Import the dataset

```
import pandas as pd

# Load the dataset
df = pd.read_csv("/content/retail_sales_dataset.csv")
```

```
# Explore the structure and content of the dataset
print(df.info())
print(df.describe())
print(df.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Transaction ID    1000 non-null   int64
 1   Date              1000 non-null   object
 2   Customer ID       1000 non-null   object
 3   Gender            1000 non-null   object
 4   Age               1000 non-null   int64
 5   Product Category  1000 non-null   object
 6   Quantity          1000 non-null   int64
 7   Price per Unit    1000 non-null   int64
 8   Total Amount      1000 non-null   int64
dtypes: int64(5), object(4)
memory usage: 70.4+ KB
None
       Transaction ID         Age     Quantity  Price per Unit  Total Amount
count     1000.000000  1000.00000  1000.000000     1000.000000   1000.000000
mean       500.500000    41.39200     2.514000      179.890000    456.000000
std        288.819436    13.68143     1.132734      189.681356    559.997632
min          1.000000    18.00000     1.000000       25.000000     25.000000
25%        250.750000    29.00000     1.000000       30.000000     60.000000
50%        500.500000    42.00000     3.000000       50.000000    135.000000
75%        750.250000    53.00000     4.000000      300.000000    900.000000
max       1000.000000    64.00000     4.000000      500.000000   2000.000000
   Transaction ID        Date Customer ID  Gender  Age Product Category  \
0               1  2023-11-24     CUST001    Male   34           Beauty
1               2  2023-02-27     CUST002  Female   26         Clothing
2               3  2023-01-13     CUST003    Male   50      Electronics
3               4  2023-05-21     CUST004    Male   37         Clothing
4               5  2023-05-06     CUST005    Male   30           Beauty

   Quantity  Price per Unit  Total Amount
0         3              50           150
1         2             500          1000
2         1              30            30
3         1             500           500
4         2              50           100
```

```
df.columns
```

```
Index(['Transaction ID', 'Date', 'Customer ID', 'Gender', 'Age',
       'Product Category', 'Quantity', 'Price per Unit', 'Total Amount'],
      dtype='object')
```

Grouping by product category

```
region_sales = df.groupby('Product Category')['Total Amount'].sum().reset_index()
region_sales = region_sales.sort_values(by='Total Amount', ascending=False)
print(region_sales)
```

```
  Product Category  Total Amount
2      Electronics        156905
1         Clothing        155580
0           Beauty        143515
```

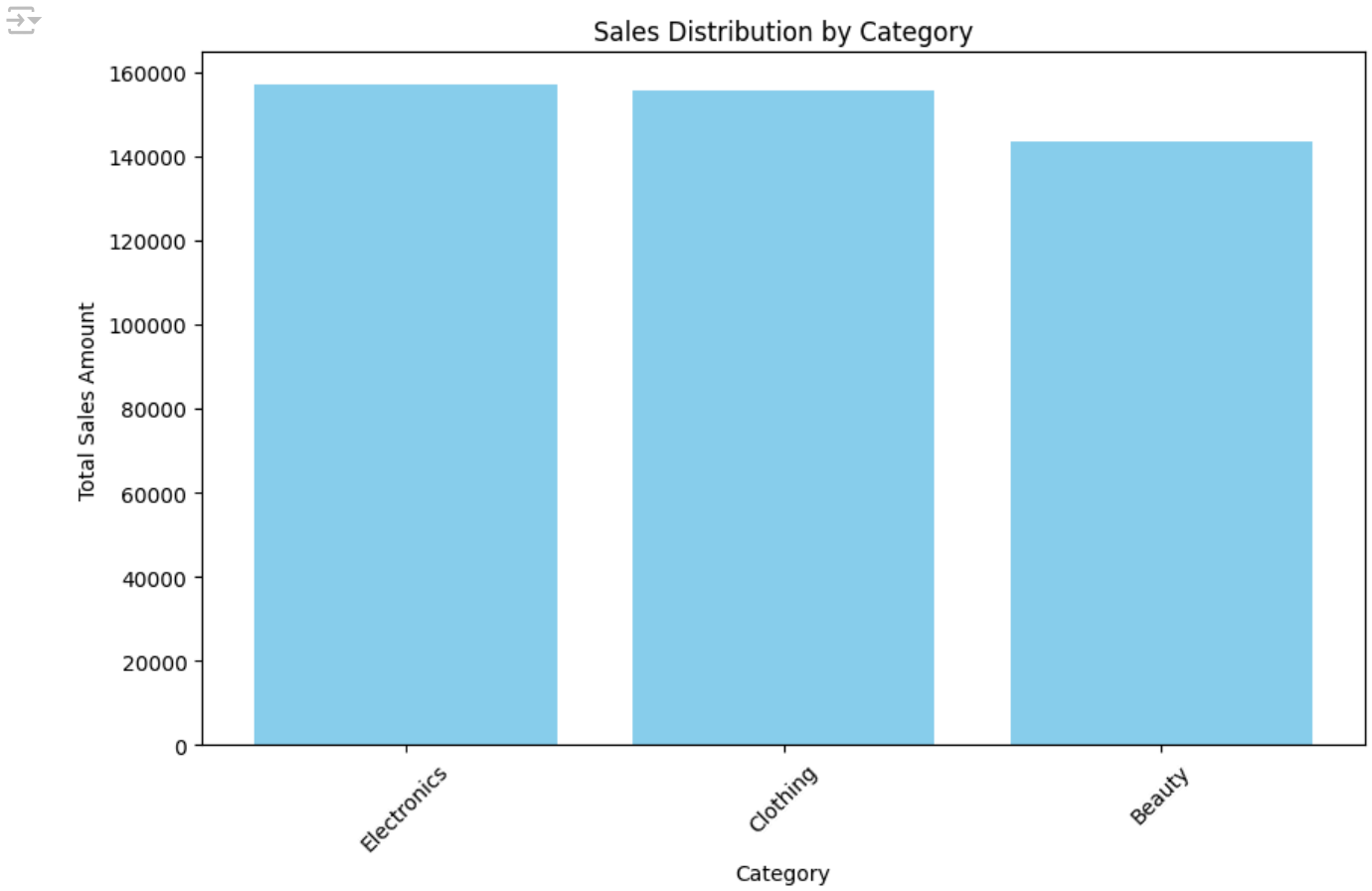
```
top_regions = region_sales.head()
print(top_regions)
```

```
  Product Category  Total Amount
2      Electronics        156905
1         Clothing        155580
0           Beauty        143515
```

Visualizing the data grouped by category

```
import matplotlib.pyplot as plt

# Bar plot
plt.figure(figsize=(10,6))
plt.bar(region_sales['Product Category'], region_sales['Total Amount'], color='skyblue')
plt.title('Sales Distribution by Category')
plt.xlabel('Category')
plt.ylabel('Total Sales Amount')
plt.xticks(rotation=45)
plt.show()
```




```
# Pie chart
```

```
plt.figure(figsize=(8,8))
plt.pie(region_sales['Total Amount'], labels=region_sales['Product Category'], autopct='%1.1f%%', colors=plt
plt.title('Sales Distribution by Category')
plt.show()
```