

Practical No 1

Aim : Creation of Dimensions and Fact tables.

Solution :

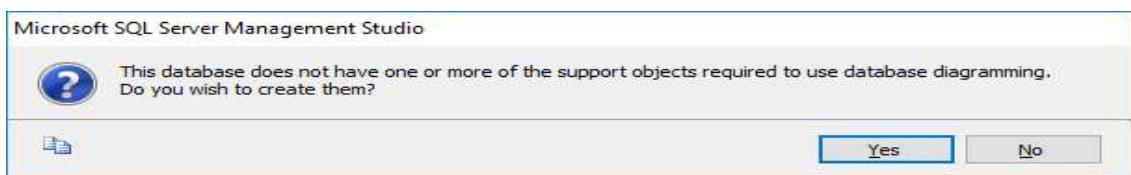
Open Application -> Microsoft SQL Server 2008 R2 -> SQL Server Management Studio

- 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

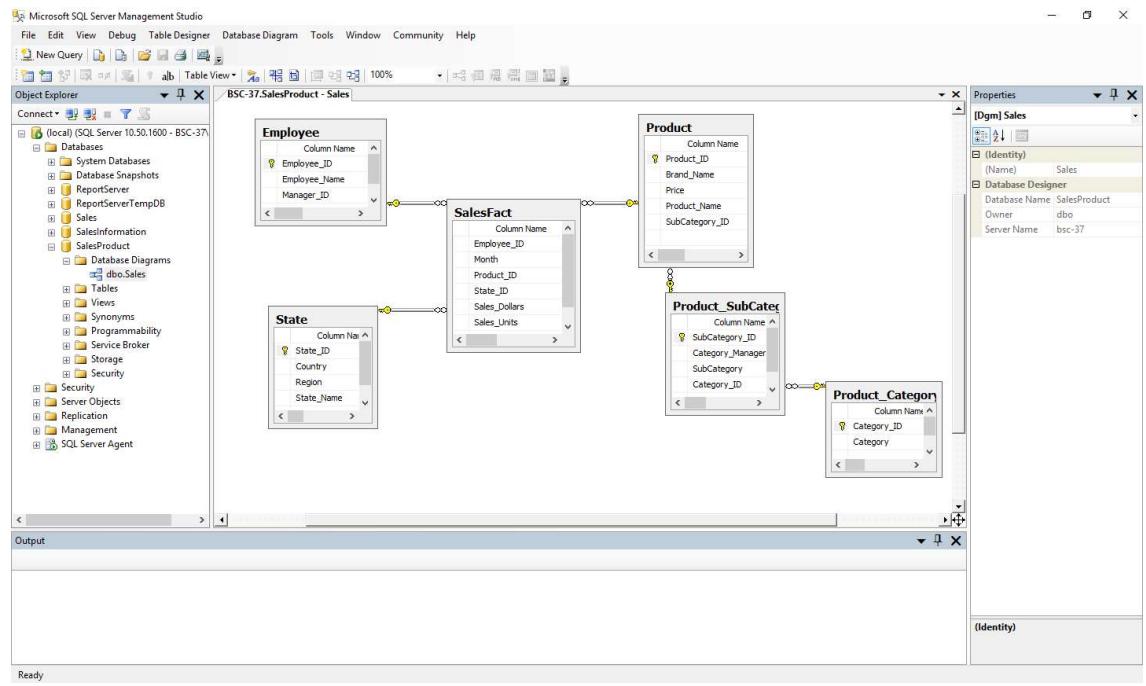
Expand the “SalesProduct” database folder.

- 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.)**

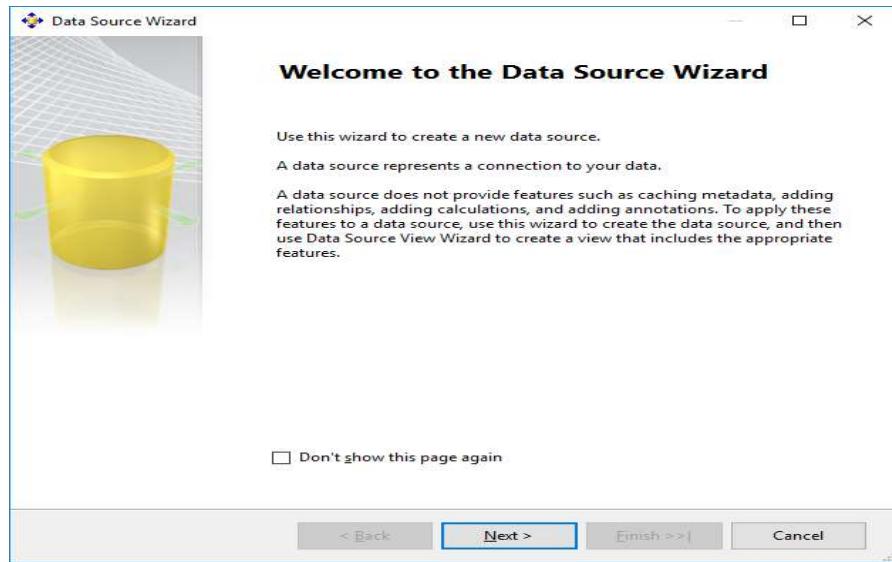
Practical No 2

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

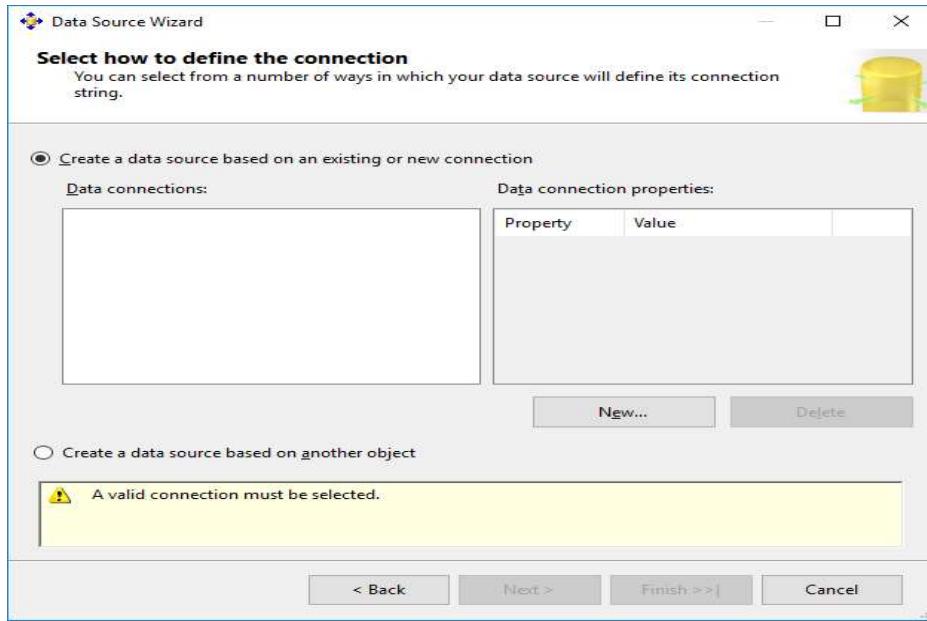
Solution :

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**

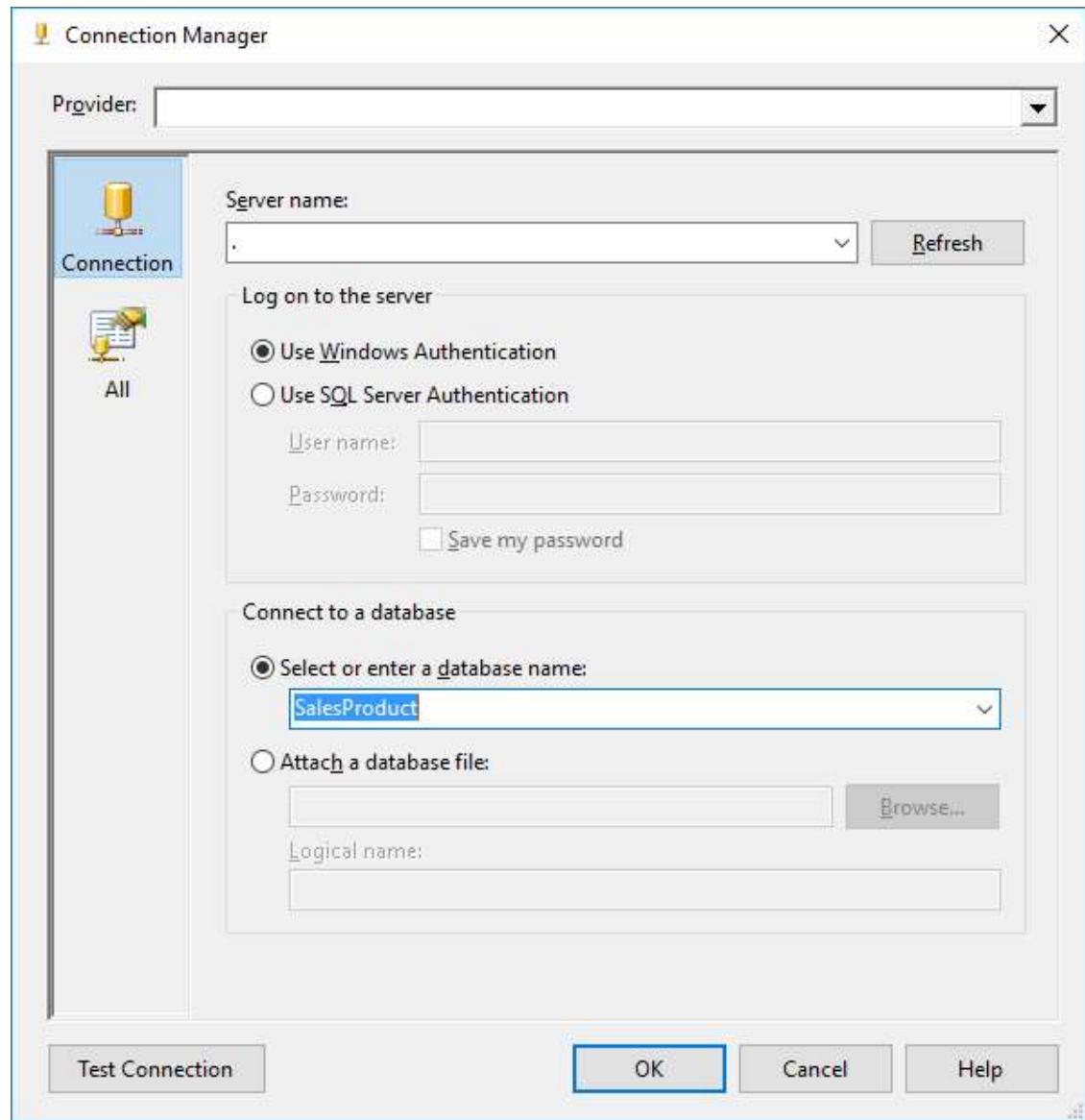


Click on Next.



Click on New.

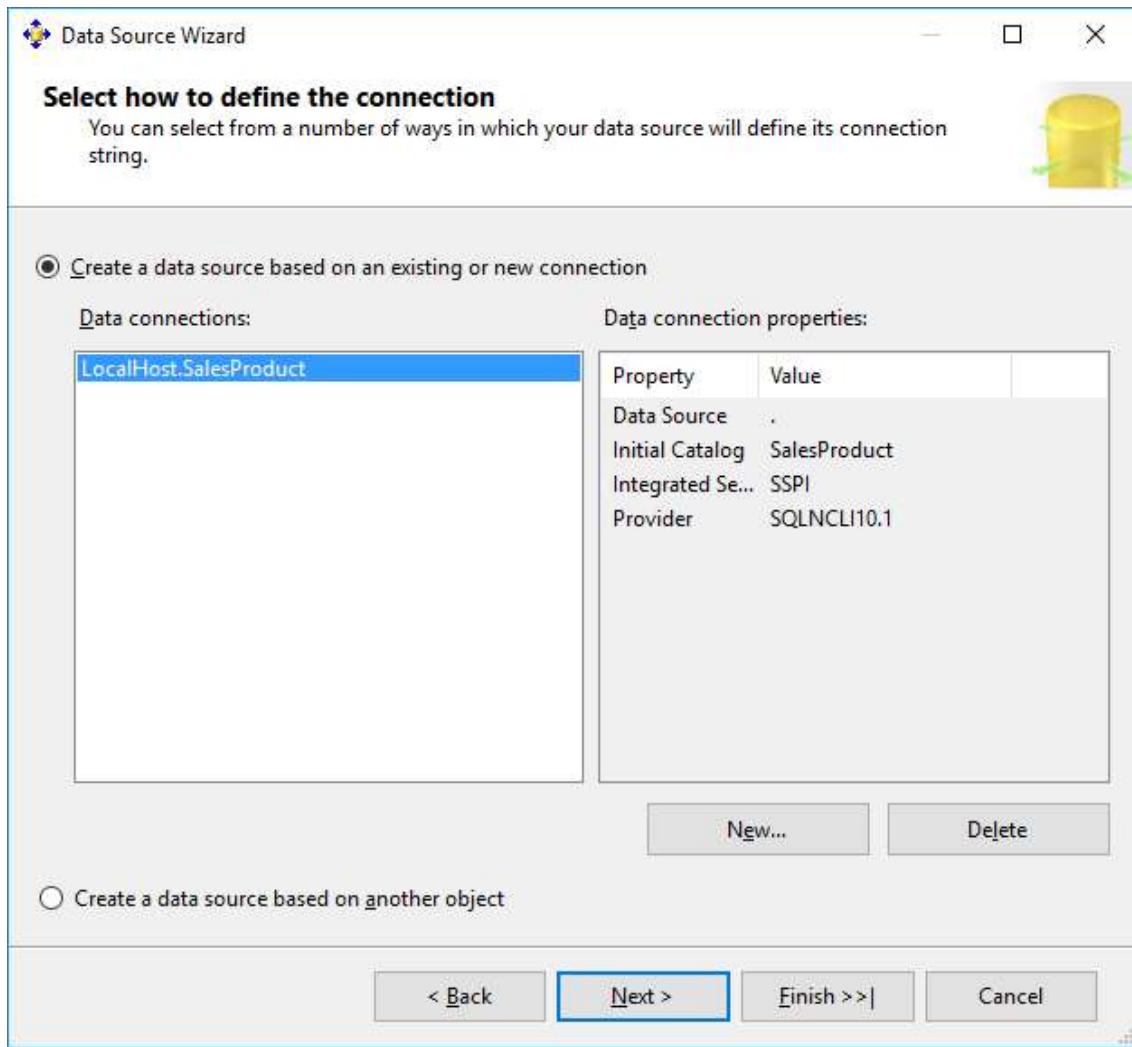
3. Choose Provider as “Microsoft OLEDB Provider for SQL Server” , Server Name as “.”, Select database name as “SalesProduct”.(Created in SQL Server Management studio).



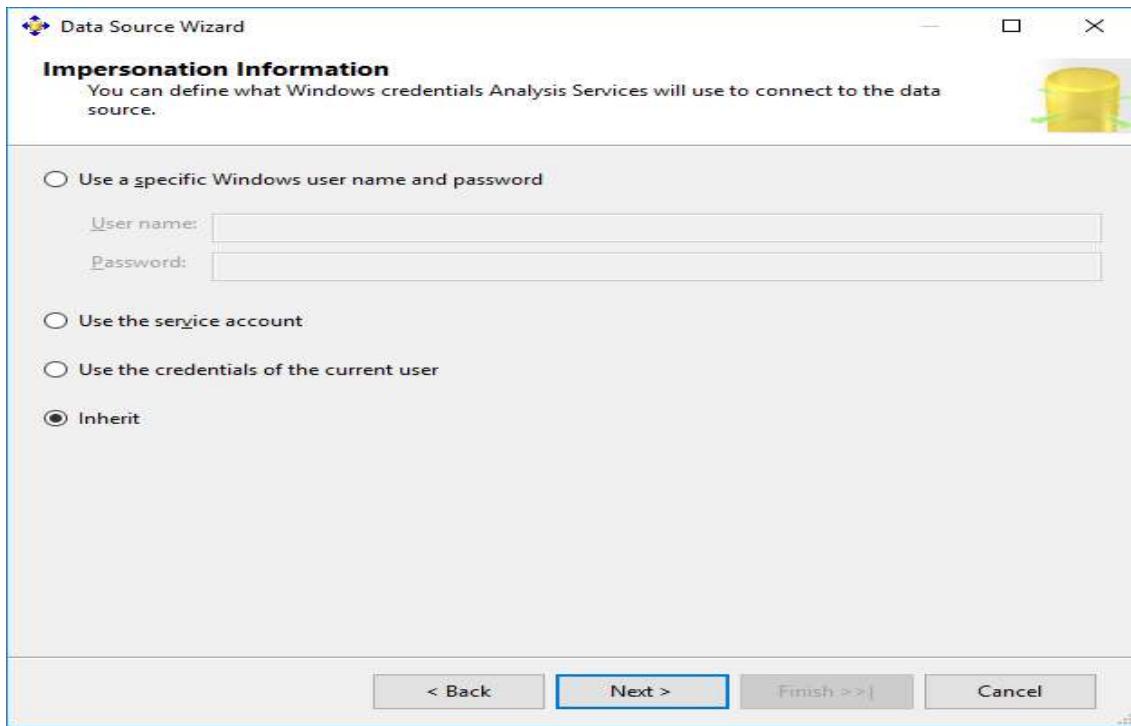
4. Click on Test Connection.



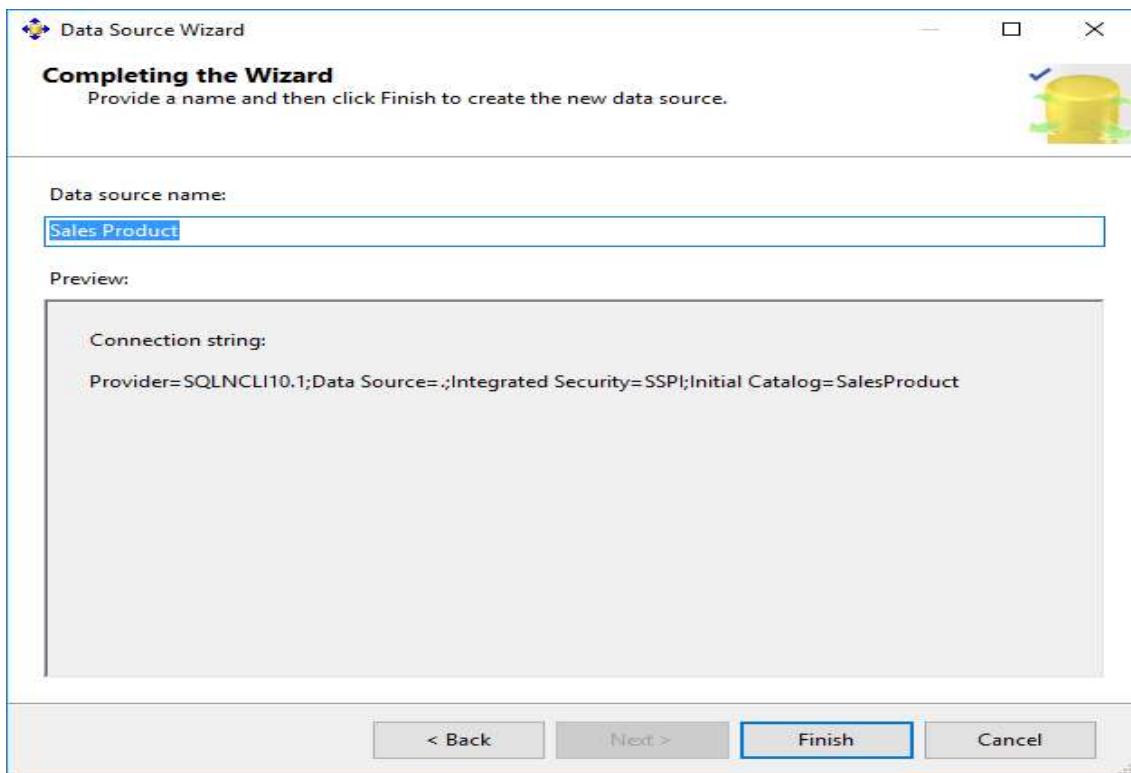
Click on OK.



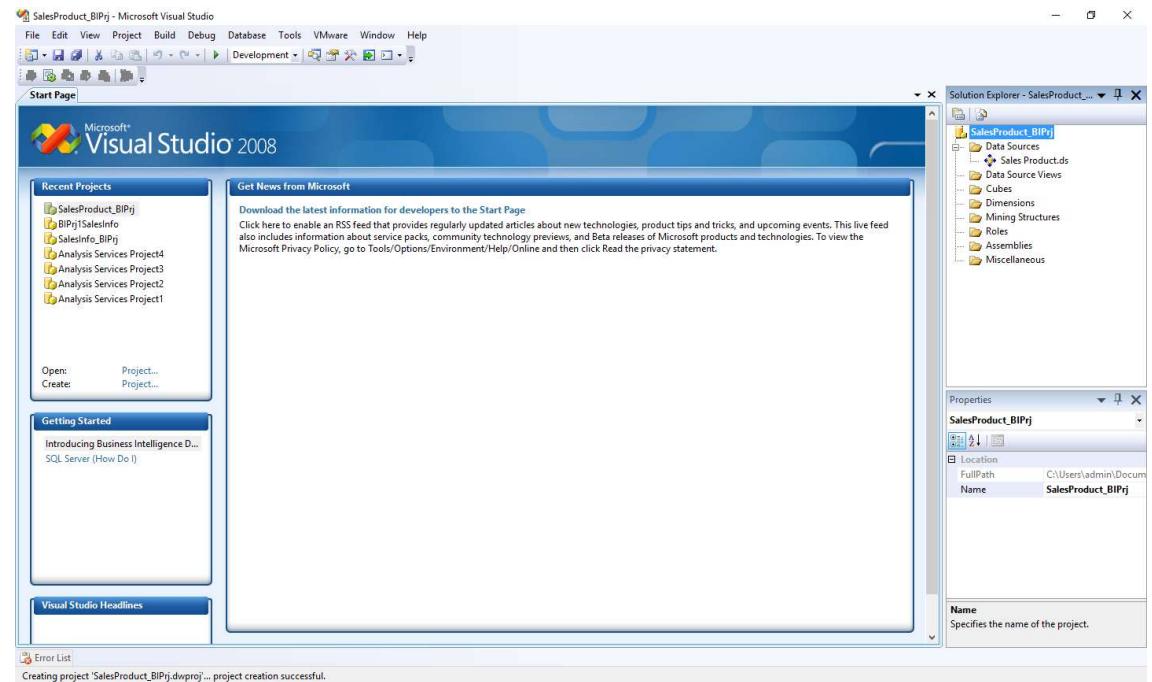
- Click on Next**
- 5. Choose “Inherit” option.**



- Click on Next.**
6. **Click on Finish.**



Name Data Source as “Sales Product”.

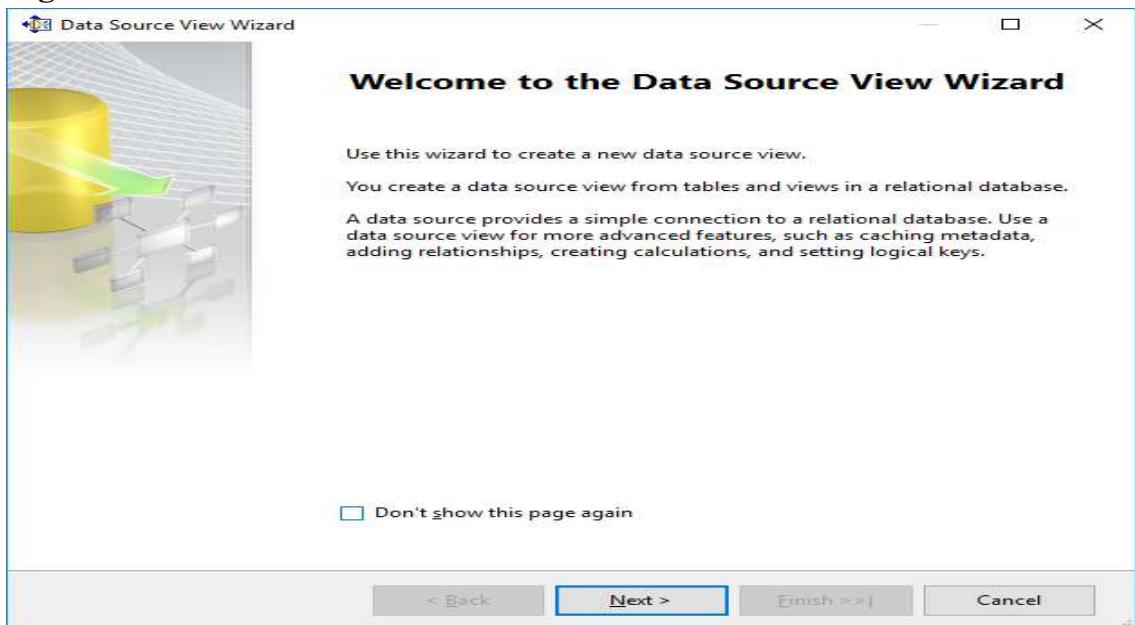


Practical No 3

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

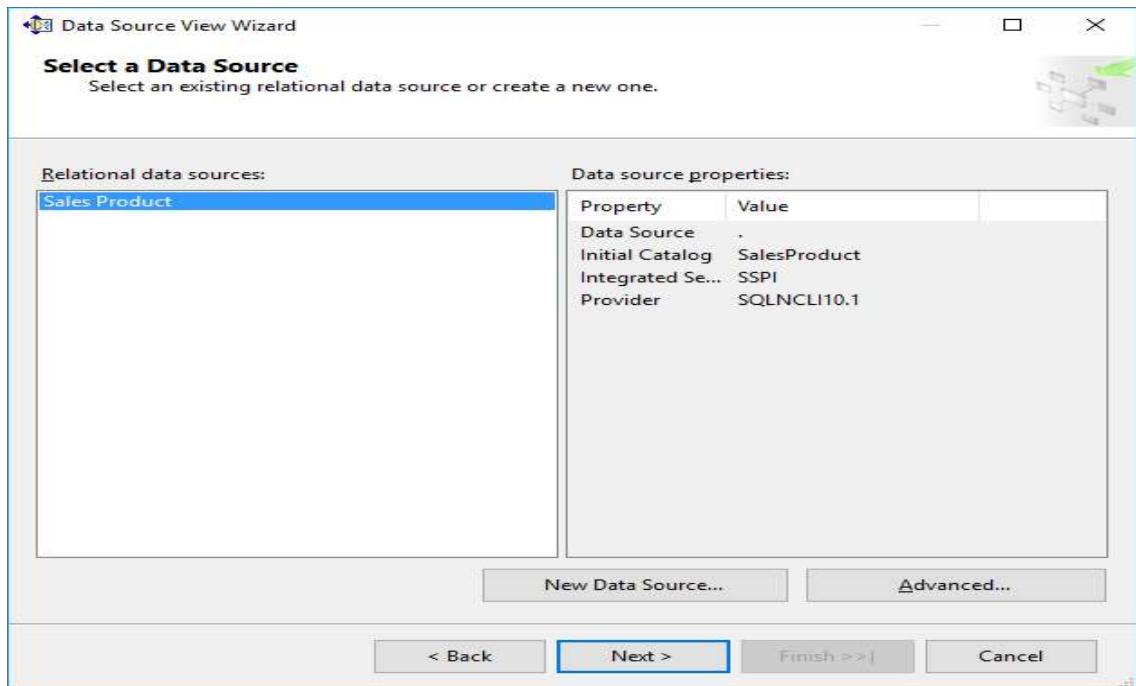
Solution :

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

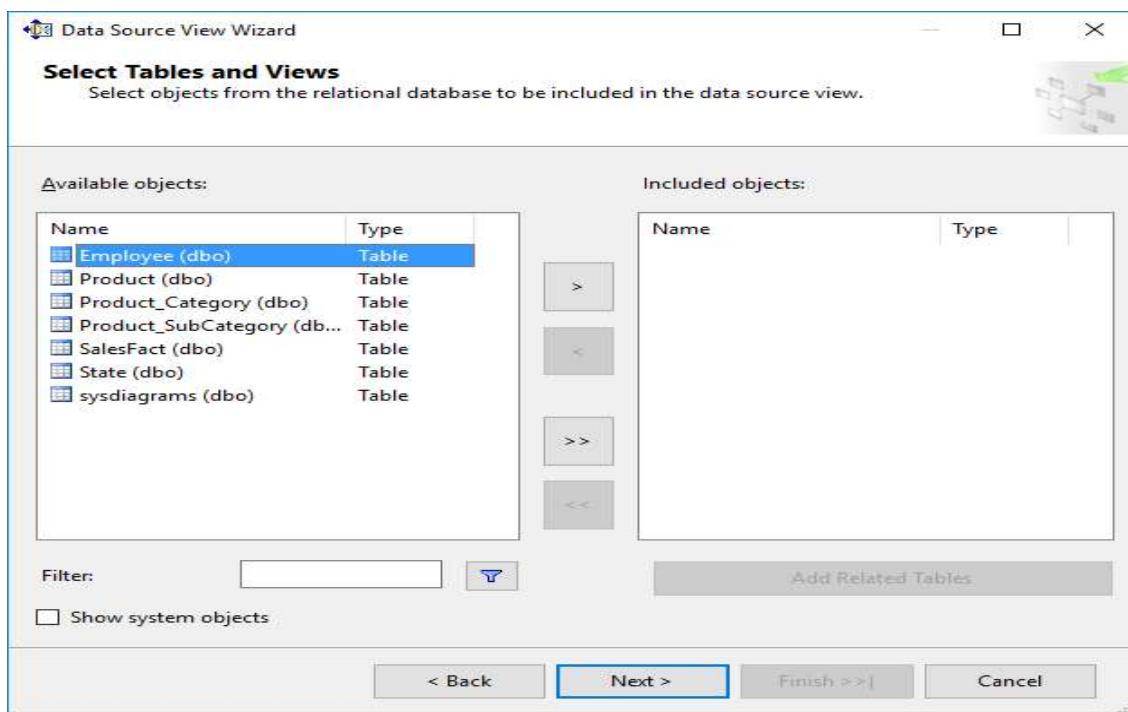


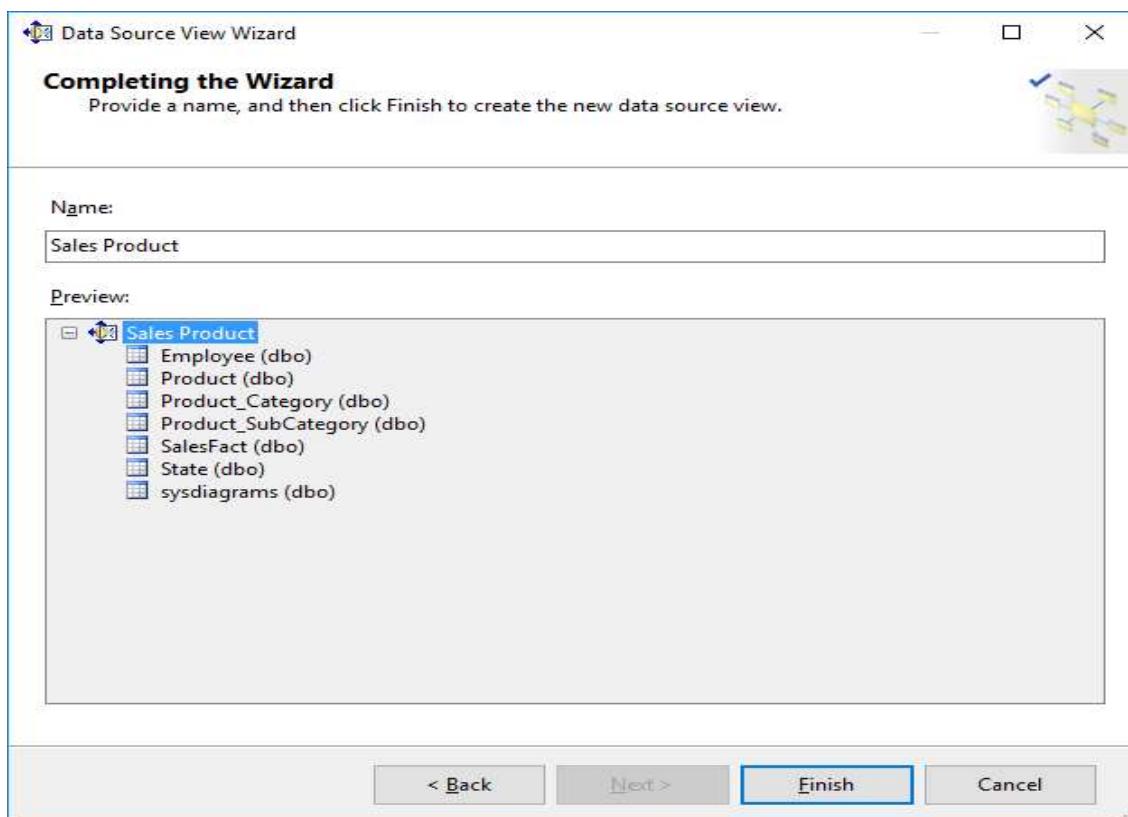
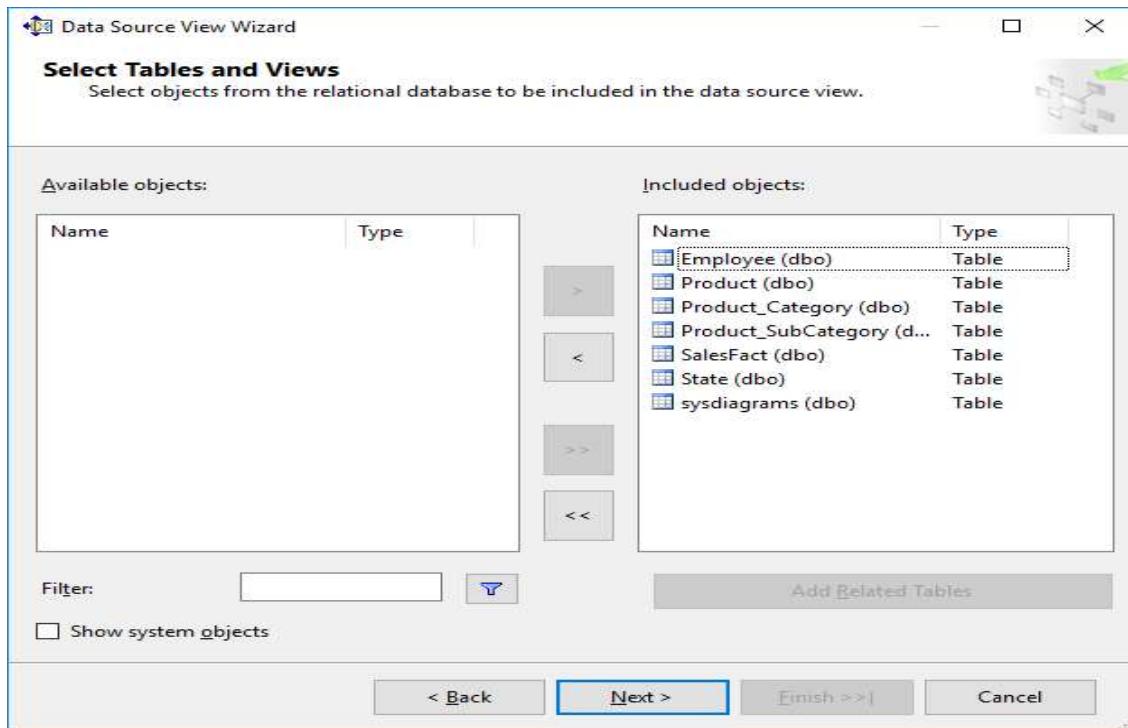
Click on Next.

- 2. Click on Next.**



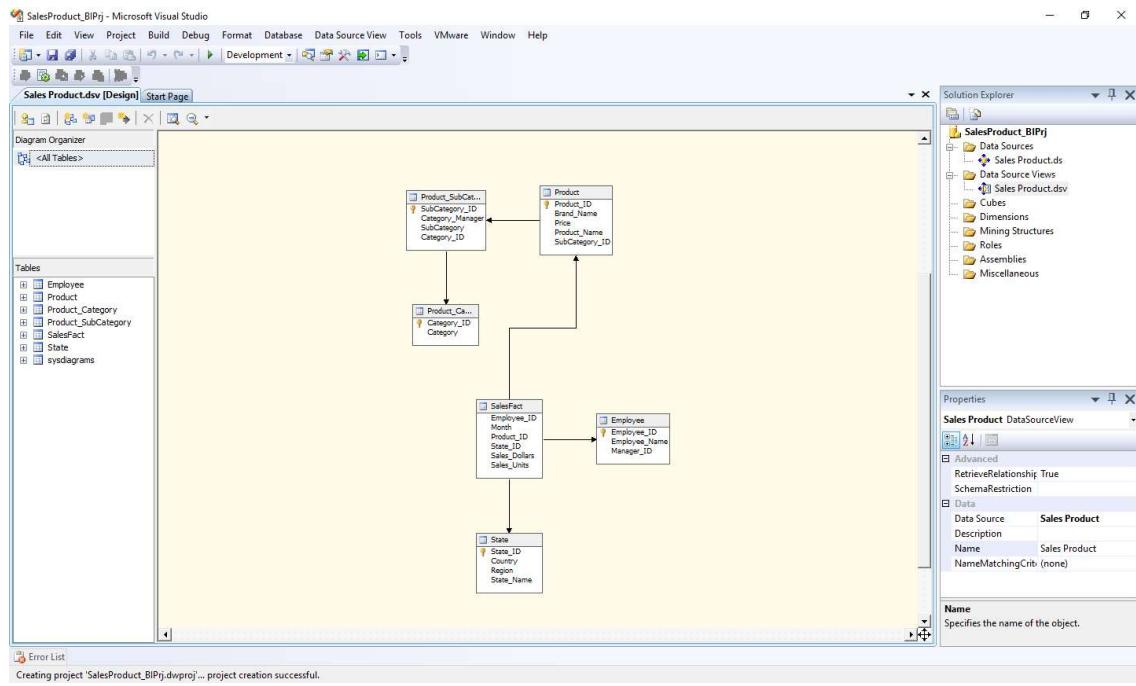
3. Select Tables and Views.





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.

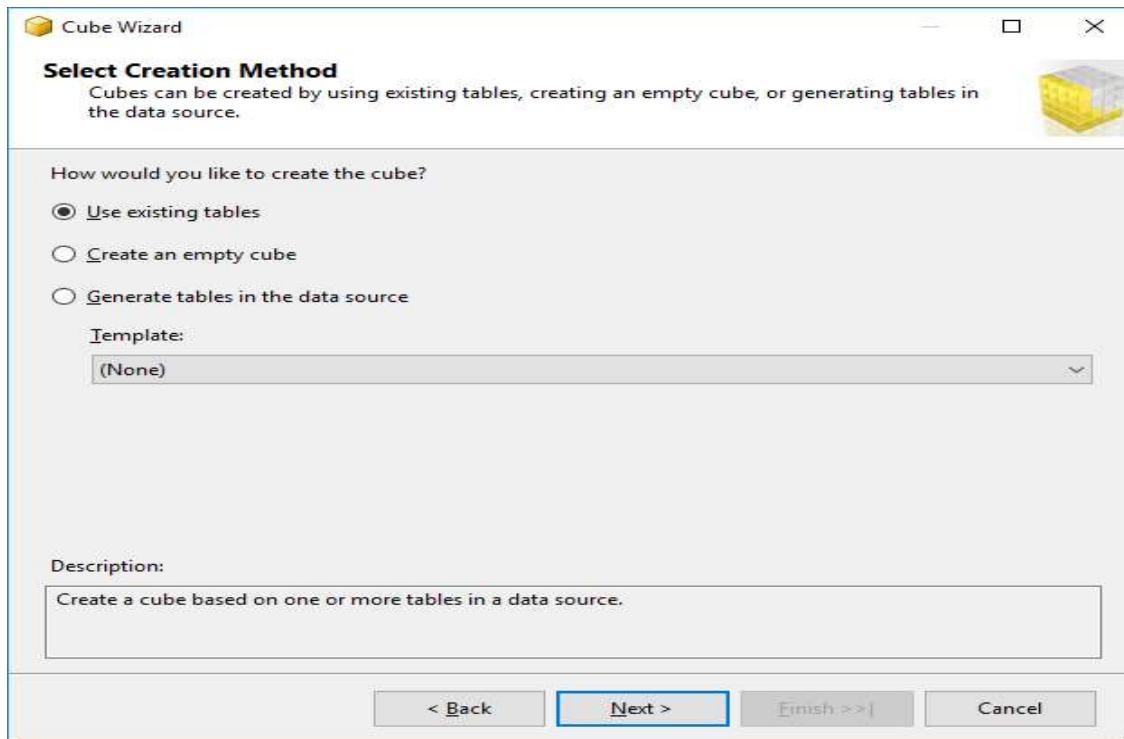
Solution :

- 1. Right click on Cubes -> New Cube.**

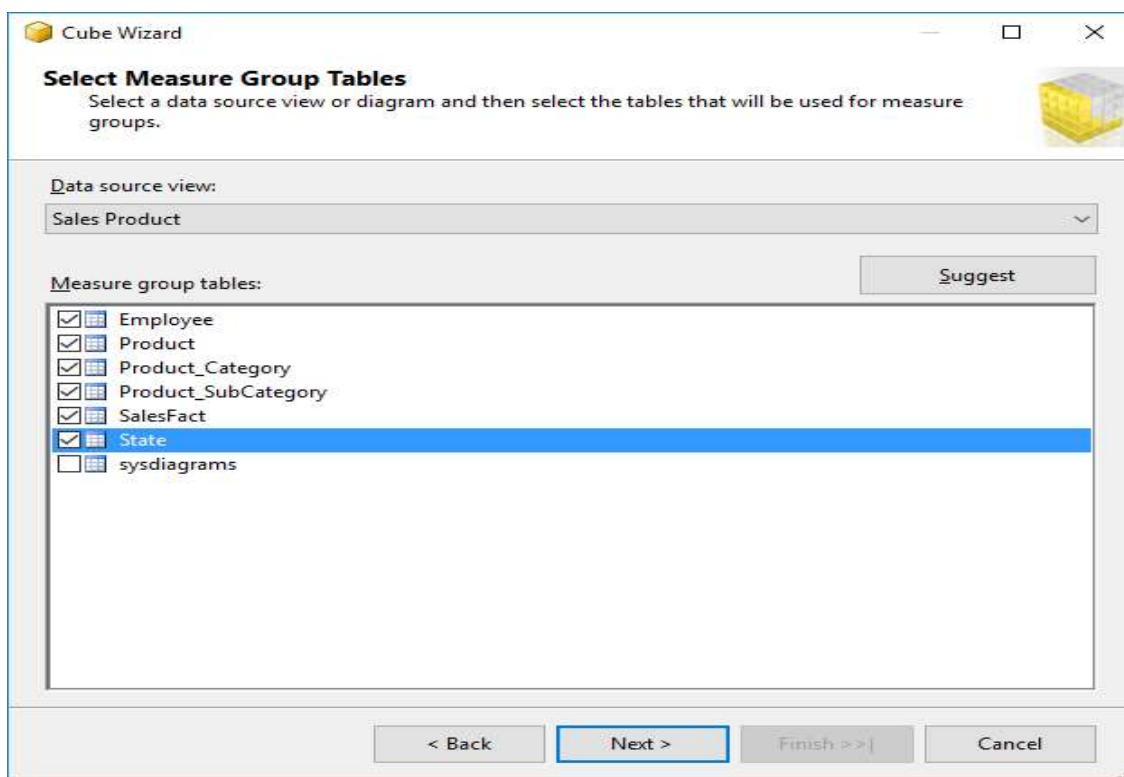


Click on Next.

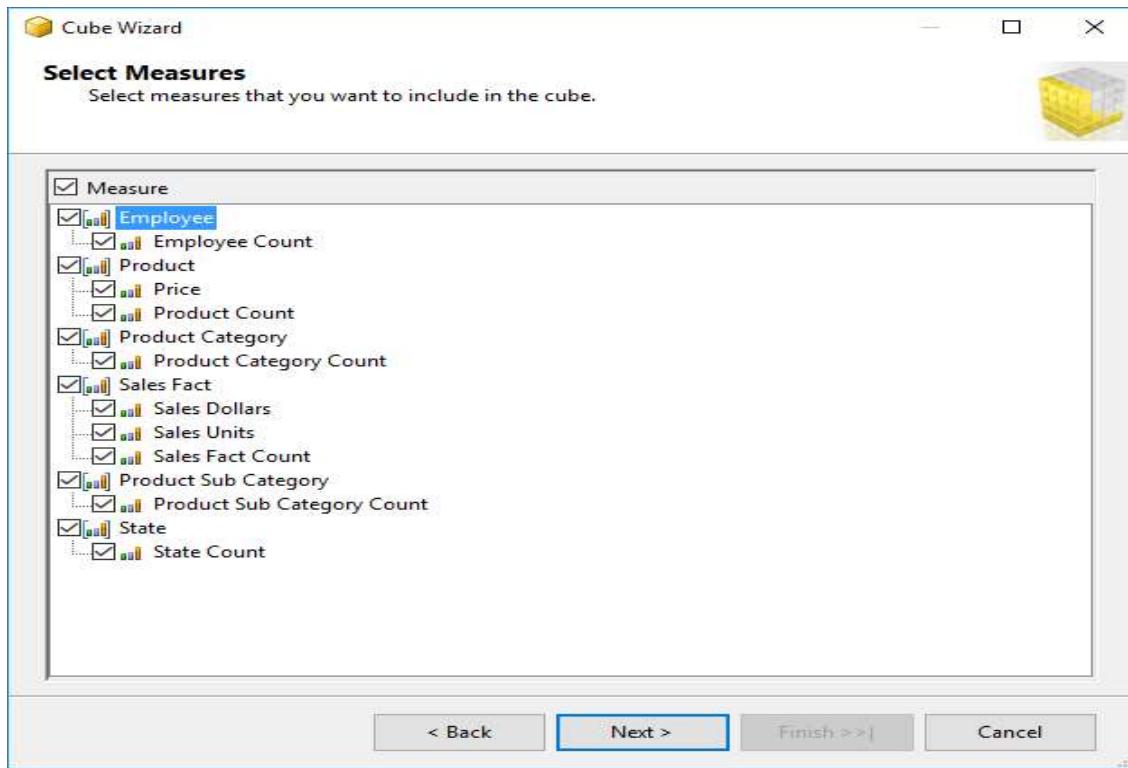
- 2. Select First option “Use existing tables”. Click on Next.**



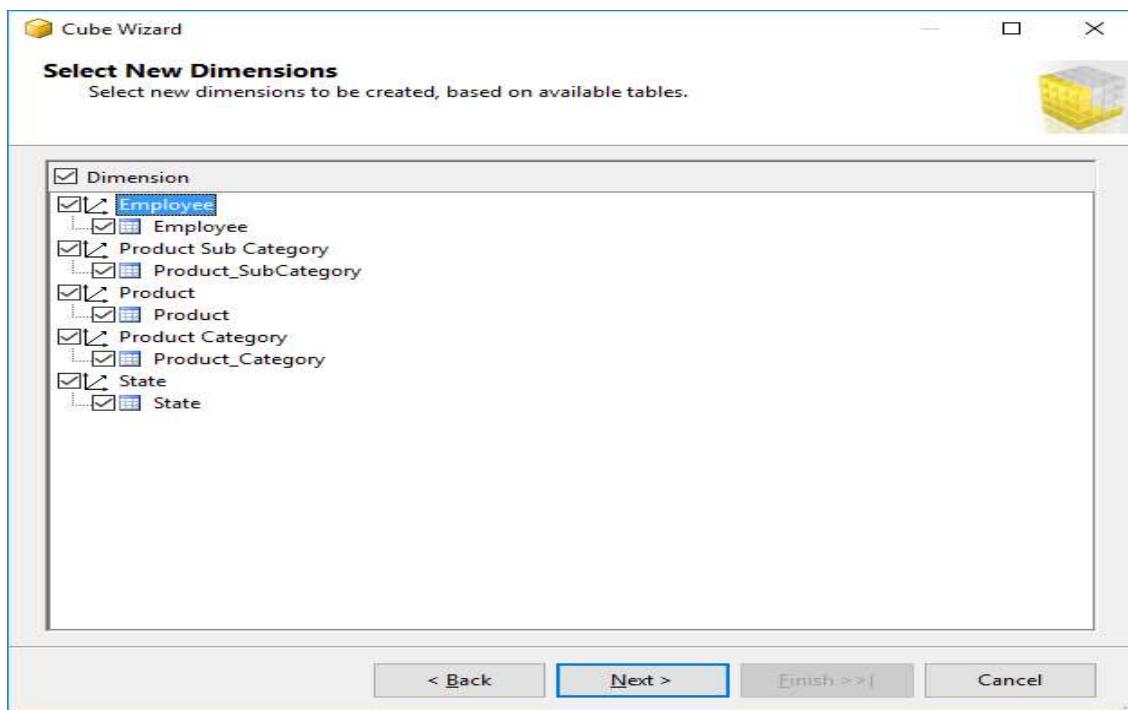
3. Select Data Source View as “Sales Product” and Select all the tables.



Click on Next.

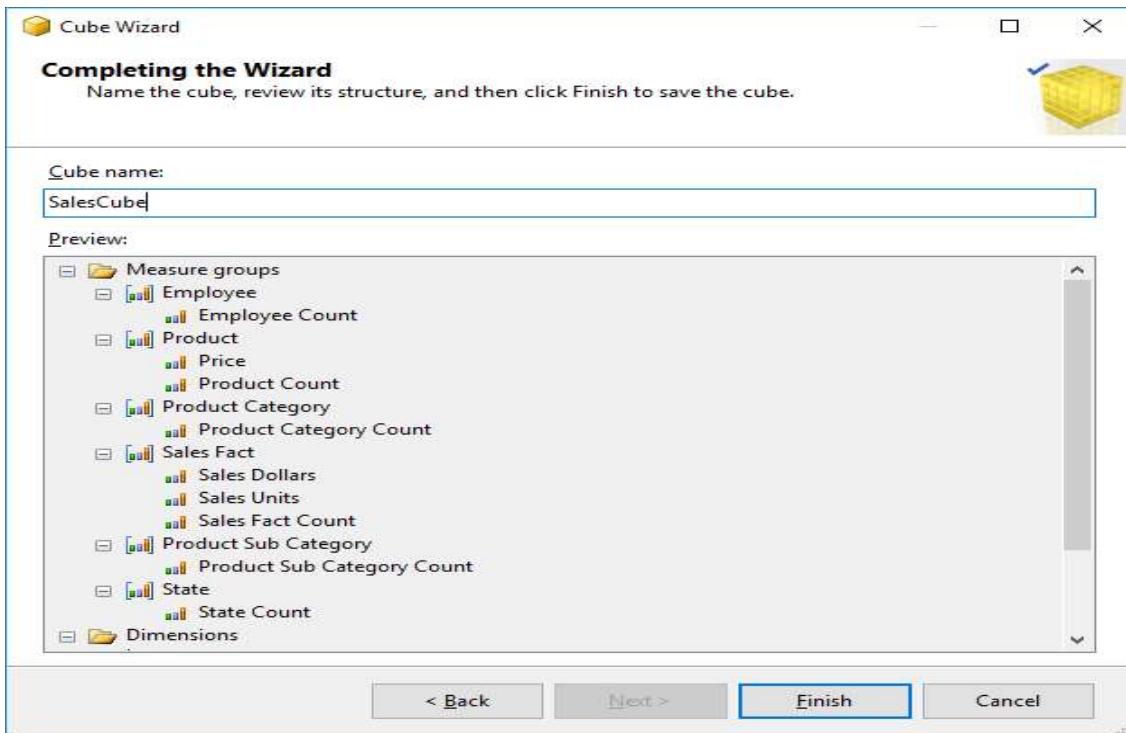


Click on Next.



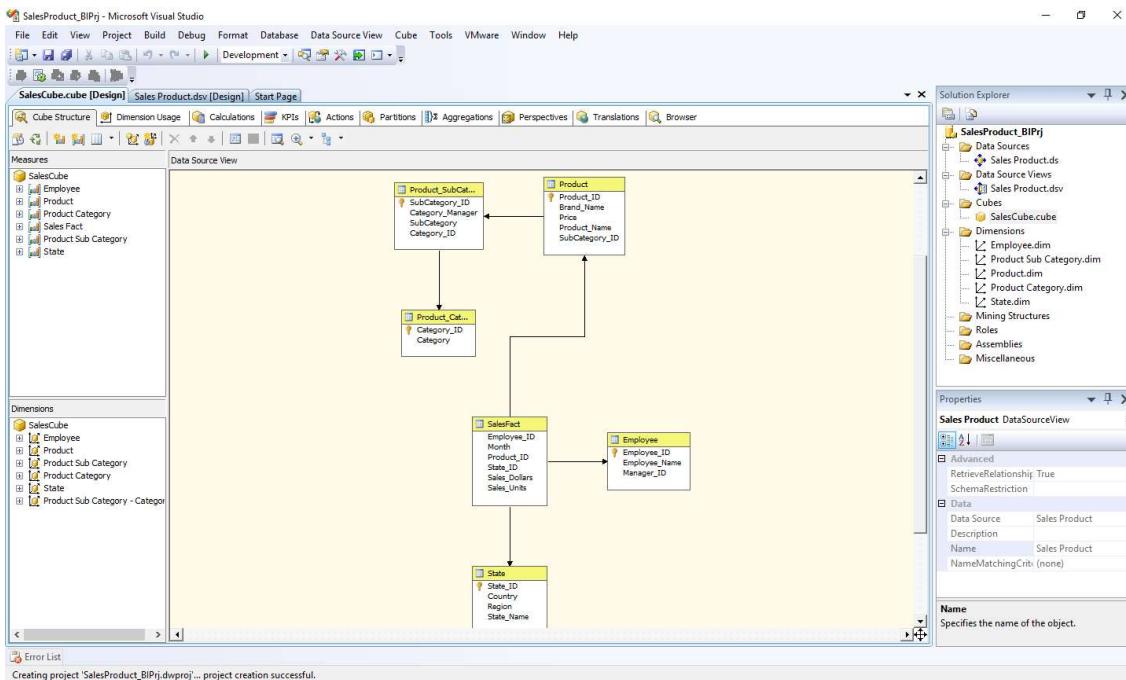
Click on Next.

4. Name Cube as “SalesCube”.

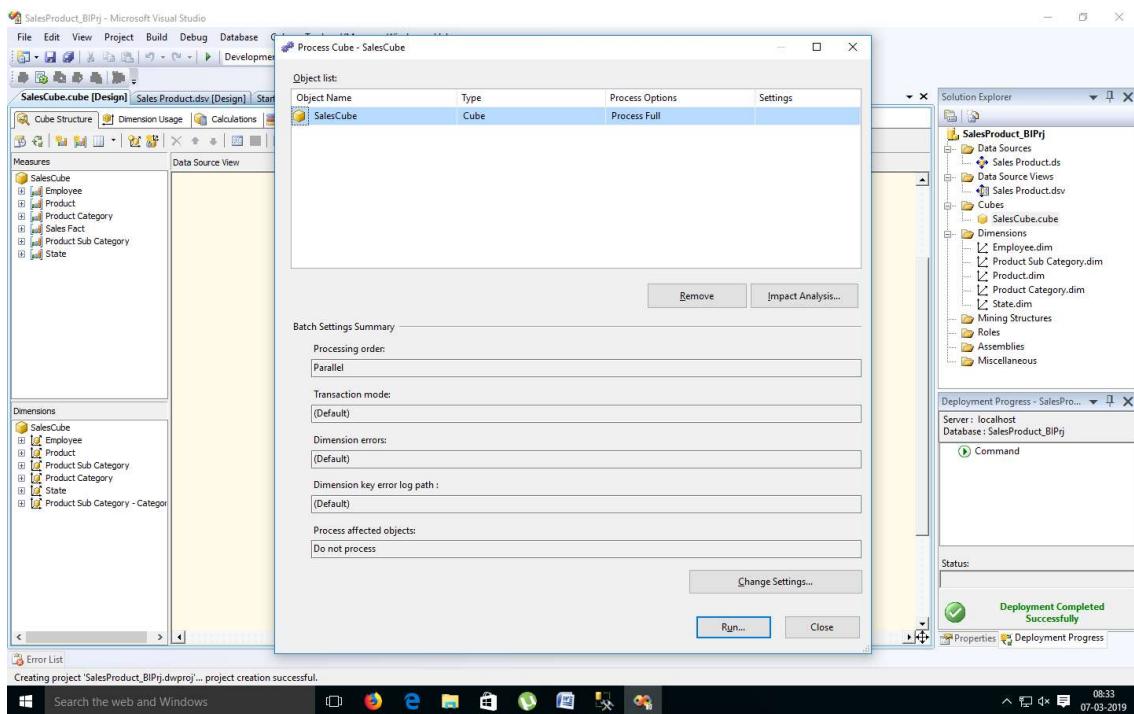
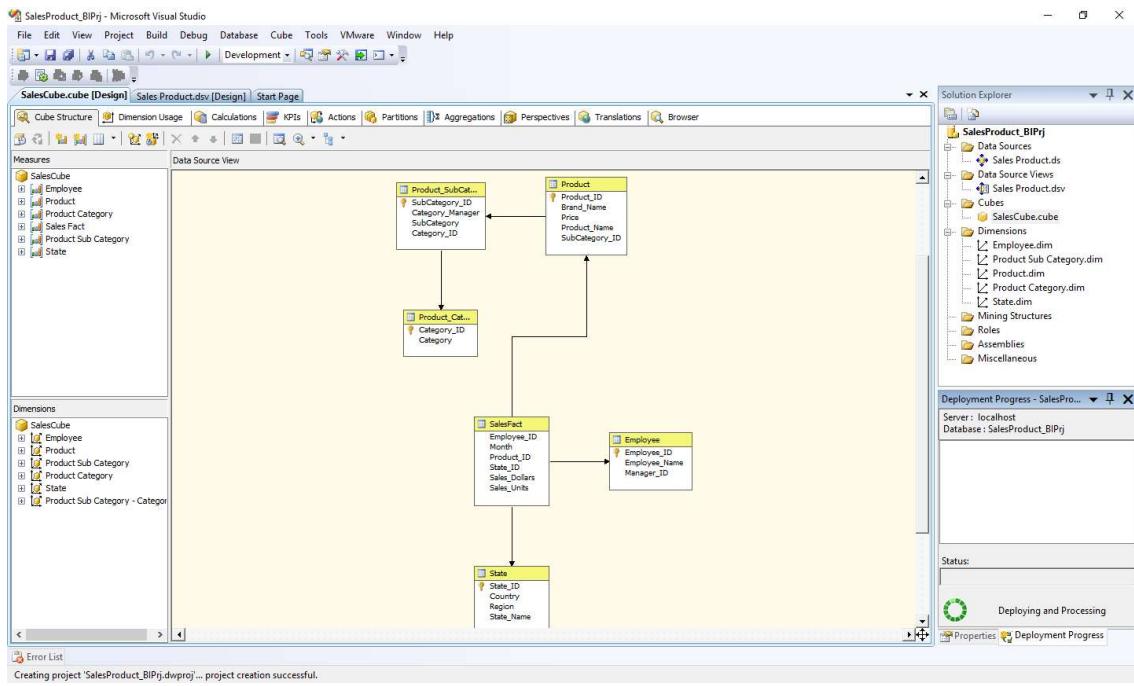


Click on Finish.

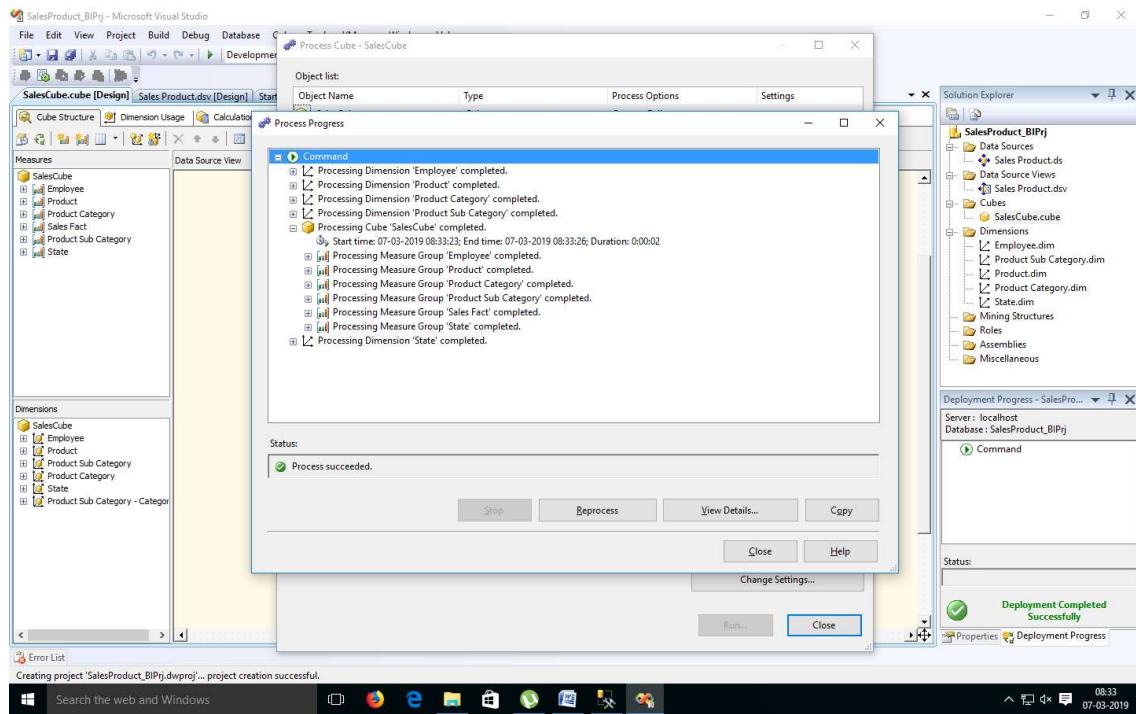
5. Finally, we will get the Cube View as well Dimensions View like :



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



7. Click on Run.

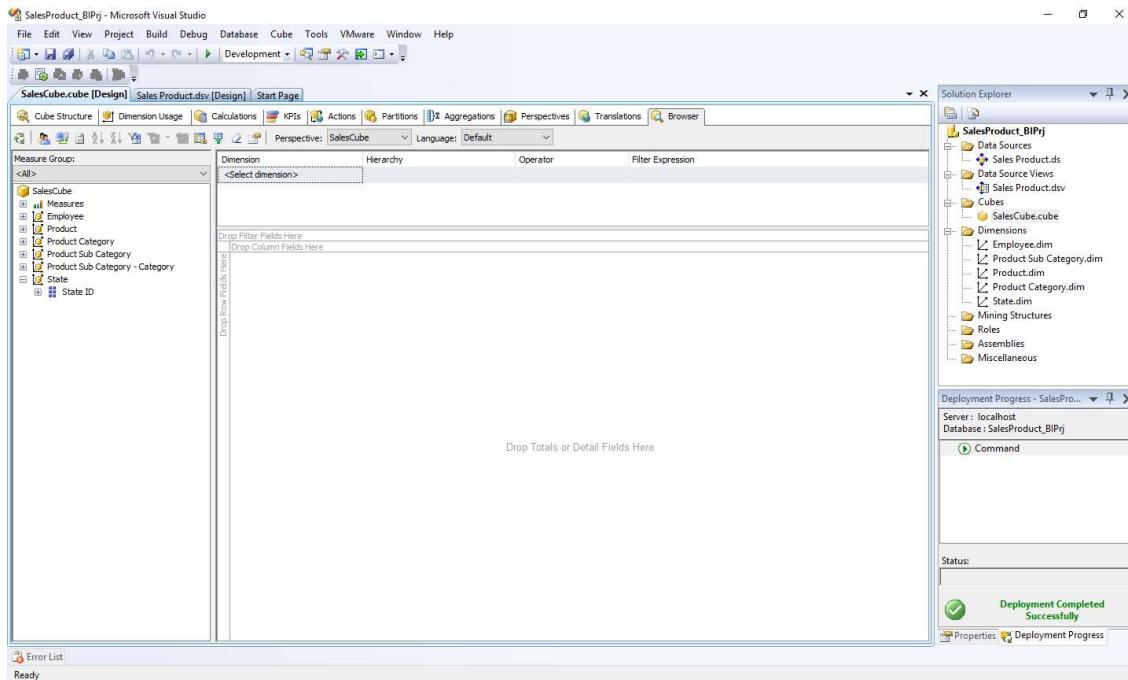


Practical No 5

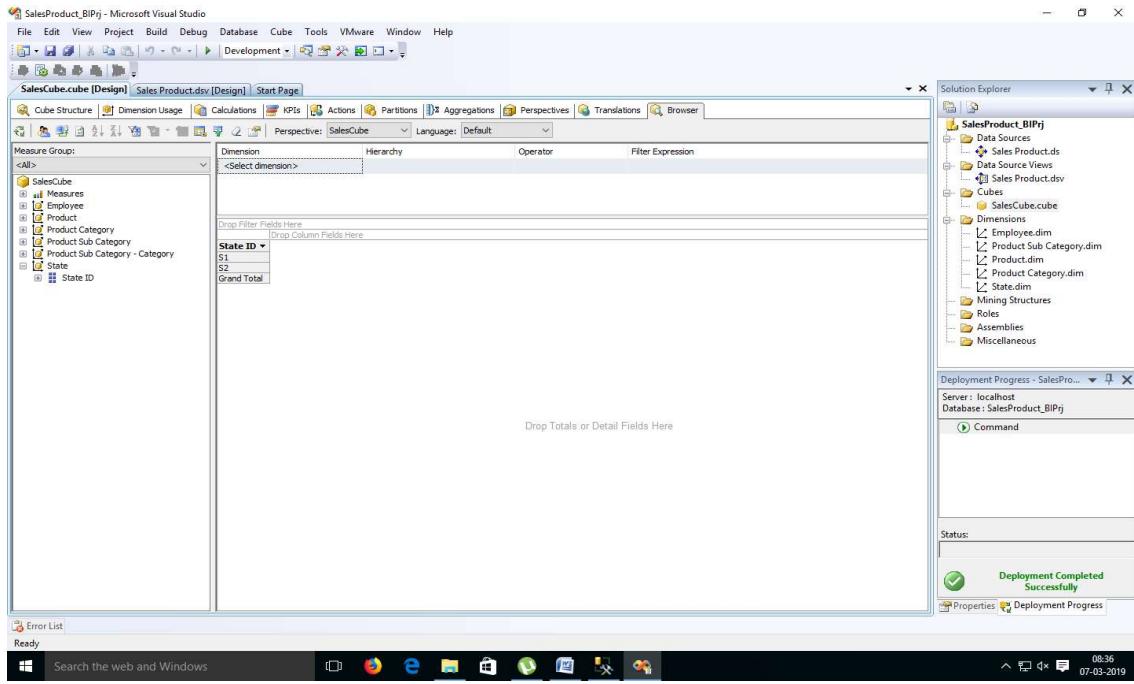
Aim : View cube data in multidimensional Format.

Solution :

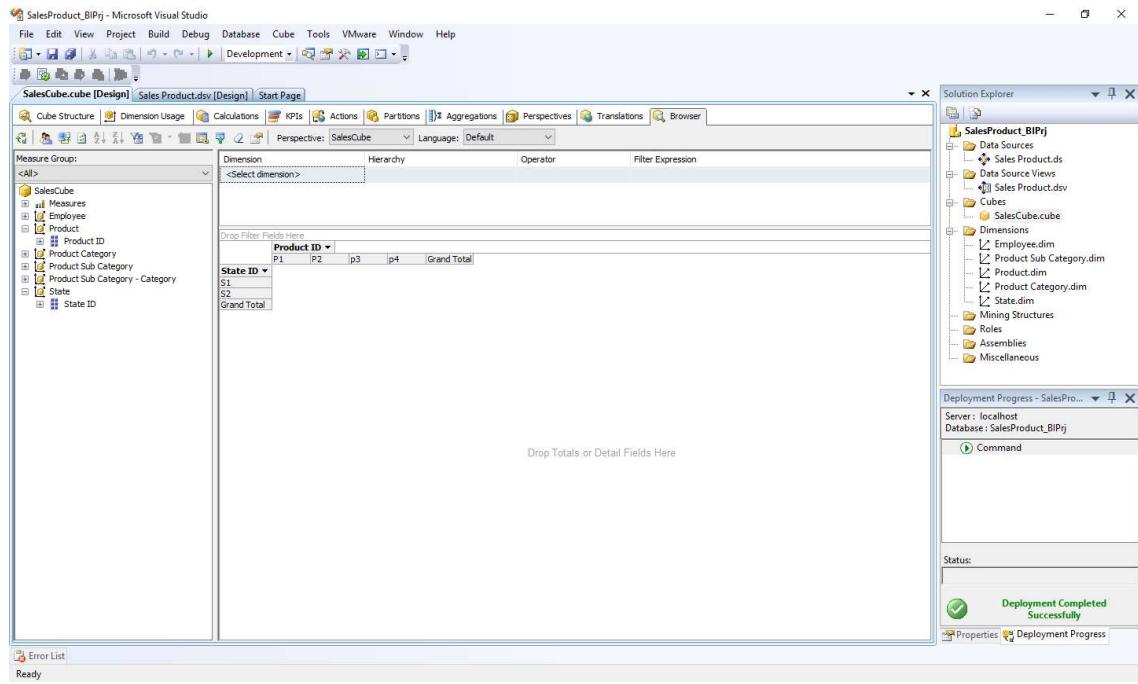
1. Double Click on “SalesCube”. Go to the “Browser” Tab.



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.

The screenshot shows the Microsoft Visual Studio interface for a BIP project named SalesProduct_BIPrj. The main window displays the SalesCube cube design. On the left, the cube structure is shown with various dimensions like SalesCube, Measures, Employee, Product, etc. In the center, a data grid shows sales data with columns for Product ID, State ID, Sales Dollars, Sales Units, and Grand Total. The Solution Explorer on the right lists the project structure, including Data Sources, Data Source Views, Cubes, Dimensions, and other components. The Deployment Progress window indicates a successful deployment to localhost.

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

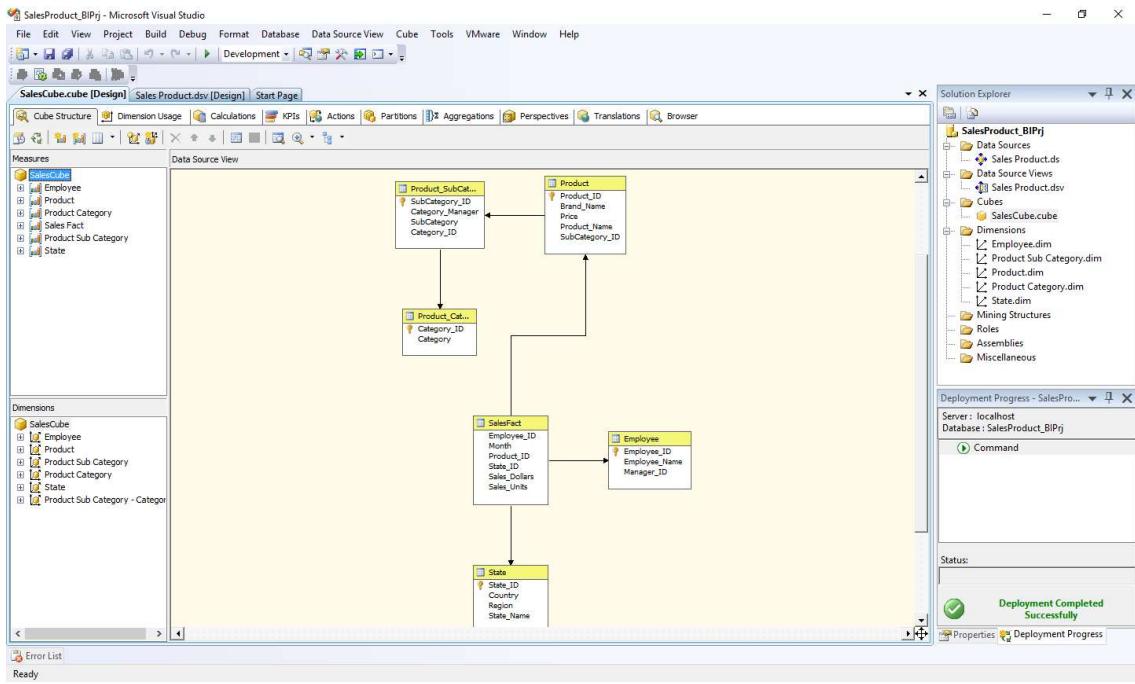
This screenshot shows the same Microsoft Visual Studio interface after the 'Employee_ID' dimension has been added to the Row Area of the SalesCube cube. The data grid now includes the Employee_ID column in the row headers, showing sales data for specific employees (E1, E3, etc.). The rest of the interface, including the cube structure, dimensions, and deployment status, remains the same as in the previous screenshot.

Practical No 6

Aim : Working with measures in the cube.

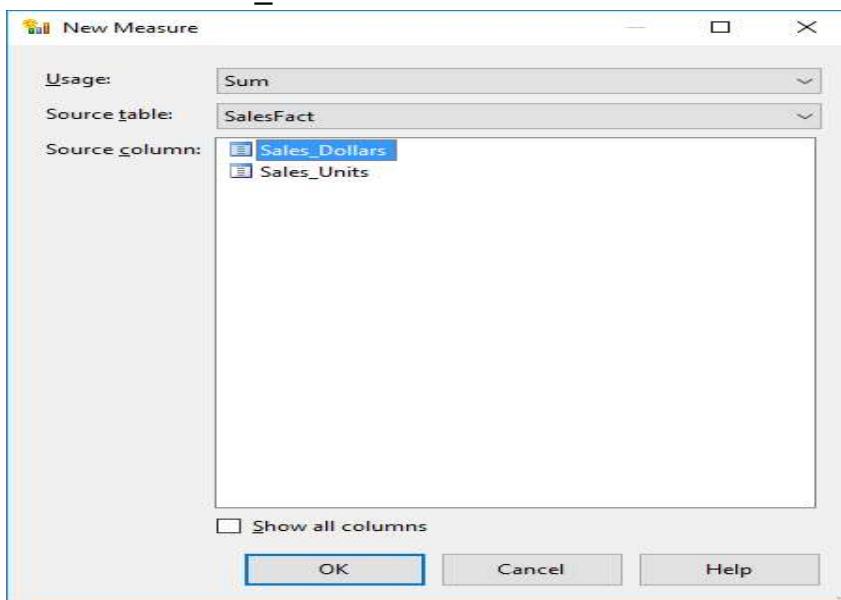
Solution :

1. Double click on ‘SalesCube’. Go to cube structure.



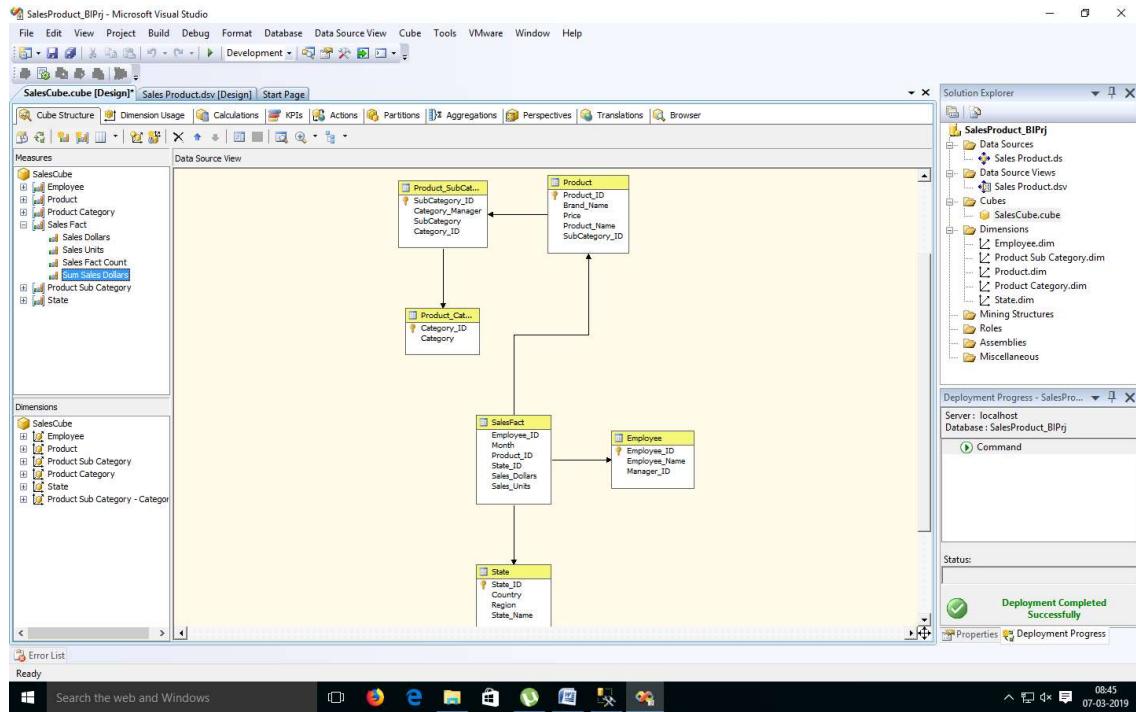
2. Right click on SalesCube -> New Measure.

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.

SalesProduct_BIPrj - Microsoft Visual Studio

SalesCube.cube [Design] | Sales Product.csv [Design] | Start Page

Cube Structure | Dimension Usage | Calculations | KPIs | Actions | Partitions | Aggregations | Perspectives | Translations | Browser

Perspective: SalesCube | Language: Default

Measure Group Dimension Hierarchy Operator Filter Expression

<All> <Select dimension>

Drop Filter Fields Here

	Product ID			Grand Total									
State ID	Employee ID	P1	P2	p3	Sales Dollars	Sales Units	Sum Sales Dc	Sales Dollars	Sales Units	Sum Sales Dc	Sales Dollars	Sales Units	Sum Sales Dc
E1		20000	200	20000						20000	200	20000	
E2					45000	3	45000	45000	3	45000	3	45000	
Total		20000	200	20000	45000	3	45000	45000	3	45000	3	45000	
S2	E2				80000	2	80000	80000	2	80000	2	80000	
Total		20000	200	20000	80000	2	80000	45000	3	45000	145000	205	145000
Grand Total		20000	200	20000	80000	2	80000	45000	3	45000	145000	205	145000

SalesProduct_BIPrj

- Data Sources
 - Sales Product.ds
 - Data Source Views
 - Sales Product.csv
- Cubes
 - SalesCube.cube
- Dimensions
 - Employee.dim
 - Product Sub Category.dim
 - Product.dim
 - Product Category.dim
 - State.dim
- Mining Structures
- Roles
- Assemblies
- Miscellaneous

Deployment Progress - SalesPro... Deployment Progress

Server : localhost Database : SalesProduct_BIPrj

Command

Status:

✓ Deployment Completed Successfully

Properties Deployment Progress

Error List

Ready

Search the web and Windows

08:47 07-03-2019

Practical No 7

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

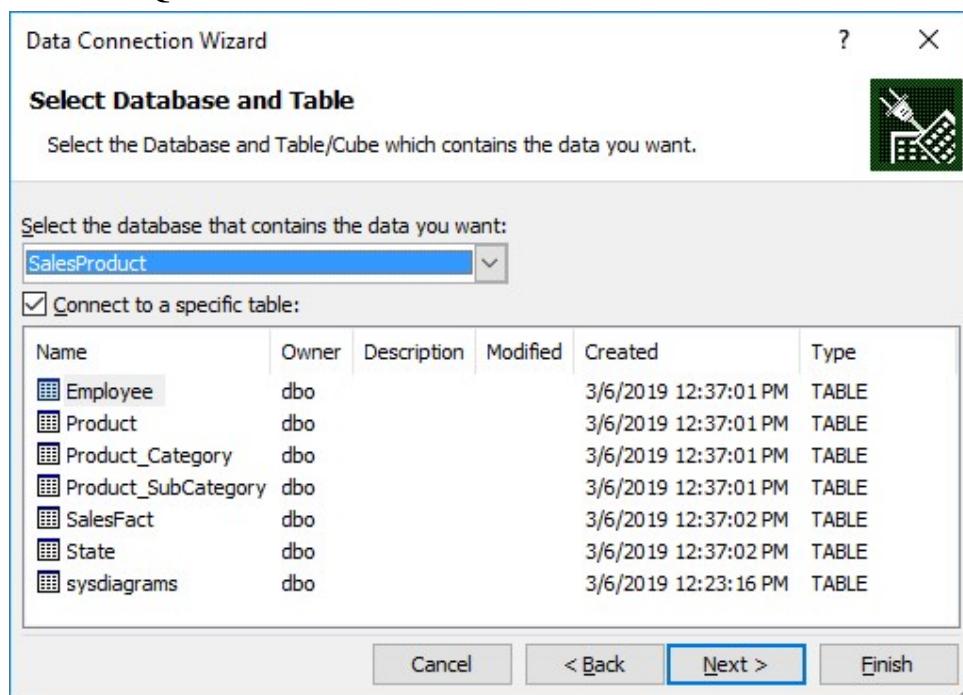
Solution :

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

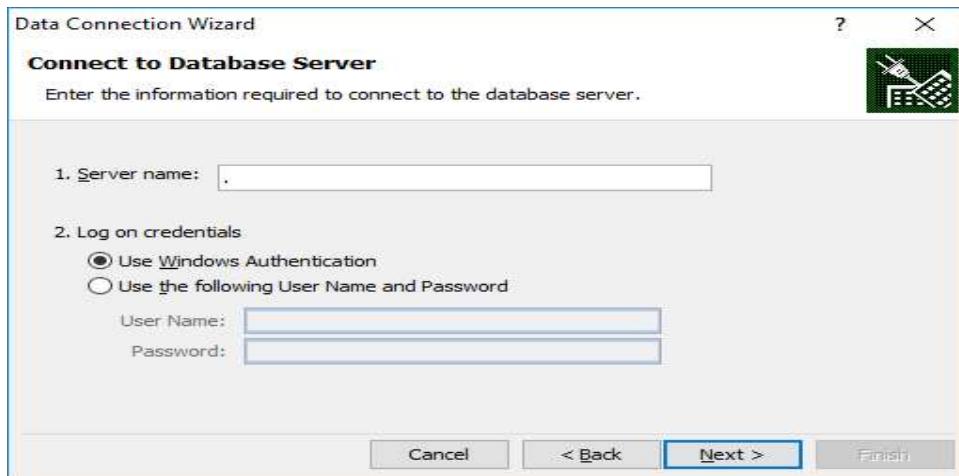


Click on Next.

Choose SQL Database -> “SalesProduct”

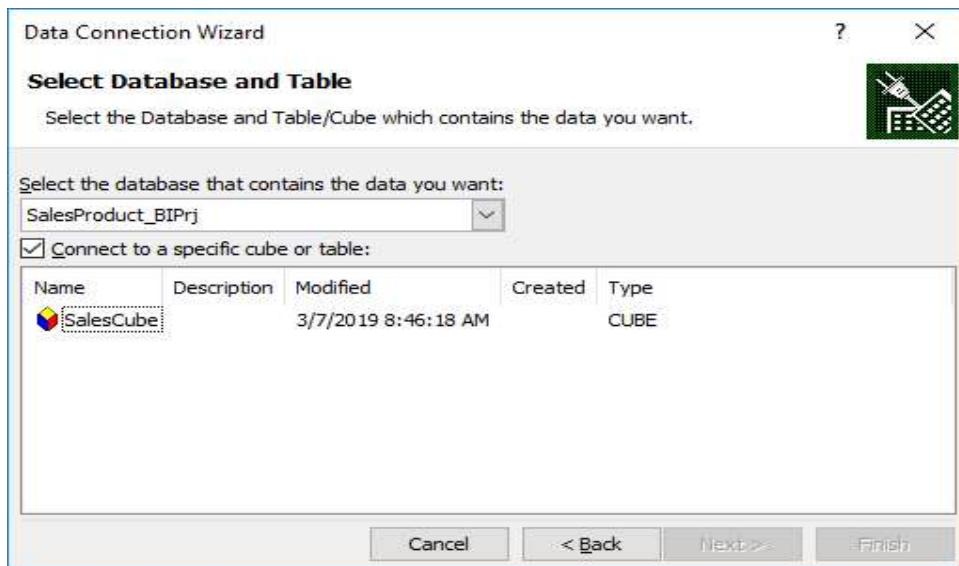


2.2. From Analysis Services -> Type Server name as “.”

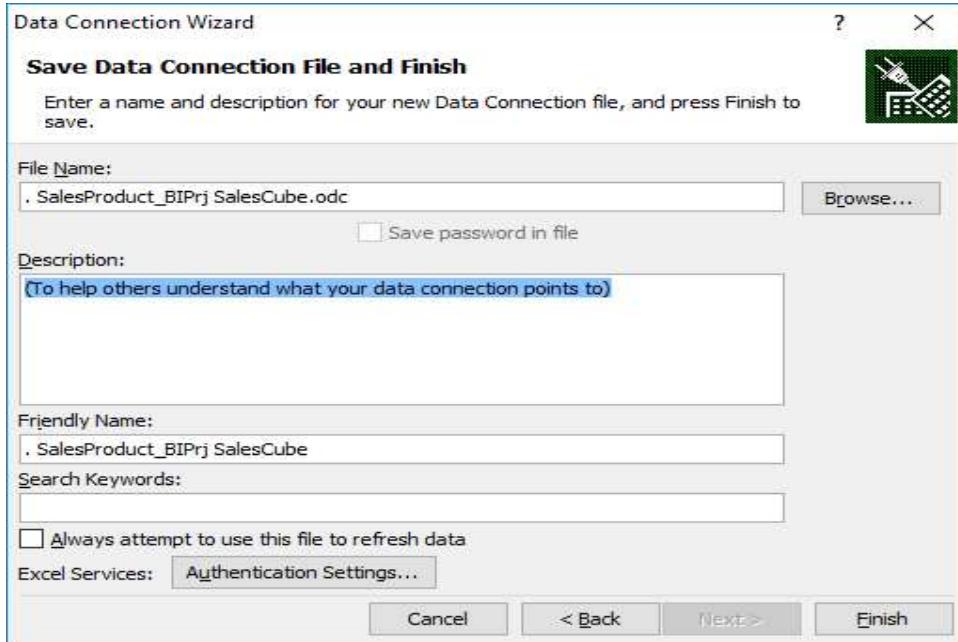


Click on Next.

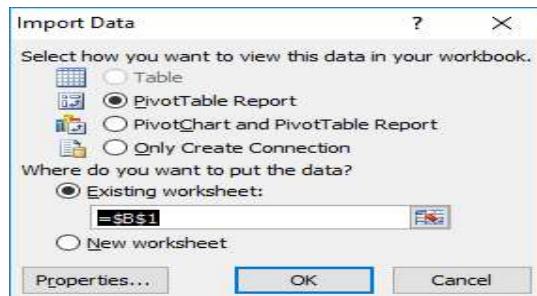
Choose Analysis Database as “SalesProduct_BIPrj”. Click on Next.



Click on OK.

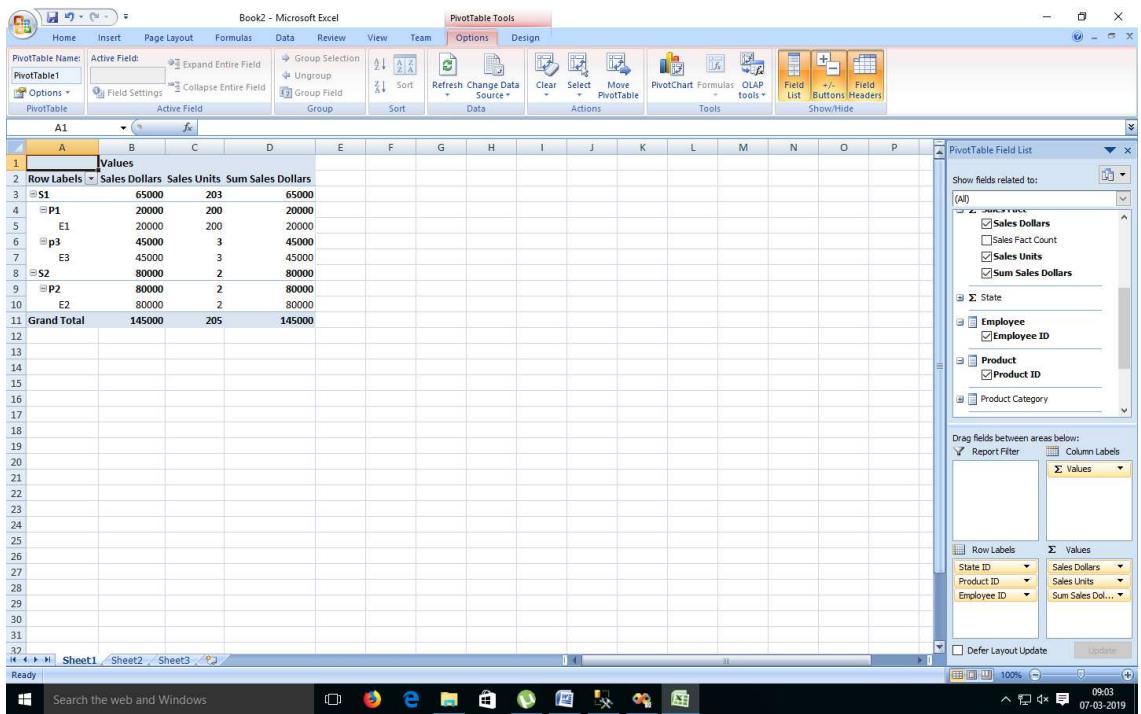


Click on Finish.

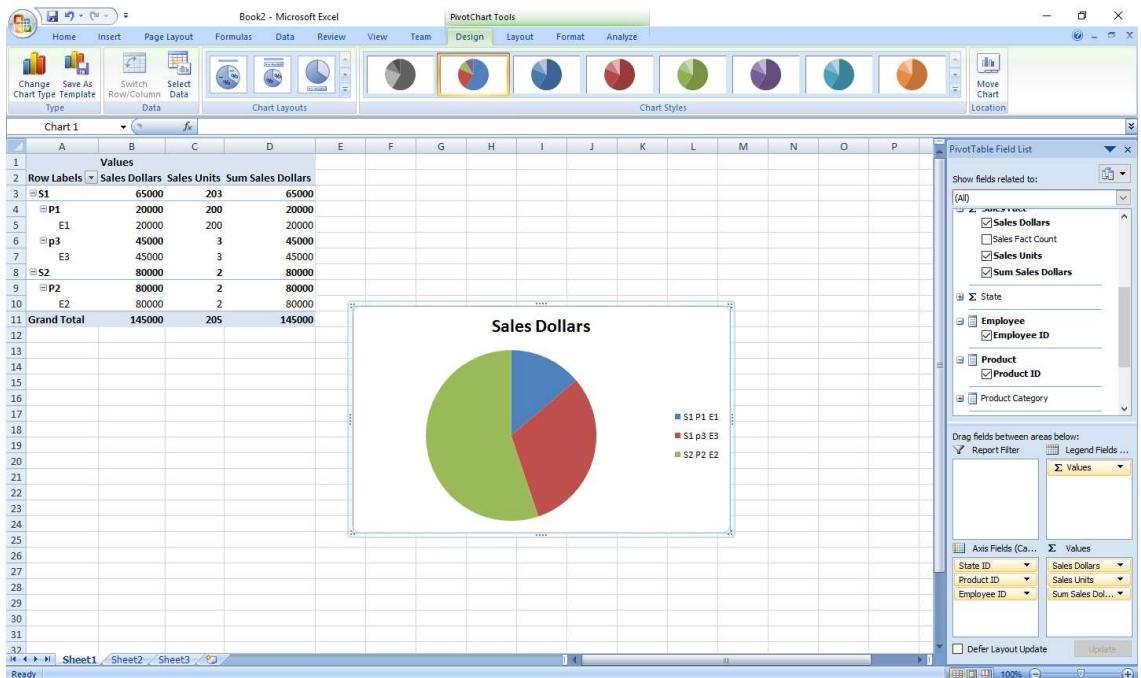


Click on OK.

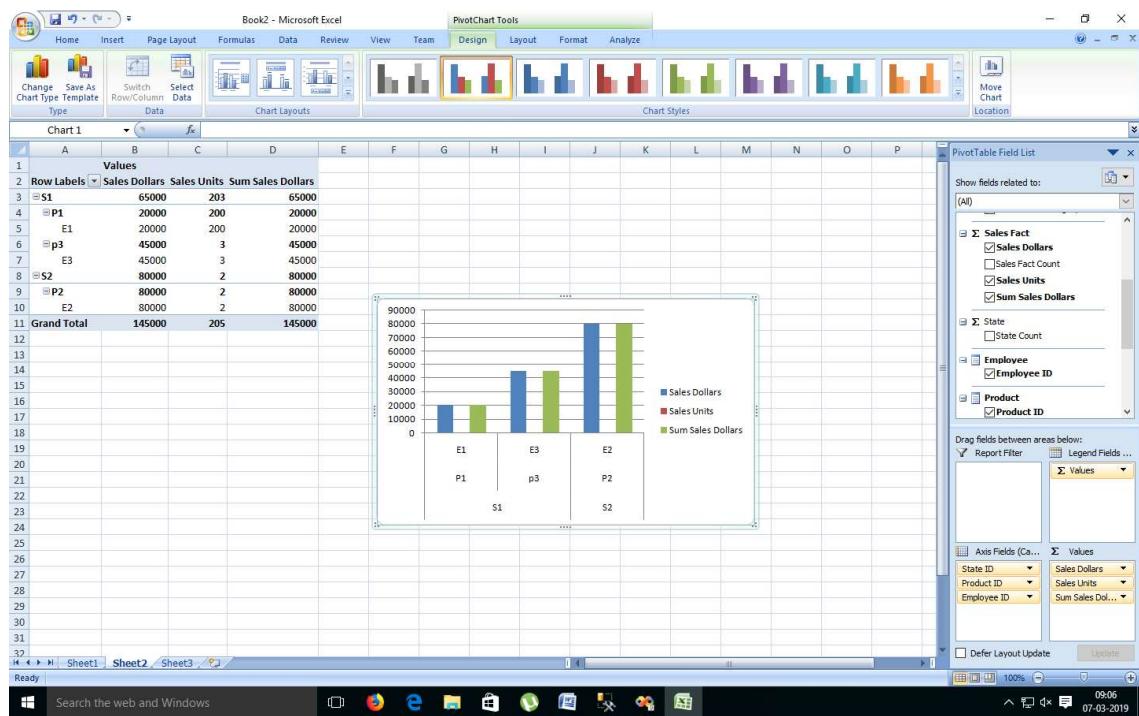
3. Select State_ID, Product_ID, Employee_ID and Measures as Sales_Dollars, Sales_Units and Sum Sales Dollars.



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 database named 'master' containing various system and user-defined objects like databases, tables, and stored procedures. The central Results pane displays the output of a T-SQL query:

```
SQLQuery4.sql (LSC-37\admin(55)) [SQLQuery4sql - (SG-37\admin(62))] ****
***** Script for SelectTopNRows command from SSMS *****
SELECT [Employee_ID]
      ,[Employee_Name]
      ,[Manager_ID]
 FROM [SalesProduct].[dbo].[Employee]
```

The results pane shows the following 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." The Properties pane on the right shows connection details such as the server name (local), connection time, and session tracing information. The Output pane below the results pane is currently empty.

**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 a connection to '(local) (SQL Server 10.50.1600 - BSC-37\...\'. The 'master' database is selected. The 'Tables' node under 'SalesProduct' contains the 'Product' table. A query window titled 'SQLQuery7.sql - ...SC-37\admin (58)' is open, displaying the following T-SQL code:

```
***** Script for SelectTopNRows command from SSMS *****
SELECT [Product_ID]
      ,[Brand_Name]
      ,[Price]
      ,[Product_Name]
      ,[SubCategory_ID]
   FROM [SalesProduct].[dbo].[Product]
```

The results pane shows the following 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 shows connection 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]**

```

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

```

	State_ID	Country	Region	State_Name
1	S1	India	Mumbai	Maharastra
2	S2	India	GN	Gujrat

Query executed successfully.

Properties pane:

- Aggregate Status: Connection fail, Elapsed time: 00:00:00.031, Finish time: 08-03-2019 09:30:32, Name: (local), Rows returned: 2, Start time: 08-03-2019 09:30:32, State: Open
- Connection: Connection nnn (local) (BSC-37\admin)
- Connection Details: Connection elap 00:00:00.031, Connection finis 08-03-2019 09:30:32, Connection row: 2, Connection start 08-03-2019 09:30:32, Connection stat: Open, Display name: (local), Login name: BSC-37\admin, Server name: (local), Server version: 10.50.1600, Session Tracing l, SPID: 53
- Name: The name of the connection.

3.4. SELECT

```

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

```

```

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

```

	SubCategory_ID	Category_Manager	SubCategory	Category_ID
1	SC1	Level1	Accessories	C1
2	SC2	Level2	Electronic	C2

Query executed successfully.

Properties pane:

- Aggregate Status: Connection fail, Elapsed time: 00:00:00.038, Finish time: 08-03-2019 09:32:37, Name: (local), Rows returned: 2, Start time: 08-03-2019 09:32:37, State: Open
- Connection: Connection nnn (local) (BSC-37\admin)
- Connection Details: Connection elap 00:00:00.038, Connection finis 08-03-2019 09:32:37, Connection row: 2, Connection start 08-03-2019 09:32:37, Connection stat: Open, Display name: (local), Login name: BSC-37\admin, Server name: (local), Server version: 10.50.1600, Session Tracing l, SPID: 60
- Name: The name of the connection.

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

```

Microsoft SQL Server Management Studio
File Edit View Query Debug Tools Window Community Help
New Query | Databases Object Explorer Results Messages Properties
Connect > (local) SQL Server 10.50.1600 - BSC-A
Object Explorer
Databases System Databases ReportServer ReportServerTempDB Sales SalesInformation SalesProduct Database Diagrams Tables System Tables dbo.Employee dbo.Product dbo.Product_Categor dbo.Product_SubCat dbo.SalesFact dbo.State Views Synonyms Programmability Service Broker Storage Security Security
Results
SELECT [Category_ID]
      [Category]
  FROM [SalesProduct].[dbo].[Product_Category]
  
```

Properties

Current connection parameters

Aggregate Status

Connection fail.

Elapsed time 00:00:00.031

Finish time 08-03-2019 09:33:38

Name (local)

Rows returned 2

Start time 08-03-2019 09:33:38

State Open

Connection

Connection name (local) (BSC-37\admin)

Connection Details

Connection elap 00:00:00.031

Connection finis 08-03-2019 09:33:38

Connection row 2

Connection start 08-03-2019 09:33:38

Connection stat Open

Display name (local)

Login name BSC-37\admin

Server name (local)

Server version 10.50.1600

Session Tracing l

SPID 61

Name The name of the connection.

Query executed successfully. (local) (10.50 RTM) | BSC-37\admin (61) | master | 00:00:00 | 2 rows

Output

Ready

Ln 4 Col 47 Ch 47 INS

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

```

Microsoft SQL Server Management Studio
File Edit View Query Debug Tools Window Community Help
New Query | Databases Object Explorer Results Messages Properties
Connect > (local) SQL Server 10.50.1600 - BSC-A
Object Explorer
Databases System Databases ReportServer ReportServerTempDB Sales SalesInformation SalesProduct Database Diagrams Tables System Tables dbo.Employee dbo.Product dbo.Product_Categor dbo.Product_SubCat dbo.SalesFact dbo.State Views Synonyms Programmability Service Broker Storage Security Security
Results
SELECT [Employee_ID]
      , [Month]
      , [Product_ID]
      , [State_ID]
      , [Sales_Dollars]
      , [Sales_Units]
  FROM [SalesProduct].[dbo].[SalesFact]
  
```

Properties

Current connection parameters

Aggregate Status

Connection fail.

Elapsed time 00:00:00.052

Finish time 08-03-2019 09:34:39

Name (local)

Rows returned 3

Start time 08-03-2019 09:34:39

State Open

Connection

Connection name (local) (BSC-37\admin)

Connection Details

Connection elap 00:00:00.052

Connection finis 08-03-2019 09:34:39

Connection row 3

Connection start 08-03-2019 09:34:39

Connection stat Open

Display name (local)

Login name BSC-37\admin

Server name (local)

Server version 10.50.1600

Session Tracing l

SPID 65

Name The name of the connection.

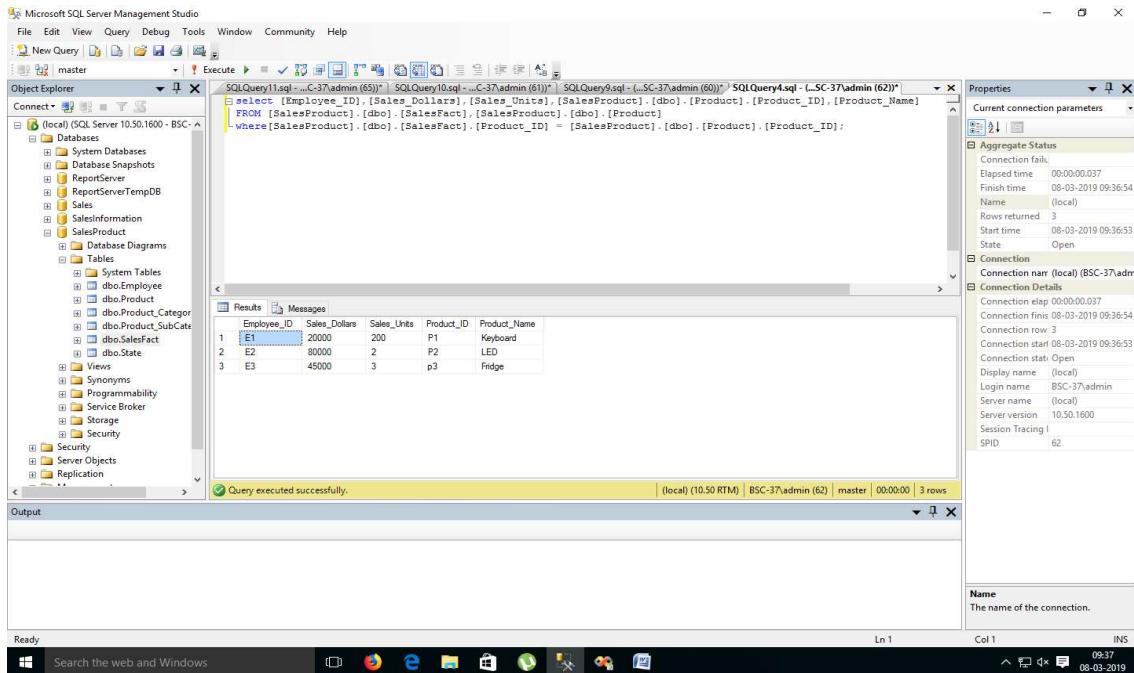
Query executed successfully. (local) (10.50 RTM) | BSC-37\admin (65) | master | 00:00:00 | 3 rows

Output

Ready

Ln 8 Col 40 Ch 40 INS

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];`



```
Microsoft SQL Server Management Studio
File Edit View Query Debug Tools Window Community Help
New Query Connect Object Explorer Execute File Edit View Query Debug Tools Window Community Help
Object Explorer
Connect (local) (SQL Server 10.50.1600 - BSC-37\)
Databases System Databases Database Snapshots ReportServer ReportServerTempDB Sales SalesInformation SalesProduct Database Diagrams Tables System Tables dbo.Employee dbo.Product dbo.Product_Categor dbo.Product_SubCateg dbo.SalesFact dbo.State Views Synonyms Programmability Service Broker Storage Security Security Server Objects Replication
Results Messages
Employee_ID Sales_Dollars Sales_Units Product_ID Product_Name
1 E1 20000 200 P1 Keyboard
2 E2 80000 2 P2 LED
3 E3 45000 3 p3 Fridge
Query executed successfully.
Properties Current connection parameters
Aggregate Status
Connection fail.
Elapsed time 00:00:00.037
Finish time 08-03-2019 09:36:54
Name (local)
Rows returned 3
Start time 08-03-2019 09:36:53
State Open
Connection Connection name (local) (BSC-37\admin)
Connection Details
Connection elap 00:00:00.037
Connection finis 08-03-2019 09:36:54
Connection row: 3
Connection start 08-03-2019 09:36:53
Connection stat: Open
Display name (local)
Login name BSC-37\admin
Server name (local)
Server version 10.50.1600
Session Tracing l
SPID 62
Name
The name of the connection.
Ln 1 Col 1 INS
09:37 08-03-2019
Output
Ready
Search the web and Windows

```

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];`

```

Microsoft SQL Server Management Studio
File Edit View Query Debug Tools Window Community Help
New Query | Execute | Run | Stop | Break | Refresh | Stop All | Break All | Properties | Results | Messages | Output
Object Explorer
Connect > (local) SQL Server 10.50.1600 - BSC->
  Databases
    System Databases
    Database Snapshots
    ReportServer
    ReportServerTempDB
    Sales
    SalesInformation
    SalesProduct
      Database Diagrams
      Tables
        System Tables
        dbo.Employee
        dbo.Product
        dbo.Product_Categor
        dbo.Product_SubCate
        dbo.SalesFact
        dbo.State
      Views
      Synonyms
      Programmability
      Service Broker
      Storage
      Security
    Security
    Server Objects
    Replication
Results
Messages
Employee_ID Sales_Dollars Sales_Units State_ID State_Name
1   E1       20000     200      S1      Maha
2   E2       80000     2       S2      Gujarat
3   E3       45000     3       S1      Maha
Query executed successfully.

```

Properties

Current connection parameters

Aggregate Status

Connection fail.

Elapsed time 00:00:00.031

Finish time 08-03-2019 09:43:12

Name (local)

Rows returned 3

Start time 08-03-2019 09:43:12

State Open

Connection Connection name (local) (BSC-37\admi

Connection Details

Connection elap 00:00:00.031

Connection finis 08-03-2019 09:43:12

Connection row 3

Connection start 08-03-2019 09:43:12

Connection stat Open

Display name (local)

Login name BSC-37\admi

Server name (local)

Server version 10.50.1600

Session Tracing l

SPID 53

Name

The name of the connection.

**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];**

```

Microsoft SQL Server Management Studio
File Edit View Query Debug Tools Window Community Help
New Query | Execute | Run | Stop | Break | Refresh | Stop All | Break All | Properties | Results | Messages | Output
Object Explorer
Connect > (local) SQL Server 10.50.1600 - BSC->
  Databases
    System Databases
    Database Snapshots
    ReportServer
    ReportServerTempDB
    Sales
    SalesInformation
    SalesProduct
      Database Diagrams
      Tables
        System Tables
        dbo.Employee
        dbo.Product
        dbo.Product_Categor
        dbo.Product_SubCate
        dbo.SalesFact
        dbo.State
      Views
      Synonyms
      Programmability
      Service Broker
      Storage
      Security
    Security
    Server Objects
    Replication
Results
Messages
Employee_ID Employee_Name Sales_Dollars Sales_Units
1   E1       Rahul      20000     200
2   E2       Arkt       80000     2
3   E3       Sujt       45000     3
Query executed successfully.

```

Properties

Current connection parameters

Aggregate Status

Connection fail.

Elapsed time 00:00:00.046

Finish time 08-03-2019 09:48:24

Name (local)

Rows returned 3

Start time 08-03-2019 09:48:24

State Open

Connection Connection name (local) (BSC-37\admi

Connection Details

Connection elap 00:00:00.046

Connection finis 08-03-2019 09:48:24

Connection row 3

Connection start 08-03-2019 09:48:24

Connection stat Open

Display name (local)

Login name BSC-37\admi

Server name (local)

Server version 10.50.1600

Session Tracing l

SPID 53

Name

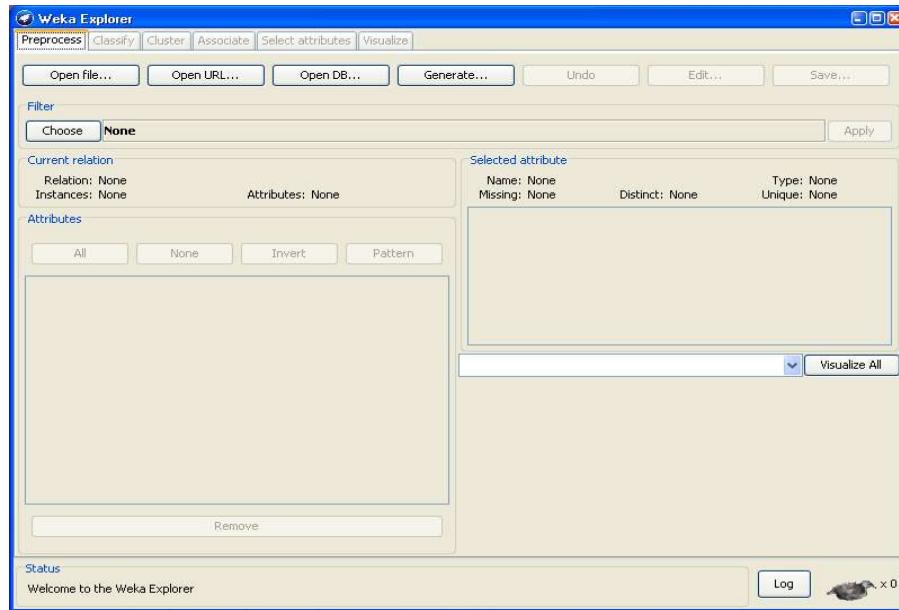
The name of the connection.

Practical No - 9

Aim : Data PreProcessing

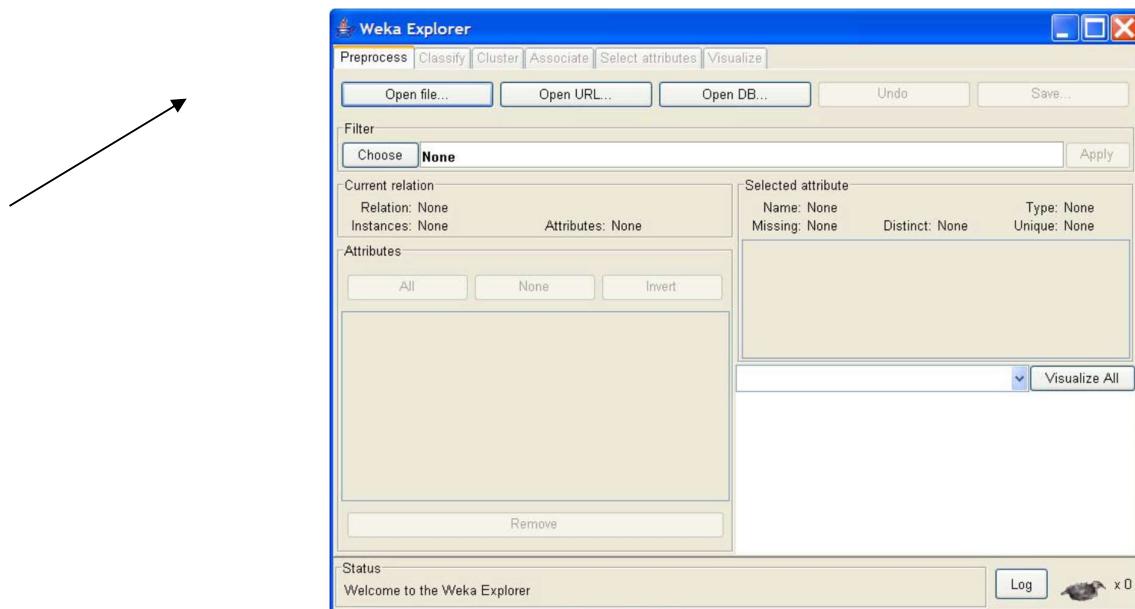
Solution :

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



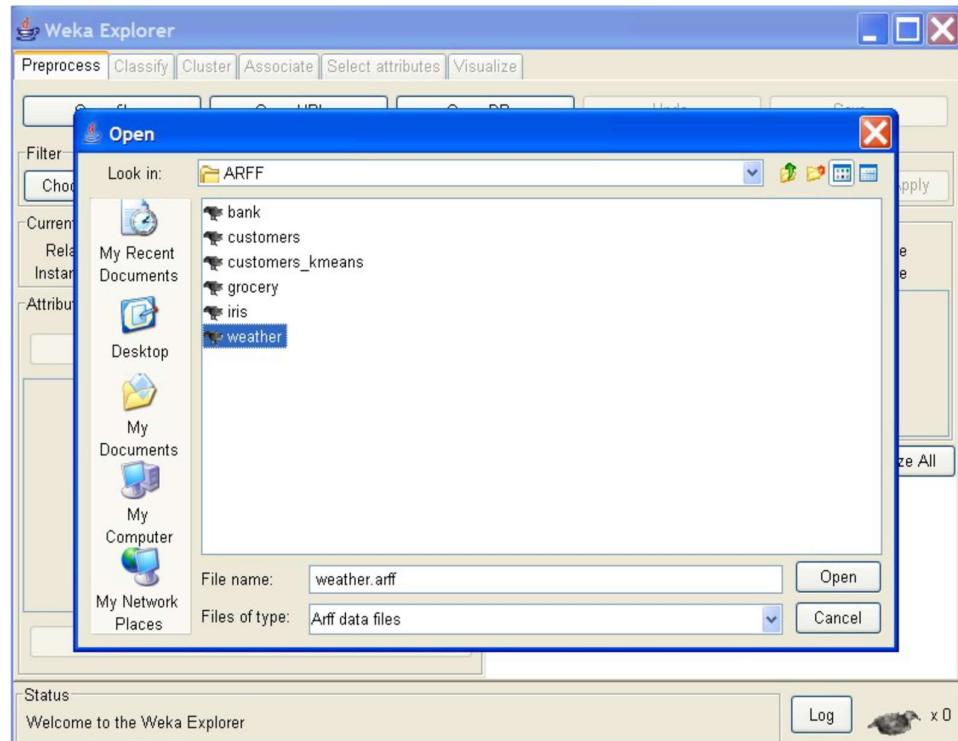
Opening file from a local file system

Click on ‘Open file...’ button



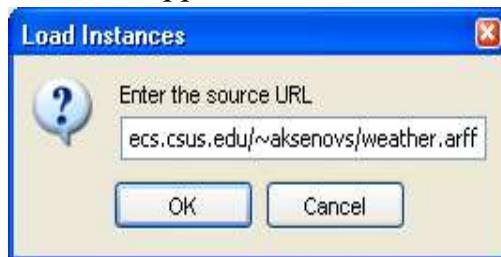
It brings up a dialog box allowing you to browse for the data file on the local file system, choose

“weather.arff” file.



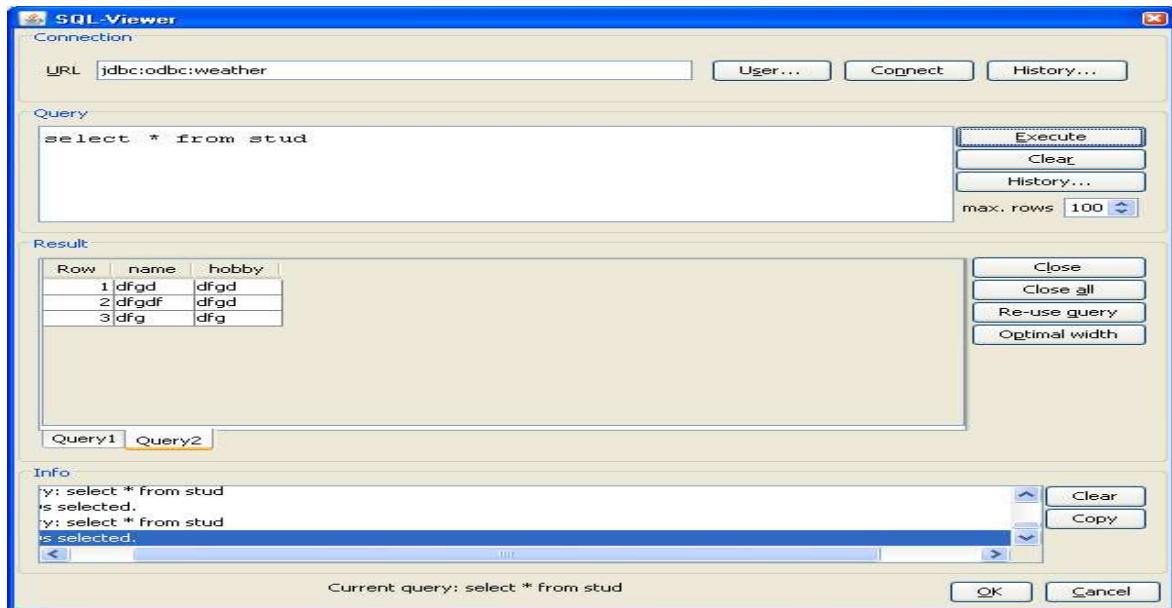
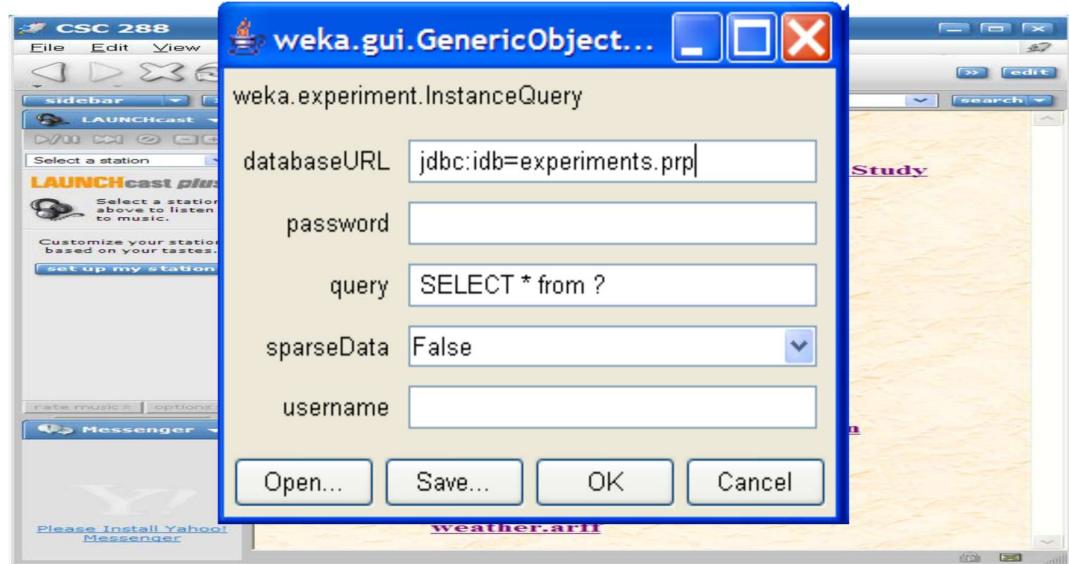
Opening file from a web site

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



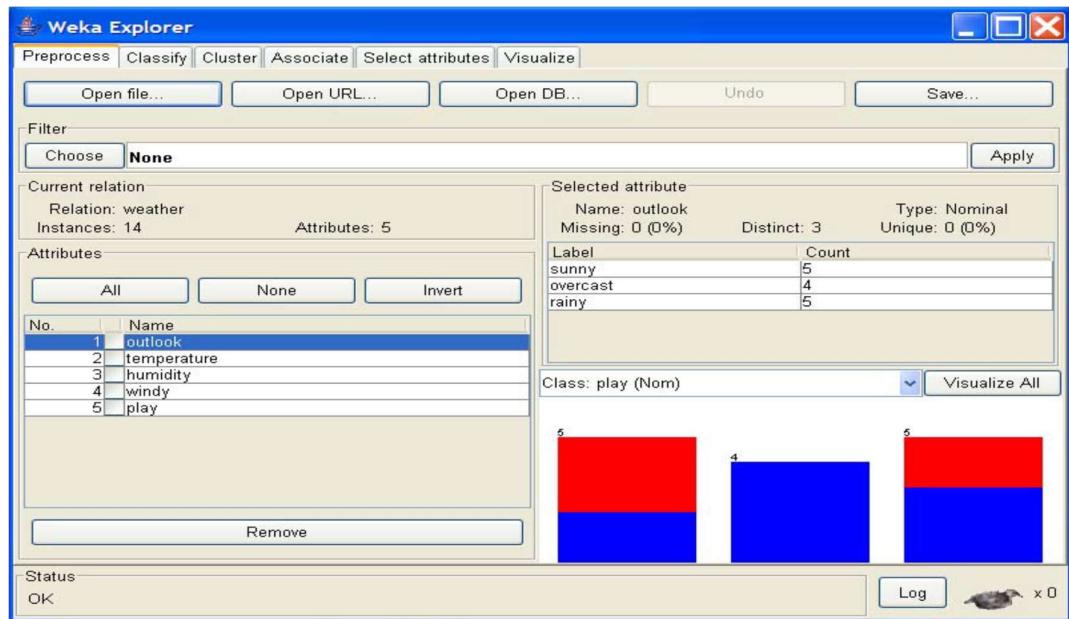
following website:

Reading data from a database:

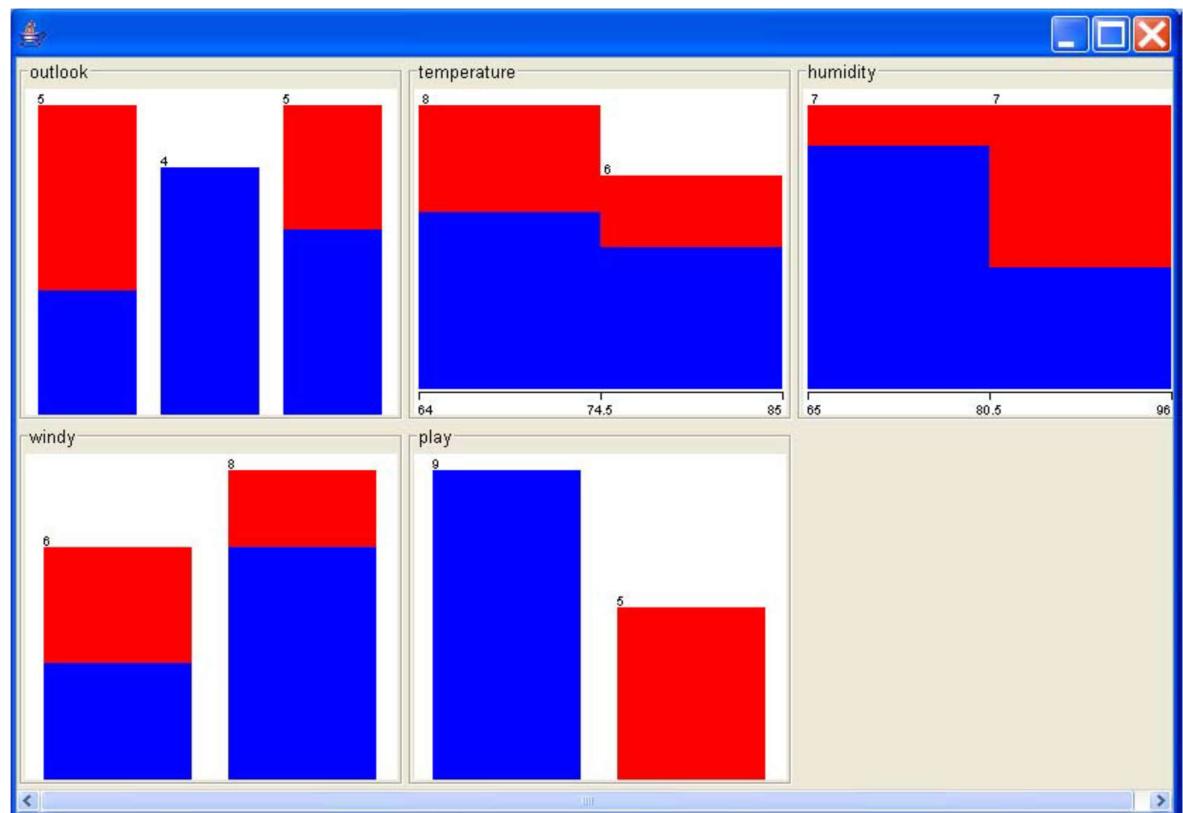


Loading data

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



Visualize Attributes:



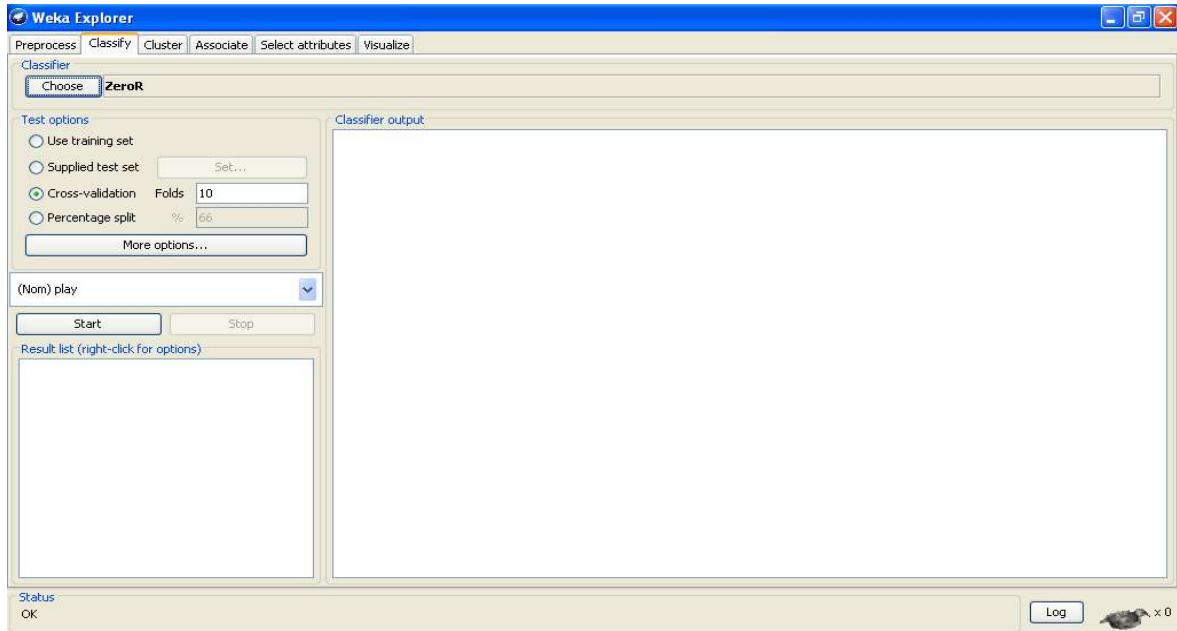
visualize all attributes by clicking on ‘Visualize All’ button.,,

Practical No - 10

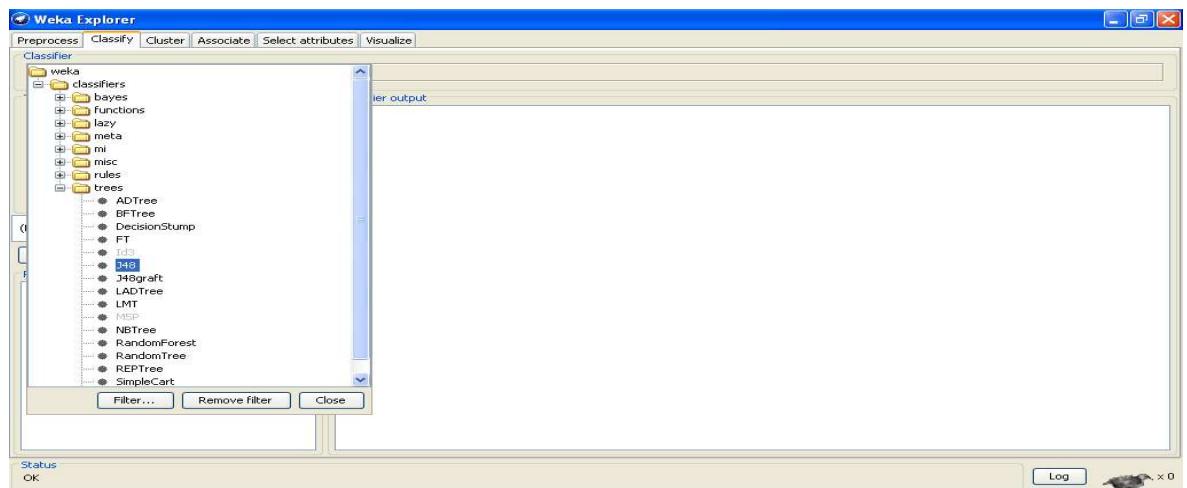
Aim : Classification problems.

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.

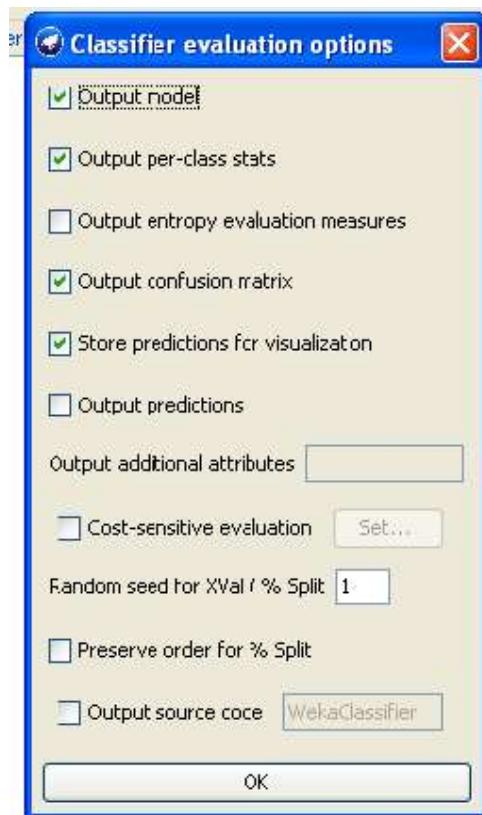


Check ‘Percentage split’ radio-button and keep it as default 66%. Click on ‘More options...’ button.

make sure that the Following options are checked :

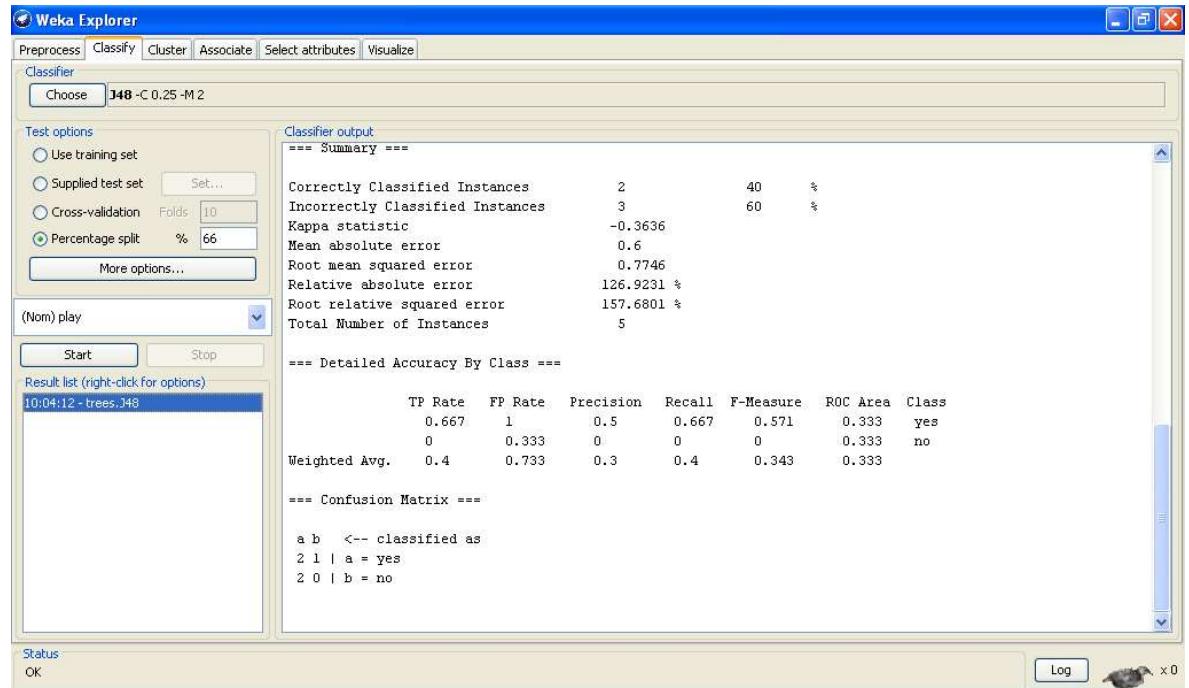
1. Output model.
2. Output per-class stats.
3. Output confusion matrix
4. Store predictions for visualization.
5. Set ‘Random seed for Xval / % Split’ to 1.

.



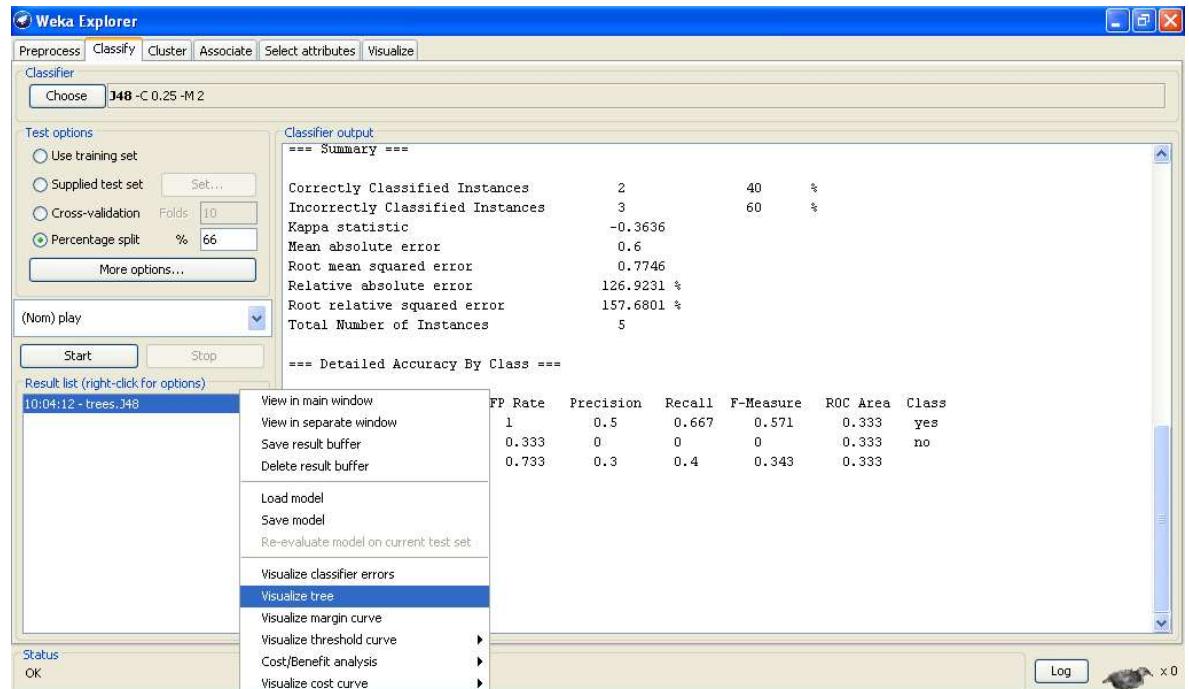
Once the options have been specified, you can run the classification algorithm. Click on

‘Start’ button

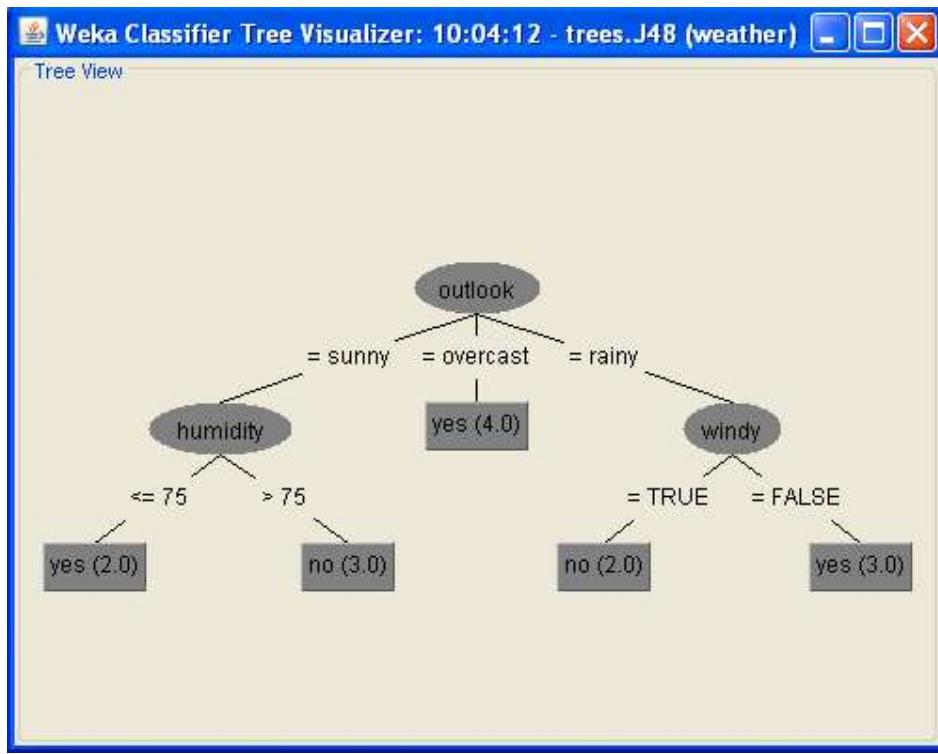


To see a graphical representation of the classification tree. Right-click on the

entry in ‘Result list’ for which you would like to visualize a tree.

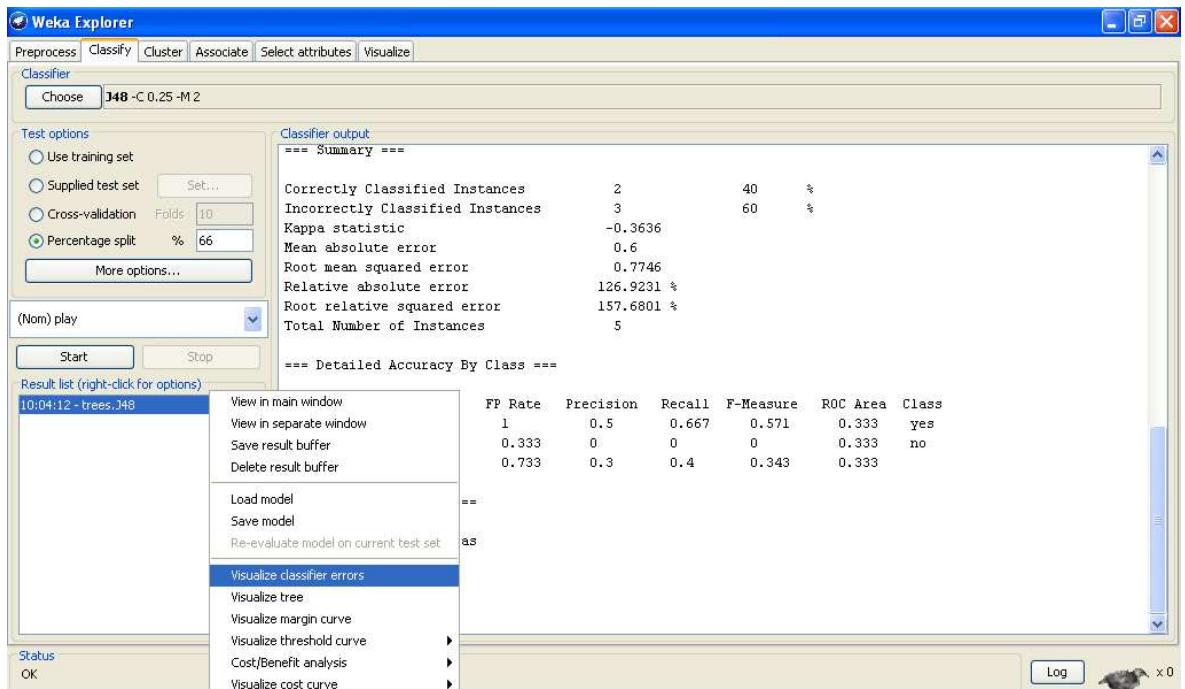


Select the item ‘Visualize tree’; a new window comes up to the screen displaying the tree.

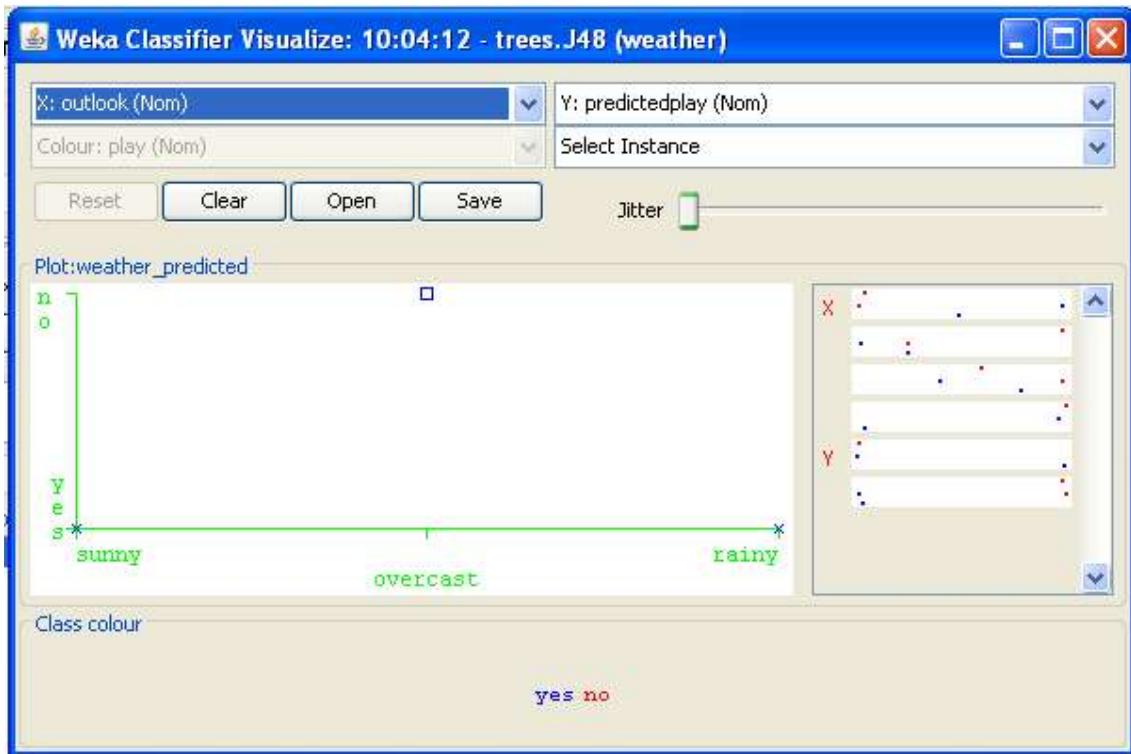


To visualize classification errors. Right-click on the entry in ‘Result list’ again

and select ‘Visualize classifier errors’ from the menu:



‘Visualize’ window displaying graph appears on the screen.

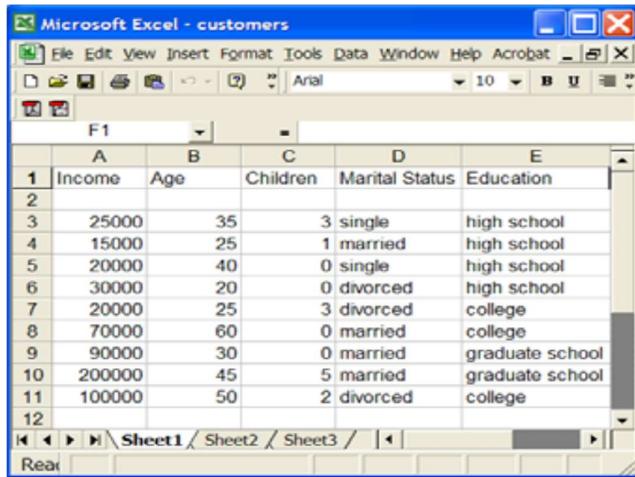


Practical N0 - 11

Aim : Clustering Analysis.

Solution :

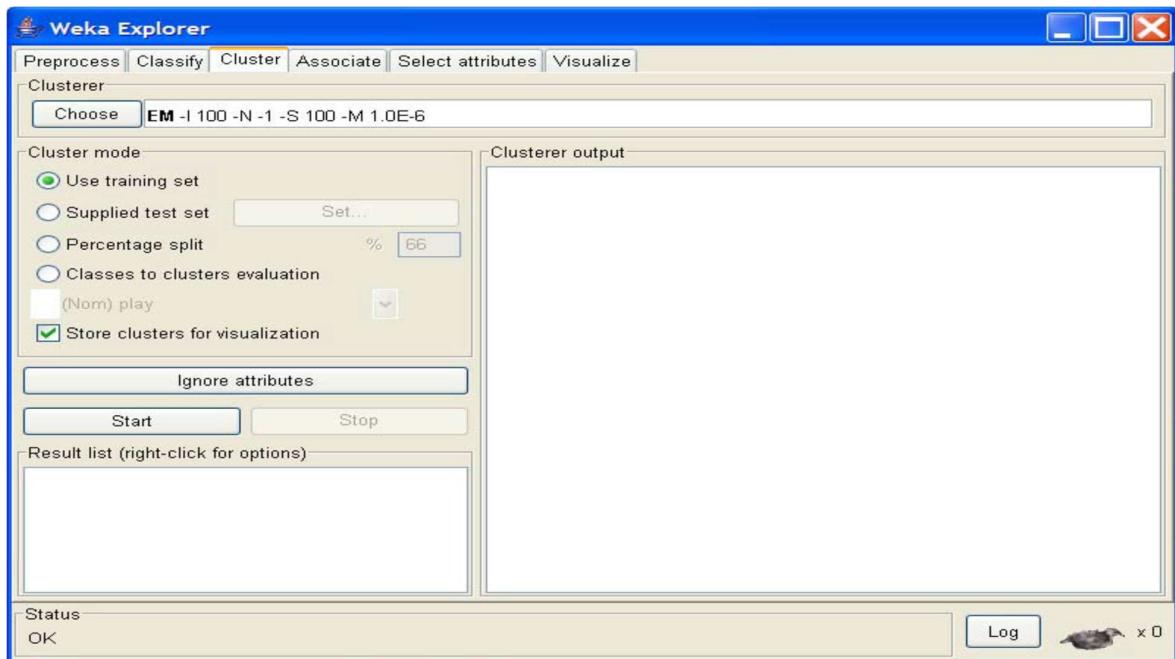
we will use customer data [6] that is contained in “customers.arff” file and analyze it with k-means clustering scheme.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - customers". The data is organized into columns labeled A through E. Column A contains row numbers from 1 to 12. Column B is labeled "Income", column C is "Age", column D is "Children", column E is "Marital Status", and column F is "Education". The data entries are as follows:

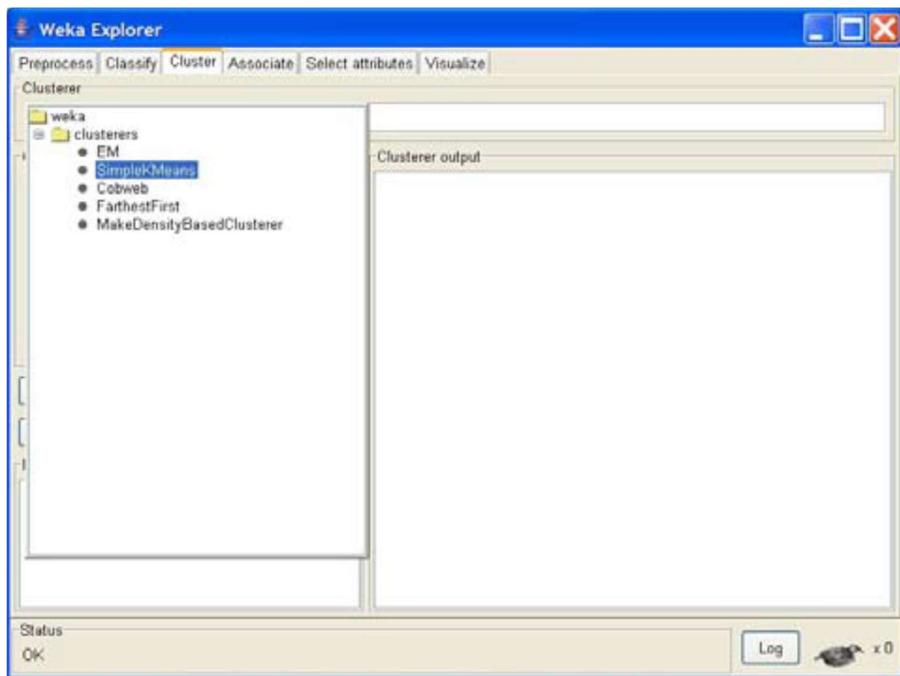
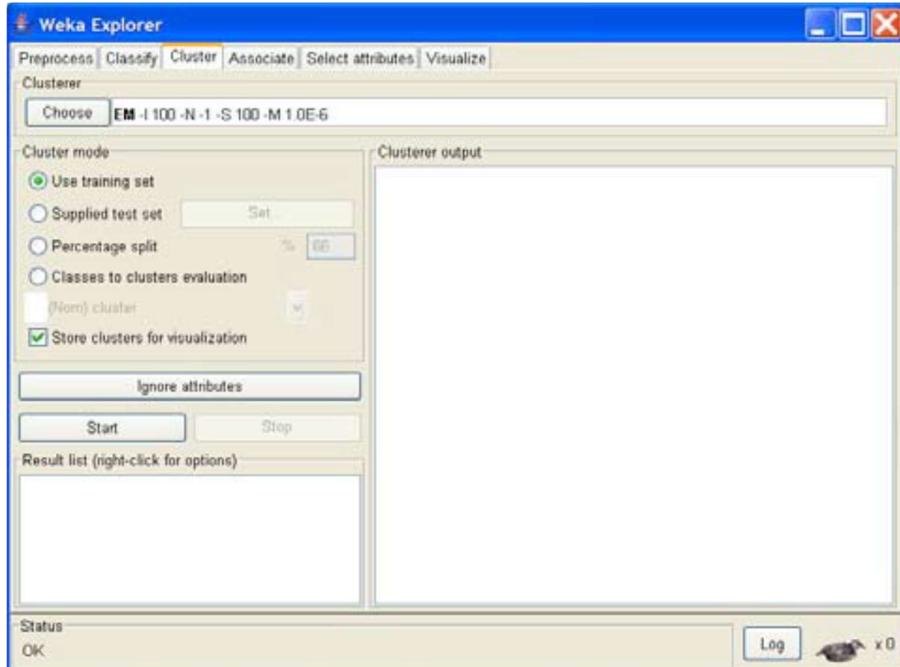
F1	A	B	C	D	E
1	Income	Age	Children	Marital Status	Education
2					
3	25000	35	3	single	high school
4	15000	25	1	married	high school
5	20000	40	0	single	high school
6	30000	20	0	divorced	high school
7	20000	25	3	divorced	college
8	70000	60	0	married	college
9	90000	30	0	married	graduate school
10	200000	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.

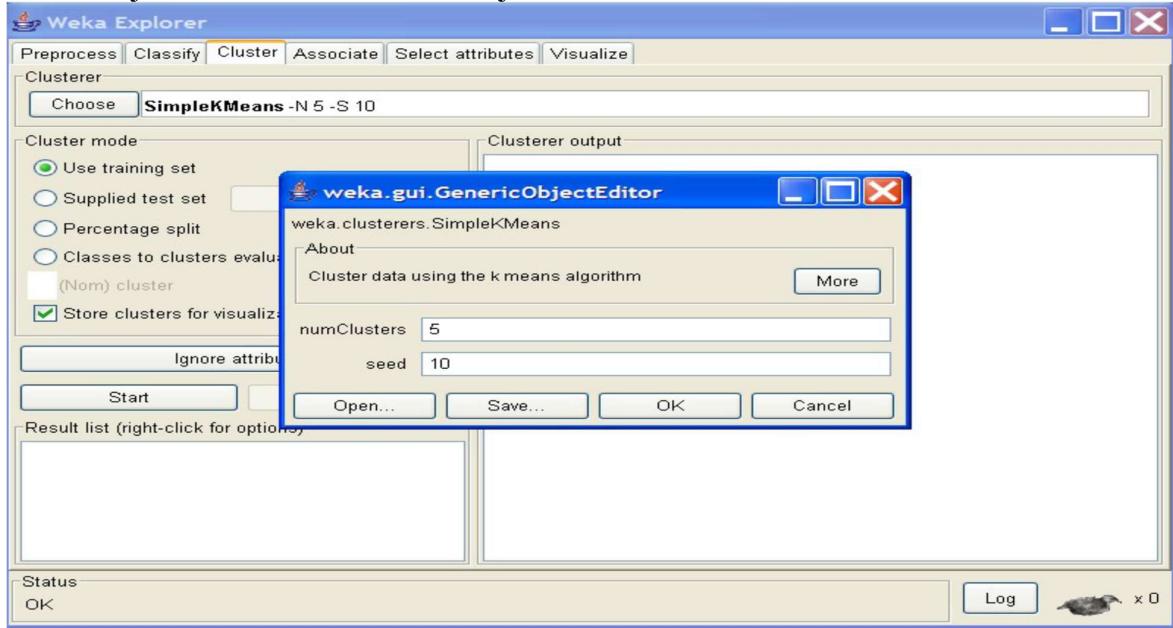


In the ‘Clusterer’ box click on ‘Choose’ button. In pull-down menu select WEKA □

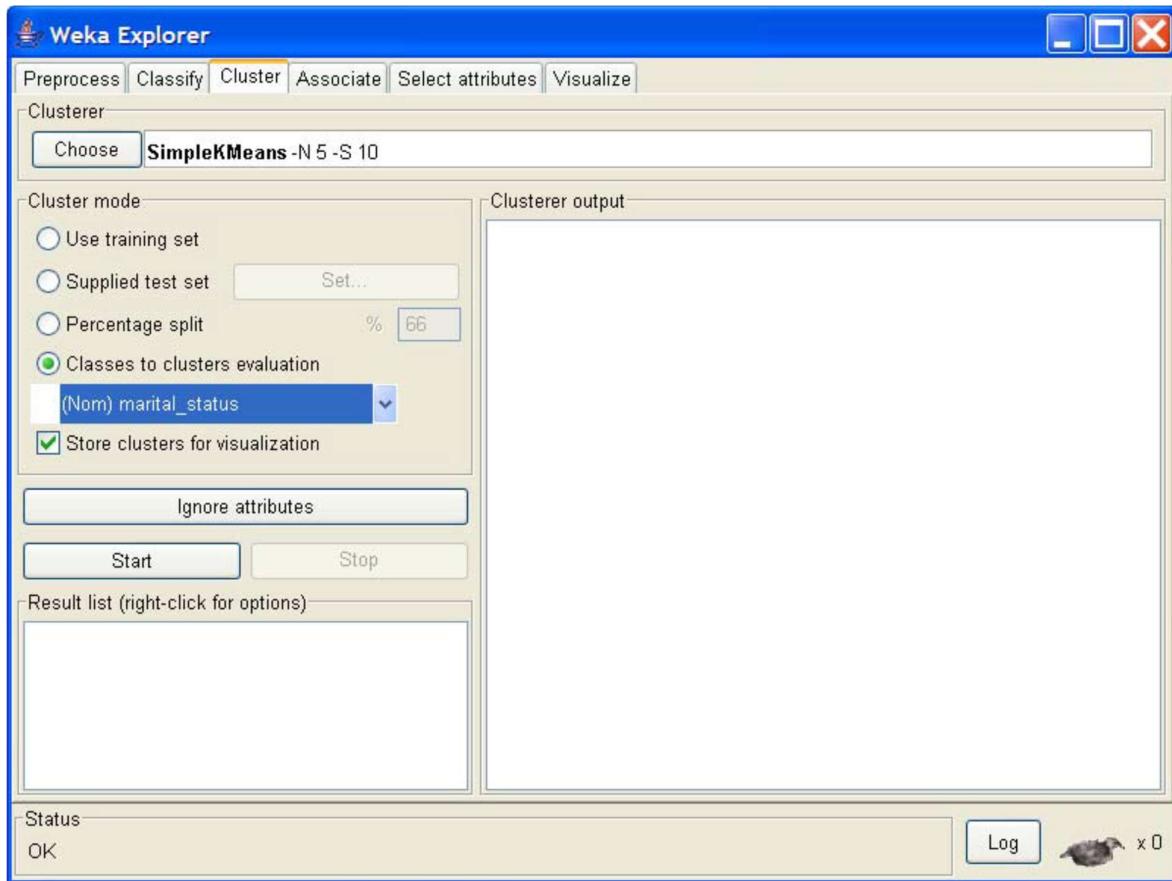
Clusterers, and select the cluster scheme ‘SimpleKMeans’. Some implementations of K-means only allow numerical values for attributes.



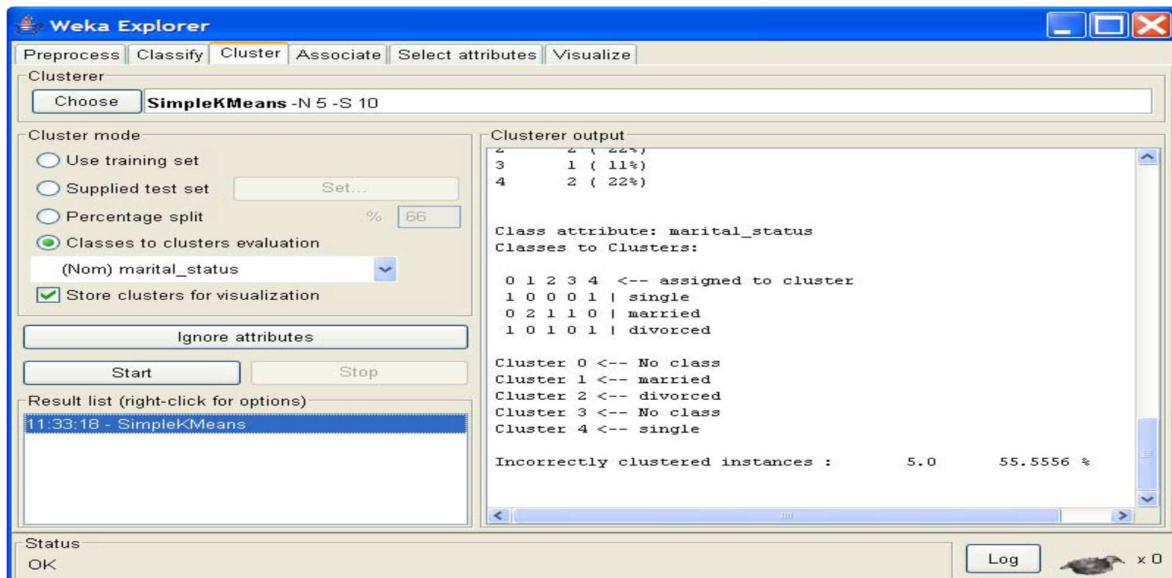
right-click on the algorithm “weak.gui.GenericObjectEditor” comes up to the screen. Set the value in “numClusters” box to 5(instead of default 2) because you have five clusters in your .arff file.



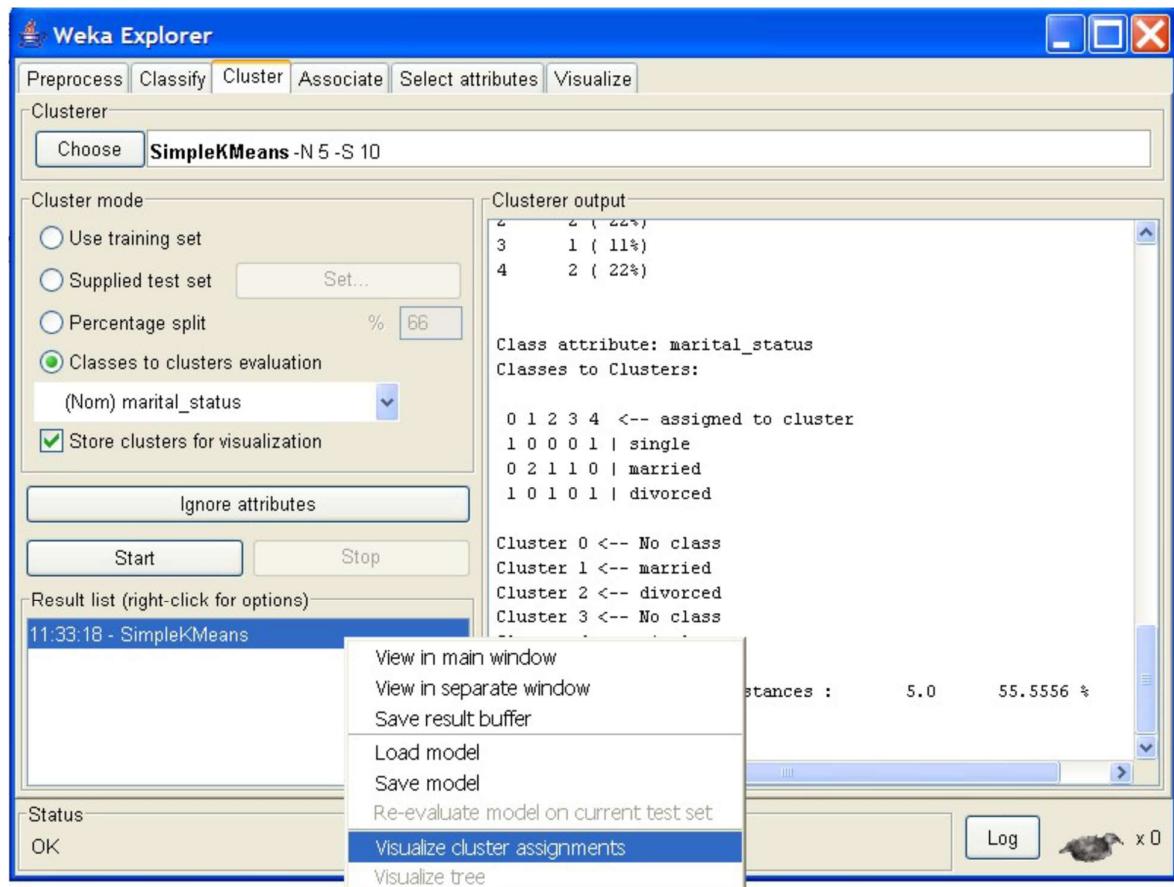
Click on ‘Classes to cluster evaluation’ radio-button in ‘Cluster mode’ box and select ‘marital_status’ in the pull-down box below.



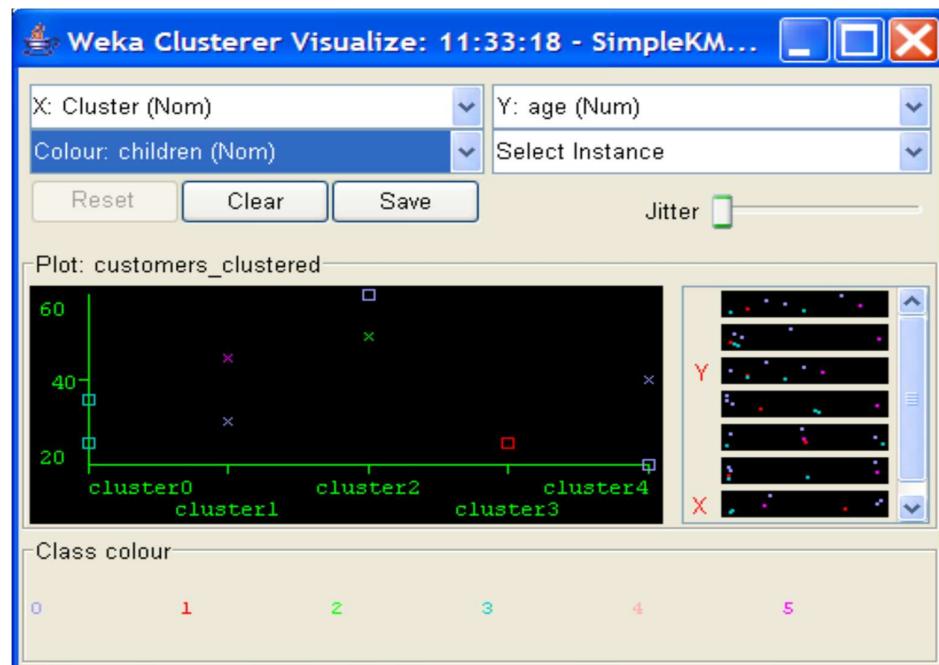
Click on the ‘Start’ button to execute the algorithm.



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

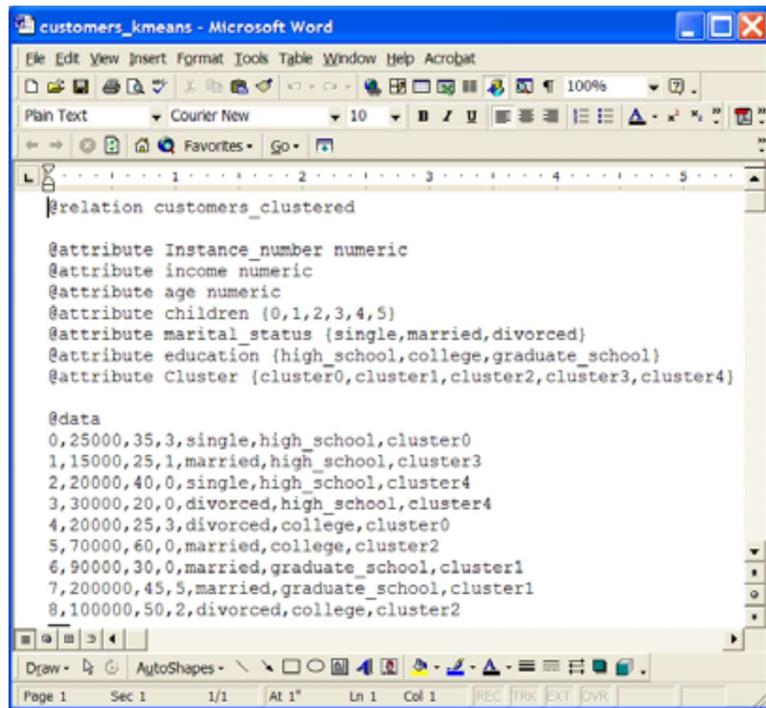


'Weka Clusterer Visualize' window.



there is a new attribute appeared in the file – ‘cluster’ that was added by

WEKA. This attribute represents the clustering done by WEKA.



The screenshot shows a Microsoft Word document window titled "customers_kmeans - Microsoft Word". The document contains an ARFF (Attribute-Relationship File Format) dataset. The code is as follows:

```
@relation 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
7,200000,45,5,married,graduate_school,cluster1
8,100000,50,2,divorced,college,cluster2
```

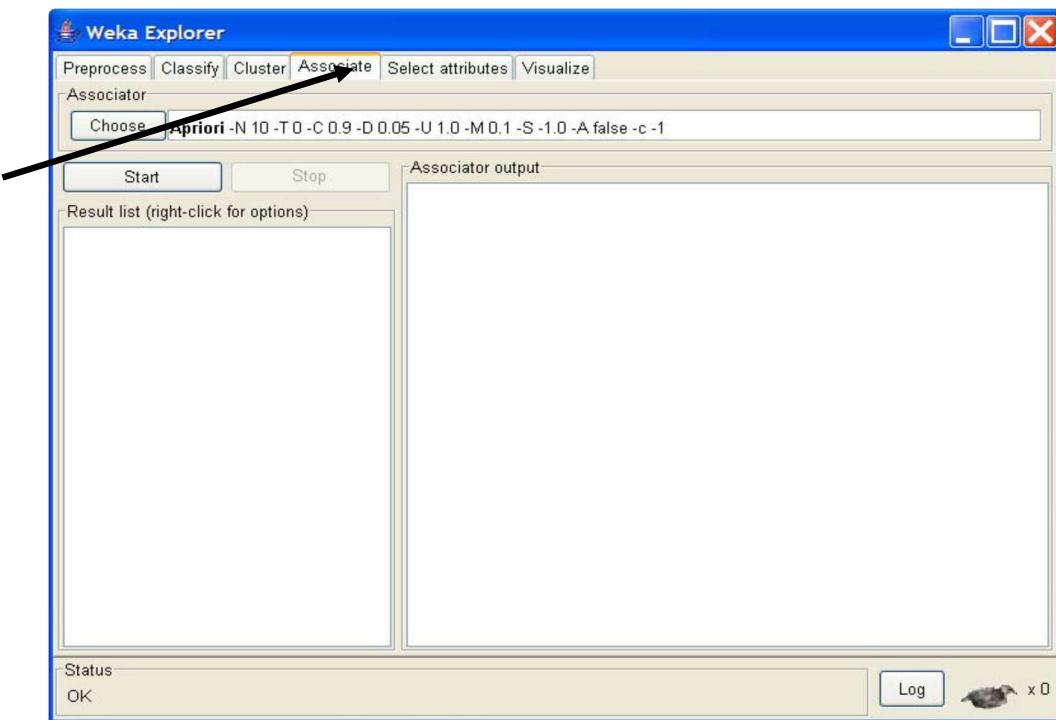
Practical No-12

Aim : Association Rule Mining.

Solution :

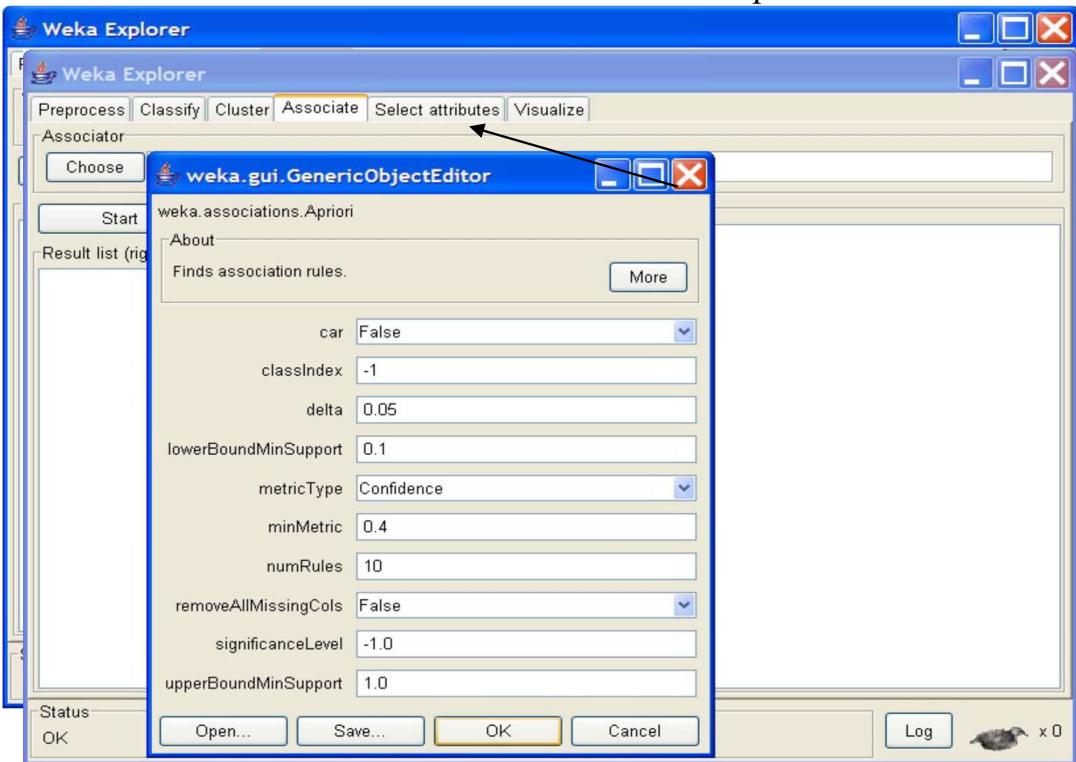
Choosing Association Scheme

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

