Sales Data Analyzer.ipynb

In [1]:
```
import os
print(os.getcwd())
```

```
C:\Users\nilsa\PycharmProjects\PythonProject
```

In [2]:
```
import pandas as pd
df = pd.read_csv("SalesDataAnalyzer.csv")
print(df.head())
```

```
         date region   product  sales_amount
0  2025-01-01  North       Pen           120
1  2025-01-02  South  Notebook            85
2  2025-01-03   East       Pen            60
3  2025-01-04   West  Notebook            90
4  2025-01-05  North    Marker            70
```

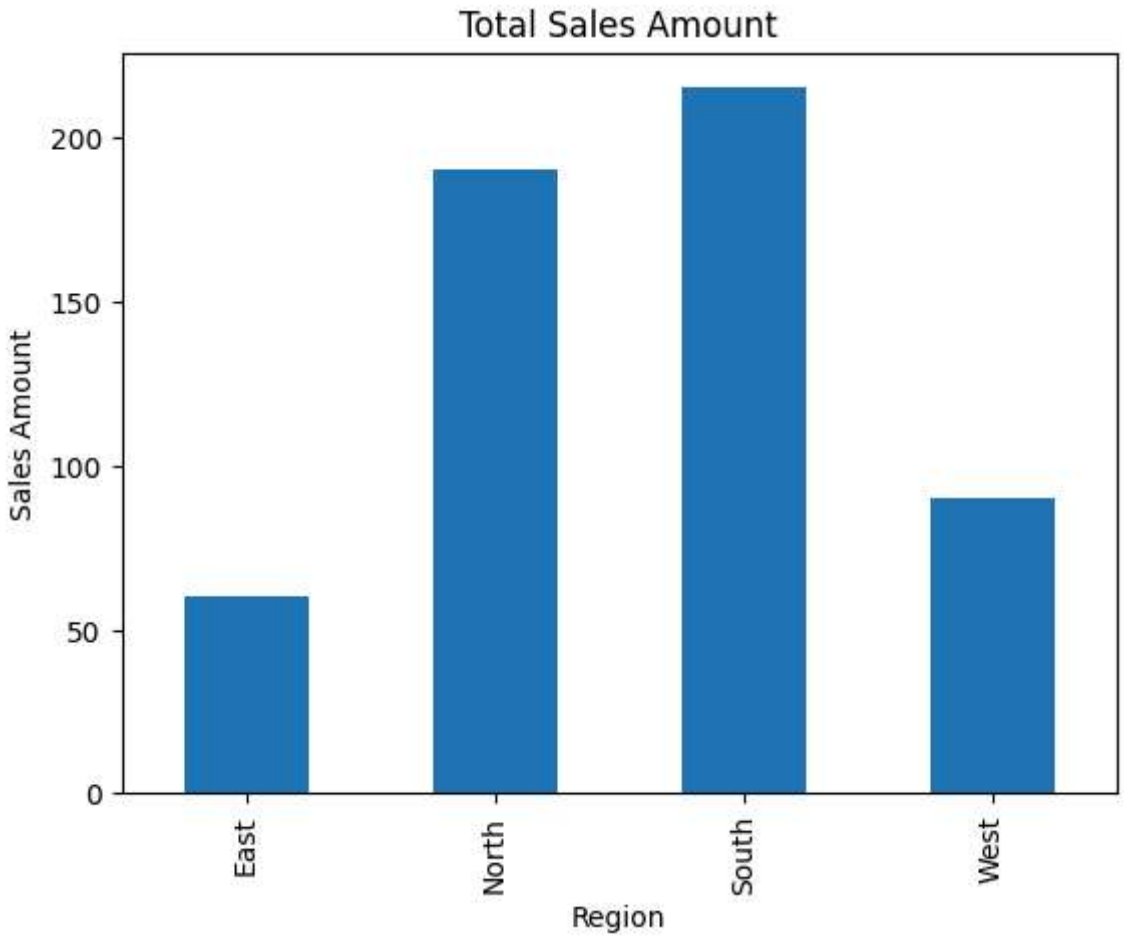
In [3]:
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('SalesDataAnalyzer.csv')
df['sales_amount'] = pd.to_numeric(df['sales_amount'], errors='coerce')

region_sales = df.groupby('region')['sales_amount'].sum()
print(region_sales)

region_sales.plot(kind='bar')
plt.title('Total Sales Amount')
plt.xlabel('Region')
plt.ylabel('Sales Amount')
plt.show()
```

```
region
East      60
North    190
South    215
West      90
Name: sales_amount, dtype: int64
```



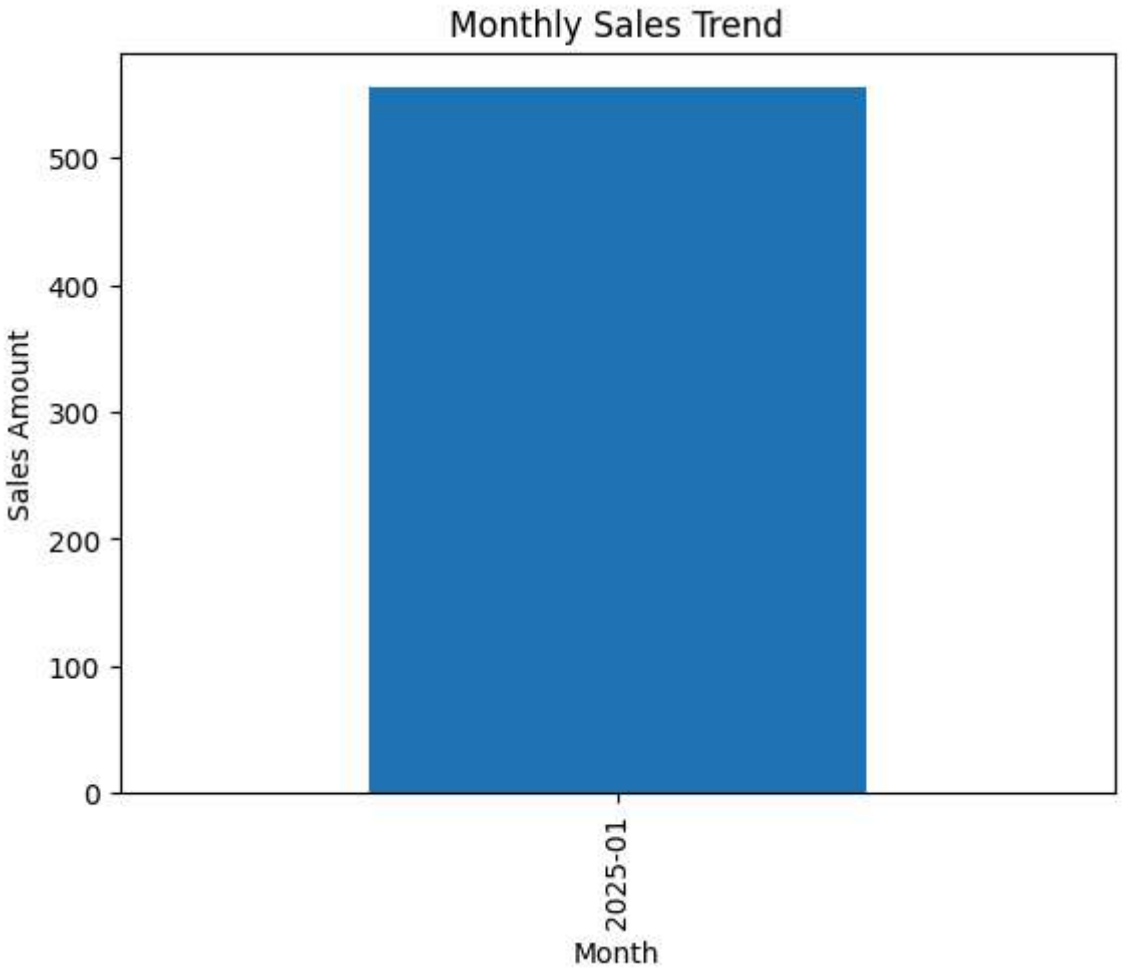
In [4]:
```
top_product = df.groupby(['region', 'product'])['sales_amount'].sum().reset_index()
print(top_product.sort_values(['region', 'sales_amount'], ascending=[True, False]))
```

```
  region   product  sales_amount
0   East       Pen            60
2  North       Pen           120
1  North    Marker            70
4  South       Pen           130
3  South  Notebook            85
5   West  Notebook            90
```

In [5]:
```
best_region_per_product = df.groupby(['product', 'region'])['sales_amount'].sum().r
max_per_product = best_region_per_product.sort_values(['product', 'sales_amount'],
print(max_per_product)
```

```
    product region  sales_amount
0    Marker  North            70
2  Notebook   West            90
5       Pen  South           130
```

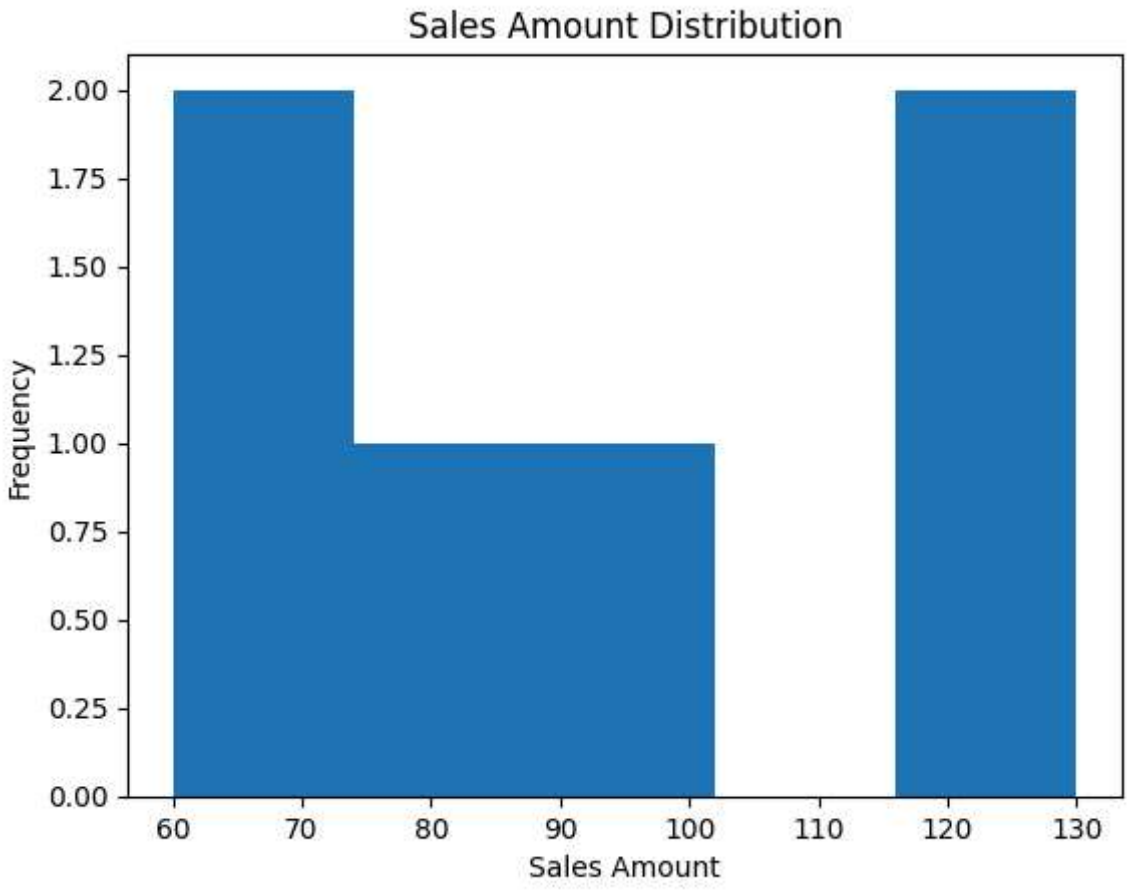
In [6]:
```
df['month'] = pd.to_datetime(df['date']).dt.to_period('M')
monthly_sales = df.groupby('month')['sales_amount'].sum()
monthly_sales.plot(kind='bar')
plt.title('Monthly Sales Trend')
plt.xlabel('Month')
plt.ylabel('Sales Amount')
plt.show()
```



In [7]:
```
top_product.to_csv('top_product_analysis.csv', index=False)
print("Analysis exported to top_product_analysis.csv!")
```

```
Analysis exported to top_product_analysis.csv!
```

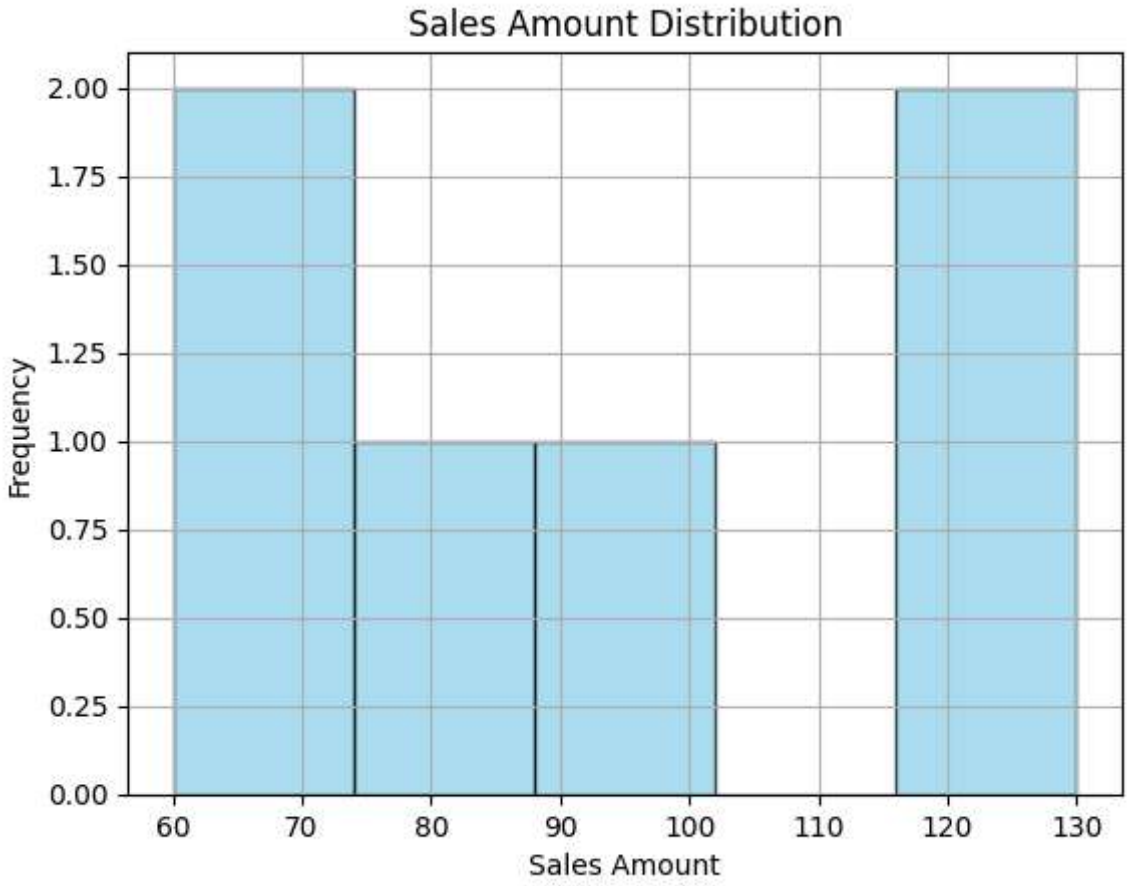
In [8]:
```
df['sales_amount'].plot(kind='hist', bins=5)
plt.title('Sales Amount Distribution')
plt.xlabel('Sales Amount')
plt.ylabel('Frequency')
plt.show()
```



In [9]:
```
print("Basic Summary:")
print(f"Total Sales: {df['sales_amount'].sum()}")
print(f"Average Sales: {df['sales_amount'].mean()}")
print(f"Max Sale: {df['sales_amount'].max()} by {df.loc[df['sales_amount'].idxmax()
print(f"Min Sale: {df['sales_amount'].min()} by {df.loc[df['sales_amount'].idxmin()
```

```
Basic Summary:
Total Sales: 555
Average Sales: 92.5
Max Sale: 130 by date            2025-01-06
region                 South
product                  Pen
sales_amount             130
month                2025-01
Name: 5, dtype: object
Min Sale: 60 by date            2025-01-03
region                  East
product                  Pen
sales_amount              60
month                2025-01
Name: 2, dtype: object
```

In [10]:
```
df['sales_amount'].plot(kind='hist', bins=5, color='skyblue', edgecolor='black', al
plt.title('Sales Amount Distribution')
plt.xlabel('Sales Amount')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```


Sales Amount Distribution
Frequency
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
60
70
80
90
100
110
120
130
Sales Amount