

INFO 5602

Information Visualization

Group Project Reporting



From Beans to Business: Data-Driven Insights from Coffee Shop Sales



Team Members

Rishekesan Senthil Kumar Vanathi

Sruthijha Pagolu

Vikash Baabhu Ramesh Baabhu

Table of Contents

- Project Overview & Context
- Problem Definition & Research Questions
- Project Motivation & Relevance
- Methodology
- Data Sources
- Data Analysis Approach
- Exploratory Data Analysis (EDA)
- Key Insights & Findings
- Conclusion

Context

- This project analyzes **149,116 transactions** from a New York City coffee-shop chain with three locations: Hell's Kitchen, Astoria, and Lower Manhattan, collected between **January 1 and June 30, 2023**.
- Each transaction includes **date, time, store, product details, quantity, and unit price**, allowing us to derive trends in revenue, demand, and purchasing behavior.
- The business offers a range of **Coffee, Tea, Bakery items, Drinking Chocolate, Coffee Beans**, and branded products.
- Our goal is to transform raw sales logs into **actionable visual insights** that inform **staffing, inventory, and marketing decisions**.

Problem Definition

Problem: The coffee shop chain possesses extensive transaction data but lacks a clear visual narrative that reveals *when* customers buy, *what they buy*, and *which stores or products* drive revenue.

Research Questions:

- 1. Temporal Trends:** How do sales vary over time (**by month, day of week, and hour of day**) across the six-month period?
- 2. Store Performance:** How do the three **store locations** differ in revenue, order volume, and average order value (AOV)?
- 3. Product Drivers:** Which **product categories** and **product types** generate the most revenue and sales volume?
- 4. Business Actions:** Based on these patterns, what **recommendations** can improve staffing, inventory, and promotional strategies?

Project Motivation

Operational Efficiency: Staffing schedules and store hours should align with *actual customer traffic patterns* such as the morning rush versus slower afternoon hours to optimize labor costs and service quality.

Inventory Optimization: Understanding which **categories** (e.g., Coffee, Tea, Bakery) perform best at specific stores enables smarter procurement and reduces waste from overstocking low-demand items.

Targeted Marketing: Knowing **low-traffic hours** or **underperforming product types** helps design promotions that attract customers during off-peak times and balance demand across the day.

Visualization Value: This dataset demonstrates how **effective data visualization** converts raw sales logs into clear, actionable insights that support better decision-making, not just aesthetically pleasing charts.

Proposed Methodology

1. Data Preparation

- Loaded **149,116 transactions** (11 variables) from three NYC locations.
- Cleaned data, handled missing values, and derived fields:

$$sales = qty \times unit\ price, day\ of\ week, month, hour$$

2. Exploratory Data Analysis (EDA)

- Aggregated by **time, store, and product category/type**.
- Computed key metrics: **revenue, quantity sold, transactions, and AOV**.
- Designed visualizations to uncover temporal, store, and product patterns.

3. Visualization & Insight Generation

- Applied information visualization principles for **clarity and accuracy**.
- Interpreted visuals to generate **business insights** on staffing, inventory, and marketing.

Data source

- **Dataset:** *Coffee Shop Sales - Transactions* (Kaggle).
- **Time Range:** *January 1 - June 30, 2023* (6 months).
- **Records:** *149,116 transactions* across three NYC stores - **Hell's Kitchen, Astoria, and Lower Manhattan.**
- **Key Fields:**
 - *Transaction:* date, time, quantity, and unit price
 - *Store:* location identifiers
 - *Product:* category, type, and details
- **Derived Metric:**
 - **Sales (Revenue)** = *transaction_qty × unit_price*

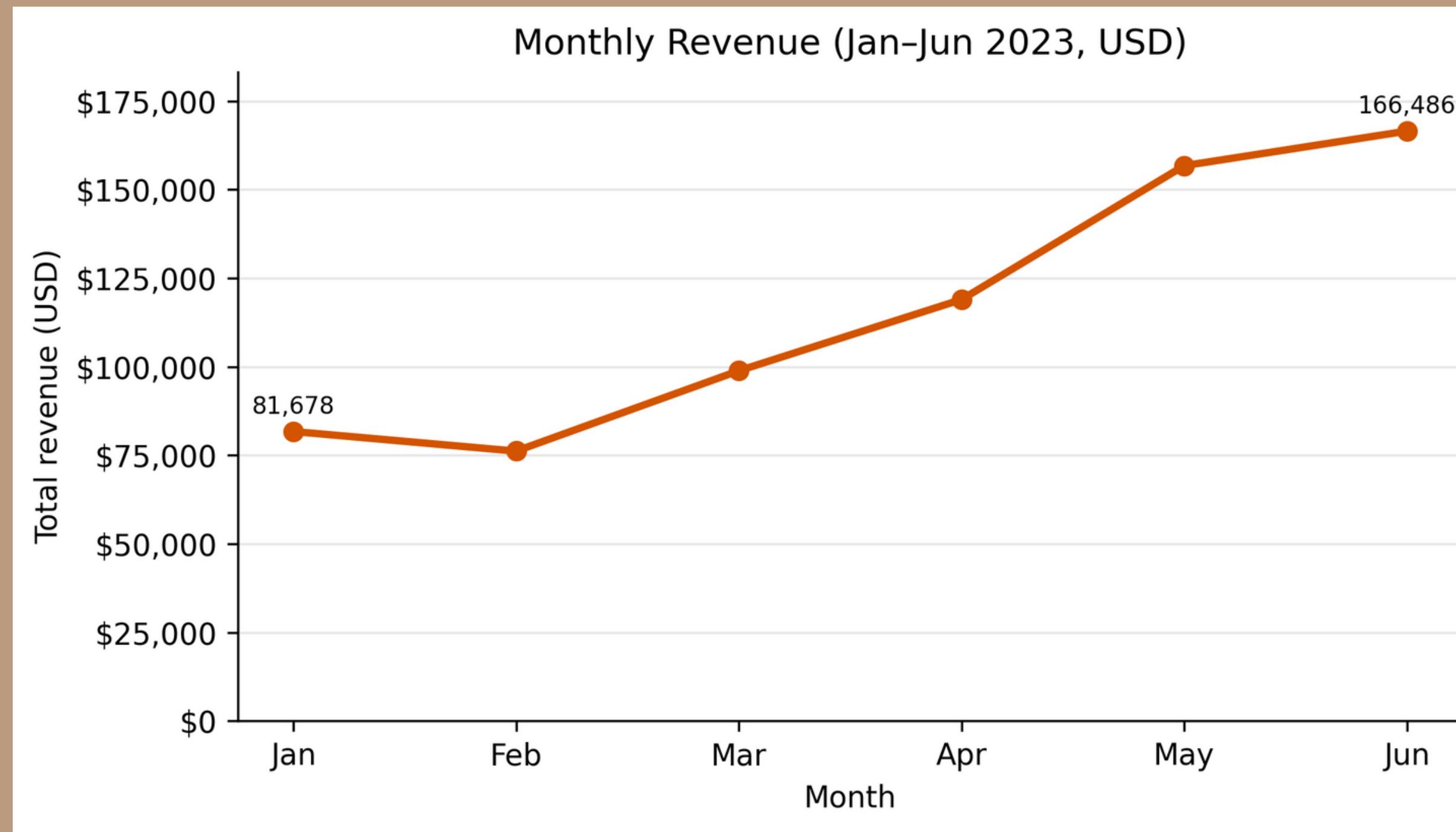
Data Analysis

- **Time Series Trends:** Analyzed **monthly**, **day-of-week**, and **hour-of-day** patterns to reveal demand cycles and peak sales periods.
- **Store Performance Comparison:** Compared **revenue**, **transaction counts**, and **average order value (AOV)** across the three NYC locations to identify operational differences.
- **Product Mix Analysis:** Evaluated **revenue share** by product category (Coffee, Tea, Bakery, etc.) and identified top-performing product types.
- **Cross-Analysis:** Combined multiple dimensions: **time × product category** and **store × product mix**, to discover deeper behavioral and sales patterns.

EDA - Monthly Trend

- **Revenue Growth:** Monthly revenue shows a strong **upward trajectory**, increasing from **\$81.7K in January** to **\$166.5K in June** - more than double in six months.
- **Customer Growth:** Transaction counts rise consistently, indicating genuine increases in customer volume rather than price effects.
- **Peak Months:** May and June record the highest revenue, suggesting **seasonal demand** or successful marketing during early summer.
- **Demand Drivers:** Parallel growth in revenue and transaction trends confirms **demand-driven scaling** rather than one-time spikes.

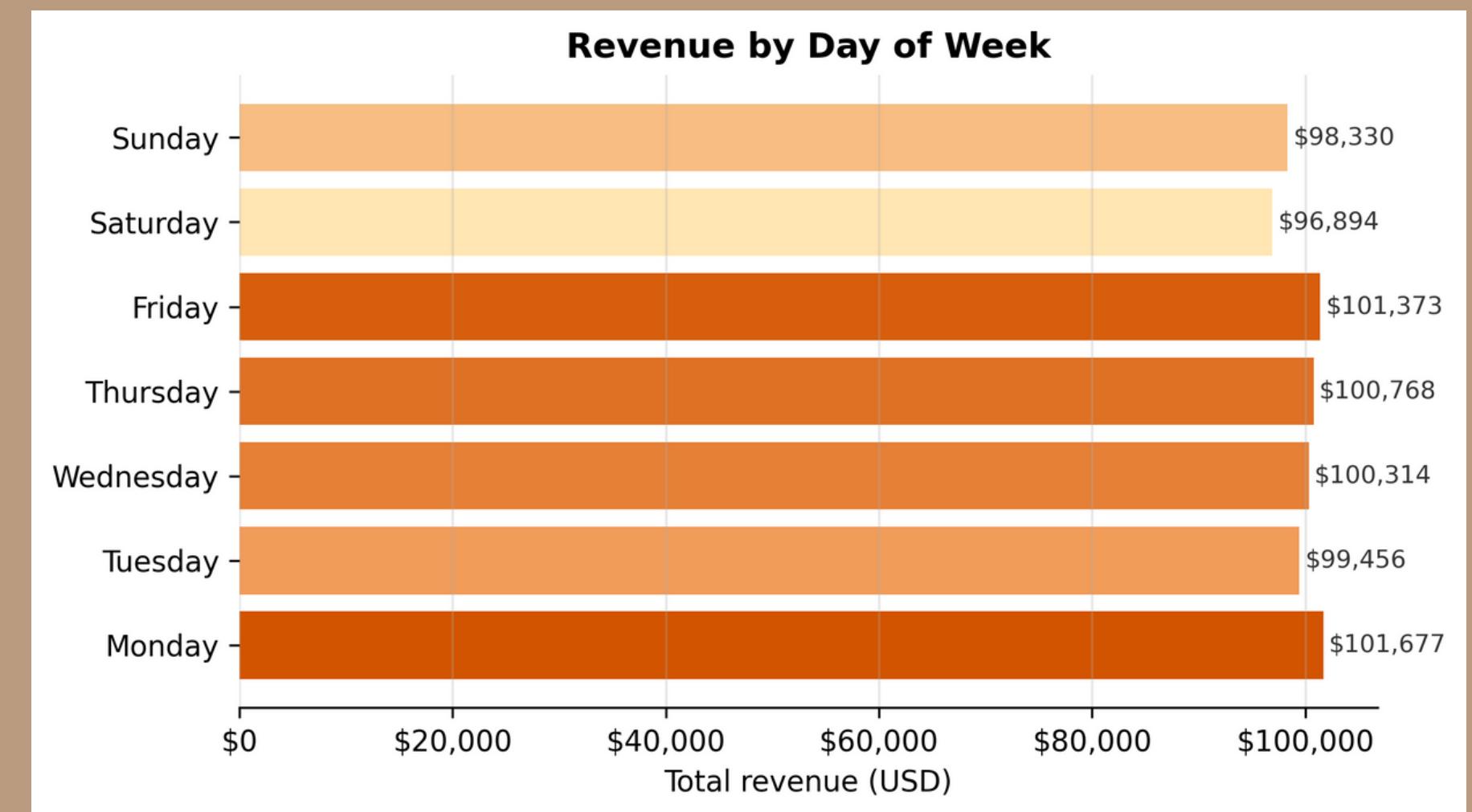
EDA - Monthly Trend



EDA - Day-of-Week Trend

Day-of-Week Trends: Sales remain steady across all days, with only minor differences between weekdays and weekends.

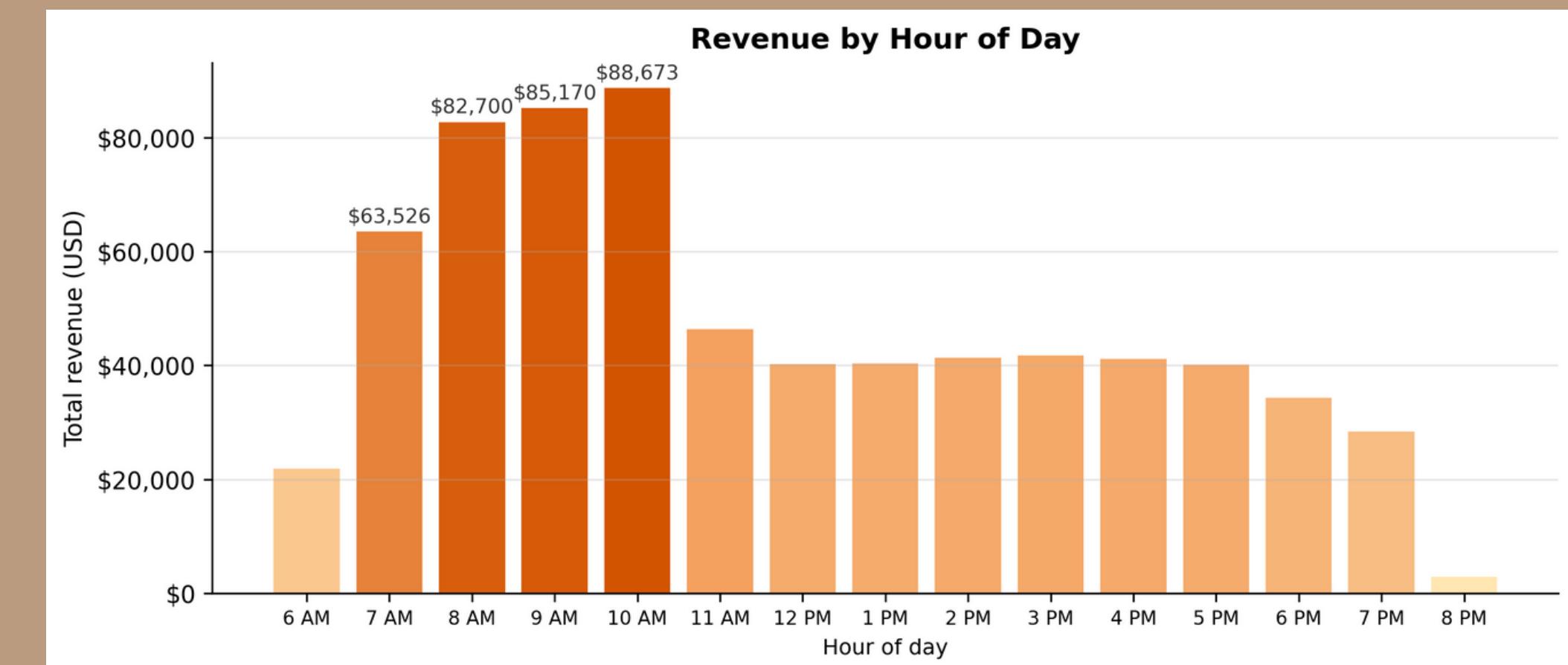
- **Monday** and **Friday** show slightly higher revenue.
- **Saturday** is marginally lower, indicating consistent but not weekend-dependent demand.



EDA - Hour-of-Day Trend

Time-of-Day Trends:

- Strong morning concentration from 7 AM – 11 AM, contributing over 50 % of total revenue.
- 10 AM is the single highest-earning hour, followed by 8 - 9 AM, reflecting commuter coffee demand.
- Afternoon sales are steady but moderate, signaling potential for off-peak promotions.

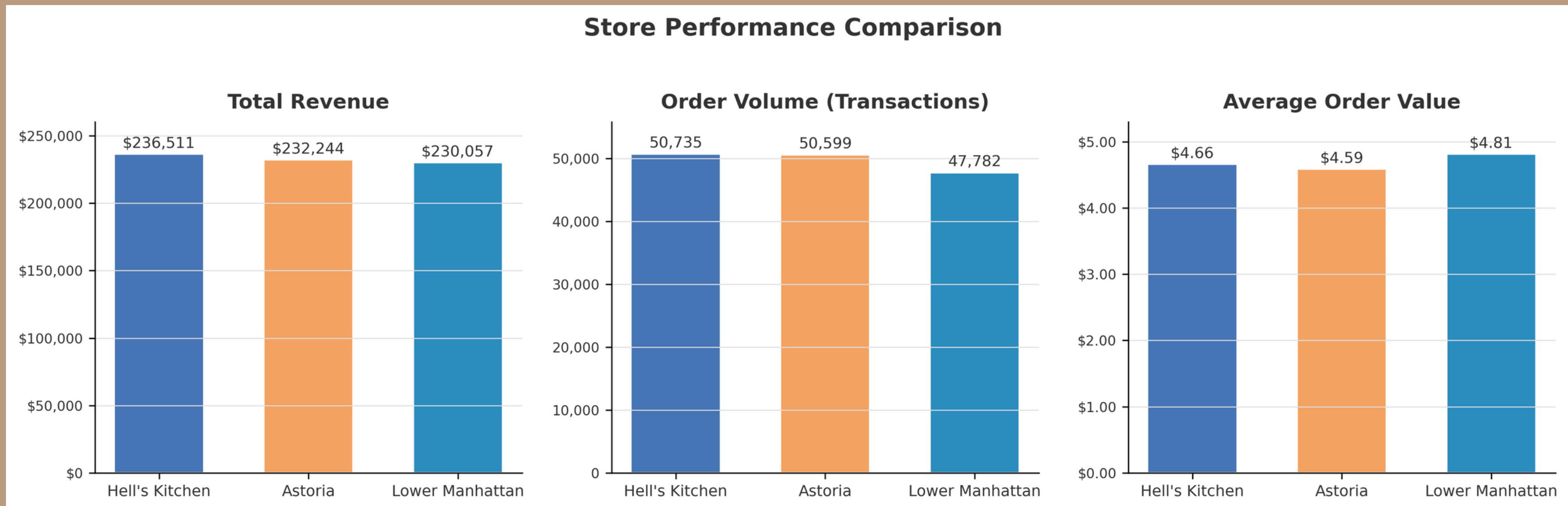


EDA - Store Performance by Location

- **Balanced Revenue Split:** All three stores contribute almost equally to total revenue (~33% each of \$698.8K).
- **Top Performer:** Hell's Kitchen leads slightly at ≈ \$236.5K, followed by Astoria (~ \$232.2K) and Lower Manhattan (~ \$230.1K).
- **Customer Reach:** Comparable transaction volumes across all stores indicate consistent demand and similar customer bases.

Key Takeaway: Store performance is balanced, suggesting that location is not a major differentiator - operations and customer traffic remain strong across all regions.

EDA - Store Performance by Location

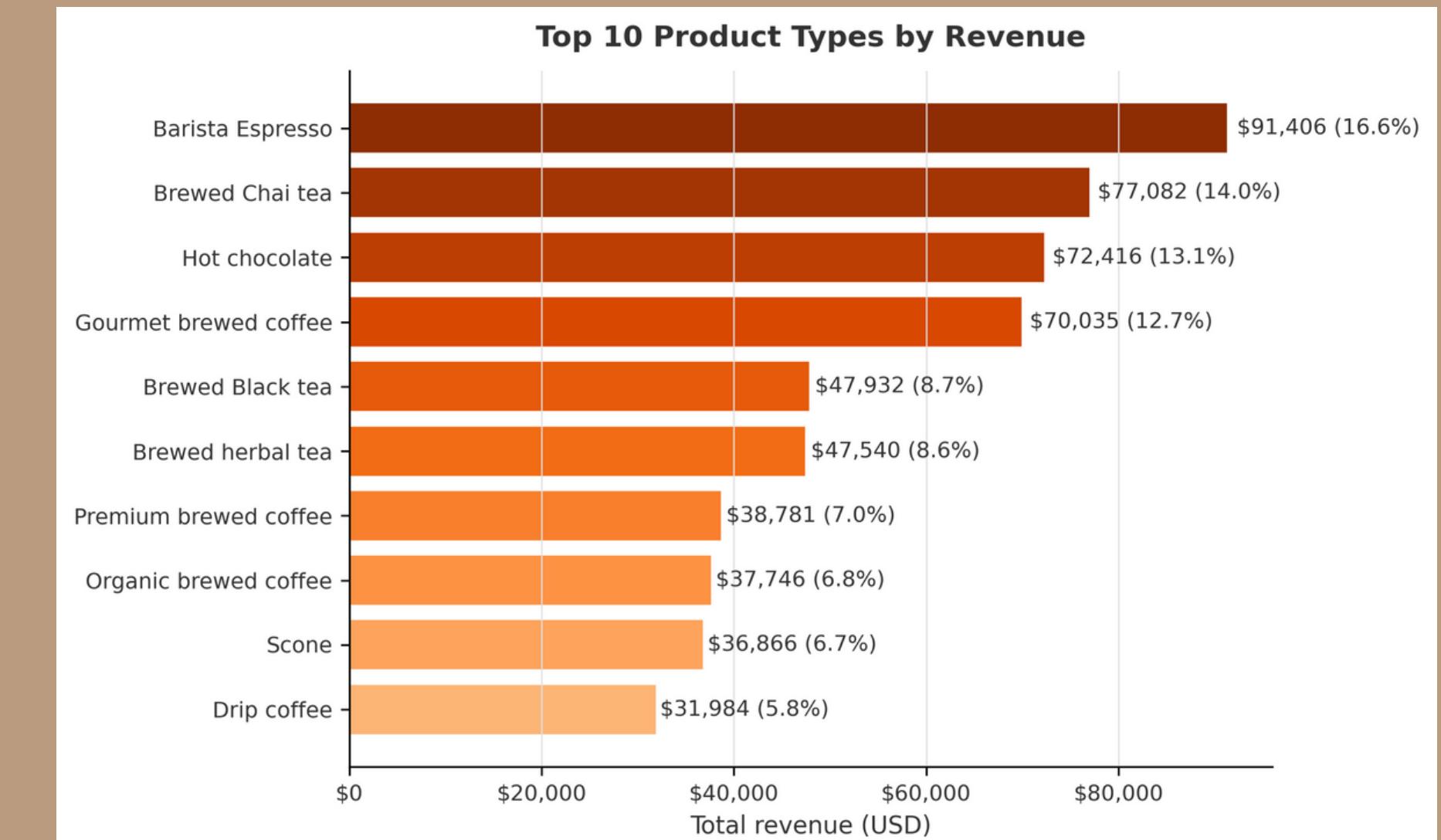
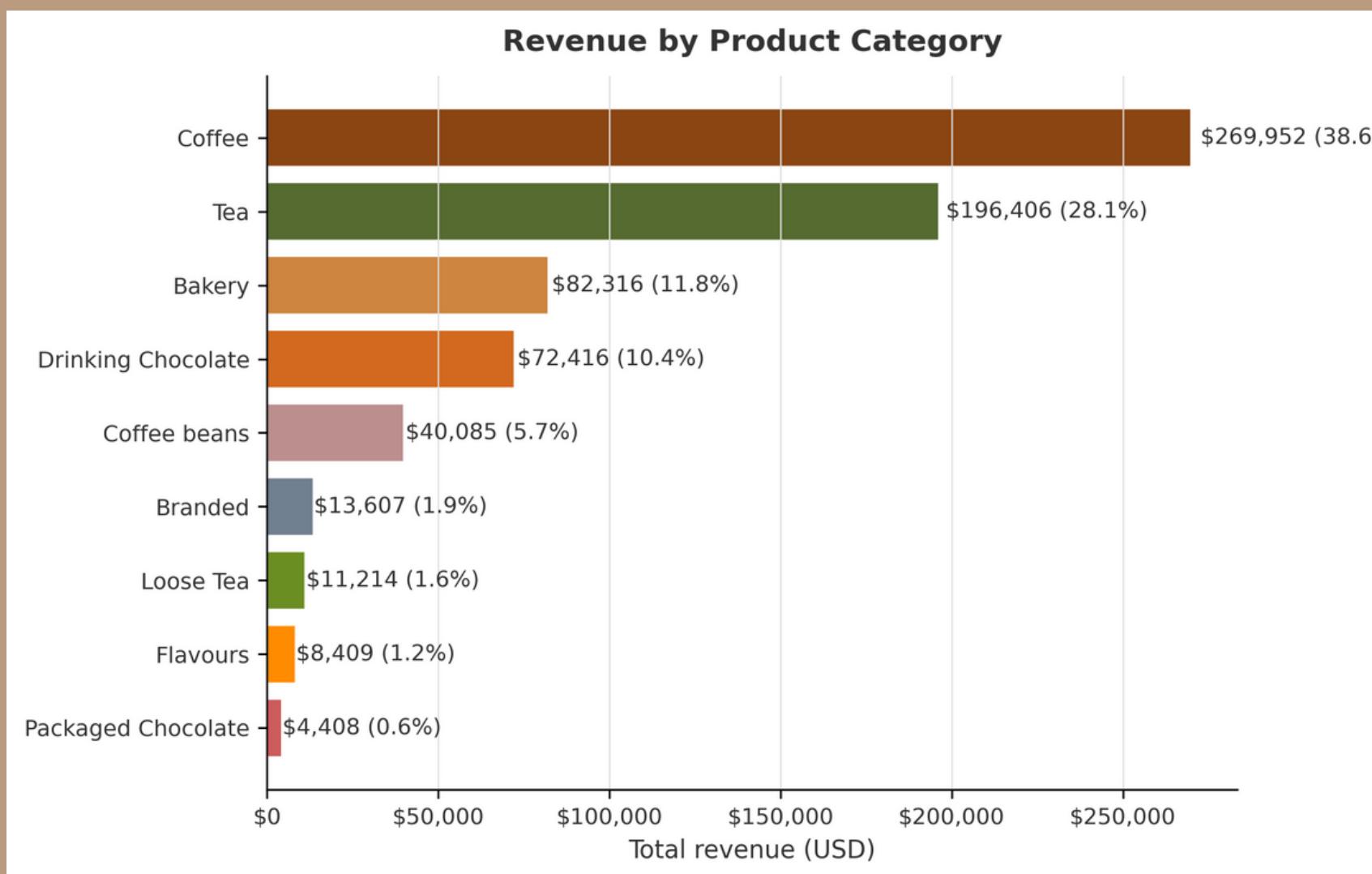


EDA - Product Category Performance

- **Core Revenue Drivers:** Coffee (\$269.9K, ~38.6%) and Tea (\$196.4K, ~28.1%) are the main revenue pillars, together contributing nearly **67% of total sales**.
- **Supporting Categories:** Bakery (\$82.3K, ~11.8%) and Drinking Chocolate (\$72.4K, ~10.4%) add meaningful cross-sell opportunities and drive combo purchases.
- **Niche Offerings:** Coffee Beans (~5.7%) and smaller packaged goods contribute less overall but may carry **higher profit margins** and brand value.

Key Takeaway: The shop's success relies heavily on **beverage sales**, while complementary food and specialty items support **upselling and diversification**.

EDA - Product Category Performance



Summary of Key Insights - Time & Traffic

- Revenue doubled from **\$81.7 K** in January to **\$166.5 K** in June, showing strong upward growth.
- Transaction volume increased proportionally, confirming genuine customer expansion rather than price effects.
- **May and June** marked the highest-performing months, indicating seasonal momentum.
- Sales remained consistent across all weekdays, with no major weekday-weekend variation.
- More than **50 % of total revenue** occurred between **7 AM and 11 AM**, highlighting the dominance of morning demand.
- **10 AM** was the peak hour for both transactions and revenue.

Key takeaway: Morning hours drive business performance and should guide staffing and promotional planning.

Key Insights - Store & Product Insights

- All three stores generated similar revenue, each contributing roughly **one-third of total sales (\$698.8K)**.
- **Lower Manhattan** recorded the **highest average order value (AOV)**, suggesting stronger premium-product purchasing behavior.
- **Coffee and Tea** categories dominated overall performance, accounting for about **two-thirds of total revenue**.
- **Bakery** and **Drinking Chocolate** items served as effective complementary products for cross-selling opportunities.
- Top-selling product types included **Barista Espresso**, **Brewed Chai Tea**, **Hot Chocolate**, and **Gourmet Brewed Coffee**.
- Product preferences were consistent across stores, indicating a **uniform customer profile**.
- Afternoon and evening sales were moderate, offering scope for **targeted off-peak promotions**.

Conclusion

RQ1 - How do sales vary over time?

- Sales showed a **strong upward trend**, with revenue more than **doubling** from January (\$81.7K) to June (\$166.5K).
- Morning hours (**7–11 AM**) consistently generated **over half of total revenue**, confirming demand peaks during commuter times.

RQ2 - How do store locations differ?

- All three stores performed **similarly in total revenue**, indicating balanced demand.
- **Lower Manhattan** had the **highest AOV**, suggesting greater premium product purchasing.

RQ3 - Which products drive performance?

- **Coffee** and **Tea** are the core revenue categories, together accounting for about **two-thirds of total sales**.
- Top-performing product types include **Barista Espresso**, **Brewed Chai Tea**, and **Hot Chocolate**; these items should be kept **highly available** and prominently featured across all stores.

Conclusion

RQ4 - What are the business recommendations?

Staffing:

- Maintain **maximum staff between 7-11 AM**, especially during **8-10 AM peak hours**.
- Reduce evening staffing while ensuring steady service quality.

Inventory & Menu:

- Prioritize **Coffee, Tea, Bakery, and Drinking Chocolate** stock.
- Promote **coffee beans** and **premium drinks** more at **Lower Manhattan**, given its higher AOV.

Marketing & Promotions:

- Introduce **combo offers (coffee + bakery)** during morning peaks.
- Run **targeted afternoon promotions** to boost off-peak sales.
- **Highlight best-selling items in menus, loyalty apps, and in-store signage.**

Thank You

