

Practical No 3

Aim: Perform the various operations on the dataset .1.Handle missing values in the datasets. 2. Combine the cleaned datasets into a single dataset.

Software used: IDLE

Theory:-

Dataset: Dataset 1 (dataset1.csv)

id,age,income,gender

1,25,50000,Male

2,30,,Female

3,22,45000,

4,,80000,Female

5,28,52000,Male

6,40,70000,Female

7,50,90000,Male

8,45,85000,Female

9,33,62000,Male

10,27,48000,Female 11,23,40000,Male

12,35,60000,Female 13,31,65000,Male

14,38,80000,Female

15,29,50000,

Dataset 2 (dataset2.csv)

id,age,income,gender

16,26,55000,Female

17,32,61000,Male

18,24,42000,Female 19,34,71000,Male

20,29,49000,Female

Code:

1 . Handle missing values in the datasets.

```
import pandas as pd
```

```
# Load datasets df1 =
```

```
pd.read_csv('dataset1.csv') df2 =
```

```
pd.read_csv('dataset2.csv')
```

```
# Display the original data
```

```
print("Original Data Part 1:")
```

```
print(df1) print("\nOriginal
```

```
Data Part 2:") print(df2)
```

```
# Handling missing values
```

```
# Fill missing 'age' and 'income' with the mean of the column
```

```
df1['age'].fillna(df1['age'].mean(), inplace=True)
```

```
df1['income'].fillna(df1['income'].mean(), inplace=True) # Fill
```

```
missing 'gender' with the mode of the column
```

```
df1['gender'].fillna(df1['gender'].mode()[0],
```

```
inplace=True) # Display cleaned data
```

```
print("\nCleaned Data Part 1:")
```

```
print(df1);
```

2. Combine the cleaned datasets into a single dataset.

Code:

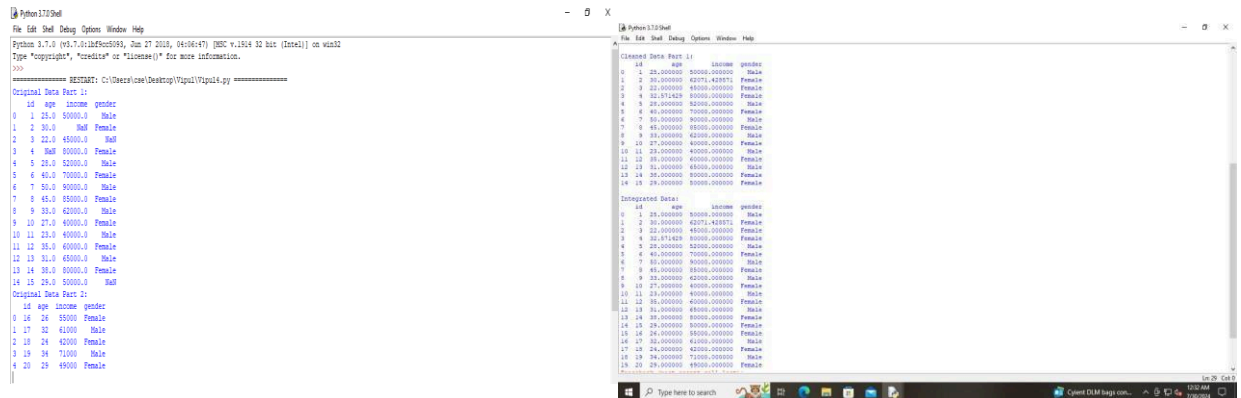
```
# Concatenate the datasets df = pd.concat([df1, df2], ignore_index=True)
```

```
# Display integrated data
```

```
print("\nIntegrated Data:")
```

```
print(df);
```

Result:-



The image shows two side-by-side screenshots of a Python 3.7.0 shell window. The left window displays the output of a script that reads a CSV file and prints the data. The right window displays the output of a script that reads a CSV file and prints the data after cleaning.

Original Data Part 1:

age	income	gender
1	25.0	Male
2	30.0	Male
3	22.0	Male
4	35.0	Female
5	28.0	Male
6	40.0	Female
7	50.0	Male
8	45.0	Female
9	33.0	Male
10	27.0	Female
11	23.0	Male
12	35.0	Female
13	31.0	Male
14	38.0	Female
15	25.0	Male

Original Data Part 2:

age	income	gender
16	25	Female
17	32	Male
18	24	Female
19	34	Male
20	29	Female

Cleaned Data Part 1:

age	income	gender
1	25.000000	Male
2	30.000000	Female
3	22.000000	Female
4	32.974629	Female
5	25.000000	Male
6	40.000000	Female
7	50.000000	Male
8	45.000000	Female
9	33.000000	Male
10	27.000000	Female
11	23.000000	Male
12	35.000000	Female
13	31.000000	Male
14	38.000000	Female
15	25.000000	Male

Cleaned Data Part 2:

age	income	gender
16	25.000000	Female
17	32.000000	Female
18	24.000000	Female
19	34.000000	Male
20	29.000000	Female

Conclusion:-

In this way, we studied performing the various operations on the dataset.