```
import pandas as pd

# Step 1: Create Sales Dataset with duplicate transaction entries
data = {
    "Transaction_ID": [101, 102, 103, 104, 101, 102],
    "Customer": ["Alice", "Bob", "Charlie", "David", "Alice", "Bob"],
    "Amount": [250, 300, 400, 150, 250, 300],
    "Date": ["2025-09-01", "2025-09-02", "2025-09-03", "2025-09-04",
             "2025-09-01", "2025-09-02"]
}

sales_df = pd.DataFrame(data)

# Step 2: Show dataset before removing duplicates
print("Dataset BEFORE removing duplicates:")
print(sales_df)
print("\nDuplicate count:", sales_df.duplicated().sum())

# Step 3: Remove duplicates
sales_df_cleaned = sales_df.drop_duplicates()

# Step 4: Show dataset after removing duplicates
print("\nDataset AFTER removing duplicates:")
print(sales_df_cleaned)
print("\nDuplicate count:", sales_df_cleaned.duplicated().sum())
```

```
Dataset BEFORE removing duplicates:
   Transaction_ID Customer  Amount        Date
0             101    Alice     250  2025-09-01
1             102      Bob     300  2025-09-02
2             103  Charlie     400  2025-09-03
3             104    David     150  2025-09-04
4             101    Alice     250  2025-09-01
5             102      Bob     300  2025-09-02

Duplicate count: 2

Dataset AFTER removing duplicates:
   Transaction_ID Customer  Amount        Date
0             101    Alice     250  2025-09-01
1             102      Bob     300  2025-09-02
2             103  Charlie     400  2025-09-03
3             104    David     150  2025-09-04

Duplicate count: 0
```

```
import pandas as pd

# Step 1: Create Employee Dataset with incorrect data types
data = {
    "Employee_ID": [1, 2, 3],
    "Name": ["John", "Sara", "Mike"],
    "Salary": ["50000", "60000", "55000"],  # stored as string
    "Joining_Date": ["2025-01-15", "2025-03-10", "2025-04-05"]  # stored as string
}

employee_df = pd.DataFrame(data)

# Step 2: Show data types before conversion
print("BEFORE Conversion:")
print(employee_df.dtypes)

# Step 3: Convert Salary to numeric and Joining_Date to datetime
employee_df["Salary"] = pd.to_numeric(employee_df["Salary"])
employee_df["Joining_Date"] = pd.to_datetime(employee_df["Joining_Date"])

# Step 4: Show data types after conversion
print("\nAFTER Conversion:")
print(employee_df.dtypes)

# Step 5: Show final dataset
print("\nCorrected Employee Dataset:")
print(employee_df)
```

```
BEFORE Conversion:
Employee_ID      int64
Name            object
Salary          object
Joining_Date    object
dtype: object
```

```
AFTER Conversion:
Employee_ID              int64
Name                    object
Salary                   int64
Joining_Date    datetime64[ns]
dtype: object

Corrected Employee Dataset:
   Employee_ID  Name  Salary Joining_Date
0            1  John   50000   2025-01-15
1            2  Sara   60000   2025-03-10
2            3  Mike   55000   2025-04-05
```