```
df = pd.read_csv('train.csv')
df.head()
```

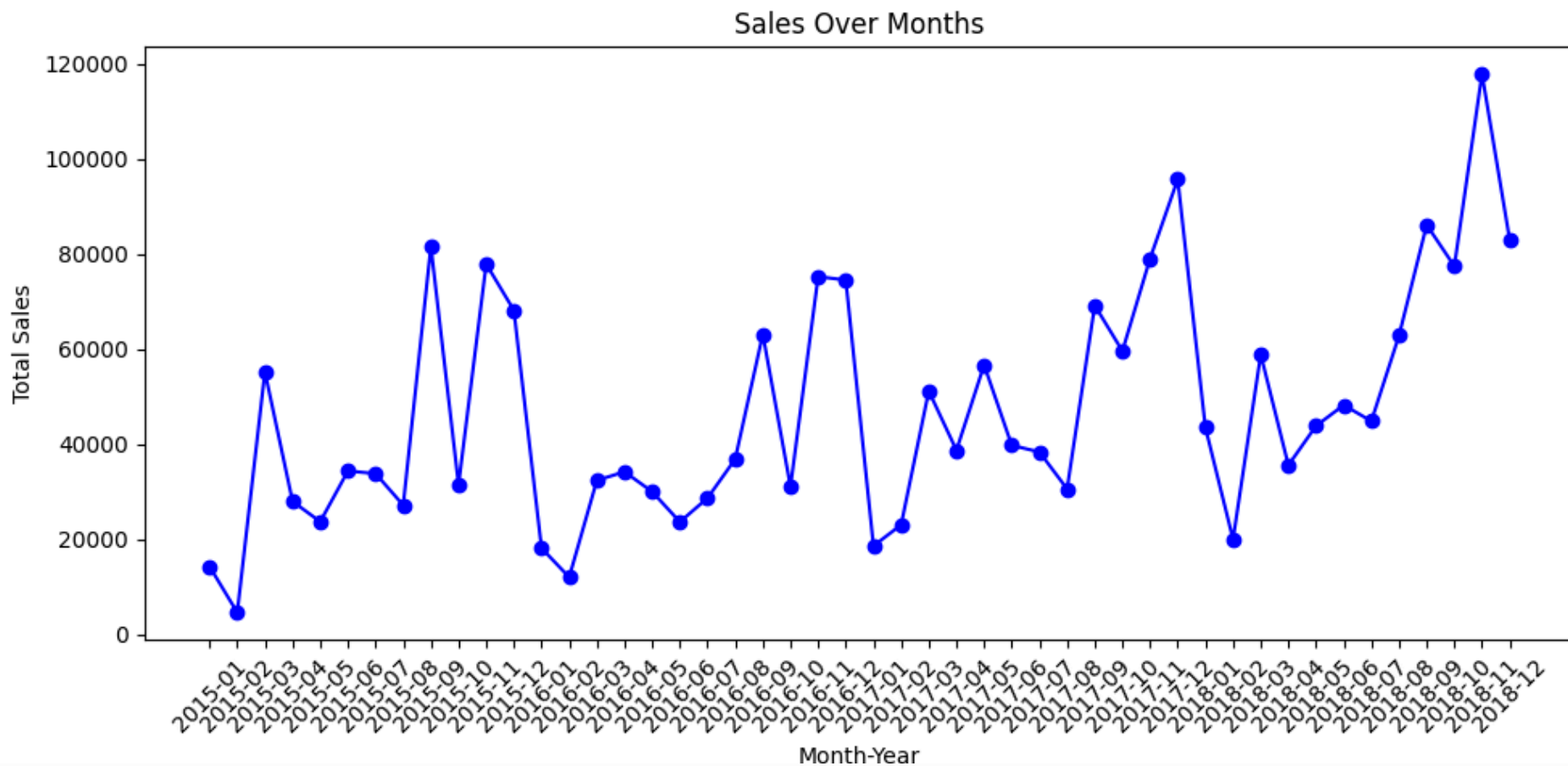|	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID
0	1	CA-2017-152156	08/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420.0	South	FUR-BO-10001798
1	2	CA-2017-152156	08/11/2017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

Next steps: Generate code with df   View recommended plots   New interactive sheet

```
df['Order Date'] = pd.to_datetime(df['Order Date'], format='%d/%m/%Y') # Specifying the correct format for day/month/year
df['Month-Year'] = df['Order Date'].dt.to_period('M').astype(str)
```

```python
import matplotlib.pyplot as plt

monthly_sales = df.groupby('Month-Year')['Sales'].sum().reset_index()

plt.figure(figsize=(10, 5))
plt.plot(monthly_sales['Month-Year'], monthly_sales['Sales'], marker='o', color='blue')
plt.xticks(rotation=45)
plt.title('Sales Over Months')
plt.xlabel('Month-Year')
plt.ylabel('Total Sales')
plt.tight_layout()
plt.show()
```



```python
region_sales = df.groupby('Region')['Sales'].sum().reset_index()
```

tryin
to
pars
the
'Ord
Date
colu
usir
the
defa
forn
'%m
(mo
but
it
enc
a
date
in a
diffe
forn
likel
'day
bas
on
the
exa
date
'15/
This
is
hap
bec
the
date

```python
plt.figure(figsize=(7, 4))
plt.bar(region_sales['Region'], region_sales['Sales'], color='green')
plt.title('Sales by Region')
plt.xlabel('Region')
plt.ylabel('Sales')
plt.tight_layout()
plt.show()
```


Sales by Region
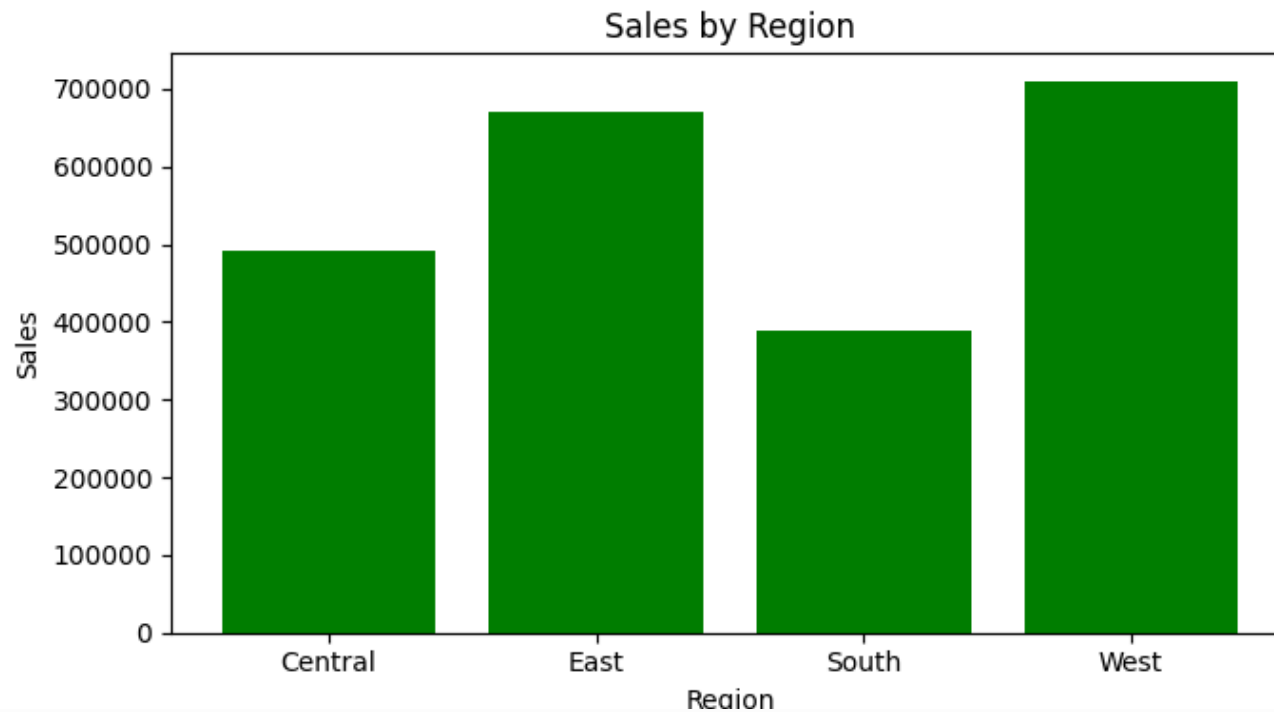
```python
category_sales = df.groupby('Category')['Sales'].sum()


plt.figure(figsize=(6,6))
plt.pie(category_sales, labels=category_sales.index, autopct='%1.1f%%', startangle=140, wedgeprops={'width': 0.4})
plt.title('Sales by Category')
plt.show()
```
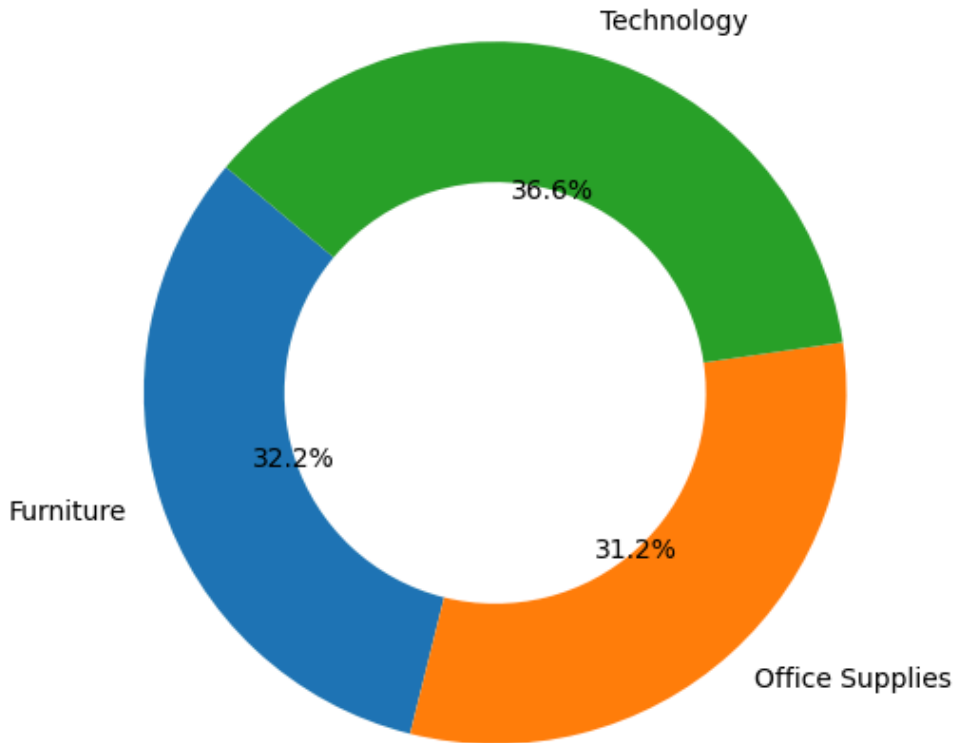
Sales by Category

Technology

36.6%

32.2%

Furniture

31.2%

Office Supplies

Insights:

1. Sales steadily increased from January to July 2023, peaking in Ju
2. The West region had the highest total sales, followed by the East
3. Technology category contributed the highest share to total sales.
4. The South region had the lowest overall sales during the year.

Insights:

1. Sales steadily increased from January to July 2023, peaking in June.
2. The West region had the highest total sales, followed by the East.
3. Technology category contributed the highest share to total sales.
4. The South region had the lowest overall sales during the year.

E

0 / 2000