DATA CLEANING & ESSENTIAL FUNCTIONS – THEORY ANSWERS

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

Data cleaning is the process of identifying errors, removing duplicates, correcting formatting issues, and organizing raw data so it becomes accurate and ready for analysis. Clean data improves accuracy, prevents wrong conclusions, ensures reliable reports, and helps in making correct decisions.

What are the potential consequences of analyzing unclean or messy data?

Messy data can lead to incorrect results, misleading charts, wrong business decisions, duplicate records affecting totals, missing values breaking formulas, and unreliable insights.

Explain the common steps involved in cleaning and organizing data.

- Remove duplicate records
- Fix spelling and formatting issues
- Handle missing values
- Remove extra spaces using TRIM
- Standardize number and date formats
- Apply validation rules
- Sort and organize the dataset

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

TRIM removes extra spaces from text. LEFT extracts characters from the beginning of a string. RIGHT extracts characters from the end. MID extracts characters from the middle. CONCAT combines multiple text values into one. These functions help standardize messy datasets.

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

TODAY returns the current date and updates automatically. It is useful for tracking deadlines, calculating remaining days, updating reports, and managing time-based datasets.

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

TODAY returns only the current date, while NOW returns both the current date and time. TODAY is used when only date matters such as deadlines or aging reports, while NOW is used for timestamps or logging events.

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

The date will automatically update to the new current date because TODAY is a dynamic function that recalculates whenever the workbook refreshes.

8(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?

Use the shortcut CTRL + ; to insert a fixed current date.