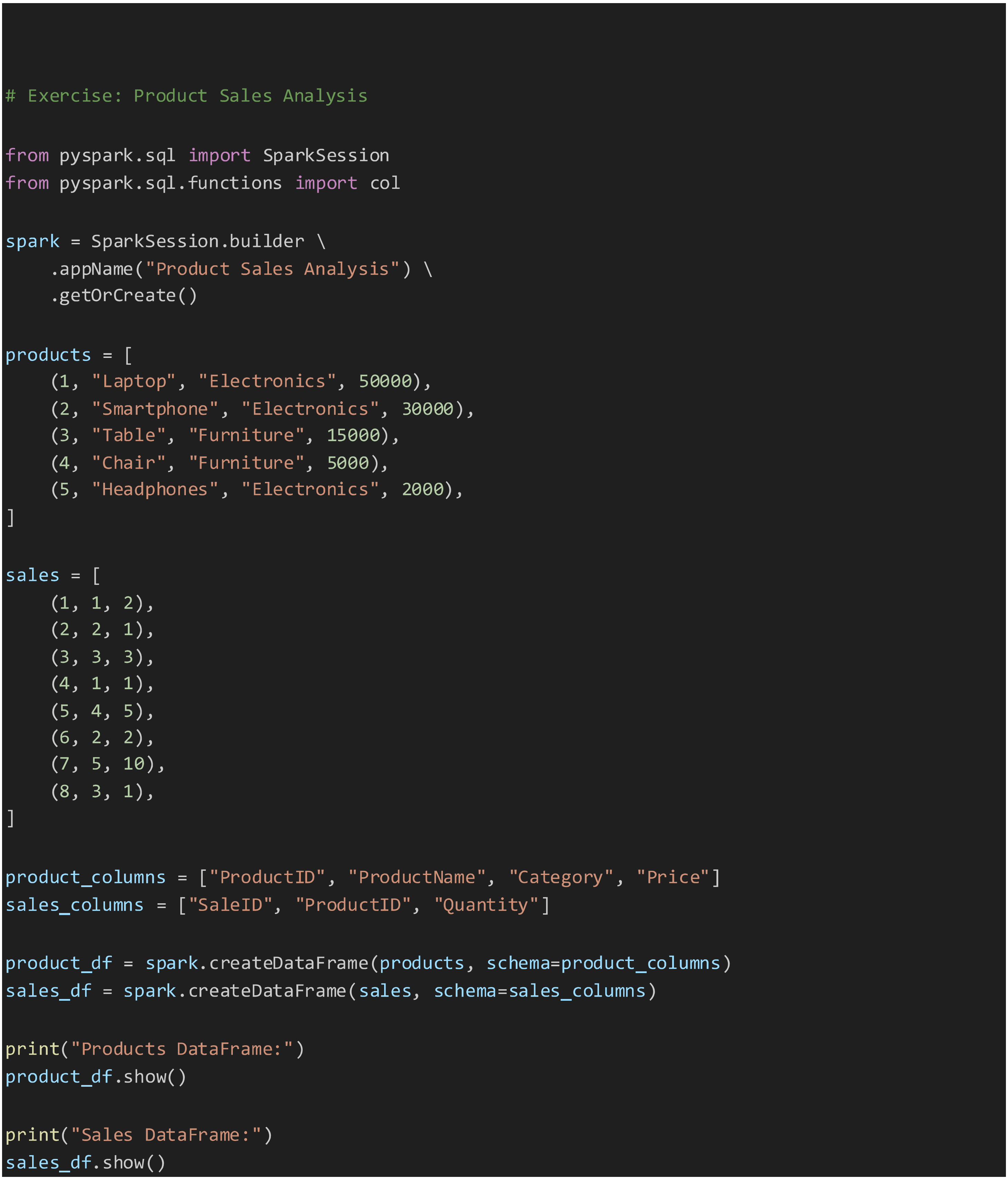
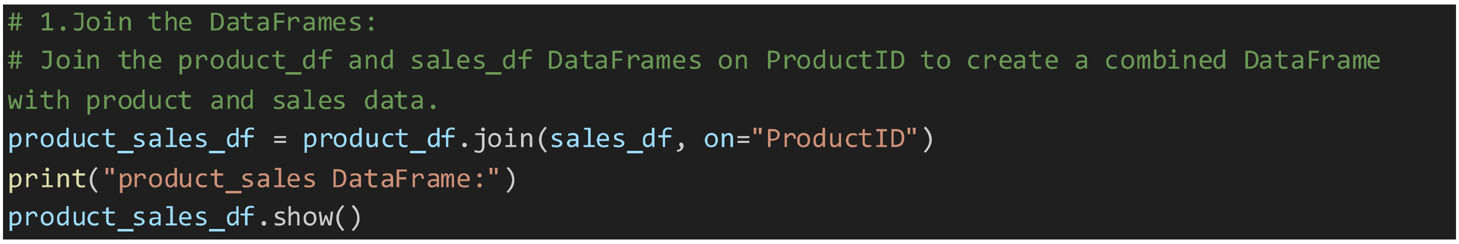
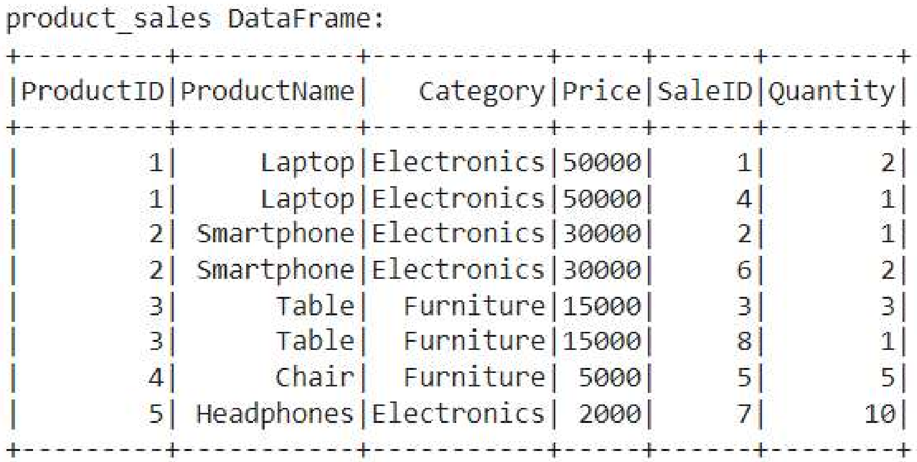


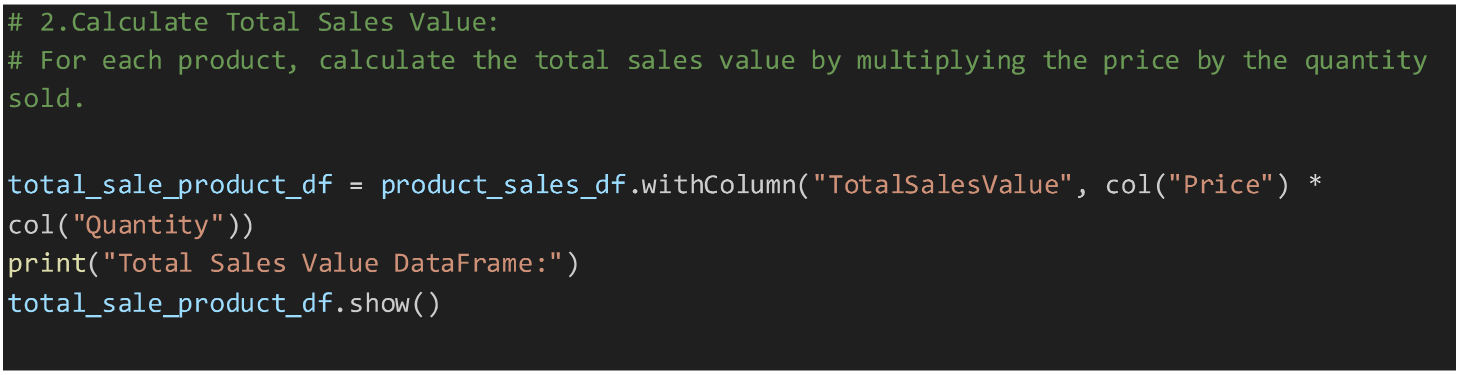
```python
# Exercise: Product Sales Analysis

from pyspark.sql import SparkSession
from pyspark.sql.functions import col

spark = SparkSession.builder \
    .appName("Product Sales Analysis") \
    .getOrCreate()

products = [
    (1, "Laptop", "Electronics", 50000),
    (2, "Smartphone", "Electronics", 30000),
    (3, "Table", "Furniture", 15000),
    (4, "Chair", "Furniture", 5000),
    (5, "Headphones", "Electronics", 2000),
]

sales = [
    (1, 1, 2),
    (2, 2, 1),
    (3, 3, 3),
    (4, 1, 1),
    (5, 4, 5),
    (6, 2, 2),
    (7, 5, 10),
    (8, 3, 1),
]

product_columns = ["ProductID", "ProductName", "Category", "Price"]
sales_columns = ["SaleID", "ProductID", "Quantity"]

product_df = spark.createDataFrame(products, schema=product_columns)
sales_df = spark.createDataFrame(sales, schema=sales_columns)

print("Products DataFrame:")
product_df.show()

print("Sales DataFrame:")
sales_df.show()
```



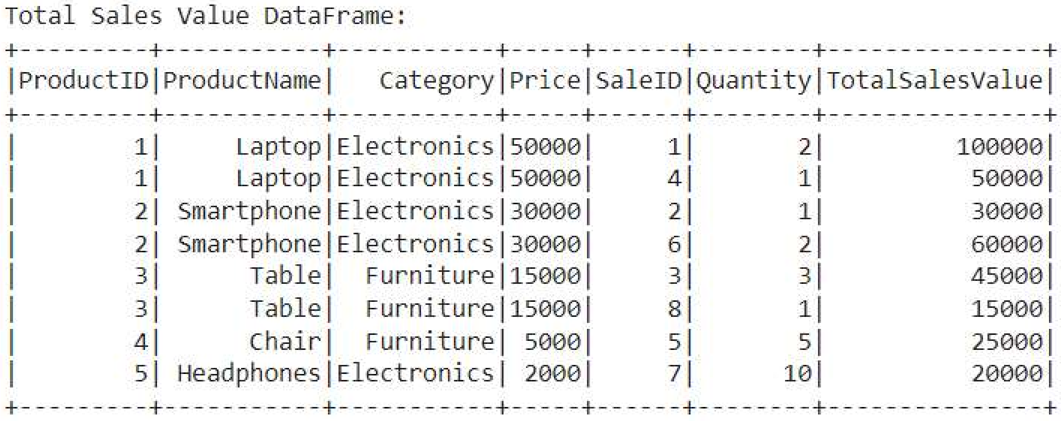
```
# 1.Join the DataFrames:
# Join the product_df and sales_df DataFrames on ProductID to create a combined DataFrame
with product and sales data.
product_sales_df = product_df.join(sales_df, on="ProductID")
print("product_sales DataFrame:")
product_sales_df.show()
```



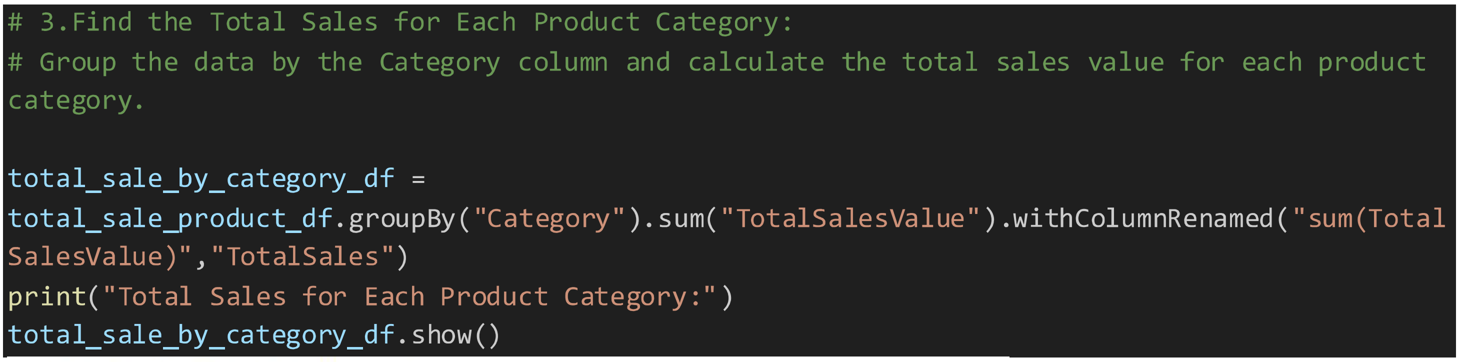
```
product_sales DataFrame:
+---------+-----------+-----------+-----+------+--------+
|ProductID|ProductName|   Category|Price|SaleID|Quantity|
+---------+-----------+-----------+-----+------+--------+
|        1|     Laptop|Electronics|50000|     1|       2|
|        1|     Laptop|Electronics|50000|     4|       1|
|        2| Smartphone|Electronics|30000|     2|       1|
|        2| Smartphone|Electronics|30000|     6|       2|
|        3|      Table|  Furniture|15000|     3|       3|
|        3|      Table|  Furniture|15000|     8|       1|
|        4|      Chair|  Furniture| 5000|     5|       5|
|        5| Headphones|Electronics| 2000|     7|      10|
+---------+-----------+-----------+-----+------+--------+
```



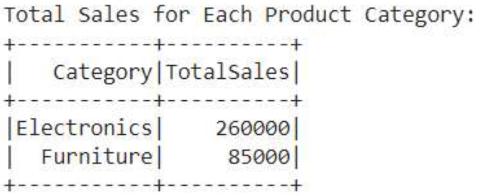
```
# 2.Calculate Total Sales Value:
# For each product, calculate the total sales value by multiplying the price by the quantity
sold.

total_sale_product_df = product_sales_df.withColumn("TotalSalesValue", col("Price") *
col("Quantity"))
print("Total Sales Value DataFrame:")
total_sale_product_df.show()
```



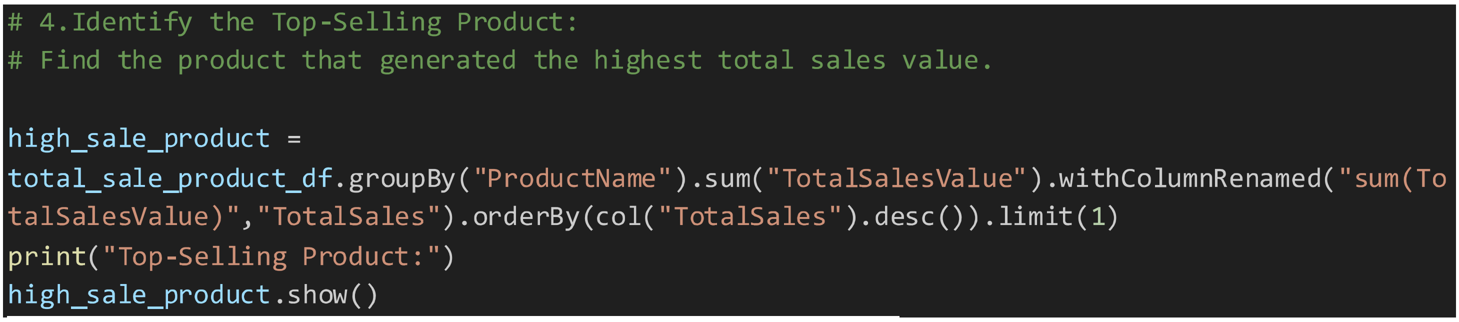
```
Total Sales Value DataFrame:
+---------+-----------+-----------+-----+------+--------+---------------+
|ProductID|ProductName|   Category|Price|SaleID|Quantity|TotalSalesValue|
+---------+-----------+-----------+-----+------+--------+---------------+
|        1|     Laptop|Electronics|50000|     1|       2|         100000|
|        1|     Laptop|Electronics|50000|     4|       1|          50000|
|        2| Smartphone|Electronics|30000|     2|       1|          30000|
|        2| Smartphone|Electronics|30000|     6|       2|          60000|
|        3|      Table|  Furniture|15000|     3|       3|          45000|
|        3|      Table|  Furniture|15000|     8|       1|          15000|
|        4|      Chair|  Furniture| 5000|     5|       5|          25000|
|        5| Headphones|Electronics| 2000|     7|      10|          20000|
+---------+-----------+-----------+-----+------+--------+---------------+
```



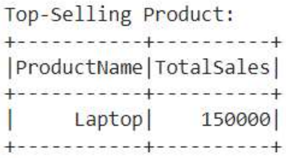
```
# 3.Find the Total Sales for Each Product Category:
# Group the data by the Category column and calculate the total sales value for each product
category.

total_sale_by_category_df =
total_sale_product_df.groupBy("Category").sum("TotalSalesValue").withColumnRenamed("sum(Total
SalesValue)","TotalSales")
print("Total Sales for Each Product Category:")
total_sale_by_category_df.show()
```



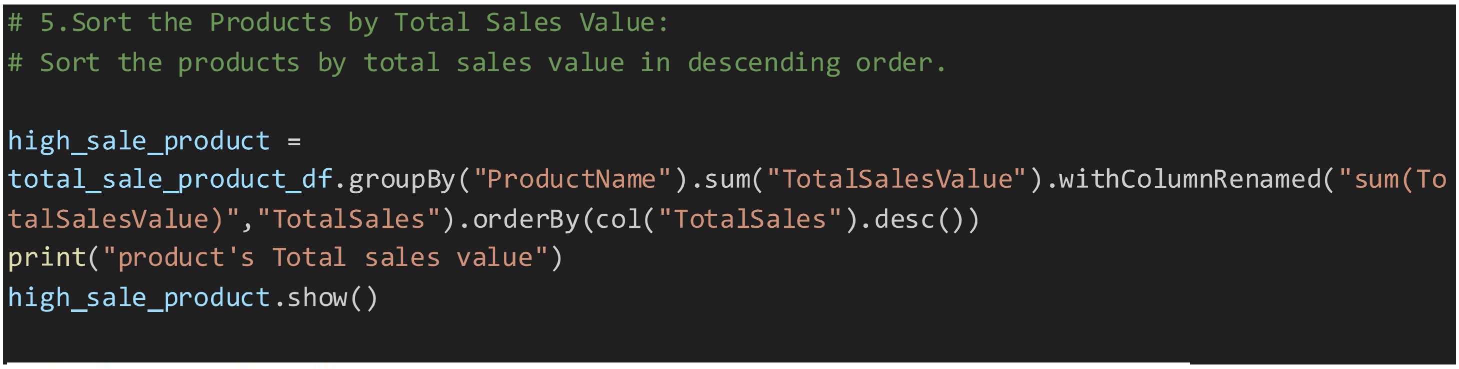
```
Total Sales for Each Product Category:
+-----------+----------+
|   Category|TotalSales|
+-----------+----------+
|Electronics|    260000|
|  Furniture|     85000|
+-----------+----------+
```



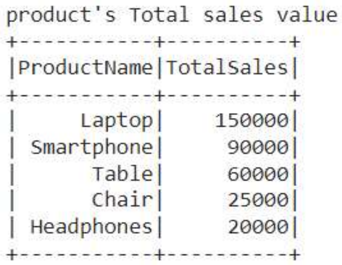
```
# 4.Identify the Top-Selling Product:
# Find the product that generated the highest total sales value.

high_sale_product =
total_sale_product_df.groupBy("ProductName").sum("TotalSalesValue").withColumnRenamed("sum(To
talSalesValue)","TotalSales").orderBy(col("TotalSales").desc()).limit(1)
print("Top-Selling Product:")
high_sale_product.show()
```



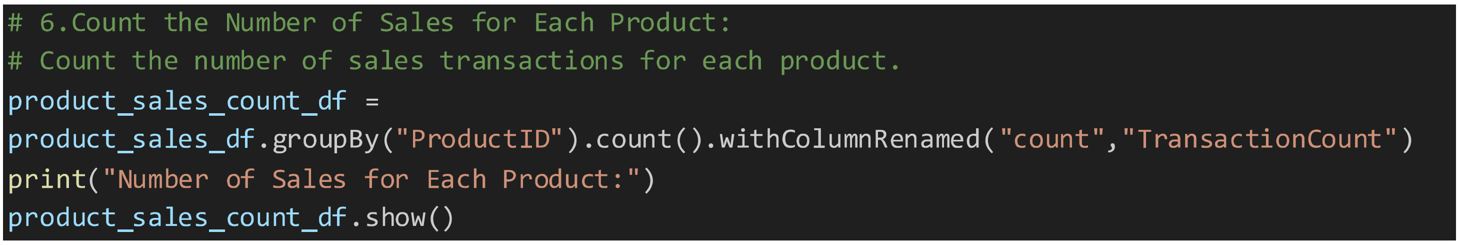
```
Top-Selling Product:
+-----------+----------+
|ProductName|TotalSales|
+-----------+----------+
|     Laptop|    150000|
+-----------+----------+
```



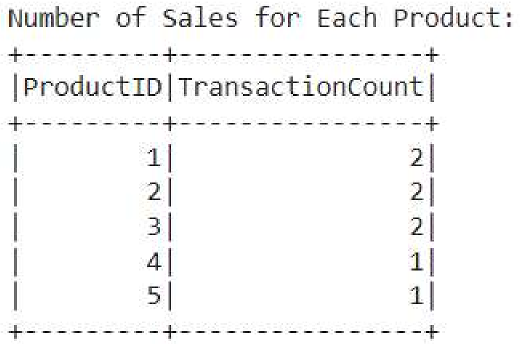
```
# 5.Sort the Products by Total Sales Value:
# Sort the products by total sales value in descending order.

high_sale_product =
total_sale_product_df.groupBy("ProductName").sum("TotalSalesValue").withColumnRenamed("sum(To
talSalesValue)","TotalSales").orderBy(col("TotalSales").desc())
print("product's Total sales value")
high_sale_product.show()
```



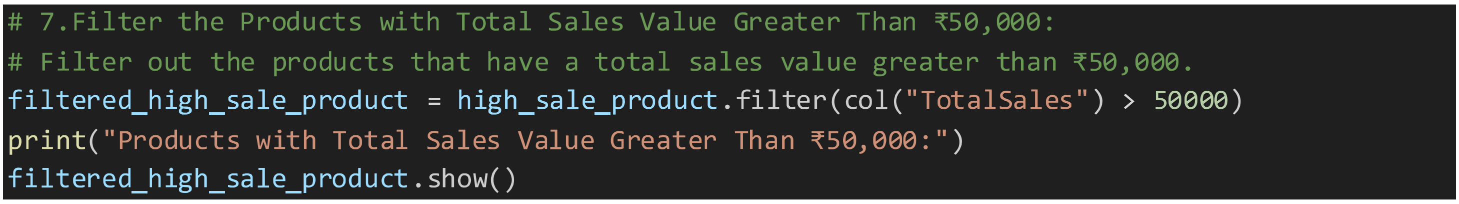
```
product's Total sales value
+-----------+----------+
|ProductName|TotalSales|
+-----------+----------+
|     Laptop|    150000|
| Smartphone|     90000|
|      Table|     60000|
|      Chair|     25000|
| Headphones|     20000|
+-----------+----------+
```



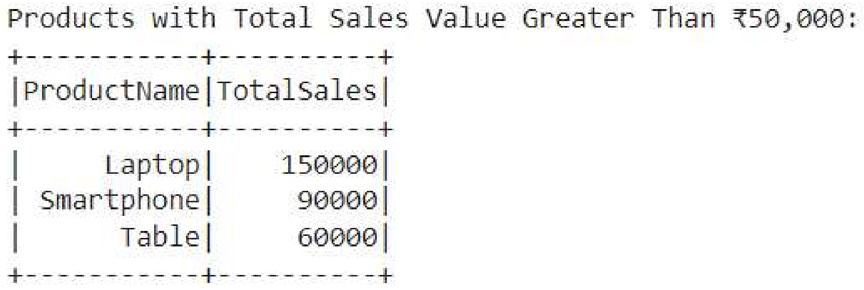
```
# 6.Count the Number of Sales for Each Product:
# Count the number of sales transactions for each product.
product_sales_count_df =
product_sales_df.groupBy("ProductID").count().withColumnRenamed("count","TransactionCount")
print("Number of Sales for Each Product:")
product_sales_count_df.show()
```



```
Number of Sales for Each Product:
+---------+----------------+
|ProductID|TransactionCount|
+---------+----------------+
|        1|               2|
|        2|               2|
|        3|               2|
|        4|               1|
|        5|               1|
+---------+----------------+
```



```
# 7.Filter the Products with Total Sales Value Greater Than ₹50,000:
# Filter out the products that have a total sales value greater than ₹50,000.
filtered_high_sale_product = high_sale_product.filter(col("TotalSales") > 50000)
print("Products with Total Sales Value Greater Than ₹50,000:")
filtered_high_sale_product.show()
```



```
Products with Total Sales Value Greater Than ₹50,000:
+-----------+----------+
|ProductName|TotalSales|
+-----------+----------+
|     Laptop|    150000|
| Smartphone|     90000|
|      Table|     60000|
+-----------+----------+
```