

Data Cleaning and Pre-processing

1. **Question:** What is data cleaning, and why is it important in data analysis?
 - **Hint:** Discuss the role of data cleaning in ensuring data quality and reliability.
2. **Question:** Explain the different methods for handling missing data. Provide examples.
 - **Hint:** Include methods such as removal, imputation, and interpolation.
3. **Question:** What is data transformation, and how is it applied in data pre-processing?
 - **Hint:** Discuss normalization, standardization, and scaling.
4. **Question:** How do you remove duplicates from a dataset? Provide a Python example.
 - **Hint:** Use the `drop_duplicates` function in Pandas.
5. **Question:** What are some common techniques for transforming data using a function or mapping?
 - **Hint:** Include examples like applying a log transformation or mapping categories to numerical values.
6. **Question:** How do you replace values in a dataset? Provide a Python example.
 - **Hint:** Use the `replace` function in Pandas.
7. **Question:** What are outliers, and how can they be detected and filtered from a dataset?
 - **Hint:** Discuss methods such as Z-scores, IQR, and visualization techniques.
8. **Question:** Explain vectorized string functions in Pandas. Provide examples.
 - **Hint:** Include functions like `str.replace`, `str.contains`, and `str.split`.

Data Visualization

9. **Question:** What is data visualization, and why is it important in data analysis?
 - **Hint:** Discuss how visualization helps in understanding and communicating data insights.

10. **Question:** Describe static graphical techniques and provide examples.
 - **Hint:** Include examples like histograms, bar charts, and line plots.
11. **Question:** Explain multivariate graphical techniques with examples.
 - **Hint:** Include examples like scatter plot matrices, parallel coordinates plots, and heatmaps.
12. **Question:** How can you customize plots in data visualization? Provide examples.
 - **Hint:** Discuss customization options like colors, labels, titles, and annotations.
13. **Question:** How do you create line plots using Pandas? Provide a Python example.
 - **Hint:** Use the `plot` function in Pandas.
14. **Question:** Describe how to create bar plots using Pandas. Provide a Python example.
 - **Hint:** Use the `bar` function in Pandas.
15. **Question:** What are histograms, and how do you create them using Pandas? Provide a Python example.
 - **Hint:** Use the `hist` function in Pandas.
16. **Question:** Explain density plots and how to create them using Pandas. Provide a Python example.
 - **Hint:** Use the `kde` function in Pandas.
17. **Question:** How do you create scatter or point plots using Pandas? Provide a Python example.
 - **Hint:** Use the `scatter` function in Pandas.
18. **Question:** Describe how to create static visualizations using Matplotlib. Provide examples.
 - **Hint:** Discuss functions like `plot`, `bar`, and `hist`.
19. **Question:** How can you create animated visualizations using Matplotlib? Provide examples.
 - **Hint:** Use the `FuncAnimation` class in Matplotlib.
20. **Question:** Explain how to create interactive visualizations using Matplotlib. Provide examples.
 - **Hint:** Discuss the use of libraries like Plotly with Matplotlib.

Exploratory Data Analysis (EDA)

21. **Question:** What is Exploratory Data Analysis (EDA), and why is it important?
 - **Hint:** Discuss the role of EDA in understanding data before formal modeling.
22. **Question:** Explain the EDA approach and its key steps.
 - **Hint:** Include steps like data summarization, visualization, and identifying patterns.
23. **Question:** Define descriptive statistics for numeric data and provide examples.
 - **Hint:** Discuss mean, median, mode, standard deviation, and variance.
24. **Question:** How do you measure central tendency? Provide examples.
 - **Hint:** Include mean, median, and mode.
25. **Question:** Explain how to measure variance and range. Provide examples.
 - **Hint:** Include formulas and examples.
26. **Question:** What are percentiles, and how are they used in data analysis? Provide examples.
 - **Hint:** Include examples like the 25th, 50th, and 75th percentiles.
27. **Question:** Define measures of normality and provide examples.
 - **Hint:** Discuss skewness and kurtosis.
28. **Question:** How do you count categorical data in a dataset? Provide a Python example.
 - **Hint:** Use the `value_counts` function in Pandas.
29. **Question:** Explain how to create contingency tables and their importance. Provide a Python example.
 - **Hint:** Use the `crosstab` function in Pandas.
30. **Question:** What are boxplots, and how are they used in EDA? Provide a Python example.
 - **Hint:** Use the `boxplot` function in Pandas.