Data Cleaning Interview Q&A (Excel-Based)

1. What are missing values and how do you handle them?

Missing values are blank cells. I handle them by either removing rows or filling them using 'Not Available', 'Unknown', or logical default values.

2. How do you treat duplicate records?

I use the Remove Duplicates tool or COUNTIF to find and delete or review duplicates as needed.

3. Difference between dropna() and fillna()?

In Excel terms:

- dropna() = delete rows with blanks

- fillna() = fill blanks using formulas or default values

4. What is outlier treatment and why is it important?

Outliers are unusual values. I find them using box plots or IQR, and treat them by capping, removing, or analyzing separately.

5. Explain the process of standardizing data.

I use the formula (value - mean) / standard deviation in Excel to bring all values to the same scale.

6. How do you handle inconsistent data formats (e.g., date/time)?

I use Excel functions like TEXT, DATEVALUE, and proper formatting to make all data consistent.

7. What are common data cleaning challenges?

Missing data, duplicates, wrong formats, typos, and inconsistent values are common. I fix them using filters, formulas, and Power Query.

8. How can you check data quality?

I check with Pivot Tables, summary stats, conditional formatting, and validation rules to spot errors and gaps.