

# ASSIGNMENT-1

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

Name: Diya Maria Sanjay

Regno: 21BML0113

Branch: Btech ECE with specialization in Biomedical  
Engineering

Campus: VIT Vellore

## Task – 1

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

### Input


```
import pandas as pd
technologies =
{'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift",
,"TypeScript","HTML","MATLAB"],
  'Fee'
:[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
  'Duration':['30days','40days','35days','40days','60days','30days','45days','35days','30days','50days'],
  'Discount':[15,23,12,25,45,15,18,20,14,15], 'Final
Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200
]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration',
'Discount','Final Amount'], index = index_labels)
print(df)
```

```
import pandas as pd
technologies = {'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift","TypeScript","HTML","MATLAB"],
  'Fee':[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
  'Duration':['30days','40days','35days','40days','60days','30days','45days','35days','30days','50days'],
  'Discount':[15,23,12,25,45,15,18,20,14,15], 'Final Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration', 'Discount','Final Amount'], index = index_labels)
print(df)
```

### Output

	Courses	Fee	Duration	Discount	Final Amount
r1	Spark	20000	30days	15	17000.00
r2	PySpark	25000	40days	23	19250.00
r3	Hadoop	26000	35days	12	22880.00
r4	Python	22000	40days	25	16500.00
r5	pandas	24000	60days	45	13200.00
r6	Rust	45000	30days	15	38250.00
r7	Swift	30000	45days	18	24600.00
r8	TypeScript	20000	35days	20	16000.00
r9	HTML	19632	30days	14	16883.52
r10	MATLAB	32000	50days	15	27200.00

	Courses	Fee	Duration	Discount	Final Amount
r1	Spark	20000	30days	15	17000.00
r2	PySpark	25000	40days	23	19250.00
r3	Hadoop	26000	35days	12	22880.00
r4	Python	22000	40days	25	16500.00
r5	pandas	24000	60days	45	13200.00
r6	Rust	45000	30days	15	38250.00
r7	Swift	30000	45days	18	24600.00
r8	TypeScript	20000	35days	20	16000.00
r9	HTML	19632	30days	14	16883.52
r10	MATLAB	32000	50days	15	27200.00


**ASSIGNMENT 1** ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```

import pandas as pd
technologies = {'Courses': ['Spark', 'PySpark', 'Hadoop', 'Python', 'pandas', 'Rust', 'Swift', 'TypeScript', 'HTML', 'MATLAB'],
                'Fee': [20000, 25000, 26000, 22000, 24000, 45000, 30000, 20000, 19632, 32000],
                'Duration': ['30days', '40days', '35days', '40days', '60days', '30days', '45days', '35days', '30days', '50days'],
                'Discount': [15, 23, 12, 25, 45, 15, 18, 20, 14, 15], 'Final Amount': [17000, 19250, 22880, 16500, 13200, 38250, 24600, 16000, 16883.52, 27200]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration', 'Discount','Final Amount'], index = index_labels)
print(df)

```

	Courses	Fee	Duration	Discount	Final Amount
r1	Spark	20000	30days	15	17000.00
r2	PySpark	25000	40days	23	19250.00
r3	Hadoop	26000	35days	12	22880.00
r4	Python	22000	40days	25	16500.00
r5	pandas	24000	60days	45	13200.00
r6	Rust	45000	30days	15	38250.00
r7	Swift	30000	45days	18	24600.00
r8	TypeScript	20000	35days	20	16000.00
r9	HTML	19632	30days	14	16883.52
r10	MATLAB	32000	50days	15	27200.00

## Task-2

### Check the info of df

#### Input

```
import pandas as pd
technologies =
{'Courses': ["Spark", "PySpark", "Hadoop", "Python", "pandas", "Rust", "Swift",
"TypeScript", "HTML", "MATLAB"],
'Fee'
: [20000, 25000, 26000, 22000, 24000, 45000, 30000, 20000, 19632, 32000],
'Duration': ['30days', '40days', '35days', '40days', '60days', '30days', '
45days', '35days', '30days', '50days'],
'Discount': [15, 23, 12, 25, 45, 15, 18, 20, 14, 15], 'Final
Amount': [17000, 19250, 22880, 16500, 13200, 38250, 24600, 16000, 16883.52, 27200
]}
index_labels=['r1', 'r2', 'r3', 'r4', 'r5', 'r6', 'r7', 'r8', 'r9', 'r10']
df = pd.DataFrame(technologies, columns = ['Courses', 'Fee', 'Duration',
'Discount', 'Final Amount'], index = 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   Courses         10 non-null    object
1   Fee             10 non-null    int64
2   Duration        10 non-null    object
3   Discount        10 non-null    int64
4   Final Amount    10 non-null    float64
dtypes: float64(1), int64(2), object(2)
memory usage: 480.0+ bytes
```

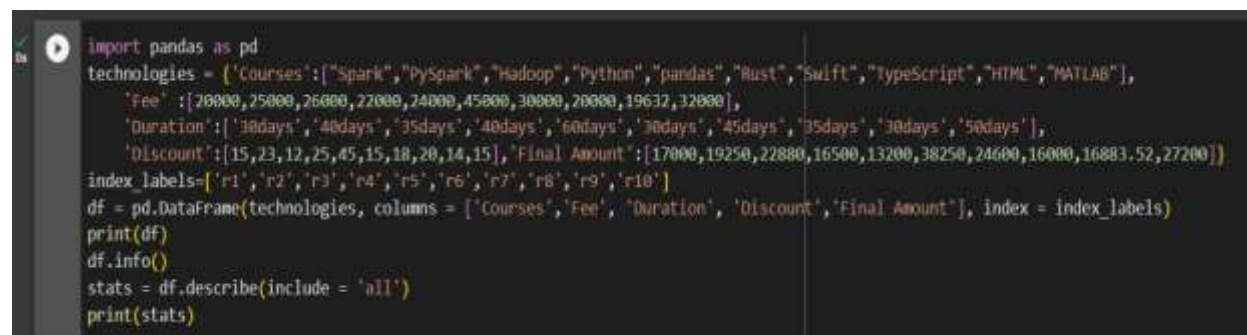
```
<class 'pandas.core.frame.DataFrame'>  
Index: 10 entries, r1 to r10  
Data columns (total 5 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   Courses         10 non-null    object  
1   Fee              10 non-null    int64  
2   Duration         10 non-null    object  
3   Discount         10 non-null    int64  
4   Final Amount    10 non-null    float64  
dtypes: float64(1), int64(2), object(2)  
memory usage: 480.0+ bytes
```

## Task -3

Check the descriptive statistics of ‘df’

### Input

```
import pandas as pd
technologies =
{'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift",
"TypeScript","HTML","MATLAB"],
'Fee'
:[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
'Duration':['30days','40days','35days','40days','60days','30days','45days','35days','30days','50days'],
'Discount':[15,23,12,25,45,15,18,20,14,15], 'Final
Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200
]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration',
'Discount','Final Amount'], index = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
```



```
import pandas as pd
technologies = {'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift","TypeScript","HTML","MATLAB"],
'Fee' :[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
'Duration':['30days','40days','35days','40days','60days','30days','45days','35days','30days','50days'],
'Discount':[15,23,12,25,45,15,18,20,14,15], 'Final Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration', 'Discount','Final Amount'], index = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
```

### Output

	Courses	Fee	Duration	Discount	Final Amount
count	10	10.000000	10	10.000000	10.000000
unique	10	NaN	6	NaN	NaN
top	Spark	NaN	30days	NaN	NaN
freq	1	NaN	3	NaN	NaN
mean	NaN	26363.200000	NaN	20.200000	21176.352000
std	NaN	7783.403152	NaN	9.647107	7410.831467
min	NaN	19632.000000	NaN	12.000000	13200.000000
25%	NaN	20500.000000	NaN	15.000000	16595.880000
50%	NaN	24500.000000	NaN	16.500000	18125.000000
75%	NaN	29000.000000	NaN	22.250000	24170.000000
max	NaN	45000.000000	NaN	45.000000	38250.000000

	Courses	Fee	Duration	Discount	Final Amount
count	10	10.000000	10	10.000000	10.000000
unique	10	NaN	6	NaN	NaN
top	Spark	NaN	30days	NaN	NaN
freq	1	NaN	3	NaN	NaN
mean	NaN	26363.200000	NaN	20.200000	21176.352000
std	NaN	7783.403152	NaN	9.647107	7410.831467
min	NaN	19632.000000	NaN	12.000000	13200.000000
25%	NaN	20500.000000	NaN	15.000000	16595.880000
50%	NaN	24500.000000	NaN	16.500000	18125.000000
75%	NaN	29000.000000	NaN	22.250000	24170.000000
max	NaN	45000.000000	NaN	45.000000	38250.000000

## Task -4

Check the 4<sup>th</sup> index observation with 'loc' slicing operator.

### Input

```
import pandas as pd
technologies =
{'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift",
"TypeScript","HTML","MATLAB"],
'Fee'
:[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
'Duration':['30days','40days','35days','40days','60days','30days','
45days','35days','30days','50days'],
'Discount':[15,23,12,25,45,15,18,20,14,15],'Final
Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200
]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration',
'Discount','Final Amount'], index = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
df.loc['r5']
```



```
import pandas as pd
technologies = {'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift","TypeScript","HTML","MATLAB"],
'Fee' :[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
'Duration':['30days','40days','35days','40days','60days','30days','45days','35days','30days','50days'],
'Discount':[15,23,12,25,45,15,18,20,14,15],'Final Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration', 'Discount','Final Amount'], index = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
df.loc['r5']
```

### Output

Courses	pandas
Fee	24000
Duration	60days
Discount	45
Final Amount	13200.0
Name: r5, dtype: object	



## Task -5

Check the null values in your 'df'.

**Input**

```
import pandas as pd

technologies =
{'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift",
,"TypeScript","HTML","MATLAB"],
  'Fee'
:[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
  'Duration':['30days','40days','35days','40days','60days','30days','
45days','35days','30days','50days'],
  'Discount':[15,23,12,25,45,15,18,20,14,15], 'Final
Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200
]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration',
'Discount','Final Amount'], index = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
df.loc['r5']
df.isnull()
```

```
import pandas as pd
technologies = {'Courses':["Spark","PySpark","Hadoop","Python","pandas","Rust","Swift","TypeScript","HTML","MATLAB"],
  'Fee' :[20000,25000,26000,22000,24000,45000,30000,20000,19632,32000],
  'Duration':['30days','40days','35days','40days','60days','30days','45days','35days','30days','50days'],
  'Discount':[15,23,12,25,45,15,18,20,14,15], 'Final Amount':[17000,19250,22880,16500,13200,38250,24600,16000,16883.52,27200]}
index_labels=['r1','r2','r3','r4','r5','r6','r7','r8','r9','r10']
df = pd.DataFrame(technologies, columns = ['Courses','Fee', 'Duration', 'Discount','Final Amount'], index = index_labels)
print(df)
df.info()
stats = df.describe(include = 'all')
print(stats)
df.loc['r5']
df.isnull()
```

## Output

	Courses	Fee	Duration	Discount	Final Amount
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