

Superstore

Dataset Analysis



Table of Content

1.

1

Introduction to power BI

2

Data description

3

Problem or opportunity formulation

4

Proposed solution

5

Conclusion

6

References

Introduction

Power BI is a business intelligence tool by Microsoft that transforms raw data into interactive reports and dashboards. It supports real-time data connectivity, AI-powered analytics, and seamless integration with Excel, Azure, and Teams. With its user-friendly interface and strong collaboration features, Power BI helps organizations make data-driven decisions efficiently.



Data description



The dataset represents sales transactions data from a retail superstore.

Snapshots

Profit	Discount	Quantity	Sales	Product Name	Sub-Category	Category	Product ID	Region	Postal Code	State	City	Country	Segment	Customer	Customer ID	Ship Mode	Ship Date	Order Date	Order ID	Row ID	
41.9136	0	2	261.96	Bush Somerset	Bookcases	Furniture	FUR-BO-10	South	42420	Kentucky	Henderson	United States	Consumer	Claire Gutierrez	CG-12520	Second Class	11/11/2016	11/8/2016	CA-2016-1	1	2
219.582	0	3	731.94	Hon Delux	Chairs	Furniture	FUR-CH-10	South	42420	Kentucky	Henderson	United States	Consumer	Claire Gutierrez	CG-12520	Second Class	11/11/2016	11/8/2016	CA-2016-1	2	3
6.8714	0	2	14.62	Self-Adhes	Labels	Office Supplies	OFF-LA-10	West	90036	California	Los Angeles	United States	Corporate	Darrin Van	DV-13045	Second Class	6/16/2016	6/12/2016	CA-2016-1	3	4
-383.031	0.45	5	957.5775	Bretford	Tables	Furniture	FUR-TA-10	South	33311	Florida	Fort Lauderdale	United States	Consumer	Sean O'Donoghue	SO-20335	Standard Class	10/18/2015	10/11/2015	US-2015-1	4	5
2.5164	0.2	2	22.368	Eldon Fold	Storage	Office Supplies	OFF-ST-10	South	33311	Florida	Fort Lauderdale	United States	Consumer	Sean O'Donoghue	SO-20335	Standard Class	10/18/2015	10/11/2015	US-2015-1	5	6
14.1694	0	7	48.86	Eldon Express	Furnishing	Furniture	FUR-FU-10	West	90032	California	Los Angeles	United States	Consumer	Brosina Hector	BH-11710	Standard Class	6/14/2014	6/9/2014	CA-2014-1	6	7
1.9656	0	4	7.28	Newell 32	Art	Office Supplies	OFF-AR-10	West	90032	California	Los Angeles	United States	Consumer	Brosina Hector	BH-11710	Standard Class	6/14/2014	6/9/2014	CA-2014-1	7	8
90.7152	0.2	6	907.152	Mitel 5320	Phones	Technology	TEC-PH-10	West	90032	California	Los Angeles	United States	Consumer	Brosina Hector	BH-11710	Standard Class	6/14/2014	6/9/2014	CA-2014-1	8	9
5.7825	0.2	3	18.504	DXL Angle	Binders	Office Supplies	OFF-BI-10	West	90032	California	Los Angeles	United States	Consumer	Brosina Hector	BH-11710	Standard Class	6/14/2014	6/9/2014	CA-2014-1	9	10
26.6304	0	1	55.48	Eaton Pre	Paper	Office Supplies	OFF-PA-10	East	11757	New York	Lindenhurst	United States	Corporate	Maya Hernandez	MH-17785	Standard Class	12/24/2015	12/20/2015	CA-2015-1	471	472
3.4092	0	3	340.92	SAFCO Bol	Storage	Office Supplies	OFF-ST-10	West	94110	California	San Francisco	United States	Consumer	Jeremy Piss	JP-15520	Second Class	11/11/2014	11/9/2014	CA-2014-1	472	473
10.4784	0.15	2	222.666	Sauder Mis	Bookcases	Furniture	FUR-BO-10	West	94110	California	San Francisco	United States	Consumer	Jeremy Piss	JP-15520	Second Class	11/11/2014	11/9/2014	CA-2014-1	473	474
87.996	0.2	4	703.968	Samsung C	Phones	Technology	TEC-PH-10	West	94110	California	San Francisco	United States	Consumer	Jeremy Piss	JP-15520	Second Class	11/11/2014	11/9/2014	CA-2014-1	474	475
24.9804	0	6	92.52	Decoflex H	Storage	Office Supplies	OFF-ST-10	West	94110	California	San Francisco	United States	Consumer	Jeremy Piss	JP-15520	Second Class	11/11/2014	11/9/2014	CA-2014-1	475	476
28.819	0	7	62.65	Recycled E	Paper	Office Supplies	OFF-PA-10	West	94110	California	San Francisco	United States	Consumer	Jeremy Piss	JP-15520	Second Class	11/11/2014	11/9/2014	CA-2014-1	476	477
45.528	0	5	94.85	Xerox 1912	Paper	Office Supplies	OFF-PA-10	West	94110	California	San Francisco	United States	Consumer	Jeremy Piss	JP-15520	Second Class	11/11/2014	11/9/2014	CA-2014-1	477	478
7.182	0.2	6	95.76	Cisco SPA	Phones	Technology	TEC-PH-10	West	90008	California	Los Angeles	United States	Corporate	Kristen Harris	KH-16690	Standard Class	7/19/2016	7/12/2016	CA-2016-1	478	479
19.296	0	3	40.2	Electrix 20	Furnishing	Furniture	FUR-FU-10	East	12180	New York	Troy	United States	Consumer	Jeremy Ellis	JE-15475	Standard Class	11/2/2016	10/27/2016	CA-2016-1	479	480
6.615	0	5	14.7	Prang Colo	Art	Office Supplies	OFF-AR-10	East	10024	New York	New York City	United States	Corporate	John Grady	JG-15805	Standard Class	7/2/2016	6/26/2016	CA-2016-1	480	481

Data Type

Column Name	Data Type
Row ID	int64
Order ID	object “String”
Order Date	object “String”
Ship Date	object “String”
Ship Mode	object “String”
Customer ID	object “String”
Customer Name	object “String”
Segment	object “String”
Country	object “String”
City	object “String”
State	object “String”
Postal Code	int64
Region	object “String”
Product ID	object “String”
Category	object “String”
Sub-Category	object “String”
Product Name	object “String”
Sales	float64
Quantity	int64
Discount	float64
Profit	float64

Column Name	Data Type
Order Date	object “String”
Ship Date	object “String”
Ship Mode	object “String”
Segment	object “String”
City	object “String”
State	object “String”
Region	object “String”
Category	object “String”
Sub-Category	object “String”
Sales	float64
Quantity	int64
Discount	float64
Profit	float64

Attribute description

Column Name	Description
Row ID	Unique ID for each row
Order ID	Unique order ID for each Customer
Order Date	Order Date of the product
Ship Date	Shipping Date of the product
Ship Mode	Shipping mode specified by the customer
Customer ID	Unique ID to identify each Customer
Customer Name	Name of the Customer
Segment	The segment where the Customer belongs
Country	Country of residence of the Customer
City	City of residence of the Customer
State	State of residence of the Customer
Postal Code	Postal Code of every Customer
Region	Region where the Customer belong
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 of the product
Profit	Profit/Loss incurred

Column Name	Description
Order Date	Order Date of the product
Ship Date	Shipping Date of the product
Ship Mode	Shipping mode specified by the customer
Segment	The segment where the Customer belongs
Country	Country of residence of the Customer
City	City of residence of the Customer
State	State of residence of the Customer
Region	Region where the Customer belong
Category	Category of the product ordered
Sub-Category	Sub-Category of the product ordered
Sales	Sales of the product
Quantity	Quantity of the product
Discount	Discount of the product
Profit	Profit/Loss incurred

RELATIONS BETWEEN COLUMNS:

Correlation Matrix

	Sales	Quantity	Discount	Profit
Sales	1	0.200795	-0.02819	0.479064
Quantity	0.200795	1	0.008623	0.066253
Discount	-0.02819	0.008623	1	-0.21949
Profit	0.479064	0.066253	-0.21949	1

1.

Sales and Profit

Positive moderate correlation (0.48)
Indicates higher sales generally lead to higher profits, though not always proportionally.

3.

Quantity and Sales:

Weak positive correlation (0.20)
Selling more items slightly tends to increase overall sales.

2.

Discount and Profit

Negative correlation (-0.22)
Higher discounts tend to reduce profit margins, highlighting that increasing discounts can negatively affect profitability.

4.

Discount and Sales

Very weak negative correlation (-0.03)
Discounts do not significantly boost sales volume, implying discount strategies might need reassessment

RELATIONS BETWEEN COLUMNS:

chi-Square analysis

Ship Mode and Segment

Strong evidence (p-value = 0.00009) indicates shipping preferences differ significantly across customer segments.



Ship Mode and Region

Significant (p-value = 0.00455), suggesting that different regions prefer distinct shipping modes.



Category and Sub-Category

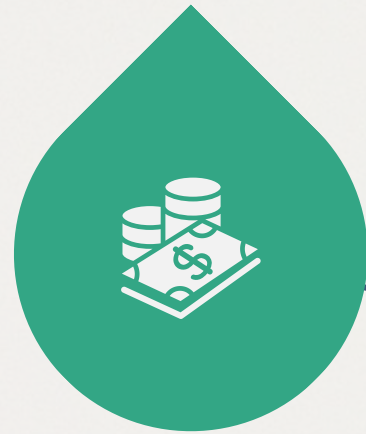
Highly significant relationship (p-value ≈ 0.0) naturally expected as sub-categories directly relate to their parent categories.



DATA ROLE IN DECISION-MAKING

Profitability Management

Columns such as Profit, Sales, Discount, and Quantity are critical for identifying profitable and non-profitable products or categories.



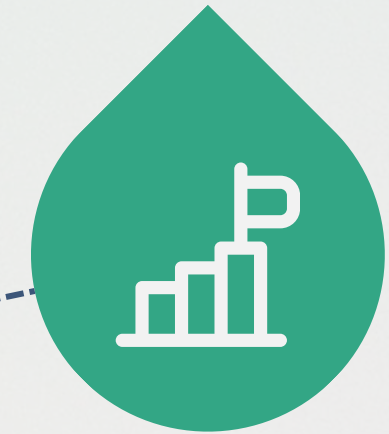
Sales and Shipping Optimization

Using Order Date, Ship Date, and Ship Mode, logistical processes can be optimized by understanding shipment timelines and customer preferences, ultimately improving service efficiency and customer satisfaction.



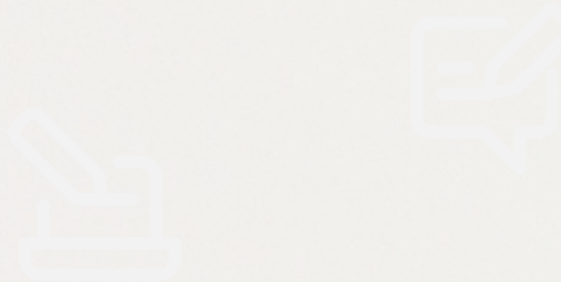
Performance Monitoring and Forecasting

Historical sales data (Order Date and Sales) can be utilized to forecast future demand, aiding decisions on resource planning, inventory management, and sales targets.



Inventory and Product

By using Category, Sub-Category, and Product Name, businesses can pinpoint which products are performing well, require restocking, or should be phased out.



Geographical and Market Expansion

Data on City, State, Region, and Postal Code highlights geographic areas with high or low sales.



THE DISCOUNT DILEMMA: LOSSING PROFIT IN THE PURSUIT OF SALES

Problem Statement

- High Sales \neq High Profit
- Discounts are reducing overall profit
- Data review revealed the issue

Why IT Matters ?

- Profit is keeps the business running
- Uncontrolled discounting $=$ Long - term loss

WHO'S AFFECTED ?



Sales Team

Think High Sales = Success
Unaware of shrinking profits



Business Impact

Harder to plan, forecast, or
stay profitable



Finance Team

Struggles to justify high sales with
low profits

IMPACT & BENEFITS OF THE PROBLEM

1. Lower Profit

1.

Better
Pricing
Strategy

4.

Better
Team
Collaboration

4. Poor Budgeting

2.

Smarter
Promotions

5.

Clearer
Financial
Reports

5. Wasted Resource

2. Unsold Stock

3.

Improved
Profit
Margins

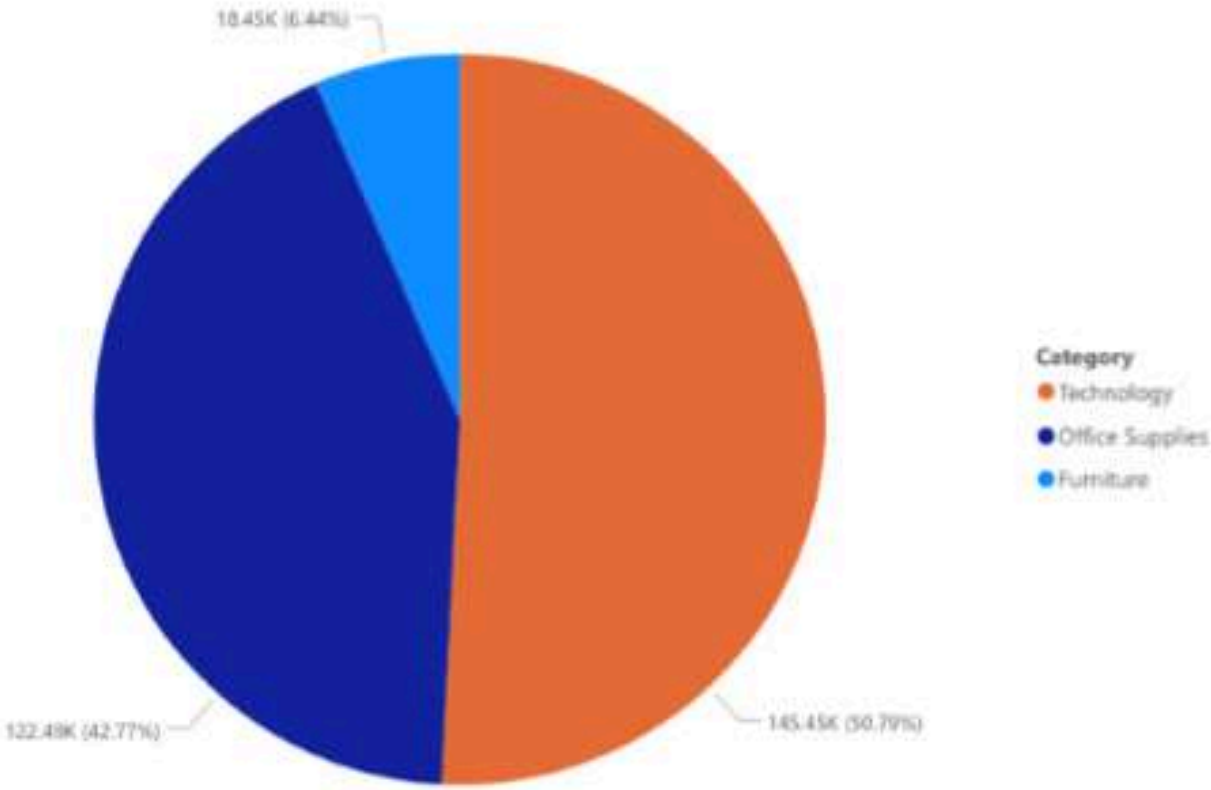
6.

Sustainable
Growth

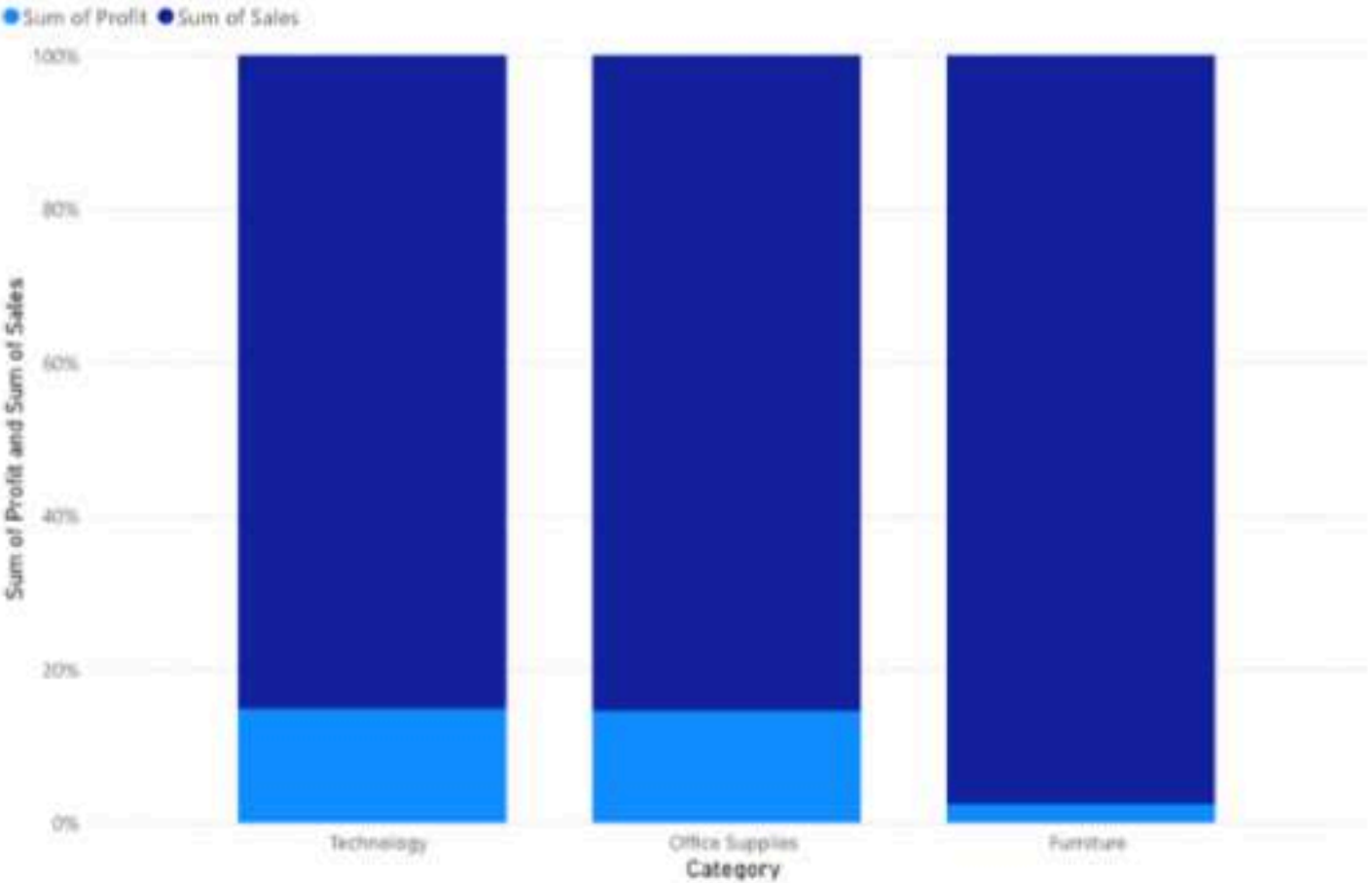
6. Harder to Compute

3. Cash Flow Issue

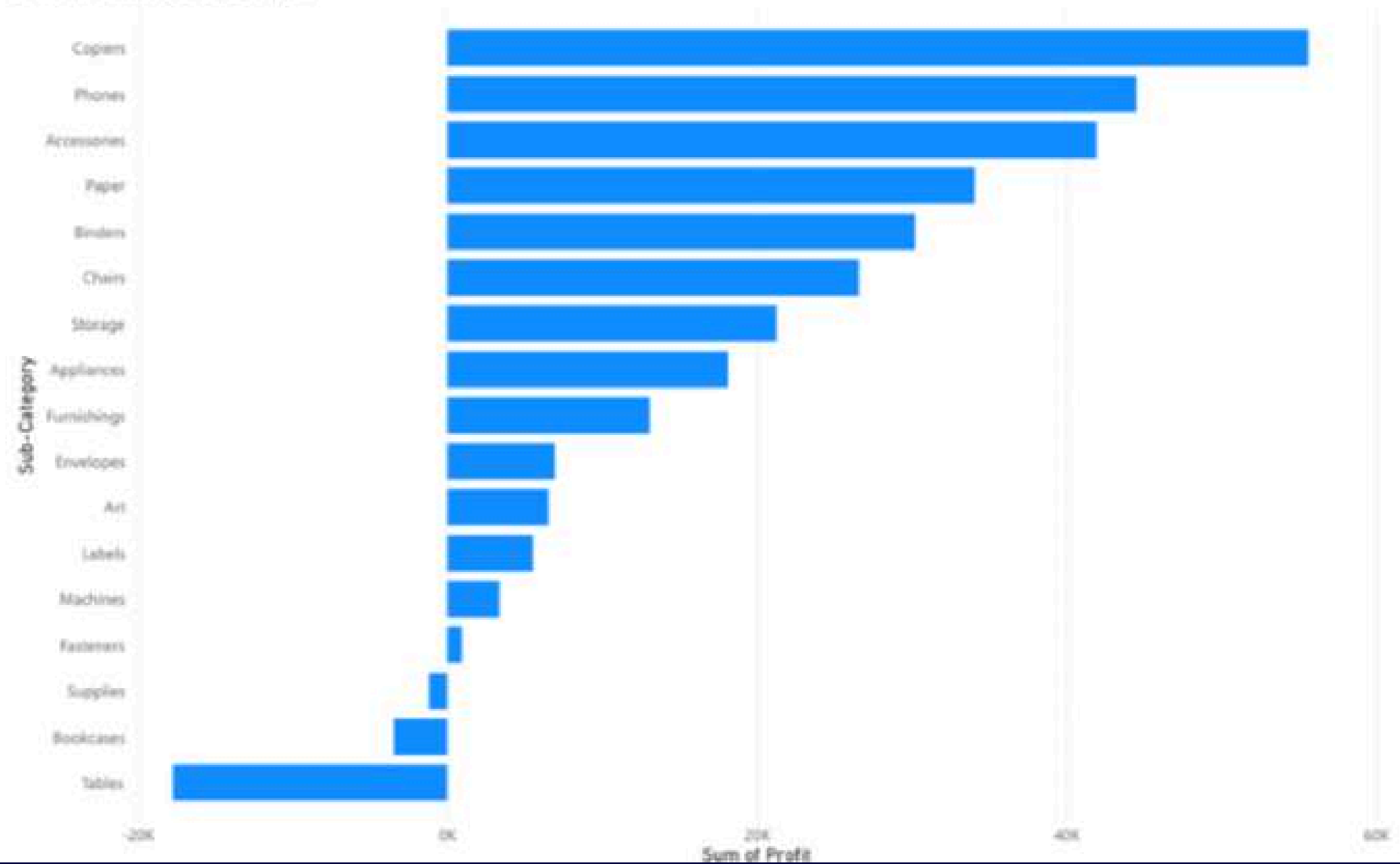
Sum of Profit by Category

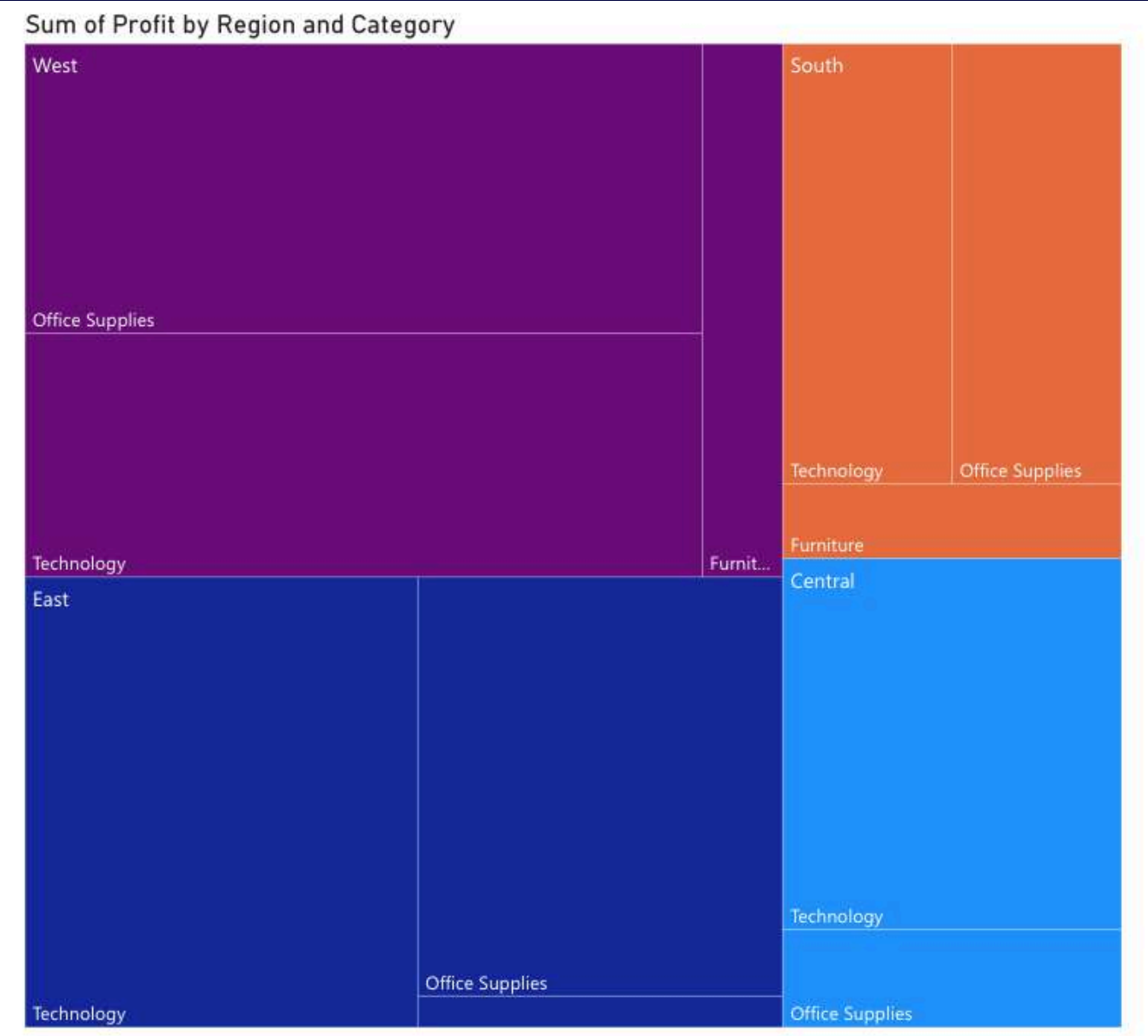
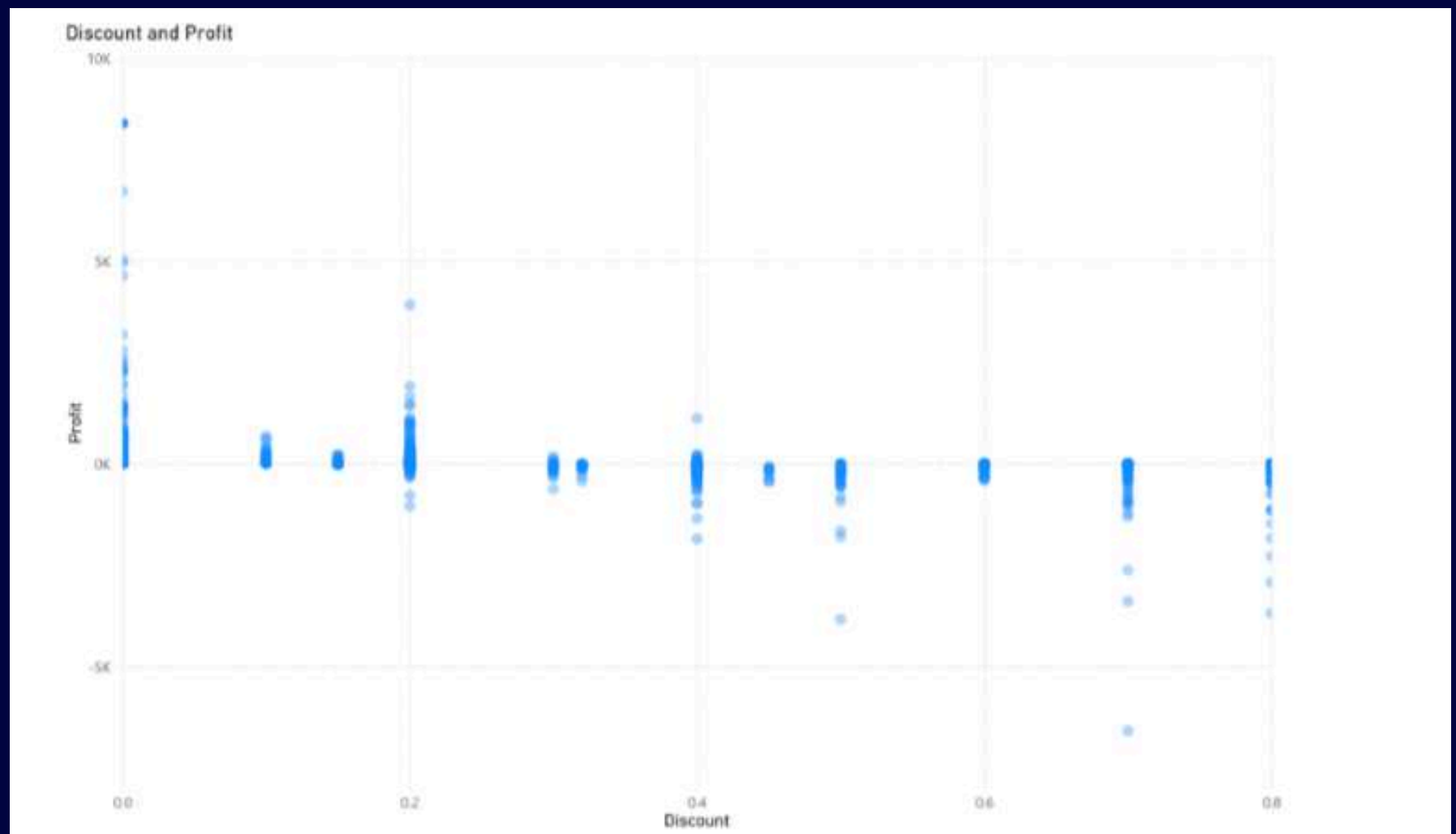


Sum of Profit and Sum of Sales by Category



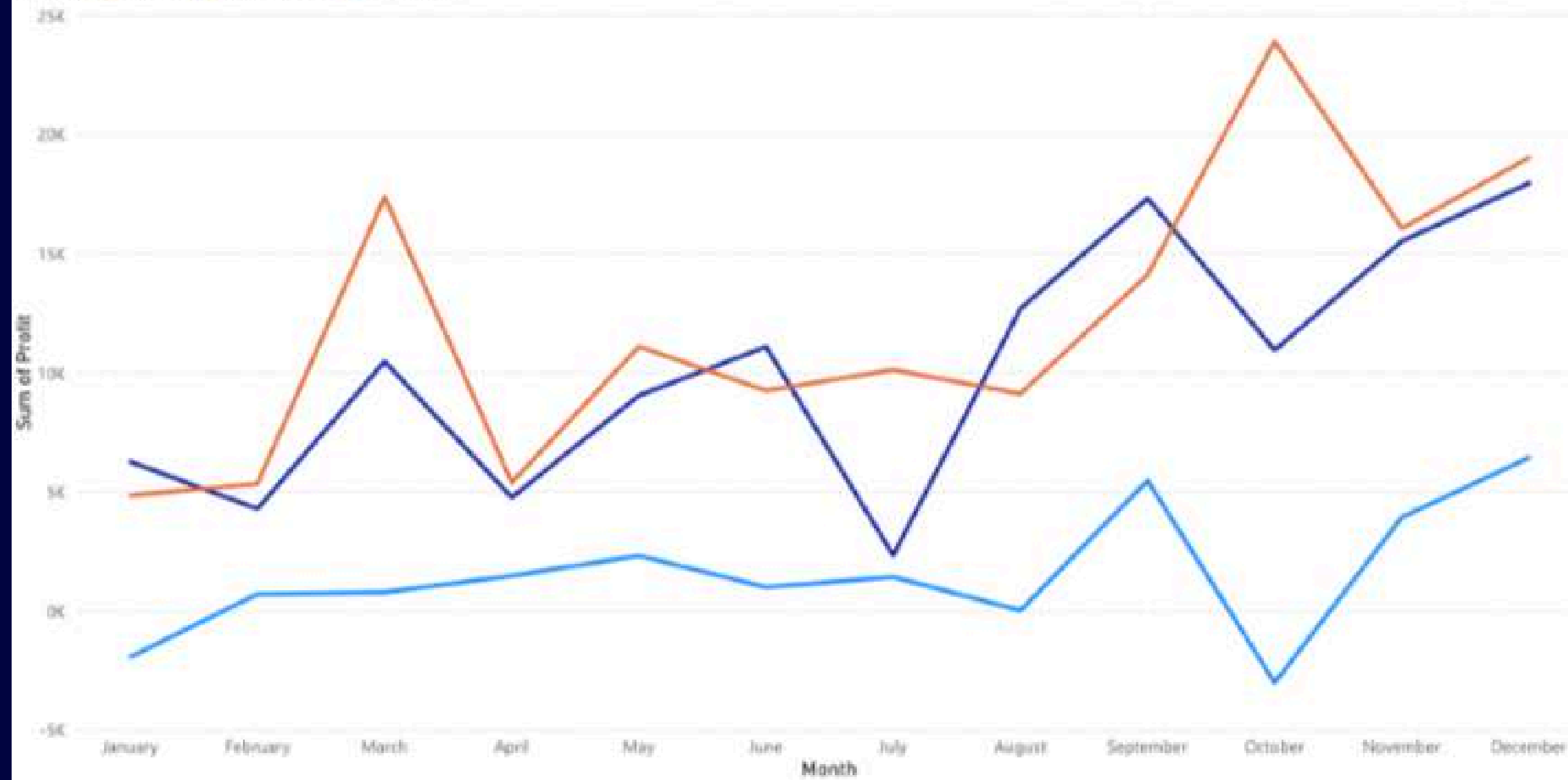
Sum of Profit by Sub-Category





Sum of Profit by Month and Category

Category ● Furniture ● Office Supplies ● Technology



PROPOSED SOLUTIONS

1. Discount Threshold for Furniture

- Furniture category often loses money because of big discounts
- We should set a maximum discount limit for furniture
- This limit should be based on past profit data
- Helps protect profit and reduce financial losses

2. Promote High-Margin Products

- Not all products bring the same profit, even if they sell well
- We should focus marketing on products with high-profit margins
- This can be found by analyzing past sales data
- Promoting profitable products will increase total profit

PROPOSED SOLUTIONS

3. Regional Pricing Strategy

- Some regions always perform worse than others
- One reason is that we use the same discounts everywhere
- Solution: use different discounts for different regions
- Lower discounts in weak areas, keep strong pricing in good areas
- This can improve profits in low-performing locations

4. Real-Time Monitoring with Power BI

- Right now, there's no regular way to check if discounts are working
- We should use Power BI dashboards to monitor performance every month
- Add AI tools to spot patterns and predict future profits
- This helps us make smarter, faster decisions

Conclusion

Our Power BI analysis of the Superstore dataset revealed that high discounts, especially in Furniture, reduce profits despite strong sales. Smarter discount strategies can boost profits, highlighting how data tools help uncover hidden issues and guide better decisions.



REFERENCES:

- [1] Alsmadi, I. (2021). Using Microsoft Power BI in analyzing and visualizing academic performance data. *International Journal of Advanced Computer Science and Applications*, 12(3), 575-580
- [2] V. Vadlamudi, "Superstore Dataset," Kaggle, 2019. [Online]. Available: <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>.
- [3] Microsoft, "Use AI Insights in Power BI Desktop - Power Query," Microsoft Learn, 2023. [Online]. Available: <https://learn.microsoft.com/en-us/power-query/ai-insights-overview>
- [4] GeeksforGeeks, "Create a Correlation Matrix using Python," Jan. 2, 2025. [Online]. Available: <https://www.geeksforgeeks.org/create-a-correlation-matrix-using-python/>
- [5] Kent State University Libraries, "SPSS Tutorials: Chi-Square Test of Independence," Kent State University, Apr. 2025. [Online]. Available: <https://libguides.library.kent.edu/spss/chisquare>