

Class : MSC CS PART I

Sem : I

Subject : Data Warehousing & Data Mining

Paper : IV

Academic Year : 2020-21

Roll No : 03

INDEX

S.NO	Date	Topic	Page No	Sign
1		Creation of Dimensions and Fact tables.		
2		Create Data Source using SSAS (SQL Server Analysis Services.)		
3		Create Data Source View using SSAS (SQL Server Analysis Services.)		
4		Create cube using SSAS(SQL Server Analysis Services.) and process the cube.		
5		View cube data in multidimensional Format. (Excel Pivot Chart.)		
6		Working with measures in the cube.		
7		Creating an Excel Pivot Table and Pivot Chart by using the OLAP cube data.		
8		Firing Queries on Tables.		
9		Calculation & KPI		
10		Data pre-processing.		
11		Data discretization.		
12		Classification problems.		
13		Clustering Analysis.		
14		Association Rule Mining.		
15		Data visualization.		

Project Name : - Sales Product

Practical No : 1

Aim : Creation of Dimensions and Fact tables.

Solution :

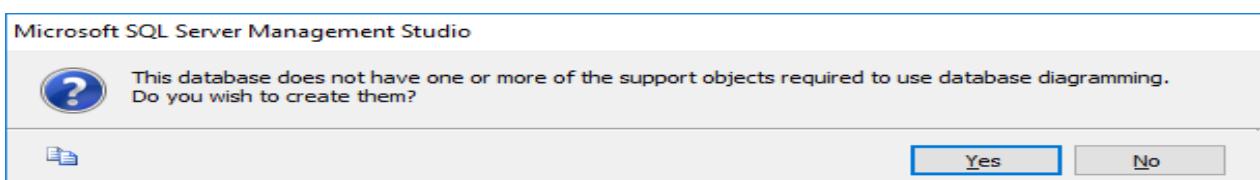
Open Application -> Microsoft SQL Server 2008 R2 -> SQL Server Management Studio

1. Select Connect Tab -> Database Engine -> Select Server Name(local)
2. Right Click the Database -> New Database
3. Types "SalesProduct" as the database name, click on OK to close the dialog box and to create the database.

Create a Database Diagrams :-

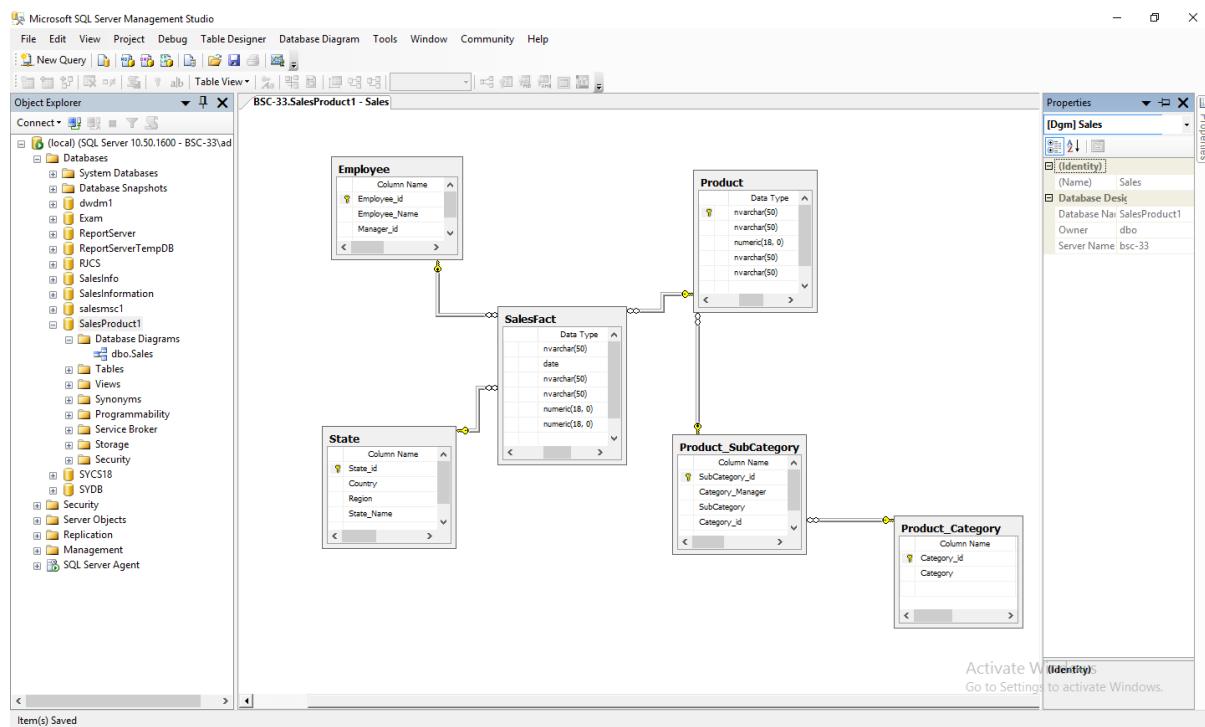
Expand the "SalesProduct" database folder.

1. Click on Database Diagrams to expand it



On click of it, above Dialog box appears, click on Yes to close it.

2. Right Click on Database Diagrams -> New Database Diagrams
3. Create fact and Dimension Tables. (Right click on surface, choose New Table to add tables on Database Diagrams.)



4. Establish relationship between fact and dimension tables.

5. Save Database Diagrams with name as “Sales”. (After saving Database Diagrams fact and dimension tables are automatically placed in Table tab.)

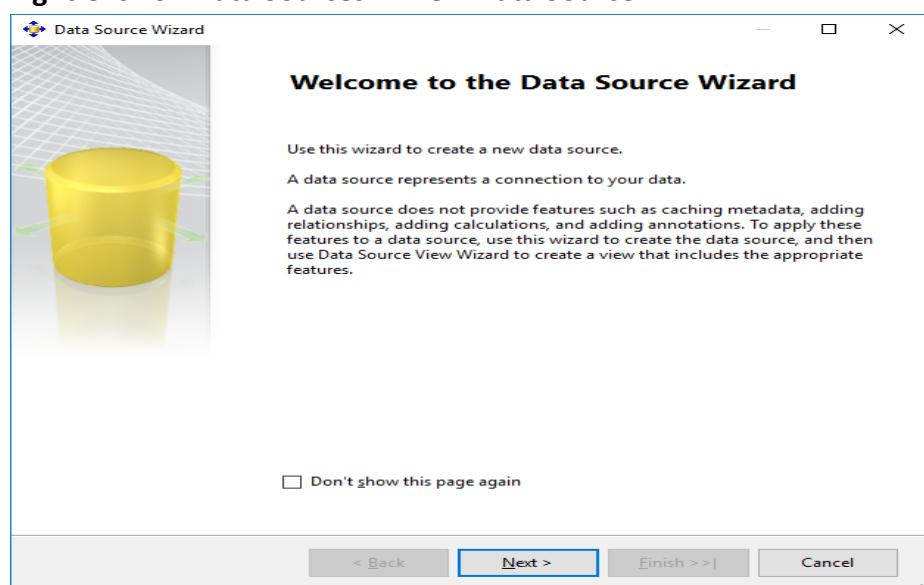
Practical No: - 2

Aim : Create Data Source using SSAS(SQL Server Analysis Services.)

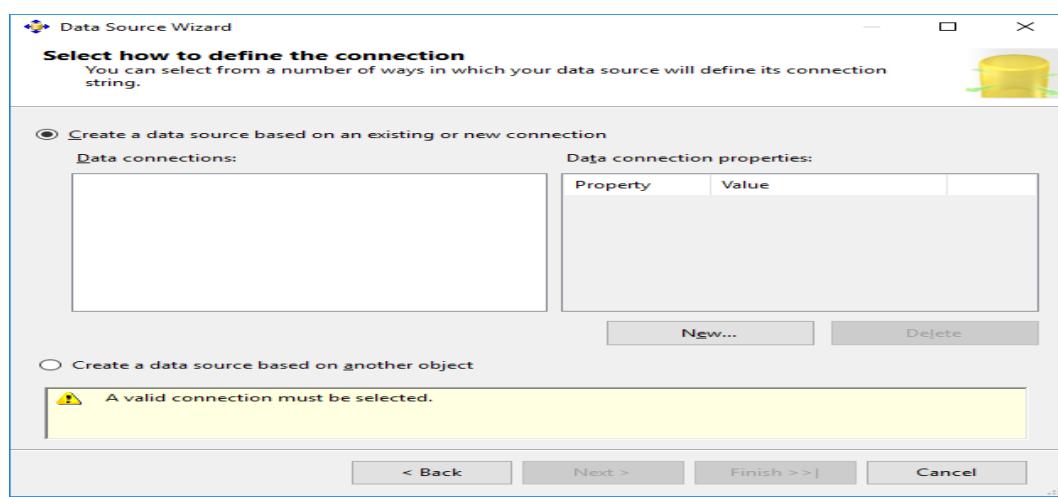
Solution :

Open Application -> Microsoft SQL Server 2008 R2 -> SQL Server Business Intelligence Development Studio

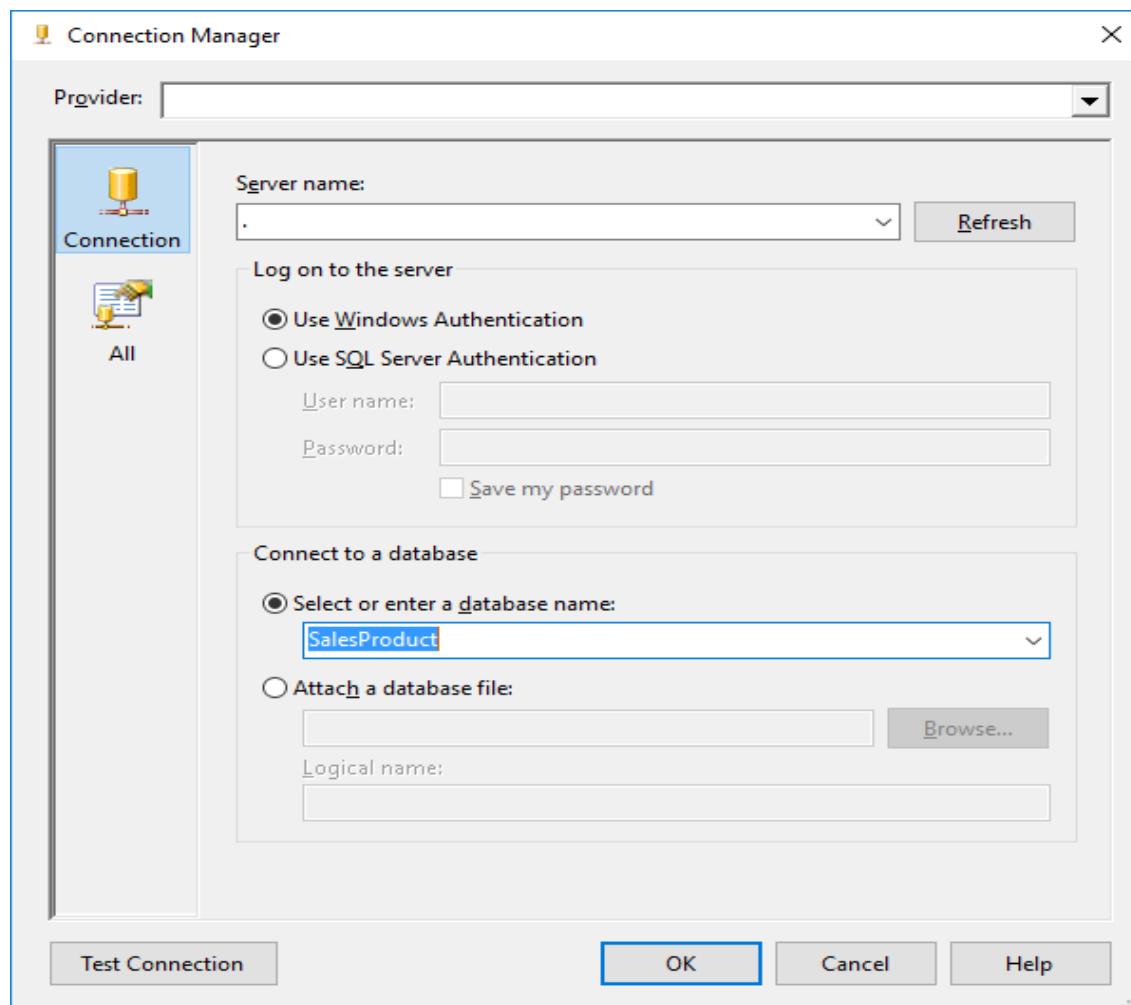
1. Select File -> New Project -> Choose Analysis Service Project -> Name it as "SalesProduct_BIPrj" and click on OK.
2. Right Click on Data Sources -> New Data Source



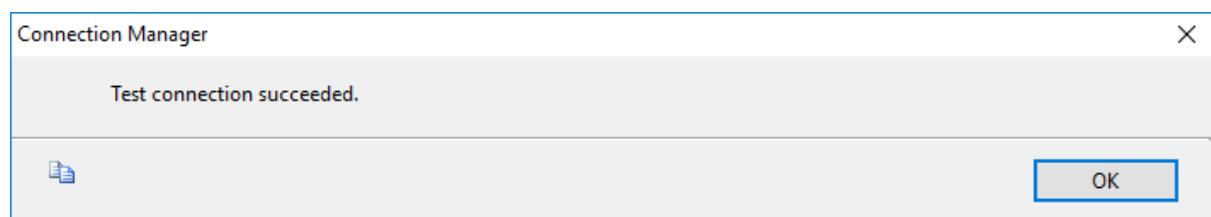
Click on Next .



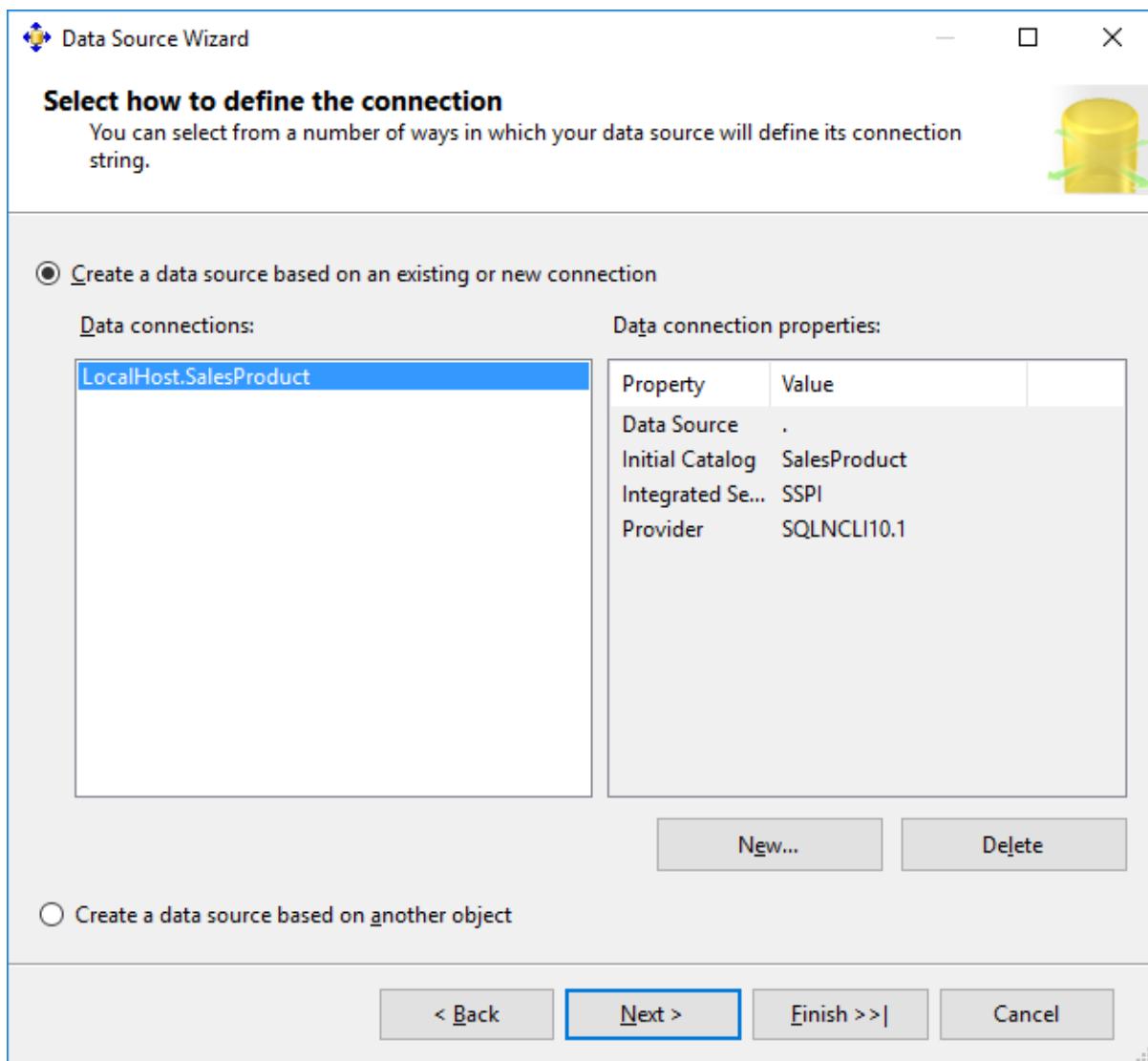
3. Choose Provider as “Microsoft OLEDB Provider for SQL Server”, Server Name as “.”, Select database name as “SalesProduct”.(Created in SQL Server Management studio).



4. Click on Test Connection.



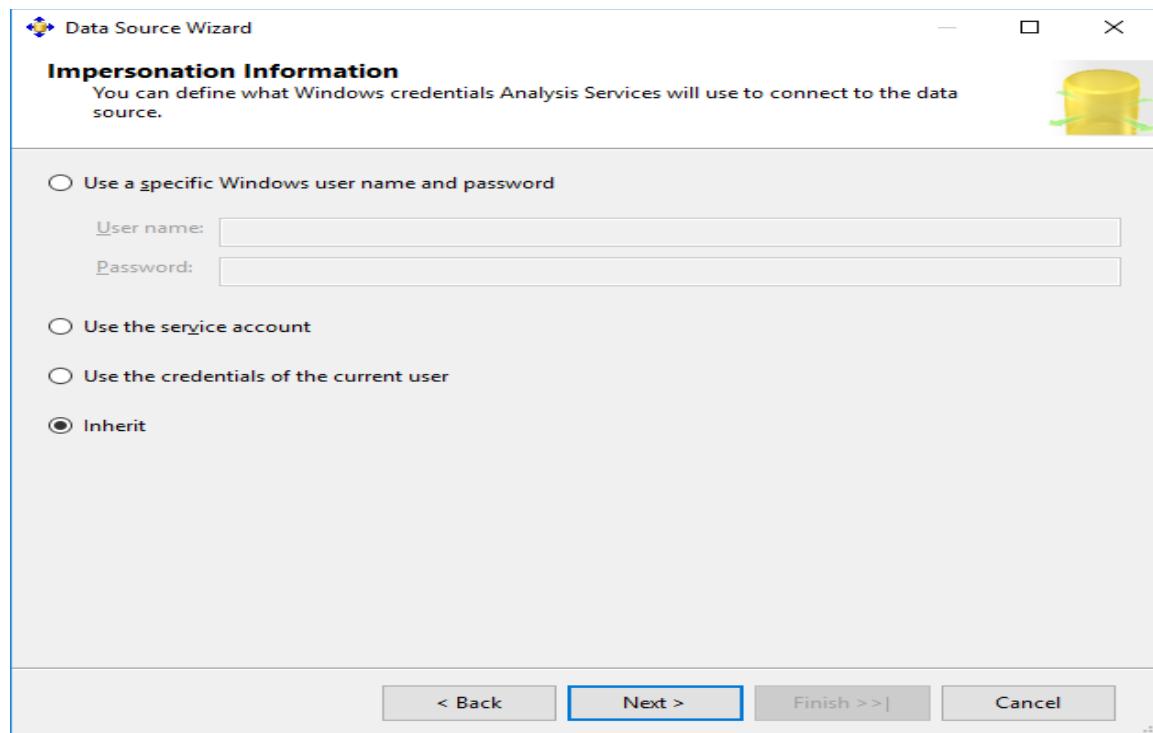
Click on Next



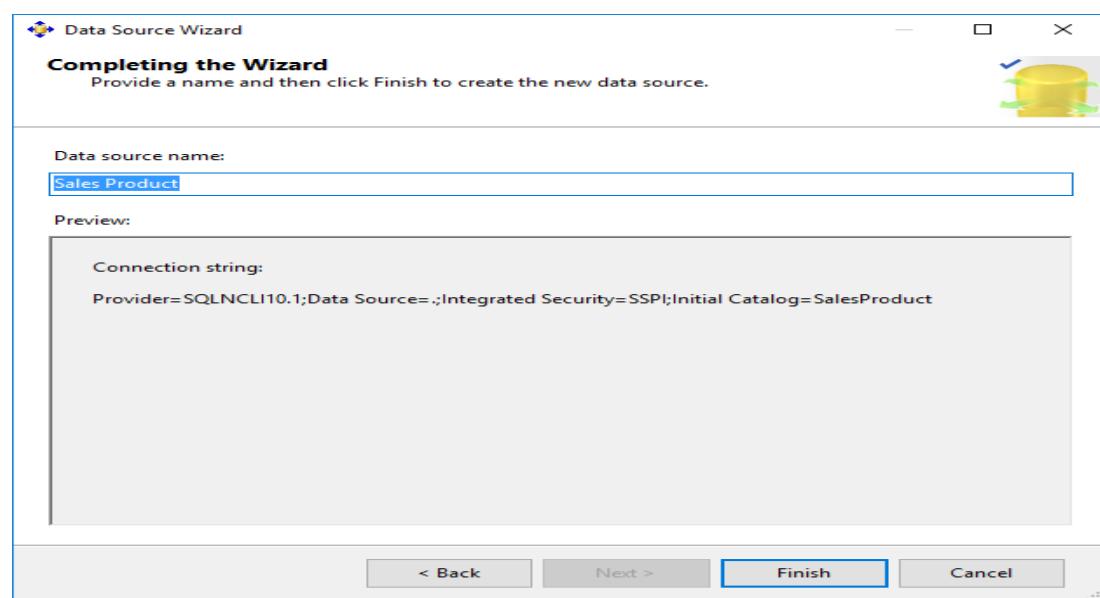
Click on Next

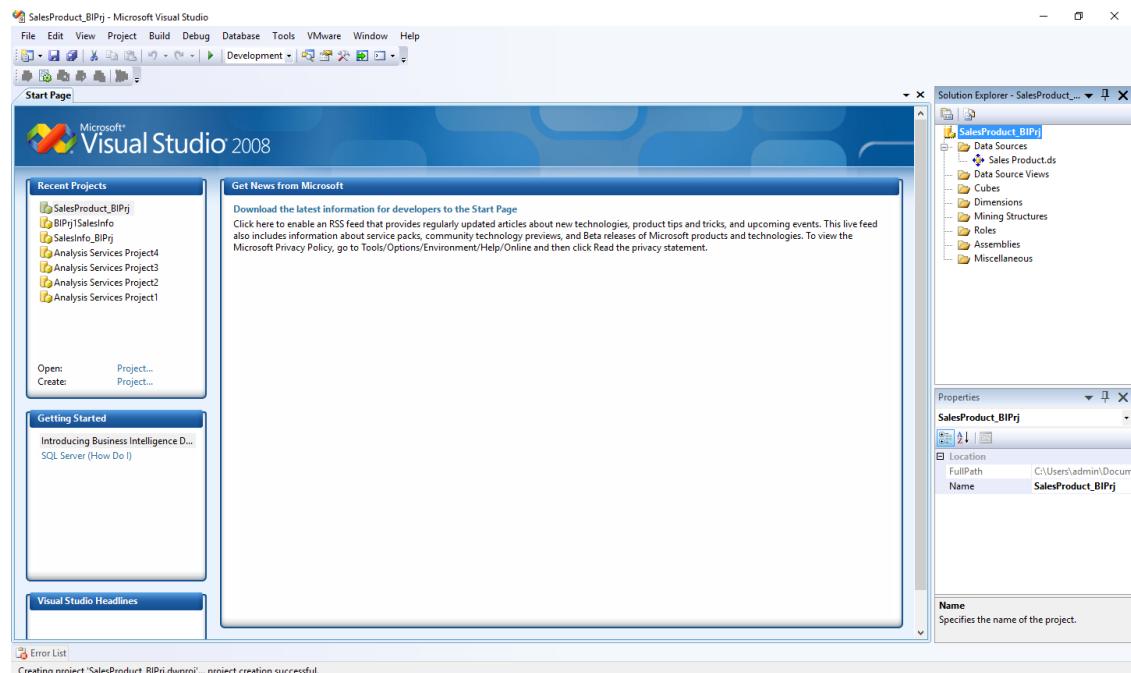
5. Choose “Inherit” Option .

Click on Next.



6. Click On Finish .



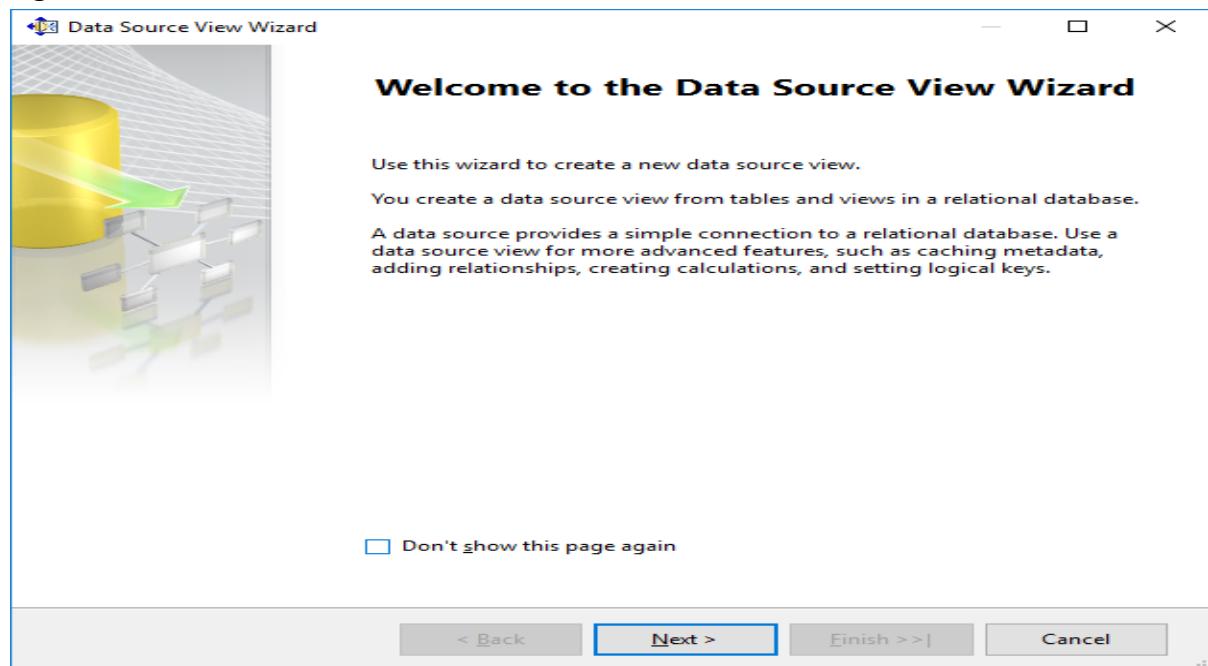
Name Data Source as “Sales Product”.

Practical No 3

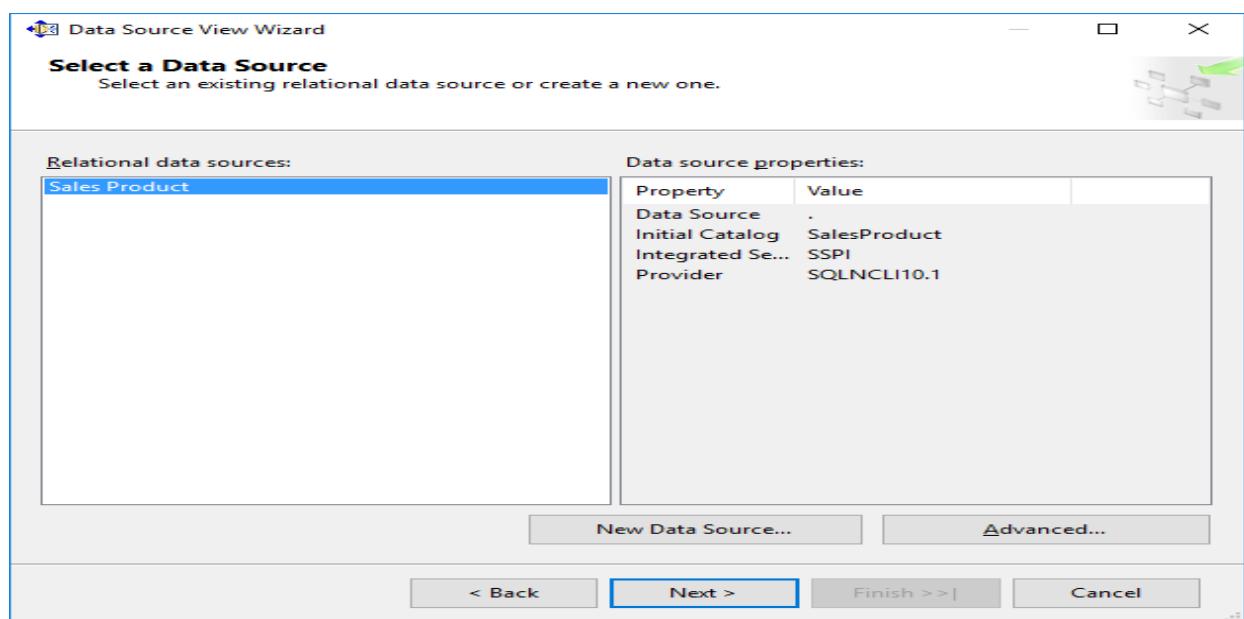
Aim : Create Data Source View using SSAS(SQL Server Analysis Services.)

Solution :

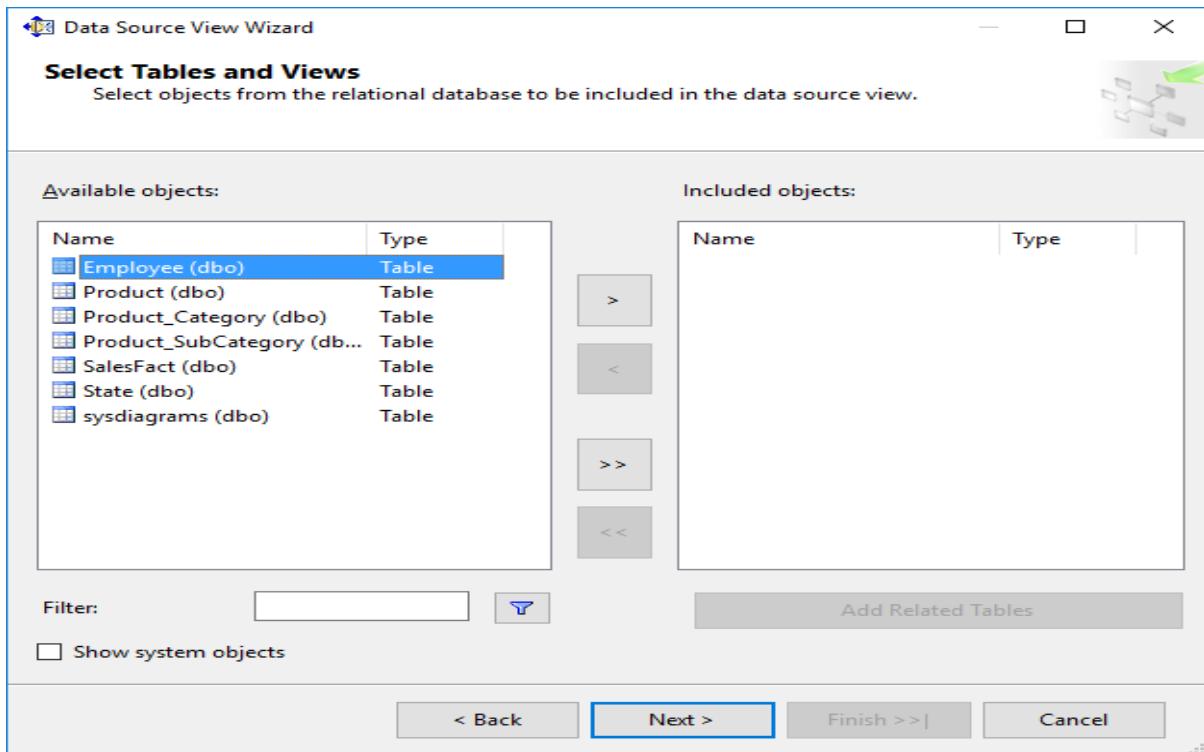
1. Right click on Data Source View -> New Data Source View



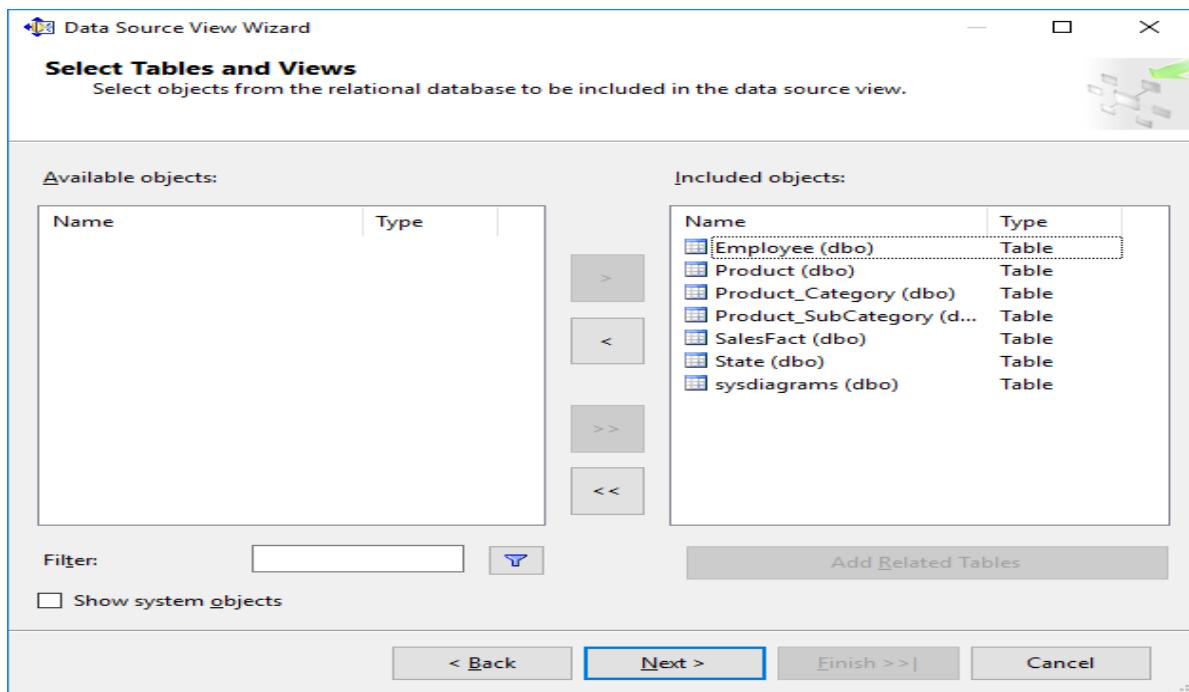
2. Click on Next.

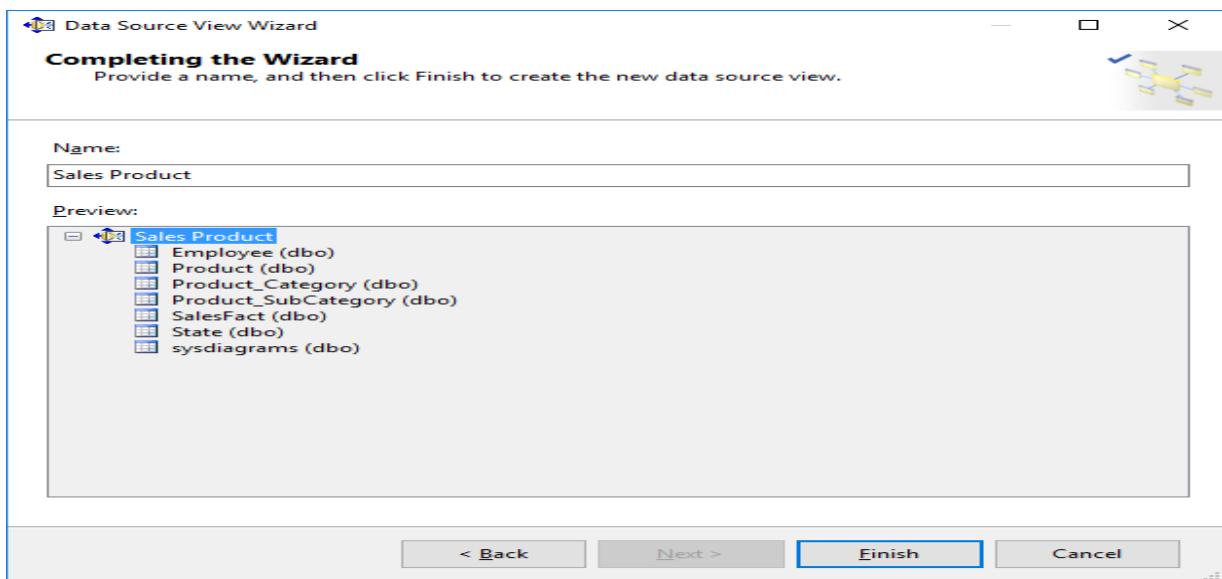


3. Select Tables and Views.



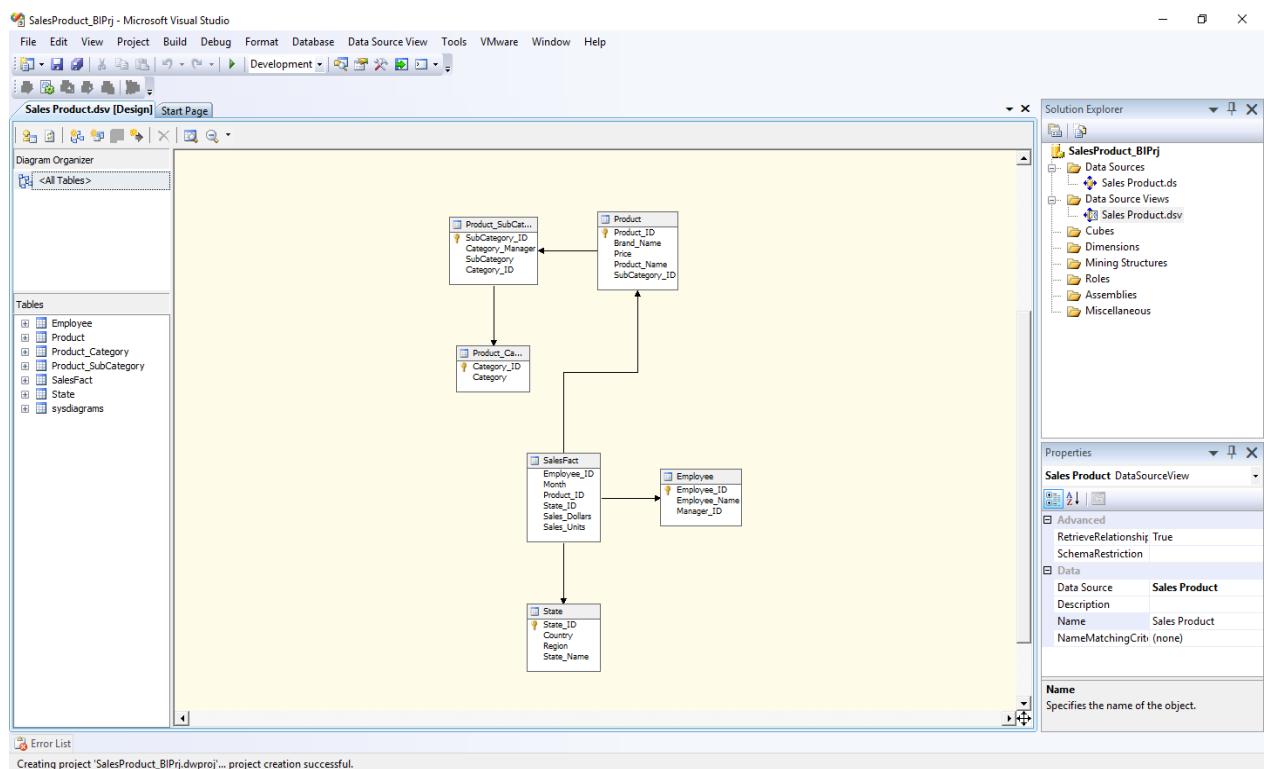
Click on next





Click on Finish.

4. Finally, we will get the Data Source View like :

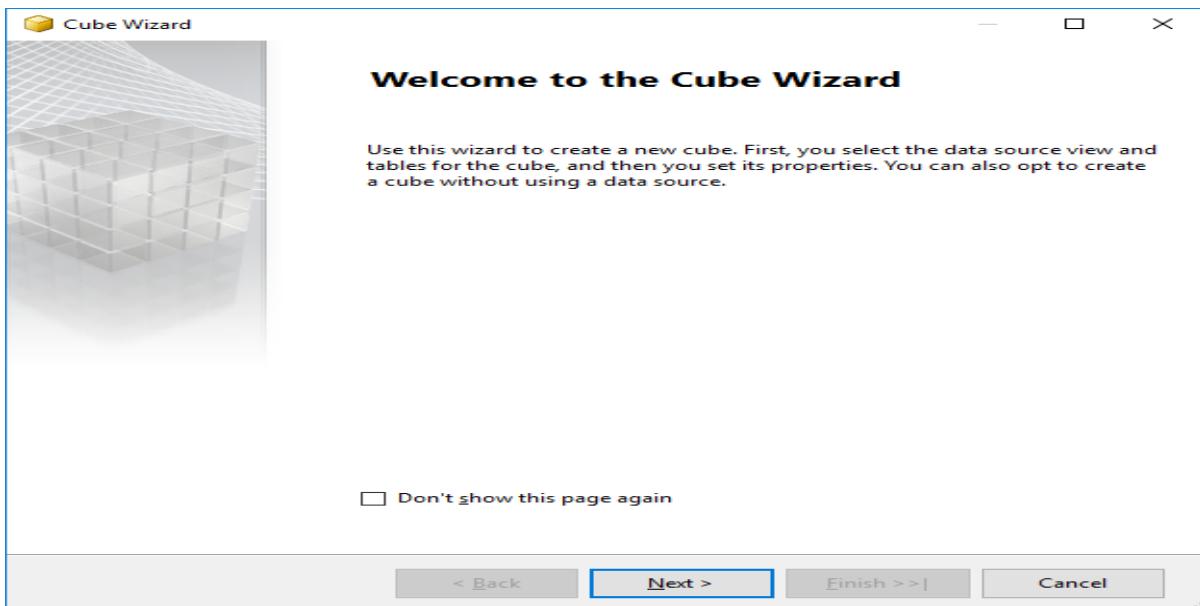


Practical No 4

Aim : Create cube using SSAS(SQL Server Analysis Services.) and process the cube.

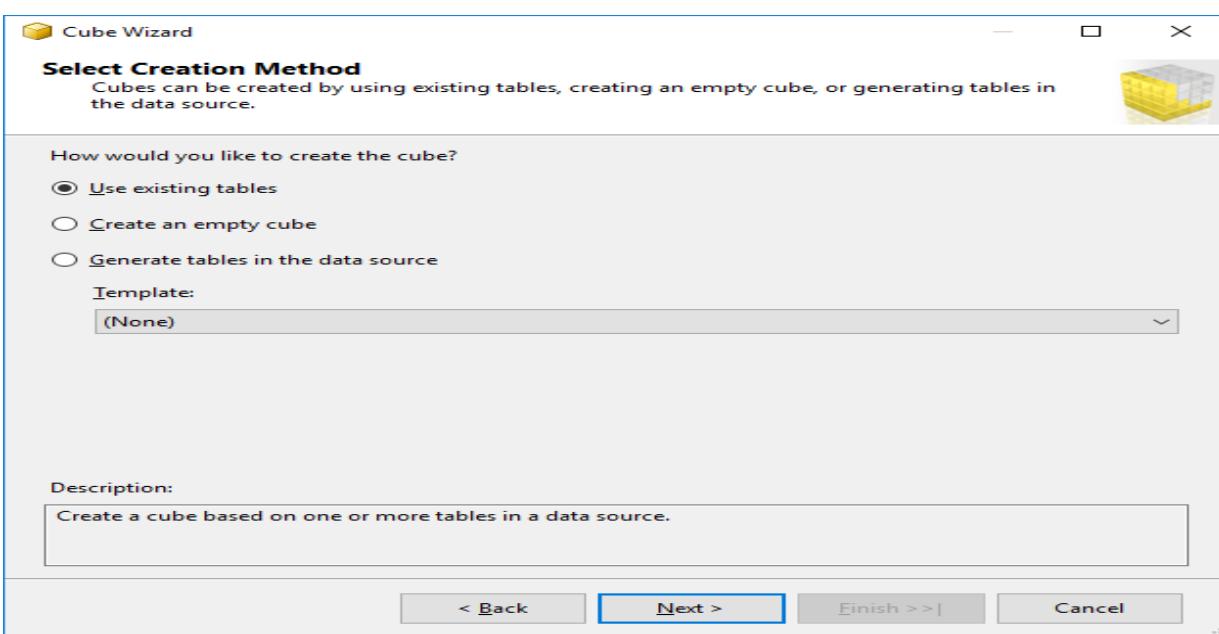
Solution :

1. Right click on Cubes -> New Cube.

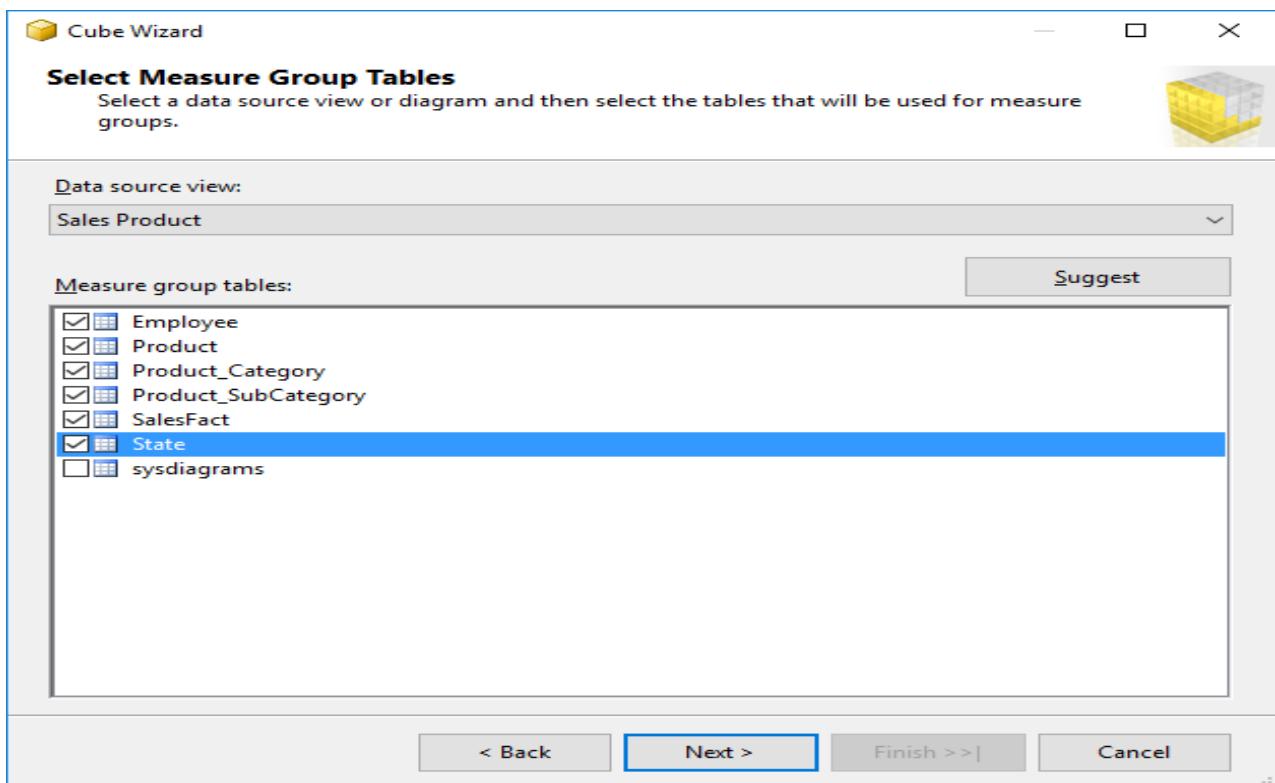


Click on Next.

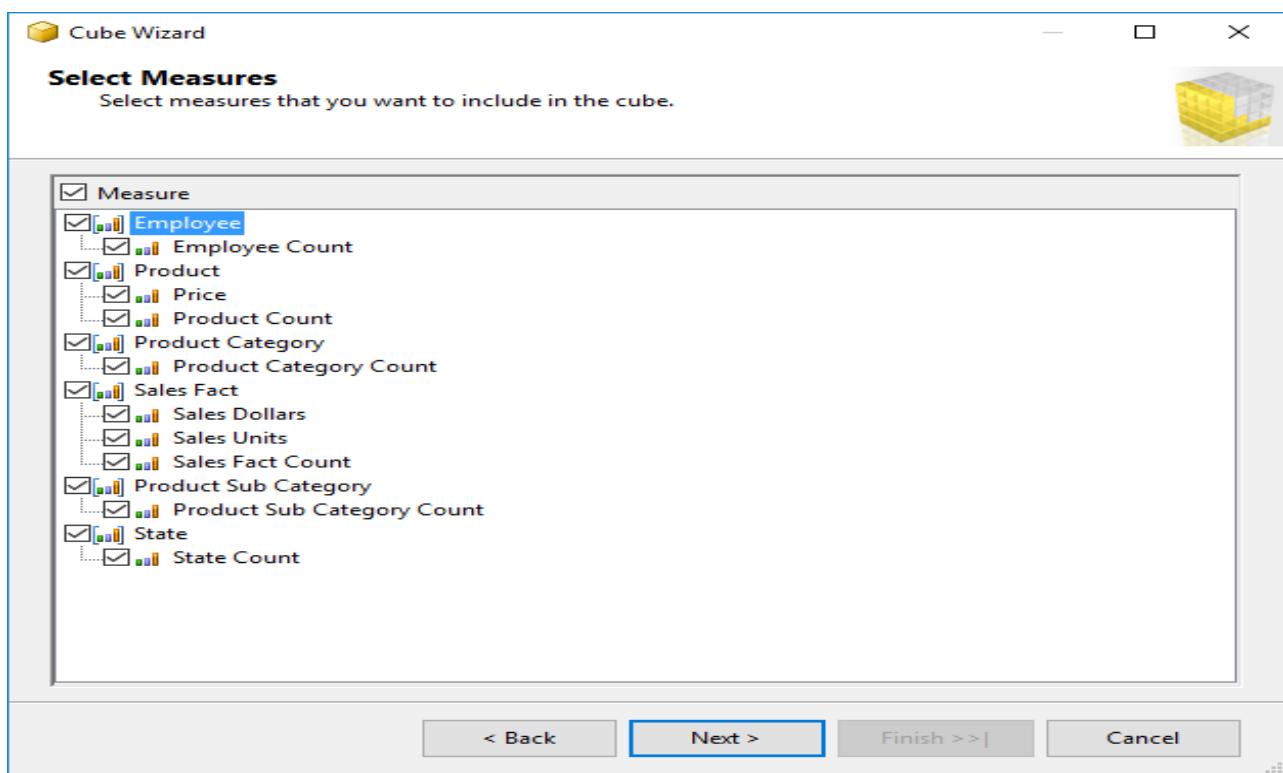
2. Select First option “Use existing tables”. Click on Next.



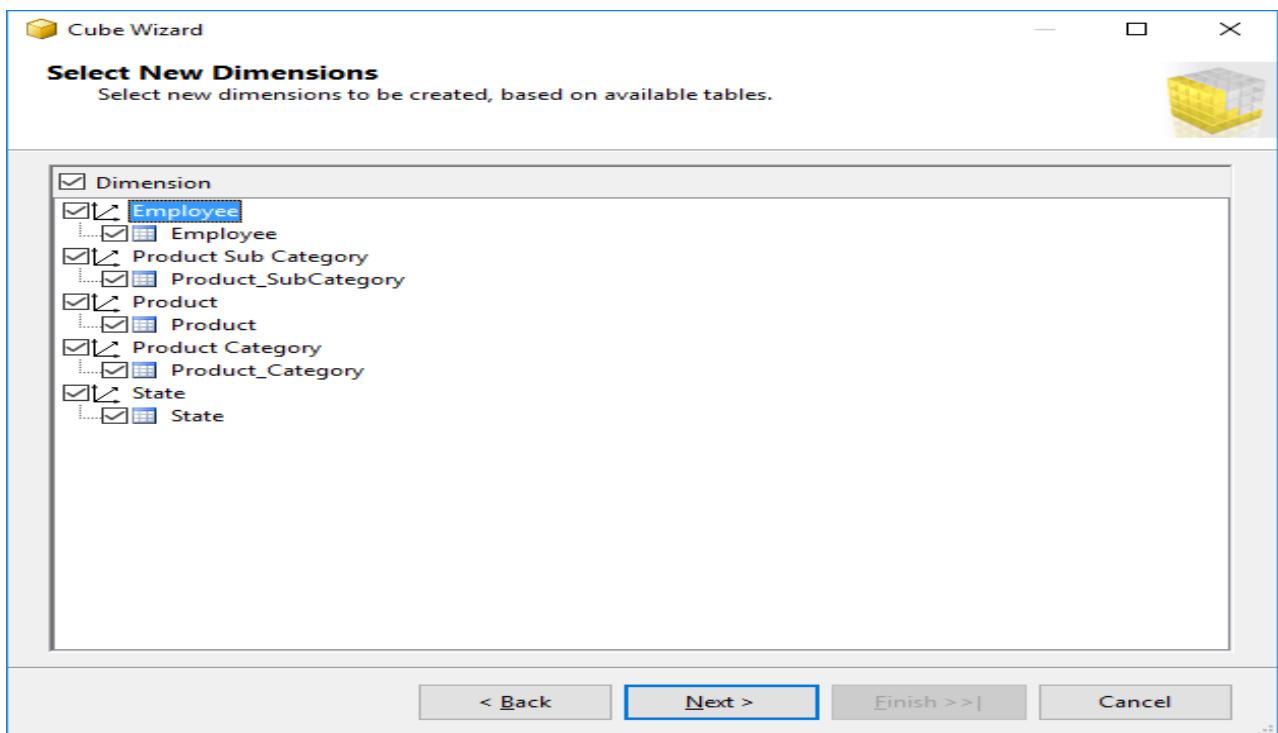
3. Select Data Source View as “Sales Product” and Select all the tables.



Click on Next.

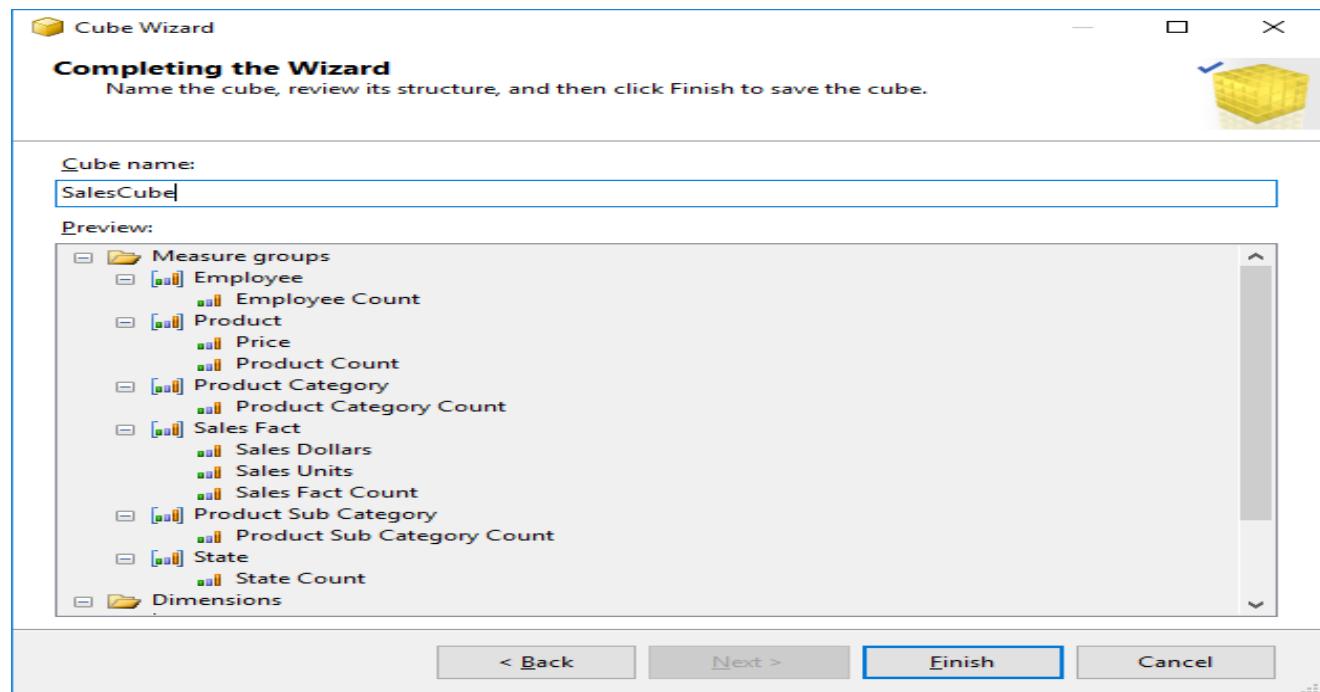


Click on Next.



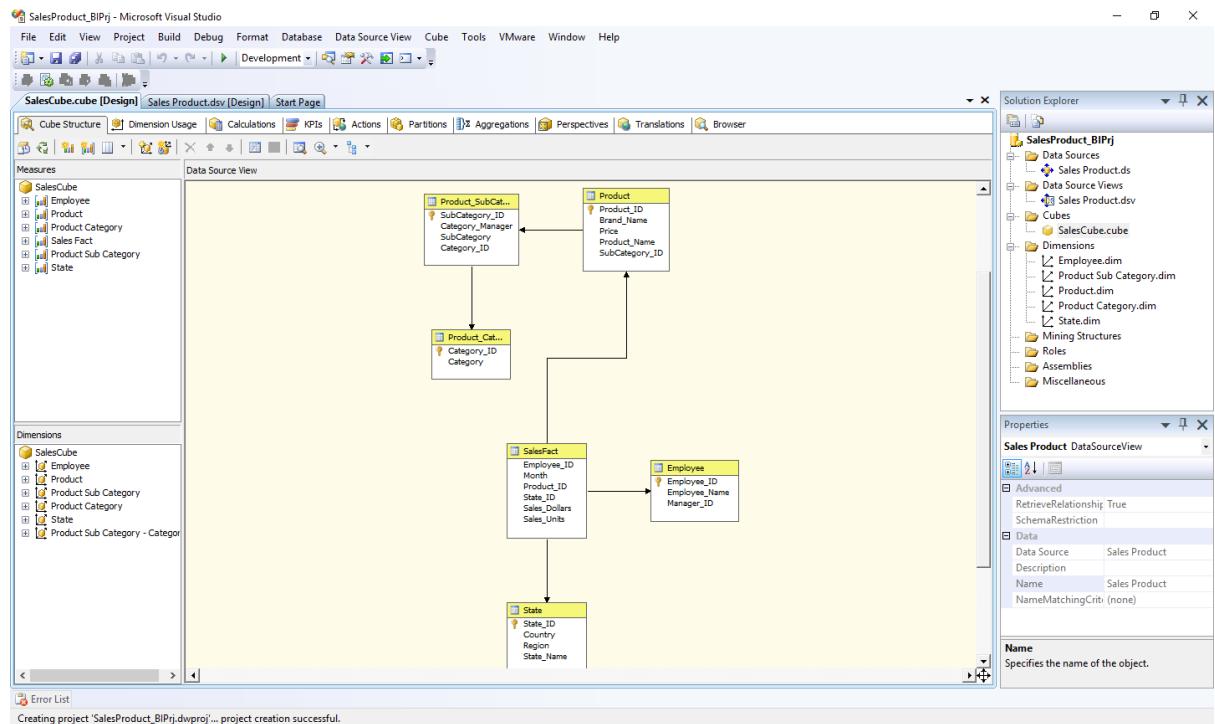
Click on Next.

4. Name Cube as “SalesCube”.

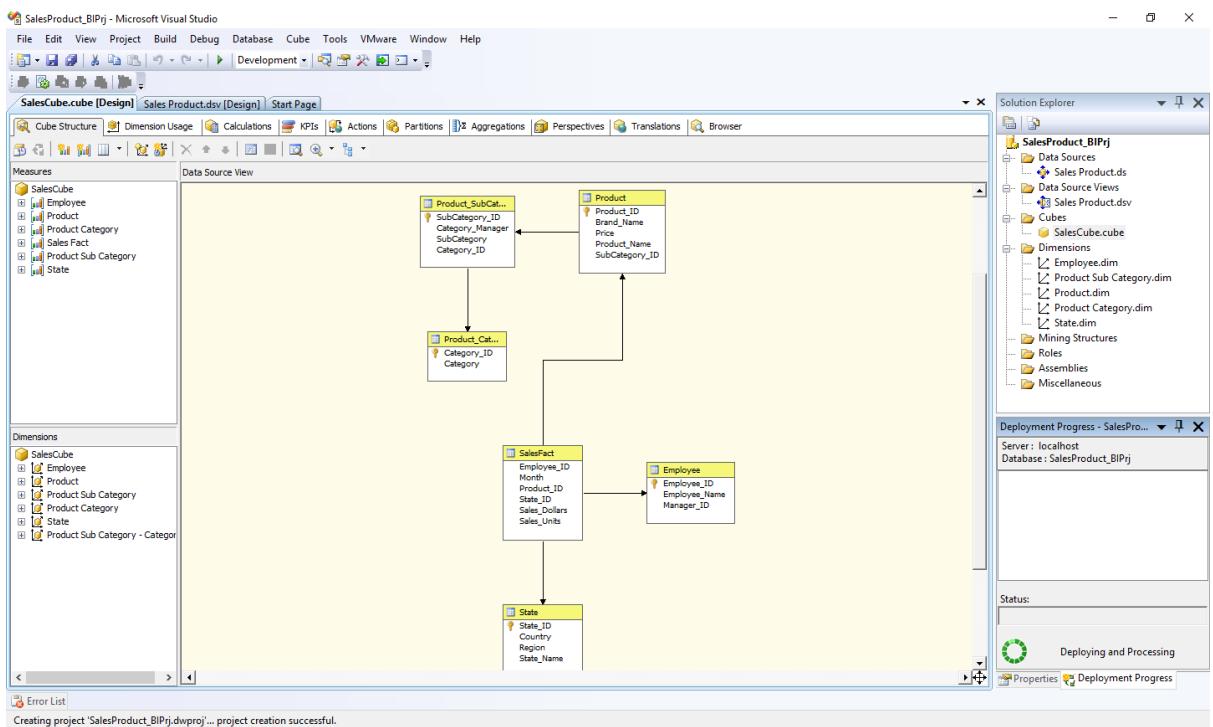


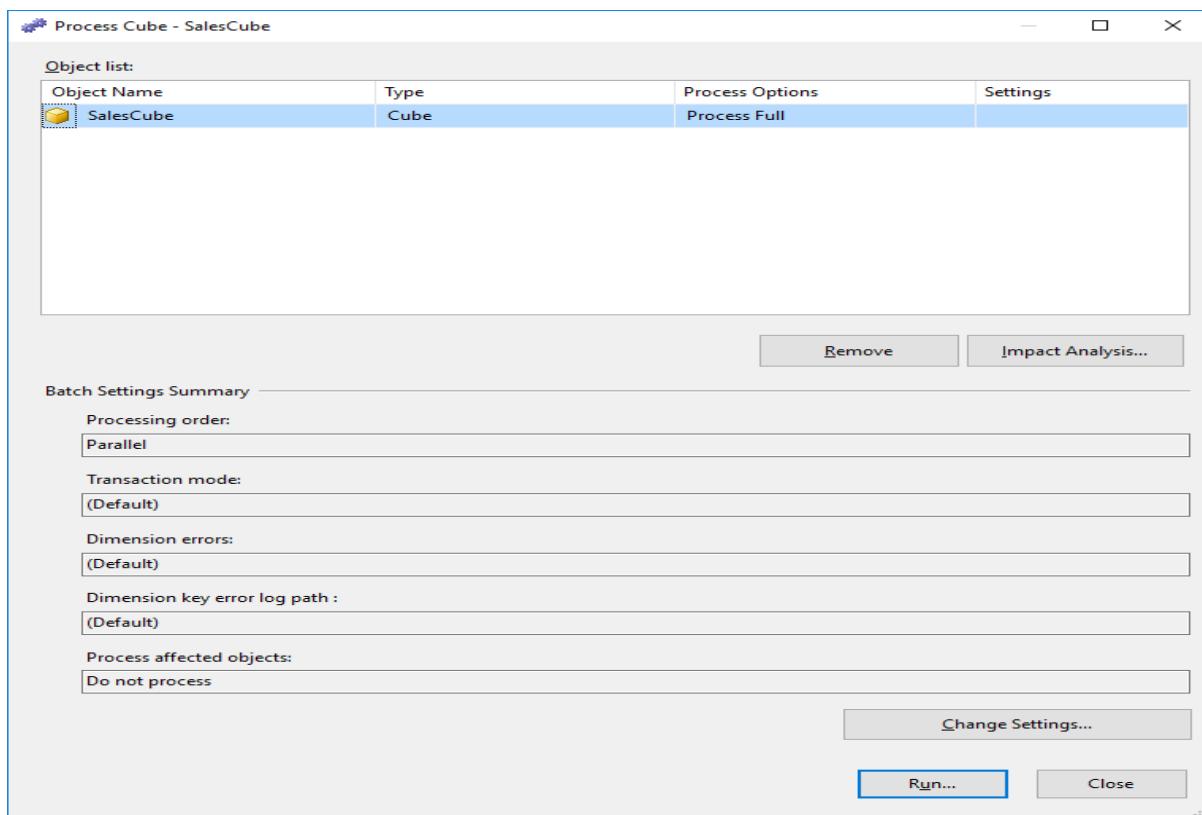
Click on Finish.

5. Finally, we will get the Cube View as well Dimensions View like :

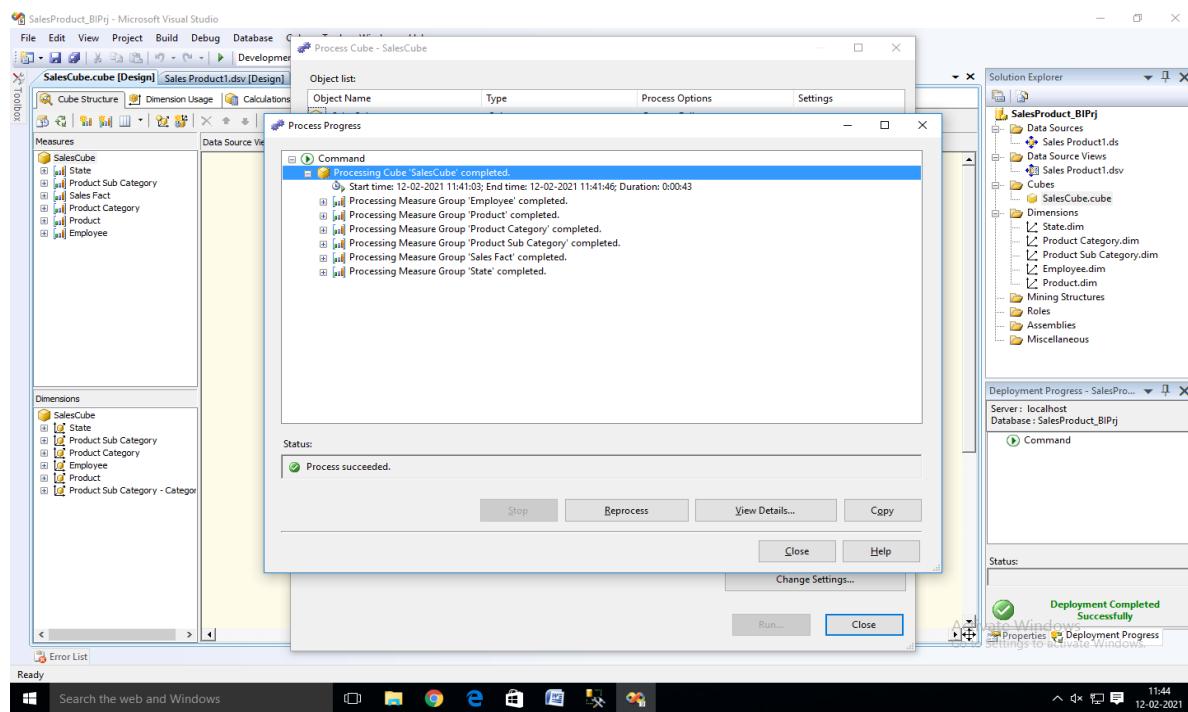


6. Finally, Process cube by Right click on SalesCube -> Process .





7. Click On Run

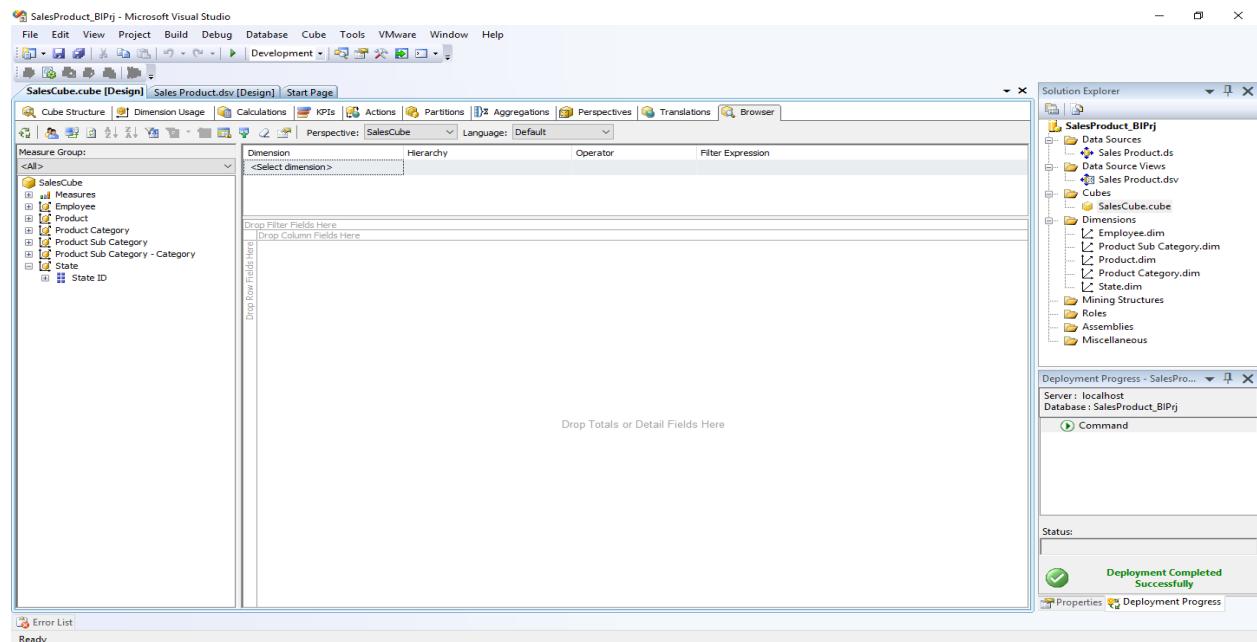


Practical No 5

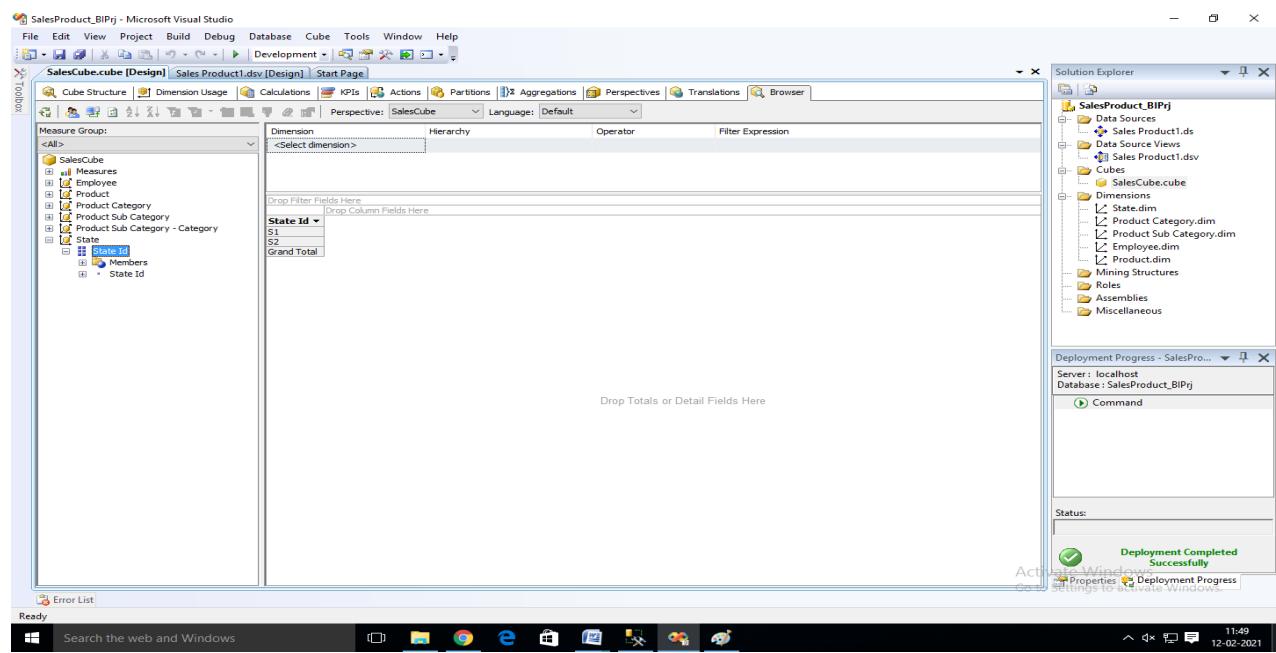
Aim : View cube data in multidimensional Format.

Solution :

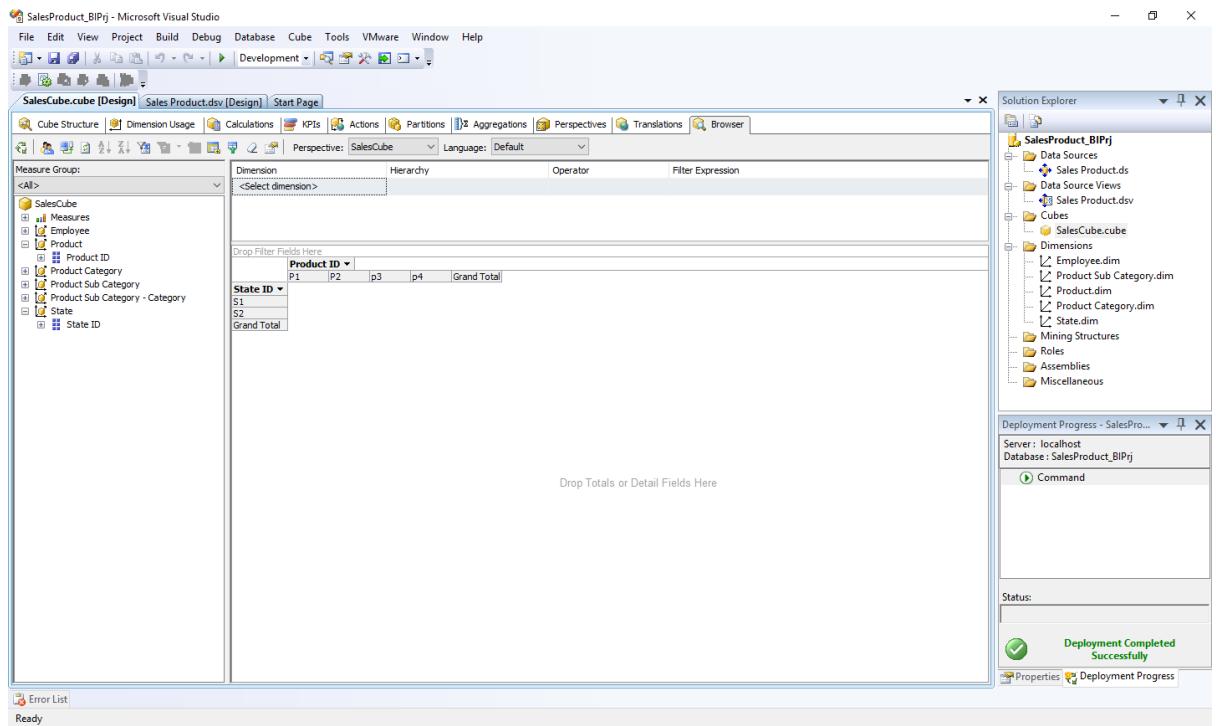
1. Double Click on “SalesCube”. Go to the “Browser” Tab.



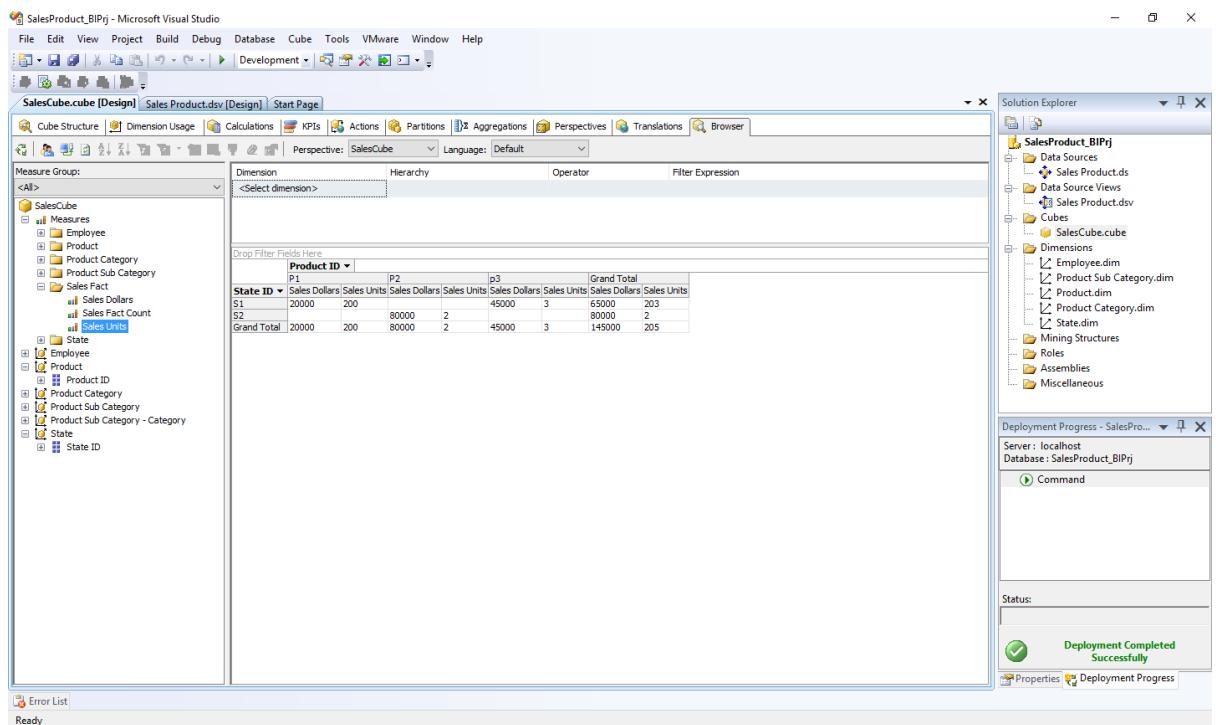
2. Go to the “State Dimension”. Right Click on ‘State_ID’ -> Add to Row Area.



3. Go to the “Product Dimension”. Right Click on ‘Product_ID’ -> Add to Column Area.



4. Go to ‘Measures’. Select ‘SalesFact’ -> Right Click on “Sales Dollars” as well “Sales Units” -> Add to Data area.



5. Go to the “Employee Dimension”. Right Click on ‘Employee_ID’ -> Add to Row Area.

The screenshot shows the Microsoft Visual Studio interface for a BI project named "SalesProduct_BIPrj". The main window displays the "SalesCube.cube [Design]" tab. On the left, the "Measure Group" pane lists various measures like Sales Dollars, Sales Units, and Sales Fact Count. The "Dimensions" pane shows the Employee dimension expanded, with "Employee ID" selected. A data grid in the center shows sales data for employees E1, E2, and E3 across three products (P1, P2, P3) with their respective sales dollars and units. The right side of the screen includes the "Solution Explorer" showing the project structure, and a "Deployment Progress" status bar indicating "Deployment Completed Successfully".

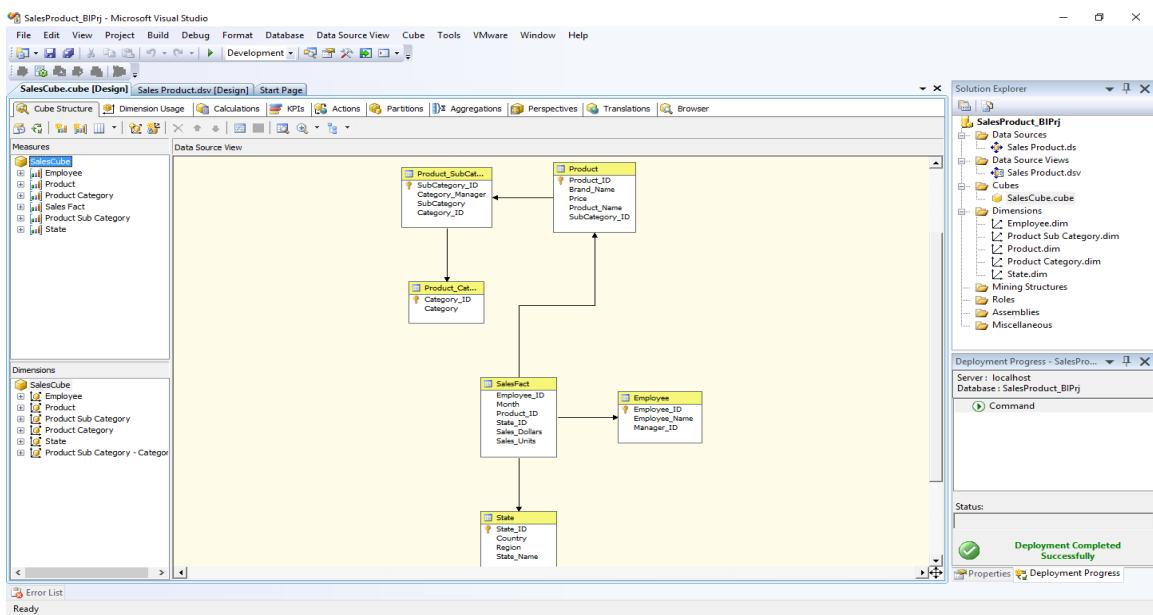
State ID	Employee ID	P1	P2	p3	Grand Total
S1	E1	20000	200		20000 200
	E3			45000 3	45000 3
	Total	20000	200	45000 3	65000 203
S2	E2			80000 2	80000 2
	Total			80000 2	80000 2
	Grand Total	20000	200	80000 2	145000 205

Practical No 6

Aim : Working with measures in the cube.

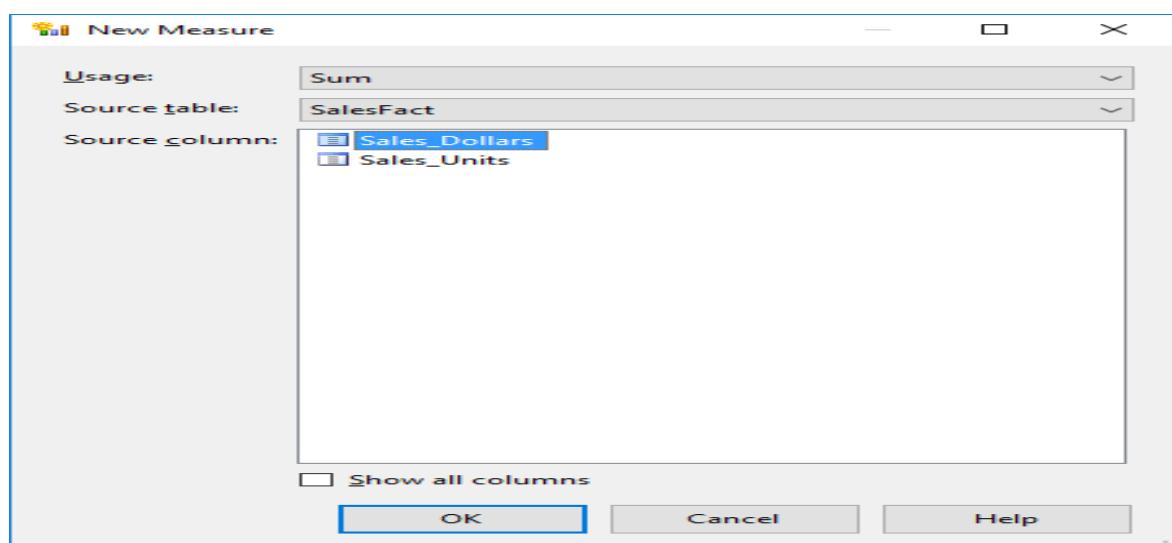
Solution :

1. Double click on 'SalesCube'. Go to cube structure.



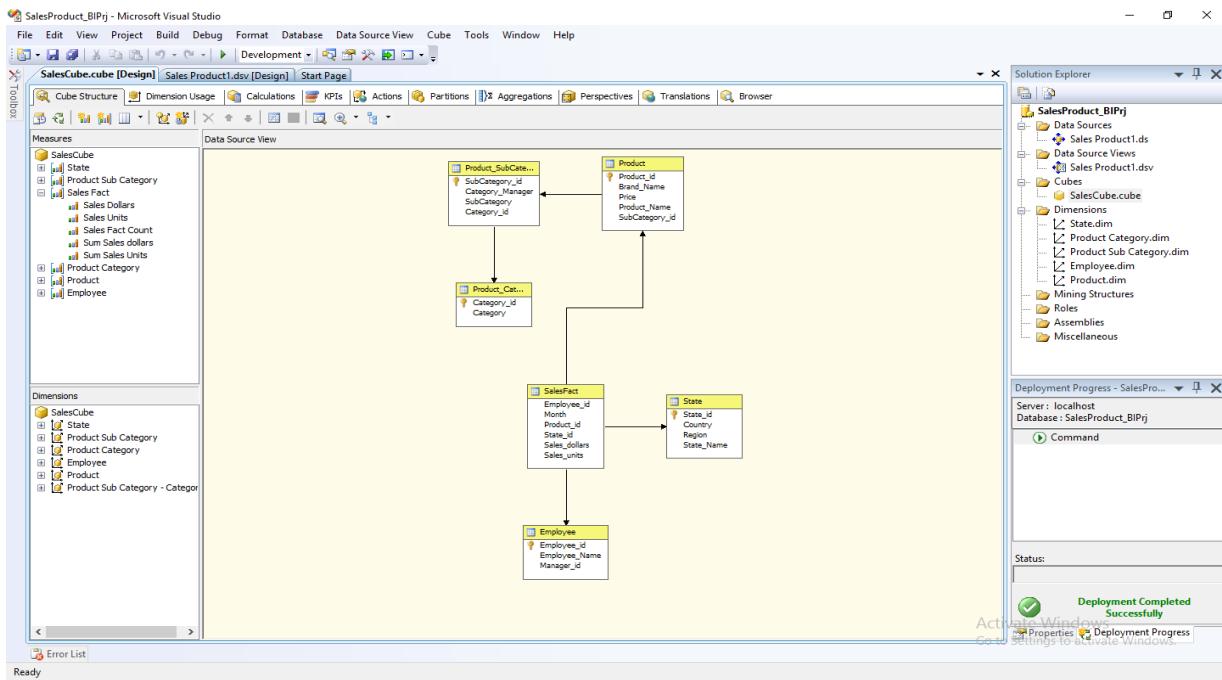
2. Right click on SalesCube -> New Measure.

Select Usage = "Sum" , Source table = "SalesFact" and Source Column = "Sales_Dollars".

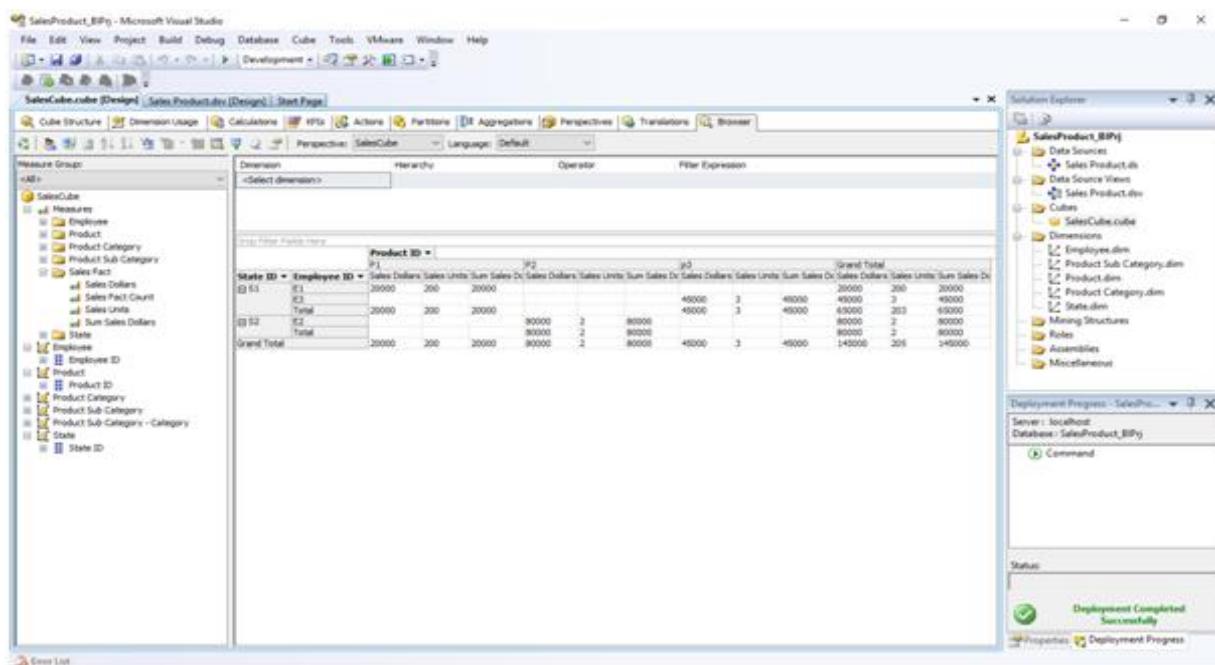


Click on OK.

3. Rename Measure as “Sum Sales Dollars”.



4. Process Cube and Go to Browser and Reconnect it. Right Click on “Sum Sales Dollars” -> Add to Data Area

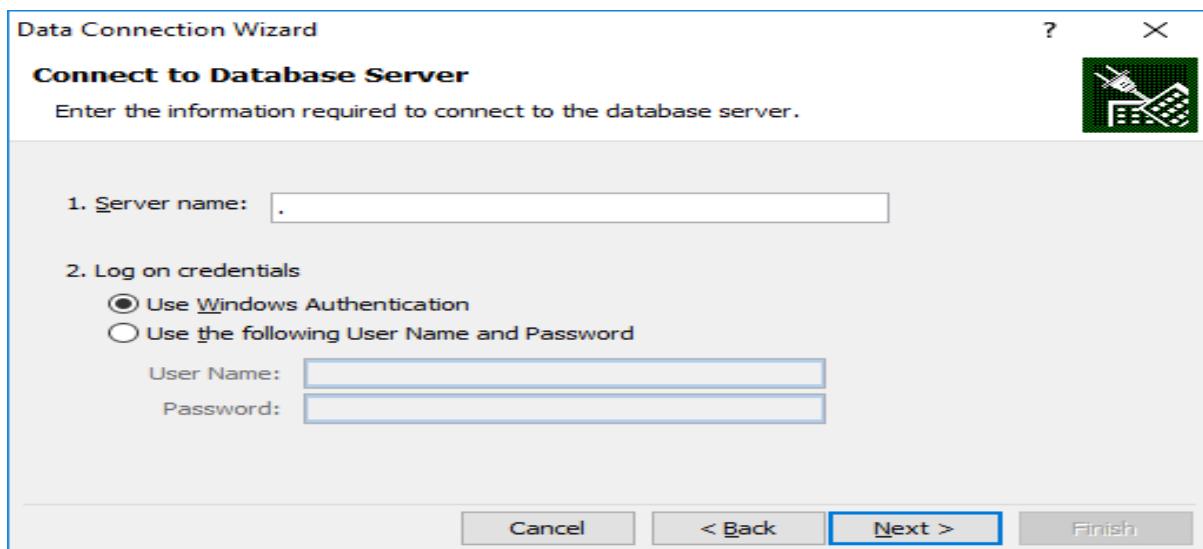


Practical No 7

Aim : Creating an Excel Pivot Table and Pivot Chart by using the OLAP cube data.

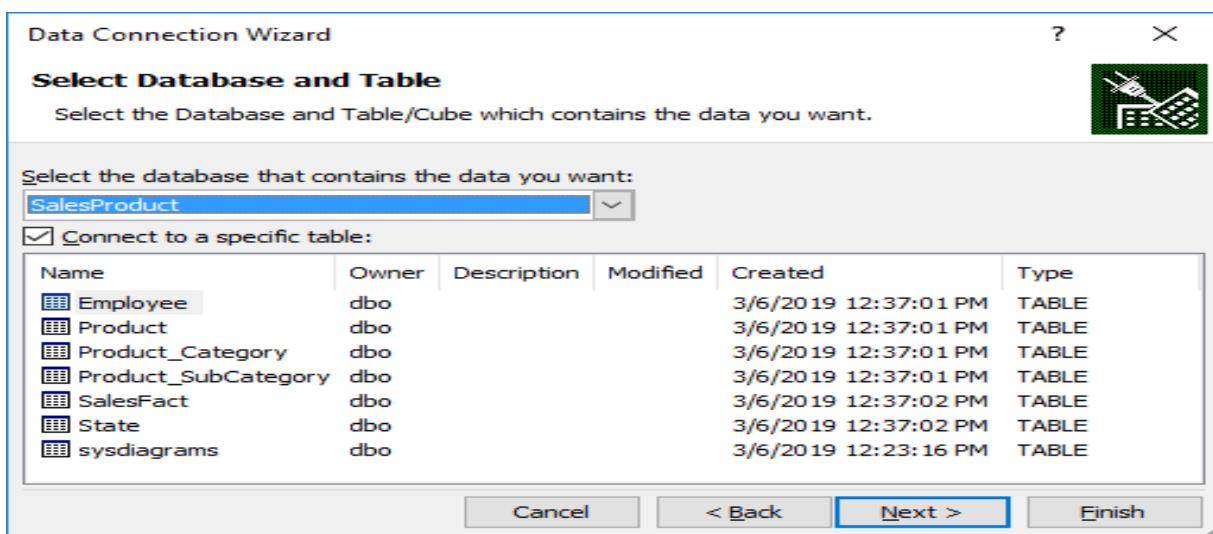
Solution :

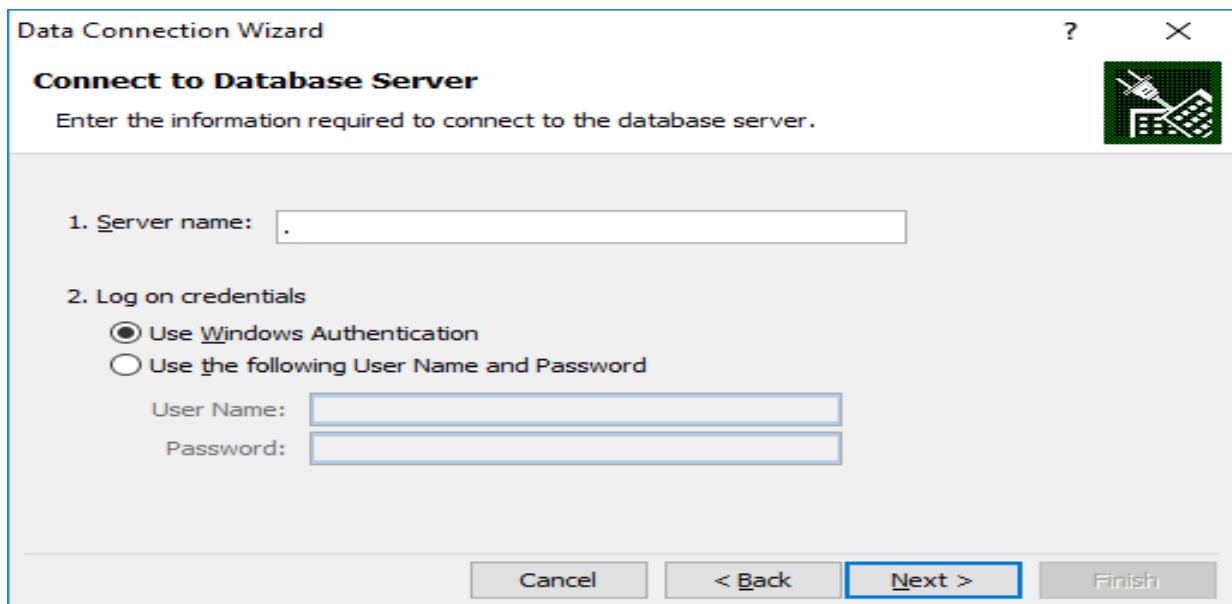
1. Open MS-Excel. Click on Data Menu.
2. Go to From Other Sources.
- 2.1. From SQL Server -> Type Server name as “.”



Click on Next.

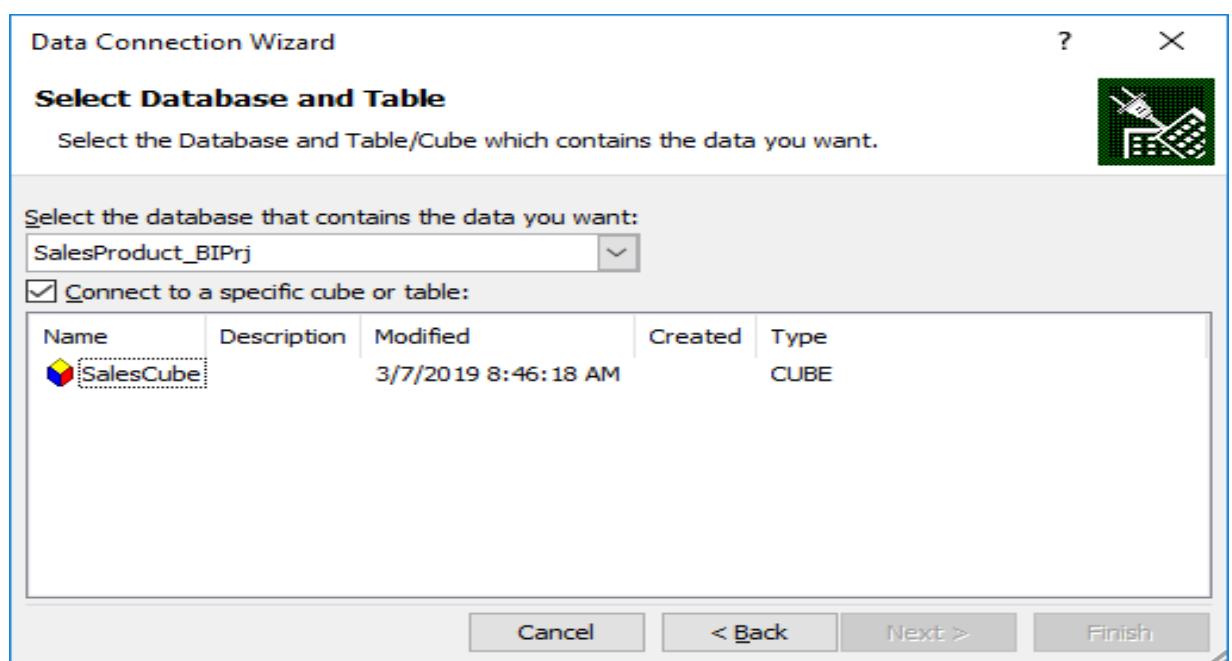
Choose SQL Database -> “SalesProduct”



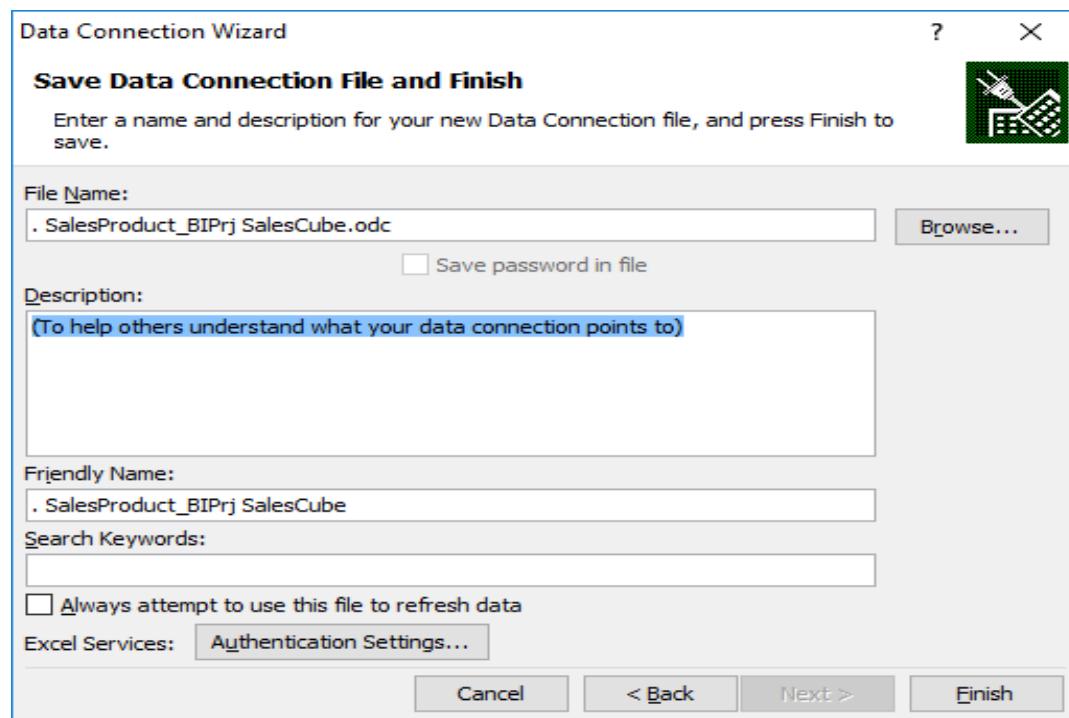
2.2. From Analysis Services -> Type Server name as “.”

Click on Next.

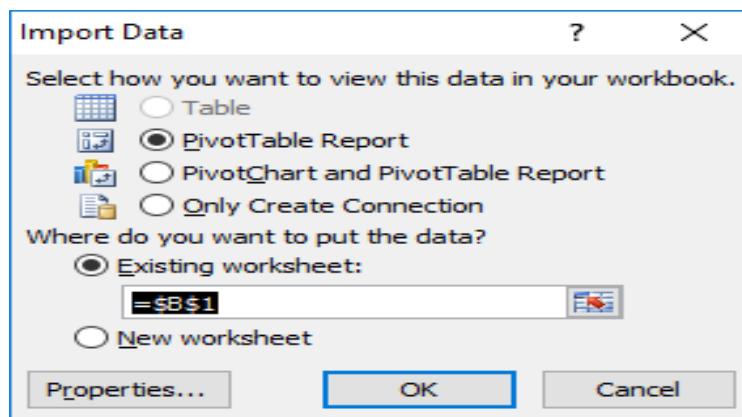
Choose Analysis Database as “SalesProduct_BIPrj”. Click on Next.



Click on OK.

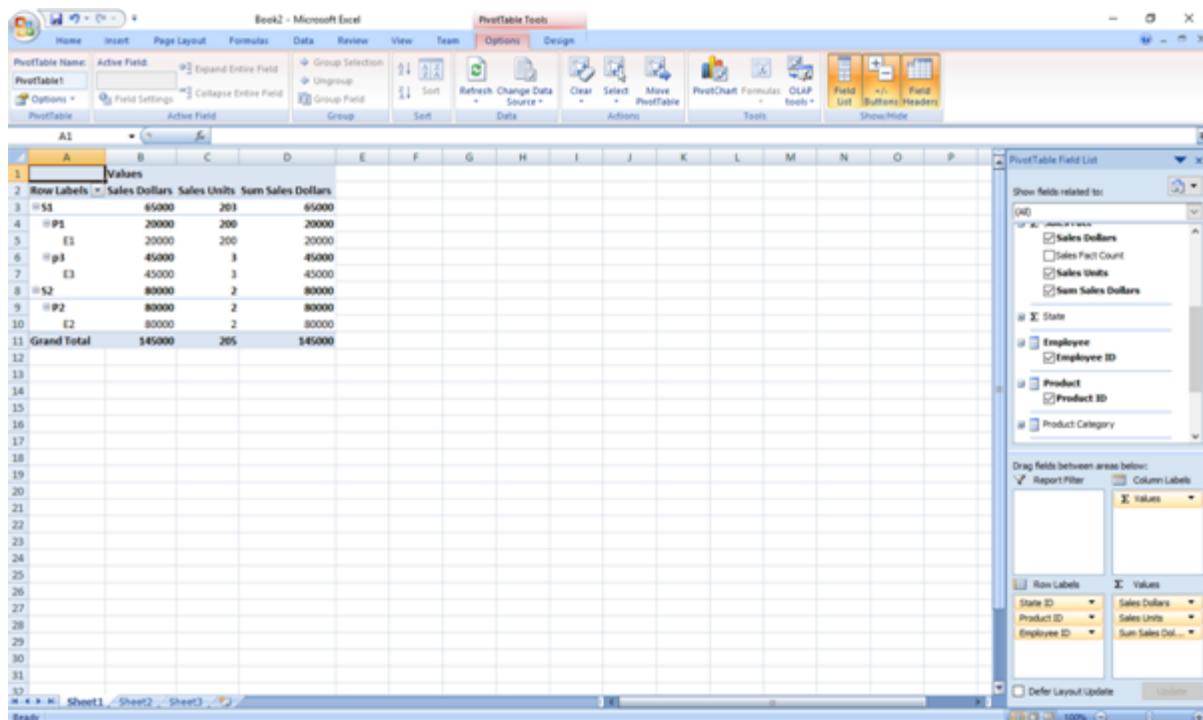


Click on Finish.

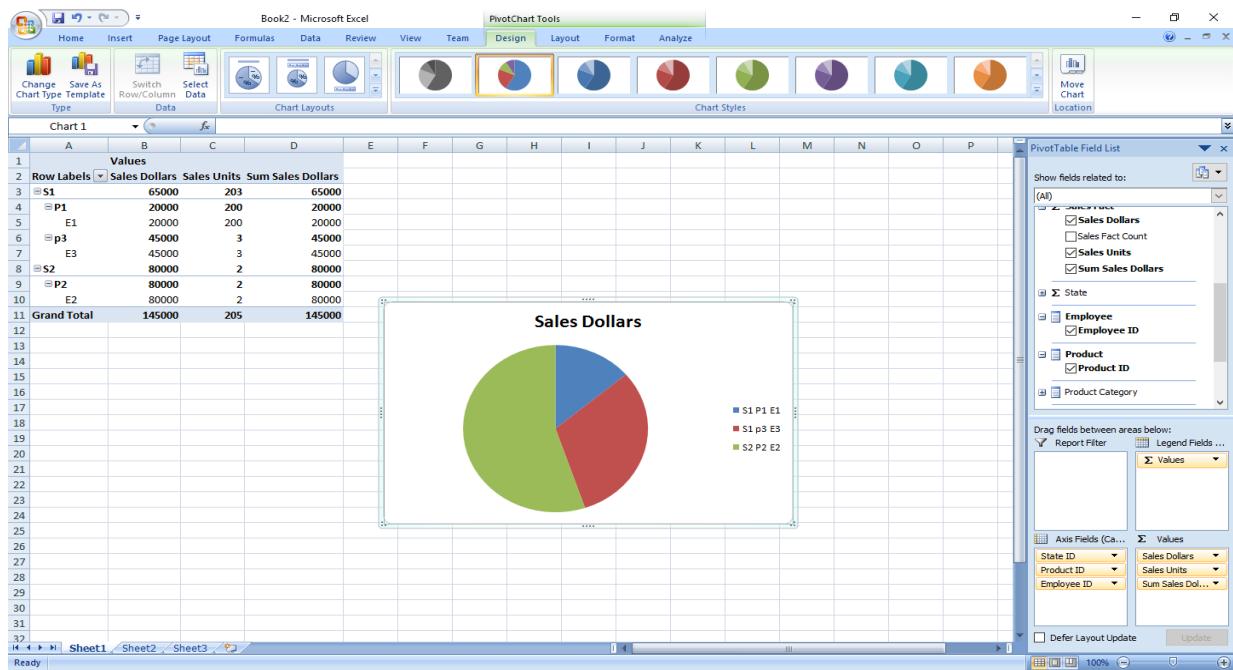


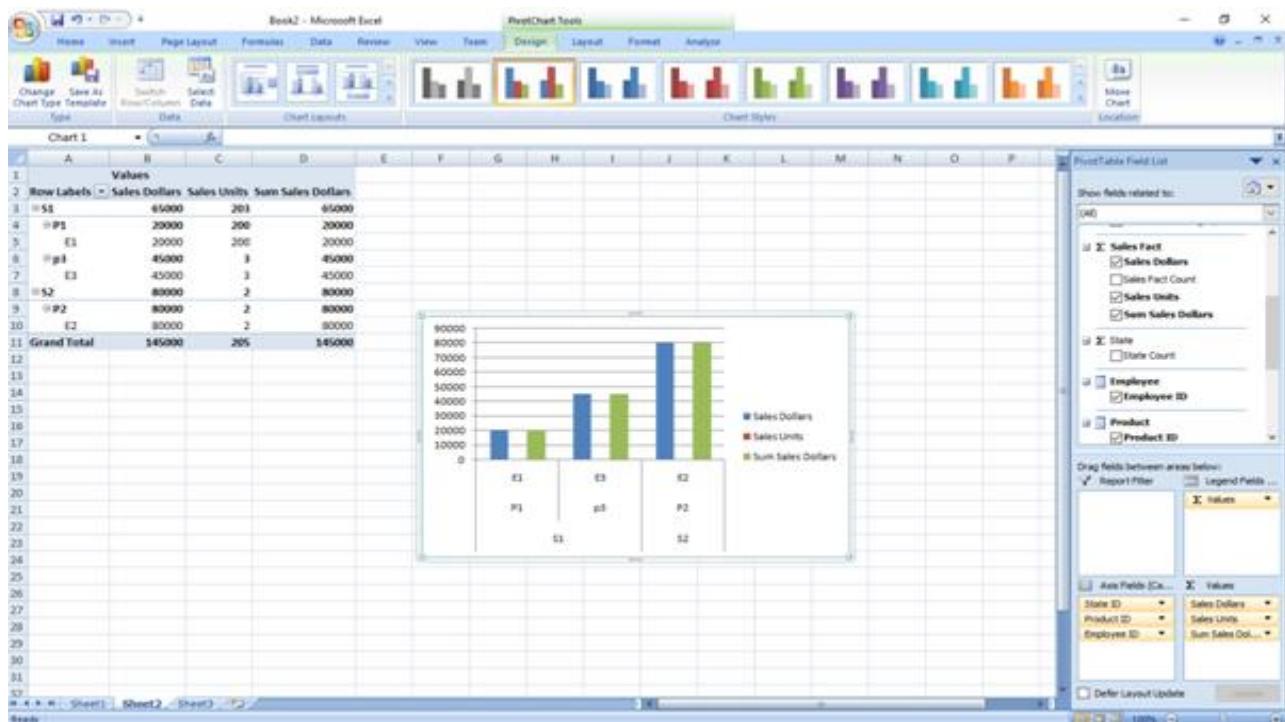
Click on OK.

3. Select State_ID, Product_ID,Employee_ID and Measures as Sales_Dollars, Sales_Units and Sum Sales Dollars.



4. Select Result Area. Go to Insert Menu. Select Pie Chart option



5. Select Result Area. Go to Insert Menu. Select Column option

Practical No 8

Aim : Firing Queries on Tables.

Solution :

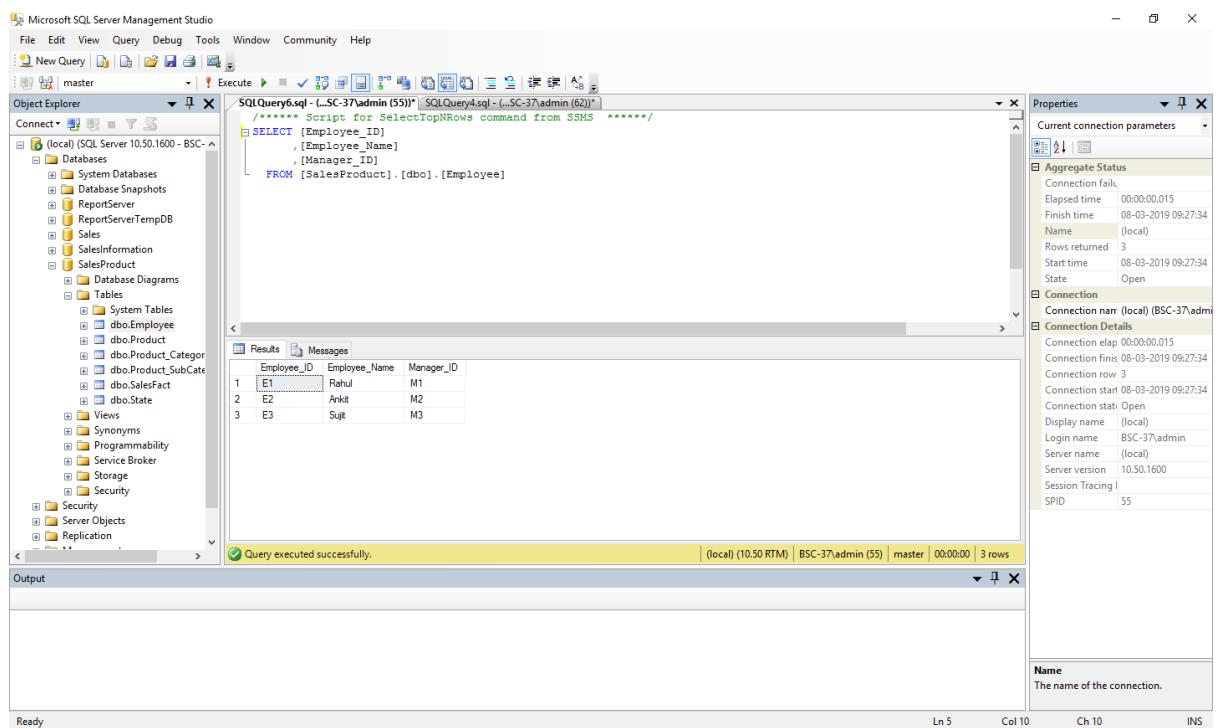
Open Application -> Microsoft SQL Server 2008 R2 -> SQL Server Management Studio

1. Select Connect Tab -> Database Engine -> Select Server Name(local)

2. Expand ‘Database’ -> Expand ‘SalesProduct’ -> Expand Tables.

3. Fire following queries :

3.1. `SELECT [Employee_ID], [Employee_Name] , [Manager_ID]
FROM [SalesProduct].[dbo].[Employee]`



The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows a connection to '(local) (SQL Server 10.50.1600 - BSC-37\...'. The 'Tables' node under 'SalesProduct' is expanded, showing 'Employee'. The 'Properties' window on the right shows connection details like 'Connection name (local)', 'Display name (local)', and 'Login name BSC-37\admin'. The central pane displays the results of the following query:

```
/*
***** Script for SelectTopNRows command from SSMS *****/
SELECT [Employee_ID]
      ,[Employee_Name]
      ,[Manager_ID]
 FROM [SalesProduct].[dbo].[Employee]
```

The results grid shows three rows of data:

	Employee_ID	Employee_Name	Manager_ID
1	E1	Rahul	M1
2	E2	Ankit	M2
3	E3	Sujit	M3

The status bar at the bottom indicates 'Query executed successfully.' and shows the session details: '(local) (10.50 RTM) | BSC-37\admin (55) | master | 00:00:00 | 3 rows'.

3.2 `SELECT [Product_ID], [Brand_Name], [Price], [Product_Name]
, [SubCategory_ID] FROM [SalesProduct].[dbo].[Product]`

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the SalesProduct database and its tables like Employee, Product, and State. The central pane displays the results of the following query:

```
SELECT [Product_ID]
      ,[Brand_Name]
      ,[Price]
      ,[Product_Name]
      ,[SubCategory_ID]
  FROM [SalesProduct].[dbo].[Product]
```

The results grid shows four rows of data:

Product_ID	Brand_Name	Price	Product_Name	SubCategory_ID
P1	LG	200	Keyboard	SC1
P2	Samsung	40000	LED	SC2
p3	LG	15000	Fridge	SC2
p4	Whirlpool	2000	WM	SC2

The status bar at the bottom indicates "Query executed successfully." and provides session details: (local) (10.50 RTM) | BSC-37\admin (58) | master | 00:00:00 | 4 rows.

3.3. `SELECT [State_ID], [Country], [Region], [State_Name]
FROM [SalesProduct].[dbo].[State]`

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the SalesProduct database and its tables like Employee, Product, and State. The central pane displays the results of the following query:

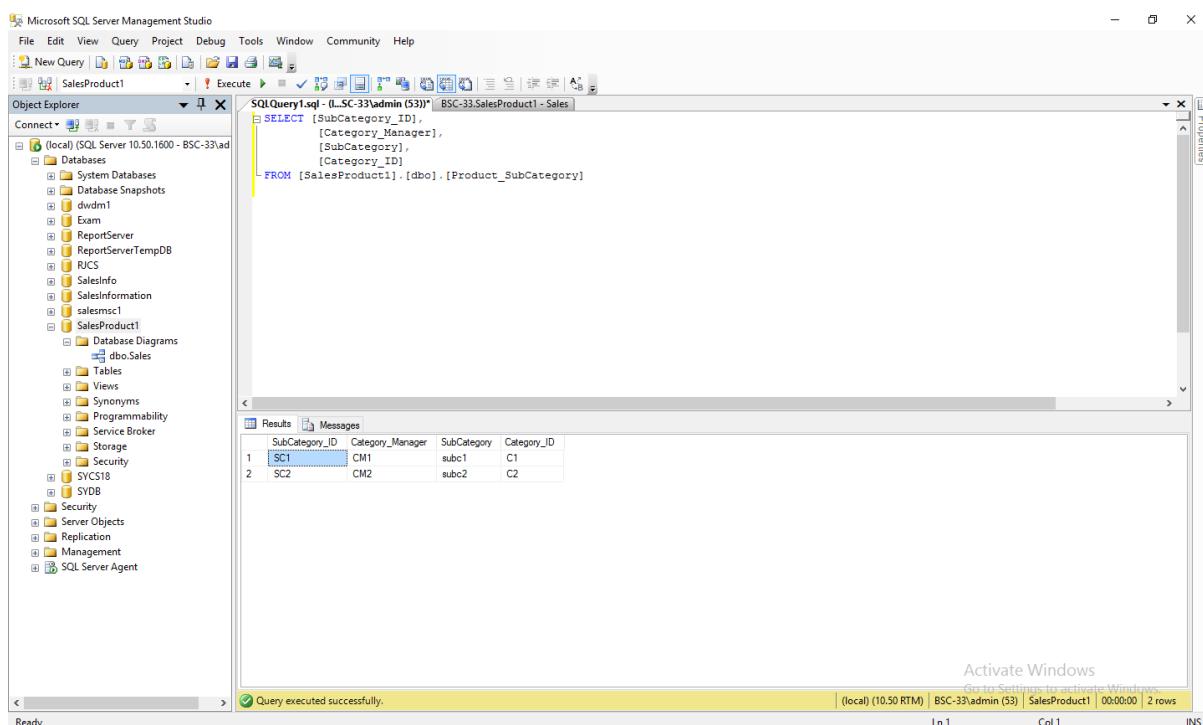
```
SELECT [State_ID]
      ,[Country]
      ,[Region]
      ,[State_Name]
  FROM [SalesProduct].[dbo].[State]
```

The results grid shows two rows of data:

State_ID	Country	Region	State_Name
S1	India	Mumba	Maha
S2	India	GN	Gujrat

The status bar at the bottom indicates "Query executed successfully." and provides session details: (local) (10.50 RTM) | BSC-37\admin (53) | master | 00:00:00 | 2 rows.

3.4. `SELECT [SubCategory_ID], [Category_Manager], [SubCategory], [Category_ID]
FROM [SalesProduct]. [dbo]. [Product_SubCategory]`



The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows a connection to '(local) (SQL Server 10.50.1600 - BSC-33)\admin'. The 'SalesProduct1' database is selected. The 'Tables' node under SalesProduct1 contains 'Sales', 'SubCategory', and 'Category'. The 'SubCategory' table is selected in the 'Object Explorer' tree. A query window titled 'SQLQuery1.sql' is open, displaying the following SQL code:

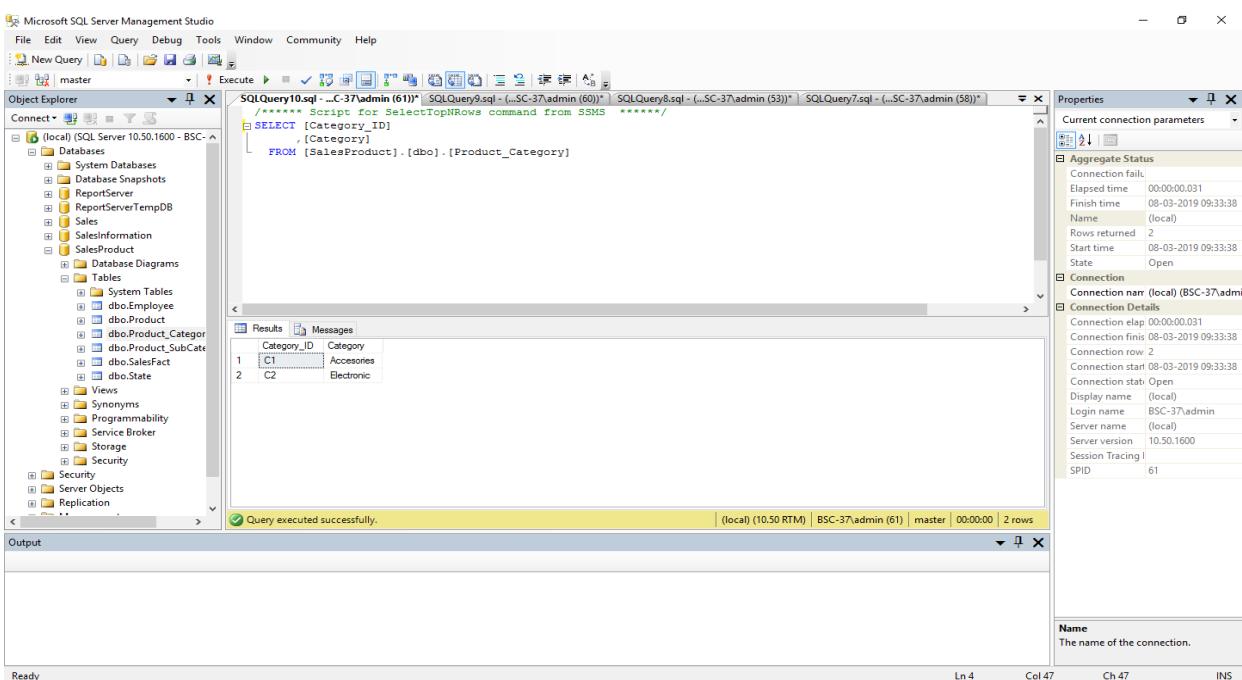
```
SELECT [SubCategory_ID],
       [Category_Manager],
       [SubCategory],
       [Category_ID]
  FROM [SalesProduct1].[dbo].[Product_SubCategory]
```

The results pane shows the following data:

SubCategory_ID	Category_Manager	SubCategory	Category_ID
SC1	CM1	subc1	C1
SC2	CM2	subc2	C2

The status bar at the bottom indicates 'Query executed successfully.' and shows the session details: '(local) (10.50 RTM) | BSC-33\admin (53) | SalesProduct1 | 00:00:00 | 2 rows'.

3.5. `SELECT [Category_ID], [Category]
FROM [SalesProduct]. [dbo]. [Product_Category]`



The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows a connection to '(local) (SQL Server 10.50.1600 - BSC-37)\admin'. The 'master' database is selected. The 'Tables' node under SalesProduct contains 'Category', 'Product', and 'Product_Cat'. The 'Category' table is selected in the 'Object Explorer' tree. A query window titled 'SQLQuery10.sql' is open, displaying the following SQL code:

```
SELECT [Category_ID],
       [Category]
  FROM [SalesProduct].[dbo].[Product_Category]
```

The results pane shows the following data:

Category_ID	Category
C1	Accessories
C2	Electronic

A 'Properties' window is open on the right side, showing connection details such as 'Connection name: (local)', 'Connection elap: 00:00:00.031', and 'Session Tracing I: SPID 61'.

3.6. `SELECT [Employee_ID], [Month], [Product_ID], [State_ID], [Sales_Dollars],
[Sales_Units] FROM [SalesProduct].[dbo].[SalesFact]`

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows a database named 'SalesProduct1' containing various objects like databases, tables, and stored procedures. The central pane displays a query window with the following SQL code:

```
SELECT [Employee_ID],
       [Month],
       [Product_ID],
       [State_ID],
       [Sales_Dollars],
       [Sales_Units]
  FROM [SalesProduct1].[dbo].[SalesFact]
```

Below the query window is a results grid showing two rows of data:

Employee_ID	Month	Product_ID	State_ID	Sales_Dollars	Sales_Units
E1	2018-05-23	P1	S1	4	4000
E2	2017-06-02	P2	S2	5	1000

The status bar at the bottom indicates "Query executed successfully." and provides other session details.

3.7. `select [Employee_ID], [Sales_Dollars], [Sales_Units],
[SalesProduct].[dbo].[Product].[Product_ID], [Product_Name]
FROM [SalesProduct].[dbo].[SalesFact], [SalesProduct].[dbo].[Product]
where [SalesProduct].[dbo].[SalesFact].[Product_ID] =
[SalesProduct].[dbo].[Product].[Product_ID];`

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows a database named 'SalesProduct1' containing various objects like databases, tables, and stored procedures. The central pane displays a query window with the following SQL code:

```
select [Employee_ID], [Sales_Dollars], [Sales_Units],
       [SalesProduct1].[dbo].[Product].[Product_ID], [Product_Name]
  FROM [SalesProduct1].[dbo].[SalesFact], [SalesProduct1].[dbo].[Product]
 where [SalesProduct1].[dbo].[SalesFact].[Product_ID] =[SalesProduct1].[dbo].[Product].[Product_ID];
```

Below the query window is a results grid showing two rows of data:

Employee_ID	Sales_Dollars	Sales_Units	Product_ID	Product_Name
E1	4	4000	P1	TV
E2	5	1000	P2	Mobile

The status bar at the bottom indicates "Query executed successfully." and provides other session details.

```
3.8. select [Employee_ID], [Sales_Dollars], [Sales_Units],
       [SalesProduct].[dbo].[State].[State_ID], [State_Name]
  FROM [SalesProduct].[dbo].[SalesFact], [SalesProduct].[dbo].[State]
 where [SalesProduct].[dbo].[SalesFact].[State_ID] =
       [SalesProduct].[dbo].[State].[State_ID];
```

Employee_ID	Sales_Dollars	Sales_Units	State_ID	State_Name
E1	4	4000	S1	Maharashtra
E2	5	1000	S2	Maharashtra

```
3.9. select [SalesProduct].[dbo].[Employee].[Employee_ID], [Employee_Name],
       [Sales_Dollars], [Sales_Units]
  FROM [SalesProduct].[dbo].[Employee] , [SalesProduct].[dbo].[SalesFact]
 where [SalesProduct].[dbo].[Employee].[Employee_ID]=
       [SalesProduct].[dbo].[SalesFact].[Employee_ID] ;
```

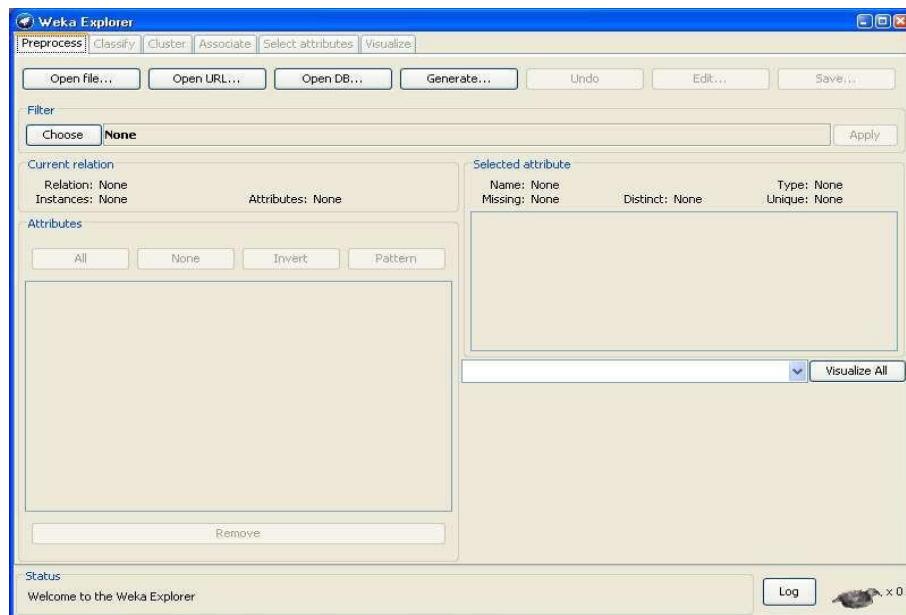
Employee_ID	Employee_Name	Sales_Dollars	Sales_Units
E1	ABC	4	4000
E2	XYZ	5	1000

Practical No - 9

Aim : Data PreProcessing

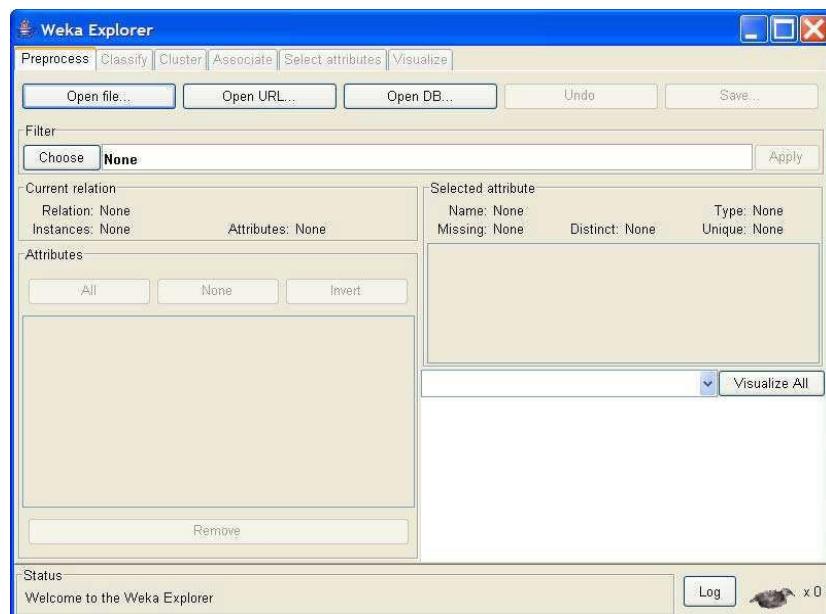
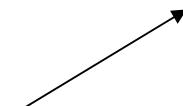
Solution :

Only the first tab, 'Preprocess', is active at the moment because there is no dataset open.

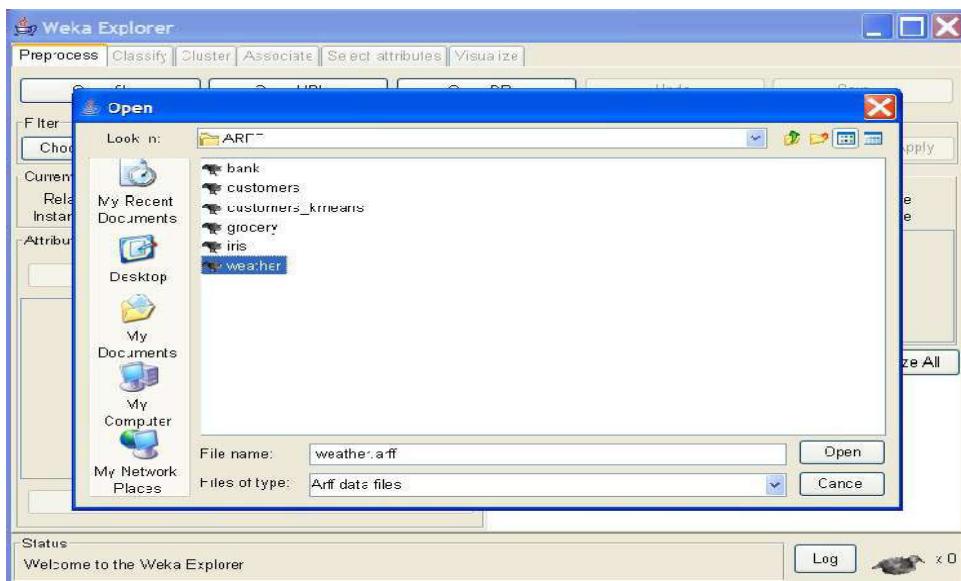


Opening file from a local file system

Click on 'Open file...' button

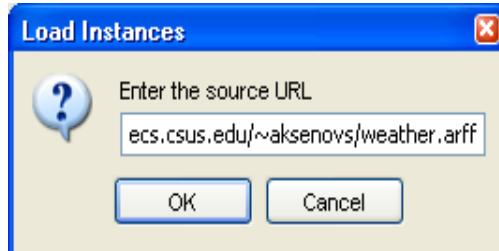


It brings up a dialog box allowing you to browse for the data file on the local file system, choose “weather.arff” file.

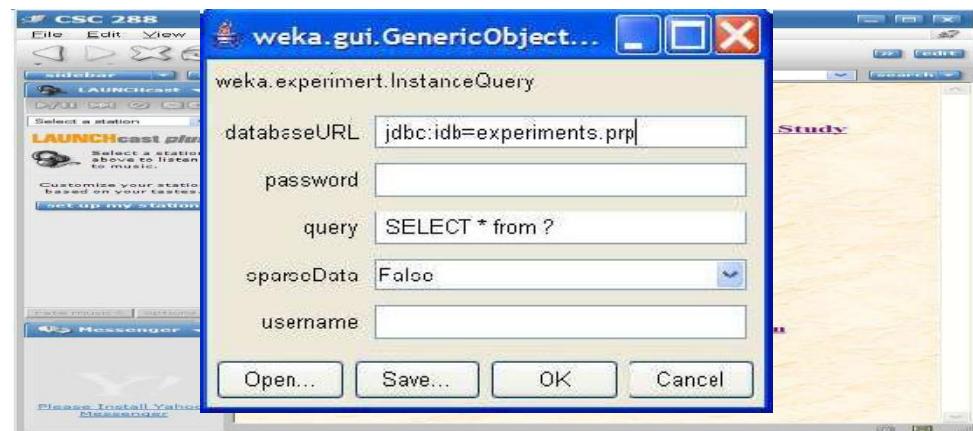


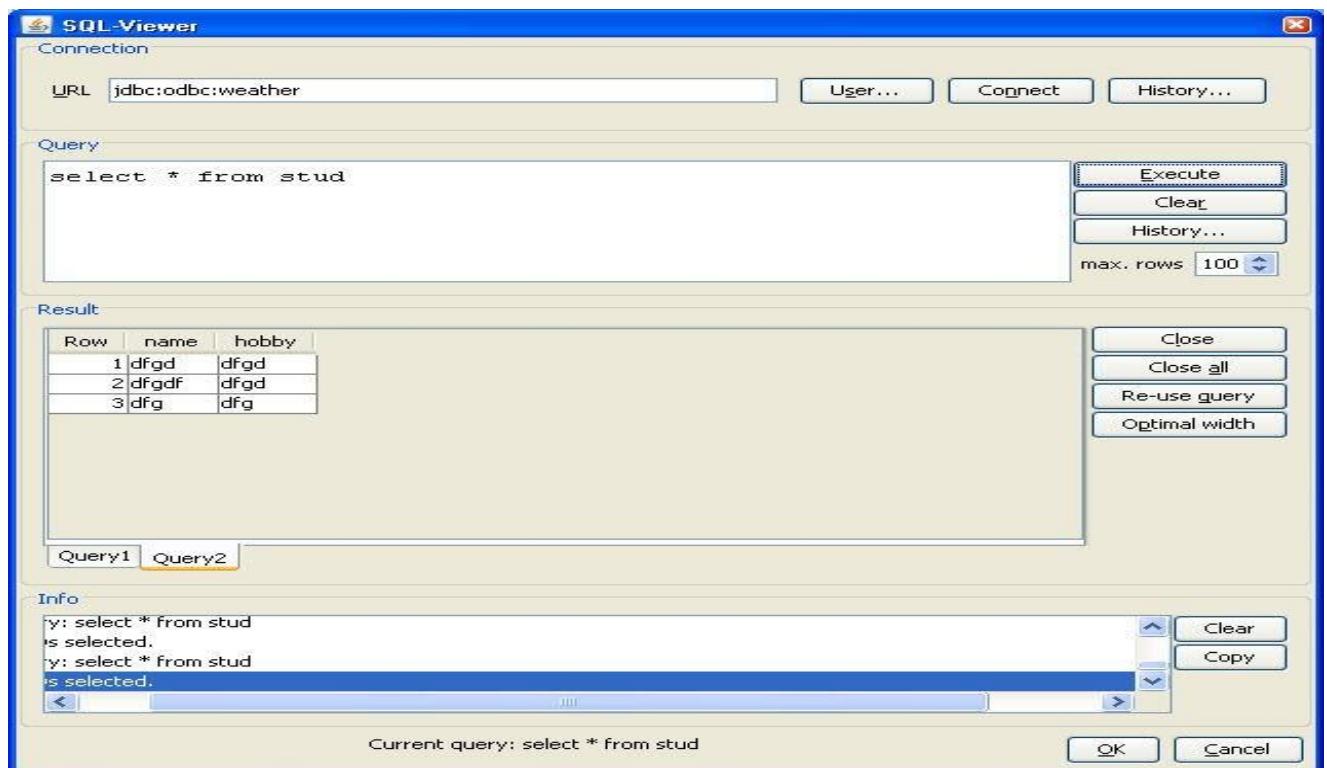
Opening file from a web site

A file can be opened from a website. Suppose, that “weather.arff ” is on the following website:



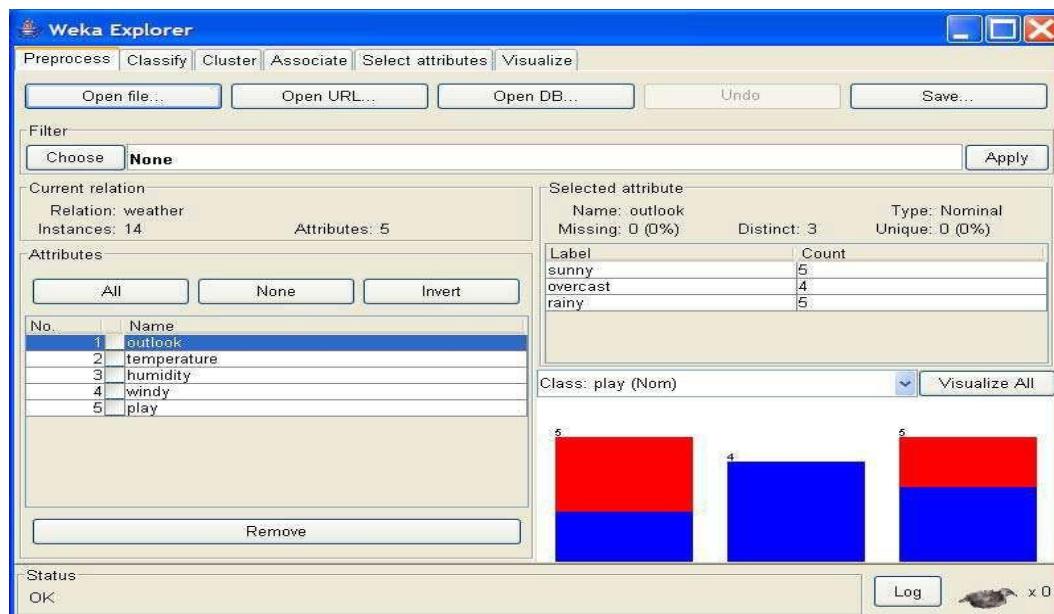
Reading data from a database:



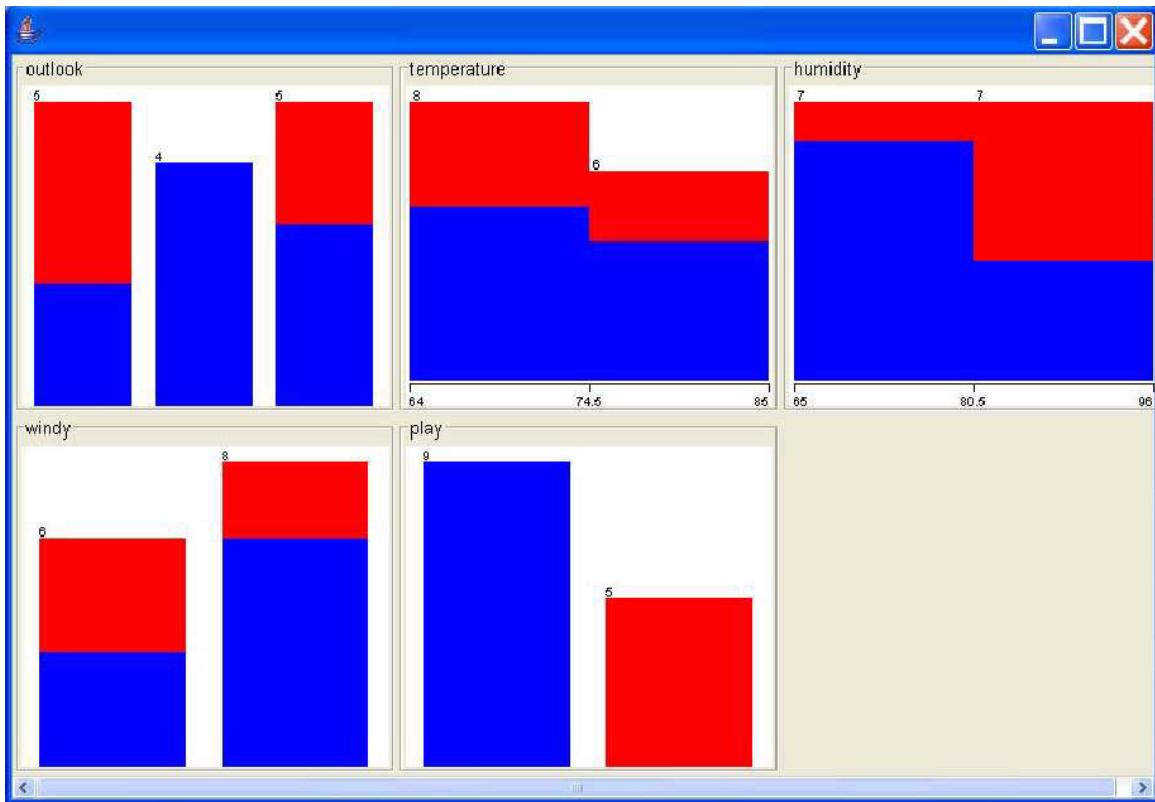


Loading data

The most common and easiest way of loading data into WEKA is from ARFF file, using Open File button.



\Visualize Attributes:



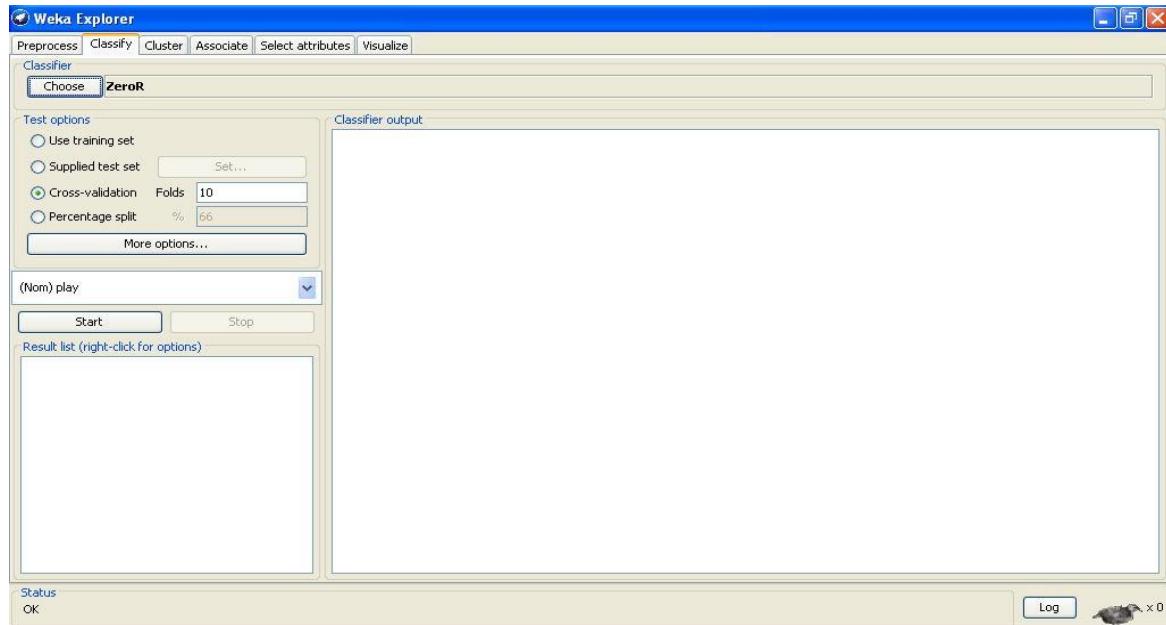
visualize all attributes by clicking on ‘Visualize All’ button.

Practical No - 10

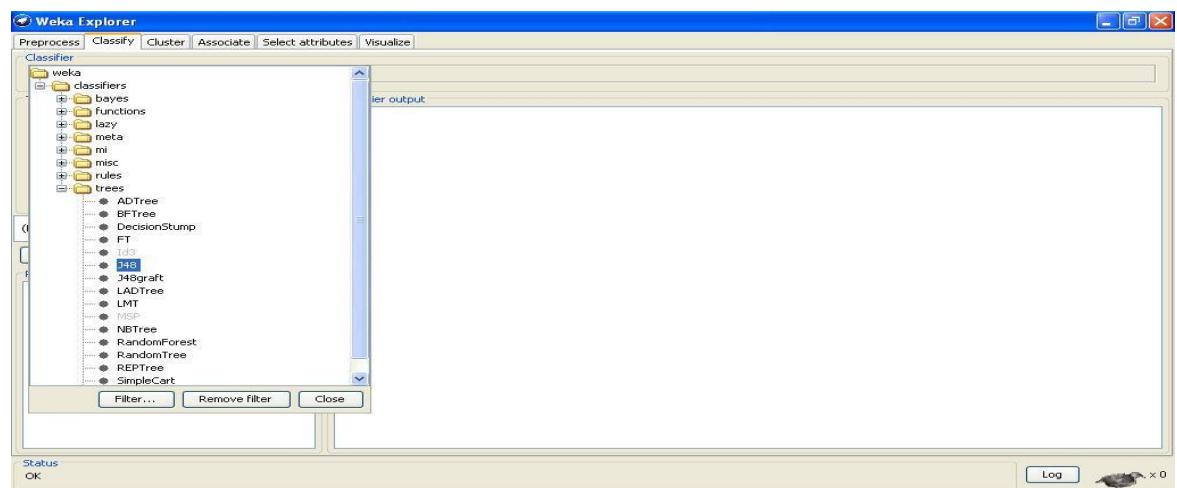
Aim : Classification problems.

Solution :

Once you have your data set loaded, all the tabs are available to you. Click on the ‘Classify’ tab.



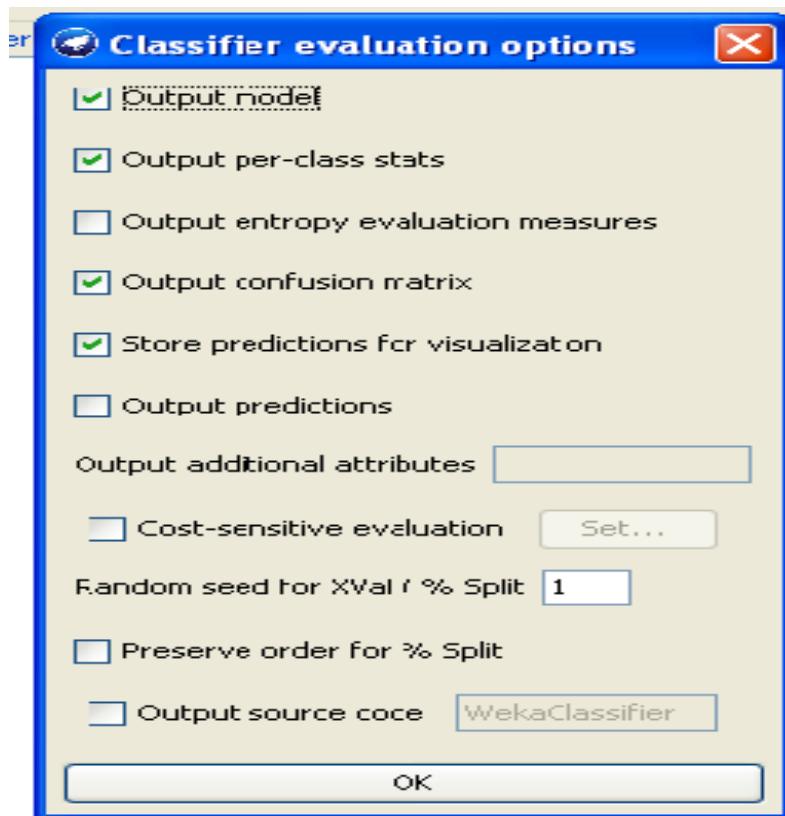
Click on ‘Choose’ button in the ‘Classifier’ box just below the tabs and select C4.5 classifier WEKA □ Classifiers □ Trees □ J48.



Check ‘Percentage split’ radio-button and keep it as default 66%. Click on ‘More options...’ button.

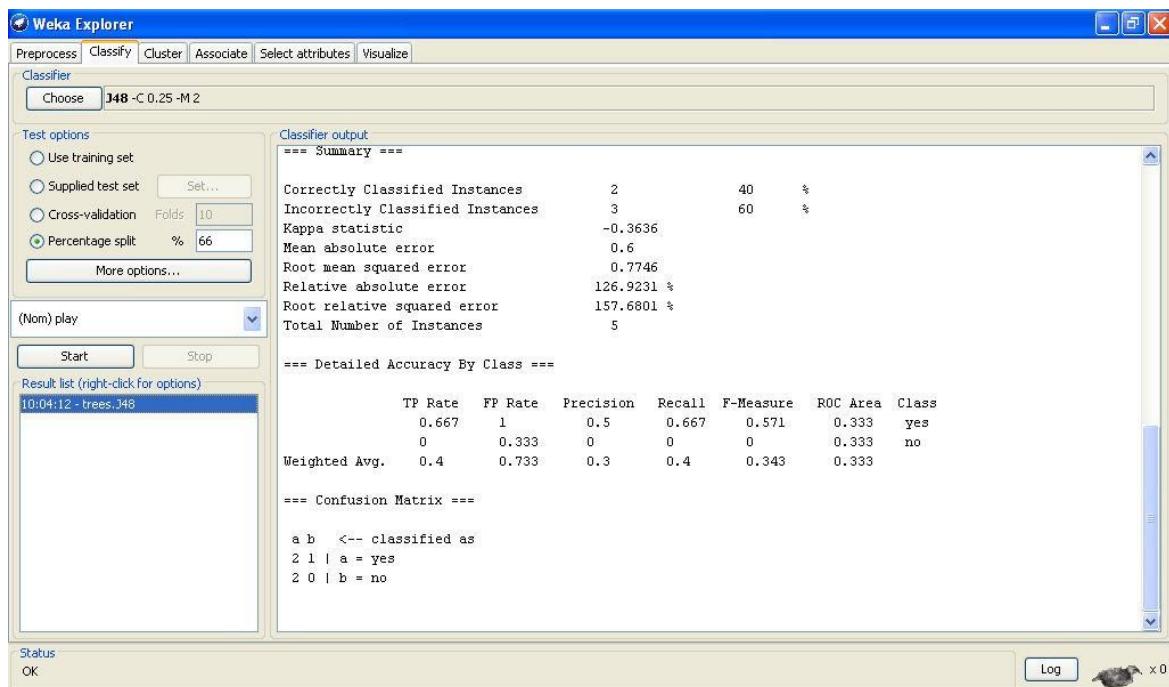
make sure that the Following options are checked :

1. Output model.
2. Output per-class stats.
3. Output confusion matrix
4. Store predictions for visualization.
5. Set ‘Random seed for Xval / % Split’ to 1.

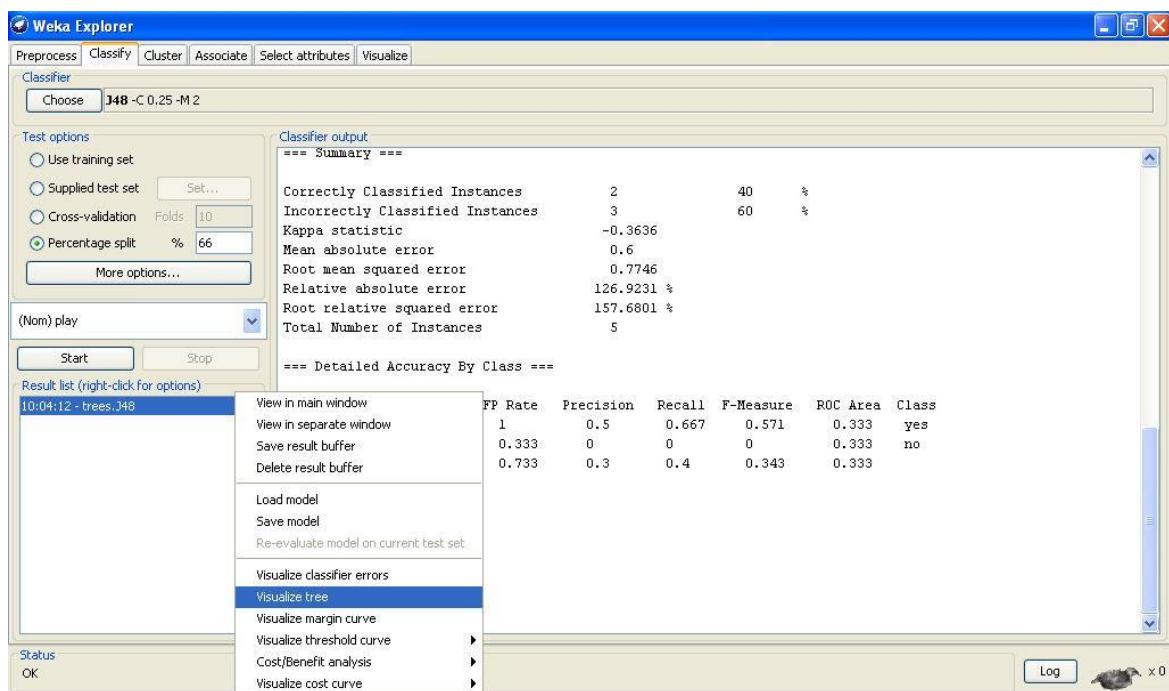


Once the options have been specified, you can run the classification algorithm. Click on

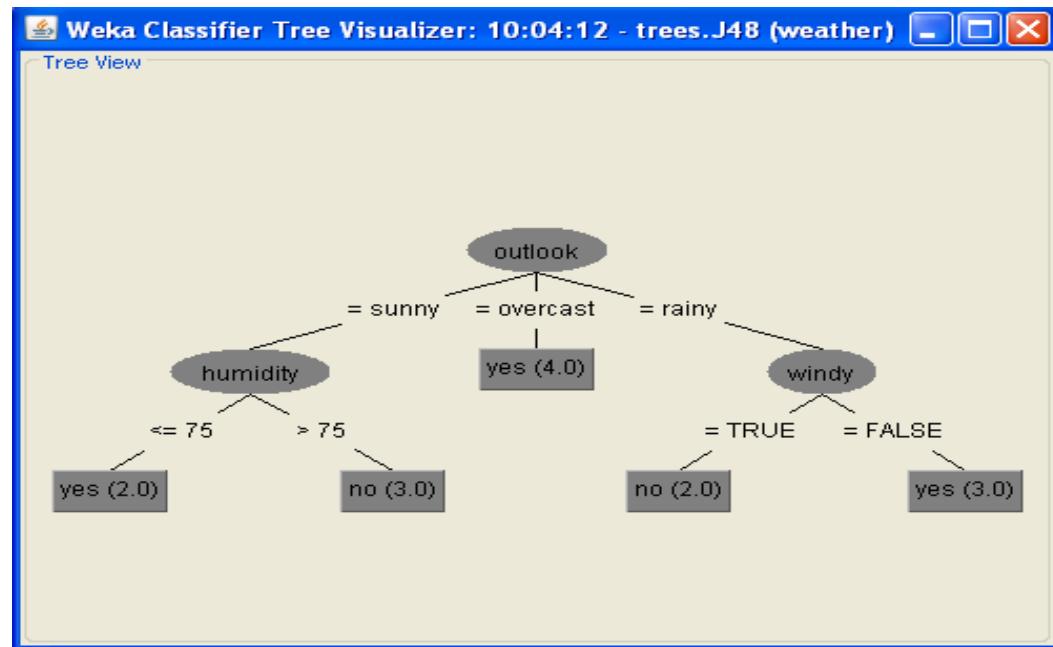
‘Start’ button



To see a graphical representation of the classification tree. Right-click on the entry in ‘Result list’ for which you would like to visualize a tree.

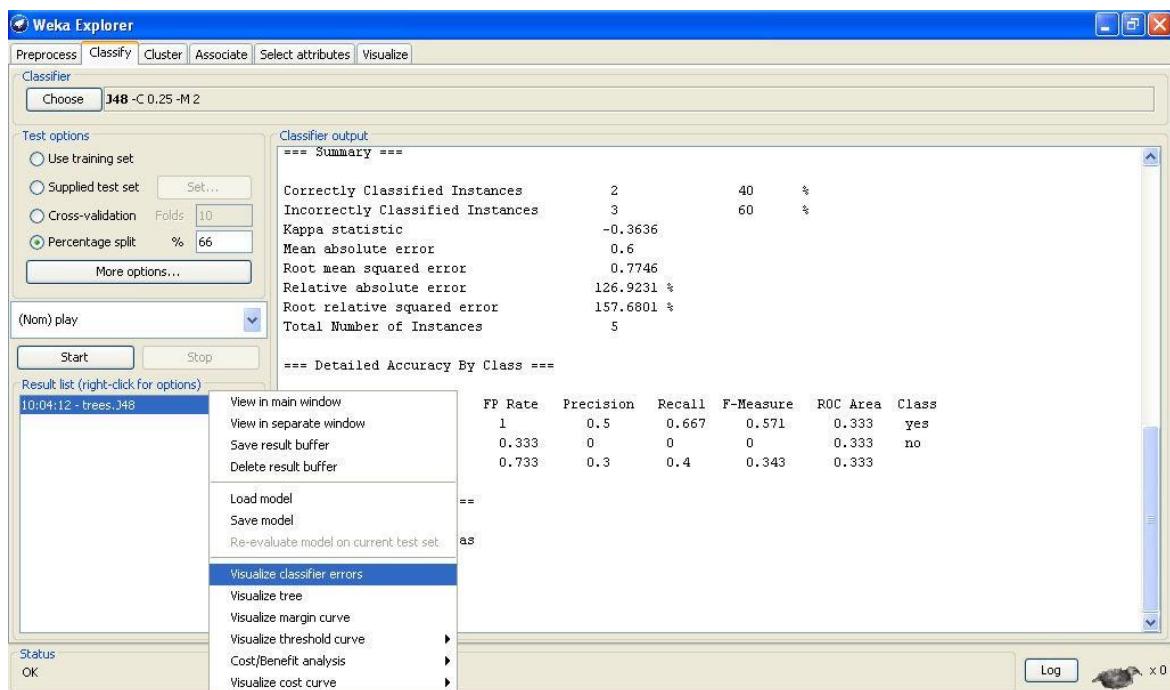


Select the item ‘Visualize tree’; a new window comes up to the screen displaying the tree.

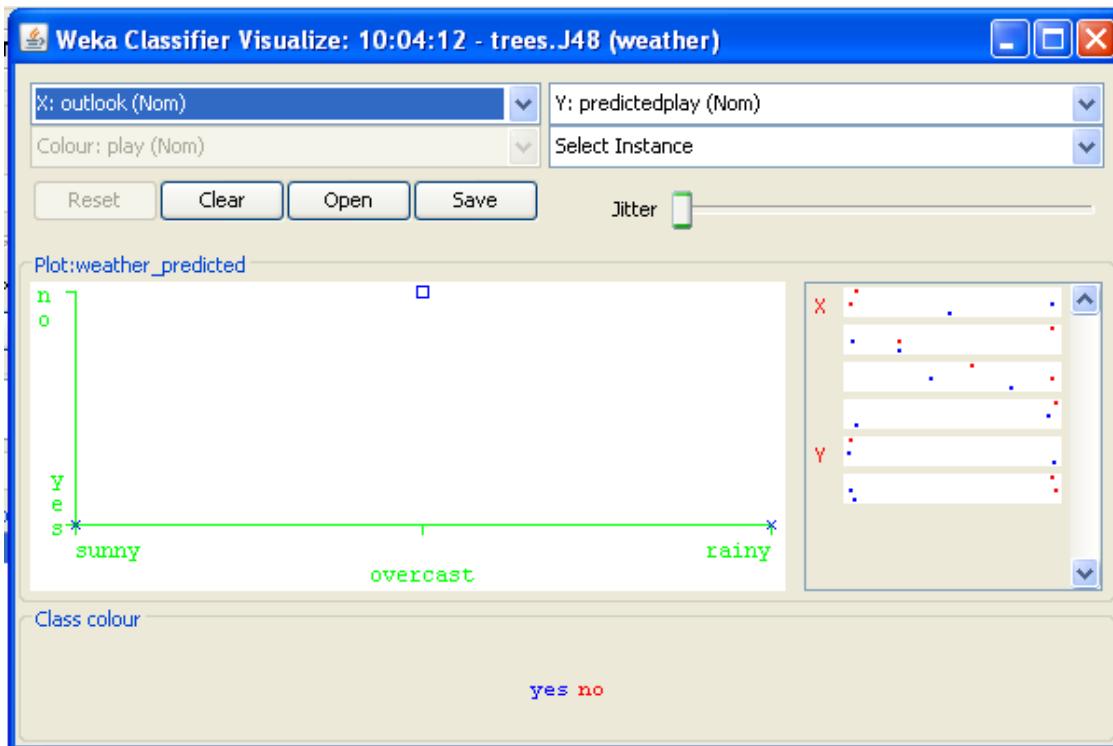


To visualize classification errors. Right-click on the entry in ‘Result list’ again

and select ‘Visualize classifier errors’ from the menu:



'Visualize' window displaying graph appears on the screen.



Practical N0 - 11

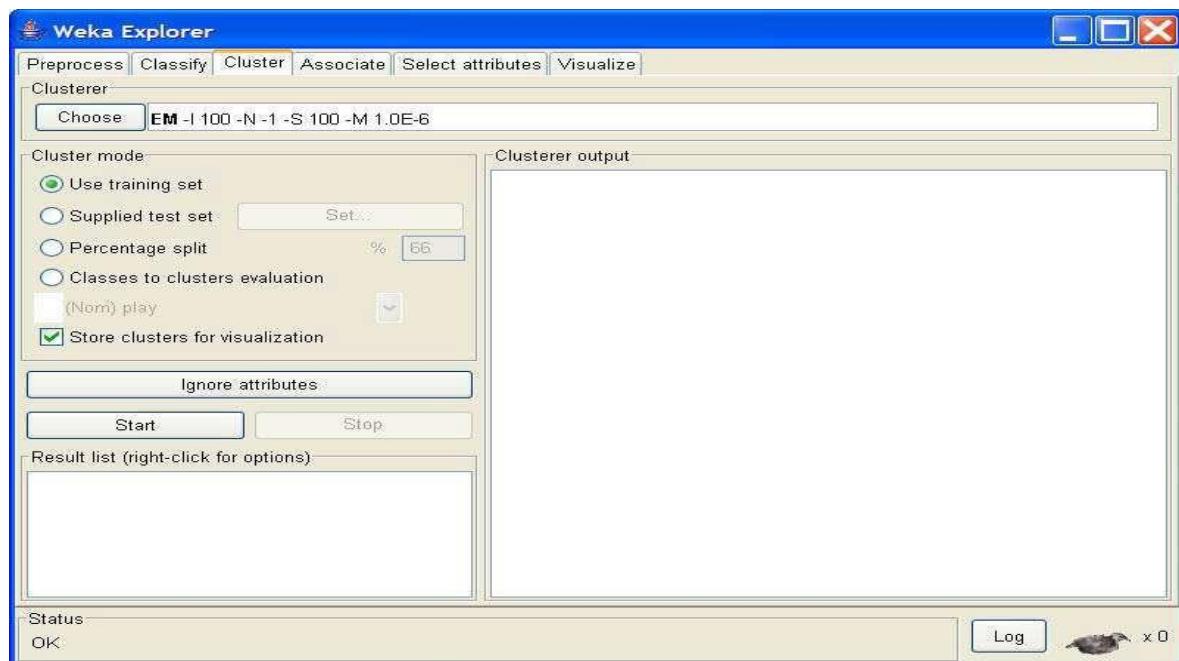
Aim : Clustering Analysis.

Solution :

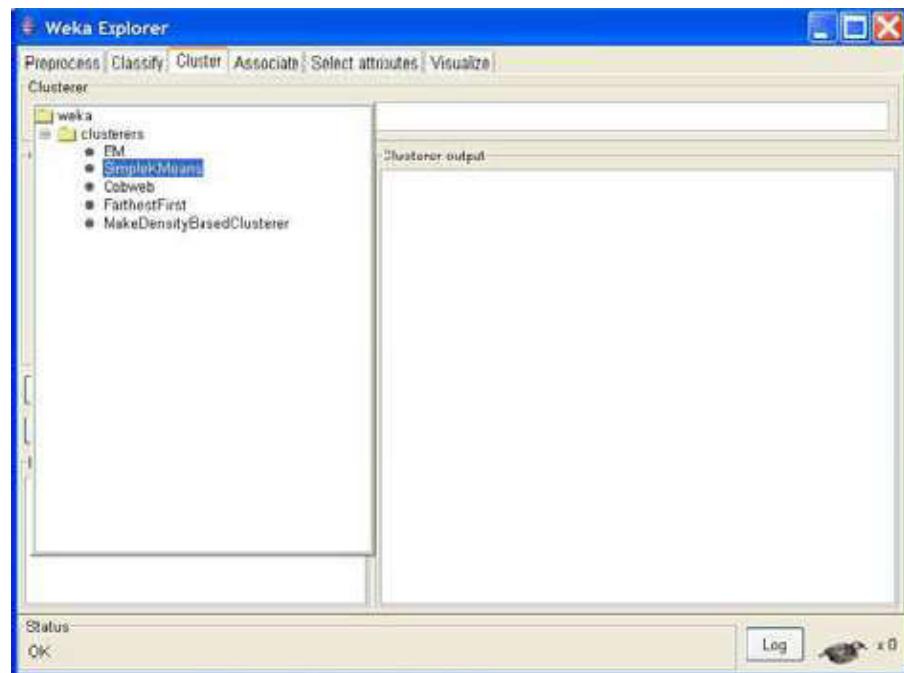
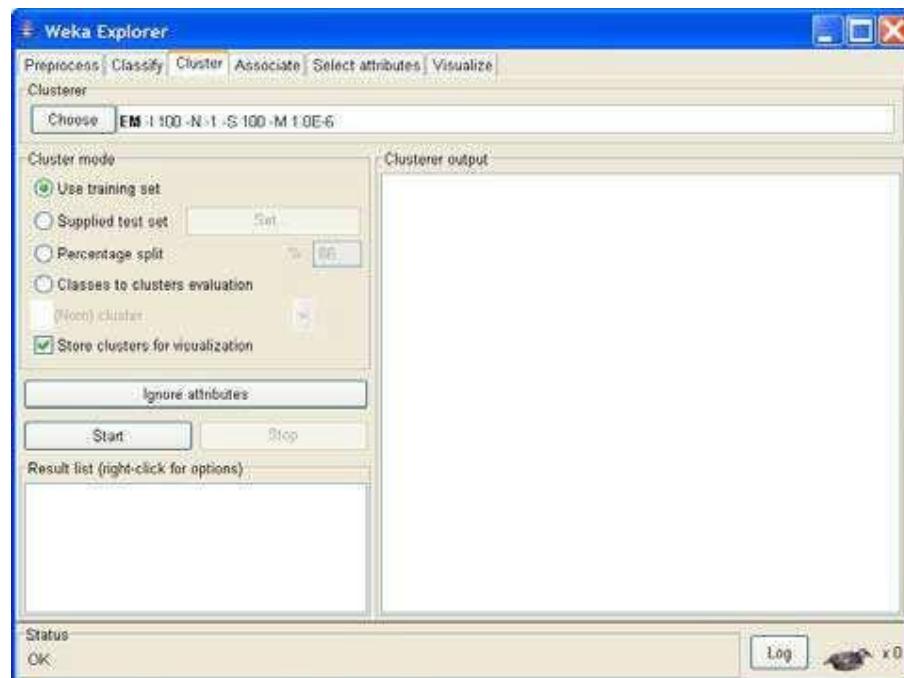
we will use customer data [6] that is contained in “customers.arff” file and analyze it with k-means clustering scheme.

	A	B	C	D	E
1	Income	Age	Children	Marital Status	Education
2					
3	25000	35	3	single	high school
4	15000	25	1	married	high school
5	20000	40	0	single	high school
6	30000	20	0	divorced	high school
7	20000	25	3	divorced	college
8	70000	60	0	married	college
9	90000	30	0	married	graduate school
10	200000	45	5	married	graduate school
11	100000	50	2	divorced	college
12					

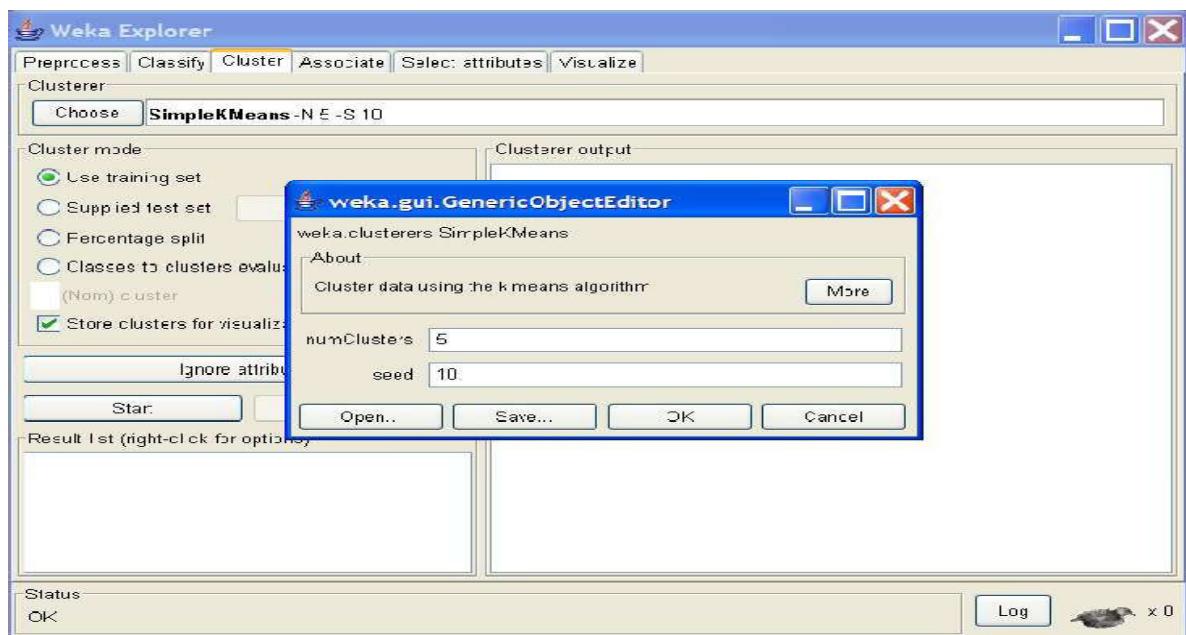
In ‘Preprocess’ window click on ‘Open file...’ button and select “customers.arff” file. Click ‘Cluster’ tab at the top of WEKA Explorer window.



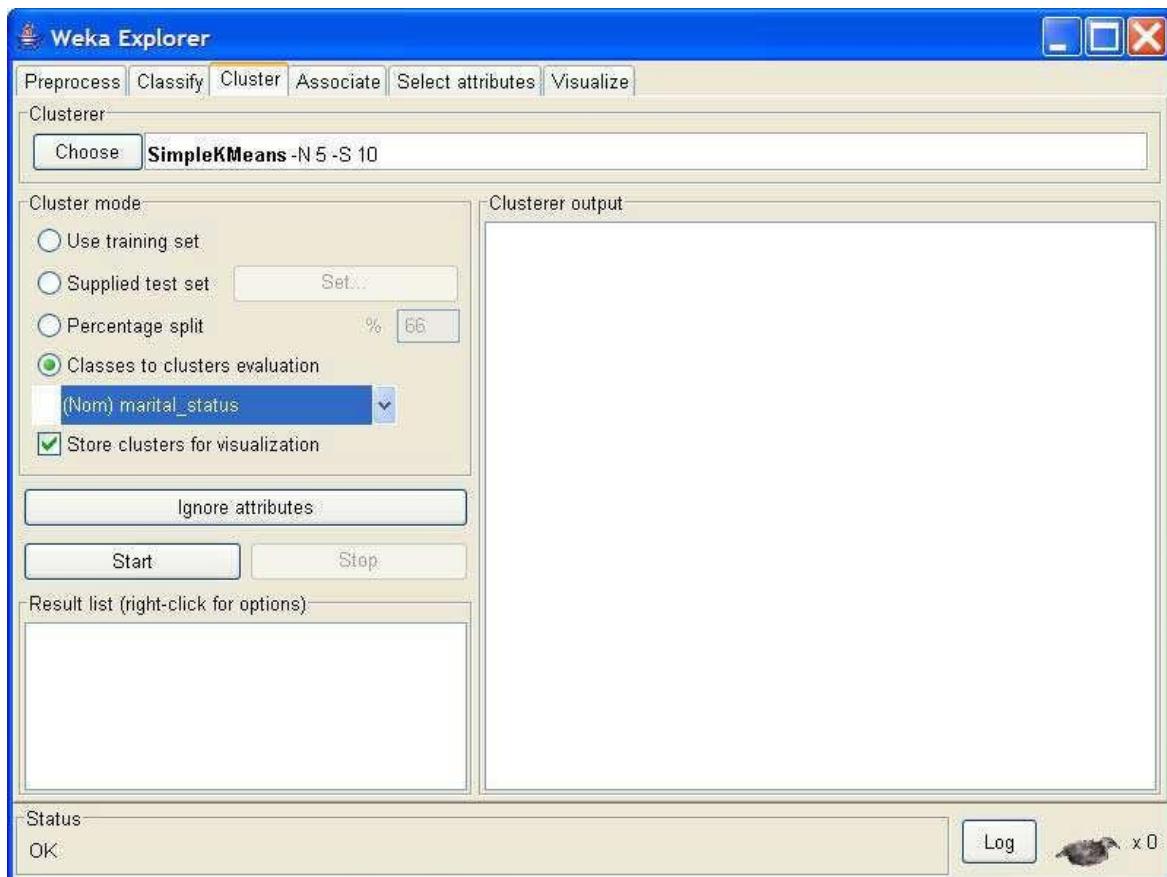
In the ‘Clusterer’ box click on ‘Choose’ button. In pull-down menu select WEKA □ Clusterers, and select the cluster scheme ‘SimpleKMeans’. Some implementations of K-means only allow numerical values for attributes.



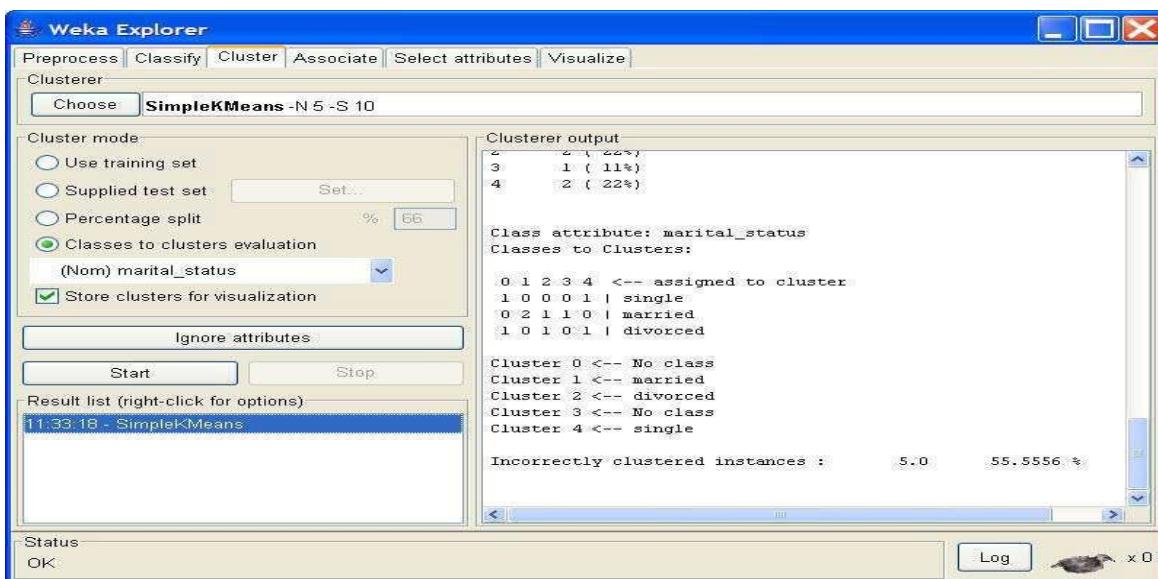
right-click on the algorithm "weak.gui.GenericObjectEditor" comes up to the screen. Set the value in "numClusters" box to 5(instead of default 2) because you have five clusters in your .arff file.



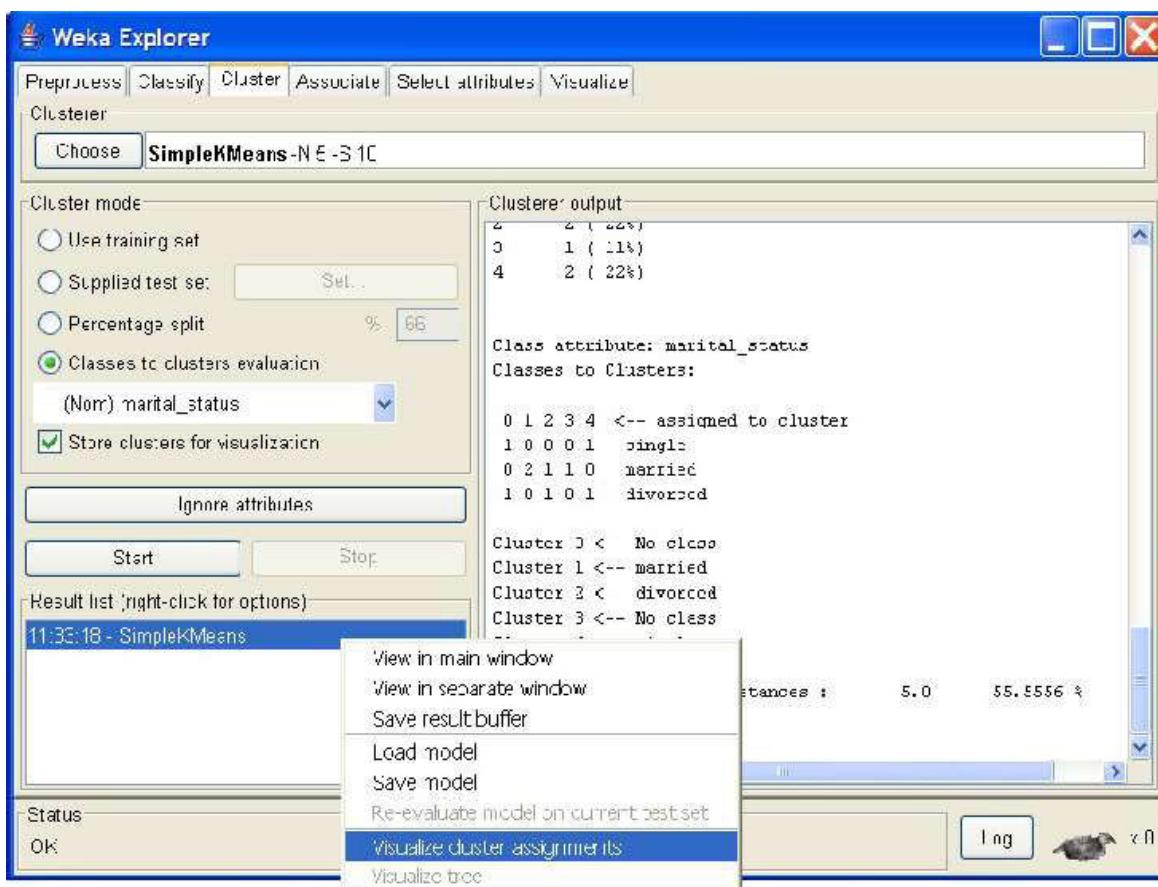
Click on ‘Classes to cluster evaluation’ radio-button in ‘Cluster mode’ box and select ‘marital_status’ in the pull-down box below.



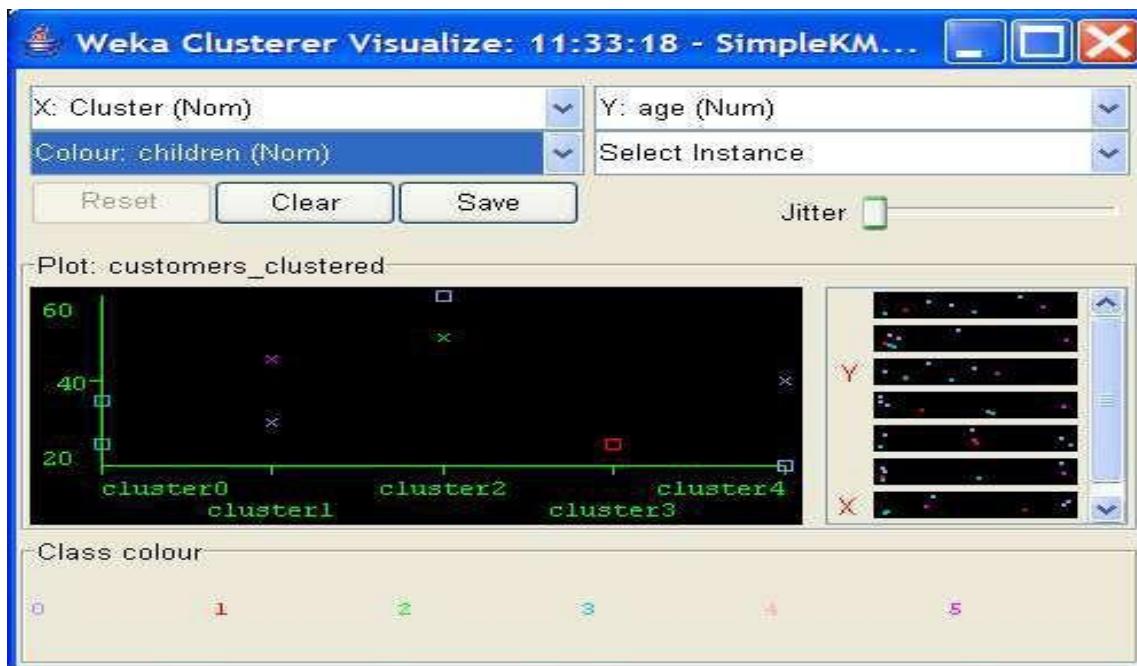
Click on the ‘Start’ button to execute the algorithm.



Right-click on the entry in the ‘Result list’ and select ‘Visualize cluster assignments’ in the pull-down window.



‘Weka Clusterer Visualize’ window.



there is a new attribute appeared in the file – ‘cluster’ that was added by WEKA. This attribute represents the clustering done by WEKA.

```

customers_kmeans - Microsoft Word
File Edit View Insert Format Tools Table Window Help Acrobat
Plain Text CounterNew 10 100% 100%
Favorites Go
Brelation customers_clustered
@attribute Instance_number numeric
@attribute income numeric
@attribute age numeric
@attribute children {0,1,2,3,4,5}
@attribute marital_status {single,married,divorced}
@attribute education {high_school,college,graduate_school}
@attribute Cluster {cluster0,cluster1,cluster2,cluster3,cluster4}

@data
0,25000,35,3,single,high_school,cluster0
1,15000,25,1,married,high_school,cluster3
2,20000,40,0,single,high_school,cluster4
3,30000,20,0,divorced,high_school,cluster4
4,20000,25,3,divorced,college,cluster0
5,70000,60,0,married,college,cluster2
6,90000,30,0,married,graduate_school,cluster1
7,20000,45,5,married,graduate_school,cluster1
8,100000,50,2,divorced,college,cluster2

```

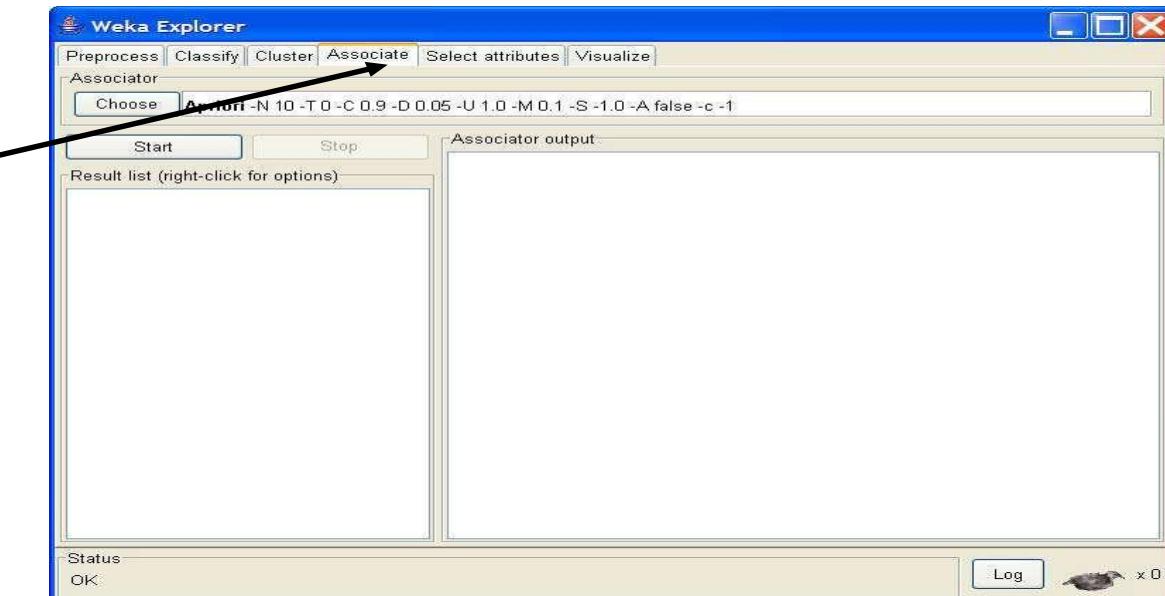
Practical No-12

Aim : Association Rule Mining.

Solution :

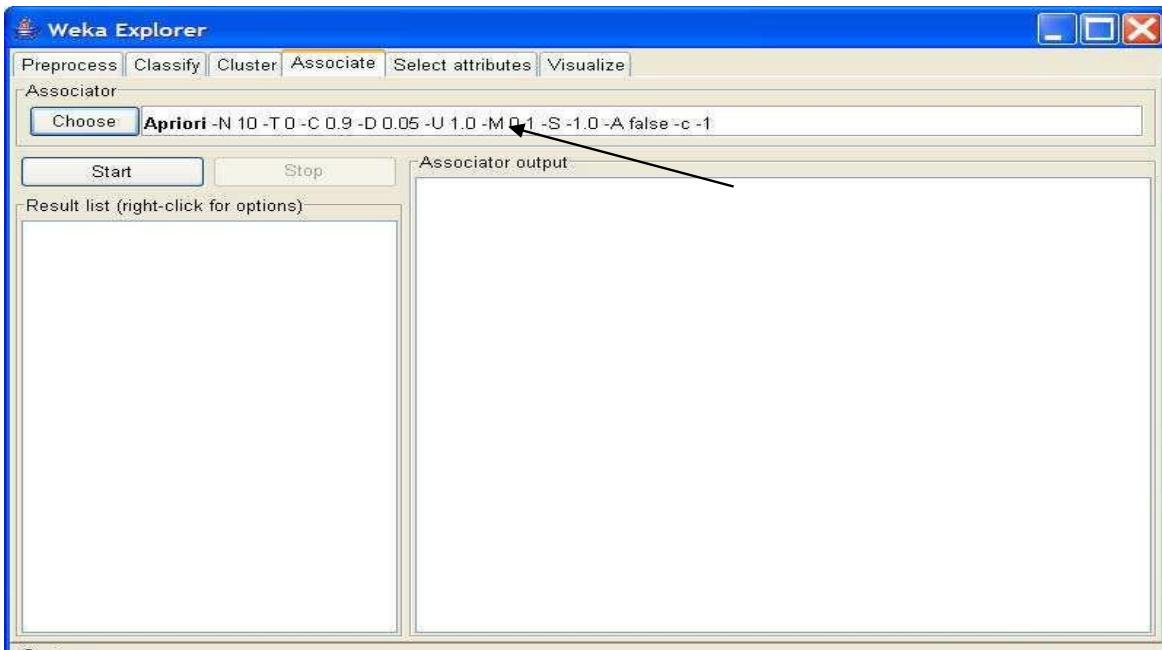
Choosing Association Scheme

Click ‘Associate’ tab at the top of ‘WEKA Explorer’ window. It brings up interface for the Apriori algorithm.

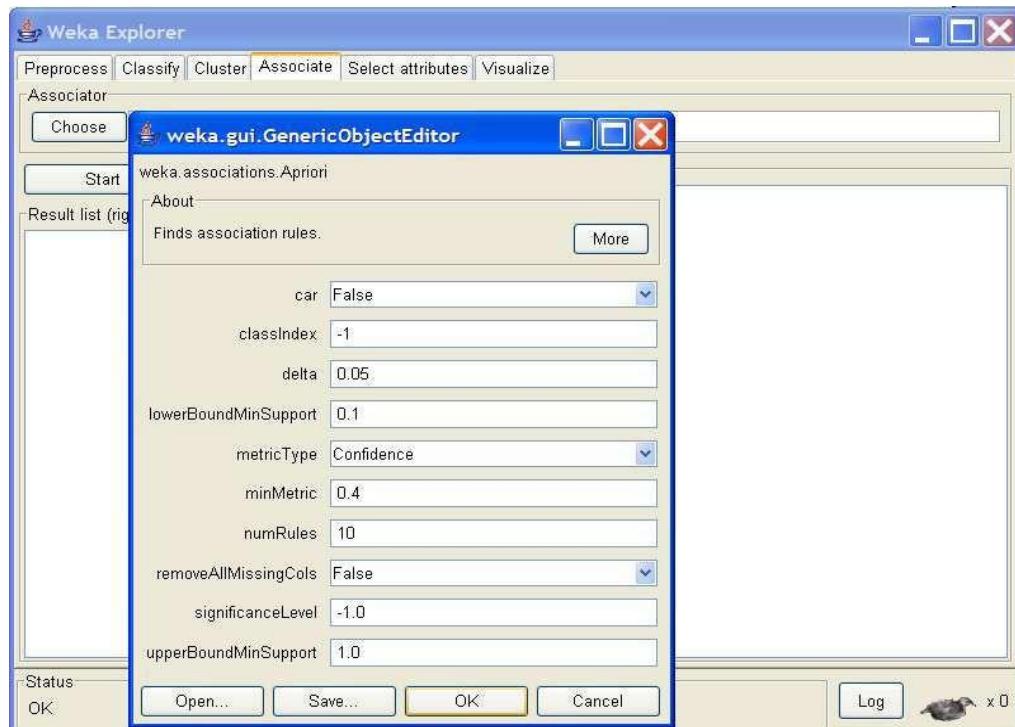


Setting Test Options

Check the text field in the ‘Associator’ box at the top of the window



Right-click on the ‘Associator’ box, ‘GenericObjectEditor’ appears on your screen



Click on the ‘Start’ button to execute the algorithm

