Python Coding Challenge

Execute Data cleaning programs & Panda joins in Python Execute Data cleaning 3 programs:

Import Data:

```
import pandas as pd

df = pd.read csv("products.csv.txt")
```

Program 1: Handling Missing Values

```
print("Missing values before cleaning:\n", df.isnull().sum())

df['product_name'] = df['product_name'].fillna("Unknown")

df['product_price'] = df['product_price'].fillna(df['product_price'].mean())

df['product_color'] = df['product_color'].fillna(df['product_color'].mode()[0])

print("\n After handling missing values:\n", df)
```

```
PS C:\Users\Bavatharani\OneDrive\Desktop\Python> & C:\Users\Bavatharani\AppData/Local/Program
 /Python/Python311/python.exe "c:/Users/Bavatharani/OneDrive/Desktop/Python/Coding Challenge.py
• Missing values before cleaning:
  product id
                  0
 product name
 product_price
 product_color
                 4
 dtype: int64
  After handling missing values:
      product id product name product price product color
                    Dress A
                                    2500.0
                    Dress B
                                    2770.0
                                                    Blue
                                                    Blue
                    Unknown
                                    1800.0
                    Dress D
                                    3200.0
                                                   Green
             4
                                                  Yellow
                    Dress E
                                    2770.0
              6
                    Dress F
                                    2700.0
                                                    Blue
                                                     Red
                    Dress G
                                    3100.0
                                                    Blue
             8
                    Dress H
                                    2900.0
 8
                    Unknown
                                    2600.0
                                                   Black
             10
                    Dress J
                                    2770.0
                                                   White
 10
             11
                    Dress K
                                     3000.0
                                                    Pink
                     Dress L
                                     2770.0
                                                  Purple
                                    2400.0
                                                  Orange
                    Unknown
                                                    Blue
             14
                     Dress N
                                     3500.0
 13
 14
             15
                     Dress 0
                                    2770.0
                                                    Blue
```

Program 2: Removing Duplicate Records

```
print("Duplicates before cleaning:", df.duplicated().sum())
df = df.drop_duplicates()
print("Duplicates after cleaning:", df.duplicated().sum())
```

Output:

```
Duplicates before cleaning: 0
Duplicates after cleaning: 0
```

Program 3: Changing Data Types

```
print("Original data types:\n", df.dtypes)

df['product_id'] = df['product_id'].astype(int)

df['product_price'] = pd.to_numeric(df['product_price'])

print("\nData types after conversion:\n", df.dtypes)
```

```
Original data types:
 product id
                    int64
                  object
product name
product price
                 float64
product color
                  object
dtype: object
Data types after conversion:
 product id
                    int64
product name
                  object
product price
                 float64
product color
                  object
dtype: object
```

Joins

```
import pandas as pd
# Step 1: Load both CSV files
products_df = pd.read_csv("products.csv.txt")  # Make sure file name is correct
stock_df = pd.read_csv("products_stock.csv.txt")  # Related stock file
products_df.rename(columns={'product_id': 'id'}, inplace=True)
stock_df.rename(columns={'product_id': 'id'}, inplace=True)

# Step 2: INNER JOIN
inner_join = pd.merge(products_df, stock_df, on='id', how='inner')
print("\n INNER JOIN:\n", inner_join)
print("• Returns only rows with matching id in both files.\n")
```

id 1	product name				
1		product_price	<pre>product_color</pre>	stock_qty	warehouse
	Dress A	2500.0	Red	50	Chennai
2	Dress B	NaN	Blue	20	Mumbai
3	NaN	1800.0	NaN	0	Delhi
4	Dress D	3200.0	Green	35	Bangalore
5	Dress E	NaN	Yellow	10	Hyderabad
6	Dress F	2700.0	Blue	25	Chennai
7	Dress G	3100.0	Red	15	Pune
8	Dress H	2900.0	NaN	60	Mumbai
9	NaN	2600.0	Black	40	Delhi
10	Dress J	NaN	White	0	Kolkata
11	Dress K	3000.0	Pink	12	Bangalore
12	Dress L	NaN	Purple	30	Hyderabad
13	NaN	2400.0	Orange	22	Chennai
14	Dress N	3500.0	NaN	18	Pune
15	Dress 0	NaN	NaN	0	Ahmedabad
	4 5 6 7 8 9 10 11 12 13 14	4 Dress D 5 Dress E 6 Dress F 7 Dress G 8 Dress H 9 NaN 10 Dress J 11 Dress K 12 Dress L 13 NaN 14 Dress N 15 Dress O	4 Dress D 3200.0 5 Dress E NaN 6 Dress F 2700.0 7 Dress G 3100.0 8 Dress H 2900.0 9 NaN 2600.0 10 Dress J NaN 11 Dress K 3000.0 12 Dress L NaN 13 NaN 2400.0 14 Dress N 3500.0 15 Dress O NaN	4 Dress D 3200.0 Green 5 Dress E NaN Yellow 6 Dress F 2700.0 Blue 7 Dress G 3100.0 Red 8 Dress H 2900.0 NaN 9 NaN 2600.0 Black 10 Dress J NaN White 11 Dress K 3000.0 Pink 12 Dress L NaN Purple 13 NaN 2400.0 Orange 14 Dress N 3500.0 NaN 15 Dress O NaN NaN	4 Dress D 3200.0 Green 35 5 Dress E NaN Yellow 10 6 Dress F 2700.0 Blue 25 7 Dress G 3100.0 Red 15 8 Dress H 2900.0 NaN 60 9 NaN 2600.0 Black 40 10 Dress J NaN White 0 11 Dress K 3000.0 Pink 12 12 Dress L NaN Purple 30 13 NaN 2400.0 Orange 22 14 Dress N 3500.0 NaN 18

```
# Step 3: LEFT JOIN
left_join = pd.merge(products_df, stock_df, on='id', how='left')
print(" LEFT JOIN:\n", left_join)
print("• All products retained, and matching stock info added.\n")
```

Output:

```
LEFT JOTN:
     id product name product price product color stock qty warehouse
                            2500.0
                                             Red
                                                         50
                                                              Chennai
           Dress B
                                            Blue
                                                               Mumbai
1
                              NaN
                                                         20
               NaN
                            1800.0
                                            NaN
                                                                 Delhi
    4
           Dress D
                            3200.0
                                          Green
                                                         35 Bangalore
            Dress E
                             NaN
                                          Yellow
                                                         10 Hyderabad
    6
           Dress F
                            2700.0
                                           Blue
                                                         25
                                                              Chennai
6
           Dress G
                           3100.0
                                            Red
                                                                  Pune
    8
           Dress H
                            2900.0
                                            NaN
                                                        60
                                                               Mumbai
               NaN
                            2600.0
                                           Black
                                                         40
                                                                Delhi
9
    10
           Dress J
                                          White
                                                         0
                                                               Kolkata
10
           Dress K
                            3000.0
                                           Pink
                                                         12 Bangalore
11
   12
            Dress L
                              NaN
                                          Purple
                                                         30 Hyderabad
12
   13
                            2400.0
                                                         22
                                                              Chennai
               NaN
                                          Orange
            Dress N
                            3500.0
                                                         18
                                                                  Pune
14
   15
            Dress 0
                              NaN
                                             NaN
                                                         Ø Ahmedabad

    All products retained, and matching stock info added.
```

```
# Step 4: RIGHT JOIN

right_join = pd.merge(products_df, stock_df, on='id', how='right')

print("RIGHT JOIN:\n", right_join)

print("• All stock entries retained, even if product is missing.\n")
```

```
RIGHT JOIN:
     id product_name
                      product_price product_color stock_qty
                                                               warehouse
                            2500.0
                                             Red
                                                                Chennai
0
            Dress A
                                                          50
            Dress B
                               NaN
                                            Blue
                                                          20
                                                                 Mumbai
     2
                NaN
                            1800.0
                                             NaN
                                                          0
                                                                  Delhi
            Dress D
                            3200.0
                                           Green
                                                              Bangalore
                                                              Hyderabad
    5
                                          Yellow.
4
           Dress E
                               NaN
                                                          10
    6
           Dress F
                            2700.0
                                            Blue
                                                          25
                                                                Chennai
            Dress G
                            3100.0
                                             Red
                                                          15
                                                                   Pune
                                                          60
    8
            Dress H
                            2900.0
                                             NaN
                                                                 Mumbai
8
    9
                                                          40
                                                                 Delhi
                NaN
                            2600.0
                                           Black
    10
           Dress J
                                           White
                                                          0
                                                                Kolkata
                               NaN
10
    11
                                            Pink
                                                              Bangalore
            Dress K
                            3000.0
                                                          12
11
    12
            Dress L
                               NaN
                                           Purple
                                                          30
                                                              Hyderabad
12
    13
                            2400.0
                                                          22
                                                                Chennai
                NaN
                                          Orange
            Dress N
                            3500.0
                                                                   Pune
13
    14
                                              NaN
                                                          18
                                                              Ahmedabad
    15
            Dress O
                               NaN
                                             NaN
 All stock entries retained, even if product is missing.
```

```
# Step 5: OUTER JOIN
outer_join = pd.merge(products_df, stock_df, on='id', how='outer')
print("OUTER JOIN:\n", outer_join)
```

print("• All records from both files shown; unmatched ones will have NaN.\n")

	id	product name	product price	product color	stock qty	warehouse
0	1	Dress A	2500.0	Red	50	Chennai
1	2	Dress B	NaN	Blue	20	Mumbai
2	3	NaN	1800.0	NaN	0	Delhi
3	4	Dress D	3200.0	Green	35	Bangalore
4	5	Dress E	NaN	Yellow	10	Hyderabad
5	6	Dress F	2700.0	Blue	25	Chennai
6	7	Dress G	3100.0	Red	15	Pune
7	8	Dress H	2900.0	NaN	60	Mumbai
8	9	NaN	2600.0	Black	40	Delhi
9	10	Dress J	NaN	White	0	Kolkata
10	11	Dress K	3000.0	Pink	12	Bangalore
11	12	Dress L	NaN	Purple	30	Hyderabad
12	13	NaN	2400.0	Orange	22	Chennai
13	14	Dress N	3500.0	NaN	18	Pune
14	15	Dress O	NaN	NaN	0	Ahmedabad