1. Importing Data:Open Excel. Click on "File" > "Open" and select the file containing your dataset. Ensure the data is organized in columns with headers in the first row.
2. Data Overview:Scroll through the dataset to visually inspect the data.Look for empty cells, strange characters, inconsistencies, and outliers.
3. Handling Missing Values:Identify missing values: Use the "Find" function (Ctrl + F) and search for "N/A," "NaN," or blank cells.Delete Rows/Columns: Right-click on the row or column header and select "Delete" to remove entire rows/columns with missing values.
4. Fill Missing Values:For numerical data: Select the column, go to "Data" > "Data Tools" > "Data Validation." Choose "Input Message" to add context. Then go to "Data" > "Data Tools" > "Data Validation" > "Input Message" to fill missing values.For categorical data: Replace missing values with the most frequent category.
5. Handling Duplicates:Identify duplicates: Select the range of data, go to "Data" > "Data Tools" > "Remove Duplicates." Choose columns to identify duplicates and click "OK" to remove them.Evaluate duplicates: If duplicates are valid, decide how to handle them (e.g., keep the first occurrence).
6. Data Formatting:Dates and Numbers: Select the column, right-click, choose "Format Cells," and select the desired date or number format.Text: Use "LOWER," "UPPER," or "PROPER" functions to standardize text formatting.
7. Data Validation and Accuracy:Data Types: Select the column, go to "Data" > "Data Tools" > "Data Validation." Choose the appropriate data type (date, number, text) and set validation rules.Validation Rules: Set rules for allowed values, date ranges, or numerical constraints.
8. Handling Outliers:Identify outliers: Create box plots or histograms to visualize data distribution.Decide: Use filters or conditional formatting to highlight outliers. Decide to keep or remove them based on analysis goals.
9. Data Transformation:Calculated Columns: Use Excel formulas (e.g., SUM, AVERAGE) to create new columns based on existing data.Pivot Tables: Select the data range, go to "Insert" > "Pivot Table," and configure rows, columns, and values based on analysis requirements.Aggregation: Use functions like SUM, AVERAGE, MIN, MAX to aggregate data based on categories.
10. Data Visualization:Create Charts and Graphs: Select the data range, go to "Insert" > "Charts" to create various chart types (bar, line, pie, etc.). Choose appropriate chart types based on data patterns.
11. Documentation:Document Changes: Create a new worksheet for documentation. Note down the changes made, including deleted rows/columns, filled missing values, etc.Data Dictionary: Create a new worksheet for the data dictionary. List column names, descriptions, and valid values.
12. Quality Assurance:Validate Results: Double-check calculations, aggregations, and visualizations to ensure accuracy.Peer Review: Have a colleague review the cleaned dataset and documentation for errors.

13. Export and Save:Save As: Click on "File" > "Save As" and save the cleaned dataset with a new name to preserve the original data.Export: If needed, export the cleaned data to CSV, Excel, or other formats.
14. Documentation and Reporting:Document Steps: In the documentation worksheet, summarize the cleaning steps performed.Prepare Report: Create a new worksheet for the report. Include key findings, insights, and visualizations derived from the cleaned dataset.