Submission Date: 30.08.2023

ASSIGNMENT -1

Artificial Intelligence & Machine Learning in collaboration with Google (Applied Data Science)

Name: Rishikesh S

Regno: 21BME0159

Branch: Btech Mechanical Engineering

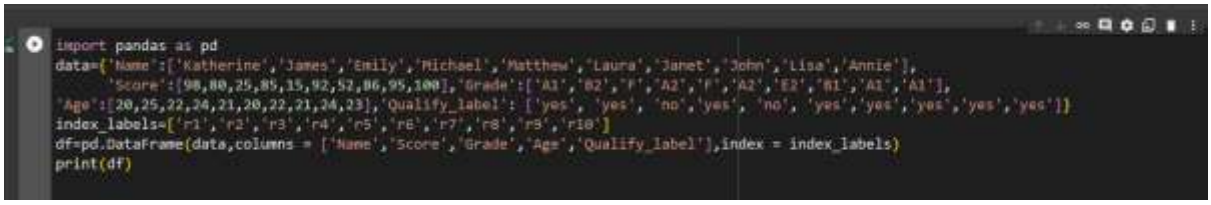
Campus: VIT Vellore

Task -1

Create a pandas dataframe (Dataframe name as ‘df’)
(10 observations and 5 features).

Input

```
import pandas as pd
data={'Name':['Katherine','James','Emily','Michael'
,'Matthew','Laura','Janet','John','Lisa','Annie'],
      'Score':[98,80,25,85,15,92,52,86,95,100],'Gra
de':['A1','B2','F','A2','F','A2','E2','B1','A1','A1
'],
'Age':[20,25,22,24,21,20,22,21,24,23],'Qualify_labe
l': ['yes', 'yes', 'no','yes', 'no',
'yes','yes','yes','yes','yes']}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r
8','r9','r10']
df=pd.DataFrame(data,columns =
['Name','Score','Grade','Age','Qualify_label'],inde
x = index_labels)
print(df)
```



Output

```
        Name     Score Grade  Age    Qualify_label
r1   Katherine     98    A1   20            yes
r2       James     80    B2   25            yes
r3       Emily     25     F   22             no
r4     Michael     85    A2   24            yes
r5     Matthew     15     F   21             no
r6       Laura     92    A2   20            yes
r7       Janet     52    E2   22            yes
r8        John     86    B1   21            yes
r9        Lisa     95    A1   24            yes
r10      Annie    100    A1   23            yes
```

	Name	Score	Grade	Age	Qualify_label
r1	Katherine	98	A1	20	yes
r2	James	80	B2	25	yes
r3	Emily	25	F	22	no
r4	Michael	85	A2	24	yes
r5	Matthew	15	F	21	no
r6	Laura	92	A2	20	yes
r7	Janet	52	E2	22	yes
r8	John	86	B1	21	yes
r9	Lisa	95	A1	24	yes
r10	Annie	100	A1	23	yes

Task -2

Check the info of df .

Input

```
import pandas as pd
data={'Name':['Katherine','James','Emily','Michael'
,'Matthew','Laura','Janet','John','Lisa','Annie'],
      'Score':[98,80,25,85,15,92,52,86,95,100],'Gra
de':['A1','B2','F','A2','F','A2','E2','B1','A1','A1
'],
'Age':[20,25,22,24,21,20,22,21,24,23],'Qualify_labe
l': ['yes', 'yes', 'no','yes', 'no',
'yes','yes','yes','yes','yes']}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r
8','r9','r10']
df=pd.DataFrame(data,columns =
['Name','Score','Grade','Age','Qualify_label'],inde
x = index_labels)
print(df)
df.info()
```

Output

```
<class 'pandas.core.frame.DataFrame'>
Index: 10 entries, r1 to r10
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Name            10 non-null     object
 1   Score           10 non-null     int64
 2   Grade           10 non-null     object
 3   Age             10 non-null     int64
 4   Qualify_label   10 non-null     object
dtypes: int64(2), object(3)
memory usage: 480.0+ bytes
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 10 entries, r1 to r10
Data columns (total 5 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Name           10 non-null     object
 1   Score          10 non-null     int64
 2   Grade          10 non-null     object
 3   Age            10 non-null     int64
 4   Qualify_label  10 non-null     object
dtypes: int64(2), object(3)
memory usage: 480.0+ bytes
```

Task -3

Check the descriptive statistics of ‘df”

Input

```
import pandas as pd
data={'Name':['Katherine','James','Emily','Michael'
,'Matthew','Laura','Janet','John','Lisa','Annie'],
      'Score':[98,80,25,85,15,92,52,86,95,100],'Gra
de':['A1','B2','F','A2','F','A2','E2','B1','A1','A1
'],
'Age':[20,25,22,24,21,20,22,21,24,23],'Qualify_labe
l': ['yes', 'yes', 'no','yes', 'no',
'yes','yes','yes','yes','yes']}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r
8','r9','r10']
df=pd.DataFrame(data,columns =
['Name','Score','Grade','Age','Qualify_label'],inde
x = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
```

Output

	Name	Score	Grade	Age	Qualify_label
count	10	10.000000	10	10.00000	10
unique	10	NaN	6	NaN	2
top	Katherine	NaN	A1	NaN	yes
freq	1	NaN	3	NaN	8
mean	NaN	72.800000	NaN	22.20000	NaN
std	NaN	31.036896	NaN	1.75119	NaN
min	NaN	15.000000	NaN	20.00000	NaN
25%	NaN	59.000000	NaN	21.00000	NaN
50%	NaN	85.500000	NaN	22.00000	NaN
75%	NaN	94.250000	NaN	23.75000	NaN
max	NaN	100.000000	NaN	25.00000	NaN

	Name	Score	Grade	Age	Qualify_label
count	10	10.000000	10	10.00000	10
unique	10	NaN	6	NaN	2
top	Katherine	NaN	A1	NaN	yes
freq	1	NaN	3	NaN	8
mean	NaN	72.800000	NaN	22.20000	NaN
std	NaN	31.036896	NaN	1.75119	NaN
min	NaN	15.000000	NaN	20.00000	NaN
25%	NaN	59.000000	NaN	21.00000	NaN
50%	NaN	85.500000	NaN	22.00000	NaN
75%	NaN	94.250000	NaN	23.75000	NaN
max	NaN	100.000000	NaN	25.00000	NaN

Task -4

Check the 4th index observation with ‘loc’ slicing operator.

Input

```
import pandas as pd
data={'Name':['Katherine','James','Emily','Michael'
,'Matthew','Laura','Janet','John','Lisa','Annie'],
      'Score':[98,80,25,85,15,92,52,86,95,100],'Gra
de':['A1','B2','F','A2','F','A2','E2','B1','A1','A1
'],
'Age':[20,25,22,24,21,20,22,21,24,23],'Qualify_labe
l': ['yes', 'yes', 'no','yes', 'no',
'yes','yes','yes','yes','yes']}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r
8','r9','r10']
df=pd.DataFrame(data,columns =
['Name','Score','Grade','Age','Qualify_label'],inde
x = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
df.loc['r5']
```

Output

```
Name              Matthew
Score                  15
Grade                   F
Age                    21
Qualify_label          no
Name: r5, dtype: object
```

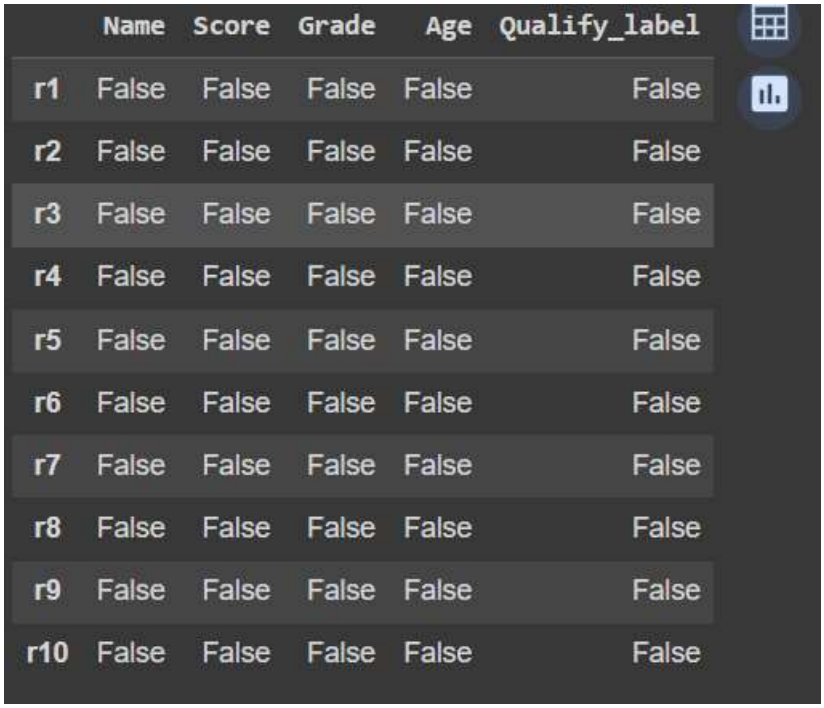
Task -5

Check the null values in your ‘df’.

Input

```
import pandas as pd
data={'Name':['Katherine','James','Emily','Michael'
,'Matthew','Laura','Janet','John','Lisa','Annie'],
      'Score':[98,80,25,85,15,92,52,86,95,100],'Gra
de':['A1','B2','F','A2','F','A2','E2','B1','A1','A1
'],
'Age':[20,25,22,24,21,20,22,21,24,23],'Qualify_labe
l': ['yes', 'yes', 'no','yes', 'no',
'yes','yes','yes','yes','yes']}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r
8','r9','r10']
df=pd.DataFrame(data,columns =
['Name','Score','Grade','Age','Qualify_label'],inde
x = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
df.loc['r5']
df.isnull()
```

Output



	Name	Score	Grade	Age	Qualify_label
r1	False	False	False	False	False
r2	False	False	False	False	False
r3	False	False	False	False	False
r4	False	False	False	False	False
r5	False	False	False	False	False
r6	False	False	False	False	False
r7	False	False	False	False	False
r8	False	False	False	False	False
r9	False	False	False	False	False
r10	False	False	False	False	False