Here's a comprehensive guide on **Data Cleaning in Excel for Big Datasets** with **real-world business questions**, **step-by-step instructions**, and **cleaning tasks** using sample datasets:

## 📌 Goal:

Learn to clean and prepare large datasets in Excel for analysis using real-world business scenarios.

## ✅ Topics Covered:

Common data cleaning issues

Step-by-step Excel cleaning tasks

Real-world business questions

Sample dataset links

Assignments

## 📊 Dataset Used (Examples):

[Superstore Sales Data](https://www.kaggle.com/datasets/vivek468/superstore-dataset-final)

[Customer Data (Dirty)](https://people.sc.fsu.edu/~jburkardt/data/csv/hw_200.csv) (Can be edited to include duplicates, blanks, typos)

## 💼 Real-World Business Questions

| **Business Question** | **Cleaning Requirement** |
| --- | --- |
| “Which region had the highest sales in 2023?” | Remove blank rows, fix region names |
| “What is the average shipping time by category?” | Clean date fields, remove outliers |
| “Are there duplicate orders?” | Remove duplicate rows |
| “What are the top customers by revenue?” | Standardize customer names |

## 🧽 Cleaning Process in Excel (with Examples)

| **Step** | **Cleaning Task** | **Excel Feature** | **Example** |
| --- | --- | --- | --- |
| 1 | Remove blank rows/columns | Go To Special → Blanks | Delete all empty rows |
| 2 | Remove Duplicates | Data → Remove Duplicates | Remove duplicate Order IDs |
| 3 | Fix Typos / Inconsistent Text | Flash Fill, Find & Replace | Replace "Soth" with "South" |
| 4 | Standardize Case | =PROPER(A2) | Standardize names |
| 5 | Trim Extra Spaces | =TRIM(A2) | Clean up address or customer name |
| 6 | Date Format Cleanup | Format Cells → Date | Ensure date columns are readable |
| 7 | Convert Text to Columns | Data → Text to Columns | Split Full Name to First and Last Name |
| 8 | Fix Errors | =IFERROR() | Handle divide-by-zero |
| 9 | Handle Missing Values | Filter + Fill Down / Replace | Fill missing Region with "Unknown" |
| 10 | Filter Outliers | Sort, Conditional Formatting | Remove sales above 3 SD |
| 11 | Convert Data Types | VALUE(), TEXT(), INT() | Convert string "100" to number |
| 12 | Create a Clean Copy | Copy & Paste Values | Final clean dataset |

## 🎯 Assignment Template (You Can Use This for Class)

### ****Assignment 1 – In-Class****

**Dataset**: Superstore Sales  
**Tasks**:

Remove all blank rows and columns

Fix inconsistent region names ("Soth", "Noth", etc.)

Remove duplicates in Order ID

Standardize Product Name case (Title Case)

Trim spaces in Customer Name

### ****Assignment 2 – Practice Set 1****

**Dataset**: Dirty Customer Data  
**Tasks**:

Use TRIM(), PROPER(), CLEAN() on name and address

Split Full Name to First and Last using “Text to Columns”

Use IFERROR() to handle division in a new column "Spend per Visit"

Format Date of Birth to dd-mmm-yyyy

Filter all customers from California and count

### ****Assignment 3 – Practice Set 2****

**Dataset**: Custom Product Sales  
**Business Questions**:

Which products are priced unusually high? (Outliers)

Are there duplicates in product codes?

Which states have blank sales data?

**Cleaning Tasks**:

Identify and delete outlier sales values

Remove duplicates in Product ID

Fill blank Sales entries with 0 or “Unknown”

## 🗂 Downloadable Template

I can generate a customized Excel workbook with:

Dirty data

Predefined cleaning tasks

Solution sheet

Let me know if you’d like that.

Would you like me to **generate a custom Excel file** for these cleaning tasks and assignments?