

Q: What is NumPy, and why is it widely used in Python

A: NumPy is a Python library for numerical computations. It is widely used for its efficient handling of large arrays and matrices, along with mathematical functions.

Q: How does broadcasting work in NumPy

A: Broadcasting allows NumPy to perform operations on arrays of different shapes by automatically expanding the smaller array to match the shape of the larger one.

Q: What is a Pandas DataFrame

A: A Pandas DataFrame is a 2D labeled data structure similar to a table, with rows and columns, used for data manipulation and analysis.

Q: Explain the use of the groupby() method in Pandas

A: groupby() is used to split data into groups based on some criteria, then apply a function like sum, mean, or count on each group.

Q: Why is Seaborn preferred for statistical visualizations

A: Seaborn provides high-level functions to create attractive and informative statistical graphics with less code compared to Matplotlib.

Q: What are the differences between NumPy arrays and Python lists

A: NumPy arrays are more memory-efficient, support vectorized operations, and are faster for numerical computations, while lists are more flexible but slower.

Q: What is a heatmap, and when should it be used

A: A heatmap is a graphical representation of data where values are represented by colors. It is used to visualize correlations or intensity of values.

Q: What does the term 'vectorized operation' mean in NumPy

A: Vectorized operation refers to performing operations on entire arrays without explicit loops, making computations faster and more concise.

Q: How does Matplotlib differ from Plotly

A: Matplotlib is a static visualization library, while Plotly provides interactive and web-based visualizations.

Q: What is the significance of hierarchical indexing in Pandas

A: Hierarchical indexing allows multiple levels of indices, enabling easier handling of higher-dimensional data in a 2D DataFrame.

Q: What is the role of Seaborn's pairplot() function

A: pairplot() creates pairwise scatter plots and histograms to visualize relationships between multiple variables in a dataset.

Q: What is the purpose of the describe() function in Pandas

A: The describe() function provides summary statistics such as mean, median, min, max, and standard deviation of DataFrame columns.

Q: Why is handling missing data important in Pandas

A: Handling missing data prevents errors, ensures accurate analysis, and improves model performance.

Q: What are the benefits of using Plotly for data visualization

A: Plotly provides interactive, web-ready plots with zoom, hover, and export capabilities, making it useful for dashboards and presentations.

Q: How does NumPy handle multidimensional arrays

A: NumPy provides ndarray objects that can store data in multiple dimensions and supports operations like slicing, reshaping, and broadcasting.

Q: What is the role of Bokeh in data visualization

A: Bokeh is a Python library for creating interactive and web-ready visualizations with easy integration into web apps.

Q: Explain the difference between apply() and map() in Pandas

A: apply() can be used on both DataFrame and Series with functions applied across rows or columns, while map() works only on Series for element-wise operations.

Q: What are some advanced features of NumPy

A: Advanced features include broadcasting, masked arrays, memory mapping, linear algebra operations, and FFT (Fast Fourier Transform).

Q: How does Pandas simplify time series analysis

A: Pandas provides datetime objects, resampling, shifting, rolling windows, and time-based indexing for easy time series analysis.

Q: What is the role of a pivot table in Pandas

A: Pivot tables allow summarizing and reorganizing data by aggregating values based on categories, similar to Excel pivot tables.

Q: Why is NumPy's array slicing faster than Python's list slicing

A: NumPy slicing is faster because arrays are stored in contiguous memory blocks and implemented in C, while Python lists are objects with overhead.

Q: What are some common use cases for Seaborn

A: Seaborn is commonly used for heatmaps, categorical plots, regression plots, pairwise plots, and distribution visualizations.