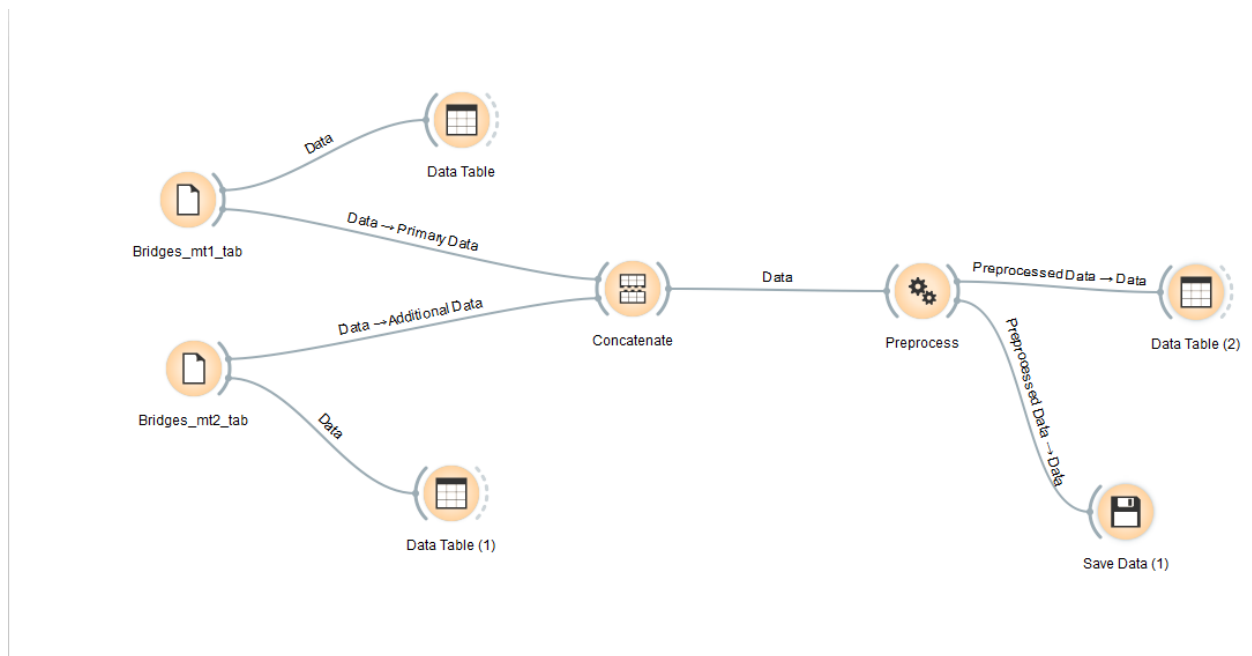


EXPERIMENT NO : 06

Aim: To perform ETL process for building a data warehouse using ORANGE tool.

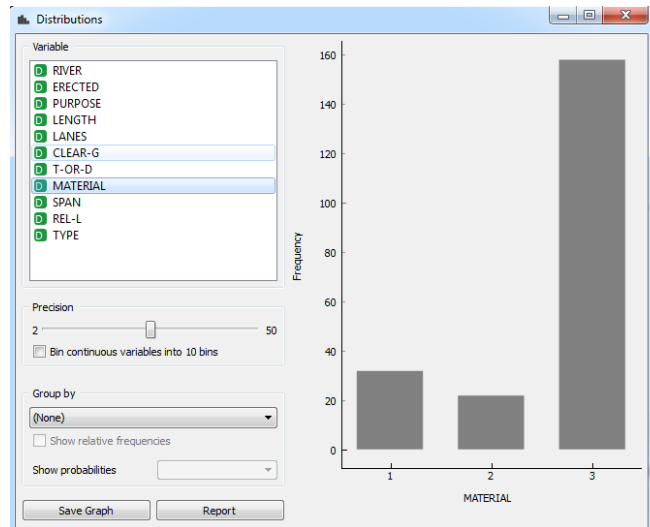
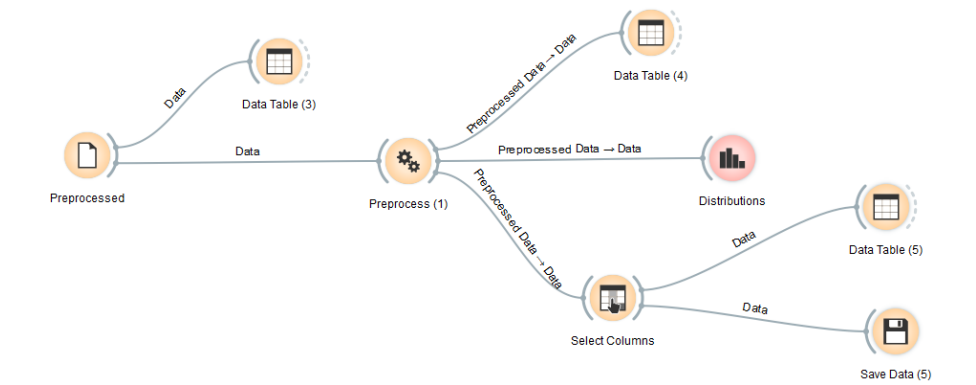
1. Consider bridges.mt1.tab and bridges.mt2.tab files as input and do as directed:

- Concatenate the files
- Apply preprocessing for missing values
- Save the new file
- Open the new file and view it.



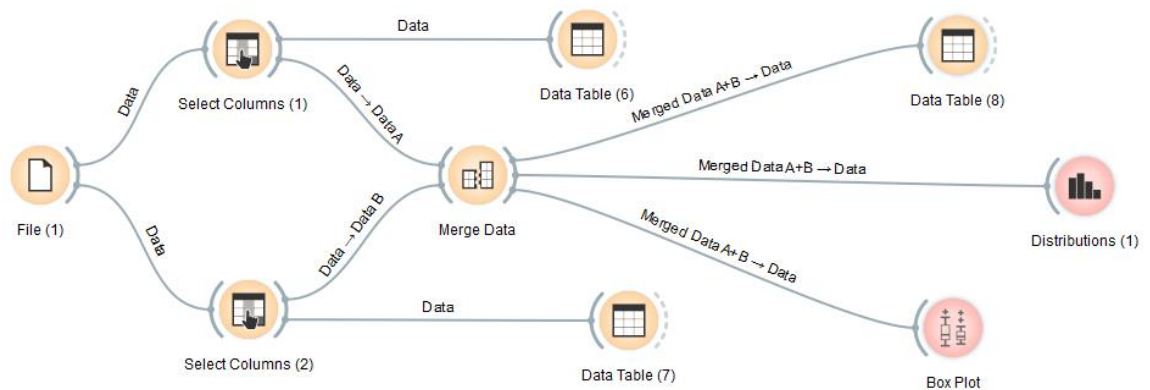
2. Use the above file to

- Discretize the variables and find the type of river that is maximum in number, which type of material is majorly used?
- Consider only erected, type, river, purpose and length columns and save it as file- F1.tab

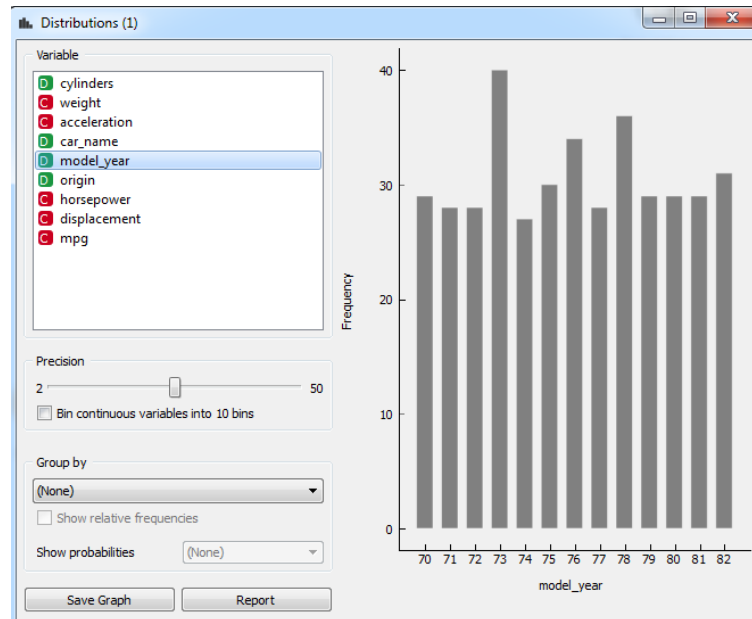


3. Do as directed:

- Import auto-mpg.tab file. Make 2 parts of the file columns.
 1. cylinders, weight, acceleration, model_year and car_name.
 2. mpg, origin, horsepower, displacement and car_name.
- Consider these 2 parts and merge to get the original file.



1. In which year maximum cars were built?



2. Write a python script to display first 3 rows of auto-mpg.

```
import Orange
data= Orange.data.Table("auto-mpg")
print(data.domain)
for d in data[:3]:
    print (d)
```

Output :

Running script:

```
[cylinders, displacement, horsepower, weight, acceleration, model_year, origin, car_name | mpg]
[8, 307.0, 130, 3504, 12.0, 70, 1, chevrolet chevelle malibu | 18.0]
[8, 350.0, 165, 3693, 11.5, 70, 1, buick skylark 320 | 15.0]
[8, 318.0, 150, 3436, 11.0, 70, 1, plymouth satellite | 18.0]
```

3. Write a python script to find the maximum displacement, its index and the corresponding row.

```
import Orange
data= Orange.data.Table("auto-mpg")
print(data.domain)
for d in data[:3]:
```

```

print (d)
p=[1]
for i in range(0,398):
p.append(data[i]['displacement'])

print('max value is:')
print(max(p))
print('index of max value is:')
i=p.index(max(p))
print(i)

data= Orange.data.Table("auto-mpg")
print(data.domain)
for q in data[i-1]:
print (q)

```

Output :

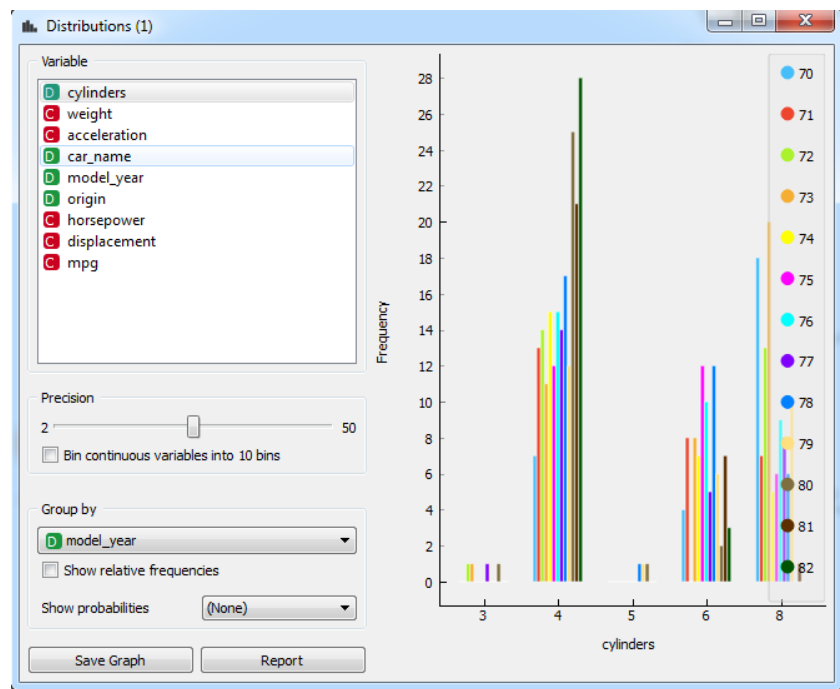
Running script:

```

[cylinders, displacement, horsepower, weight, acceleration, model_year, origin, car_name | mpg]
[8, 307.0, 130, 3504, 12.0, 70, 1, chevrolet chevelle malibu | 18.0]
[8, 350.0, 165, 3693, 11.5, 70, 1, buick skylark 320 | 15.0]
[8, 318.0, 150, 3436, 11.0, 70, 1, plymouth satellite | 18.0]
max value is:
455.0
index of max value is:
9
[cylinders, displacement, horsepower, weight, acceleration, model_year, origin, car_name | mpg]
4.0
455.0
225.0
4425.0
10.0
0.0
0.0
241.0
14.0

```

4. Conclude about the number of cylinders. Add grouping of model_year in box plot to make conclusions.



5. Write a python script to find the minimum displacement, its index and the corresponding row.

```
import Orange

data= Orange.data.Table("auto-mpg")
print(data.domain)
p=[398]
for i in range(0,398):
    p.append(data[i]['displacement'])
print('min value is:')
```

Output :

Running script:

[cylinders, displacement, horsepower, weight, acceleration, model_year, origin, car_name | mpg]

min value is:

68.0

index of min value is:

118

1.0

68.0

49.0

1867.0

19.5

3.0

1.0

124.0

29.0