

**NAME :- Soham Landge**

**Class : MSC I Sem : I**

**Subject : Data Warehousing & Data Mining(BI) Paper : IV**

**Academic Year : 2022-23 Roll No : 545**

### **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**

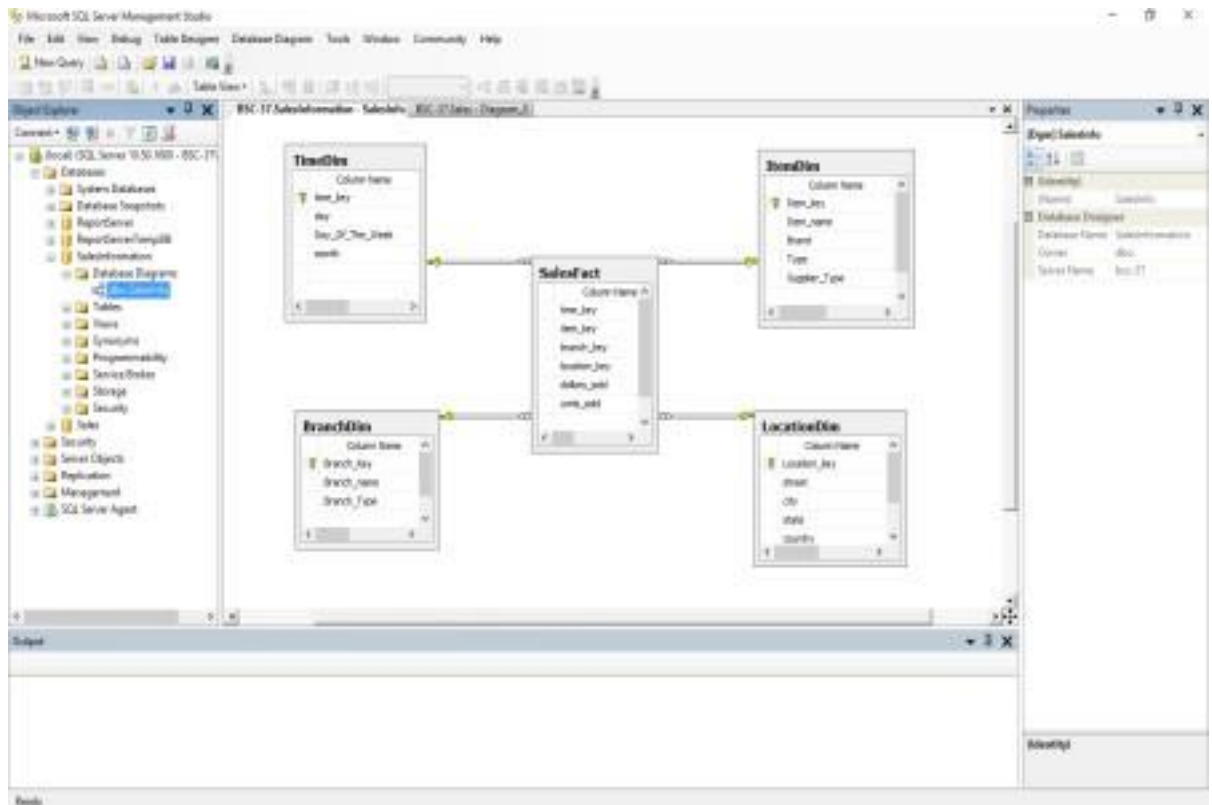
**Expand the "SalesInformation" database folder.**

- 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 :**

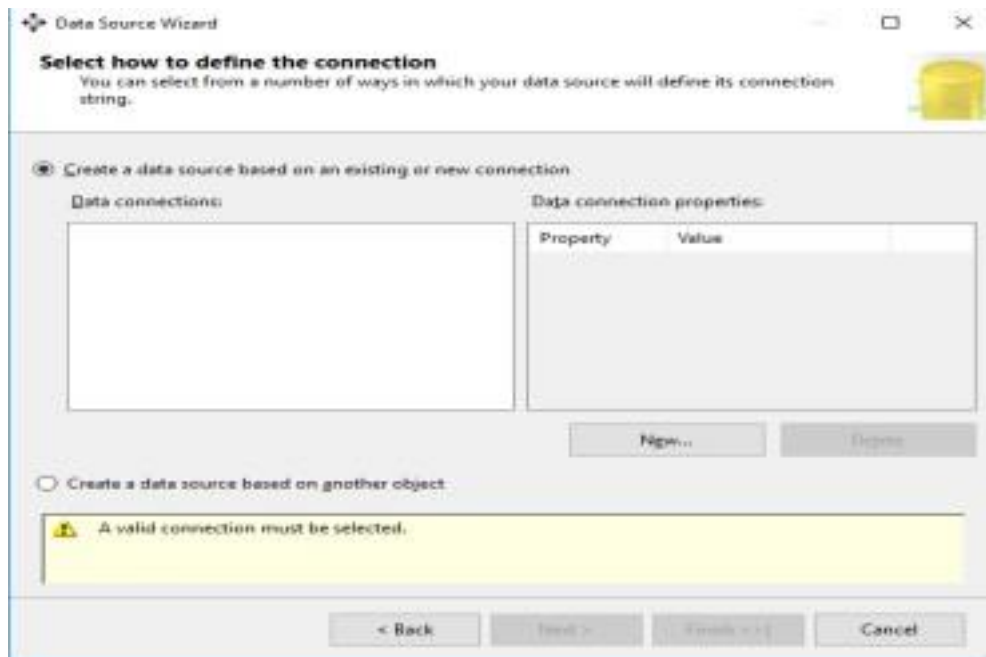
**Open Application -> Microsoft SQL Server 2008 R2 -> SQL Server Business Intelligence Development Studio**

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



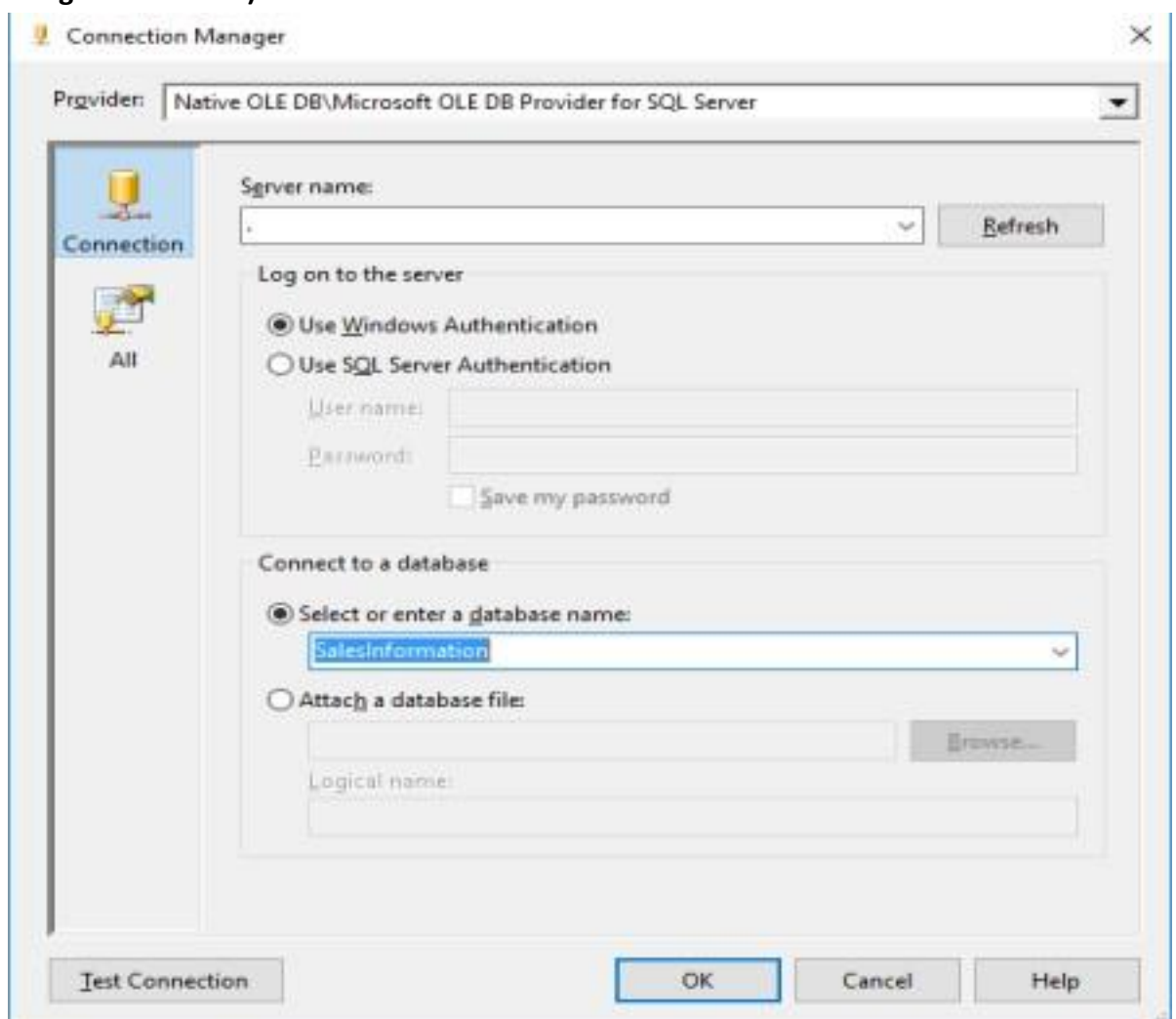
Click on Next.



The Data Source Wizard dialog box is shown. It has a title bar with a plus icon, the text "Data Source Wizard", and standard window controls. The main content area is titled "Select how to define the connection" with a subtitle "You can select from a number of ways in which your data source will define its connection string." There are two radio buttons: the first is selected and labeled "Create a data source based on an existing or new connection"; the second is labeled "Create a data source based on another object". Below the first radio button, there is a "Data connections:" list box (currently empty) and a "Data connection properties:" table with columns "Property" and "Value". Below these are "New..." and "Remove" buttons. Below the second radio button, there is a yellow message box with a warning icon and the text "A valid connection must be selected." At the bottom are "< Back", "Next >", "Finish >>", and "Cancel" buttons.

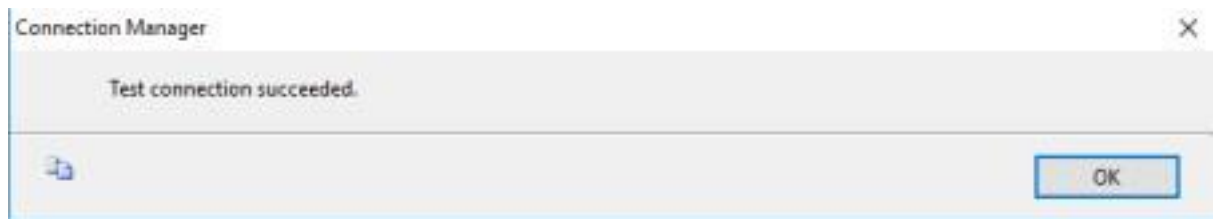
Click on New.

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

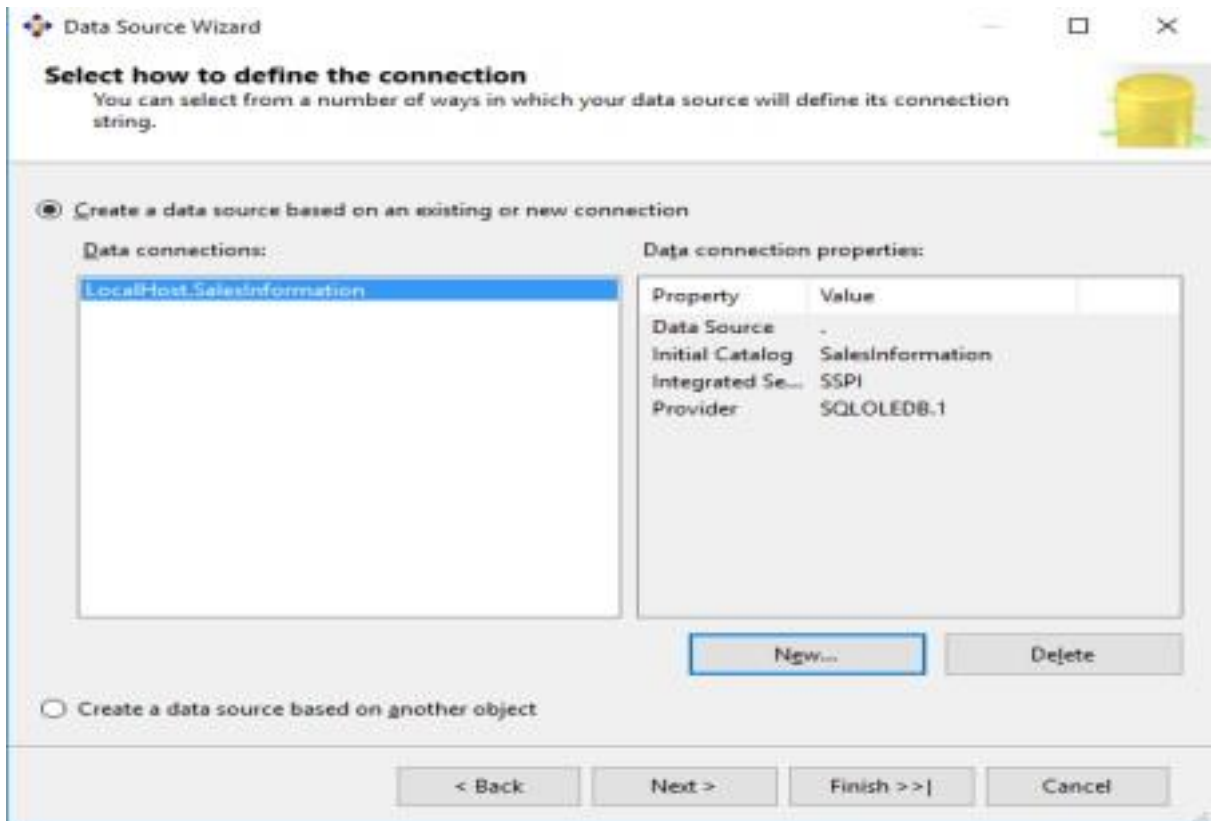


The Connection Manager dialog box is shown. It has a title bar with a plus icon, the text "Connection Manager", and standard window controls. The "Provider:" dropdown is set to "Native OLE DB\Microsoft OLE DB Provider for SQL Server". On the left is a tree view with "Connection" (selected) and "All". The "Server name:" dropdown is set to "." with a "Refresh" button. Under "Log on to the server", "Use Windows Authentication" is selected. Below are "User name:" and "Password:" text boxes, and a "Save my password" checkbox. Under "Connect to a database", "Select or enter a database name:" is selected, and the dropdown is set to "SalesInformation". Below are "Attach a database file:" (disabled), a "Browse..." button, and a "Logical name:" text box. At the bottom are "Test Connection", "OK", "Cancel", and "Help" buttons.

4. Click on Test Connection.

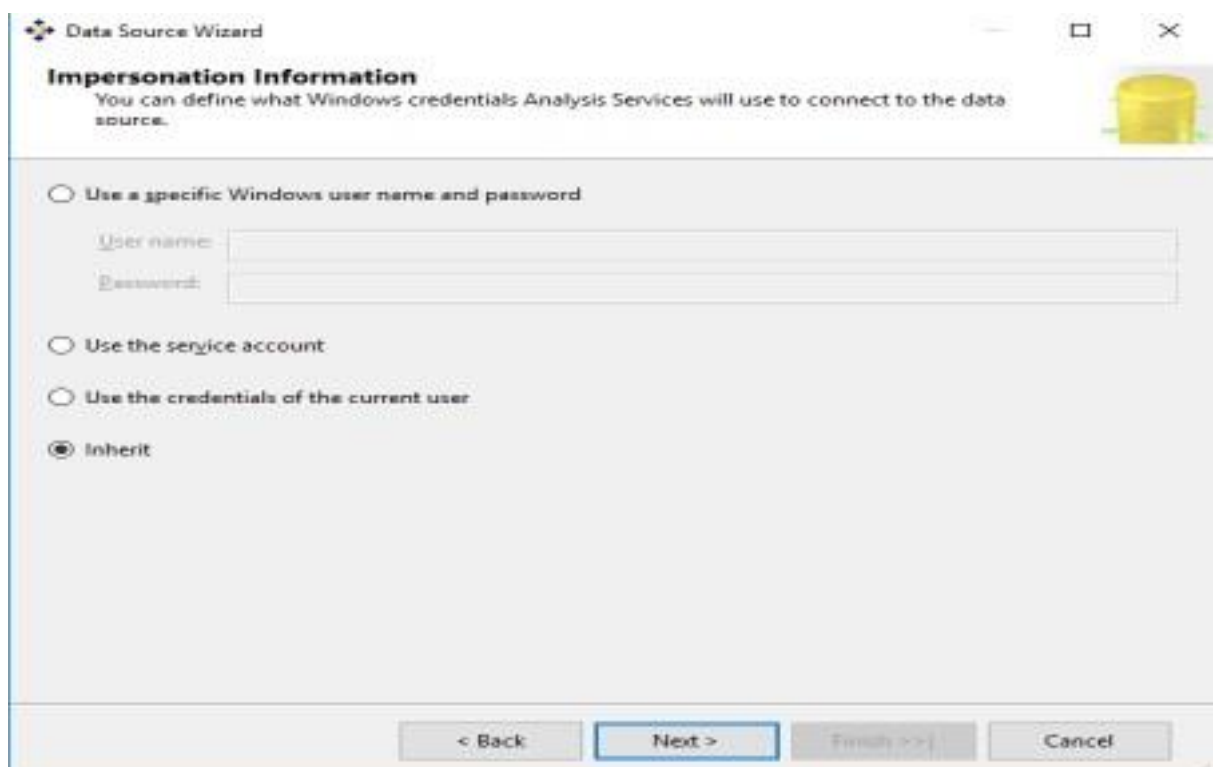


Click on OK.



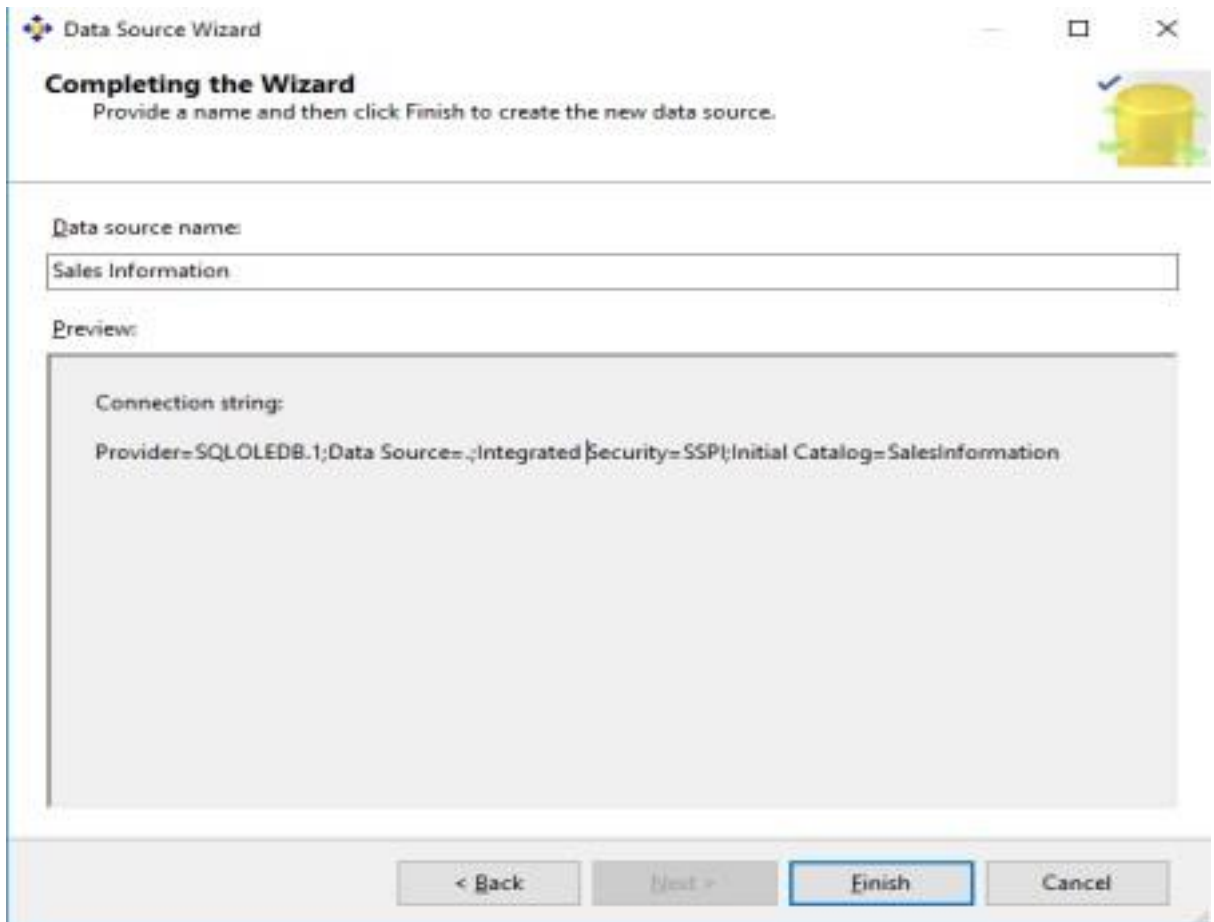
Click on Next

5. Choose "Inherit" option.



Click on Next.

6. Click on Finish.



The screenshot shows the 'Data Source Wizard' window at the 'Completing the Wizard' step. The title bar reads 'Data Source Wizard'. The main heading is 'Completing the Wizard' with a sub-instruction: 'Provide a name and then click Finish to create the new data source.' Below this, there is a text box labeled 'Data source name:' containing the text 'Sales Information'. Underneath is a 'Preview:' section showing the connection string: 'Provider=SQLOLEDB.1;Data Source=.;Integrated Security=SSPI;Initial Catalog=SalesInformation'. At the bottom, there are four buttons: '< Back', 'Next >', 'Finish' (which is highlighted with a blue border), and 'Cancel'.

Name Data Source as “Sales Information”.

### Practical No 3

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

Solution :

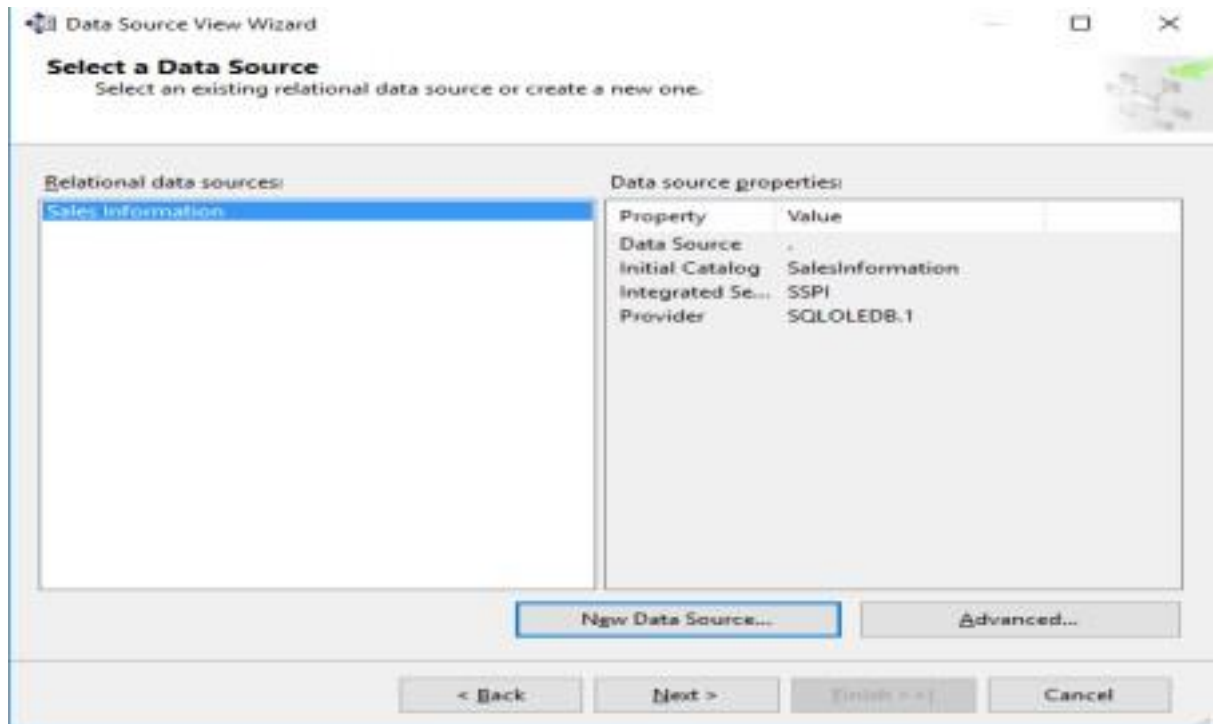
1. Right click on Data Source View -> New Data Source View



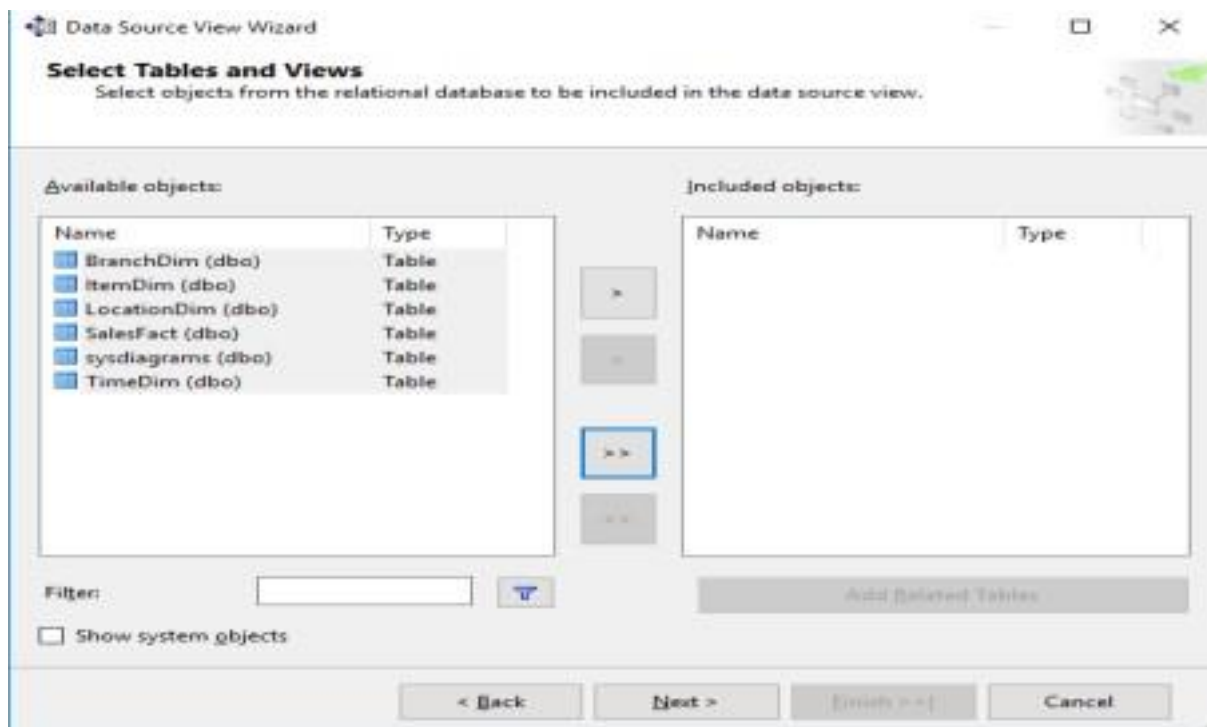
The screenshot shows the 'Data Source View Wizard' window at the 'Welcome' step. The title bar reads 'Data Source View Wizard'. The main heading is 'Welcome to the Data Source View Wizard'. Below this, there is explanatory text: 'Use this wizard to create a new data source view. You create a data source view from tables and views in a relational database. A data source provides a simple connection to a relational database. Use a data source view for more advanced features, such as caching metadata, adding relationships, creating calculations, and setting logical keys.' At the bottom left, there is a checkbox labeled 'Don't show this page again'. At the bottom, there are four buttons: '< Back', 'Next >' (highlighted with a blue border), 'Finish >>', and 'Cancel'.

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

**Data Source View Wizard**

### Completing the Wizard

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

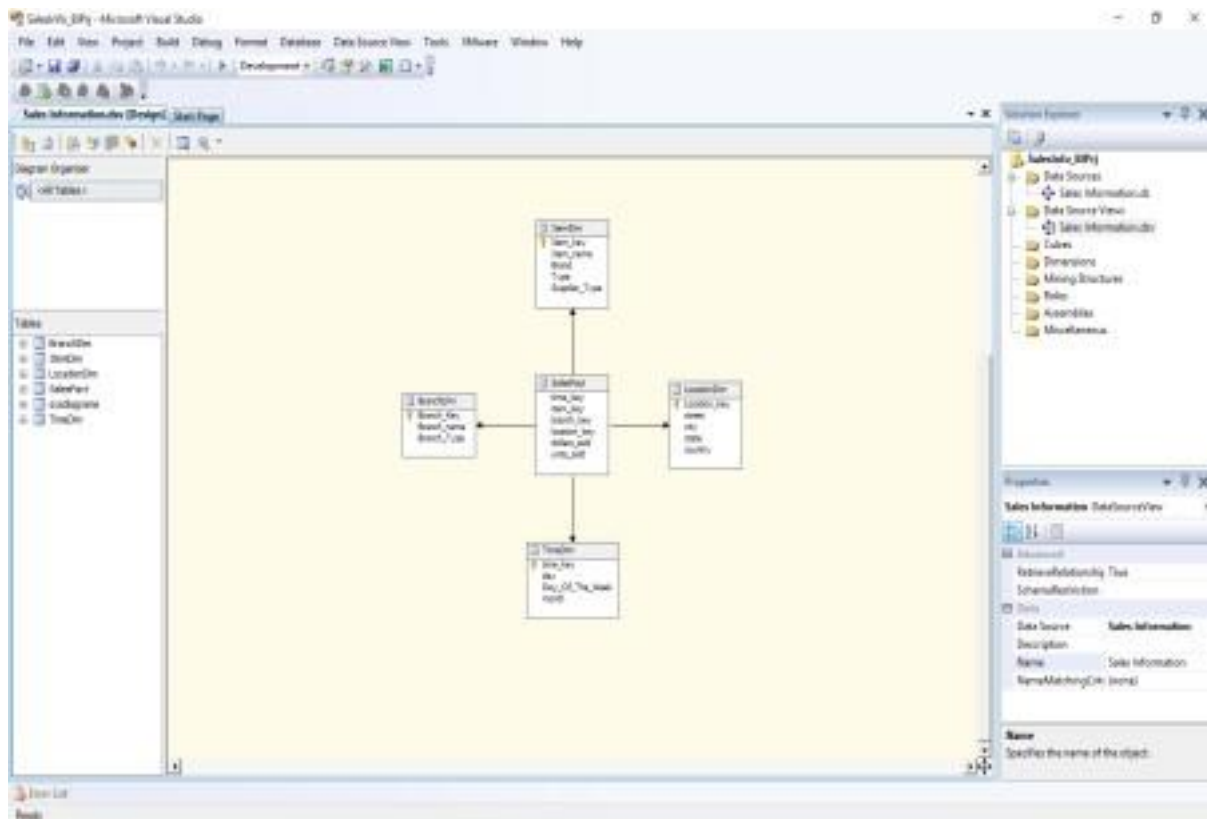
Name:

Preview:

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

Click on Finish.

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

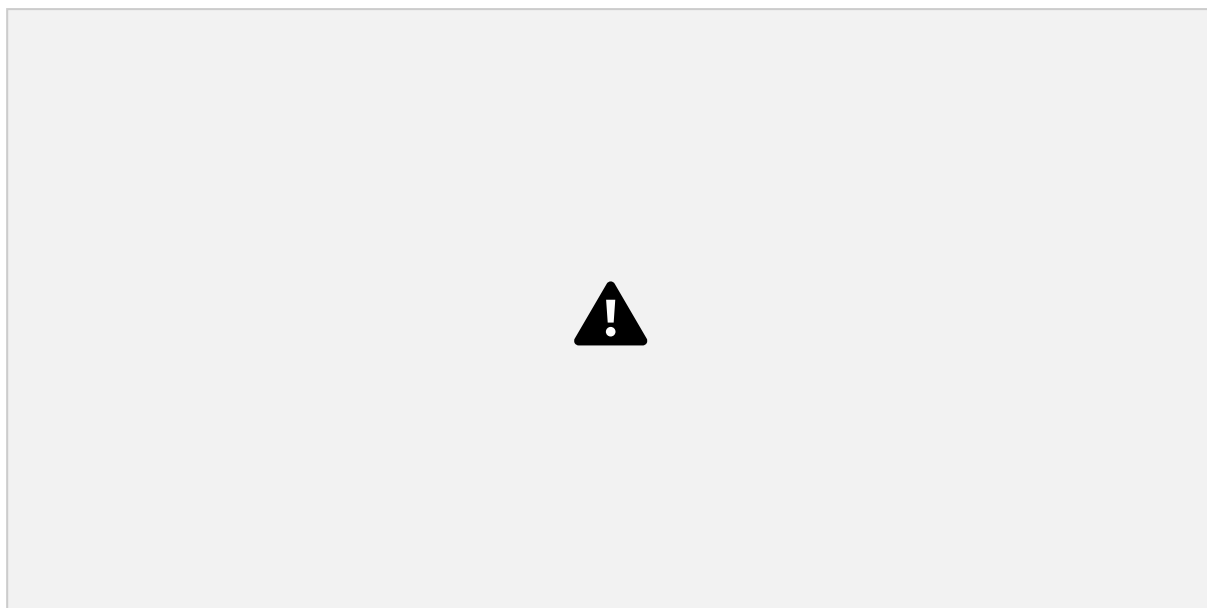


## Practical No 4

**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 ColumnArea.**



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



**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 DollarsSold”  
-> 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]`



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



**3.3.** `SELECT` [Location\_key],[street],[city],[state],[country] `FROM`  
[SalesInformation].[dbo].[LocationDim]



**3.4.** `SELECT` [time\_key],[item\_key],[branch\_key],[location\_key]  
,[dollars\_sold],[units\_sold]  
`FROM` [SalesInformation].[dbo].[SalesFact]



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



**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];`



```
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];
```



```
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];
```





### **Practical No 9**

**Aim : Calculation & KPI**

### **Practical No - 10**

**Aim : Data PreProcessing**

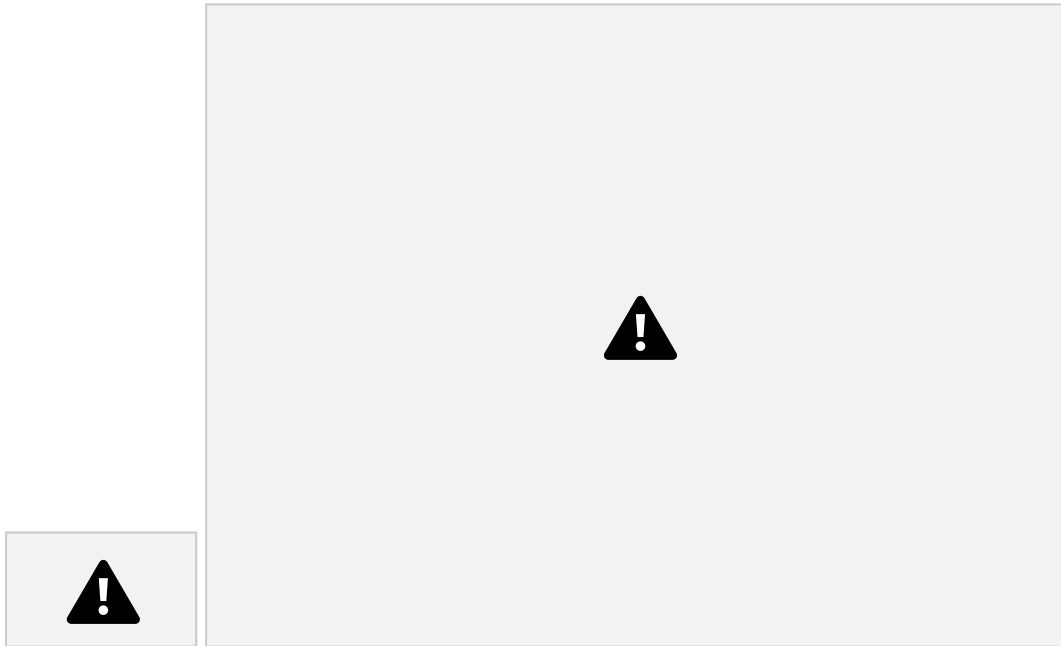
**Solution :**

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



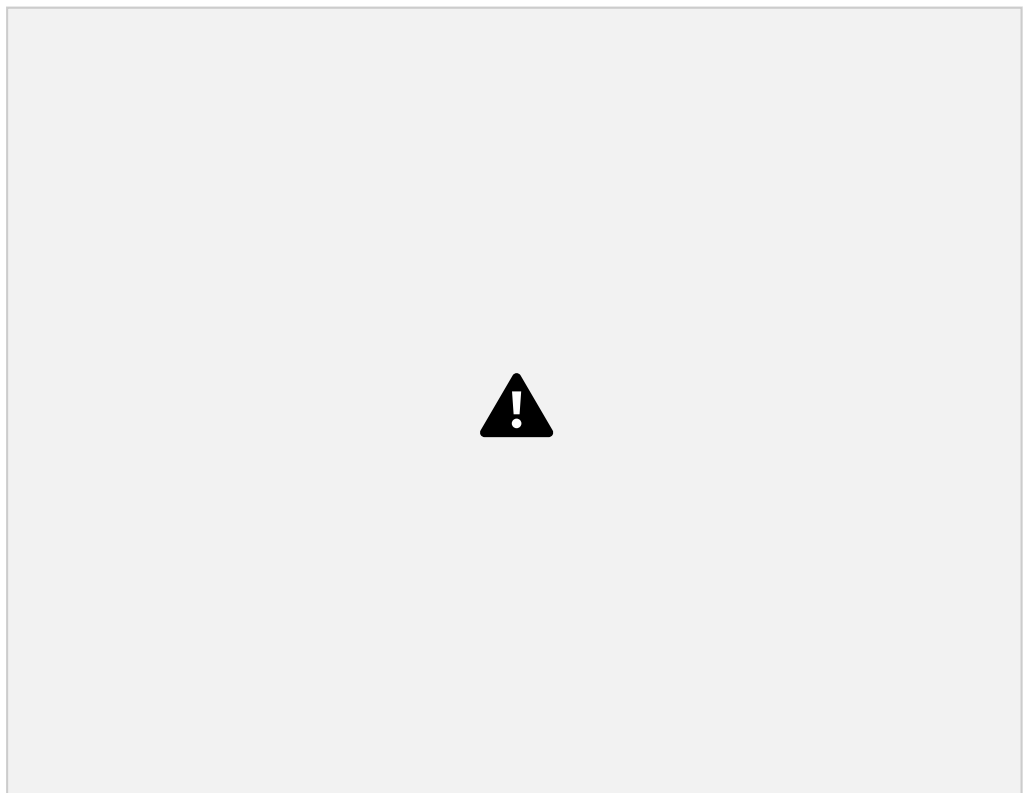
**Opening file from a local file system**

Click on 'Open file...' button

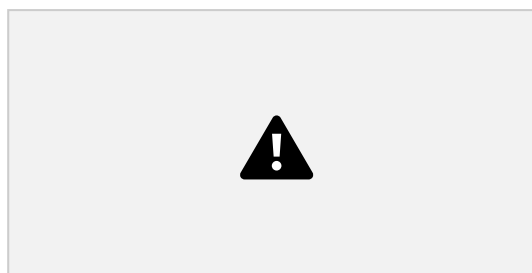


It brings up a dialog box allowing you to browse for the data file on the local file system, choose “weather.arff” file.

#### **Opening file from a web site**



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

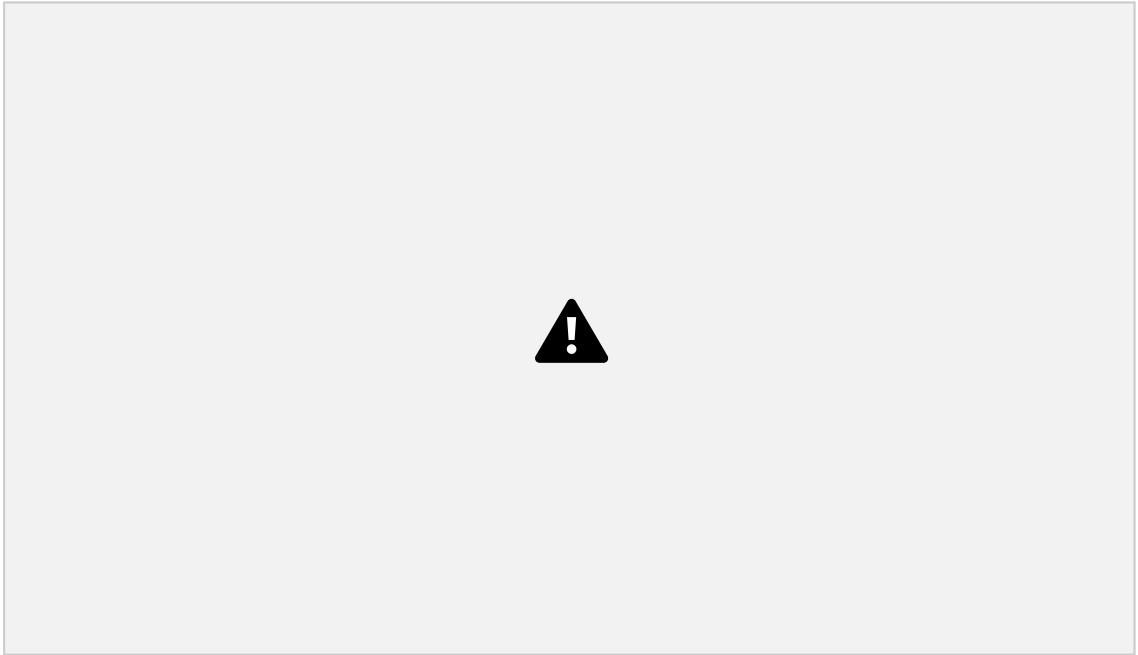


#### **Reading data from a database:**

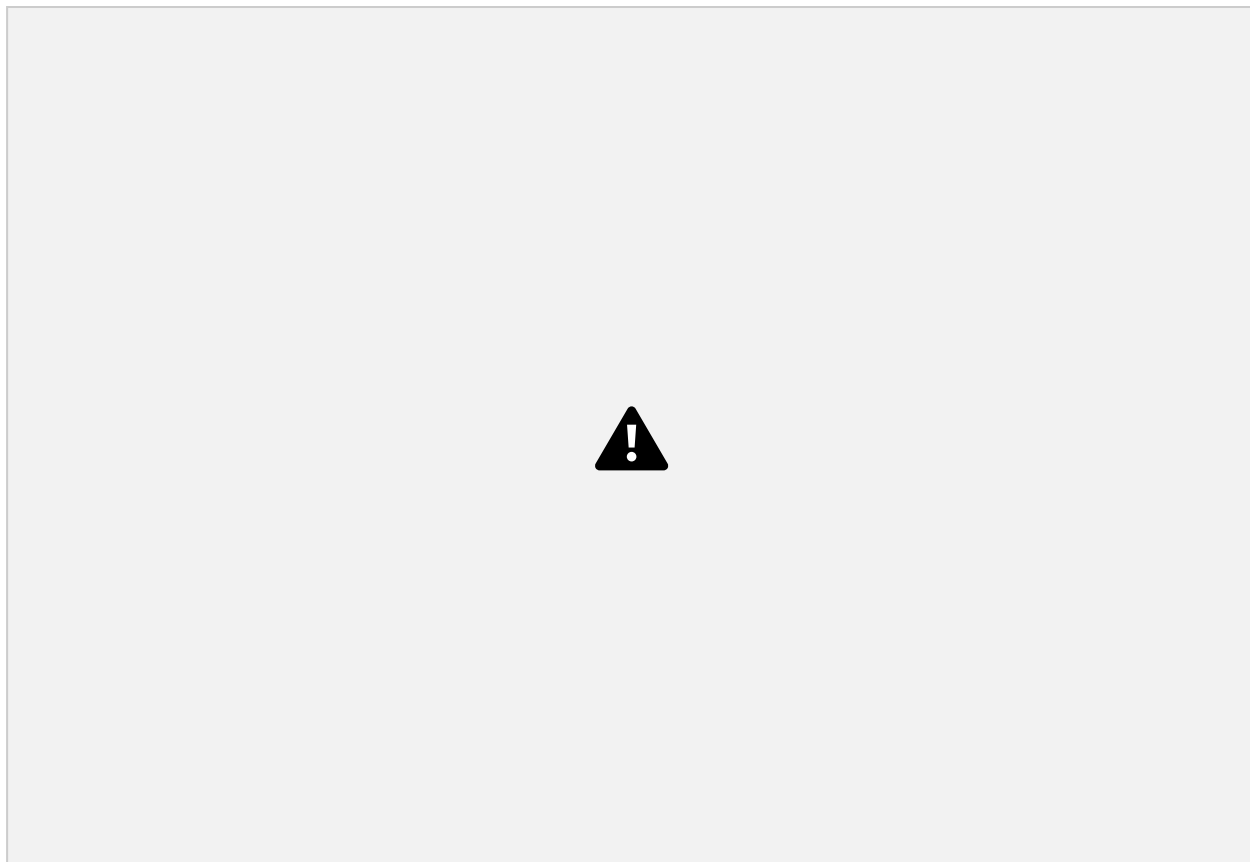


### **Loading data**

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.,

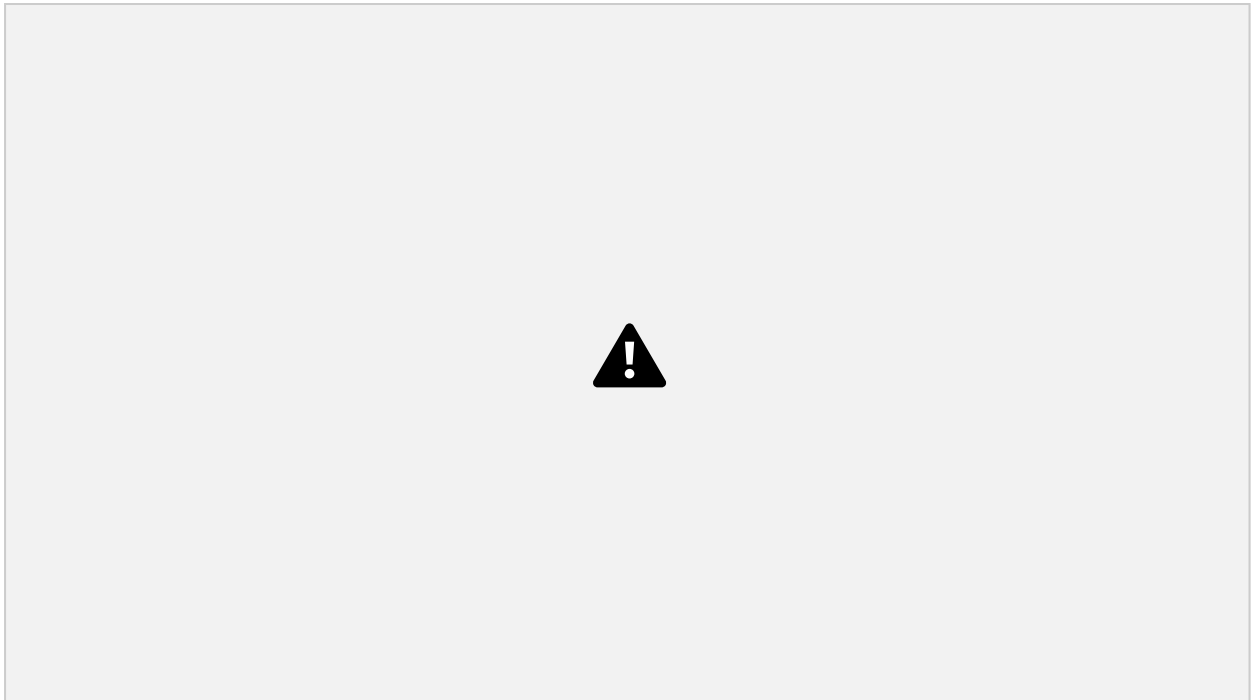
## **Practical No – 11**

**Aim : Data discretization.**

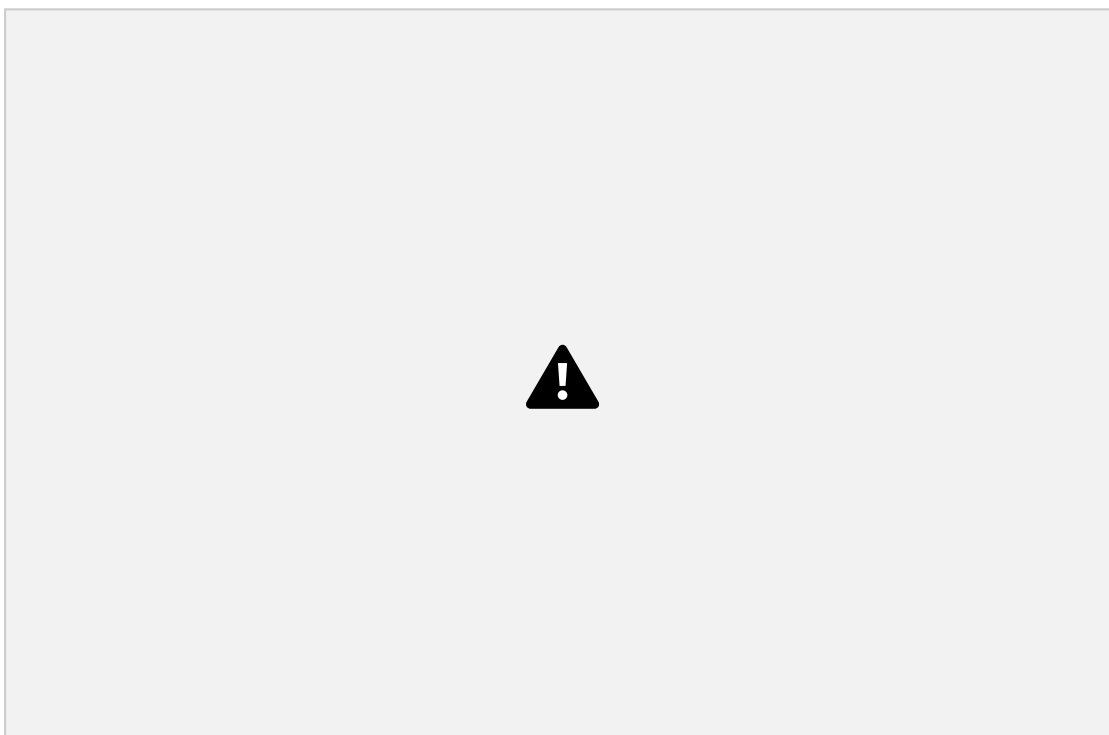
**Solution :**

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

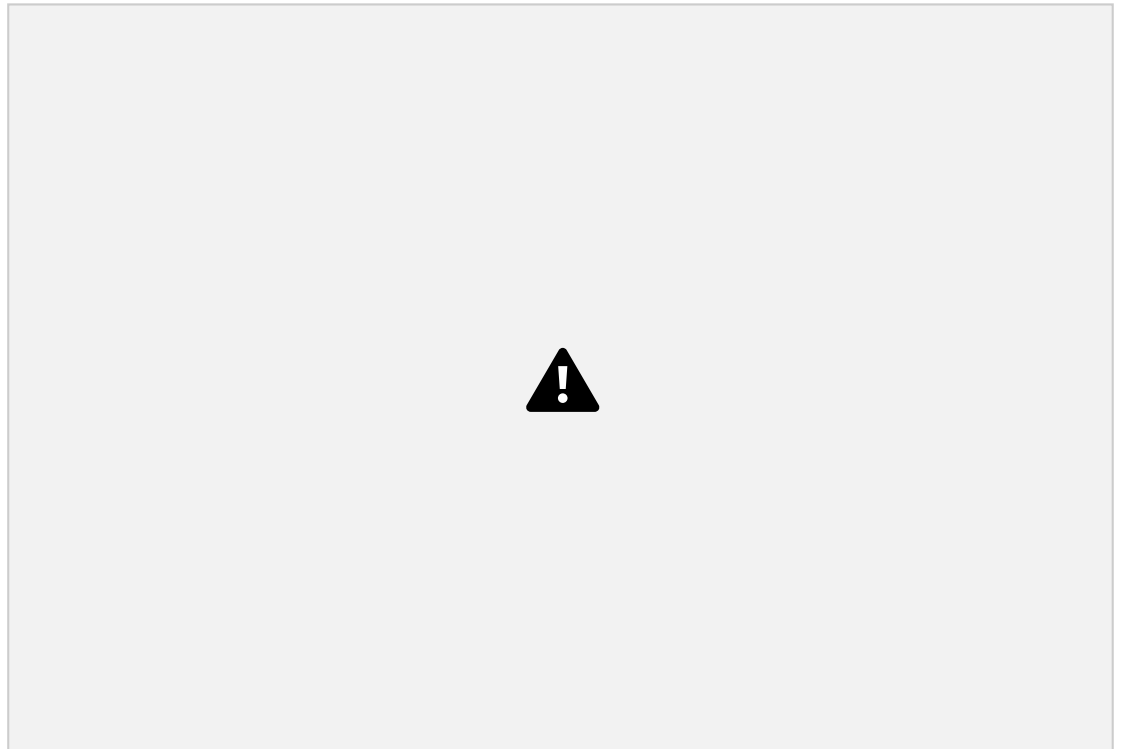
This will show pull-down menu with a list of available filters. Select Supervised Attribute Discretize and click on 'Apply' button.



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.

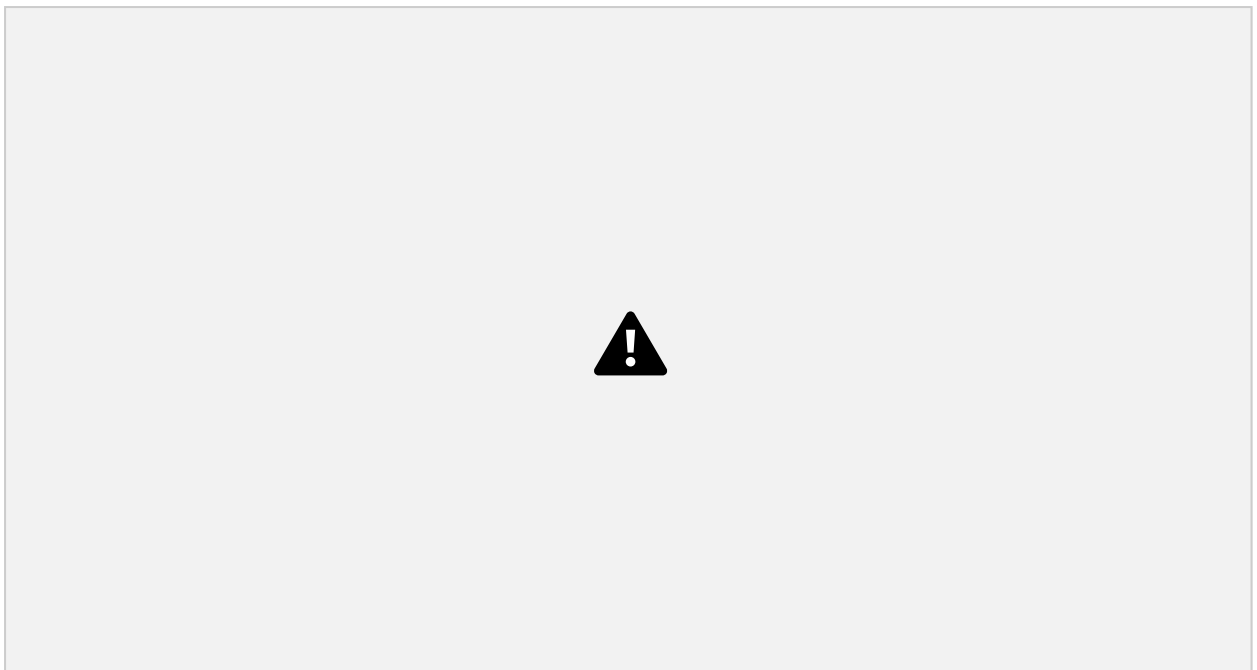


### **Practical No - 12**

**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

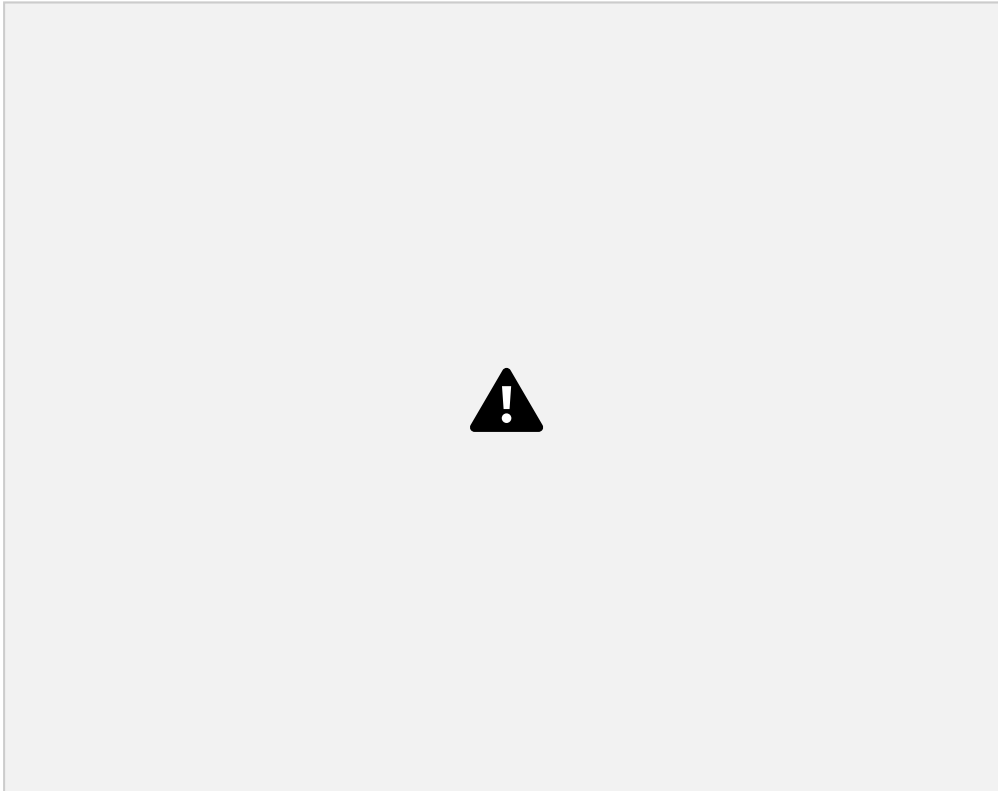


To see a graphical representation of the classification tree. Right-click on the entry in 'Result list' for which you would like to visualize a tree.

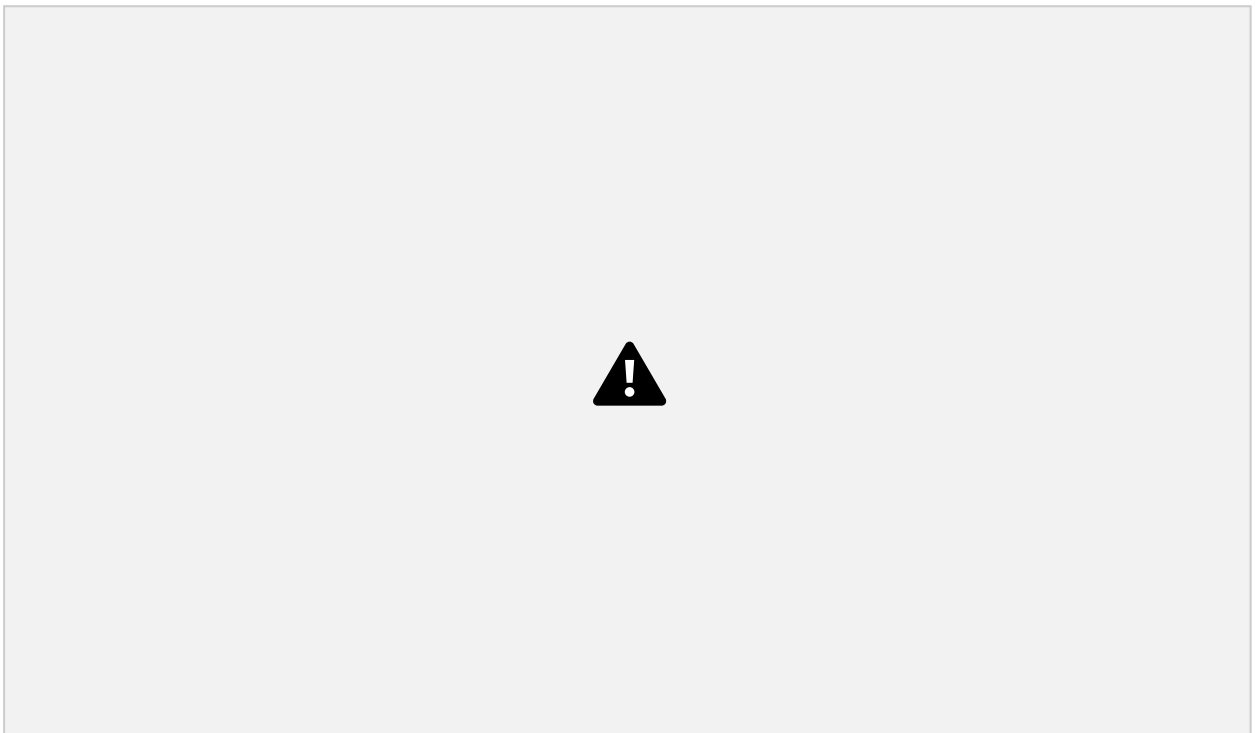


Select the item 'Visualize tree'; a new window comes up to the screen displaying the tree.





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.



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.**



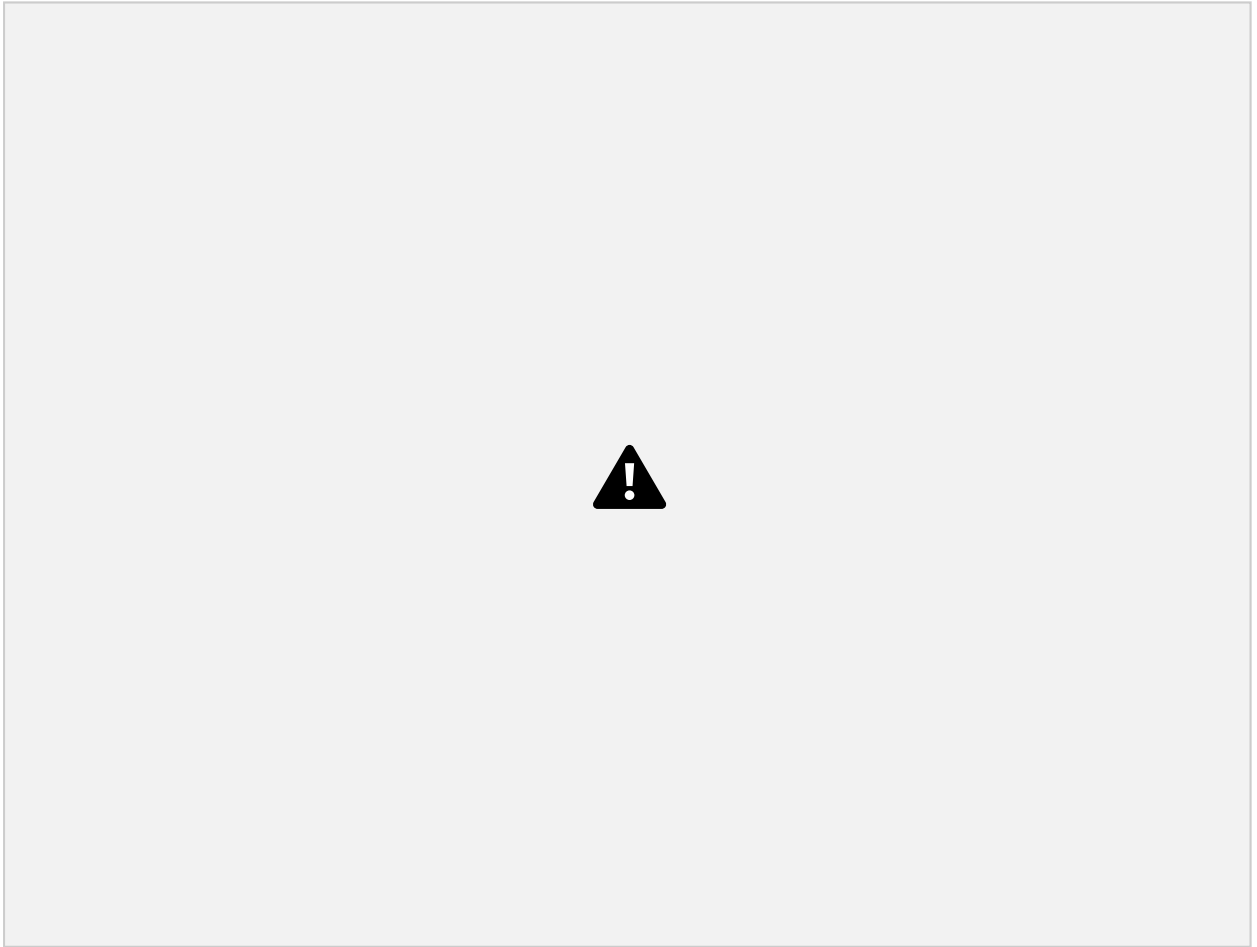
**Click on ‘Classes to cluster evaluation’ radio-button in ‘Cluster mode’ box and select ‘marital\_status’ in the pull-down box below.**



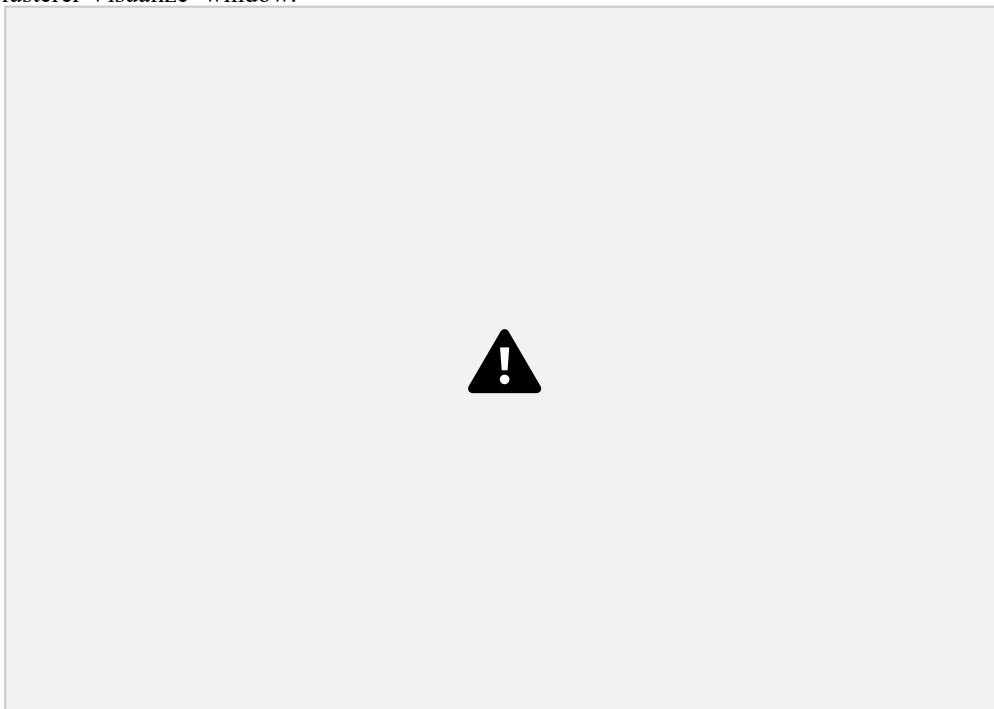
Click on the 'Start' button to execute the algorithm.



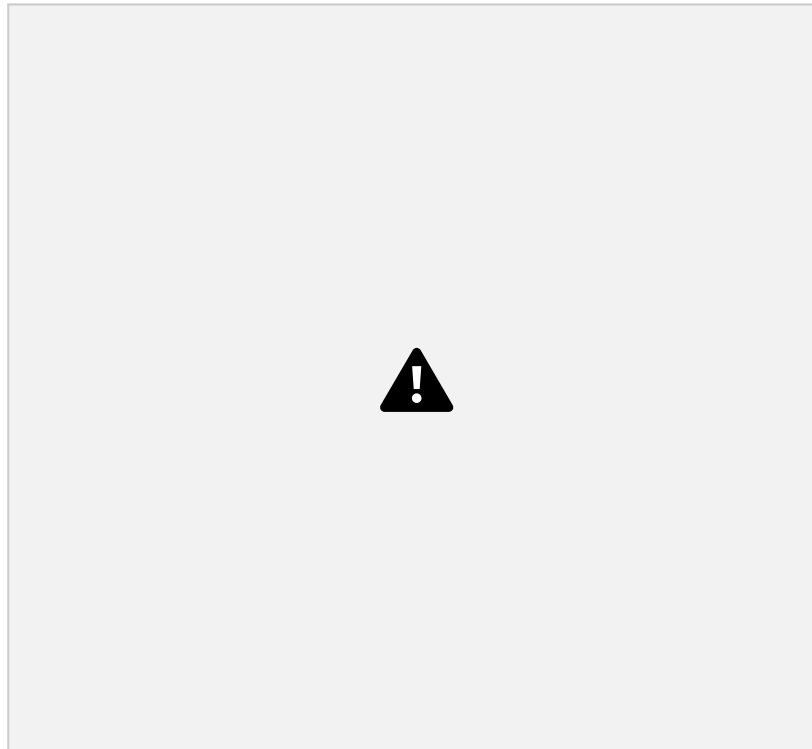
Right-click on the entry in the 'Result list' and select 'Visualize cluster assignments' in the pull-down window.



‘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.



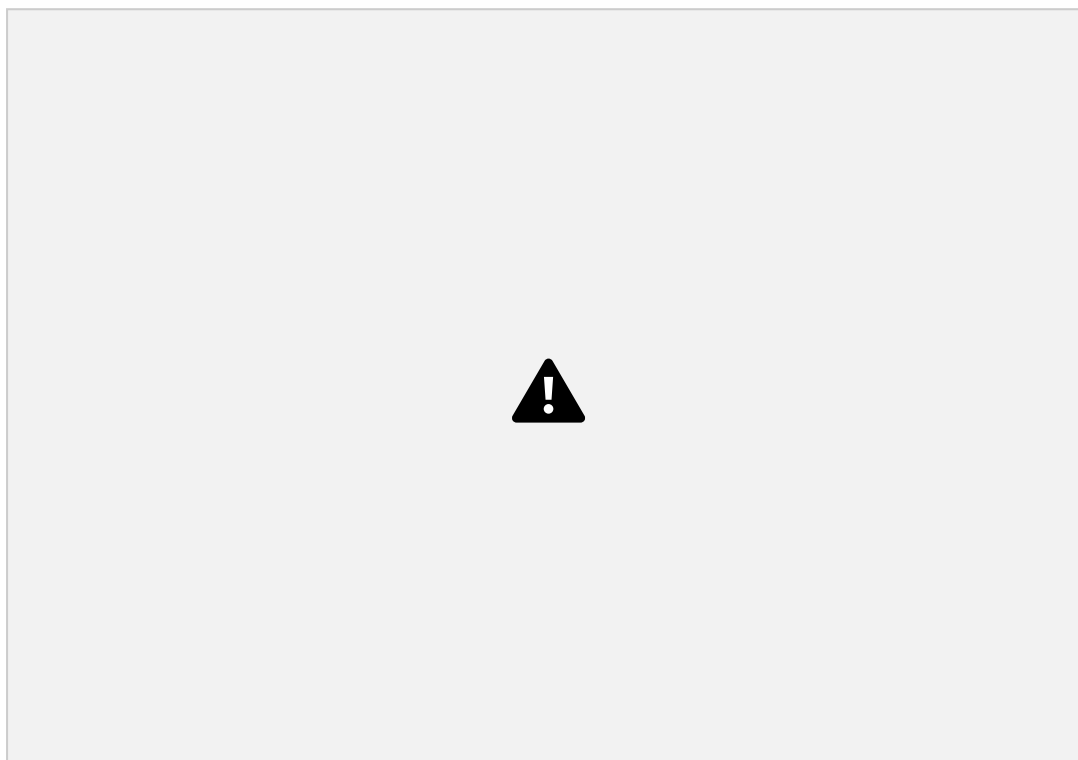
## Practical No-14

**Aim :** Association Rule Mining.

**Solution :**

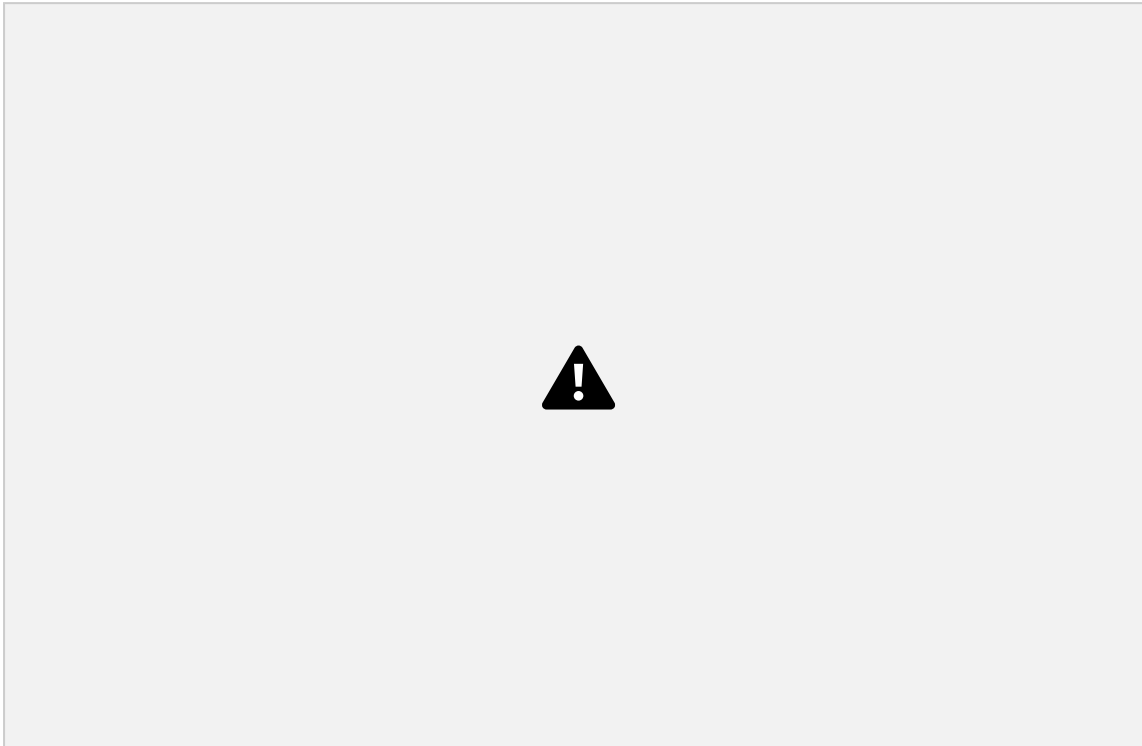
### Choosing Association Scheme

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

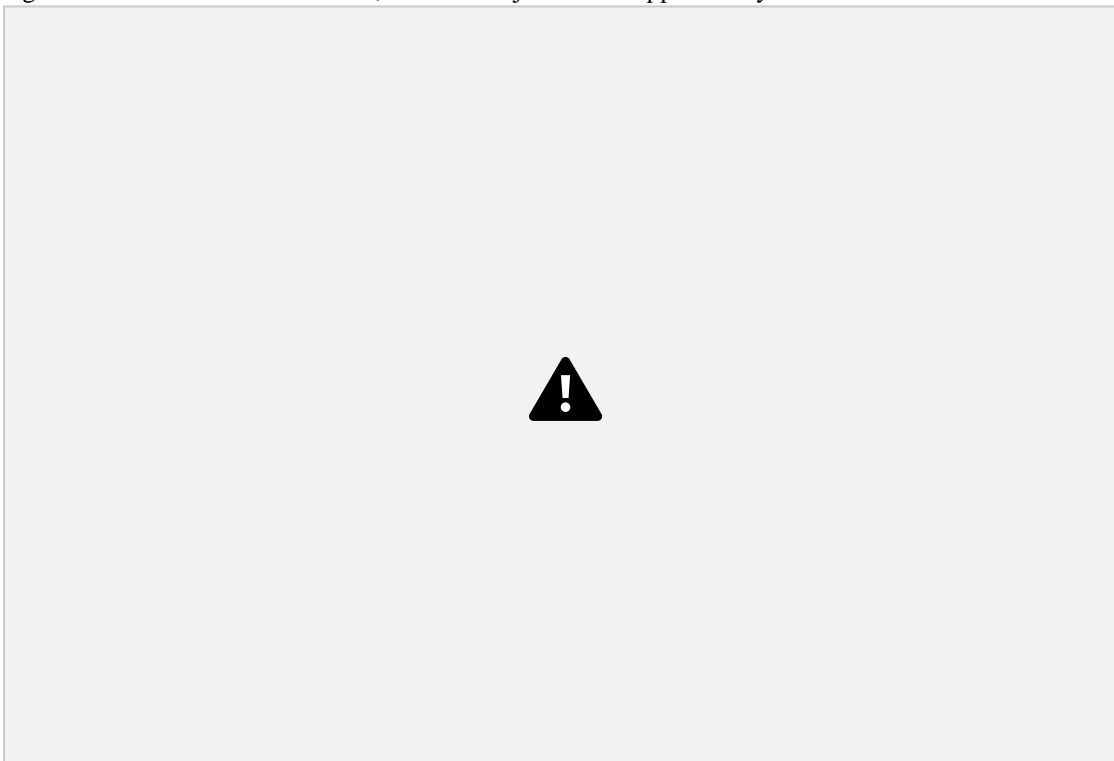


**Setting Test Options**

Check the text field in the 'Associator' box at the top of the window

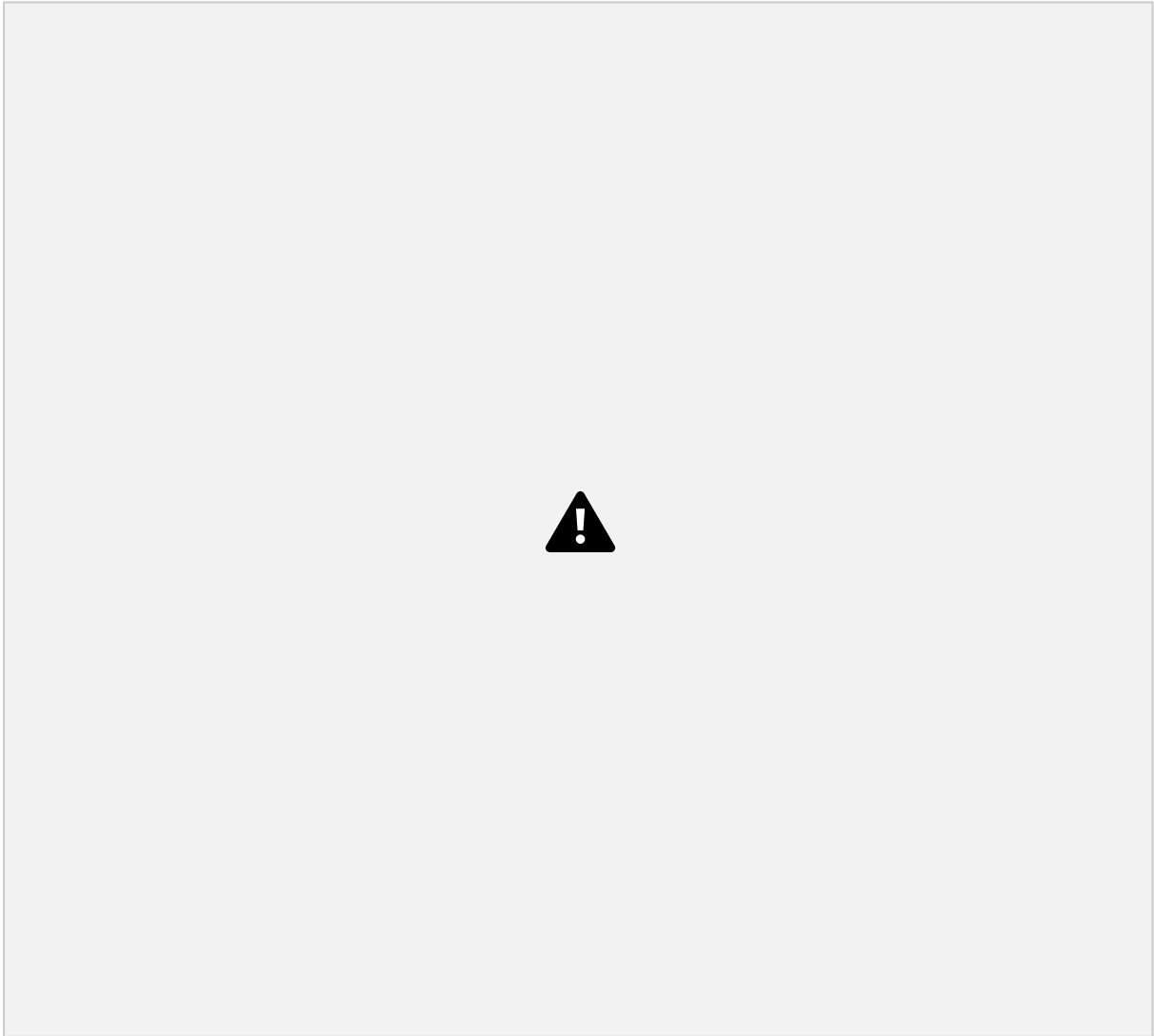


Right-click on the 'Associator' box, 'GenericObjectEditor' appears on your screen



Click on the 'Start' button to execute the algorithm





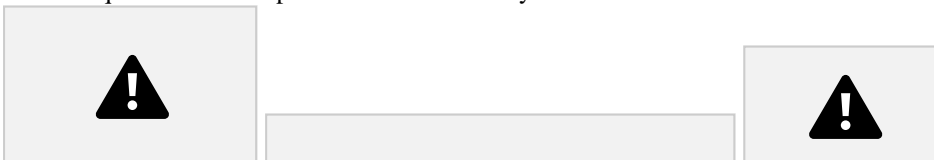
### **Practical No-15**

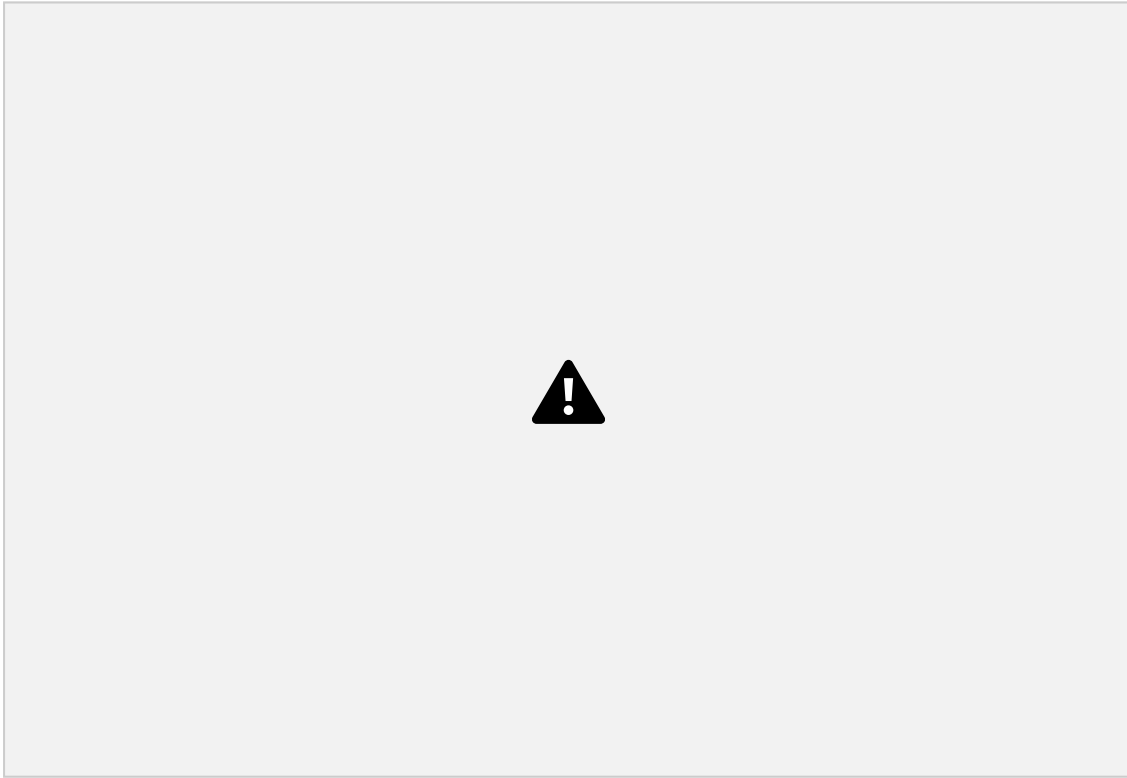
**Aim : Data Visualization**

**Solution :**

To open Visualization screen, click 'Visualize' tab.

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



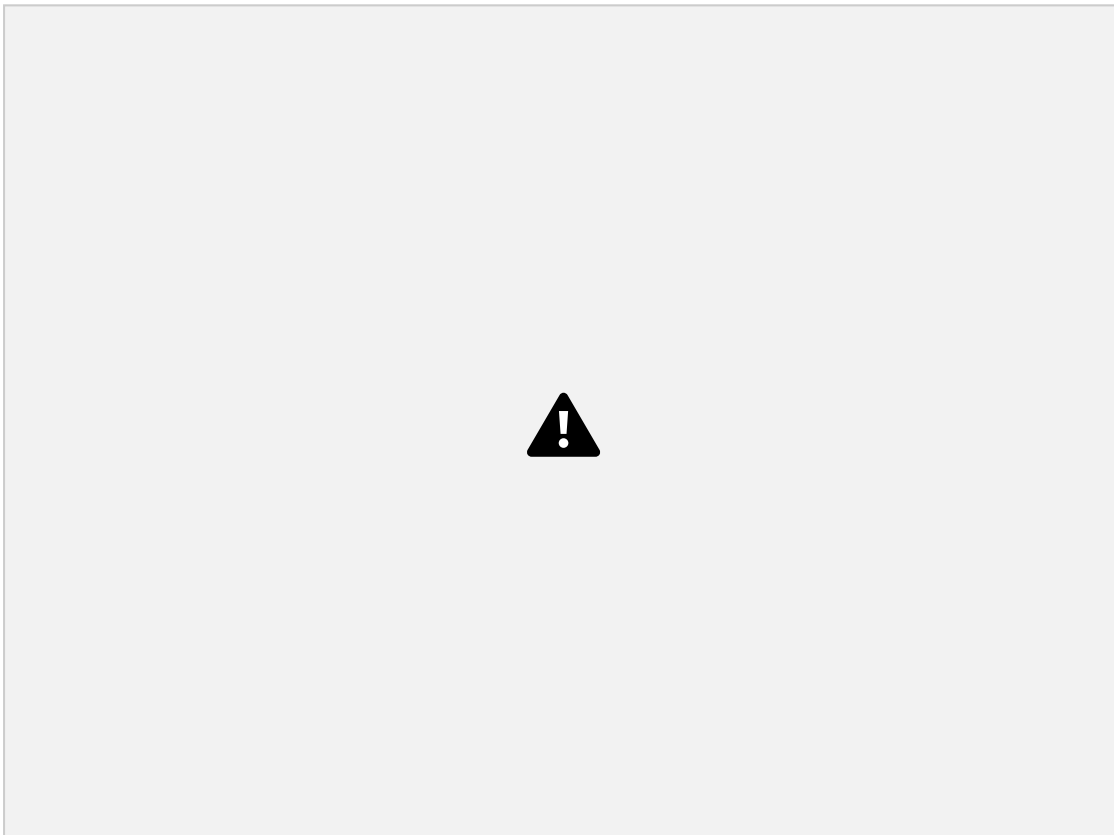


A

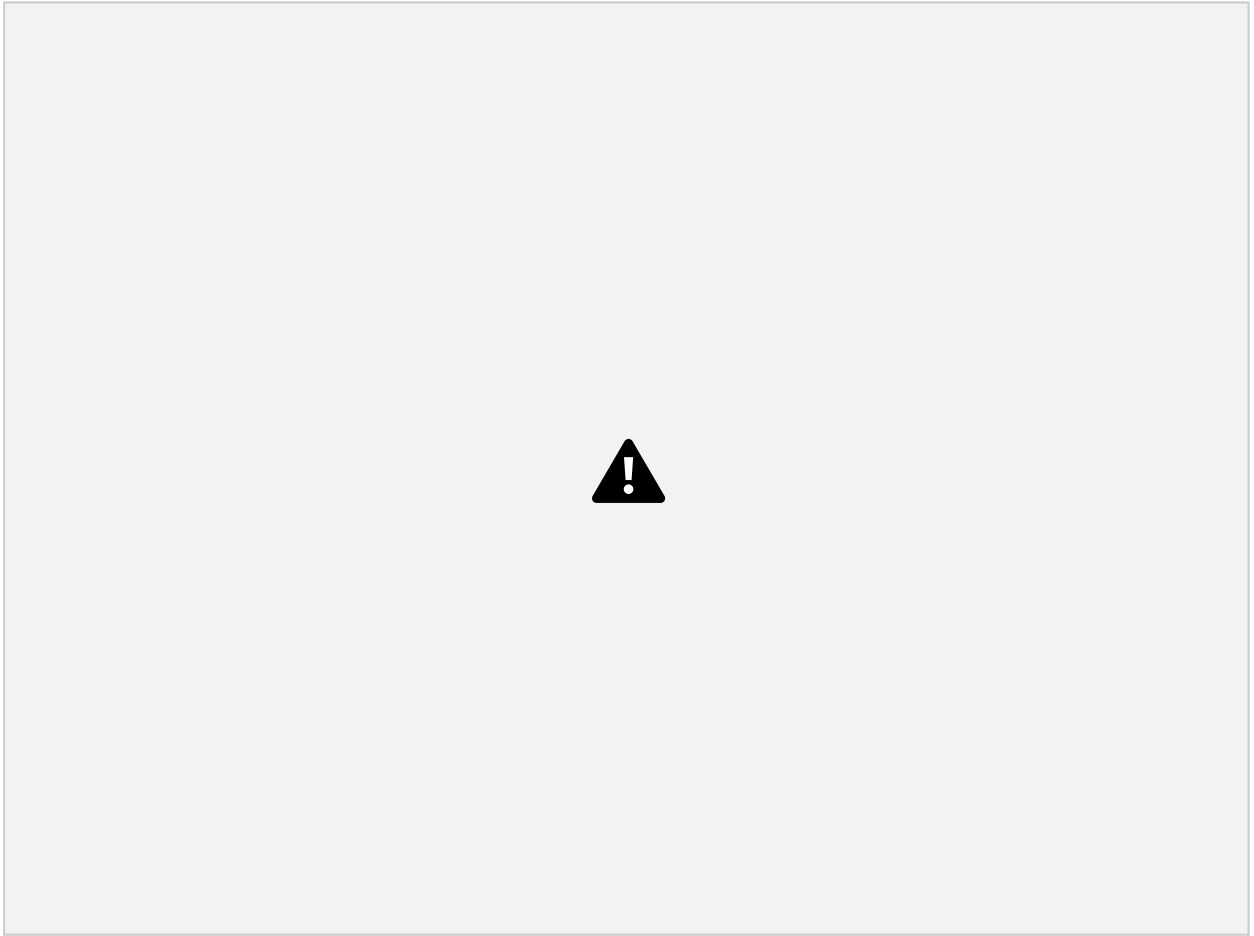
‘Visualizing weather’ window appears on the screen.

### **Changing the View**

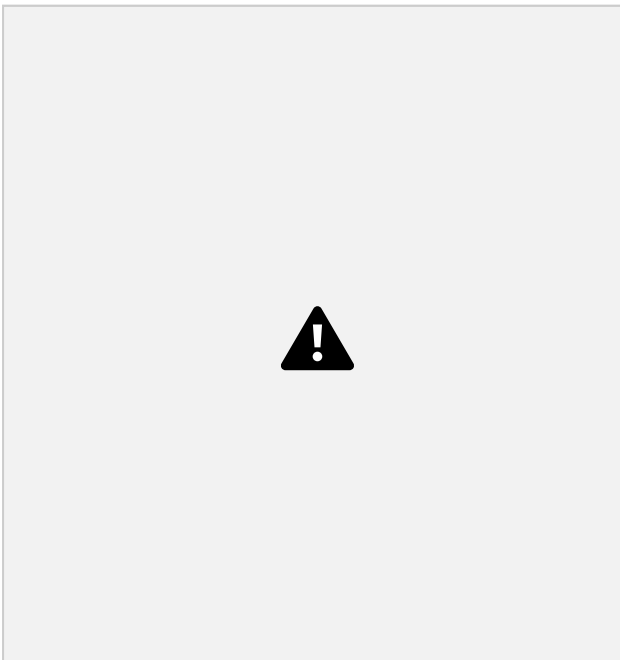
Keep sliding ‘Jitter’, a random



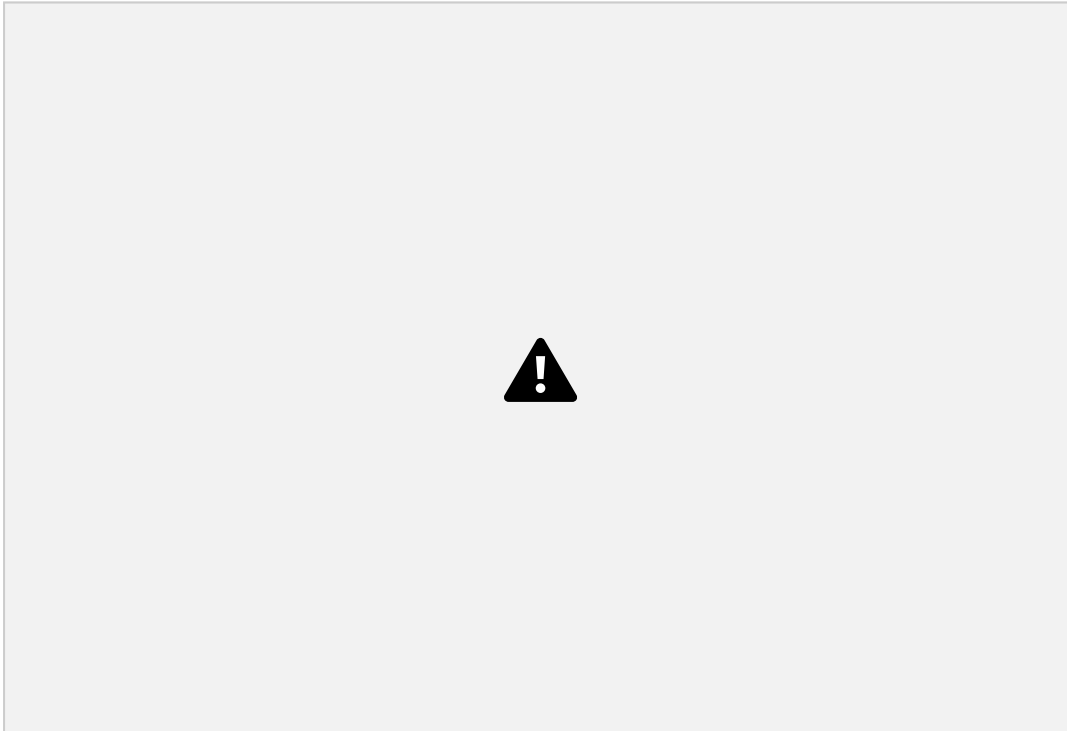
displacement given to all points in the plot, to the right, until you can spot concentration points



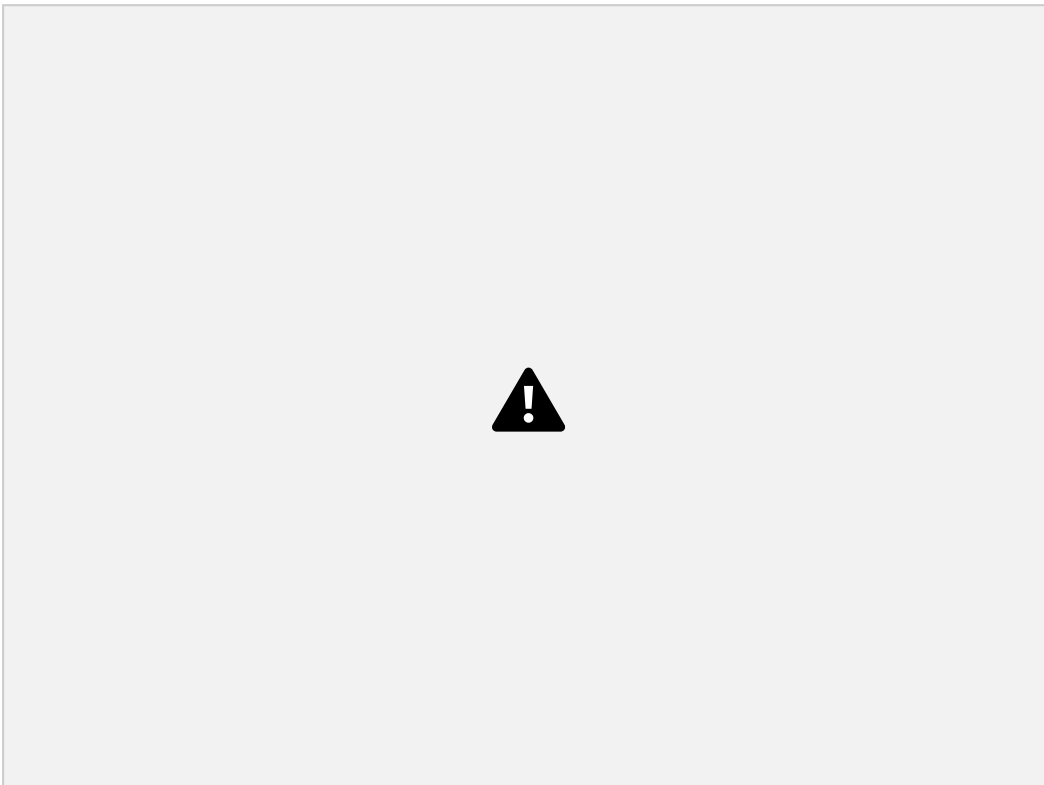
**Selecting Instances:** Click on an individual data point



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



4. **Polyline.** To distinguish the points on one side from the once 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

