The dataset taken contains transaction records of cafe sales, including item details,quantity, pricing, payment methods, and locations. The objective of this analysis is to clean the data, handle missing values and outliers, and perform statistical and visual analyses using exploratory data analysis to derive insights.

1) Data Cleaning:

Handling Missing Values:

- Numerical Columns: Filled missing values using the mean of the respective columns.
- Categorical Columns: Replaced missing values with the mode (most frequent value).

Handling duplicate records: None were found.

Handling Erroneous Values:

- Replaced "UNKNOWN" and "ERROR" values in categorical columns with their respective mode.

Detecting and Treating Outliers:

- Used the Interquartile Range (IQR) method to detect outliers in numerical columns.
- Any value outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR) was considered an outlier.

2)Exploratory Data Analysis(EDA):

Univariate Analysis:

- Computed mean, median, variance, skewness, and standard deviation for numerical columns.
- Observed distribution patterns in numerical features like Total Spent, Price Per Unit, and Quantity using histogram and box plot.
- Used bar plots to visualize the distribution of categorical features like Item,Payment Method and Location.

Bivariate Analysis:

- Correlation Matrix represented using a heatmap for numerical data.
- Scatter Plots: Analyzed relationships between numerical variables, such as Price Per Unit vs. Total Spent.
- Box Plots: Compared numerical distributions across categorical features (e.g., Total Spent across Payment Method).
- Violin Plots: Compared the density of numerical features across different categories.

Multivariate Analysis:

- Pair Plots: Explored multiple numerical relationships simultaneously, grouped by categorical features.

- Correlation Matrix & Heatmaps:Identified relationships between numerical variables.
- Grouped bar plot and grouped box plot: Comparison of multiple features to identify results.

Insights:

- Coffee and Cake are the most frequently purchased.
- Higher-priced items like Smoothies and Salads show greater spending variations.
- Customers spend similar amounts across Credit Card, Cash, and Digital Wallet.
- No specific preference for high or low-value transactions per payment method.
- The heatmap indicates that Total Spent and Quantity Ordered have the strongest positive correlation.
- Other variables show weak correlations, meaning spending habits are mostly driven by order quantity.
- No significant difference in spending patterns based on payment method.
- The box plot reveals that total spending remains consistent across payment methods.
- Some outliers indicate exceptionally high spending transactions.