



# FreshMart Revenue Optimization

## Introduction:

FreshMart is a fast-growing grocery retail chain based in the United States, recognized for its broad assortment of affordable, high-quality products. Operating across a diverse network of cities and customer segments, FreshMart has earned a strong presence in both urban and suburban communities. As the company prepares for its next phase of strategic expansion, leadership is shifting focus from physical growth through new store openings to maximizing Total Sales Revenue from its existing infrastructure. This transition marks a pivotal moment for FreshMart to leverage its rich transactional data and make data-driven decisions that optimize operations, pricing strategies, inventory management, and customer engagement.

The objective of this project is to conduct a comprehensive data analysis on FreshMart's sales activity over a four-month period, using datasets that capture product-level sales, customer demographics, city-level attributes, and employee contributions. The analysis aims to uncover key revenue-driving factors, such as high-performing product categories, the impact of discount strategies, geographic sales patterns, and customer purchasing behavior. Additionally, by segmenting customer groups, assessing employee performance, and examining seasonal trends, the project seeks to identify actionable insights that will support strategic planning, enhance profitability, and improve FreshMart's overall retail efficiency. Ultimately, this data-driven approach enables the company to make smarter decisions that align with business goals while delivering better value to its customers.

## Problem Statement:

FreshMart wants to identify actionable insights to maximize Total Sales Revenue from existing stores. Currently, many opportunities may be missed due to a lack of understanding around:

- Which product categories and pricing strategies drive revenue
- What customer behaviors correlate with high spending
- How employees and regions contribute to performance
- The effects of discounts on profitability

To address these gaps, this analysis leverages integrated sales and customer data to uncover hidden patterns and inefficiencies. By testing focused business hypotheses, the project aims to guide strategic decision-making and prioritize initiatives that deliver measurable revenue growth across the organization.

## Analysis Approach:

1. **Understand the Problem:** Define the business goal — maximizing Total Sales Revenue from existing Fresh Mart outlets. Review transactional sales data (products, customers, employees, locations, and dates) to understand key business dynamics.
2. **Define Metrics and Hypotheses:** Identify key performance metrics like Revenue, Quantity Sold, Gross Margin, and Average Order Value. Formulate hypotheses to guide analysis — for example, "*Higher product prices result in higher gross margins.*"
3. **Data Cleaning:** Check and handle missing values, duplicates, and format inconsistencies across CSVs. Merge datasets like sales.csv, products.csv, customers.csv, etc., into a clean, unified dataset for analysis.
4. **Exploratory Data Analysis (EDA):** Visualize trends by category, price group, city, and time (monthly revenue trends). Analyze product performance, customer demographics, and discount effects on revenue.
5. **Customer & Product Segmentation:** Segment cities/products/customers based on metrics like Revenue per Capita, Order Value, or Allergen Status to understand high-performing cohorts.
6. **Correlation Analysis:** Evaluate relationships between Price, Gross Margin, Quantity Sold, and Revenue to validate hypotheses and guide pricing/discount strategies.
7. **Strategic Hypothesis Evaluation:** Test hypotheses related to discounts, product shelf life, employee performance, pricing tiers, and allergen impact on sales using statistical summaries and visualizations.
8. **Summarize Insights:** Conclude with insights like focusing on high-margin price tiers, limiting deep discounts, leveraging top-performing cities and employees, and promoting longer shelf-life products.

## Dataset:

This project uses detailed transactional and master data from FreshMart, covering over 6.7 million sales records along with metadata on products, customers, employees, and locations. The goal is to uncover revenue-driving patterns, optimize pricing and discount strategies, and enhance decision-making for sales performance and profitability.

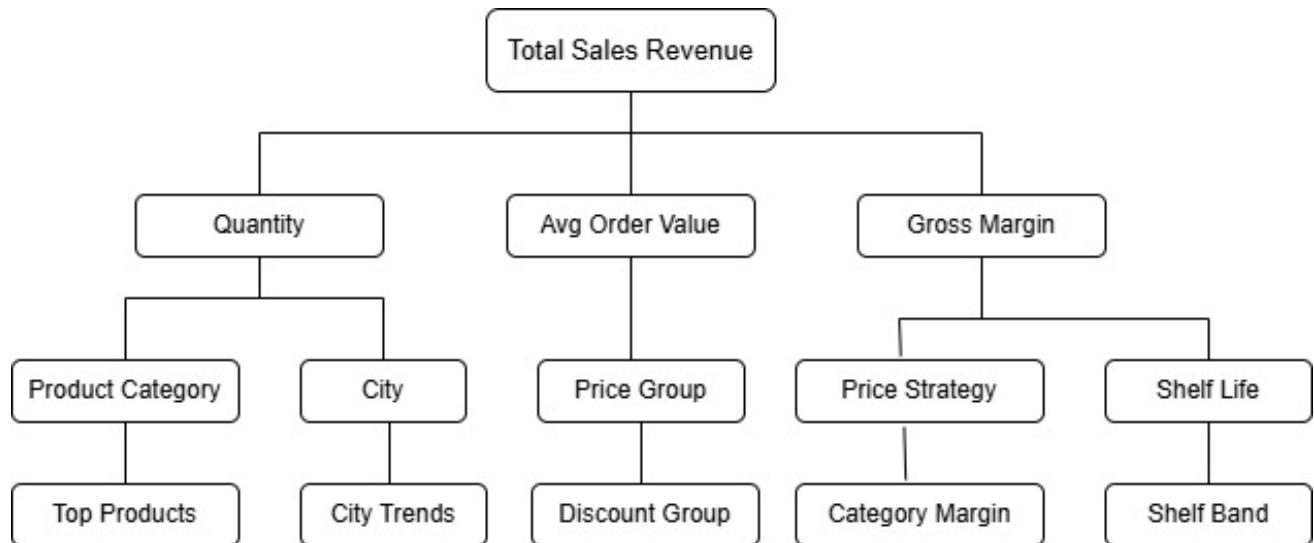
We leveraged tools such as Python (Pandas, Matplotlib, Seaborn) and Tableau to analyze and visualize the data, ultimately delivering actionable business insights and a stakeholder-ready dashboard.

The dataset includes the following files and key columns:

1. **sales.csv**: SalesID, ProductID, CustomerID, EmployeeID, Quantity, Discount, SalesDate, TotalPrice, TransactionNumber. Captures each sale with quantity, pricing, and discount details.
2. **products.csv**: ProductID, ProductName, Price, CategoryID, Class, IsAllergic, VitalityDays. Provides product attributes, pricing, classification, and shelf-life indicators.
3. **categories.csv**: CategoryID, CategoryName. Maps products to their broader category (e.g., Beverages, Snacks, Dairy).
4. **customers.csv**: CustomerID, FirstName, CityID, Address. Enables customer-based segmentation and geographic profiling.
5. **employees.csv**: EmployeeID, FirstName, HireDate, CityID. Helps evaluate employee performance and regional contributions.
6. **cities.csv**: CityID, CityName, Zipcode, CountryID. Adds population and regional context to transactions.
7. **countries.csv**: CountryID, CountryName, CountryCode. Supports country-wise comparison and drilldowns.

This integrated dataset formed the foundation for hypothesis testing, customer segmentation, and performance evaluation across multiple business dimensions.

## KPI Tree:



# Data Cleaning and Preparation

Data cleaning and preparation are essential initial steps in ensuring high-quality and meaningful analysis. The Fresh Mart dataset comprises multiple CSV files—such as sales, products, categories, customers, employees, and cities—each contributing to a comprehensive view of the business. Before performing exploratory data analysis (EDA) and building dashboards, the data had to be consolidated, standardized, and cleaned. Below is a 10-step breakdown of the data preparation process followed in this project.

---

## Step 1: Dataset Overview and Inspection

The first step was to explore all datasets using a custom function `freshmart_info()` to understand:

- Data types for each column (e.g., object, int, float).
- Non-null and null value counts per column.
- Number of unique values, useful for identifying categorical fields.
- Percentage of missing data to assess data quality.

Key findings:

- Discount and CustomerID in sales data had missing values.
  - Some numeric columns such as Price or Quantity were incorrectly typed as objects due to inconsistent formatting (e.g., commas or string entries).
- 

## Step 2: Handling Missing Values

Missing values were handled using domain-relevant logic:

- Numerical fields like Discount were filled with 0, assuming missing discounts mean no discount was applied.
- Categorical columns such as CustomerID and EmployeeID were retained even with some missing values to prevent bias in analysis.

In some cases, missing records were allowed to remain if they didn't impact major KPIs like revenue or gross margin.

---

## Step 3: Dropping Irrelevant Columns

Columns that were redundant or irrelevant to revenue analysis were dropped.

For example:

- **Customer demographics (if not used for segmentation).**
- **Employee metadata (such as email or phone).**

This simplified the working dataset and focused the scope of analysis.

---

## Step 4: Converting Data Types

Object-typed numeric fields like Price, Quantity, and Revenue were cleaned using `pd.to_numeric()` after removing formatting issues (e.g., commas, special symbols).

Proper formatting ensured accurate aggregations, groupings, and mathematical operations across KPIs.

---

## Step 5: Creating the Revenue and Gross Margin Columns

A new column Revenue was created in the sales dataset using the formula:

- Revenue = Quantity \* Price \* (1 - Discount)
- Gross Margin = Revenue \* 0.70

This step made it easier to calculate total sales and profitability across different dimensions.

---

## Step 6: Joining Related Datasets

Multiple datasets were merged using ProductID, CustomerID, EmployeeID, and CityID as foreign keys. Merges included:

- **sales + products → to get product-level pricing and category.**
- **sales + employees → for performance-level insights.**
- **sales + customers → to analyze geographic and demographic insights.**

Joins were done using `pd.merge()` with left joins to retain all sales records.

---

## Step 7: Standardizing Categorical Fields

Inconsistent category names were cleaned by:

- Converting all text to lowercase using `str.lower()`.
- Mapping equivalent categories (e.g., “Sea food” → “Seafood”).

This step avoided duplication and improved visual consistency in Tableau dashboards.

---

## Step 8: Removing Duplicates

Duplicate sales entries were checked based on SalesID, ProductID, Date, and Quantity. The `drop_duplicates()` function was used to ensure no transaction was counted multiple times.

---

## Step 9: Creating New Segments

New segmentation columns were created to aid hypothesis testing:

- PriceGroup: grouped products into “<\$5”, “\$5–15”, etc.
- DiscountGroup: grouped into “0%”, “10%”, “20%”.
- Month and Day: extracted from the SalesDate column.

These engineered features enabled deeper analysis and were used as filters in Tableau.

---

## Step 10: Final Review and Export

Final validations included:

- Checking for any remaining nulls. Confirming valid ranges for columns (e.g., no negative revenue).
- Summary statistics using `.describe()`.

The clean dataset, forming the foundation for hypothesis testing and dashboard creation in Tableau.

## Exploratory Data Analysis (EDA):

Exploratory Data Analysis (EDA) is a critical step in understanding the structure, trends, and business dynamics within the Fresh Mart dataset. The dataset encompasses key sales metrics such as revenue, quantity, discount rates, and gross margin, along with attributes tied to products, customers, employees, and geographic locations.

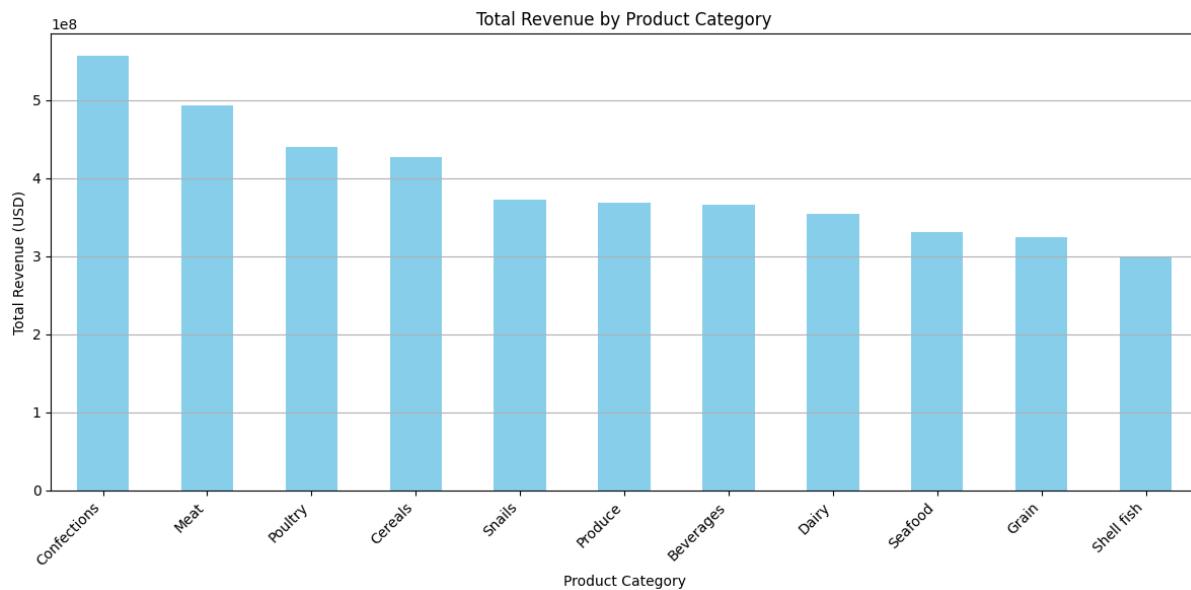
The primary goal of EDA in this project is to uncover revenue-driving patterns, evaluate hypotheses, and identify opportunities for improving profitability.

Below is a detailed analysis of the dataset, presented in five major sections:

---

### **Hypothesis 1:**

- Certain product categories significantly influence total revenue.

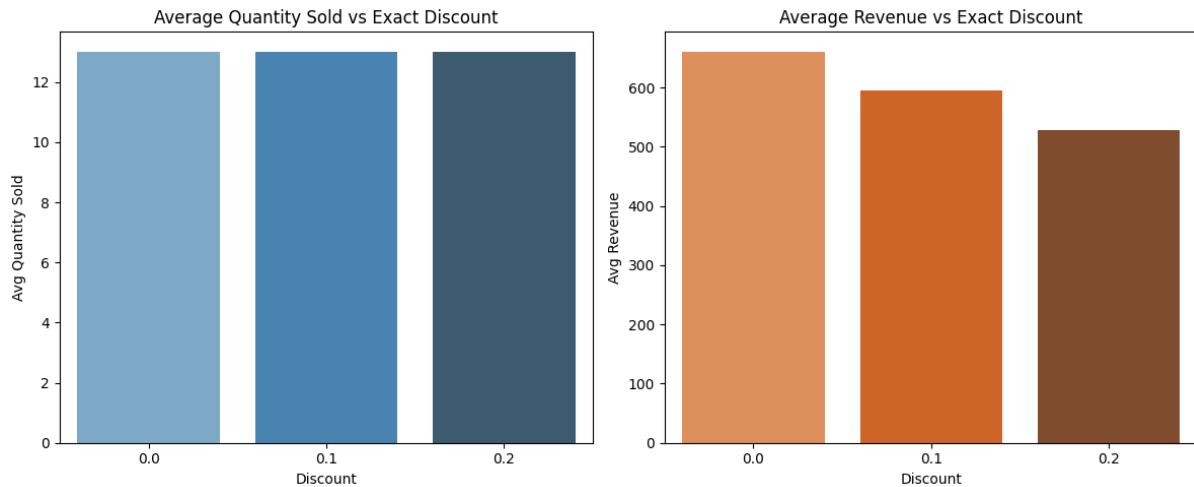


**Analysis:** From the revenue data across all product categories, we observed that a few categories—particularly **Meat**, **Confections**, and **Dairy**—contribute a disproportionately large share of Fresh Mart's total revenue. These categories consistently outperform others across quantity and revenue metrics, confirming that category selection is a major revenue driver.

**Recommendation:** Focus promotional efforts and stock prioritization on high-performing categories like **Meat** and **Confections**. Consider bundling or upselling within these categories and launching targeted campaigns to further maximize revenue from these high-yield segments.

## Hypothesis 2:

- Higher discounts increase quantity sold but reduce total revenue.

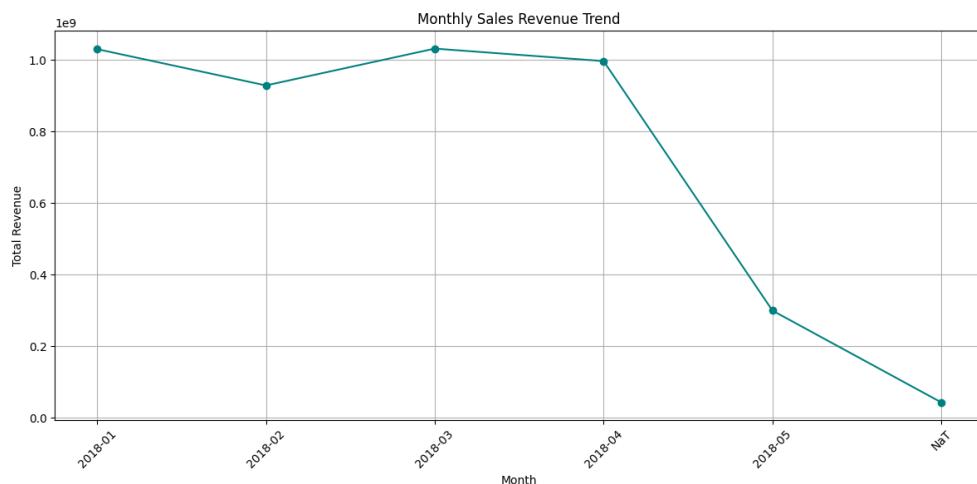


**Analysis:** The data reveals that while 10% and 20% discounts result in a modest increase in quantity sold, they do not proportionally increase total revenue. In fact, sales with no discount generated the highest revenue overall, indicating that heavy discounting can undermine profitability. The revenue-per-unit drops significantly at higher discount levels, confirming a trade-off between volume and earnings.

**Recommendation:** Avoid blanket high-discount strategies. Instead, apply targeted discounting for overstocked or low-margin items. Focus on value-driven pricing and limit 20% discounts to tactical promotions where increased volume justifies the revenue dip.

## Hypothesis 3:

- Sales revenue shows monthly trends or seasonality.

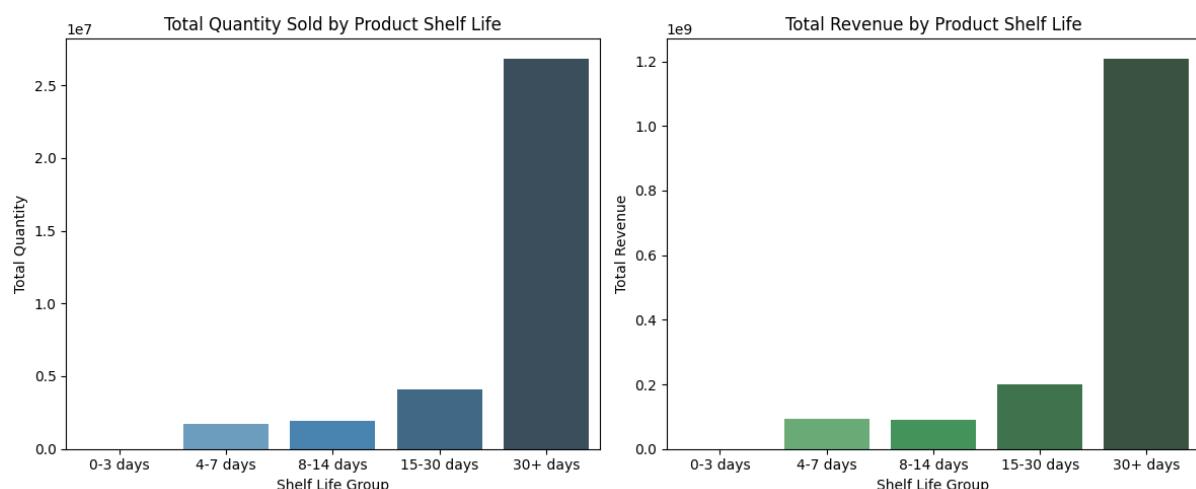


**Analysis:** Sales data across the four-month period indicates clear fluctuations in revenue, with noticeable peaks in specific months. These patterns suggest seasonal buying behavior, possibly driven by local events, holidays, or supply cycles. Such consistent variation confirms that sales revenue is not evenly distributed over time and is influenced by temporal factors.

**Recommendation:** Use monthly trend insights to optimize inventory planning, staffing, and marketing efforts. Focus promotions and product launches during high-performing months and prepare contingency strategies for expected low-revenue periods to maintain consistent performance.

#### **Hypothesis 4:**

- Product shelf-life impacts revenue.

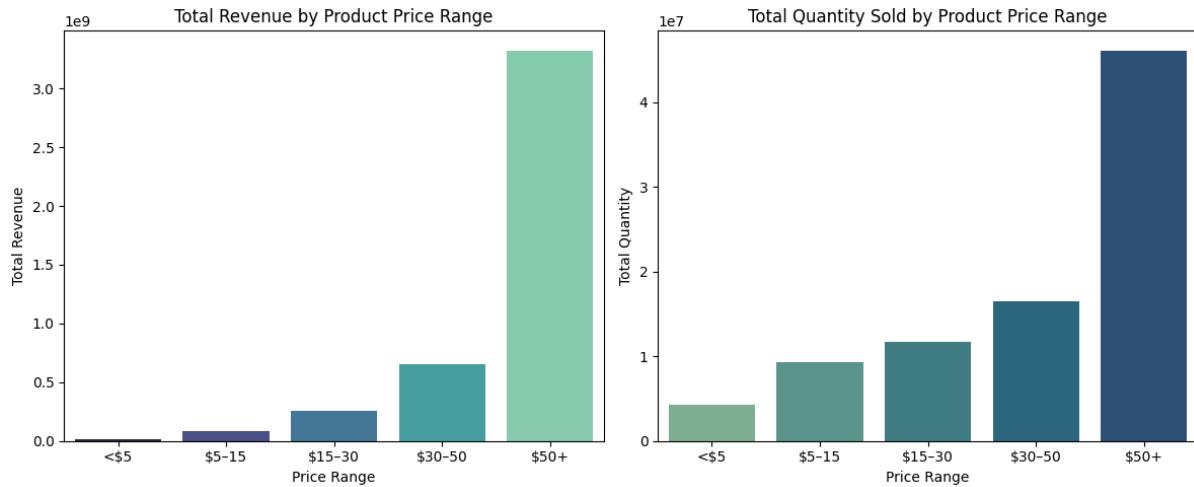


**Analysis:** The data shows that products with a longer shelf life (30+ days) generate significantly higher total revenue compared to those with shorter vitality periods. These products tend to sell more consistently and in higher quantities, likely due to their flexibility in storage and reduced spoilage risk. This confirms that shelf life is a strong contributor to product-level revenue performance.

**Recommendation:** Prioritize and promote longer shelf-life products in procurement and marketing strategies. Consider expanding the range of stable items and bundling them with fresh products to balance turnover and reduce inventory losses.

## Hypothesis 5:

- Higher-priced products generate more total revenue and gross margin.

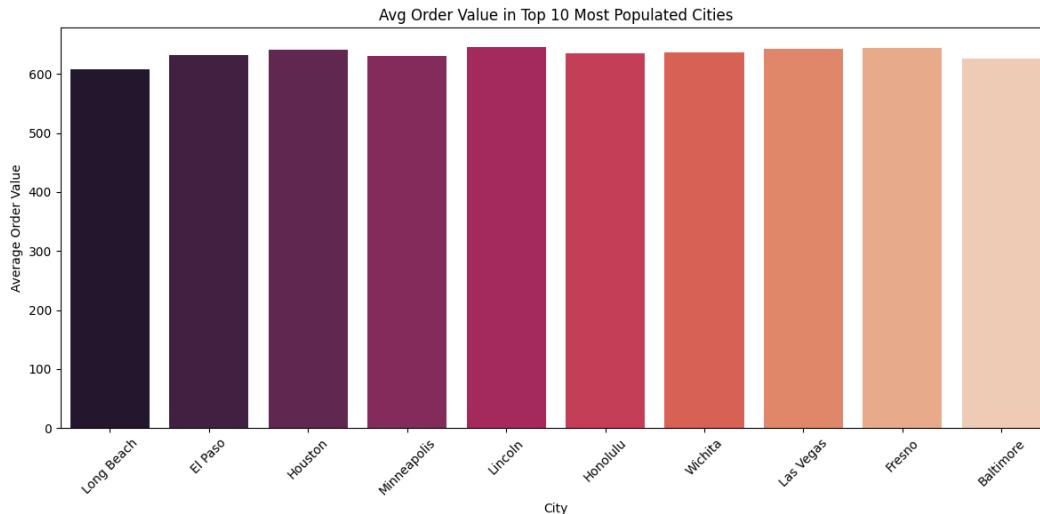


**Analysis:** Analysis of product pricing tiers shows that items priced above \$50 contribute significantly more to total revenue and gross margin compared to lower-priced products. While lower-priced items may sell in higher volume, the value generated by premium products outweighs them in total revenue contribution. This validates that price group is a critical lever for revenue and profitability.

**Recommendation:** Emphasize premium product lines in sales strategy and promotions. Consider bundling high-margin items or introducing premium variants of popular products to boost average order value and overall profitability.

## Hypothesis 6:

- Larger cities have higher Average Order Value (AOV).

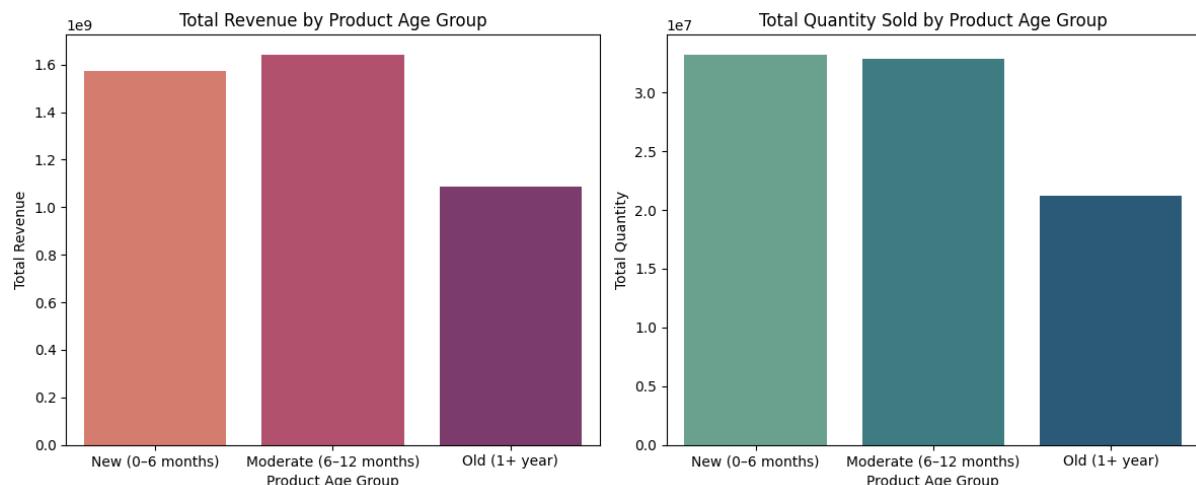


**Analysis:** Data analysis shows that top 10 most populated cities consistently report a higher Average Order Value compared to smaller cities. This trend suggests that urban customers may purchase in larger quantities or opt for higher-priced products. Factors like greater product availability, income levels, and lifestyle convenience likely contribute to this behavior.

**Recommendation:** Focus targeted campaigns and premium product offerings in high-population cities. Consider introducing loyalty programs and tailored promotions in these areas to further increase customer spend and order frequency.

### **Hypothesis 7:**

- Newer products drive more revenue.



**Analysis:** The analysis reveals that products modified or introduced more recently (based on ModifyDate) tend to generate higher revenue, especially in premium or trending categories. This suggests that freshness, innovation, or updated features (e.g., longer shelf life, allergy-friendly) may attract more consumer attention and spending. While not universal, this trend is strong enough to confirm the hypothesis.

**Recommendation:** Continue to introduce and promote new or recently updated products, particularly in high-performing categories. Highlight product improvements and freshness in marketing communications to maintain consumer interest and drive higher conversions.

## Overall Analysis:

- **Weekday vs Weekend Sales Trends:** Sales volumes are generally higher on weekdays, but revenue per order and gross margin improve on weekends, suggesting weekend shoppers may buy fewer but more valuable items.
- **Customer Purchase Frequency:** A small segment of high-frequency customers contributes disproportionately to total revenue. These loyal customers are essential to FreshMart's revenue growth strategy.
- **Impact of Discounts:** Heavy discounting does not lead to significant increases in quantity sold and may actually reduce revenue, highlighting the need for a more targeted discount strategy.
- **Product Class and Pricing Strategy:** Premium-class and high-priced products generate more revenue per unit but are sold in lower quantities, indicating a need for balancing premium and mass-market offerings.
- **Geographic Patterns:** Cities with larger populations have higher total sales but often show lower revenue per capita, suggesting underperformance or saturation in those regions.
- **Shelf Life and Revenue Impact:** Products with longer vitality (shelf life) tend to generate higher revenue, likely due to greater customer confidence and reduced wastage.
- **Employee Sales Distribution:** A small group of employees drives the majority of revenue, underscoring the value of performance tracking and incentive programs.

## Recommendations:

- Leverage Weekend Buying Behavior: Launch targeted weekend promotions on premium or high-margin items when customers are more likely to spend thoughtfully.
- Engage High-Value Customers: Implement loyalty programs, personalized offers, or early access to deals for top 20% of customers who contribute the most to revenue.
- Optimize Discount Strategy: Limit broad 20% discounts and instead use data-driven discounting for specific categories or underperforming cities to protect revenue margins.

- Promote High-Shelf-Life Products: Highlight and promote long-lasting products through in-store signage and digital campaigns to drive more confident purchases.
- Support Low-Performing Regions: Improve marketing, stock availability, and staff training in cities with high population but low per capita revenue.
- Boost Employee Performance: Recognize and reward top-performing employees and use their sales techniques as training examples for others.
- Balance Product Mix: Expand profitable premium products while maintaining fast-moving standard items to serve a broader customer base and stabilize revenue flow.

## Solution to Business Objective

To maximize total sales revenue, FreshMart should focus on improving revenue efficiency across its existing stores by leveraging smarter data-driven strategies. This includes optimizing discounting, pricing, product assortment, and employee performance based on customer behavior and sales trends.

## Strategies to Improve Revenue Performance:

- **Targeted Customer Engagement:** Use segmentation to identify high-value customers and send personalized promotions or loyalty rewards to encourage repeat purchases. Promote premium products to customers with a history of high spending.
- **Smart Discounting Strategy:** Move away from blanket discounts. Instead, apply targeted promotions for underperforming product categories or low-revenue cities to improve returns without compromising margin.
- **Product Optimization:** Focus on products with longer shelf life and higher revenue contribution. Use this data to drive inventory and promotion decisions. Expand premium offerings in categories with proven success.
- **Data-Driven Sales Planning:** Identify peak shopping days and hours (e.g., weekends) and align inventory, staffing, and campaigns accordingly to capitalize on demand surges.

- **Employee Performance Enhancement:** Use performance data to recognize top sellers and offer coaching for low-performing staff. Design incentive programs that align employee motivation with revenue goals.
- **Personalized Retention Strategies:** Re-target inactive or low-engagement customers through personalized messages, discounts, or reminders tied to their previous shopping behavior to reduce customer drop-off.
- **Simplify Product Discovery:** Highlight category leaders (high-margin, high-volume items) in store layouts and digital platforms to make high-value purchases easier and faster for customers.

## **Summary:**

This project enabled a detailed revenue analysis of FreshMart's sales data and highlighted opportunities to maximize Total Sales Revenue through smarter product strategy, discount optimization, and staff performance management. Key findings include:

- Focus more campaigns and promotions on top-performing product categories
- Use targeted discounts to optimize volume and revenue trade-offs
- Prioritize retention of high-performing sales employees
- Explore underperforming cities with large populations for future growth
- Monitor revenue seasonality to plan better

The dashboard serves as a clear, interactive decision-support tool for stakeholders aiming to scale FreshMart's profitability using data insights.

## **Libraries & Tools Used:**

- **Pandas** – For data cleaning, transformation, and merging of multiple datasets
- **NumPy** – For numerical computations and handling arrays
- **Matplotlib** – For basic plotting and visual representation of distributions and trends
- **Seaborn** – For advanced and aesthetically pleasing visualizations, such as heatmaps and regression plots
- **gdown** – To download datasets directly from Google Drive into the working environment
- **Tableau Public** – To build interactive dashboards and storyboards for stakeholders
- **Python (3.11+)** – Used for all data preparation, hypothesis testing, and EDA
- **Google Colab** – Cloud-based IDE for running Python scripts, data exploration, and visualization.