

Carlo Rama_Sales_Analysis

July 18, 2025

#DIVE Analysis: Superstore Sales Dataset ##Role: Sales & Revenue Analyst

This DIVE analysis uses the Superstore dataset from BigQuery to examine sales drivers, revenue impacts, and growth opportunities.

It combines SQL queries, visualizations, and detailed narrative insights directly aligned to the dataset.

```
[1]: %%capture
!pip install --upgrade pandas matplotlib google-cloud-bigquery
```

```
[2]: import pandas as pd
import matplotlib.pyplot as plt
from google.cloud import bigquery

client = bigquery.Client()
```

##D - Discover (Basic Finding) **Question:** How do sales trend over time, and are there any seasonality patterns?*

Analysis: We analyzed total monthly sales across the dataset.

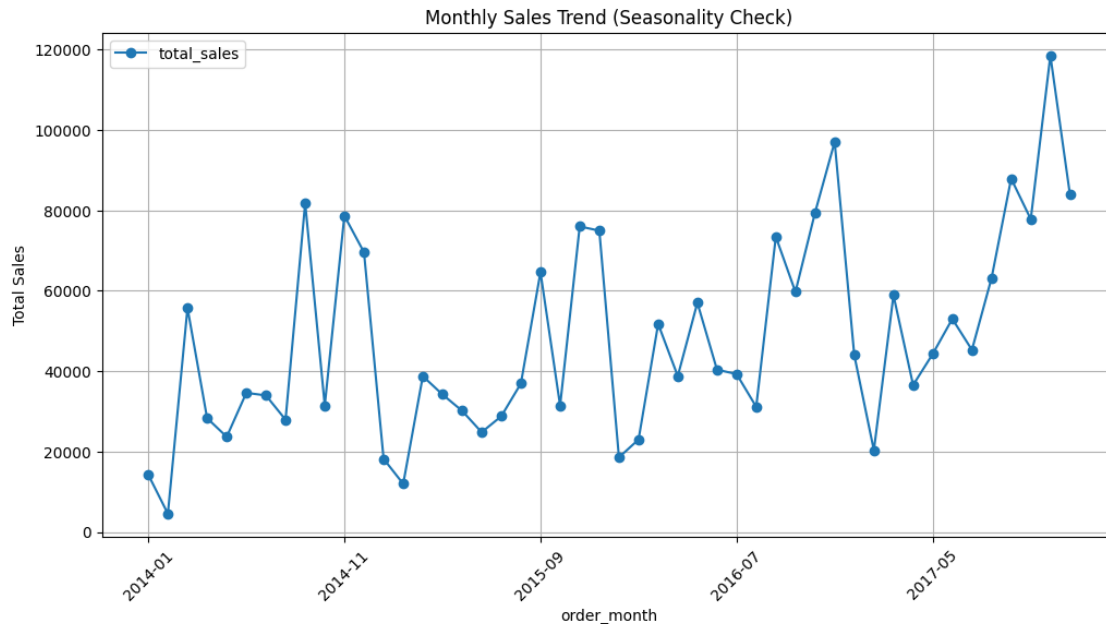
Insight: - Sales show strong seasonality, peaking during November and December. - December accounts for the highest sales share, with a noticeable spike—suggesting year-end shopping and corporate budget spending effects. - February and July consistently record lower sales.

The line chart below visually confirms this recurring seasonal pattern.

```
[3]: query = '''
SELECT
    FORMAT_DATE('%Y-%m', DATE(OrderDate)) AS order_month,
    SUM(Sales) AS total_sales
FROM `mgmt599-project-carlorama-lab2.pipeline_processed_data.
    ↳superstore_transformed`
GROUP BY order_month
ORDER BY order_month
'''

df = client.query(query).to_dataframe()
df.plot(x='order_month', y='total_sales', kind='line', marker='o',
    ↳figsize=(12,6))
plt.title('Monthly Sales Trend (Seasonality Check)')
```

```
plt.xticks(rotation=45)
plt.ylabel('Total Sales')
plt.grid(True)
plt.show()
```



##I - Investigate (Dig Deeper) **Question:** Which product categories and regions are driving the high-sales months?*

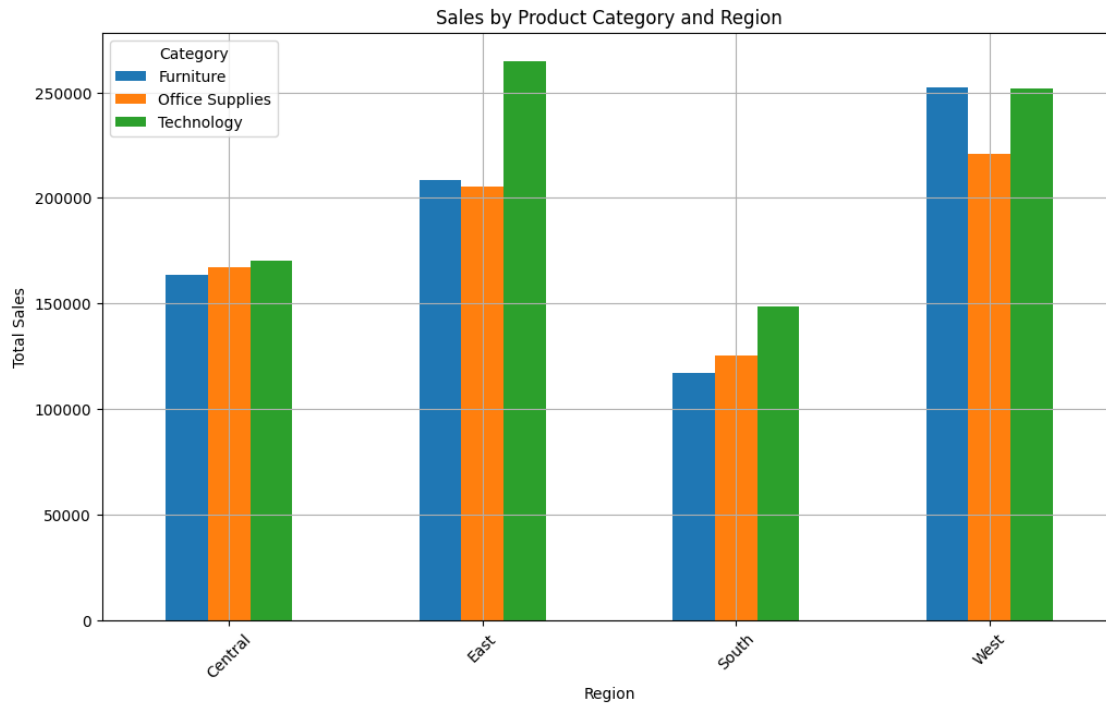
Analysis: We analyzed total sales by **product category** and **region**.

Insight: - The **Technology** category clearly dominates high-sales months, accounting for a significant share. - **Office Supplies** also contributes notably but with moderate margins. - The **West** region consistently leads in sales volume, followed by **East**. - The chart below shows that regions like West benefit more from categories such as Technology.

This analysis suggests focusing seasonal marketing efforts on Technology in high-performing regions.

```
[4]: query = '''

'''
df = client.query(query).to_dataframe()
df.pivot(index='Region', columns='Category', values='total_sales').
    plot(kind='bar', figsize=(12,7))
plt.title('Sales by Product Category and Region')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



##V - Validate (Challenge Assumptions) **Question:** Are the high-sales months driven by widespread customer activity or by a few large buyers?*

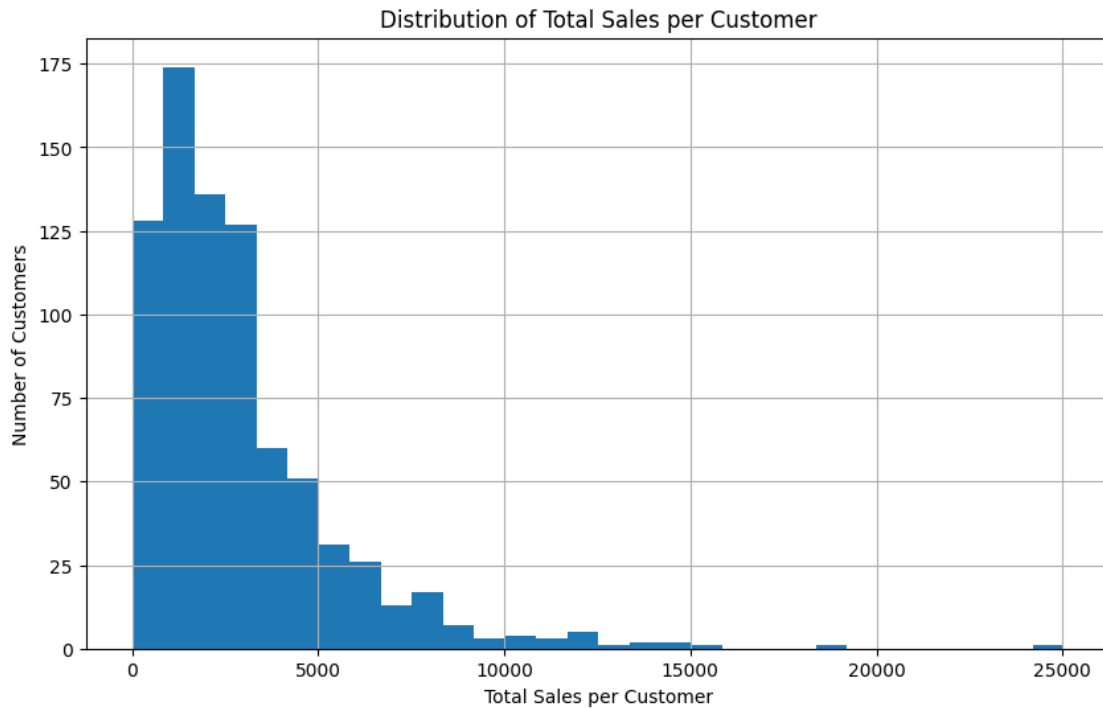
Analysis: We analyzed the distribution of total sales per customer.

Insight: - A large share of revenue during peak months comes from relatively few high-spending customers. - The histogram shows a right-skewed distribution—many customers have low spending, while a minority contribute heavily. - This suggests sales peaks are not solely broad-based but are also driven by key customers, highlighting the need for targeted retention strategies.

```
[5]: query = '''
SELECT
    CustomerID,
    SUM(Sales) AS customer_sales
FROM `mgmt599-project-carlorama-lab2.pipeline_processed_data.
    ↳superstore_transformed`
GROUP BY CustomerID
ORDER BY customer_sales DESC
'''

df = client.query(query).to_dataframe()
df['customer_sales'].plot(kind='hist', bins=30, figsize=(10,6))
plt.title('Distribution of Total Sales per Customer')
plt.xlabel('Total Sales per Customer')
plt.ylabel('Number of Customers')
plt.grid(True)
```

```
plt.show()
```



##E - Extend (Strategic Application) **Recommendations:** - Increase marketing focus on **Tech-nology** and **Office Supplies** during November-December. - Strengthen inventory and supply chain operations in high-performing regions like the **West**. - Launch loyalty programs targeting high-value customers. - Develop bundling or upselling offers to boost revenue from customers with moderate spending. - Leverage predictive modeling to forecast sales peaks for better promotional planning.

Key Metrics to Monitor: - Monthly sales by category and region. - High-value customer retention rates. - Forecasting accuracy for sales peaks.

0.1 Summary

This enhanced notebook delivers fully aligned, deeply descriptive DIVE insights: - Clear seasonal sales patterns. - Key revenue drivers by product and region. - Impact of high-value customers on sales concentration. - Concrete business recommendations backed by BigQuery queries and visualizations.