Name: Prarabdh Srivastava

Batch Name: Data Analytics Dec Live Batch

Assignment Name: Data Cleaning & Essential Functions

Assignment 2

Question 1) What is data cleaning, and why is it important in data analysis?

- What are the potential consequences of analyzing unclean or messy data?
- Explain the common steps involved in cleaning and organizing data.

Answer)

Data cleaning means removing wrong, missing, or duplicate data.

Example: Removing duplicate customer names

Correcting spelling mistakes (e.g., “Notebok” → “Notebook”)

Importance:

Data cleaning is important because clean data gives correct results.

- Helps in accurate analysis
- Improves data quality
- Avoids wrong conclusions

Consequences of Using Unclean or Messy Data:

If data is not cleaned, it can cause:

- Wrong analysis results
- Incorrect business decisions
- Misleading charts and reports

Common Steps in Data Cleaning and Organizing Data

1. **Remove duplicate data**
   (Same record appearing multiple times)
2. **Handle missing values**
   (Fill, delete, or replace missing data)
3. **Correct errors**
   (Spelling mistakes, wrong formats)
4. **Standardize data**
   (Same date format, same units)
5. **Remove irrelevant data**
   (Unnecessary columns or rows)
6. **Check data consistency**
   (Same values should mean the same thing)

Question 2) How would you sort the following dataset first by "Department" (A-Z) and then by "Salary" (Largest to Smallest)? Write a step-by-step approach.

Answer)

Step-by-Step Approach

1. Select the entire dataset
   (Include all rows and columns so data does not mismatch)
2. Go to the **Data tab** in Excel / Google Sheets
3. Click on **Sort**
4. In the Sort dialog box:

- Sort by: Department
- Order***: A to Z***

5. Click on Add Level
6. For the second level:

- Then by: Salary

- Order: ***Largest to Smallest***

7. Click OK

Question 3) Explain the use of text functions such as TRIM , LEFT, RIGHT, MID, and CONCAT in data cleaning**.**

Answer)

TRIM

Removes unwanted spaces before, after, and between words

Example:
" John " → "John"

LEFT

Extracts characters from the left side of text.

Example:
=LEFT("AKTU2024",4) → AKTU

RIGHT

Extracts characters from the right side of text.

Example:
=RIGHT("Roll123",3) → 123

MID

Extracts text from the middle of a string.

Example:
=MID("CS-AIML",4,4) → AIML

CONCAT

Joins multiple text values into one.

Example:
First Name + Last Name → Full Name
=CONCAT("Amit"," Kumar") → Amit Kumar

Question 4) What is the role of date functions like TODAY in managing datasets?

Answer)

- Helps track daily reports
- Useful for age calculation
- Used in deadline monitoring
- Updates automatically every day

Question 5) Apply Data Validation to restrict Quantity values to only whole numbers between 1 and 10.

Configure an input message that appears when a user selects a cell in the "Quantity" column explaining: "Please enter a whole number between 1 and 10."

Set up an error alert message that triggers if the user enters a number less than 1 or greater than 10, showing: "Invalid input! The quantity must be a whole number between 1 and 10."

Write a step-by-step approach for this question

Answer) ***STEP 1: Select the Quantity Column***

1. Click on the Quantity column cells
   (for example: E2:E50)

STEP 2: Open Data Validation

2. Go to the Data tab
3. Click on Data Validation

STEP 3: Set Validation Rules

4. In the Settings tab:
    - Allow: Whole number
    - Data: Between
    - Minimum: 1
    - Maximum: 10

This restricts quantity values to whole numbers between 1 and 10.

STEP 4: Configure Input Message

5. Click on the Input Message tab
6. Check Show input message when cell is selected
7. Enter:
    - Title: Quantity Instruction
    - Input Message:
    - Please enter a whole number between 1 and 10.

This message appears when the user selects the Quantity cell.

STEP 5: Configure Error Alert

8. Click on the Error Alert tab
9. Set:
    - Style: Stop

- Title: Invalid Input
- Error Message:
- Invalid input! The quantity must be a whole number between 1 and 10.

This message appears when the user enters an invalid value.

STEP 6: Apply

10. Click OK

Question 8) Understanding TODAY() and NOW()

a. What is the difference between TODAY() and NOW() in Excel? Provide an example of when you would use each function.

b. If cell A1 contains the date 2025-06-10, write a formula using TODAY() to determine how many days are left until that date

c. Write an Excel formula using NOW() to display the current date and time in the format MM/DD/YYYY HH:MM AM/PM.́

d. If a cell contains =TODAY(), what will happen when the worksheet is reopened the next day? Explain

e. You want to store a static date (today's date) in a cell without it changing every day. What keyboard shortcut should you use

Answer)

A) TODAY() → Used to calculate age, deadlines, or days remaining
Example:

=TODAY()

NOW() → Used to record login time, report generation time
Example:

=NOW()

B)

Formula:

=A1 - TODAY()

This gives the number of days remaining from today until that date.

C) Formula:

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

This formats date and time properly.

D) If a cell contains:

=TODAY()

- The date will automatically update to the new current date
- TODAY() is a dynamic (volatile) function

So the date changes every day.

E) Keyboard Shortcut:

Ctrl + ;

This inserts today's date as a fixed value, not a formula.