SALES DATA ANALYSIS REPORT

1. Introduction

This report presents a comprehensive analysis of the firm's sales dataset. The objective of this project was to clean the raw sales data using Python, perform exploratory data analysis (EDA), visualize key metrics, and generate business insights to support strategic decision-making.

The analysis focuses on revenue performance, transaction volume, product performance, payment methods, sales channels, and time-based trends.

2. Dataset Description

The dataset contains transactional sales records of the firm. Key variables include:

- Transaction ID
- Transaction Date
- Item
- Quantity
- Price per Unit
- Total Spent
- Payment Method
- Location

The dataset represents individual sales transactions across multiple products, payment channels, and sales locations.

3. Data Cleaning Process

Data cleaning was conducted using Python (Pandas, NumPy).

3.1 Data Inspection

Initial inspection included:

- Checking data structure using `df.info()`
- Summary statistics using `df.describe()`
- Identifying missing values using `df.isnull().sum()`
- Detecting duplicate records

3.2 Handling Missing Values

- Missing categorical values (e.g., Payment Method, Location) were replaced with **"Unknown"**.
- Numeric fields were converted using `pd.to_numeric()` with `errors='coerce'`.
- Date columns were converted to datetime format using `pd.to_datetime()`.

A significant proportion of transactions contained **Unknown values in Payment Method (31%) and Location (40%)**, which indicates a data recording gap.

3.3 Data Type Corrections

- Transaction Date converted to `datetime64`
- Total Spent converted to numeric
- Categorical columns standardized

3.4 Duplicate Removal

Duplicate records were identified and removed to ensure data integrity.

3.5 Feature Engineering

New variables were created to enhance analysis:

- Month (extracted from Transaction Date)
- Day of Week
- Year (if applicable)

These features enabled time-based trend analysis.

4. Exploratory Data Analysis (EDA)

4.1 Key Performance Indicators (KPIs)

- **Distinct Items Sold:** 9
- **Total Transactions:** 9,540
- **Total Revenue Generated:** 76.69K

The firm demonstrates consistent transaction activity with moderate revenue performance.

4.2 Revenue by Product

Analysis shows:

- **Salad** is the highest revenue-generating product.
- **Smoothie and Sandwich** follow closely.
- **Coffee, Tea, and Cookie** generate comparatively lower revenue.

- A portion of revenue is categorized under **Unknown**, suggesting classification issues.

This indicates revenue concentration in a few high-performing items.

4.3 Revenue by Payment Method

Payment methods are distributed as follows:

- Credit Card – 23%
- Digital Wallet – 23%
- Cash – 23%
- Unknown – 31%

Customer payment preference appears balanced across available channels. However, the high percentage of Unknown entries reduces analytical reliability.

4.4 Revenue by Location

- In-Store – 30%
- Takeaway – 30%
- Unknown – 40%

The largest share of sales is attributed to Unknown location data, limiting accurate evaluation of channel performance.

4.5 Monthly Sales Trend

Spending remains relatively stable throughout the year with moderate fluctuations:

- Peak observed in **June**
- Lower performance recorded in **February and September**

This suggests consistent demand with slight seasonal variation.

4.6 Weekly Sales Trend

- **Monday** recorded the highest spending.
- Weekend performance (Saturday & Sunday) remains strong.
- Midweek (Wednesday–Thursday) shows comparatively lower activity.

This indicates stronger customer engagement at the beginning and end of the week.

5. Key Findings

1. Revenue is concentrated in a few top-performing products (Salad, Smoothie, Sandwich).

2. Transaction volume is high relative to product variety, suggesting repeat purchases.
3. Payment methods show no dominant preference.
4. A significant portion of records are classified as Unknown (Payment: 31%, Location: 40%).
5. Sales remain stable across months with a mid-year peak.
6. Weekly performance is strongest on Mondays and weekends.

6. Business Implications

- The firm has strong demand consistency.
- Product mix optimization may increase profitability.
- Data quality issues may limit accurate strategic decisions.
- Marketing efforts could target midweek sales improvement.

7. Recommendations

1. Improve data capture processes to reduce “Unknown” classifications.
2. Focus promotions on high-performing products to maximize revenue.
3. Introduce midweek discounts or campaigns to increase Wednesday–Thursday sales.
4. Conduct further analysis on customer segmentation if data becomes available.
5. Implement standardized data entry controls in the POS system.

8. Conclusion

The sales dataset reveals a stable revenue structure with identifiable top-performing products and consistent transaction volume. While overall performance is steady, data quality improvements are necessary to enhance the reliability of future analyses and support stronger business decision-making.

With improved data governance and targeted strategic actions, the firm can optimize operational efficiency and revenue growth.