```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load your uploaded file (correct file name used)
df = pd.read_csv("superstore.csv.csv")

# Show the first few rows
df.head()
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	C
0	1	CA-2017-152156	08-11-2017	11-11-2017	Second Class	CG-12520	Claire Gute	Consumer	Spain	Barcelona	...	42420.0	South	FUR-BO-10001798	Furniture	Bo
1	2	CA-2017-152156	08-11-2017	11-11-2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420.0	South	FUR-CH-10000454	Furniture	
2	3	CA-2017-138688	12-06-2017	16-06-2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	90036.0	West	OFF-LA-10000240	Office Supplies	
3	4	US-2016-108966	11-10-2016	18-10-2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	Germany	Munich	...	33311.0	South	FUR-TA-10000577	Furniture	
4	5	US-2016-108966	11-10-2016	18-10-2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311.0	South	OFF-ST-10000760	Office Supplies	

5 rows × 21 columns

```
# Drop rows with missing or zero profit or quantity
df = df.dropna(subset=["Profit", "Quantity"])
df = df[(df["Profit"] != 0) & (df["Quantity"] != 0)]

# Quick check
df[["Quantity", "Profit"]].describe()
```

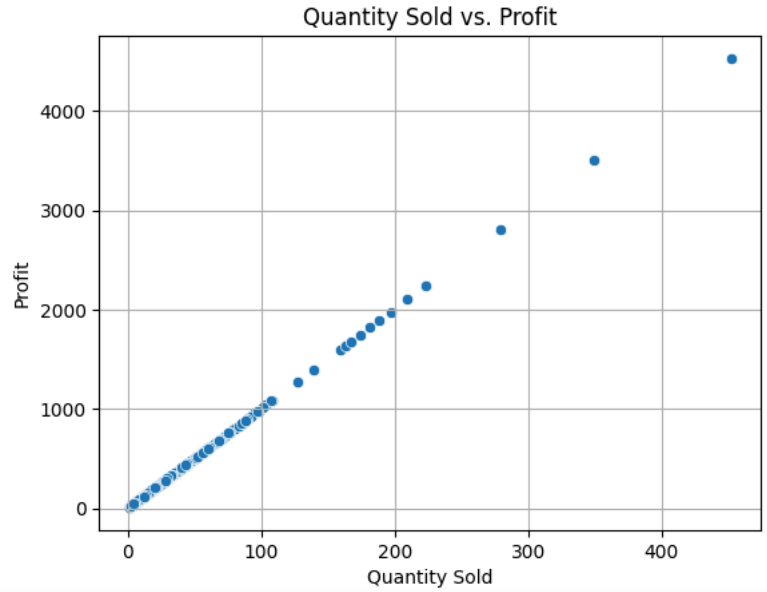
	Quantity	Profit
count	5046.000000	5046.000000
mean	8.106025	85.918852
std	16.472979	165.053795
min	1.000000	10.000000
25%	1.000000	19.216000
50%	4.000000	40.354400
75%	9.000000	92.762400
max	452.000000	4527.696000

```
# Correlation matrix
correlation = df[["Quantity", "Profit"]].corr()
print("Correlation between Quantity and Profit:")
print(correlation)

# Scatter plot
sns.scatterplot(x="Quantity", y="Profit", data=df)
plt.title("Quantity Sold vs. Profit")
plt.xlabel("Quantity Sold")
plt.ylabel("Profit")
plt.grid(True)
plt.show()
```

What can I help you build?

```
Correlation between Quantity and Profit:
            Quantity   Profit
Quantity    1.00000  0.99984
Profit      0.99984  1.00000
```




```
grouped = df.groupby("Sub-Category").agg({
    "Quantity": "sum",
    "Profit": "sum"
})
grouped["Profit_per_Unit"] = grouped["Profit"] / grouped["Quantity"]
grouped = grouped.sort_values("Profit_per_Unit")

grouped.plot(kind="barh", y="Profit_per_Unit", legend=False, figsize=(8, 6), title="Profit per Unit by Sub-Category")
plt.xlabel("Profit per Unit Sold")
plt.tight_layout()
plt.show()
```