Select Related Tables', 'Load' (highlighted in green), 'Transform Data', and 'Cancel'.

ProductID	ProductName	SupplierID	CategoryID	QuantityPerUnit
1	Chai	1	1	10
2	Chang	1	1	24
3	Aniseed Syrup	1	2	12
4	Chef Anton's Cajun Seasoning	2	2	48
5	Chef Anton's Gumbo Mix	2	2	36
6	Grandma's Boysenberry Spread	3	2	12
7	Uncle Bob's Organic Dried Pears	3	7	12
8	Northwoods Cranberry Sauce	3	2	12
9	Mishi Kobe Niku	4	6	18
10	Ikura	4	8	12
11	Queso Cabrales	5	4	11
12	Queso Manchego La Pastora	5	4	10
13	Konbu	6	8	21
14	Tofu	6	7	40
15	Genen Shouyu	6	2	24
16	Pavlova	7	3	32
17	Alice Mutton	7	6	20
18	Carnarvon Tigers	7	8	16
19	Teatime Chocolate Biscuits	8	3	10
20	Sir Rodney's Marmalade	8	3	30
21	Sir Rodney's Scones	8	3	24
22	Gustaf's Knäckebröd	9	5	24
23	Tunnbröd	9	5	12

Name: Dharmit Shah.

Roll no: 31011020018.

Remove other columns to only display columns of interest in Query Editor, Select the ProductID, ProductName ,QuantityPerUnit and UnitsInStock columns.

The screenshot shows the Microsoft Power Query Editor interface. The title bar says "Untitled - Power Query Editor". The ribbon has tabs like File, Home, Transform, Add Column, View, Tools, and Help. In the Home tab, there are icons for Close & Apply, New, Recent, Enter Data, Data source settings, Manage Parameters, Refresh Preview, Properties, Advanced Editor, and Manage. The Manage section has dropdowns for Choose Columns, Remove Columns, Keep Rows, Remove Rows, and Remove Other Columns. Below the ribbon is a "Queries [1]" pane containing a single item named "Products". The main area shows a table with columns: ProductID, ProductName, SupplierID, and CategoryID. The "Remove Columns" option is highlighted in the context menu. The table data lists 28 products from 1 to 27, each with its name, supplier ID, and category ID.

ProductID	ProductName	SupplierID	CategoryID
1	Chai	1	
2	Chang	1	
3	Aniseed Syrup	1	
4	Chef Anton's Cajun Seasoning	2	
5	Chef Anton's Gumbo Mix	2	
6	Grandma's Boysenberry Spread	3	
7	Uncle Bob's Organic Dried Pears	3	
8	Northwoods Cranberry Sauce	3	
9	Mishi Kobe Niku	4	
10	Ikura	4	
11	Queso Cabrales	5	
12	Queso Manchego La Pastora	5	
13	Konbu	6	
14	Tofu	6	
15	Genen Shouyu	6	
16	Pavlova	7	
17	Alice Mutton	7	
18	Carnarvon Tigers	7	
19	Teatime Chocolate Biscuits	8	
20	Sir Rodney's Marmalade	8	
21	Sir Rodney's Scones	8	
22	Gustaf's Knäckebröd	9	
23	Tunnbröd	9	
24	Guaraná Fantástica	10	
25	NuNuCa Nuß-Nougat-Creme	11	
26	Gumbär Gummibärchen	11	
27	Schoggi Schokolade	11	

Change the data type of the UnitsInStock column

Select the UnitInStock column

Name: Dharmit Shah.

Roll no: 31011020018.

The screenshot shows the Power Query Editor interface with the 'Products' table loaded. A context menu is open over the 'UnitsInStock' column header, showing various data type conversion options. The 'Whole Number' option is selected.

Expand the Orders table

Once you have loaded a data source ,you can click on Recent Sources to select desirable table (Orders)

The screenshot shows the Power BI Desktop interface with the 'Recent sources' pane open. It lists recent data sources and pending changes. The message 'Build visuals with your data' is displayed at the bottom.

After selecting the URL,Navigator window will appear from which you can select Orders table ->Click on Load

Name: Dharmit Shah.

Roll no: 31011020018.

The screenshot shows the Power BI Navigator window. On the left, there is a tree view of available tables from a Northwind OData source. The 'Orders' table is selected, indicated by a checked checkbox icon next to it. The main pane displays the 'Orders' table data with columns: OrderID, CustomerID, EmployeeID, OrderDate, and RequiredDate. The data consists of 42 rows of order information. At the bottom of the window, there are buttons for 'Select Related Tables', 'Load', 'Transform Data', and 'Cancel'.

OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate
10248	VINET	5	04-07-1996 00:00:00	01-08-1996
10249	TOMSP	6	05-07-1996 00:00:00	16-08-1996
10250	HANAR	4	08-07-1996 00:00:00	05-08-1996
10251	VICTE	3	08-07-1996 00:00:00	05-08-1996
10252	SUPRD	4	09-07-1996 00:00:00	06-08-1996
10253	HANAR	3	10-07-1996 00:00:00	24-07-1996
10254	CHOPS	5	11-07-1996 00:00:00	08-08-1996
10255	RICSU	9	12-07-1996 00:00:00	09-08-1996
10256	WELLI	3	15-07-1996 00:00:00	12-08-1996
10257	HILAA	4	16-07-1996 00:00:00	13-08-1996
10258	ERNSH	1	17-07-1996 00:00:00	14-08-1996
10259	CENTC	4	18-07-1996 00:00:00	15-08-1996
10260	OTTIK	4	19-07-1996 00:00:00	16-08-1996
10261	QUEDDE	4	19-07-1996 00:00:00	16-08-1996
10262	RATTC	8	22-07-1996 00:00:00	19-08-1996
10263	ERNSH	9	23-07-1996 00:00:00	20-08-1996
10264	FOLKO	6	24-07-1996 00:00:00	21-08-1996
10265	BLONP	2	25-07-1996 00:00:00	22-08-1996
10266	WARTH	3	26-07-1996 00:00:00	06-09-1996
10267	FRANK	4	29-07-1996 00:00:00	26-08-1996
10268	GROSR	8	30-07-1996 00:00:00	27-08-1996
10269	WHITC	5	31-07-1996 00:00:00	14-08-1996
10270	WARTH	1	01-08-1996 00:00:00	29-08-1996

Query Editor Window will appear

In the Query View ,scroll to the Order_Details column

In the Order_Details column,select the expand icon .

In the Expand drop-down:

Select(Select All Columns) to clear all columns

Select ProductID,UnitPrice and Quantity.

Click OK.

Name: Dharmit Shah.

Roll no: 31011020018.

The screenshot shows the Power Query Editor interface with the 'Order_Details' query selected. A context menu is open over the 'Order_Details' table, with the 'Expand' option selected. The 'Expand' dialog box is displayed, showing the 'Aggregate' radio button is selected. Under 'Search Columns to Expand', the 'Product' column is checked. Other columns like OrderID, ProductID, UnitPrice, Quantity, and Discount are also checked. The 'OK' button is highlighted. The main pane shows the expanded 'Order_Details' table with additional columns for OrderID, ProductID, UnitPrice, Quantity, and Discount. The status bar at the bottom right indicates 'PREVIEW DOWNLOADED AT 13:28' and the date '29-12-2022'.

This screenshot shows the Power Query Editor after clicking 'OK' in the previous dialog. The 'Order_Details' query has been successfully expanded, adding new columns: OrderID, ProductID, UnitPrice, Quantity, and Discount. The expanded table now has 23 columns and 28 rows. The status bar at the bottom right shows '14:10' and the date '29-12-2022'.

After Clicking on OK following screen appears with combined columns.

Name: Dharmit Shah.

Roll no: 31011020018.

A screenshot of the Power Query Editor interface. At the top, there is a formula bar with the code: `= Table.ExpandTableColumn(Orders_table, "Order_Details", {"ProductID", "UnitPrice", "Quantity"}, {"Order_Details.ProductID", "1.2 Order_Details.UnitPrice", "1.2 Order_Details.Quantity", "Shipper"})`. Below the formula bar is a table with five columns: Employee, Order_Details.ProductID, Order_Details.UnitPrice, Order_Details.Quantity, and Shipper. The table contains 25 rows, each labeled 'Record' in the Employee column. The Order_Details.UnitPrice column contains values like 11, 42, 72, etc. The Order_Details.Quantity column contains values like 14, 9.8, 34.8, etc. The Shipper column contains values like 12, 10, 5, etc.

	Employee	Order_Details.ProductID	1.2 Order_Details.UnitPrice	1.2 Order_Details.Quantity	Shipper
1	Record		11	14	12
2	Record		42	9.8	10
3	Record		72	34.8	5
4	Record		14	18.6	9
5	Record		51	42.4	40
6	Record		41	7.7	10
7	Record		51	42.4	35
8	Record		65	16.8	15
9	Record		22	16.8	6
10	Record		57	15.6	15
11	Record		65	16.8	20
12	Record		20	64.8	40
13	Record		33	2	25
14	Record		60	27.2	40
15	Record		31	10	20
16	Record		39	14.4	42
17	Record		49	16	40
18	Record		24	3.6	15
19	Record		55	19.2	21
20	Record		74	8	21
21	Record		2	15.2	20
22	Record		16	13.9	35
23	Record		36	15.2	25
24	Record		59	44	30
25	Record		53	26.2	15
26	Record		77	10.4	12
27	Record		27	35.1	25

Calculate the lines total for each Order_Details row

In the Add Column ribbon tab, Click Add Custom Column

In the Custom Column dialog box ,in the Custom Column Formula textbox, enter [Order_Details.UnitPrice]*[Order_Details.Quantity] by selecting from available columns and click on insert for each column

In the New column name textbox, enter LineTotal

Click OK.

A screenshot of the Power Query Editor interface. The ribbon is visible at the top with the 'Add Column' tab selected. The formula bar shows the formula: `= Table.ExpandTableColumn(Orders_table, "Order_Details", {"ProductID", "UnitPrice", "Quantity"}, {"Order_Details.ProductID", "1.2 Order_Details.UnitPrice", "1.2 Order_Details.Quantity", "Shipper"})`. In the main area, there are two queries: 'Products' and 'Orders'. The 'Orders' query is currently selected. A preview of the data is shown with columns: Customer, Employee, Order_Details.ProductID, 1.2 Order_Details.UnitPrice, and 1.2 Order_Details.Quantity. The data shows three records with values corresponding to the table above.

Customer	Employee	Order_Details.ProductID	1.2 Order_Details.UnitPrice	1.2 Order_Details.Quantity
1 Record	Record	11	14	
2 Record	Record	42	9.8	
3 Record	Record	72	34.8	

Name: Dharmit Shah.

Roll no: 31011020018.

Custom Column

Add a column that is computed from the other columns.

New column name

Custom

Custom column formula ⓘ

=

Available columns

ShipCountry
Customer
Employee
Order_Details.ProductID
Order_Details.UnitPrice
Order_Details.Quantity
Shipper

<< Insert

[Learn about Power Query formulas](#)

OK

Cancel

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

ShipCountry
Customer
Employee
Order_Details.ProductID
Order_Details.UnitPrice
Order_Details.Quantity
Shipper

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

		2		15.2		20	Re
		16		13.9		35	Re

Name: Dharmit Shah.

Roll no: 31011020018.

Custom Column

Add a column that is computed from the other columns.

New column name

LineTotal

Custom column formula ⓘ

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

Available columns

ShipCountry
Customer
Employee
Order_Details.ProductID
Order_Details.UnitPrice
Order_Details.Quantity
Shipper

<< Insert

[Learn about Power Query formulas](#)

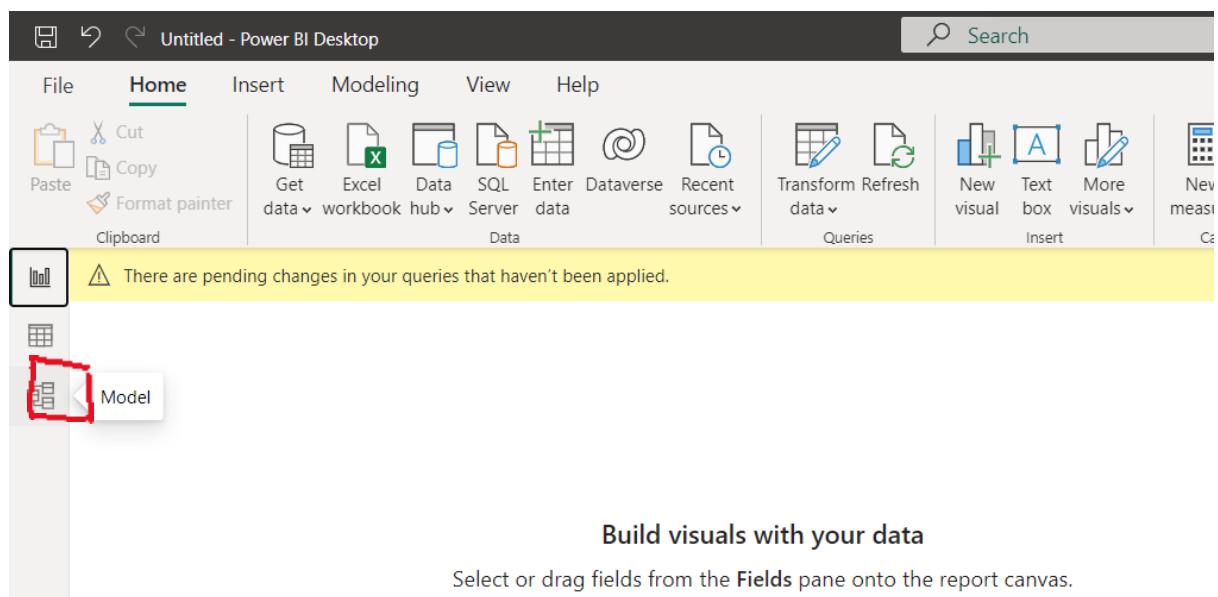
✓ No syntax errors have been detected.

OK

Cancel

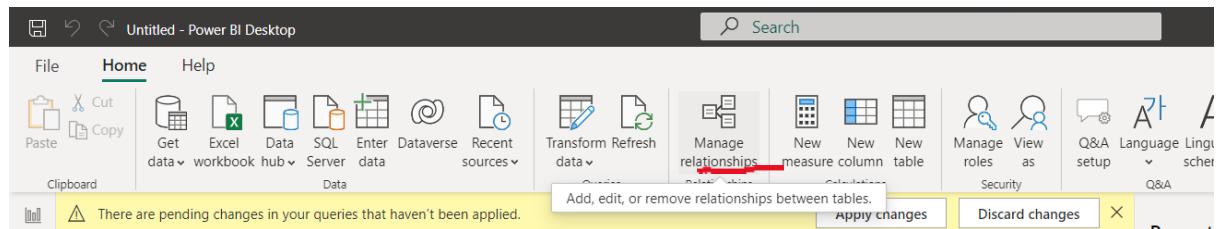
Once the data is loaded ,select the Manage Relationship button Home ribbon from Untitled-Power BI Desktop window

Click on Model-> Manage relationships

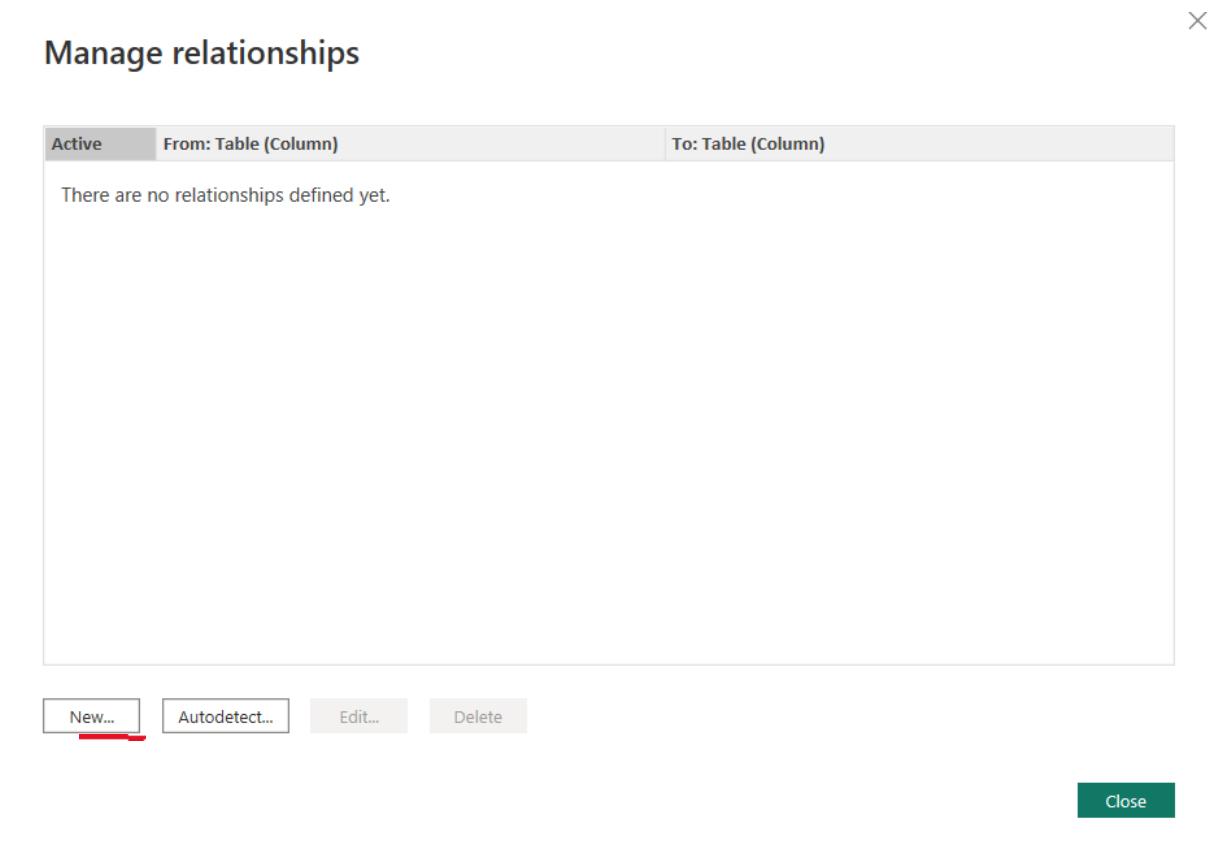


Name: Dharmit Shah.

Roll no: 31011020018.



Click on new->Create relationship ->click on ok



Name: Dharmit Shah.

Roll no: 31011020018.

X

Create relationship

Select tables and columns that are related.

Products

ProductID	ProductName	SupplierID	CategoryID	QuantityPerUnit	UnitPrice	UnitsInStock	UnitsOnOrder	ReorderLevel	Discontinued
1	Chai	1	1	10 boxes x 20 bags	18	39		10	
2	Chang	1	1	24 - 12 oz bottles	19	17		15	
3	Aniseed Syrup	1	2	12 - 550 ml bottles	10	13		15	

Orders

OrderID	CustomerID	EmployeeID	OrderDate	RequiredDate	ShippedDate	ShipVia	Freight
10249	TOMSP	6	05-07-1996 00:00:00	16-08-1996 00:00:00	10-07-1996 00:00:00	1	10.00
10260	OTTIK	4	19-07-1996 00:00:00	16-08-1996 00:00:00	29-07-1996 00:00:00	1	10.00
10267	FRANK	4	29-07-1996 00:00:00	26-08-1996 00:00:00	06-08-1996 00:00:00	1	10.00

Cardinality

Cross filter direction

One to many (1:*)

▼ Single ▼

Make this relationship active

Apply security filter in both directions

Assume referential integrity

OK

Cancel

OUTPUT

Name: Dharmit Shah.

Roll no: 31011020018.

The screenshot shows the Power BI Desktop interface with the 'Home' tab selected. A yellow status bar at the top indicates pending changes. Two tables are displayed: 'Products' and 'Orders'. A relationship is established between them, with 'Products' having a primary key constraint (PK) and 'Orders' having a foreign key constraint (FK). The 'Products' table contains columns: CategoryID, Discontinued, ProductID, ProductName, QuantityPerUnit, ReorderLevel, SupplierID, UnitPrice, and UnitsInStock. The 'Orders' table contains columns: CustomerID, EmployeeID, Freight, OrderDate, OrderID, RequiredDate, ShipAddress, ShipCity, and ShipCountry.

Products

- CategoryID
- Discontinued
- ProductID
- ProductName
- QuantityPerUnit
- ReorderLevel
- SupplierID
- UnitPrice
- UnitsInStock

Orders

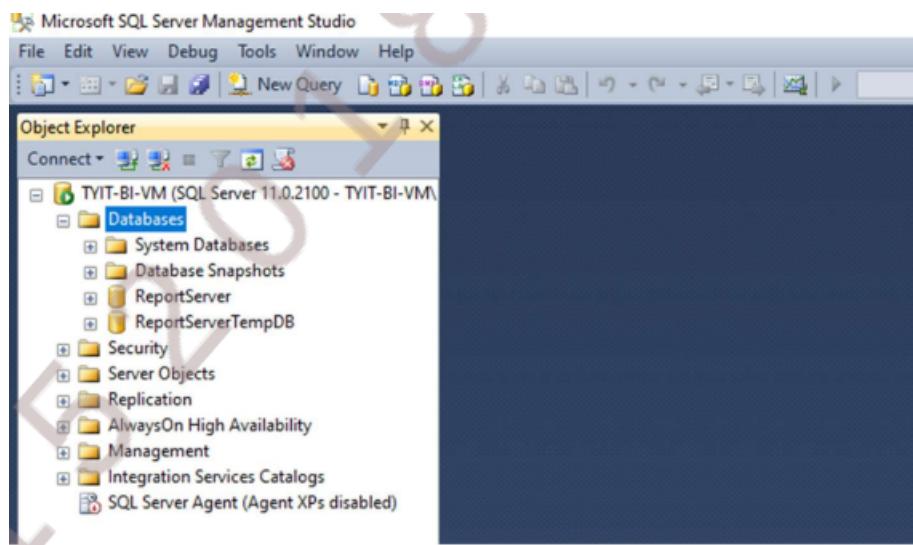
- CustomerID
- EmployeeID
- Freight
- OrderDate
- OrderID
- RequiredDate
- ShipAddress
- ShipCity
- ShipCountry

Name: Dharmit Shah.

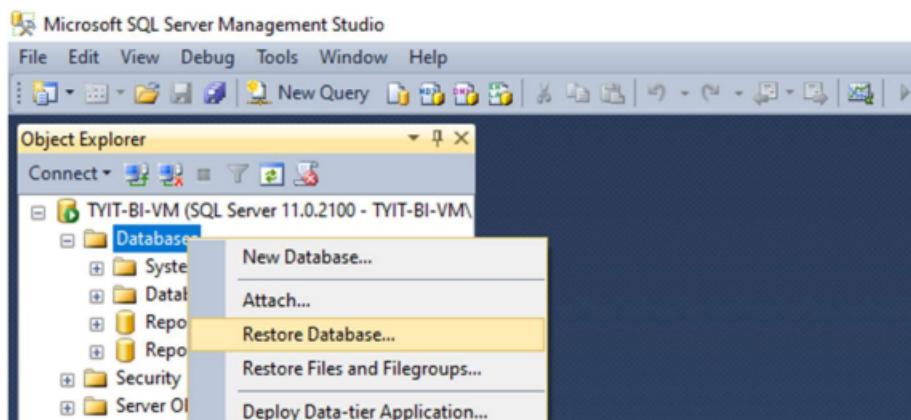
Roll no: 31011020018.

2B) ETL PROCESS IN SQL SERVER

1. Open SQL Server Management Studio to restore backup file



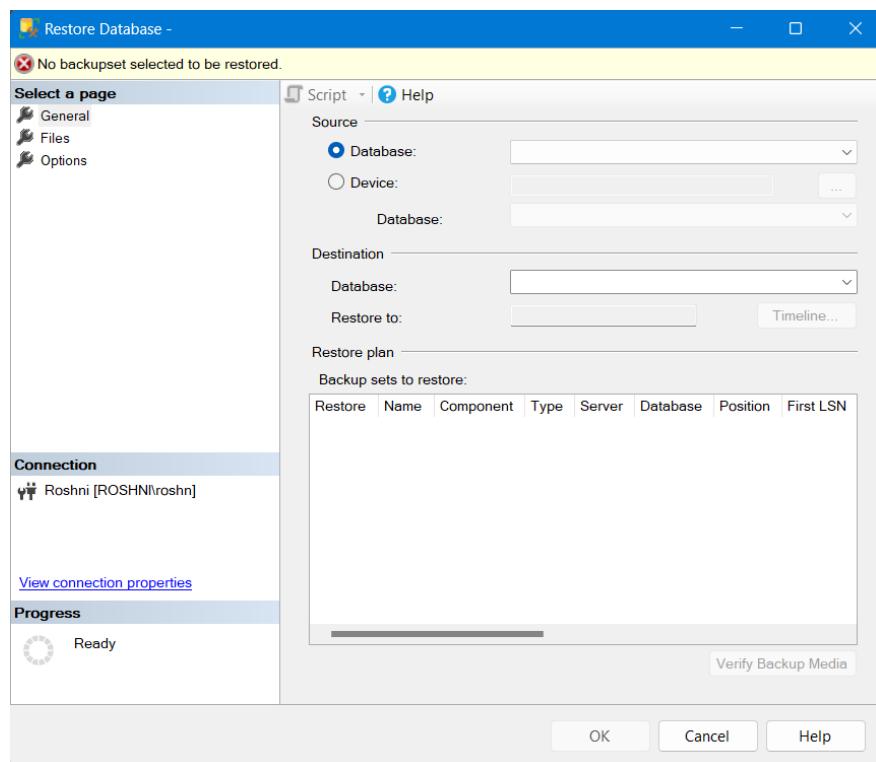
2. Right click on Databases → Restore Database



3. Select Device → click on ... icon towards end of device box

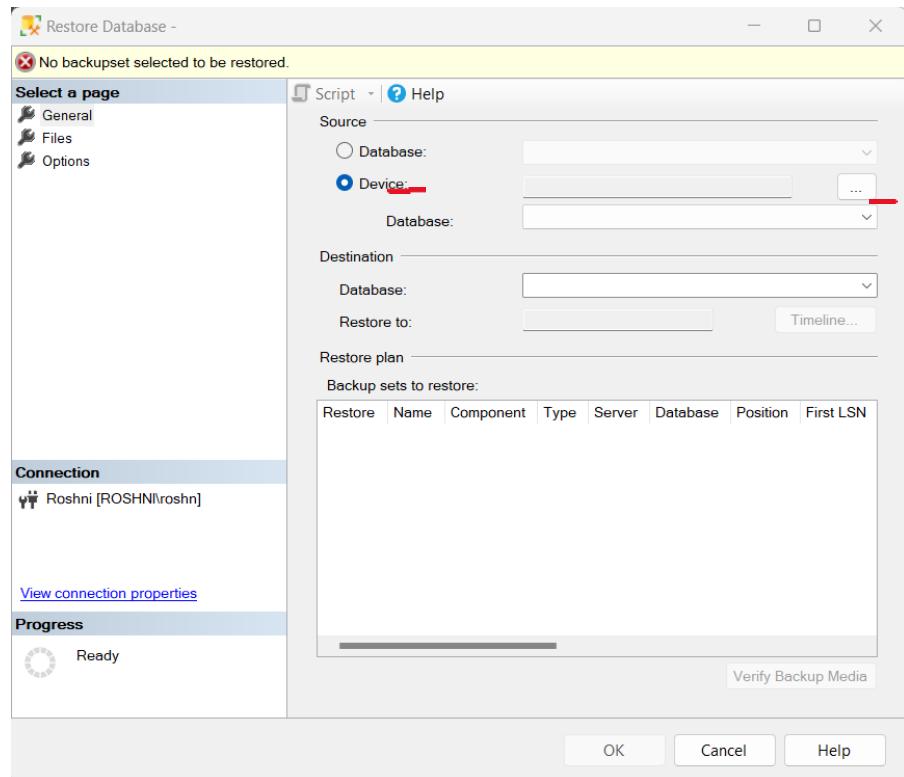
Name: Dharmit Shah.

Roll no: 31011020018.

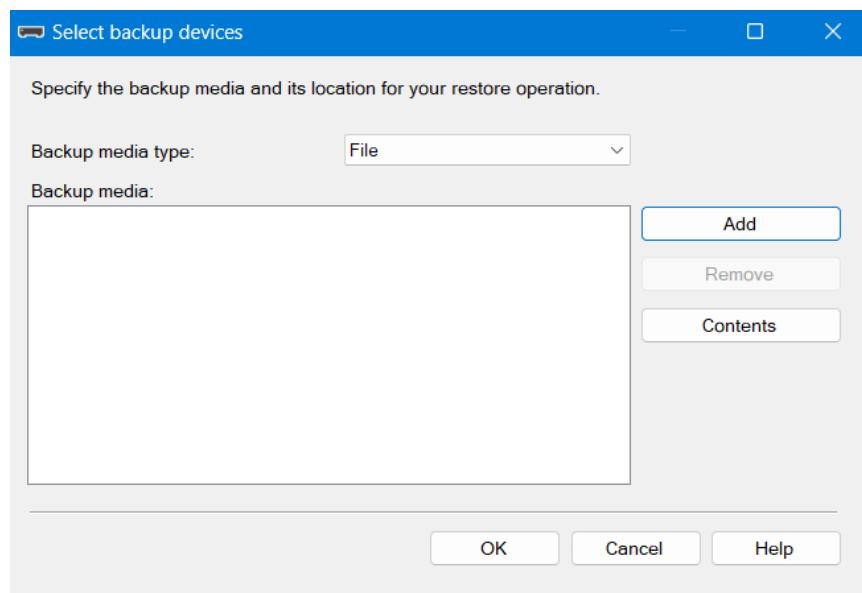


Name: Dharmit Shah.

Roll no: 31011020018.

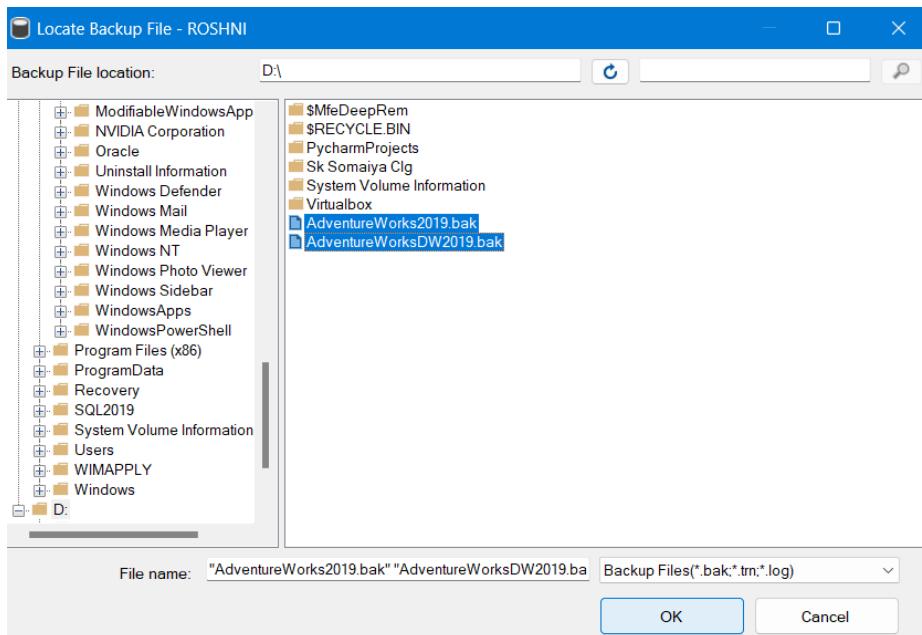


4. Click on Add -> Select on folder where you have downloaded your backup files -> Click Ok

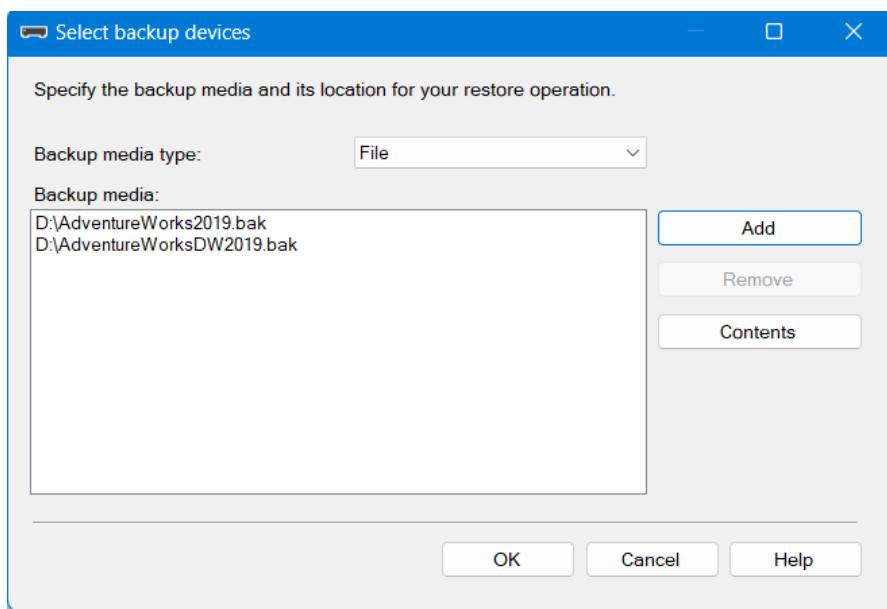


Name: Dharmit Shah.

Roll no: 31011020018.

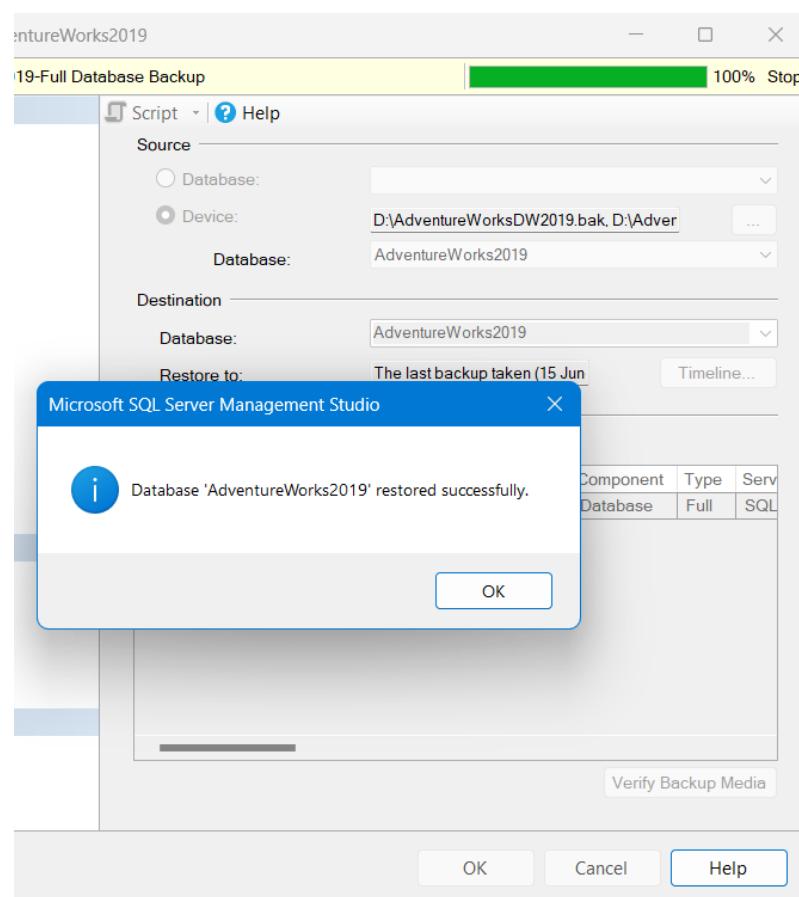
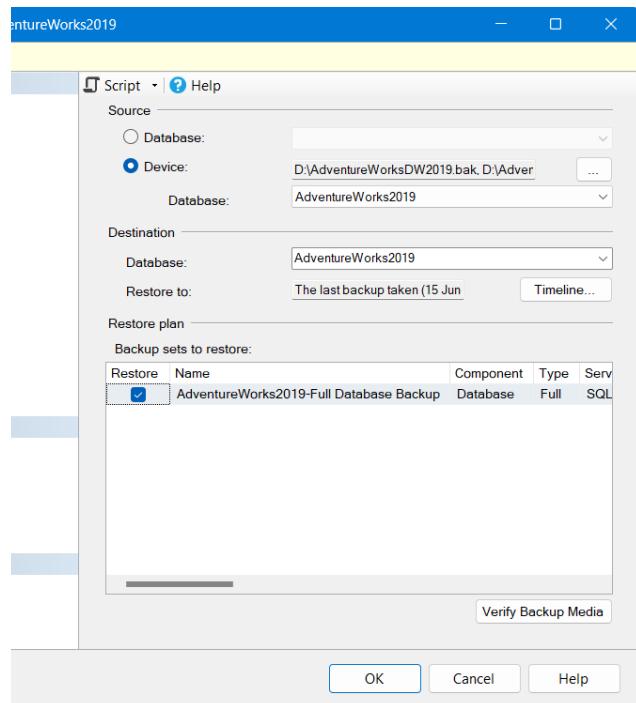


□ Click Ok



Name: Dharmit Shah.

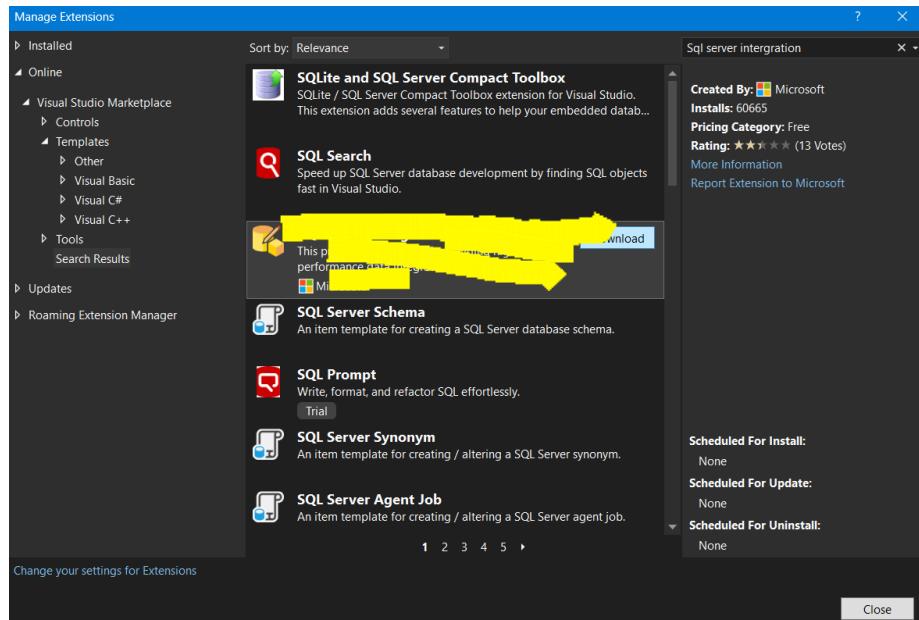
Roll no: 31011020018.



Name: Dharmit Shah.

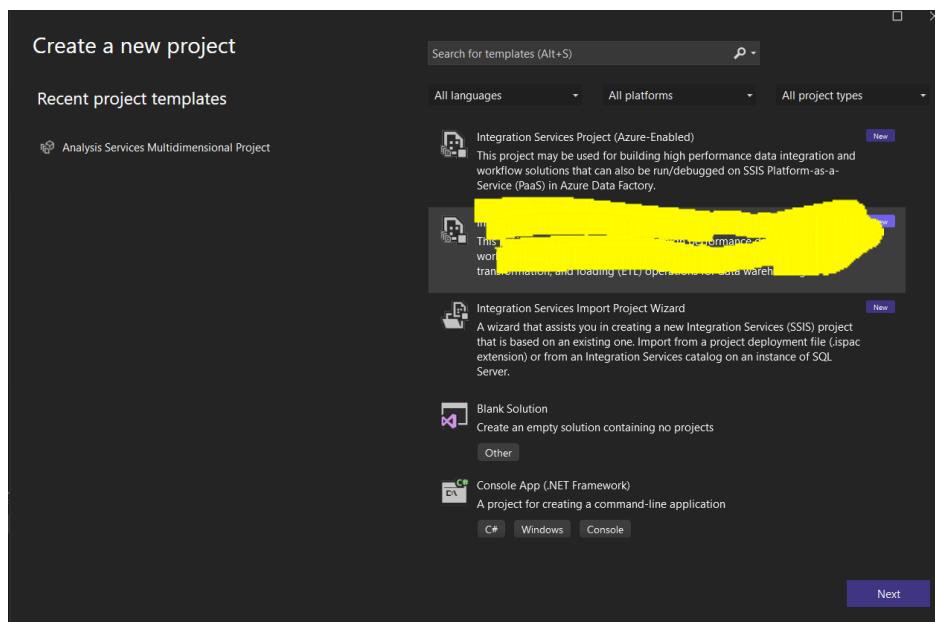
Roll no: 31011020018.

4. Open Visual Studio 2022-> Download & install extension (SQL Server Integration Service Project)



5. Open Visual Studio -> Create a new project -> Integration Services

Project & give appropriate project name.

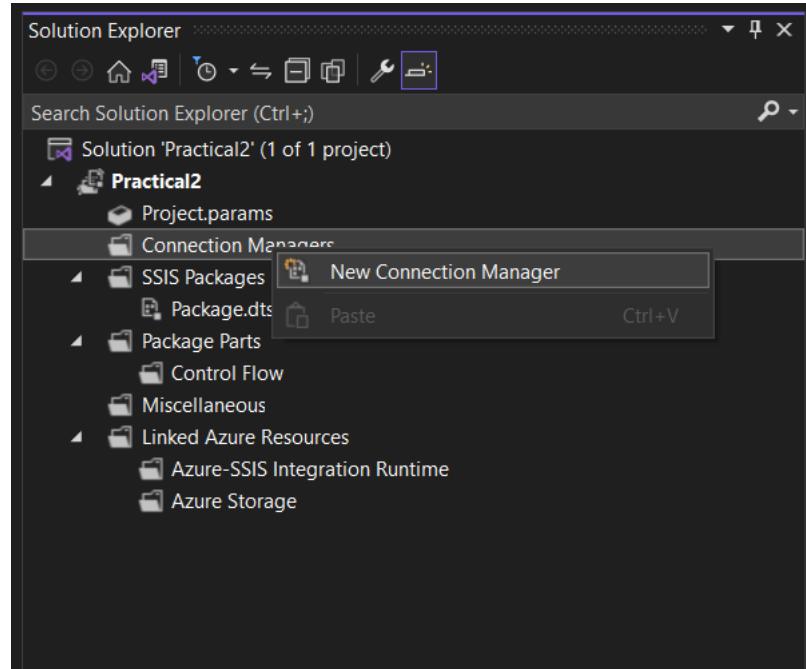


6. Right click on Connection Managers in solution explorer and click on New

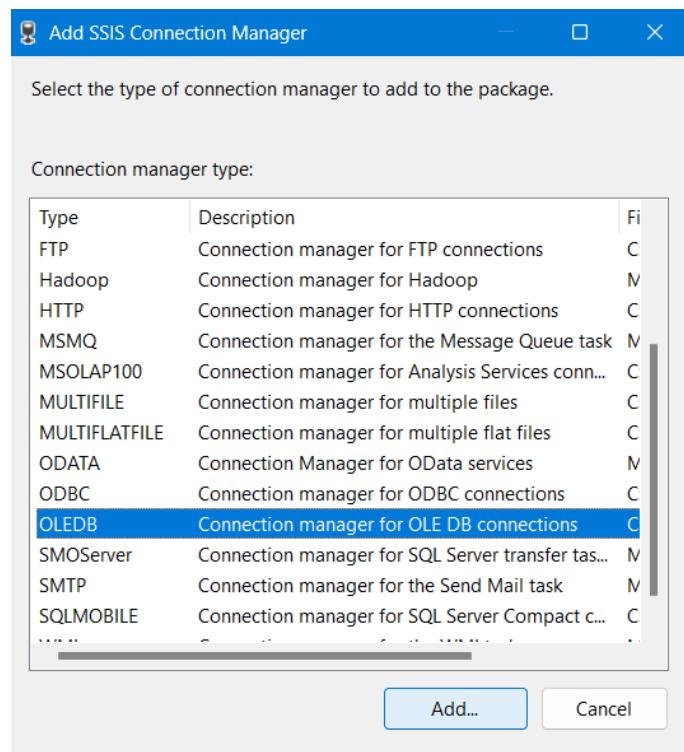
Connection Manager.

Name: Dharmit Shah.

Roll no: 31011020018.



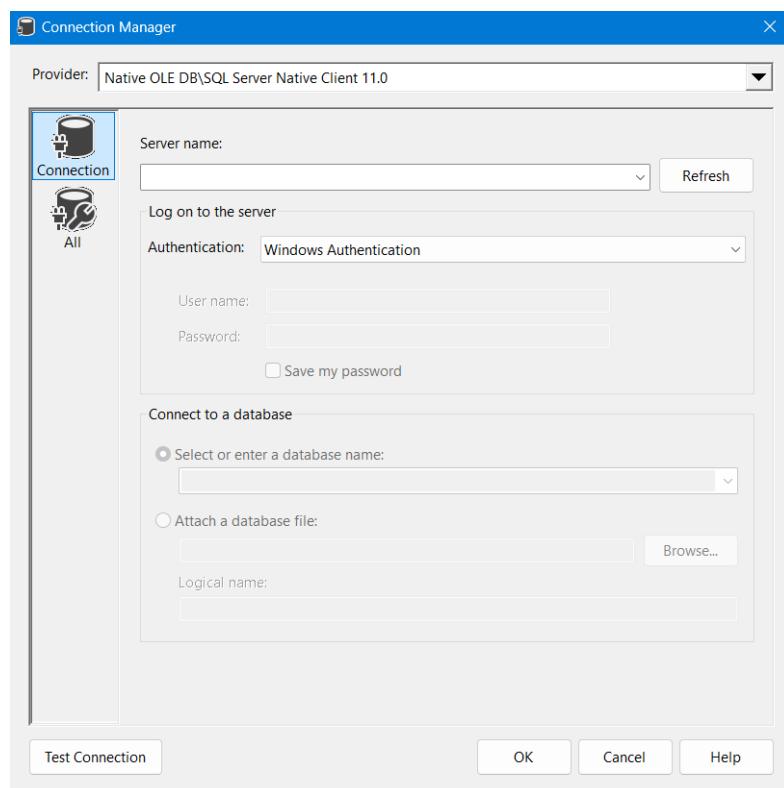
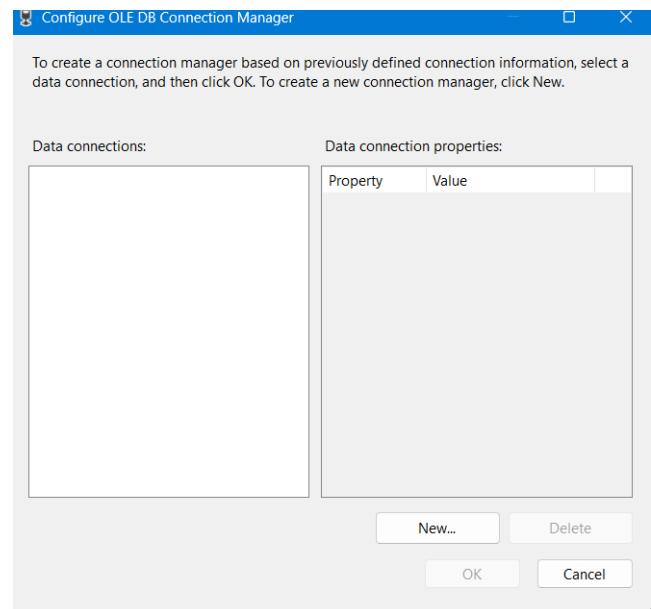
7. Configure OLEDB Connection Manager window appears → Click on Add



□ Click on New

Name: Dharmit Shah.

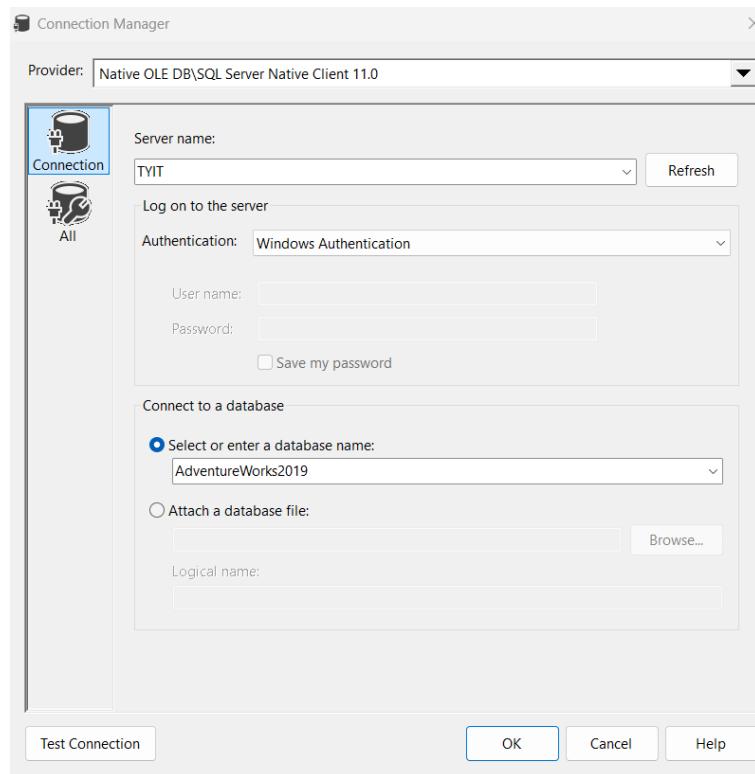
Roll no: 31011020018.



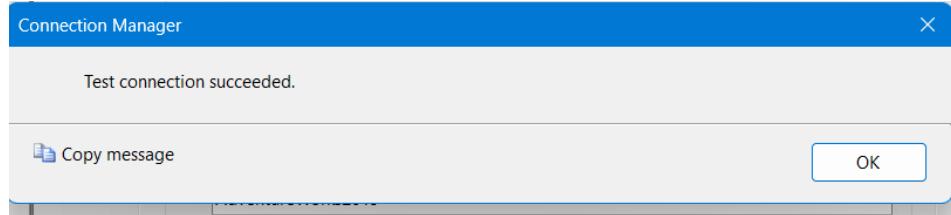
8. Select Server name(Your PC Name) from drop down and database name and click on Test connection.

Name: Dharmit Shah.

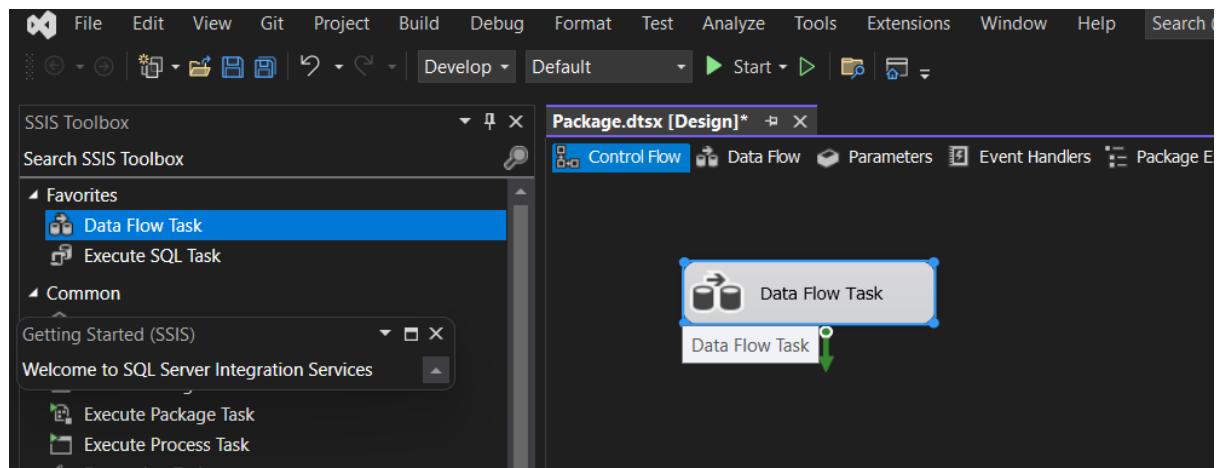
Roll no: 31011020018.



□ Click ok



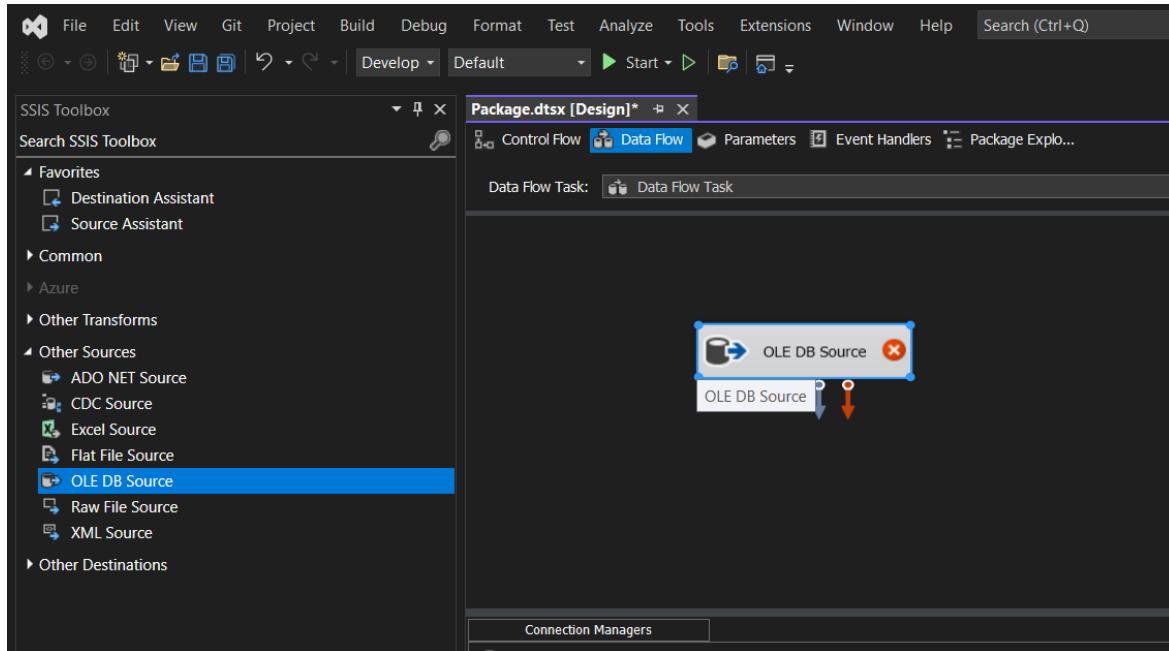
9. Drag and drop Data Flow Task in Control Flow tab



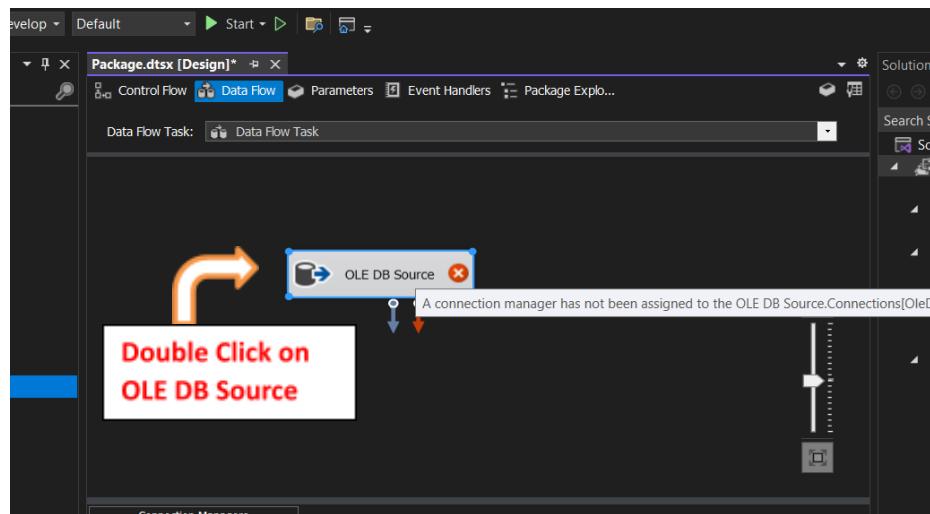
Name: Dharmit Shah.

Roll no: 31011020018.

10. Drag OLE DB Source from Other Sources and drop into Data Flow tab

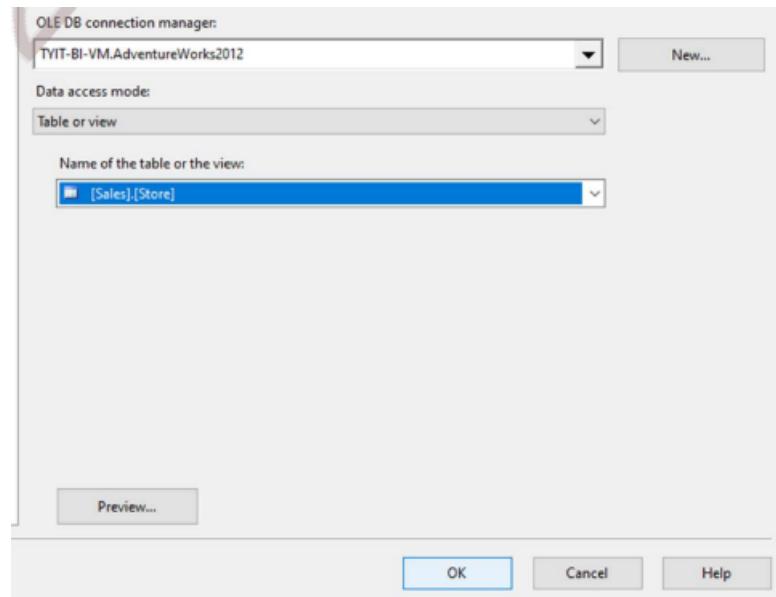


11. Double click on OLE DB source → OLE DB Source Editor appears → click on New to add connection manager. Select [Sales].[Store] table from drop down → ok

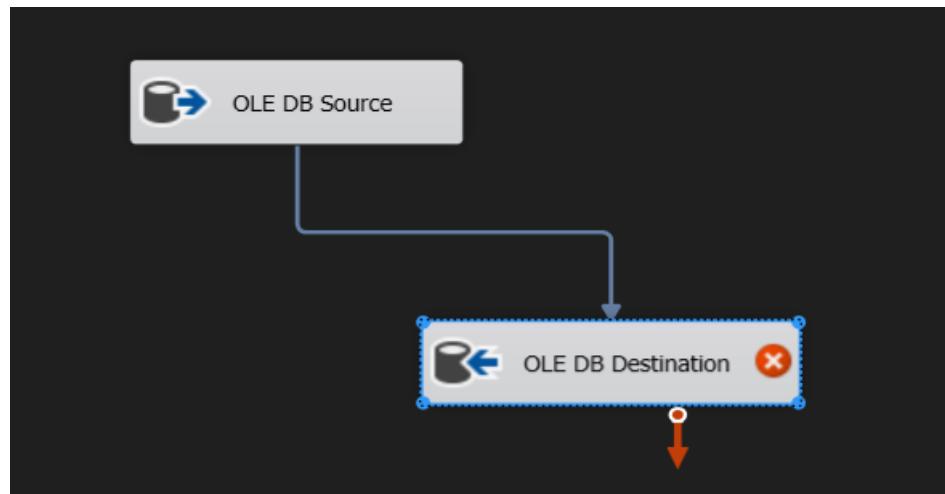
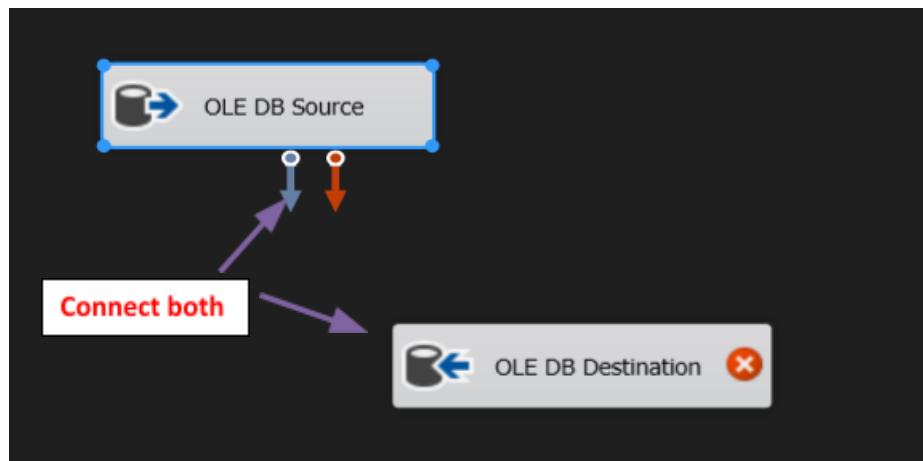


Name: Dharmit Shah.

Roll no: 31011020018.



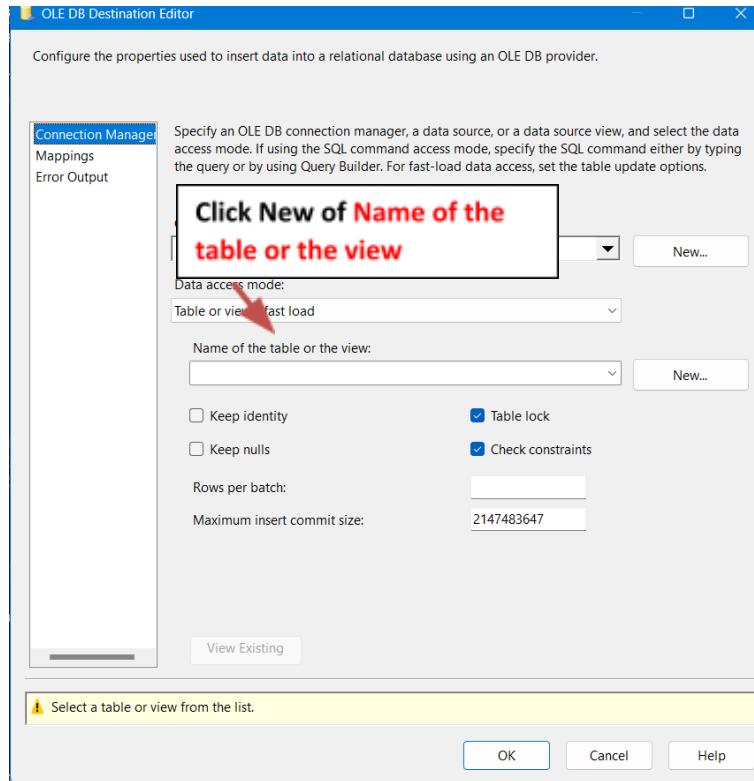
12. Drag OLE DB destination from Other Destination in data flow tab and connect both



Name: Dharmit Shah.

Roll no: 31011020018.

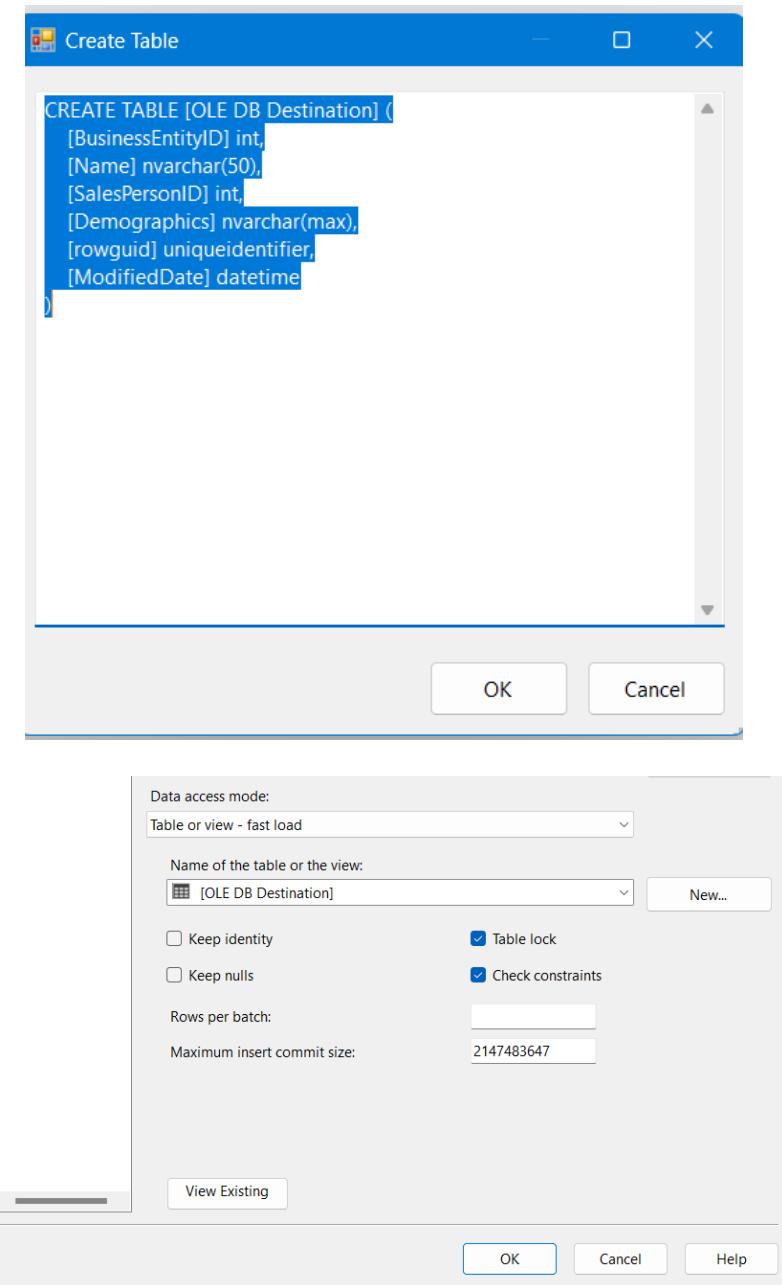
13. Double click on OLE DB destination Click on New to run the query to get [OLEDB Destination] in Name of the table or the view.



Click Ok

Name: Dharmit Shah.

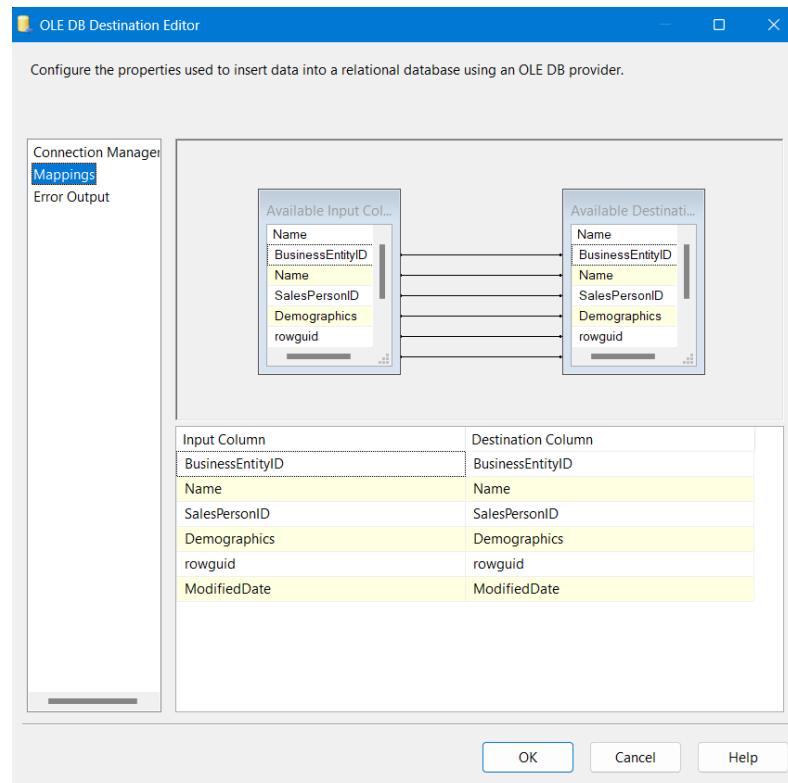
Roll no: 31011020018.



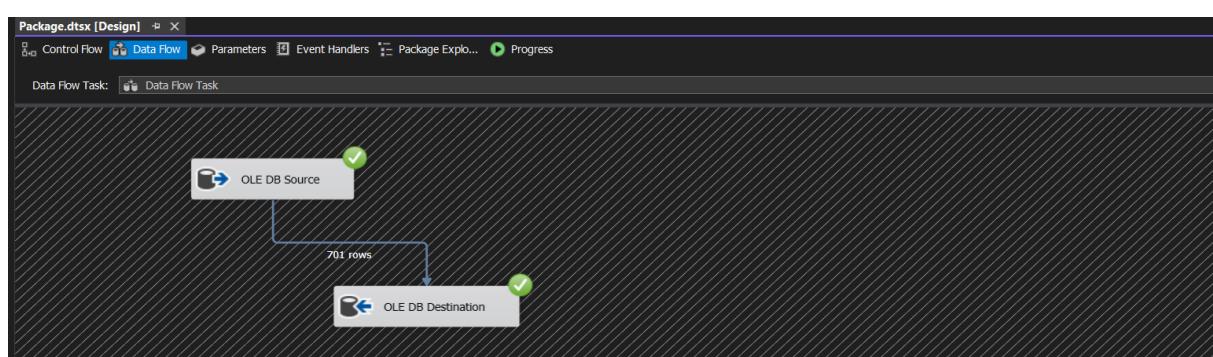
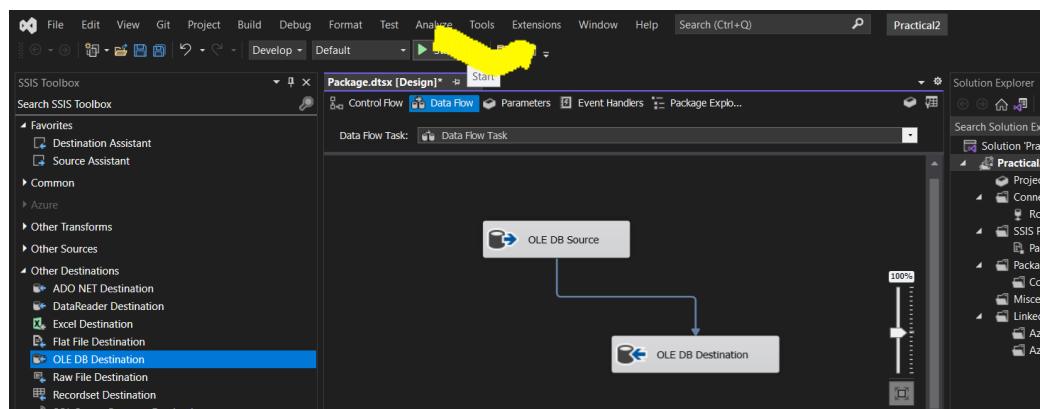
□ Click on Mappings

Name: Dharmit Shah.

Roll no: 31011020018.



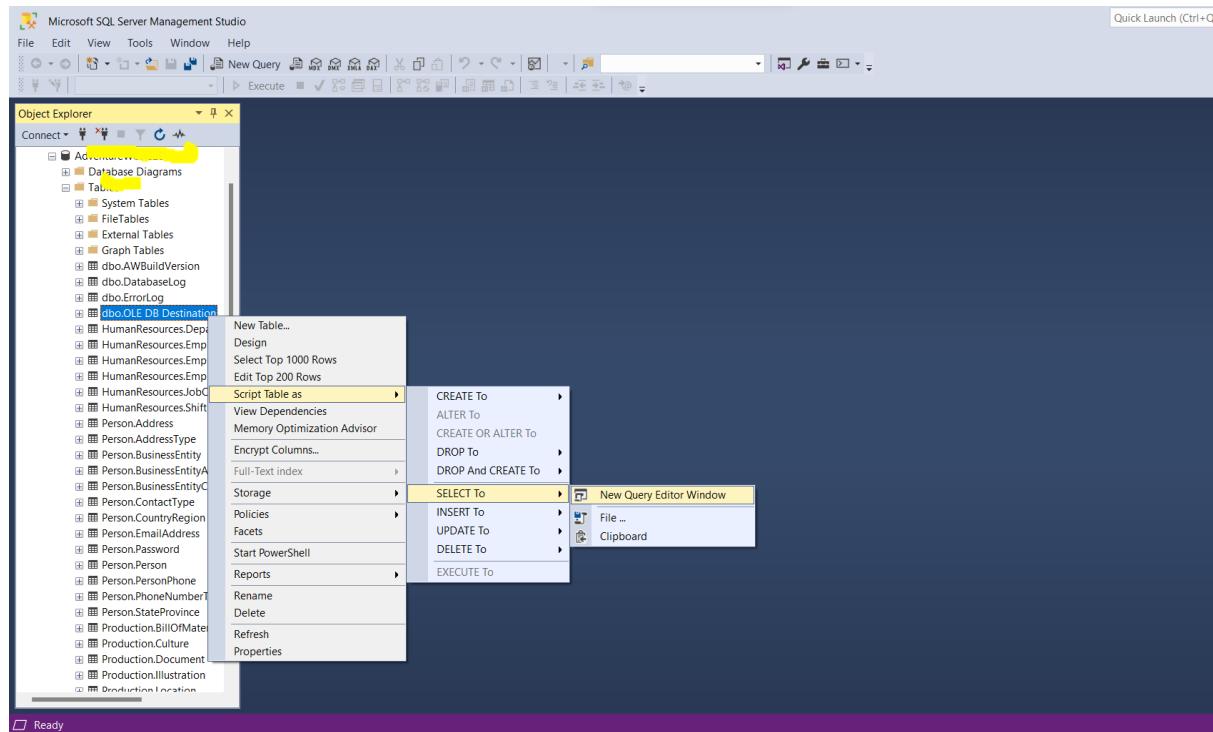
14. Click on start



Name: Dharmit Shah.

Roll no: 31011020018.

**15. Go to SQL Server Management Studio In database tab → AdventureWorks2019
→Tables → Right click on [dbo].[OLE DB Destination] → Scrip Table as → SELECT To
→ New Query Editor Window**



16. Execute following query to get output.

```
SQLQuery1.sql - TY...-BI-VM\admin (51)) X
USE [AdventureWorks2012]
GO

SELECT [BusinessEntityID]
,[Name]
,[SalesPersonID]
,[Demographics]
,[rowguid]
,[ModifiedDate]
FROM [dbo].[ODE DB Destination]
GO
```

Name: Dharmit Shah.

Roll no: 31011020018.

The screenshot shows a SQL Server Management Studio window with the 'Results' tab selected. A green status bar at the bottom indicates 'Query executed successfully.' The results grid displays 17 rows of data from a table with columns: BusinessEntityID, Name, SalesPersonID, Demographics, and rowguid. The data includes various bike store names like 'Next-Door Bike Store', 'Professional Sales and Service', and 'Valley Toy Shop', along with their respective IDs and demographic XML snippets.

	BusinessEntityID	Name	SalesPersonID	Demographics	rowguid
1	292	Next-Door Bike Store	279	<StoreSurvey xmlns="http://schemas.microsoft.co...	A22517E3-848D-4EBE-B9D9-7437F34323I
2	294	Professional Sales and Service	276	<StoreSurvey xmlns="http://schemas.microsoft.co...	B50CA50B-C601-4A13-B07E-2C63862D71I
3	296	Riders Company	277	<StoreSurvey xmlns="http://schemas.microsoft.co...	337C3688-1339-4E1A-A08A-B54B23566E4
4	298	The Bike Mechanics	275	<StoreSurvey xmlns="http://schemas.microsoft.co...	7894F278-F0C8-4D16-BD75-213FDBF130:
5	300	Nationwide Supply	286	<StoreSurvey xmlns="http://schemas.microsoft.co...	C3FC9705-A8C4-4F3A-9550-E82FA4B7B64
6	302	Area Bike Accessories	281	<StoreSurvey xmlns="http://schemas.microsoft.co...	368BE6DD-30E5-49BB-9A86-71FD49C58F
7	304	Bicycle Accessories and Kits	283	<StoreSurvey xmlns="http://schemas.microsoft.co...	35F40636-5105-49D5-869E-27E23118915
8	306	Clamps & Brackets Co.	275	<StoreSurvey xmlns="http://schemas.microsoft.co...	64D06BFC-D060-405C-8C60-C067FE7067
9	308	Valley Bicycle Specialists	277	<StoreSurvey xmlns="http://schemas.microsoft.co...	59386B0C-652E-4668-B44B-4E171179333
10	310	New Bikes Company	279	<StoreSurvey xmlns="http://schemas.microsoft.co...	47E4B6BD-5CD1-45A3-A231-79D930381C
11	312	Vinyl and Plastic Goods Corporation	282	<StoreSurvey xmlns="http://schemas.microsoft.co...	DC610525-E373-49B1-B786-EA040EC250
12	314	Top of the Line Bikes	288	<StoreSurvey xmlns="http://schemas.microsoft.co...	E290E93F-A980-4BA3-86C3-9858F15C8A6
13	316	Fun Toys and Bikes	281	<StoreSurvey xmlns="http://schemas.microsoft.co...	6CDCF941-4192-49C7-994A-5ADBA534E0!
14	318	Great Bikes	283	<StoreSurvey xmlns="http://schemas.microsoft.co...	956FBC35-5E0D-4175-8045-E0BE380BA3-
15	320	Metropolitan Sales and Rental	275	<StoreSurvey xmlns="http://schemas.microsoft.co...	0CB4FF2-5047-40F7-8848-B59F7A3F3EE
16	322	Irregulars Outlet	288	<StoreSurvey xmlns="http://schemas.microsoft.co...	CDE66279-83D8-4340-A83C-E86E15514A
17	324	Valley Toy Shop	287	<StoreSurvey xmlns="http://schemas.microsoft.co...	6A1BEA66-DCB7-4ECE-9C02-2706C12E92

Practical 3

Aim:- Create the cube with suitable dimension and fact tables based on OLAP

Steps: -

1. Creating Data Warehouse

a. Open SQL Server Management Studio 2019

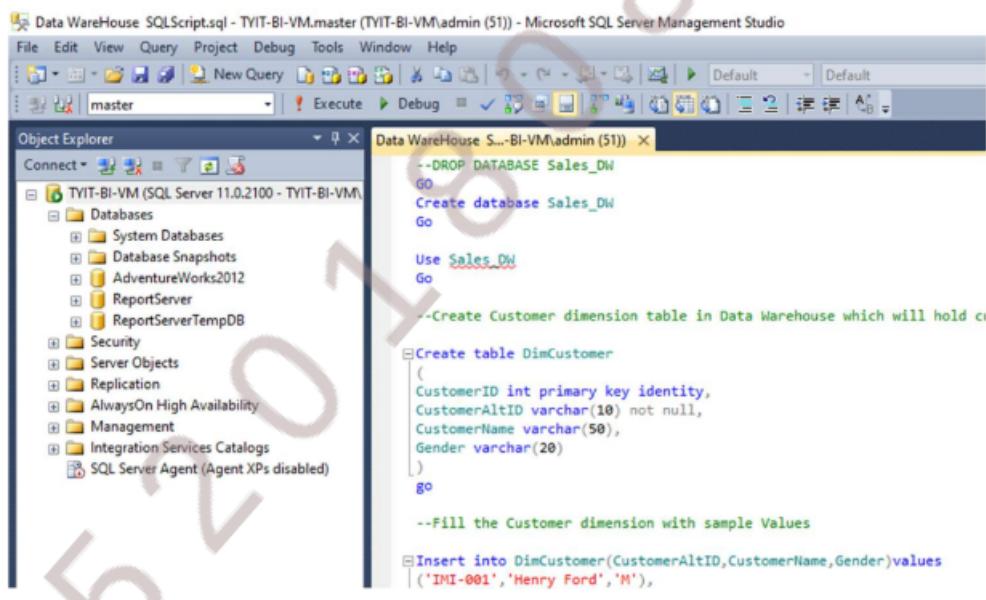
b. Connect Database Engine

Name: Dharmit Shah.

Roll no: 31011020018.

c. Open New Query editor.

d. New SQL Query Editor will be opened containing Sales_DW Database.

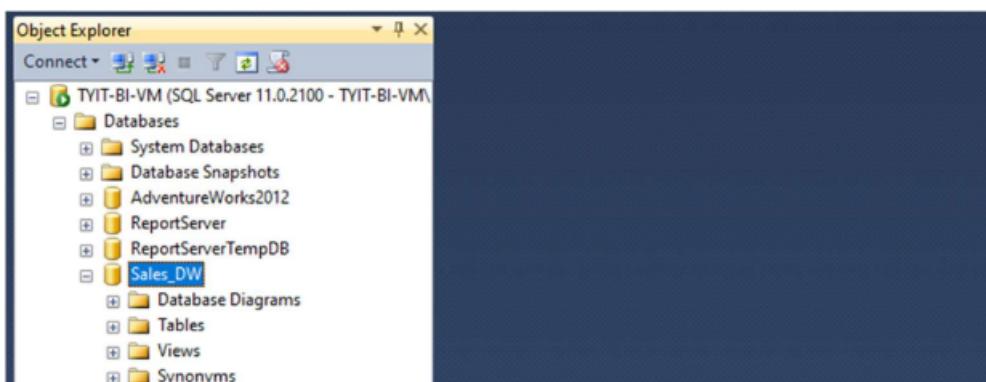


The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the connection to 'TYIT-BI-VM (SQL Server 11.0.2100 - TYIT-BI-VM\master)'. The 'Databases' node is expanded, showing 'System Databases', 'Database Snapshots', 'AdventureWorks2012', 'ReportServer', 'ReportServerTempDB', 'Security', 'Server Objects', 'Replication', 'AlwaysOn High Availability', 'Management', 'Integration Services Catalogs', and 'SQL Server Agent (Agent XPs disabled)'. The 'Query' tab in the center contains a script to create the 'Sales_DW' database and a 'DimCustomer' table, followed by sample data insertions. The status bar at the bottom indicates '1 row(s) affected'.

```
--DROP DATABASE Sales_DW  
GO  
Create database Sales_DW  
Go  
  
Use Sales_DW  
Go  
  
--Create Customer dimension table in Data Warehouse which will hold customer information  
Create table DimCustomer  
(  
CustomerID int primary key identity,  
CustomerAltID varchar(10) not null,  
CustomerName varchar(50),  
Gender varchar(20)  
)  
go  
  
--Fill the Customer dimension with sample Values  
Insert into DimCustomer(CustomerAltID,CustomerName,Gender)values  
('IMI-001','Henry Ford','M')
```

e. Click on execute or press F5 by selecting query one by one or directly click on Execute.

f. After completing execution save and close SQL Server Management studio & Reopen to see Sales_DW in Databases Tab.

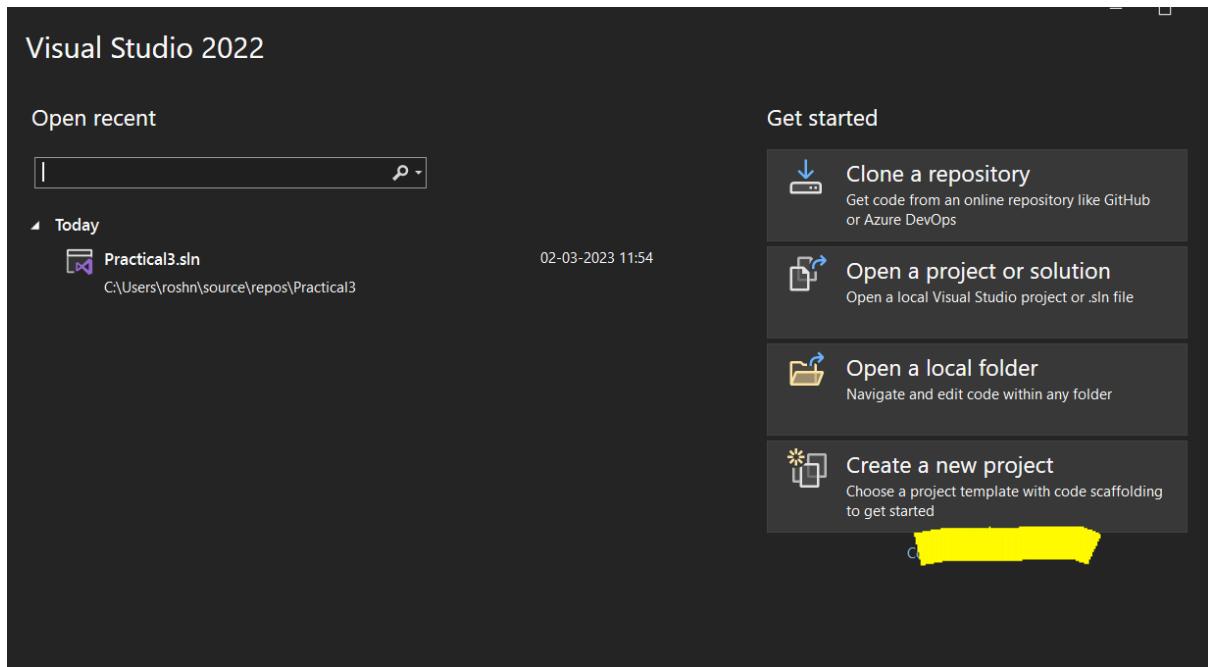


2. Adding Microsoft Analysis Services 2022 extension in visual studios

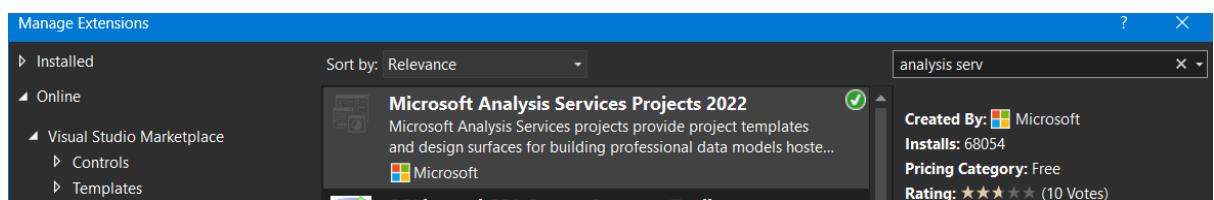
a. Open Visual studio 2022 and continue without code

Name: Dharmit Shah.

Roll no: 31011020018.



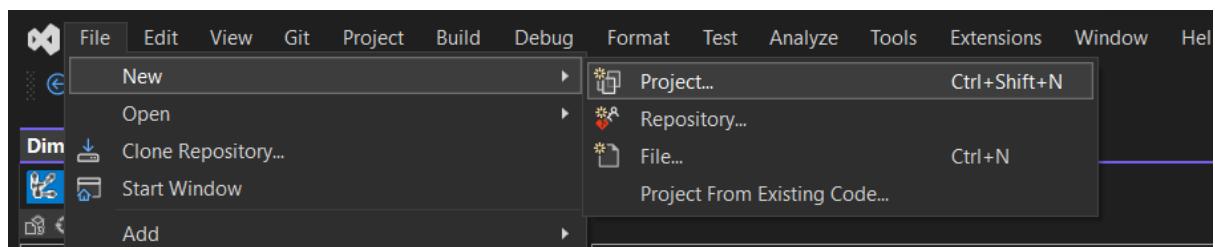
- b. Go to Extension >Manage Extensions >Online >Search for Analysis Services download and install the extension**



3. Creating NEW DATA SOURCE

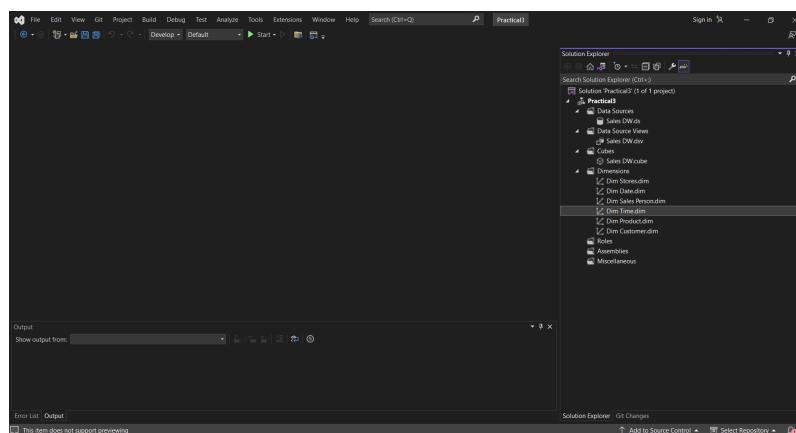
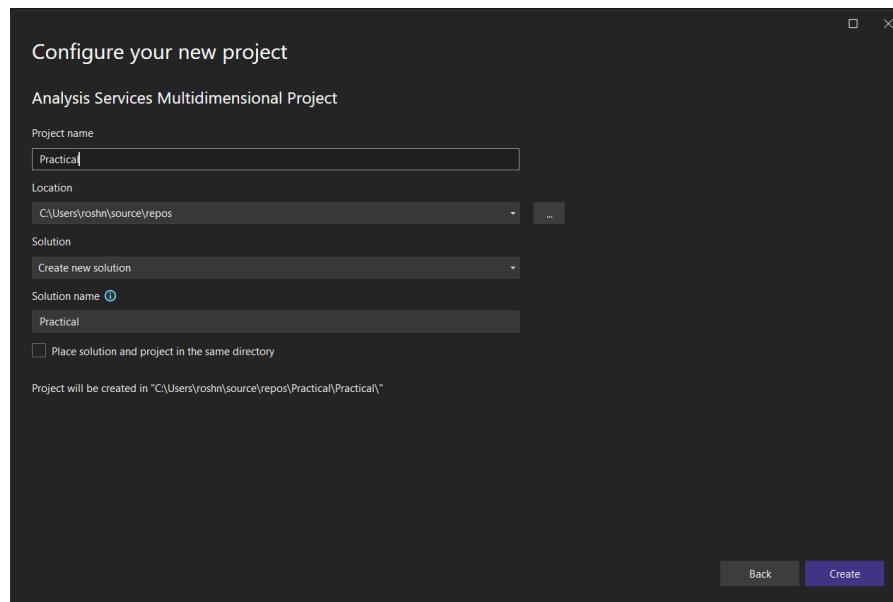
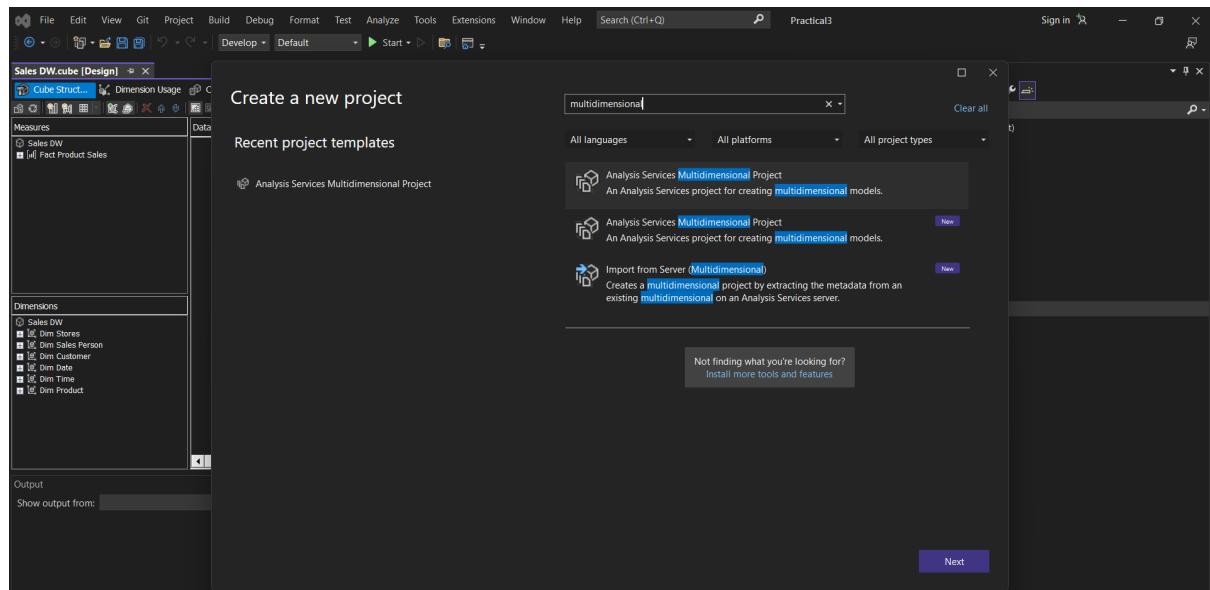
- a. Open Visual Studios.**

Click on File ->New->Project-> Analysis Services Multidimensional Project ->Next->appropriate project name ->Create.



Name: Dharmit Shah.

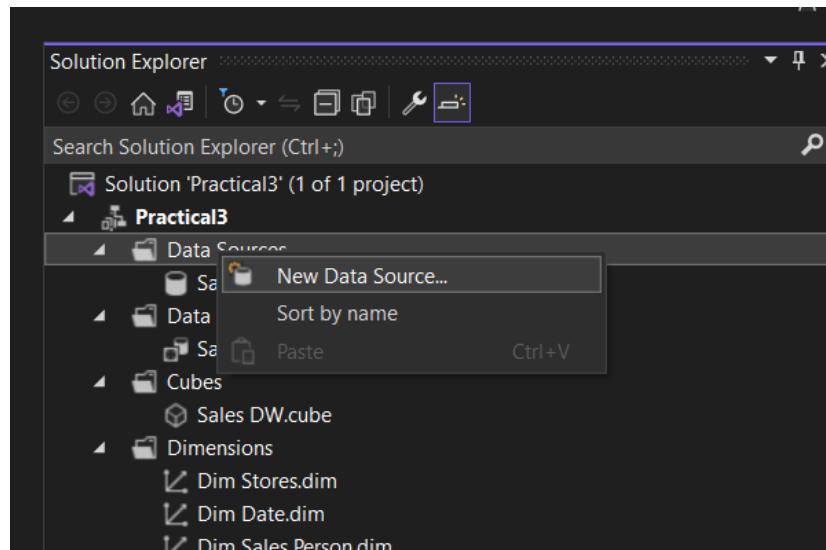
Roll no: 31011020018.



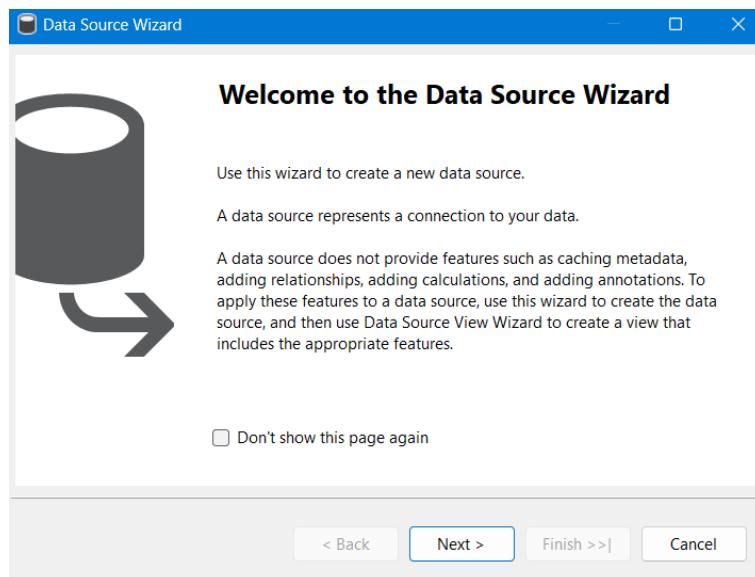
b. Right click on Data Sources in solution explorer ->New Data Source

Name: Dharmit Shah.

Roll no: 31011020018.



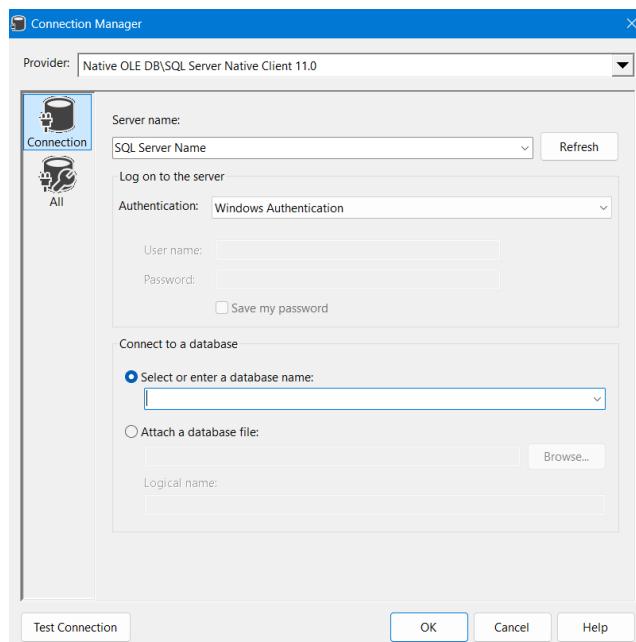
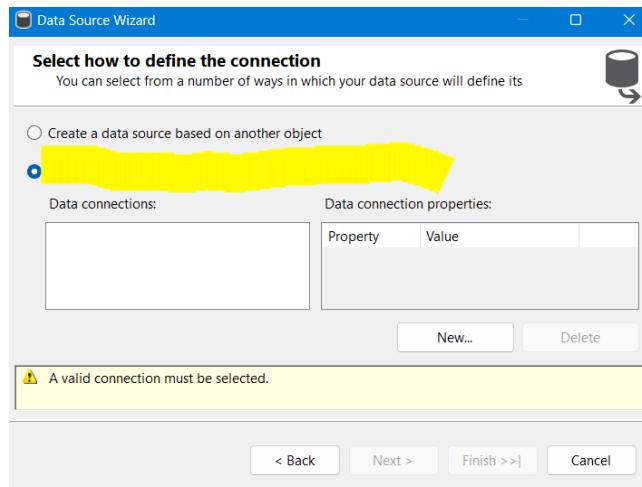
c. Data Source Wizard appears ->Next



**d. Select Server Name (Your PC Name)->select Use Windows Authentication
->Select or enter a database name (Sales_DW) -> Click Ok->Test Connection**

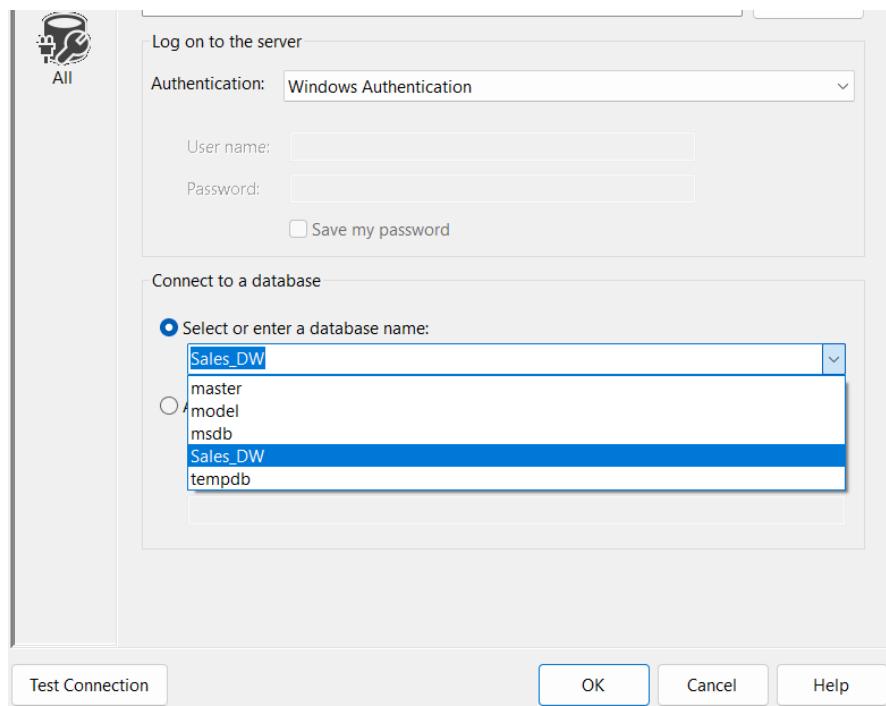
Name: Dharmit Shah.

Roll no: 31011020018.



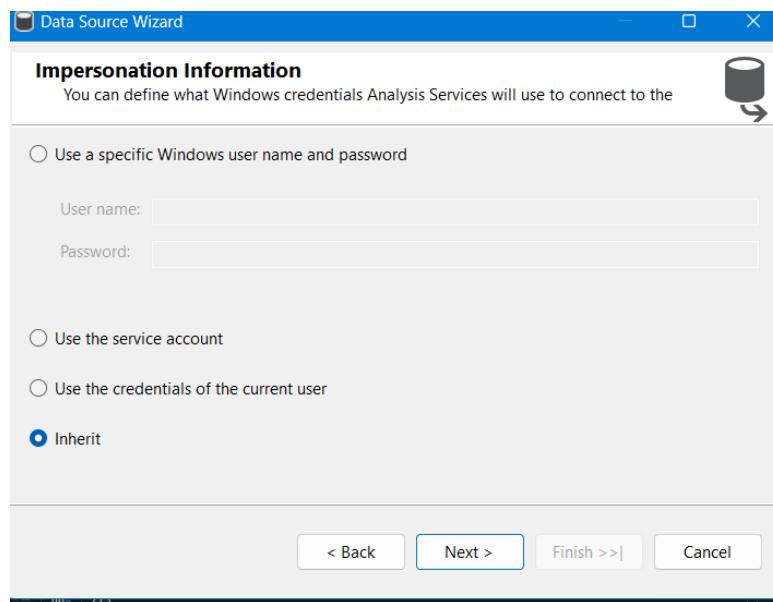
Name: Dharmit Shah.

Roll no: 31011020018.



e. Click Next

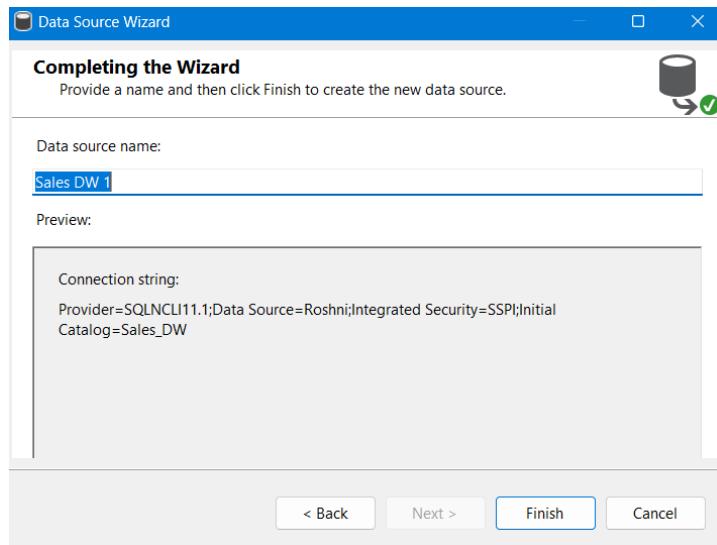
f. Select Inherit ->Next



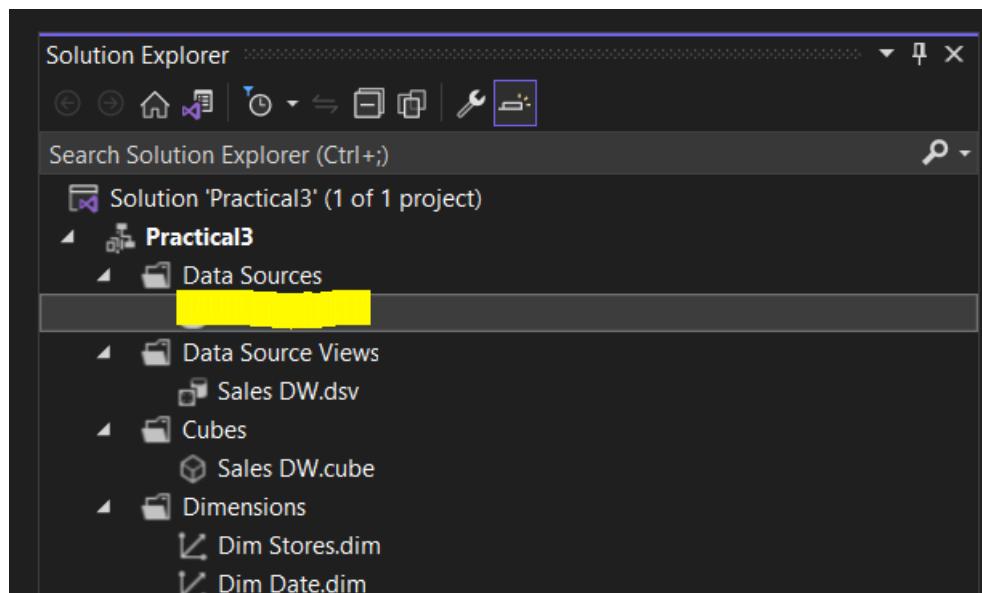
g. Click Finish

Name: Dharmit Shah.

Roll no: 31011020018.

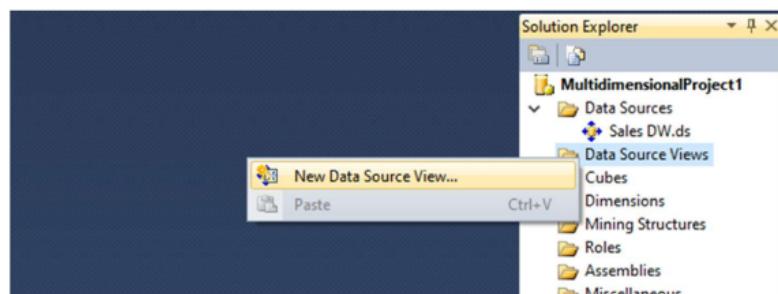


k. Sales DW.ds gets created under Data Sources in Solution Explorer



4. Creating New Data Source View

- In Solution explorer right click on Data Source View ->Select New Data Source View

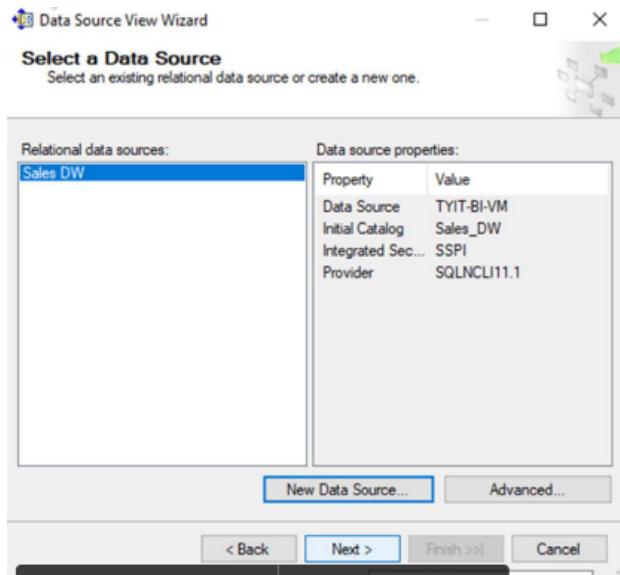


Name: Dharmit Shah.

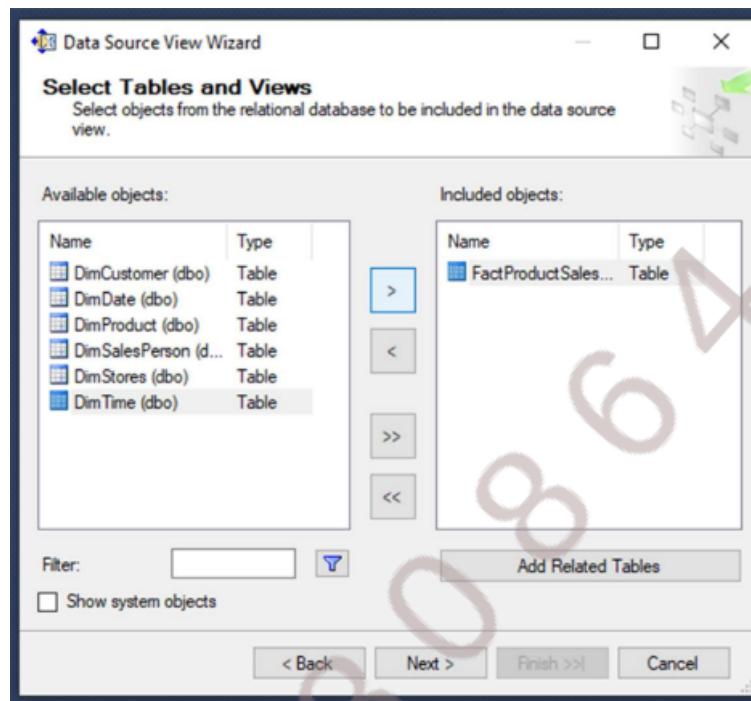
Roll no: 31011020018.

b. Click Next

c. Click Next



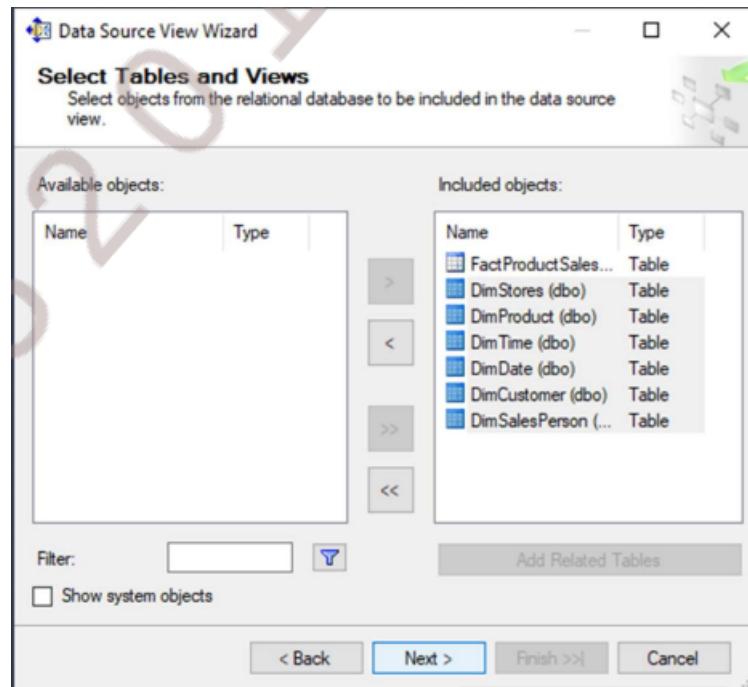
d. Select FactProductSales(dbo) from Available objects and put in Includes Objects by clicking on (>)



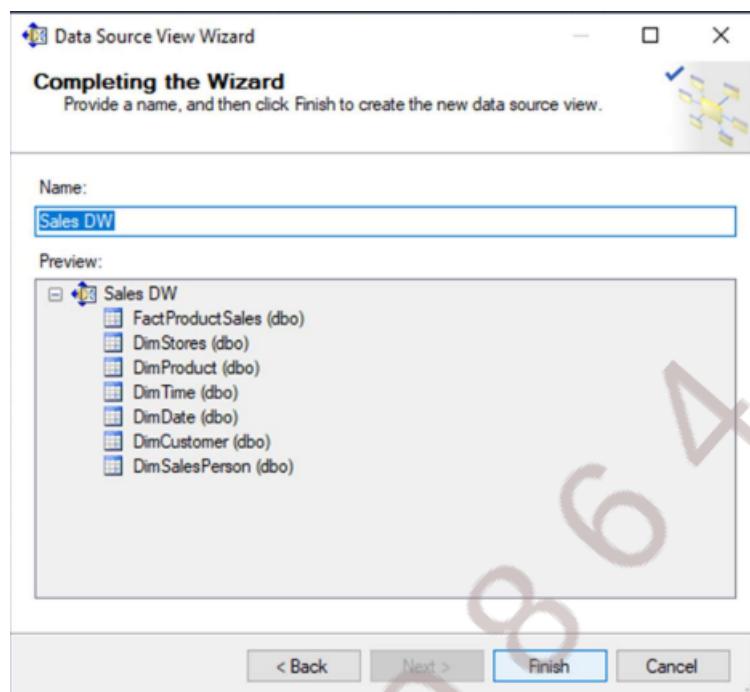
e. Click on Add Related Tables

Name: Dharmit Shah.

Roll no: 31011020018.



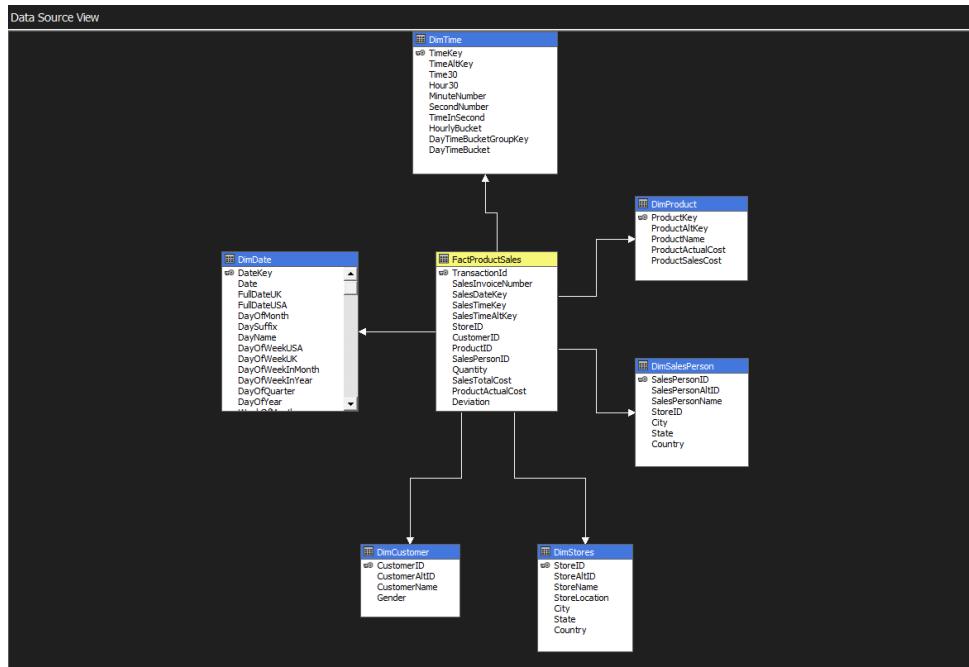
f. Click Next



g. Click Finish

Name: Dharmit Shah.

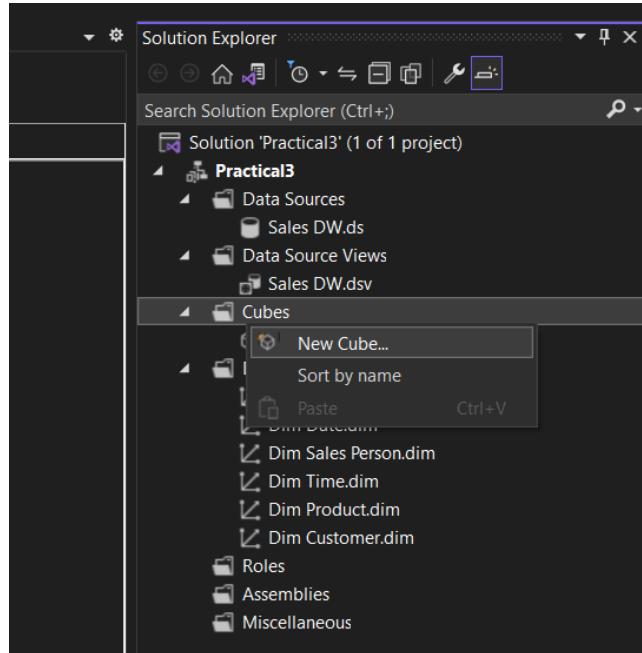
Roll no: 31011020018.



h. Sales DW.dsv appears in Data Source Views in Solution Explorer.

5. Creating New Cube

a. Right click on Cubes -> New Cube



b. Select Use existing tables in Select Creation Method ->Next

c. In Select Measure Group Tables → Select FactProductSales → Click Next

d. In Select Measures → check all measures → Next

e. In Select New Dimensions → Check all Dimensions → Next

f. Click on Finish

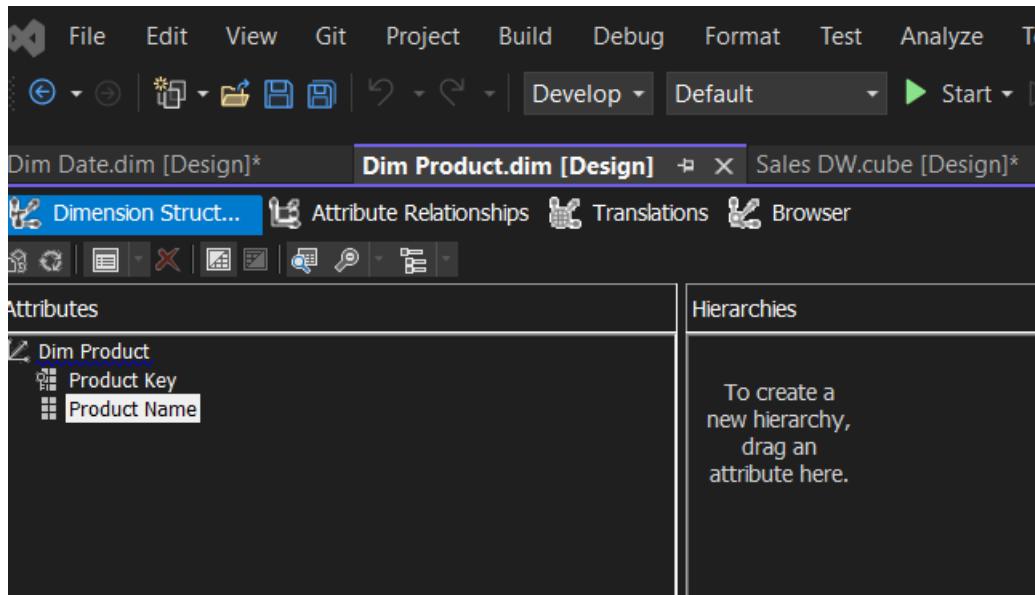
Name: Dharmit Shah.

Roll no: 31011020018.

g. Sales_DW.cube is created.

6. Dimension Modification

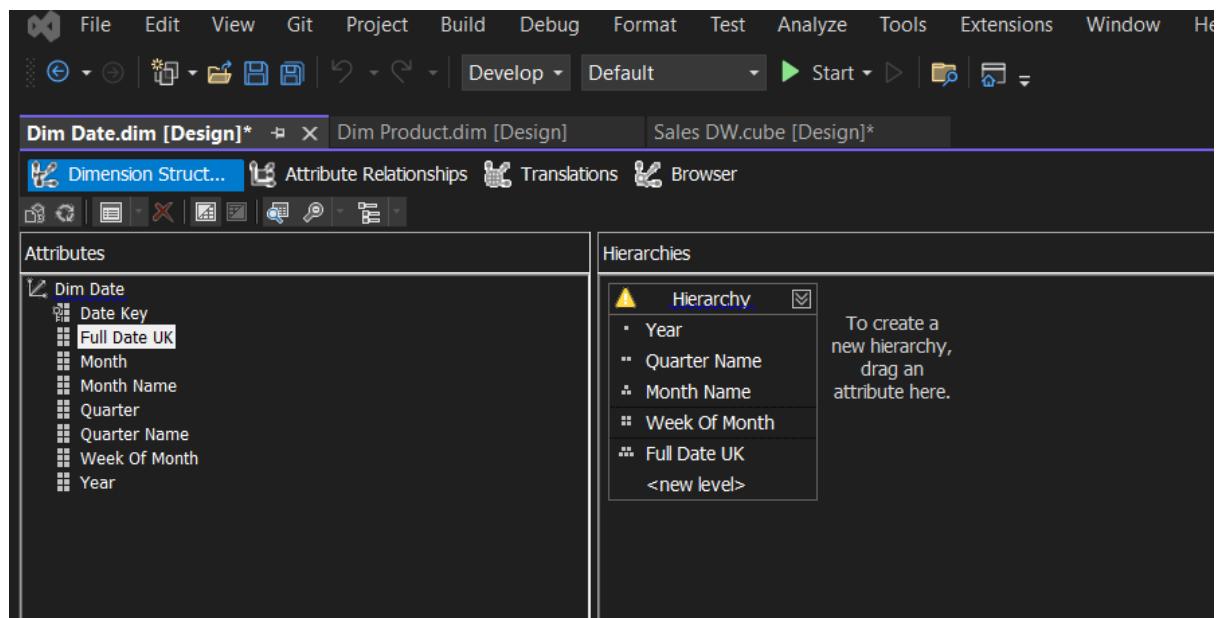
- a. In dimension tab → Double Click Dim Product.dim**
- b. Drag and Drop Product Name from Table in Data Source View and Add in Attribute Pane at left side**



- c. Creating Attribute Hierarchy in Date Dimension. Double click On Dim Date dimension -> Drag and Drop Fields from Table shown in Data Source View to Attributes-> Drag and Drop attributes from leftmost pane of attributes to middle pane of Hierarchy. Drag fields in sequence from Attributes to Hierarchy window (Year, Quarter Name, Month Name, Week of the Month, Full Date UK)**

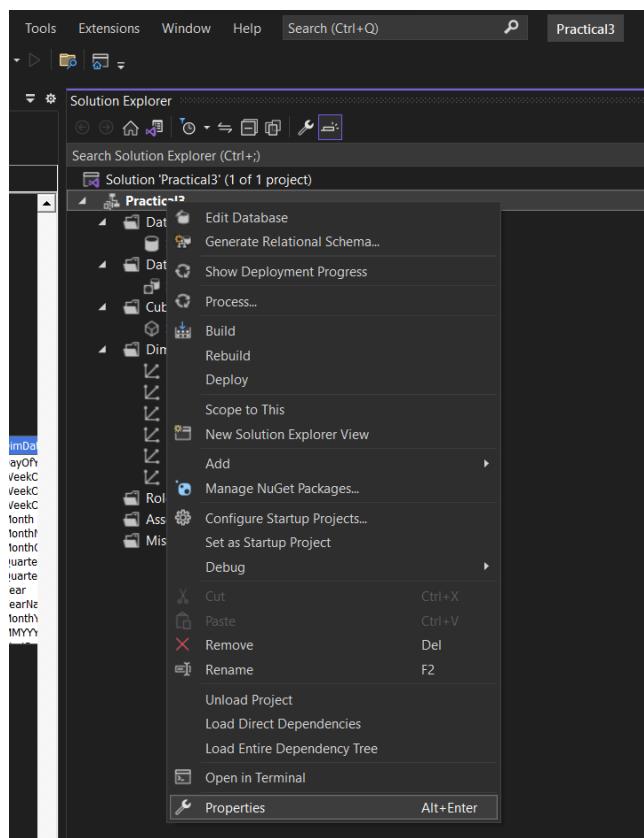
Name: Dharmit Shah.

Roll no: 31011020018.



7. Deploy Cube

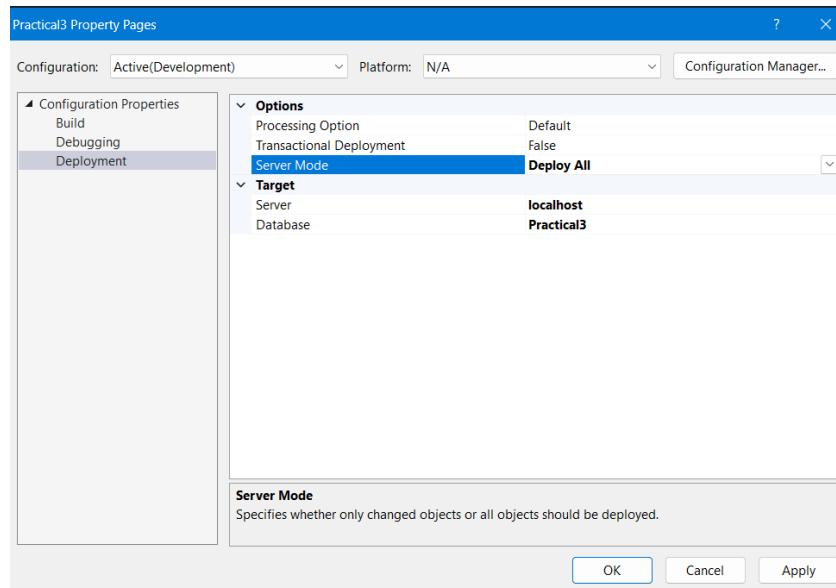
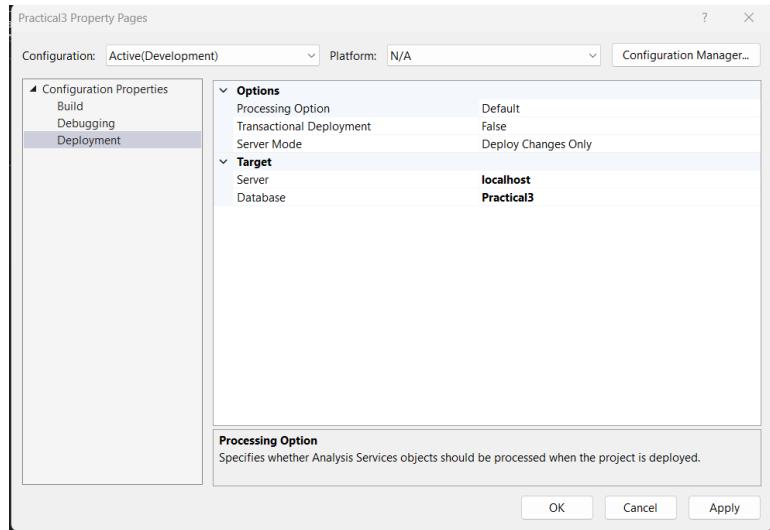
a. Right click on Project name ->Properties



b. Deployment[Server Mode(Deploy All)]->Apply->Ok

Name: Dharmit Shah.

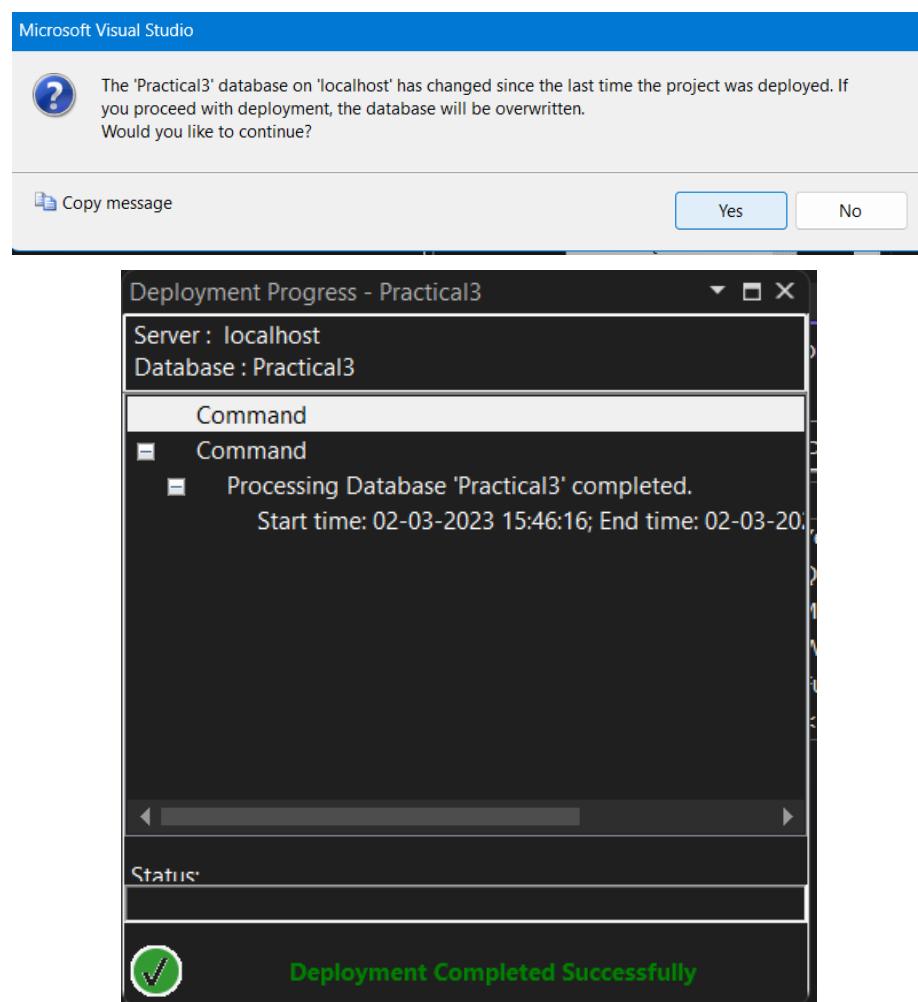
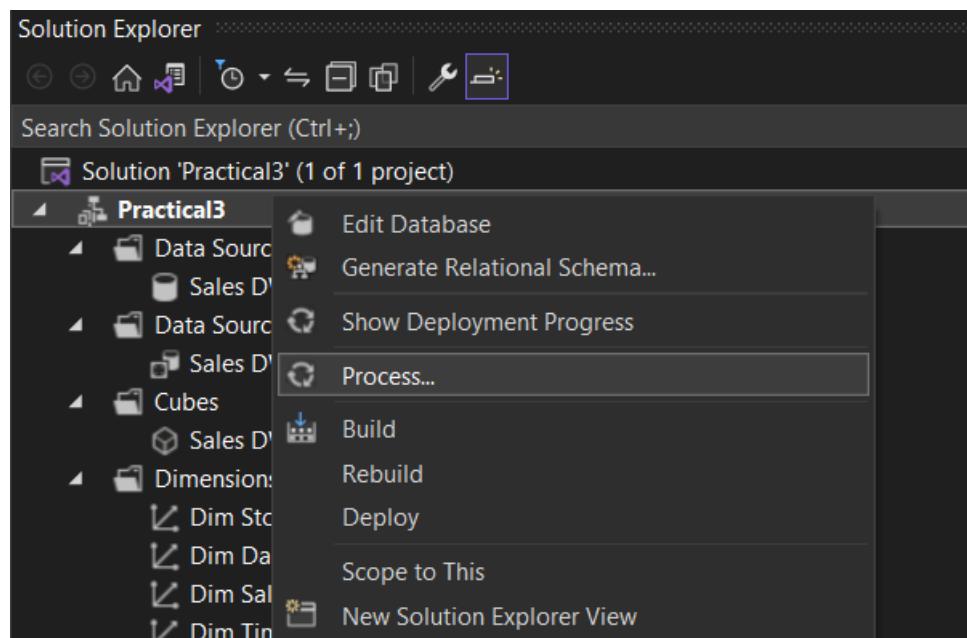
Roll no: 31011020018.



c. Right click on project name ->Process

Name: Dharmit Shah.

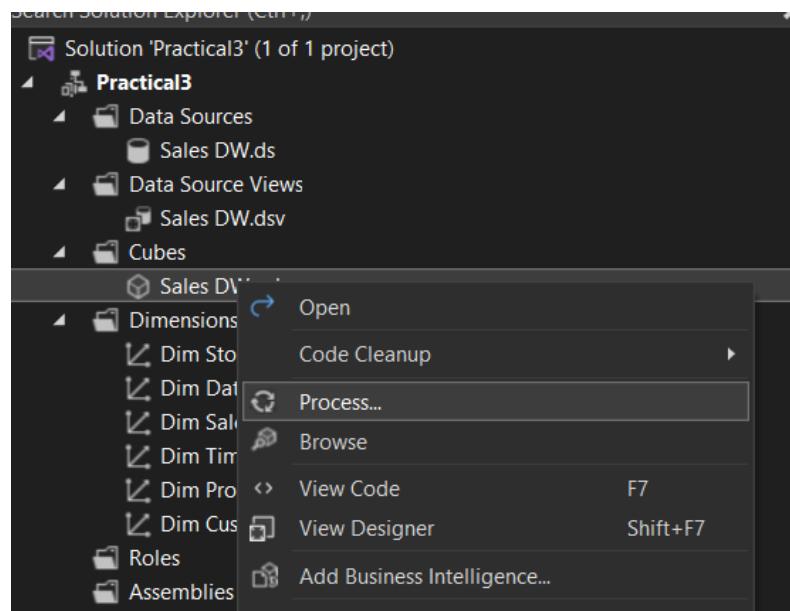
Roll no: 31011020018.



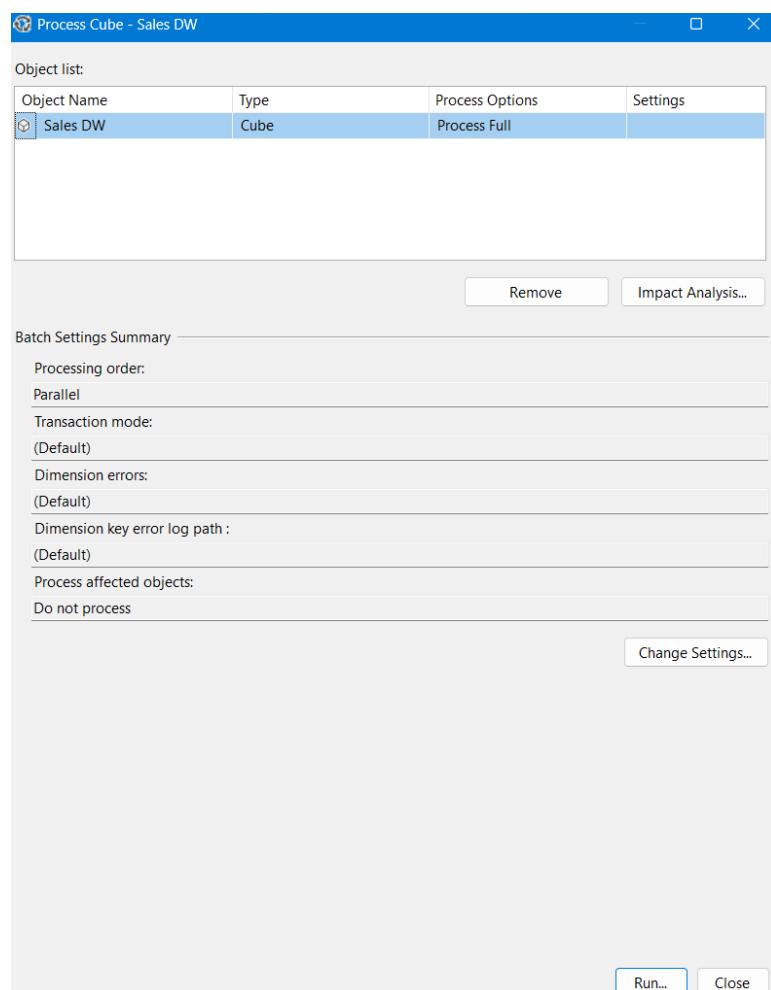
- d. To process cube right click on Sales_DW(cube) ->Process

Name: Dharmit Shah.

Roll no: 31011020018.

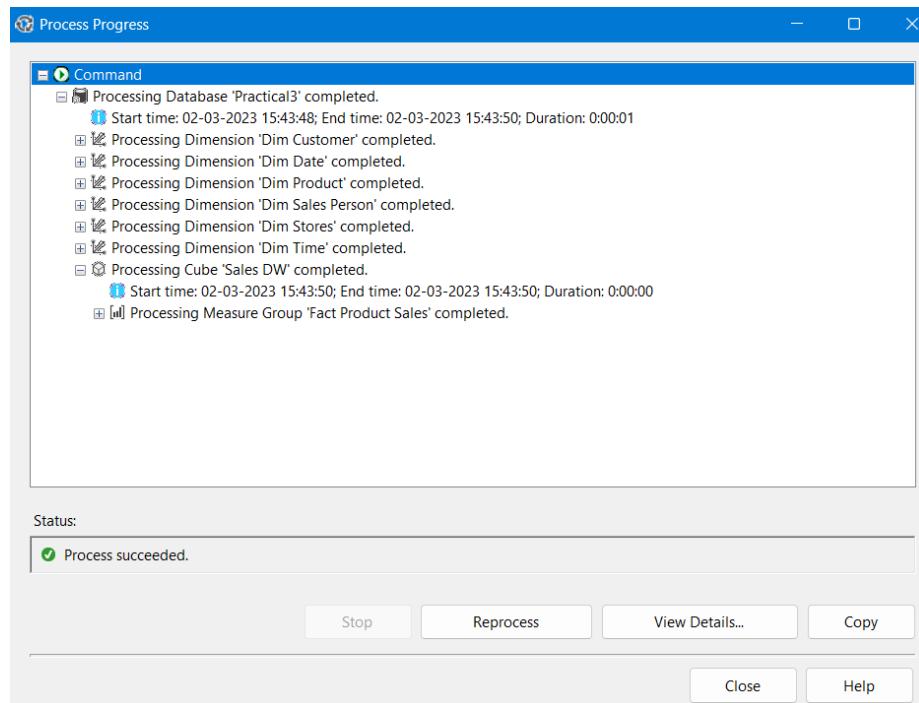


e. Click run

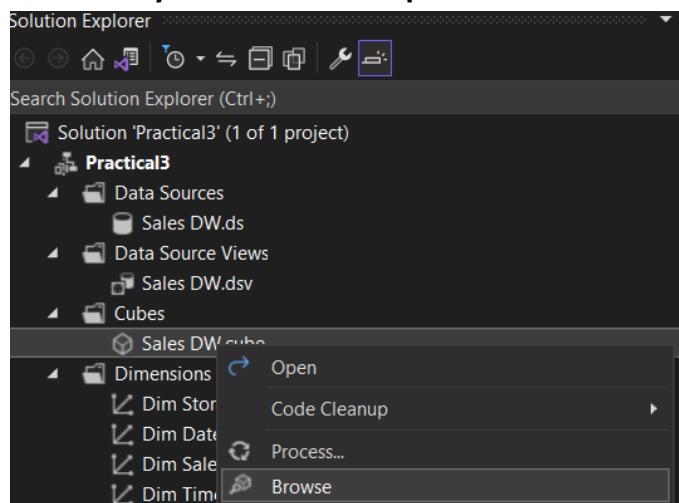


Name: Dharmit Shah.

Roll no: 31011020018.

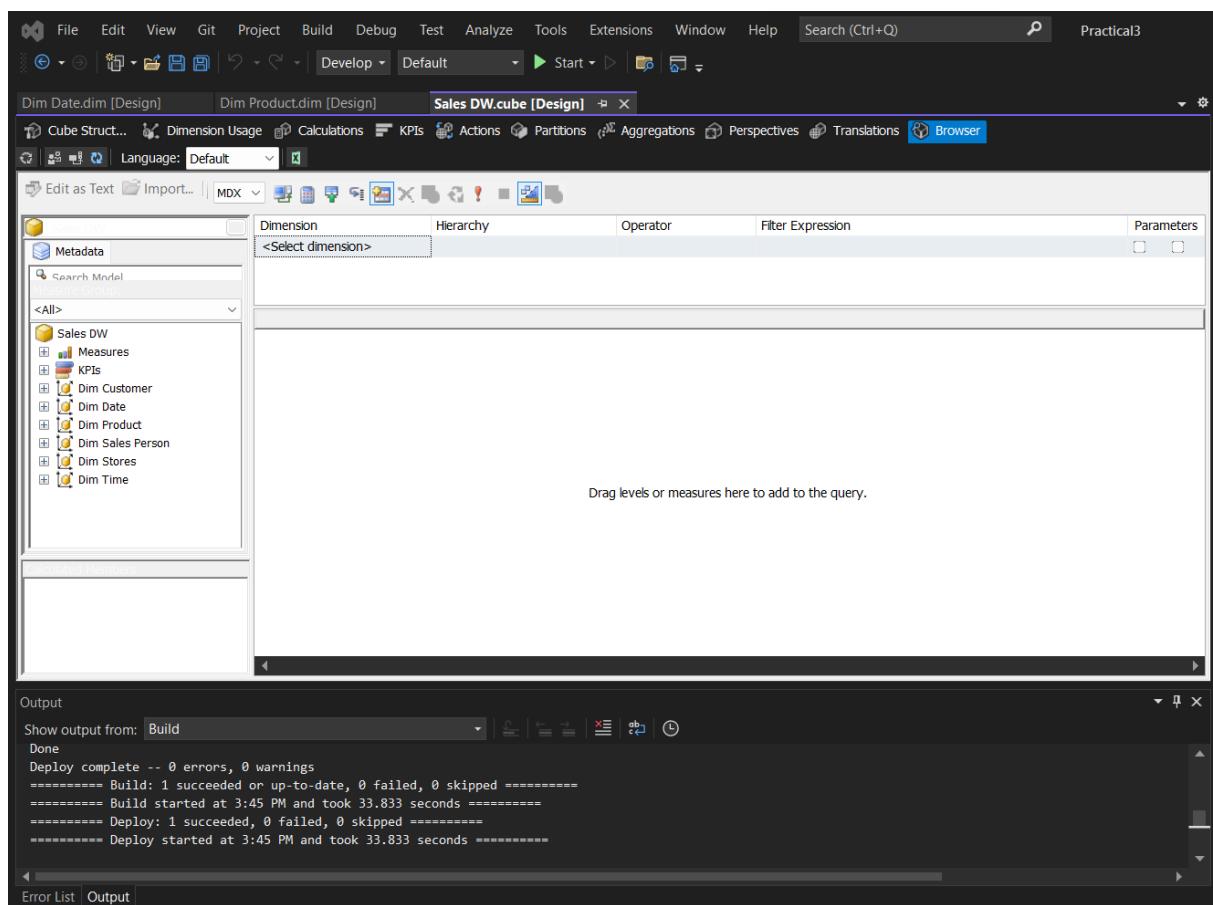
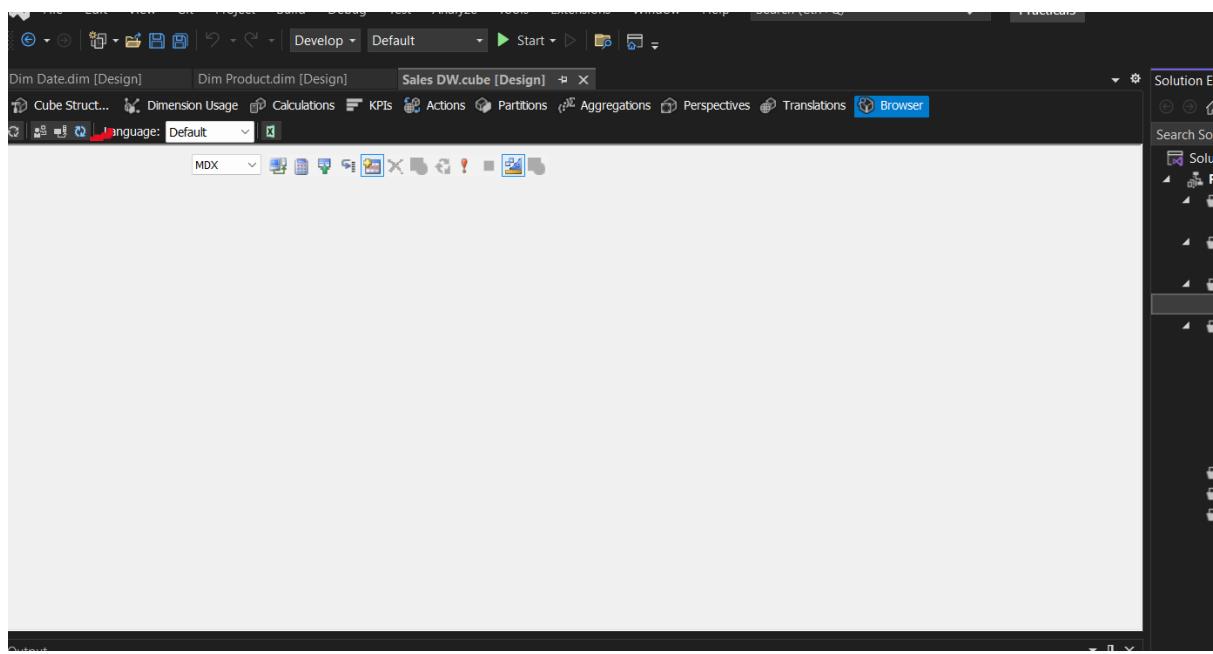


f. Browse the cube for analysis in solution explorer



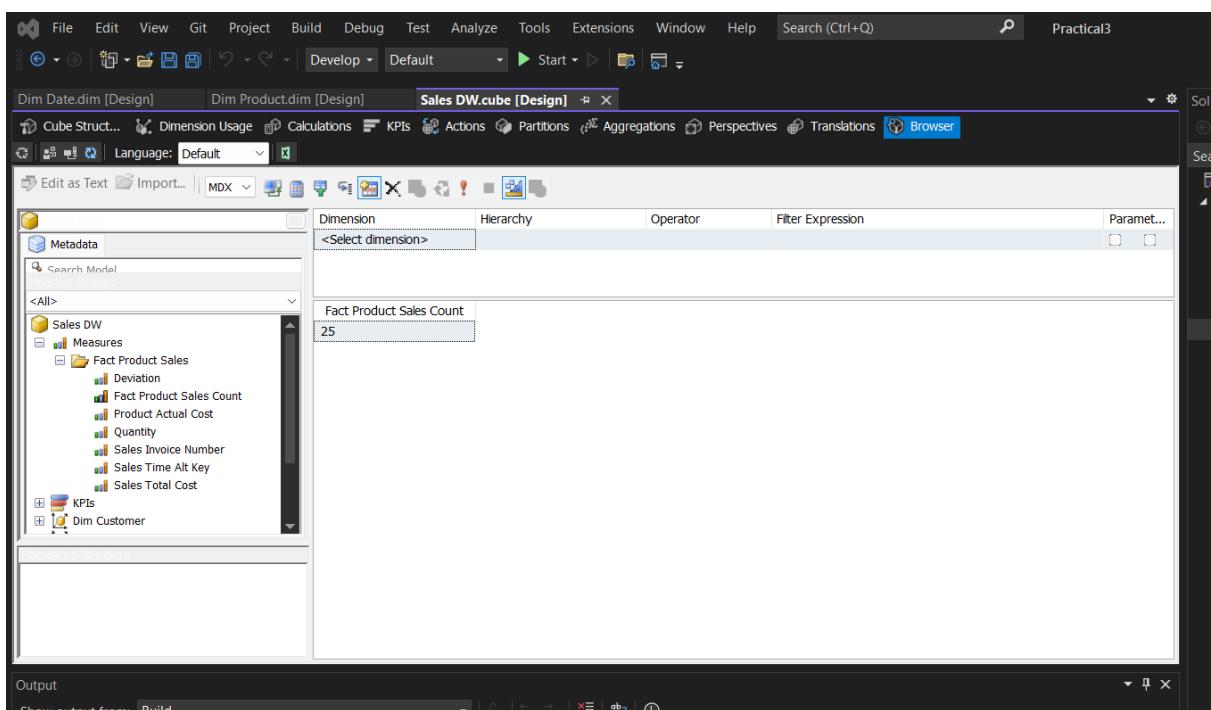
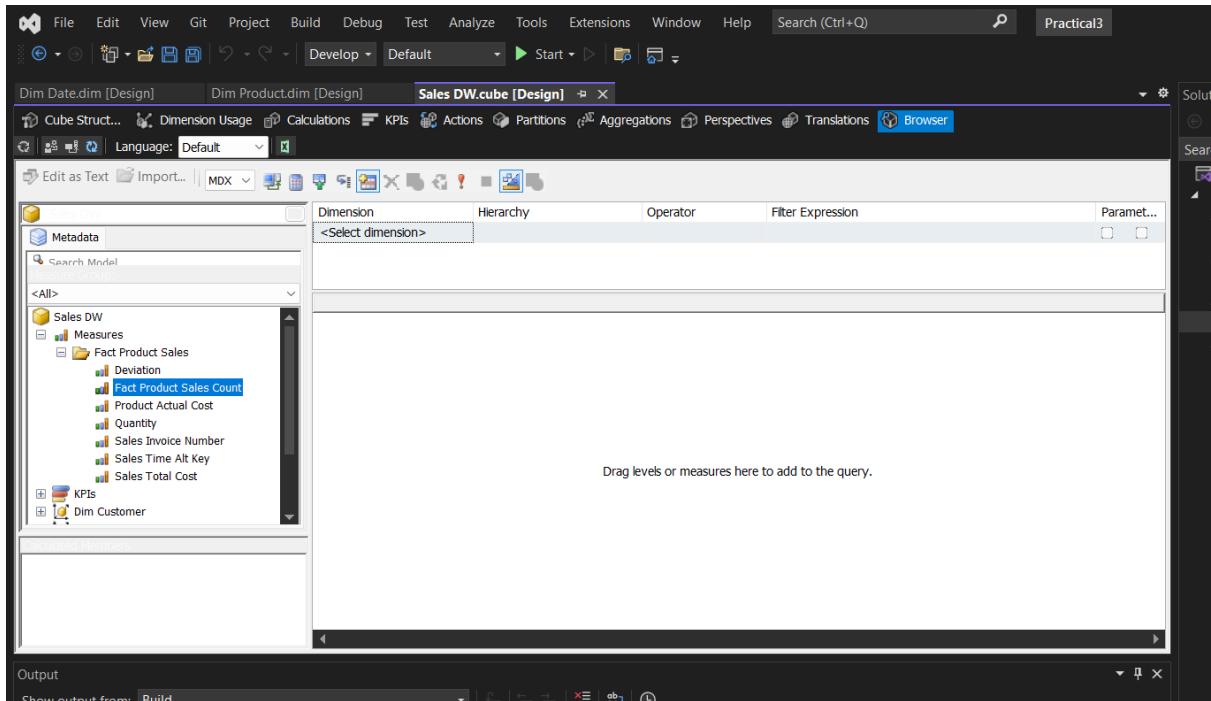
Name: Dharmit Shah.

Roll no: 31011020018.



Name: Dharmit Shah.

Roll no: 31011020018.



Name: Dharmit Shah.

Roll no: 31011020018.

Practical 4

Aim: Apply the what-if Analysis for data visualization.

Step 1: Open Excel Sheet ->Enter data {Total Books(100) %Sold for Highest Price(60.00%) Highest[Unit Profit(50)] Lowest[Unit Profit(20)]}

The screenshot shows an Excel spreadsheet titled "Book2 - Excel". The table has columns labeled A through L. Row 1 contains labels: "Total Books", "%Sold for Highest Price", "No.of books", and "Unit Profit". Row 2 contains values: 100, 60.00%, =A2*B2, and 50. Row 3 contains labels: "Highest" and "Lowest". Row 4 contains values: 20 and Total Price. The formula =A2*B2 is highlighted in green, indicating it is selected or being edited. The formula bar at the top also displays =A2*B2.

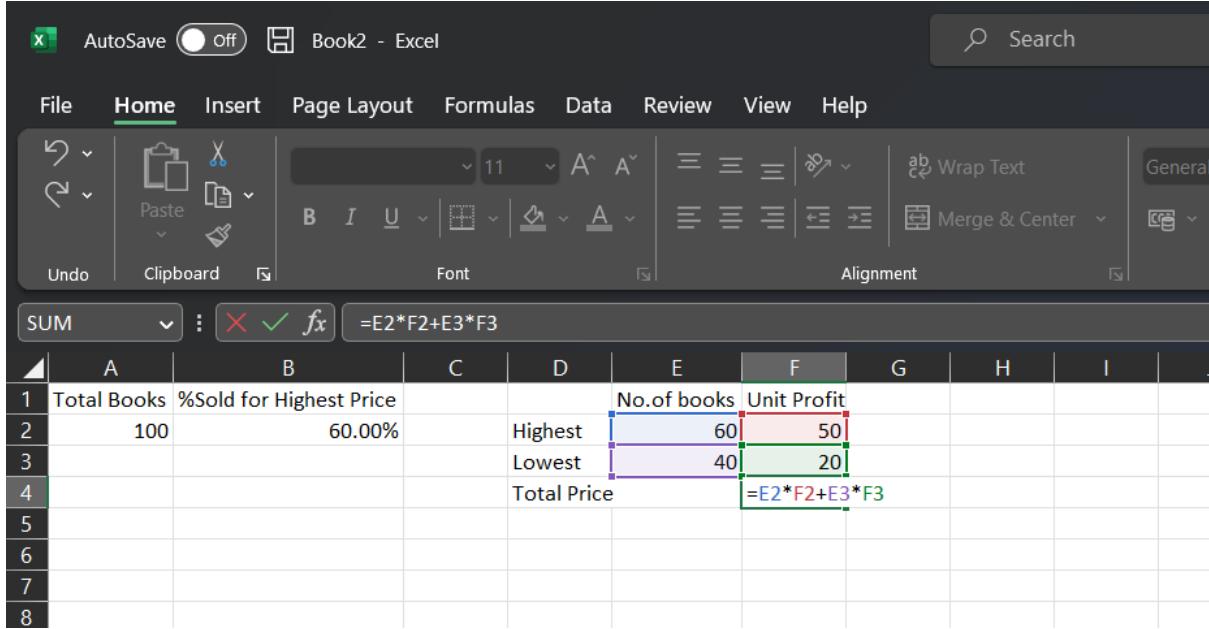
A	B	C	D	E	F	G	H	I	J	K	L
Total Books	%Sold for Highest Price			No.of books	Unit Profit						
100	60.00%		Highest	=A2*B2	50						
			Lowest		20						
				Total Price							

The screenshot shows an Excel spreadsheet titled "Book2 - Excel". The table has columns labeled A through K. Row 1 contains labels: "Total Books", "%Sold for Highest Price", "No.of books", and "Unit Profit". Row 2 contains values: 100, 60.00%, 60, and 50. Row 3 contains labels: "Highest" and "Lowest". Row 4 contains values: =A2*(1-B2) and 20. The formula =A2*(1-B2) is highlighted in green, indicating it is selected or being edited. The formula bar at the top also displays =A2*(1-B2).

A	B	C	D	E	F	G	H	I	J	K
Total Books	%Sold for Highest Price			No.of books	Unit Profit					
100	60.00%		Highest	60	50					
			Lowest	=A2*(1-B2)	20					
				Total Price						

Name: Dharmit Shah.

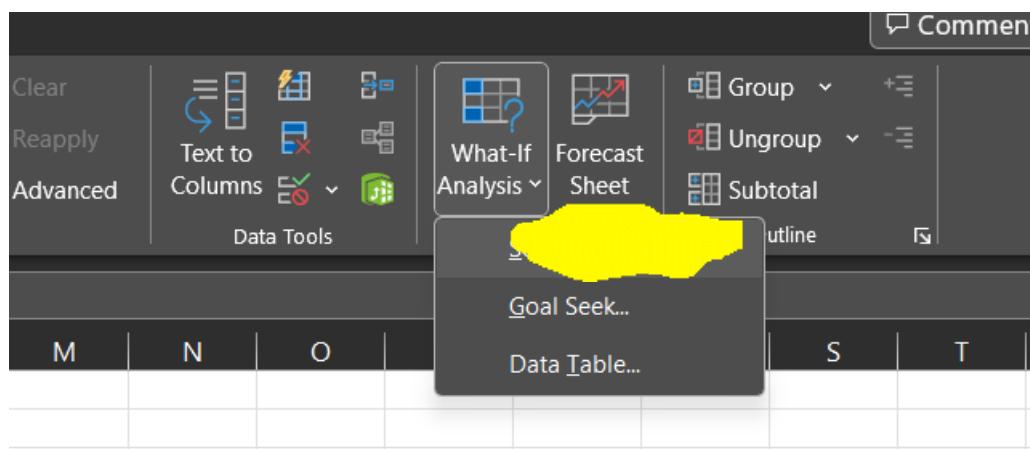
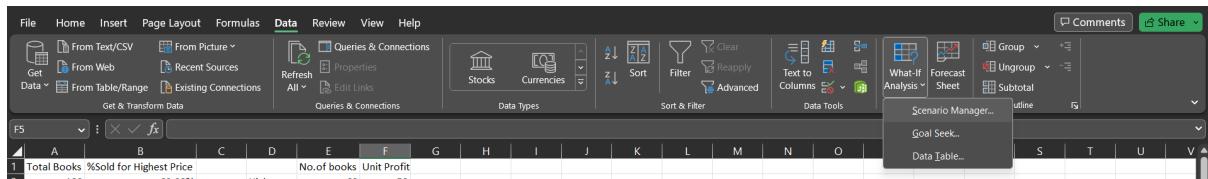
Roll no: 31011020018.



A screenshot of Microsoft Excel showing a simple formula in cell F3. The formula is =E2*F2+E3*F3, which calculates the total profit based on the number of books sold at different prices. The spreadsheet has columns A through I and rows 1 through 8. Row 1 contains column headers: 'Total Books' in A1, '%Sold for Highest Price' in B1, 'No.of books' in E1, and 'Unit Profit' in F1. Rows 2 and 3 show data for 'Highest' and 'Lowest' sales categories respectively. Row 4 shows the total price calculation.

	A	B	C	D	E	F	G	H	I
1	Total Books	%Sold for Highest Price			No.of books	Unit Profit			
2	100	60.00%		Highest	60	50			
3				Lowest	40	20			
4				Total Price	=E2*F2+E3*F3				
5									
6									
7									
8									

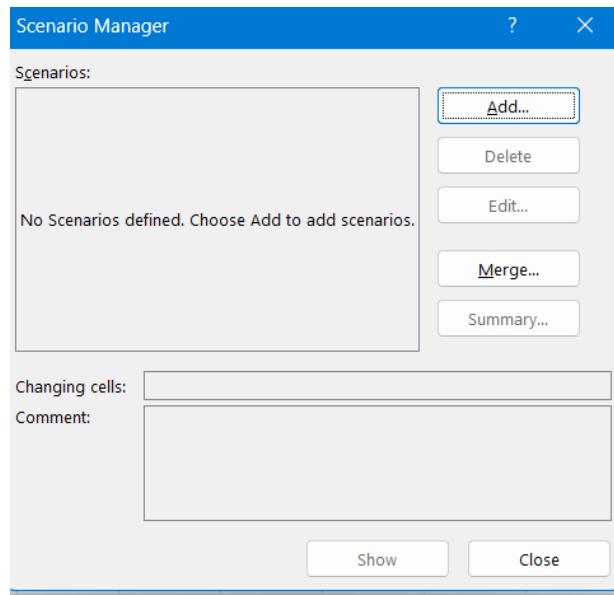
Step 2: In Excel, Click on the Data tab, in the Forecast group, click What-If Analysis & Scenario Manager



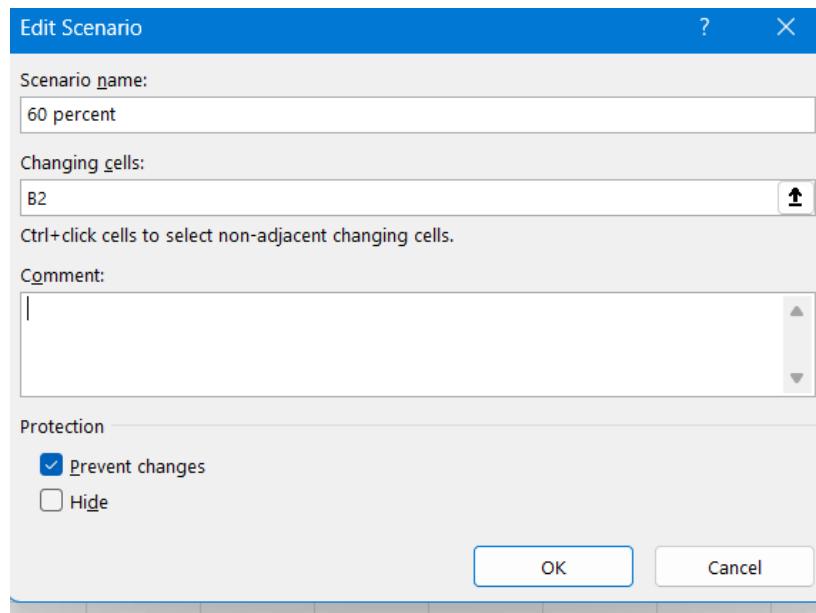
Step 3: Add a scenario by clicking on Add.

Name: Dharmit Shah.

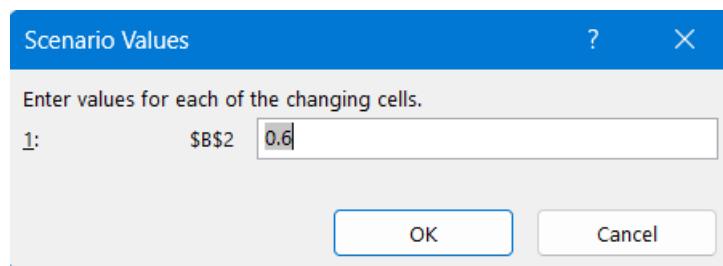
Roll no: 31011020018.



Step 4: Type a name (60percent), select cell \$F\$4 (% sold for the highest price) for the Changing cells and click on OK.



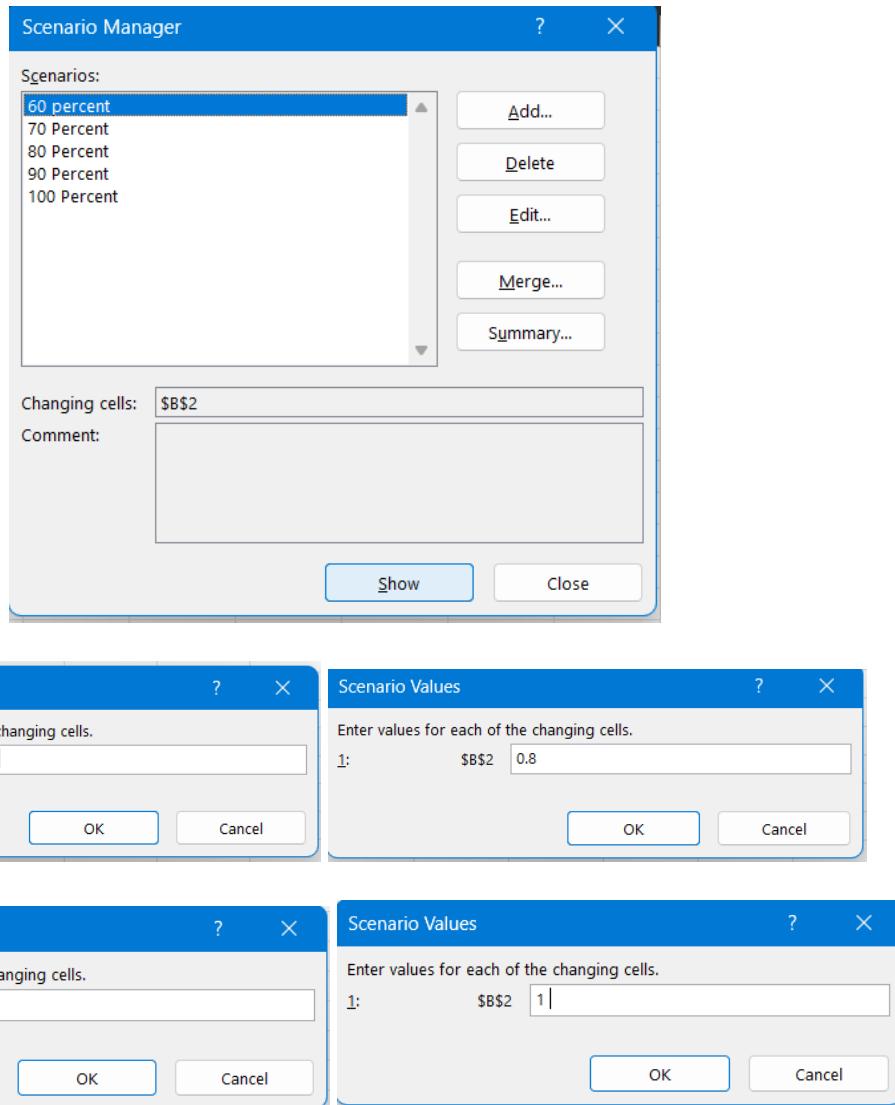
Step 5: Enter the corresponding value 0.6 and click on OK again.



Name: Dharmit Shah.

Roll no: 31011020018.

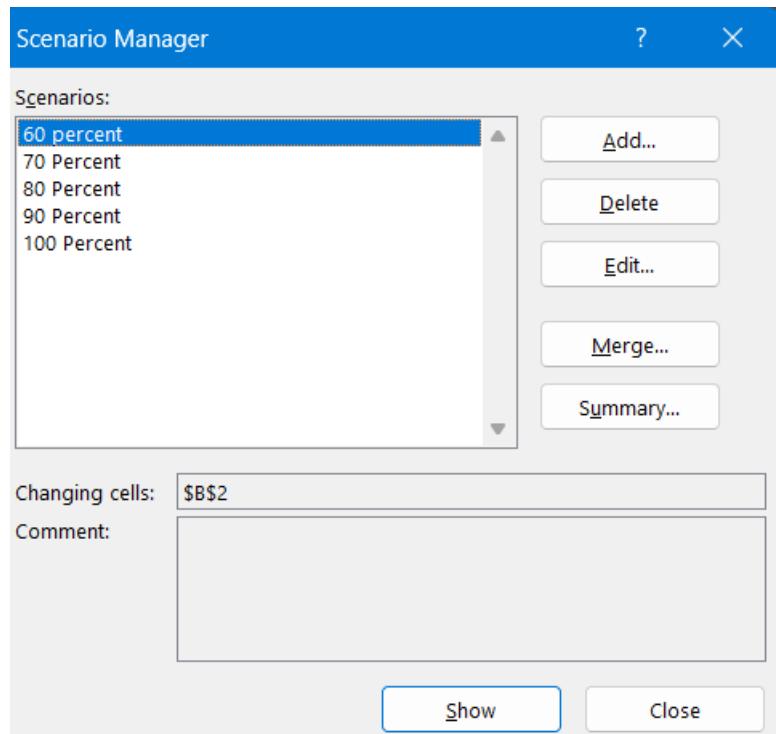
Step 6: Next, add 4 other scenarios (70%, 80%, 90% and 100%) Finally, your Scenario Manager should be consistent with the picture below.



Step 7: Click On Summary

Name: Dharmit Shah.

Roll no: 31011020018.



An Excel spreadsheet shows a table with columns 'No.of books' and 'Unit Profit'. Rows include 'Highest' (60, 50), 'Lowest' (40, 20), and 'Total Price' (3800). The 'Total Price' cell is highlighted with a green dashed border. A 'Scenario Summary' dialog box is open over the spreadsheet. It has 'Report type' set to 'Scenario summary' (radio button selected), a 'Result cells:' field containing '=F\$4', and 'OK' and 'Cancel' buttons at the bottom.

A	B	C	D	E	F	G	H	I	J	K	L	M
Scenario Summary												
Current Values:												
Changing Cells:	\$B\$2	60.00%	60.00%	70.00%	80.00%	90.00%	100.00%					
Result Cells:	\$F\$4	3800	3800	4100	4400	4700	5000					

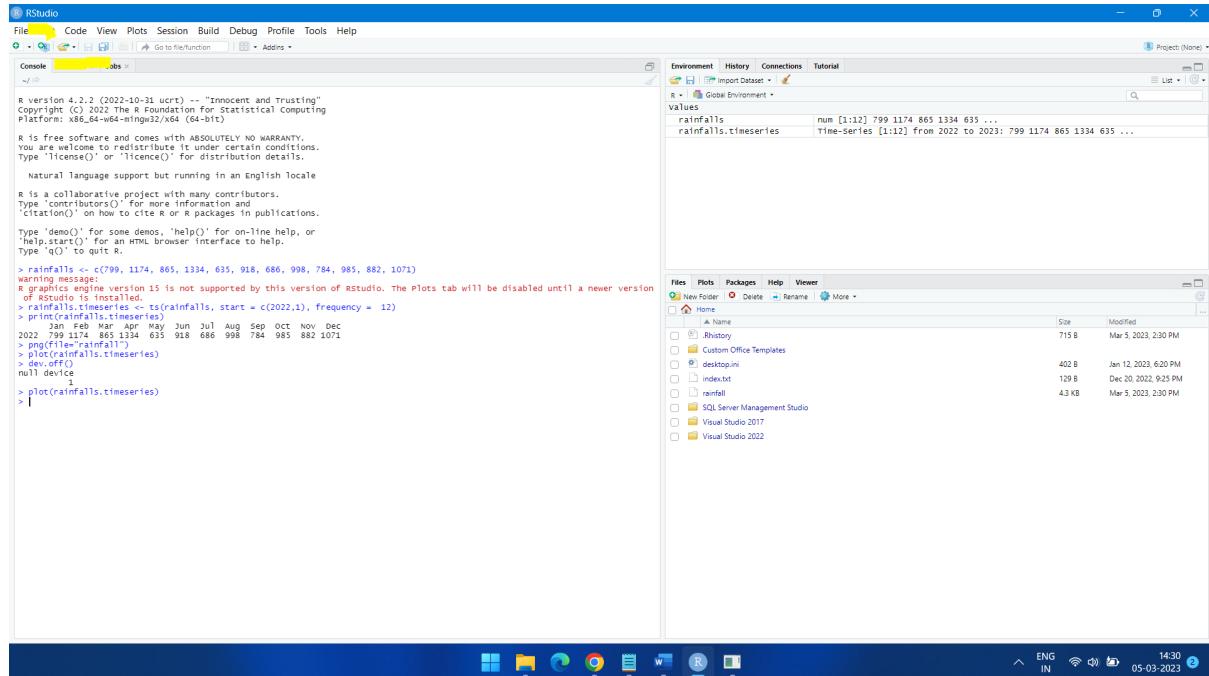
Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.

Name: Dharmit Shah.

Roll no: 31011020018.

Practical 5

Aim: Perform the data classification using classification algorithm



Code

```
rainfalls <- c(799, 1174, 865, 1334, 635, 918, 686, 998, 784, 985, 882, 1071)
```

```
rainfalls.timeseries <- ts(rainfalls, start = c(2022,1), frequency = 12)
```

```
print(rainfalls.timeseries)
```

```
png(file="rainfall")
```

```
plot(rainfalls.timeseries)
```

```
dev.off()
```

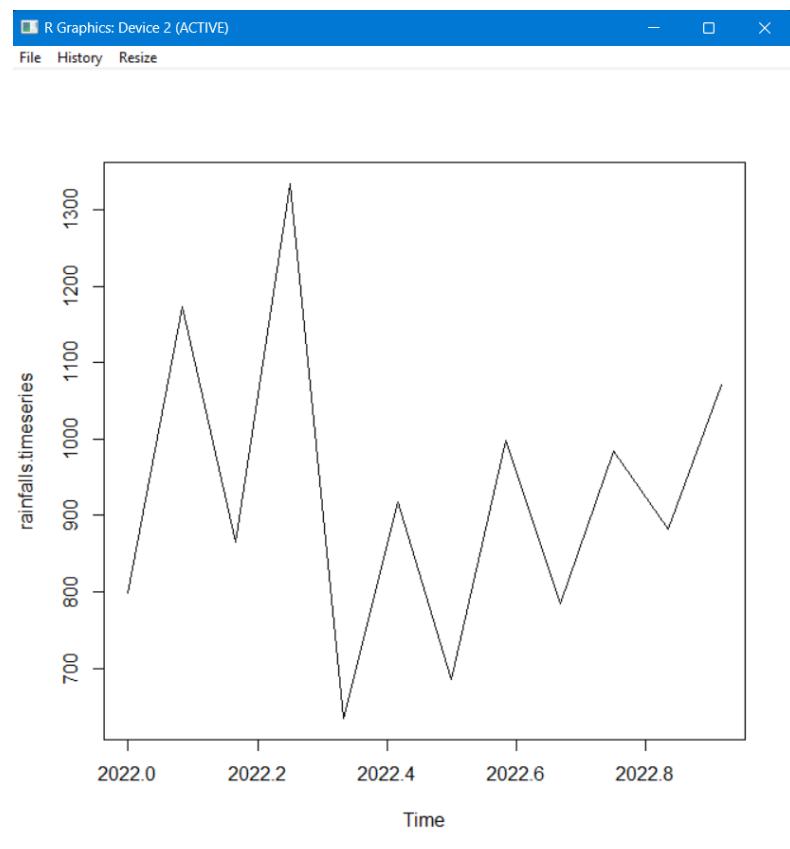
```
plot(rainfalls.timeseries)
```

Output

```
> rainfalls <- c(799, 1174, 865, 1334, 635, 918, 686, 998, 784, 985, 882, 1071)
> rainfalls.timeseries <- ts(rainfalls, start = c(2022,1), frequency = 12)
> print(rainfalls.timeseries)
   Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
2022 799 1174 865 1334 635 918 686 998 784 985 882 1071
> png(file="rainfall")
> plot(rainfalls.timeseries)
> dev.off()
windows
  2
> plot(rainfalls.timeseries)
> |
```

Name: Dharmit Shah.

Roll no: 31011020018.



Name: Dharmit Shah.

Roll no: 31011020018.

Practical 6

Aim: K-Means clustering using R

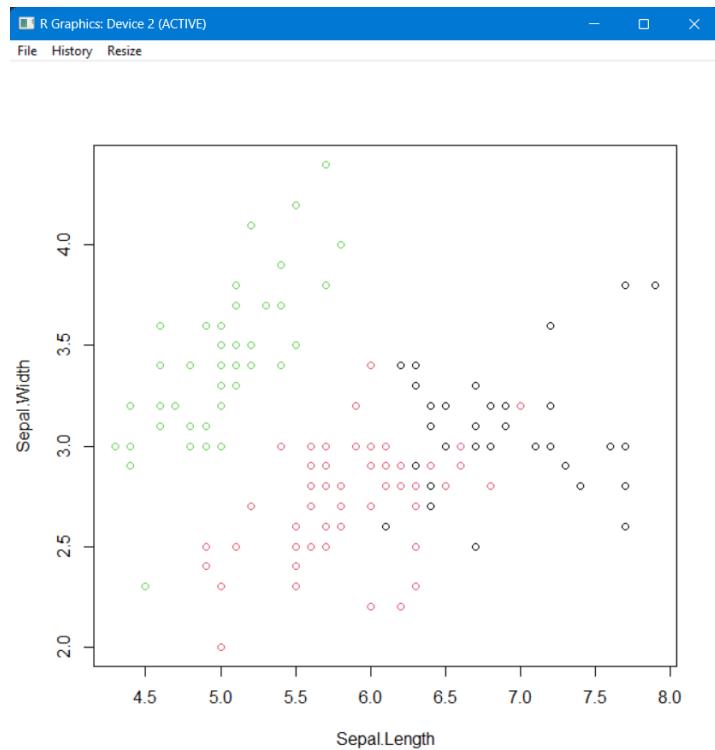
Code

```
newiris <- iris  
  
newiris$Species <- NULL  
  
kc <- kmeans(newiris,3)  
  
kc  
  
table(iris$Species, kc$cluster)  
  
plot(newiris[c("Sepal.Length", "Sepal.Width")], col=kc$cluster)  
  
points(kc$centers[, c("Sepal.Length", "Sepal.Width")], col=1:3, pch = 8, cex = 2)  
  
dev.off()  
  
plot(newiris[c("Sepal.Length", "Sepal.Width")], col=kc$cluster)
```

Output

Name: Dharmit Shah.

Roll no: 31011020018.



Name: Dharmit Shah.

Roll no: 31011020018.

Practical 7

Aim: Prediction Using Linear Regression

Code

```
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(relation)
```

```
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))
```

```
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)
```

```
a <-data.frame(x=170)
```

```
result<-predict(relation,a)
```

```
print(result)
```

```
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)
```

```
png(file="linearregression.png")
```

```
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")
```

Name: Dharmit Shah.

Roll no: 31011020018.

dev.off()

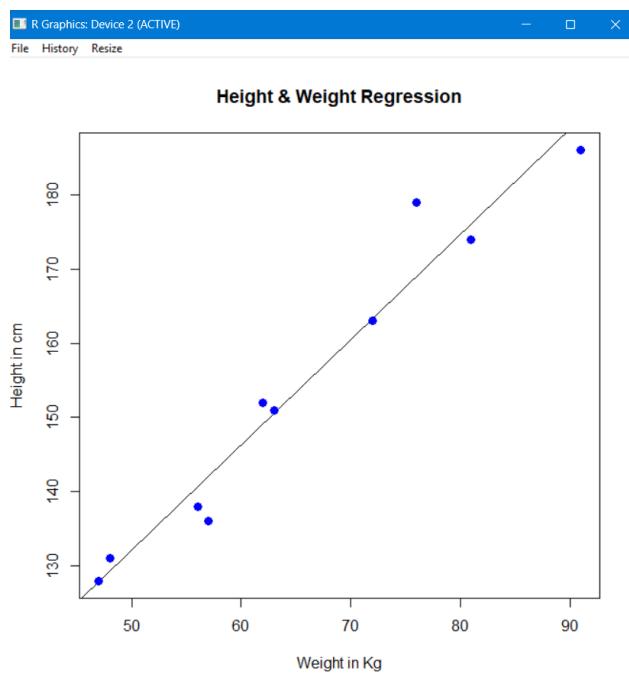
```
plot(y,x,col="blue",main="Height & Weight  
Regression",abline(lm(x~y)),cex=1.3,pch=16,xlab="Weight in Kg",ylab="Height in  
cm")
```

Output

```
> 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(relation)  
  
Call:  
lm(formula = y ~ x)  
  
Coefficients:  
(Intercept) x  
-38.4551 0.6746  
  
>  
> 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  
  
>  
> 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)  
> a <-data.frame(x=170)  
> result<-predict(relation,a)  
> print(result)  
 1  
76.22869  
  
>  
> 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)  
> png(file="linearregression.png")  
> 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")  
> dev.off()  
null device  
 1  
> 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")  
>
```

Name: Dharmit Shah.

Roll no: 31011020018.



Name: Dharmit Shah.

Roll no: 31011020018.

Practical 8

Aim: Perform the logistic regression on the given data warehouse data

Code

```
quality <- read.csv("C/quality.csv")
str(quality)
table(quality$PoorCare)
98/131
install.packages("caTools")
library(caTools)
set.seed(88)
split=sample.split(quality$PoorCare,SplitRatio=0.75)
split
qualityTrain=subset(quality,split==TRUE)
qualityTest=subset(quality,split==FALSE)
nrow(qualityTrain)
QualityLog = glm(PoorCare ~ OfficeVisits +
Narcotics,data=qualityTrain,family=binomial)
summary(QualityLog)
predictTrain=predict(QualityLog,type="response")
summary(predictTrain)
tapply(predictTrain,qualityTrain$PoorCare,mean)
table(qualityTrain$PoorCare,predictTrain>0.5)
table(qualityTrain$PoorCare,predictTrain>0.7)
8/25
73/74
table(qualityTrain$PoorCare,predictTrain>0.2)
16/25
```

Name: Dharmit Shah.

Roll no: 31011020018.

54/74

```
install.packages("ROCR")
```

```
library(ROCR)
```

```
ROCRpred = prediction(predictTrain,qualityTrain$PoorCare)
```

```
ROCRperf = performance(ROCRpred,"tpr","fpr")
```

```
plot(ROCRperf)
```

```
plot(ROCRperf,colorize=TRUE)
```

```
plot(ROCRperf,colorize=TRUE,print.cutoffs.at=seq(0,1,by=0.1),text.adj=c(-0.2,1.7))
```

Output

```
> quality <- read.csv("C:/Users/roshn/Downloads/quality.csv")
> str(quality)
'data.frame': 131 obs. of 14 variables:
 $ MemberID      : int 1 2 3 4 5 6 7 8 9 10 ...
 $ InpatientDays : int 0 1 0 0 8 2 16 2 2 4 ...
 $ ERVisits       : int 0 1 0 1 2 0 1 0 1 2 ...
 $ OfficeVisits   : int 18 6 5 19 19 9 8 8 4 0 ...
 $ Narcotics      : int 1 1 3 0 3 2 1 0 3 2 ...
 $ DaysSinceLastERVisit: num 731 411 731 158 449 ...
 $ Pain           : int 10 0 10 34 10 6 4 5 5 2 ...
 $ TotalVisits    : int 18 8 5 20 29 11 25 10 7 6 ...
 $ ProviderCount  : int 21 27 16 14 24 40 19 11 28 21 ...
 $ MedicalClaims  : int 93 19 27 59 51 53 40 28 20 17 ...
 $ ClaimLines     : int 222 115 148 242 204 156 261 87 98 66 ...
 $ StartedOnCombination: logi FALSE FALSE FALSE FALSE FALSE FALSE ...
 $ AcuteDrugGapSmall : int 0 1 5 0 0 4 0 0 0 0 ...
 $ PoorCare        : int 0 0 0 0 0 1 0 0 1 0 ...
> table(quality$PoorCare)

 0 1
98 33
> 98/131
[1] 0.7480916

--> 98/131
[1] 0.7480916
> install.packages("caTools")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/roshn/AppData/Local/R/win-library/4.2'
(as 'lib' is unspecified)
also installing the dependency 'bitops'

Warning in install.packages :
  the 'wininet' method is deprecated for http:// and https:// URLs
  try the 'https://cran.rstudio.com/bin/windows/contrib/4.2/bitops_1.0-7.zip'
Content type 'application/zip' length 31679 bytes (30 KB)
downloaded 30 KB

Warning in install.packages :
  the 'wininet' method is deprecated for http:// and https:// URLs
  trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/caTools_1.18.2.zip'
Content type 'application/zip' length 246150 bytes (240 KB)
downloaded 240 KB

package 'bitops' successfully unpacked and MD5 sums checked
package 'caTools' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
  C:/Users/roshn/AppData/Local/Temp/RtmpsofwrX downloaded_packages

C:/Users/roshn/AppData/Local/Temp/RtmpsofwrX downloaded_packages
> library(caTools)
> set.seed(88)
> split=sample.split(quality$PoorCare,splitRatio=0.75)
> split
 [0] TRUE TRUE TRUE TRUE FALSE TRUE FALSE TRUE FALSE TRUE TRUE TRUE FALSE FALSE FALSE TRUE
UE TRUE
[37] FALSE TRUE TRUE TRUE FALSE TRUE FALSE TRUE FALSE TRUE FALSE TRUE TRUE FALSE FALSE TRUE F
SE TRUE
[73] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE
UE FALSE
[109] TRUE FALSE FALSE TRUE TRUE FALSE TRUE
> qualityTrain=subset(quality,split==TRUE)
> qualityTest=subset(quality,split==FALSE)
> nrow(qualityTrain)
[1] 99
>
```

Name: Dharmit Shah.

Roll no: 31011020018.

```

> QualityLog = glm(PoorCare ~ Officevisits + Narcotics,data=qualityTrain,family=binomial)
> summary(QualityLog)

call:
glm(formula = PoorCare ~ Officevisits + Narcotics, family = binomial,
     data = qualityTrain)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-2.06303 -0.63155 -0.50503 -0.09689  2.16686 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -2.64613   0.52357 -5.054 4.33e-07 ***
Officevisits  0.08212   0.03055  2.688  0.00718 **  
Narcotics    0.07630   0.03205  2.381  0.01728 *   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 111.888 on 98 degrees of freedom
Residual deviance: 89.127 on 96 degrees of freedom
AIC: 95.127

Number of Fisher Scoring iterations: 4

> predictTrain=predict(QualityLog,type="response")
> summary(predictTrain)
   Min. 1st Qu. Median Mean 3rd Qu. Max. 
0.06623 0.11912 0.15967 0.25253 0.26765 0.98456 
> tapply(predictTrain,qualityTrain$PoorCare,mean)
 0          1        
0.1894512 0.4392246 
> table(qualityTrain$PoorCare,predictTrain>0.5)

  FALSE TRUE
0     70   4
1     15  10

> table(qualityTrain$PoorCare,predictTrain>0.7)

  FALSE TRUE
0     73   1
1     17   8

> 8/25
[1] 0.32
> 73/74
[1] 0.9864865
> table(qualityTrain$PoorCare,predictTrain>0.2)

  FALSE TRUE
0     54   20
1      9   16

> 16/25
[1] 0.64
> 54/74
[1] 0.7297297

L4J v. 1.4.2.429
> install.packages("ROCR")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding.

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/users/roshn/AppData/Local/R/win-library/4.2'
(as 'lib' is unspecified)
also installing the dependencies 'gtools', 'gplots'

Warning in install.packages :
  the 'wininet' method is deprecated for http:// and https:// URLs
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/gtools_3.9.4.zip'
Content type 'application/zip' length 359791 bytes (351 kB)
downloaded 351 kB

Warning in install.packages :
  the 'wininet' method is deprecated for http:// and https:// URLs
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/gplots_3.1.3.zip'
Content type 'application/zip' length 603202 bytes (589 kB)
downloaded 589 kB

Warning in install.packages :
  the 'wininet' method is deprecated for http:// and https:// URLs
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/ROCR_1.0-11.zip'
Content type 'application/zip' length 453427 bytes (442 kB)
downloaded 442 kB

package 'gtools' successfully unpacked and MD5 sums checked
package 'gplots' successfully unpacked and MD5 sums checked
package 'ROCR' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/users/roshn/AppData/Local/Temp/Rtmpsofrwx/downloaded_packages
> library(ROCR)
>

```

Name: Dharmit Shah.

Roll no: 31011020018.

```
> ROCRpred = prediction(predictTrain,qualityTrain$PoorCare)
> ROCRperf = performance(ROCRpred,"tpr","fpr")
> plot(ROCRperf)
> plot(ROCRperf,colorize=TRUE)
> plot(ROCRperf,colorize=TRUE,print.cutoffs.at=seq(0,1,by=0.1),text.adj=c(-0.2,1.7))
> |
```

