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 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)
```