Project: Dirty Cafe Sales Data Analysis

Overview
The *Dirty Cafe Sales Dataset* contains 10,000 rows of synthetic sales transactions from a fictional cafe. Unlike clean, curated datasets, this dataset is intentionally **messy**, containing missing values, inconsistent entries, and errors across multiple columns. Its purpose is to simulate real-world business data challenges and provide a hands-on opportunity for practicing **data cleaning, wrangling, and exploratory data analysis (EDA)**.

Objectives
1. Data Cleaning & Preprocessing
- Handle missing values in categorical and numerical columns.
- Standardize inconsistent entries (e.g., payment methods, item names, locations).
- Correct invalid or corrupted values (e.g., transaction dates, quantities, or prices).

2. Feature Engineering
- Derive new features such as daily sales, average spend per customer, and item popularity.
- Compute interdependent values: total_spent = quantity × price_per_unit.
- Create time-based features for trend analysis (e.g., month, weekday).

3. Exploratory Data Analysis (EDA)
- Identify sales trends over time.
- Compare item popularity and revenue contributions.
- Analyze customer purchasing behavior by location and payment method.
- Highlight anomalies and outliers in transaction records.

Skills Practiced
- **Data Cleaning:** Handling nulls, fixing invalid entries, imputing values.
- **Data Wrangling:** Standardizing formats, restructuring tables, feature extraction.
- **Exploratory Data Analysis (EDA):** Statistical summaries, visualizations, and trend identification.
- **Feature Engineering:** Deriving meaningful insights from raw transaction data.

Expected Outcomes
By the end of this project, the dataset will be fully cleaned and structured, allowing for:
- A **cleaned sales dataset** ready for analysis or modeling.
- **Visual insights** into cafe sales performance, customer behavior, and item demand.
- A **reproducible data pipeline** showcasing end-to-end handling of raw business data.