MEDI-CAPS UNIVERSITY
Data Science

Question Bank

UNIT-I

These are a few sample questions, these does not mean that the questions in the university exams are restricted to only below questions.

1. What is data science.
2. Write down different types of data we have.
3. Define structured, semi structured, and unstructured data with example.
4. Why do we need data science? explain.
5. Explain importance of data science with real life example.
6. Who are data scientists?
7. Explain briefly about Big Data.
8. Differentiate between big data and data science.
9. Differentiate between business intelligence and data science.
10. Write down some of the main components of data science.
11. Write down the data science hierarchy of need with necessary figure.
12. List the different data science techniques.
13. Write different data science framework explain any three.
14. Differentiate between qualitative data and quantitative data, providing examples for each type.
15. Describe the skills that are typically required for a career as a data scientist.
16. Define nominal data and ordinal data and provide examples of each.
17. Explore the various domains where data scientists can make a significant impact, providing examples and explaining the relevance of data science in those domains.
18. Differentiate between qualitative and quantitative data. Provide detailed examples of each type and discuss the importance of understanding data types in statistical analysis.
19. Explain the distinctions between discrete data and continuous data. Use real-world examples to illustrate the differences and discuss how these data types are used in statistical analysis.
20. Define structured data and explain why it is typically stored in a database. Provide an example of structured data.
21. Describe semi-structured data and mention its advantages in terms of analysis. Provide an example of semi-structured data.
22. What is unstructured data, and why is it not suitable for traditional relational databases? Give examples of unstructured data.
23. Explain the fundamental concept of data science and its role in extracting insights from various types of data.
24. Explain the concept of "Big Data" and outline the key characteristics that define data as "big data."
25. Differentiate between structured, semi-structured, and unstructured data. Provide examples of each type and explain their relevance in data analysis.

26. Discuss the primary roles of Data Scientists and their significance in the field of data analysis.
27. Elaborate on the components of Data Science. How does Data Strategy differ from Data Engineering, and what role does each play in the data science process.
28. Compare and contrast structured, semi-structured, and unstructured data, highlighting their characteristics and use cases. Provide examples for each type and explain how organizations can leverage them for analysis.
29. Discuss the multidisciplinary nature of data science and how it applies principles from fields like computer science, mathematics, and statistics to analyze large datasets. Provide real-world examples of how data science is used in different industries.
30. Explore the mathematical and statistical foundations of machine learning used in data science. Discuss the importance of linear algebra, probability theory, multivariate calculus, and other mathematical concepts in machine learning and data analysis.
31. Describe the evolution of Data Science and its significance in today's data-driven world. Provide real-life examples of how businesses and organizations utilize Data Science for decision-making.
32. Compare and contrast the concepts of Business Intelligence (BI) and Business Analytics (BA). Explain the key differences, and discuss the scenarios in which each is more suitable for use.
33. Examine the four characteristics of "Big Data" (Volume, Velocity, Variety, and Complexity). Provide detailed explanations of each characteristic and discuss their implications for data analysis.
34. Explore the relationship between Data Science and Machine Learning. How do these fields overlap, and what distinguishes them from each other? Provide examples of applications in which both are utilized.

UNIT-II

1) What is mean, median and mode?
2) What is a sample space? Give an example.
3) Write a note on different methods determining the probability of an event.
4) What is an empirical method? Give an example.
5) Write a note on Mean, Median and Mode with necessary example
6) What is classical method? Give an example.
7) Write note on Normal distribution.
8) What is Random Variable?
9) What is probability? Give an example
10) What are populations and samples? Explain
11) What is the difference between Random Variables and Sample space
12) What are the methods to determine the probability of an event? Explain
13) Explain variance and standard deviation with necessary example of its usage in data science.

14) Discuss the measures of central tendency, including Mean, Median, and Mode. Provide real-world examples where each of these measures is useful in data analysis.
15) Elaborate on the concept of a Random Variable in probability. How does it relate to the outcomes of a random experiment? Provide practical applications of Random Variables in real-life scenarios.
16) Explain the characteristics of a Normal Distribution and how it differs from other data distributions. Describe the relationships between Mean, Median, and Mode in a Normal Distribution
17) What is sample space in probability?
18) What is variance and standard deviation?
19) Explain skewness and kurtosis in detail.
20) Compare and contrast the classical, empirical, and subjective methods of determining the probability of an event. Provide examples of situations where each method would be appropriate.
21) Explain with an example, how Skewness and Kurtosis are going to be useful in data analysis?
22) What are descriptive, predictive and prescriptive analytics?
23) Explain, how Markov chains useful in data science with the help of necessary example?
24) What is the trajectory of a Markov chain?
25) Write a note on the Markov process with the help of an example.
26) What is Markov property.

UNIT-III & IV

1. Write a note on Exploratory Data Analysis
2. What are the purposes of Exploratory Data Analysis?
3. What are variable, value, observation and tabular data?
4. How do we handle missing data in the data set?
5. What are the different types of Exploratory Data Analysis?
6. Write a note on univariate non-graphical Exploratory Data Analysis
7. Write a note on spread.
8. What is the role of Inter Quartile Range? Explain
9. How do we calculate the Inter Quartile Range for odd number of samples?
10. How do we calculate the Inter Quartile Range for even number of samples?
11. What is univariate graphical Exploratory Data Analysis?
12. Write a note on Histogram with example.
13. Write a note on stem and leaf plot with example.
14. What is Box plot? How is it used in Univariate Exploratory Data Analysis?
15. Explain QQ Plot? How is it used in Univariate Exploratory Data Analysis?
16. Write a note on Covariance and Correlation.
17. Explain how the risk analysis takes place in a univariate analysis?

18. Explain with an example how the Covariance and Correlation used in multi variate analysis?
19. Write a note on Multivariate Analysis.
20. What is cross tabulation? Explain
21. What is Correlation?
22. What is Covariance?
23. Write a note on ANOVA.
24. Explain with an example, how the one-way ANOVA works?
25. Write a note on Multi Variate graphical EDA.
26. Explain the use of scatter plots in Multivariate analysis.
27. Explain the use of Box plots in Multivariate analysis.
28. Explain how the loss of money in business can be identified in Exploratory Data Analysis?
29. Explain some of the important and common steps when performing EDA.
30. Define EDA
31. What are Variable, Value and Observation?
32. What is the aim of Exploratory Data Analysis?
33. How to find correlation?
34. How to find covariance?
35. Explain any one of the industry project of data science in detail.

UNIT-V

1) Compare some of the available Data Science tools
2) Write a note on Python libraries for Data Science
3) What is the purpose of Pandas? Give examples
4) Write a note on Data Frame data types
5) How to do a Data Frame processing?
6) How to do the grouping, filtering, and slicing. Give examples
7) How to handle missing values using python
8) How the descriptive statistics works in python
9) Why RDBMS is not suitable for BIG DATA
10) How to deal with the scalability in handling Big Data
11) Write a note on NOSQL
12) Write a note on BASE Transactions
13) Compare ACID properties with BASE
14) Write a note on NOSQL properties
15) Write a note on NOSQL database types
16) Write a note CAP theorem
17) Write a note on challenges and Scope of Data Science project management