**Slip1 & Slip 11**

Q2A)

import pandas as pd

import matplotlib.pyplot as plt

d = pd.read\_csv('C:\\Users\\DELL\\Untitled Folder\\Iris.csv')

ax=plt.subplots(1,1,figsize=(10,8)) //defines size of chart area

d['Species'].value\_counts().plot.pie() //counts distinct values in dataset

plt.title("Iris Species %")

plt.show()

Q2B)

import pandas as p

df = pd.read\_csv('C:\\Users\\DELL\\winequality-red.csv')

df.shape # no.of rows & cols

df.describe() #stats data

df.info() #features

df.dtypes

**Slip2 & slip6**

Q2 A)

import pandas as p

import numpy as n

d=p.read\_csv('D:\yogita\ss.csv')

v=d['age'].mean()

v1=d['salary'].mean()

d['age'].fillna(v,inplace=True)

d['salary'].fillna(v1,inplace=True)

print(d)

Q2B)

import numpy as np

import matplotlib.pyplot as plt

import pandas as p

df=p.DataFrame({'name':['kunal','rekha','satish','ashish','radha'],

'age':[20,23,22,20,21],

'per':[98,80,95,92,85],

'salary':[100000,300000,20000,300000,80000] })

df.plot(x="name",y="salary")

plt.show()

Q2C)

import pandas as p

df=p.read\_csv("ht&wt.csv")

print("first 10 rows \n",df.head(10))

print("\n random 20 rows\n",df.sample(20))

print("\n shape \n" ,df.shape)

**Slip 3**

Q2A)

import pandas as p

d=p.read\_csv('C:\\Users\\DELL\\Untitled Folder\\Iris.csv')

#remove id field from iris dataset

new\_data = d[["SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm"]]

print(new\_data)

plt.figure(figsize = (10, 7))

new\_data.boxplot()

Q2B)

import pandas as p

df = pd.read\_csv('C:\\Users\\DELL\\ht&wt.csv')

df.shape # no.of rows & cols

df.describe() #stats data

df.info() #features

df.dtypes

**Slip 4 and Slip5**

Q2A)

import matplotlib.pyplot as plt

import numpy as np

x = np.random.randn(50)

y = np.random.randn(50)

plt.plot(x,y)

plt.show()

plt.scatter(x,y)

plt.show()

plt.hist(x)

plt.show()

plt.boxplot(y, vert=False)

plt.show()

Q2b)

import pandas as p

df = pd.read\_csv('C:\\Users\\DELL\\User\_Data.csv')

df.shape # no.of rows & cols

df.describe() #stats data

df.info() #features

df.dtypes

**Slip 7 &slip29**

**Q2)**

import pandas as p

from sklearn import preprocessing

d = pd.read\_csv('D:\\yogita\\Data.csv')

label\_encoder = preprocessing.LabelEncoder()

d['purchased']= label\_encoder.fit\_transform(d['purchased'])

one\_hot\_encoded\_data = p.get\_dummies(d, columns = ['country'])

print(one\_hot\_encoded\_data)

**Slip 9 &slip 15**

Q2A)

import matplotlib.pyplot as plt

import numpy as np

from matplotlib import colors

from matplotlib.ticker import PercentFormatter

no\_of\_balls=50

x = np.random.randn(50)

y = np.random.randn(50)

colors = [np.random.randint(1, 4) for i in range(no\_of\_balls)]

plt.plot(x,y)

plt.show()

plt.scatter(x,y,c=colors)

plt.show()

Q2B)

from matplotlib import pyplot as plt

import numpy as np

# Creating dataset

subjects = ['TCS', 'Data Science', 'OS',

'JAVA', 'PHP', 'Python']

marks = [23, 17, 35, 29, 12, 33]

# Creating plot

fig = plt.figure(figsize =(10, 7))

plt.pie(marks, labels = subjects)

# show plot

plt.show()

Q2C)

import pandas as p

df = pd.read\_csv('C:\\Users\\DELL\\winequality-red.csv')

print("\n",df.shape) # no.of rows & cols

print("\n",df.describe()) #stats data

df.head(3)

**Slip 10**

Q2A)

import pandas as p

df=p.read\_csv("ht&wt.csv")

print("mean is \n",df.mean)

print("median is \n",df.median)

Q2B)

def distancesum (x, y, n):

sum = 0

# for each point, finding distance

# to rest of the point

for i in range(n):

for j in range(i+1,n):

sum += (abs(x[i] - x[j]) +

abs(y[i] - y[j]))

return sum

x = [ -1, 1, 3, 2 ]

y = [ 5, 6, 5, 3 ]

n = len(x)

print(distancesum(x, y, n) )

**Slip 12**

Q2A)

import matplotlib.pyplot as plt

import numpy as np

x = np.random.randn(50)

y = np.random.randn(50)

plt.plot(x,y)

plt.show()

plt.scatter(x,y)

plt.show()

plt.hist(x)

plt.show()

plt.boxplot(y, vert=False)

plt.show()

Q2B)

import pandas as p

df=p.DataFrame({'name':['kunal','rekha','satish','ashish','radha'],

'dept':['production','computer','manufacturing',None,'manufacturing'],

'salary':[100000,300000,20000,300000,80000] })

print(df)

d=df.dropna()

print(d)

**Slip 13**

Q2A)

import pandas as p

import matplotlib.pyplot as plt

d=p.read\_csv('C:\\Users\\DELL\\Untitled Folder\\Iris.csv')

fig = d[d.Species=='Iris-setosa'].plot.bar(x='PetalLengthCm',y='PetalWidthCm',color='orange', label='Setosa')

d[d.Species=='Iris-versicolor'].plot.bar(x='PetalLengthCm',y='PetalWidthCm',color='blue', label='versicolor',ax=fig)

d[d.Species=='Iris-virginica'].plot.bar(x='PetalLengthCm',y='PetalWidthCm',color='green', label='virginica', ax=fig)

fig.set\_xlabel("Petal Length")

fig.set\_ylabel("Petal Width")

fig.set\_title(" Petal Length VS Petal Width")

#fig=plt.gcf()

#fig.set\_size\_inches(12,8)

plt.show()

Q2B)

import numpy as n

d=n.array([[0,1],[2,3]])

print(d.max())

print(d.min())

**Slip14**

Q2A)

**import** numpy as np

# Original array

array **=** np.arange(5)

**print**(array)

weights **=** np.arange(10, 15)

print(weights)

# Weighted average of the given array

res1 **=** np.average(array, weights**=**weights)

print(res1)

Q2B)

import pandas as p

df = pd.read\_csv('C:\\Users\\DELL\\Advertising.csv')

df.shape # no.of rows & cols

df.describe() #stats data

df.info() #features

df.dtypes

Slip 16

Q2A)

from matplotlib import pyplot as plt

import numpy as np

# Creating dataset

subjects = ['TCS', 'Data Science', 'OS',

'JAVA', 'PHP', 'Python']

marks = [23, 17, 35, 29, 12, 33]

# Creating plot

fig = plt.figure(figsize =(10, 7))

plt.pie(marks, labels = subjects)

csv

# show plot

plt.show()

Q2B)

import pandas as p

import numpy as n

df=p.DataFrame({'name':['kunal','rekha','satish','ashish','radha'],

'age':[20,23,22,20,21],

'per':[98,80,95,92,85]})

print(n.average(df['age']))

print(n.average(df['per']))

slip 17

Q2B)

import pandas as p

df=p.DataFrame({'name':['kunal','rekha','satish','ashish','radha'],

'age':[20,23,22,20,21],

'salary':[100000,300000,20000,300000,80000] })

df

Q2A)

import pandas as p

import matplotlib.pyplot as plt

d=p.read\_csv('C:\\Users\\DELL\\Untitled Folder\\Iris.csv')

fig = d[d.Species=='Iris-setosa'].plot.scatter(x='PetalLengthCm',y='PetalWidthCm',color='orange', label='Setosa')

d[d.Species=='Iris-versicolor'].plot.scatter(x='PetalLengthCm',y='PetalWidthCm',color='blue', label='versicolor',ax=fig)

fig.set\_xlabel("Petal Length")

fig.set\_ylabel("Petal Width")

fig.set\_title(" Petal Length VS Petal Width")

plt.show()

**Slip 18**

Q2A)

import pandas as p

d=p.read\_csv('C:\\Users\\DELL\\Untitled Folder\\Iris.csv')

#remove id field from iris dataset

new\_data = d[["SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm"]]

print(new\_data)

plt.figure(figsize = (10, 7))

new\_data.boxplot()

Q2B)

import pandas as p

df = pd.read\_csv('C:\\Users\\DELL\\ht&wt.csv')

print(df.head(5))

print(df.tail(5))

print(df.sample(10))

**Slip 19 & Slip 28**

Q2A)

import pandas as p

df=p.DataFrame(columns =['name','age','per'])

df.loc[0]=['rajesh',20,95]

df.loc[1]=['suresh',21,85]

df.loc[2]=['avinash',20,90]

df.loc[3]=['kunal',21,75]

df.loc[4]=['sakshi',20,80]

df.loc[6]=['xxx',np.nan,95]

df.loc[7]=['suresh',21,85]

df.loc[8]=['archana',22,91]

df.loc[9]=['kunal',20,np.nan]

print(df)

print(df.shape)

print(df.describe)

print(df.info())

print(df.dtypes)

df["remark"]=None

df

**Slip 20**

Q2A)

import matplotlib.pyplot as plt

import numpy as np

x = np.random.randn(50)

y = np.random.randn(50)

plt.plot(x,y)

plt.show()

plt.scatter(x,y)

plt.show()

plt.hist(x)

plt.show()

Q2B)

plt.boxplot(y, vert=False)

plt.show()

**Slip 21 and 24**

Q2A)

import pandas as p

import matplotlib.pyplot as plt

d=p.read\_csv('C:\\Users\\DELL\\Untitled Folder\\Iris.csv')

d[d.Species=='Iris-setosa'].plot.bar(x='PetalLengthCm',y='PetalWidthCm',color='orange', label='Setosa')

d[d.Species=='Iris-versicolor'].plot.bar(x='PetalLengthCm',y='PetalWidthCm',color='blue', label='versicolor')

d[d.Species=='Iris-virginica'].plot.bar(x='PetalLengthCm',y='PetalWidthCm',color='green', label='virginica')

fig.set\_xlabel("Petal Length")

fig.set\_ylabel("Petal Width")

fig.set\_title(" Petal Length VS Petal Width")

#fig=plt.gcf()

#fig.set\_size\_inches(12,8)

plt.show()

Q2B)

import pandas as p

import matplotlib.pyplot as plt

d=p.read\_csv('C:\\Users\\DELL\\Untitled Folder\\Iris.csv')

d[d.Species=='Iris-setosa'].plot.hist(x='PetalLengthCm',y='PetalWidthCm',color='orange', label='Setosa')

d[d.Species=='Iris-versicolor'].plot.hist(x='PetalLengthCm',y='PetalWidthCm',color='blue', label='versicolor')

d[d.Species=='Iris-virginica'].plot.hist(x='PetalLengthCm',y='PetalWidthCm',color='green', label='virginica')

fig.set\_xlabel("Petal Length")

fig.set\_ylabel("Petal Width")

fig.set\_title(" Petal Length VS Petal Width")

#fig=plt.gcf()

#fig.set\_size\_inches(12,8)

plt.show()

**Slip 25 & slip 26 &Slip 30**

Q2A)

import matplotlib.pyplot as plt

import numpy as np

x = np.random.randn(50)

y = np.random.randn(50)

plt.plot(x,y)

plt.show()

plt.scatter(x,y,color=’green’)

plt.show()

plt.hist(x,color=’yellow’)

plt.show()

plt.boxplot(y, vert=False)

plt.show()

Q2B)

from matplotlib import pyplot as plt

import numpy as np

# Creating dataset

subjects = ['TCS', 'Data Science', 'OS',

'JAVA', 'PHP', 'Python']

marks = [23, 17, 35, 29, 12, 33]

# Creating plot

fig = plt.figure(figsize =(10, 7))

plt.pie(marks, labels = subjects)

# show plot

plt.show()

**Slip 27**

Q2A)

import pandas as p

from sklearn import preprocessing

d = pd.read\_csv('D:\\yogita\\Data.csv')

label\_encoder = preprocessing.LabelEncoder()

d['purchased']= label\_encoder.fit\_transform(d['purchased'])

one\_hot\_encoded\_data = p.get\_dummies(d, columns = ['country'])

print(one\_hot\_encoded\_data)