

## Lab 17

**Name :-** Aryan Dilipbhai Langhanoja

**Date :-** 26-09-2023

**Enrollment No :-** 92200133030

**CO1: To write, test, and debug simple Python programs**

**CO2: To implement Python programs with conditional, loops and functions**

**Task 1:- Creating a Series by passing a list of values, letting pandas create a default RangeIndex.**

**Python Code:**

```
import numpy as np
import pandas as pd
s = pd.Series([1, 3, 5, np.nan, 6, 8])
print(s)
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe
0    1.0
1    3.0
2    5.0
3    NaN
4    6.0
5    8.0
dtype: float64
```

**Task 2:- Creating a DataFrame by passing a NumPy array with a datetime index using date\_range() and labeled columns**

**Python Code:**

```
import numpy as np
import pandas as pd
dates = pd.date_range("20130101", periods=6)
df = pd.DataFrame(np.random.randn(6, 4), index=dates, columns=list("ABCD"))
print(df)
```

**Output:**

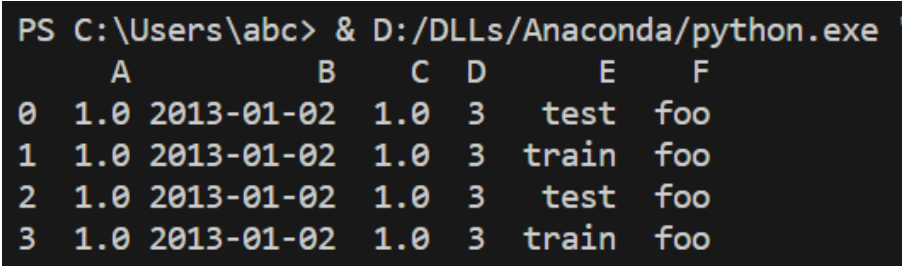
```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe "d:/Aryan/Semester - 3
          A          B          C          D
2013-01-01  1.284167  0.352124  1.528883  0.408306
2013-01-02 -0.022408 -0.713907 -0.581931 -0.546099
2013-01-03  0.141676 -0.535073 -0.656605 -0.062410
2013-01-04  0.017682  1.436127  0.062835  0.270190
2013-01-05  1.693776  1.248533 -0.077483 -0.120428
2013-01-06  0.232976 -2.050982  1.044301  0.511698
```

**Task 3:- Creating a DataFrame by passing a dictionary of objects where the keys are the column labels and the values are the column values.**

**Python Code:**

```
import numpy as np
import pandas as pd
df2 = pd.DataFrame(
    {
        "A": 1.0,
        "B": pd.Timestamp("20130102"),
        "C": pd.Series(1, index=list(range(4)), dtype="float32"),
        "D": np.array([3] * 4, dtype="int32"),
        "E": pd.Categorical(["test", "train", "test", "train"]),
        "F": "foo",
    }
)
print(df2)
```

**Output:**



	A	B	C	D	E	F
0	1.0	2013-01-02	1.0	3	test	foo
1	1.0	2013-01-02	1.0	3	train	foo
2	1.0	2013-01-02	1.0	3	test	foo
3	1.0	2013-01-02	1.0	3	train	foo

**Task 4:- Checking The DataTypes of Each Column In Dataframe**

**Python Code:**

```
import numpy as np
import pandas as pd
df2 = pd.DataFrame(
    {
        "A": 1.0,
        "B": pd.Timestamp("20130102"),
        "C": pd.Series(1, index=list(range(4)), dtype="float32"),
        "D": np.array([3] * 4, dtype="int32"),
        "E": pd.Categorical(["test", "train", "test", "train"]),
        "F": "foo",
    }
)
print(df2.dtypes)
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe
A          float64
B    datetime64[ns]
C          float32
D          int32
E        category
F          object
dtype: object
```

**Task 5:- Implement head Function****Python Code:**

```
import numpy as np
import pandas as pd
df2 = pd.DataFrame(
    {
        "A": 1.0,
        "B": pd.Timestamp("20130102"),
        "C": pd.Series(1, index=list(range(4)), dtype="float32"),
        "D": np.array([3] * 4, dtype="int32"),
        "E": pd.Categorical(["test", "train", "test", "train"]),
        "F": "foo",
    }
)
print(df2.head())
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe
   A          B    C  D    E    F
0  1.0 2013-01-02  1.0  3  test  foo
1  1.0 2013-01-02  1.0  3  train foo
2  1.0 2013-01-02  1.0  3  test  foo
3  1.0 2013-01-02  1.0  3  train foo
```

**Task 6:- Implement tail Function****Python Code:**

```
import numpy as np
import pandas as pd
df2 = pd.DataFrame(
    {
        "A": 1.0,
        "B": pd.Timestamp("20130102"),
```

```
"C": pd.Series(1, index=list(range(4)), dtype="float32"),
"D": np.array([3] * 4, dtype="int32"),
"E": pd.Categorical(["test", "train", "test", "train"]),
"F": "foo",
}
)
print(df2.tail(3))
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe
      A      B      C      D      E      F
1  1.0 2013-01-02  1.0  3  train  foo
2  1.0 2013-01-02  1.0  3   test  foo
3  1.0 2013-01-02  1.0  3  train  foo
```

**Task 7:- Implement index Function****Python Code:**

```
import numpy as np
import pandas as pd
df2 = pd.DataFrame(
{
    "A": 1.0,
    "B": pd.Timestamp("20130102"),
    "C": pd.Series(1, index=list(range(4)), dtype="float32"),
    "D": np.array([3] * 4, dtype="int32"),
    "E": pd.Categorical(["test", "train", "test", "train"]),
    "F": "foo",
}
)

print(df2.index)
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe
Int64Index([0, 1, 2, 3], dtype='int64')
```

**Task 8:- Implement columns Function****Python Code:**

```
import numpy as np
import pandas as pd
df2 = pd.DataFrame(
{
    "A": 1.0,
```

```
"B": pd.Timestamp("20130102"),
"C": pd.Series(1, index=list(range(4)), dtype="float32"),
"D": np.array([3] * 4, dtype="int32"),
"E": pd.Categorical(["test", "train", "test", "train"]),
"F": "foo",
}
)
print(df2.columns)
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe "d:/Aryan/Semester - 3,
Index(['A', 'B', 'C', 'D', 'E', 'F'], dtype='object')
```

## Post Lab

**Task 1:- Converting Dataframe into Numpy Array****Python Code:**

```
import numpy as np
import pandas as pd
df2 = pd.DataFrame(
{
    "A": 1.0,
    "B": pd.Timestamp("20130102"),
    "C": pd.Series(1, index=list(range(4)), dtype="float32"),
    "D": np.array([3] * 4, dtype="int32"),
    "E": pd.Categorical(["test", "train", "test", "train"]),
    "F": "foo",
}
)

print(df2.to_numpy())
```

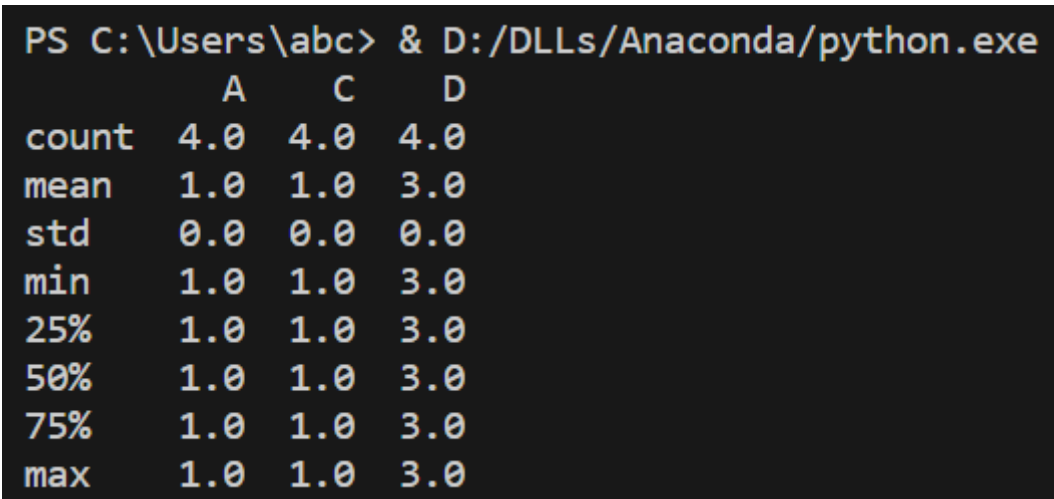
**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe "d:/Aryan/Semester - 3,
[[1.0 Timestamp('2013-01-02 00:00:00') 1.0 3 'test' 'foo']
 [1.0 Timestamp('2013-01-02 00:00:00') 1.0 3 'train' 'foo']
 [1.0 Timestamp('2013-01-02 00:00:00') 1.0 3 'test' 'foo']
 [1.0 Timestamp('2013-01-02 00:00:00') 1.0 3 'train' 'foo']]
```

**Task 2:- Implementing describe() Function****Python Code:**

```
import numpy as np
import pandas as pd
df = pd.DataFrame(
{
```

```
"A": 1.0,  
"B": pd.Timestamp("20130102"),  
  
"C": pd.Series(1, index=list(range(4)), dtype="float32"),  
"D": np.array([3] * 4, dtype="int32"),  
"E": pd.Categorical(["test", "train", "test", "train"]),  
"F": "foo",  
}  
)  
print(df.describe())
```

**Output:**

	A	C	D
count	4.0	4.0	4.0
mean	1.0	1.0	3.0
std	0.0	0.0	0.0
min	1.0	1.0	3.0
25%	1.0	1.0	3.0
50%	1.0	1.0	3.0
75%	1.0	1.0	3.0
max	1.0	1.0	3.0

**Task 3:- Transposing your data****Python Code:**

```
import numpy as np  
import pandas as pd  
df = pd.DataFrame(  
    {  
        "A": 1.0,  
        "B": pd.Timestamp("20130102"),  
        "C": pd.Series(1, index=list(range(4)), dtype="float32"),  
        "D": np.array([3] * 4, dtype="int32"),  
        "E": pd.Categorical(["test", "train", "test", "train"]),  
        "F": "foo",  
    }  
)  
  
print(df.T)
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe "d:/Aryan/Semester - 3/Programming With Python"
      0      1      2      3
A      1.0      1.0      1.0      1.0
B  2013-01-02 00:00:00  2013-01-02 00:00:00  2013-01-02 00:00:00  2013-01-02 00:00:00
C      1.0      1.0      1.0      1.0
D      3      3      3      3
E      test      train      test      train
F      foo      foo      foo      foo
```

**Task 4:- Sorting your data (using [DataFrame.sort\\_index\(\)](#) & [DataFrame.sort\\_values\(\)](#))**
**Python Code:**

```
import numpy as np
import pandas as pd
df = pd.DataFrame(
    {
        "A": 1.0,
        "B": pd.Timestamp("20130102"),
        "C": pd.Series(1, index=list(range(4)), dtype="float32"),
        "D": np.array([3] * 4, dtype="int32"),
        "E": pd.Categorical(["test", "train", "test", "train"]),
        "F": "foo",
    }
)

print(df.sort_index(axis=1, ascending=False))
print(df.sort_values(by="B"))
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe "d:/Aryan/Semester - 3/Programming With Python"
      F      E      D      C      B      A
0  foo  test  3  1.0  2013-01-02  1.0
1  foo  train 3  1.0  2013-01-02  1.0
2  foo  test  3  1.0  2013-01-02  1.0
3  foo  train 3  1.0  2013-01-02  1.0
      A      B      C      D      E      F
0  1.0  2013-01-02  1.0  3  test  foo
1  1.0  2013-01-02  1.0  3  train  foo
2  1.0  2013-01-02  1.0  3  test  foo
3  1.0  2013-01-02  1.0  3  train  foo
```

**Task 5:- Selection (getitem[],slice:,label, position)**
**Python Code:**

```
import numpy as np
import pandas as pd
df = pd.DataFrame(
    {
```

```
"A": 1.0,  
"B": pd.Timestamp("20130102"),  
"C": pd.Series(1, index=list(range(4)), dtype="float32"),  
"D": np.array([3] * 4, dtype="int32"),  
"E": pd.Categorical(["test", "train", "test", "train"]),  
"F": "foo",  
}  
)  
  
print(df.loc["20130102":"20130104", ["A", "B"]])  
print(df.iloc[3])  
print(df[df["A"] > 0])
```

**Output:**

```
PS C:\Users\abc> & D:/DLLs/Anaconda/python.exe "d:/Aryan/Semester - 3  
Empty DataFrame  
Columns: [A, B]  
Index: []  
A          1.0  
B    2013-01-02 00:00:00  
C          1.0  
D          3  
E        train  
F         foo  
Name: 3, dtype: object  
   A      B      C  D  E  F  
0  1.0  2013-01-02  1.0  3  test  foo  
1  1.0  2013-01-02  1.0  3  train  foo  
2  1.0  2013-01-02  1.0  3  test  foo  
3  1.0  2013-01-02  1.0  3  train  foo
```