

Dataset Cleaning Report

The goal of this task was to clean and prepare the dataset for analysis by handling missing values and correcting data .

Steps Followed :

- I have imported the necessary Python libraries required for data handling and cleaning, such as Pandas , NumPy and Matplot.
- Imported dataset using pandas.

Then, performed basic operations to understand the dataset:

- Used head() and tail() to view sample records Also, Reviewed overall dataset structure
- Checked dataset shape and data types.

Checking Missing and Duplicate Values

- Checked for missing values and for duplicate rows in the dataset
- There were missing values in two columns: Item Weight and Outlet Size , Also there are no duplicate Values.

Handling Missing Values

- For the Outlet column, missing values were replaced with "Unknown"
- For Item Weight, I noticed that the same items had the same weight but some records had null values . For this , I took help from stack overflow there I got solution to use a transform function to fill missing item weights using existing values of the same item.

- Some values were still missing, so I filled them using the mean value of the column.
- Since mean values had long decimal points, I rounded them to 3 decimal places.
- In the Fat Content column, values such as "LF" and "reg" were replaced with "Low Fat" and "Regular" as Low Fat and Regular values were already there.
- The Item MRP column values were rounded to 2 decimal places to make the calculation easier.

Now as the dataset is cleaned , I created visualizations to understand the patterns using bar graphs , line graph and pie charts.