Coffee Shop Sales: Model Results and Analysis

This document outlines the results of two predictive models built to analyze the coffee shop sales data:

1. **A Decision Tree Classifier** to predict sales by time of day.
2. **A KMeans Clustering Model** to segment customers for targeted marketing.

Model 1: Sales Prediction (Decision Tree)

1. Classification Report

- **Accuracy: 27%** The model's overall accuracy is approximately 27%. While this figure appears low, it is important to consider the following factors:
  - **Task Difficulty:** Predicting one of eight distinct coffee types is an inherently complex classification task.
  - **Insufficient Features:** The model only used time-based features (hour, day of week, month). Customer preference is highly personal and not solely dictated by the time of day.
  - **Data Imbalance:** As shown by the f1-scores, the model is slightly better at predicting common items like 'Americano' (0.33) and 'Americano with Milk' (0.34), but performs very poorly (0.00) on rare items like 'Espresso'.

2. Example Prediction

- **Prediction:** `Example Prediction for 10 AM on a Friday in March: Latte`
- **Interpretation:** The model predicts that a customer at 10 AM on a Friday in March is most likely to buy a **Latte**. This aligns with our previous heatmap analysis, which identified 'Latte' and 'Americano with Milk' as popular choices during this peak morning hour.

3. How to Use This Model

This model should be viewed as an **automated supplement to the sales heatmaps**. It can be used to provide a quick "best guess" for what to prepare for an upcoming hour. Its low accuracy means it should **not be exclusively relied upon**, but it can serve as a useful "guide" for production planning and inventory management.

Model 2: Customer Clustering (KMeans)

To segment customers, we used an RFM (Recency, Frequency, Monetary) model based on the following features:

- **Recency:** `days_since_last_visit`
- **Frequency:** `total_visits`
- **Monetary:** `total_spent`

1. The "Elbow Plot" (kmeans_elbow_plot.png)

This plot shows the model's "Inertia" (a measure of cluster density). We look for the "elbow," the point where adding more clusters yields diminishing returns. Based on the plot, a **k-value of 4** was selected to provide a granular and meaningful segmentation of the customer base.

2. Cluster Analysis (Group Means)

This table shows the most critical output: the average characteristics of our four customer segments. These values represent the "profile" of each group.

Cluster	days_since_last_visit	total_visits	total_spent
0	272.09	1.58	50.79
1	99.88	1.88	60.24
2	71.50	74.75	2259.01
3	82.66	15.77	497.53

3. Interpretation of Customer Segments

Cluster 0: "Churned / One-Time Customers"

- **Profile:** This group has not visited in an average of **272 days** (high recency), has a very low visit frequency (avg. 1.58 visits), and the lowest monetary value.
- **Marketing Strategy:** This group is almost entirely lost. A deep-discount "we miss you" campaign (e.g., 50% off) could be attempted, but marketing budget should be limited here.

Cluster 1: "Recent New Customers / At-Risk"

- **Profile:** This group is composed of new or low-frequency customers who visited more recently than Cluster 0 (avg. 100 days ago) but have not formed a habit (avg. 1.88 visits).
- **Marketing Strategy:** This is the **"At-Risk"** group. The goal is to convert them into regulars. Target them with a "20% off your second visit" or "bring a friend" coupon to encourage a return visit and move them into Cluster 3.

Cluster 2: "VIPs / Super-Loyal"

- **Profile:** This is a small but elite group, defined by extremely high frequency (avg. **74.75 visits**) and extremely high spending (avg. **$2259.01**). They visit relatively often (avg. 71.5 days).
- **Marketing Strategy:** These are your super-fans. **Do not give them discounts**; they don't need them. Give them **status**: free size upgrades, first access to new products, or holiday cards. Their loyalty is priceless and should be rewarded with recognition, not discounts.

Cluster 3: "Loyal Regulars"

- **Profile:** This is your core, high-value customer base. They are frequent (avg. **15.77 visits**), spend significantly (avg. **$497.53**), and visit regularly.

- **Marketing Strategy:** This is the "morning coffee" crowd. A "Buy 5, Get 1 Free" loyalty card or a