

ABSTRACT

This project presents a comprehensive analysis of coffee sales data collected from a vending machine to uncover actionable insights into customer purchasing patterns, peak sales periods, and product performance. The dataset spans from March to July 2024, capturing detailed transactional information such as purchase date and time, payment method, product type, and sales amount. The primary objective is to use exploratory data analysis (EDA) to answer key business questions, including:

- Which coffee products are most popular among customers?
- What are the busiest hours of the day and days of the week for coffee sales?
- How do sales trends evolve month over month?
- Which products generate the highest revenue?
- What are the dominant payment preferences among customers?

The data cleaning process involved multiple stages to ensure high-quality, reliable data for analysis. Missing values in the card column were identified as cash transactions and retained for payment preference analysis. Date and time fields were converted into appropriate datetime formats, enabling the extraction of additional time-based features such as month, day of the week, and hour of purchase. Categorical fields like coffee_name were standardized to ensure consistency in analysis. Duplicate entries were checked and confirmed absent, while unnecessary columns such as card (due to anonymization) were removed for clarity.

The EDA revealed several important insights. **Americano with Milk** and **Latte** emerged as the top-selling coffee types, together accounting for almost half of all transactions. While Latte was not the highest in transaction count, it generated the most revenue, indicating a higher price point per sale. Hourly sales patterns showed two distinct peaks: one around **10:00 AM** (morning caffeine rush) and another at **7:00 PM** (evening preference). Weekly sales analysis revealed that **Tuesdays** experienced the highest transaction volume, with weekends showing slightly lower activity compared to weekdays. A steady upward trend was observed in monthly sales, particularly for Americano with Milk and Latte, suggesting growing customer engagement with the vending machine over the five-month period.

The payment method analysis showed that more than **92%** of all purchases were made using a card, reflecting a strong preference for cashless transactions. This information is valuable for future payment system optimization and promotional campaigns. Overall, the findings from this project provide clear guidance for inventory management, product stocking strategies, and marketing efforts. These insights can also serve as a solid foundation for predictive modeling to forecast future sales trends and customer behavior, enabling data-driven decision-making for improved business outcomes.

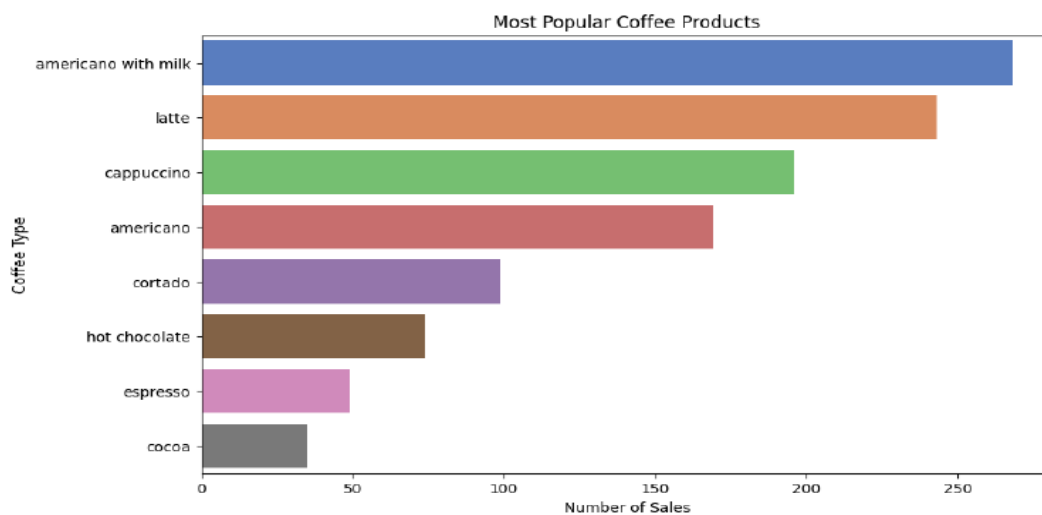
1) Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a critical step in the data analytics process that involves summarizing the main characteristics of a dataset and uncovering underlying patterns, relationships, and trends. In this project, EDA was performed to transform raw coffee sales transaction data into actionable insights. The analysis began with understanding the structure of the dataset, examining data types, identifying missing values, and checking for duplicates. Data cleaning steps included converting date and time fields into proper datetime formats, standardizing categorical values such as coffee product names, and engineering new features like purchase hour, day of the week, and month. These enhancements allowed for deeper time-based trend analysis and segmentation of sales patterns.

The visual exploration phase involved creating plots and charts to reveal customer purchasing behavior and product performance. Count plots were used to determine the most popular coffee types, while bar and line charts displayed revenue contributions, hourly and weekly sales patterns, and monthly growth trends. Payment method preferences were also visualized to assess the proportion of card versus cash transactions. The EDA revealed that Americano with Milk and Latte dominated sales volumes, Latte generated the highest revenue, peak demand occurred around 10 AM and 7 PM, Tuesdays recorded the highest weekly sales, and over 92% of purchases were made using cards. These findings provide a clear foundation for optimizing inventory management, scheduling refills, and designing targeted marketing strategies.

1.1 Most Popular Coffee Products

- **Which coffee types are most frequently purchased?**
- Americano with Milk and Latte are the top-selling coffee types, together accounting for nearly half of all transactions. Cappuccino and Americano follow as the second tier of popular choices. Less preferred drinks include Cortado, Hot Chocolate, Espresso, and Cocoa.

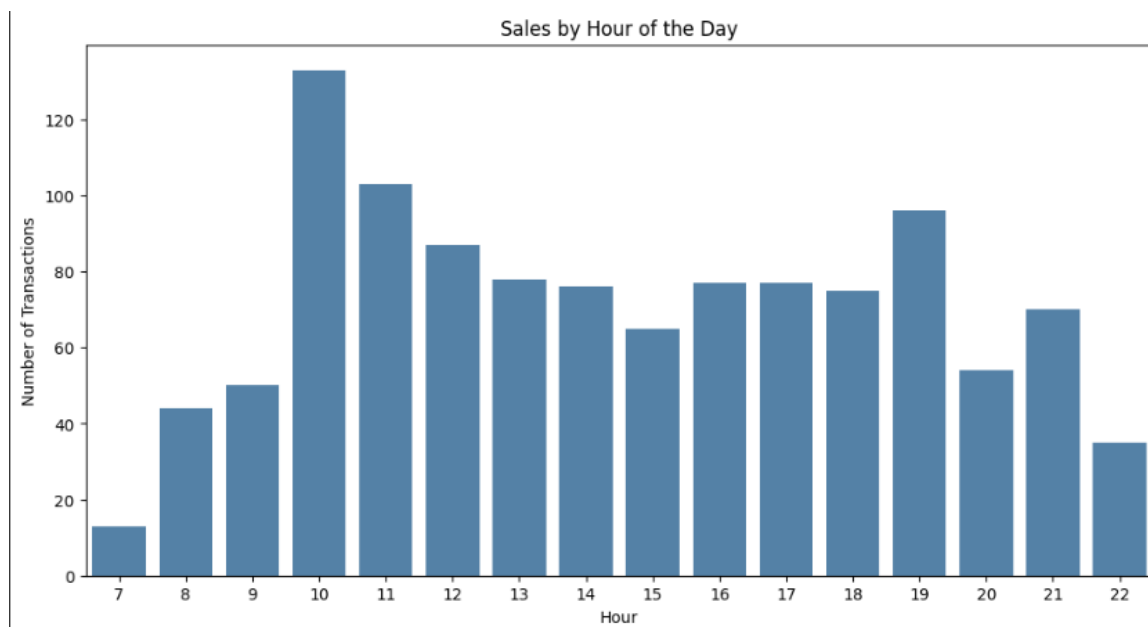


The bar chart illustrates the most popular coffee products sold by the vending machine, ranked by the number of sales. **Americano with Milk** leads as the top-selling beverage, followed closely by

Latte, with both products significantly outperforming all other options and together accounting for a large share of total transactions. **Cappuccino** and **Americano** form the second tier of popular drinks, while **Cortado** and **Hot Chocolate** show moderate sales levels. **Espresso** and **Cocoa** are the least popular choices, with comparatively low purchase volumes. This distribution highlights a strong customer preference for milk-based coffees, suggesting potential areas for focused marketing and inventory prioritization.

1.2 Hourly Sales Trend

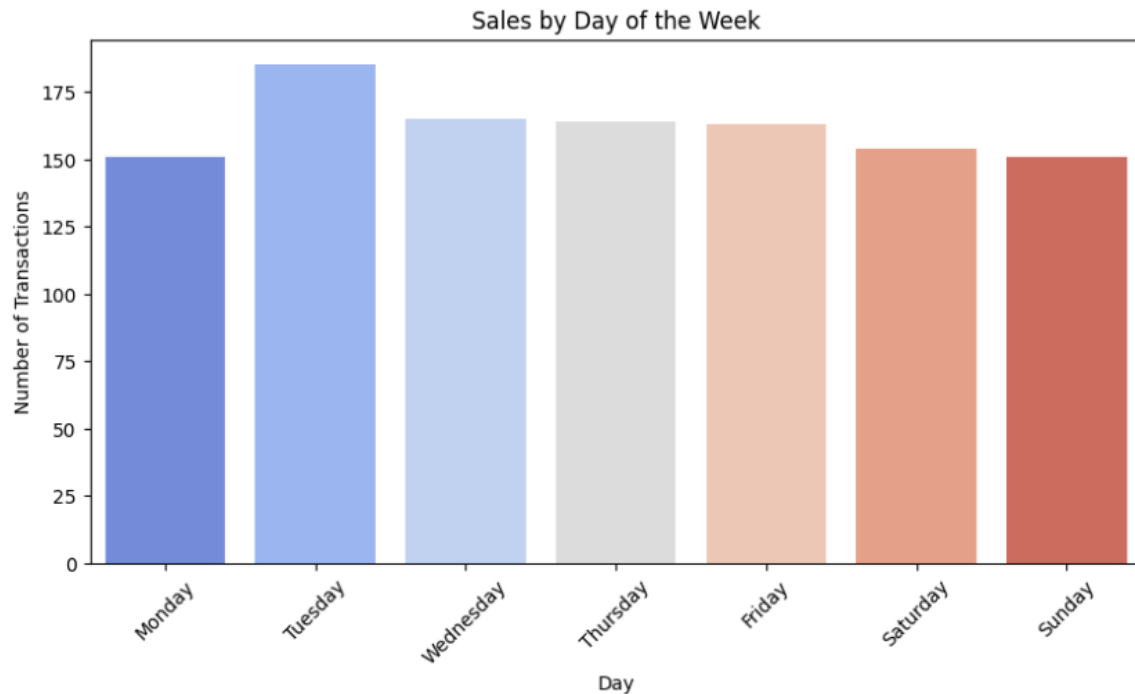
- When do most purchases happen during the day?
- Sales peak around 10 AM and again at 7 PM, suggesting two key coffee demand windows: a morning caffeine rush and an evening wind-down. These trends are consistent across most coffee products.



The bar chart shows the distribution of coffee sales by hour of the day, revealing clear peaks in customer activity. The most significant surge in transactions occurs at **10:00 AM**, indicating a strong morning coffee rush, followed by a smaller but notable peak around **7:00 PM**, suggesting an evening preference among customers. Sales gradually increase from early morning, reaching the highest point mid-morning before tapering off into the afternoon. Evening sales pick up again before declining towards late night. These patterns highlight two key demand windows that could guide staffing, inventory management, and promotional timing.

1.3 Weekday Sales Trend

- Which days of the week are busiest?
- Tuesday records the highest number of transactions, indicating a possible midweek coffee surge. Other weekdays have relatively similar volumes, while weekends show slightly lower activity.

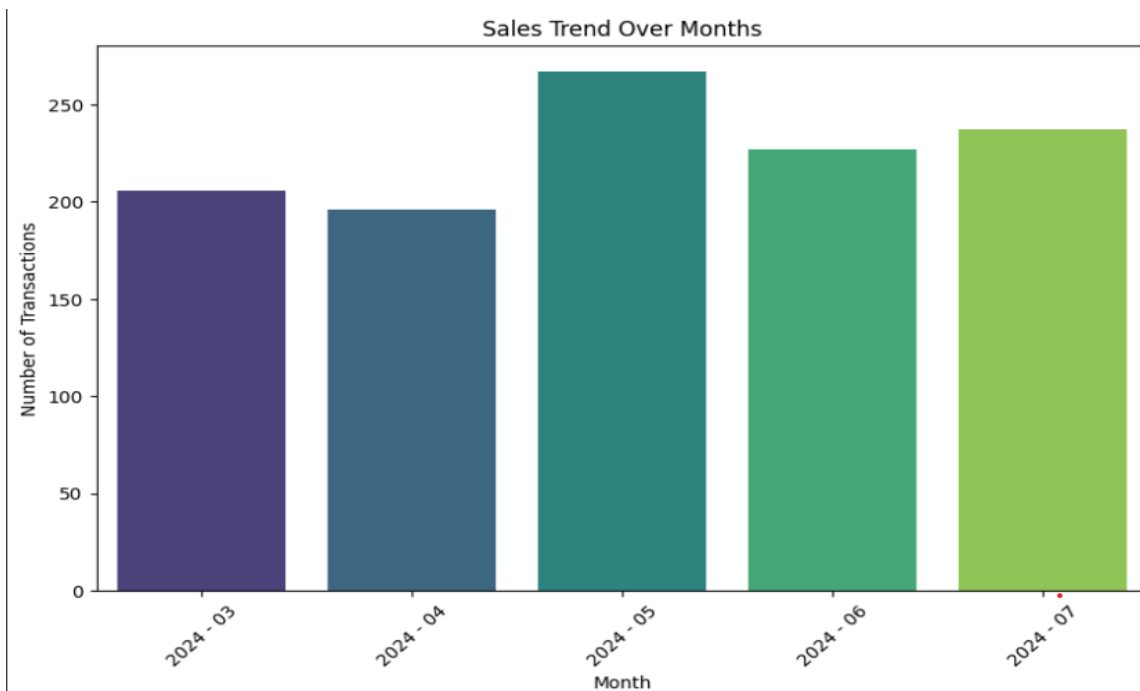


The bar chart illustrates coffee sales distribution across the days of the week, showing that **Tuesday** records the highest number of transactions, suggesting a midweek surge in demand. Sales remain relatively consistent from Wednesday to Friday, indicating steady weekday performance. **Monday**, **Saturday**, and **Sunday** see slightly lower transaction counts, with weekends experiencing a small dip compared to weekdays. This pattern suggests that coffee consumption is strongest during the workweek, particularly early to midweek, which could inform targeted promotions or inventory planning for these peak periods.

1.4 Monthly Sales Trend

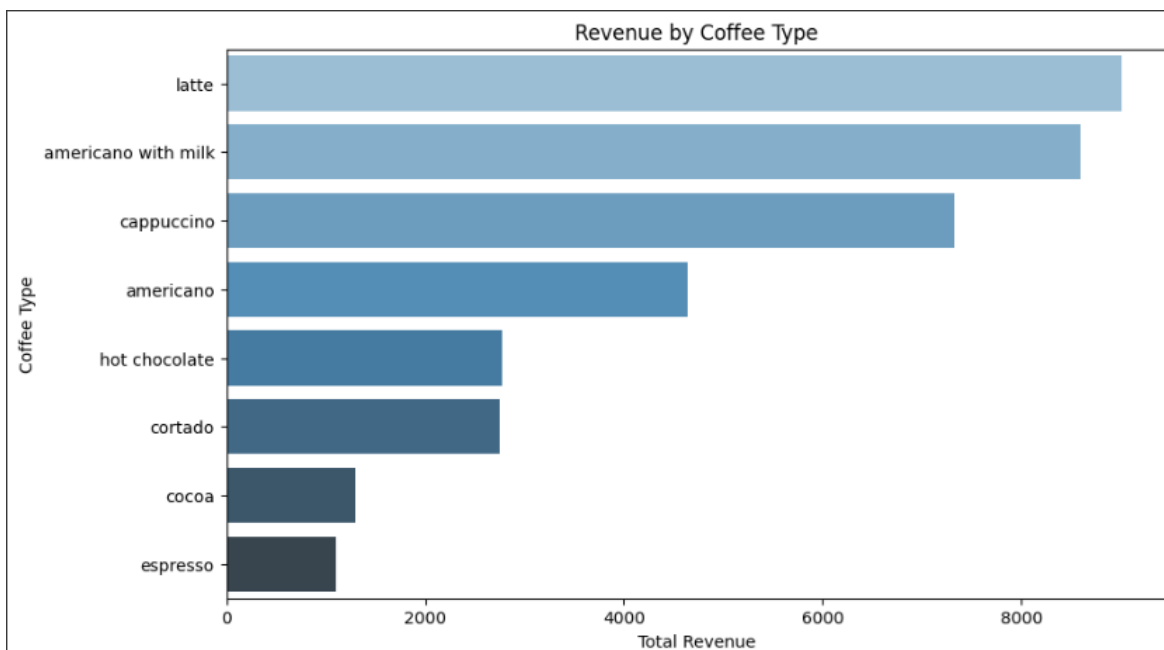
- How has sales evolved over time?
- From March to July 2024, there's an overall upward trend in sales, with Americano with Milk and Latte showing steady growth. This suggests increasing popularity or vending machine usage over time.

The bar chart presents the monthly sales trend from March to July 2024, revealing fluctuations in transaction volumes over the five-month period. Sales dipped slightly from **March** to **April**, followed by a significant surge in **May**, which recorded the highest number of transactions during the observed period. Although sales decreased in **June**, they remained higher than the early months, and **July** saw a slight recovery. This pattern suggests that May experienced exceptional demand possibly due to seasonal factors or promotions while overall, the trend indicates growing engagement with the vending machine compared to the start of the dataset.



1.5 Revenue by Product

- Which coffee types generate the highest revenue?
- Latte contributes the highest total revenue, despite not having the most transactions indicating a relatively higher price point. Americano with Milk also performs well in terms of revenue, while Espresso and Cocoa contribute the least.

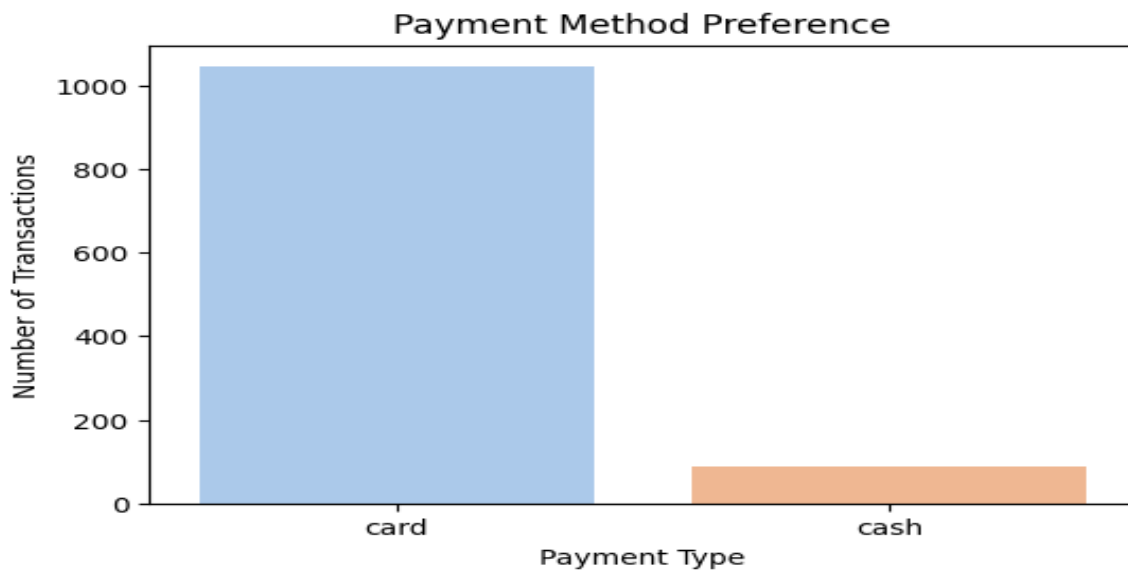


The bar chart illustrates the total revenue generated by each coffee type, highlighting **Latte** as the top revenue contributor, followed closely by **Americano with Milk**. Despite not having the

highest transaction count, Latte's leading position suggests a higher average price per sale. **Cappuccino** and **Americano** also make substantial contributions, while **Hot Chocolate** and **Cortado** generate moderate revenue levels. **Cocoa** and **Espresso** are the lowest earners, reflecting their smaller sales volumes. This revenue distribution indicates that premium-priced or consistently popular items like Latte and Americano with Milk play a critical role in driving overall sales income, making them strategic products to prioritize in inventory and promotions.

1.6 Payment Method Analysis

- Do customers prefer card or cash?
- Over 92% of transactions are made using card, highlighting a strong preference for digital payments. All missing card values correspond to cash transactions, which account for only ~8% of sales.



The bar chart displays customer payment preferences, showing a strong dominance of **card transactions**, which account for the vast majority of purchases. In contrast, **cash payments** make up only a small fraction of total sales. This significant skew toward cashless payments highlights a clear customer preference for digital transactions, suggesting that maintaining efficient card payment systems is essential for smooth operations and customer satisfaction.

2) Conclusion And Future Scope

The comprehensive analysis of coffee sales data from March to July 2024 has offered a detailed understanding of consumer preferences, peak purchasing times, and revenue contributions across different coffee products. The results highlight that Americano with Milk and Latte dominate the market in terms of sales volume, with Latte emerging as the highest revenue generator, reflecting its strong customer appeal and higher price point. Hourly sales trends show a significant morning peak at 10:00 AM and a secondary evening peak at 7:00 PM, which can be strategically targeted for restocking and promotions. Tuesdays stand out as the busiest day of the week, while weekend sales slightly decline, suggesting differences in weekday and weekend coffee consumption habits. Monthly sales analysis reveals a remarkable surge in May, possibly driven by seasonal demand or promotional activities, followed by a steady trend. Payment method analysis shows that over 92% of transactions are card-based, confirming a shift toward cashless transactions and underscoring the need to maintain seamless digital payment systems.

For the **future scope**, the project can evolve by integrating advanced analytics and machine learning techniques. Time series forecasting models like ARIMA, Prophet, or LSTM could be employed to predict sales trends, enabling proactive inventory and supply chain management. Including external data such as weather patterns, local events, holidays, and seasonal trends could uncover additional demand drivers. With enhanced customer-level data, segmentation and clustering techniques could identify distinct customer profiles, allowing for personalized offers and loyalty programs. Expanding the dataset to include multiple vending machine locations would enable comparative performance analysis and regional strategy customization. Additionally, real-time analytics dashboards could be developed for continuous monitoring, helping businesses respond quickly to fluctuations in demand. Over time, coupling these insights with automated inventory management and targeted marketing campaigns can lead to significant cost optimization, increased customer satisfaction, and sustained revenue growth.