DMWA Lab

WEEK-4

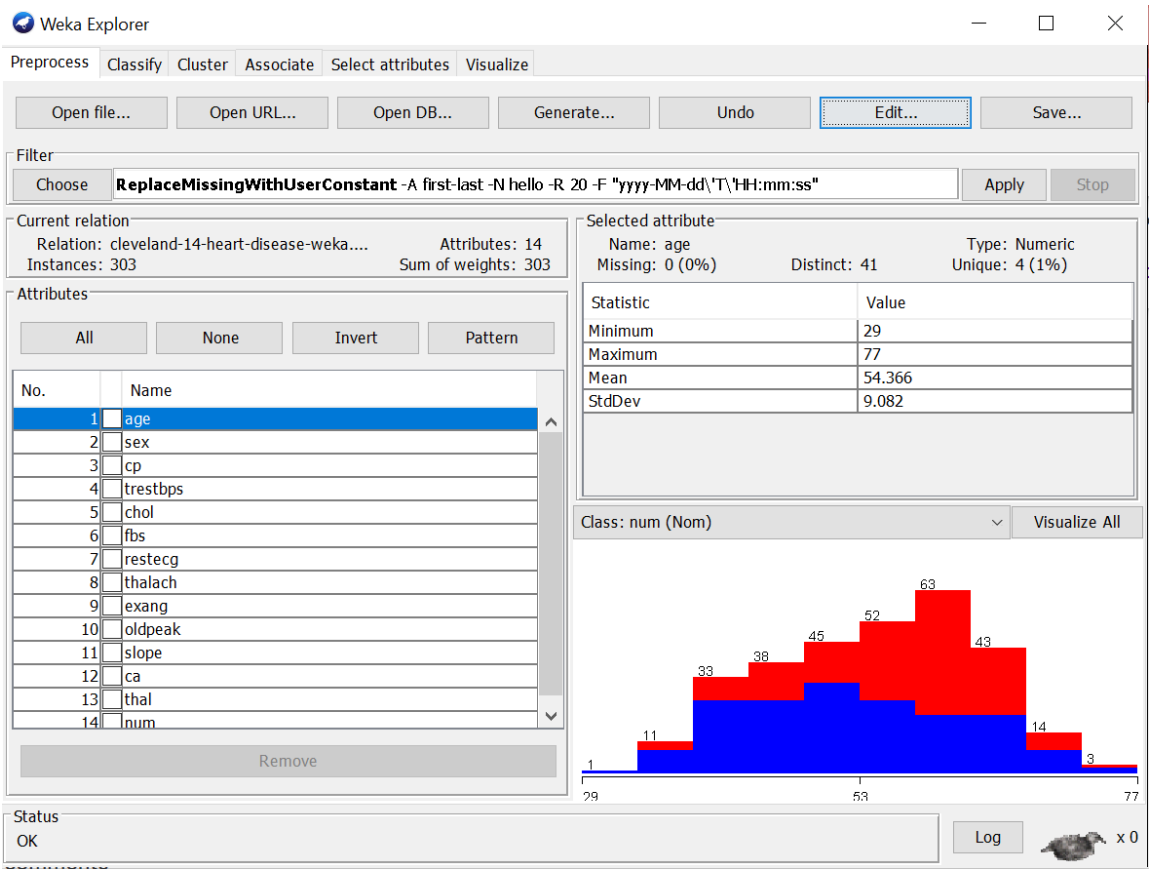
Teghdeep Kapoor

18104050

B12

Qs-1

a) ReplaceMissingWithUserConstant



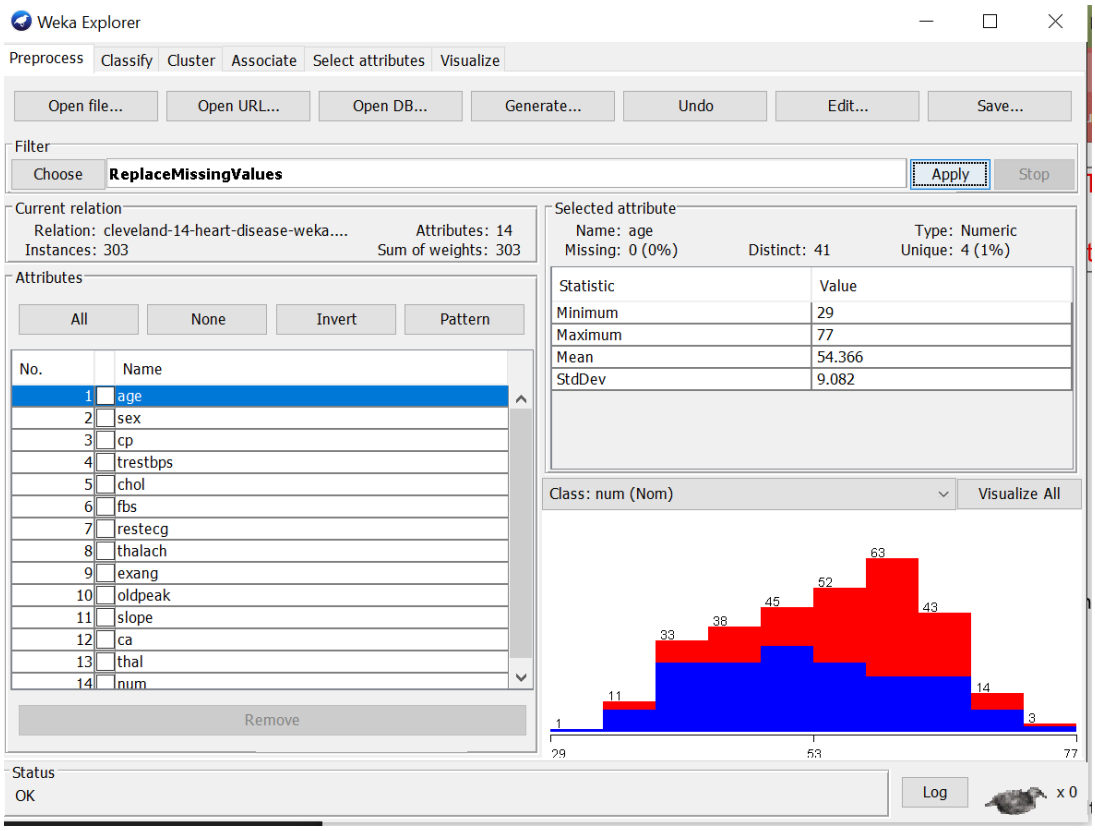
Viewer

Relation: cleveland-14-heart-disease-weka.filters.unsupervised.attribute.ReplaceMissingWithUserConstant-Afirst-last-Nhello-R20-Fyyyy-MM-dd'T'HH:mm:ss

No.	1: age Numeric	2: sex Nominal	3: cp Nominal	4: trestbps Numeric	5: chol Numeric	6: fbs Nominal	7: restecg Nominal	8: thalach Numeric	9: exang Nominal	10: oldpeak Numeric	11: slope Nominal	12: ca Numeric	13: thal Nominal	14: **num** Nominal
1	63.0	male	typ_an...	145.0	233.0	t	left_vent...	150.0	no	2.3	down	0.0	fixed_d...	(50
2	67.0	male	asympt	160.0	286.0	f	left_vent...	108.0	yes	1.5	flat	3.0	normal	)50_1
3	67.0	male	asympt	120.0	229.0	f	left_vent...	129.0	yes	2.6	flat	2.0	revers...	)50_1
4	37.0	male	non_a...	130.0	250.0	f	normal	187.0	no	3.5	down	0.0	normal	(50
5	41.0	female	atyp_a...	130.0	204.0	f	left_vent...	172.0	no	1.4	up	0.0	normal	(50
6	56.0	male	atyp_a...	120.0	236.0	f	normal	178.0	no	0.8	up	0.0	normal	(50
7	62.0	female	asympt	140.0	268.0	f	left_vent...	160.0	no	3.6	down	2.0	normal	)50_1
8	57.0	female	asympt	120.0	354.0	f	normal	163.0	yes	0.6	up	0.0	normal	(50
9	63.0	male	asympt	130.0	254.0	f	left_vent...	147.0	no	1.4	flat	1.0	revers...	)50_1
10	53.0	male	asympt	140.0	203.0	t	left_vent...	155.0	yes	3.1	down	0.0	revers...	)50_1
11	57.0	male	asympt	140.0	192.0	f	normal	148.0	no	0.4	flat	0.0	fixed_d...	(50
12	56.0	female	atyp_a...	140.0	294.0	f	left_vent...	153.0	no	1.3	flat	0.0	normal	(50
13	56.0	male	non_a...	130.0	256.0	t	left_vent...	142.0	yes	0.6	flat	1.0	fixed_d...	)50_1
14	44.0	male	atyp_a...	120.0	263.0	f	normal	173.0	no	0.0	up	0.0	revers...	(50
15	52.0	male	non_a...	172.0	199.0	t	normal	162.0	no	0.5	up	0.0	revers...	(50
16	57.0	male	non_a...	150.0	168.0	f	normal	174.0	no	1.6	up			
17	48.0	male	atyp_a...	110.0	229.0	f	normal	168.0	no	1.0	down			
18	54.0	male	asympt	140.0	239.0	f	normal	160.0	no	1.2	up	0.0	normal	(50
19	48.0	female	non_a...	130.0	275.0	f	normal	139.0	no	0.2	up	0.0	normal	(50
20	49.0	male	atyp_a...	130.0	266.0	f	normal	171.0	no	0.6	up	0.0	normal	(50
21	64.0	male	typ_an...	110.0	211.0	f	left_vent...	144.0	yes	1.8	flat	0.0	normal	(50
22	58.0	female	typ_an...	150.0	283.0	t	left_vent...	162.0	no	1.0	up	0.0	normal	(50
23	58.0	male	atyp_a...	120.0	284.0	f	left_vent...	160.0	no	1.8	flat	0.0	normal	)50_1
24	58.0	male	non_a...	132.0	224.0	f	left_vent...	173.0	no	3.2	up	2.0	revers...	)50_1
25	60.0	male	asympt	130.0	206.0	f	left_vent...	132.0	yes	2.4	flat	2.0	revers...	)50_1
26	50.0	female	non_a...	120.0	219.0	f	normal	158.0	no	1.6	flat	0.0	normal	(50
27	58.0	female	non_a...	120.0	340.0	f	normal	172.0	no	0.0	up	0.0	normal	(50
28	66.0	female	typ_an...	150.0	226.0	f	normal	114.0	no	2.6	down	0.0	normal	(50
29	43.0	male	asympt	150.0	247.0	f	normal	171.0	no	1.5	up	0.0	normal	(50
30	40.0	male	asympt	110.0	167.0	f	left_vent...	114.0	yes	2.0	flat	0.0	revers...	)50_1

Right click (or left+alt) f

b) ReplaceMissingValues



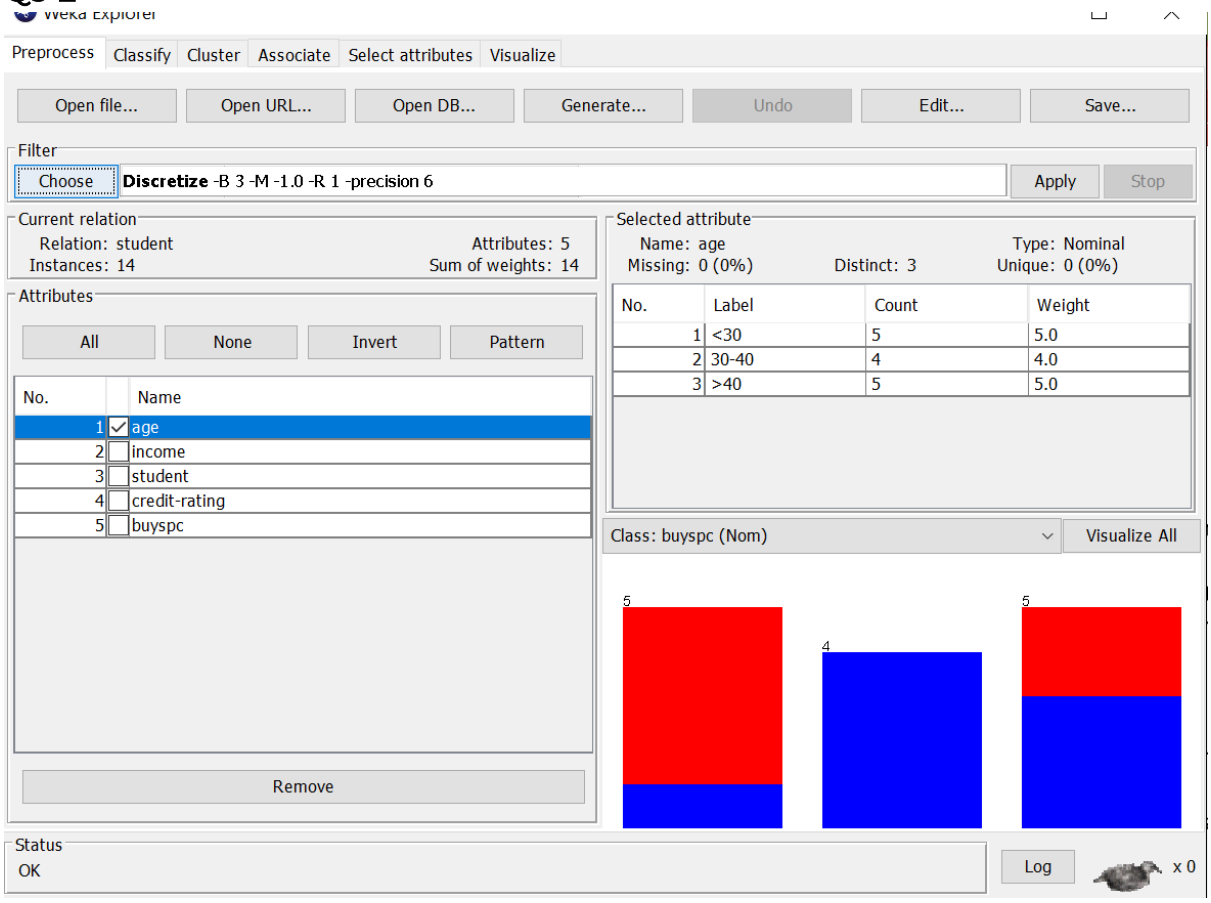
c) RemoveMissingValues

elation: cleveland-14-heart-disease-weka.filters.unsupervised.attribute.NumericCleaner-min1.0E-9-min-defaultNaN-max1.7976931348623157E308-max-default1.7976

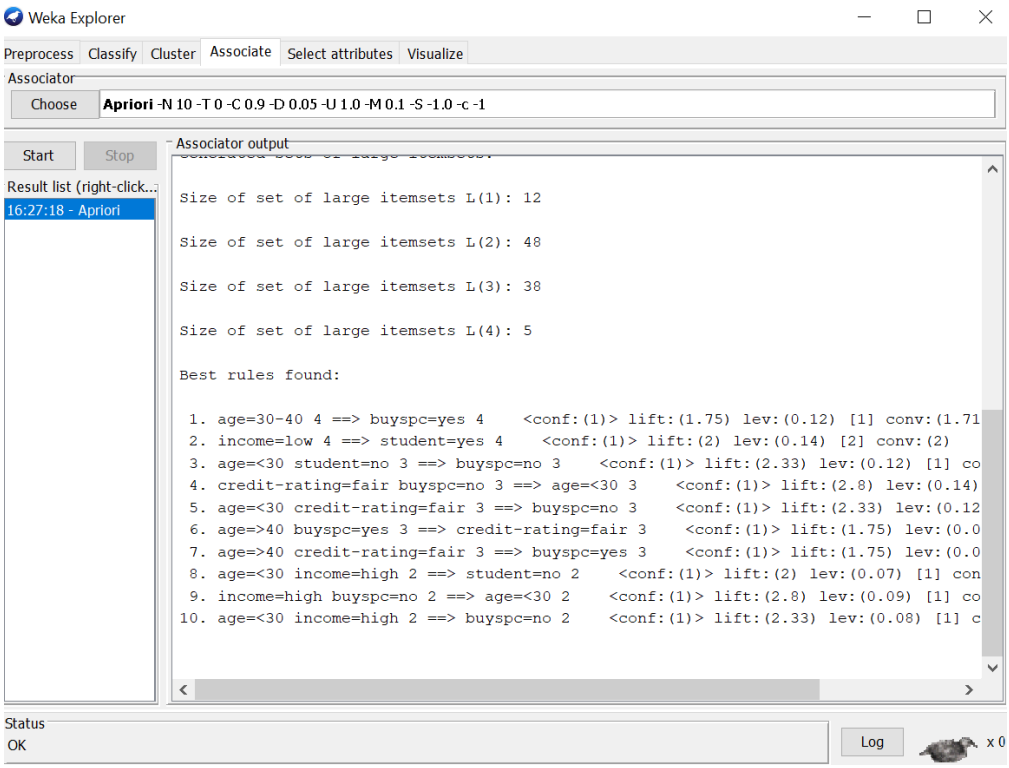
No.	1: age Numeric	2: sex Nominal	3: cp Nominal	4: trestbps Numeric	5: chol Numeric	6: fbs Nominal	7: restecg Nominal	8: thalach Numeric	9: exang Nominal	10: oldpeak Numeric	11: slope Nominal	12: ca Numeric	13: thal Nominal	14: **num** Nominal
4	49.0	male	non_a...	118.0	149.0	f	left_vent...	126.0	no	0.8	up	3.0	normal	)50_1
5	74.0	female	atyp_a...	120.0	269.0	f	left_vent...	121.0	yes	0.2	up	1.0	normal	(50
6	54.0	female	non_a...	160.0	201.0	f	normal	163.0	no	0.0	up	1.0	normal	(50
7	54.0	male	asympt	122.0	286.0	f	left_vent...	116.0	yes	3.2	flat	2.0	normal	)50_1
8	61.0	male	typ_an...	134.0	234.0	f	normal	145.0	no	2.6	flat	2.0	normal	)50_1
9	58.0	male	asympt	100.0	234.0	f	normal	156.0	no	0.1	up	1.0	revers...	)50_1
00	47.0	male	asympt	110.0	275.0	f	left_vent...	118.0	yes	1.0	flat	1.0	normal	)50_1
01	52.0	male	asympt	125.0	212.0	f	normal	168.0	no	1.0	up	2.0	revers...	)50_1
02	58.0	male	asympt	146.0	218.0	f	normal	105.0	no	2.0	flat	1.0	revers...	)50_1
03	64.0	male	asympt	128.0	263.0	f	normal	105.0	yes	0.2	flat	1.0	revers...	(50
04	67.0	female	asympt	106.0	223.0	f	normal	142.0	no	0.3	up	2.0	normal	(50
05	44.0	female	non_a...	118.0	242.0	f	normal	149.0	no	0.3	flat	1.0	normal	(50
06	58.0	female	atyp_a...	136.0	319.0	t	left_vent...	152.0	no	0.0	up	2.0	normal	)50_1
07	61.0	male	asympt	138.0	166.0	f	left_vent...	125.0	yes	3.6	flat	1.0	normal	)50_1
08	59.0	male	non_a...	126.0	218.0	t	normal	134.0	no	2.2	flat	1.0	fixed_d...	)50_1
09	61.0	male	asympt	140.0	207.0	f	left_vent...	138.0	yes	1.9	up	1.0	revers...	)50_1
10	46.0	male	asympt	140.0	311.0	f	normal	120.0	yes	1.8	flat	2.0	revers...	)50_1
11	59.0	male	typ_an...	134.0	204.0	f	normal	162.0	no	0.8	up	2.0	normal	)50_1
12	66.0	female	non_a...	146.0	278.0	f	left_vent...	152.0	no	0.0	flat	1.0	normal	(50
13	57.0	male	atyp_a...	154.0	232.0	f	left_vent...	164.0	no	0.0	up	1.0	normal	)50_1
14	57.0	male	asympt	110.0	335.0	f	normal	143.0	yes	3.0	flat	1.0	revers...	)50_1
15	55.0	female	asympt	128.0	205.0	f	st_t_wav...	130.0	yes	2.0	flat	1.0	revers...	)50_1
16	61.0	male	asympt	148.0	203.0	f	normal	161.0	no	0.0	up	1.0	revers...	)50_1
17	58.0	male	asympt	114.0	318.0	f	st_t_wav...	140.0	no	4.4	down	3.0	fixed_d...	)50_1
18	58.0	female	asympt	170.0	225.0	t	left_vent...	146.0	yes	2.8	flat	2.0	fixed_d...	)50_1
19	63.0	male	asympt	140.0	187.0	f	left_vent...	144.0	yes	4.0	up	2.0	revers...	)50_1
20	59.0	male	asympt	164.0	176.0	t	left_vent...	90.0	no	1.0	flat	2.0	fixed_d...	)50_1
21	68.0	male	asympt	144.0	193.0	t	normal	141.0	no	3.4	flat	2.0	revers...	)50_1
22	57.0	male	asympt	130.0	131.0	f	normal	115.0	yes	1.2	flat	1.0	revers...	)50_1
23	57.0	female	atyp_a...	130.0	236.0	f	left_vent...	174.0	no	0.0	flat	1.0	normal	)50_1

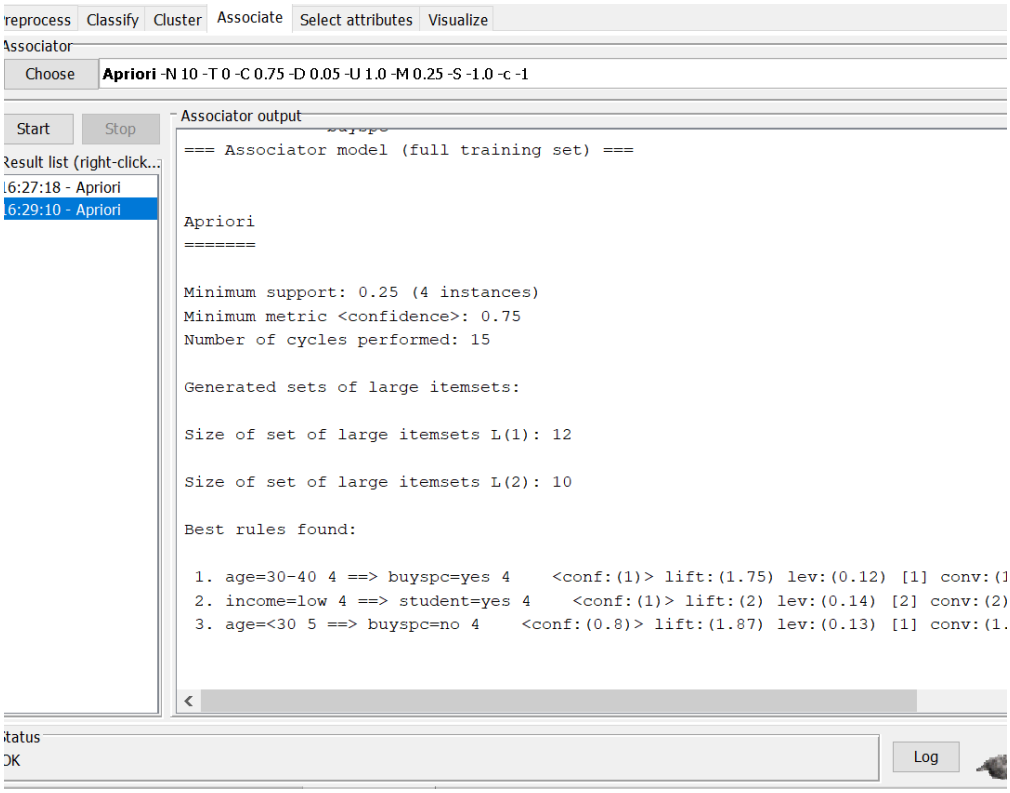
Add instance

Qs-2



Qs-3


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-click...)
16:27:18 - Apriori
Associator output
Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 48
Size of set of large itemsets L(3): 38
Size of set of large itemsets L(4): 5
Best rules found:
1. age=30-40 4 ==> buyspc=yes 4 <conf:(1)> lift:(1.75) lev:(0.12) [1] conv:(1.71
2. income=low 4 ==> student=yes 4 <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. age=<30 student=no 3 ==> buyspc=no 3 <conf:(1)> lift:(2.33) lev:(0.12) [1] co
4. credit-rating=fair buyspc=no 3 ==> age=<30 3 <conf:(1)> lift:(2.8) lev:(0.14)
5. age=<30 credit-rating=fair 3 ==> buyspc=no 3 <conf:(1)> lift:(2.33) lev:(0.12
6. age=>40 buyspc=yes 3 ==> credit-rating=fair 3 <conf:(1)> lift:(1.75) lev:(0.0
7. age=>40 credit-rating=fair 3 ==> buyspc=yes 3 <conf:(1)> lift:(1.75) lev:(0.0
8. age=<30 income=high 2 ==> student=no 2 <conf:(1)> lift:(2) lev:(0.07) [1] con
9. income=high buyspc=no 2 ==> age=<30 2 <conf:(1)> lift:(2.8) lev:(0.09) [1] co
10. age=<30 income=high 2 ==> buyspc=no 2 <conf:(1)> lift:(2.33) lev:(0.08) [1] c
Status
OK
Log
x 0


reprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.75 -D 0.05 -U 1.0 -M 0.25 -S -1.0 -c -1
Start
Stop
Result list (right-click...)
16:27:18 - Apriori
16:29:10 - Apriori
Associator output
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.25 (4 instances)
Minimum metric <confidence>: 0.75
Number of cycles performed: 15
Generated sets of large itemsets:
Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 10
Best rules found:
1. age=30-40 4 ==> buyspc=yes 4 <conf:(1)> lift:(1.75) lev:(0.12) [1] conv:(1
2. income=low 4 ==> student=yes 4 <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. age=<30 5 ==> buyspc=no 4 <conf:(0.8)> lift:(1.87) lev:(0.13) [1] conv:(1.
tatus
OK
Log

Qs-4

a)

```
n_num = [1, 2, 3, 4, 5]
n = len(n_num)

get_sum = sum(n_num)
mean = get_sum / n

print("Mean / Average is: " + str(mean))

n_num.sort()

if n % 2 == 0:
    median1 = n_num[n//2]
    median2 = n_num[n//2 - 1]
    median = (median1 + median2)/2
else:
    median = n_num[n//2]
print("Median is: " + str(median))
data = Counter(n_num)
get_mode = dict(data)
mode = [k for k, v in get_mode.items() if v == max(list(data.values()))]

if len(mode) == n:
    get_mode = "No mode found"
else:
    get_mode = "Mode is / are: " + ', '.join(map(str, mode))

print(get_mode)
```

b)

```
a = dataset.data

b = np.zeros(150)

for i in range (150):

    b[i]=a[i,1]

b=np.sort(b)  #sort the array

bin1=np.zeros((30,5))

bin2=np.zeros((30,5))

bin3=np.zeros((30,5))

for i in range (0,150,5):

    k=int(i/5)

    mean=(b[i] + b[i+1] + b[i+2] + b[i+3] + b[i+4])/5

    for j in range(5):
```

```
        bin1[k,j]=mean
print("Bin Mean: \n",bin1)
for i in range (0,150,5):
    k=int(i/5)
    for j in range (5):
        if (b[i+j]-b[i]) < (b[i+4]-b[i+j]):
            bin2[k,j]=b[i]
        else:
            bin2[k,j]=b[i+4]
print("Bin Boundaries: \n",bin2)
for i in range (0,150,5):
    k=int(i/5)
    for j in range (5):
        bin3[k,j]=b[i+2]
print("Bin Median: \n",bin3)
```