### **Basic Questions**

1. What is Pandas and why is it used?

**Pandas** is an **open-source** Python library designed for data **manipulation**, **analysis**, and **cleaning**. It provides easy-to-use data structures and data analysis tools, making it a **go-to library** for working with structured data (like tables, spreadsheets, or SQL query results).

### **🚀 Why is it used?**

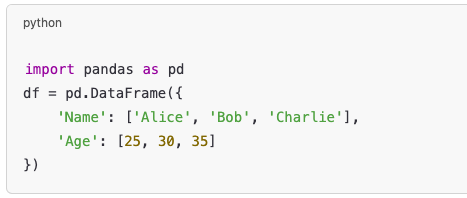
* **Data Cleaning:** Handling missing data, duplicates, and type conversions.
* **Data Transformation:** Adding new columns, renaming, or restructuring.
* **Data Analysis:** Aggregations, statistical summaries, and group operations.
* **Data Exploration:** Quick visualization and descriptive statistics.
* **File I/O:** Read and write from multiple file types (CSV, Excel, SQL, etc.).

1. What are the main data structures in Pandas?

A **Series** is a **one-dimensional** labeled array capable of holding data of any type (integers, floats, strings, etc.).

A **DataFrame** is a **two-dimensional** table, similar to a spreadsheet or SQL table.

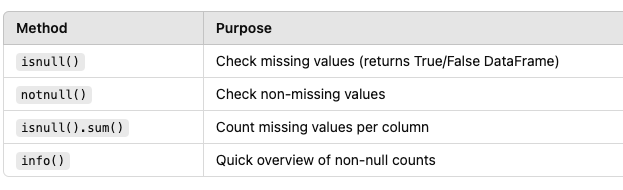
1. How do you create a DataFrame in Pandas?



1. How to read and write CSV files using Pandas?

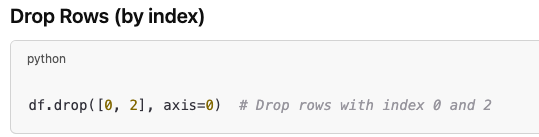
df = pd.read\_csv(“url”)

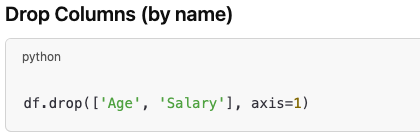
1. How do you check for missing values in a DataFrame?





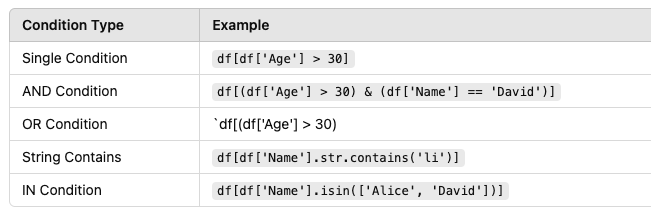
1. How do you drop rows or columns from a DataFrame?







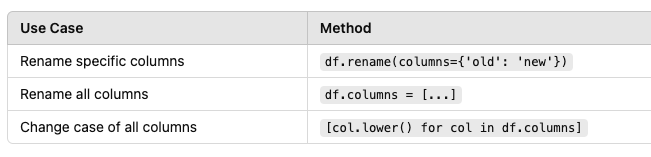
1. What is the difference between loc[] and iloc[]?
2. How do you filter rows in a DataFrame based on a condition?



1. How do you sort a DataFrame by column values?

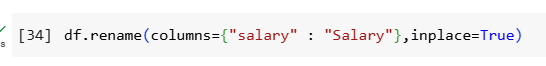


1. How can you rename columns in a DataFrame?



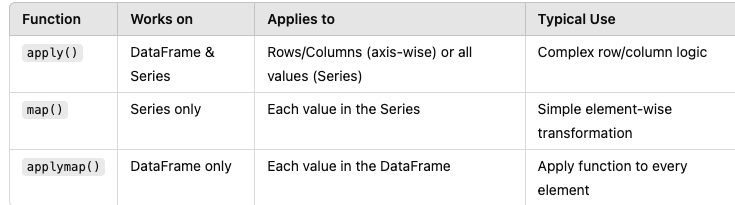
### 



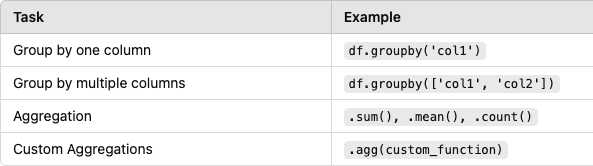


### **Intermediate Questions**

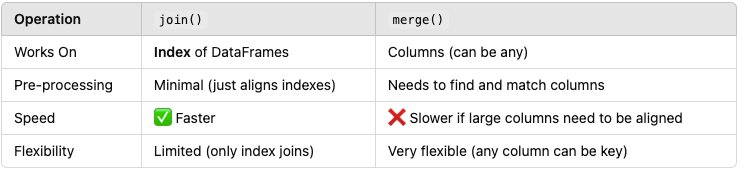
1. What is the difference between apply(), map(), and applymap()?



1. How to group data in Pandas? Explain groupby().

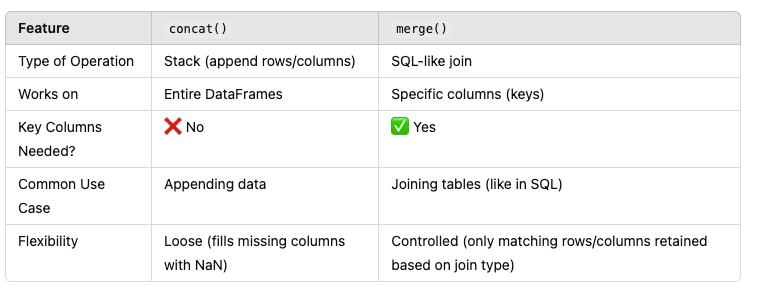


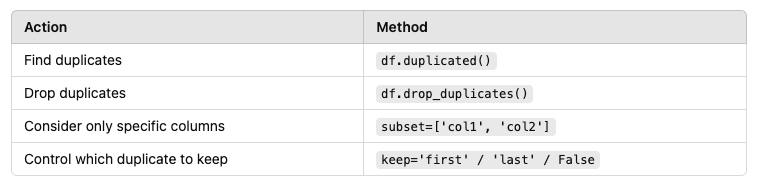
1. How to merge/join two DataFrames?



1. What is the difference between concat() and merge()?

Concat is between two series, merge can be between two dfs



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

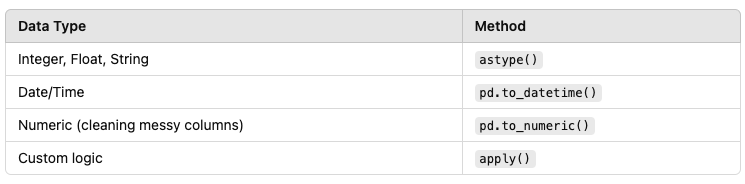
# Drop duplicates based on all columns, keeping the first occurrence df\_unique = df\_drop\_duplicated(df) print("DataFrame with duplicates removed (keeping first):\n", df\_unique)

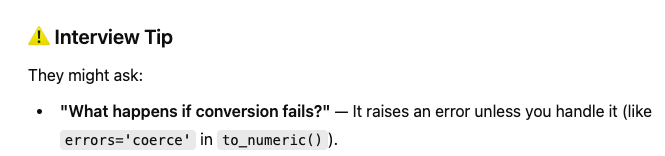
# Drop duplicates based on 'col1', keeping the last occurrence df\_unique\_col1\_last = df\_drop\_duplicated(df, subset=['col1'], keep='last') print("\nDataFrame with duplicates removed (col1, keeping last):\n", df\_unique\_col1\_last)

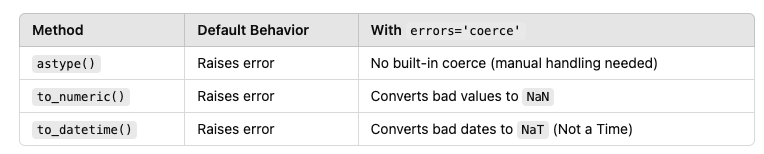
# Drop all duplicates based on 'col1' df\_unique\_col1\_all = df\_drop\_duplicated(df, subset=['col1'], keep=False) print("\nDataFrame with all duplicates removed (col1):\n", df\_unique\_col1\_all)

# Inplace example df\_copy = df.copy() df\_drop\_duplicated(df\_copy, inplace=True) print("\nDataFrame with duplicates removed inplace:\n", df\_copy)

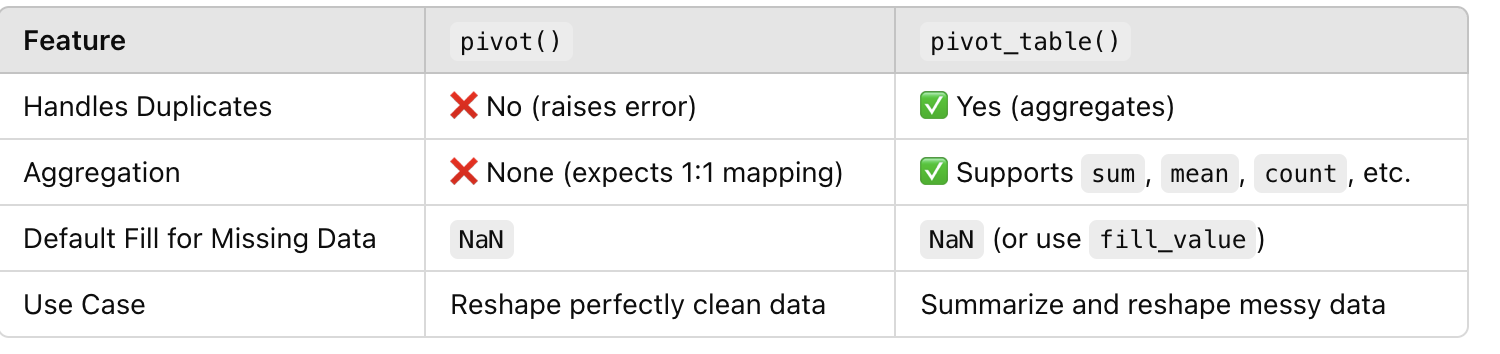
1. How do you change the data type of a column in Pandas?

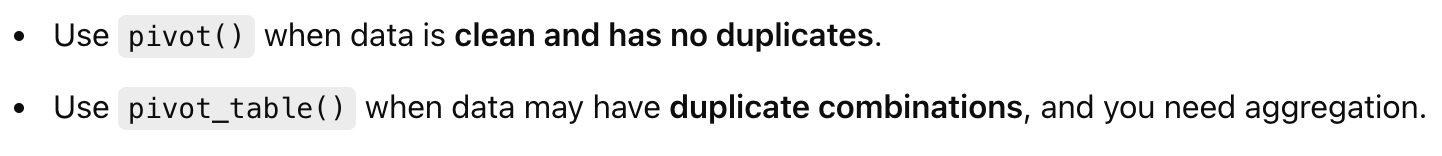




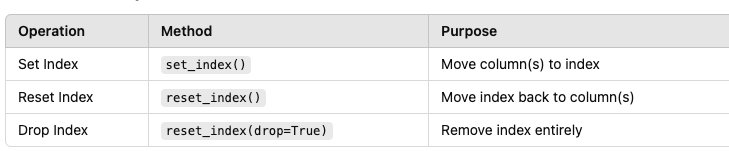


1. What’s the difference between pivot() and pivot\_table()?

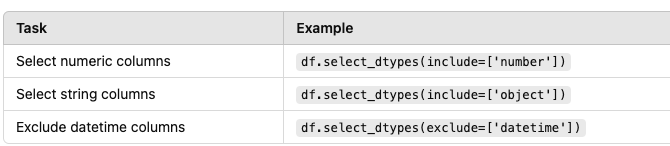




1. Explain how to reset and set an index in a DataFrame.

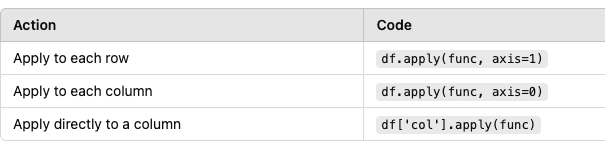


1. How do you select columns based on data types?

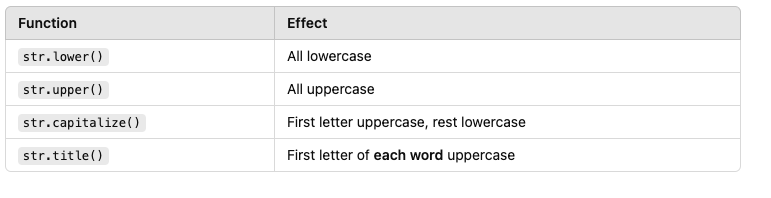




1. How to apply a custom function to each row or column in a DataFrame?

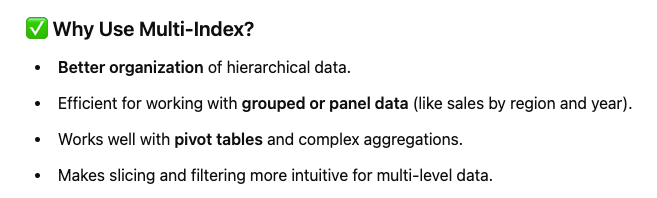


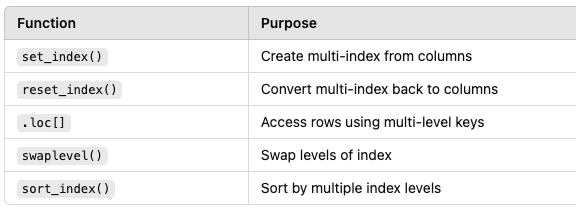
1. Difference between title() and capitalize()

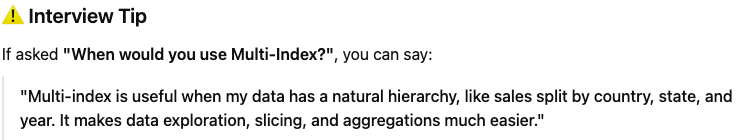


### **Advanced Questions**

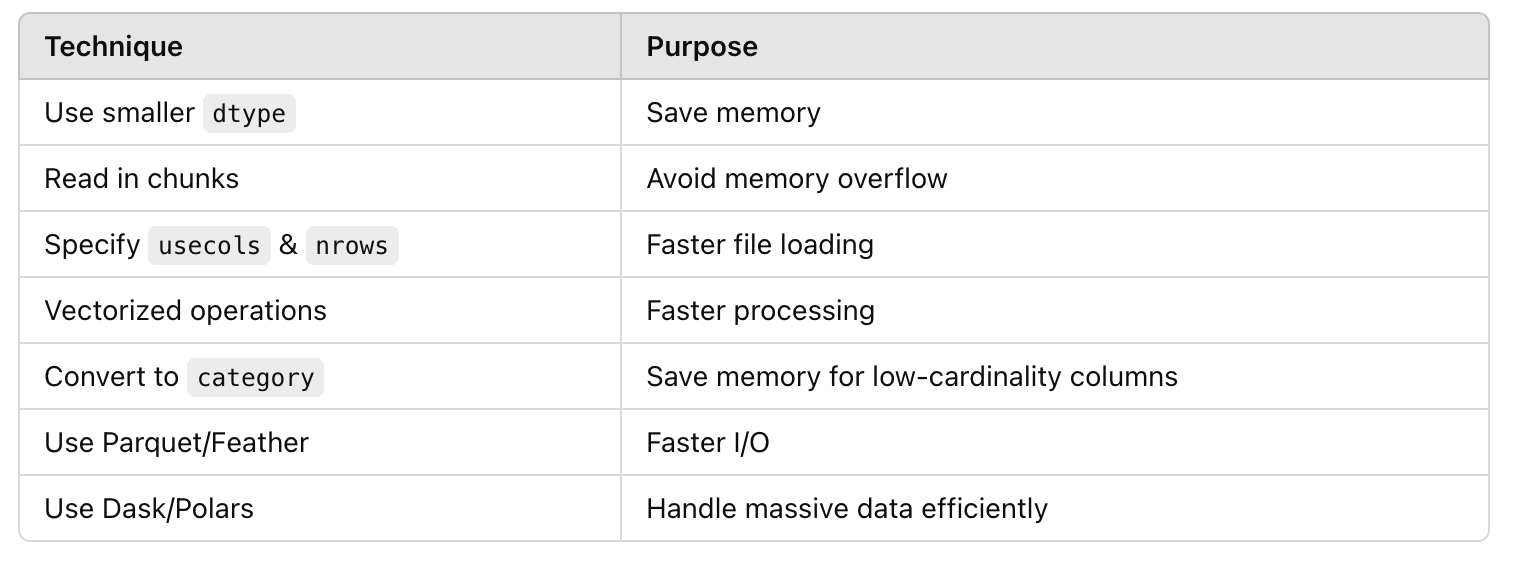
1. What is multi-indexing in Pandas?



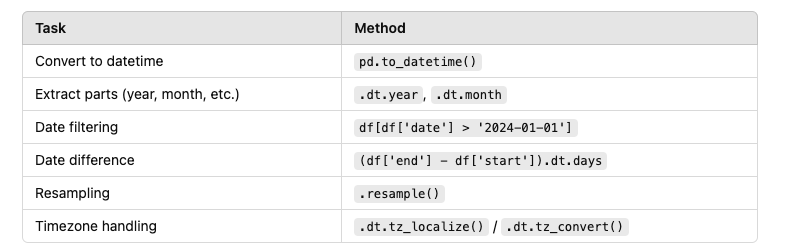




1. How do you efficiently work with very large datasets in Pandas?



1. How do you handle datetime data in Pandas?

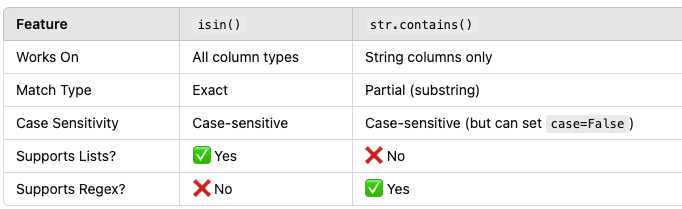


1. How do you perform window functions like rolling mean in Pandas?



1. How to read specific rows/columns from a large file in Pandas?
2. What are some ways to improve Pandas performance for large data?

1. How to reshape data using melt() and stack()/unstack()?
2. How to detect and handle outliers using Pandas?
3. Explain the difference between isin() and str.contains().



1. How do you save a DataFrame to Excel with multiple sheets?

### **Scenario-Based/Case Study Questions**

1. How would you find the top 5 most occurring values in a column?
2. You have a DataFrame with duplicate rows — how will you keep only the first occurrence?
3. You need to calculate cumulative sales by product — how would you do it?
4. How would you split a column containing comma-separated values into multiple columns?
5. How would you find correlation between numeric columns in a DataFrame?