Running the 15 Cafe

An Analysis of Coffee Shop Data and Resulting Strategies

ECE143 | November 20, 2025

Introducing the Coffee Sales Dataset

Introduction of Dataset and our Objectives

The core data utilized for this analysis is a year-long record of coffee shop transactions, detailing purchase time, payment type, customer ID, and amount spent from kaggle

Based on the Data Visualization and Modeling, our objectives are:

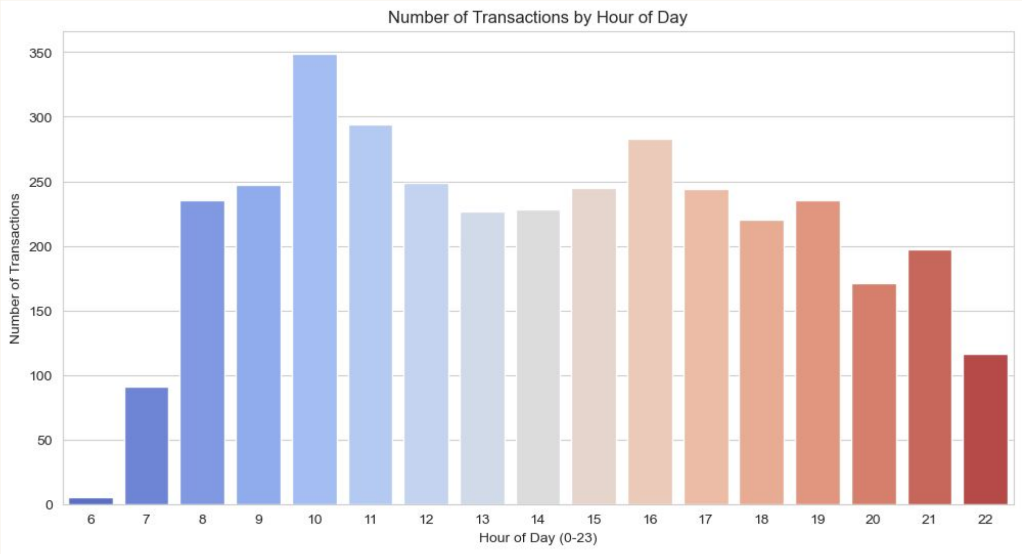
- **Optimizing Operational Setup**
- **Increasing Profitability**
- **Customizing Menu Strategy**

UC San Diego

Briefe Data Visualization

Transaction by Hour of Day

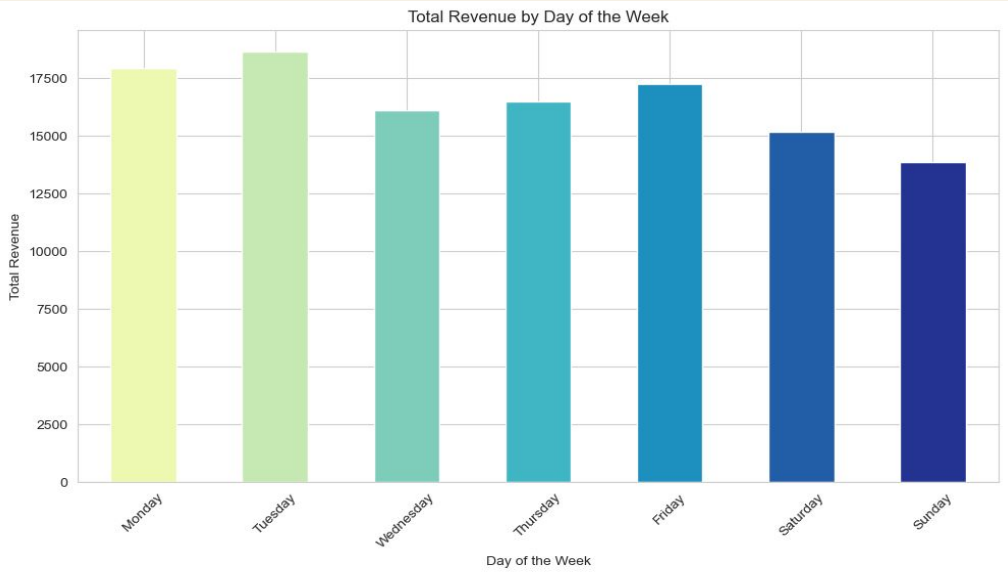
- **Morning Peak:** Transaction volume reaches its first and most concentrated peak between **10:00 and 11:00**
- **Afternoon Peak:** A secondary, smaller peak occurs between **15:00 and 16:00**, aligning with the typical afternoon coffee and snack time
- **Off-Peak Periods:** Transaction volume drops significantly during the opening hours (before 9:00) and in the late afternoon or evening (after 17:00)



Briefe Data Visualization

Transaction by Hour of Day

- **Revenue Peak: Tuesday** generates the highest total revenue, closely followed by **Monday**. This indicates that the primary consumption demand is concentrated at the **beginning of the work week**
- **Revenue Trough: Sunday** yields the lowest revenue across all days, followed by Saturday, confirming that weekend revenue levels are significantly lower than weekdays



Setup the Coffee Shop

How Can a Coffee Shop Optimize Operations?

Understanding the Considerations for Coffee Shop Operations

- **Material Costs**
  - A coffee shop has the challenge of considering the material costs for its product.
    - e.g. How much milk to get? How much coffee beans to order?

- **Staffing Costs**
  - Wages for staffing are important for profit optimization
    - e.g. Overstaffing vs. Understaffing

UC San Diego

Stocking the Milk and Coffee

Analysis of Popular Drinks Based on Material

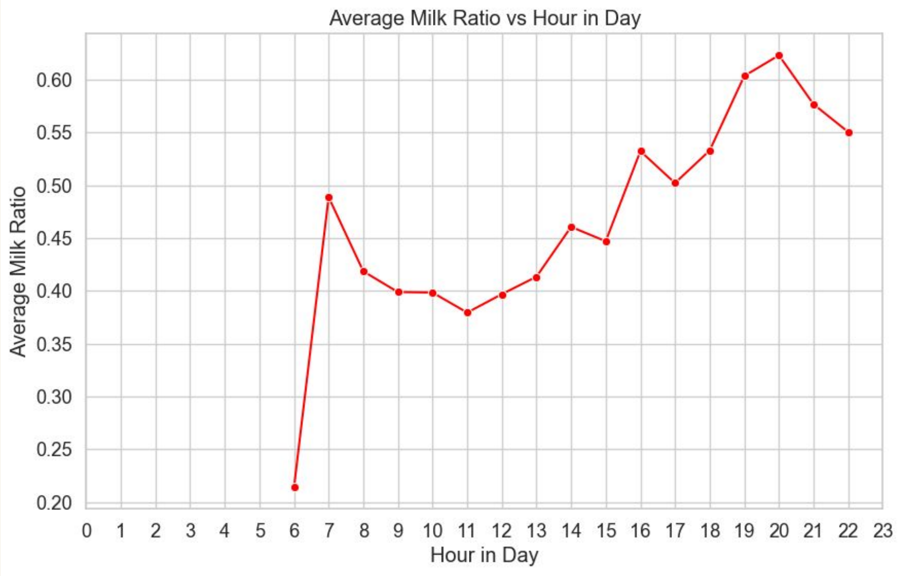
- Drinks can be viewed from a material perspective with regards to the composition of their ingredients; i.e the Ratio of Milk to Coffee
    - Americanos are all coffee
    - Flat Whites are 3 parts milk to 1 part coffee
- Analysis of material composition of sales provides insights on:
    - Ingredient-based analysis for supply and pricing pricing purposes
    - Customer tastes

Stocking the Milk and Coffee

Analysis of Popular Drinks Based on Material

- **Average Milk Ratio** per hour as a custom metric for the general “taste” of the customer base for a given time in the day in terms of ingredients

- **Time-based analysis** to associate milk-ratios with hours in a day
  - Informs shop of popularity of milk-forward versus coffee-forward drinks



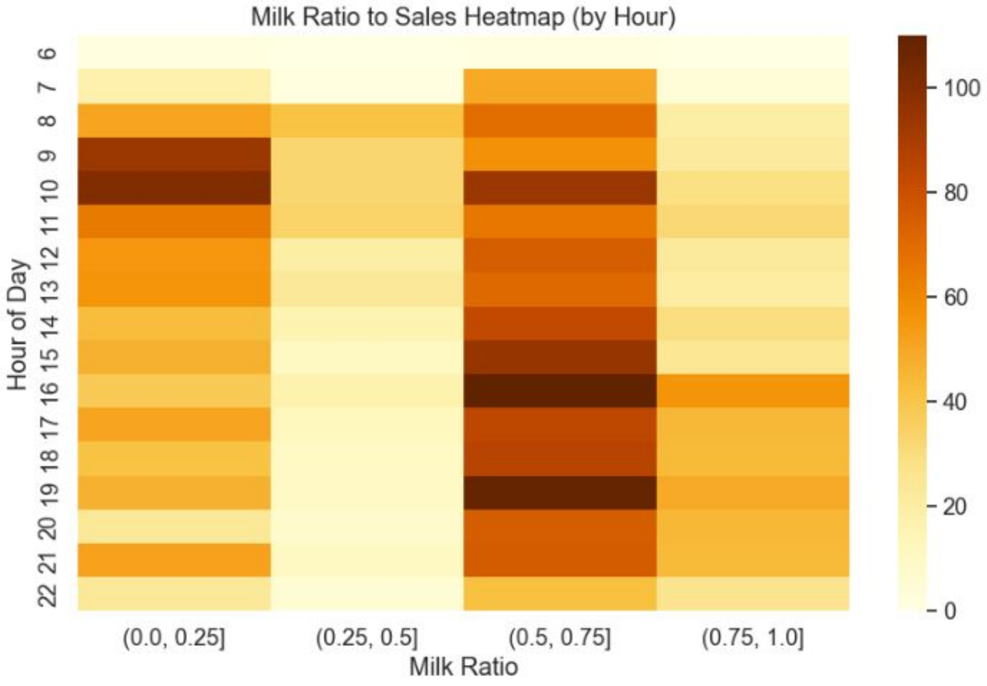
UC San Diego

Stocking the Milk and Coffee

Analysis of Popular Drinks Based on Material

Takeaways for Shop Strategies:

- Less prep for coffee beans in the evening to maximize bean freshness for the next day
- Stocking milk becomes imperative past the afternoon



Staffing

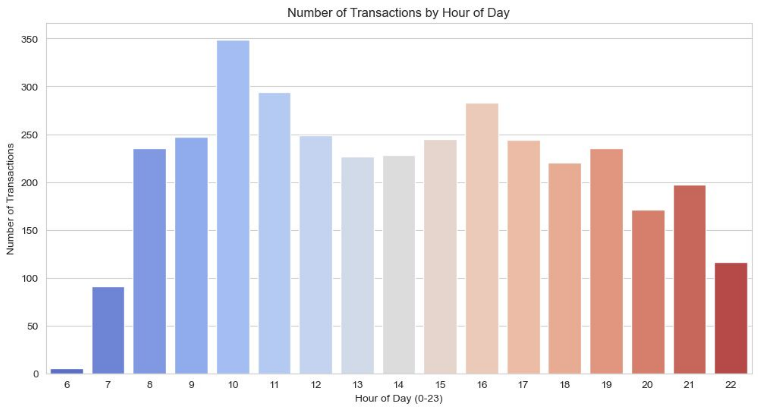
Analysis of Shop Business by Time

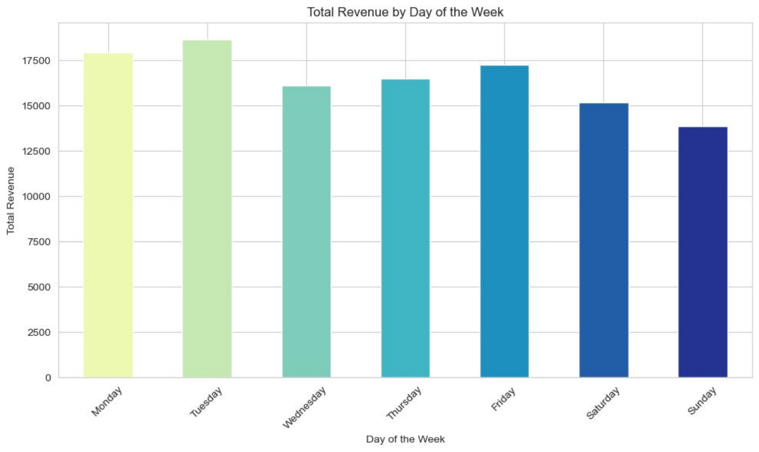
Analysis on when the shop is most/least popular allows a shop to optimize human resources!

Staffing & Efficiency: Barista and staff scheduling should concentrate the majority of human resources between **10:00 and 16:00**.

Early Week Strategy: Given that **Monday and Tuesday** are the highest revenue days, ensure maximum staffing and the highest inventory levels on these days to capture sales opportunities and guarantee service quality.

Weekend Adjustment: Operations management should review the low Sunday revenue. If profits cannot cover costs due to low traffic, consider reducing Sunday operating hours or adjusting staffing levels.





UC San Diego

“Show me the money”

Increasing profits over-time through promotional strategies

UC San Diego

Studying how customers spend

We study the composition of sales on three levels:

1. How do sales vary over time (seasonally, daily etc)
2. How do individual customers spend
3. How does the drink influence sales (drink price, popularity of drink)

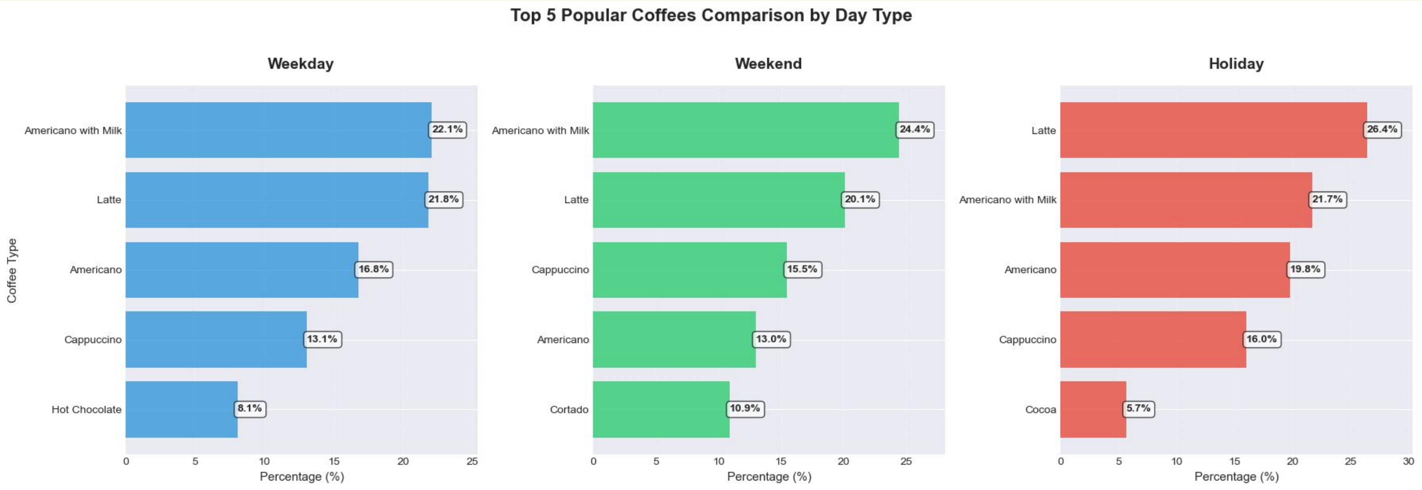
Based on our observations we devise strategies based on:

1. Customer Rewards
2. Promotional Campaigns to run

UC San Diego

Stocking Strategies based on the Day Type

- **Core best-seller: Americano with Milk** and **Latte** stay in the top. → Keep these as always-in-stock, high-priority items (beans, milk)
- **Weekend pattern**: **Americano with Milk** peaks and **Cortado** enters the top 5. → Plan for slightly higher milk + espresso usage and extra Cortado ingredients for "treat yourself" orders.



- **Holiday pattern**: **Latte** jumps to the clear #1 with strong **Americano** and **Americano with Milk** demand.
- → Boost latte capacity (milk, flavor syrups) and increase espresso bean stock; keep only light cocoa for niche orders.

UC San Diego

Understanding our Customer Base

Unsupervised Clustering for Customer Analysis

- Clustering customers into groups provides insights to our Customer Base.
  - Dataset includes anonymized IDs for individual customers.
- Goal is to group customer base based on certain features that are helpful for targeted promotions/advertising
  - Help shops understand the habits of their client to optimize marketing strategies.

Algorithm: K-Means Clustering

- Partition dataset into a predefined set of groups based on similarity, minimize *inertia* based on clustering.
- Provides “centers” to each cluster (criteria by which each client in the cluster is “closest” to.

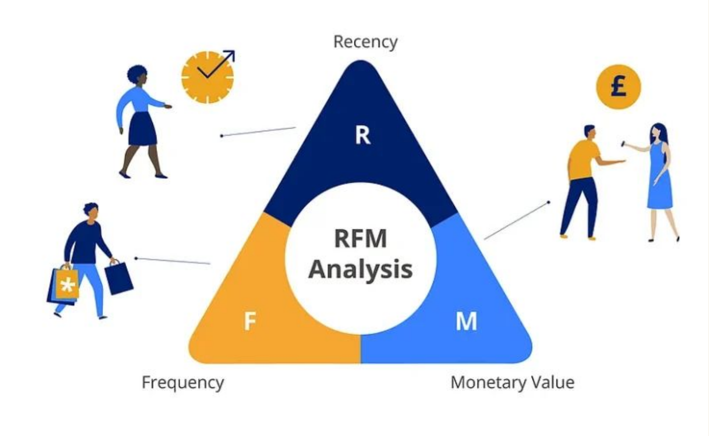
UC San Diego

Understanding our Customer Base

Unsupervised Clustering for Customer Analysis

What is our Clustering Metric? (i.e. what defines a cluster “center”?)

- Solution: RFM Analysis
  - Marketing technique to group customers based on three criteria
    - **R**ecency - How recent was latest visit
    - **F**requency - How often customer visits
    - **M**onetary Value - How much client spends



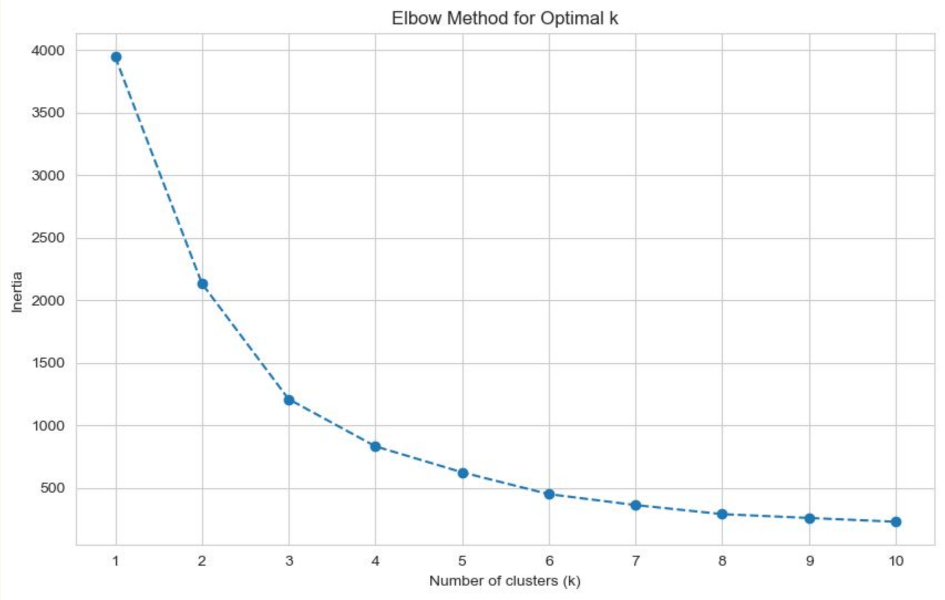
Choosing a Suitable Number of Groups

Elbow Plot

Observations:

- Inertia drops significantly as number of clusters increase from 1 to 3.
- Inertia starts to flatten after k=4, meaning additional clusters provide diminishing returns.

3 & 4 are optimal values for the number of clusters.



We finally choose **3** as the k value to maximize the separation between groups. In addition, it is reasonable to classify customers into 3 categories in terms of the frequency of visit.

UC San Diego

Clustering Result

Cluster 1: Churned / Single-Visit Customers

This group is a mixture of customers who have **highly lapsed (Churned)** or were **single-visit** patrons. They have not visited in almost nine months on average and represent the lowest value segment

- **Low-Cost Recall:** Avoid allocating a significant budget to this group. We may attempt a single, highly attractive deep-discount offer (e.g., **"We Miss You: 50% Off Your Next Order"**) as a last-ditch effort
- **Avoid Regular Promotions:** Reserve regular promotions (e.g., "Buy One Get One Free," loyalty points) for more active segments

Metric	Value
Recency (R)	270.22 days
Frequency (F)	1.65 visits
Monetary (M)	$ 53.40

Clustering Result

Cluster 2: At-Risk / Low-Frequency Customers

This represents a **significant opportunity**. These customers have shown interest by visiting 2-3 times but have lapsed for over three months (96 days). They have not yet formed a habit and are at high **risk of churning**

- **Habit Formation:** Implement a targeted **"Habit Building Program."** For example: "Complete 4 visits in the next 30 days to earn a free drink."
- **Personalized Recommendation:** Push related new products or upgrade offers based on their past purchase history to guide them back to the shop
- **Limited-Time Recall:** Send offers like "Free Size Upgrade" or "Complimentary Dessert“

Metric	Value
Recency (R)	96.51 days
Frequency (F)	2.71 visits
Monetary (M)	$ 86.06

UC San Diego

Clustering Result

Cluster 3: Super VIP Customers

The **“highest-value” customers**. While small in size, frequency and spending are exponentially higher than the other groups

- **Reward Status, Not Discounts: Absolutely avoid giving them price-based discounts,** discounts may diminish their VIP experience
- **Exclusive Treatment:** Offer **exclusive experiences**, such as early access to new product tastings
- **Maintain Feedback Loop:** Invite them to provide feedback on service or new menu items, making them feel like an integral part of the business's decision-making process

Metric	Value
Recency (R)	71.50 days
Frequency (F)	74.75 visits
Monetary (M)	$ 2259.01

UC San Diego

How do the customers drink

Feature Engineering & User Profiling

1. **Customer History** (The "**Memory**"):

   capture a customer's personal habits and spending capacity: customer_favorite_coffee, last_coffee, customer_visit_count, customer_avg_spend.

2. **Cyclical Time Encoding** (**"When is the purchase made?"**):
   - **Basic temporal attributes**: hour, day_of_week_num, month_num, is_weekend
   - **Cyclical encodings (Sin/Cos)**: hour_sin/cos, day_of_week_sin/cos, month_sin/cos
3. **Contextual & Interaction** (**"What is the context?"**):
   - **Price environment**, historical average prices for that time period: avg_price_by_hour, avg_price_by_month
   - **Interaction features**: hour_weekend, month_weekend

How do the customers drink

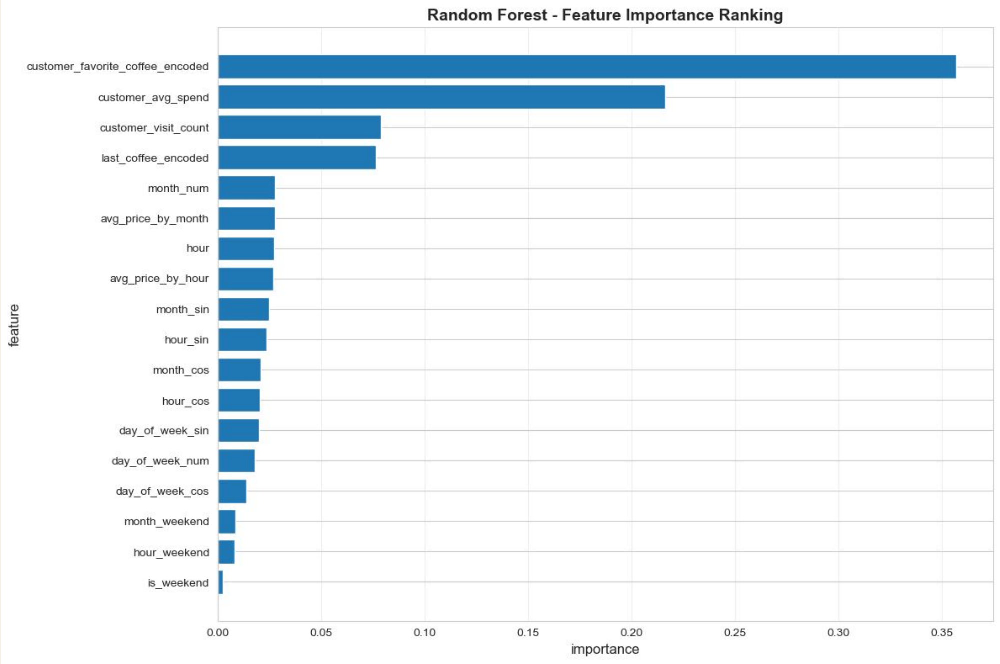
What drives the coffee choice?

Based on **Random Forest** Feature Analysis

- Top Predictor (36%): **customer_favorite_coffee_encoded**. The strongest signal is simply **what the customer usually buys**. Customers are creatures of habit.
- Secondary Predictor (22%): **customer_avg_spend**. **Price sensitivity** plays a major role in decision-making.

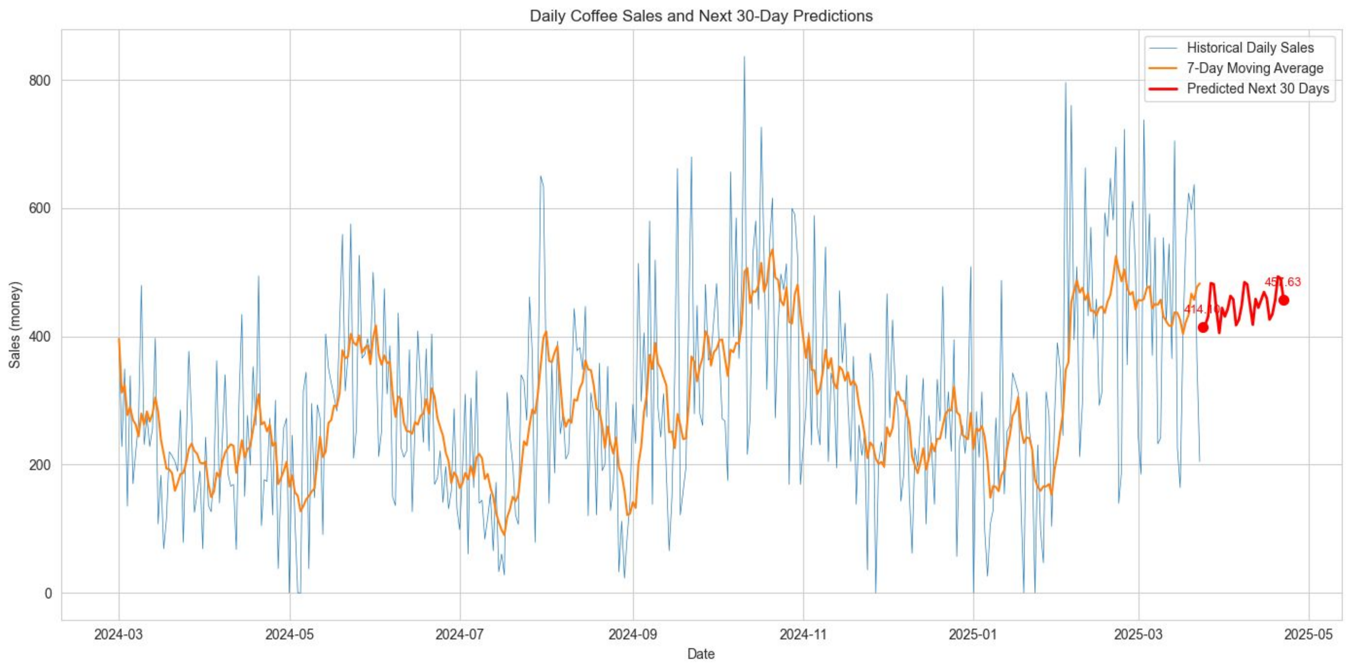
Takeaway for Business:

- **Time is less relevant**: Static time features (Hour, Month) contributed < 3% each.
- Recommendation Strategy: Personalization algorithms should **prioritize User History** over Context (Time/Season) to maximize accuracy.



How do sales vary over time (seasonally, daily etc)

Can we predict the sales (based on historical data)



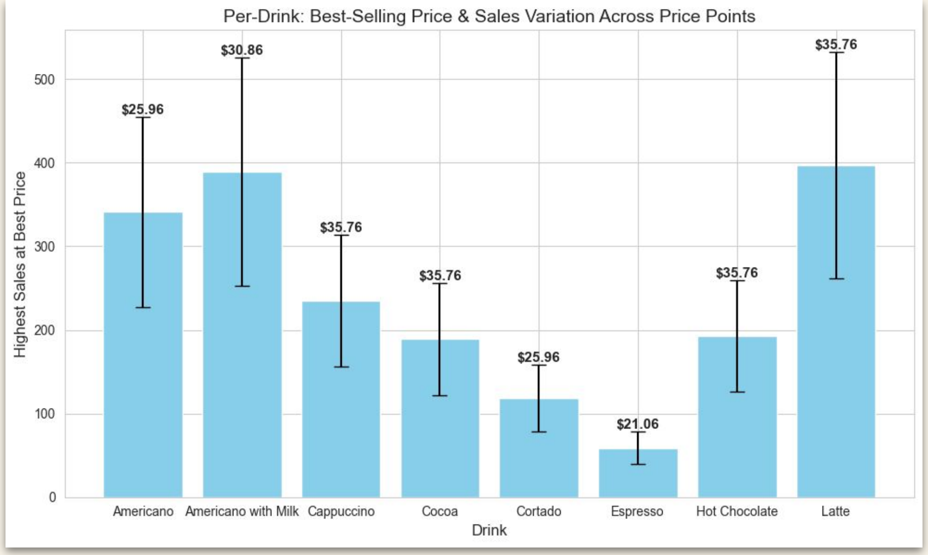
We used Seasonal ARIMA to forecast daily sales for the next month.

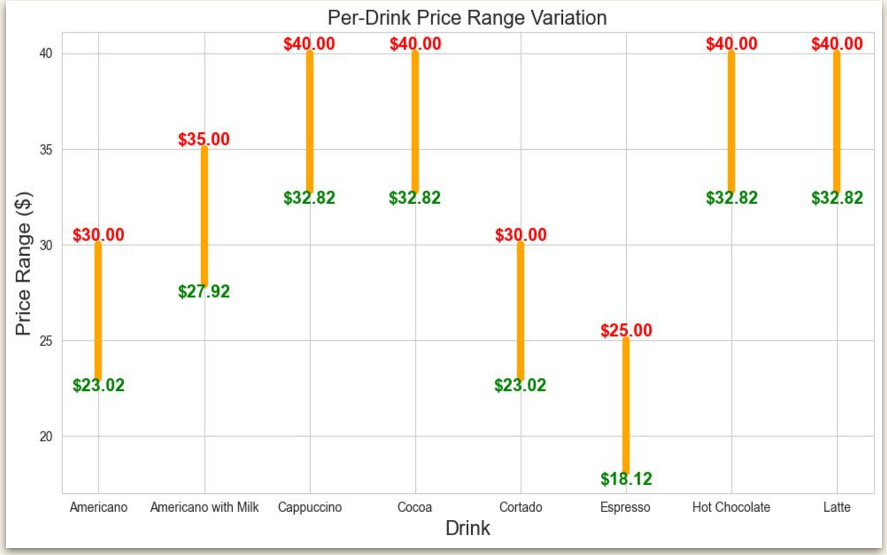
If profit margins are known, we can estimate which drink to promote to maximise profits

UC San Diego

How does the drink price influence sales

We want to identify the profit margins (if any) for our drinks





We can identify profit margins for each drink.
For instance cortado sell most at 25.96, but we easily can vary from 23 (-2) to 30 (+5)

Which drink to promote today (for a fixed profit margin)

We define a gap scoring strategy based on two features:

1. **Sales forecasted** (can be based on rolling averages e.g. 3d,5d,7d, 30d)
2. **Profit margin** (what price to sell that drink)

We score each drink as: (predicted_sales - today's sales) * profit_margin

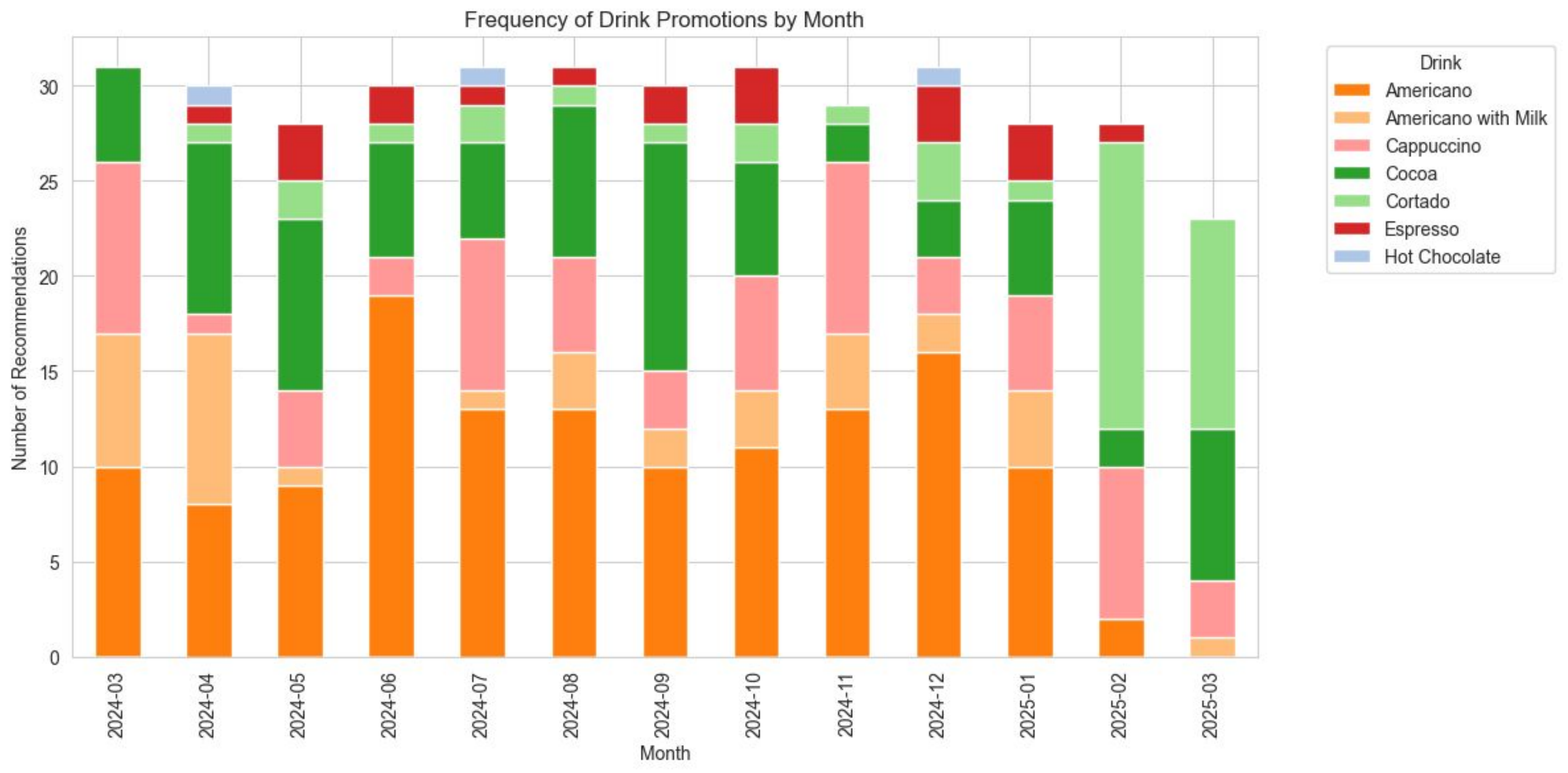
Drink	Predicted Sale	Today's sale	Profit Margin	Score
Americano	5	6	2	-2
Cortado	2	1	2	2
Hot Chocolate	**3**	**0**	**2**	**6**
Espresso	1	0	2	2

UC San Diego

Promotional Offer frequency (Monthly)

All profit margins are 2.0

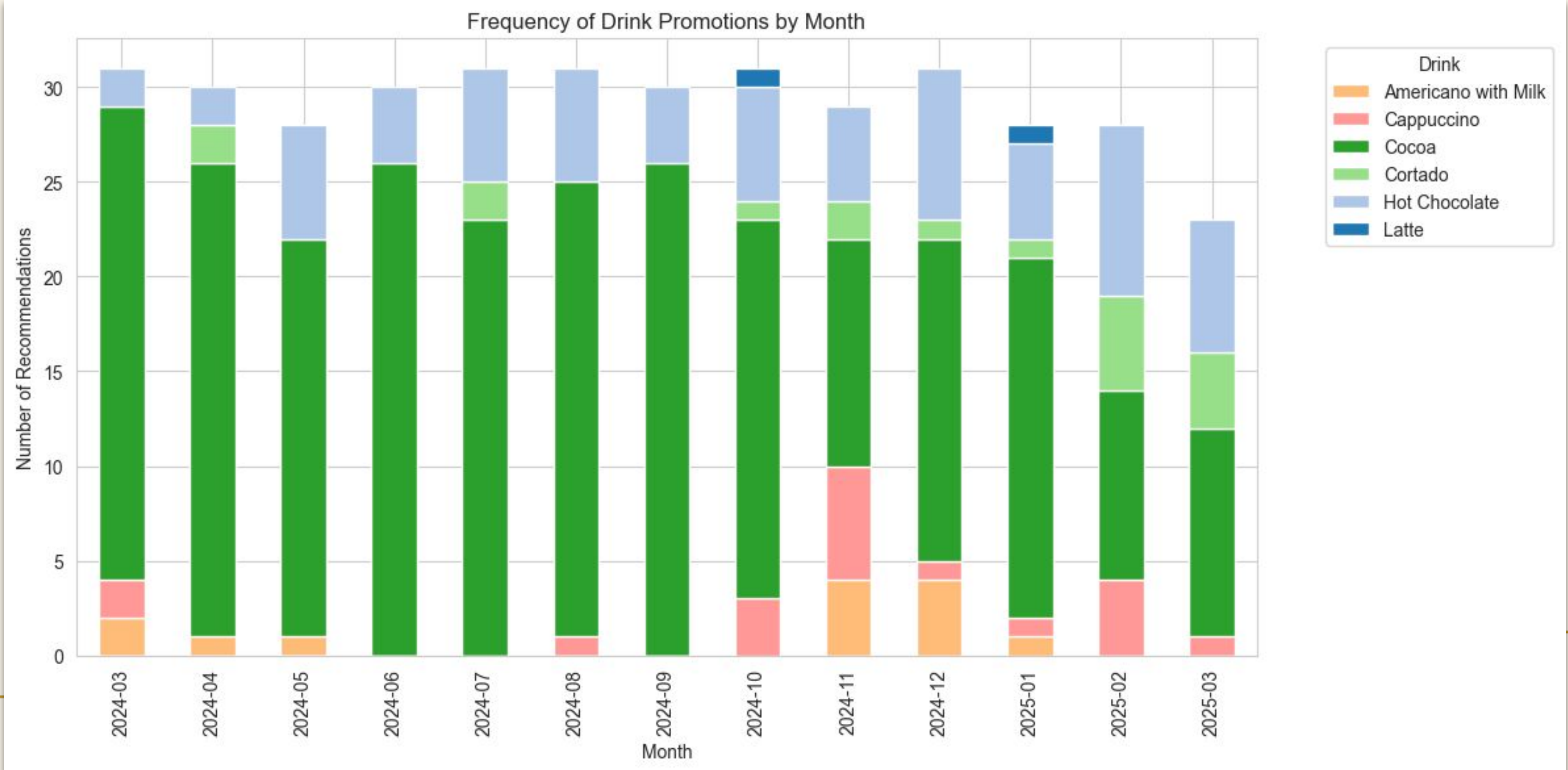
Under normal conditions balance your promotions as follows. Usually no point in promoting espresso, but Free refill is beneficial



Promotional Offer frequency (Monthly)

Increasing the price of milk based drinks (4 for 100% milk; 3 for <100% milk; 0 for rest)

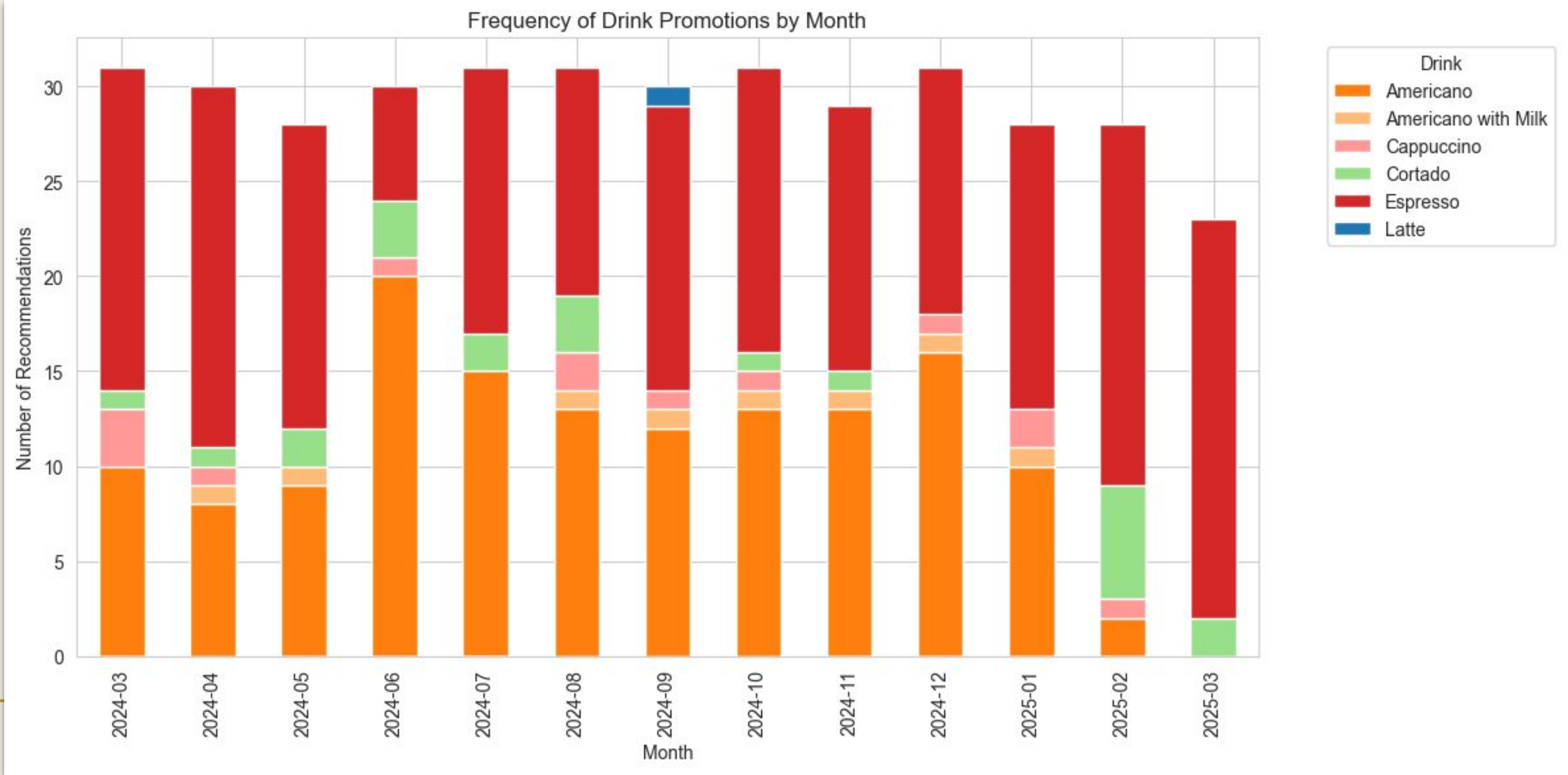
If milk gets expensive - promote Cocoa, Hot Chocolate

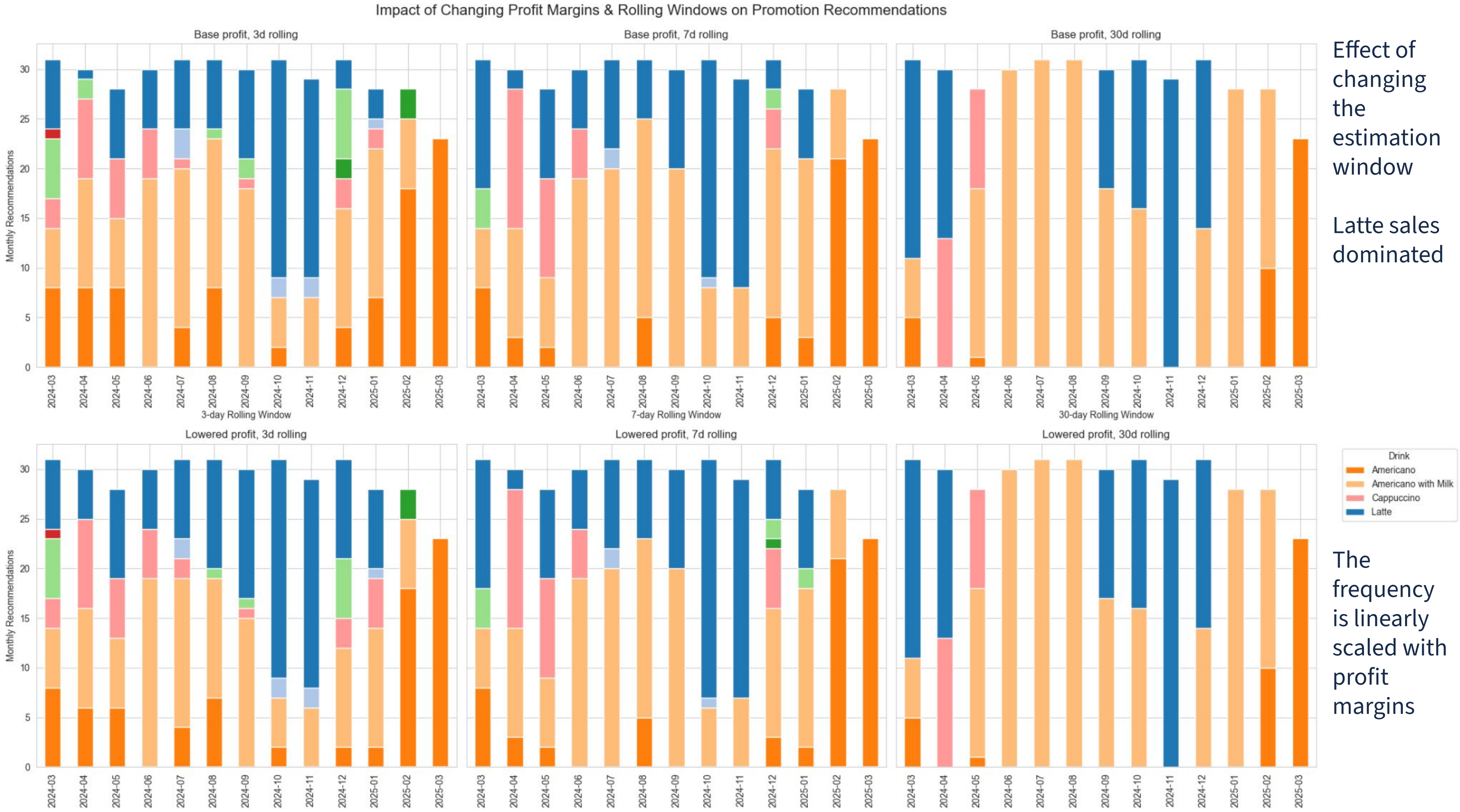


Promotional Offer frequency (Monthly)

Increasing the price of coffee shots (4 for americano, espresso, 3 for milk based coffee)

If coffee gets expensive - promote Americano, Espresso





Effect of changing the estimation window

Latte sales dominated

The frequency is linearly scaled with profit margins

Summary

Recap of Results

- Engineering custom features like Milk-to-Coffee Ratio and visualizing with Line Plots/Heat-maps to justify optimal stocking strategies based on user-tastes.
- Time-based bar charts to visualize popular hours for staffing strategies.
- Break-down of popular drink types by day type (holidays, weekdays, weekends) through bar chart to justify seasonal promotions for profit.
- K-Means Clustering to group clients based on RFM criteria for targeted promotions and understanding general habits of clients.
- Random-Forest Feature Analysis to rank feature importance of client behaviors and understand factors that lead to sales.
- Seasonal ARIMA to understand and predict sales based and test variants of profit margins.

UC San Diego