```
from google.colab import files
uploaded = files.upload()
```

```
Choose Files Online Sales Data.csv
• Online Sales Data.csv(text/csv) - 21746 bytes, last modified: 12/8/2025 - 100% done
Saving Online Sales Data.csv to Online Sales Data.csv
```

```
!pip install pandas matplotlib
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (3.10.0)
Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.3.3)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (4.59.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.4.9)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (11.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (3.2.3)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

```
import sqlite3
import pandas as pd
import matplotlib.pyplot as plt
```

```
CSV_FILE = "Online Sales Data.csv"
DB_FILE = "sales_data.db"
CHART_FILE = "sales_chart.png"
```

```
df_csv = pd.read_csv(CSV_FILE)
```

```
df_csv.head()
```

	Transaction ID	Date	Product Category	Product Name	Units Sold	Unit Price	Total Revenue	Region	Payment Method
0	10001	01-01-2024	Electronics	iPhone 14 Pro	2	999.99	1999.98	North America	Credit Card
1	10002	02-01-2024	Home Appliances	Dyson V11 Vacuum	1	499.99	499.99	Europe	PayPal
2	10003	03-01-2024	Clothing	Levi's 501 Jeans	3	69.99	209.97	Asia	Debit Card

Next steps: Generate code with df_csv | View recommended plots | New interactive sheet

```
conn = sqlite3.connect(DB_FILE)
df_csv.to_sql("sales", conn, if_exists="replace", index=False)
```

```
240
```

```
#Total quantity & revenue by product
query1 = """
SELECT "Product Name" AS product,
       SUM("Units Sold") AS total_qty,
       ROUND(SUM("Total Revenue"), 2) AS revenue
FROM sales
GROUP BY "Product Name"
ORDER BY revenue DESC;
"""
pd.read_sql_query(query1, conn)
```

	product	total_qty	revenue
0	Canon EOS R5 Camera	1	3899.99
1	LG OLED TV	2	2599.98
2	MacBook Pro 16-inch	1	2499.99
3	Apple MacBook Pro 16-inch	1	2399.00
4	iPhone 14 Pro	2	1999.98
...	...	...	...
227	Neutrogena Hydro Boost Water Gel	1	16.99
228	Biore UV Aqua Rich Watery Essence Sunscreen	1	15.00
229	The Ordinary Hyaluronic Acid Serum	1	6.80
230	The Ordinary Caffeine Solution 5% + EGCG	1	6.70
231	The Ordinary Niacinamide Serum	1	6.50

232 rows × 3 columns

```
#Revenue by product category
query2 = """
SELECT "Product Category" AS category,
       ROUND(SUM("Total Revenue"), 2) AS total_revenue
FROM sales
GROUP BY "Product Category"
ORDER BY total_revenue DESC;
"""
pd.read_sql_query(query2, conn)
```

	category	total_revenue
0	Electronics	34982.41
1	Home Appliances	18646.16
2	Sports	14326.52
3	Clothing	8128.93
4	Beauty Products	2621.90
5	Books	1861.93

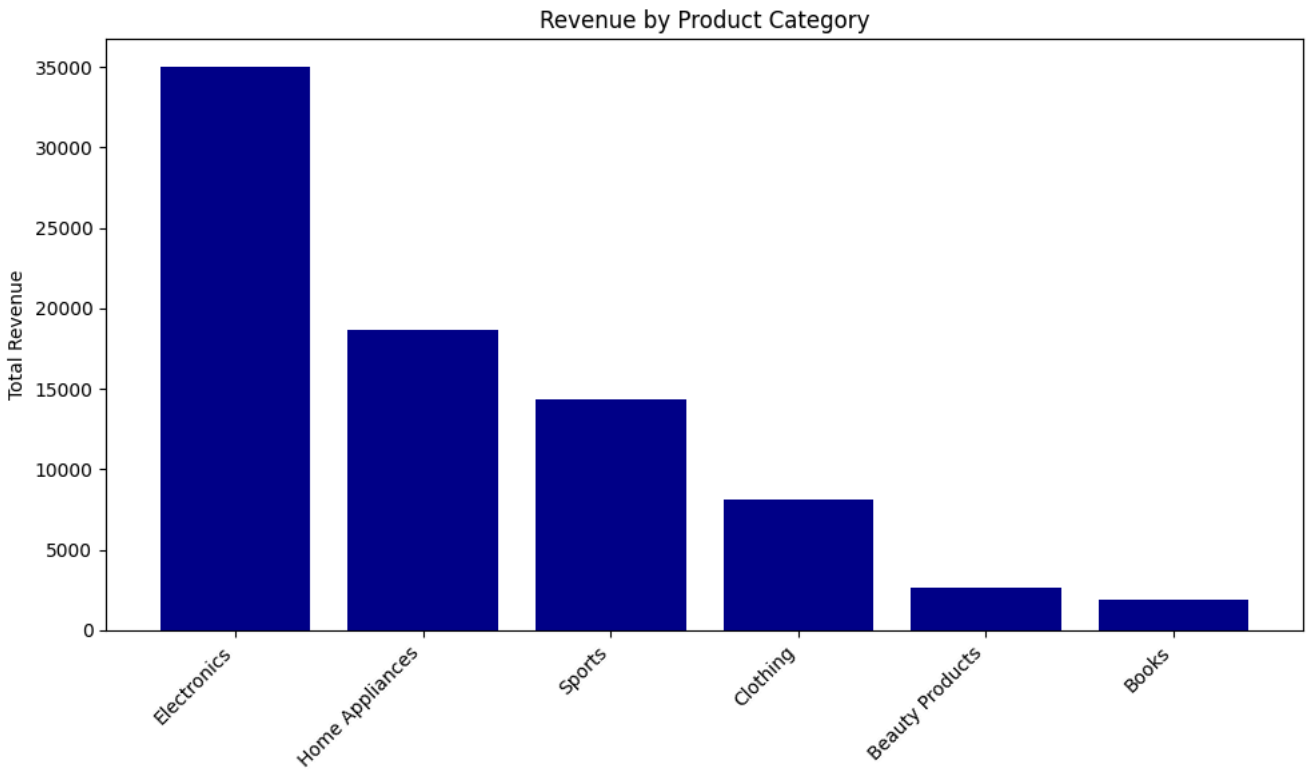
```
#Sales by region
query3 = """
SELECT Region,
       SUM("Units Sold") AS total_units,
       ROUND(SUM("Total Revenue"), 2) AS total_revenue
FROM sales
GROUP BY Region
ORDER BY total_revenue DESC;
"""
pd.read_sql_query(query3, conn)
```

	Region	total_units	total_revenue
0	North America	180	36844.34
1	Asia	233	22455.45
2	Europe	105	21268.06

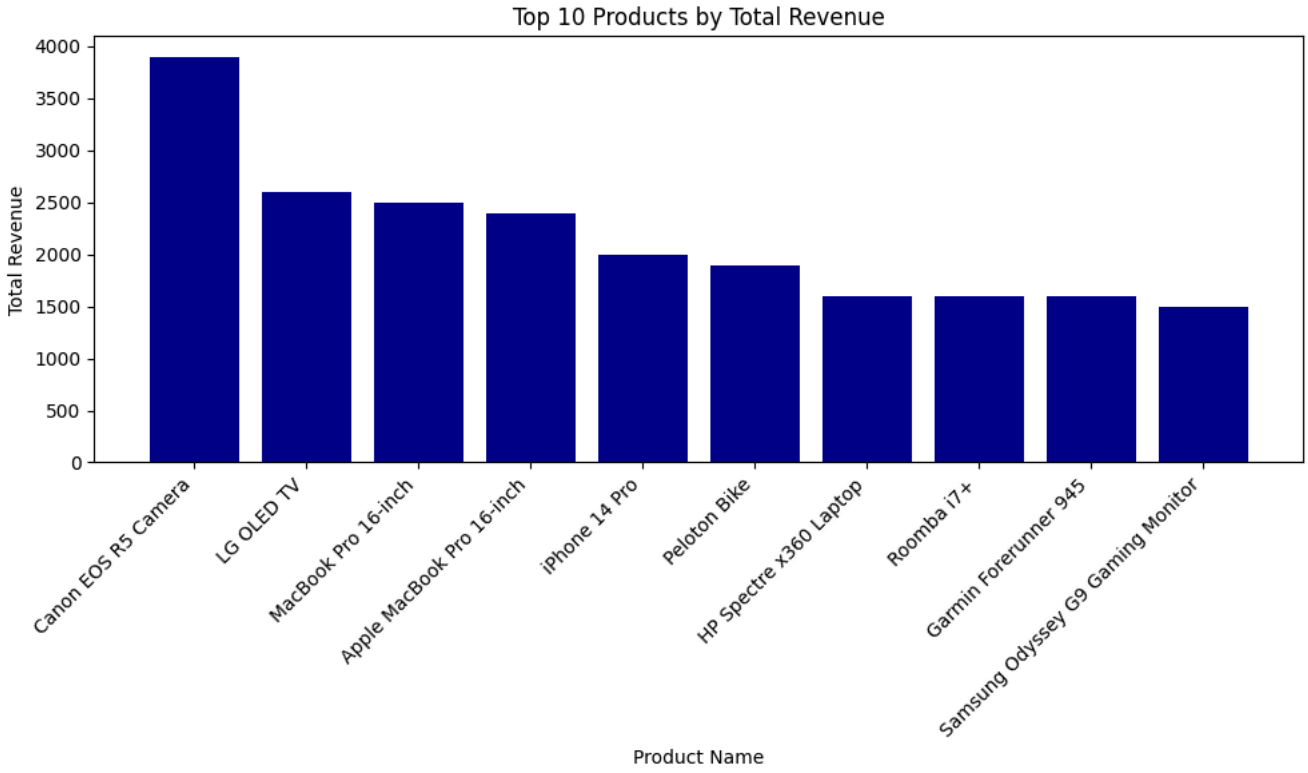
```
#Top 5 best-selling products
query4 = """
SELECT "Product Name",
       SUM("Units Sold") AS total_units
FROM sales
GROUP BY "Product Name"
ORDER BY total_units DESC
LIMIT 5;
"""
pd.read_sql_query(query4, conn)
```

	Product Name	total_units
0	Hanes ComfortSoft T-Shirt	10
1	The Catcher in the Rye by J.D. Salinger	7
2	Yeti Rambler Tumbler	6
3	Spalding NBA Street Basketball	6
4	Nike Air Force 1	6

```
category_revenue = df_csv.groupby("Product Category")["Total Revenue"].sum().reset_index()


category_revenue = category_revenue.sort_values(by="Total Revenue", ascending=False)


plt.figure(figsize=(10, 6))
plt.bar(category_revenue["Product Category"], category_revenue["Total Revenue"], color="Darkblue")
plt.ylabel("Total Revenue")
plt.title("Revenue by Product Category")
plt.xticks(rotation=45, ha="right")
plt.tight_layout()
plt.show()
```




```
product_revenue = df_csv.groupby('Product Name')["Total Revenue"].sum().reset_index()


product_revenue = product_revenue.sort_values(by="Total Revenue", ascending=False).head(10)


plt.figure(figsize=(10, 6))
plt.bar(product_revenue['Product Name'], product_revenue['Total Revenue'], color='Darkblue')
plt.ylabel("Total Revenue")
plt.xlabel("Product Name")
plt.title("Top 10 Products by Total Revenue")
plt.xticks(rotation=45, ha="right")
plt.tight_layout()
plt.show()
```




```
category_revenue = df_csv.groupby('Product Category')['Total Revenue'].sum()


plt.figure(figsize=(8, 8))
plt.pie(category_revenue, labels=category_revenue.index, autopct='%1.1f%%', startangle=140)
plt.title("Revenue Share by Category")
plt.show()
```