

Class: MSC-I

SEM: I

Subject: Data Mining & Data Warehousing (DMDW)

Paper: III

Academic Year: 2022-2023

Roll No: 517

INDEX

S.NO	Date	Topic	Page. No	Sign
1		Creation of Dimensions and Fact tables.	2	
2		Create Data Source using SSAS (SQL Server Analysis Services.)	4	
3		Create Data Source View using SSAS (SQL Server Analysis Services.)	9	
4		Create cube using SSAS(SQL Server Analysis Services.) and process the cube.	13	
5		View cube data in multidimensional Format. (Excel Pivot Chart.)	18	
6		Working with measures in the cube.	21	
7		Creating an Excel Pivot Table and Pivot Chart by using the OLAP cube data.	23	
8		Firing Queries on Tables.	28	
9		Calculation & KPI	-	
10		Data pre-processing.	32	
11		Data discretization.	37	
12		Classification problems.	39	
13		Clustering Analysis.	44	
14		Association Rule Mining.	49	
15		Data visualization.	52	

Practical No 1

Aim : Creation of Dimensions and Fact tables.

Solution :

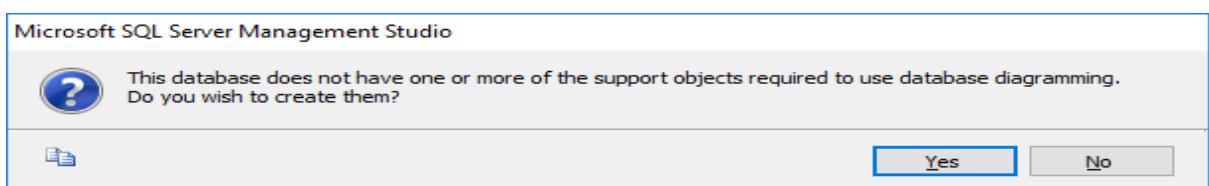
Open Application -> Microsoft SQL Server 2008 R2 -> SQL Server Management Studio

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

Create a Database Diagrams

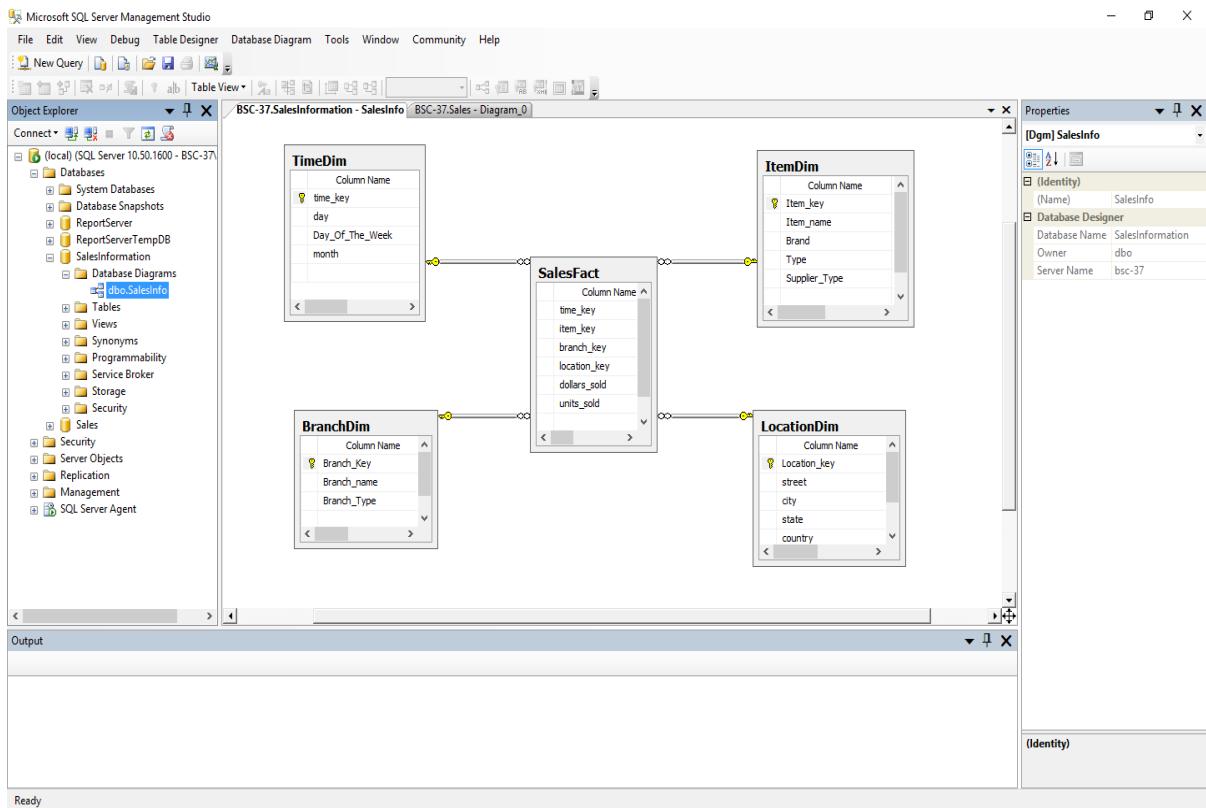
Expand the “SalesInformation” database folder.

- 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 “SalesInfo”. (After saving Database Diagrams fact and dimension tables are automatically placed in Table tab.)**

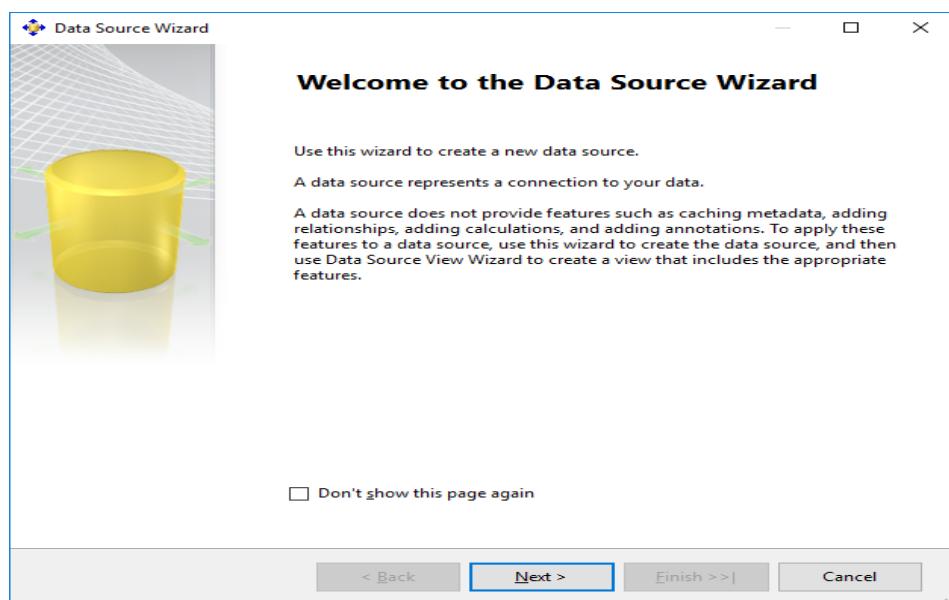
Practical No 2

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

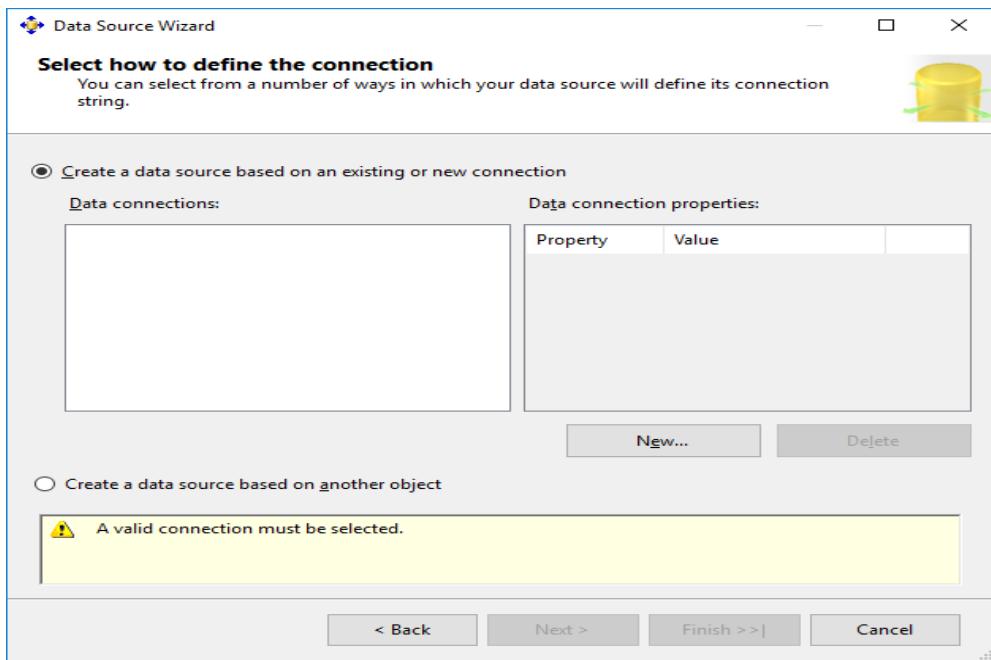
Solution :

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 "SalesInfo_BIPrj" and click on OK.
2. Right Click on Data Sources -> New Data Source

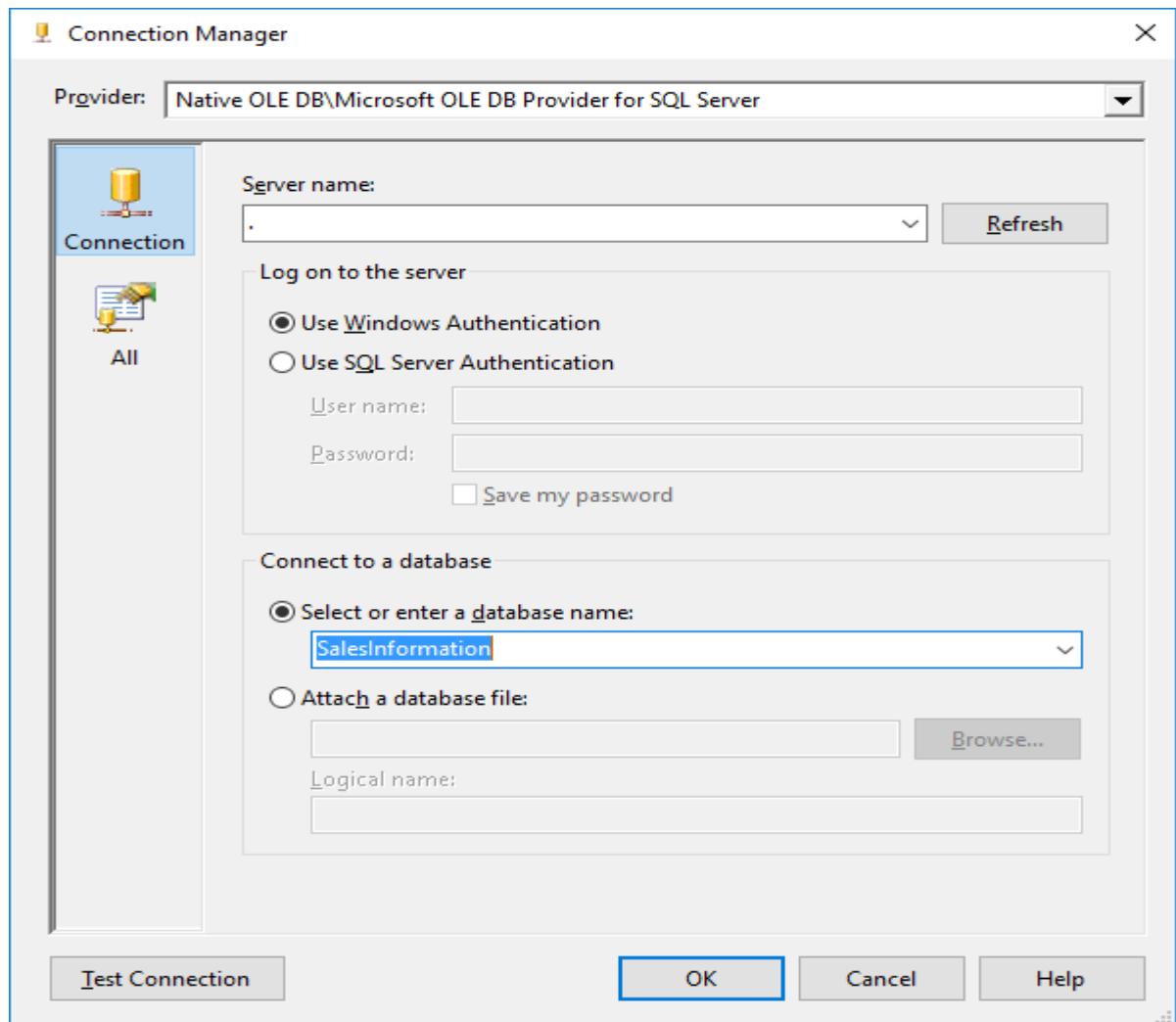


Click on Next.

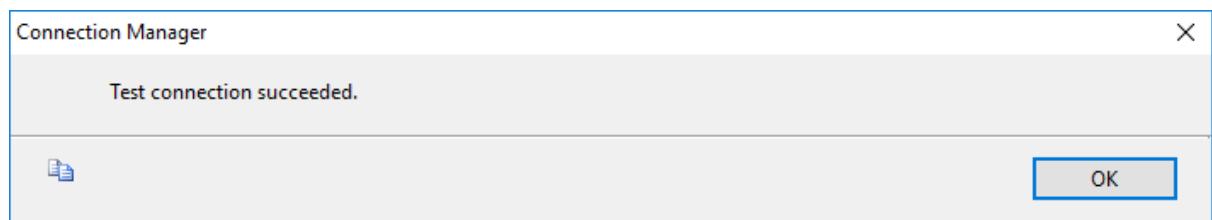


Click on New.

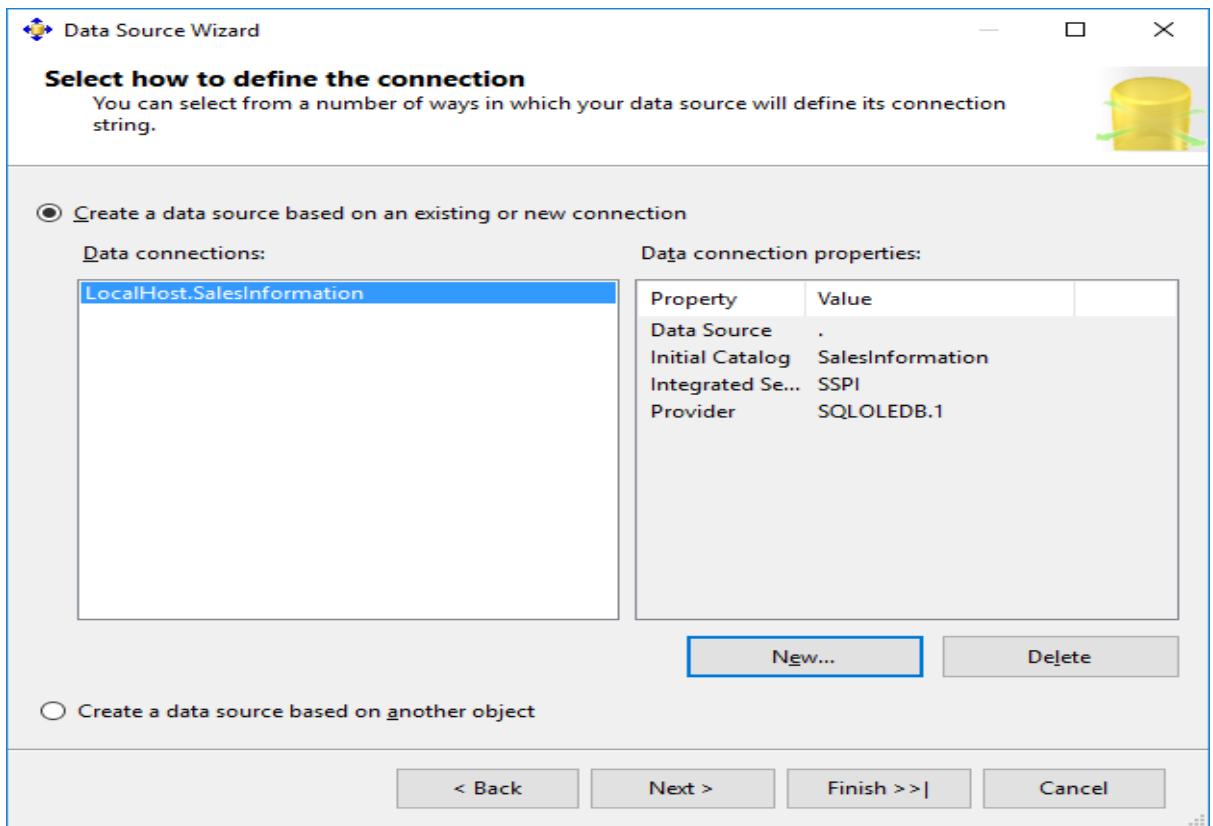
3. Choose Provider as “Microsoft OLEDB Provider for SQL Server” , Server Name as “.”, Select database name as “SalesInformation”.(Created in SQL Server Management studio).



4. Click on Test Connection.

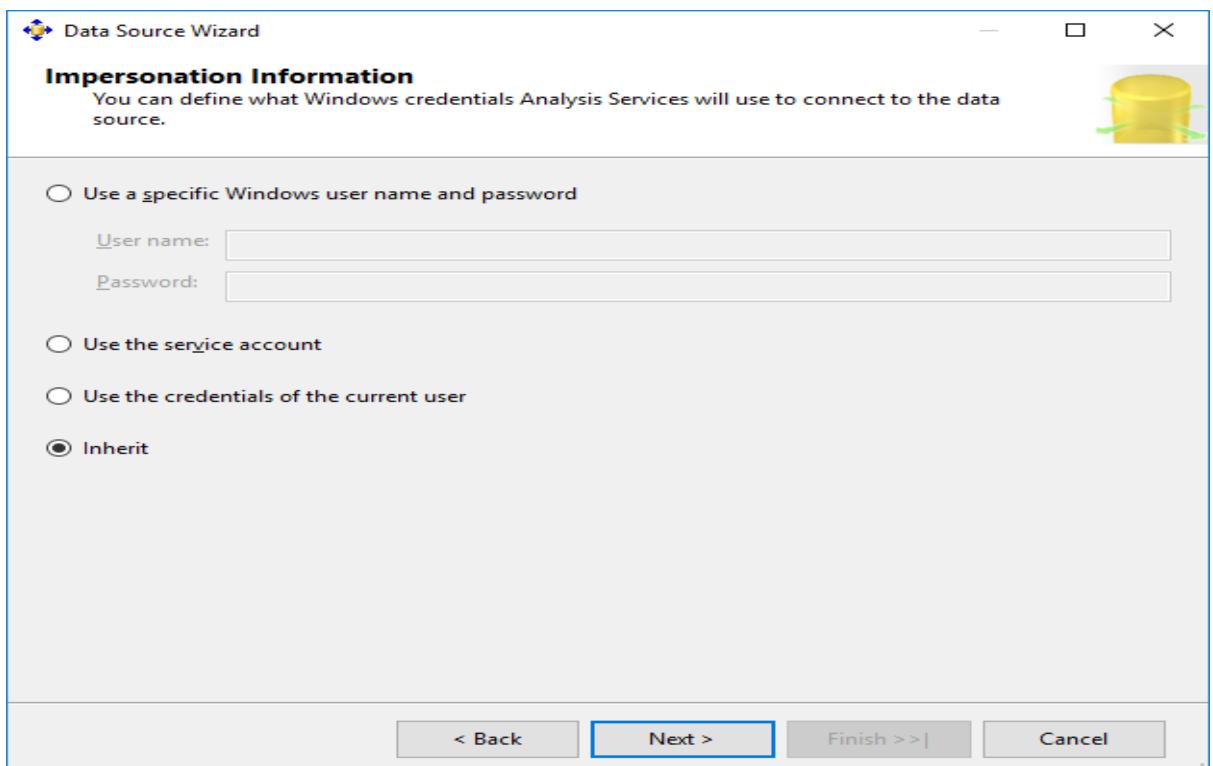


Click on OK.



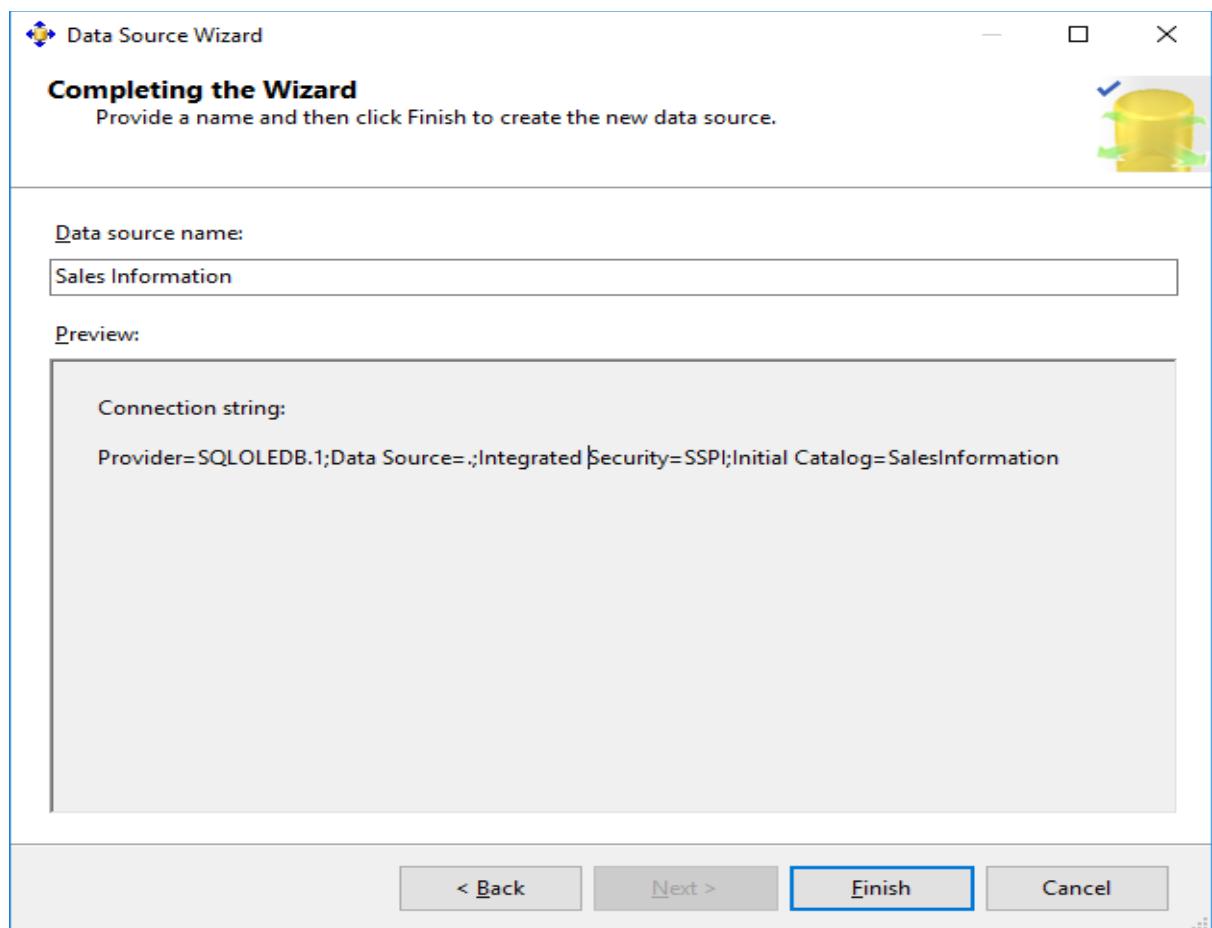
Click on Next

5. Choose “Inherit” option.



Click on Next.

6. Click on Finish.



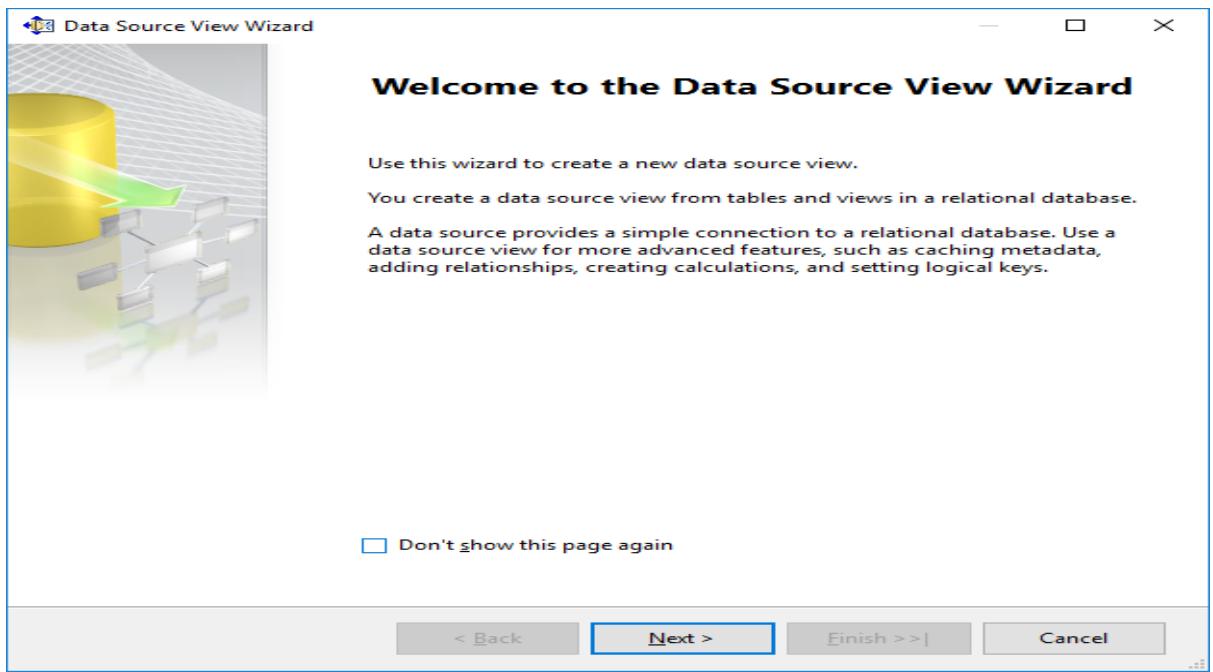
Name Data Source as “Sales Information”.

Practical No 3

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

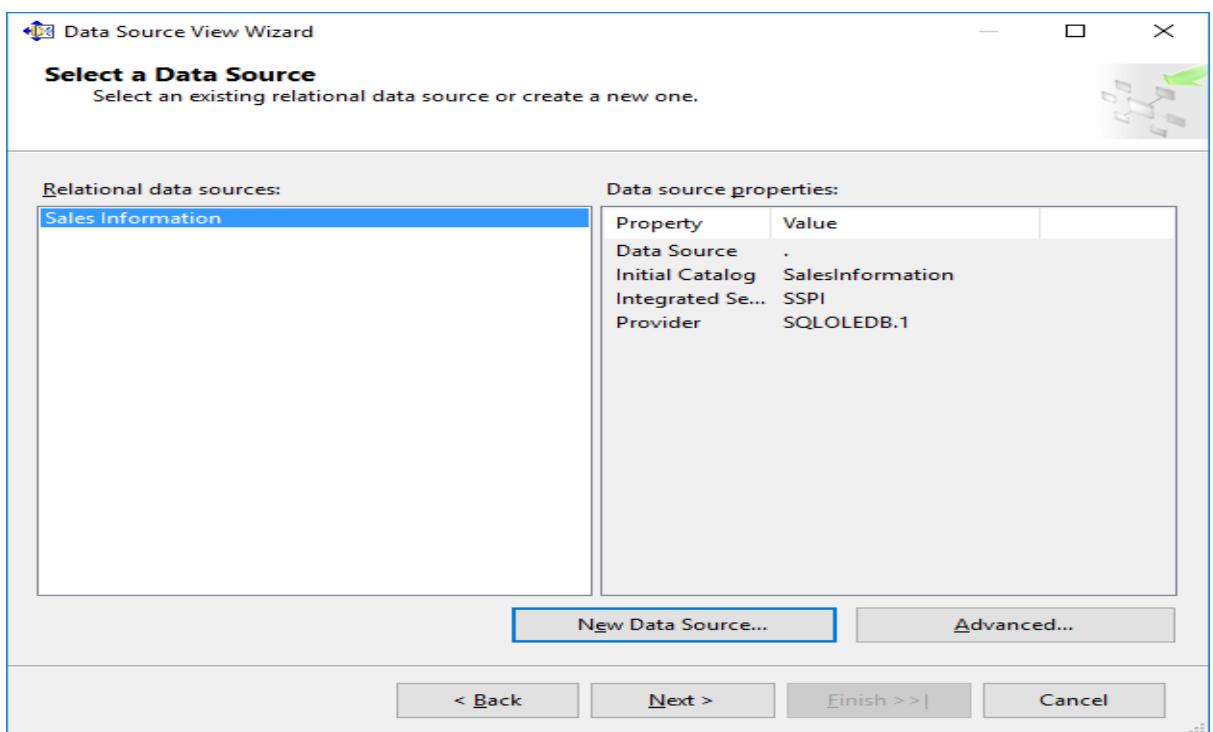
Solution :

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

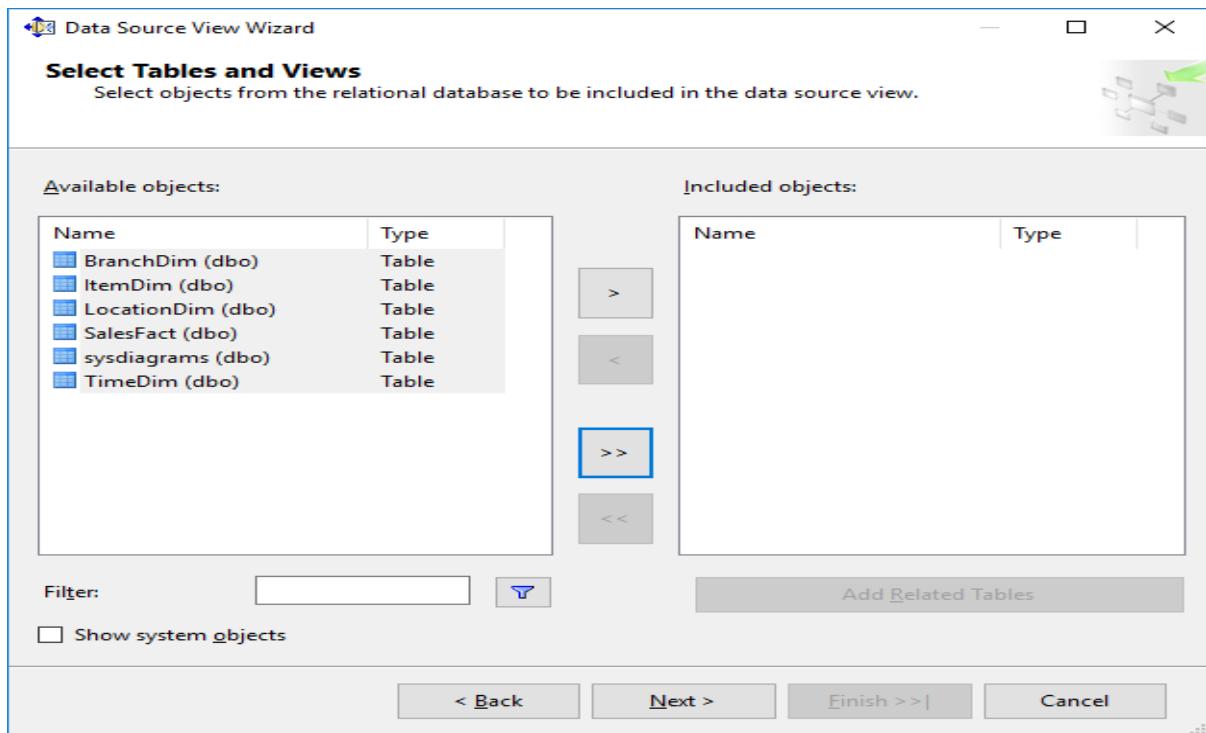


Click on Next.

2. Click on Next.



3. Select Tables and Views.



 Data Source View Wizard

Select Tables and Views
Select objects from the relational database to be included in the data source view.

Available objects:

Name	Type

Included objects:

Name	Type
BranchDim (dbo)	Table
ItemDim (dbo)	Table
LocationDim (dbo)	Table
SalesFact (dbo)	Table
sysdiagrams (dbo)	Table
TimeDim (dbo)	Table

> < >> <<

Filter: 

Add Related Tables

Show system objects

< Back **Next >** Finish >> Cancel

 Data Source View Wizard

Completing the Wizard
Provide a name, and then click Finish to create the new data source view.

Name: **Sales Information**

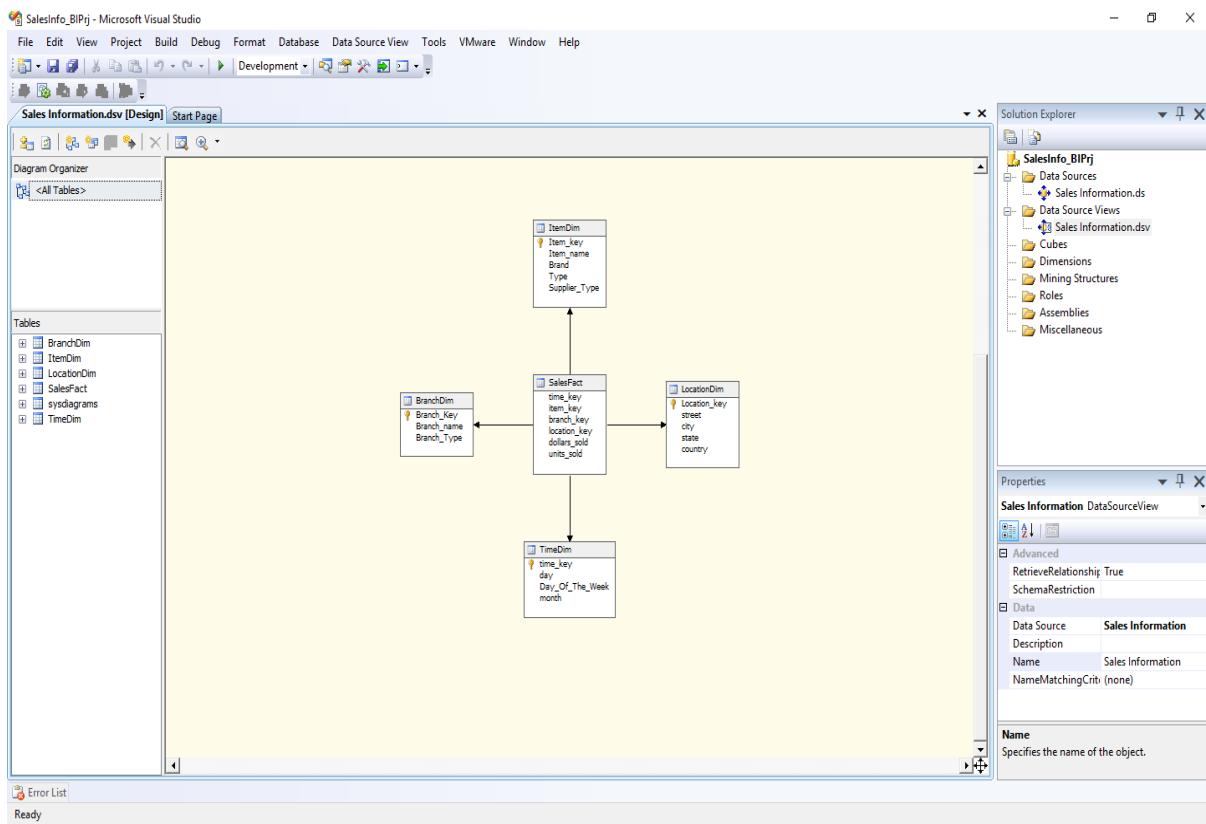
Preview:

- [-] Sales Information
 - BranchDim (dbo)
 - ItemDim (dbo)
 - LocationDim (dbo)
 - SalesFact (dbo)
 - sysdiagrams (dbo)
 - TimeDim (dbo)

< Back **Finish** Cancel

Click on Finish.

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

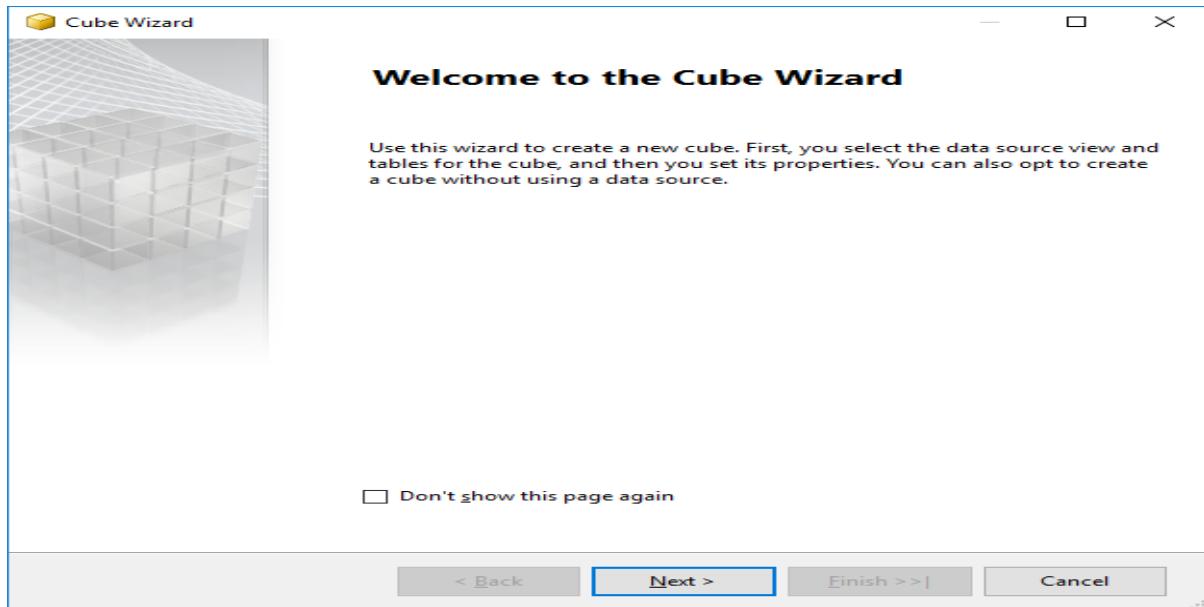


Practical No 4

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

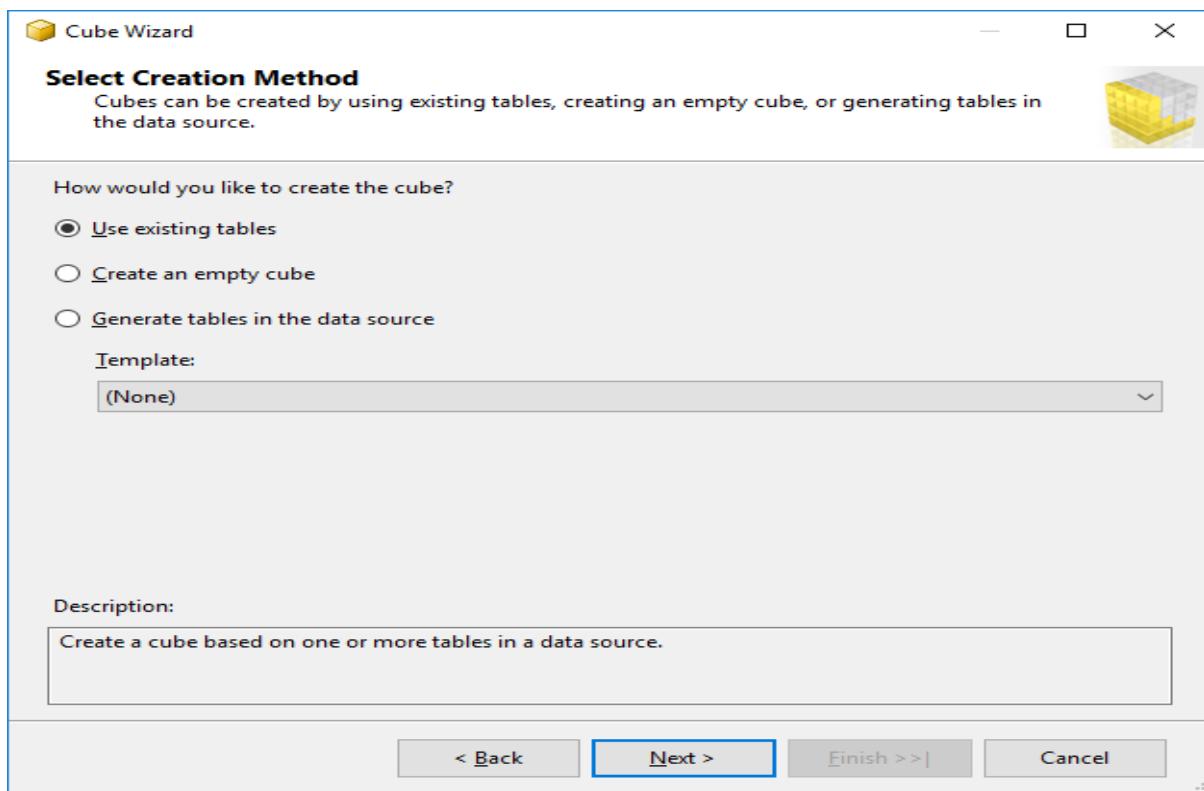
Solution :

1. Right click on Cubes -> New Cube.

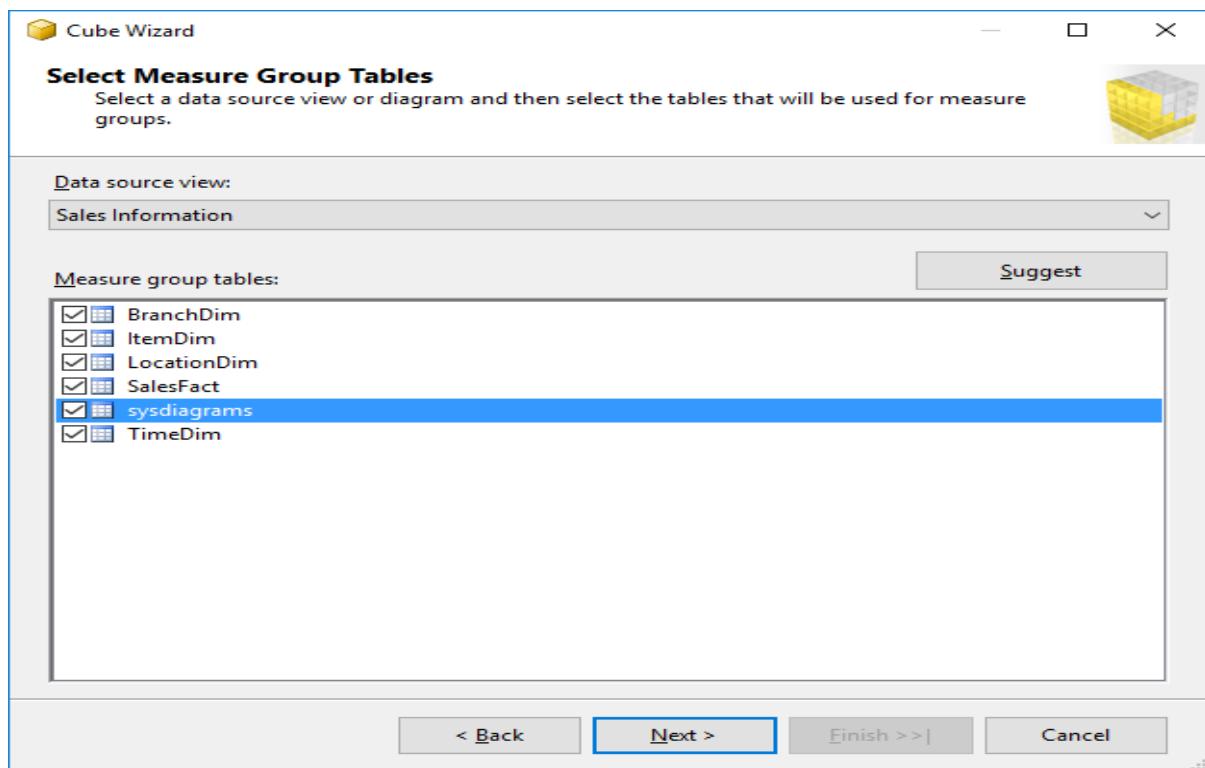


Click on Next.

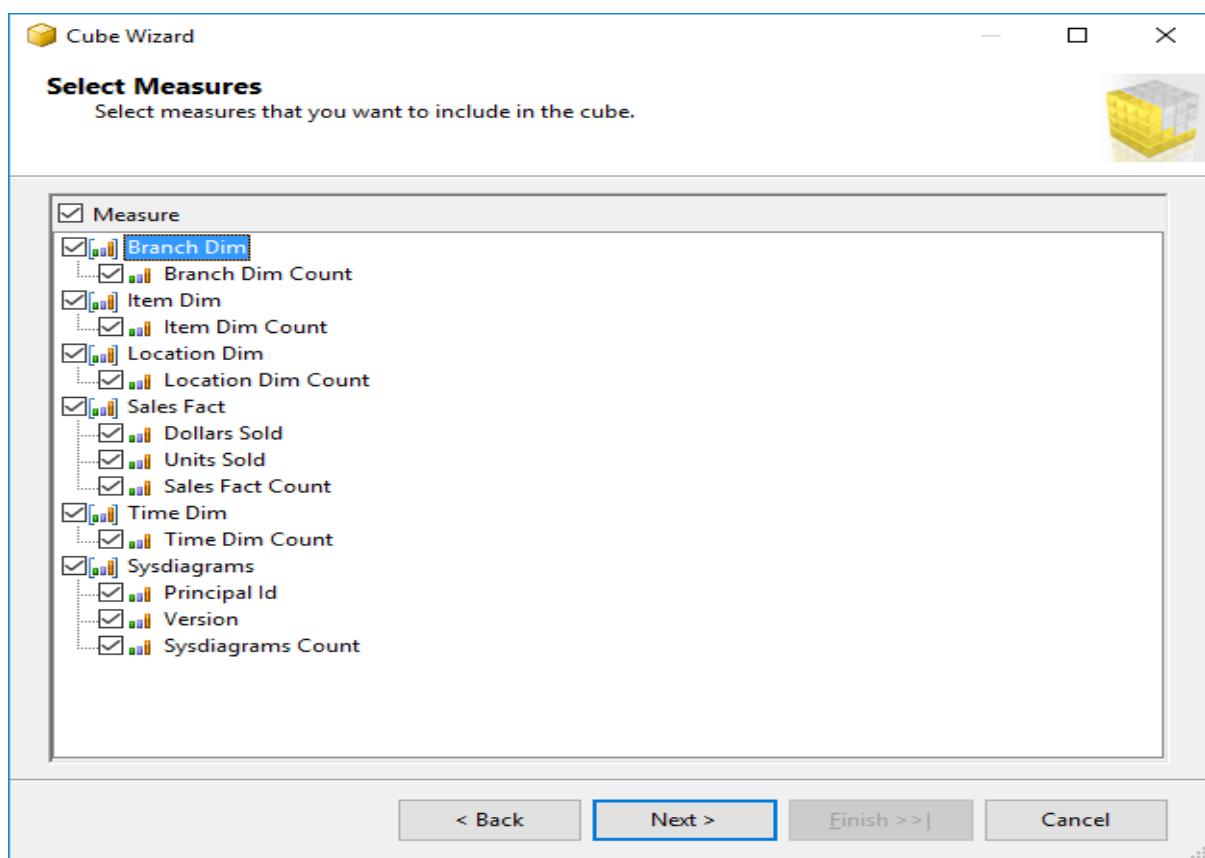
2. Select First option “Use existing tables”. Click on Next.



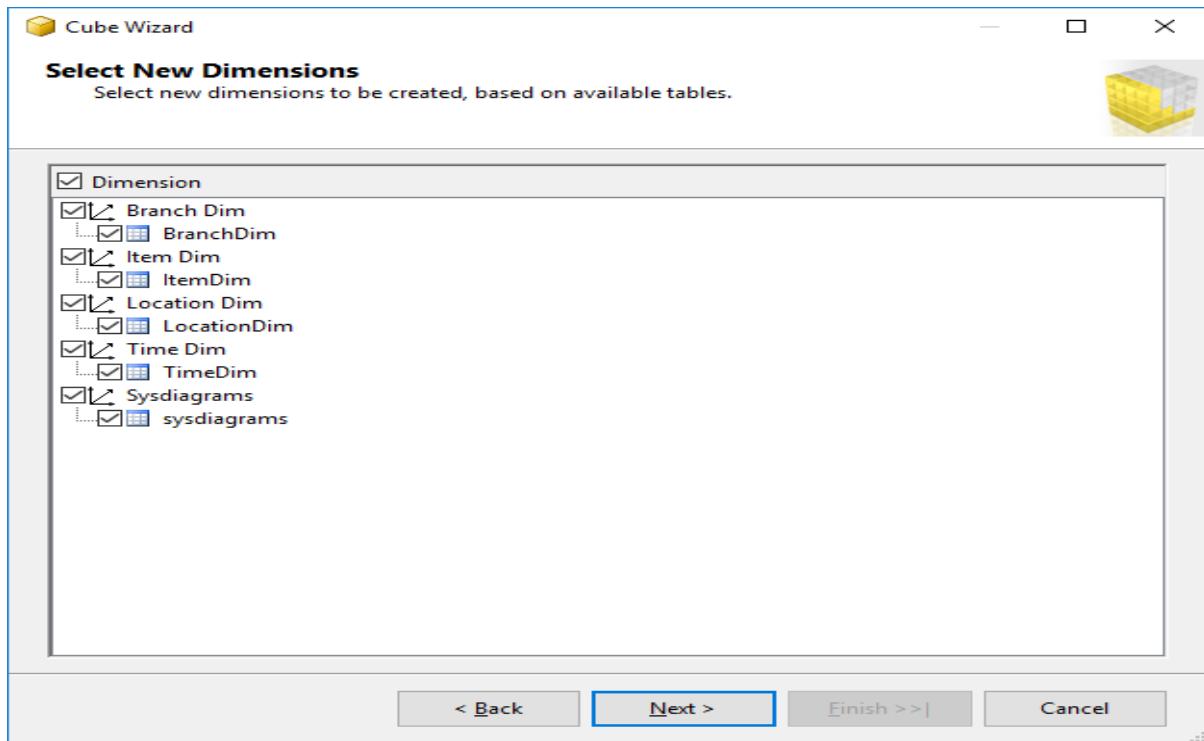
3. Select Data Source View as “Sales Information” and Select all the tables.



Click on Next.

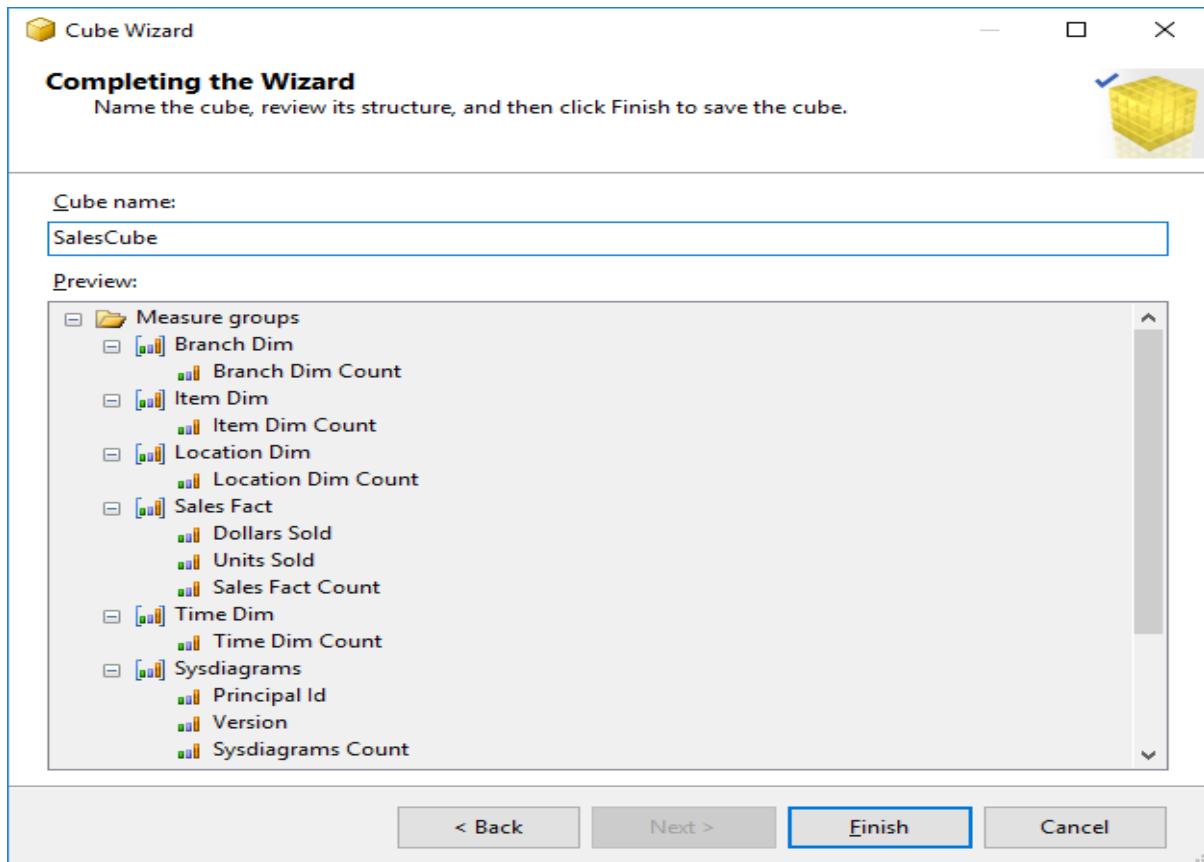


Click on Next.



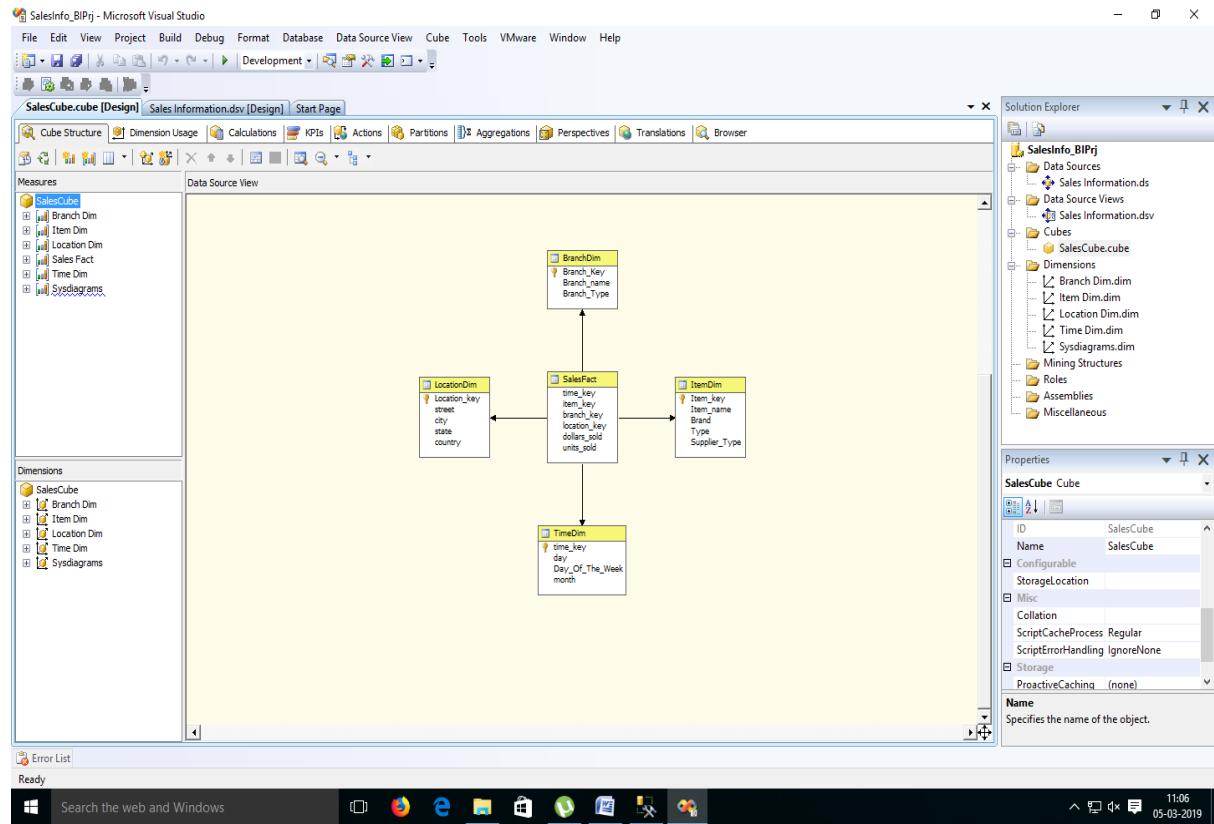
Click on Next.

4. Name Cube as “SalesCube”.

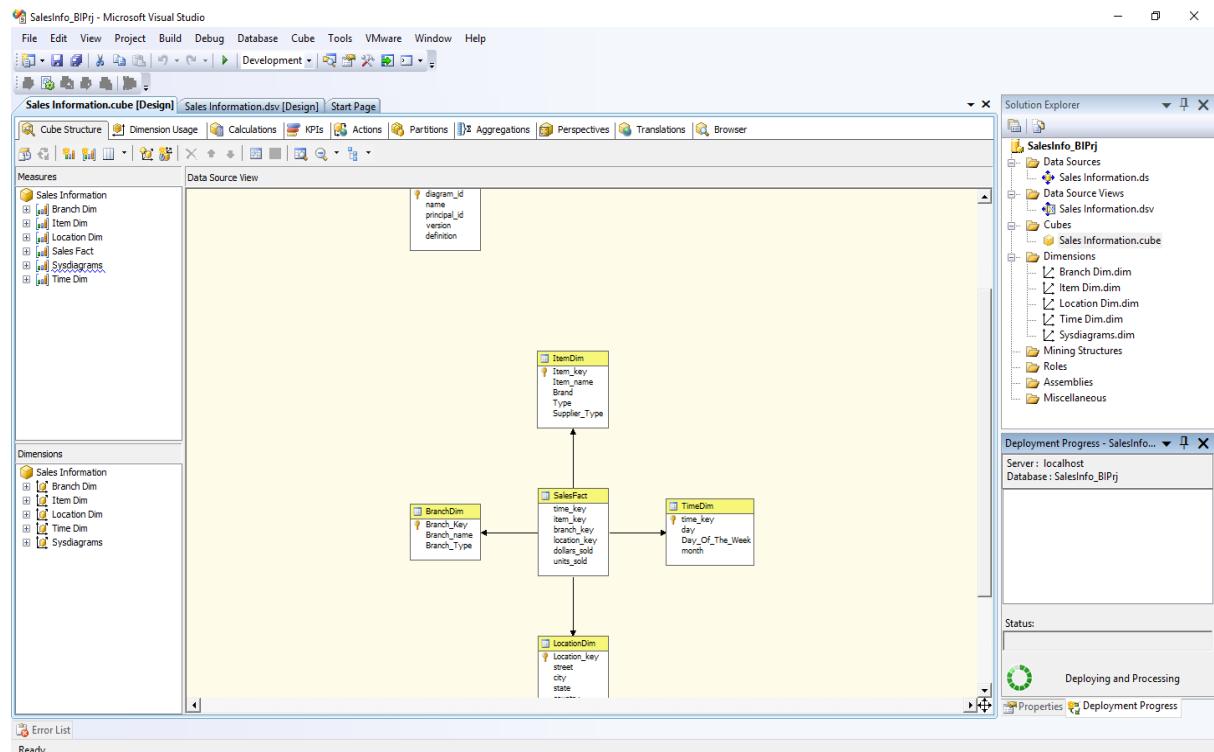


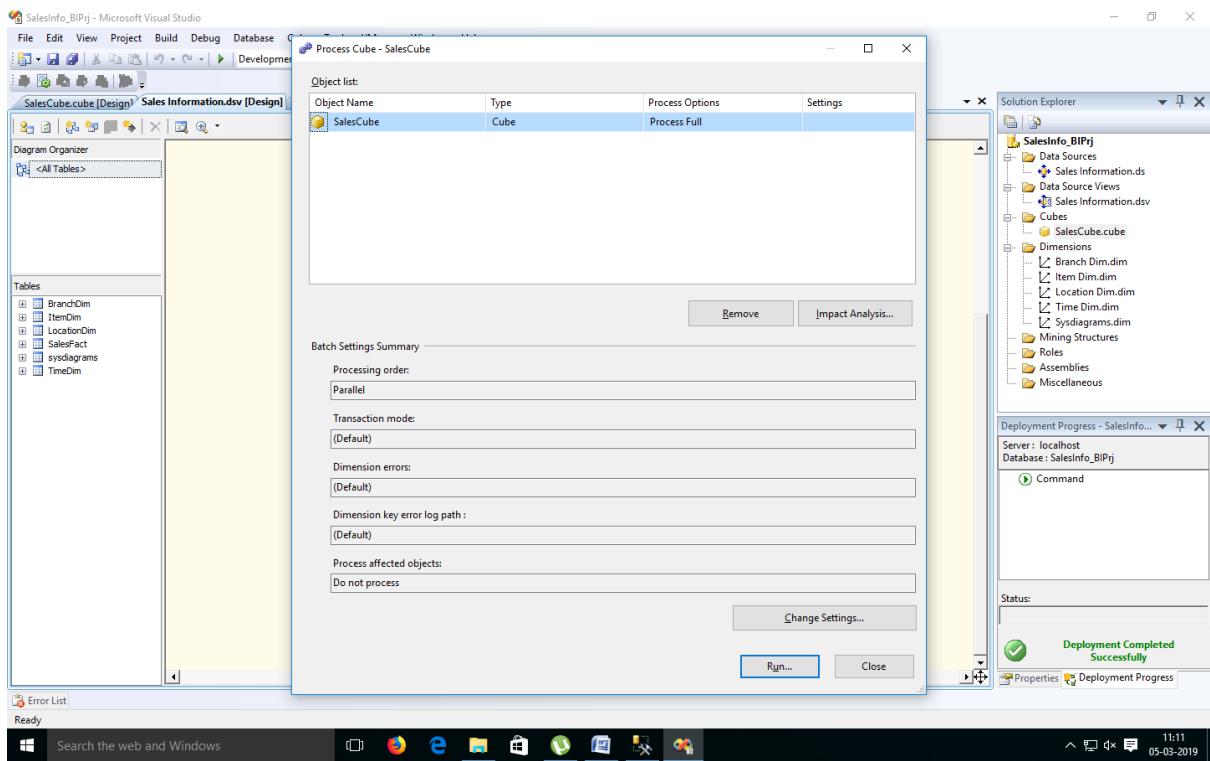
Click on Finish.

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

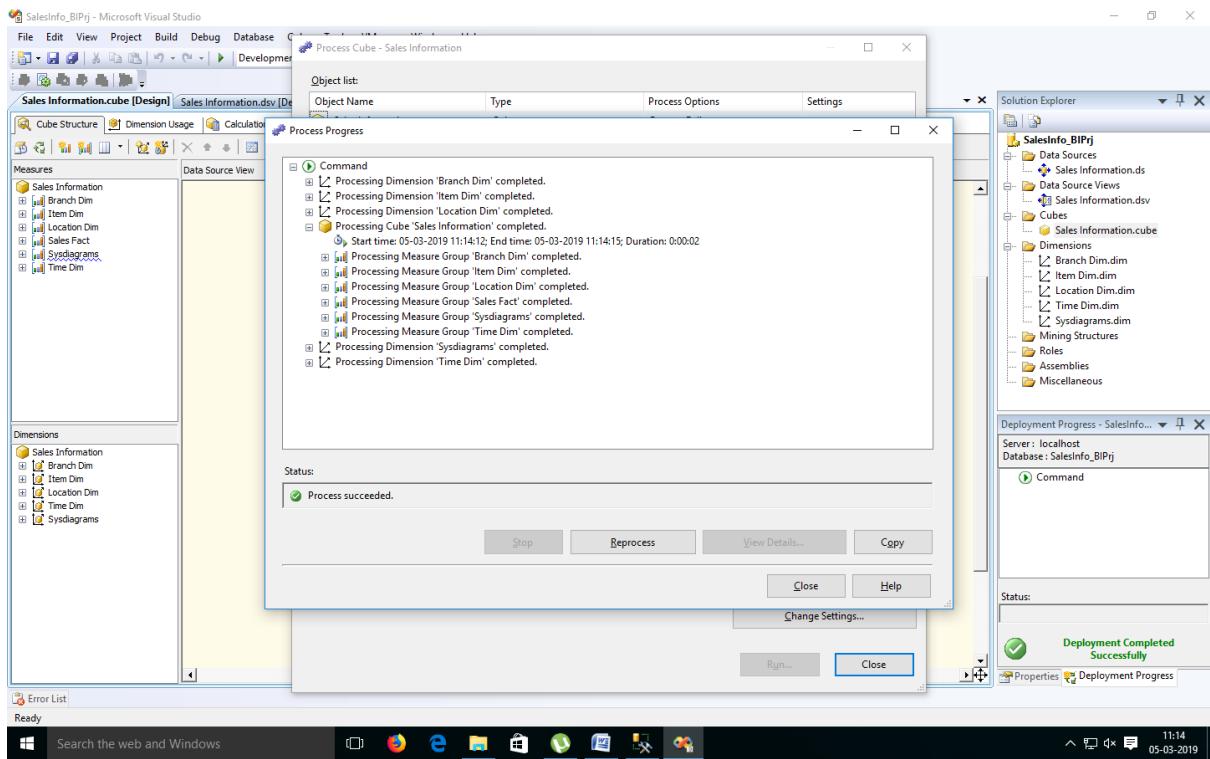


6. Finally, Process cube by Right click on SalesCube -> Process .





7. Click on Run.

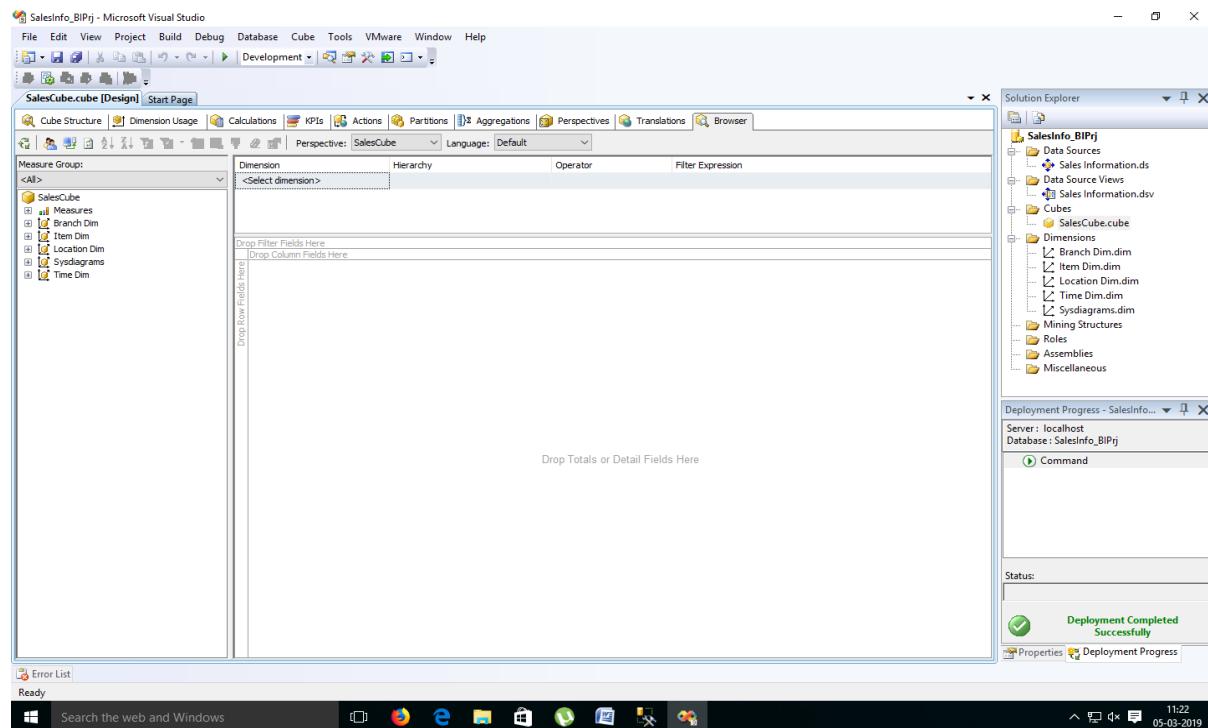


Practical No 5

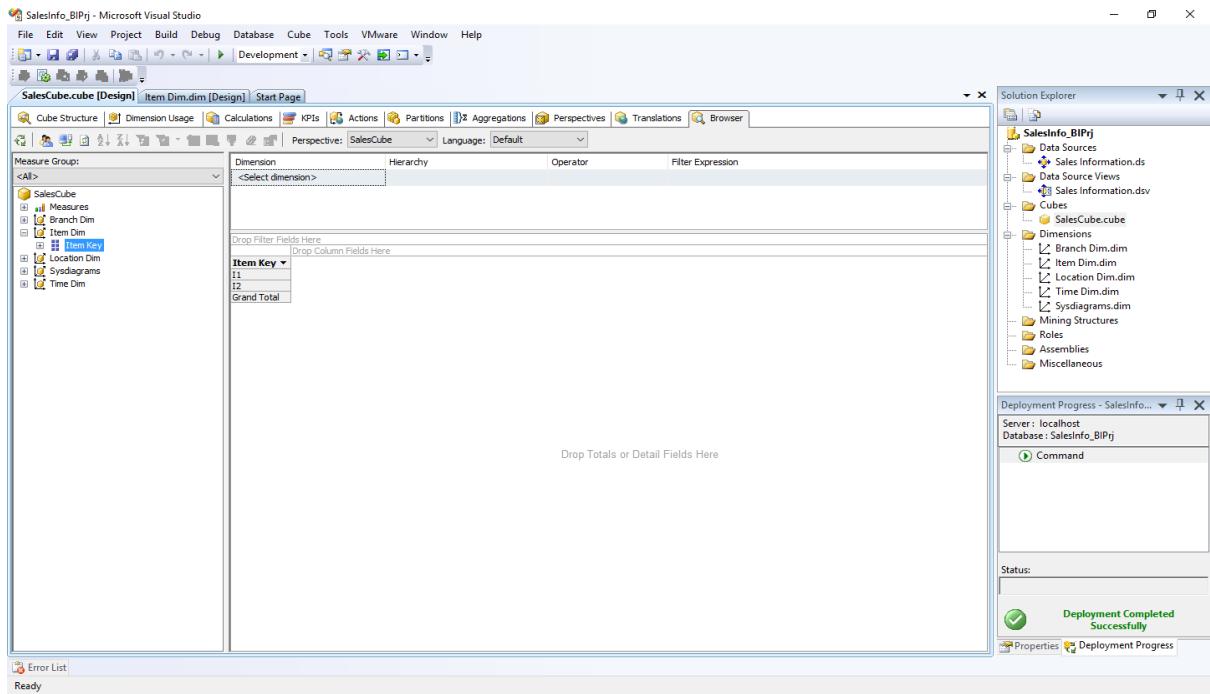
Aim : View cube data in multidimensional Format.

Solution :

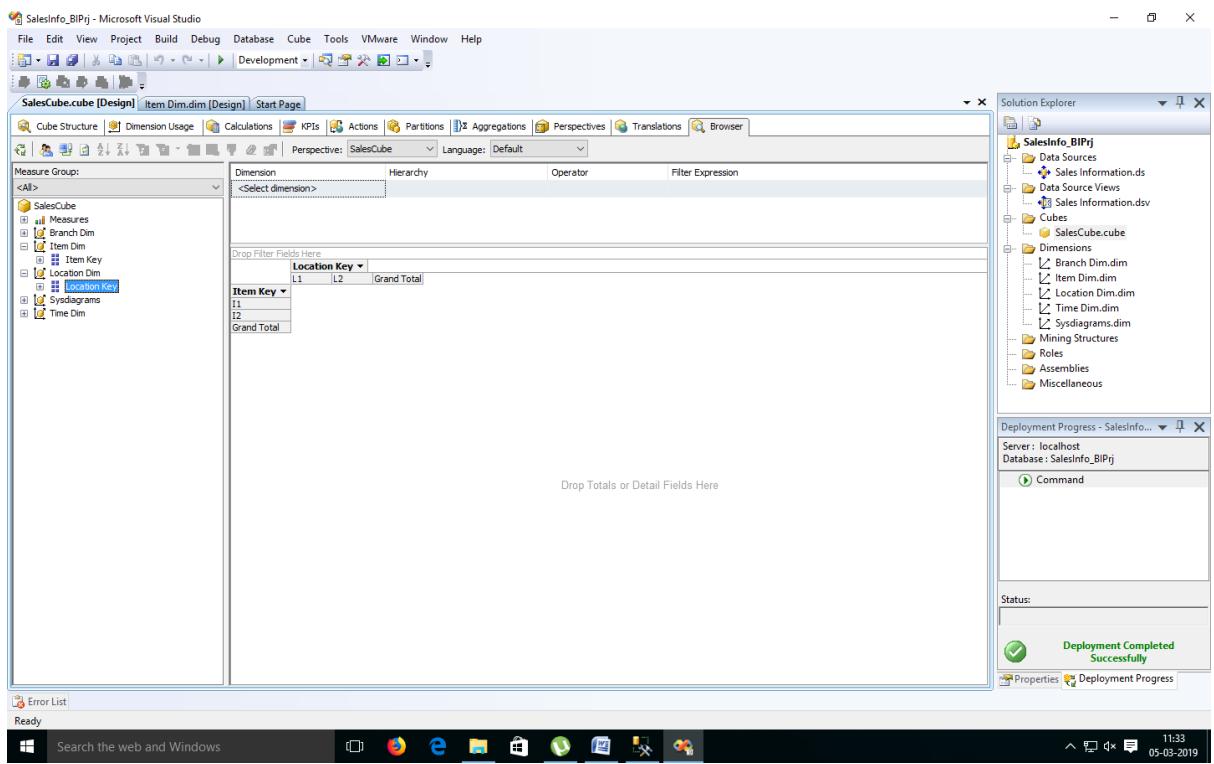
1. Double Click on “SalesCube”. Go to the “Browser” Tab.



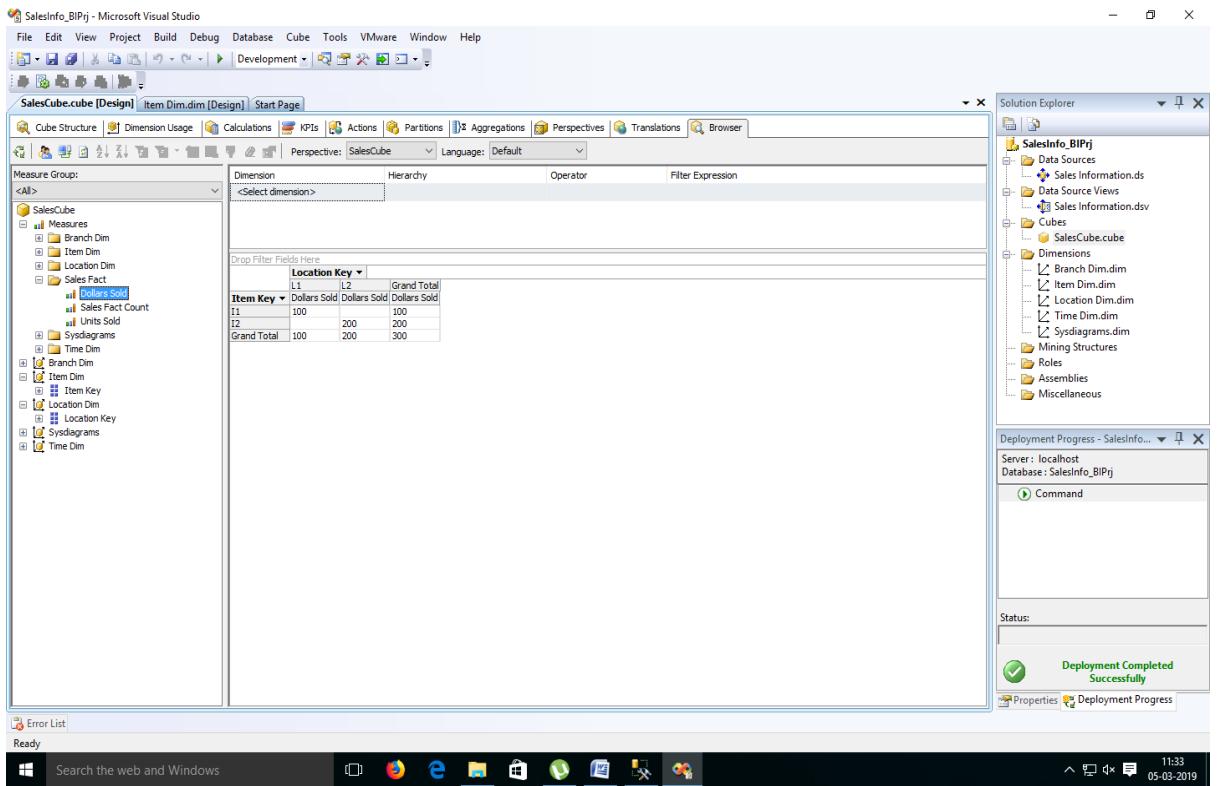
2. Go to the “Item Dimension”. Right Click on ‘Item Key’ -> Add to Row Area.



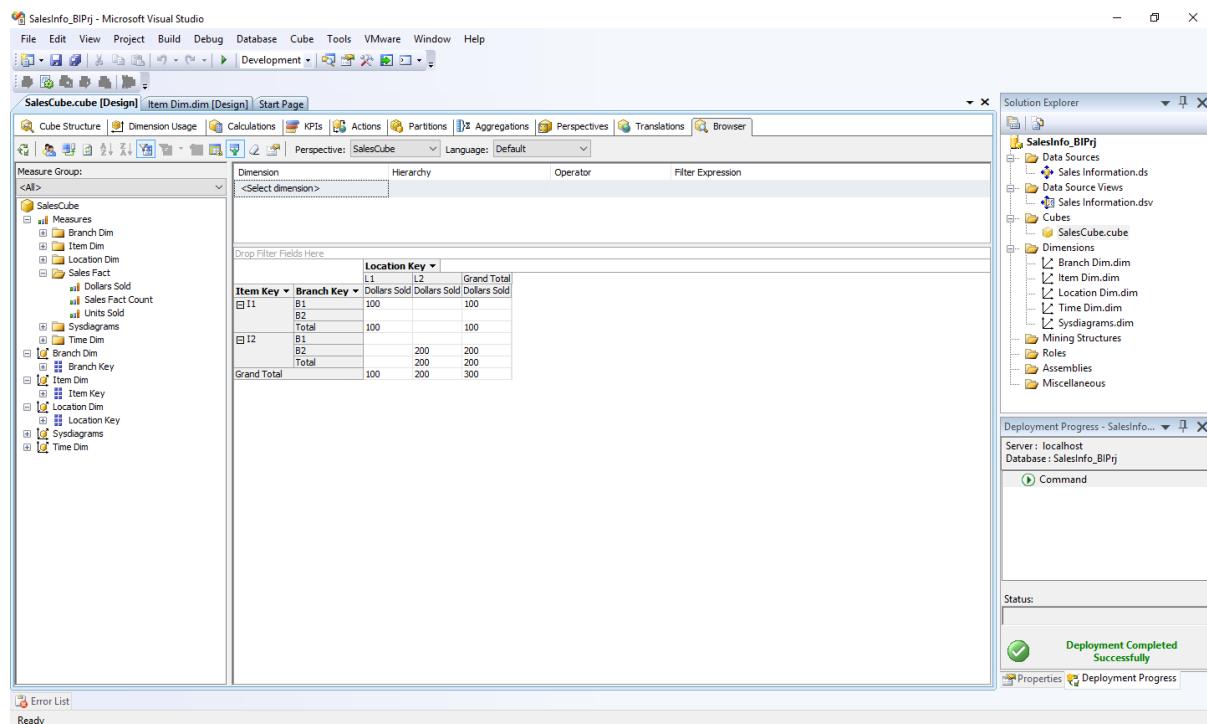
3. Go to the “Location Dimension”. Right Click on ‘Location Key’ -> Add to Column Area.



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



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

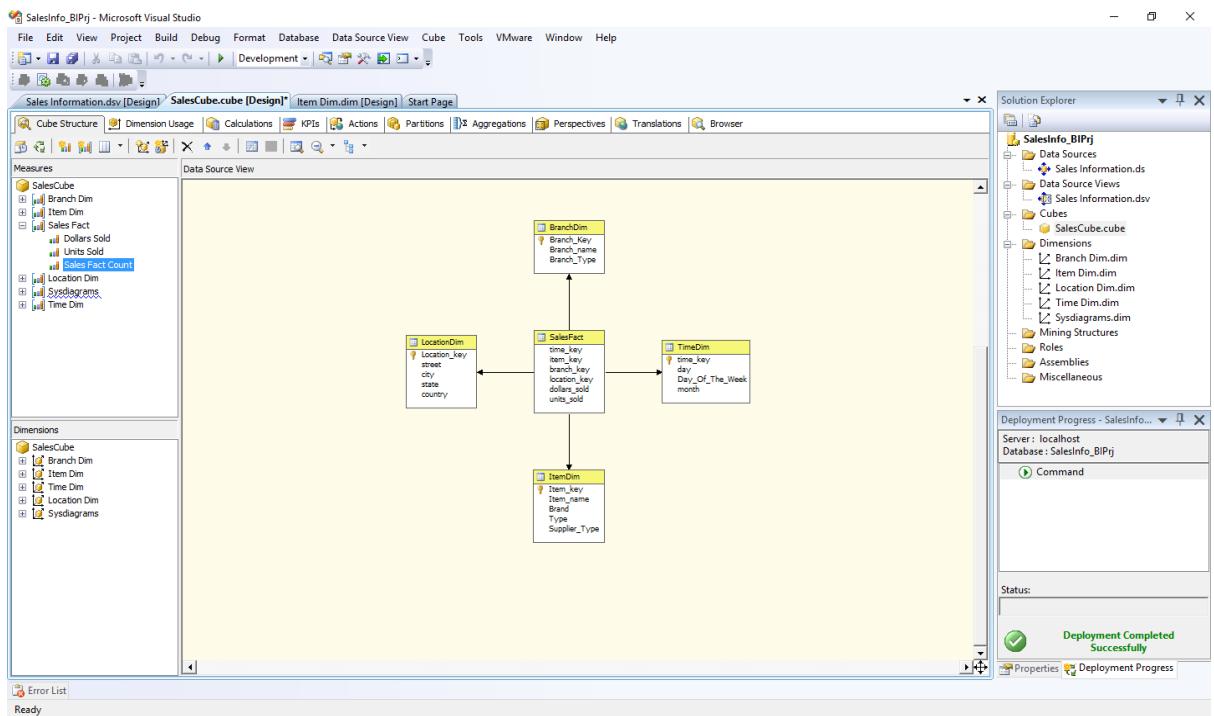


Practical No 6

Aim : Working with measures in the cube.

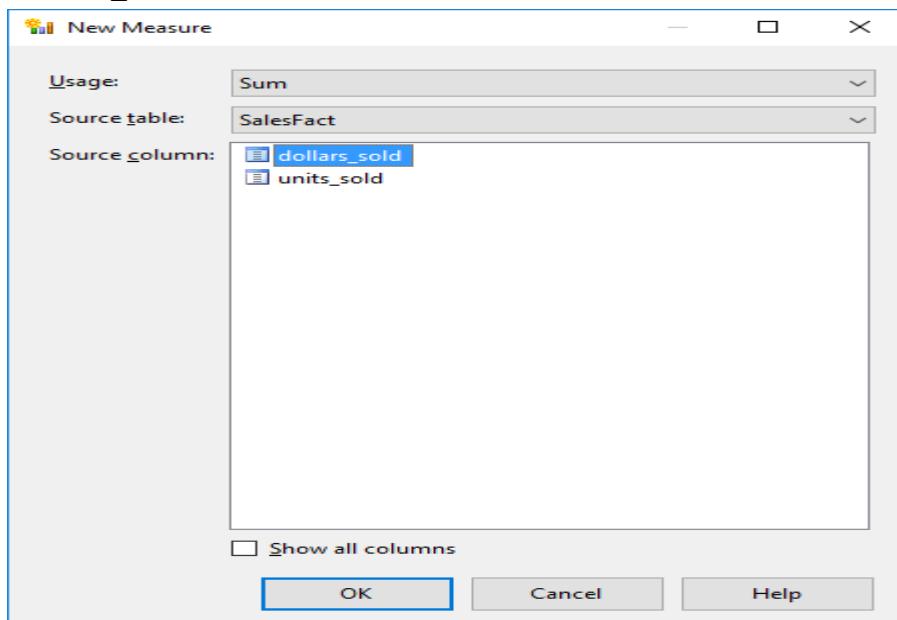
Solution :

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



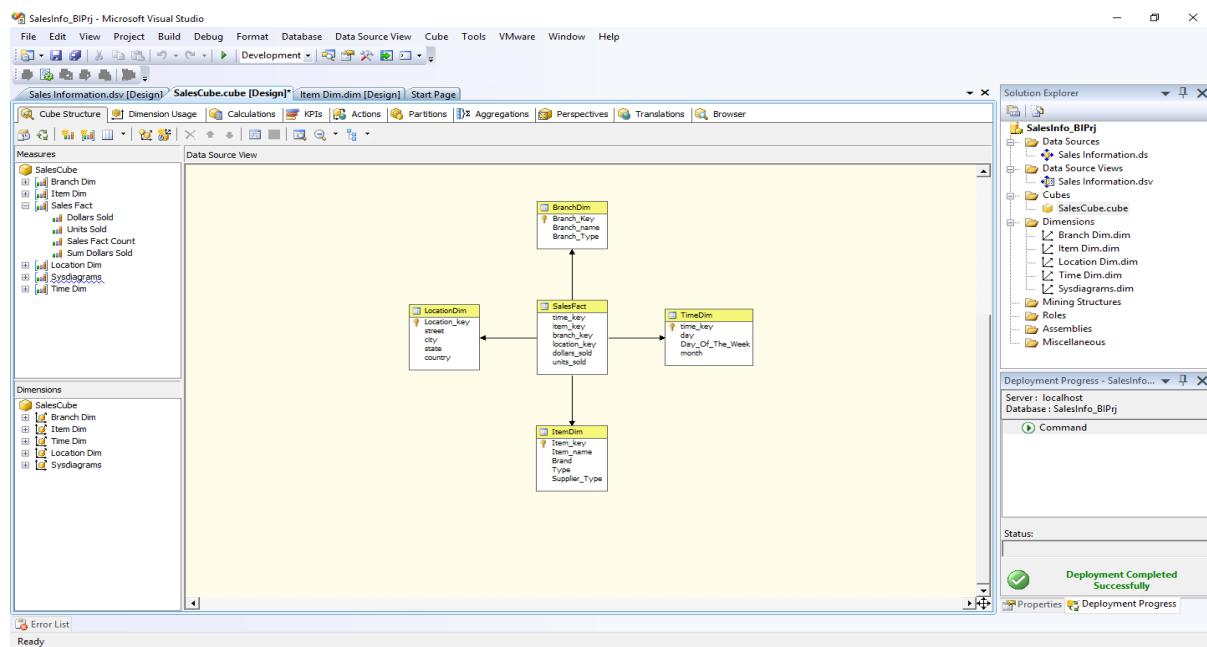
2. Right click on SalesCube -> New Measure.

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

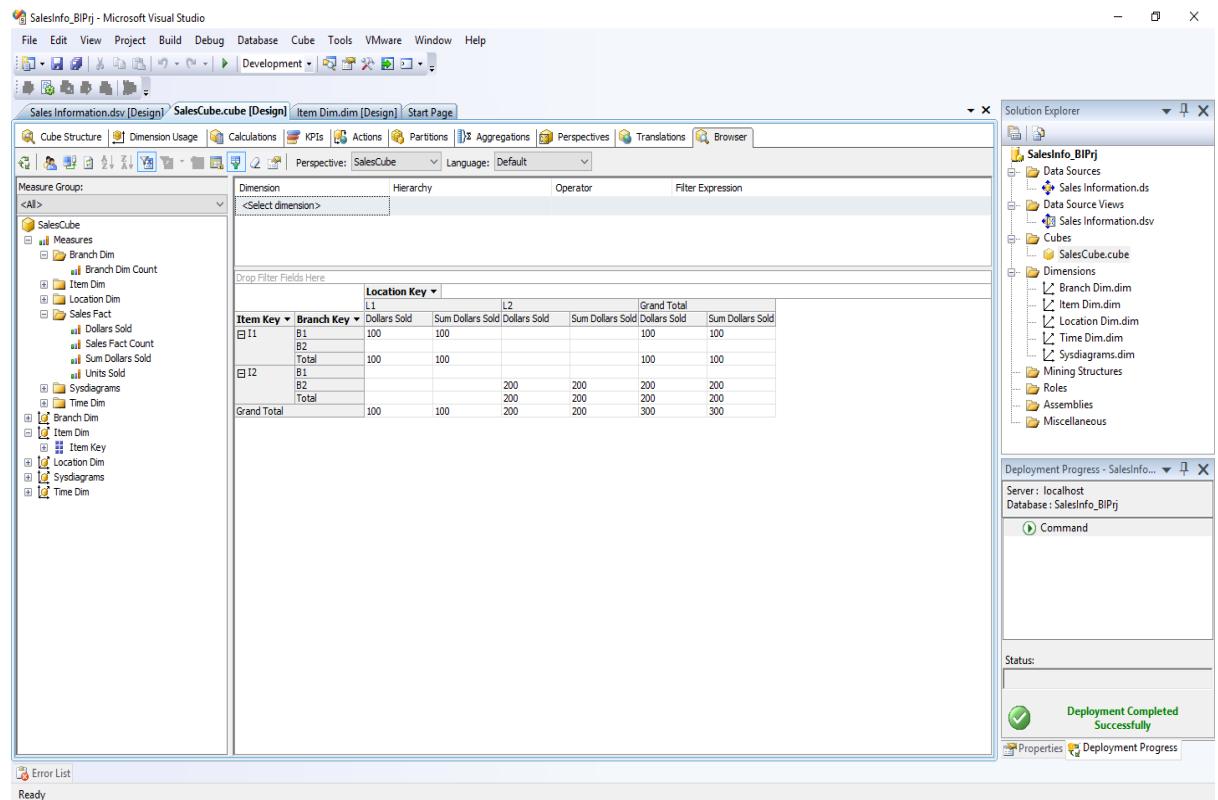


Click on OK.

3. Rename Measure as “Sum Dollars sold”.



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



Practical No 7

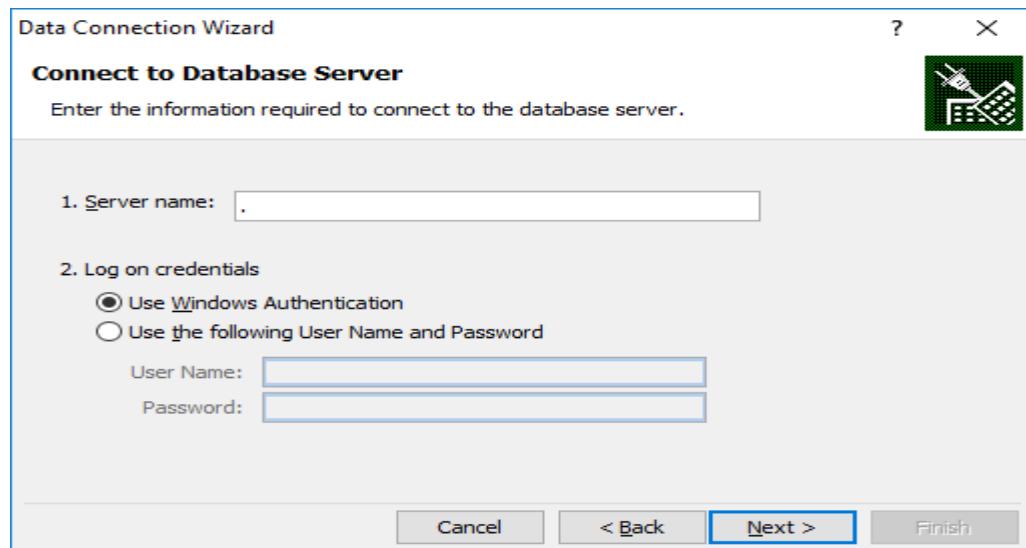
Aim : Creating an Excel Pivot Table and Pivot Chart by using the OLAP cube data.

Solution :

1. Open MS-Excel. Click on Data Menu.

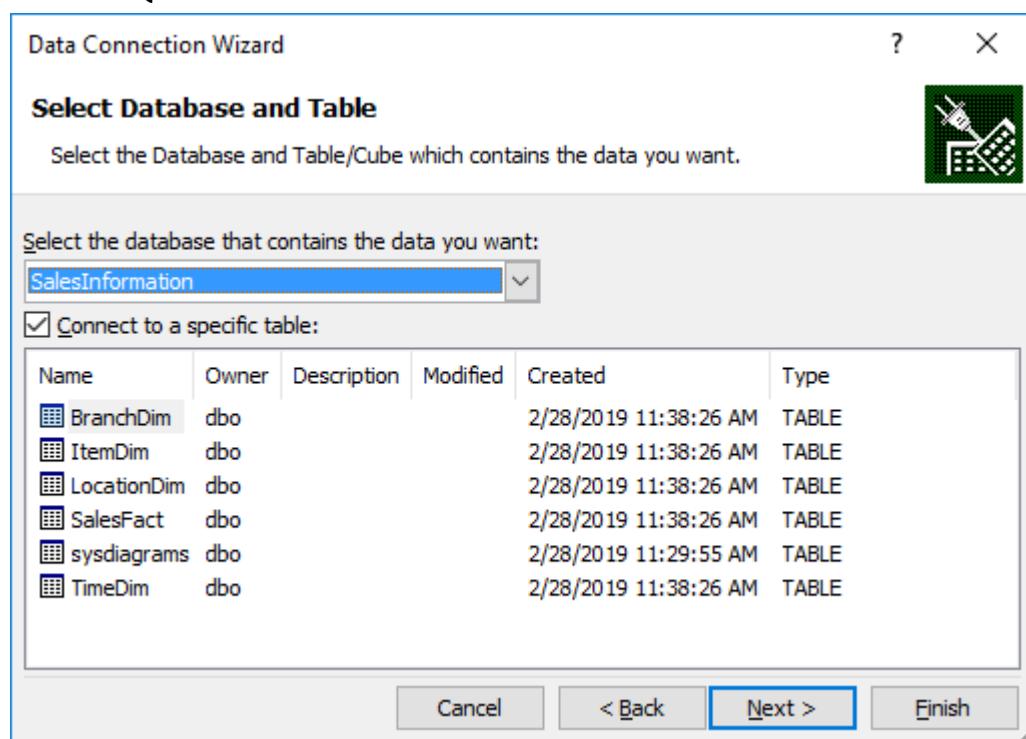
2. Go to From Other Sources.

2.1. From SQL Server -> Type Server name as “.”

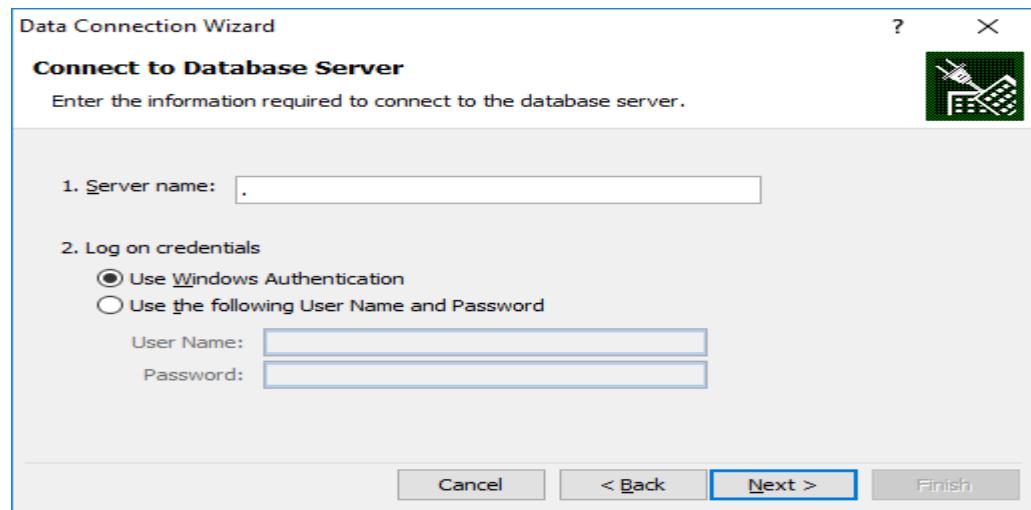


Click on Next.

Choose SQL Database -> “SalesInformation”

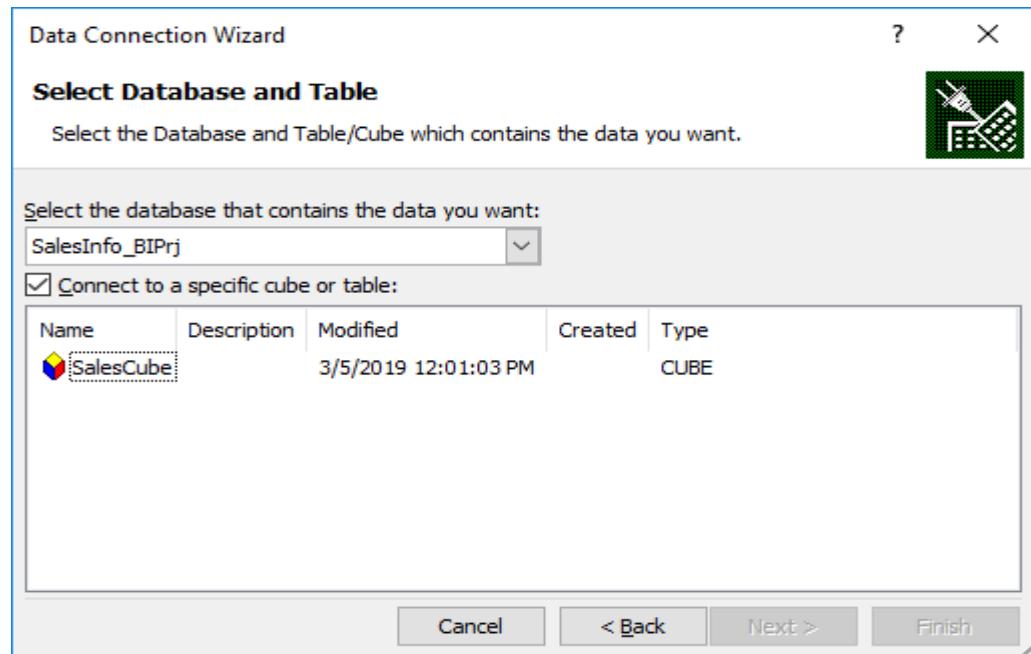


2.2. From Analysis Services -> Type Server name as “.”

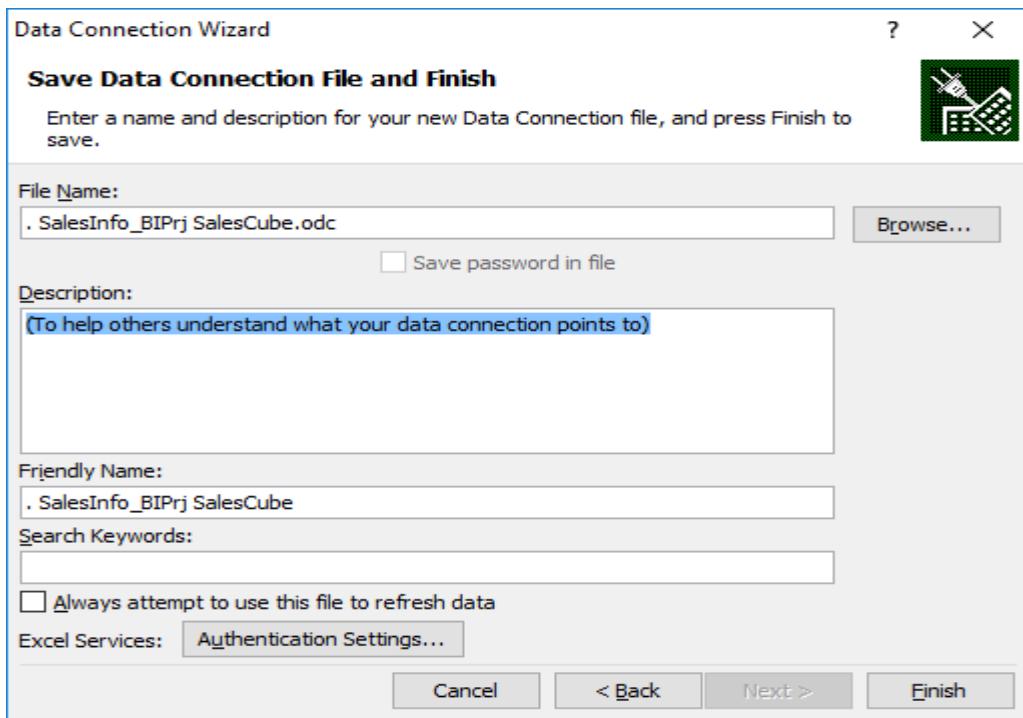


Click on Next.

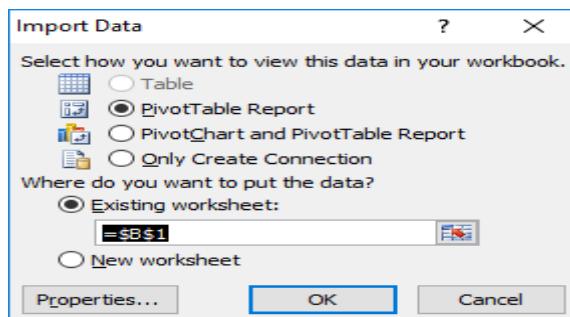
Choose Analysis Database as "SalesInfo_BIPrj". Click on Next.



Click on OK

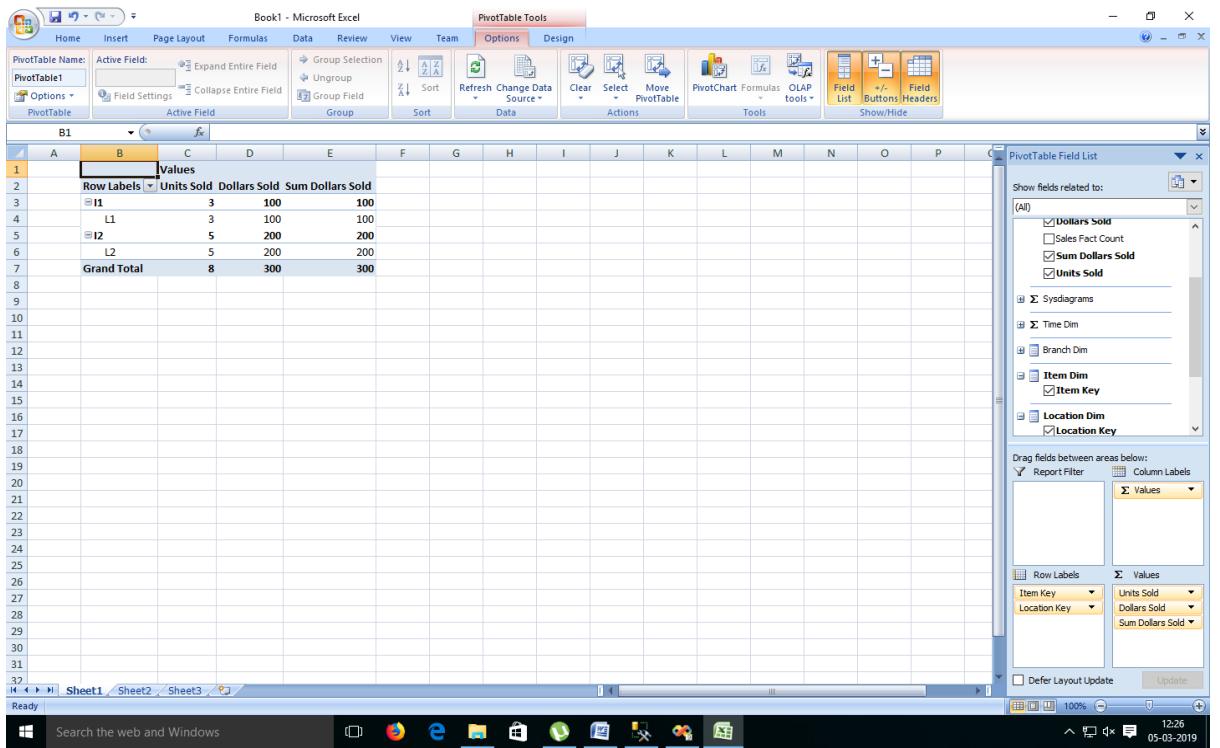


Click on Finish.

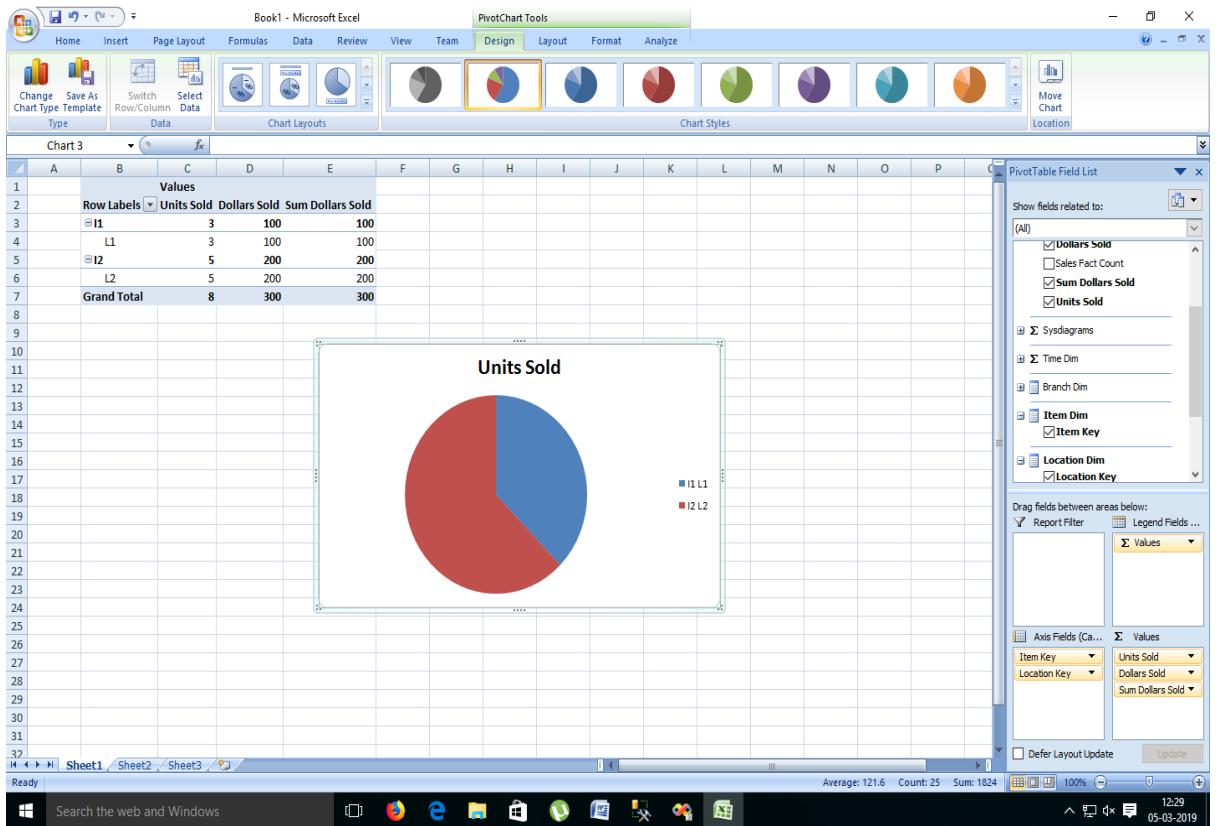


Click on OK.

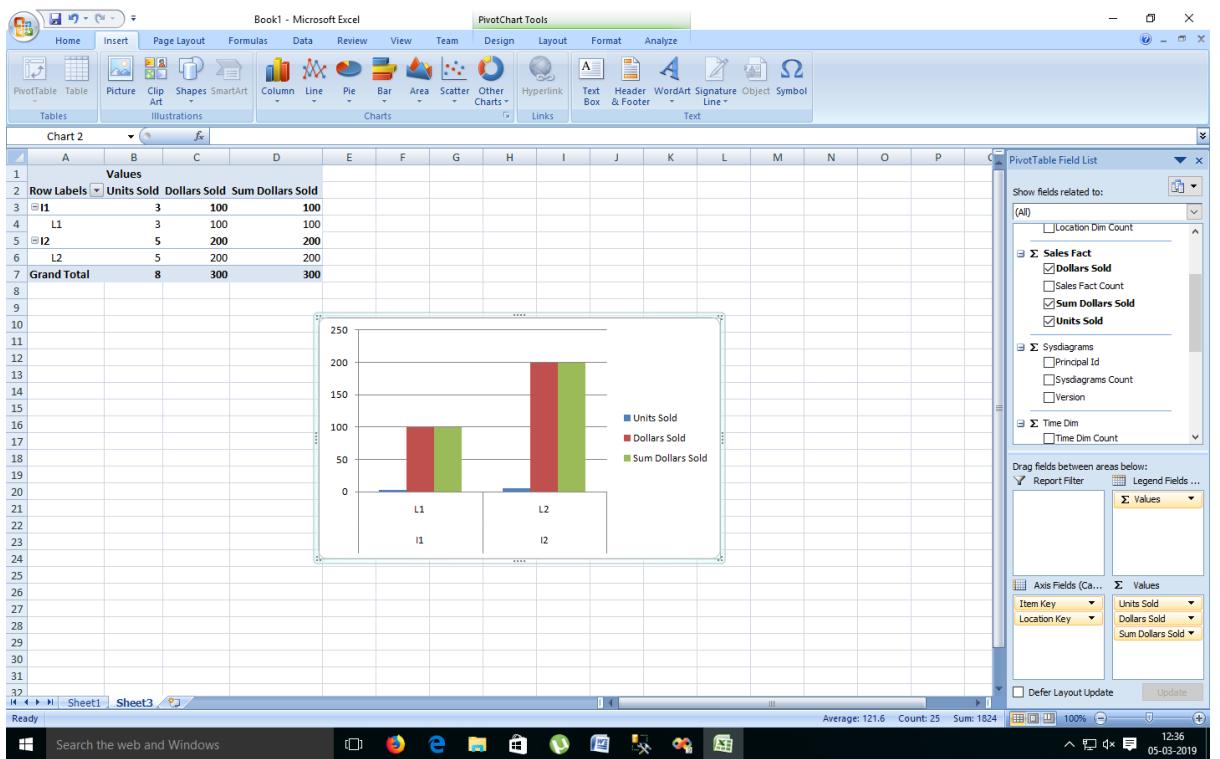
3. Select Item Key, Location Key and Measures as Dollars Sold, Units Sold and Sum Dollars Sold



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 'SalesInformation' -> Expand Tables.

3. Fire following queries :

**3.1. SELECT [Branch_Key], [Branch_name], [Branch_Type]
FROM [SalesInformation].[dbo].[BranchDim]**

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the 'SalesInformation' database is selected. In the center pane, a query window displays the following T-SQL code:

```
SELECT [Branch_Key]
      ,[Branch_name]
      ,[Branch_Type]
  FROM [SalesInformation].[dbo].[BranchDim]
```

The results pane shows the output of the query:

	Branch_Key	Branch_name	Branch_Type
1	B1	RJ	Vendor
2	B2	JR	Vendor

The status bar at the bottom indicates "Query executed successfully." The Properties pane on the right shows connection details for the current session.

**3.2. SELECT [Item_key], [Item_name], [Brand], [Type], [Supplier_Type]
FROM [SalesInformation].[dbo].[ItemDim]**

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the 'SalesInformation' database is selected. In the center pane, a query window displays the following T-SQL code:

```
SELECT [Item_key]
      ,[Item_name]
      ,[Brand]
      ,[Type]
      ,[Supplier_Type]
  FROM [SalesInformation].[dbo].[ItemDim]
```

The results pane shows the output of the query:

	Item_key	Item_name	Brand	Type	Supplier_Type
1	11	Laptop	LG	Accessories	VT
2	12	Mouse	LG	Accessories	VT

The status bar at the bottom indicates "Query executed successfully." The Properties pane on the right shows connection details for the current session.

3.3. `SELECT [Location_key], [street], [city], [state], [country]
FROM [SalesInformation].[dbo].[LocationDim]`

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the SalesInformation database and its tables like LocationDim, SalesFact, and TimeDim. The central pane displays the following T-SQL code:

```
***** Script for SelectTopNRows command from SSMS *****
SELECT [Location_key]
      ,[street]
      ,[city]
      ,[state]
      ,[country]
 FROM [SalesInformation].[dbo].[LocationDim]
```

The Results pane shows the output of the query:

Location_key	street	city	state	country
L1	LBS	Mumbai	Maha	India
L2	JM	Thane	Maha	India

The status bar at the bottom right indicates the query was executed successfully at 08-03-2019 09:57:38.

3.4. `SELECT [time_key], [item_key], [branch_key], [location_key]
, [dollars_sold], [units_sold]
FROM [SalesInformation].[dbo].[SalesFact]`

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the SalesInformation database and its tables like SalesFact, LocationDim, and ItemDim. The central pane displays the following T-SQL code:

```
***** Script for SelectTopNRows command from SSMS *****
SELECT [time_key]
      ,[item_key]
      ,[branch_key]
      ,[location_key]
      ,[dollars_sold]
      ,[units_sold]
 FROM [SalesInformation].[dbo].[SalesFact]
```

The Results pane shows the output of the query:

time_key	item_key	branch_key	location_key	dollars_sold	units_sold
T1	I1	B1	L1	100	3
T2	I2	B2	L2	200	5

The status bar at the bottom right indicates the query was executed successfully at 08-03-2019 09:58:08.

3.5. `SELECT [time_key], [day], [Day_Of_The_Week], [month]
FROM [SalesInformation].[dbo].[TimeDim]`

The screenshot shows the Microsoft SQL Server Management Studio interface. The query window contains the following SQL code:

```
/*
***** Script for SelectTopNRows command from SSMS *****/
SELECT [time_key]
      ,[day]
      ,[Day_Of_The_Week]
      ,[month]
 FROM [SalesInformation].[dbo].[TimeDim]
```

The results pane displays two rows of data:

	time_key	day	Day_Of_The_Week	month
1	T1	2018-02-01 00:00:00.000	Monday	2018-02-01 00:00:00.000
2	T2	2019-03-03 00:00:00.000	Tuesday	2019-03-03 00:00:00.000

The status bar at the bottom right indicates the session started at 08-03-2019 10:00:08 and ended at 08-03-2019 10:00:08.

3.6. `SELECT [SalesInformation].[dbo].[BranchDim].[Branch_Key],
[Branch_name], [dollars_sold], [units_sold]
FROM [SalesInformation].[dbo].[BranchDim],
[SalesInformation].[dbo].[SalesFact]
where [SalesInformation].[dbo].[BranchDim].[Branch_Key]=
[SalesInformation].[dbo].[SalesFact].[Branch_Key];`

The screenshot shows the Microsoft SQL Server Management Studio interface. The query window contains the following SQL code:

```
SELECT [SalesInformation].[dbo].[BranchDim].[Branch_Key],  
[Branch_name],  
[dollars_sold],  
[units_sold]  
FROM [SalesInformation].[dbo].[BranchDim], [SalesInformation].[dbo].[SalesFact]  
where [SalesInformation].[dbo].[BranchDim].[Branch_Key]=  
[SalesInformation].[dbo].[SalesFact].[Branch_Key];
```

The results pane displays two rows of data:

	Branch_Key	Branch_name	dollars_sold	units_sold
1	B1	RJ	100	3
2	B2	JR	200	5

The status bar at the bottom right indicates the session started at 08-03-2019 10:07:23 and ended at 08-03-2019 10:07:23.

```

3.7. SELECT [SalesInformation].[dbo].[ItemDim].[Item_Key], [item_Name]
      , [Type], [dollars_sold], [units_sold]
   FROM [SalesInformation].[dbo].[ItemDim],
        [SalesInformation].[dbo].[SalesFact]
  Where [SalesInformation].[dbo].[ItemDim].[Item_key]=
    [SalesInformation].[dbo].[SalesFact].[item_key];

```

The screenshot shows the Microsoft SQL Server Management Studio interface. The query window contains the following SQL code:

```

SELECT [SalesInformation].[dbo].[ItemDim].[Item_Key], [item_Name]
      , [Type], [dollars_sold], [units_sold]
   FROM [SalesInformation].[dbo].[ItemDim],
        [SalesInformation].[dbo].[SalesFact]
  Where [SalesInformation].[dbo].[ItemDim].[Item_key]=
    [SalesInformation].[dbo].[SalesFact].[item_key];

```

The results pane displays the following data:

Item_Key	item_Name	Type	dollars_sold	units_sold
11	Laptop	Accessories	100	3
12	Mouse	Accessories	200	5

The status bar at the bottom indicates "Query executed successfully." and shows connection details: (local) (10.50 RTM) | BSC-37\admin (56) | master | 00:00:00 | 2 rows.

```

3.8. SELECT [SalesInformation].[dbo].[LocationDim].[Location_key]
      , [city], [item_Key], [dollars_sold], [units_sold]
   FROM [SalesInformation].[dbo].[LocationDim],
        [SalesInformation].[dbo].[SalesFact]
  Where [SalesInformation].[dbo].[LocationDim].[Location_key]=
    [SalesInformation].[dbo].[SalesFact].[location_key];

```

The screenshot shows the Microsoft SQL Server Management Studio interface. The query window contains the following SQL code:

```

SELECT [SalesInformation].[dbo].[LocationDim].[Location_key]
      , [city], [item_Key], [dollars_sold], [units_sold]
   FROM [SalesInformation].[dbo].[LocationDim],
        [SalesInformation].[dbo].[SalesFact]
  Where [SalesInformation].[dbo].[LocationDim].[Location_key]=
    [SalesInformation].[dbo].[SalesFact].[location_key];

```

The results pane displays the following data:

Location_key	city	item_Key	dollars_sold	units_sold
L1	Mumbai	11	100	3
L2	Thane	12	200	5

The status bar at the bottom indicates "Query executed successfully." and shows connection details: (local) (10.50 RTM) | BSC-37\admin (56) | master | 00:00:00 | 2 rows. Additionally, there is a message in the output pane:

```

Auto-attach :0 process '1044' [SQL] (local) on machine '(local)' succeeded.
The thread '(local) (56)' (0x1650) has exited with code 0 (0x0).
The thread '(local) (56)' (0x1650) has waited with code 0 (0x0).
The program '1044' [SQL] (local) has exited with code 0 (0x0).

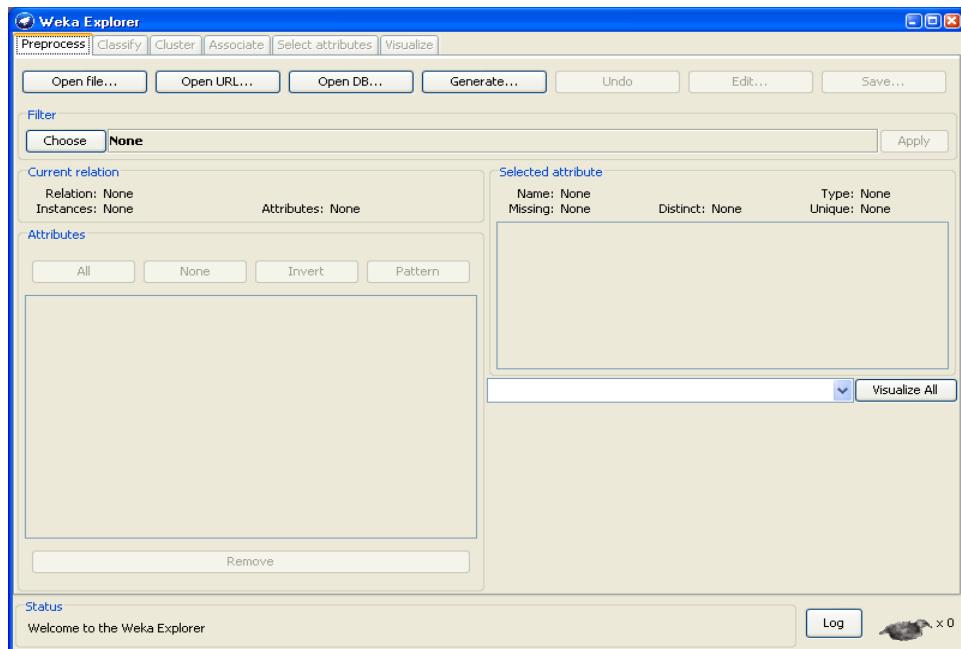
```

Practical No - 10

Aim : Data PreProcessing

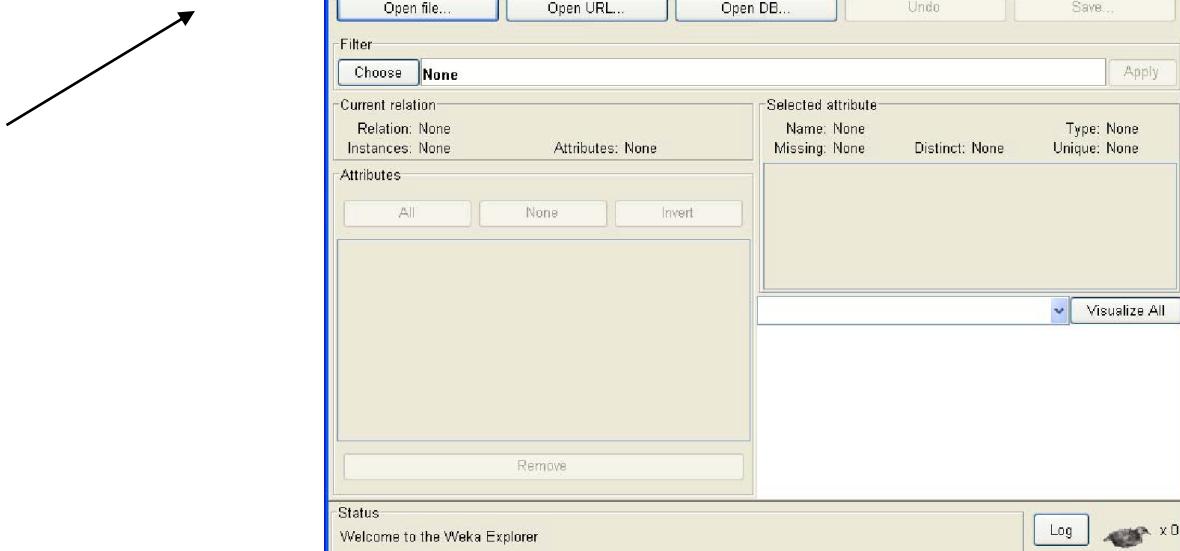
Solution :

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



Opening file from a local file system

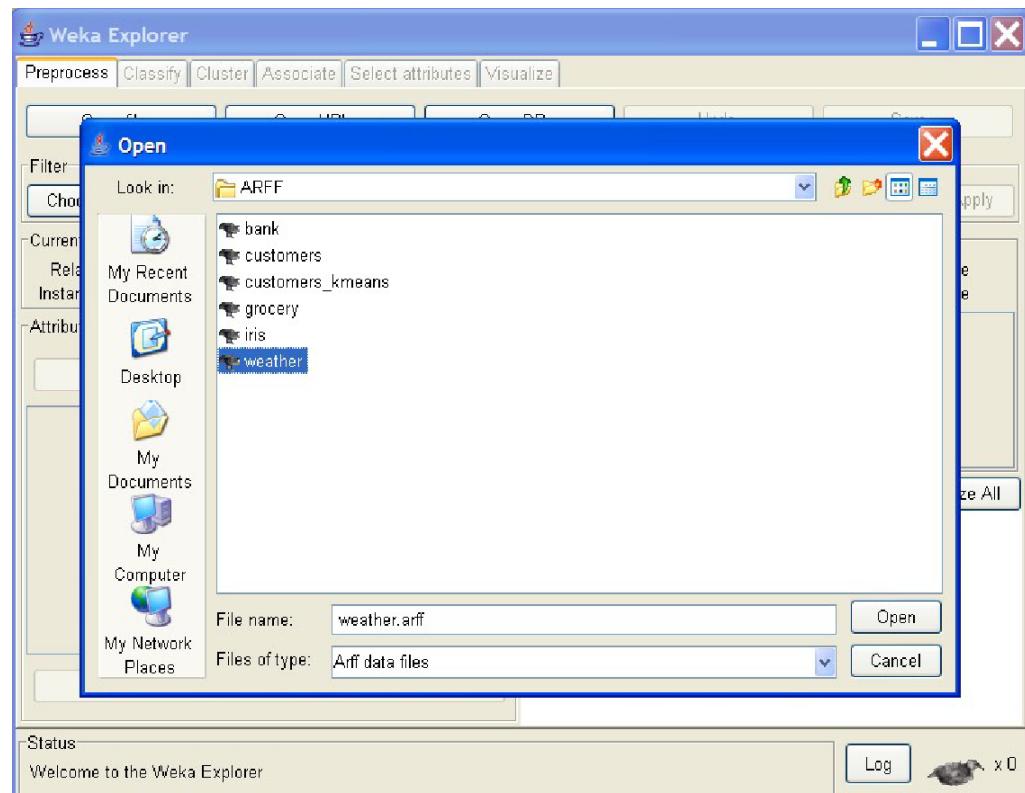
Click on ‘Open file...’ button



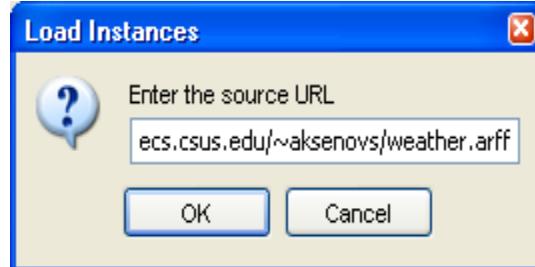
It brings up a dialog box allowing you to browse for the data file on the local file system, choose

“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:

The screenshot shows a web browser window titled "CSC 288" with the URL <http://gaia.ecs.csus.edu/~akzenovs>. The page content includes:

- Assignments:**
 - [Assignment 1 - Vocabulary and Case Study](#)
 - [Assignment 2 - Classification](#)
 - [Assignment 3 - Clustering](#)
 - [Assignment 4 - Association Rules](#)
 - [Assignment 5 - Credibility](#)
- Paper Review:**
 - [Summary](#)
 - [PowerPoint Presentation](#)
- Term Project:**
 - [Project Poposal](#)
 - [Project Progress Report](#)
 - [Project Progress Report Presentation](#)
- [WEKA Tutorial](#)
- [WEKA Tutorial Presentation](#)
- [weather.arff](#)

Below the browser window is a separate application window titled "SQL-Viewer". The "Query" tab contains the SQL command:

```
select * from stud
```

The "Result" tab displays the following table:

Row	name	hobby
1	dfgd	dfgd
2	drgdf	dfgd
3	dfg	dfg

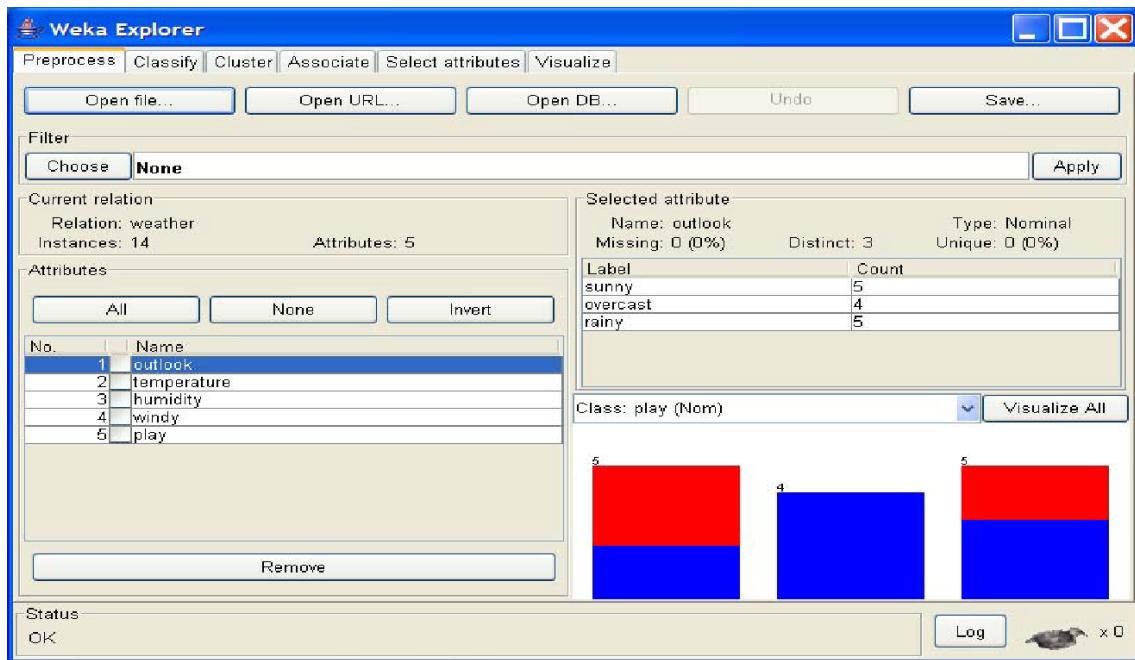
The "Info" tab shows the history of queries:

```
y: select * from stud  
is selected.  
y: select * from stud  
is selected.
```

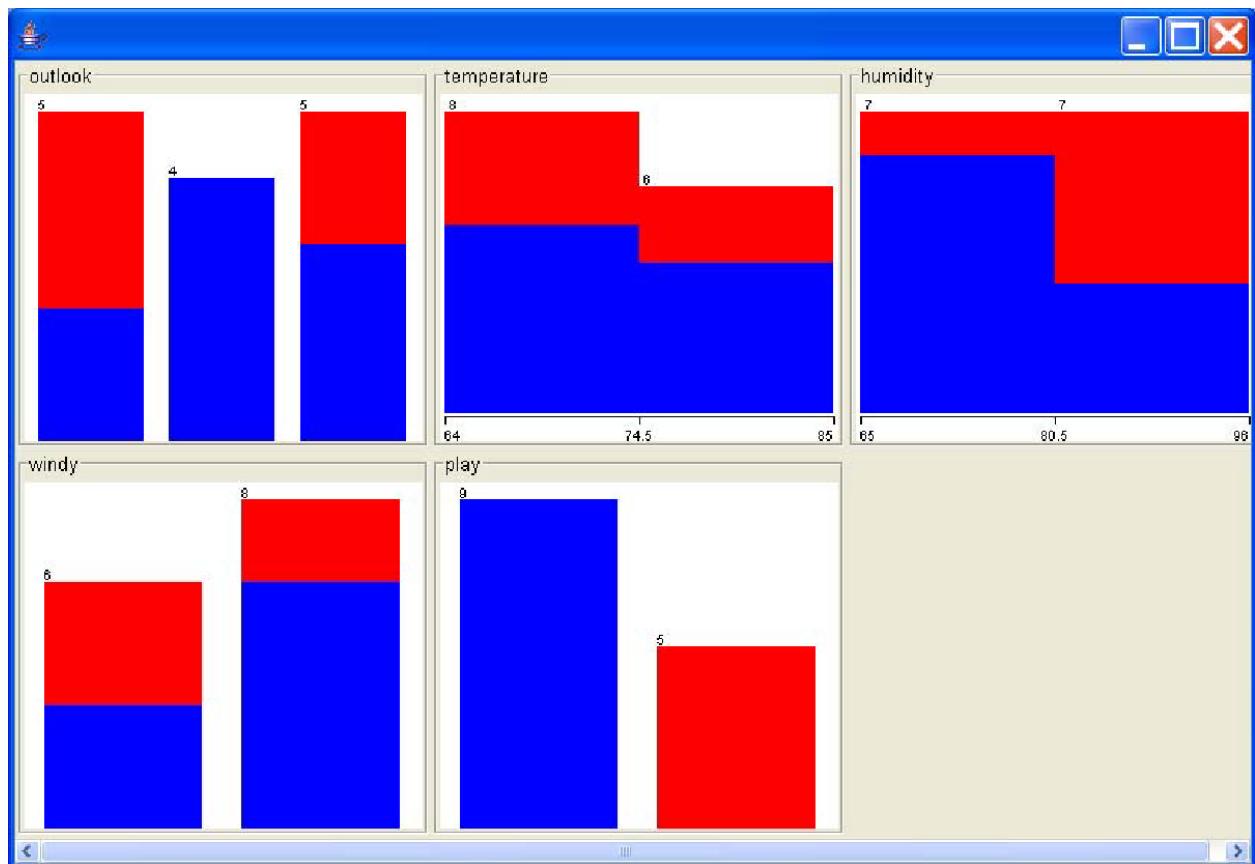
Buttons at the bottom right of the SQL-Viewer window include "Close", "Close all", "Re-use query", "Optimal width", "Clear", and "Copy".

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.,

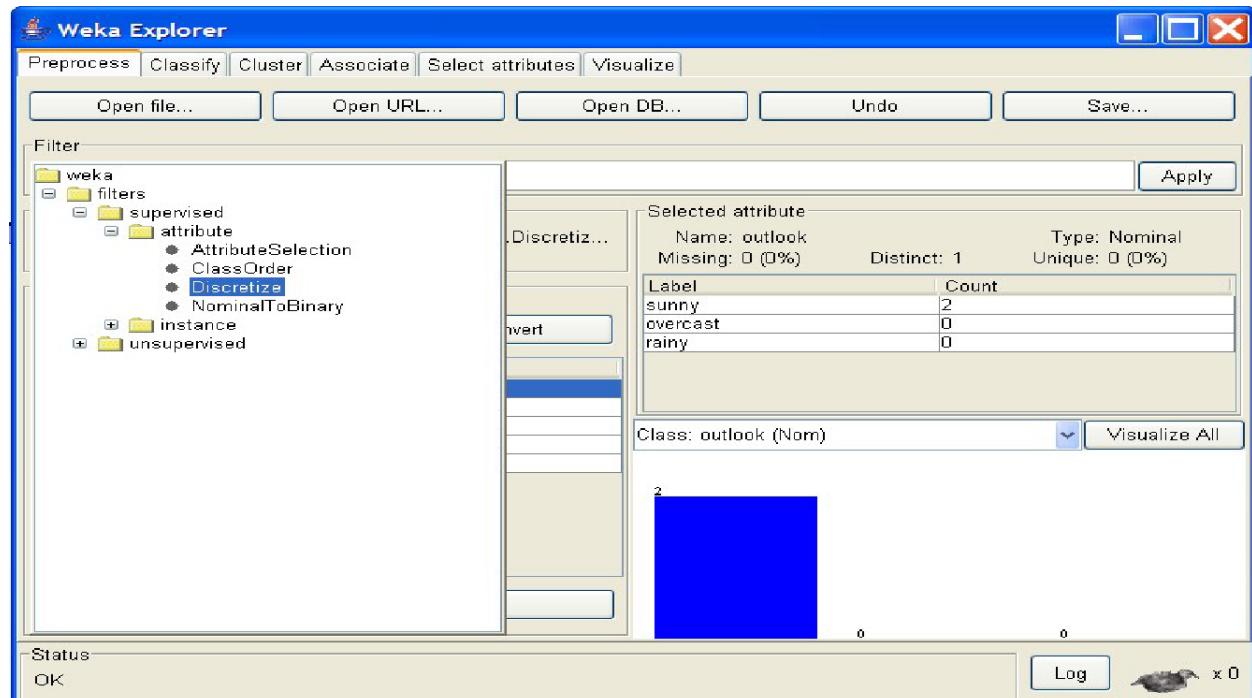
Practical No – 11

Aim : Data discretization.

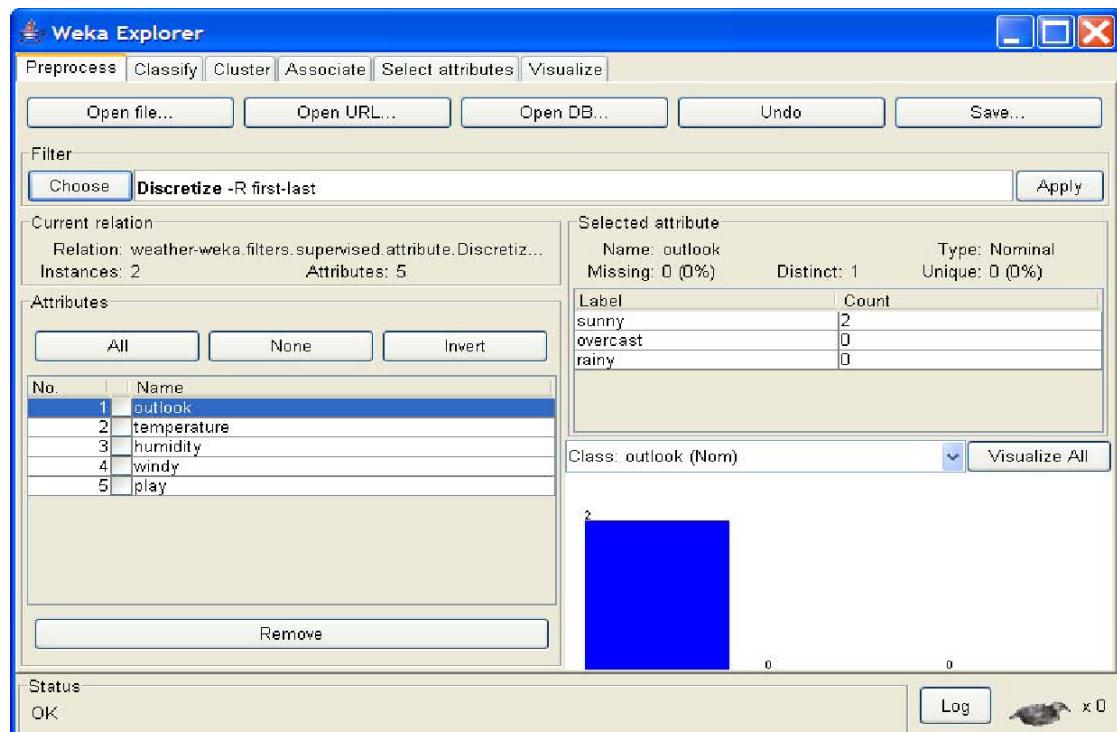
Solution :

In ‘Filters’ window, click on the ‘Choose’ button.

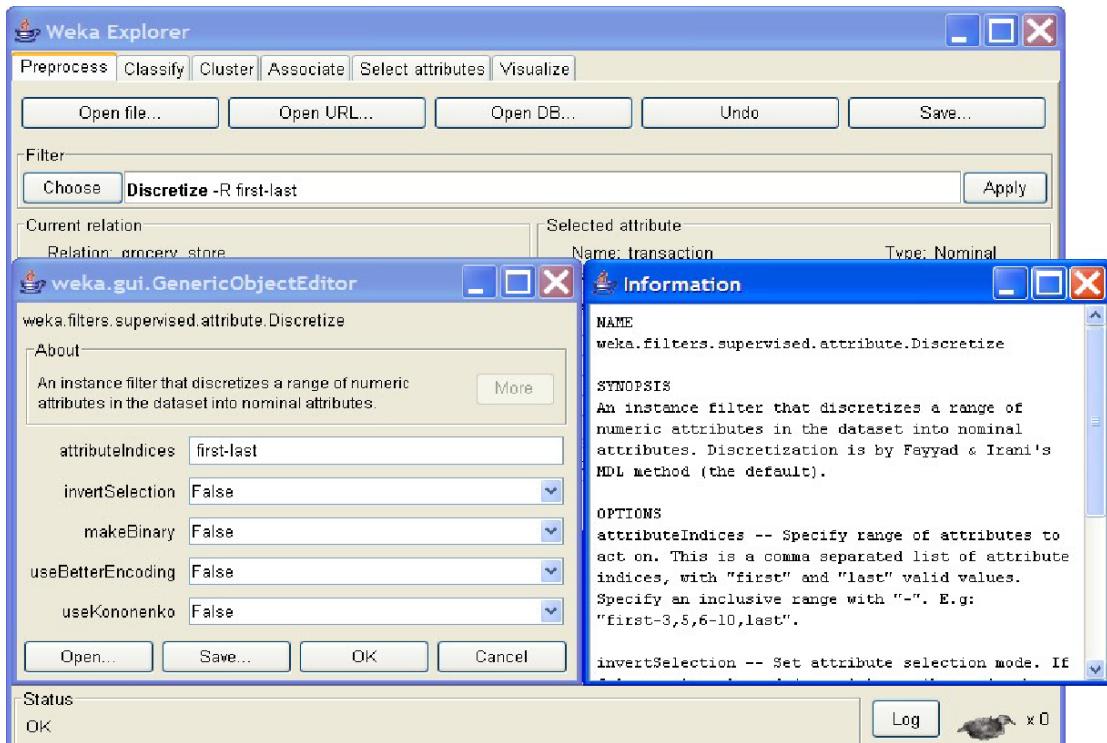
This will show pull-down menu with a list of available filters. Select Supervised Attribute Discretize and click on ‘Apply’ button.



The filter will convert Numeric values into Nominal.
the fields in the window changes to reflect available options.



a ‘GenericObjectEditor’ dialog box comes up on your screen.
The box lets you to choose the filter configuration options.

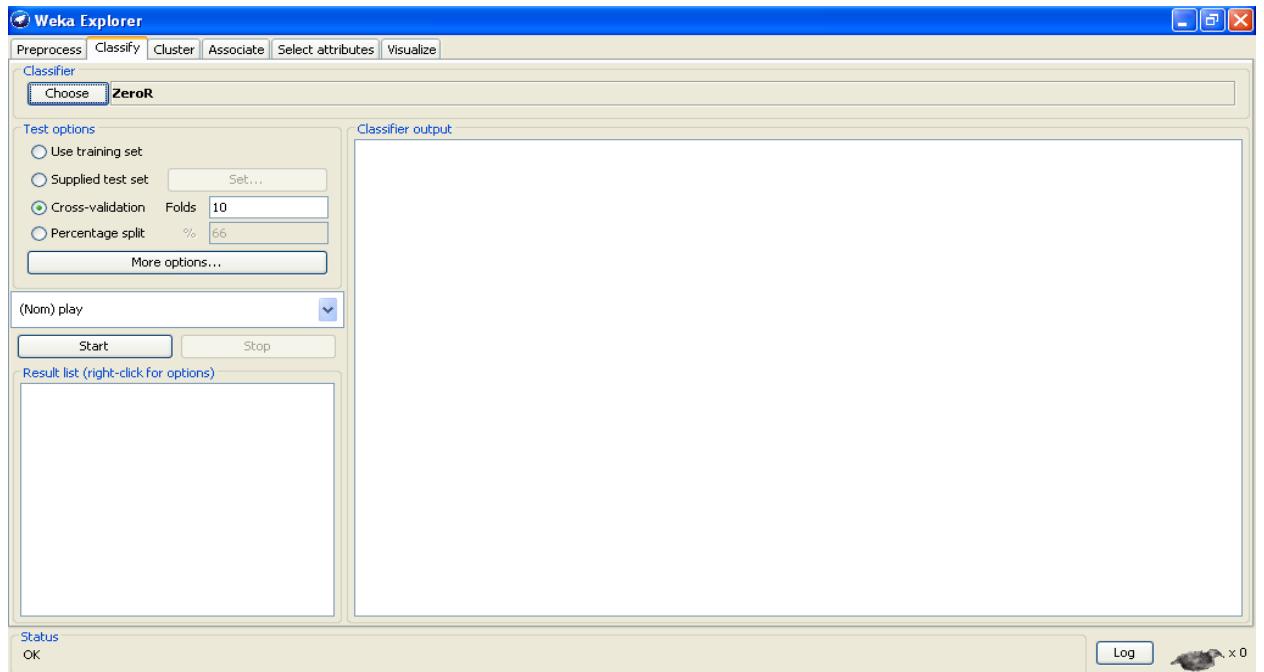


Practical No - 12

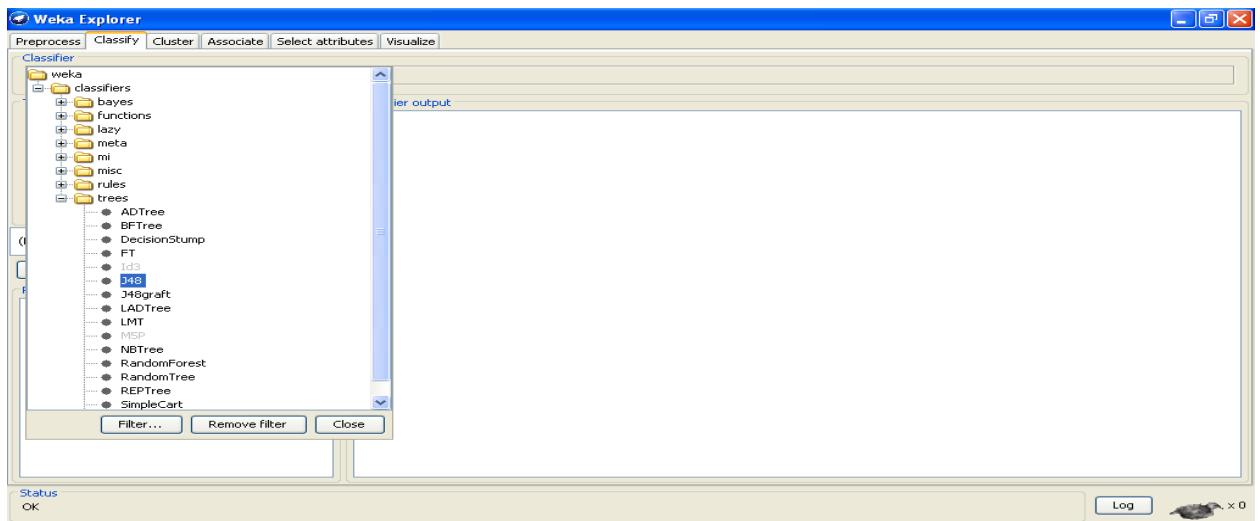
Aim : Classification problems.

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.

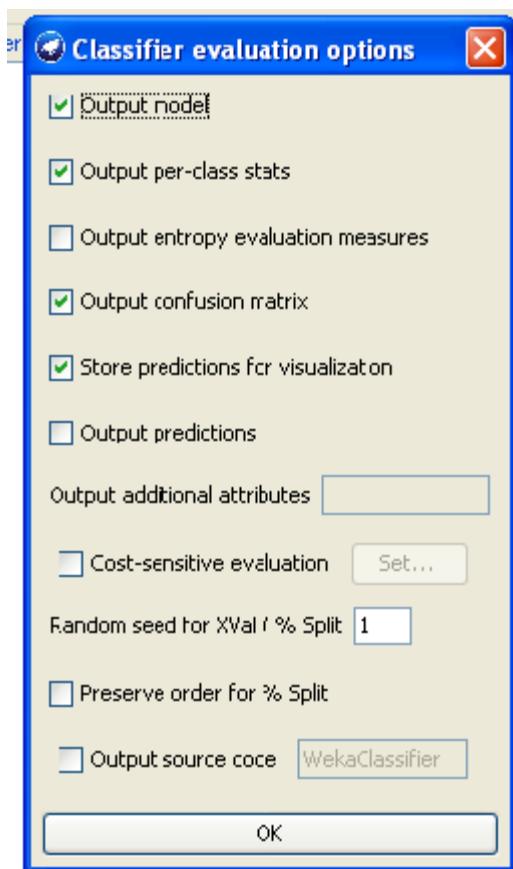


Check ‘Percentage split’ radio-button and keep it as default 66%. Click on ‘More options...’ button.

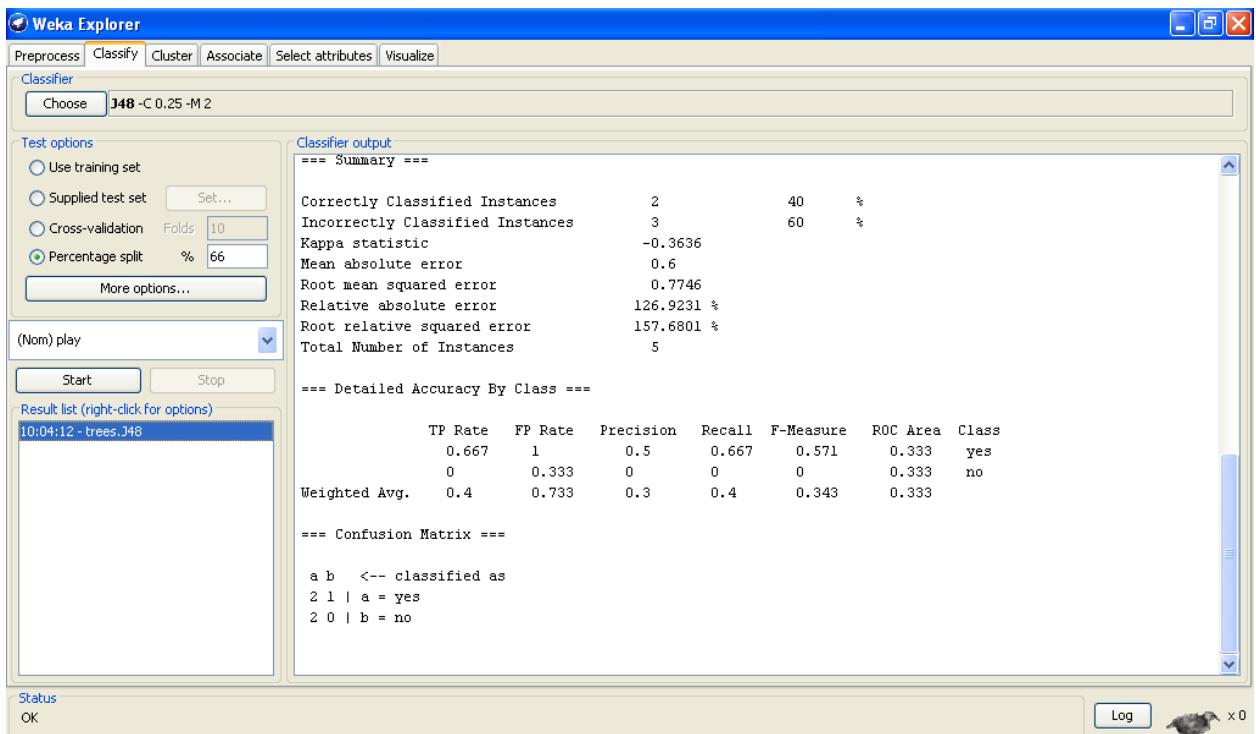
make sure that the Following options are checked :

1. Output model.

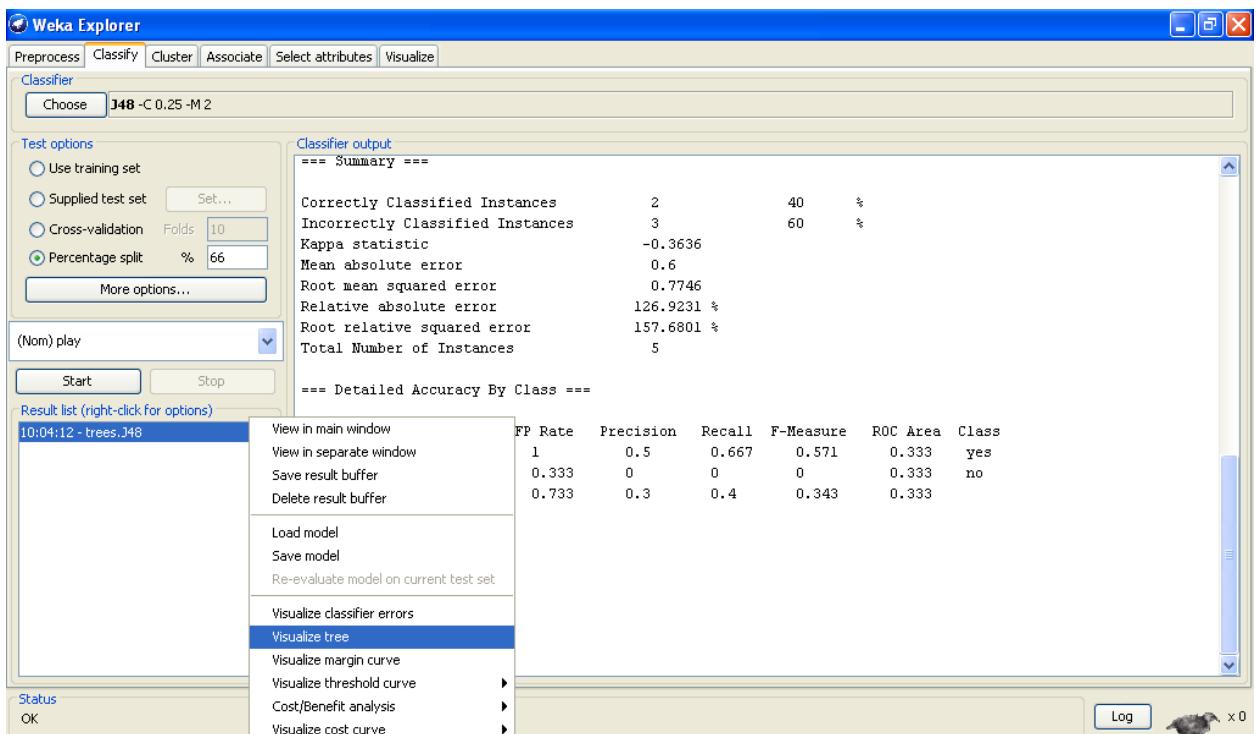
2. Output per-class stats.
3. Output confusion matrix
4. Store predictions for visualization.
5. Set ‘Random seed for Xval / % Split’ to 1.



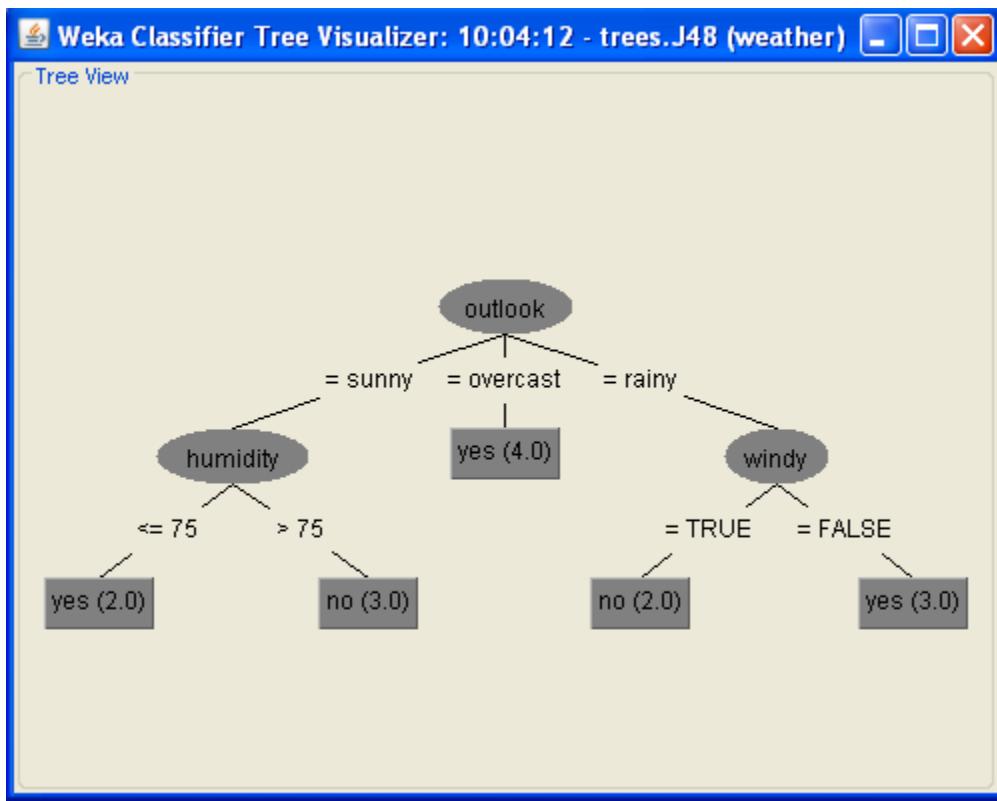
Once the options have been specified, you can run the classification algorithm. Click on ‘Start’ button



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.

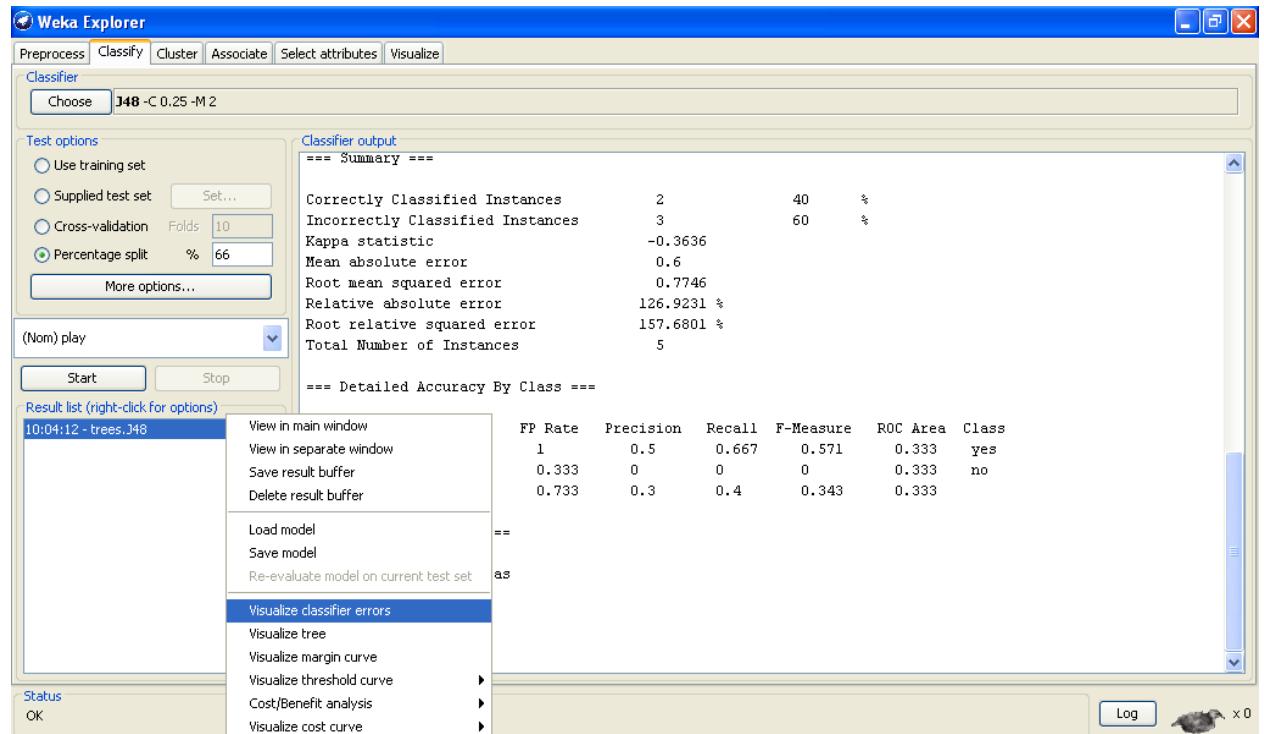


Select the item ‘Visualize tree’; a new window comes up to the screen displaying the tree.

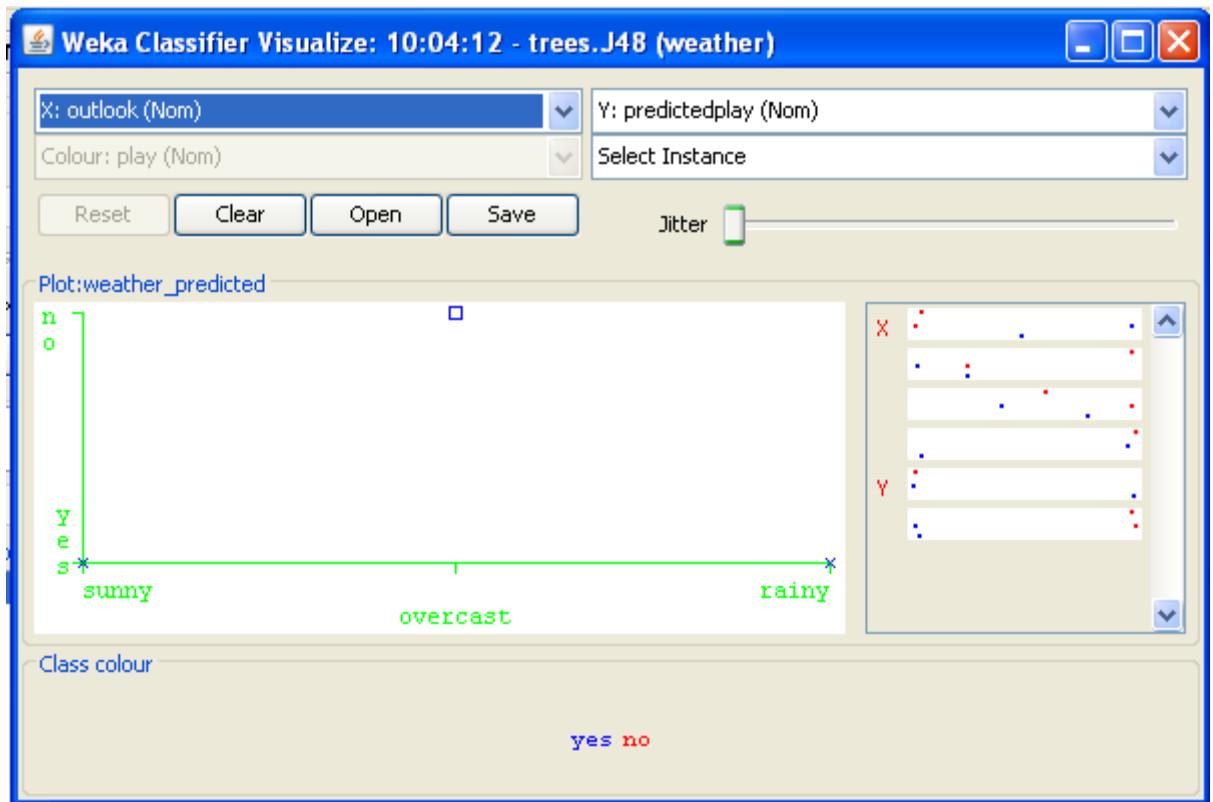


To visualize classification errors. Right-click on the entry in ‘Result list’ again

and select ‘Visualize classifier errors’ from the menu:



‘Visualize’ window displaying graph appears on the screen.



Practical N0 - 13

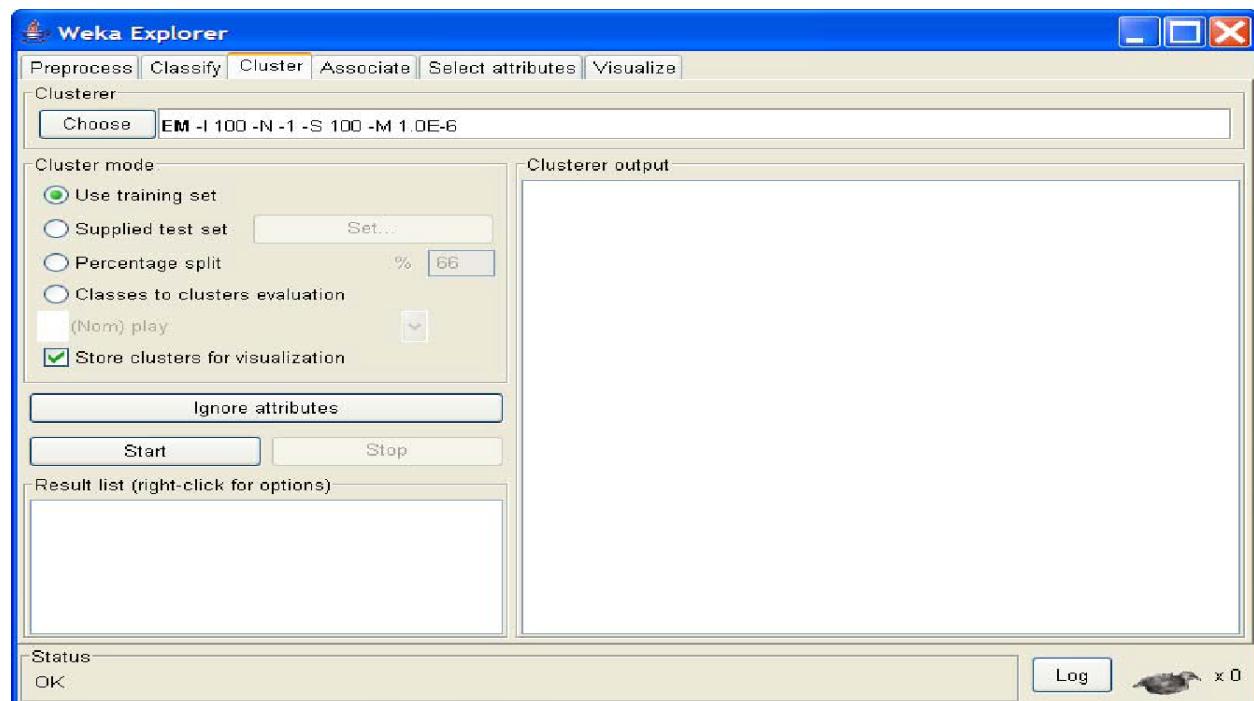
Aim : Clustering Analysis.

Solution :

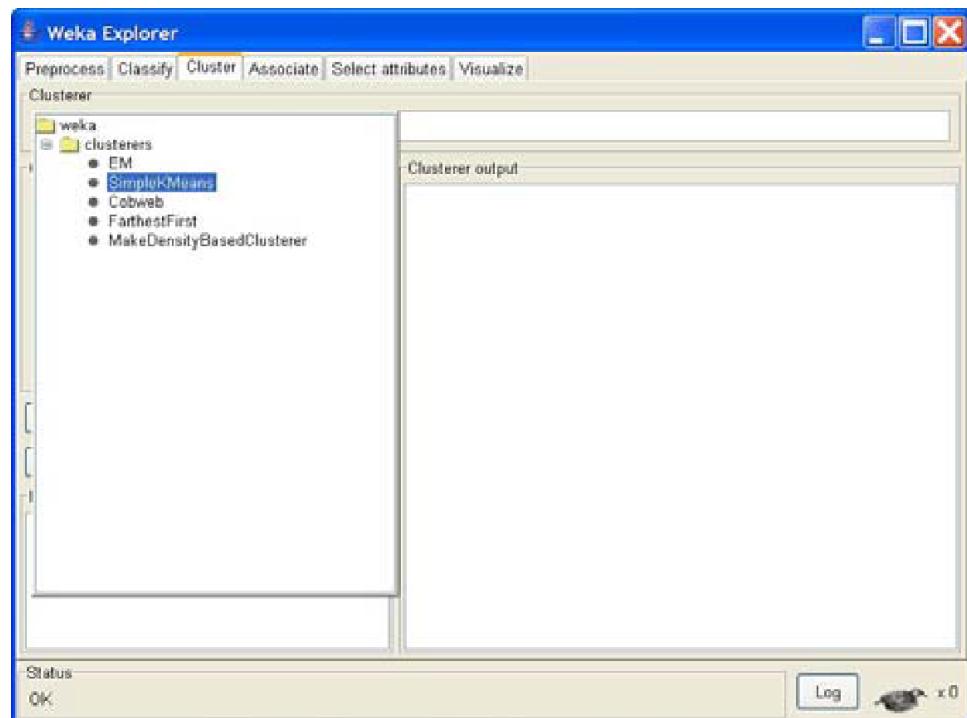
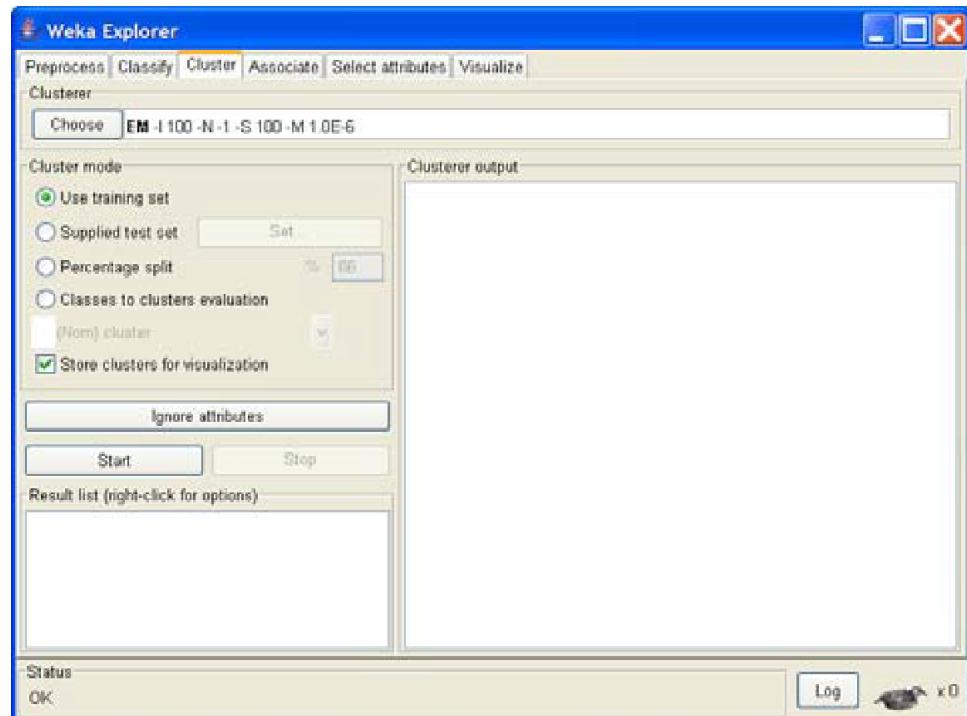
we will use customer data [6] that is contained in “customers.arff” file and analyze it with k-means clustering scheme.

	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					

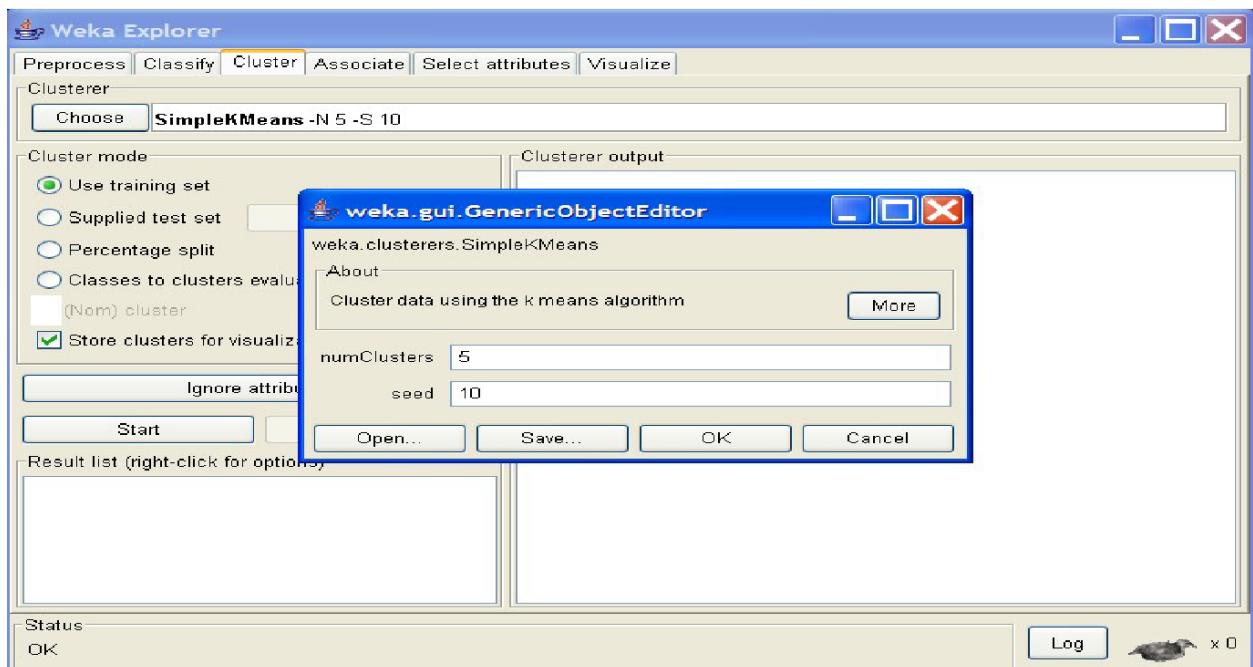
In ‘Preprocess’ window click on ‘Open file...’ button and select “customers.arff” file. Click ‘Cluster’ tab at the top of WEKA Explorer window.



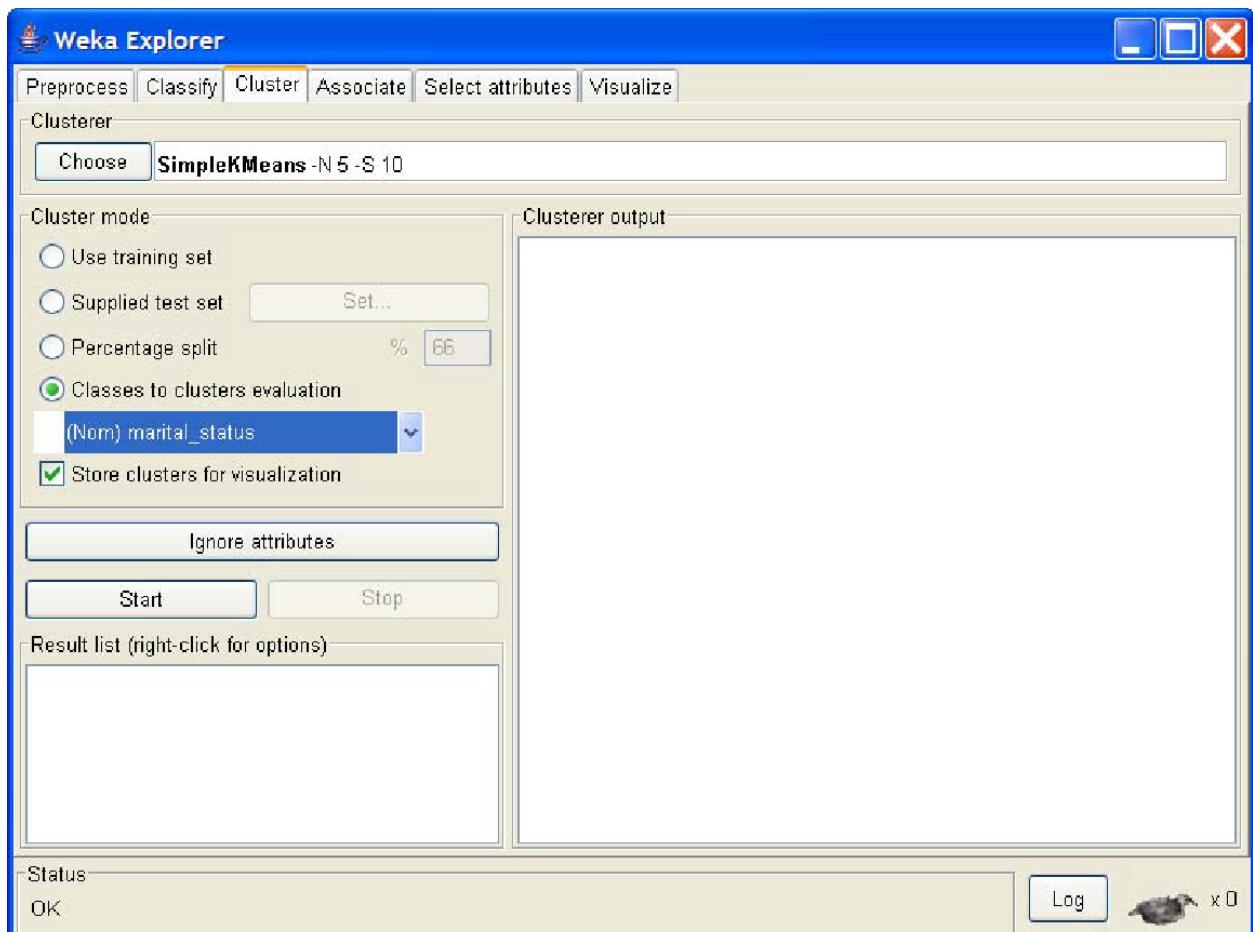
In the ‘Clusterer’ box click on ‘Choose’ button. In pull-down menu select WEKA □ Clusterers, and select the cluster scheme ‘SimpleKMeans’. Some implementations of K-means only allow numerical values for attributes.



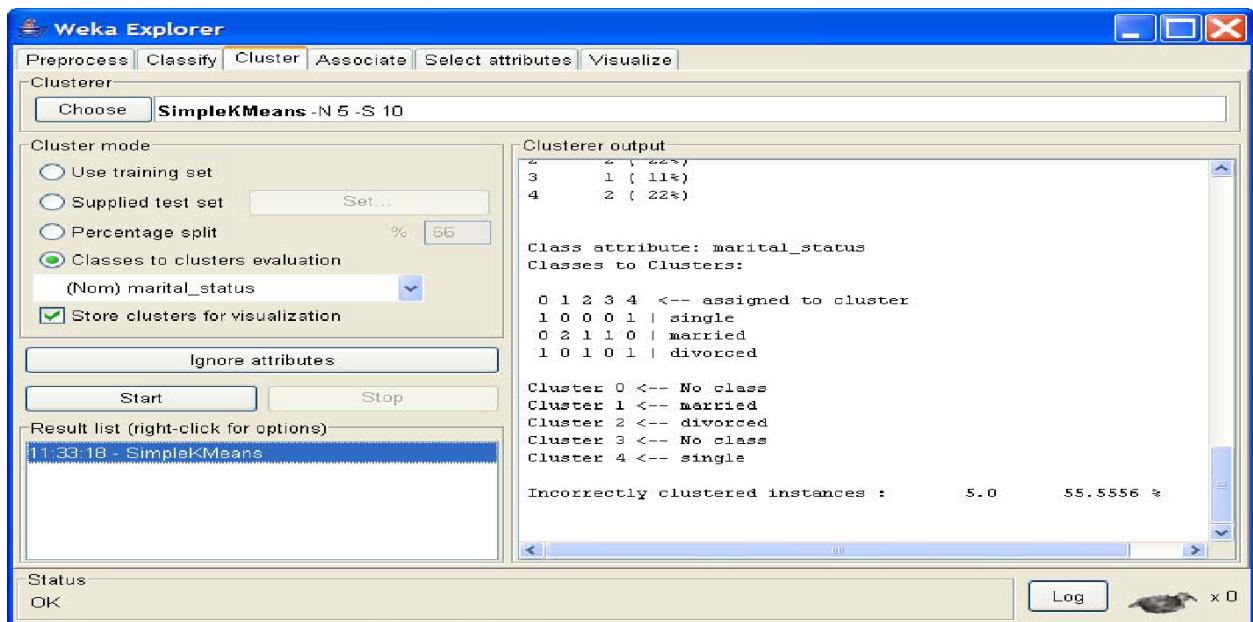
right-click on the algorithm “weak.gui.GenericObjectEditor” comes up to the screen. Set the value in “numClusters” box to 5(instead of default 2) because you have five clusters in your .arff file.



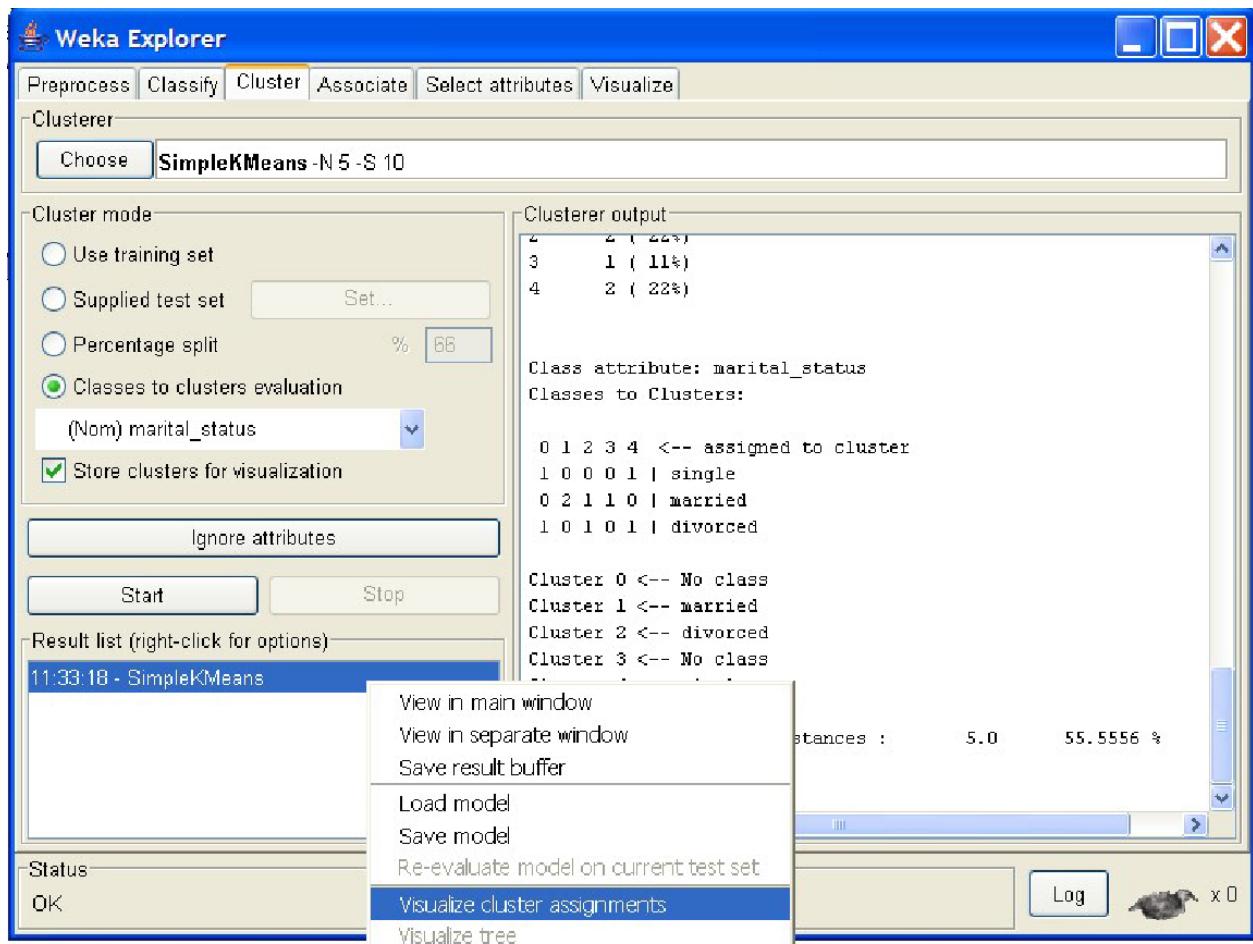
Click on ‘Classes to cluster evaluation’ radio-button in ‘Cluster mode’ box and select ‘marital_status’ in the pull-down box below.



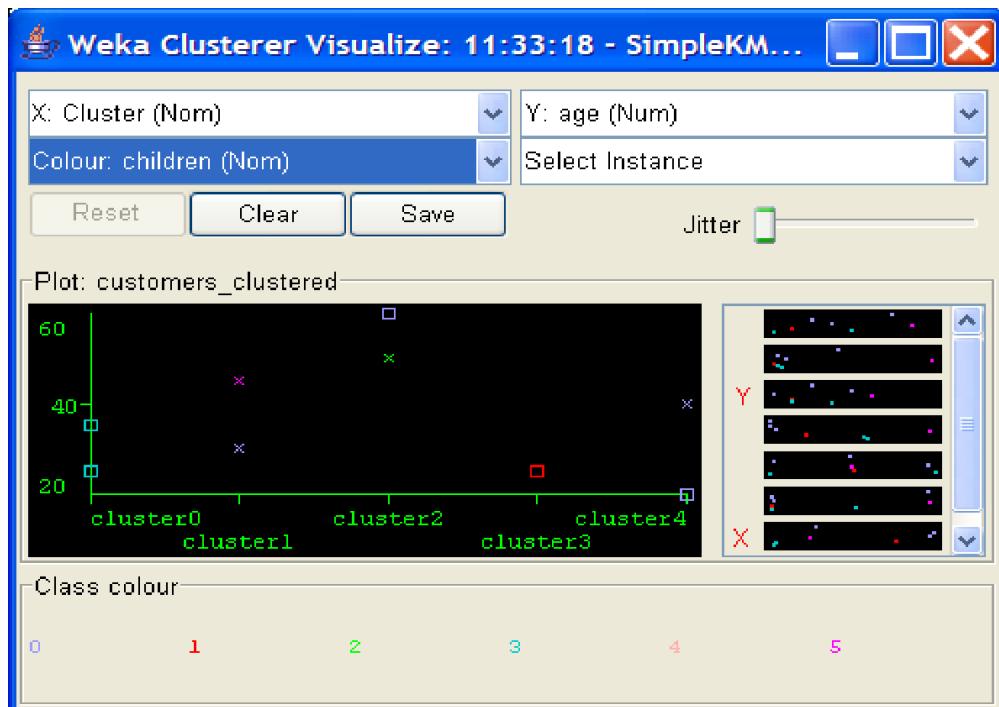
Click on the ‘Start’ button to execute the algorithm.



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.

The figure shows a Microsoft Word document titled "customers_kmeans - Microsoft Word". The content of the document 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

```

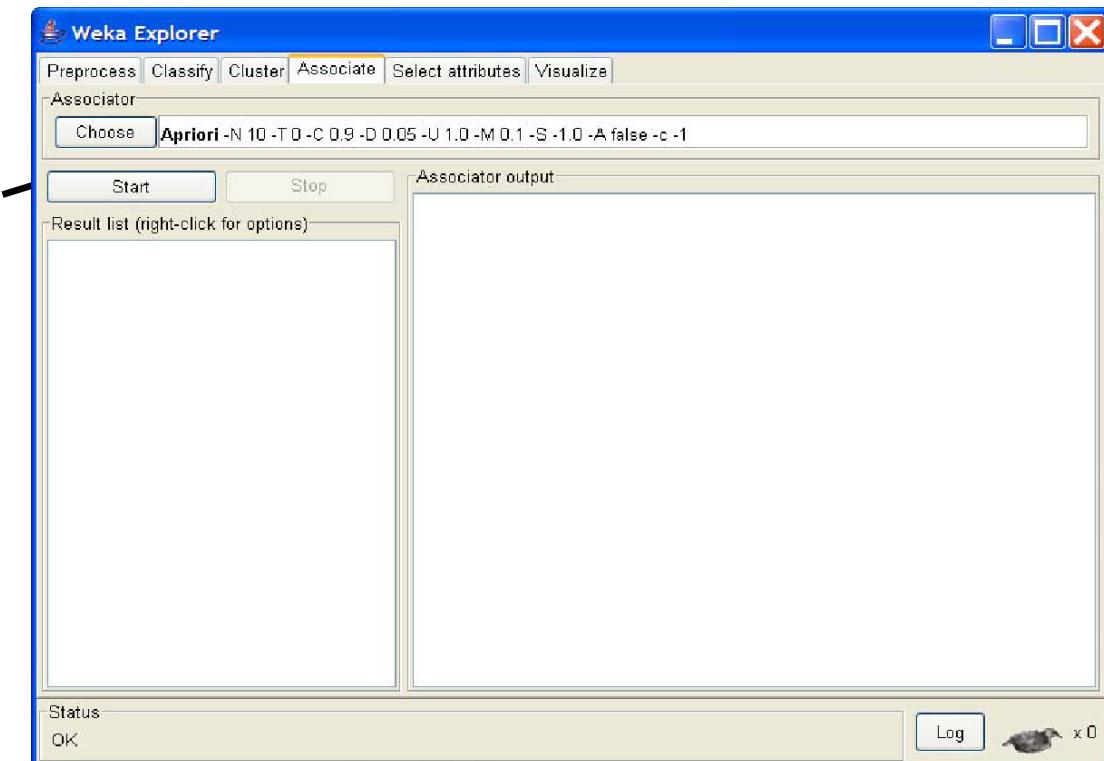
Practical No-14

Aim : Association Rule Mining.

Solution :

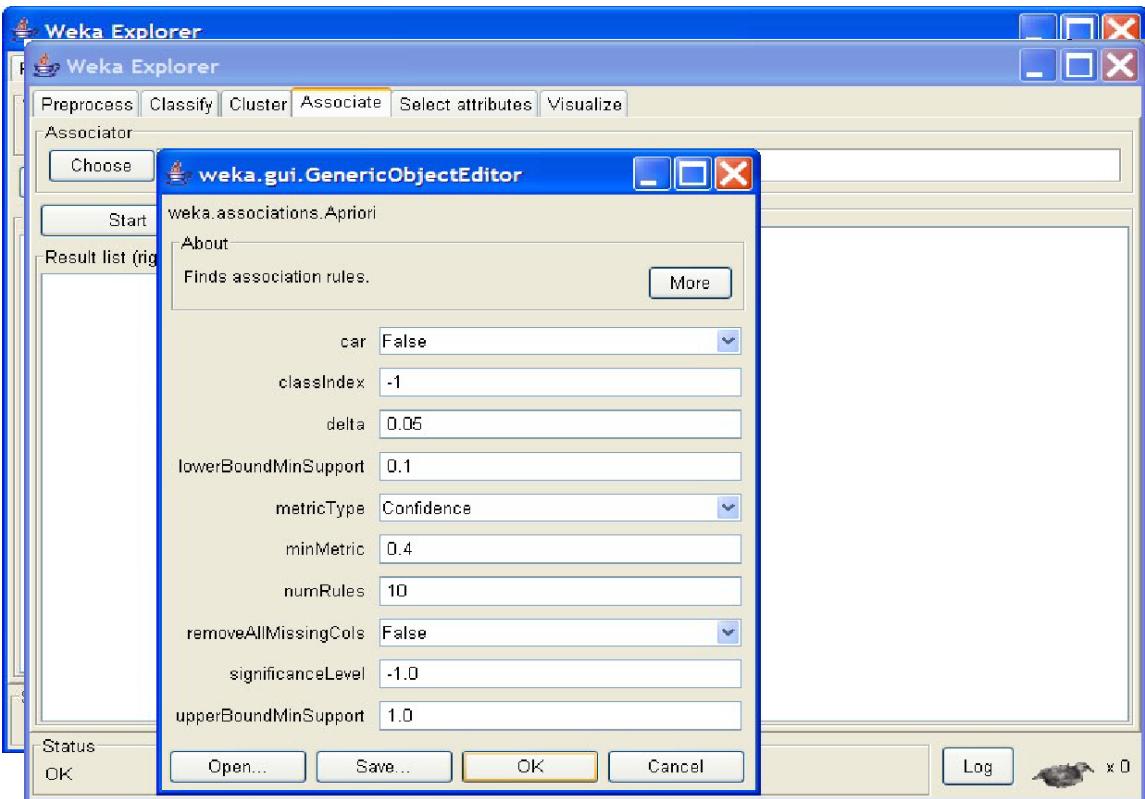
Choosing Association Scheme

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



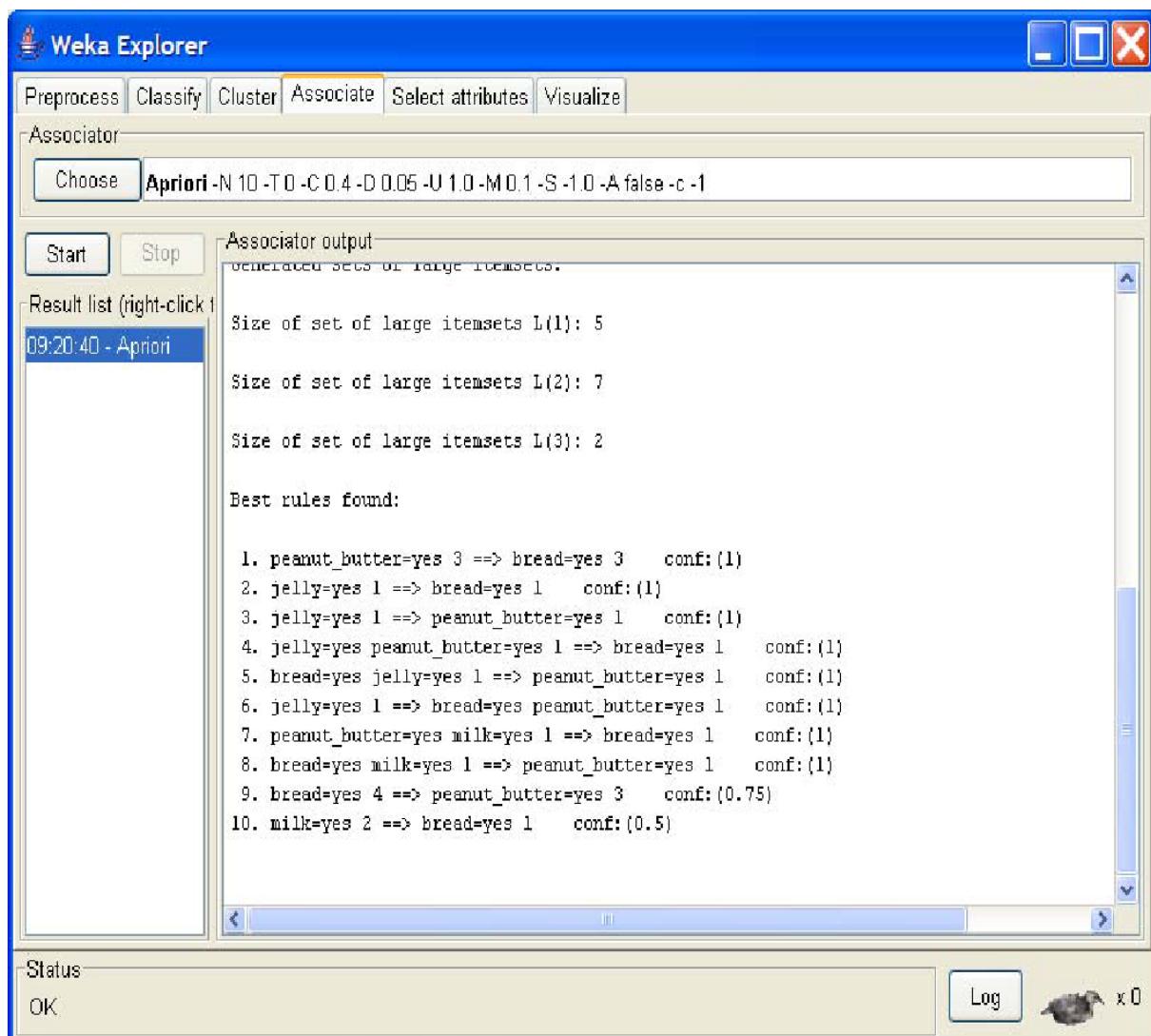
Setting Test Options

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



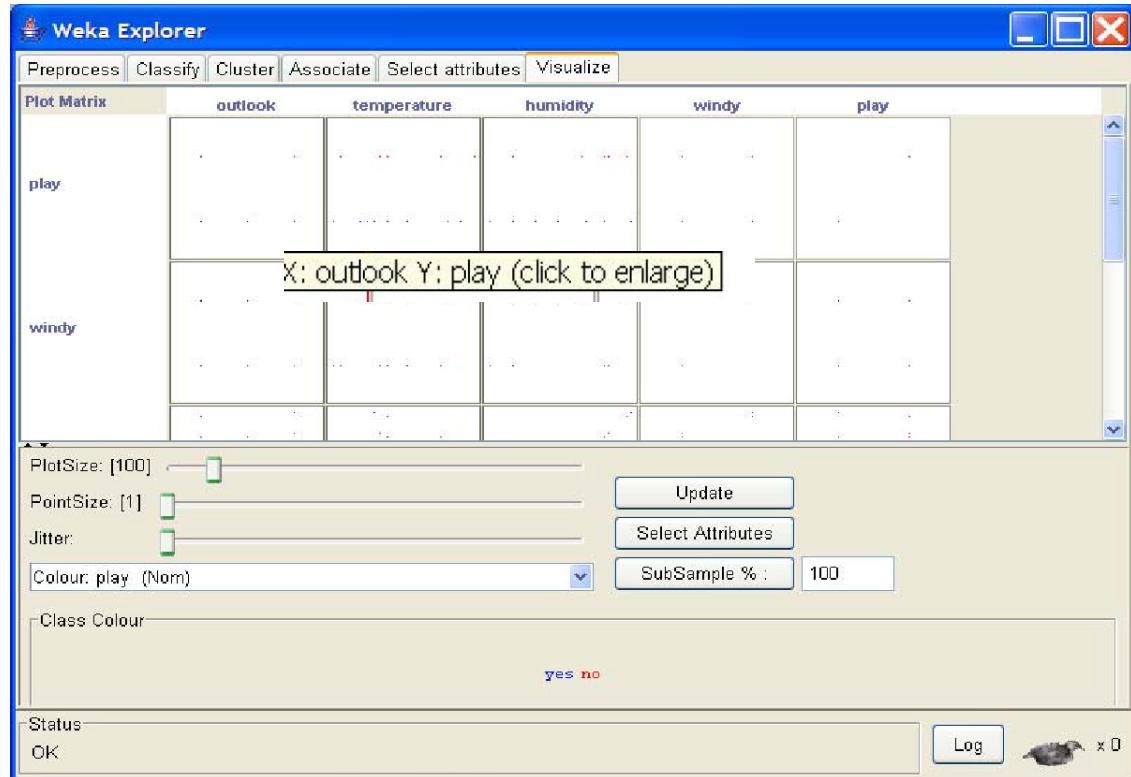
Practical No-15

Aim : Data Visualization

Solution :

To open Visualization screen, click ‘Visualize’ tab.

Select a square that corresponds to the attributes you would like to visualize.

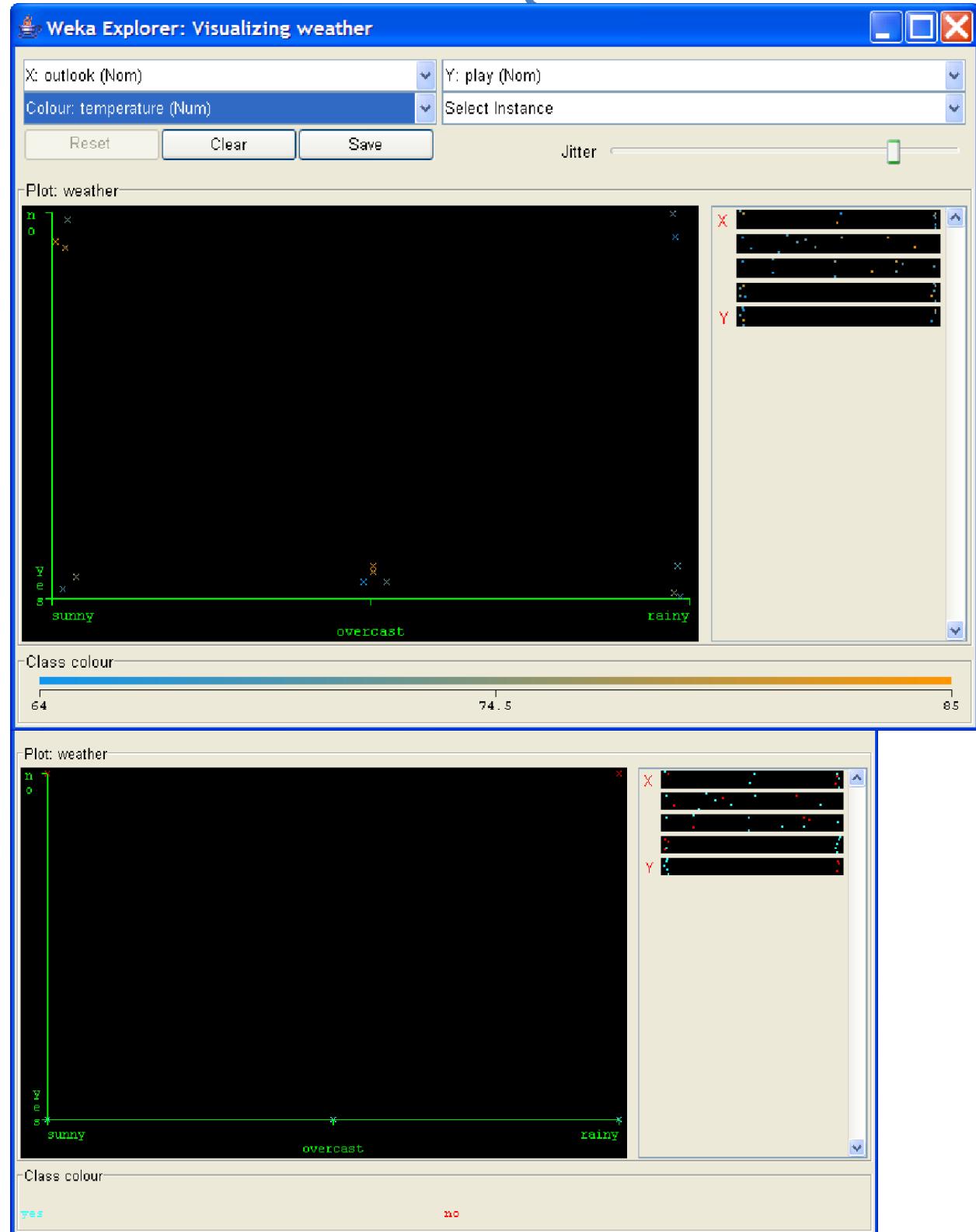


A ‘Visualizing weather’ window appears on the screen.

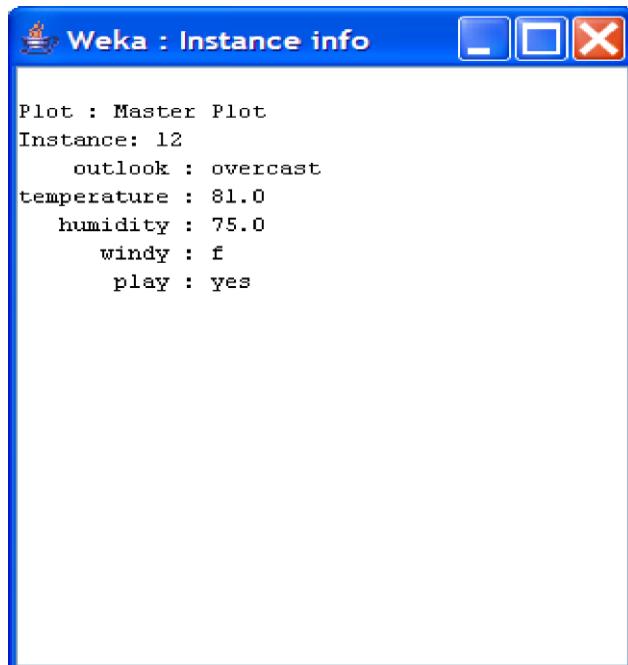
Changing the View

Keep sliding ‘Jitter’, a random

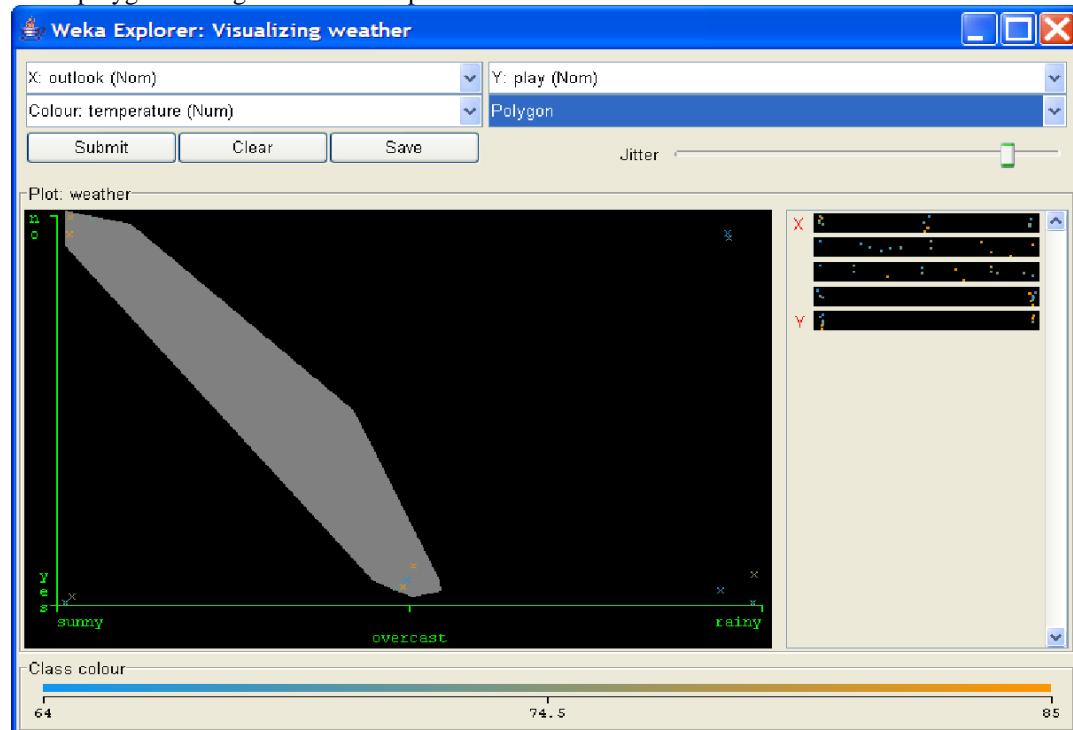
displacement given to all points in the plot, to the right, until you can spot concentration points



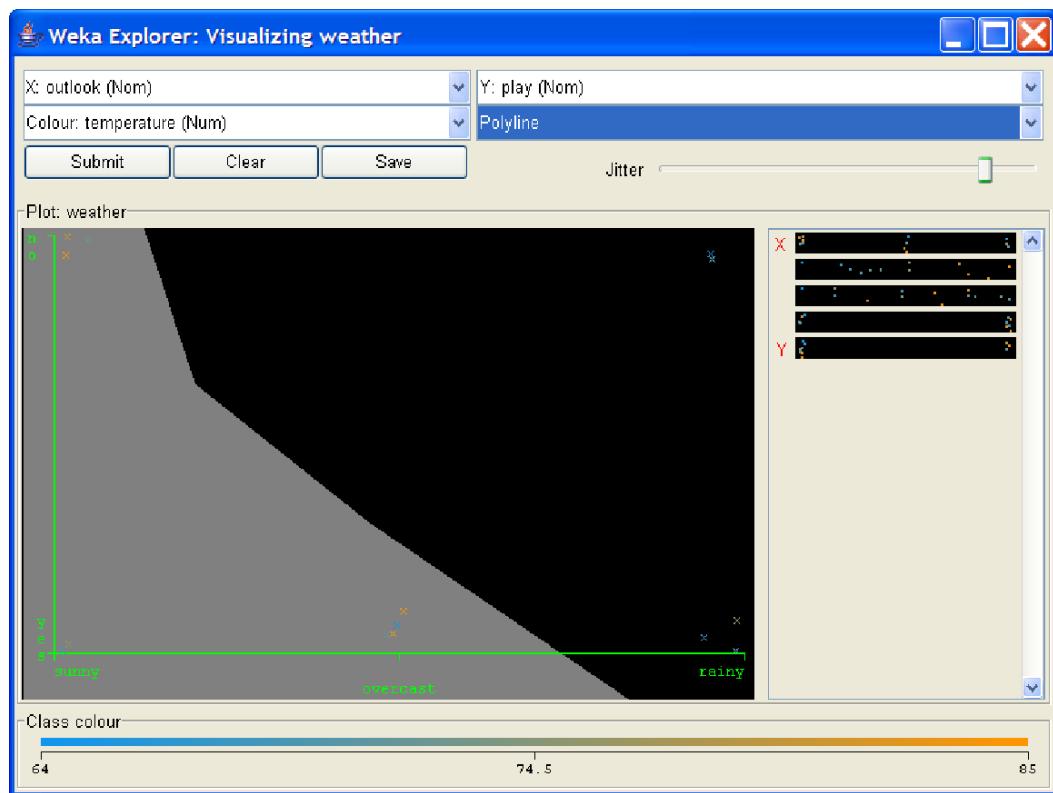
Selecting Instances: Click on an individual data point



3. **Polygon.** You can select several points by building a free-form polygon. Left-click on the graph to add vertices to the polygon and right-click to complete it.



4. **Polyline.** To distinguish the points on one side from the ones on another, you can build a polyline. Left-click on the graph to add vertices to the polyline and right-click to finish.



Rectangle. You can create a rectangle by dragging it around the points

