

## 53 Pandas interview questions to ask data analyst and scientist candidates

### Questions

---

1. What is Pandas, and why is it used?
2. What are DataFrames in Pandas?
3. How would you handle missing data in a Pandas DataFrame?
4. Can you explain the difference between `loc` and `iloc` in Pandas?
5. What is a Series in Pandas?
6. How do you merge two DataFrames in Pandas?
7. What are some common operations you can perform on a Pandas DataFrame?
8. How would you convert a Pandas DataFrame to a CSV file?
9. What is the significance of the `groupby()` function in Pandas?
10. How do you handle large datasets in Pandas?
11. How do you filter rows in a DataFrame based on a specific condition?
12. Can you explain what a Pivot Table is in Pandas and how to create one?
13. What is the purpose of the `apply()` function in Pandas, and can you provide an example of its use?
14. How would you concatenate two DataFrames vertically and horizontally in Pandas?
15. What is the difference between the `astype()` and `convert_dtypes()` methods in Pandas?
16. How can you sort a DataFrame by multiple columns?
17. Can you explain how to handle date and time data in Pandas?
18. What are the advantages of using the `query()` method in Pandas?
19. How do you create a new column in a DataFrame based on existing columns?
20. What steps would you take to visualize data from a Pandas DataFrame?
21. How would you identify and remove duplicate rows in a DataFrame?
22. Can you explain the concept of indexing in Pandas and why it is useful?
23. What is chaining in Pandas, and why should it be avoided?
24. How would you handle categorical data in a DataFrame?
25. What are the common ways to group data in Pandas?
26. How do you perform data cleaning in Pandas?
27. Can you explain the difference between a shallow copy and a deep copy in Pandas?
28. What is the purpose of the `pd.cut()` function, and when would you use it?
29. How do you handle time series data in Pandas?
30. What are some common pitfalls when using Pandas, and how can they be avoided?
31. How would you reshape a wide DataFrame to a long format?
32. Explain how you would use the `melt()` function in Pandas.
33. What's the difference between `map()`, `apply()`, and `applymap()` in Pandas?
34. How can you efficiently handle outliers in a DataFrame column?
35. Describe a situation where you'd use the `pivot()` function.
36. How would you calculate rolling averages on a time series DataFrame?
37. Explain the concept of method chaining in Pandas and provide an example.
38. How can you create a new column based on conditional logic from other columns?
39. What's the purpose of the `agg()` function and how would you use it?
40. How would you handle multi-level indexing in a DataFrame?
41. Explain how to perform a custom sort on a DataFrame using a specific order of categories.
42. How can you efficiently find and replace values in a DataFrame?
43. Describe how you'd use the `cut()` function to bin continuous data.
44. How would you handle time zone conversions in a DataFrame with datetime data?
45. How would you describe the process of grouping data in Pandas and why is it useful?
46. What techniques would you use to identify and handle outliers in a Pandas DataFrame?
47. Can you explain the concept of a rolling window in Pandas and provide an example of its use?
48. How do you ensure data quality while performing data analysis tasks with Pandas?
49. What strategies would you use to improve the performance of data operations in Pandas?
50. How do you merge multiple DataFrames with different columns in Pandas?
51. How do you handle time zone conversions in a DataFrame with datetime data?
52. What methods would you use to visualize data from a Pandas DataFrame?
53. How would you go about reshaping a DataFrame from wide to long format?