Subject Name: DATA ANALYTICS

Unit-Wise Question Bank

Q.No	QUESTION	MARKS	SECTION	UNIT
1.	Define Data Analytics. Why Data Analytics is Crucial for Success in the Modern Era?	8	Section-I	
2.	Explain the importance of data analytics in decision-making across industries.	8	Section-I	1
3.	Explain the four primary types of data analysis and their key purposes?	8	Section-I	1
4.	Design a data analytics workflow for a company aiming to improve customer satisfaction. Highlight the steps and tools involved.	8	Section-I	1
5.	Implement Python's built-in data structures with examples program.	8	Section-I	1
6.	Identify the key functionalities of Jupyter Notebook that make it an effective tool for data analytics, and justify its use over other IDEs.	8	Section-I	1
7.	Illustrate Python Libraries for Data Visualization with suitable examples a) Matplotlib b) Seaborn	8	Section-I	1
8.	Assume a dataset to calculate Central tendency. What are the limitations of each measure in representing the data?	8	Section-I	1
9.	Assume a dataset to calculate Dispersion. What are the limitations of each measure in representing the data?	8	Section-I	1
10.	Explain the key features of Pandas Series and DataFrame with suitable examples.	8	Section-I	1
11.	Define and explain Probability and describe the types of Probability.	8	Section-II	2
12.	Explain the Bayesian Rule and Bayes' Theorem applications in various fields.	8	Section-II	2
13.	Describe and interpret the following terms: a) Conditional Probability b) Relative frequency method	8	Section-II	2

<table>
<tr><td>14.</td><td>Why is sampling important in research? Discuss how the choice between Simple Random and Cluster Sampling can impact the results.</td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>15.</td><td>Describe and explain the Chi-Square Test and demonstrate how it works.</td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>16.</td><td>Describe and explain the types of Student t-Tests, and illustrate with an example</td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>17.</td><td>A company wants to analyze the relationship between employees' education level and their job satisfaction. The data is presented in a contingency table that categorizes employees based on their education levels (High School, Bachelor's, and Master's) and their job satisfaction levels (Satisfied, Neutral, and Dissatisfied). Using the Chi-Square Test for Independence, determine whether there is a statistically significant relationship between education level and job satisfaction.
<table><tr><th>Education Level</th><th>Satisfied</th><th>Neutral</th><th>Dissatisfied</th></tr><tr><td>High School</td><td>50</td><td>30</td><td>20</td></tr><tr><td>Bachelor's</td><td>40</td><td>25</td><td>35</td></tr><tr><td>Master's</td><td>30</td><td>20</td><td>50</td></tr></table></td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>18.</td><td>A company wants to evaluate the effectiveness of its training program on employee performance. To assess this, the performance scores of 8 employees were recorded both before and after the training. Using the Student's t-test, determine whether the training program has led to a statistically significant improvement in employee performance.
Given Performance Scores:
Before Training: [65, 70, 68, 72, 75, 78, 74, 71]
After Training: [78, 82, 80, 85, 87, 90, 86, 83]
Perform a paired t-test and interpret the results.</td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>19.</td><td>Explain the Central Limit Theorem (CLT) and its significance in statistics. How does CLT help in constructing confidence intervals?</td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>20.</td><td>What is Sample Space and an Event in probability theory? Explain with suitable examples.</td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>21.</td><td>What is a Random Variable? Differentiate between Discrete and Continuous Random Variables with examples.</td><td>8</td><td>Section-II</td><td>2</td></tr>
<tr><td>22.</td><td>Define the following terms related to hypothesis testing:
a) Null hypothesis (H_0)
b) Alternative hypothesis (H_1)</td><td>8</td><td>Section-III</td><td>3</td></tr>
</table>

23.	Explain why hypothesis testing is important in statistical decision-making. Provide an example of its real-world application.	8	Section-III	3
24.	What is a Type I Error, and what are the impacts of a False Positive in statistical hypothesis testing?	8	Section-III	3
25.	What is a Type II Error, and what are the impacts of a False Negative in statistical hypothesis testing?	8	Section-III	3
26.	How to Reduce False Positives and False Negatives?	8	Section-III	3
27.	How would you use hypothesis testing to determine if a new drug is more effective than an existing one?	8	Section-III	3
28.	A company claims that their battery lasts an average of 50 hours. A researcher tests a sample of 30 batteries and finds an average life of 48 hours with a standard deviation of 5 hours. Formulate the null and alternative hypotheses for this study.	8	Section-III	3
29.	Identify and explain the different approaches to hypothesis testing.	8	Section-III	3
30.	Explain Student's t-test with suitable example	8	Section-III	3
31.	a) What is Analysis of Variance (ANOVA)? When to Use ANOVA? (4M) b) Provide some Important Terms Related to ANOVA. (4M)	8	Section-I	4
32.	a) What is one-way ANOVA? 2 M b) How is one-way ANOVA used? And what are some limitations to consider? 6 M	8	Section-I	4
33.	Provide mathematical operation, Performs a One-Way ANOVA (Analysis of Variance) test to determine if there is a significant difference in test scores among three different teaching methods (Method A, Method B, and Method C) Method A: 85, 88, 90, 92, 87 Method B: 78, 80, 82, 85, 81 Method C: 88, 85, 84, 87, 90	8	Section-I	4
34.	Write a python program to implement the following scenario, Performs a One-Way ANOVA (Analysis of Variance) test to determine if there is a significant difference in test scores among three different teaching methods (Method A, Method B, and Method C) Method A: 85, 88, 90, 92, 87 Method B: 78, 80, 82, 85, 81 Method C: 88, 85, 84, 87, 90	8	Section-I	4

35.	What is Two-Way ANOVA? And How Two-Way ANOVA Works?	8	Section-I	4
36.	Explain Post Hoc Test ANOVA and Provide Application of Post Hoc Tests in ANOVA	8	Section-I	4
37.	Explain Simple Linear Regression with suitable example.	8	Section-I	4
38.	Explain Multiple Linear Regression with suitable example.	8	Section-I	4
39.	Develop a Python program utilizing Multiple Linear Regression to predict home prices in Monroe Township, NJ (USA). Given the attributes of square footage area, number of bedrooms, and age of the home, the program should predict the prices for the following properties: a. 3000 sqft area, 3 bedrooms, 40 years old b. 2500 sqft area, 4 bedrooms, 5 years old	8	Section-I	4
40.	Explain Maximum Likelihood Estimation.	8	Section-I	4
41.	a) Explain Confusion Matrix? (4M) b) Why Do We Need a Confusion Matrix? (4M)	8	Section-V	5
42.	What are the performance of diagnostic tests in ROC	8	Section-V	5
43.	Explain AUC-ROC Curve in Machine Learning	8	Section-V	5
44.	a) How AUC-ROC curve can be used for the Multi-class Model? (4M) b) Applications of AUC-ROC Curve (4M)	8	Section-V	5
45.	Explain Types of Clustering Methods	8	Section-V	5
46.	a) Explain Hierarchical Clustering (4M) b) Explain Partition Clustering (4M)	8	Section-V	5
47.	a) Explain Density-based Clustering (4M) b) Explain Distribution-Based Clustering (4 M)	8	Section-V	5
48.	Explain CART algorithm	8	Section-V	5
49.	What are the Applications of Clustering?	8	Section-V	5
50.	Explain different machine learning algorithms with applications.	8	Section-V	5