**Coding Assessment Python**

**By Afreen Sultana A**

**Data Cleaning**

**To remove duplicates**

import pandas as pd

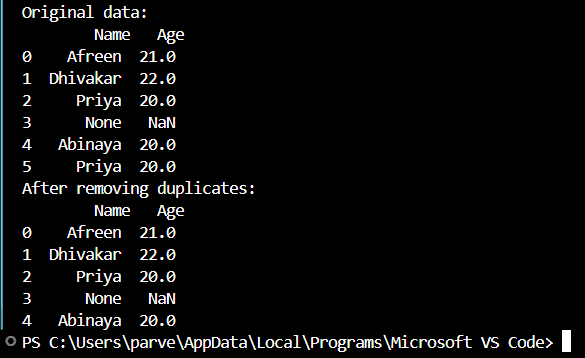
data={'Name':['Afreen', 'Dhivakar', 'Priya', None, 'Abinaya','Priya'],'Age':[21, 22, 20, None, 20, 20]}

df=pd.DataFrame(data)

print("Original data:\n", df)

df\_unique=df.drop\_duplicates()

print("After removing duplicates:\n", df\_unique)



**To remove missing datas**

import pandas as pd

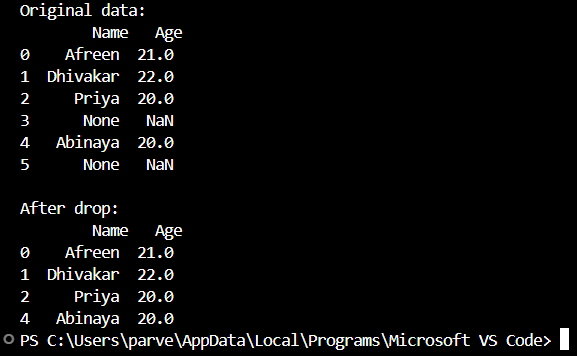
data={'Name':['Afreen', 'Dhivakar', 'Priya', None, 'Abinaya', None],'Age':[21, 22, 20, None, 20, None]}

df=pd.DataFrame(data)

print("Original data:\n", df)

df\_cleaned=df.dropna()

print("\nAfter drop:\n",df\_cleaned)

****

**To replace data**

import pandas as pd

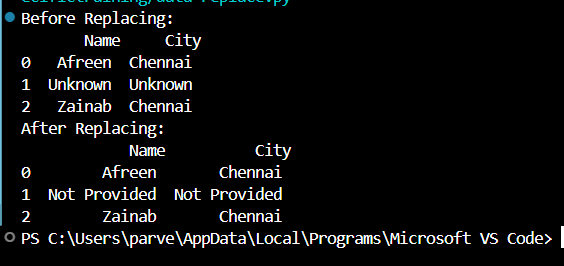
data={'Name':['Afreen','Unknown', 'Zainab'],'City':['Chennai','Unknown','Chennai']}

df=pd.DataFrame(data)

print("Before Replacing:\n",df)

df\_replace=df.replace('Unknown','Not Provided')

print("After Replacing:\n",df\_replace)



**Joins in Python**

import pandas as pd

Students=pd.DataFrame({'Student\_ID':[1,2,3,4,5],'Name':['Afreen', 'Priya','Suganthi','Abi','Zainab']})

Scores=pd.DataFrame({'Student\_ID':[1,2,3,6,7],'score':[90,92,86,80,93]})

print("Student Table:\n",Students)

print("Scores Table:\n",Scores)

#inner join

inner\_join=pd.merge(Students, Scores, on='Student\_ID', how='inner')

print("Inner Join:\n",inner\_join)

#left join

left\_join=pd.merge(Students, Scores, on='Student\_ID', how='left')

print("Left Join:\n",left\_join)

#right join

right\_join=pd.merge(Students, Scores, on='Student\_ID', how='right')

print("Right Join:\n",right\_join)

#outer join

outer\_join=pd.merge(Students, Scores, on='Student\_ID', how='outer')

print("Outer Join:\n",outer\_join)

|  |  |
| --- | --- |
|  |  |