Pooja Agarwal
1905330

Lab 1

Q1: In the dataset “data.csv”

CODE:

```
1  import pandas as pd
2
3  df = pd.read_csv("Data.csv")
4
5  print(df)
```

OUTPUT:

```
In [6]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab1/q1.py', wdir='C:/Users/KIIT/
Desktop/Assignments/TNT/Lab1')
   Country   Age   Salary Purchased
0   France  44.0  72000.0        No
1    Spain  27.0  48000.0       Yes
2  Germany  30.0      NaN        No
3    Spain  38.0  61000.0        No
4  Germany  40.0  70000.0       Yes
5   France  35.0  58000.0       Yes
6    Spain   NaN  52000.0        No
7   France  48.0  79000.0       Yes
8  Germany  50.0  83000.0        No
9      NaN  37.0  67000.0       Yes
```

Pooja Agarwal
1905330

Q2: Add a new column : Salary_class
A for loop is implemented and the observations are separated into three categories:
o Salary
• greater than 70000 - class0
• between 61000-70000 -class1
• between 48000-60000 -class2
• The classes have been stored in a new column 'Salary Class'

CODE:

```
import pandas as pd

df = pd.read_csv("Data.csv")

sal_class = []

for i in range(10):

    sal = df['Salary'][i]

    if sal>70000:
        sal_class.append('class0')
    elif sal>=61000:
        sal_class.append('class1')
    elif sal >= 48000:
        sal_class.append('class2')
    else:
        sal_class.append('')

df['Salary_class'] = sal_class

print(df)
```

OUTPUT:

```
In [7]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab1/untitled0.py', wdir='C:/Users/
KIIT/Desktop/Assignments/TNT/Lab1')
   Country   Age   Salary Purchased Salary_class
0   France  44.0  72000.0        No       class0
1    Spain  27.0  48000.0       Yes       class2
2  Germany  30.0      NaN        No
3    Spain  38.0  61000.0        No       class1
4  Germany  40.0  70000.0       Yes       class1
5   France  35.0  58000.0       Yes       class2
6    Spain   NaN  52000.0        No       class2
7   France  48.0  79000.0       Yes       class0
8  Germany  50.0  83000.0        No       class0
9      NaN  37.0  67000.0       Yes       class1
```

Pooja Agarwal
1905330

Q3: Implement above using both for and while loop

CODE:

```
import pandas as pd

df = pd.read_csv("Data.csv")

sal_class = []

i = 0
while i < 10:

    sal = df['Salary'][i]

    if sal>70000:
        sal_class.append('class0')
    elif sal>=61000:
        sal_class.append('class1')
    elif sal >= 48000:
        sal_class.append('class2')
    else:
        sal_class.append('')

    i += 1

df['Salary_class'] = sal_class

print(df)
```

OUTPUT:

```
In [8]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab1/untitled1.py', wdir='C:/Users/
KIIT/Desktop/Assignments/TNT/Lab1')
   Country   Age   Salary Purchased Salary_class
0   France  44.0  72000.0        No       class0
1    Spain  27.0  48000.0       Yes       class2
2  Germany  30.0      NaN        No
3    Spain  38.0  61000.0        No       class1
4  Germany  40.0  70000.0       Yes       class1
5   France  35.0  58000.0       Yes       class2
6    Spain   NaN  52000.0        No       class2
7   France  48.0  79000.0       Yes       class0
8  Germany  50.0  83000.0        No       class0
9      NaN  37.0  67000.0       Yes       class1
```

Q4: Count the number of each class (class 0, class1,class2) in your dataset.

CODE:

```
import pandas as pd

df = pd.read_csv("Data.csv")

sal_class = []

i = 0
while i < 10:

    sal = df['Salary'][i]

    if sal>70000:
        sal_class.append('class0')
    elif sal>=61000:
        sal_class.append('class1')
    elif sal >= 48000:
        sal_class.append('class2')
    else:
        sal_class.append('')

    i += 1

df['Salary_class'] = sal_class


c0 = len(df[df['Salary_class'] == 'class0'])
c1 = len(df[df['Salary_class'] == 'class1'])
c2 = len(df[df['Salary_class'] == 'class2'])

print(f'class0 = {c0}, class1 = {c1}, class2 = {c2}')
```

OUPUT:

```
In [9]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab1/untitled2.py', wdir='C:/Users/
KIIT/Desktop/Assignments/TNT/Lab1')
class0 = 3, class1 = 3, class2 = 3
```

Pooja Agarwal
1905330

Q5: Insert a new column Age_Converted:
Use function c_convert to add in the new column the converted values fromcolumn “Age” :
dataset[“Age_Converted”]=dataset[“Age”]*12

CODE:

```
import pandas as pd

df = pd.read_csv("Data.csv")

sal_class = []

i = 0
while i < 10:

    sal = df['Salary'][i]

    if sal>70000:
        sal_class.append('class0')
    elif sal>=61000:
        sal_class.append('class1')
    elif sal >= 48000:
        sal_class.append('class2')
    else:
        sal_class.append('')

    i += 1

df['Salary_class'] = sal_class

age_con = df['Age']*12

df['Age_Converted'] = age_con

print(df)
```

OUTPUT:

```
In [10]: runfile('C:/Users/KIIT/Desktop/Assignments/TNT/Lab1/untitled3.py', wdir='C:/Users/
KIIT/Desktop/Assignments/TNT/Lab1')
   Country   Age   Salary Purchased Salary_class  Age_Converted
0   France  44.0  72000.0        No       class0          528.0
1    Spain  27.0  48000.0       Yes       class2          324.0
2  Germany  30.0      NaN        No                       360.0
3    Spain  38.0  61000.0        No       class1          456.0
4  Germany  40.0  70000.0       Yes       class1          480.0
5   France  35.0  58000.0       Yes       class2          420.0
6    Spain   NaN  52000.0        No       class2            NaN
7   France  48.0  79000.0       Yes       class0          576.0
8  Germany  50.0  83000.0        No       class0          600.0
9      NaN  37.0  67000.0       Yes       class1          444.0
```