**HARSH KASLIWAL**

**2003085**

**C21**

**EXPERIMENT 12**

**AIM:** Programs Based on Pandas in Python

# THEORY:

**PANDAS**

Pandas is a Python library for **data analysis**. Started by Wes McKinney in 2008 out of a need for a powerful and flexible quantitative analysis tool, pandas has grown into one of the most popular Python libraries.

### Dataframes

**Pandas DataFrame** is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the **data**, **rows**, and **columns**.

**Series**

A series in Python is a kind of one-dimensional array of any data type that we specified in the pandas module. The only difference you can find was, each value in a Python pandas series is associated with the index. The default index value of the Python pandas Series is from 0 to number – 1, or you can specify your own index values.

## Task 1.

Write a pandas program to:

1. add, subtract, multiple and divide two pandas series
2. compare the elements of the two Pandas Series.
3. convert a dictionary to a Pandas series.
4. convert a NumPy array to a Pandas series.

# Code:

# arithmetic operations on panda series import numpy as np

import pandas as pd

ls1 = [1, 2, 3, 4, 5, 6]

ls2 = [5, 2, 6, 4, 30, 12]

ser1 = pd.Series(ls1)

ser2 = pd.Series(ls2)

print("ser1:\n", ser1) print("ser2:\n", ser2)

# addition print("\nAddition: ") ser = ser1 + ser2 print(ser)

# subtraction print("\nSubtraction: ") serSub = ser1 - ser2

# multiplication print("\nMultiplication: ") serMul = ser1 \* ser2 print(serMul)

# Division print("\nDivision: ") serDiv = ser1 / ser2 print(serDiv)

# Comparing both series print("\nComparing both series: ") print(ser1 == ser2)

# Panda series from dictionary

print("\nConverting a Dictionary to Panda series: ") dictTem = {

'a': 11,

'b': 22,

'c': 33,

'd': 44,

'e': 55,

'f': 66,

'g': 77,

}

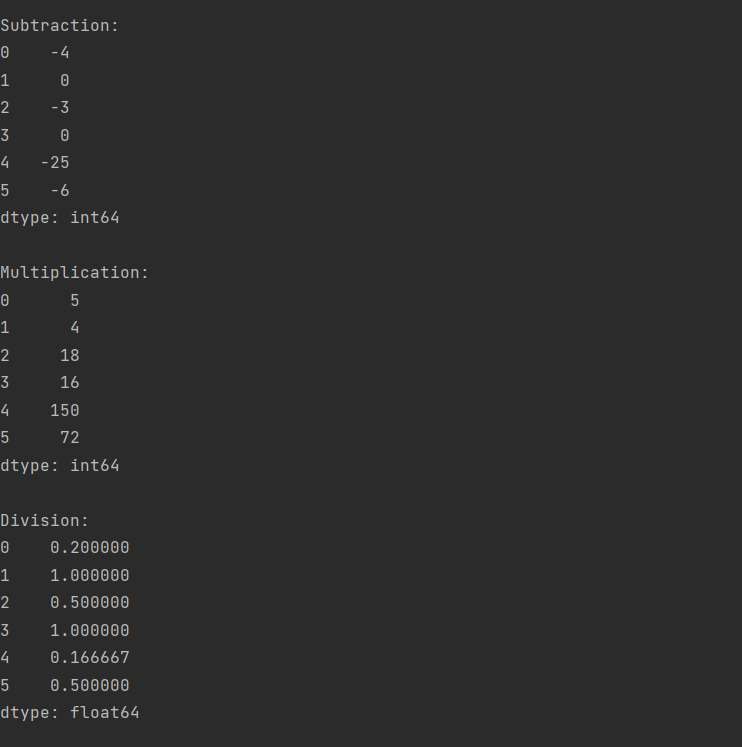
serTem = pd.Series(dictTem) print("Dict: ",dictTem) print("Series: ") print(serTem)

# Converting a numpy array to pandas series print("\nConverting a numpy array to pandas series: ")

arr = np.array([1,2,3,5,6,9]) print("Numpy Array: ",arr) serAr = pd.Series(arr) print("Series: ") print(serAr)

**OUTPUT**







**Task 2:**

Write a program to read csv file in a dataframe, replace missing values with any value, drop the row if all values are missing or contain null values.

### CODE:

import pandas as pd import numpy as np

# creating dataframes as df1 and df2

df1 = pd.DataFrame({'ID': [1, 2, 3, 5, 7, 8],

'Name': ['Sam', 'John', 'Bridge',

'Edge', 'Joe', 'Hope']})

df2 = pd.DataFrame({'ID': [1, 2, 4, 5, 6, 8, 9],

'Marks': [67, 92, 75, 83, 69, 56, 81]})

# merging df1 and df2 by ID

# i.e. the rows with common ID's get # Operations on CSV files

import numpy as np import pandas as pd

flights = pd.read\_csv("flights.csv") print(flights)

# Checking for missing values print("\n",flights.notnull())

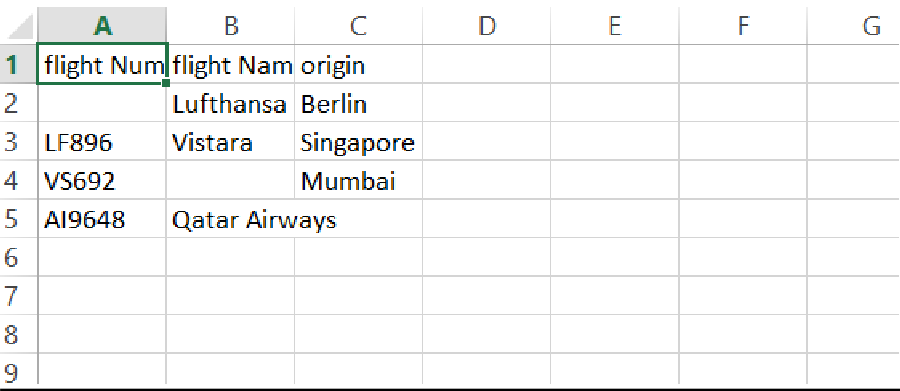
# filling missing values

# flights["flight Num"].fillna("NoNumber",inplace=True) # flights.replace(to\_replace=np.nan,value="NoValue") flights = flights.fillna("NoValue")

# dropping the row if there is a NULL Value flights = flights.dropna()

print("\n",flights)

**Csv File:**

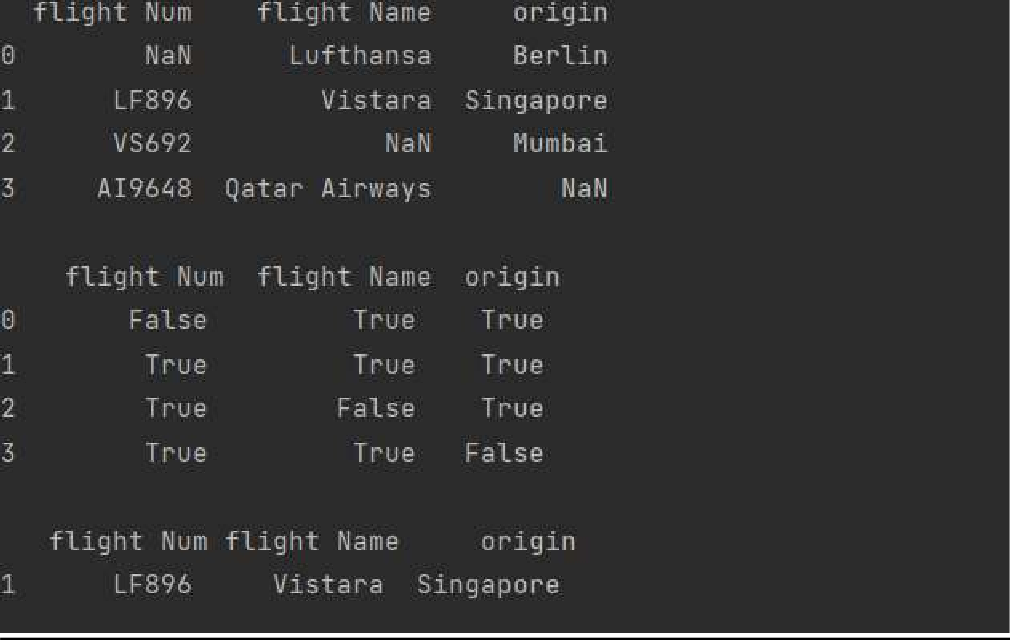


**OutPut:**

### Replacing Null Values with random values



**Droping Row if it will have null values**



# Task 3:

Write a program to demonstrate merging of Frames:

1. on the basis of id
2. using how

### Code:

# merged i.e. {1,2,5,8} print("Merging based on ID")

df = pd.merge(df1, df2, on="ID") print(df)

# merging df1 and df2 by ID

# i.e. the rows with common ID's get merged # with all the ID's of left dataframe i.e. df1

# and NaN for columns of df2 where ID do not match print("Merging with how")

df = pd.merge(df1, df2, on="ID", how="left") print("\n",df)

**OUTPUT:**

