ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

DATA WAREHOUSE AND DATA MINING LAB

ARFF CREATION

By

Nadella Sunil

M.Tech., (PhD) UGC NET & AP SET (Computer Science & Applications)

TS & AP SET (Mathematical Sciences)

Department of Computer Science and Engineering

Aditya College of Engineering & Technology, Surampalem, Andhra Pradesh

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

Agenda

- Procedure to create ARFF File (Attribute relation File Format) explained in this presentation
- As the name suggests it described a list of instances sharing a set of attributes.
- these files are supported by WEKA machine Learning tool,
- The *arff* files are used for the purpose of various operations related to data preprocessing, data cleaning etc.

Structure of ARFF file

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

- ARFF file contains 2 sections
  1. Header Section
  2. Data Section
- All the *keywords in ARFF file start with @* symbol.

Header Section

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

- This section contains various information related to the dataset like the name of the relation, columns, and type of columns. The header section contains 2 parts

1. **Table/relation** and
2. **attribute** part.

Header Section

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

- *@relation :used to give the table name*
- *@attribute: used to give a column name*

DATA TYPES

- ***nominal:*** *represented inside curly brackets (Like constants)*
- ***string :*** *data type which accepts only string value*
- ***numeric:*** *used to store numbers*
- ***date:*** *used to store date*

Header Section – Syntax

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

- @relation tablename
- @attribute column_name type

Header Section –Example

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

@relation "employee"

@attribute f_name string

@attribute l_name string

@attribute contact_num numeric

@attribute dept {HR,IT,MANAGEMENT,MAINTAINANCE}

@attribute DOB date dd-mm-yyyy

@attribute city string

Here dept column is having nominal data type so it can only accept above mentioned types of data only,

Data Section

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

- Data section is used to represents the data or entries for available columns. (according to the order in header section data would be inserted).
- data section starts with @data, and this section must be added after Header section. only single record can be written in single line.
- **@data:** Used to start data section
- %: % sign is used to represent the comment in file.

Data Section - Syntax

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

@data

<record1>

<record2>

•

•

<record N>

all the Records must be in the same format as their attributes are defined in Header section Like

Data Section - Example

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

1,naman,N,1234556678,IT,02-08-2000,rjt

2,yash,M,1234556679,HR,04-05-2001,amd
3,kishan,G,1214556678,MANAGEMENT,02-11-2001,pbr

4,?,?,5234556678,IT,03-05-2000,amd

- ✓We separate values by comma(,) and to
- ✓ represent the empty or missing value for a particular column we use the (?)sign.

Emp.arff

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

```
@relation "employee"
@attribute id numeric
@attribute f_name string
@attribute l_name string
@attribute contact_num numeric
@attribute dept {HR,IT,MANAGEMENT,MAINTAINANCE}
@attribute DOB date dd-mm-yyyy
@attribute city string
@data
1,naman,N,1234556678,IT,02-08-2000,rjt
2,yash,M,1234556679,HR,04-05-2001,amd
3,kishan,G,1214556678,MANAGEMENT,02-11-2001,pbr
4,?,?,5234556678,IT,03-05-2000,amd
```

How to Create and open arff file - Procedure

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

Step 1: Open any text editor and paste the above code.

Step 2: Save the file with emp_dm.arff file extension

Step 3: Open weka tool

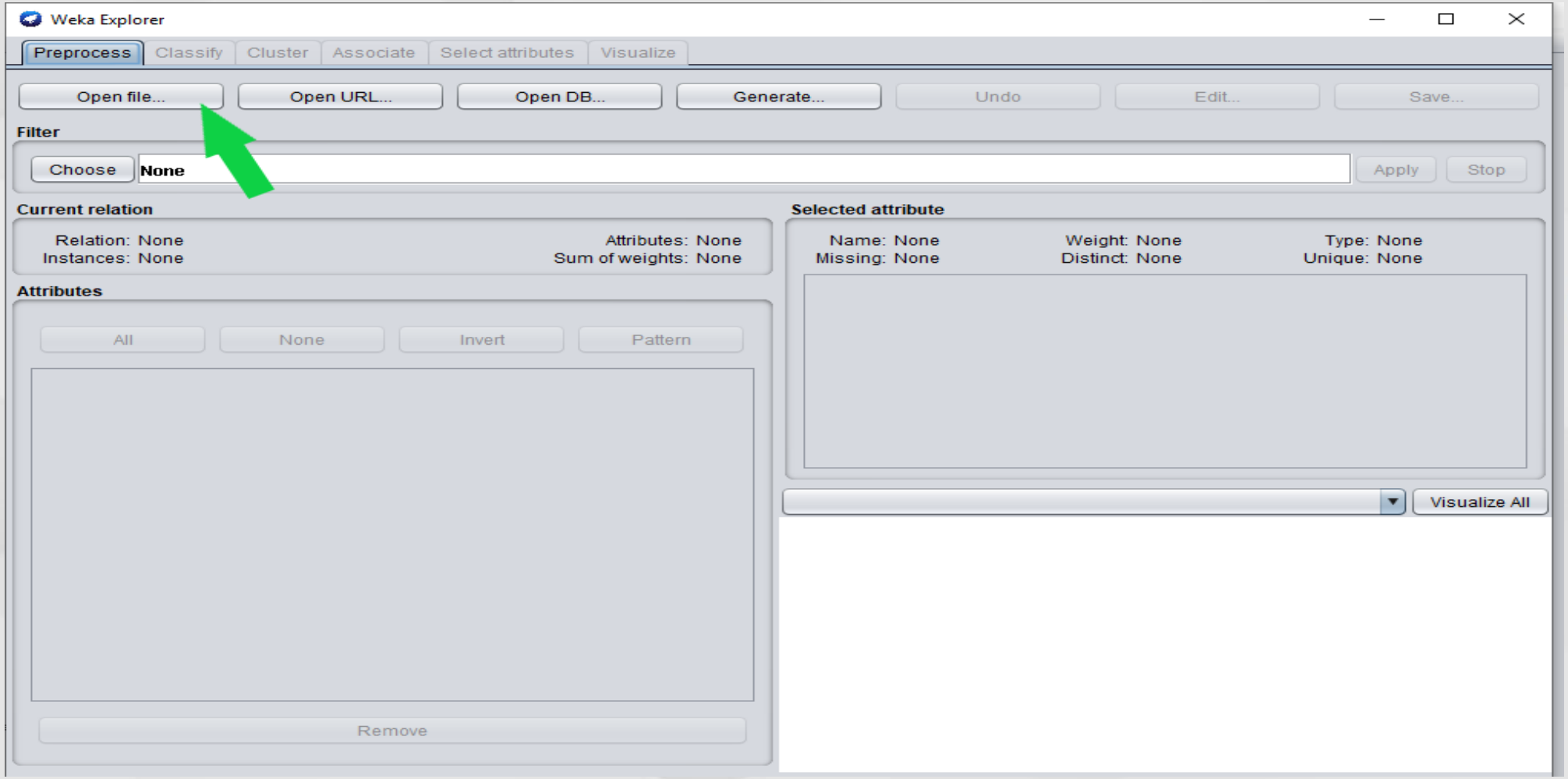
Step 4: Click on Explorer

Open WEKA tool and select Explore

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY

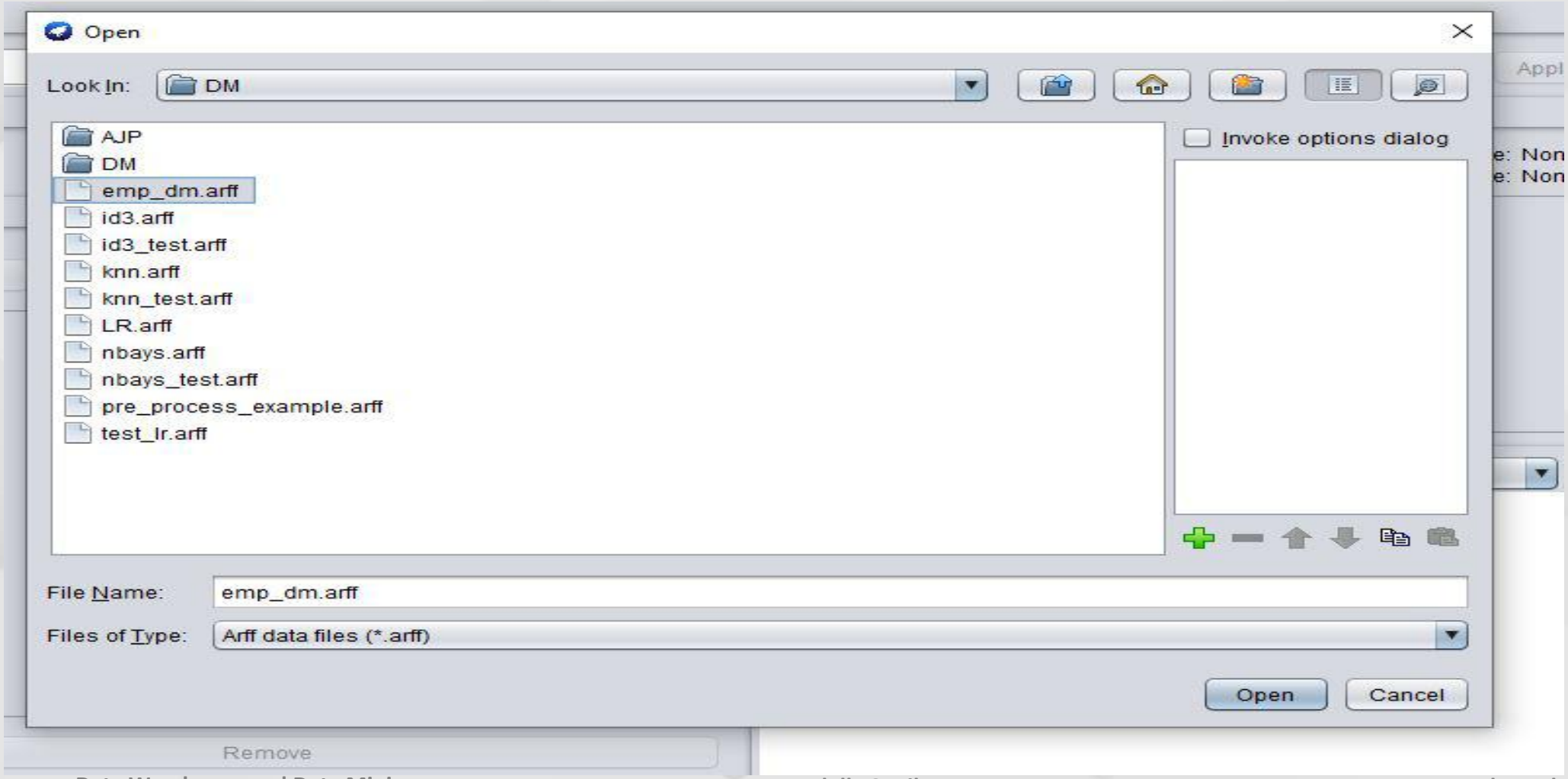
Step – 5 :
Then click on Open file

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY



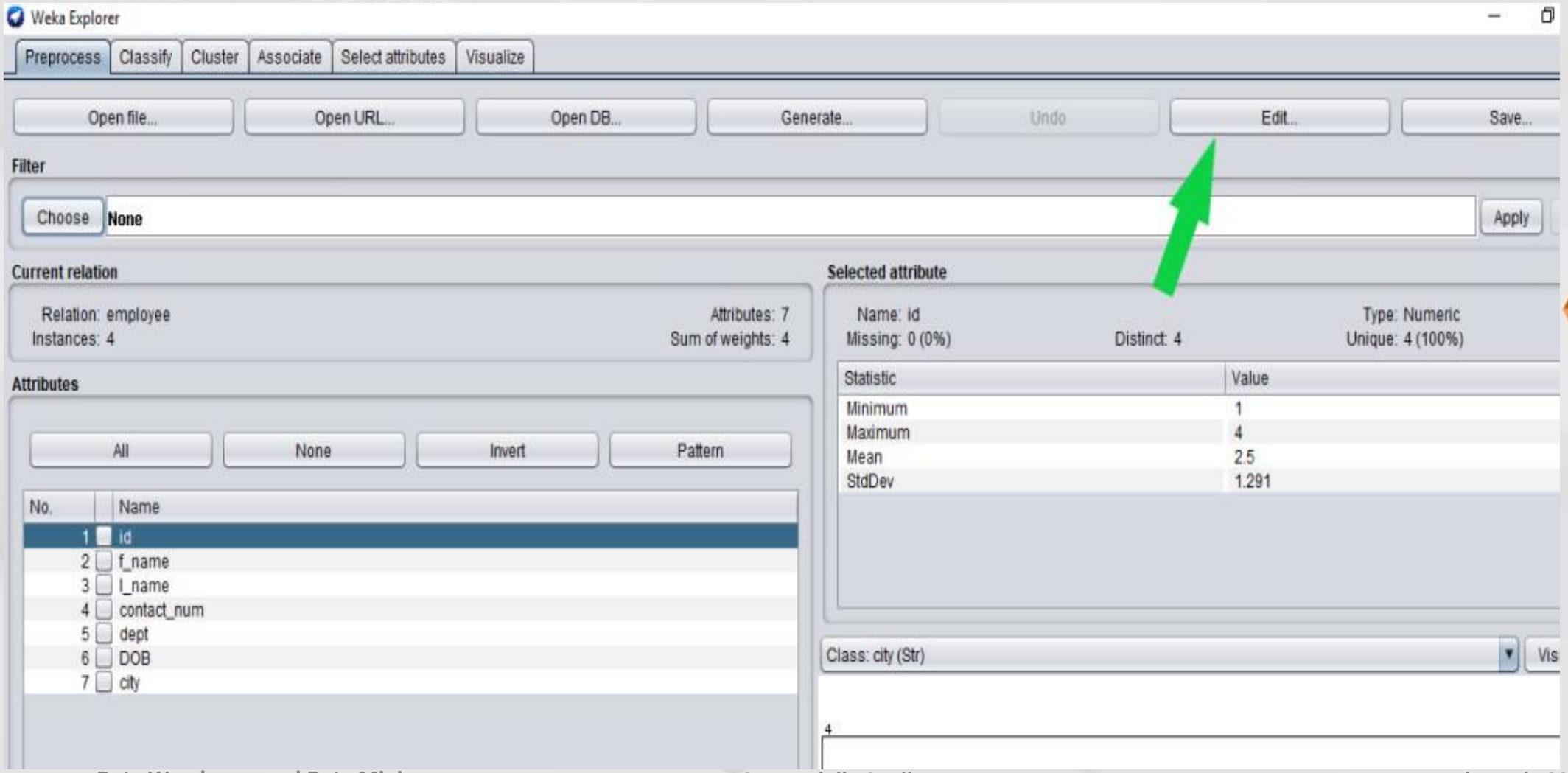
Step – 6 :
Select/Locate arff file from disk then click On Open.

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY



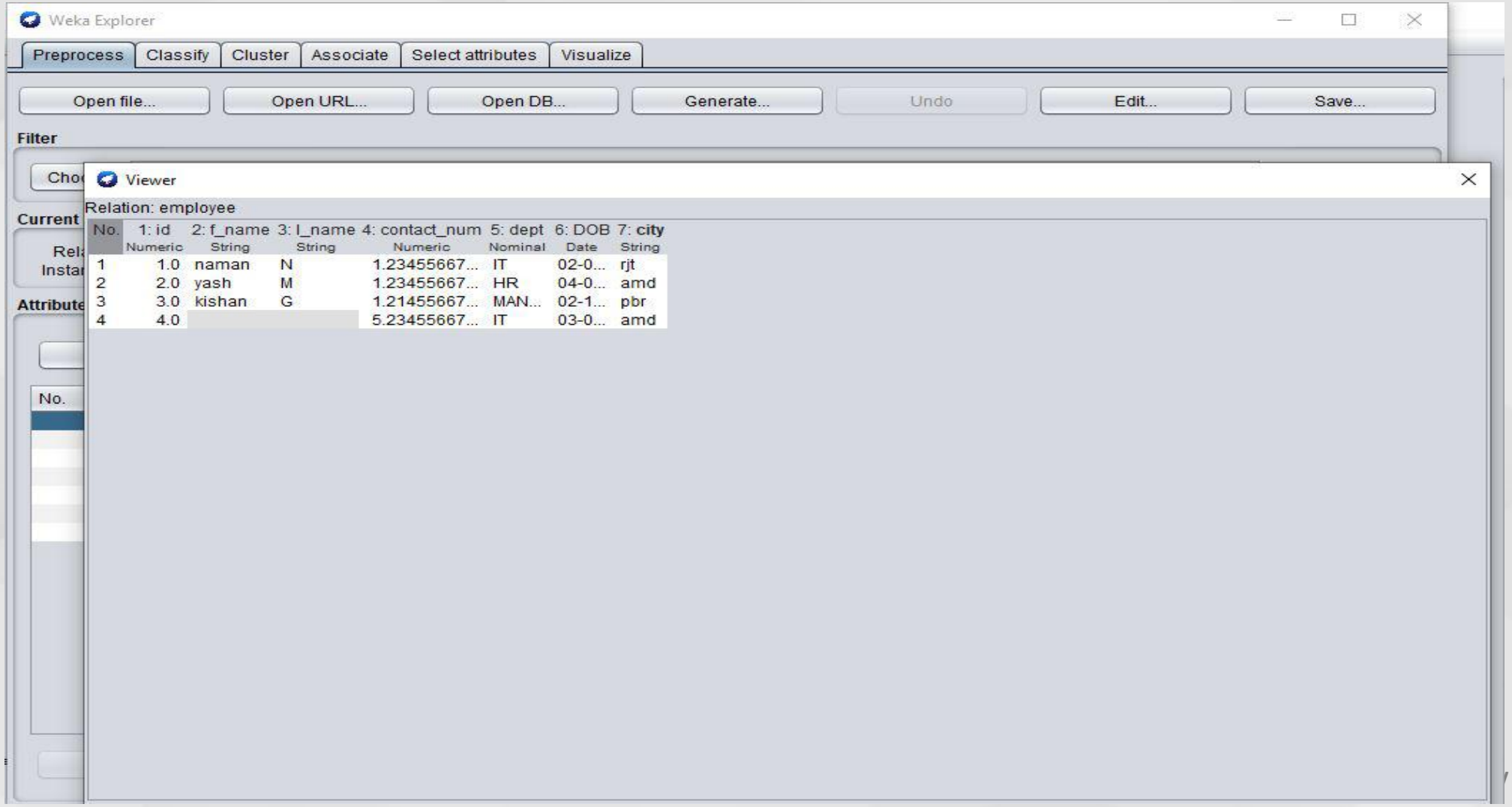
Step – 7 :
file is now Loaded now click on Edit from Preprocess Tab

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY



Step – 8 : viewing the emp.arff dataset

ADITYA
COLLEGE OF ENGINEERING & TECHNOLOGY



No.	1: id Numeric	2: f_name String	3: l_name String	4: contact_num Numeric	5: dept Nominal	6: DOB Date	7: city String
1	1.0	naman	N	1.23455667...	IT	02-0...	rjt
2	2.0	yash	M	1.23455667...	HR	04-0...	amd
3	3.0	kishan	G	1.21455667...	MAN...	02-1...	pbr
4	4.0			5.23455667...	IT	03-0...	amd


Thank You!