**Name: Vishnu Vijayakumar**

**Roll No:53**

**Batch:MCA-B**

**Date:01-09-2022**

**DATA SCIENCE LAB**

**Experiment No.: 3**

**Aim**

 Programs to handle data using pandas.

**Question**

**Q1 - Pandas  Series**

1. How to create Series with nd array
2. How to create Series with Mutable index
3. Creating a series from a Dictionary
4. Print all the values of the Series by multiplying them by 2.
5. Print Square of all the values of the series.
6. Print all the values of the Series that are greater than2
7. Addition of two series
8. Print the first and last 5 elements of a series
9. Print the values from index 0 to 5
10. Selection Using loc, iloc index label
11. Retrieve subsets of data using slicing

**Q2 Dataframe**

1. create Dataframe From Series
2. DataFrame from List of Dictionaries
3. Display the first 5 rows of data frame
4. Select the last two columns of the data frame
5. Add two data frames
6. Demonstrate deletion, and renaming of columns
7. Demonstrate concat, Merge operations in data frame
8. Write a Pandas program to join the two given dataframes along rows and assign all data

**Test Data:**

student\_data1:

  student\_id              name  marks

0         S1  Danniella Fenton    200

1         S2      Ryder Storey    210

2         S3      Bryce Jensen    190

3         S4         Ed Bernal    222

4         S5       Kwame Morin    199

student\_data2:

  student\_id              name  marks

0         S4  Scarlette Fisher    201

1         S5  Carla Williamson    200

2         S6       Dante Morse    198

3         S7    Kaiser William    219

4         S8   Madeeha Preston    201

**Procedure and Output**

#1.How to create Series with nd array

import pandas as pd

import numpy as np

arr=np.array([10,15,20,25])

s = pd.Series(arr)

print(s)

**out put**

0 10

1 15

2 20

3 25

dtype: int64

2.How to create Series with Mutable index

import pandas as pd

s = pd.Series(50, index =[1,2,3,4])

print(s)

Output

1 50

2 50

3 50

4 50

dtype: int64

3Creating a series from a Dictionary

import pandas as pd

d = {'Name': 'MS Dhoni','Team':'India','Runs':17500}

s = pd.Series(d)

print(s)

Output  
Name MS Dhoni

Team India

Runs 17500

dtype: object

4. Print all the values of the Series by multiplying them by 2

import pandas as pd

s=pd.Series([1,2,3,4,5])

print('Multiply all the values in a series by 2')

print('------------')

print(s\*2)

Output

Multiply all the values in a series by 2

------------

0 2

1 4

2 6

3 8

4 10

dtype: int64

5.Print Square of all the values of the series

print('Find all the squres of all the values in a series')

print('-------------')

print(s\*\*2)

Output

Find all the squres of all the values in a series

-------------

1 1

2 4

3 9

4 16

5 25

dtype: int64

6 Print all the values of the Series that are greater than2

print('Find all the values that are greater than 2')

print('-------------')

print(s[s>2])

Output

Find all the values that are greater than 2

-------------

2 3

3 4

4 5

dtype: int64

7.Addition of two series

import pandas as pd

s2=pd.Series([10,20,30,40,50],index=['a','b','c','d','e'])

s3=pd.Series([17,25,30,40,50],index=['a','b','c','d','e'])

print('To add  the series1 and series2')

print('------------')

print(s2+s3)

print(s2.add(s1,fill\_value=0))

Output

To add the series1 and series2

------------

a 27

b 45

c 60

d 80

e 100

8. Print the first and last 5 elements of a series

import pandas as pd

import numpy as np

arr = np.array([10,20,30,40,50,60,70,80])

s=pd.Series(arr)

print(s.head())

Output

0 10

1 20

2 30

3 40

4 50

dtype: int64

9. Print the values from index 0 to 5

import pandas as pd

import numpy as np

arr = np.array([10,20,30,40,50,60,70,80])

s=pd.Series(arr)

print(s.head())

print(s.head(3))

Output

0 10

1 20

2 30

3 40

4 50

dtype: int64

0 10

1 20

2 30

dtype: int64

CodeText

10.Selection Using loc, iloc index label

import pandas as pd

import numpy as np

arr=np.array([12,17,18,25,33,45,7,10,3])

s=pd.Series(arr)

print(s.loc[:6])

print(s.iloc[5:6])

Output

0 12

1 17

2 18

3 25

4 33

5 45

6 7

dtype: int64

5 45

dtype: int64

11.Retrieve subsets of data using slicing

import pandas as pd

import numpy as np

arr=np.array([10,20,30,40,50,60,70])

s=pd.Series(arr,index=['A','B','C','D','E','F','G'])

print(s)

print(s[1:5:2])

print(s[0:6:2])

Output

A 10

B 20

C 30

D 40

E 50

F 60

G 70

dtype: int64

B 20

D 40

dtype: int64

A 10

C 30

E 50

dtype: int64

1.create Dataframe From Series

import pandas as pd

name=pd.Series(['MS Dhoni','Virat Kohli'])

team=pd.Series(['CSK','RCB'])

dic={'Name':name ,'Team':team}

df=pd.DataFrame(dic)

print(df)

Output

Name Team

0 MS Dhoni CSK

1 Virat Kohli RCB

2 DataFrame from List of Dictionaries

import pandas as pd

l=[{'Name':'Vishnu','SirName':'Vijayakumar'}]

dfl=pd.DataFrame(l)

print(dfl)

Output

Name SirName

0 Vishnu Vijayakumar

3.Display the first 5 rows of data frame

import pandas as pd

l=[{'Name':'Vishnu','SirName':'Vijayakumar'},

   {'Name':'Sachin','SirName':'Tendulkar'},

   {'Name':'Virat','SirName':'Kohli'},

    {'Name':'MS','SirName':'Dhoni'},

    {'Name':'Adam','SirName':'Gilchrist'}]

dfl = pd.DataFrame(l)

print(dfl)

for(row\_index,row\_value) in dfl.iterrows():

   print('\nrow\_index is::',row\_index)

   print('row index is')

   print(row\_index)

Output

Name SirName

0 Vishnu Vijayakumar

1 Sachin Tendulkar

2 Virat Kohli

3 MS Dhoni

4 Adam Gilchrist

row\_index is:: 0

row index is

0

row\_index is:: 1

row index is

1

row\_index is:: 2

row index is

2

row\_index is:: 3

row index is

3

row\_index is:: 4

row index is

4

4.Select the last two columns of the data frame

import pandas as pd

empdata = {'empid':[1,2,3,4,5,6], 'ename':['Virat','Sachin','Raina','Dhoni','Rohit','Yuvi']}

df=pd.DataFrame(empdata)

print(df)

df.loc[0:5]

print(df.tail(2))

Output

empid ename

0 1 Virat

1 2 Sachin

2 3 Raina

3 4 Dhoni

4 5 Rohit

5 6 Yuvi

empid ename

4 5 Rohit

5 6 Yuvi

5.Add to DataFrames

import pandas as pd

s= pd.Series([5,50,20,70])

df= pd.DataFrame(s)

df.columns= ['List1']

df['List2']= 10

df['SumList']=df['List1']+df['List2']

print(df)

Output

List1 List2 SumList

0 5 10 15

1 50 10 60

2 20 10 30

3 70 10 80

CodeText

6. Demonstrate deletion, and renaming of columns

 import pandas as pd dic1= {'id':['1','2','3','4','5'],'value1':['A','C','E','G','I'],'value2':['B','D','F','H','J']} dic2= {'id':['2','3','6','7','8'],'value1':['K','M','O','Q','S'],'value2':['L','N','P','R','T']} dic3= {'id':['1','2','3','4','5','7','8','9','10','11'],'value3':[12,13,14,15,16,17,15,12,13,23]} df1=pd.DataFrame(dic1) df2=pd.DataFrame(dic2) df3=pd.concat([df1,df2]) df4=pd.DataFrame(dic3) df5=pd.merge(df3,df4,on='id') print(df5)

id value1 value2 value3 0 1 A B 12 1 2 C D 13 2 2 K L 13 3 3 E F 14 4 3 M N 14 5 4 G H 15 6 5 I J 16 7 7 Q R 17 8 8 S T 15

7 Demonstrate concat, Merge operations in data frame

import pandas as pd

s= pd.Series([10,20,30,40])

df=pd.DataFrame(s)

df.columns=['List1']

df['List2']=40

df1=df.drop('List2',axis=1)

df2=df.drop(index=[2,3],axis=0)

print(df)

print(" After deletion::")

print(df1)

print (" After row deletion::")

print(df2)

out put

List1  List2

0     10     40

1     20     40

2     30     40

3     40     40

 After deletion::

   List1

0     10

1     20

2     30

3     40

 After row deletion::

   List1  List2

0     10     40

1. 20     40

8.Write a Pandas program to join the two given dataframes along rows and assign all data

**Test Data:**

student\_data1:

  student\_id              name  marks

0         S1  Danniella Fenton    200

1         S2      Ryder Storey    210

2         S3      Bryce Jensen    190

3         S4         Ed Bernal    222

4         S5       Kwame Morin    199

student\_data2:

  student\_id              name  marks

0         S4  Scarlette Fisher    201

1         S5  Carla Williamson    200

2         S6       Dante Morse    198

3         S7    Kaiser William    219

4         S8   Madeeha Preston    201

Out put

Original DataFrames:

  student\_id              name  marks

0         S1  Danniella Fenton    200

1         S2      Ryder Storey    210

2         S3      Bryce Jensen    190

3         S4         Ed Bernal    222

4         S5       Kwame Morin    199

-------------------------------------

  student\_id              name  marks

0         S4  Scarlette Fisher    201

1         S5  Carla Williamson    200

2         S6       Dante Morse    198

3         S7    Kaiser William    219

4         S8   Madeeha Preston    201

Join the said two dataframes along rows:

  student\_id              name  marks

0         S1  Danniella Fenton    200

1         S2      Ryder Storey    210

2         S3      Bryce Jensen    190

3         S4         Ed Bernal    222

4         S5       Kwame Morin    199

0         S4  Scarlette Fisher    201

1         S5  Carla Williamson    200

2         S6       Dante Morse    198

3         S7    Kaiser William    219

4         S8   Madeeha Preston    201