***1.Descriptive Analytics for Numerical Columns***

Column: Volume

Mean: 5.066666666666666. This is the average value of the Volume column

Median: 4.0. This is the middle value of the Volume when sorted. It is less sensitive to outliers than the mean.

Mode: 3. This is the most frequent value in the Volume column.

Standard Deviation: 4.231602391213926. This measures the amount of variation or dispersion of the Volume values.

--------------------

Column: Avg Price

mean: 10453.433333333332

median: 1450.0

mode: 400

std: 18079.90483993645

--------------------

Column: Total Sales Value

mean: 33812.83555555555

median: 5700.0

mode: 24300

std: 50535.07417255328

--------------------

Column: Discount Rate (%)

mean: 15.155241895330914

median: 16.57776564929597

mode: 5.007822189204133

std: 4.220602153325511

--------------------

Column: Discount Amount

mean: 3346.499424342415

median: 988.9337332382784

mode: 69.17794228822787

std: 4509.9029633539985

--------------------

Column: Net Sales Value

mean: 30466.336131270138

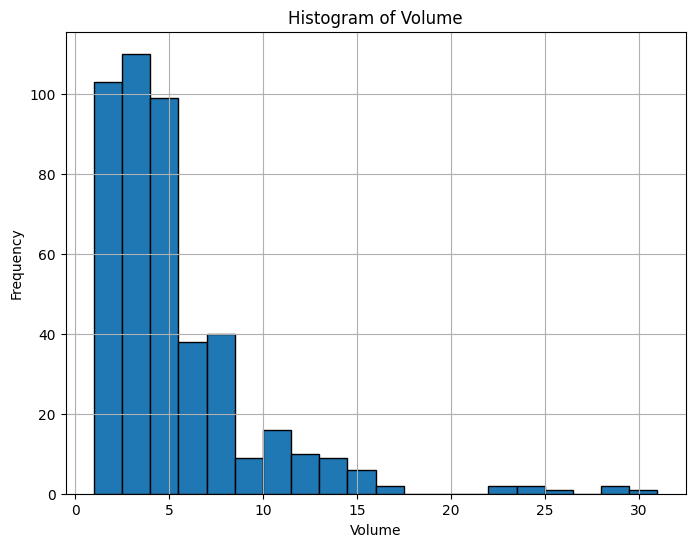
median: 4677.788059301853

mode: 326.97480055470817

std: 46358.656624066934

***2.Data Visualization***

**Histograms**

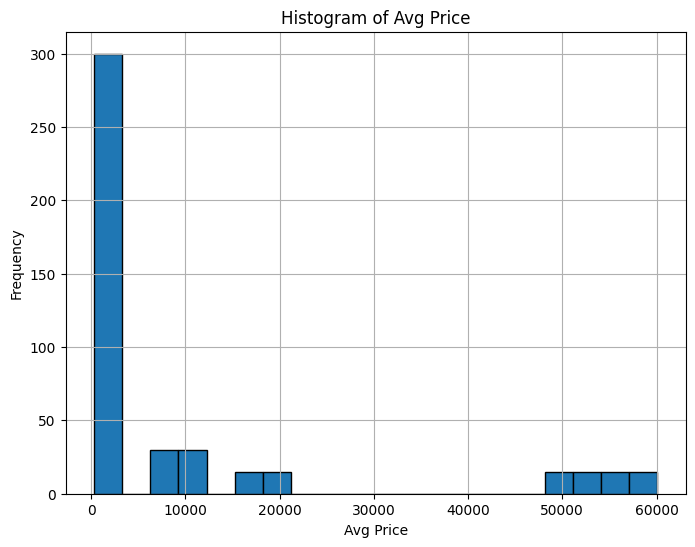


Skewness of Volume: 2.731723626557714

The distribution of Volume is right-skewed (positively skewed).

Number of outliers in Volume: 44

Potential outliers detected. Further investigation might be needed.

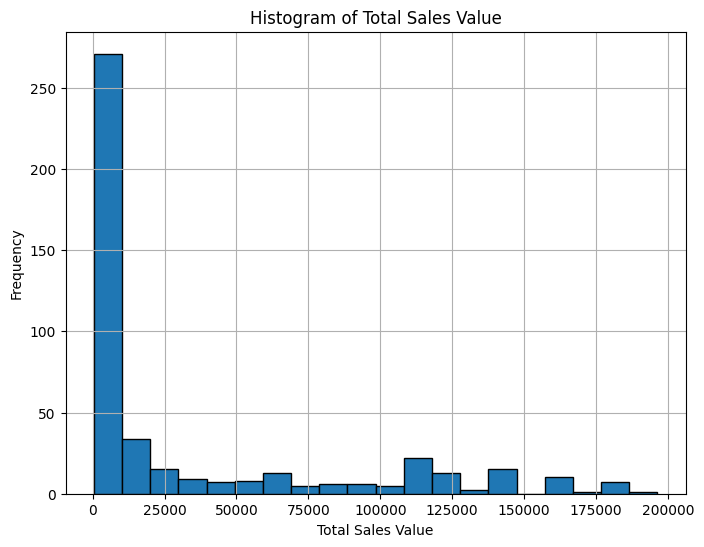


Skewness of Avg Price: 1.9088730767759727

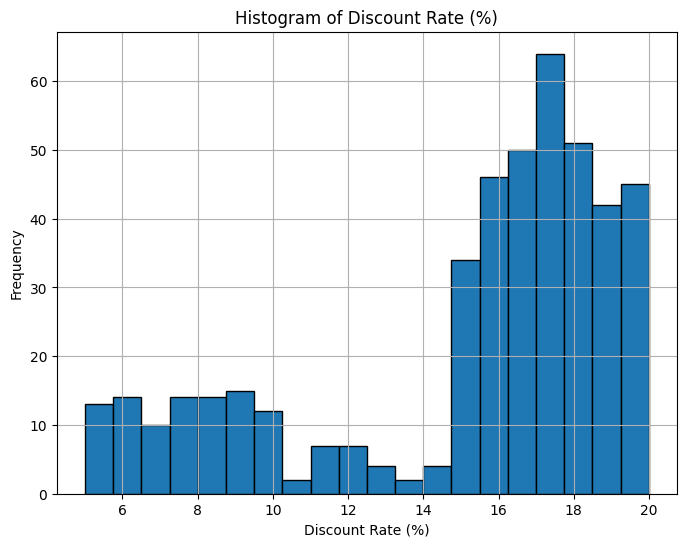
The distribution of Avg Price is right-skewed (positively skewed).

Number of outliers in Avg Price: 60

Potential outliers detected. Further investigation might be needed.

 Skewness of Total Sales Value: 1.5347293238977937

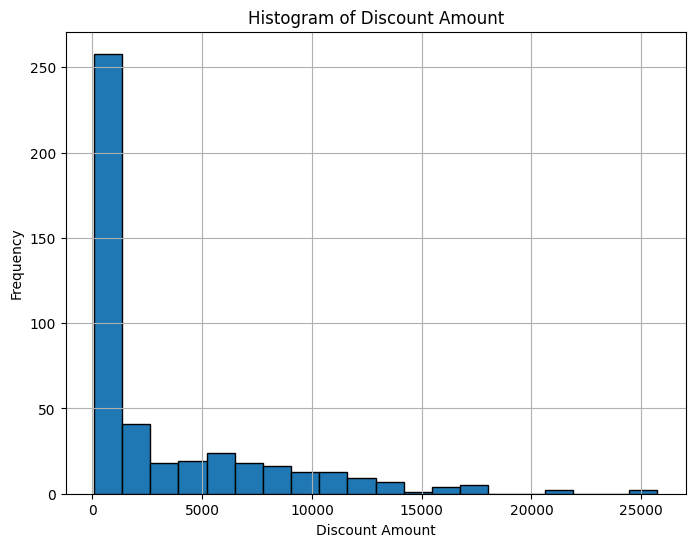
The distribution of Total Sales Value is right-skewed (positively skewed)Number of outliers in Total Sales Value: 36



Skewness of Discount Rate (%): -1.0622943696988825

The distribution of Discount Rate (%) is left-skewed (negatively skewed).

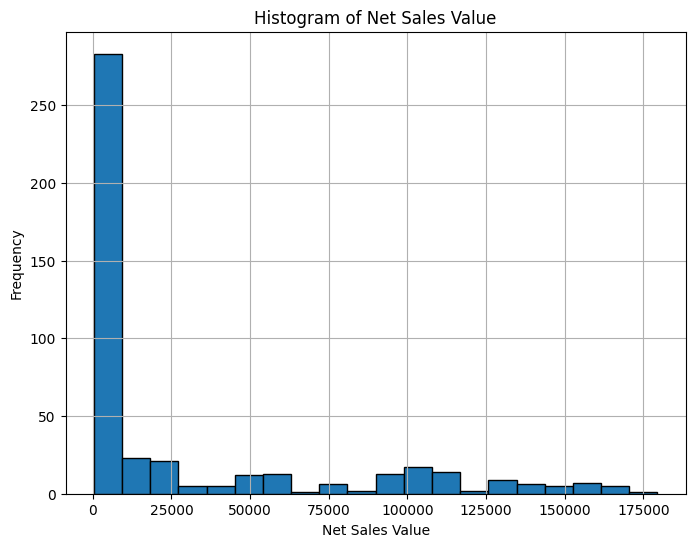
Number of outliers in Discount Rate (%): 45



Skewness of Discount Amount: 1.9130377339538271

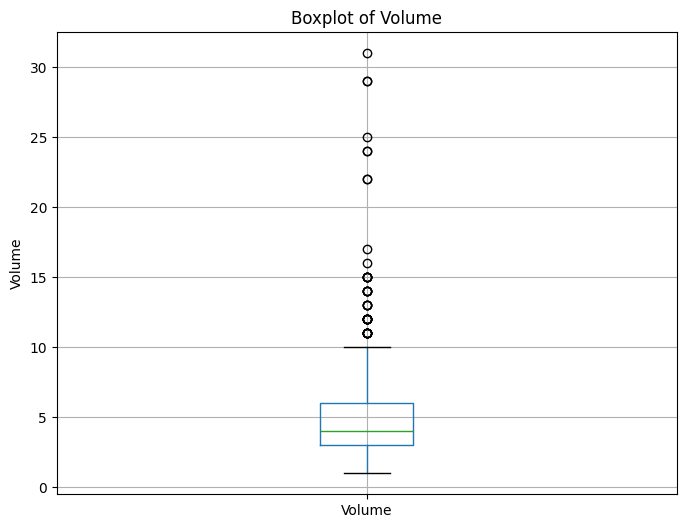
The distribution of Discount Amount is right-skewed (positively skewed).

Number of outliers in Discount Amount: 24



Skewness of Net Sales Value: 1.5408217575344376.The distribution of Net Sales Value is right-skewed (positively skewed). Number of outliers in Net Sales Value: 35

**Boxplots**

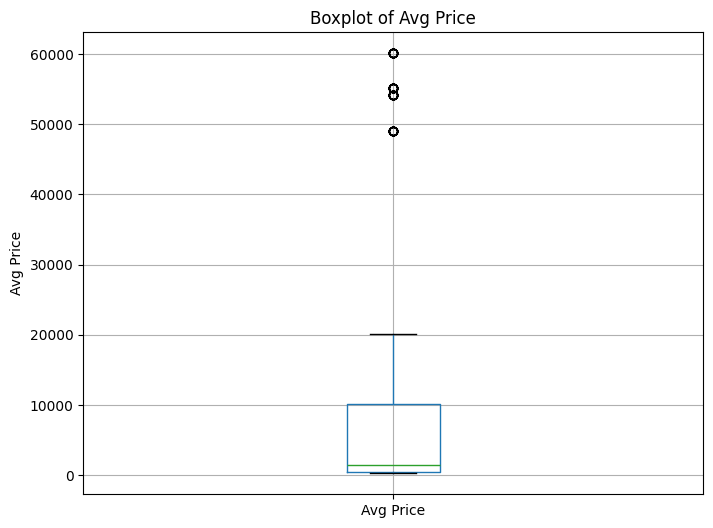


Column: Volume

Interquartile Range (IQR): 3.0

Lower Bound for Outliers: -1.5

Upper Bound for Outliers: 10.5

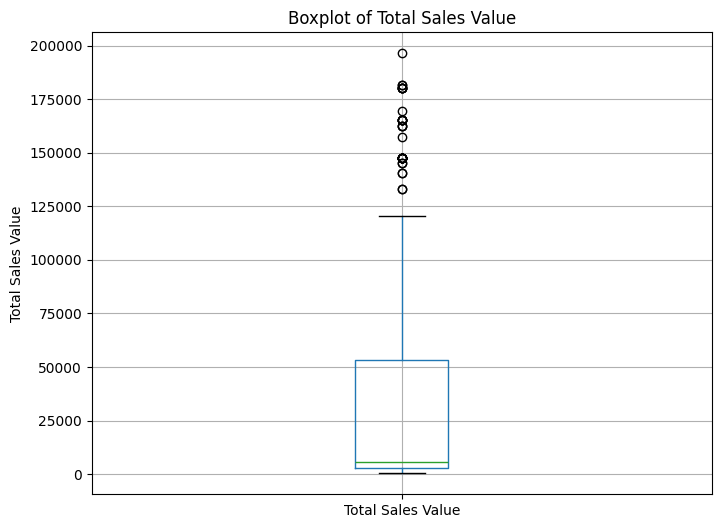


Column: Avg Price

Interquartile Range (IQR): 9635.0

Lower Bound for Outliers: -13987.5

Upper Bound for Outliers: 24552.5

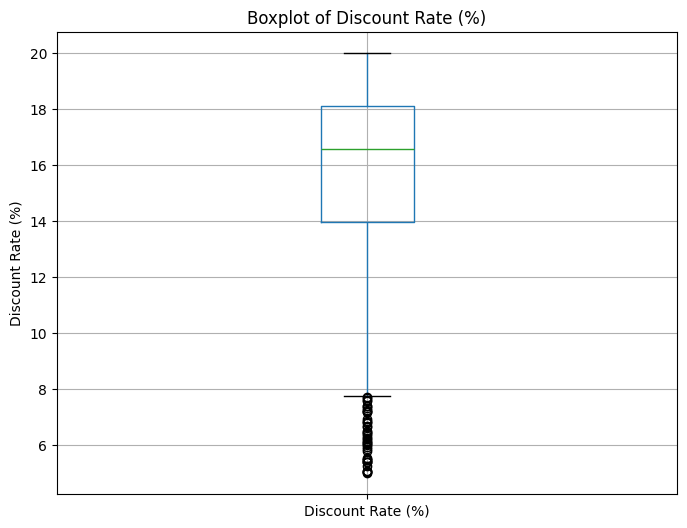


Column: Total Sales Value

Interquartile Range (IQR): 50500.0

Lower Bound for Outliers: -73050.0

Upper Bound for Outliers: 128950.

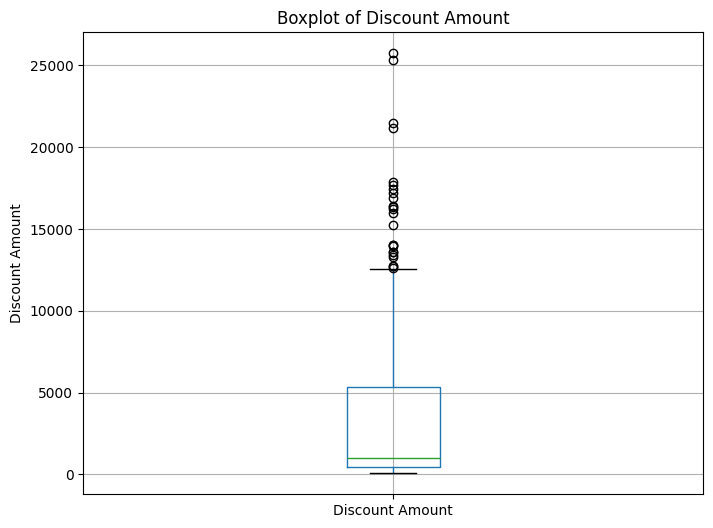


Column: Discount Rate (%)

Interquartile Range (IQR): 4.149655934050962

Lower Bound for Outliers: 7.740578642625298

Upper Bound for Outliers: 24.339202378829146

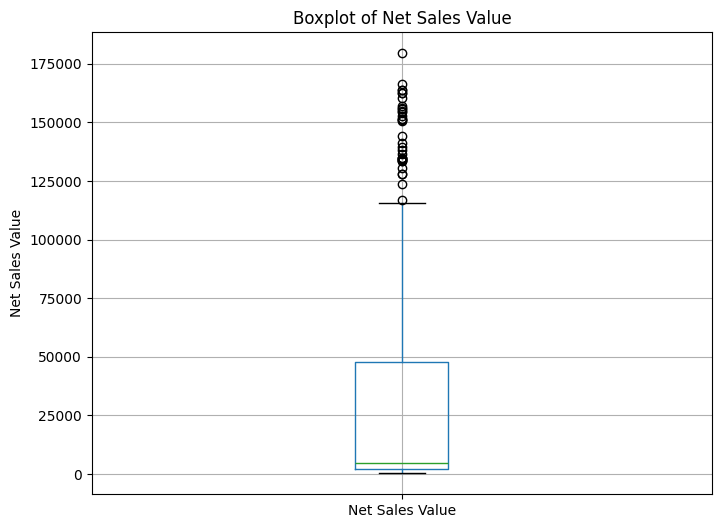


Column: Discount Amount

Interquartile Range (IQR): 4856.03612280112

Lower Bound for Outliers: -6823.594880316146

Upper Bound for Outliers: 12600.54961088833



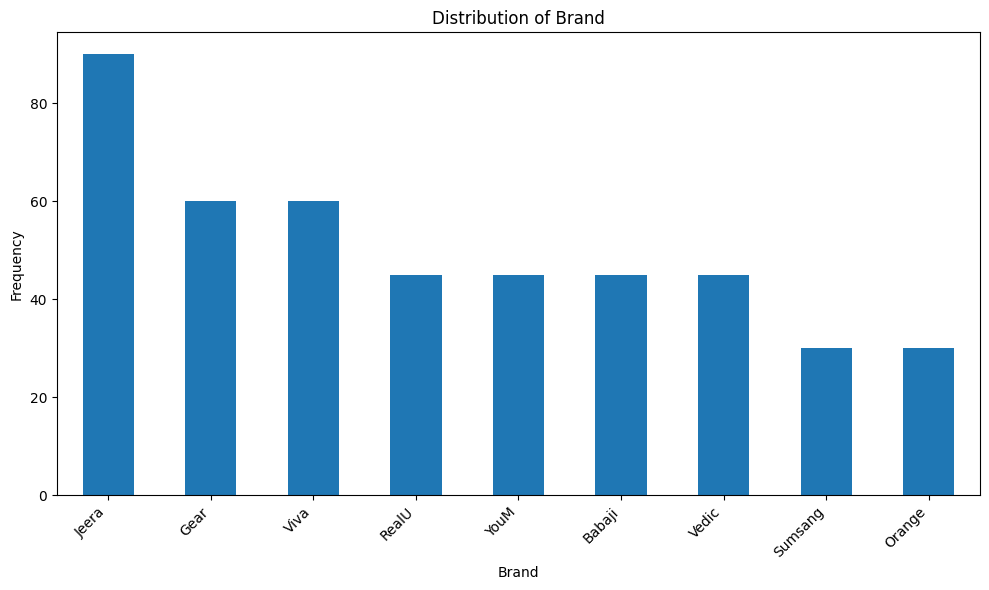
Column: Net Sales Value

Interquartile Range (IQR): 45645.70420627015

Lower Bound for Outliers: -66266.347664084

Upper Bound for Outliers: 116316.46916099661

**Bar Chart Analysis for Categorical Column**



Distribution of Brand:

Brand

Jeera 90

Gear 60

Viva 60

RealU 45

YouM 45

Babaji 45

Vedic 45

Sumsang 30

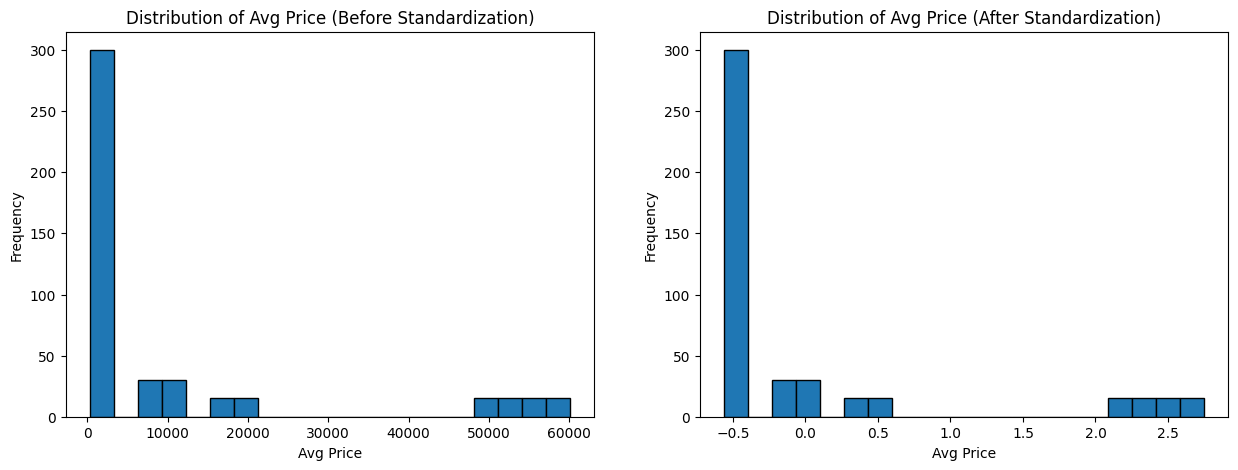
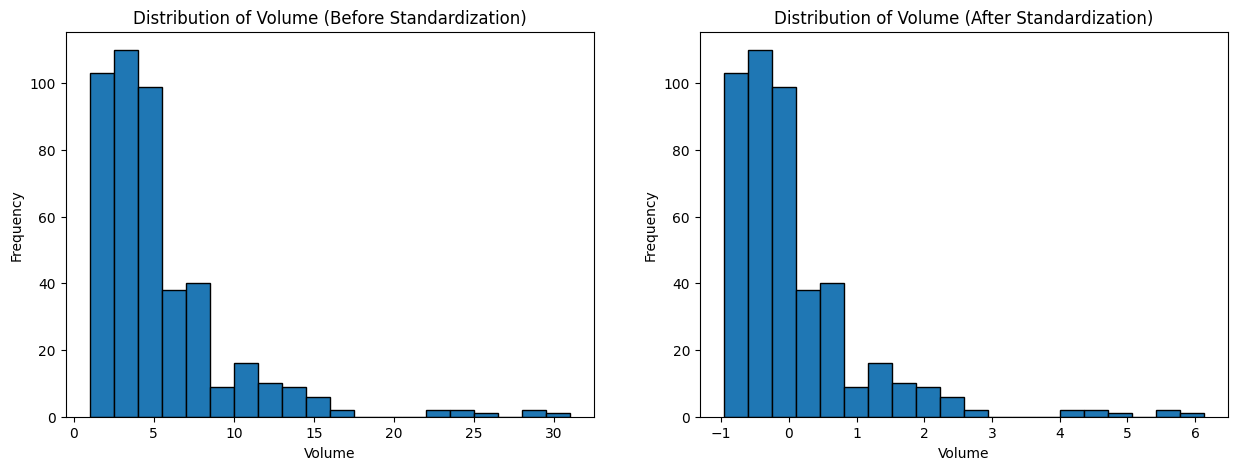
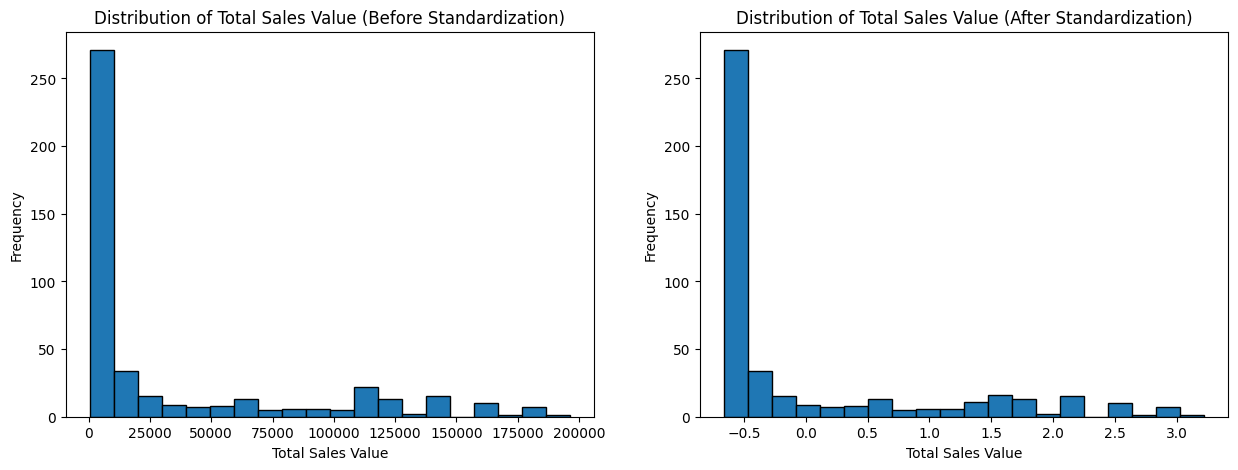
Orange 30

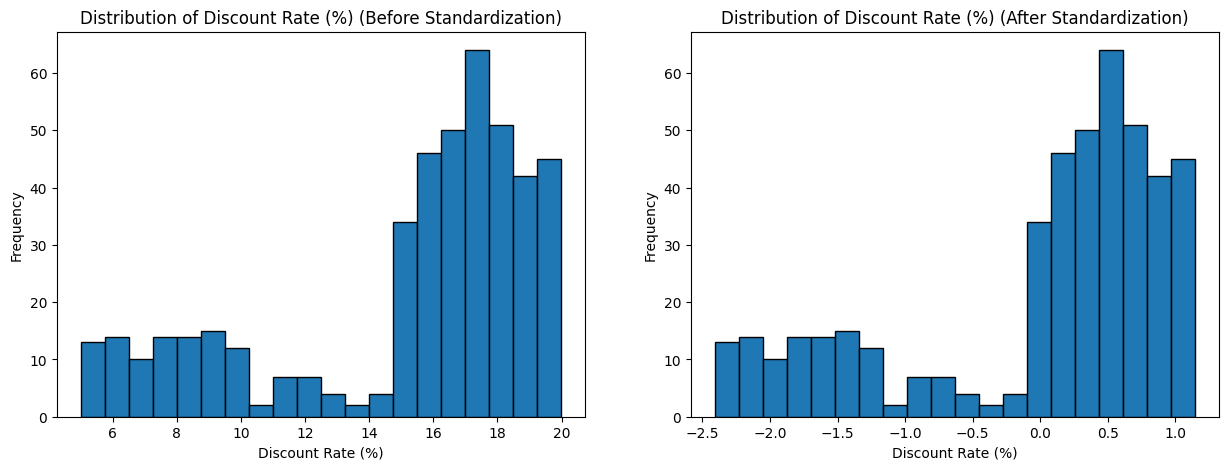
Name: count, dtype: int64

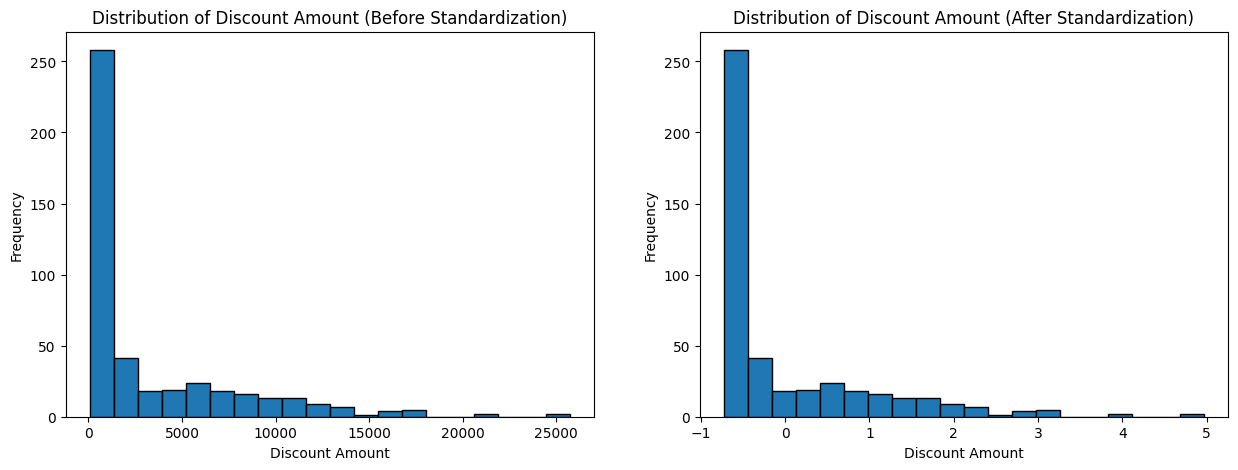
The most frequent category in 'Brand' is 'Jeera' with 90 occurrences.

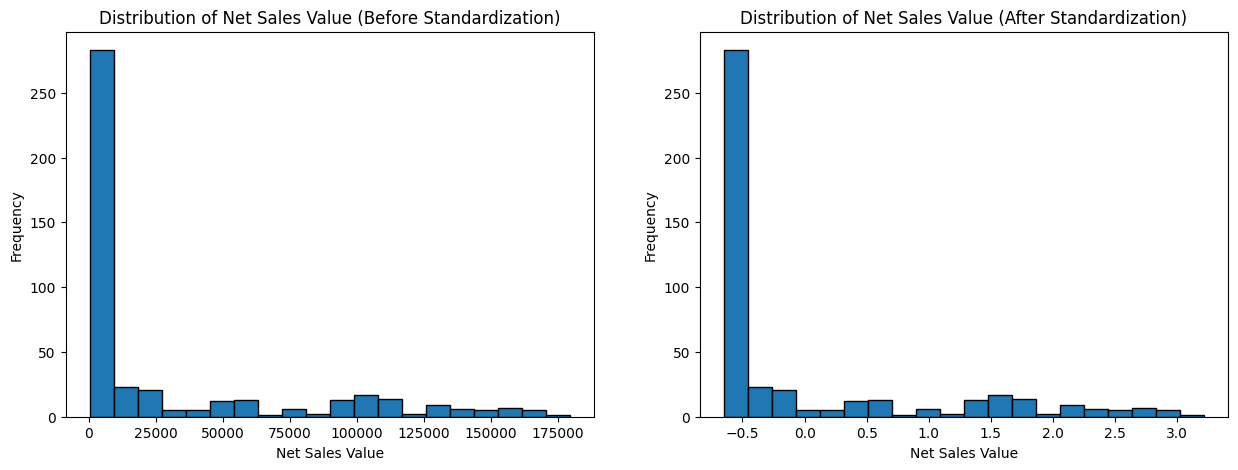
#### **3.Standardization of Numerical Variables**

Standardization, also known as **z-score normalization**, is a statistical technique used to scale and center a dataset so that it has a mean of 0 and a standard deviation of 1. It is particularly useful when data features have different scales and units, as it makes the data more comparable and often improves the performance of machine learning models.







#### **5.Conversion of Categorical Data into Dummy Variables**

In machine learning, many algorithms require numerical input to function properly. Categorical data, which consists of labels or categories (e.g., "Red," "Blue," "Green" for colors), needs to be transformed into a numerical format for such algorithms. **One-hot encoding** is a commonly used method for this purpose. It converts categorical variables into a format that can be provided to machine learning algorithms while preserving the categorical nature of the data.

Volume Avg Price Total Sales Value Discount Rate (%) Discount Amount \

0 15 12100 181500 11.654820 21153.498820

1 10 10100 101000 11.560498 11676.102961

2 7 16100 112700 9.456886 10657.910157

3 6 20100 120600 6.935385 8364.074702

4 3 8100 24300 17.995663 4372.946230

Net Sales Value Date\_02-04-2021 Date\_03-04-2021 Date\_04-04-2021 \

0 160346.501180 False False False

1 89323.897039 False False False

2 102042.089843 False False False

3 112235.925298 False False False

4 19927.053770 False False False

Date\_05-04-2021 ... Model\_Vedic Cream Model\_Vedic Oil \

0 False ... False False

1 False ... False False

2 False ... False False

3 False ... False False

4 False ... False False

Model\_Vedic Shampoo Model\_W-Casuals Model\_W-Inners Model\_W-Lounge \

0 False False False False

1 False False False False

2 False False False False

3 False False False False

4 False False False False

Model\_W-Western Model\_YM-98 Model\_YM-99 Model\_YM-99 Plus

0 False False False False

1 False False False False

2 False False True False

3 False False False True

4 False True False False