INT375(Data Science Toolbox: Python Programming)

PROJECT REPORT : *Seasonality Sales*

Submitted by- Subham Dhiman
Registration No- 12309083
Section -K23SG

Under the Guidance of: Dr. Manpreet Singh Sehgal

School of Computer Science & Engineering
Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that Subham Dhiman bearing Registration no. 12309083 has completed INT375(Data Science Toolbox: Python Programming) project titled- Seasonality Sales under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

DECLARATION

I, Subham Dhiman, student of B.Tech. under CSE Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date – 12/04/2025
Registration no. -12309083

Signature
Name: Subham Dhiman

Acknowledgement

I would like to extend my sincere appreciation to anyone who assisted and provided useful guidance in this process of the project. I wish to express my appreciation to the data providers and organizations that provided the dataset, which allowed for an investigation into the allocation of public funding, across regions, organizations, and demographic characteristics. I also acknowledge all the developers and contributors of open-source software, such as Pandas, Matplotlib, and Seaborn, which provided outstanding functionality to assist my data cleaning, statistical calculations, and visualizations. Finally, I thank the on-going encouragement of peers, family, and friends, who have provided feedback and encouragement throughout the various stages of the research project

INTRODUCTION

The **Seasonality Sales Dataset** provides valuable insights into consumer purchasing behavior across different dimensions such as month, season, weather conditions, promotions, customer segments, regions, and event types. This dataset is ideal for analyzing how various factors influence retail sales and can help businesses optimize their marketing strategies, promotional planning, and inventory management.

In this analysis, the dataset was explored through five main objectives using Python libraries like **Pandas**, **Matplotlib**, and **Seaborn**. First, seasonal trends in demand were identified by visualizing monthly and weather-based sales, revealing that months like December and January, as well as Winter and Monsoon seasons, show peak sales—likely due to holidays and indoor shopping patterns. Second, the impact of promotions was evaluated, showing that products on promotion generally have higher average sales than non-promoted ones, suggesting the strategic importance of discounting. Third, future demand forecasting was initiated using trend plots of monthly revenue, with a recommendation to apply time series models like ARIMA or Prophet for predictive modeling. Fourth, customer segmentation was analyzed, highlighting that young adults tend to spend more than adults, which can inform targeted marketing campaigns. Lastly, regional and event-based performance was assessed, showing that regions like West and South perform better and that events such as Christmas and New Year sales significantly boost revenue. Throughout the analysis, key data visualization techniques including bar charts, line graphs, histograms, and correlation heatmaps were used to support data-driven insights.

2. Source of Dataset

This project analyzes retail sales patterns using the **Seasonality Demand Dataset**, available on Kaggle. The dataset includes transactional and contextual information such as sales amount, month, weather, promotional status, customer segment, region, and event types. The purpose of this analysis is to uncover seasonal and behavioral trends in consumer demand, evaluate the effectiveness of marketing strategies, and segment customers for better targeting.
Dataset link: https://www.kaggle.com/datasets/rajsumit17/seasonality-demand-dataset

3. Dataset Preprocessing

The dataset was first imported from a CSV file using **Pandas**. Initial steps included examining the structure of the dataset with functions like **.head()** and **.info()**, followed by data cleaning and formatting to ensure consistency. Columns such as **Sales** were converted to numeric types, while the **Month** column was parsed to a consistent format to enable time series analysis. Missing values in key fields (such as **Sales**, **Month**, **Weather**, or **Promotion**) were either imputed or dropped, depending on the context, to maintain data integrity. The cleaned data was then used for analysis and visualization using tools like **Matplotlib**, **Seaborn**, and **NumPy**.

4. Analysis on Dataset

1) General Description

This project focuses on analyzing **retail sales data** to uncover trends influenced by seasonal demand, promotional activities, regional preferences, and customer demographics. Using Python, the dataset was explored to identify high-performing months, weather conditions contributing to increased sales, the impact of marketing efforts, and consumer behavior segmented by age and region.

The analysis includes detailed visualizations such as line charts, bar plots, histograms, and correlation heatmaps, which highlight key insights like peak sales months (e.g., December, January), the positive effect of promotions, and regional or event-based performance (e.g., during New Year or Christmas sales).

The main goal is to generate actionable insights to support **data-driven marketing and inventory planning**, helping businesses make informed decisions to maximize profit and customer engagement.

2) Specific Requirements

To perform this project effectively, the following tools and libraries were used:

- **Software & IDE**:
  - Python 3.x
  - Jupyter Notebook (for interactive development and step-wise visualization)
- **Python Libraries**:
  - **pandas** – for data cleaning, preprocessing, and manipulation
  - **matplotlib** – for basic plots and customization
  - **seaborn** – for advanced visualizations and statistical graphics
  - **numpy** – for numerical computations and array handling
  - **datetime** – for handling date-time formats during monthly analysis

3) Analysis Results & Visualization

The analysis produced the following major findings, supported by visual plots and diagrams:

Seasonal Trends:

Line plots showed peak sales in December and January, indicating holiday shopping surges. Weather-wise, Winter and Monsoon seasons drove higher demand.

Promotional Impact:

Bar plots revealed that products with promotions consistently recorded higher average sales than non-promoted items.

Forecasting Readiness:

Time-series graphs displayed regular patterns, indicating the dataset's suitability for future sales forecasting using models like ARIMA or Prophet.

Customer Segmentation:

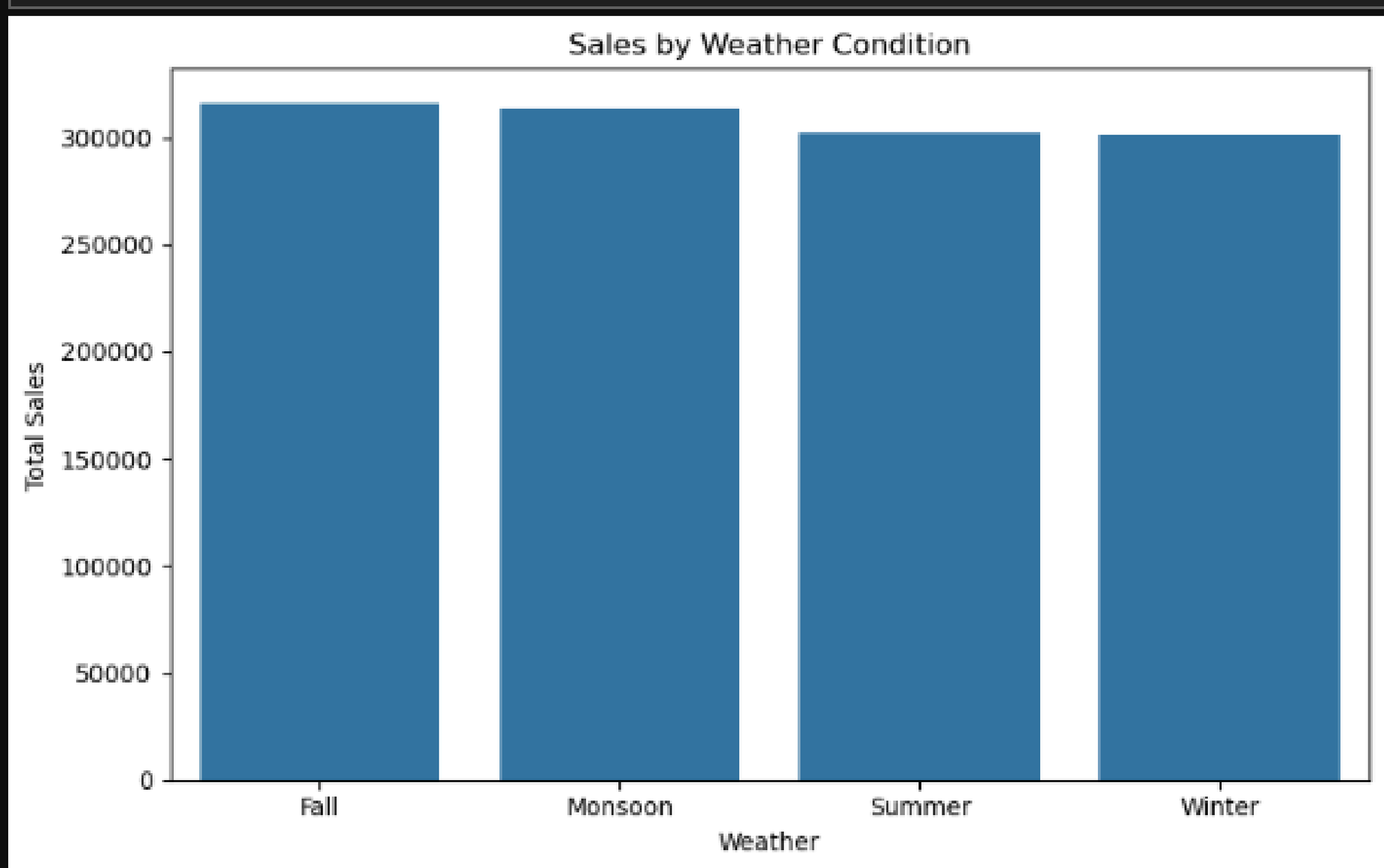
Histograms and bar charts showed that Young Adults were the most active consumer segment, especially during promotional events.

Regional & Event-Based Analysis:

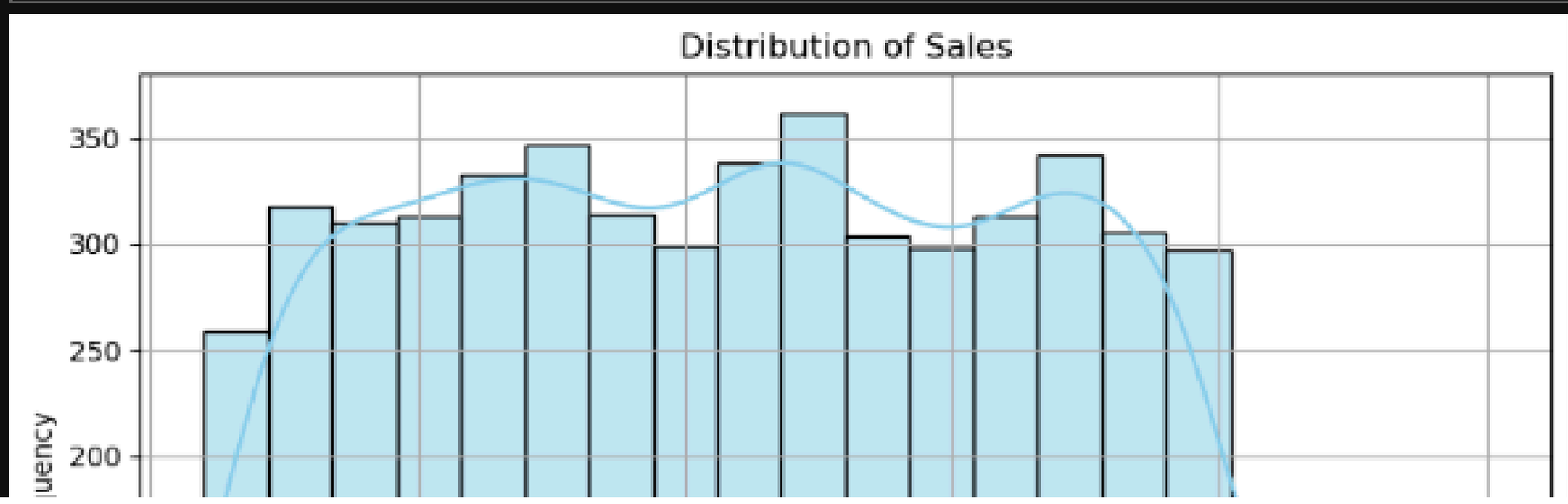
Visualizations highlighted that regions like West and South contributed significantly to total sales, with event-based promotions (e.g., New Year Sale) leading to notable performance spikes.

Objective 1: Identify Seasonal Trends in Demand Monthly Sales Trend shows peak demand in months like December and January, hinting at holiday/festival season demand.
Weather-based Sales reveal that Winter and Monsoon see the highest sales, possibly due to indoor or seasonal shopping patterns.

[196]:
```
plt.figure(figsize=(8, 5))
colors = sns.color_palette("Set2", len(weather_sales))
sns.barplot(data=weather_sales, x='Weather', y='Sales')
plt.title("Sales by Weather Condition")
plt.xlabel("Weather")
plt.ylabel("Total Sales")
plt.tight_layout()
plt.show()
```



[198]:
```
plt.figure(figsize=(8, 5))
sns.histplot(df['Sales'], bins=20, kde=True, color='skyblue')
plt.title("Distribution of Sales")
plt.xlabel("Sales")
plt.ylabel("Frequency")
plt.grid(True)
plt.tight_layout()
plt.show()
```



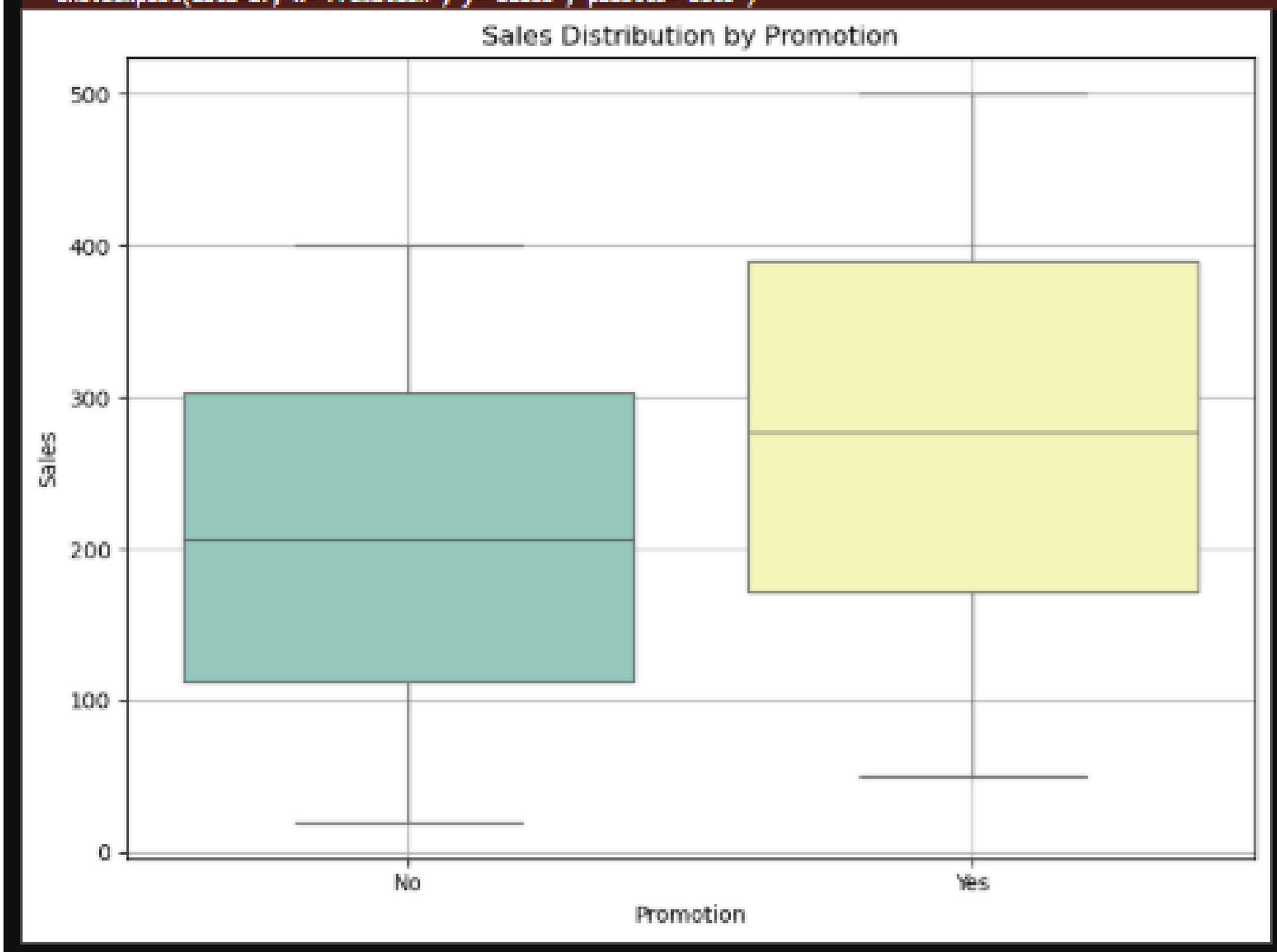
Objective 2: Evaluate the Impact of Promotions and Advertising Products with "Yes" in Promotion show higher average sales than non-promoted items.
This indicates promotions are effective in boosting sales and should be used strategically.

```
plt.figure(figsize=(8, 6))
sns.boxplot(data=df, x='Promotion', y='Sales', palette='Set3')
plt.title("Sales Distribution by Promotion")
plt.xlabel("Promotion")
plt.ylabel("Sales")
plt.grid(True)
plt.tight_layout()
plt.show()
```

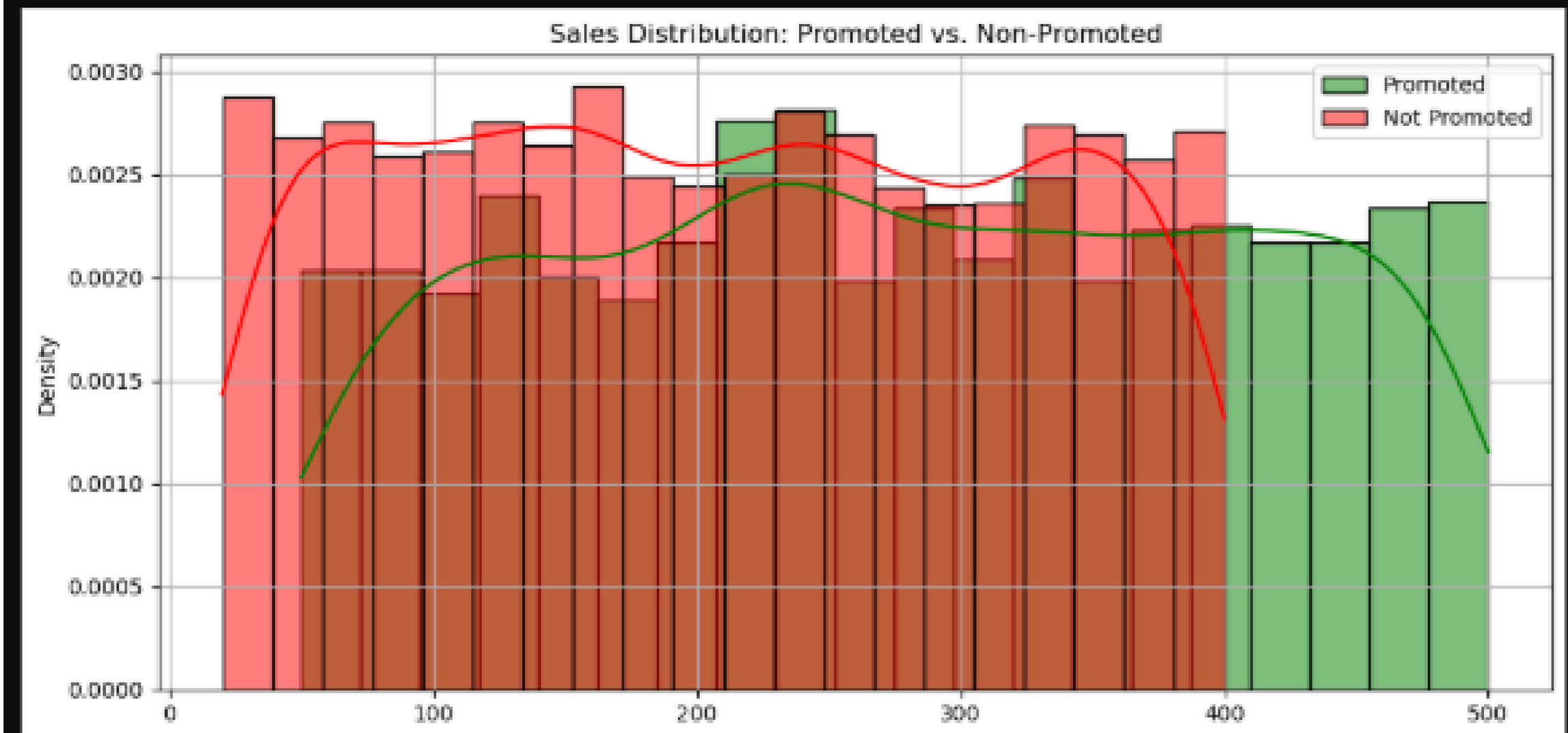
```
C:\Users\dhima\AppData\Local\Temp\ipykernel_23660\2362558684.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.boxplot(data=df, x='Promotion', y='Sales', palette='Set3')
```

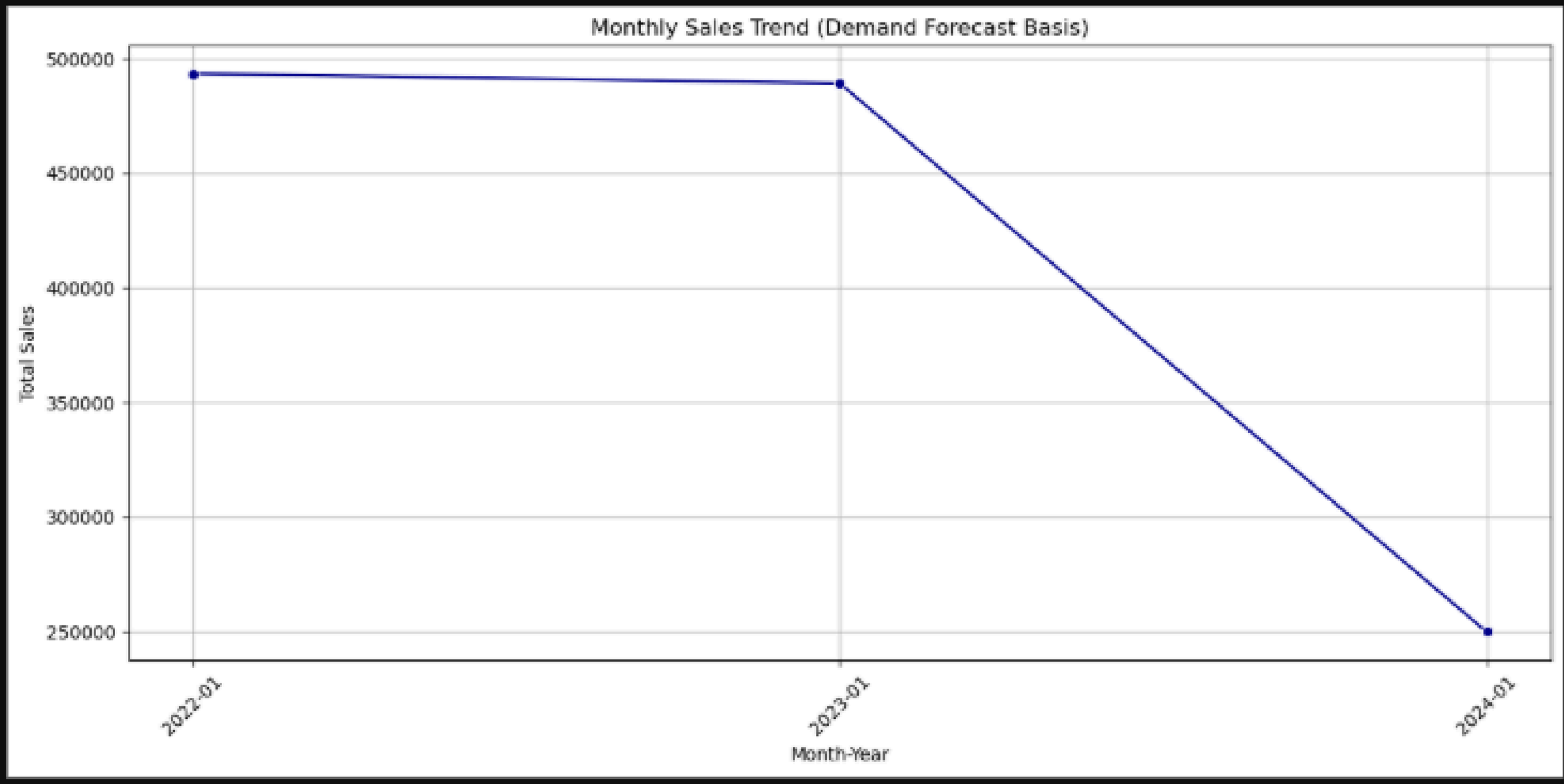



```
plt.figure(figsize=(10, 5))
sns.histplot(data=df[df['Promotion'] == 'Yes'], x='Sales', color='green', label='Promoted', kde=True, stat='density', bins=20)
sns.histplot(data=df[df['Promotion'] == 'No'], x='Sales', color='red', label='Not Promoted', kde=True, stat='density', bins=20)
plt.title("Sales Distribution: Promoted vs. Non-Promoted")
plt.xlabel("Sales")
plt.ylabel("Density")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



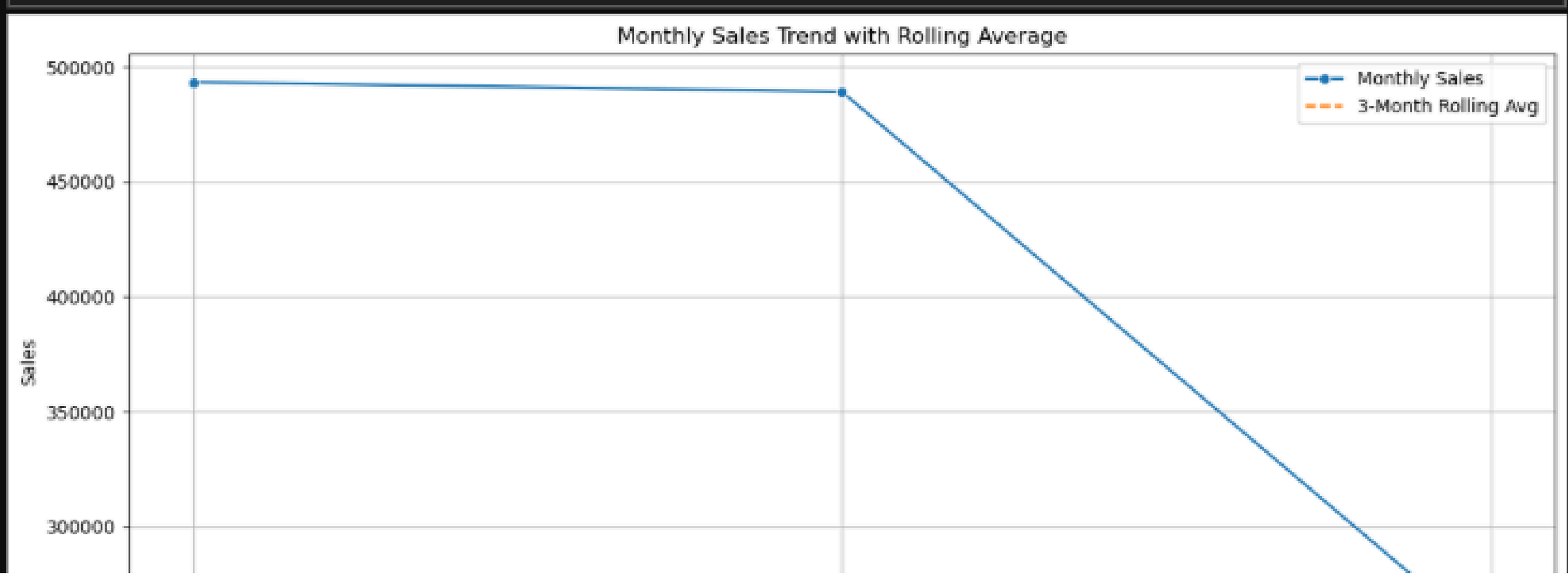
Objective 3: Forecast Future Demand The monthly revenue trend can guide forecasting—especially noting regular spikes around holidays.
Use time series models like ARIMA, Prophet, or Exponential Smoothing for better predictions.

```
plt.xlabel("Month-Year")
plt.ylabel("Total Sales")
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()
plt.show()
```



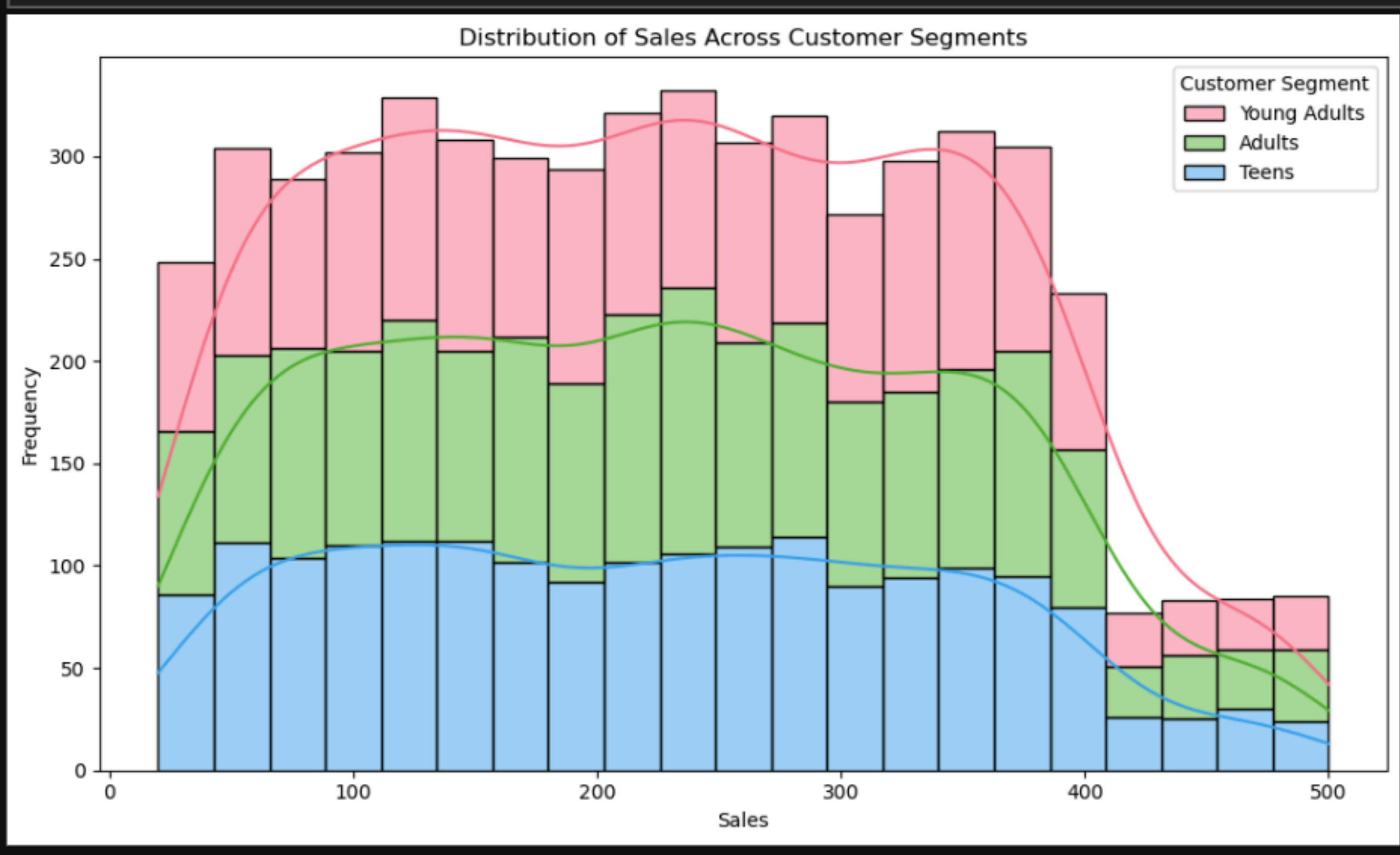

```
247]: monthly_sales['Rolling_Sales'] = monthly_sales['Sales'].rolling(window=3).mean()

plt.figure(figsize=(12, 6))
sns.lineplot(x='Month_Year', y='Sales', data=monthly_sales, label='Monthly Sales', marker='o')
sns.lineplot(x='Month_Year', y='Rolling_Sales', data=monthly_sales, label='3-Month Rolling Avg', linestyle='--')
plt.title("Monthly Sales Trend with Rolling Average")
plt.xlabel("Month-Year")
plt.ylabel("Sales")
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

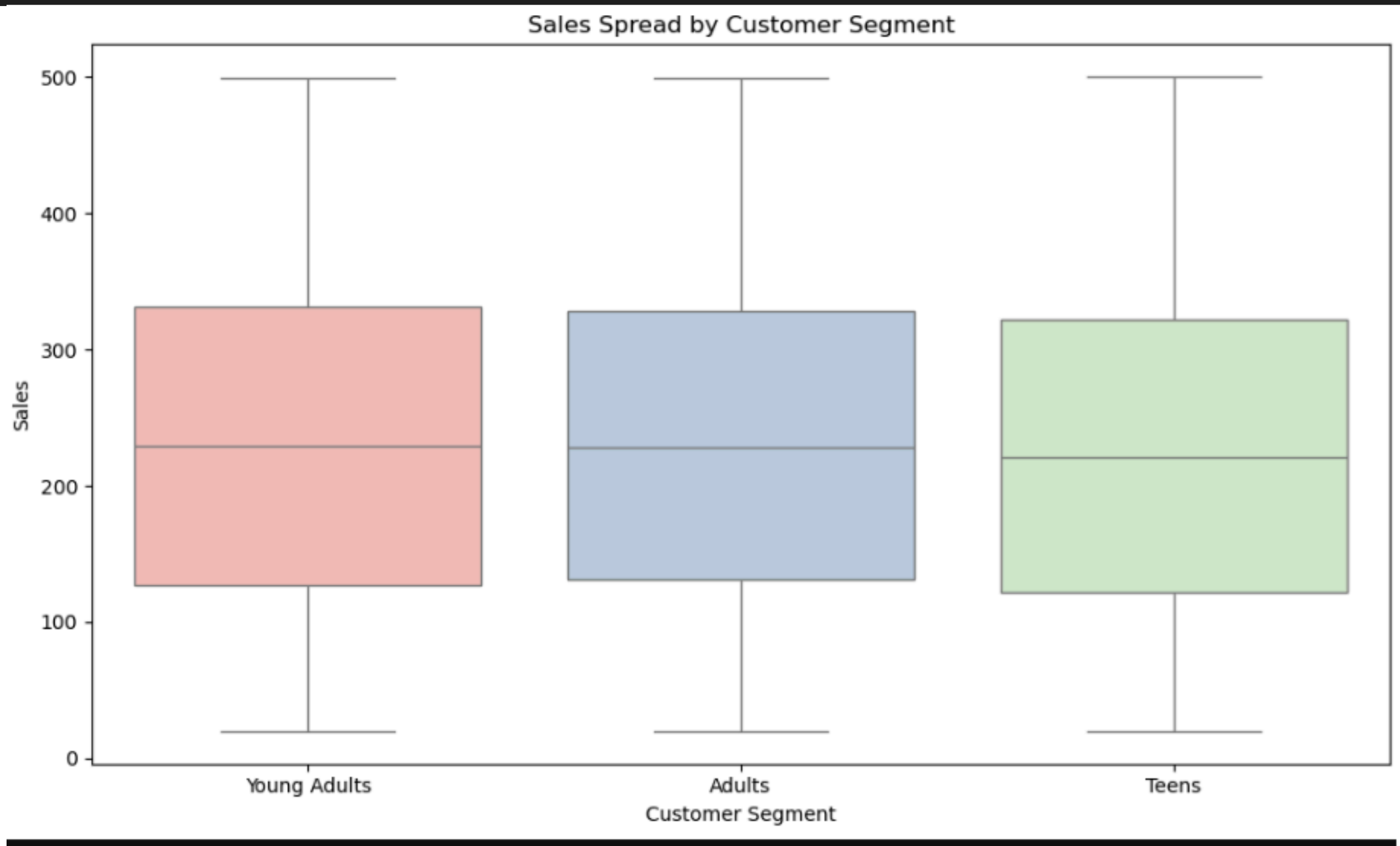


Objective 4: Understand Customer Segmentation Young Adults tend to drive higher average sales than Adults.
Custom campaigns targeting this group can yield better results—especially during events.

```
sns.histplot(data=df, x='Sales', hue='Customer Segment', kde=True, multiple='stack', palette='husl')
plt.title("Distribution of Sales Across Customer Segments")
plt.xlabel("Sales")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```



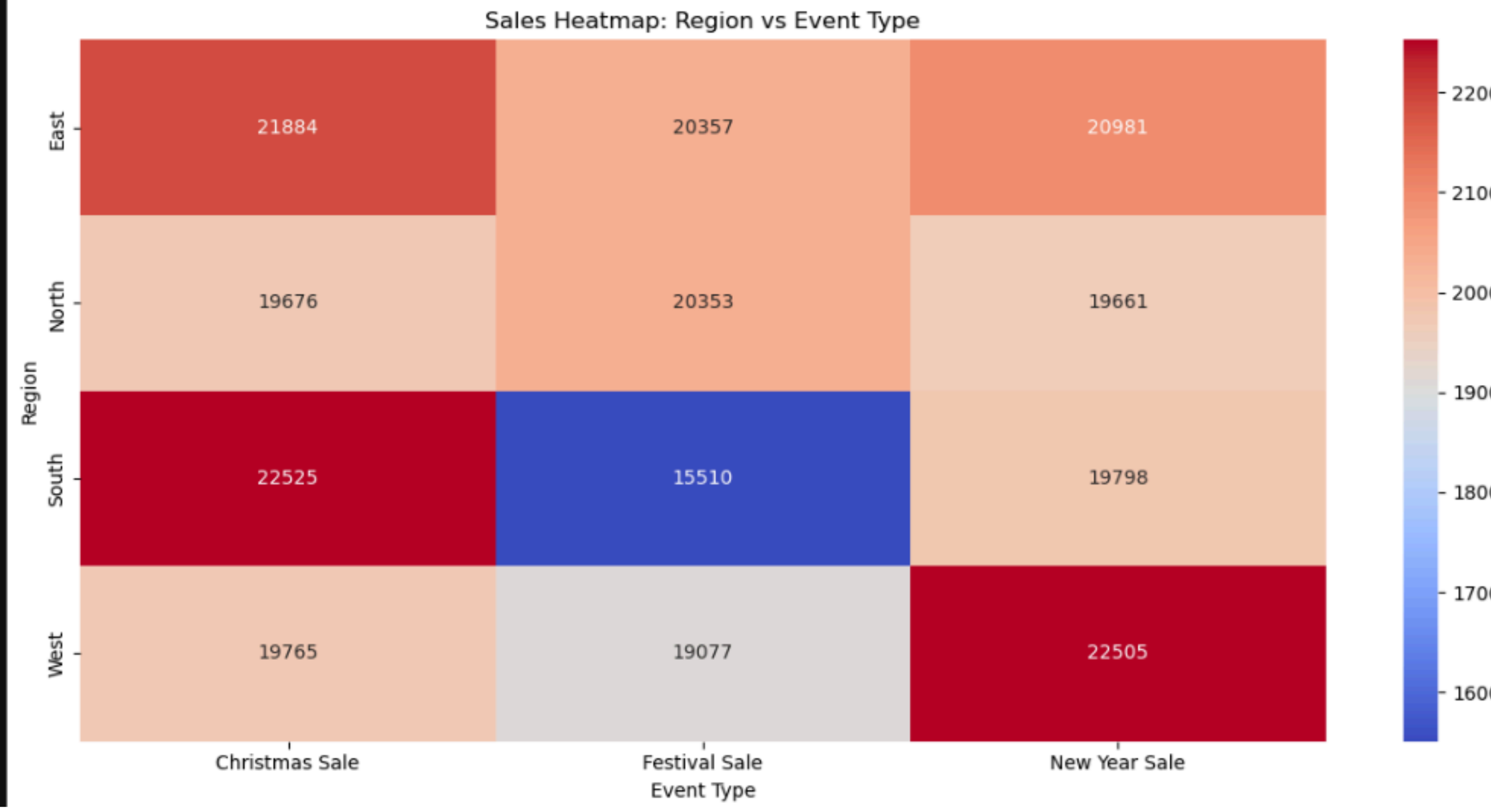

```
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='Customer Segment', y='Sales', palette='Pastel1')
plt.title("Sales Spread by Customer Segment")
plt.xlabel("Customer Segment")
plt.ylabel("Sales")
plt.tight_layout()
plt.show()
```



Objective 5: Regional and Event-based Performance Regions like West and South show higher sales—indicating stronger performance there. Event Types like "Christmas Sale" and "New Year Sale" drive up product performance significantly.

```
[325]: pivot = df.pivot_table(index='Region', columns='Event Type', values='Sales', aggfunc='sum', fill_value=0)

       plt.figure(figsize=(12, 6))
       sns.heatmap(pivot, cmap='coolwarm', annot=True, fmt='.0f')
       plt.title("Sales Heatmap: Region vs Event Type")
       plt.tight_layout()
       plt.show()
```



Future Scope

This project establishes a strong groundwork for understanding seasonal sales patterns, customer segmentation, and regional sales performance. In the future, the analysis can be expanded by incorporating advanced time-series forecasting models such as ARIMA, Prophet, or LSTM neural networks to improve accuracy in predicting future demand. Additionally, integrating real-time sales data through APIs or connected databases could enable dynamic dashboards and facilitate timely business decisions. Geographic insights can be deepened by including geo-spatial analysis through map-based visualizations, helping businesses tailor strategies based on regional trends. Furthermore, applying clustering algorithms like K-means or DBSCAN could enhance customer segmentation beyond age groups, uncovering behavior patterns based on purchase frequency or category preferences. Lastly, analyzing multi-channel engagement and social media influence can provide a broader perspective on promotional effectiveness and customer interaction, making the system more robust and aligned with modern marketing dynamics.

Conclusion

This project provided a detailed and insightful analysis of seasonal sales trends using the **Seasonality Demand Dataset** sourced from Kaggle. By applying data preprocessing, exploratory data analysis (EDA), and visualization techniques with Python, the project successfully uncovered patterns in customer behavior, regional preferences, and the influence of external factors such as promotions, events, and weather conditions.

The results show clear evidence of **peak demand during winter months** and festive seasons like **Christmas and New Year**, as well as the **positive impact of promotional strategies** on sales. Segmentation analysis revealed that **Young Adults are a key demographic**, while regions like **West and South consistently outperform others** in sales figures.

The project highlights the value of data science in driving business decisions. The visualizations made it easier to understand sales dynamics and provided actionable insights that can be used for campaign targeting, inventory planning, and demand forecasting. This analysis demonstrates how **Python and its data science libraries** empower businesses to make informed, evidence-based decisions by turning raw data into meaningful strategies.

References

1. UK Government Open Data Portal – Big Lottery Fund Grants Data (2004 onwards): http://data.gov.uk/dataset/blf-grants-data-2004-onwards
2. The Pandas Development Team. pandas: powerful Python data analysis toolkit. https://pandas.pydata.org/
3. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. https://matplotlib.org/
4. Waskom, M. et al. Seaborn: Statistical Data Visualization. https://seaborn.pydata.org/
5. Python Software Foundation. Python Language Reference, version 3.x. https://www.python.org/