Task 3 - DataFrames

1. What is Dataframe?

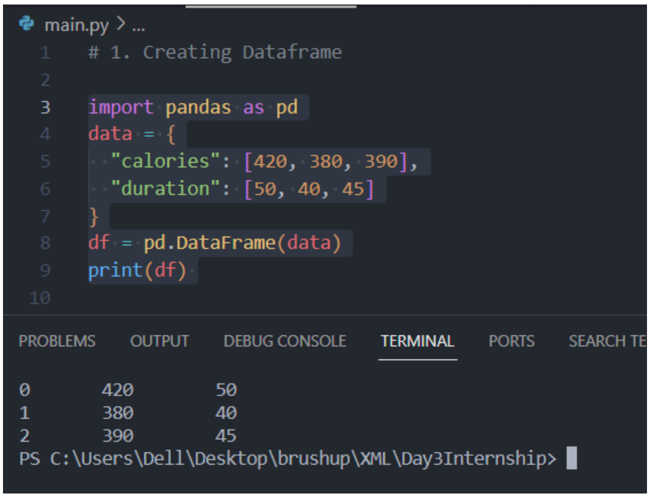
   A Pandas DataFrame is a 2 dimensional data structure, like a 2 dimensional array, or a table with rows and columns.

2. Applications of Dataframe

   Used to efficiently manage and analyze large datasets of financial transactions, stock market data, and economic indicators to make informed investment decisions

3. Creating Dataframe

```
import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data)
print(df)
```



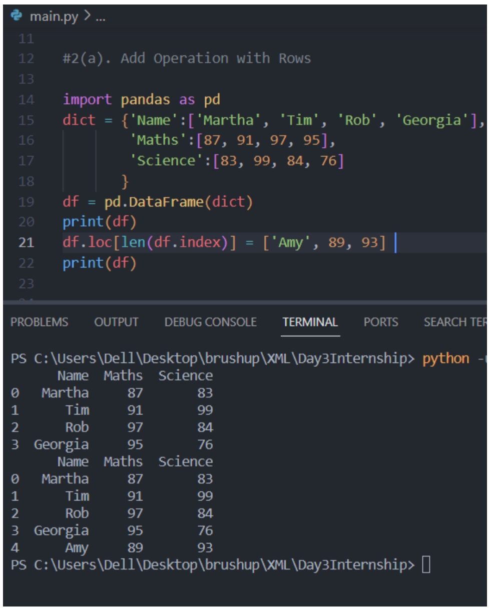
```
# 1. Creating Dataframe

import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data)
print(df)

```

```
0       420       50
1       380       40
2       390       45
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

4. Add, remove operation with rows and columns

   a. Add Operation with Rows

```
import pandas as pd
dict = {'Name':['Martha', 'Tim', 'Rob', 'Georgia'],
        'Maths':[87, 91, 97, 95],
        'Science':[83, 99, 84, 76]
       }
df = pd.DataFrame(dict)
print(df)
df.loc[len(df.index)] = ['Amy', 89, 93]
print(df)
```

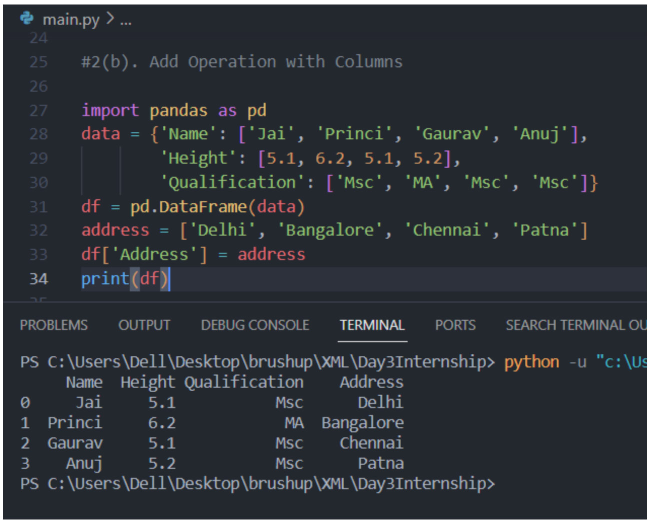


```
#2(a). Add Operation with Rows

import pandas as pd
dict = {'Name':['Martha', 'Tim', 'Rob', 'Georgia'],
        'Maths':[87, 91, 97, 95],
        'Science':[83, 99, 84, 76]
       }
df = pd.DataFrame(dict)
print(df)
df.loc[len(df.index)] = ['Amy', 89, 93]
print(df)
```

```
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship> python -
      Name  Maths  Science
0   Martha     87       83
1      Tim     91       99
2      Rob     97       84
3  Georgia     95       76
      Name  Maths  Science
0   Martha     87       83
1      Tim     91       99
2      Rob     97       84
3  Georgia     95       76
4      Amy     89       93
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

b. Add Operation with Columns

```
import pandas as pd
data = {'Name': ['Jai', 'Princi', 'Gaurav', 'Anuj'],
        'Height': [5.1, 6.2, 5.1, 5.2],
        'Qualification': ['Msc', 'MA', 'Msc', 'Msc']}
df = pd.DataFrame(data)
address = ['Delhi', 'Bangalore', 'Chennai', 'Patna']
df['Address'] = address
print(df)
```

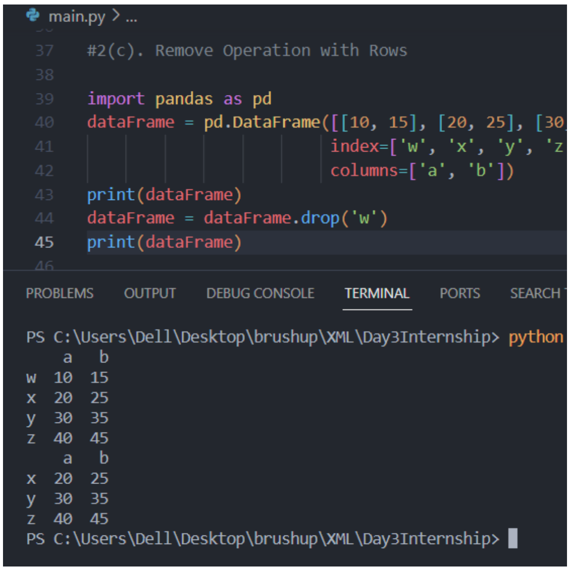


```
#2(b). Add Operation with Columns

import pandas as pd
data = {'Name': ['Jai', 'Princi', 'Gaurav', 'Anuj'],
        'Height': [5.1, 6.2, 5.1, 5.2],
        'Qualification': ['Msc', 'MA', 'Msc', 'Msc']}
df = pd.DataFrame(data)
address = ['Delhi', 'Bangalore', 'Chennai', 'Patna']
df['Address'] = address
print(df)
```

```
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship> python -u "c:\Us
     Name  Height Qualification    Address
0     Jai     5.1           Msc      Delhi
1  Princi     6.2            MA  Bangalore
2  Gaurav     5.1           Msc    Chennai
3    Anuj     5.2           Msc      Patna
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

c. Remove Operation with Rows

```
import pandas as pd
dataFrame = pd.DataFrame([[10, 15], [20, 25], [30, 35], [40, 45]],
                         index=['w', 'x', 'y', 'z'],
                         columns=['a', 'b'])
print(dataFrame)
dataFrame = dataFrame.drop('w')
print(dataFrame)
```

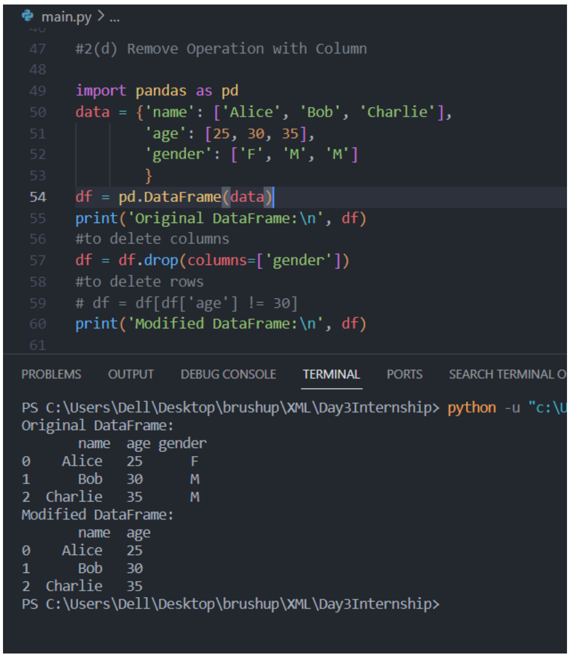


```
#2(c). Remove Operation with Rows

import pandas as pd
dataFrame = pd.DataFrame([[10, 15], [20, 25], [30
                         index=['w', 'x', 'y', 'z
                         columns=['a', 'b'])
print(dataFrame)
dataFrame = dataFrame.drop('w')
print(dataFrame)
```

```
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship> python
    a   b
w  10  15
x  20  25
y  30  35
z  40  45
    a   b
x  20  25
y  30  35
z  40  45
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

d. Remove Operation with Column

```
import pandas as pd
data = {'name': ['Alice', 'Bob', 'Charlie'],
        'age': [25, 30, 35],
        'gender': ['F', 'M', 'M']
        }
df = pd.DataFrame(data)
print('Original DataFrame:\n', df)
#to delete columns
df = df.drop(columns=['gender'])
#to delete rows
# df = df[df['age'] != 30]
print('Modified DataFrame:\n', df)
```

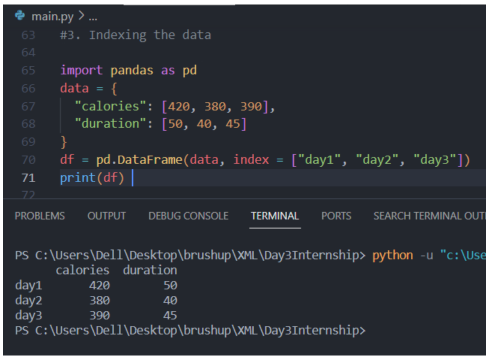


```python
#2(d) Remove Operation with Column

import pandas as pd
data = {'name': ['Alice', 'Bob', 'Charlie'],
        'age': [25, 30, 35],
        'gender': ['F', 'M', 'M']
        }
df = pd.DataFrame(data)
print('Original DataFrame:\n', df)
#to delete columns
df = df.drop(columns=['gender'])
#to delete rows
# df = df[df['age'] != 30]
print('Modified DataFrame:\n', df)
```

```
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship> python -u "c:\U
Original DataFrame:
       name  age gender
0    Alice   25      F
1      Bob   30      M
2  Charlie   35      M
Modified DataFrame:
       name  age
0    Alice   25
1      Bob   30
2  Charlie   35
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

5. Indexing the data

```python
import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data, index = ["day1", "day2", "day3"])
print(df)
```



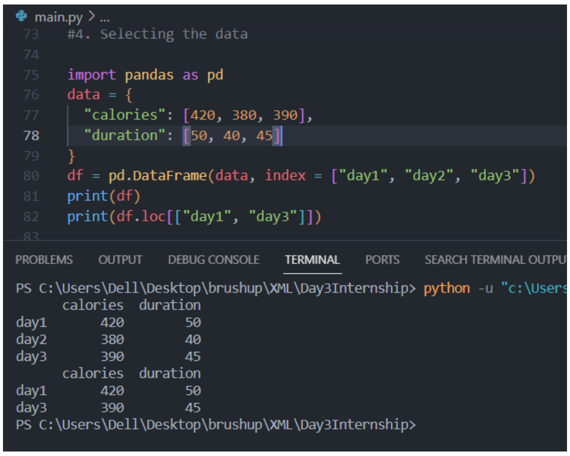
```python
#3. Indexing the data

import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data, index = ["day1", "day2", "day3"])
print(df)
```

```
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship> python -u "c:\Use
      calories  duration
day1       420        50
day2       380        40
day3       390        45
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

6. Selecting the data

```python
import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data, index = ["day1", "day2", "day3"])
```

```
print(df)
print(df.loc[["day1", "day3"]])
```

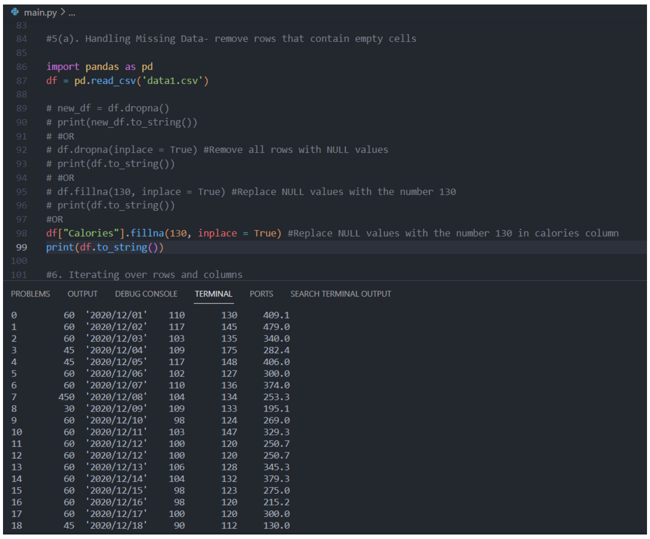


```
#4. Selecting the data

import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
df = pd.DataFrame(data, index = ["day1", "day2", "day3"])
print(df)
print(df.loc[["day1", "day3"]])
```

```
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship> python -u "c:\Users
      calories  duration
day1       420        50
day2       380        40
day3       390        45
      calories  duration
day1       420        50
day3       390        45
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

7. Handling Missing Data

```
import pandas as pd
df = pd.read_csv('data1.csv')

new_df = df.dropna()
print(new_df.to_string())
#OR
df.dropna(inplace = True) #Remove all rows with NULL values
print(df.to_string())
#OR
df.fillna(130, inplace = True) #Replace NULL values with the number 130
print(df.to_string())
#OR
df["Calories"].fillna(130, inplace = True) #Replace NULL values with the number 130 in calories column
print(df.to_string())
```

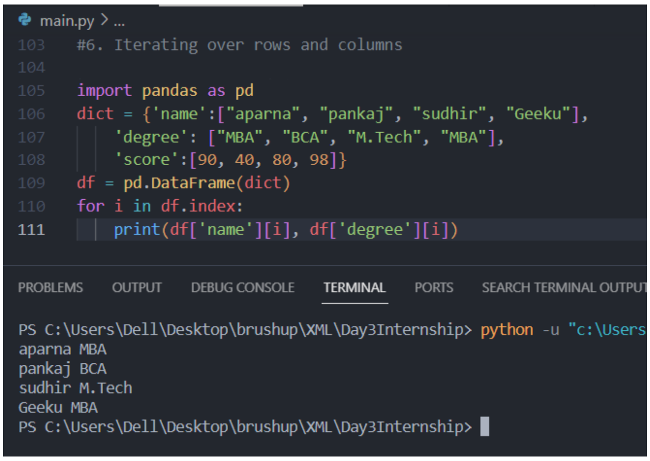


```
#5(a). Handling Missing Data- remove rows that contain empty cells

import pandas as pd
df = pd.read_csv('data1.csv')

# new_df = df.dropna()
# print(new_df.to_string())
# #OR
# df.dropna(inplace = True) #Remove all rows with NULL values
# print(df.to_string())
# #OR
# df.fillna(130, inplace = True) #Replace NULL values with the number 130
# print(df.to_string())
#OR
df["Calories"].fillna(130, inplace = True) #Replace NULL values with the number 130 in calories column
print(df.to_string())

#6. Iterating over rows and columns
```

```
0       60  '2020/12/01'   110      130    409.1
1       60  '2020/12/02'   117      145    479.0
2       60  '2020/12/03'   103      135    340.0
3       45  '2020/12/04'   109      175    282.4
4       45  '2020/12/05'   117      148    406.0
5       60  '2020/12/06'   102      127    300.0
6       60  '2020/12/07'   110      136    374.0
7      450  '2020/12/08'   104      134    253.3
8       30  '2020/12/09'   109      133    195.1
9       60  '2020/12/10'    98      124    269.0
10      60  '2020/12/11'   103      147    329.3
11      60  '2020/12/12'   100      120    250.7
12      60  '2020/12/12'   100      120    250.7
13      60  '2020/12/13'   106      128    345.3
14      60  '2020/12/14'   104      132    379.3
15      60  '2020/12/15'    98      123    275.0
16      60  '2020/12/16'    98      120    215.2
17      60  '2020/12/17'   100      120    300.0
18      45  '2020/12/18'    90      112    130.0
```

8. Iterating over rows and columns

```
import pandas as pd
dict = {'name':["aparna", "pankaj", "sudhir", "Geeku"],
    'degree': ["MBA", "BCA", "M.Tech", "MBA"],
    'score':[90, 40, 80, 98]}
df = pd.DataFrame(dict)
for i in df.index:
    print(df['name'][i], df['degree'][i])
```



```
#6. Iterating over rows and columns

import pandas as pd
dict = {'name':["aparna", "pankaj", "sudhir", "Geeku"],
    'degree': ["MBA", "BCA", "M.Tech", "MBA"],
    'score':[90, 40, 80, 98]}
df = pd.DataFrame(dict)
for i in df.index:
    print(df['name'][i], df['degree'][i])
```

```
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship> python -u "c:\Users
aparna MBA
pankaj BCA
sudhir M.Tech
Geeku MBA
PS C:\Users\Dell\Desktop\brushup\XML\Day3Internship>
```

Team 2

Abhijna

Anusha S

Varshaa Shetty

Shalini L