

## Python code for Amazon Sales Data

# Install necessary libraries

```
!pip install pandas matplotlib seaborn
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
from google.colab import files
```

# Upload CSV file

```
uploaded = files.upload()
```

# Load the data

```
file_path = list(uploaded.keys())[0]
```

```
data = pd.read_csv(file_path)
```

# Display the first few rows of the dataframe

```
print(data.head())
```

# Convert the Date column to datetime type

```
data['Date'] = pd.to_datetime(data['Date'], format='%Y-%m-%d')
```

# Handle missing values (e.g., removing rows with NaN in Sales)

```
data = data.dropna(subset=['Sales'])
```

# Extract Year and Month for aggregation

```
data['Year'] = data['Date'].dt.year
```

```
data['Month'] = data['Date'].dt.to_period('M')
```

# Aggregate data by Month

```
monthly_sales = data.groupby('Month')['Sales'].sum().reset_index()
```

```
monthly_sales['Month'] = monthly_sales['Month'].astype(str)
```

# Plot Monthly Sales Trend

```
plt.figure(figsize=(12, 6))
```

```
sns.lineplot(data=monthly_sales, x='Month', y='Sales')
```

```
plt.title('Monthly Sales Trend')
```

```
plt.xlabel('Month')
```

```
plt.ylabel('Total Sales')
```

```
plt.xticks(rotation=90)
```

```
plt.tight_layout()
```

```
plt.show()
```

# Aggregate data by Year

```
yearly_sales = data.groupby('Year')['Sales'].sum().reset_index()
```

# Plot Yearly Sales Trend

```
plt.figure(figsize=(12, 6))
```

```
sns.barplot(data=yearly_sales, x='Year', y='Sales')
```

```
plt.title('Yearly Sales Trend')
```

```
plt.xlabel('Year')
```

```
plt.ylabel('Total Sales')
```

```
plt.tight_layout()
plt.show()
```

```
# Aggregate data by Product
product_sales = data.groupby('Product')['Sales'].sum().reset_index()
product_sales = product_sales.sort_values(by='Sales', ascending=False)
```

```
# Plot Sales Distribution by Product
plt.figure(figsize=(12, 6))
sns.barplot(data=product_sales, x='Product', y='Sales')
plt.title('Sales Distribution by Product')
plt.xlabel('Product')
plt.ylabel('Total Sales')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```

```
# Key Metrics
average_monthly_sales = monthly_sales['Sales'].mean()
top_products = product_sales.head(5)
```

```
# Print results
print(f'Average Monthly Sales: ${average_monthly_sales:,.2f}')
print('Top 5 Products by Total Sales:')
print(top_products)
```