Experiment No. 14

Aim: To implement programs based on Pandas library in Python.

Questions:

1. Implement a python program to create and print Pandas Series (without index, with index, using dictionary).

```
import pandas as pd
import numpy as np
data = np.array(['s', 'g', 'u'])
ser = pd.Series(data)
print(ser)
```

```
0    s
1    g
2    u
dtype: object
```

```
data = np.array(['s', 'g', 'u'])
ser = pd.Series(data, index=[41, 42, 43])
print(ser)
```

```
41    s
42    g
43    u
dtype: object
```

```
dict = {'Sanjay': 10, 'Ghodawat': 20, 'University': 30}
ser = pd.Series(dict)
print(ser)
```

```
Sanjay        10
Ghodawat      20
University    30
dtype: int64
```

2. Implement a python program to create and print Pandas DataFrame. (without named indexes, with named indexes, with specific row retrieval)

```
: import pandas as pd
data = {'Name': ['abc', 'def', 'ghi'],
        'Age': [20, 21, 19]}
df = pd.DataFrame(data)
print(df)
```

```
  Name  Age
0  abc   20
1  def   21
2  ghi   19
```

```
: data = {'Name': ['abc', 'def', 'ghi'],
        'Age': [20, 21, 19]}
df = pd.DataFrame(data, index=['rank1', 'rank2', 'rank3'])
print(df)
```

```
      Name  Age
rank1  abc   20
rank2  def   21
rank3  ghi   19
```

```
: print(df.loc['rank2'])
```

```
Name    def
Age      21
Name: rank2, dtype: object
```

3. Implement a python program to read a CSV file and apply head(), tail() and info() methods.

```
df = pd.read_csv("abc.csv")
print(df.to_string())
```

```
   Roll No Name  Cat-1  Cat-2
0        1    a     12     12
1        2    b     14     14
2        3    c     11     11
3        4    d     23     23
4        5    e     21     21
5        6    f     16     16
6        7    g     19     19
7        8    h     13     13
8        9    I     12     12
9       10    j     10     10
```

```
print(df.head())
```

```
   Roll No Name  Cat-1  Cat-2
0        1    a     12     12
1        2    b     14     14
2        3    c     11     11
3        4    d     23     23
4        5    e     21     21
```

```
print(df.tail())
```

```
   Roll No Name  Cat-1  Cat-2
5        6    f     16     16
6        7    g     19     19
7        8    h     13     13
8        9    I     12     12
9       10    j     10     10
```

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 4 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Roll No  10 non-null     int64
 1   Name     10 non-null     object
 2   Cat-1    10 non-null     int64
 3   Cat-2    10 non-null     int64
dtypes: int64(3), object(1)
memory usage: 448.0+ bytes
None
```

4. Implement a python program to read a CSV file and handle null values by using all 3 techniques.

```
df = pd.read_csv("abc.csv")
print(df.to_string())
```

```
   Roll No Name  Cat-1  Cat-2
0        1    a   12.0   12.0
1        2    b   14.0   14.0
2        3    c   11.0    NaN
3        4    d   23.0   23.0
4        5    e   21.0   21.0
5        6    f    NaN   16.0
6        7    g   19.0   19.0
7        8    h   13.0   13.0
8        9    I   12.0   12.0
9       10    j   10.0   10.0
```

```
newDf = df.dropna()
print(newDf.to_string())
```

```
   Roll No Name  Cat-1  Cat-2
0        1    a   12.0   12.0
1        2    b   14.0   14.0
3        4    d   23.0   23.0
4        5    e   21.0   21.0
6        7    g   19.0   19.0
7        8    h   13.0   13.0
8        9    I   12.0   12.0
9       10    j   10.0   10.0
```

```
df.fillna(23, inplace=True)
print(df.to_string())
```

```
   Roll No Name  Cat-1  Cat-2
0        1    a   12.0   12.0
1        2    b   14.0   14.0
2        3    c   11.0   23.0
3        4    d   23.0   23.0
4        5    e   21.0   21.0
5        6    f   23.0   16.0
6        7    g   19.0   19.0
7        8    h   13.0   13.0
8        9    I   12.0   12.0
9       10    j   10.0   10.0
```

```
df = pd.read_csv("abc.csv")
df['Cat-1'].fillna(23, inplace=True)
print(df.to_string())
```

```
   Roll No Name  Cat-1  Cat-2
0        1    a   12.0   12.0
1        2    b   14.0   14.0
2        3    c   11.0    NaN
3        4    d   23.0   23.0
4        5    e   21.0   21.0
5        6    f   23.0   16.0
6        7    g   19.0   19.0
7        8    h   13.0   13.0
8        9    I   12.0   12.0
9       10    j   10.0   10.0
```

5. Implement a python program to read a CSV file and handle duplicate rows.

```
df = pd.read_csv("abc.csv")
df.duplicated()
```

```
0    False
1    False
2    False
3    False
4    False
5     True
6    False
7    False
8    False
9    False
dtype: bool
```

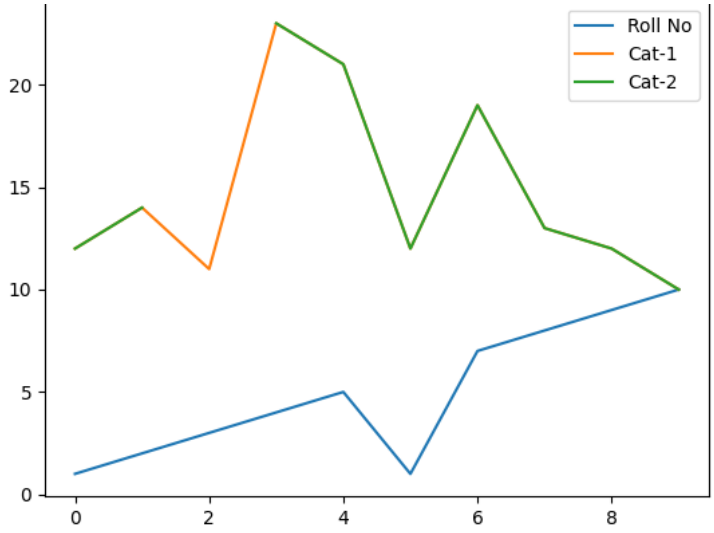
```
df.drop_duplicates(inplace=True)
```

```
print(df.to_string())
```

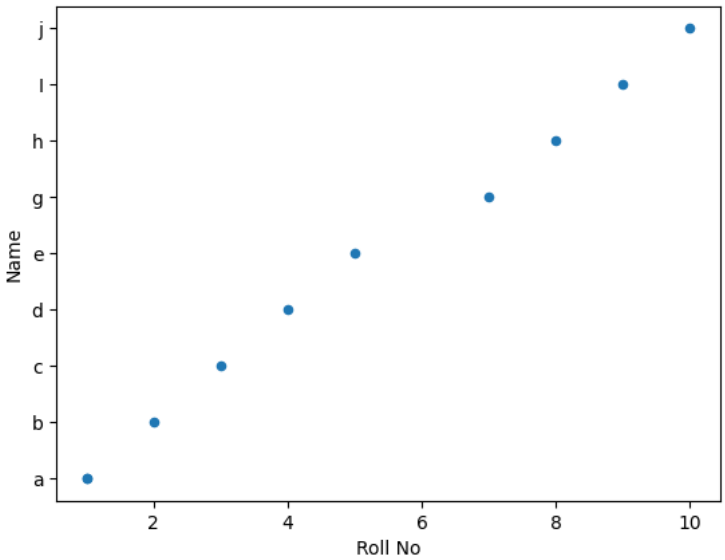
```
   Roll No Name  Cat-1  Cat-2
0        1    a     12   12.0
1        2    b     14   14.0
2        3    c     11    NaN
3        4    d     23   23.0
4        5    e     21   21.0
6        7    g     19   19.0
7        8    h     13   13.0
8        9    I     12   12.0
9       10    j     10   10.0
```

6. Implement a python program to read a CSV file and plot the same using graph, scatter plot and histogram.

```
import sys
import matplotlib
matplotlib.use('Agg')

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("abc.csv")
df.plot()
plt.show()
# plt.savefig(sys.stdout.buffer)
# sys.stdout.flush()
plt.savefig('a.png')
```




```
import sys
import matplotlib
matplotlib.use('Agg')

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("abc.csv")
df.plot(kind='scatter', x='Roll No', y='Name')
plt.show()
# plt.savefig(sys.stdout.buffer)
# sys.stdout.flush()
plt.savefig('a.png')
```



```python
import sys
import matplotlib
matplotlib.use('Agg')

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("abc.csv")
df.plot(kind='hist')
plt.show()
# plt.savefig(sys.stdout.buffer)
# sys.stdout.flush()
plt.savefig('a.png')
```