

Coffee Sales Analysis & Prediction

Team members:

Mudhawi Alshiha: Data Cleaning, Exploratory Data Analysis (EDA), Plotly Visualizations, Business Insights.

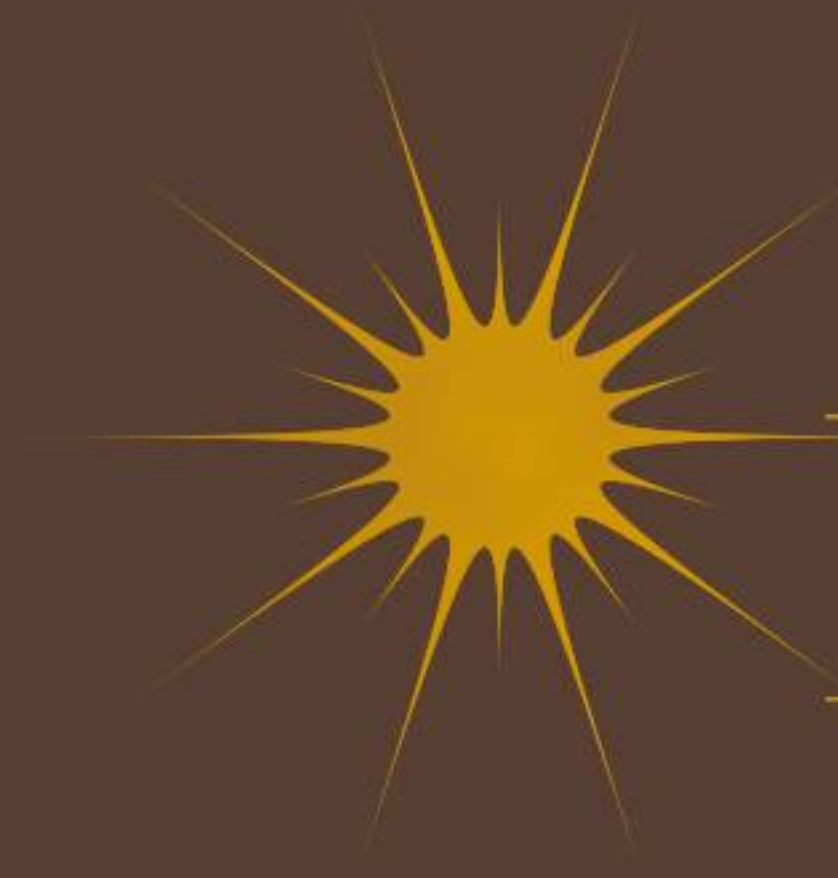
Musab Alabdulatif: Data Cleaning, Exploratory Data Analysis (EDA), Plotly Visualizations, Business Insights.

Mousa Hamoud: Machine Learning Modeling (OLS, Linear Regression, XGBoost), Model Evaluation, Streamlit Dashboard

Instructor: Hany Elshafy



Agenda



Problem Statement **01**

Objective **02**

Datasets Overview **03**

Methodology **04**

EDA Highlights (Visuals) **05**

Modeling Results **06**

Dashboard **07**

Key Insights & Business Recommendations **08**

Problem Statement

Coffee shops generate large amounts of sales transaction data, but patterns in timing, product preference, and payment behavior are not always clear.

Key Questions:

- Which days, times, and months generate the most sales?
- What coffee types are most popular?
- How do payment methods affect sales?
- Can machine learning predict future sales?

Why it matters?

- Optimize staffing & inventory
- Plan promotions around peak demand
- Improve sales forecasting

Objectives

- Analyze customer spending by weekday, time of day, and season
- Identify the most popular and most profitable coffee types
- Compare weekday vs weekend, and card vs cash payments
- Build regression & ML models to predict sales
- Deploy an interactive Streamlit dashboard for results

Dataset Overview

COLUMN	TYPE	DESCRIPTION
Money	Numeric	Amount of money spent per transaction (SAR)
coffee_name	Categorical	Coffee type ordered (Latte, Americano, etc.)
cash_type	Categorical	Payment method (card or cash)
hour_of_day	Numeric	Hour of the transaction (0–23)
Time_of_Day	Categorical	Period (Morning, Afternoon, Evening, Night)
Weekday	Categorical	Day name (Mon–Sun)
WeekdaySort	Numeric	Numeric order of day (1=Mon ... 7=Sun)

Dataset Overview

COLUMN	TYPE	DESCRIPTION
Month_name	Categorical	Month name (Jan–Dec)
MonthSort	Numeric	Numeric order of month (1=Jan ... 12=Dec)
Date	Date	Transaction date (YYYY-MM-DD)
Time	Time	Transaction time (HH\:MM\:SS)

Source: Kaggle

Size: around 3,500 transactions (2025 data)

Quality Notes:

- Some categorical values standardized (spaces/case)
- Minimal missing values

Methodology

EDA → Modeling → Dashboard

Steps:

1. Data Cleaning (NumPy, pandas)
2. EDA (NumPy calculations + Plotly interactive visuals)
3. Modeling (statsmodels, scikit-learn, XGBoost)
4. Evaluation (R^2 , MAE comparison)
5. Dashboard deployment (Streamlit app.py)
6. Upload the work into GitHub

Questions?

Q1. Sales by Weekday

Which day of the week generates the most sales?

Q2. Popular Coffee Types

Which coffee is ordered most often?

Q3. Weekday vs Weekend

Compare average spending

Q4. Monthly/Seasonal Trends

Which month has the highest sales?

Q5. Best Coffee Type in Each Season

Which coffee type is the best-seller in Season?

Q6. Hourly Trends

Which hour of the day has the most sales?

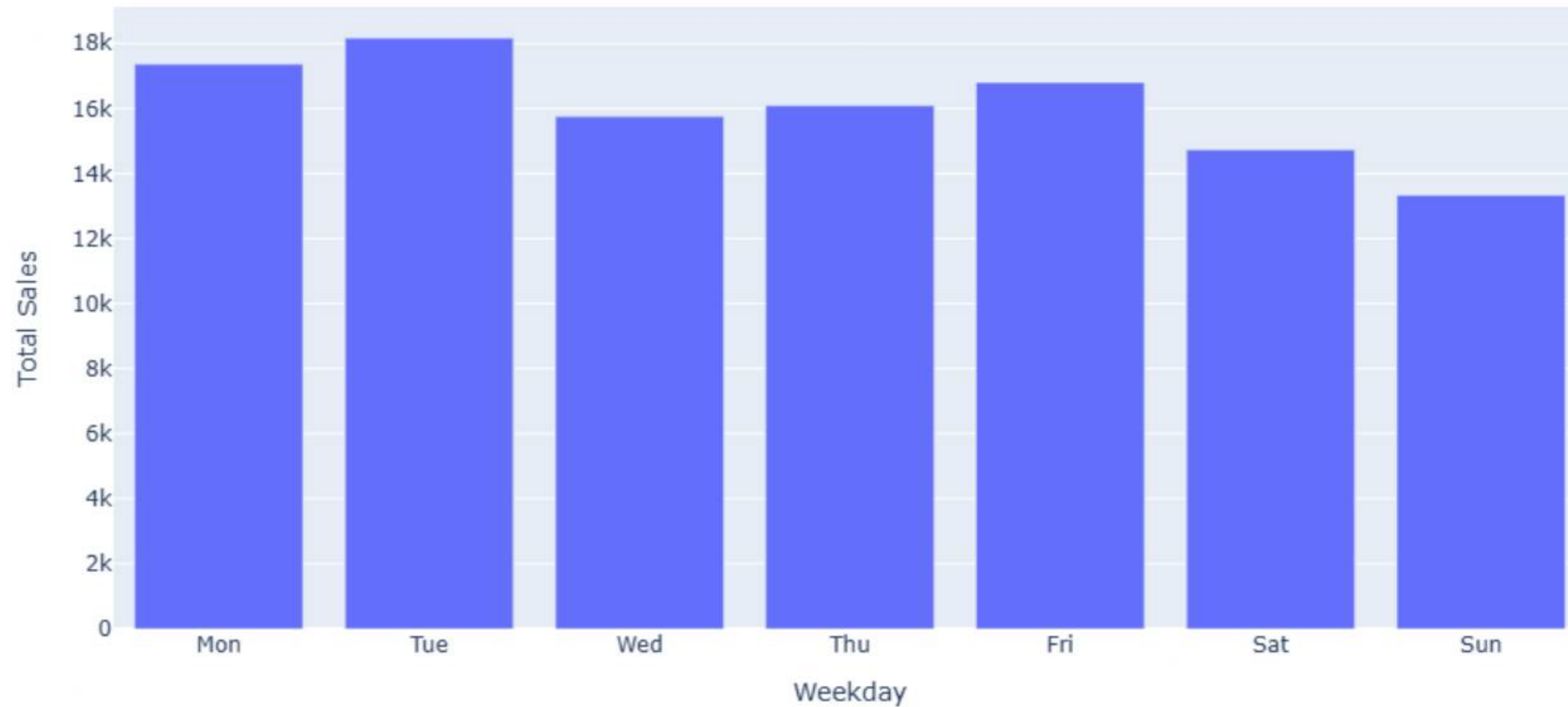


Q1. Sales by Weekday

Which day of the week generates the most sales?

Finding: Tuesday — total sales 18,168.38

Total Sales by Weekday (All Years)

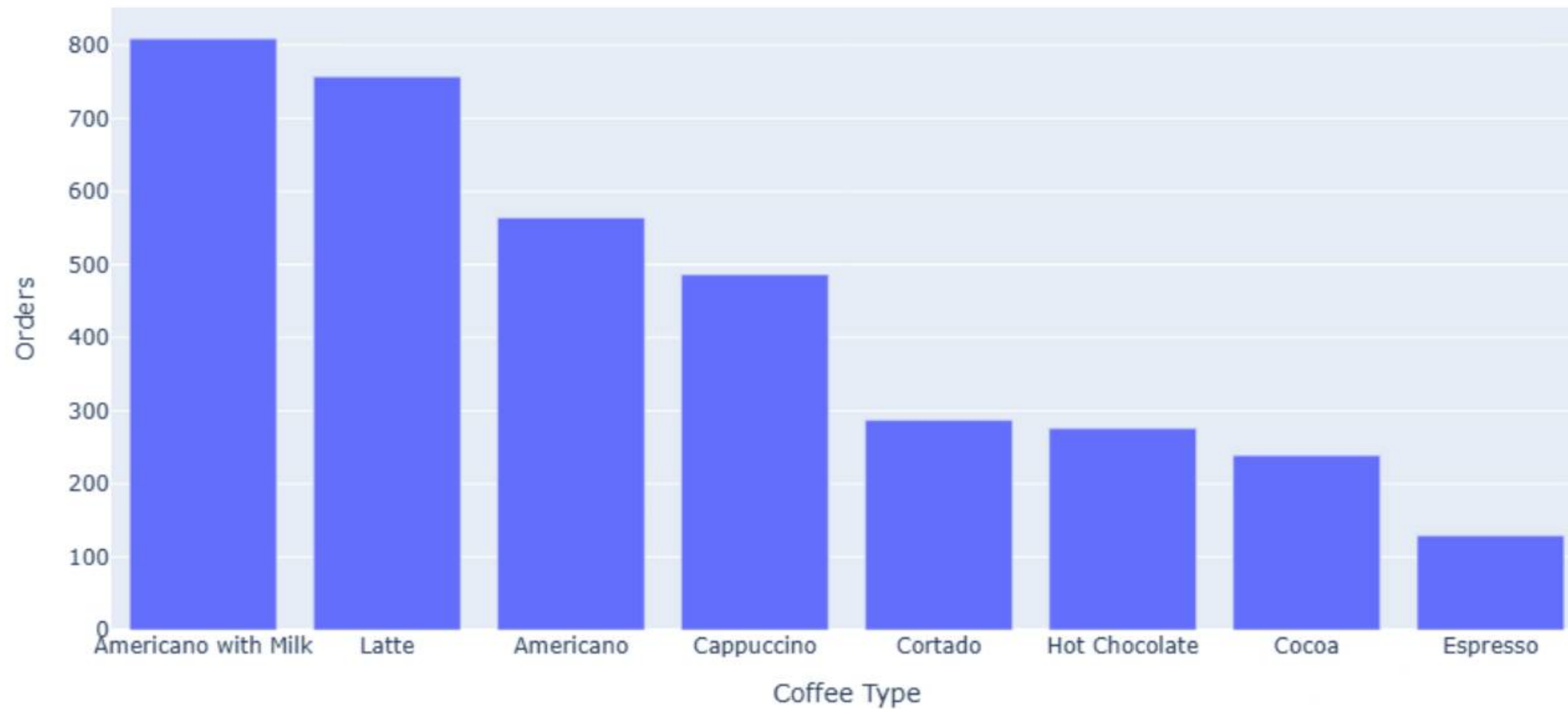


Q2. Popular Coffee Types

Which coffee is ordered most often?

Finding: Americano with Milk — 809 orders

Coffee Popularity (Order Count)

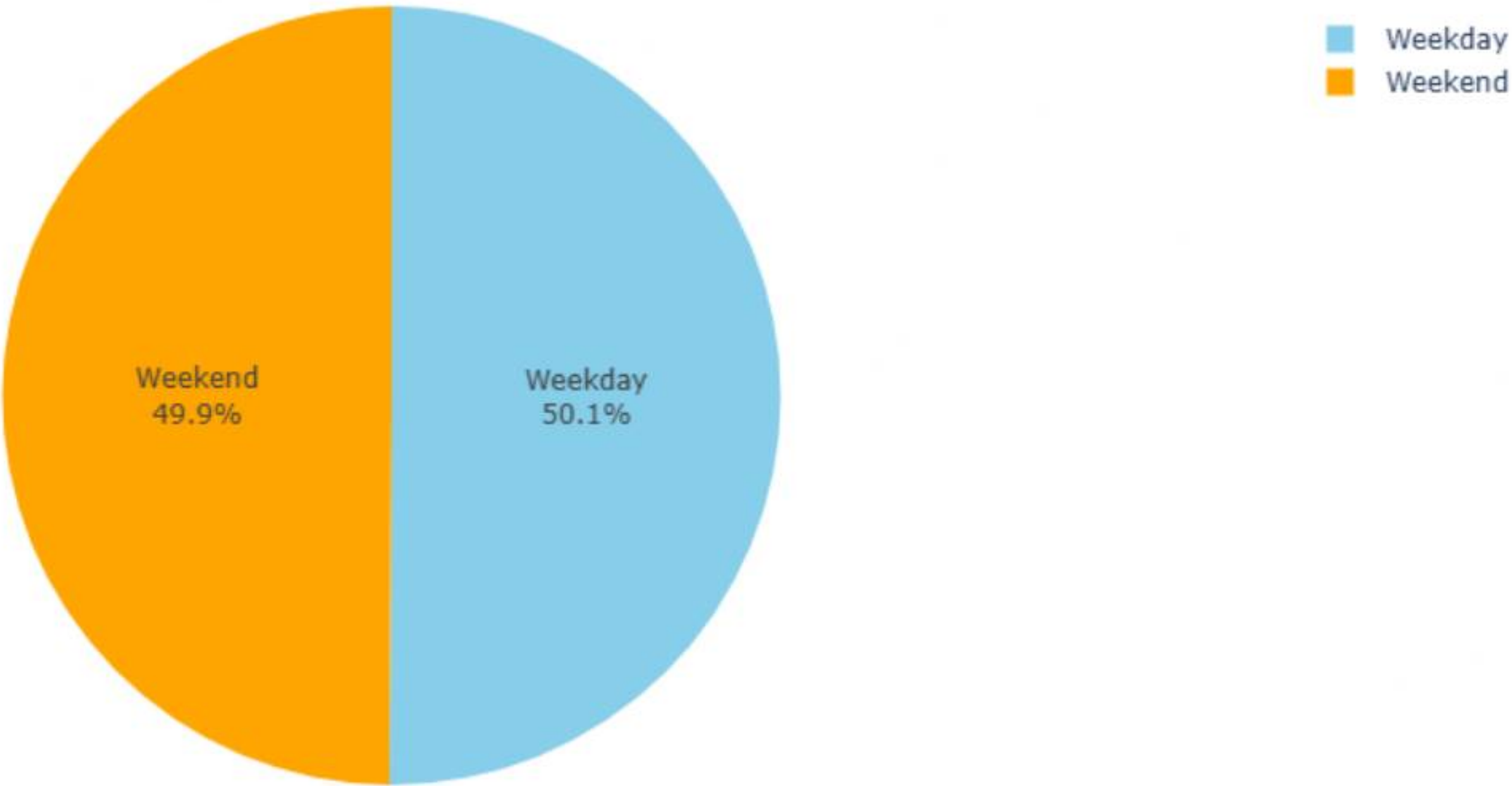


Q3. Weekday vs Weekend

Compare average spending

Finding: Weekdays (31.71) slightly higher than Weekends (31.47)

Average Sales: Weekdays vs Weekends

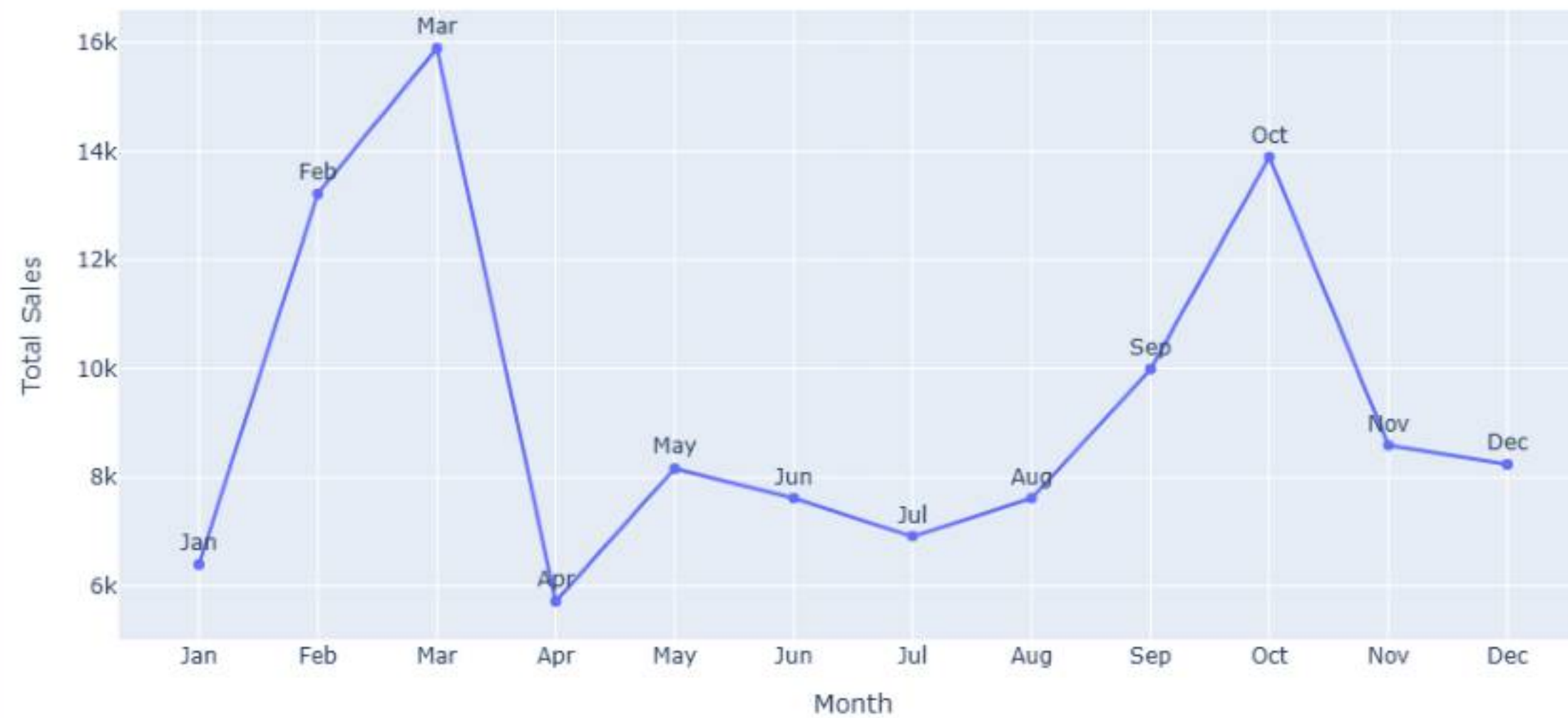


Q4. Monthly/Seasonal Trends

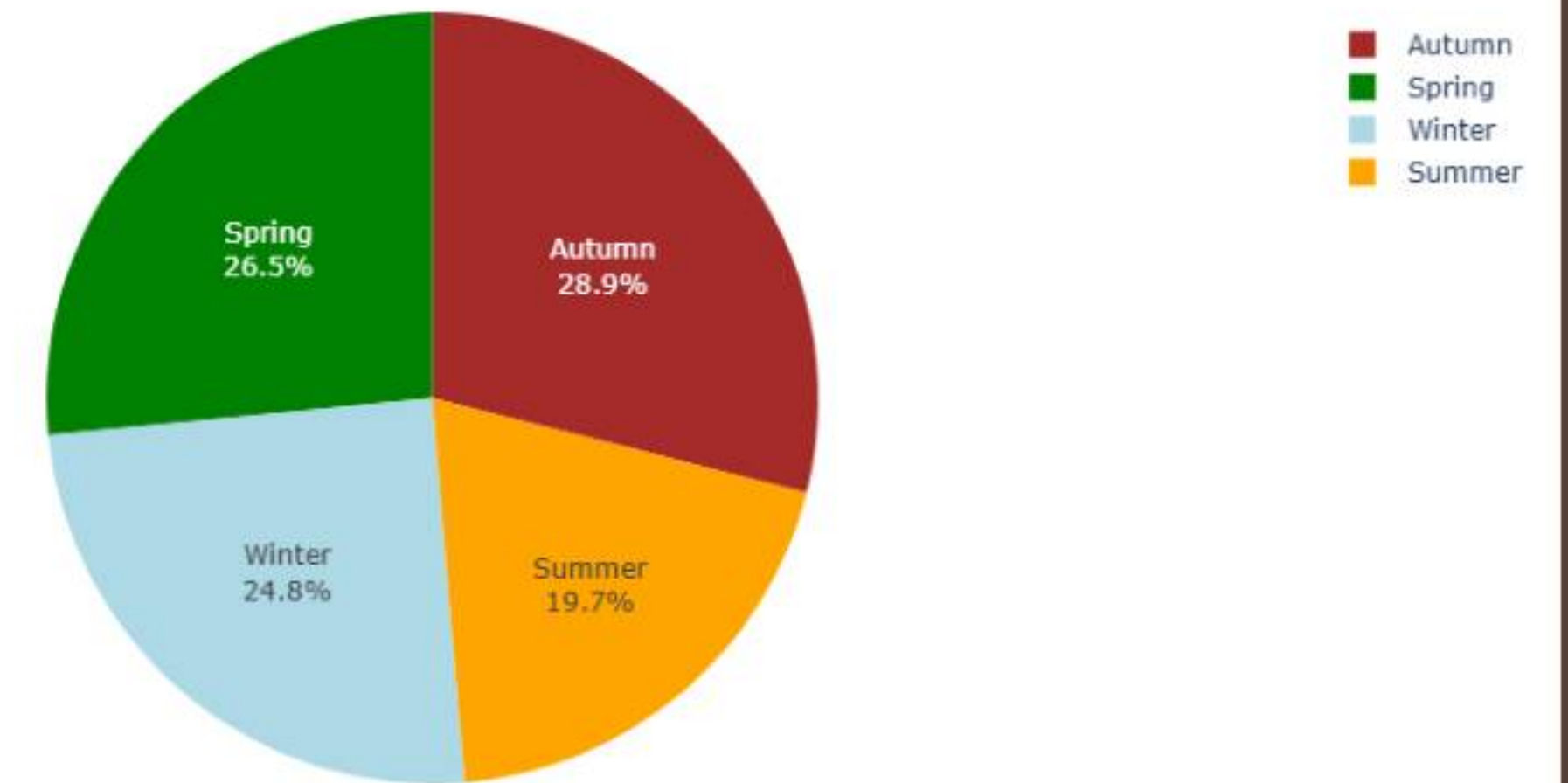
Which month has the highest sales?

Finding: March — total 15,891.64

Monthly Coffee Sales Trend



Seasonal Share of Coffee Sales



Q5. Best Coffee Type in Each Season

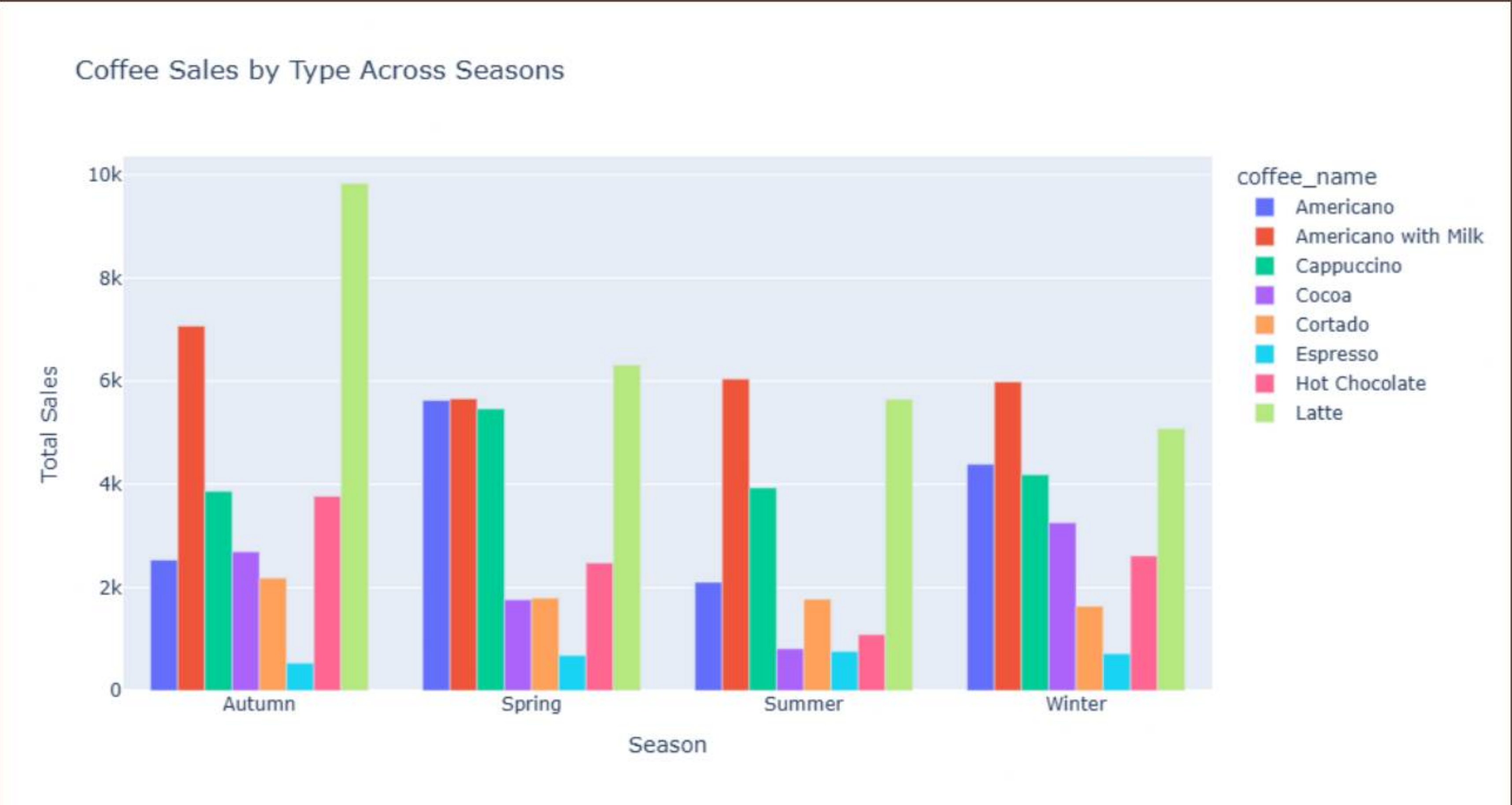
Which coffee type is the best-seller in Season?

Autumn: Latte → 9837.36

Spring: Latte → 6312.94

Summer: Americano with Milk → 6040.16

Winter: Americano with Milk → 5986.84

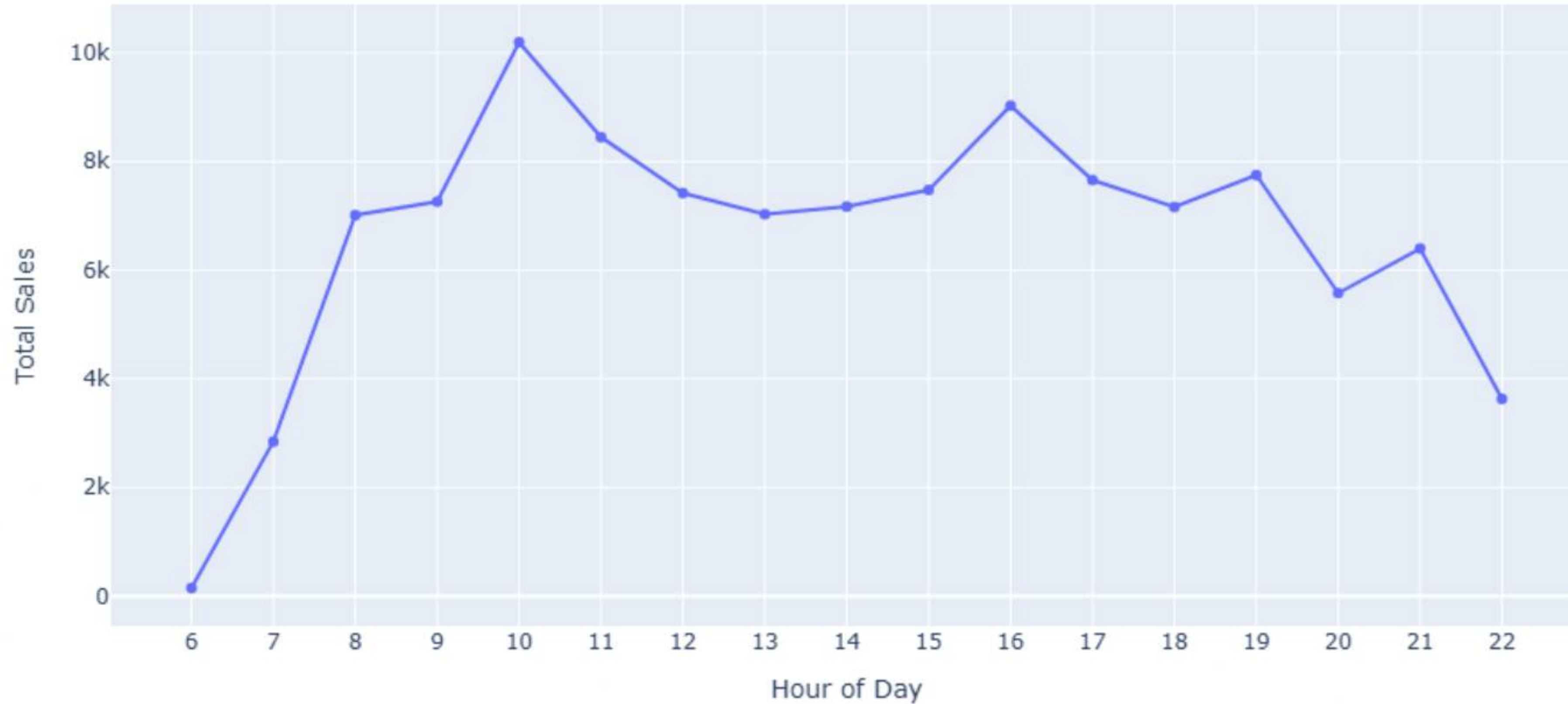


Q6. Hourly Trends


Which hour of the day has the most sales?

Finding: 10:00 AM — total 10,198.52

Total Sales by Hour of Day (All Years)

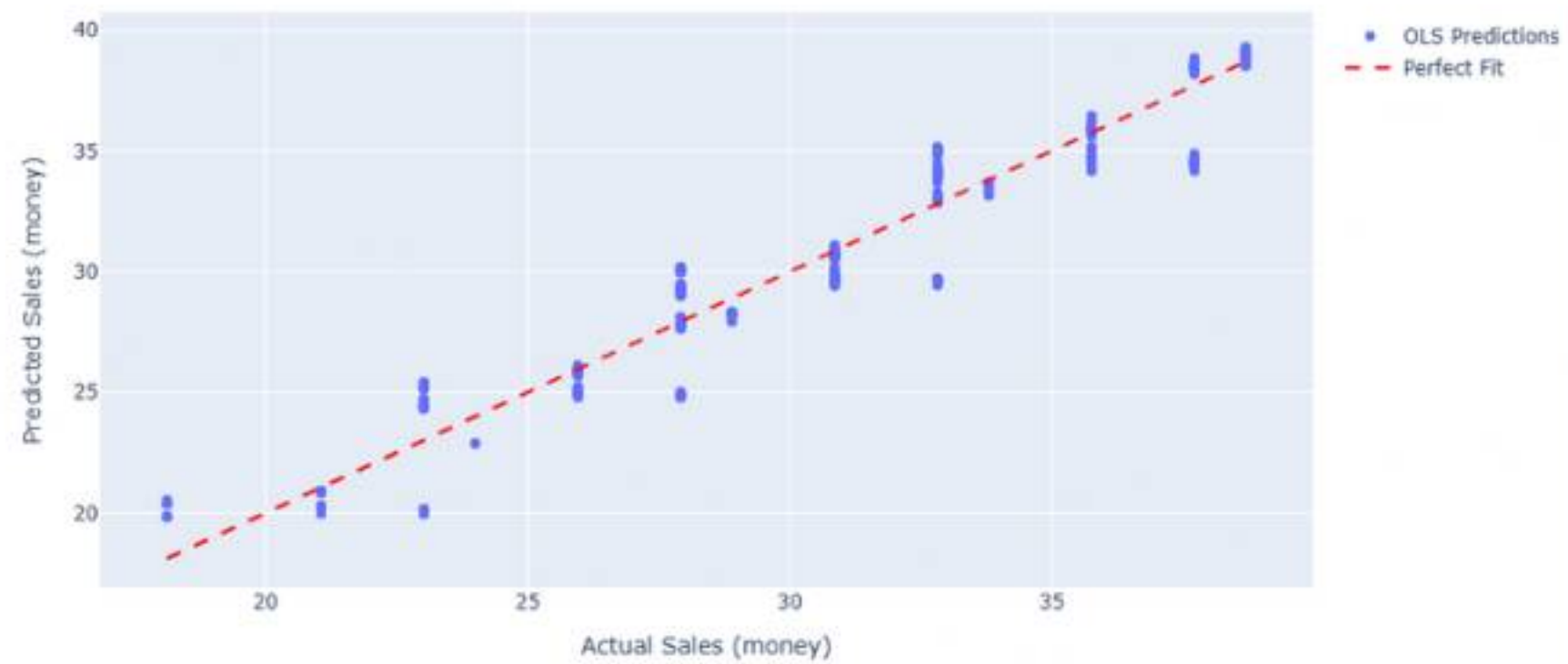


Modeling Results

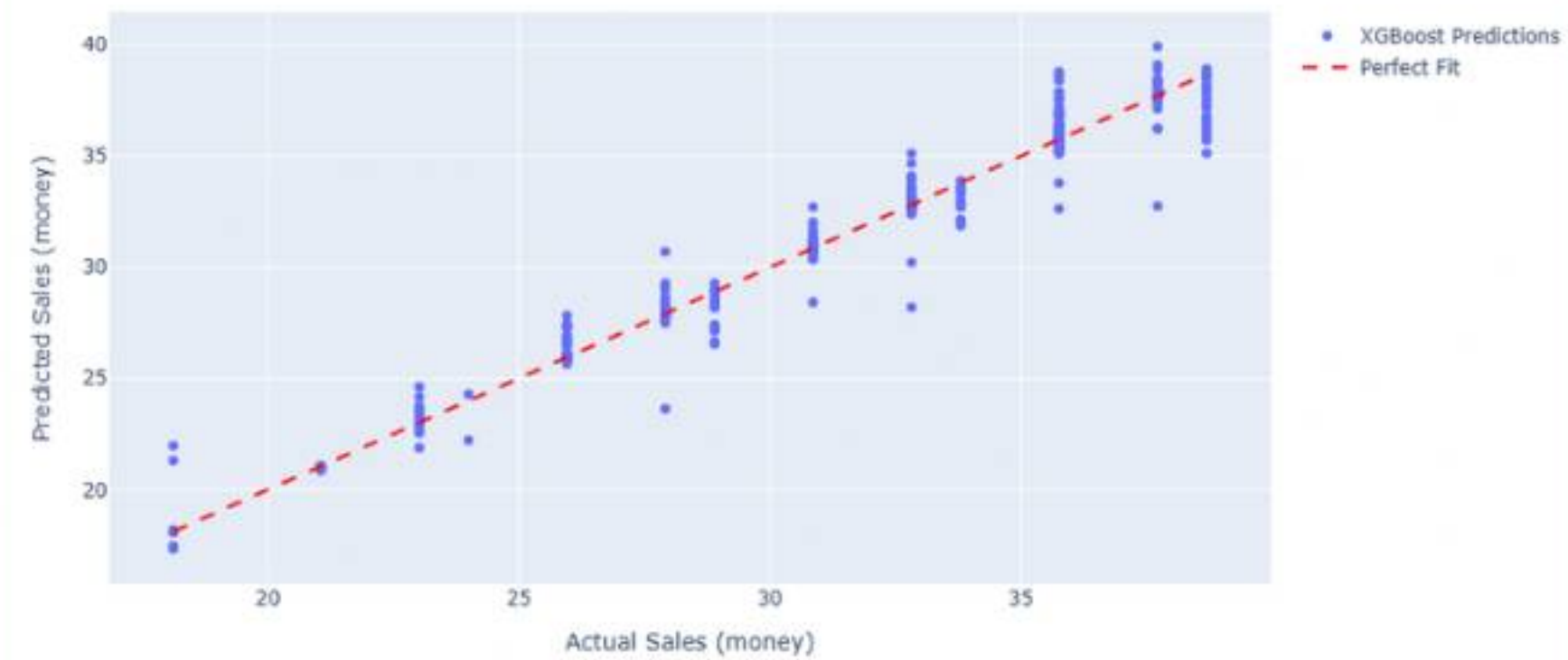
- Models tested: OLS, Linear Regression, XGBoost
 - Features: time, weekday, month, coffee type, payment type
 - Target: money (sales amount)
 - Metrics: R^2 , MAE
 - Best Model: XGBoost
- 

Model Comparison

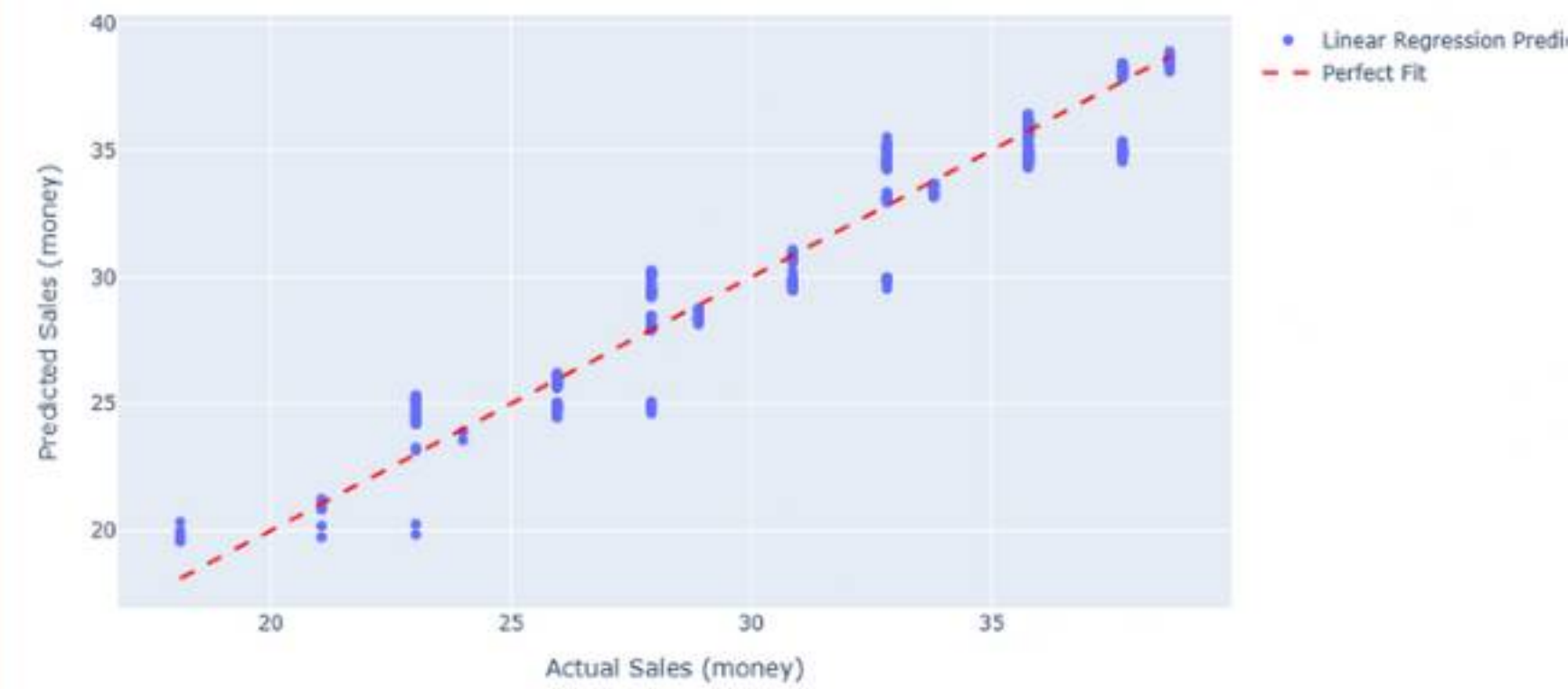
Actual vs Predicted Sales (OLS Model)



Actual vs Predicted Sales (XGBoost Model)



Actual vs Predicted Sales (Linear Regression Model)



Dashboard

- Built with Streamlit (app.py)

Features:

- Upload CSV datasets
- Explore interactive EDA visuals
- Run predictive models
- Display insights by coffee type, payment method, weekday, etc.

→ Key Insight

- Tuesday is the top sales day → allocate staff accordingly
- Americano with Milk is most popular → bundle offers
- Night spending highest → target late customers with promos
- Card payments dominate → partner with banks for loyalty deals
- March peak sales → plan seasonal promotions
- 10:00 AM sales peak → promote morning rush bundles

References

- Kaggle. (2024). Coffee Sales Dataset. Kaggle. <https://www.kaggle.com/>
- Hugging Face. (2024). Coffee Sales Dataset. Hugging Face. <https://huggingface.co/>
- MUSAB10000. (2024). Project-Coffee-Sales (GitHub repository). GitHub. <https://github.com/MUSAB10000/Project-Coffee-Sales>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56). SciPy. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 92–96). SciPy.
- Streamlit Inc. (2024). Streamlit documentation. Streamlit. <https://docs.streamlit.io/>

Thank you All

A large, flowing, yellow abstract shape, resembling a stylized ribbon or a wave, curves across the lower right portion of the image. It has a thick, organic form with multiple curves and undulations. The background is a solid, dark brown color.