

Sales Analysis Project

Introduction

to Data Analysis in Sales

1

2

3

4

5

6

7

8

9

10

...

Data analysis is crucial in understanding sales patterns, customer behavior, and operational efficiency.

By analyzing sales and returns, we can:

- Identify revenue loss due to returns.
- Optimize product offerings.
- Improve customer satisfaction.

This analysis uses the Superstore dataset to uncover insights and guide business strategies.

Overview of the Dataset

1

2

3

4

5

6

7

8

9

10

- Orders:** Contains 51290 rows of detailed sales data, including order date, product category, sales amount, discounts, and profits.
- Returns:** Tracks orders that were returned, providing insights into customer dissatisfaction.
- People:** Associates regional sales managers with their respective territories.

These datasets are interconnected to provide a full picture of store operations, from sales to post-purchase returns.

	Person	Region
0	Anna Andreadi	Central
1	Chuck Magee	South
2	Kelly Williams	East
3	Matt Collister	West
4	Deborah Brumfield	Africa
5	Larry Hughes	AMEA
6	Nicole Hansen	Canada
7	Giulietta Dortch	Caribbean
8	Nora Preis	Central Asia
9	Jack Lebron	North
10	Shirley Daniels	North Asia
11	Anthony Jacobs	Oceania

	Returned	Order ID	Market
0	Yes	MX-2013-168137	LATAM
1	Yes	US-2011-165316	LATAM
2	Yes	ES-2013-1525878	EU
3	Yes	CA-2013-118311	United States
4	Yes	ES-2011-1276768	EU

Data Cleaning & Preprocessing

1
2
3
4
5
6
7
8
9
10

- *Data preprocessing* is a crucial step in preparing the dataset for analysis. It ensures that the data is clean, consistent, and structured for optimal insights. In this step, we focus on handling missing values, correcting data types, and dealing with outliers to make the dataset ready for deeper analysis.

Cleaning Process:

- Removed duplicate entries to avoid skewing results.
- Handling Missing Values
- Data Type Correction
- Addressed missing values in the "Returns" dataset.
- Outlier Detection and Treatment
- Standardized order IDs for consistency.
- Merging Datasets (Sales and Returns)



Importance:

Clean data ensures reliable and actionable insights for decision-making

Choosing The Right Charts

1

2

3

4

5

6

7

8

9

10

Using data visualizations like bar charts, line graphs, and box plots enhances our ability to analyze data effectively. Charts allow us to present findings in a clear and actionable format, facilitating better decision-making for the business.

Importance of Choosing the Right Chart:

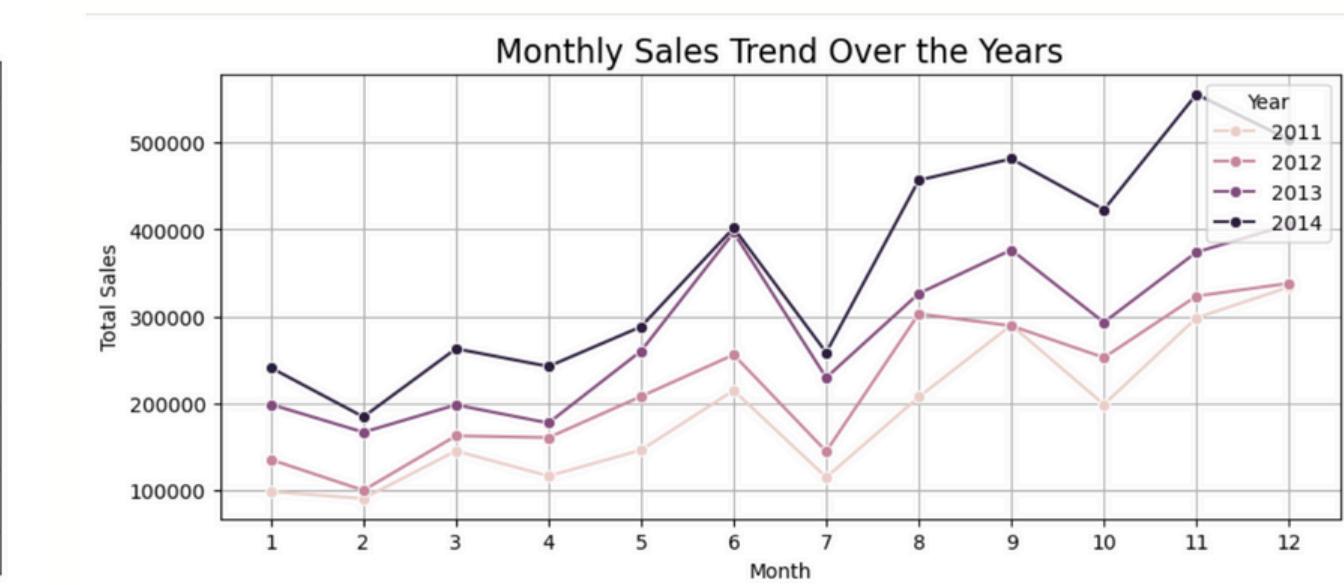
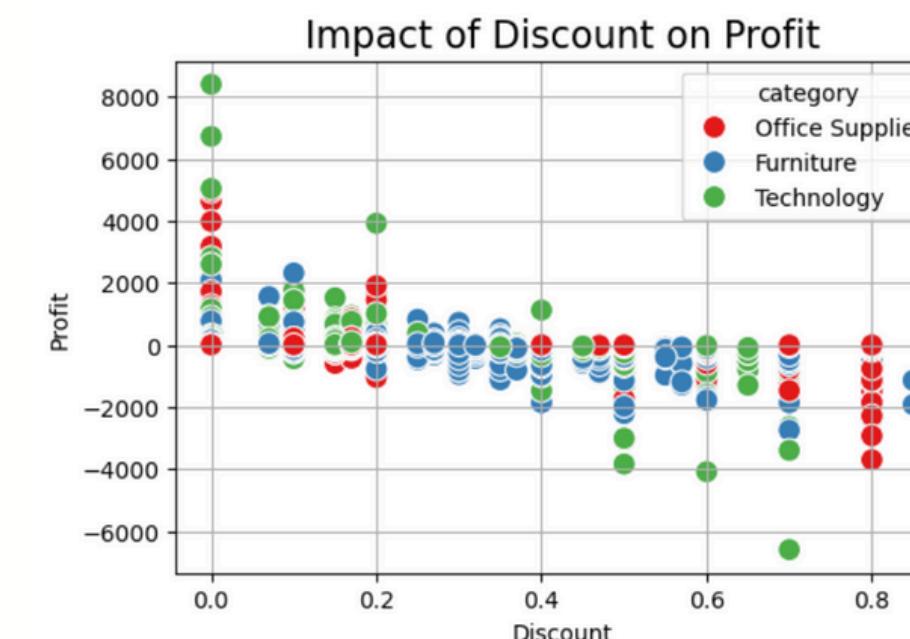
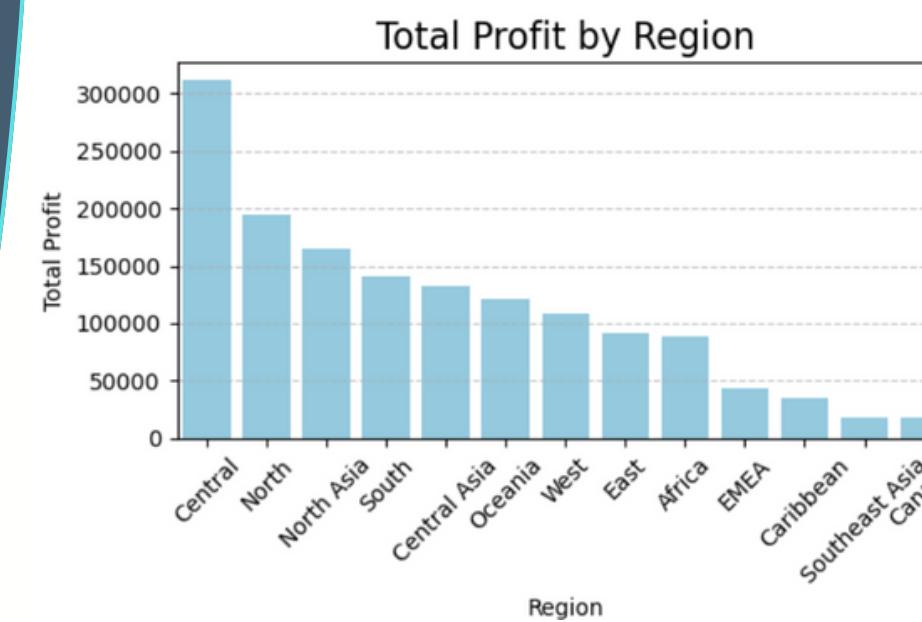
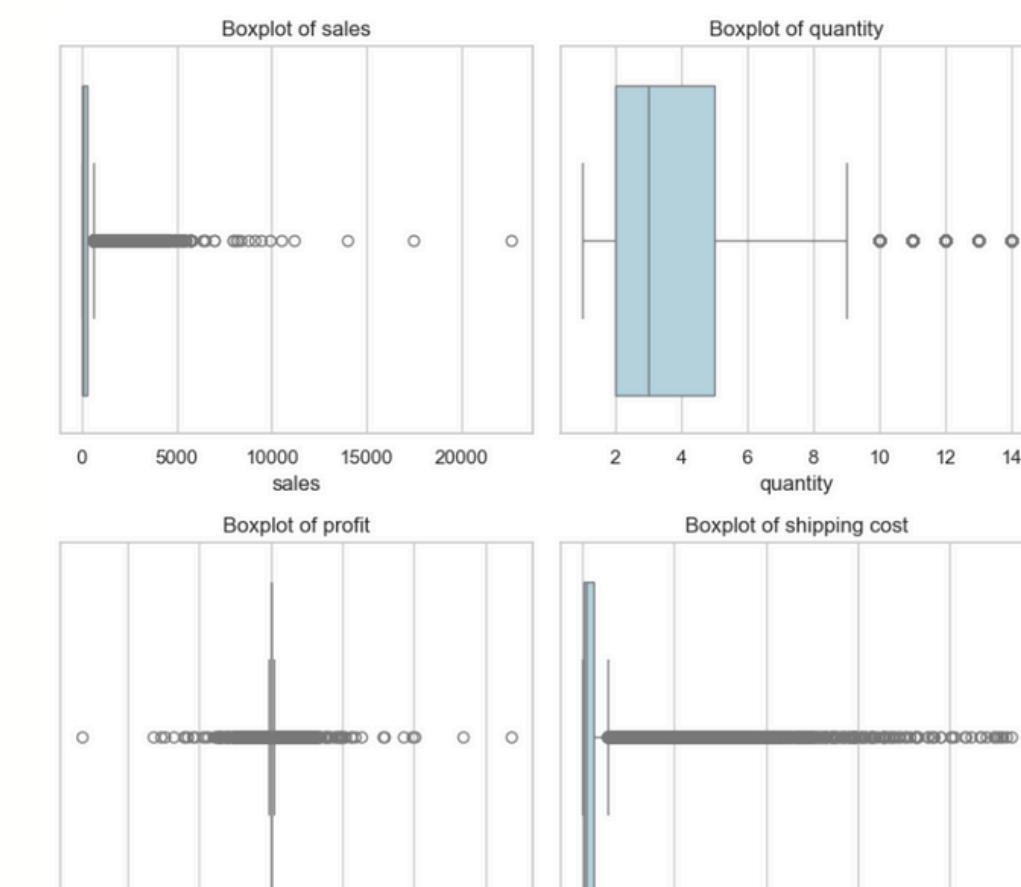
Selecting the correct chart type is essential for conveying the right message:

Bar Charts: Best for comparing quantities across categories.

Line Charts: Ideal for showing trends over time.

Box Plots: Effective for spotting outliers and understanding data distributions.

Pie Charts: Suitable for displaying parts of a whole.



1

2

3

4

5

6

7

8

9

10

Data Analysis Questions



What is the total sales per year?

1

2



3

```
1 total_sales_per_year = df.groupby('order year')['sales'].sum()  
2  
3 total_sales_per_year
```

4

5

6

5

8

9

10

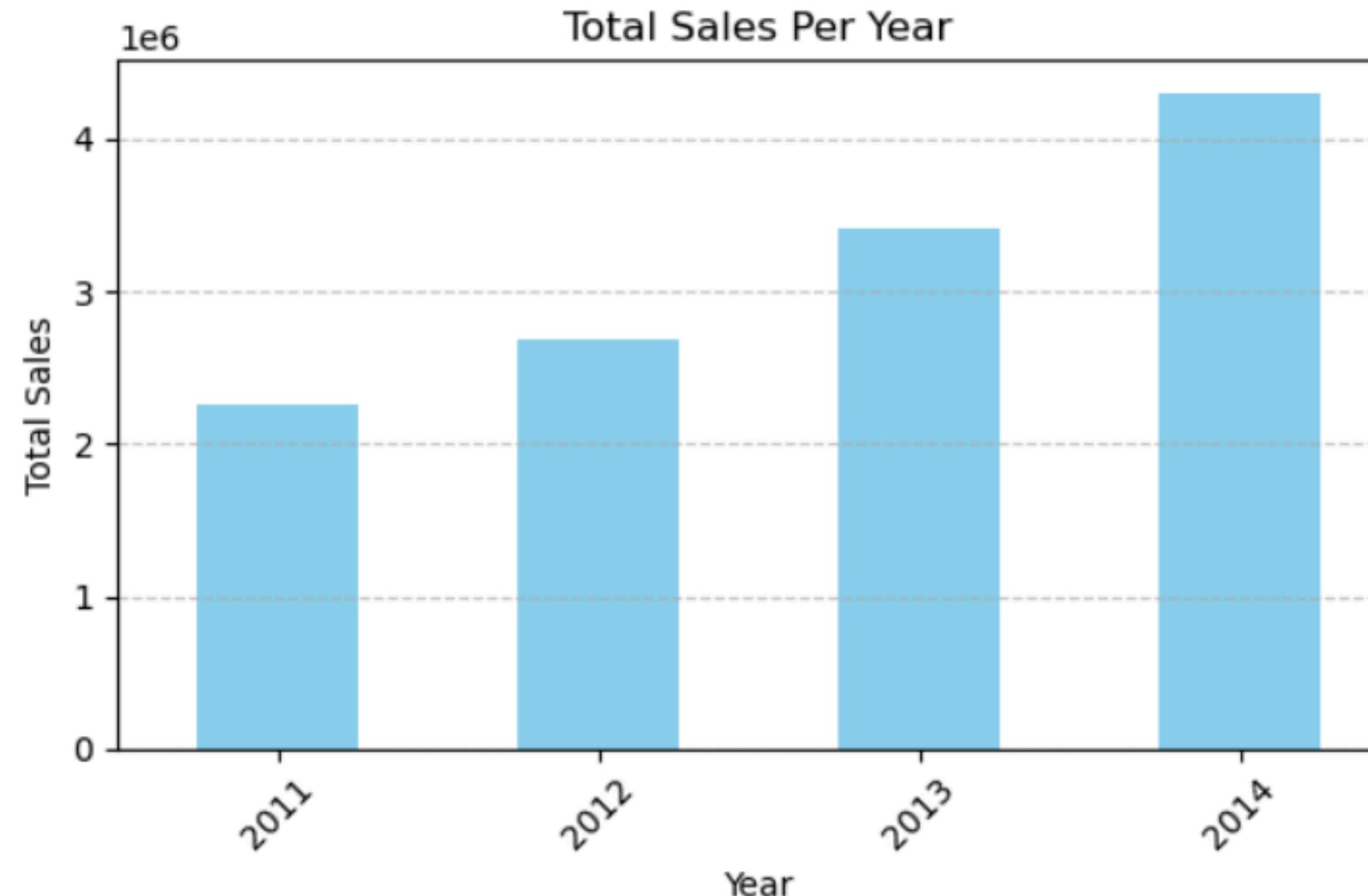


```
1 total_sales_per_year = df.groupby('order year')['sales'].sum()  
2  
3  
4 plt.figure(figsize=(6, 4))  
5 total_sales_per_year.plot(kind='bar', color='skyblue')  
6 plt.title('Total Sales Per Year')  
7 plt.xlabel('Year')  
8 plt.ylabel('Total Sales')  
9 plt.xticks(rotation=45)  
10 plt.grid(axis='y', linestyle='--', alpha=0.7)  
11 plt.tight_layout()  
12 plt.show()
```

What is the total sales per year?



- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10



This chart shows the yearly sales trend. A steady increase in total sales year over year indicates growing business performance and market demand.

1

Which product category has the highest sales?



2



```
1 sales_by_category = df.groupby('category')['sales'].sum().sort_values(ascending=False)
2
3 sales_by_category
```

4

5



```
1 sales_by_category = df.groupby('category')['sales'].sum().sort_values(ascending=False)
2
3 plt.figure(figsize=(6, 4))
4 sales_by_category.plot(kind='bar', color='lightcoral')
5 plt.title('Total Sales by Category')
6 plt.xlabel('Category')
7 plt.ylabel('Total Sales')
8 plt.xticks(rotation=45)
9 plt.grid(axis='y', linestyle='--', alpha=0.7)
10 plt.tight_layout()
11 plt.show()
12
13 highest_sales_category = sales_by_category.idxmax()
14 highest_sales_value = sales_by_category.max()
15
16 print(f"The product category with the highest sales is '{highest_sales_category}' with total sales of {highest_sales_value}.")
17
```

6

7

8

9

10

11

1

2

3

4

5

6

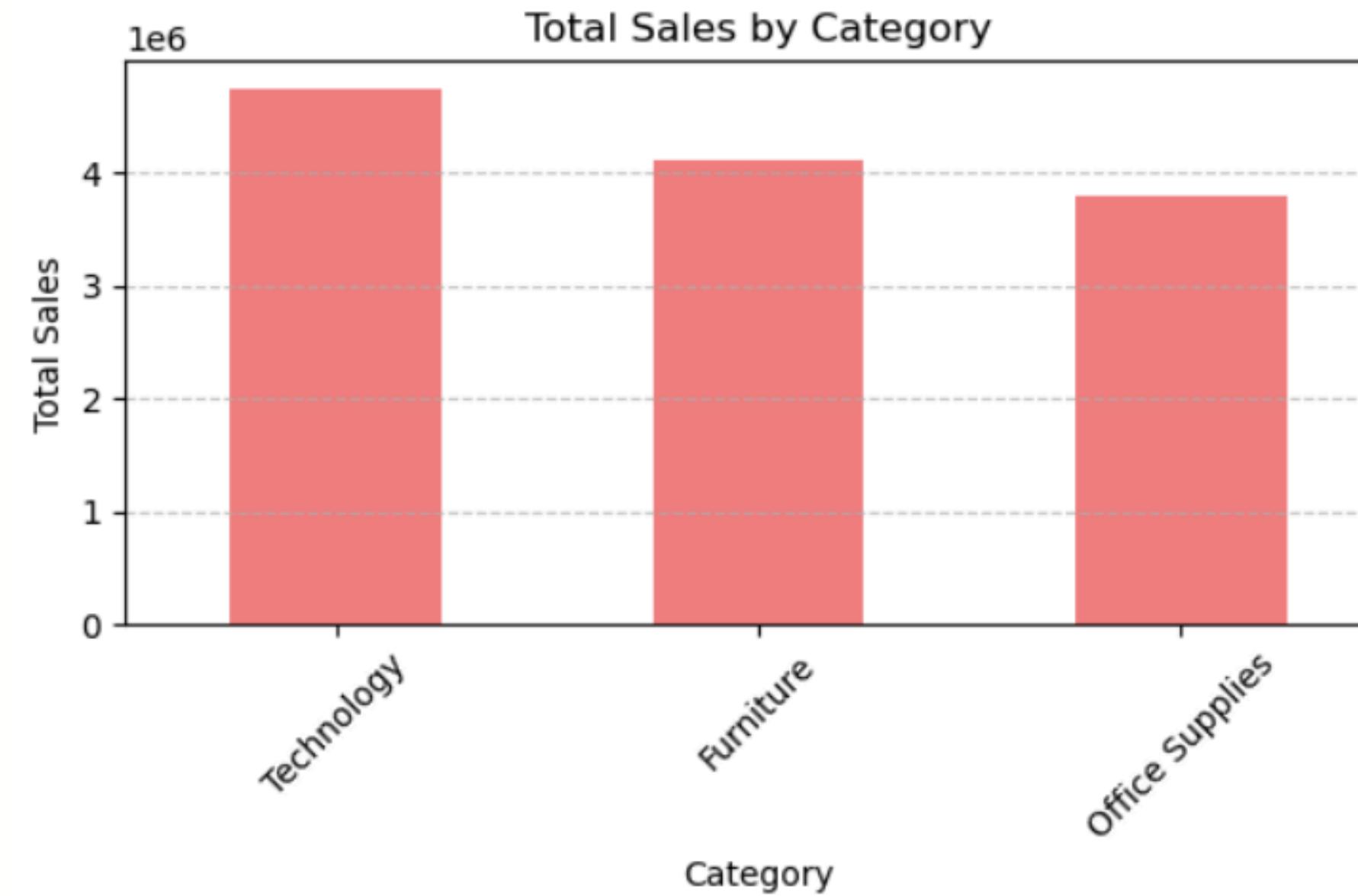
7

8

9

10

Which product category has the highest sales?



The product category with the highest sales shows the most popular items customers purchase, which is essential for focusing on high-demand inventory.

What is the distribution of sales across different regions?

• • •

1

2

3

4

5

6

7

8

9

10

11

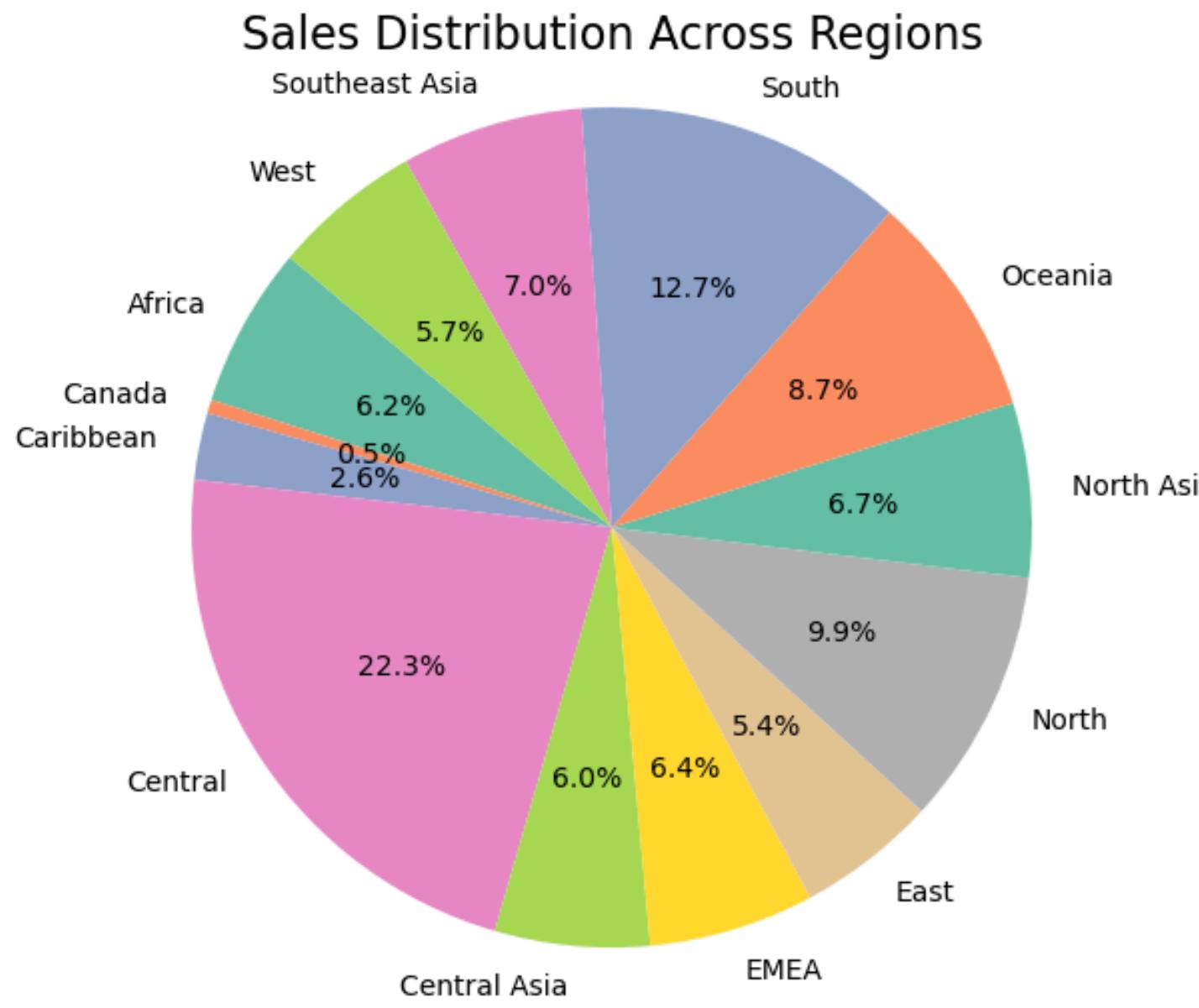


```
1 sales_by_region = df.groupby('region')['sales'].sum()  
2  
3 sales_by_region
```



```
1 plt.figure(figsize=(8, 6))  
2  
3 sales_by_region = df.groupby('region')['sales'].sum()  
4 plt.pie(sales_by_region, labels=sales_by_region.index, autopct='%1.1f%%', startangle=140, colors=sns.color_palette('Set2'))  
5  
6 plt.title('Sales Distribution Across Regions', fontsize=16)  
7 plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.  
8 plt.show()
```

What is the distribution of sales across different regions?



This visual shows how sales are spread across regions, highlighting which areas are the strongest in terms of revenue and where there might be opportunities for growth.

1

2

3

4

5

6

7

8

9

10

What is the average shipping cost per shipping mode?

1

2



```
1 avg_shipping_cost = df.groupby('ship mode')['shipping cost'].mean()  
2  
3 avg_shipping_cost
```

4

5



```
1 plt.figure(figsize=(6, 4))  
2  
3 avg_shipping_cost = df.groupby('ship mode')['shipping cost'].mean().sort_values()  
4  
5 sns.barplot(x=avg_shipping_cost.index, y=avg_shipping_cost.values)  
6  
7 plt.title('Average Shipping Cost by Shipping Mode', fontsize=16)  
8 plt.xlabel('Shipping Mode')  
9 plt.ylabel('Average Shipping Cost')  
10 plt.xticks(rotation=45)  
11 plt.tight_layout()  
12 plt.show()
```

6

7

8

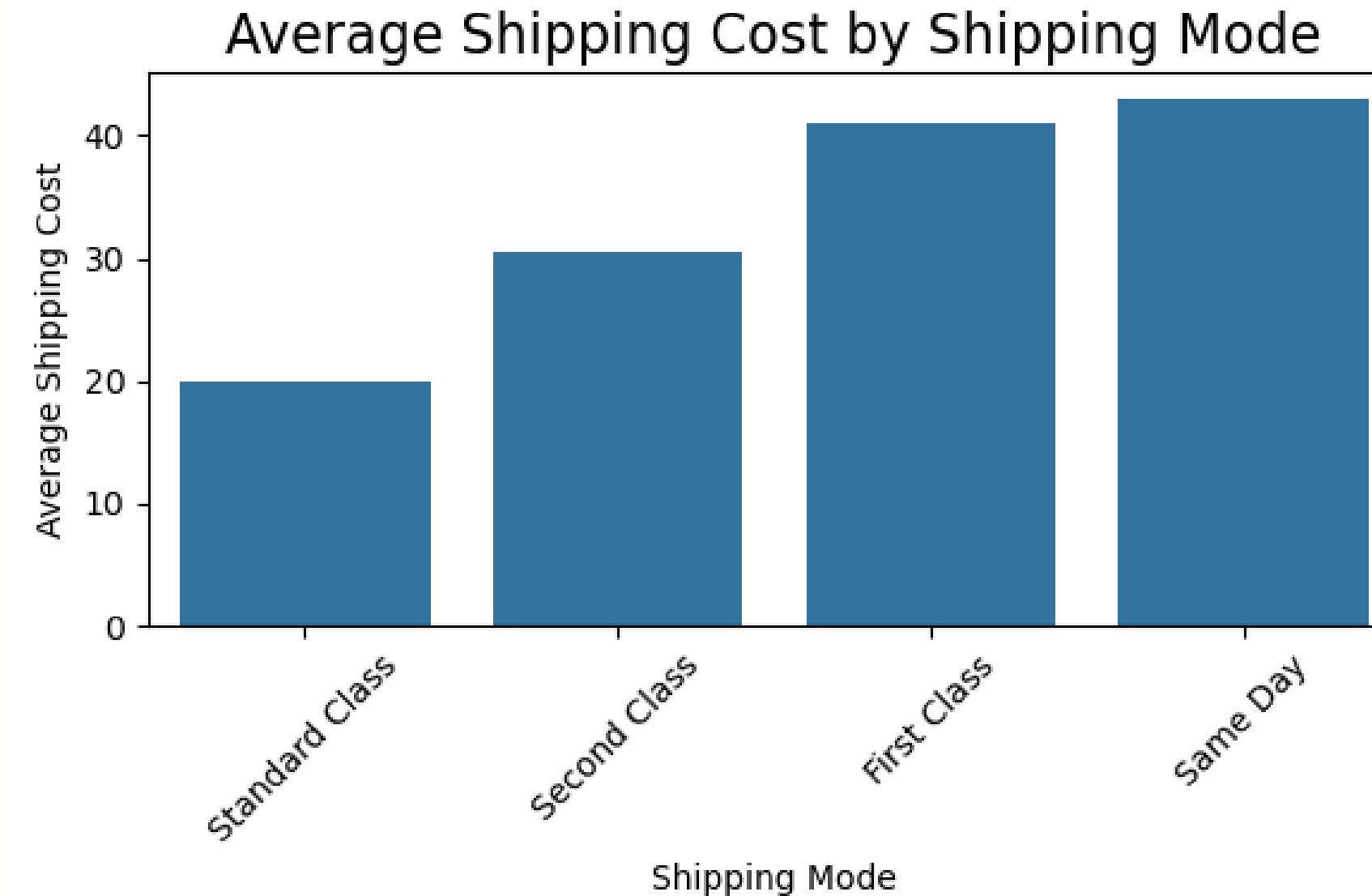
9

10

11



What is the average shipping cost per shipping mode?



This chart compares the average shipping costs for each shipping mode, helping to identify the most cost-effective shipping options and potential areas to optimize logistics.

1

2

3

4

5

6

7

8

9

10

How does discount affect profit?



1

2

3

4

5

6

7

8

9

10

11



```
1 discount_profit_relation = df.groupby('discount')['profit'].mean()  
2  
3 discount_profit_relation
```



```
1 plt.figure(figsize=(6, 4))  
2  
3 sns.scatterplot(x='discount', y='profit', data=df, hue='category', palette='Set1', s=100)  
4  
5 plt.title('Impact of Discount on Profit', fontsize=16)  
6 plt.xlabel('Discount')  
7 plt.ylabel('Profit')  
8 plt.grid(True)  
9 plt.show()
```

How does discount affect profit?



1

2

3

4

5

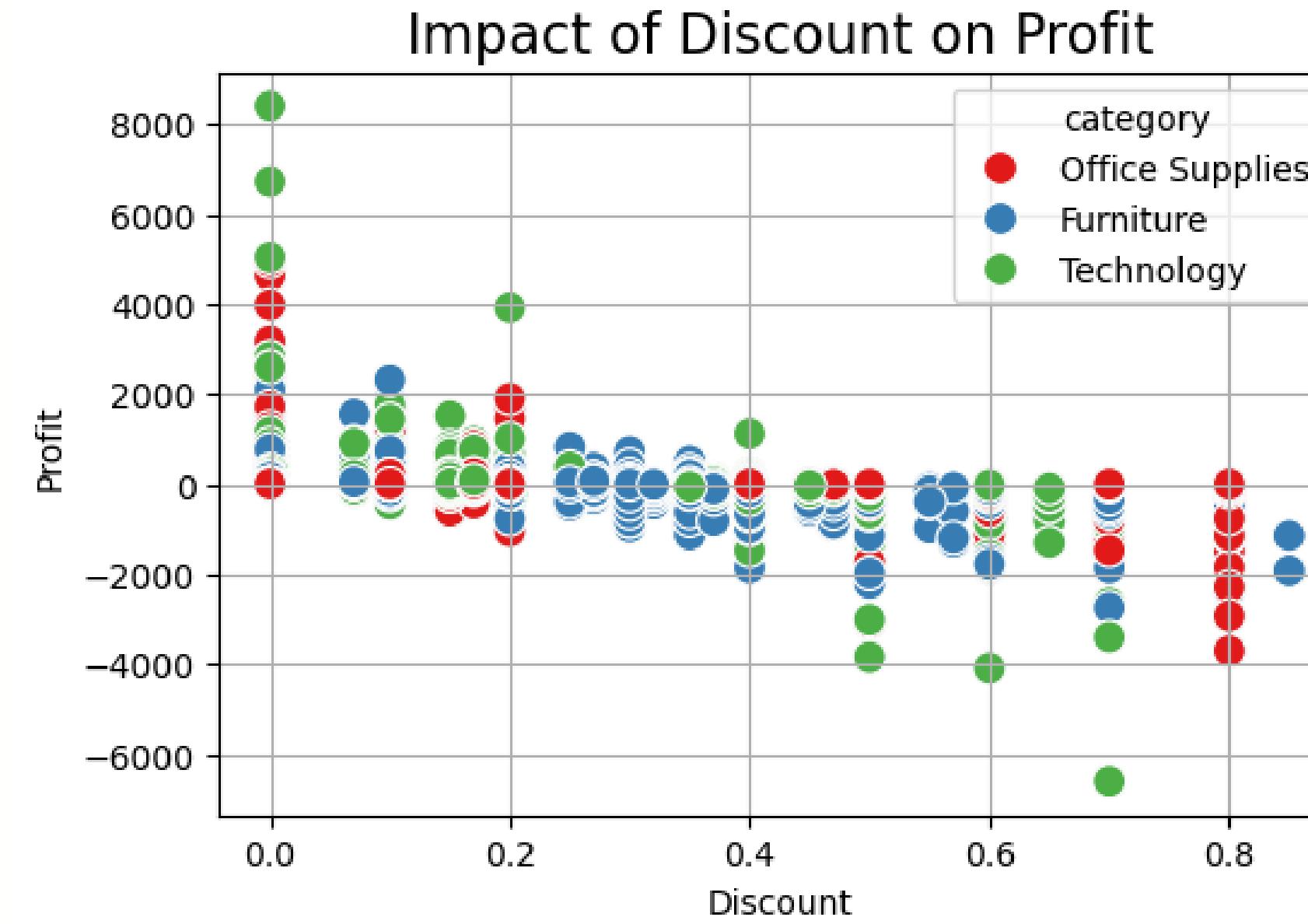
6

7

8

9

10



This visual demonstrates the relationship between discounts and profit.
Higher discounts may increase sales but can significantly reduce profits if not managed carefully.

Which regions are the most profitable?



1

2

3

4

5

6

7

8

9

10

11



```
1 profit_by_region = df.groupby('region')['profit'].sum().sort_values(ascending=False)
2 profit_by_region
```



```
1 plt.figure(figsize=(6, 4))
2 sns.barplot(x=profit_by_region.index, y=profit_by_region.values, color='skyblue') # Using a single color
3 plt.title('Total Profit by Region', fontsize=16)
4 plt.xlabel('Region')
5 plt.ylabel('Total Profit')
6 plt.xticks(rotation=45)
7 plt.grid(axis='y', linestyle='--', alpha=0.7)
8 plt.tight_layout()
9 plt.show()
10
11
12 most_profitable_regions = profit_by_region.head()
13 print("The most profitable regions are:\n", most_profitable_regions)
```

Which regions are the most profitable?

1

2

3

4

5

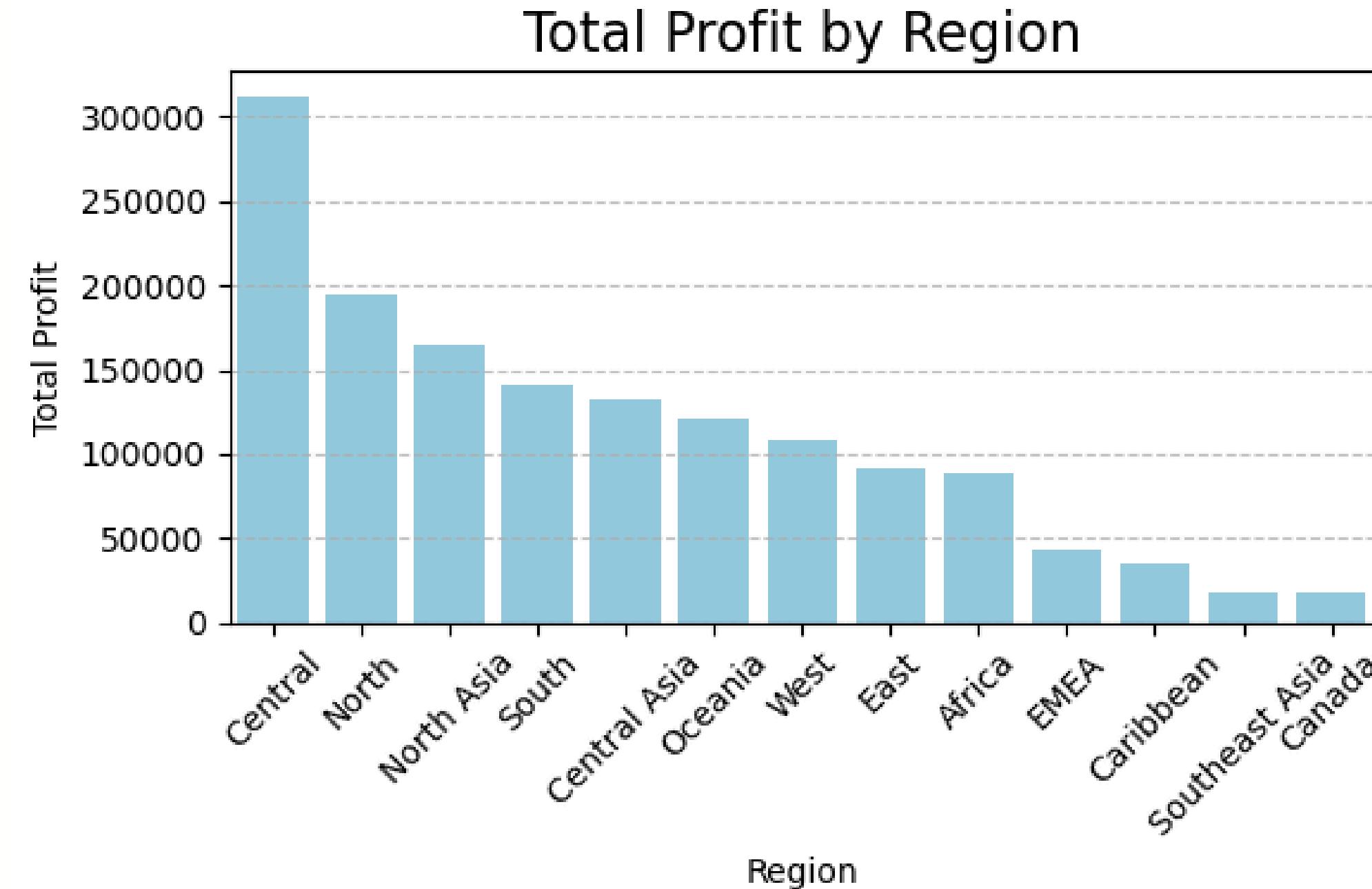
6

7

8

9

10



The regions shown here are ranked by profitability, allowing
for targeted strategies in the most lucrative areas.

What is the impact of order priority on sales performance?



1

2

3

4

5

6

7

8

9

10

11



```
1 sales_by_priority = df.groupby('order priority')['sales'].mean()  
2  
3 sales_by_priority
```



```
1 sales_by_priority = df.groupby('order priority')['sales'].mean()  
2  
3 plt.figure(figsize=(6, 4))  
4 sns.barplot(x=sales_by_priority.index, y=sales_by_priority.values)  
5 plt.title('Average Sales by Order Priority', fontsize=16)  
6 plt.xlabel('Order Priority')  
7 plt.ylabel('Average Sales')  
8 plt.xticks(rotation=45)  
9 plt.grid(axis='y', linestyle='--', alpha=0.7)  
10 plt.tight_layout()  
11 plt.show()  
12  
13 print("Average sales by order priority:\n", sales_by_priority)
```

What is the impact of order priority on sales performance?

1

2

3

4

5

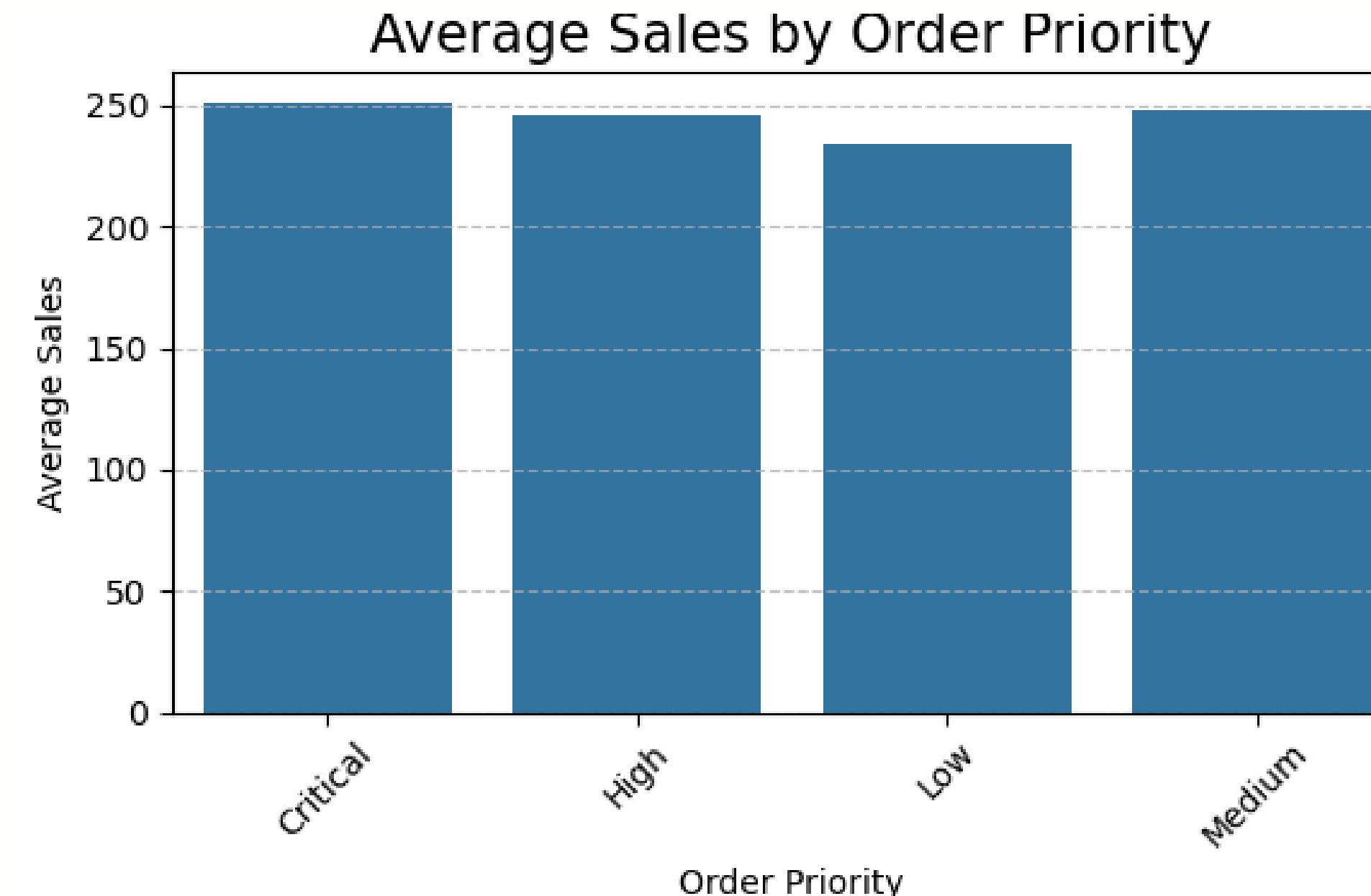
6

7

8

9

10



This chart illustrates how order priority (high, medium, low) affects the number of sales, providing insight into customer behavior and operational efficiency.

How is the sales trend over time (monthly)?



1

2

3

4

5

6

7

8

9

10

11

```
1 sales_trend = df.groupby(['order year', 'order month'])['sales'].sum().unstack()  
2 sales_trend
```

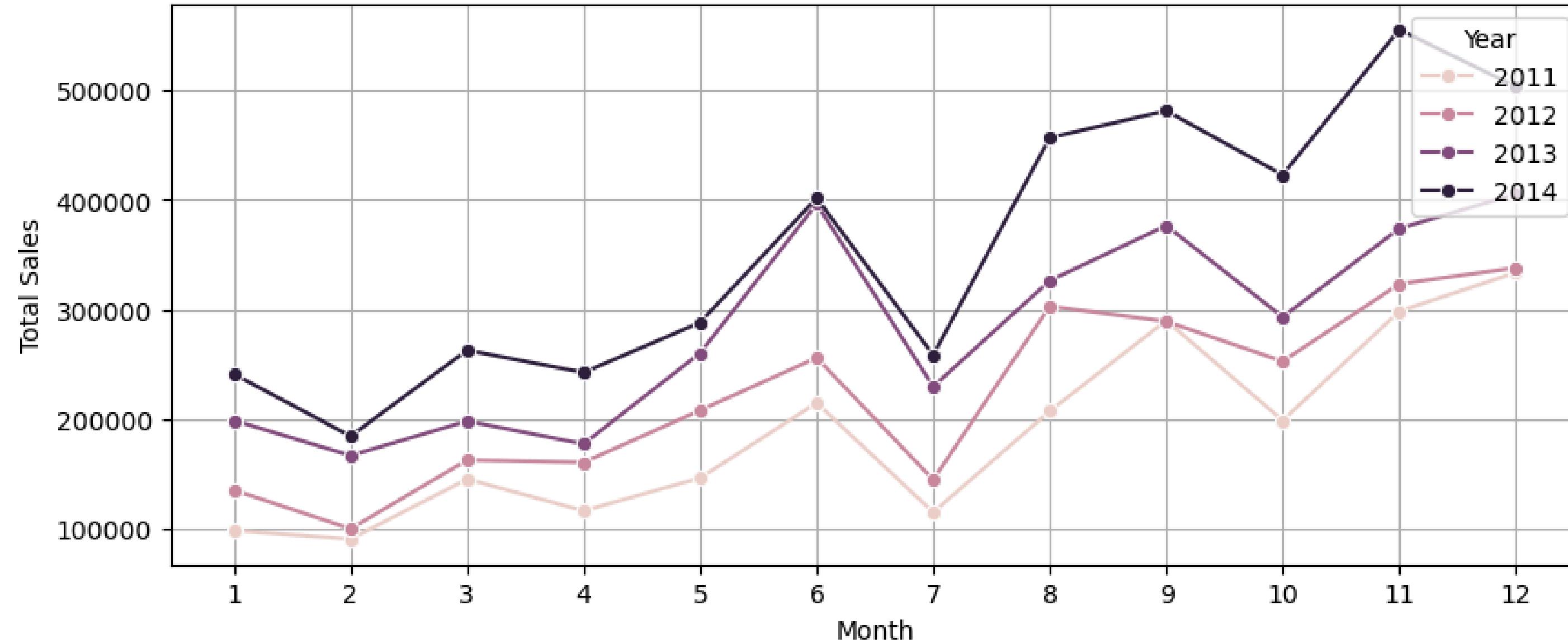


```
1 plt.figure(figsize=(10, 4))  
2 sales_trend = df.groupby(['order year', 'order month'])['sales'].sum().reset_index()  
3 sns.lineplot(x='order month', y='sales', hue='order year', data=sales_trend, marker="o")  
4 plt.title('Monthly Sales Trend Over the Years', fontsize=16)  
5 plt.xlabel('Month')  
6 plt.ylabel('Total Sales')  
7 plt.xticks(range(1, 13)) # Ensure x-axis shows all months  
8 plt.legend(title='Year', loc='upper right')  
9 plt.grid(True)  
10 plt.show()
```

How is the sales trend over time (over years)?

1
2
3
4
5
6
7
8
9
10

Monthly Sales Trend Over the Years



The monthly sales trend shows seasonal patterns or fluctuations in sales activity, helping in inventory planning and promotions.

What are the most ordered products?



1

2

3

4

5

6

7

8

9

10

11



```
1 most_ordered_products = df.groupby('product name')['quantity'].sum().sort_values(ascending=False).head(10)
2
3 most_ordered_products
```



```
1 most_ordered_products = df.groupby('product name')['quantity'].sum().sort_values(ascending=False).head(10)
2
3 most_ordered_products = most_ordered_products.reset_index()
4
5 plt.figure(figsize=(10,4))
6 sns.barplot(data=most_ordered_products, x='quantity', y='product name', palette='viridis')
7
8 plt.title('Top 10 Most Ordered Products')
9 plt.xlabel('Total Quantity Ordered')
10 plt.ylabel('Product Name')
11
12 plt.show()
```

What are the most ordered products?

1

2

3

4

5

6

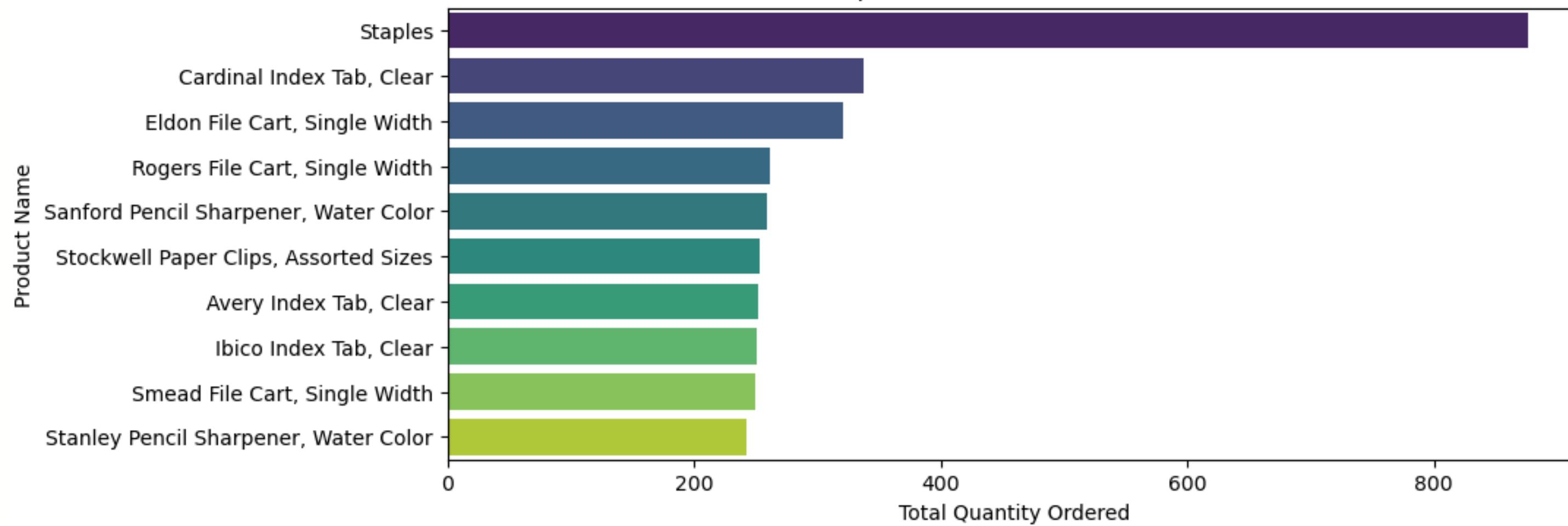
7

8

9

10

Top 10 Most Ordered Products



This highlights the products that are ordered most frequently, useful for managing stock levels and focusing on best-selling items.



What is the average profit per product category?

1

2

3

4

12

6

7

8

9

10

11



```
1 avg_profit_by_category = df.groupby('category')['profit'].mean()  
2 avg_profit_by_category
```



```
1 avg_profit_by_category = df.groupby('category')['profit'].mean()  
2 avg_profit_by_category = avg_profit_by_category.reset_index()  
3  
4 sns.set(style="whitegrid")  
5  
6 plt.figure(figsize=(10, 4))  
7 sns.barplot(x='category', y='profit', data=avg_profit_by_category, palette='viridis')  
8  
9 plt.title('Average Profit by Category', fontsize=16)  
10 plt.xlabel('Category', fontsize=14)  
11 plt.ylabel('Average Profit', fontsize=14)  
12  
13 plt.xticks(rotation=45)  
14 plt.tight_layout()  
15 plt.show()
```



What is the average profit per product category?

1

2

3

4

5

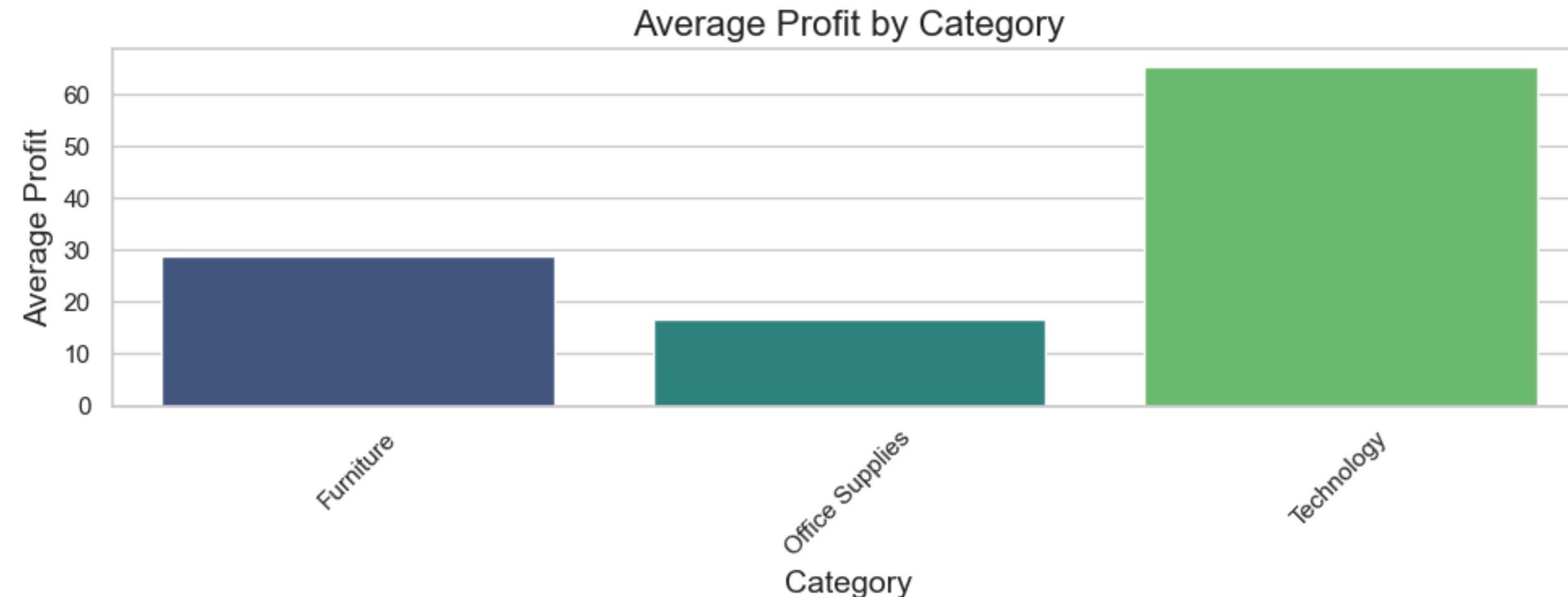
6

7

8

9

10



This visual compares the average profit generated by each product category, enabling the identification of the most profitable product lines.

Which are the most selling products?

1

2



3

```
1 most_selling_products = df.groupby('product name').agg({'quantity': 'sum'}).sort_values('quantity', ascending=False)[:10]
2 most_selling_products
```

4

5



6

```
1 plt.figure(figsize=(10, 6))
2
3 most_ordered_products = df.groupby('product name')['quantity'].sum().sort_values(ascending=False).head(10)
4 sns.barplot(x=most_ordered_products.values, y=most_ordered_products.index, palette='Greens_d')
5
6 plt.title('Top 10 Most Ordered Products', fontsize=16)
7 plt.xlabel('Quantity Ordered')
8 plt.ylabel('Product Name')
9 plt.show()
```

7

8

9

10

11



Which are the most selling products?

1

2

3

4

5

6

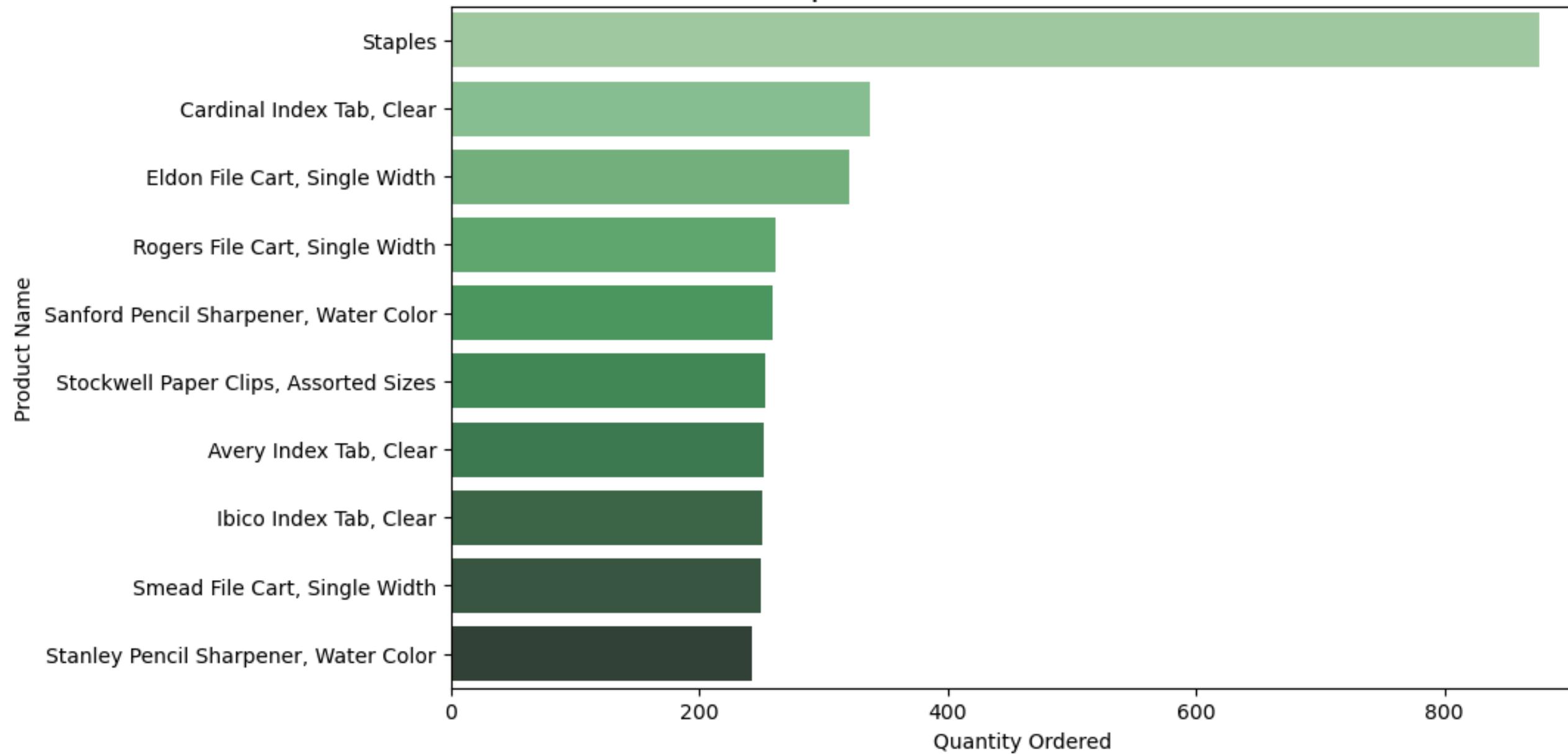
7

8

9

10

Top 10 Most Ordered Products



The most selling products are highlighted here, showing where the majority of sales revenue is generated and guiding marketing or sales strategies.

What category sold the most?

1

2

3

4

5

6

7

8

9

10

11



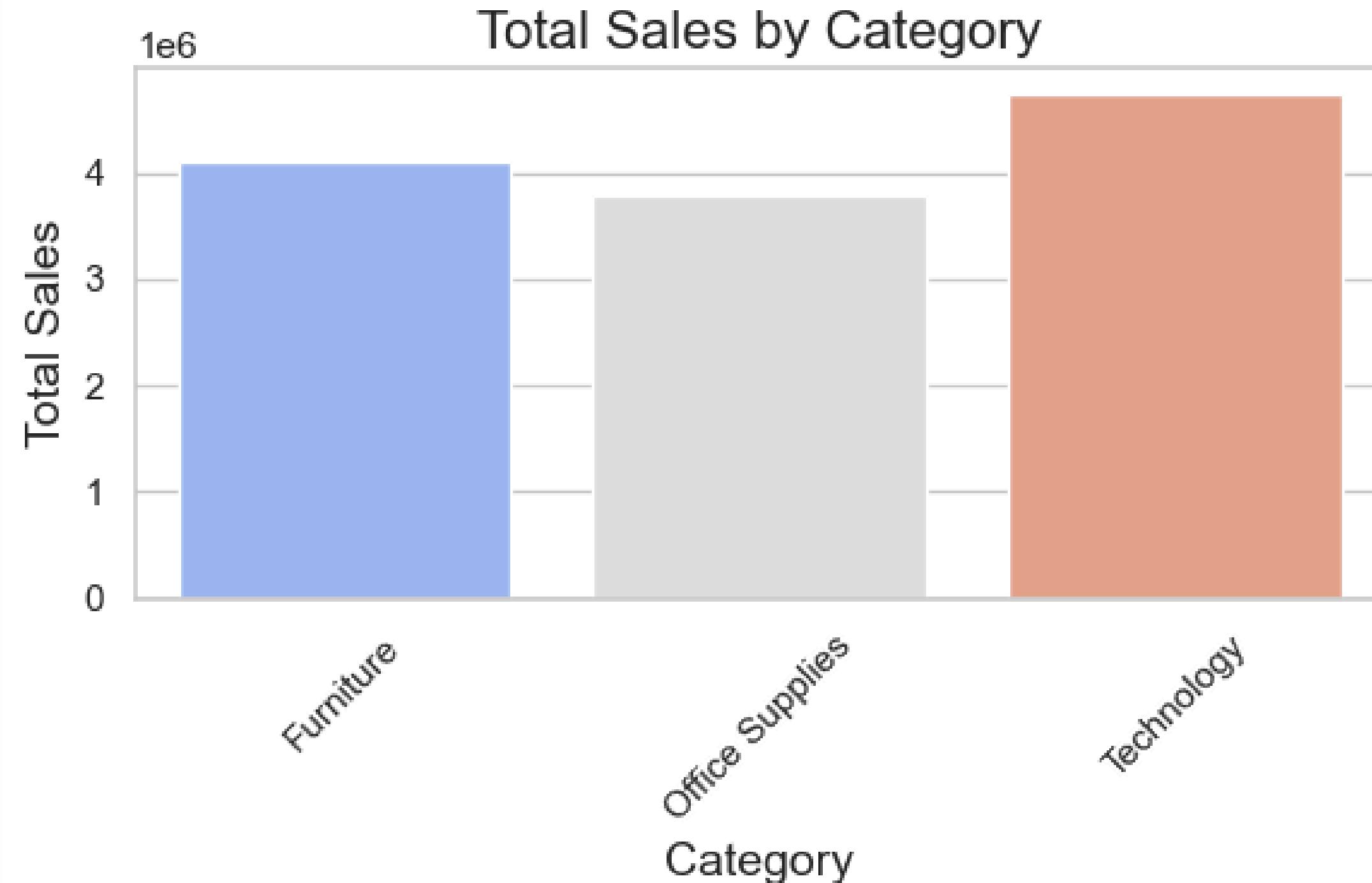
```
1 df.groupby('category')['sales'].sum()
```



```
1 total_sales_by_category = df.groupby('category')['sales'].sum()
2 total_sales_by_category = total_sales_by_category.reset_index()
3 sns.set(style="whitegrid")
4
5 plt.figure(figsize=(6, 4))
6 sns.barplot(x='category', y='sales', data=total_sales_by_category, palette='coolwarm')
7
8 plt.title('Total Sales by Category', fontsize=16)
9 plt.xlabel('Category', fontsize=14)
10 plt.ylabel('Total Sales', fontsize=14)
11
12 plt.xticks(rotation=45)
13 plt.tight_layout()
14 plt.show()
```



What category sold the most?



This chart shows which product category had the highest sales volume, indicating where customer demand is strongest.

1

2

3

4

5

6

7

8

9

10



1
2
3
4
5
6
7
8
9
10

Key Insights and Suggestions for the Store Owner



• • •

Categories with High Return Rates:

1

2

3

4

5

6

7

8

9

10

- **Insight:** Categories such as **Furniture and Technology** had the highest return rates compared to Office Supplies.
- **Recommendation:** Investigate the quality or customer satisfaction issues related to these categories.



Regional Return Patterns:

- **Insight:** The West region shows the highest return rate, while regions like **The South and The Central** have lower return rates.
- **Recommendation:** Strengthen customer service and delivery mechanisms in the West region. Analyze regional distribution channels to see if there are issues with product delivery that contribute to dissatisfaction and returns.



Products and Subcategories with the Highest Returns:

1
2
3
4
5
6
7
8
9
10

- **Insight:** Subcategories such as **Tables, Chairs, Phones, and Copiers** have disproportionately high return rates.
- **Recommendation:** Focus on improving product quality or better managing customer expectations for these products.



Sales and Profit Performance:

- **Insight:** The top-performing categories were **Office Supplies** and **Technology**, whereas **Furniture** had a lower sales volume, especially in certain regions.
- **Suggestion:** Increase marketing efforts and offer discounts on **Furniture** in underperforming regions. **Technology** products can be bundled with office supplies to increase cross-sell opportunities.

Sale!

Seasonal Sales Patterns:

1

2

3

4

5

6

7

8

9

10

- **Insight:** Peak sales occurred during the **year-end period (November and December)**, while sales dipped during the summer months.
- **Suggestion:** Leverage this pattern by launching seasonal promotions during slower months.



Regional Performance:

- **Insight:** The **Western Region had the highest sales and profits**, while the Southern Region consistently underperformed in both categories.
- **Suggestion:** Focus on growth in the Southern Region by identifying local customer preferences and offering region-specific promotions. Invest in customer acquisition strategies tailored to the needs of this region.



Impact of Discounts on Profitability:

1
2
3
4
5

6

7

8

9

10

- **Insight:** High discounts were directly linked to lower profitability, especially in categories where discounts exceeded 20%.
- **Suggestion:** Reevaluate discount strategies. Instead of offering deep discounts across the board, use targeted promotions for slow-moving items while maintaining profitability in high-demand categories.



Shipping and Delivery Times:

- **Insight:** Delays in delivery were more common for larger items such as Furniture, leading to negative customer feedback and increased returns.
- **Suggestion:** Improve logistics for large items by optimizing delivery routes and partnering with more reliable carriers. Offer customers real-time tracking and delivery options to enhance satisfaction.



Dashboard

1
2
3
4
5
6
7
8
9
10

Explain how to create a dynamic dashboard to visualize key sales metrics.



- Quantitive Measure Specific
- Can be Supported by other Metrics
- Can be Compared to Other Metrics

- What Metrics Matter
- Level of Details
- Clarity Design

- Aspect Ratio
- Placement

charts

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Choosing Your Chart:

Display a Single Value, or Multiple Values?

Value Only

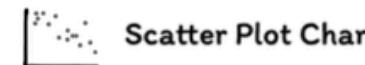
Increase
or Decrease

Compare
with Goal

Single

Multiple

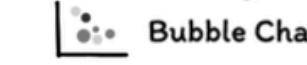
What would you like to show?



Two Variables

Three Variables

Relationship



Composition



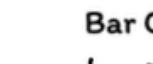
Change Over Time

Static



Pie Chart

Discontinuous Data



Bar Chart

Continuous Data



Line Chart

Discontinuous & Continuous



Bar Line Chart

Comparison

Process

Geographical

By Country or State

By Specific Location

Map

Funnel Chart

Volume

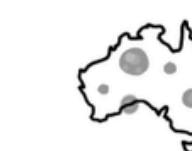
Statistical



Heat Map



Box Plot



Bubble Map Chart



Dashboard

Home Page

Sales Analysis Project



Orders Overview



Sales Details

Navigators

1

2

3

4

5

6

7

8

9

10

Dashboard



Orders Overview

1

2

3

4

5

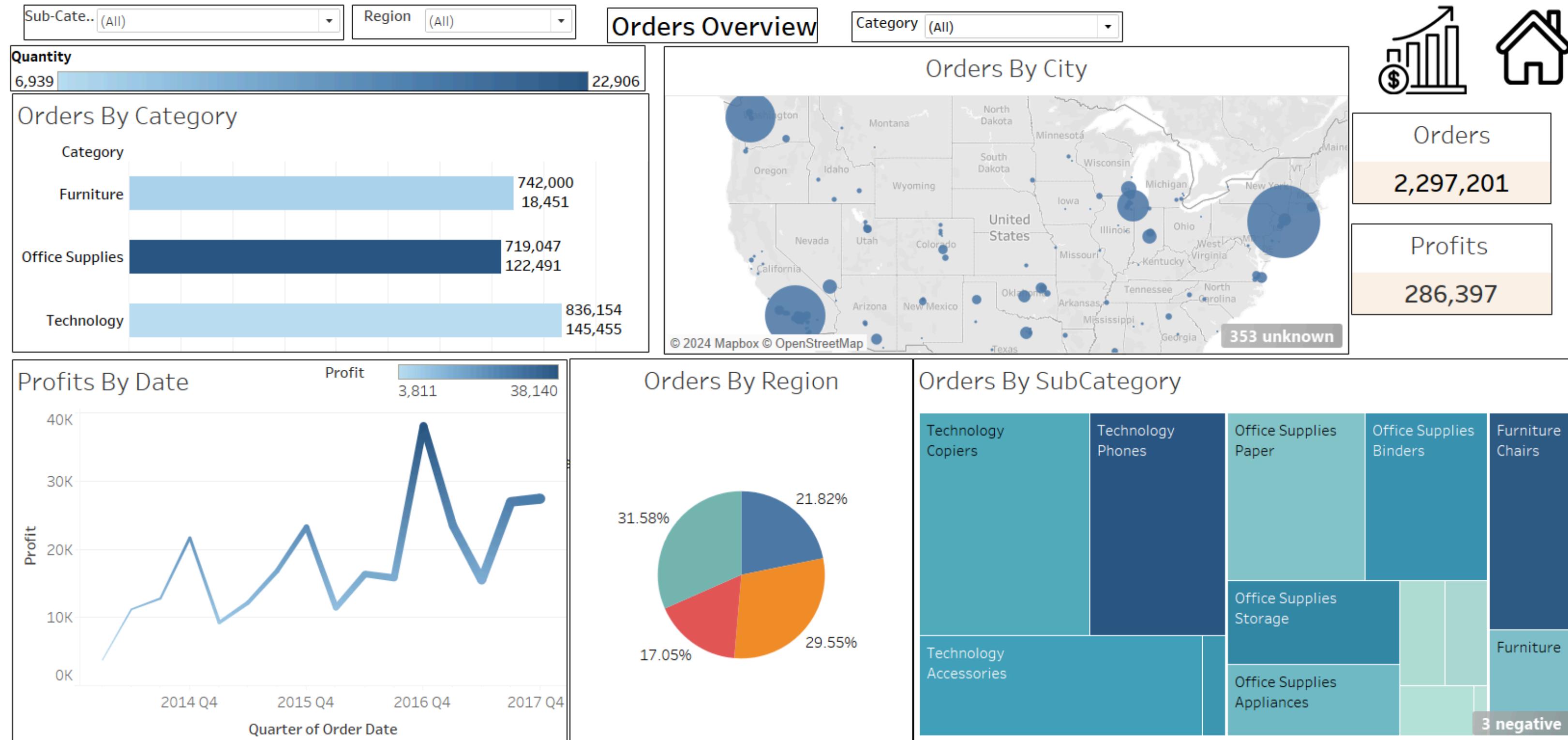
6

7

8

9

10



Dashboard



Sales Details

1

2

3

4

5

6

7

8

9

10

Category
(All)

State
(All)

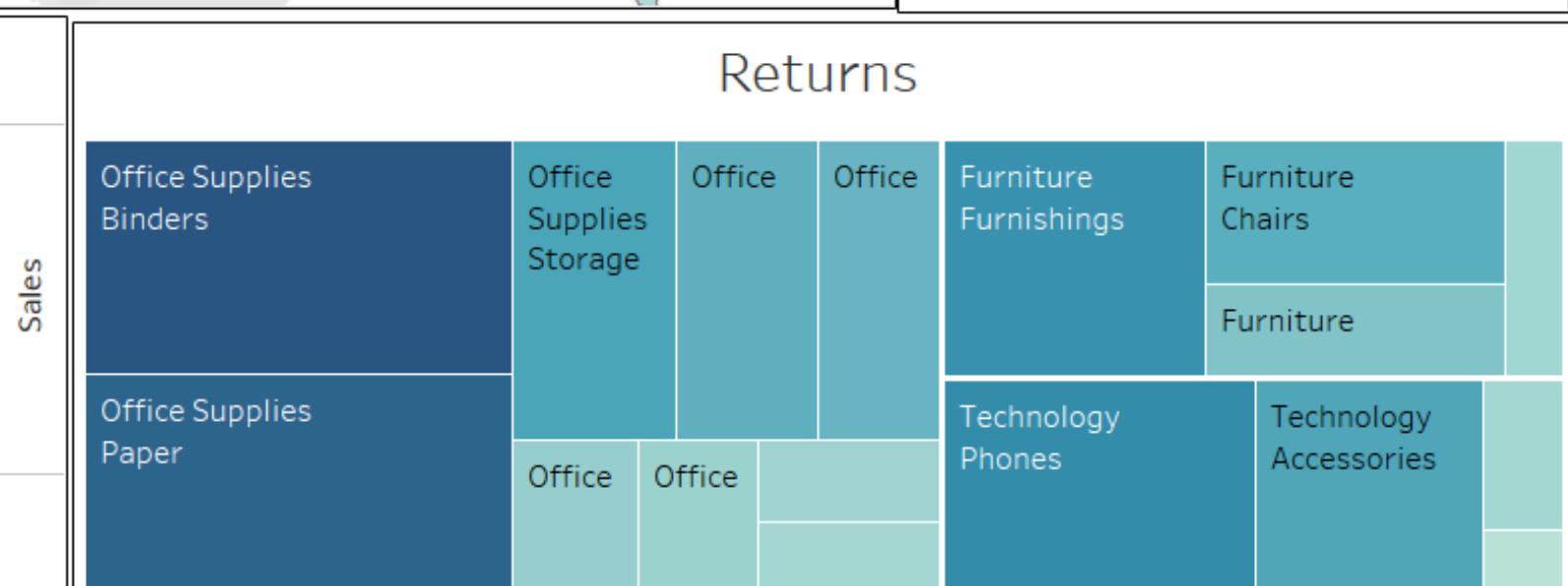
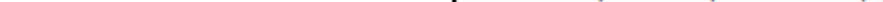
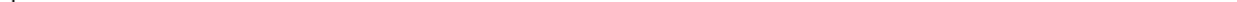
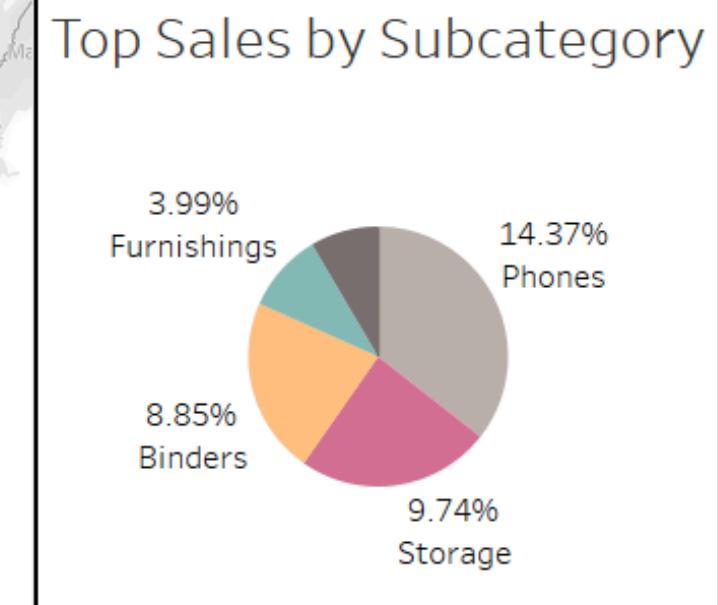
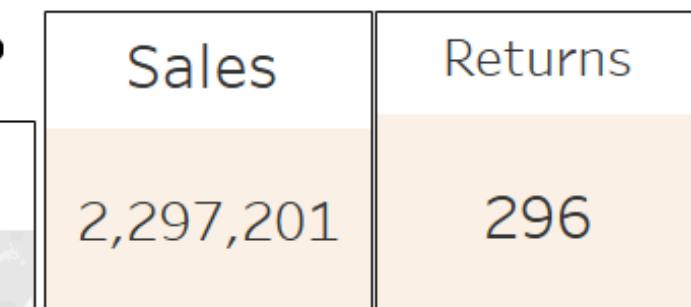
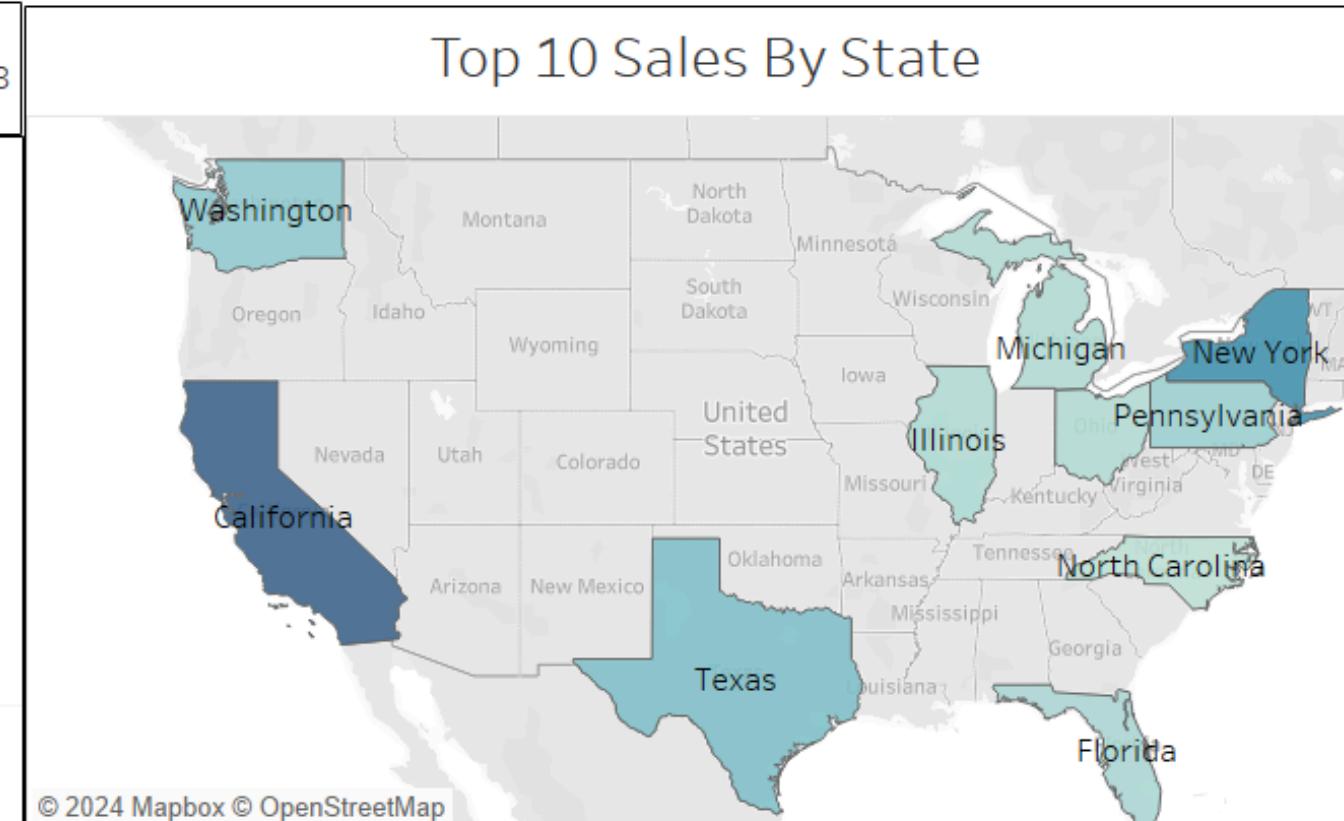
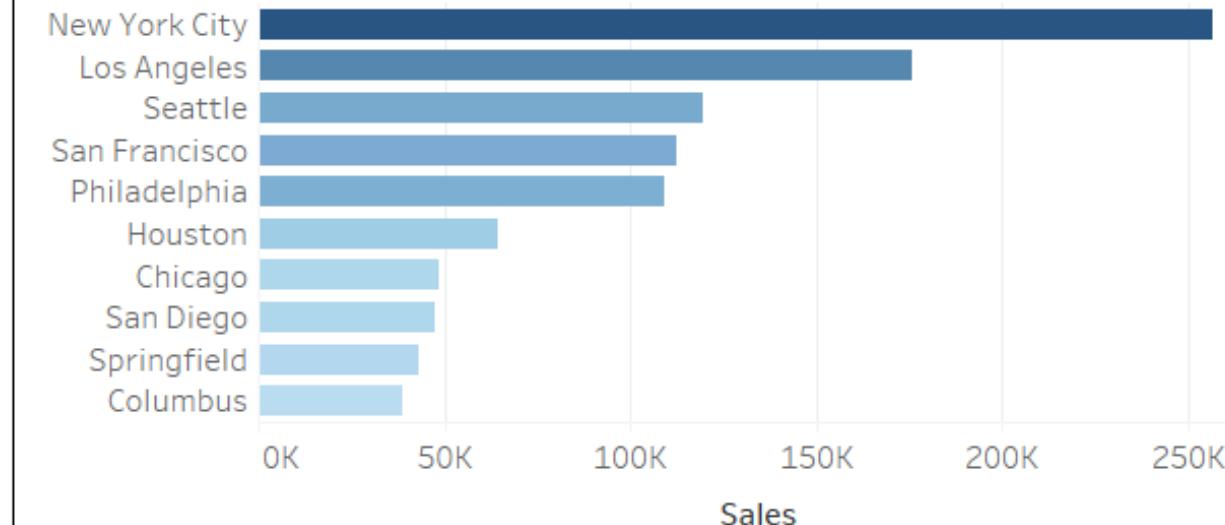
Sales Details

Year of Order Date
2014 2017



Top 10 Sales Cities

City =



Dashboard



1

2

3

4

5

6

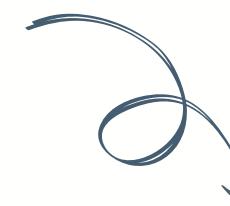
7

8

9

10

Dashboard Link



[https://public.tableau.com/app/profile/amr.elkhamisy/viz/
SalesAnalysisProject_17300263789400/SalesDetails](https://public.tableau.com/app/profile/amr.elkhamisy/viz/SalesAnalysisProject_17300263789400/SalesDetails)

THANK

YOU!

ABOUT
OUR TEAM



- Amr Awad
- Mohamamed Ali
- Bassel Yousef

- Maryam Eslam
- Menna Ateya

1

2

3

4

5

6

7

8

9

10