

Assignment 2: Data Cleaning and Excel Functions

1. What is Data Cleaning & Why It Matters

Definition: Data cleaning is the process of detecting and correcting (or removing) inaccurate, incomplete, or irrelevant data to make it accurate and reliable.

Importance:

- Ensures accurate analysis and decision-making.
- Prevents errors caused by inconsistent or incorrect data.
- Improves efficiency in reporting and automation.

Consequences of Messy Data:

- Wrong conclusions and poor decisions.
 - Inaccurate analysis and reporting.
 - Wasted time fixing avoidable errors.
-

2. Common Steps in Data Cleaning

1. Back up the original dataset.
 2. Inspect data (types, missing values, duplicates).
 3. Remove duplicates.
 4. Handle missing values.
 5. Fix data types (numbers, dates, text).
 6. Trim spaces and standardize text.
 7. Validate ranges and correct errors.
 8. Handle outliers.
 9. Split/merge columns if required.
 10. Document cleaning steps.
-

3. Sorting Dataset by Department and Salary

Manual Method:

1. Select all data.
2. Go to **Data → Sort**.
3. Sort by **Department (A-Z)**.
4. Add level → Sort by **Salary (Largest to Smallest)**.

Excel Formula (Dynamic):

```
=SORT(A2:E11, {3,4}, {1,-1})
```

(Sorts first by column 3 ascending, then column 4 descending.)

4. Common Text Functions in Excel

Function	Description	Example
TRIM(text)	Removes extra spaces	=TRIM(" Excel Tips ") → Excel Tips
LEFT(text, n)	Returns first <i>n</i> characters	=LEFT("ExcelTips",5) → Excel
RIGHT(text, n)	Returns last <i>n</i> characters	=RIGHT("Data.xlsx",4) → xlsx
MID(text,start,n)	Returns substring	=MID("ExcelTips",6,4) → Tips
CONCAT(text1,text2,...)	Joins text	=CONCAT(A2," ",B2)

5. Example Exercises

Task	Formula	Result
First 5 chars of "ExcelTipsAreGreat"	=LEFT("ExcelTipsAreGreat",5)	Excel
Last 4 chars of "DataAnalysis.xlsx"	=RIGHT("DataAnalysis.xlsx",4)	xlsx
Extract "Tips"	=MID("ExcelTipsAreGreat",6,4)	Tips
Count characters in "Hello World!"	=LEN("Hello World!")	12
Extract 6 chars from middle	=MID("12345-67890-ABCDE",6,6)	-67890

6. Combining Text (Concatenation)

- **Using CONCAT:** =CONCAT("Hello"," ", "World") → Hello World
- **Using TEXTJOIN:** =TEXTJOIN(", ",TRUE,"Apple","Banana","Cherry") → Apple,Banana,Cherry
- **Using & Operator:** ="2025" & ": " & "Excel" → 2025: Excel
- **Combine First + Last Name:** =A2 & " " & B2

7. Data Validation for Quantity (1-10)

1. Select the Quantity column (e.g., C2:C100).

2. Go to **Data → Data Validation**.
 3. Settings:
 4. Allow: Whole number
 5. Data: Between
 6. Minimum: 1
 7. Maximum: 10
 8. Input Message: *"Please enter a number between 1 and 10."*
 9. Error Alert: *"Invalid input! The quantity must be between 1 and 10."*
-

8. Date Functions: TODAY() & NOW()

Function	Returns	Example
TODAY()	Current date only	=TODAY() → 09-Feb-2026
NOW()	Current date & time	=NOW() → 09-Feb-2026 12:34 PM

a. Days until 10 June 2025:

=A1 - TODAY()

(If A1 = 10-Jun-2025)

b. Static formatted date & time:

=TEXT(NOW(),"MM/DD/YYYY hh:mm AM/PM")

c. When reopened: =TODAY() updates automatically to the new date.

d. To insert a static date: Press **Ctrl + ;**

9. Summary of Key Text Functions

Function	Purpose
TRIM	Removes extra spaces
LEFT	Extracts characters from left
RIGHT	Extracts characters from right
MID	Extracts characters from middle
CONCAT	Joins multiple text values

