Using SQL

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

```python
import pandas as pd
import sqlite3

# Load the CSV
df = pd.read_csv("sample_retail_data.csv")

# Convert 'Order Date' to datetime
df['Order Date'] = pd.to_datetime(df['Order Date'])

# Connect to an in-memory SQLite database
conn = sqlite3.connect(":memory:")

# Write the DataFrame to a SQL table
df.to_sql("retail_data", conn, index=False, if_exists='replace')
```

200

```python
query1 = """
SELECT Category,
       ROUND(SUM(Sales), 2) AS Total_Sales,
       ROUND(SUM(Profit), 2) AS Total_Profit
FROM retail_data
GROUP BY Category
ORDER BY Total_Profit DESC;
"""

pd.read_sql(query1, conn)
```
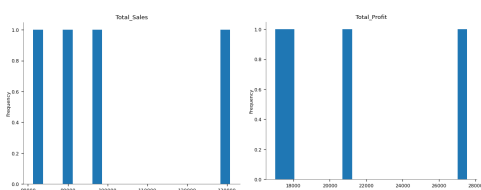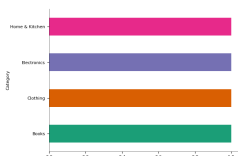
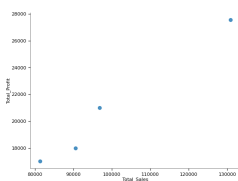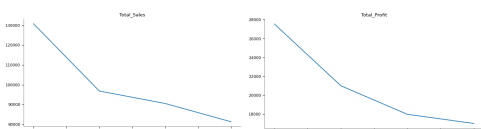|	Category	Total_Sales	Total_Profit
0	Books	130797.95	27550.81
1	Home & Kitchen	96823.12	21006.98
2	Clothing	90530.25	17984.34
3	Electronics	81319.82	17013.68

Distributions

Categorical distributions

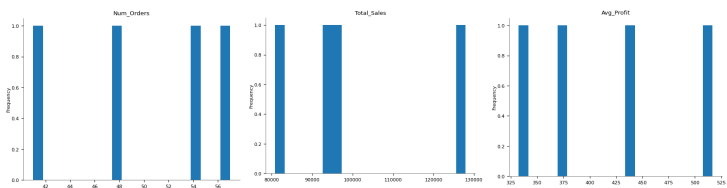

2-d distributions

Values

Start coding or generate with AI.

```python
query2 = """
SELECT Region,
       COUNT(DISTINCT "Order ID") AS Num_Orders,
       ROUND(SUM(Sales), 2) AS Total_Sales,
       ROUND(AVG(Profit), 2) AS Avg_Profit
FROM retail_data
GROUP BY Region;
"""

pd.read_sql(query2, conn)
```
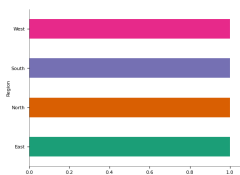
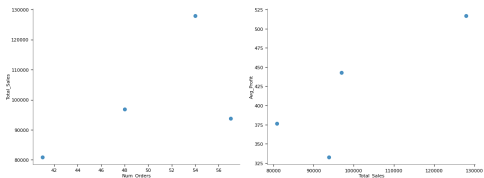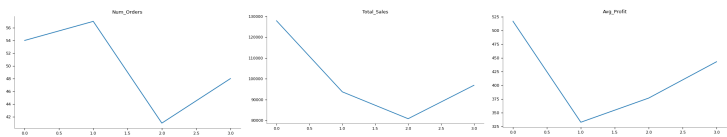|	Region	Num_Orders	Total_Sales	Avg_Profit
0	East	54	127925.74	516.68
1	North	57	93770.90	332.55
2	South	41	80854.78	376.62
3	West	48	96919.72	442.87

Distributions

Categorical distributions

2-d distributions

Values
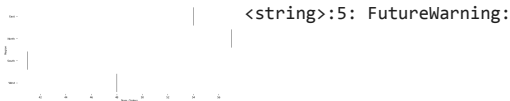

Faceted distributions

```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.
```



```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.
```



```
<string>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.
```
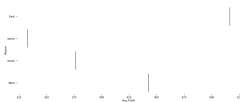
```
query3 = """
SELECT strftime('%Y-%m', "Order Date") AS Month,
       ROUND(SUM(Sales), 2) AS Monthly_Sales,
       ROUND(SUM(Profit), 2) AS Monthly_Profit
FROM retail_data
GROUP BY Month
ORDER BY Month;
"""

monthly_df = pd.read_sql(query3, conn)
monthly_df
```

	Month	Monthly_Sales	Monthly_Profit
0	2024-07	17109.38	3751.10
1	2024-08	34988.68	7320.09
2	2024-09	22451.66	5661.79
3	2024-10	39882.86	8104.74
4	2024-11	47231.73	8059.02
5	2024-12	38608.28	8501.07
6	2025-01	44779.59	9369.68
7	2025-02	35337.71	6501.31
8	2025-03	25435.13	5000.66
9	2025-04	40873.38	9627.58
10	2025-05	27317.26	6178.85
11	2025-06	25455.48	5479.92

Distributions

2-d distributions

Time series

Values


```
import matplotlib.pyplot as plt

monthly_df.plot(x='Month', y=['Monthly_Sales', 'Monthly_Profit'], kind='line', marker='o')
plt.title("Monthly Sales and Profit Trend")
plt.grid(True)
plt.xticks(rotation=45)
plt.show()
```

Monthly Sales and Profit Trend