G. H. Raisoni College Of Engineering And Management, Wagholi Pune

2021- 2022

Group B :-Assignment no :-6

Department	CE [SUMMER 2022 (Online)]		
Term / Section	III/B	Date Of **submission**	11-12-2021
Subject Name /Code	Python for Data Science / UCSP204		
Roll No.	SCOB77	Name	Pratham Rajkumar pitty
Registration Number	2020AC0E1100107		

Group B :- Assignment No 6 (13)

Aim :→ Write a Python program to create a Dataframe of 6 rows from Dictionary. First create a Tourism_visitors dictionary from four lists of "cities", "visitors", "signups" and "weekdays". convert the dictionary to Dataframe importing pandas library. And create row_labels as the short forms of the country names. And display the data fram.

Theory :→

Pandas DataFrame :

Pandas DataFram is a 2-D Size-Mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A DataFrame is a 2-D data structure. ie. data is aligned in a tabular fashion in rows and Column. pandas DataFrame consists of 3 principal components, the data, rows & Column

The basic operations which can be performed on pandas dataframe :

- Creating a DataFrame
- Working with missing data
- Dealing with Rows & Columns
- iterating over rows & column
- indexing and Selecting Dataframe

(1) Creating a pandas Dataframe

In the real world pandas Dataframe is created by loading the data set from existing storage, storage can be SQL Database, CSV File & Excel File. pandas DataFrame can be created in different ~~ways~~ From lists, comp dictionary.

and from a list of dictionary, etc.

Example.

```
import pandas as pd.
List = [' Apple, 'orange' Mango, 'Banana
dF = pd.Dataframe (List)
print (dF)
```

Creating data frame from dict of ndarray/lists:-

To creat DataFrame from dict of narray/lists, all the narray must be some length. If index is passed then the length index should be equal to the length of arrays. If no index is passed, then by default index will be range (n) where n is the array length.

Ex.

```
import pandas as pd
data = {'Name': ['Apple', 'Tom', 'nick', 'krish'],
        'Age': [20, 21, 19, 18]}
df = Pd. Data Frame (data)
print ( df)
```

(2) Dealing with rows & column

A DataFram is a 2-D data Structure, i.e. data is aligned in a tabular fashion in rows and column. Basic operation like adding, deleting, selecting, & renaming) can be performed on it.

column Selection:→ In order to select a column in pandas DataFrame, we can access the columns by calling them by their columns name.

Ex.

```
import pandas as pd          # import Pandas Package
data={'Name':['Jai','prince','Raj'],          # Define
      'Age':[27,24,23]
      'Address':['Delhi','Kanpor','pune']
      'qualification':['MSc','MA','MCA']}
df= pd.DataFrame(data)       # convert dictionary to DataFra
print(df[['Name','Age']])    # select 2 columns
```

Output	Name	Age
0	Jay	27
1	prince	24
2	Raj	23

Row Selection :→ Pandas provide a unique method to retrive rows from a DataFrame. 'Dataframe.loc[]' method is used to retrive rows from pandas Dataframe. Rows can also be selected by passing integer location to an 'iloc[]' Function.

Ex

```
import pandas as pd.
data = pd.read_csv("nba.csv", index_col="Name")
# making data frame from csv file
# retrieving row by loc method
First = data.loc["Avery Bradley"]
Second = data.loc["R.J. Hunter"]
print(First,"\n",Second)
```

Note:→ # Link For nba.csv → https://media.geeksforgeeks.org/wp-content/uploads/nba.csv

(3) Indexing and Selecting Data:→

indexing in pandas means

Simply selecting particular rows and columns of data from a DataFrame. Indexing could mean selecting all the rows and some of the columns, some of the rows and all of the columns, or some of each of the rows & columns. It is also known as Subset selection.

Indexing a Dataframe using indexing operator [] :→
Indexing operator is used to refer to the square brackets following an object.
The .loc & .iloc indexers also use the indexing operator to make selection. In this indexing operator to refer to df[].

Ex

Selecting a single column

```
# In order to select single column simply put the name
# of column in between.
import pandas as pd
data = Pd.read_csv("nba.csv", index_col="Name")
First = data["Age"]  ← retriving column
print(First)
```

.loc

```
First = data.loc["Avery Bradley"]   # single row
second = data.loc["R.J. Hunter"]
print(First, "\n\n", second)
```

.iloc

```
row2 = data.iloc[3]   # Single row
print[row2]
```

In [7]:

```
print("*************SCOB77_Pratham pitty_Group B_Assignment_6*************")
print("\n------------------------------------------------------------")

import pandas as pd

Tourism_visitors = {"cities":['Amravati', 'Jalgaon', 'Pune', 'Paratwada','Nagpur','Dombival
"signups":['1/10/2020','2/10/2020','3/10/2020','4/10/2020','5/10/2020','6/10/2020'],"Weekda

print(Tourism_visitors.keys())

Brics = pd.DataFrame(Tourism_visitors)
print(Brics)
print("-------------------------------------------------\n")
row_labels = ['1', '2', '3', '4', '5', '6']

df = pd.DataFrame(Tourism_visitors, index = row_labels)
print(df)
```

```
*************SCOB77_Pratham pitty_Group B_Assignment_6*************

------------------------------------------------------------
dict_keys(['cities', 'Visitors', 'signups', 'Weekdays'])
      cities Visitors    signups   Weekdays
0   Amravati  Pratham  1/10/2020     Monday
1    Jalgaon    Rajas  2/10/2020    Tuesday
2       Pune    Sahil  3/10/2020  Wednesday
3  Paratwada    Shyam  4/10/2020   Thursday
4     Nagpur  Sourabh  5/10/2020     Friday
5  Dombivali    Anmol  6/10/2020   Saturday
-------------------------------------------------

      cities Visitors    signups   Weekdays
1   Amravati  Pratham  1/10/2020     Monday
2    Jalgaon    Rajas  2/10/2020    Tuesday
3       Pune    Sahil  3/10/2020  Wednesday
4  Paratwada    Shyam  4/10/2020   Thursday
5     Nagpur  Sourabh  5/10/2020     Friday
6  Dombivali    Anmol  6/10/2020   Saturday
```

In [ ]: