

NAME :- MANALI YASHWANT BHASKAR.

Class : MSC I

Sem : I

Subject : Data Warehousing & Data Mining(BI)

Paper : IV

Academic Year : 2022-23

Roll No : 528

Practical No 1

Aim : Creation of Dimensions and Fact tables.

Solution :

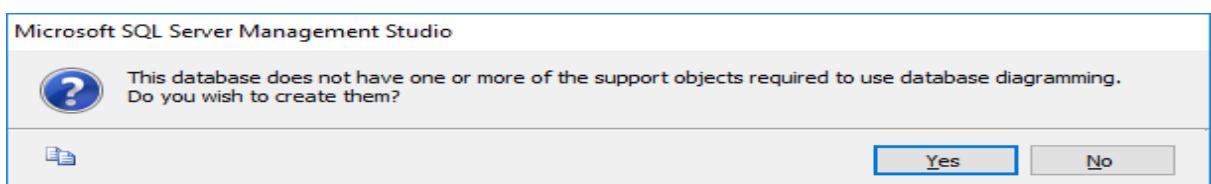
Open Application -> Microsoft SQL Server 2008 R2 -> SQL Server Management Studio

- 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.

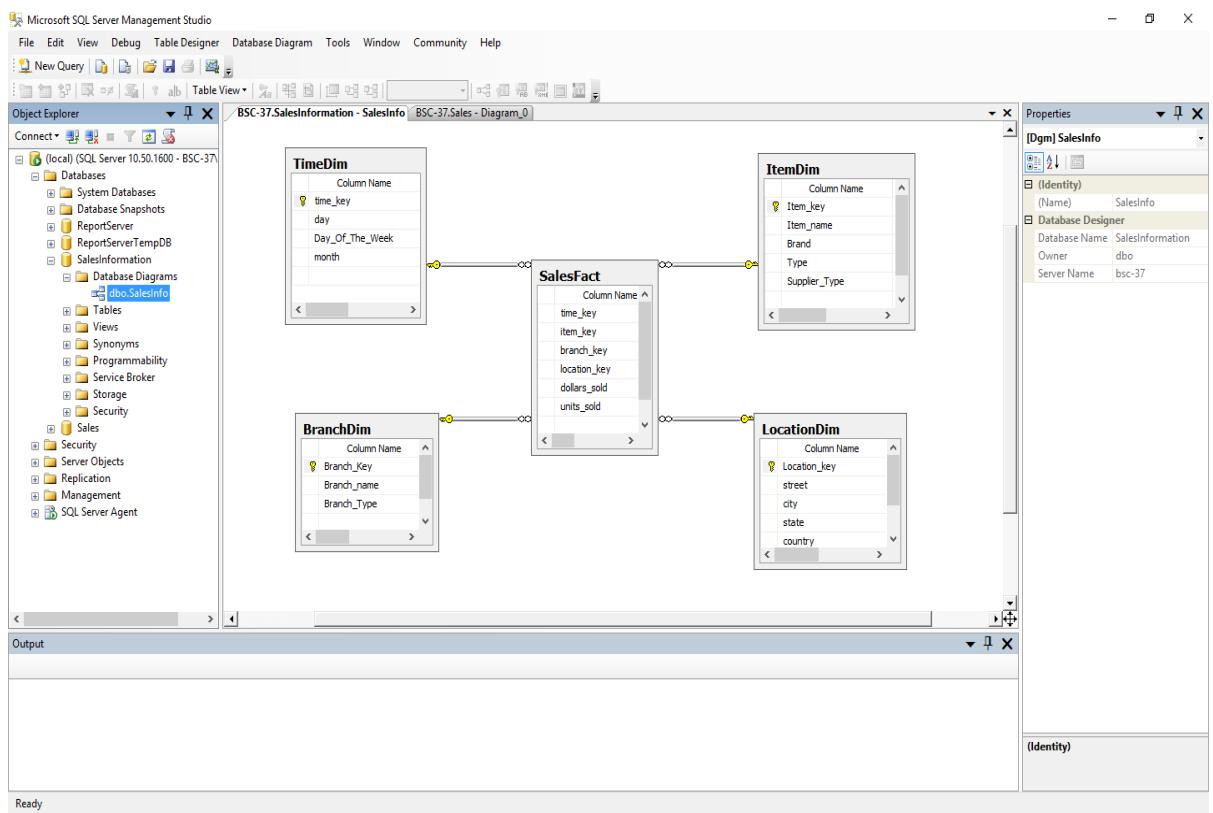
- 1. Click on Database Diagrams to expand it**



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.)

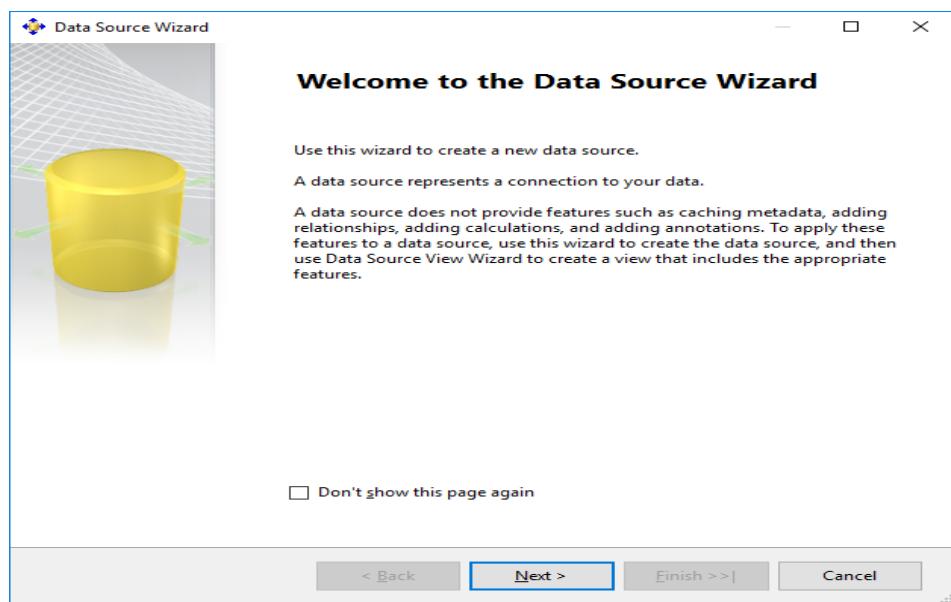
Practical No 2

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

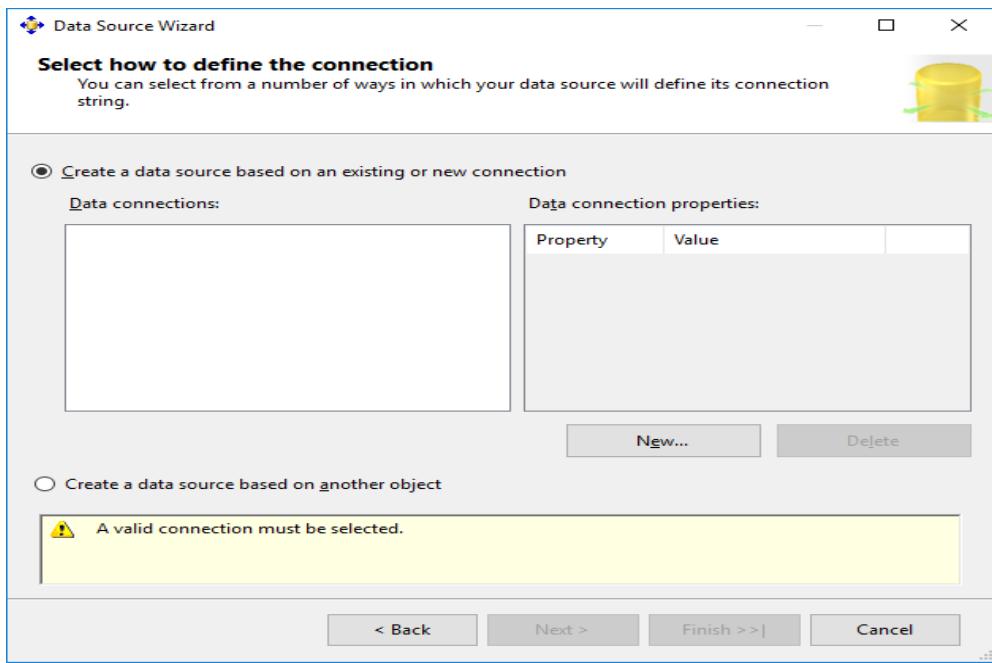
Solution :

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**

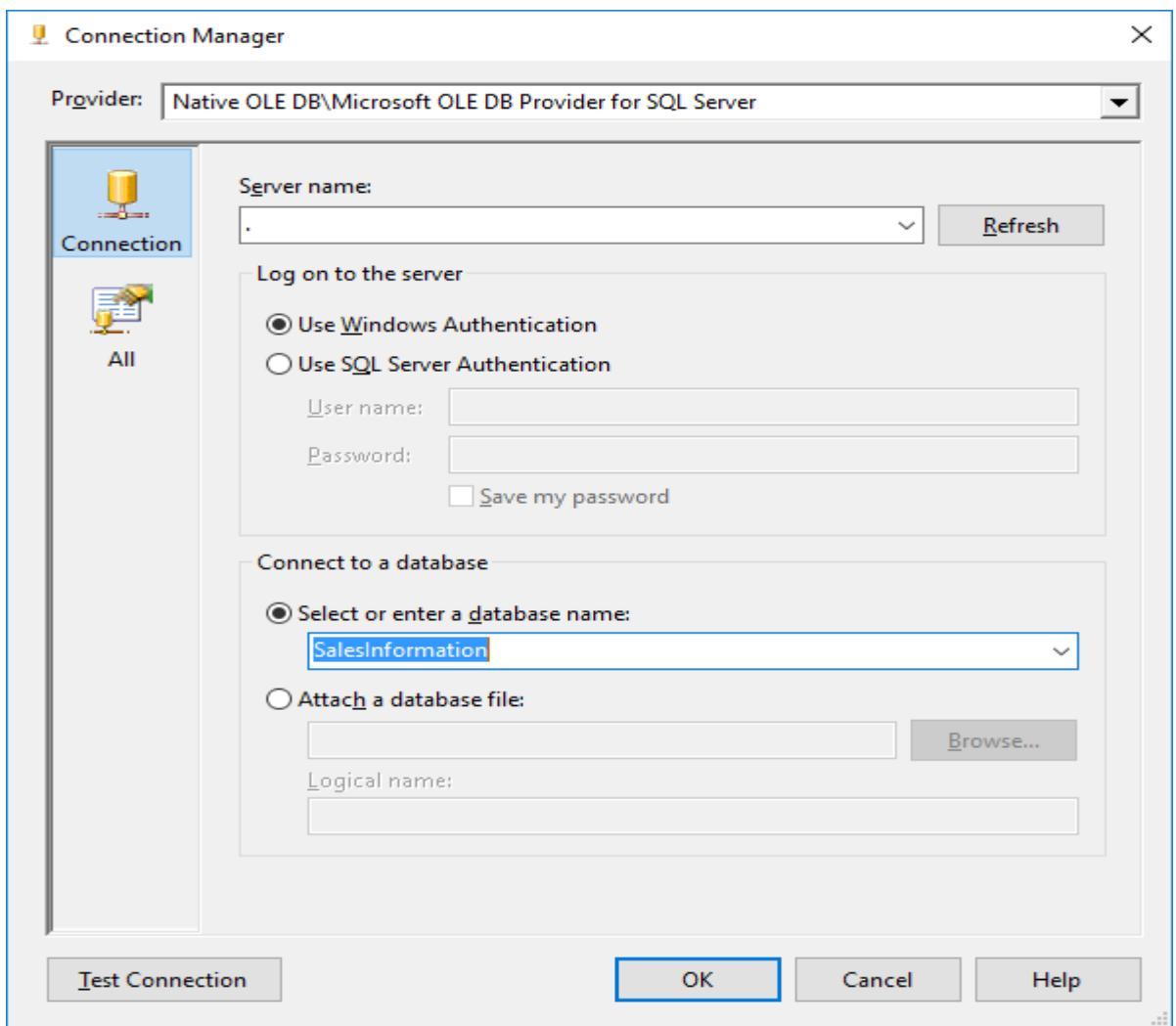


Click on Next.

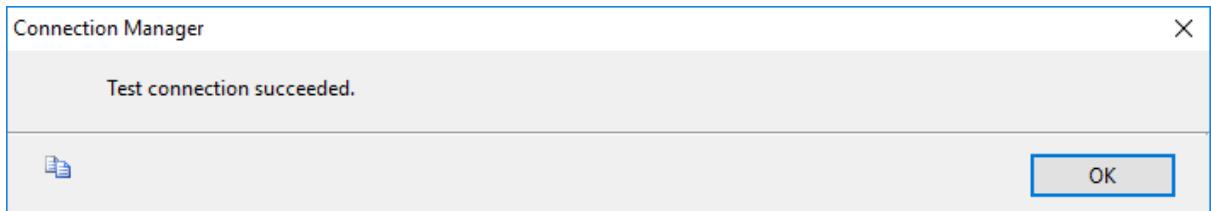


Click on New.

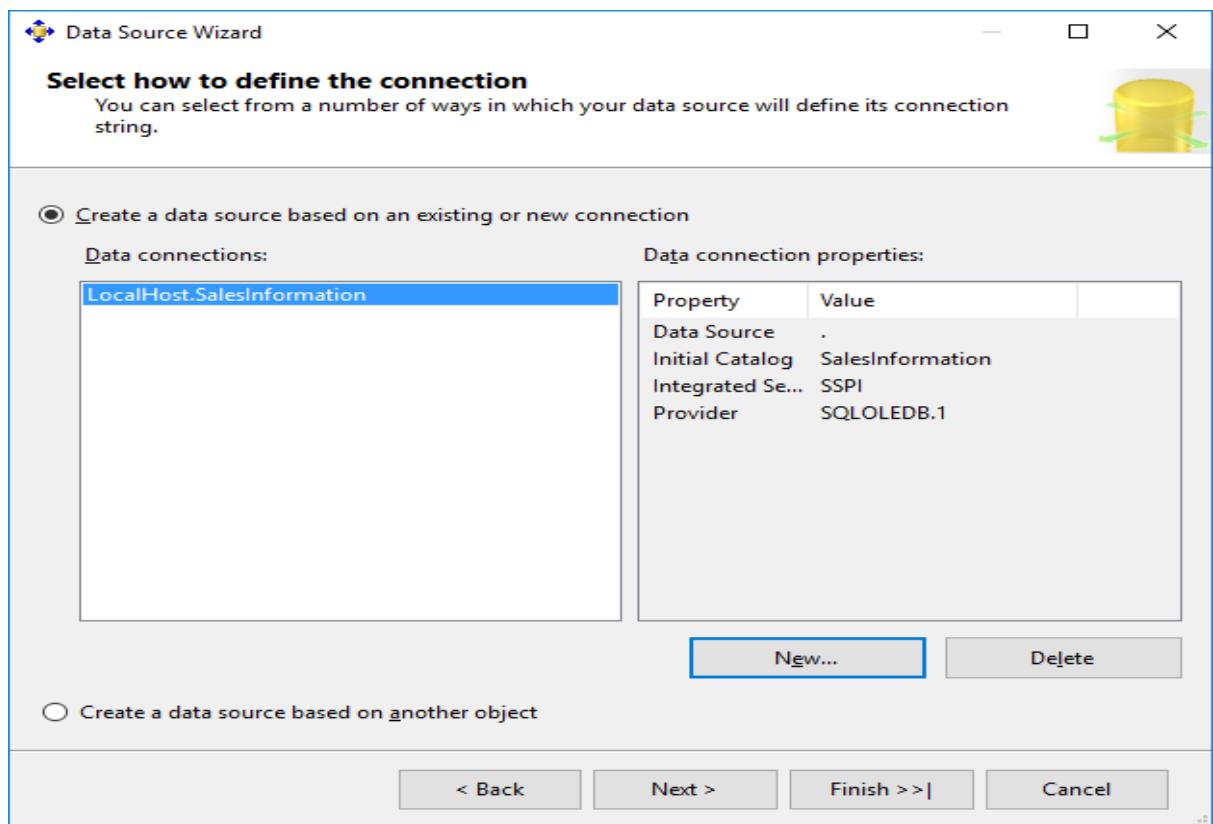
3. Choose Provider as “Microsoft OLEDB Provider for SQL Server”, Server Name as “.”, Select database name as “SalesInformation”.(Created in SQL Server Management studio).



4. Click on Test Connection.

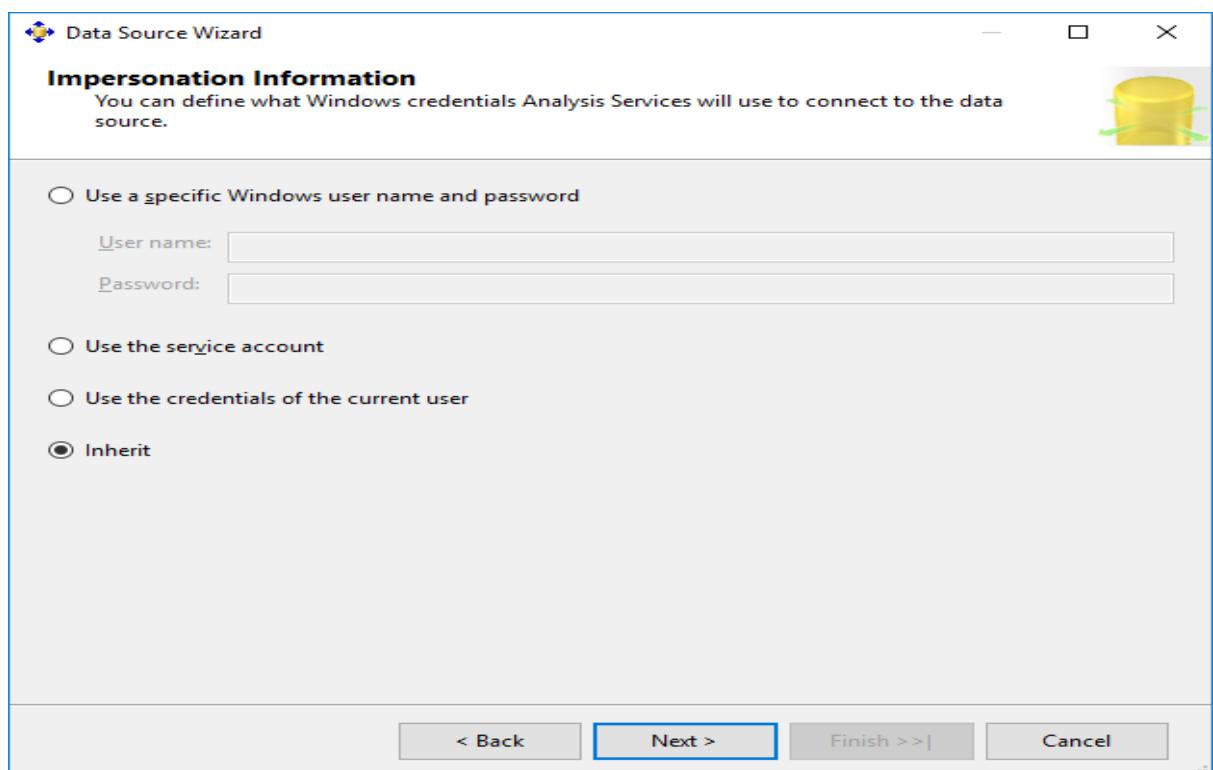


Click on OK.



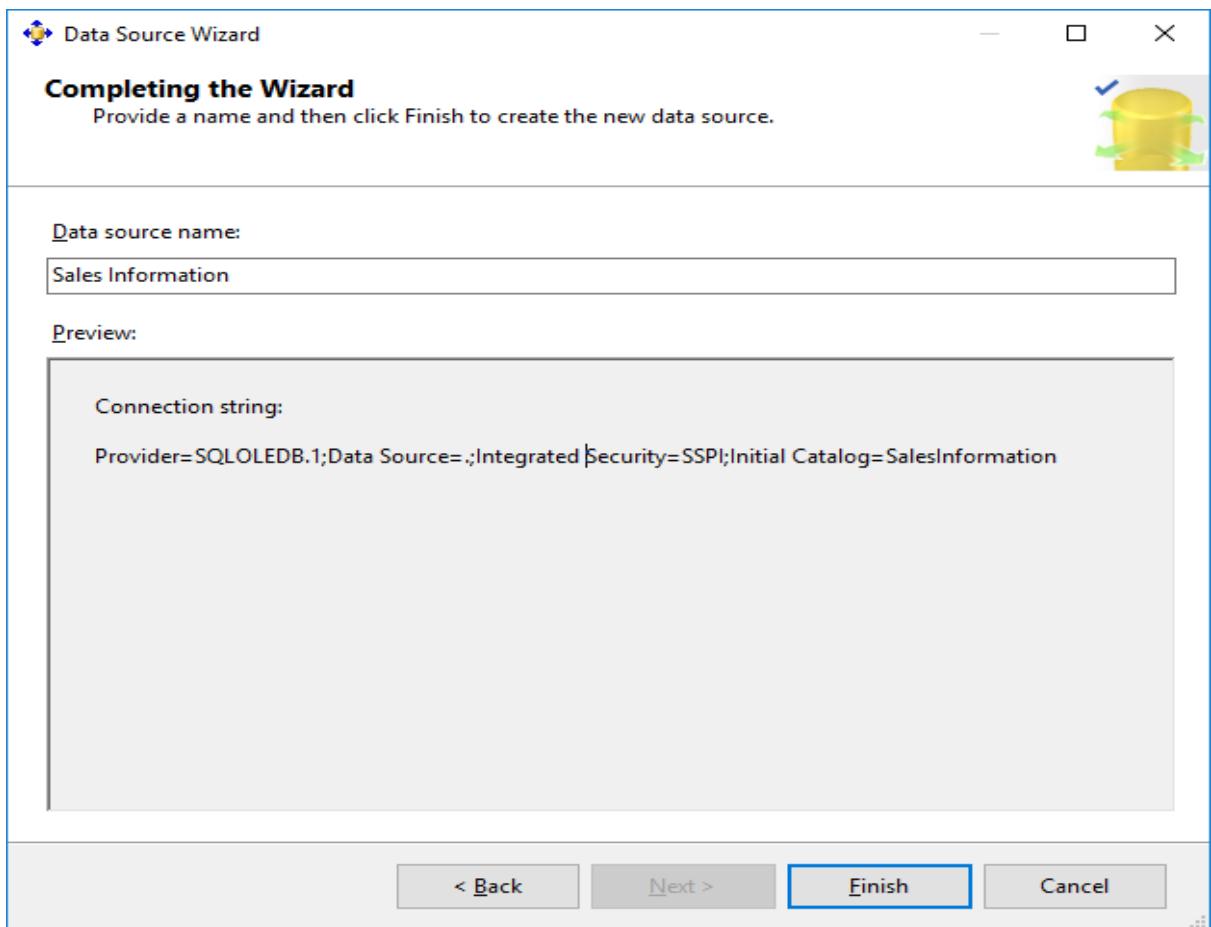
Click on Next

5. Choose “Inherit” option.



Click on Next.

6. Click on Finish.



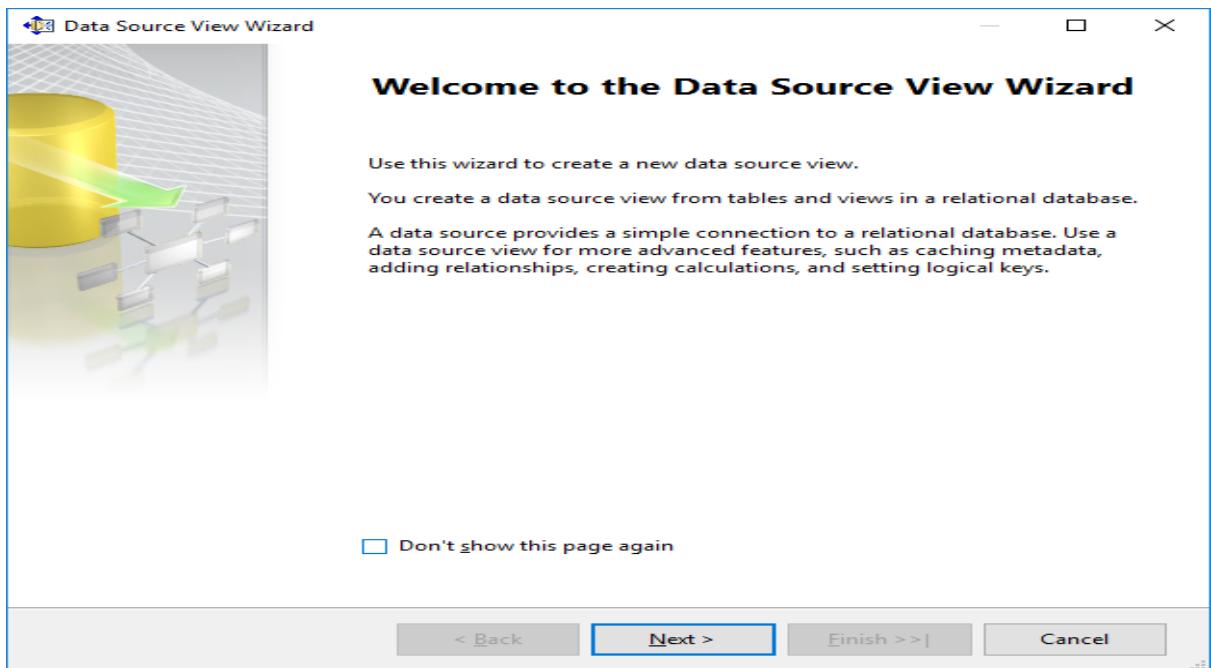
Name Data Source as “Sales Information”.

Practical No 3

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

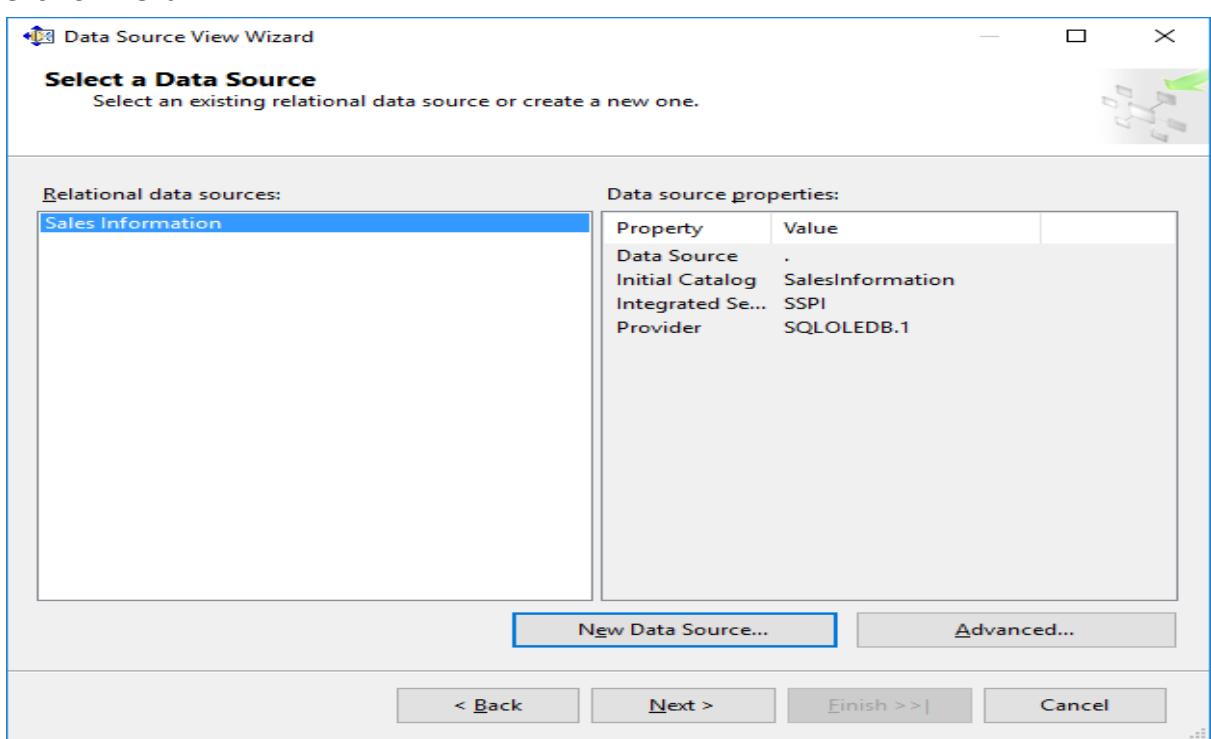
Solution :

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

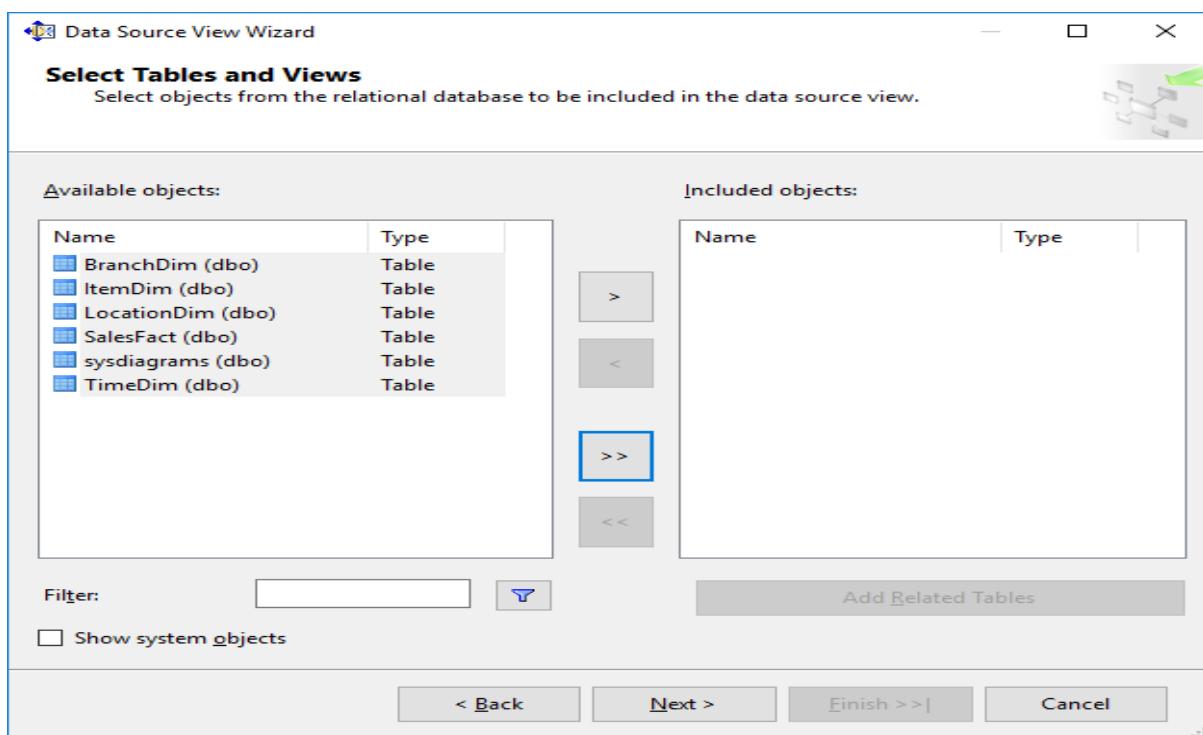


Click on Next.

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: 

Show system objects

Add Related Tables

< 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**

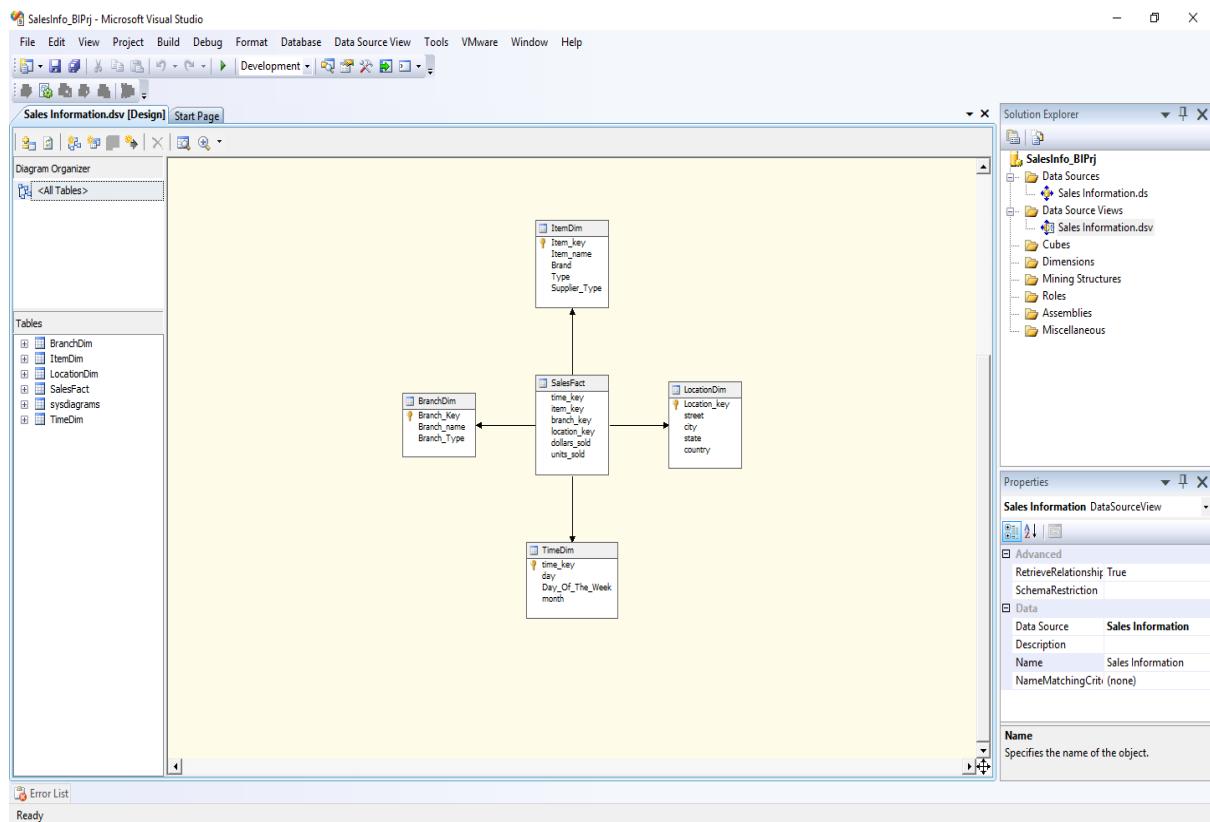
Preview:

-  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 :

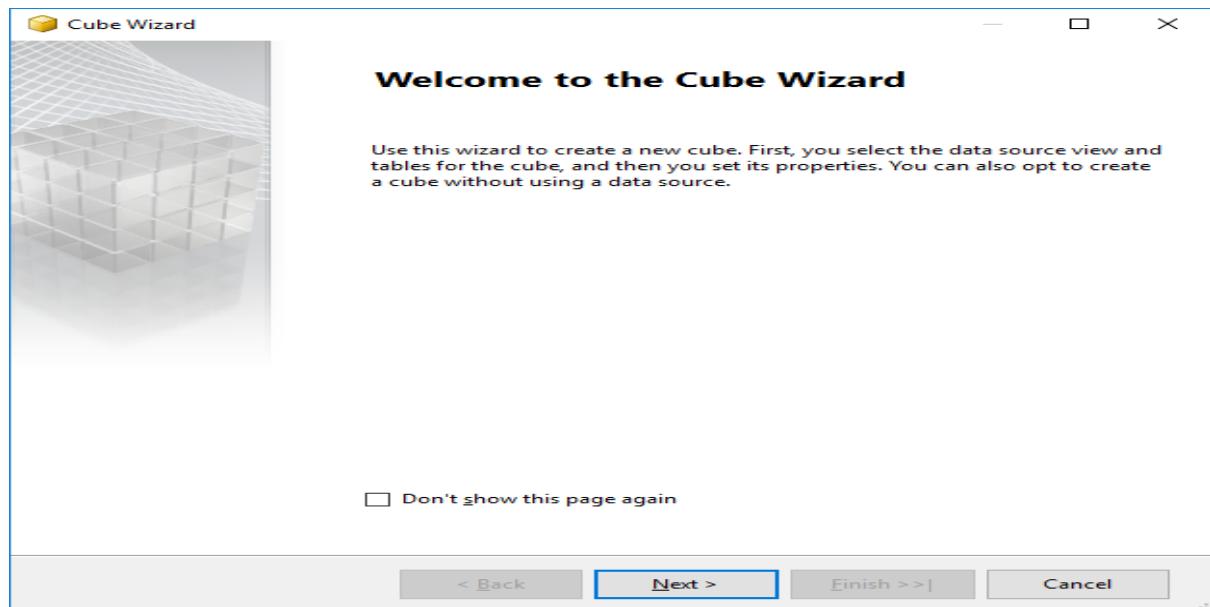


Practical No 4

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

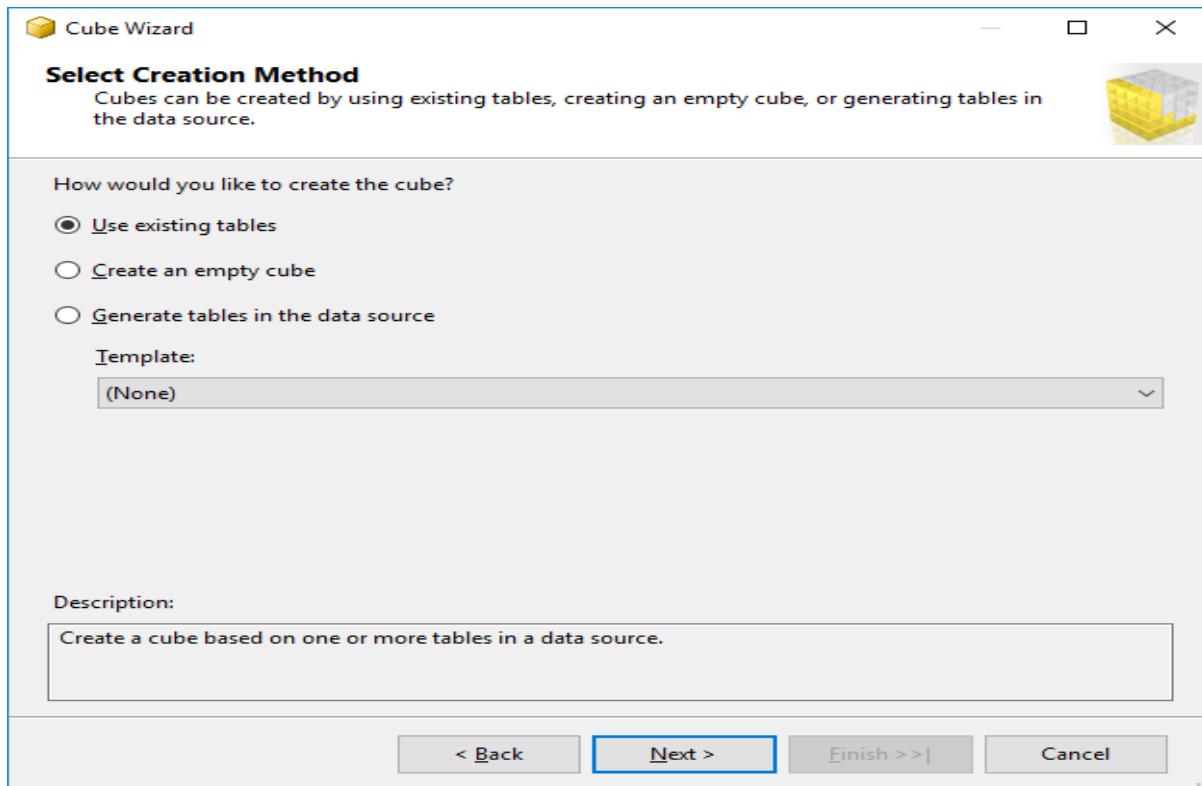
Solution :

1. Right click on Cubes -> New Cube.

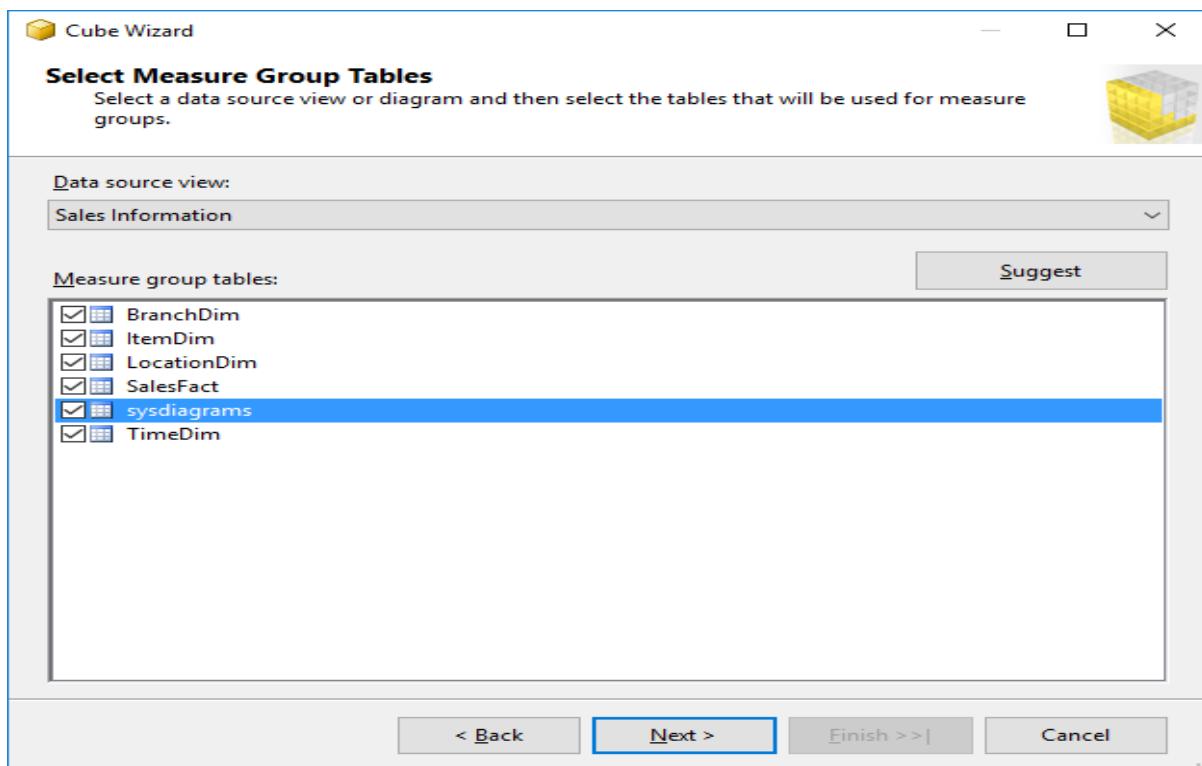


Click on Next.

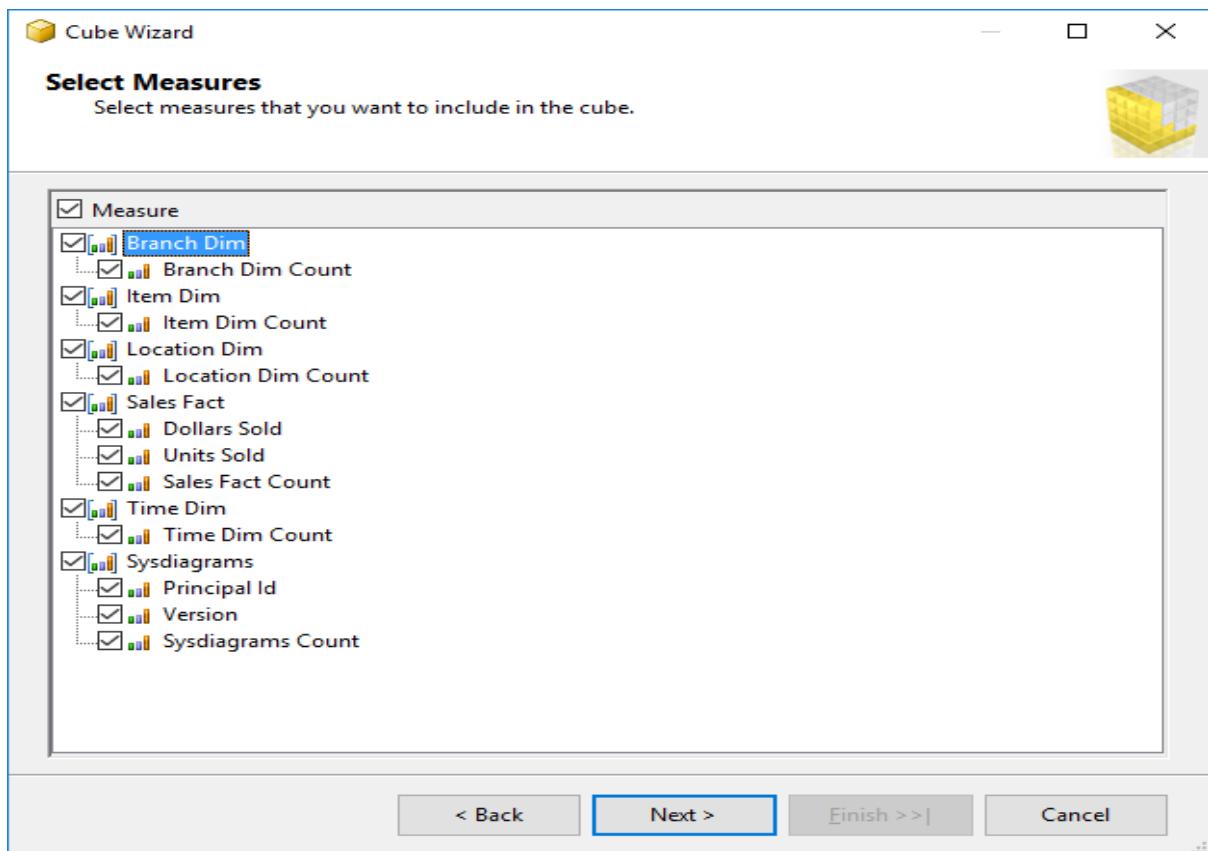
2. Select First option “Use existing tables”. Click on Next.



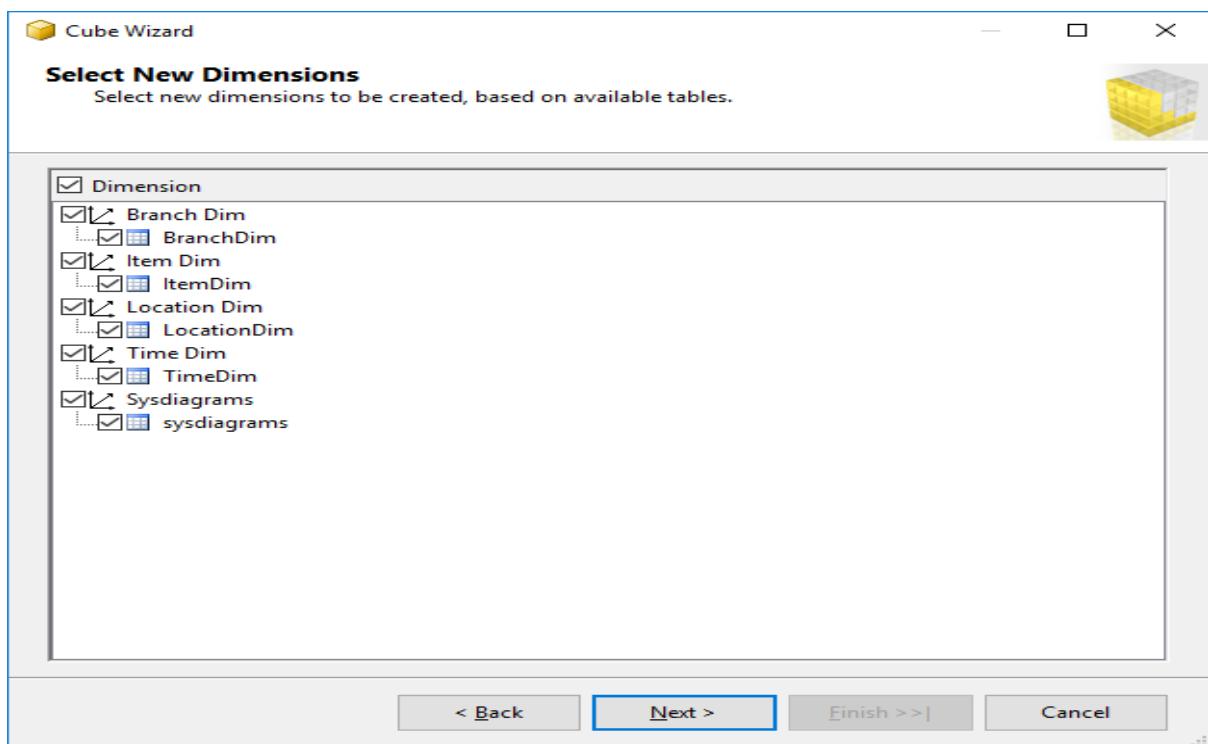
3. Select Data Source View as “Sales Information” and Select all the tables.



Click on Next.

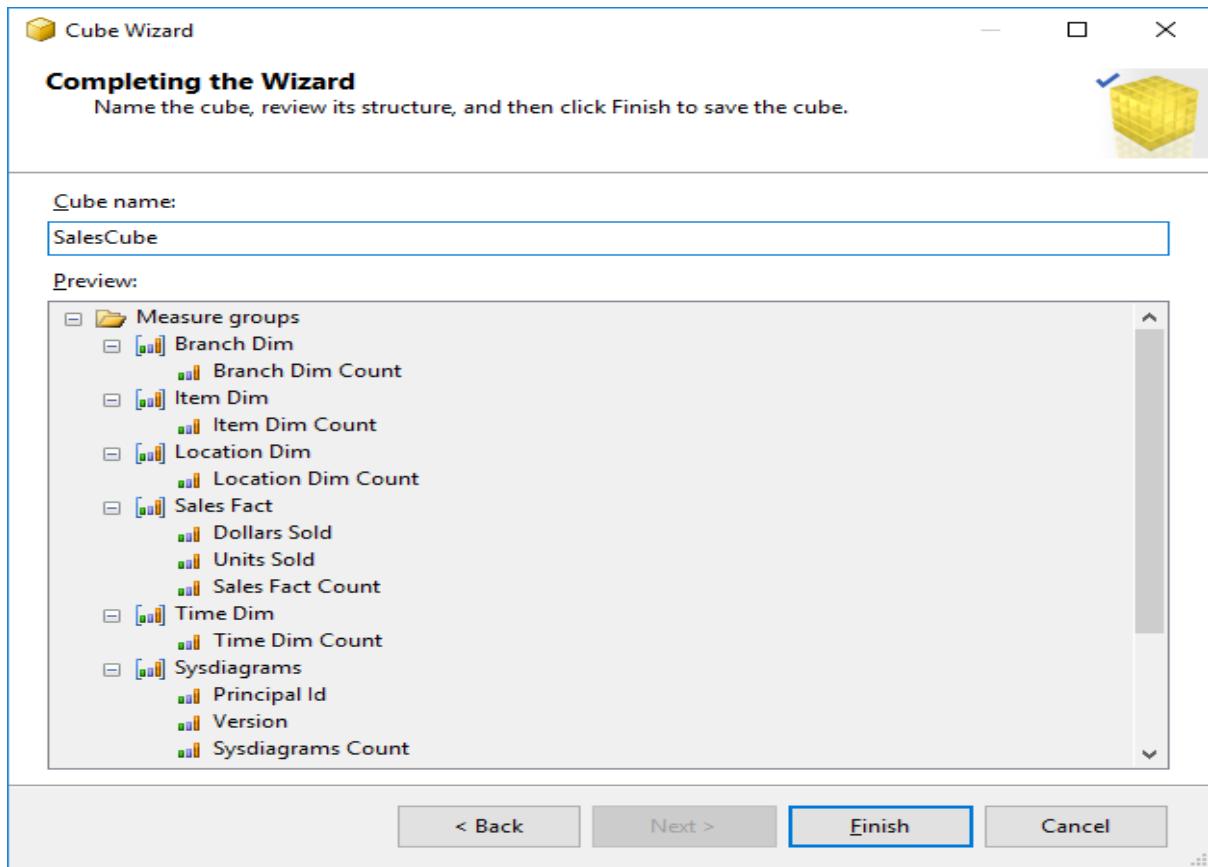


Click on Next.



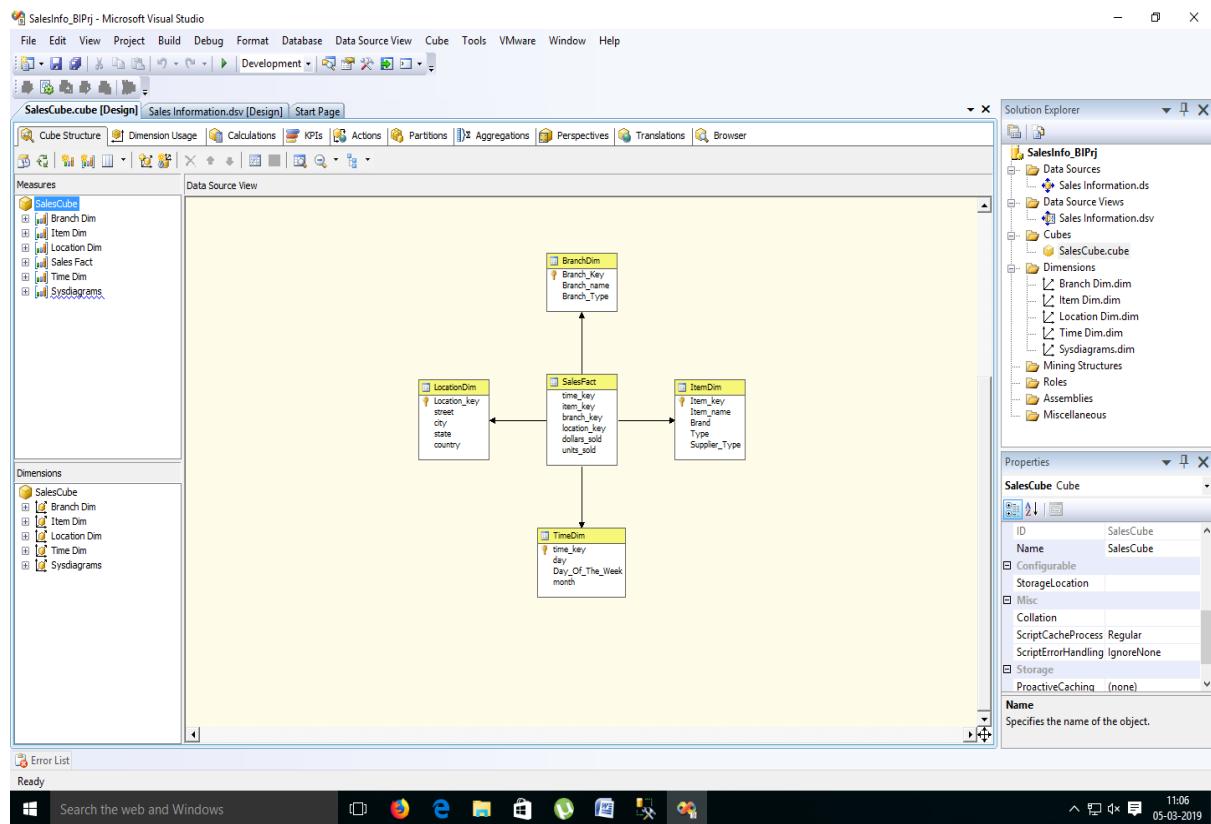
Click on Next.

4. Name Cube as “SalesCube”.

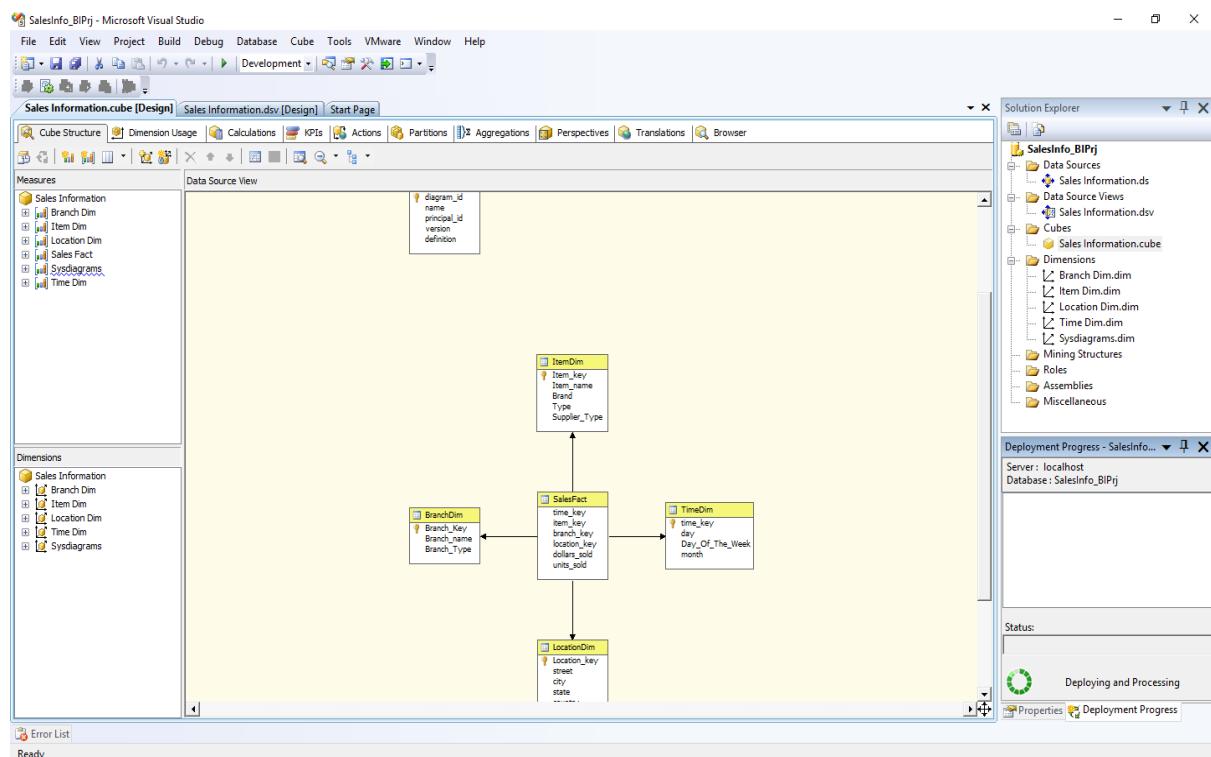


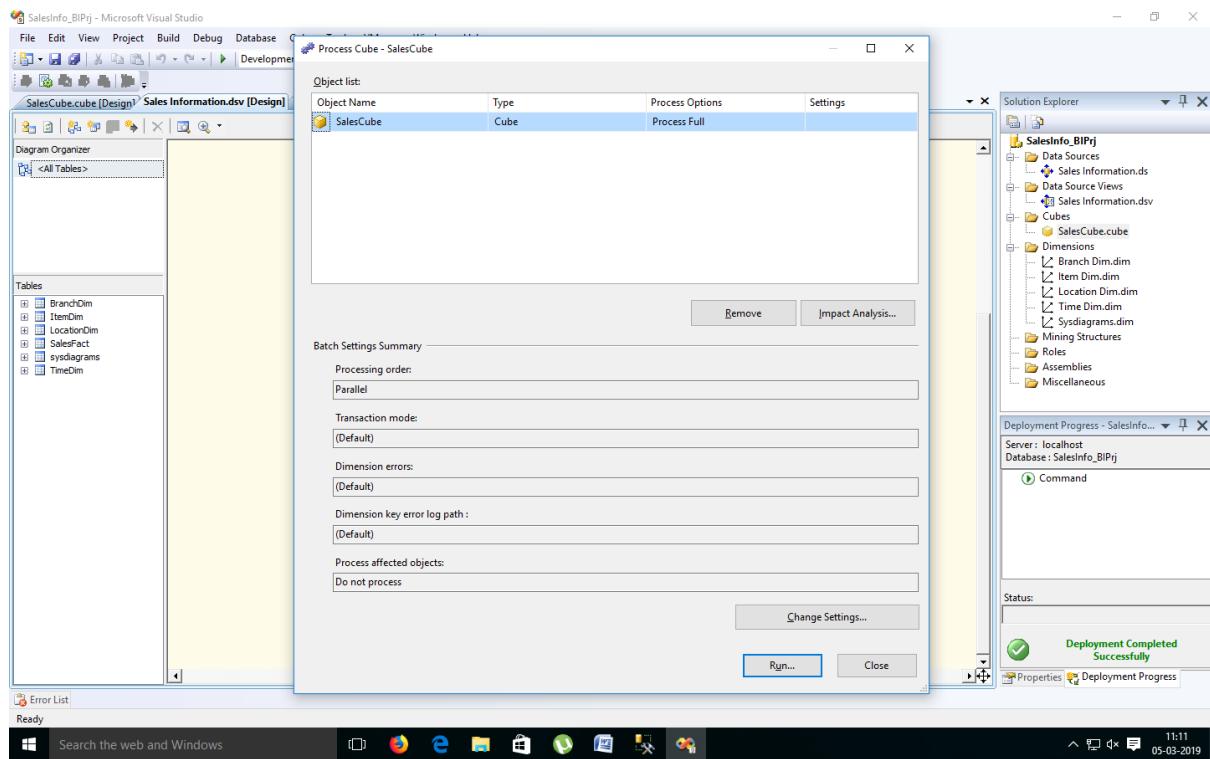
Click on Finish.

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

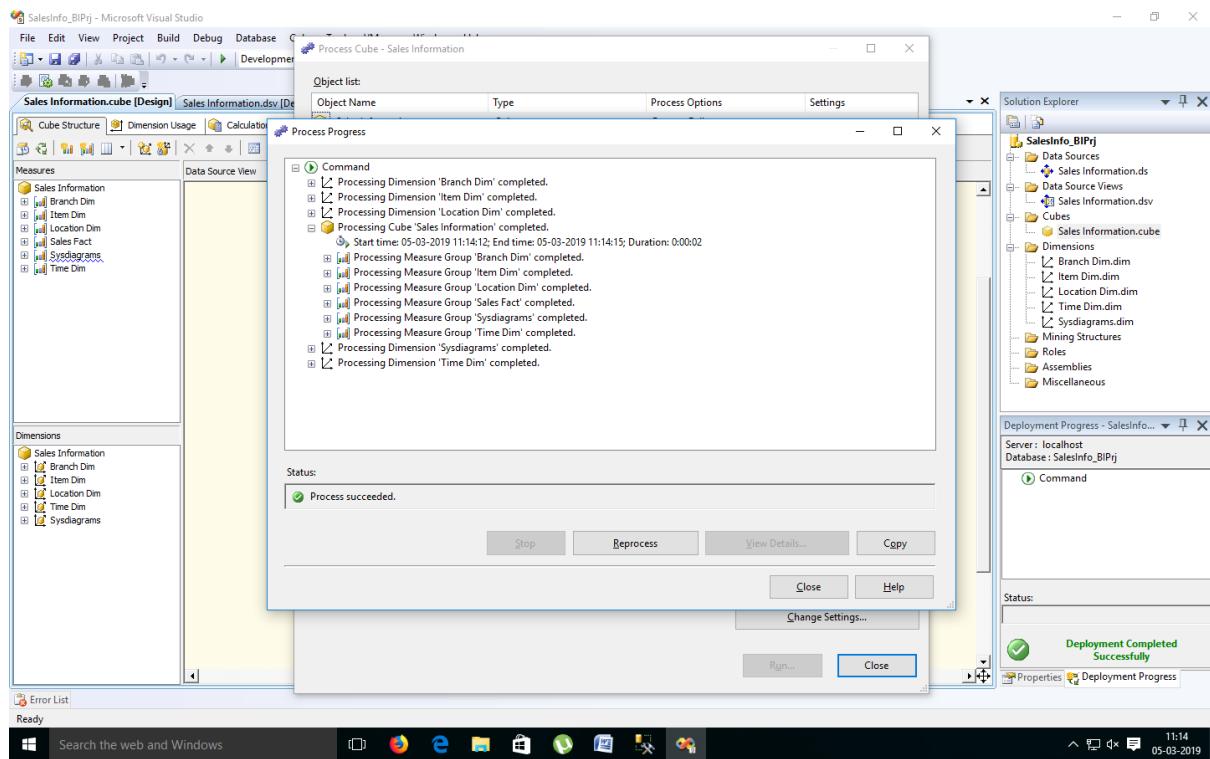


6. Finally, Process cube by Right click on SalesCube -> Process .





7. Click on Run.

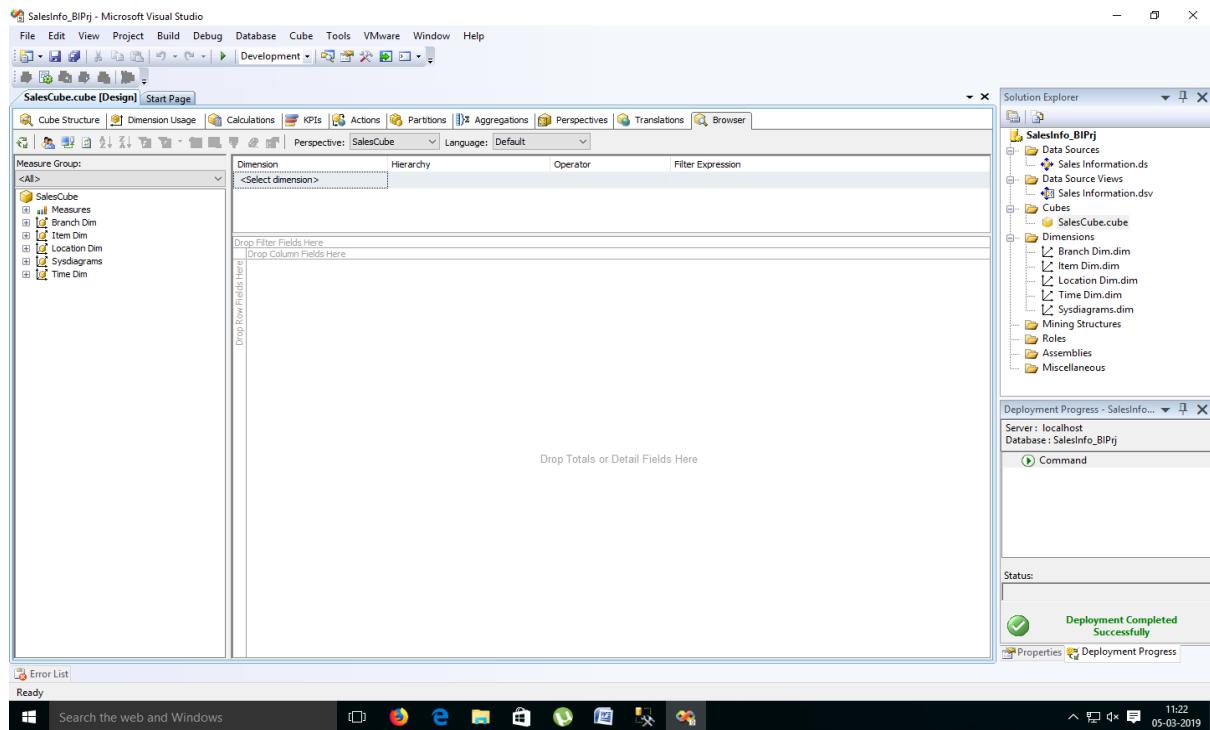


Practical No 5

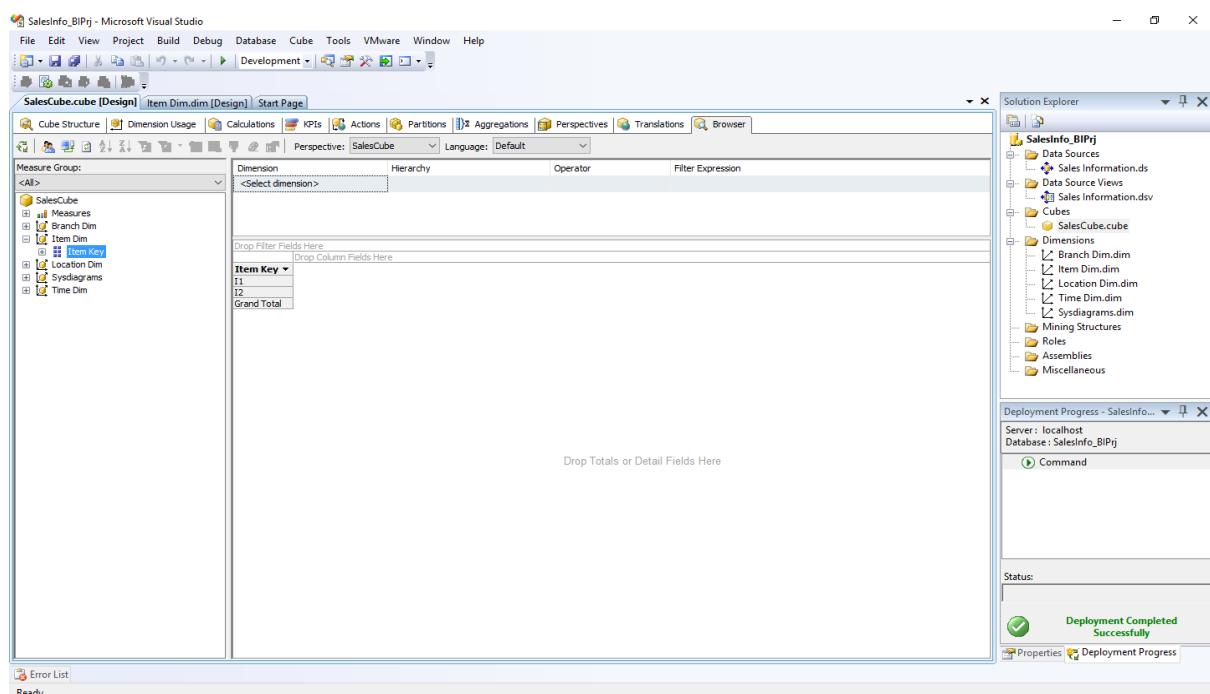
Aim : View cube data in multidimensional Format.

Solution :

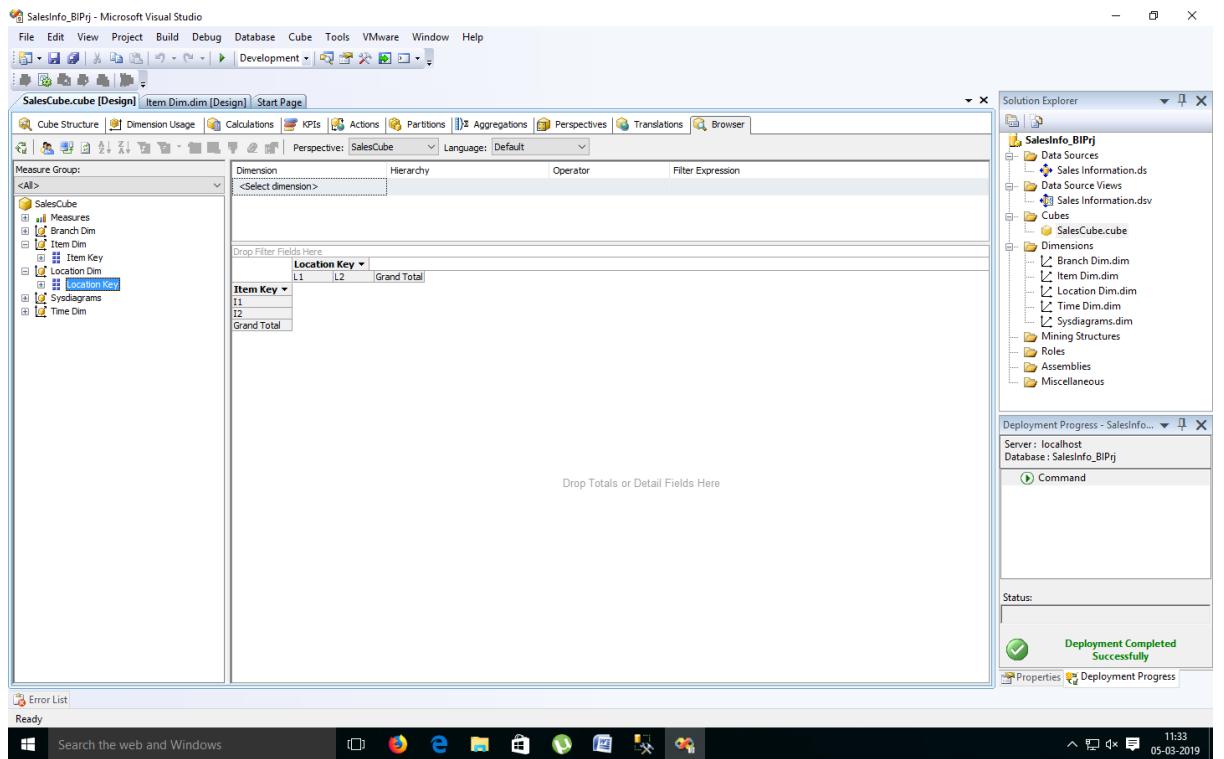
1. Double Click on “SalesCube”. Go to the “Browser” Tab.



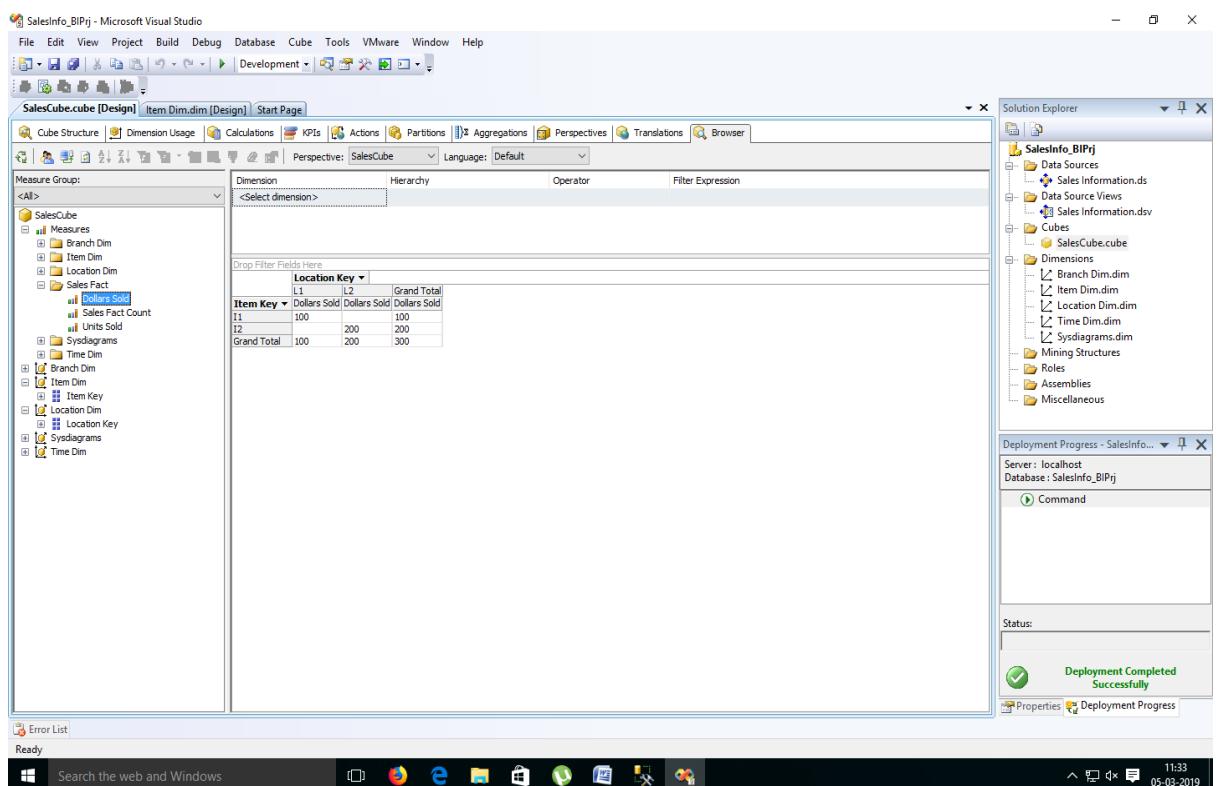
2. Go to the “Item Dimension”. Right Click on ‘Item Key’ -> Add to Row Area.



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.

The screenshot shows the Microsoft Visual Studio interface for a BI project named SalesInfo_BIPrj. The main window is the Cube Structure designer for SalesCube.cube. In the 'Dimensions' pane on the left, the 'Branch Dim' dimension is selected. The main pane displays a data grid for the 'Sales Fact' measure group, filtered by 'Item Key' and 'Branch Key'. The grid shows sales data for two branches (B1, B2) across three locations (L1, L2). The 'Solution Explorer' pane on the right shows the project structure, and the 'Deployment Progress' pane at the bottom indicates deployment completed successfully.

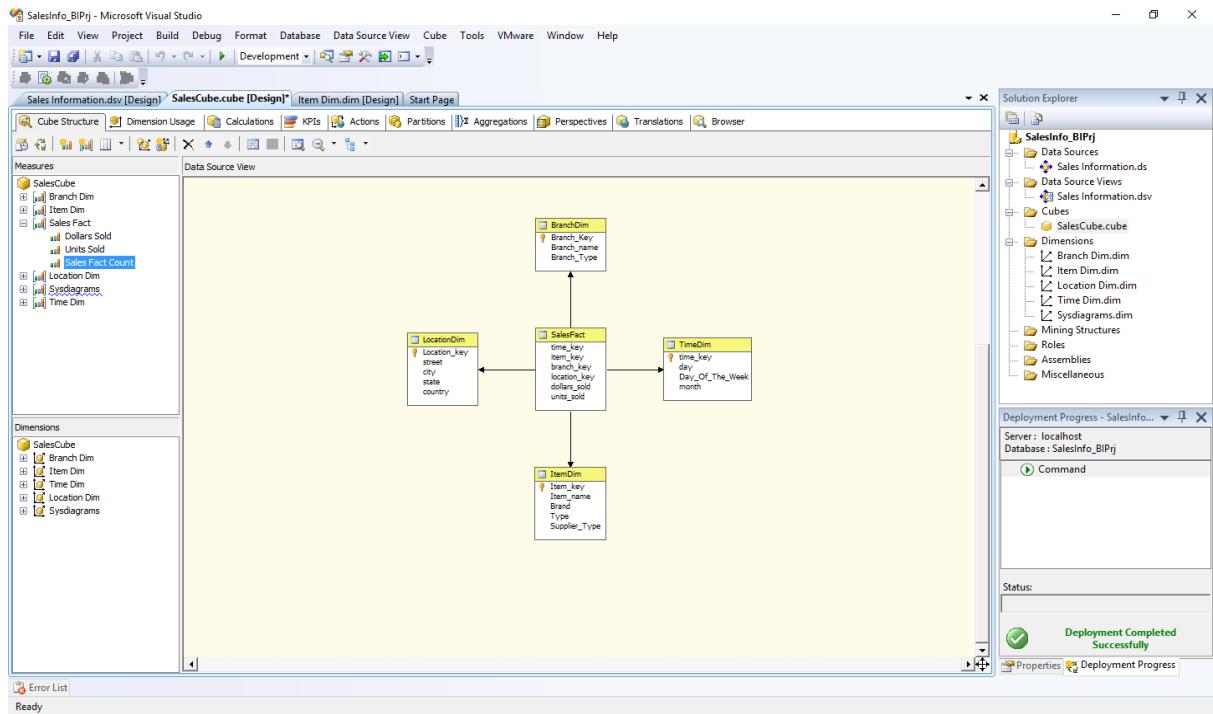
Item Key	Branch Key	Location Key	L1	L2	Grand Total
I1	B1		100		100
	B2				
	Total		100		100
I2	B1			200	200
	B2				
	Total		200	200	200
	Grand Total		100	200	300

Practical No 6

Aim : Working with measures in the cube.

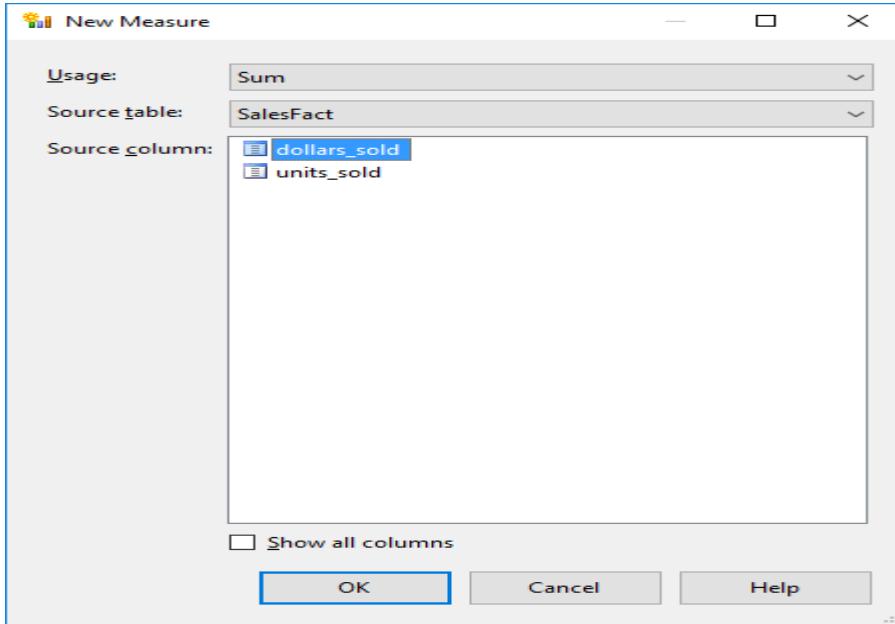
Solution :

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



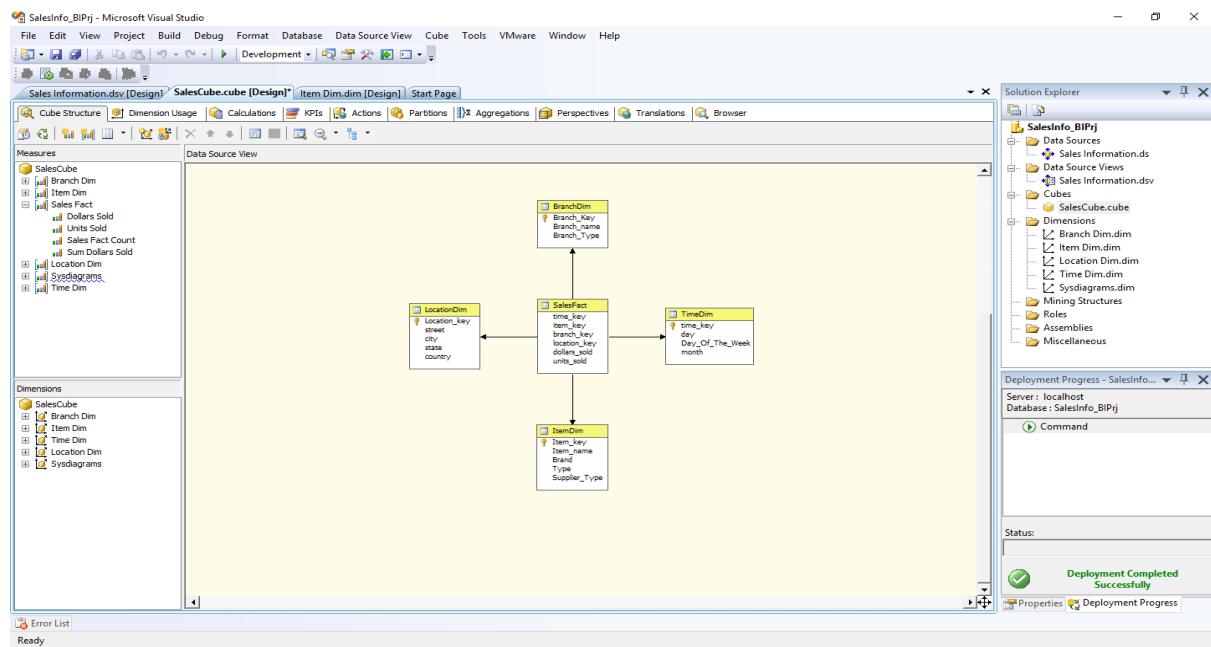
2. Right click on SalesCube -> New Measure.

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

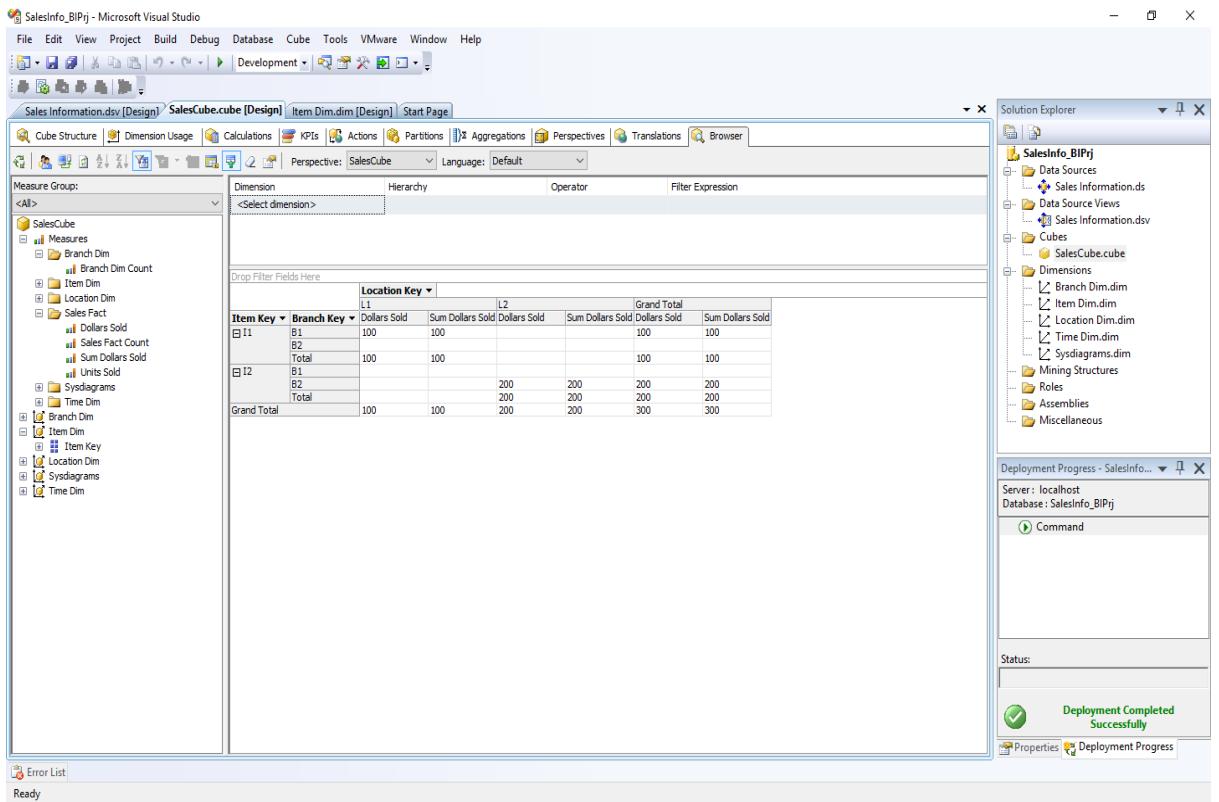


Click on OK.

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.

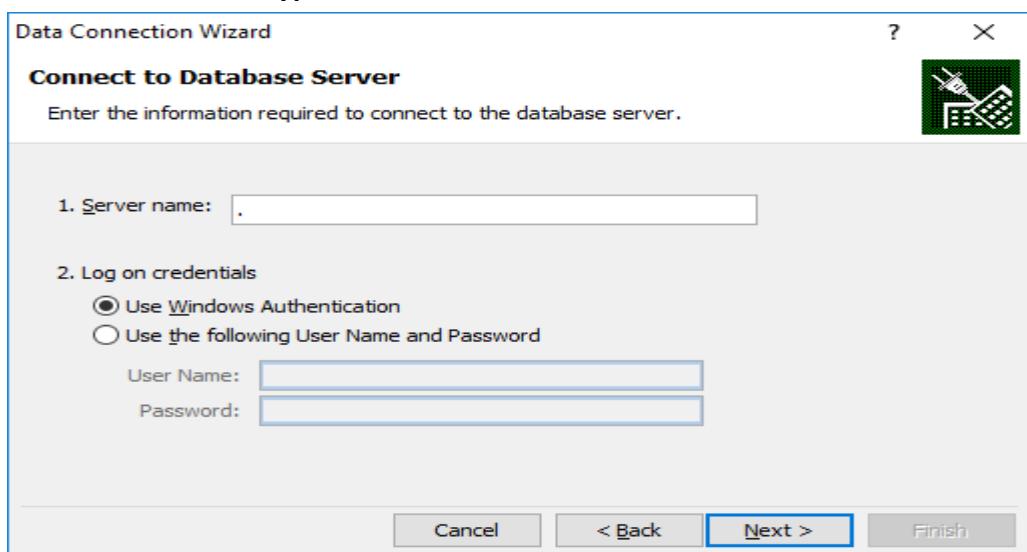


Practical No 7

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

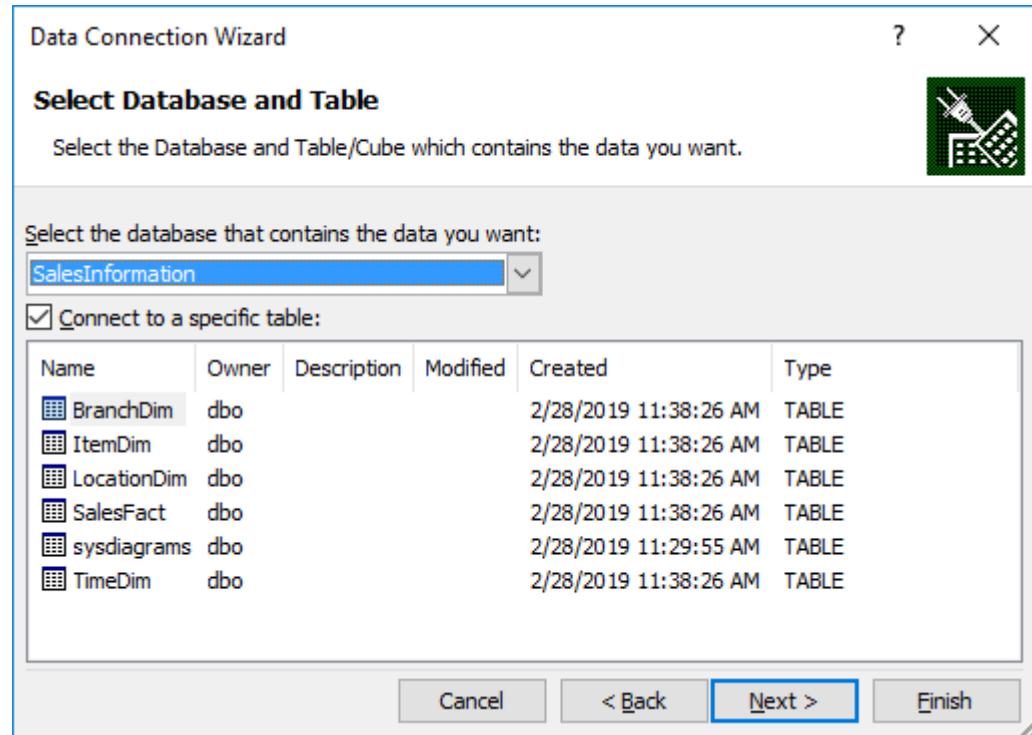
Solution :

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

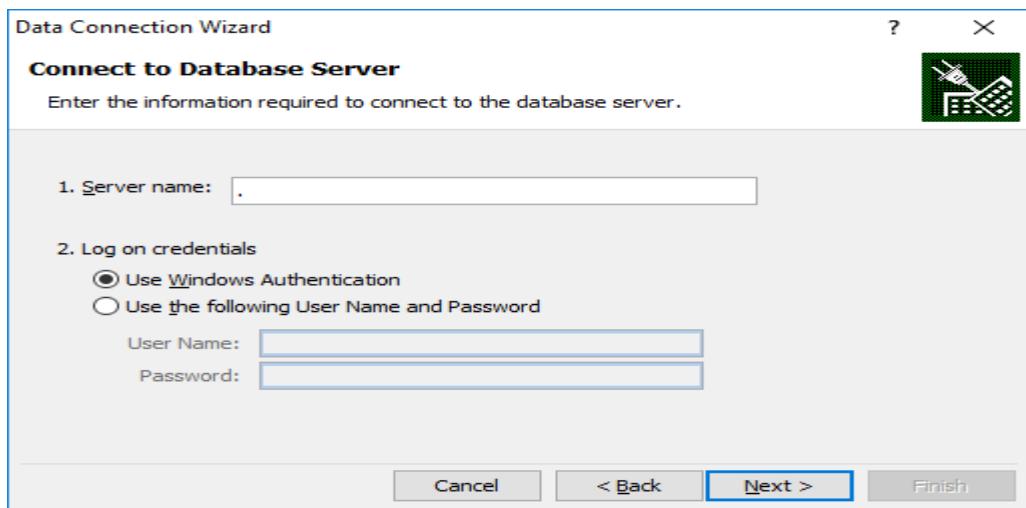


Click on Next.

Choose SQL Database -> “SalesInformation”

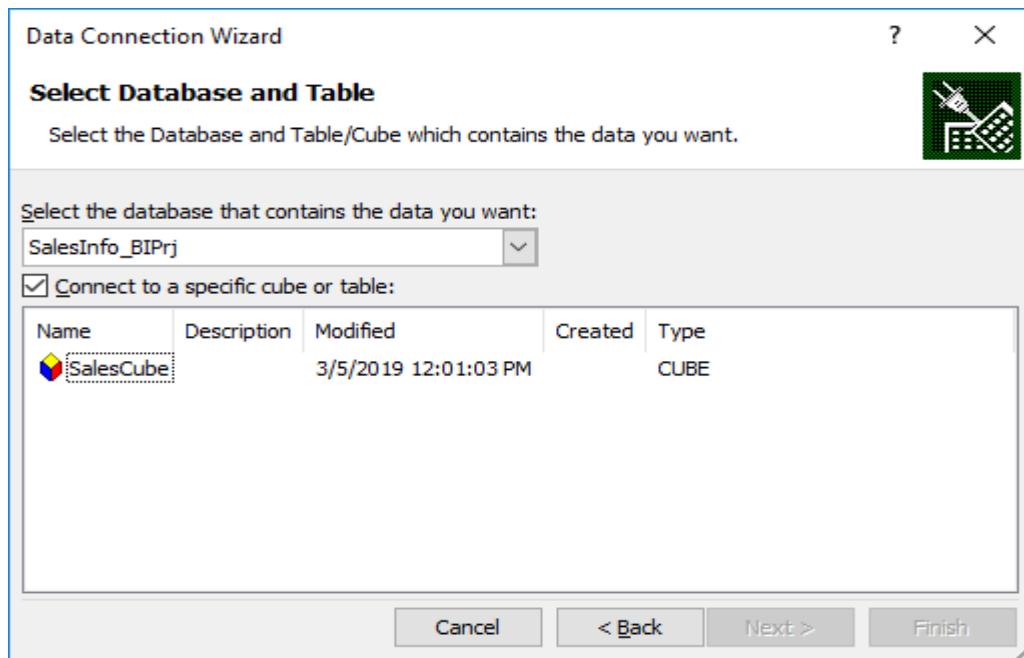


2.2. From Analysis Services -> Type Server name as ":"

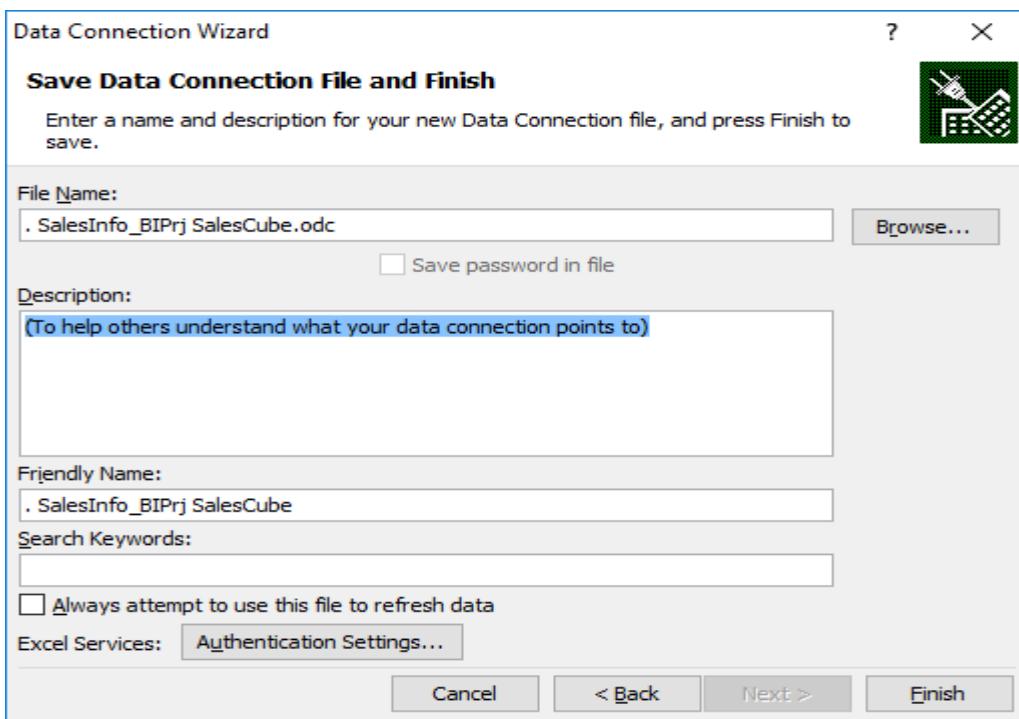


Click on Next.

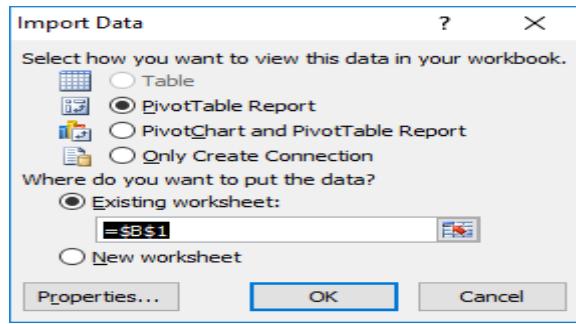
Choose Analysis Database as “SalesInfo_BIPrj”. Click on Next.



Click on OK



Click on Finish.



Click on OK.

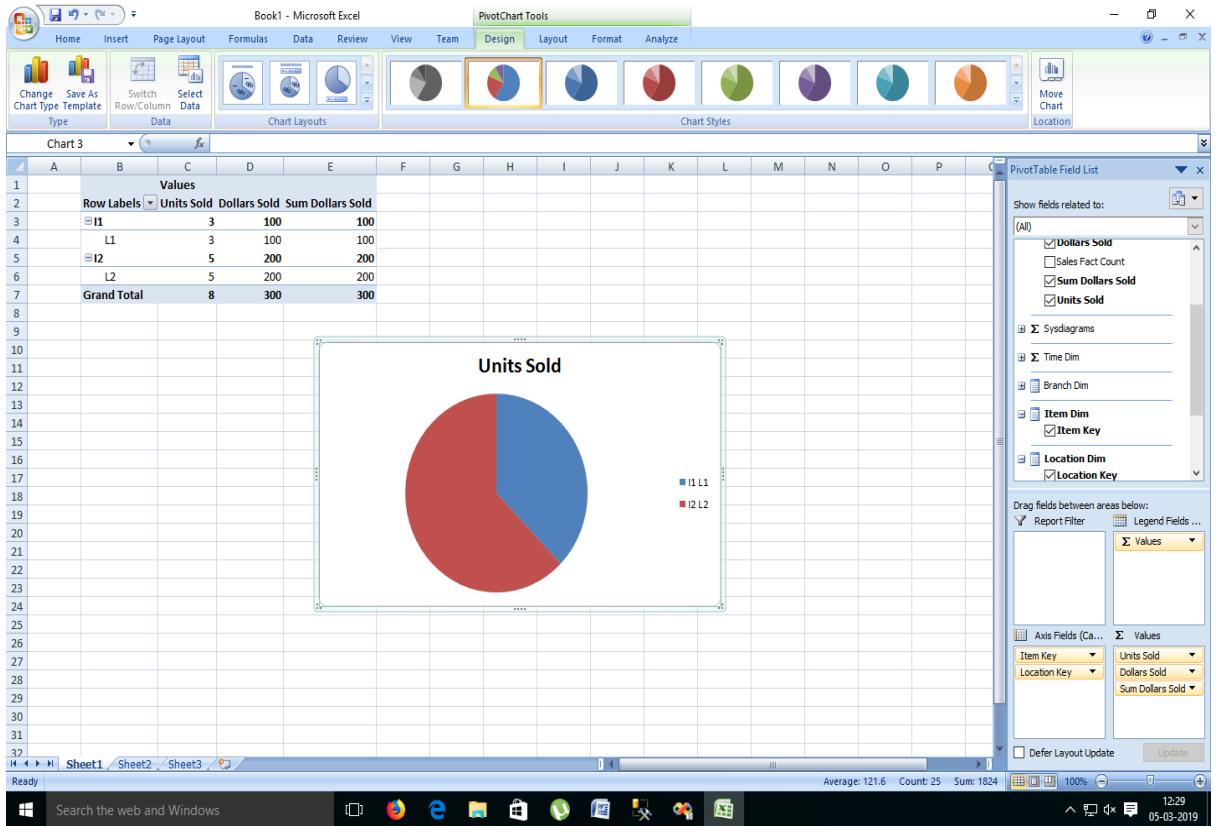
3. Select Item Key, Location Key and Measures as Dollars Sold, Units Sold and Sum Dollars Sold

The screenshot shows a Microsoft Excel window titled "Book1 - Microsoft Excel". The ribbon is at the top with tabs: Home, Insert, Page Layout, Formulas, Data, Review, View, Team, Options, Design, and PivotTable Tools (which is selected). The main area displays a PivotTable named "PivotTable1" with the following data:

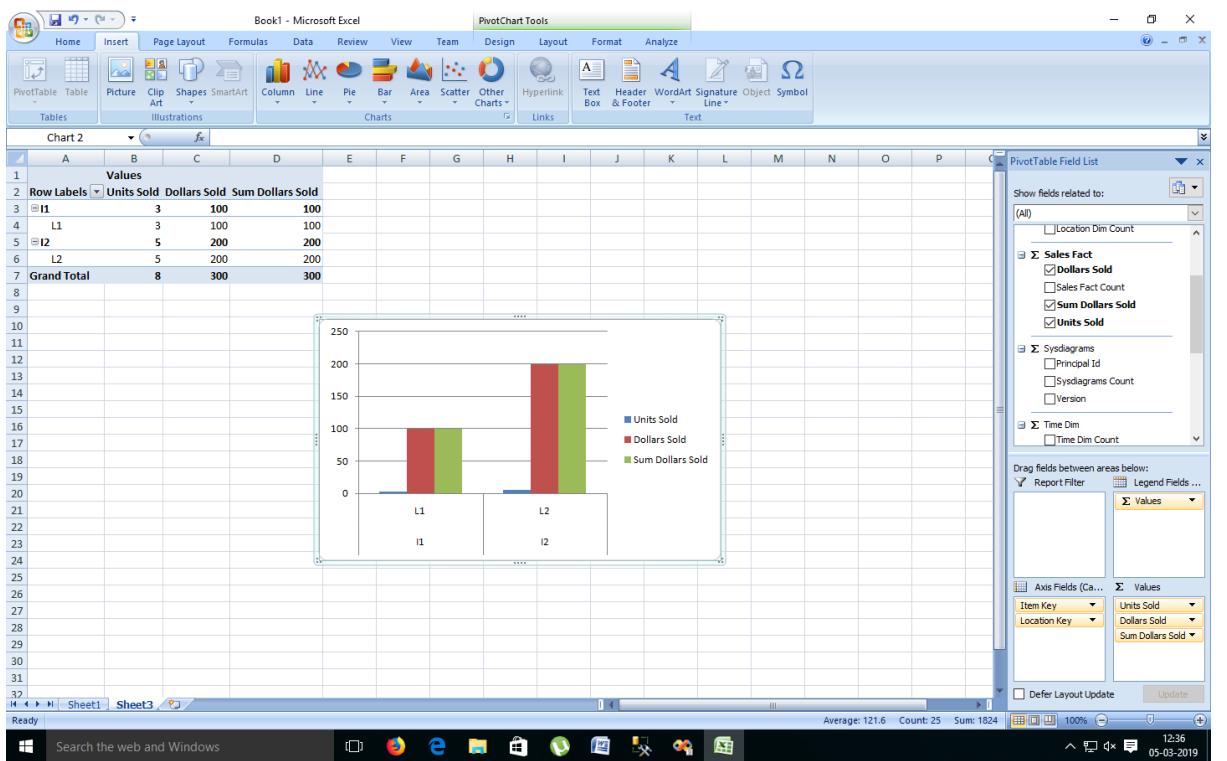
	Units Sold	Dollars Sold	Sum Dollars Sold
I1	3	100	100
L1	5	200	200
I2	5	200	200
Grand Total	8	300	300

To the right of the table is the "PivotTable Field List" pane. It shows fields related to "All" and includes "Dollars Sold", "Sum Dollars Sold", "Units Sold", "Σ Sydograms", "Σ Time Dim", "Branch Dim", "Item Dim" (with "Item Key" checked), and "Location Dim" (with "Location Key" checked). Below the field list are sections for "Report Filter" and "Column Labels", both containing "Values". The status bar at the bottom shows the date and time: "12:26 05-03-2019".

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]`**

Microsoft SQL Server Management Studio

File Edit View Query Debug Tools Window Community Help

New Query | Execute | Results | Messages | Properties

Object Explorer

Connect (local) SQL Server 10.50.1600 - BSC-37\admn (53) master

Databases

Tables

System Tables

Views

Synonyms

Programmability

Service Broker

Storage

Security

Replication

Object Explorer

SQLQuery16.sql - C:\37\admn (53)*

```
***** Script for SelectTopNRows command from SSMS *****
SELECT
    [Branch_key]
    ,[Branch_name]
    ,[Branch_Type]
FROM [SalesInformation].[dbo].[BranchDim]
```

Results

	Branch_Key	Branch_Name	Branch_Type
1	B1	RJ	Vendor
2	B2	JR	Vendor

Messages

Query executed successfully.

Properties

Aggregate Status

Connection fail.

Elapsed time 00:00:00.031

Finish time 08-03-2019 09:52:20

Rows returned 2

Start time 08-03-2019 09:52:20

State Open

Connection name (local)

Connection Details

Connection elap 00:00:00.031

Connection finis 08-03-2019 09:52:20

Connection row: 2

Connection start 08-03-2019 09:52:20

Connection stat: Open

Display name (local)

Login name BSC-37\admn

Server name (local)

Server version 10.50.1600

Session Tracing I

SPID 53

Name The name of the connection.

Output

Ready

Ln 5 Col 44 Ch 44 INS

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

Microsoft SQL Server Management Studio

File Edit View Query Debug Tools Window Community Help

New Query | Execute | Results | Messages | Properties

Object Explorer

Connect (local) SQL Server 10.50.1600 - BSC-37\admn (55) master

Databases

Tables

System Tables

Views

Synonyms

Programmability

Service Broker

Storage

Security

Replication

Object Explorer

SQLQuery17.sql - C:\37\admn (55)* SQLQuery16.sql - C:\37\admn (53)*

```
***** Script for SelectTopNRows command from SSMS *****
SELECT
    [Item_key]
    ,[Item_name]
    ,[Brand]
    ,[Type]
    ,[Supplier_Type]
FROM [SalesInformation].[dbo].[ItemDim]
```

Results

	Item_Key	Item_Name	Brand	Type	Supplier_Type
1	Laptop	LG	Accessories	VT	
2	Mouse	LG	Accessories	VT	

Messages

Query executed successfully.

Properties

Aggregate Status

Connection fail.

Elapsed time 00:00:00.037

Finish time 08-03-2019 09:54:15

Name (local)

Rows returned 2

Start time 08-03-2019 09:54:15

State Open

Connection name (local)

Connection Details

Connection elap 00:00:00.037

Connection finis 08-03-2019 09:54:15

Connection row: 2

Connection start 08-03-2019 09:54:15

Connection stat: Open

Display name (local)

Login name BSC-37\admn

Server name (local)

Server version 10.50.1600

Session Tracing I

SPID 55

Name The name of the connection.

Output

Ready

Ln 7 Col 42 Ch 42 INS

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

```

Microsoft SQL Server Management Studio -> master | Execute > SQLQuery18.sql | SQLQuery17.sql | SQLQuery16.sql | SQLQuery15.sql | SQLQuery14.sql | SQLQuery13.sql | SQLQuery12.sql | SQLQuery11.sql | SQLQuery10.sql | SQLQuery9.sql | SQLQuery8.sql | SQLQuery7.sql | SQLQuery6.sql | SQLQuery5.sql | SQLQuery4.sql | SQLQuery3.sql | SQLQuery2.sql | SQLQuery1.sql | Properties
File Edit View Query Debug Tools Window Community Help
Object Explorer
Connect > (local) [SQL Server 10.50.1600 - BSC -> master | Databases | System Databases | Database Snapshots | ReportServer | ReportServerTempDB | Sales | SalesInformation | Database Diagrams | dbo.SalesInfo | Tables | System Tables | dbo.BranchDim | dbo.ItemDim | dbo.LocationDim | dbo.SaleFact | dbo.SaleFact | dbo.TimeDim | Views | Synonyms | Programmability | Service Broker | Storage | Security | SalesProduct | Security | Server Objects | Replication
SQLQuery18.sql -> C:\37\admin (57)* | SQLQuery17.sql -> C:\37\admin (55)* | SQLQuery16.sql -> C:\37\admin (53)*
***** Script for SelectTopNRows command from SSMS *****/
SELECT [Location_key]
      ,[street]
      ,[city]
      ,[state]
      ,[country]
   FROM [SalesInformation].[dbo].[LocationDim]

Results Messages
Location_key street city state country
1 L1 LBS Mumbai Maharashtra India
2 L2 JM Thane Maharashtra India

Query executed successfully.

```

Properties

Current connection parameters

Aggregate Status

Connection fail.

Elapsed time 00:00:00.053

Finish time 08-03-2019 09:57:38

Name (local)

Rows returned 2

Start time 08-03-2019 09:57:38

State Open

Connection

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

Connection Details

Connection elap 00:00:00.053

Connection finis 08-03-2019 09:57:38

Connection row 2

Connection start 08-03-2019 09:57:38

Connection stat: Open

Display name (local)

Login name BSC-37\admin

Server name (local)

Server version 10.50.1600

Session Tracing I

SPID 57

Name The name of the connection.

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

```

Microsoft SQL Server Management Studio -> master | Execute > SQLQuery19.sql | SQLQuery18.sql | SQLQuery17.sql | SQLQuery16.sql | SQLQuery15.sql | SQLQuery14.sql | SQLQuery13.sql | SQLQuery12.sql | SQLQuery11.sql | SQLQuery10.sql | SQLQuery9.sql | SQLQuery8.sql | SQLQuery7.sql | SQLQuery6.sql | SQLQuery5.sql | SQLQuery4.sql | SQLQuery3.sql | SQLQuery2.sql | SQLQuery1.sql | Properties
File Edit View Query Debug Tools Window Community Help
Object Explorer
Connect > (local) [SQL Server 10.50.1600 - BSC -> master | Databases | System Databases | Database Snapshots | ReportServer | ReportServerTempDB | Sales | SalesInformation | Database Diagrams | dbo.SalesInfo | Tables | System Tables | dbo.BranchDim | dbo.ItemDim | dbo.LocationDim | dbo.SaleFact | dbo.SaleFact | dbo.TimeDim | Views | Synonyms | Programmability | Service Broker | Storage | Security | SalesProduct | Security | Server Objects | Replication
SQLQuery19.sql -> C:\37\admin (58)* | SQLQuery18.sql -> C:\37\admin (57)* | SQLQuery17.sql -> C:\37\admin (55)* | SQLQuery16.sql -> C:\37\admin (53)*
***** Script for SelectTopNRows command from SSMS *****/
SELECT [time_key]
      ,[item_key]
      ,[branch_key]
      ,[location_key]
      ,[dollars_sold]
      ,[units_sold]
   FROM [SalesInformation].[dbo].[SalesFact]

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

Query executed successfully.

```

Properties

Current connection parameters

Aggregate Status

Connection fail.

Elapsed time 00:00:00.051

Finish time 08-03-2019 09:58:08

Name (local)

Rows returned 2

Start time 08-03-2019 09:58:08

State Open

Connection

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

Connection Details

Connection elap 00:00:00.051

Connection finis 08-03-2019 09:58:08

Connection row 2

Connection start 08-03-2019 09:58:08

Connection stat: Open

Display name (local)

Login name BSC-37\admin

Server name (local)

Server version 10.50.1600

Session Tracing I

SPID 58

Name The name of the connection.

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

```

Microsoft SQL Server Management Studio
File Edit View Query Debug Tools Window Community Help
New Query Execute
Object Explorer
Connect (local) (SQL Server 10.50.1600 - BSC-A)
Databases System Databases Database Snapshots ReportServer ReportServerTempDB Sales SalesInformation Database Diagrams dbo.SalesInfo Tables System Tables dbo.BranchDim dbo.ItemDim dbo.LocationDim dbo.SaleFact dbo.TimeDim Views Synonyms Programmability Service Broker Storage Security SalesProduct Security Server Objects Replication
Results Messages
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
Query executed successfully.
Properties
Current connection parameters
Aggregate Status
Connection fail.
Elapsed time 00:00:00.046
Finish time 08-03-2019 10:00:08
Name (local)
Rows returned 2
Start time 08-03-2019 10:00:08
State Open
Connection
Connection name (local) (BSC-37\admi...
Connection Details
Connection elap 00:00:00.046
Connection finis 08-03-2019 10:00:08
Connection row 2
Connection start 08-03-2019 10:00:08
Connection stat Open
Display name (local)
Login name BSC-37\admin
Server name (local)
Server version 10.50.1600
Session Tracing I
SPID 61
Name
The name of the connection.

```

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];`

```

Microsoft SQL Server Management Studio
File Edit View Query Debug Tools Window Community Help
New Query Execute
Object Explorer
Connect (local) (SQL Server 10.50.1600 - BSC-A)
Databases System Databases Database Snapshots ReportServer ReportServerTempDB Sales SalesInformation Database Diagrams dbo.SalesInfo Tables System Tables dbo.BranchDim dbo.ItemDim dbo.LocationDim dbo.SaleFact dbo.TimeDim Views Synonyms Programmability Service Broker Storage Security SalesProduct Security Server Objects Replication
Results Messages
Branch_Key Branch_name dollars_sold units_sold
1 B1 RJ 100 3
2 B2 JR 200 5
Query executed successfully.
Properties
Current connection parameters
Aggregate Status
Connection fail.
Elapsed time 00:00:00.015
Finish time 08-03-2019 10:07:23
Name (local)
Rows returned 2
Start time 08-03-2019 10:07:23
State Open
Connection
Connection name (local) (BSC-37\admi...
Connection Details
Connection elap 00:00:00.015
Connection finis 08-03-2019 10:07:23
Connection row 2
Connection start 08-03-2019 10:07:23
Connection stat Open
Display name (local)
Login name BSC-37\admin
Server name (local)
Server version 10.50.1600
Session Tracing I
SPID 57
Name
The name of the connection.

```

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 T-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
I1	Laptop	Accessories	100	3
I2	Mouse	Accessories	200	5

The status bar at the bottom indicates "Query executed successfully." and provides session 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 T-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	I1	100	3
L2	Thane	I2	200	5

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

At the bottom of the screen, a Windows taskbar shows the date and time as 11:13 08-03-2019.

Practical No 9

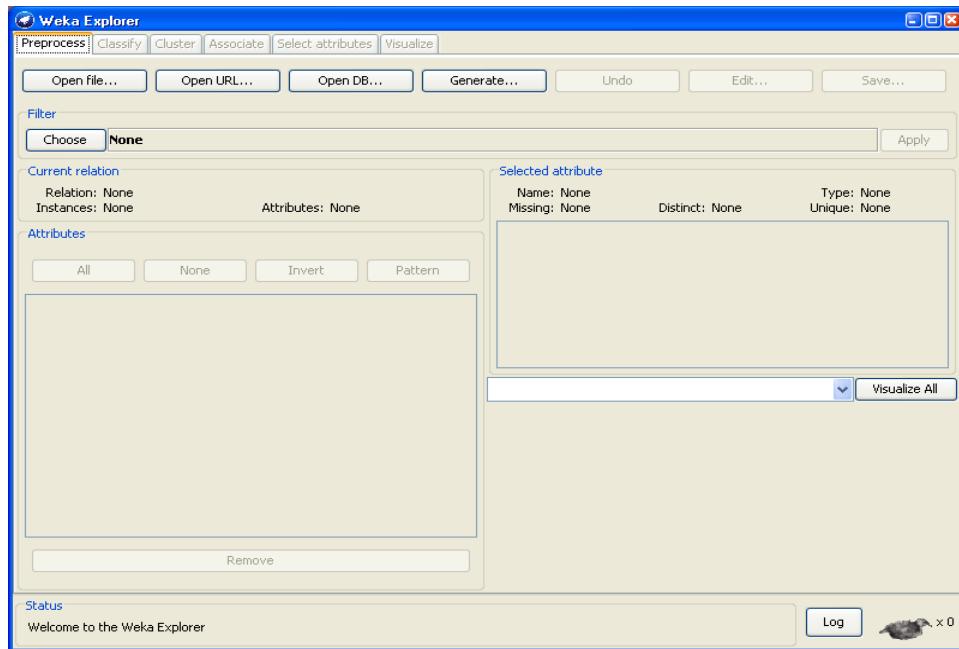
Aim : Calculation & KPI

Practical No - 10

Aim : Data PreProcessing

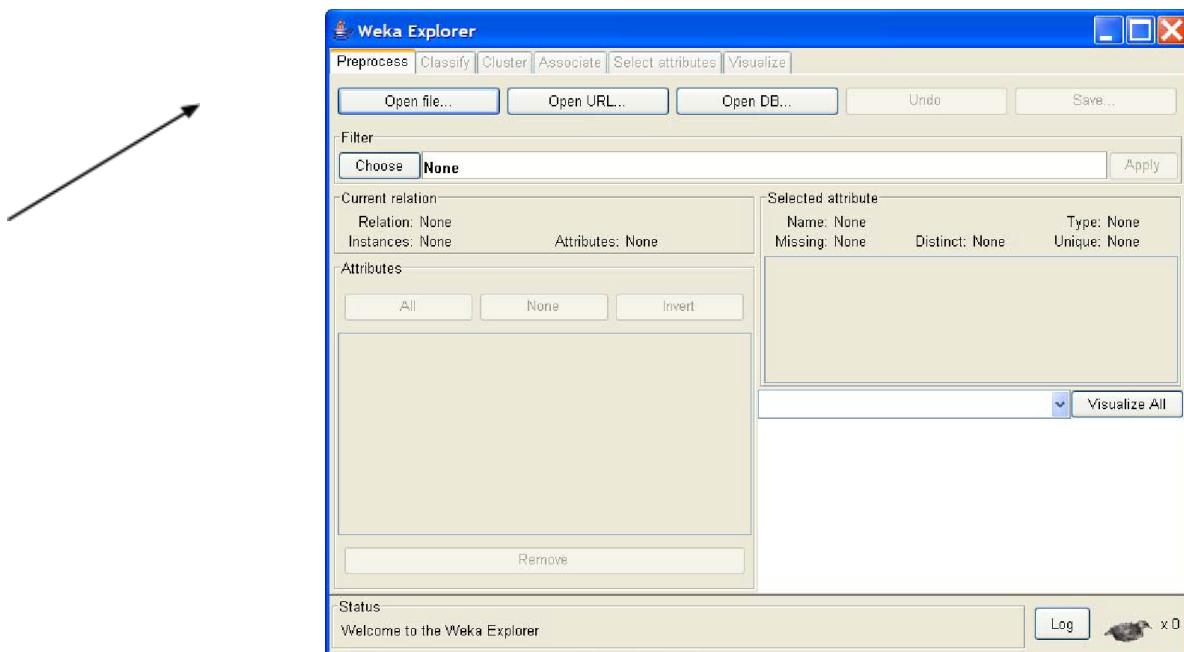
Solution :

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



Opening file from a local file system

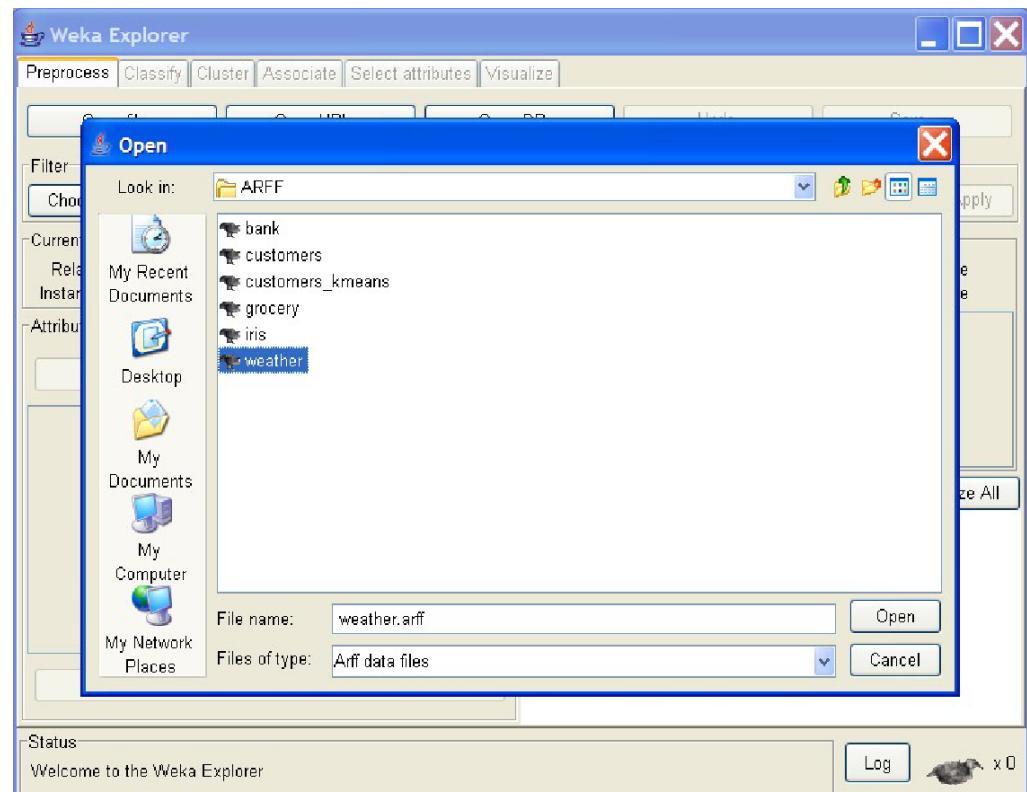
Click on ‘Open file…’ button



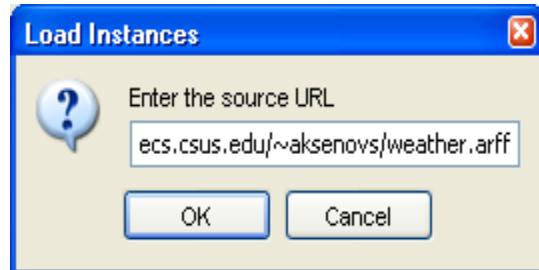
It brings up a dialog box allowing you to browse for the data file on the local file system, choose

“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:

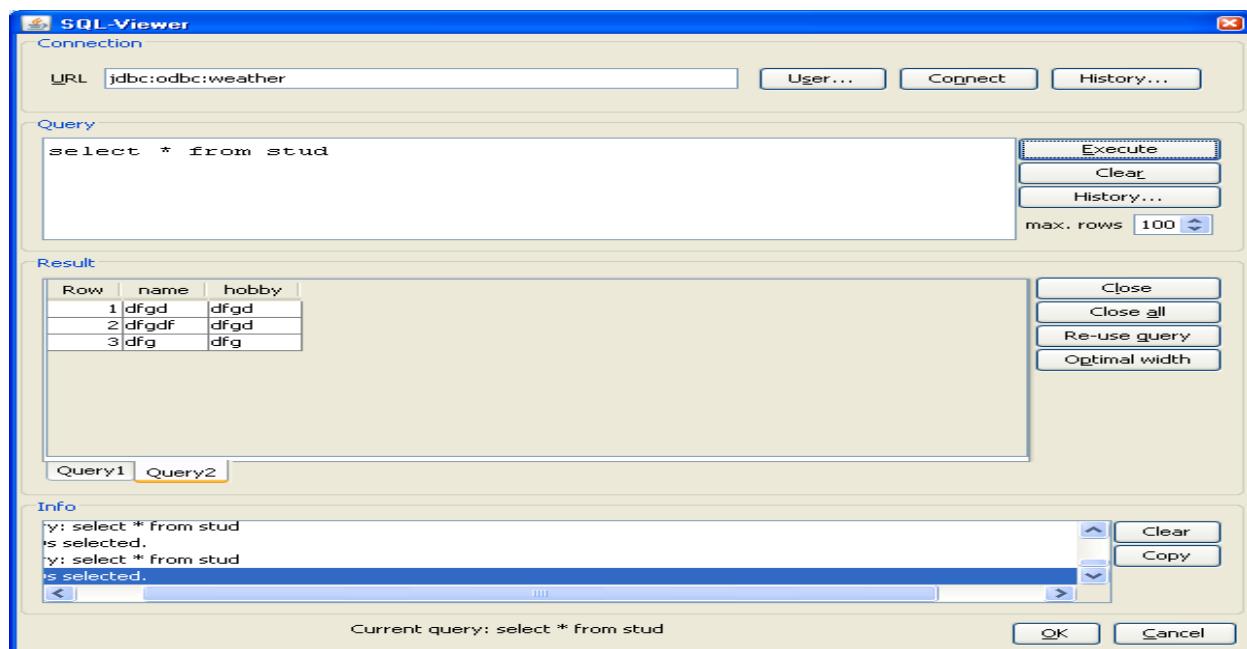


Reading data from a database:

The screenshot shows a web browser window with the following content:

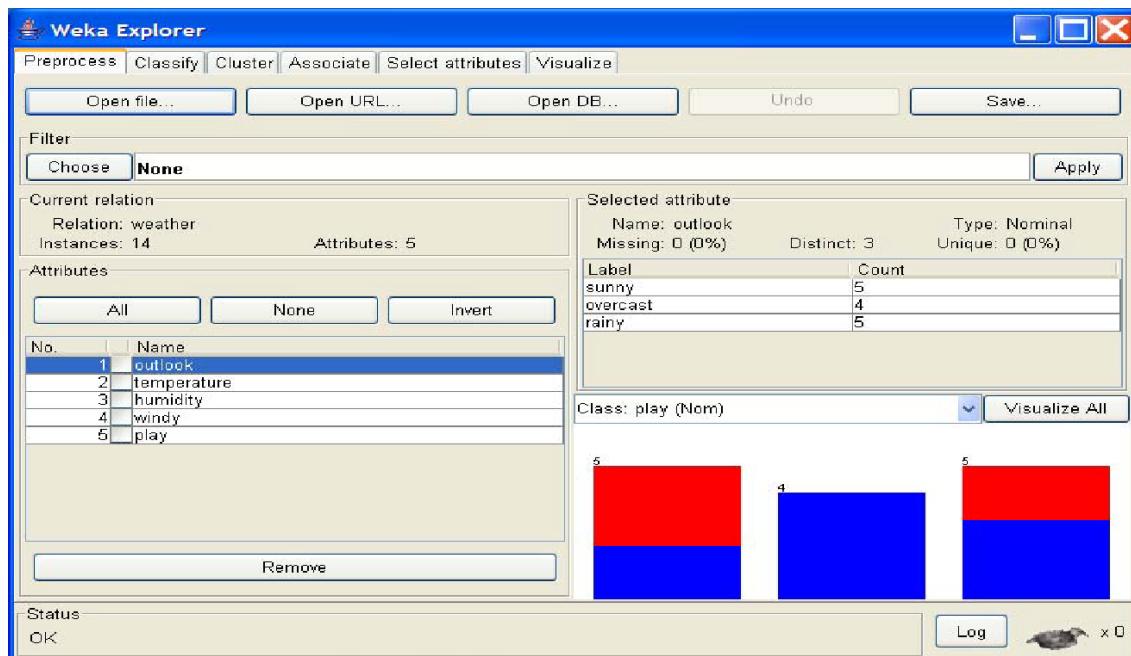
- Assignments:**
 - Assignment 1 - Vocabulary and Case Study
 - Assignment 2 - Classification
 - Assignment 3 - Clustering
 - Assignment 4 - Association Rules
 - Assignment 5 - Credibility
- Paper Review:**
 - Summary
 - PowerPoint Presentation
- Term Project:**
 - Project Poposal
 - Project Progress Report
 - Project Progress Report Presentation
- WEKA Tutorial**
- WEKA Tutorial Presentation**

At the bottom of the page, there is a link to 'weather.arff'.

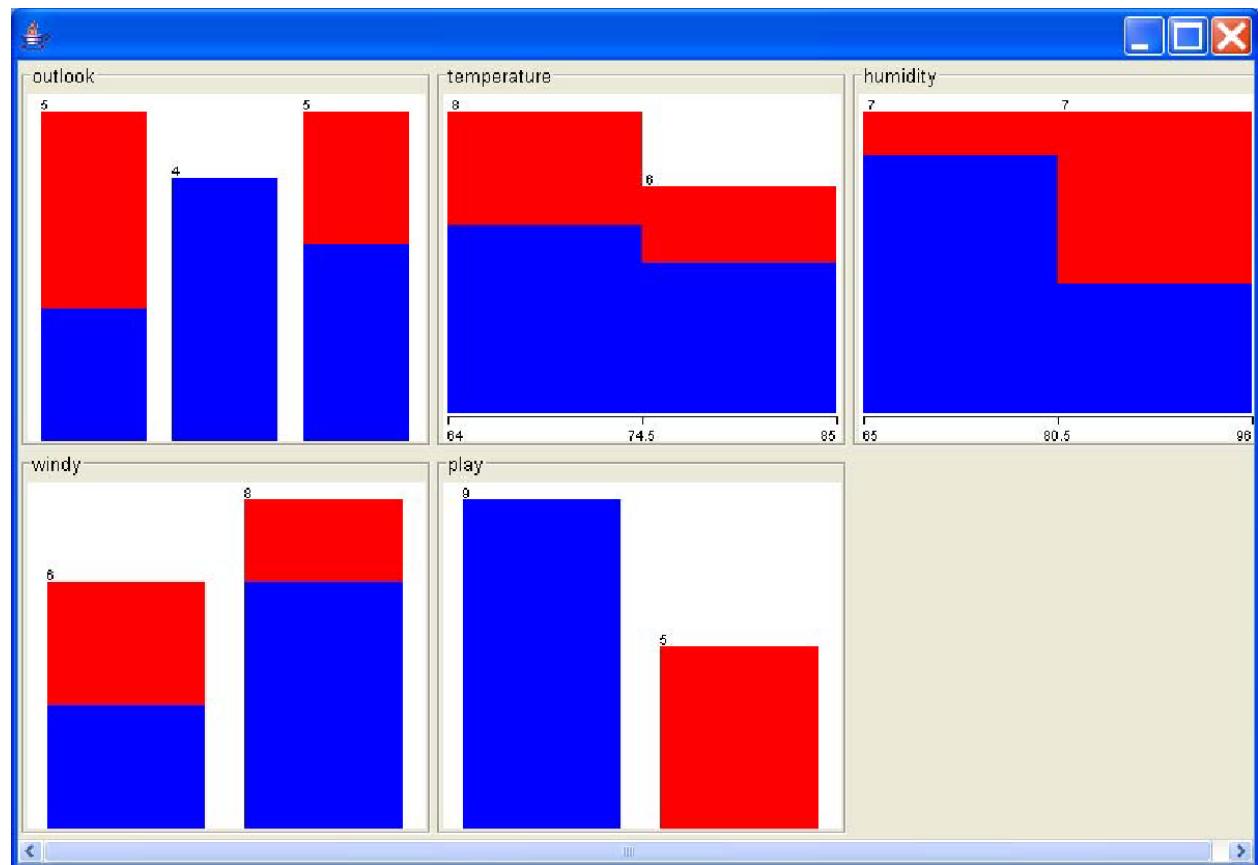


Loading data

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.,,

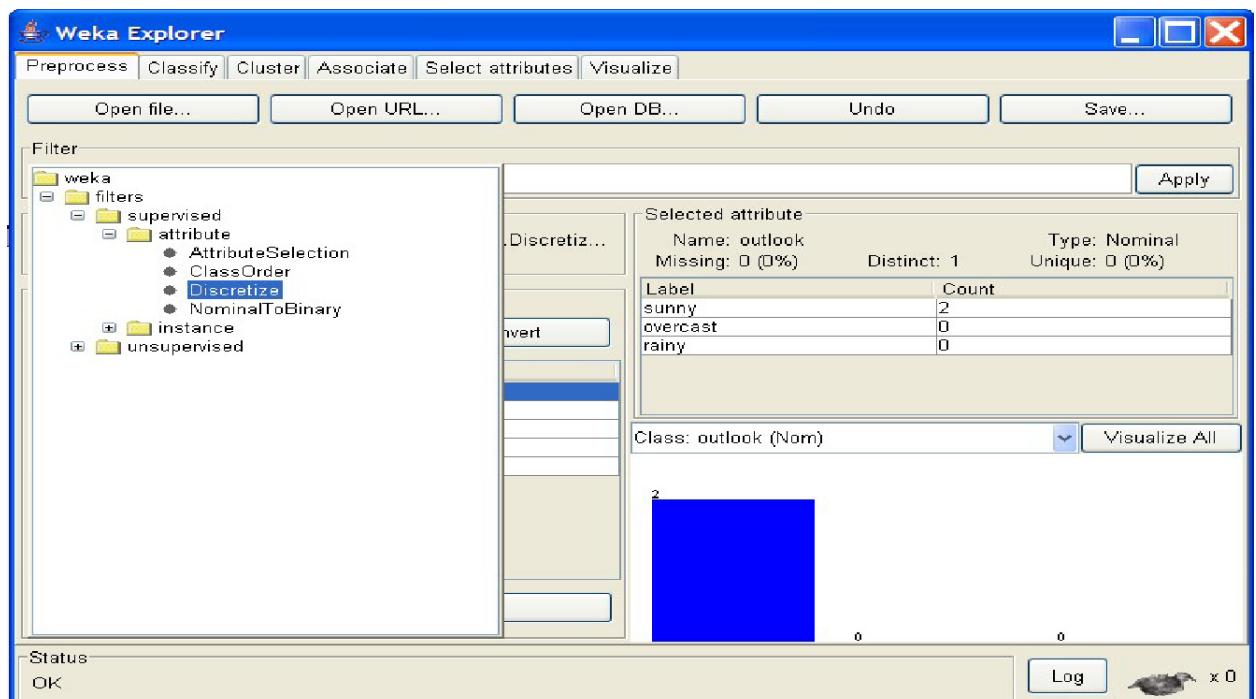
Practical No – 11

Aim : Data discretization.

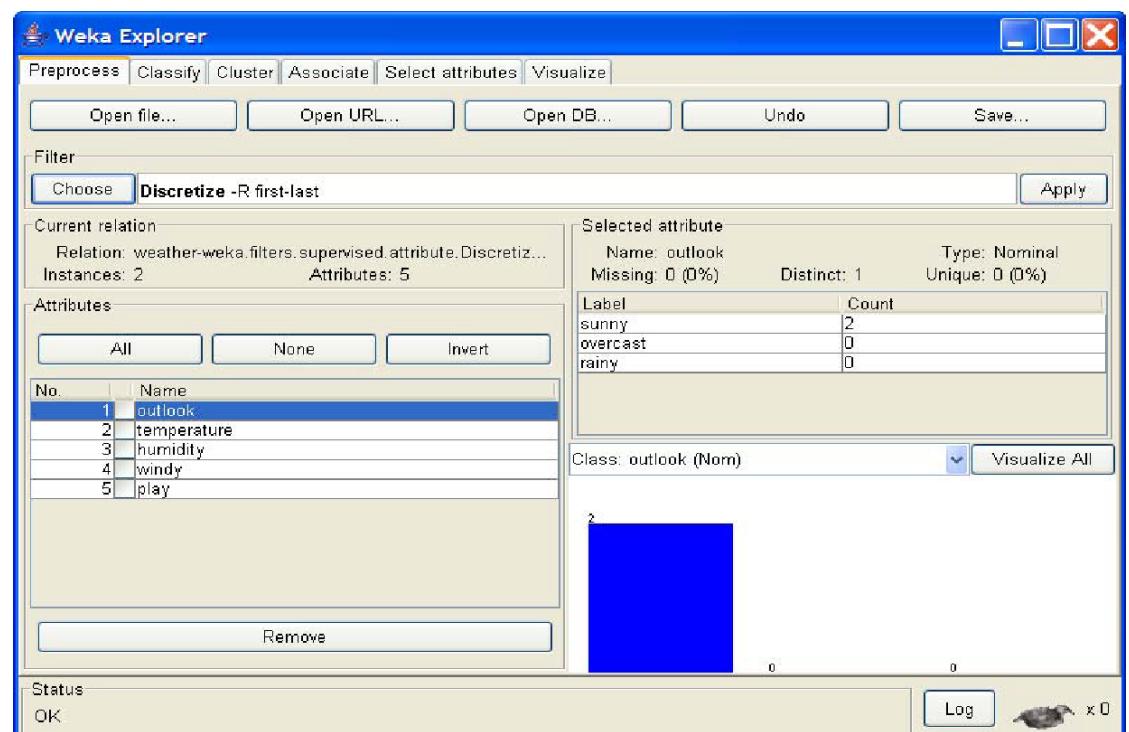
Solution :

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

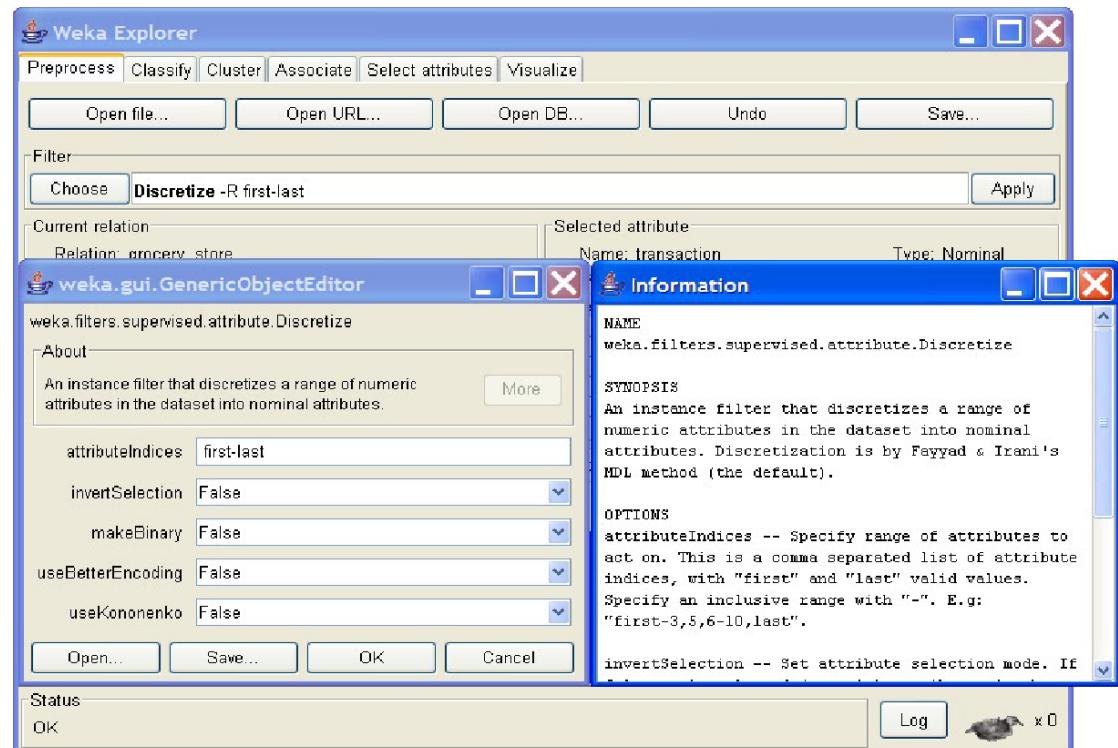
This will show pull-down menu with a list of available filters. Select Supervised Attribute Discretize and click on ‘Apply’ button.



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 choose the filter configuration options.

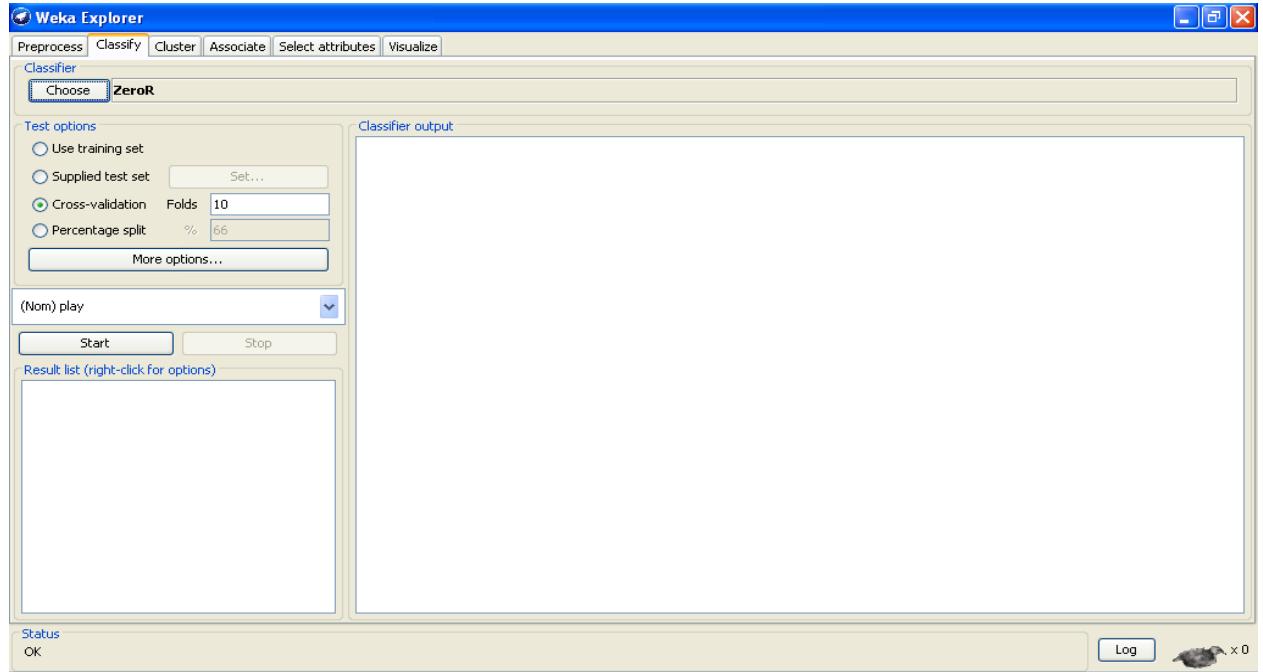


Practical No - 12

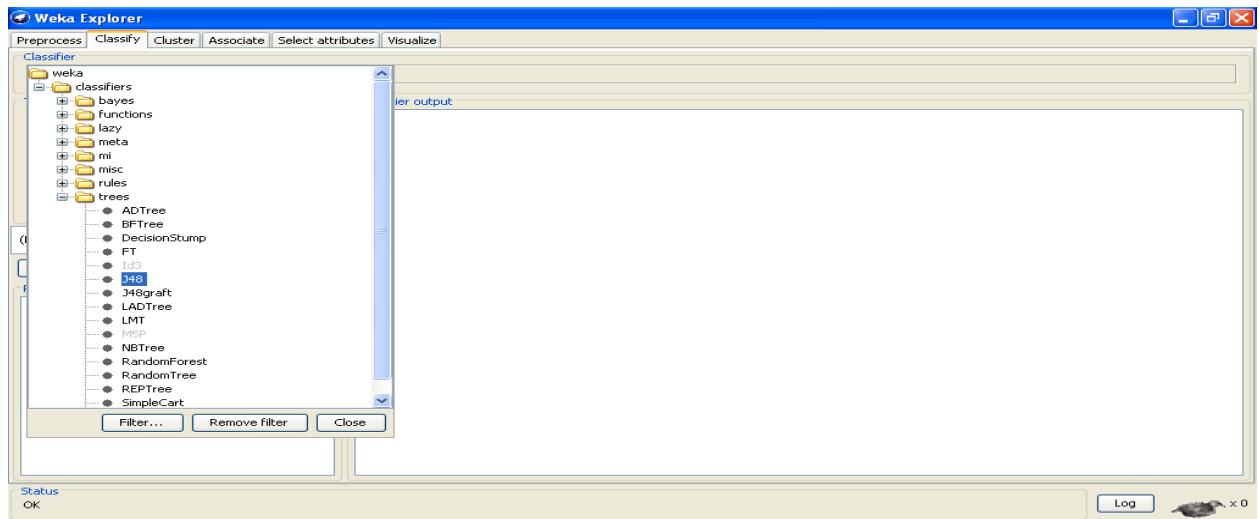
Aim : Classification problems.

Solution :

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



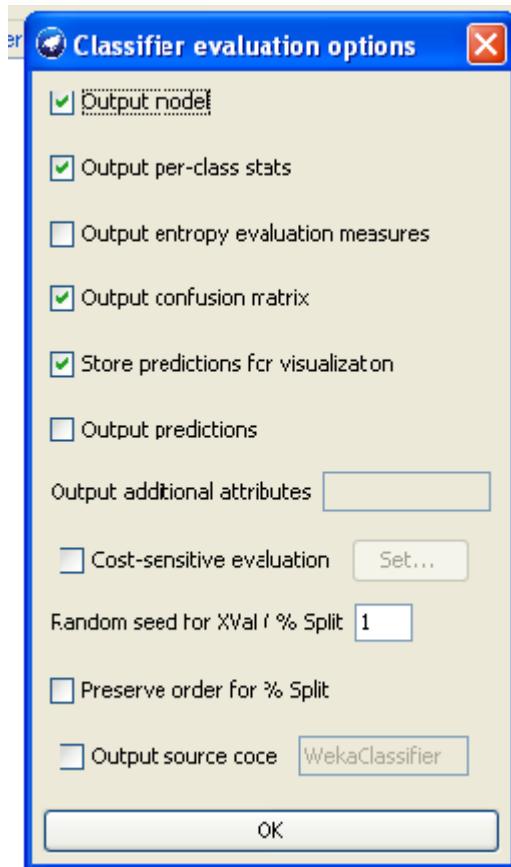
Click on ‘Choose’ button in the ‘Classifier’ box just below the tabs and select C4.5 classifier WEKA Classifiers Trees J48.



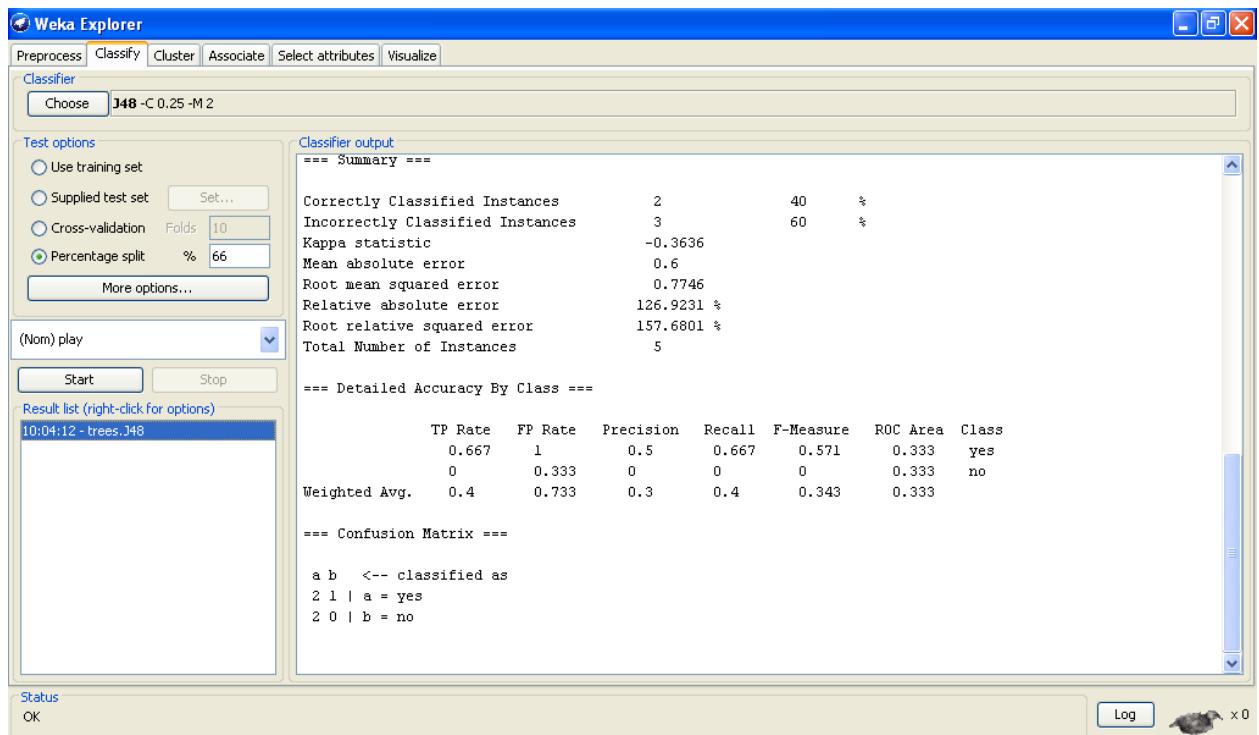
Check ‘Percentage split’ radio-button and keep it as default 66%. Click on ‘More options...’ button.

make sure that the Following options are checked :

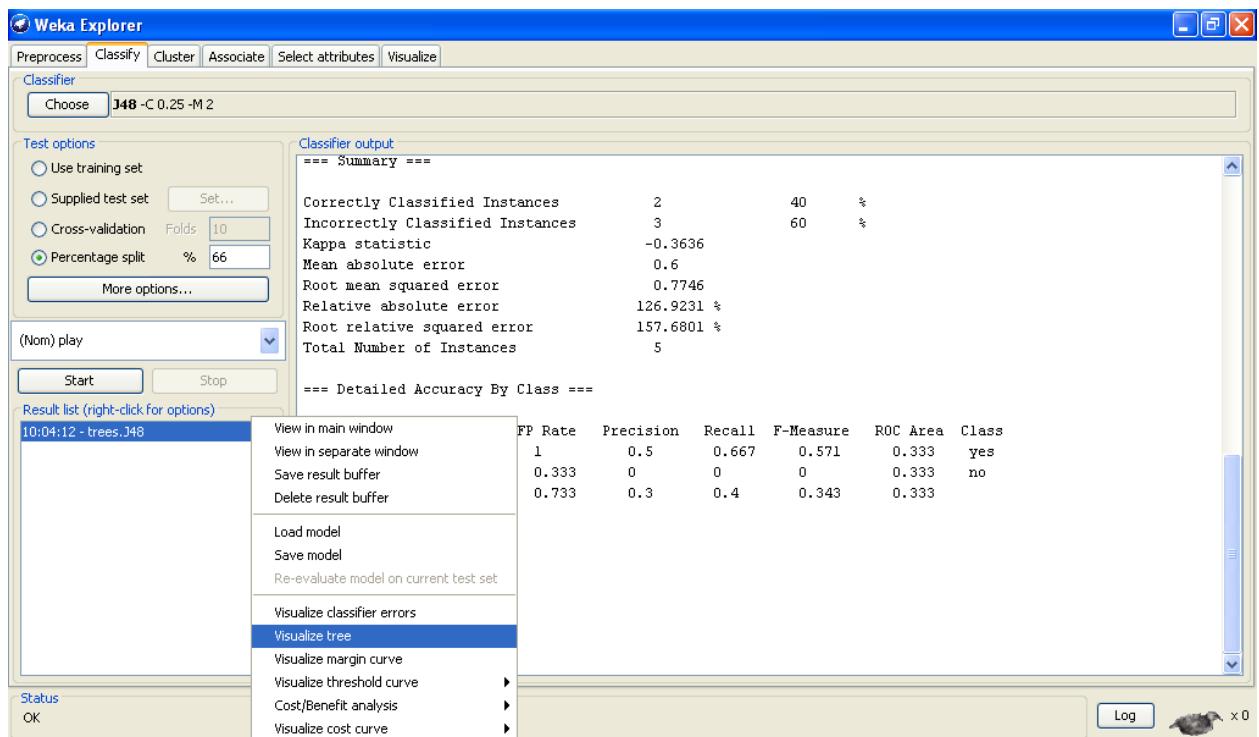
1. Output model.
2. Output per-class stats.
3. Output confusion matrix
4. Store predictions for visualization.
5. Set ‘Random seed for Xval / % Split’ to 1.



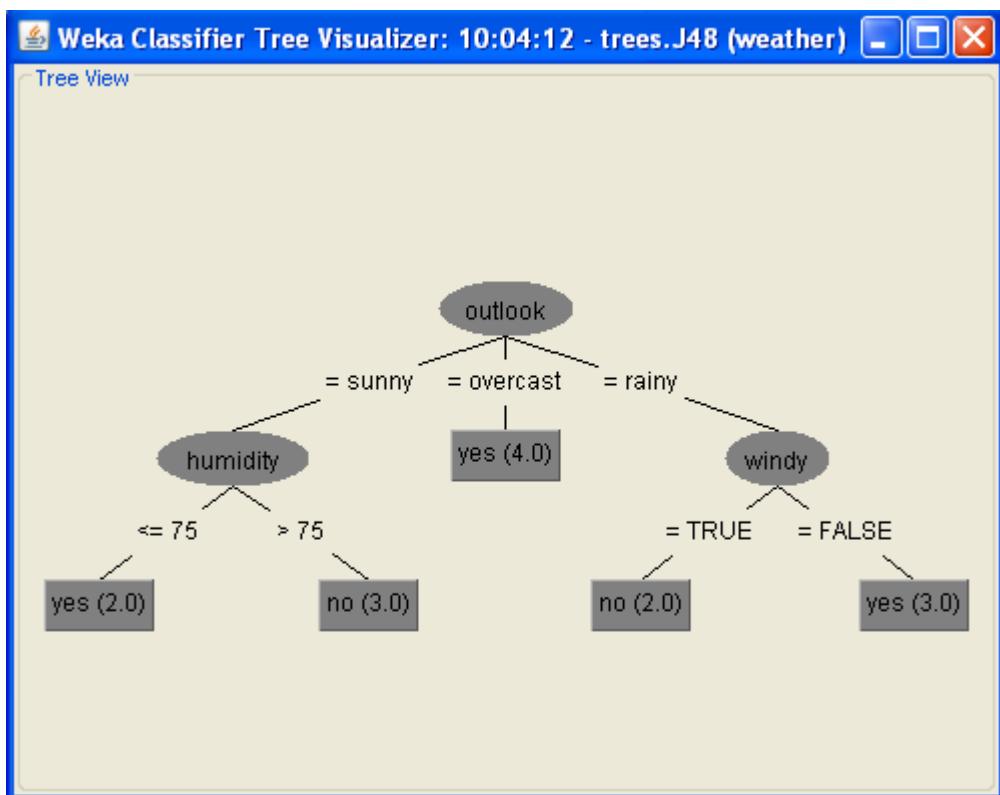
Once the options have been specified, you can run the classification algorithm. Click on ‘Start’ button



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.

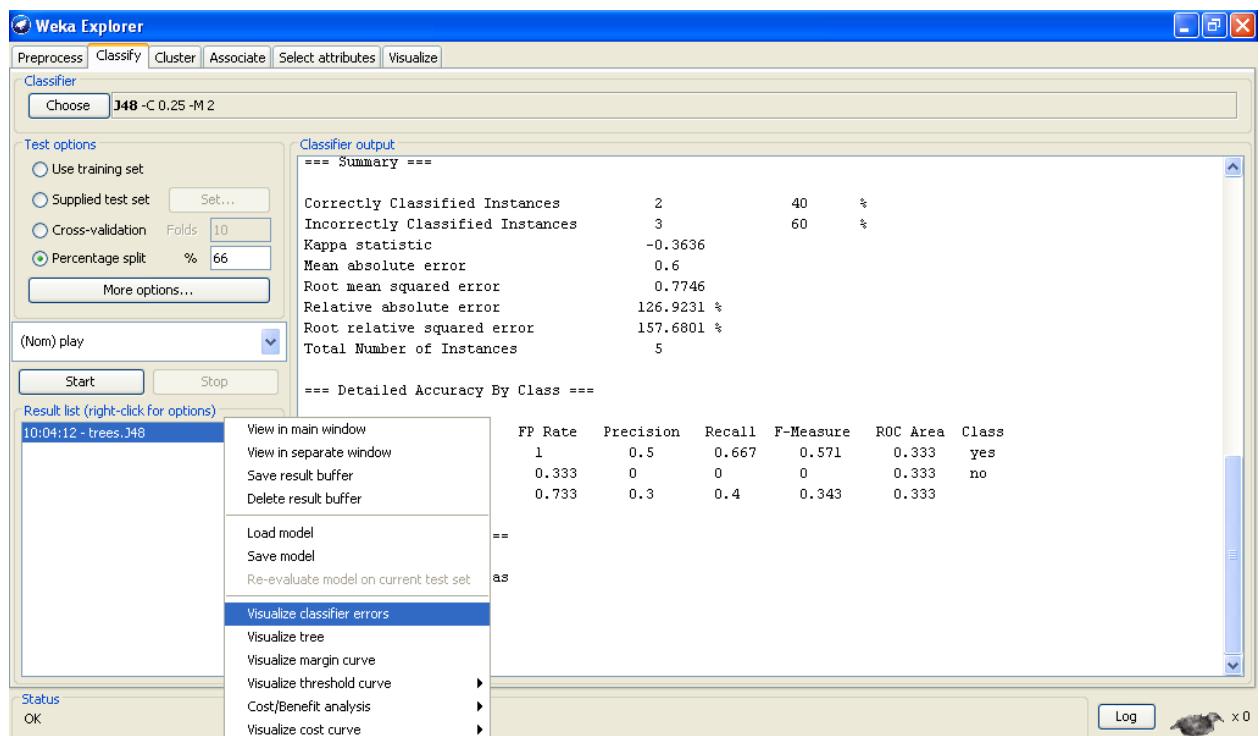


Select the item ‘Visualize tree’; a new window comes up to the screen displaying the tree.

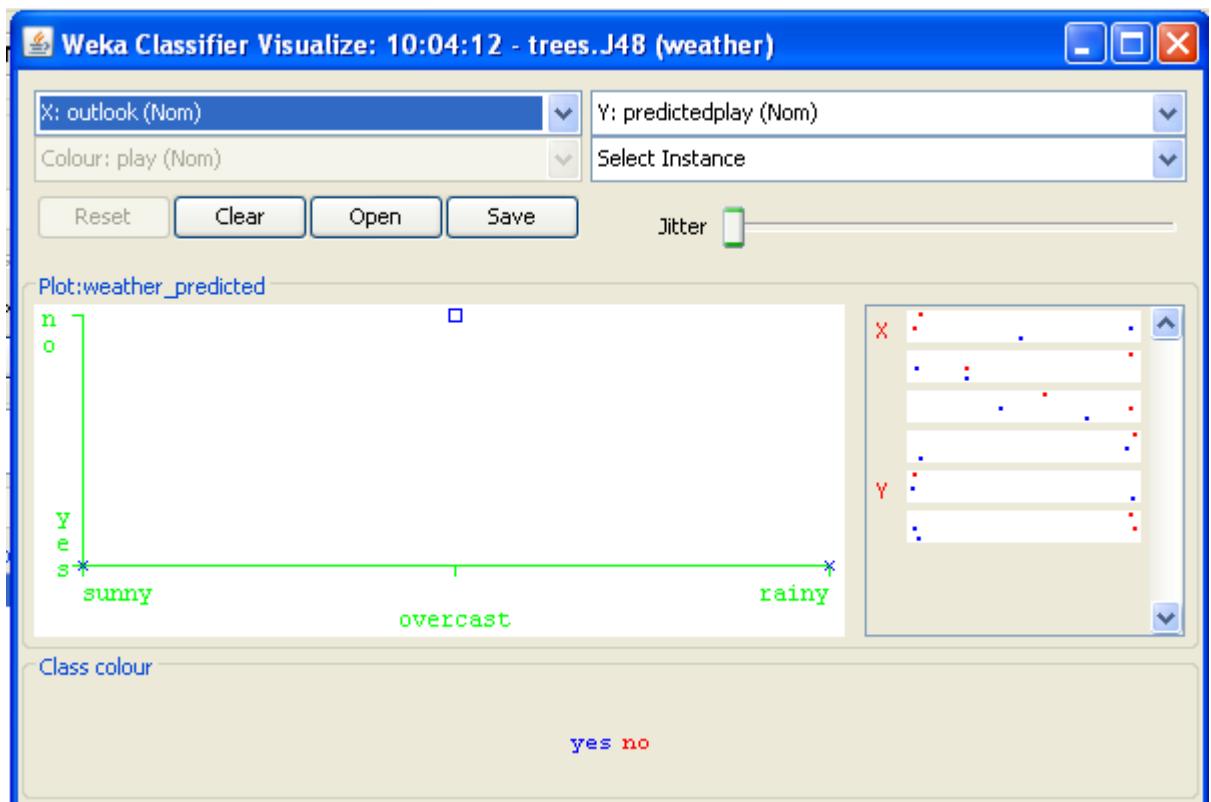


To visualize classification errors. Right-click on the entry in ‘Result list’ again

and select ‘Visualize classifier errors’ from the menu:



‘Visualize’ window displaying graph appears on the screen.

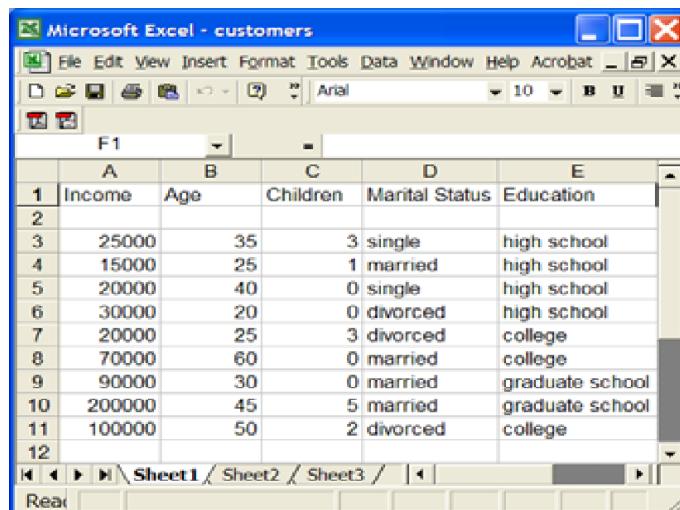


Practical N0 - 13

Aim : Clustering Analysis.

Solution :

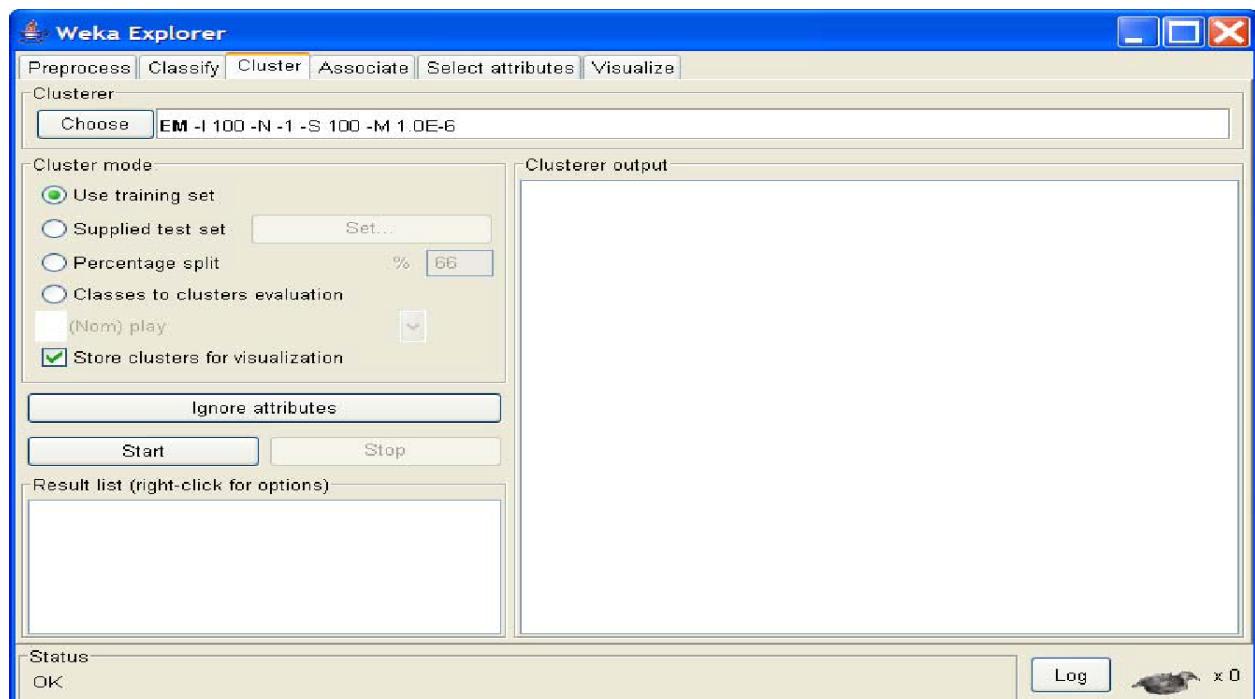
we will use customer data [6] that is contained in “customers.arff” file and analyze it with k-means clustering scheme.



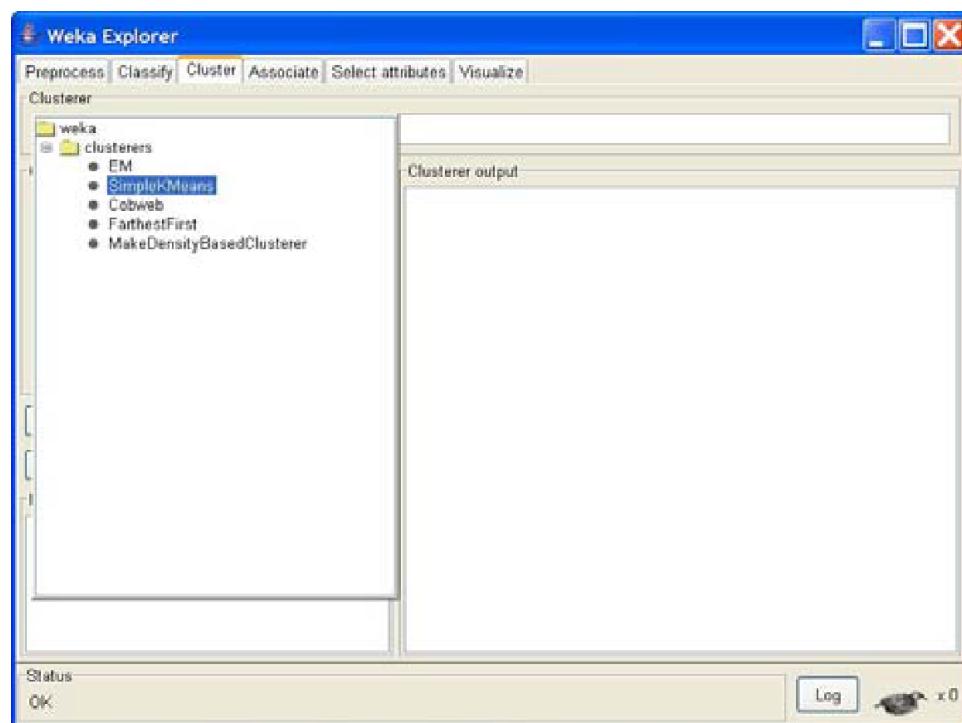
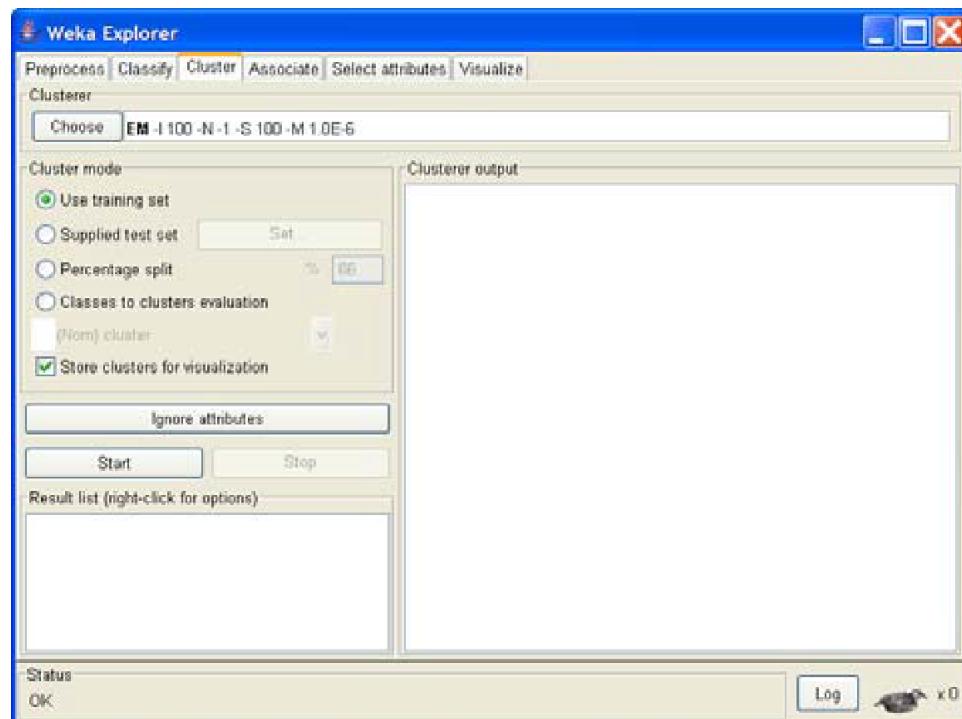
The screenshot shows a Microsoft Excel spreadsheet titled "Microsoft Excel - customers". The data is organized into columns labeled A through E. Column A is "Income", column B is "Age", column C is "Children", column D is "Marital Status", and column E is "Education". The data consists of 11 rows of customer information, starting from row 2. Row 1 contains the column headers.

	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					

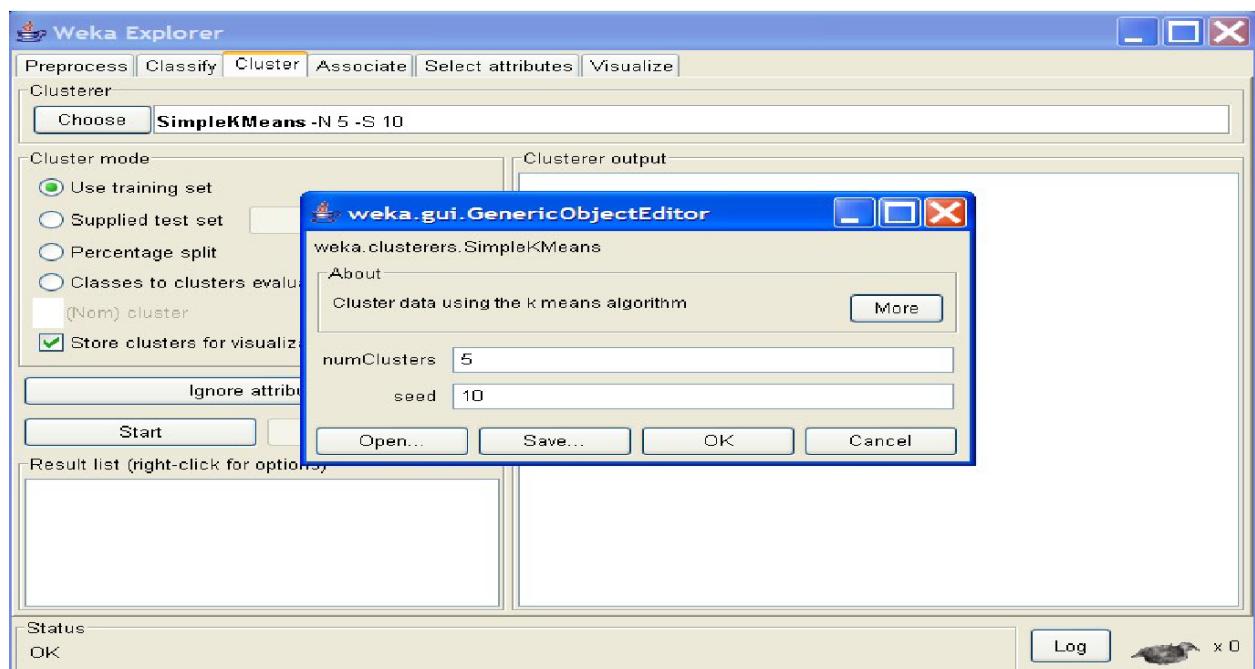
In ‘Preprocess’ window click on ‘Open file...’ button and select “customers.arff” file. Click ‘Cluster’ tab at the top of WEKA Explorer window.



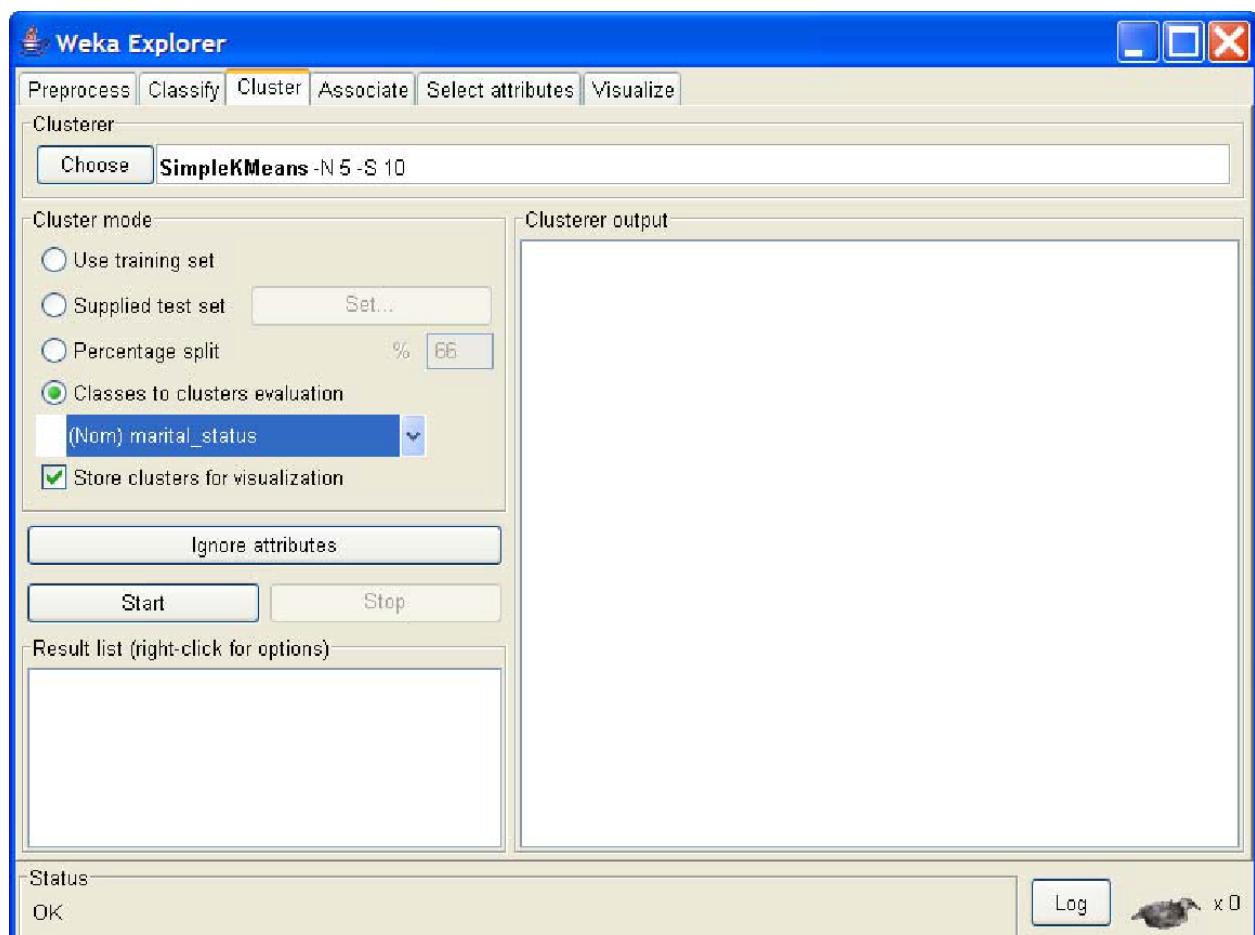
In the ‘Clusterer’ box click on ‘Choose’ button. In pull-down menu select WEKA Clusterers, and select the cluster scheme ‘SimpleKMeans’. Some implementations of K-means only allow numerical values for attributes.



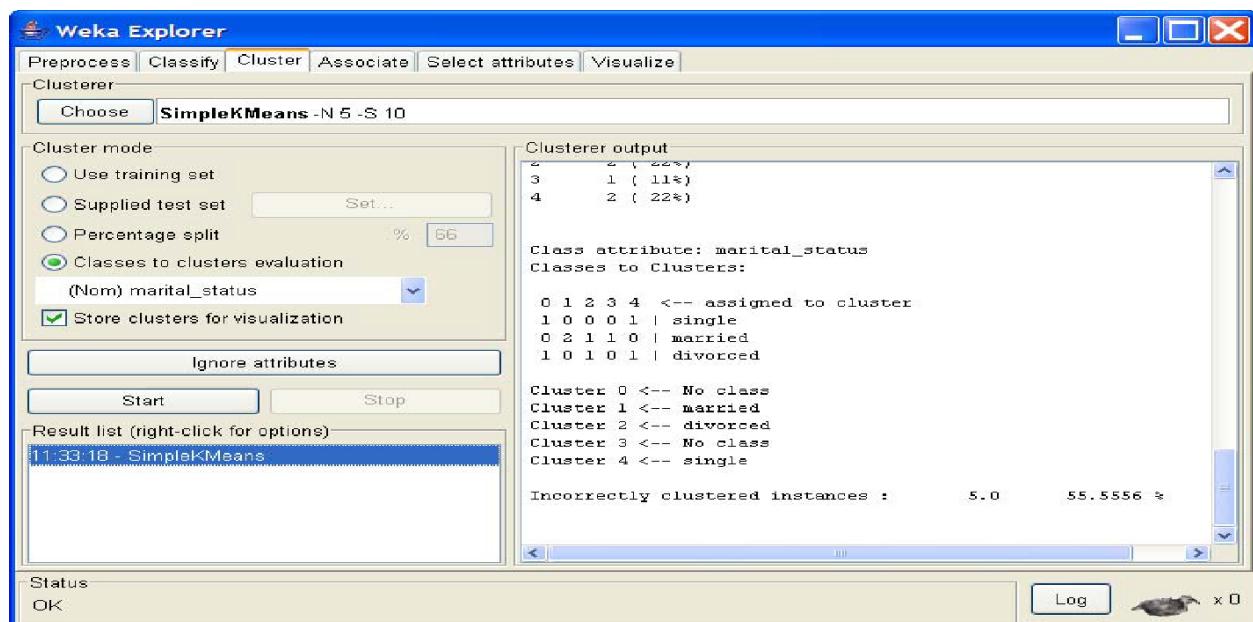
right-click on the algorithm “weak.gui.GenericObjectEditor” comes up to the screen. Set the value in “numClusters” box to 5(instead of default 2) because you have five clusters in your .arff file.



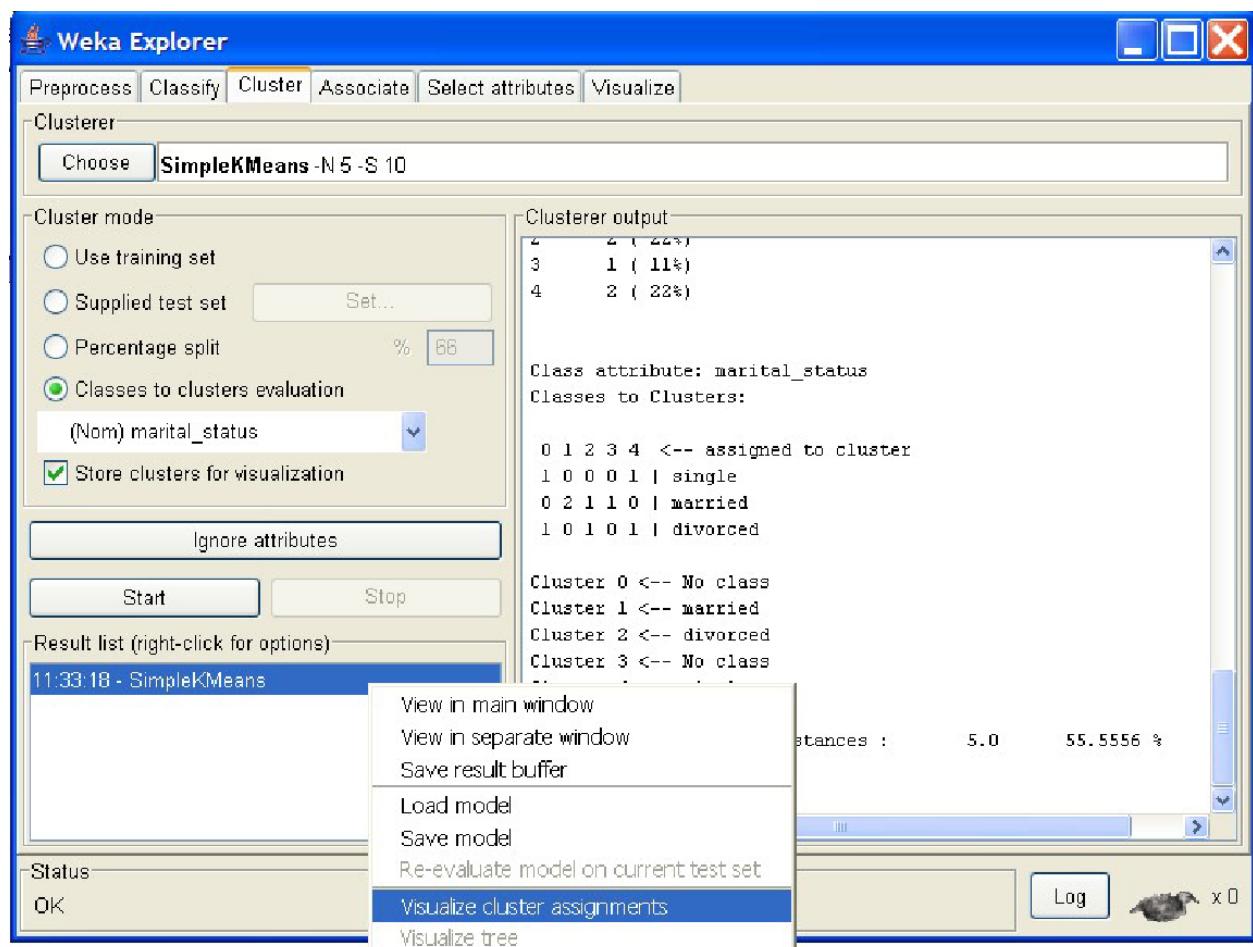
Click on ‘Classes to cluster evaluation’ radio-button in ‘Cluster mode’ box and select ‘marital_status’ in the pull-down box below.



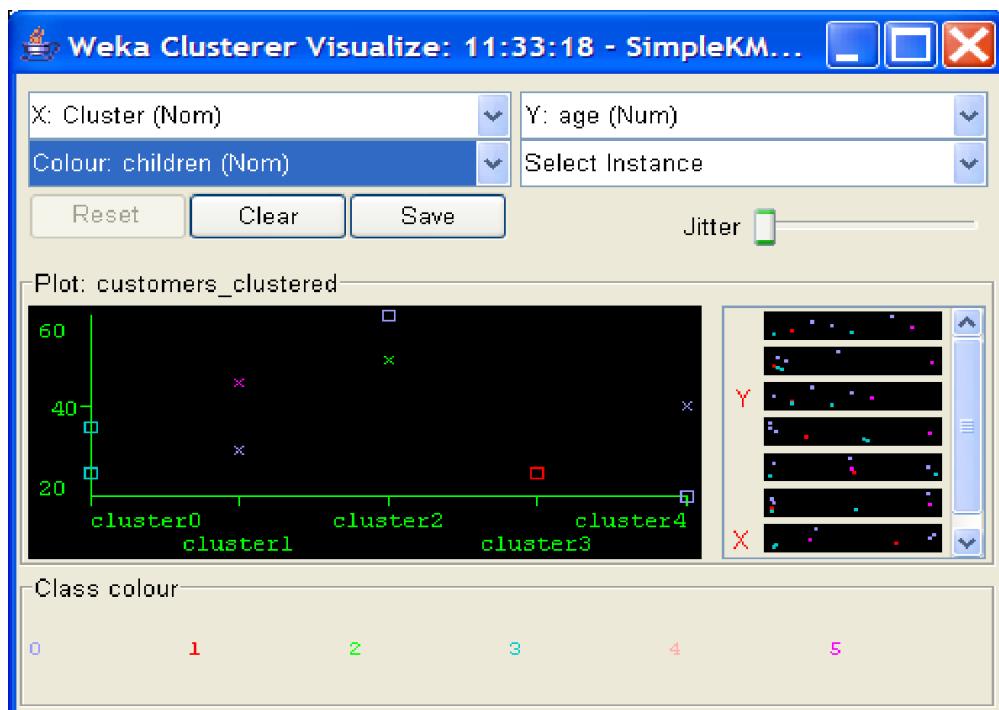
Click on the ‘Start’ button to execute the algorithm.



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

```

@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

```

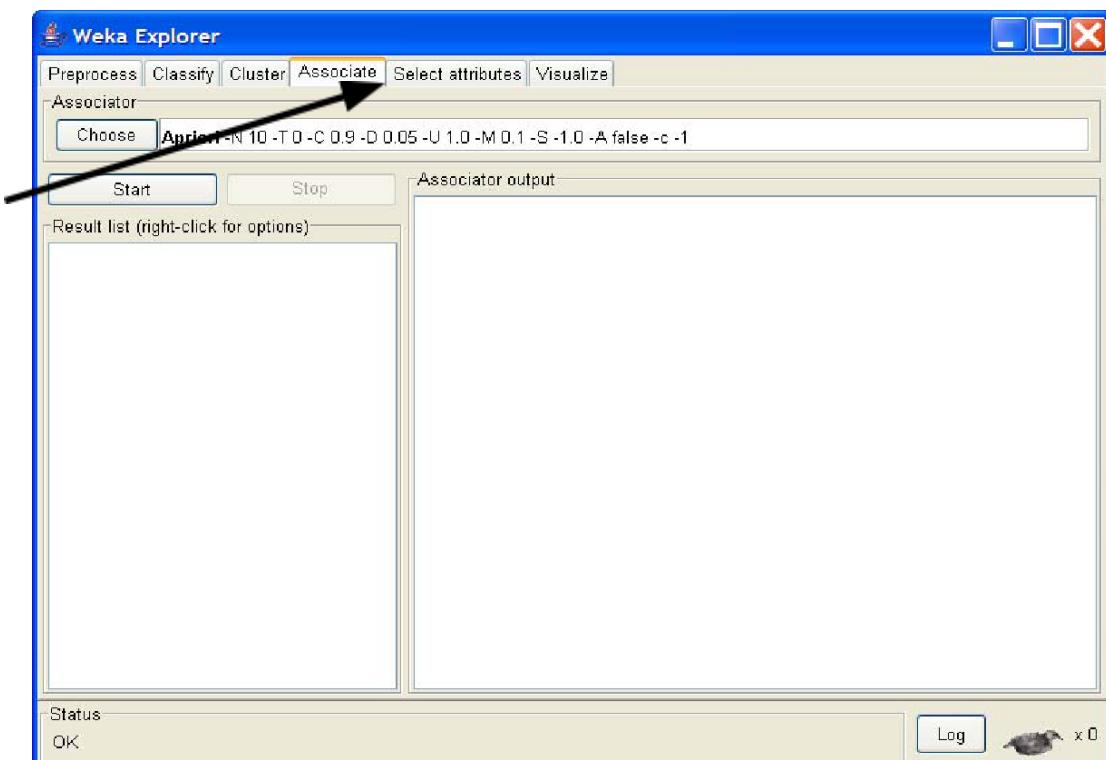
Practical No-14

Aim : Association Rule Mining.

Solution :

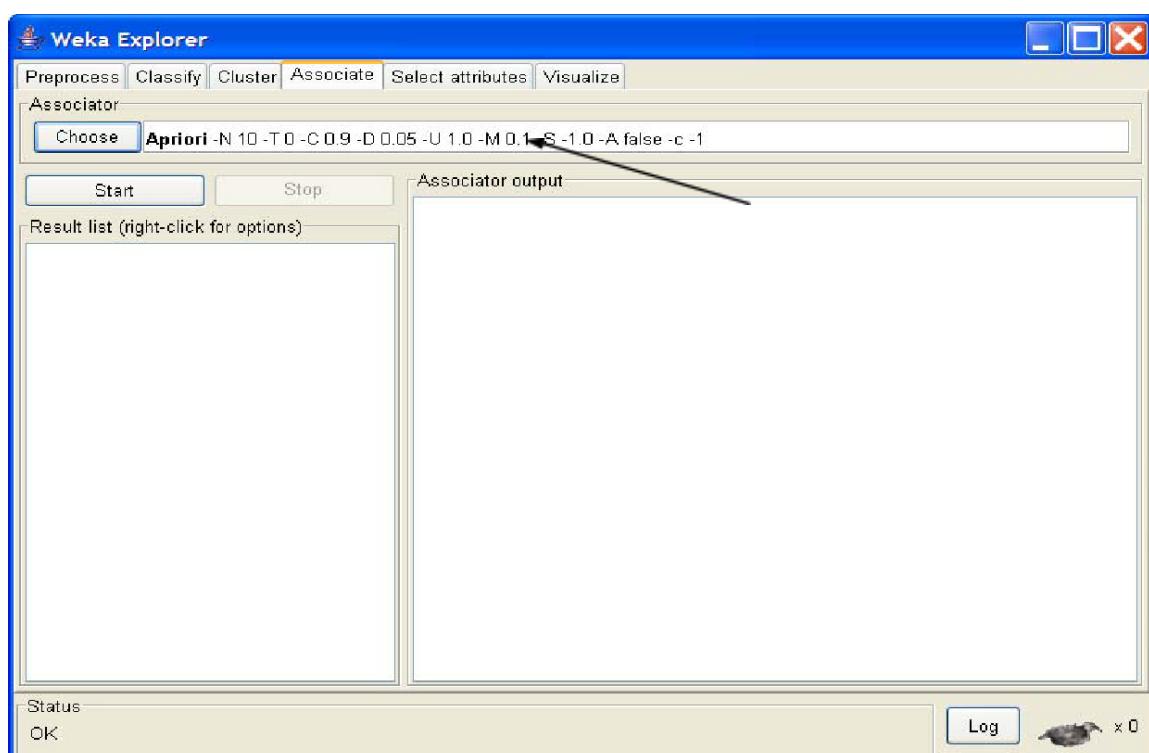
Choosing Association Scheme

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

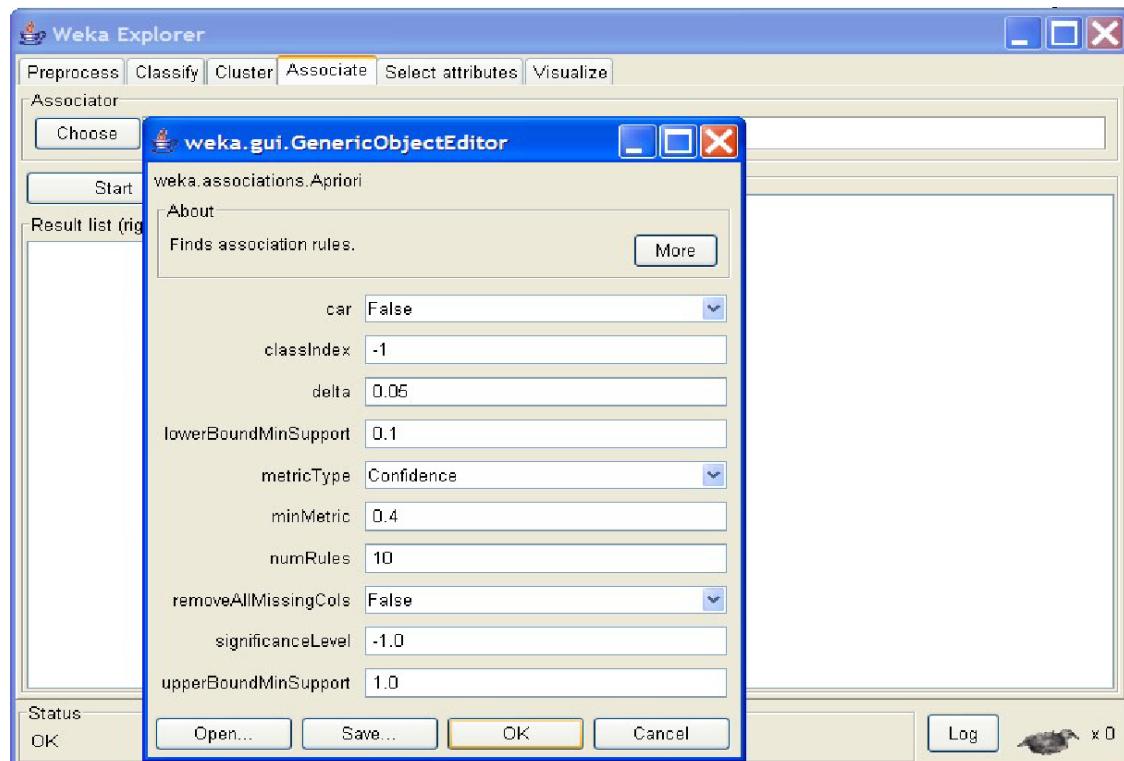


Setting Test Options

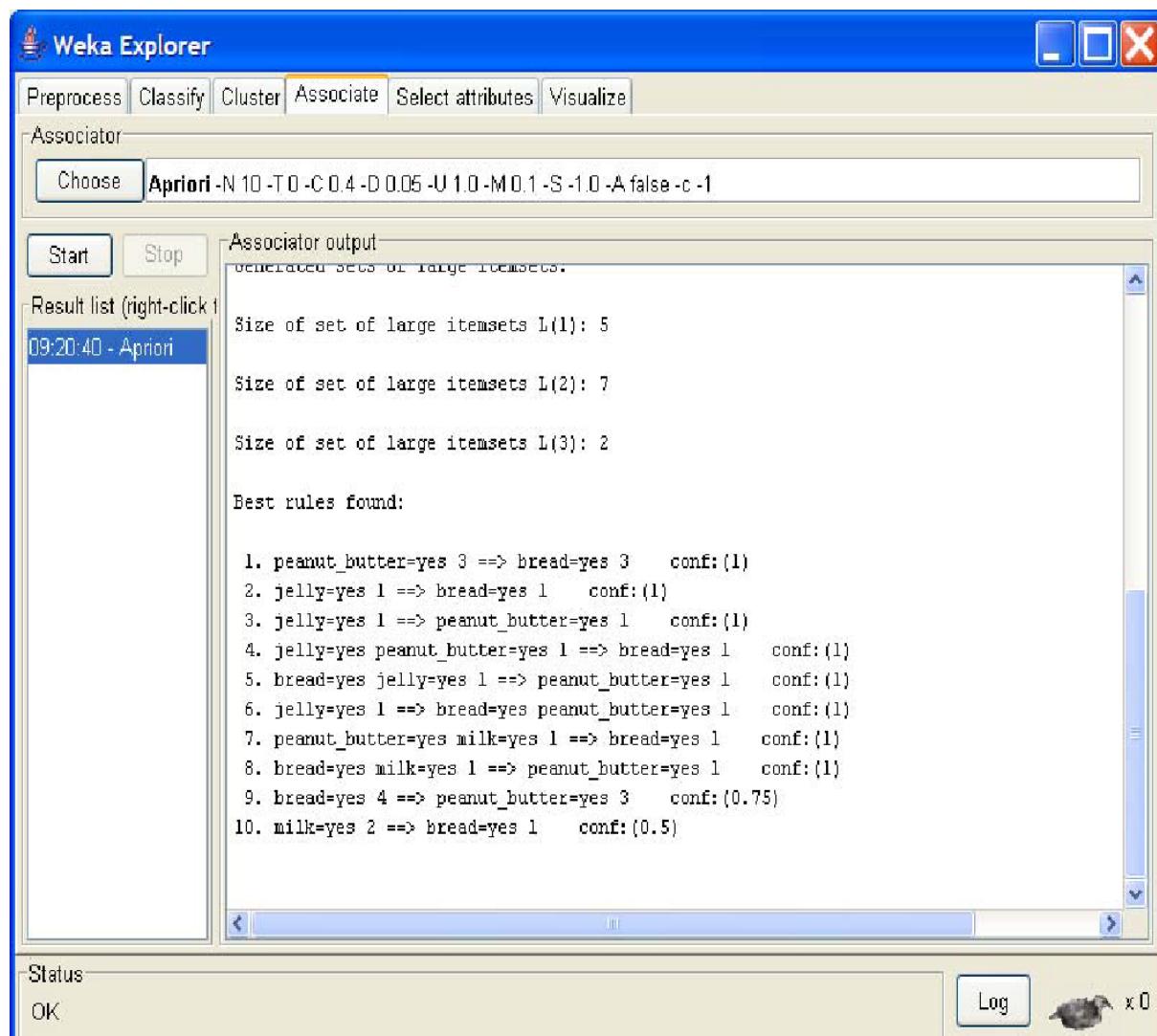
Check the text field in the ‘Associator’ box at the top of the window



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



Click on the ‘Start’ button to execute the algorithm



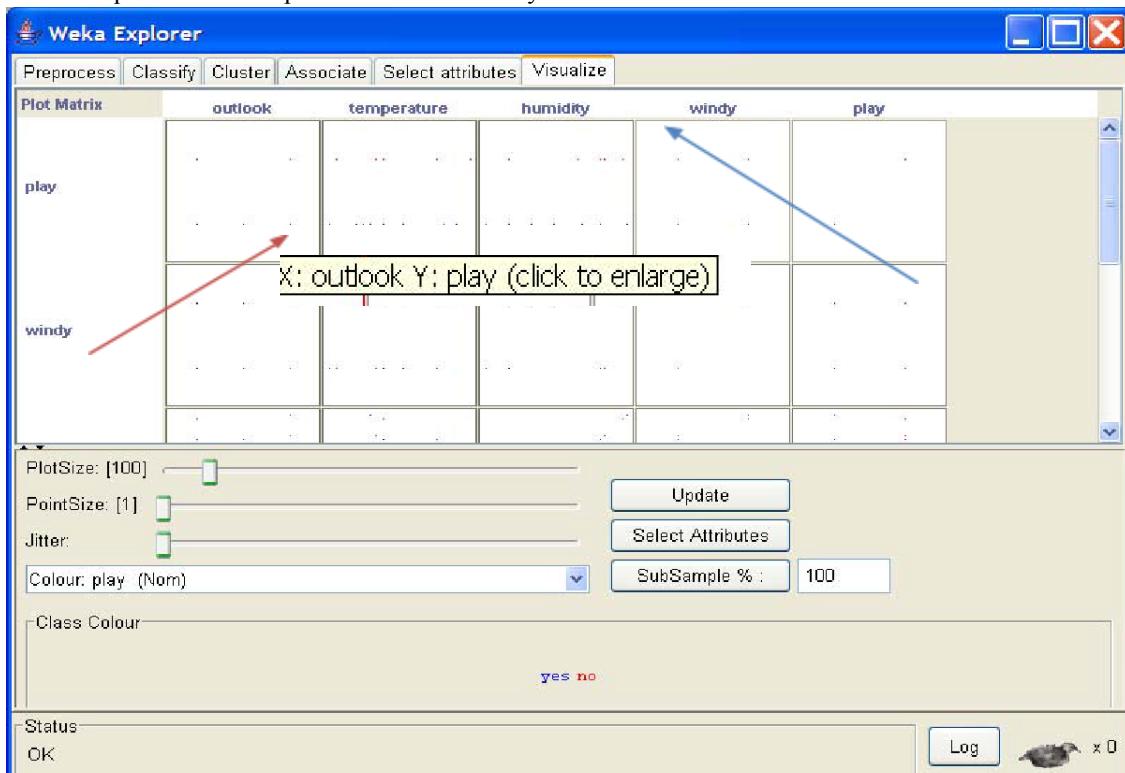
Practical No-15

Aim : Data Visualization

Solution :

To open Visualization screen, click ‘Visualize’ tab.

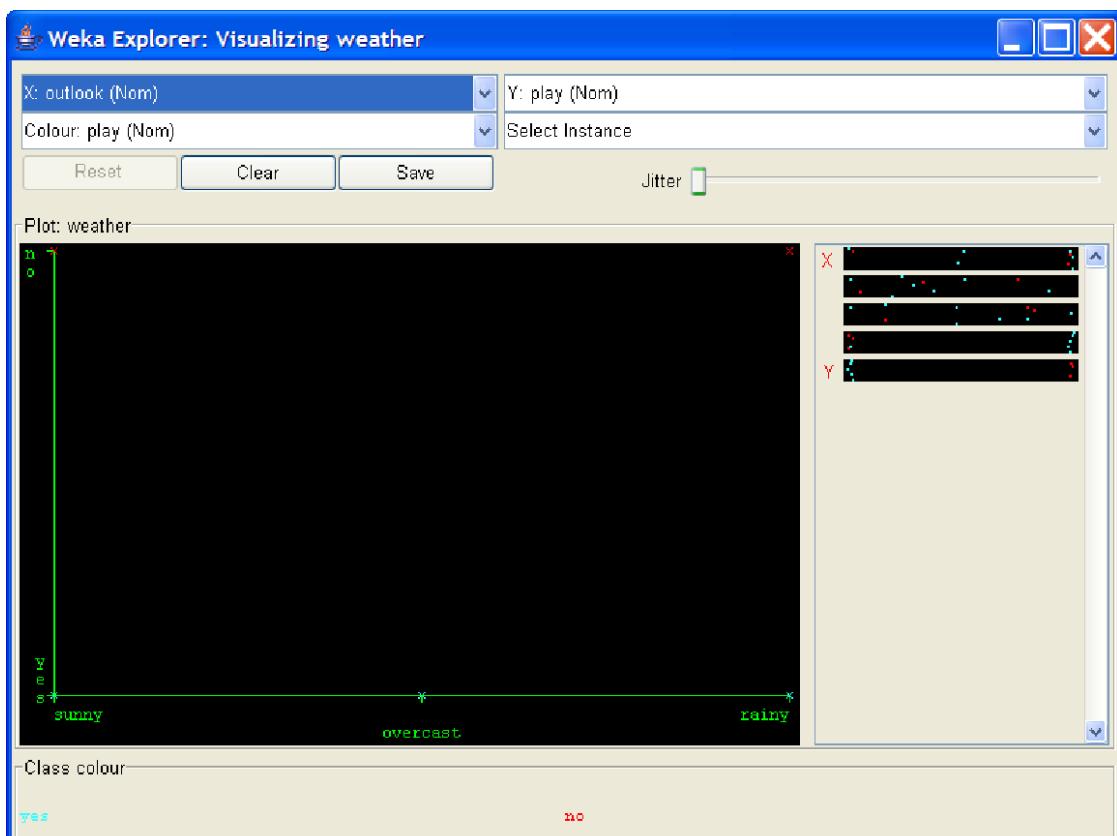
Select a square that corresponds to the attributes you would like to visualize.



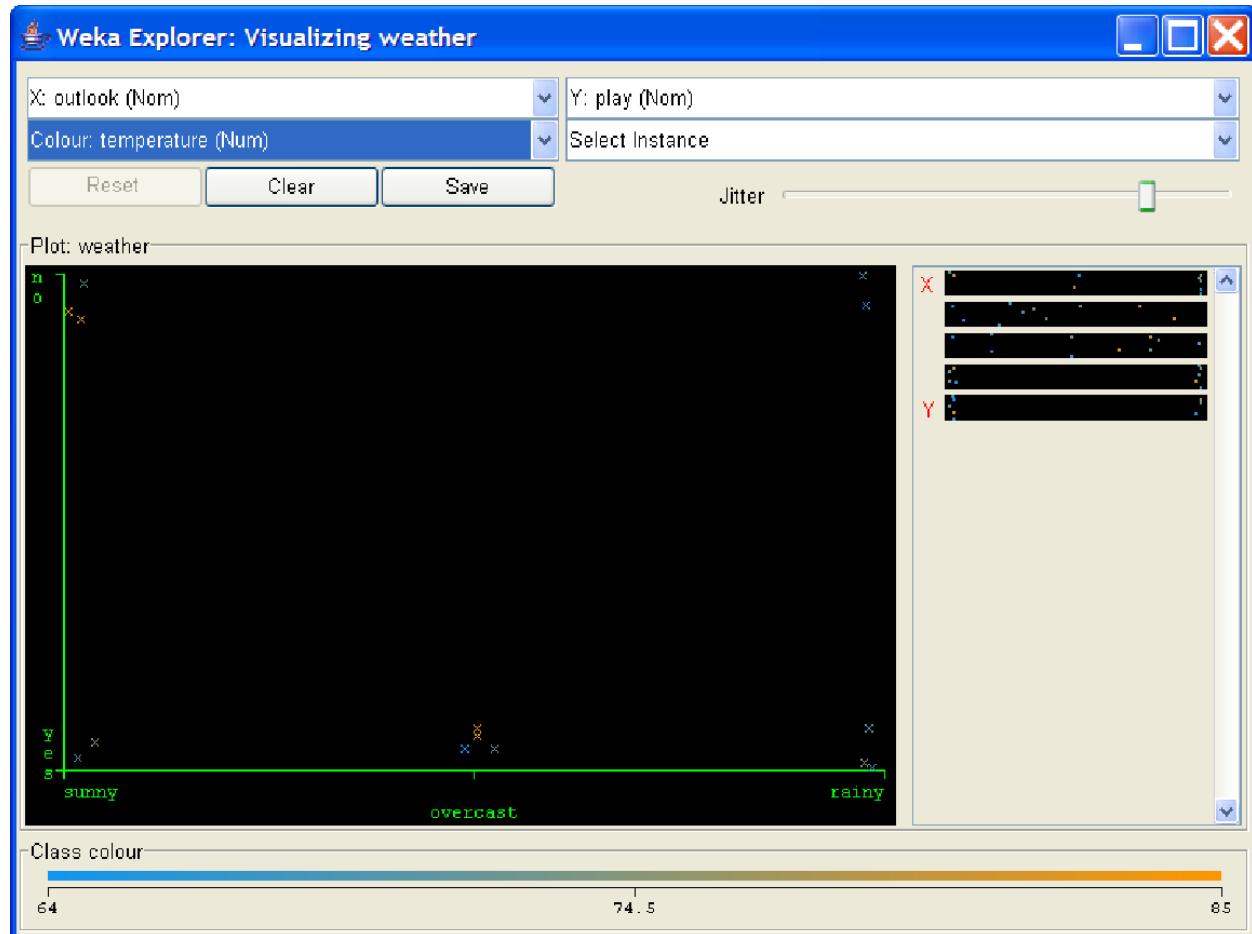
A ‘Visualizing weather’ window appears on the screen.

Changing the View

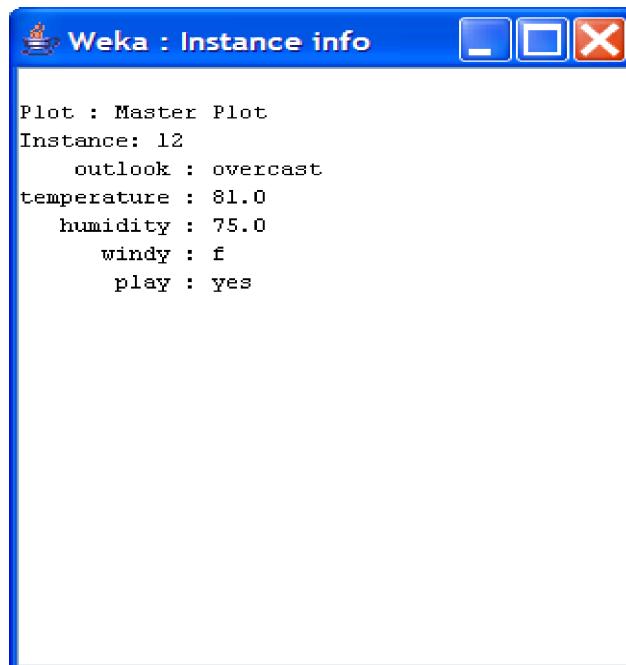
Keep sliding ‘Jitter’, a random



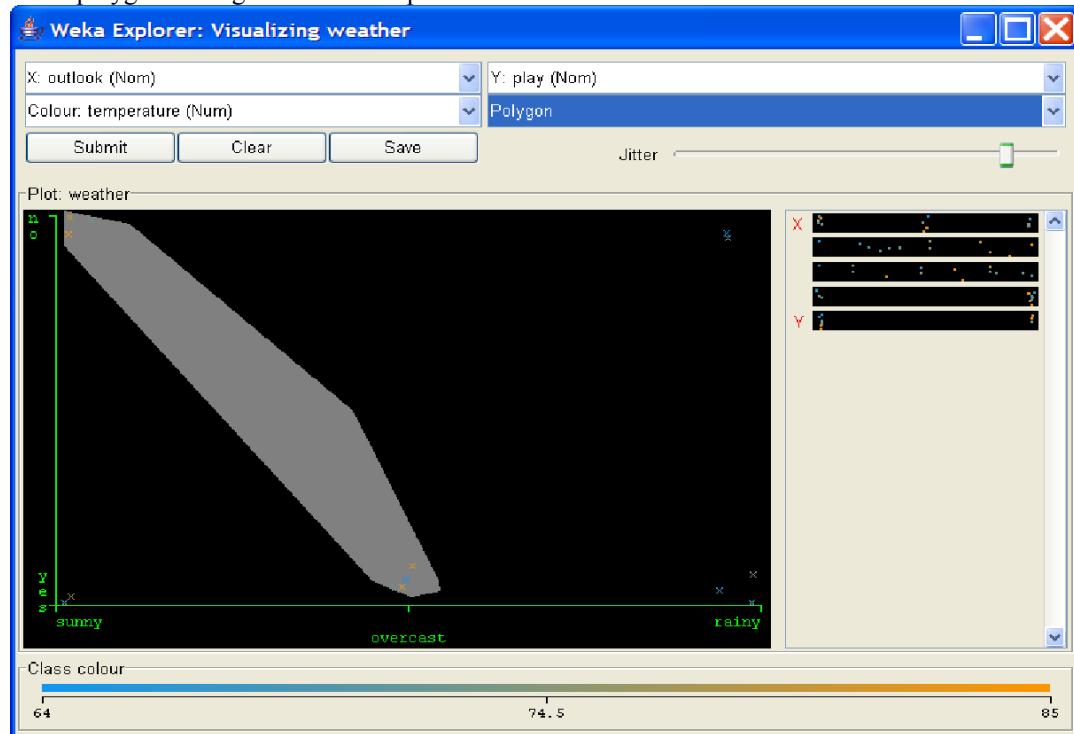
displacement given to all points in the plot, to the right, until you can spot concentration points



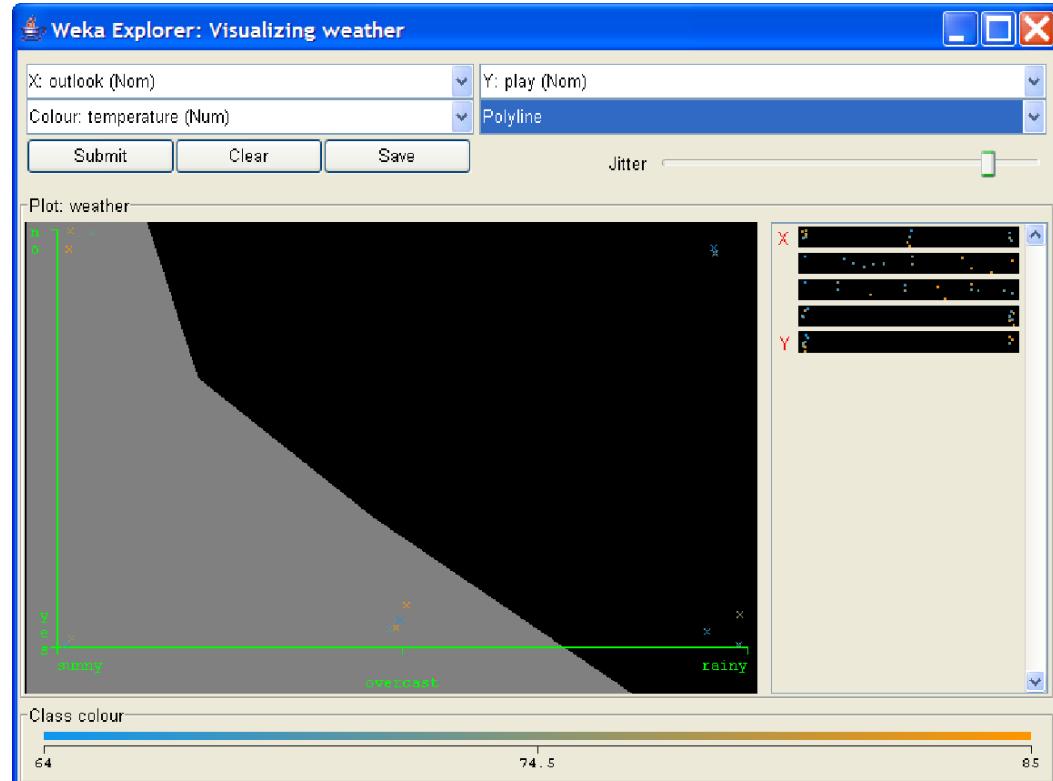
Selecting Instances: Click on an individual data point



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

