PYTHON CODING ASSESSMENT

Execute Data cleaning programs & Panda joins in Python

Execute Data cleaning 3 programs:

IMPORT DATA:

```
import pandas as pd

# Load the CSV file

df = pd.read_csv("/content/products.csv")

df.head()
```

```
# Load the CSV file
df = pd.read_csv("/content/products.csv")
df.head()
```

	id	product_name	product_price	product_material	product_color
0	1	Intelligent Fresh Chips	655.0	Concrete	mint green
1	2	Practical Fresh Sausages	911.0	Cotton	indigo
2	3	Refined Steel Car	690.0	Rubber	gold
3	4	Gorgeous Plastic Pants	492.0	Soft	plum
4	5	Sleek Cotton Chair	33.0	Fresh	black

Program 1: Handling Missing Values

```
# Show missing values before cleaning

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


# Fill missing product_price with mean

df['product_price'].fillna(df['product_price'].mean(), inplace=True)


# Fill missing product_color with mode
```

```
df['product_color'].fillna(df['product_color'].mode()[0], inplace=True)

# Display after filling
print("\nAfter handling missing values:\n", df.isnull().sum())
print(df.head())
```

```
Missing values before cleaning:
 id                0
product_name       0
product_price      1
product_material   0
product_color      1
dtype: int64

After handling missing values:
 id                0
product_name       0
product_price      0
product_material   0
product_color      0
dtype: int64
   id             product_name  product_price product_material product_color
0   1   Intelligent Fresh Chips          655.0         Concrete    mint green
1   2  Practical Fresh Sausages          911.0           Cotton        indigo
2   3         Refined Steel Car          690.0           Rubber          gold
3   4    Gorgeous Plastic Pants          492.0             Soft          plum
4   5        Sleek Cotton Chair           33.0            Fresh         black
```

Program 2: Removing Duplicate Records

```
# Check duplicates before cleaning
print("Duplicates before cleaning:", df.duplicated().sum())

# Drop duplicate rows
df = df.drop_duplicates()

# Check again
print("Duplicates after cleaning:", df.duplicated().sum())
```

```
Duplicates before cleaning: 1
Duplicates after cleaning: 0
```

Program 3: Changing Data Types

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

# Convert id to integer
df['id'] = df['id'].astype(int)

# Ensure product_price is float
df['product_price'] = pd.to_numeric(df['product_price'])

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

```
Original data types:
 id                    int64
product_name          object
product_price        float64
product_material      object
product_color         object
dtype: object

Data types after conversion:
 id                    int64
product_name          object
product_price        float64
product_material      object
product_color         object
dtype: object
```

Panda joins in Python

IMPORT DATA:

```
import pandas as pd
# Load both CSV files
products_df = pd.read_csv("/content/products.csv")
stock_df = pd.read_csv("/content/product_stock.csv")
# Preview data
print("Products:\n", products_df.head())
print("\nStock Info:\n", stock_df.head())
```

```
Products:
    id              product_name  product_price product_material product_color
0   1   Intelligent Fresh Chips          655.0         Concrete    mint green
1   2  Practical Fresh Sausages          911.0           Cotton        indigo
2   3         Refined Steel Car          690.0           Rubber          gold
3   4    Gorgeous Plastic Pants          492.0             Soft          plum
4   5        Sleek Cotton Chair           33.0            Fresh         black

Stock Info:
    id  stock_available  supplier
0   1             True   ABC Ltd
1   2            False  XYZ Corp
2   3             True   ABC Ltd
3   4             True   LMN Inc
4   6            False  XYZ Corp
```

INNER JOIN

```
inner_join = pd.merge(products_df, stock_df, on='id', how='inner')
print("\nINNER JOIN:\n", inner_join)
```

- Returns only rows with matching id in both files.

```
INNER JOIN:
    id              product_name  product_price product_material product_color  \
0   1   Intelligent Fresh Chips          655.0         Concrete    mint green
1   2  Practical Fresh Sausages          911.0           Cotton         indigo
2   3         Refined Steel Car          690.0           Rubber           gold
3   4    Gorgeous Plastic Pants          492.0             Soft           plum
4   6     Awesome Wooden Towels          474.0          Plastic         orange
5   8      Incredible Steel Hat           78.0           Rubber         violet
6  10      Generic Wooden Pizza           84.0           Frozen         indigo
7  12   Unbranded Plastic Salad           89.0           Wooden           pink
8  15      Ergonomic Cotton Hat           43.0           Rubber    mint green
9  15      Ergonomic Cotton Hat           43.0           Rubber    mint green

   stock_available      supplier
0             True       ABC Ltd
1            False       XYZ Corp
2             True       ABC Ltd
3             True       LMN Inc
4            False       XYZ Corp
5             True  Global Trade
6            False       ABC Ltd
7             True       LMN Inc
8            False       XYZ Corp
9            False       XYZ Corp
```

LEFT JOIN

left_join = pd.merge(products_df, stock_df, on='id', how='left')

print("\nLEFT JOIN:\n", left_join)

- All products retained, and matching stock info added.

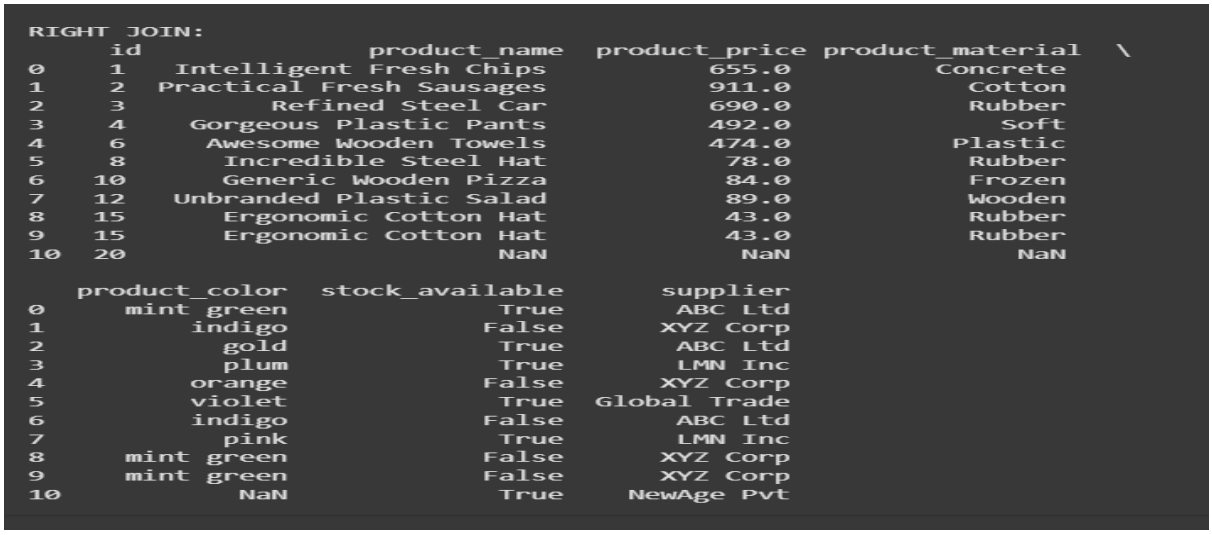
```
LEFT JOIN:
     id              product_name  product_price product_material  \
0    1   Intelligent Fresh Chips          655.0         Concrete
1    2  Practical Fresh Sausages          911.0           Cotton
2    3         Refined Steel Car          690.0           Rubber
3    4    Gorgeous Plastic Pants          492.0             Soft
4    5        Sleek Cotton Chair           33.0            Fresh
5    6     Awesome Wooden Towels          474.0          Plastic
6    7      Practical Soft Shoes          500.0           Rubber
7    8      Incredible Steel Hat           78.0           Rubber
8    9       Awesome Wooden Ball           28.0             Soft
9   10      Generic Wooden Pizza           84.0           Frozen
10  11   Unbranded Wooden Cheese           26.0             Soft
11  12   Unbranded Plastic Salad           89.0           Wooden
12  13  Gorgeous Cotton Keyboard           37.0         Concrete
13  14    Incredible Steel Shirt            NaN            Metal
14  15      Ergonomic Cotton Hat           43.0           Rubber
15  15      Ergonomic Cotton Hat           43.0           Rubber

   product_color stock_available      supplier
0     mint green             True       ABC Ltd
1         indigo            False       XYZ Corp
2           gold             True       ABC Ltd
3           plum             True       LMN Inc
4          black              NaN           NaN
5         orange            False       XYZ Corp
6           pink              NaN           NaN
7         violet             True  Global Trade
8            NaN              NaN           NaN
9         indigo            False       ABC Ltd
10         black              NaN           NaN
11           pink             True       LMN Inc
12       sky blue              NaN           NaN
13          white              NaN           NaN
14    mint green            False       XYZ Corp
15    mint green            False       XYZ Corp
```

RIGHT JOIN

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

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

- All stock entries retained, even if product is missing.



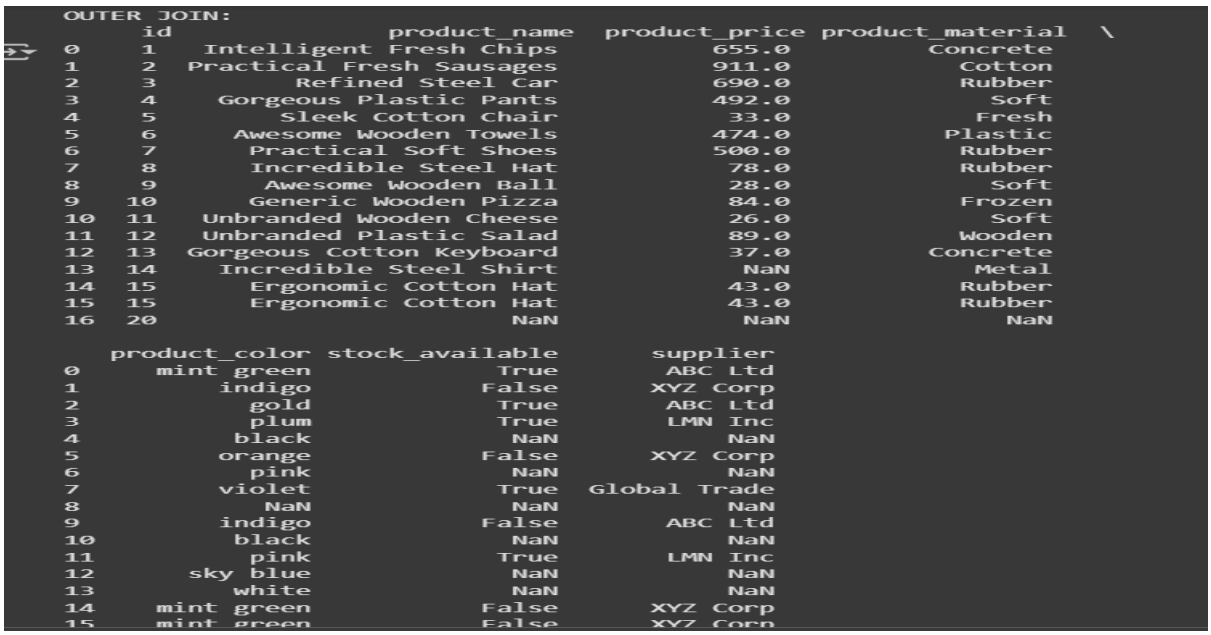
```
RIGHT JOIN:
      id                product_name  product_price product_material  \
0     1   Intelligent Fresh Chips           655.0          Concrete
1     2  Practical Fresh Sausages           911.0            Cotton
2     3          Refined Steel Car           690.0            Rubber
3     4    Gorgeous Plastic Pants            492.0              Soft
4     6     Awesome Wooden Towels           474.0           Plastic
5     8       Incredible Steel Hat            78.0            Rubber
6    10      Generic Wooden Pizza             84.0            Frozen
7    12   Unbranded Plastic Salad             89.0            Wooden
8    15       Ergonomic Cotton Hat            43.0            Rubber
9    15       Ergonomic Cotton Hat            43.0            Rubber
10   20                        NaN             NaN               NaN

    product_color  stock_available       supplier
0      mint green             True       ABC Ltd
1          indigo            False       XYZ Corp
2            gold             True       ABC Ltd
3            plum             True       LMN Inc
4          orange            False       XYZ Corp
5          violet             True  Global Trade
6          indigo            False       ABC Ltd
7            pink             True       LMN Inc
8      mint green            False       XYZ Corp
9      mint green            False       XYZ Corp
10            NaN             True    NewAge Pvt
```

OUTER JOIN

outer_join = pd.merge(products_df, stock_df, on='id', how='outer')

print("\nOUTER JOIN:\n", outer_join)

- All records from both files shown; unmatched ones will have NaN.



```
OUTER JOIN:
      id                product_name  product_price product_material  \
0     1   Intelligent Fresh Chips           655.0          Concrete
1     2  Practical Fresh Sausages           911.0            Cotton
2     3          Refined Steel Car           690.0            Rubber
3     4    Gorgeous Plastic Pants            492.0              Soft
4     5         Sleek Cotton Chair            33.0             Fresh
5     6     Awesome Wooden Towels           474.0           Plastic
6     7      Practical Soft Shoes           500.0            Rubber
7     8       Incredible Steel Hat            78.0            Rubber
8     9       Awesome Wooden Ball             28.0              Soft
9    10      Generic Wooden Pizza             84.0            Frozen
10   11   Unbranded Wooden Cheese             26.0              Soft
11   12   Unbranded Plastic Salad             89.0            Wooden
12   13  Gorgeous Cotton Keyboard             37.0          Concrete
13   14    Incredible Steel Shirt              NaN             Metal
14   15       Ergonomic Cotton Hat            43.0            Rubber
15   15       Ergonomic Cotton Hat            43.0            Rubber
16   20                        NaN             NaN               NaN

    product_color stock_available       supplier
0      mint green             True       ABC Ltd
1          indigo            False       XYZ Corp
2            gold             True       ABC Ltd
3            plum             True       LMN Inc
4           black              NaN            NaN
5          orange            False       XYZ Corp
6            pink              NaN            NaN
7          violet             True  Global Trade
8             NaN              NaN            NaN
9          indigo            False       ABC Ltd
10          black              NaN            NaN
11           pink             True       LMN Inc
12       sky blue              NaN            NaN
13          white              NaN            NaN
14     mint green            False       XYZ Corp
15     mint green            False       XYZ Corp
```