Project Title: Coffee Sales Data Analysis

Internship Role: Data Analyst Intern

Project Overview: This project aims to analyze coffee sales data from a vending machine to uncover insights into customer purchasing patterns, sales trends, and product preferences. The findings will assist in optimizing inventory management, marketing strategies, and business operations.

Tools & Technologies:

- **Programming Languages:** Python
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- **Database:** SQL
- **Other Tools:** Microsoft Excel

Project Objectives:

1. Perform Time Series Exploratory Data Analysis (EDA).
2. Predict next day/week/month sales using machine learning models.
3. Analyze specific customer purchase behaviors.
4. Provide actionable business recommendations.

Dataset Information:

- **Source:** Provided dataset capturing daily transactions from March 2024 to the present.
- **Attributes:** Date, Store, Product, Sales, Quantity, Price, Cash Type, Card, Customer ID.

Project Workflow:

1. **Data Collection:**
    - Load the dataset using Python (pandas).
    - Understand data structure and initial statistics.
2. **Data Cleaning & Preprocessing:**
    - Handle missing values (e.g., impute with median/mode).
    - Convert date fields to datetime objects.
    - Remove outliers using Z-score.
    - Feature engineering: Extract month, year, day of the week, and hour from date/time.
3. **Exploratory Data Analysis (EDA):**
    - Sales trends over time (line plots).

- Sales distribution by product, store, and customer segment.
- Identify peak sales hours and days.

4. **Machine Learning Modeling:**
    - **Model Selection:** Linear Regression, Decision Tree Regressor.
    - **Feature Engineering:** One-hot encoding for categorical variables.
    - **Model Training:** Train-test split, model fitting.
    - **Evaluation:** Use metrics like Mean Squared Error (MSE) and R-squared (R^2).
5. **Customer Purchase Behavior Analysis:**
    - Segment customers based on purchase frequency and preferences.
    - Analyze purchase trends for specific customer IDs.
6. **Business Insights & Recommendations:**
    - Identify best-selling products and peak hours.
    - Recommend inventory adjustments based on demand forecasts.
    - Suggest promotional strategies for underperforming products.

Deliverables:

- Cleaned dataset with derived features.
- Data visualizations highlighting key trends.
- Predictive model for future sales forecasts.
- Comprehensive project report summarizing findings and business recommendations.

Learning Outcomes:

- Hands-on experience with data analysis and visualization tools.
- Application of machine learning models for sales forecasting.
- Enhanced understanding of business data interpretation.

Here are the key observations from the dataset:

- **Missing Values:** The card column has 89 missing values, which likely correspond to cash transactions.
- **Unique Values:**
- **Payment Methods (cash_type):** 2 (card, cash)

- **Coffee Types (coffee_name):** 8, with *Americano with Milk* being the most popular (268 transactions).
- **Transactions:** 1133 entries, spanning 150 unique dates.
- **Revenue Range (money):** From 18.12 to 40.00, with an average of 33.11.

I'll proceed with data cleaning (handling missing values), feature engineering (e.g., extracting time features), and then dive into exploratory data analysis (EDA).

The dataset has been cleaned, and new features have been added:

- **month**: Extracted in YYYY-MM format.
- **weekday**: Day of the week (e.g., Friday).
- **hour**: Transaction hour.

References:

- Dataset link and code repositories for reference.
- Documentation of libraries and tools used in the project.