Question Bank_DADV

Unit 1: Introduction to Data Analytics

2-Mark Questions

1. Define data analytics.
2. What is the importance of data in decision making?
3. List the types of data analytics.
4. Differentiate between data analysis and data analytics.
5. What are the main elements of data analytics?
6. Who is a data analyst?
7. Who is a data scientist?
8. Write any two examples of how data analytics is used in real life.
9. What do you mean by descriptive analytics?
10. Define predictive analytics.

5-Mark Questions

11. Explain the importance and role of data in modern organizations.
12. Discuss the types of data analytics with suitable examples.
13. Compare and contrast data analyst and data scientist.
14. Describe the classification of data analytics.
15. Explain the key elements of data analytics with a neat diagram.

Unit 2: Python Fundamentals and Statistics

2-Mark Questions

16. Why is Python preferred for data analytics?
17. List the features of Python that make it suitable for data science.
18. What are the levels of data measurement?
19. Define mean, median, and mode.
20. Define standard deviation and variance.
21. What is dispersion in statistics?
22. Define population and sample.
23. What is a confidence interval?

24. Define sampling error.
25. What is the significance of sample mean distribution?

5-Mark Questions

26. Explain the importance of Python in data analytics.
27. Discuss the levels of data measurement with suitable examples.
28. Explain the measures of central tendency and dispersion.
29. Describe population, sample, and variance with examples.
30. Explain the concept of confidence interval estimation in statistics.

Unit 3: Probability and Types of Testing

2-Mark Questions

31. Define probability.
32. What is a probability distribution?
33. List the types of probability distribution.
34. Define hypothesis testing.
35. What are the steps in hypothesis testing?
36. What is a null hypothesis?
37. What is an alternative hypothesis?
38. What is a t-test used for?
39. What is an ANOVA test?
40. What is a Chi-square test?

5-Mark Questions

41. Explain probability and probability distribution with examples.
42. Discuss sampling and sampling distribution in detail.
43. Explain the concept and steps of hypothesis testing with an example.
44. Compare t-test, ANOVA test, and Chi-square test.
45. Explain one-way ANOVA with its assumptions and interpretation.

Unit 4: Regression, Classification, and Clustering

2-Mark Questions

46. What is regression analysis?
47. Differentiate between linear and logistic regression.
48. Define independent and dependent variables.
49. What is clustering?
50. What is K-Means clustering?
51. What is hierarchical clustering?
52. What is a decision tree?
53. Define confusion matrix.
54. What is supervised learning?
55. What is unsupervised learning?

5-Mark Questions

56. Explain linear regression with an example and equation.
57. Describe logistic regression and its applications.
58. Explain the steps involved in K-Means clustering.
59. Differentiate between K-Means and hierarchical clustering.
60. Explain decision tree algorithm and its working.
61. Describe the components and interpretation of a confusion matrix.
62. Compare regression, classification, and clustering approaches.
63. Discuss the importance of model evaluation metrics in classification.
64. Explain how clustering helps in customer segmentation.
65. Write short notes on supervised and unsupervised learning with examples.

Unit 5: Data Visualization Using Power BI

2-Mark Questions

66. What is data visualization?
67. What is Power BI?
68. List any four features of Power BI.
69. Define data transformation.
70. What is a data model in Power BI?
71. What are the types of visualizations available in Power BI?

72. What are dashboards and reports in Power BI?
73. How can data be imported from different sources in Power BI?
74. What is Power Query Editor used for?
75. What is meant by publishing a Power BI report?

5-Mark Questions

76. Explain the importance of visualization and analytic tools in data analytics.
77. Discuss the architecture and components of Power BI.
78. Describe the process of getting data from different sources in Power BI.
79. Explain data transformation and data modeling in Power BI.
80. List and explain various types of visualizations in Power BI with examples.
81. Explain the steps involved in creating and publishing reports in Power BI.
82. Discuss how dashboards and analytical reports are created using Power BI.
83. Explain the use cases of Power BI dashboards in industry.
84. Discuss the role of Power BI in business decision making.
85. Compare Power BI with other visualization tools like Tableau or Excel.

Integrated and Analytical Questions

2-Mark Questions

86. Define data-driven decision making.
87. What is the relationship between data analytics and business intelligence?
88. What is big data analytics?
89. What are KPIs (Key Performance Indicators)?
90. What is data cleaning?

5-Mark Questions

91. Explain the life cycle of data analytics.
92. Discuss the challenges in data collection and preprocessing.
93. Explain the importance of statistics in data analytics.
94. Discuss the role of visualization in data-driven decision making.

95. Describe a real-life use case where data analytics improved organizational efficiency.
96. Explain how Python and Power BI complement each other in analytics.
97. Write short notes on:
    a) Data Preprocessing
    b) Data Transformation
    c) Data Modeling
98. Describe the process of building an end-to-end data analytics project.
99. Explain the ethical issues and privacy concerns in data analytics.
100. Summarize the complete workflow of data analytics from data collection to visualization.

Numerical Problems:

1. In a classification problem, a **logistic regression** model gives the following probabilities for five observations:
   [0.8, 0.3, 0.6, 0.9, 0.2].
   If the cutoff is **0.5**, determine which observations are classified as positive and compute **accuracy**, given the actual labels are [1, 0, 1, 1, 0].
2. Perform one iteration of **K-Means clustering** for data points (2,4), (3,5), (10,8), (11,9) with initial centroids (2,4) and (10,8).
   Assign points to clusters and compute new centroids.
3. In **K-Means clustering**, the coordinates of two cluster centers are (2,3) and (8,7). Find the **Euclidean distance** between them.
4. Consider the following **confusion matrix**:

	Predicted Positive	**Predicted Negative**
Actual Positive	50	10
Actual Negative	5	35

   Calculate **Accuracy** and **Precision.**

5. A company predicts sales (Y) based on advertising expenditure (X) using the regression line:
   Y = 5 + 0.8X.
   Find the **predicted sales** when advertising expenditure is ₹12,000.
   Also, find the **residual** if actual sales are ₹14,000.

6. Find the mean, median, and mode of the data set: **5, 7, 3, 7, 9, 7, 2, 10.**

7. Calculate the **range** and **standard deviation** for the following values: **12, 15, 10, 18, 20.**
8. company records the monthly sales (in ₹ thousands) as:
   25, 28, 30, 32, 27, 26, 29, 31, 34, 30.
   Calculate the **mean, variance, and standard deviation.**

9. The marks of 10 students are given below:
   25, 30, 40, 35, 20, 45, 50, 55, 60, 40.
   Find the **coefficient of variation** and interpret the result.
10. A bag contains **5 red, 3 blue, and 2 green balls.** One ball is drawn at random.
    Find the **probability of drawing a red ball.**
11. A discrete random variable X has the following probability distribution:

X	0	1	2	3	4
P(X)=	.1	.2	.3	.3	.1

Find **E(X)** (mean), **Variance(X)**, and **Standard Deviation(X)**.