



"YAT 200"

Superstore Sales Analysis

DEPI Final Project

28 September, 2024

Dr/Soha Nagy



Contents:

Introduction	3
Tools and Technologies	4
Data	5
Data cleaning	9
Visualization	10
Recommendations	23
Conclusion	25



INTRODUCTION

Understanding regional variations in consumer behavior is critical for businesses seeking to optimize their product offerings and marketing strategies.

we seek to analyze sales data to compare consumer preferences between urban and rural areas and between east and west regions.

we aim to uncover insights that shed light on the factors influencing consumer behavior in different geographic settings.



Tools and Technologies:

"Using PYTHON"

Pandas:

Matplotlib and Seaborn:

cikit-learn (Simple Imputer):



Data Description:



The dataset contains sales transaction records for a retail superstore operating in the United States. It covers a period from 2015 to 2018 and includes customer demographics, product details, and shipping information. This rich dataset helps analyze consumer preferences, trends, and sales patterns across different regions in the US.



Key Features:

This is a small abbreviation for the most important columns we use:

- **Sales:**

The sales revenue generated from each transaction (Numerical). This feature is essential for assessing the financial performance of different products, regions, and customer segments, and is a primary metric for the analysis.

- **Category:**

: The main category of the product sold (Categorical), such as "Furniture," "Office Supplies," or "Technology." This feature is crucial for analyzing sales performance across different product categories.



- **Order Date:**

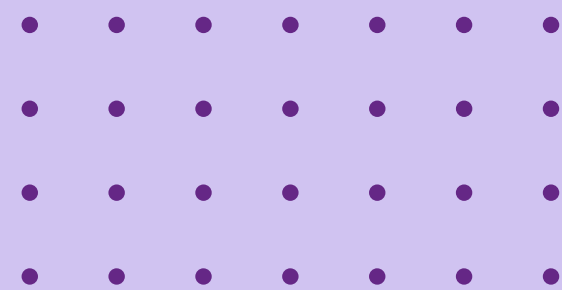
The date on which the order was placed (Datetime). This feature enables time-series analysis, helping to identify trends such as seasonality and changes in consumer behavior over time, from 2015 to 2018.

- **Region:**

The region of the United States in which the transaction took place (Categorical), such as "East," "West," "South," and "Central." This feature allows for regional analysis of sales trends and consumer behavior across different parts of the country

- **City:**

The city where the customer is located (Categorical). This feature plays a central role in comparing consumer behavior between urban and rural areas, allowing for location-based segmentation of sales data.



Data Challenges:



Postal Code:

The column exhibited missing values, duplicate records . posing the risk of skewing the analysis, which required further attention.

Ship Date and Order Date:

The Ship Date and Order Date columns displayed inconsistent formatting, complicating temporal analysis..



Region:

Represented broad categories or more granular data, necessitating an examination of the unique values within each column.



Data Cleaning:

To prepare the dataset for analysis, a series of data cleaning steps were implemented. The dataset was initially loaded, with the index set to the Row ID for easier manipulation. A thorough examination identified columns with missing values, leading to calculations of the percentage of missing data for each feature. Notably, the Postal Code column contained incomplete entries, specifically in rows corresponding to Vermont. To address these null values, we utilized the Simple Imputer from Scikit-learn with the "constant" strategy, filling them with the value 05401. Since Postal Codes begin with "0" and do not require numerical operations, the values were cast to strings to ensure proper formatting. The dataset was further evaluated for unique values to differentiate between categorical and granular data. Duplicate entries were identified and removed to maintain data integrity. Additionally, the Order Date and Ship Date columns were standardized due to inconsistent formats, converting them to a unified datetime format. The cleaned dataset was then saved for further analysis, ensuring accuracy and reliability.

Visualization and Driven-Insights:

The total sales disparity shows that urban areas significantly outperform rural areas in terms of total sales

To diversify the sales base, efforts could be made to increase sales in rural areas, potentially through targeted promotions, improved logistics, or localized product offerings.

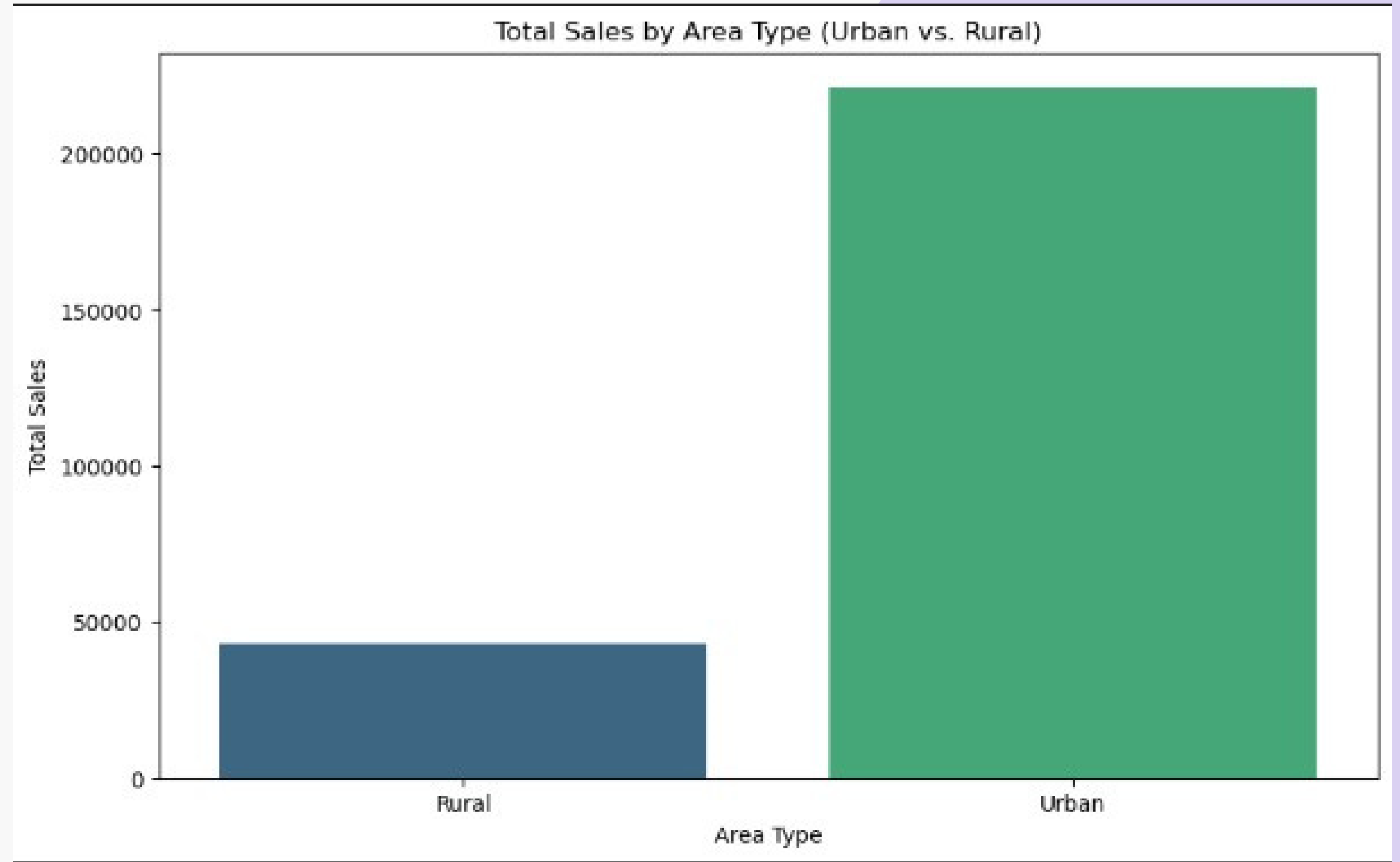


Chart (1)

Chart (2)

Rural areas have a higher average sales per order. In contrast, urban areas have a lower average sales per order. Despite the lower total sales in rural areas, the higher average sales per order present an interesting growth opportunity.

Increasing the average order size in urban regions through bundled promotions or incentives could further boost revenue

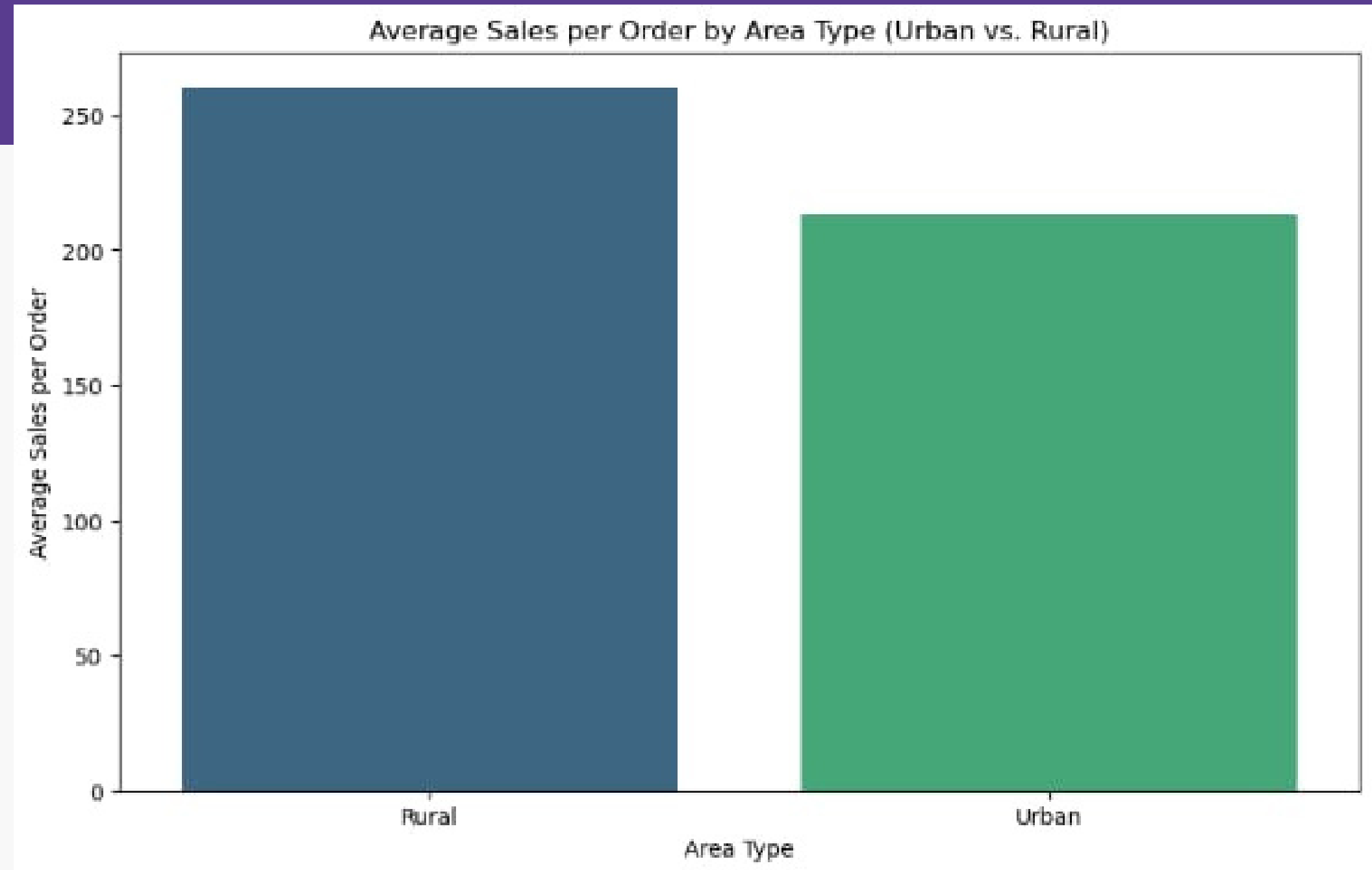


Chart (3)

urban areas exhibit higher sales across all product categories compared to rural areas. Technology products have the largest sales gap between urban and rural regions. While urban areas also lead in furniture and office supplies sales, the differences are less significant.

Businesses can prioritize urban areas for maximum sales but should also consider strategies to tap into the rural market potential. Product differentiation, tailored marketing, and appropriate distribution channels are essential for success in both regions

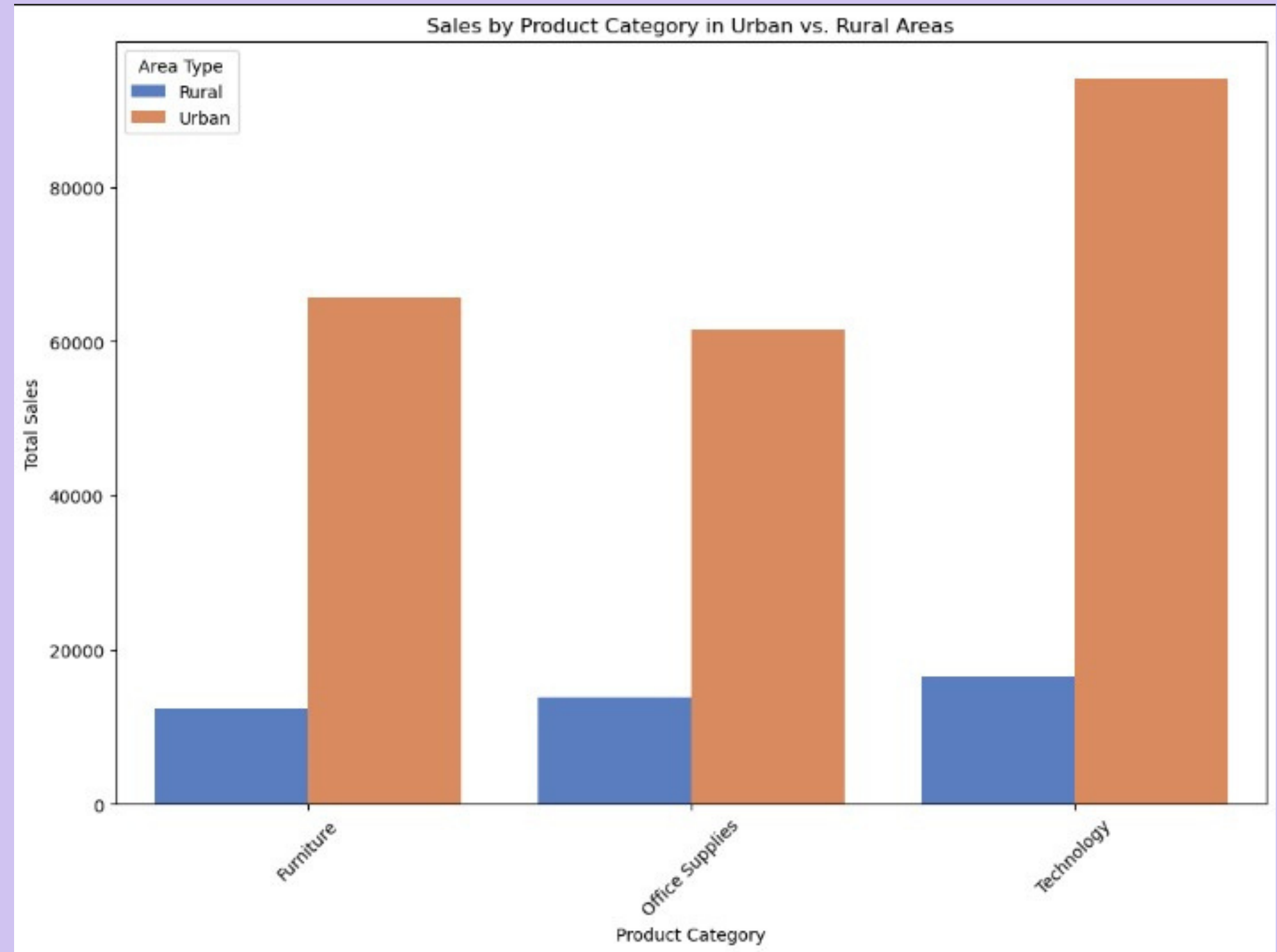


Chart (4)

The West region has the highest total sales, exceeding 700,000, making it the top-performing region in terms of sales.

Lastly, the South region has the lowest sales, falling below 500,000, indicating potential challenges. We need to focus on the South region, exploring the reasons behind its performance and developing strategies to boost sales, which could be beneficial. Additionally, it is essential to sustain performance in the West and East regions, as these areas are performing well; continued engagement offers, or targeted marketing could ensure that they maintain their high sales figures.

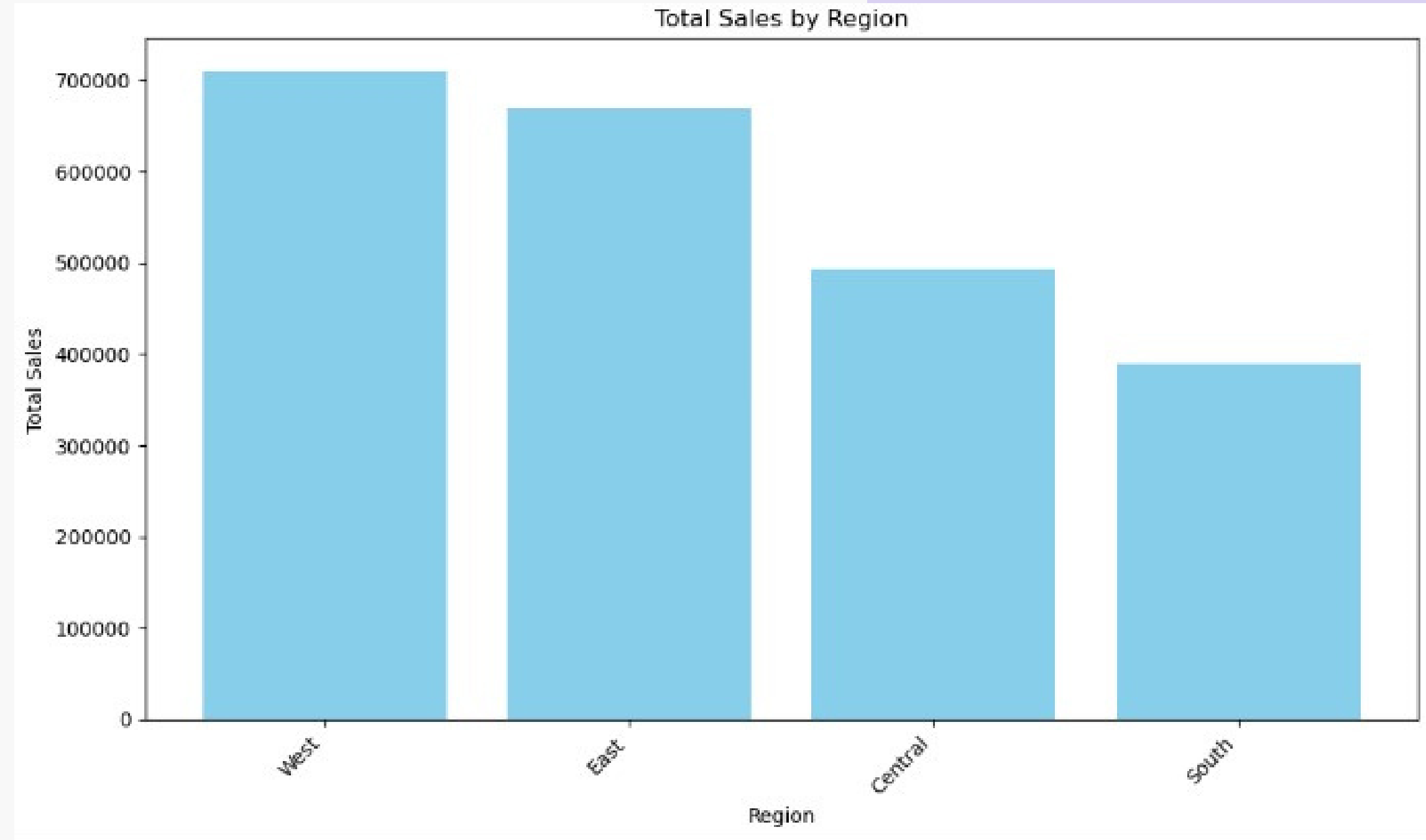


Chart (5)

The heatmap shows that the East region exhibits a strong preference for technology, with a lower interest in office supplies.

The Central region shows relatively balanced demand across the categories, with no strong spikes. Lastly, the South region indicates a higher interest in technology but comparatively lower interest in furniture and office supplies

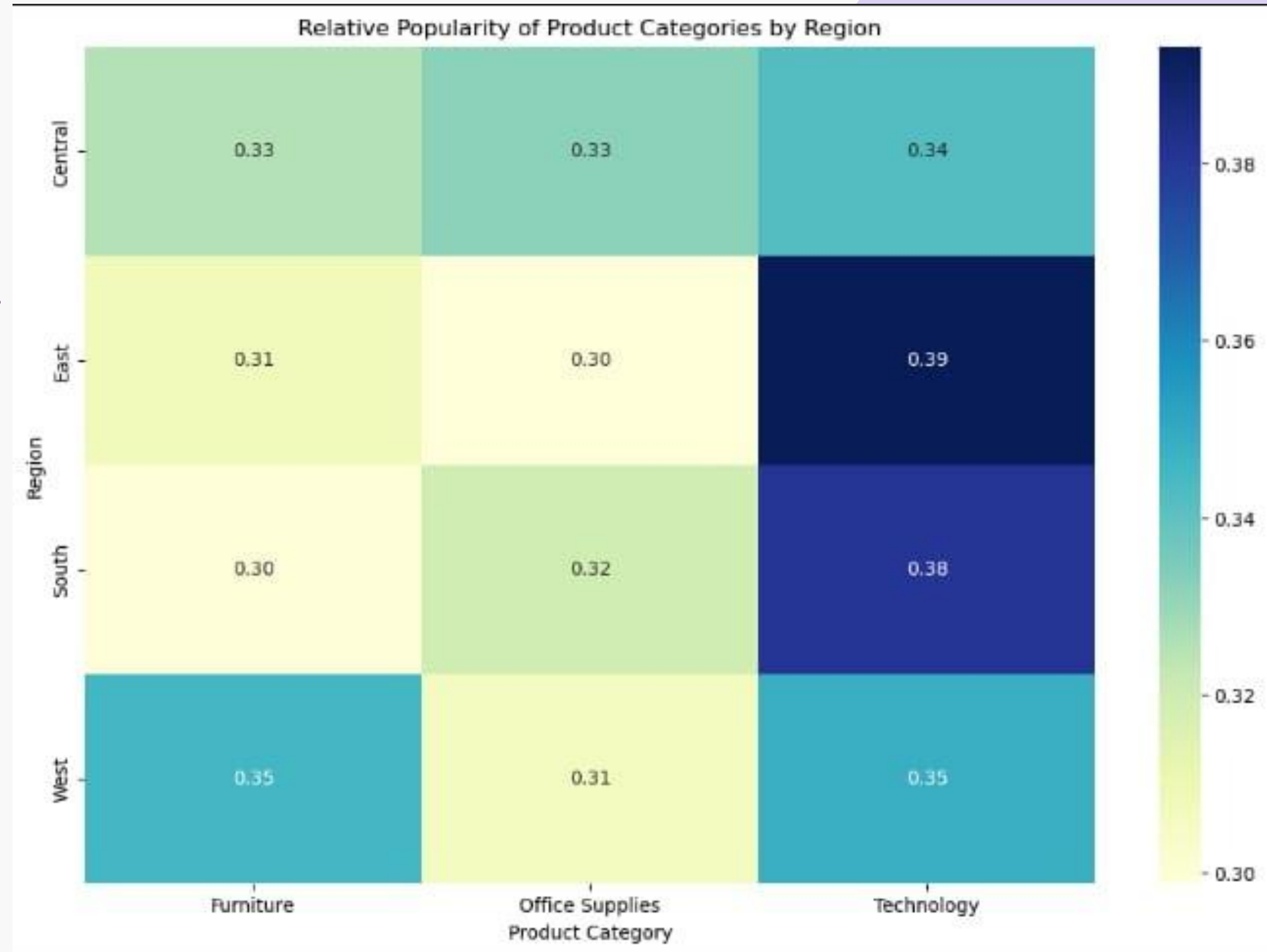


Chart (6)

Sales Over Time vs. Region

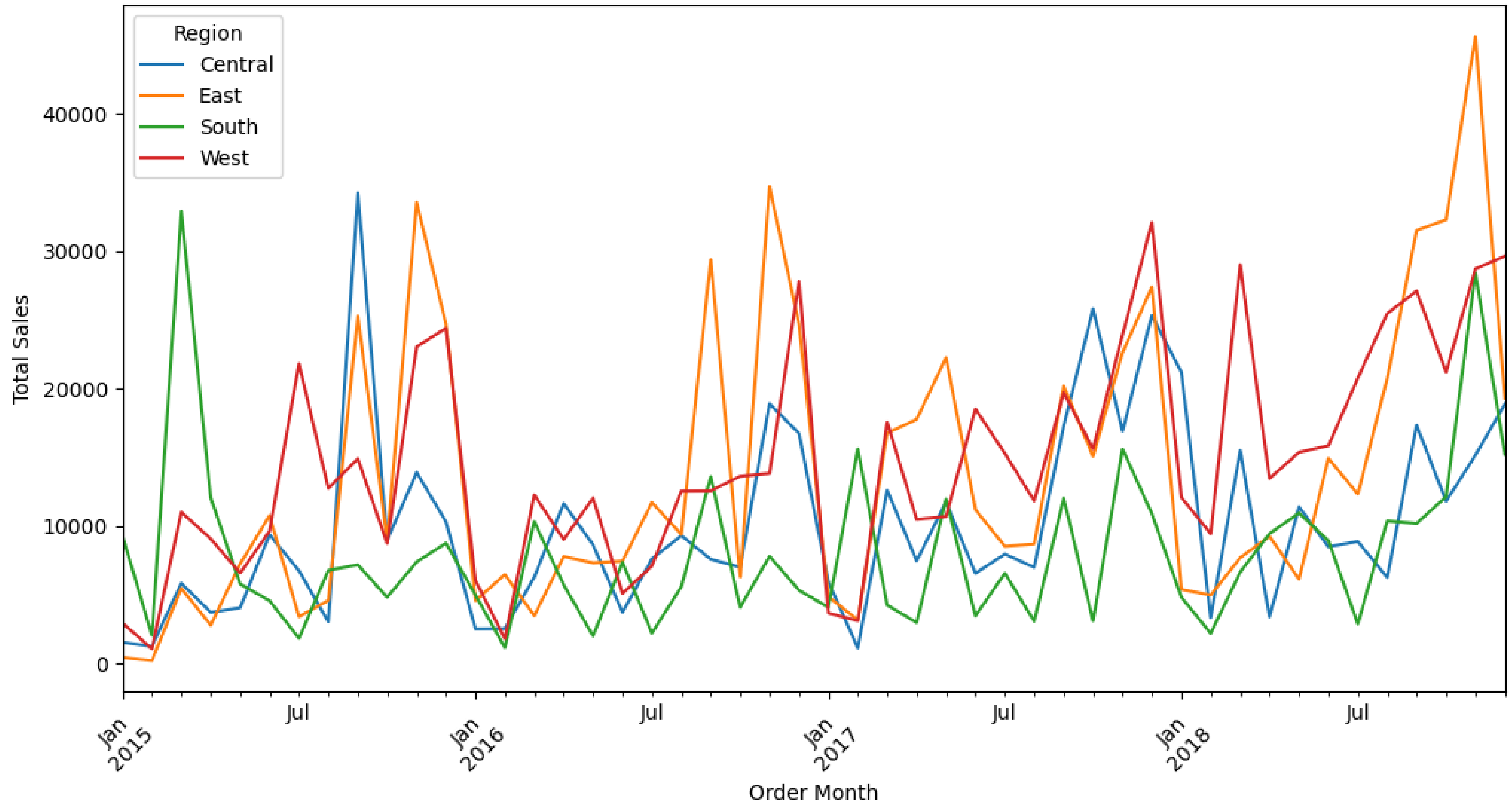
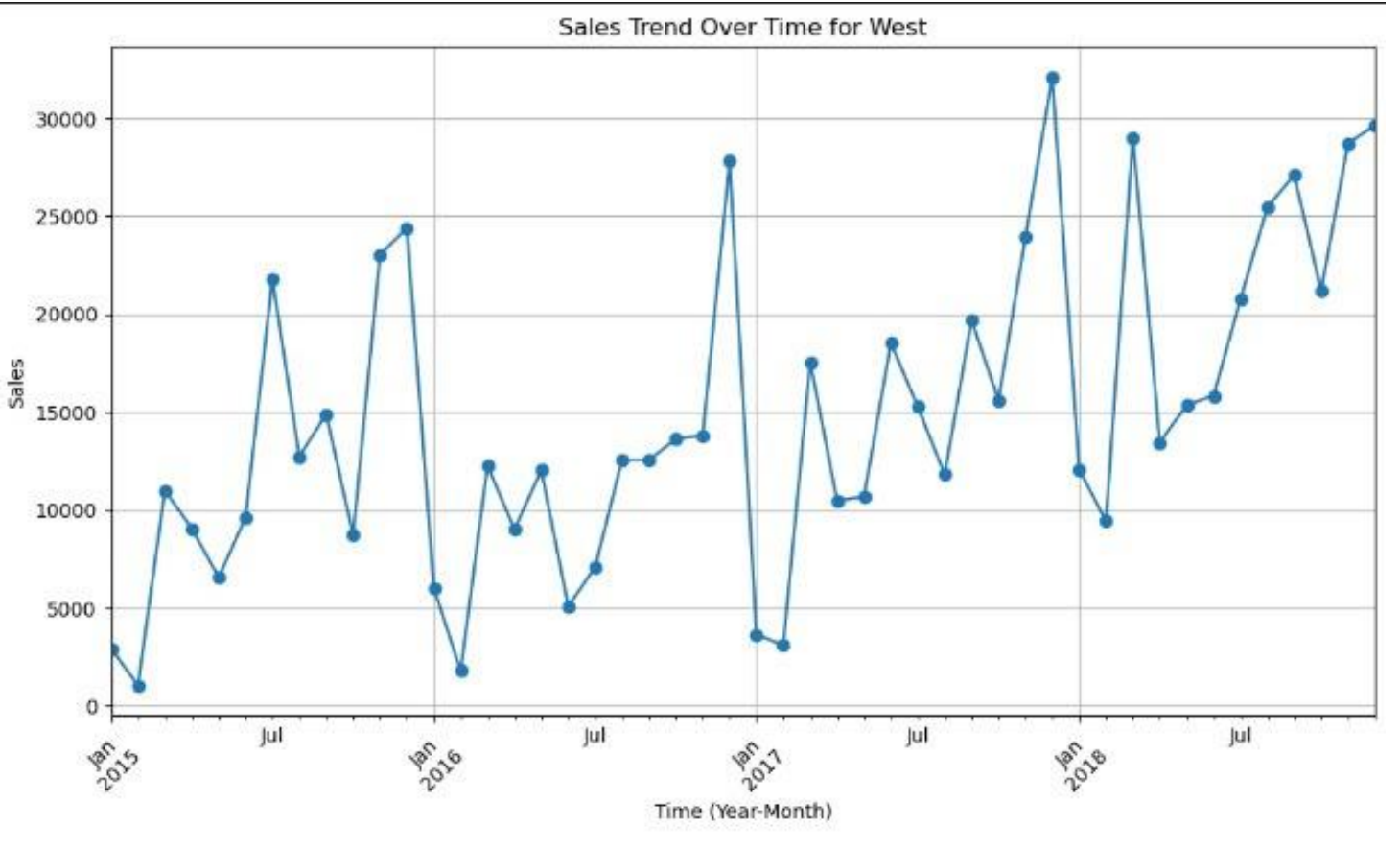
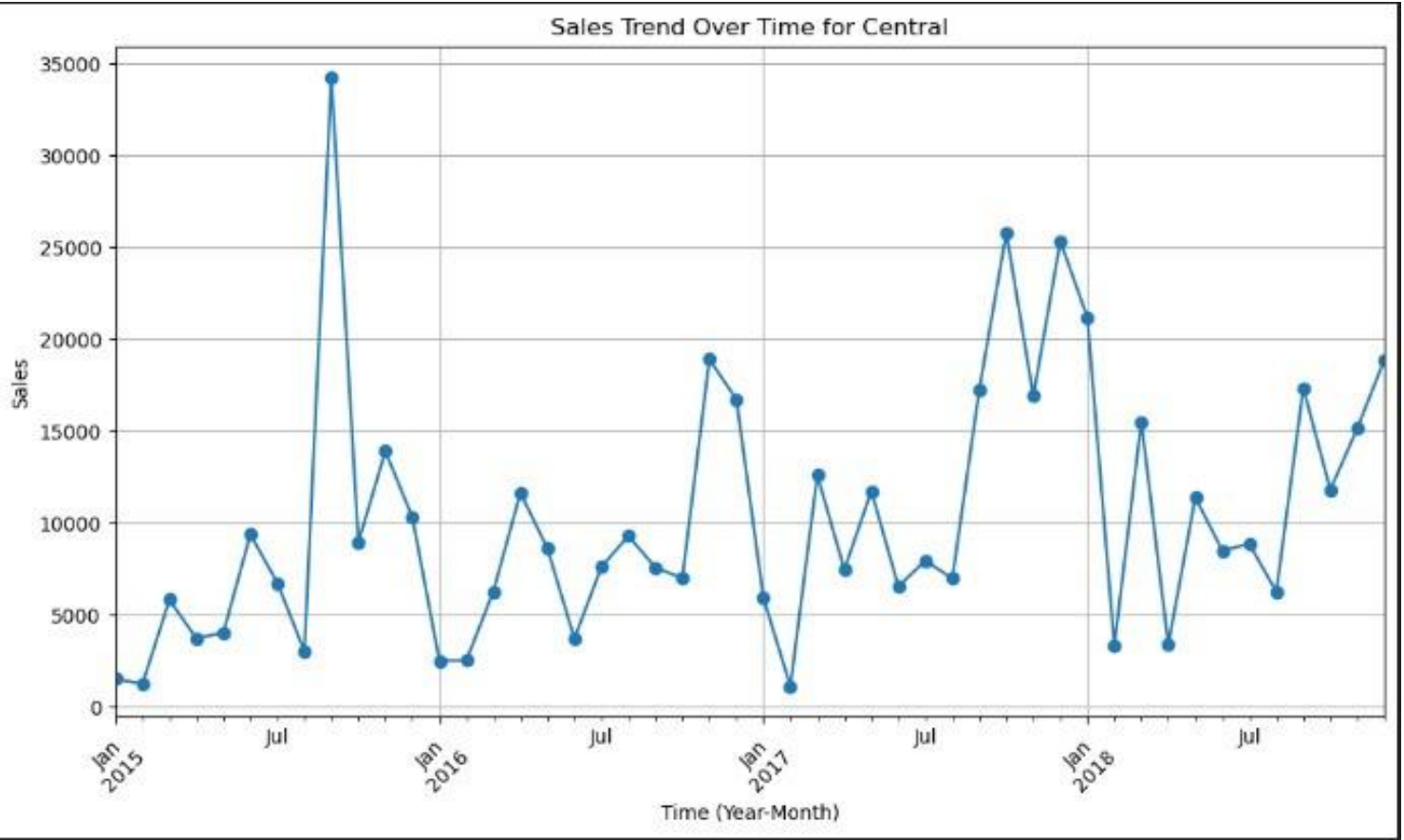
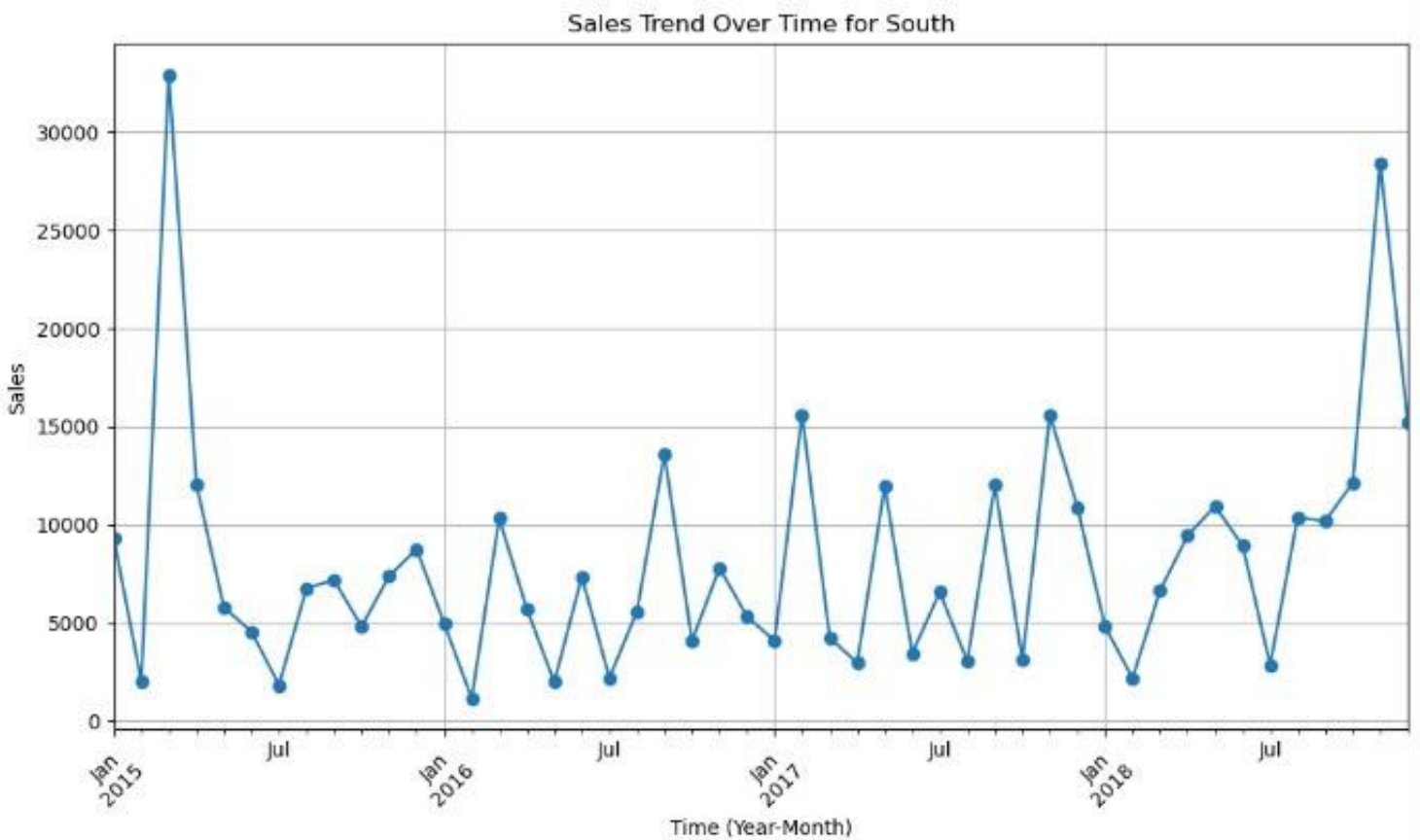
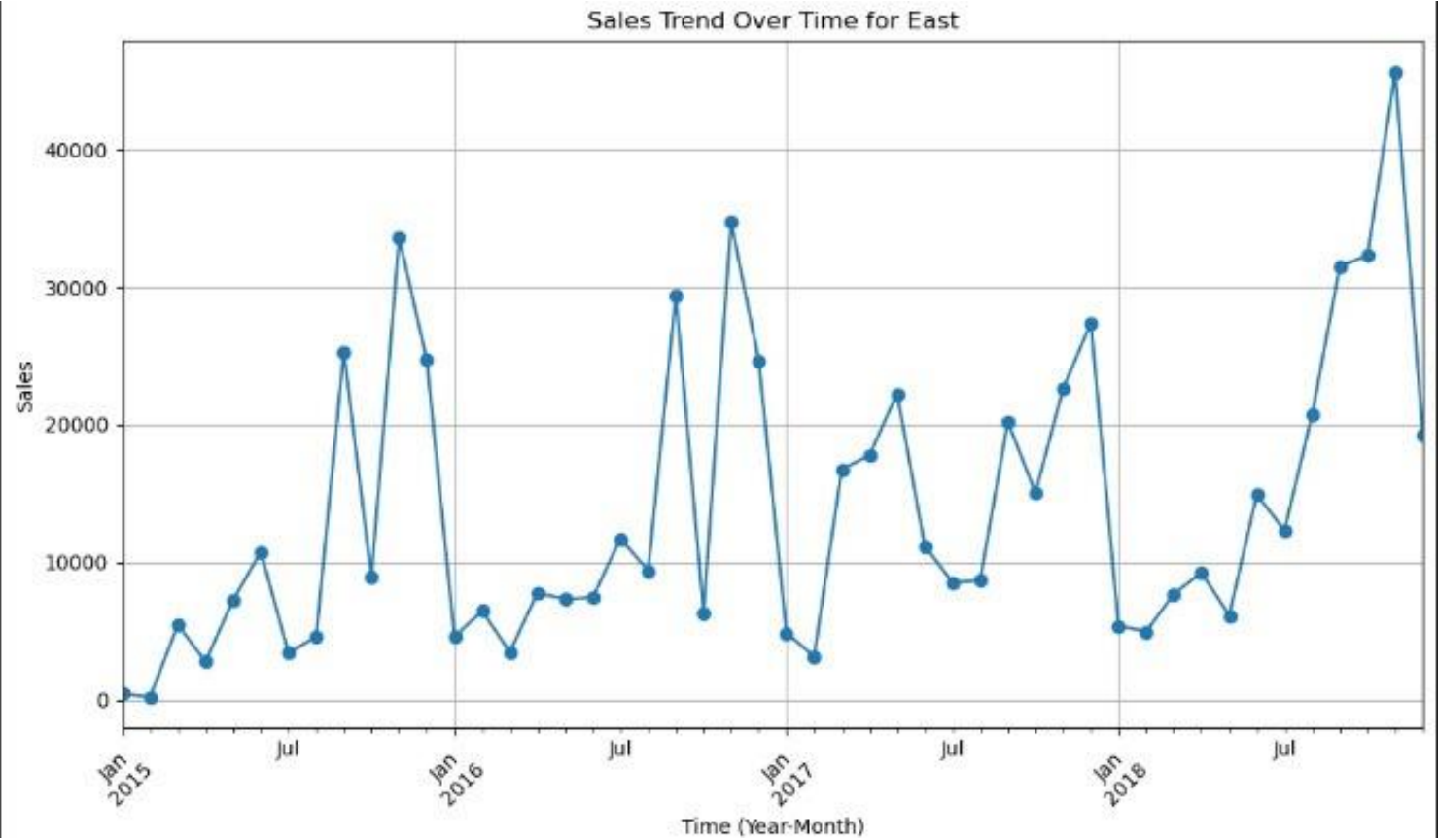
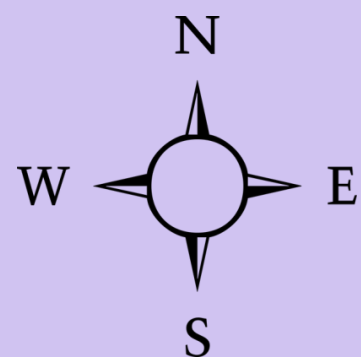


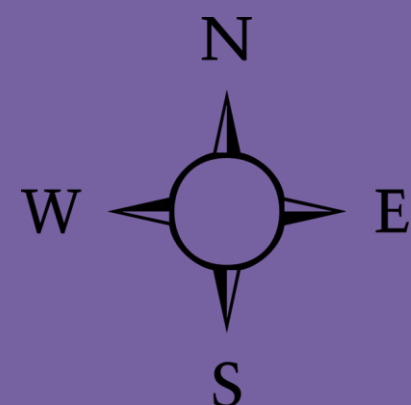
Chart (6)





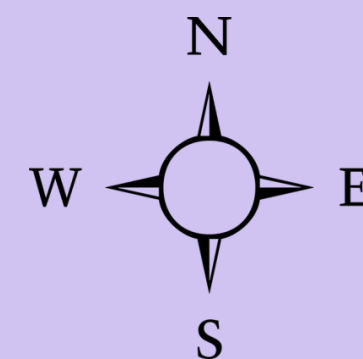
In the East region

sales have shown fluctuations but demonstrate an overall upward trend, with a possible hint of seasonality. The highest point in sales was observed in November 2018.



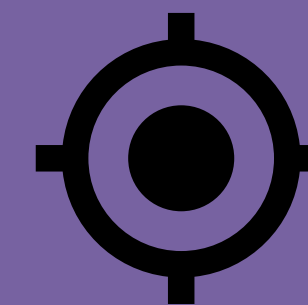
In the South region

displays highly volatile sales with significant peaks and troughs. While there is a slight upward trend, the fluctuations are much more pronounced than in the East, with the peak occurring in March 2015.



In the West region

sales have experienced significant fluctuations but maintain an upward trajectory, with a slightly more pronounced seasonal pattern compared to the East. Like the East, the West region reached its sales peak in December 2017.



In the Central region

upward trend is visible, although the volatility remains high. The peak for this region also occurred in September 2015.

Chart (6)

When comparing the regions, the South and Central regions show the highest levels of volatility, with dramatic swings in sales. In contrast, the East and West regions exhibit less extreme fluctuations, suggesting more stability. Seasonal patterns, while present in all regions, appear more pronounced in the East and West. Overall, all regions display a positive upward trend in sales, indicating growth over the analyzed period. However, the rate of growth and the stability of sales trends differ, with the East and West regions showing more consistent growth, while the South and Central regions experience more volatility.

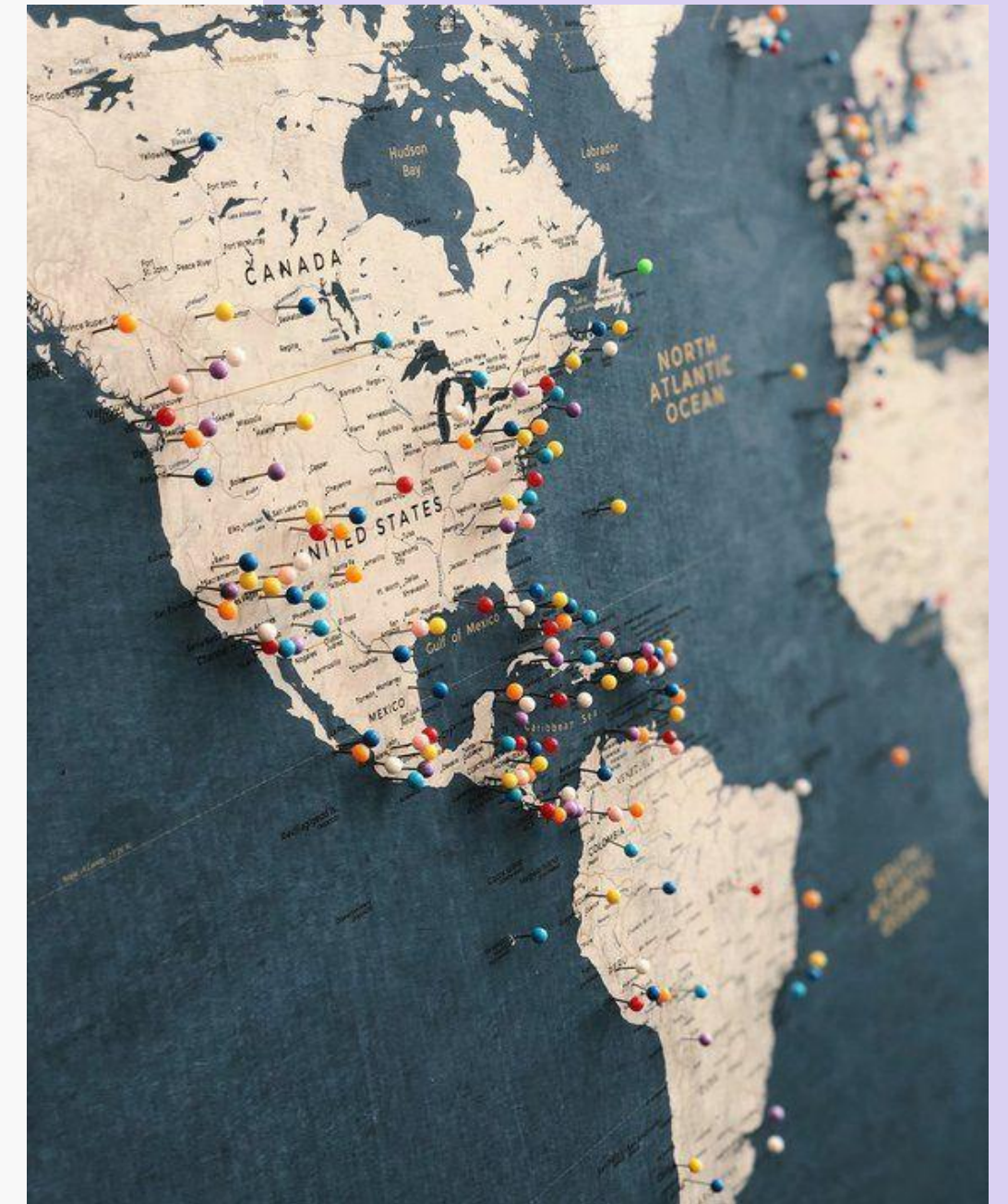


Chart (7)

Each category shows fluctuations, with peaks around the year-end, suggesting seasonality in consumer demand.

Furniture sales are more stable but spike occasionally, likely due to bulk purchases by businesses. Office Supplies follow a consistent upward trend, reflecting steady demand.

Technology, however, shows sharp increases, particularly toward 2018, indicating growing investment in tech products.

These insights suggest that the company should prepare for seasonal sales surges, especially in technology, and align inventory and marketing strategies accordingly.

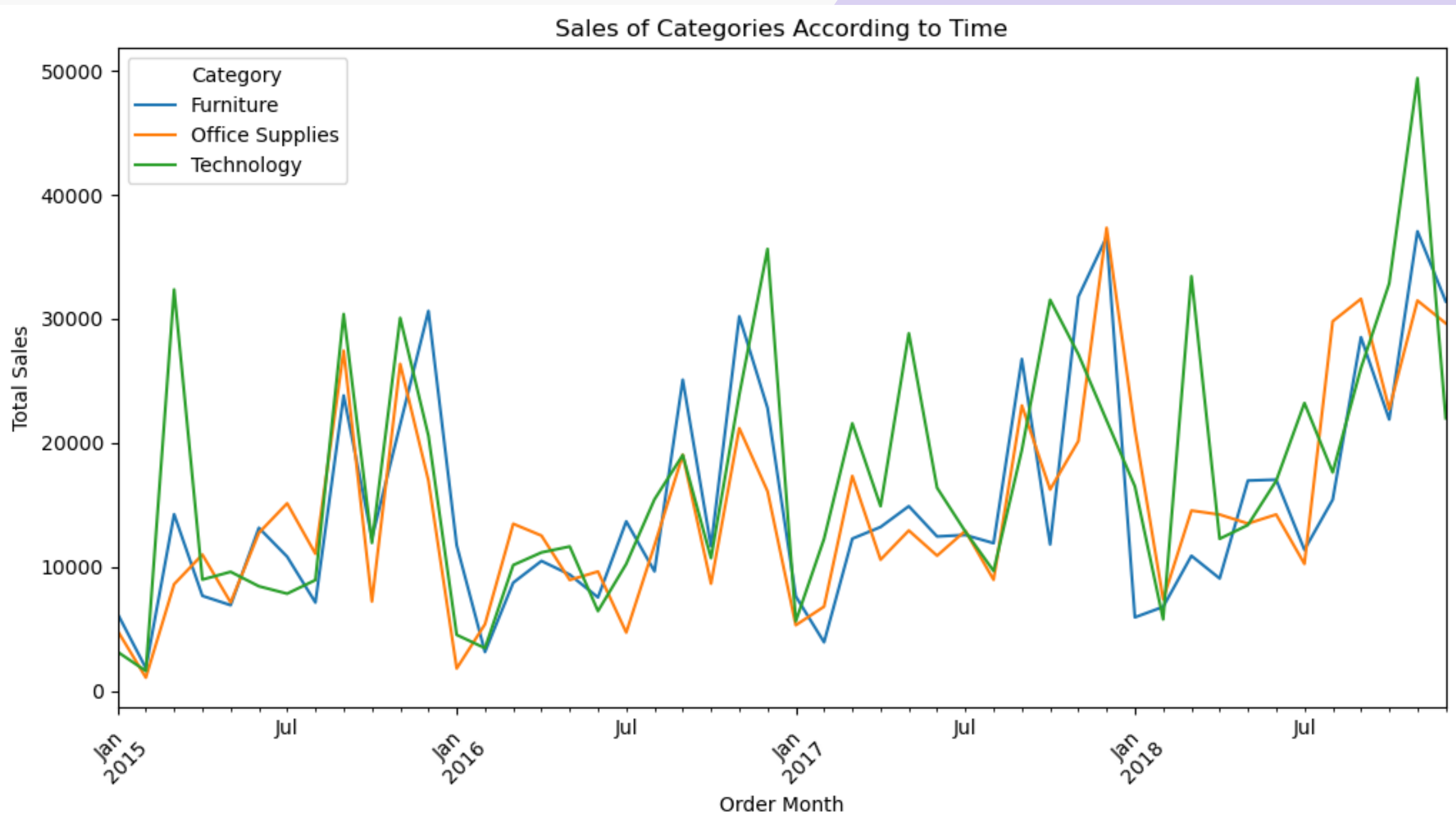


Chart (7)

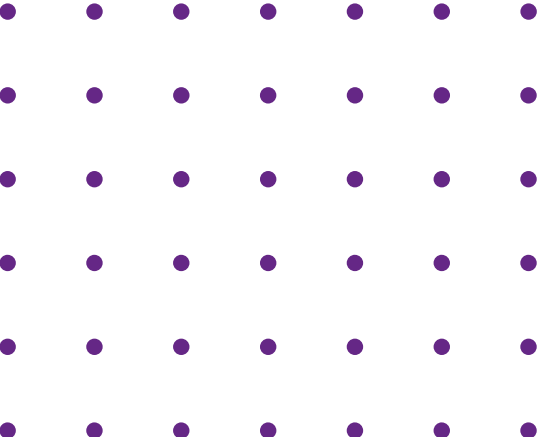
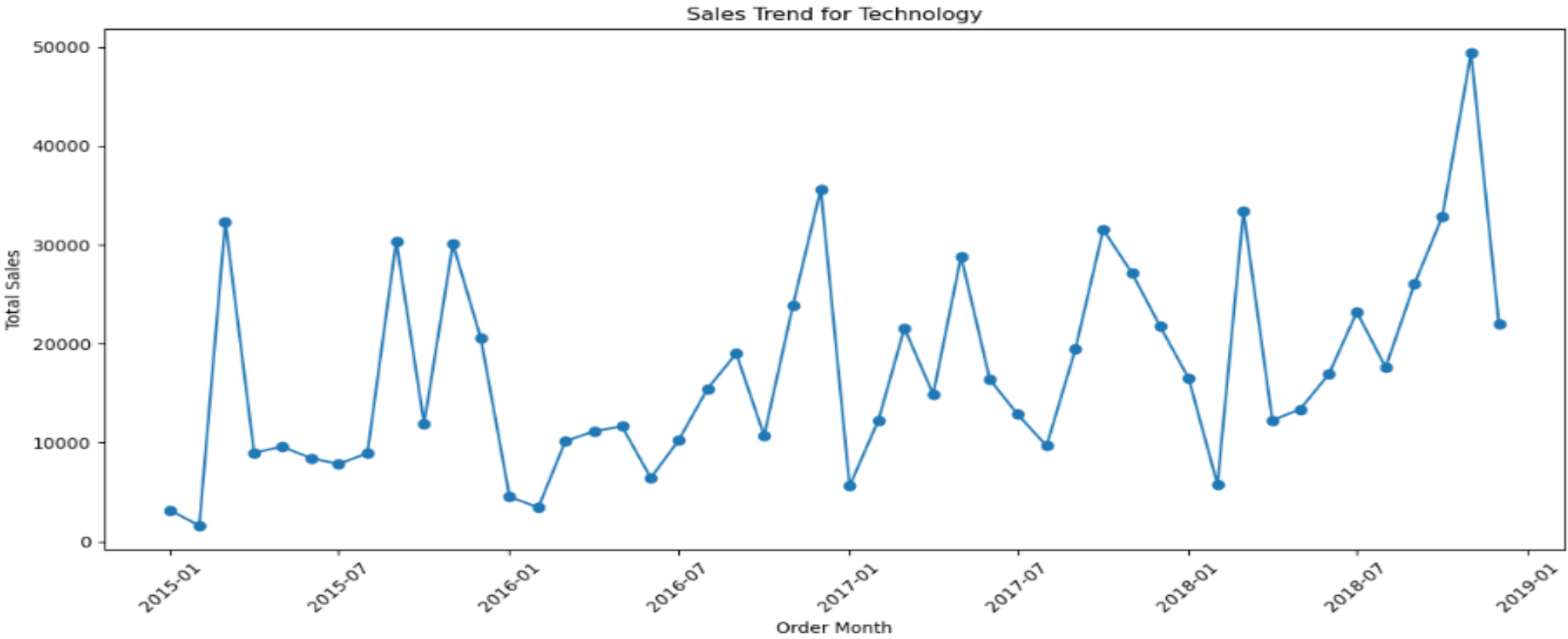
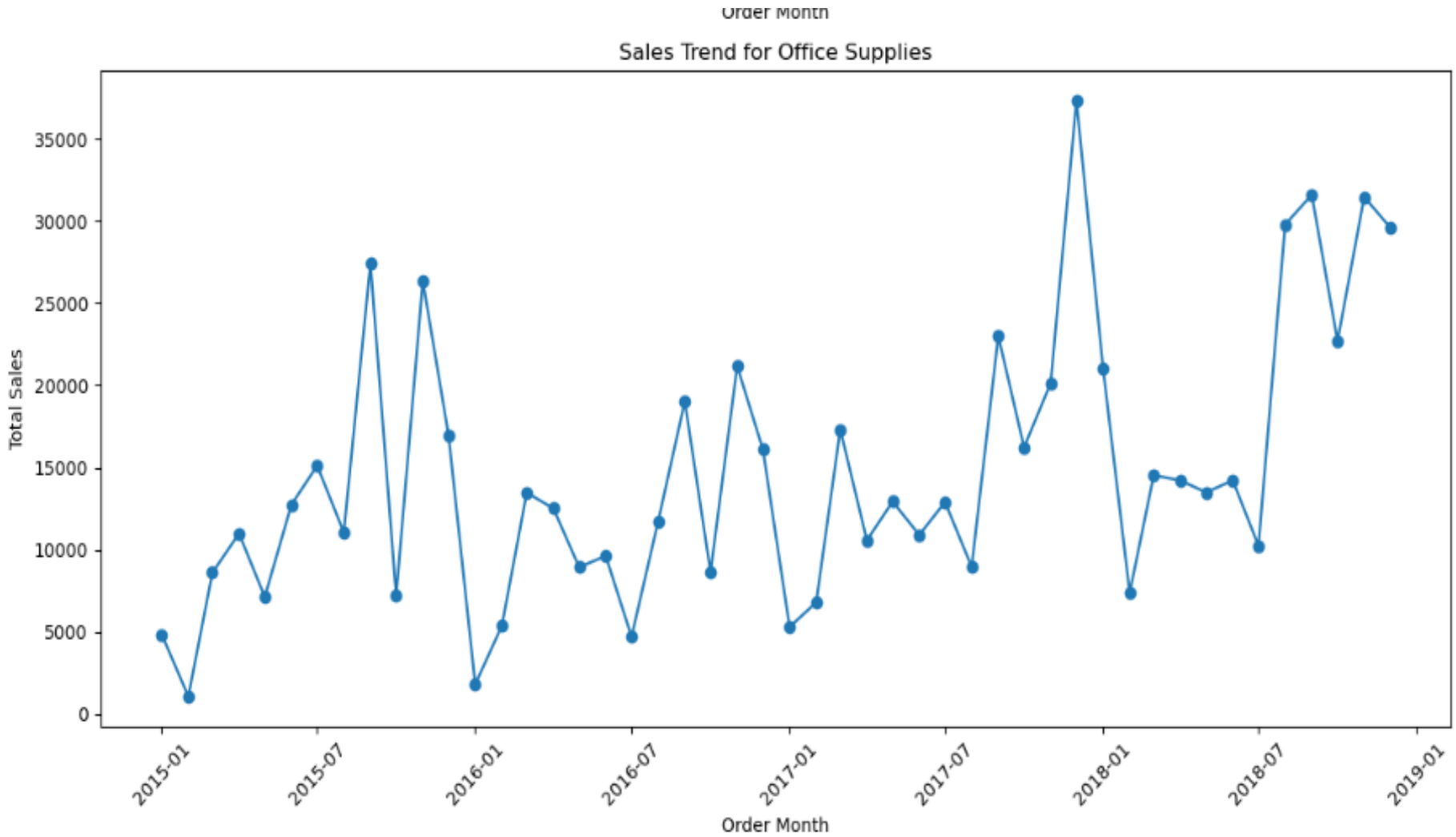
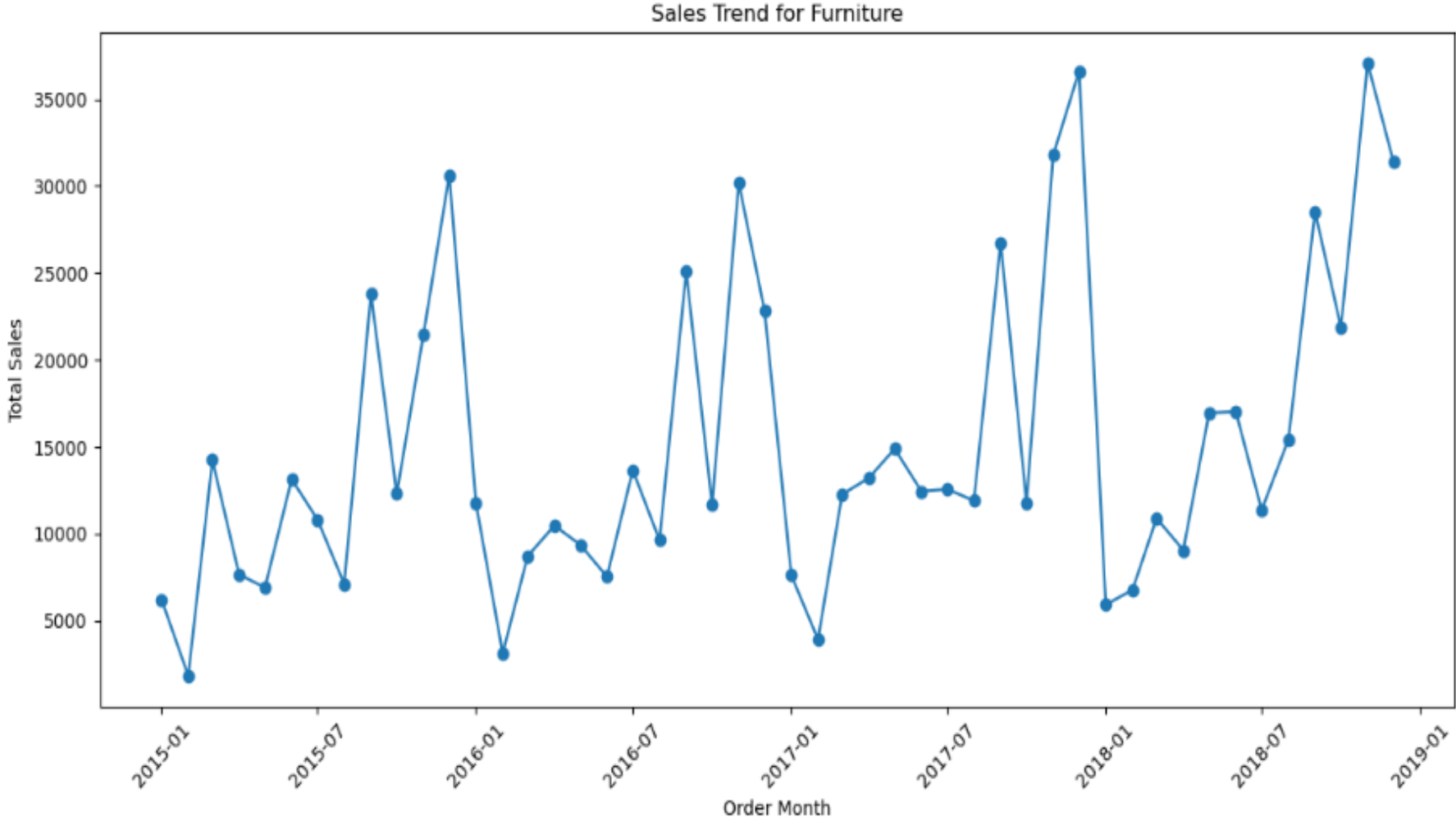
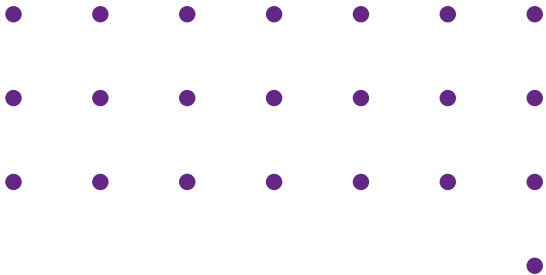


Chart (8)

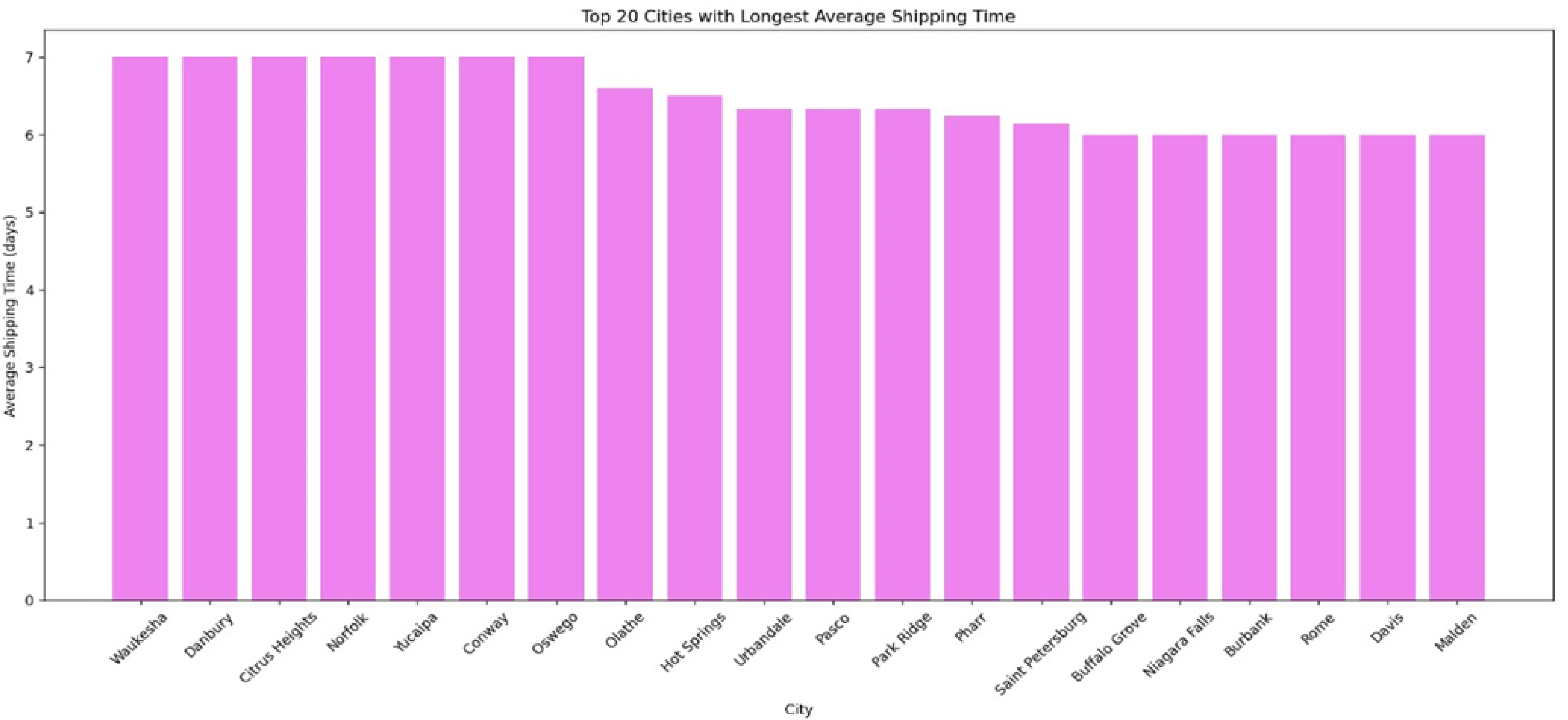


chart (8) shows the top 20 cities with the longest average shipping times, with values ranging from 6 to 7 days. Most of these cities appear to be rural or suburban, indicating that less urbanized areas face longer shipping delays, likely due to limited infrastructure and distance from major distribution hubs. Urban cities, while present, show slightly faster shipping times. Improving logistics in rural areas could reduce these delays.

Recommendations

- **Urban Areas:**

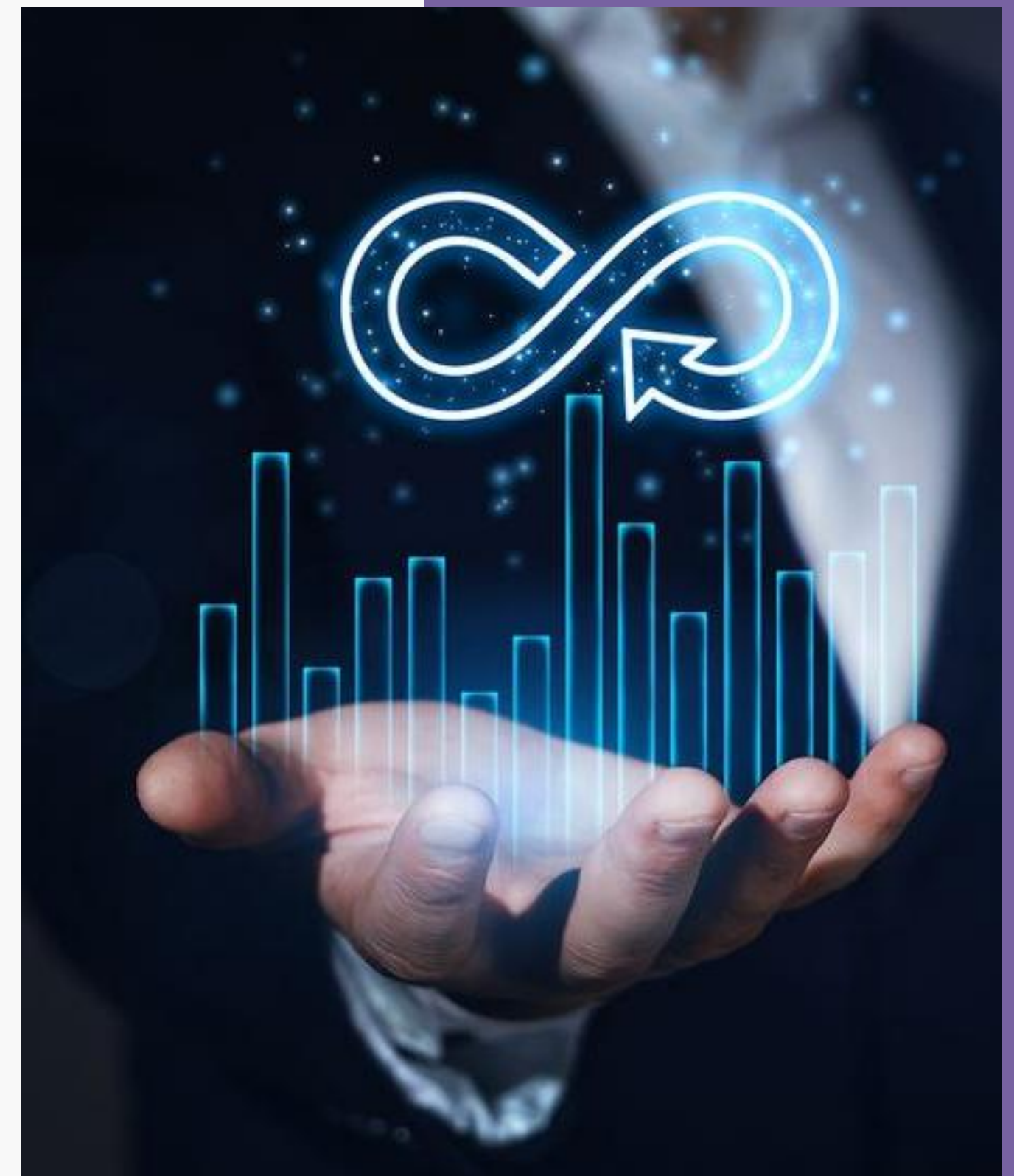
Urban areas lead in tech sales. To boost revenue, expand tech offerings and run targeted marketing during peak seasons. Use promotions and loyalty programs to encourage larger purchases.

- **Rural Markets:**

Rural areas contribute less to total sales but have higher average sales per order. To leverage this, expand practical, high-value products and improve logistics for faster shipping. Targeted rural promotions can boost purchase frequency.

- **Regional Disparities:**

The West region leads in sales, while the South underperforms. To improve the South, investigate the causes and launch targeted promotions. Maintain strong sales in the West and East with customer engagement and seasonal deals. In the Central region, stabilize demand with consistent marketing and strategic products.



- **Sales Volatility Management:**

To manage sales volatility, use consistent off-peak promotions in regions with sharp fluctuations. Implement seasonal strategies to promote steady year-round growth, especially in high-demand areas like technology.

- **Product Differentiation:**

Urban consumers lead in tech sales, so the business should prioritize expanding tech offerings in urban areas while introducing them in rural regions. Tailoring products to meet the distinct needs of each region ensures alignment with local demand. Improving logistics in rural areas, where longer shipping times are common, will enhance service quality and unlock sales growth potential.

By following these recommendations, the business can diversify its sales base, drive growth in underperforming regions, and create a sustainable and balanced approach to revenue generation across all geographic areas. This strategy will not only strengthen performance in high-performing regions but also unlock new opportunities in areas with untapped potential

Conclusion



The analysis of consumer preferences and sales patterns across the United States reveals significant regional disparities. Urban areas, particularly in technology, lead in total sales, while rural regions demonstrate higher average sales per order. The West region outperforms others, while the South faces challenges. To improve performance, businesses should focus on product differentiation, optimized distribution, and managing sales volatility. Targeted promotions and strategic product offerings in both urban and rural areas, along with addressing regional challenges, can foster balanced growth. By emphasizing regional dynamics, logistical improvements, and marketing adjustments, businesses can capitalize on opportunities in underperforming areas while maintaining momentum in stronger markets.



Team Members



Nada
Leader



Roqia
Member



Mona
Member



Marwa H.
Member

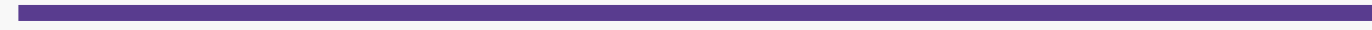


Dalia
Member



Marwa O.
Member





Thank you

