Improving Sales Performance of Cafe Dumaguete

By: Mark Gelson E. Panganoron

Background

Cafe Dumaguete is a mid-sized coffee shop chain with three branches located in Dumaguete City. Over the past year, the cafe noticed inconsistent sales performance: some months show high revenue, while others drop significantly. Mr. Cofino, the owner of the Cafe, wants to understand what drives sales and how analytics can support better decision-making.

Dataset

	Date	Branch	Customers	Weather	Promotion	Top_Product	Sales
0	2025-01-01	South	173	Cloudy	Yes	Milk Tea	23659.42
1	2025-01-02	North	90	Sunny	No	Milk Tea	12978.45
2	2025-01-03	South	64	Cloudy	No	Milk Tea	8623.59
3	2025-01-04	South	94	Sunny	No	Pastry	12443.58
4	2025-01-05	North	114	Sunny	Yes	Americano	15906.84

• Daily sales (₱)
• Number of customers
• Weather conditions (sunny, rainy, cloudy)
• Day of the week
• Marketing promotions (yes/no)
• Most purchased product category (latte, frappe, pastry, etc.)

1. Descriptive Analytics

a). What are the monthly and weekly sales trends?

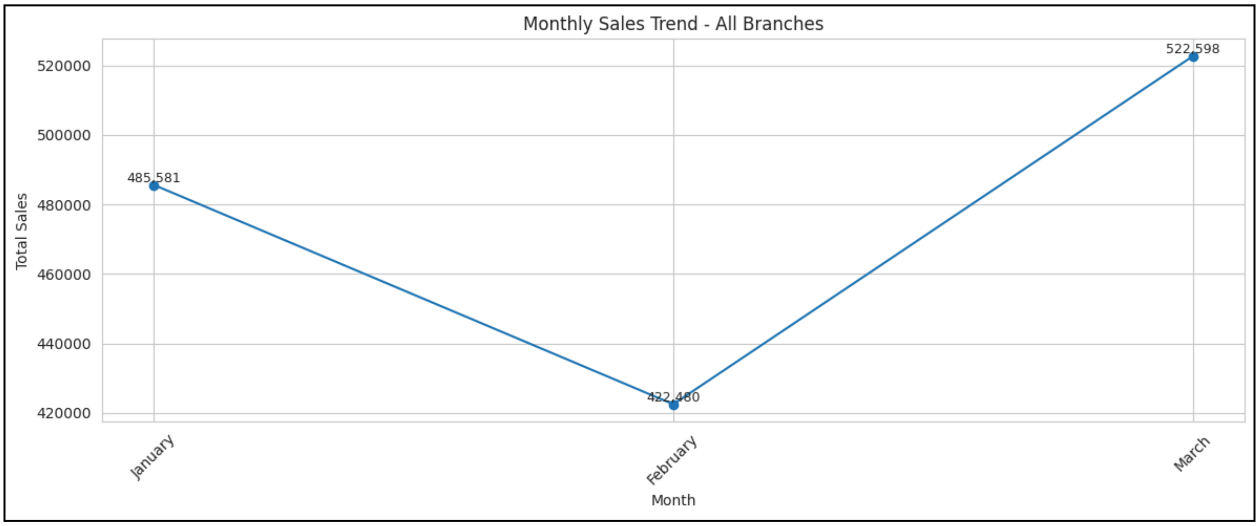

Figure 1. *Monthly Sales Trend (All Branches)*

During the first month of operation, total monthly sales across all branches were approximately ₱485,000. In the second month, sales decreased to around ₱422,000, then increased to ₱522,000 in the third month.

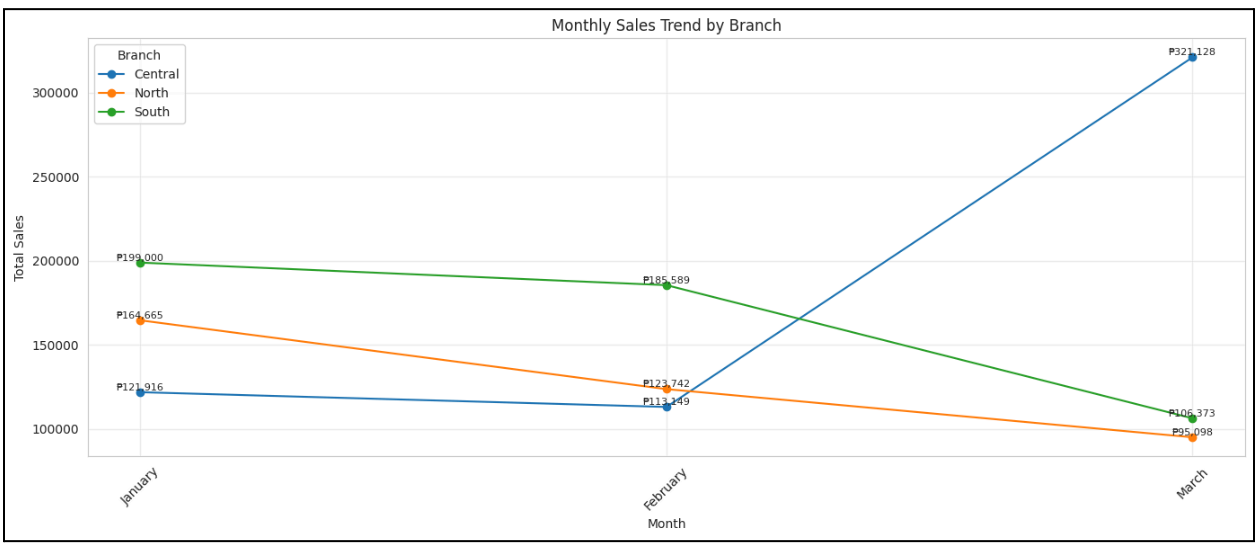

Figure 2. *Monthly Sales Trend (Per Branch)*

At the branch level, the South branch recorded the highest sales in the first month, followed by the North and Central branches. The same ranking persisted in the second month. In the third month, the Central branch rose to the top position, while sales in the South and North branches declined.

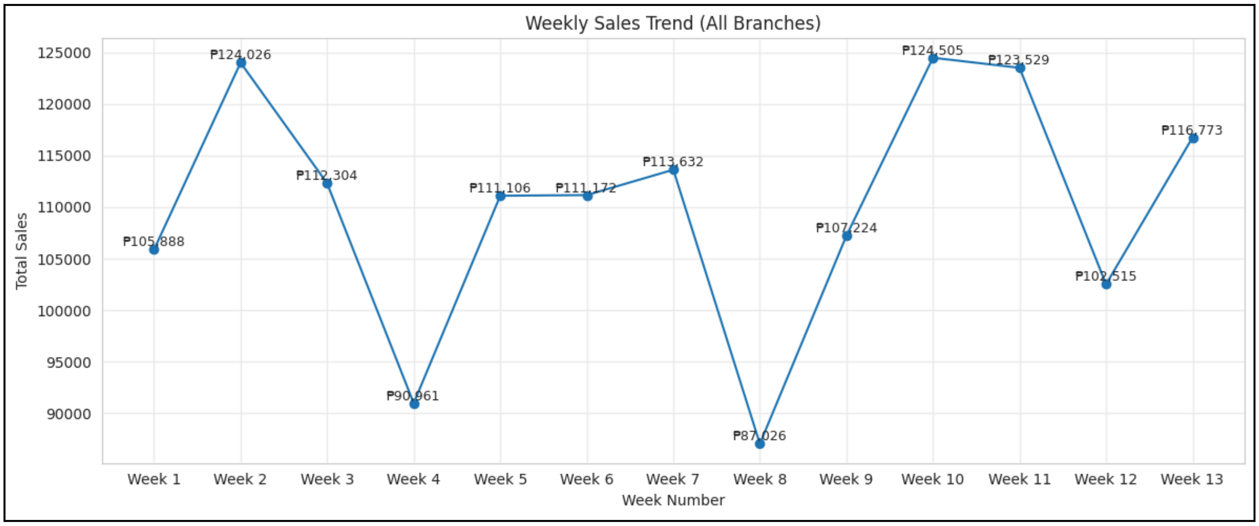

Figure 3. *Weekly Sales Trend (All Branches)*

The weekly sales trend shows fluctuations in sales, with increases and decreases occurring across different weeks. Sales peaked in weeks 2 and 10 at approximately ₱124,000 and reached a low in week 8 at ₱87,000. This indicates that total weekly sales vary substantially, with no consistent upward or downward trend over time. The pattern of peaks and troughs suggests that sales performance is uneven across weeks.

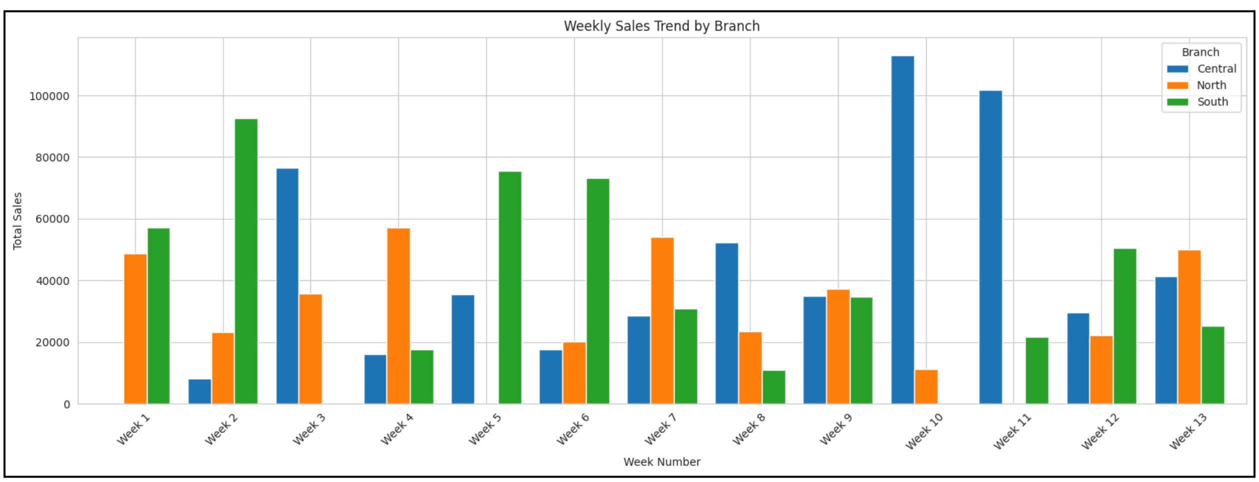

Figure 4. *Weekly Sales Trend (Per Branch)*

The weekly sales trend per branch shows that some branches recorded zero sales during certain weeks, indicating they were not operational during those periods. The Central branch consistently recorded sales in almost all weeks, while the South and

North branches had intermittent weeks with no sales. In addition, the Central branch had the highest weekly sales during weeks 10 and 11. These observations show that branch-level performance varies both in terms of sales volume and operational consistency across weeks.

b). Which branch performs the best?

Figure 5. *Branch Performance Overview*

This analysis compares branch performance across four categories: total sales, total customers, number of opening days, and number of promotions.

The Central branch leads in all four categories, recording the highest total sales, the highest number of customers, and the most operating days. In addition, it also conducted the most promotional activities. The South branch ranks second in all categories, showing moderate performance compared to the Central branch. The North branch consistently records the lowest values, with the fewest total sales, the least number of customers, the fewest operating days, and the lowest number of promotions.

c). Which products contribute most to total sales?

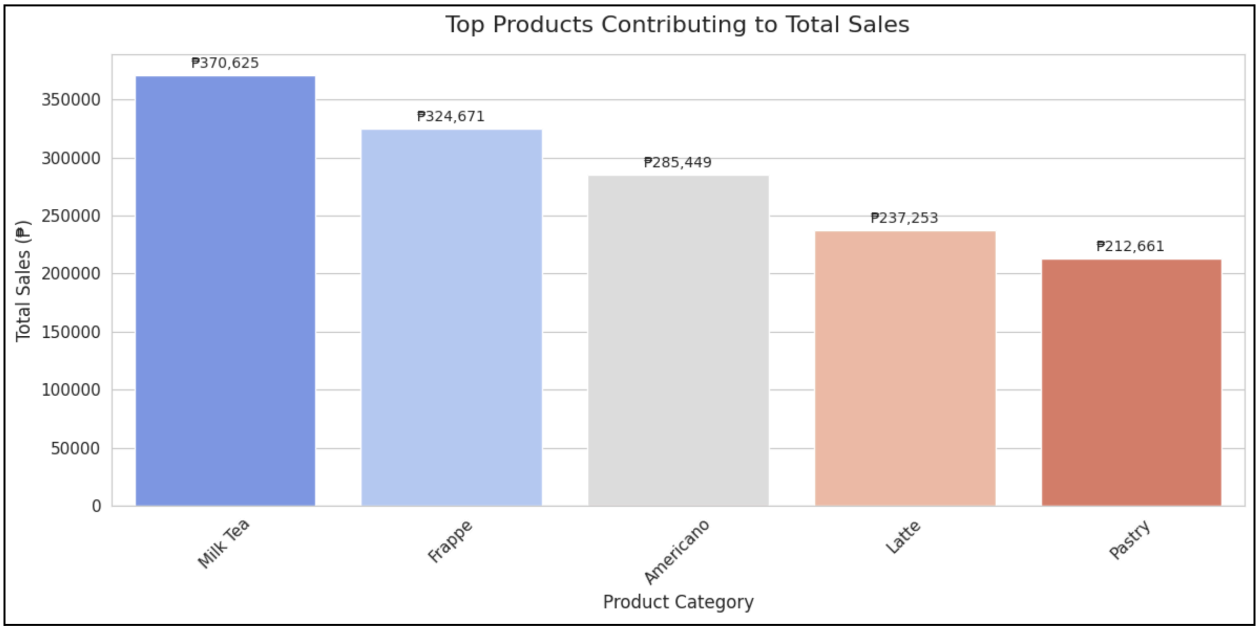

Figure 6. *Top Products Contributing to Total Sales (All branches)*

The data shows that Milk Tea is the top-performing product, contributing approximately ₱370,000 to total sales across all branches. Frappe ranks second in sales contribution, while Pastry has the lowest contribution, totaling around ₱212,000.

The distribution of sales among products indicates a clear hierarchy in customer preferences, with Milk Tea accounting for the largest share of revenue. Frappe and other beverage categories contribute moderately, whereas Pastry contributes the least to overall sales. This comparison highlights which product categories are driving total revenue across all branches.

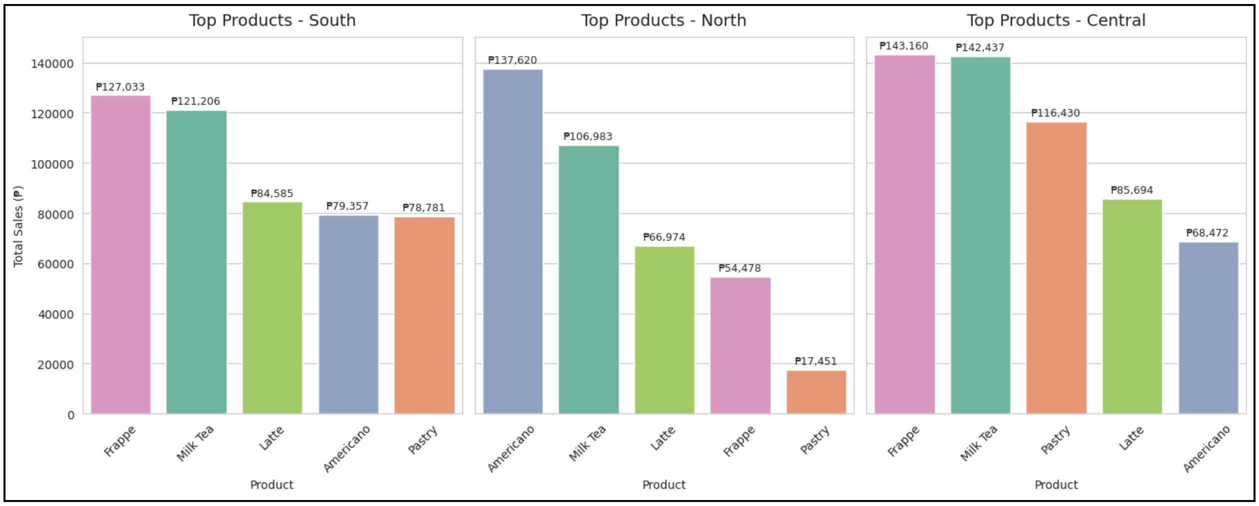

Figure 7. *Top Products Contributing to Total Sales (Per branch)*

The top-selling products vary across branches. In the South branch, Frappe is the highest contributor to sales, totaling approximately ₱127,000, followed by Milk Tea. In the North branch, Americano leads with around ₱137,000, followed by Milk Tea. In the Central branch, Frappe tops sales at approximately ₱143,000, closely followed by Milk Tea.

If Frappe had performed better in the North branch, it would have been the overall top product across all branches. Pastry consistently ranks last in two branches, while in the Central branch it ranks slightly higher.

These observations indicate that while Milk Tea and Frappe are generally strong performers, branch-specific preferences affect the ranking of top products, resulting in variation in contribution to total sales at the branch level.

d). Does weather affect customer volume or sales?

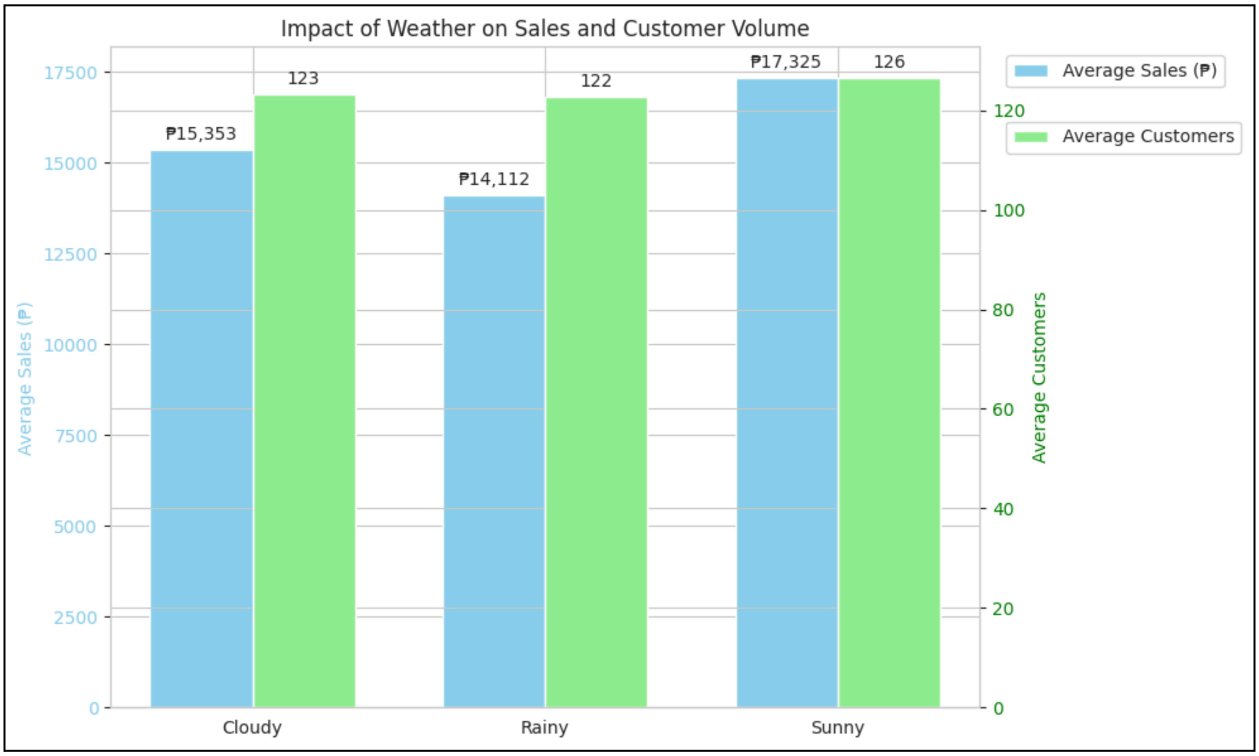

Figure 8. *Impact of Weather on Sales and Customer Volume*

The data shows variation in average sales across different weather conditions. Sunny days recorded the highest average sales, while rainy days had the lowest.In terms of customer volume, the differences are minimal. The average number of customers is 126 on sunny days, 123 on cloudy days, and 122 on rainy days. This indicates that while weather affects total sales, it has little impact on the number of customers visiting the branches.

e). How do promotions impact daily sales?

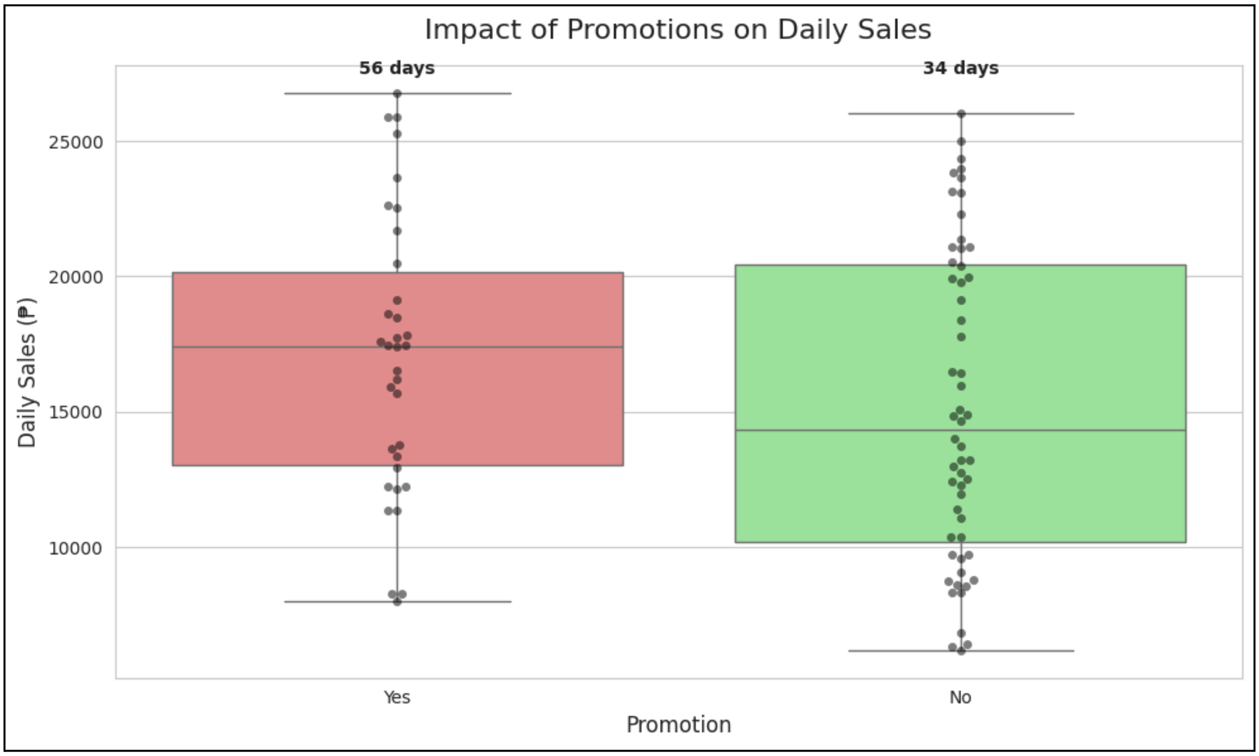

Figure 8. *Impact of Promotions on Daily Sales*

The figure shows that average daily sales are similar on days with and without promotions. While there is a difference in the number of days for each category, both exhibit comparable sales performance. This indicates that promotions did not noticeably affect daily sales during the observed period.

Summary

Sales performance across the branches shows a dynamic pattern, with some weeks and months experiencing fluctuations and occasional periods where certain branches were not operational. Among the branches, the Central branch consistently leads in sales, customer visits, operating days, and promotional activities, while the North branch records the lowest performance across all metrics. Milk Tea and Frappe emerge as the most popular products overall, yet each branch shows its own preferences, and Pastry consistently contributes the least to revenue. Weather plays a role in sales, with sunny days generating higher revenue, although the number of customers remains largely unaffected by changing conditions. Promotional activities, despite being conducted on some days, do not appear to significantly influence daily sales, as performance remains similar regardless of promotions.

2. Predictive Analytics

Model Trained: Linear Regression

[1]
```python
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error
from sklearn.preprocessing import OneHotEncoder
```

[2]
```python
df = pd.read_csv('cafe_dumaguete_dataset.csv')
df['Date'] = pd.to_datetime(df['Date'])
```

[3] 0s
```python
df.head()
```

	Date	Branch	Customers	Weather	Promotion	Top_Product	Sales
0	2025-01-01	South	173	Cloudy	Yes	Milk Tea	23659.42
1	2025-01-02	North	90	Sunny	No	Milk Tea	12978.45
2	2025-01-03	South	64	Cloudy	No	Milk Tea	8623.59
3	2025-01-04	South	94	Sunny	No	Pastry	12443.58
4	2025-01-05	North	114	Sunny	Yes	Americano	15906.84

Next steps: Generate code with df New interactive sheet

[7]
```python
categorical_features = ['Branch','Weather','Promotion']
df_encoded = pd.get_dummies(df, columns=categorical_features, drop_first=True)

```

[8]
```python
X = df_encoded.drop(columns=['Date','Sales','Top_Product'])
y = df_encoded['Sales']
```

[9]
```python
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, shuffle=False)
```

[10]
```python
model = LinearRegression()
model.fit(X_train, y_train)
```

```
LinearRegression
LinearRegression()
```

[11]

```python
y_pred = model.predict(X_test)
```

[13]

```python
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print(f"Mean Absolute Error (MAE): ₱{mae:,.2f}")
print(f"Root Mean Squared Error (RMSE): ₱{rmse:,.2f}")
```

```
Mean Absolute Error (MAE): ₱1,106.36
Root Mean Squared Error (RMSE): ₱1,346.38
```

[14]

```python
last_week = X.tail(7)
next_week_pred = model.predict(last_week)
print("\nPredicted Sales for Next Week:")
print(next_week_pred)
```

```

Predicted Sales for Next Week:
[ 9058.19872246 10808.97068409 24185.94718165 24383.03426604
 19822.10185871 17072.79026406 19744.6022366 ]
```

[15]

```python
coeff_df = pd.DataFrame({
    'Feature': X.columns,
    'Coefficient': model.coef_
})
coeff_df['Absolute'] = coeff_df['Coefficient'].abs()
strong_predictors = coeff_df.sort_values(by='Absolute', ascending=False)
print("\nStrongest Predictors of Sales:")
print(strong_predictors.head(10))
```

```

Strongest Predictors of Sales:
         Feature  Coefficient     Absolute
4  Weather_Sunny  1509.526037  1509.526037
5  Promotion_Yes  1089.664474  1089.664474
3  Weather_Rainy -1012.164852  1012.164852
1   Branch_North   210.625788   210.625788
0      Customers   120.751778   120.751778
2   Branch_South    76.335307    76.335307
```

[16]

```python
import matplotlib.pyplot as plt

plt.figure(figsize=(12,6))
plt.plot(y_test.values, label='Actual Sales', marker='o', color='blue')
plt.plot(y_pred, label='Predicted Sales', marker='x', color='red')
plt.title("Actual vs Predicted Daily Sales")
plt.xlabel("Test Data Points")
plt.ylabel("Sales (₱)")
plt.legend()
plt.grid(alpha=0.3)
plt.tight_layout()
plt.show()
```

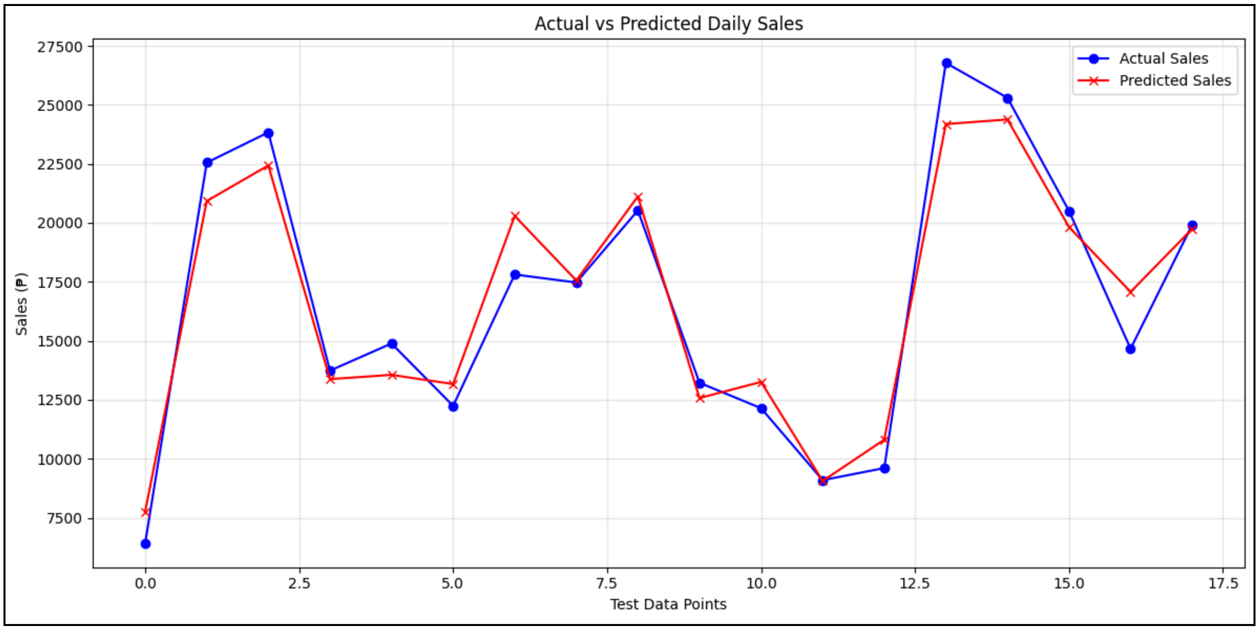

Figure 9. *Actual vs Predicted Daily Sales*

Our Linear Regression model effectively captured the trends in daily sales, as evidenced by the performance metrics. The Mean Absolute Error (MAE) of ₱1,106.36 indicates that, on average, our predictions were off by approximately ₱1,106.36 from the actual sales. The Root Mean Squared Error (RMSE) of ₱1,346.38 provides a measure of the magnitude of the errors, giving more weight to larger errors. Both metrics suggest that the model offers a reasonably accurate prediction of daily sales.

The 'Actual vs Predicted Daily Sales' plot visually confirms the model's performance. While there are some discrepancies, the predicted sales closely follow the pattern of actual sales, indicating a good fit. The model is generally able to track the peaks and troughs of sales, which is crucial for forecasting.

Strongest Predictors of Sales:

The analysis of the model's coefficients revealed several key factors influencing daily sales:

- Weather_Sunny was the strongest positive predictor, suggesting that sunny weather significantly boosts sales. For each sunny day, sales increased by approximately ₱1,509.53.
- Promotion_Yes also had a substantial positive impact, indicating that running promotions leads to higher sales. Promotions are associated with an increase of roughly ₱1,089.66 in sales.

- Conversely, Weather_Rainy was a strong negative predictor, meaning rainy weather tends to decrease sales by about ₱1,012.16.
- Branch_North, Customers, and Branch_South also played a role, though with smaller coefficients compared to weather and promotions. An increase in customer count naturally correlates with higher sales.

3. Prescriptive Analytics

Strategies to Increase Sales by 10% Next Month

Based on the analysis discussed above (daily sales, branch performance, product contribution, weather effects and the predictive model), the following strategies can achieve a 10% increase in sales next month.

1. **Regular Operational Days.**
   Consistent operations can significantly improve daily sales performance. Some branches experienced weeks with no sales, resulting in lost revenue. Ensuring all branches operate consistently will increase customer volume and foster customer loyalty.

2. **Maximize and Optimize Promotions on special occasions.**
   Promotional days did not show a significant effect on daily sales, indicating that promotion execution should be improved. Focus promotions on special dates or high-traffic periods rather than spreading them frequently, maximizing their impact on sales.

3. **Focus on best-selling products.**
   Milk Tea and Frappe consistently drive the highest sales across branches, except in the North branch where Americano leads. Promotional campaigns highlighting these best-sellers can maximize revenue. Additionally, consider targeted promotions for lower-performing products, such as Pastry, during off-peak hours to reduce waste and boost sales.

4. **Leverage Predictive Insights.**
   The Linear Regression model indicates that weather, promotions, branch, and customer volume are strong predictors of sales. Use daily weather forecasts to plan inventory and staffing, and schedule marketing or promotional efforts around expected high-traffic days. Provide additional support to branches with historically lower sales.

5. **Follow the best practices of the best performing branch**
   The Central branch consistently achieves the highest performance. Underperforming branches can adopt strategies from the Central branch, such as extended operating hours, targeted promotions, optimized product placement, and enhancements to store ambiance or location-specific offerings

4. The Modelling Lifecycle Reflection

Reflection on the Analytics Process

1. **Business Understanding**
   The main objective was to analyze the sales performance of Cafe Dumaguete, identify the key factors driving performance, and develop strategies to improve sales.

2. **Data Understanding**
   The dataset consisted of daily records across three branches, including the date, branch, number of customers, weather conditions, promotions, top-selling product, and daily sales. This provided a comprehensive view of operational and sales patterns.

3. **Data Preparation**
   The dataset was largely ready for analysis. Preparation steps included converting dates to proper datetime objects, ensuring numeric values for sales and customers, and encoding categorical variables such as promotions and weather into numerical formats for modeling.

4. **Modeling**
   A Linear Regression model was trained to forecast next week's sales. Predictor variables included customers, weather, promotions, branch, and top product. The model identified key factors positively correlated with sales as well as variables that negatively impacted sales. These insights are actionable for operational planning and marketing strategies.

5. **Evaluation**
   The model was evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), indicating reasonably accurate predictions. A line chart comparing actual vs predicted sales confirmed that the model effectively captured sales trends, tracking peaks and troughs across the testing period.

6. **Deployment (hypothetical recommendation)**
   The model can be used by Cafe Dumaguete to forecast daily sales, enabling data-driven decisions for inventory, staffing, promotions, and operational planning to improve overall performance.