Ex. No: 6

Aim: *Perform following operations using pandas.*

Theory:

- Pandas is a Python library used for working with data sets.
- It has functions for analyzing, cleaning, exploring, and manipulating data.
- This library is built on top of the NumPy library.
- Pandas is fast and it has high performance & productivity for users.

If you want work with pandas then install Pandas on your machine, just type the below command on your terminal:

```
$pip install pandas
```

Now you need to import the library:

```
$import pandas as pd
```

pd is the defacto abbreviation for Pandas used by the data science community

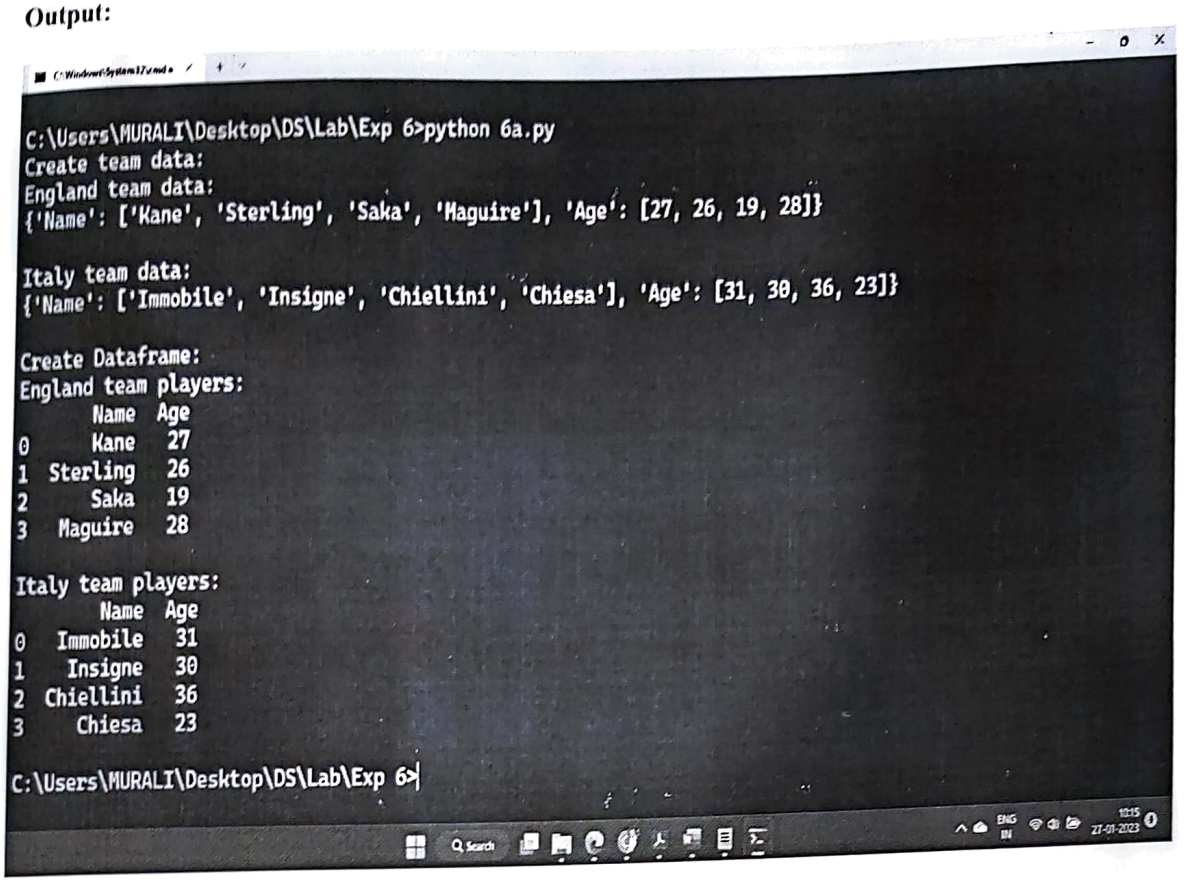
6 (A):
Creating dataframe:

We shall create a Football data frame that stores the record of 4 players each from Euro Cup 2020's finalists – England and Italy.

Example:

```
import pandas as pd

print("Create team data:")
data_england = {'Name': ['Kane', 'Sterling', 'Saka', 'Maguire'], 'Age': [27, 26, 19, 28]}
data_italy = {'Name': ['Immobile', 'Insigne', 'Chiellini', 'Chiesa'], 'Age': [31, 30, 36, 23]}

print("England team data:")
print(data_england)

print("\nItaly team data:")
print(data_italy)

print("\nCreate Dataframe:")
df_england = pd.DataFrame(data_england)
df_italy = pd.DataFrame(data_italy)
print("England team players:")
print(df_england)

print("\nItaly team players:")
print(df_italy)
```

Output:


C:\Users\MURALI\Desktop\DS\Lab\Exp 6>python 6a.py
Create team data:
England team data:
{'Name': ['Kane', 'Sterling', 'Saka', 'Maguire'], 'Age': [27, 26, 19, 28]}
Italy team data:
{'Name': ['Immobile', 'Insigne', 'Chiellini', 'Chiesa'], 'Age': [31, 30, 36, 23]}
Create Dataframe:
England team players:
Name Age
0 Kane 27
1 Sterling 26
2 Saka 19
3 Maguire 28
Italy team players:
Name Age
0 Immobile 31
1 Insigne 30
2 Chiellini 36
3 Chiesa 23
C:\Users\MURALI\Desktop\DS\Lab\Exp 6>

6 (B):
concat():

Let's start by concatenating our two data frames. The word "concatenate" means to "link together in series". Now that we have created two data frames, let's try and "concat" them.

We do this by implementing the concat() function.

Example:

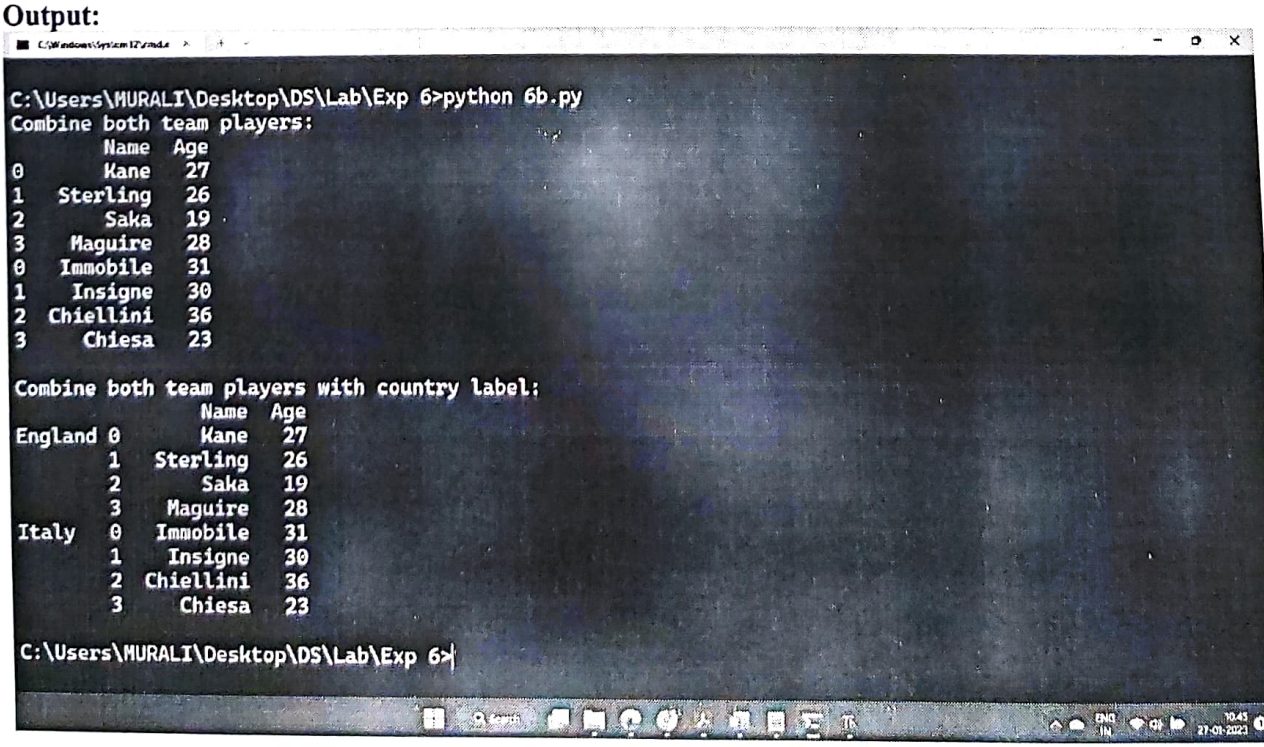
dataframe.py

```
import pandas as pd
def england():
    data_england = {'Name': ['Kane', 'Sterling', 'Saka', 'Maguire'], 'Age': [27, 26, 19, 28]}
    df_england = pd.DataFrame(data_england)
    return df_england

def italy():
    data_italy = {'Name': ['Immobile', 'Insigne', 'Chiellini', 'Chiesa'], 'Age': [31, 30, 36, 23]}
    df_italy = pd.DataFrame(data_italy)
    return df_italy
```

6b.py

```
import pandas as pd
from dataframe import *
df_england=england()
df_italy=italy()
print("Combine both team players:")
frames=[df_england, df_italy]
both_teams=pd.concat(frames)
print(both_teams)
print("\nCombine both team players with country label:")
both_teams_label=pd.concat(frames, keys=["England", "Italy"])
print(both_teams_label)
```

Output:



6 (C):
Setting conditions:

Conditional statements basically define conditions for data frame columns. There may be situations where you have to filter out various data by applying certain column conditions (numeric or non-numeric).

Imagine we want to filter experienced players from our squad. Let's say, we want to filter those players whose age is greater than or equal to 30.

Now, let's try to do some string filtration. We want to filter those players whose name starts with "S". This implementation can be done by pandas' startswith() function.

Example:

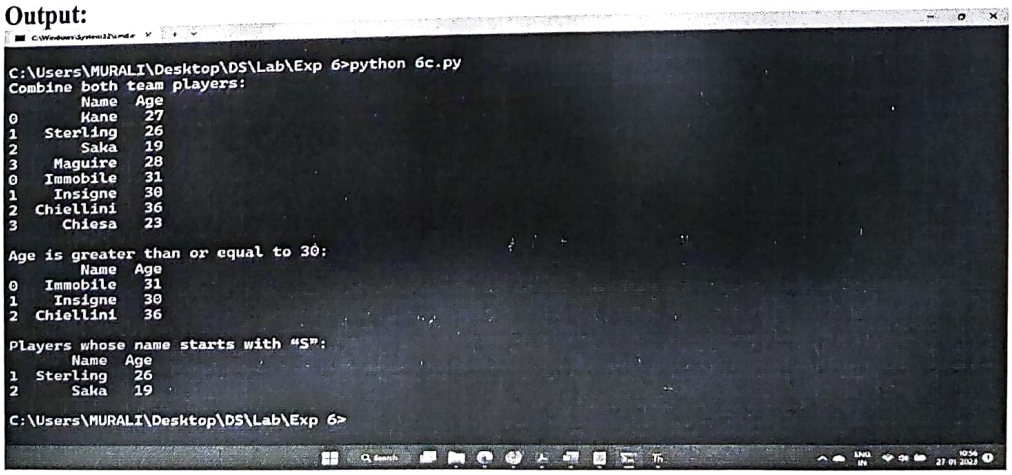
dataframe.py

```
import pandas as pd
def england():
    data_england = {'Name': ['Kane', 'Sterling', 'Saka', 'Maguire'], 'Age': [27, 26, 19, 28]}
    df_england = pd.DataFrame(data_england)
    return df_england
def italy():
    data_italy = {'Name': ['Immobile', 'Insigne', 'Chiellini', 'Chiesa'], 'Age': [31, 30, 36, 23]}
    df_italy = pd.DataFrame(data_italy)
    return df_italy
```

6c.py

```
import pandas as pd
from dataframe import *
df_england=england()
df_italy=italy()
print("Combine both team players:")
frames=[df_england, df_italy]
both_teams=pd.concat(frames)
print(both_teams)
print("\nAge is greater than or equal to 30:")
print(both_teams[both_teams["Age"] >= 30])
print("\nPlayers whose name starts with "S":")
print(both_teams[both_teams["Name"].str.startswith('S')])
```

Output:



6 (D):
Adding a new column:

Let's try adding more data to England team data frame and try to repeat implementing the concat() function after updating the data for England.

Example:

dataframe.py

```
import pandas as pd
def england():
    data_england = {'Name': ['Kane', 'Sterling', 'Saka', 'Maguire'], 'Age': [27, 26, 19, 28]}
    df_england = pd.DataFrame(data_england)
    return df_england
def italy():
    data_italy = {'Name': ['Immobile', 'Insigne', 'Chiellini', 'Chiesa'], 'Age': [31, 30, 36, 23]}
    df_italy = pd.DataFrame(data_italy)
    return df_italy
```

6d.py

```
import pandas as pd
from dataframe import *
df_england=england()
df_italy=italy()
club = ['Tottenham', 'Man City', 'Arsenal', 'Man Utd']
print("'Associated Club' is new column name:")
df_england['Associated Clubs'] = club
print(df_england)
print("\nCombine both team players:")
frames=[df_england, df_italy]
both_teams=pd.concat(frames)
print(both_teams)
```

Output: