

|  |  |
| --- | --- |
| Report of Café Sales | |
| Client name | DeepSeek |
| ANALYst name | Ftrah Rahmi Putri |
| Date | 27-29 Aug 2025 |

## Executive summary

Based on the Café Sales dataset, we identified that **Juice** is the best-selling item, with total sales of 6,435 units throughout 2023, and daily and monthly average sales of 69 and 178 units, respectively. Customers most frequently used **digital wallets** as their payment method and preferred **takeaway** over enjoying their drinks in-store.

Sales distribution throughout 2023 was generally stable with several growth spikes, resulting in total revenue of **$87,768**. However, there were **2,845 issues** in the dataset, including transactions with a value of $0, and data recorded as *error* or *unknown* in transactions and item names.

## Background and Objectives

This analysis was conducted to help the client with:

* **Sales Performance:**
  + Identify the best-selling items over one year.
  + Analyze the total and average revenue of the Café in 2023.
  + Identify daily and monthly sales trends
* **Customer Behavior:**
  + Determine the most preferred payment method.
  + Reveal consumption pattern tendencies (takeaway vs. dine-in).
  + Map customer order demands.
* **Operational Insights:**
  + Analyze items frequently causing issues in sales.
  + Map customer location preference for consumption

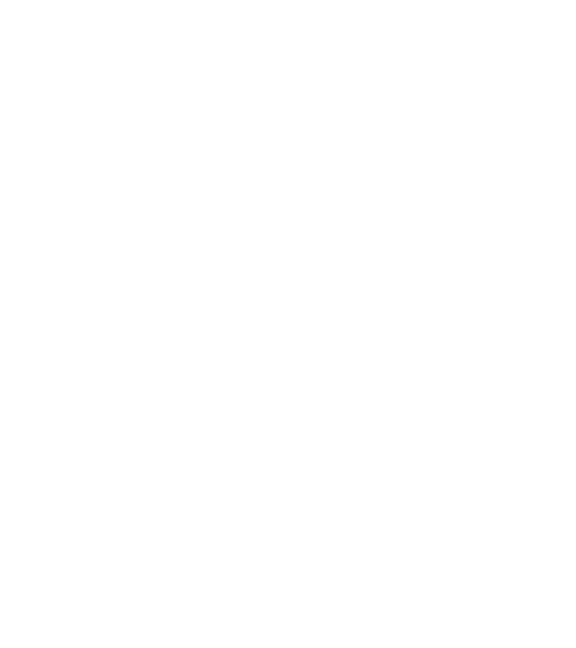
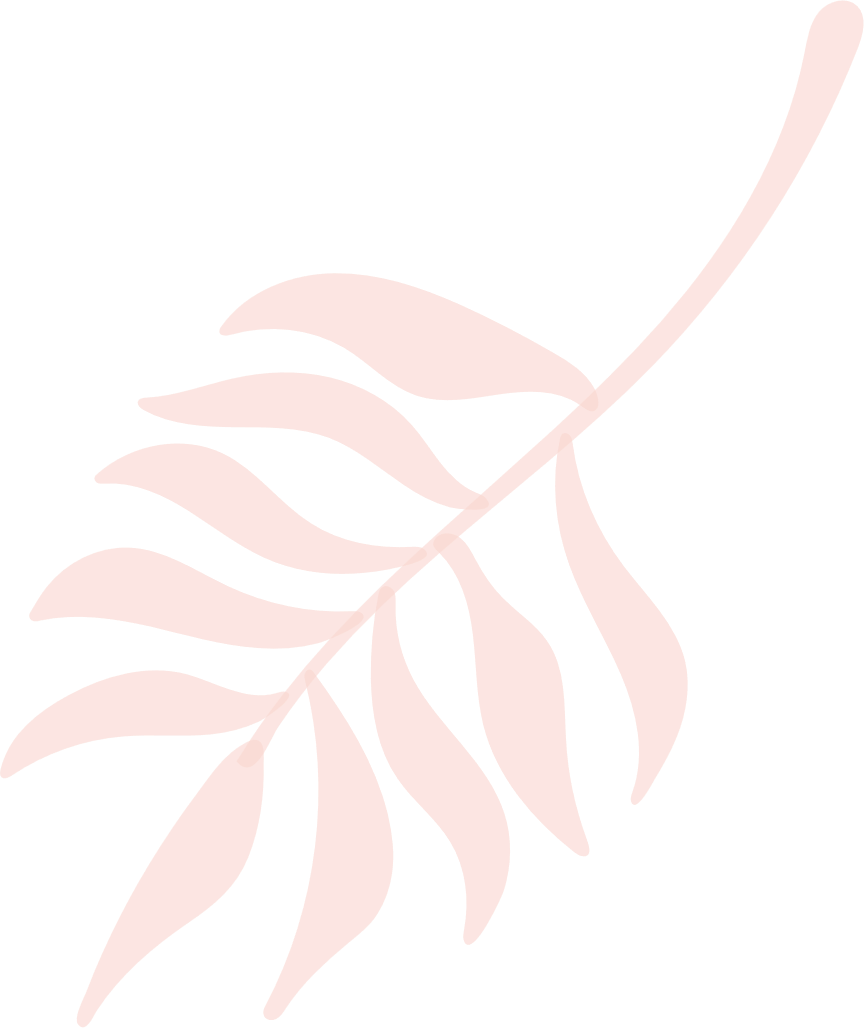
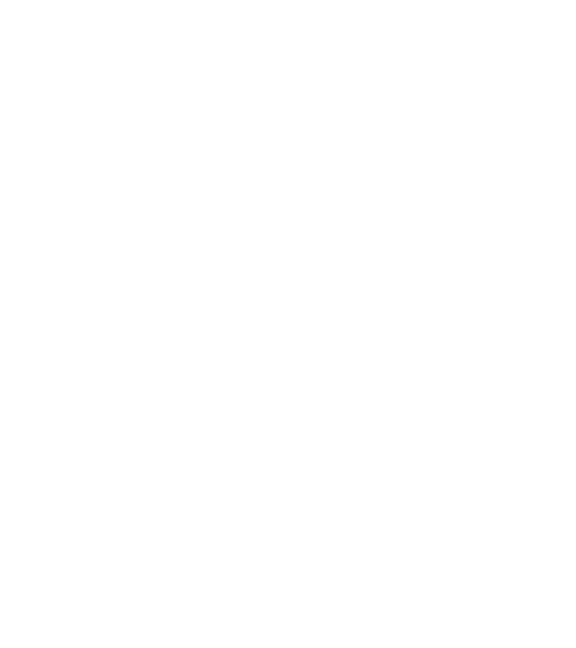
## Metodology

1. **Introduction**

This project aims to analyze the café's sales performance throughout 2023 to identify:

* + Best-selling products and revenue performance.
  + Consumer behavior.
  + Operational issues in transactions.

1. **Data Source**
   * Data Source: Kaggle Dataset ([Cafe Sales - Dirty Data for Cleaning Training](https://www.kaggle.com/datasets/ahmedmohamed2003/cafe-sales-dirty-data-for-cleaning-training))
   * Data Period: January 1 – December 31, 2023
   * Governance: Data has been anonymized and processed in accordance with GDPR regulations.
2. **Data Preparation and Processing**
   * **Integration**: Importing and validating raw data using SQL to ensure the structure, data types, and error handling met analysis needs.
   * **Cleaning**:
     + Placeholder Error Handling: Data containing "**ERROR**" and "**UNKNOWN**" in various columns was converted to **NULL**values (2,845 records).
     + Missing Values Handling:
       - **quantitiy** and **price\_unit** columns: Filled with the *median* value.
       - **total\_spend** column: Filled by multiplying quantity and price\_unit.
       - **transaction\_date** column: Filled with the *median* value.
       - Categorical columns: Filled with the *mode* (most frequent value).
     + **Outlie**r: No specific handling was applied as the total\_spend values are a natural product of quantity and price\_unit.
   * **Transformation**
     + Changed the transaction\_date data type from text (object) to datetime format.
     + Created new metrics: trans\_year, trans\_month, trans\_date
   * **Statistical Techniques**
     + Descriptive Statistics (median, mode)
     + Time series analysis for trend identification.
     + Frequency analysis for consumer preferences.
   * **Segmentation:** Customer segmentation based on purchasing behavior.
   * **Tools**
     + SQL for data querying.
     + Python (Pandas, NumPy, Matplotlib, Seaborn, Scipy) with MySQL connector.
     + Excel for data validation

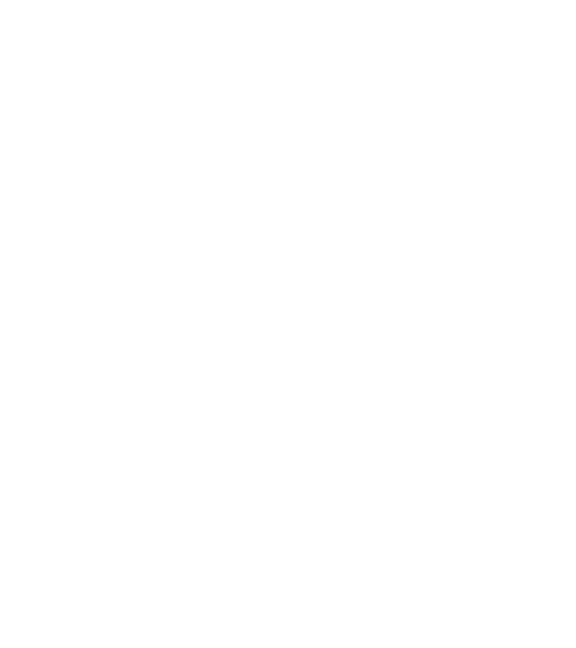
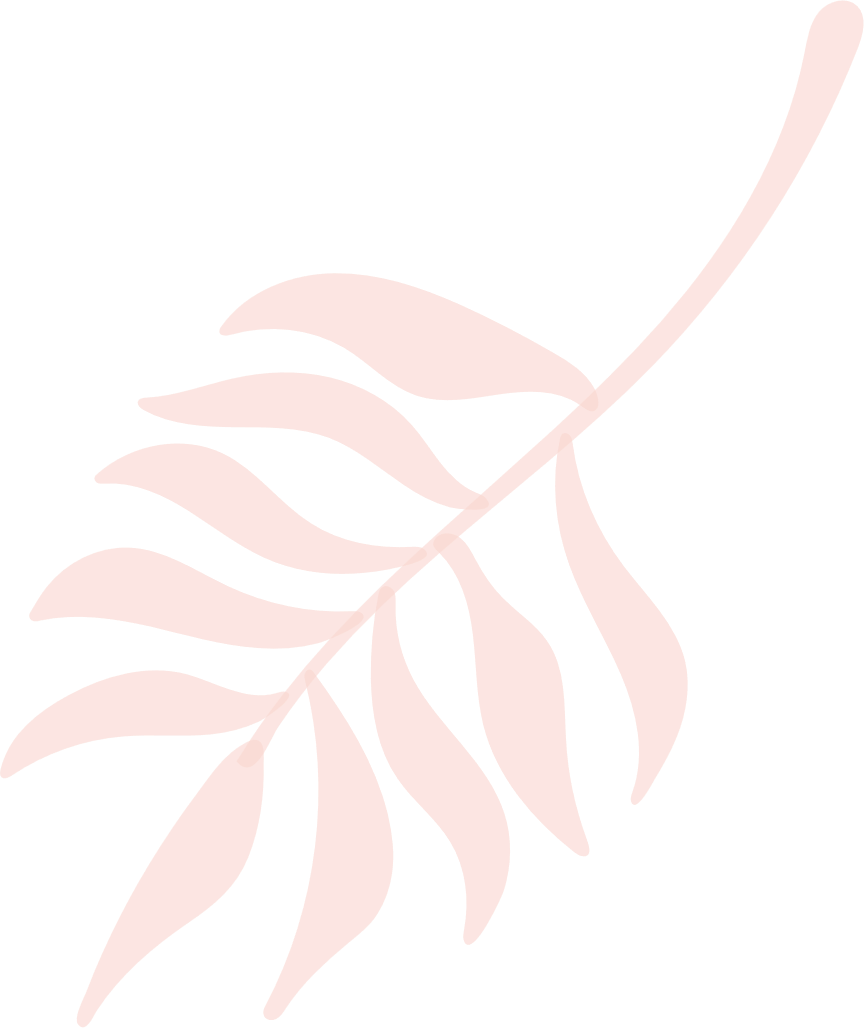
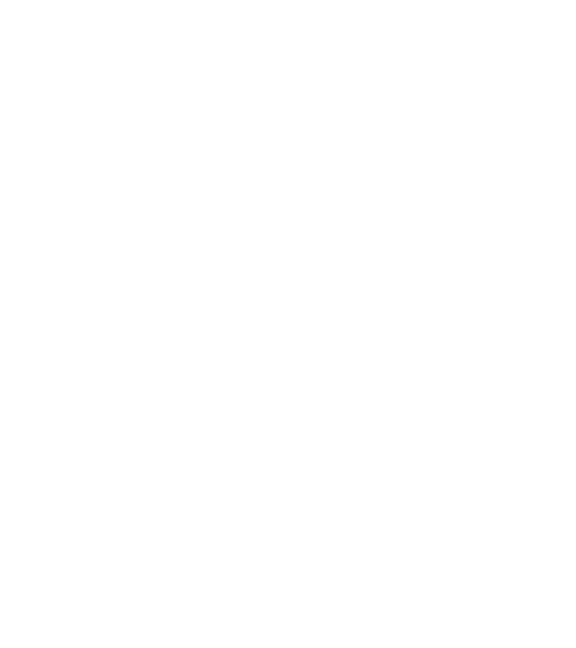


## Findings

1. **Best-Selling Product**: Juice is the most popular item (6,435 units sold).
2. **Payment Preference**: Digital wallets are most dominant (54.69%), followed by card (22.73%) and cash (22.58%).
3. **Consumption Pattern**: 70% of customers chose takeaway, 30% dine-in.
4. **Sales Trend**: Stable throughout the year with peaks mid-year.
5. **Operational Issues**: 2,845 problematic transactions (**Null**, **ERROR** and **UNKNOWN** data).

## Recommendation

1. **Flagship Product Optimization**
   1. Increase stock and promotions for Juice products.
   2. Develop new variants based on consumer preferences (e.g., juice mixes or coffee latte).
2. **Enhanced Digital Experience**
   1. Enhance and promote digital wallet features for easier transactions.
   2. Implement a digital loyalty program.
3. **Improve Operational Efficiency**
   1. Implement a data validation system at the point of entry to reduce errors.
   2. Provide staff training to ensure data input accuracy.
4. **Customer Experience Development**
   1. Optimize takeaway service with better and more practical packaging.
   2. Reorganize the space to enhance the dine-in experience
5. **Continuous Monitoring**
   1. Develop a real-time dashboard to monitor sales and operational performance.
   2. Establish a regular monthly reporting system for periodic evaluation



**Note:** Recommendations are based on 2023 data. Periodic evaluation is needed to adapt to current market dynamics.