QUESTIONS BANK

1. Describe at least three techniques used to handle missing data. What are the advantages and disadvantages of each?
2. Describe two distinct approaches to dealing with a dataset's missing data. Discuss how each technique might affect the outcomes of data analysis in your response.
3. What are the common steps involved in the data wrangling process? Explain each step briefly and mention why it is important in data analysis.
4. Differentiate between data cleaning and data transformation with examples. Include examples of missing value handling and data normalization.
5. Give examples of when to utilise each of the two popular categories of data transformation techniques.
6. In data preprocessing, distinguish between standardisation and scaling.
7. Why does data analysis require data visualisation? Give examples of two different kinds of data visualisations and the situations in which they might work well. Give examples to back up your response.
8. How can outliers be detected and treated during data preparation? Mention both statistical and visualization-based methods.
9. What is Python's Matplotlib and how is it used to visualise data? Explain the fundamental elements of a Matplotlib plot.
10. Describe how to use Matplotlib in Python to produce a 3D surface plot. What are the main roles and factors at play?
11. What are the key principles of effective data visualization? Explain how these principles help in better communication of insights.
12. You are given a dataset with inconsistent categorical values (e.g., "Male", "male", "MALE"). How would you clean and standardize this data for analysis? Write a sample approach or pseudocode.
13. Why is data wrangling a crucial phase in the data analysis process, and what does it entail? Give an explanation of at least three typical data wrangling tasks and how they help get data ready for modelling or analysis.
14. Compare and contrast bar charts, histograms, and box plots. When would you use each one?
15. Give an explanation of the data cleaning procedure and its significance for guaranteeing accurate and trustworthy data analysis outcomes. Talk about popular data cleaning methods include data transformation, outlier identification, and handling missing values.
16. How can you use visualizations to identify relationships between two or more variables? Give examples using scatter plots, heatmaps, or pair plots.

17. What are some common mistakes in data visualization, and how can they be avoided? Mention issues like misleading axes, cluttered charts, and poor color choices.
18. Talk about how crucial sophisticated visualisation is to data analysis. Give examples of how to employ at least two sophisticated visualisation approaches (such heatmaps, pair plots, or 3D visualisations) to help make sense of complicated datasets.
19. What is the role of color and labeling in improving the readability and interpretability of a visualization?
20. Describe a real-world scenario where an interactive dashboard would be more beneficial than a static report. Mention tools like Tableau, Power BI, or Plotly Dash if relevant.
21. Controlling the plot's line attributes (colour, style, width, and markers) is crucial for enhancing the chart's readability and appearance in Matplotlib. Explain how to use Matplotlib to manipulate different line attributes in a plot. Give an example of a basic line plot where you can alter the markers, width, style, and line colour.

:::