**Product Demand Prediction Report**

1. **Data Exploration**

## Dataset Overview

The dataset comprises the following key attributes:

1. Store ID
2. Total Price
3. Base Price
4. Units Sold

## Data Statistics

**Total Price:** The dataset contains a range of total prices, with unique values and varying frequencies.

**Base Price:** Similar to total price, base price exhibits diverse values and frequencies.

**Units Sold:** The units sold also demonstrate a wide range of values.

**Missing Values:** No missing values were observed in the dataset, indicating a well-structured dataset.

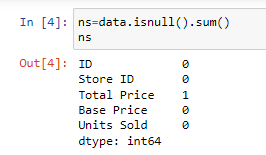
**Outliers**: Outliers were identified through box plots, showcasing extreme values in both total price and units sold.

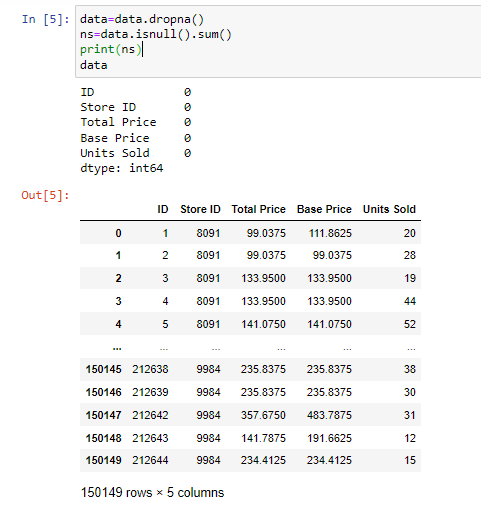
1. **Exploratory Data Analysis**

# 2.1 Initial Exploration

**•** Loaded the dataset and conducted an initial exploration using describe() and info().

• Ensured there are no missing values in the dataset.





# 2.2 Discount Analysis

A new column, 'Discount,' has been added to the dataset to quantify the percentage discount offered on each sale. The calculation is performed using the following formula:

Discount=((Base Price−Total Price)/Base Price)×100

**Insights:**

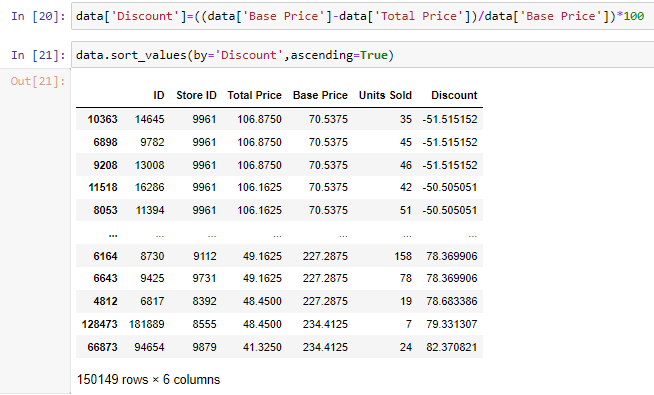
**Understanding Discount Distribution:**

The 'Discount' column provides insights into the extent of price reductions for each sale.

It allows us to analyse the distribution of discounts across the dataset.

**Sorting by Discount Percentage:**

Sorting the dataset based on the 'Discount' column in ascending order reveals products with the lowest percentage discounts. This can be indicative of items typically sold closer to their base prices.



**Identifying Significant Discounts:**

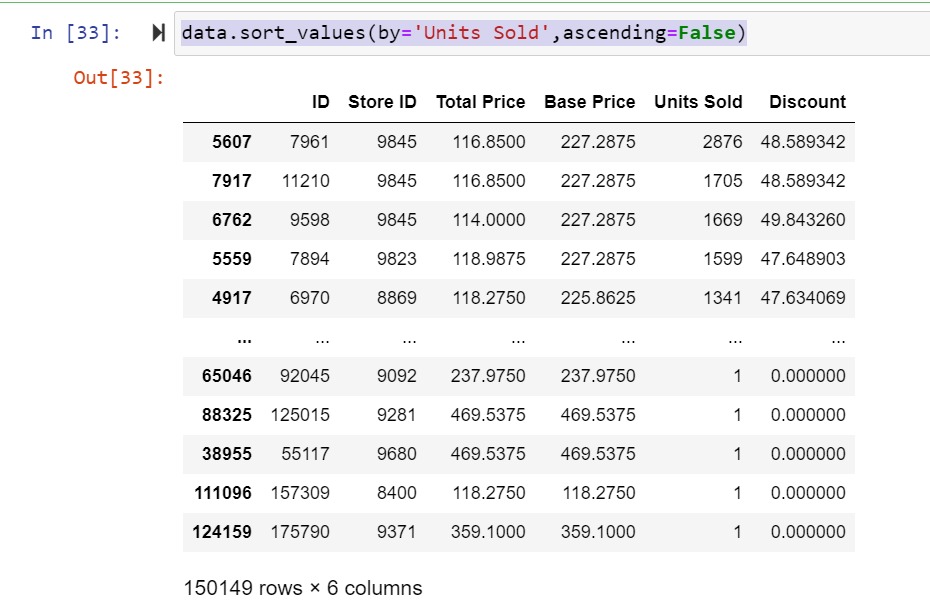
Higher percentage discounts highlight products with more substantial price reductions.

**Purpose:**

The addition of the 'Discount' column aids in the analysis of pricing strategies, customer behaviour, and discounting policies. It offers a quantitative measure of the discount provided for each sale compared to the original or base price. This information can be valuable for decision-making related to product pricing and marketing strategies.

# 2.3 Identifying Top-Selling Products:

By sorting the dataset in descending order based on 'Units Sold,' you can easily identify and list the top-selling products. These are the products that have the highest sales volume.

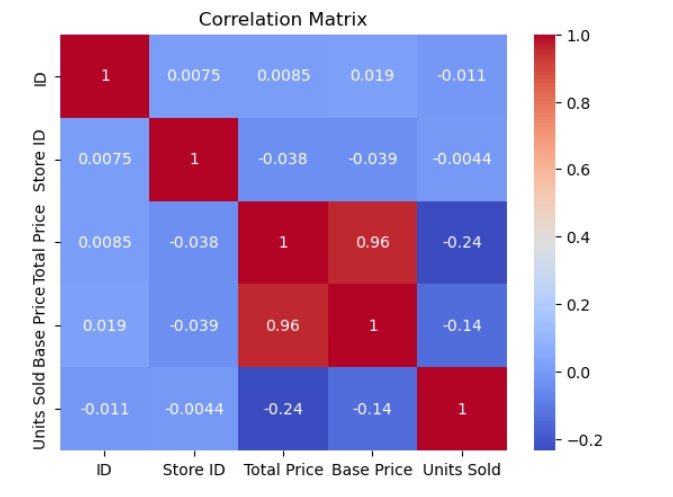


1. **Visualisation of Insights**

**3.1. Correlation Analysis**

**Heatmap Visualization**

To gain insights into the relationships between different attributes within the dataset, a correlation matrix was computed and visualized using a heatmap.



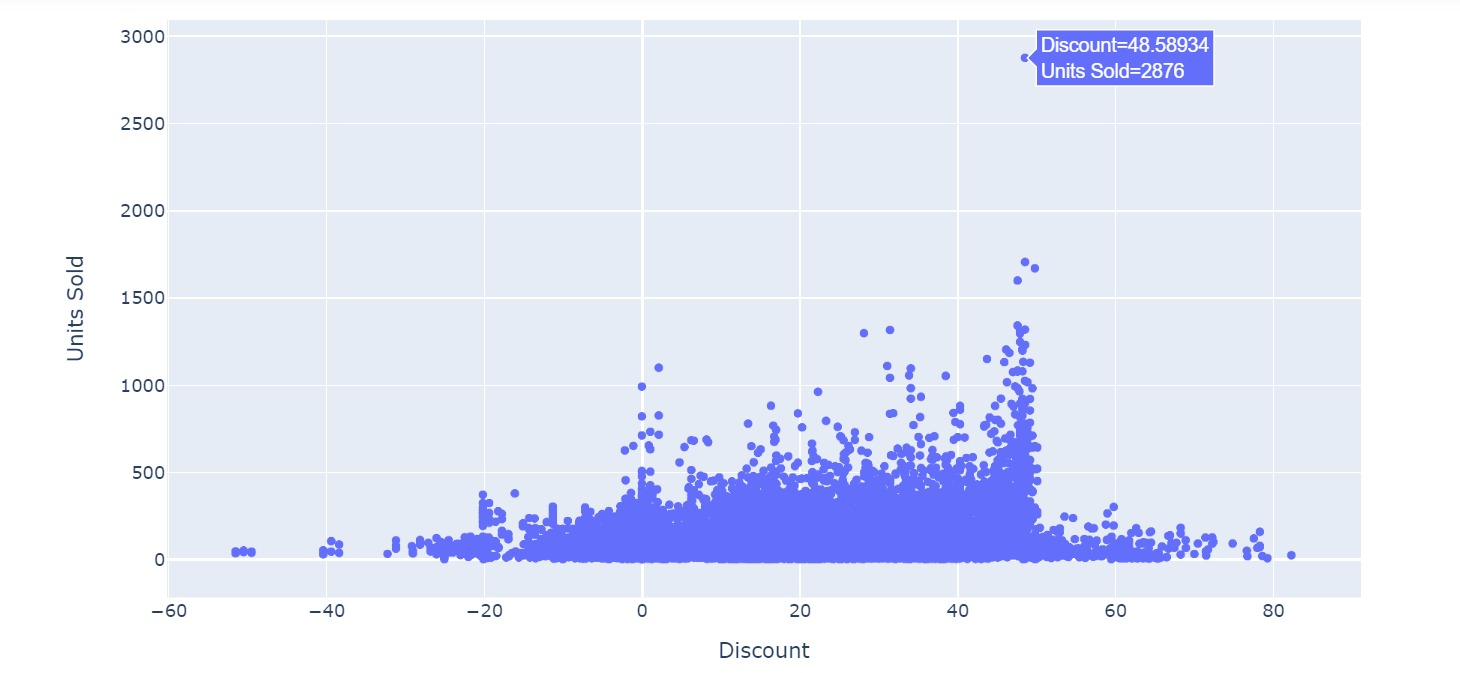
Total Price and Base Price:

The most notable observation is the exceptionally strong positive correlation (0.96) between 'Total Price' and 'Base Price.'

**Observation:** This strong relationship between total price and base price, there is a possibility of missing data value, which when thought about could be the discount values for each and every product.

**3.2 Scatter Plot**

The scatter plot between Units Sold and Discount reveals a distinct point where the highest number of units, 2876, was sold. This occurred at a discount of approximately 48.59%. This point stands out as a significant outlier in terms of both discount percentage and units sold.



**Patterns Observed**:

* Discount Range of 47-49%:

Another noteworthy observation is that units sold remain consistently high, above 1500, for products with discounts ranging from 47% to 49%. This broader range suggests a potential sweet spot where offering discounts can significantly boost sales.

* Discount Range of 43-49%:

Within the narrower range of 43-49% discounts, the scatter plot indicates a slightly lower but still substantial range of units sold, between 1000 and 1500. This suggests that even at slightly lower discounts, there is a positive impact on sales volume.

* Discount Range of 47-49%:

Products with discounts in the range of 47-49% exhibit a consistent trend of high unit sales. The highest sale point falls within this range, indicating that customers may be particularly responsive to discounts in this bracket.



**Conclusion:**

Through our exploratory data analysis, a noteworthy pattern has emerged in the dataset, indicating a compelling relationship between the discount offered and the corresponding units sold. Specifically, when the discount falls within 43 - 49% range, we consistently observe that the units sold surpass the 1000 threshold. This observation strongly suggests that there is a positive correlation between offering discounts within this range and a heightened demand for the product.

In conclusion, the dataset strongly suggests that maintaining discounts within the identified range is associated with a substantial increase in product demand, positioning it as a key determinant in influencing consumer purchasing decisions. This insight can inform strategic decision-making, allowing businesses to tailor pricing strategies to optimize both profitability and market competitiveness.