

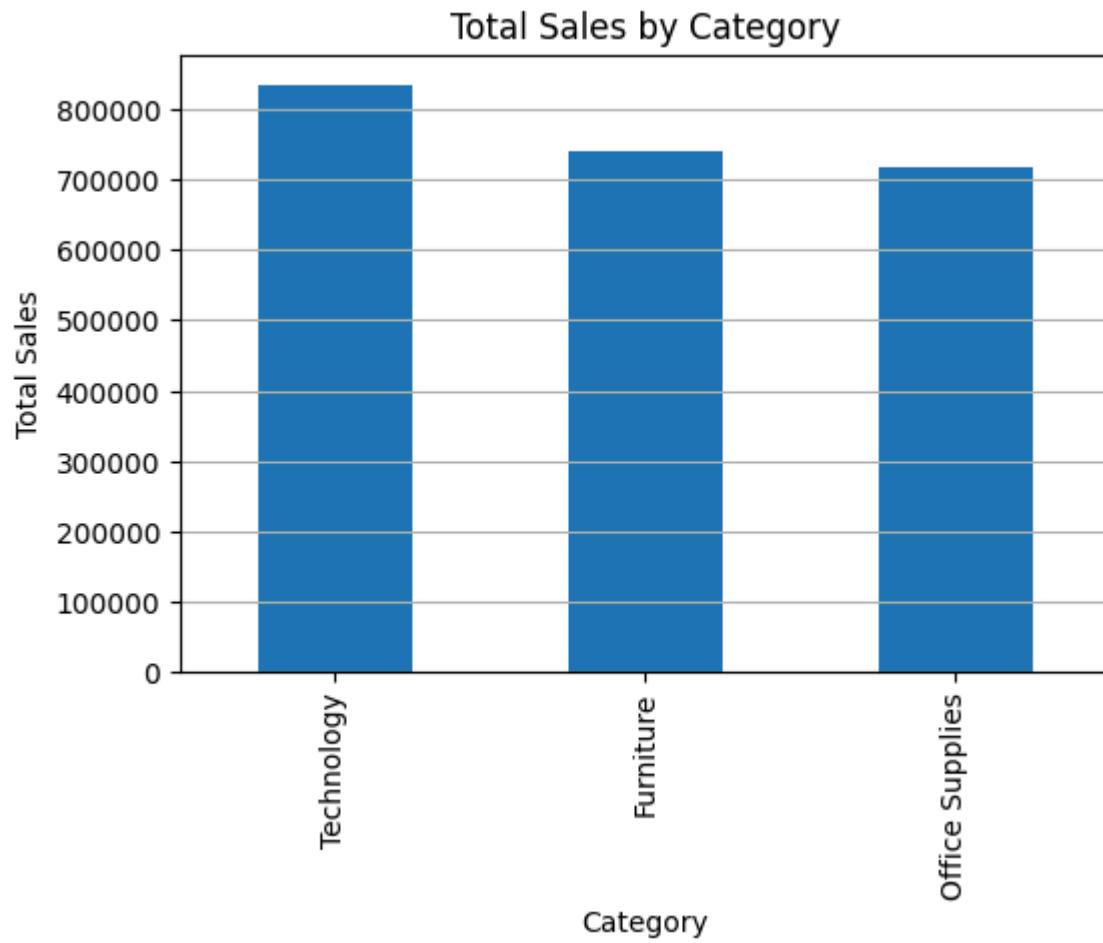
```
import pandas as pd  
import matplotlib.pyplot as plt
```

```
df = pd.read_csv("Sample - Superstore.csv", encoding="latin1")  
df.head()
```

Row	Order	Order	Ship Date	Ship	Customer	Customer	Segment	Country	City	Postal	Region	P
df.shape												
df.info()												
152150												
<class 'pandas.core.frame.DataFrame'>												
RangeIndex: 9994 entries, 0 to 9993												
Data columns (total 21 columns):												
#	Column	CA	Non-Null Count	Dtype	Second							
1	1	2016-11/8/2016	11/11/2016	int64	Class	CG-12520	Claire Gute	Consumer	United States	Henderson	... 42420	South FL 10
2	2	152156	9994 non-null	int64								
3	0	Row ID	9994 non-null	object								
4	1	Order ID	9994 non-null	object								
5	2	Order Date	9994 non-null	object								
6	3	Ship Date	9994 non-null	object								
7	4	Ship Mode	9994 non-null	object								
8	5	Customer ID	2016-6/13/2016	6/16/2016	object	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	... 90036	West CA 10
9	6	Customer Name	9994 non-null	object								
10	7	Segment	9994 non-null	object								
11	8	Country	9994 non-null	object								
12	9	City	9994 non-null	object								
13	10	State US-	9994 non-null	object								
14	11	Postal Code	2015-10/1/2015	10/18/2015	Standard int64	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	... 33311	South FL 10
15	12	Region	9994 non-null	object								
16	13	Product ID	9994 non-null	object								
17	14	Category	9994 non-null	object								
18	15	Sub-Category	9994 non-null	object								
19	16	Product Name	2015-10/1/2015	10/18/2015	Standard object	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	... 33311	South CA 10
20	17	Sale ID	9994 non-null	float64								
21	18	Quantity	9994 non-null	int64								
22	19	Discount	9994 non-null	float64								
23	20	Profit	9994 non-null	float64								
24	21	dtypes:	float64(3), int64(3), object(15)									
25	22	memory usage:	1.6+ MB									

```
top_categories = (
    df.groupby("Category")["Sales"]
    .sum()
    .sort_values(ascending=False)
```

```
)  
  
top_categories.plot(  
    kind="bar",  
    title="Total Sales by Category",  
    figsize=(6,4)  
)  
  
plt.xlabel("Category")  
plt.ylabel("Total Sales")  
plt.grid(axis="y")  
plt.show()
```

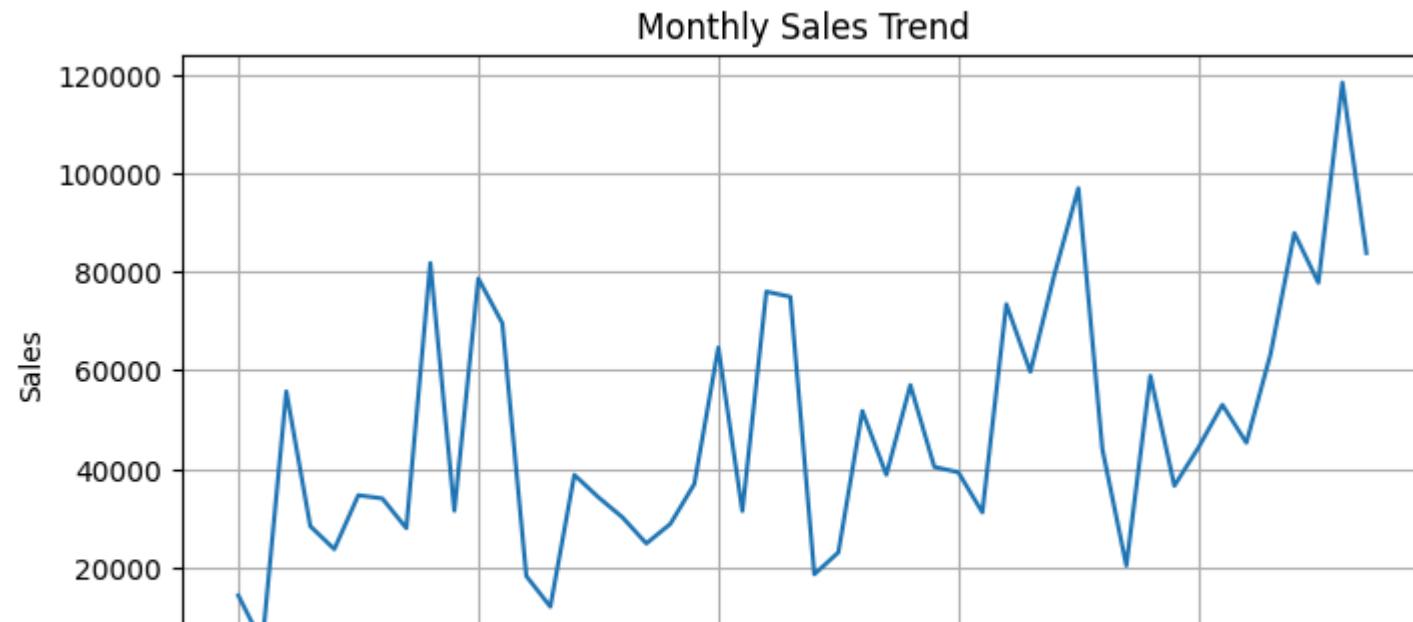


```
df["Order Date"] = pd.to_datetime(df["Order Date"])
```

```
sales_trend = (  
    df.groupby(df["Order Date"].dt.to_period("M"))["Sales"]  
    .sum()  
)
```

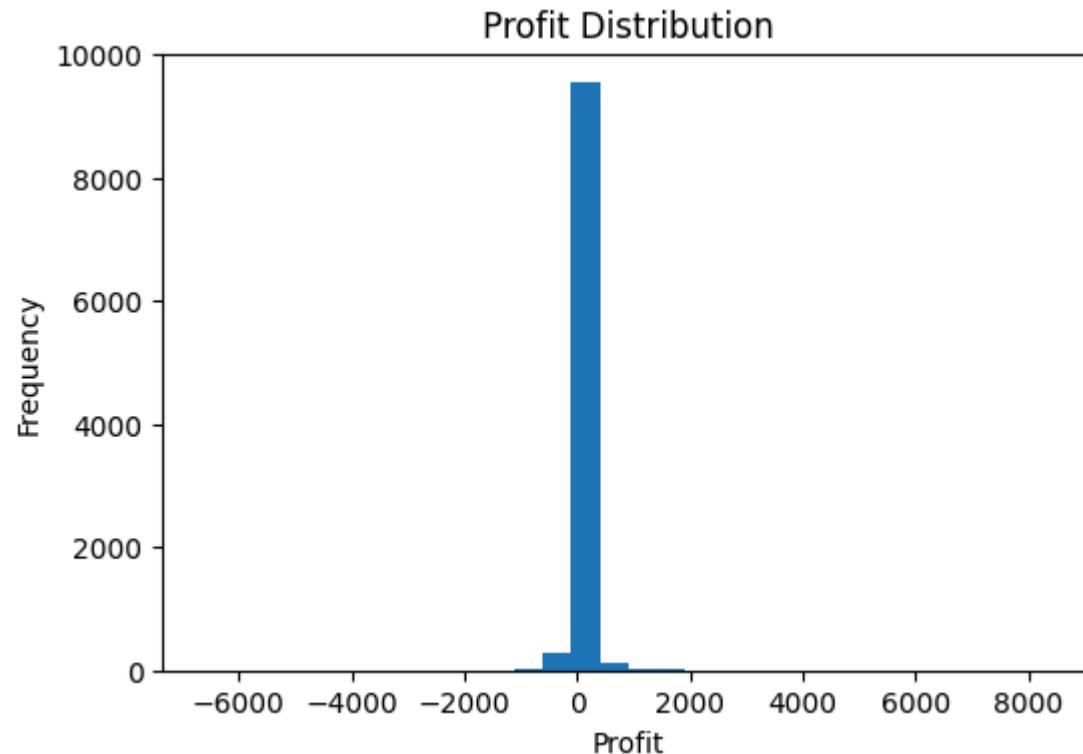
```
sales_trend.index = sales_trend.index.astype(str)
```

```
sales_trend.plot(  
    kind="line",  
    title="Monthly Sales Trend",  
    figsize=(8,4)  
)  
  
plt.xlabel("Month")  
plt.ylabel("Sales")  
plt.grid()  
plt.show()
```

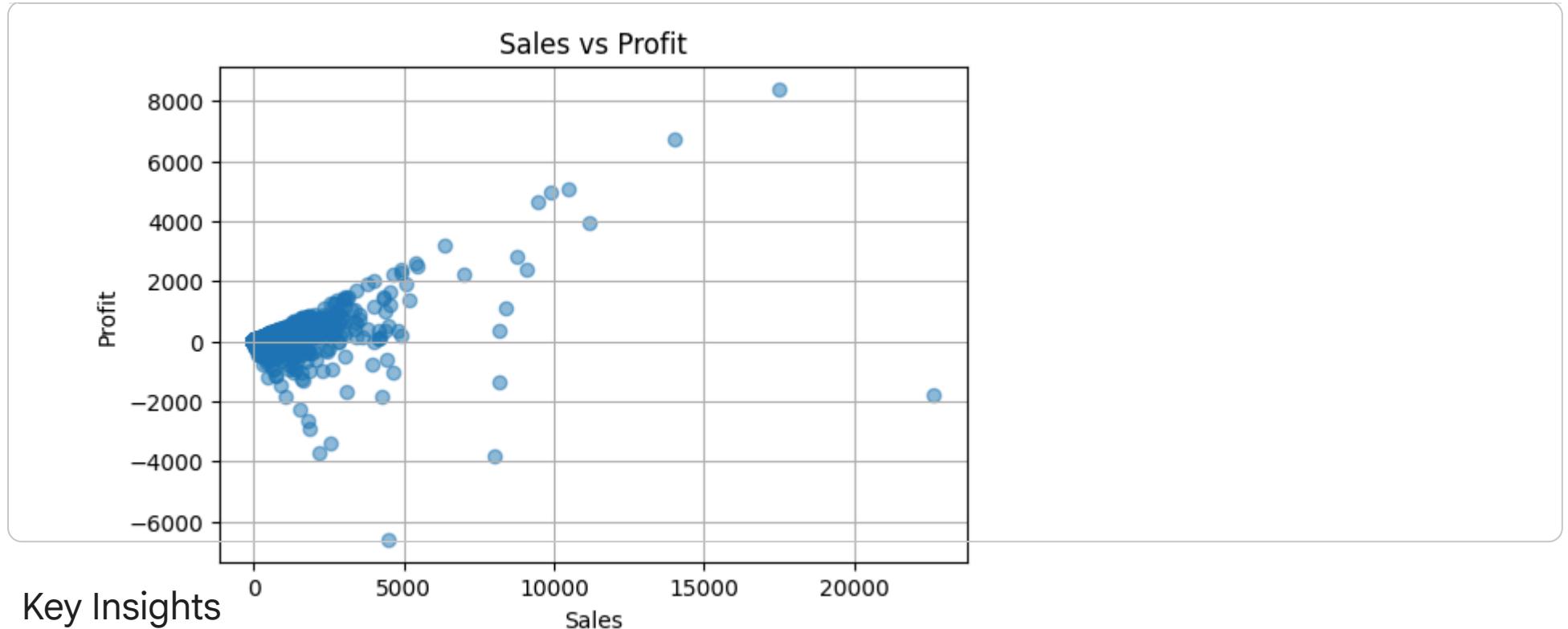


```
df["Profit"].plot(  
    kind="hist",  
    bins=30,  
    title="Profit Distribution",  
    figsize=(6,4)  
)
```

```
plt.xlabel("Profit")
plt.show()
```



```
plt.figure(figsize=(6,4))
plt.scatter(df["Sales"], df["Profit"], alpha=0.5)
plt.title("Sales vs Profit")
plt.xlabel("Sales")
plt.ylabel("Profit")
plt.grid()
plt.show()
```



Key Insights

1. Technology category generates the highest total sales among all categories.
2. Monthly sales show seasonal spikes, indicating high-demand periods.