

2-Mark Questions

1. What is the primary use of NumPy in data science?
2. Define the role of Pandas in handling tabular data.
3. Name one advantage of using Matplotlib for data visualization.
4. What is the purpose of Seaborn in Python data science workflows?
5. What is an Integrated Development Environment (IDE) for Python? Give one example.
6. Explain the term "data filtering" in Pandas with a simple example.
7. How does the `sort_values()` function work in Pandas?
8. What is a heatmap, and how is it useful in data analysis?
9. Name two types of plots used to visualize data distributions.
10. What is the purpose of data ranking in Pandas?
11. Name the four fundamental libraries in data science: NumPy, Pandas, Matplotlib, and Seaborn, and state their primary purposes.
12. Define an Integrated Development Environment (IDE) and give one example for Python data science.
13. What is the role of NumPy in handling numerical data?
14. Explain the purpose of Pandas for data manipulation.
15. What is data reading in Pandas, and name a common function used for it?
16. Define data selecting in Pandas with an example.
17. What is filtering data in Pandas?
18. Explain sorting data using Pandas.
19. What is a basic plot in Matplotlib, and give one type?
20. How is Seaborn used for visualizing data distributions?
21. Define outliers in data analysis.
22. What is a heatmap, and which library is commonly used to create it?
23. Explain grouping data in Pandas.
24. What is data rearranging in Pandas?
25. Define ranking data in Pandas.
26. Name two methods for detecting outliers using plots.
27. What is data manipulating in the context of Pandas?
28. Give an example of a basic plot for distributions.
29. Why is an IDE like Jupyter Notebook popular in data science?

3-Mark Questions

1. Explain how NumPy arrays differ from Python lists in data science applications.
2. Describe the process of reading a CSV file using Pandas with a code example.
3. How does Matplotlib help in creating a scatter plot? Provide a brief code snippet.
4. What is the role of Seaborn in enhancing Matplotlib visualizations? Give an example.
5. Explain how Jupyter Notebook supports data science tasks as an IDE.
6. Describe the `groupby()` function in Pandas with a simple example.

7. How can a box plot in Seaborn help detect outliers in a dataset?
8. What is data rearranging in Pandas, and provide an example using pivot() or melt().
9. Explain how histograms are used to visualize data distributions in Matplotlib.
10. How does the rank() function in Pandas work for data ranking? Provide an example.
11. Briefly explain the introduction to Python in data science, including its key features.
12. Describe the differences between NumPy and Pandas in data science applications.
13. What is the role of Matplotlib and Seaborn in data visualization? Provide one example for each.
14. Explain how an IDE like Jupyter Notebook facilitates data science workflows.
15. Describe the process of reading a dataset using Pandas, with a code example for CSV files.
16. How can you select and filter data in Pandas? Provide a simple code snippet.
17. Explain data sorting and grouping in Pandas with examples.
18. What is data rearranging and ranking in Pandas? Illustrate with brief examples.
19. Describe basic plots in Matplotlib, such as line and bar plots, with their uses.
20. How does Seaborn help in visualizing data distributions? Provide a code example for a histogram.
21. Explain the concept of detecting outliers using box plots in Seaborn.
22. What is a heatmap in data science, and how is it created using Seaborn? Include a code snippet.
23. Discuss data manipulation operations in Pandas, including selecting and filtering.
24. How can grouping and aggregation be performed in Pandas? Provide an example.
25. Explain the importance of sorting and ranking in data analysis with Pandas examples.
26. Describe how to rearrange data using pivot tables in Pandas.
27. What methods can be used to visualize and detect outliers in distributions?
28. Compare Matplotlib and Seaborn for creating heatmaps.
29. Explain the integration of NumPy with Pandas for data operations.
30. How does an IDE support code execution and visualization in data

science? **5-Mark Questions**

1. Provide an introduction to Python and its role in data science, highlighting key features and advantages.
2. Discuss the fundamental libraries NumPy, Pandas, Matplotlib, and Seaborn, explaining their interconnections in data science pipelines.
3. Explain the concept of an Integrated Development Environment (IDE) in Python. Compare two IDEs like Jupyter Notebook and VS Code for data science tasks.
4. Describe the complete process of data operations in Pandas: reading, selecting, filtering, and manipulating data, with code examples.
5. How can sorting, grouping, rearranging, and ranking be applied in Pandas for exploratory data analysis? Provide a detailed example using a sample dataset.
6. Discuss basic plots in Matplotlib and Seaborn, including line, scatter, and bar plots, with code snippets and use cases.

7. Explain the methods for visualizing data distributions using histograms and kernel density plots in Seaborn, including code examples.
8. Describe techniques for detecting outliers in datasets using box plots and scatter plots. Provide code examples with Matplotlib and Seaborn.
9. What is a heatmap, and how is it used to visualize correlations or patterns in data? Include a detailed code example using Seaborn and discuss its applications.
10. Integrate all data operations: Explain a workflow from reading data to ranking and grouping, using Pandas and NumPy, with code.
11. Discuss the role of plotting in data science: from basic plots to advanced visualizations like heatmaps for outlier detection.
12. How do NumPy and Pandas support data manipulation, and how can visualizations in Matplotlib/Seaborn aid in understanding manipulated data? Provide an example workflow.
13. Explain the importance of IDEs in managing data science projects, with a focus on features for code, data, and visualization integration.
14. Describe a comprehensive approach to handling outliers: detection via plots, filtering in Pandas, and visualization confirmation. Include code.
15. Discuss how heatmaps can be used in conjunction with data operations like grouping and rearranging to analyze complex datasets. Provide examples.
16. Provide an end-to-end example: Introduce Python libraries, use an IDE setup, perform all data operations, and visualize with plots including distributions and heatmaps.
17. Compare the capabilities of Matplotlib and Seaborn for basic plots, distribution visualization, outlier detection, and heatmaps.
18. Explain data ranking and rearranging in depth, with Pandas code examples, and visualize the results using appropriate plots.
19. Discuss the integration of all topics: From Python introduction to advanced plotting for insights in data science.
20. How can data science libraries handle real-world data tasks? Illustrate with examples covering all operations and plotting techniques mentioned.