**1. Handling Missing Values (Item Weight)**

**Issue:** Item Weight column has 1463 missing values.  
**Fix:** Fill missing values with the **mean** of the column.

**🔧 Excel Formula:**

excel

CopyEdit

=IF(ISBLANK([@Item\_Weight]), AVERAGE([Item\_Weight]), [@Item\_Weight])

**How to apply:**

* Insert a new column: Item Weight (Cleaned)
* Use the above formula in the new column.
* Then **Copy → Paste as Values** to overwrite the original Item Weight.

**2. Remove Duplicate Rows**

**Action:**

* Select all data.
* Go to Data → Click on **Remove Duplicates**.
* Select **all columns** → Click OK.

This will drop any repeated rows based on all columns.

**3. Standardize Text Values (Item Fat Content)**

**Issue:** Inconsistent entries like LF, low fat, reg.

**Excel Formula:**

=IF(OR([@Item\_Fat\_Content]="LF",[@Item\_Fat\_Content]="low fat"), "Low Fat",

IF([@Item\_Fat\_Content]="reg", "Regular",[@Item\_Fat\_Content]))

**Steps:**

* Add a new column: Item Fat Content (Cleaned)
* Use the formula above to normalize the values.
* Copy and paste it as values to replace the old column.

**4. Convert Establishment Year to Proper Date Format**

**Action:** Convert the Outlet Establishment Year from numeric to a date format like 01-01-YYYY.

**Excel Formula:**

excel

=DATE([@Outlet\_Establishment\_Year],1,1)

**Format:**

* Use Format Cells → Date → dd-mm-yyyy to display as 01-01-2012.

**5. Rename Column Headers**

**Fixes Applied:**

* Convert all headers to lowercase.
* Replace spaces with underscores (\_).
* Remove leading/trailing whitespaces.

**Manual Steps:**

* Double-click each header and rename:
  + Item Fat Content → item\_fat\_content
  + Item Type → item\_type
  + Outlet Size → outlet\_size
  + etc.

**6. 🔢 Fix Data Types**

**Rating:**

* Should be Integer.

**🔧 Excel Formula (if some ratings are in decimal):**

excel

=INT([@Rating])

**Sales:**

* Ensure column is formatted as Number with 2 decimal places.

**Steps:**

* Use Data → Text to Columns if the column is misformatted.
* Use Format Cells → Number.

**✅ Final Columns After Cleaning:**

|  |  |
| --- | --- |
| **Clean Column Name** | **Notes** |
| item\_fat\_content | Cleaned, standardized casing |
| item\_identifier | No changes needed |
| item\_type | No changes needed |
| outlet\_establishment\_year | Converted to date |
| outlet\_identifier | No changes needed |
| outlet\_location\_type | No changes needed |
| outlet\_size | No changes needed |
| outlet\_type | No changes needed |
| item\_visibility | No changes needed |
| item\_weight | Missing values filled with mean |
| sales | Verified as float |
| rating | Converted to integer |