# Online Retail Sales Report

## Dataset Description

This is a transactional dataset that includes all purchases made between 01/12/2010 and 09/12/2011 by customers of a UK-based and registered non-store online retailer. The company primarily sells unique, gift-related products through its online platform, with the majority of its customers being wholesalers. Each transaction includes details such as invoice number, stock code, item description, quantity, invoice date, unit price, customer ID, and country of purchase.

This dataset was generously donated to the UCI Machine Learning Repository by Daqing Chen on November 5, 2015. It is made publicly available under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, allowing free use, distribution, and adaptation of the data provided proper attribution is given. The dataset serves as a rich resource for exploring customer segmentation, sales trends, and data preprocessing techniques.The following is a description of the dataset columns.

| Variable Name | Type | Description |
|---|---|---|
| InvoiceNo | Categorical | a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter "c", it indicates a cancellation |
| StockCode | Categorical | a 5-digit integral number uniquely assigned to each distinct product |
| Description | Categorical | product name |
| Quantity | Integer | the quantities of each product (item) per transaction |
| InvoiceDate | Date | the day and time when each transaction was generated |
| UnitPrice | Continuous | product price per unit |
| CustomerID | Catgeorical | a 5-digit integral number uniquely assigned to each customer |
| Country | Categorical | the name of the country where each customer resides |

# Key Takeaways:

- The dataset contains 541,909 rows and 8 columns
- The dataset contains 38 unique countries.
- Unique Countries are: United Kingdom, France, Australia, Netherlands, Germany, Norway, EIRE, Switzerland, Spain, Poland, Portugal, Italy, Belgium, Lithuania, Japan, Iceland, Channel Islands, Denmark, Cyprus, Sweden, Austria, Israel, Finland, Bahrain, Greece, Hong Kong, Singapore, Lebanon, United Arab Emirates, Saudi Arabia, Czech Republic, Canada, Unspecified, Brazil, USA, European Community, Malta, RSA

*Missing Values Summary Table:*

| Column Name | Missing Values (%) |
|---|---|
| CustomerID | 24.93 |
| Description | 0.27 |
| StockCode | 0.0 |
| InvoiceNo | 0.0 |
| Quantity | 0.0 |
| InvoiceDate | 0.0 |

|            |       |
| ---------- | ----- |
| UnitPrice  | 0.0   |
| Country    | 0.0   |

**\*\*Above report provides a raw overview of the dataset and has not been fully cleaned or finalized.**

*The following visualizations have been generated after applying key data cleaning steps. Specifically, all records with null values were removed,and only transactions with positive quantities and unit prices (i.e., Quantity > 0 and UnitPrice > 0) were retained to ensure data accuracy.*

## Top 5 Countries by Record Count:
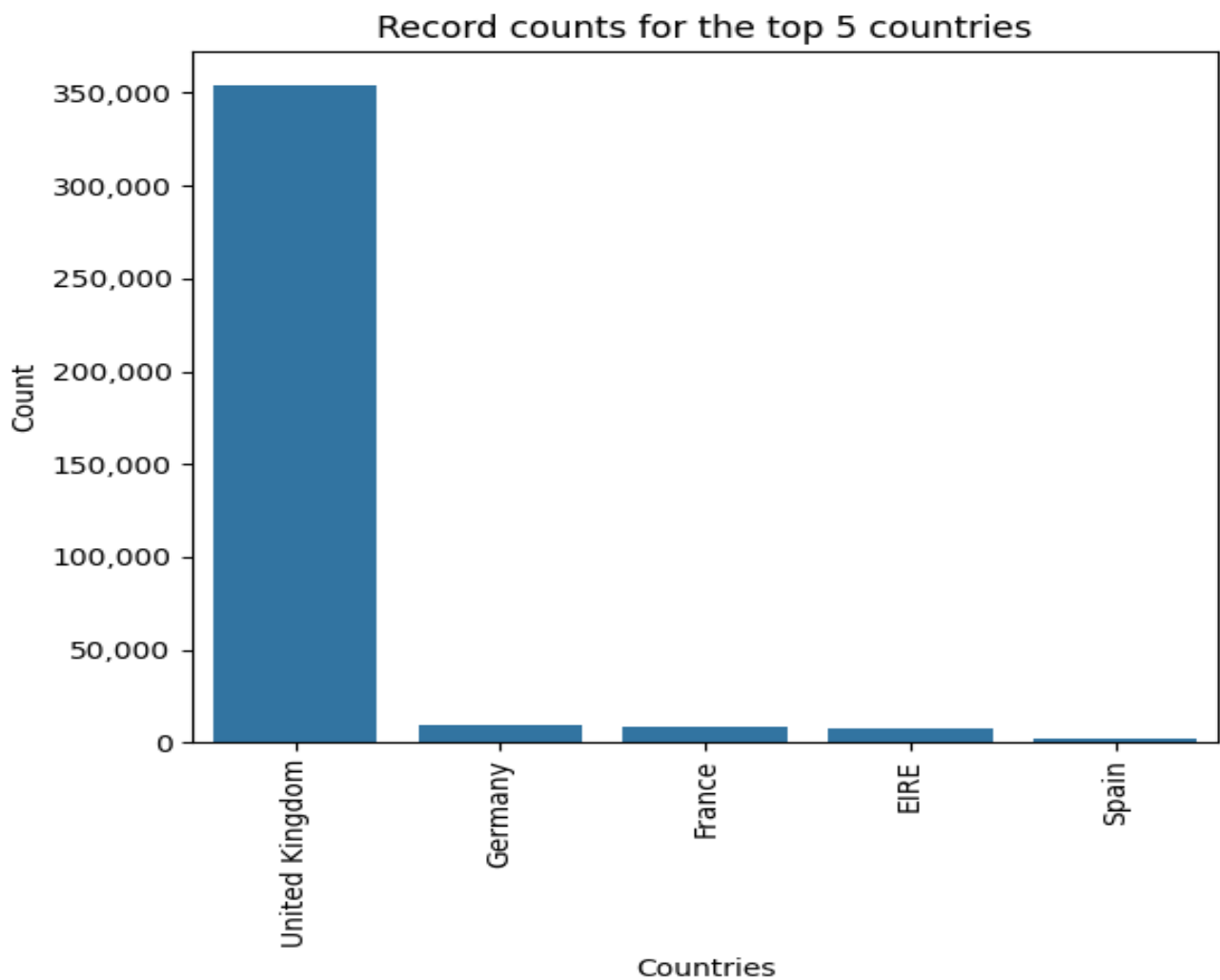


**Analysis:** Most transactions occurred in the United Kingdom, followed by Germany and other countries.

**Total Quantaties Sold Per Year:**

Total Quantity sold per Year
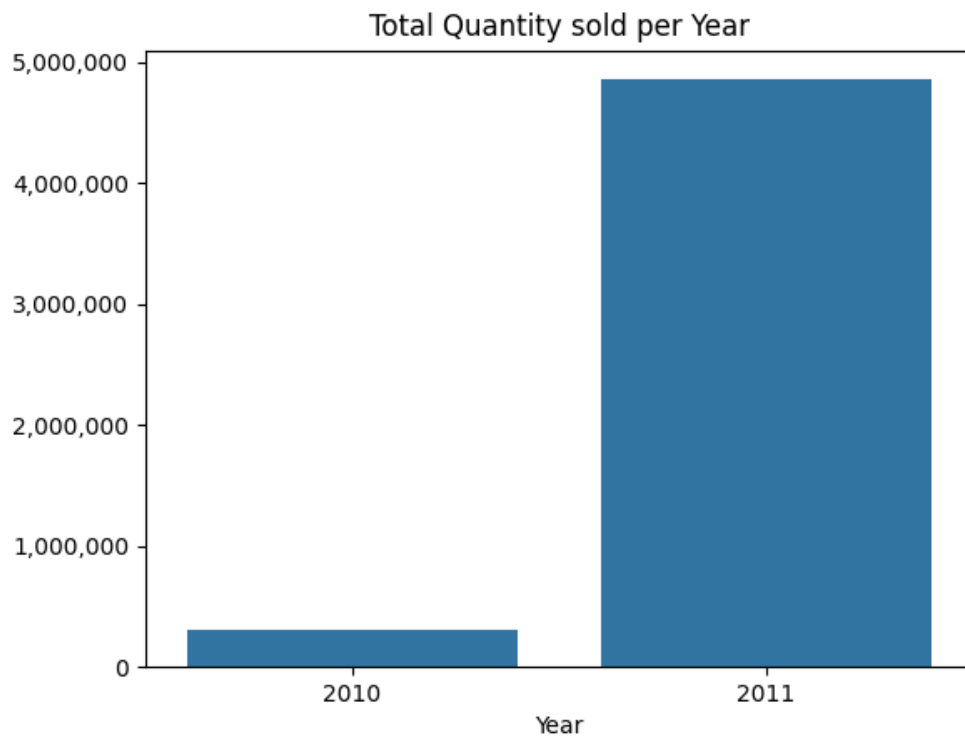


**Analysis:** The majority of transactions occurred in 2011, showing a significant increase compared to 2010. This suggests a rising customer base and growing business activity during that period, possibly due to seasonal demand and improved online retail presence. The trend highlights the importance of maintaining inventory and marketing strategies to meet peak demand periods effectively.

# Monthly Sales Quantities (2011):
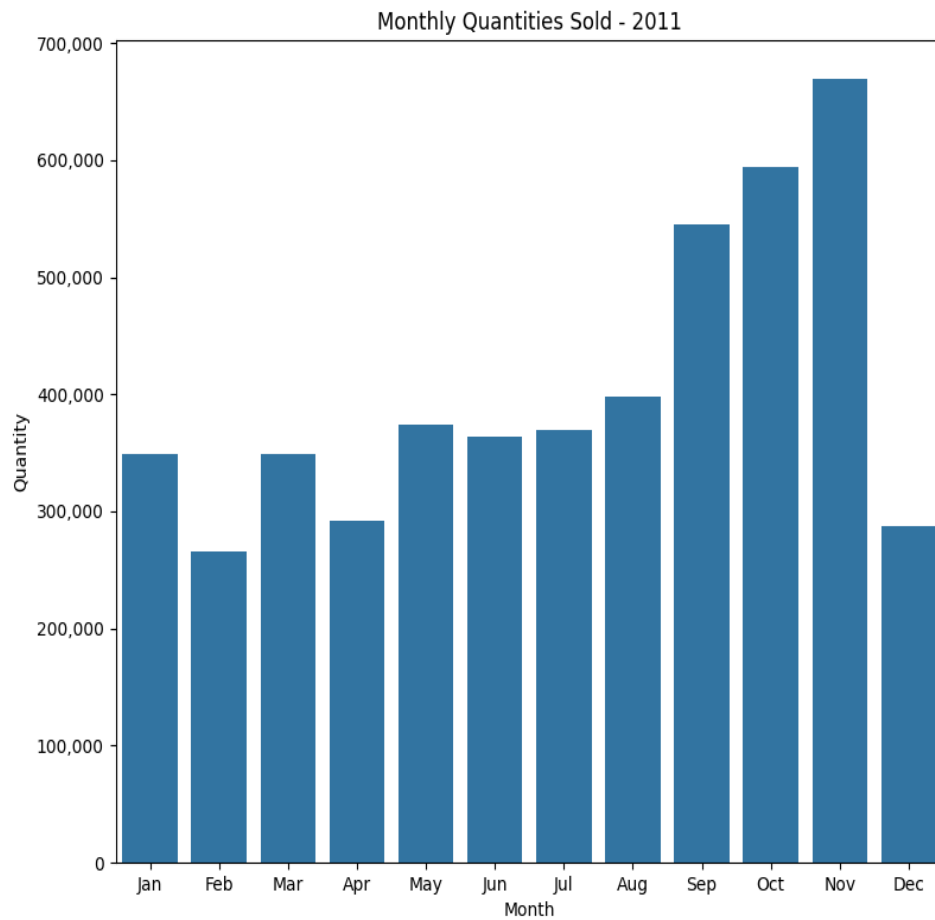


**Analysis:** Most quantities were sold between September and November of 2011, indicating a seasonal surge in product demand. This peak may be driven by holiday shopping behavior, promotional campaigns, or bulk purchasing by wholesalers. Recognizing such patterns is essential for aligning inventory planning, sales strategies, and customer targeting during high-demand periods.

# Correlation between Quantity & Unit Price (For Sep,Oct,Nov):
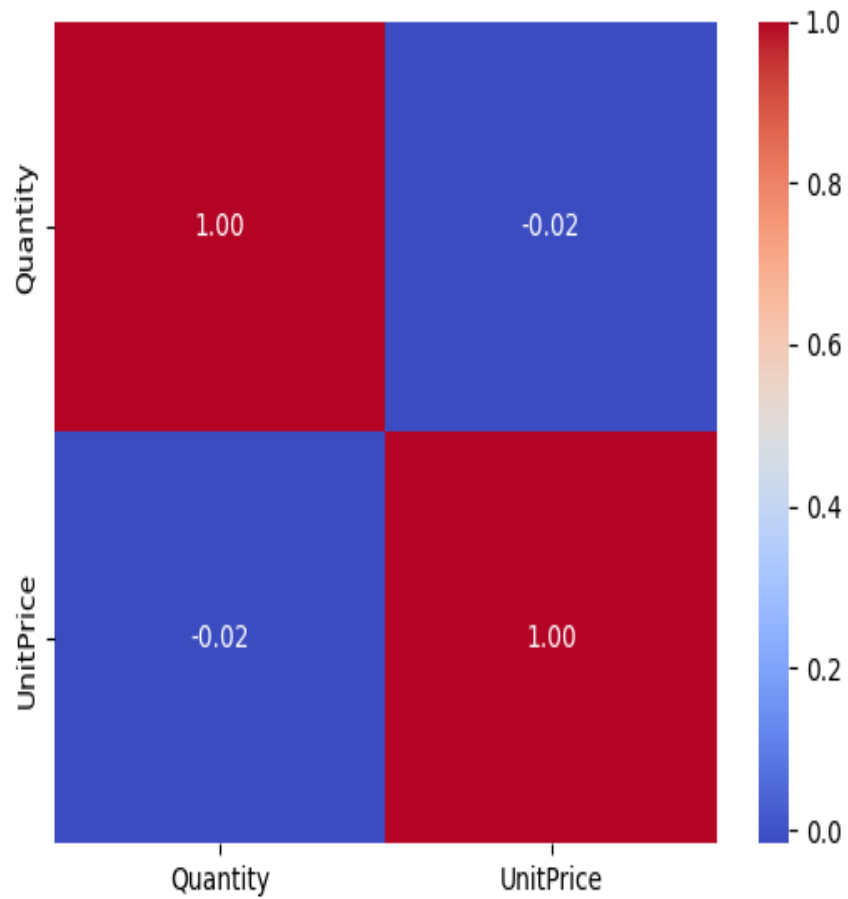


**Analysis:** The correlation analysis between Quantity and UnitPrice for the months of September, October, and November revealed a negative relationship. This suggests that lower-priced items were typically purchased in higher quantities, which is common in wholesale or promotional buying behavior. Understanding this inverse correlation can help optimize pricing strategies, especially during high-volume sales periods.

# Highest Purchasing Customers (by Quantity):
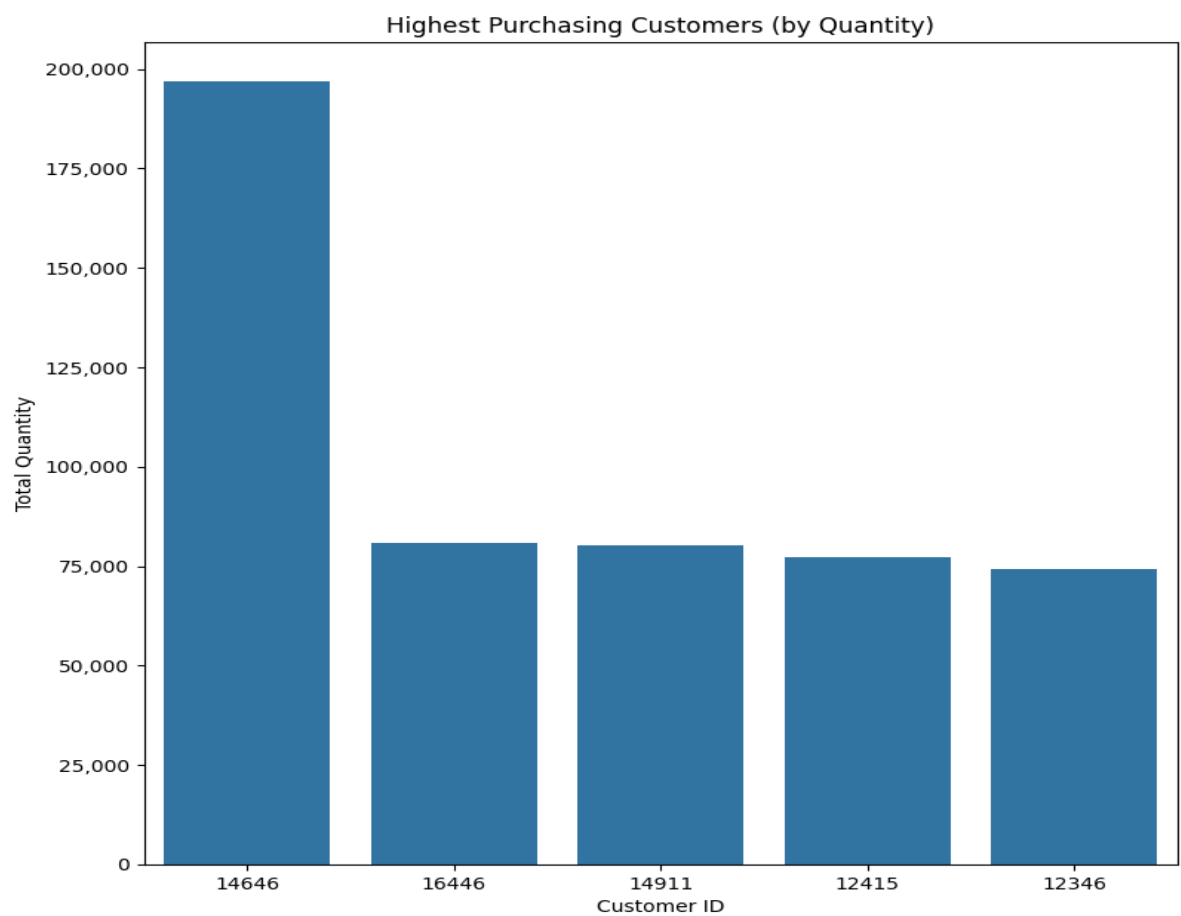


**Analysis:** The analysis of top purchasing customers by quantity reveals that a small subset of customers contributes a disproportionately large share of total sales volume. These high-volume buyers are likely to be wholesale or repeat customers. Identifying and nurturing such key accounts can help strengthen customer loyalty and significantly impact revenue through targeted marketing, personalized offers, or bulk purchase incentives.

# Transactions by Weekday and Time of Day:
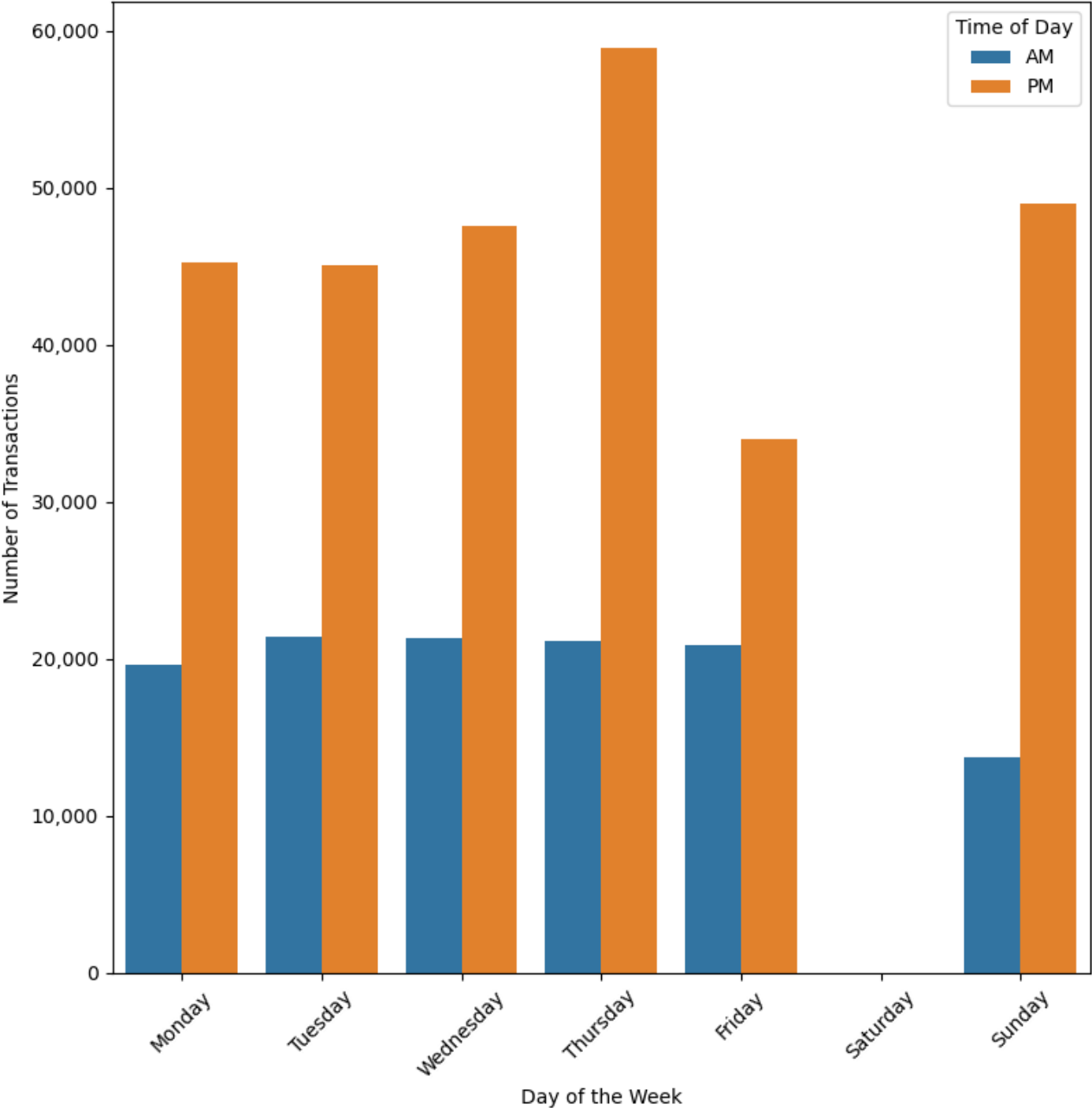


**Analysis:** Most purchases occurred on Thursday afternoons, indicating a mid-to-late week surge in customer activity. This trend may reflect business-driven buying cycles, as many wholesale customers place bulk orders before the weekend. The consistent spike on Thursdays could also be aligned with marketing campaigns or restocking behaviors. Understanding this peak period enables better planning for inventory management, customer support staffing, and targeted promotions to

capitalize on high-engagement time windows. Additionally, scheduling system maintenance or data updates outside of these peak hours could ensure seamless customer experience.