COFFEE SALES

Created by Rajat Singh

OVERVIEW

ABOUT DATASET

This dataset contains detailed records of coffee sales from a vending machine.

The vending machine is the work of a dataset author who is committed to providing an open dataset to the community.
It is intended for analysis of purchasing patterns, sales trends, and customer preferences related to coffee products

The dataset spans from March 2024 to Present time, capturing daily transaction data. And new information continues to be added

TASKS:

- Time Series Exploratory Data Analysis
- Next day/week/month sales
- Specific customer purchases

FLOW CHART

Data Ingestion

- Load Data → Initial Observations → Identify Missing Values and Categorical Features

EDA

- Analyze Sales Trends → Examine Payment Methods → Identify Popular Coffee Types

Preprocessing

- Handle Missing Values → Encode Categorical Features → Split Data into Training and Testing Sets

Model Training

- Test Models (Linear Regression, Decision Tree, Gradient Boosting, etc.)

HyperParameter Tuning

- Perform Hyperparameter Tuning → Select Best Model

Model Evaluation

r2_score or MAE

Streamlit

AWS EC2

Docker image

CI / CD

Github Action

Github Repo

Github/workflow

yaml

INTRODUCTION

- Objective: The project's main goal is to predict coffee sales, allowing for data-driven decisions regarding inventory, promotions, and product offerings.
- Approach: A structured workflow was implemented from initial data loading, through detailed analysis, and model training, to the deployment of a user-friendly prediction app.

PROJECT OVERVIEW

- Dataset Overview: The dataset includes sales records with features such as date, datetime, cash_type, card, money, and coffee_name.
- Initial Findings: Missing values were found in the card column, and several columns are categorical, requiring preprocessing for use in machine learning models

DATA CLEANING

A

Missing Values: The card column had 89 missing values, which were replaced with the mode (most frequently occurring value) to maintain data integrity.

B

Categorical Encoding: Variables like cash_type and coffee_name were encoded to numerical formats using Label Encoding or OneHot Encoding, making them suitable for model training.

C

Data Split: The dataset was split into training and testing sets to validate model performance effectively

EXPLORATORY DATA ANALYSIS

1

Sales Trends: A time series analysis showed specific patterns in coffee sales, with noticeable peaks and valleys, indicating periods of high and low demand.

2

Payment Methods: Observations revealed a preference for certain payment methods, which could be leveraged to optimize payment processing strategies

3

Popular Coffee Types: Analysis of coffee_name indicated which coffee products were top sellers, valuable for inventory management and targeted promotions.

4

Key Insight: Strong correlation between sales volume and certain times/dates suggests the potential for seasonality or daily trends in coffee sales

MODEL TRAINING

1

Model Comparison: Various regression models, including Linear Regression, Decision Trees, and Gradient Boosting, were tested.

2

Hyperparameter Tuning: Techniques such as RandomizedSearchCV helped optimize model parameters, enhancing accuracy without overfitting.

3

Best Model Selection: The Gradient Boosting model provided the best predictive performance, showing a balance between training accuracy and generalizability.

MODEL EVALUATION

A

Performance Metrics: The chosen model achieved an R-squared value of Numeric Values, indicating a strong fit to the data, and a low Mean Absolute Error, highlighting its reliability in predicting sales figures.

B

Comparative Analysis: The Gradient Boosting model outperformed other models in terms of both accuracy and stability, confirming its suitability for the task.

PREDICTION APPLICATION

1

User Friendly Dashboard: The app provides an easy to use interface for sales predictions, including visualizations that make complex data accessible.

2

Key Functionalities: Users can input specific details to predict sales outcomes, with additional options to visualize trends, helping stakeholders make informed business decisions.

3

Impact on Decision Making: The app enables quick insights into future sales patterns, supporting proactive business strategies like inventory management and marketing campaigns

CONCLUSION AND FUTURE WORK

A

Summary of Results: The model accurately predicts sales, providing actionable insights into coffee sales trends and product performance.

B

Future Enhancements: Future work could incorporate more features (e.g., weather, holidays) to further refine predictions and add insights into seasonal trends.

PROJECT SUMMARY

On the other hand, Cocoa and Espresso aren't selling much at all, so they could really use some advertising to attract more customers. Cortado is also not doing well in sales.

Top Sellers: Americano with Milk, Latte, and Cappuccino lead in sales, while Cocoa and Espresso underperform, suggesting a need for targeted advertising

Customer traffic peaks around 10:00 AM, especially for Latte. In the evening, between 6:00 PM and 8:00 PM, beverages like Cappuccino and Hot Chocolate become more popular

Sales Trends: Sales for Americano with Milk and Latte show an upward trajectory, particularly as colder months approach after May, when coffee consumption typically increases.

Weekends generate higher overall sales, indicating a preference for coffee consumption during these times.

Americano is most popular on Mondays, while Latte sees peak sales on Thursdays

KEY INSIGHTS

These findings can support decisions in:
Inventory management to avoid stockouts.
Optimizing vending machine layouts for higher engagement.
Pinpointing the best restocking times to meet demand more effectively.

THANK YOU