Name: Siddhi Parekh
Reg No. 221071047
Batch : C
SY COMPS

Experiment No. : 10

AIM:
Write an application to analysis the data using NumPy, Pandas libraries

THEORY:
Pandas is an open-source library that is built on top of NumPy library. It is a Python package that offers various data structures and operations for manipulating numerical data and time series. It is mainly popular for importing and analyzing data much easier. Pandas is fast and it has high-performance & productivity for users.

Numpy is a general-purpose array-processing package. It provides tools for working with these arrays. It is the fundamental package for scientific computing with Python.

CODE:

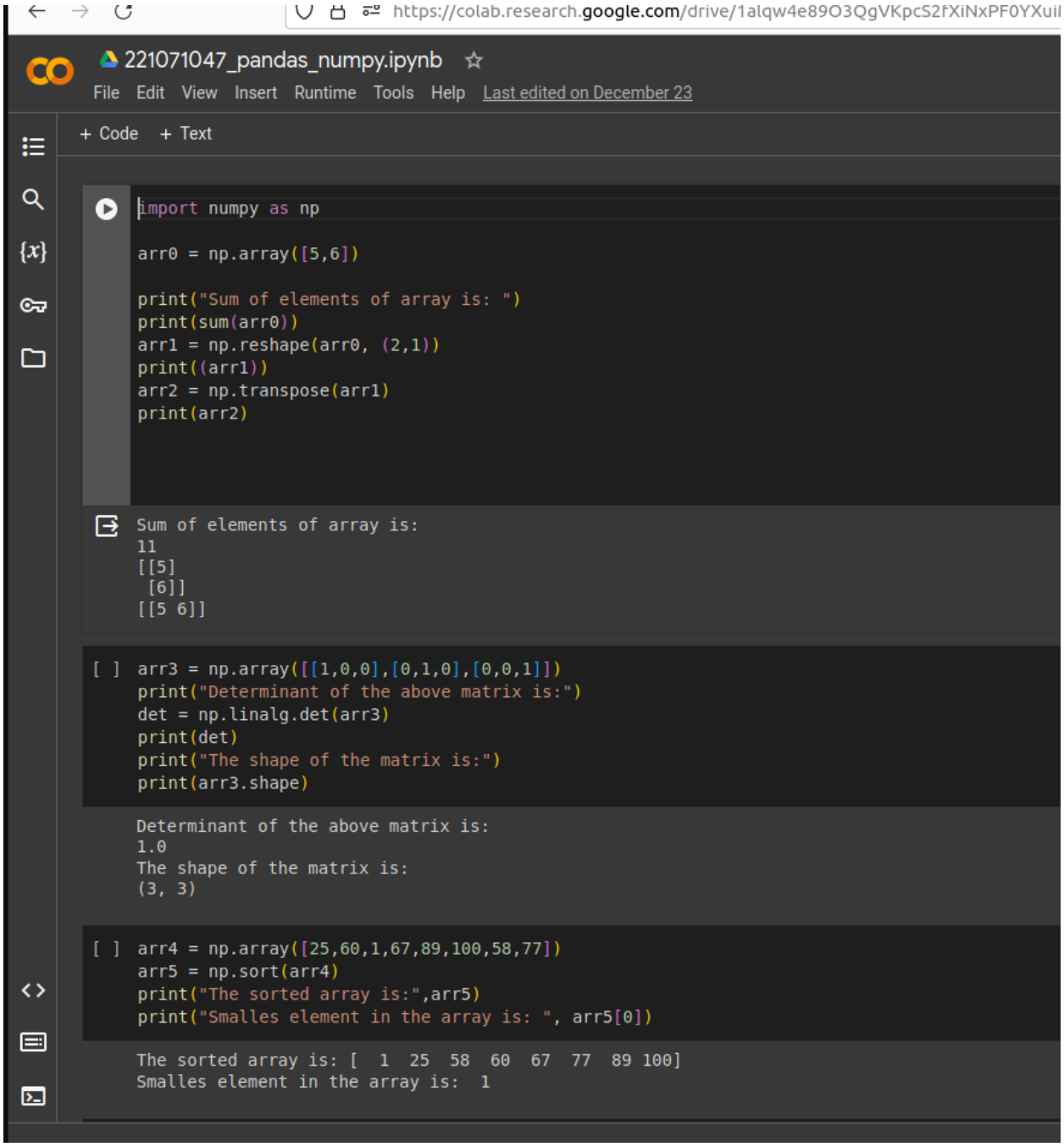
```
# -*- coding: utf-8 -*-

"""221071047_pandas_numpy.ipynb



Automatically generated by Colaboratory.



Original file is located at
```

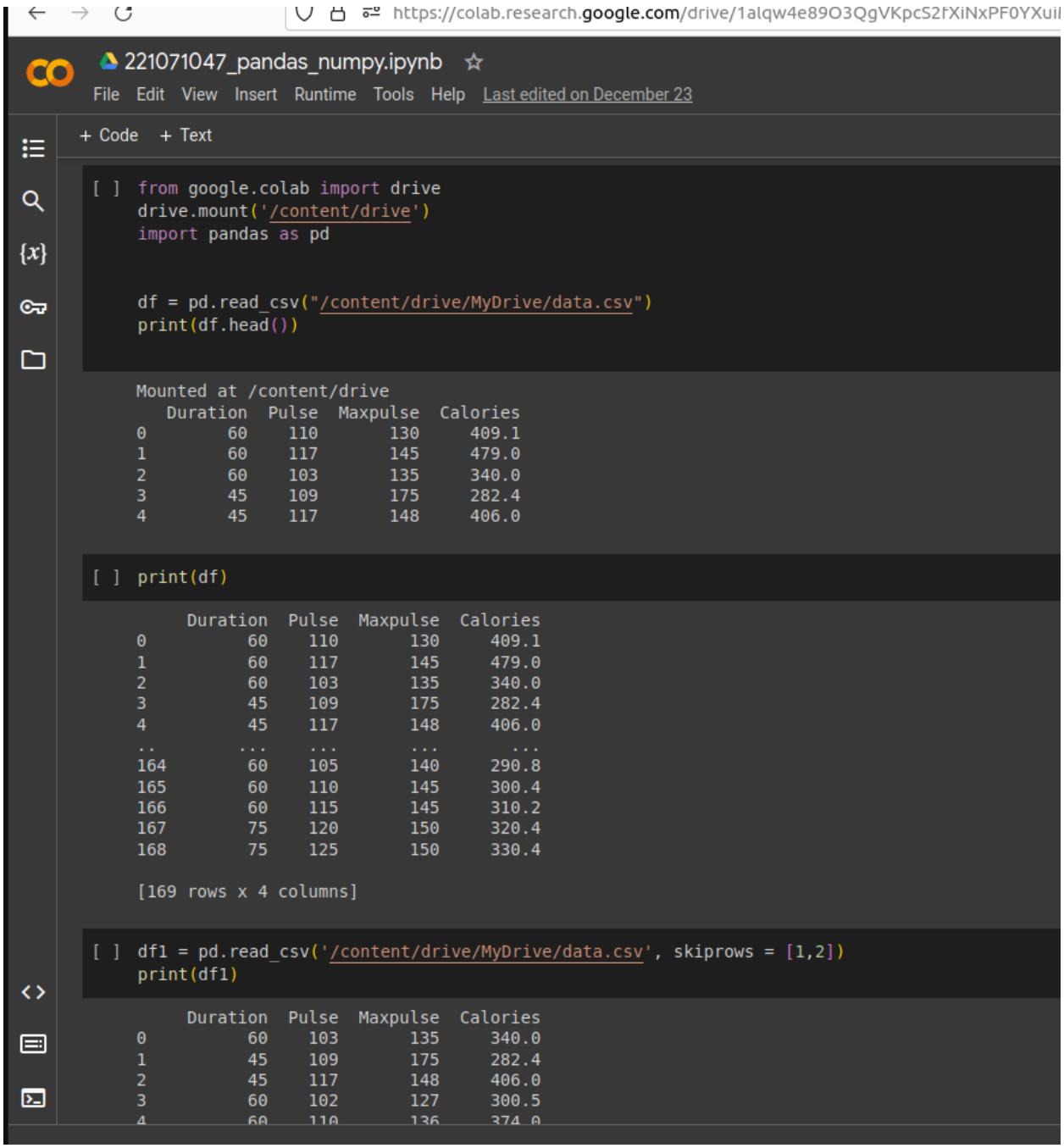
```
https://colab.research.google.com/drive/1alqw4e89O3QgVKpcS2fXiNxPF0YXuiMb

"""



import numpy as np



arr0 = np.array([5,6])



print("Sum of elements of array is: ")

print(sum(arr0))

arr1 = np.reshape(arr0, (2,1))

print((arr1))

arr2 = np.transpose(arr1)

print(arr2)



arr3 = np.array([[1,0,0],[0,1,0],[0,0,1]])

print("Determinant of the above matrix is:")

det = np.linalg.det(arr3)

print(det)

print("The shape of the matrix is:")
```

```
print(arr3.shape)



arr4 = np.array([25,60,1,67,89,100,58,77])

arr5 = np.sort(arr4)

print("The sorted array is:",arr5)

print("Smalles element in the array is: ", arr5[0])



from google.colab import drive

drive.mount('/content/drive')

import pandas as pd




df = pd.read_csv("/content/drive/MyDrive/data.csv")

print(df.head())



print(df)



df1 = pd.read_csv('/content/drive/MyDrive/data.csv', skiprows = [1,2])

print(df1)
```

```
df.mean(axis = 0)



df.mean(axis = 1)



new_row = {'Duration': '7', 'Pulse': '7', 'Maxpulse':'7', 'Calories': '7'}

df.loc[len(df)] = new_row

df = df.append(new_row , ignore_index=True)



print(df)



cell = df.iloc[169]['Duration']

print(cell)

cell1 = df.iloc[0]['Duration']

print(cell1)



df= pd.read_csv('/content/drive/MyDrive/data.csv')

df.drop('Duration',axis=1,inplace=True)

print(df)
```

OUTPUT:


https://colab.research.google.com/drive/1alqw4e89O3QgVKpcS2fXiNxPF0YXuil
221071047_pandas_numpy.ipynb
File Edit View Insert Runtime Tools Help Last edited on December 23
+ Code + Text
import numpy as np
arr0 = np.array([5,6])
print("Sum of elements of array is: ")
print(sum(arr0))
arr1 = np.reshape(arr0, (2,1))
print((arr1))
arr2 = np.transpose(arr1)
print(arr2)
Sum of elements of array is:
11
[[5]
[6]]
[[5 6]]
arr3 = np.array([[1,0,0],[0,1,0],[0,0,1]])
print("Determinant of the above matrix is:")
det = np.linalg.det(arr3)
print(det)
print("The shape of the matrix is:")
print(arr3.shape)
Determinant of the above matrix is:
1.0
The shape of the matrix is:
(3, 3)
arr4 = np.array([25,60,1,67,89,100,58,77])
arr5 = np.sort(arr4)
print("The sorted array is:",arr5)
print("Smalles element in the array is: ", arr5[0])
The sorted array is: [1 25 58 60 67 77 89 100]
Smalles element in the array is: 1


https://colab.research.google.com/drive/1alqw4e89O3QgVKpcS2fXiNxPF0YXuil
221071047_pandas_numpy.ipynb
File Edit View Insert Runtime Tools Help Last edited on December 23
+ Code + Text
[ ] from google.colab import drive
drive.mount('/content/drive')
import pandas as pd
df = pd.read_csv("/content/drive/MyDrive/data.csv")
print(df.head())
Mounted at /content/drive
Duration Pulse Maxpulse Calories
0 60 110 130 409.1
1 60 117 145 479.0
2 60 103 135 340.0
3 45 109 175 282.4
4 45 117 148 406.0
[ ] print(df)
Duration Pulse Maxpulse Calories
0 60 110 130 409.1
1 60 117 145 479.0
2 60 103 135 340.0
3 45 109 175 282.4
4 45 117 148 406.0
.. ..
164 60 105 140 290.8
165 60 110 145 300.4
166 60 115 145 310.2
167 75 120 150 320.4
168 75 125 150 330.4
[169 rows x 4 columns]
[ ] df1 = pd.read_csv('/content/drive/MyDrive/data.csv', skiprows = [1,2])
print(df1)
Duration Pulse Maxpulse Calories
0 60 103 135 340.0
1 45 109 175 282.4
2 45 117 148 406.0
3 60 102 127 300.5

https://colab.research.google.com/drive/1alqw4e89O3QgVKpcS2fXiNxPF0YXuil

221071047_pandas_numpy.ipynb ☆

File Edit View Insert Runtime Tools Help Last edited on December 23

+ Code + Text

```
[ ] df1 = pd.read_csv('/content/drive/MyDrive/data.csv', skiprows = [1,2])
    print(df1)
```

```
     Duration  Pulse  Maxpulse  Calories
0          60    103       135     340.0
1          45    109       175     282.4
2          45    117       148     406.0
3          60    102       127     300.5
4          60    110       136     374.0
..        ...    ...       ...       ...
162        60    105       140     290.8
163        60    110       145     300.4
164        60    115       145     310.2
165        75    120       150     320.4
166        75    125       150     330.4

[167 rows x 4 columns]
```

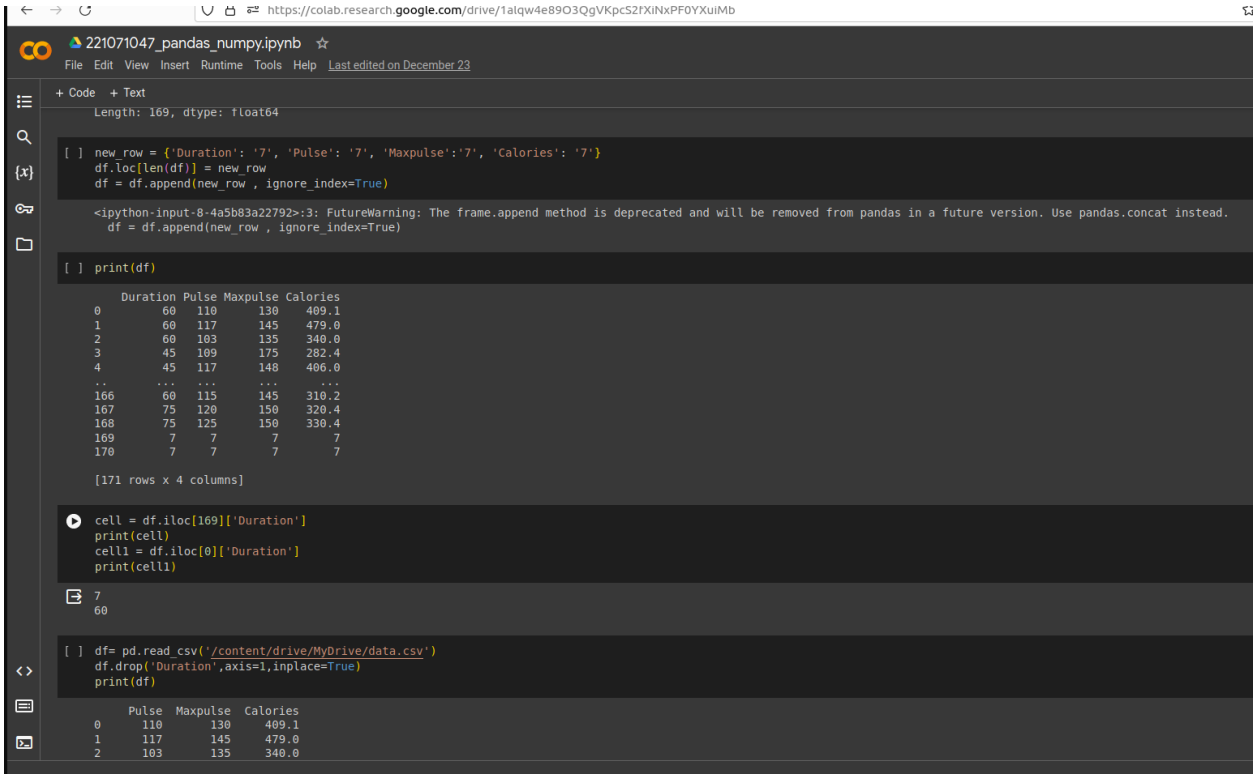
```
df.mean(axis = 0)
```

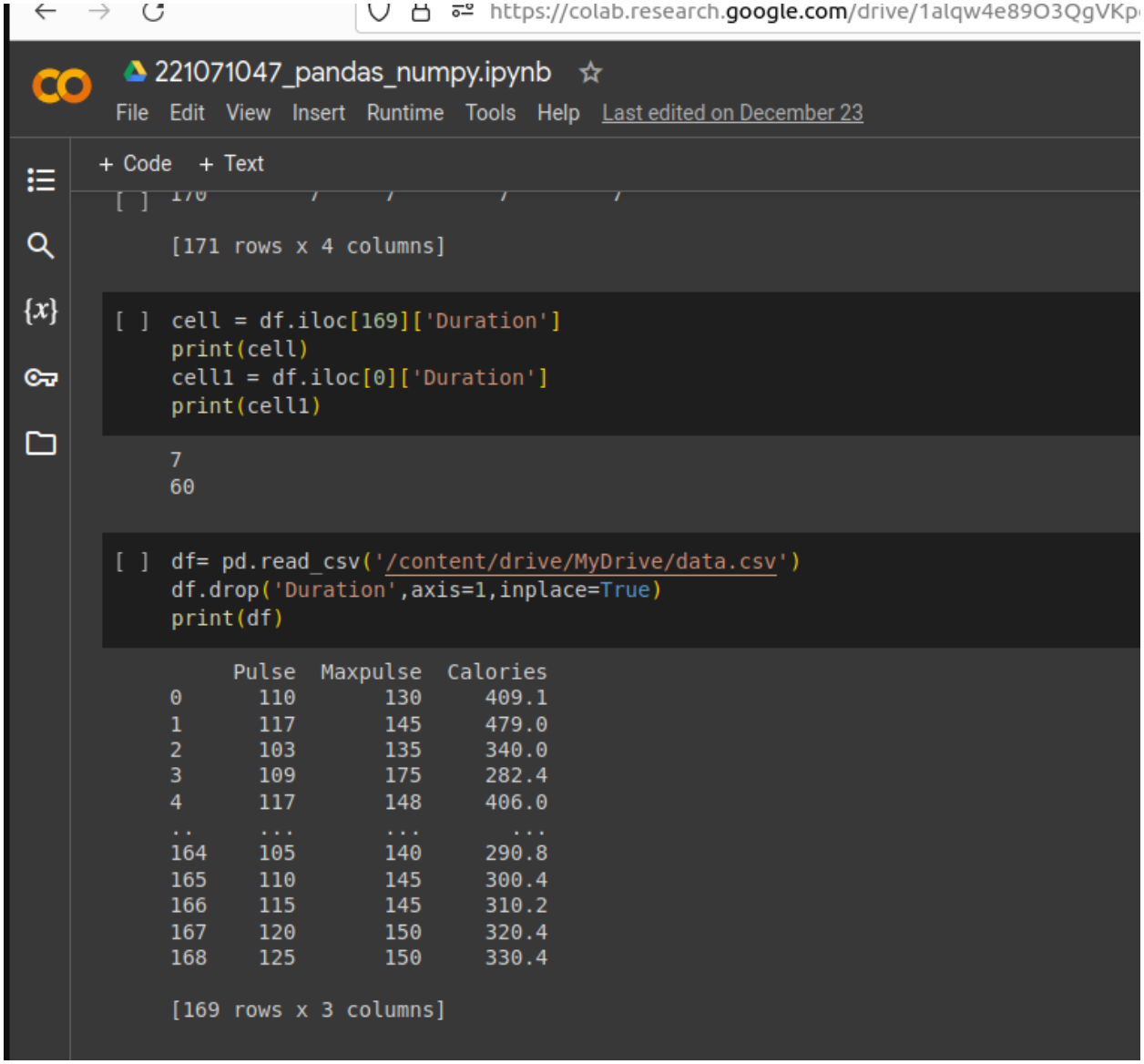
```
Duration     63.846154
Pulse       107.461538
Maxpulse    134.047337
Calories    375.800000
dtype: float64
```

```
[ ] df.mean(axis = 1)
```

```
0      177.275
1      200.250
2      159.500
3      152.850
4      179.000
        ...
164    148.950
165    153.850
166    157.550
167    166.350
168    170.100
Length: 169, dtype: float64
```

```
[ ] new_row = {'Duration': '7', 'Pulse': '7', 'Maxpulse':'7', 'Calories': '7'}
    df.loc[len(df)] = new_row
```


https://colab.research.google.com/drive/1alqw4e89O3QgVKpcS2fXiNxPF0YXuiMb
221071047_pandas_numpy.ipynb
File Edit View Insert Runtime Tools Help Last edited on December 23
+ Code + Text
Length: 169, dtype: float64
new_row = {'Duration': '7', 'Pulse': '7', 'Maxpulse':'7', 'Calories': '7'}
df.loc[len(df)] = new_row
df = df.append(new_row , ignore_index=True)
<ipython-input-8-4a5b83a22792>:3: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead.
df = df.append(new_row , ignore_index=True)
print(df)
Duration Pulse Maxpulse Calories
0 60 110 130 409.1
1 60 117 145 479.0
2 60 103 135 340.0
3 45 109 175 282.4
4 45 117 148 406.0
... ...
166 60 115 145 310.2
167 75 120 150 320.4
168 75 125 150 330.4
169 7 7 7 7
170 7 7 7 7
[171 rows x 4 columns]
cell = df.iloc[169]['Duration']
print(cell)
cell1 = df.iloc[0]['Duration']
print(cell1)
7
60
df= pd.read_csv('/content/drive/MyDrive/data.csv')
df.drop('Duration',axis=1,inplace=True)
print(df)
Pulse Maxpulse Calories
0 110 130 409.1
1 117 145 479.0
2 103 135 340.0



CONCLUSION:

In this experiment, we learnt about numpy and pandas libraries and how to use them to perform various operations on a given set of data.