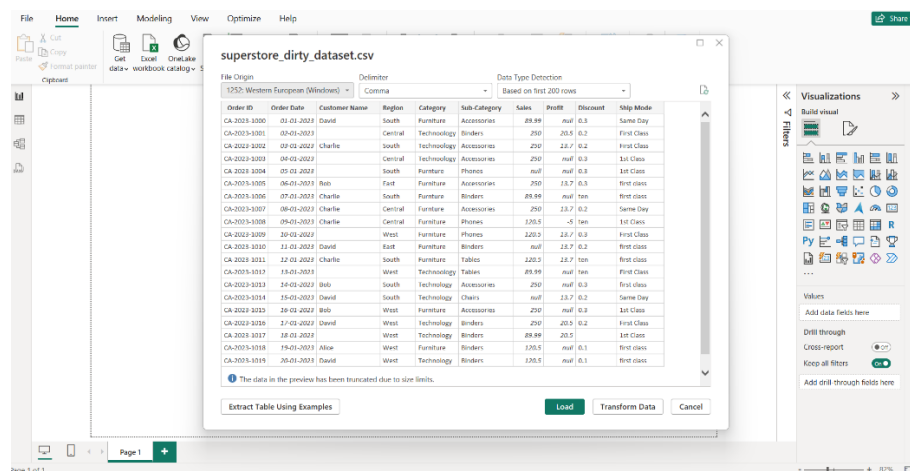


# Perform the Extraction Transformation and Loading (ETL) process to construct the database in the SQL server / Power BI.

## STEP 1: EXTRACT – Load the Dataset into Power BI

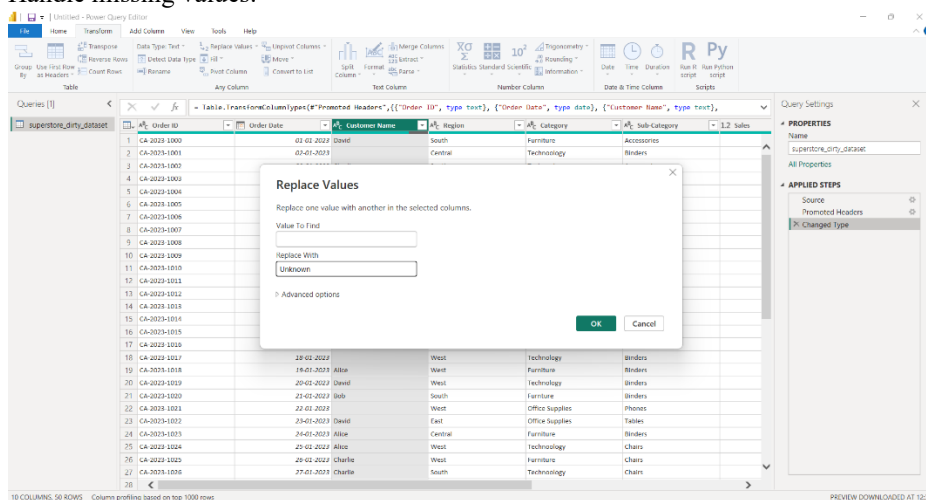
1. Open Power BI Desktop.
2. Click on Home → Get Data → Text/CSV.
3. Select the file: superstore\_dirty\_dataset.csv.
4. In the preview window, click Load.



## STEP 2: TRANSFORM – Clean the Data in Power Query Editor

Click Transform Data (from Home tab) to enter the Power Query Editor.

### A. Handle missing values:



### B. Handle Null values:

- Replace null with **mean** of the column.
- Go to transform → Under Statistics tab select **average**. The mean of that column will be calculated.
- Select column → Transform tab → Replace Values → Replace null with the calculated mean.

Untitled - Power Query Editor

Queries [1] **superstore\_city\_dataset**

Table.ReplaceValue("Changed Type","","Unknown",Replacer.ReplaceValue)

Line	Category	Sub-Category	Sales	Profit	Discount	Ship Mode
1	Furniture	Accessories	89.89	null	0.3	Same Day
2	Technology	Binders	250	203.7	0.2	First Class
3	Technology	Accessories	250	23.7	0.2	First Class
4	Technology	Accessories	250	null	0.3	1st Class
5	Furniture	Phones	null	null	0.3	1st Class
6	Furniture	Accessories	250	23.7	0.3	First class
7	Furniture	Binders	89.89	null	1st	First class
8	Furniture	Accessories	250	13.7	0.2	Same Day
9	Furniture	Phones	130.5	4	1st	1st Class
10	Furniture	Phones	120.5	23.7	0.3	First Class
11	Furniture	Binders	null	23.7	0.2	First class
12	Furniture	Tables	191.5	13.7	1st	1st Class
13	Technology	Tables	89.89	null	1st	First Class
14	Technology	Accessories	250	null	0.3	First class
15	Technology	Chairs	null	23.7	0.2	Same Day
16	Furniture	Accessories	250	null	0.3	1st Class
17	Technology	Binders	250	20.5	0.2	First Class
18	Technology	Binders	89.89	25.5	1st	1st Class
19	Furniture	Binders	203.5	null	0.2	First class
20	Technology	Binders	120.5	null	0.1	First class
21	Furniture	Binders	250	23.7	0.3	First Class
22	Office Supplies	Phones	null	0	0.3	Same Day
23	Office Supplies	Tables	120.5	null	1st	Standard Class
24	Furniture	Binders	null	25.5	0.2	First Class
25	Technology	Chairs	120.5	13.7	1st	Standard Class
26	Furniture	Chairs	null	null	0.2	Standard Class
27	Technology	Chairs	89.95	25.5	1st	Same Day

10 COLUMNS, 50 ROWS. Columns profiling based on top 1000 rows. PREVIEW DOWNLOADED AT 12:45

Untitled - Power Query Editor

Queries [1] **superstore\_city\_dataset**

Table.Average([Replaced Value][Sales])

156.10465464464598

READY

Query Settings

Properties

Name

superstore\_city\_dataset

Applied Steps

Source

Promoted Headers

Changed Type

Replaced Value

Calculated Average

Untitled - Power Query Editor

Queries [1] **superstore\_city\_dataset**

Table.ReplaceValue("Changed Type","","Unknown",Replacer.ReplaceValue,("Customer Name"))

Replace Values

Replace one value with another in the selected columns.

Value To Find

Null

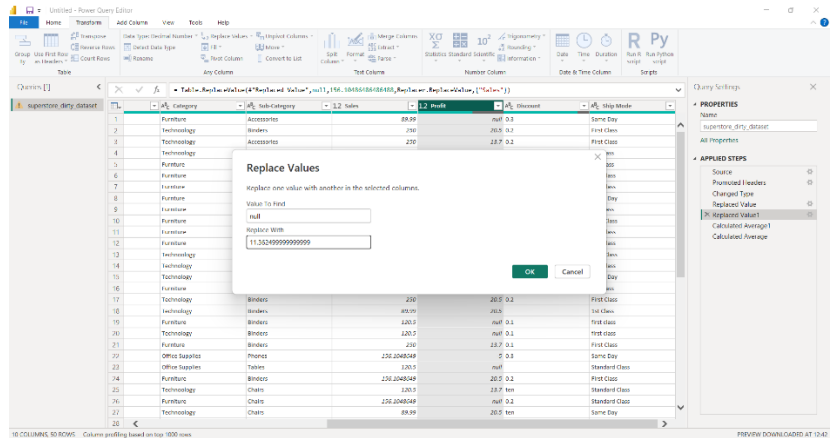
Replace With

156.10465464464598

OK Cancel

Line	Region	Category	Sub-Category	Sales	Profit	Discount
1	South	Furniture	Accessories	89.89	null	0.3
2	Central	Technology	Binders	250	203.7	0.2
3	South	Technology	Accessories	250	23.7	0.2
4	Central				null	0.3
5	South				null	0.3
6	East				null	1st
7	South				null	1st
8	Central				23.7	0.2
9	West				0	1st
10	West				23.7	0.3
11	East				23.7	0.2
12	South				23.7	1st
13	West				null	1st
14	South				null	0.3
15	West				23.7	0.2
16	West				null	0.3
17	West	Technology	Binders	250	203.7	0.2
18	West	Furniture	Binders	89.89	25.5	1st
19	West	Technology	Binders	120.5	null	0.3
20	West	Technology	Binders	120.5	null	0.1
21	South	Furniture	Binders	250	23.7	0.3
22	West	Office Supplies	Phones	null	0	0.3
23	East	Office Supplies	Tables	120.5	null	1st
24	Central	Furniture	Binders	null	25.5	0.2
25	West	Technology	Chairs	120.5	13.7	1st
26	West	Furniture	Chairs	null	null	0.2
27	South	Technology	Chairs	89.95	25.5	1st

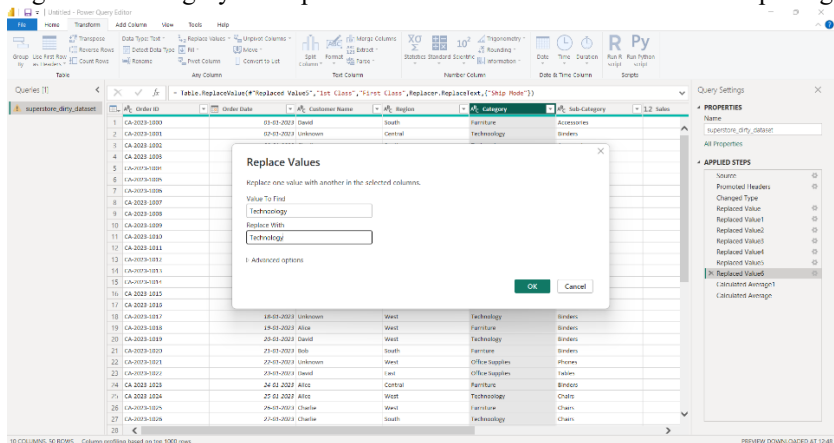
10 COLUMNS, 50 ROWS. Columns profiling based on top 1000 rows. PREVIEW DOWNLOADED AT 12:45



## C. Fix Text Errors and Inconsistencies:

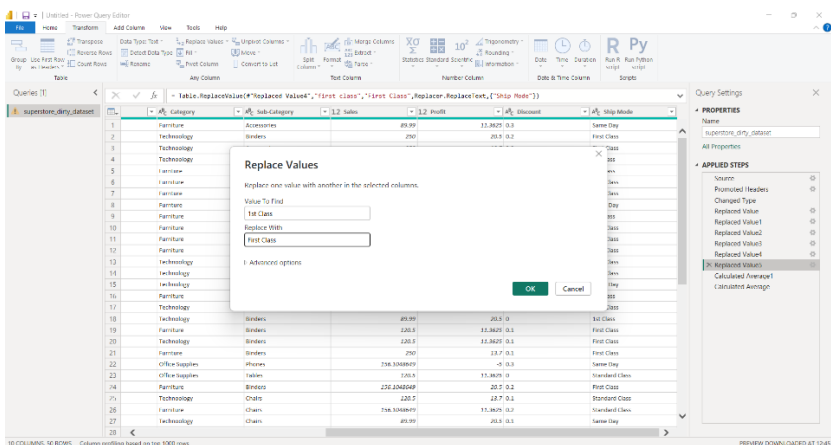
### 1. Category Column (fix typos like "Technology", "Furniture"):

- Replace "Technology" → "Technology"
- Replace "Furniture" → "Furniture"
- Method:
- Right-click Category → Replace Values → Enter incorrect & correct spelling



### 2. Ship Mode Column:

- Standardize formats:
- Replace "1st Class" and "first class" → "First Class"
- Right-click → Replace Values (repeat for each inconsistency)



## D. Fix Data Types

### 1. Discount Column:

- Some values like "ten" or null are not numeric.
- Step 1: Replace "ten" with 0.1 or your desired value.
- Step 2: Change column type to Decimal Number
- Click on the column → Data Type → Decimal Number

