**DATA CLEANING AND PREPARATION - Excel**

**Ex.No. :4b 230901039**

**Date :04-03-2025**

**AIM:**

To do data cleaning and preparation using Excel data.

**SOFTWARE USED:**

Jupyter notebook

**DESCRIPTION:**

1. Read the Excel file.
2. Do data cleaning process.
3. Write the updated data set into the excel file.

**PROGRAM:**

**import** pandas **as** pd

pd**.**set\_option('display.max\_columns',10)

print("Original data:\n")

a**=**pd**.**read\_excel("/Users/student/Downloads/Marks.xlsx")

print(a)

f**=**a['Name']**.**str**.**strip()

f**.**to\_excel("/Users/student/Downloads/Name.xlsx")

print(f)

print("\nReplace Value:\n")

k**=**a**.**fillna(method**=**'pad')

print(k)

*#print("\nUpdated list:\n")*

k**.**to\_excel("Updated.xlsx")

print("\nDrop one particular column and its values:\n")

k**.**drop(['Social'],axis**=**1,inplace**=True**)

k**.**to\_excel("Updated.xlsx")

print(k)

print("\nDrop NaN rows:\n")

q**=**pd**.**read\_excel("/Users/student/Downloads/Marks.xlsx")

print(q)

x**=**q**.**dropna()

print("\nUpdated list:\n")

x**.**to\_excel("/Users/student/Downloads/Marks.xlsx")

print(x)

**import** pandas **as** pd

print("\nReplace values:\n")

n**=**pd**.**read\_excel("/Users/student/Downloads/Marks.xlsx")

print("Original data:\n")

print(n)

y**=**n**.**replace({89:88})

print("\nReplace dataset:\n")

print(y)

y**.**to\_excel("Updated.xlsx")

print(n)

print("\nExtracting isin() function:\n")

new**=**n['English']**.**isin([88])

print(n[new])

**OUTPUT:**

Original data:

Roll no Name English Tamil Maths Science Social Total

0 1 A 90.0 90.0 99.0 100.0 89.0 468

1 2 B 99.0 90.0 79.0 89.0 70.0 427

2 3 C 89.0 90.0 79.0 89.0 70.0 417

3 4 D 79.0 90.0 79.0 89.0 70.0 407

4 5 E 88.0 90.0 79.0 89.0 70.0 416

5 6 F 90.0 90.0 79.0 89.0 70.0 418

6 7 G 98.0 90.0 79.0 89.0 70.0 426

7 8 H 98.0 90.0 79.0 89.0 70.0 426

8 9 I 99.0 90.0 79.0 89.0 70.0 427

9 10 J 79.0 90.0 79.0 89.0 70.0 407

10 11 K 80.0 90.0 79.0 89.0 70.0 408

11 12 L NaN NaN NaN NaN NaN 0

12 13 M 88.0 90.0 79.0 89.0 70.0 416

13 14 N 99.0 90.0 79.0 89.0 70.0 427

14 15 O 79.0 90.0 79.0 89.0 70.0 407

15 16 P 79.0 90.0 79.0 89.0 70.0 407

16 17 Q 79.0 90.0 79.0 89.0 70.0 407

17 18 R 79.0 90.0 79.0 89.0 70.0 407

18 19 S 79.0 90.0 79.0 89.0 70.0 407

19 20 T 79.0 90.0 79.0 89.0 70.0 407

0 A

1 B

2 C

3 D

4 E

5 F

6 G

7 H

8 I

9 J

10 K

11 L

12 M

13 N

14 O

15 P

16 Q

17 R

18 S

19 T

Name: Name, dtype: object

Replace Value:

Roll no Name English Tamil Maths Science Social Total

0 1 A 90.0 90.0 99.0 100.0 89.0 468

1 2 B 99.0 90.0 79.0 89.0 70.0 427

2 3 C 89.0 90.0 79.0 89.0 70.0 417

3 4 D 79.0 90.0 79.0 89.0 70.0 407

4 5 E 88.0 90.0 79.0 89.0 70.0 416

5 6 F 90.0 90.0 79.0 89.0 70.0 418

6 7 G 98.0 90.0 79.0 89.0 70.0 426

7 8 H 98.0 90.0 79.0 89.0 70.0 426

8 9 I 99.0 90.0 79.0 89.0 70.0 427

9 10 J 79.0 90.0 79.0 89.0 70.0 407

10 11 K 80.0 90.0 79.0 89.0 70.0 408

11 12 L 80.0 90.0 79.0 89.0 70.0 0

12 13 M 88.0 90.0 79.0 89.0 70.0 416

13 14 N 99.0 90.0 79.0 89.0 70.0 427

14 15 O 79.0 90.0 79.0 89.0 70.0 407

15 16 P 79.0 90.0 79.0 89.0 70.0 407

16 17 Q 79.0 90.0 79.0 89.0 70.0 407

17 18 R 79.0 90.0 79.0 89.0 70.0 407

18 19 S 79.0 90.0 79.0 89.0 70.0 407

19 20 T 79.0 90.0 79.0 89.0 70.0 407

Drop one particular column and its values:

Roll no Name English Tamil Maths Science Total

0 1 A 90.0 90.0 99.0 100.0 468

1 2 B 99.0 90.0 79.0 89.0 427

2 3 C 89.0 90.0 79.0 89.0 417

3 4 D 79.0 90.0 79.0 89.0 407

4 5 E 88.0 90.0 79.0 89.0 416

5 6 F 90.0 90.0 79.0 89.0 418

6 7 G 98.0 90.0 79.0 89.0 426

7 8 H 98.0 90.0 79.0 89.0 426

8 9 I 99.0 90.0 79.0 89.0 427

9 10 J 79.0 90.0 79.0 89.0 407

10 11 K 80.0 90.0 79.0 89.0 408

11 12 L 80.0 90.0 79.0 89.0 0

12 13 M 88.0 90.0 79.0 89.0 416

13 14 N 99.0 90.0 79.0 89.0 427

14 15 O 79.0 90.0 79.0 89.0 407

15 16 P 79.0 90.0 79.0 89.0 407

16 17 Q 79.0 90.0 79.0 89.0 407

17 18 R 79.0 90.0 79.0 89.0 407

18 19 S 79.0 90.0 79.0 89.0 407

19 20 T 79.0 90.0 79.0 89.0 407

Drop NaN rows:

Roll no Name English Tamil Maths Science Social Total

0 1 A 90.0 90.0 99.0 100.0 89.0 468

1 2 B 99.0 90.0 79.0 89.0 70.0 427

2 3 C 89.0 90.0 79.0 89.0 70.0 417

3 4 D 79.0 90.0 79.0 89.0 70.0 407

4 5 E 88.0 90.0 79.0 89.0 70.0 416

5 6 F 90.0 90.0 79.0 89.0 70.0 418

6 7 G 98.0 90.0 79.0 89.0 70.0 426

7 8 H 98.0 90.0 79.0 89.0 70.0 426

8 9 I 99.0 90.0 79.0 89.0 70.0 427

9 10 J 79.0 90.0 79.0 89.0 70.0 407

10 11 K 80.0 90.0 79.0 89.0 70.0 408

11 12 L NaN NaN NaN NaN NaN 0

12 13 M 88.0 90.0 79.0 89.0 70.0 416

13 14 N 99.0 90.0 79.0 89.0 70.0 427

14 15 O 79.0 90.0 79.0 89.0 70.0 407

15 16 P 79.0 90.0 79.0 89.0 70.0 407

16 17 Q 79.0 90.0 79.0 89.0 70.0 407

17 18 R 79.0 90.0 79.0 89.0 70.0 407

18 19 S 79.0 90.0 79.0 89.0 70.0 407

19 20 T 79.0 90.0 79.0 89.0 70.0 407

Updated list:

Roll no Name English Tamil Maths Science Social Total

0 1 A 90.0 90.0 99.0 100.0 89.0 468

1 2 B 99.0 90.0 79.0 89.0 70.0 427

2 3 C 89.0 90.0 79.0 89.0 70.0 417

3 4 D 79.0 90.0 79.0 89.0 70.0 407

4 5 E 88.0 90.0 79.0 89.0 70.0 416

5 6 F 90.0 90.0 79.0 89.0 70.0 418

6 7 G 98.0 90.0 79.0 89.0 70.0 426

7 8 H 98.0 90.0 79.0 89.0 70.0 426

8 9 I 99.0 90.0 79.0 89.0 70.0 427

9 10 J 79.0 90.0 79.0 89.0 70.0 407

10 11 K 80.0 90.0 79.0 89.0 70.0 408

12 13 M 88.0 90.0 79.0 89.0 70.0 416

13 14 N 99.0 90.0 79.0 89.0 70.0 427

14 15 O 79.0 90.0 79.0 89.0 70.0 407

15 16 P 79.0 90.0 79.0 89.0 70.0 407

16 17 Q 79.0 90.0 79.0 89.0 70.0 407

17 18 R 79.0 90.0 79.0 89.0 70.0 407

18 19 S 79.0 90.0 79.0 89.0 70.0 407

19 20 T 79.0 90.0 79.0 89.0 70.0 407

Replace values:

Original data:

Unnamed: 0 Roll no Name English Tamil Maths Science Social Total

0 0 1 A 90 90 99 100 89 468

1 1 2 B 99 90 79 89 70 427

2 2 3 C 89 90 79 89 70 417

3 3 4 D 79 90 79 89 70 407

4 4 5 E 88 90 79 89 70 416

5 5 6 F 90 90 79 89 70 418

6 6 7 G 98 90 79 89 70 426

7 7 8 H 98 90 79 89 70 426

8 8 9 I 99 90 79 89 70 427

9 9 10 J 79 90 79 89 70 407

10 10 11 K 80 90 79 89 70 408

11 12 13 M 88 90 79 89 70 416

12 13 14 N 99 90 79 89 70 427

13 14 15 O 79 90 79 89 70 407

14 15 16 P 79 90 79 89 70 407

15 16 17 Q 79 90 79 89 70 407

16 17 18 R 79 90 79 89 70 407

17 18 19 S 79 90 79 89 70 407

18 19 20 T 79 90 79 89 70 407

Replace dataset:

Unnamed: 0 Roll no Name English Tamil Maths Science Social Total

0 0 1 A 90 90 99 100 88 468

1 1 2 B 99 90 79 88 70 427

2 2 3 C 88 90 79 88 70 417

3 3 4 D 79 90 79 88 70 407

4 4 5 E 88 90 79 88 70 416

5 5 6 F 90 90 79 88 70 418

6 6 7 G 98 90 79 88 70 426

7 7 8 H 98 90 79 88 70 426

8 8 9 I 99 90 79 88 70 427

9 9 10 J 79 90 79 88 70 407

10 10 11 K 80 90 79 88 70 408

11 12 13 M 88 90 79 88 70 416

12 13 14 N 99 90 79 88 70 427

13 14 15 O 79 90 79 88 70 407

14 15 16 P 79 90 79 88 70 407

15 16 17 Q 79 90 79 88 70 407

16 17 18 R 79 90 79 88 70 407

17 18 19 S 79 90 79 88 70 407

18 19 20 T 79 90 79 88 70 407

Unnamed: 0 Roll no Name English Tamil Maths Science Social Total

0 0 1 A 90 90 99 100 89 468

1 1 2 B 99 90 79 89 70 427

2 2 3 C 89 90 79 89 70 417

3 3 4 D 79 90 79 89 70 407

4 4 5 E 88 90 79 89 70 416

5 5 6 F 90 90 79 89 70 418

6 6 7 G 98 90 79 89 70 426

7 7 8 H 98 90 79 89 70 426

8 8 9 I 99 90 79 89 70 427

9 9 10 J 79 90 79 89 70 407

10 10 11 K 80 90 79 89 70 408

11 12 13 M 88 90 79 89 70 416

12 13 14 N 99 90 79 89 70 427

13 14 15 O 79 90 79 89 70 407

14 15 16 P 79 90 79 89 70 407

15 16 17 Q 79 90 79 89 70 407

16 17 18 R 79 90 79 89 70 407

17 18 19 S 79 90 79 89 70 407

18 19 20 T 79 90 79 89 70 407

Extracting isin() function:

Unnamed: 0 Roll no Name English Tamil Maths Science Social Total

4 4 5 E 88 90 79 89 70 416

11 12 13 M 88 90 79 89 70 416

**Result:**

Thus data cleaning and preparation was done using excel.