

PRESENTATION OF **EDA** OF SUPER STORE **SALES DATA**

This analysis explores the Super Store Dataset to uncover insights about sales, profit, and trends across various categories and regions. Using Python and popular data analysis libraries such as Pandas, Matplotlib, and Seaborn, we delve into the data to identify key patterns and trends.

OVERALL SUMMARY

The analysis reveals positive growth trends in orders, sales, and profit over the years. The Consumer segment is the most significant contributor to both sales and profit.

Geographically, California, New York, and Texas are the top-performing states. Office Supplies dominate in terms of order volume and stock sales, but Technology leads in profitability. Standard Class shipping is the most utilized shipping method. This comprehensive insight can guide strategic decisions to further enhance performance and address areas with potential for improvement.

DATA CLEANING AND PREPARATION

Key Points:-

- **Missing Values:** The dataset was checked and confirmed to have no missing values, ensuring data integrity and completeness.
- **Redundancy:** The 'Row_ID' column, which serves only as an index or sequence identifier, was removed from the dataset.
- **Feature Engineering:** Extracted the year from the 'Order Date' column and created new columns 'Order Year' and 'Ship Year' to enhance the dataset with additional temporal features.
- **Conversion :**'Sales' and 'Profit' columns were converted to integers for accurate calculations and consistent data representation.

EDA

Starting Univariate Analysis

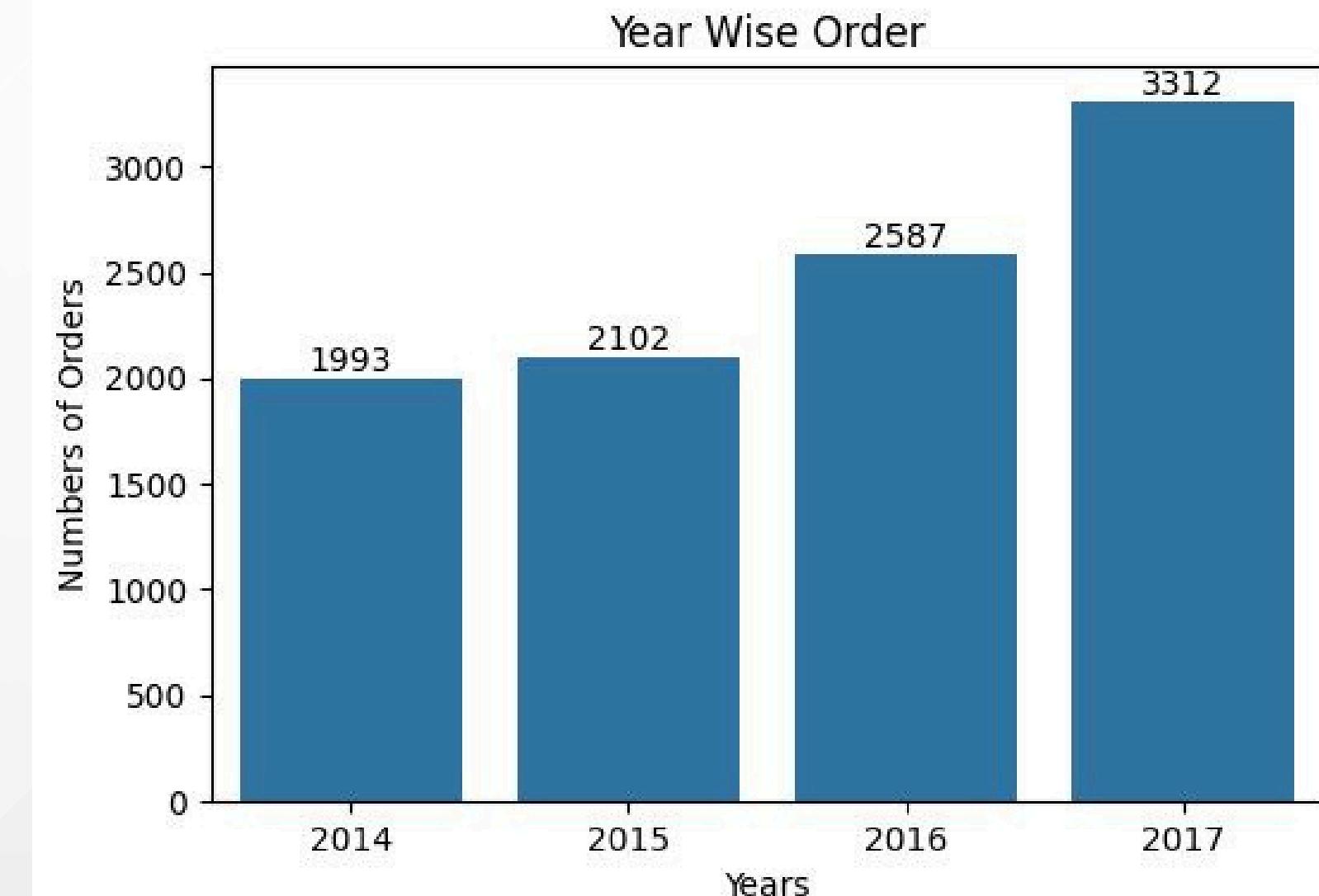
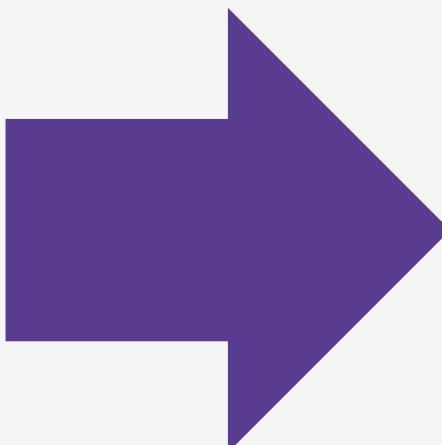
- **Year-Wise Orders:** Identified trends in the number of orders over the years, highlighting any growth or decline in sales.
- **Category-Wise Orders:** Analyzed order distribution across different product categories, revealing which categories are popular among customers.
- **Consumer Segment Wise Orders:** Explored order patterns based on consumer segments, providing insights into customer preferences and behavior.
- **Order Shipped By Company:** Examined the distribution of orders shipped by the company, indicating the company's operational efficiency and workload.
- **State-Wise Orders:** Investigated order distribution across different states, uncovering geographic trends and demand variations.

YEAR WISE ORDERS :-

The data indicated a positive trajectory, with the number of orders increasing steadily from year to year. This growth trend suggests successful business expansion and customer acquisition efforts.

```
# Check the Year Wise Order Trends..
plt.figure(figsize = (6 , 4))
ax = sns.countplot(x = 'Order Year' , data = df )
plt.title('Year Wise Order')
plt.xlabel('Years')
plt.ylabel('Numbers of Orders')

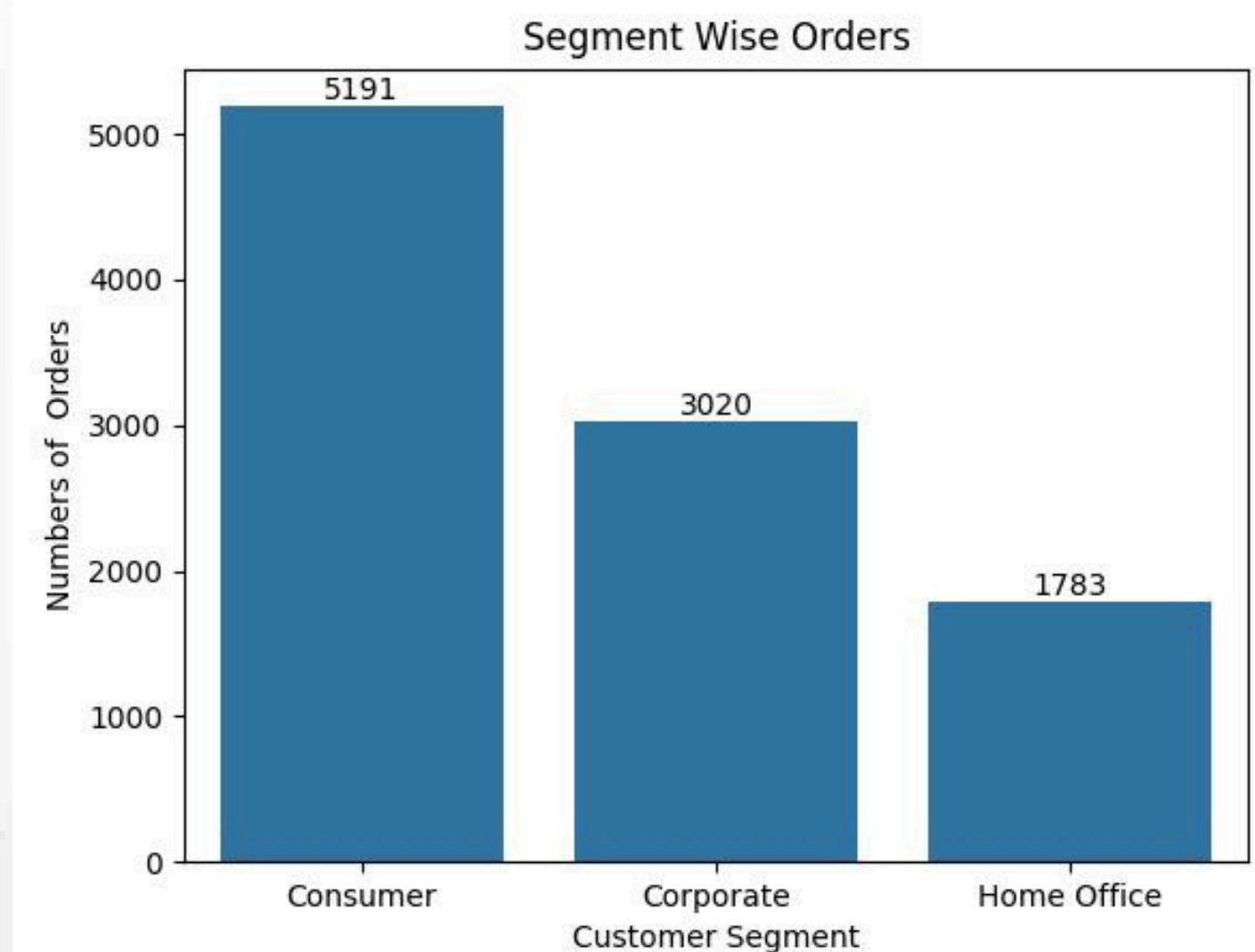
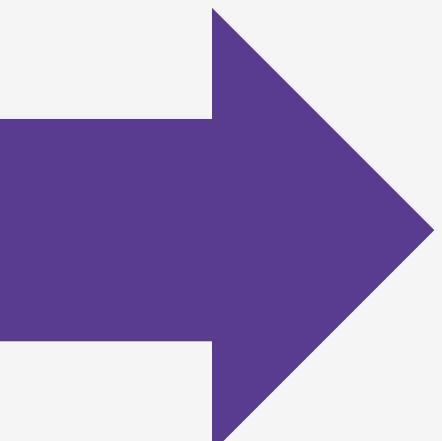
# For Numbers in the Bars
for i in ax.containers :
    ax.bar_label(i)
```



SEGMENT WISE ORDERS :-

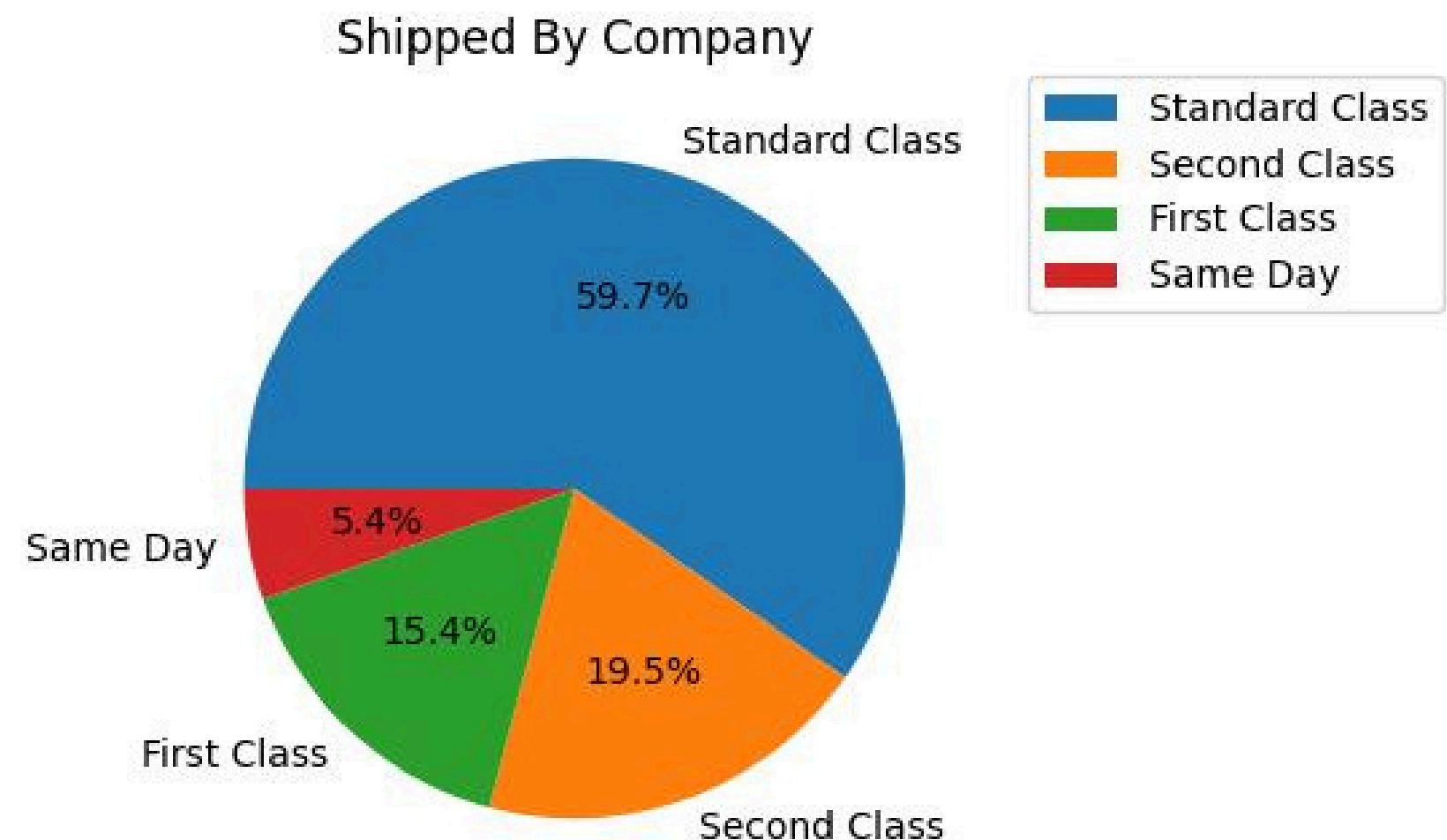
The Consumer segment generates the highest number of orders, followed by the Corporate segment, with the Home Office segment contributing the fewest orders.

```
# See the Order Counts By Segment Wise  
ax = sns.countplot(x = 'Segment' , data = df)  
plt.title('Segment Wise Orders')  
plt.xlabel('Customer Segment')  
plt.ylabel('Numbers of Orders')  
  
for unique in ax.containers :  
    ax.bar_label(unique)
```



SHIPPED BY COMPANY

"Standard Class shipping is the most preferred option, accounting for 59.7% of orders, followed by Second Class at 19.5% and First Class at 15.4%. Same Day shipping is the least used, handling only 5.4% of orders."

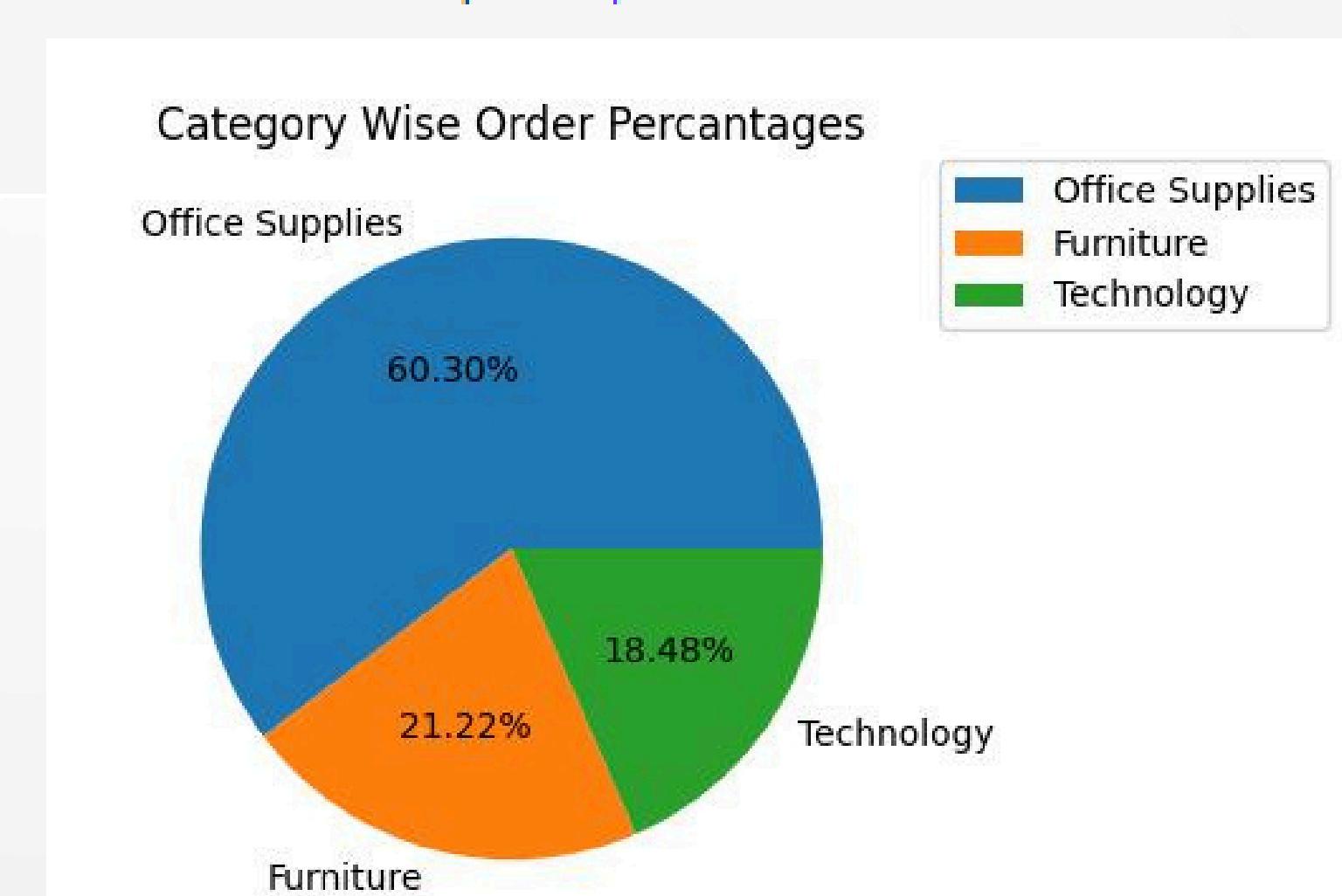


```
plt.figure(figsize = (7, 4))
plt.pie(df['Ship Mode'].value_counts().values , labels = df['Ship Mode'].value_counts().index
        , autopct = '%.1f%%' , counterclock=False, startangle=180)
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0)
plt.title('Shipped By Company')
plt.savefig('Shipped_By_Company.jpg')
```

CATEGORY WISE ORDERS

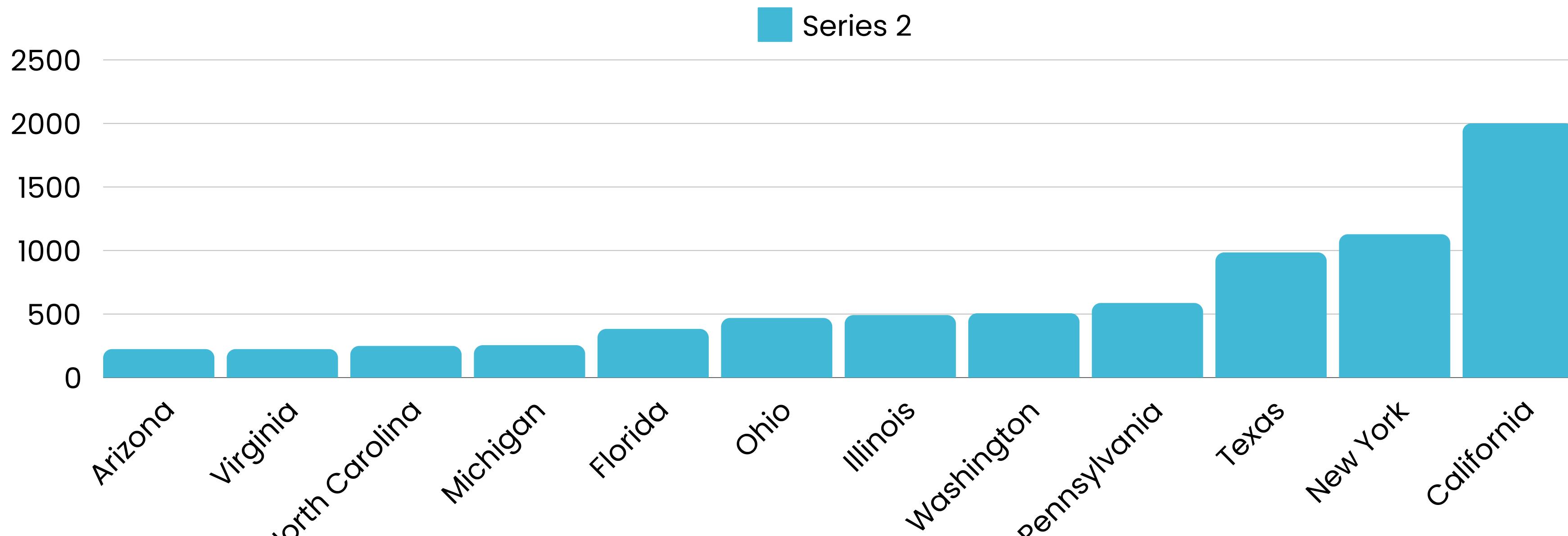
```
# Make the Pie Plot...
plt.figure(figsize = (7,4))
plt.pie(df['Category'].value_counts().values
, labels = df['Category'].value_counts().index , autopct = '%0.2f%%')
plt.title("Category Wise Order Percentages")
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0)
plt.savefig('Category_Wise_Order.jpg')
plt.show()
```

"Office Supply is the dominant category, comprising 60.30% of orders, followed by Furniture with 21.22%. The remaining orders are sourced from the Technology category."



STATE-WISE ORDERS :-

The highest number of orders are coming from California, followed by New York and Texas, making them the top three states with the highest order volumes.



Starting Bivariate Analysis

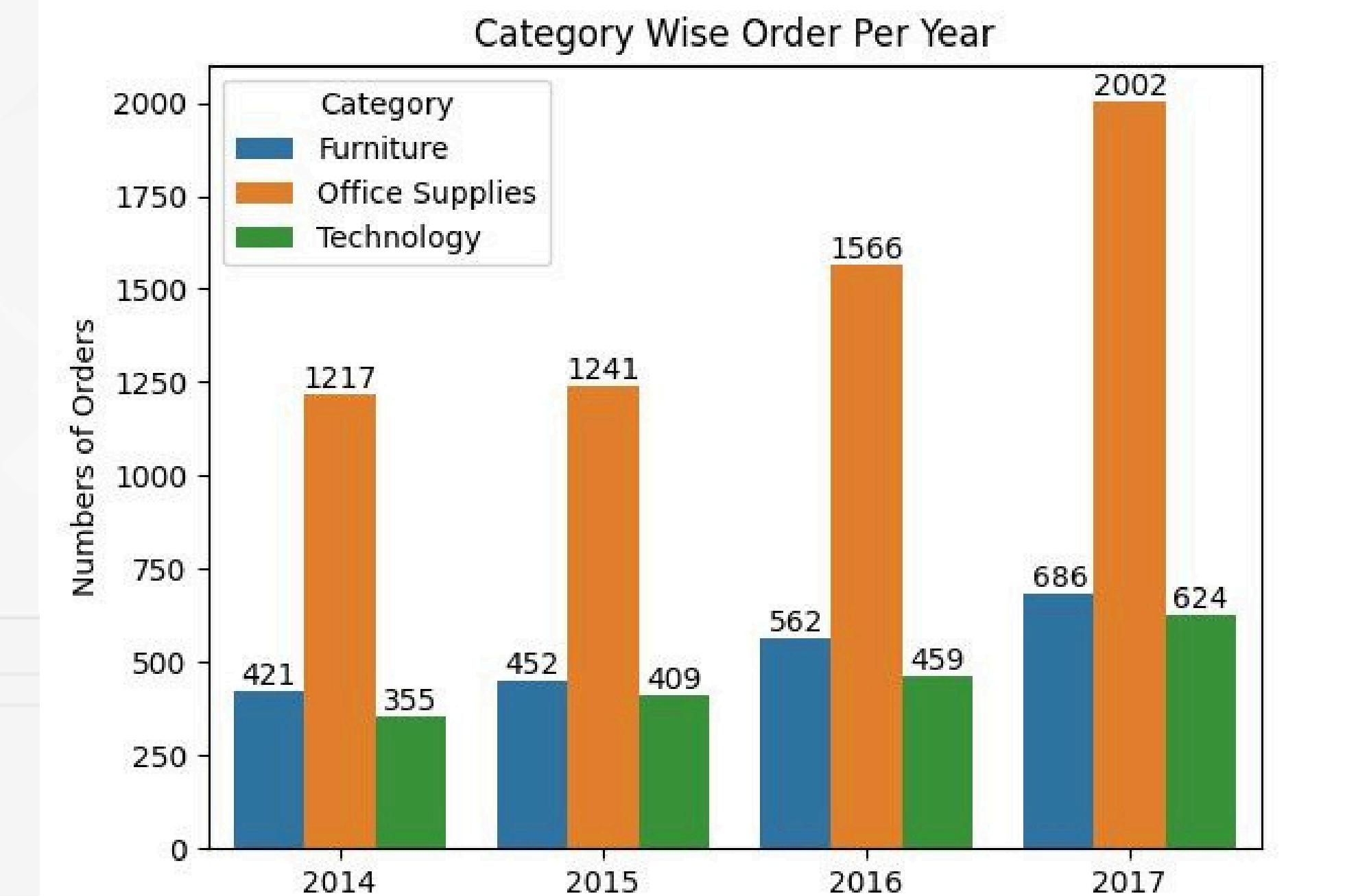
- **Number of Orders per Category Year-Wise:** Distribution of orders across product categories (Office Supply, Furniture, Technology) analyzed year by year.
- **Sales per Year:** Total sales figures examined annually.
- **Profit per Year:** Trends in profitability observed year over year..
- **Total Stock Sales per Year:** Distribution of stock sales across product categories examined.
- **Sales vs. Profit per Years:** Relationship between sales and profit analyzed annually.

NUMBER OF ORDERS PER CATEGORY

YEAR-WISE

Over the years, it has been observed that all three categories—Office Supplies, Furniture, and Technology—experience a steady increase in the number of orders. However, among these categories, Office Supplies consistently garners the highest volume of orders. This trend indicates a robust and growing demand for office supplies, surpassing the order numbers of both the Furniture and Technology categories.

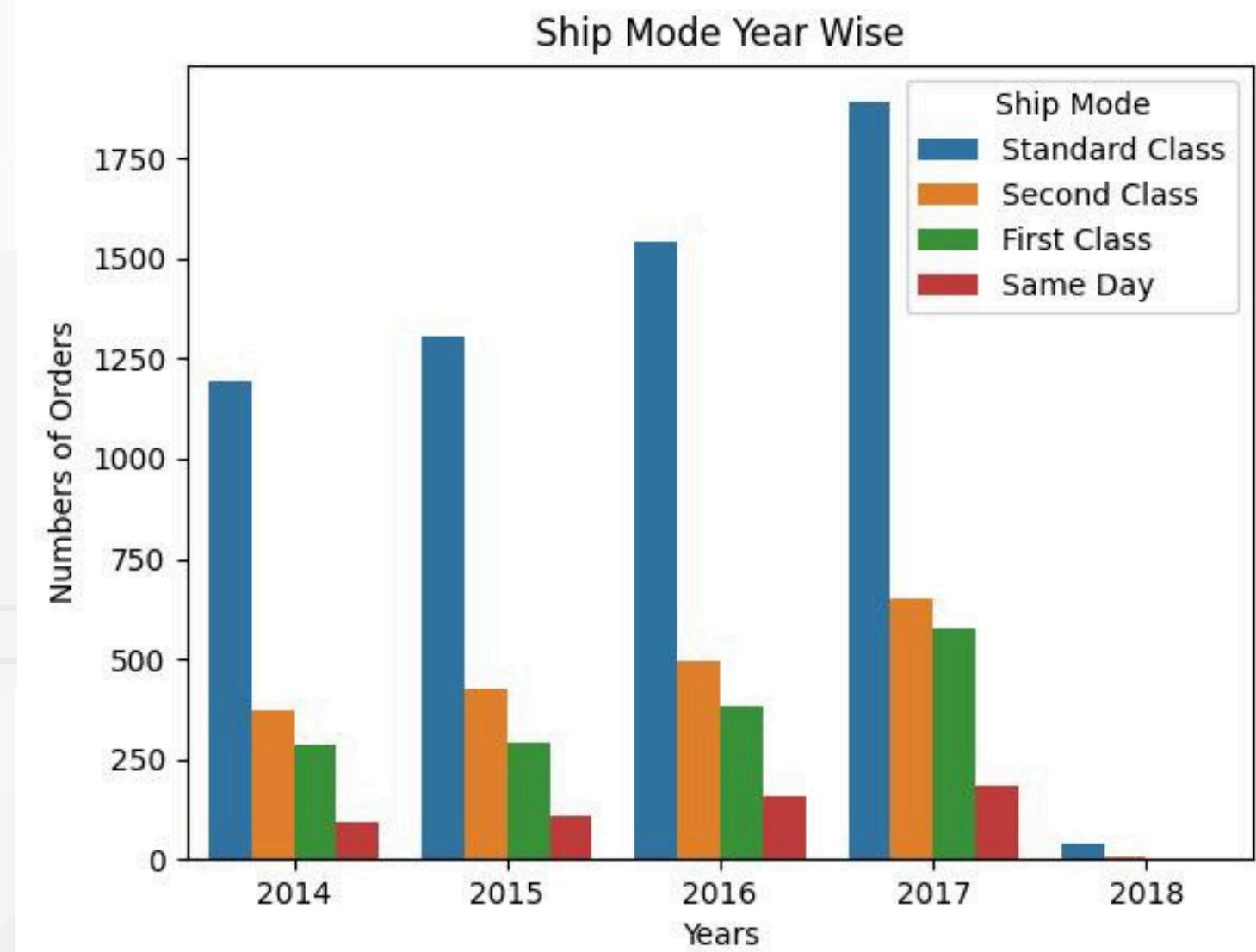
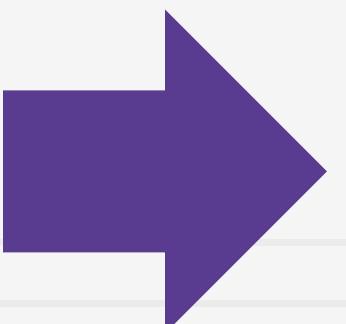
```
# Ship Mode Year Wise
sns.countplot(x = 'Ship Year' , data = df , hue = 'Ship Mode')
plt.title('Ship Mode Year Wise')
plt.xlabel('Years')
plt.ylabel('Numbers of Orders')
plt.savefig('Ship_Mode_Year_Wise.jpg')
plt.show()
```



SHIP MODE YEAR WISE

Over time, our shipping preferences have evolved from Standard to Second Class, and then to First Class, with Same Day Shipping being the least favored. This trend reflects a gradual inclination towards faster delivery options while balancing cost and efficiency

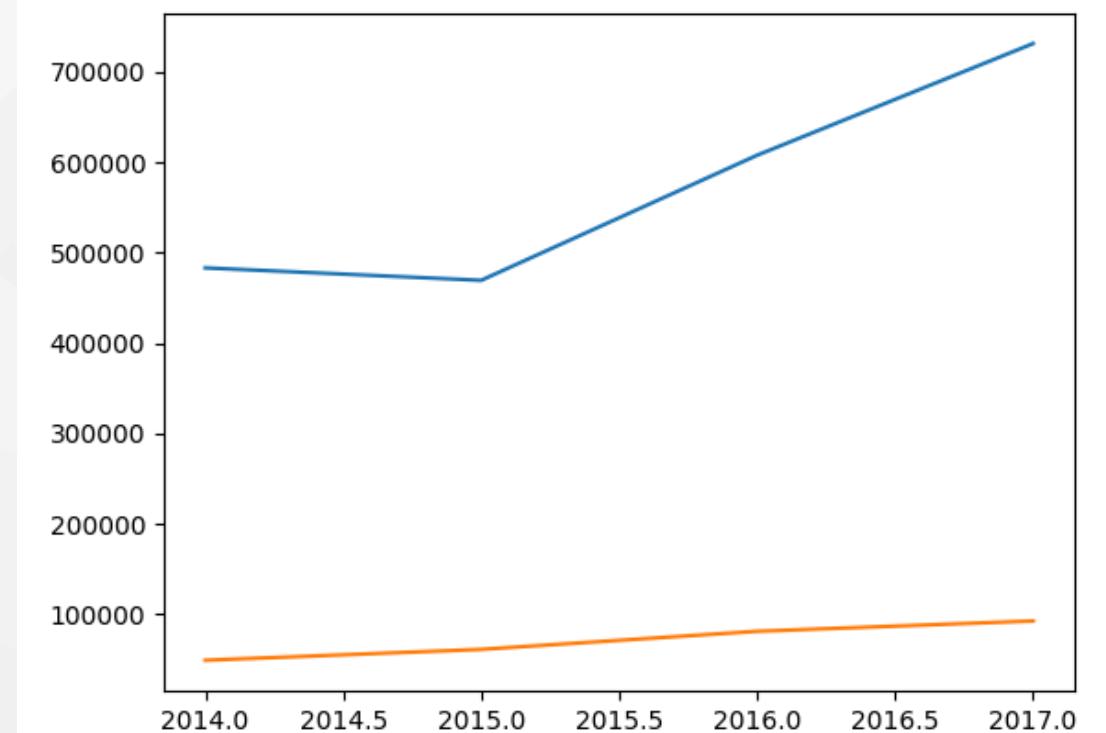
```
# Ship Mode Year Wise
sns.countplot(x = 'Ship Year' , data = df , hue = 'Ship Mode')
plt.title('Ship Mode Year Wise')
plt.xlabel('Years')
plt.ylabel('Numbers of Orders')
plt.savefig('Ship_Mode_Year_Wise.jpg')
plt.show()
```



SALES VS PROFIT

Our profit margin is considerably low compared to sales volume; selling 450,000 items only yields a profit of \$40,000, which is unsatisfactory. Increasing our profit margin is imperative for our business's sustainability and growth.

```
# Sales Vs Profit Per Year
sales_vs_profit = df.groupby('Order Year')[['Sales' , 'Profit']].sum().reset_index()
plt.plot(sales_vs_profit['Order Year'] , sales_vs_profit['Sales'] , label = 'Sales')
plt.plot(sales_vs_profit['Order Year'] , sales_vs_profit['Profit'] , label = 'Profit')
plt.savefig('Sales_VS_Profit')
```

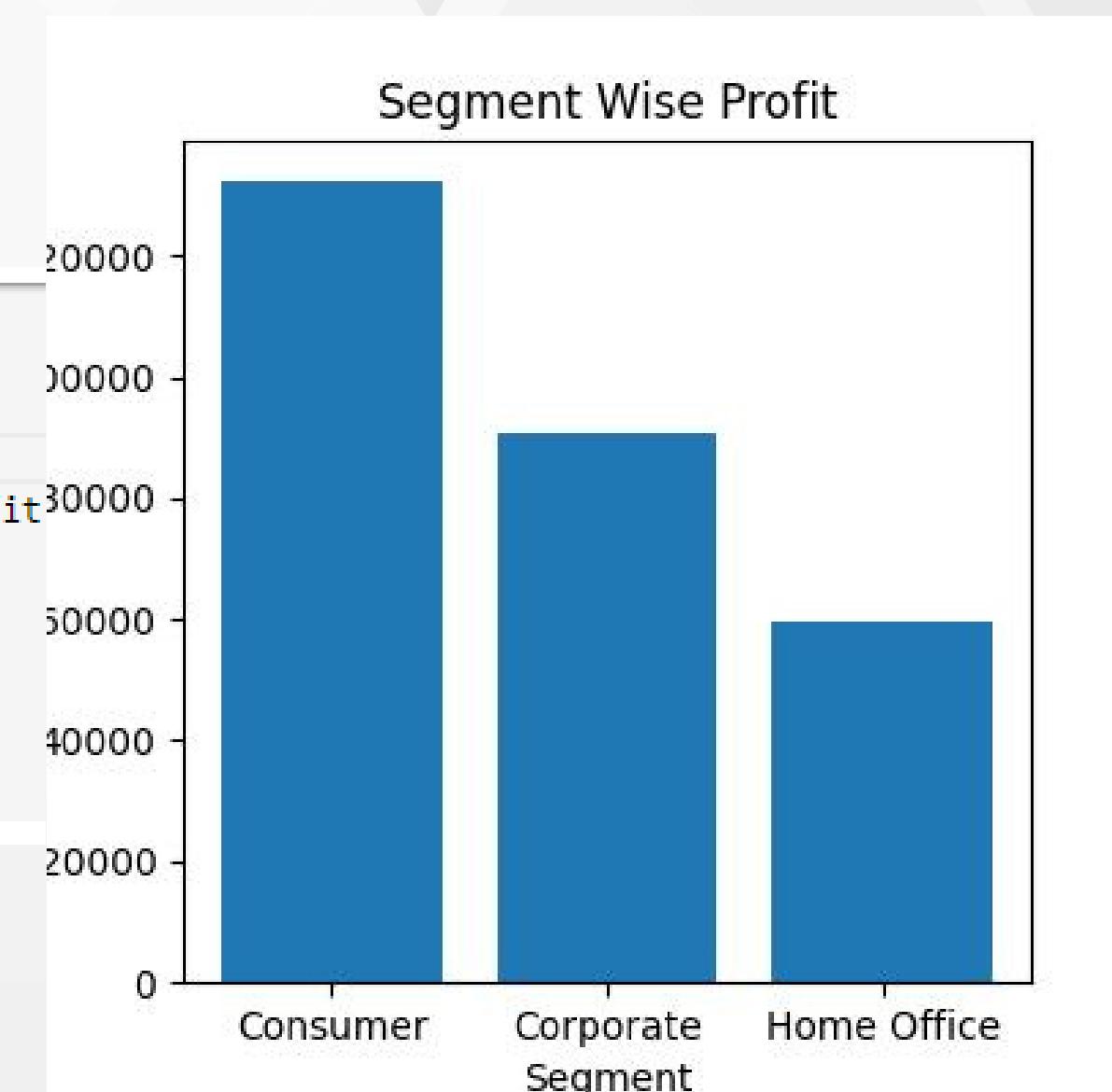


SEGMENT WISE PROFIT

The analysis reveals that the majority of profits stem from the Consumer Segment, followed by the Corporate Segment, with the Home Office Segment contributing the least. This observation underscores the importance of focusing on strategies to enhance profitability within the Home Office Segment while sustaining the momentum in the Consumer and Corporate Segments.

```
# Segment Wise Profit
seg_wise_profit = df.groupby('Segment')['Profit'].sum().reset_index()

plt.figure(figsize = (4, 4)) Loading...
plt.bar(x = seg_wise_profit['Segment'] , height = seg_wise_profit['Profit'] , data = seg_wise_profit)
plt.title('Segment Wise Profit')
plt.xlabel('Segment')
plt.ylabel('Profit')
plt.savefig('Segment_Wise_Profit.jpg')
plt.show()
```



CONCLUSIONS

- Positive growth trends in orders, sales, and profit over the years.
- The consumer segment is the most significant contributor to both sales and profit.
- Top-performing states: California, New York, and Texas.
- Office Supplies dominate in order volume and stock sales.
- The technology category leads to profitability.
- Standard Class is the most utilized shipping method.
- Most orders take 4 days to ship, followed by 5 and 6 days, negatively impacting customer satisfaction.
- Insights can guide strategic decisions to enhance performance, improve customer satisfaction, and address areas for improvement.