1. Python Programs for creation and manipulation of Data Frames using Pandas Library

**Code:**

# CREATION OF DATAFRAMES

import pandas as pd

data = pd.DataFrame({"Patient ID": ['55691','55692','55693','55694','55695','55696','55697','55698','55699','55700'],

                    "Gender": ['M','F','F','M','M','F','F','M','F','F'],

                     "Age": ['25','27','29','31','33','45','55','40','58','60'],

                     "Zipcode": ['520003','520023','520013','530003','520023','520526','521301','531309','520003','526732']

                     })

print(data)

**OUTPUT**

Patient ID Gender Age Zipcode

0 55691 M 25 520003

1 55692 F 27 520023

2 55693 F 29 520013

3 55694 M 31 530003

4 55695 M 33 520023

5 55696 F 45 520526

6 55697 F 55 521301

7 55698 M 40 531309

8 55699 F 58 520003

9 55700 F 60 526732

# MANIPULATION OF DATAFRAMES

# Add New Column to pandas DataFrame

Test = ["positive", "positive", "negative", "negative", "negative", "positive","negative","positive","positive","negative"]

print(Test)           # Create list

data\_add = data.assign(Test = Test)

print(data\_add)

['positive', 'positive', 'negative', 'negative', 'negative', 'positive', 'negative', 'positive', 'positive', 'negative']

Patient ID Gender Age Zipcode Test

0 55691 M 25 520003 positive

1 55692 F 27 520023 positive

2 55693 F 29 520013 negative

3 55694 M 31 530003 negative

4 55695 M 33 520023 negative

5 55696 F 45 520526 positive

6 55697 F 55 521301 negative

7 55698 M 40 531309 positive

8 55699 F 58 520003 positive

9 55700 F 60 526732 negative

# Rename Columns of pandas DataFrame

data\_rename = data.copy()

data\_rename.columns = ["P\_ID", "p\_Gender", "P\_Age", "P\_Zipcode",]

print(data\_rename)

P\_ID p\_Gender P\_Age P\_Zipcode

0 55691 M 25 520003

1 55692 F 27 520023

2 55693 F 29 520013

3 55694 M 31 530003

4 55695 M 33 520023

5 55696 F 45 520526

6 55697 F 55 521301

7 55698 M 40 531309

8 55699 F 58 520003

9 55700 F 60 526732

Patient ID Gender Age Zipcode

0 55691 M 25 520003

1 55692 F 27 520023

2 55693 F 29 520013

3 55694 M 31 530003

4 55695 M 33 520023

7 55698 M 40 531309

5 55696 F 45 520526

6 55697 F 55 521301

8 55699 F 58 520003

9 55700 F 60 526732

['55701', 'F', '72', '521301']

# Sort Rows of pandas DataFrame

data\_sort = data.copy()

data\_sort = data\_sort.sort\_values("Age")

print(data\_sort)

Patient ID Gender Age Zipcode

0 55691 M 25 520003

1 55692 F 27 520023

2 55693 F 29 520013

3 55694 M 31 530003

4 55695 M 33 520023

5 55696 F 45 520526

6 55697 F 55 521301

7 55698 M 40 531309

8 55699 F 58 520003

9 55700 F 60 526732

10 55701 F 72 521301

#Add New Row to pandas DataFrame

new\_row = ["55701", "F", "72", "521301"]

print(new\_row)

data\_new3 = data.copy()

data\_new3.loc[10] = new\_row

print(data\_new3)

0 4

1 4

2 4

3 4

4 4

5 4

6 4

7 4

8 4

9 4

dtype: int64

# Count the no.of columns

print(data.count(1))

# Remove Column from pandas DataFrame

data\_drop = data.drop("Zipcode",axis=1)   # Drop variable from DataFrame

print(data\_drop)

# for showing the dimension of the dataframe

print('Shape')

print(data.shape)

OUTPUT

 Shape

(10, 4)

 # for showing the statistical info of the dataframe

print("\n\nDescribe")

data.describe()

Describe

|  | **Patient ID** | **Gender** | **Age** | **Zipcode** |
| --- | --- | --- | --- | --- |
| **count** | 10 | 10 | 10 | 10 |
| **unique** | 10 | 2 | 10 | 8 |
| **top** | 55691 | F | 25 | 520003 |
| **freq** | 1 | 6 | 1 | 2 |