Task:0

Aim: To practice sub queries and joins in Ms. Sql Server Management Studio.

Procedure:

First we create a database and tables in it and then we will insert records into tables.

```
SQLQuery1.sql - L...RNPTP1\sql (53))*
use b4;
create table saleorder
 (
salesorderid integer primary key,
customerid int,
cname varchar(10),
orddate date,
ordammount int);
create table products
 (
pid int primary key,
pname varchar(10),
stock int);
create table saledetails
 (
salesdetailid int primary key,
salesorderid int,
pid int,
foreign key(salesorderid) references salesorder(salesorderid),
unitprice integer,
foreign key(pid) references product(pid));
insert into saleorder
values(1,11,'keerthi','2018-07-22',100),
(2,22,'anjali','2018-07-26',200),
(3,33,'sunaina','2018-09-08',300);

select * from saleorder;
insert into products
values(51,'book',21),
(58,'usb',23);
select * from products;
insert into saledetails
values(31,1,51,100),
(32,1,58,90),
(33,2,51,200),
(34,2,58,150);
select * from saledetails;
```

Results | Messages

	salesorderid	customerid	cname	orddate	ordammount
1	1	11	keerthi	2018-07-22	100
2	2	22	anjali	2018-07-26	200
3	3	33	sunaina	2018-09-08	300

Results | Messages

	pid	pname	stock
1	51	book	21
2	58	usb	23

Results | Messages

	salesdetailid	salesorderid	pid	unitprice
1	31	1	51	100
2	32	1	58	90
3	33	2	51	200
4	34	2	58	150

Now,we will apply sub queries ,corelated sub queries and join operations on them.

```
select cname from saleorder where salesorderid in(select salesorderid from saledetails);
select cname from saleorder where salesorderid in(select salesorderid from saledetails where unitprice=100);
select cname from saleorder s1
where s1.salesorderid < all(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < all(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname,unitprice from saleorder s1
inner join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

Results | Messages

cname
keerthi
anjali

```
select cname from saleorder where salesorderid in(select salesorderid from saledetails where unitprice=100);
select cname from saleorder s1
where s1.salesorderid < all(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < all(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname,unitprice from saleorder s1
inner join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

Results | Messages

cname
keerthi

```
select cname from saleorder s1
where s1.salesorderid < all(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < all(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname,unitprice from saleorder s1
inner join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```


```
select cname from saleorder s1
where s1.salesorderid < all(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname,unitprice from saleorder s1
inner join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

```
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2);
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname,unitprice from saleorder s1
inner join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```


```
select cname from saleorder s1
where s1.salesorderid < any(select s2.salesorderid from saledetails s2 where s2.unitprice=150);
select cname,unitprice from saleorder s1
inner join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

```
select cname,unitprice from saleorder s1
inner join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

Results | Messages

cname	unitprice
keerthi	100
keerthi	90
anjali	200
anjali	150

```
select cname,unitprice from saleorder s1
left join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

Results | Messages

cname	unitprice
keerthi	100
keerthi	90
anjali	200
anjali	150
sunaina	NULL

```
select cname,unitprice from saleorder s1
right join saledetails s2 on s2.salesorderid=s1.salesorderid;
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

Results | Messages

cname	unitprice
keerthi	100
keerthi	90
anjali	200
anjali	150

```
select cname,unitprice from saleorder s1
full outer join saledetails s2 on s2.salesorderid=s1.salesorderid;
```

Results | Messages

cname	unitprice
keerthi	100
keerthi	90
anjali	200
anjali	150
sunaina	NULL

Task:1

Aim: Basic concepts of ware house

1. Differences between data and information:

Characteristic	**Data**	**Information**
Definition (Oxford Dictionaries)	Facts and statistics collected together for reference or analysis	Facts provided or learned about something or someone Data as processed, stored, or transmitted by a computer
Refers to	Raw Data	Analyzed Data
Description	Qualitative Or Quantitative Variables that can be used to make ideas or conclusions	A group of data which carries news and meaning
In the form of	Numbers, letters, or a set of characters.	Ideas and inferences
Collected via	Measurements, experiments, etc.	Linking data and making inferences
Represented in	A structure, such as tabular data, data tree, a data graph, etc.	Language, ideas, and thoughts based on the data
Analysis	Not analyzed	Always analyzed
Meaning	Carries no specific meaning	Carries meaning that has been assigned by interpreting data
Interrelation	Information that is collected	Data that has been processed

2. Differences between RDBMS and Data Warehousing:

RDMS:

- It is used for online transaction processing.
- The view of data is flat relational.
- The most frequent type of access is read/write.

Data Warehousing:

- It is used for online analytical processing.
- The view of data is multidimensional.
- It mostly use the read access for stored data.

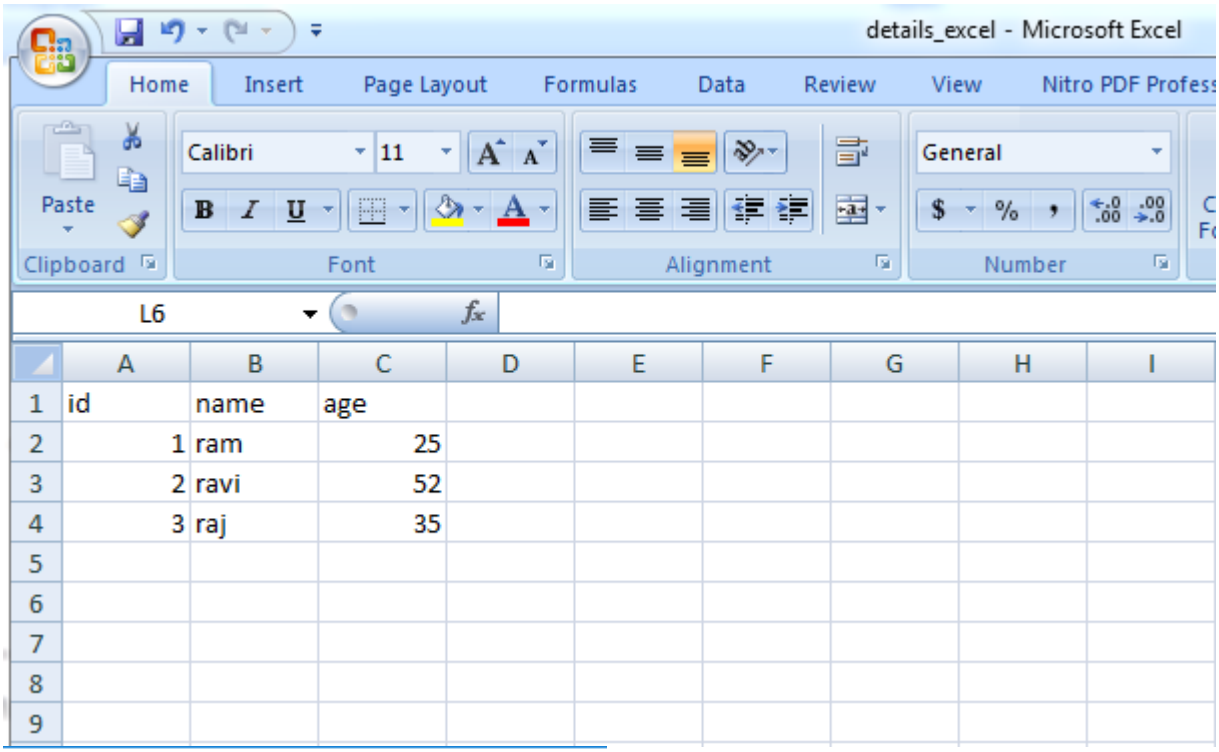
3. Five uses of Data Warehousing:
   - Used as an integral part for enterprise management.
   - For generating reports.
   - To analyze summarized and detailed data.
   - Business executives use the data in data warehouse to perform data analysis and to make strategic decisions.
   - For performing multidimensional analysis and sophisticated operations.
4. Tools of Data warehousing:
   - Amazon Redshift
   - Google Bigquery
   - Qlik
   - Fivetran
   - Blendo
5. ETL –extract, transform, load.
6. Use of ETL: For extracting data from source systems and bringing it into the data warehouse.
7. Basic concepts of Data Warehousing:
   - According to Bill Inmon, a data warehouse is subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision making process.
   - There is no frequent updating done in a data warehouse.
   - It possesses historical data, which helps the organization to analyze its business.
8. How is viewing data in a table is different from viewing data with respect to data warehouse:
   - In database, the output appears in rows and columns.
   - In data warehouse, it appears in multidimensional tables.

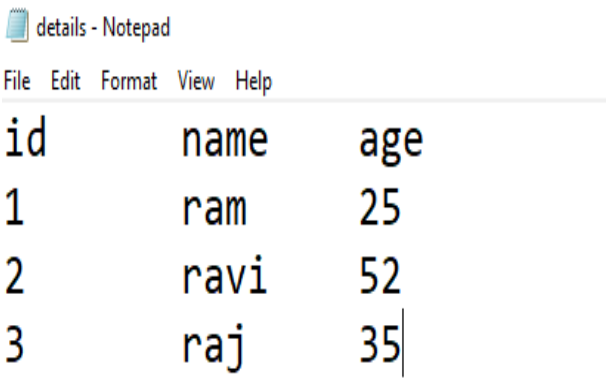
Task 2:

Aim: To load flat and CSV files into a database

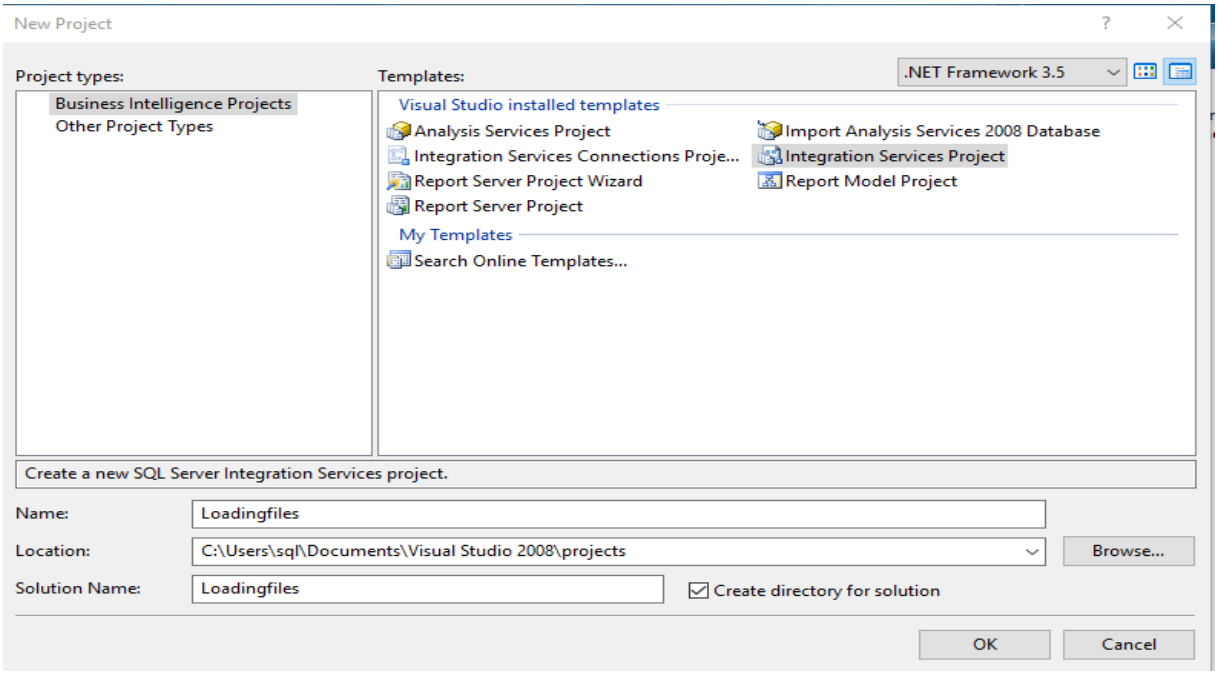
Procedure:

1. Create flat file and CSV file with sample data.

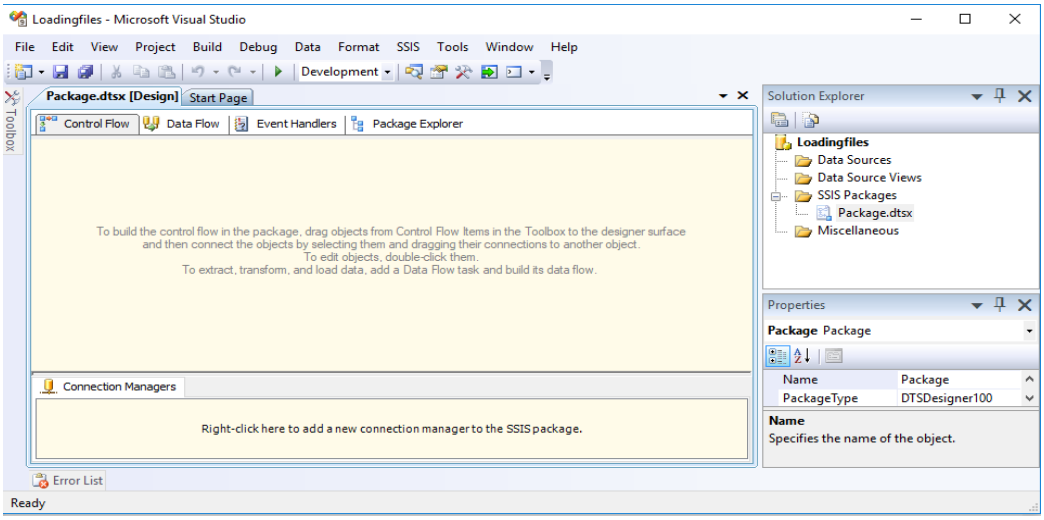


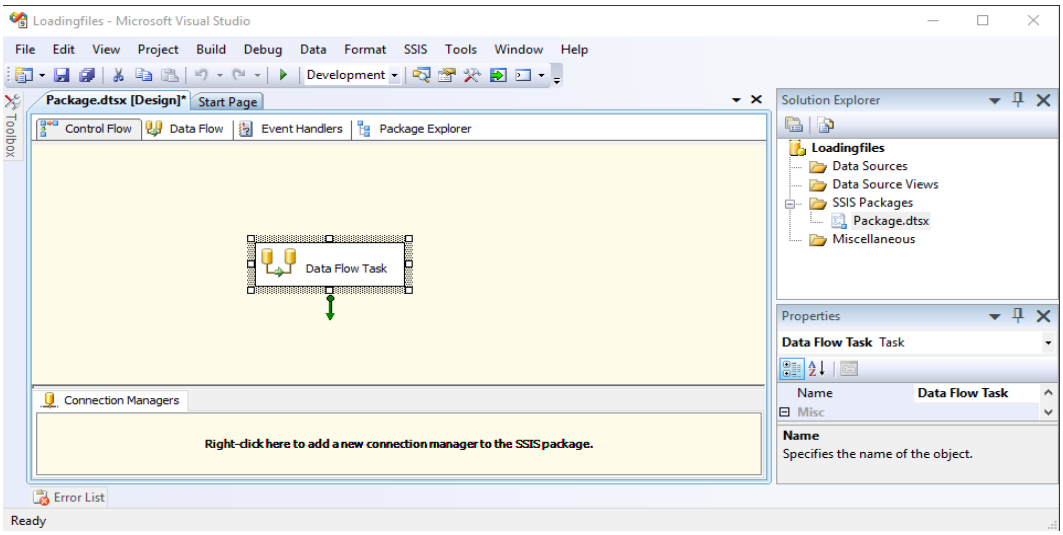


2.Now open a new project in sql business intellience studio and do as follows:

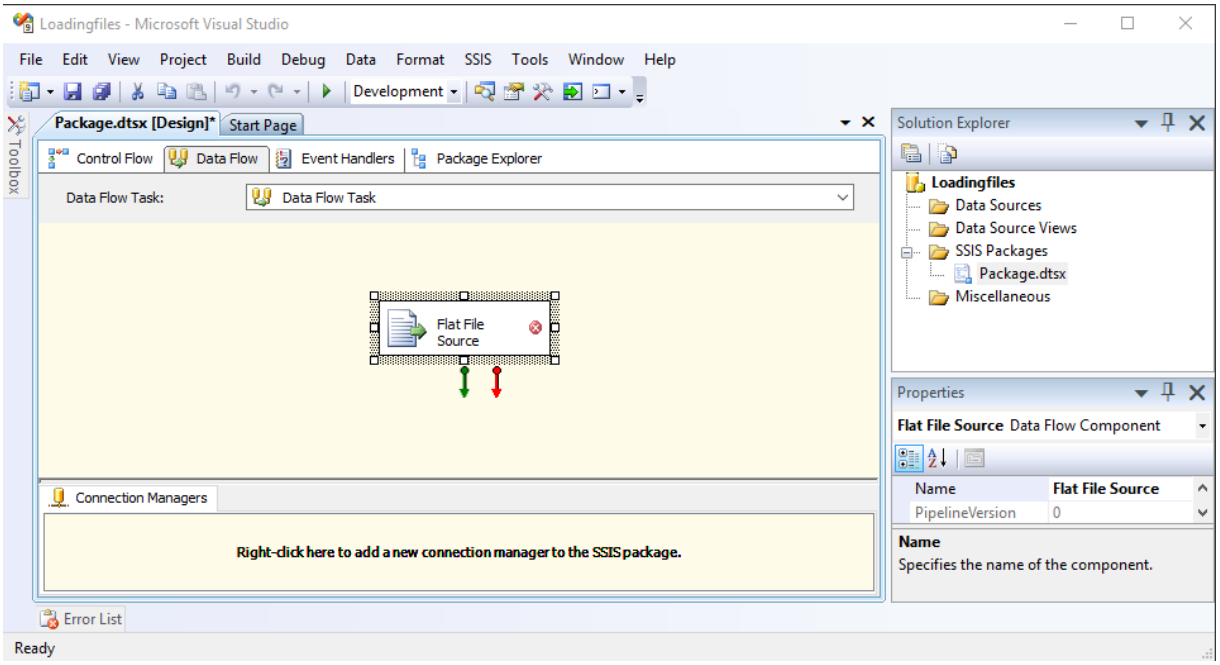


3.Now drag the data flow task from control flow items in toolbox

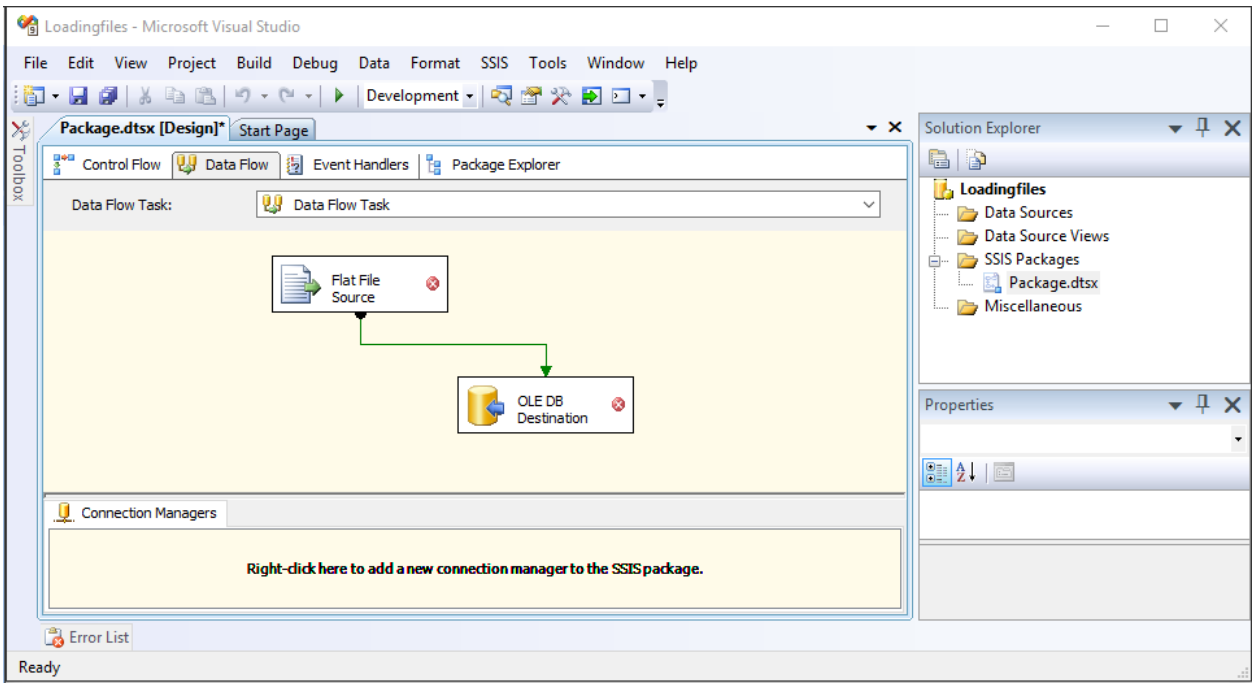




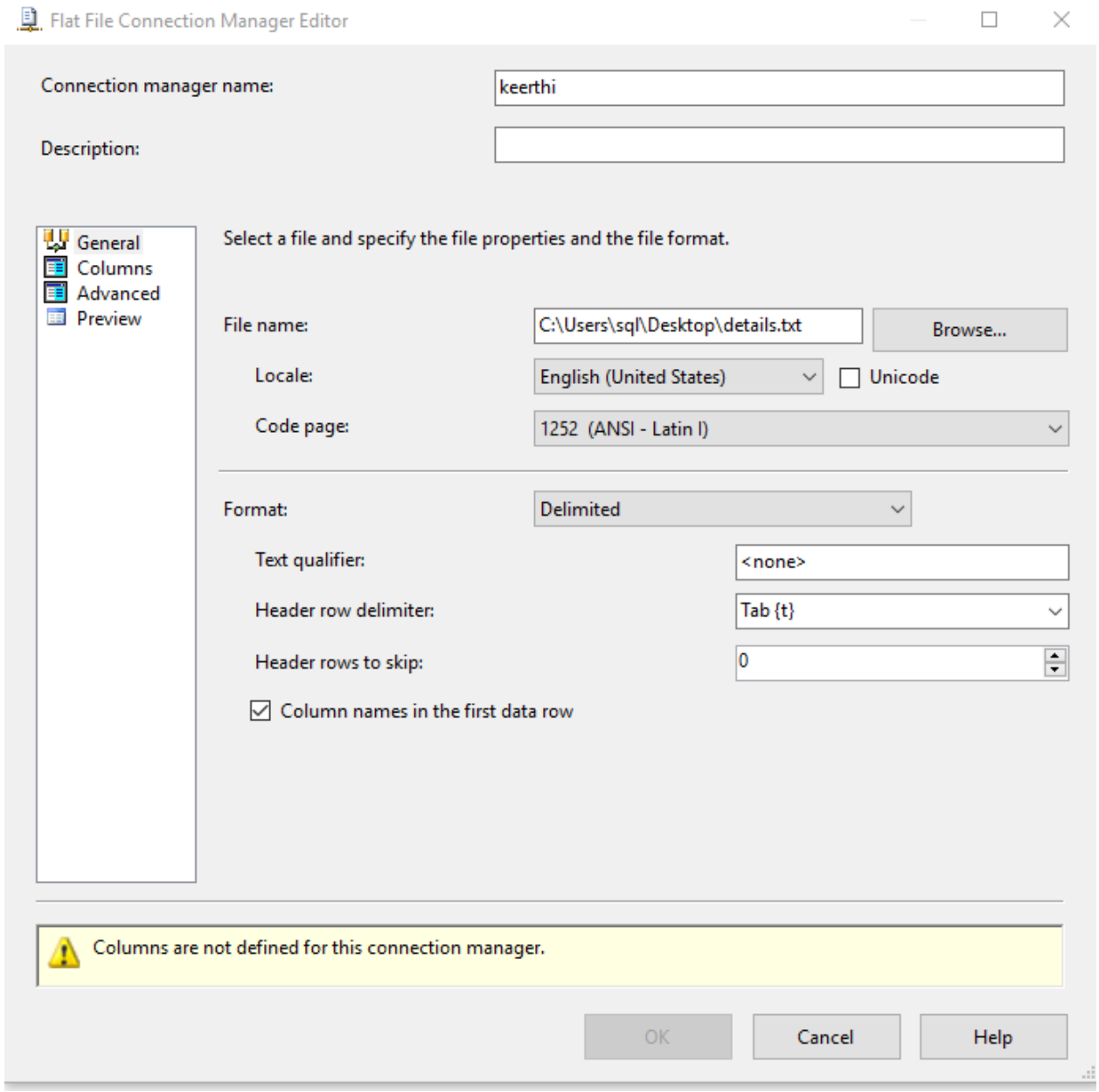
4.Now double click on data flow task and then select flat file source.

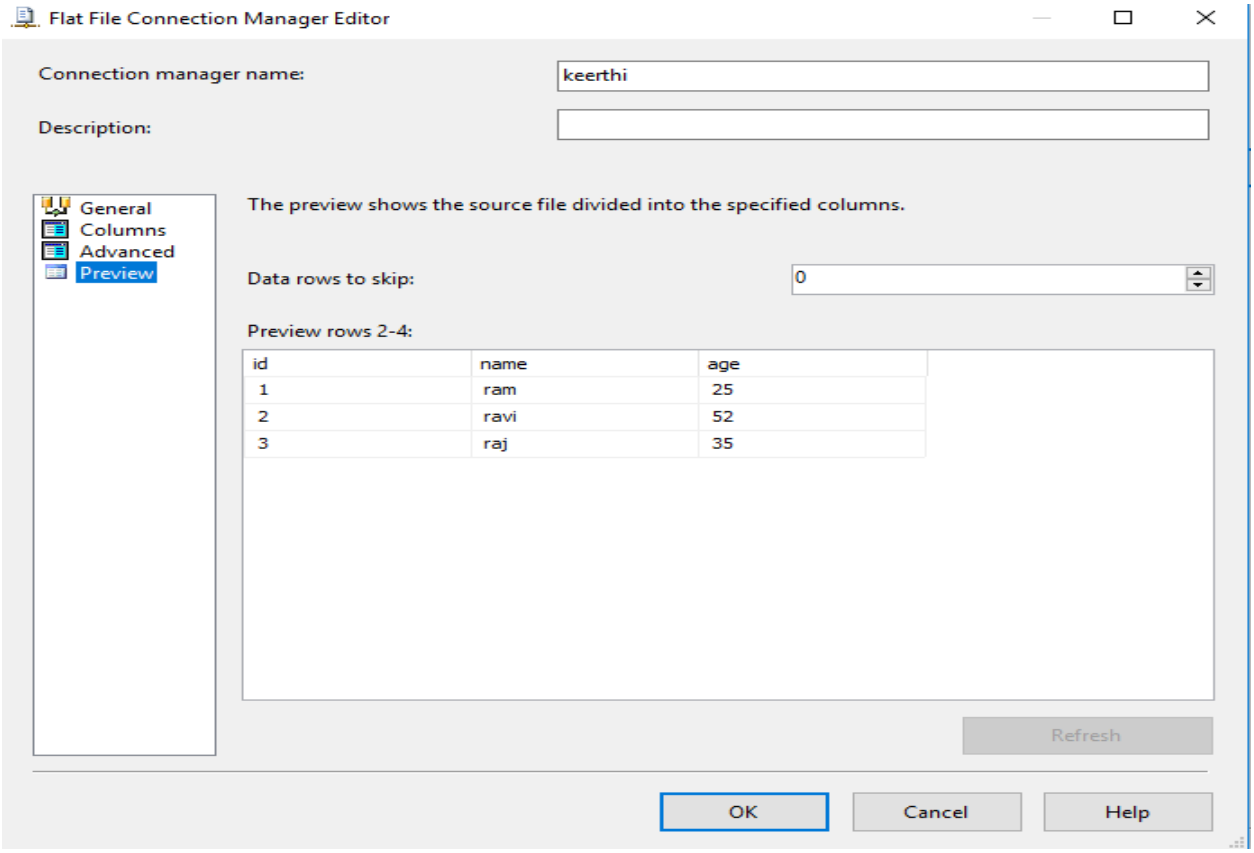


5.now drag the OLEDB destionation source and link it with flat file source.

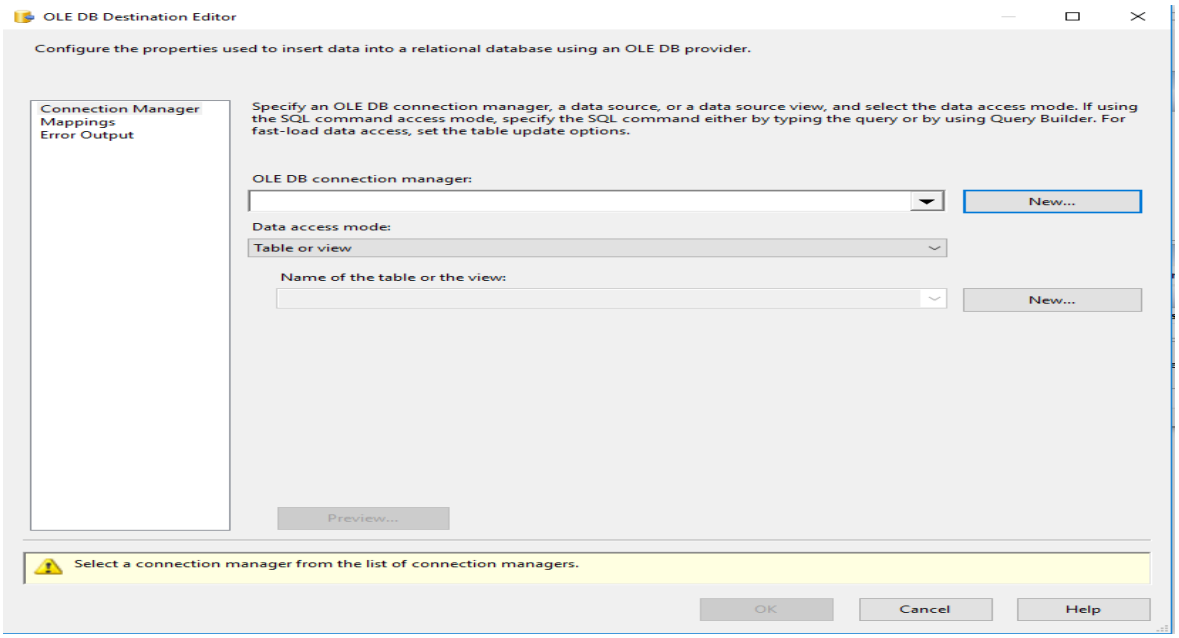


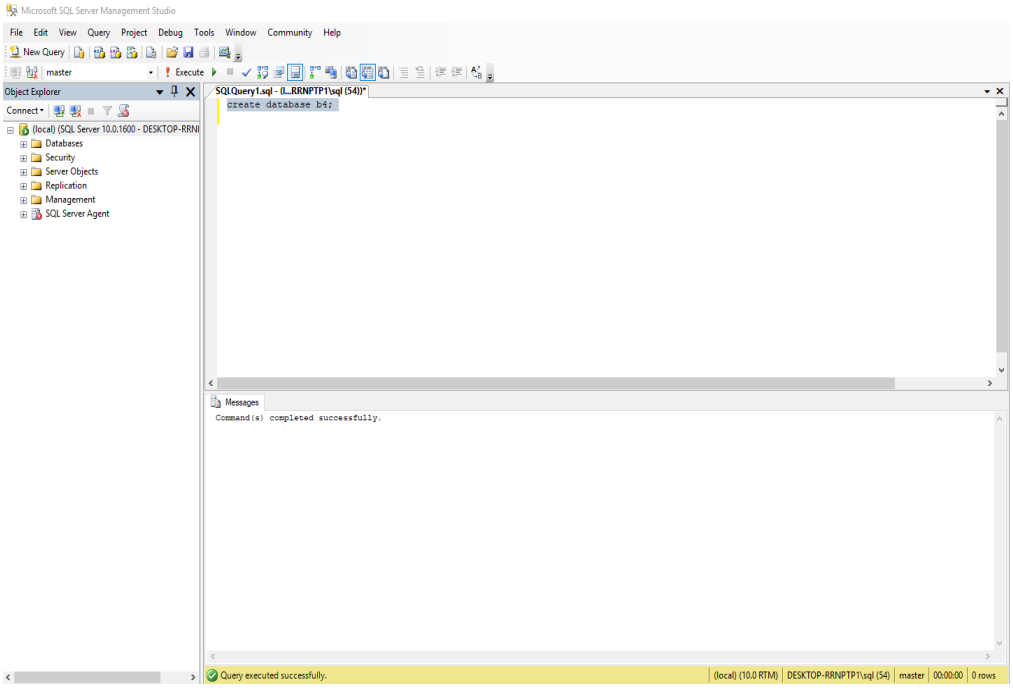
6.Now, double click on flat file source and give a connection manager name followed by selecting your file and after checking our data in preview option click ok.

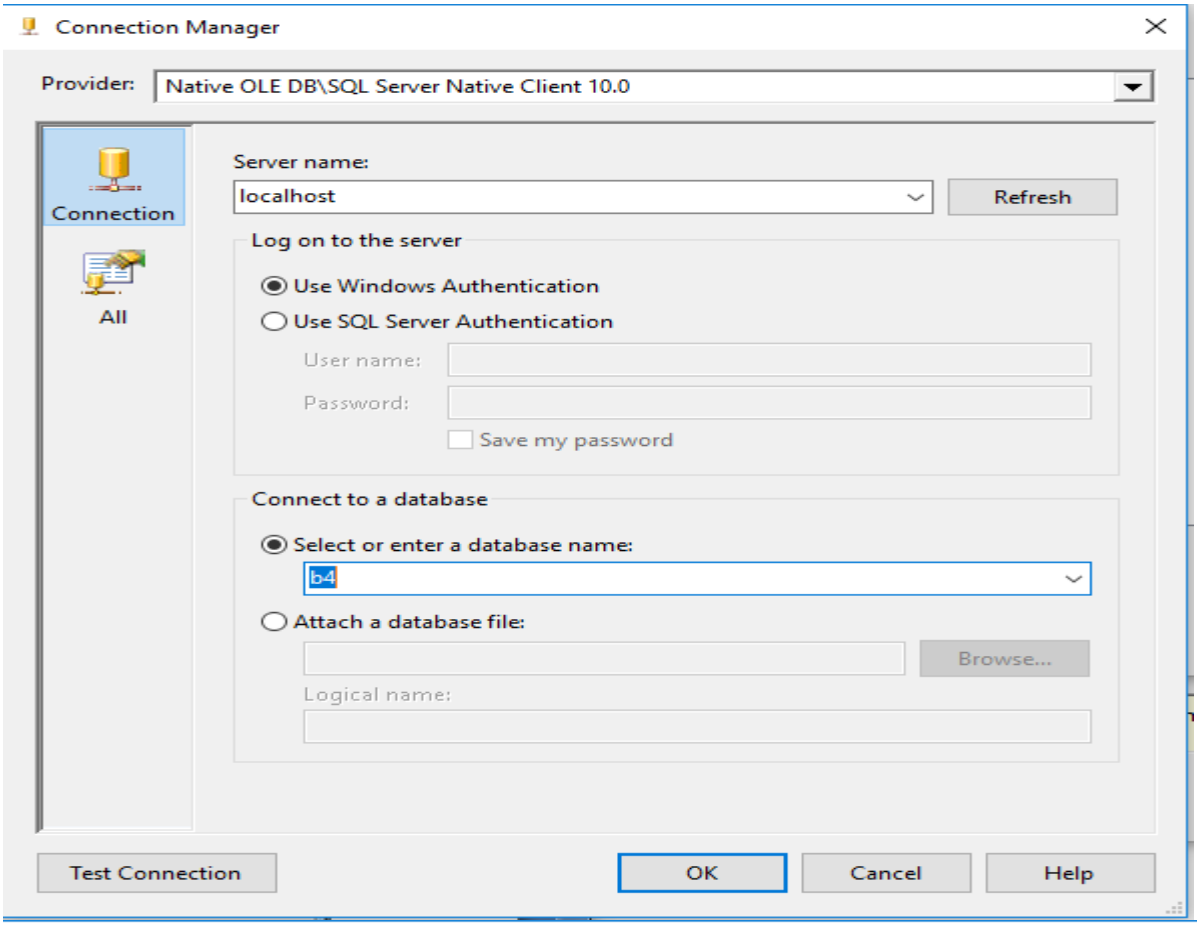

Flat File Connection Manager Editor
Connection manager name:
keerthi
Description:
General
Columns
Advanced
Preview
Select a file and specify the file properties and the file format.
File name:
C:\Users\sql\Desktop\details.txt
Browse...
Locale:
English (United States)
Unicode
Code page:
1252 (ANSI - Latin I)
Format:
Delimited
Text qualifier:
<none>
Header row delimiter:
Tab {t}
Header rows to skip:
0
Column names in the first data row
Columns are not defined for this connection manager.
OK
Cancel
Help

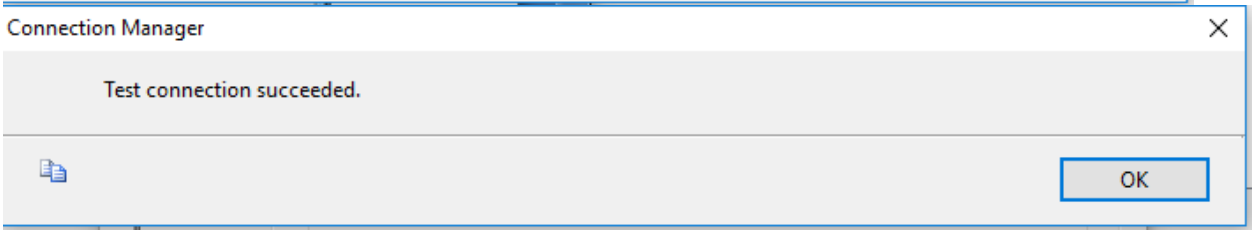


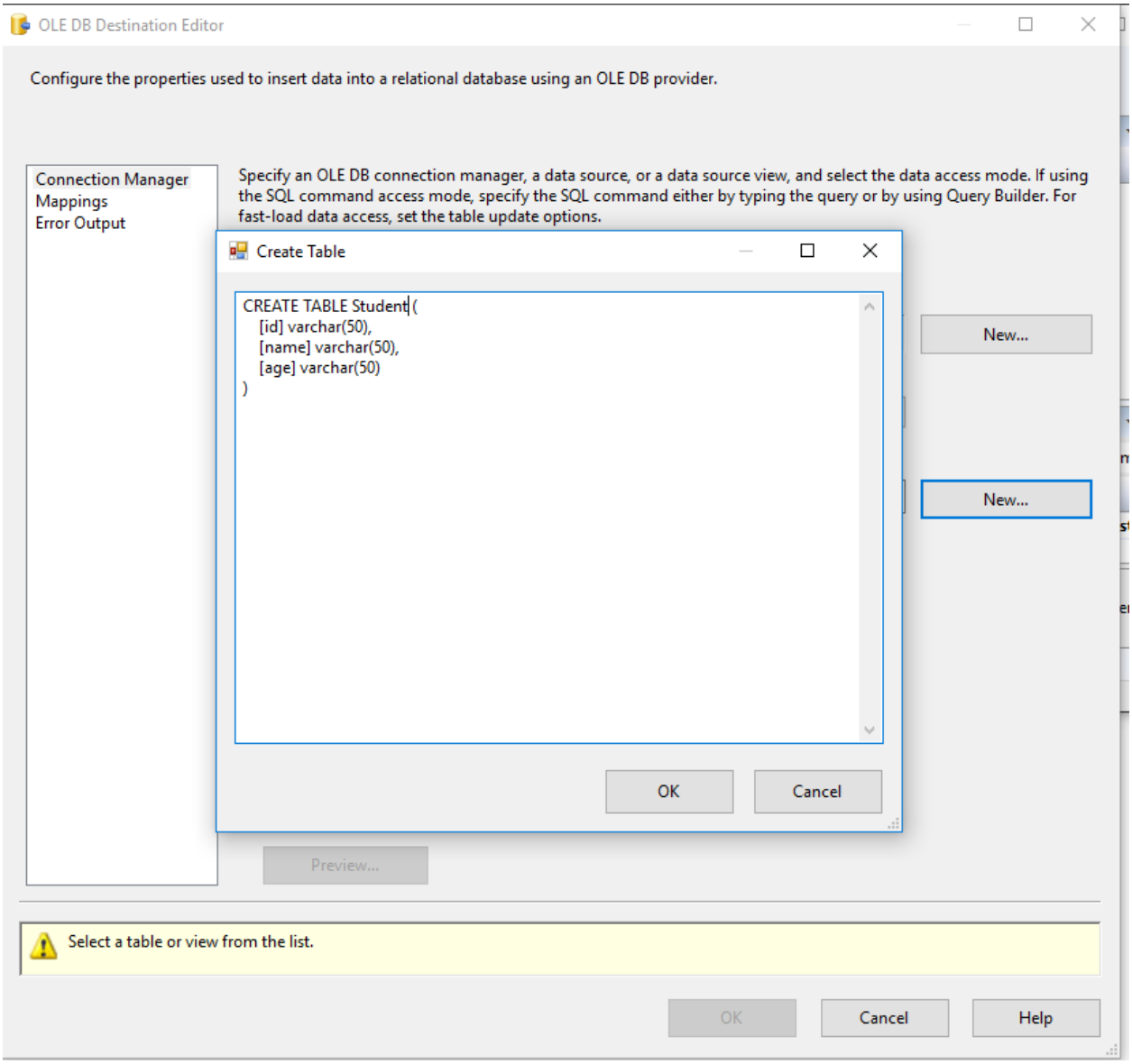
7.Now activate OLEDB destination source by selecting the server name and database name, followed by a test connection. And also give your table in database, a name.



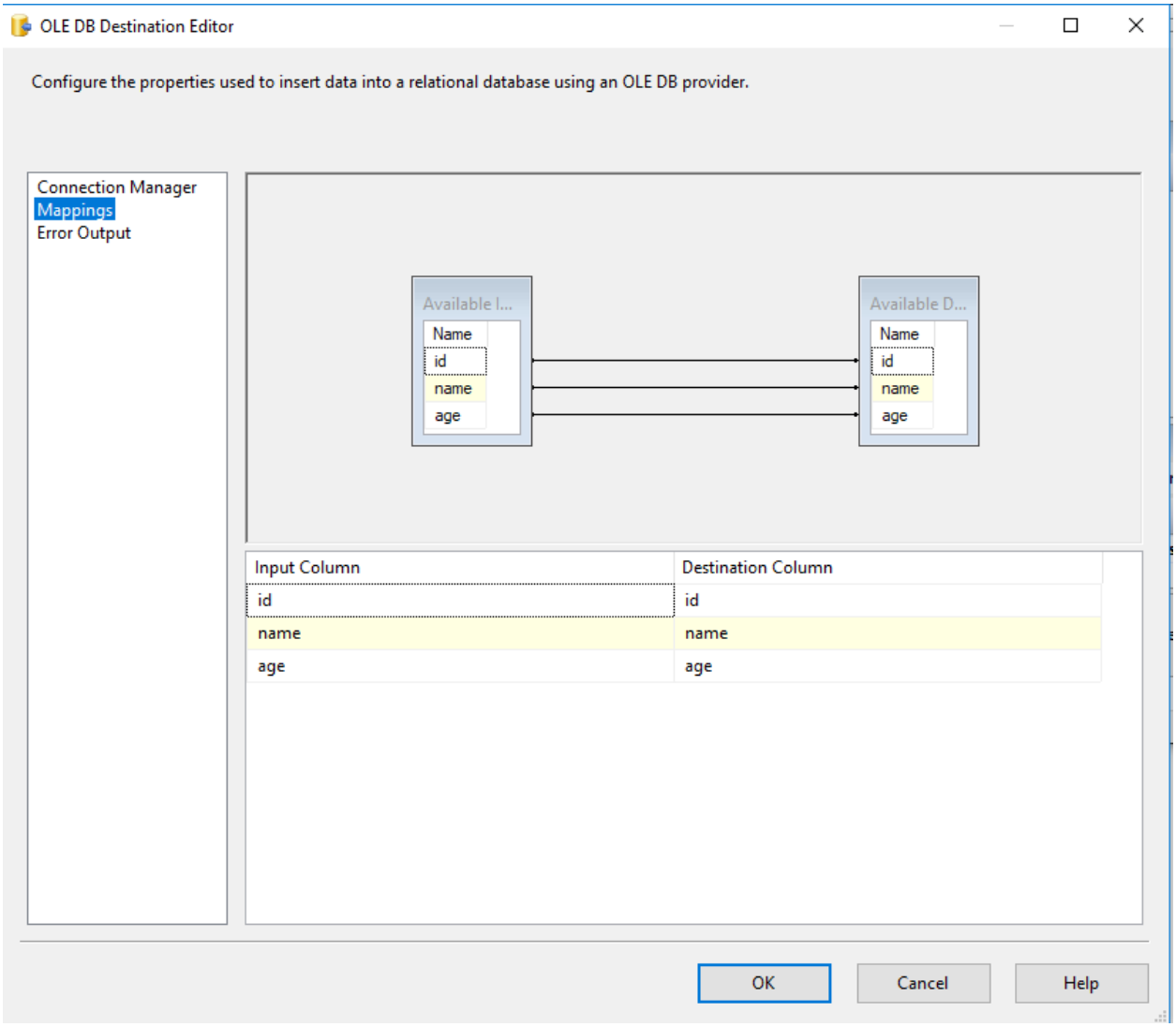

Microsoft SQL Server Management Studio
File Edit View Query Project Debug Tools Window Community Help
New Query
Execute
Object Explorer
Connect
(local) (SQL Server 10.0.1600 - DESKTOP-RRNI
Databases
Security
Server Objects
Replication
Management
SQL Server Agent
SQLQuery1.sql - (l...RRNPTP1\sql (54))*
create database b4;
Messages
Command(s) completed successfully.
Query executed successfully.
(local) (10.0 RTM)
DESKTOP-RRNPTP1\sql (54)
master
00:00:00
0 rows


Connection Manager
Provider:
Native OLE DB\SQL Server Native Client 10.0
Connection
All
Server name:
localhost
Refresh
Log on to the server
Use Windows Authentication
Use SQL Server Authentication
User name:
Password:
Save my password
Connect to a database
Select or enter a database name:
b4
Attach a database file:
Browse...
Logical name:
Test Connection
OK
Cancel
Help

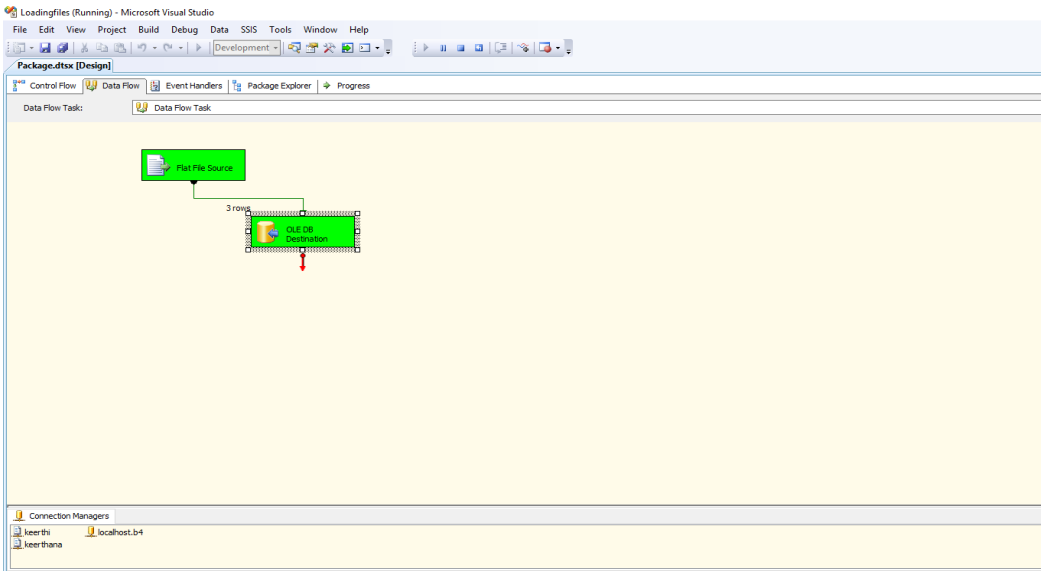

Connection Manager
Test connection succeeded.
OK



8. Now check your data, whether they are mapped to their respective column names in mappings option. And then click ok.

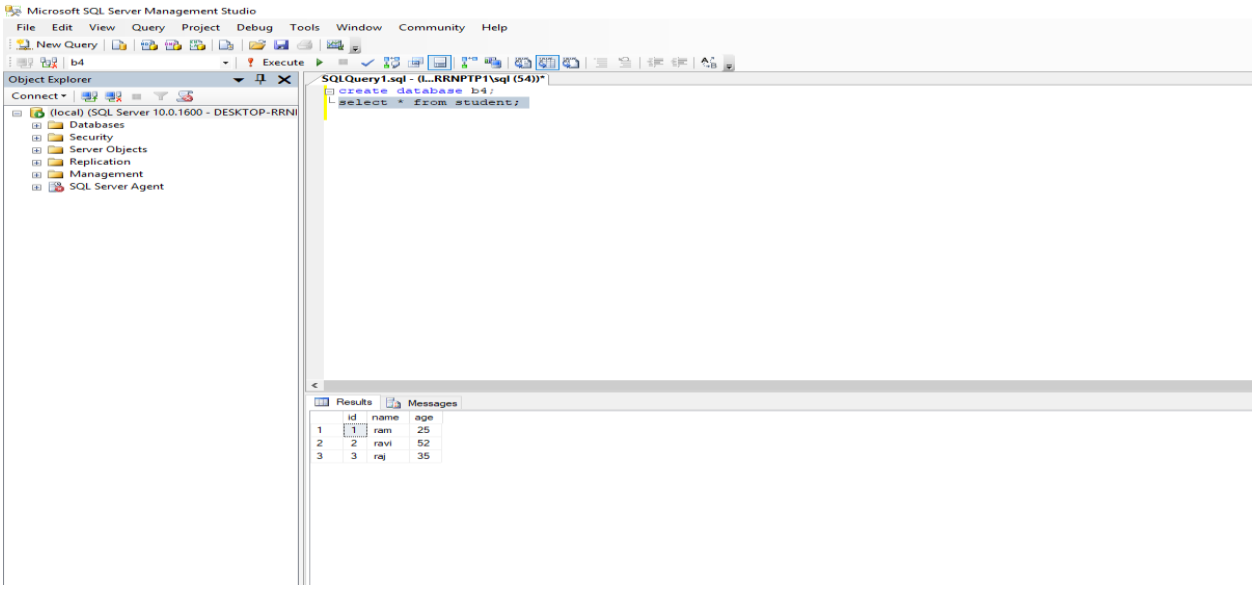


9. Now press F5 for debugging.



Output:

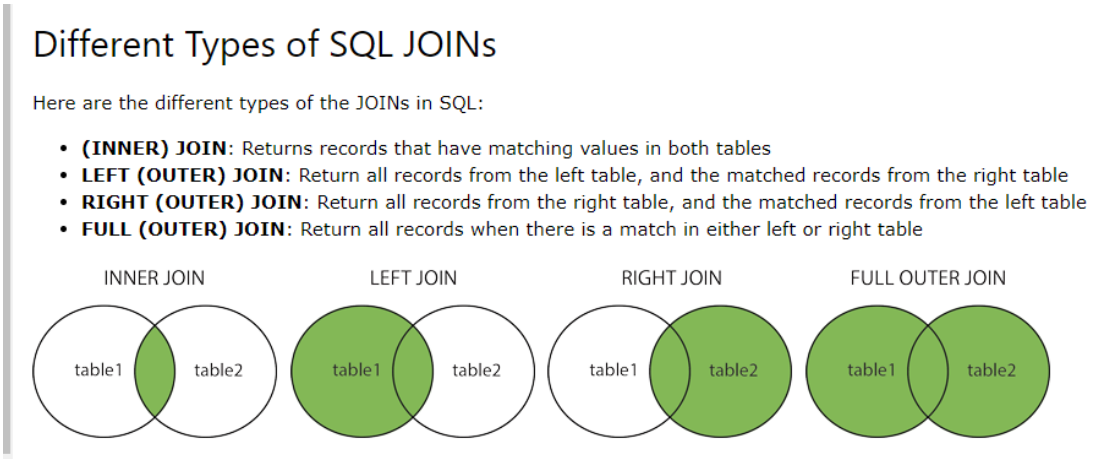
Now in Microsoft sql server management studio write the query for retrieving the data.



Task:3

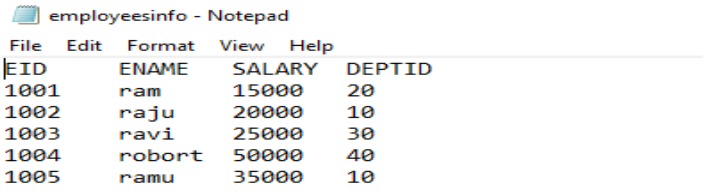
Aim: To perform various types of join operations using SSIS and loading data into sql server database.

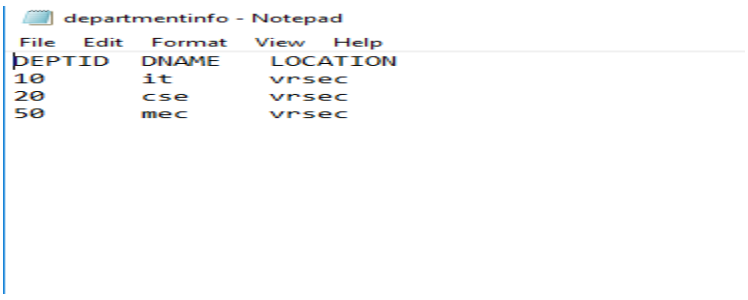
Theory:



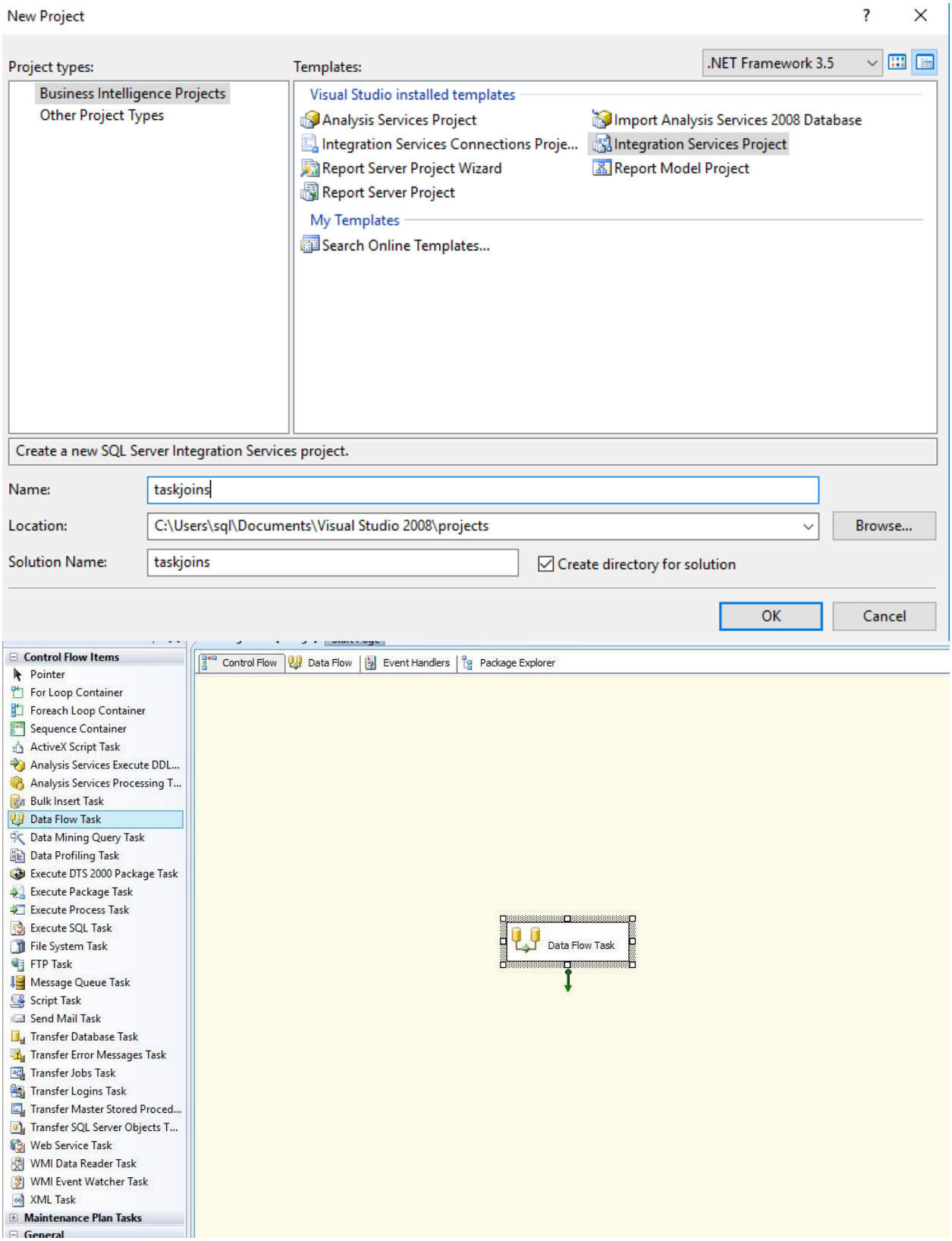
Procedure:

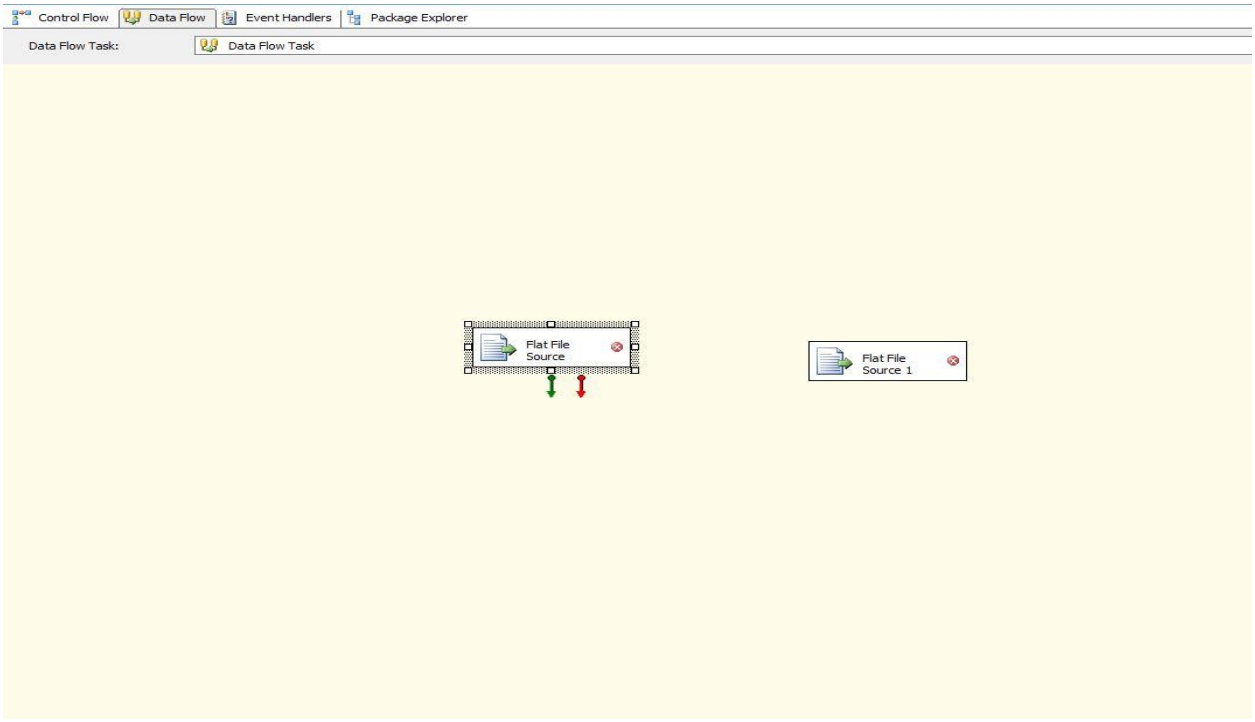
1.At first create two flat files with data having a common column in each of them.



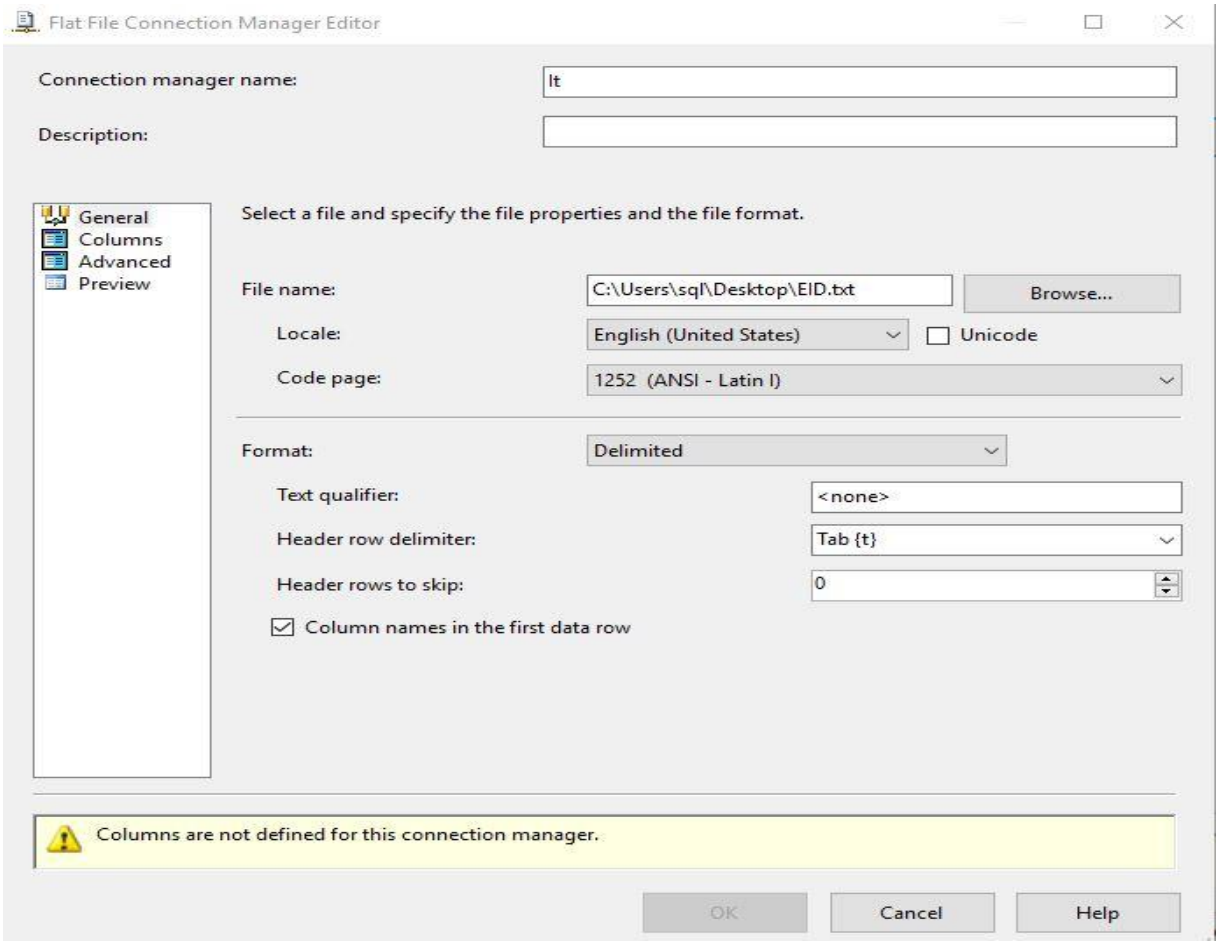


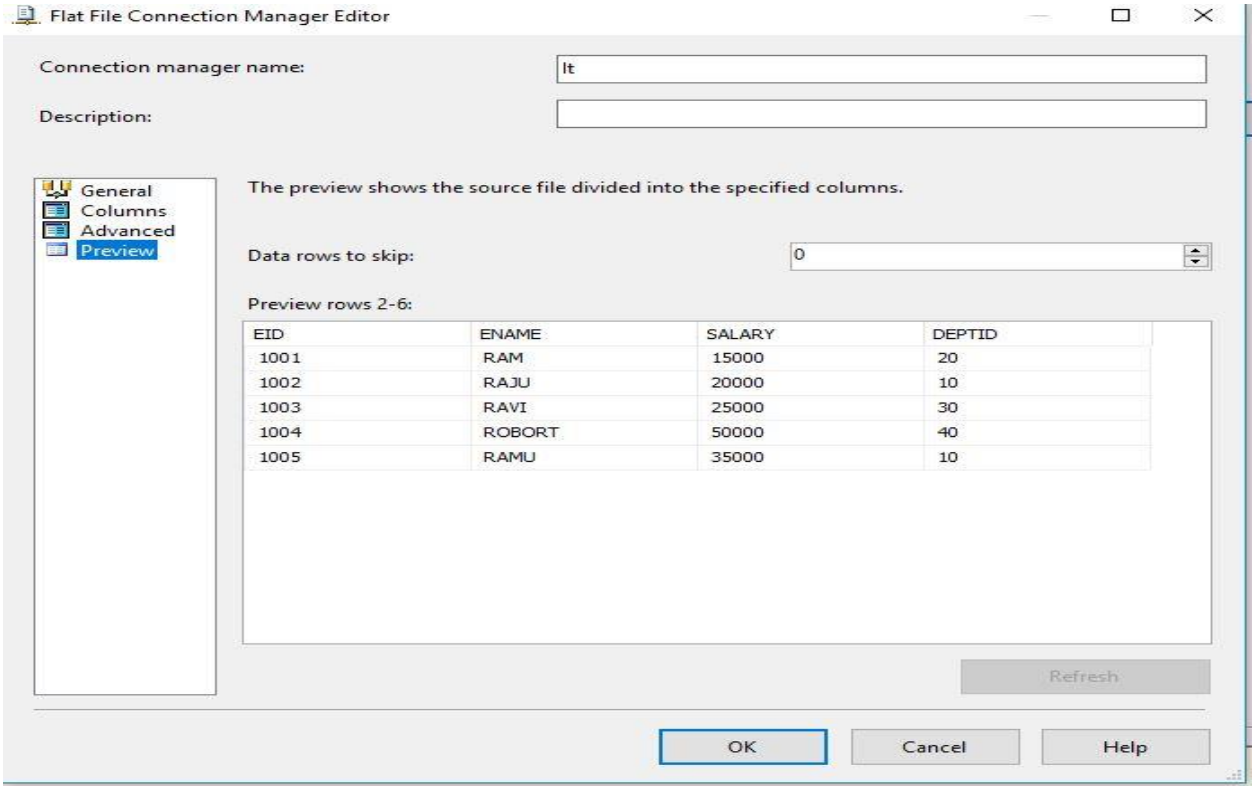
2. Open new project and in control flow drag the data flow task from the toolbox and in it drag the two flat file sources.

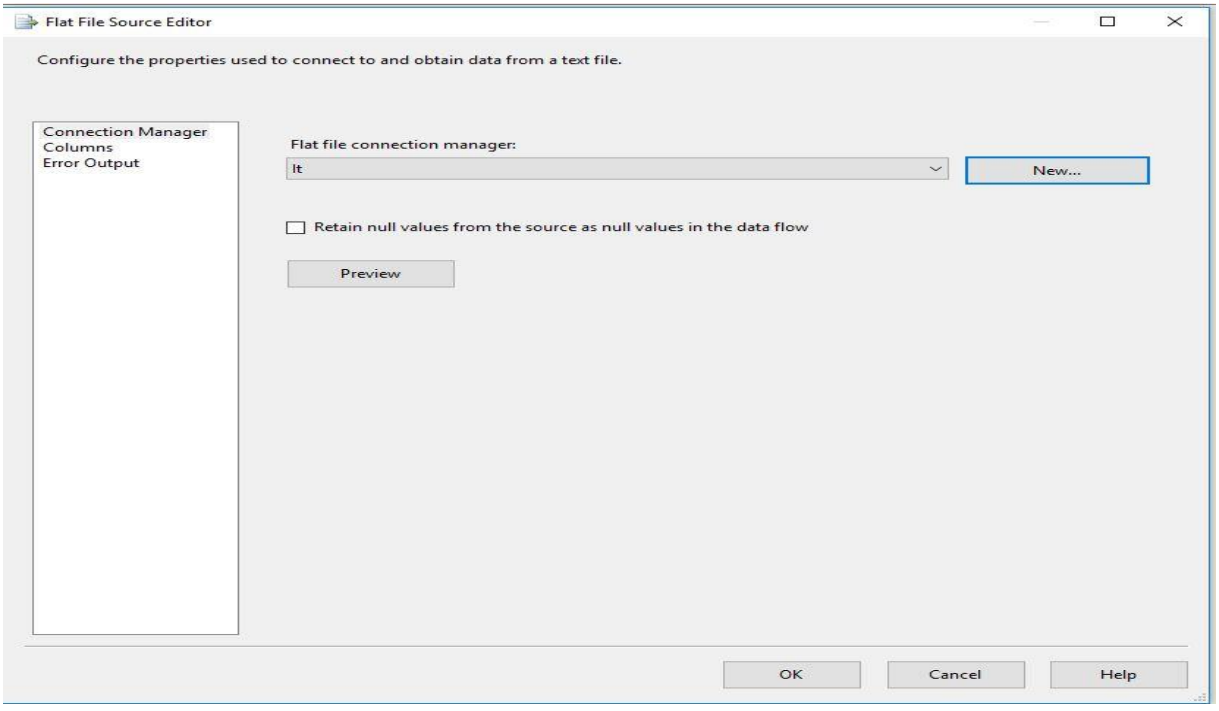

New Project
Project types:
Business Intelligence Projects
Other Project Types
Templates:
.NET Framework 3.5
Visual Studio installed templates
Analysis Services Project
Import Analysis Services 2008 Database
Integration Services Connections Proje...
Integration Services Project
Report Server Project Wizard
Report Model Project
Report Server Project
My Templates
Search Online Templates...
Create a new SQL Server Integration Services project.
Name:
taskjoins
Location:
C:\Users\sql\Documents\Visual Studio 2008\projects
Browse...
Solution Name:
taskjoins
Create directory for solution
OK
Cancel
Control Flow Items
Pointer
For Loop Container
Foreach Loop Container
Sequence Container
ActiveX Script Task
Analysis Services Execute DDL...
Analysis Services Processing T...
Bulk Insert Task
Data Flow Task
Data Mining Query Task
Data Profiling Task
Execute DTS 2000 Package Task
Execute Package Task
Execute Process Task
Execute SQL Task
File System Task
FTP Task
Message Queue Task
Script Task
Send Mail Task
Transfer Database Task
Transfer Error Messages Task
Transfer Jobs Task
Transfer Logins Task
Transfer Master Stored Proced...
Transfer SQL Server Objects T...
Web Service Task
WMI Data Reader Task
WMI Event Watcher Task
XML Task
Maintenance Plan Tasks
General
Control Flow
Data Flow
Event Handlers
Package Explorer
Data Flow Task

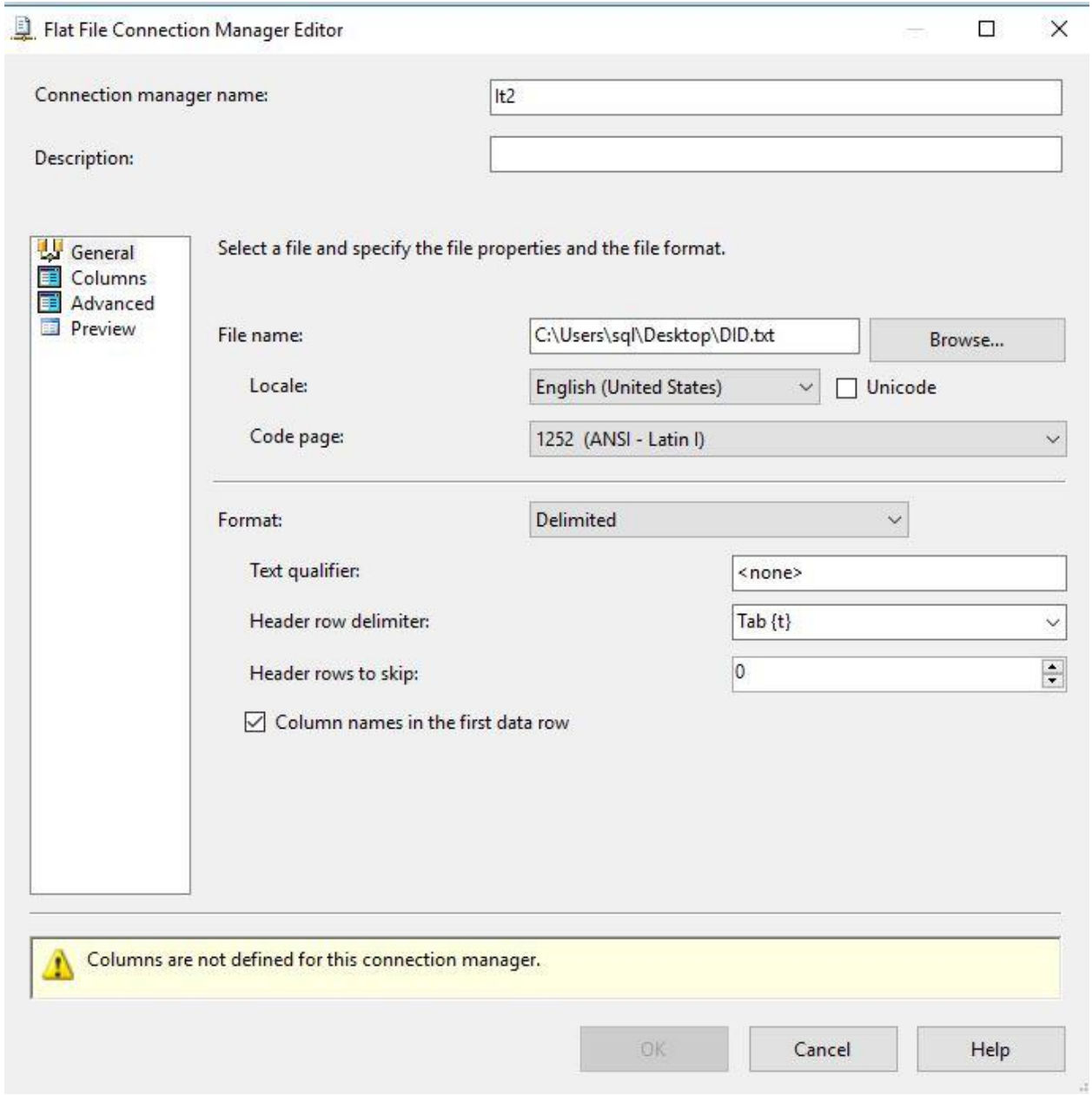


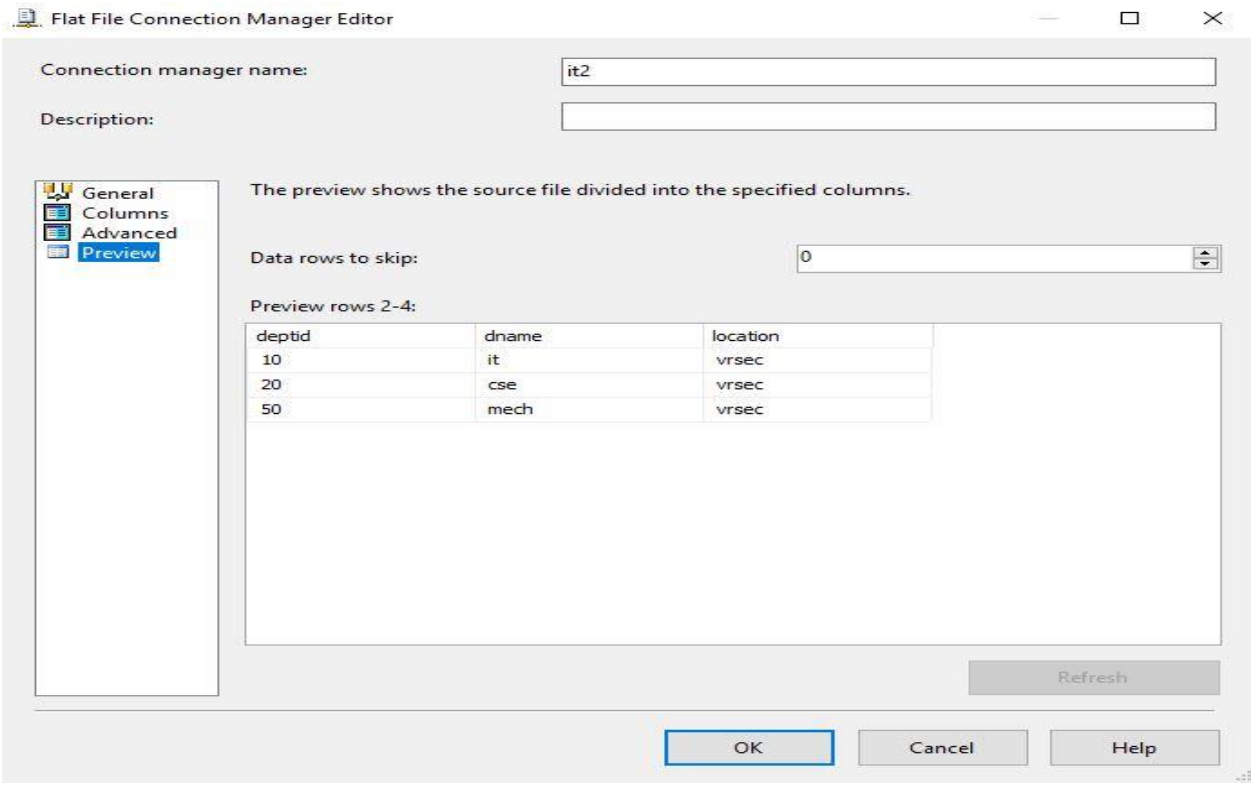
3. Now import the files in two flat file sources.

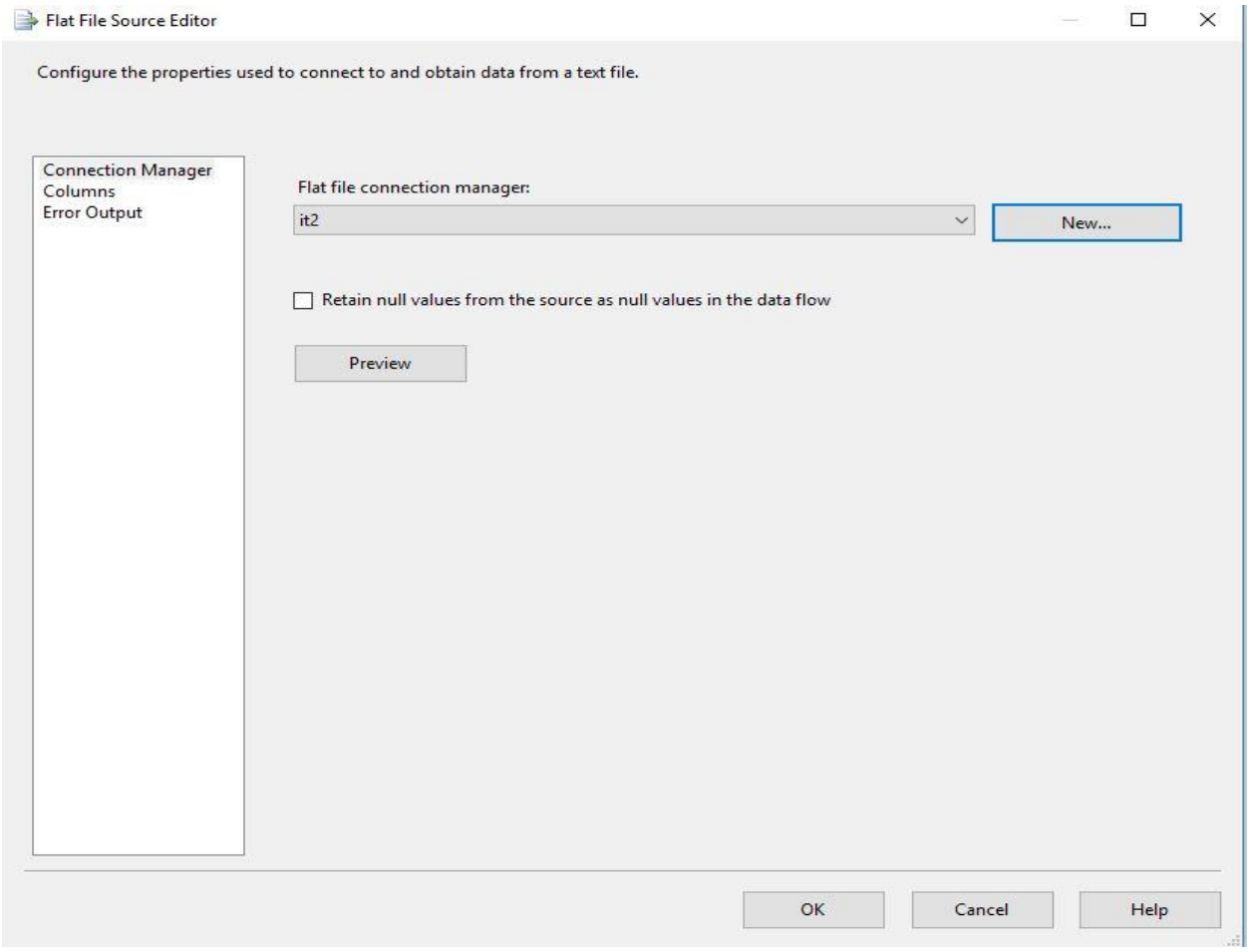



Flat File Connection Manager Editor
Connection manager name:
lt
Description:
General
Columns
Advanced
Preview
The preview shows the source file divided into the specified columns.
Data rows to skip:
0
Preview rows 2-6:
EID ENAME SALARY DEPTID
1001 RAM 15000 20
1002 RAJU 20000 10
1003 RAVI 25000 30
1004 ROBORT 50000 40
1005 RAMU 35000 10
Refresh
OK
Cancel
Help

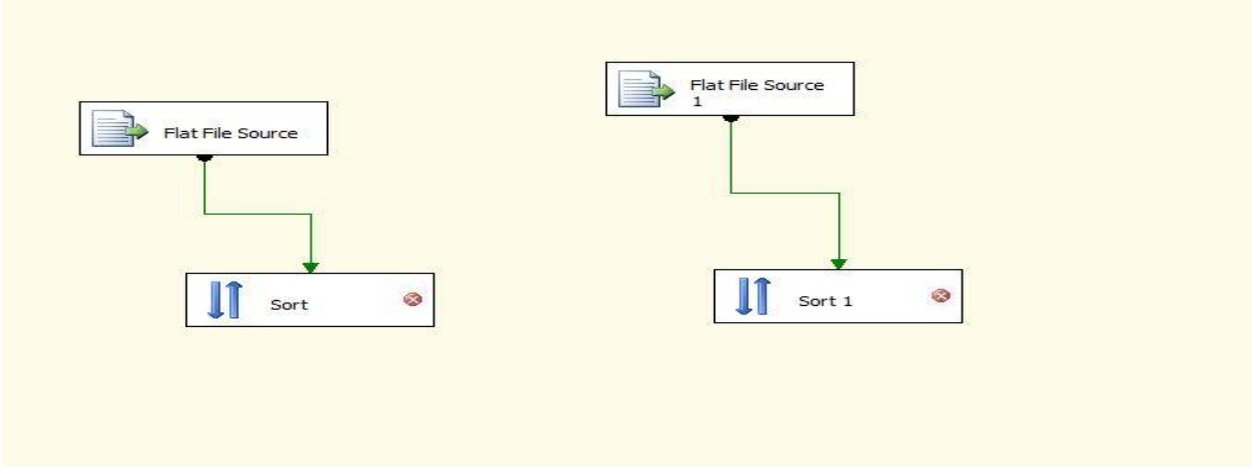

Flat File Source Editor
Configure the properties used to connect to and obtain data from a text file.
Connection Manager
Columns
Error Output
Flat file connection manager:
lt
New...
Retain null values from the source as null values in the data flow
Preview
OK
Cancel
Help

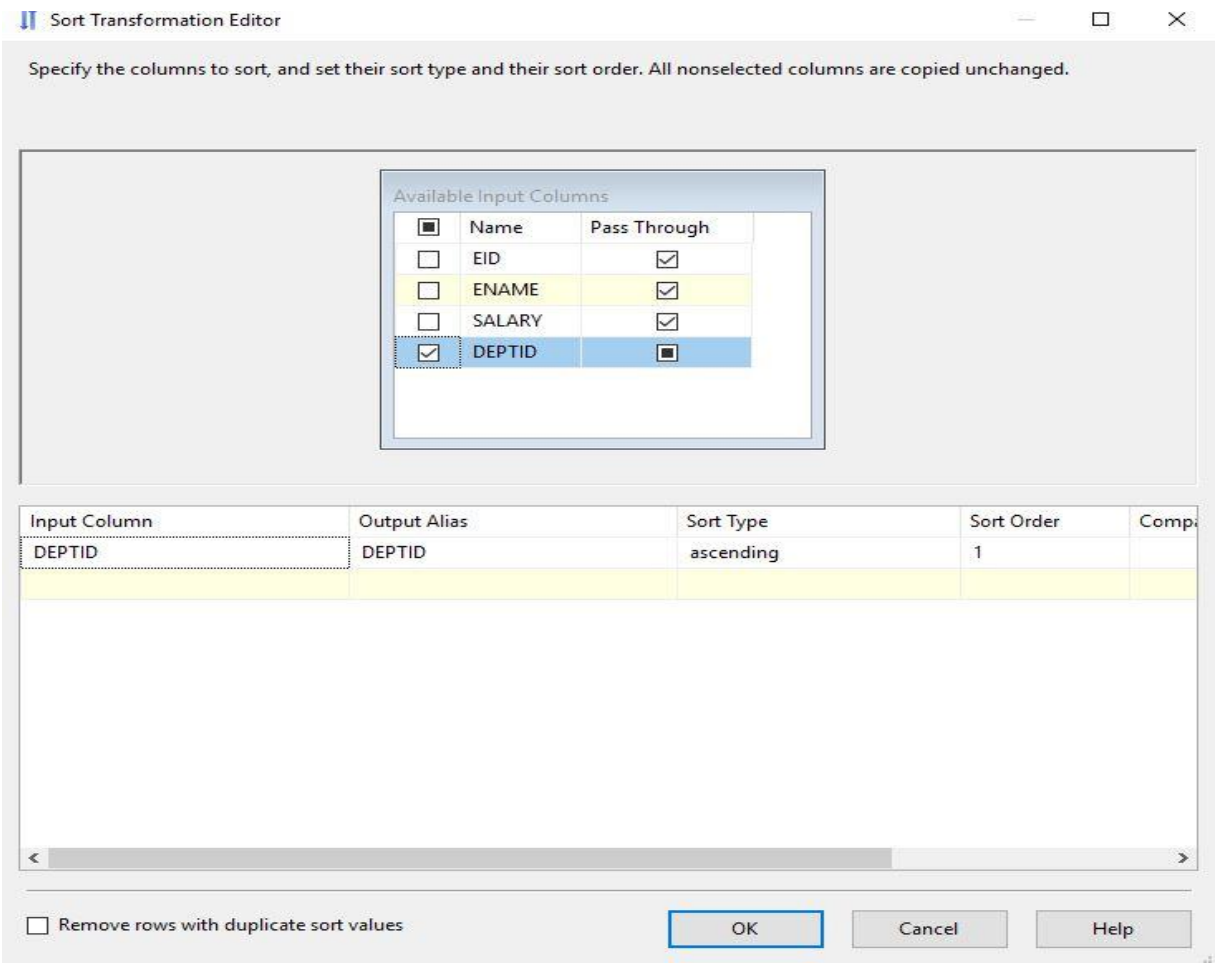

Flat File Connection Manager Editor
Connection manager name:
lt2
Description:
General
Columns
Advanced
Preview
Select a file and specify the file properties and the file format.
File name:
C:\Users\sql\Desktop\DID.txt
Browse...
Locale:
English (United States)
Unicode
Code page:
1252 (ANSI - Latin I)
Format:
Delimited
Text qualifier:
<none>
Header row delimiter:
Tab {t}
Header rows to skip:
0
Column names in the first data row
Columns are not defined for this connection manager.
OK
Cancel
Help

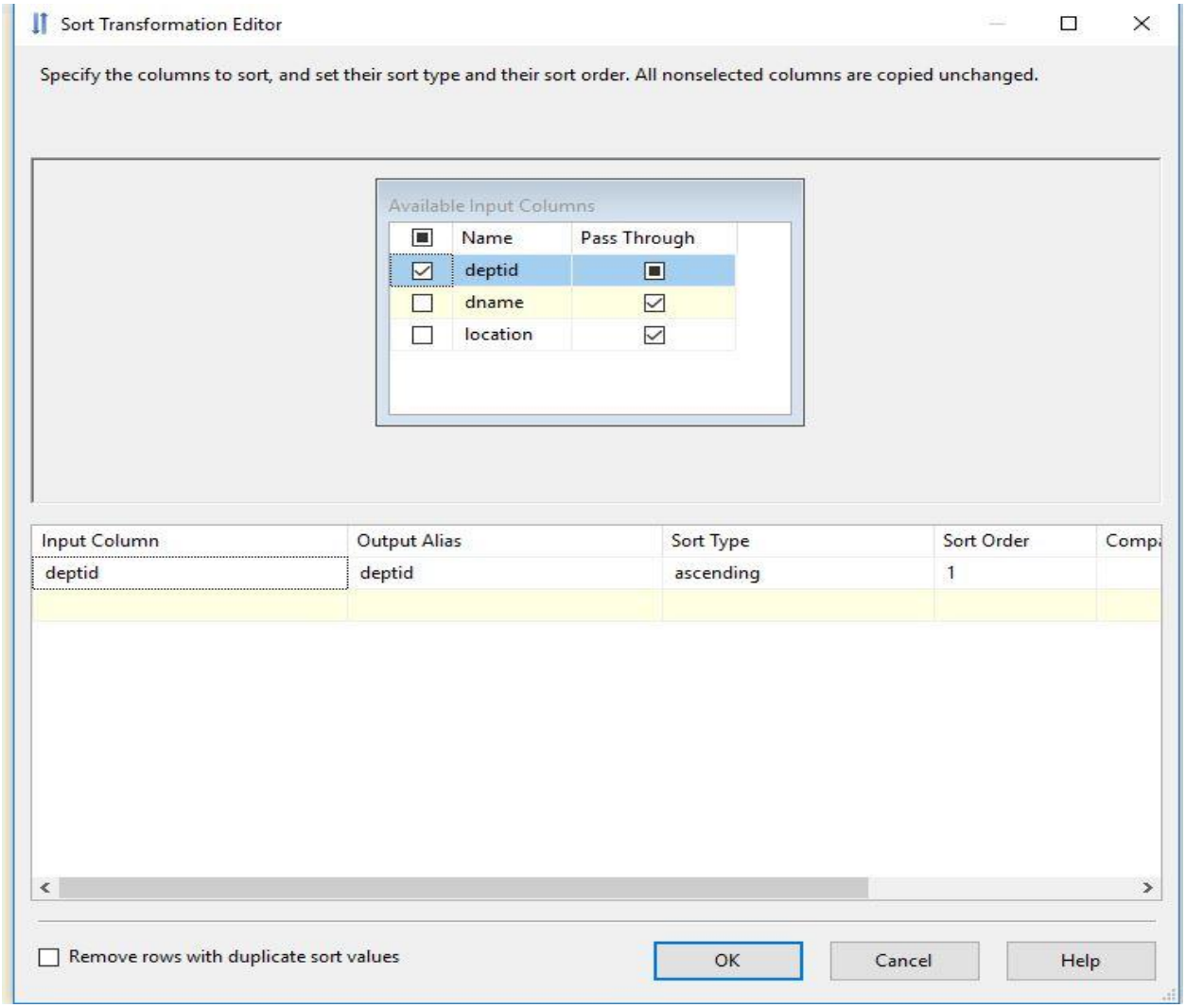

Flat File Connection Manager Editor
Connection manager name:
it2
Description:
General
Columns
Advanced
Preview
The preview shows the source file divided into the specified columns.
Data rows to skip:
0
Preview rows 2-4:
deptid
dname
location
10
it
vrsec
20
cse
vrsec
50
mech
vrsec
Refresh
OK
Cancel
Help

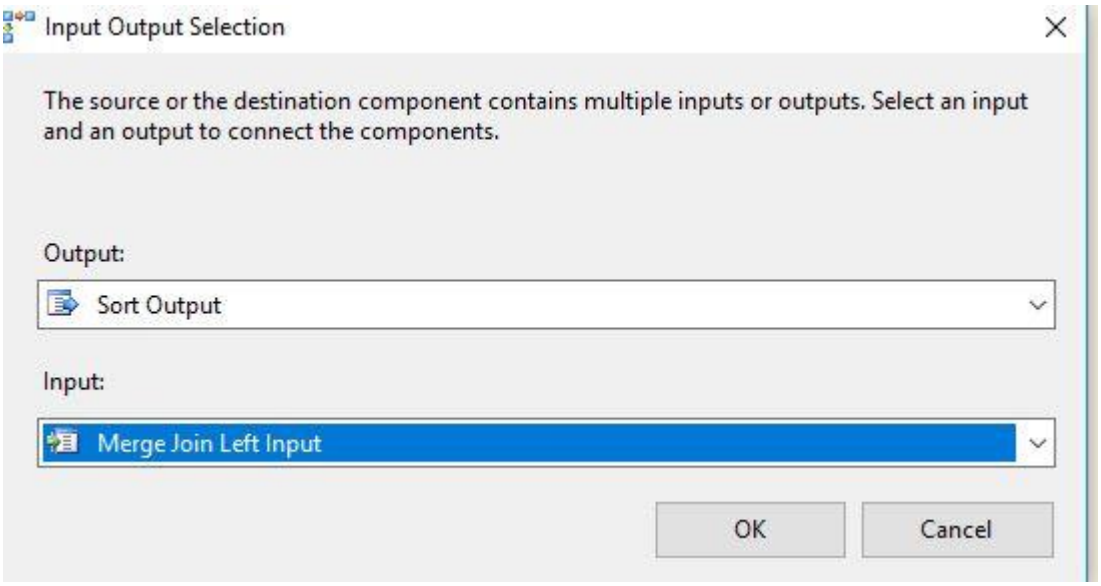

Flat File Source Editor
Configure the properties used to connect to and obtain data from a text file.
Connection Manager
Columns
Error Output
Flat file connection manager:
it2
New...
Retain null values from the source as null values in the data flow
Preview
OK
Cancel
Help

4.Now drag two sorts and join them with the flat file sources.And after double clicking on them, we have to select the attribute, we are supposed to link.

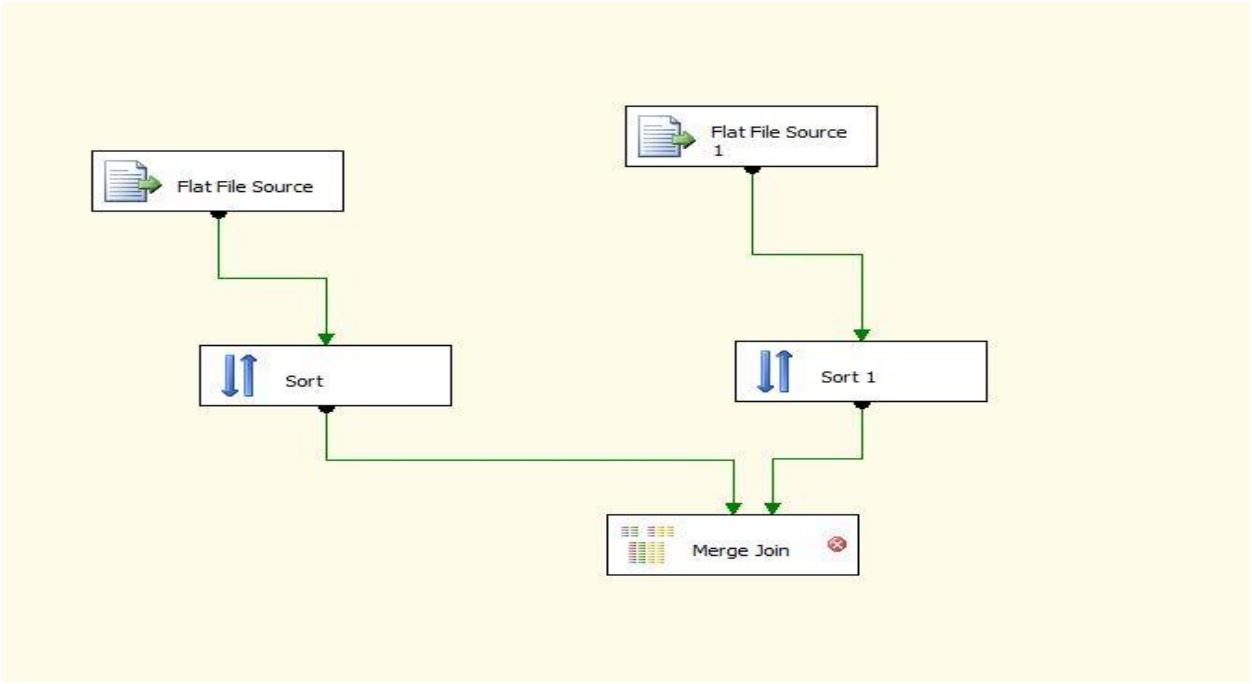


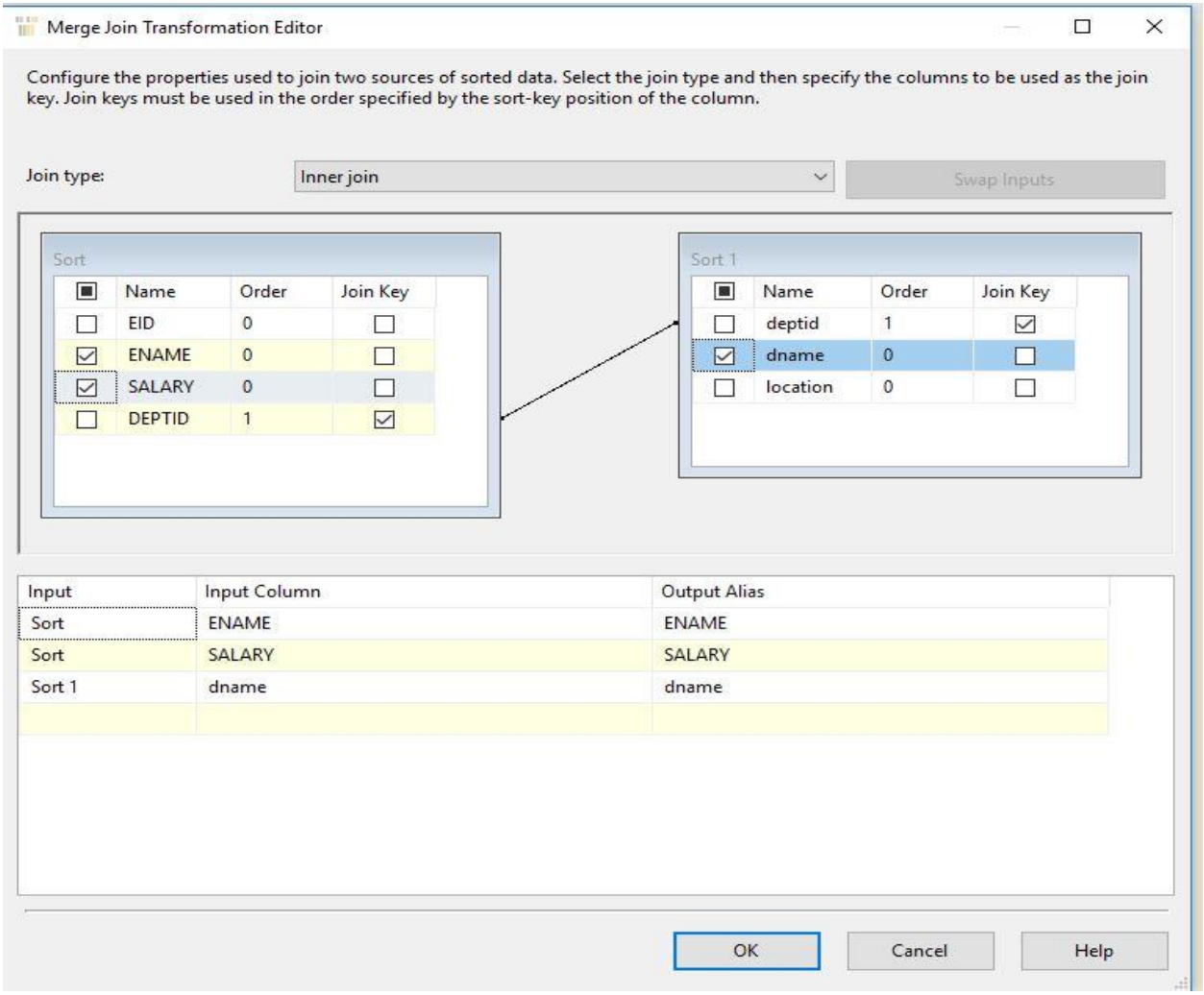




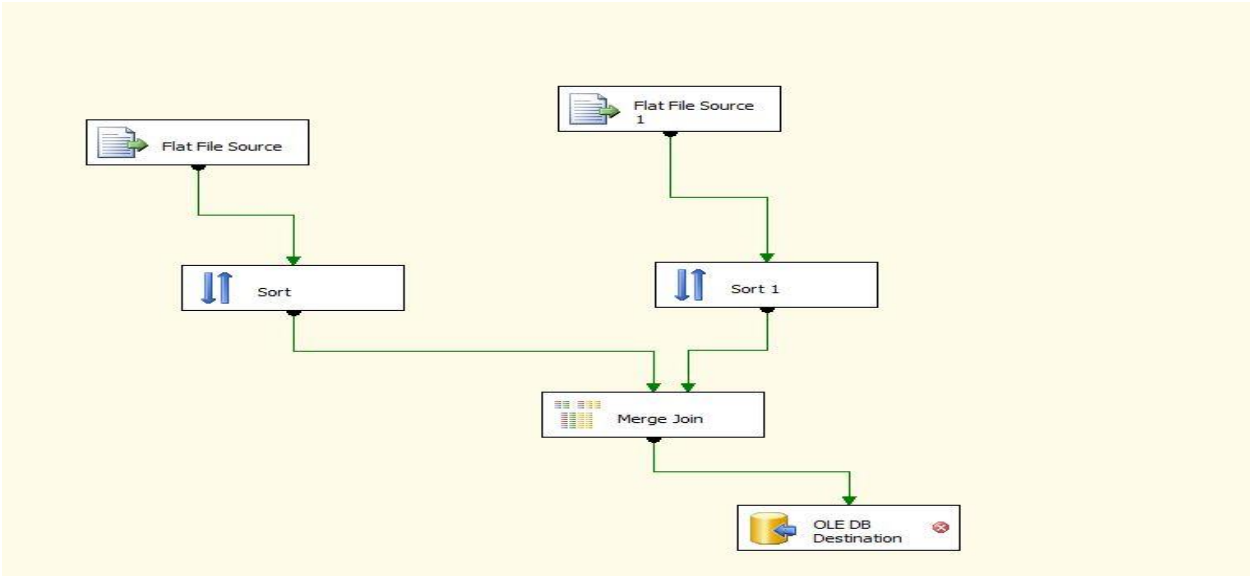


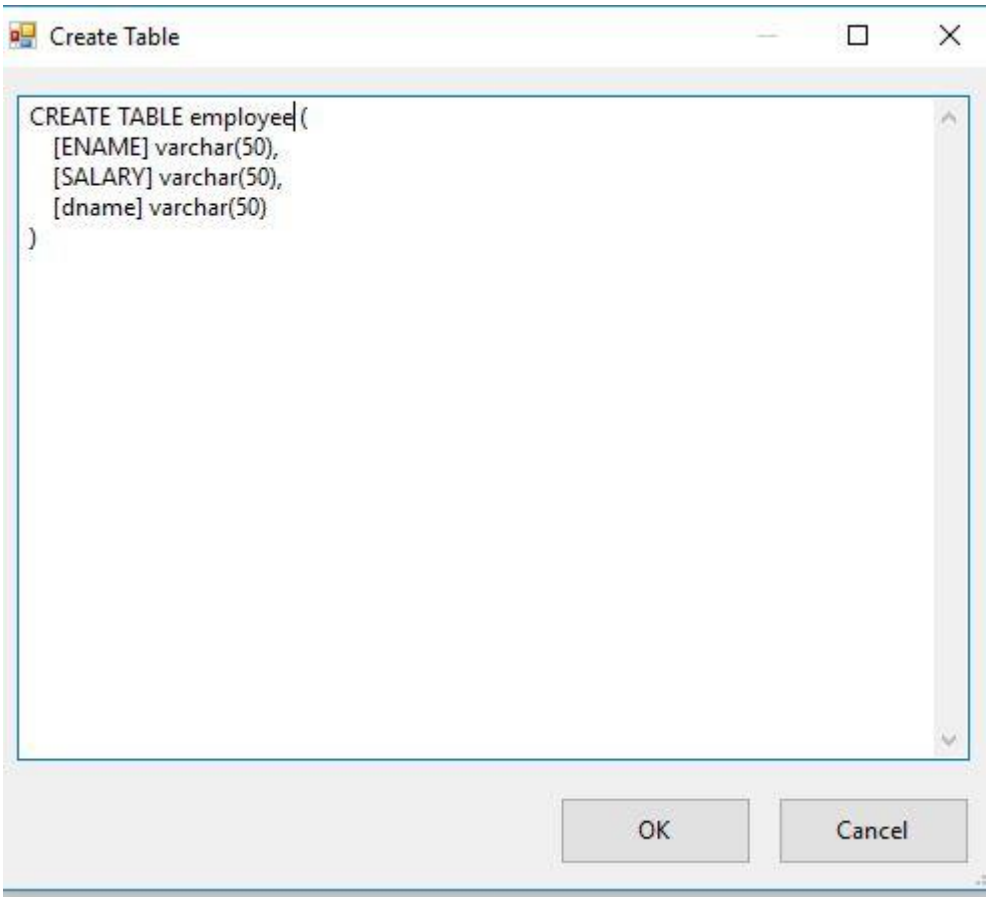
5.Now drag the merge join and then select the column names that you want in your output and also select the type of join as per your requirements.

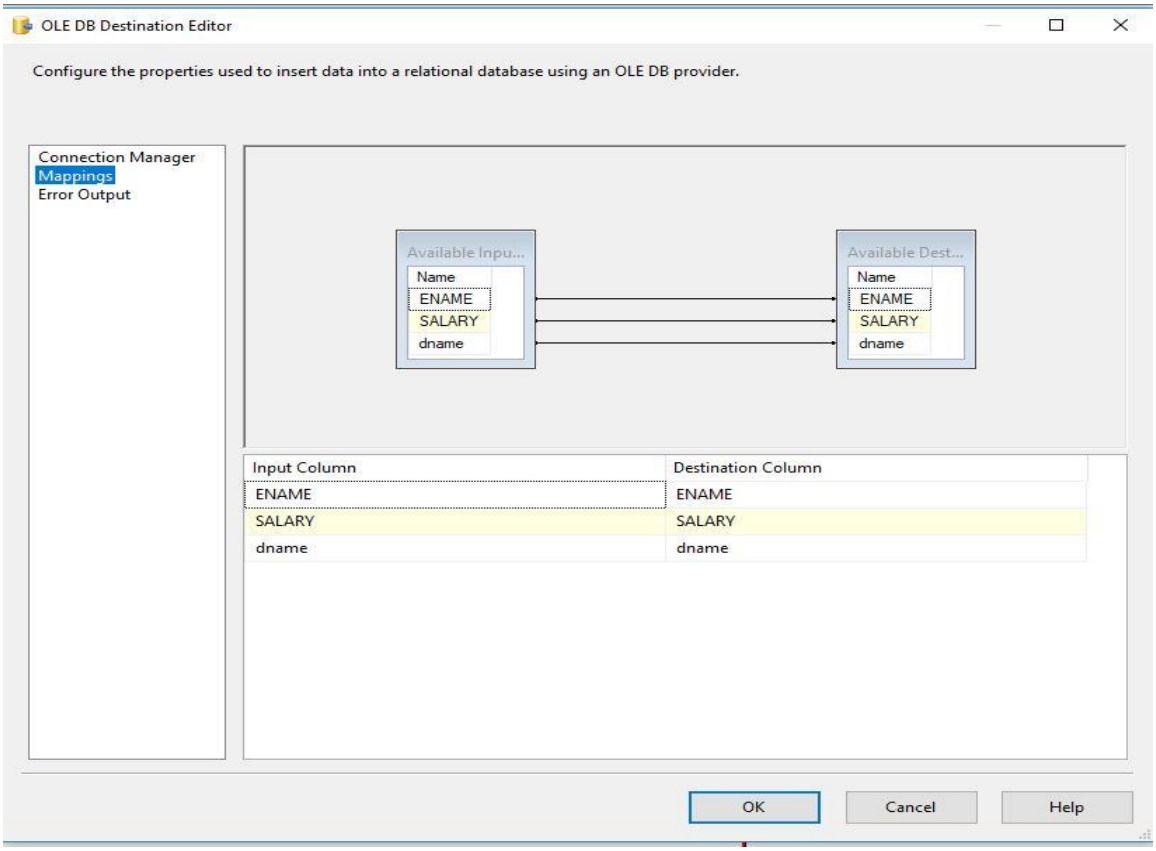


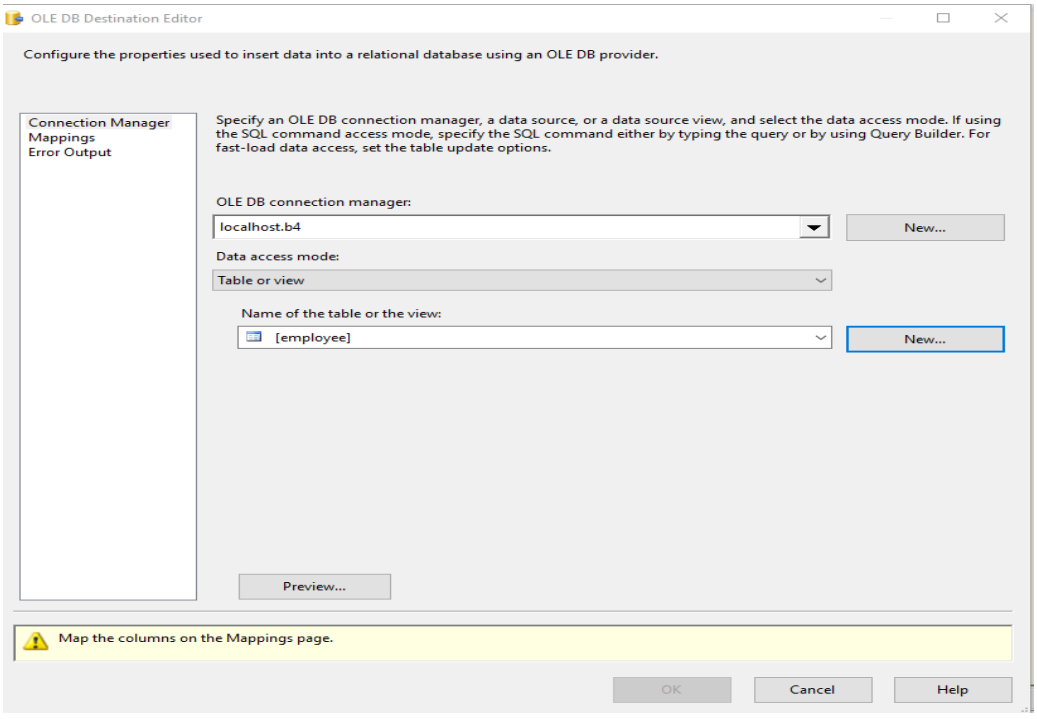


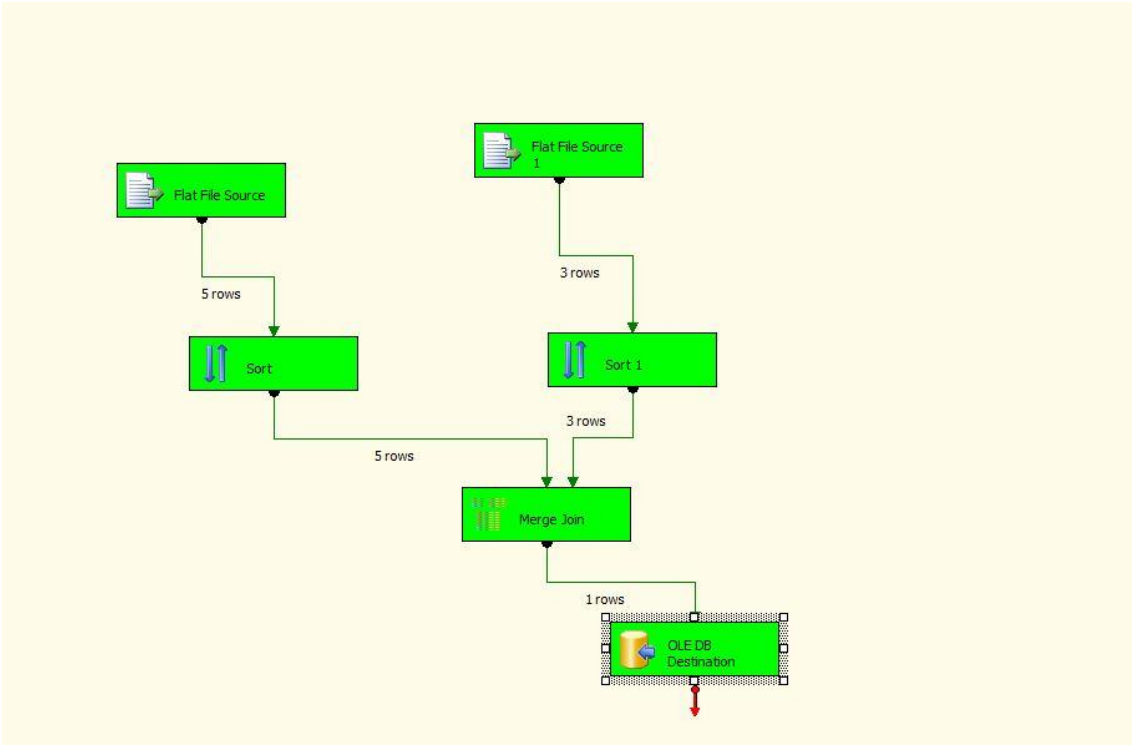
6.Now drag the OLEDB destination and then give your table a name and check your mappings and at last press F5 to debug.


Flat File Source
Flat File Source 1
Sort
Sort 1
Merge Join
OLE DB Destination


Create Table
CREATE TABLE employee (
[ENAME] varchar(50),
[SALARY] varchar(50),
[dname] varchar(50)
)
OK
Cancel

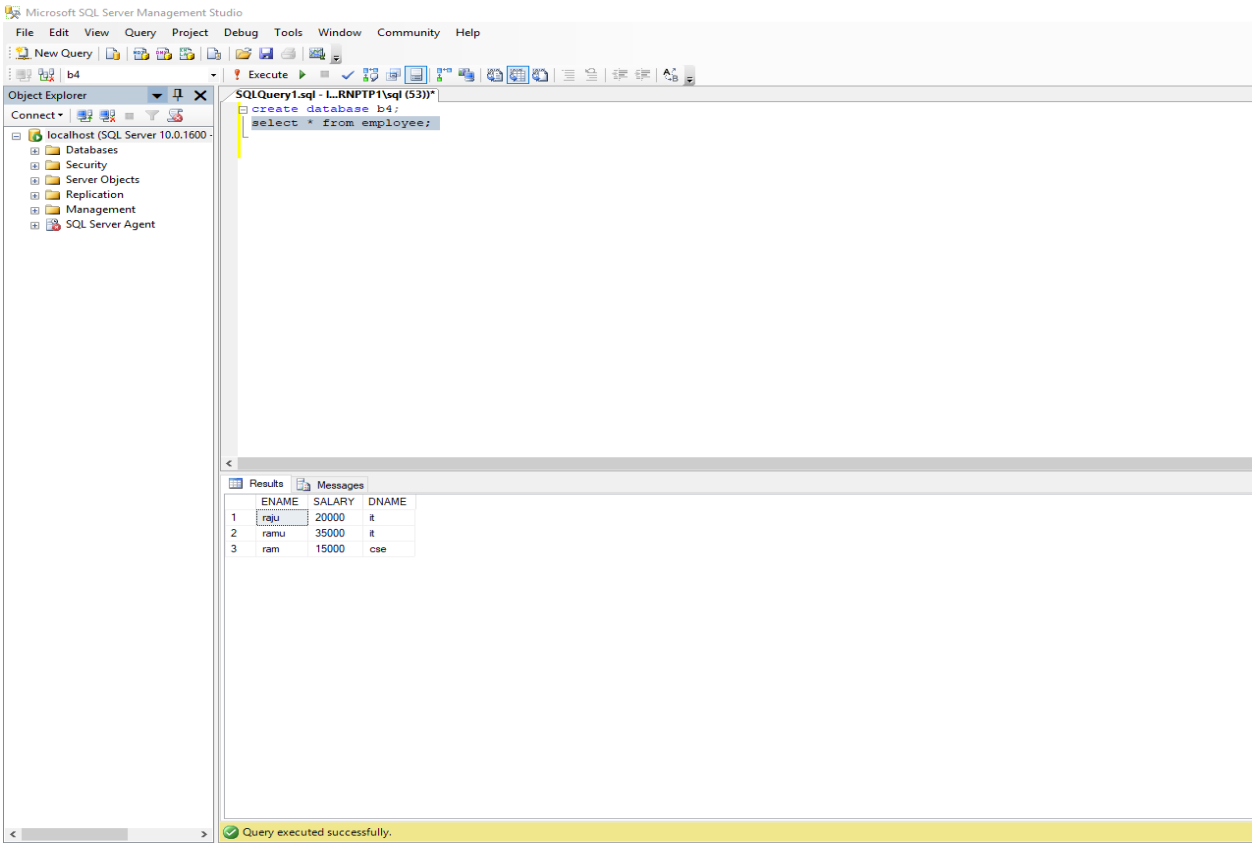

OLE DB Destination Editor
Configure the properties used to insert data into a relational database using an OLE DB provider.
Connection Manager
Mappings
Error Output
Available Inpu...
Name
ENAME
SALARY
dname
Available Dest...
Name
ENAME
SALARY
dname
Input Column
Destination Column
ENAME
ENAME
SALARY
SALARY
dname
dname
OK
Cancel
Help


OLE DB Destination Editor
Configure the properties used to insert data into a relational database using an OLE DB provider.
Connection Manager
Mappings
Error Output
Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.
OLE DB connection manager:
localhost.b4
New...
Data access mode:
Table or view
Name of the table or the view:
[employee]
New...
Preview...
Map the columns on the Mappings page.
OK
Cancel
Help

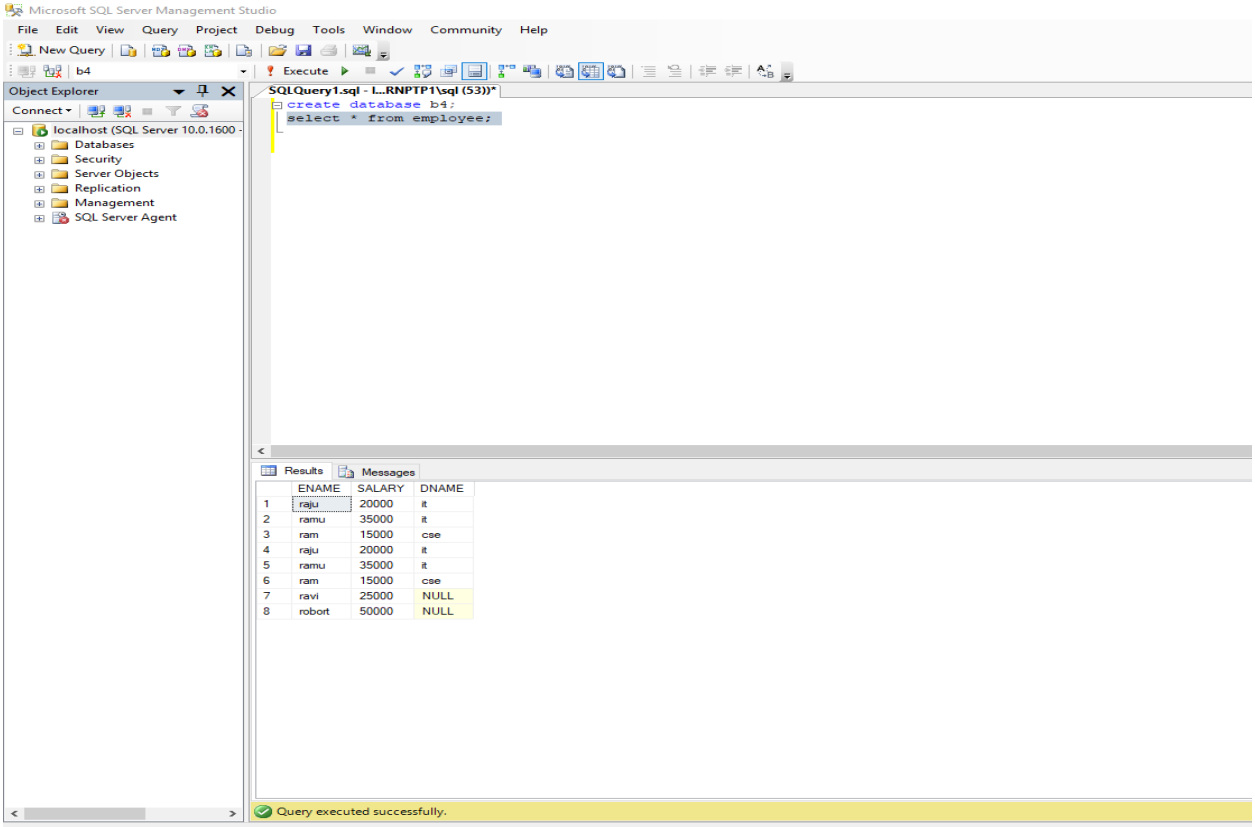


Output:

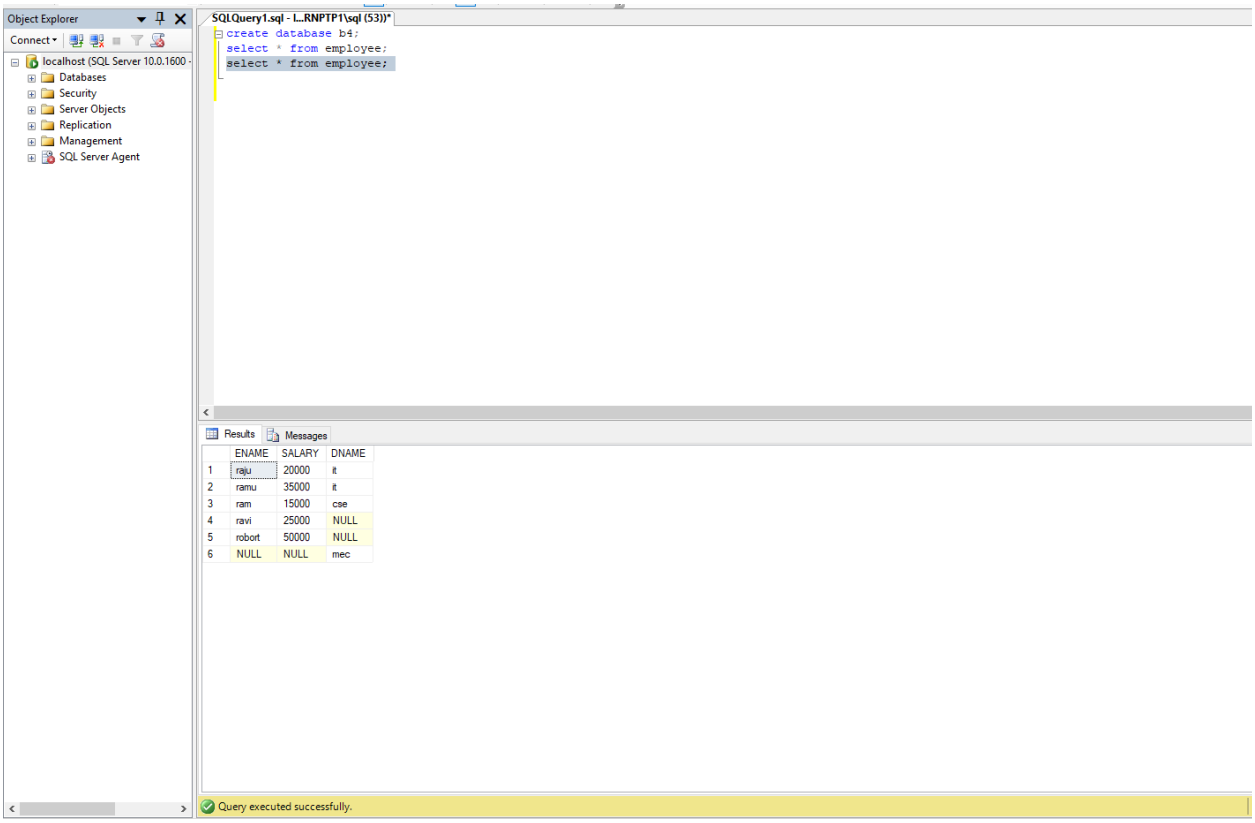
Inner join:



Left Outer Join:



Full Outer Join:



Task:4

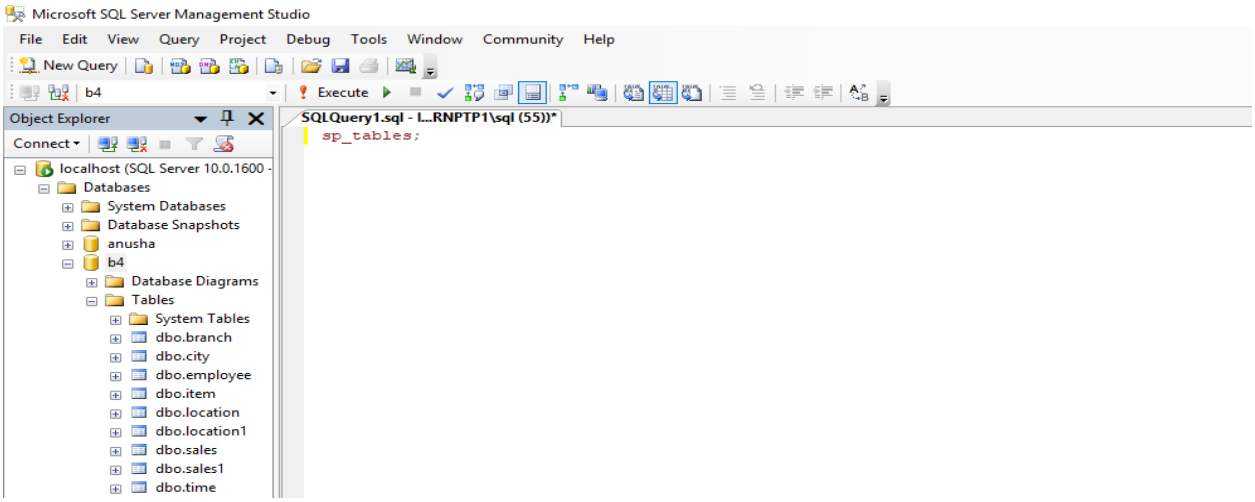
Aim:

To try to create your own database and implement the concepts of star,snowflake and fact constellation schemas in Microsoft Business Intelligence Studio(Analysis Service Project)

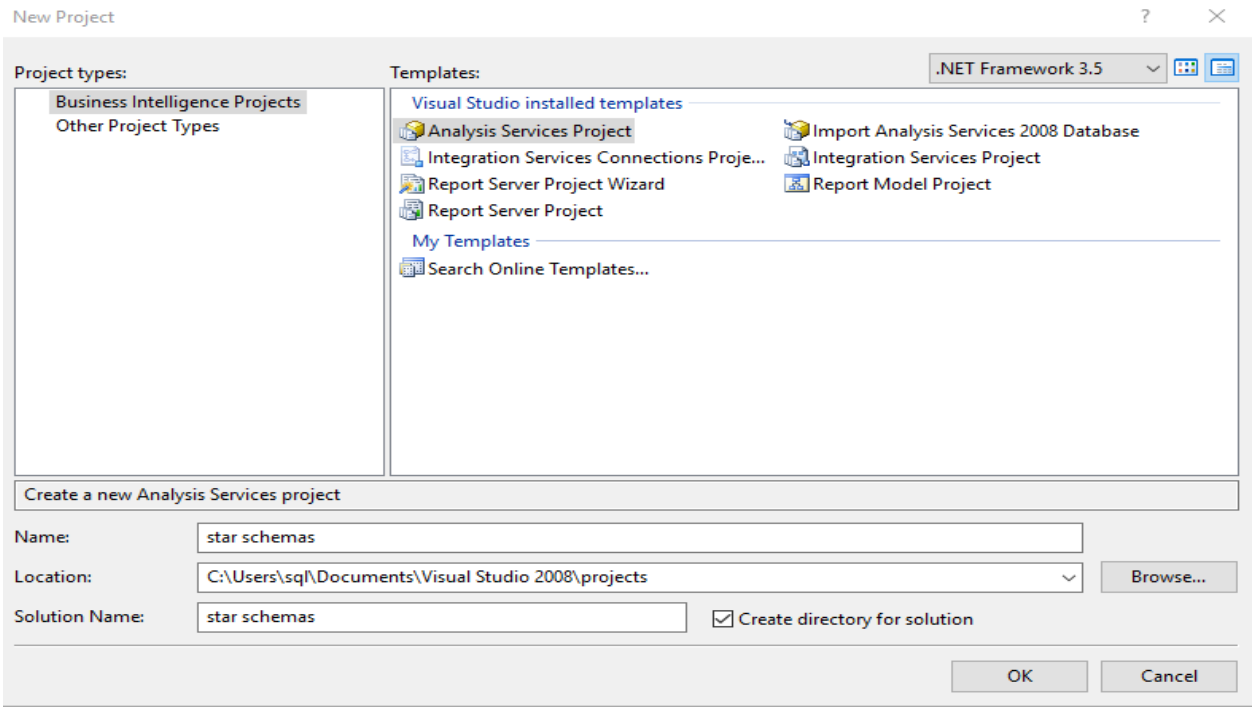
Procedure:

Star schema:

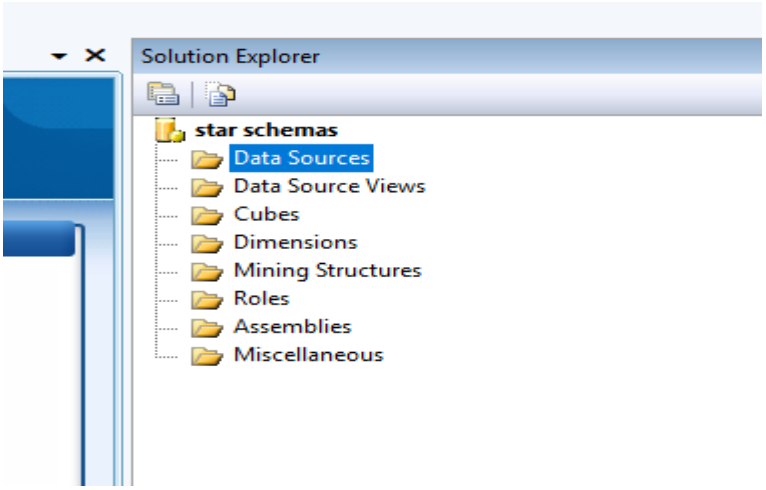
1.first create a database with required tables.

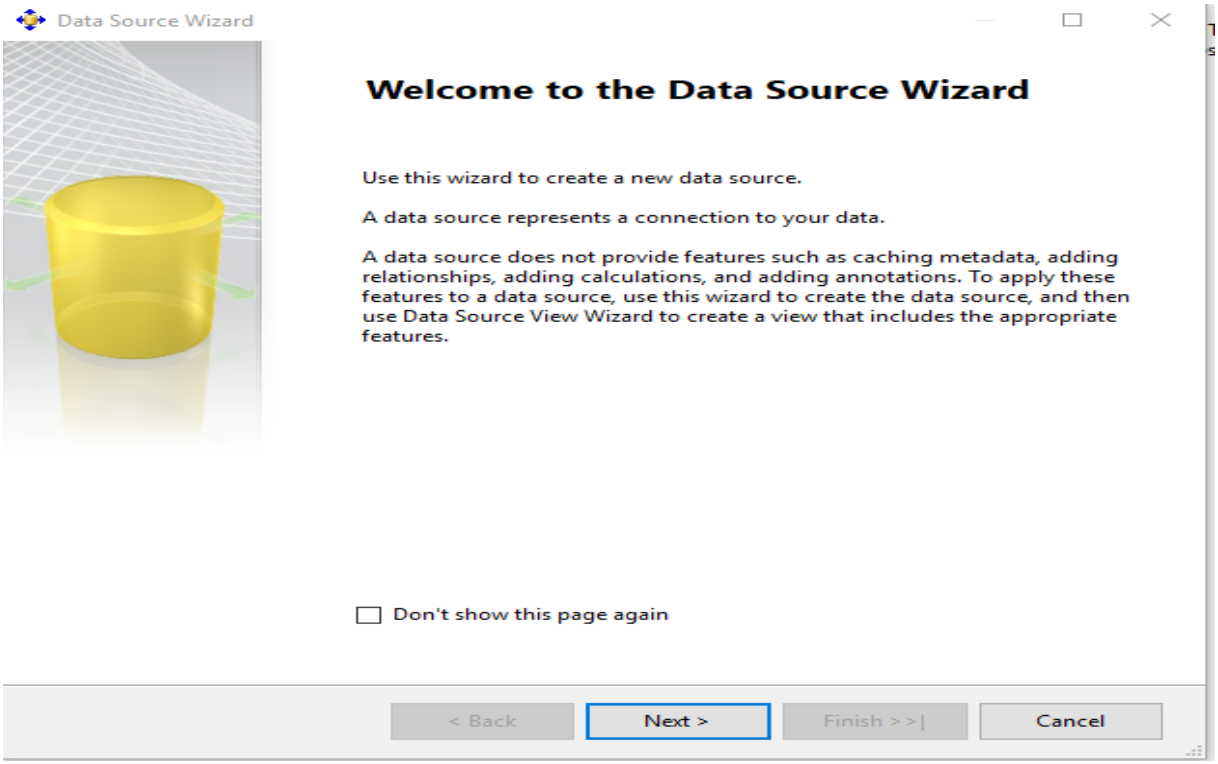


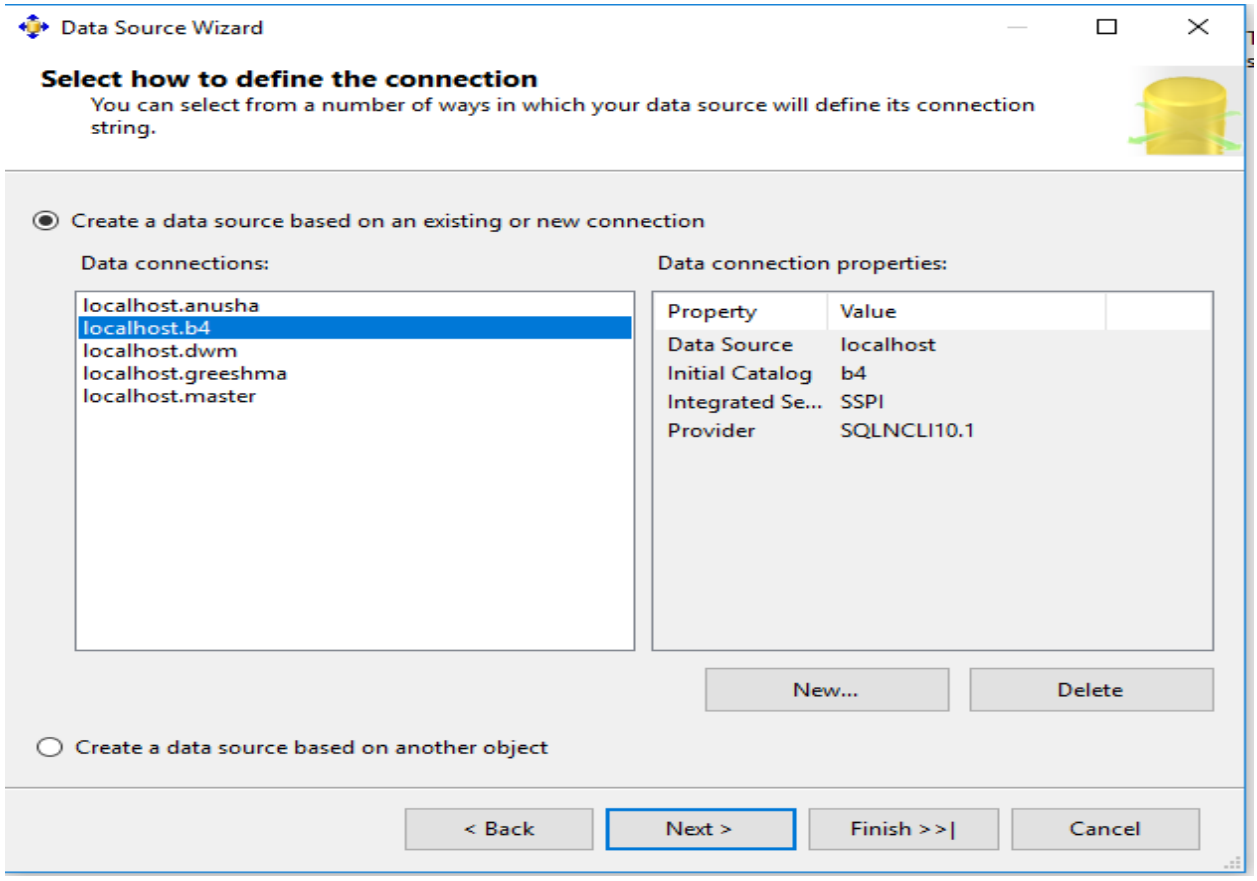
2.Now open a new project in BI studio and select analysis service project.

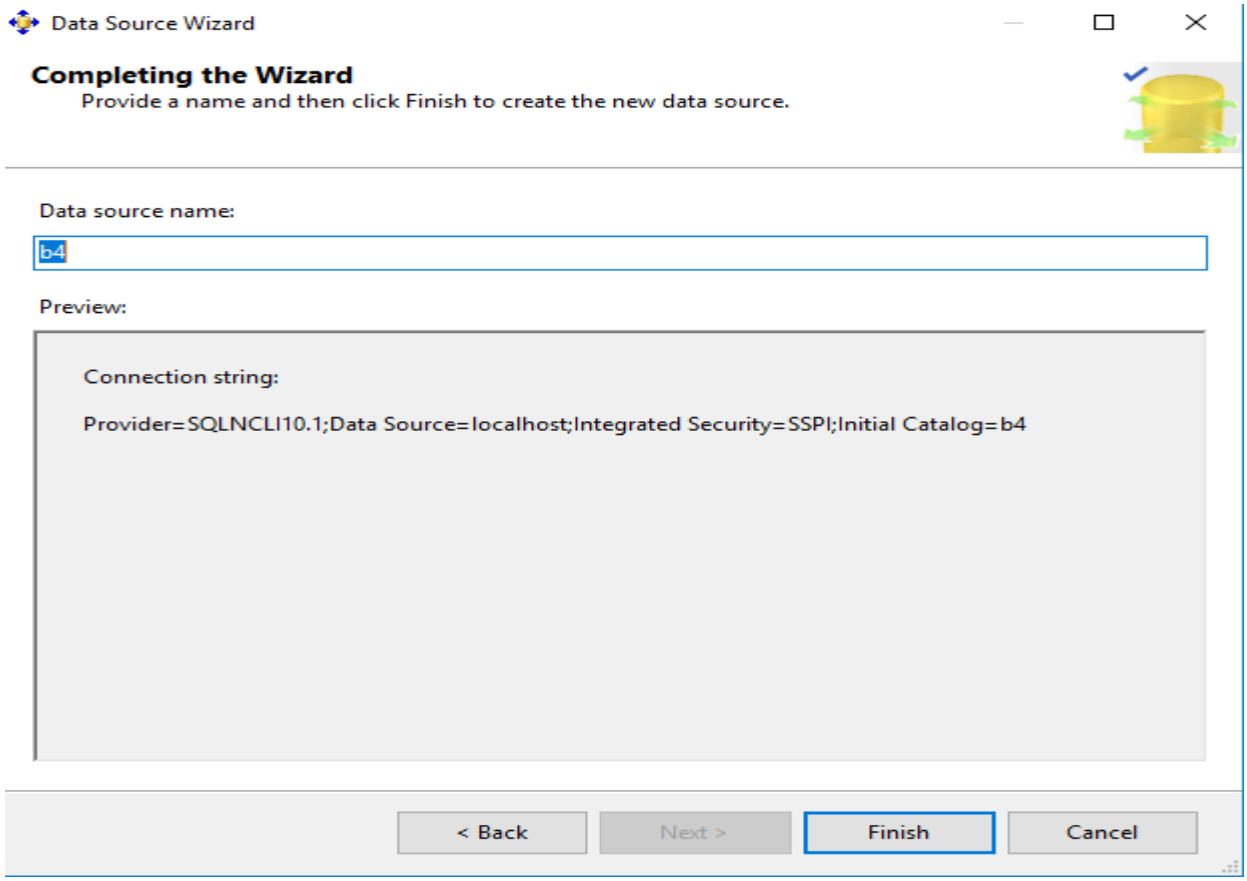


3.Select datasource and right click on it.Then click on new datasource,and then complete the datasource wizard by selecting the server name and database name on which action is to be performed.

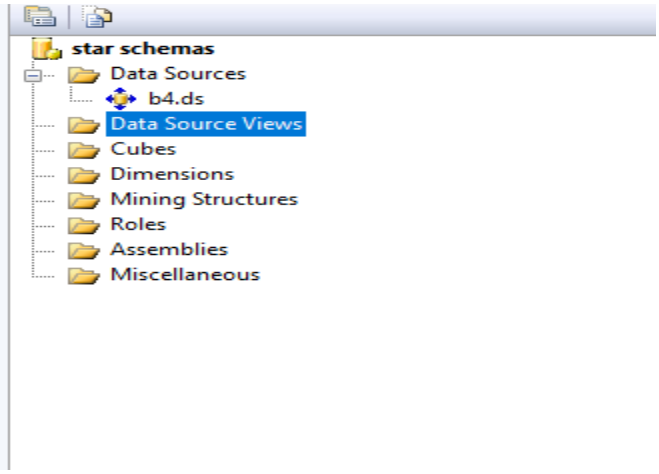


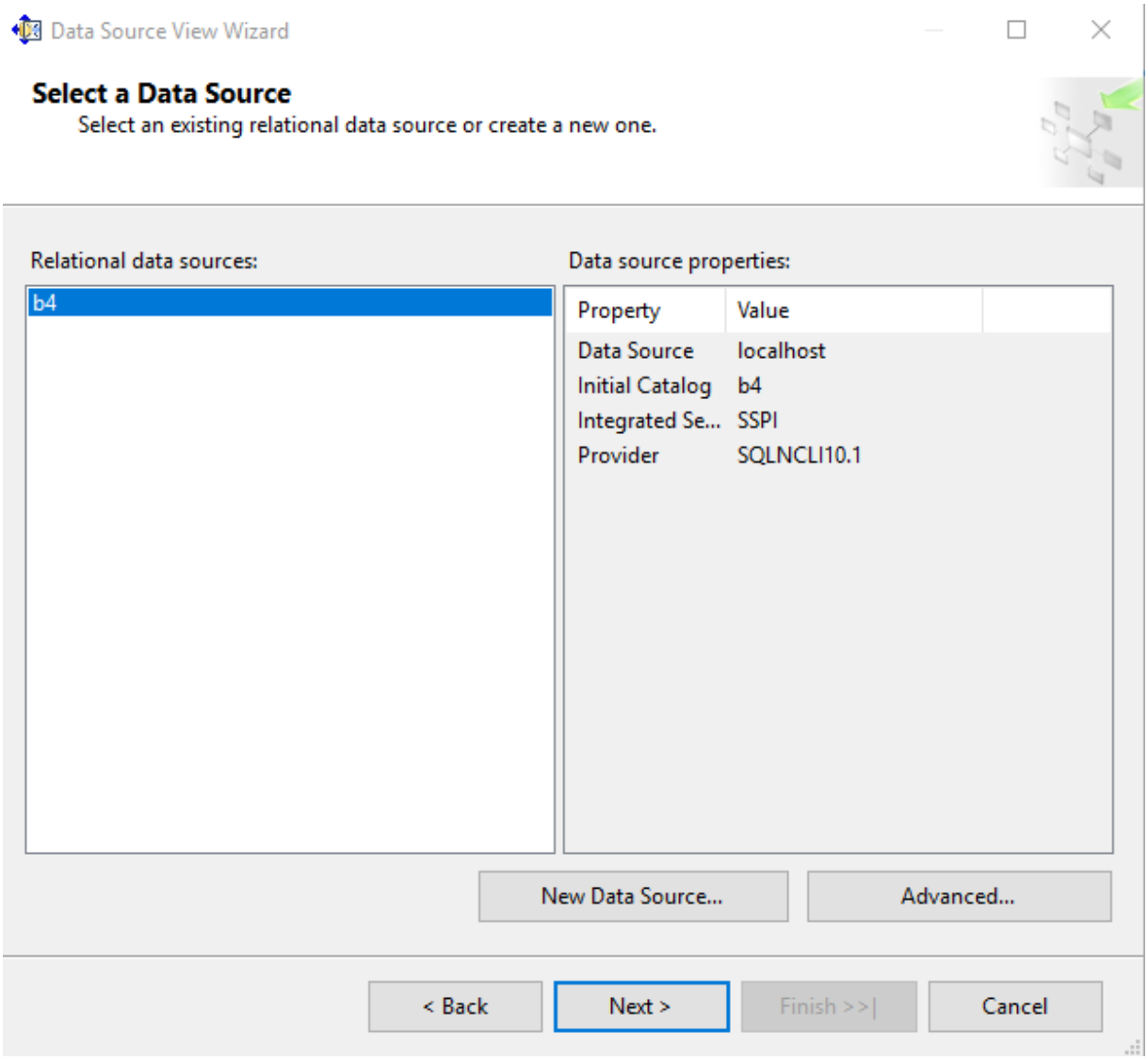


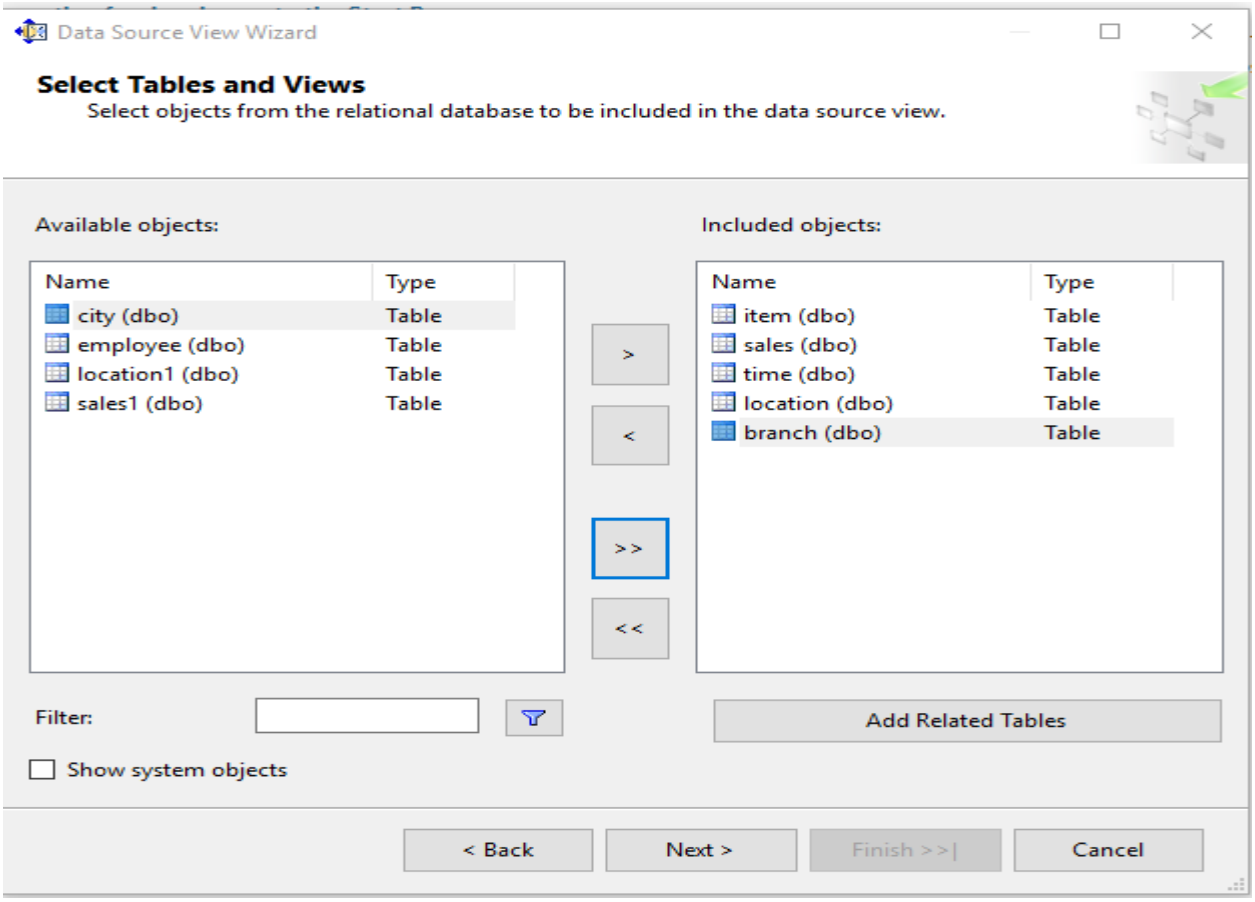

Data Source Wizard
Select how to define the connection
You can select from a number of ways in which your data source will define its connection string.
Create a data source based on an existing or new connection
Data connections:
localhost.anusha
localhost.b4
localhost.dwm
localhost.greeshma
localhost.master
Data connection properties:
Property
Value
Data Source
localhost
Initial Catalog
b4
Integrated Se...
SSPI
Provider
SQLNCLI10.1
New...
Delete
Create a data source based on another object
< Back
Next >
Finish >>|
Cancel

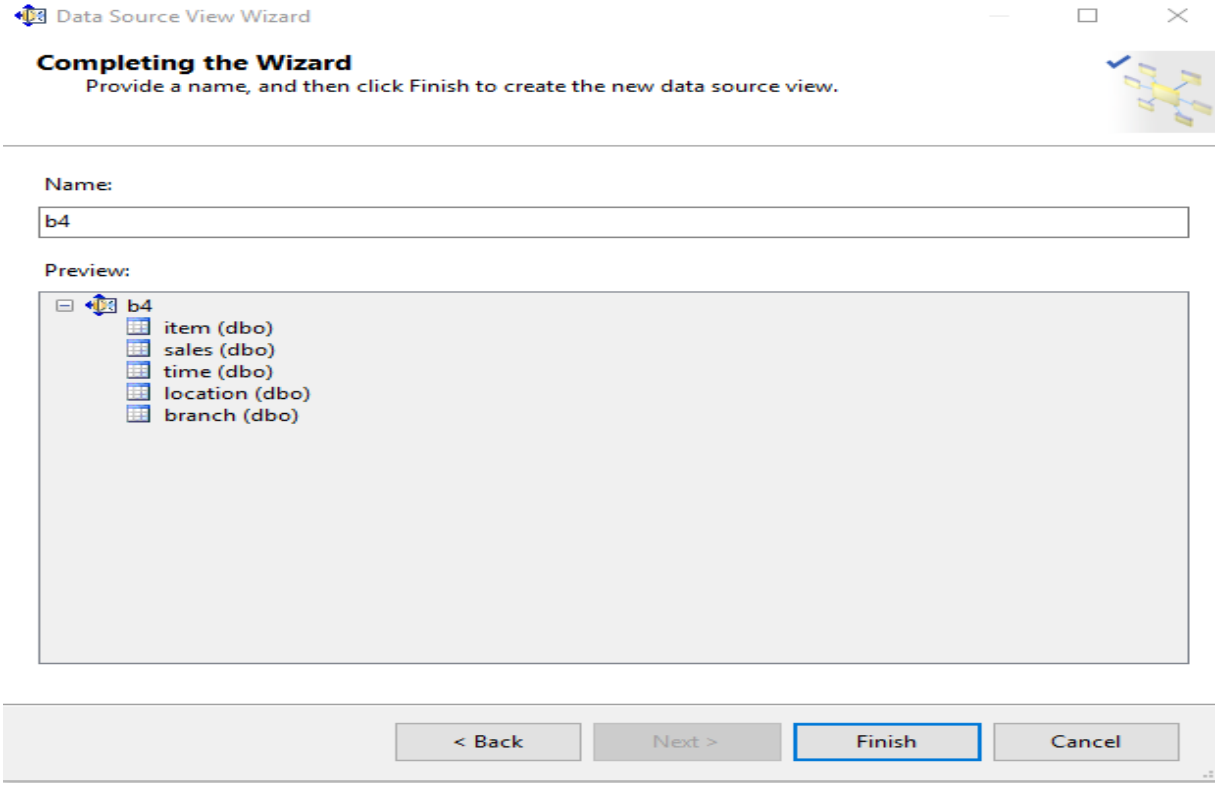

Data Source Wizard
Completing the Wizard
Provide a name and then click Finish to create the new data source.
Data source name:
b4
Preview:
Connection string:
Provider=SQLNCLI10.1;Data Source=localhost;Integrated Security=SSPI;Initial Catalog=b4
< Back
Next >
Finish
Cancel

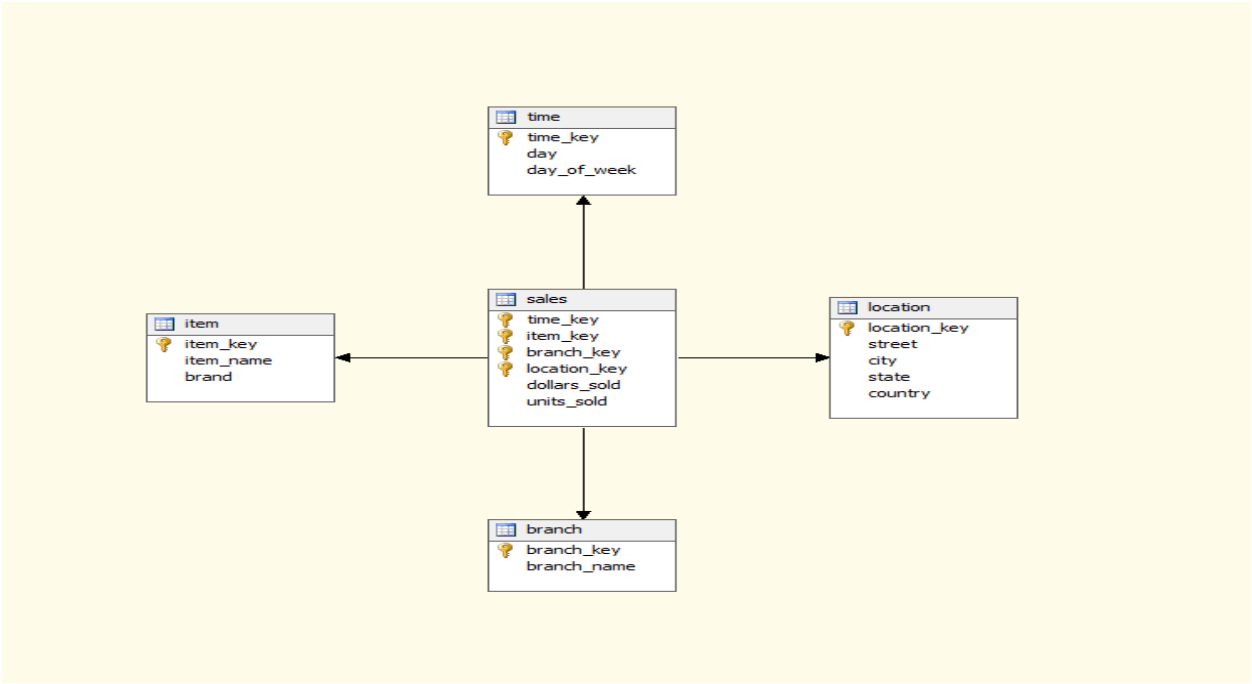
4. Select datasource view and right click on it.Then click on new datasource view,and then complete the datasource view wizard by selecting the tables used for star schema.



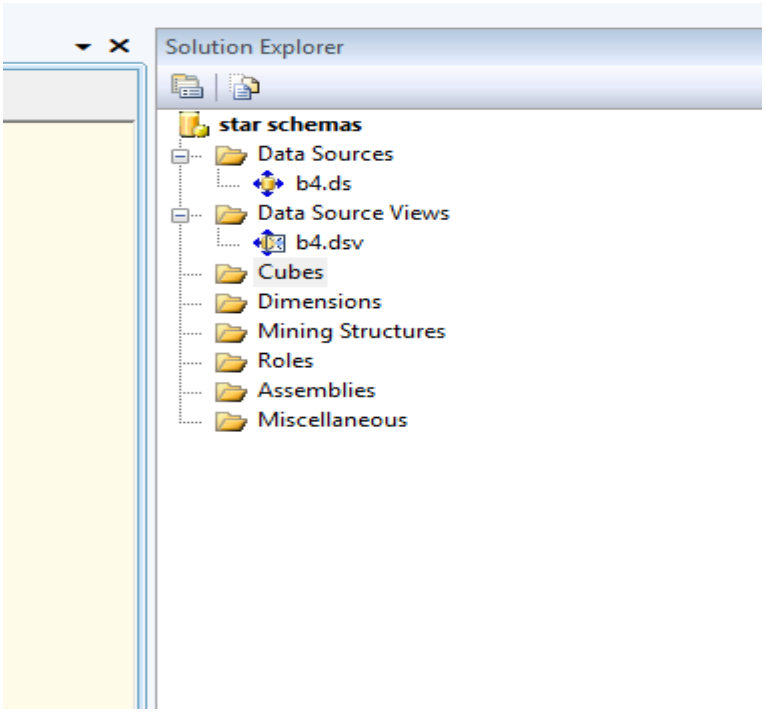


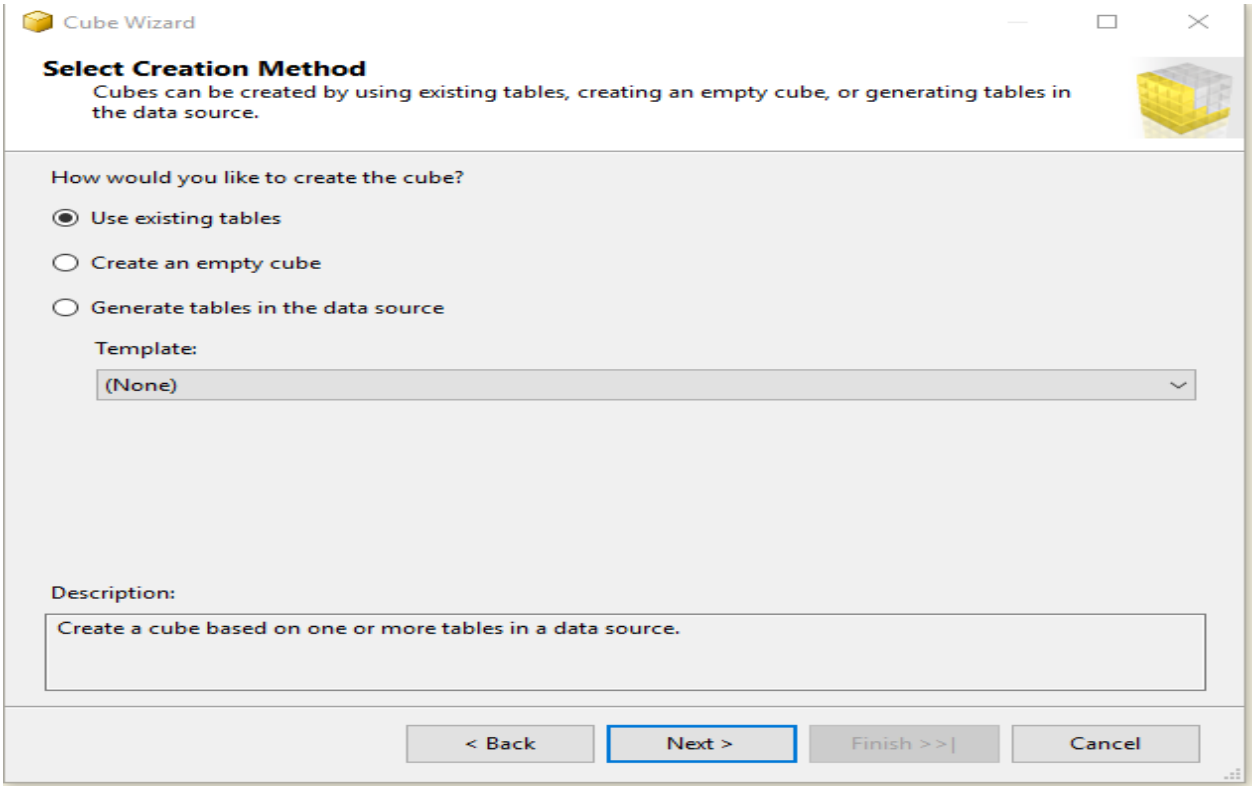

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
city (dbo)
Table
employee (dbo)
Table
location1 (dbo)
Table
sales1 (dbo)
Table
>
<
>>
<<
Included objects:
Name
Type
item (dbo)
Table
sales (dbo)
Table
time (dbo)
Table
location (dbo)
Table
branch (dbo)
Table
Filter:
Add Related Tables
Show system objects
< Back
Next >
Finish >>|
Cancel

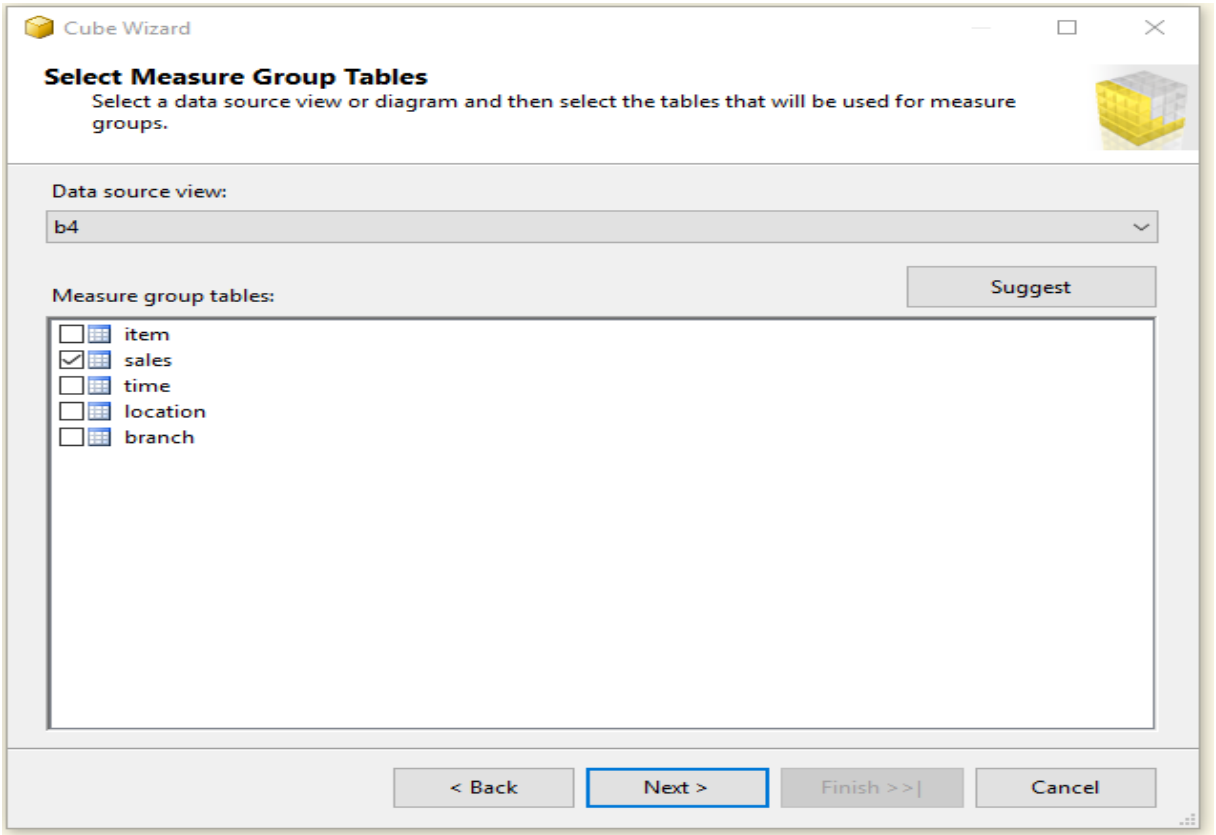

Data Source View Wizard
Completing the Wizard
Provide a name, and then click Finish to create the new data source view.
Name:
b4
Preview:
b4
item (dbo)
sales (dbo)
time (dbo)
location (dbo)
branch (dbo)
< Back
Next >
Finish
Cancel

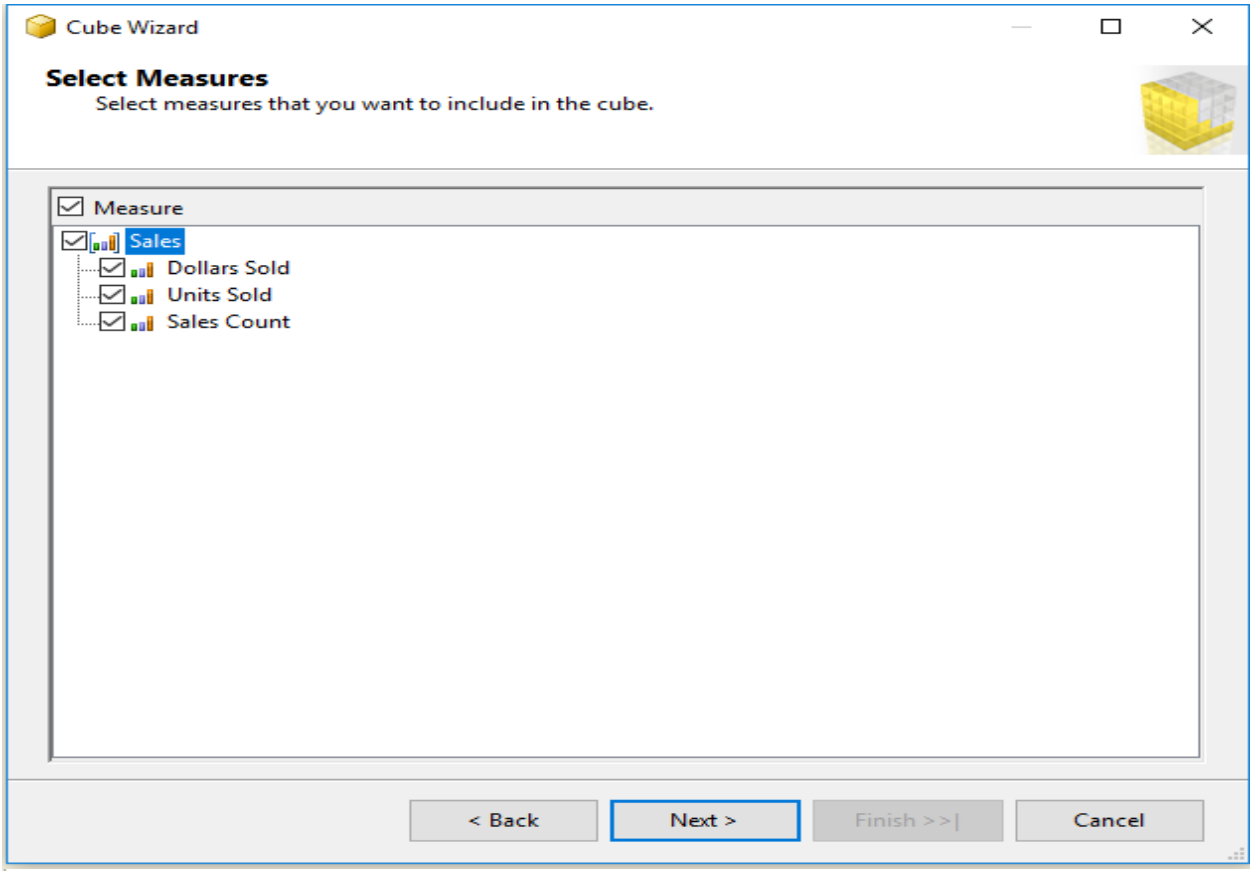


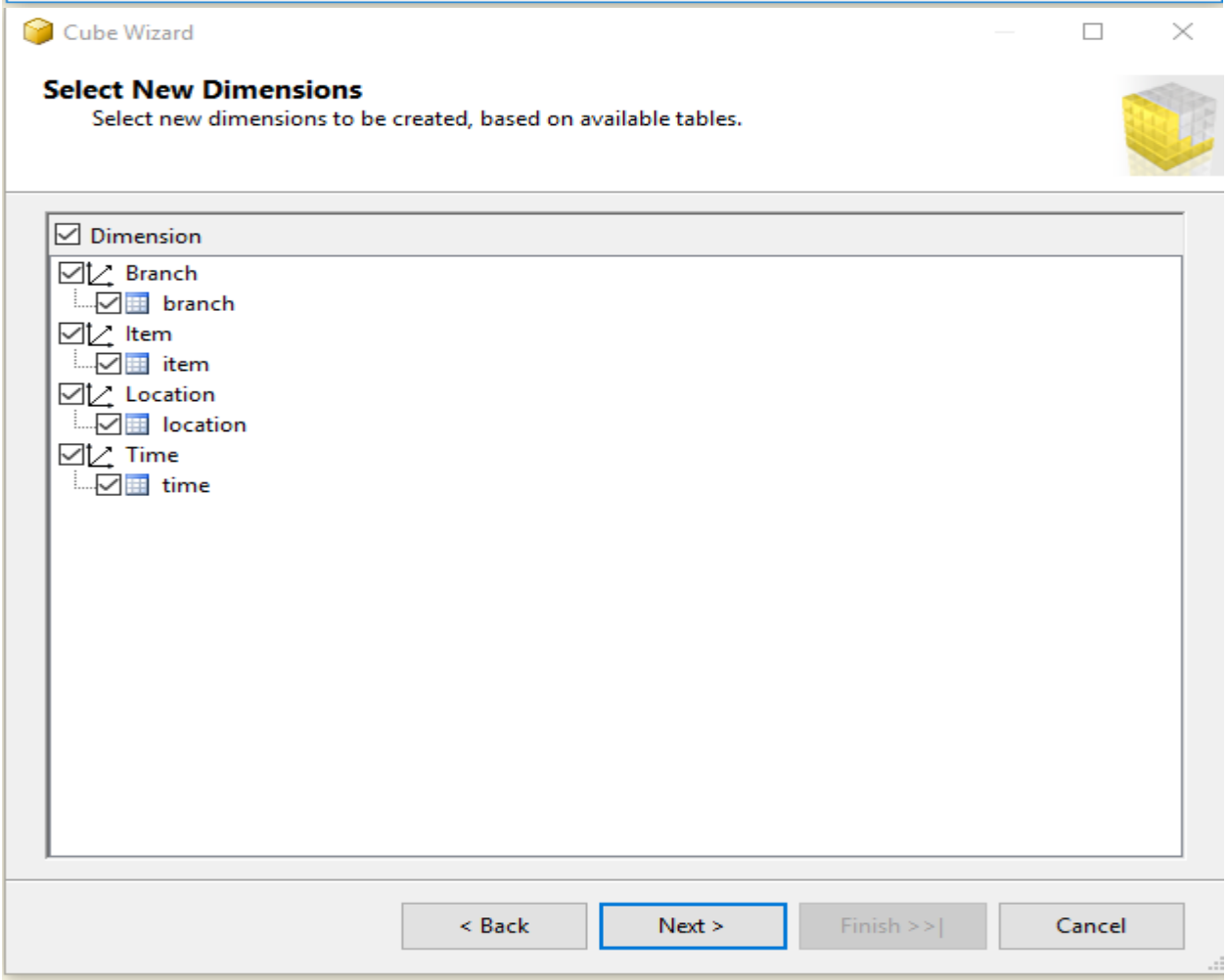
5. Select cubes and right click on it.Then click on new cube,and then complete the cube wizard by selecting the fact tables used for star schema.

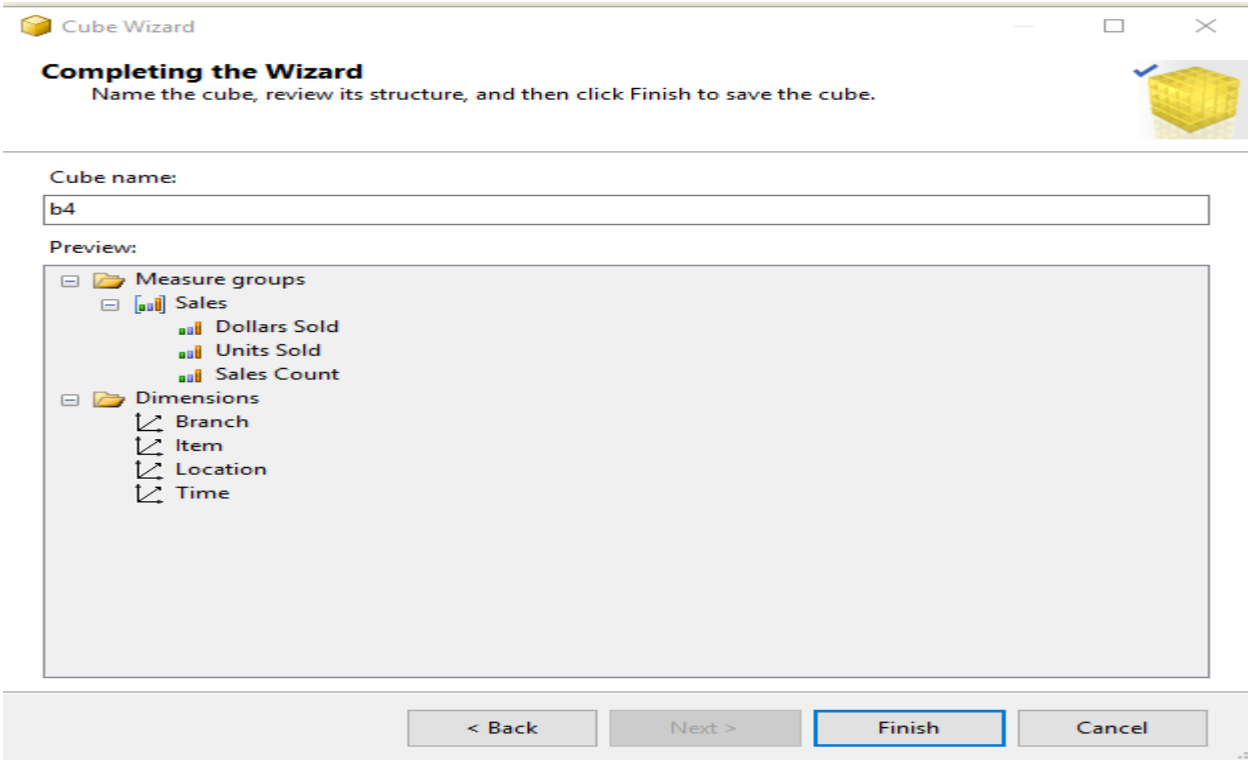



Cube Wizard
Select Creation Method
Cubes can be created by using existing tables, creating an empty cube, or generating tables in the data source.
How would you like to create the cube?
Use existing tables
Create an empty cube
Generate tables in the data source
Template:
(None)
Description:
Create a cube based on one or more tables in a data source.
< Back
Next >
Finish >>|
Cancel

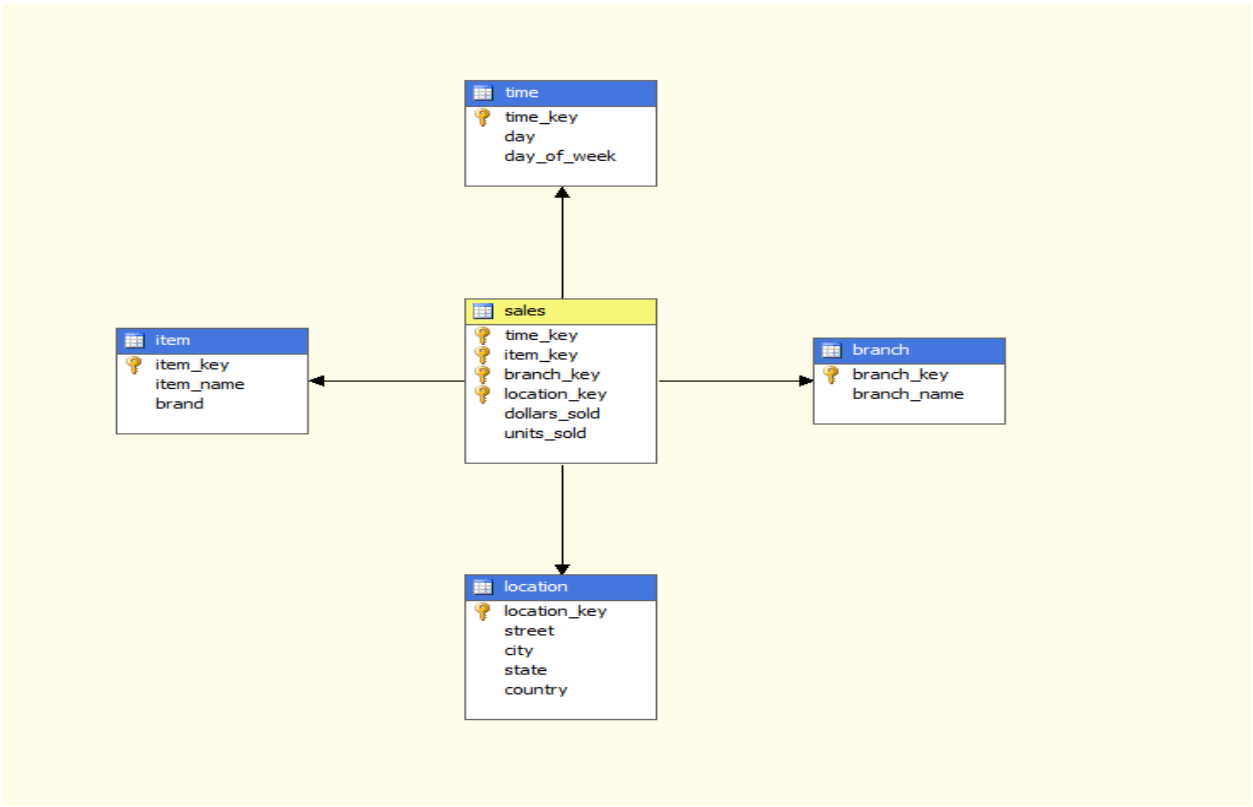

Cube Wizard
Select Measure Group Tables
Select a data source view or diagram and then select the tables that will be used for measure groups.
Data source view:
b4
Measure group tables:
Suggest
item
sales
time
location
branch
< Back
Next >
Finish >>|
Cancel


Cube Wizard
Select Measures
Select measures that you want to include in the cube.
Measure
Sales
Dollars Sold
Units Sold
Sales Count
< Back
Next >
Finish >>|
Cancel


Cube Wizard
Select New Dimensions
Select new dimensions to be created, based on available tables.
Dimension
Branch
branch
Item
item
Location
location
Time
time
< Back
Next >
Finish >>|
Cancel

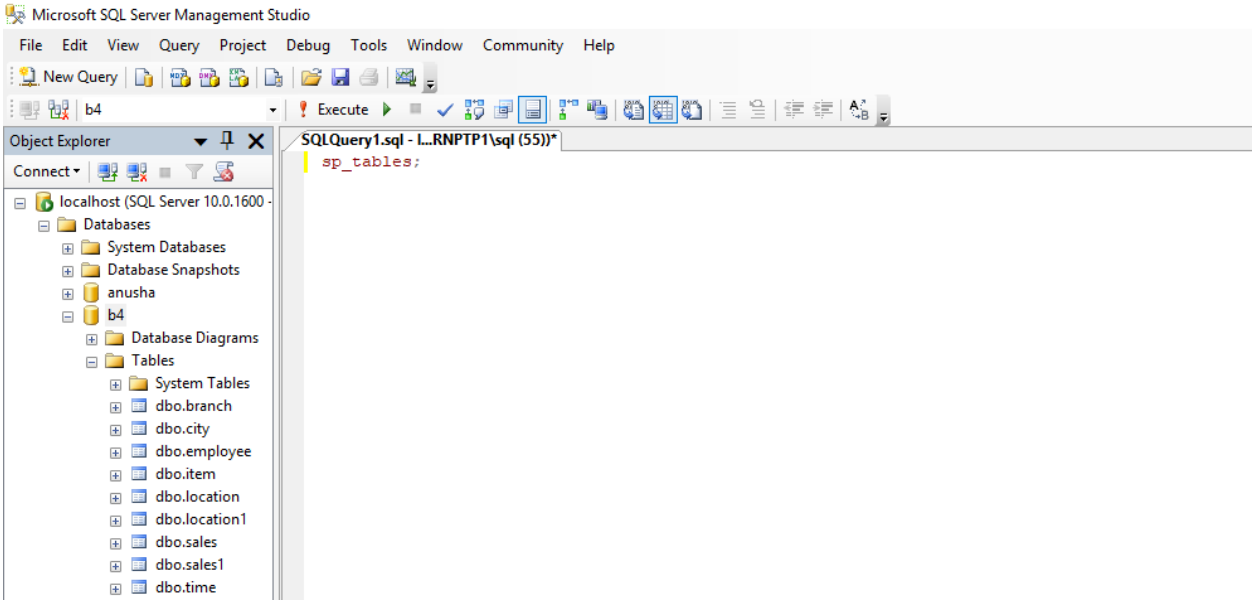


Output for star schema:



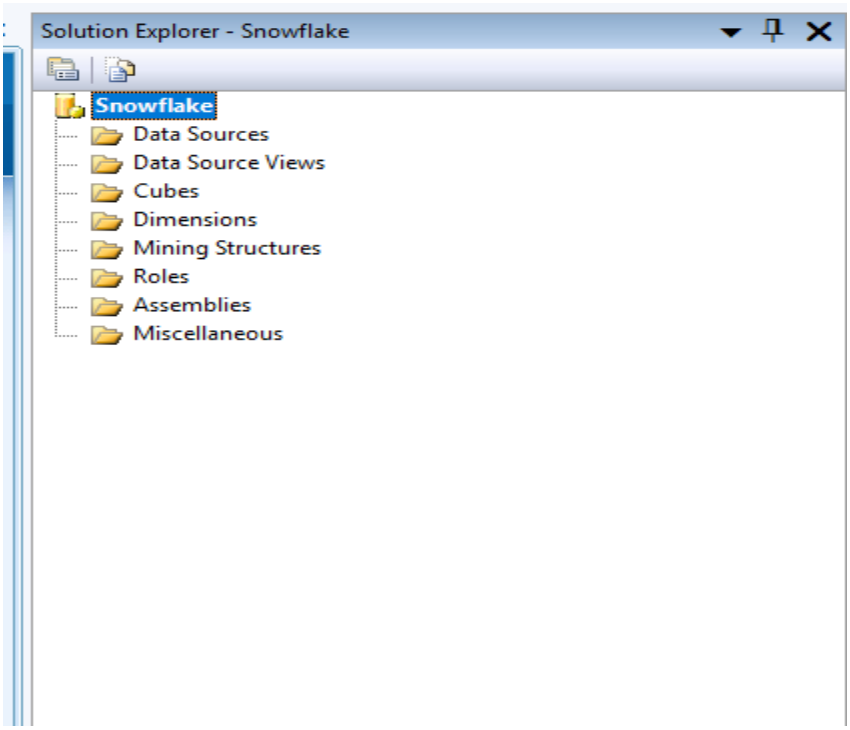
Snow flake schema:

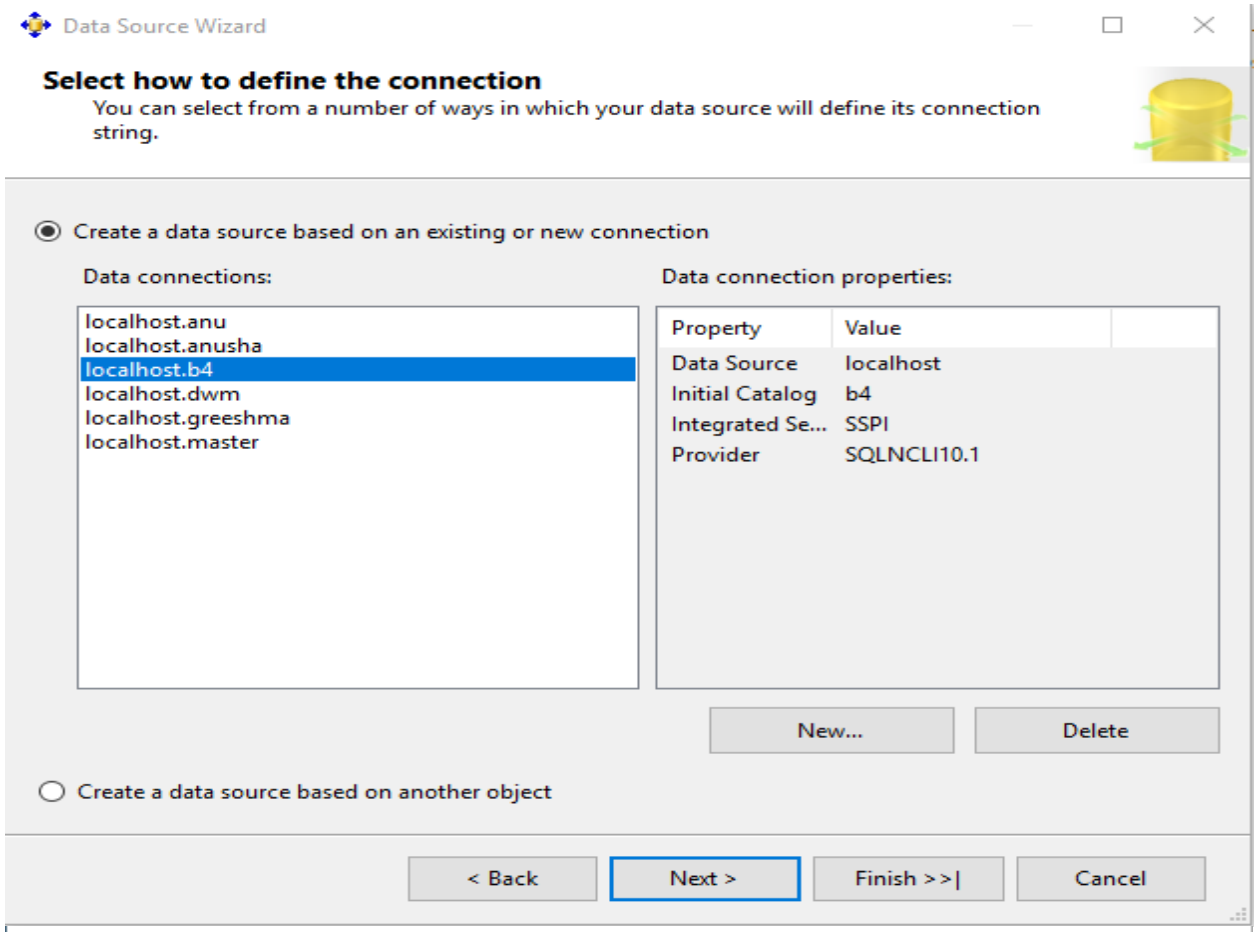
1.first create a database with required tables.

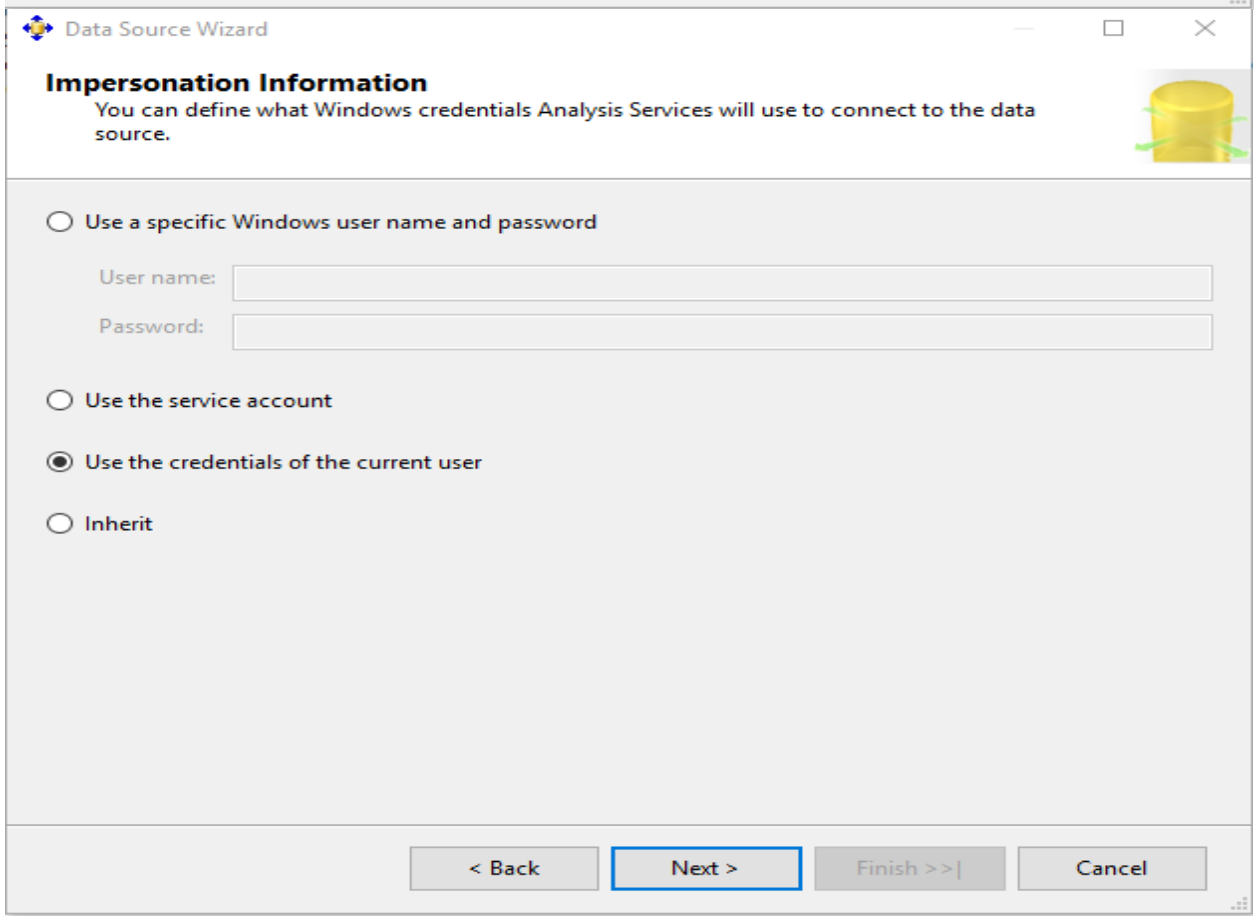


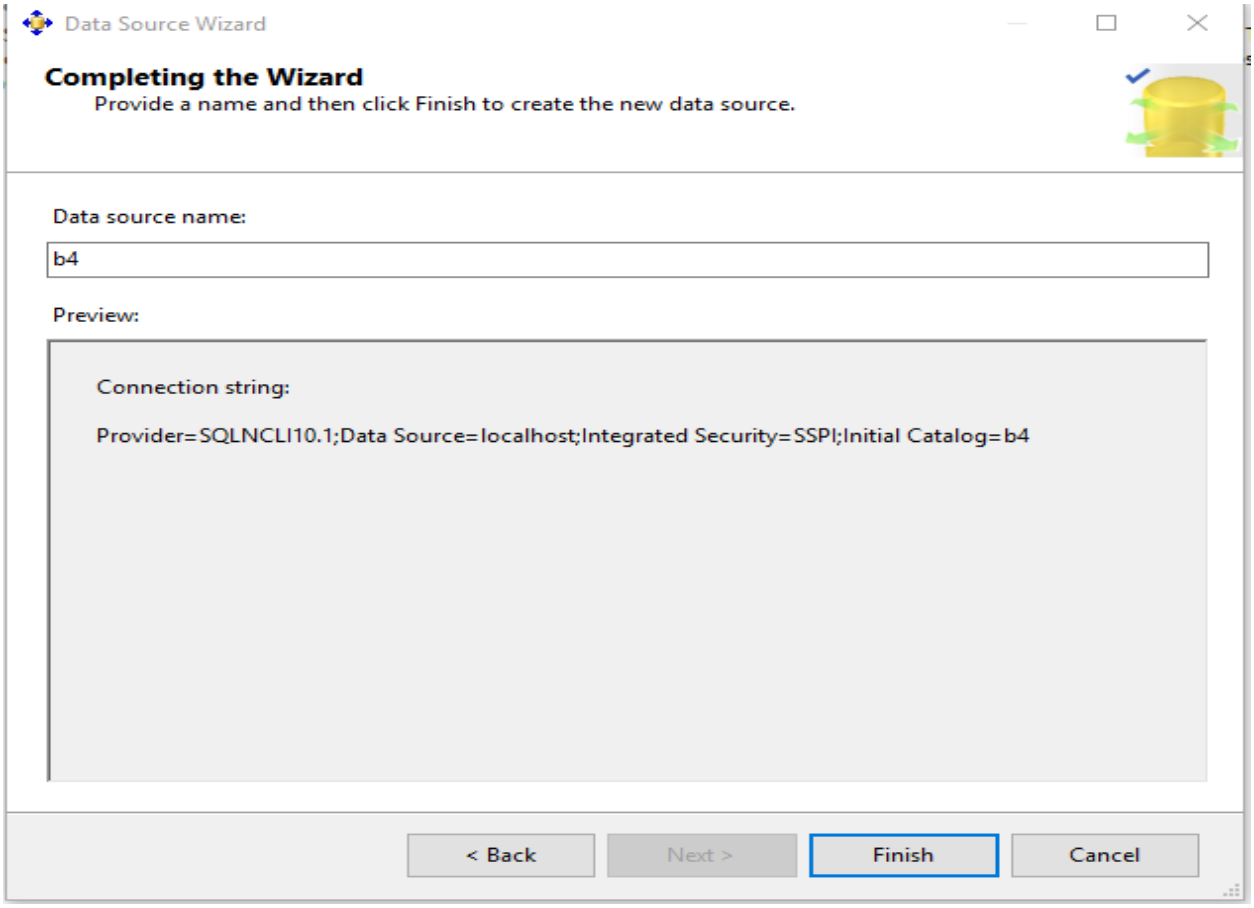
2.Now open a new project in BI studio and select analysis service project.

3. Select datasource and right click on it.Then click on new datasource,and then complete the datasource wizard by selecting the server name and database name on which action is to be performed.

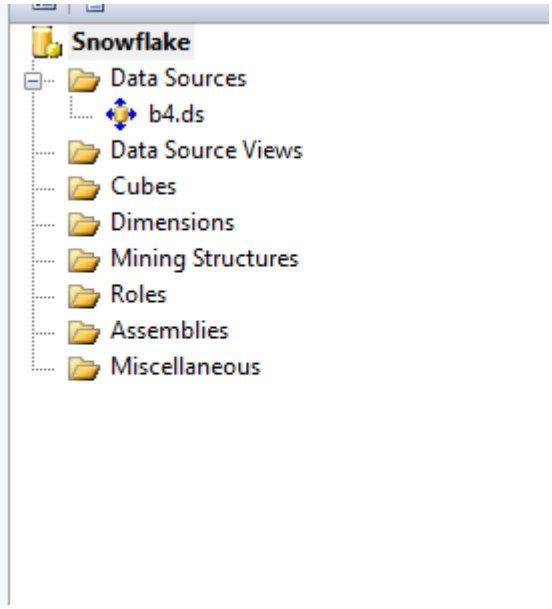


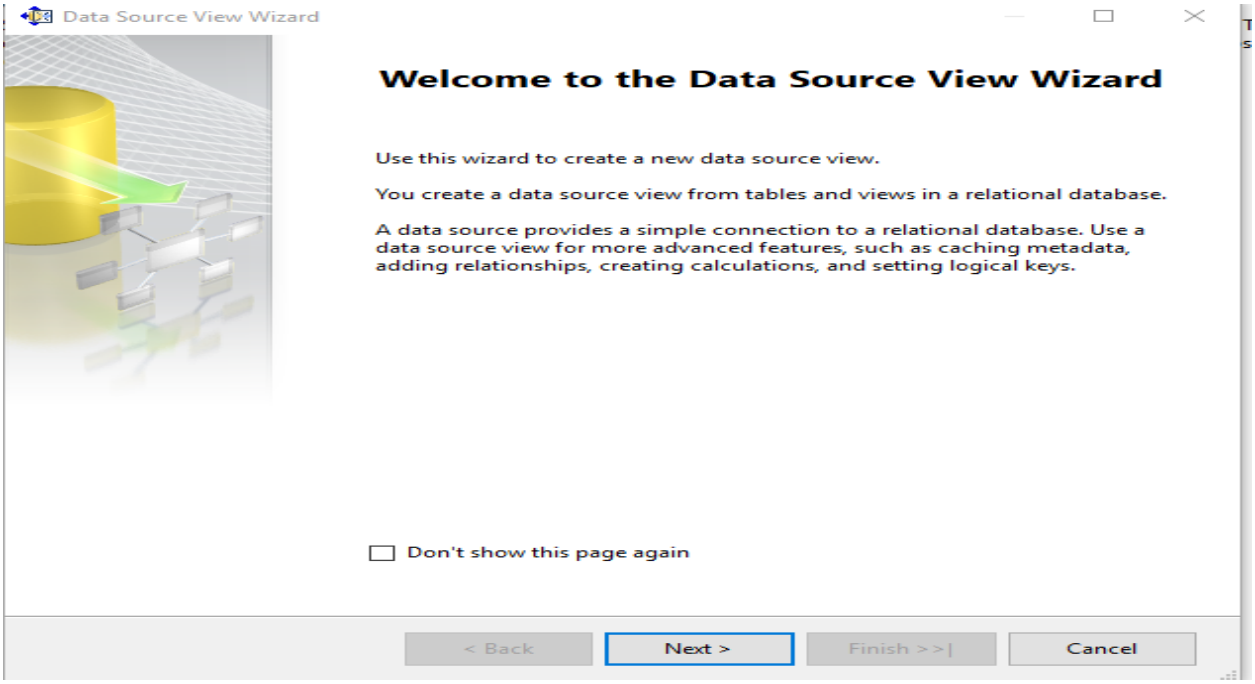

Data Source Wizard
Select how to define the connection
You can select from a number of ways in which your data source will define its connection string.
Create a data source based on an existing or new connection
Data connections:
localhost.anu
localhost.anusha
localhost.b4
localhost.dwm
localhost.greeshma
localhost.master
Data connection properties:
Property
Value
Data Source
localhost
Initial Catalog
b4
Integrated Se...
SSPI
Provider
SQLNCLI10.1
New...
Delete
Create a data source based on another object
< Back
Next >
Finish >>|
Cancel

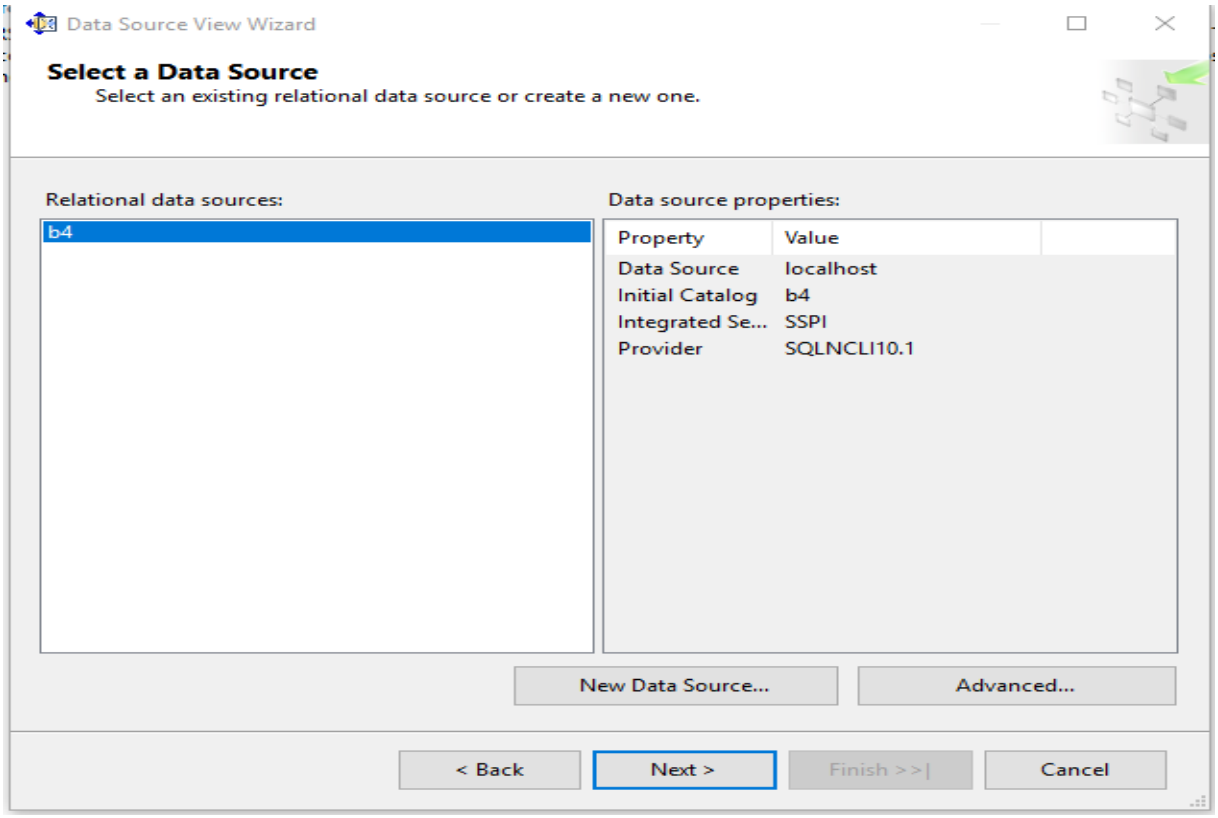

Data Source Wizard
Impersonation Information
You can define what Windows credentials Analysis Services will use to connect to the data source.
Use a specific Windows user name and password
User name:
Password:
Use the service account
Use the credentials of the current user
Inherit
< Back
Next >
Finish >>|
Cancel

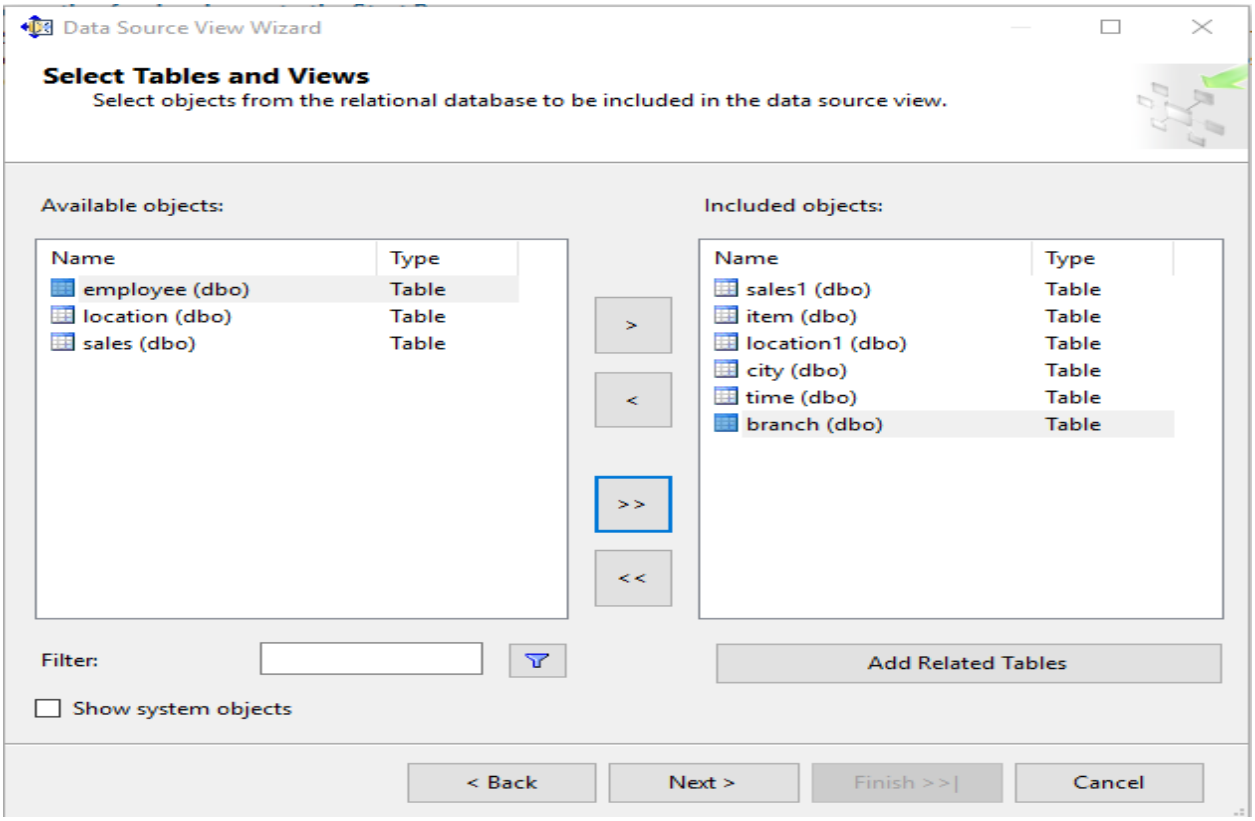


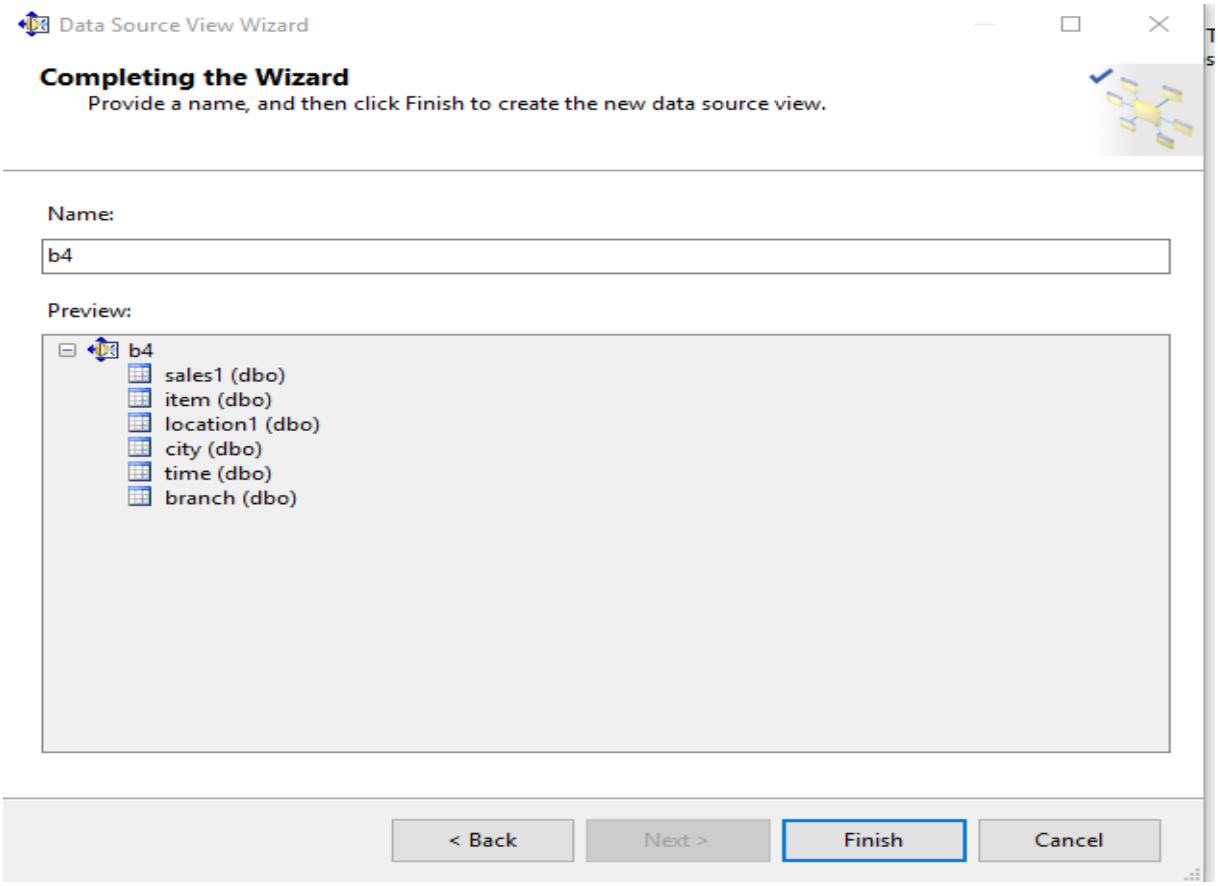
4. Select datasource view and right click on it.Then click on new datasource view,and then complete the datasource view wizard by selecting the tables used for snow flake schema.

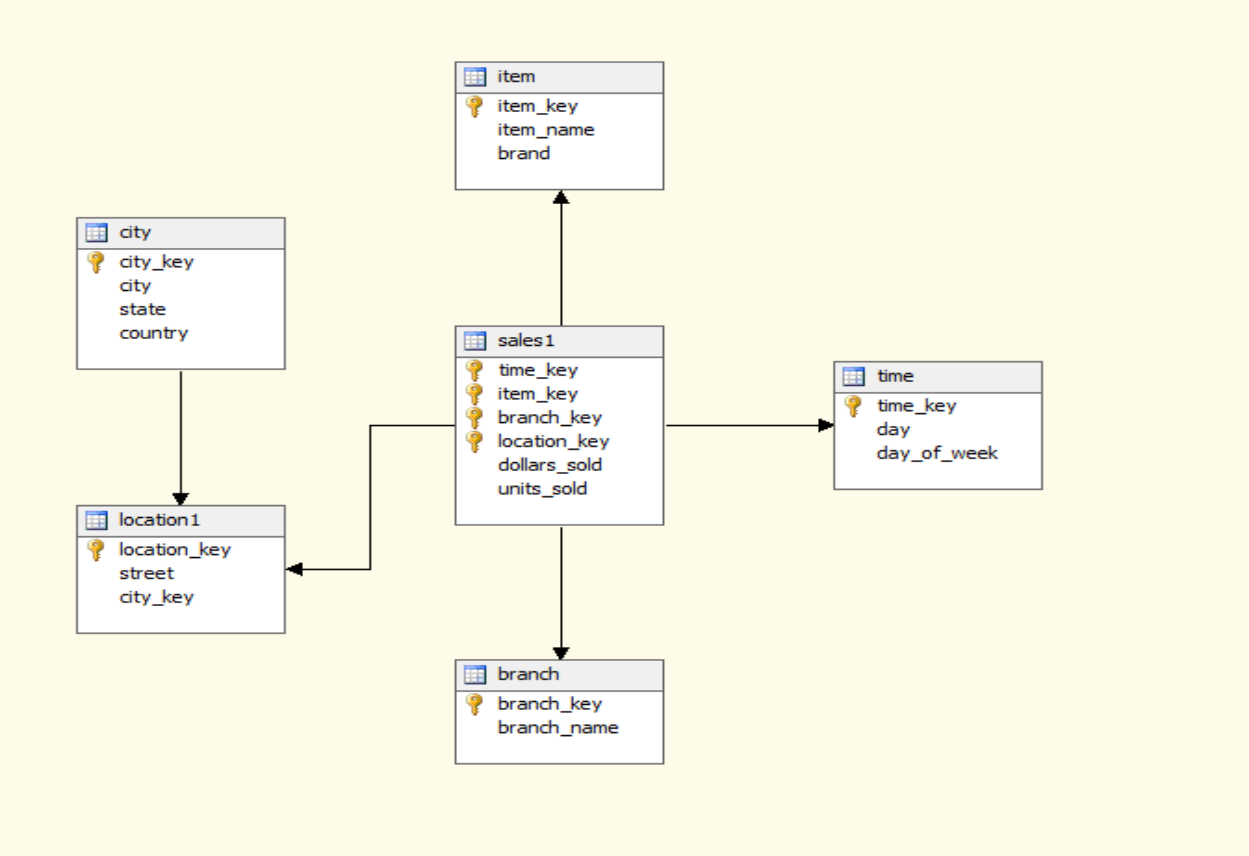



Data Source View Wizard
Welcome to the Data Source View Wizard
Use this wizard to create a new data source view.
You create a data source view from tables and views in a relational database.
A data source provides a simple connection to a relational database. Use a data source view for more advanced features, such as caching metadata, adding relationships, creating calculations, and setting logical keys.
Don't show this page again
< Back
Next >
Finish >>|
Cancel

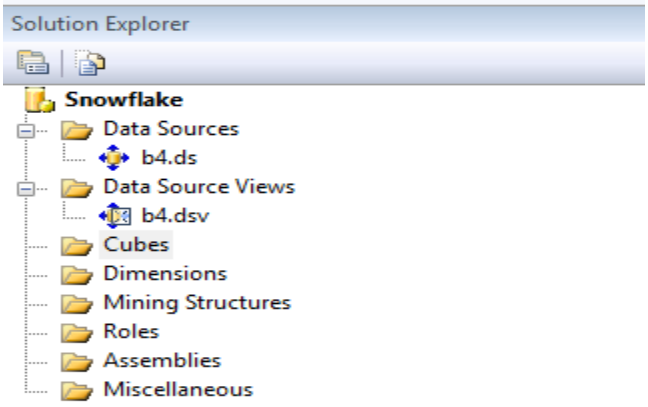

Data Source View Wizard
Select a Data Source
Select an existing relational data source or create a new one.
Relational data sources:
b4
Data source properties:
Property
Value
Data Source
localhost
Initial Catalog
b4
Integrated Se...
SSPI
Provider
SQLNCLI10.1
New Data Source...
Advanced...
< Back
Next >
Finish >>|
Cancel

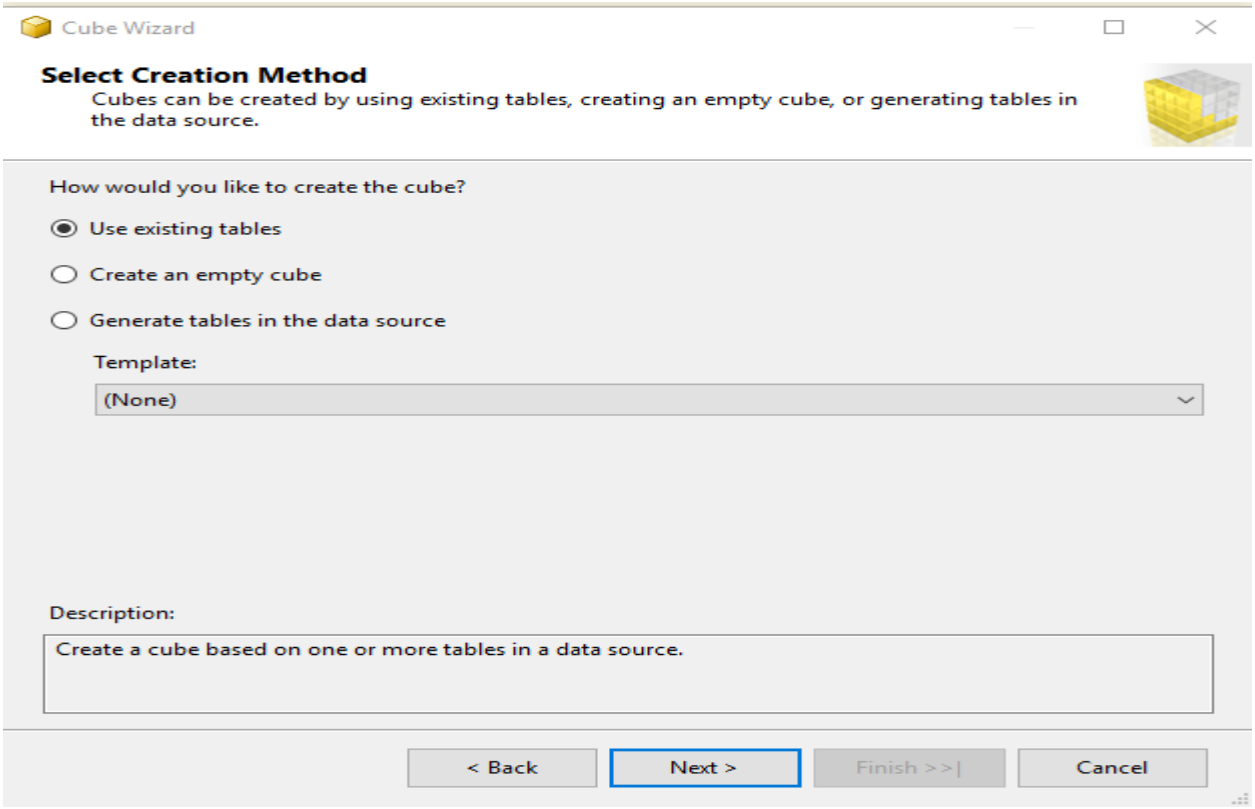

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
employee (dbo)
Table
location (dbo)
Table
sales (dbo)
Table
>
<
>>
<<
Included objects:
Name
Type
sales1 (dbo)
Table
item (dbo)
Table
location1 (dbo)
Table
city (dbo)
Table
time (dbo)
Table
branch (dbo)
Table
Filter:
Add Related Tables
Show system objects
< Back
Next >
Finish >>|
Cancel

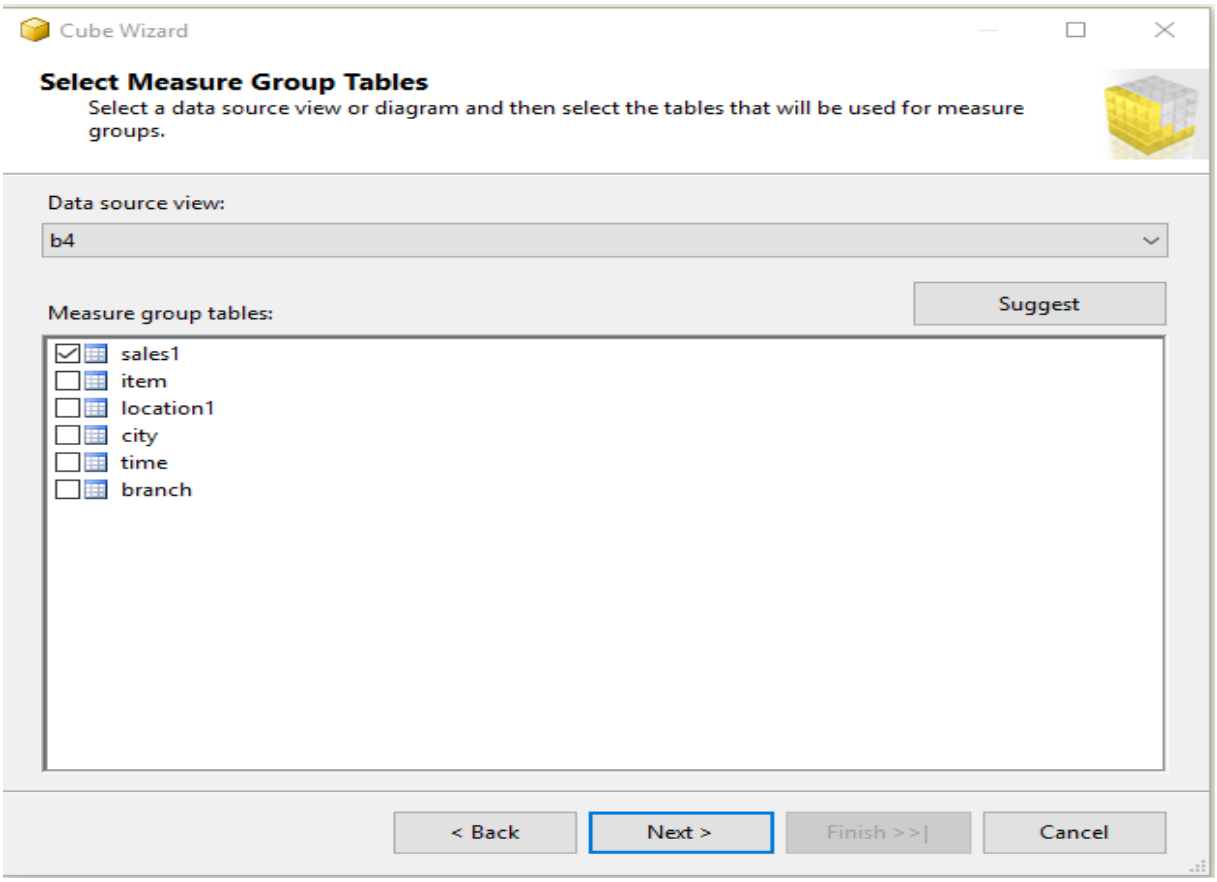

Data Source View Wizard
Completing the Wizard
Provide a name, and then click Finish to create the new data source view.
Name:
b4
Preview:
b4
sales1 (dbo)
item (dbo)
location1 (dbo)
city (dbo)
time (dbo)
branch (dbo)
< Back
Next >
Finish
Cancel

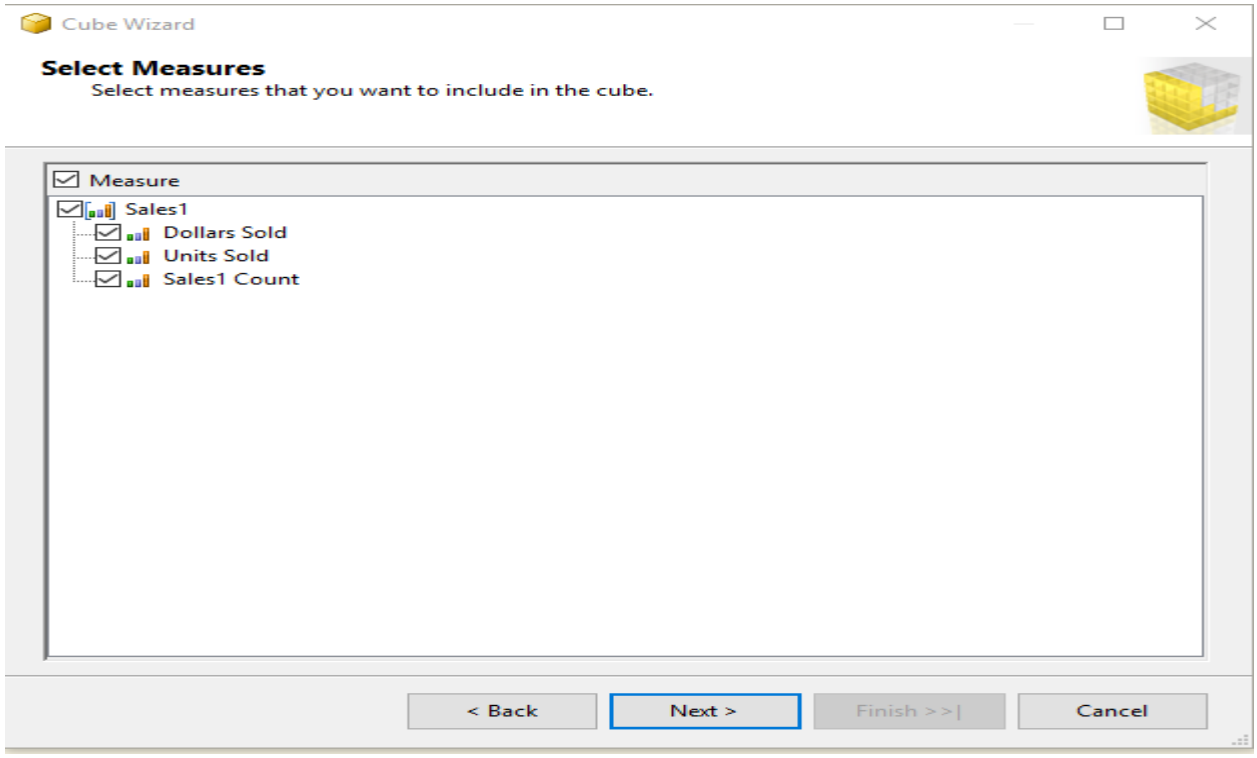


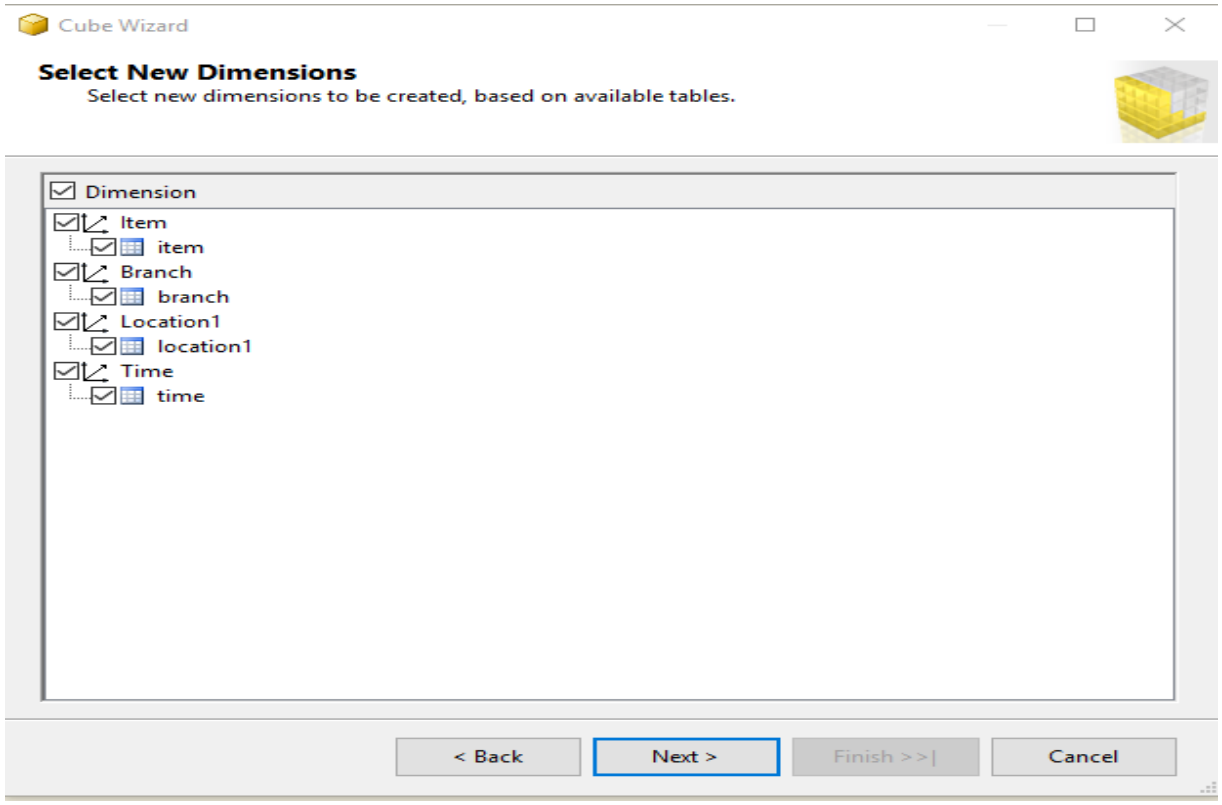
5. Select cubes and right click on it.Then click on new cube,and then complete the cube wizard by selecting the fact tables used for snow flake schema.

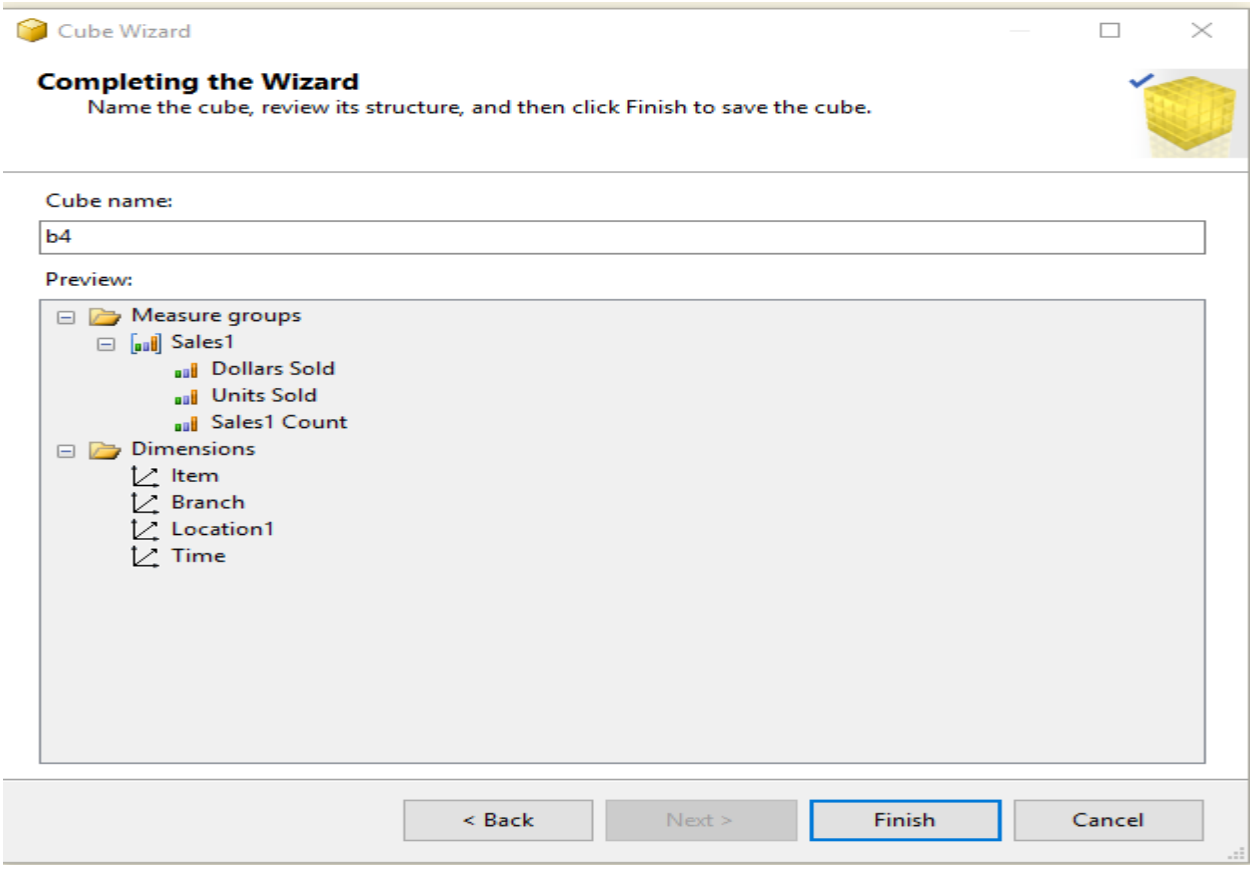



Cube Wizard
Select Creation Method
Cubes can be created by using existing tables, creating an empty cube, or generating tables in the data source.
How would you like to create the cube?
Use existing tables
Create an empty cube
Generate tables in the data source
Template:
(None)
Description:
Create a cube based on one or more tables in a data source.
< Back
Next >
Finish >>|
Cancel

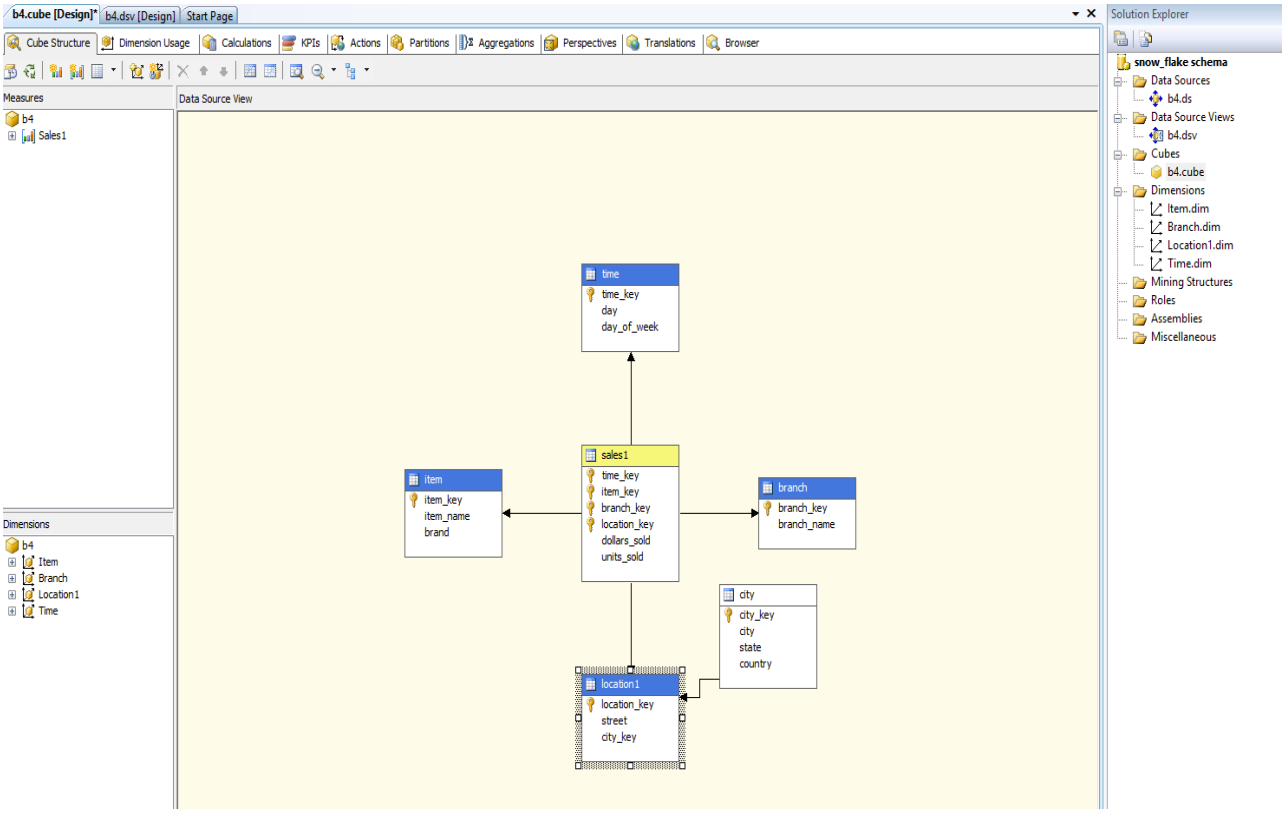

Cube Wizard
Select Measure Group Tables
Select a data source view or diagram and then select the tables that will be used for measure groups.
Data source view:
b4
Measure group tables:
Suggest
sales1
item
location1
city
time
branch
< Back
Next >
Finish >>|
Cancel


Cube Wizard
Select Measures
Select measures that you want to include in the cube.
Measure
Sales1
Dollars Sold
Units Sold
Sales1 Count
< Back
Next >
Finish >>|
Cancel


Cube Wizard
Select New Dimensions
Select new dimensions to be created, based on available tables.
Dimension
Item
item
Branch
branch
Location1
location1
Time
time
< Back
Next >
Finish >>|
Cancel

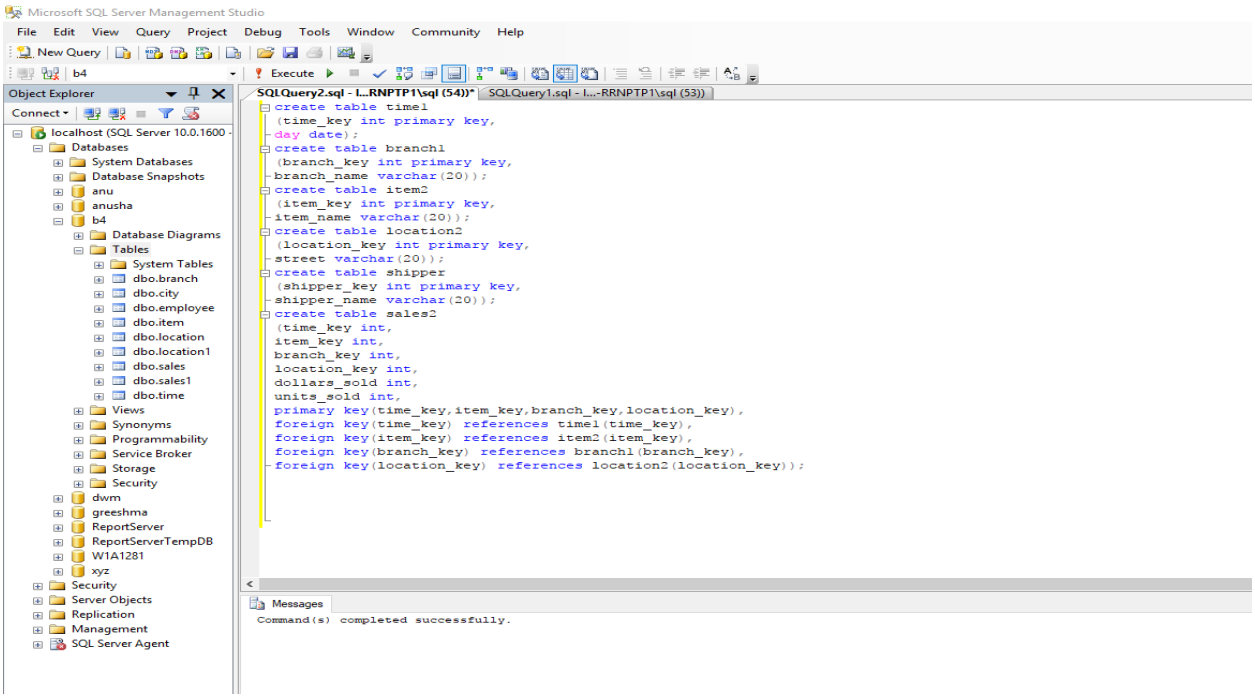


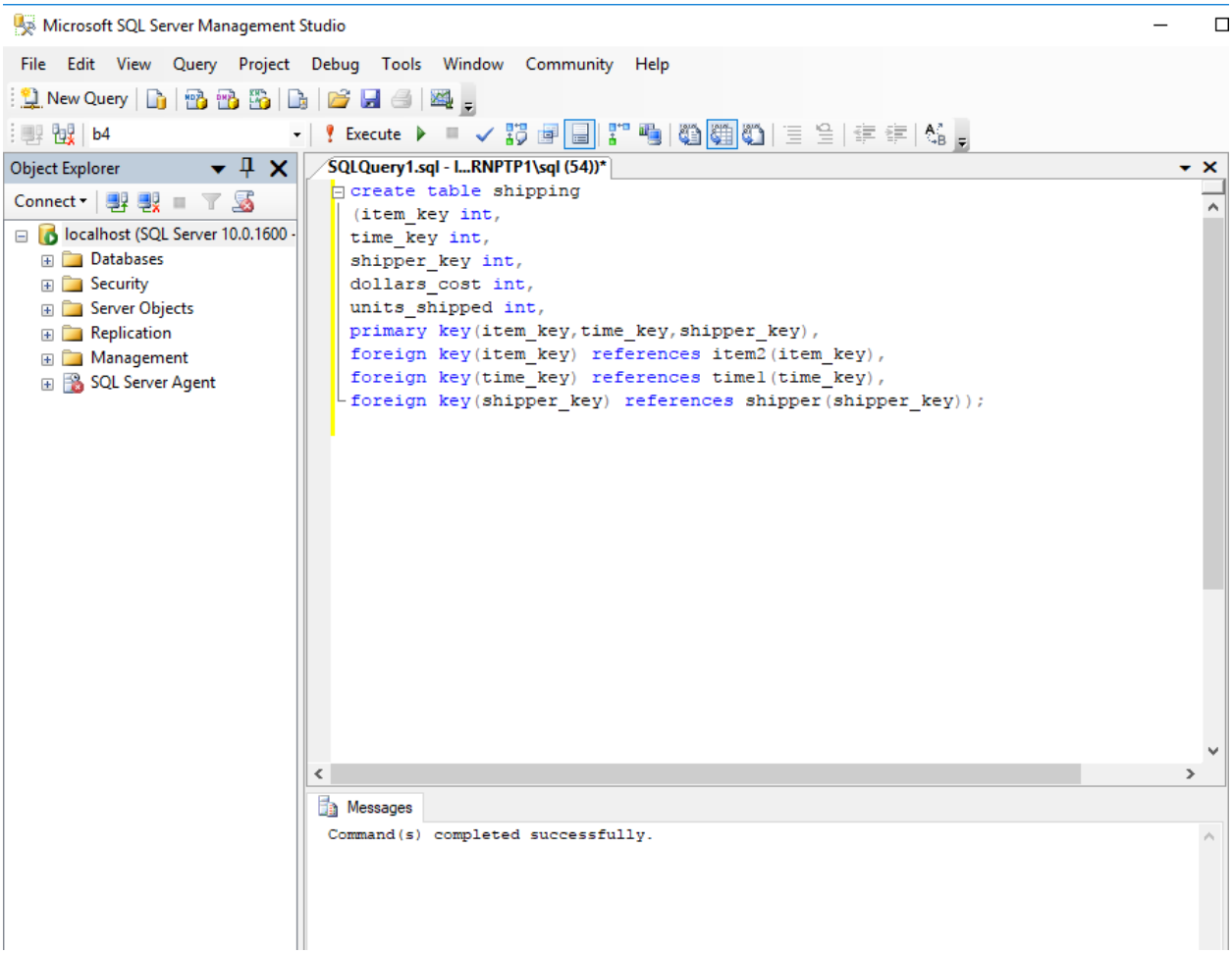
Output for snow flake schema:



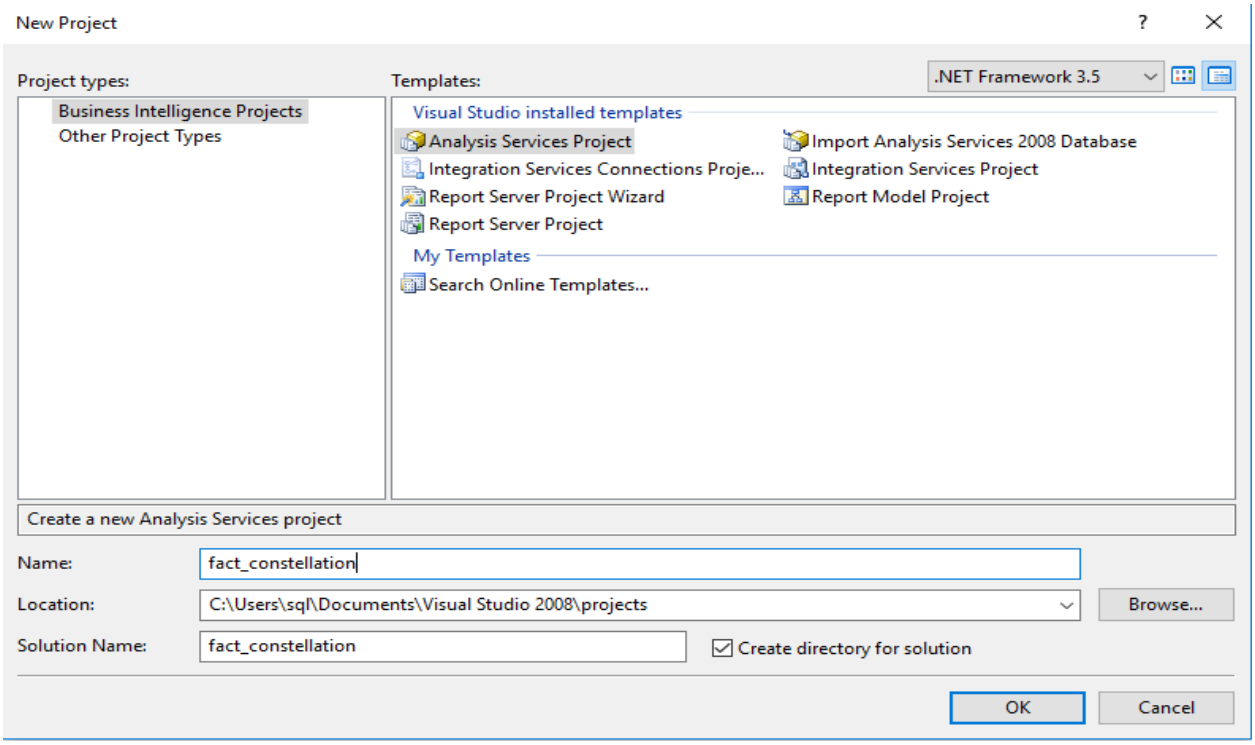
Fact constellation:

1.first create a database with required tables.

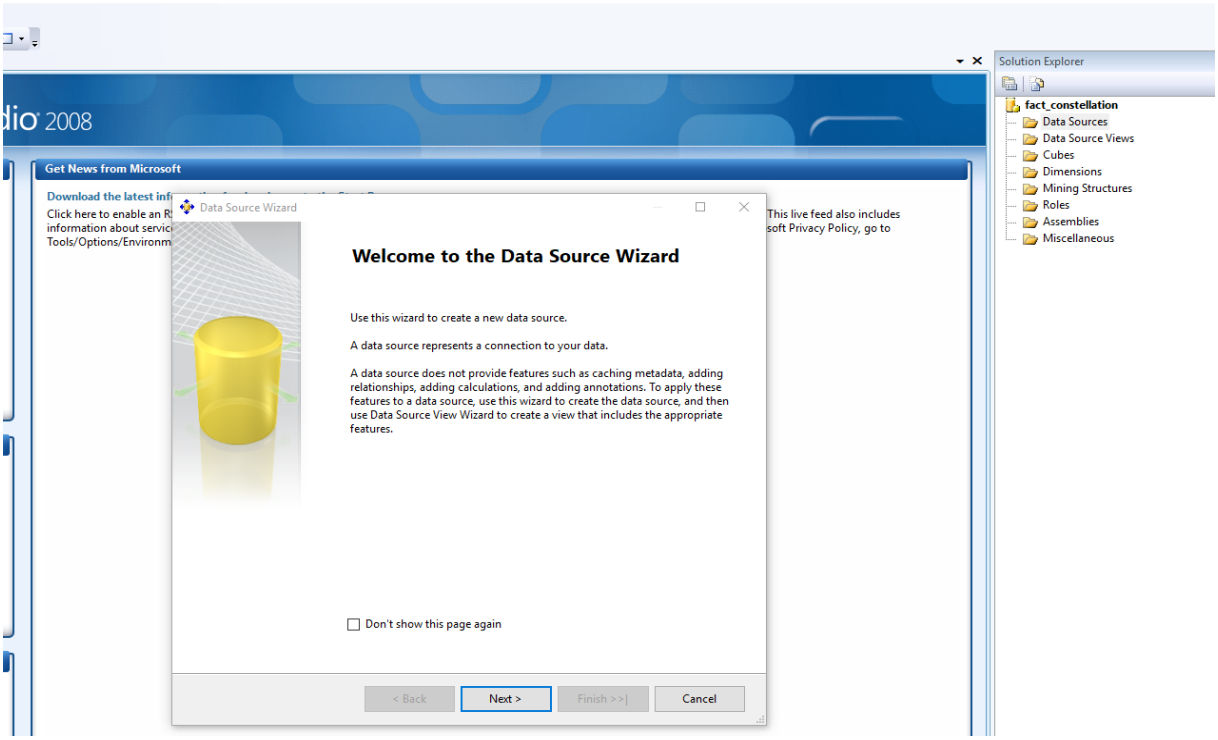


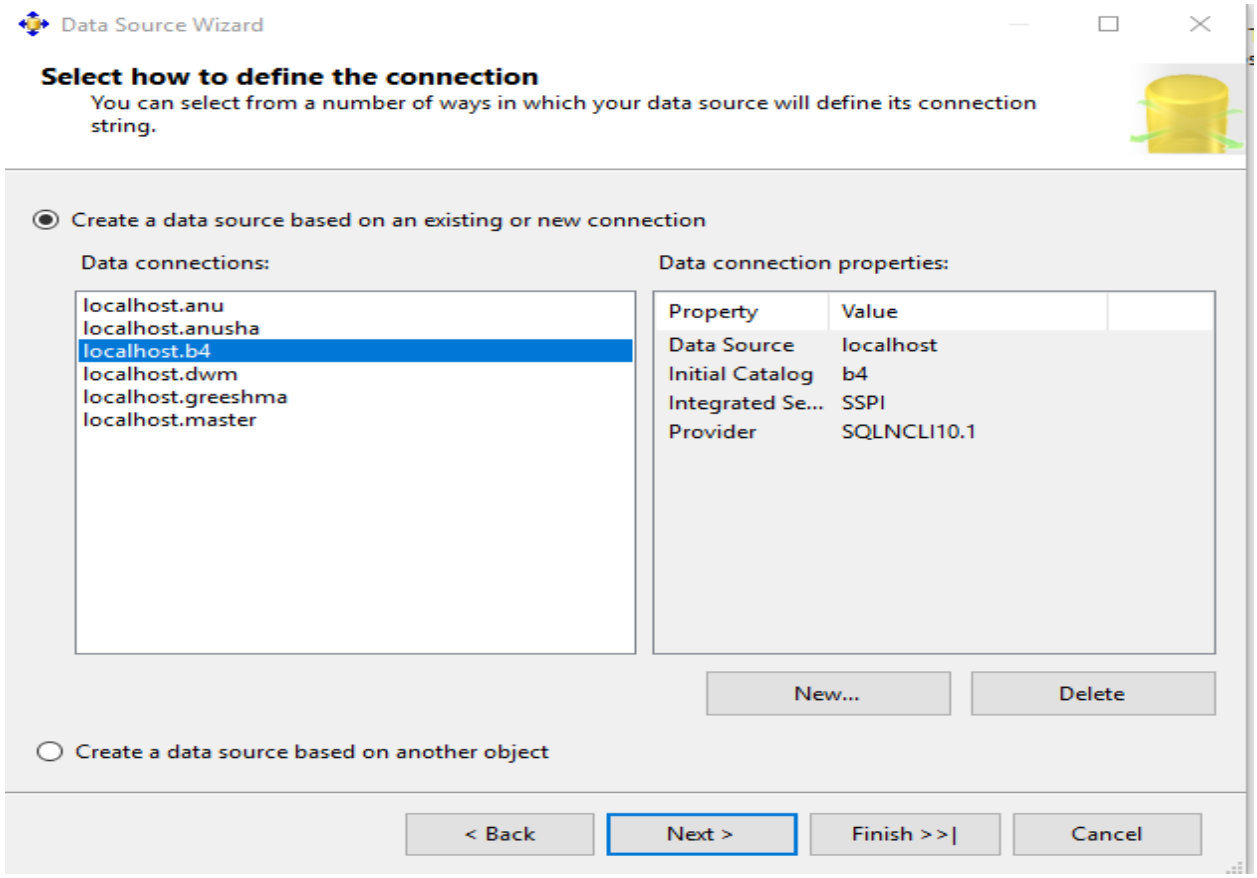


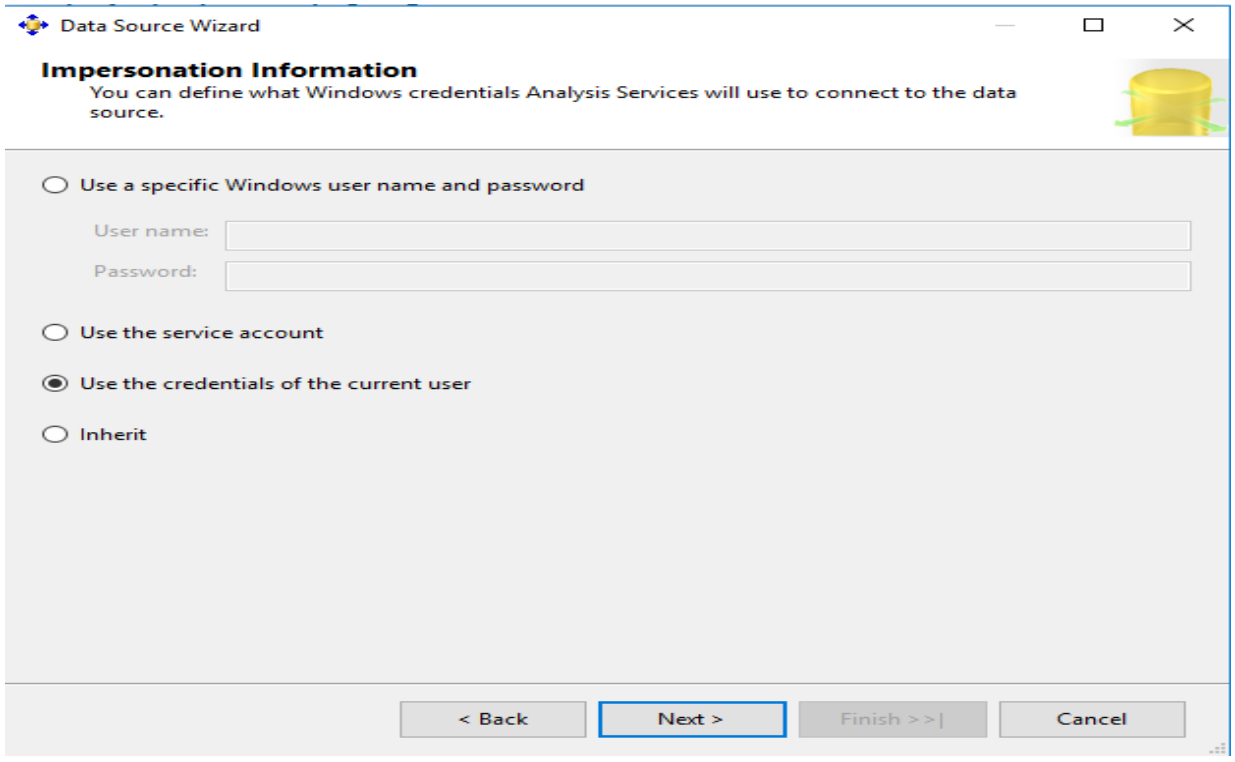
2. Now open a new project in BI studio and select analysis service project.

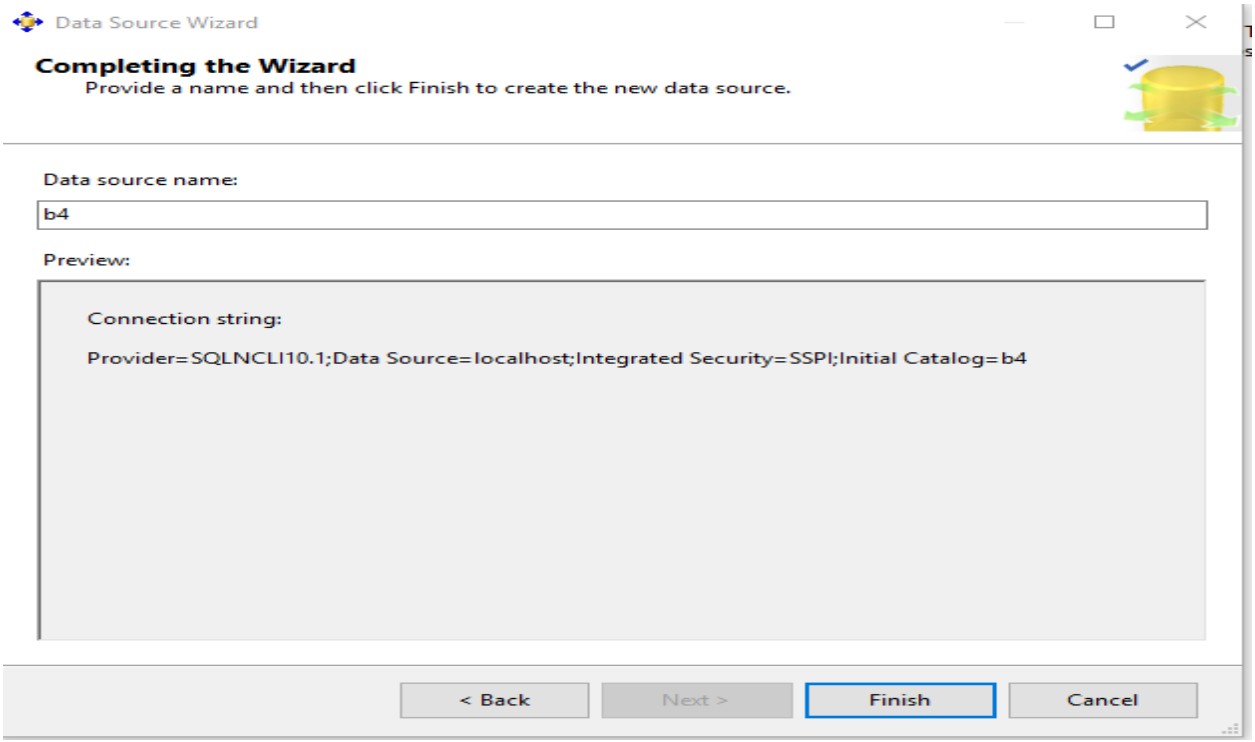


3. Select datasource and right click on it.Then click on new datasource,and then complete the datasource wizard by selecting the server name and database name on which action is to be performed.

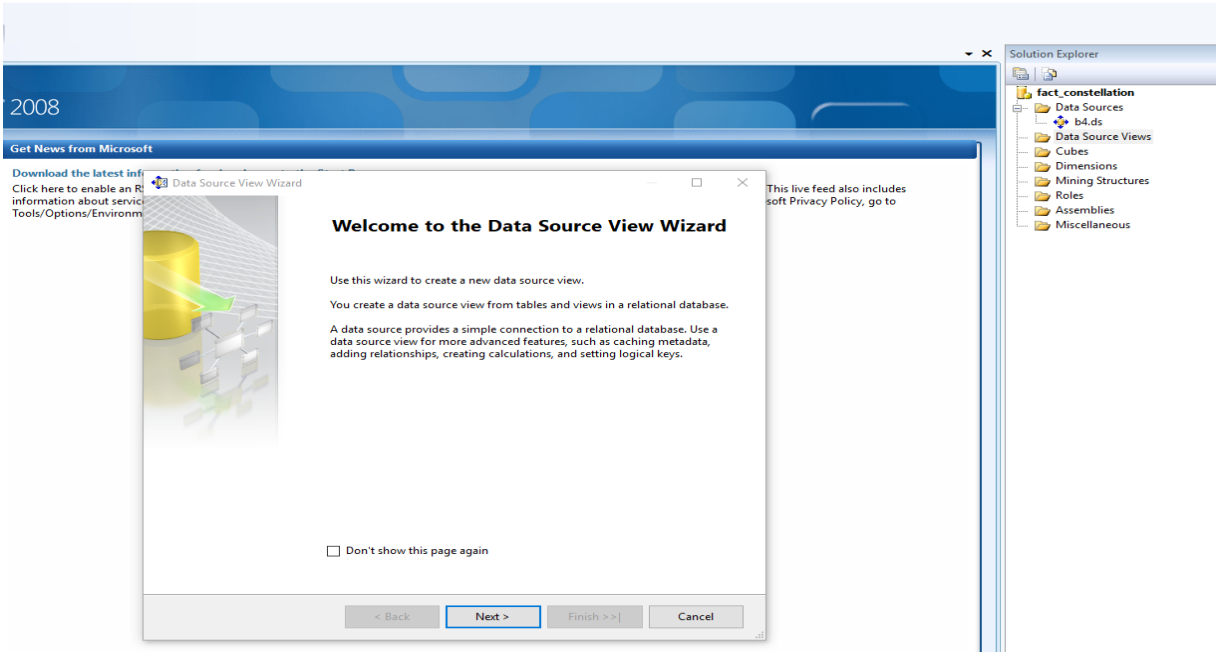


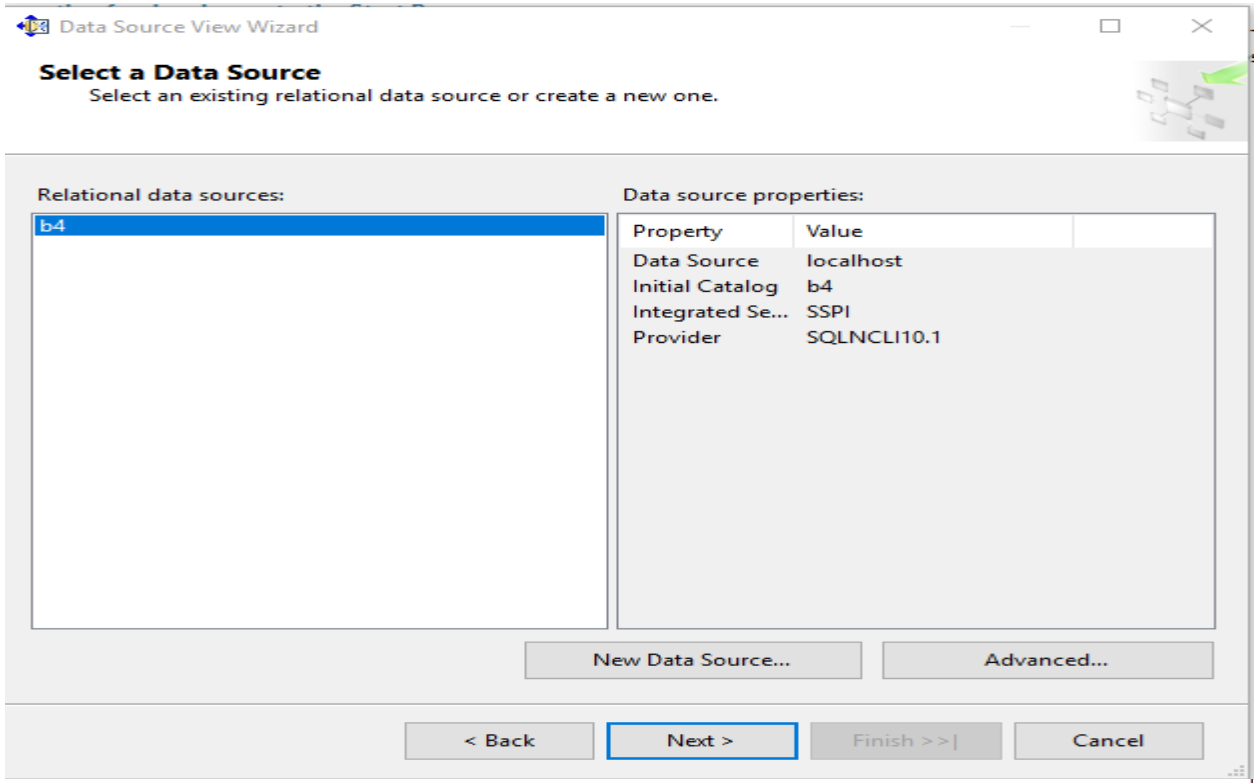

Data Source Wizard
Select how to define the connection
You can select from a number of ways in which your data source will define its connection string.
Create a data source based on an existing or new connection
Data connections:
localhost.anu
localhost.anusha
localhost.b4
localhost.dwm
localhost.greeshma
localhost.master
Data connection properties:
Property
Value
Data Source
localhost
Initial Catalog
b4
Integrated Se...
SSPI
Provider
SQLNCLI10.1
New...
Delete
Create a data source based on another object
< Back
Next >
Finish >>|
Cancel

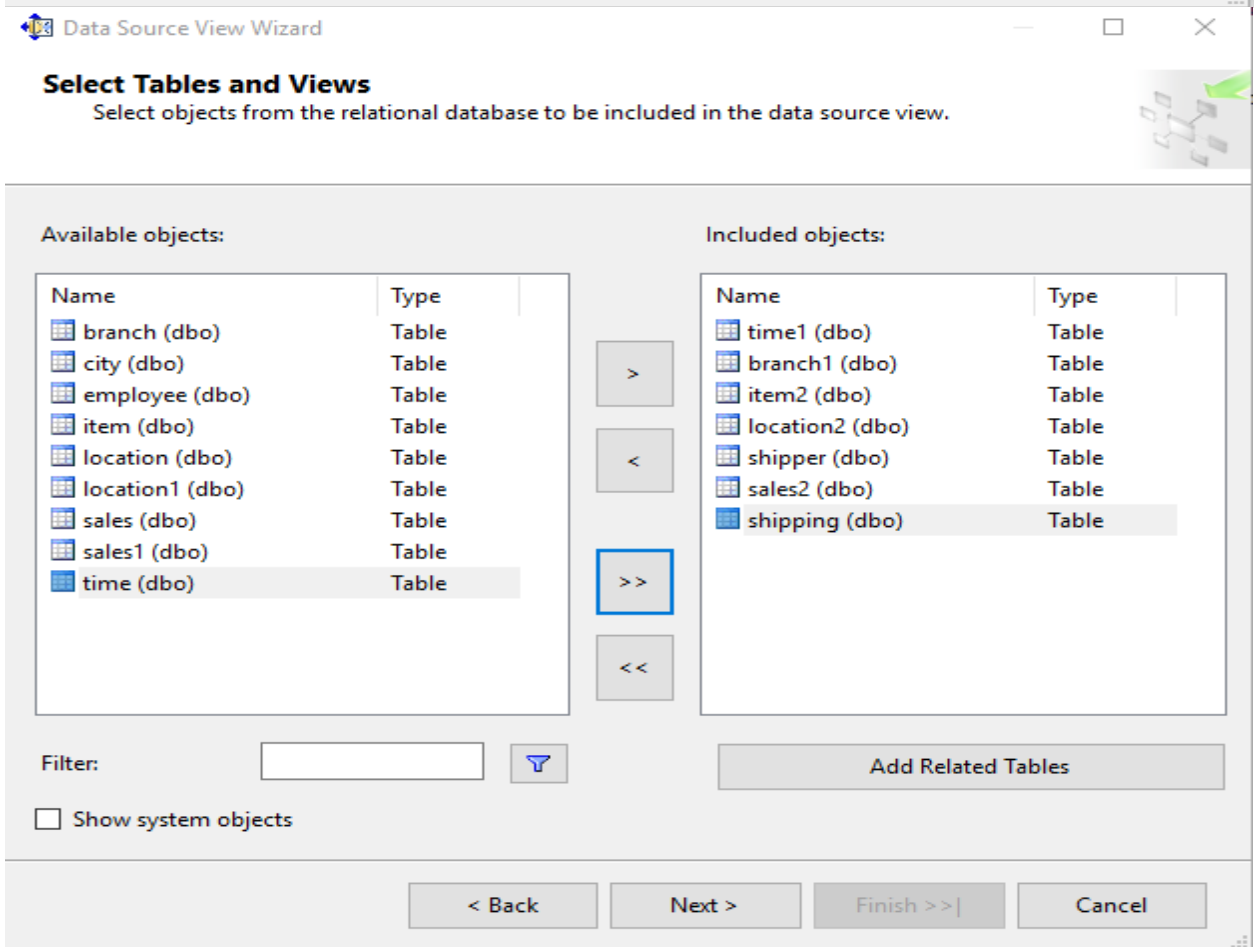

Data Source Wizard
Impersonation Information
You can define what Windows credentials Analysis Services will use to connect to the data source.
Use a specific Windows user name and password
User name:
Password:
Use the service account
Use the credentials of the current user
Inherit
< Back
Next >
Finish >>|
Cancel

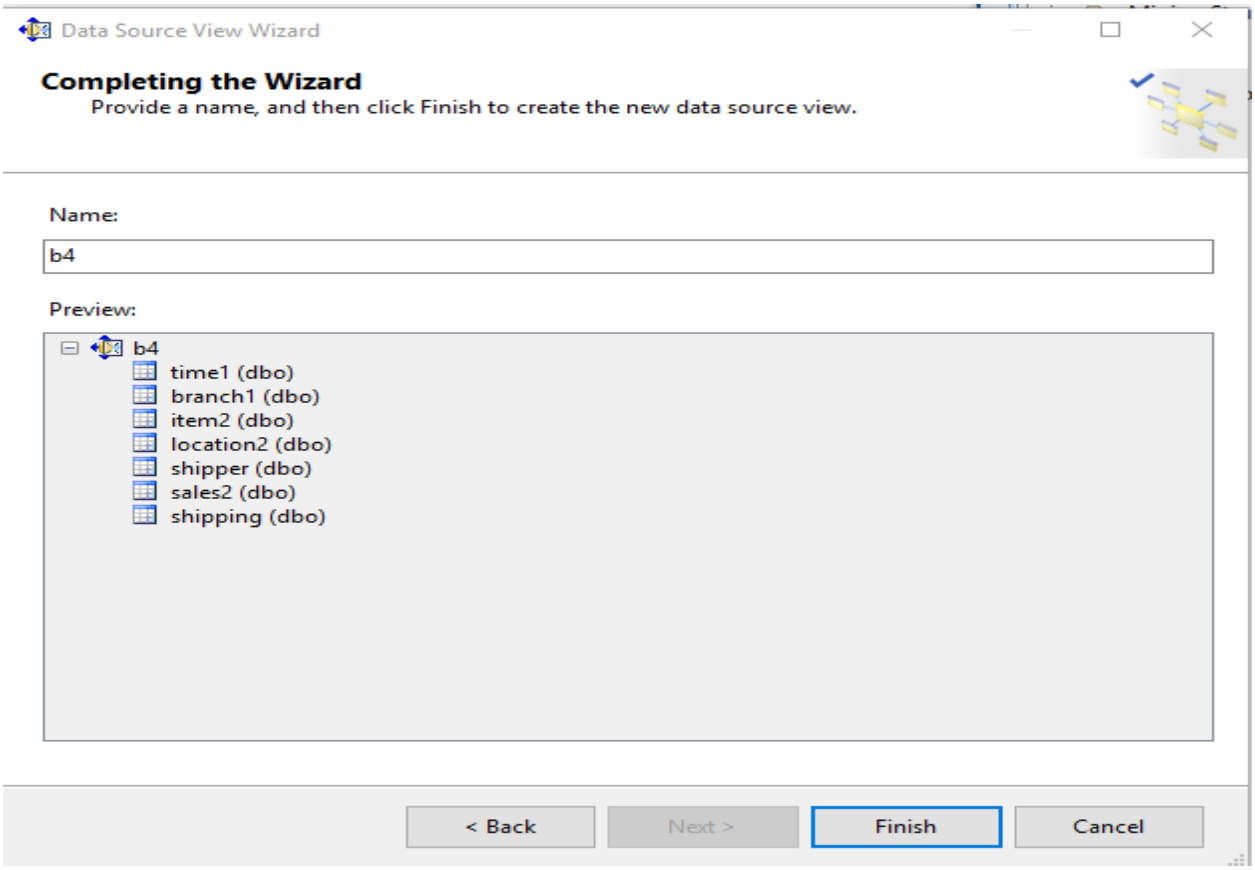


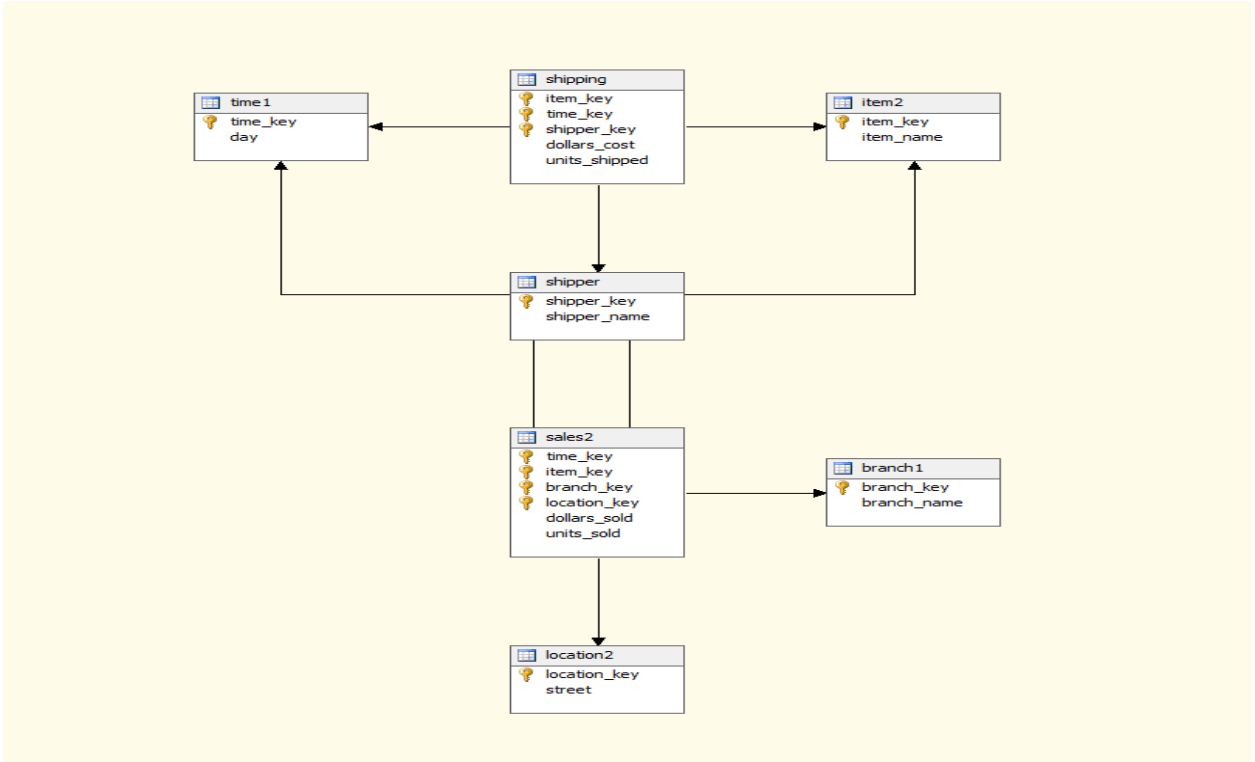
4. Select datasource view and right click on it.Then click on new datasource view,and then complete the datasource view wizard by selecting the tables used for fact constellation schema.



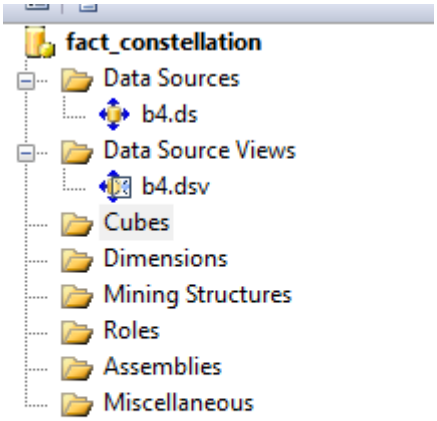

Data Source View Wizard
Select a Data Source
Select an existing relational data source or create a new one.
Relational data sources:
b4
Data source properties:
Property
Value
Data Source
localhost
Initial Catalog
b4
Integrated Se...
SSPI
Provider
SQLNCLI10.1
New Data Source...
Advanced...
< Back
Next >
Finish >>|
Cancel

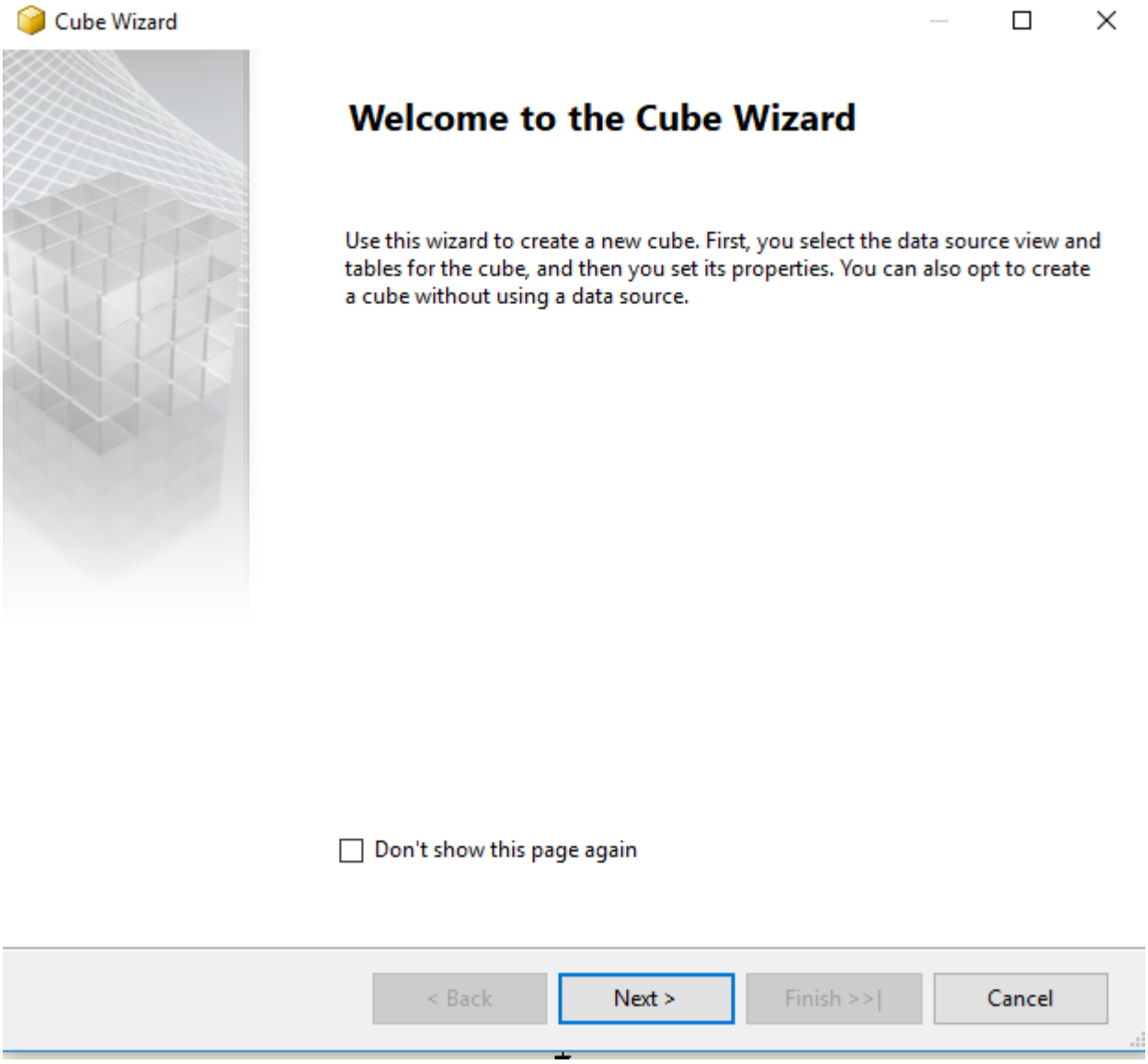

Data Source View Wizard
Select Tables and Views
Select objects from the relational database to be included in the data source view.
Available objects:
Name
Type
branch (dbo)
Table
city (dbo)
Table
employee (dbo)
Table
item (dbo)
Table
location (dbo)
Table
location1 (dbo)
Table
sales (dbo)
Table
sales1 (dbo)
Table
time (dbo)
Table
>
<
>>
<<
Included objects:
Name
Type
time1 (dbo)
Table
branch1 (dbo)
Table
item2 (dbo)
Table
location2 (dbo)
Table
shipper (dbo)
Table
sales2 (dbo)
Table
shipping (dbo)
Table
Filter:
Add Related Tables
Show system objects
< Back
Next >
Finish >>|
Cancel

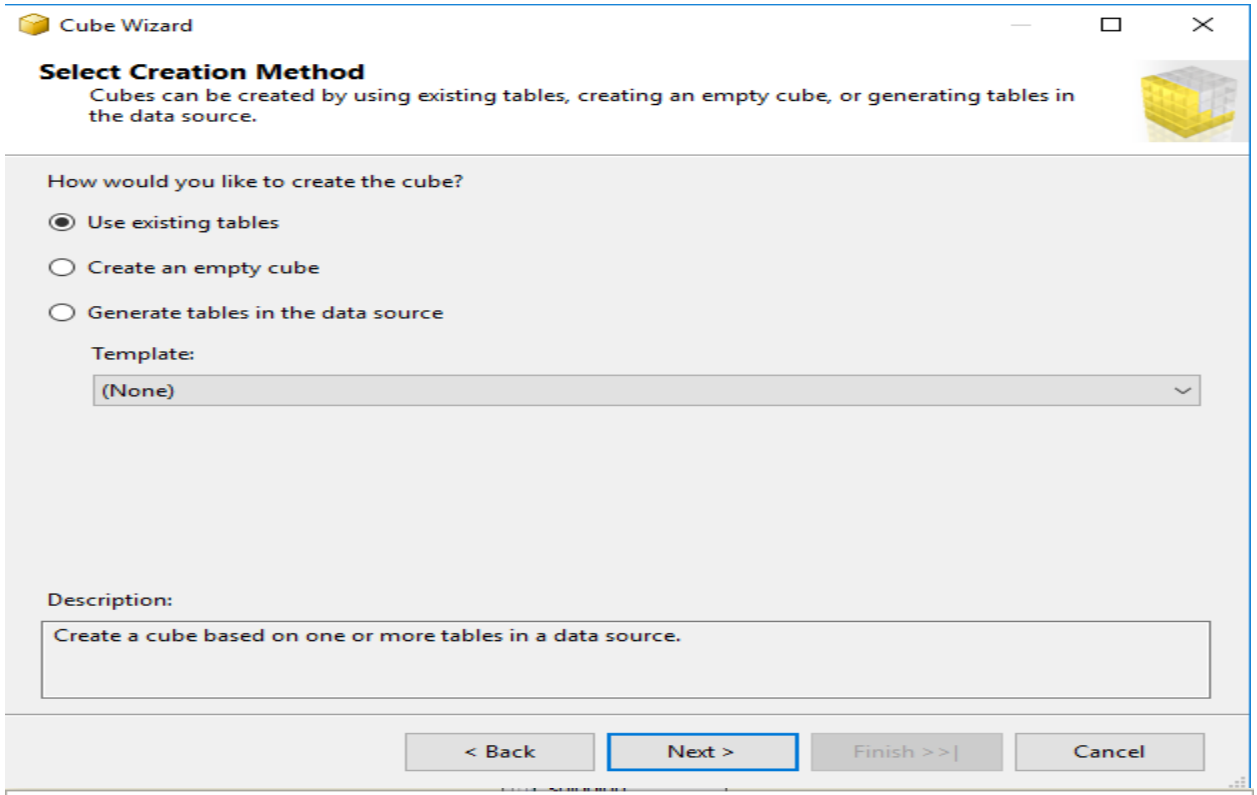


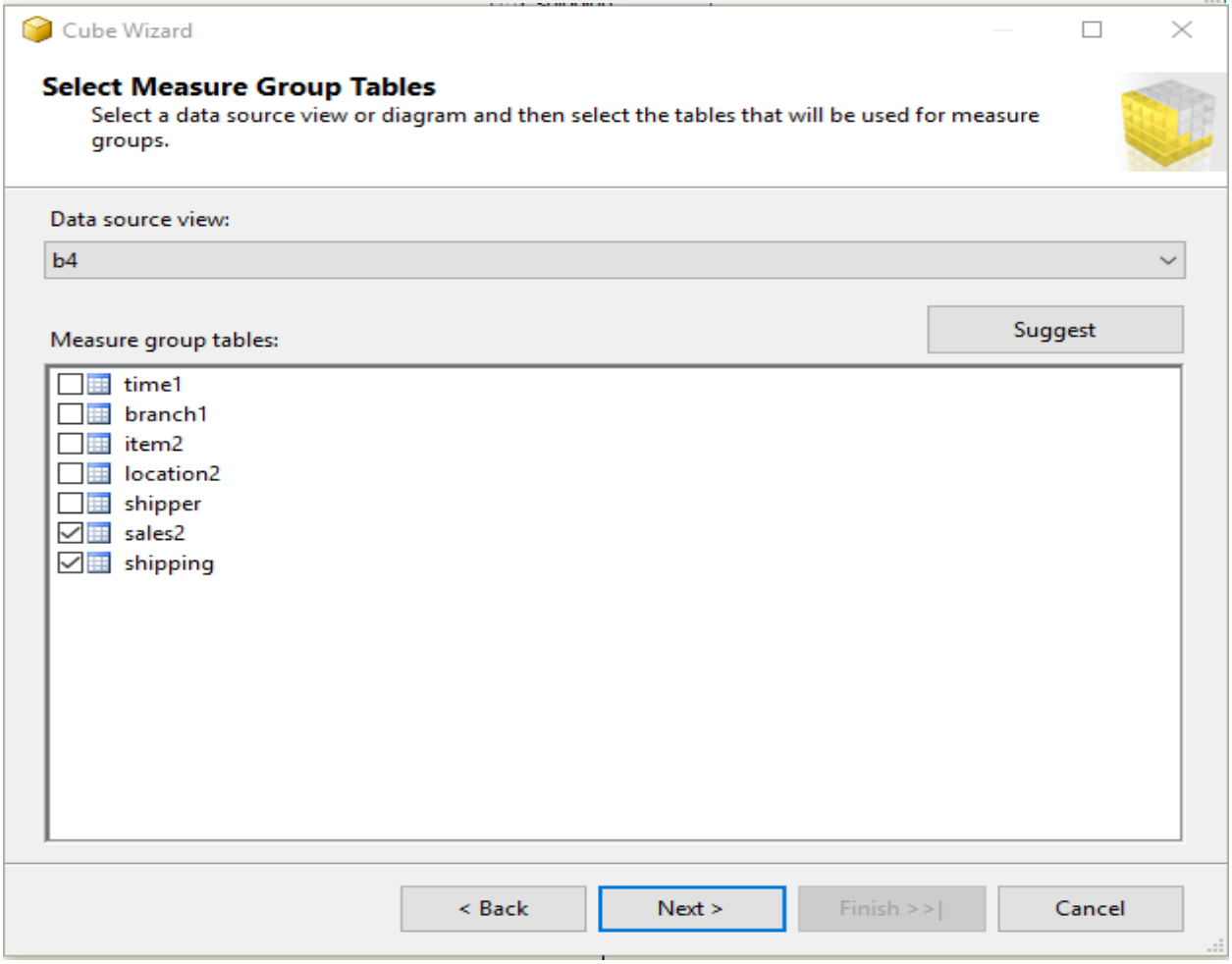


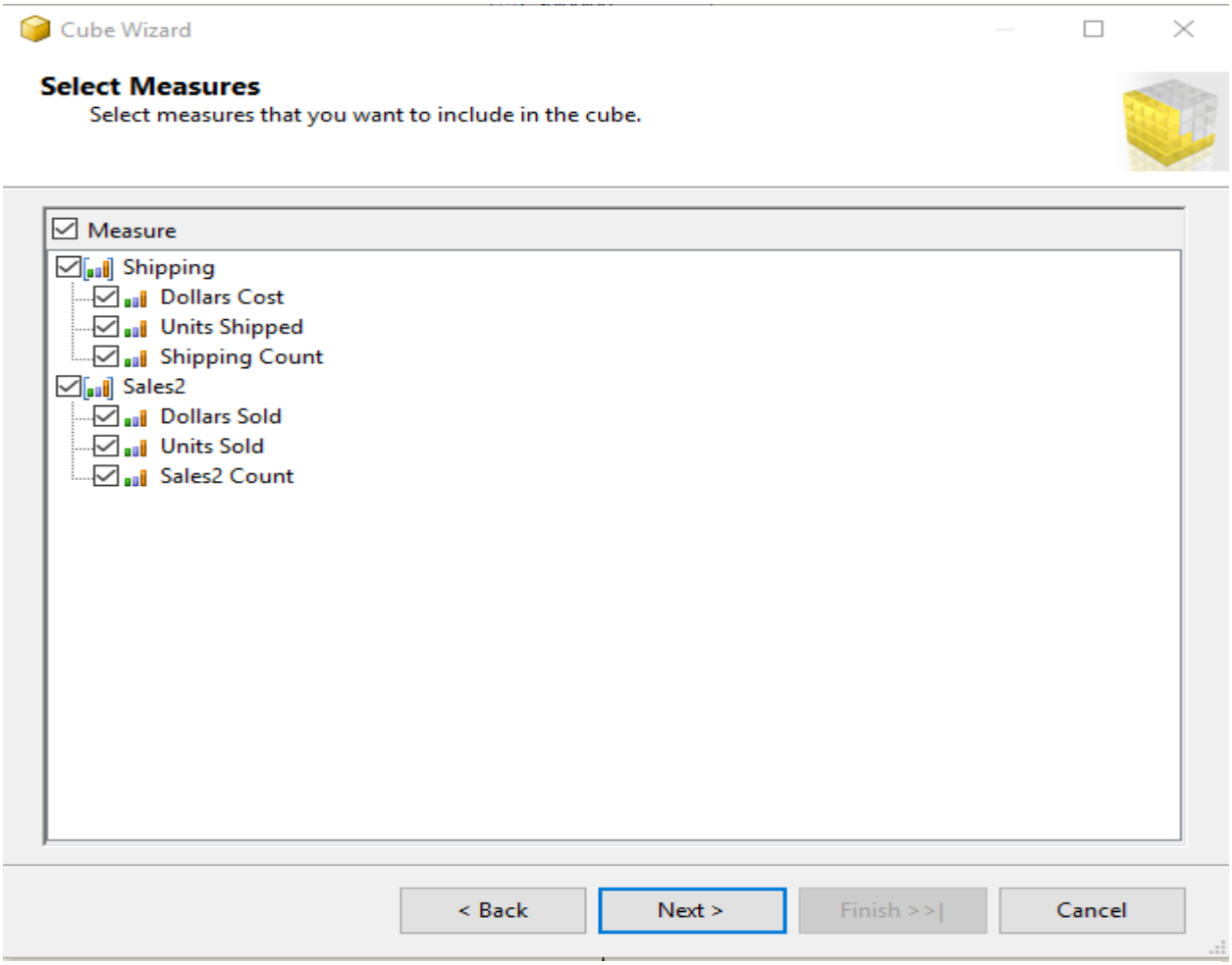
5. Select cubes and right click on it.Then click on new cube,and then complete the cube wizard by selecting the fact tables used for fact constellation schema.

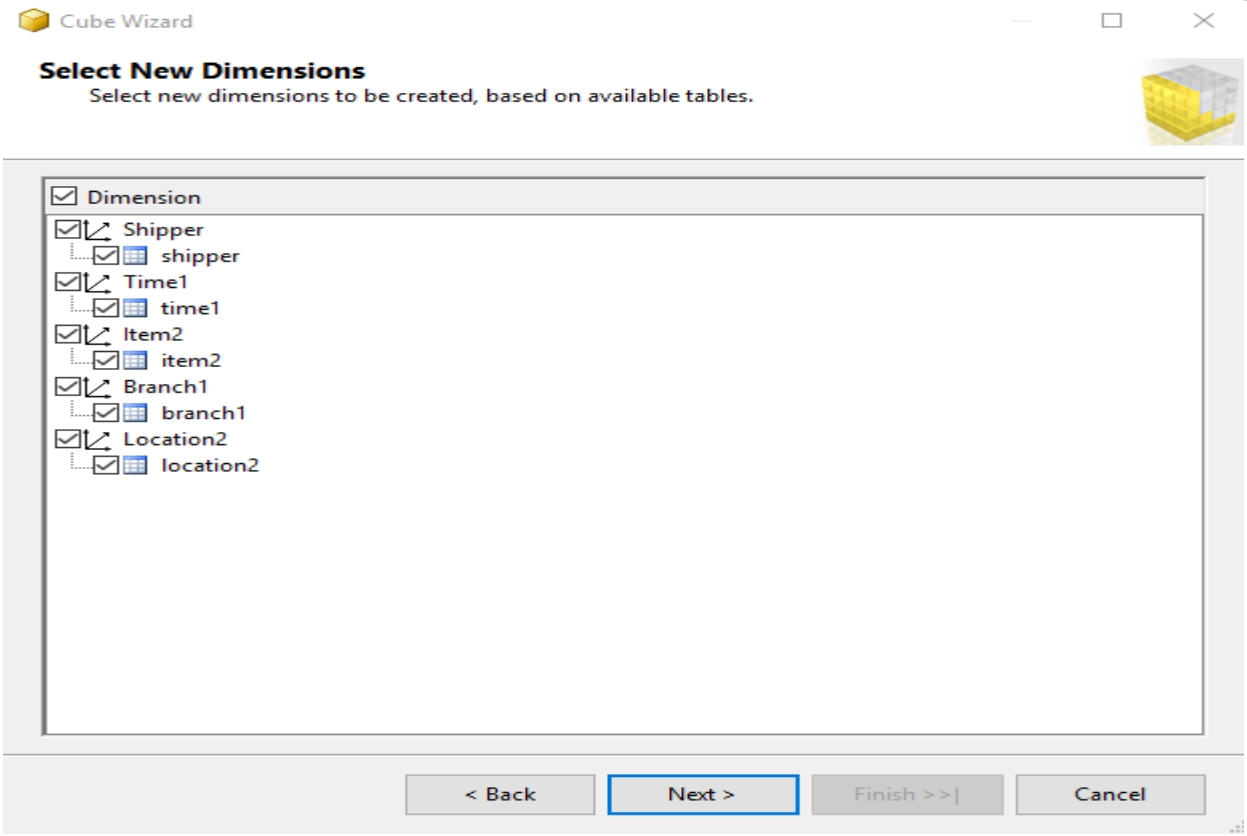

fact_constellation
Data Sources
b4.ds
Data Source Views
b4.dsv
Cubes
Dimensions
Mining Structures
Roles
Assemblies
Miscellaneous

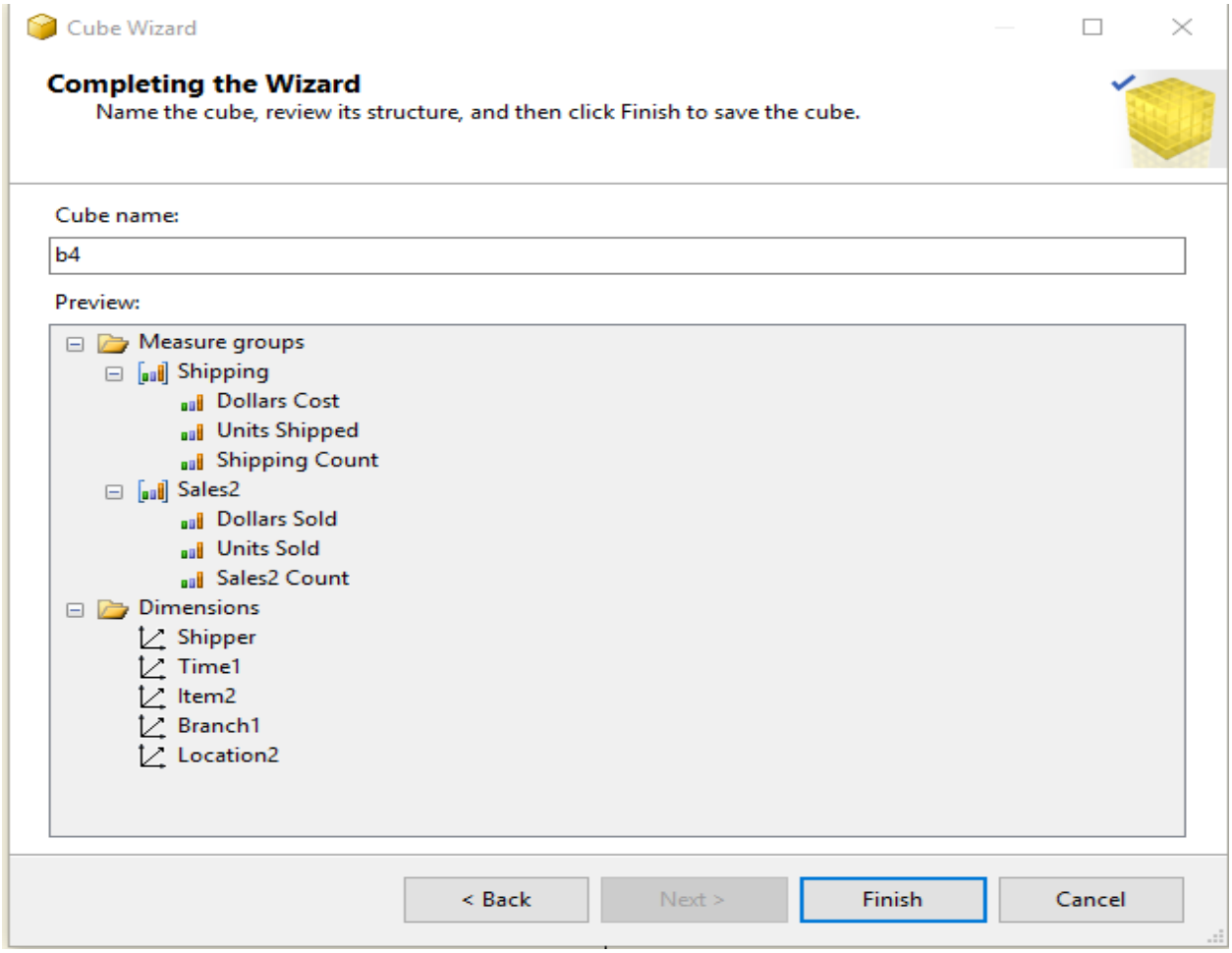

Cube Wizard
Welcome to the Cube Wizard
Use this wizard to create a new cube. First, you select the data source view and tables for the cube, and then you set its properties. You can also opt to create a cube without using a data source.
Don't show this page again
< Back
Next >
Finish >>|
Cancel


Cube Wizard
Select Creation Method
Cubes can be created by using existing tables, creating an empty cube, or generating tables in the data source.
How would you like to create the cube?
Use existing tables
Create an empty cube
Generate tables in the data source
Template:
(None)
Description:
Create a cube based on one or more tables in a data source.
< Back
Next >
Finish >>|
Cancel

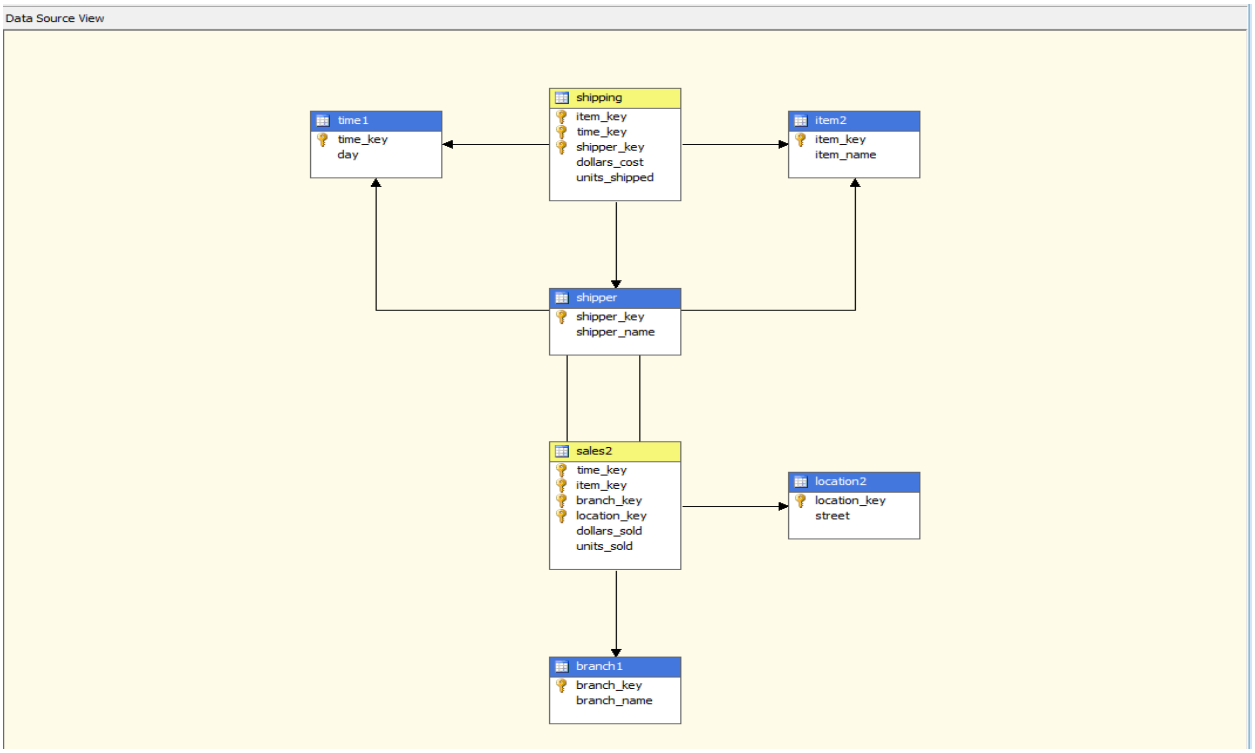

Cube Wizard
Select Measure Group Tables
Select a data source view or diagram and then select the tables that will be used for measure groups.
Data source view:
b4
Measure group tables:
Suggest
time1
branch1
item2
location2
shipper
sales2
shipping
< Back
Next >
Finish >>|
Cancel


Cube Wizard
Select Measures
Select measures that you want to include in the cube.
Measure
Shipping
Dollars Cost
Units Shipped
Shipping Count
Sales2
Dollars Sold
Units Sold
Sales2 Count
< Back
Next >
Finish >>|
Cancel


Cube Wizard
Select New Dimensions
Select new dimensions to be created, based on available tables.
Dimension
Shipper
shipper
Time1
time1
Item2
item2
Branch1
branch1
Location2
location2
< Back
Next >
Finish >>|
Cancel



Output for fact constellation schema:



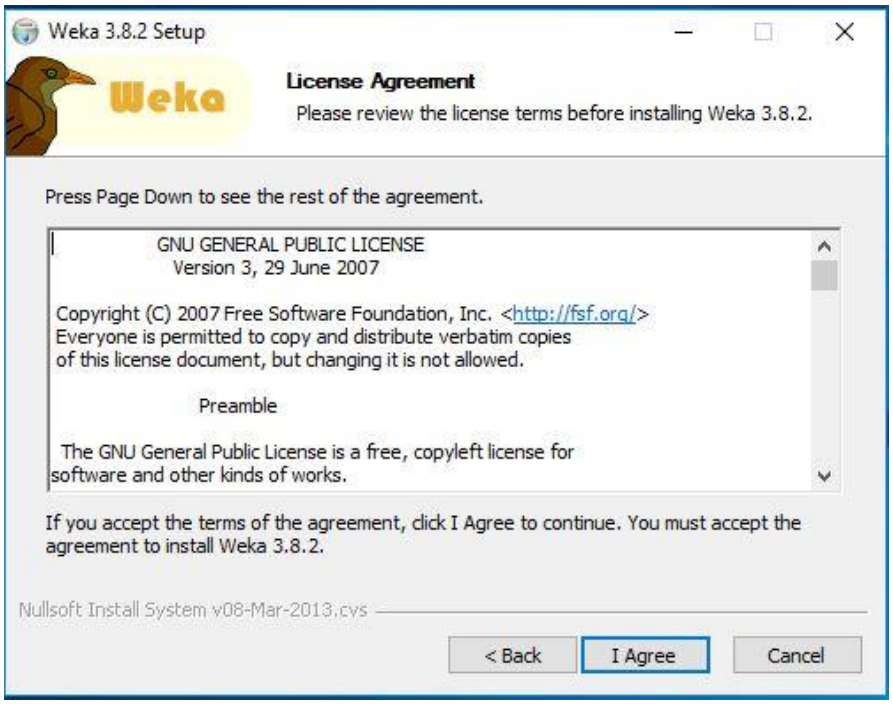
Task:5

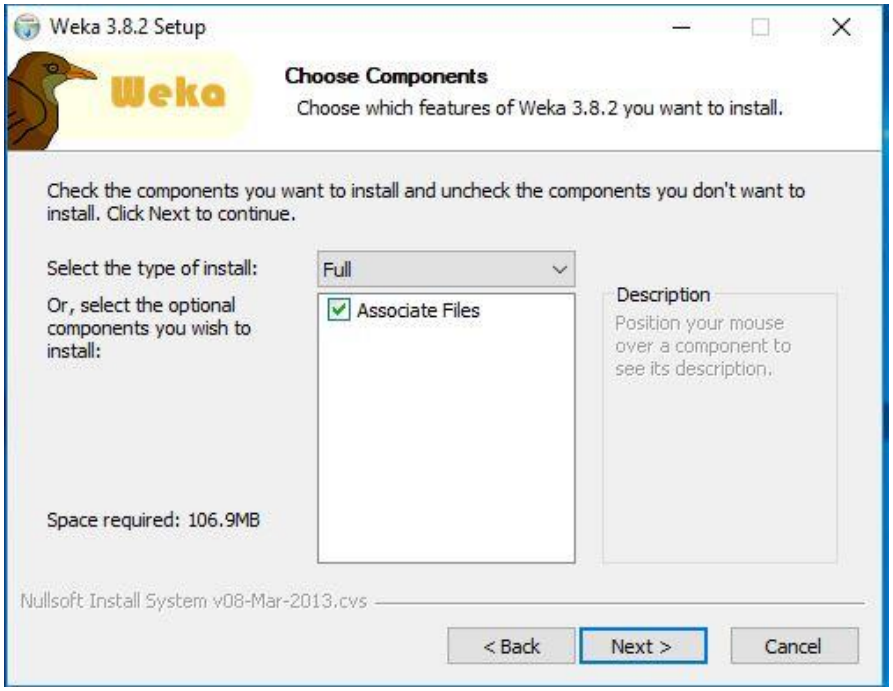
Aim: To download weka tool.

Procedure:

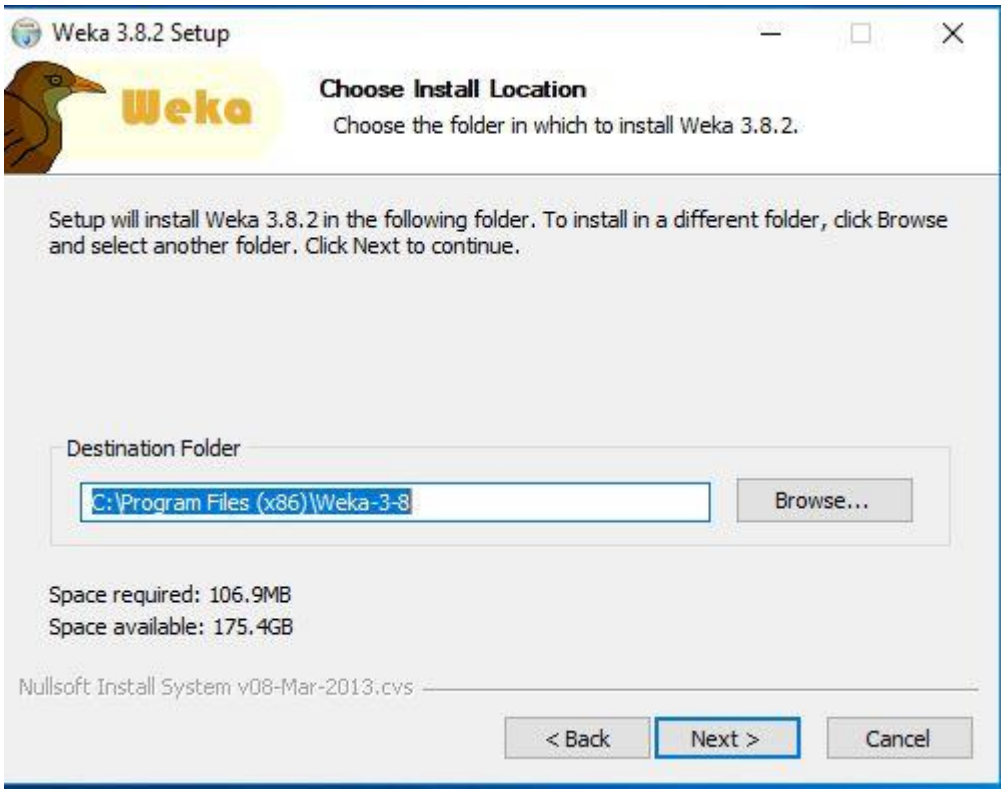
1. Download the weka 3.8.2 from the link and double click on it.A dialogue box appears, click next.

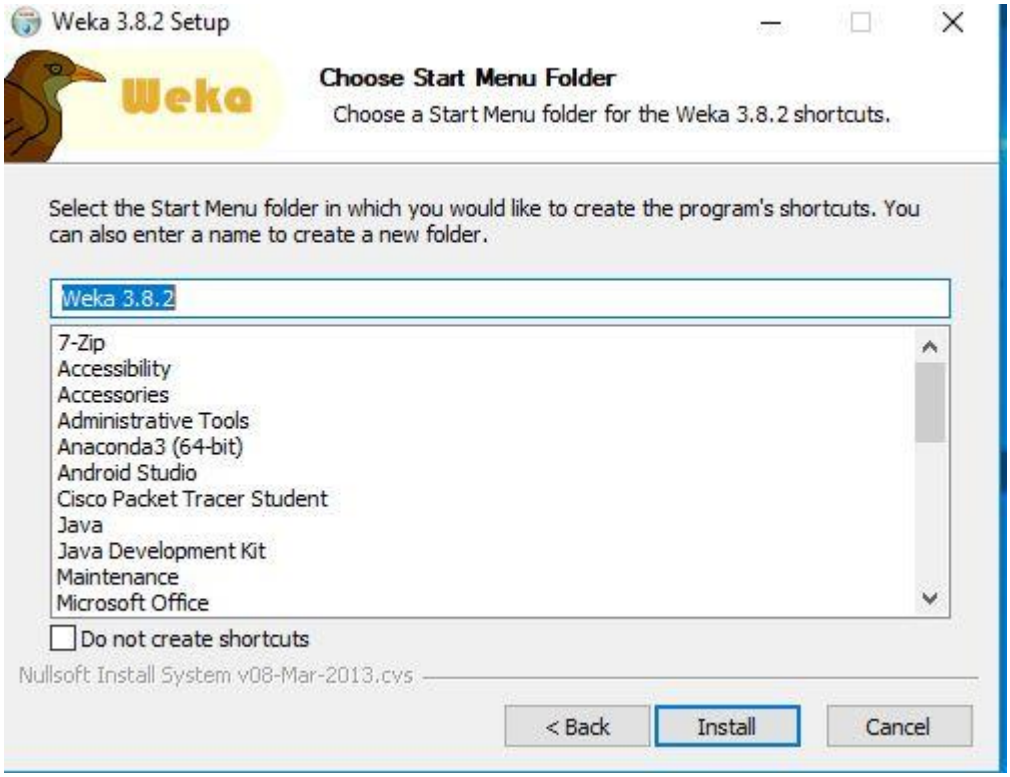
2. Agree the license Agreement.Now select the components like “Associate files”,”Install JRE” and then click on next.

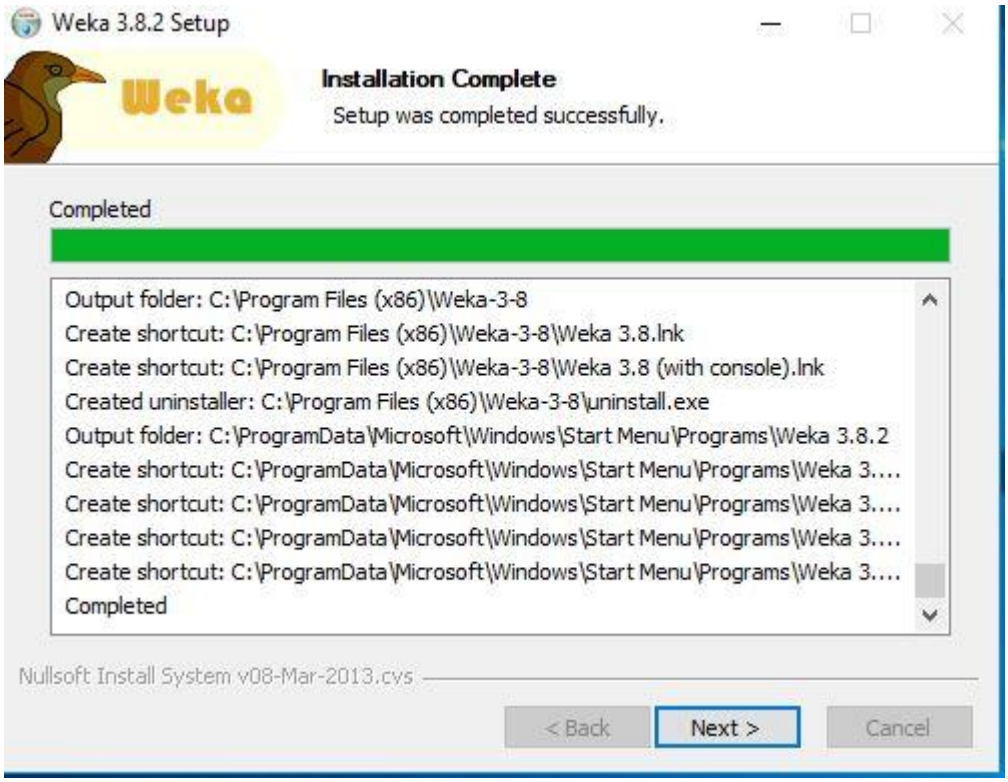


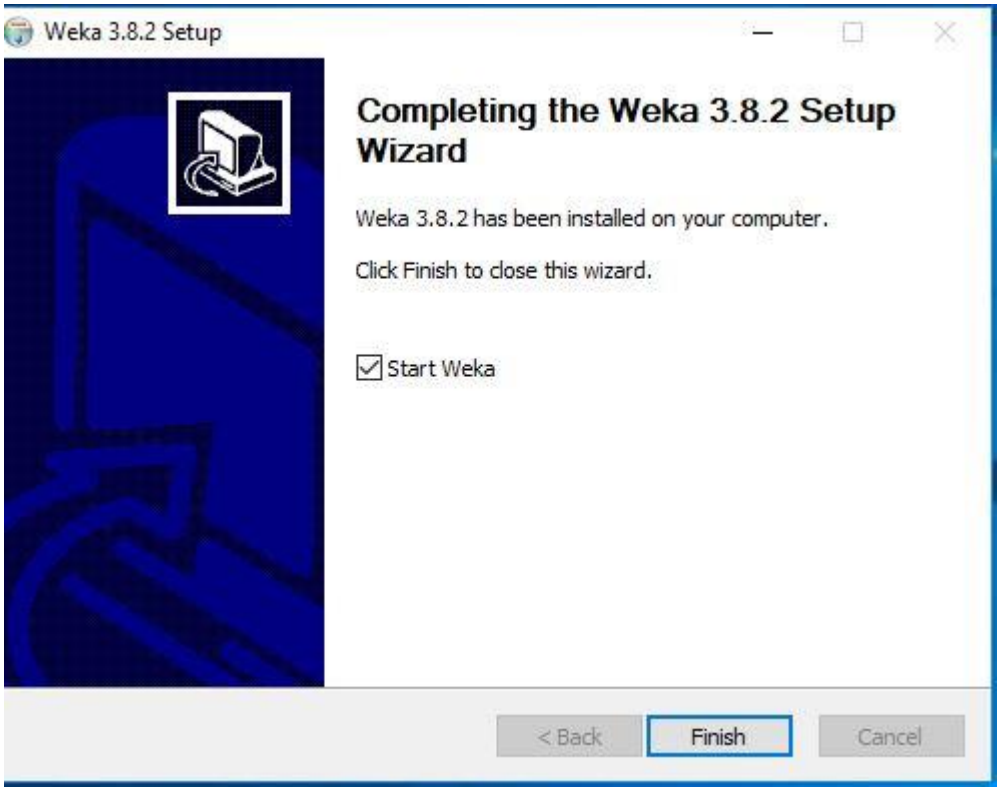


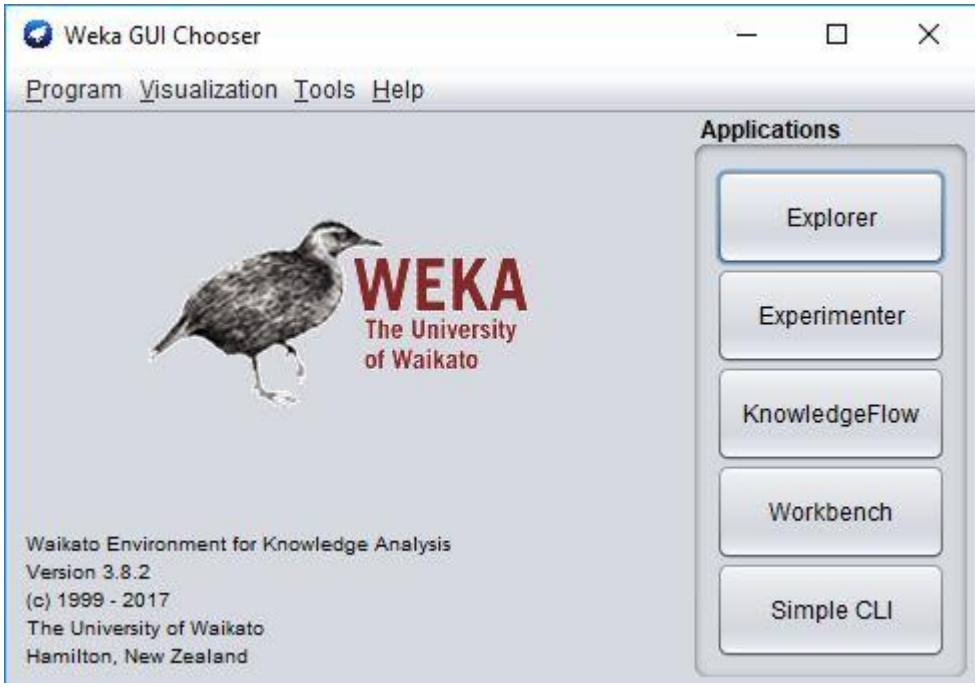
3. Now choose install location and then click on next.Now, choose a start menu folder for the weka 3.8.2 shortcuts and then click on Install.At last weka is succesfully installed.




Weka 3.8.2 Setup
Weka
Choose Start Menu Folder
Choose a Start Menu folder for the Weka 3.8.2 shortcuts.
Select the Start Menu folder in which you would like to create the program's shortcuts. You can also enter a name to create a new folder.
Weka 3.8.2
7-Zip
Accessibility
Accessories
Administrative Tools
Anaconda3 (64-bit)
Android Studio
Cisco Packet Tracer Student
Java
Java Development Kit
Maintenance
Microsoft Office
Do not create shortcuts
Nullsoft Install System v08-Mar-2013.cvs
< Back
Install
Cancel


Weka 3.8.2 Setup
Weka
Installation Complete
Setup was completed successfully.
Completed
Output folder: C:\Program Files (x86)\Weka-3-8
Create shortcut: C:\Program Files (x86)\Weka-3-8\Weka 3.8.lnk
Create shortcut: C:\Program Files (x86)\Weka-3-8\Weka 3.8 (with console).lnk
Created uninstaller: C:\Program Files (x86)\Weka-3-8\uninstall.exe
Output folder: C:\ProgramData\Microsoft\Windows\Start Menu\Programs\Weka 3.8.2
Create shortcut: C:\ProgramData\Microsoft\Windows\Start Menu\Programs\Weka 3....
Create shortcut: C:\ProgramData\Microsoft\Windows\Start Menu\Programs\Weka 3....
Create shortcut: C:\ProgramData\Microsoft\Windows\Start Menu\Programs\Weka 3....
Create shortcut: C:\ProgramData\Microsoft\Windows\Start Menu\Programs\Weka 3....
Completed
Nullsoft Install System v08-Mar-2013.cvs
< Back
Next >
Cancel





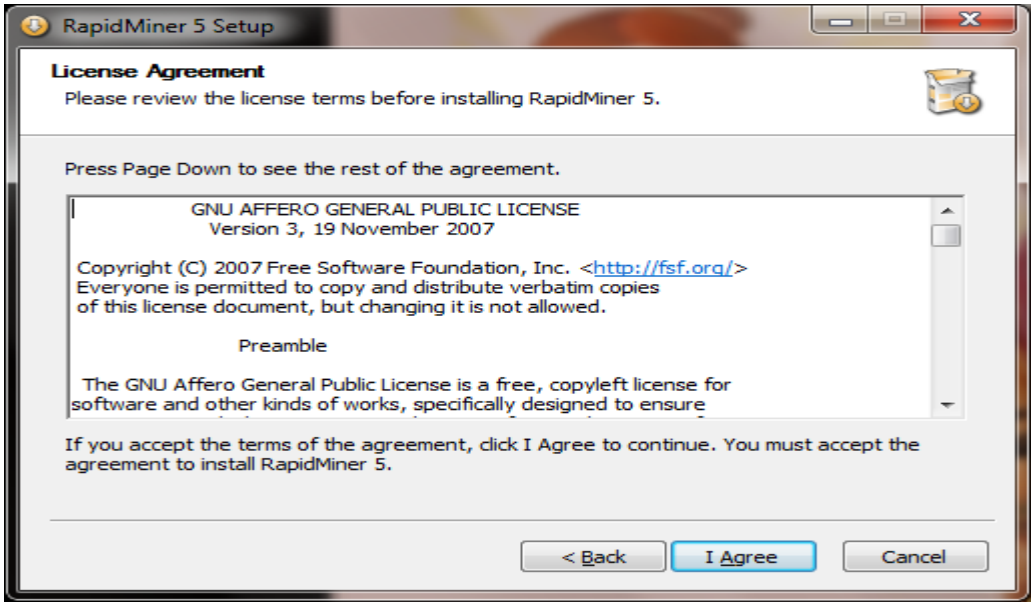
Result: Weka Explorer is installed successfully.

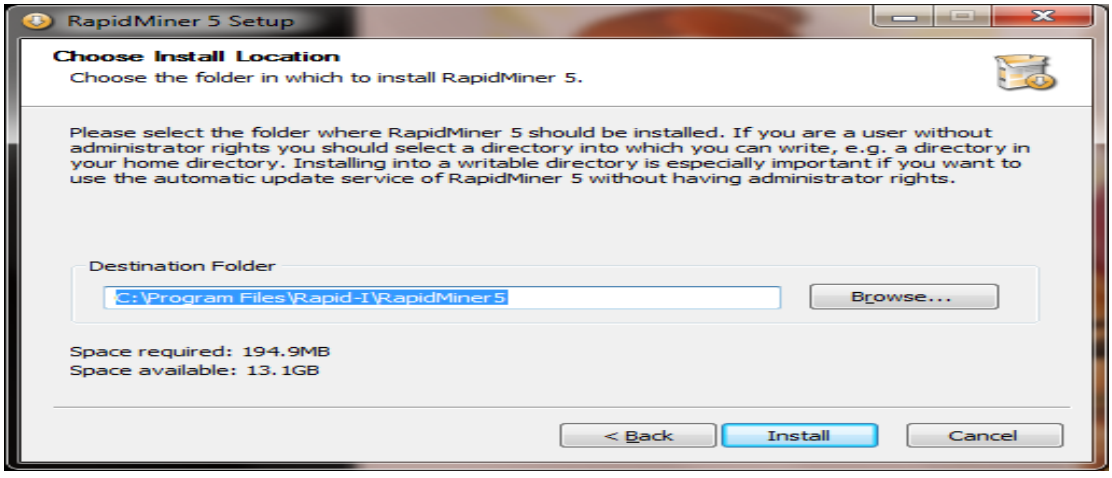
Aim: To install Rapid miner tool.

Procedure:

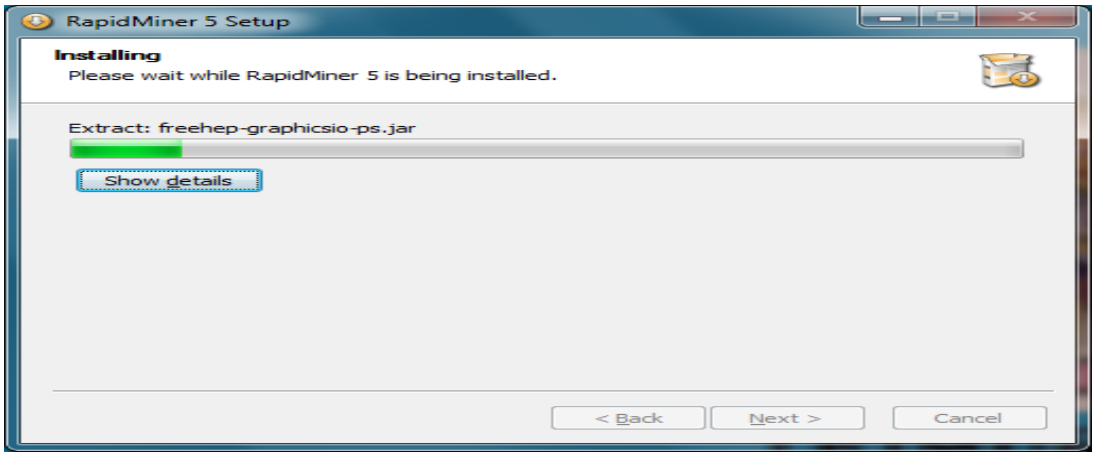
1. First download the Rapid miner and now open the setup and click next.

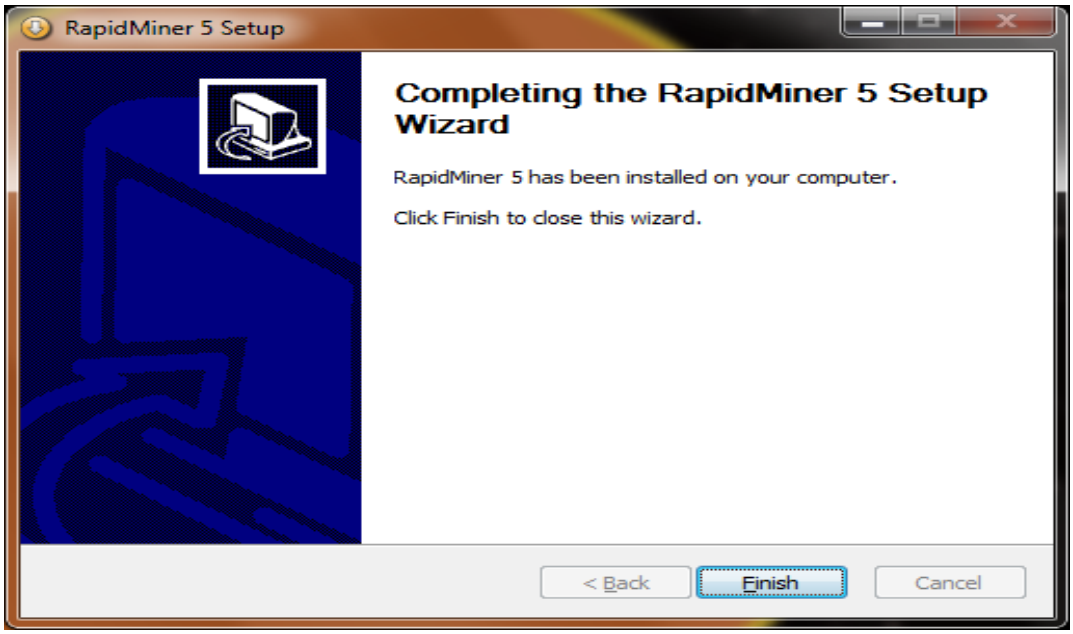
2. Agree the License and set path to store the software and click install.





3. Now the installation process starts, after completion of that click next and then click on “Finish” to finish the installation.





Result: Rapid miner is installed successfully.

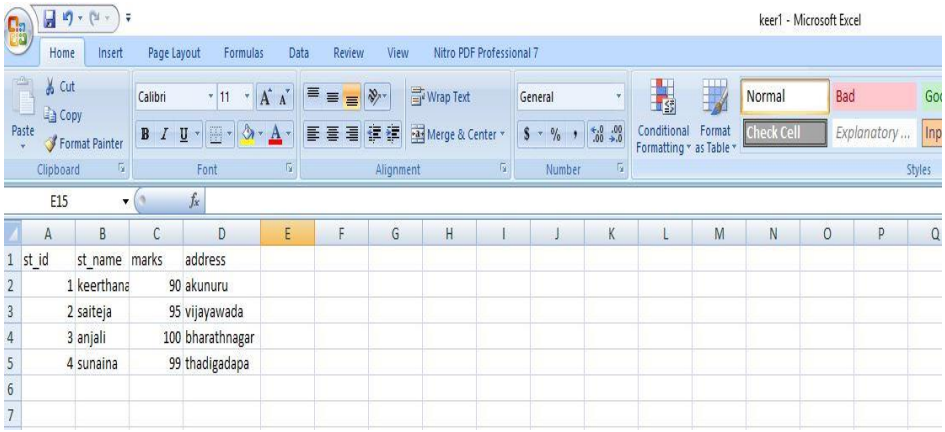
Task-6

Aim: To convert CSV file into ARFF file.

Description: A CSV is a comma separated values file, which allows data to be saved in a table structured format. CSVs look like a garden-variety spreadsheet but with a .csv extension (Traditionally they take the form of a text file containing information separated by commas, hence the name).

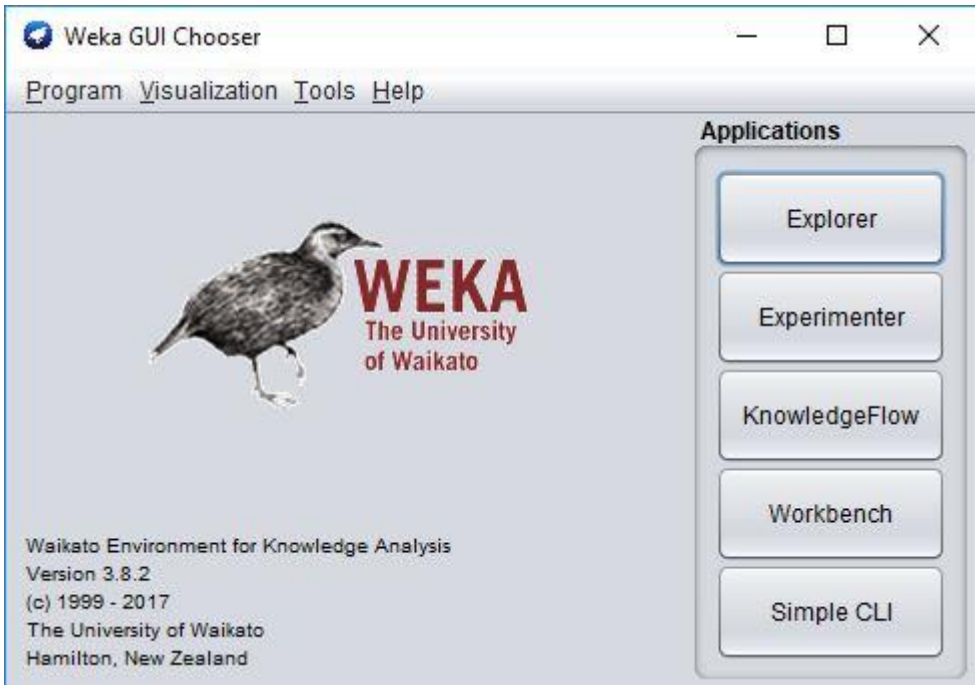
Procedure:

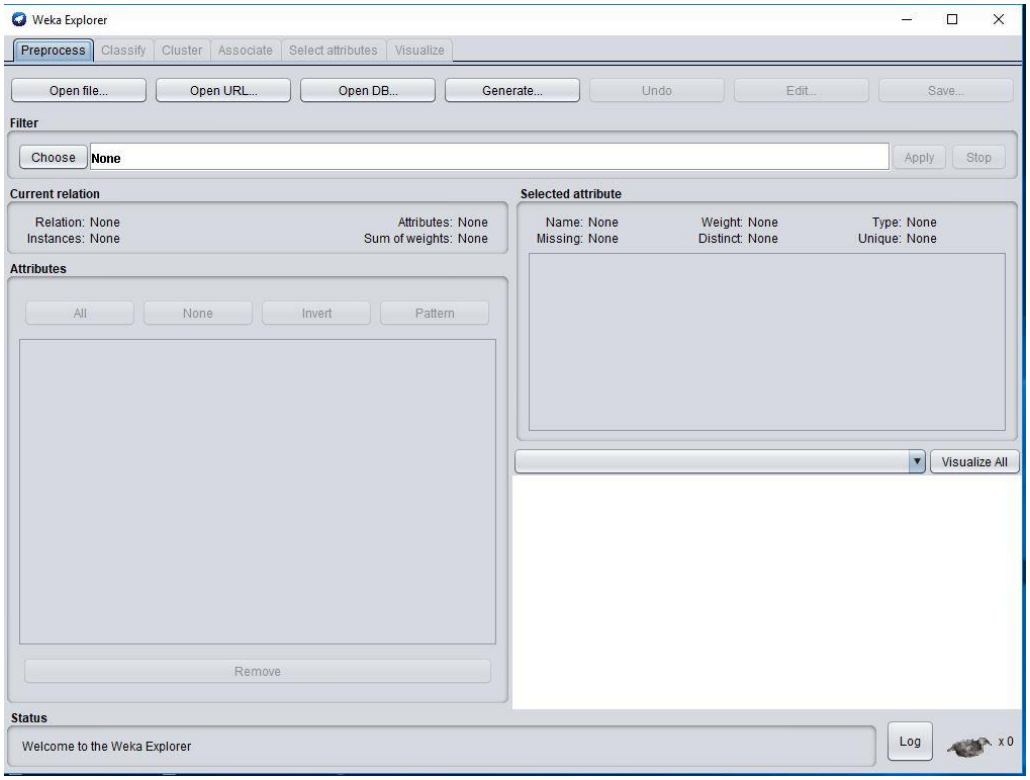
1. First create a csv file.

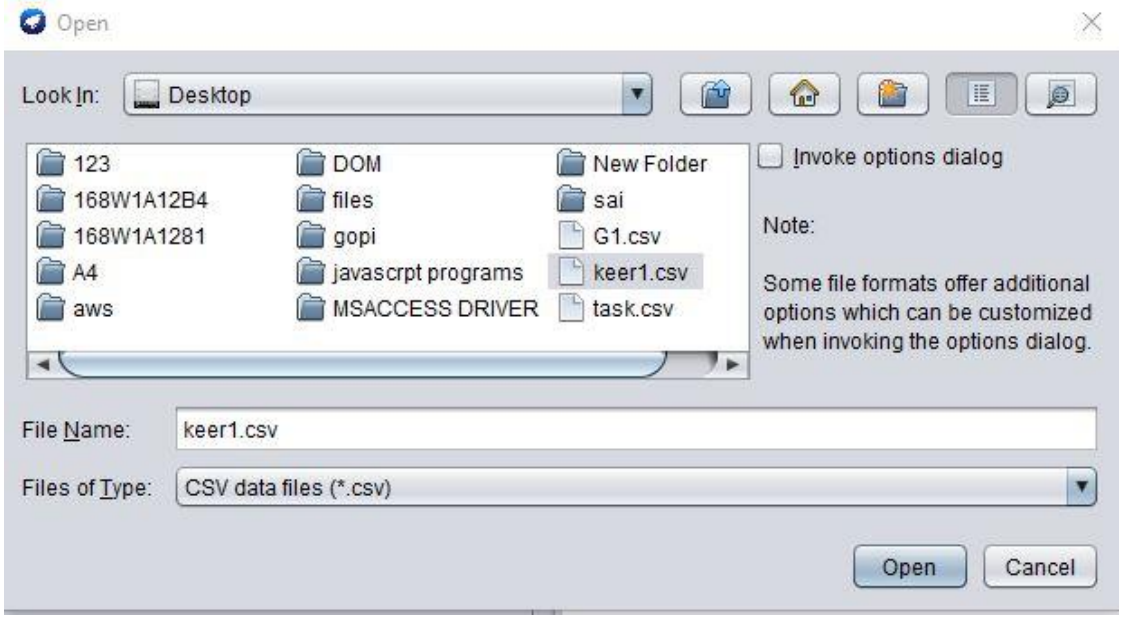


	A	B	C	D
1	st_id	st_name	marks	address
2	1	keerthana	90	akunuru
3	2	saiteja	95	vijayawada
4	3	anjali	100	bharathnagar
5	4	sunaina	99	thadigadapa
6				
7				

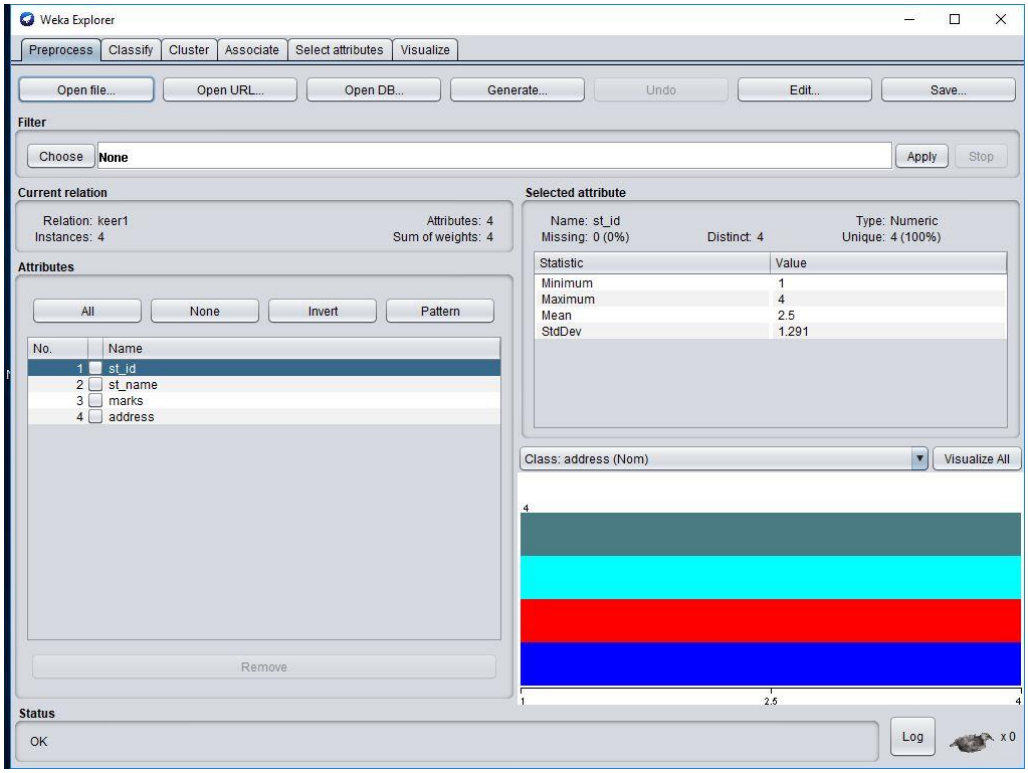
2. Now open this file in weka by using the following steps.

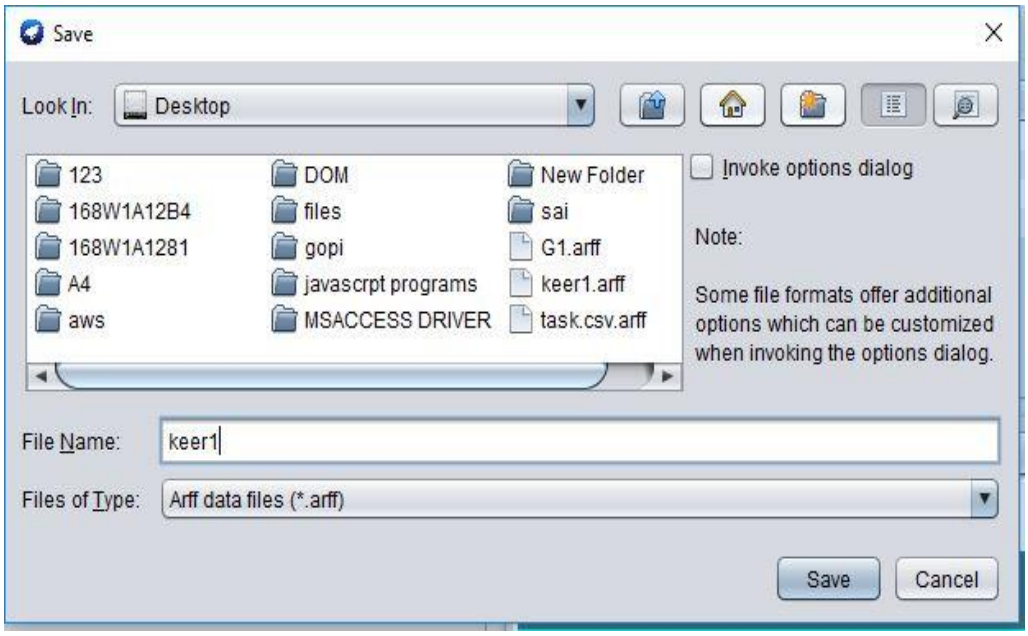






3. Now save the file with .arff extension by clicking on “save” button.





Result: Successfully converted csv file to arff file.

Task-7

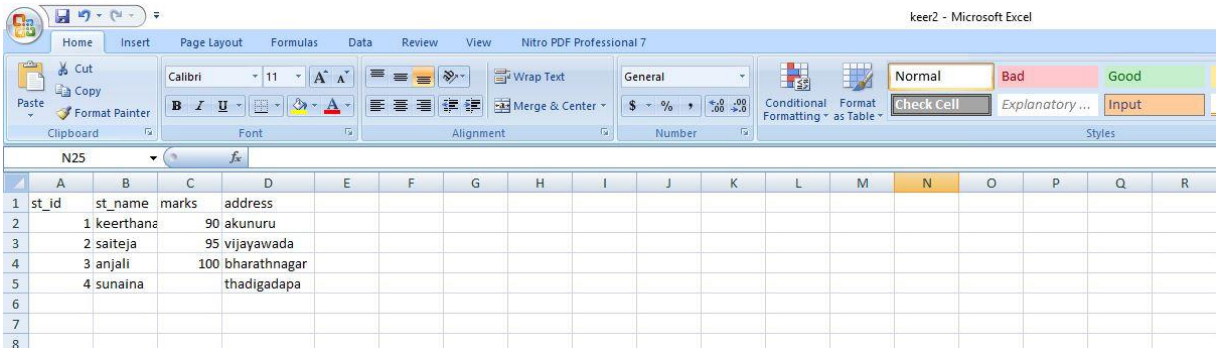
Aim: To apply pre processing techniques on the data in WEKA.

Description:

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

Input file:

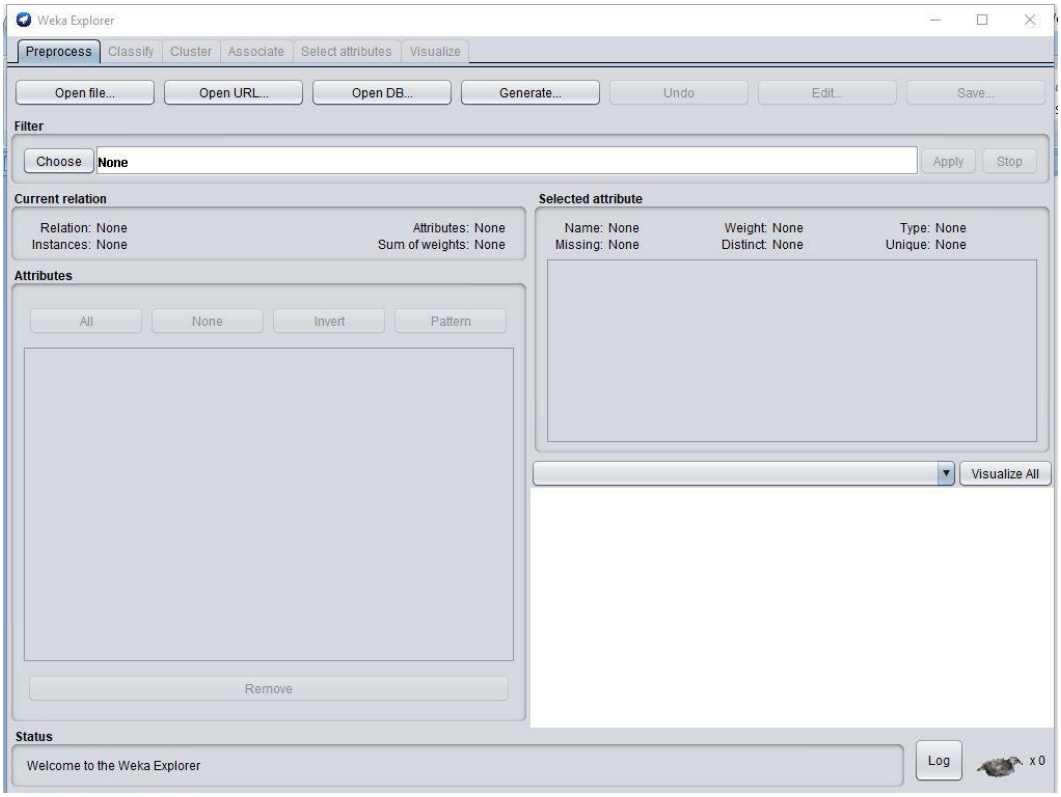
First create a file with.csv extension.

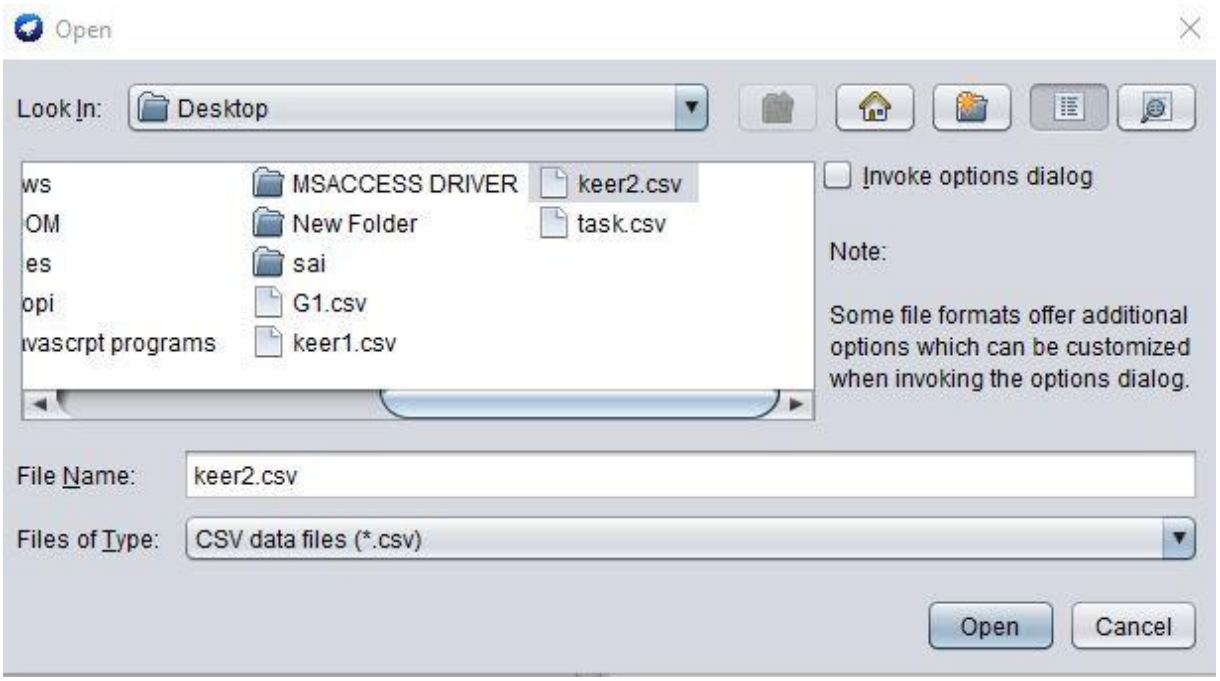


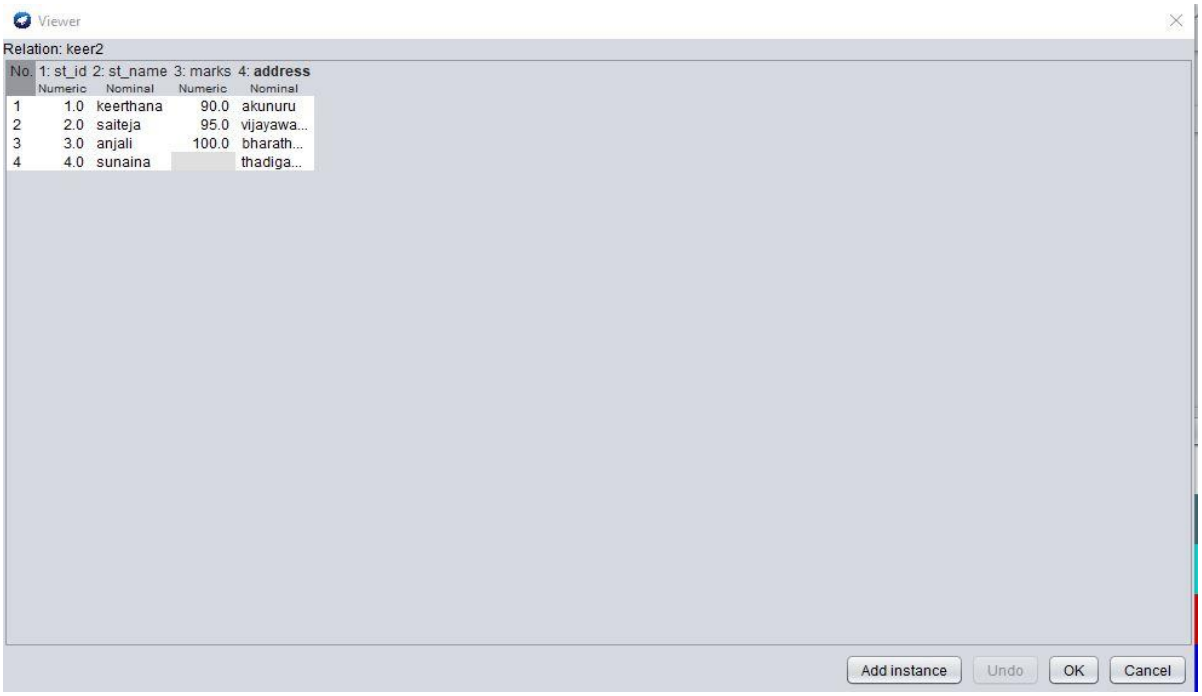
	A	B	C	D
1	st_id	st_name	marks	address
2	1	keerthana	90	akunuru
3	2	saiteja	95	vijayawada
4	3	anjali	100	bharathnagar
5	4	sunaina		thadigadapa
6				
7				

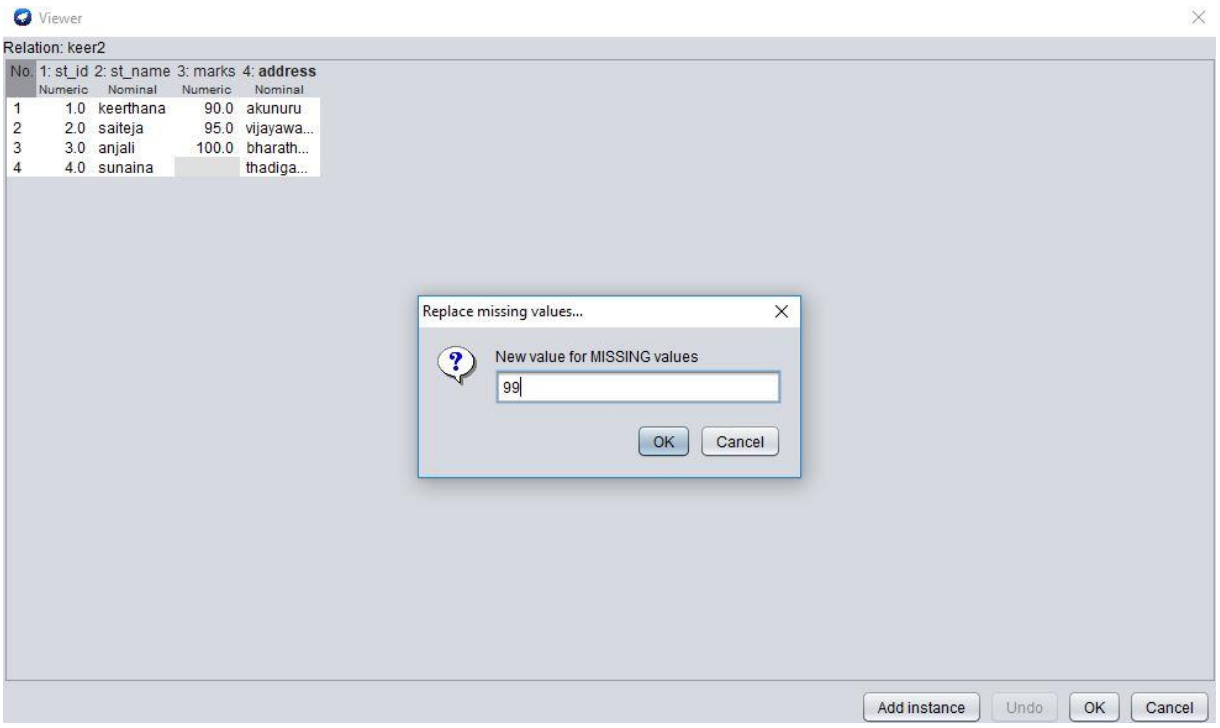
Procedure:

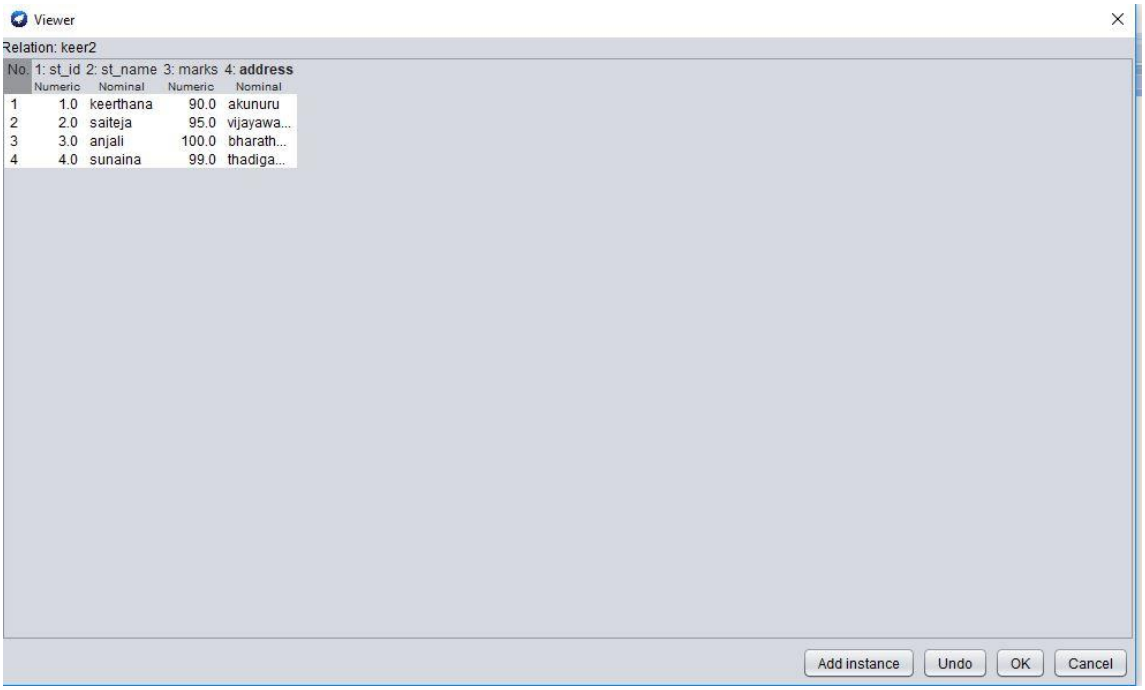
Missing Values: First select the file and then click on “edit” so that viewer is displayed, left click on the empty column and select “set missing values to” and then enter value for it and then click on ok.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: None
Instances: None
Attributes: None
Sum of weights: None
Selected attribute
Name: None
Missing: None
Weight: None
Distinct: None
Type: None
Unique: None
Attributes
All
None
Invert
Pattern
Remove
Visualize All
Status
Welcome to the Weka Explorer
Log
x 0


Open
Look In:
Desktop
ws
OM
es
opi
vascrpt programs
MSACCESS DRIVER
New Folder
sai
G1.csv
keer1.csv
keer2.csv
task.csv
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
File Name:
keer2.csv
Files of Type:
CSV data files (*.csv)
Open
Cancel

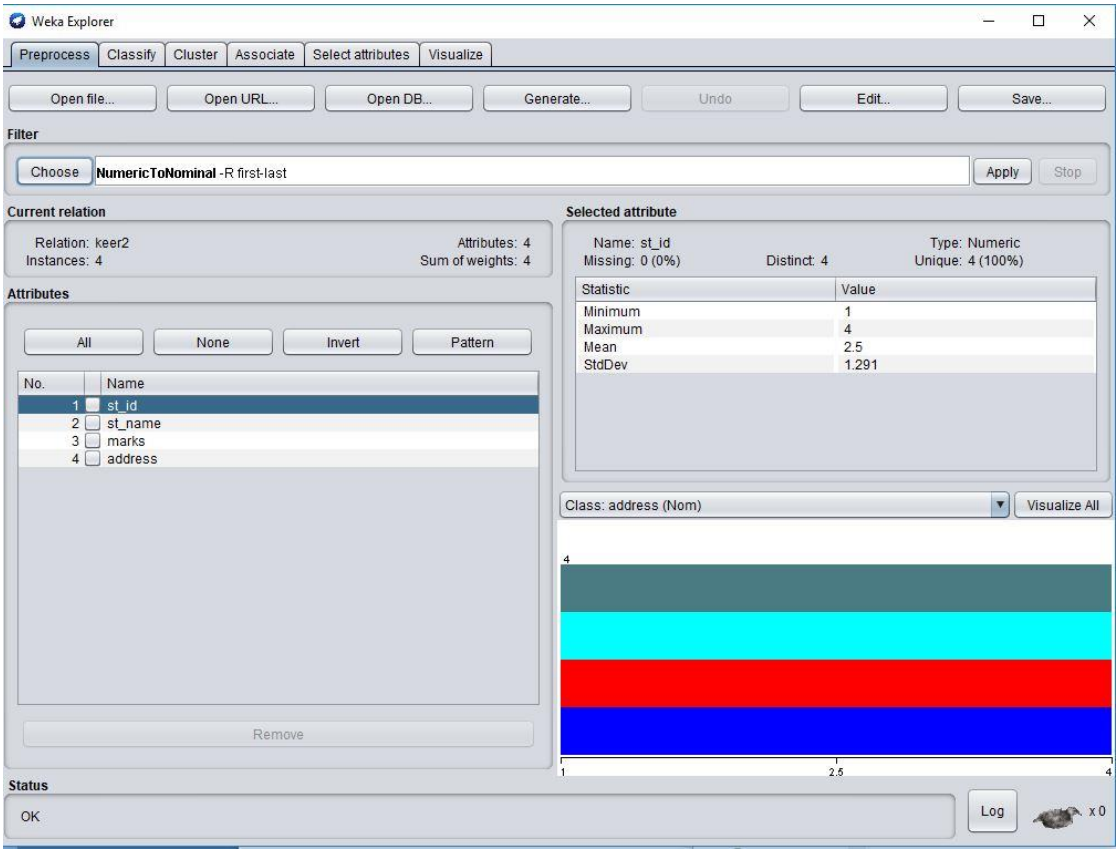

Viewer
Relation: keer2
No. 1: st_id Numeric 2: st_name Nominal 3: marks Numeric 4: address Nominal
1 1.0 keerthana 90.0 akunuru
2 2.0 saiteja 95.0 vijayawa...
3 3.0 anjali 100.0 bharath...
4 4.0 sunaina thadiga...
Add instance
Undo
OK
Cancel

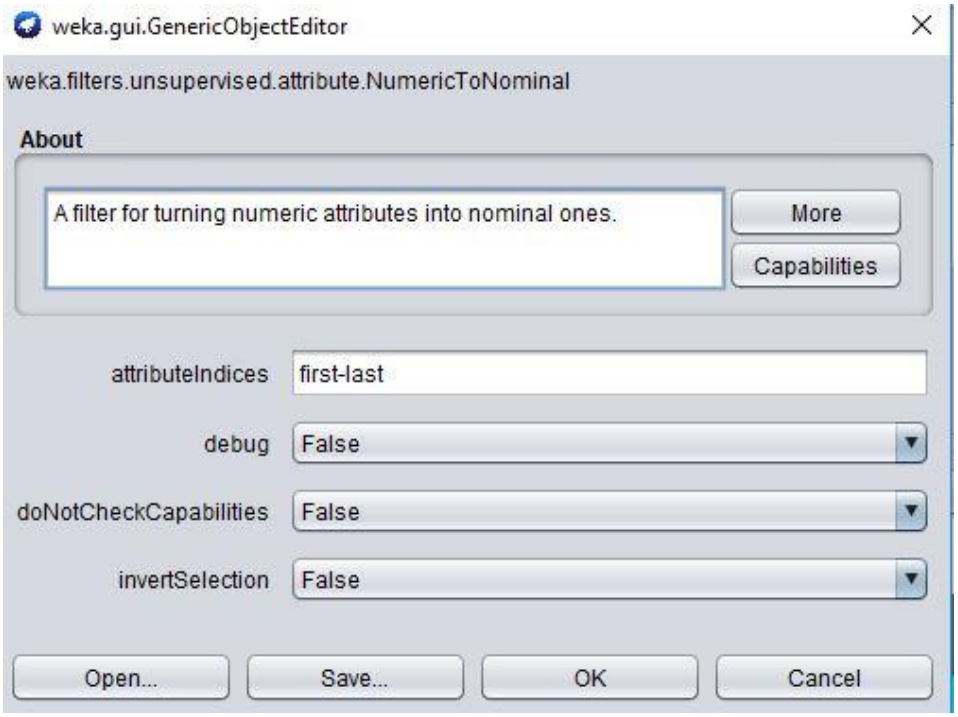

Viewer
Relation: keer2
No. 1: st_id Numeric 2: st_name Nominal 3: marks Numeric 4: address Nominal
1 1.0 keerthana 90.0 akunuru
2 2.0 saiteja 95.0 vijayawa...
3 3.0 anjali 100.0 bharath...
4 4.0 sunaina thadiga...
Replace missing values...
New value for MISSING values
99
OK
Cancel
Add instance
Undo
OK
Cancel

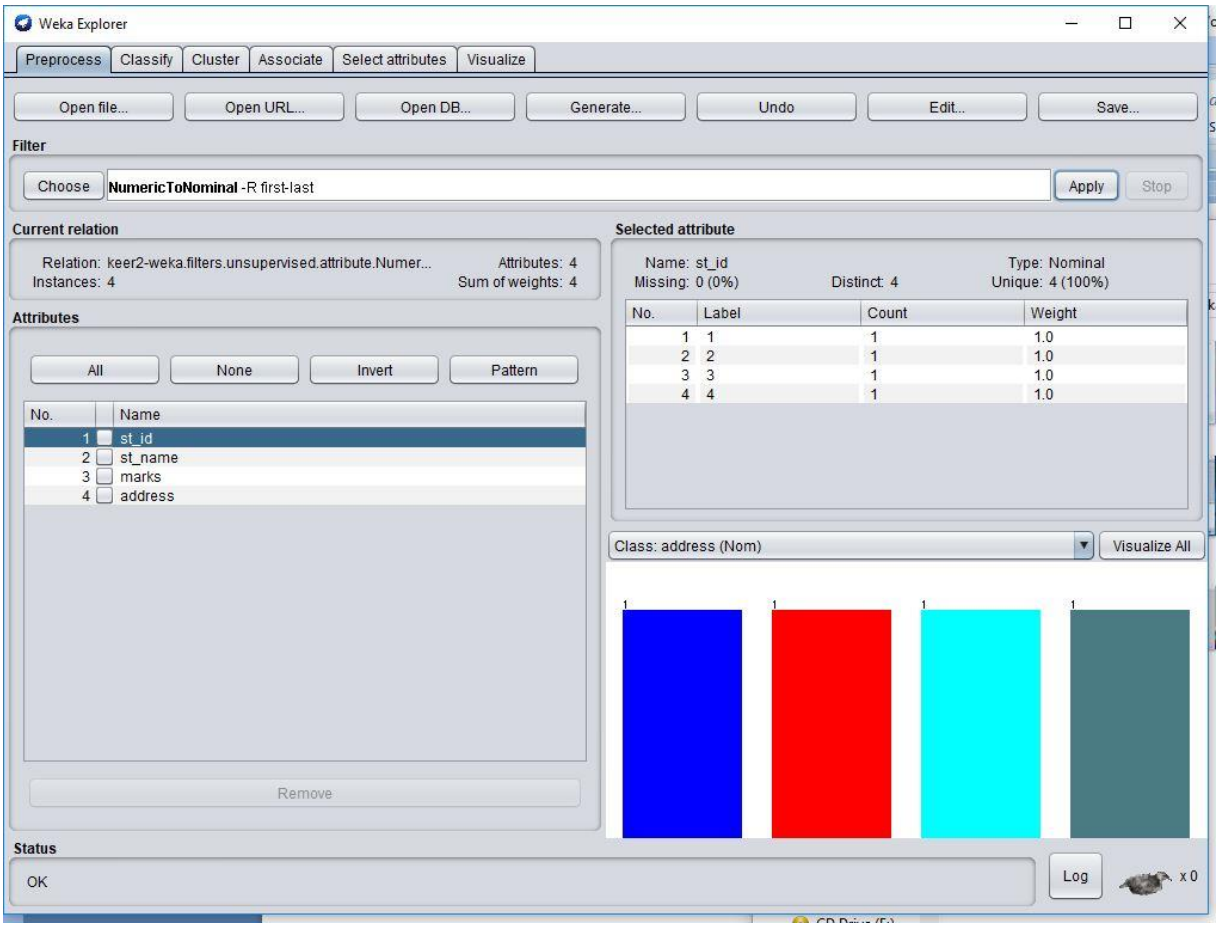


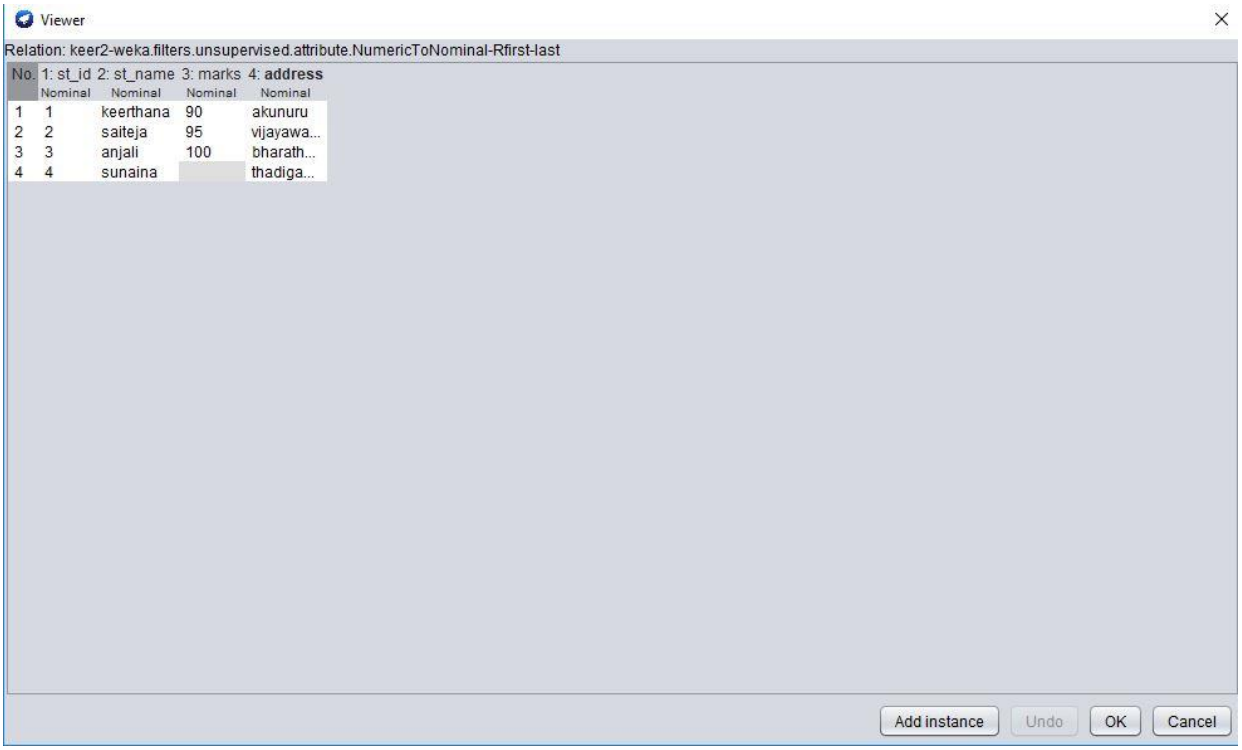
Numeric to Nominal:

After selecting the file then go to choose> filters> unsupervised> attribute> numeric_to_nominal. Then click on textbox beside choose and after verifying it click ok and apply. Later then we can observer output in viewer by clicking on "edit".



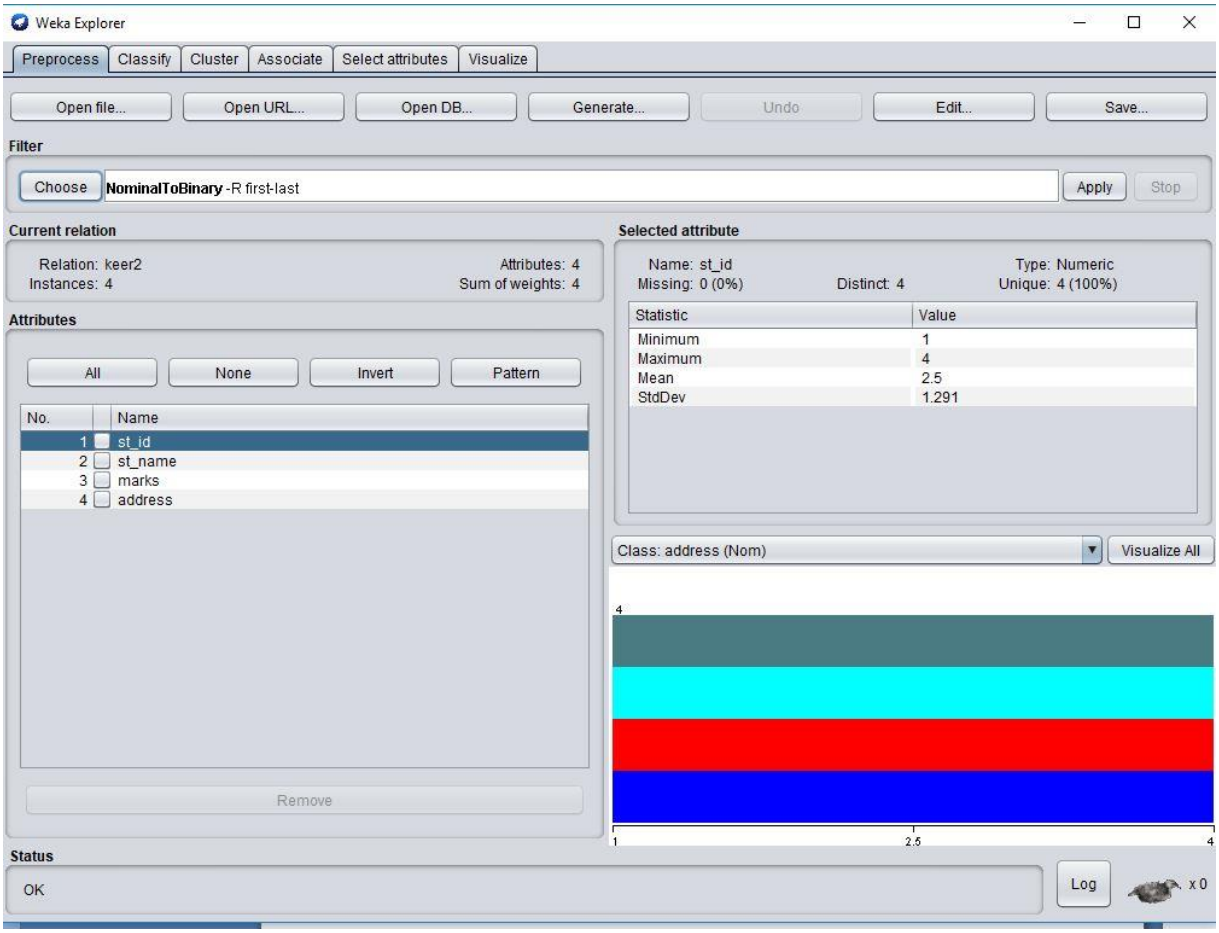

weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.NumericToNominal
About
A filter for turning numeric attributes into nominal ones.
More
Capabilities
attributeIndices
first-last
debug
False
doNotCheckCapabilities
False
invertSelection
False
Open...
Save...
OK
Cancel

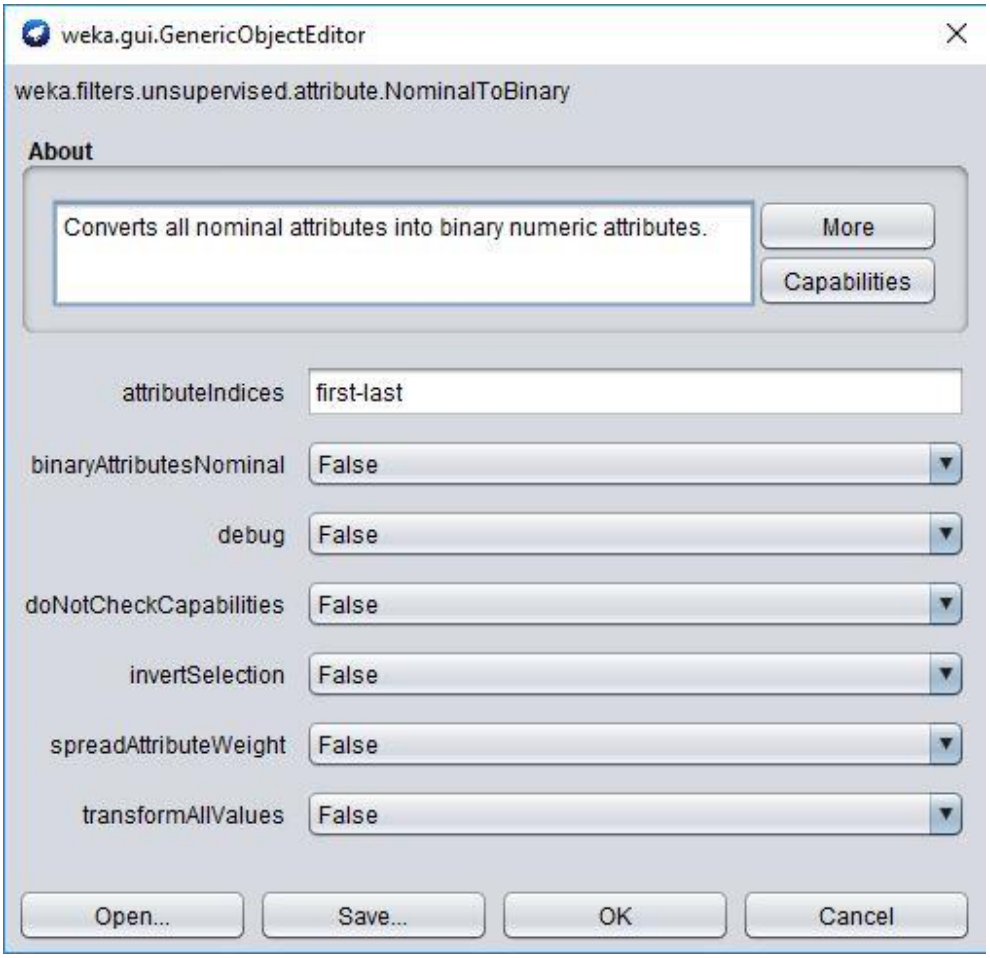

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
NumericToNominal -R first-last
Apply
Stop
Current relation
Relation: keer2-weka.filters.unsupervised.attribute.Numer...
Instances: 4
Attributes: 4
Sum of weights: 4
Selected attribute
Name: st_id
Missing: 0 (0%)
Distinct: 4
Type: Nominal
Unique: 4 (100%)
No. Label Count Weight
1 1 1 1.0
2 2 1 1.0
3 3 1 1.0
4 4 1 1.0
Attributes
All
None
Invert
Pattern
No. Name
1 st_id
2 st_name
3 marks
4 address
Remove
Class: address (Nom)
Visualize All
Status
OK
Log
x 0

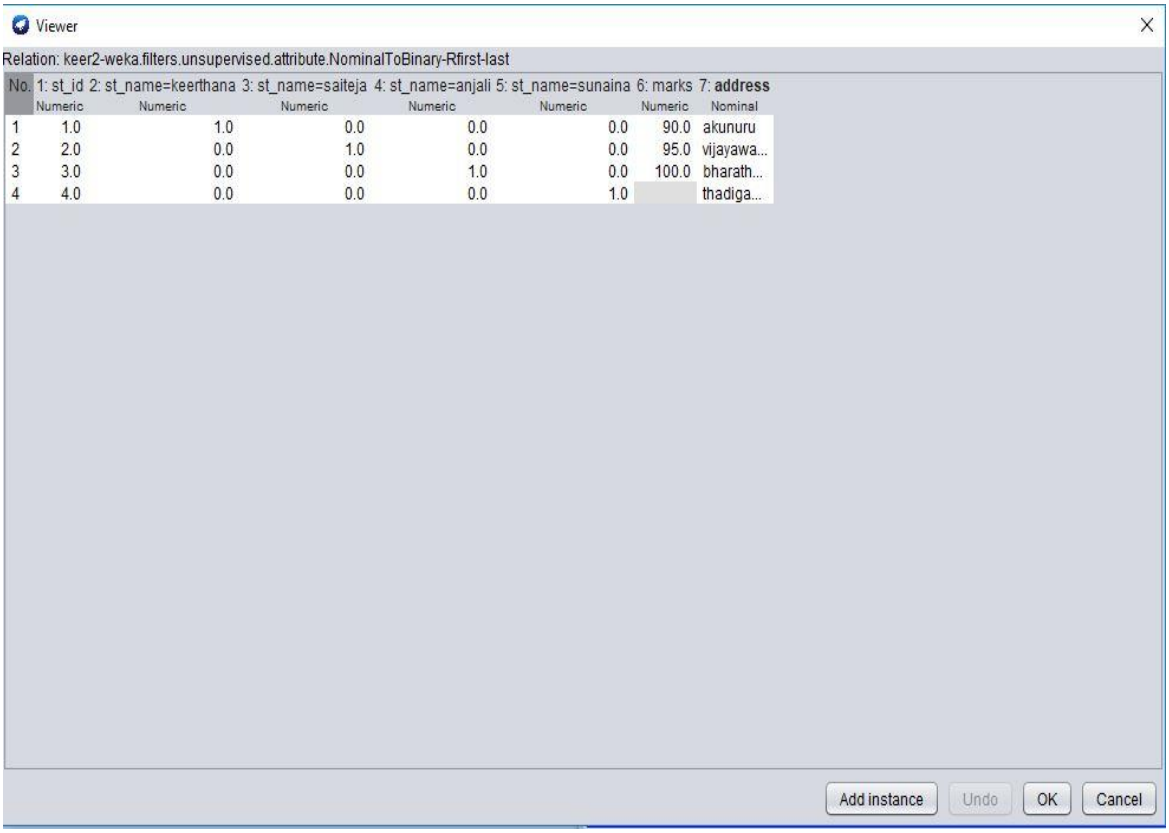


Nominal to Binary:

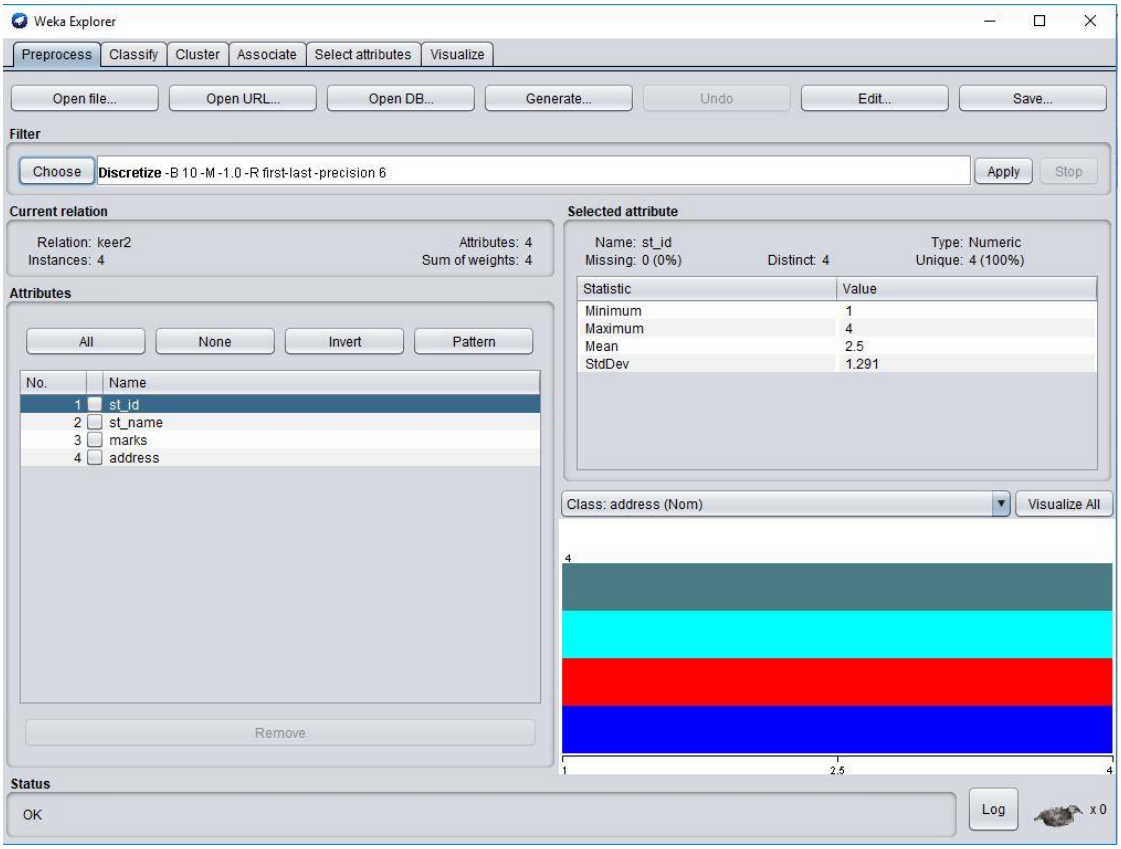
Select nominal to binary.

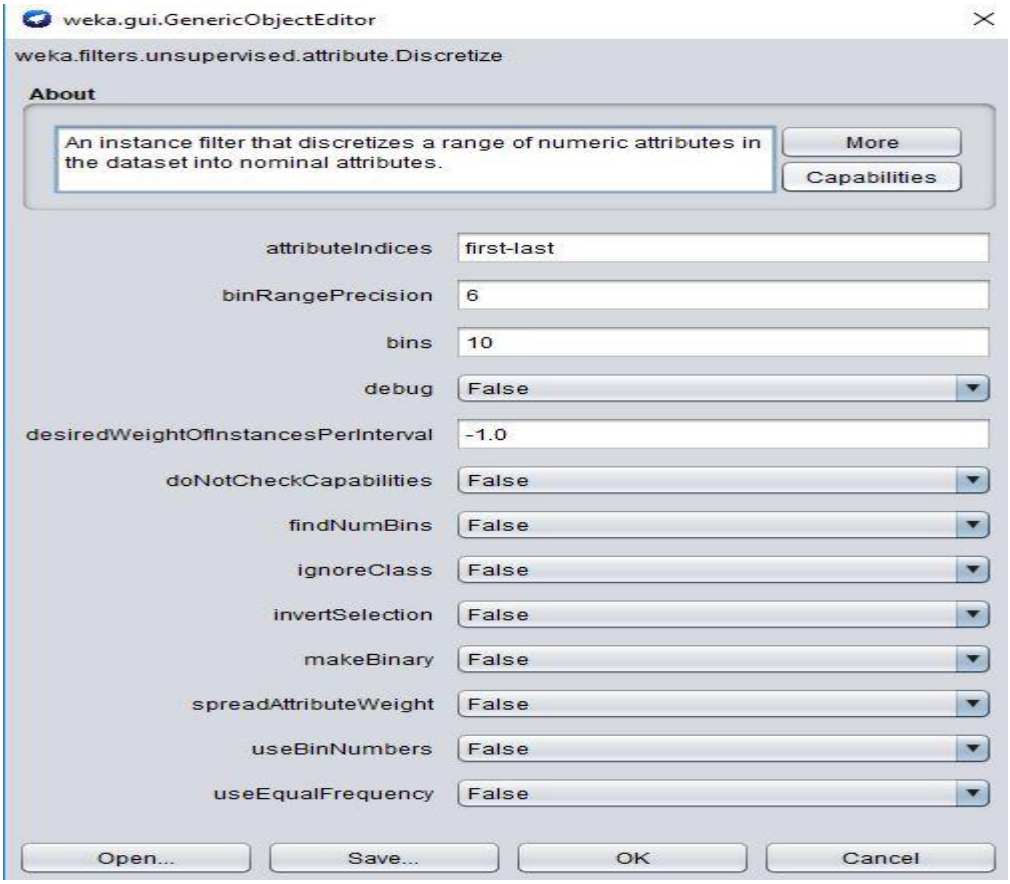


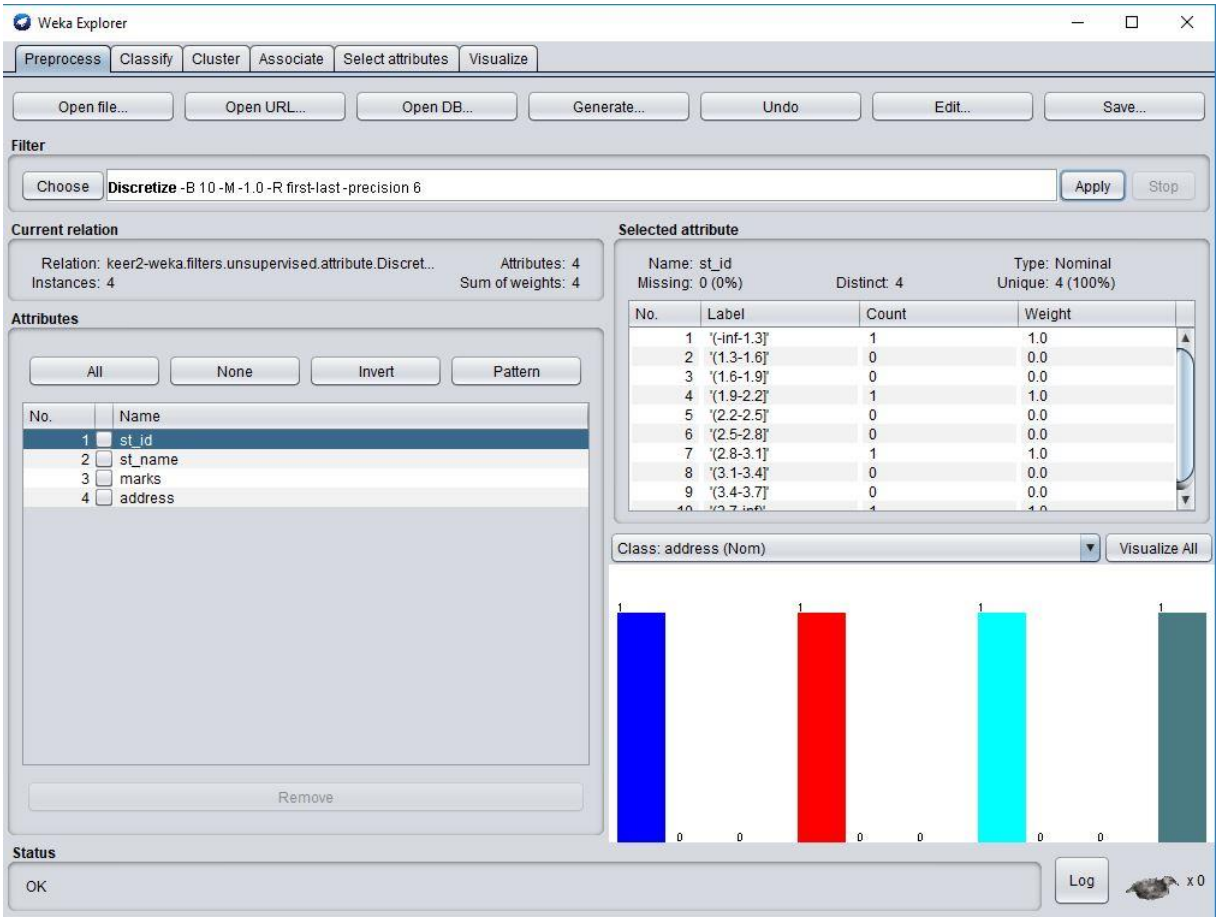

weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.NominalToBinary
About
Converts all nominal attributes into binary numeric attributes.
More
Capabilities
attributeIndices
first-last
binaryAttributesNominal
False
debug
False
doNotCheckCapabilities
False
invertSelection
False
spreadAttributeWeight
False
transformAllValues
False
Open...
Save...
OK
Cancel

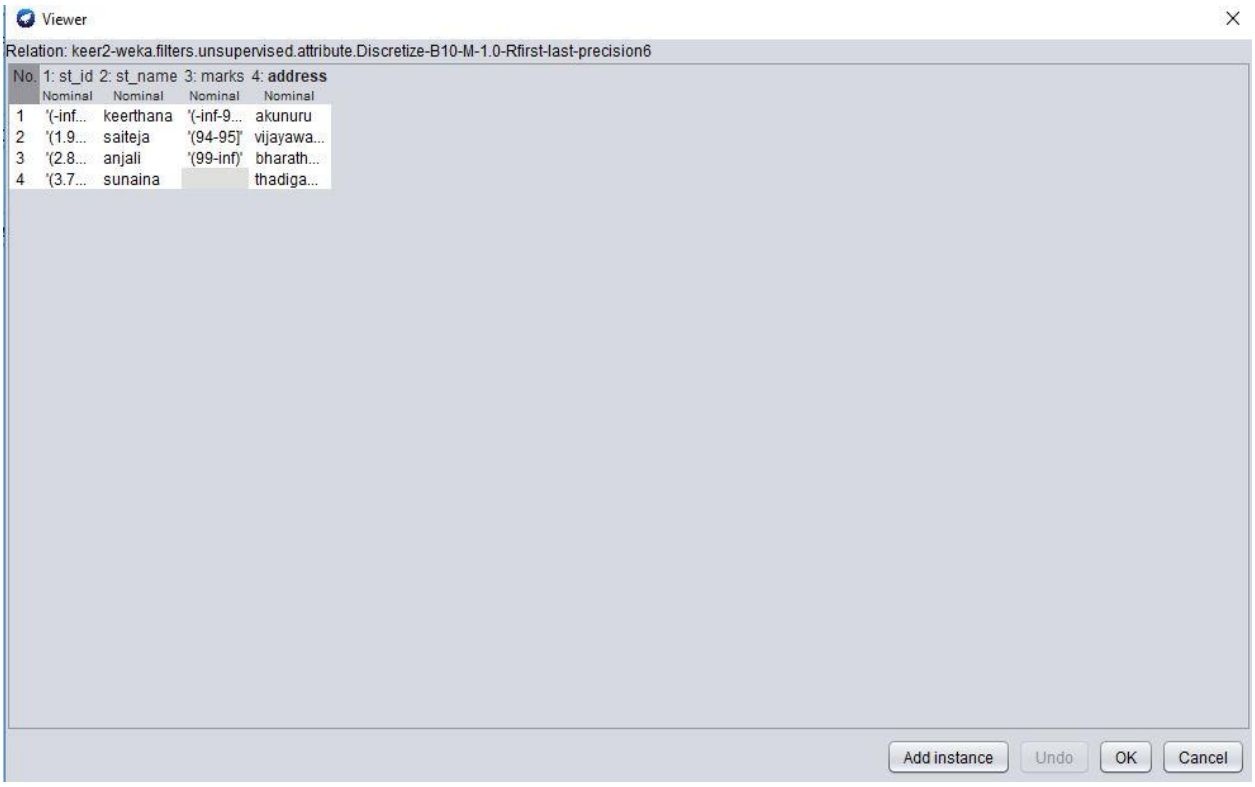

Viewer
Relation: keer2-weka.filters.unsupervised.attribute.NominalToBinary-Rfirst-last
No. 1: st_id 2: st_name=keerthana 3: st_name=saiteja 4: st_name=anjali 5: st_name=sunaina 6: marks 7: address
Numeric Numeric Numeric Numeric Numeric Numeric Nominal
1 1.0 1.0 0.0 0.0 0.0 90.0 akunuru
2 2.0 0.0 1.0 0.0 0.0 95.0 vijayawa...
3 3.0 0.0 0.0 1.0 0.0 100.0 bharath...
4 4.0 0.0 0.0 0.0 1.0 thadiga...
Add instance
Undo
OK
Cancel

Discretize:

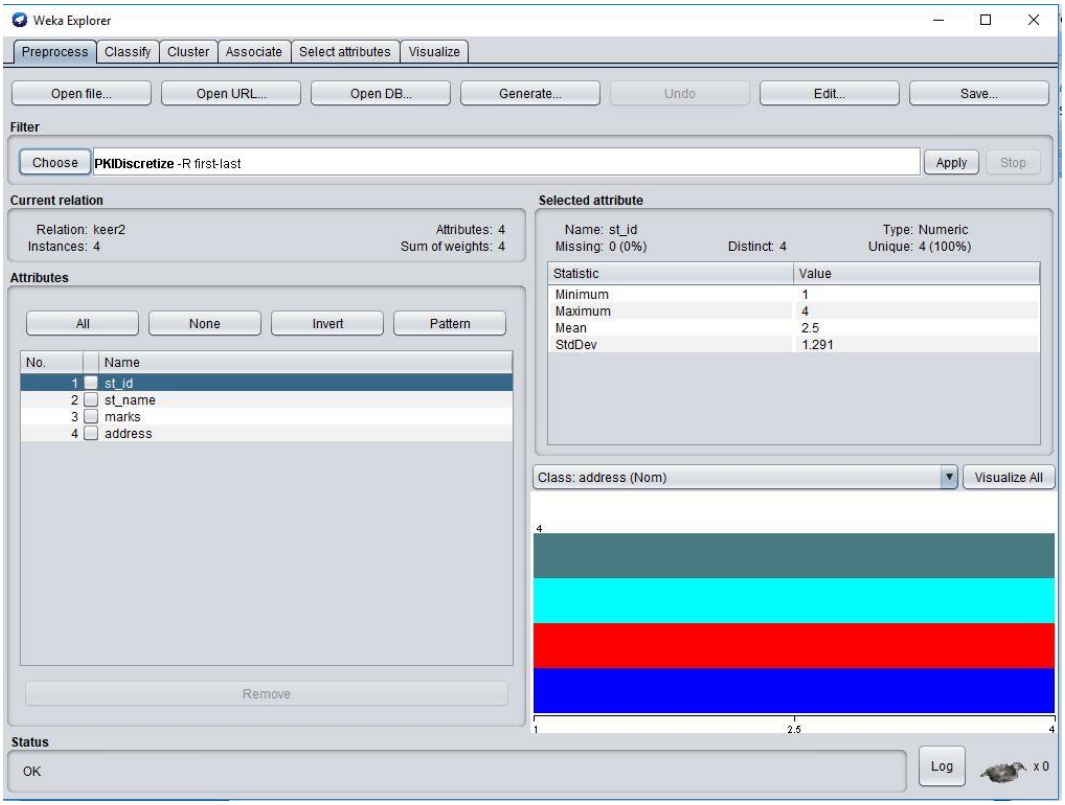

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Discretize -B 10 -M -1.0 -R first-last -precision 6
Apply
Stop
Current relation
Relation: keer2
Instances: 4
Attributes: 4
Sum of weights: 4
Selected attribute
Name: st_id
Missing: 0 (0%)
Distinct: 4
Type: Numeric
Unique: 4 (100%)
Statistic
Value
Minimum
1
Maximum
4
Mean
2.5
StdDev
1.291
Attributes
All
None
Invert
Pattern
No.
Name
1 st_id
2 st_name
3 marks
4 address
Remove
Class: address (Nom)
Visualize All
Status
OK
Log
x 0

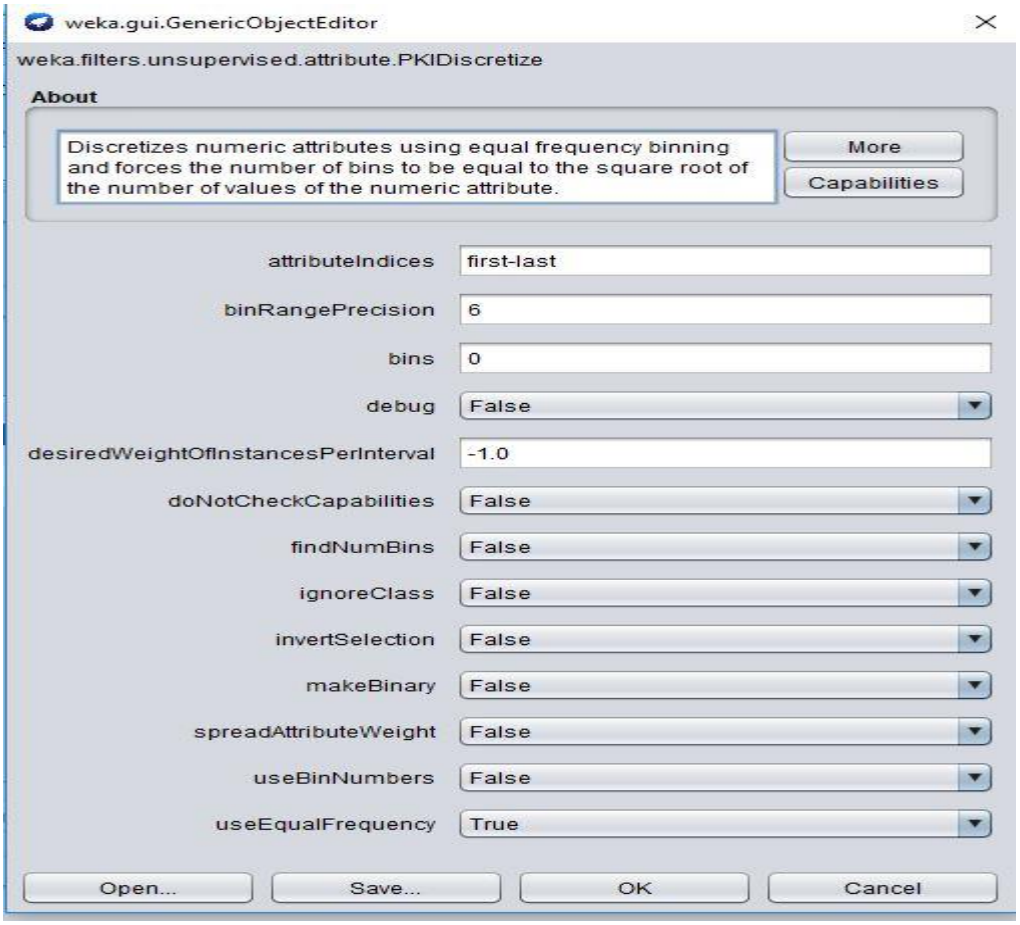

weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Discretize
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
More
Capabilities
attributeIndices
first-last
binRangePrecision
6
bins
10
debug
False
desiredWeightOfInstancesPerInterval
-1.0
doNotCheckCapabilities
False
findNumBins
False
ignoreClass
False
invertSelection
False
makeBinary
False
spreadAttributeWeight
False
useBinNumbers
False
useEqualFrequency
False
Open...
Save...
OK
Cancel

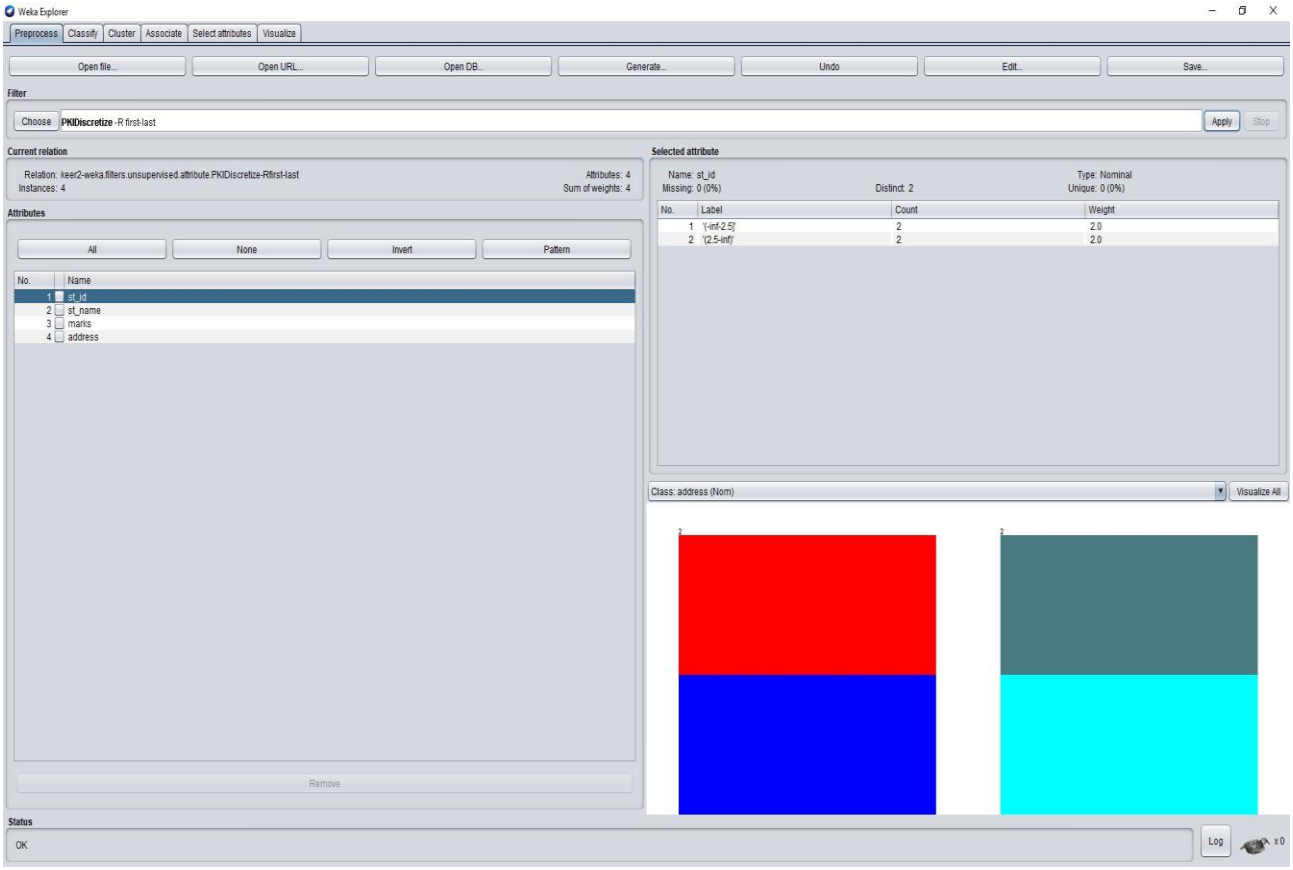

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Discretize -B 10 -M -1.0 -R first-last -precision 6
Apply
Stop
Current relation
Relation: keer2-weka.filters.unsupervised.attribute.Discret...
Attributes: 4
Instances: 4
Sum of weights: 4
Selected attribute
Name: st_id
Type: Nominal
Missing: 0 (0%)
Distinct: 4
Unique: 4 (100%)
No. Label Count Weight
1 '(-inf-1.3]' 1 1.0
2 '(1.3-1.6]' 0 0.0
3 '(1.6-1.9]' 0 0.0
4 '(1.9-2.2]' 1 1.0
5 '(2.2-2.5]' 0 0.0
6 '(2.5-2.8]' 0 0.0
7 '(2.8-3.1]' 1 1.0
8 '(3.1-3.4]' 0 0.0
9 '(3.4-3.7]' 0 0.0
Attributes
All
None
Invert
Pattern
No. Name
1 st_id
2 st_name
3 marks
4 address
Remove
Class: address (Nom)
Visualize All
Status
OK
Log
x 0

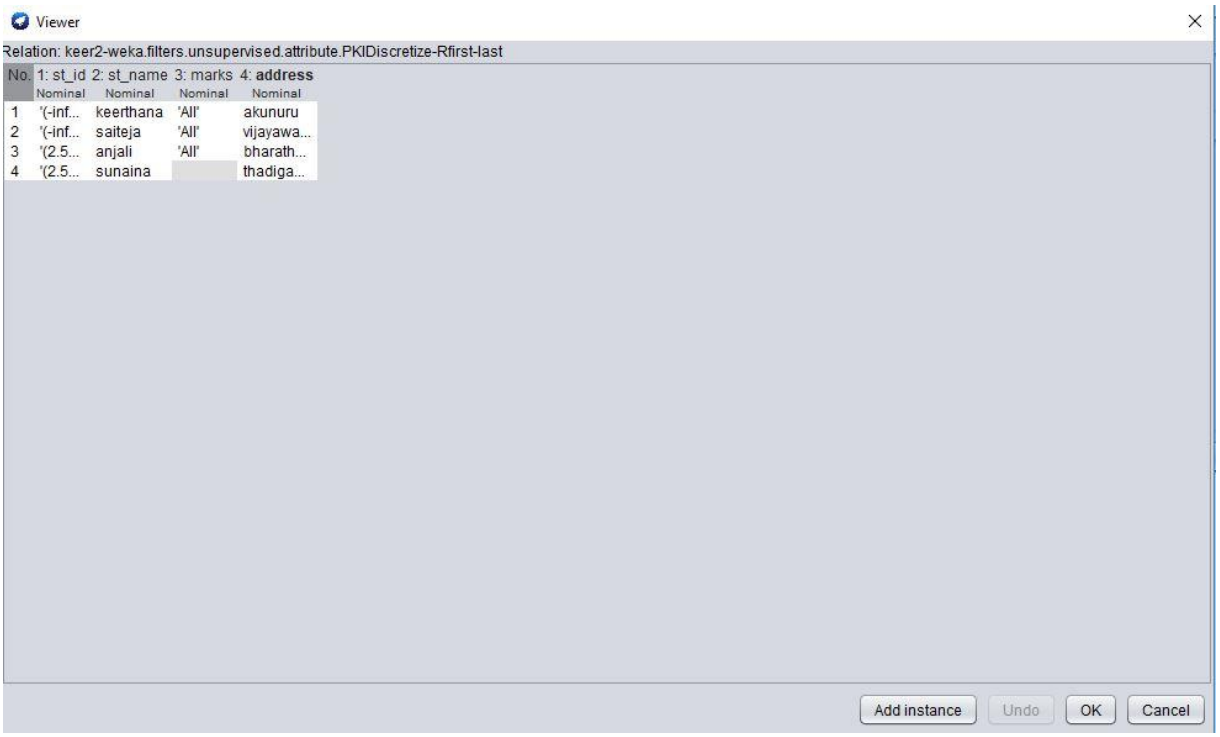

Viewer
Relation: keer2-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6
No. 1: st_id 2: st_name 3: marks 4: address
Nominal Nominal Nominal Nominal
1 '(-inf... keerthana '(-inf-9... akunuru
2 '(1.9... saiteja '(94-95]' vijayawa...
3 '(2.8... anjali '(99-inf)' bharath...
4 '(3.7... sunaina thadiga...
Add instance
Undo
OK
Cancel

PK Discretize:

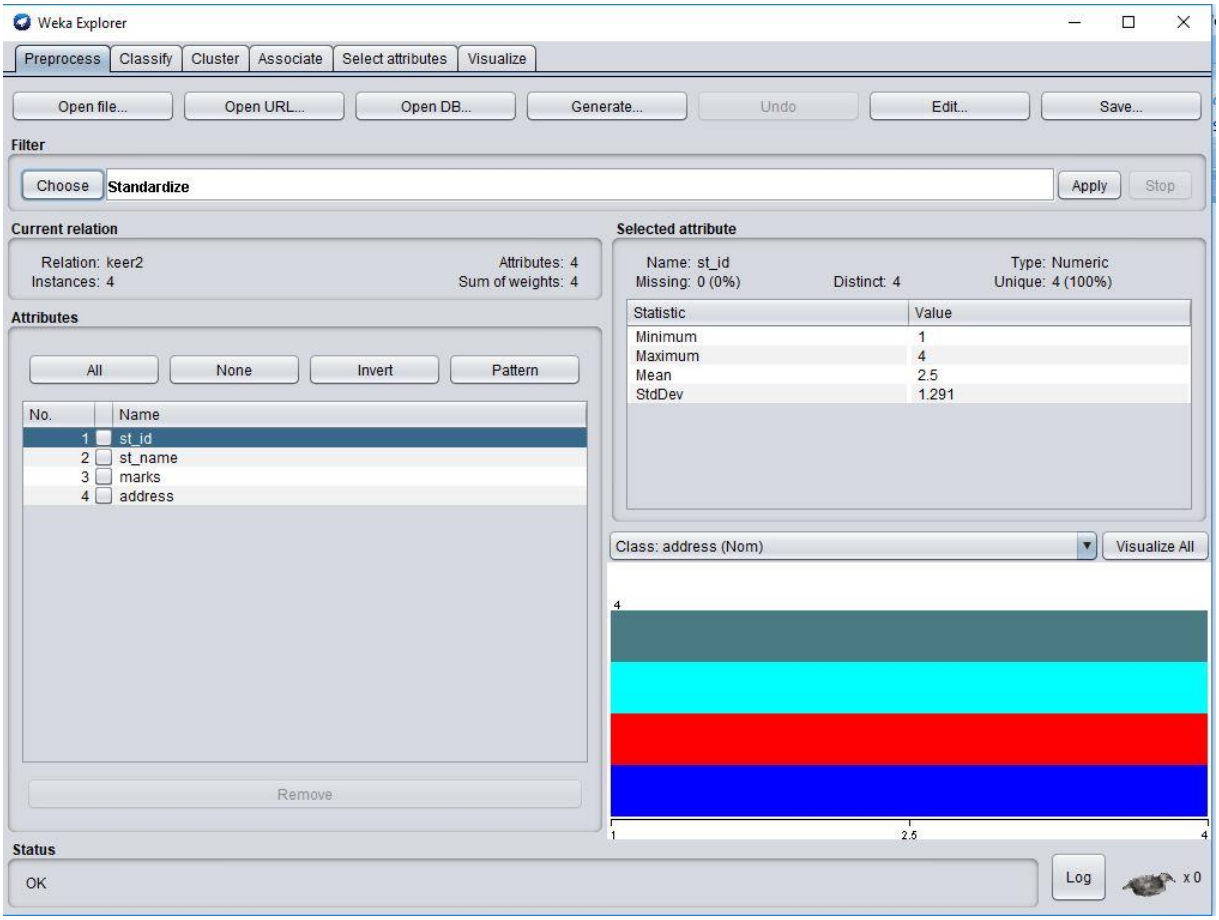


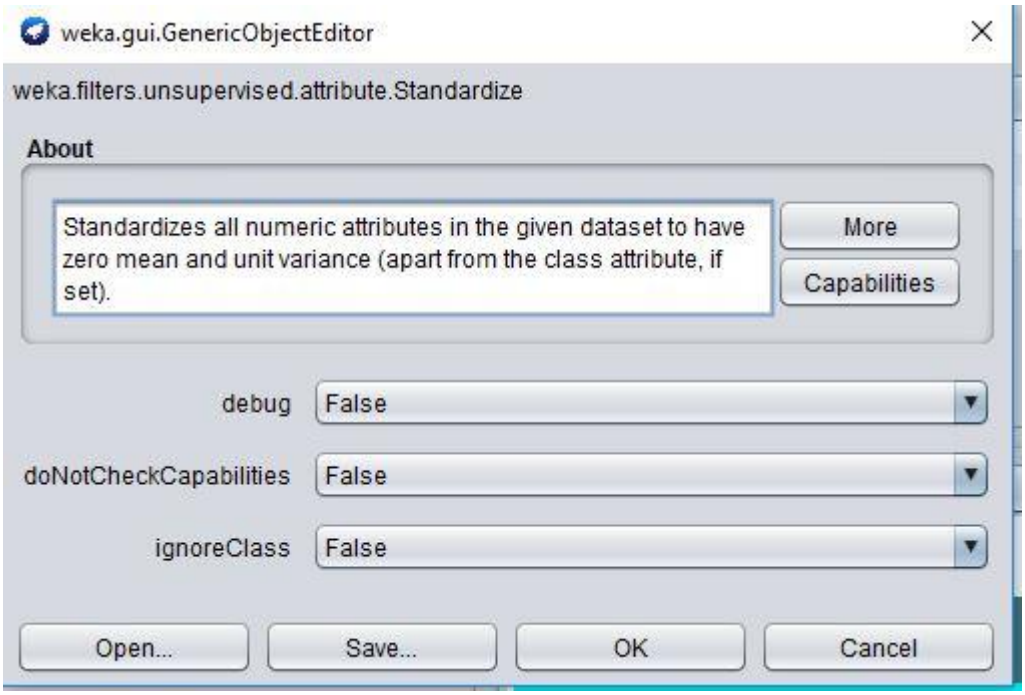


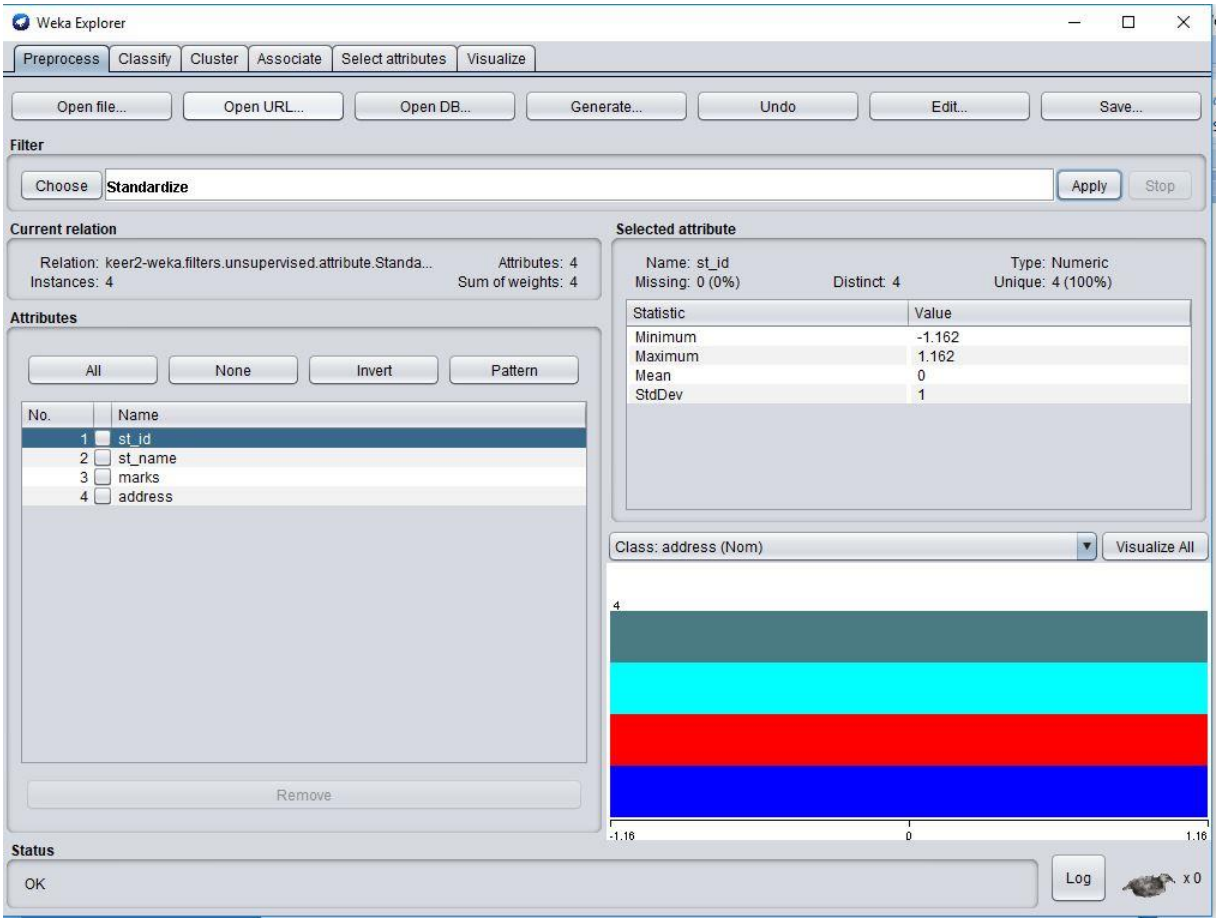

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
PKIDiscretize -R first-last
Apply
Stop
Current relation
Relation: keer2-weka.filters.unsupervised.attribute.PKIDiscretize-Rfirst-last
Instances: 4
Attributes: 4
Sum of weights: 4
Selected attribute
Name: st_id
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No. Label Count Weight
1 '(-inf-2.5]' 2 2.0
2 '(2.5-inf)' 2 2.0
Attributes
All
None
Invert
Pattern
No. Name
1 st_id
2 st_name
3 marks
4 address
Remove
Class: address (Nom)
Visualize All
Status
OK
Log
x 0

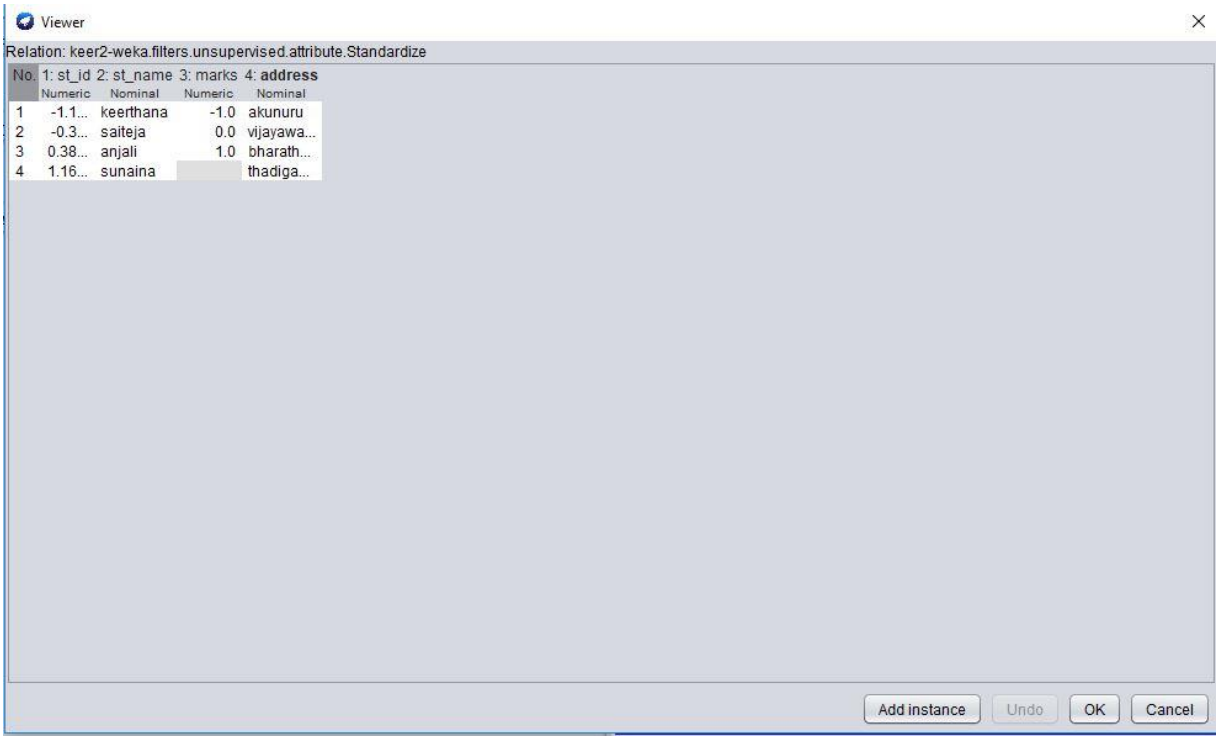

Viewer
Relation: keer2-weka.filters.unsupervised.attribute.PKIDiscretize-Rfirst-last
No. 1: st_id Nominal 2: st_name Nominal 3: marks Nominal 4: address Nominal
1 '(-inf... keerthana 'All' akunuru
2 '(-inf... saiteja 'All' vijayawa...
3 '(2.5... anjali 'All' bharath...
4 '(2.5... sunaina thadiga...
Add instance
Undo
OK
Cancel

Standardize:



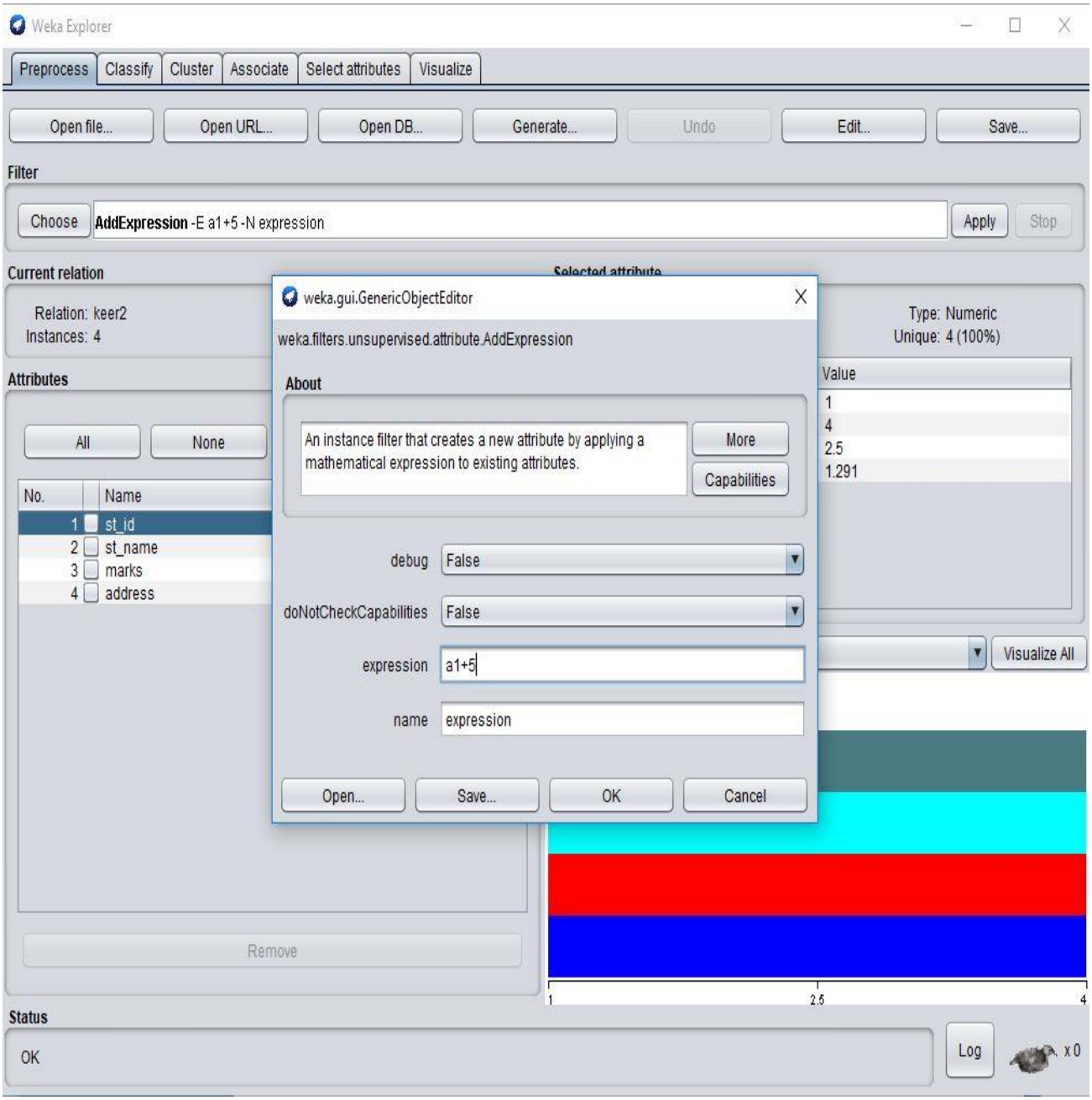


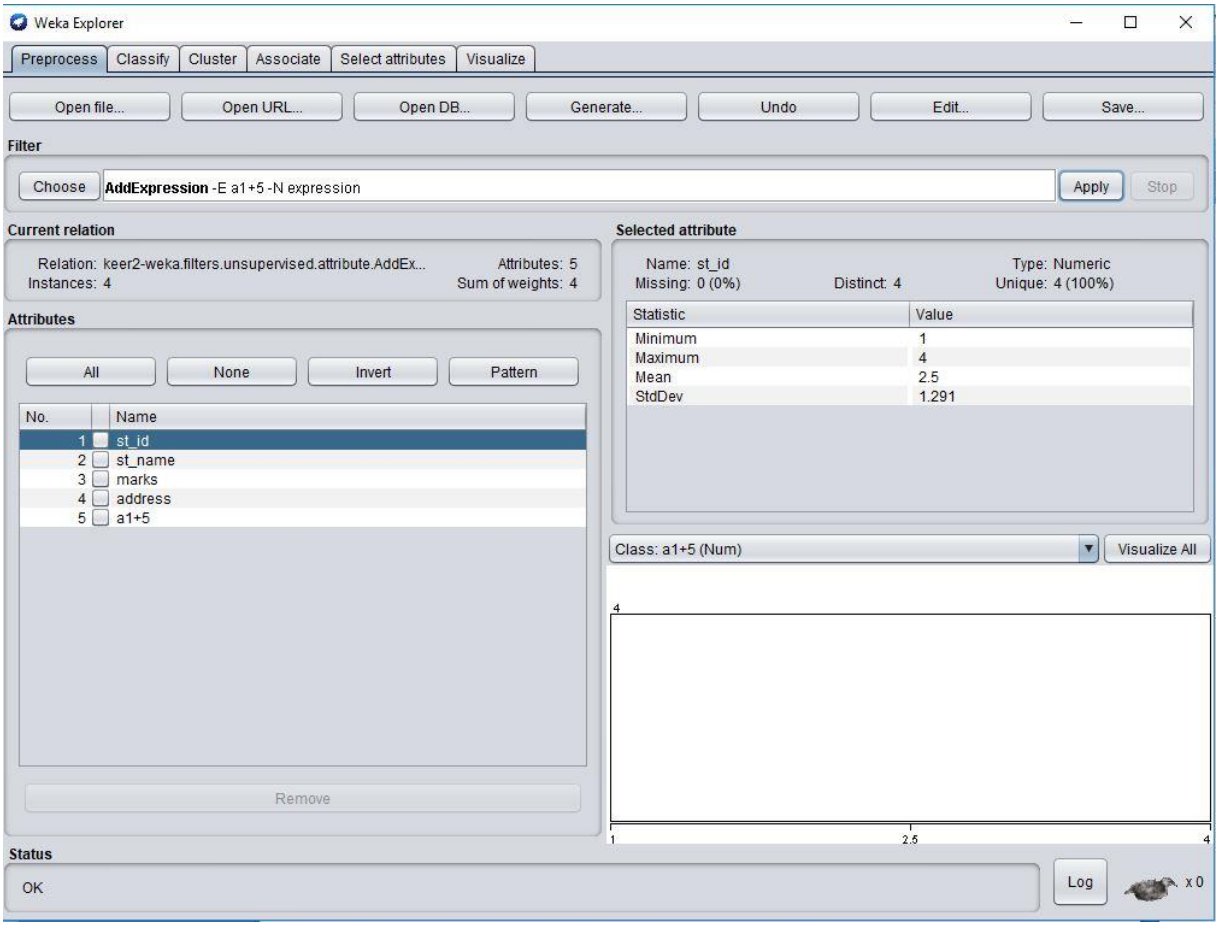

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Standardize
Apply
Stop
Current relation
Relation: keer2-weka.filters.unsupervised.attribute.Standa...
Instances: 4
Attributes: 4
Sum of weights: 4
Selected attribute
Name: st_id
Missing: 0 (0%)
Distinct: 4
Type: Numeric
Unique: 4 (100%)
Statistic
Value
Minimum
-1.162
Maximum
1.162
Mean
0
StdDev
1
Attributes
All
None
Invert
Pattern
No.
Name
1 st_id
2 st_name
3 marks
4 address
Remove
Class: address (Nom)
Visualize All
4
-1.16
0
1.16
Status
OK
Log
x 0

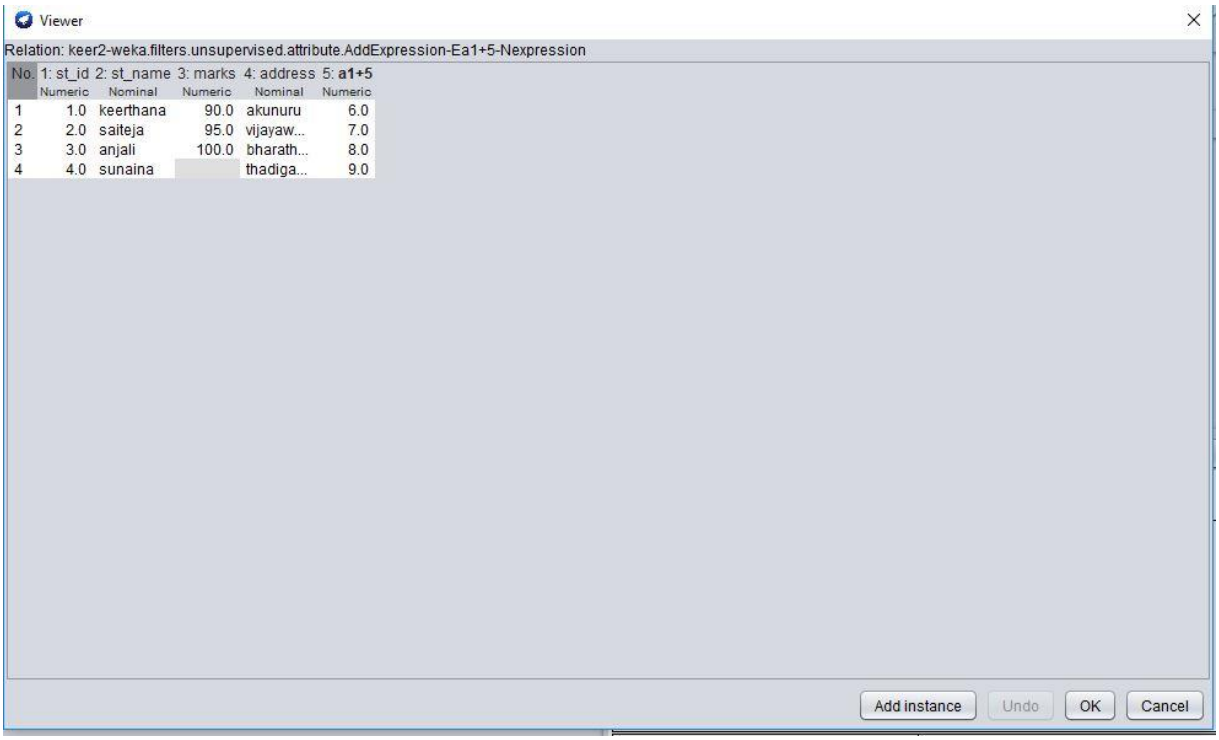

Viewer
Relation: keer2-weka.filters.unsupervised.attribute.Standardize
No.
1: st_id
Numeric
2: st_name
Nominal
3: marks
Numeric
4: address
Nominal
1 -1.1... keerthana -1.0 akunuru
2 -0.3... saiteja 0.0 vijayawa...
3 0.38... anjali 1.0 bharath...
4 1.16... sunaina thadiga...
Add instance
Undo
OK
Cancel

Add Expression:

To add any mathematical expression to evaluate some data in an attribute.







Result: Applying pre processing techniques on the data is done successfully.

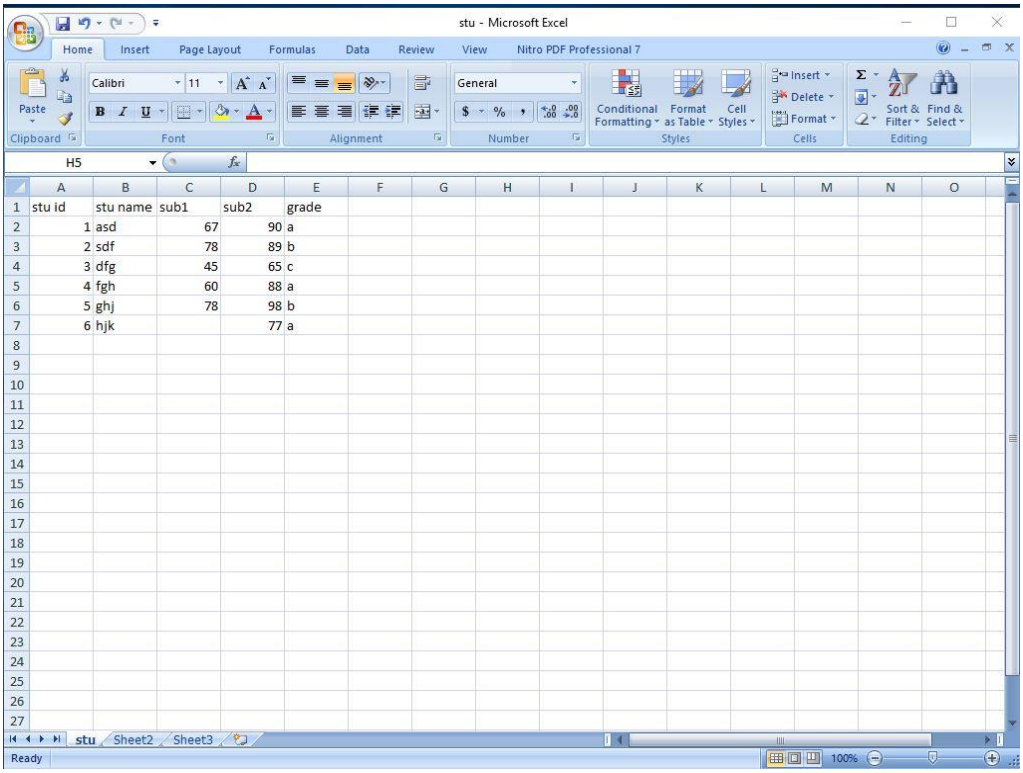
Task-8

Aim: To apply pre processing techniques on the data in Rapid miner

Description:

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

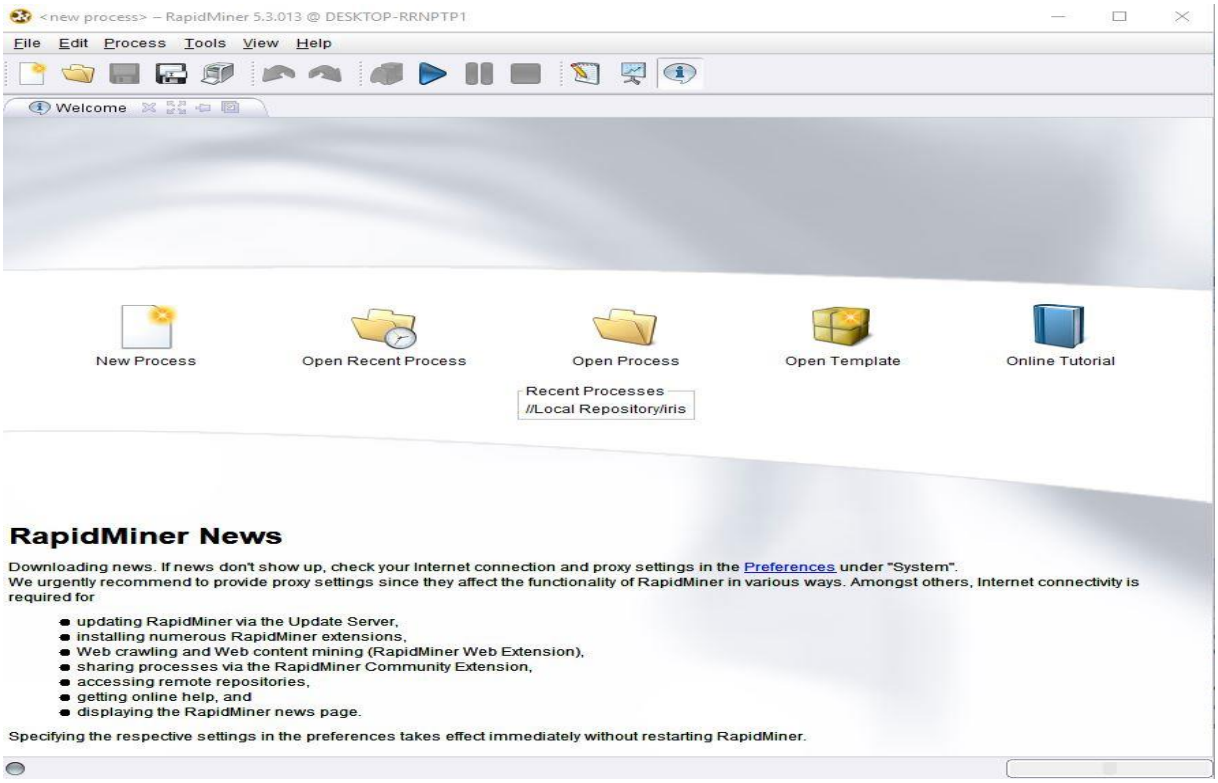
Input file:

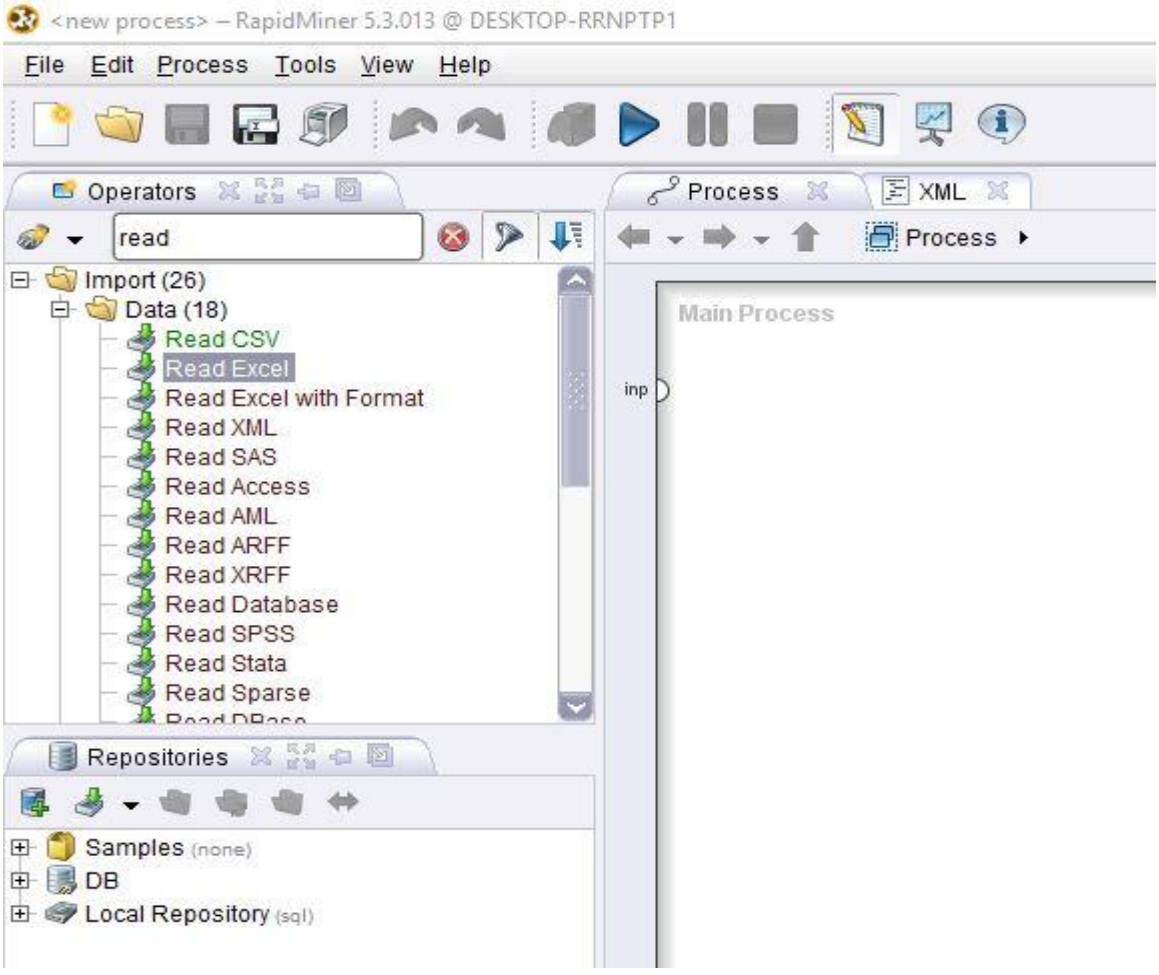


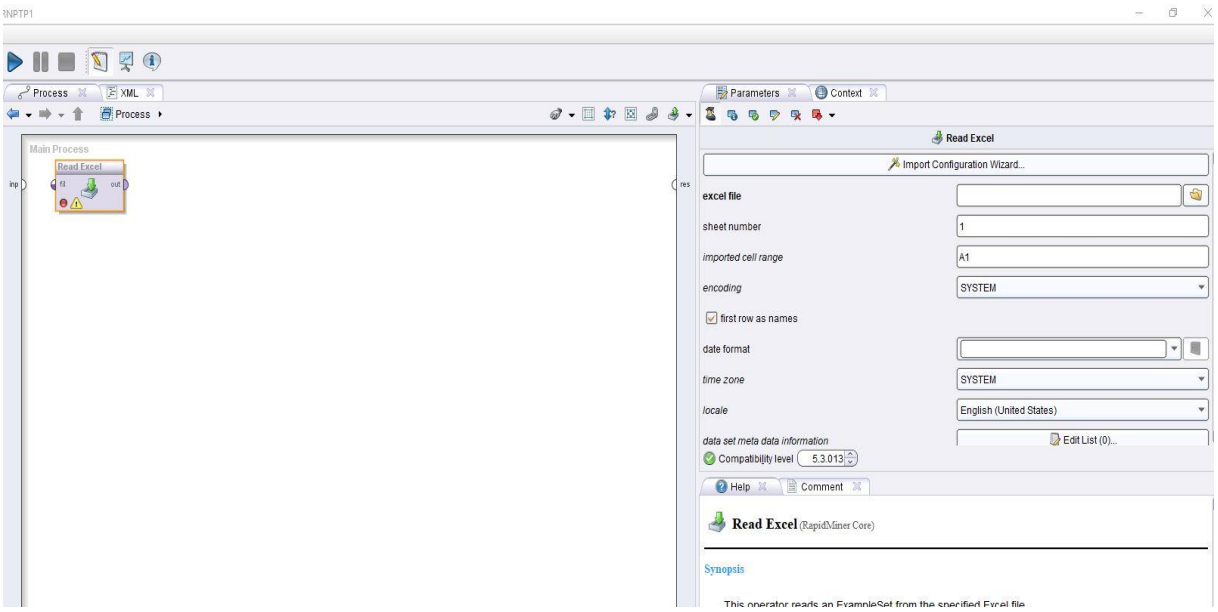
	A	B	C	D	E
1	stu id	stu name	sub1	sub2	grade
2	1	asd	67	90	a
3	2	sdf	78	89	b
4	3	dfg	45	65	c
5	4	fgh	60	88	a
6	5	ghj	78	98	b
7	6	hjk		77	a

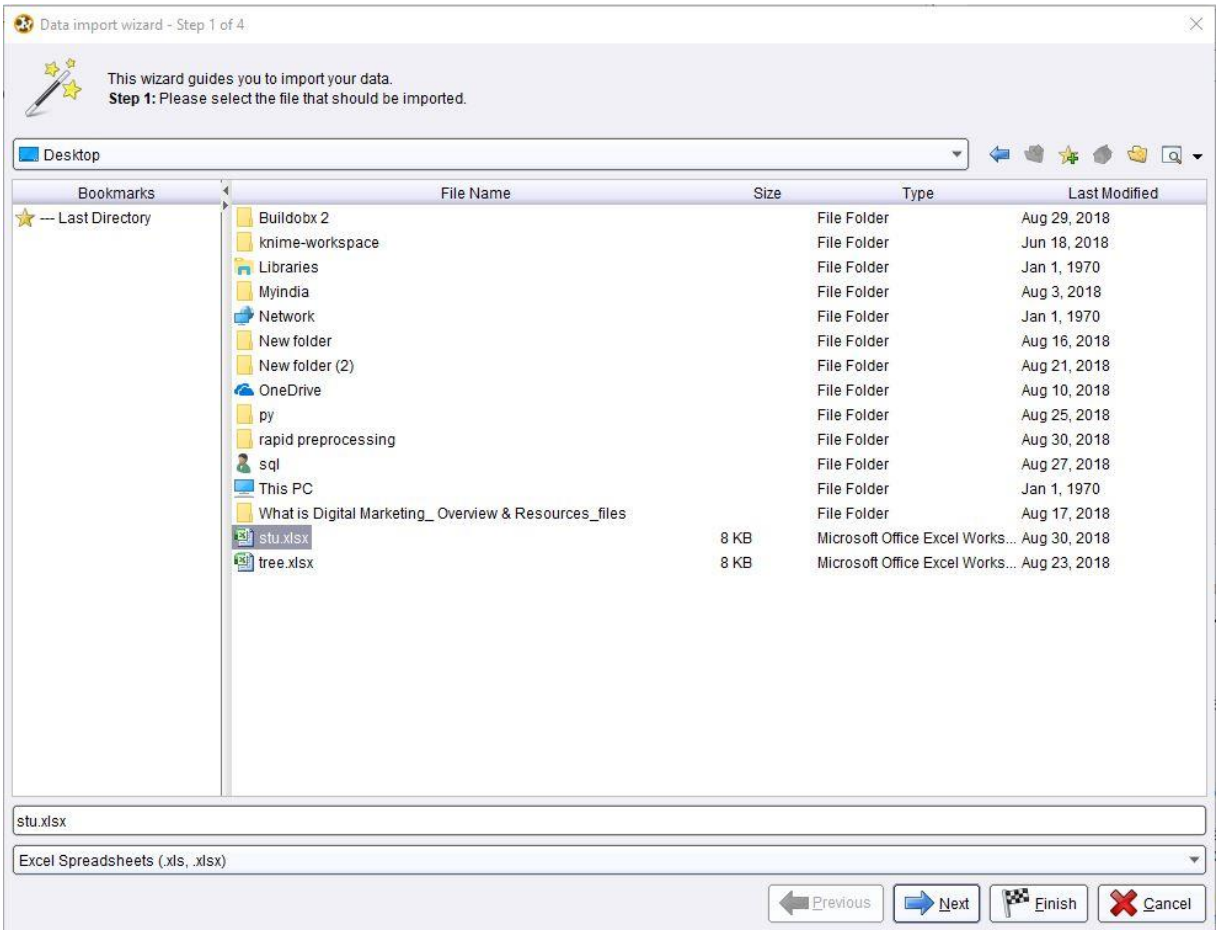
Procedure:

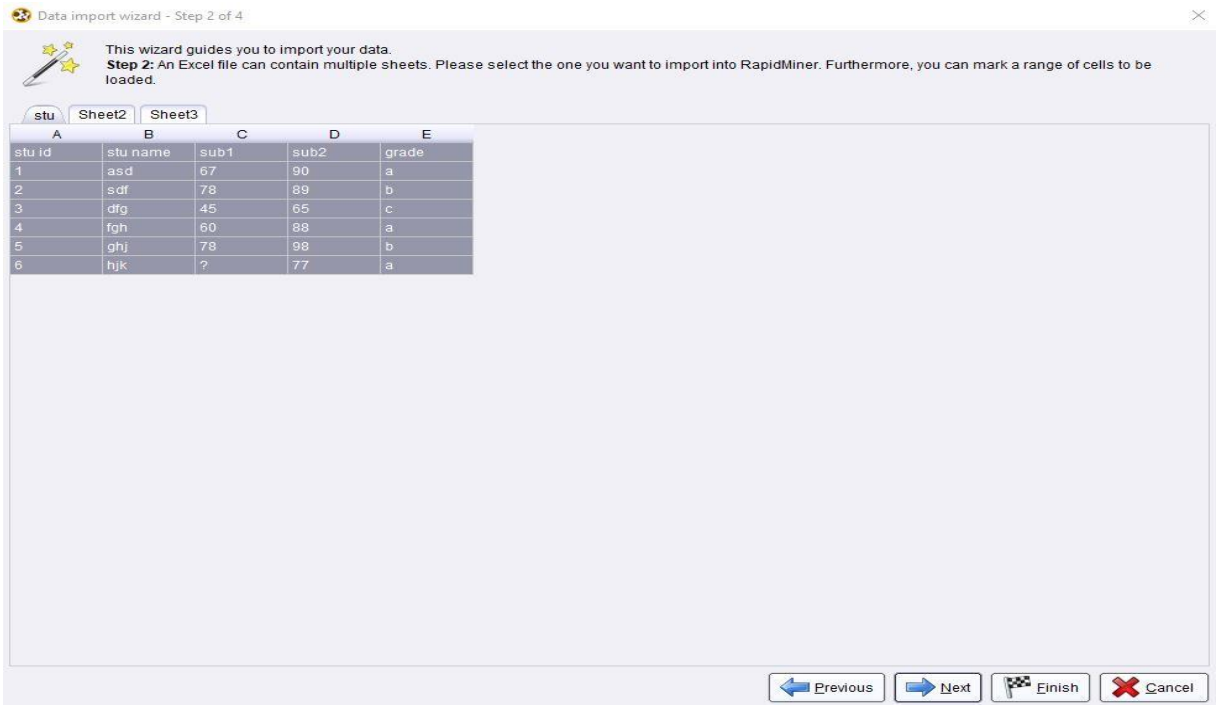
Replace Missing Values: First, create a new process and then drag and drop “read excel” and now import excel file by clicking on “import configuration wizard” and later drag and drop the “replace missing values” and then by clicking on “edit” select the attribute where the missing value is present and select the option with which the missing value should be replaced. At last join the input and output connections.

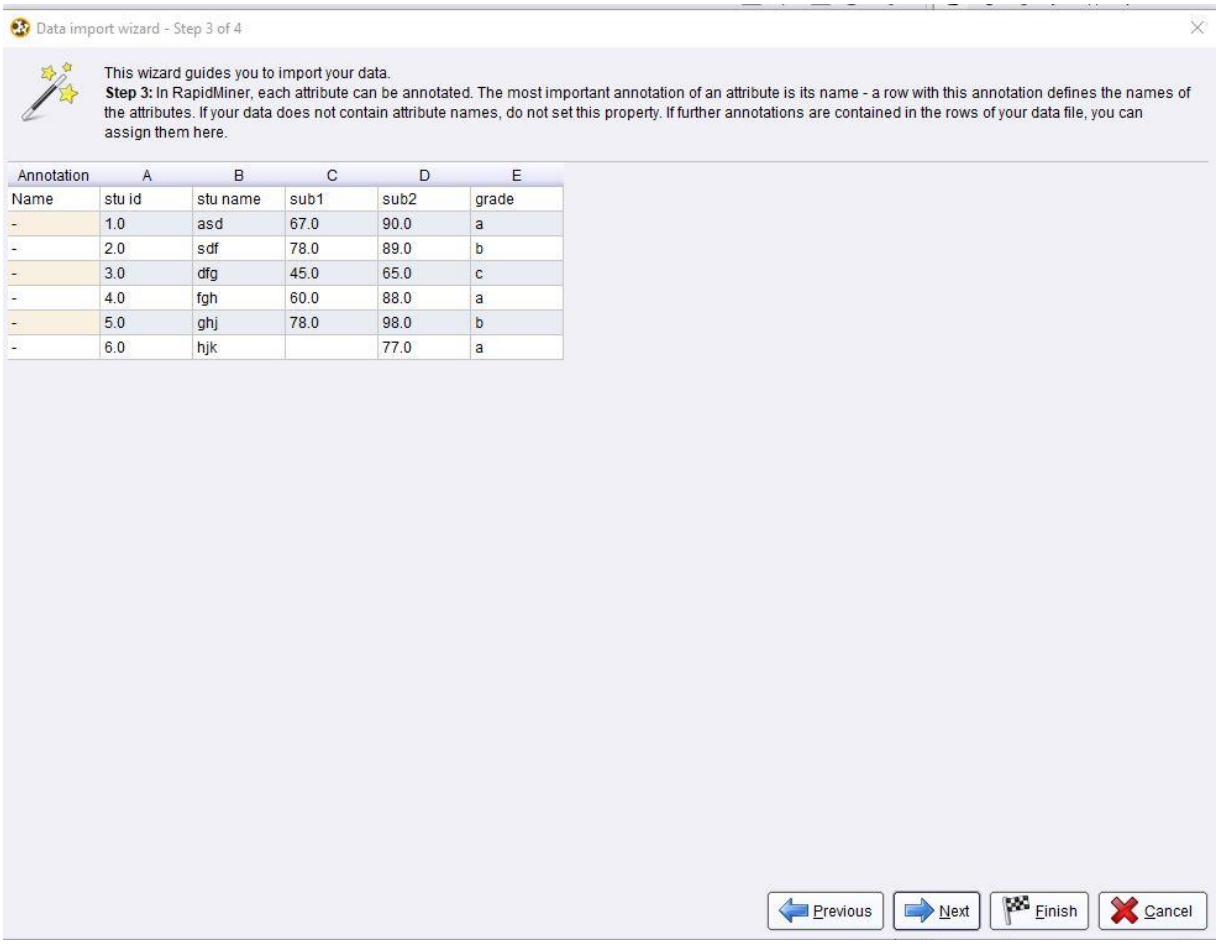

<new process> – RapidMiner 5.3.013 @ DESKTOP-RRNPTP1
File Edit Process Tools View Help
Welcome
New Process
Open Recent Process
Open Process
Open Template
Online Tutorial
Recent Processes
//Local Repository/iris
RapidMiner News
Downloading news. If news don't show up, check your Internet connection and proxy settings in the Preferences under "System".
We urgently recommend to provide proxy settings since they affect the functionality of RapidMiner in various ways. Amongst others, Internet connectivity is required for
updating RapidMiner via the Update Server,
installing numerous RapidMiner extensions,
Web crawling and Web content mining (RapidMiner Web Extension),
sharing processes via the RapidMiner Community Extension,
accessing remote repositories,
getting online help, and
displaying the RapidMiner news page.
Specifying the respective settings in the preferences takes effect immediately without restarting RapidMiner.

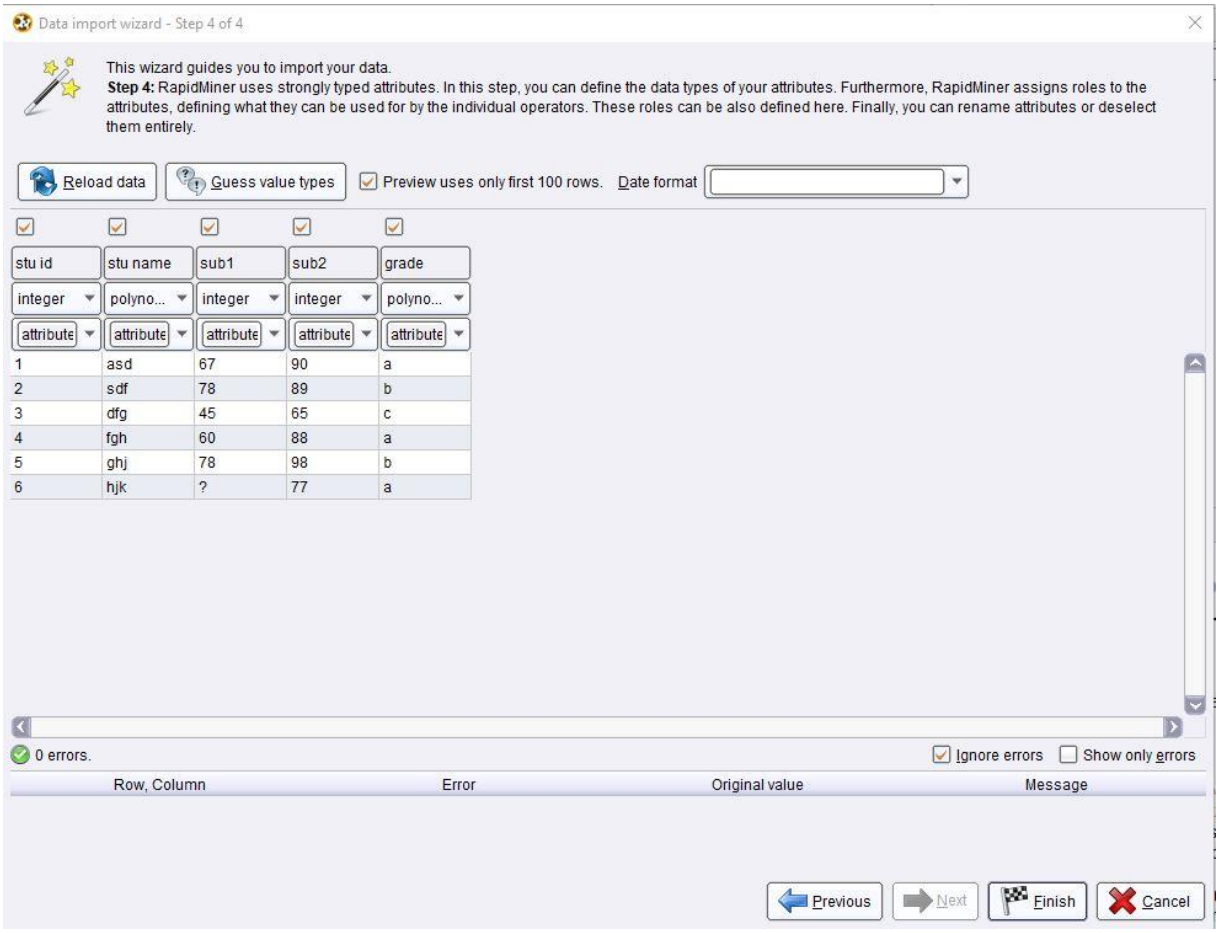

<new process> – RapidMiner 5.3.013 @ DESKTOP-RRNPTP1
File Edit Process Tools View Help
Operators
read
Import (26)
Data (18)
Read CSV
Read Excel
Read Excel with Format
Read XML
Read SAS
Read Access
Read AML
Read ARFF
Read XRFF
Read Database
Read SPSS
Read Stata
Read Sparse
Repositories
Samples (none)
DB
Local Repository (sql)
Process
XML
Process
Main Process
inp

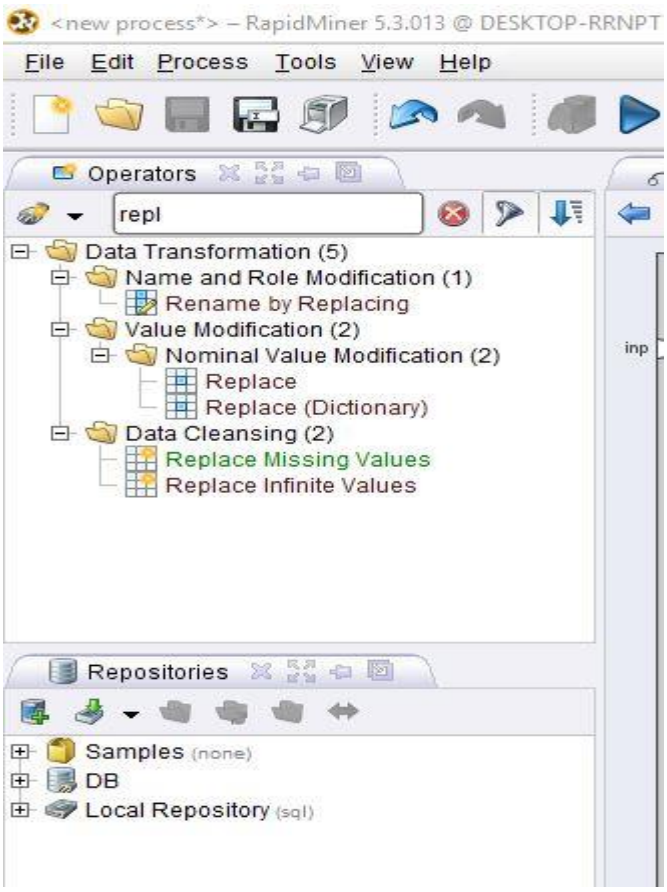

Process
XML
Process
Main Process
Read Excel
Parameters
Context
Read Excel
Import Configuration Wizard...
excel file
sheet number
1
imported cell range
A1
encoding
SYSTEM
first row as names
date format
time zone
SYSTEM
locale
English (United States)
data set meta data information
Edit List (0)...
Compatibility level
5.3.013
Help
Comment
Read Excel (RapidMiner Core)
Synopsis

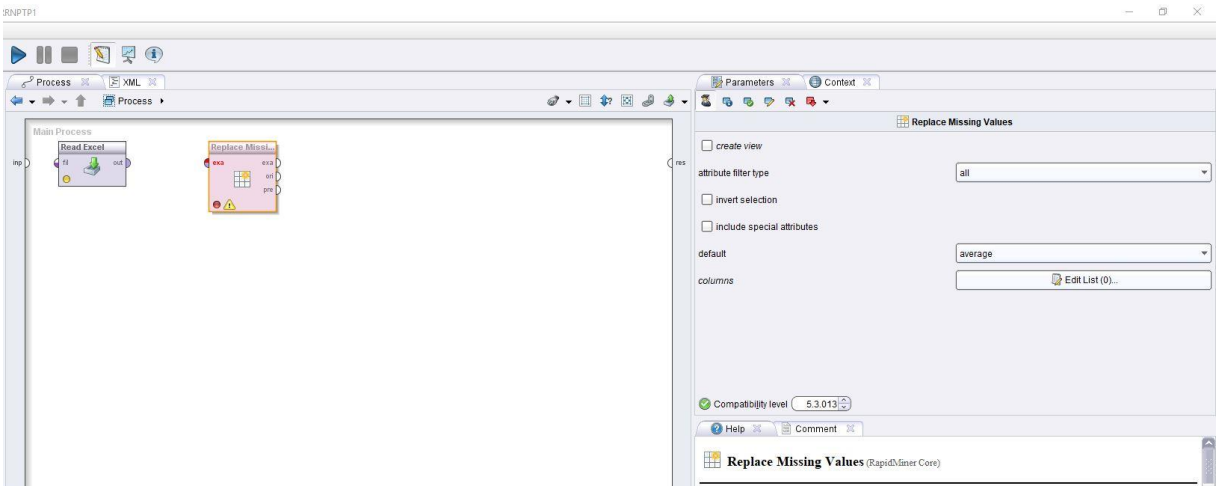

Data import wizard - Step 1 of 4
This wizard guides you to import your data.
Step 1: Please select the file that should be imported.
Desktop
Bookmarks
--- Last Directory
File Name
Size
Type
Last Modified
Buildobx 2 File Folder Aug 29, 2018
knime-workspace File Folder Jun 18, 2018
Libraries File Folder Jan 1, 1970
Myindia File Folder Aug 3, 2018
Network File Folder Jan 1, 1970
New folder File Folder Aug 16, 2018
New folder (2) File Folder Aug 21, 2018
OneDrive File Folder Aug 10, 2018
py File Folder Aug 25, 2018
rapid preprocessing File Folder Aug 30, 2018
sql File Folder Aug 27, 2018
This PC File Folder Jan 1, 1970
What is Digital Marketing_ Overview & Resources_files File Folder Aug 17, 2018
stu.xlsx 8 KB Microsoft Office Excel Works... Aug 30, 2018
tree.xlsx 8 KB Microsoft Office Excel Works... Aug 23, 2018
stu.xlsx
Excel Spreadsheets (.xls, .xlsx)
Previous
Next
Finish
Cancel

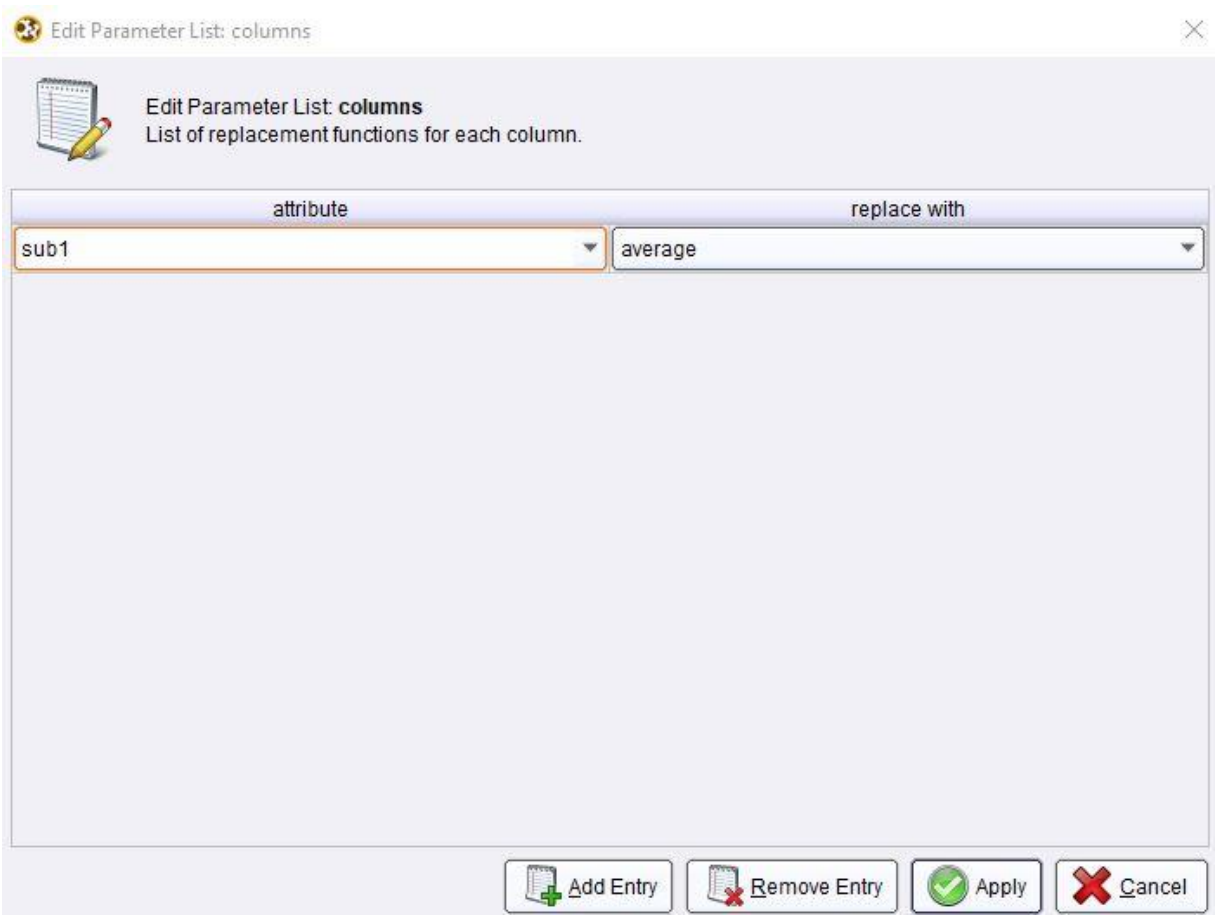

Data import wizard - Step 2 of 4
This wizard guides you to import your data.
Step 2: An Excel file can contain multiple sheets. Please select the one you want to import into RapidMiner. Furthermore, you can mark a range of cells to be loaded.
stu
Sheet2
Sheet3
A B C D E
stu id | stu name | sub1 | sub2 | grade
1 | asd | 67 | 90 | a
2 | sdf | 78 | 89 | b
3 | dfg | 45 | 65 | c
4 | fgh | 60 | 88 | a
5 | ghj | 78 | 98 | b
6 | hjk | ? | 77 | a
Previous
Next
Finish
Cancel

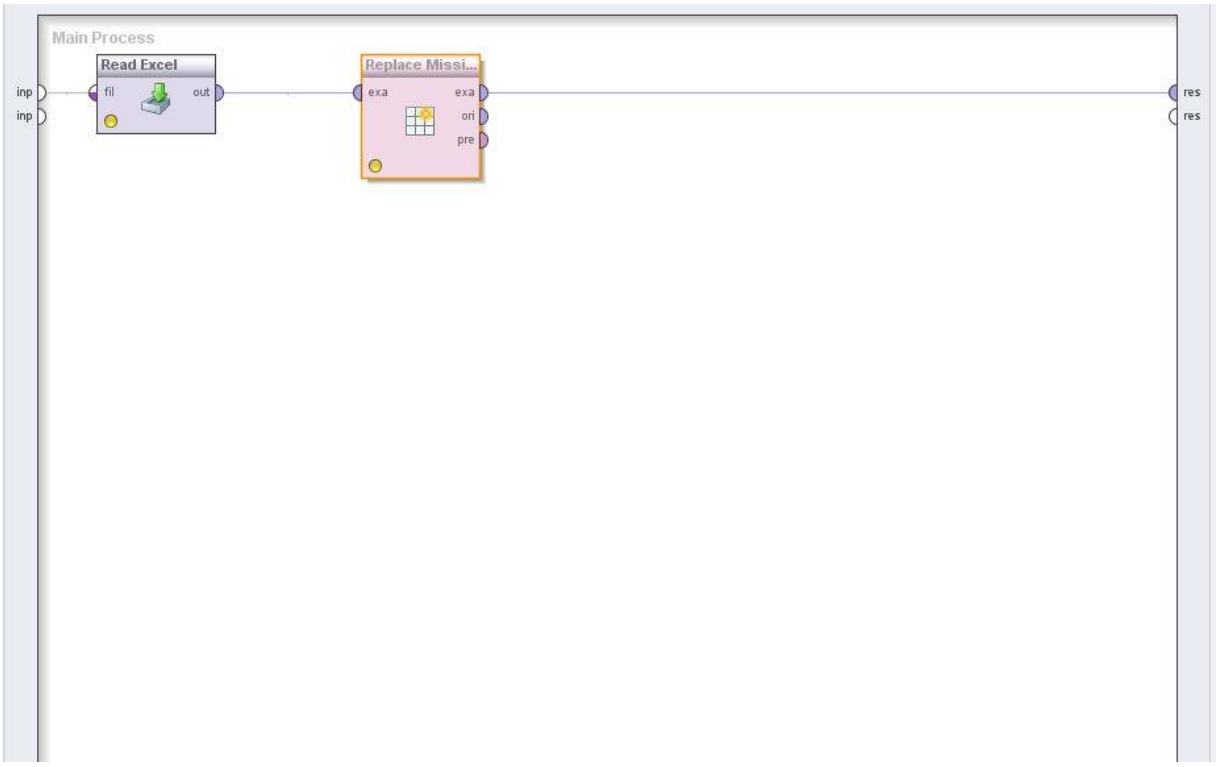

Data import wizard - Step 3 of 4
This wizard guides you to import your data.
Step 3: In RapidMiner, each attribute can be annotated. The most important annotation of an attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.
Annotation | A | B | C | D | E
Name | stu id | stu name | sub1 | sub2 | grade
- | 1.0 | asd | 67.0 | 90.0 | a
- | 2.0 | sdf | 78.0 | 89.0 | b
- | 3.0 | dfg | 45.0 | 65.0 | c
- | 4.0 | fgh | 60.0 | 88.0 | a
- | 5.0 | ghj | 78.0 | 98.0 | b
- | 6.0 | hjk | | 77.0 | a
Previous
Next
Finish
Cancel

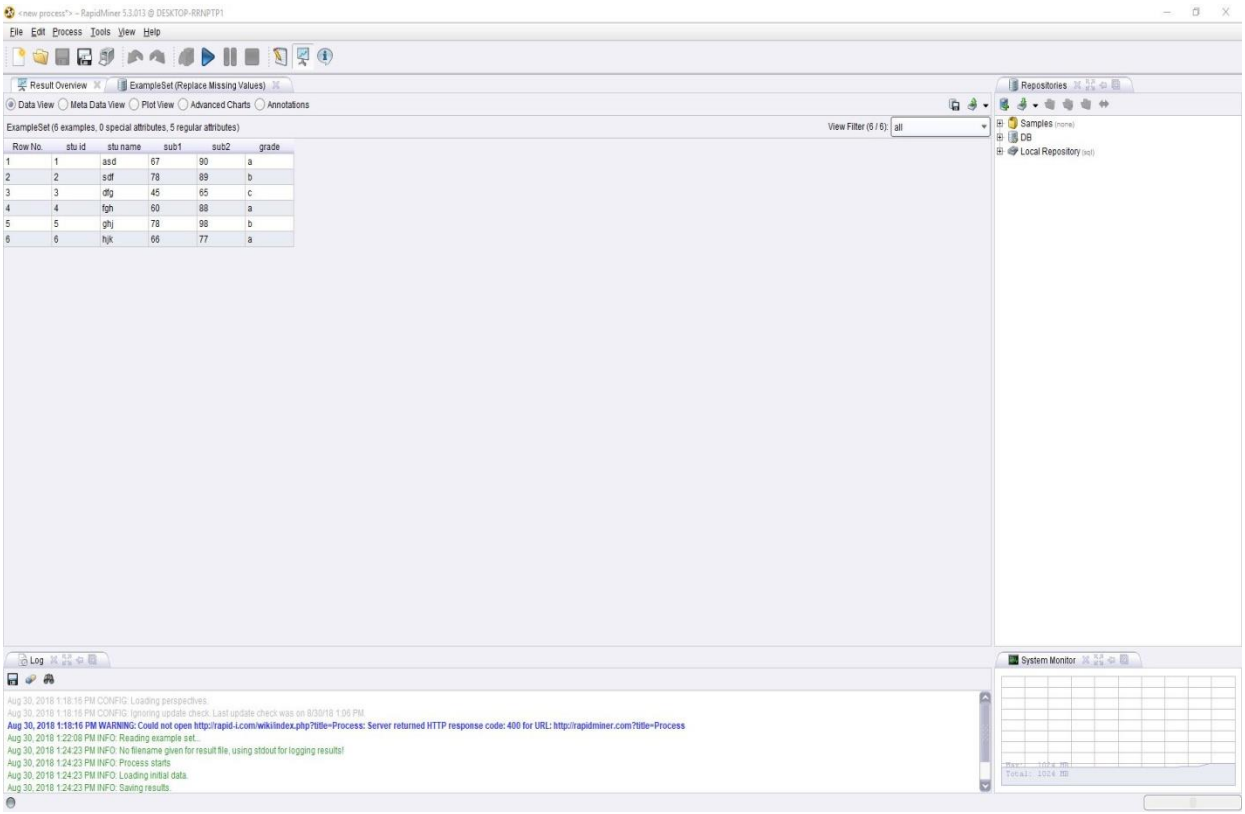

Data import wizard - Step 4 of 4
This wizard guides you to import your data.
Step 4: RapidMiner uses strongly typed attributes. In this step, you can define the data types of your attributes. Furthermore, RapidMiner assigns roles to the attributes, defining what they can be used for by the individual operators. These roles can be also defined here. Finally, you can rename attributes or deselect them entirely.
Reload data
Guess value types
Preview uses only first 100 rows.
Date format
stu id | stu name | sub1 | sub2 | grade
integer | polyno... | integer | integer | polyno...
attribute
1 asd 67 90 a
2 sdf 78 89 b
3 dfg 45 65 c
4 fgh 60 88 a
5 ghj 78 98 b
6 hjk ? 77 a
0 errors.
Ignore errors
Show only errors
Row, Column
Error
Original value
Message
Previous
Next
Finish
Cancel


<new process*> – RapidMiner 5.3.013 @ DESKTOP-RRNPT
File Edit Process Tools View Help
Operators
repl
Data Transformation (5)
Name and Role Modification (1)
Rename by Replacing
Value Modification (2)
Nominal Value Modification (2)
Replace
Replace (Dictionary)
Data Cleansing (2)
Replace Missing Values
Replace Infinite Values
inp
Repositories
Samples (none)
DB
Local Repository (sql)


Process
XML
Process
Main Process
Read Excel
Replace Missi...
Parameters
Context
Replace Missing Values
create view
attribute filter type
all
invert selection
include special attributes
default
average
columns
Edit List (0)...
Compatibility level
5.3.013
Help
Comment
Replace Missing Values (RapidMiner Core)

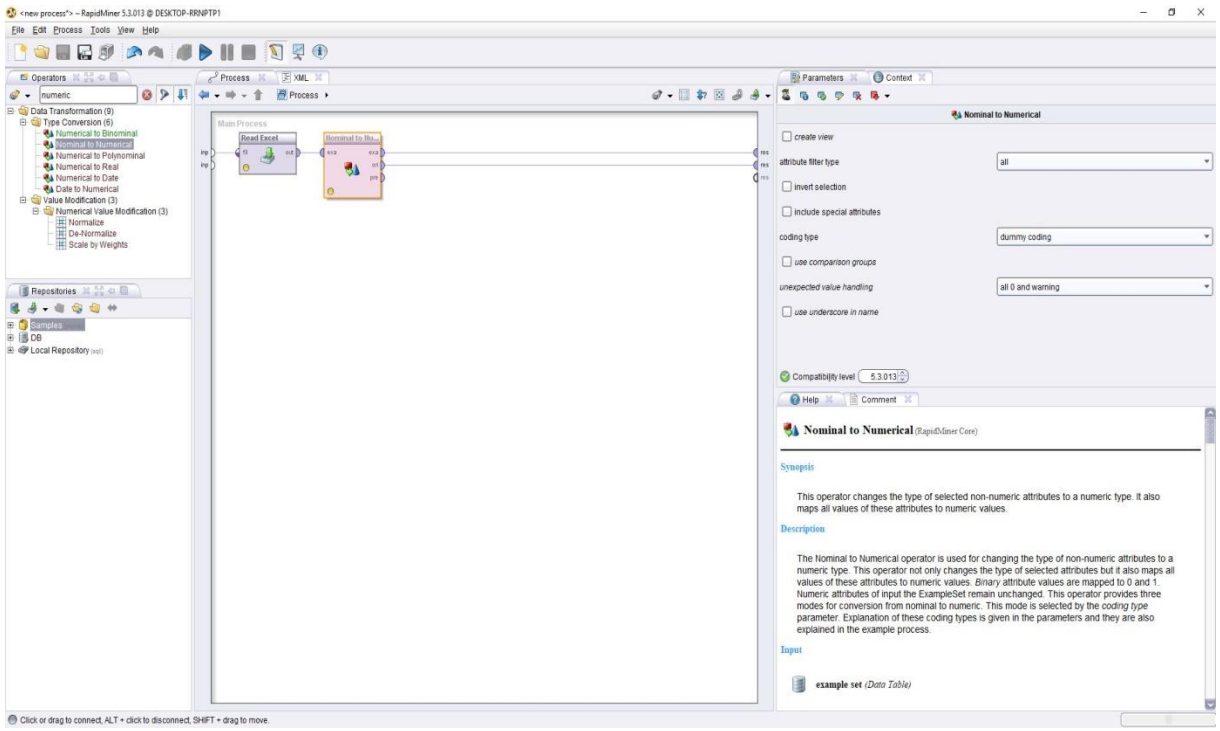

Edit Parameter List: columns
Edit Parameter List: columns
List of replacement functions for each column.
attribute
replace with
sub1
average
Add Entry
Remove Entry
Apply
Cancel

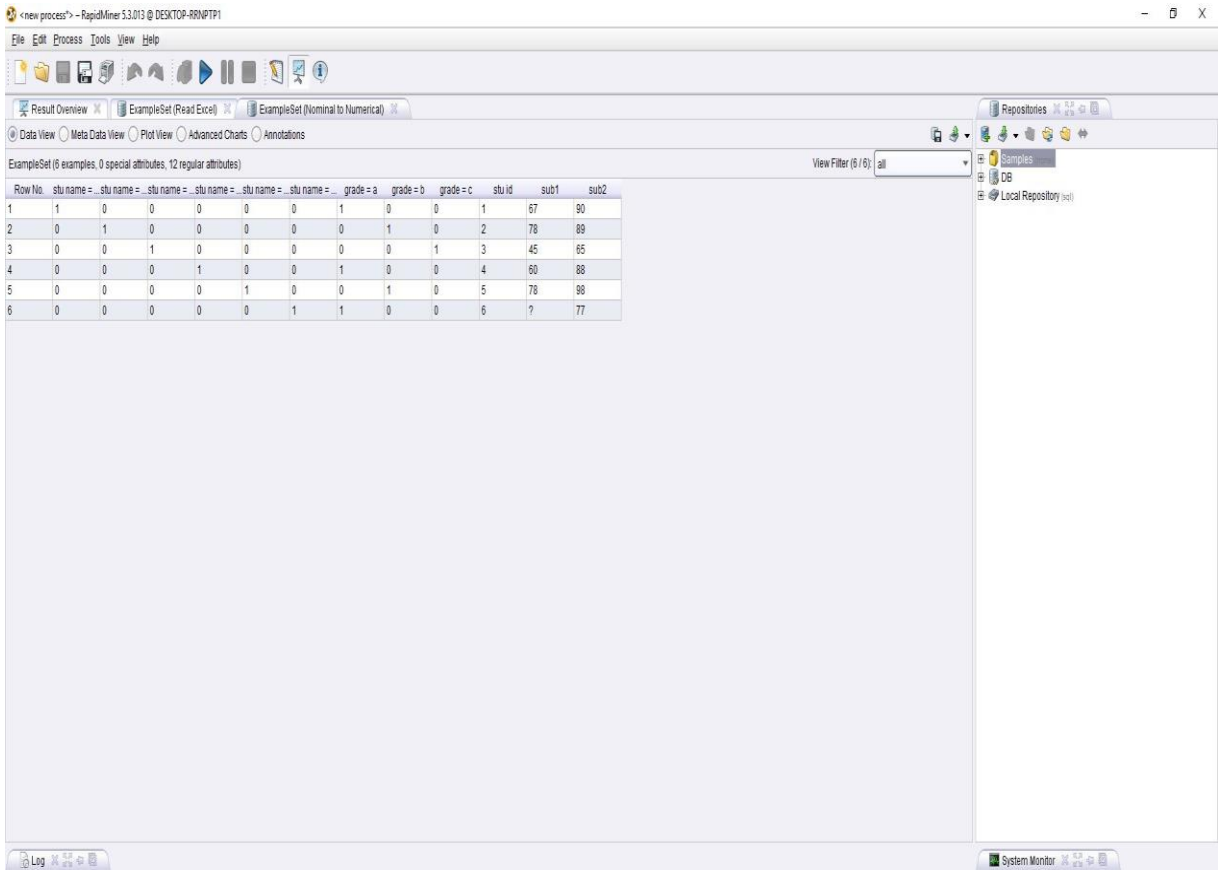




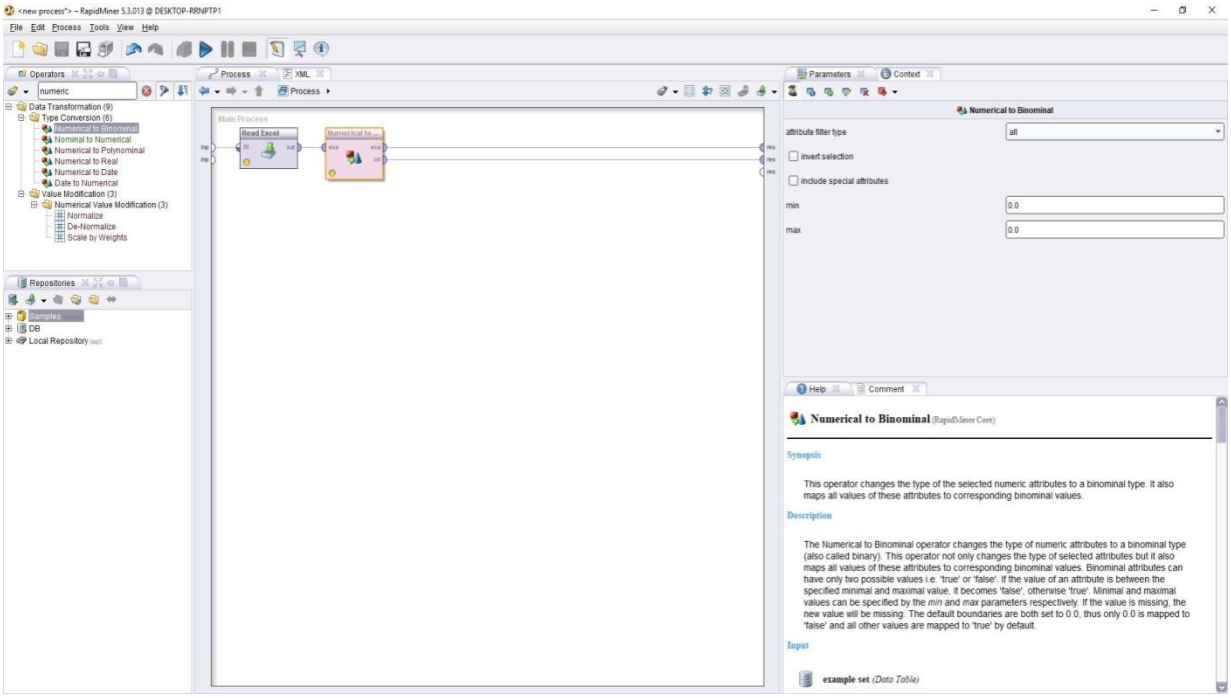
Row No.	stu id	stu name	sub1	sub2	grade
1	1	asd	67	90	a
2	2	sdf	78	89	b
3	3	dfg	45	65	c
4	4	fgh	60	88	a
5	5	ghj	78	98	b
6	6	hjk	66	77	a

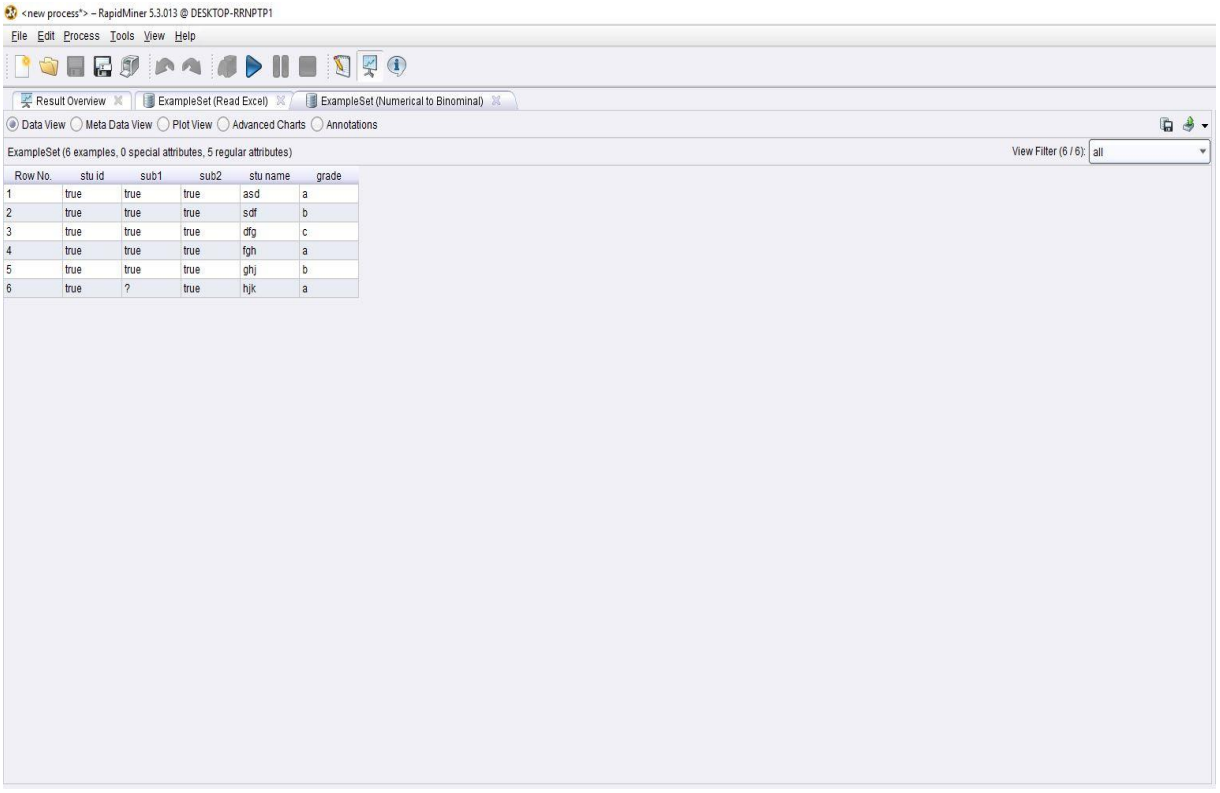
Nominal to Numeric:





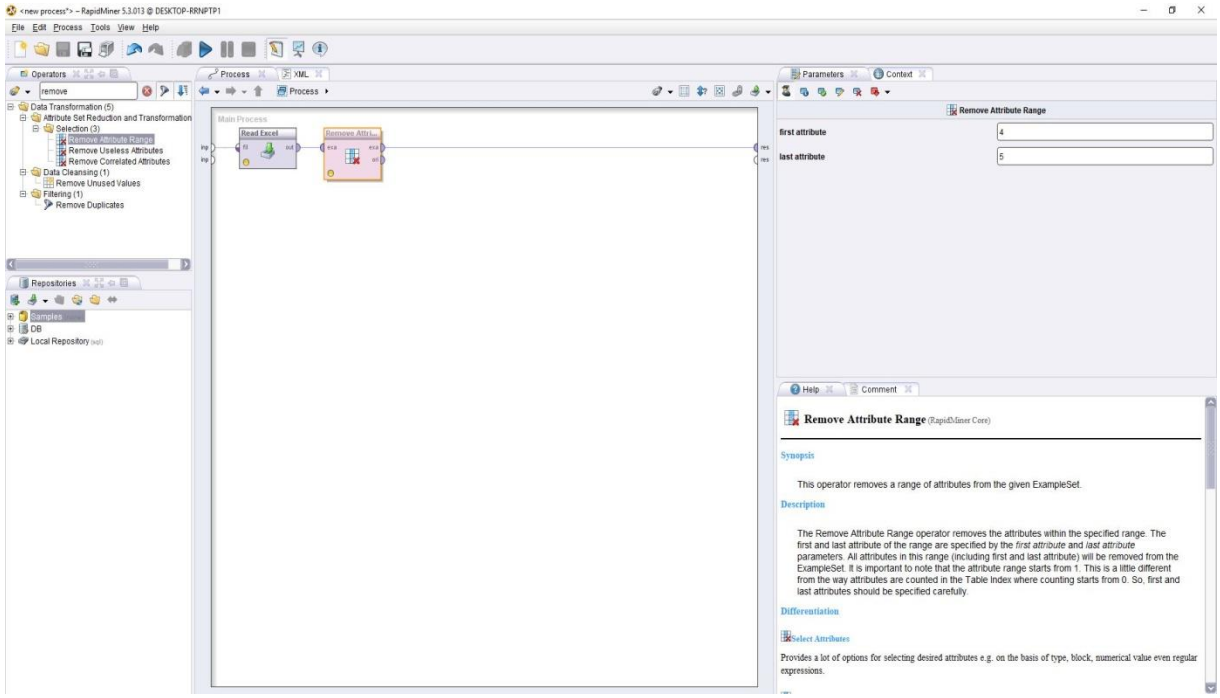
Numeric to Binomial: Converts the numerical attributes to binomial values.

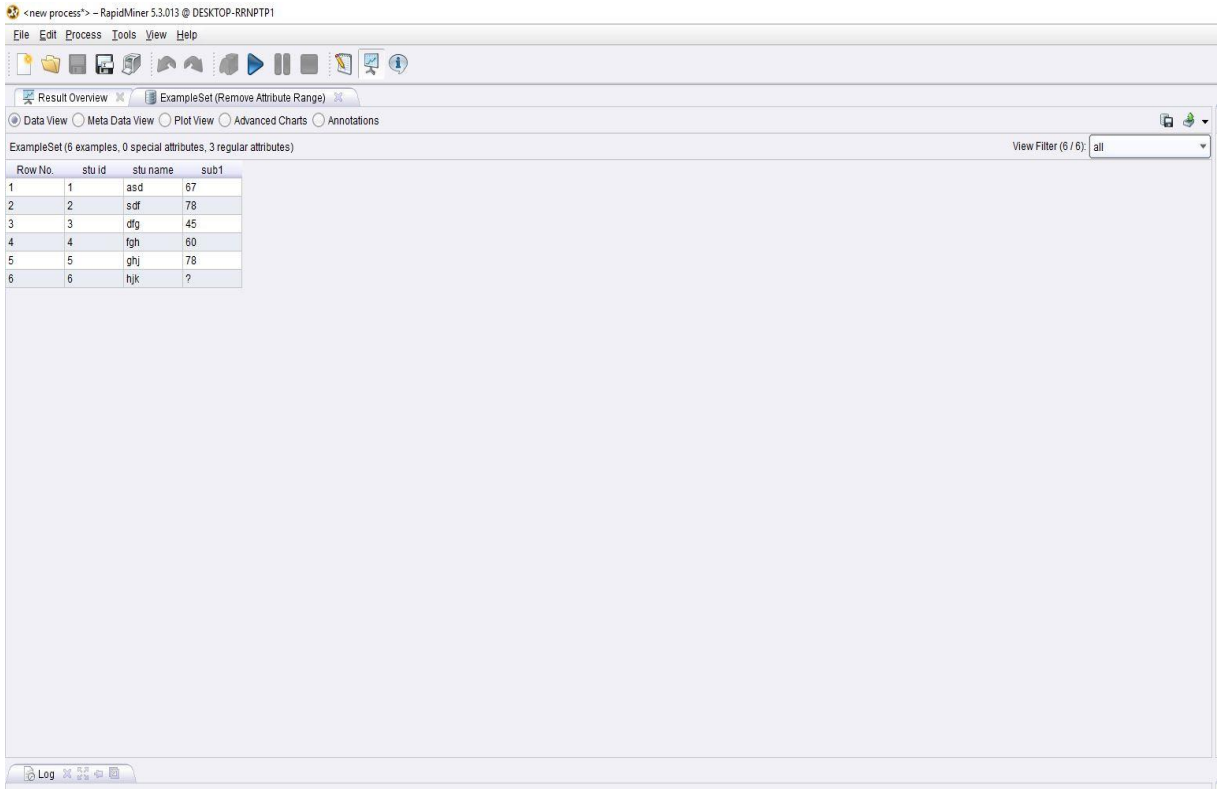




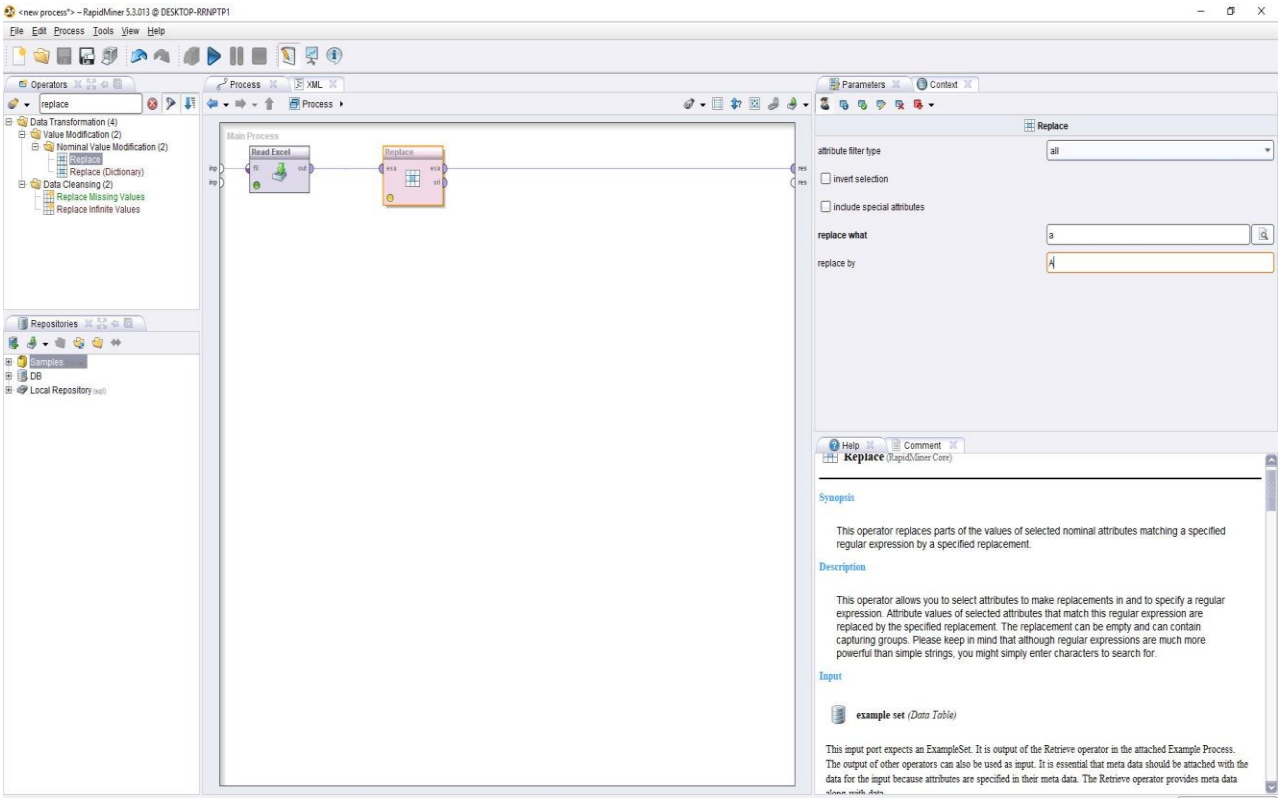
Row No.	stu id	sub1	sub2	stu name	grade
1	true	true	true	asd	a
2	true	true	true	sdf	b
3	true	true	true	dfg	c
4	true	true	true	fgh	a
5	true	true	true	ghj	b
6	true	?	true	hjk	a

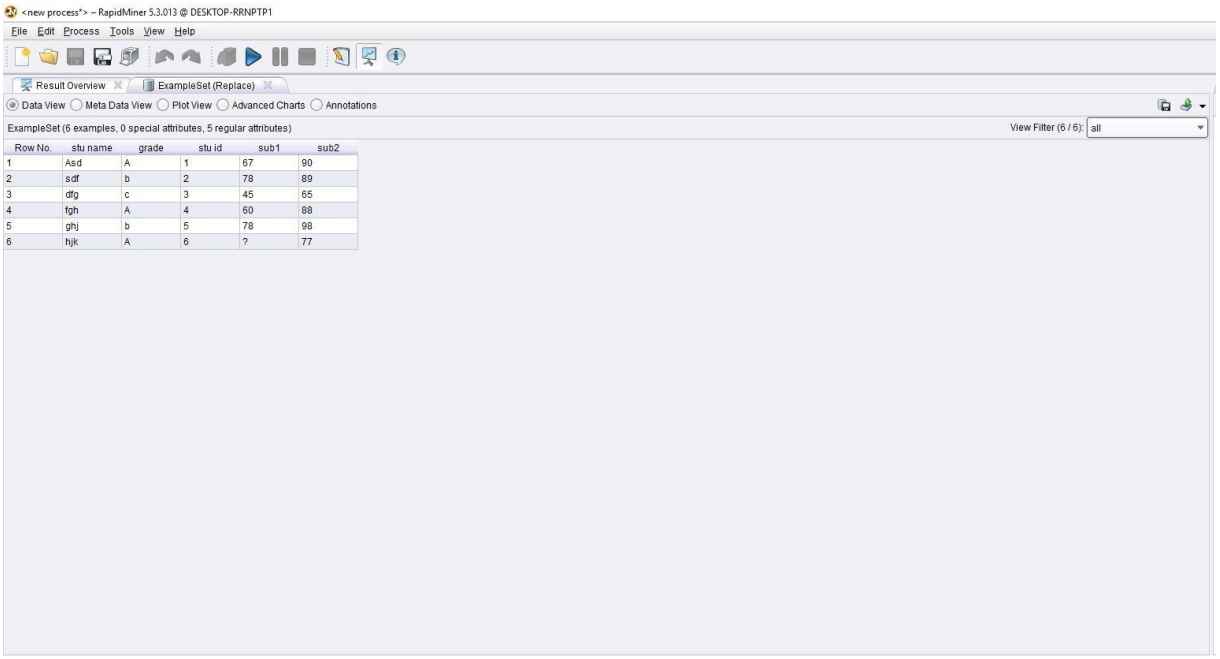
Remove Attributes: Removes the selected attributes from the data.





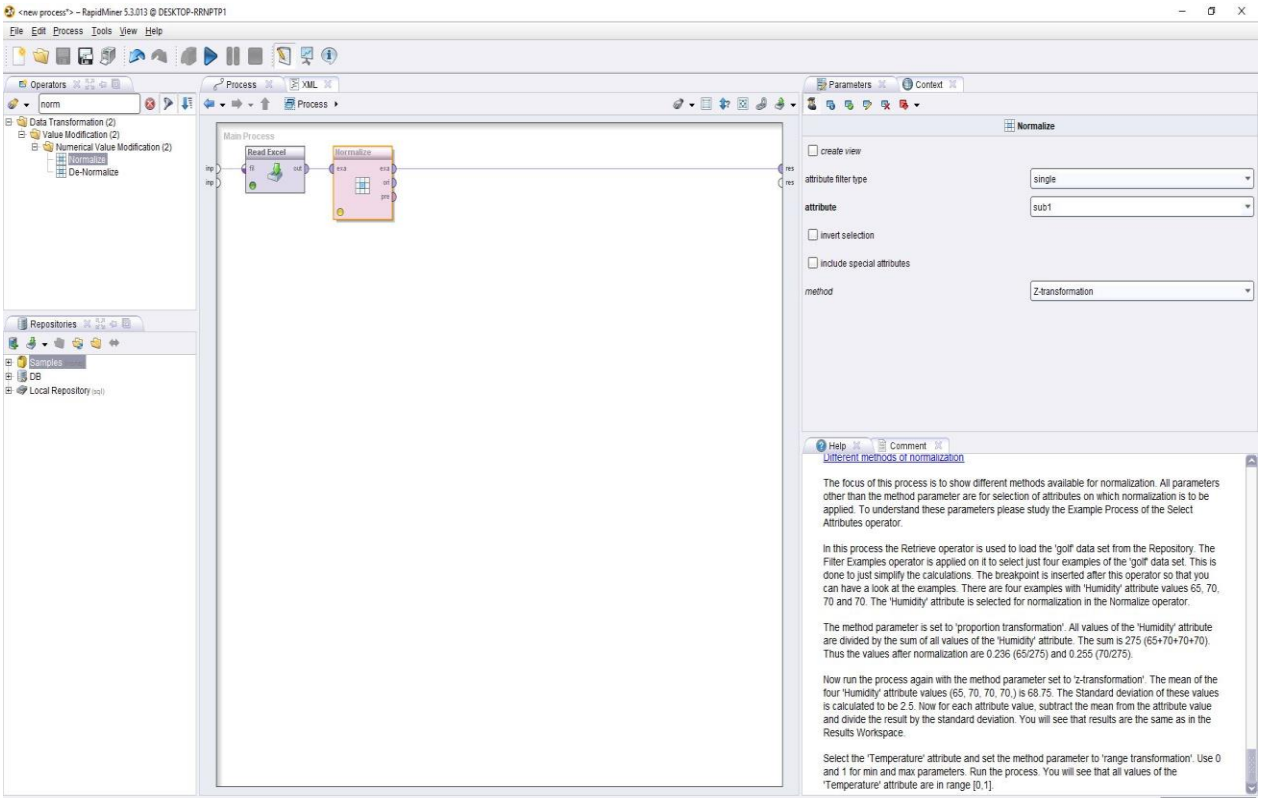
Replace: Replace the particular attribute's data values with the given values.

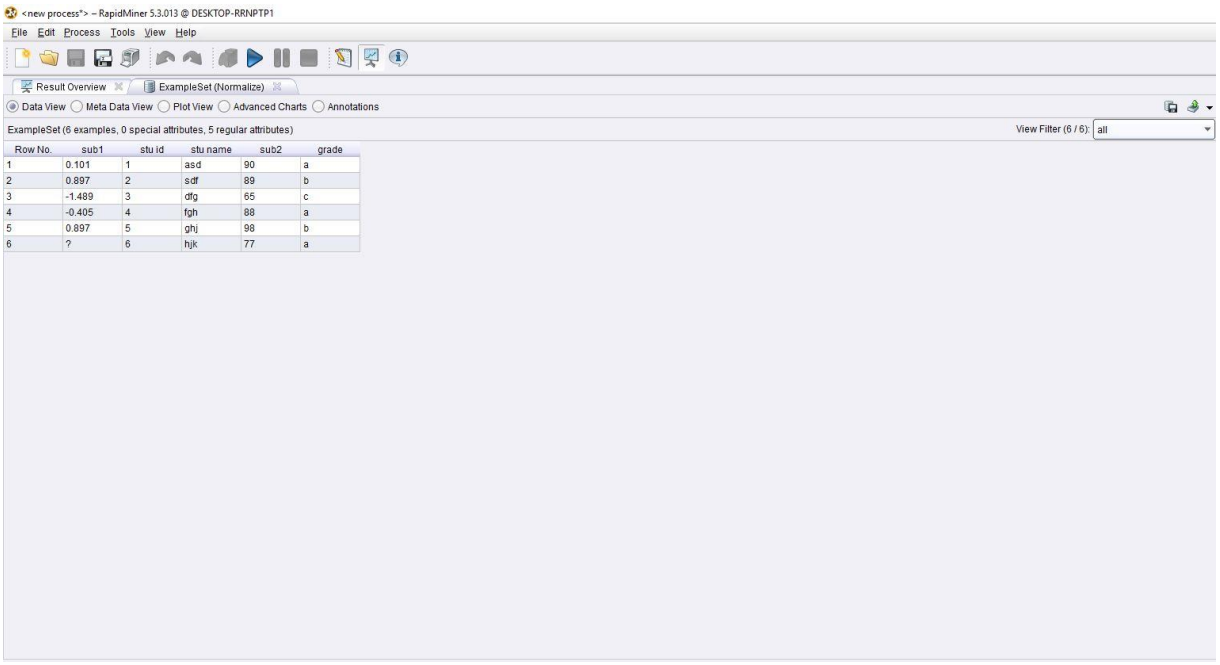




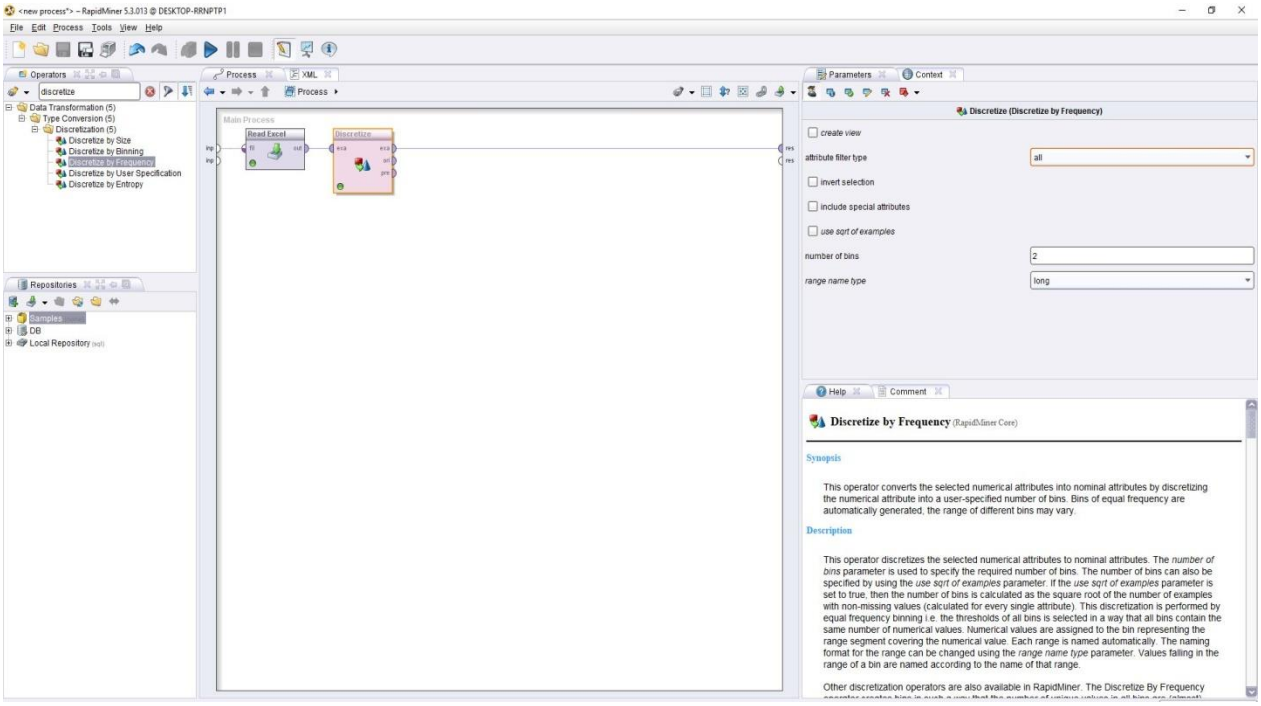
Row No.	stu name	grade	stu id	sub1	sub2
1	Asd	A	1	67	90
2	sdf	b	2	78	89
3	dfg	c	3	45	65
4	fgh	A	4	60	88
5	ghj	b	5	78	98
6	hjk	A	6	?	77

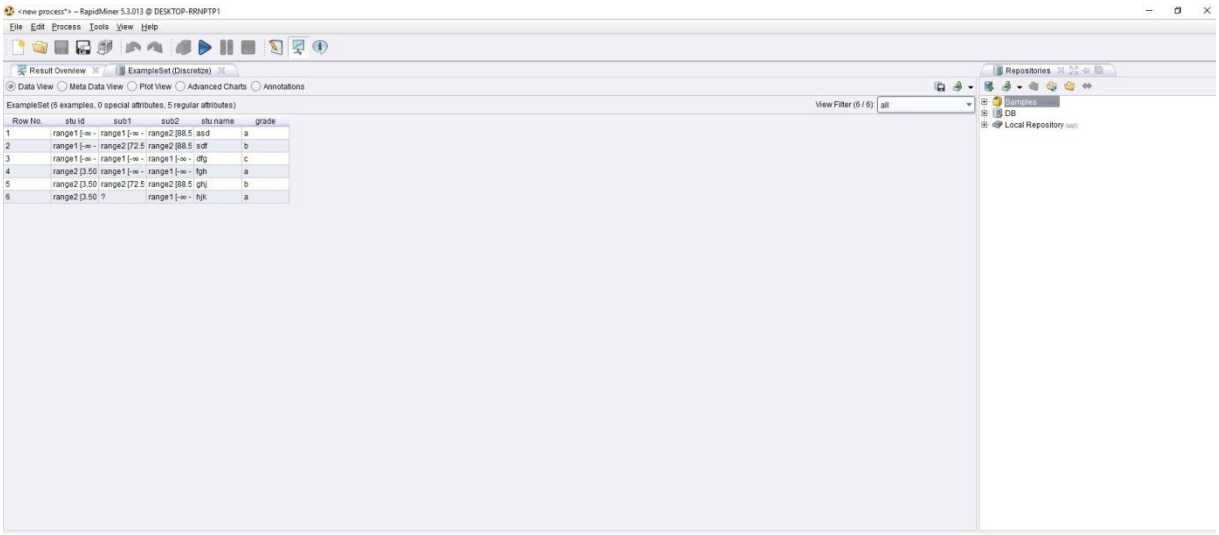
Normalize: Normalize the given attribute values into 0.0 to 1.0 range.





Discretize:





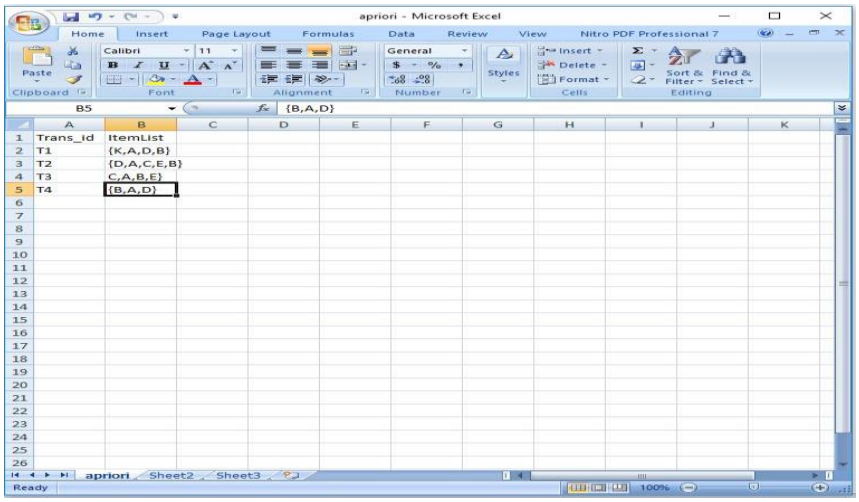
Result: Pre-processing in rapid miner is done successfully.

Task:9

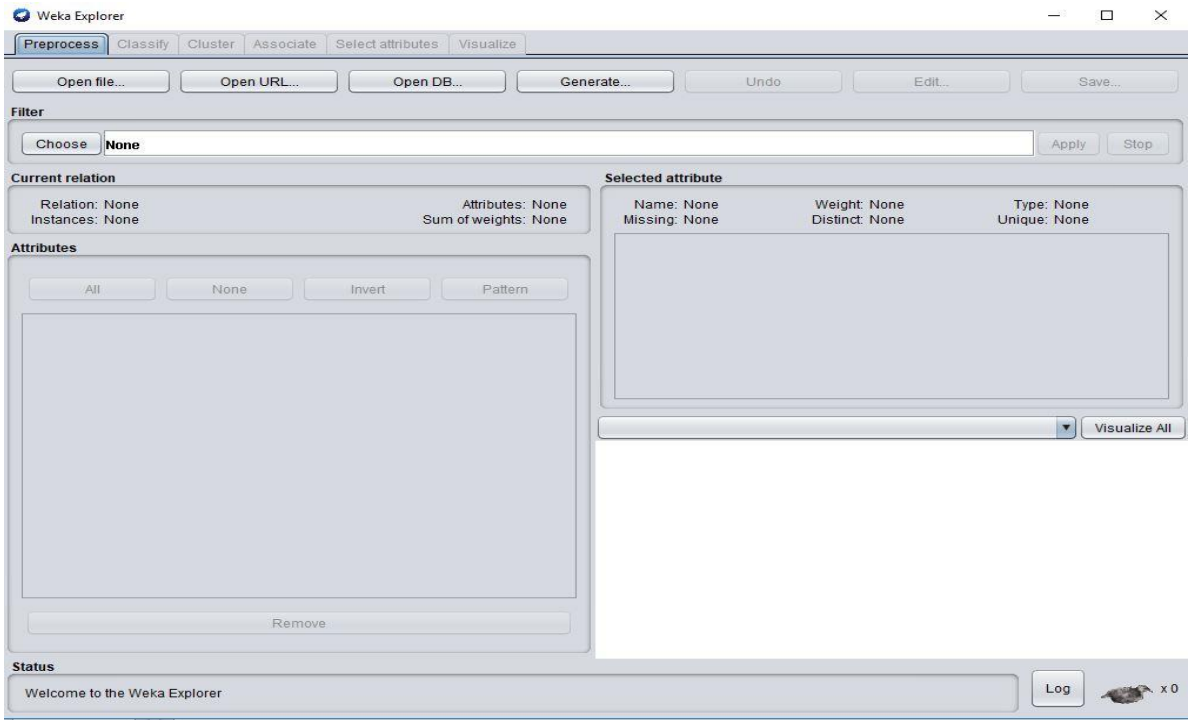
Aim: To implement apriori algorithm for the given data.

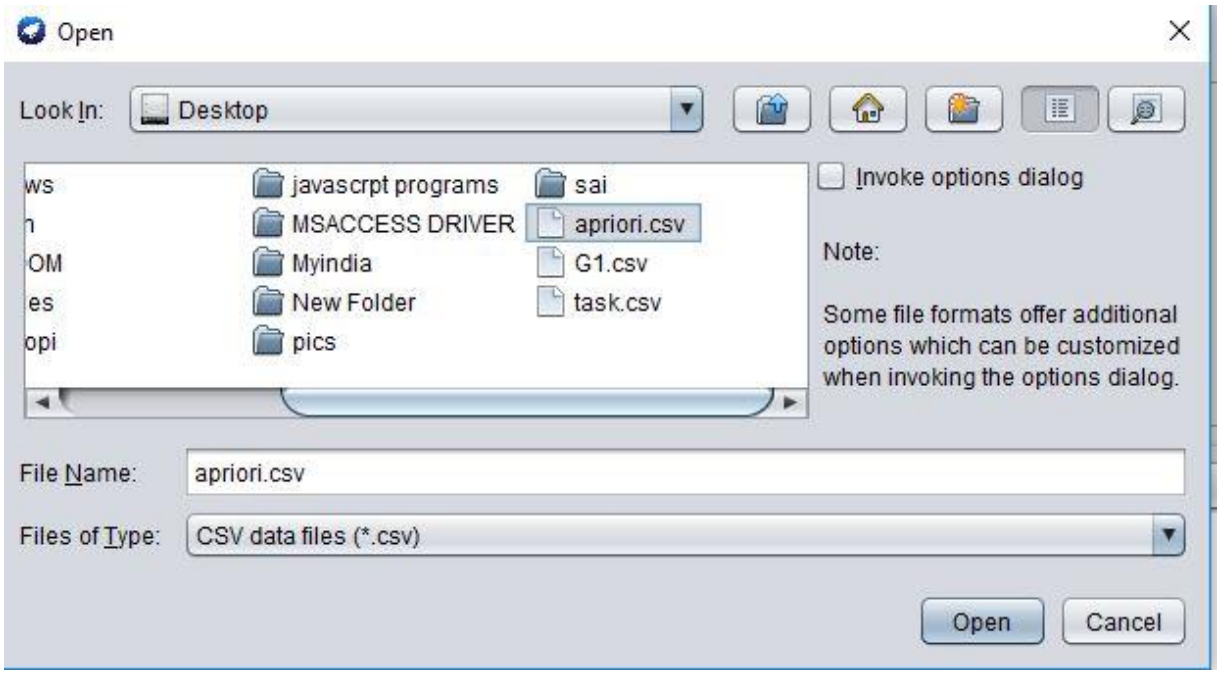
Procedure:

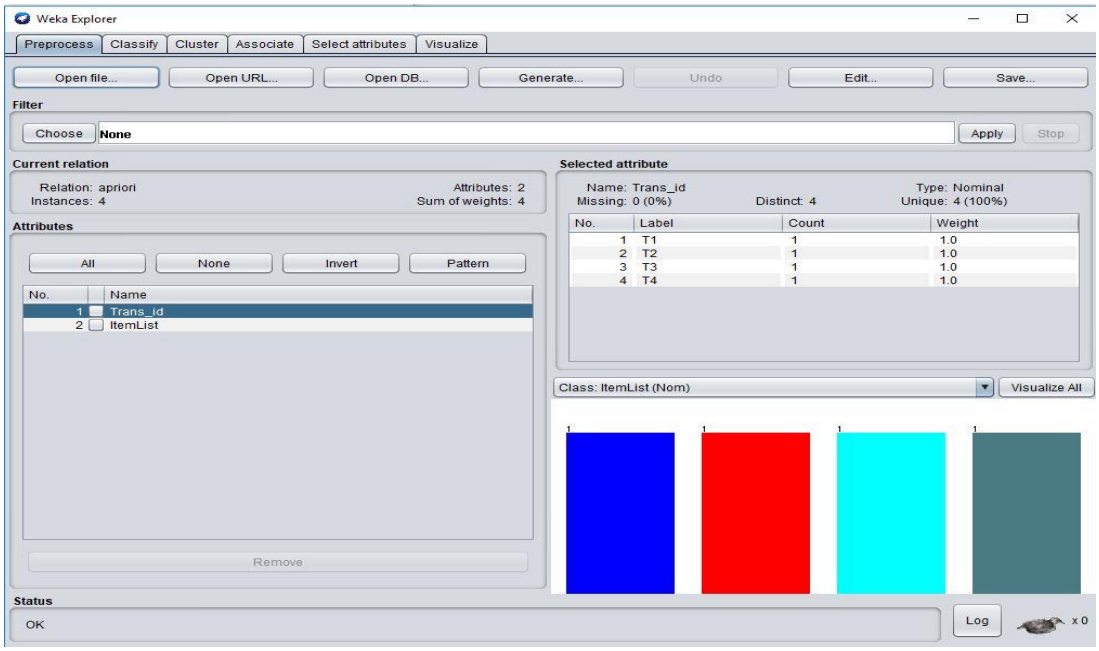
1. First create a csv file.

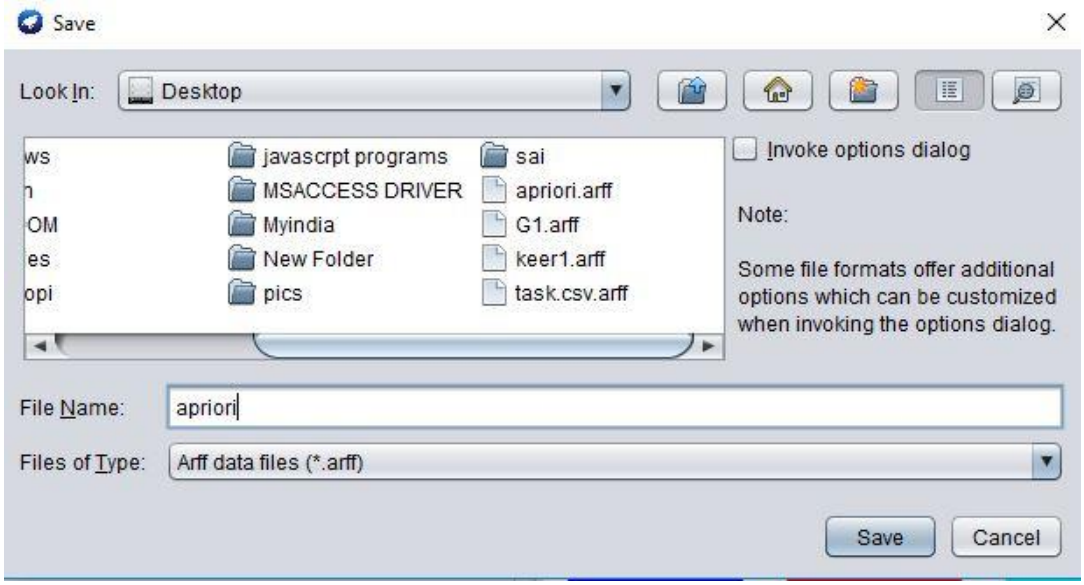


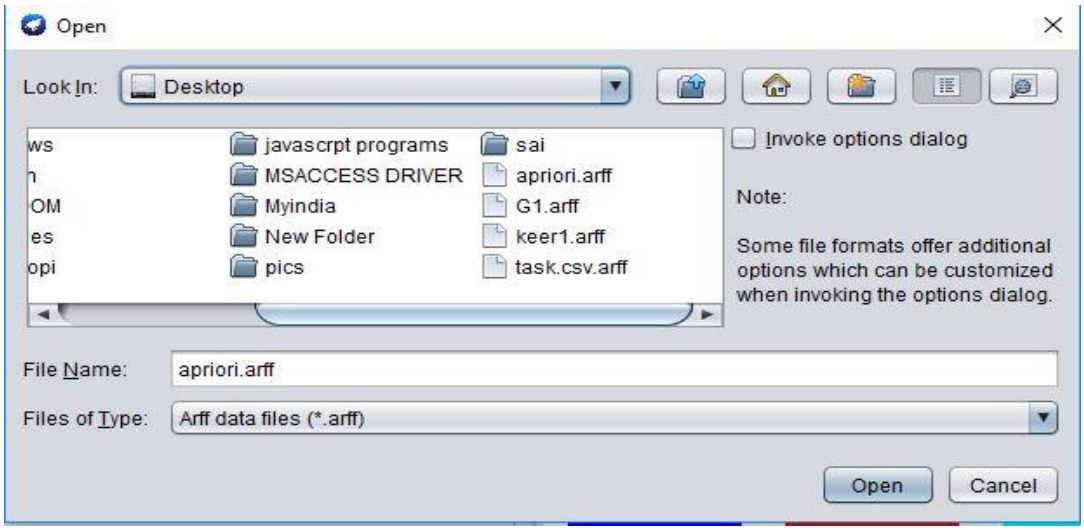
2. Now open weka and convert that csv file into arff file and then open it.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: None
Instances: None
Attributes: None
Sum of weights: None
Selected attribute
Name: None
Missing: None
Weight: None
Distinct: None
Type: None
Unique: None
Attributes
All
None
Invert
Pattern
Remove
Visualize All
Status
Welcome to the Weka Explorer
Log
x 0

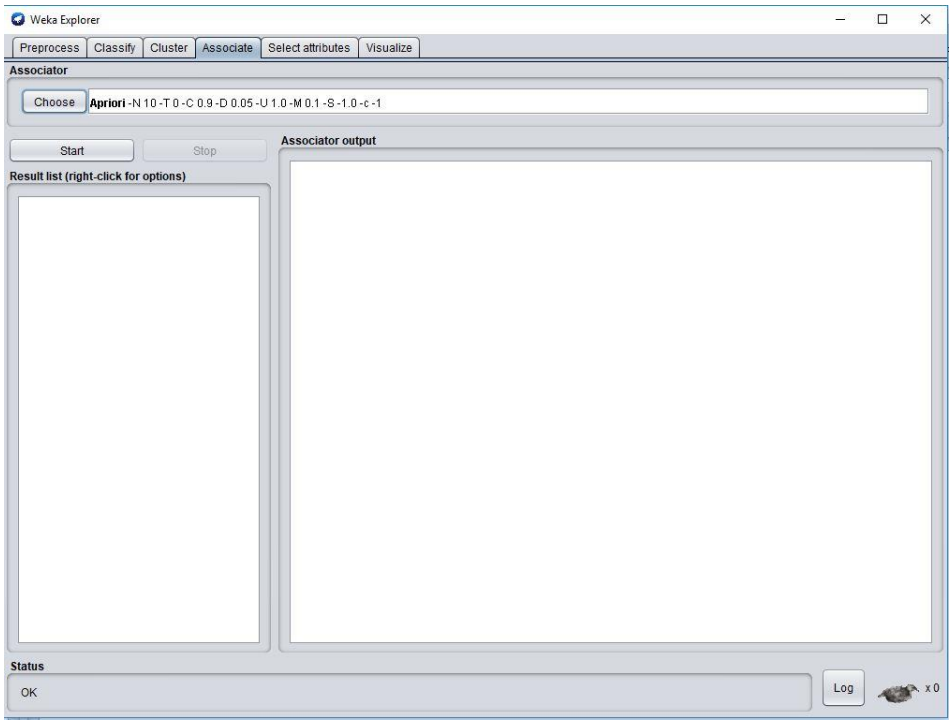

Open
Look In:
Desktop
ws
n
OM
es
opi
javascrpt programs
MSACCESS DRIVER
Myindia
New Folder
pics
sai
apriori.csv
G1.csv
task.csv
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
File Name:
apriori.csv
Files of Type:
CSV data files (*.csv)
Open
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: apriori
Instances: 4
Attributes: 2
Sum of weights: 4
Attributes
All
None
Invert
Pattern
No.
Name
1 Trans_id
2 ItemList
Remove
Selected attribute
Name: Trans_id
Missing: 0 (0%)
Distinct: 4
Type: Nominal
Unique: 4 (100%)
No.
Label
Count
Weight
1 T1 1 1.0
2 T2 1 1.0
3 T3 1 1.0
4 T4 1 1.0
Class: ItemList (Nom)
Visualize All
Status
OK
Log
x 0

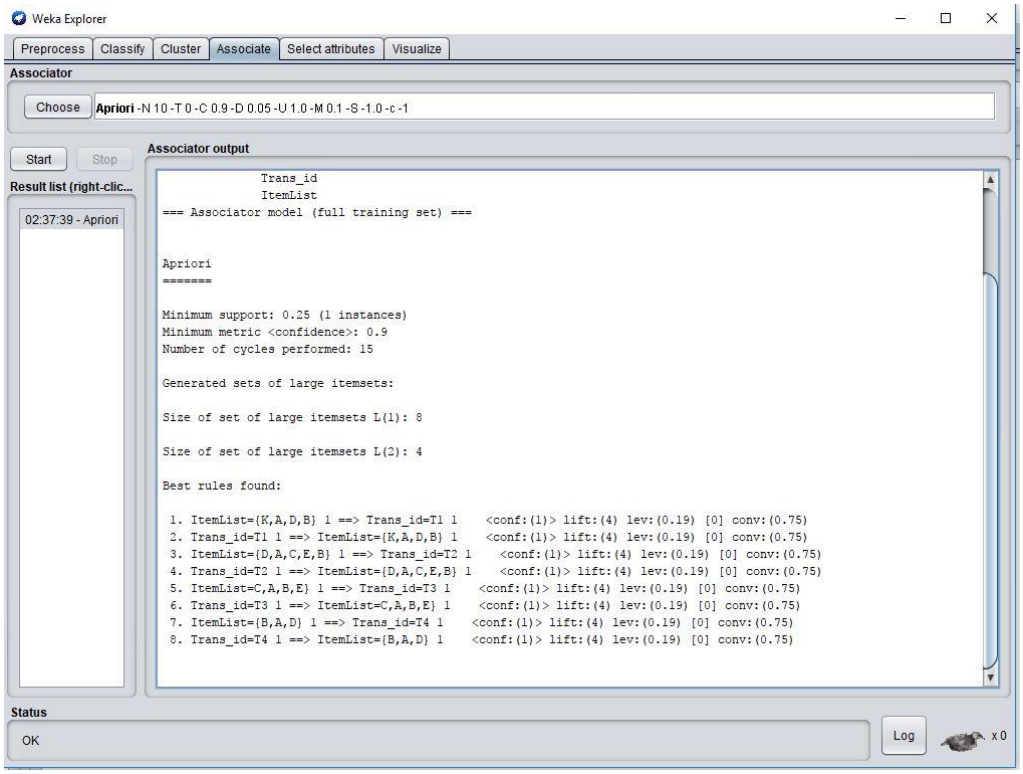

Save
Look In:
Desktop
ws
n
OM
es
opi
javascrpt programs
MSACCESS DRIVER
Myindia
New Folder
pics
sai
apriori.arff
G1.arff
keer1.arff
task.csv.arff
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
File Name:
apriori
Files of Type:
Arff data files (*.arff)
Save
Cancel


Open
Look In:
Desktop
ws
n
OM
es
opi
javascrpt programs
MSACCESS DRIVER
Myindia
New Folder
pics
sai
apriori.arff
G1.arff
keer1.arff
task.csv.arff
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
File Name:
apriori.arff
Files of Type:
Arff data files (*.arff)
Open
Cancel

3. Now go to “Associate” and then to “Apriori algorithm” and if necessary preprocessing techniques are used before only. Later start the process and rules are generated to the frequent items of given dataset.



Output:

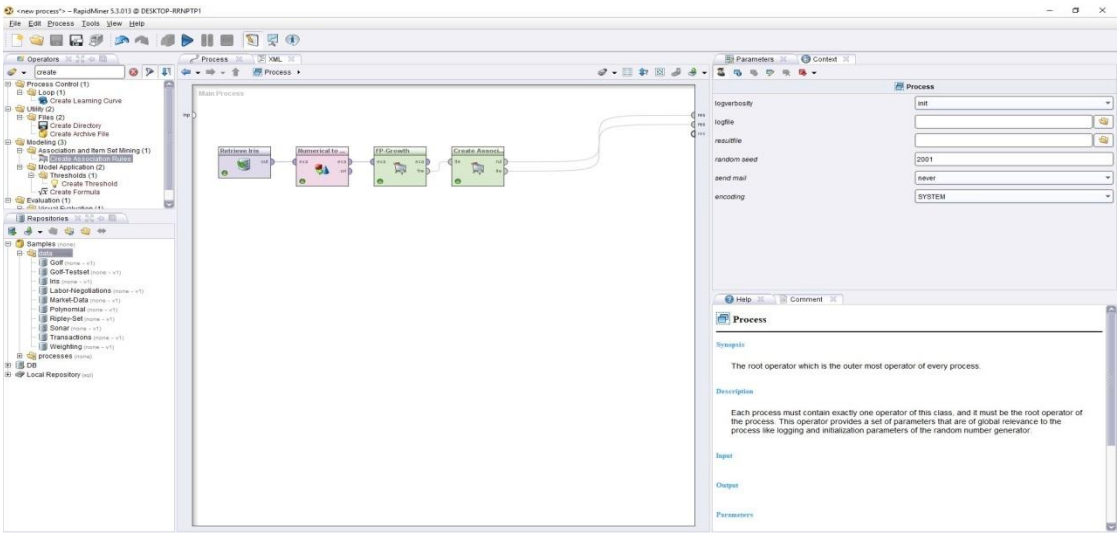


Task-10

Aim: To implement fp growth algorithm for a given dataset.

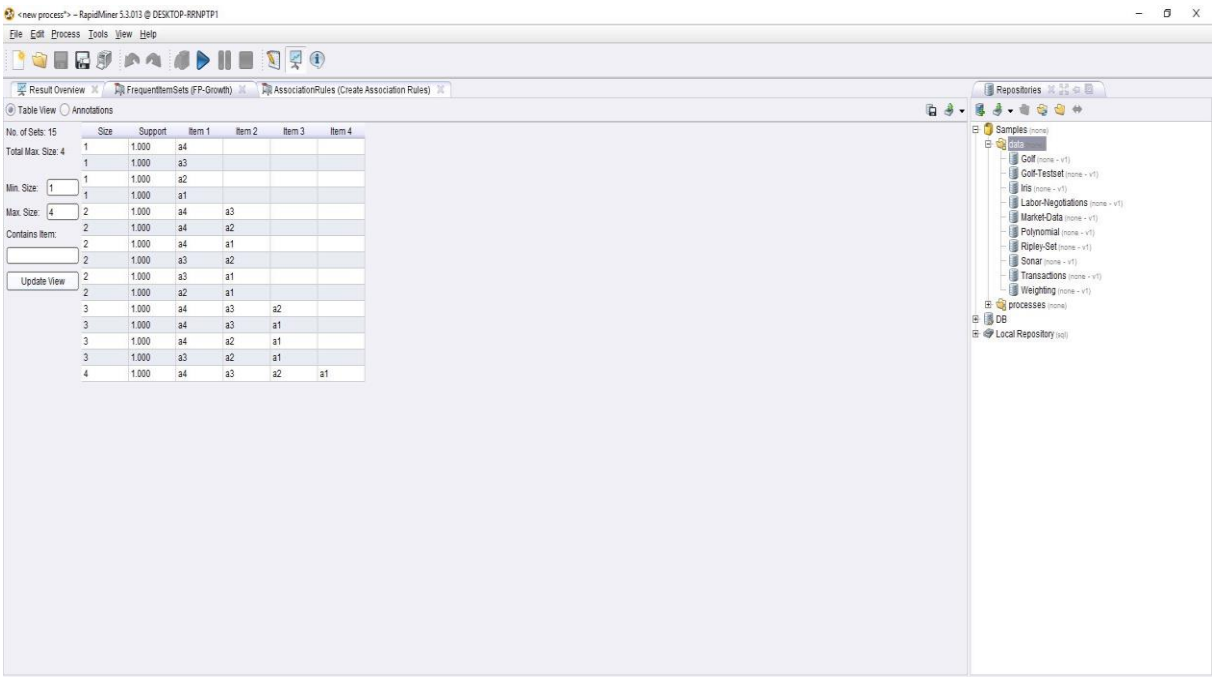
Procedure:

1. First retrieve iris dataset by selecting samples>data>iris in repositories.
2. Now drag and drop the Numerical to Binomial operator and also Fp growth and create association rules operators.
3. Now join the input and output connections properly. And finally click on run.



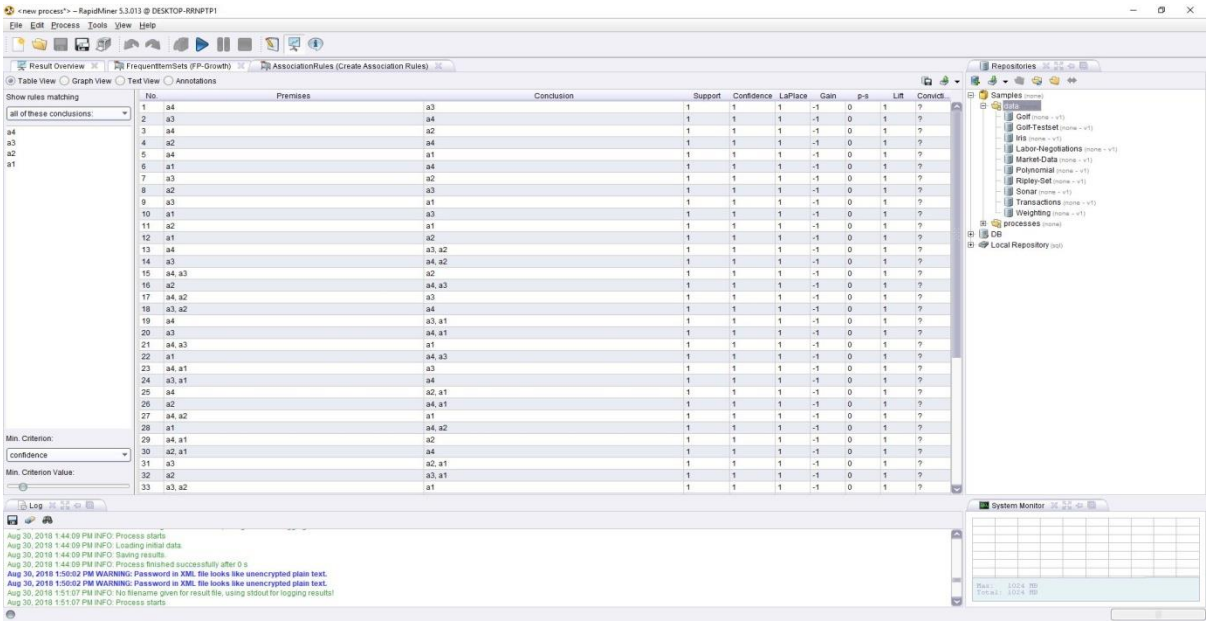
Output:

Frequent items sets:

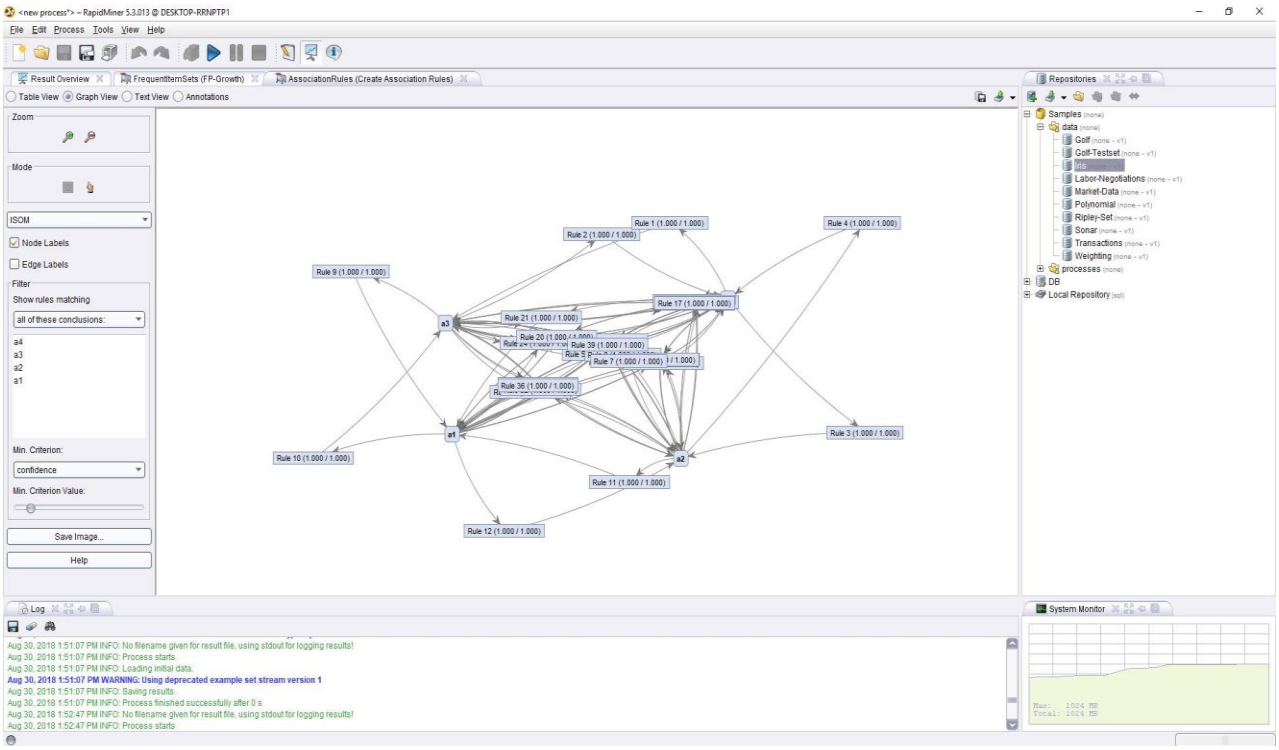


Size	Support	Item 1	Item 2	Item 3	Item 4
1	1.000	a4			
1	1.000	a3			
1	1.000	a2			
1	1.000	a1			
2	1.000	a4	a3		
2	1.000	a4	a2		
2	1.000	a4	a1		
2	1.000	a3	a2		
2	1.000	a3	a1		
2	1.000	a2	a1		
3	1.000	a4	a3	a2	
3	1.000	a4	a3	a1	
3	1.000	a4	a2	a1	
3	1.000	a3	a2	a1	
4	1.000	a4	a3	a2	a1

Association Rules:


Result Overview
FrequentItemSets (FP-Growth)
AssociationRules (Create Association Rules)
Table View
Graph View
Text View
Annotations
Show rules matching
all of these conclusions:
a4
a3
a2
a1
Min. Criterion:
confidence
Min. Criterion Value:
No. Premises Conclusion Support Confidence LaPlace Gain p-s Lift Convicti...
1 a4 a3 1 1 1 -1 0 1 ?
2 a3 a4 1 1 1 -1 0 1 ?
3 a4 a2 1 1 1 -1 0 1 ?
4 a2 a4 1 1 1 -1 0 1 ?
5 a4 a1 1 1 1 -1 0 1 ?
6 a1 a4 1 1 1 -1 0 1 ?
7 a3 a2 1 1 1 -1 0 1 ?
8 a2 a3 1 1 1 -1 0 1 ?
9 a3 a1 1 1 1 -1 0 1 ?
10 a1 a3 1 1 1 -1 0 1 ?
11 a2 a1 1 1 1 -1 0 1 ?
12 a1 a2 1 1 1 -1 0 1 ?
13 a4 a3, a2 1 1 1 -1 0 1 ?
14 a3 a4, a2 1 1 1 -1 0 1 ?
15 a4, a3 a2 1 1 1 -1 0 1 ?
16 a2 a4, a3 1 1 1 -1 0 1 ?
17 a4, a2 a3 1 1 1 -1 0 1 ?
18 a3, a2 a4 1 1 1 -1 0 1 ?
19 a4 a3, a1 1 1 1 -1 0 1 ?
20 a3 a4, a1 1 1 1 -1 0 1 ?
21 a4, a3 a1 1 1 1 -1 0 1 ?
22 a1 a4, a3 1 1 1 -1 0 1 ?
23 a4, a1 a3 1 1 1 -1 0 1 ?
24 a3, a1 a4 1 1 1 -1 0 1 ?
25 a4 a2, a1 1 1 1 -1 0 1 ?
26 a2 a4, a1 1 1 1 -1 0 1 ?
27 a4, a2 a1 1 1 1 -1 0 1 ?
28 a1 a4, a2 1 1 1 -1 0 1 ?
29 a4, a1 a2 1 1 1 -1 0 1 ?
30 a2, a1 a4 1 1 1 -1 0 1 ?
31 a3 a2, a1 1 1 1 -1 0 1 ?
32 a2 a3, a1 1 1 1 -1 0 1 ?
33 a3, a2 a1 1 1 1 -1 0 1 ?
Repositories
Samples (none)
data (none)
Golf (none - v1)
Golf-Testset (none - v1)
Iris (none - v1)
Labor-Negotiations (none - v1)
Market-Data (none - v1)
Polynomial (none - v1)
Ripley-Set (none - v1)
Sonar (none - v1)
Transactions (none - v1)
Weighting (none - v1)
processes (none)
DB
Local Repository (sql)
Log
System Monitor
Aug 30, 2018 1:44:09 PM INFO: Process starts.
Aug 30, 2018 1:44:09 PM INFO: Loading initial data.
Aug 30, 2018 1:44:09 PM INFO: Saving results.
Aug 30, 2018 1:44:09 PM INFO: Process finished successfully after 0 s
Aug 30, 2018 1:50:02 PM WARNING: Password in XML file looks like unencrypted plain text.
Aug 30, 2018 1:50:02 PM WARNING: Password in XML file looks like unencrypted plain text.
Aug 30, 2018 1:51:07 PM INFO: No filename given for result file, using stdout for logging results!
Aug 30, 2018 1:51:07 PM INFO: Process starts.

Graph View:

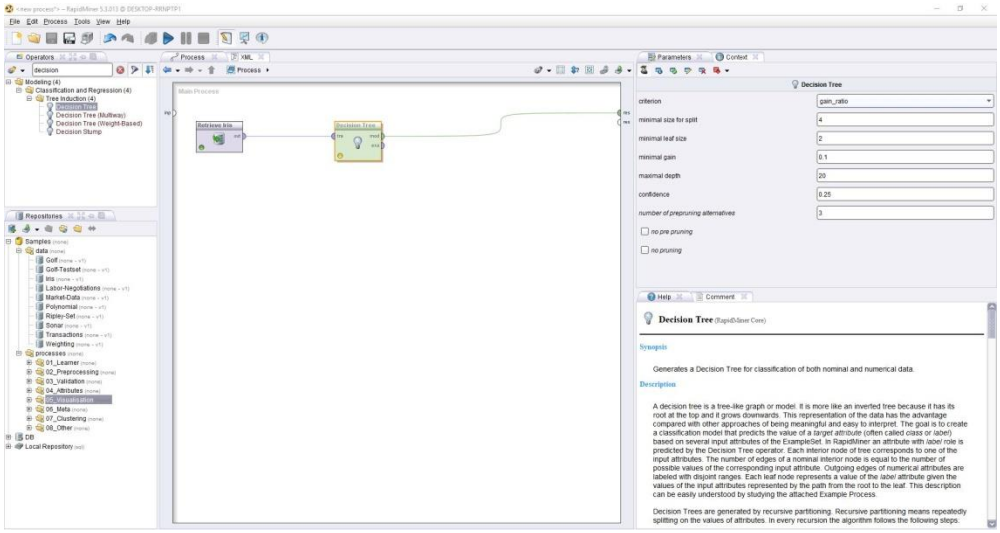

Result Overview
FrequentItemSets (FP-Growth)
AssociationRules (Create Association Rules)
Table View
Graph View
Text View
Annotations
Zoom
Mode
ISOM
Node Labels
Edge Labels
Filter
Show rules matching
all of these conclusions:
a4
a3
a2
a1
Min. Criterion:
confidence
Min. Criterion Value:
Save Image...
Help
Rule 1 (1.000 / 1.000)
Rule 2 (1.000 / 1.000)
Rule 3 (1.000 / 1.000)
Rule 4 (1.000 / 1.000)
Rule 9 (1.000 / 1.000)
Rule 10 (1.000 / 1.000)
Rule 11 (1.000 / 1.000)
Rule 12 (1.000 / 1.000)
Rule 17 (1.000 / 1.000)
Rule 21 (1.000 / 1.000)
Rule 20 (1.000 / 1.000)
Rule 39 (1.000 / 1.000)
Rule 7 (1.000 / 1.000)
Rule 36 (1.000 / 1.000)
a3
a1
a2
Repositories
Samples (none)
data (none)
Golf (none - v1)
Golf-Testset (none - v1)
Iris (none - v1)
Labor-Negotiations (none - v1)
Market-Data (none - v1)
Polynomial (none - v1)
Ripley-Set (none - v1)
Sonar (none - v1)
Transactions (none - v1)
Weighting (none - v1)
processes (none)
DB
Local Repository (sql)
Log
System Monitor
Aug 30, 2018 1:51:07 PM INFO: No filename given for result file, using stdout for logging results!
Aug 30, 2018 1:51:07 PM INFO: Process starts.
Aug 30, 2018 1:51:07 PM INFO: Loading initial data.
Aug 30, 2018 1:51:07 PM WARNING: Using deprecated example set stream version 1
Aug 30, 2018 1:51:07 PM INFO: Saving results.
Aug 30, 2018 1:51:07 PM INFO: Process finished successfully after 0 s
Aug 30, 2018 1:52:47 PM INFO: No filename given for result file, using stdout for logging results!
Aug 30, 2018 1:52:47 PM INFO: Process starts.

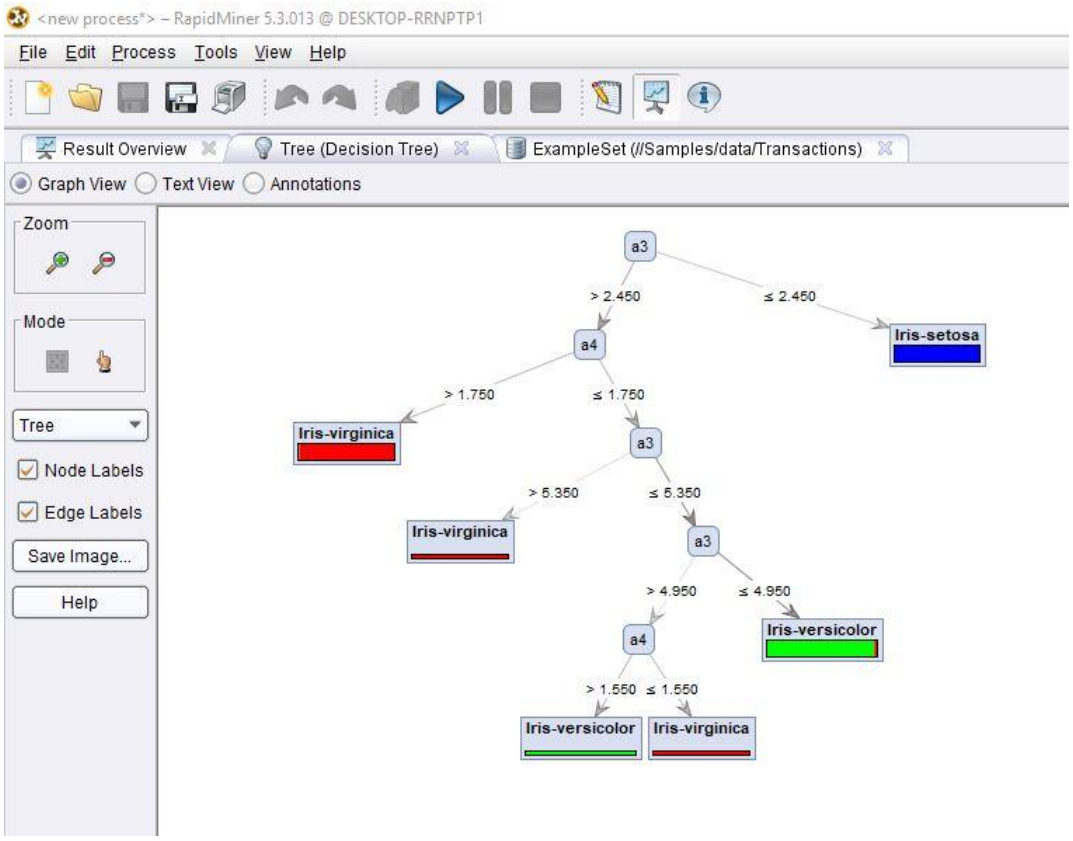
Task:11

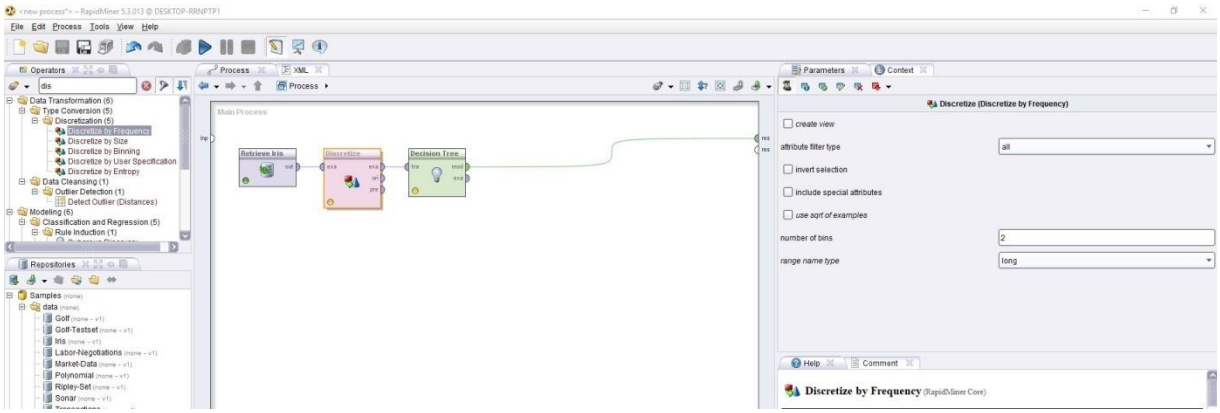
Aim: To implement decision tree using rapid miner tool.

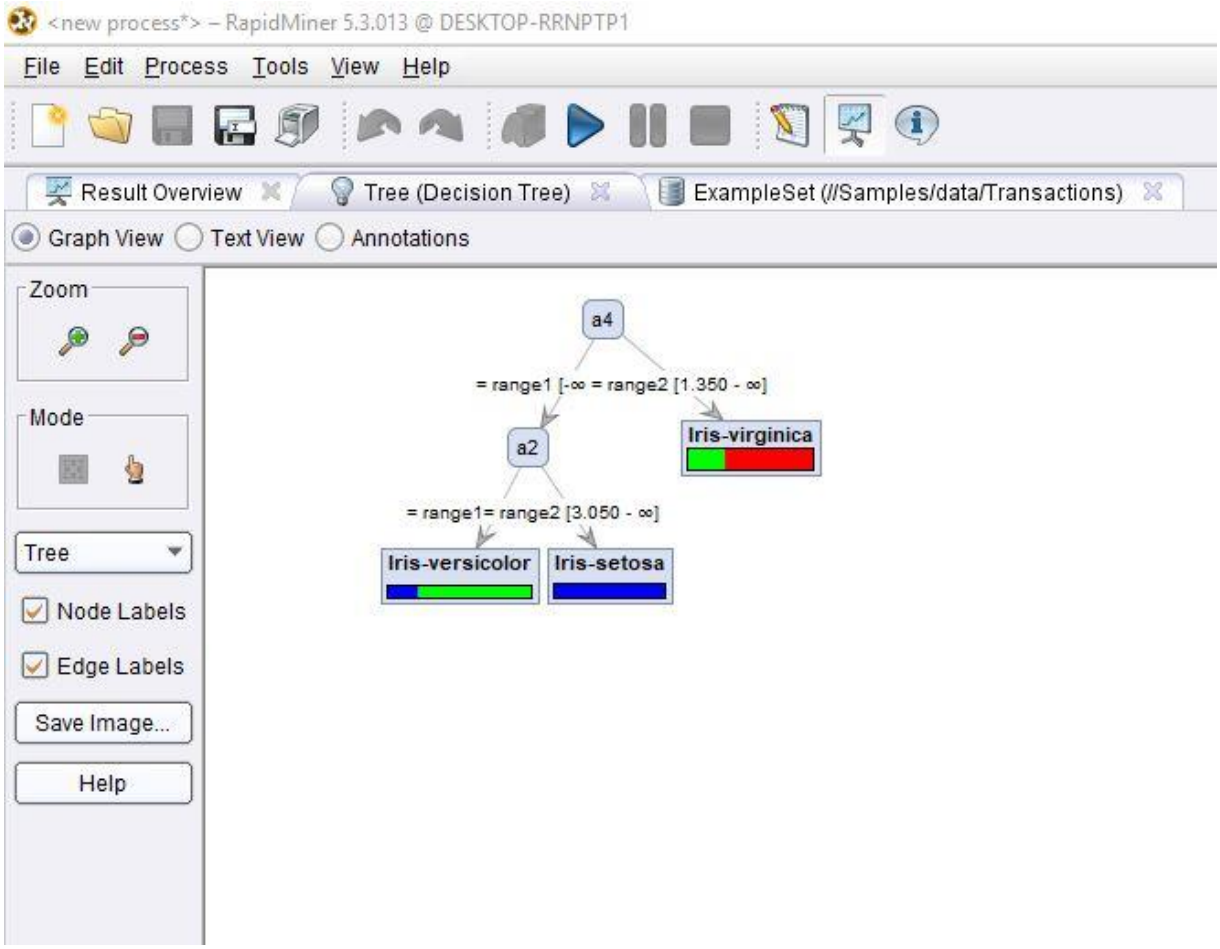
Procedure:

1. First create a new process and then retrieve the iris dataset. If necessary discretize the dataset by frequency.
2. Now drag and drop the decision tree operator.
3. Finally connect the input and output connections and click on run.







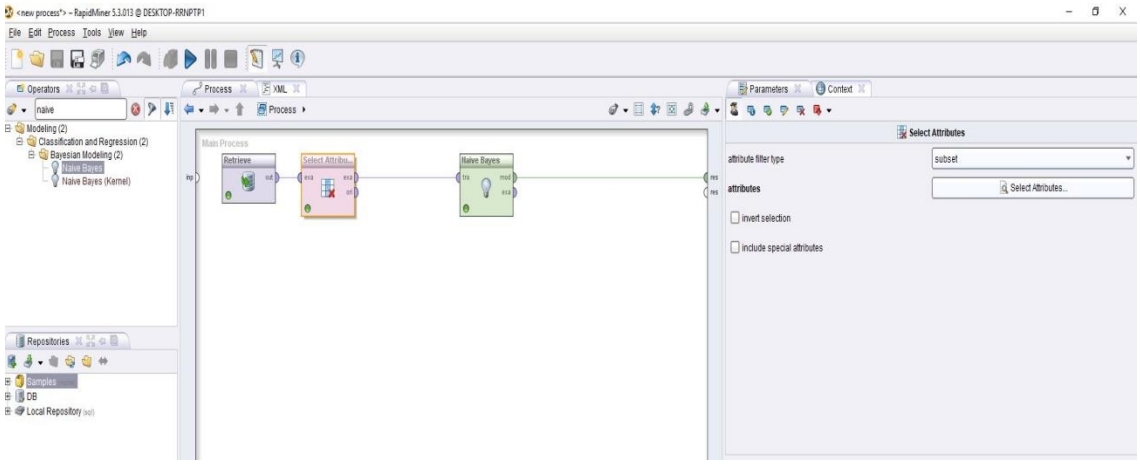

<new process*> – RapidMiner 5.3.013 @ DESKTOP-RRNPTP1
File Edit Process Tools View Help
Result Overview
Tree (Decision Tree)
ExampleSet (//Samples/data/Transactions)
Graph View
Text View
Annotations
Zoom
Mode
Tree
Node Labels
Edge Labels
Save Image...
Help
a4
= range1 [-∞ = range2 [1.350 - ∞]
a2
Iris-virginica
= range1= range2 [3.050 - ∞]
Iris-versicolor
Iris-setosa

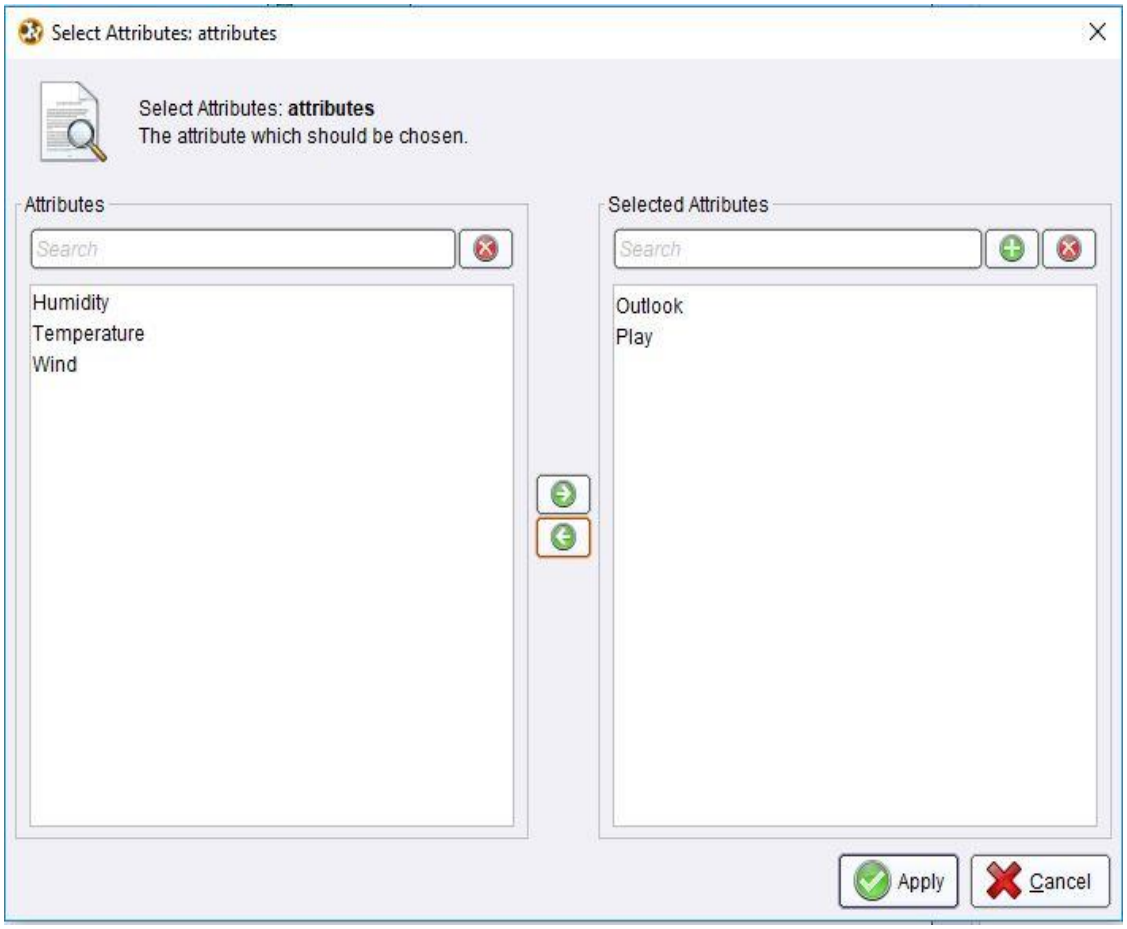
Task:12

Aim: To implement Bayesian classification using naïve bayes algorithm.

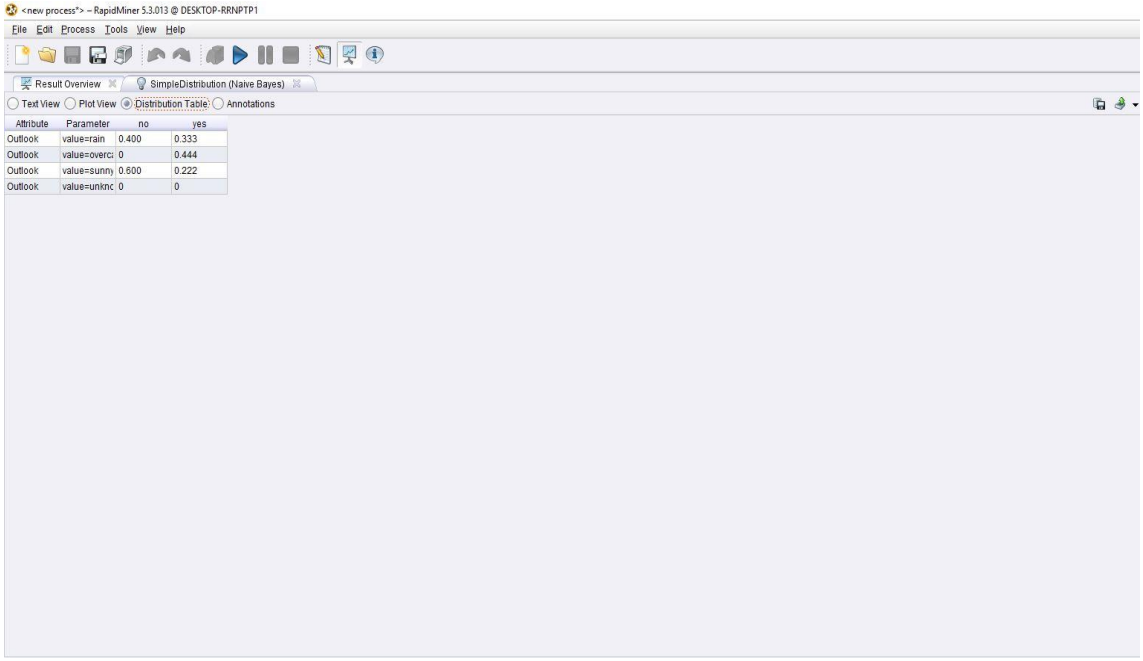
Procedure:

1. First open the new process and retrieve the iris dataset.
2. Now drag and drop the “select attributes” and naive bayes operators.
3. In “select attributes”, select the attribute which you desired.
4. Now join the input and output connections properly.





Output:



Result: Successfully implemented naive bayes algorithm.

Task:13

Aim: To implement k-means clustering in rapid miner.

Description:

Cluster: Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster and dissimilar objects are grouped in another cluster.

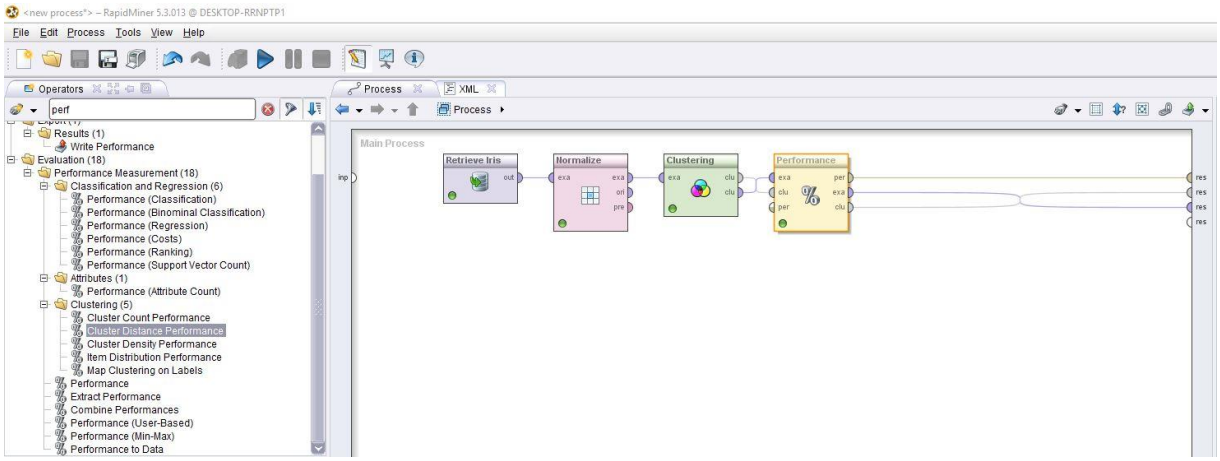
1. While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups.
2. The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.
3. Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.

The following points throw light on why clustering is required in data mining :

1. Scalability – We need highly scalable clustering algorithms to deal with large databases.
2. Ability to deal with different kinds of attributes − Algorithms should be capable to be applied on any kind of data such as interval-based (numerical) data, categorical, and binary data.
3. Discovery of clusters with attribute shape – The clustering algorithm should be capable of detecting clusters of arbitrary shape. They should not be bounded to only distance measures that tend to find spherical cluster of small sizes.
4. High dimensionality − The clustering algorithm should not only be able to handle low-dimensional data but also the high dimensional space.

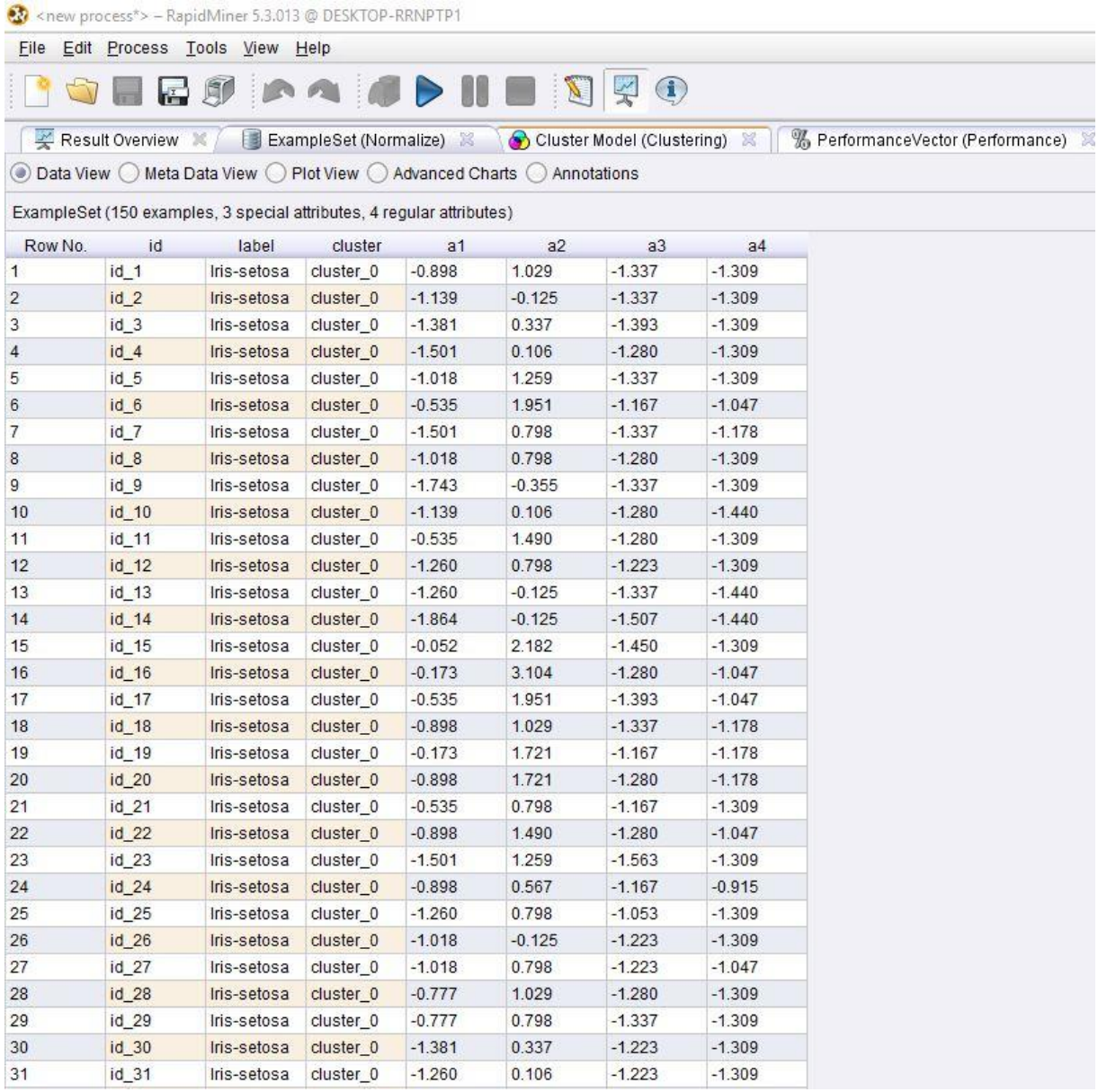
Procedure:

1. First of all open rapid miner and retrieve the iris dataset.
2. Now normalize the data by drag and dropping the “normalize”.
3. Now drag and drop the K-means clustering and then cluster distance performance which is used for performance evaluation of centroid based clustering methods.
4. Then connect the input and output connections properly.



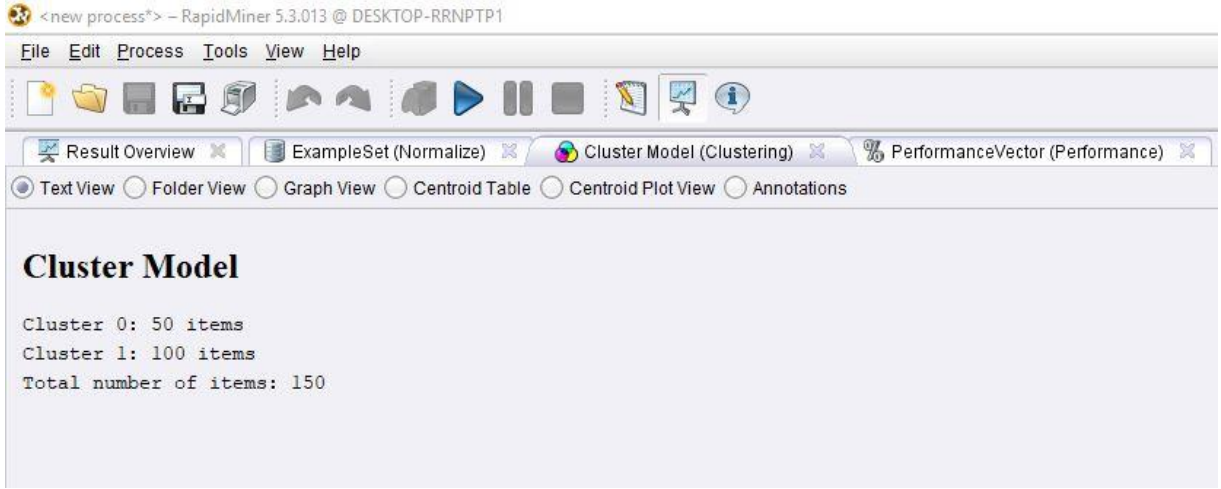
Output:

Example set:

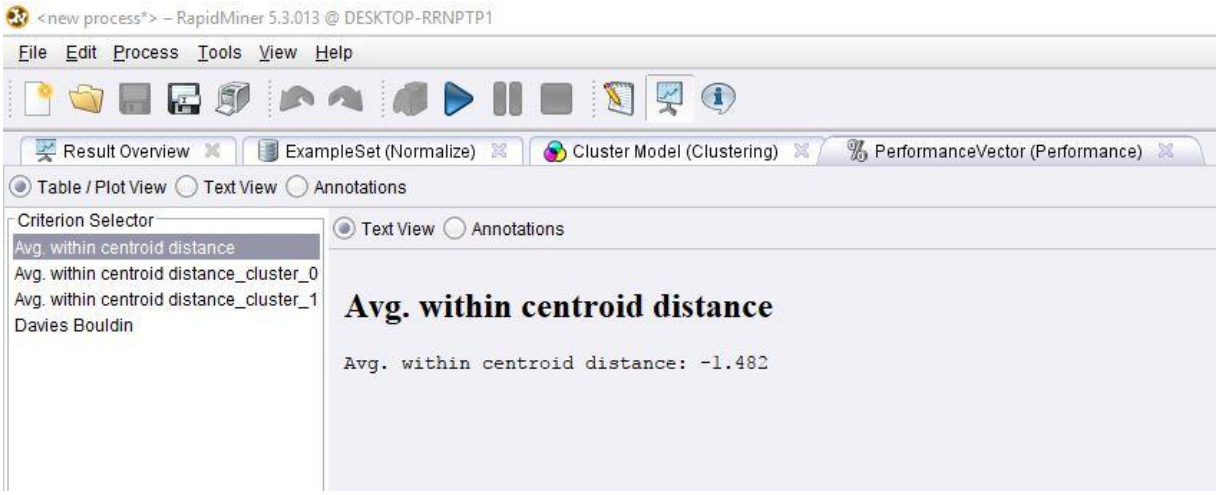


Row No.	id	label	cluster	a1	a2	a3	a4
1	id_1	Iris-setosa	cluster_0	-0.898	1.029	-1.337	-1.309
2	id_2	Iris-setosa	cluster_0	-1.139	-0.125	-1.337	-1.309
3	id_3	Iris-setosa	cluster_0	-1.381	0.337	-1.393	-1.309
4	id_4	Iris-setosa	cluster_0	-1.501	0.106	-1.280	-1.309
5	id_5	Iris-setosa	cluster_0	-1.018	1.259	-1.337	-1.309
6	id_6	Iris-setosa	cluster_0	-0.535	1.951	-1.167	-1.047
7	id_7	Iris-setosa	cluster_0	-1.501	0.798	-1.337	-1.178
8	id_8	Iris-setosa	cluster_0	-1.018	0.798	-1.280	-1.309
9	id_9	Iris-setosa	cluster_0	-1.743	-0.355	-1.337	-1.309
10	id_10	Iris-setosa	cluster_0	-1.139	0.106	-1.280	-1.440
11	id_11	Iris-setosa	cluster_0	-0.535	1.490	-1.280	-1.309
12	id_12	Iris-setosa	cluster_0	-1.260	0.798	-1.223	-1.309
13	id_13	Iris-setosa	cluster_0	-1.260	-0.125	-1.337	-1.440
14	id_14	Iris-setosa	cluster_0	-1.864	-0.125	-1.507	-1.440
15	id_15	Iris-setosa	cluster_0	-0.052	2.182	-1.450	-1.309
16	id_16	Iris-setosa	cluster_0	-0.173	3.104	-1.280	-1.047
17	id_17	Iris-setosa	cluster_0	-0.535	1.951	-1.393	-1.047
18	id_18	Iris-setosa	cluster_0	-0.898	1.029	-1.337	-1.178
19	id_19	Iris-setosa	cluster_0	-0.173	1.721	-1.167	-1.178
20	id_20	Iris-setosa	cluster_0	-0.898	1.721	-1.280	-1.178
21	id_21	Iris-setosa	cluster_0	-0.535	0.798	-1.167	-1.309
22	id_22	Iris-setosa	cluster_0	-0.898	1.490	-1.280	-1.047
23	id_23	Iris-setosa	cluster_0	-1.501	1.259	-1.563	-1.309
24	id_24	Iris-setosa	cluster_0	-0.898	0.567	-1.167	-0.915
25	id_25	Iris-setosa	cluster_0	-1.260	0.798	-1.053	-1.309
26	id_26	Iris-setosa	cluster_0	-1.018	-0.125	-1.223	-1.309
27	id_27	Iris-setosa	cluster_0	-1.018	0.798	-1.223	-1.047
28	id_28	Iris-setosa	cluster_0	-0.777	1.029	-1.280	-1.309
29	id_29	Iris-setosa	cluster_0	-0.777	0.798	-1.337	-1.309
30	id_30	Iris-setosa	cluster_0	-1.381	0.337	-1.223	-1.309
31	id_31	Iris-setosa	cluster_0	-1.260	0.106	-1.223	-1.309

Cluster Model:



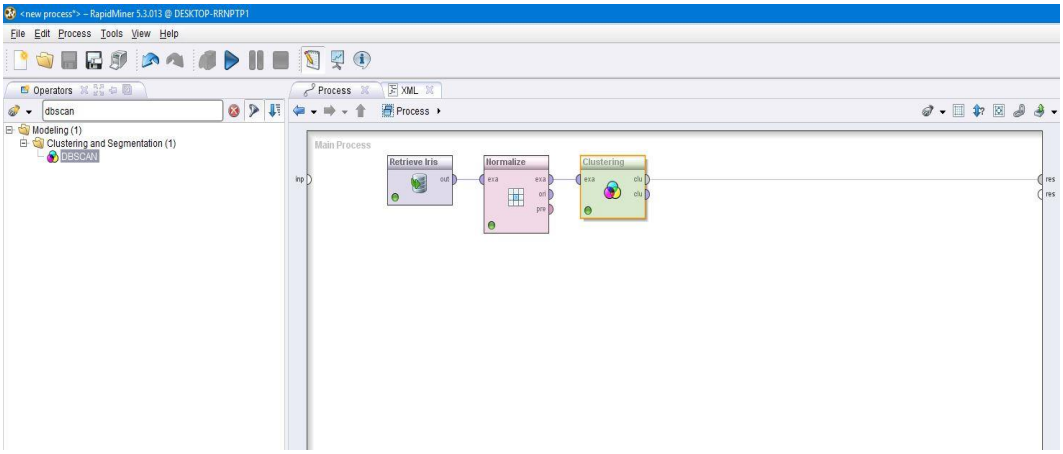
Performance Vector:



Aim: To implement DBscan clustering in rapid miner.

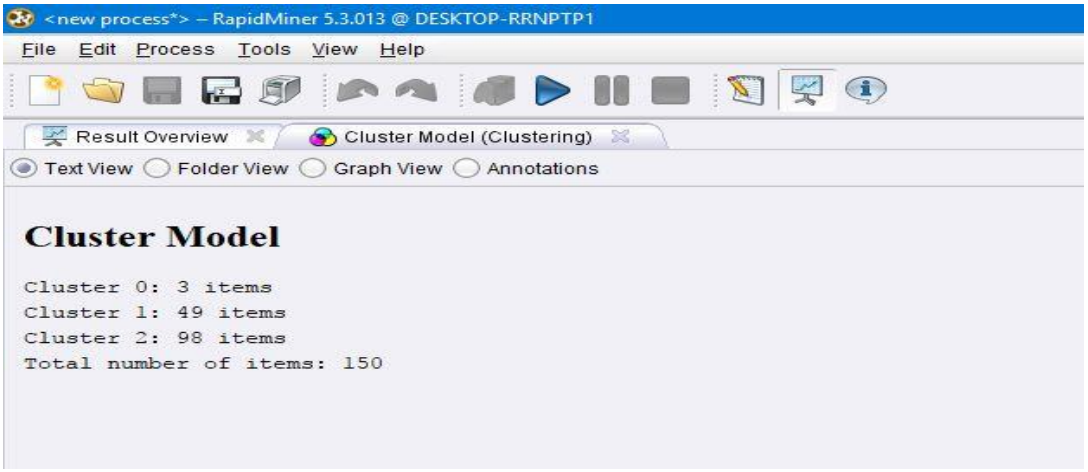
Procedure:

1. First of all open rapid miner and retrieve the iris dataset.
2. Now normalize the data by drag and dropping the “normalize”.
3. Now drag and drop the DBscan clustering.
4. Then connect the input and output connections properly.

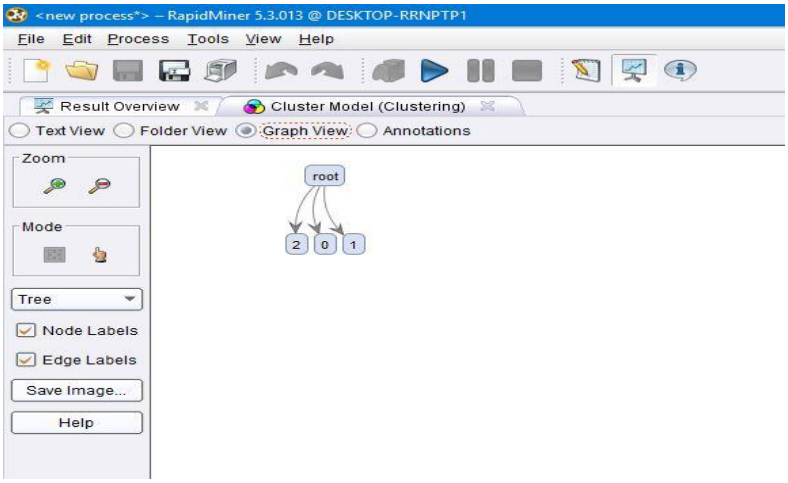


Output:

Text View:



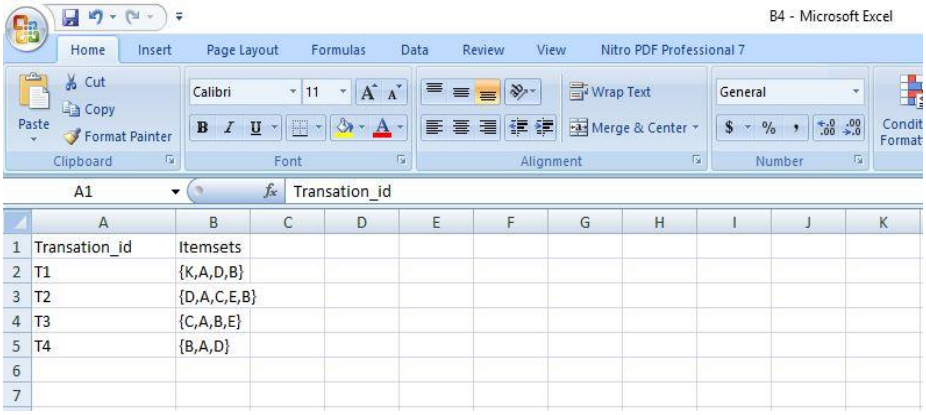
Graph View:



Task-14:
Case study-1:

Aim: Market basket analysis is a modelling technique based upon a theory that if you buy a certain group of items you are more likely to buy another group of items. This technique may allow the retailer to understand the purchase behaviour of a buyer. This information may help the retailer to know the buyer's needs and change the store's layout accordingly. Using differential analysis comparison of results between different stores, between customers in different demographic groups can be done.

Input File:

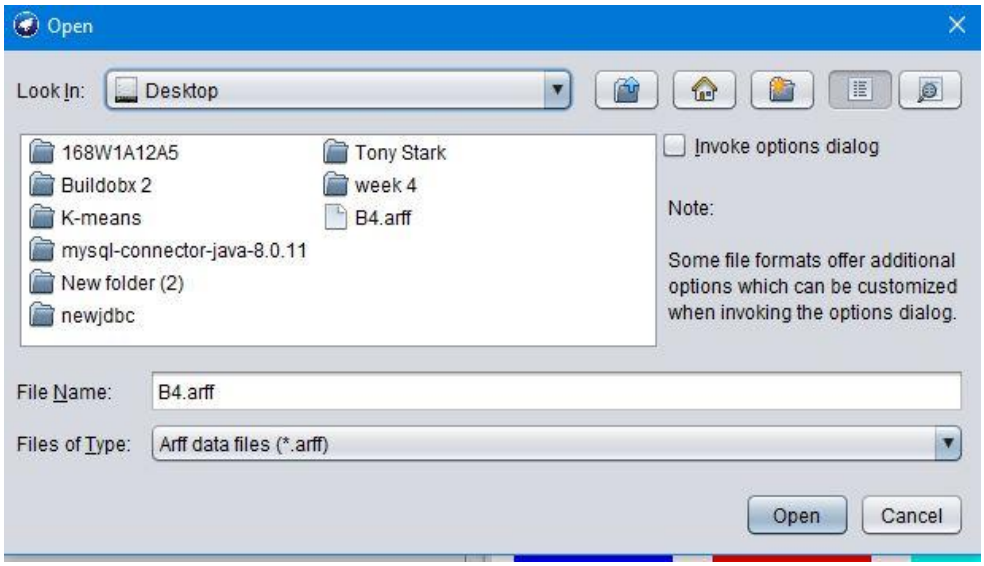


	A	B	C	D	E	F	G	H	I	J	K
1	Transation_id	Itemsets									
2	T1	{K,A,D,B}									
3	T2	{D,A,C,E,B}									
4	T3	{C,A,B,E}									
5	T4	{B,A,D}									
6											
7											

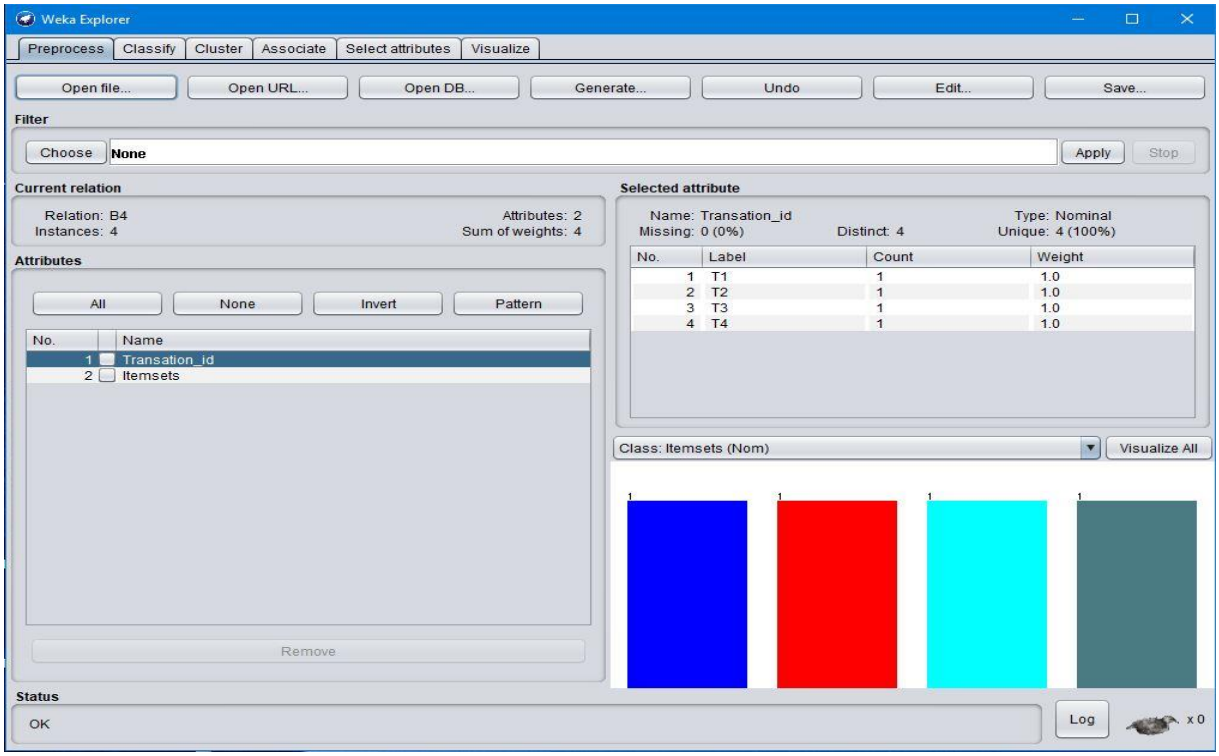
Procedure:

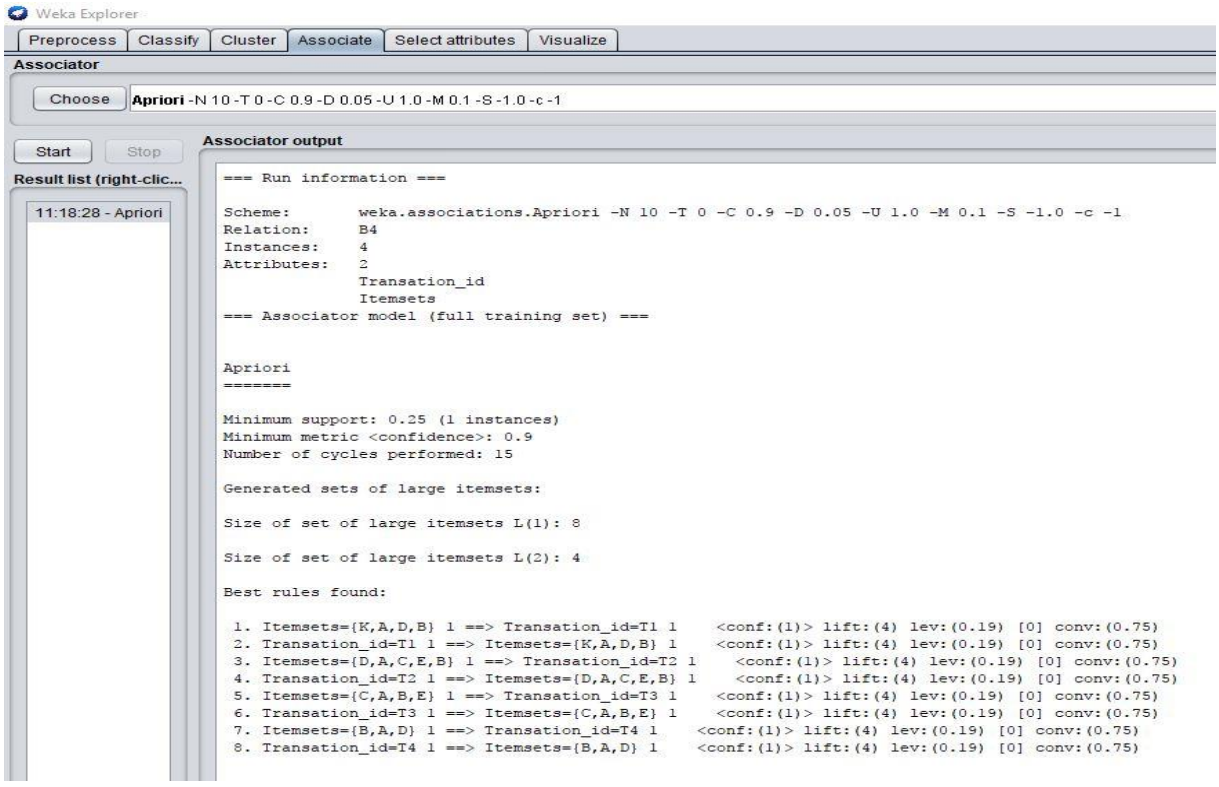
Apriori:

1. Open weka and open the file that we have created and save it with .arff extension.



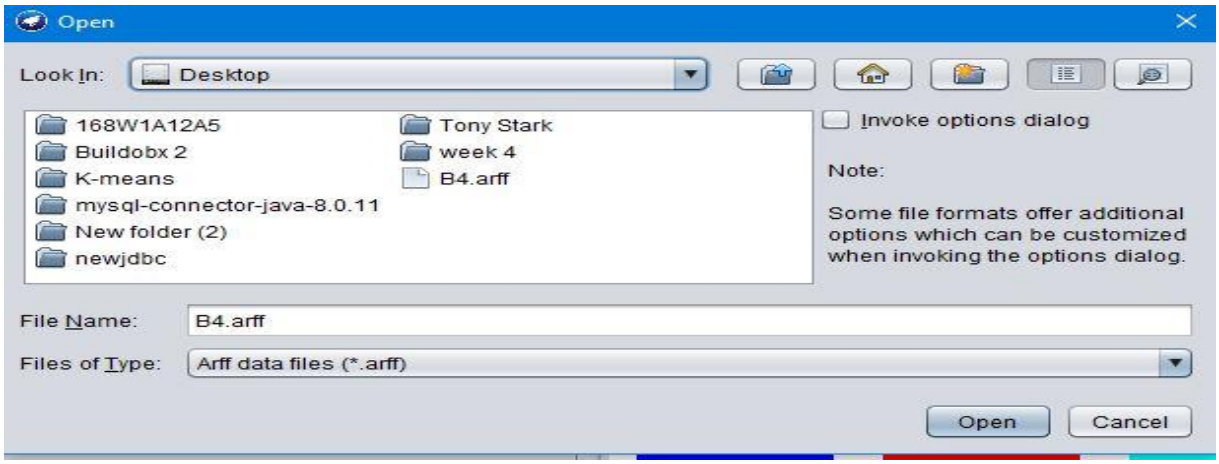
2. Now open the arff file and go to associate and then choose apriori and start the process.

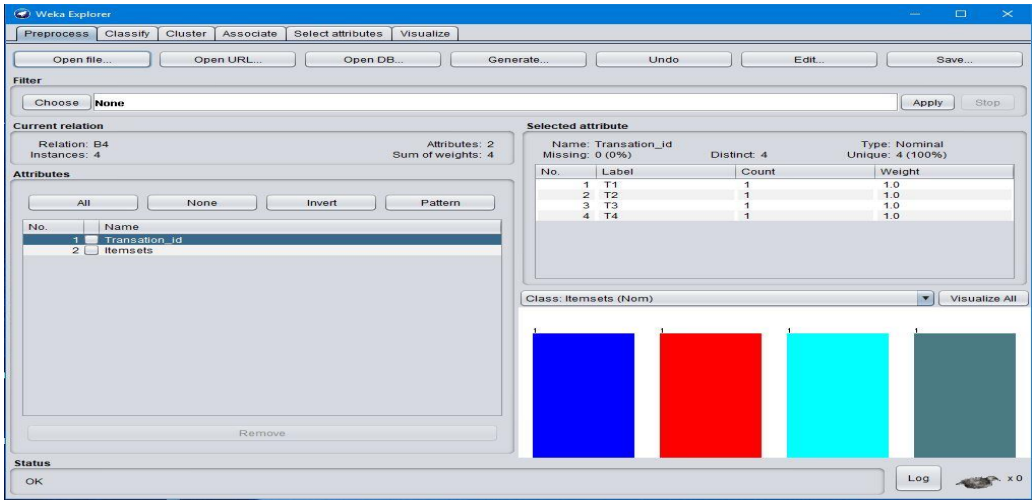

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: B4
Instances: 4
Attributes: 2
Sum of weights: 4
Selected attribute
Name: Transation_id
Missing: 0 (0%)
Distinct: 4
Type: Nominal
Unique: 4 (100%)
No.
Label
Count
Weight
1 T1 1 1.0
2 T2 1 1.0
3 T3 1 1.0
4 T4 1 1.0
Attributes
All
None
Invert
Pattern
No.
Name
1 Transation_id
2 Itemsets
Remove
Class: Itemsets (Nom)
Visualize All
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Start
Stop
Result list (right-clic...
11:18:28 - Apriori
Associator output
=== Run information ===
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: B4
Instances: 4
Attributes: 2
Transation_id
Itemsets
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.25 (1 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 15
Generated sets of large itemsets:
Size of set of large itemsets L(1): 8
Size of set of large itemsets L(2): 4
Best rules found:
1. Itemsets={K,A,D,B} 1 ==> Transation_id=T1 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)
2. Transation_id=T1 1 ==> Itemsets={K,A,D,B} 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)
3. Itemsets={D,A,C,E,B} 1 ==> Transation_id=T2 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)
4. Transation_id=T2 1 ==> Itemsets={D,A,C,E,B} 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)
5. Itemsets={C,A,B,E} 1 ==> Transation_id=T3 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)
6. Transation_id=T3 1 ==> Itemsets={C,A,B,E} 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)
7. Itemsets={B,A,D} 1 ==> Transation_id=T4 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)
8. Transation_id=T4 1 ==> Itemsets={B,A,D} 1 <conf:(1)> lift:(4) lev:(0.19) [0] conv:(0.75)

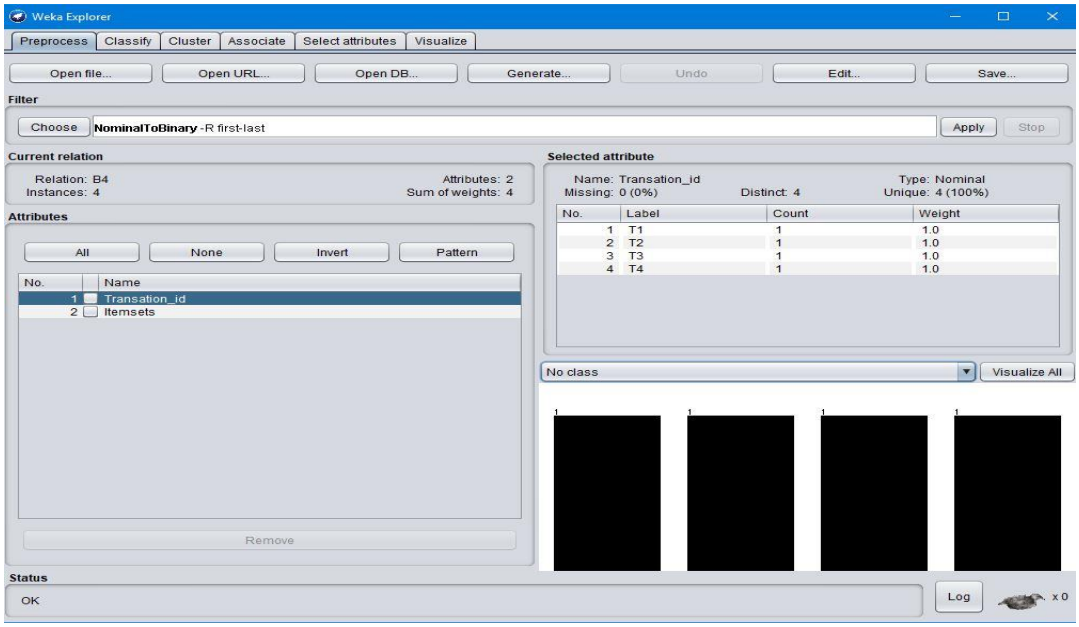
Fp growth:

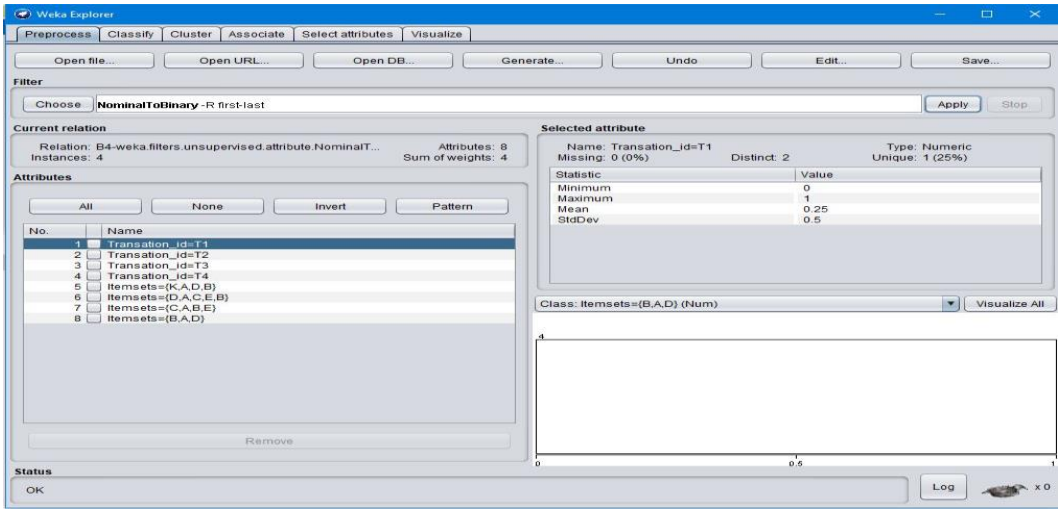
1. Open the arff file and then apply preprocessing techniques for converting all the data values to binary.



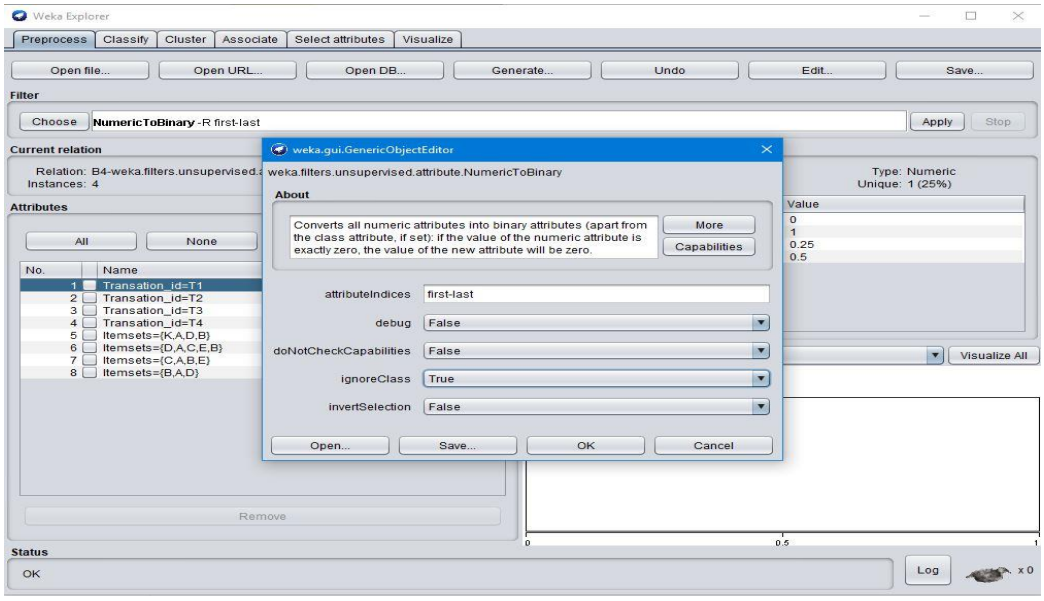


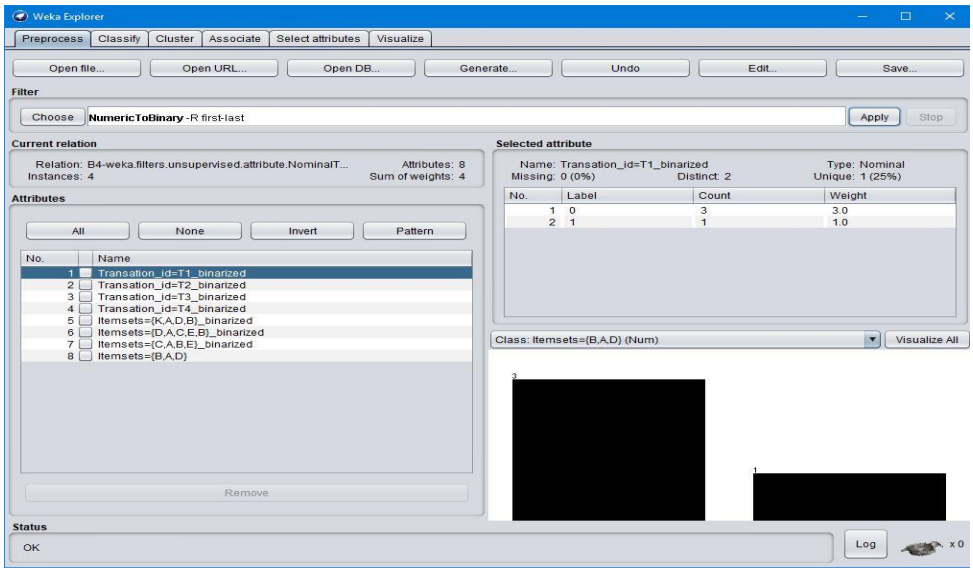
2. So apply “Nominal to Binary” and choose “no class”.

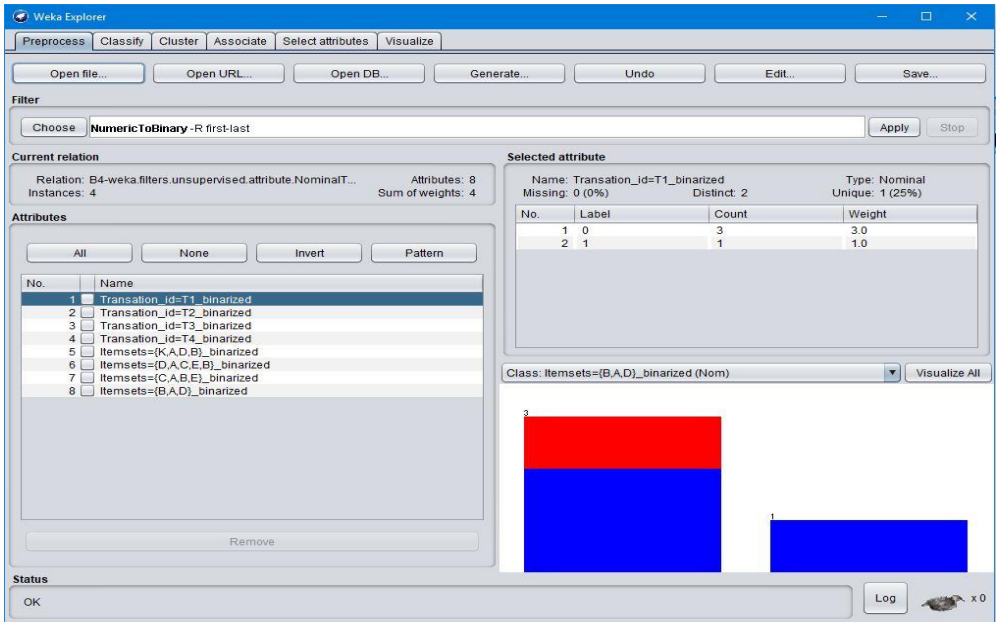




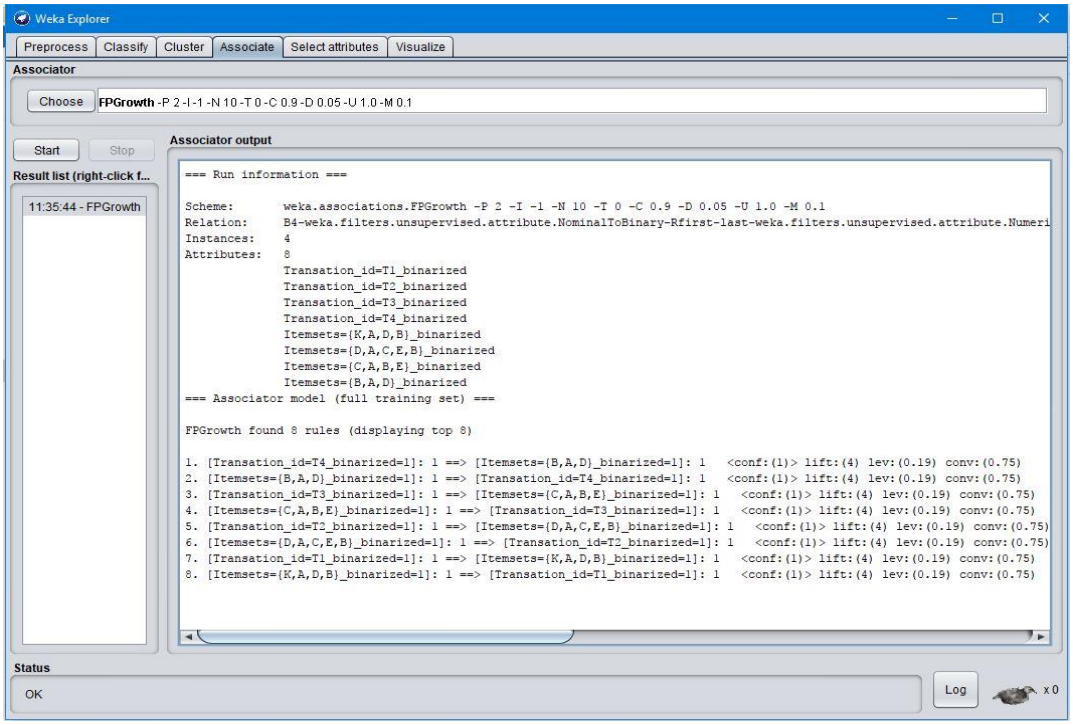
3. Also apply “Numeric to Binary” and then in it select ignore class= “true”.







4. Finally go to associate and then select fp growth and start the process.



Comparision between Apriori and Fp growth:

1. Because of generation of candidates in apriori large memory space is required than fp growth.
2. Multiple scans are required in apriori for generating candidate sets where as in fp growth we might scan only twice.
3. Execution time is more in apriori because of candidate sets which is less in fp growth.
4. Therefore,Fp growth is better than apriori especially in case of larger datasets.