Supplement Sales Analysis

Aziz Prabowo

About the project

Background

In the competitive health and wellness retail industry, delivering effective and targeted promotions is essential to driving sales and maintaining customer engagement. WOMart, a nationwide supplement retail chain with over 350 stores across 100+ cities, faces challenges in optimizing its promotional strategies. Inconsistent promotional performance across store types and regions, along with underutilized historical data, makes it difficult to assess which campaigns truly drive sales. This raises a business question, which promotional strategies actually drive sales.

Goal

The goal of this project is to analyze the impact of discounts on sales performance across different store types and regions.

Objective

Identify the most influential factors driving promotional effectiveness and quantify their impact on sales performance based on store type and regions.

Dataset: Supplement Sales
Notebook: Jupyter Notebook

About the data

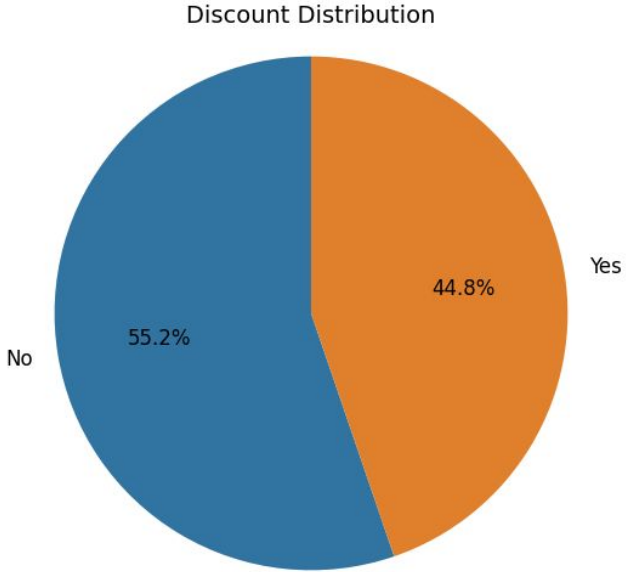
There are two numeric, one string, and one datetime attributes in the dataset, and the rest are categorical. Also, there seems to be no missing values in the dataset. So, we can proceed to the analysis process.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 188340 entries, 0 to 188339
Data columns (total 10 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   id             188340 non-null  object
 1   store_id       188340 non-null  category
 2   store_type     188340 non-null  category
 3   location_type  188340 non-null  category
 4   region_code    188340 non-null  category
 5   date           188340 non-null  datetime64[ns]
 6   holiday        188340 non-null  category
 7   discount       188340 non-null  category
 8   order          188340 non-null  int64
 9   sales          188340 non-null  float64
dtypes: category(6), datetime64[ns](1), float64(1), int64(1), object(1)
memory usage: 7.0+ MB
```

- 6 categorical attributes
- 2 numerical attributes
- 1 datetime attributes
- 1 string

Discounts drive 44.8% of sales transactions

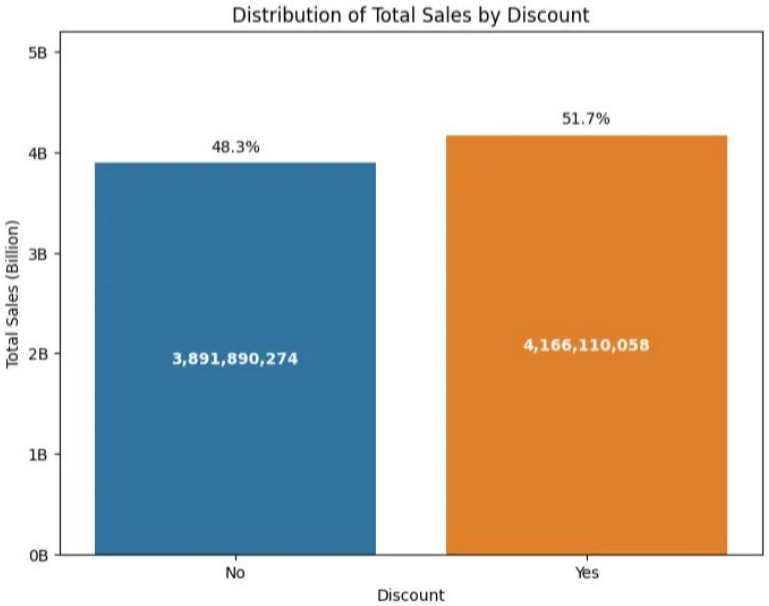
84,289 transactions of the total **188,340** were made with discounts. This represents nearly half of all sales, prompting further analysis into how discounts influence overall sales performance.

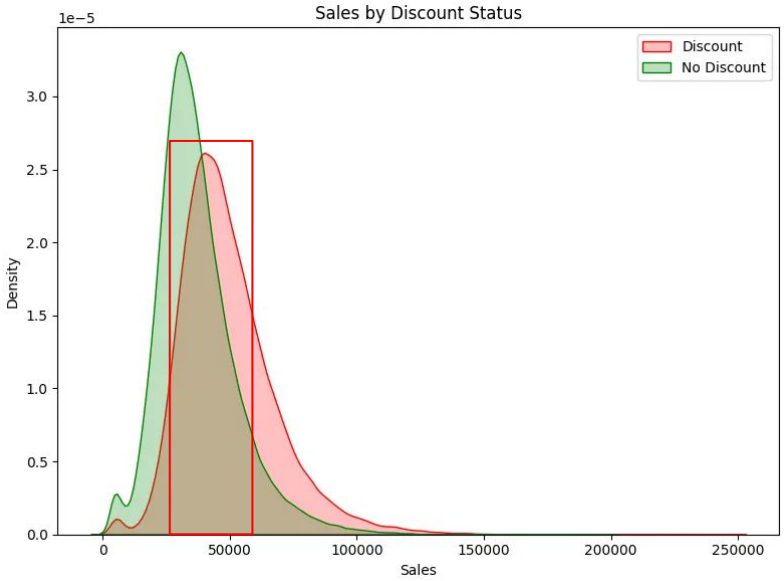


Follow-up: *Do discounts significantly increase sales volume overall?*

Discounted Transactions Drive 51.7% of Total Sales Revenue, Accounting for 4.2 Billion

We found that the discounted transaction contribute to **51.7%** of the total sales revenue (**4.2B**). Overall, transactions involving discounts show higher sales revenue than those without. Additionally, higher-value transactions (over **40,000**) are more likely to involve discounts.

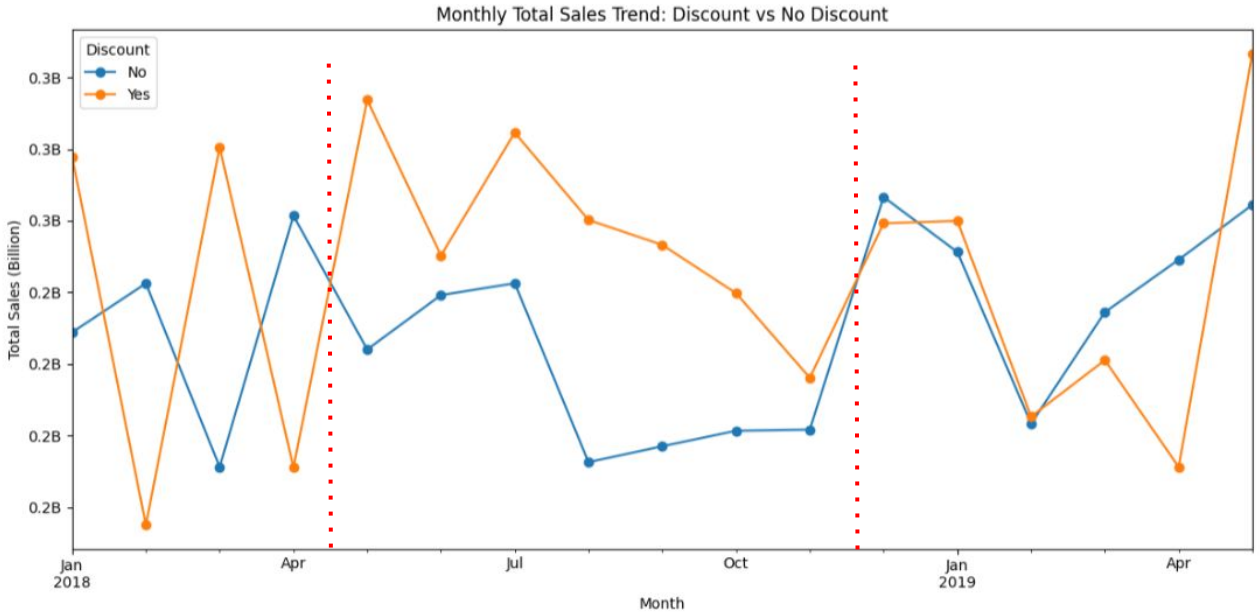




Follow-up: *What about the monthly sales trend?*

2.2 Billion in Discounted Deals Surpass Non-Discounted Deals from Q2 to Q4

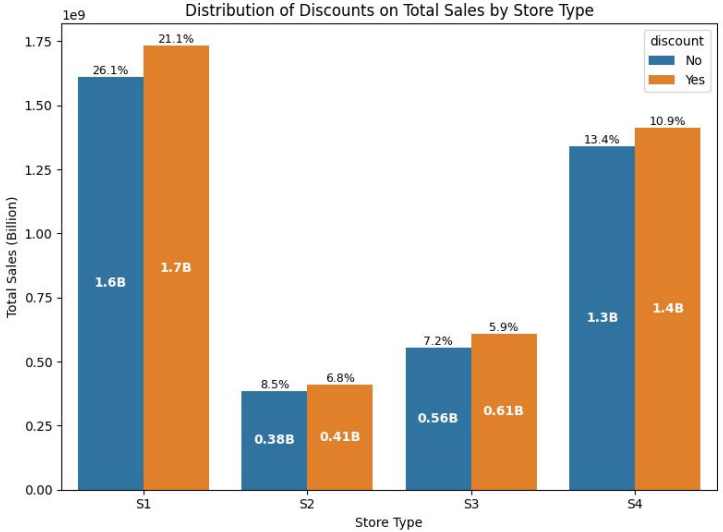
Sales trends fluctuate throughout the year, with **46K discounted transactions** beginning to dominate from the second quarter onward. This shift indicates that discounts become a more prominent driver of sales later in the year, likely due to seasonal promotions or strategic sales periods.

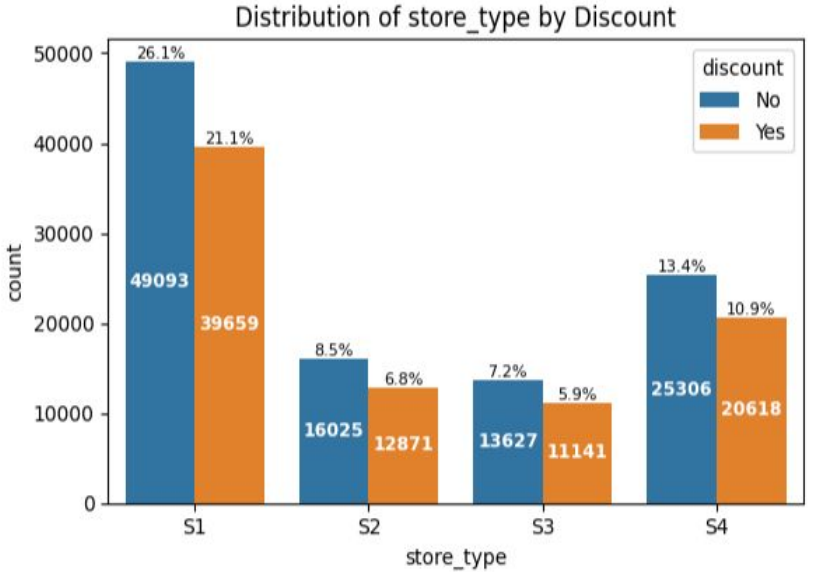


Follow-up: *How does the impact of discounts differ across store type?*

S1 & S2 Store Types Drive Majority of Sales, Generating 6 Billion in Total Revenue

S1 and **S2 store types** contribute significantly to overall sales, with **discounted transactions generating 3.1 billion in revenue**, while **non-discounted transactions account for 2.9 billion**. This indicates that while discounts play a key role, non-discounted sales remain strong, highlighting the diverse strategies driving sales across these store types.

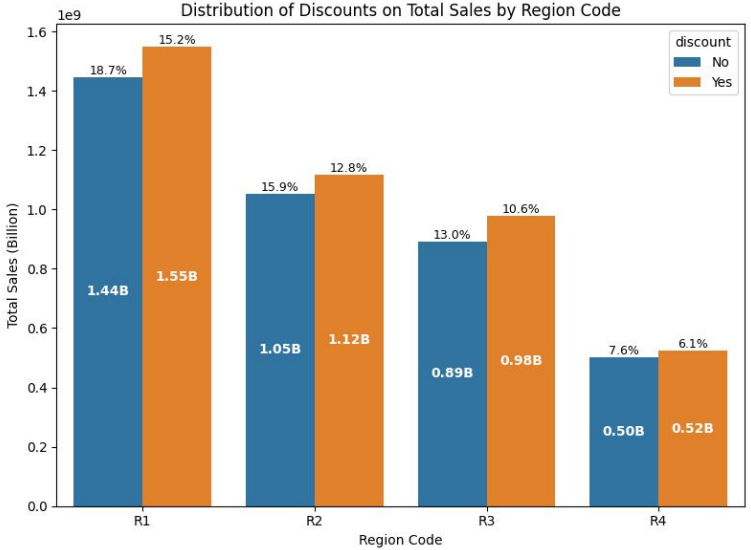


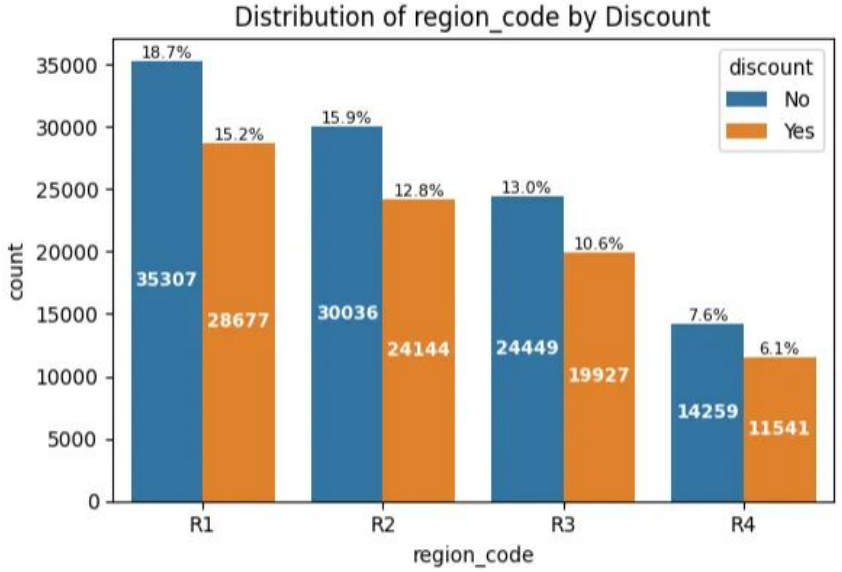


Follow-up: *What about the impact of discounts across region code?*

R1 Region Leads in Sales with 2.9 Billion in Revenue, R4 Trails with 1.2 Billion

The **R1 region dominates sales performance**, generating a total of **2.9 billion in revenue**, while the **R4 region lags behind with just 1.2 billion**. This disparity suggests a significant difference in sales strategy, customer engagement, or regional demand between the two regions, warranting further analysis to understand the factors driving these outcomes.



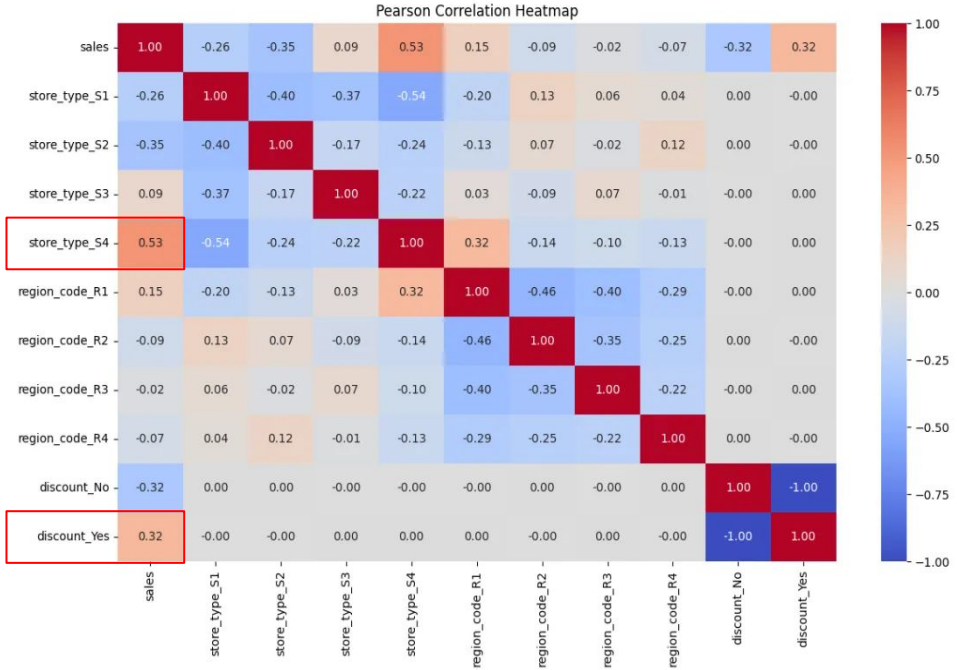


Follow-up: *So, what's the most correlated variables with Discount in statistical way?*

Store Type S4 and Discounted transactions are key drivers of sales performance.

Among these variables, two key drivers of sales stand out:

- **Store Type S4** show a much higher likelihood of sales performance (**correlation: +0.53**) compared to the other store types, particularly **Store Type 2**, which are less associated with higher sales (**correlation: –0.35**).
- **Discounted transactions** are also boost sales (**correlation: +0.32**), this aligns with earlier insights that discounted transactions account for **51.7%** of total revenue.



Summary

- **44.8% of transactions** involved discounts (84,289 out of 188,340), driving **51.7% of total revenue**.
- **Store Type S4** exhibits the strongest positive correlation with sales, significantly outperforming other store types.
- **Discounted transactions** are positively correlated with sales, contributing more revenue per transaction, while **non-discounted transactions**, although more frequent, generate less revenue.
- From **mid-Q2 to year-end**, discounted transactions dominate, revealing a **seasonal sales pattern** where discounts become a more prominent driver of revenue. **2.2 billion in discounted deals** surpassed **non-discounted deals** from Q2 to Q4.
- **Store Types S1 and S2** are negatively correlated with sales, suggesting underperformance in comparison to other store types.
- Across all regions, the pattern holds: discounted sales are fewer but generate significantly higher revenue, with **R1 region** leading with **2.9 billion in revenue**, and **R4** trailing at **1.2 billion**.
- **S1 & S2 store types** together account for the majority of sales revenue, generating a total of **6 billion**, with **3.1 billion** from discounted transactions and **2.9 billion** from non-discounted transactions.

Recommendations

1. **Prioritize and expand Store Type S4 operations** during high-demand seasons (Q3–Q4) to capitalize on strong sales performance.
2. **Enhance discount strategies**, particularly in **Store Type S4**, to maximize revenue from high-value purchases.
3. **Re-evaluate underperforming store types (S1 & S2)**—adjust product mix, layout, or promotional efforts to boost sales.
4. **Launch targeted seasonal campaigns** with threshold-based discounts to drive higher transaction values during mid-to-late year.
5. **Reduce or minimize discounting in Q1–early Q2** to protect profit margins when discount impact is lower.
6. **Segment promotions by store type and region**, using performance data and correlation insights for smarter targeting.
7. **Tailor pricing strategies** based on discount effectiveness and store type profitability.

Thanks