DATA MINING LAB ASSIGNMENT 1

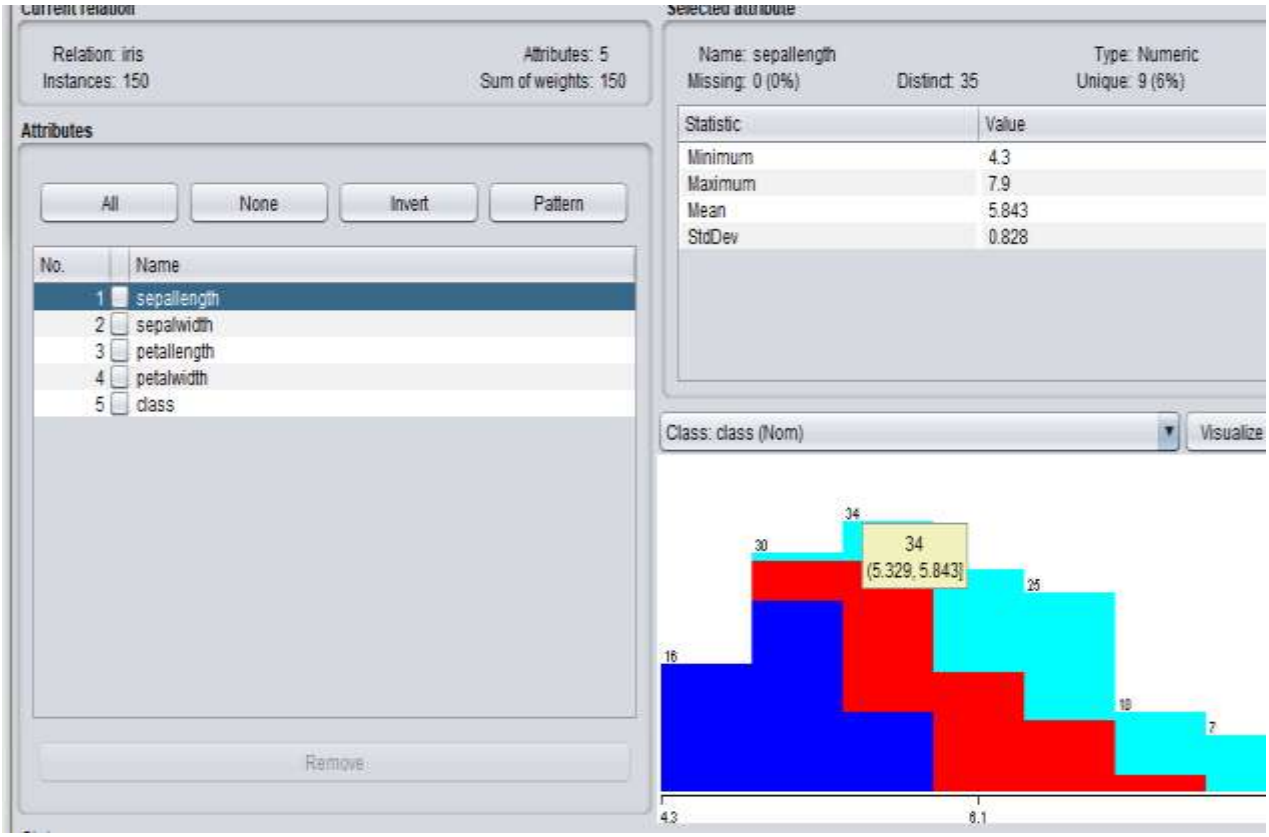
Load each dataset and observe the following:

Observations of Iris dataset:

1.Loading Iris dataset:

i. List the attribute names and its types:

ATTRIBUTE NAME	ATTRIBUTE TYPE
sepallength	Numeric
sepalwidth	Numeric
petallength	Numeric
petalwidth	Numeric
class	Nominal

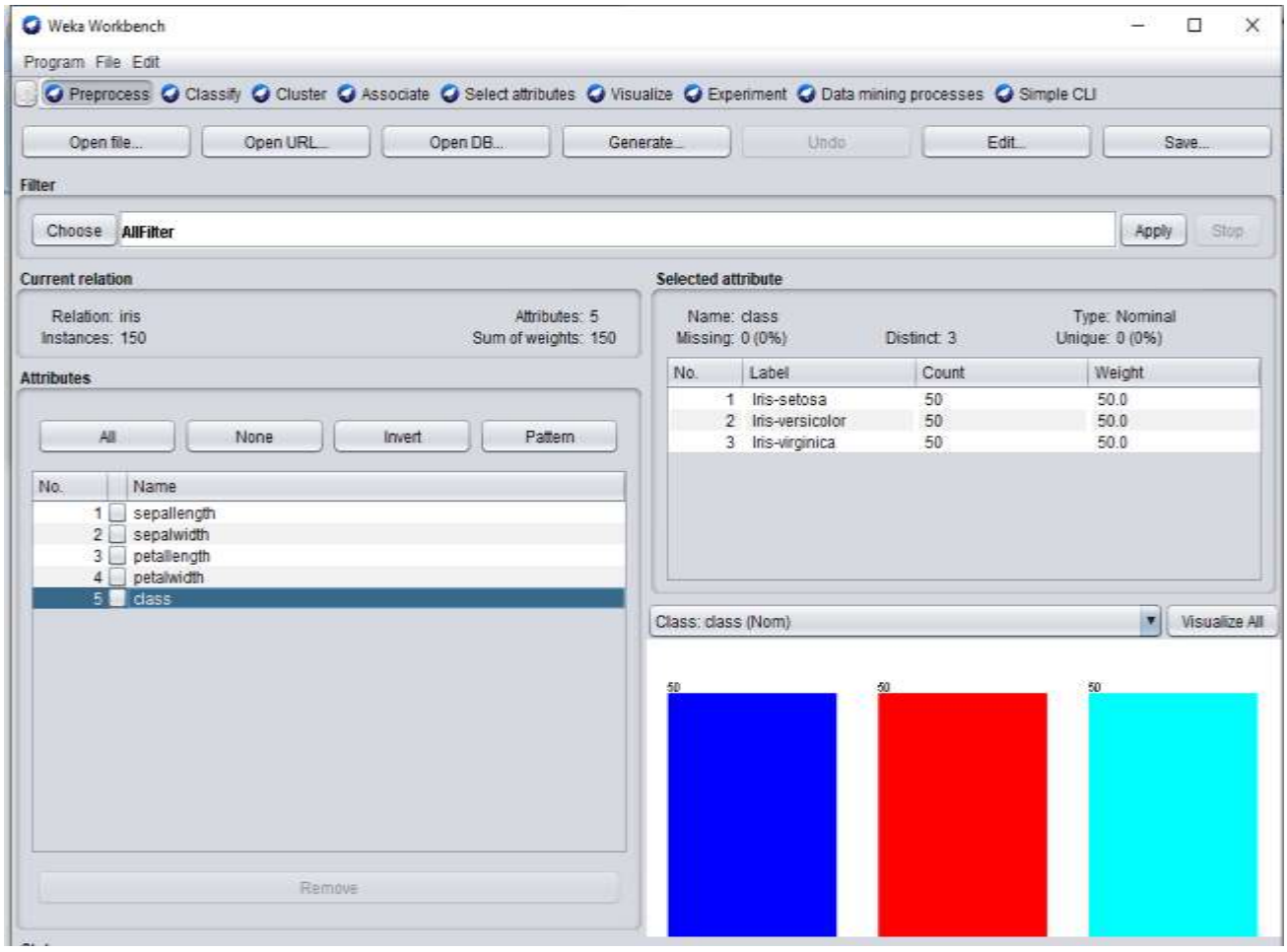


ii. <u>Number of records in each dataset</u>: 150(instances)

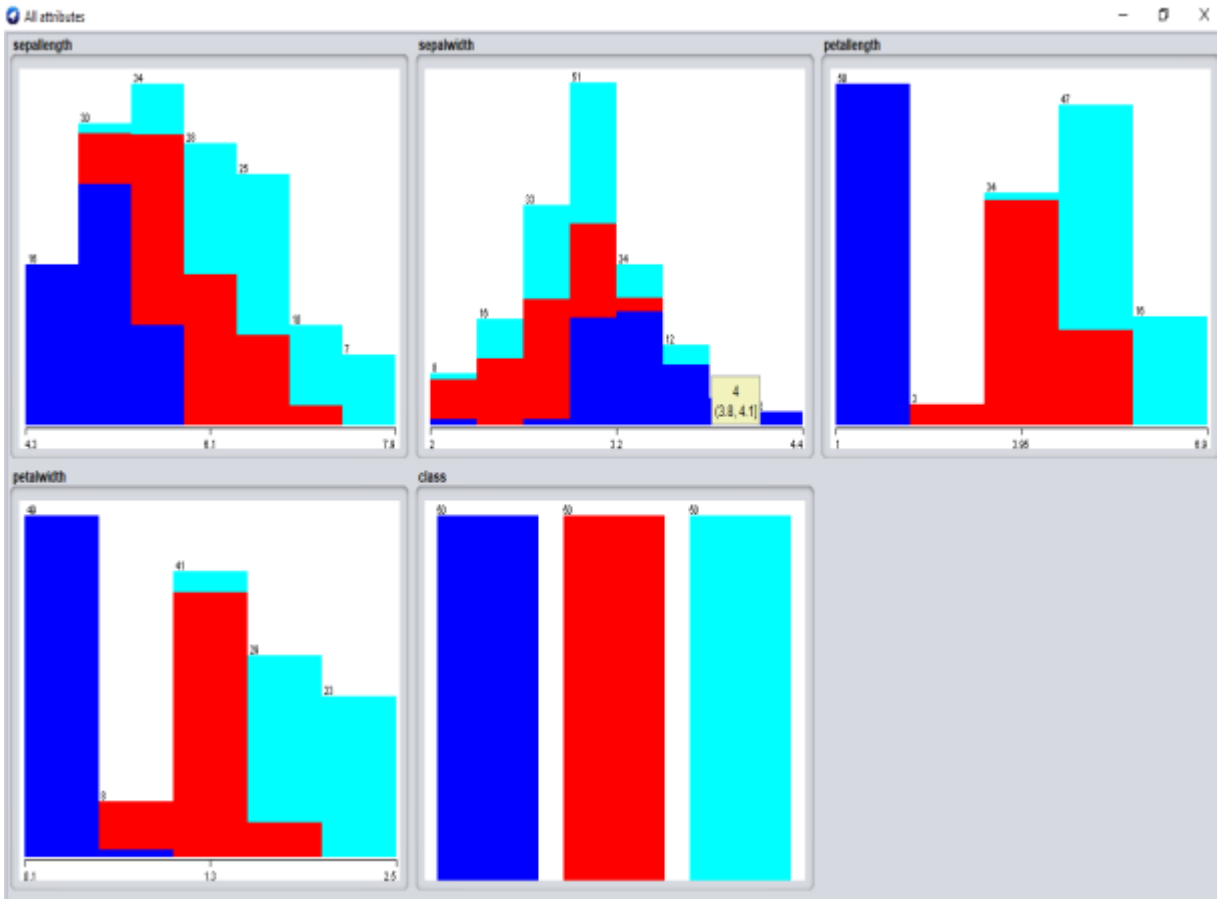
iii. <u>Identify the class attribute</u>(if any):

Only one class attribute is identified in the iris dataset

Class Attribute	Its Type
class	Nominal



iv. Plot Histogram:



V. Determine the number of records in each class:

Class attribute has three labels. They are:

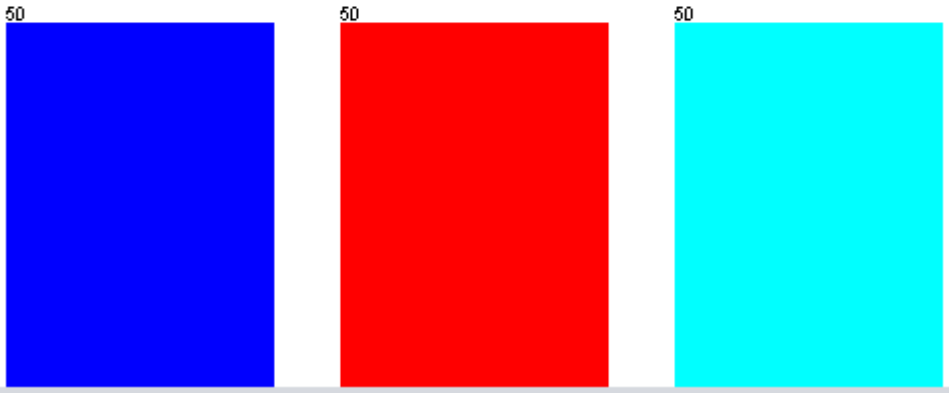
- Iris-setosa – 50 records
- Iris-versicolor – 50 records
- Iris-virginica – 50 records

Selected attribute

Name: class Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	Iris-setosa	50	50.0
2	Iris-versicolor	50	50.0
3	Iris-virginica	50	50.0

Class: class (Nom) Visualize All



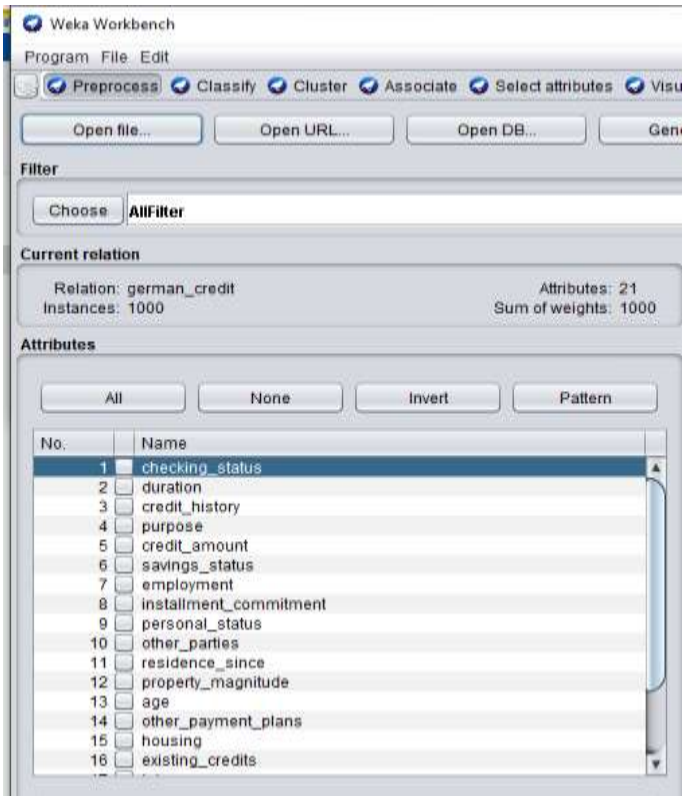
Load each dataset and observe the following:

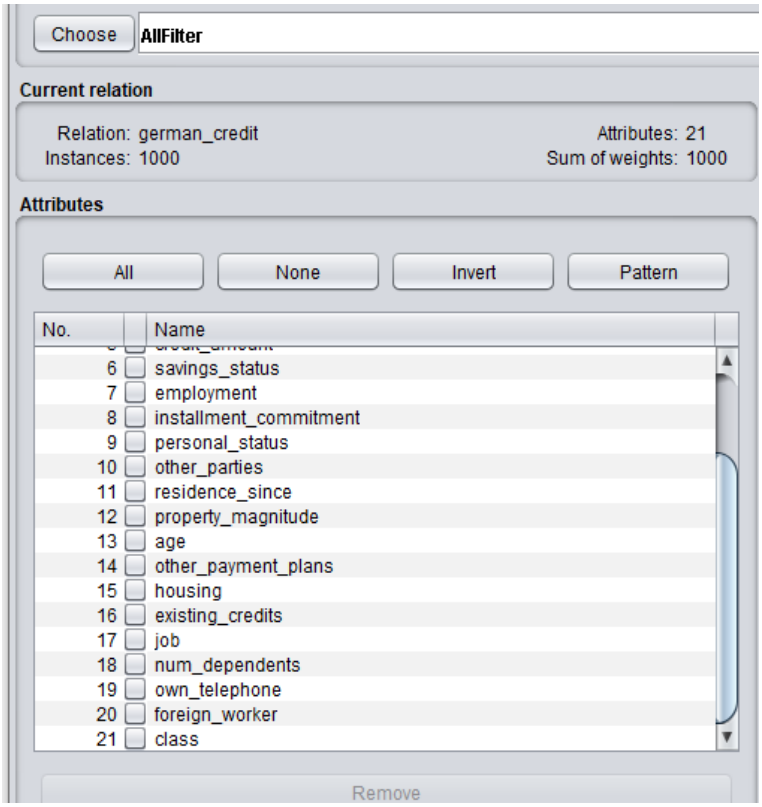
Observations of GermanCredit Dataset:

1. Loading Germancredit dataset:

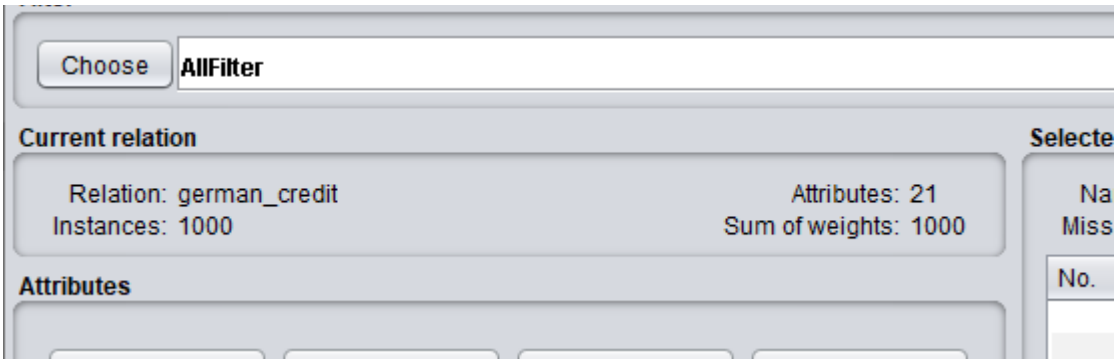
i. List the attribute names and its types:

Attribute name	Its type
checking_status	Nominal
duration	Numeric
credit_history	Nominal
purpose	Nominal
credit_amount	Numeric
savings_status	Nominal
employment	Nominal
installment_commitment	Numeric
personal_status	Nominal
other_parties	Nominal
residence_since	Numeric
property_magnitude	Nominal
age	Numeric
other_payment_plans	Nominal
housing	Nominal
existing_credits	Numeric
job	Nominal
num_dependents	Numeric
own_telephone	Nominal
foreign_worker	Nominal
class	Nominal


Weka Workbench
Program File Edit
Preprocess Classify Cluster Associate Select attributes
Open file... Open URL... Open DB...
Filter
Choose AllFilter
Current relation
Relation: german_credit
Instances: 1000
Attributes: 21
Sum of weights: 1000
Attributes
All None Invert Pattern
No. Name
1 checking_status
2 duration
3 credit_history
4 purpose
5 credit_amount
6 savings_status
7 employment
8 installment_commitment
9 personal_status
10 other_parties
11 residence_since
12 property_magnitude
13 age
14 other_payment_plans
15 housing
16 existing_credits


Choose AllFilter
Current relation
Relation: german_credit
Instances: 1000
Attributes: 21
Sum of weights: 1000
Attributes
All None Invert Pattern
No. Name
6 savings_status
7 employment
8 installment_commitment
9 personal_status
10 other_parties
11 residence_since
12 property_magnitude
13 age
14 other_payment_plans
15 housing
16 existing_credits
17 job
18 num_dependents
19 own_telephone
20 foreign_worker
21 class
Remove

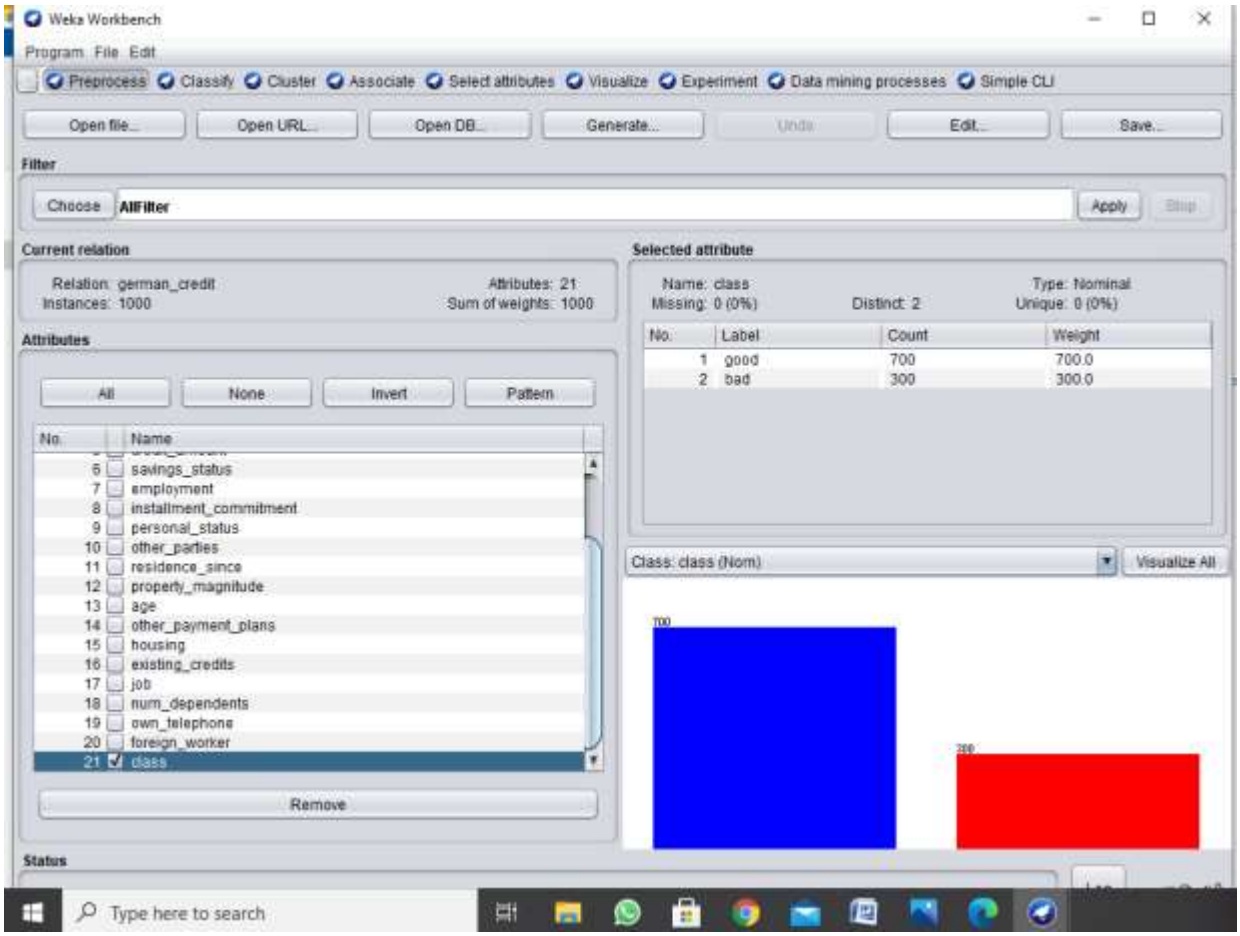
ii. <u>Number of records in each dataset</u>: 1000(instances)



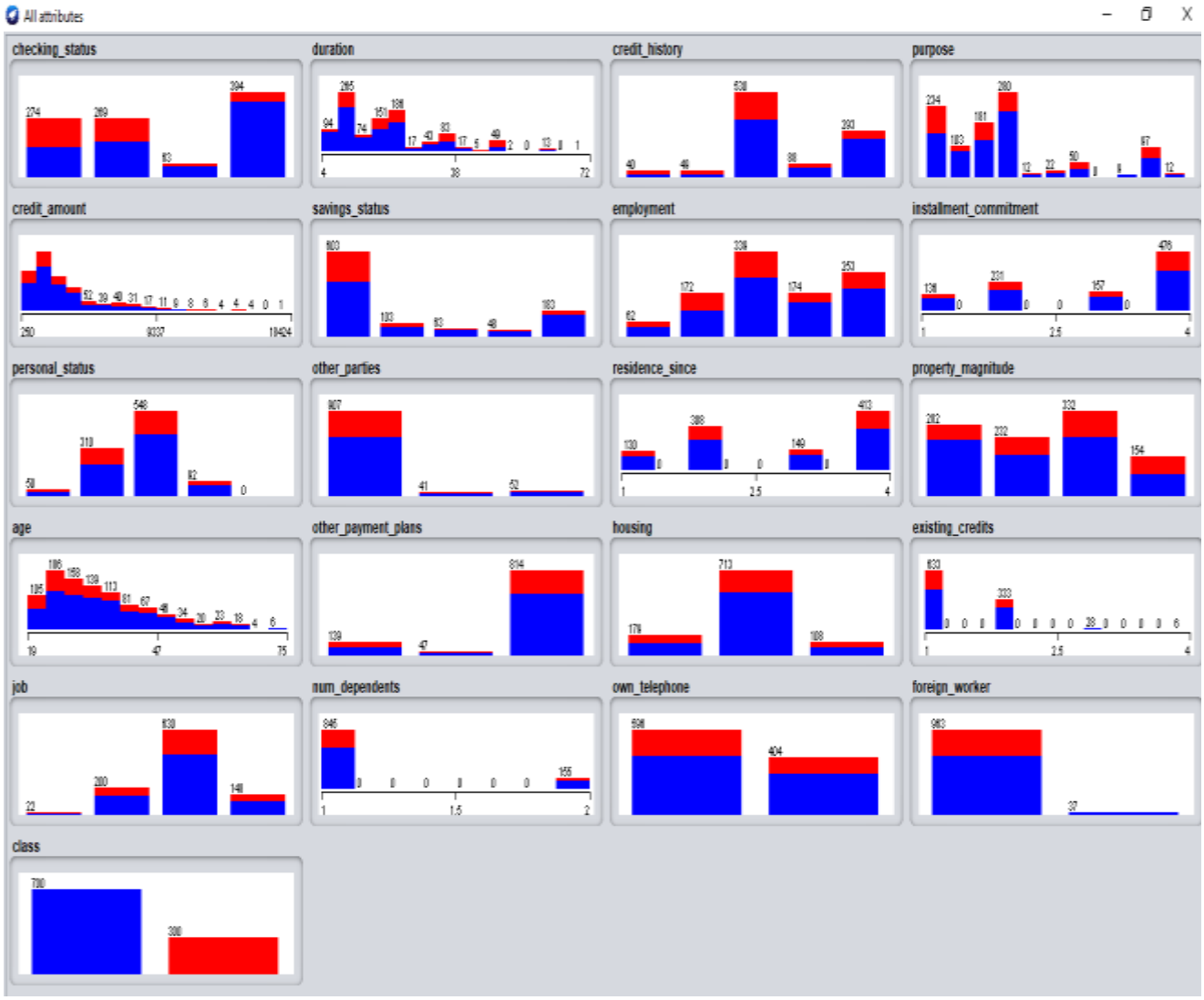
iii. <u>Identify the class attribute</u>(if any):

<u>The class attributes identified are</u>:

class - Nominal



iv. Plot Histogram:



V. Determine the number of records in each class:

Number of records for class:

i. good - 700

ii. bad - 300


lize
Experiment
Data mining processes
Simple CLI
rate...
Undo
Edit...
Save...
Apply
Stop
Selected attribute
Name: class
Type: Nominal
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)
No.
Label
Count
Weight
1
good
700
700.0
2
bad
300
300.0
Class: class (Nom)
Visualize All
700
300