

In [5]:

import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt

In [6]:

data =pd.read\_csv('C:/Users/Baihaki/Downloads/datamining-master/datamining-master/Uas/dataset1.csv',delimiter=";")

In [7]:

data.head()

Out[7]:

	Age	Income	Student	Credit_rating	Class (buy_computer)
0	<=30	High	No	Fair	No
1	<=30	High	No	Excellent	No
2	31..40	High	No	Fair	Yes
3	> 40	Medium	No	Fair	Yes
4	> 40	Low	Yes	Fair	Yes

In [8]:

data.tail(10)

Out[8]:

	Age	Income	Student	Credit_rating	Class (buy_computer)
41	> 40	Low	Yes	Fair	No
42	31..40	Low	Yes	Fair	Yes
43	31..40	Low	Yes	Excellent	No
44	<= 30	High	No	Excellent	No
45	<= 30	Medium	Yes	Excellent	Yes
46	> 40	Low	Yes	Fair	Yes
47	<= 30	Low	Yes	Fair	Yes
48	31..40	Medium	No	Fair	No
49	31..40	High	Yes	Excellent	Yes
50	> 40	Medium	No	Excellent	No

In [9]:

data['Age'].value\_counts()

Out[9]:

> 40        17  
<= 30       15  
31..40      14  
<=30        5  
Name: Age, dtype: int64

In [10]:

data['Income'].value\_counts()

Out[10]:

Low        21  
Medium    19  
High       11  
Name: Income, dtype: int64

In [11]:

data['Student'].value\_counts()

Out[11]:

Yes        27  
No         24  
Name: Student, dtype: int64

In [12]:

data['Credit\_rating'].value\_counts()

Out[12]:

Fair        31  
Excellent   20  
Name: Credit\_rating, dtype: int64

In [13]:

data['Class (buy\_computer)'].value\_counts()

Out[13]:

Yes        29  
No         22  
Name: Class (buy\_computer), dtype: int64

In [14]:

data.shape

Out[14]:

(51, 5)

In [15]:

PYes = 27/51  
PNo = 24/51

In [16]:

pd.crosstab(data['Age'],data['Income'])

Out[16]:

	Income	High	Low	Medium
Age				
31..40		5	5	4
<= 30		1	6	8
<=30		5	0	0
> 40		0	10	7

In [17]:

pd.crosstab(data['Age'],data['Student'])

Out[17]:

	Student	No	Yes
Age			
31..40		7	7
<= 30		6	9
<=30		5	0
> 40		6	11

In [18]:

pd.crosstab(data['Age'],data['Credit\_rating'])

Out[18]:

	Credit_rating	Excellent	Fair
Age			
31..40		7	7
<= 30		4	11
<=30		3	2
> 40		6	11

In [19]:

pd.crosstab(data['Income'],data['Class (buy\_computer)'])

Out[19]:

	Class (buy_computer)	No	Yes
Income			
High		6	5
Low		11	10
Medium		5	14

In [20]:

pd.crosstab(data['Income'],data['Credit\_rating'])

Out[20]:

	Credit_rating	Excellent	Fair
Income			
High		5	6
Low		8	13
Medium		7	12

In [21]:

pd.crosstab(data['Income'],data['Age'])

Out[21]:

	Age	31..40	<= 30	<=30	> 40
Income					
High		5	1	5	0
Low		5	6	0	10
Medium		4	8	0	7

In [22]:

PHighNo = 6/22  
PLowNo = 11/22  
PMediumNo = 5/22  
  
PHighYes = 5/29  
PLowYes = 10/29  
PMediumYes = 5/29  
  
PHigh = 11/52  
PLow = 21/51  
PMedium = 19/51  
  
print (PHighNo)  
  
0.2727272727272727

In [23]:

print (PHighYes)  
  
0.1724137931034483

In [24]:

print (PHigh)  
  
0.21153846153846154

In [25]:

print (PLowNo)  
  
0.5

In [26]:

print (PLowYes)  
  
0.3448275862068966

In [27]:

print (PLow)  
  
0.4117647058823529

In [28]:

print (PMediumYes)  
  
0.1724137931034483

In [29]:

print (PMediumNo)  
  
0.22727272727272727

In [30]:

print (PMedium)  
  
0.37254901960784315

In [31]:

data.describe()

Out[31]:

	Age	Income	Student	Credit_rating	Class (buy_computer)
count	51	51	51	51	51
unique	4	3	2	2	2
top	> 40	Low	Yes	Fair	Yes
freq	17	21	27	31	29

In [33]:

%notebook "C:/Users/Baihaki/Downloads/datamining-master/datamining-master/Uas/JawabanNo1a.ipynb"

In [34]:

data.to\_excel('C:/Users/Baihaki/Downloads/datamining-master/datamining-master/Uas/JawabanNo1.xls')

In [ ]: