Linear Algebra and Starbucks

Seasonal drinks are a hallmark of Starbucks' menu, featuring popular holiday offerings like the Caramel brulee latte and new limited-time offerings like spicy refreshers. As a Starbucks barista, I've noticed distinct trends in the popularity of seasonal drinks. Holiday drinks, like the Sugar Cookie Almondmilk Latte, tend to gain popularity during the cooler fall and winter months, while iced refreshers and frappuccinos dominate sales during the warmer spring and summer seasons. These patterns highlight the importance of understanding when to introduce and phase out seasonal offerings to maximize sales and minimize waste.

For example, when the company reintroduced macadamia-flavored drinks earlier in the spring, they became fan favorites. But as we got closer to the end of summer, there was a noticeable decrease in sales and an increasing surplus of macadamia syrup. The drink was later removed from the menu due to it becoming unprofitable. Last month in November, my store gave out several bottles of unopened macadamia syrup to its baristas; around 12 bottles. Since our syrup pumps are calibrated to 0.25 ounces and each bottle holds 33.8 ounces (1L) of syrup, meaning one bottle contains just around 135 pumps. If we use the macadamia cream cold brew as a basis, 1-4 pumps of syrup are used for this $5.75-6.65 dollar drink, from sizes tall to trenta.

The revenue per pump for a tall size: $5.75/1 pump = $5.75, and a trenta size: $6.65/4 pumps ~= $1.66. For 12 bottles, around 1620 pumps**, minimum revenue loss** ($1.66/pump): 1,620 * 1.66 = **$2,689.20**, and the **maximum revenue loss** was ($5.75/pump): 1,620 * 5.75 = **$9,315.00** (*important note: customers are able to add as many pumps without charge*)

That's a lot of money lost from just one store in San Jose, highlighting the importance of optimizing how and when seasonal drinks/ingredients should be phased out. My project aims to address this challenge by using linear regression to analyze sales data and predict the best timing to phase out seasonal drinks. Although this is a personal project, it may provide insights that help Starbucks make informed, data-based decisions by leveraging the use of linear algebra. *Linear Regression* is used to model the relationship between one of more independent variables and a dependent variable. This modeling technique relies on the linear algebra principles of *Matrix Operations, Matrix inversion, and Matrix Transposes,* when expressing the linear regression model in matrix notation. We use *Matrix Notation* in cases where we need to handle multiple variables and large datasets, represented mathematically as:

$$\mathbf{y = X\beta + \varepsilon}$$

y: Vector of dependent variables (observed sales for all weeks)
X: A matrix containing independent variables (e.g., week numbers, temperature, etc.)
β: Coefficient vector: intercept and slopes; represents the weight or **impact** of variables
ε: Residual (error) vector

To find $\beta = (X^TX)^{-1}X^Ty$

X^T: Transpose of the matrix.
$(X^TX)^{-1}$: Inverse of the X^T matrix multiplied by the original X matrix.

X^Ty: Dot product of the transpose X matrix and the dependent variable vector.

I will use Python to calculate the regression coefficient vector (β) using matrix operations, predict weekly sales of our holiday drinks, and compare predicted values with actual sales. The formula to compute β involves matrix transposition, multiplication, and inversion, which is a lot of computation when working with multiple variables and datasets, necessitating the use of software like **NumPy** for matrix computations and **Matplotlib** for visualization. For this project, I am analyzing weekly sales data for a seasonal drink using the following *Variables* below represented as integer lists. Unfortunately, I dont have access to stabucks' actual data for factors such as foot traffic or drink sales. So, I will be generating a simulation designed to mimic actual sales patterns for a generic Starbucks seasonal drink, based on my personal experience as a barista:

- **Week**: To track trends over 14 weeks starting from November 7 (starbucks' holiday launch).
- **Temperature**: Shows how San Jose weather might impact drink sales; List containing average temperature from weeks 1-14.
- **Holiday Week**: To capture spikes in sales during holidays, weeks containing Starbucks holiday launch day, Thanksgiving, Christmas, and New Years will be set to 1 for their respective index. A 0 for all other non holiday weeks.
- **Daily Traffic**: To measure the number of customers visiting the store. Simulated weekly traffic.

Presenting the Linear Algebra with Results:

My program utilizes numpy functions to stack arrays as columns side by side to form a 2D array, as well as creting the bias column. Effectively forming the matrix **X,** which represents all the independent variables though this line:

X = np.column_stack((np.ones(len(Week)), Week, Temperature, HolidayWeek, DailyTraffic))

When it comes to solving for $\boldsymbol{\beta = (X^TX)^{-1}X^Ty}$, my program finds the transpose of X, where rows become columns and vice versa. It also uses numpy linear algebra function .inv to find the inverse of $\mathbf{X^TX}$**.** This provides the optimal coefficients (**β**) that minimize the difference between the predicted and actual values. These coefficients help identify the relationship between the independent variables and the dependent variable (Sales). Calculated with lines:

XT = X.T # Transpose of X

beta = np.linalg.inv(XT @ X) @ XT @ y # regression coefficients β

Finally, to solve for: **yPrediction = Xβ** I calculated the predicted sales values for each week based on the independent variables.The model successfully captures the expected trends, such as: *Declining sales over time, Sales spikes during holidays, and Reduced sales in warmer weeks.* Calculated with the line:

yPrediction= X @ beta

Code output:

Normal Matrix X with labels (Independent Variables):

row	Bias	Week	Temperature	HolidayWeek	DailyTraffic
0	1.0	1.0	63.0	1.0	900.0
1	1.0	2.0	58.0	0.0	700.0
2	1.0	3.0	57.0	0.0	600.0
3	1.0	4.0	56.0	1.0	850.0
4	1.0	5.0	55.0	0.0	650.0
5	1.0	6.0	54.0	0.0	680.0
6	1.0	7.0	53.0	0.0	720.0
7	1.0	8.0	52.0	1.0	1100.0
8	1.0	9.0	52.0	1.0	1000.0
9	1.0	10.0	52.0	0.0	640.0
10	1.0	11.0	53.0	0.0	610.0
11	1.0	12.0	53.0	0.0	610.0
12	1.0	13.0	54.0	0.0	630.0
13	1.0	14.0	56.0	0.0	580.0

Regression Coefficient Vector β = [3.65044755e+02, -8.25618837e+00, -4.96072943e+00 3.79226174e+01 2.18230223e-01]

Interpretation:

Intercept (**Bias**): β_0 = 365.04

While it's not realistic for all variables to be zero, the intercept provides a *starting point* for the regression. (baseline sales)

Week Coefficient: β_1 = -8.26 (Effect of week on sales)

For each additional week, sales *decrease* by approximately 8.26 units or dollars, assuming all other variables remain constant. This reflects the seasonal nature of the drink, where Ive noticed sales tend to taper off after peak holiday periods.

Temperature Coefficient: $\beta_2 = -4.96$ (Effect of temperature on sales)

For every 1° increase in temperature, sales *decrease* by approximately 4.96 dollars, assuming all other variables remain constant. We can see that colder weather drives demand for hot seasonal drinks like the Sugar Cookie Almondmilk Latte, while warmer weather reduces sales.

HolidayWeek Coefficient: $\beta_3 = 37.92$ (Effect of holidays on sales)

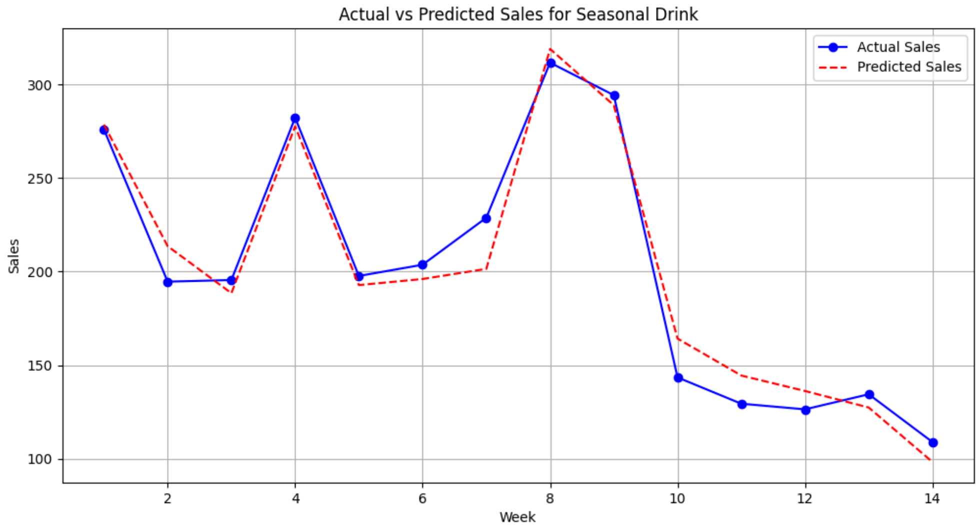
During holiday weeks, sales seemingly *increase* by approximately 37.92 dollars, assuming all other variables remain constant. This shows just how important holiday marketing and customer behavior are around the holidays, contributing to huge boosts in sales. This could mean the holiday time is perfect for launching festive drinks

DailyTraffic Coefficient: $\beta_4 = 0.2182$ (Effect of foot traffic on sales)

For each additional customer visiting the store, sales *increase* by approximately 0.2182 dollars, assuming all other variables remain constant. While the individual impact of a single customer is very small, as foot traffic accumulates it will have a much more meaningful effect on sales. For example: 1,000 extra customers would contribute roughly 218 dollars to sales.

To summarize: Using linear regression, I managed to model a simulation of Starbucks' seasonal drink sales based on factors such as Week, Temperature, HolidayWeek, and DailyTraffic. The regression vector β reveals some trends, providing a practical tool for predicting and optimizing seasonal drink performance; Key takeaways:

- The negative temperature coefficient shows that warmer weather reduces sales, suggesting prioritizing hot drinks during colder weeks.
- The importance of holidays indicates the need for more marketing during these weeks to maximize sales boosts.
- Sales decline each week after the holidays, suggesting that seasonal drinks should be phased out post-January.


Actual vs Predicted Sales for Seasonal Drink
Actual Sales
Predicted Sales
Sales
Week
300
250
200
150
100
2
4
6
8
10
12
14