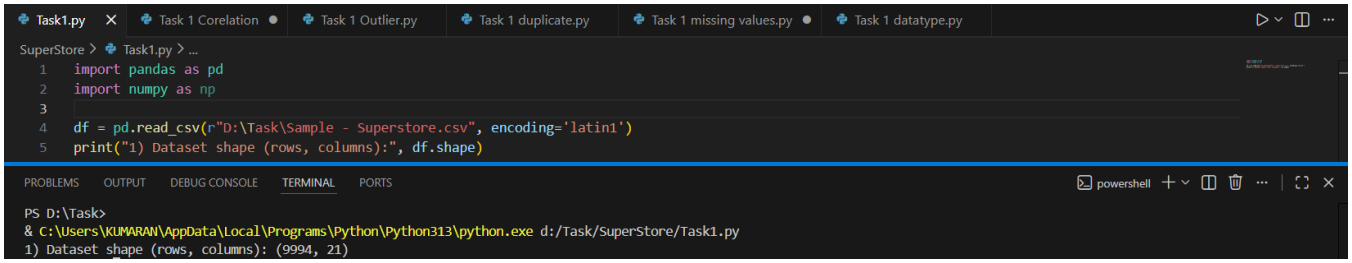
DATA PROFILING REPORT : SAMPLE SUPERSTORE DATASET

Dataset: Sample_Superstore.csv

1. Dataset Structure

- The Superstore dataset contains 9,994 rows and 21 columns.



2. Column Names & Data Types

- Row ID (int)
- Order ID (object)
- Order Date (object)
- Ship Date (object)
- Ship Mode (object)
- Customer ID (object)
- Customer Name (object)
- Segment (object)
- Country (object)
- City (object)
- State (object)
- Postal Code (int)
- Region (object)
- Product ID (object)
- Category (object)
- Sub-Category (object)
- Product Name (object)
- Sales (float)
- Quantity (int)
- Discount (float)
- Profit (float)

Task1.py | Task 1 Corelation | Task 1 Outlier.py | Task 1 duplicate.py | Task 1 missing values.py | Task 1 datatype.py

SuperStore > Task 1 datatype.py > ...

```python
import pandas as pd
import numpy as np

df = pd.read_csv(r"D:\Task\Sample - Superstore.csv", encoding='latin1')
print("\n2) Columns and Data Types:")
print(df.dtypes)

```

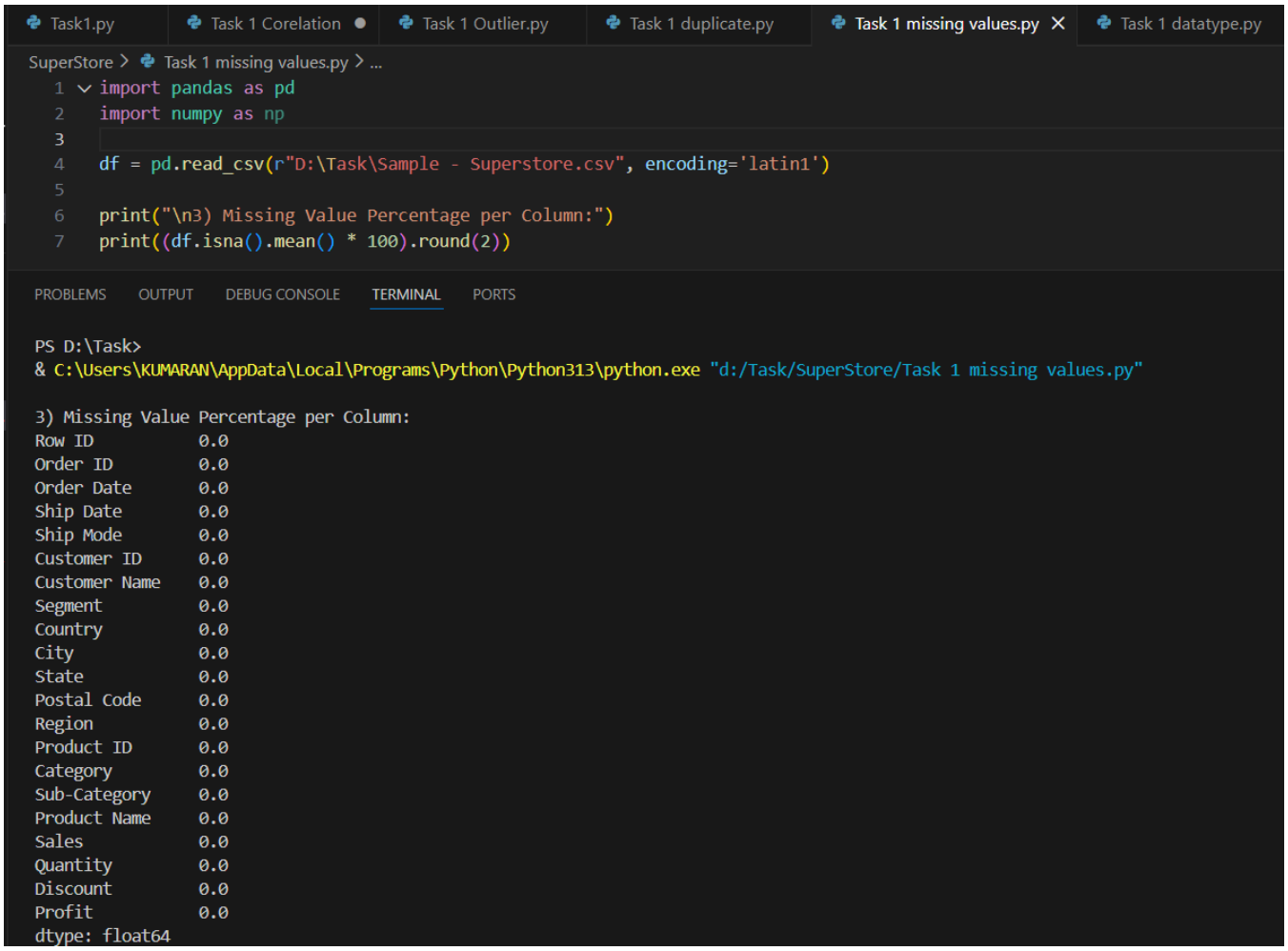
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS D:\Task>
& C:\Users\KUMARAN\AppData\Local\Programs\Python\Python313\python.exe "d:/Task/SuperStore/Task 1 datatype.py"

2) Columns and Data Types:
Row ID            int64
Order ID         object
Order Date       object
Ship Date        object
Ship Mode        object
Customer ID      object
Customer Name    object
Segment          object
Country          object
City             object
State            object
Postal Code       int64
Region           object
Product ID       object
Category         object
Sub-Category     object
Product Name     object
Sales           float64
Quantity          int64
Discount        float64
Profit          float64
dtype: object
```

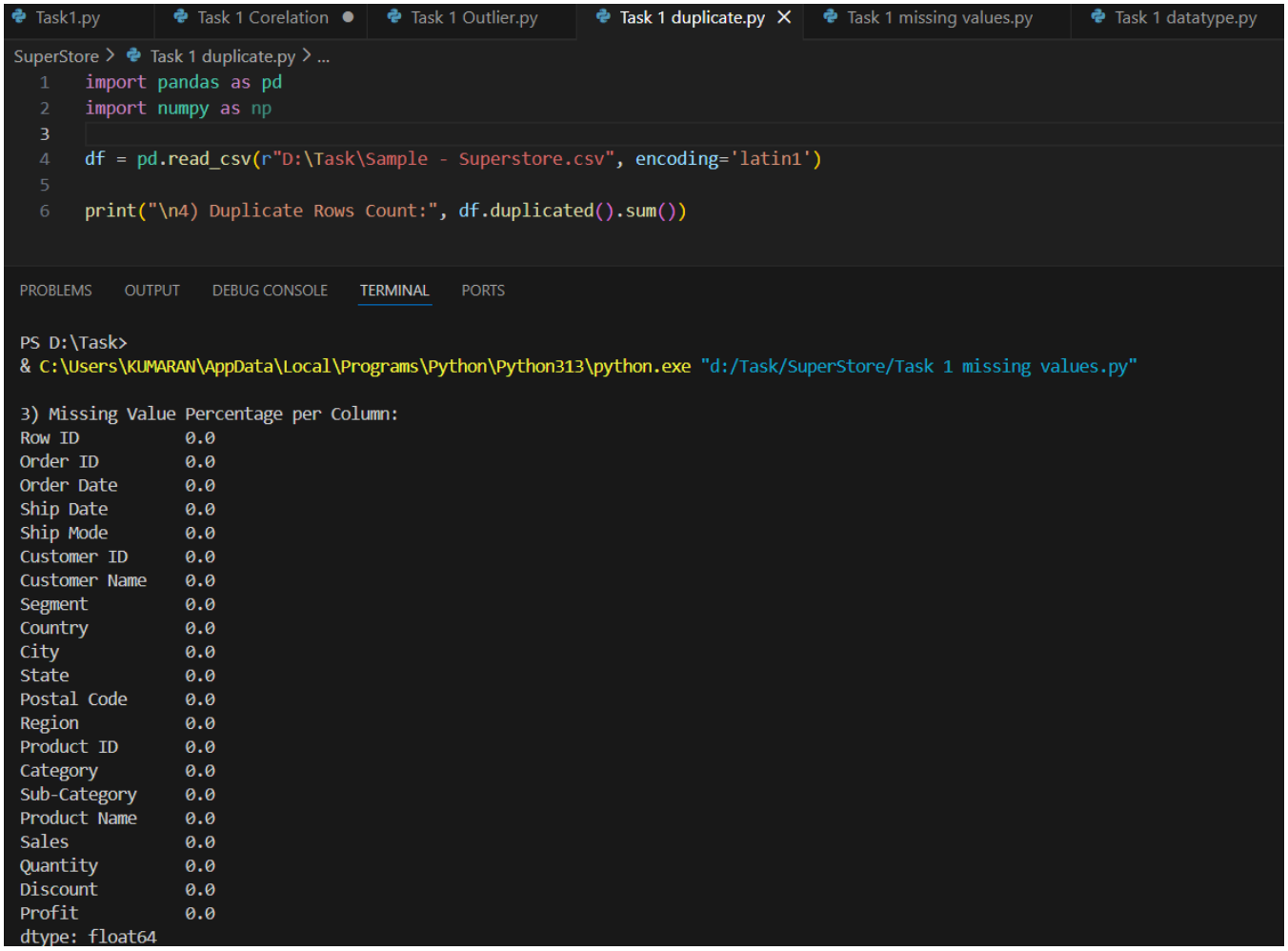
3. Missing Value Analysis

- The dataset contains 0% missing values across all columns.
- No missing values found in key analytical columns such as Sales, Profit, Quantity, Discount, Category, Segment, Region.



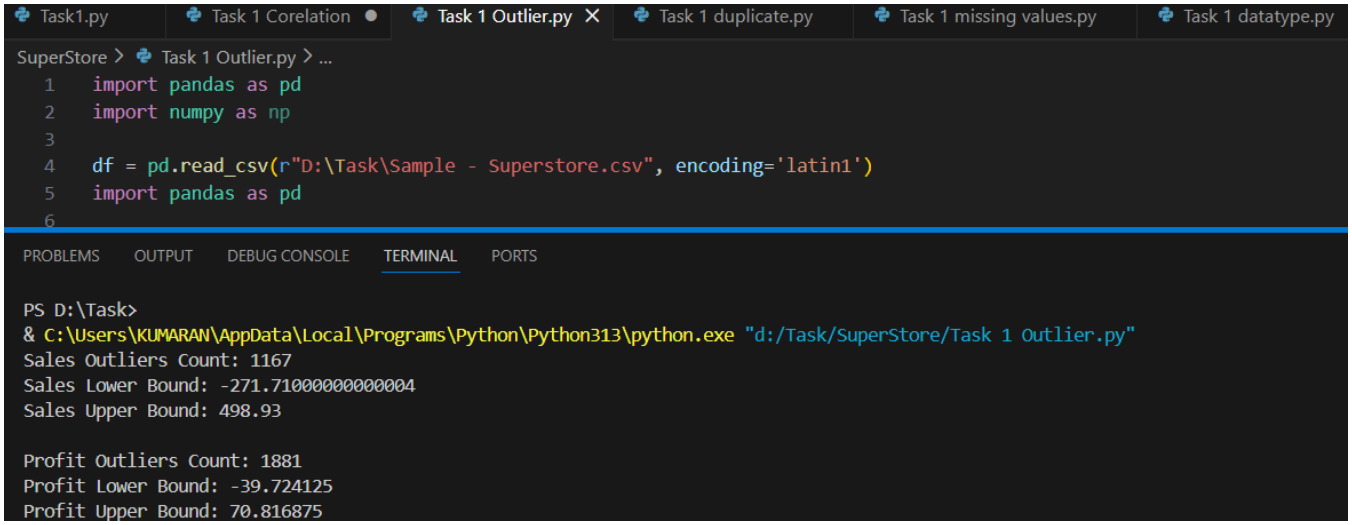
4. Duplicate Records

- Duplicate rows detected: 0
- Final dataset contains only unique order-line entries.



5. Outlier Analysis

- Sales Outliers: 1,167
- Profit Outliers: 1,881



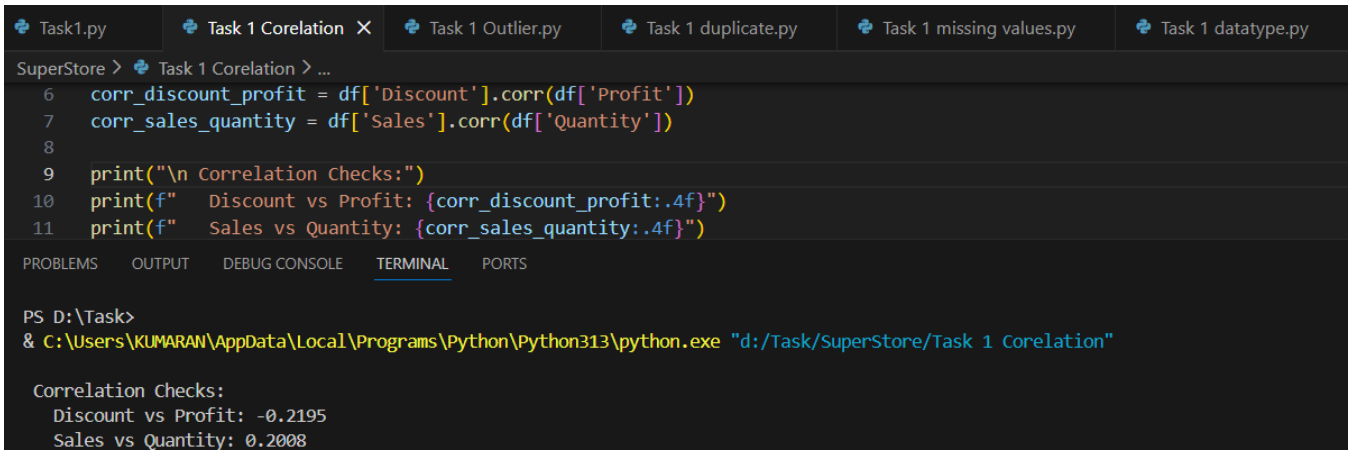
6. Relationship Analysis (Correlation)

Discount vs Profit (–0.21 correlation)

- Small negative correlation confirms that increasing discount generally reduces profitability.

Sales vs Quantity (+0.20 correlation)

- Positive but weak relationship: increasing quantity does not always mean higher sales value.



7. Summary of Insights

- Dataset is clean with no missing values or duplicates.
- Sales & Profit contain heavy outliers.
- Discounting reduces profitability.
- Quantity has weak influence on Sales.
- Profit volatility is high.