IN PARTNERSHIP WITH PLYMOUTH UNIVERSITY

Name: Fernando Naveen	
Student Reference Number: 10899483	
Module Code: PUSL 2077	Module Name: Data Science in Python
Coursework Title: PUSL 2077 Final Project Report	
Deadline Date: 2nd April 2024	Member of staff responsible for coursework:
Programme: BSc (Honours) Data Science	

Please note that University Academic Regulations are available under Rules and Regulations on the University website www.plymouth.ac.uk/studenthandbook.

Group 4

ID No	Name
10899483	Fernando Naveen
10899488	Senanayake Senanayake
10899178	Konganige N Anthony
10899177	Urulugastenne Amarakone
10899495	Thehara Weerasiri

We confirm that we have read and understood the Plymouth University regulations relating to Assessment Offences and that we are aware of the possible penalties for any breach of these regulations. We confirm that this is the independent work of the group.

Signed on behalf of the group: FW Naveen

Individual assignment: ***I confirm that I have read and understood the Plymouth University regulations relating to Assessment Offences and that I am aware of the possible penalties for any breach of these regulations. I confirm that this is my own independent work.***

Signed :

Use of translation software: failure to declare that translation software or a similar writing aid has been used will be treated as an assessment offence.

I *have used/not used translation software.

If used, please state name of software………………………………………………………………

Overall mark _____% Assessors Initials ______ Date_________

PUSL 2077

Data Science in Python

Final Project Report

Group 4

Contents

Introduction

Sustained success in the athletic apparel industry requires a deep understanding of retail operations due to the intense competition and constantly changing consumer preferences. Considering this requirement, our project begins a thorough analysis of Adidas's retail operations, paying close attention to sales and profitability.

A giant in the sportswear sector, Adidas is always looking to improve and develop its retail strategy in response to changing consumer expectations and market trends. Our goal is to provide Adidas with useful information that comes from a close examination of the financial health of its retail division.

The project aims to shed light on Adidas's retail operations as they stand today by carefully examining important indicators to identify both the company's strong points and potential improvement areas. We hope to provide Adidas with vital guidance that will allow it to improve its business operations by closely examining profitability and sales performance.

In the end, our primary goal is to give Adidas a road map for skillfully negotiating the cutthroat landscape of the athletic wear industry. By providing strategic advice and insights, we hope to enable Adidas to improve productivity, encourage expansion, and forge a future of long-term success in the fiercely competitive market.

About the dataset

The dataset that is now being examined provides a thorough understanding of Adidas's sales performance from 2020 to 2021. Through a thorough analysis of the available data, the following primary categories can be identified:

1. **Sales Method:**

   The dataset outlines the many approaches Adidas uses to complete sales transactions. It is possible to assess the efficacy of various sales strategies by having a thorough understanding of the distribution of sales channels.

2. **Financial Information:**

   This section provides information about Adidas's financial performance with a particular emphasis on indicators related to operating profit and operating margin. When evaluating the profitability and operational effectiveness of the business, these financial indicators offer insightful context.

3. **Retailer Information:**

   This section includes information about six major Adidas-affiliated stores, providing a detailed overview of each one's contributions to the sales environment.

4. **Location Information:**

   Comprehensive geographic insights are offered, encompassing data on states, cities, and regions. This division makes it easier to investigate local sales patterns and customer inclinations.

5. **Product Information:**

   Important information can be found on product categories, unit costs, sold units, and overall sales numbers. A detailed examination of product performance, pricing tactics, and revenue creation across several product categories is made possible by this detailed data.

Descriptive Statistics

```
[6] df.describe()
```

	Price per Unit	Total Sales	Operating Profit	Operating Margin
count	9648.000000	9648.000000	9648.000000	9648.000000
mean	40.215692	1543.442993	2115.808458	28.304104
std	14.788030	823.963683	1247.109315	9.703310
min	0.000000	0.000000	0.000000	0.000000
25%	30.000000	904.750000	956.750000	21.000000
50%	40.000000	1500.000000	2044.500000	27.000000
75%	50.000000	2189.000000	3286.000000	35.000000
max	93.000000	3137.000000	4186.000000	65.000000

With the 'describe ()' function in python we can obtain the descriptive statistics related to out dataset. Here we can identify the key metrics which are considered in our dataset. The non-categorical columns are mentioned in the table as 'Price per unit', 'Total sales', 'Operating Profit' and 'Operating Margin'.

The first row indicates that there are 9648 data points in each column. The mean of each column indicates the average value of each column in the dataset. From this, we can see the average price per unit is $40 and that there has been $1543 of total sales overall. Then we have the operating profit and operating margin values as $1247 and $9 respectively.

Std row indicates the standard deviation of rows. The general idea of this is how the values have deviated around the mean values which we discussed above.

Then in the min row we have the minimum values of each non-categorical column. The minimum values for every column have been 0 according to this table.

25%,50%,75% rows indicate three different portions of our dataset. 25% denotes the values related to first quarter of our dataset and then 50% means the values belonging to the first half of the dataset. Here the values of the first quarter is also considered. Then we take 75% portion of our dataset and the values related to the third quarter is included here.

At last we have the maximum values of each non-categorical variable of our dataset. Here we can see the maximum price per unit is $93 and the highest number of total sales is $3137 while the maximum operating profit and operating margins are $4186 and $65 respectively.

Data Preprocessing

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
path = "/content/drive/MyDrive/Dataset/pythonproject.csv"
df=pd.read_csv(path)
```

```
df.rename(columns={'Unnamed: 0': 'Retailer'}, inplace=True)
df.rename(columns={'Unnamed: 1': 'Retailer_id'}, inplace=True)
df.rename(columns={'Unnamed: 2': 'Invoice Date'}, inplace=True)
df.rename(columns={'Unnamed: 3': 'Region'}, inplace=True)
df.rename(columns={'Unnamed: 4': 'State'}, inplace=True)
df.rename(columns={'Unnamed: 5': 'City'}, inplace=True)
df.rename(columns={'Unnamed: 6': 'Product'}, inplace=True)
df.rename(columns={'Unnamed: 7': 'Price per Unit'}, inplace=True)
df.rename(columns={'Unnamed: 8': 'Units Sold'}, inplace=True)
df.rename(columns={'Unnamed: 9': 'Total Sales'}, inplace=True)
df.rename(columns={'Unnamed: 10': 'Operating Profit'}, inplace=True)
df.rename(columns={'Unnamed: 11': 'Operating Margin'}, inplace=True)
df.rename(columns={'Unnamed: 12': 'Sales Method'}, inplace=True)
```

```
df.head()
```

	Retailer	Retailer_id	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method
0	NaN	Adidas Sales Database	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Retailer	Retailer ID	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method
4	Foot Locker	1185732	1/1/2020	Northeast	New York	New York	Men's Street Footwear	$50.00	1,200	$600,000	$300,000	50%	In-store

```
indices_to_delete = list(range(0, 4))
df.drop(indices_to_delete, inplace=True)
```

```
df.head()
```

	Retailer	Retailer_id	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method
4	Foot Locker	1185732	1/1/2020	Northeast	New York	New York	Men's Street Footwear	$50.00	1,200	$600,000	$300,000	50%	In-store
5	Foot Locker	1185732	1/2/2020	Northeast	New York	New York	Men's Athletic Footwear	$50.00	1,000	$500,000	$150,000	30%	In-store
6	Foot Locker	1185732	1/3/2020	Northeast	New York	New York	Women's Street Footwear	$40.00	1,000	$400,000	$140,000	35%	In-store
7	Foot Locker	1185732	1/4/2020	Northeast	New York	New York	Women's Athletic Footwear	$45.00	850	$382,500	$133,875	35%	In-store
8	Foot Locker	1185732	1/5/2020	Northeast	New York	New York	Men's Apparel	$60.00	900	$540,000	$162,000	30%	In-store

```python
columns_to_encode = [ 'State', 'Price per Unit','Total Sales','Operating Profit','Operating Margin','Units Sold']

label_encoder = LabelEncoder()

for column in columns_to_encode:
    df[column] = label_encoder.fit_transform(df[column])
```

```python
df
```

	Retailer	Retailer_id	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method
4	Foot Locker	1185732	1/1/2020	Northeast	31	New York	Men's Street Footwear	45	11	2472	2475	36	In-store
5	Foot Locker	1185732	1/2/2020	Northeast	31	New York	Men's Athletic Footwear	45	1	2188	954	16	In-store
6	Foot Locker	1185732	1/3/2020	Northeast	31	New York	Women's Street Footwear	35	1	1898	924	21	In-store
7	Foot Locker	1185732	1/4/2020	Northeast	31	New York	Women's Athletic Footwear	40	341	1637	906	21	In-store
8	Foot Locker	1185732	1/5/2020	Northeast	31	New York	Men's Apparel	55	348	2209	988	16	In-store
...	...	...	...	...	...	...	...	...	...	...	...	...	...
9647	Foot Locker	1185732	1/24/2021	Northeast	28	Manchester	Men's Apparel	45	305	1362	4036	14	Outlet
9648	Foot Locker	1185732	1/24/2021	Northeast	28	Manchester	Women's Apparel	36	20	1721	294	18	Outlet
9649	Foot Locker	1185732	2/22/2021	Northeast	28	Manchester	Men's Street Footwear	36	98	2630	1616	23	Outlet
9650	Foot Locker	1185732	2/22/2021	Northeast	28	Manchester	Men's Athletic Footwear	37	316	1194	189	28	Outlet
9651	Foot Locker	1185732	2/22/2021	Northeast	28	Manchester	Women's Street Footwear	24	338	1043	3609	13	Outlet

9648 rows × 13 columns

```python
df.isnull().sum()
```

```
Retailer            0
Retailer_id         0
Invoice Date        0
Region              0
State               0
City                0
Product             0
Price per Unit      0
Units Sold          0
Total Sales         0
Operating Profit    0
Operating Margin    0
Sales Method        0
dtype: int64
```

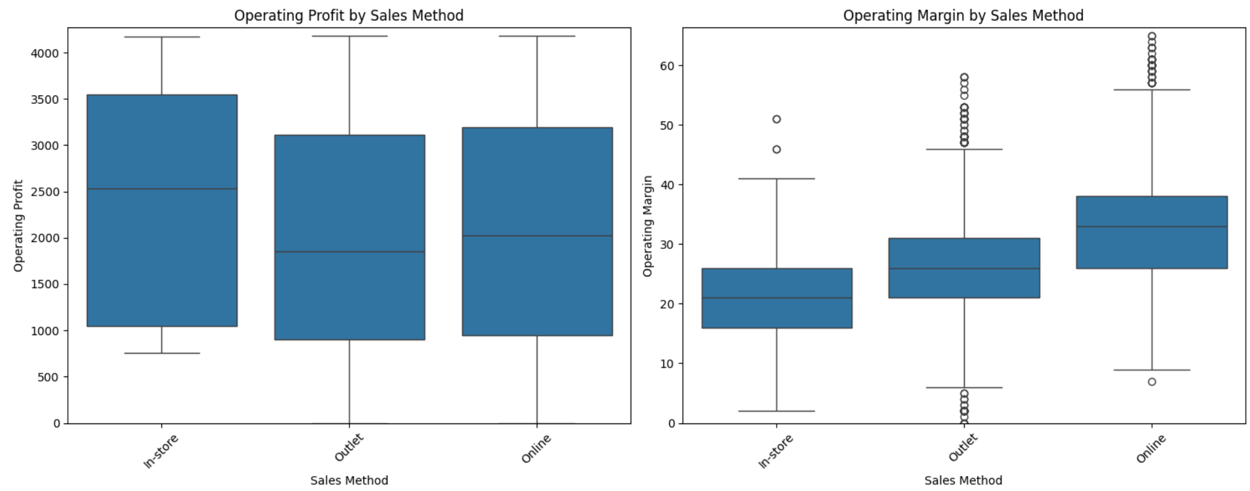
Data visualization

I. Sales methods

The data set lists the many methods Adidas use to finish sales transactions. A detailed grasp of the distribution of sales channels makes it feasible to evaluate the effectiveness of different sales techniques.

1. Boxplot visualizes the distribution of operating profit and operating margin by sales method.

```
fig, axes = plt.subplots(1, 2, figsize=(15, 6))

sns.boxplot(ax=axes[0], x='Sales Method', y='Operating Profit', data=df)
axes[0].set_xlabel('Sales Method')
axes[0].set_ylabel('Operating Profit')
axes[0].set_title('Operating Profit by Sales Method')
axes[0].set_ylim(0, max(df['Operating Profit']) * 1.02)
axes[0].tick_params(axis='x', rotation=45)

sns.boxplot(ax=axes[1], x='Sales Method', y='Operating Margin', data=df)
axes[1].set_xlabel('Sales Method')
axes[1].set_ylabel('Operating Margin')
axes[1].set_title('Operating Margin by Sales Method')
axes[1].set_ylim(0, max(df['Operating Margin']) * 1.02)
axes[1].tick_params(axis='x', rotation=45)

plt.tight_layout()
plt.show()
```

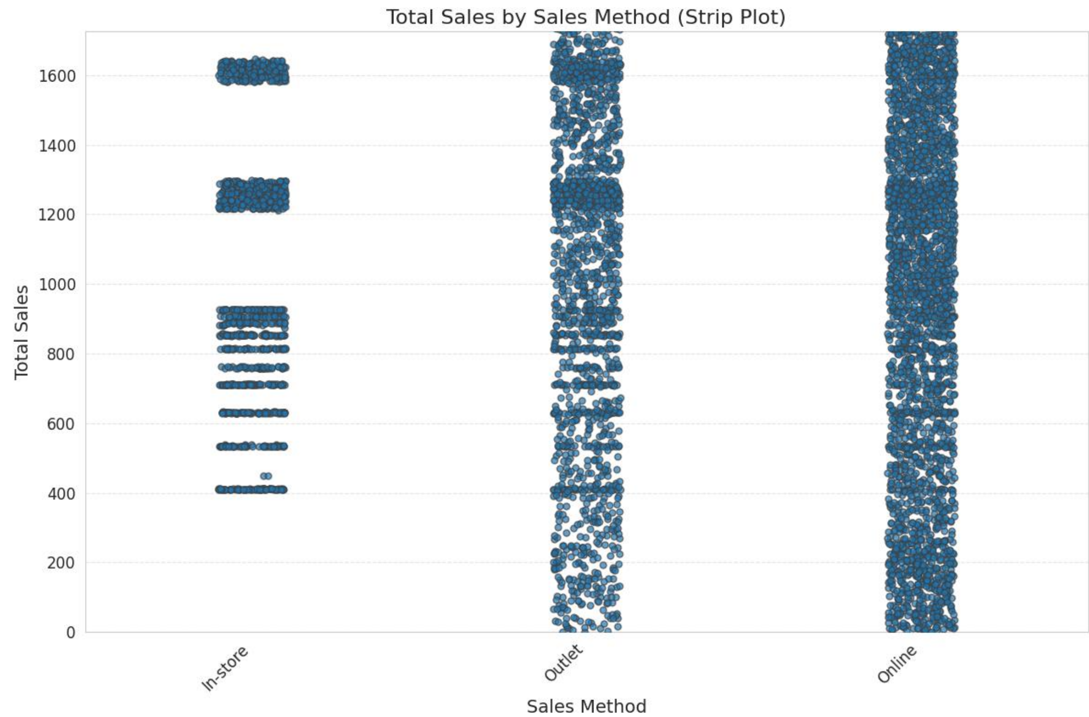


It is clear from the box-plot analysis that operating profits from in-store sales are higher than operating profits from other sales channels. When comparing the boxplots of outlet and online sale methods, the one for in-store sales shows a broader interquartile range and a higher median. This implies that in-store sales are consistently and relatively more profitable.

On the other hand, the boxplot for the online sales method shows that it generates the largest operational margins, even if in-store sales show higher operating profits. Online sales show a narrower spread of data points around the median, indicating a bigger proportion of sales with higher profit margins, even though operating profits may be lower than in-store sales.

2. <u>Total sales by Sales method (Strip plot)</u>

```
sns.set_style("whitegrid")

plt.figure(figsize=(12, 8))
sns.stripplot(x='Sales Method', y='Total Sales', data=df, jitter=True, dodge=True, linewidth=1, edgecolor='gray', alpha=0.7)

plt.xlabel('Sales Method', fontsize=14)
plt.ylabel('Total Sales', fontsize=14)
plt.title('Total Sales by Sales Method (Strip Plot)', fontsize=16)

plt.ylim(0, max(df['Total Sales']) * 0.55)

plt.xticks(rotation=45, ha='right', fontsize=12)

plt.yticks(fontsize=12)

plt.grid(axis='y', linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()
```



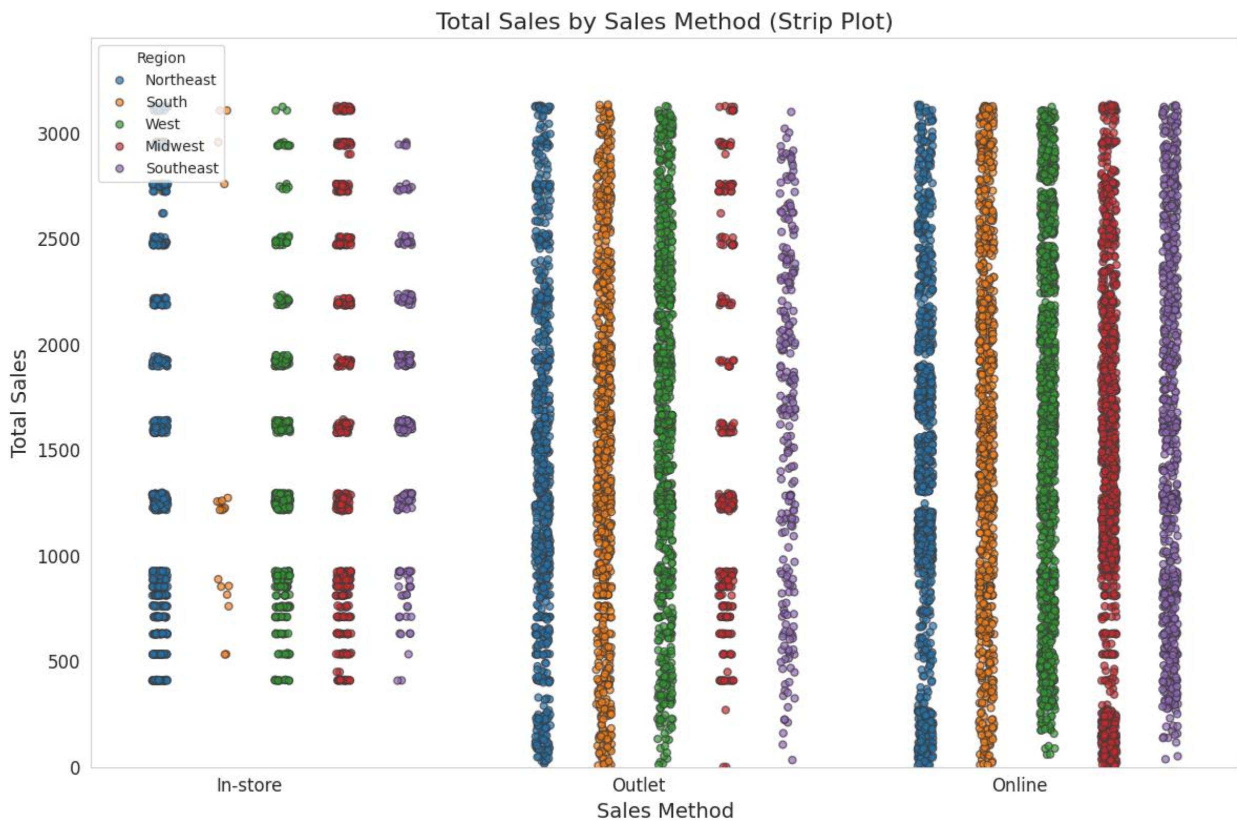
The strip plot visualizes the distribution of total sales across different sales channels, namely in-store, outlet, and online. It's clear from analysis that these channels have different distributions of overall sales.

The strip plot indicates a more concentrated distribution in the case of in-store sales, indicating that overall sales typically fall within a specific range. This concentration suggests that sales volume within physical retail outlets follows a more regular pattern.

On the other hand, the strip plot shows how total sales for outlets and online platforms are distributed more widely and unevenly. There is a greater range of sales volumes inside these channels, as seen by the total sales that are recorded throughout different ranges. This wider dispersion points to higher levels of flexibility and unpredictability in sales performance, which may be impacted by variables including regional reach, marketing initiatives, and online environments.

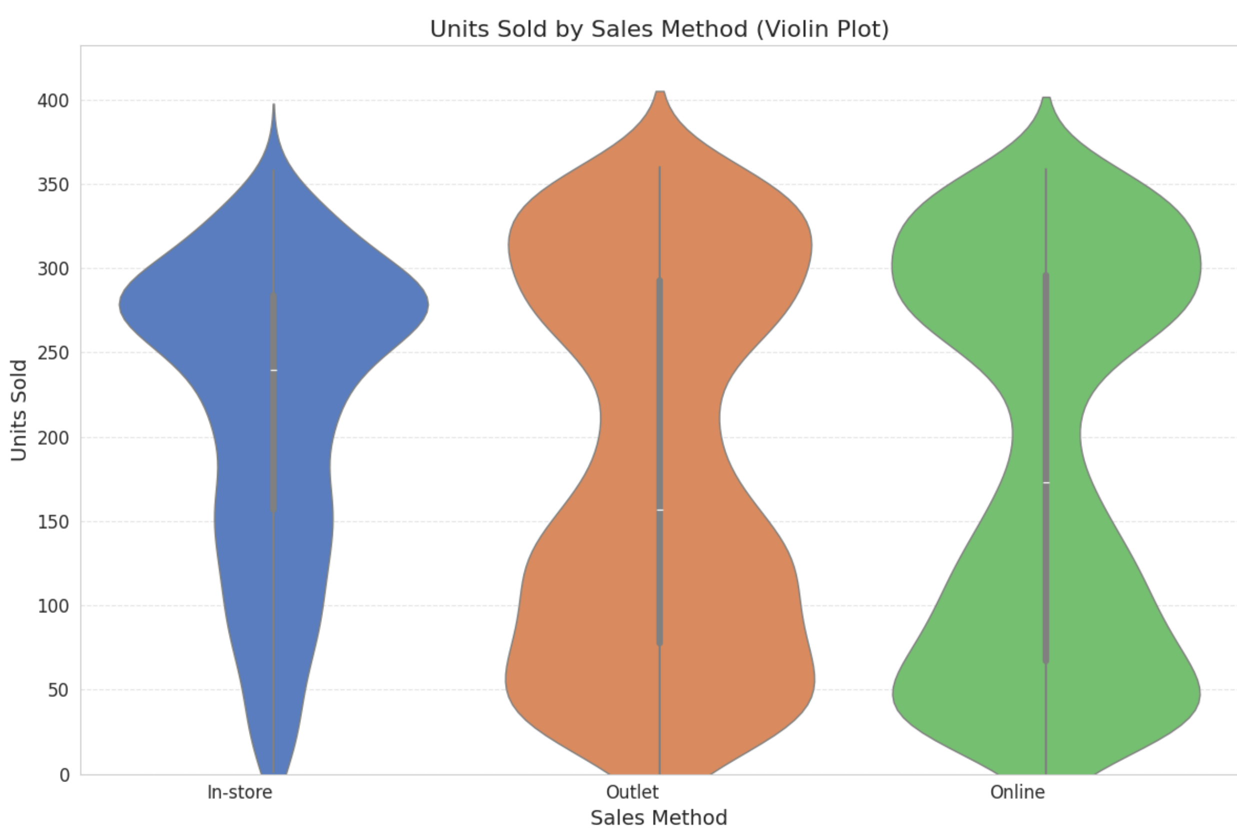
Furthermore, the following strip plot shows how the total sales in each sales method region wise.

```
sns.set_style("whitegrid")

plt.figure(figsize=(12, 8))
sns.stripplot(x='Sales Method', y='Total Sales', hue='Region', data=df, jitter=True, dodge=True, linewidth=1, edgecolor='gray', alpha=0.7)

plt.xlabel('Sales Method', fontsize=14)
plt.ylabel('Total Sales', fontsize=14)
plt.title('Total Sales by Sales Method (Strip Plot)', fontsize=16)

plt.ylim(0, max(df['Total Sales']) * 1.1)

plt.xticks(rotation=0, ha='right', fontsize=12)

plt.yticks(fontsize=12)

plt.legend(title='Region')

plt.grid(False)

plt.tight_layout()
plt.show()
```



3. Violin plot with total sales by sales method

```
plt.figure(figsize=(12, 8))
sns.violinplot(x='Sales Method', y='Units Sold', data=df, linewidth=1, edgecolor='gray', palette='muted')

plt.xlabel('Sales Method', fontsize=14)
plt.ylabel('Units Sold', fontsize=14)
plt.title('Units Sold by Sales Method (Violin Plot)', fontsize=16)

plt.ylim(0, max(df['Units Sold']) * 1.2)

plt.xticks(rotation=0, ha='right', fontsize=12)

plt.yticks(fontsize=12)

plt.grid(axis='y', linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()
```



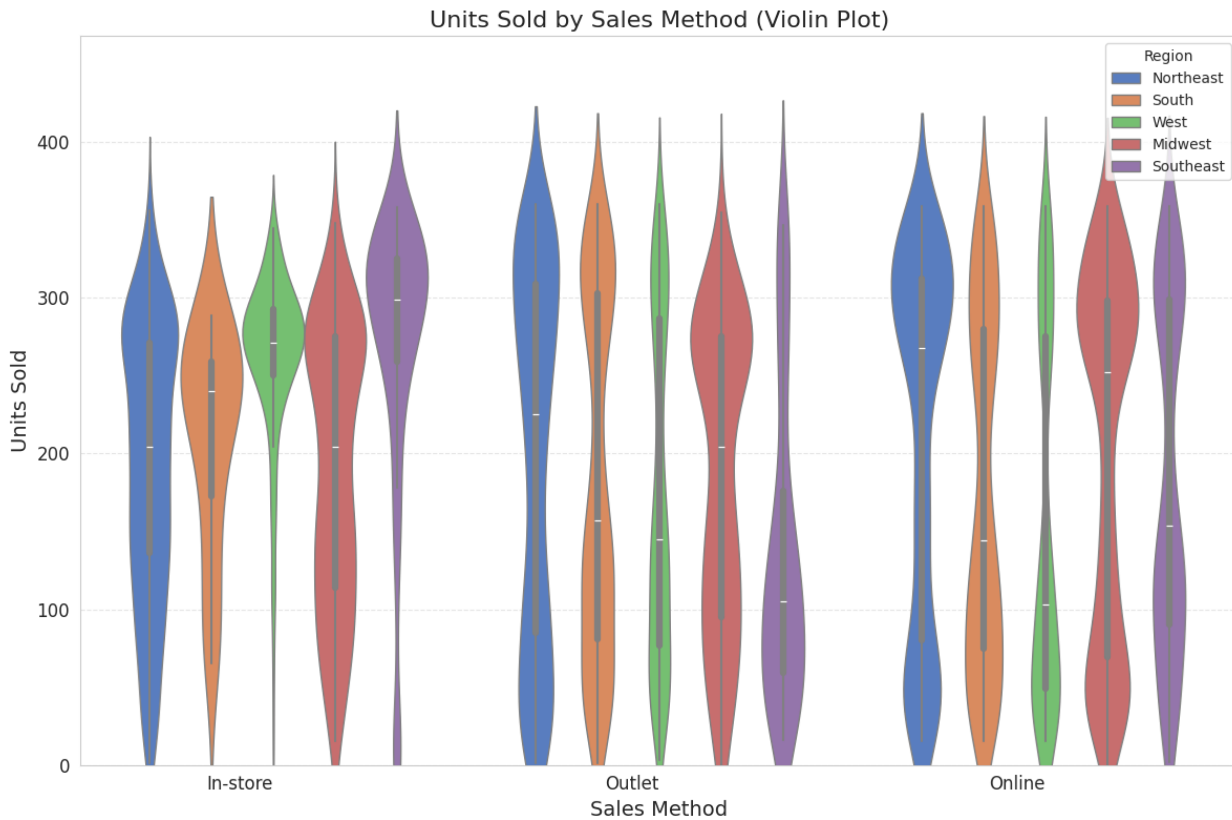
The violin plot provides valuable insights into how units are distributed among different selling methods, such as online, in-store, and outlet sales.

First, the violin plot makes it clear that, in comparison to the other methods of sales, the in-store sales method shows a noticeably higher number of units sold. A larger density of units sold throughout a range of prices is shown by the violin plot for in-store sales, which has a wider and more prominent shape. This implies that in-store sales typically draw more clients or transactions, which raises the number of units sold.

On the other hand, as the violin plot shows, outlet sales and online sales have similar unit distributions. The violins' width and form of their online and store sales seem quite similar, suggesting a comparable distribution and density of units sold at various price points. Additionally, the violin plot indicates that the online and outlet sales channels have almost the same mean number of units sold.

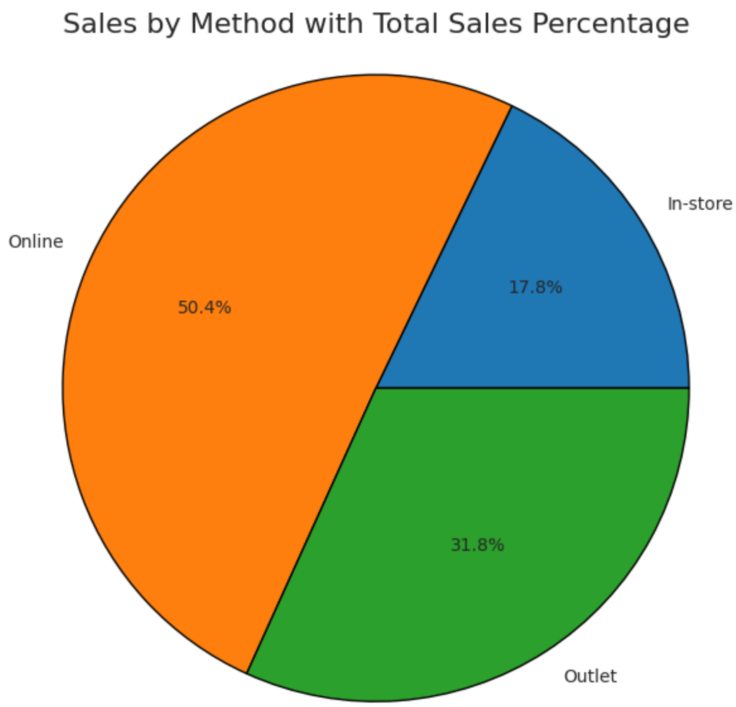
Furthermore, the following violin shows how the units sold in each sales method region wise.

```
sns.set_style("whitegrid")

plt.figure(figsize=(12, 8))
sns.violinplot(x='Sales Method', y='Units Sold', data=df, linewidth=1, edgecolor='gray',hue='Region', palette='muted')

plt.xlabel('Sales Method', fontsize=14)
plt.ylabel('Units Sold', fontsize=14)
plt.title('Units Sold by Sales Method (Violin Plot)', fontsize=16)

plt.ylim(0, max(df['Units Sold']) * 1.3)

plt.xticks(rotation=0, ha='right', fontsize=12)

plt.yticks(fontsize=12)

plt.grid(axis='y', linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()
```



4. <u>Sales by method with total sales percentage.</u>

```
sales_by_method = df.groupby('Sales Method')['Total Sales'].sum()

colors = plt.cm.tab10.colors

plt.figure(figsize=(10, 6))
plt.pie(sales_by_method, labels=sales_by_method.index, autopct='%1.1f%%', startangle=0, colors=colors, wedgeprops={'edgecolor': 'black'})

plt.title('Sales by Method with Total Sales Percentage', fontsize=16)

plt.axis('equal')

plt.tight_layout()
plt.show()
```



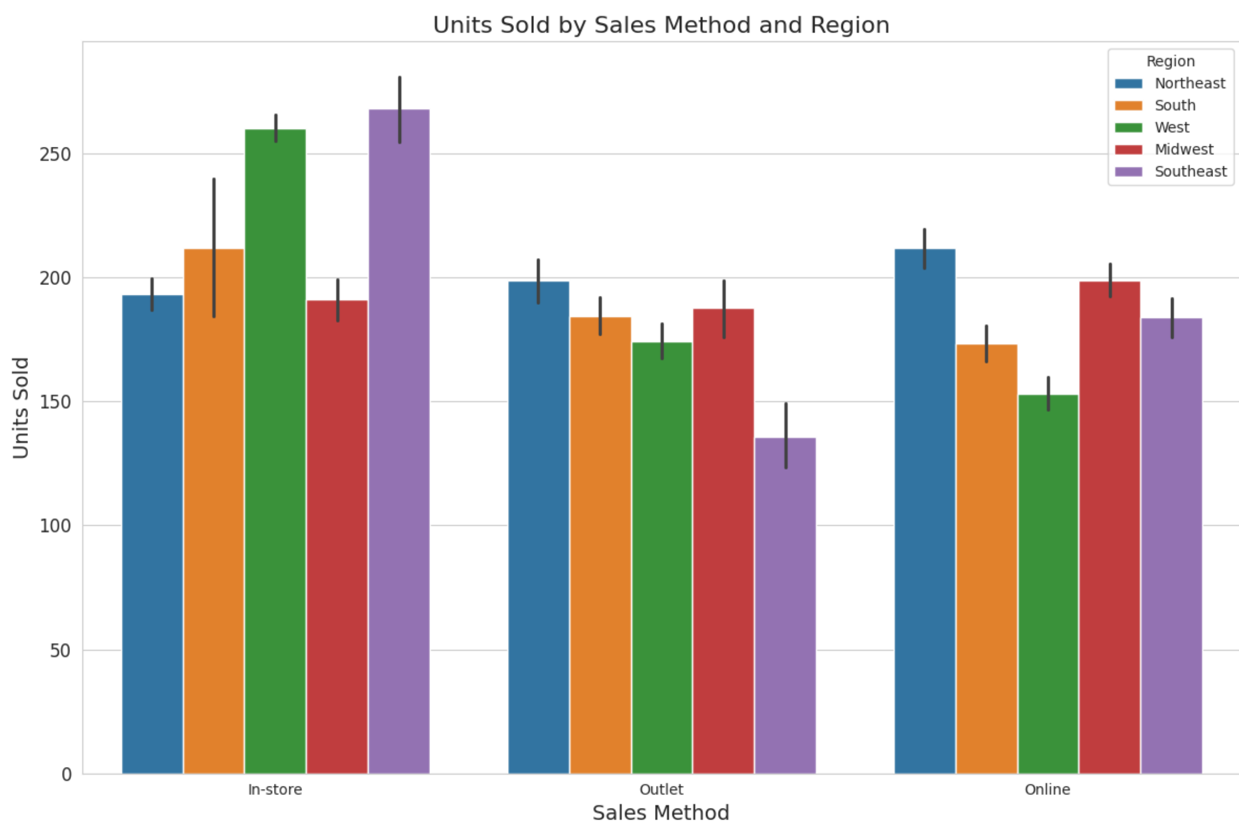
According to the chart, the sales method that has the highest percentage of total sales is online sales that takes up the largest amount, which is 50.5% of the entire sales volume. This suggests that the online sales approach has made a considerable impact on overall sales.

In addition, the pie chart shows that 17.8% of total sales come from in-store purchases. This implies that in-store transactions account for a modest fraction of total sales, which is a significant but relatively smaller portion of total sales volume.

However, with 31.8% of the overall sales volume, outlet sales make up a bigger share of the total sales. This shows that outlet sales make up a sizeable portion of overall revenue, underscoring the role that this sales channel plays in boosting sales results.

5. **Units sold by sales method in region.**

```
plt.figure(figsize=(12, 8))
sns.barplot(x='Sales Method', y='Units Sold', hue='Region', data=df)

plt.xlabel('Sales Method', fontsize=14)
plt.ylabel('Units Sold', fontsize=14)
plt.title('Units Sold by Sales Method and Region', fontsize=16)

plt.yticks(fontsize=12)

plt.legend(title='Region')

plt.tight_layout()
plt.show()
```



A comparison of units sold by sales technique across several regions is shown in the grouped barplot. Plot analysis reveals several noteworthy findings about sales performance across regions and sales strategies.

The southwest, Midwest, and northwest regions stand out as the top-performing locations for the in-store sales strategy, with the highest units sold. Remarkably, the Midwest and northwest areas trail closely behind the southwest region in terms of units sold. Even while the west region reports significantly lower unit sales than the southwest, it nevertheless shows a sizable number of in-store transactions.

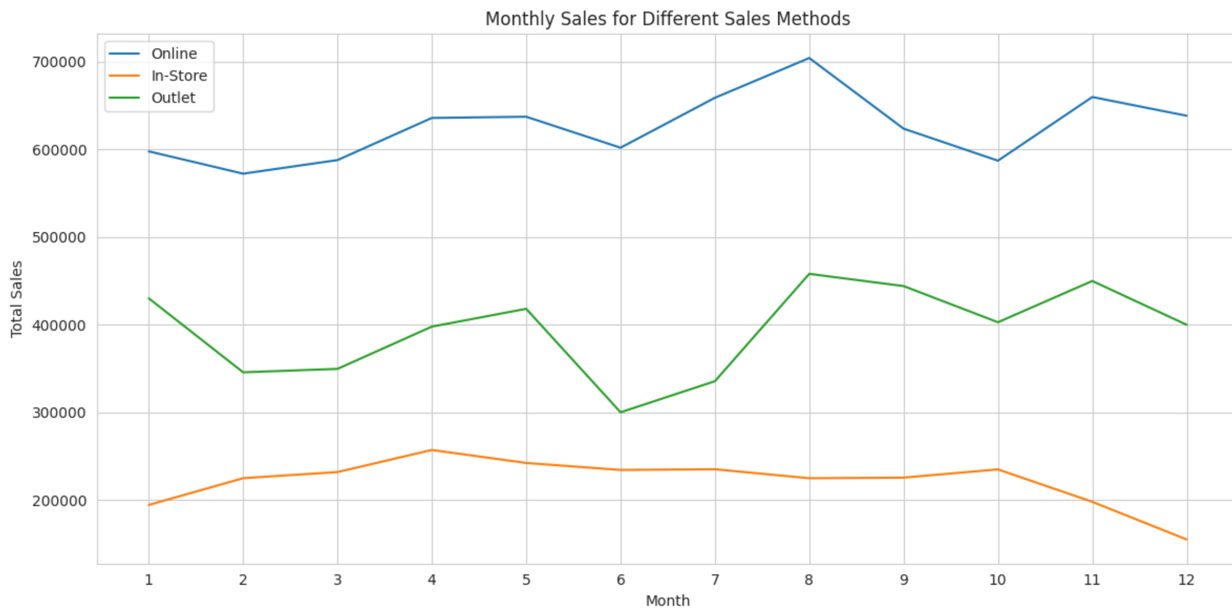
Furthermore, although the difference is quite small, the south region reported selling a little more unit than the northeast.

On the other hand, when compared to in-store and online techniques, the outlet sales approach consistently shows fewer units sold across all geographies. In terms of units sold, in-store transactions easily outperform the other two sales techniques, demonstrating their consistent performance across geographic boundaries.

In terms of units sold, it highlights the superiority of in-store sales while also highlighting the relatively poor performance of outlet and online selling strategies.

6. Monthly Sales for Different Sales Method

```
fig, ax = plt.subplots(figsize=(12, 6))

df_online['Month'] = pd.to_datetime(df_online['Invoice Date']).dt.month
monthly_sales_online = df_online.groupby('Month')['Total Sales'].sum()
ax.plot(monthly_sales_online.index, monthly_sales_online.values, label='Online')

df_in_store['Month'] = pd.to_datetime(df_in_store['Invoice Date']).dt.month
monthly_sales_in_store = df_in_store.groupby('Month')['Total Sales'].sum()
ax.plot(monthly_sales_in_store.index, monthly_sales_in_store.values, label='In-Store')

df_outlet['Month'] = pd.to_datetime(df_outlet['Invoice Date']).dt.month
monthly_sales_outlet = df_outlet.groupby('Month')['Total Sales'].sum()
ax.plot(monthly_sales_outlet.index, monthly_sales_outlet.values, label='Outlet')

ax.set_title('Monthly Sales for Different Sales Methods')
ax.set_xlabel('Month')
ax.set_ylabel('Total Sales')

ax.set_xticks(monthly_sales_online.index)

ax.legend()

plt.tight_layout()
plt.show()
```

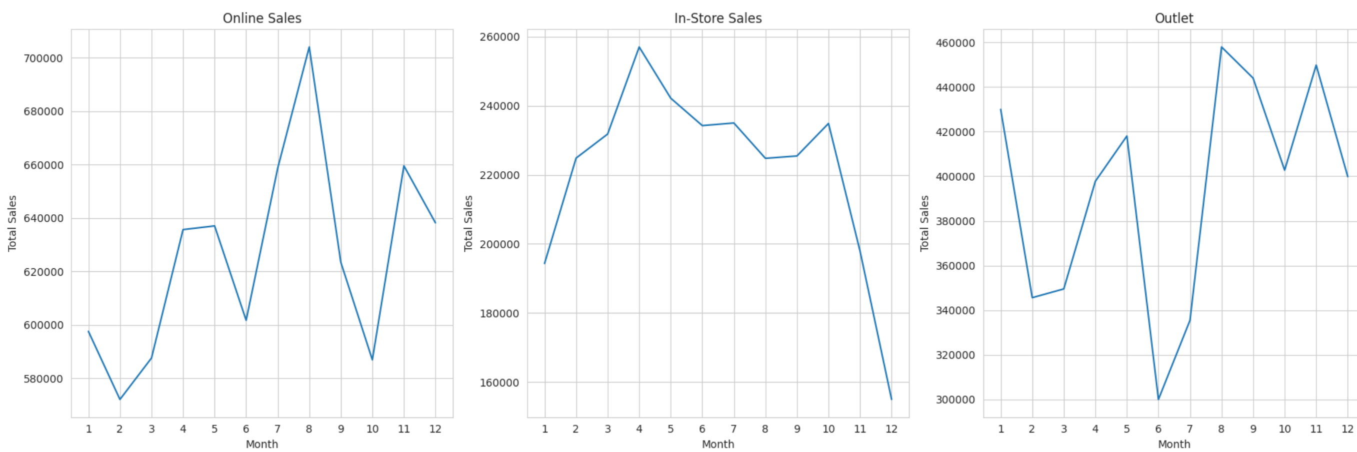


Online sales regularly show the largest monthly sales volume throughout the year, according to the line plot. Interestingly, online sales reached their highest point in August, suggesting a notable spike in sales activity during this time. Outlet sales come in second place to online sales in terms of monthly sales volume contribution. Nonetheless, there is a clear drop in outlet sales in June, indicating a reduction in sales performance during that time. When comparing in-store sales to online and outlet sales techniques, in-store sales consistently show the lowest monthly sales volume. Though they have remained relatively consistent throughout the year, December saw a little decline in in-store sales.

The line plot shows changes in sales performance over time and displays the monthly sales trends for various sales techniques. It highlights how online sales generate the most money every month, but it also highlights the year-over-year variations and patterns in outlet and in-store sales.

From the below show visualization we can get a better understanding about the trends in the monthly sales in each sales method.

```
fig, axes = plt.subplots(1, 3, figsize=(18, 6))

df_online['Month'] = pd.to_datetime(df_online['Invoice Date']).dt.month
monthly_sales_online = df_online.groupby('Month')['Total Sales'].sum()
axes[0].plot(monthly_sales_online.index, monthly_sales_online.values)
axes[0].set_title('Online Sales')
axes[0].set_xlabel('Month')
axes[0].set_ylabel('Total Sales')
axes[0].set_xticks(monthly_sales_online.index)

df_in_store['Month'] = pd.to_datetime(df_in_store['Invoice Date']).dt.month
monthly_sales_in_store = df_in_store.groupby('Month')['Total Sales'].sum()
axes[1].plot(monthly_sales_in_store.index, monthly_sales_in_store.values)
axes[1].set_title('In-Store Sales')
axes[1].set_xlabel('Month')
axes[1].set_ylabel('Total Sales')
axes[1].set_xticks(monthly_sales_in_store.index)

df_outlet['Month'] = pd.to_datetime(df_outlet['Invoice Date']).dt.month
monthly_sales_outlet = df_outlet.groupby('Month')['Total Sales'].sum()
axes[2].plot(monthly_sales_outlet.index, monthly_sales_outlet.values)
axes[2].set_title('Outlet')
axes[2].set_xlabel('Month')
axes[2].set_ylabel('Total Sales')
axes[2].set_xticks(monthly_sales_outlet.index)

plt.tight_layout()
plt.show()
```

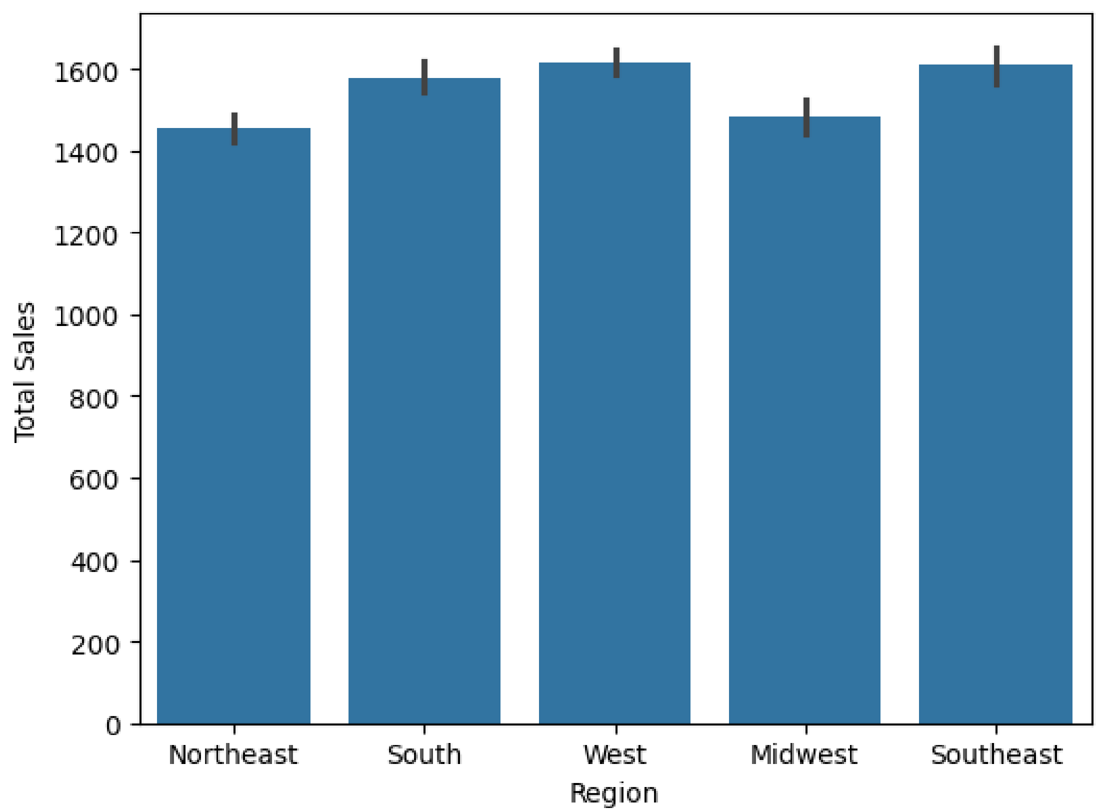


II. Financial Information:

This section provides information about Adidas's financial performance with a particular emphasis on indicators related to operating profit and operating margin. When evaluating the profitability and operational effectiveness of the business, these financial indicators offer insightful context.

1. Region wise sales of the company

```
sns.barplot(x='Region', y='Total Sales', data=df)
plt.show()
```



The whole sales of Adidas items across the US appear to be represented in the bar chart . The Northeast, South, West, Midwest, and Southeast are the regions listed on the chart's x-axis. The overall sales data are displayed on the y-axis. The bars have a blue tint.

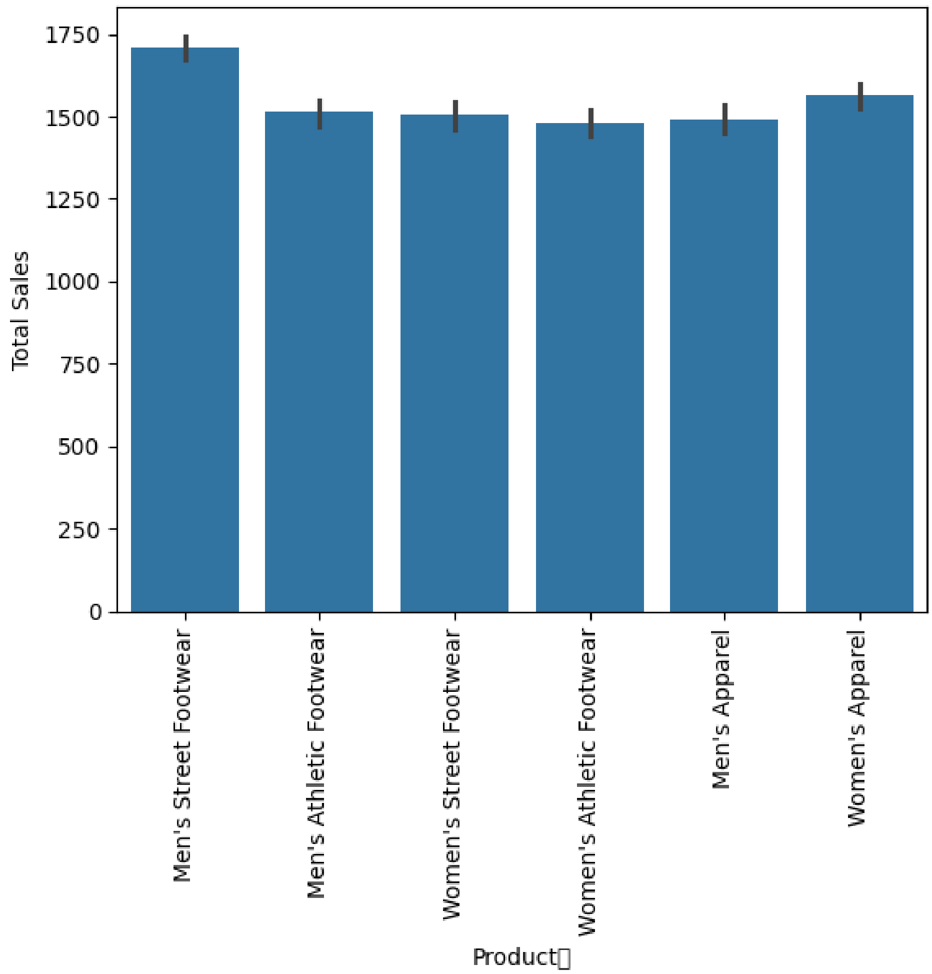
The Northeast seems to have the most overall sales, with the South, Midwest, West, and Southeast following behind.

2. Total sales of different product categories

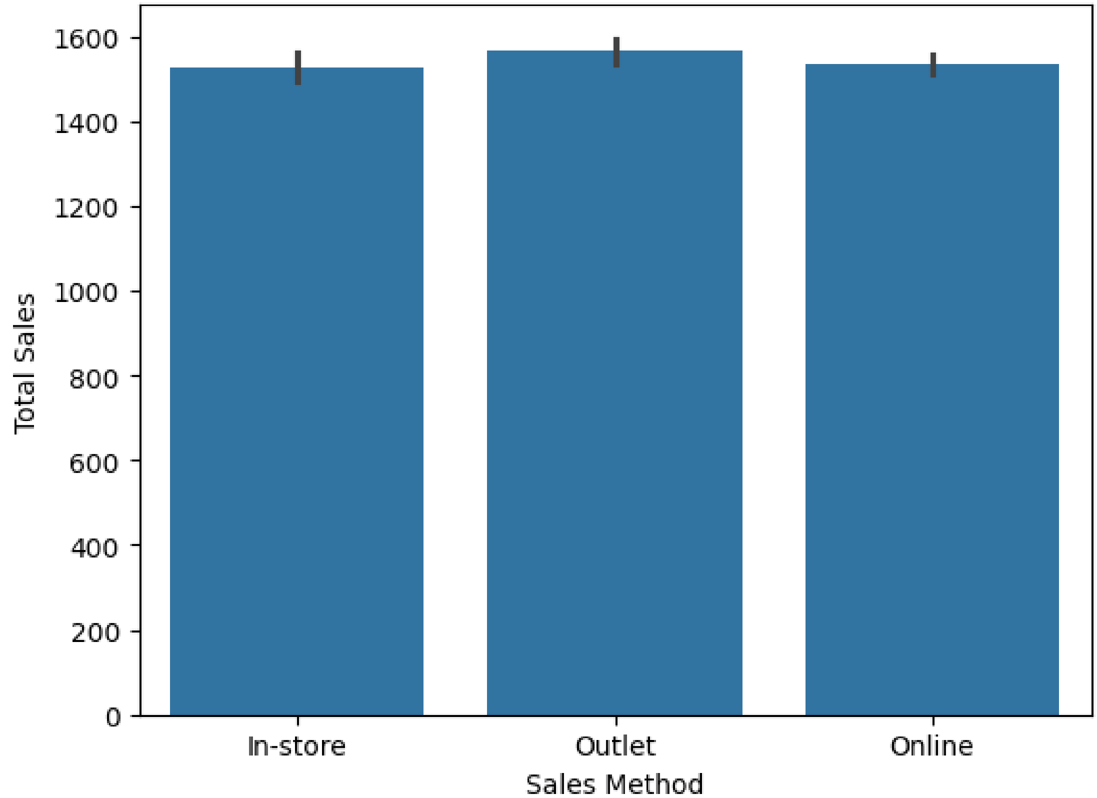
In this below bar graph that displays the overall sales figures for the various product categories. Men's street footwear, men's athletic footwear, women's street footwear, women's athletic footwear, men's apparel, and women's apparel are the product categories listed on the x-axis. The overall sales data appears on the y-axis.

According to the graphic, men's sports footwear, women's street footwear, women's clothing, and women's street footwear are the next greatest categories in terms of total sales, after women's clothes.

```
[ ]

    sns.barplot(x='Product\t', y='Total Sales', data=df)
    plt.xticks(rotation=90)
    plt.show()
```


1750
1500
1250
1000
750
500
250
0
Total Sales
Men's Street Footwear
Men's Athletic Footwear
Women's Street Footwear
Women's Athletic Footwear
Men's Apparel
Women's Apparel
Product

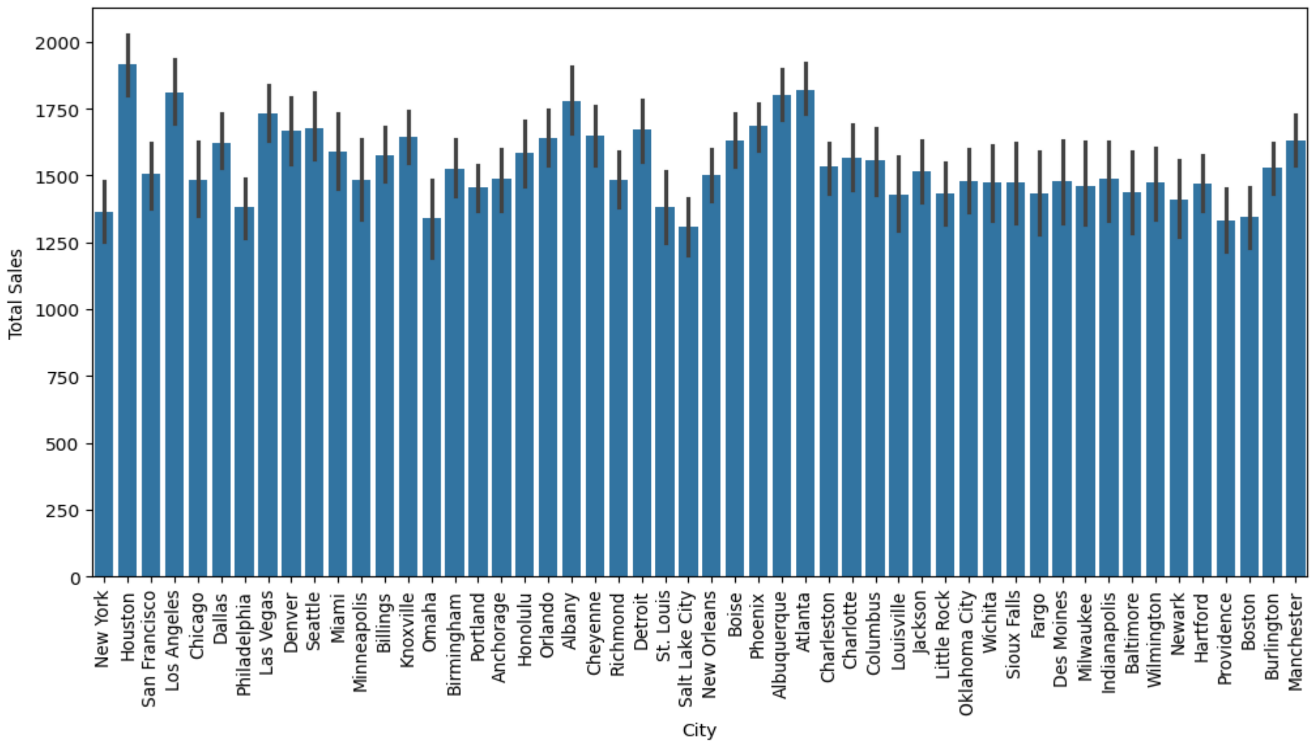
3. Behavior of total sales with different sales methods



This bar graph displaying the total sales made from online, in-store, and outlet sales. The graph's y-axis displays the total sales, while the graph's x-axis specifies the sales technique. Over the course of the graph's display, online sales seem to have peaked.
The second-highest sales appear to be in-store.
Over the course of the graph's display, outlet sales seem to have been the lowest.

```
sns.barplot(x='Sales Method', y='Total Sales', data=df)
plt.show()
```

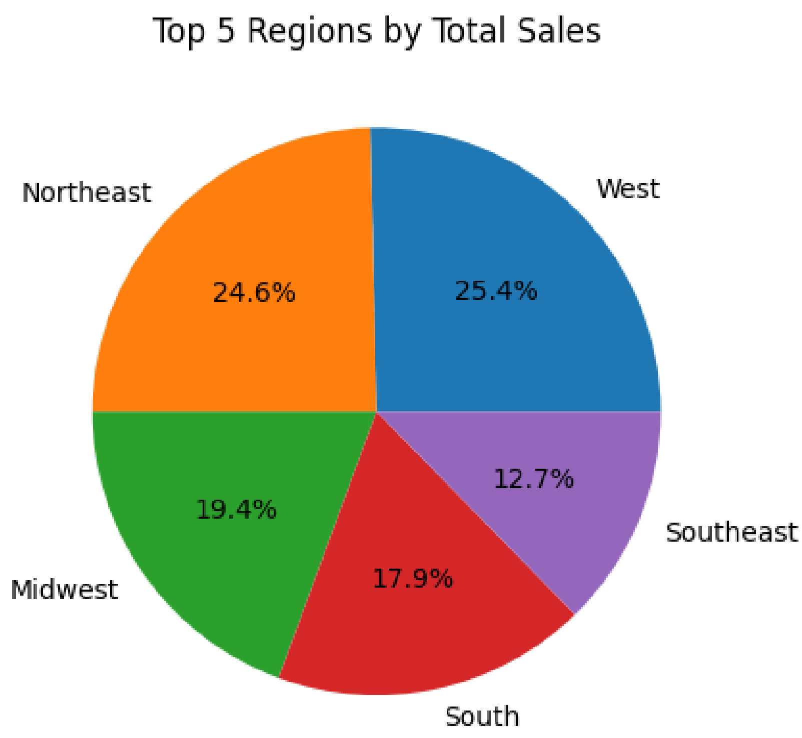
4. City wise sales in Adidas



The above bar graph shows how adidas sales preforming in different city areas .According this bar chart we can seen Housten is the best sales preforming city .And we can see Orlando is weakest preforming city in sales .Addidas can get more sales with forcing on lowest preforming areas and running marketing campaigns it that specific areas .

```
plt.figure(figsize=(12, 6))
sns.barplot(x='City', y='Total Sales', data=df)
plt.xticks(rotation=90)
plt.show()
```

5. Top Five Regions according to their total sales



The total sales are split into five slices, each of which represents a region, in the pie chart. The name of the area and its share of the overall sales are labeled on each slice.

With a bright blue tint and the label "West," the largest slice of the pie chart represents 25.4% of the total sales.
With a light green tint and the label "Northeast," the second-largest slice makes up 24.6% of the total sales.

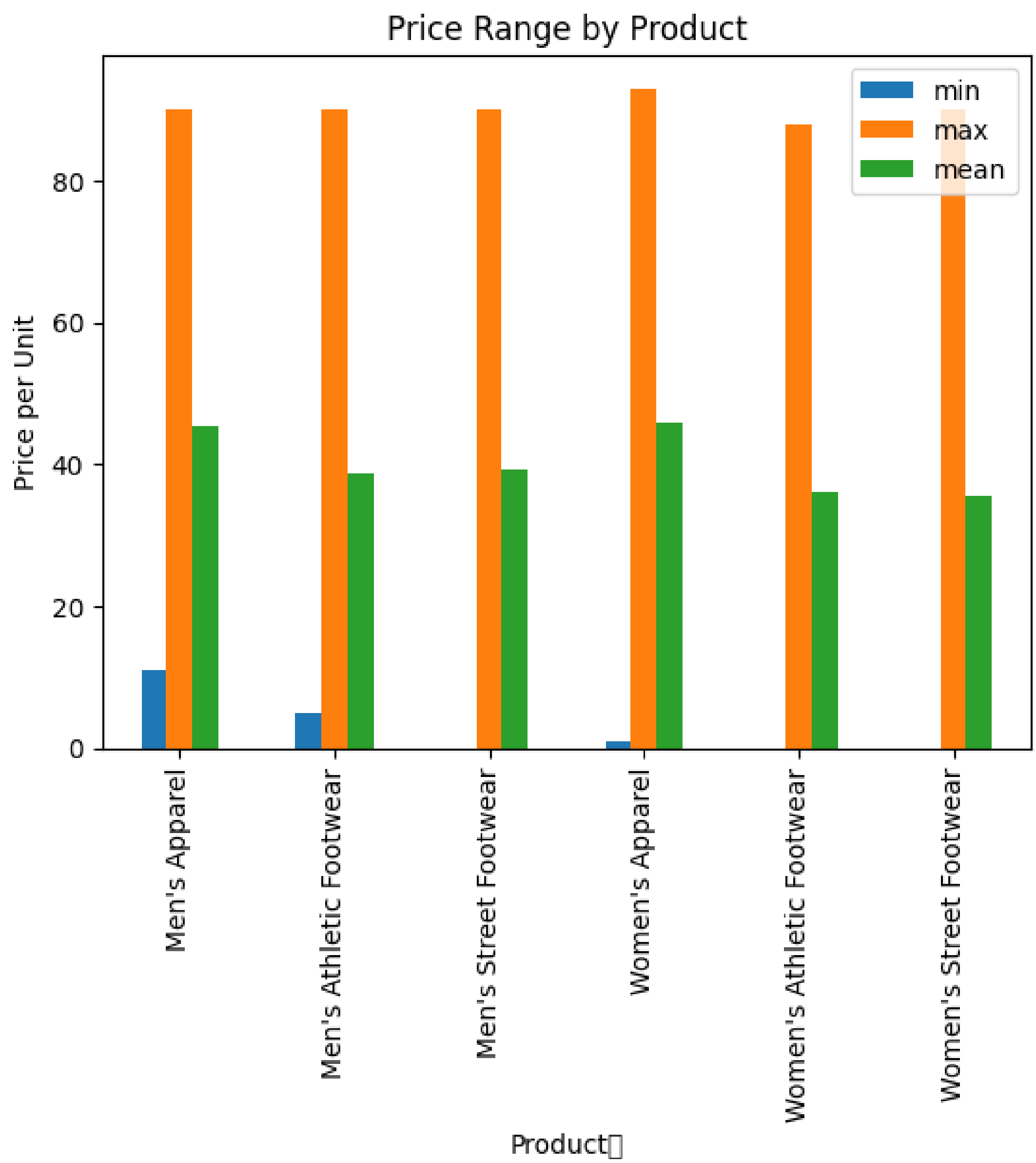
With a pink tint and the name "Midwest," the third-largest slice makes up 19.4% of the overall sales.
Yellow-labeled "South" is the fourth-largest slice, making up 17.9% of the overall sales.
With a dark blue tint and the label "Southeast," the smallest slice represents 12.7% of the total sales.
Finally, the pie figure illustrates that the West area.

```
top_5_regions = df['Region'].value_counts().head(5)
plt.pie(top_5_regions, labels=top_5_regions.index, autopct="%1.1f%%")
plt.title("Top 5 Regions by Total Sales")
plt.show()
```

6. How different product preform with their prices

Men's clothing, men's athletic footwear, men's street footwear, women's clothing, women's athletic footwear, and women's street footwear are the product categories listed on the x-axis of the below bar chart.
The price range is represented on the y-axis. Although there are markings for the minimum and maximum prices.
Every bar has three colored portions separated from one another. The lowest price for each product category is shown in the left most part, which is highlighted in blue. The range between the lowest and highest price is shown in the center portion, which is colored gray. The orange portion at the rightmost position indicates the highest price for every category of products.

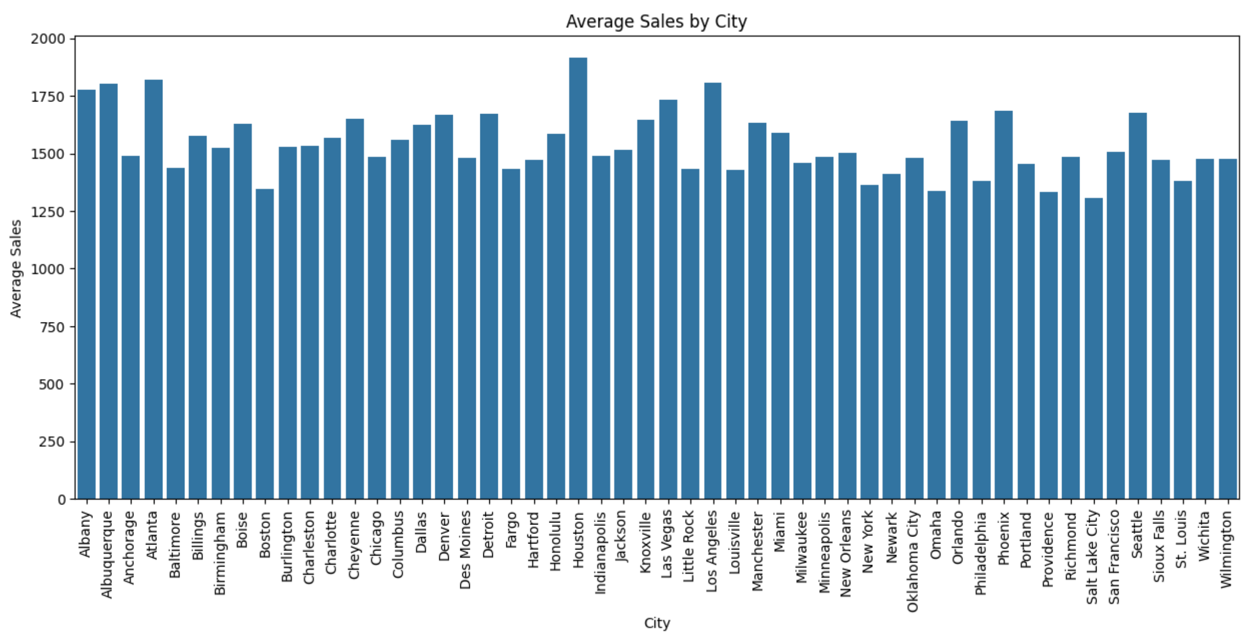
Men's athletic footwear appears to have the highest maximum price on the list, followed by women's athletic footwear, men's and women's clothing, men's and women's street footwear, and men's and women's clothes.




```
price_range_df = df[['Price per Unit', 'Product\t']]

grouped_df = price_range_df.groupby('Product\t')['Price per Unit'].agg(['min', 'max', 'mean'])

print(grouped_df)
```

```
[ ]

    grouped_df.plot(kind='bar', ylabel='Price per Unit', title='Price Range by Product')
    plt.xticks(rotation=90)
    plt.show()
```

7. Average total sales vs city



The average sales by city are displayed in this bar chart. The chart is titled "Average Sales by City".

The cities are listed on the chart's x-axis. I see that 40 cities are listed in alphabetical order, beginning with Albany and concluding with Wichita.
The average sales numbers are displayed on the y-axis; however, the scale is not indicated. Tick markers are located from 250 to 1750 along the axis.

The height of each bar in the graphic, which represents a city, relates to the average sales for that city in 2000. larger average sales are indicated by larger bars.

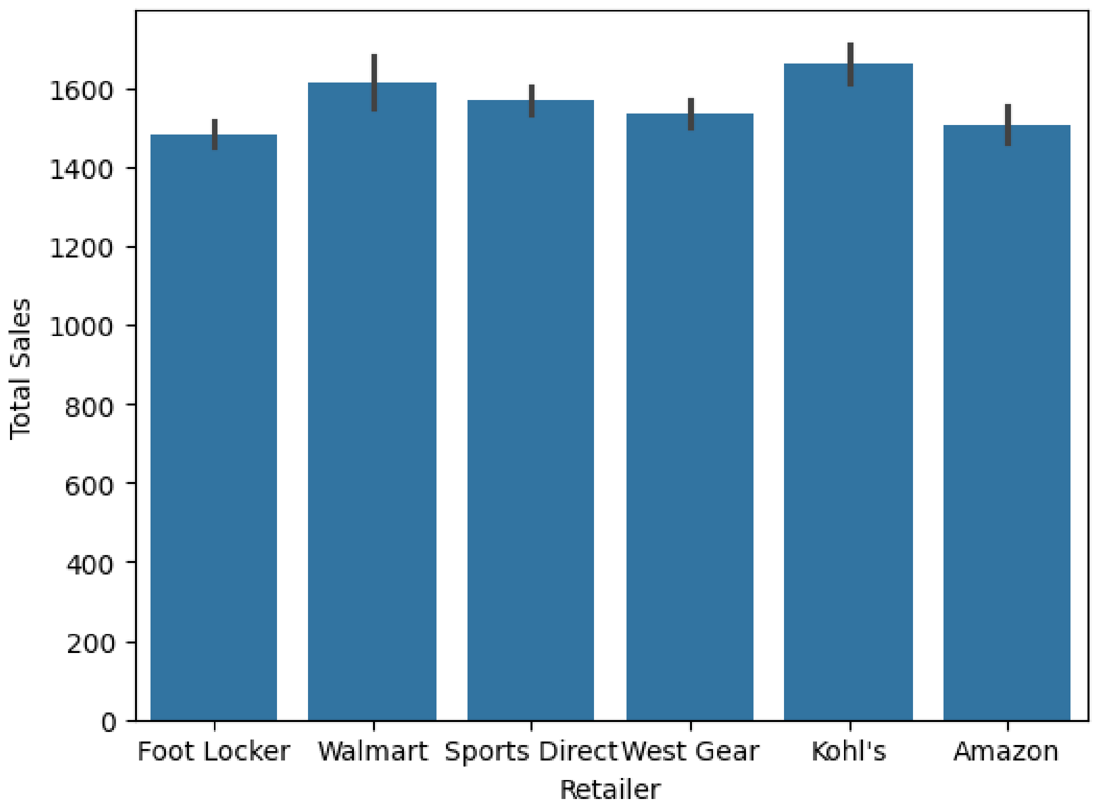
```
plt.figure(figsize=(15, 6))
sns.barplot(x='City', y='mean', data=grouped_df1)
plt.xticks(rotation=90)
plt.ylabel('Average Sales')
plt.xlabel('City')
plt.title('Average Sales by City')
plt.show()
```

III. Retailers Information

This section includes information about six major Adidas-affiliated stores, providing a detailed overview of each one's contributions to the sales environment.

1. Retailer wise total sales using a bar plot.

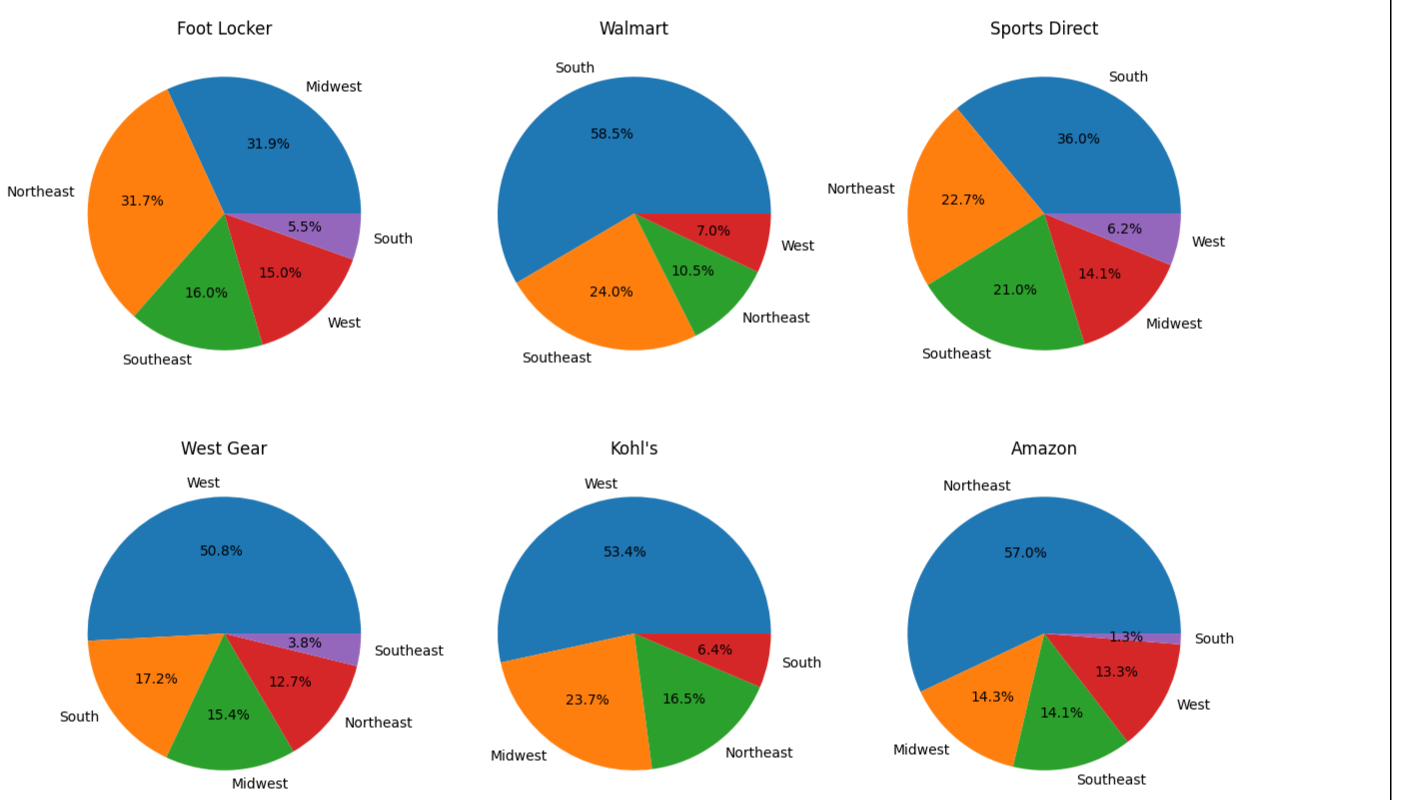
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.barplot(x='Retailer', y='Total Sales', data=df)
plt.show()
```



This graph shows the total sales by each Adidas retailer. By analyzing this graph, we can identify that "Kohl's" is the retailer which highest number of total sales followed by "Walmart" and "Sports Direct". On the other hand, we can see that "Foot Locker" retailer with the least number of total sales when compared to the other retailers.

2. Retailer performance region wise using pie charts.

```
import matplotlib.pyplot as plt
# Getting each retailer from 'retailer' column
retailers = df['Retailer'].unique()

# Create subplots
fig, axes = plt.subplots(2, 3, figsize=(15, 10))

# Iterate over retailers and plot pie charts
for i, retailer in enumerate(retailers):
# Filter data for current retailer
  df_retailer = df[df['Retailer'] == retailer]

# Get region-wise sales
  region_sales = df_retailer['Region'].value_counts()

# Plot pie chart
  axes[i // 3, i % 3].pie(region_sales, labels=region_sales.index, autopct="%1.1f%%")
  axes[i // 3, i % 3].set_title(retailer)

# Display plot
plt.show()
```



The above pie charts depict the sales performance of each retailer in each region. In this dataset, there are 5 different regions in which the retailers are operating namely Northeast, West, Southeast, South and Midwest.

Foot Locker: We can see that most of the sales are from Midwest region. But percentage of Northeast region is almost the same as Midwest because it only has a slight difference. We can see that Foot Locker has least number of sales in the South region which is a very low percentage compared to the other regions in which Foot Locker operates.

Walmart: It has a very high percentage of sales in the southern region which above 50%. So more than half of the sales of Walmart are from that region which is a dominating gap between Southern region and southeast region and is the second highest sales region of Walmart.

Sports Direct: According to the pie chart the highest sales of Sports Direct are from Southern region followed by Northeast and Southeast regions. We can see that it has not performed well in the West region which is a very low percentage compared to the other regions.

West Gear: Half of the sales of West Gear are from West region according to this pie chart. Second highest and third highest sales are from South and Midwest regions respectively. Sales of South and Midwest regions are relatively similar because the gap between the percentage is low. The other highlighting fact of this pie chart is that the sales of Southeast region is very low which is only 3.8%. So, we can see that it has not performed well in the Southeast region.

Kohl's: West region is the region with the highest number of sales for Kohl's followed by Midwest and Northeast regions. West region also has more than half the percentage of sales compared to the other regions in which Kohl's operates. The region with the lowest sales for Kohl's is South.

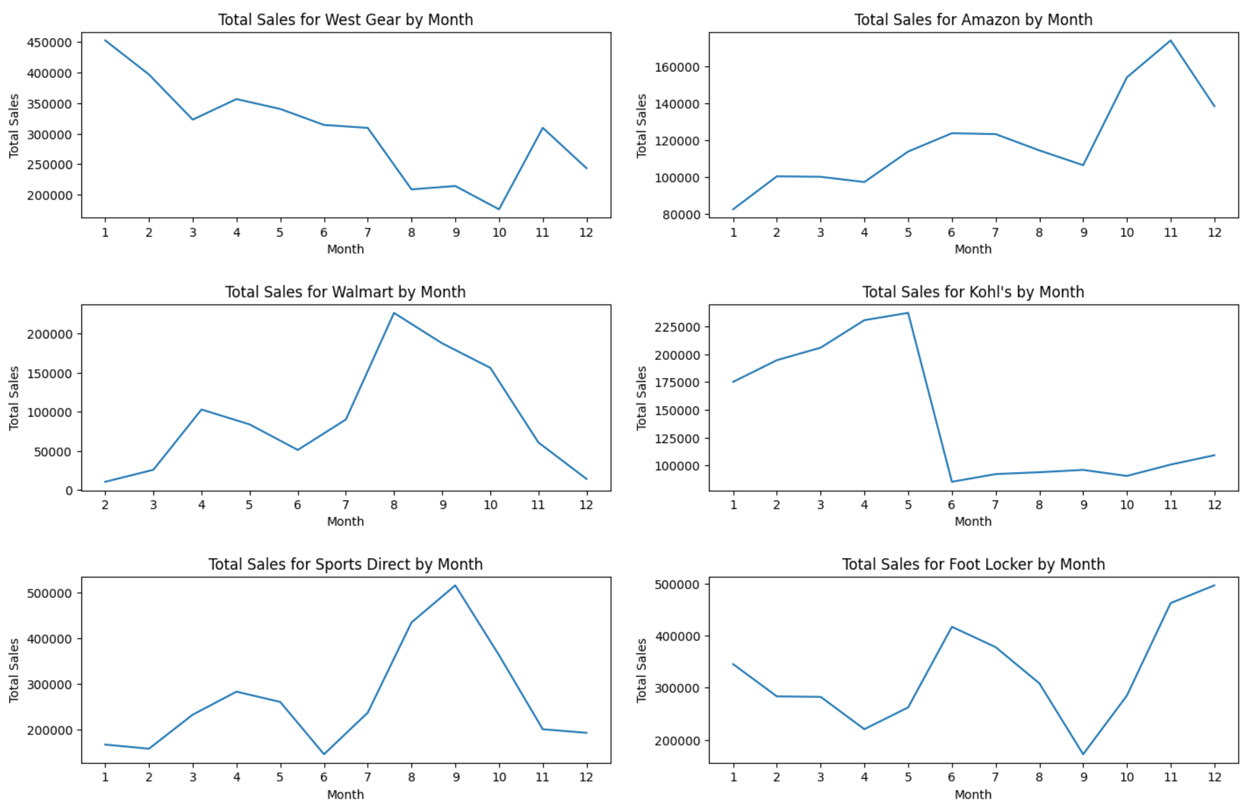
Amazon: Amazons' highest sales are from Northeast region which is a dominant feature of this pie chart with a percentage of 57%. Midwest and Southeast regions have almost similar amounts of percentages with a very slight difference between them. South region has the lowest percentage of sales with 1% which is also a highlighting fact in this pie chart.

3. Retailer sales by month using line plots.

```
import matplotlib.pyplot as plt
fig, axes = plt.subplots(3, 2, figsize=(15, 10))
fig.tight_layout(pad=5.0)

df_west_gear = df[df['Retailer'] == 'West Gear']
df_west_gear['Month'] = pd.to_datetime(df_west_gear['Invoice Date']).dt.month
df_west_gear_monthly_sales = df_west_gear.groupby('Month')['Total Sales'].sum()
axes[0, 0].plot(df_west_gear_monthly_sales.index, df_west_gear_monthly_sales.values)
axes[0, 0].set_xlabel('Month')
axes[0, 0].set_ylabel('Total Sales')
axes[0, 0].set_title('Total Sales for West Gear by Month')
axes[0, 0].set_xticks(df_west_gear_monthly_sales.index)

df_amazon = df[df['Retailer'] == 'Amazon']
df_amazon['Month'] = pd.to_datetime(df_amazon['Invoice Date']).dt.month
df_amazon_monthly_sales = df_amazon.groupby('Month')['Total Sales'].sum()
axes[0, 1].plot(df_amazon_monthly_sales.index, df_amazon_monthly_sales.values)
axes[0, 1].set_xlabel('Month')
axes[0, 1].set_ylabel('Total Sales')
axes[0, 1].set_title('Total Sales for Amazon by Month')
axes[0, 1].set_xticks(df_amazon_monthly_sales.index)
```

```
df_walmart = df[df['Retailer'] == 'Walmart']
df_walmart['Month'] = pd.to_datetime(df_walmart['Invoice Date']).dt.month
df_walmart_monthly_sales = df_walmart.groupby('Month')['Total Sales'].sum()
axes[1, 0].plot(df_walmart_monthly_sales.index, df_walmart_monthly_sales.values)
axes[1, 0].set_xlabel('Month')
axes[1, 0].set_ylabel('Total Sales')
axes[1, 0].set_title('Total Sales for Walmart by Month')
axes[1, 0].set_xticks(df_walmart_monthly_sales.index)

df_kohls = df[df['Retailer'] == "Kohl's"]
df_kohls['Month'] = pd.to_datetime(df_kohls['Invoice Date']).dt.month
df_kohls_monthly_sales = df_kohls.groupby('Month')['Total Sales'].sum()
axes[1, 1].plot(df_kohls_monthly_sales.index, df_kohls_monthly_sales.values)
axes[1, 1].set_xlabel('Month')
axes[1, 1].set_ylabel('Total Sales')
axes[1, 1].set_title("Total Sales for Kohl's by Month")
axes[1, 1].set_xticks(df_kohls_monthly_sales.index)

df_sports_direct = df[df['Retailer'] == 'Sports Direct']
df_sports_direct['Month'] = pd.to_datetime(df_sports_direct['Invoice Date']).dt.month
df_sports_direct_monthly_sales = df_sports_direct.groupby('Month')['Total Sales'].sum()
axes[2, 0].plot(df_sports_direct_monthly_sales.index, df_sports_direct_monthly_sales.values)
axes[2, 0].set_xlabel('Month')
axes[2, 0].set_ylabel('Total Sales')
axes[2, 0].set_title('Total Sales for Sports Direct by Month')
axes[2, 0].set_xticks(df_sports_direct_monthly_sales.index)
```

```
df_foot_locker = df[df['Retailer'] == 'Foot Locker']
df_foot_locker['Month'] = pd.to_datetime(df_foot_locker['Invoice Date']).dt.month
df_foot_locker_monthly_sales = df_foot_locker.groupby('Month')['Total Sales'].sum()
axes[2, 1].plot(df_foot_locker_monthly_sales.index, df_foot_locker_monthly_sales.values)
axes[2, 1].set_xlabel('Month')
axes[2, 1].set_ylabel('Total Sales')
axes[2, 1].set_title('Total Sales for Foot Locker by Month')
axes[2, 1].set_xticks(df_foot_locker_monthly_sales.index)

plt.show()
```



The above line plots depict the sales performance of each retailer in by month.

West Gear: In the line plot depicting the sales performance of West Gear, we can see that in the first month it has recorded the highest total sales, which is over 450000. But gradually it decreased over the following months. In the latter part of the line plot during the 10th and 11th month, there is a slight increase in sales but not as much as in the first month. This can be a main reason because it's the start of the holiday season.

Amazon: In the line plot depicting the sales performance of Amazon, we can see that in the first month it has recorded a very low amount of sales, but gradually it has increased over the next months. We can see a significant increase in the sales after the 9^{th} month and has recorded the highest number of total sales in the 11^{th} month which is over 165000.

Walmart: In the line plot depicting the sales performance of Walmart, we can see that in the first month it has recorded a very low amount of sales, but gradually it has increased over the next months specially total sales have significantly increased in the 3^{rd} and 4^{th} months while in the 8^{th} month Walmart has recorded the highest number of total sales which is over 200000.

Kohl's: In the line plot depicting the sales performance of Kohl's, we can see that it has started to increase the total number of sales from 175000 and has reached the highest total number of sales which is over 225000 in the 5^{th} month. Then onwards a significant downfall of sales can be seen which has dropped below 110000. This is a highlighting fact because this has happened suddenly within a month and has never really being able to maintain a good amount of total number of sales in the coming months.

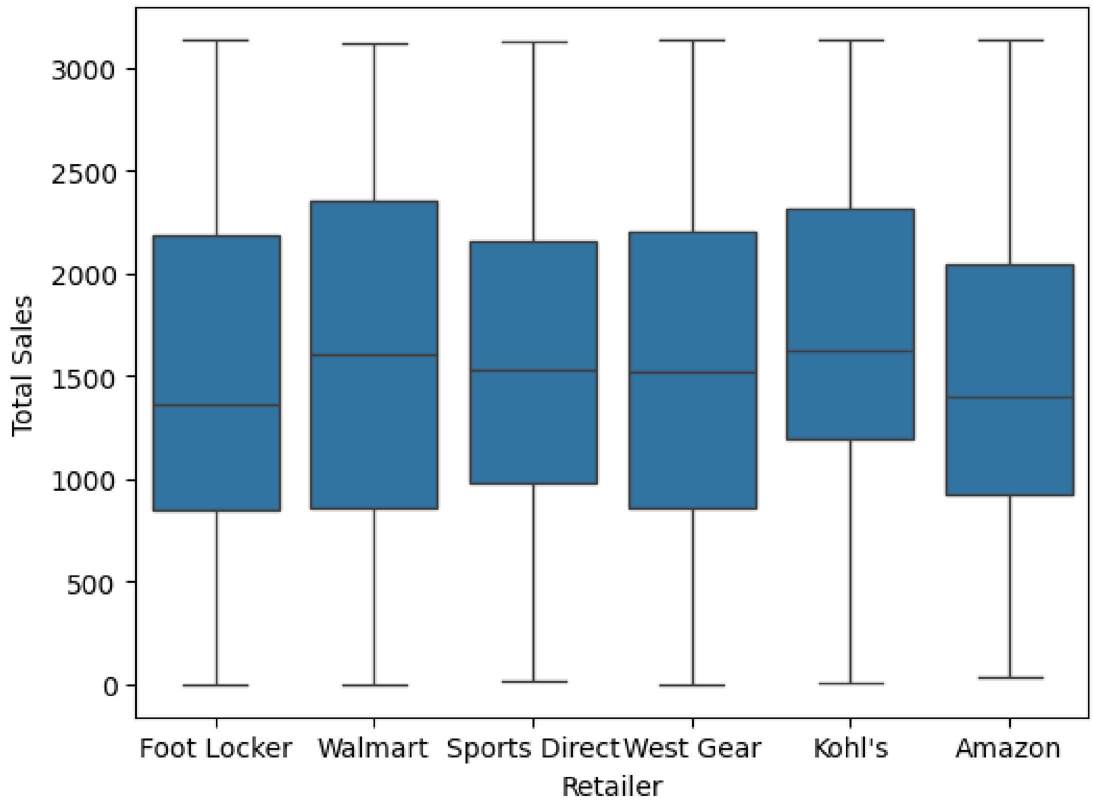
Sports Direct: In the line plot depicting the sales performance of Sports Direct, we can see a slight increase in sales in the first few months of the year and then after the 5^{th} month, we can see a consistent increase in the total sales. This has reached the highest total sales in the 9^{th} month which is over 500000 but in the latter part of the year it has decreased gradually.

Amazon: In the line plot depicting the sales performance of Amazon, we can see that in the first month it has recorded a very low amount of sales, but gradually it has increased over the next months. We can see a significant increase in the sales after the 9^{th} month and has recorded the highest number of total sales in the 11^{th} month which is over 165000.

Foot Locker: In the line plot depicting the sales performance of Foot Locker, we can see that after the 4^{th} month it has slightly increased and then by the 9^{th} month it has recorded the lowest total sales which is below 200000. But afterwards we can see a continuous increase in total sales in the latter part of the year with the holiday season and also it has recorded the highest total sales over 500000.

4. Retailer wise performance using boxplots.

```
sns.boxplot(x='Retailer', y='Total Sales', data=df)
plt.show()
```



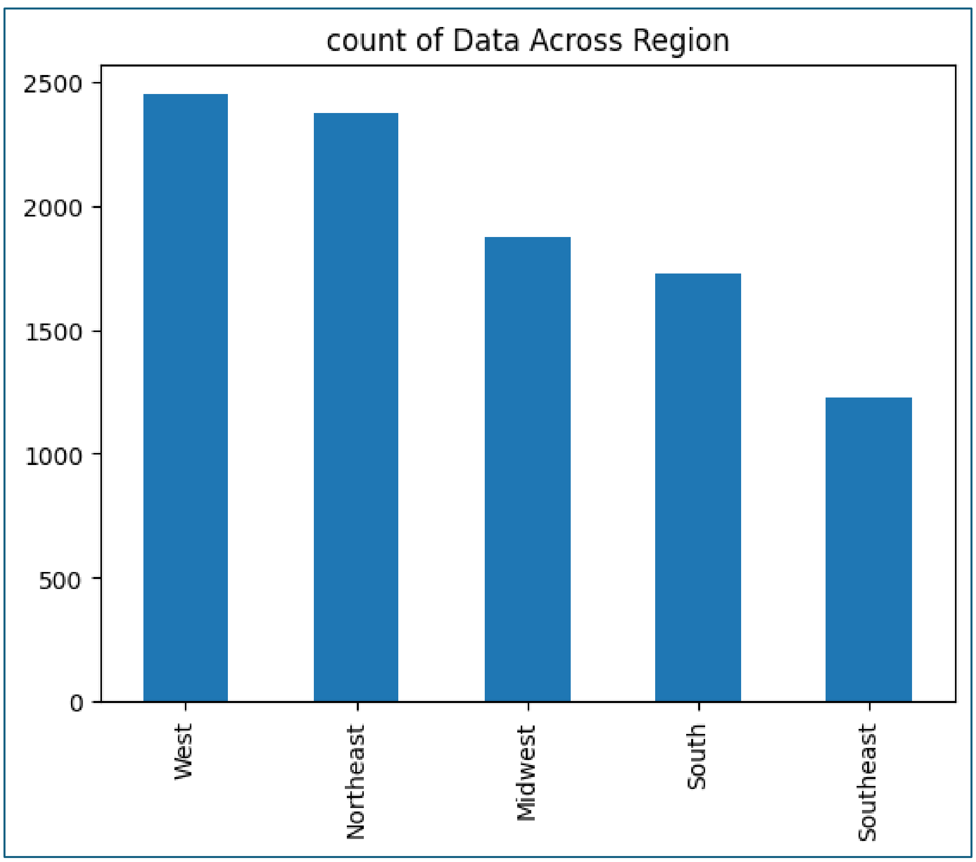
From the above boxplots, we can understand that Walmart has the highest spread or variability when compared to the other retailers. The highest median number of sales has been maintained by Kohl’s and Walmart, which is over 1500.

IV. Location Information:

Comprehensive geographic insights are offered, encompassing data on states, cities, and regions. This division makes it easier to investigate local sales patterns and customer inclinations.

1. Count plot of region column

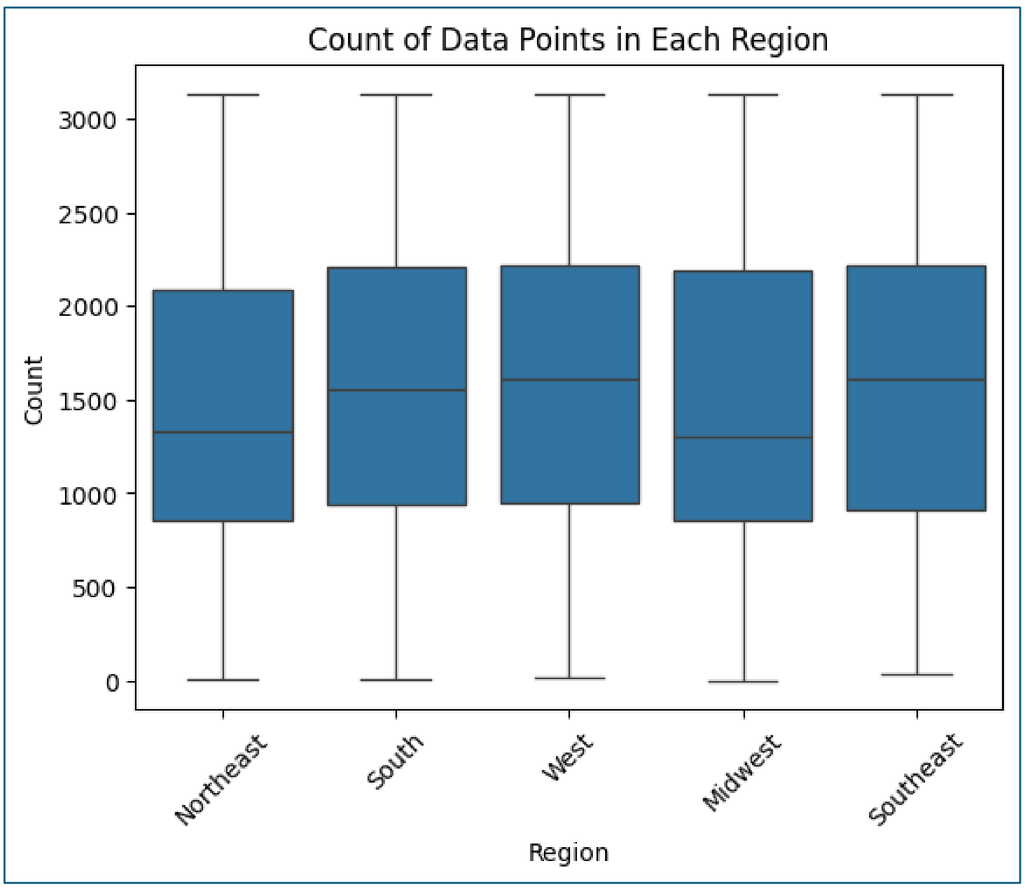
```
region_counts = df['Region'].value_counts()
region_counts.plot(kind='bar')
plt.title('count of Data Across Region')
plt.ylabel('')
plt.show()
```



The resulting bar plot displays the number of data points connected to each region on the y-axis and each region on the x- axis. The height of each bar indicates how many data points from that specific region there are in the dataset. This visualization shows that “West” region has more, and “Southeast” have less data points, giving a brief summary of the distribution of data across various regions. It can be useful for figuring out the regional distribution of the sales dataset.

2. <u>Boxplot of each region</u>

```
sns.boxplot(data=df, x='Region' , y='Total Sales')
plt.title('Count of Data Points in Each Region')
plt.xlabel('Region')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



The distribution of total sales across several regions can be compared according to this visualization. It facilitates the identification of regional variations in sales performances, such as variances in median sales, spread (variability) , and the number of outliers. Such insights can inform decision-making process related to resource allocation, marketing strategies, and regional targeting.

The breakdown of the boxplot to understand what it reveals about sales in different regions.

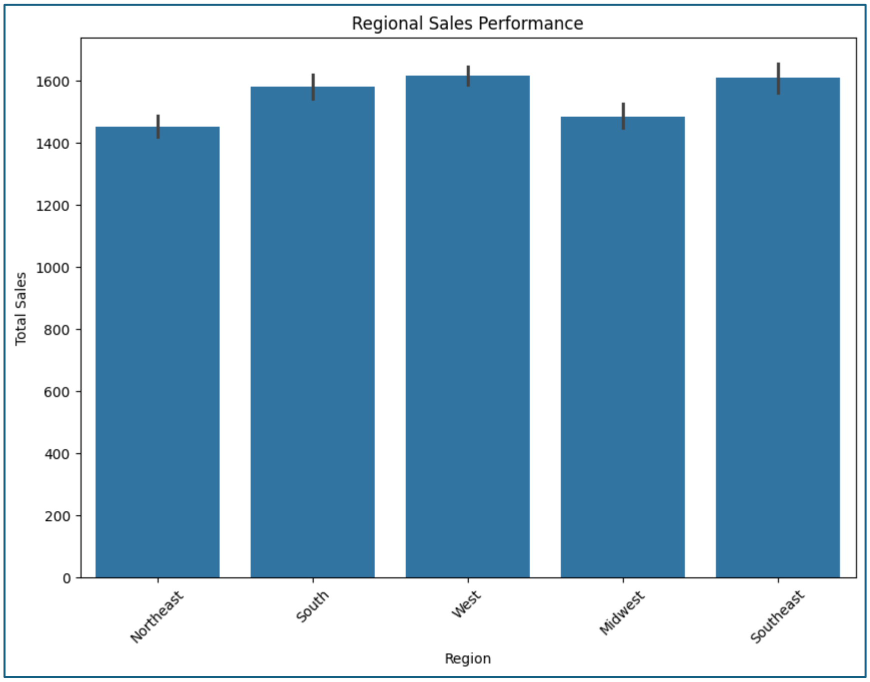
Center line: The median sales for that region are shown by the line in the center of each box. Since the median represents the 50^{th} percentile, half of the sales data in that area are below and half are above it. The median sales amount for the south region is around 1600.

Box: The interquartile range is shown by the box(IQR). The range containing the middle 50% of the data points is known as the IQR. The center data spread is shown by the height of the box. A large box indicates a wider spread in sales figures and a small box indicates that the sales figures are more tightly clustered around the median. The southeast region's box is the largest, indicating a greater variation in sales data there than in the other regions.

Outliers: Outliers are data points that fall outside the whiskers. They can represent unusually high or low sales figures in a practical region. If we get the "Northeast" region as a example there isn't clear outliers for the Northeast region.

3. Regional Sales Performance

```
plt.figure(figsize=(10, 7))
sns.barplot(data=df, x='Region' , y='Total Sales')
plt.title('Regional Sales Performance')
plt.xlabel('Region')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.show()
```

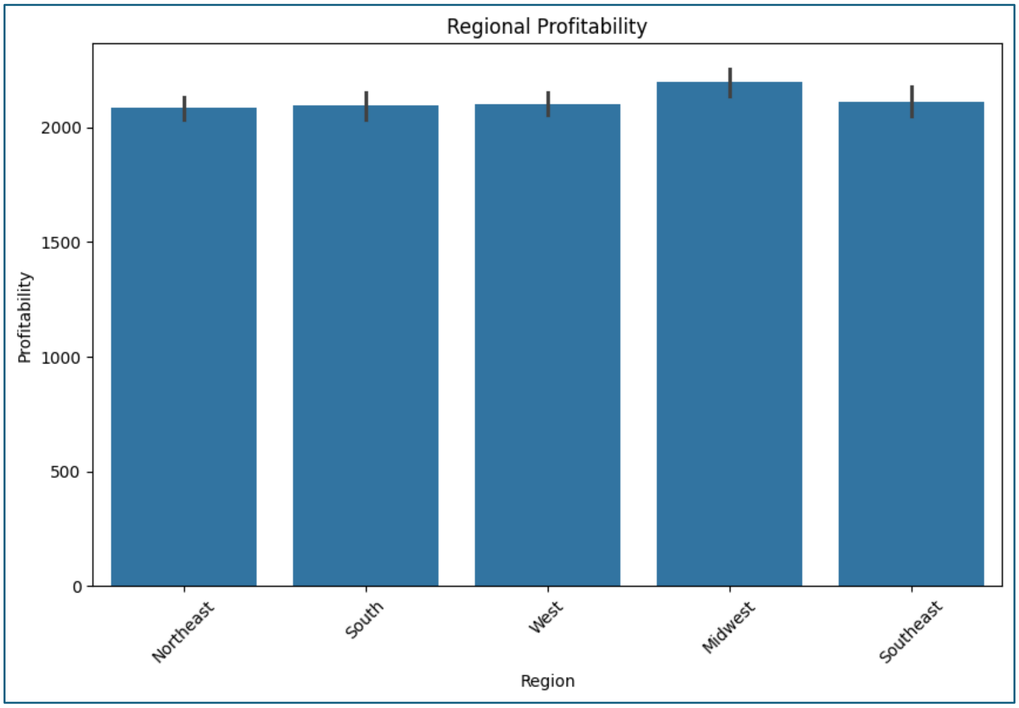


Due to its representation of various geographic locations where sales take place, the “Region” column can have a substantial impact on sales performance. Regional differences in market demand, consumer behavior, economic conditions and demography can all have an impact on sales level.

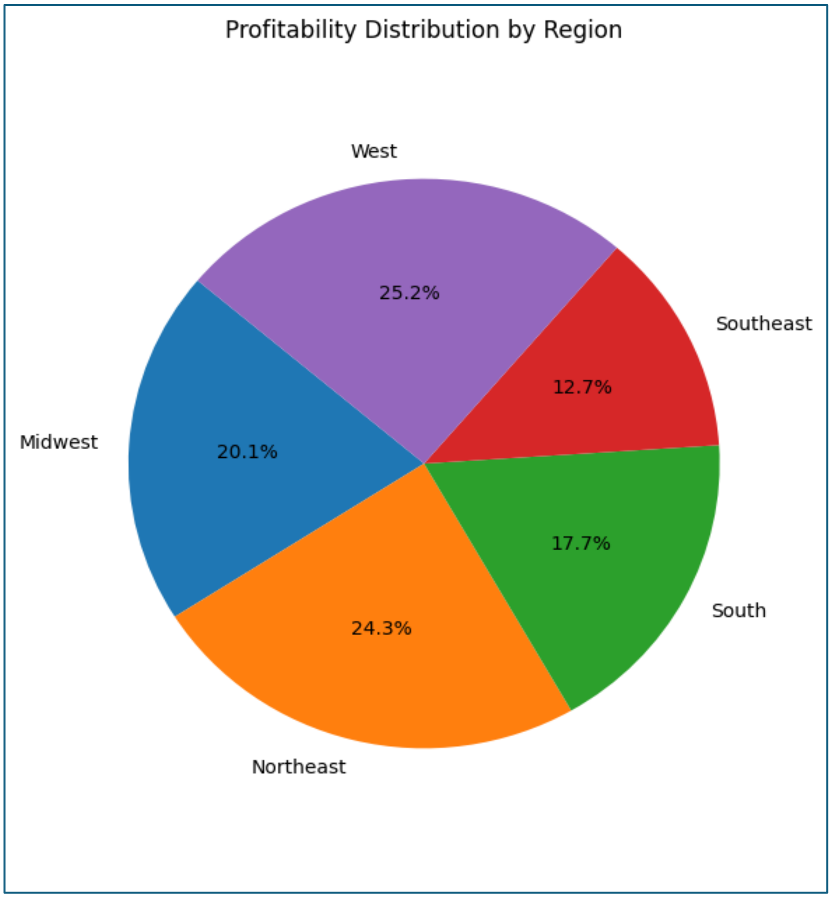
The bar graph shows the number of sales in different regions of the world. The horizontal axis shows from left to right there are Northeast, South, West, Midwest and Southeast. The vertical axis of the chart shows the number of sales. The scale goes from 0 to 1600. This plot can be used to visualize sales performances across different regions. This visualization can help identify regions with high and low sales performances. According to the plot west region is showing that high count of sales than in any other region and low count of sales shows in region southeast.

4. Regional Profitability

```
[ ] plt.figure(figsize=(10, 6))
    sns.barplot(data=df,x='Region' , y='Operating Profit')
    plt.title('Regional Profitability')
    plt.xlabel('Region')
    plt.ylabel('Profitability')
    plt.xticks(rotation=45)
    plt.show()
```




```
region_profitability = df.groupby('Region')['Operating Profit'].sum()

plt.figure(figsize=(6, 8))
plt.pie(region_profitability, labels=region_profitability.index, autopct='%1.1f%%', startangle=140)
plt.title('Profitability Distribution by Region')
plt.axis('equal')
plt.show()
```

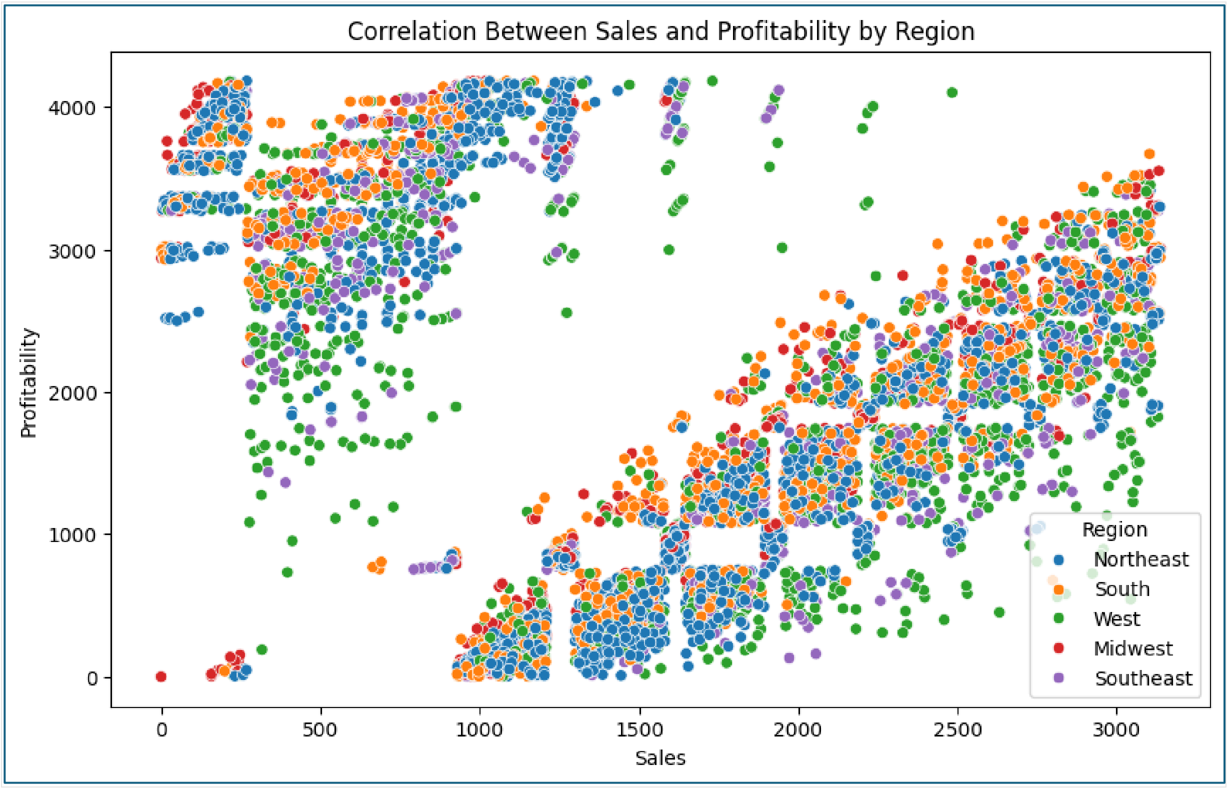


Similar to this “region” column, which shows the geographic distribution of sales and business operation, can have an impact on profitability. Differentiating costs, pricing tactics, market dynamics are some of the factors that might affect profitability levels indifferent geographic areas.

By using the above bar plot we can get the highest profitability region in all the relevant regions, that is Midwest. It’s more than 2000. Northeast, South and West have similar higher bars. And using the pie chart we can visualize profitability by region, showing the proportion of total profitability contributed by region. 25.2% amount shows high profitability, which is in west region.

5. Correlation between Sales and Profitability By Region

```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df, x='Total Sales', y='Operating Profit', hue='Region')
plt.title('Correlation Between Sales and Profitability by Region')
plt.xlabel('Sales')
plt.ylabel('Profitability')
plt.legend(title='Region')
plt.show()
```

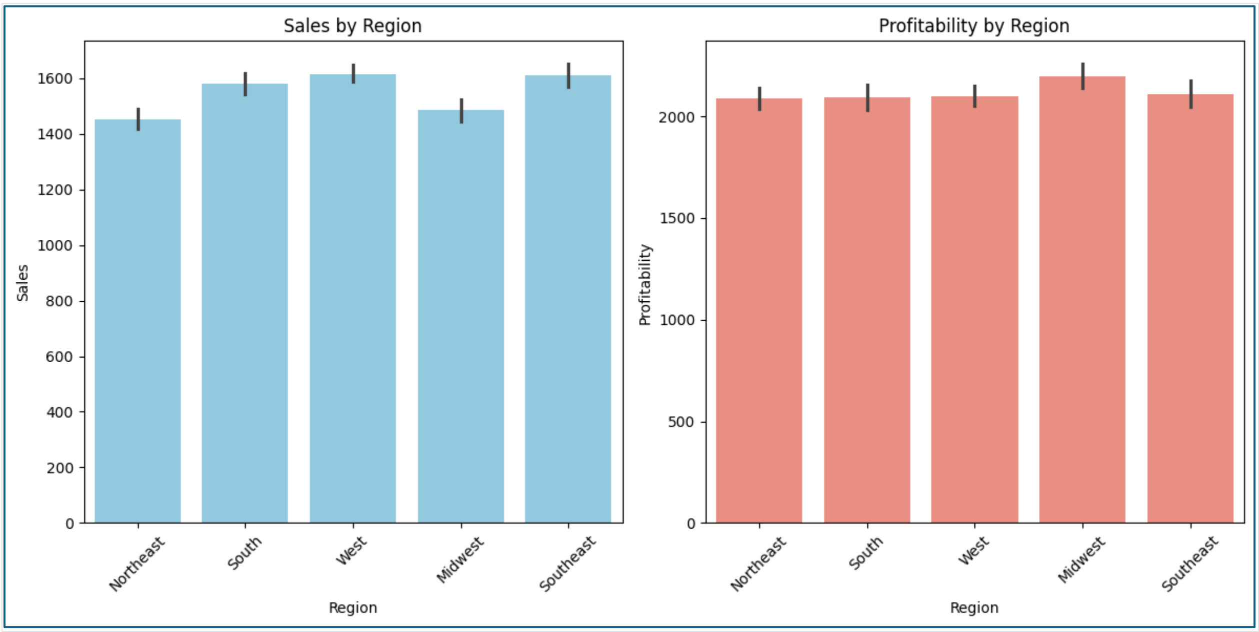


Examining the correlation between sales and profitability in every area can yield valuable information about the efficiency of sales tactics and operational efficiency. Because profitability is impacted by a number of variables, including cost structure, price and discounts, increasing sales volume does not automatically equate to higher profitability.

Based on the above scatter plot, cannot really identify if whether there is a positive, negative or no correlation between sales and profitability by region.

6. Regional Comparison and Benchmarking

```python
plt.figure(figsize=(12, 6))

# Plotting sales
plt.subplot(1, 2, 1)
sns.barplot(data=df, x='Region', y='Total Sales', color='skyblue')
plt.title('Sales by Region')
plt.xlabel('Region')
plt.ylabel('Sales')
plt.xticks(rotation=45)

# Plotting profitability
plt.subplot(1, 2, 2)
sns.barplot(data=df, x='Region', y='Operating Profit', color='salmon')
plt.title('Profitability by Region')
plt.xlabel('Region')
plt.ylabel('Profitability')
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```



Comparing and measuring different region's sales and profitability makes it easier to determine which ones are performing better or worse than expected. Strategies to enhance performance in other regions can be informed by an understanding of the factors that contribute to success in high-performing regions.

The ability to compare sales and profitability side by side makes it simple and quick to asses performance across several regions.

Here are some things that we can get from this visualization:

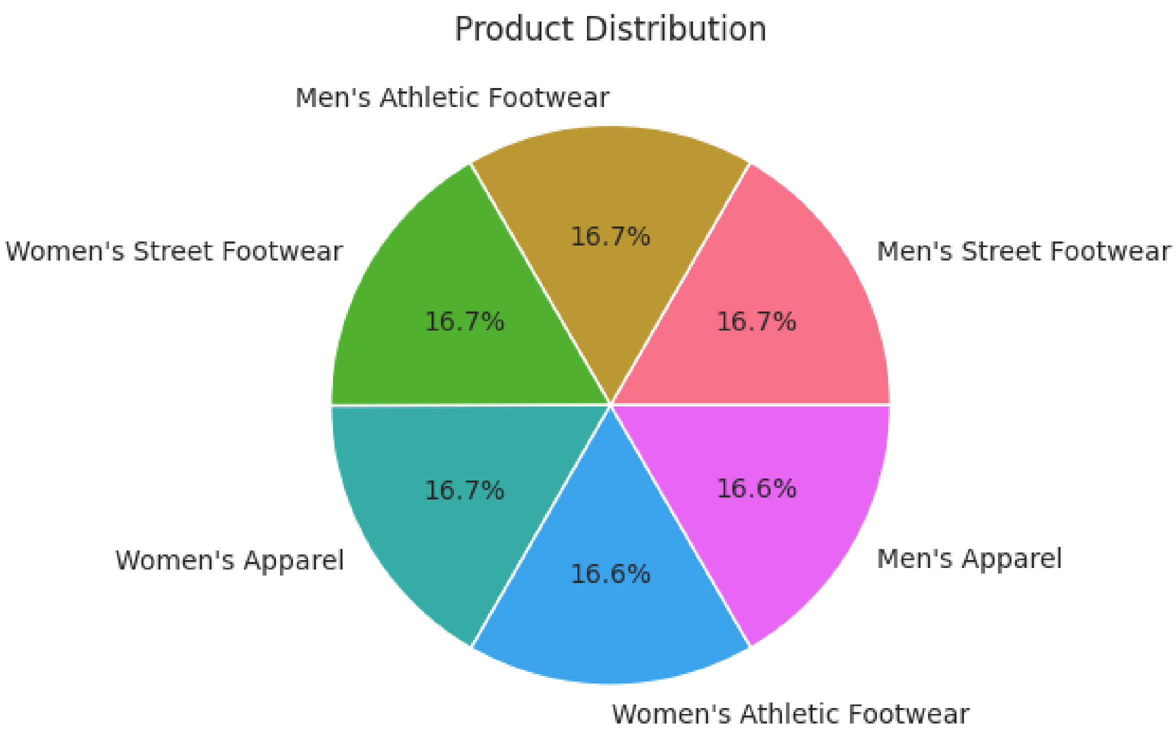
Identify high performing regions: According to the example, the southeast and West have the most sales and profitability. These areas may be in the lead in a few areas that affect performance in general.

Identify underperforming regions: the graph indicates the region with the lowest sales and profitability appears to be the Northeast.

Benchmarking: by understanding what factors are driving success in the high performing regions (Southeast and West), strategies can be implemented to improve performance in the other region (Northeast).

V. Product Information

1. Product Distribution Pie Chart



The distribution of various shoe and clothing types is depicted in the pie chart. This is an explanation of the data in the chart:

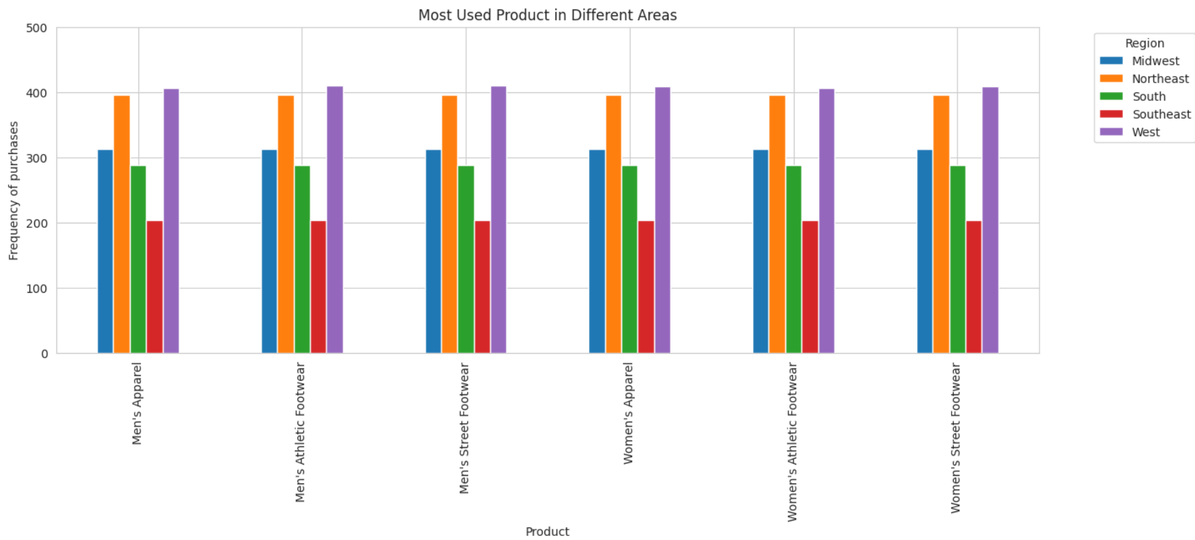
- There are six slices in the pie chart, each of which represents a distinct product category.
- The two biggest slices, accounting for 16.7% of the pie, are for men's and women's street shoes.
- The next four slices, which make up 16.6% of the pie each, are for men's, women's, and women's athletic footwear, clothing, and accessories.

All things considered, we can state that the categories of street footwear for men and women each account for a greater share of the pie chart than any other. This implies that in this specific data set, these shoe types might be more popular than athletic footwear or clothing.

```
product_counts = df['Product\t'].value_counts()
```

```
labels = product_counts.index.to_list()
colors = sns.color_palette('husl', len(labels))

plt.pie(product_counts, labels=labels, autopct='%1.1f%%', colors=colors)
plt.title('Product Distribution')
plt.show
```

2. Most Used Products by Region Stacked Bar Plots

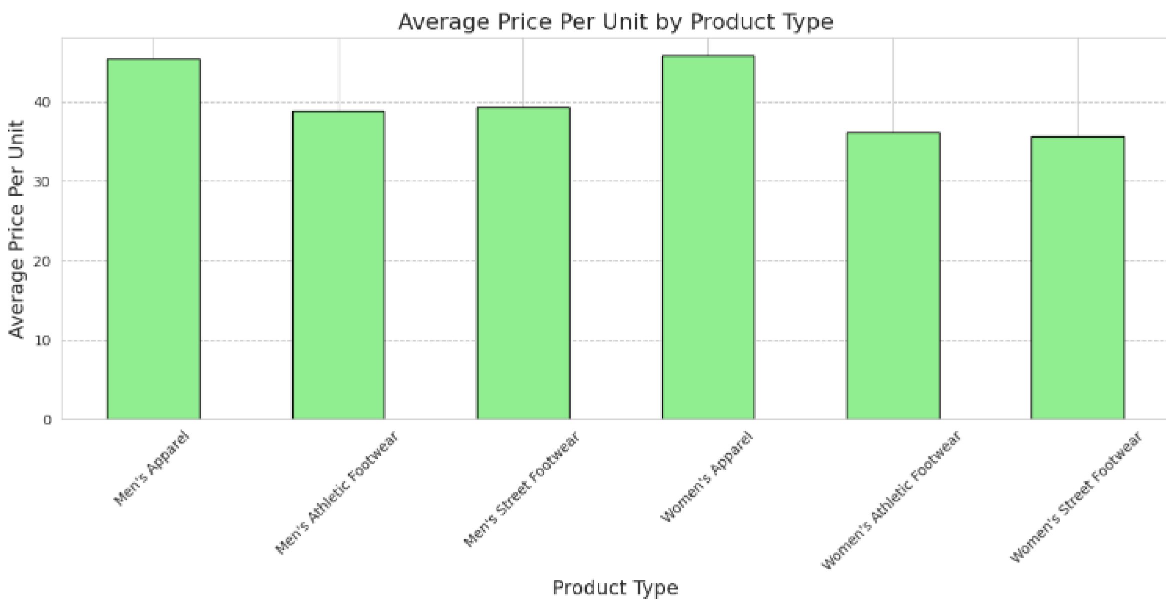


- Men's Apparel: The Midwest has the highest frequency of purchases in this category, suggesting that men's apparel is the most popular product overall.
- Women's Apparel: This category appears to be the second most popular overall, with the Midwest having the highest frequency of purchases.
- Sports Footwear for Men: Based on the graph, it seems that Sports Footwear for Men could be the third most popular product overall. Purchases of this category appear to be most common in the Midwest and Southeast.
- Women's Athletic Footwear: Compared to the West, the Midwest, Northeast, and Southeast seem to be more popular places for women to wear athletic footwear.
- Street Footwear: Due to the stacked bars, it is challenging to determine with certainty from this graph which style of street footwear (men's or women's) is most popular.

All things considered, we can state that, statewide, athletic clothing and footwear appear to be more in demand than street footwear. When it comes to street footwear purchases, the Midwest and Southeast seem to make more of them than the West.

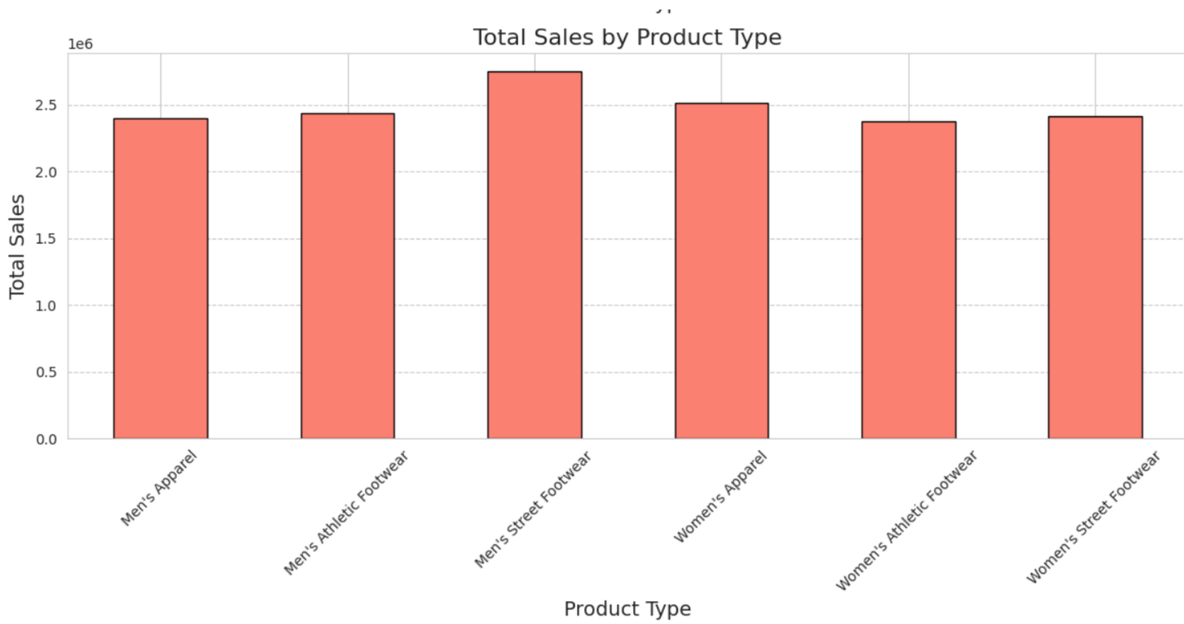
```
product_region_crosstab = pd.crosstab(df['Product\t'], df['Region'])

# Plot the crosstab as a bar chart
product_region_crosstab.plot(kind="bar", figsize=(15, 5))

# Increase the spacing between the bars
plt.ylim(bottom=0, top=500)

# Add title and labels
plt.title('Most Used Product in Different Areas')
plt.xlabel('Product')
plt.ylabel('Frequency')

# Show the plot
plt.show()
```

3. Bar Plots for Average Price Per Unit and Total Sales By product



Here are some details regarding the typical cost per item for various categories of apparel and footwear:

- Comparable Prices: The average price per unit for all product types is in the close range of $10 to $40.
- Most Expensive: At about $40 per unit, men's athletic footwear is the priciest product.
- The least expensive products are those for men and women, with each unit costing about $10.

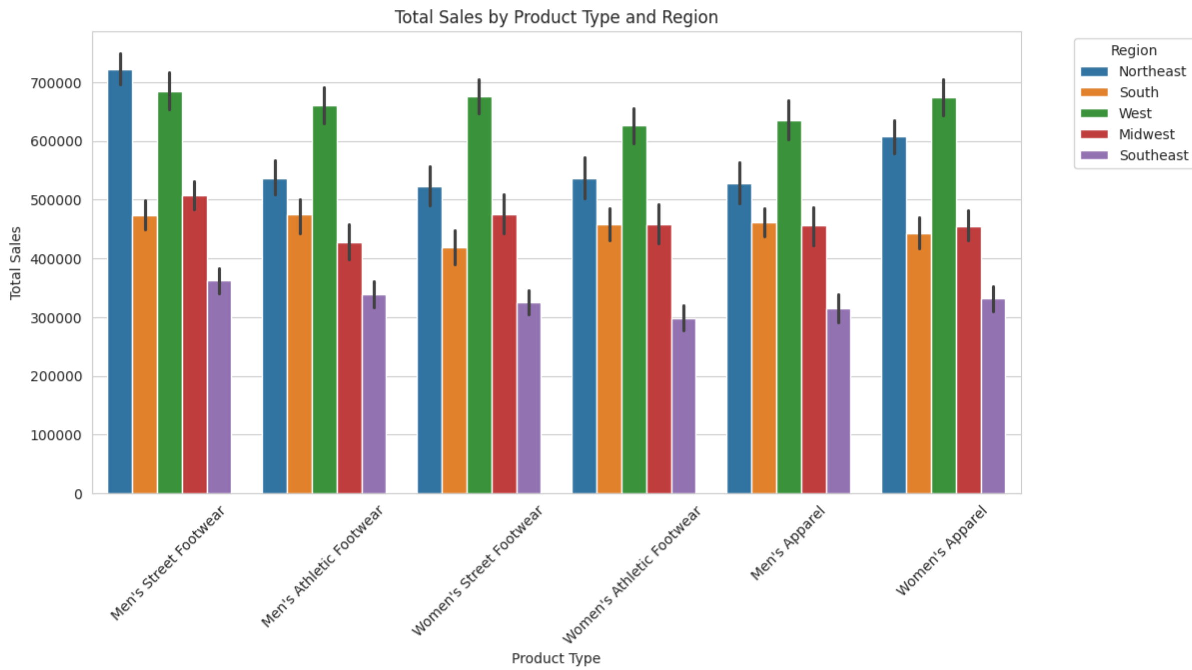


- The graph indicates that the two most popular product categories are women's apparel and men's street footwear, with women's street footwear coming in third.
- Products with the lowest popularity: Based on the graph, women's athletic footwear seems to be the least popular category of products.

```
import seaborn as sns
import matplotlib.pyplot as plt

# Set the figure size
plt.figure(figsize=(12, 6))

# Set style
sns.set_style("whitegrid")

# Create subplots
fig, axes = plt.subplots(3, 1, figsize=(12, 18))

# Plot for Product Counts
product_counts.plot(kind='bar', ax=axes[0], color='skyblue', edgecolor='black')
axes[0].set_title('Number of Products Sold by Product Type', fontsize=16)
```

```
axes[0].set_xlabel('Product Type', fontsize=14)
axes[0].set_ylabel('Number of Products Sold', fontsize=14)
axes[0].tick_params(axis='x', labelrotation=45)
axes[0].grid(axis='y', linestyle='--')

# Plot for Average Price Per Unit
average_price_per_unit.plot(kind='bar', ax=axes[1], color='lightgreen', edgecolor='black')
axes[1].set_title('Average Price Per Unit by Product Type', fontsize=16)
axes[1].set_xlabel('Product Type', fontsize=14)
axes[1].set_ylabel('Average Price Per Unit', fontsize=14)
axes[1].tick_params(axis='x', labelrotation=45)
axes[1].grid(axis='y', linestyle='--')

# Plot for Total Sales
total_sales.plot(kind='bar', ax=axes[2], color='salmon', edgecolor='black')
axes[2].set_title('Total Sales by Product Type', fontsize=16)
axes[2].set_xlabel('Product Type', fontsize=14)
axes[2].set_ylabel('Total Sales', fontsize=14)
axes[2].tick_params(axis='x', labelrotation=45)
axes[2].grid(axis='y', linestyle='--')

plt.tight_layout()
plt.show()
```

4. Stacked Bar plot for Total Sales by Product Type and Region



Sales of Each Product Type:

- Athletic Footwear: It appears that across all regions, men's athletic footwear is the most popular style, closely followed by women's athletic footwear.

- Street Footwear: In the South and West, sales of women's street footwear are comparable to those of men's athletic footwear, suggesting that it is the third most popular style of footwear.

- Sales of street footwear for men seem to be lower than sales of any other kind of footwear in every region.

Sales by Region:

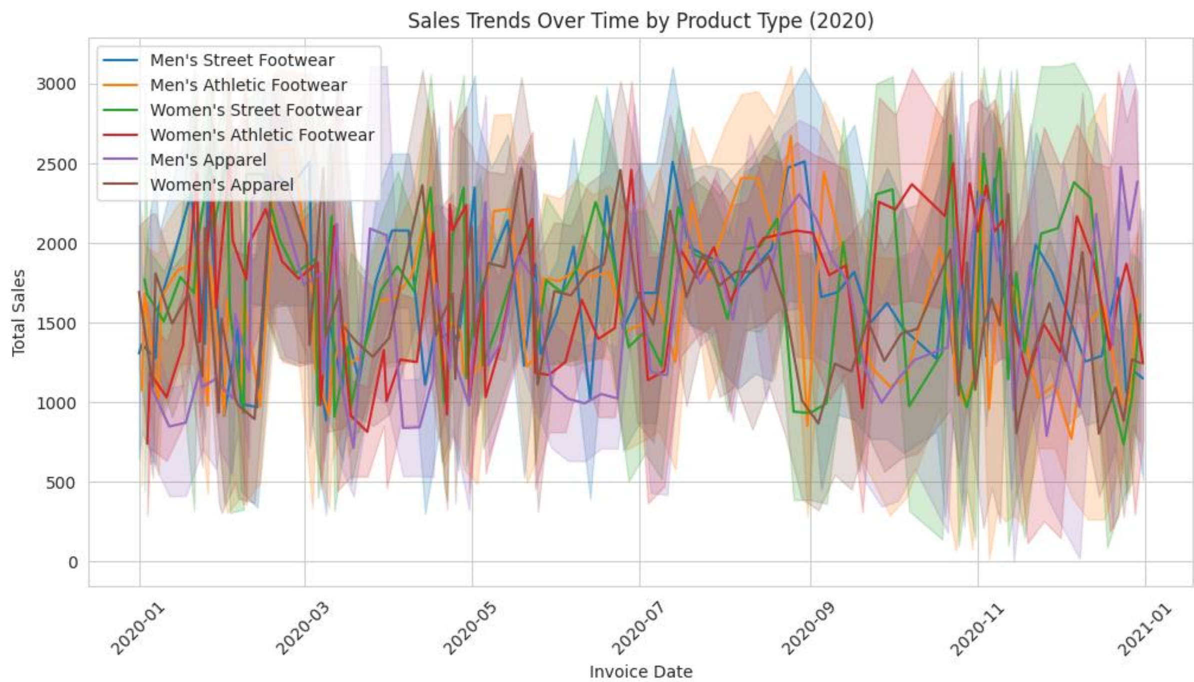
- Based on available data, it appears that the Northeast region has the highest sales of athletic footwear for both men and women.

- In comparison to the Northeast and Midwest, the South and West seem to have lower sales figures for athletic footwear.

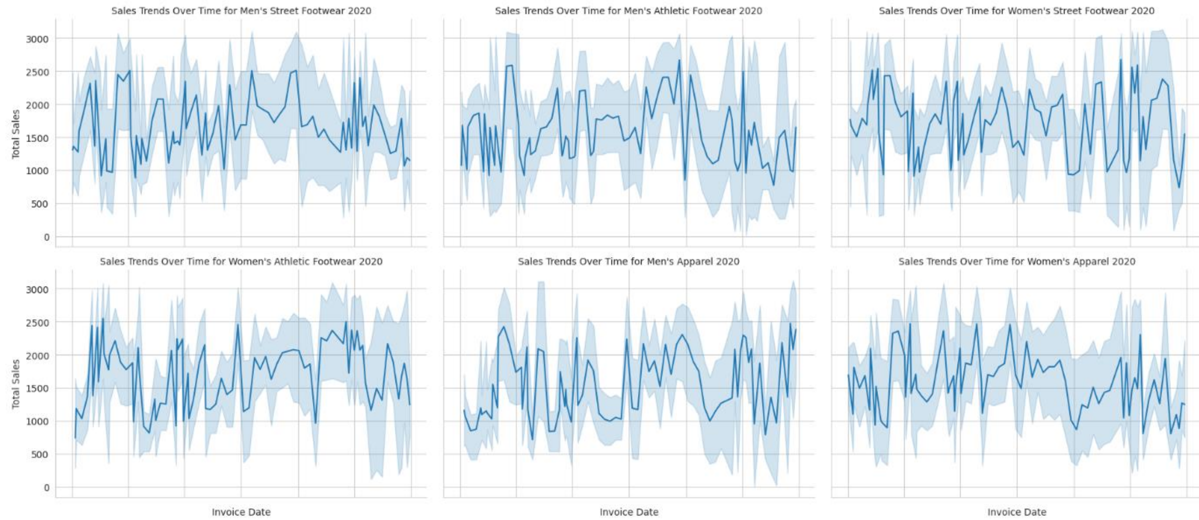
- The number of sales of women's street shoes seems to be fairly constant throughout all regions.

```
#Stacked Barplot of Total Sales by Product Type and Region:
plt.figure(figsize=(12, 6))
sns.barplot(x='Product\t', y='Total Sales', hue='Region', data=df, estimator=sum)
plt.title('Total Sales by Product Type and Region')
plt.xlabel('Product Type')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.legend(title='Region', loc='upper right')
plt.show()
```

5. Line Plot for Sales Trends Over Time By Product Type 2020




```
df_2020 = df[df['Invoice Date'].dt.year == 2020]

plt.figure(figsize=(12, 6))
sns.lineplot(x='Invoice Date', y='Total Sales', hue='Product\t', data=df_2020)
plt.title('Sales Trends Over Time by Product Type (2020)')
plt.xlabel('Invoice Date')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.legend(loc='upper left')
plt.show()
```



Here are some insights into the sales trends of various product types in 2020, based on the "Sales Trends Over Time for Different Product Types (2020)":

Overall Trends:

For the majority of product categories, sales seem to have increased throughout the year. Sales of athletic footwear for women increased at the fastest rate, followed by those of women's clothing and men's athletic footwear.

Declining Sales:

Men's Street Footwear is the only product category that seems to be experiencing a decline in sales over time.

Comparative Analysis by Product Type:

- In January and February of 2020, the most popular product category was Men's Street Footwear.
- But in March 2020, women's athletic footwear overtook men's street footwear in terms of sales, and it stayed at the top for the remainder of the year.

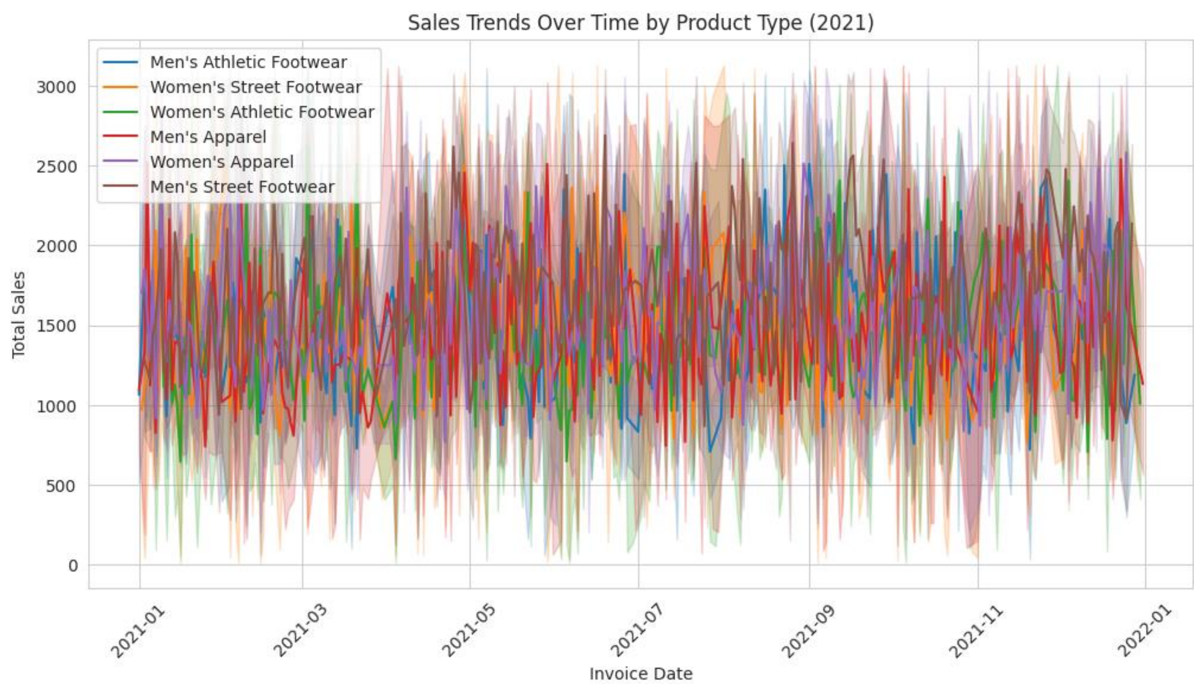
- Although women's apparel sales increased gradually throughout the year, they fell short of men's or women's athletic footwear sales.

- Sales of men's athletic footwear rose all year long, and by November 2020, they had overtaken those of men's street footwear and men's apparel.

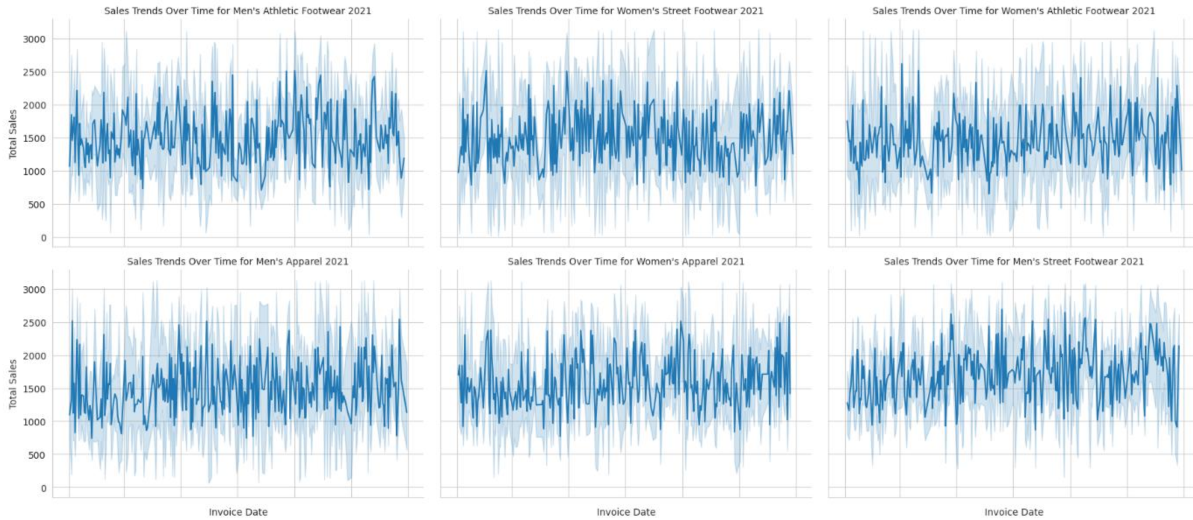
- Sales of men's apparel stayed mostly unchanged in 2020.

```
import seaborn as sns
import matplotlib.pyplot as plt

# Filter data for the year 2020
df_2020 = df[df['Invoice Date'].dt.year == 2020]

# Get unique product types
product_types = ["Men's Street Footwear", "Men's Athletic Footwear",
                 "Women's Street Footwear", "Women's Athletic Footwear",
                 "Men's Apparel", "Women's Apparel"]

# Create separate line plots for each product type
g = sns.FacetGrid(df_2020, col='Product\t', col_wrap=3, height=4, aspect=1.5)
g.map(sns.lineplot, 'Invoice Date', 'Total Sales')

# Set titles and labels
g.set_titles("Sales Trends Over Time for {col_name}")
g.set_axis_labels("Invoice Date", "Total Sales")
g.set_xticklabels(rotation=45)

plt.show()
```

6. Line Plot for Sales Trends Over Time By Product Type 2021




```
df_2021 = df[df['Invoice Date'].dt.year == 2021]

plt.figure(figsize=(12, 6))
sns.lineplot(x='Invoice Date', y='Total Sales', hue='Product\t', data=df_2021)
plt.title('Sales Trends Over Time by Product Type (2021)')
plt.xlabel('Invoice Date')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.legend(loc='upper left')
plt.show()
```



Here are some insights into the sales trends of various product types in 2020, based on the "Sales Trends Over Time for Different Product Types (2021)":

Overall Trends:
It seems that most shoe types saw an increase in sales throughout the course of the year. Sales of athletic footwear for women increased at the fastest rate, with street footwear for women and men's athletic footwear following suit.

Diminished Sales:
It seems that the only product category experiencing a decline in sales over time is men's clothing.

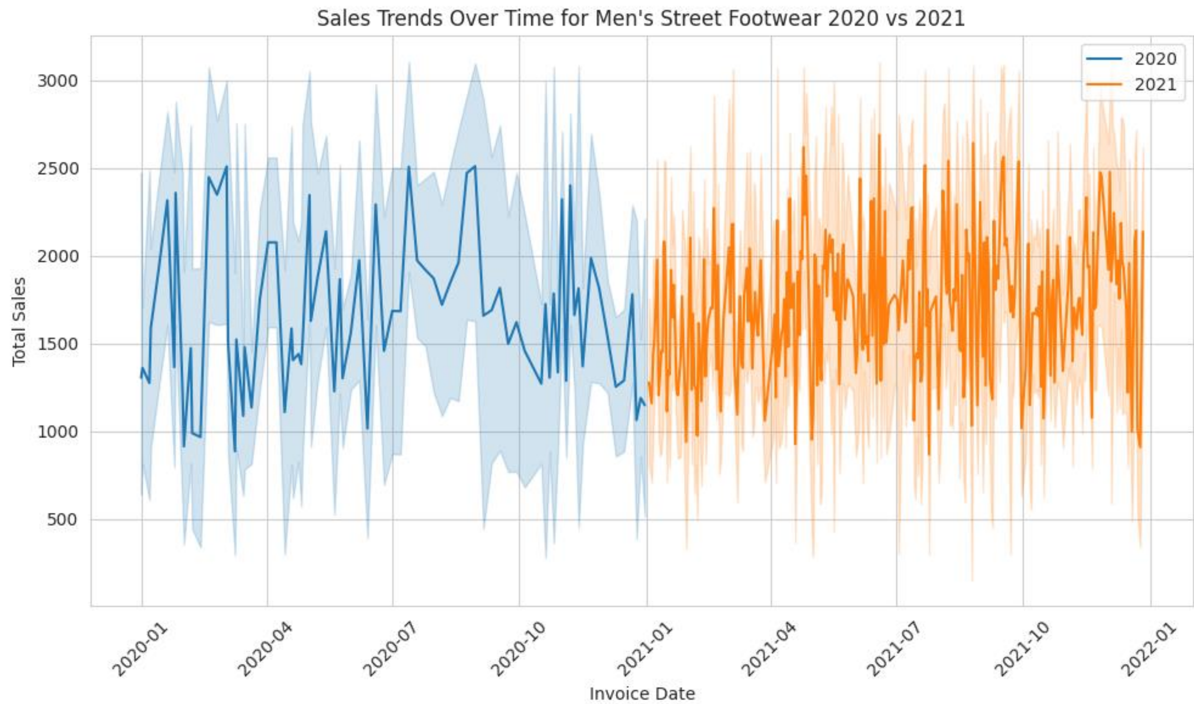
Comparative Analysis by Product Type:

- The most popular product category in January 2021 was men's apparel.
- Nonetheless, women's athletic footwear overtook men's apparel in sales by February 2021 and held the top spot for the remainder of the year.
- Women's street footwear sales increased year-over-year and by July 2021 they had surpassed both men's street footwear and men's apparel sales.

- Sales of men's athletic footwear rose all year long, surpassing those of men's apparel by August 2021.

```
import seaborn as sns
import matplotlib.pyplot as plt

# Filter data for the year 2021
df_2021 = df[df['Invoice Date'].dt.year == 2021]

# Get unique product types
product_types = ["Men's Street Footwear", "Men's Athletic Footwear",
                 "Women's Street Footwear", "Women's Athletic Footwear",
                 "Men's Apparel", "Women's Apparel"]

# Create separate line plots for each product type
g = sns.FacetGrid(df_2021, col='Product\t', col_wrap=3, height=4, aspect=1.5)
g.map(sns.lineplot, 'Invoice Date', 'Total Sales')

# Set titles and labels
g.set_titles("Sales Trends Over Time for {col_name} 2021")
g.set_axis_labels("Invoice Date", "Total Sales")
g.set_xticklabels(rotation=45)

plt.show()
```

7. Analysis of Product type 2020 and 2021

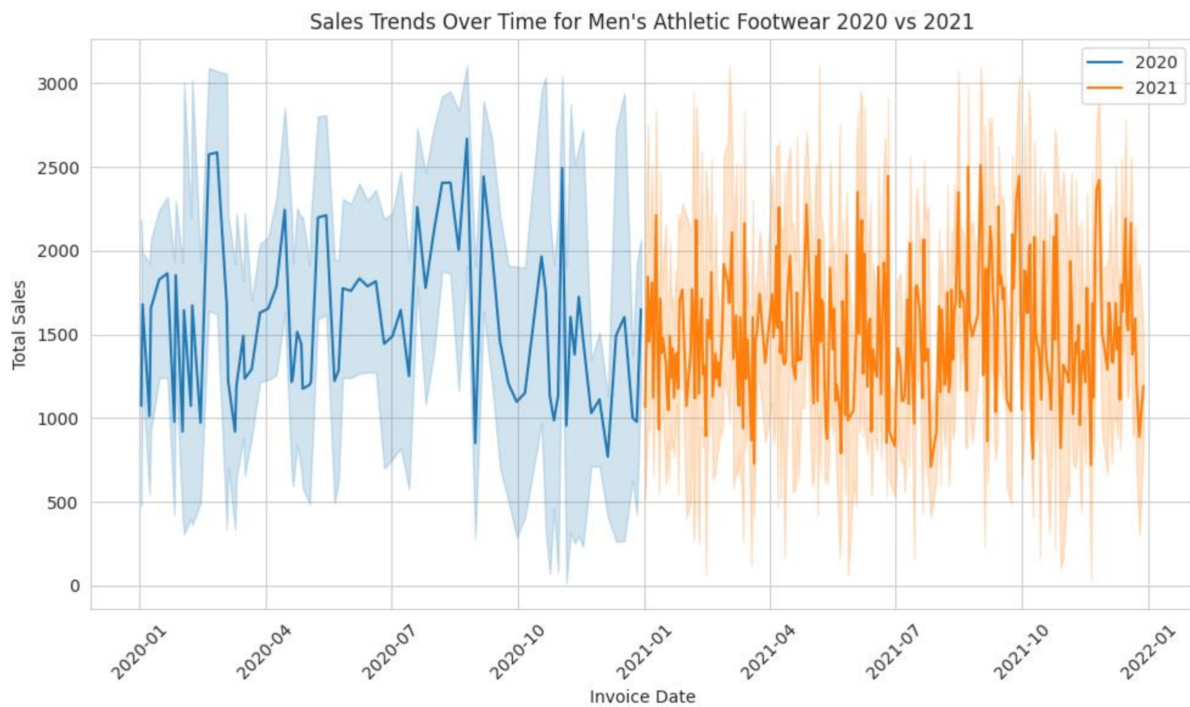


- Seasonal Fluctuations:

  A seasonal pattern in sales is evident in the graph. Sales increase significantly in Q3 (July-September) of both 2020 and 2021, peaking at that time of year. This implies that there is a direct relationship between the arrival of warmer months and the rise in demand for street shoes for men. Possible causes for this include:
  - Fall fashion shopping for back-to-school.
  - Summertime outdoor activities should be prioritized.
  - Release of footwear companies' newest summer collections.

- Notably, both years' sales declined after reaching their Q3 peaks. This could suggest:
  - Consumers spend less on shoes once summer ends.
  - Fewer stocks available following the period of highest purchase.

- Growth from Year to Year:
  When comparing 2020 to 2021, there is a positive trend even though both years show seasonal fluctuations. Every quarter in 2021 seems to have a higher overall sales value. This might result from:
  - Spending by consumers increased when pandemic (COVID) restrictions relaxed.
  - Increasing acceptance of particular street shoe trends.
  - Possible transition to online shoe shopping.



- Seasonal Fluctuations:

  The graph suggests a potential seasonal pattern, with sales possibly peaking in both years' Q3 (July–September). Nevertheless, it is challenging to unequivocally confirm a consistent seasonal trend. The following explanations explain why a seasonal peak in Q3 might be developing:
  - Fall sports season shopping for back-to-school outfits.
  - During the summer, put an emphasis on outdoor activities and fitness regimens.
  - Summer athletic events are marked by the release of new athletic footwear.

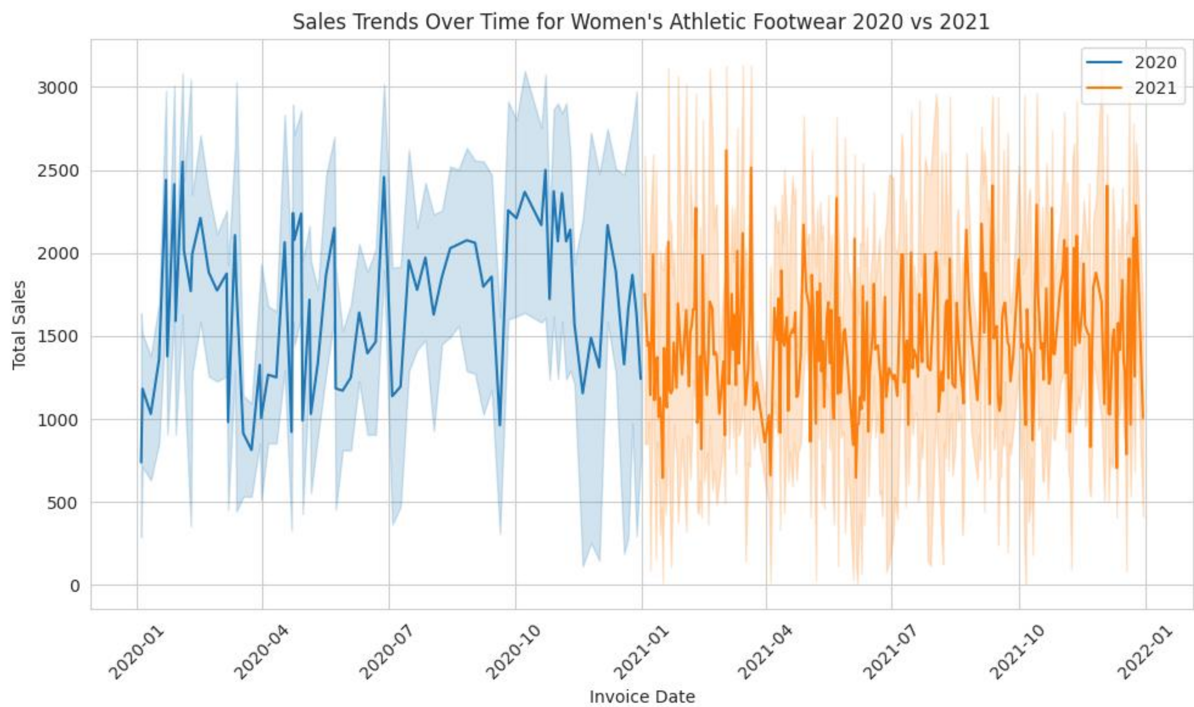
- It is necessary to investigate further the drop in sales that both years experienced after the possible Q3 peak. Some explanations that could be given are:
    - After the summer, consumers spent less on athletic footwear.
    - After the peak buying period, there is less stock available.

- Year-over-Year Growth:
  When comparing 2020 and 2021, there is a positive trend despite the uncertainty surrounding seasonality. For every quarter for which data is available in 2021, the overall sales value seems to be higher. There are a few possible reasons for this:
    - A greater focus on fitness and health as pandemic restrictions loosened.
    - Growing athleisure fashion trends are driving up the frequency of purchases of athletic footwear.
    - Possibility of an increase in footwear online sales, with a greater variety and more convenience.

- Seasonal Fluctuations:

  The graph points to a potential seasonal trend in women's street shoe sales, with summertime (around July) peaking as the sales may be. The supporting data and rationale for this possibility are broken down as follows:
    - July Sales Peaks: Two possible July sales peaks in 2020 and 2021 are depicted in the graph. This might just be a coincidence, but it suggests a recurrent trend.
    - Warm Weather Footwear: The summer months usually see a rise in the sales of sandals, open-toed shoes, and other warmer-weather footwear. Sales during these months may increase as a result of this.
    - Trends in Summer Shopping: Buying binges during the summer may be a factor in the peak. People may be buying new shoes for outdoor activities, trips, or just to refresh their summer attire.

- Year-over-Year Growth:
    - From July 2020 to July 2021, sales could rise by 20.00%, according to the scant data that is currently available. This indicates a possible annual increase in the sales of street footwear for women.

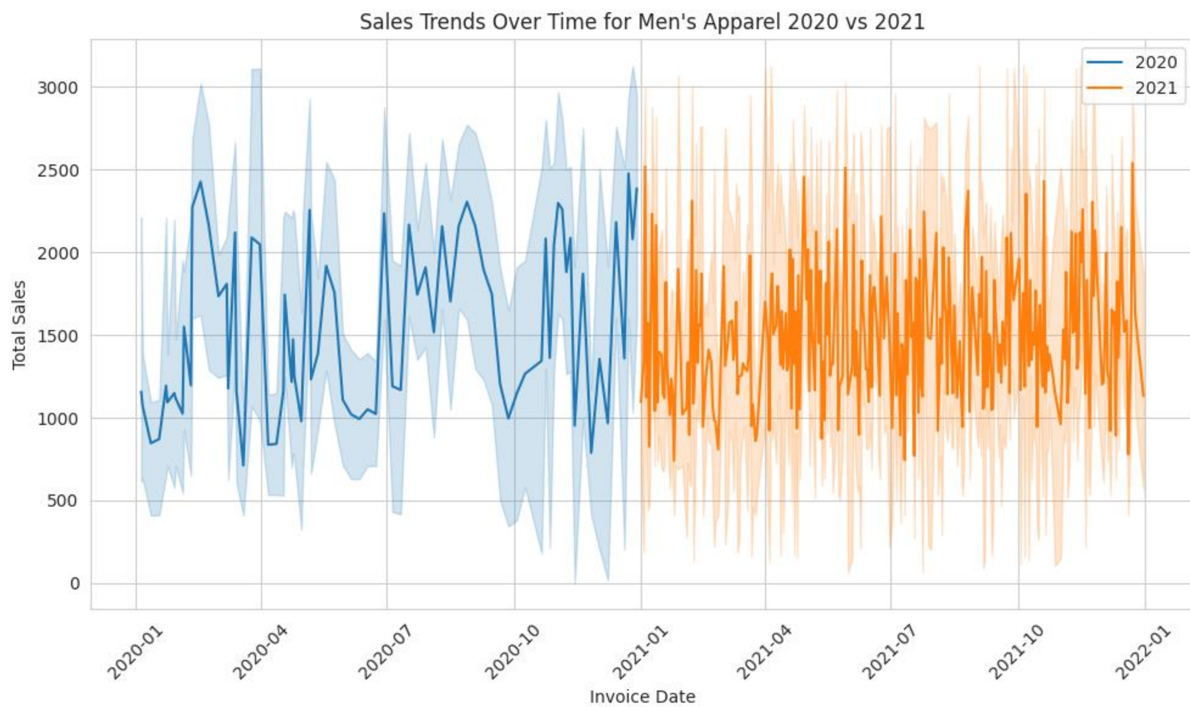


- Seasonal Fluctuations:

  Sales may be showing a seasonal trend, peaking in July 2020 and July 2021, respectively. Although it's hard to tell for sure from this graph, sales may be rising during the summer due to seasonal trends.
  It's possible that sales varied throughout the year, possibly peaking in July 2020 and July 2021 but possibly falling in other months as well. There are several possible reasons for the increase in sales between April 2020 and July 2020, including the following:
  - seasonal consumer behaviour, with consumers buying sports shoes for outdoor summertime activities.
  - 2020 saw a spike in sales of athletic wear following the lockdown as consumers sought out at-home workout options.

- Year-over-Year Growth:

  - Potential Year-over-Year Growth: It's possible that sales increased from the previous year. July 2021 sales seem to be higher than July 2020 sales. This might point to a general rise in the market for athletic footwear for women.



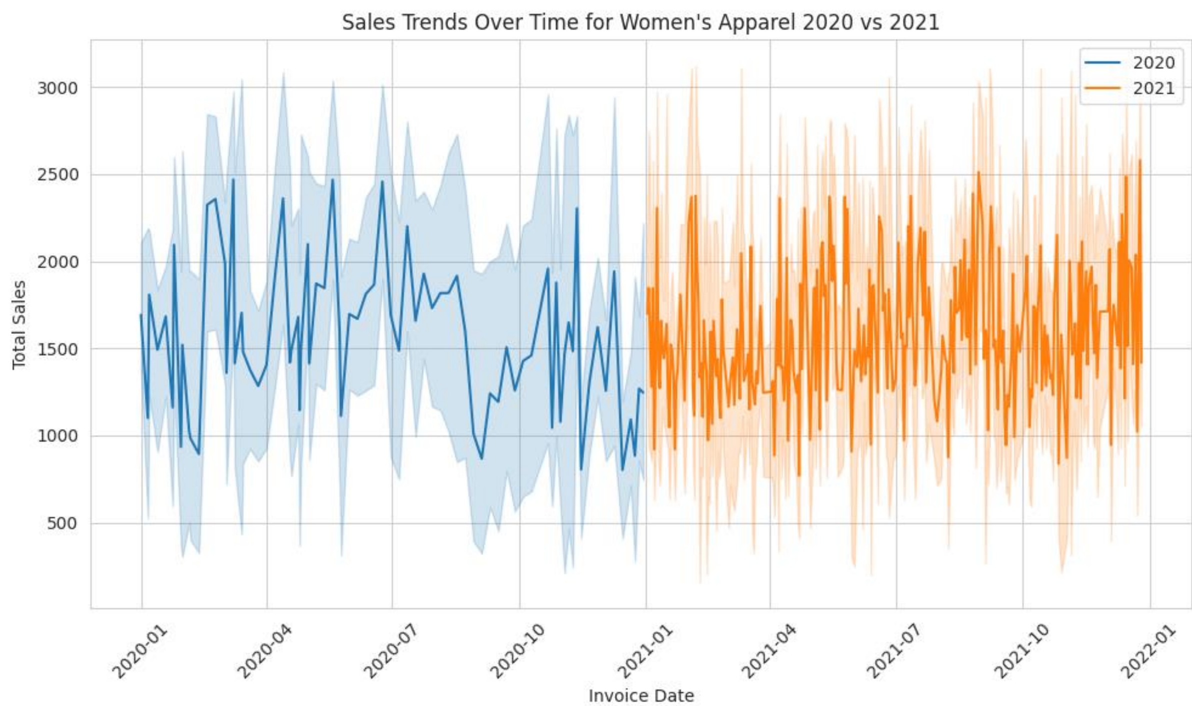
- Seasonal Fluctuations:

  Sales may follow a seasonal pattern, peaking in December of 2020 and 2021. There are a few possible reasons for this, including:
  - December is the time to shop for dresses to give as gifts or to wear to a party.
  - Greater emphasis on dressing up for holiday events at the end of the year.

- Year-over-Year Growth:

  For the most part, sales in 2021 seem to be higher overall than they were in 2020. There are several possible reasons for this, including:
  - Spending by consumers increased when pandemic restrictions relaxed.
  - Growing acceptance of online clothing shopping.
  - Trend shift in Favor of dresses in clothing.

- Summer Months: In both years, sales are typically lowest from June to August. This might occur because of:
  - In hot weather, people are less likely to purchase dresses.
  - Pay attention to summertime getaways and pursuits that don't always call for dresses.



- Seasonal Fluctuations:

    - December 2021 and December 2020. This might be the result of buying dresses for parties or as gifts during the holidays. In both years, sales tend to be lower in the summer (June–August), probably because fewer people purchase dresses in the heat.

    - Sales for both years appear to have peaked in the summer, from June to August. This is probably due to the fact that swimwear is more frequently purchased for summertime activities and vacations.
    - In both sales seem to peak in the fall and winter (October–March). This is probably due to the fact that sweaters are more popular in the winter.

- Year-over-Year Growth:
    - may exhibit greater sales in 2021 than in 2020, especially in December of both years. Sales in 2021 could surpass those in 2020, particularly in the fall and winter.

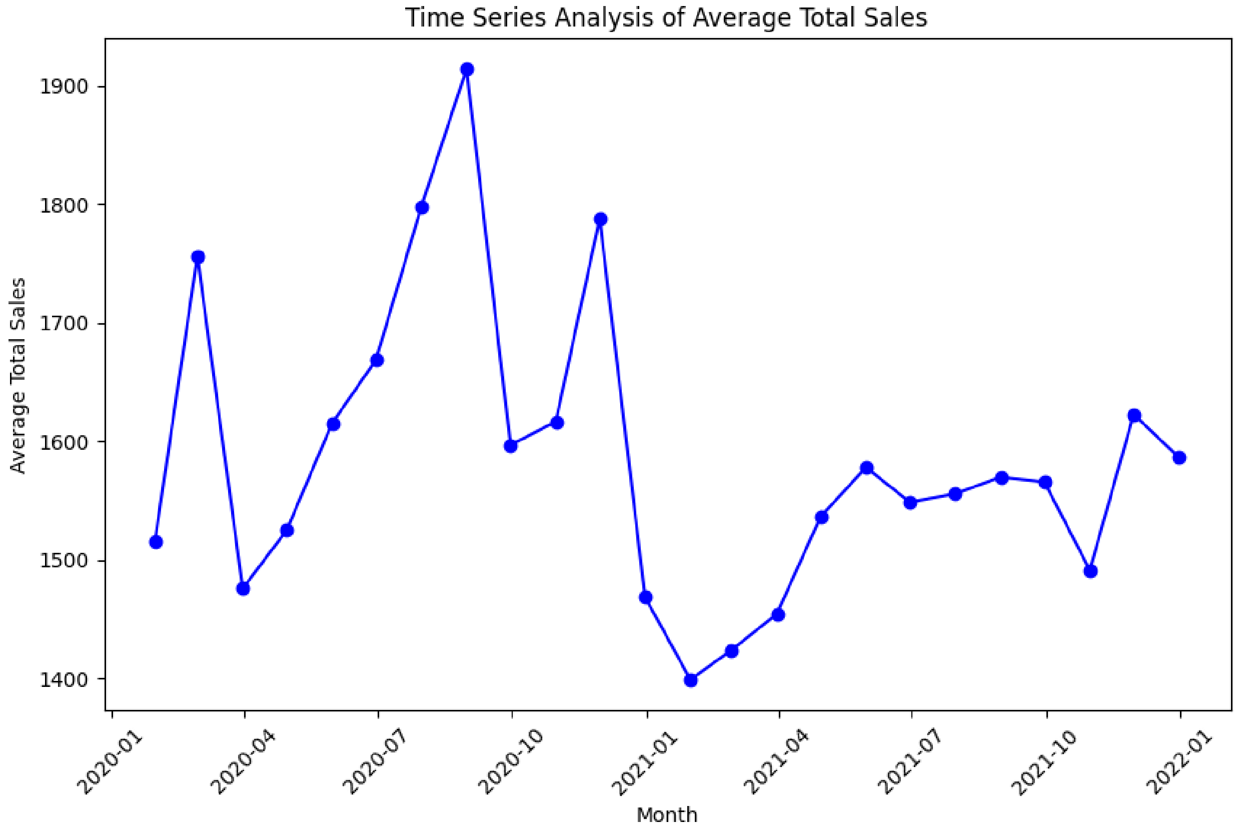
Codes for above Line Plots

```
import seaborn as sns
import matplotlib.pyplot as plt

# Filter data for the year 2020 and 2021
df_2020 = df[df['Invoice Date'].dt.year == 2020]
df_2021 = df[df['Invoice Date'].dt.year == 2021]

# Get unique product types
product_types = ["Men's Street Footwear", "Men's Athletic Footwear",
                 "Women's Street Footwear", "Women's Athletic Footwear",
                 "Men's Apparel", "Women's Apparel"]

# Create separate line plots for each product type
for product_type in product_types:
    plt.figure(figsize=(10, 6))

    # Filter data for the current product type in 2020 and 2021
    df_product_2020 = df_2020[df_2020['Product\t'] == product_type]
    df_product_2021 = df_2021[df_2021['Product\t'] == product_type]

    # Plot sales trends for 2020
    sns.lineplot(data=df_product_2020, x='Invoice Date', y='Total Sales',
label='2020')

    # Plot sales trends for 2021
    sns.lineplot(data=df_product_2021, x='Invoice Date', y='Total Sales',
label='2021')

    # Set title and labels
    plt.title(f'Sales Trends Over Time for {product_type}')
    plt.xlabel('Invoice Date')
    plt.ylabel('Total Sales')

    # Rotate x-axis labels for better readability
    plt.xticks(rotation=45)

    # Show legend
    plt.legend()
```

```
# Show plot
plt.tight_layout()
plt.show()
```

Models

1. Time series Analysis of year and average total sales of the company



The average total sales throughout time are displayed in a time series plot. "Time Series Analysis of Average Total Sales" is the title of the chart.

Time is displayed on the chart's x-axis. The time scale looks to encompass a two-year period divided into months. It begins in January 2020 and runs through December 2021.
The average total sales numbers are displayed on the y-axis. The scale ranges from 1400 to 2000 dollars and is marked in dollars.

On the graph, there is only one line drawn. Over the course of the two years, there seems to be some fluctuation in the blue line.

Regarding the average total sales throughout time, the following findings are made:
The data seems to show a seasonal trend, with average total sales being lower in the spring and summer (April to July) and greater in the later fall and winter months (October to December).
Over the course of the two years, average total sales appear to be modestly rising.

```
df_time_series = df[['Invoice Date', 'Total Sales']]

df_time_series['Invoice Date'] = pd.to_datetime(df_time_series['Invoice Date'])

grouped_df = df_time_series.groupby(pd.Grouper(key='Invoice Date', freq='M'))['Total Sales'].mean()

grouped_df = grouped_df.reset_index()
plt.figure(figsize=(10, 6))

plt.plot(grouped_df['Invoice Date'], grouped_df['Total Sales'],linestyle='-', marker='o', color='blue')
plt.xlabel('Month')
plt.ylabel('Average Total Sales')
plt.title('Time Series Analysis of Average Total Sales')
plt.xticks(rotation=45)
plt.show()
```

ARIMA moving average and total average sales

Below time series plot with an ARIMA moving average overlayed on the real sales data, displaying the average total sales. "2000 Time Series Analysis of Average Total Sales with ARIMA Moving Average" is the title of the chart.

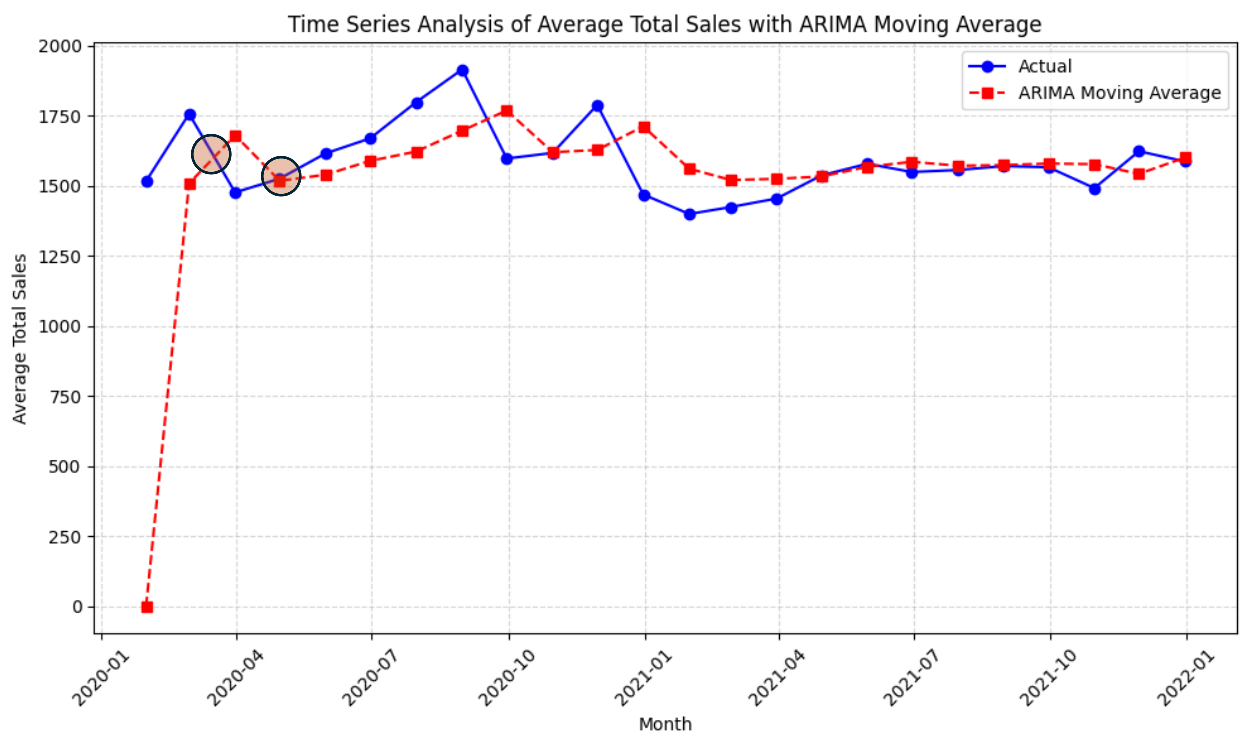
Time is displayed on the chart's x-axis. There are tick marks on the axis that most likely correspond to months, but the precise units for the time scale are not indicated.
The average total sales numbers are displayed on the y-axis. The scale has a dollar sign on it and ranges from 250 to 1750.

On the graph, there are two lines plotted:
The actual average total sales data are displayed as a blue line.
The sales data's ARIMA moving average is displayed as an orange line.

Autoregressive Integrated Moving Average is referred to as ARIMA. This statistical technique is used to model and predict data from time series. The graph shows us the following information:

The real sales data (blue line) varies during the course of the duration.
The ARIMA moving average (orange line) helps to smooth out these swings and provides a more overall trend in the data.
The moving average indicates a potential upward tendency in average total sales over time, despite the fact that the actual sales data exhibits ups and downs.

We can see a accurate pattern from ARIMA model and the total sales curve .when the average curve crosses the model curve suddenly sales of the company goes down .when the model crossed the average line total sales of the company go up.



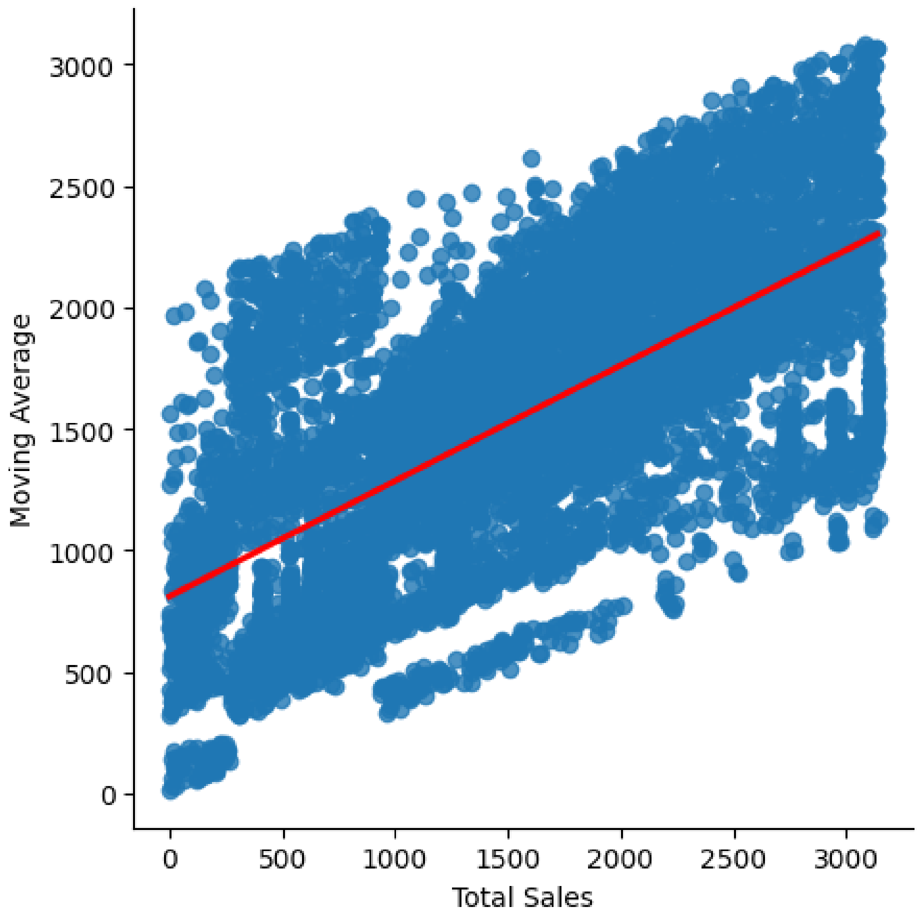
```python
df_time_series = df[['Invoice Date', 'Total Sales']]
df_time_series['Invoice Date'] = pd.to_datetime(df_time_series['Invoice Date'])
grouped_df = df_time_series.groupby(pd.Grouper(key='Invoice Date', freq='M'))['Total Sales'].mean()
grouped_df = grouped_df.reset_index()

model = ARIMA(grouped_df['Total Sales'], order=(1,1,1))  # ARIMA(1,1,1) model
model_fit = model.fit()
predictions = model_fit.predict(start=0, end=len(grouped_df)-1)

plt.figure(figsize=(10, 6))

plt.plot(grouped_df['Invoice Date'], grouped_df['Total Sales'], linestyle='-', marker='o', color='blue', label='Actual')


plt.plot(grouped_df['Invoice Date'], predictions, linestyle='--', marker='s', color='red', label='ARIMA Moving Average')

plt.xlabel('Month')
plt.ylabel('Average Total Sales')
plt.title('Time Series Analysis of Average Total Sales with ARIMA Moving Average')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```

2. Linear regression

The plot seems to illustrate how a linear regression model was fitted to a dataset with total sales as the dependent variable and the moving average as the independent variable.

The moving average is shown by the x-axis, which has a scale from 0 to around 3000.
The entire sales are represented by the y-axis, which has a scale from 500 to around 3000.

Plotted on the graph are multiple data points that seem to be dispersed around a straight line.
The linear regression model indicates that the straight line is the best-fit line.
The best-fit line's positive slope indicates that the moving average and total sales have a positive linear connection. Put differently, when the moving average rises, overall sales often increases as well




```python
sns.lmplot(x='Total Sales', y='Moving Average', data=df, line_kws={'color': 'red'})

plt.show()
```

Overall Conclusions and Recommendations

A. Sales Method

a) Sales Channel Performance:

- When compared to outlet and online sales techniques, in-store sales continuously show a higher volume of units sold, demonstrating the ongoing importance of physical retail locations in generating sales.

- When compared to in-store sales, outlet and internet sales strategies show lower unit sales. However, outlet sales do marginally better than online sales.

b) Regional Disparities:

- Variations in sales effectiveness between regions underline how crucial it is to customize sales tactics for regions.

- In terms of units sold for both in-store and online sales methods, the Southwest region tops the list, indicating a robust market presence in this area.

c) Seasonal Trends:

- Seasonal fluctuations in sales are evident, with certain months experiencing peaks or declines in sales volume across all sales methods.

- Notably, online sales exhibit a significant peak in August, indicating a surge in online shopping activity during this period.

d) Opportunities for Improvement:

- Outlet sales have room to grow, especially if they want to reverse the June decrease. Reviving outlet sales performance might require more research and calculated actions.

- Targeted attempts to increase sales volume could have a positive impact on in-store sales while preserving relative stability, particularly in slower months like December.

e) Online Sales Enhancement:

- Maximize use of the steady increase in sales volume seen in online channels by increasing the investments in digital marketing initiatives and enhancing the online buying experience.

- Use focused marketing efforts and specials to take advantage of the surge in online sales, particularly in high-demand months like August.

f) Outlet Sales Revitalization:

- Analyze outlet locations, product offers, and promotional activities in detail to address the June fall in outlet sales.

- To attract more clients and boost revenue, think about changing up the outlet's stock or offering special offers and discounts.

g) In-store Sales Optimization:

- Examine ways to improve in-store sales performance through localized marketing efforts, better store layouts, and enhanced customer service, especially in slower months like December.
- To enable tailored product placement and promotional methods, use data analytics to uncover trends and preferences among in-store shoppers.

B. Financial Information

a) Focus on Underperforming Regions and Cities:

- To boost sales in some places, such as Orlando, identify regions and cities with poorer sales performance and devote money to focused marketing campaigns and promotions.
- Give untapped prospective sectors, like those with less market penetration, priority.

b) Product Mix Optimization:

- Examine each product category's sales data and modify the product mix as necessary. Invest in well-liked markets such as athletic footwear for men and women, but also look for ways to increase sales in other markets by developing new products and implementing creative marketing techniques.

c) Enhance Retailer Relationships:

- To reach a larger client base, fortify alliances with major retailers like Walmart and Amazon and investigate ways to work with up-and-coming merchants to develop distribution networks.

d) Improve Sales Methods:

- Capitalize on the expanding trend of online buying by making investments in online sales channels and improving the e-commerce experience. Put tactics in place to increase in-store sales by offering specials and individualized customer service.

e) Seasonal Sales Strategies:

- Create seasonal sales plans by utilizing time series analysis findings. To optimize sales potential, coordinate marketing campaigns and inventory levels with seasonal variations in customer demand. Pay particular attention to peak seasons.

f) Utilize Predictive Analytics:

- Keep using sophisticated analytics methods to predict future sales trends and make data-driven choices, such as linear regression and ARIMA. Optimize results by integrating predictive modeling into resource allocation and sales planning procedures.

g) Customer Engagement and Loyalty Programs:

- To cultivate enduring relationships with customers, implement customer engagement initiatives and loyalty programs. To improve client happiness and promote repeat business, provide incentives, awards, and tailored recommendations.

C. Retailers Information

a) Retailer Ranking and Market Dominance:

- Sports Direct, Walmart, and Kohl's are the next best-performing retailers with the highest overall sales. This demonstrates Kohl's robust market presence and consumer appeal.
- Foot Locker's lower overall sales than its competitors point to possible difficulties in competing with them in the Adidas market area.

b) Regional Performance Analysis:

- Walmart dominates the Southern region, demonstrating a solid foothold in that market. Retailers show differing degrees of performance across different regions. This emphasizes how crucial it is to target regions and comprehend local customer preferences.
- The Midwest accounts for most Foot Locker's sales, indicating a potential strength that might be expanded upon. Its poor sales in the South, however, point to areas where market penetration tactics may be strengthened.

c) Seasonal Trends:

- Seasonal elements like holidays have an impact on retailer sales, which fluctuate throughout the course of the month.
- Sales of West Gear surge during the holidays, indicating the significance of matching marketing initiatives with consumer purchasing trends.
- Sales on Amazon grow significantly near the end of the year, suggesting the possible influence of marketing campaigns or promotional events at that time.

d) Boxplot Analysis:

- Walmart shows the greatest spread or variability in sales, pointing to variations in the company's overall performance throughout time.4
- When compared to other stores, Kohl's and Walmart continue to have the greatest median sales amounts, indicating sustained performance.

e) **Regional Targeting methods:**

- To capitalize on regional strengths and overcome deficiencies, retailers should customize their marketing and sales methods. To understand local consumer preferences and modify product offerings appropriately, this may entail performing market research.

f) **Seasonal Marketing Campaigns:**

- To increase sales at peak times, retailers can create focused marketing campaigns in line with seasonal patterns. This could involve offering special discounts, incentives, and promos to draw customers during busy shopping seasons.

g) **Inventory Management:**

- Stores should streamline their inventory control procedures to guarantee adequate stock levels during periods of high sales while reducing surplus inventory during slower times. Precise sales forecasting and effective supply chain management are necessary for this.

h) **Comparative Evaluation:**

- Retailers need to keep a close eye on market trends and rival performance to spot market possibilities and dangers. Making strategic decisions based on this can help you stay competitive.

i) **Customer Experience and Engagement:**

- Improving the general customer experience can lead to more sales and more devoted customers, both online and in-store. This entails offering individualized shopping experiences, top-notch customer support, and smooth transaction procedures.
- Retailers may enhance their sales performance, fortify their market position, and foster long-term expansion by putting these suggestions into practice in the cutthroat Adidas retail environment.

D. Regions Information

a) Regional distribution

- The distribution of data across area is shown by the count plot, where the West has the most data points and the Southeast the fewest.

b) Sales performance

- The bar graph shows the performance of sales by region, with the West having the highest number of sales. The identification of high and low performing location can be guided by this information.

c) Profitability

- When profitability is broken down by region, the Midwest comes out on top, with the West making a major contribution as well. Pricing techniques can be improved by having a thorough understanding of these patterns.

d) Correlation analysis

- The scatter plot fails to clearly show how sales and profitability are correlated.

e) Regional comparison

- Benchmarking and identifying opportunities for improvement are made possible by comparing sales and profitability across region. The Northeast has a lot of improvement, while the West and Southeast stand out as high performing regions.

f) Resources allocation

- Allocate resources based on regional sales performance , focusing on regions with high potential for growth.

g) Marketing strategies

- While conducting focused activities to enhance sales in underperforming regions, marketing strategies can be adjusted to take advantages on high performing areas.

h) Operational efficiency

- Analyze the elements influencing profitability in areas with strong performance, then apply best practices throughout the company to increase profitability overall.

i) Continuous monitoring

- Keep updated on regional performance measures on a regular basis to ensure that strategies are aligned with changing market realities.

E. Products Information

a) Seasonal Campaigns and Promotions:

- In Q3, focus on back-to-school clothes; in the summer, highlight outdoor activities.
- Think of releasing limited-edition products at popular times.

b) Embrace Athleisure Trends:

- Offer a range of stylish and practical footwear styles.
- Emphasize adaptability in advertising strategies.

c) Online Sales Optimization:

- Boost mobile friendliness and usability of websites.
- Make use of focused digital marketing strategies.

d) Summer-Specific Marketing Approaches:

- In summer marketing, emphasize comfort, breathability, and adaptability.
- Display appropriateness for outdoor pursuits.

e) Encourage Activities Related to the Seasons:

- Take part in or sponsor summer fitness and sports competitions.
- Educate women who are active about the brand.

f) Wide Range of Products Offered:

- All year long, stock an assortment of dresses, swimwear, sportswear, and outerwear.

g) Autumn/Winter Collection Highlights:

- Emphasize the autumn and winter collections when sales are at their highest.
- Display must-have products for the winter, such as coats and sweaters.