Pandas Interview Questions with Answers

Basics

1. What is Pandas in Python?

Pandas is an open-source Python library primarily used for data manipulation, analysis, and cleaning. It offers two primary data structures: Series (1-dimensional) and DataFrame (2-dimensional), which facilitate efficient handling of structured data. Pandas is built on top of NumPy and supports operations such as data loading, visualization, transformation, and aggregation.

2. What are the main data structures in Pandas?

- Series: A one-dimensional labeled array capable of holding any data type. It's similar to a column in a spreadsheet or database.
- DataFrame: A two-dimensional, tabular data structure with labeled axes (rows and columns). It can hold data of different types.
- Panel: A rarely used 3D data structure capable of storing heterogeneous data, mainly for legacy purposes.

3. How do you create a DataFrame in Pandas?

You can create a DataFrame using a variety of inputs, such as:

```
import pandas as pd
data = {'Name': ['Alice', 'Bob'], 'Age': [25, 30]}
df = pd.DataFrame(data)
```

This creates a DataFrame with columns 'Name' and 'Age'.

4. How do you check the first and last five rows of a DataFrame?

- First five rows: df.head()

- Last five rows: df.tail()

These methods provide a quick look at the data.

5. What is the difference between .loc[] and .iloc[]?

- .loc[] is used for label-based indexing, i.e., selecting data by row/column names.

- .iloc[] is used for positional indexing, i.e., selecting data by row/column positions.

... (Include the rest of the content from previous message here for full PDF) ...