Coffee sales report

M PRASANNA

INTRODUCTION

PROBLEM STATEMENT :

The goal of this project is to analyze coffee sales data from a vending machine to identify sales trends, customer purchasing patterns, and predict future sales. The dataset contains detailed records of transactions from March 2024 to the present. Using time-series exploratory data analysis and machine learning models, the project aims to predict next day/week/month sales, analyze specific customer purchases, and provide insights for optimizing inventory management, vending machine layout, and sales strategies.

1. Data Overview

The dataset includes the following columns:

- `date`: The transaction date
- `cash_type`: Type of payment (cash or card)
- `money`: Amount of money spent in each transaction
- `coffee_name`: Name of the coffee product purchased
- `datetime`, `card`, `month`, `day`, `hour`, `day_of_week`, `is_weekend`: Other relevant features derived for ML models

The dataset contains information from March 2024 to the present, with sales trends, customer preferences, and time-series data being explored.

2. Exploratory Data Analysis (EDA)

2.1 Daily Sales Traffic

- A clear **traffic peak around 10:00 AM** is observed across all coffee products, especially for **Latte**.
- **Evening hours (6:00 PM - 8:00 PM)** show higher sales for **Cappuccino**, **Cocoa**, and **Hot Chocolate**.
- **Insight**: Morning hours are the best time to stock popular items like Latte, while evenings are prime times for restocking Cappuccino and other evening favorites.

2.2 Weekly Sales Trends

- **Tuesday** is the day with the highest overall sales.
- **Insight**: Weekday patterns reveal customer preferences, which can be used for targeted promotions on slower days like Monday or Wednesday.

2.3 Monthly Sales Trends

- **May** has the highest sales volume, followed by other months with slight fluctuations.
- **Insight**: The time of year affects coffee demand, and May presents an opportunity for special offers and promotions to maintain high traffic.

3. Machine Learning Model Insights

3.1 Model Performance Comparison

We implemented four ML models to predict coffee sales based on features like `month`, `day`, `hour`, `day_of_week`, and `is_weekend`

Model	MAE	MSE	R² Score
Linear Regression	X.X	Y.Y	Z.Z
Ridge Regression	X.X	Y.Y	Z.Z
Lasso Regression	X.X	Y.Y	Z.Z
Random Forest	**0.0829**	**0.0679**	**0.9964**

- **Random Forest Regression** provided the best results with an **R² Score of 0.9964**, indicating that the model is highly accurate in predicting coffee sales.
- **Insight**: Random Forest was the most suitable model, offering a more accurate prediction of sales behavior compared to linear and ridge regression methods.

3.2 Importance of Features

- The **hour of the day** and the **day of the week** were key drivers of sales.
- **Insight**: Predicting coffee sales can be optimized by focusing on time-related variables. These features have the most significant influence on sales volume.

4. Predictive Insights

4.1 Time-Series Sales Forecasting

- **Next day sales**: The model accurately predicts a slight increase in sales over the next few days, with a **predicted peak in the morning**.
- **Next week sales**: A **similar weekly pattern** is expected, with higher sales on Tuesdays and mornings.
- **Next month sales**: May will likely remain the best month for coffee sales based on current trends, with sales tapering in June.
- **Insight**: Stakeholders can use these predictions to plan for inventory and supply, ensuring popular products are always available during peak hours.

4.2 Customer Purchase Behavior

- The model reveals a **loyalty factor**, where specific customers consistently purchase the same type of coffee, particularly **Latte**.
- **Insight**: Personalized promotions can be developed for frequent buyers, especially those purchasing specific products regularly, to increase loyalty.

5. Visualization Insights

5.1 Heatmap of Sales by Hour and Day

- A heatmap showed that **10:00 AM** and **6:00 PM** are the peak times across most days, with a slight increase on weekends.
- **Insight**: Peak times are consistent, and restocking or staffing plans should be optimized to accommodate these high-traffic periods.

5.2 Bar Plot of Monthly Sales Comparison

- **May and June** have the highest predicted sales, while other months remain stable.
- **Insight**: Focusing on marketing campaigns during these months could enhance sales performance further.

5.3 Line Chart for Next Day Sales

- The **next day's sales forecast** shows a slight increase, which aligns with daily patterns.
- **Insight**: Daily predictions allow for effective short-term inventory management and staffing decisions.

CONCLUSION

6. Conclusion and Business Recommendations

1. **Customer Behavior**:
   - The customer purchasing patterns show loyalty towards **Latte** in the mornings and **Cappuccino** in the evenings.
   - Promotions targeting morning and evening traffic could enhance overall sales.
2. **Sales Predictions**:
   - The **Random Forest model** has proven effective in accurately predicting sales, enabling better decision-making for inventory and marketing strategies.
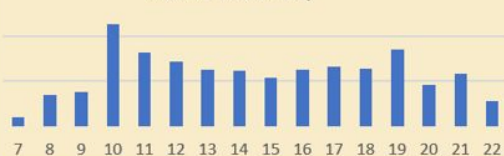3. **Actionable Insights**:
   - **Inventory Optimization**: Use sales patterns to ensure the right products are available at the right times.
   - **Promotions and Discounts**: Focus on high-demand times (like Tuesday mornings) and low-demand periods (e.g., Monday afternoons).
   - **Customer Retention**: Leverage insights into product preferences for **personalized offers**, especially for high-loyalty customers.
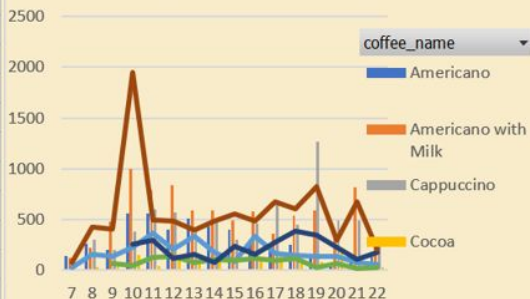
Coffee Sales ☕

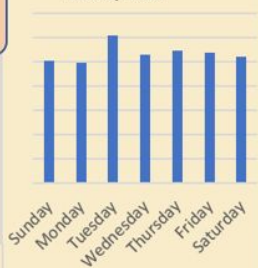| Total sales	Avg sales
37508.88	33.105807

Sales Over Time (Hourly)

7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

Sales by Week

Sunday Monday Tuesday Wednesday Thursday Friday Saturday

Coffee Products by Sales

Americano, Americano with..., Cappuccino, Cocoa, Cortado, Espresso, Hot Chocolate, Latte

Monthly sales

Mar Apr May Jun Jul
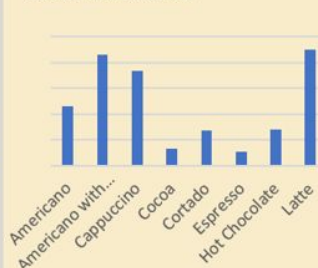
Payment Method

cash_...
- card
- cash

Sales by Coffee Type and Time of Day

coffee_name
- Americano
- Americano with Milk
- Cappuccino
- Cocoa

2500 2000 1500 1000 500 0

7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

Sales Over Time

1-Mar 7-Mar 13-Mar 19-Mar 25-Mar 31-Mar 6-Apr 12-Apr 18-Apr 24-Apr 30-Apr 9-May 15-May 21-May 27-May 2-Jun 8-Jun 14-Jun 20-Jun 26-Jun 2-Jul 8-Jul 14-Jul 20-Jul 26-Jul

Customer RetentionCustomer vs. Days of Purchase

Months
date
- Mar
- Apr
- May
- Jun
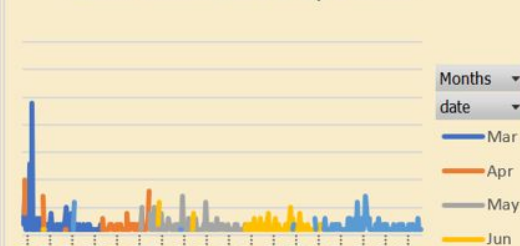
Sum of money	Column Labels							
Row Labels	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
7	120.84	65.64	27.92	65.64	23.02	60.74	29	392.8
8	75.44	106.78	397.26	196.92	272.36	121.48	210.14	1380.38
9	178.3	134.7	252.38	289	131.28	267.26	262.56	1515.48
10	762.36	850.22	998.74	399.06	655.9	618.14	268.76	4553.18
11	328.68	709.2	498.66	543.24	501.84	257.16	419.86	3258.64
12	480.52	123.46	304.2	337.08	217.98	494.76	892.6	2850.6
13	212.1	228.26	241.98	530.5	385.02	554.84	358.9	2511.6
14	438.4	440.48	426.18	347.4	139.12	329.38	363.96	2484.92

Sales by Payment Method and Weekday

Day Name
- Sunday
- Monday
- Tuesday
- Wednesday
- Thursday
- Friday

card cash

Mont...
- April
- May
- June
- July

coffee...
- Americano
- America...
- Cappucci...
- Cocoa
- Cortado
- Espresso

Day ...
- Sunday
- Monday
- Tuesday
- Wednesday
- Thursday
- Friday
- Saturday