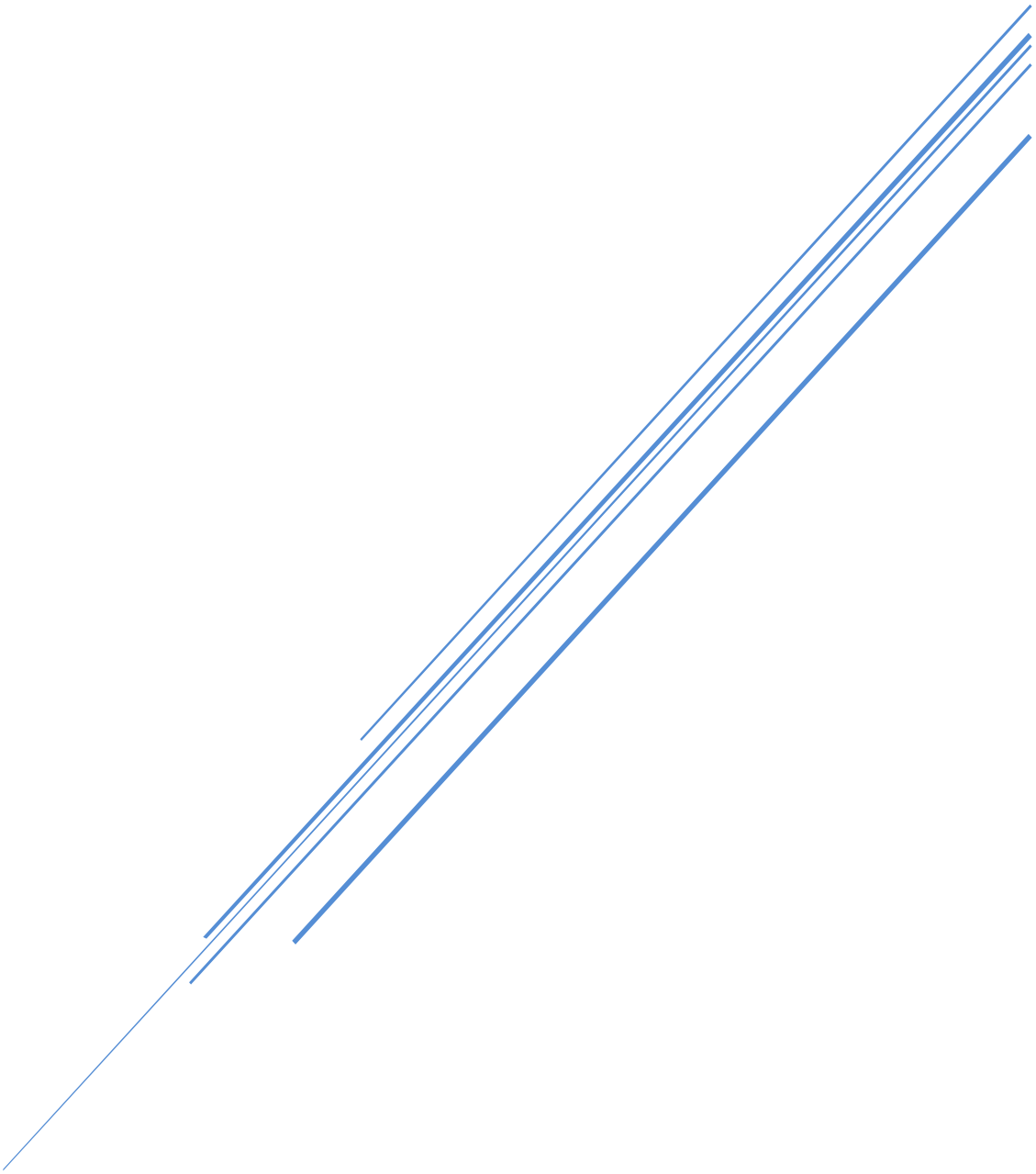
SUPERSTORE SUCCESS: BOOSTING GROWTH WITH STRATEGIC DATA ANALYSIS



Sushant Thapa Magar
MSc. Business Analytics

Contents

1. INTRODUCTION

1.1 Problem Background:

In the ever-changing retail sector, the Superstore Giant is tackling the challenge of staying ahead in a market characterized by constantly shifting customer preferences, a wide range of product performances, and diverse regional market conditions. The store possesses an extensive collection of data encompassing sales, profit margins, customer transactions, and regional sales outcomes. The central challenge lies in effectively analysing this wealth of information to derive meaningful insights. The objective is to pinpoint the most lucrative products, identify key customer segments, and grasp the nuances of regional market trends. Through the application of descriptive analytics, the store aims to reveal patterns in sales and profit, understand customer buying behaviours, and conduct an in-depth analysis of regional market behaviour. This critical analysis is essential for refining the product mix, crafting targeted marketing strategies to engage customers, and customizing approaches to meet regional market needs. Gaining these insights is vital for improving sales, strengthening customer relations, and shaping regional tactics, ultimately enhancing the store's profitability and competitive stance in the marketplace.

1.2 Problem Statement:

The Superstore Giant needs to analyse its extensive retail data to optimize product offerings, enhance customer engagement, and adapt to regional market trends for improved profitability and competitiveness.

1.3 Aim and objectives:

Boost Superstore Giant's market position through strategic insights. Utilize descriptive analytics for a deep understanding of historical sales and profit trends. Optimize target regions, customer segments, and product portfolios, fostering growth.

A. Product Portfolio Optimization and Sales Performance:

- **Product Portfolio Evaluation**: Assess individual product performance and profitability; make strategic decisions on optimizing or phasing out specific products.

- **Historical Sales and Profit Analysis**: Perform time-series analysis to identify trends, patterns, and areas needing improvement in sales and profits.

B. Regional Market Analysis and Strategy Development:

- **Regional Sales and Profit Performance Analysis:** Evaluate sales and profit performance across different regions to identify high-performing and underperforming markets. Analyze market penetration in different regions and states; identify underperforming concerning areas and those with potential for growth and expansion.

C. Customer-Centric Analysis and Engagement Strategy:

- **In-Depth Customer Segmentation**: Segment customers based on transactional data to understand distinct behavior patterns, order frequencies, and preferences.
- **Tailored Marketing and Engagement Approaches:** Develop marketing strategies focused on different customer segments to enhance engagement and satisfaction.

2. TOOLS USED

Tools and Techniques used for Data Analysis:

In our dataset, ***pandas***, a versatile Python library, played a pivotal role in simplifying tasks like loading, preprocessing, and analysis based on ***NumPy***. Leveraging its capabilities in handling missing values, restructuring data, and conducting group-by operations,

Matplotlib and Seaborn enhance the analytical toolbox, with Seaborn excelling in statistical data visualization and intuitive interfaces, while Matplotlib provides versatile plot styles. Both are crucial in crafting various plots for comprehensive data analysis of our dataset.

Additionally, usage of Microsoft excel has been done for creating tables to fuel clarity and basic statistical theories have been used for data analysis.

Furthermore, find the figure below describes the tools and techniques have been used in this analysis:

Tools	Area Used	Techniques and Visualizations Used
Pandas	Data Loading/Exploration	read_csv, df.head(), "df.info(), df.shape, df.columns
	Data Preprocessing	Handling missing values- isnull(), Handling duplicates- df.duplicated(), Dropping Values- dropna(), df.drop, Column Renaming
	Date time Operations	pd.to_datetime(df['Order Date'])
	KPIs Calculation	Various operations
	Grouping & Sorting	df.groupby, sort_values
	Unit Metric Calculations	'Unit Discounted Sale Price,' 'Unit Original Sale Price,' 'Unit Profit,' 'Unit Cost Price,' and 'Profit Margin'
	Datetime Indexing	df.set_index
	Time Series Analysis	Resampling for monthly trends- df['Sales'].resample('M').sum()
Matplotlib	Charts	Line chart-plt.plot, Area Chart-plt.fill_between, Bar charts-plt.bar
	Annotations	bar_plot.text
	Customer Segmentation	Grouped Bar Charts
	Discount Analysis	Correlation Matrix, Heatmap
	Product Analysis	Clustered Bar Charts, Top Products Identification
	Shipping Analysis	Bar Charts, Box Plots
Seaborn	Data Visualization	Bar Charts, Line Charts, Area Charts, Clustered Bar Charts- sns.barplot
	Correlation Analysis	Correlation Matrix, Heatmap, sns.heatmap

Fig. 1

3. WORK CARRIED OUT

3.1 Data collection

The decision to select the Kaggle dataset from the Tableau website stems from its suitability in addressing the business challenges faced by the Superstore . The dataset captures crucial details such as order information, customer demographics, and key financial metrics like sales and profit. While the dataset presents an enticing opportunity for predictive modeling, particularly using regression analysis, it's important to note that, at this juncture, I haven't acquired the requisite skills for such advanced analytics. Therefore, the focus will be on leveraging descriptive analytics techniques to extract meaningful insights and inform strategic decision-making. The dataset's availability on Kaggle, a reputable collaborative platform, ensures accessibility for learning and educational purposes. Additionally, the ethical consideration of crediting the original creators from the Tableau website adds to the dataset's reliability. In essence, the choice of this dataset aligns with the current learning objectives, emphasizing descriptive analytics while acknowledging the potential for future advanced analyses.

3.2 Data pre-processing

This process involves several key stages to ensure that the data is accurate, relevant, and structured for effective analysis.

a) Library and Data Import:
Crucial libraries like NumPy, Pandas, Matplotlib, and Seaborn are imported for data manipulation and visualization.

b) Initial Data Exploration:
Commands such as df.head(), df.info(), df.shape, and df.columns offer an initial grasp of the dataset's makeup, content, and data types.

c) Data Cleaning:
Managing Missing Data: Using df.isnull().sum(), missing values are identified, and df.dropna(inplace=True) is employed to eliminate rows with missing data, ensuring data completeness.

d) Duplicate Removal: After identifying duplicates (df.duplicated().value_counts()), the 'Row ID' column, which hinders duplicate detection, is dropped, ensuring data uniqueness.

e) Feature Selection:
Domain Knowledge-based Selection: Columns like 'Customer ID,' 'Country,' and 'Postal Code' are removed based on their relevance to the analysis, guided by an understanding of the data's context and specific analytical goals.

f) Data Enrichment:
Introduction of New Metrics: Metrics such as 'Unit Discounted Sale Price,' 'Unit Original Sale Price,' 'Unit Profit,' 'Unit Cost Price,' and 'Profit Margin' are added, providing deeper insights into profitability and sales dynamics.

g) Column Renaming:
Columns are renamed (e.g., 'Segment' to 'Customer Segment') for improved clarity and interpretability.

The included code snippets illustrate the practical implementation of each step, reflecting a comprehensive understanding of both the data and the business context.

3.3 Data Analysis and Inferences

A. Key Performance Indicators (KPIs):

Before we start analyzing based on the objectives we have set, we need to understand the *Key Performing Indicators (KPI)* relevant to the data set. The inferences provide a starting point for deeper analysis.

KPIs	Values
Total Sales	2,296,919
Total Quantity Sold	37,871
Total Profit	286,409
Average Profit Margin	12.47%
Total Customers	793.00
Total Orders	5,009
Average Discount Rate	16.00%
Average Quantity Per Order	7.16
Average Sales Per Order	459
Average Profit Per Order	57.18

Fig. 2

Inferences:

- ***12.47%*** profit margin reflects efficient cost management, emphasizing the potential in high-margin products for growth.
- Customer retention-***5,009 orders from 793*** customers highlight repeat business. Prioritize customer loyalty through rewards and personalized services.
- Order Size-Encourage larger orders ***(7.16 items/order, 458.56 sales/order)*** through promotions and bundling deals.

B. Product Portfolio Optimization and Sales Performance:

1.1 Product Portfolio Optimization:

Bar charts excel in optimizing product portfolios by providing visual clarity for comparing sales and profit across product categories and sub-categories. They aid decision-making and trend analysis, simplifying complex data.

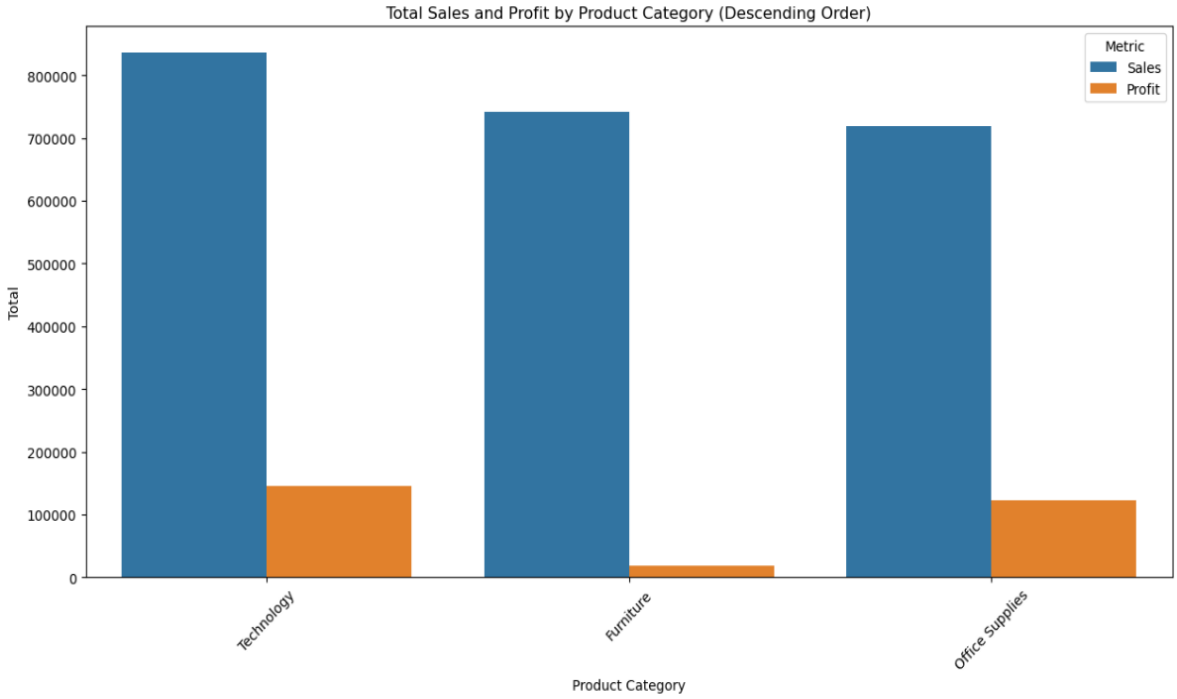

Total Sales and Profit by Product Category (Descending Order)
Metric
Sales
Profit
Total
800000
700000
600000
500000
400000
300000
200000
100000
0
Technology
Furniture
Office Supplies
Product Category

Fig. 3

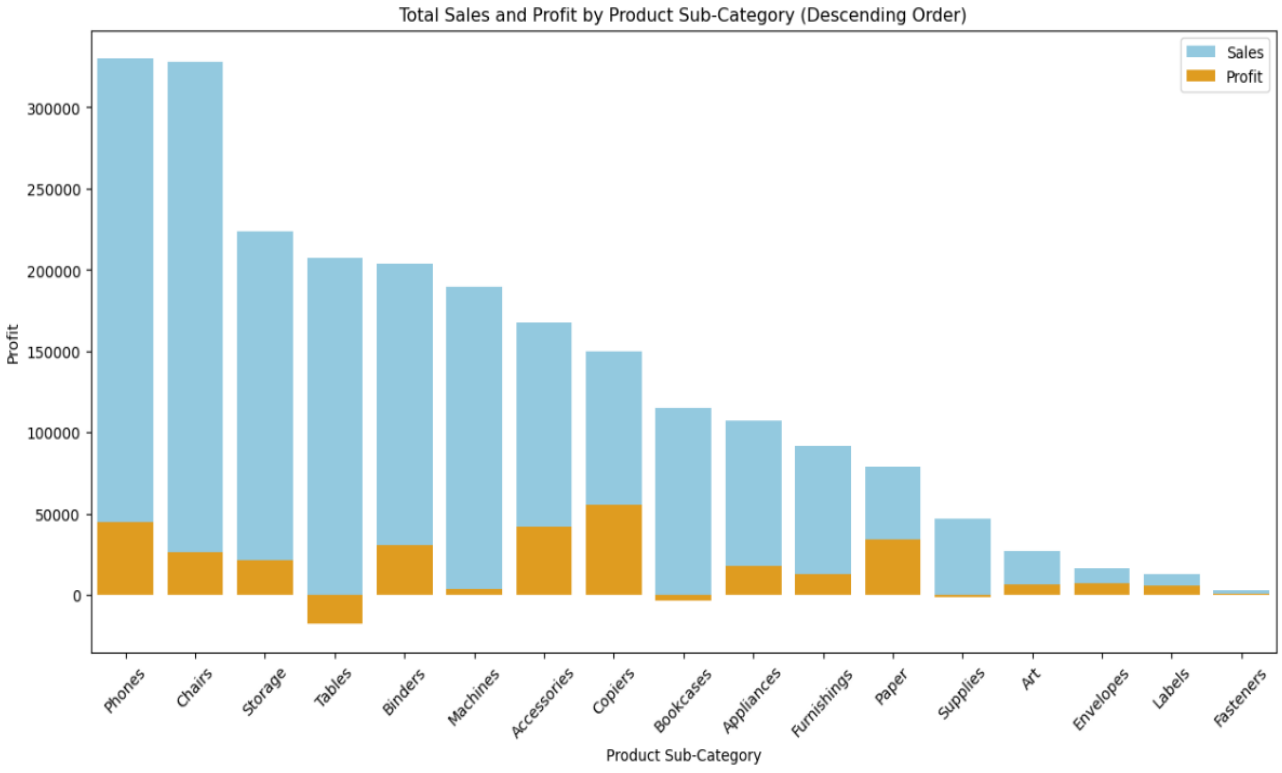

Total Sales and Profit by Product Sub-Category (Descending Order)
Sales
Profit
Profit
300000
250000
200000
150000
100000
50000
0
Phones
Chairs
Storage
Tables
Binders
Machines
Accessories
Copiers
Bookcases
Appliances
Furnishings
Paper
Supplies
Art
Envelopes
Labels
Fasteners
Product Sub-Category

Fig. 4

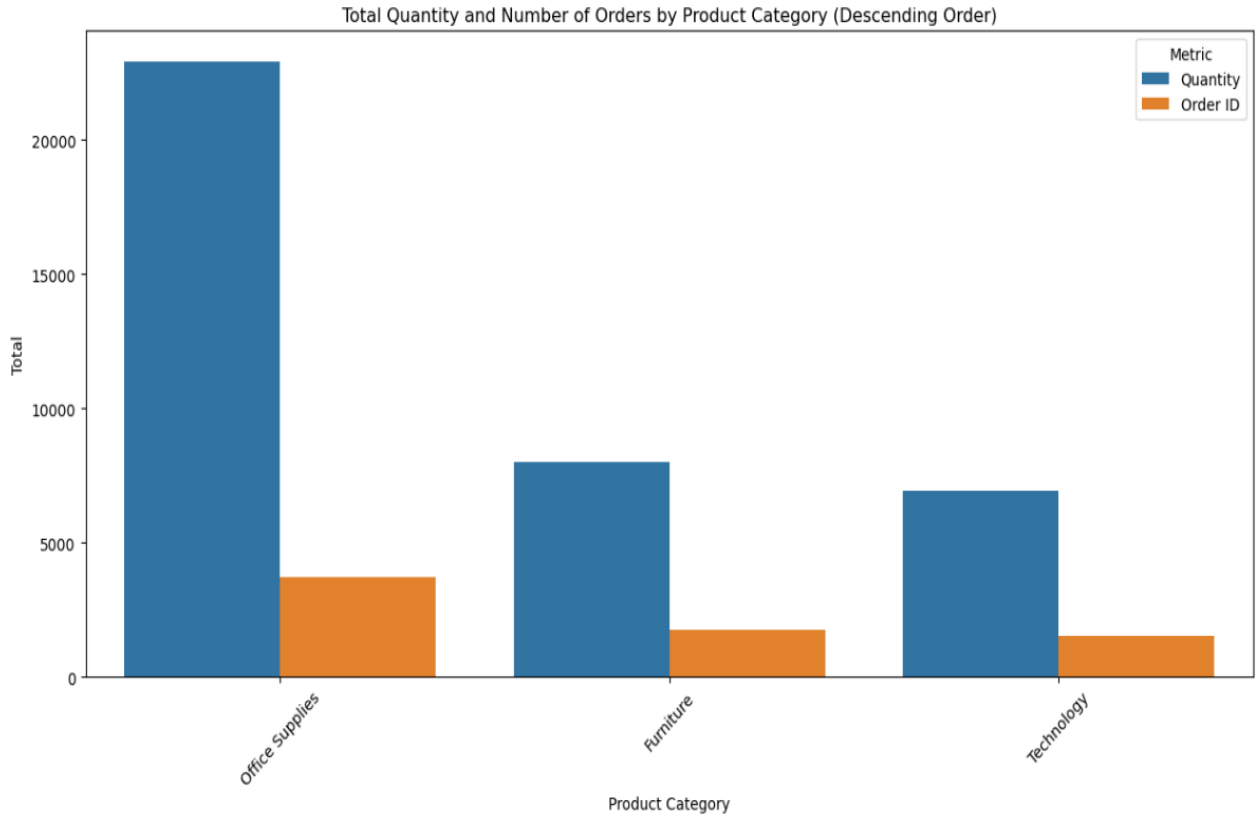

Total Quantity and Number of Orders by Product Category (Descending Order)
Metric
Quantity
Order ID
Total
20000
15000
10000
5000
0
Office Supplies
Furniture
Technology
Product Category

Fig. 5

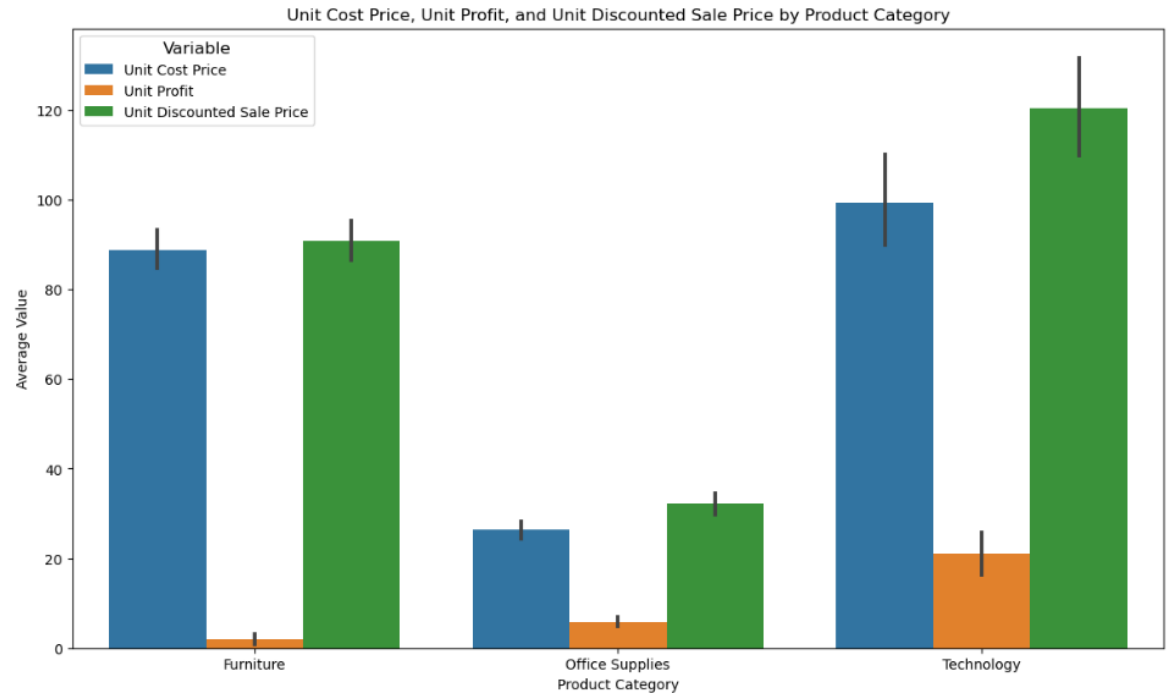

Unit Cost Price, Unit Profit, and Unit Discounted Sale Price by Product Category
Variable
Unit Cost Price
Unit Profit
Unit Discounted Sale Price
Average Value
120
100
80
60
40
20
0
Furniture
Office Supplies
Technology
Product Category

Fig. 6

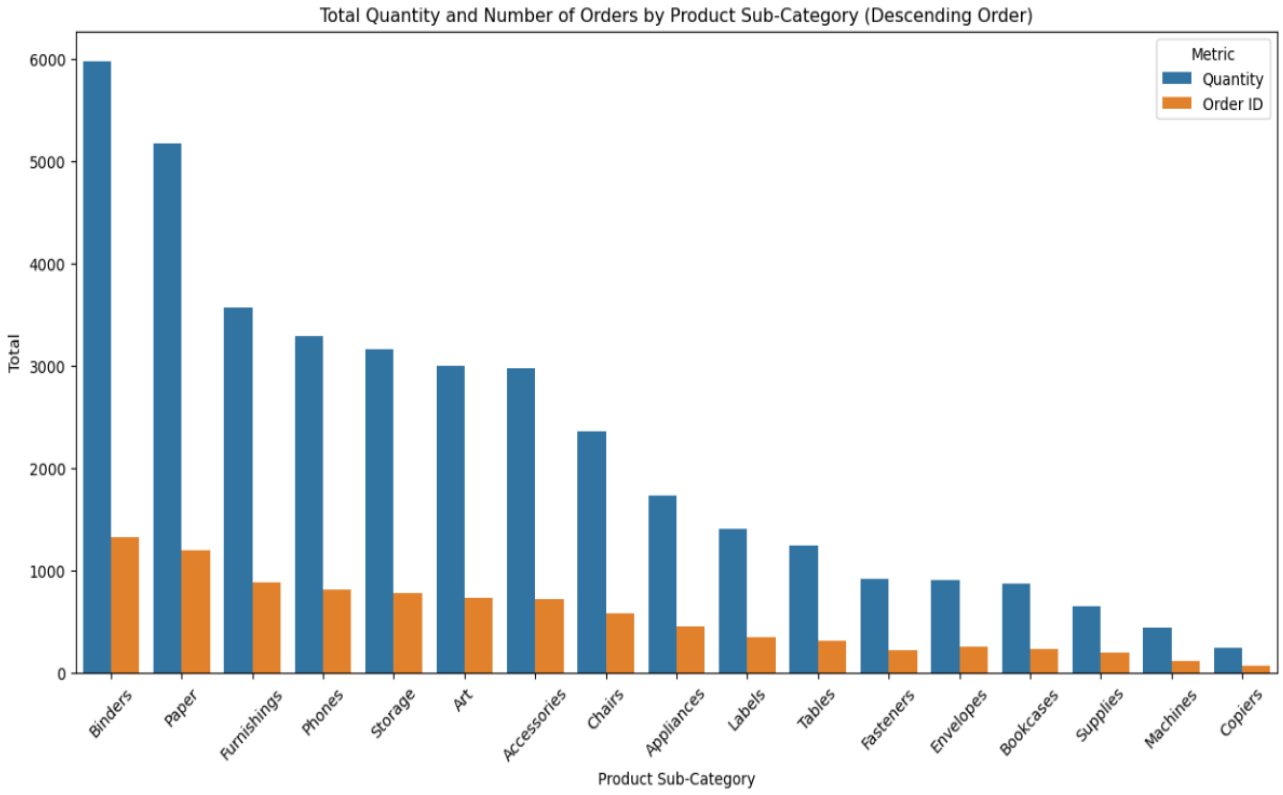

Total Quantity and Number of Orders by Product Sub-Category (Descending Order)
Metric
Quantity
Order ID
Total
Product Sub-Category
Binders
Paper
Furnishings
Phones
Storage
Art
Accessories
Chairs
Appliances
Labels
Tables
Fasteners
Envelopes
Bookcases
Supplies
Machines
Copiers

Fig. 7

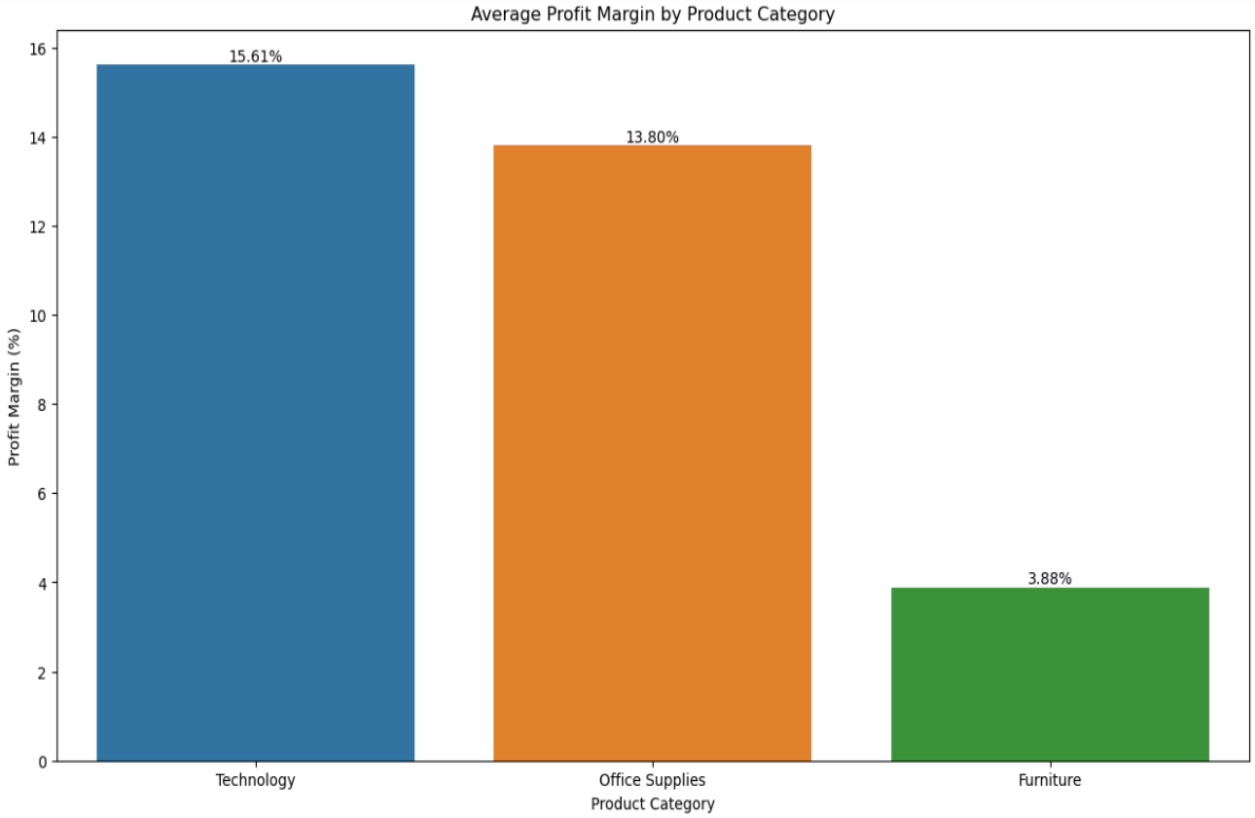

Average Profit Margin by Product Category
Profit Margin (%)
15.61%
13.80%
3.88%
Technology
Office Supplies
Furniture
Product Category

Fig. 8

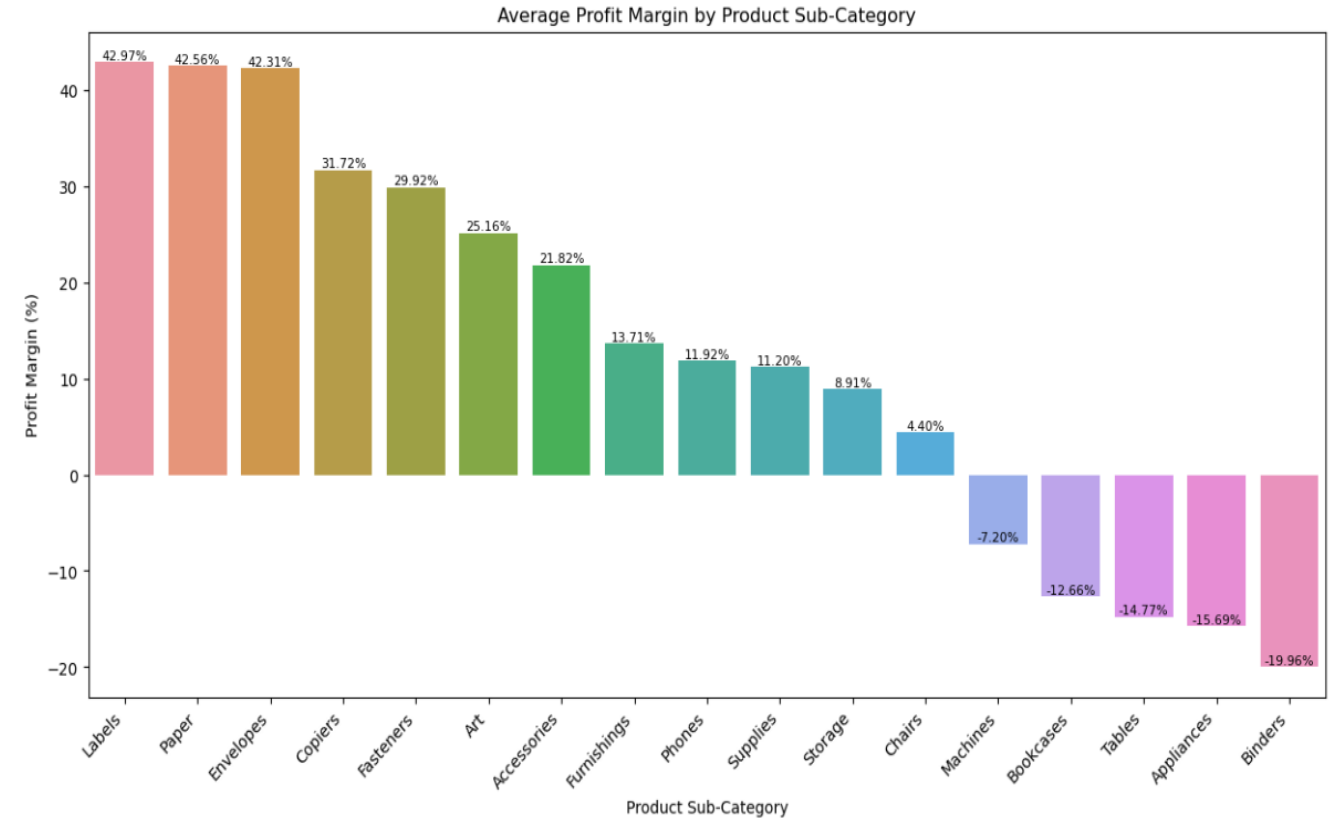
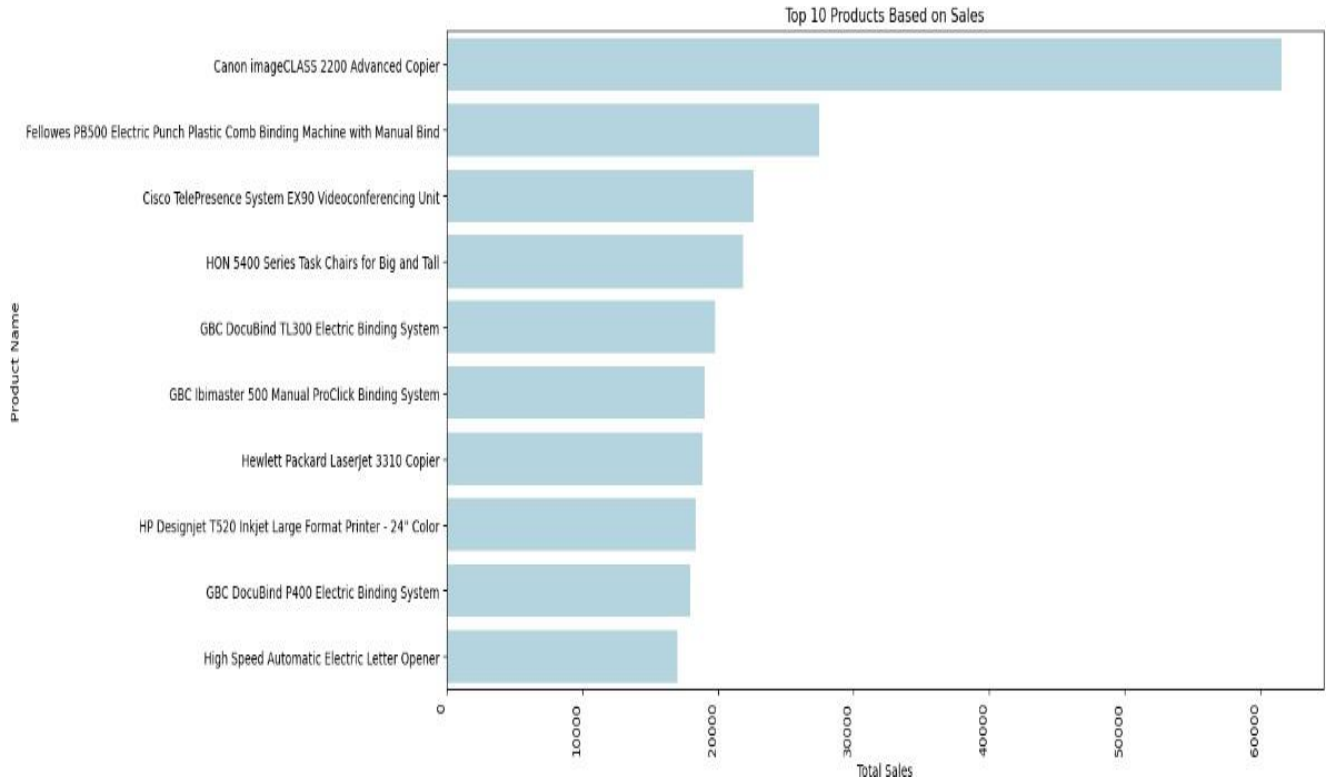

Average Profit Margin by Product Sub-Category
Profit Margin (%)
Product Sub-Category
42.97%
42.56%
42.31%
31.72%
29.92%
25.16%
21.82%
13.71%
11.92%
11.20%
8.91%
4.40%
-7.20%
-12.66%
-14.77%
-15.69%
-19.96%
Labels
Paper
Envelopes
Copiers
Fasteners
Art
Accessories
Furnishings
Phones
Supplies
Storage
Chairs
Machines
Bookcases
Tables
Appliances
Binders

Fig. 9


Top 10 Products Based on Sales
Product Name
Total Sales
Canon imageCLASS 2200 Advanced Copier
Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind
Cisco TelePresence System EX90 Videoconferencing Unit
HON 5400 Series Task Chairs for Big and Tall
GBC DocuBind TL300 Electric Binding System
GBC Ibimaster 500 Manual ProClick Binding System
Hewlett Packard LaserJet 3310 Copier
HP Designjet T520 Inkjet Large Format Printer - 24" Color
GBC DocuBind P400 Electric Binding System
High Speed Automatic Electric Letter Opener
0
10000
20000
30000
40000
50000
60000

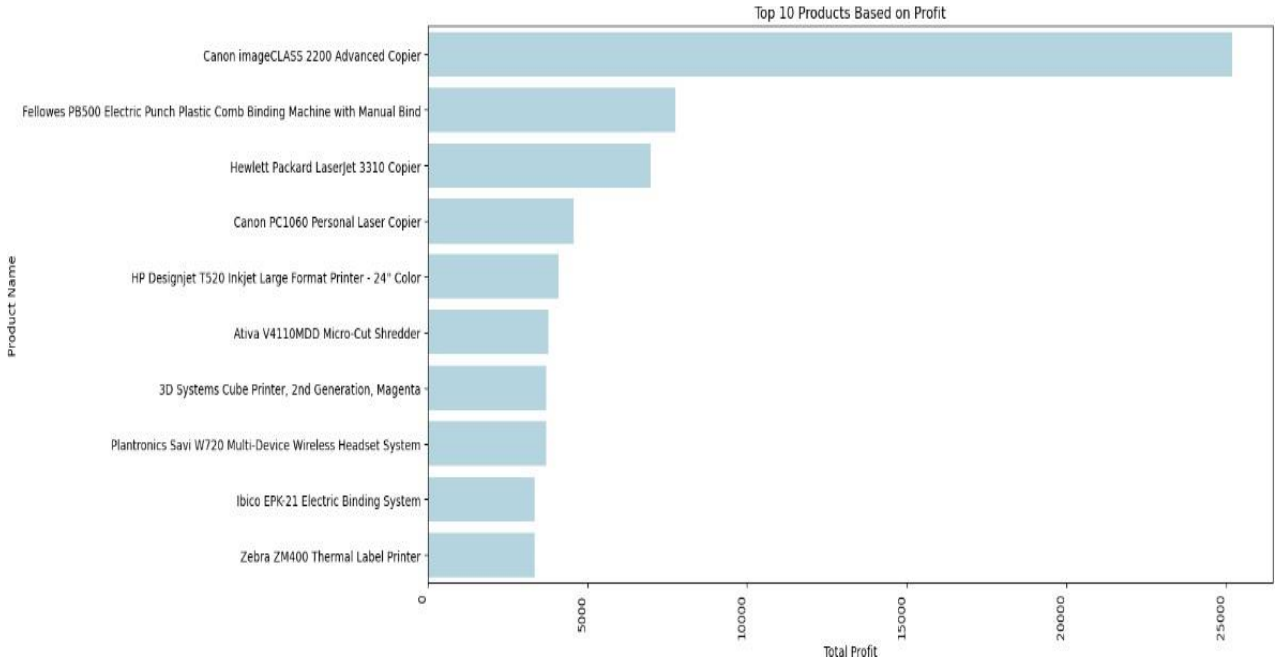
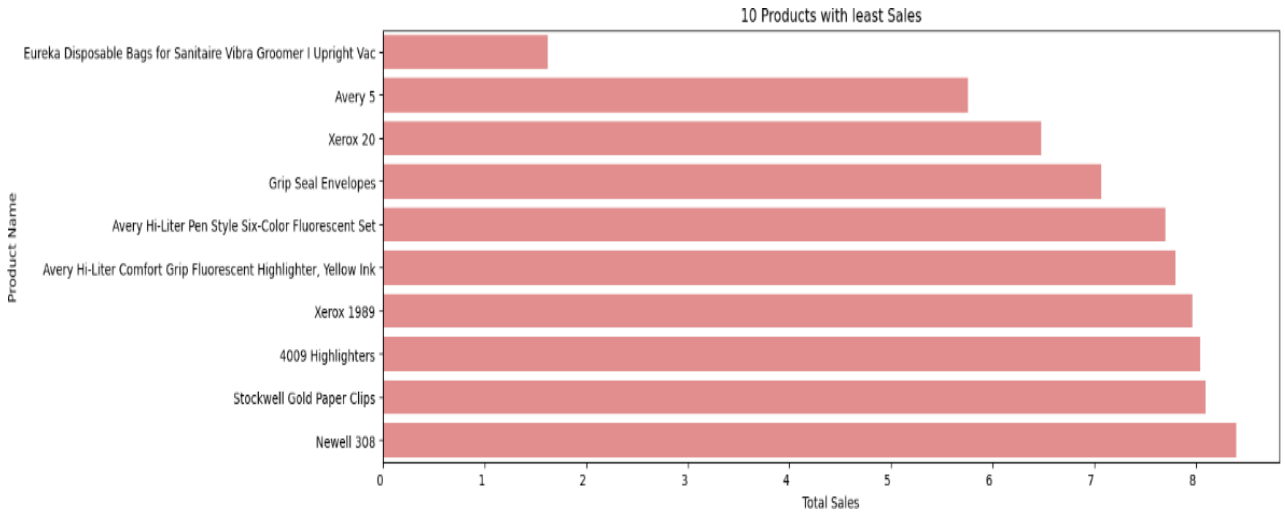

Top 10 Products Based on Profit
Canon imageCLASS 2200 Advanced Copier
Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind
Hewlett Packard LaserJet 3310 Copier
Canon PC1060 Personal Laser Copier
HP Designjet T520 Inkjet Large Format Printer - 24" Color
Ativa V4110MDD Micro-Cut Shredder
3D Systems Cube Printer, 2nd Generation, Magenta
Plantronics Savi W720 Multi-Device Wireless Headset System
Ibico EPK-21 Electric Binding System
Zebra ZM400 Thermal Label Printer
Product Name
0
5000
10000
15000
20000
25000
Total Profit

Fig. 10


10 Products with least Sales
Eureka Disposable Bags for Sanitaire Vibra Groomer I Upright Vac
Avery 5
Xerox 20
Grip Seal Envelopes
Avery Hi-Liter Pen Style Six-Color Fluorescent Set
Avery Hi-Liter Comfort Grip Fluorescent Highlighter, Yellow Ink
Xerox 1989
4009 Highlighters
Stockwell Gold Paper Clips
Newell 308
Product Name
0
1
2
3
4
5
6
7
8
Total Sales

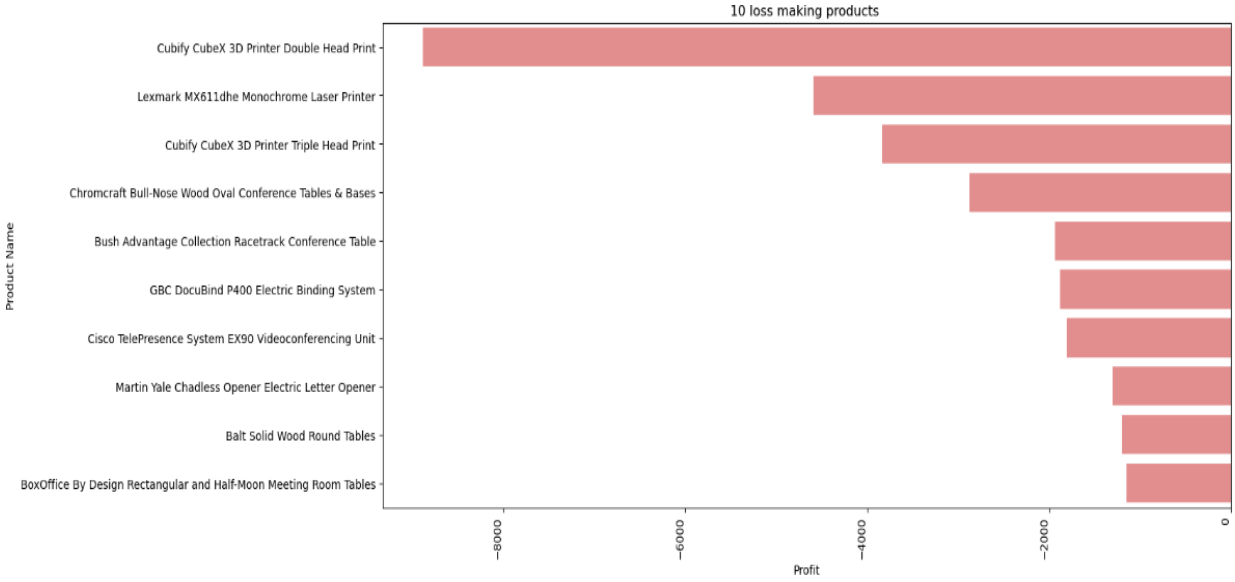


Fig. 11

Inferences:

- **Strategic Emphasis on Technology Category:** The data strongly suggests that focusing on the "Technology" category can be a key driver for both high sales and profitability.

- Sub-categories like "Copiers" and "Technology" demonstrate both high sales and profit margins,but others like "Machines," "Bookcases," and "Tables" pose profitability challenges.

- Top loss-making items highlights areas for improvement such as targeted marketing or operational adjustments.

1.2 Sales and Profit Analysis:

These visualizations use a *line plot and area chart* to show trends over time, a valuable technique for tracking changes in sales and profits and for identifying patterns or irregularities. The use of histogram reveals distribution patterns, identifies outliers, and assesses data variability effectively.

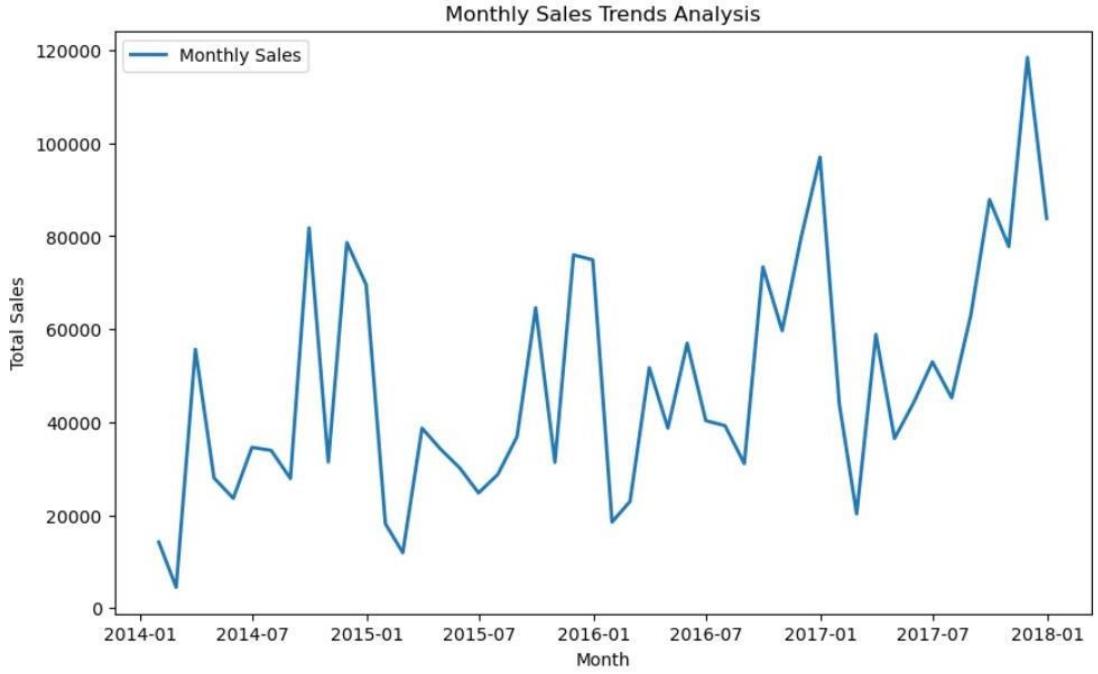


Fig. 12

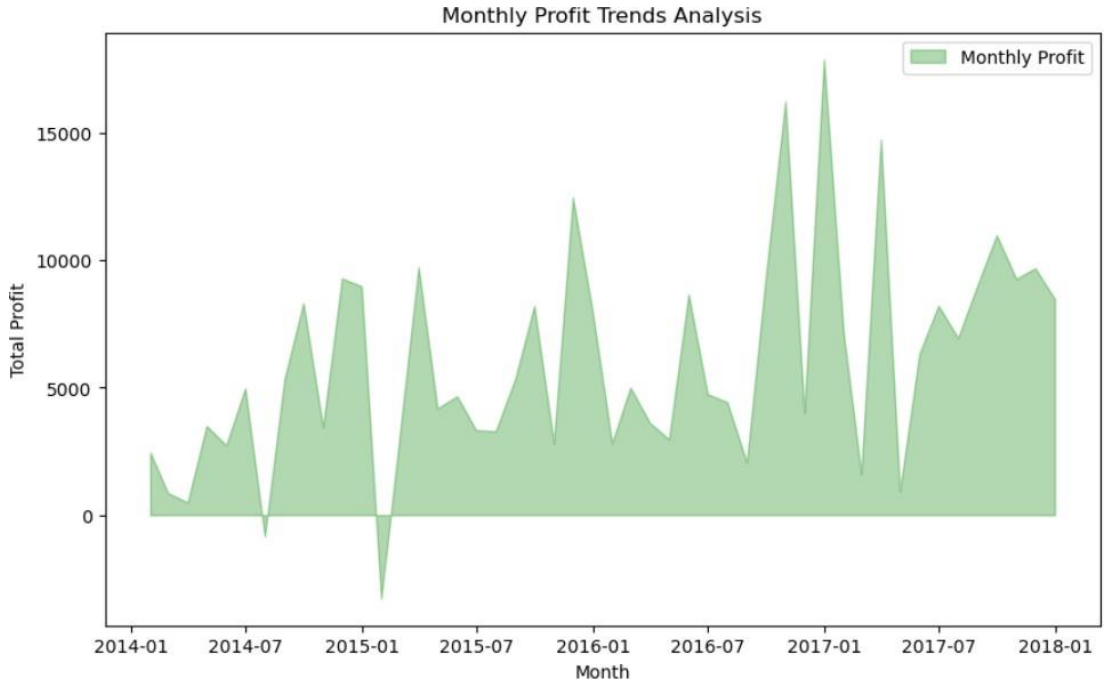


Fig. 13

Particulars	Value	Date
Highest Sales	118,447.82	Nov-17
Lowest Sales	4,519.89	Feb-14
Maximum Profit	17,885.31	Dec-16
Minimum Profit	-3,281.01	Jan-15

Fig. 14

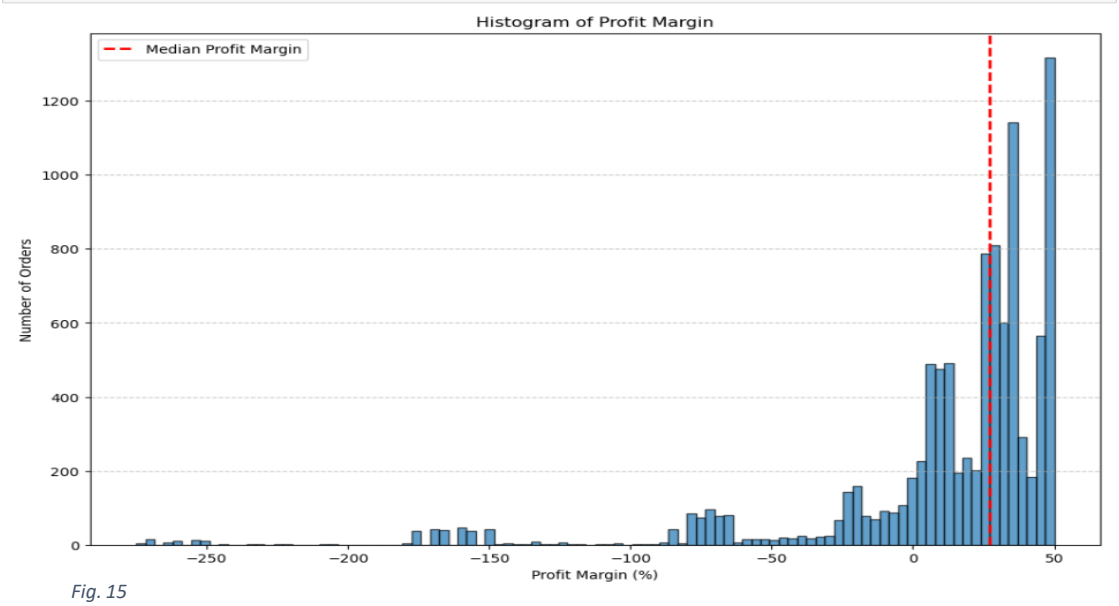


Fig. 15

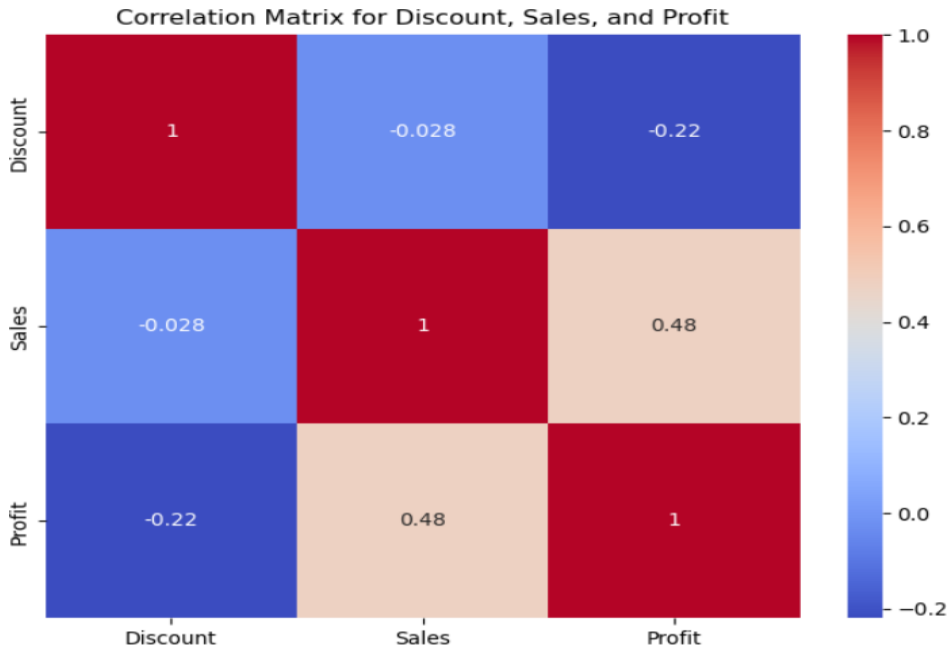

Correlation Matrix for Discount, Sales, and Profit
Discount
Sales
Profit
1
-0.028
-0.22
-0.028
1
0.48
-0.22
0.48
1
Discount
Sales
Profit
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Fig. 16

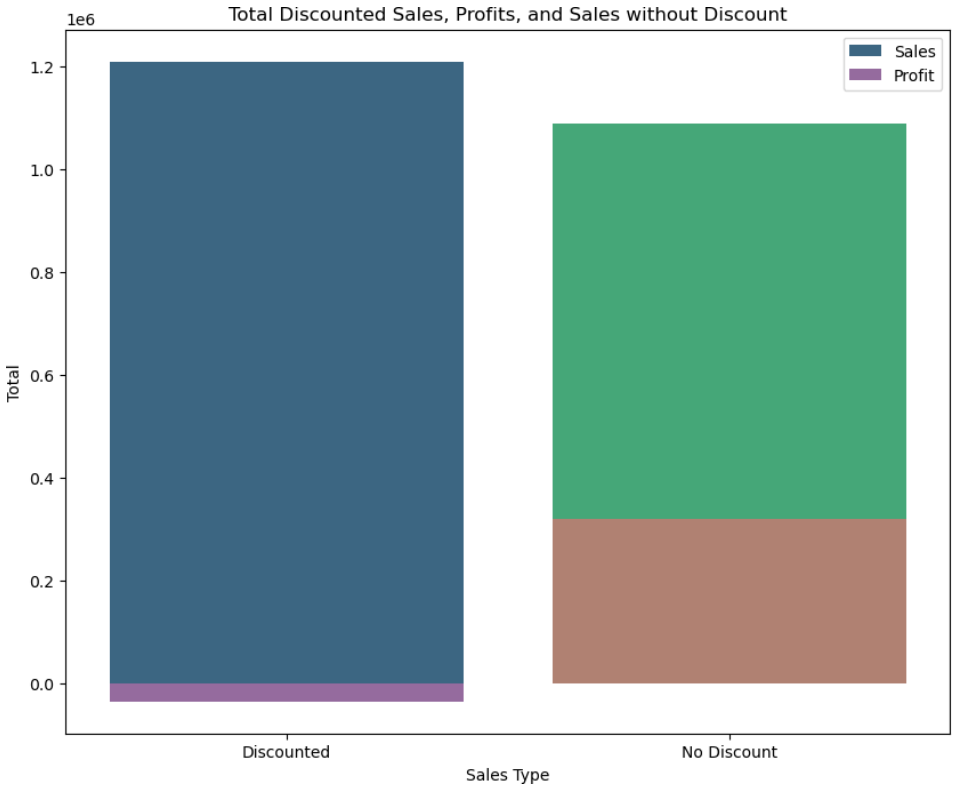

Total Discounted Sales, Profits, and Sales without Discount
1e6
Total
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sales
Profit
Discounted
No Discount
Sales Type

Fig. 17

Inferences:

- **Sales-Profit Discrepancy:** High sales in November 2017 didn't correspond to the maximum profit month, indicating factors like costs or discounts impacting profitability.

- **Seasonal Trends and Growth:** The data reveals recurring monthly sales and profit patterns, showing seasonality. Additionally, there's a noticeable long-term trend of increasing sales and profits.

- The negative correlation between discount and profit, coupled with the stacked bar chart, emphasizes the need for a more strategic approach to discounting. While discounts may boost sales, the data suggests that they often erode profit margins.

- There is a varied profit margin distribution, emphasizing a *concentration around **50%.*** Approximately half of the orders exhibit non-negative profit margins, but a significant portion incurs losses.

C. Regional Market Analysis and Strategy Development:

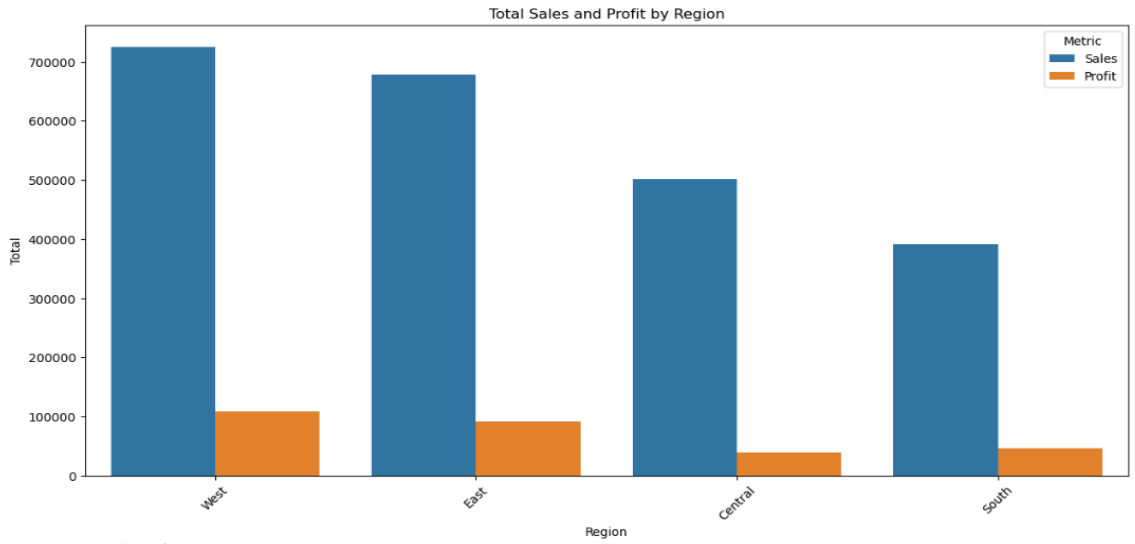


Fig. 18

Region	Sales	Profit	Profitability
West	725,457.82	108,418.45	14.94%
East	678,499.87	91,534.84	13.49%
Central	501,239.89	39,706.36	7.92%
South	391,721.91	46,749.43	11.93%

Fig. 19

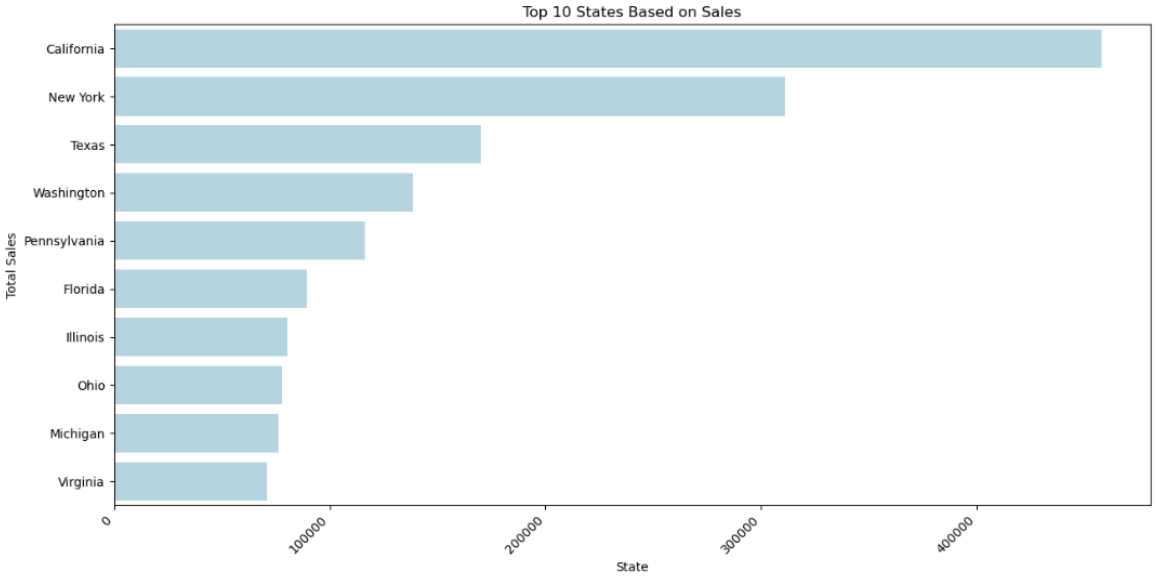


Fig. 20

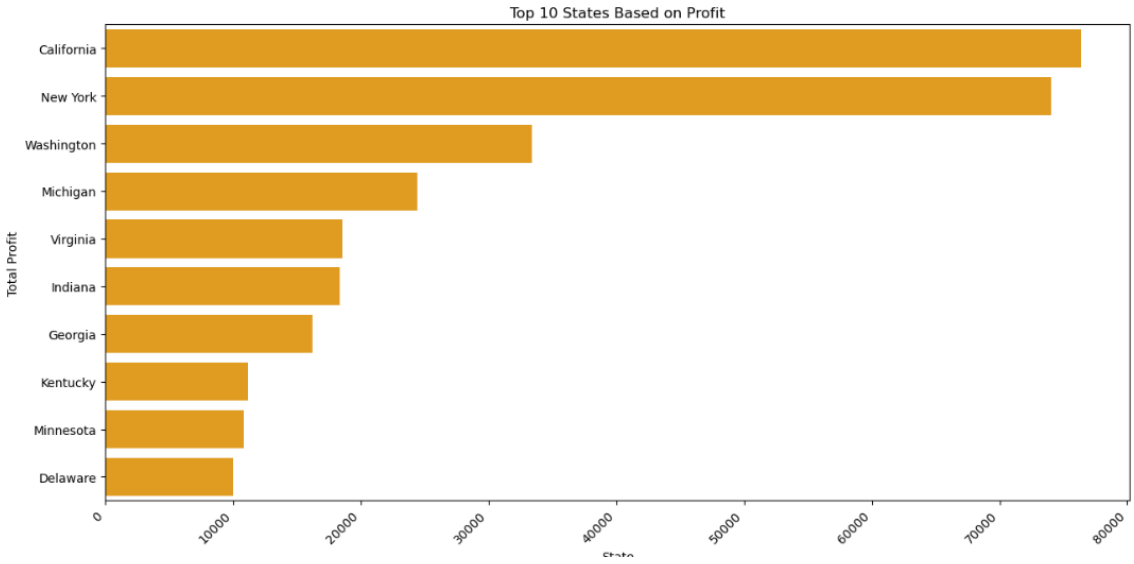


Fig. 21

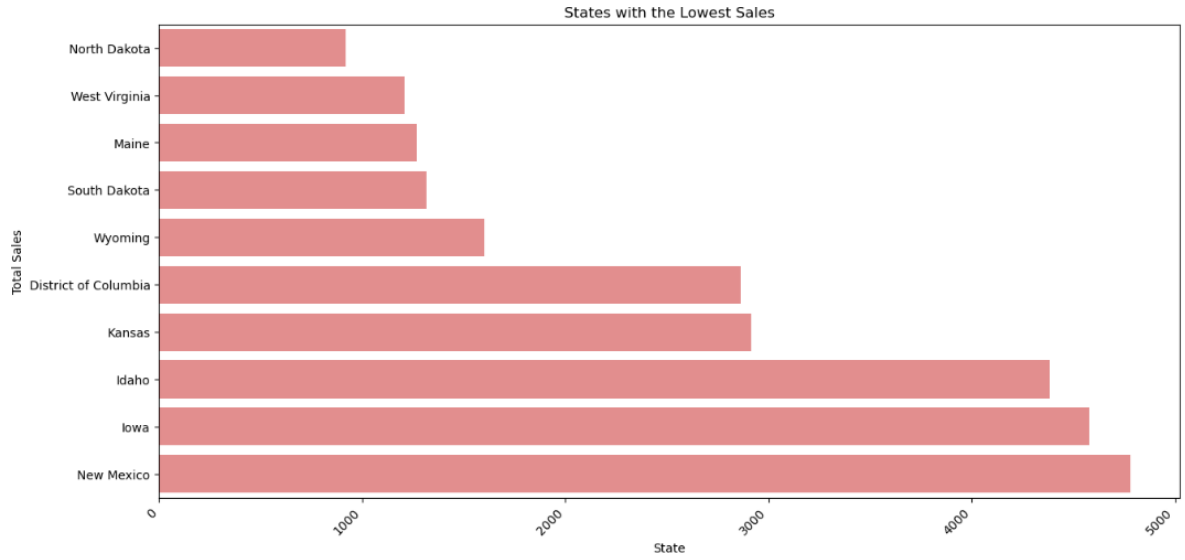


Fig. 22

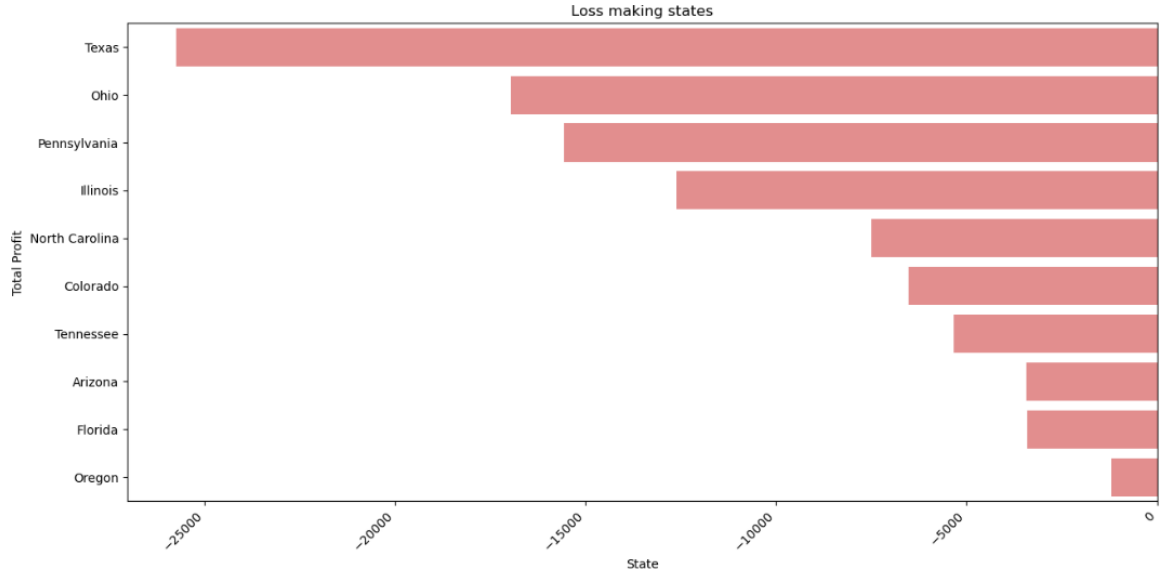


Fig. 23

States	Profit	Sales	Profitability
New York	76,381.39	457,687.63	16.69%
California	74,038.55	310,876.27	23.82%
All other states	135,989.14	1,528,355.59	8.90%
Total	286,409.08	2,296,919.49	12.47%

Profit Share of Top 2 States: 52.52%

Sales Share of Top 2 States: 33.46%

Fig. 24

Inferences:

- **Concentrated High Performance**: California and New York generate ***over 50 %of the total profit*** and over ***30% of total sales*** with higher profitability ratios.

- **Opportunities for Growth:** Other states have significant sales potential, with double sales yet significantly less profitability than the top two states combined.

- **Underperformers:** Low sales and loss-making states reveal underperforming regions, guiding improvement efforts or resource reallocation toward more profitable areas.

D. Customer-Centric Analysis and Engagement Strategy:

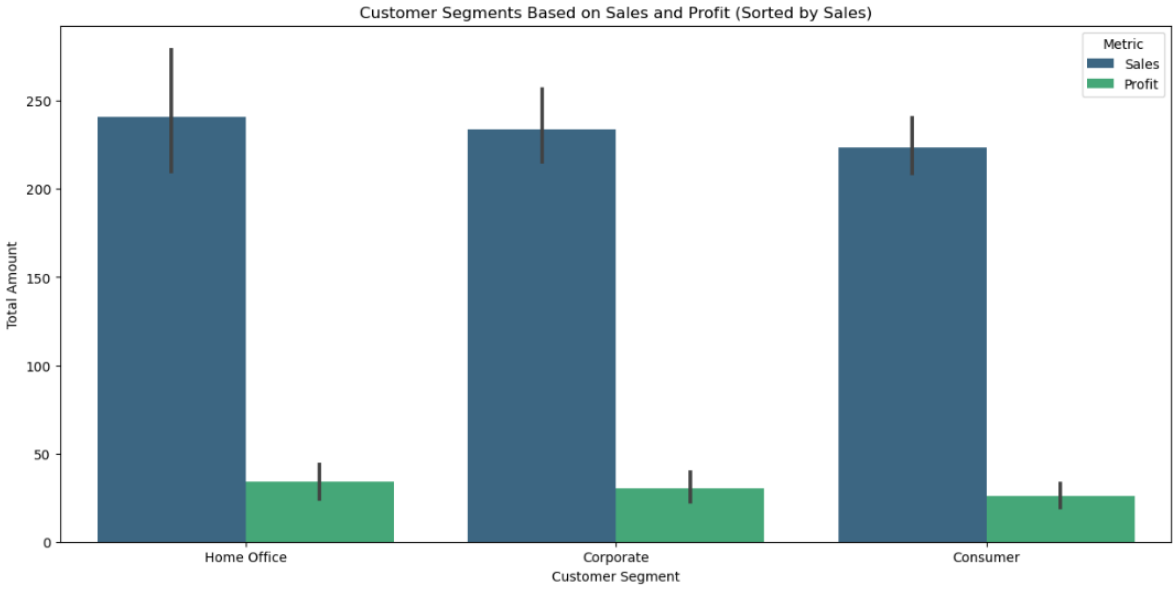


Fig. 25

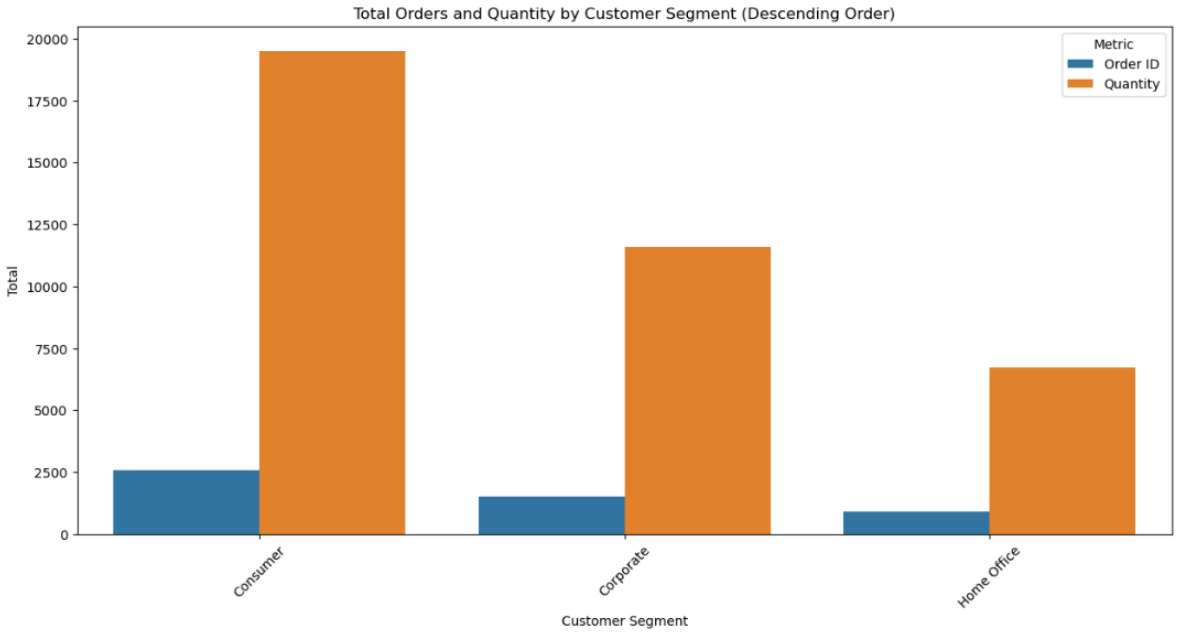

Total Orders and Quantity by Customer Segment (Descending Order)
Metric
Order ID
Quantity
Total
20000
17500
15000
12500
10000
7500
5000
2500
0
Consumer
Corporate
Home Office
Customer Segment

Fig. 26

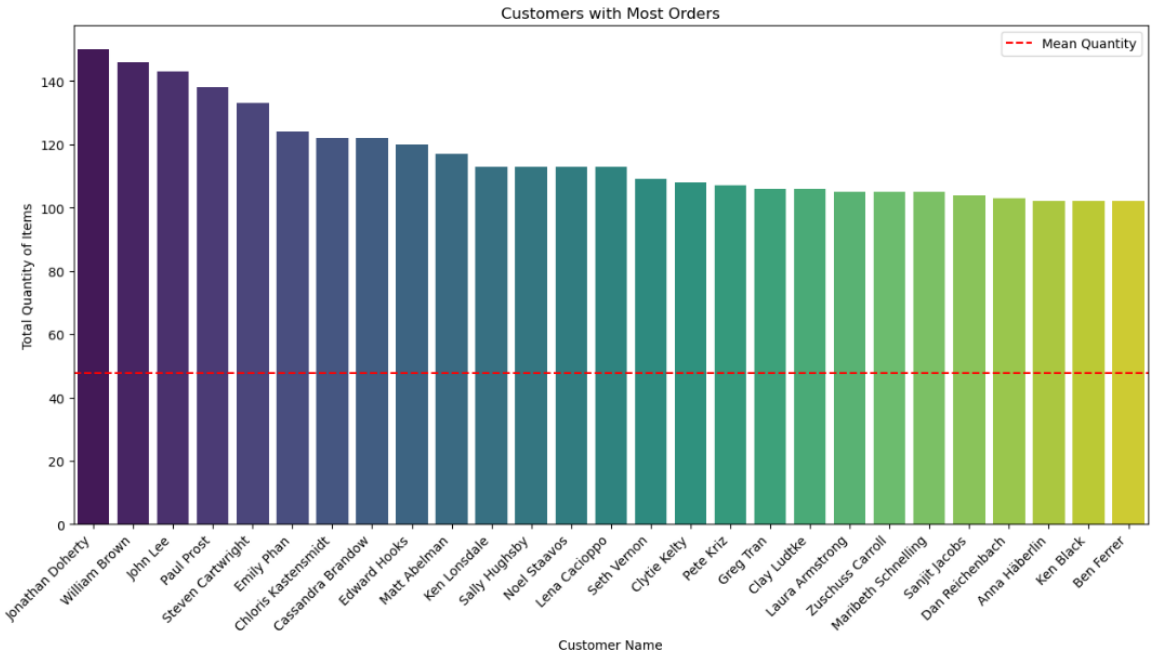

Customers with Most Orders
Mean Quantity
Total Quantity of Items
140
120
100
80
60
40
20
0
Jonathan Doherty
William Brown
John Lee
Paul Prost
Steven Cartwright
Emily Phan
Chloris Kastensmidt
Cassandra Brandow
Edward Hooks
Matt Abelman
Ken Lonsdale
Sally Hughsby
Noel Staavos
Lena Cacioppo
Seth Vernon
Clytie Kelly
Pete Kriz
Greg Tran
Clay Ludtke
Laura Armstrong
Zuschuss Carroll
Maribeth Schnelling
Sanjit Jacobs
Dan Reichenbach
Anna Häberlin
Ken Black
Ben Ferrer
Customer Name

Fig. 27

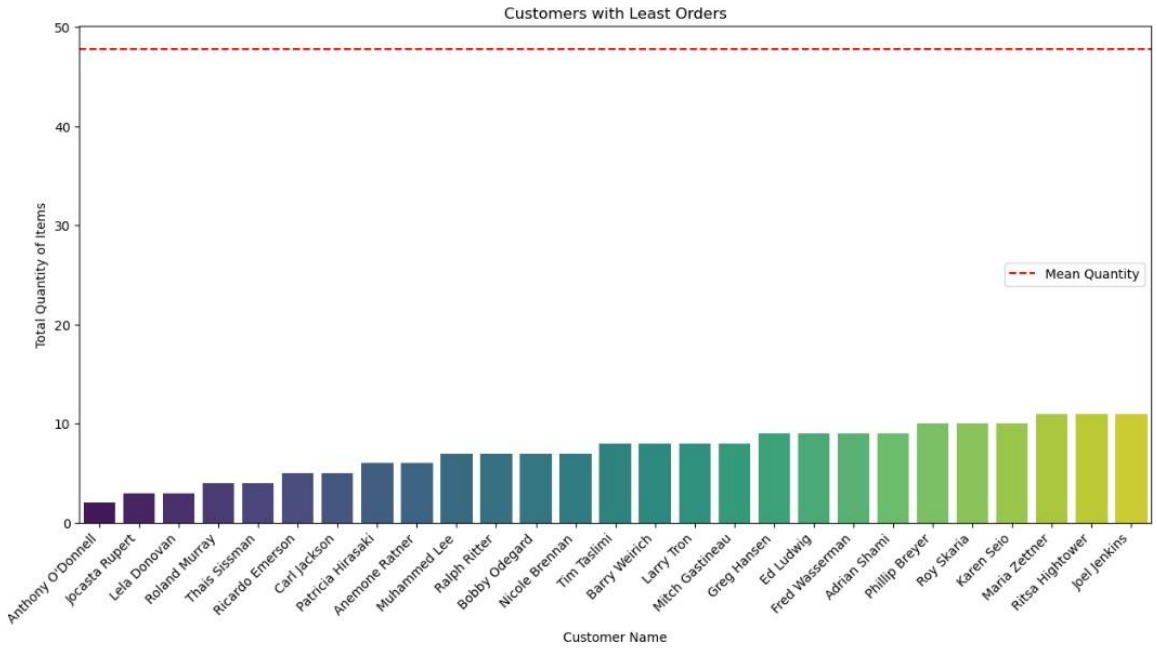

Customers with Least Orders
Total Quantity of Items
Mean Quantity
Customer Name
Anthony O'Donnell
Jocasta Rupert
Lela Donovan
Roland Murray
Thais Sissman
Ricardo Emerson
Carl Jackson
Patricia Hirasaki
Anemone Ratner
Muhammed Lee
Ralph Ritter
Bobby Odegard
Nicole Brennan
Tim Taslimi
Barry Weirich
Larry Tron
Mitch Gastineau
Greg Hansen
Ed Ludwig
Fred Wasserman
Adrian Shami
Phillip Breyer
Roy Skaria
Karen Seio
Maria Zettner
Ritsa Hightower
Joel Jenkins

Fig. 28

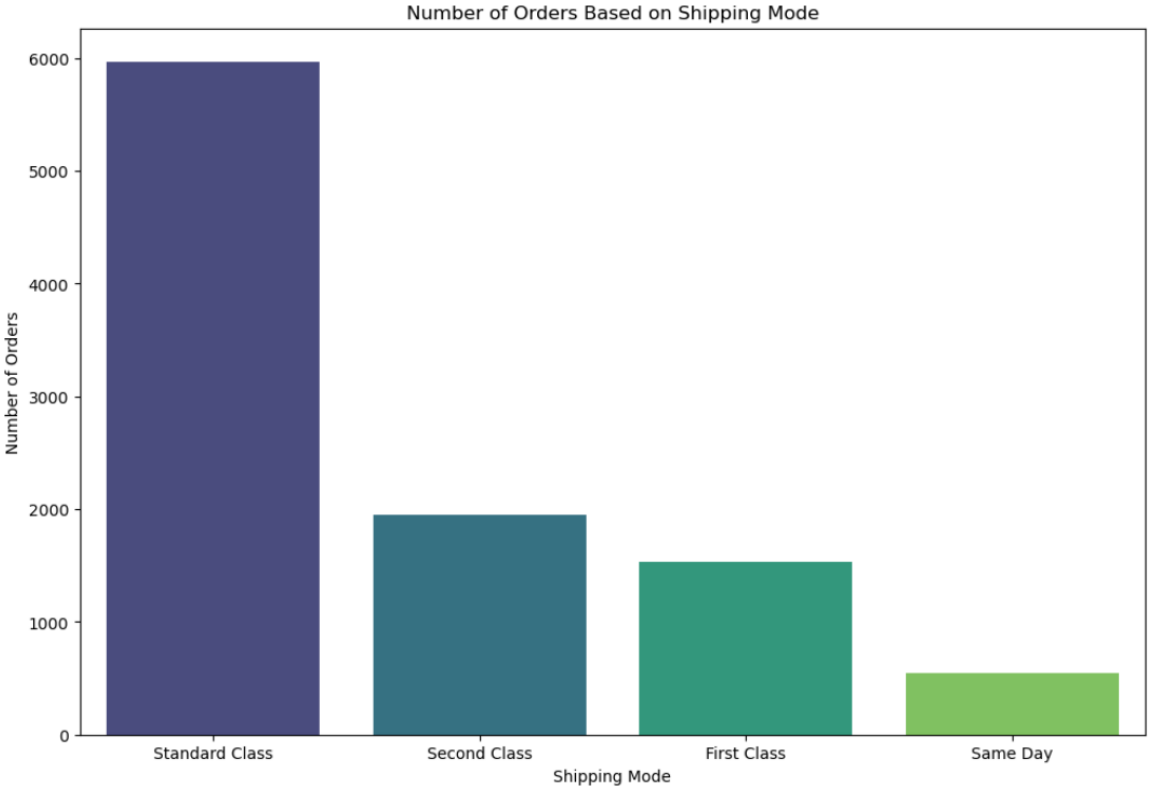

Number of Orders Based on Shipping Mode
Number of Orders
Shipping Mode
Standard Class
Second Class
First Class
Same Day

Fig. 29

Inferences:

- **Consumer Dominance:** Consumer segment leads in orders and quantities, highlighting a key market dominance leaving room for improvement for other sectors.

- **Cost-Effective Shipping:** Standard Class shipping prevails, indicating a cost-focused customer preference.

- **Segment Optimization:** Target mid-tier for engagement; gather top customer feedback for personalized strategies and growth.

4. Results:

Considering the inferences drawn already, there are critical insights for the superstore fuelling strategic optimization for enhanced profitability and competitiveness. In product portfolio evaluation, high performing categories and products help, emphasizing a successful high-margin, low-volume approach. Recommendations include intensified marketing and inventory allocation efforts for these categories, alongside targeted adjustments for low-sales and loss-making products. The seasonal influences in sales, emphasize the need to understand and capitalize on these trends. The negative correlation between discount and profit suggests a strategic revamp of discounting approaches, aligning with the importance of non-discounted sales in overall profitability. The wide profit margin distribution prompts a focused product mix optimization, prioritizing high-margin products.

Regional market analysis identifies performance variations, with the Northeast and West excelling while the South lags, also two states occupying most of profitability. Tailored strategies for underperforming regions and opportunities in underpenetrated rural markets are recommended. This involves restructuring marketing approaches and expanding distribution networks for growth while intensifying focus in best states. Customer-centric analysis reveals behaviour patterns, allowing tailored engagement strategies. High-order customers offer loyalty program opportunities, and shipping preference insights suggest cost-effective options. Profitability analysis by customer segments recommends upselling for the Corporate segment and differentiated marketing for the Consumer segment. In summary, the results provide a comprehensive roadmap, emphasizing strategic focus on high-margin products, optimized discounting, regional growth strategies, and customer-centric approaches success.

5. Conclusions:

The insights from the Superstore Giant's retail data hold substantial managerial implications for practitioners. For product portfolio managers, the emphasis on the Technology category should guide resource allocation to products with the highest profit margins. By focusing on high-margin, low-volume strategies, the superstore can maximize profitability. Efficient inventory management, especially for high-value items, could reduce holding costs and increase cash flow. The management needs to scrutinize low-sales and loss-making products, using data to inform the discontinuation or optimization of such items.

These insights must be leveraged by marketing department to align promotional strategies with seasonal sales patterns, ensuring campaigns are targeted and efficient. Regional managers should note the disparities in regional performance and allocate marketing efforts

accordingly to improve underperforming markets. This may involve a more localized product mix and marketing strategy, especially in areas like the South region. Customer engagement department should consider segment-specific approaches. The data supports the implementation of loyalty programs and targeted offers to engage both high-volume and infrequent customers, potentially boosting profitability. Shipping options also present a tactical opportunity, suggesting that promoting cost-effective shipping could resonate with a larger customer base.

Excluding supply chain and inventory data limits understanding of logistical efficiency and stock management, potentially leading to supply shortfalls or excess. Lack of competitor insights hinders competitive strategy formation. Ignoring e-commerce trends and customer reviews can result in missed online market opportunities and a poor grasp of customer satisfaction and product feedback. Additionally, regional performance analysis might be affected by external factors not considered in the data, such as economic shifts or competitive actions.

Future research for Superstore Giant should integrate predictive analytics for market trends, delve into supply chain and inventory for logistical efficiency, and include competitor benchmarking for strategic positioning. Emphasizing e-commerce trends, digital strategies, and customer sentiment analysis is vital to understand online behavior and preferences. Incorporating real-time data analytics and exploring diverse pricing strategies will facilitate agile decision-making and a comprehensive understanding of dynamic market conditions.

6. REFERENCES:

Stepanek, H. (2020) *Thinking in pandas: How to use the Python Data Analysis Library the right way*. New York: Apress.

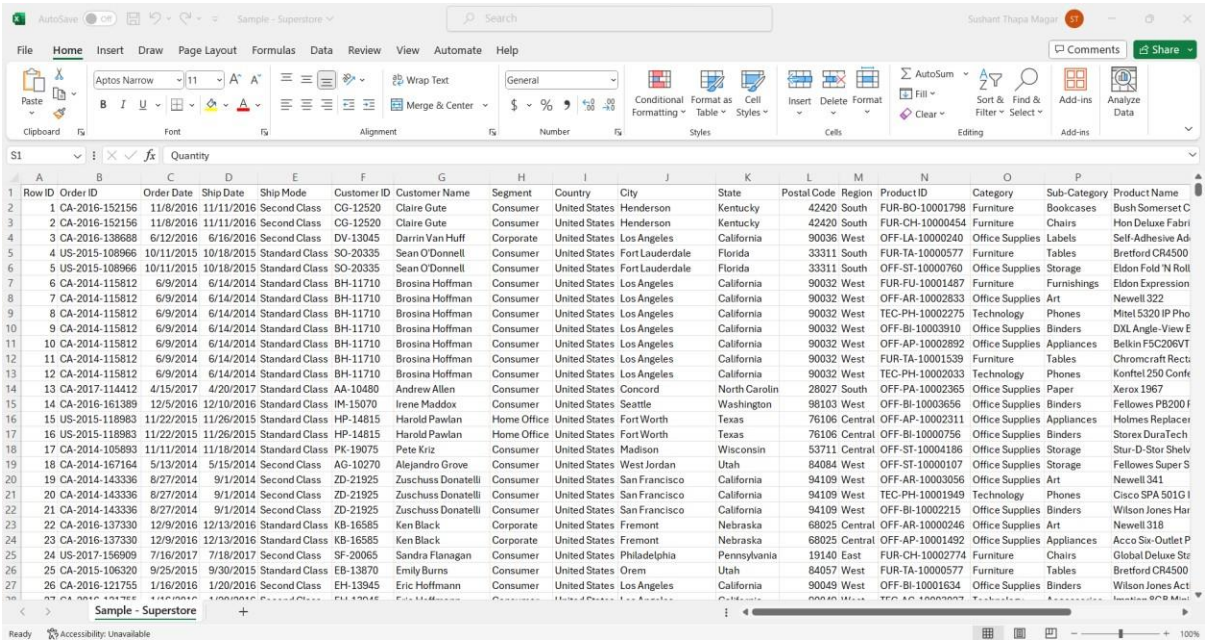
Phuong, V.T.H. (2017) *Python: Data analytics and visualization: Understand, evaluate, visualize data, a course in three modules*. Birmingham, UK: Packt Publishing.

Jafari, R. (2022) *Hands-on data preprocessing in python: Learn how to effectively prepare data for successful data analytics*. Birmingham: Packt Publishing.

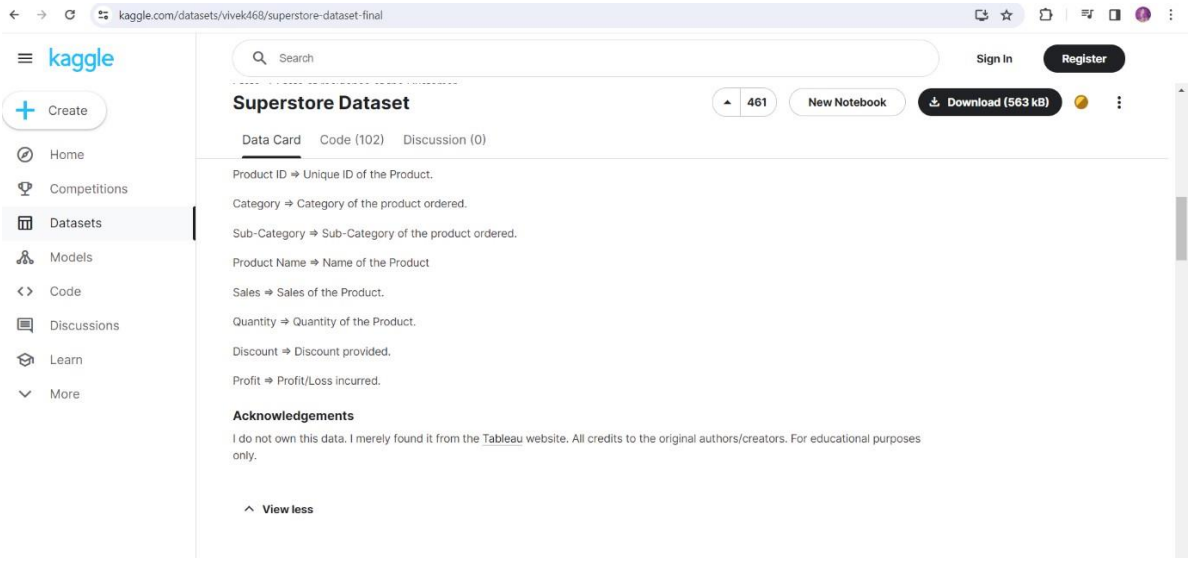
Haleem, F. and Jehangir, M. (2017) 'Strategic management practices by Morrison plc, UK. analysis, lessons and implications', *Middle East Journal of Business*, 12(3), pp. 3–9. doi:10.5742/mejb.2017.92969.

M. Singh, B. Ghutla, R. L. Jnr, A. F. Mohammed, and M. A. Rashid, "Walmart's sales data analysis-A big data analytics perspective," in *Proceedings of the 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, pp. 114–119, Mana Island, Fiji, December 2017.

7. Appendices (Data & Code)



Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name
1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset C
2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabri
3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Ad
4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500
5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll
6	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-10001487	Furniture	Furnishings	Eldon Expression
7	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-10002833	Office Supplies	Art	Newell 322
8	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10002275	Technology	Phones	Mitel 5320 IP Pho
9	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-10003910	Office Supplies	Binders	DXL Angle-View E
10	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10002892	Office Supplies	Appliances	Belkin F5C206VT
11	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-10001539	Furniture	Tables	Chromcraft Rect
12	CA-2014-115812	6/9/2014	6/14/2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10002033	Technology	Phones	Konftel 250 Conf
13	CA-2017-114412	4/15/2017	4/20/2017	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolin	28027	South	OFF-PA-10002365	Office Supplies	Paper	Xerox 1967
14	CA-2016-161389	12/5/2016	12/10/2016	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-10003656	Office Supplies	Binders	Fellowes PB200
15	US-2015-118983	11/22/2015	11/26/2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-10002311	Office Supplies	Appliances	Holmes Replacer
16	US-2015-118983	11/22/2015	11/26/2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10000756	Office Supplies	Binders	Storex DuraTech
17	CA-2014-105893	11/11/2014	11/18/2014	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10004186	Office Supplies	Storage	Stur-D-Stor Shelv
18	CA-2014-167164	5/13/2014	5/15/2014	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-10000107	Office Supplies	Storage	Fellowes Super S
19	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AR-10003056	Office Supplies	Art	Newell 341
20	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	TEC-PH-10001949	Technology	Phones	Cisco SPA 501G
21	CA-2014-143336	8/27/2014	9/1/2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-10002215	Office Supplies	Binders	Wilson Jones Har
22	CA-2016-137330	12/9/2016	12/13/2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-10000246	Office Supplies	Art	Newell 318
23	CA-2016-137330	12/9/2016	12/13/2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-10001492	Office Supplies	Appliances	Acco Six-Outlet P
24	US-2017-156909	7/16/2017	7/18/2017	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-10002774	Furniture	Chairs	Global Deluxe Sta
25	CA-2015-106320	9/25/2015	9/30/2015	Standard Class	EB-13870	Emily Burns	Consumer	United States	Orem	Utah	84057	West	FUR-TA-10000577	Furniture	Tables	Bretford CR4500
26	CA-2016-121755	1/16/2016	1/20/2016	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	California	90049	West	OFF-BI-10001634	Office Supplies	Binders	Wilson Jones Act



kaggle.com/datasets/vivek468/superstore-dataset-final

kaggle

Search

Sign In | Register

Create

Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

Superstore Dataset

461 | New Notebook | Download (563 kB)

Data Card | Code (102) | Discussion (0)

Product ID ⇒ Unique ID of the Product.

Category ⇒ Category of the product ordered.

Sub-Category ⇒ Sub-Category of the product ordered.

Product Name ⇒ Name of the Product

Sales ⇒ Sales of the Product.

Quantity ⇒ Quantity of the Product.

Discount ⇒ Discount provided.

Profit ⇒ Profit/Loss incurred.

Acknowledgements

I do not own this data. I merely found it from the Tableau website. All credits to the original authors/creators. For educational purposes only.

View less

kaggle.com/datasets/vivek468/superstore-dataset-final

kaggle

Search

Sign In | Register

Create

Home
Competitions
Datasets
Models
Code
Discussions
Learn
More

VIVEK CHOWDHURY · UPDATED 2 YEARS AGO

461 | New Notebook | Download (563 kB)

Superstore Dataset

Dataset containing Sales & Profits of a Superstore

Data Card | Code (102) | Discussion (0)

About Dataset

Context

With growing demands and cut-throat competitions in the market, a Superstore Giant is seeking your knowledge in understanding what works best for them. They would like to understand which products, regions, categories and customer segments they should target or avoid.

You can even take this a step further and try and build a Regression model to predict Sales or Profit.

Go crazy with the dataset, but also make sure to provide some business insights to improve.

Usability 10.00

License Other (specified in description)

Expected update frequency Never

Tags Business

jupyter superstoredataset - Copy - Copy Last Checkpoint: 4 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [1]:
```
#IMPORTING RELEVANT LIBRARIES

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import os
import seaborn as sns
```

In [2]:
```
#IMPORTING THE CSV FILE (SUPERSTORE DATASET)
df = pd.read_csv("C:\\Users\\magar\\OneDrive\\Desktop\\RHUL\\Assignments\\Businss Analytics Language\\Sample - Superstore.csv
```

In [3]:
```
#UNDERSTANDING THE CHARACTERISTICS OF THE GIVEN DATA SET
df.head() #Displaying the first few rows of the DataFrame to get a quick overview.
```

Out[3]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	Sub-Category
0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-BO-10001798	Furniture	Bookcases
1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-CH-10000454	Furniture	Chairs
2	3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	90036	West	OFF-LA-10000240	Office Supplies	Labels
3	4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	FUR-TA-10000577	Furniture	Tables
4	5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	OFF-ST-10000760	Office Supplies	Storage

jupyter superstoredataset - Copy - Copy Last Checkpoint: 5 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3 (ipykernel)

Run Code

4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	Sean O'Donnell	Consumer	Fort Lauderdale	Florida	South	OFF-ST-10000760	...	Eldon Fold 'N Roll Cart System	22.3680	2	0.20

5 rows × 22 columns

In [21]:
```
#Displaying some of the important KPIs
# Calculating KPIs
total_sales = df['Sales'].sum()
total_orders = df['Order ID'].nunique()
total_quantity_sold = df['Quantity'].sum()
total_profit = df['Profit'].sum()
profit_margin = (total_profit / total_sales) * 100
number_of_customers = df['Customer Name'].nunique()

# Preparing data for visualization
print("Total Sales:",total_sales)
print("Total Quantity Sold:",total_quantity_sold)
print("Profit Margin:", profit_margin)
print("Total Profit:",total_profit)
print("Total customers:",number_of_customers)
print("Total orders:",total_orders)

KPIs = {'KPI': ['Total Sales', 'Total Quantity Sold', 'Total Profit','Profit Margin', 'Total Customers', 'Total Orders'],
    'Value': [total_sales, total_quantity_sold, total_profit,profit_margin, number_of_customers, total_orders]}

kpi_df = pd.DataFrame(KPIs)
```

```
Total Sales: 2296919.4883000003
Total Quantity Sold: 37871
Profit Margin: 12.469269469779174
Total Profit: 286409.0805
Total customers: 793
Total orders: 5009
```

In [22]:
```
# Displaying some of the important KPIs
# Calculating KPIs

average_discount_rate = df['Discount'].mean()
average_quantity_per_order = df['Quantity'].sum() / df['Order ID'].nunique()
```

jupyter superstoredataset - Copy - Copy Last Checkpoint: 4 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

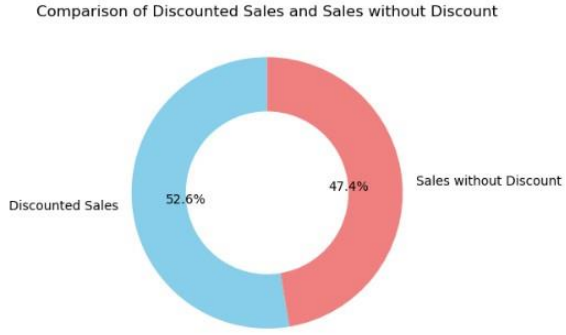
Trusted Python 3 (ipykernel)

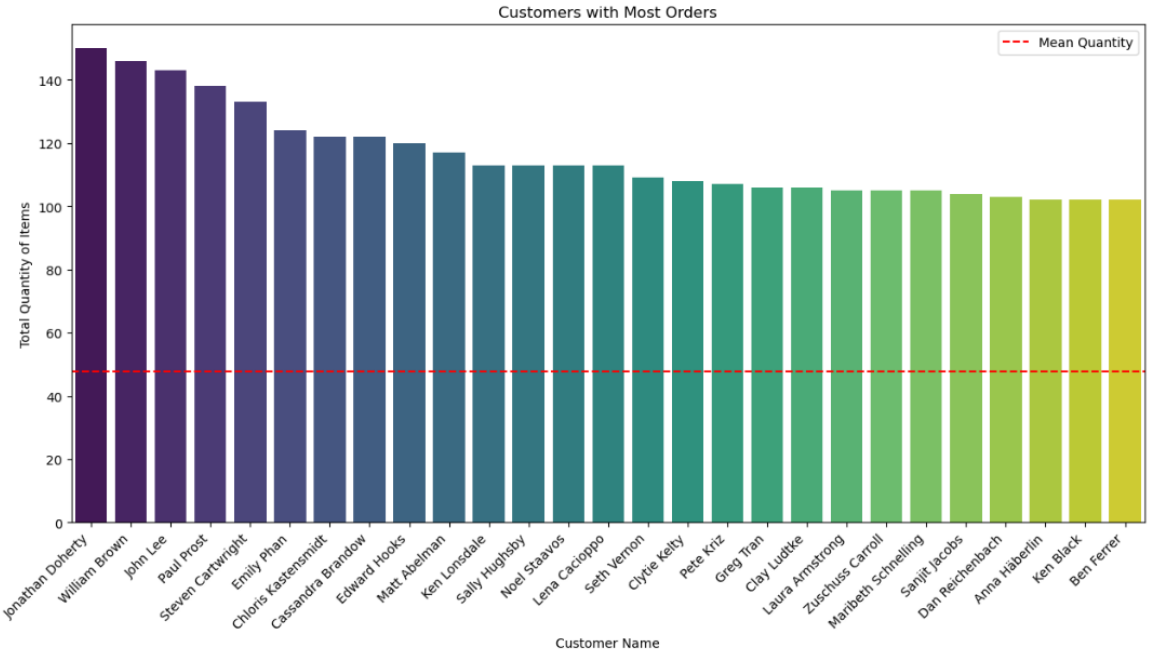
Run Code

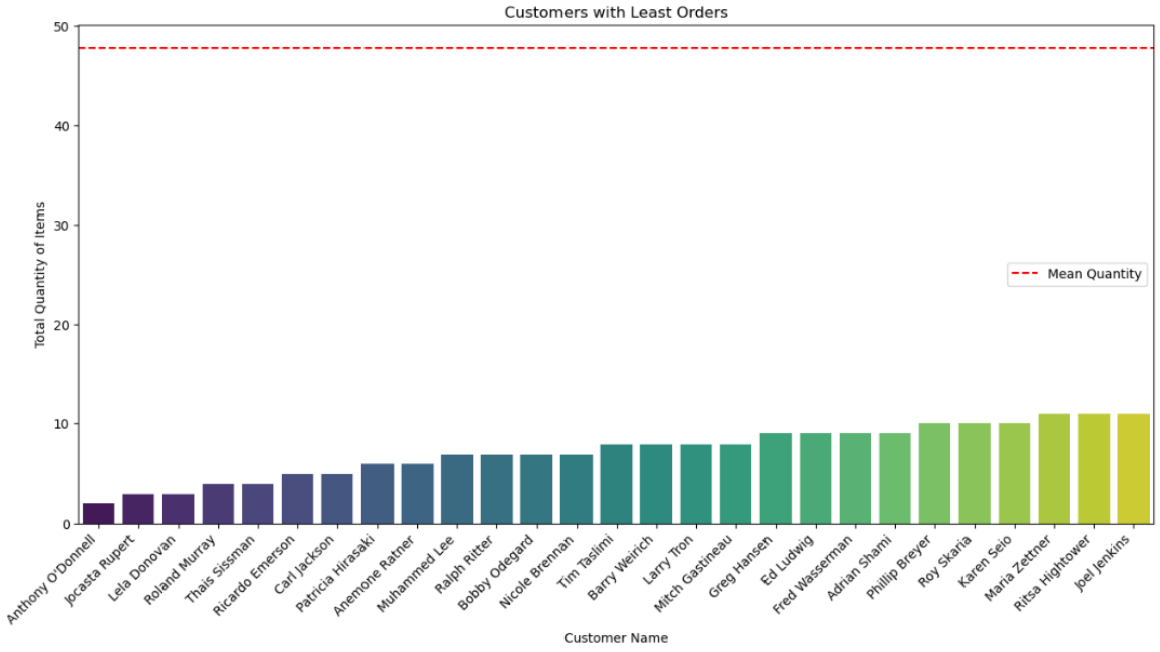
In [54]:
```
# Calculate correlation coefficient between Profit and Discount
correlation_profit_discount = df['Profit'].corr(df['Discount'])

# Calculate correlation coefficient between Sales and Discount
correlation_sales_discount = df['Sales'].corr(df['Discount'])

print(f'Correlation between Profit and Discount: {correlation_profit_discount:.2f}')
print(f'Correlation between Sales and Discount: {correlation_sales_discount:.2f}')
```

```
Correlation between Profit and Discount: -0.22
Correlation between Sales and Discount: -0.03
```

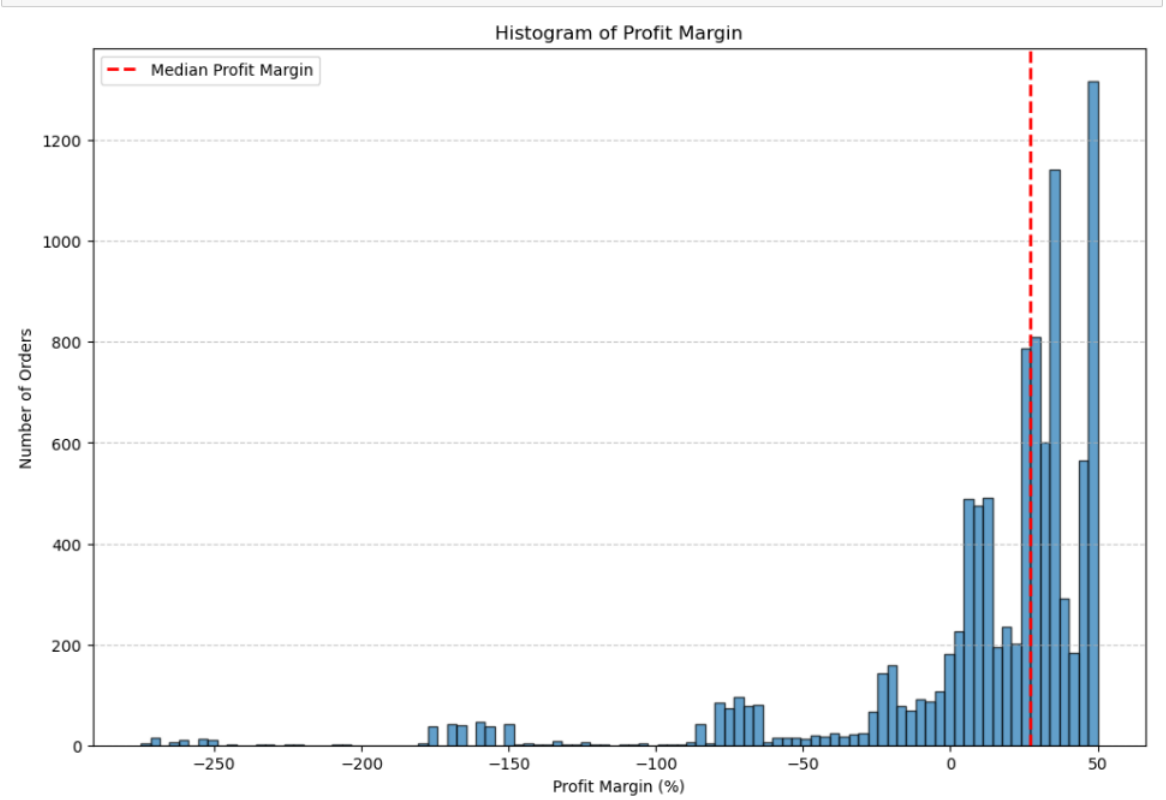
In [55]:
```
# Calculate total sales with discount and without discount
total_sales_with_discount = df[df['Discount'] > 0]['Sales'].sum()
total_sales_without_discount = df[df['Discount'] == 0]['Sales'].sum()

# Create a pie chart
labels = ['Discounted Sales', 'Sales without Discount']
sizes = [total_sales_with_discount, total_sales_without_discount]
colors = ['skyblue', 'lightcoral']

plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=90, colors=colors, wedgeprops=dict(width=0.4))

# Adding a title
plt.title('Comparison of Discounted Sales and Sales without Discount')

plt.show()
```




Customers with Most Orders
Mean Quantity
Total Quantity of Items
0
20
40
60
80
100
120
140
Jonathan Doherty
William Brown
John Lee
Paul Prost
Steven Cartwright
Emily Phan
Chloris Kastensmidt
Cassandra Brandow
Edward Hooks
Matt Abelman
Ken Lonsdale
Sally Hughsby
Noel Staavos
Lena Cacioppo
Seth Vernon
Clytie Kelty
Pete Kriz
Greg Tran
Clay Ludtke
Laura Armstrong
Zuschuss Carroll
Maribeth Schnelling
Sanjit Jacobs
Dan Reichenbach
Anna Häberlin
Ken Black
Ben Ferrer
Customer Name


Customers with Least Orders
Mean Quantity
Total Quantity of Items
0
10
20
30
40
50
Anthony O'Donnell
Jocasta Rupert
Lela Donovan
Roland Murray
Thais Sissman
Ricardo Emerson
Carl Jackson
Patricia Hirasaki
Anemone Ratner
Muhammed Lee
Ralph Ritter
Bobby Odegard
Nicole Brennan
Tim Taslimi
Barry Weirich
Larry Tron
Mitch Gastineau
Greg Hansen
Ed Ludwig
Fred Wasserman
Adrian Shami
Phillip Breyer
Roy Skaria
Karen Seio
Maria Zettner
Ritsa Hightower
Joel Jenkins
Customer Name

KPIs	Values
Total Sales	2,296,919
Total Quantity Sold	37,871
Total Profit	286,409
Average Profit Margin	12.47%
Total Customers	793.00
Total Orders	5,009
Average Discount Rate	16.00%
Average Quantity Per Order	7.16
Average Sales Per Order	459
Average Profit Per Order	57.18



Region	Sales	Profit	Profitability
West	725,457.82	108,418.45	14.94%
East	678,499.87	91,534.84	13.49%
Central	501,239.89	39,706.36	7.92%
South	391,721.91	46,749.43	11.93%