**EDA Case Study: “Analyzing Cafe Sales Data”**

This is class activity. In which we perform following options

Finding missing values

Deals with missing values

Finding outlier and deal with them

Create different visualization like line, pie chart, heat map

One hot encoding technique.

Before dealing with missing value we convert data type of Quantity Price per unity and total spent for calculation. Because on object no mathematical operation can perform.

Transaction id: Keep as it is because no missing values.

Item: We replace it with Unknown because data is categorical

Quantity: Find median and replace missing with it.

Price per unit: Replace with median

Total spent: after replacing quantity and price per unit. Multiplay these 2 to find total spent

Payment method: Replace with unknown

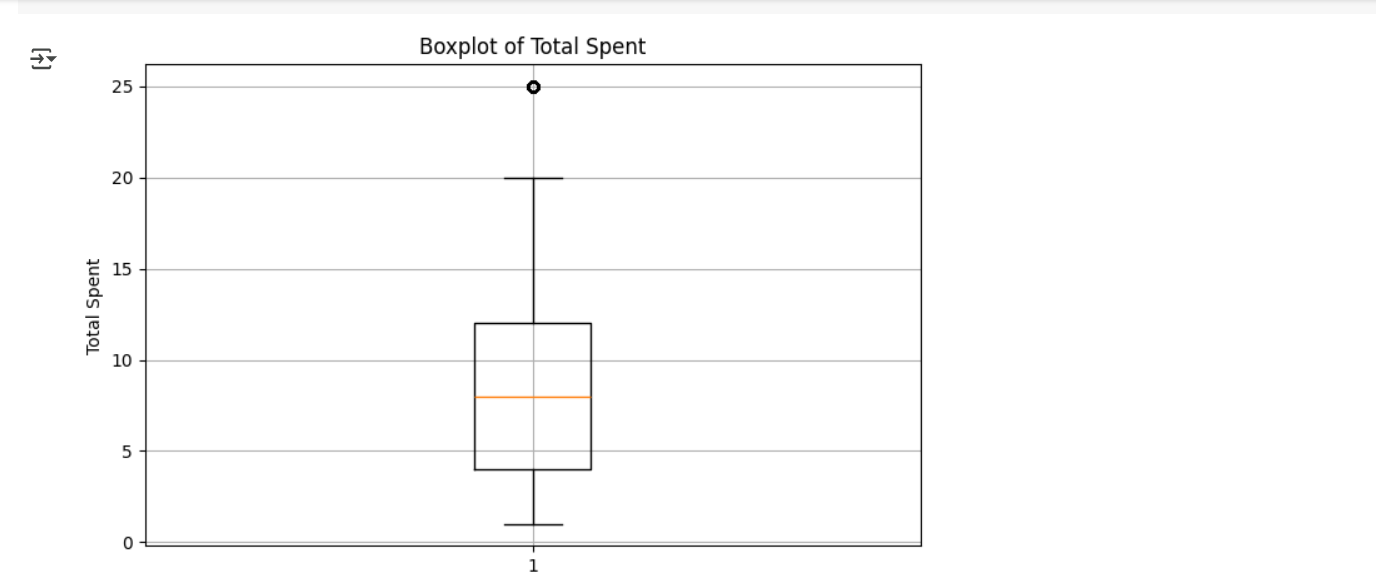
Location: Drop column because more than 30% data missing

Transaction date: Drop row because less than 2% data missing save to remove.

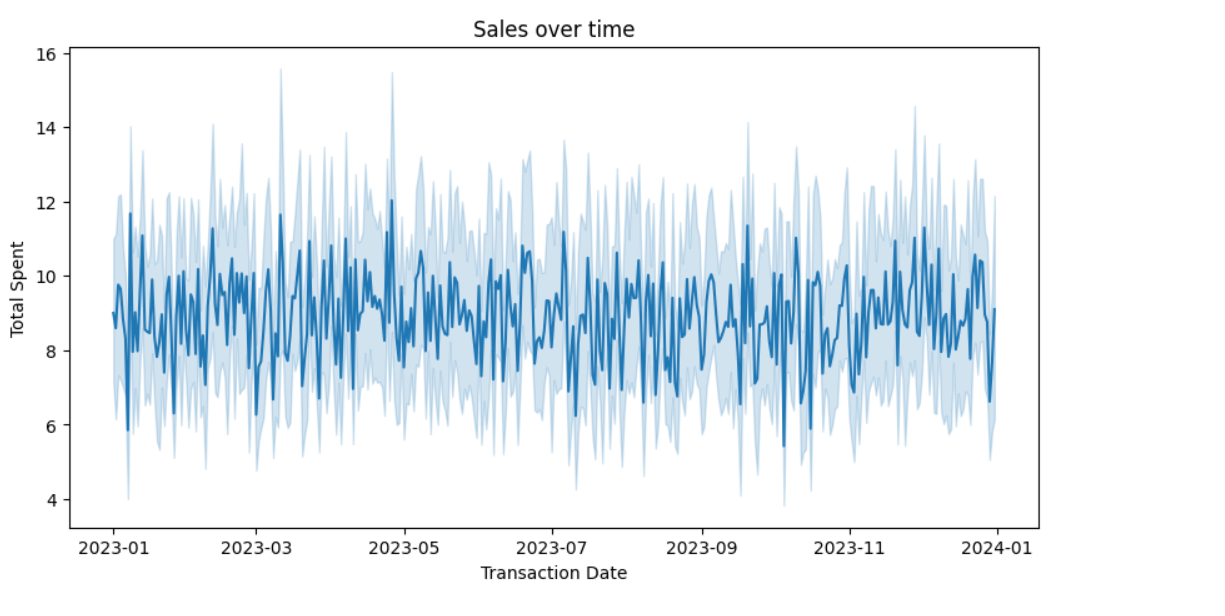
**Outlier Report:**

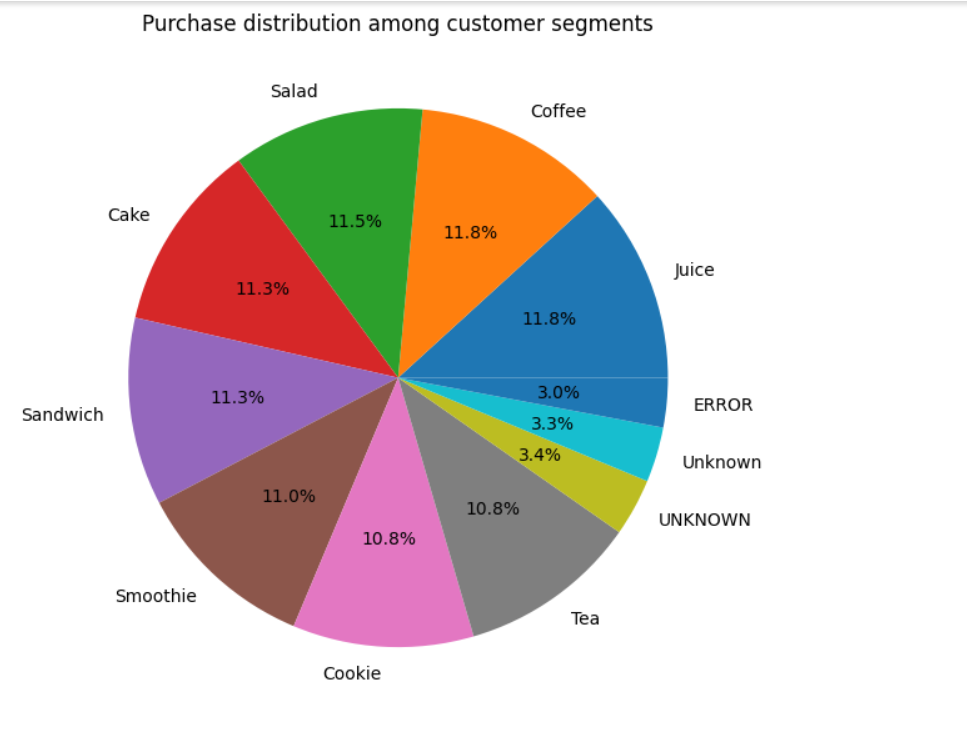
Find outlier in Total spent and cap it by using lower bound and upper bound.

**Visualizations:**



Box plot shows the outlier.





**Feature Summary:**New column of **Total Sale** add.  
we add **Hour** column to check the peak hour.