BCA (DS & AI) Question Bank for Second Sessional Test and Final Paper 2024-25
Fourth Semester

BCAD14202: Data Science

---

1. Define Data Science and list its key components.
2. Explain the significance of Data Science in today's industries.
3. What are the various roles in a Data Science team? Explain each of them.
4. Explain the Data Science lifecycle.
5. Differentiate between Data Science and Business Intelligence.
6. Write short notes on descriptive, predictive, and prescriptive analytics.
7. Describe the CRISP-DM methodology.
8. Explain the purpose of IBM Watson Studio in Data Science projects.
9. What is the importance of data collection in a Data Science project?
10. Explain any two Data Science domains with examples.
11. Describe how IBM Cloud is used for Data Science.
12. What is the function of a Jupyter Notebook in Data Science?
13. Write a short note on IBM Watson Studio components.
14. Explain the role of data engineers vs. data scientists.
15. How does cloud technology ensure data privacy and compliance?
16. What is cloud computing? How is it used in Data Science?
17. Explain any two cloud platforms used in Data Science projects.
18. What is a cloud-based Data Science lifecycle? Describe its phases.
19. How can data be stored and retrieved securely in the cloud?
20. Explain the role of cloud infrastructure in hosting and scaling Data Science models.
21. What are integrated environments in the cloud? Give examples.
22. List tools that support data exploration in cloud environments.
23. What are the benefits of using cloud platforms for collaborative Data Science?
24. Explain the difference between on-premise and cloud-based data analytics.
25. What are some common cloud storage formats and their use in Data Science?
26. Describe challenges in preparing data on the cloud.
27. What is data wrangling? How is it performed in cloud environments?
28. Explain the importance of understanding business needs before model building.

29. What steps should be followed to understand a company's business?
30. What are the various techniques used for transforming data during dataset preparation?
31. What are data representation techniques? Explain with examples.
32. Explain data transformation with respect to unstructured data.
33. Describe any three data transformation tools.
34. What is the role of statistics in Data Science?
35. Explain mean, median, mode, and standard deviation with examples.
36. Write a Python code to generate random numbers using NumPy and calculate mean & standard deviation.
37. What are common types of graphs used for data visualization?
38. Explain the importance of data modeling in machine learning.
39. What are some popular open-source Data Science frameworks?
40. Write a Python code using Pandas to drop rows with missing values.
41. Explain model deployment and the need for it.
42. How does AutoAI help in building models automatically?
43. Write a Python program using Pandas to group data by 'Department' and calculate average salary.
44. Write a Matplotlib program to display a bar chart for 4 quarters.
45. Write short notes on Numy, pandas, Matplotlib, and seaborn.
46. Use Seaborn to create a boxplot from a DataFrame.
47. Describe broadcasting in NumPy along with its types. Why is it important in array computations?
48. Write a NumPy program to create a 2x5 array, extract specific rows and columns using slicing, reshape it, and calculate its total sum, min value, and variance.
49. How does the merge() function work in Pandas? Describe its role in joining DataFrames.
50. Write a Pandas program to merge two DataFrames and filter by a condition. Assume any columns name.
51. Describe how melt() and its reverse operation (pivot() or pivot_table()) are used in Pandas for reshaping data.
52. Write a Python program using Pandas to read a CSV file and display only rows where salary <= 30000.
53. What are the various types of plots provided by Matplotlib? Explain with examples.
54. Differentiate between supervised and unsupervised learning.
55. Write short notes on regression and classification. Support your explanation with suitable code examples.
56. What are binary and multi-class classification problems? Give examples.
57. Compare linear regression and logistic regression.
58. Write a Python program to implement Linear Regression using sklearn on a sample dataset. Assume any datasets.
59. Explain the importance of Machine Learning in extracting insights from data
60. Describe how the Gradient Descent algorithm works in optimizing machine learning models.
61. Compare Regression and Classification models with examples.
62. Write a Python code to split data into train and test sets using Scikit-learn.
63. Define overfitting and explain how it can be avoided.
64. What is the Machine Learning Framework in IBM Cloud?
65. Write a Python program to implement any classification algorithm using sklearn on a sample dataset. Assume any datasets

---