

# Story: How Data Helped Supermart Understand Its Grocery Business

## Setting the scene

Supermart is a grocery delivery app serving customers across Tamil Nadu. Every day, thousands of orders flow in: oils, masalas, beverages, grains, fruits, vegetables, and more. But leadership had a problem:

They could see total sales in their system, but they could not answer simple questions like:

- Which categories actually drive most of our revenue?
- Which cities and regions are most valuable?
- Are our discounts helping or quietly killing our profit?
- How do sales change across months and years?

All this information was **hidden inside a CSV file** with 9,994 orders and 11 columns.

That's where your analysis came in.

---

## Cleaning up the chaos

You started in Python using pandas, numpy, matplotlib, and seaborn.

Step by step, you:

- Loaded the `Supermart-Grocery-Sales-Retail-Analytics-Dataset.csv` file into a DataFrame.
- Converted the messy `Order Date` column, which mixed formats like `06-06-2017` and `3/28/2015`, into a clean datetime column using `pd.to_datetime(..., dayfirst=True).[1]`
- Created helpful time features:
  - `month_no` to see which months perform best.
  - `Month` (name) and `year` to track trends over time.
- Verified that key fields like `Sales`, `Discount`, and `Profit` were numeric and had no missing values.

Now the raw CSV had turned into an analysis-ready dataset.

---

## What the data revealed

Using groupby and visualizations, you turned thousands of rows into a few clear stories.

### 1) Which products drive sales?

Grouping by `Category` and `Sub Category`, you found that:

- A small set of categories contributed a large share of total sales, including daily staples and high-frequency grocery items.
- Top sub-categories (like key staples and popular grocery lines) dominated the revenue chart, while several niche items contributed only a small slice.

This showed Supermart that **not all products are equal**—a few categories keep the business running.

## 2) Where is the money coming from?

Looking at **City** and **Region**, your charts showed that:

- A handful of cities generated a big chunk of total sales.
- Certain regions consistently outperformed others, both in sales and profit.

This helped leadership see **which locations they should double down on**, and which ones might need a different strategy.

## 3) How do sales change over time?

Using the extracted **month\_no** and **year**:

- Monthly trends showed that some months were clearly stronger than others, indicating seasonal demand patterns.
- Yearly totals confirmed that later years contributed a large portion of overall revenue, reflecting strong business growth over time.

This turned the timeline from a guess into a **clear story of growth and seasonality**.

## 4) Are discounts really helping?

By plotting **Discount** against **Profit** and looking at the correlation between **Sales**, **Discount**, and **Profit**:

- You saw that discounting did boost sales volume in some areas.
- But higher discounts were often linked with **lower profit**, especially in specific categories.

This showed that discounts were a **double-edged sword**—good for volume, but dangerous for margins if not controlled.

---

## 💡 Turning analysis into decisions

From your EDA, you translated numbers into simple, business-friendly actions:

- **Category focus:**
  - Prioritize high-sales, high-profit categories.
  - Re-evaluate low-profit categories even if their sales look “good.”
- **City & region strategy:**
  - Treat top cities as “core markets” with more attention and better service.
  - Use mid-tier and emerging cities as growth engines where small investments can have big impact.
- **Seasonal planning:**
  - Prepare inventory and staffing for stronger months identified in the monthly trend charts.
  - Use low months for experiments or promotions.

- **Smarter discounts:**

- Avoid blanket high discounts.
  - Use smaller, controlled discounts in categories where profit is very sensitive.
-