



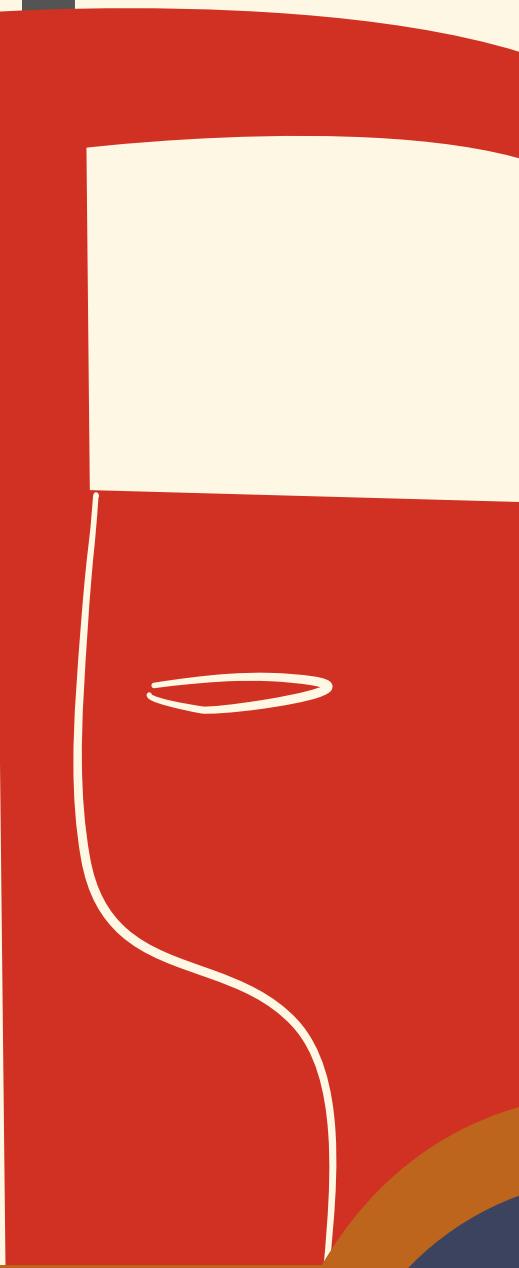
CAPSTONE

AWS Saas Analysis

Pijar HM

Topic Outline

- 1 Problem Statement
- 2 EDA
- 3 Pre-processing
- 4 Analysis
- 5 Result & Recommendation



1. Problem Statement

There's lot of product that can be sell in specific segment or industry

But product is not always really sell well in mean of time based on period of time of trend

Because of that we should identify and analyze what kind of product that will give us huge profit beside it's quantity



Goal & Objectives

Goal

What kind of product that make lot of profit and when?

Objectives

1. Identification of product
2. Identification of range of time
3. Identification of sales, profit, and quantity based on product



Data Overview

Dataset is 9994 data divided into 3 metadata (id, date), 4 parameter and 6 category

6 category

10999 data

DATASET OF AWS SAAS

3
metadata

4 parameter



2. EDA



EDA

RangeIndex: 9994 entries, 0 to 9993

Data columns (total 19 columns):

#	Column	Non-Null Count	Dtype
0	Row ID	9994 non-null	int64
1	Order ID	9994 non-null	object
2	Order Date	9994 non-null	object
3	Date Key	9994 non-null	int64
4	Contact Name	9994 non-null	object
5	Country	9994 non-null	object
6	City	9994 non-null	object
7	Region	9994 non-null	object
8	Subregion	9994 non-null	object
9	Customer	9994 non-null	object
10	Customer ID	9994 non-null	int64
11	Industry	9994 non-null	object
12	Segment	9994 non-null	object
13	Product	9994 non-null	object
14	License	9994 non-null	object
15	Sales	9994 non-null	float64
16	Quantity	9994 non-null	int64
17	Discount	9994 non-null	float64
18	Profit	9994 non-null	float64

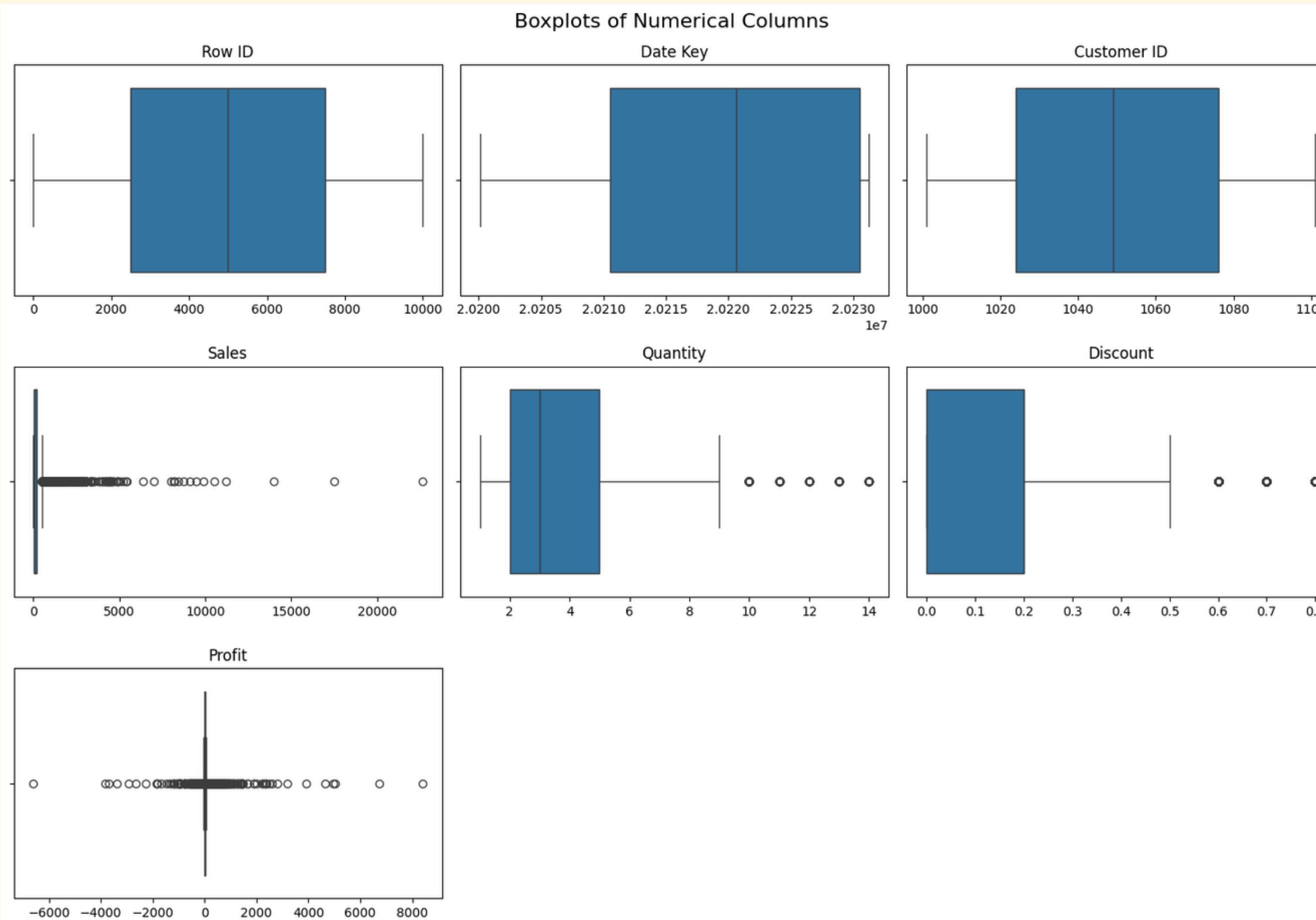
dtypes: float64(3), int64(4), object(12)

memory usage: 1.4+ MB

	Row ID	Date Key	Customer ID	Sales	Quantity	Discount	Profit
count	9994.000000	9.994000e+03	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	2.021803e+07	1049.771963	229.858001	3.789574	0.156203	28.656896
std	2885.163629	1.123898e+04	29.719388	623.245101	2.225110	0.206452	234.260108
min	1.000000	2.020010e+07	1001.000000	0.444000	1.000000	0.000000	-6599.978000
25%	2499.250000	2.021052e+07	1024.000000	17.280000	2.000000	0.000000	1.728750
50%	4997.500000	2.022063e+07	1049.000000	54.490000	3.000000	0.200000	8.666500
75%	7495.750000	2.023052e+07	1076.000000	209.940000	5.000000	0.200000	29.364000
max	9994.000000	2.023123e+07	1101.000000	22638.480000	14.000000	0.800000	8399.976000

- Indications of outliers in the numerical columns
- lower bound (25%) has a value of 17, but the maximum value is much higher at 22,638.
- Similarly, in the profit column, the lower bound is 1.7, while the maximum value is 8,399.

EDA



Column 'Row ID':
Total values: 9994
Outliers: 0

Column 'Date Key':
Total values: 9994
Outliers: 0

Column 'Customer ID':
Total values: 9994
Outliers: 0

Column 'Sales':
Total values: 9994
Outliers: 1167

Column 'Quantity':
Total values: 9994
Outliers: 170

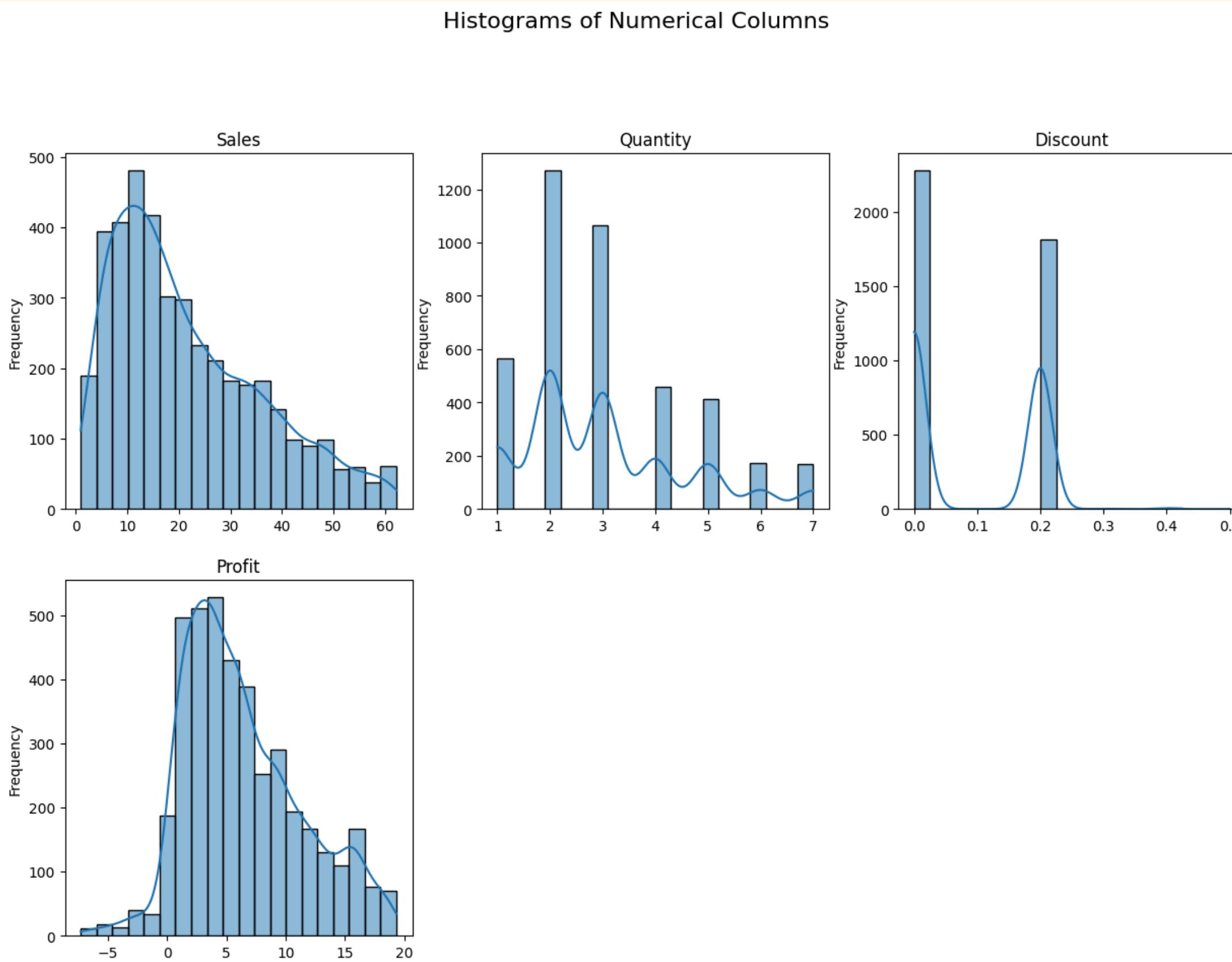
Column 'Discount':
Total values: 9994
Outliers: 856

Column 'Profit':
Total values: 9994
Outliers: 1881

2. Pre-processing



Outlier Eliminate



```
print("Df shape:", df.shape)
print("Cleaned Df shape:", df_cleaned.shape)
```

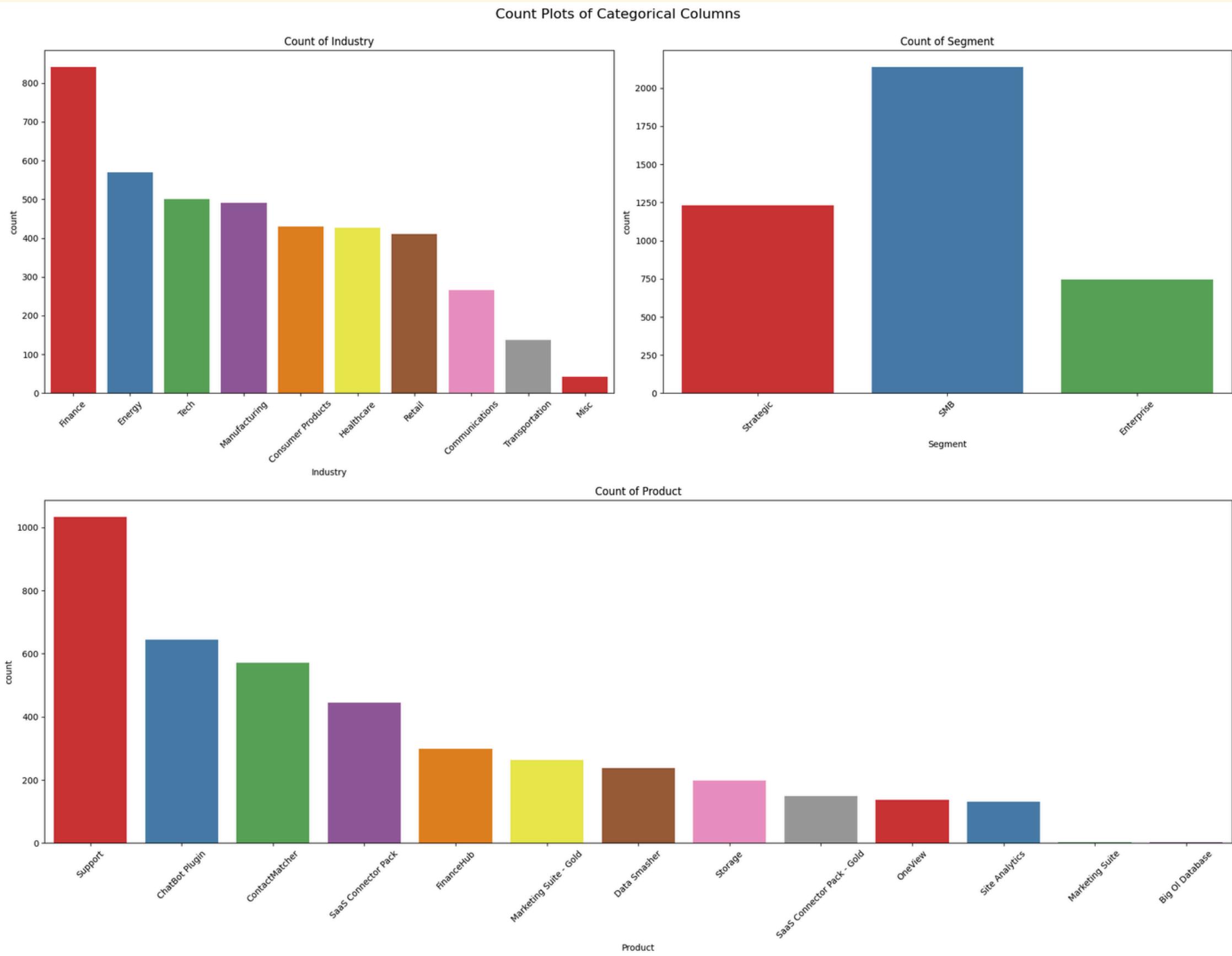
```
Df shape: (9994, 19)
Cleaned Df shape: (4115, 19)
```

```
Column 'Row ID':
    Total values: 4115
    Outliers: 0
Column 'Date Key':
    Total values: 4115
    Outliers: 0
Column 'Customer ID':
    Total values: 4115
    Outliers: 0
Column 'Sales':
    Total values: 4115
    Outliers: 0
Column 'Quantity':
    Total values: 4115
    Outliers: 0
Column 'Discount':
    Total values: 4115
    Outliers: 0
Column 'Profit':
    Total values: 4115
    Outliers: 0
```

4. Analysis

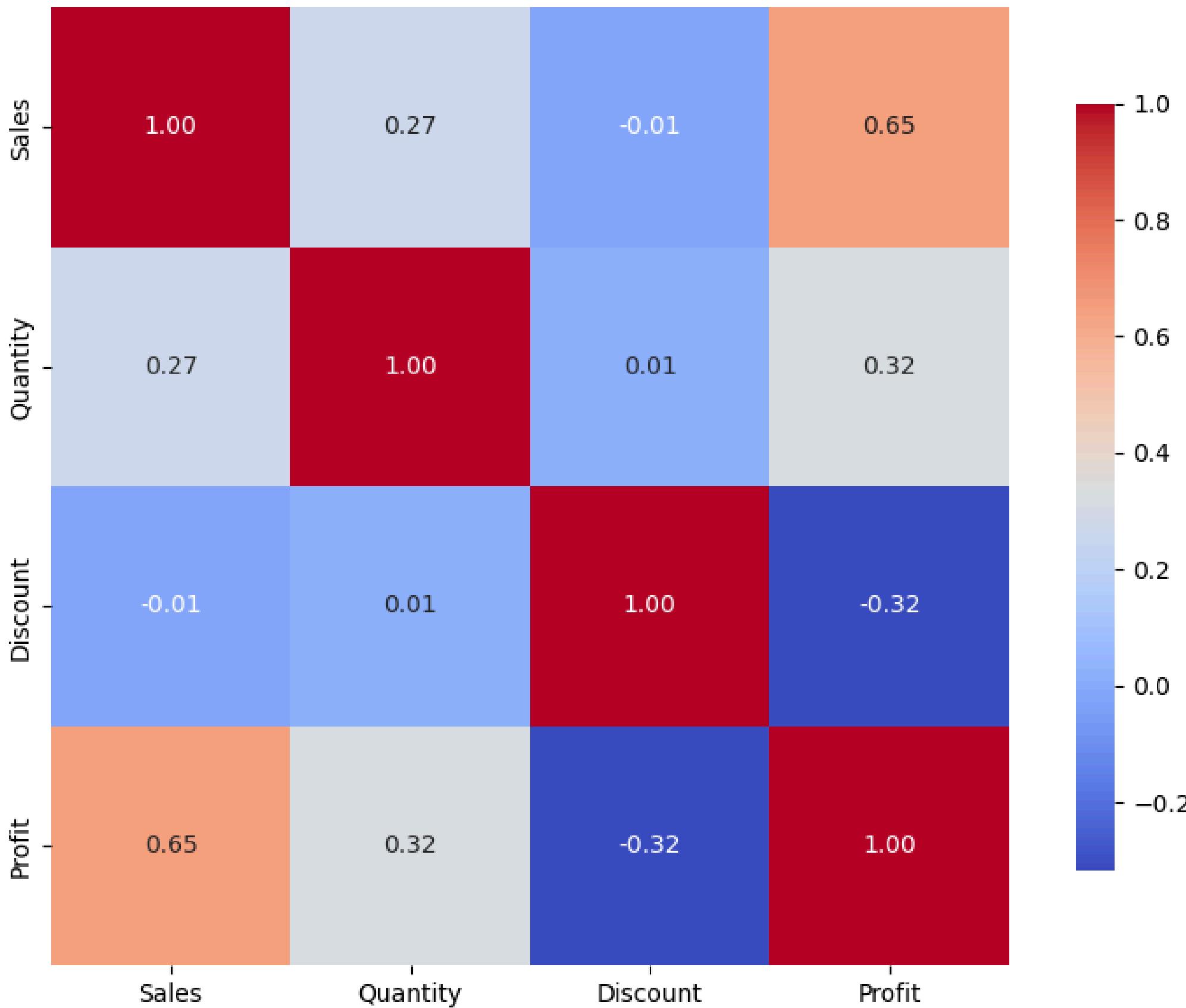


Univariate



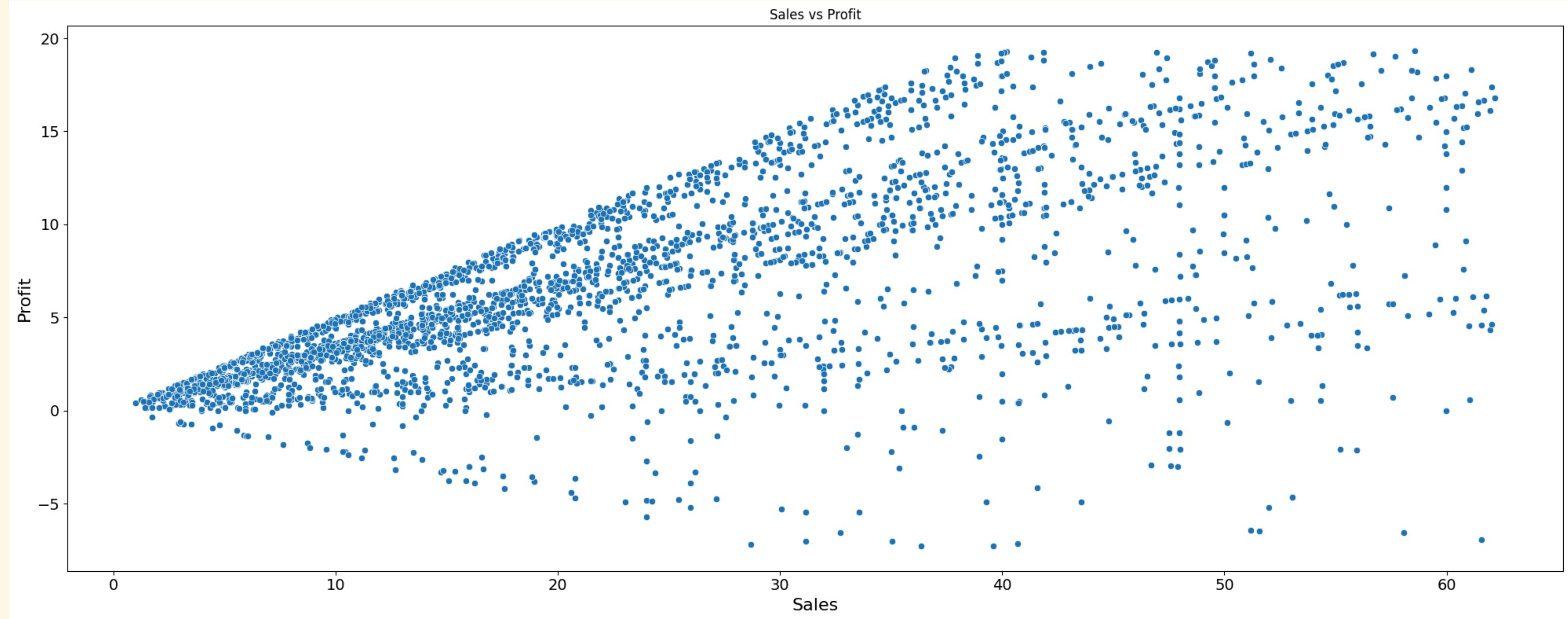
- Finance
- Energy
- Tech
- Support
- Chatbot
- ContactMatcher

Multivariate



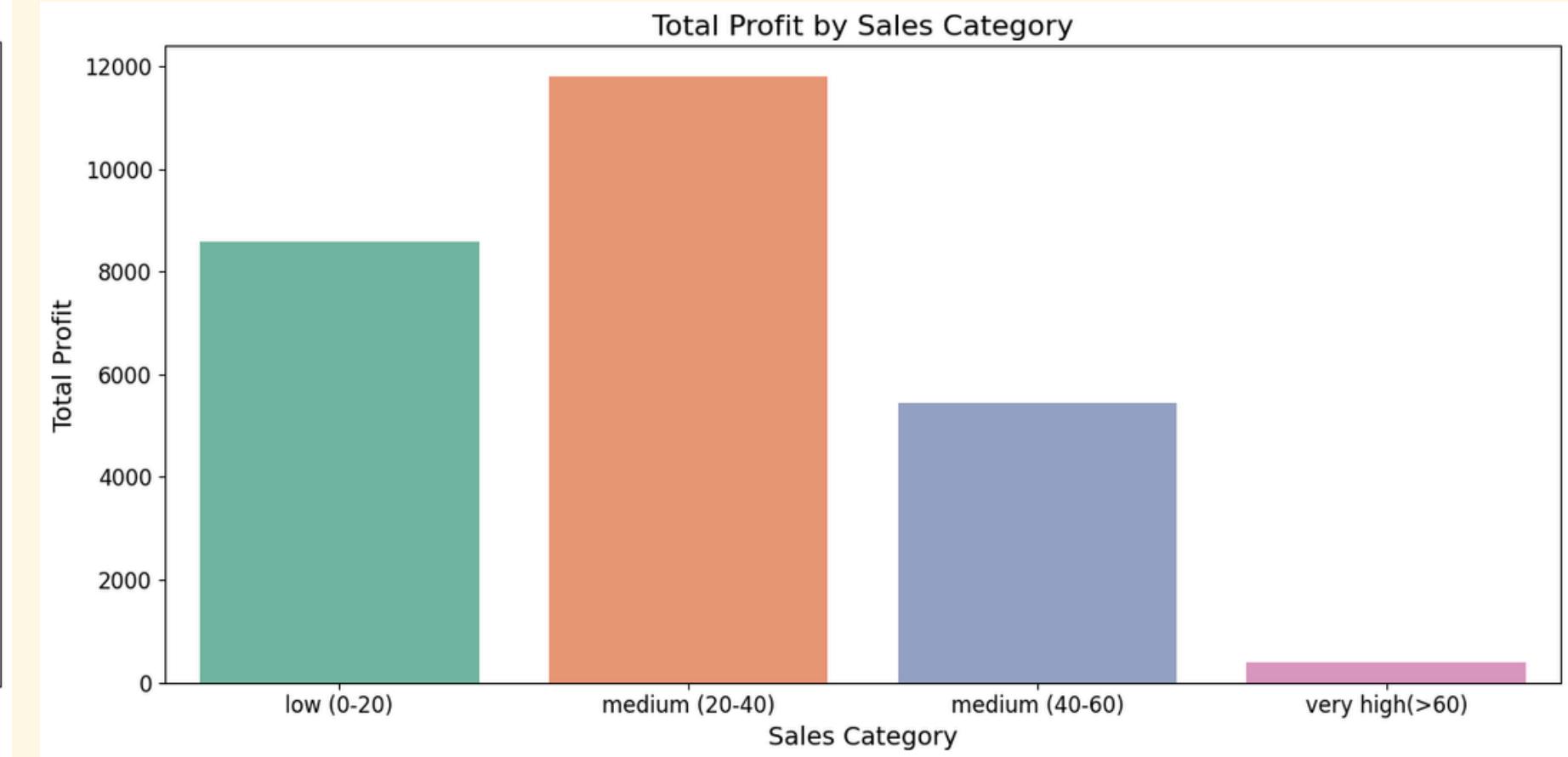
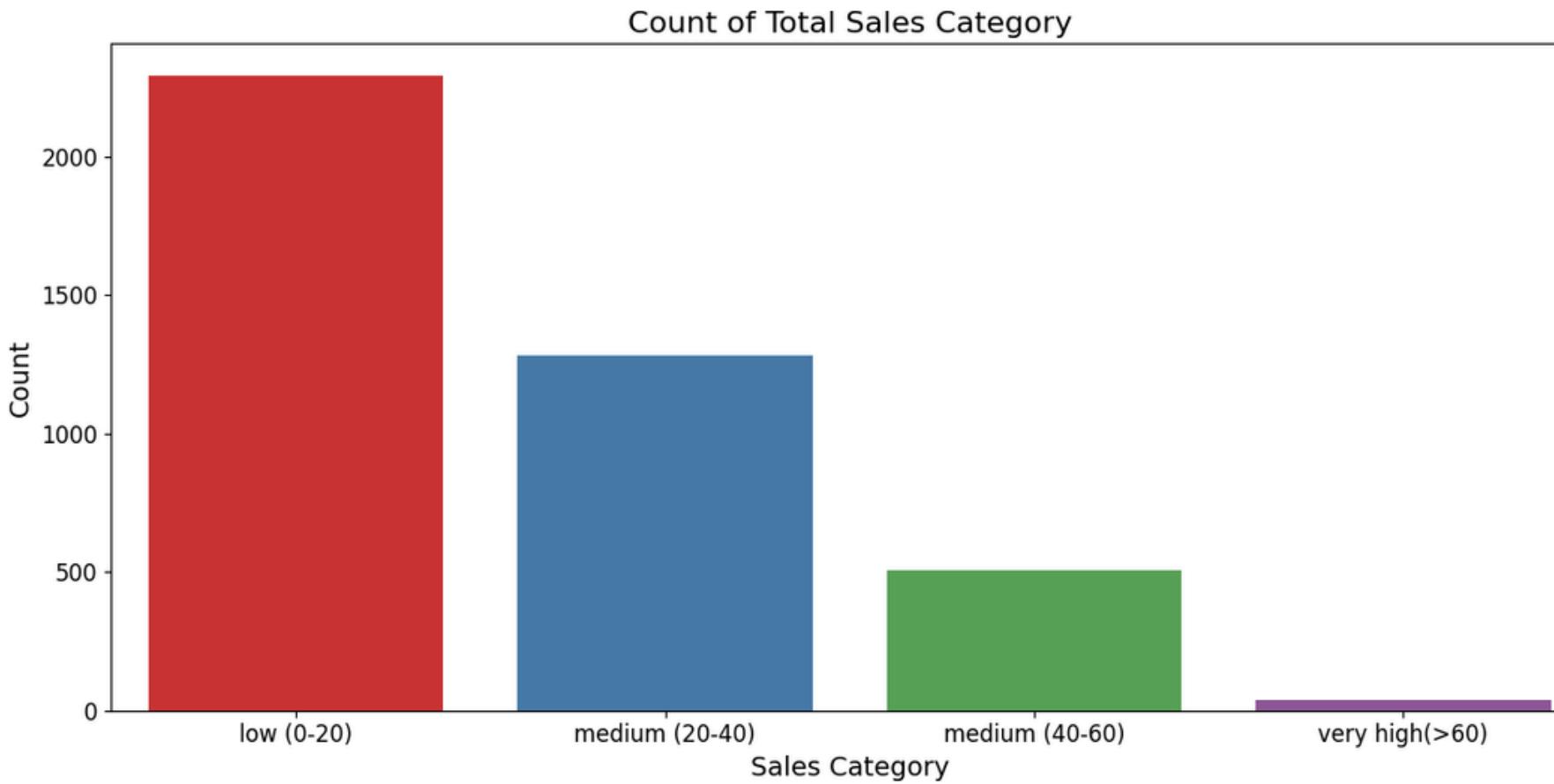
- This correlation matrix shows that `Sales` and `Profit` have a strong positive correlation (0.65), indicating that an increase in sales tends to lead to an increase in profit.
- The correlation between `Quantity` and `Profit` is also positive but weaker (0.32), suggesting that the number of items sold has a slight impact on profit.
- Meanwhile, `Discount` has a negative correlation with `Profit` (-0.32), indicating that an increase in discounts tends to decrease profit.
- There is no significant correlation between `Discount` and `Sales`, meaning that discounts do not significantly affect overall sales volume.

Multivariate



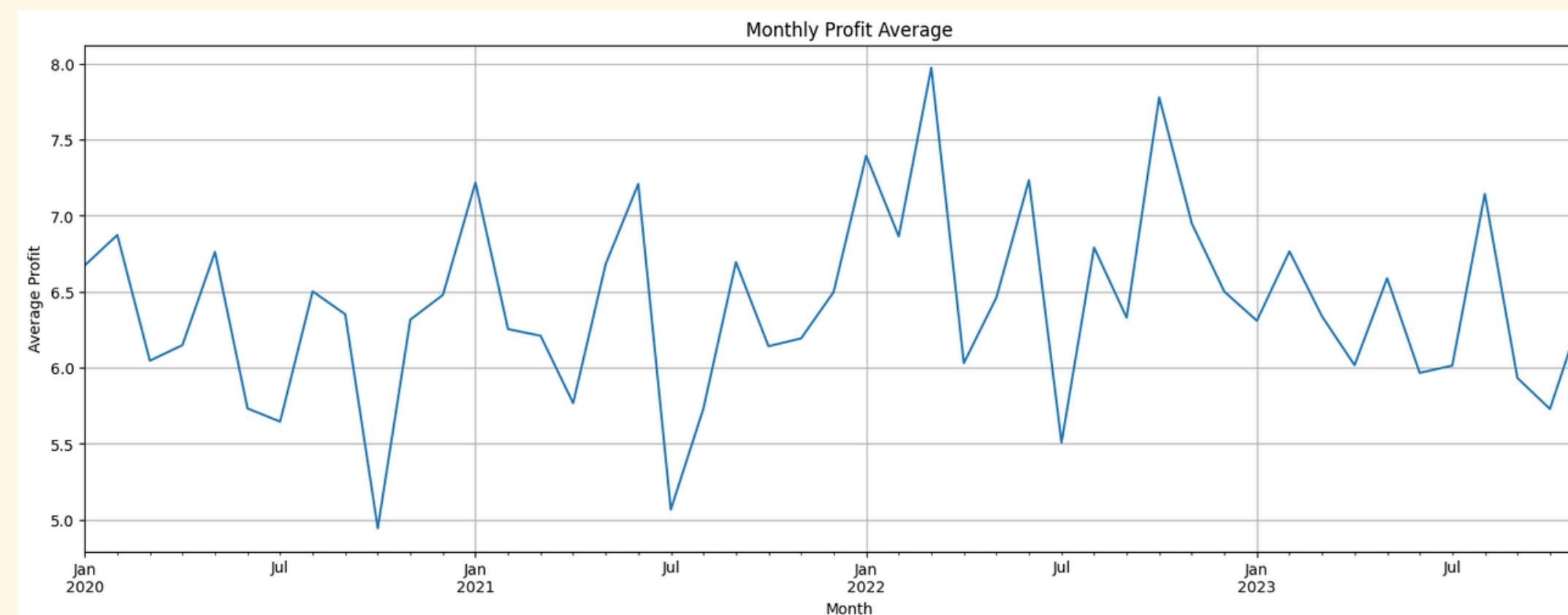
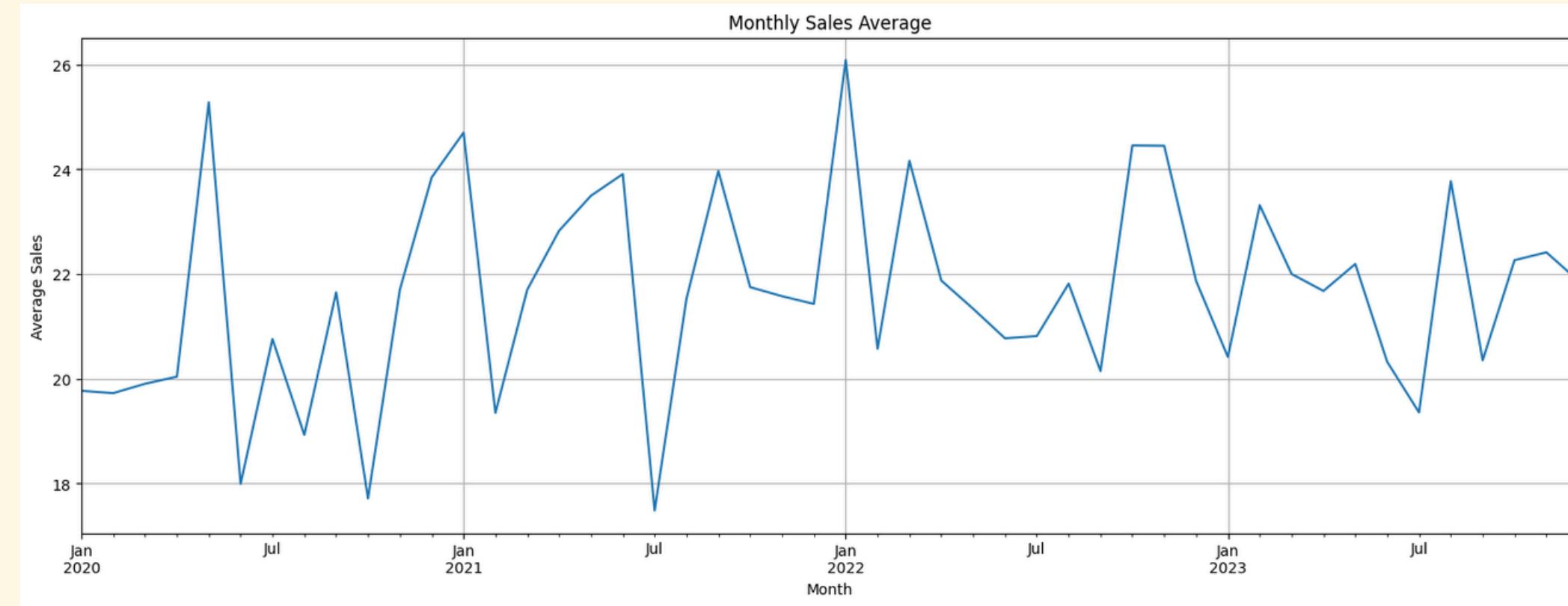
- Range of 10 – 20 has potential

Advance analysis



- The visual above shows that the sales category "medium" contributes the most to total profit. Thus, products in the "medium" range, despite having lower sales quantity compared to the "low" category, are able to generate the highest profit.

Advance analysis



Advance analysis

Based on Sales AVG and Profit AVG, the top 3 month for sales product is:

1. **March 2022:**

- - Sales Avg: 24.2
- - Profit Avg: 8.0
- - Sales Rate Change: 17.4%
- - Profit Rate Change: 16.2%

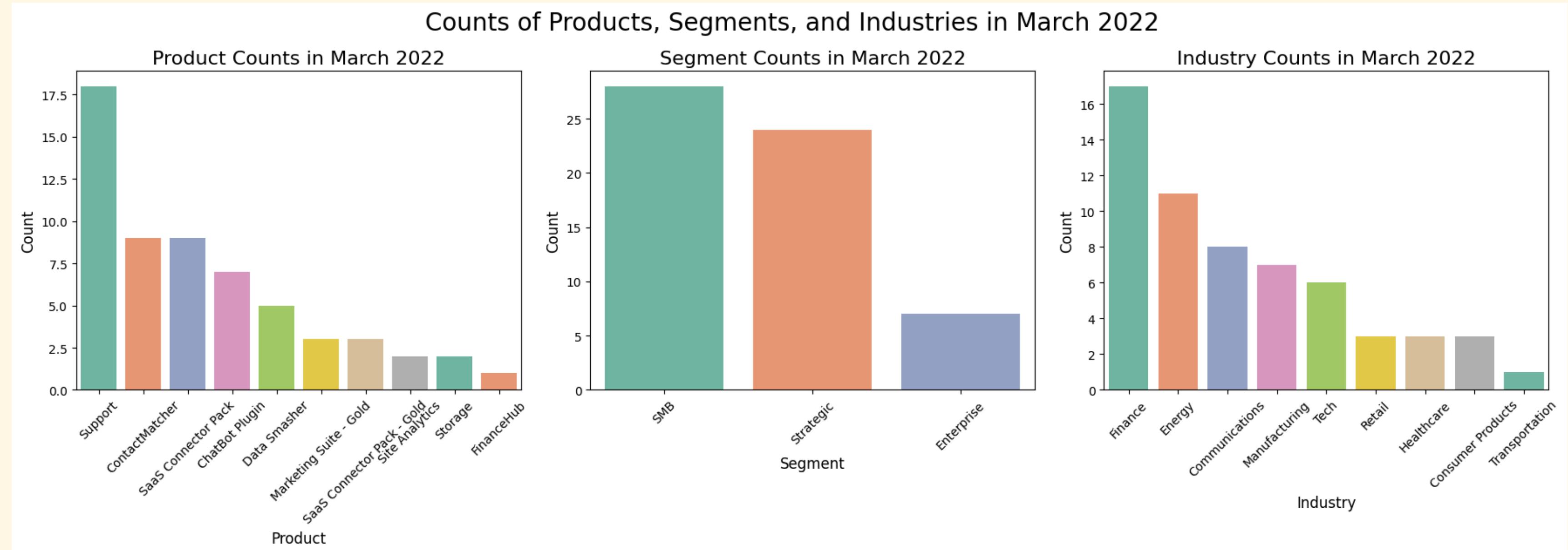
2. **October 2022:**

- - Sales Avg: 24.5
- - Profit Avg: 7.8
- - Sales Rate Change: 21.4%
- - Profit Rate Change: 22.9%

3. **January 2022:**

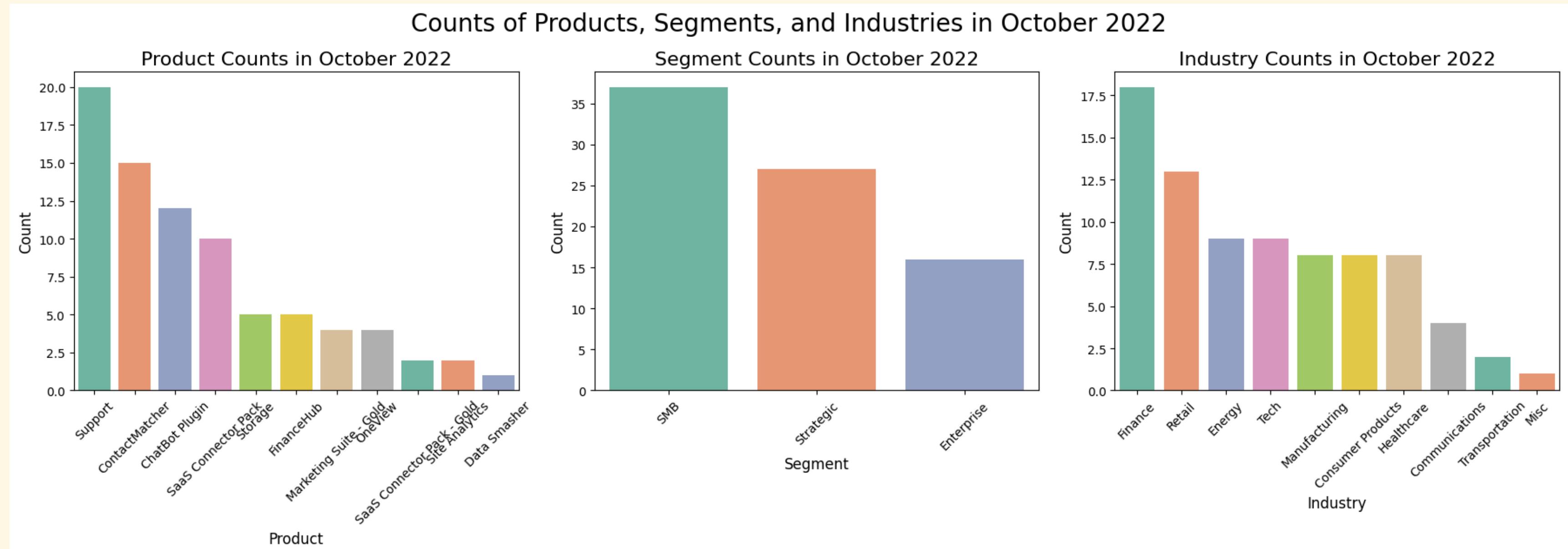
- - Sales Avg: 26.1
- - Profit Avg: 7.4
- - Sales Rate Change: 21.7%
- - Profit Rate Change: 13.8%

Advance analysis



- In March 2022, "Software" led, with "Consulting" and "Data Services" following, indicating growth from January.

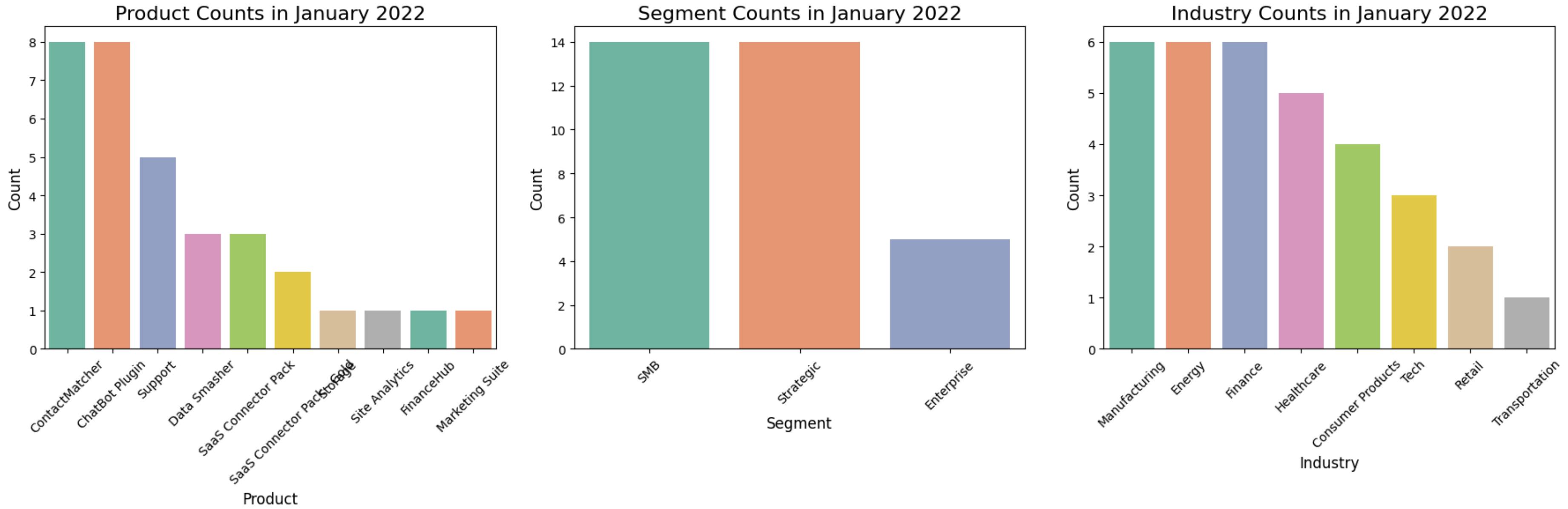
Advance analysis



- By October 2022, "Software" remained dominant, showing consistent popularity for these three products.

Advance analysis

Counts of Products, Segments, and Industries in January 2022



-In January 2022, "Consulting" had the highest count, followed by "Software" and "Data Services."

5. Business Recommendation



Recommendation

Segment Counts:

- Across all months, "SMB" and "Strategic" segments were higher than "Enterprise," with "SMB" consistently the largest. This indicates a primary focus on small-to-medium businesses and strategic clients.

Industry Counts:

- "Finance" and "Energy" led consistently, indicating strong demand, with "Healthcare" and "Manufacturing" also stable but lower in count. "Transportation" had consistently low numbers.

Changes Over Time:

- From January to March, there was an increase in counts for some categories like "Software" and "Finance."
- From March to October, counts remained stable, suggesting steady demand across these segments and industries.

Month-Specific Recommendations

- **January**: Focus on selling Software, Consulting, and Data Services, targeting SMB, Strategic segments, and Finance and Energy industries.
- **March**: Leverage high demand for Software and Consulting, and explore new products or promotions for smaller industries.
- **October**: Continue focusing on dominant products like Software, with possible additional promotions for supporting products or services in established industries.

Thank you for listening!

Don't hesitate to ask any questions!

