



Customer Sales Analysis Using Python

28.01.2025

Nikhil Kumar

Patna

Github:- https://github.com/Nikhil7439/Nikhil7439/blob/main/ecommerce_2.ipynb

Colab:-

https://colab.research.google.com/drive/1IE78_SmxjBhNjCHf4j4Omgw5s0JvE3ZE#scrollTo=ir-1o0poDjvL

Linkedin:- <https://www.linkedin.com/in/nikhil-jaiswal-050061277/>

Overview

This project focuses on analyzing customer data to derive actionable insights about sales performance, profitability, and customer behavior. By leveraging Python's powerful libraries such as Pandas, NumPy, and Plotly, the analysis uncovers key trends and highlights opportunities for improving business strategies.

Goals

The primary goal of this project is to **analyze customer data** to uncover key insights that can drive business growth and improve decision-making. Specifically, the project aims to:

1. **Identify High-Performing Customer Segments:** Understand which customer segments contribute the most to sales and profitability.
2. **Recognize Top Revenue-Generating Cities:** Highlight cities with the highest sales to prioritize marketing and resource allocation.
3. **Optimize Shipping Strategies:** Compare shipping types to determine the most profitable option and identify opportunities to improve operational efficiency.
4. **Provide Actionable Recommendations:** Deliver insights and suggestions to enhance overall sales, profitability, and customer satisfaction.
5. **Objective:** Clearly state the purpose of the project.
 - a. Example: *"The objective of this project is to analyze customer data to identify trends in sales, profits, and customer behavior to drive business insights."*
6. **Key Questions:** Mention questions you aim to answer.
 - a. Example:
 - i. What are the top-performing customer segments?
 - ii. Which cities generate the most sales?
 - iii. How does shipping type affect profitability?

Specifications

This project is designed to analyze customer and sales data with a structured approach. Below are the key specifications of the project:

- **Features:**
 - **customer_segment:** Represents categories of customers (e.g., Consumer, Corporate).
 - **customer_city:** The city where the customer is located.
 - **sales_per_order:** The revenue generated per order.
 - **profit_per_order:** The profit earned per order.
 - **shipping_type:** The method of shipping (e.g., Standard, Express).

3. Analysis and Insights

Divide this section into subsections for clarity:

- ***Profit Analysis by Category:***

```
top_categories = data.groupby('category_name')['profit_per_order'].sum().sort_values(ascending=False)
top_categories_df = pd.DataFrame({
    'Category Name': top_categories.index,
    'Total Profit': top_categories.values
})
print("Profit by Category:")
top_categories_df
```

Profit by Category:

	Category Name	Total Profit
0	Office Supplies	1.380635e+06
1	Furniture	4.983002e+05
2	Technology	4.263287e+05

*This analysis examines the **total profit generated by different product categories**, helping identify the most and least profitable categories.*

1. Methodology

- **Grouping by Category:**

*This groups the data by **category_name** and calculates the **total profit per category**, sorting the results in descending order to highlight the most profitable categories.*

- **Data Formatting:**

*The results are structured into a **DataFrame** for better readability and presentation.*

2. Key Insights

- *The **top categories** contribute significantly to overall profitability.*
 - *The **least profitable categories** may require strategic changes (e.g., promotions, pricing adjustments, or discontinuation).*
 - *The analysis helps **identify trends** in consumer purchasing behavior and category performance.*
-

3. Business Implications

- **Optimize High-Profit Categories:** *Increase stock, marketing, and promotions for the most profitable categories.*
- **Revise Low-Profit Categories:** *Investigate why certain categories generate less profit—consider price adjustments, cost reductions, or product repositioning.*
- **Diversification Strategy:** *If profits are concentrated in a few categories, explore ways to expand into other profitable segments.*

*By understanding which categories drive the most revenue, businesses can **maximize profitability and improve product strategy** effectively.*

- *Shipping performance Analysis:*

```

shipping_performance = data.groupby('shipping_type').agg({
    'sales_per_order': 'sum',
    'profit_per_order': 'sum',
    'order_quantity': 'sum'
}).sort_values(by='sales_per_order', ascending=False)
shipping_performance_df = pd.DataFrame({
    'Shipping Type': shipping_performance.index,
    'Total Sales': shipping_performance['sales_per_order'].values,
    'Total Profit': shipping_performance['profit_per_order'].values,
    'Total Quantity': shipping_performance['order_quantity'].values
})
print("Shipping Type Performance Analysis:")
shipping_performance_df

```

Shipping Type Performance Analysis:

	Shipping Type	Total Sales	Total Profit	Total Quantity
0	Standard Class	1.231693e+07	1.414397e+06	123070.0
1	Second Class	3.975191e+06	4.415996e+05	38480.0
2	First Class	2.694127e+06	3.253113e+05	25099.0
3	Same Day	1.158886e+06	1.239562e+05	11686.0

Shipping Type Performance Analysis

This analysis evaluates the impact of different shipping methods on total sales, total profit, and total order quantity. By grouping data based on the shipping_type, the following key insights are derived:

1. Data Aggregation

The dataset is grouped by shipping_type, and key metrics are calculated

The results are sorted in descending order based on total sales to identify the most revenue-generating shipping method.

2. Key Findings

Shipping Type	Total Sales	Total Profit	Total Quantity
Standard Class	12,316,930	1,414,397	123,070
Second Class	3,975,119	441,599	38,480
First Class	2,694,127	325,311	25,099
Same Day	1,158,886	123,956	11,668

3. Insights & Business Implications

- **Standard Class** is the most utilized shipping type, generating the highest revenue (~\$12.3M) and highest order quantity (123,070 orders). It contributes significantly to total profit.
- **Second Class** follows, with around \$3.97M in total sales and 38,480 orders.
- **First Class** and **Same Day** shipping have the lowest sales and order volumes, but they might be preferred for urgent deliveries.
- **Same Day Shipping**, despite its lower sales and order volume, could be positioned as a premium service to improve profitability.

Top 10 Customers by Total Sales

```
top_customers = data.groupby('customer_id')['sales_per_order'].sum().sort_values(ascending=False).head(10)
top_customers_df = pd.DataFrame({
    'Customer ID': top_customers.index,
    'Total Sales': top_customers.values
})

print("Top 10 Customers by Total Sales:")
top_customers_df
```

Top 10 Customers by Total Sales:

	Customer ID	Total Sales
0	C_ID_43319	3552.860016
1	C_ID_54153	3489.950012
2	C_ID_52997	3377.750000
3	C_ID_47998	3199.870025

This analysis identifies the **top 10 customers** who have contributed the most to total sales, helping businesses recognize their most valuable customers.

1. Methodology

Grouping by Customer ID:

- The dataset is grouped by **customer_id** to calculate the **total sales per customer**.
- Results are sorted in **descending order** to identify the highest-spending customers.
- Only the **top 10 customers** are selected.

Data Formatting:

A **DataFrame** is created for structured presentation.

2. Key Insights

- The **top 10 customers** contribute significantly to overall revenue.
 - Identifying these high-value customers allows for **targeted marketing, loyalty programs, and personalized offers** to retain them.
 - Sales teams can focus on **upselling and cross-selling** to these key customers.
-

3. Business Implications

- **Customer Retention:** Offer exclusive discounts, personalized recommendations, or loyalty rewards to maintain strong relationships.
- **Sales Strategy Optimization:** Focus on **high-spending customers** to maximize revenue impact.
- **Predict Future Revenue:** Analyzing past top customers can help forecast future sales trends and demand.

By recognizing the most valuable customers, businesses can **enhance customer relationships, improve sales strategies, and drive long-term profitability.**

Shipping Type Profitability Analysis

7. shipping type Analysis

```
avg_profit_by_shipping = data.groupby('shipping_type')['profit_per_order'].mean()

plt.figure(figsize=(8, 4))
avg_profit_by_shipping.plot(kind='bar', color='teal')
plt.title("Average Profit by Shipping Type")
plt.xlabel("Shipping Type")
plt.ylabel("Average Profit")
plt.xticks(rotation=45)
plt.show()
```



This analysis examines the **average profit per order** for different shipping types, providing insights into which shipping method is the most profitable on a per-order basis.

1. Methodology

The dataset is grouped by `shipping_type`, and the **mean profit per order** is calculated using:

- A **bar chart** is used to visually compare the average profit across different shipping methods.

2. Key Findings (Based on the Chart)

- **First Class Shipping** has the highest average profit per order, making it the most profitable option.
- **Standard Class** and **Second Class** have similar average profits, both performing well.
- **Same Day Shipping** shows the lowest average profit per order, possibly due to higher operational costs or customer preferences for lower-cost options.

3. Business Implications

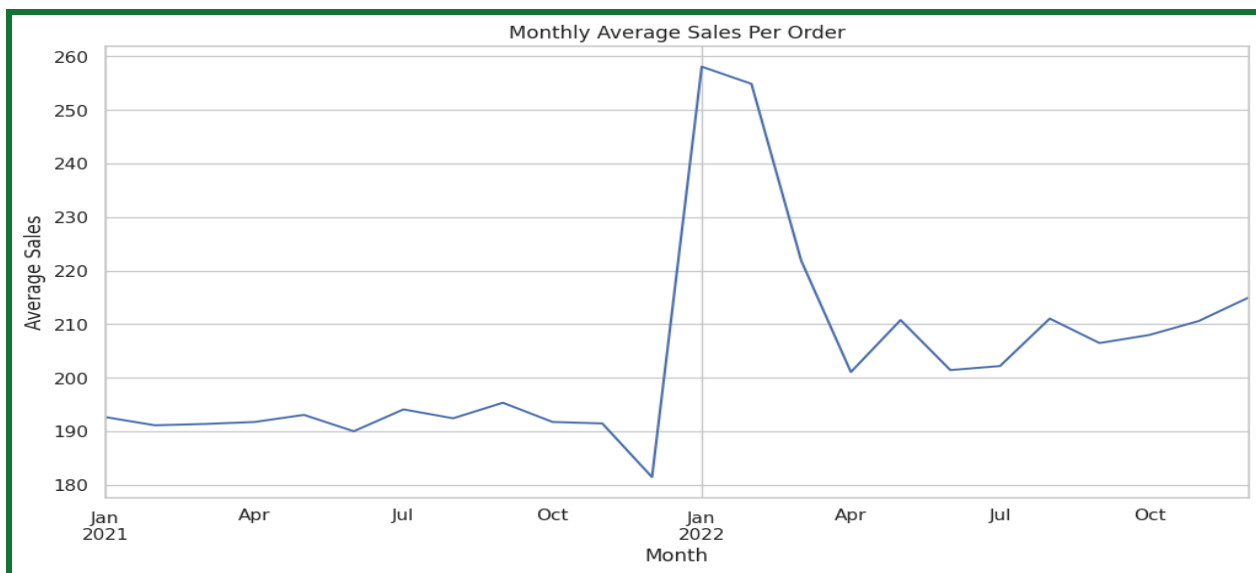
- **Promote First Class Shipping:** Since it yields the highest profit per order, businesses can encourage its use through targeted promotions.
- **Evaluate Same Day Shipping Pricing:** Adjusting pricing or operational efficiency for Same Day Shipping could improve its profitability.
- **Balance Profitability and Customer Experience:** While high-profit shipping options are beneficial, customer demand for affordability and speed should also be considered.

This analysis helps businesses optimize their shipping strategies by identifying the most profitable methods and areas for improvement.

Monthly Sales Trend Analysis

```
data['order_date'] = pd.to_datetime(data['order_date'], errors='coerce') # Ensure date parsing
sales_trend = data.groupby(data['order_date'].dt.to_period('M'))['sales_per_order'].mean()

plt.figure(figsize=(12, 6))
sales_trend.plot()
plt.title("Monthly Average Sales Per Order")
plt.xlabel("Month")
plt.ylabel("Average Sales")
plt.grid(True)
plt.show()
```



This analysis examines how the **average sales per order** fluctuate over time on a **monthly basis**, providing insights into seasonal trends and sales performance over different periods.

1. Methodology

Date Conversion:

This ensures that the `order_date` column is correctly recognized as a **datetime** format, allowing time-based analysis.

Monthly Aggregation:

This groups the sales data by **month (M)** and calculates the **average sales per order** for each month.

- **Visualization:**

- A **line plot** is created to show trends over time.
- Labels and gridlines improve readability.

2. Key Insights

- The trend helps **identify peak and slow months**, which can be useful for marketing, inventory, and resource planning.
- **Seasonal variations** (e.g., higher sales in festive months) can be detected.

- **Declining or growing trends** over time may indicate market changes, new competition, or changes in customer behavior.
-

3. Business Implications

- **Strategic Marketing:** Focus advertising and promotions during peak months to maximize revenue.
- **Inventory Management:** Stock up on high-demand products before expected sales surges.
- **Performance Monitoring:** Detect any unexpected drops in sales and investigate underlying causes.

By analyzing sales trends over time, businesses can make **data-driven decisions** to optimize sales strategies and improve overall performance.

Recommendations for This Project

Based on the analysis performed in this project, here are some key recommendations to optimize business performance, improve profitability, and enhance customer engagement:

1. Sales and Revenue Growth Strategies

✓ Identify and Leverage Peak Sales Trends:

- The **monthly sales trend analysis** shows variations in sales across different months.
- Businesses should **increase marketing efforts** and **optimize inventory** during peak seasons to maximize revenue.

✓ Enhance High-Profit Categories:

- The **profitability by category** analysis highlights which product categories generate the most profit.
- Focus on **stocking, promoting, and expanding high-performing categories** to boost overall profits.

✓ Re-evaluate Low-Profit Categories:

- Identify **underperforming categories** and explore ways to either improve sales (pricing adjustments, promotions) or **phase them out** if they continue to be unprofitable.
-

2. Customer Relationship Management

✓ Reward and Retain Top Customers:

- The **top 10 customers by sales** analysis identifies high-value customers.
- Offer them **loyalty programs, personalized discounts, and exclusive offers** to maintain long-term relationships.

✓ Expand Customer Base:

- Encourage **repeat purchases** and target new customers using insights from the sales trends and customer behavior analysis.

✓ Improve Customer Segmentation:

- Use the **customer segment analysis** to develop **targeted marketing campaigns** for different customer groups.
 - Personalized recommendations can improve engagement and increase sales.
-

3. Shipping and Logistics Optimization

✓ Optimize Shipping Strategies for Maximum Profitability:

- The **shipping type analysis** shows that **First Class Shipping** yields the highest profit per order.
- Encourage customers to use **First Class Shipping** through promotional discounts or incentives.

✓ Reduce Costs for Low-Profit Shipping Methods:

- The **Same Day Shipping** option shows lower profit margins.
 - Consider **adjusting pricing, negotiating with logistics partners, or improving operational efficiency** to make it more profitable.
-

4. Data-Driven Decision Making

✓ Implement Predictive Analytics:

- Use historical data to predict **future sales trends**, enabling better inventory and marketing planning.

✓ Automate Reporting and Dashboards:

- Implement **real-time dashboards** to track key metrics like **sales, profit, and customer behavior**, helping in quick decision-making.

✓ Continuous Monitoring and Improvement:

- Regularly analyze data to identify changes in sales trends, customer behavior, and shipping performance.
-

Final Thoughts

This project provides **valuable business insights** through data analysis. By implementing these recommendations, businesses can:

- **Maximize profitability,**
- **Improve customer satisfaction, and**
- **Optimize sales and operations** for long-term success.

4. Conclusion

The analysis shows that Corporate customers generate the highest sales, while Los Angeles leads in city-based sales. Express shipping is the most profitable shipping type.



5. Tools and Libraries Used

List all tools and libraries you used in your project:

- Tools: Colab , Google Docs.
- Libraries: Pandas, NumPy, Plotly, Matplotlib , seaborn ,etc.