

```
In [1]: import pandas as pd  
import matplotlib.pyplot as plt
```

```
In [2]: df = pd.read_csv("customer_shopping_data.csv")  
df
```

```
Out[2]:   invoice_no  customer_id  gender  age  category  quantity  price  payment_method  invoice_date  shop  
0       I138884      C241288  Female   28  Clothing       5  1500.40  Credit Card  5/8/2022  
1       I317333      C111565   Male    21    Shoes       3  1800.51  Debit Card  12/12/2021 Forum  
2       I127801      C266599   Male    20  Clothing       1  300.08    Cash  9/11/2021  
3       I173702      C988172  Female   66    Shoes       5  3000.85  Credit Card  16/05/2021 Met  
4       I337046      C189076  Female   53    Books       4  60.60    Cash  24/10/2021  
...  
99452     I219422      C441542  Female   45  Souvenir       5  58.65  Credit Card  21/09/2022  
99453     I325143      C569580   Male    27  Food & Beverage       2  10.46    Cash  22/09/2021 Forum  
99454     I824010      C103292   Male    63  Food & Beverage       2  10.46  Debit Card  28/03/2021  
99455     I702964      C800631   Male    56 Technology       4  4200.00    Cash  16/03/2021  
99456     I232867      C273973  Female   36  Souvenir       3  35.19  Credit Card  15/10/2022 Mall
```

99457 rows × 10 columns

```
In [3]: print("Dataset Info:")  
print(df.info())  
print("\nFirst 5 rows:")  
print(df.head())
```

```
Dataset Info:  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 99457 entries, 0 to 99456  
Data columns (total 10 columns):  
 #   Column           Non-Null Count  Dtype     
---  --  
 0   invoice_no      99457 non-null   object    
 1   customer_id     99457 non-null   object    
 2   gender          99457 non-null   object    
 3   age             99457 non-null   int64     
 4   category        99457 non-null   object    
 5   quantity        99457 non-null   int64     
 6   price           99457 non-null   float64   
 7   payment_method   99457 non-null   object    
 8   invoice_date    99457 non-null   object    
 9   shopping_mall   99457 non-null   object    
dtypes: float64(1), int64(2), object(7)  
memory usage: 7.6+ MB  
None
```

```
First 5 rows:  
   invoice_no  customer_id  gender  age  category  quantity  price  \  
0       I138884      C241288  Female   28  Clothing       5  1500.40  
1       I317333      C111565   Male    21    Shoes       3  1800.51  
2       I127801      C266599   Male    20  Clothing       1  300.08  
3       I173702      C988172  Female   66    Shoes       5  3000.85  
4       I337046      C189076  Female   53    Books       4  60.60  
  
  payment_method  invoice_date  shopping_mall  
0   Credit Card    5/8/2022      Kanyon  
1   Debit Card     12/12/2021 Forum Istanbul  
2       Cash        9/11/2021 Metrocity  
3   Credit Card    16/05/2021 Metropol AVM  
4       Cash        24/10/2021 Kanyon
```

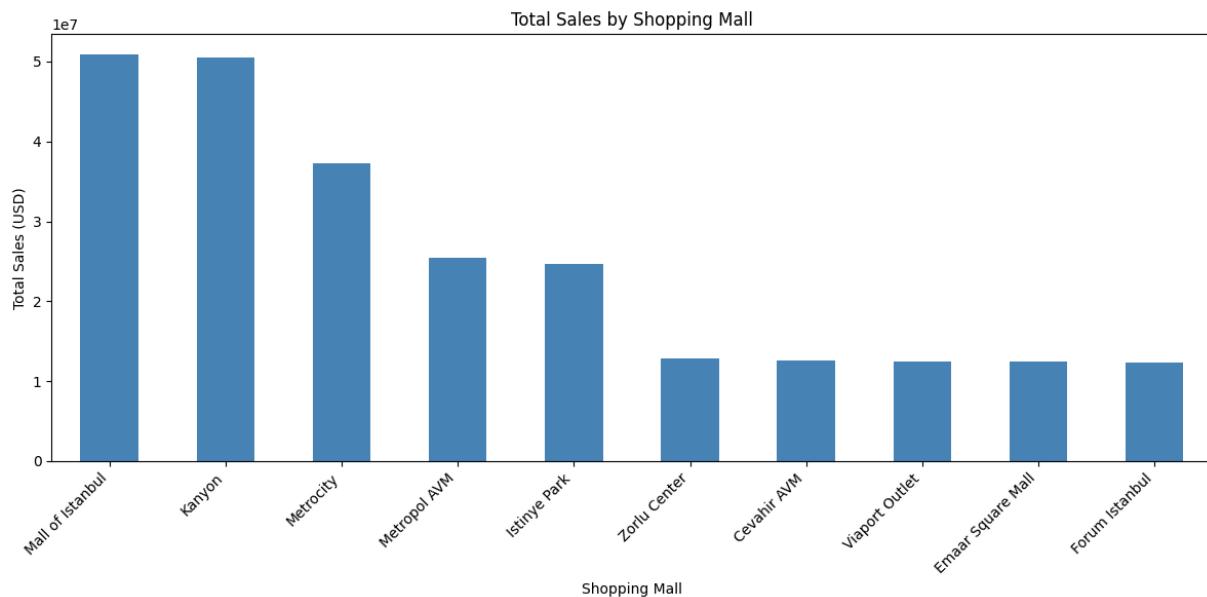
```
In [4]: print("\nColumn Names:", df.columns.tolist())
Column Names: ['invoice_no', 'customer_id', 'gender', 'age', 'category', 'quantity', 'price', 'payment_method', 'invoice_date', 'shopping_mall']

In [5]: df["Sales_Amount"] = df["quantity"] * df["price"]

In [6]: sales_by_region = (
    df.groupby("shopping_mall")["Sales_Amount"]
    .sum()
    .sort_values(ascending=False)
)
print("\nTotal Sales by Shopping Mall:")
print(sales_by_region)

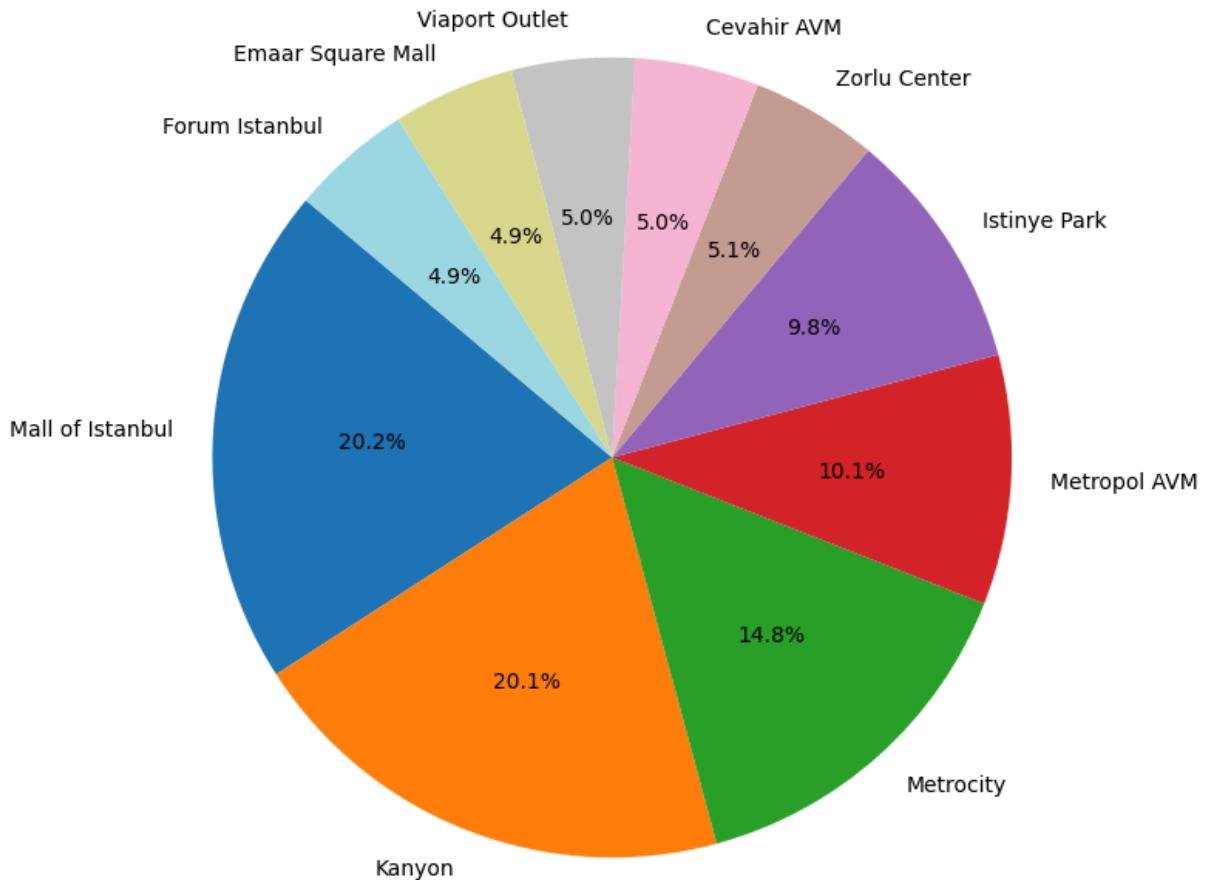
Total Sales by Shopping Mall:
shopping_mall
Mall of Istanbul      50872481.68
Kanyon                50554231.10
Metrocity              37302787.33
Metropol AVM          25379913.19
Istinye Park           24618827.68
Zorlu Center           12901053.82
Cevahir AVM            12645138.20
Viaport Outlet         12521339.72
Emaar Square Mall      12406100.29
Forum Istanbul          12303921.24
Name: Sales_Amount, dtype: float64
```

```
In [7]: plt.figure(figsize=(12, 6))
sales_by_region.plot(kind="bar", color="steelblue")
plt.title("Total Sales by Shopping Mall")
plt.xlabel("Shopping Mall")
plt.ylabel("Total Sales (USD)")
plt.xticks(rotation=45, ha="right")
plt.tight_layout()
plt.show()
```



```
In [8]: plt.figure(figsize=(8, 8))
sales_by_region.plot(
    kind="pie",
    autopct="%1.1f%%",
    startangle=140,
    colormap="tab20",
    ylabel=""
)
plt.title("Sales Distribution by Shopping Mall")
plt.tight_layout()
plt.show()
```

Sales Distribution by Shopping Mall



```
In [9]: top_malls = sales_by_region.head(10)
print("\nTop 10 Performing Shopping Malls:")
print(top_malls)
```

Top 10 Performing Shopping Malls:

shopping_mall	Sales_Amount
Mall of Istanbul	50872481.68
Kanyon	50554231.10
Metrocity	37302787.33
Metropol AVM	25379913.19
Istinye Park	24618827.68
Zorlu Center	12901053.82
Cevahir AVM	12645138.20
Viaport Outlet	12521339.72
Emaar Square Mall	12406100.29
Forum Istanbul	12303921.24

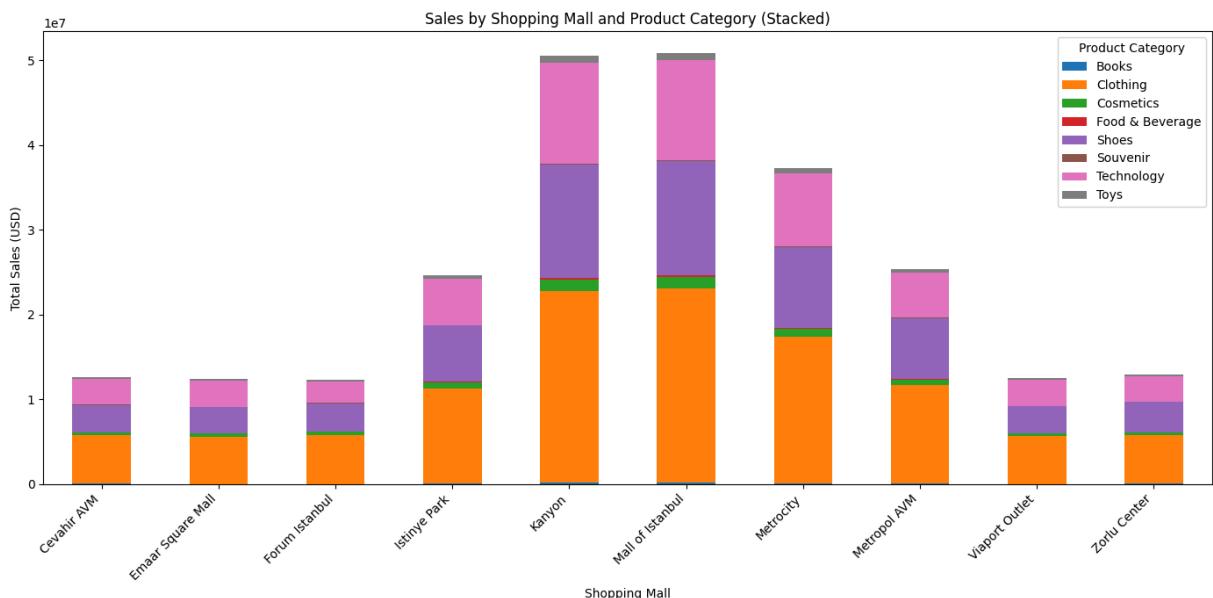
Name: Sales_Amount, dtype: float64

```
In [10]: region_category_sales = (
    df.groupby(["shopping_mall", "category"])["Sales_Amount"]
    .sum()
    .reset_index()
```

```
In [11]: region_pivot = region_category_sales.pivot(
    index="shopping_mall", columns="category", values="Sales_Amount"
)

region_pivot.plot(kind="bar", stacked=True, figsize=(14, 7))
plt.title("Sales by Shopping Mall and Product Category (Stacked)")
plt.xlabel("Shopping Mall")
plt.ylabel("Total Sales (USD)")
plt.xticks(rotation=45, ha="right")
plt.legend(title="Product Category")
```

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



```
In [12]: top_category_per_mall = (  
    region_category_sales.sort_values(["shopping_mall", "Sales_Amount"], ascending=[True, False])  
    .groupby("shopping_mall")  
    .first()  
    .reset_index()  
)  
  
print("\nTop Product Category in Each Shopping Mall:")  
print(top_category_per_mall)
```

Top Product Category in Each Shopping Mall:

	shopping_mall	category	Sales_Amount
0	Cevahir AVM	Clothing	5706321.28
1	Emaar Square Mall	Clothing	5590490.40
2	Forum Istanbul	Clothing	5792444.24
3	Istinye Park	Clothing	11253900.24
4	Kanyon	Clothing	22609527.60
5	Mall of Istanbul	Clothing	22947417.68
6	Metrocity	Clothing	17226692.56
7	Metropol AVM	Clothing	11568084.00
8	Viaport Outlet	Clothing	5604594.16
9	Zorlu Center	Clothing	5697318.88