**Pandas DataFrame**

Pandas DataFrame will be created by loading the datasets from existing storage, storage can be SQL Database, CSV file, and Excel file.

Pandas DataFrame can be created from the lists, dictionary, and from a list of dictionary etc.

A Dataframe is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. In dataframe datasets arrange in rows and columns, we can store any number of datasets in a dataframe. We can perform many operations on these datasets like arithmetic operation, columns/rows selection, columns/rows addition etc.

**Dataframe using List:**

*import pandas as pd*

*a = ['disha', 'computer', 'institute', 'experts','in', 'computer', 'training']*

*d = pd.DataFrame(a)*

*print(d)*

**DataFrame from dict of array/lists:**

*import pandas as pd*

*data = {'Name':['Akash', 'Nick', 'Pranav', 'Aman'], 'Course':[‘catia’,’photoshop’,’java’,’digital marketing’]}*

*d = pd.DataFrame(data)*

*print(d)*

**Dataframe from lists using zip()**

lists to create lists of tuples and create a dictionary from it. Then, this dictionary can be used to construct a dataframe.

This function can create pandas DataFrames by merging two lists.

*import pandas as pd*

*Name = ['Akash', 'Nick', 'Pranav', 'Aman']*

*Course=['catia','photoshop','java','digital marketing']*

*a = list(zip(Name, Course))*

*print(a)*

**Pandas Series**

Pandas Series is a one-dimensional labeled array capable of holding data of any type

*import pandas as pd*

*l = [1, 7, 2]*

*l1=['d','i','s','h','a']*

*l2=["disha","computer","institute"]*

*a=pd.Series(l)*

*b=pd.Series(l1)*

*c=pd.Series(l2)*

*print(a)*

*print(b)*

*print(c)*

**Label**

Labels need not be unique but must be a hashable type. The object supports both integer and label-based indexing and provides a host of methods for performing operations

If nothing else is specified, the values are labeled with their index number. Eg(0 1 2 …)

In order to access an element from series, we have to set values by index label. A Series is like a fixed-size dictionary in that you can get and set values by index label.

*l = ["c","c++","tally","photoshop"]*

*a = p.Series(l,index = ["coding", "coding", "accounting","designing"])*

*print(a)*

**.read\_csv()**

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*report*

**.head() method**

head() method is used to return top n rows of a data frame or series, by default it’s 5.

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*top=report.head()*

*top*

**Creating series & Calling on Series by .head()**

**Cerating a series by heading**

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*series = report["name"]*

*series*

**Calling on Series by .head()**

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*series = report["name"]*

*t=series.head()*

*t*

**Calling on Series with n parameter()**

*import pandas as pd*

*data = pd.read\_csv("d:\\python\\p1.csv")*

*n =int(input("enter number of rows you want "))*

*series = data["name"]*

*top = series.head(n = n)*

*top*

**.tail() method**

tail() method is used to return bottom n rows of a data frame or series, by default it’s 5.

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*top=report.tail()*

*top*

**Calling on Series by .tail()**

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*series = report["name"]*

*t=series.tail()*

*t*

**Dataframe.describe() method**

 describe() is used to view some basic statistical details.

*import pandas as pd*

*import re*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*ar = report["area"].describe()*

*ar*

**Rows and Columns in Pandas DataFrame**

**Columns:**

access the columns by calling them by their columns name

*import pandas as pd*

*data = {'Name':['Akash', 'Nick', 'Pranav', 'Aman'], 'Course':['catia','photoshop','java','digital marketing'],'age':[26,22,23,29]}*

*d = pd.DataFrame(data)*

*print(d[['Name', 'age']])*

**Column Addition:**

declare a new list as a column and add to a existing Dataframe.

*import pandas as pd*

*data = {'Name':['Akash', 'Nick', 'Pranav', 'Aman'], 'Course':['catia','photoshop','java','digital marketing'],'age':[26,22,23,29]}*

*d = pd.DataFrame(data)*

*print(d[['Name', 'age']])*

*add=['dangechowk','kalewadi','chinchwad','wakad']*

*d['Address']=add*

*d*

**Column Deletion:**

drop() method Columns will deleted by dropping columns with column names.

*import pandas as pd*

*data = {'Name':['Akash', 'Nick', 'Pranav', 'Aman'], 'Course':['catia','photoshop','java','digital marketing'],'age':[26,22,23,29]}*

*d = pd.DataFrame(data)*

*print(d)*

*d.drop(['age'], axis = 1, inplace = True)*

*print(d)*

**Convert dataframe to Numpy array**

*import pandas as pd*

*d = pd.DataFrame([[1, 2, 3],[4, 5, 6],[7, 8, 9],[10, 11, 12]],columns=['a', 'b', 'c'])*

*a = d.to\_numpy()*

*print(type(a))*

*a*

**Convert a particular column into numpy array**

*import pandas as pd*

*d = pd.DataFrame([[1, 2, 3],[4, 5, 6],[7, 8, 9],[10, 11, 12]],columns=['a', 'b', 'c'])*

*a = d[['a', 'c']].to\_numpy()*

*a*

**Pandas Series.to\_numpy()**

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*s=pd.Series(report['price'].head())*

*b=s.to\_numpy()*

*b*

**Parameters to Provide the dtype**

*import pandas as pd*

*report = pd.read\_csv("d:\\python\\p1.csv")*

*s=pd.Series(report['price'].head())*

*b=s.to\_numpy(dtype='int') //(dtype='float') , (dtype='str')*

*b*

**Series.as\_matrix()**

Convert the given series or dataframe object to Numpy-array representation.

*import pandas as pd*

*data = pd.Series(['C', 'C++', 'html', 'css', 'js'])*

*i = ['Course 1', 'Course 2', 'Course 3', 'Course 4', 'Course 5']*

*data.index = i*

*print(data)*

**Extracting rows using .loc[]**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv")*

*row = data.loc[2]*

*row*

**Muiltipal rows**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv")*

*row = data.iloc[[4, 5, 9, 7]]*

*row*

**Extracting rows using .iloc[]**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv")*

*row1 = data.iloc[4:10]*

*row = data.iloc[4:12:2]*

*print(row)*

*print("\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*")*

*print(row1)*

**Compare rows**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv")*

*row = data.loc[[4,5,6,7,8]]*

*row1 = data.iloc[4:9]*

*row1 == row*

**Indexing and Selecting Data with Pandas**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv", index\_col="name")*

*rec = data[[ "course", "area"]]*

*rec*

**Indexing a DataFrame using .loc[ ]**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv", index\_col="name")*

*rec1 = data.loc["chetan"]*

*rec2= data.loc["aman"]*

*print(rec1)*

*print(rec2)*

**Selecting multiple rows**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv", index\_col="name")*

*rec = data.loc[["chetan","aman"]]*

*print(rec)*

**Selecting two rows and three columns**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv",index\_col="name")*

*rec = data.loc[["chetan","aman"],['course','area','price']]*

*print(rec)*

**Selecting all of the rows and some columns**

*import pandas as pd*

*data =pd.read\_csv("d:\\python\\p1.csv",index\_col="name")*

*rec = data.loc[:,['course','area','price']]*

*print(rec)*

**Adding new column to existing DataFrame in Pandas**

**By declaring a new list as a column**

*import pandas as pd*

*data = {*

*'Name': ['sakshi', 'akash', 'Gaurav', 'disha'],*

*'address' : ['ravet', 'pimpri', 'chinchwad', 'dangechowk'],*

*'Qualification': ['Msc', 'bca', 'bsc', 'Msc']*

*}*

*new = pd.DataFrame(data)*

*print(new)*

*course= ['c','c++','html','java']*

*new['course'] = course*

*print(df)*

**By using DataFrame.insert()**

insert method allows the user to insert a column in a dataframe

Syntax:

DataFrameName.insert(loc, column, value, allow\_duplicates = False)

*import pandas as pd*

*data = {*

*'Name': ['sakshi', 'akash', 'Gaurav', 'disha'],*

*'address' : ['ravet', 'pimpri', 'chinchwad', 'dangechowk'],*

*'Qualification': ['Msc', 'bca', 'bsc', 'Msc']*

*}*

*n = pd.DataFrame(data)*

*print(n)*

*n.insert(1, 'course', ['c#','c++','java','java'], True)*

*n['course'] = course*

*print(n)*

**Using Dataframe.assign()**

Dataframe.assign() method assign new columns to a DataFrame

*import pandas as pd*

*data = {*

*'Name': ['sakshi', 'akash', 'Gaurav', 'disha'],*

*'address' : ['ravet', 'pimpri', 'chinchwad', 'dangechowk'],*

*'Qualification': ['Msc', 'bca', 'bsc', 'Msc']*

*}*

*new = pd.DataFrame(data)*

*print(new)*

*new1=new.assign(course= ['c','c++','html','java'])*

*new['course'] = course*

*print(new)*

**Difference between two dates**

*import datetime*

*# datetime(year, month, day, hour, minute, second)*

*d1= datetime.datetime(2023, 2, 13)*

*d2 = datetime.datetime(2023, 1, 26)*

*dif = a-b*

*print('Difference: ', dif)*