**☕ Cafe Sales Analysis – Project Documentation**

**1. Business Problem**

A local café business wants to better understand its sales trends, customer preferences, and payment behaviors to improve operations and increase profitability. However, the available sales data is messy, inconsistent, and contains numerous errors, making it difficult to derive clear insights without thorough data preprocessing and analysis.

**2. Objective**

The main goal of this project is to:

* Clean and preprocess the dirty café sales dataset
* Explore patterns in customer behavior and product performance
* Identify peak sales periods and high-performing products
* Develop an interactive Power BI dashboard to visualize insights for business decision-making

**3. Tools Used**

* **Python (Google Colab)**:
  + For data cleaning, preprocessing, exploratory data analysis (EDA), and feature engineering
* **Power BI**: For building an interactive dashboard to visualize insights and support decision-making

**4. Business Questions**

To help the café improve its operations and sales, the following business questions were addressed:

* What are the top-selling and least-selling items?
* Which payment methods are most commonly used?
* Which days and months generate the highest sales?
* What is the sales trend across different locations and times?
* How does customer behavior vary throughout the week?

**5. Dataset**

* **Source**: Kaggle
* **Name**: dirty\_cafe\_sales.csv
* **Size**: 10,000 rows × 8 columns
* **Nature**: Synthetic but intentionally messy with missing values, incorrect entries, and inconsistent formatting

**6. Important Features in the Dataset**

| **Feature Name** | **Description** |
| --- | --- |
| Transaction ID | Unique identifier for each transaction |
| Item | Name of the item purchased (e.g., Coffee, Cake) |
| Quantity | Number of units purchased |
| Price Per Unit | Price of a single unit |
| Total Spent | Total amount spent (Quantity × Price) |
| Payment Method | Mode of payment (e.g., Cash, Credit Card) |
| Location | Location of the transaction (In-store or Takeaway) |
| Transaction Date | Date when the transaction occurred |

**7. Data Preprocessing**

Performed using Python in Google Colab:

* **Handled Missing Values**: Replaced missing or invalid entries like UNKNOWN, None, and ERROR
* **Removed Duplicates** and corrected formatting issues
* **Standardized Data Types** for numeric and date columns
* **Feature Engineering**:
  + Created Month and Day\_Name columns for temporal analysis
  + Created Is\_Month\_End column to identify end-of-month transactions
* **Exported Cleaned Dataset** to Excel for use in Power BI

**8. Report (Power BI Dashboard)**

Created an interactive dashboard featuring:

* **KPI Cards**: Total Revenue, Total Transactions
* **Bar & Line Charts**:
  + Top Selling Items
  + Monthly & Daily Sales Trends
  + Sales by Location
* **Pie Chart**: Payment Method Breakdown
* **Heatmap & Filters**: Day-wise and item-wise performance, dynamic slicers for exploration

**9. Key Insights**

* **Top Products**: Coffee, Sandwich, and Cake are the most popular and profitable items
* **Sales Trends**: Highest sales occur on weekends and month-ends
* **Payment Preferences**: Majority of customers use digital payment methods
* **Customer Behavior**: Sales dip midweek—potential opportunity for weekday promotions
* **Operational Clarity**: In-store sales outperform takeaway in both volume and value