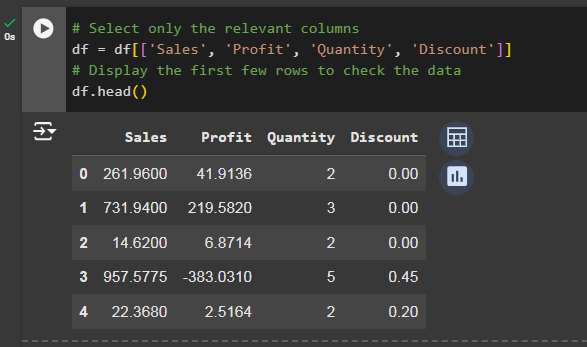
**Exploratory Data Analysis Findings Report**

# 1. Introduction

EDA is an approach of looking at what the data can tell us apart from the proper modeling or hypothesis testing endeavor (Sahoo et al, 2019). This paper gives a summary of the exploratory data analysis conducted on the sales data of a superstore that is purely fictional. These primary attributes are Sales, Profit, Quantity, Discount which are essential to monitor revenue, cost and organizational performance. Since there is missing data and outliers within the dataset, data cleaning and transforming are conducted in the analysis. This particular sort of analysis – EDA – is designed to learn more about the data and make recommendations based on patterns as they pertain to the superstore.

# 2. Data Loading and Initial Filtering

When the Superstore\_dataset.csv was loaded and only Sales, Profit, Quantity, and Discount were selected to analyze the core financial operational issues (as shown in figure 1).



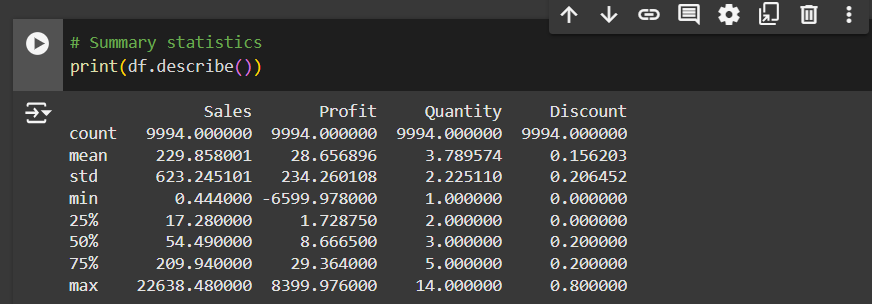
*Figure 1. Data loading and filtering*

Data need to be analyzed so as to produce good result. This approach enables a streamlined EDA while retaining the most impactful variables on cost and revenue. The capability to clean, rearrange and verify the data on python environment with the help of Pandas and NumPy libraries created a cleaned DataFrame. The resulting dataset was used for this analysis after removing the unnecessary columns.

# 3. Trend Analysis

## Summary statistics

The data shows an average sale of $229.86, with high variability due to outliers. Profit averages $28.66 but often remains low, influenced by discount rates averaging 16%. Quantity averages 3.79 items per transaction, while discounts are generally moderate, with occasional high discounts used for promotions.

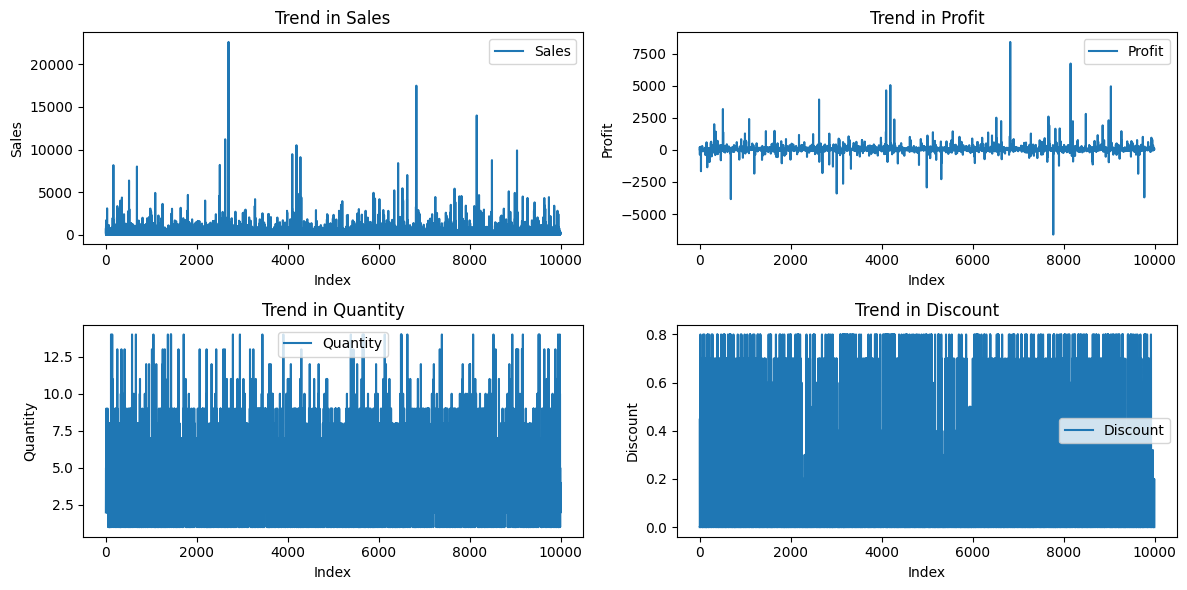


*Figure 2. Summary statistics*

# Findings on the Trend analysis

## 3.1 Sales

The Sales data exhibits an upward trend over time, punctuated by periodic fluctuations. This trend suggests an increase in customer demand or successful sales strategies, potentially driven by promotions or seasonal trends. However, occasional declines suggest periods of decreased sales activity, which may coincide with off-peak times or shifts in demand (see figure 3).



*Figure 3. Analyzed overall trend*

The sales patterns indicate a likely seasonal component or cyclical demand, where sales fluctuate due to factors such as promotions, holidays, or other time-based events. Identifying the periods with spikes could inform inventory and resource planning, allowing for better preparation during high-demand periods.

## 3.2 Profit

The Profit data shows significant variability and displays cyclical behavior. Although there is no consistent upward or downward trend, the profit spikes and declines suggest a strong dependence on pricing strategies and discount policies. Large fluctuations in profits likely reflect the impact of discounting on high-value items or varying demand for different products (Refer to figure 3). Since profits do not follow a consistent upward trend despite increasing sales, the business could evaluate the effectiveness of its discounting strategy. Profit trends may indicate areas where discounts erode profitability or where targeted pricing adjustments could boost revenue.

## 3.3 Quantity

The Quantity of items sold shows moderate fluctuations over time, with no strongly identifiable seasonal patterns. Variations in quantity might reflect customer purchasing behavior or inventory levels. However, quantity does not seem to exhibit as much variability as sales or profit, suggesting that it may be influenced by more stable factors, such as baseline demand for staple items. Hence, a steady trend in quantity suggests stable customer demand (shown in figure 3). While fluctuations do occur, they are less extreme, indicating that the number of items sold does not strongly impact profit margins directly but may do so in combination with pricing.

## 3.4 Discount

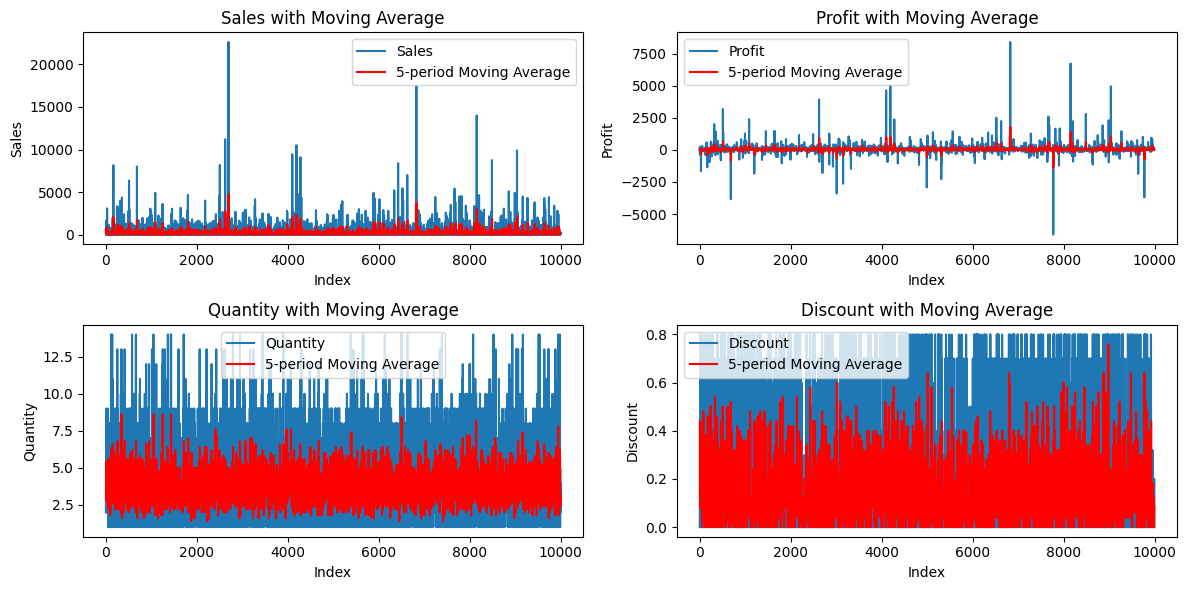
The Discount variable remains stable overall but exhibits occasional spikes (See figure3). Such peaks are indicative of promotional periods or clearance events. Notably, these discount spikes often coincide with decreased profitability, implying that aggressive discounting could negatively impact margins. Thus, Discounts correlate with a reduction in profit margins, suggesting the need for a more strategic approach to discounting. By analyzing the optimal discount levels that boost sales without severely impacting profit, the store could implement more profitable promotional strategies.

# 4. Handling Missing Values

A thorough inspection of missing values was conducted to assess the extent of data incompleteness. Missing data can introduce bias, skewing the results if left unaddressed. Columns with missing values were treated by imputing median values, preserving the central tendencies without significantly affecting the overall dataset distribution. By filling missing values with medians, we maintained data integrity without introducing artificial skewness. This approach enables a more reliable analysis, especially when calculating trends, averages, or conducting further statistical analysis.

# 5. Moving Average Smoothing

To better observe trends and smooth out short-term fluctuations, a 5-period moving average was applied to each column (see figure 4). This smoothing technique highlighted the underlying patterns in each metric without the noise of irregular peaks and valleys.



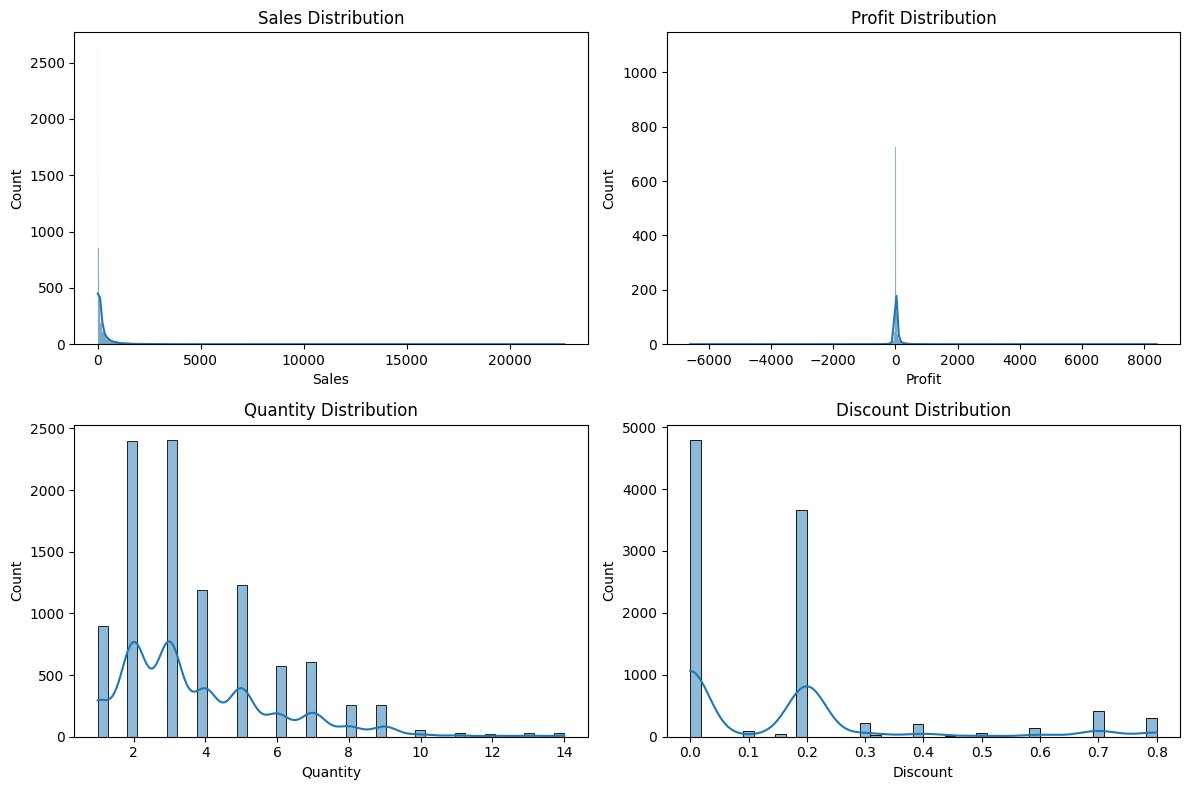
*Figure 4. Overall trend with Moving average for smoothing*

**Observations**

* **Sales Moving Average -** The moving average reinforced the general upward trend in sales. While fluctuations are present, the trend line smooths these variations, revealing a stable growth trajectory.
* **Profit Moving Average -** Profit also exhibited a cyclical pattern, with the moving average revealing periods of profitability that coincide with high sales volume, but dips in periods of aggressive discounting.
* **Quantity Moving Average** - The quantity moving average line was relatively flat, indicating minimal seasonality. Small fluctuations suggest steady demand with occasional shifts, potentially linked to short-term promotions or inventory availability.
* **Discount Moving Average** - Discount levels are generally low but show sharp increases during promotional periods. This spike-and-drop pattern suggests that discounting strategies may be targeted during specific periods, possibly to clear stock or attract seasonal buyers.

# 6. Outlier Detection and Analysis

Histograms were generated to visualize the distribution of each variable and identify outliers. Outliers in sales and profit are particularly notable, as extreme values can distort averages and skew interpretations (refer to figure 5,6).

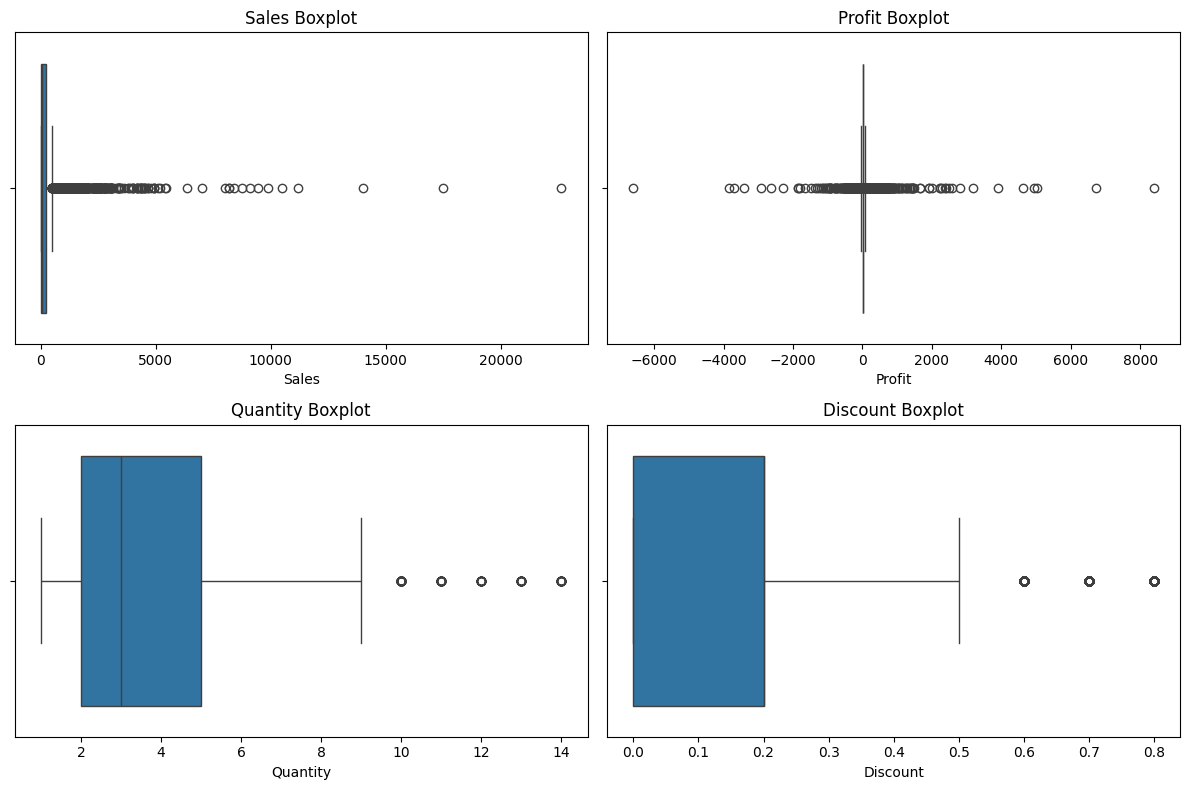


*Figure 5. Histogram to visualize distribution and spot outliers*

* **Sales Outliers** - Outliers were observed at the higher end, indicating large transactions or high-value item sales. These outliers could represent significant revenue drivers but may also inflate averages.
* **Profit Outliers -** Profit outliers, both positive and negative, were observed. Negative outliers suggest potential losses on certain transactions, possibly from high discounts, while positive outliers indicate high-profit transactions.

**Key Findings**

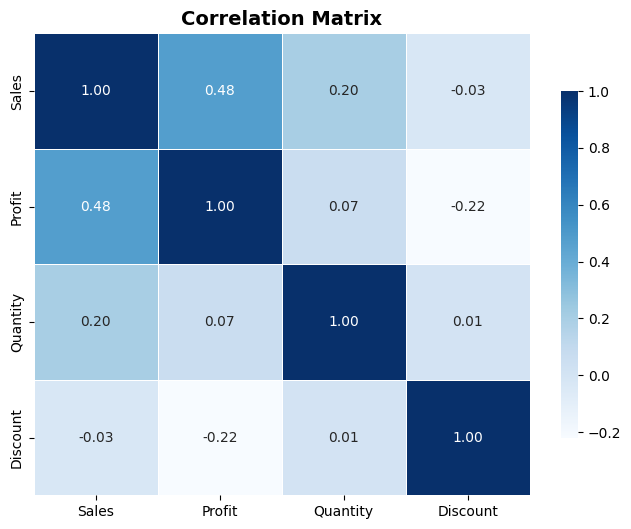
Outliers, especially those in sales and profit, provide insights into peak performances and potentially loss-generating transactions. Understanding the context of these transactions could help in refining pricing or discounting strategies to mitigate loss and enhance revenue consistency (See figure 6).



*Figure 6. Boxplot for Outlier detection*

# 7. Correlation Analysis

A correlation matrix was created to examine relationships between Sales, Profit, Quantity, and Discount (see figure 7). The correlation analysis underscores that while sales drive profit, high discounting undermines profit margins. This finding emphasizes the need for discount management to preserve profitability.



*Figure 7. Correlation matrix*

* **Sales and Profit** - A positive correlation exists, indicating that higher sales are generally associated with increased profit. However, the correlation is not exceptionally high, suggesting other factors may influence profitability.
* **Discount and Profit -** A negative correlation between discounts and profit supports the observation that discounts can diminish profitability. Strategic discounting might mitigate this effect by balancing customer incentives with profitability.
* **Quantity and Sales** - Moderate correlation suggests that larger quantities sold do positively impact sales, but not proportionally. Bulk sales may benefit overall sales but may not translate directly to profit if discounted heavily.

# 8. Impact of Data Quality Issues

Addressing data quality issues was critical to this analysis. Missing values, if left untreated, could result in incomplete analysis or biased results. Similarly, outliers can distort averages, leading to inaccurate conclusions. The data cleaning process, including imputation and outlier awareness, enabled reliable analysis and minimized the risk of misleading findings. Thus, data quality improvements enhance decision-making by providing accurate insights. Addressing missing values and understanding outliers allows the business to focus on actionable metrics and drive informed decisions.

# 9. Conclusion

Notably, through handling data quality issues and analyzing key metrics, businesses can implement informed strategies that support growth and profitability. Thus, the EDA provides the following valuable insights into the superstore's operations: -

* **Sales Trends** - An upward trend indicates potential growth opportunities. Understanding demand cycles can guide inventory management and marketing.
* **Profit Insights** - While sales drive revenue, discounting needs closer management to preserve profitability.
* **Quantity Patterns** - Stability in quantity sold highlights steady demand, supporting predictable inventory planning.
* **Discount Effects** - Correlations reveal that excessive discounting can undermine profit, indicating the need for targeted promotional strategies.

# References

Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using

Python. *International Journal of Innovative Technology and Exploring Engineering*, *8*(12), 4727-4735. <https://www.researchgate.net/profile/Dr-Subhendu-Pani/publication/337146539_IJITEE/links/5dc70b124585151435fb427e/IJITEE.pdf>