Mathematics and Statistics:  
  
1. What is the Central Limit Theorem, and why is it important in statistics?  
2. Explain the difference between population and sample.  
3. What is probability and how is it calculated?  
4. What are the measures of central tendency, and when would you use each one?  
5. Define variance and standard deviation.  
6. What is the significance of hypothesis testing in data science?  
7. Explain the p-value and its significance in hypothesis testing.  
8. What is a normal distribution, and why is it important in statistics?  
9. Describe the differences between a Z-score and a T-score.  
10. What is correlation, and how is it measured?  
11. What is the difference between covariance and correlation?  
12. What is the law of large numbers?  
  
Machine Learning:  
  
13. What is machine learning, and how is it different from traditional programming?  
14. Explain the bias-variance trade-off.  
15. What are the different types of machine learning algorithms?  
16. What is overfitting, and how can you prevent it?  
17. Describe the k-fold cross-validation technique.  
18. What is regularization, and why is it important in machine learning?  
19. Explain the concept of feature engineering.  
20. What is gradient descent, and how does it work in machine learning?  
21. What is a decision tree, and how does it work?  
22. What are ensemble methods in machine learning, and provide examples.  
23. Explain the difference between supervised and unsupervised learning.  
24. What is deep learning, and how does it differ from traditional neural networks?  
25. What is a convolutional neural network (CNN), and where is it commonly used?  
26. What is a recurrent neural network (RNN), and where is it commonly used?  
27. What is the vanishing gradient problem in deep learning?  
28. Describe the concept of transfer learning in deep learning.  
  
Data Preprocessing:  
  
29. What is data preprocessing, and why is it important in data science?  
30. Explain missing data imputation techniques.  
31. What is one-hot encoding, and when is it used?  
32. How do you handle categorical data in machine learning?  
33. Describe the process of data normalization and standardization.  
34. What is feature scaling, and why is it necessary?  
35. What is outlier detection, and how can you identify outliers in a dataset?  
  
Data Exploration:  
  
36. What is exploratory data analysis (EDA), and why is it important?  
37. Explain the concept of data distribution.  
38. What are box plots, and how are they used in EDA?  
39. What is a histogram, and what insights can you gain from it?  
40. Describe the concept of data skewness.  
41. What are scatter plots, and how are they useful in data analysis?  
42. What is a correlation matrix, and how is it used in EDA?  
43. How do you handle imbalanced datasets in machine learning?  
  
Model Evaluation:  
  
44. What are the common metrics used for evaluating classification models?  
45. Explain precision, recall, and F1-score.  
46. What is ROC curve analysis, and what does it measure?  
47. How do you choose the appropriate evaluation metric for a regression problem?  
48. Describe the concept of confusion matrix.  
49. What is cross-entropy loss, and how is it used in classification problems?  
50. Explain the concept of AUC-ROC.  
Python and Programming:  
  
51. Describe the differences between Python 2 and Python 3.  
52. What is the Global Interpreter Lock (GIL) in Python, and how does it affect multi-threading?  
53. Explain the use of decorators in Python.  
54. What are list comprehensions, and how do they work?  
55. Describe the purpose of virtual environments in Python.  
56. How can you handle exceptions in Python?  
57. What is a lambda function, and where is it typically used?  
58. Explain the difference between shallow and deep copy in Python.  
59. What is the purpose of the map() and filter() functions in Python?  
60. Describe the difference between append() and extend() methods for lists.  
  
SQL and Database Knowledge:  
  
61. What is SQL, and how is it used in data science?  
62. Explain the difference between SQL's INNER JOIN and LEFT JOIN.  
63. What is a primary key and a foreign key in a relational database?  
64. How do you write a SQL query to retrieve data from a database table?  
65. What is the purpose of the GROUP BY clause in SQL?  
66. Explain the concept of indexing in databases.  
67. What are NoSQL databases, and how are they different from SQL databases?  
  
Big Data and Distributed Computing:  
  
68. What is Hadoop, and how does it handle big data?  
69. Explain the MapReduce programming model.  
70. What is Apache Spark, and why is it popular in big data processing?  
71. Describe the concept of distributed computing.  
72. What are the advantages and disadvantages of distributed databases?  
  
Data Visualization:  
  
73. Why is data visualization important in data science?  
74. Describe the types of charts and graphs commonly used in data visualization.  
75. What is the purpose of a heatmap in data visualization?  
76. Explain the concept of storytelling through data visualization.  
77. How can you create interactive data visualizations in Python?  
  
Natural Language Processing (NLP):  
  
78. What is natural language processing, and what are its applications?  
79. Describe the steps involved in text preprocessing for NLP.  
80. What is tokenization, and why is it necessary in NLP?  
81. Explain the concept of stop words in NLP.  
82. What are n-grams, and how are they used in text analysis?  
83. What is sentiment analysis, and how is it performed using NLP techniques?  
84. What is named entity recognition (NER) in NLP?  
  
Time Series Analysis:  
  
85. What is a time series, and give examples of time series data.  
86. Explain the components of a time series (trend, seasonality, and noise).  
87. What is autocorrelation in time series analysis?  
88. How do you perform time series forecasting?  
89. What are ARIMA models, and how are they used in time series forecasting?  
90. Describe exponential smoothing methods in time series analysis.  
  
Dimensionality Reduction:  
  
91. Why is dimensionality reduction important in machine learning?  
92. Explain the concept of Principal Component Analysis (PCA).  
93. What is t-SNE, and how is it used for dimensionality reduction?  
94. Describe the curse of dimensionality.  
95. When would you use feature selection versus feature extraction for dimensionality reduction?  
  
Ethical and Business Considerations:  
  
96. What are the ethical considerations in data science?  
97. How can bias be introduced into machine learning models, and how can it be mitigated?  
98. Explain the concept of data privacy and GDPR compliance.  
99. How can data science provide value to a business?  
100. Describe a real-world project where data science had a significant impact.