USA Super Store Dataset:

Objectives:

- Analyze monthly sales data, calculate key metrics (revenue, average order size), and visulize trends over time.

Import Libraries

In [132…

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

Load Dataset

In [133…

```
df = pd.read_csv('D:/Bistartx Intership/Month-1/EDA_Super_Store_Sales/SuperStore.csv')
df.sample(5)
```

Out[133...

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code
9324	9325	CA-2018-121853	23/09/2018	29/09/2018	Standard Class	DB-13660	Duane Benoit	Consumer	United States	Los Angeles	California	90036.0
2668	2669	US-2016-139759	25/08/2016	30/08/2016	Standard Class	NL-18310	Nancy Lomonaco	Home Office	United States	Los Angeles	California	90045.0
6493	6494	CA-2017-113845	20/11/2017	25/11/2017	Standard Class	FA-14230	Frank Atkinson	Corporate	United States	Orlando	Florida	32839.0
4150	4151	CA-2018-106068	23/10/2018	28/10/2018	Standard Class	RB-19330	Randy Bradley	Consumer	United States	Austin	Texas	78745.0
4849	4850	CA-2015-107818	08/09/2015	14/09/2015	Standard Class	MC-17275	Marc Crier	Consumer	United States	Pasco	Washington	99301.0

In [134...

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9800 entries, 0 to 9799
Data columns (total 18 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Row ID         9800 non-null   int64  
 1   Order ID       9800 non-null   object 
 2   Order Date     9800 non-null   object 
 3   Ship Date      9800 non-null   object 
 4   Ship Mode      9800 non-null   object 
 5   Customer ID    9800 non-null   object 
 6   Customer Name  9800 non-null   object 
 7   Segment        9800 non-null   object 
 8   Country        9800 non-null   object 
 9   City           9800 non-null   object 
 10  State          9800 non-null   object 
 11  Postal Code    9789 non-null   float64
 12  Region         9800 non-null   object 
 13  Product ID     9800 non-null   object 
 14  Category       9800 non-null   object 
 15  Sub-Category   9800 non-null   object 
 16  Product Name   9800 non-null   object 
 17  Sales          9800 non-null   float64
dtypes: float64(2), int64(1), object(15)
memory usage: 1.3+ MB
```

In [135…

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

Out[135…

```
Row ID           0
Order ID         0
Order Date       0
Ship Date        0
Ship Mode        0
Customer ID      0
Customer Name    0
Segment          0
Country          0
City             0
State            0
Postal Code     11
Region           0
Product ID       0
Category         0
Sub-Category     0
Product Name     0
Sales            0
dtype: int64
```

Observations

1. Change `Order Date`, `Ship Date` dtype into datetime.

Errors Handling

Change dtype of order date, ship date to datetime

In [136…

```
date_columns = ['Order Date', 'Ship Date']

for col in date_columns:
    df[col] = pd.to_datetime(df[col], dayfirst = True)
    df[col] = df[col].dt.date
    df[col] = df[col].astype('datetime64[ns]')
```

Now the format is: format=Year/Month/Date

In [137…
```
df[['Order Date','Ship Date']].sample(5)
```

Out[137…

	Order Date	Ship Date
1682	2017-09-29	2017-10-01
7309	2018-06-10	2018-06-14
6723	2018-09-04	2018-09-06
2910	2016-11-05	2016-11-09
4220	2018-06-11	2018-06-13

In [138…
```
df['Row ID'].nunique()
```

Out[138… 9800

We already have the "Row ID" column in our dataset, so we set it as the index to organize the data efficiently

In [139…
```
df = df.set_index('Row ID')
df.index.name = None
```

In [140…
```
df.sample(5)
```

Out[140…

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	P
1231	CA-2018-100013	2018-11-06	2018-11-11	Standard Class	ZC-21910	Zuschuss Carroll	Consumer	United States	Los Angeles	California	90045.0	West	C 10
4988	CA-2017-149279	2017-04-24	2017-04-28	Standard Class	CL-12700	Craig Leslie	Home Office	United States	Colorado Springs	Colorado	80906.0	West	FU 10
5863	CA-2017-101525	2017-05-01	2017-05-04	Second Class	CM-12235	Chris McAfee	Consumer	United States	Little Rock	Arkansas	72209.0	South	C 10
6812	CA-2018-156237	2018-09-14	2018-09-15	First Class	PS-18760	Pamela Stobb	Consumer	United States	Philadelphia	Pennsylvania	19140.0	East	FU 10
172	CA-2015-118962	2015-08-05	2015-08-09	Standard Class	CS-12130	Chad Sievert	Consumer	United States	Los Angeles	California	90004.0	West	C 10

Since we already have Country, State, City, and Region columns, the Postal Code column is redundant and can be removed

In [141…

```
df.drop(columns=['Postal Code'], inplace=True)
```

In [142…

```
df.sample()
```

Out[142...

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Region	Product ID	Cat
9044	CA-2017-168830	2017-11-07	2017-11-13	Standard Class	ML-17395	Marina Lichtenstein	Corporate	United States	San Francisco	California	West	TEC-AC-10003911	Techr

In [143...

```
df.duplicated().sum()
```

Out[143...

```
1
```

There are no duplicate values in the data.

Summary of Data:

In [144...

```
for col in df.columns:
    print(f"Column {col}, Unique Values: {df[col].nunique()}")
```

```
Column Order ID, Unique Values: 4922
Column Order Date, Unique Values: 1230
Column Ship Date, Unique Values: 1326
Column Ship Mode, Unique Values: 4
Column Customer ID, Unique Values: 793
Column Customer Name, Unique Values: 793
Column Segment, Unique Values: 3
Column Country, Unique Values: 1
Column City, Unique Values: 529
Column State, Unique Values: 49
Column Region, Unique Values: 4
Column Product ID, Unique Values: 1861
Column Category, Unique Values: 3
Column Sub-Category, Unique Values: 17
Column Product Name, Unique Values: 1849
Column Sales, Unique Values: 5757
```

In [145…

```python
for col in df.columns:
    print(f'{df[col].value_counts()}')
    print('\n')
```

```
Order ID
CA-2018-100111    14
CA-2018-157987    12
CA-2017-165330    11
US-2017-108504    11
CA-2017-105732    10
                  ..
US-2016-110261     1
CA-2016-125710     1
US-2016-137960     1
CA-2016-124975     1
CA-2016-142202     1
Name: count, Length: 4922, dtype: int64


Order Date
2017-09-05    38
2017-11-10    35
2018-12-02    34
2018-12-01    34
2018-09-02    33
              ..
2017-02-25     1
2017-10-25     1
2015-02-21     1
2015-09-11     1
2016-05-09     1
Name: count, Length: 1230, dtype: int64


Ship Date
2018-09-26    34
2018-12-06    32
2016-12-16    31
2018-09-15    30
2018-09-06    30
              ..
2015-07-10     1
2016-03-29     1
2016-06-14     1
2018-01-10     1
2016-05-13     1
```

```
Name: count, Length: 1326, dtype: int64


Ship Mode
Standard Class    5859
Second Class      1902
First Class       1501
Same Day           538
Name: count, dtype: int64


Customer ID
WB-21850    35
MA-17560    34
PP-18955    34
JL-15835    33
CK-12205    32
            ..
JR-15700     1
CJ-11875     1
SC-20845     1
RE-19405     1
AO-10810     1
Name: count, Length: 793, dtype: int64


Customer Name
William Brown          35
Matt Abelman           34
Paul Prost             34
John Lee               33
Chloris Kastensmidt    32
                       ..
Jocasta Rupert          1
Carl Jackson            1
Sung Chung              1
Ricardo Emerson         1
Anthony O'Donnell       1
Name: count, Length: 793, dtype: int64


Segment
```

```
Consumer       5101
Corporate      2953
Home Office    1746
Name: count, dtype: int64


Country
United States    9800
Name: count, dtype: int64


City
New York City    891
Los Angeles      728
Philadelphia     532
San Francisco    500
Seattle          426
                ...
San Mateo          1
Cheyenne           1
Conway             1
Melbourne          1
Springdale         1
Name: count, Length: 529, dtype: int64


State
California              1946
New York                1097
Texas                    973
Pennsylvania             582
Washington               504
Illinois                 483
Ohio                     454
Florida                  373
Michigan                 253
North Carolina           247
Virginia                 224
Arizona                  223
Tennessee                183
Colorado                 179
Georgia                  177
```

```
Kentucky                137
Indiana                 135
Massachusetts           135
Oregon                  122
New Jersey              122
Maryland                105
Wisconsin               105
Delaware                 93
Minnesota                89
Connecticut              82
Missouri                 66
Oklahoma                 66
Alabama                  61
Arkansas                 60
Rhode Island             55
Mississippi              53
Utah                     53
South Carolina           42
Louisiana                41
Nevada                   39
Nebraska                 38
New Mexico               37
New Hampshire            27
Iowa                     26
Kansas                   24
Idaho                    21
Montana                  15
South Dakota             12
Vermont                  11
District of Columbia     10
Maine                     8
North Dakota              7
West Virginia             4
Wyoming                   1
Name: count, dtype: int64


Region
West       3140
East       2785
Central    2277
South      1598
```

```
Name: count, dtype: int64


Product ID
OFF-PA-10001970    19
TEC-AC-10003832    18
FUR-FU-10004270    16
TEC-AC-10002049    15
TEC-AC-10003628    15
                   ..
OFF-PA-10000919     1
TEC-MA-10003353     1
OFF-LA-10003388     1
OFF-EN-10004206     1
TEC-PH-10002645     1
Name: count, Length: 1861, dtype: int64


Category
Office Supplies    5909
Furniture          2078
Technology         1813
Name: count, dtype: int64


Sub-Category
Binders        1492
Paper          1338
Furnishings     931
Phones          876
Storage         832
Art             785
Accessories     756
Chairs          607
Appliances      459
Labels          357
Tables          314
Envelopes       248
Bookcases       226
Fasteners       214
Supplies        184
Machines        115
```

```
Copiers          66
Name: count, dtype: int64


Product Name
Staple envelope                                                     47
Staples                                                             46
Easy-staple paper                                                   44
Avery Non-Stick Binders                                             20
Staples in misc. colors                                             18
                                                                    ..
Xiaomi Mi3                                                           1
Universal Ultra Bright White Copier/Laser Paper, 8 1/2" x 11", Ream  1
Socket Bluetooth Cordless Hand Scanner (CHS)                         1
Logitech Illuminated Ultrathin Keyboard with Backlighting            1
LG G2                                                                1
Name: count, Length: 1849, dtype: int64


Sales
12.960     55
15.552     39
19.440     39
10.368     35
25.920     34
           ..
339.136     1
60.048      1
5.022       1
7.857       1
10.384      1
Name: count, Length: 5757, dtype: int64
```

Summary of Data

- **Shiping Modes = 4**
    - Standard class
    - Second class

    - First class
    - Same day
- **Segments = 3**
    - Consumer
    - Corporate
    - Home Office
- **Categories = 3**
    - Office Supplies
    - Furniture
    - Technology
- **Sub-Categories = 17**
- **Country = 1**
    - United States
- **States = 49**
- **Cities = 529**
- **Regions = 4**
    - West
    - East
    - Central
    - South
- **Orders = 9800**
- **Customers = 793**
- **Products = 1849**

Exploratory Data Analyysis

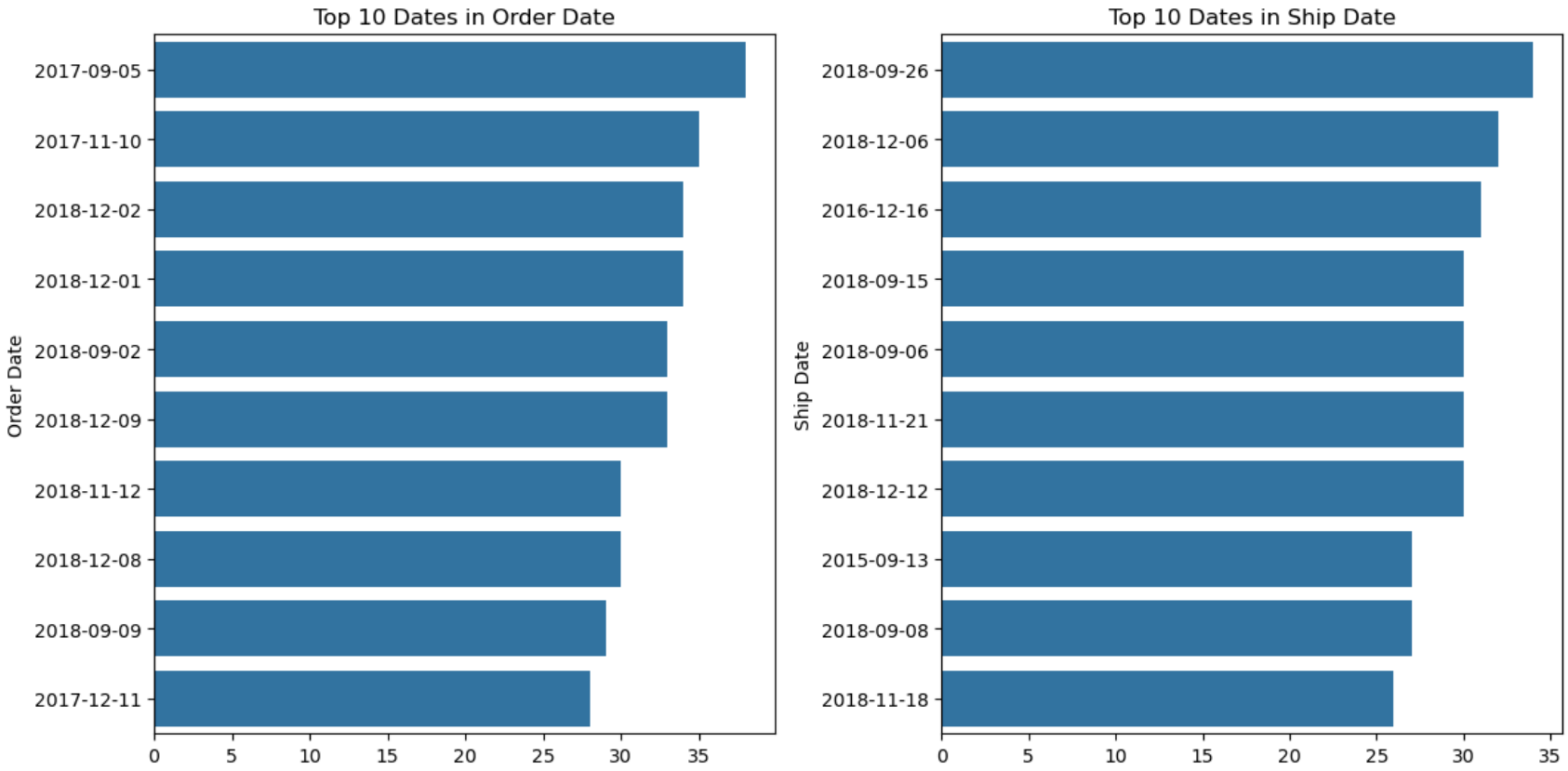
Univarite Analysis

Date Columns

In [146…
```
date_cols = ['Order Date','Ship Date']
for i in date_cols:
    print(f"{i}: Min = {df[i].min()}, Max = {df[i].max()}, Unique = {df[i].nunique()}")
```

```
Order Date: Min = 2015-01-03 00:00:00, Max = 2018-12-30 00:00:00, Unique = 1230
Ship Date: Min = 2015-01-07 00:00:00, Max = 2019-01-05 00:00:00, Unique = 1326
```

In [147…
```
plt.figure(figsize=(12, 6))
for i, col in enumerate(date_cols):
    ax = plt.subplot(1, 2, i+1)
    sns.barplot(y=df[col].value_counts().head(10).index, x=df[col].value_counts().head(10).values, ax=ax)
    plt.title(f"Top 10 Dates in {col}")

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



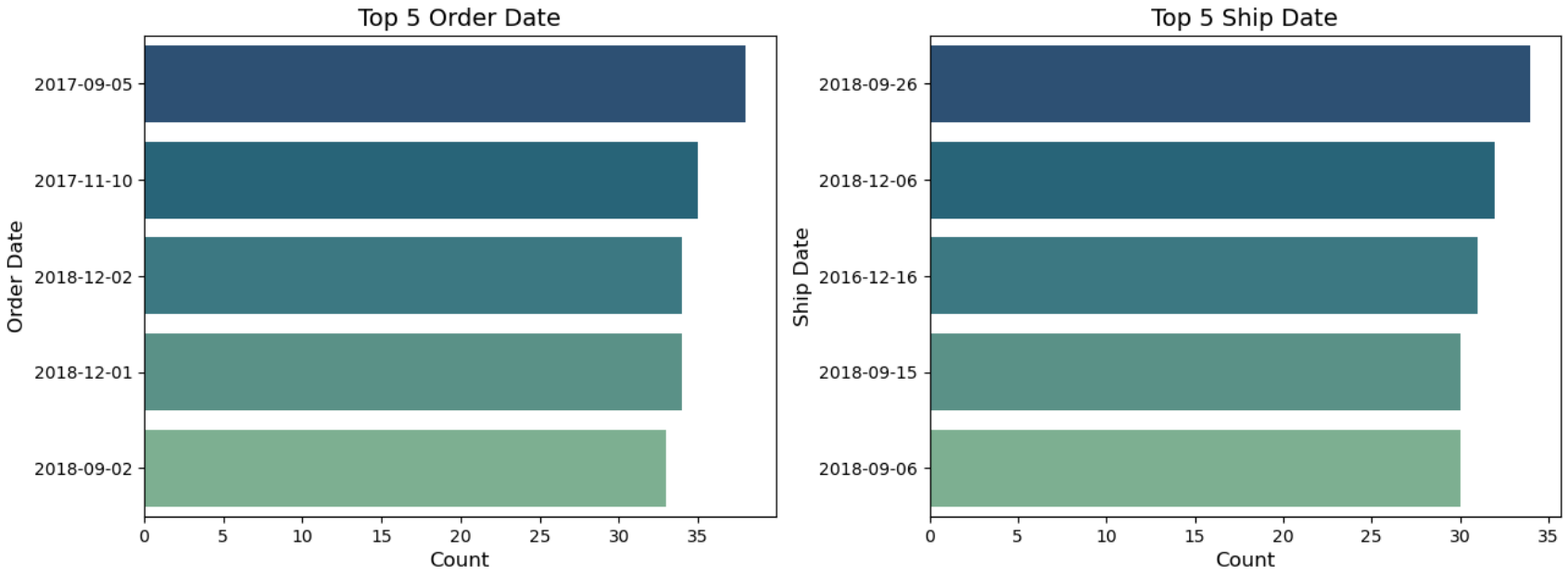
Observations

- `Order` volume is highest from September to December.
- `Orders peak` from `September to Decembe` r due to holiday shopping, Black Friday, and Christmas demand.

In [148…
```
df['Order Year'] = df['Order Date'].dt.year
df['Order Month'] = df['Order Date'].dt.month
df['Order Day'] = df['Order Date'].dt.day
df['Order Weekday'] = df['Order Date'].dt.day_name()
```

In [149…
```
plt.figure(figsize=(13, 9))

for i, col in enumerate(date_cols, 1):
    ax = plt.subplot(2, 2, i)
```

```
    top_5 = df[col].value_counts().head(5)
    sns.barplot(x=top_5.values, y=top_5.index,
                ax=ax, orient='h', order=top_5.index,
                palette='crest_r')

    ax.set_title(f"Top 5 {col}", fontsize=14)
    ax.set_xlabel("Count", fontsize=12)
    ax.set_ylabel(col, fontsize=12)

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



Observations

- Most orders were placed in `2018`.
- Order volume is highest from `September to December`.
- The busiest order days are the `20th–23rd, 26th, 2nd–5th, and 8th & 11th`.
- Orders peak on `Tuesdays, Saturdays, Sundays, and Mondays`.

Reasons

Most Orders Were Placed in 2018

- The company might have expanded, gained more customers, or introduced new products.
- 2018 could have been a peak year for e-commerce or retail sales.

Most Orders Were Placed Between September and December

- Holiday Shopping Season: Sales surge due to Christmas, Thanksgiving, and New Year.
- Black Friday & Cyber Monday: Massive discounts in November drive purchases.
- Year-End Budgets: Businesses and individuals finalize purchases before the new year.

Most Frequent Order Days: 2nd–5th, 8th, 11th, 20th–23rd, 26th

- Payday Effect: Many people shop right after getting paid (often on the 1st or 15th of the month).
- Seasonal Events: Some dates may align with sales, promotions, or holidays.

Most Orders Were Placed on Tuesdays, Saturdays, Sundays, and Mondays

- Weekend Shopping: People have more time to shop on Saturdays and Sundays.
- Monday Back-to-Work Effect: Businesses may place bulk orders at the start of the week.
- Tuesday Discounts: Many e-commerce platforms and retailers launch deals on Tuesdays.

Categorical Columns

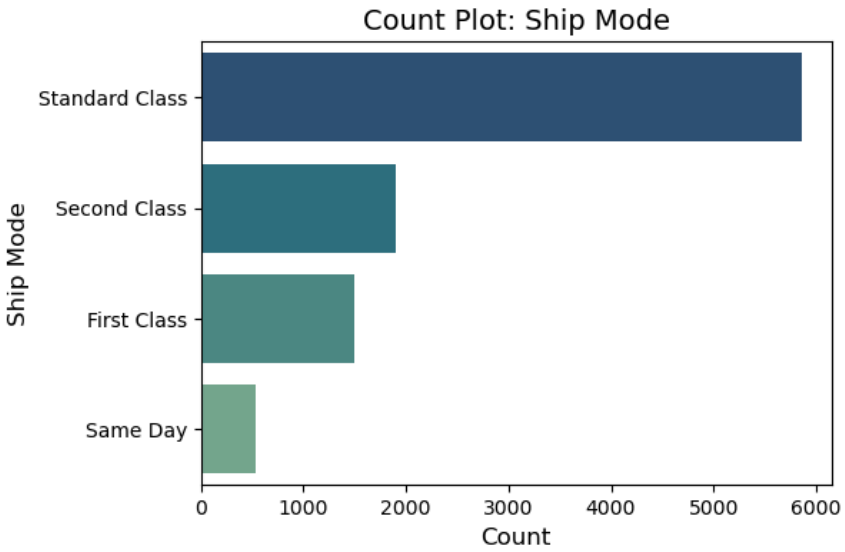
In [150…
```
df.select_dtypes(include='object').columns
```

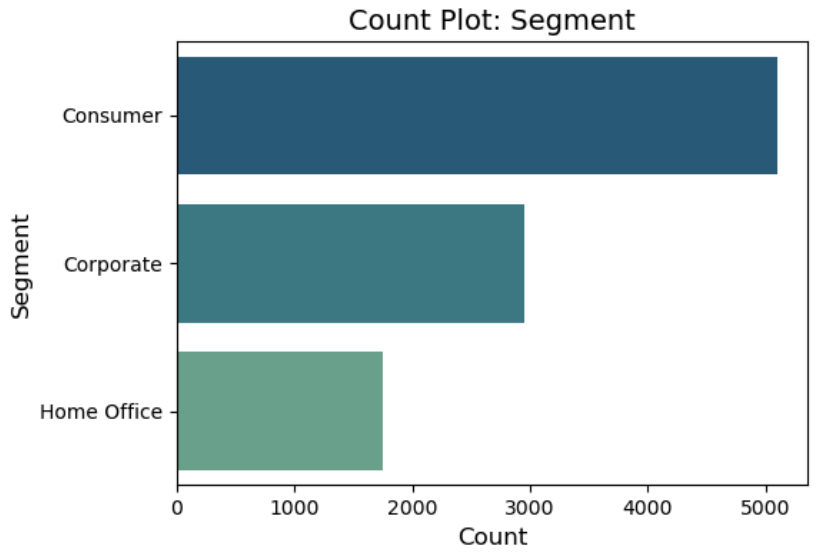
Out[150…
```
Index(['Order ID', 'Ship Mode', 'Customer ID', 'Customer Name', 'Segment',
       'Country', 'City', 'State', 'Region', 'Product ID', 'Category',
       'Sub-Category', 'Product Name', 'Order Weekday'],
      dtype='object')
```

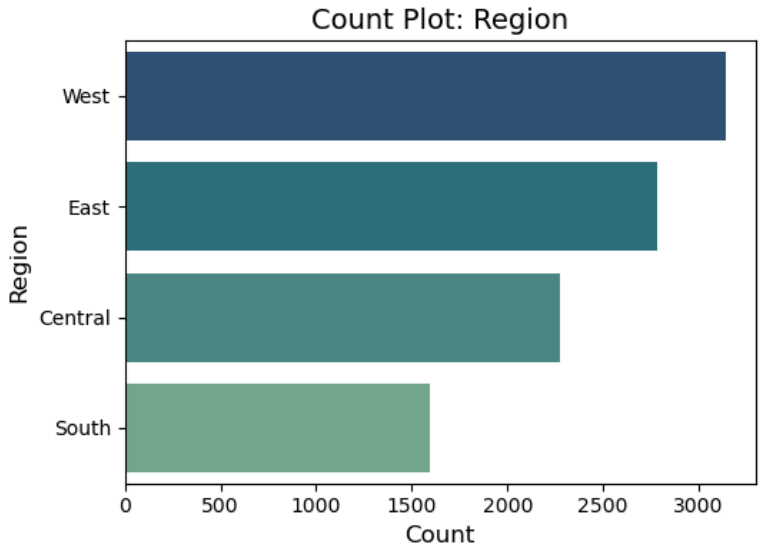
In [151…
```
cat_cols = ['Ship Mode', 'Segment', 'Region', 'Category']
```

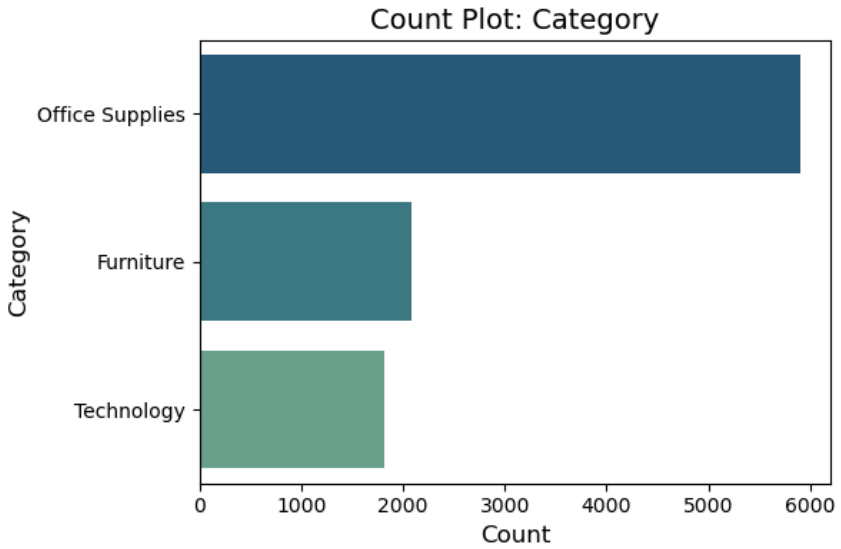
In [152…

```
plt.figure(figsize=(12, 8))

for i, col in enumerate(cat_cols):
    ax = plt.subplot(2, 2, i+1)

    sns.countplot(data=df, y=col,
                  order=df[col].value_counts().head().index,
                  palette= 'crest_r')

    ax.set_title(f"Count Plot: {col}", fontsize=14)
    ax.set_xlabel("Count", fontsize=12)
    ax.set_ylabel(col, fontsize=12)

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









Observations

- The `Standard Class` shipping mode is the most popular.
- The `West region` records the highest number of orders.
- The `Consumer segment` contributes the highest number of orders.

- The `Office Supplies` category has the highest number of orders.

Reasons

The `Standard Class` shipping mode is the most popular

- Cost-Effective: Standard Class is usually the cheapest shipping option, making it attractive for customers who want to save money.
- Bulk Orders & Businesses: Businesses ordering office supplies or other products in bulk might not be in a rush, so they opt for cost-effective shipping.
- Wide Availability: Standard shipping is generally available for all locations, whereas faster options (like Same-Day or Next-Day) might be limited to specific areas.

The `West region` records the highest number of orders

- High Population & Business Hubs: The West region likely includes major cities with strong economies (e.g., California, Washington, etc.), leading to higher consumer demand.
- Tech & Corporate Influence: Many tech companies and startups operate in the West, increasing demand for office supplies and equipment.
- Higher Purchasing Power: Some Western states have a higher average income, leading to more discretionary spending.
- Strong Distribution Networks: Well-developed supply chain infrastructure might make it easier to fulfill orders quickly, encouraging more purchases.

The `Consumer segment` contributes the highest number of orders

- Larger Customer Base: The number of individual consumers is much higher than the number of businesses or government organizations.
- Frequent Small Purchases: Consumers often make frequent, smaller orders compared to businesses that order in bulk.
- Marketing & Discounts: Businesses might negotiate bulk discounts, but consumers often pay full price, increasing total revenue.
- Impulse Buying: Unlike businesses, consumers are more likely to make impulse purchases, leading to higher overall sales.

The `Office Supplies` category has the highest number of orders

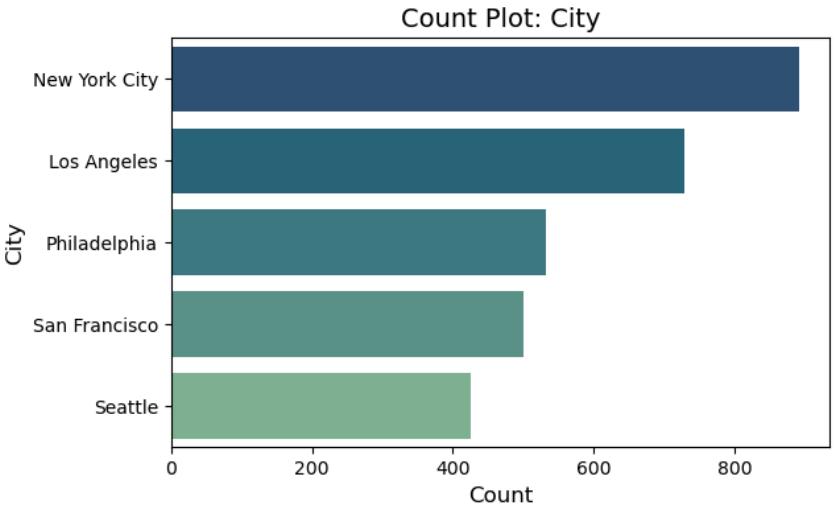
- Essential for Businesses & Individuals: Office supplies (e.g., pens, paper, notebooks) are used by both companies and individuals, ensuring consistent demand.
- Lower Price per Item → More Units Sold: Compared to furniture or electronics, office supplies are relatively cheap, so people buy them in larger quantities.
- Frequent Restocking: Unlike furniture or tech products, office supplies run out quickly, requiring frequent repurchasing.
- Schools & Institutions as Buyers: Schools, colleges, and government offices consistently buy office supplies.

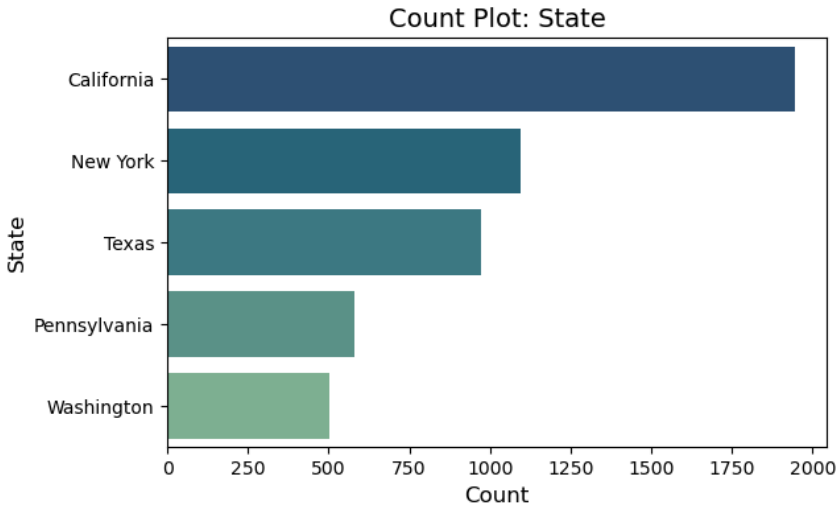
In [153…

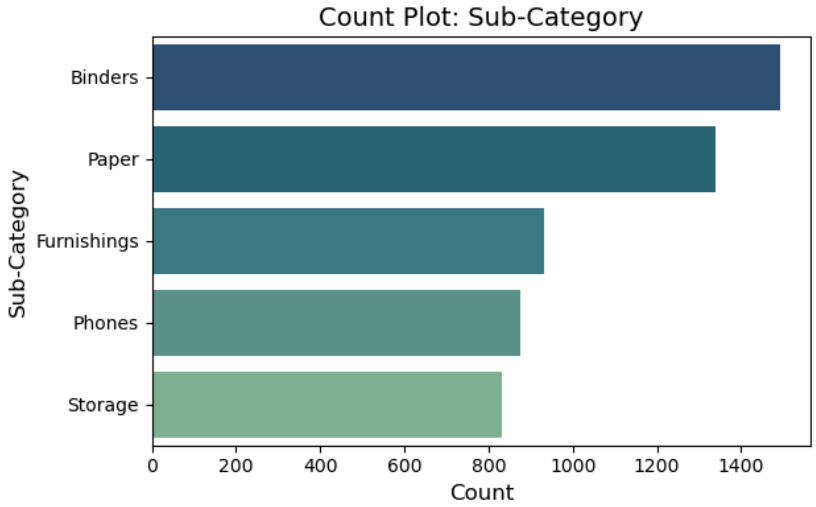
```
cat_cols = ['City','State','Sub-Category']
```

In [154…

```
plt.figure(figsize=(13, 8))

for i, col in enumerate(cat_cols):
    ax = plt.subplot(2, 2, i+1)

    sns.countplot(data=df, y=col,
                  order=df[col].value_counts().head().index,
                  palette= 'crest_r')

    ax.set_title(f"Count Plot: {col}", fontsize=14)
    ax.set_xlabel("Count", fontsize=12)
    ax.set_ylabel(col, fontsize=12)

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







Observations

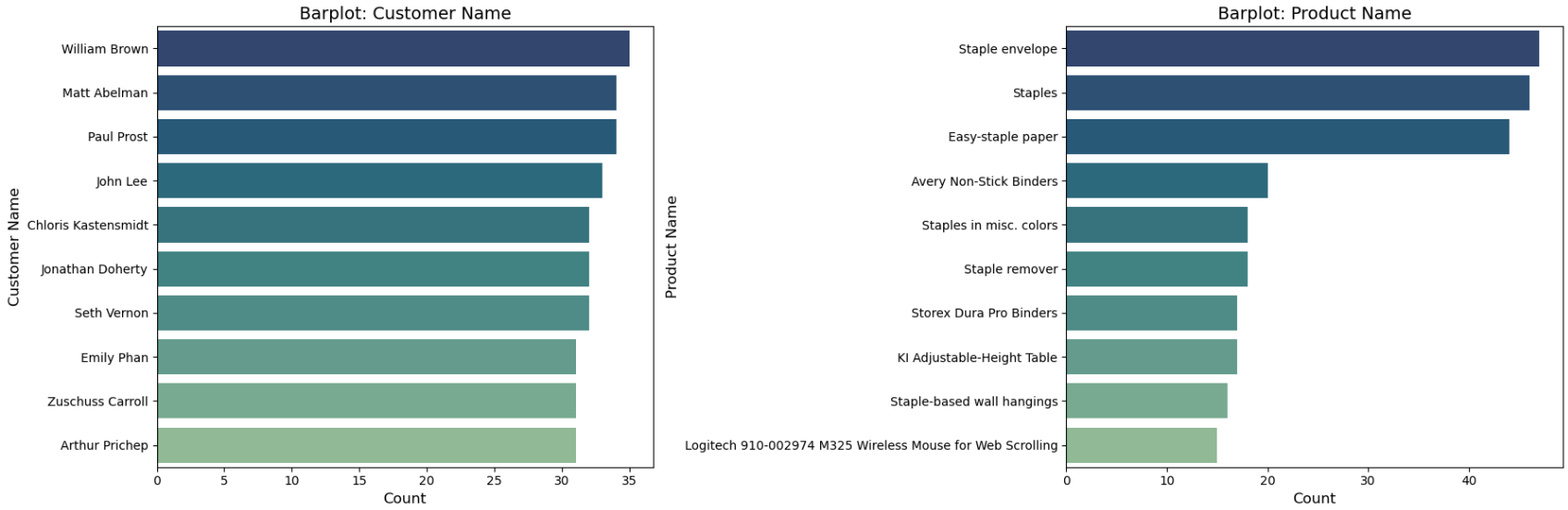
- Most of the `Orders` from `New York City` and `Los Angeles`.
- Most of the `Orders` from `California and New York state`.
- Most of the `Ordered` sub-categories are `Binders` and `Paper`.

Reasons

- The `New York City` and `Los Angeles` are the most populated cities in USA. So, the demand for `Binders` and `Paper` will be more.
- Both cities have large business districts, corporations, and startups that frequently order office supplies.
- These cities are major centers for online shopping, and residents are more likely to purchase online compared to smaller towns.

In [155…
```
customer = ['Customer Name','Product Name']
```

In [156…
```
plt.figure(figsize=(18, 6))

for i, col in enumerate(customer):
    ax = plt.subplot(1,2,i+1)
    sns.barplot(y=df[col].value_counts().head(10).index,
                x=df[col].value_counts().head(10).values, ax=ax,
                palette='crest_r')

    ax.set_title(f"Barplot: {col}", fontsize=14)
    ax.set_xlabel("Count", fontsize=12)
    ax.set_ylabel(col, fontsize=12)

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



Observations

- `William Brown to Arthur Prichep` these 10 customers are the top buyers.
- The majority of purchases belong to `Paper and Office/School Work products.`
- These include categories like `Office Supplies, sub-categories like Paper, and specific products related to school and office work.`

Bivariate Analysis

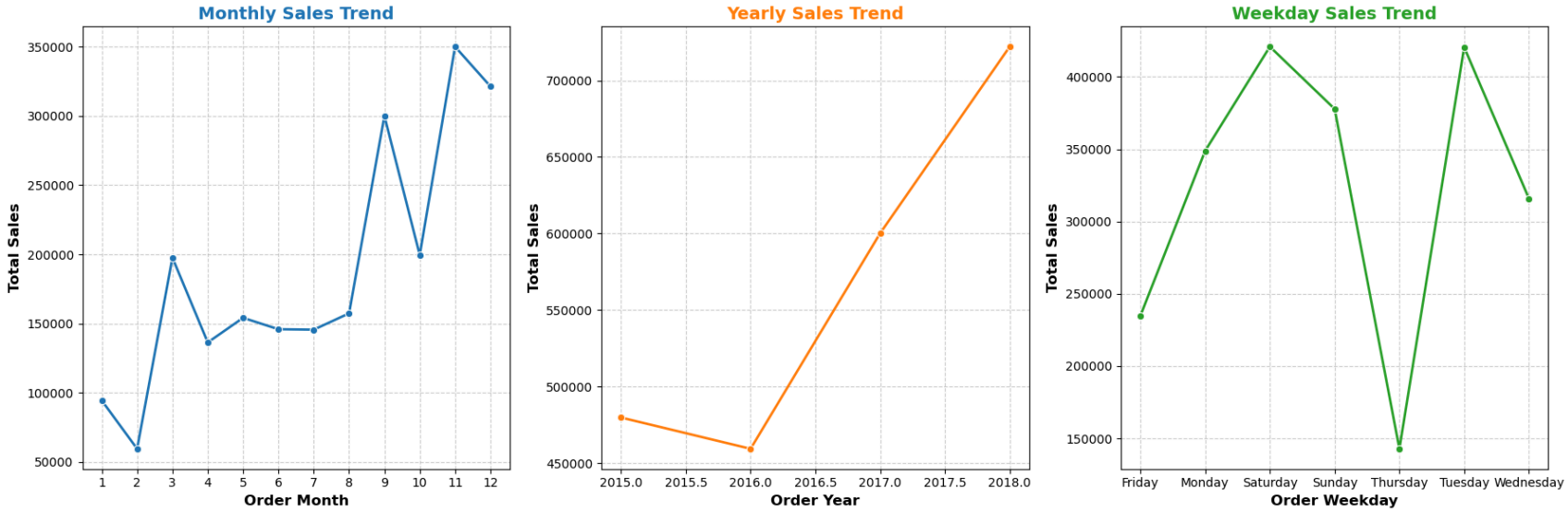
In [157…
```
total_sales =df['Sales'].sum().round(2)
print(f'Total Sales: ${total_sales}')
```

```
Total Sales: $2261536.78
```

Sales vs. Order Year/Month/Weekday Trend

In [158…
```
monthly_sales = df.groupby('Order Month')['Sales'].sum().reset_index()
yearly_sales = df.groupby('Order Year')['Sales'].sum().reset_index()
weekday_sales = df.groupby('Order Weekday')['Sales'].sum().reset_index()

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

colors = ['#1f77b4', '#ff7f0e', '#2ca02c']

# Monthly Sales Trend
sns.lineplot(x='Order Month', y='Sales', data=monthly_sales, marker='o', color=colors[0], linewidth=2, ax=axes[0])
axes[0].set_title('Monthly Sales Trend', fontsize=14, fontweight='bold', color=colors[0])
axes[0].set_xlabel('Order Month', fontsize=12, fontweight='bold')
axes[0].set_ylabel('Total Sales', fontsize=12, fontweight='bold')
axes[0].set_xticks(range(1, 13))
axes[0].tick_params(axis='x')
axes[0].grid(True, linestyle='--', alpha=0.6)

# Yearly Sales Trend
sns.lineplot(x='Order Year', y='Sales', data=yearly_sales, marker='o', color=colors[1], linewidth=2, ax=axes[1])
axes[1].set_title('Yearly Sales Trend', fontsize=14, fontweight='bold', color=colors[1])
```

```
axes[1].set_xlabel('Order Year', fontsize=12, fontweight='bold')
axes[1].set_ylabel('Total Sales', fontsize=12, fontweight='bold')
axes[1].tick_params(axis='x')
axes[1].grid(True, linestyle='--', alpha=0.6)

# Weekday Sales Trend
sns.lineplot(x='Order Weekday', y='Sales', data=weekday_sales, marker='o', color=colors[2], linewidth=2, ax=axes[2])
axes[2].set_title('Weekday Sales Trend', fontsize=14, fontweight='bold', color=colors[2])
axes[2].set_xlabel('Order Weekday', fontsize=12, fontweight='bold')
axes[2].set_ylabel('Total Sales', fontsize=12, fontweight='bold')
axes[2].tick_params(axis='x')
axes[2].grid(True, linestyle='--', alpha=0.6)

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



Observations

Seasonal Sales Trends (Monthly)

- Sales peak between September and December every year.
- Possible reasons: Holiday season, Black Friday, Christmas sales boost.

- Recommendation: Focus marketing campaigns and stock management for Q4.

Yearly Growth Analysis

- Sales volume increased rapidly after 2016.
- Indicates business expansion, better customer acquisition, or market growth.
- Suggestion: Analyze post-2016 changes (new products, promotions, regional growth).

Weekly Sales Performance

- Peak sales days: Saturday & Tuesday.
- Moderate sales: Monday & Sunday.
- Likely reasons: Weekend shopping habits, online sale spikes.
- Suggestion: Optimize promotions and advertising on peak days.

Need Improvement

Monthly Sales Trend

- Sales are lower from January to August.
- Seasonal effect: Fewer major shopping events.
- Post-holiday slump: Reduced spending after December peak.

Suggestions

- Launch off-season promotions in Q1-Q3.
- Introduce flash sales or discount bundles in slow months.
- Leverage back-to-school sales in July-August to boost revenue.

Weekly Sales Performance

- Friday, Tuesday, and Wednesday have lower sales.
- Midweek slump: People are focused on work/school.
- Less weekend impulse buying.

Suggestions

- Offer weekday discounts (e.g., "Tuesday Saver Deals").
- Target ads on social media midweek to drive engagement.
- Introduce "Flash Friday" deals to improve Friday sales.

Yearly Sales Growth

- Sales volume in 2015 & 2016 was too low.
- Early-stage business growth.
- Limited marketing or fewer product offerings.

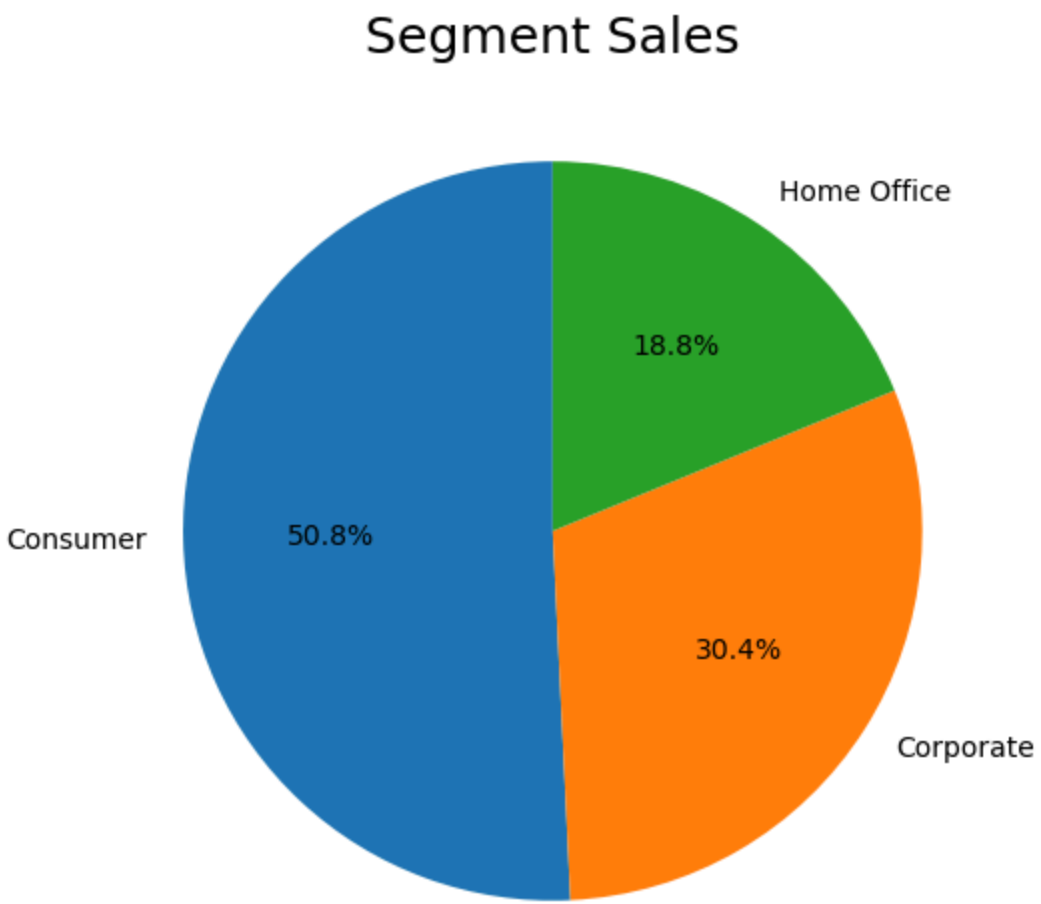
Suggestions

- Analyze what changed after 2016 to improve sales.
- Identify successful marketing strategies post-2016.
- Use historical data to find demand patterns and optimize future strategies.

Sales VS Segment

In [159…

```
plt.figure(figsize=(12,6))
plt.subplot(1,1,1)

segment = df.groupby('Segment')['Sales'].sum()

plt.pie(segment, labels=segment.index, autopct='%1.1f%%', startangle=90)
plt.title('Segment Sales', fontsize=18)
plt.show()
```



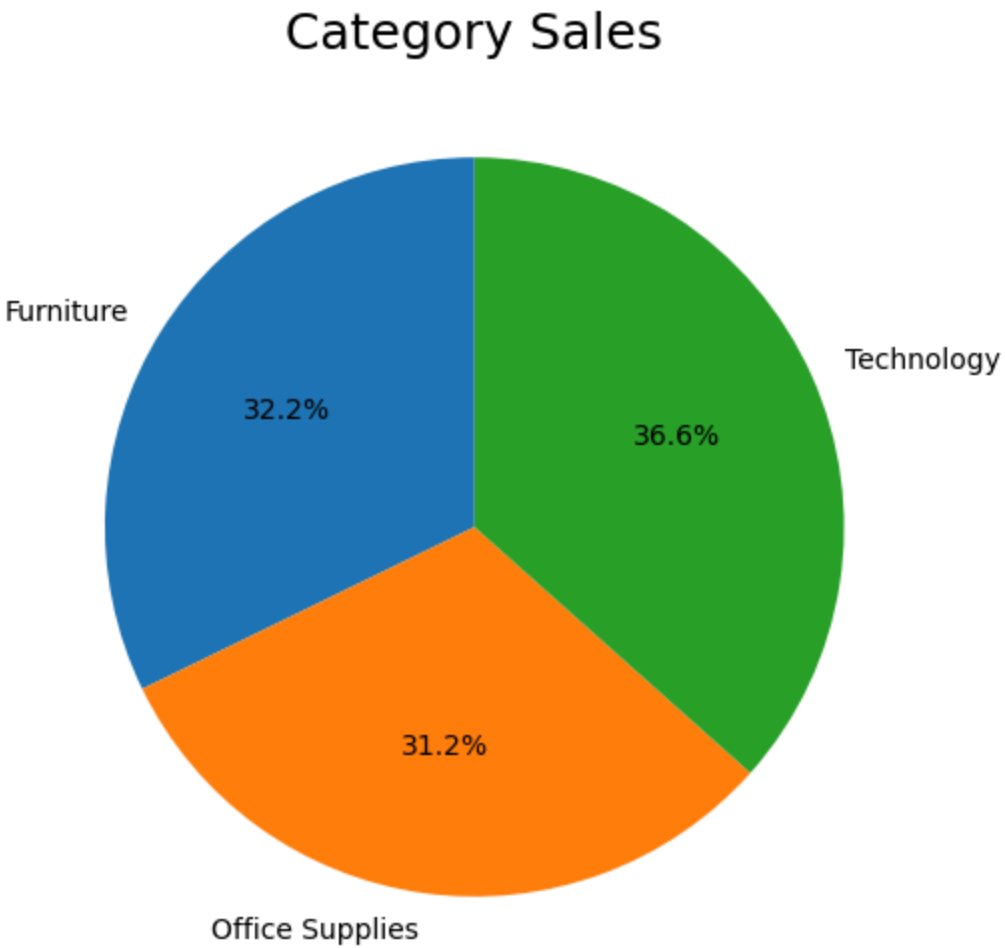
Observations

- The `Consumer segment` contributes `58.8%` of total sales, making it the dominant segment in terms of revenue.
- The `Corporate segment` accounts for `30.4%` of total sales, indicating a significant but lower contribution compared to the Consumer segment.
- The `Home Office segment` generates `18.8%` of total sales, representing the smallest share among the three segments.

Sales VS Category

In [160…

```
plt.figure(figsize=(12,6))
plt.subplot(1,1,1)

category = df.groupby('Category')['Sales'].sum()

plt.pie(category, labels=category.index, autopct='%1.1f%%', startangle=90)
plt.title('Category Sales', fontsize=18)
plt.show()
```



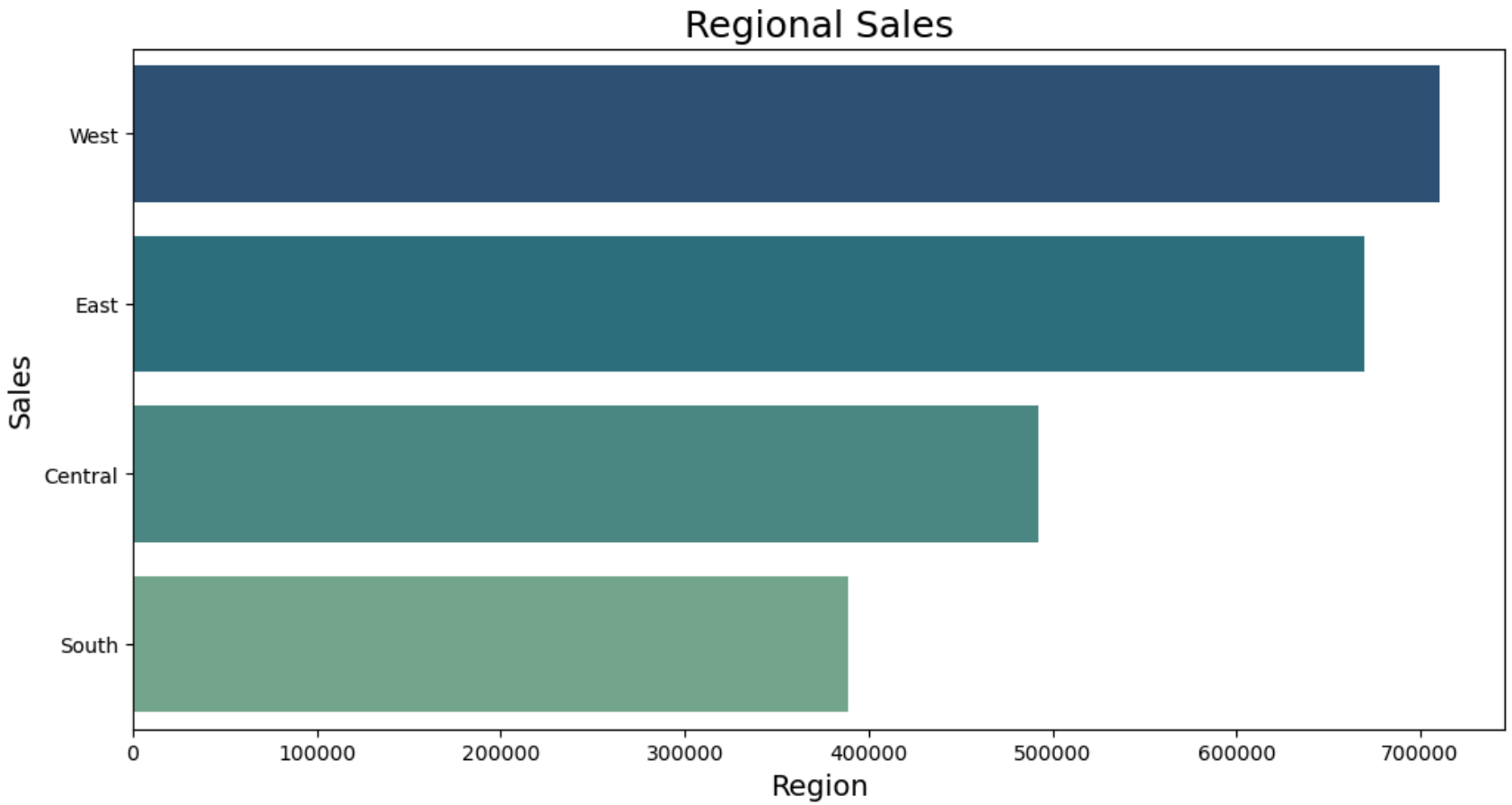
Observations

- The `Technology category` generates the highest sales, contributing `36.6%` of total revenue, indicating strong demand for tech-related products.
- The `Office Supplies category` accounts for the lowest sales at `31.2%`, despite having a high number of orders, suggesting lower-priced items or smaller profit margins in this category.

Sales VS Region

In [161…

```
plt.figure(figsize=(12,6))
plt.subplot(1,1,1)

region = df.groupby('Region')['Sales'].sum().sort_values(ascending=False)
sns.barplot(x=region.values, y=region.index, palette = 'crest_r')

plt.xlabel('Region', fontsize=14)
plt.ylabel('Sales', fontsize=14)
plt.title('Regional Sales', fontsize=18)
plt.show()
```



Observations

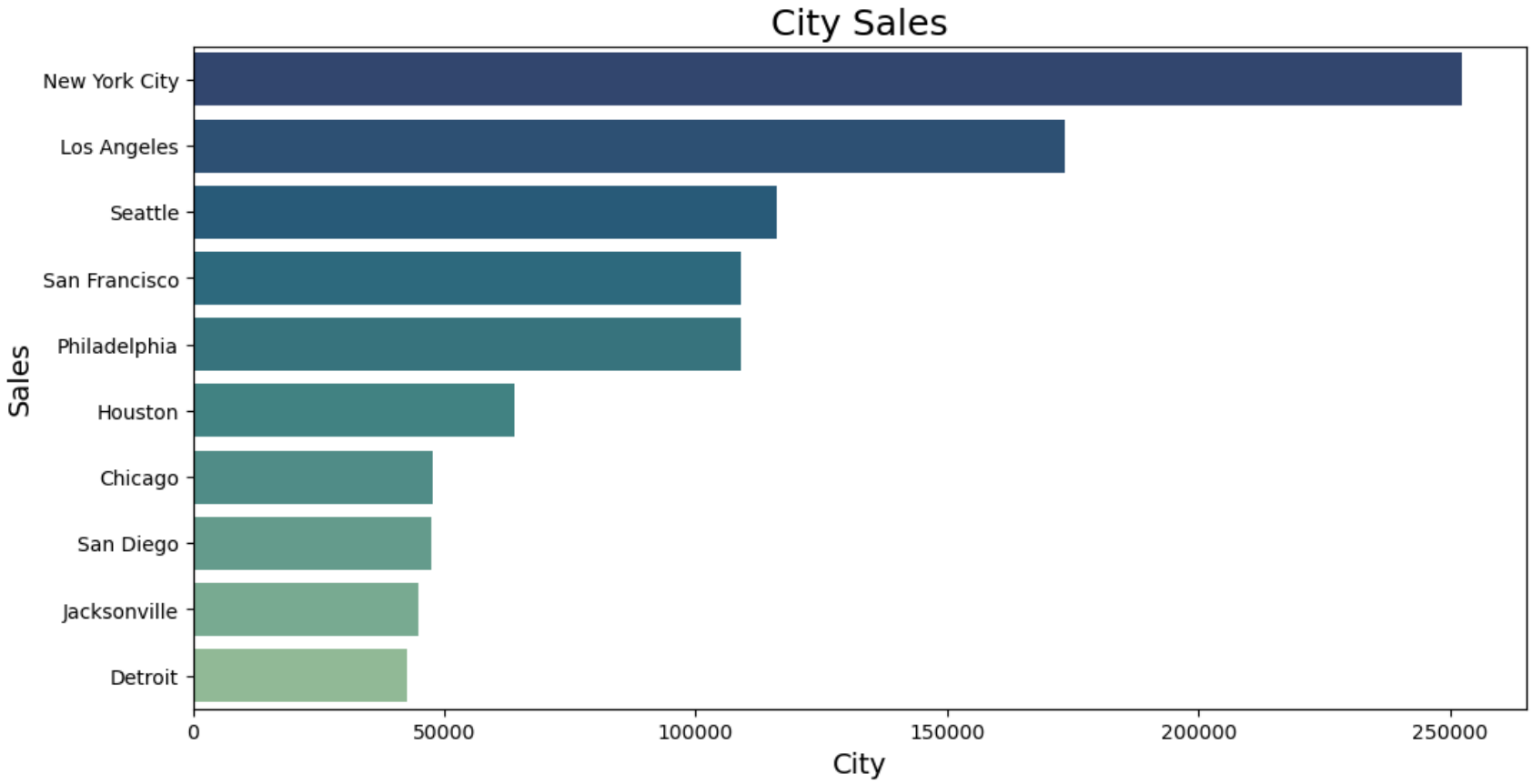
- The `West region` leads in sales, making it the highest-performing region in terms of revenue.
- The `East region` also performs well, with sales closely trailing the West, indicating strong market demand in both regions.

Sales VS City

In [162…

```
city = df.groupby('City')['Sales'].sum().sort_values(ascending=False).head(10)
plt.figure(figsize=(12,6))
plt.subplot(1,1,1)

city = df.groupby('City')['Sales'].sum().sort_values(ascending=False).head(10)
```

```
sns.barplot(x=city.values, y=city.index, palette = 'crest_r')

plt.xlabel('City', fontsize=14)
plt.ylabel('Sales', fontsize=14)
plt.title('City Sales', fontsize=18)
plt.show()
```



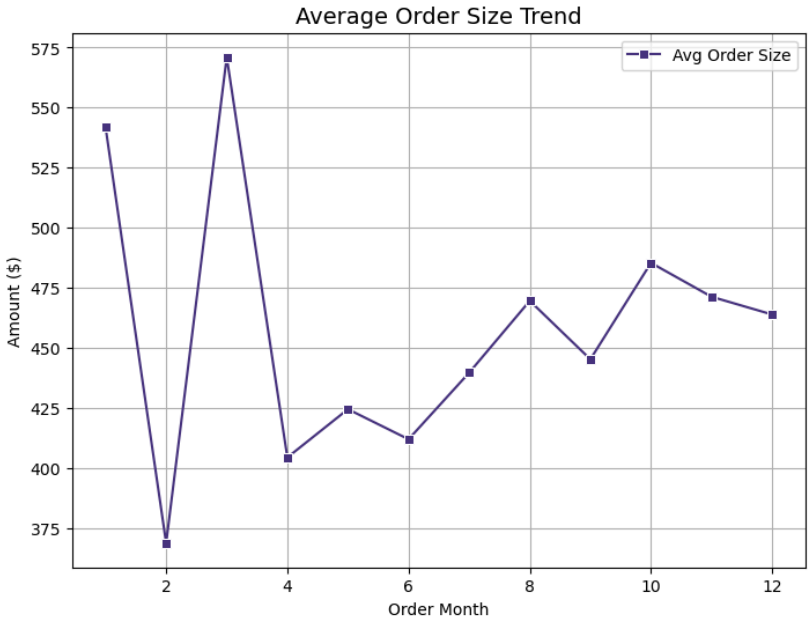
Observations

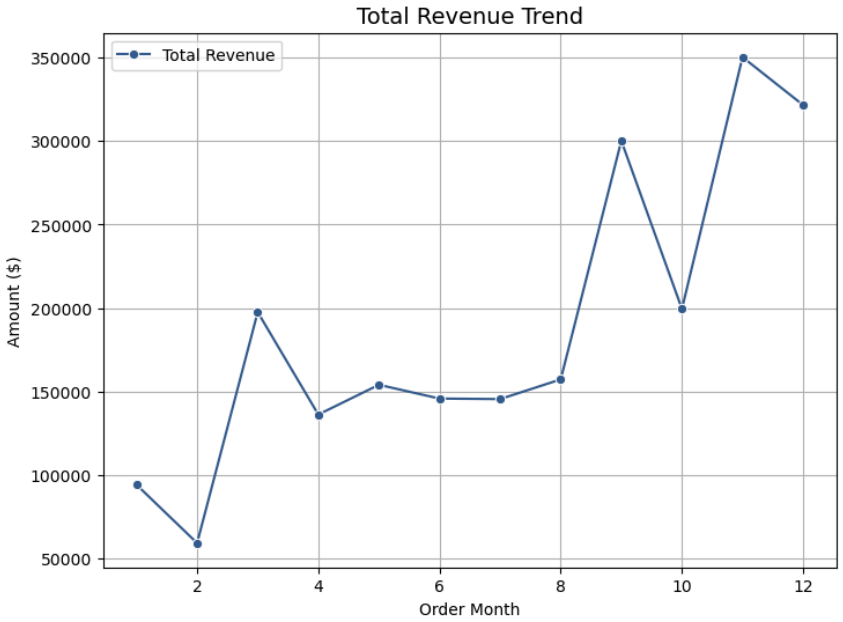
- `New York and Los Angeles` record the highest sales, making them the top-performing cities.
- These cities `outperform` others, indicating strong market demand and a high concentration of sales activity.

Average Order size & Total Revenue Trend

In [163…

```
monthly_sales = df.groupby('Order Month').agg(Revenue=('Sales', 'sum'), Order_Count=('Order ID', 'nunique')).reset_ir

monthly_sales['Avg_Order_Size'] = monthly_sales['Revenue'] / monthly_sales['Order_Count']

fig, axes = plt.subplots(1, 2, figsize=(18, 6))
colors = sns.color_palette('viridis')

sns.lineplot(data=monthly_sales, x='Order Month', y='Avg_Order_Size', marker='s', label="Avg Order Size", ax=axes[0],

axes[0].set_title("Average Order Size Trend", fontsize=14)
axes[0].set_xlabel("Order Month")
axes[0].set_ylabel("Amount ($)")
axes[0].legend()
axes[0].grid(True)

sns.lineplot(data=monthly_sales, x='Order Month', y='Revenue', marker='o', label="Total Revenue", color = colors[1],
axes[1].set_title("Total Revenue Trend", fontsize=14)
axes[1].set_xlabel("Order Month")
axes[1].set_ylabel("Amount ($)")
axes[1].legend()
axes[1].grid(True)

plt.show()
```





Observations

- The number of orders is highest in `Q1`, but total sales remain `low`, suggesting that most purchases during this period involve lower-value items.
- Orders volume remains moderate in `Q3 and Q4`, but sales peak in `Q4`, indicating that customers tend to purchase higher-value items towards the end of the year.

Insights & Recommendations

Order Trends

- Peak Order Period: September to December, driven by holiday shopping, Black Friday, and Christmas sales.
- Top Order Days: 2nd–5th, 8th, 11th, 20th–23rd, 26th, mainly due to paydays and promotions.
- Busiest Order Days: Tuesdays, Saturdays, Sundays, and Mondays, influenced by weekend shopping and corporate purchases.

Recommendation:

- Focus marketing and stock replenishment from September to December.
- Offer mid-week discounts to boost Tuesday sales.

Shipping & Customer Segments

- Most Popular Shipping Mode: Standard Class, likely due to affordability.
- Highest Orders by Region: West region, possibly due to higher population and business hubs.
- Leading Customer Segment: Consumer segment (58.8%), driven by frequent, smaller purchases.

Recommendation:

- Optimize logistics for the West region and Consumer segment to enhance delivery efficiency.

Product Performance

- Best-Selling Category: Technology (36.6% of sales), despite Office Supplies having more orders.
- Top Ordered Products: Binders and Paper, mainly purchased in New York & Los Angeles.

Recommendation:

- Upsell tech accessories alongside Office Supplies to boost revenue.

Sales Performance

- Peak Sales Period: Q4 (Sept–Dec), aligning with holiday demand.
- Q1 Sales Dip: Despite higher order volume, sales remain low, indicating low-value purchases.

Recommendation:

- Run promotions in Q1 (e.g., back-to-school and business restocking discounts) to boost revenue.

Geographic Insights

- Top Cities by Sales: New York & Los Angeles, reflecting high demand and consumer density.
- Top States by Orders: California & New York.

Recommendation:

- Target these cities for regional promotions and warehouse optimization.

Sales Growth Patterns

- Rapid Growth After 2016: Likely due to business expansion and marketing efforts.
- Low Sales on Wednesdays & Fridays: Possibly due to midweek slump.

Recommendation:

- Introduce Flash Friday or Midweek Deals to counter low sales days.