**Assignment 1**

**Association Rules**

**1.Consider following observations/data. And apply simple linear regression and find out estimated coefficients b1 and b1 Also analyse the performance of the model (Use sklearn package)**

import matplotlib.pyplot as plt

from scipy import stats

import numpy as np

x = np.array([1,2,3,4,5,6,7,8])

y = np.array([7,14,15,18,19,21,26,23])

slope, intercept, r, p, std\_err = stats.linregress(x, y)

def myfunc(x):

return slope \* x + intercept

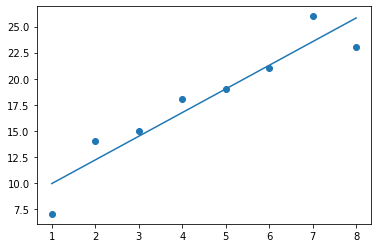
mymodel = list(map(myfunc, x))

plt.scatter(x, y)

plt.plot(x, mymodel)

plt.show()

**#output**



**2.Write python program to read “studentPerformance.csv” file. solve the following.**

**-display shape of dataset**

**-display top rows of dataset**

import numpy as np

import pandas as pd

import os

dataset=pd.read\_csv("StudentsPerformance.csv")

#To display the shape of dataset.

dataset

#To display the top rows of the dataset with their columns.

dataset.head()

#To display the number of rows randomly.

dataset.sample(5)

#To display the number of columns and names of the columns.

rows = len(dataset.axes[0])

cols = len(dataset.axes[1])

# Print the number of rows and columns

print("Number of Rows: " + str(rows))

print("Number of Columns: " + str(cols))

#name of columns

for col in dataset.columns:

print(col)

**#output**

Number of Rows: 1000

Number of Columns: 8

gender

race/ethnicity

parental level of education

lunch

test preparation course

math score

reading score

writing score

**3.write a python program for aprior algorithm using ARM and print the rule ----**

import pandas as pd

import os

dataset=pd.read\_csv("groceries2.csv")

print(dataset.head(10))

transactions = []

# Add all the items from each row in one list( Neglect the 1st columns where all the items are in number (0-9))

for i in range(0, 100):

transactions.append([str(dataset.values[i,u]) for u in range(10, 32)])

from apyori import apriori

rules = apriori(transactions, min\_support=0.0040, min\_confidence=0.2, min\_lift=3, min\_length = 2)

results = list(rules)

# See the items that were bought together with their support

results\_list = []

for i in range(0, len(results)):

results\_list.append('RULE:\t' + str(results[i][0]) + '\nSUPPORT:\t' + str(results[i][1]))

print(dataset.head())

**#output**

ModuleNotFoundError: No module named 'apyori'

**Assignment 2**

**Clustering**

1. **Write python program to read “car.csv” file.**

import pandas

from sklearn import linear\_model

df = pandas.read\_csv("cars.csv")

X = df[['Weight', 'Volume']]

y = df['CO2']

regr = linear\_model.LinearRegression()

regr.fit(X, y)

predictedCO2 = regr.predict([[2300, 1300]])

print(predictedCO2)

**#output**

ModuleNotFoundError: No module named 'sklearn'

**2.linear regression**

import pandas as pd

import matplotlib.pyplot as plt

Stock\_Market = {'Year':[2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016,2016],

'Month': [12, 11,10,9,8,7,6,5,4,3,2,1,12,11,10,9,8,7,6,5,4,3,2,1],

'Interest\_Rate':[2.75,2.5,2.5,2.5,2.5,2.5,2.5,2.25,2.25,2.25,2,2,2,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75,1.75],

'Unemployment\_Rate':[5.3,5.3,5.3,5.3,5.4,5.6,5.5,5.5,5.5,5.6,5.7,5.9,6,5.9,5.8,6.1,6.2,6.1,6.1,6.1,5.9,6.2,6.2,6.1],

'Stock\_Index\_Price':[1464,1394,1357,1293,1256,1254,1234,1195,1159,1167,1130,1075,1047,965,943,958,971,949,884,866,876,822,704,719]

}df = pd.DataFrame(Stock\_Market,columns=['Year','Month','Interest\_Rate','Unemployment\_Rate','Stock\_Index\_Price'])

plt.scatter(df['Interest\_Rate'], df['Stock\_Index\_Price'], color='red')

plt.title('Stock Index Price Vs Interest Rate', fontsize=14)

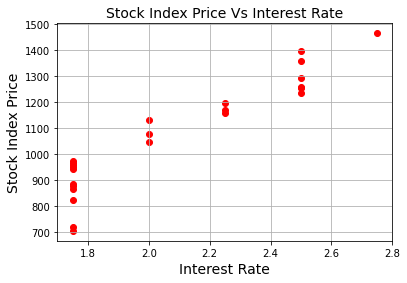
plt.xlabel('Interest Rate', fontsize=14)

plt.ylabel('Stock Index Price', fontsize=14)

plt.grid(True)

plt.show()

**#output**



**Assignment 3**

**1.K-mean algorithm**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

from sklearn.preprocessing import StandardScaler, normalize

from sklearn.decomposition import PCA

from sklearn.metrics import silhouette\_score

raw\_df = pd.read\_csv("GENERAL.csv")

raw\_df = raw\_df.drop('CUST\_ID', axis = 1)

raw\_df.fillna(method ='ffill', inplace = True)

raw\_df.head(2)

**#output**

ModuleNotFoundError: No module named 'sklearn'