Data Visualization and Predictive Analytics on Coffee Sales Dataset

Project Overview

- **Goal:**
  Analyze coffee shop sales to uncover buying patterns, peak hours, and forecast future sales using visualization and machine learning.
  Dataset:
- Source: Kaggle - Coffee Sales Dataset
- Records: 3,547
- Features: 11 (Date, Time, Coffee Type, Money, Weekday, Month, etc.)
- No missing or duplicate data

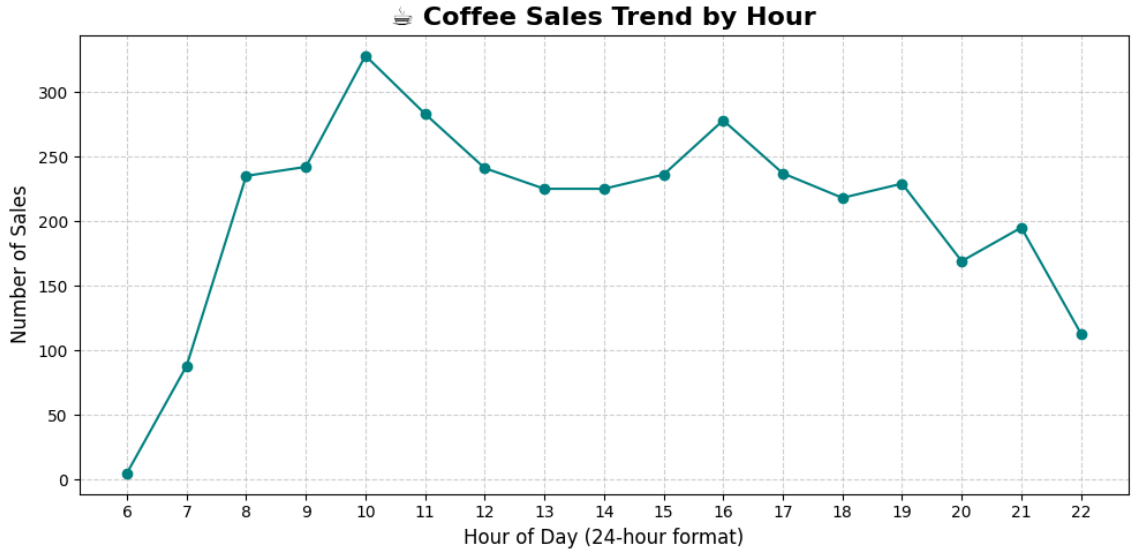
Exploratory Data Insights

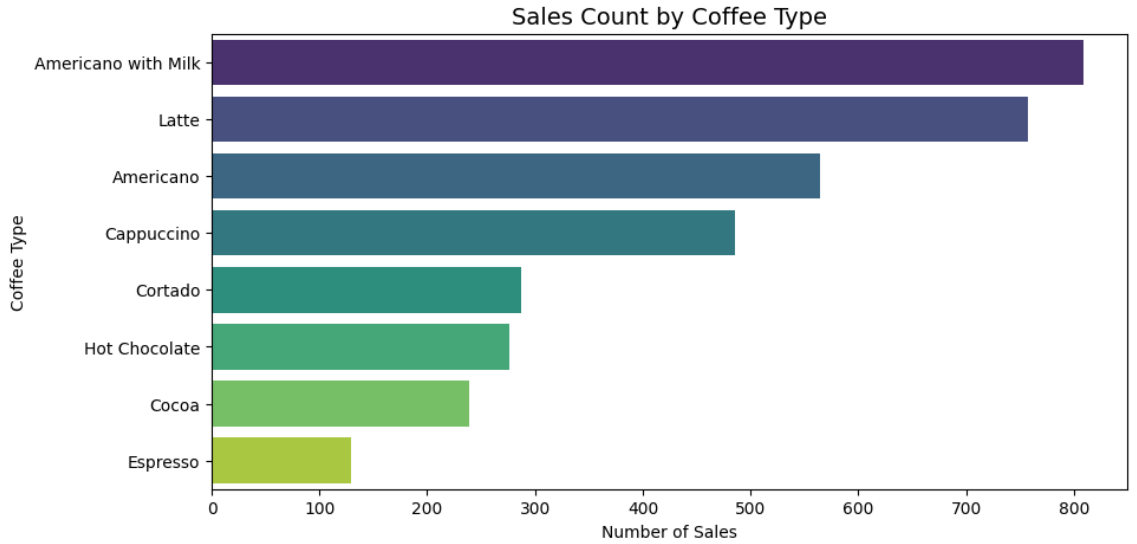
✨ **Data Overview:** Reviewed dataset structure, column names, and record count.
🔍 **Missing Values Check:** Verified dataset had no null or missing entries.
🗓️ **Date Exploration:** Converted and analyzed date column for daily/monthly trends.
💰 **Sales Distribution:** Calculated total money earned per coffee type.
🧾 **Top Products:** Identified highest revenue-generating coffee varieties.
📈 **Customer Patterns:** Observed most frequently purchased coffee types.
📊 **Data Grouping:** Grouped by coffee name to summarize overall sales performance.
💡 **Key Insight:** Certain coffee types consistently dominated both sales count and revenue.

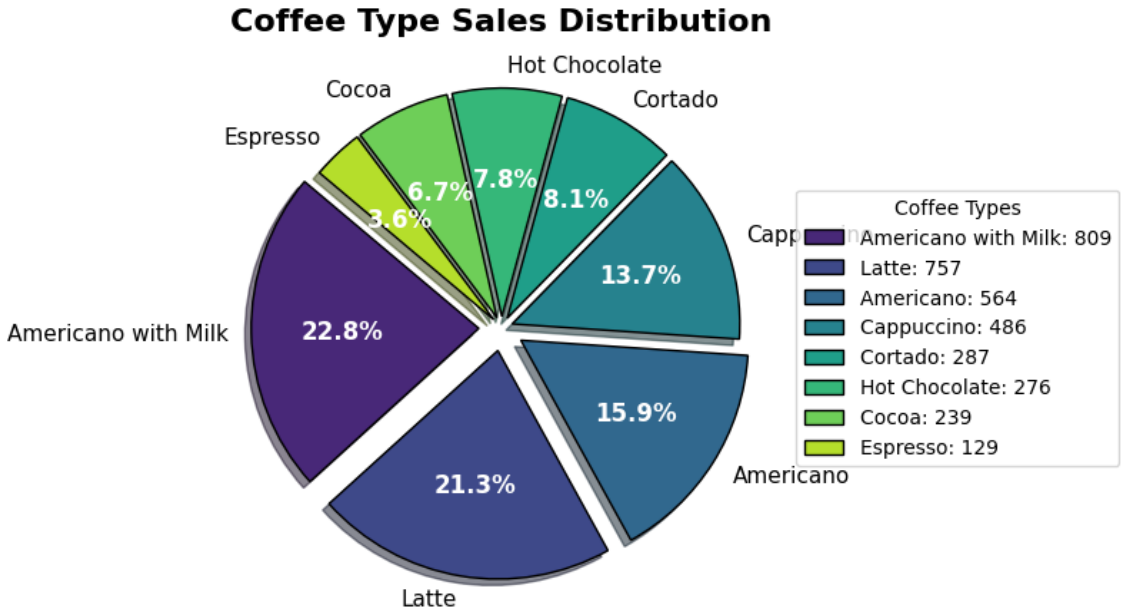
EDA helped us understand daily, weekly, and monthly sales behavior before modeling.

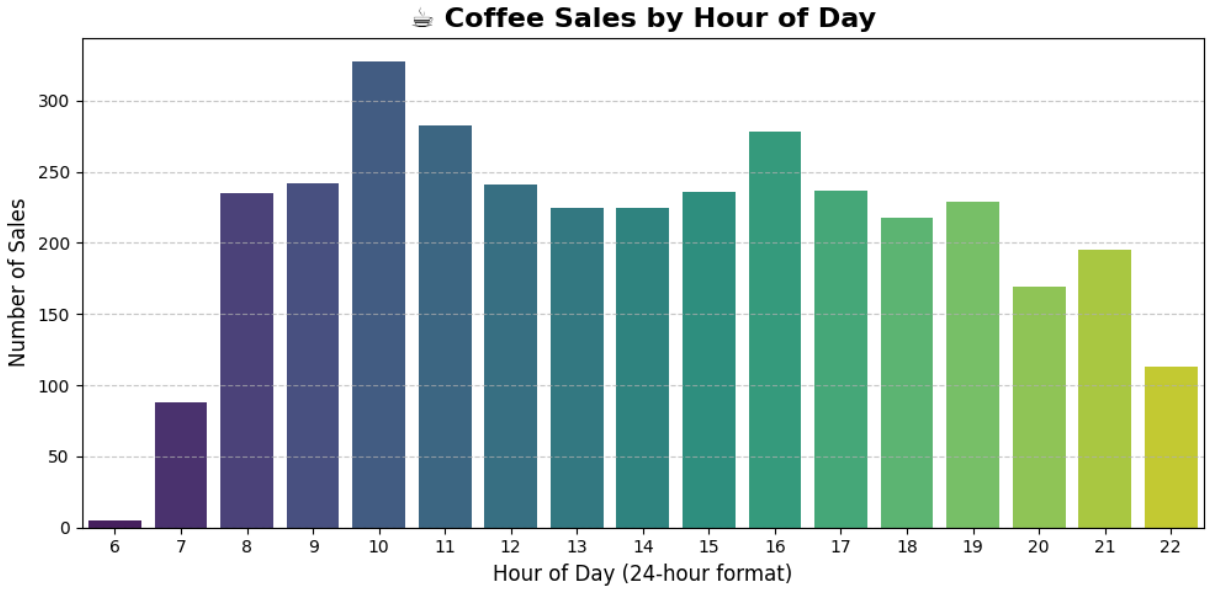
Visual Analysis

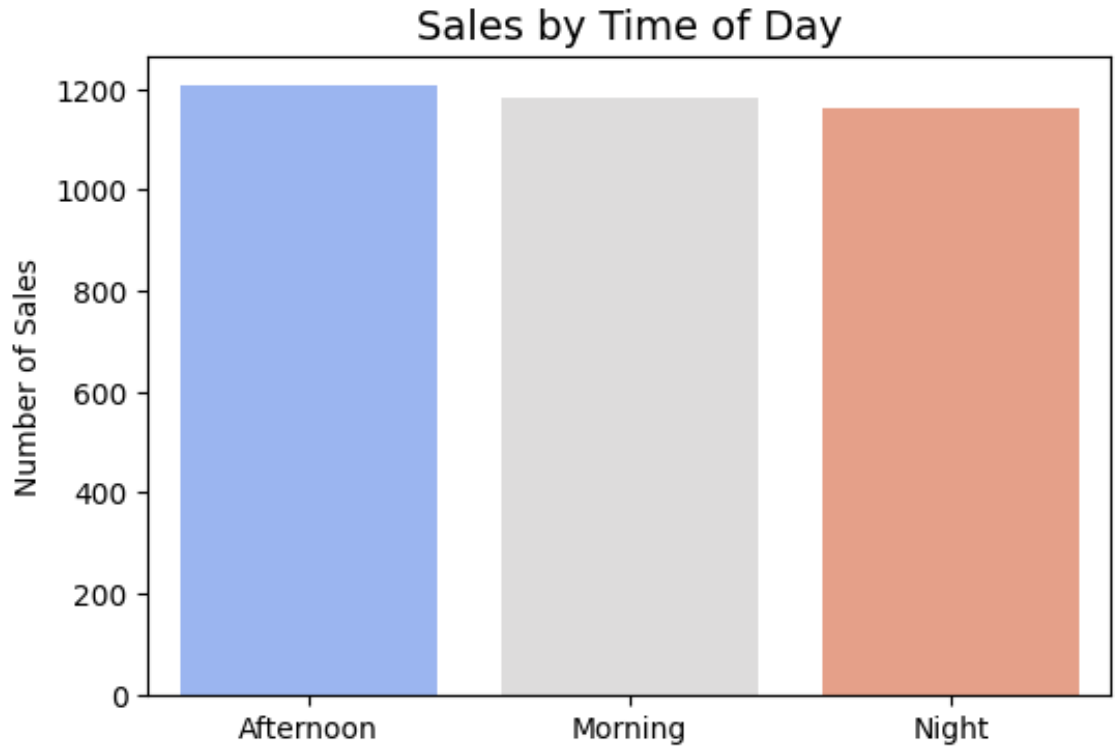
- 📊 **Sales by Coffee Type:** Bar and count plots showing top-selling coffee varieties.
  💸 **Revenue by Coffee Type:** Horizontal bar chart displaying total money earned per product.
  🕑 **Hourly Sales Trends:** Line chart revealing peak coffee sales hours.
  🗓️ **Daily & Monthly Earnings:** Aggregated plots highlighting revenue patterns across time.
  🌎 **Store Performance:** Visuals comparing top-performing store locations.
  📈 **Correlation Heatmap:** Illustrated relationships between numerical variables (e.g., money, quantity, hour).
  🤖 **Prediction Visualization:** Scatter plots and regression lines from linear regression model results.

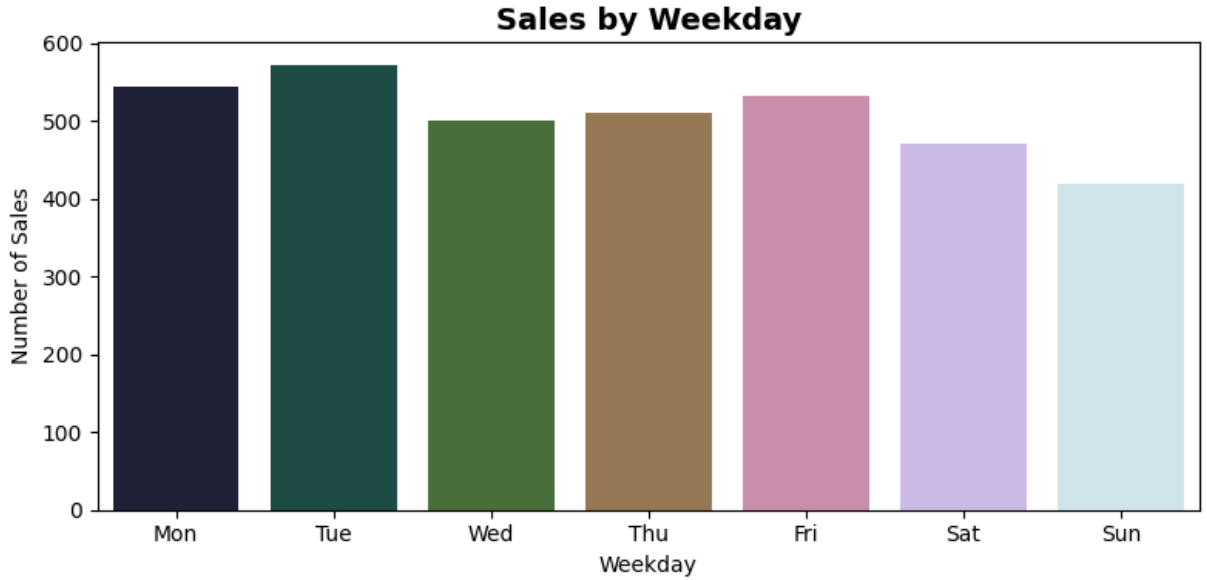

☕ Coffee Sales Trend by Hour
Number of Sales
Hour of Day (24-hour format)
0
50
100
150
200
250
300
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

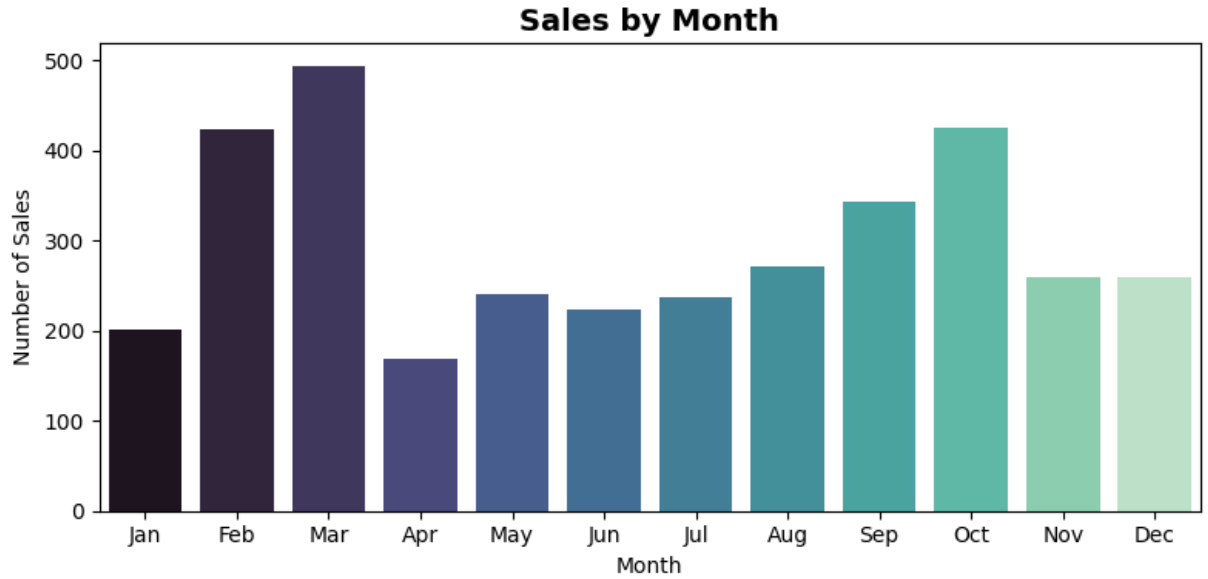

Sales Count by Coffee Type
Coffee Type
Americano with Milk
Latte
Americano
Cappuccino
Cortado
Hot Chocolate
Cocoa
Espresso
0
100
200
300
400
500
600
700
800
Number of Sales

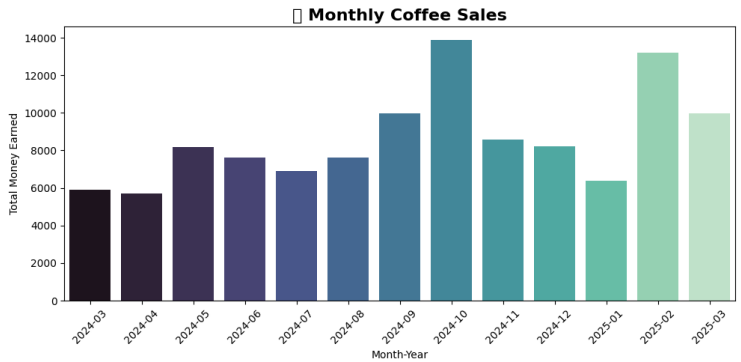

Coffee Type Sales Distribution
Hot Chocolate
Cocoa
Cortado
Espresso
Americano with Milk
Latte
Americano
Capp
22.8%
21.3%
15.9%
13.7%
8.1%
7.8%
6.7%
3.6%
Coffee Types
Americano with Milk: 809
Latte: 757
Americano: 564
Cappuccino: 486
Cortado: 287
Hot Chocolate: 276
Cocoa: 239
Espresso: 129

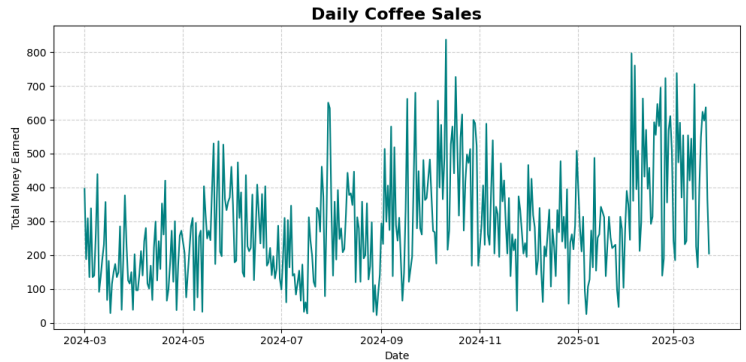

☕ Coffee Sales by Hour of Day
Number of Sales
0
50
100
150
200
250
300
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Hour of Day (24-hour format)

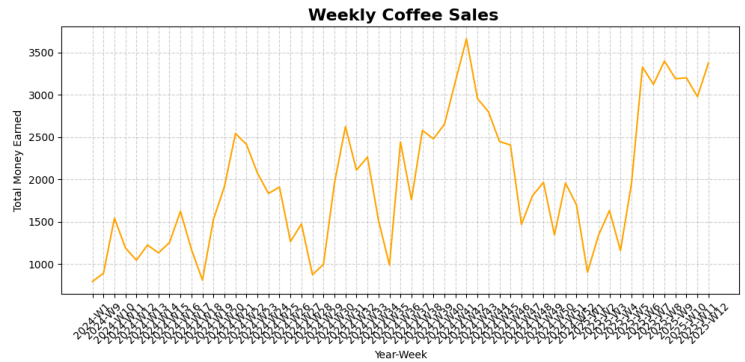

Sales by Time of Day
Number of Sales
0
200
400
600
800
1000
1200
Afternoon
Morning
Night

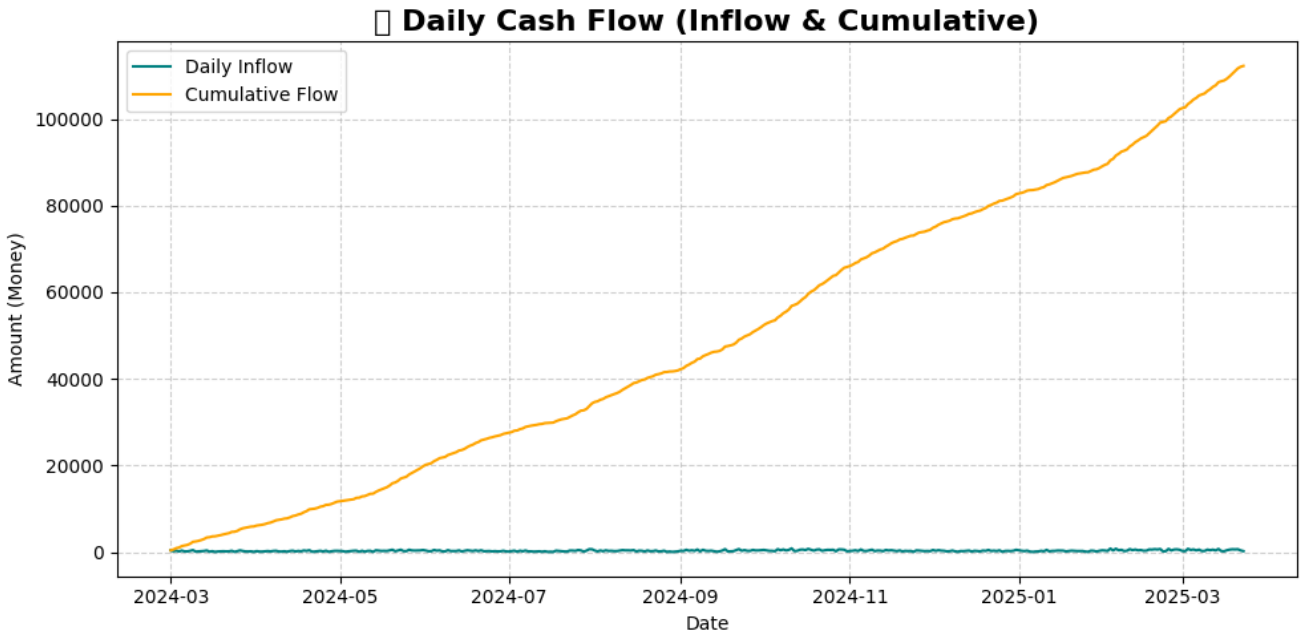

Sales by Weekday
Number of Sales
0
100
200
300
400
500
600
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Weekday

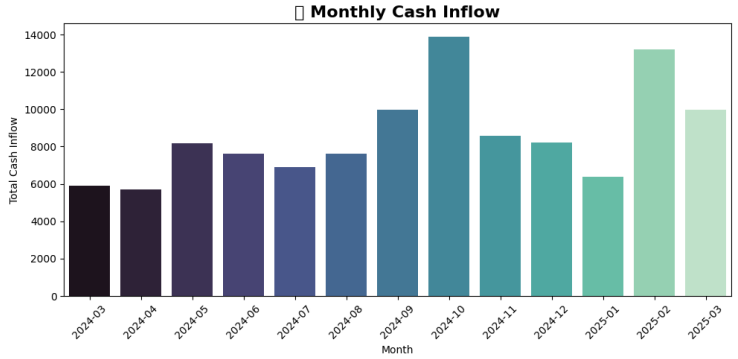

Sales by Month
Number of Sales
0
100
200
300
400
500
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

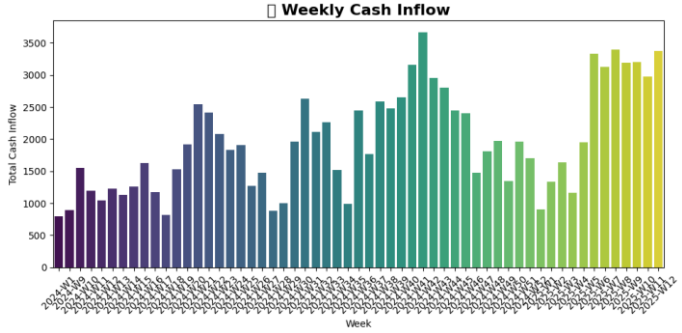

☕ Monthly Coffee Sales
Total Money Earned
Month-Year
2024-03
2024-04
2024-05
2024-06
2024-07
2024-08
2024-09
2024-10
2024-11
2024-12
2025-01
2025-02
2025-03

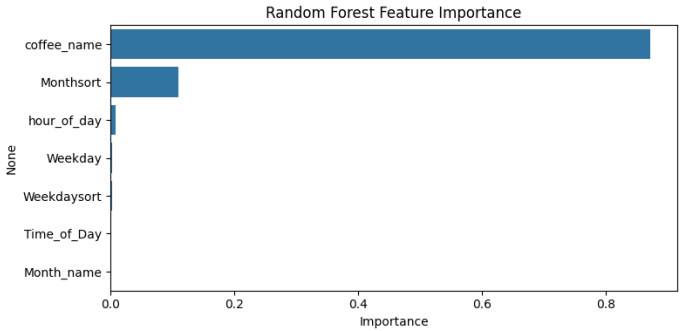

Daily Coffee Sales
Total Money Earned
Date

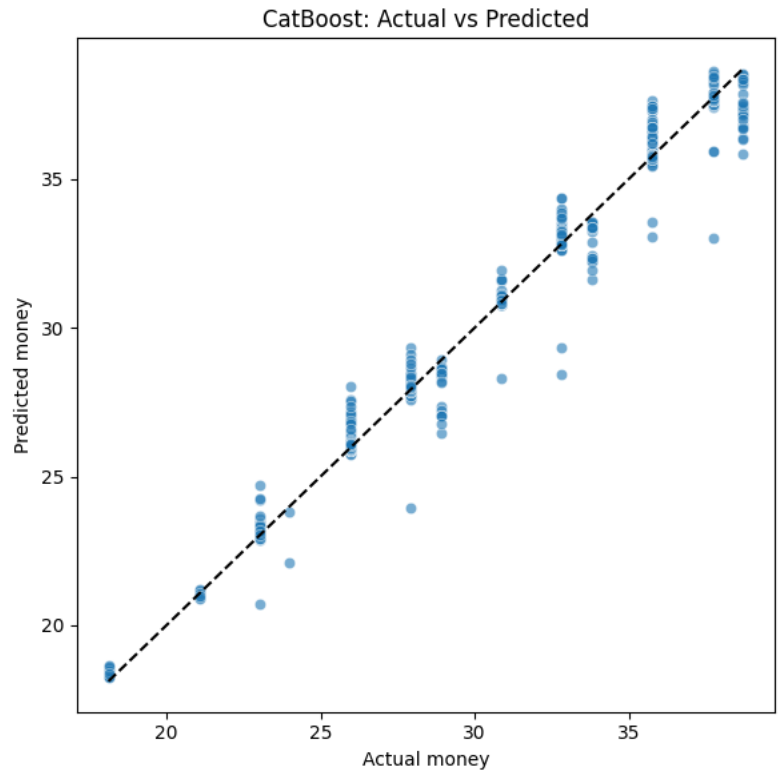

Weekly Coffee Sales
Total Money Earned
Year-Week

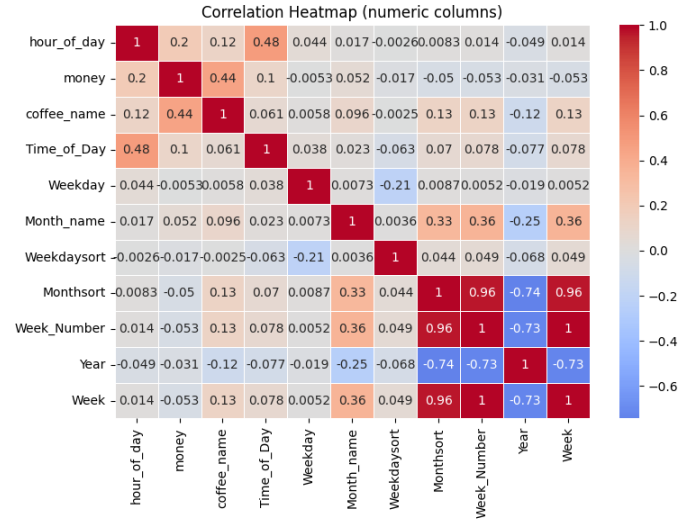

Daily Inflow
Cumulative Flow
100000
80000
60000
40000
20000
0
Amount (Money)
2024-03
2024-05
2024-07
2024-09
2024-11
2025-01
2025-03
Date

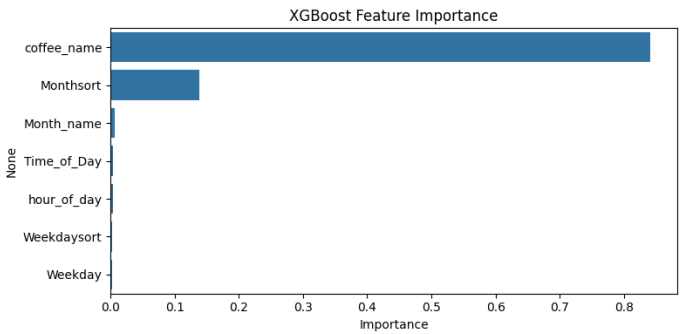

14000
12000
10000
8000
6000
4000
2000
0
Total Cash Inflow
2024-03
2024-04
2024-05
2024-06
2024-07
2024-08
2024-09
2024-10
2024-11
2024-12
2025-01
2025-02
2025-03
Month

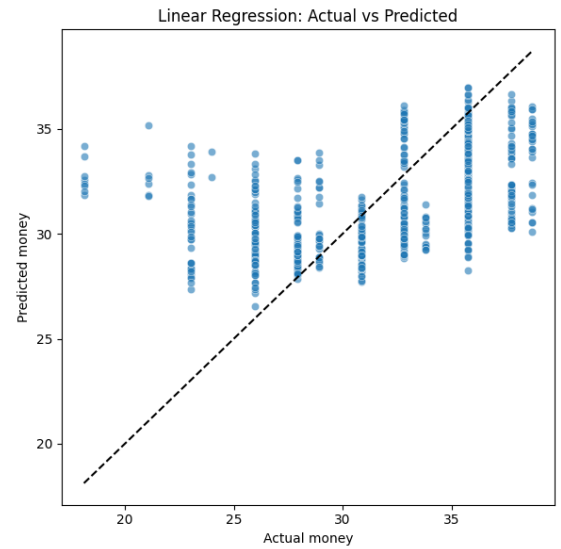

3500
3000
2500
2000
1500
1000
500
0
Total Cash Inflow
Week

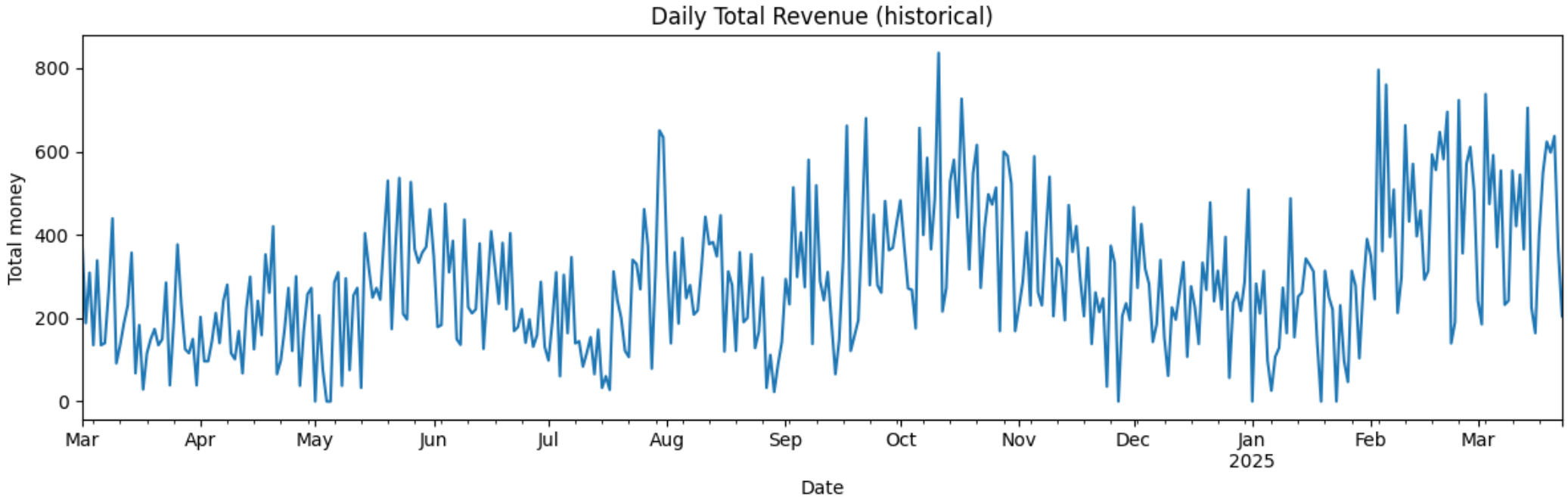

Random Forest Feature Importance
coffee_name
Monthsort
hour_of_day
Weekday
Weekdaysort
Time_of_Day
Month_name
None
0.0
0.2
0.4
0.6
0.8
Importance

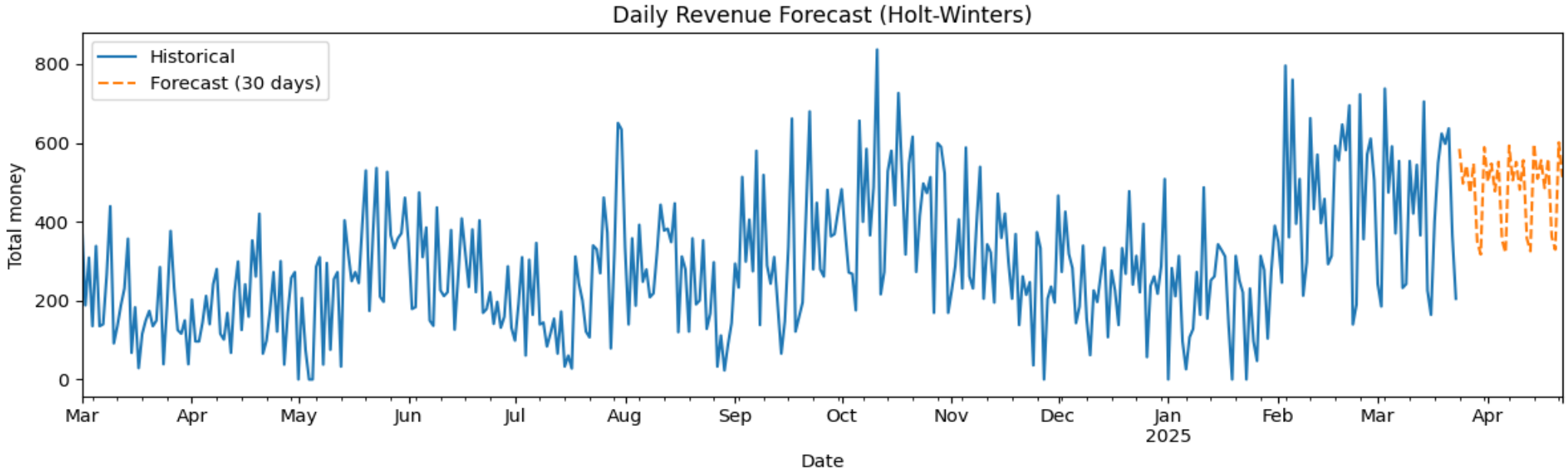

CatBoost: Actual vs Predicted
Predicted money
Actual money
20
25
30
35


Correlation Heatmap (numeric columns)
hour_of_day 1 0.2 0.12 0.48 0.044 0.017 -0.0026 0.0083 0.014 -0.049 0.014
money 0.2 1 0.44 0.1 -0.0053 0.052 -0.017 -0.05 -0.053 -0.031 -0.053
coffee_name 0.12 0.44 1 0.061 0.0058 0.096 -0.0025 0.13 0.13 -0.12 0.13
Time_of_Day 0.48 0.1 0.061 1 0.038 0.023 -0.063 0.07 0.078 -0.077 0.078
Weekday 0.044 -0.0053 0.0058 0.038 1 0.0073 -0.21 0.0087 0.0052 -0.019 0.0052
Month_name 0.017 0.052 0.096 0.023 0.0073 1 0.0036 0.33 0.36 -0.25 0.36
Weekdaysort -0.0026 -0.017 -0.0025 -0.063 -0.21 0.0036 1 0.044 0.049 -0.068 0.049
Monthsort -0.0083 -0.05 0.13 0.07 0.0087 0.33 0.044 1 0.96 -0.74 0.96
Week_Number 0.014 -0.053 0.13 0.078 0.0052 0.36 0.049 0.96 1 -0.73 1
Year -0.049 -0.031 -0.12 -0.077 -0.019 -0.25 -0.068 -0.74 -0.73 1 -0.73
Week 0.014 -0.053 0.13 0.078 0.0052 0.36 0.049 0.96 1 -0.73 1
hour_of_day
money
coffee_name
Time_of_Day
Weekday
Month_name
Weekdaysort
Monthsort
Week_Number
Year
Week
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


XGBoost Feature Importance
coffee_name
Monthsort
Month_name
Time_of_Day
hour_of_day
Weekdaysort
Weekday
None
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Importance


Linear Regression: Actual vs Predicted
Predicted money
Actual money
20
25
30
35


Daily Total Revenue (historical)
Total money
0
200
400
600
800
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2025
Feb
Mar
Date


Daily Revenue Forecast (Holt-Winters)
Historical
Forecast (30 days)
Total money
0
200
400
600
800
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2025
Feb
Mar
Apr
Date

📊 Key Insights at a Glance

- 💰 **Product Trends**
- **Top Sellers:** Bar charts highlight standout coffee types which is an ideal focus for marketing & promotions.
- **High Revenue Items:** Premium coffees drive higher margins despite lower volumes.
- 🕒 **Time-Based Patterns**
- **Peak Hours:** Strong spikes during **morning rush & lunch breaks** so adjust staffing & inventory accordingly.
- **Daily/Monthly Cycles:** Weekday sales dominate; mild seasonal peaks in **winter months**.
- 🏪 **Store & Customer Behavior**
- **Top Locations:** Certain stores consistently outperform others leveraging best practices.
- **Correlation Insights:** Clear link between **quantity sold** & **total revenue**; sales rise predictably with customer flow.
- 🤖 **Predictive Analytics**
- **Model Fit:** Regression visuals confirm accurate predictive modeling.
- **Forecast Power:** Machine learning models effectively capture sales trends and time-based variability.

Model Performance Comparison

Model	MAE ↓	RMSE ↓	R^2 ↑	Interpretation
Linear Regression	3.19	4.09	0.27	Basic model , underfits and fails to capture complexity.
Random Forest	0.26	0.69	0.979	Excellent , effectively models nonlinear relationships.
XGBoost	0.34	0.79	0.973	Great performance, slightly behind Random Forest.
CatBoost	**0.33**	**0.66**	**0.981**	**Best model overall with top accuracy and lowest error.**

📈 Forecasted Revenue (Holt–Winters Model)

- **What the Chart Shows:**
- Historical sales data (🔵 blue line) and a **30-day forecast** (🟠 orange dashed line) for March 2025.
- The model effectively captures seasonal and trend patterns.
- **Key Observations:**
- Forecasted daily revenue fluctuates between **≈350–600 units**.
- No strong upward or downward trend indicates steady-state performance with normal day-to-day variation.
- Predictive pattern mirrors previous seasonal cycles, confirming model stability and reliability.
- **Interpretation:**
- Revenue is expected to remain stable in the short term, with ongoing, predictable fluctuations.
- The store operates within a healthy consistent sales range, a sign of mature demand with steady customer engagement.

Final Takeaways

- **CatBoost** provides the most accurate predictive performance for coffee sales forecasting.
- **Time-of-day** and **coffee type** are the strongest drivers of revenue.
- **Holt–Winters forecasting** supports stable revenue expectations with realistic variability.
- Insights can directly inform **inventory management**, **staff scheduling**, and **marketing timing** for optimized operational efficiency.

Strategic Recommendations

- **Enhance Product Strategy:** Focus marketing and upselling on the top 3 performing coffee varieties.
- **Optimize Operations:** Adjust staff shifts and stock rotation around morning and lunch peaks.
- **Leverage Predictive Models:** Deploy CatBoost for continuous sales forecasting and trend monitoring.
- **Seasonal Planning:** Use Holt–Winters projections to anticipate holiday and seasonal demand surges.
- **Location Benchmarking:** Analyze top-performing stores to replicate success factors across other branches.