1. What is NumPy, and how is it used in data analysis with Python?

Answer: NumPy is a fundamental library for numerical computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays. It's essential for performing efficient numerical operations in data analysis tasks.

---

2. Explain the key differences between Pandas and NumPy.

Answer:

NumPy is focused on numerical arrays and provides a high-performance multidimensional array object. Pandas, on the other hand, offers data structures like Series and DataFrames that are more flexible and designed for labeled and relational data. Pandas builds on NumPy and is more suited for data manipulation and analysis tasks.

---

3. What is a DataFrame in Pandas?

Answer:

A DataFrame is a two-dimensional, labeled data structure in Pandas that can hold data of different types. It is similar to a spreadsheet or SQL table and is commonly used for data manipulation and analysis.

---

4. How do you handle missing data in a DataFrame?

Answer:

Missing data can be handled in Pandas using methods like `dropna()` to remove missing values or `fillna()` to fill them with specified values. For example, to fill missing values with the mean of the column:

```
df['column'].fillna(df['column'].mean(), inplace=True)
```

This approach ensures that the dataset remains complete for analysis.

---

6. How do you merge two DataFrames in pandas?

Answer:

You can merge two DataFrames using the `pd.merge()` function, which is similar to SQL JOIN operations. For example:

```
merged_df = pd.merge(df1, df2, on='common_column', how='inner')
```

This merges `df1` and `df2` on the `common_column` using an inner join. Other types of joins (inner, outer, left, right) can be specified using the `how` parameter.

---

7. What is the purpose of the `groupby()` function in pandas?

Answer:

The `groupby()` function in pandas is used to split data into groups based on some criteria. It is often followed by an aggregation function like `sum()`, `mean()`, `count()`, etc., to apply a function to each group independently. This is particularly useful for summarizing and analyzing large datasets, such as calculating the total sales for each product category.

---

8. How do you handle duplicate values in a DataFrame?

Answer

To handle duplicates in a DataFrame:

- Use `df.duplicated()` to identify duplicate rows.
- Use `df.drop_duplicates()` to remove duplicate rows.

These methods help in cleaning the dataset by ensuring that each record is unique.

---

9. What are lambda functions in Python, and how are they used?

Answer: Lambda functions are small, anonymous functions defined using the `lambda` keyword. They can take any number of arguments but can only have one expression. For example:

```
add_ten = lambda x: x + 10
```

Lambda functions are commonly used for short-term operations, such as within the `apply()` function in pandas.

---

10. How do you read and write data from/to a CSV file using Pandas?

Answer: