

1. What is SQL?
2. Who maintains the database, queries, website, frontend, backend, and other services in an application?
3. What is a database?
4. What is DBMS?
5. What is RDBMS?
6. What is the difference between DBMS and RDBMS?
7. What are the examples of DBMS and RDBMS?
8. What is the difference between SQL and MySQL?
9. What are Tables and Fields?
10. What is the basic syntax of SQL?
11. Write any query to show how the code is written in SQL?
12. What are the Constraints in SQL?
13. What is the SELECT command?
14. Explain the use of the FROM clause?
15. What and how is \* used in SQL?
16. What is the use of the WHERE clause in SQL?
17. What is the use of the ORDER BY clause in SQL?
18. What is the use of GROUP BY clause in SQL?
19. Write a SQL query to get the average marks scored by each class.
20. Can we use the WHERE command with an aggregate function?
21. What is the use of the HAVING clause in SQL?
22. What is the use of the LIMIT clause in SQL?
23. Use the LIKE clause to filter all the products that start with a and end with e
24. What are wildcards in SQL?
25. One more important interview question on this line is when you are asked to pick the product name iphone but the iphone can be written as
26. How to create a table with a single field as primary key?
27. How to create a table with multiple fields as primary key?
28. What is a Primary Key?
29. What is a UNIQUE constraint?
30. What is a Foreign Key?
31. What is a Join?
32. What are the different types of joins?
33. What is an inner join? Give its syntax.
34. What is Left Join? Give its syntax.
35. What is Right Join? Give its syntax.
36. What is Full Join? Give its syntax
37. What is self-join?
38. In an EmployeeDetails table, we have 3 columns:-

**Emp\_id, Emp\_Name, and Mgr\_id**

The Manager id is nothing but the employee id of some employee. Example

1, A, 2

2, B, 3

3, C, 4

B is the manager of A and C is the manager of B.

Output required

Employee Name	Manager Name
A	B
B	C
C	D

Write a code in SQL to get the desired output.

39. Now in the above example, we have used Inner join as a part of the self join, the above will query will work fine for all the employees who have a Manager, but the Managing Director, CEO's, etc. won't necessarily have a Manager. What part of the query will you change to make sure all the employees of the company is present in the output.

40. What is a Cross-Join?

41. What is an Index?

42. We have two columns(Revenue and Cost Price) in a table like below, get me the Profit column

Revenue	Cost Price	Profit
100	Null	100
200	20	180
300	50	250
Null	50	-50

43. How do the SQL commands flow at the back end?

44. Write a SQL query to get the second highest query using sub query.

45. Write a SQL query to find all the student names Nitin in a table  
select name from student where lower(name) like '%nitin%'
46. Write a query to get all the student with name length 10, starting with K and ending with z.
47. Write a SQL query to get the second highest query using Ranking
48. Can you use HAVING command without any aggregate function in SQL?
49. You have data on people have applied for a lottery ticket. The data consists of their name and ticket number. You have to choose winners by selecting the people present in the alternate rows (the first winner starting from row number 3). Write a query to make things easy to select the winners.
50. We have the following values

10000  
10000  
20000  
30000  
30000  
30000

What would be the result of row number, rank, and dense rank ?

51. Find all the students who either are male or live in Mumbai ( have Mumbai as a part of their address).
52. Suppose there are two columns in employee table i.e. emp id and email  
get all the unique domains like gmail.com, yahoo.com, outlook.com, etc.
53. Can you join two table without any common column?
- 54 to 58. Give the output for the following

```
SELECT 'NITIN'+1
SELECT 'NITIN'+1'
SELECT(SELECT 'NITIN')
SELECT '1'+1
SELECT 1+'1'
```

59. select case when null=null then 'Amit' else 'Rahul' end from dual.

What will be the output of the above query?

60. What is the difference between COUNT(\*) and COUNT(ColName)?
61. Help me create a table with all the employee Names and Manager Names.

Employee_Name	Employee_Id	Manager_Id
A	1	2
B	2	3
C	3	2
D	4	3

62. Write a query for collecting the names of children who pursuing their graduation in

their residential city.

NAME	AGE	ADDRESS	COLLEGE
*****	*****	*****	*****

(Hint- Look for the common city name in the columns of ADDRESS and COLLEGE)

63. What is indexing in SQL?

64. List the different types of relationships in SQL.

65. Demonstrate how to write a query to show details of an HR whose name starts with M.

66. What is OLTP?

67. What are the differences between OLTP and OLAP?

68. What is the usage of the NVL() function?

69. Explain character-manipulation functions? Explains its different types in SQL.

70. Get the number of duplicate names and their frequency

Table – Employee

Name

Nitin

Amit

Gaurav

Nitin

Amit

Output

Nitin – 2

Amit – 2

Gaurav – 1

71. Write the SQL query to get the third maximum salary of an employee from a table named employees.

72. Define records and fields in a table

73. What is the use of FETCH command?

74. What are UNION, MINUS and INTERSECT commands?

75. How to fetch alternate records(even rows) from a table?

76. You always have a big data i.e. millions of rows in your tables, how would you partition it for optimum performance?

77. Suppose in class, you have  $3n$  boys and  $2n$  girls with their names tabulated along with their weight and gender. Write a SQL query to separate students alphabetically who are over-weight (55kg for girls, 75kg for boys)
78. What is indexing in SQL? Explain with proper example.
79. Arrange the employees with respect to their Joining date, most experienced employee coming on the top followed by others.
80. What is Cursor? How to use a Cursor?
81. List the different types of relationships in SQL.
82. How to create a temp table in SQL Server?
83. Write an SQL Query find number of employees whose DOB is between 01/07/1965 to 31/12/1975. Collect the data separately for different gender.  
Use column name as sex, DOB table name Employees.
84. How can you create an empty table from an existing table? Write the steps and explain the working.
85. What is the use of IFNULL and ISNULL in SQL?
86. You have got some data in the Table 1 and Table 2  
Write a SQL query to create a Table 3 that contains the following columns- Id, First\_Name, Last\_Name, Salary
87. NoSQL vs SQL
88. What is the difference between NOW() and CURRENT\_DATE()?
89. Extract the information about all department managers who were hired between the 1st of January 2020 and the 1st of January 2021.
90. If a table contains duplicate rows, does a query result display the duplicate values by default? How can you eliminate duplicate rows from a query result?
91. How to create a stored procedure using SQL Server?
92. Which join is used to join a table with itself
93. What is the difference between CHAR and VARCHAR2 datatype in SQL?
94. List the total numbers of products of each brand(Take the name of the table as well as the column name by yourself)
95. Does the virtual table created occupy space for the operation to joining with itself?
96. What are Constraints?
97. What is the difference between DELETE and TRUNCATE statements?
98. What is CROSS JOIN UNNEST in Presto ?
99. What is LATERAL VIEW explode in Hive?

100-103. Suppose there are two tables, X and Y, X has just one column A and Y has B.

These are the two tables

X(A)	Y(B)
1	1
2	2
3	3
	4
	5

**Questions – How many rows will be populated if you do**

**-X left join Y**

**-X inner join Y**

**-X cross join Y**

**-X right join Y**

**104.**

**In the table below, some duplicate records might be present by mistake. Sort out a way to locate them and find a way to delete them.**

empid	empname	managerid	deptid	Salary
1	Emp1	0	1	6000
2	Emp2	0	5	6000
3	Emp3	1	1	2000
13	Emp13	2	5	2000
11	Emp11	2	1	2000
9	Emp9	1	5	3000
8	Emp8	3	1	3500
7	Emp7	2	5	NULL
3	Emp3	1	1	2000

**105. What is the difference between clustered and non-clustered index in SQL?**

**106. Pivot a table in SQL without using pivot function**

**Suppose there are two columns**

**Age Name**

**25 Nitin**

**30 Amit**

**27 Rishab**

**29 Ankush**

**Convert into**

**Name Nitin. Amit. Rishab. Ankush**

**Age. 25. 30. 27. 29**

**107. What is the function of OFFSET command? You might know the definition but create a scenario where you have to make use of it. Also, can the same process be done with any other method?**

**108. Get all employee detail from EmployeeDetail table whose "FirstName" not start with any single character between 'a-p'**

**109. What do you understand by query optimization?**

**110. Important conditions for joining two tables on a key?**

**111.What are Entities and Relationships?**

**112. Assume the name of table and columns**

**There are two tables with a common column. Which one will take more processing time**

**-Outer Join**

**– Full Outer Join**

**– Cartesian Join**

**113. What is NTILE with syntax? With example and use case**

**114. What is Normalization and what are the advantages of it?**

**115. Explain different types of Normalization.**

**116. Find the Nth largest salary from employee table.**

**117. How the triggers will execute if two or more triggers?**

**118. Get all the employee detail from EmployeeDetail table whose "FirstName" starts with A and contain 5 letters**

**119. Could you tell output or result of following SQL statements?**

(Hint- In some cases, there may be an error. So, try to locate them and answer accordingly)

**select 5**

**select '5'**

**select count ('5')**

**select count (5)**

**select count (\*)**

**120. How do you maintain database integrity where deletions from one table will automatically cause deletions in another table?**

**121. Find the number of people who are from Delhi and have arrived in Patna in the last 7 days**

**122. How do you handle ties in dense ranking?**

**123. Does the data stored in the stored procedure increase access time or execution time?**

**124. How many Aggregate functions are available in SQL?**

**125. When do we use Coalesce() function?**

**126. You have a table with the following 3 columns**

customer_id, order_id, order_date		
123.	987.	2/9/2020
123	1234	4/9/2020
456	1211	1/1/2019
456	2344	8/8/2020

**We want to know the number of customer who have their first two orders in the last 180 days. In the above example 123 will qualify but 456 won't qualify**



127. What is a RANK() function?

128. How do we use the DISTINCT statement? What is its use?

129. What is the ACID property in a database?

130. What are the different types of a subquery?

131. Write an SQL query to fetch the current date-time from the system.

132 to 137

There are two tables:

Patients Table

Patient ID	Patient Name	Sex	Age	Address
01	Sheela	F	23	Flat no 201, Vasavi Heights, Yakutap
02	Rehan	M	21	Building no 2, Yelahanka
03	Anay	M	56	H No 1, Panipat
04	Mahira	F	42	House no 12, Gandhinagar
05	Nishant	M	12	Sunflower Heights, Thane

PatientsCheckup Table

Patient ID	BP	Weight	Consultation Fees
01	121/80	67	300
02	142/76	78	400
03	151/75	55	300
04	160/81	61	550
05	143/67	78	700

**132. Write a query to fetch top N records using the TOP/LIMIT, ordered by ConsultationFees.**

**133. Write a SQL query to create a table where the structure is copied from other table.**

**134. Write a query to fetch even and odd rows from a table.**

**135. Write an SQL query to fetch duplicate records from Patients, without considering the primary key.**

**136. Write a query to fetch the number of patients whose weight is greater than 68.**

**137. Write a query to retrieve the list of patients from the same state.**

**138. Write a SQL query to fetch consultation fees – wise count and sort them in descending order.**

**139. Write a SQL query to retrieve patient details from the Patients table who have a weight in the PatientsCheckup table.**

**140. Write a SQL query to retrieve the last 2 records from the Patients table.**

**141. Write a SQL query to find all the patients who joined in the year 2022.**

**142. Write a SQL query to fetch 50% records from the PatientsCheckup table.**

**143. Write a query to update the patient names by removing the leading and trailing spaces.**

**144. Write a query to add email validation to your database.**

**145. Write a query to find all patient names whose name:**

**Begin with A**

**Ends with S and contains 3 alphabets**

**Staying in the state Telangana**

**146. Write a SQL query to fetch details of all patients excluding patients with name “Sheela” and “Anay”.**

**147. Write a query to fetch the total count of occurrences of a particular character – ‘x’ in the PatientName.**

**148. Write a query to retrieve the first three characters of PatientName from the Patients table.**

**149. Write a query to combine Address and state into a new column – NewAddress.**

**150. Write a query to fetch PatientIDs which are present in:**

- a. Both tables**
- b. One of the table.**

**151. Write a query to find the number of patients whose RegDate is between 01/04/2021 to 31/12/2022 and are grouped according to state.**

**152. Write a query to fetch all records from the Patients table; ordered by PatientName in ascending order, State in descending order.**

**153. How can you create an empty table from an existing table? Write the steps and explain the working.**

**154. For what purpose we are using view? How is it created to get the job done?**

**155. Query the list of CITY names from STATION that does not start with vowels and do not end with vowels. Your result cannot contain duplicates.**

**156. State the differences between views and tables.**

**157. What is the difference between BETWEEN and IN operators in SQL?**

**158. What are the different types of a subquery?**

**159. Name the operator which is used in the query for pattern matching?**

**160. Write a SQL query to find the 10th highest employee salary from an Employee table**

**161. Given a table dbo.users where the column user\_id is a unique numeric identifier, how can you efficiently select the first 100 odd user\_id values from the table?**

**162. What will be the output of the below query, given an Employee table having 10 records?**

```
BEGIN TRAN  
TRUNCATE TABLE Employees  
ROLLBACK
```

**SELECT \* FROM Employees**

**163. How do you get the last id without the max function?**

**164. What is the difference between IN and EXISTS?**

**165. Which is better CTE or temp table?**

**166. Why are CTEs slower than temp tables?**

**167. Write an SQL query to fetch the EmpIds that are present in EmployeeDetails but not in EmployeeSalary.**

**168. Write an SQL query to fetch the position of a given character(s) in a field.**

**169. Write a query to fetch only the first name(string before space) from the FullName column of the EmployeeDetails table.**

**170. Write an SQL query to find the count of the total occurrences of a particular character – 'n' in the FullName field.**

**171. Write an SQL query to fetch employee names having a salary greater than or equal to 5000 and less than or equal to 10000.**

**172. Write an SQL query to fetch duplicate records from EmployeeDetails (without considering the primary key – EmpId).**

**173. What is the difference between COALESCE() & ISNULL()?**

**174. Write SQL Query to find duplicate rows in a database? and then write SQL query to delete them?**

**175. There is a table which contains two columns Student and Marks, you need to find all the students, whose marks are greater than average marks i.e. list of above-average students.**

**176. How to Show the Max marks and min marks together from student table?**

**177. Give the order of SQL SELECT.**

**178. What do we need to check in Database Testing?**

**179. How many types of Privileges are available in SQL?**

**180. What are the types of transaction levels in SQL SERVER?**

**181. What is log shipping?**

**182. What is a T-SQL?**

**183. What is query optimization?**

**184. What is the purpose of Query Optimization?**

**185. What are the metrics for analyzing query performance for SQL Query Optimization?**

186. What is the difference between cross join and natural join?
187. What is SQL Injection?
188. What is BLOB and TEXT in MySQL?
189. Explain database white box testing and black box testing.
190. Explain the difference between OLTP and OLAP.
191. What is a stored procedure? Give an example.
192. What are the types of views in SQL?
193. Mention different types of replication in SQL Server?
194. What is a data warehouse?
195. What do you mean by data integrity?
196. What is a No SQL database? What are some examples of NO SQL database?
  
197. What is a Relational Database? What are some examples of Relational databases?
198. Which type of data is stored in NO SQL and Relational Databases?
199. NO SQL is a vertically scalable or horizontally scalable database? What about Relational Database? How does this scaling make database life easier?
200. When should we use NO SQL and where should we use Relational Database?
201. Disadvantage of Temporary keyword in Presto?
202. What is row number in SQL?
203. How is row number different from rank and dense rank?
204. What is the syntax for calculating row number in SQL?
205. Can you give an example of when you would use row number?
206. How can you use row number to group data?
207. How do you reset the row number to 1 for each group in SQL?
208. Can you explain the performance implications of using row numbers?
209. How would you handle NULL values when using row number?
210. How can you use row number to perform pagination in SQL?
211. How can you use row number to delete duplicate rows in SQL?
212. What is dense rank in SQL?
213. How is dense rank different from rank and row number?
214. What is the syntax for calculating dense rank in SQL?
215. Can you give an example of when you would use dense rank?
216. How do you handle ties in dense ranking?
217. Can you explain the difference between dense rank and percent rank?
218. How can you use dense rank to group data?
219. How do you calculate dense rank for a subset of data in SQL?
220. Can you explain the performance implications of using dense rank?
221. How would you handle NULL values when using dense rank?
222. What is the LIKE command in SQL?
223. How does the LIKE command work?
224. What are the wildcards used in the LIKE command?
225. Can you give an example of using the LIKE command?
226. How can you use the LIKE command with multiple patterns?

227. How can you use the LIKE command to search for data that contains special characters?
228. Can you use the LIKE command with a NOT operator to exclude specific patterns?
229. How can you use the LIKE command to search for data that starts with or ends with a specific pattern?
230. How can you improve the performance of a LIKE command query?
231. Can you use the LIKE command with parameters in a stored procedure?
232. What is a regular expression and why would you use it in SQL?
233. How do you use the REGEXP\_LIKE function in SQL? Can you provide an example of when you might use this function?
234. How do you use the REGEXP\_REPLACE function in SQL? Can you provide an example of when you might use this function?
235. How do you use character classes in regular expressions? Can you provide an example of when you might use a character class in SQL?
236. How do you use quantifiers in regular expressions? Can you provide an example of when you might use a quantifier in SQL?
237. How do you use anchors in regular expressions? Can you provide an example of when you might use an anchor in SQL?
238. How do you use alternation in regular expressions? Can you provide an example of when you might use alternation in SQL?
239. How do you escape special characters in regular expressions? Can you provide an example of when you might need to escape a special character in SQL?
240. How do you use regular expressions to validate input in SQL? Can you provide an example of how you might validate an email address or a phone number using regular expressions in SQL?
241. How can you use regular expressions to extract data from a string in SQL? Can you provide an example of how you might extract specific values from a string using regular expressions in SQL?
242. What is JSON?
243. How does Presto handle JSON data?
244. What is the syntax for querying JSON data in Presto?
245. How does Presto handle nested JSON data?
246. Can you join JSON data with other tables in Presto?
247. How does Presto handle JSON arrays?
248. How can you flatten nested JSON data in Presto?
249. How can you convert JSON data to a relational format in Presto?
250. How can you use the json\_table() function in Presto?
251. How can you use the json\_arrayagg() function in Presto?

**252. What is Python? What is the difference between Python and Java?**

**253. Sum of two numbers in Python**

**254. How to take input from user?**

**255. Take input from user and print the sum. This will test two concepts, first the concept of type casting and secondly how to print the sum and not concatenate the numbers.**

**256. What are the data types in Python?**

**257. What is number data type? What is type() function in Python?**

**258. What is Boolean data type in Python?**

**259. What is string data type in Python?**

**260. How to concatenate variables?**

**261. What do you mean when you say “Strings are immutable”?**

**262. What is a list in Python?**

**263. Are lists mutable ?**

**264. What are some important functions used with list?**

**265. What is a dictionary in Python?**

**266. Is dictionary zero indexed? Can we pull something like Team[0] from the above example?**

**267. What is a tuple?**

**268. What are the salient features of tuples?**

**269. What are some salient features of set?**

**270. A few data types are ordered and other are not ordered. What does that mean?**

**271. What is the function range() ?**

**272. What is type conversion?**

**273. What are collections in Python?**

**274. Write a Python program to get the factorial of any number.**

**275. Get the square of a number if the number is odd and cube if even**

**276. Sum of square of first n natural number**

**277. Check if a number is Armstrong number in Python. An Armstrong number is a number that equals the sum of cube of all the digits of the number. Ex. 153 =  $1^3 + 5^3 + 3^3$**

**278. Write a program to get the area of a rectangle (to know if the candidate understands the concept of function)**

**279. Check if a number is a prime number in Python.**

**280. Write a program to get the Prime number in a list of numbers with starting and end point**

- 281. Print the nth number in the fibonacci series.
- 282. Check if a given number is a perfect square or not
- 284. Finding sum of elements of array
- 285. Largest element of an array
- 286. Rotate an array
- 287. Split the array at a particular point
- 288. Create an array using array library
- 289. What is append(), pop() and remove() function for array?
  
- 290. How can you reverse an array?
- 291. Interchange first and last element in a list
- 292. Use another method or way to interchange the first and the last elements of a list in Python. This was asked in Flipkart interview in 2021
- 293. In a list, Swap two positions in a list (Meesho interview question)
- 294. Get the length of a list and check if a particular element is present in the list?
- 295. Convert a list in a set and use in function
- 296. Use a loop to find if an element is present in a set or not?
- 297. Reverse a list
- 298. sum of elements in a list
- 299. Another method to print the sum of all the elements in the list
- 300. Multiple all the numbers in a list
- 301. Count occurrence of an element in a list
- 302. Cumulative sum in a list
- 303. Break a list in N parts using the yield clause
- 304. Reverse a word
- 305. Check if a string is palindrome. Palindrome example - Nitin or 12321
- 306. check if a string is present in another string
- 307. How to get word frequency in a particular sentence using collections?
- 308. What is the one big advantage of set that you can use to remove duplicate alphabets from a string?
- 309. What is the use of the following command  
  
"".join(str)
- 310. Using the above two questions remove all the duplicate characters from a string
- 311. Split a sentence in all the words
- 312. Find words with length greater than 3
- 313. rotate a string at a particular
- 314. Print words of even length
  
- x = "The Data Monk is a website"
- 315.  
Print all the keys of a dictionary
- 316. Print all the values of a dictionary



- 317. Sum of all the values of a dictionary
- 318. Remove a key from dictionary
- 319. Merge two dictionaries
- 320. Flatten a dictionary
- 321. How to define a tuple and get the size of tuple variable
- 322. Create a tuple with the 4 elements, then create a list of tuple with cube of each element
- 323. Adding tuple to a list
- 324. Extract digits from a tuple.
- 325. Removeing list of length 2 from the tuple
- 326. Flatten tuple of a list
- 327. What is split() function in Python ?
- 328. What is reduce() function in Python?
- 329. What is eval() ?
- 330. What is enumerate() ?
- 331. What is round() in Python?
- 332. What is max() in Python ?
- 333. What is min() in Python?
- 334. What is a map() function in Python ?
- 335. What is getattr() in Python?
- 336. What is append() in Python?
- 337. What is range() in Python?
- 338. What is the use of slice() function in Python?
- 339. What is the format() command in Python?
- 340. What is strip() method in Python ?
- 341. What is the use of abs() function in Python?
- 342. What is the use of upper() and lower() function in Python?
- 343. What is the join() function in Python?
- 344. What is replace() function in Python?
- 345. What is sorted() function in Python?
- 346. Write a Python program to print set of duplicates in a list.
- 347. Write a Python program to extract all the digits from a text
- 348. What are the most used package for visualization in Python?
- 349. How to create a line chart?
- 350. Plot a sin graph using line plot
- 351. Graph 2 – Cos graph using line plot
- 352. Class vs Number of Students chart with proper labels and plot title

You want to change the color of the line?

353.

Multi Line Chart

## **Graph 5 – Adding multiple lines to a graph**

### **354. Bar Chart**

**“A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.”**

**355. Use random values between 1 and 100 to create the same graph.**

**356. Bar chart with random values**

**357. A 100% stacked bar chart**

**358. A simple histogram - create a list using random variables and plot it in 4 bins**

**359. A histogram made with random variables**

**360. Parallel histogram**

**361. A horizontal histogram**

**362. A line histogram**

**363. A variable width histogram**

**364. Area Chart**

**365. A basic area chart**

**366. Box and Whisker Plot**

**367. Graph Question**

**368. Following is the code with some fancy colors to help you understand each term individually.**

**369. Box Whisker Chart**

**370. Scatter plot**

**371. A scatter plot**

**372. Pie Chart**

**373. Stacked Bar graph**

**374. What are the type of information that we can display using plot?**

**375. What are the charts that show the distribution in your data set?**

**376. What are the charts that show a comparison metrics?**

**377. What are the types of relationship charts?**

**378. What are comparison charts?**

**379. How do you handle missing values in a data set? What are some common techniques for imputing missing data?**

**380. How do you handle categorical data in a data set, and what are some common techniques for encoding or transforming it?**

**381. What are some common machine learning algorithms that you have experience with, and how do you use them in Python?**

**382. How do you assess the performance of a machine learning model, and what are some common metrics used to do so?**

**383. What is cross-validation, and how is it used in machine learning?**

**384. What is regularization, and how is it used to prevent overfitting in machine learning?**

**385. How do you handle imbalanced data sets in machine learning, and what are some common techniques for addressing this issue?**

**386. How do you use Python for data visualization, and what are some common libraries or tools that you have used?**

**387. How do you handle time series data in Python, and what are some common techniques for analyzing or forecasting time series data?**

**388. What are some common Python libraries used for data visualization, and how do they differ in terms of functionality and ease of use?**

**389. How do you create a line chart or scatter plot in Python using a given data set?**

**390. How do you customize the appearance of a plot in Python, such as changing the color, title, axis labels, or legend?**

**391. What is a heatmap in data visualization, and how do you create one in Python?**

**392. What is a histogram in data visualization, and how do you create one in Python?**

**393. How do you handle missing data or outliers in a data set when creating a visualization, and how do they affect the final result?**

**394. What are some common techniques for visualizing high-dimensional data sets in Python, such as using dimensionality reduction or clustering algorithms?**

**395. How do you create interactive visualizations in Python, such as using tools like Bokeh or Plotly?**

**396. What are some common techniques for visualizing geographic data in Python, such as using choropleth maps or scatter plots with latitudes and longitudes?**

**397. How do you create animated visualizations in Python, such as using the Matplotlib animation module or GIFs?**

### **Python Tricky Interview Questions**

**398. What is the difference between a list and a tuple in Python, and when would you use each?**

**399. Can you explain the difference between `is` and `==` in Python?**

**400. What is the difference between a shallow copy and a deep copy in Python, and when would you use each?**

**401. Can you explain what a closure is in Python and provide an example?**

**402. How does Python's global interpreter lock (GIL) impact multi-threaded programming in Python?**

**403. What is the difference between a generator and a list comprehension in Python, and when would you use each?**

**404. How does Python handle memory management, and what are some common memory management issues that can arise?**

**405. Can you explain the difference between a module and a package in Python, and how would you import and use each?**

**406. What is the difference between a class method and a static method in Python, and when would you use each?**

**407. Can you explain the difference between the range and xrange functions in Python 2, and why was xrange removed in Python 3?**

**408. What is the difference between "global" and "nonlocal" in Python?**

**409. What is the difference between "deepcopy" and "shallow copy" in Python?**

**410. What is a decorator in Python?**

**411. What is a lambda function in Python?**

**412. What is the difference between a module and a package in Python?**

**413. What will be the output of the code below in Python 2?**

**414. Write a Python program to calculate mode without using inbuilt function Input**

**415. How to webscrapping in Python?**

**416. What is the difference between "is" and "=="? In Python?**

**417. How is a variable allocated memory in Python?**

**418. Define a class named car with 2 attributes, "color" and "speed". Then create an instance and return speed.**

**419. What is the difference between instance, static and class methods in python?**

**420. What is the difference between "func" and "func()"?**

**421. Does Python call by value or call by reference?**

**422. Given two strings, string1 and string2, determine if there exists a one to one character mapping between each character of string1 to string2.**

**423. Write a program to print the following pattern:**

```
1
12
123
1234
12345
```

**424. Write a program to print the following pattern:**

A  
AB  
ABC  
ABCD  
ABCDE

**425. Write a program to print the following pattern:**

1  
22  
333  
4444  
55555

**426. Write a program to print the following pattern:**

E  
DE  
CDE  
BCDE  
ABCDE

**427. Write a program to print the following pattern:**

1  
2 3  
4 5 6  
7 8 9 10

**428. Write a program to print the following pattern:**

A B C D E F  
A B C D E  
A B C D  
A B C  
A B  
A

**429. Write a program to print the following pattern:**

1  
0 1  
1 0 1  
0 1 0 1  
1 0 1 0 1

**430. Write a program to check if a number is prime or not.**

**431. Write a program to find all prime numbers in a given range.**

**432. Write a program to check if a number is an Armstrong number or not.**

**433. Write a program to check if a number is a perfect number or not.**

**434. Write a program to find the factorial of a number.**

**435. Write a program to check if a number is a palindrome or not.**

**436. Write a program to find the sum of digits of a number.**

- 437. Write a program to check if a number is an automorphic number or not.
- 438. Write a program to check if a number is a Harshad number or not.
- 439. Write a program to check if a number is a Smith number or not.
- 440. What is the difference between `print()` and `return` in Python?
- 441. Can you give an example of using f-strings in Python?
- 442. How would you handle errors in Python code when printing output?
  
- 443. What is the purpose of `sys.stdout` and how can it be used?
- 444. Can you explain the difference between `sys.stdout.write()` and `print()`?
- 445. How would you write a Python function that writes output to a file instead of to the console?
  
- 446. How would you use the logging module in Python to output messages to a file?
- 447. How can you redirect the output of a Python script to a file instead of to the console?
  
- 448. Can you explain the difference between using `print()` and logging for outputting information in a Python application?
- 449. How can you use the `pprint` module to print formatted output in Python?

## **Pandas**

- 450. What is Pandas ?**
- 451. What is python pandas used for ?**
- 452. Write Steps to install Pandas on Windows.**
- 453. What are the key features of pandas library ?**
  
- 454. What is pandas dataframe ?**
- 455. How to Import Pandas Library and also check the version of Library.**
- 456. Mention the different types of Data Structures in Pandas?**
- 457. How to read the different – different format files using pandas??**
- 458. Define Series in Pandas?**
  
- 459. How to create a Series from a numpy array ?**
  
- 460.. Create a series using disctionary**
  
- 461. What is the describe() method in pandas ?**
- 462. What is a Data Frame?**
- 463. Define the different ways a DataFrame can be created in pandas?**
- 464. Create a dataframe form a List**
- 465. Create a DataFrame from dict of ndarrays:**
  
- 466. Describe how you will get the names of columns of a DataFrame in Pandas?**
- 467. How to Covert Json(Javascript object notation) data into Dataframe ?**
- 468 .Read a file using read\_csv()**
- 469. head() and tail()**
  
- 470.xyz.info()**
- 471. You can also take a look at the shape and size of the dataset. We have 100 rows and 6 columns in the original dataset**
- 472. Choose n number of random sample from the dataset**
- 473. Get all the standard mathematical analysis of each column of data set**
- 474. Find number of distinct values – This is an important function as it will directly tell you how many categorical variables are there in your dataset**
- 475. How to find if there is any variable/column with missing values in it?**
- 476. isnull()**
- 477. Find the number of null values in each column, this set of function tells you if you can ignore a column or not**
- 478. Get the name of all the columns**
- 479. Get the nsmallest or nlargest values from a column**
- 480. Now comes loc and iloc – There are a few interviewers who tries to check your basics with loc and iloc**
- 481. Slice the date – It means cutting the dataset vertically or horizontally**

- 482. Group by in Pandas – Very useful pandas function
- 483. Sort the complete data frame according to one column
- 484. Query in data frame
- 485. Get unique values from a column
- 486. If you want to know how many space columns are taking into your computer then use memory\_usage
- 487. How do you select a column in a DataFrame?
- 488. How do you select multiple columns in a DataFrame?
- 489. How do you filter rows based on a condition in a DataFrame?
- 490. How do you merge two DataFrames in Pandas?
- 491. How do you handle missing values in a DataFrame?
- 492. How do you group data in a DataFrame?
- 493. How do you sort a DataFrame by one or more columns?
- 494. How do you apply a function to a DataFrame?
- 495. How do you rename columns in a DataFrame?
- 496. How do you create a new column in a DataFrame?
- 497. How do you remove a column from a DataFrame?
- 498. How do you export data to a CSV file using Pandas?
- 499. Suppose you have a DataFrame named sales\_data that contains columns Year, Month, and Sales. Write a code snippet to calculate the total sales for each year and output the results in a new DataFrame with columns Year and Total Sales.
- 500. Suppose you have a DataFrame named customer\_data that contains columns CustomerID, Name, Age, and City. Write a code snippet to filter the data to only include customers who are 18 years old or older and who live in either "New York" or "San Francisco".
- 501. How do you select a subset of rows and columns from a DataFrame based on certain conditions?
- 502. How do you group a DataFrame by a certain column and perform an aggregation function on each group?
- 503. How do you merge two DataFrames based on a common column?
- 504. How do you create a new column in a DataFrame that is the result of a calculation involving other columns?
- 505. How do you rename a column in a DataFrame?
- 506. How do you sort a DataFrame by one or more columns?
- 507. How do you check if a DataFrame has any missing values?
- 508. How do you fill in missing values in a DataFrame with a certain value or method?
- 509. How do you write a pandas DataFrame to a CSV file?
- 510. What are some ways to handle missing data in pandas?
- 511. How can you handle duplicates in a pandas DataFrame?



512. How can you merge two pandas DataFrames together?

513. What are some ways to improve the performance of pandas operations?

514. How can you use pandas to group data by a certain column and perform aggregate functions on the groups?

515. How can you pivot a pandas DataFrame?

516. How can you use pandas to create a time series plot?

517. How can you use pandas to handle categorical data?

518. What are some common methods for data manipulation in pandas?

519. How can you use pandas to perform statistical analysis on a dataset?

520. How can you use pandas to handle time series data?

521. What are some techniques for reshaping data in pandas?

513. What are some ways to improve the performance of pandas operations?

514. How can you use pandas to group data by a certain column and perform aggregate functions on the groups?

515. How can you pivot a pandas DataFrame?

516. How can you use pandas to create a time series plot?

517. How can you use pandas to handle categorical data?

518. What are some common methods for data manipulation in pandas?

519. How can you use pandas to perform statistical analysis on a dataset?

520. How can you use pandas to handle time series data?

521. What are some techniques for reshaping data in pandas?

522. How can you use pandas to handle text data?

523. What are some ways to handle outliers and anomalies in a pandas DataFrame?

524. How can you use pandas to handle large datasets that don't fit in memory?

525. How can you use pandas to handle imbalanced datasets?

526. How can you use pandas to handle multi-index DataFrames?

527. What are some techniques for feature engineering in pandas?

528. How can you use pandas to handle data from multiple sources or files?

529. What are some advanced visualization techniques in pandas?

530. How can you create a scatter plot in pandas?

531. How can you create a bar chart in pandas?

532. How can you create a line plot in pandas?

533. How can you create a histogram in pandas?

534. How can you create a box plot in pandas?

535. How can you create a heatmap in pandas?

536. How can you customize the appearance of a plot in pandas?

537. How can you create multiple plots on a single figure in pandas?

538. How can you convert a series of strings to a series of numbers in pandas, where some of the strings may contain non-numeric characters or missing values?

539. How can you identify and handle multicollinearity in a pandas DataFrame?

540. How can you select rows from a pandas DataFrame that satisfy a condition based on the values in multiple columns?

541. How can you compute the rolling standard deviation of a time series in pandas, where the rolling window size varies over time?

**542. How can you handle imbalanced data in a pandas DataFrame for a classification problem?**

**543. How can you handle data with a mix of categorical and continuous variables in a pandas DataFrame for a regression problem?**

**544. How can you perform anomaly detection in a pandas DataFrame using statistical methods?**

**545. How can you perform sentiment analysis on text data in a pandas DataFrame?**

**546. How can you use pandas to perform feature selection and feature extraction for a machine learning model?**

**547. How can you use pandas to perform hyperparameter tuning for a machine learning model?**

**548. How can you optimize the memory usage of a pandas DataFrame?**

**549. How can you handle missing values in time series data in a way that preserves the time series structure?**

**550. How can you handle imbalanced data in a multi-class classification problem in a way that maintains class balance during training?**

**551. How can you use pandas to perform time series forecasting using machine learning models?**

**552. How can you use pandas to handle very large datasets that don't fit into memory on a single machine?**

**553. How can you use pandas to perform natural language processing on text data in a DataFrame?**

**554. How can you use pandas to perform unsupervised learning on a DataFrame with mixed data types?**

**555. How can you use pandas to perform deep learning on a DataFrame with image data?**

**556. How can you use pandas to perform distributed computing across a cluster of machines?**

**557. How can you use pandas to handle streaming data in real-time?**

## **Numpy**

- 558. What is NumPy, and what are its main features?**
- 559. How do you create a NumPy array from a Python list?**
- 560. How do you perform element-wise arithmetic operations between two NumPy arrays?**
- 561. How do you calculate the dot product of two NumPy arrays?**
- 562. What is broadcasting in NumPy, and how does it work?**
- 563. What are some common statistical functions available in NumPy, and how do you use them?**
- 564. Define the mean function in numpy and give a simple example**
- 565. Define the median function in numpy and give a simple example**
- 567. Define the var function in numpy and give a simple example**
  
- 568. Define the min function in numpy and give a simple example**
  
- 569. Define the max function in numpy and give a simple example**
  
  
- 570. How do you select elements from a NumPy array based on a conditional expression?**
- 571. How do you reshape a NumPy array?**
- 572. Convert a multidimensional array to 1D array**
- 573. How do you perform matrix operations in NumPy, such as matrix multiplication and inversion?**
- 574. What is Matrix Inversion?**
- 575. What is Determinant calculation?**
- 576. How do you save and load NumPy arrays from disk?**
- 577. How do you concatenate two or more NumPy arrays horizontally?**
- 578. How do you concatenate two or more NumPy arrays vertically?**
  
- 579. How do you concatenate two or more NumPy arrays arbitrarily?**
- 580. How do you create a masked array in NumPy, and what is its purpose?**
- 581. What is a shallow copy in a NumPy array?**
  
- 582. What is a deep copy in NumPy?**
- 583. How do you generate random numbers in NumPy, and what are some common distributions you can sample from?**
  
- 584. How do you sort a NumPy array in ascending or descending order?**
- 585. How do you perform element-wise logical operations between two NumPy arrays?**

**586. How do you compute the Fourier transform of a signal using NumPy?**

**587. How to use NumPy with SciPy?**

**588. How to use NumPy with matplotlib?**

#### **Coding Questions**

**589. Write a NumPy code snippet to create an array of zeros with shape (3, 4).**

**590. Write a NumPy code snippet to create an array of ones with shape (2, 5).**

**591. Write a NumPy code snippet to create an array of evenly spaced values between 0 and 10 with a step size of 2.**

**592. Write a NumPy code snippet to create a random array with shape (2, 3) and values between 0 and 1.**

**593. Write a NumPy code snippet to calculate the sum of all elements in a two-dimensional array.**

**594. Write a NumPy code snippet to calculate the mean of all elements in a one-dimensional array.**

**595. Write a NumPy code snippet to calculate the standard deviation of all elements in a one-dimensional array.**

**596. Write a NumPy code snippet to calculate the dot product of two one-dimensional arrays.**

**597. Write a NumPy code snippet to reshape a one-dimensional array into a two-dimensional array with 3 rows and 2 columns.**

**598. Write a NumPy code snippet to find the index of the maximum value in a one-dimensional array.**

#### **Data Cleaning using Numpy**

**599. How do you remove missing or null values from a NumPy array?**

**600. How can you identify and remove outliers in a NumPy array?**

**601. What are some common techniques for normalizing data in a NumPy array?**

**602. How do you sort a NumPy array, and what are some of the options for customizing the sort?**

**603. What are some functions in NumPy that are commonly used for data cleaning?**

**604. How can you handle duplicate values in a NumPy array?**

**605. What is the difference between slicing and indexing in NumPy, and how are they used for data cleaning?**

**606. How can you concatenate two NumPy arrays, and what are some of the considerations when doing so?**

**607. How can you reshape a NumPy array, and what are some common use cases for doing so in data cleaning?**

**608. How do you create a NumPy array with specific dimensions and data types?**

609. How do you access specific elements in a NumPy array?
610. How can you perform basic arithmetic operations on NumPy arrays?
611. How can you create a mask for a NumPy array based on specific conditions?
612. How can you apply a function to specific elements in a NumPy array?
613. How can you combine two or more NumPy arrays to create a new array?
614. How can you save a NumPy array to a file, and how can you load a saved array back into Python?
615. How can you calculate dot products and matrix multiplication using NumPy?
616. How can you apply linear algebra operations (e.g., inverse, determinant) to a NumPy array?
617. What happens if you try to reshape a NumPy array with the wrong number of elements?
618. What is the difference between NumPy's broadcasting rules and Python's broadcasting rules?
619. How can you modify the data type of a NumPy array?
620. How can you perform element-wise comparison between two NumPy arrays with different shapes?
621. What is the difference between NumPy's views and copies, and how can you determine which one you have?
622. How can you create a custom data type in NumPy, and what are some use cases for doing so?
623. What are some best practices for optimizing performance when working with large NumPy arrays?
624. How can you handle missing or invalid data in a NumPy array?
625. What are some ways to create a NumPy array, and when would you use each one?
626. How can you access and modify individual elements of a NumPy array?
627. What is the difference between slicing and indexing in NumPy, and how can you use them to extract subsets of an array?
628. How can you perform mathematical operations on a NumPy array, and what are some common functions for doing so?
629. What is broadcasting in NumPy, and how can you use it to perform element-wise operations on arrays with different shapes?
630. How can you reshape a NumPy array, and what are some common use cases for doing so?
631. What is a masked array in NumPy, and how can you use it to handle missing data?

**632. How can you stack and concatenate NumPy arrays, and what are some use cases for doing so?**

**633. What are some best practices for optimizing performance when working with large NumPy arrays?**

**Numpy Advance Interview Questions**

**634. What are some of the performance benefits of using NumPy over pure Python when working with numerical data?**

**635. What is a view in NumPy, and how does it differ from a copy?**

**636. How can you use NumPy to perform linear algebra operations, such as matrix multiplication and solving systems of equations?**

**637. What are some of the built-in functions in NumPy for generating random numbers, and how can you use them to simulate data?**

**638. What is the difference between a structured array and a record array in NumPy, and what are some use cases for each?**

**639. How can you use NumPy to perform Fourier transforms, and what are some applications of Fourier analysis in signal processing and image processing?**

**640. How can you use NumPy to perform interpolation, and what are some use cases for doing so?**

**641. What are some of the limitations of NumPy, and how can you work around them?**

**642. What are some best practices for organizing and structuring code when working with NumPy, especially when dealing with large, complex arrays?**

**643. Write a NumPy code snippet to create an array of 100 random integers between 0 and 10.**

**644. Write a NumPy code snippet to calculate the element-wise sum of two arrays of the same shape.**

**645. Write a NumPy code snippet to compute the inner product of two one-dimensional arrays**

**646. Write a NumPy code snippet to find the indices of the maximum and minimum values in a two-dimensional array.**

**647. Write a NumPy code snippet to reshape a one-dimensional array into a two-dimensional array with 4 rows and 5 columns**

**648. Write a NumPy code snippet to calculate the correlation coefficient between two arrays of the same length.**

**649. Write a NumPy code snippet to create a diagonal matrix with the elements 1, 2, and 3 on the diagonal.**

**650. Write a NumPy code snippet to sort a one-dimensional array in ascending order.**

**651. Write a NumPy code snippet to calculate the element-wise product of two arrays of the same shape.**

**652. Write a NumPy code snippet to create a mask that selects all elements of an array that are greater than 5.**

### **Case Study**

**653. What are the metrics to evaluate a website? A search engine?**

**654. Punjabi By Nature is going through a rough phase and is wasting 30-40 Kgs of biryani on some days of the week.**

**Management is thinking to find a method to sell the left over biryani or reduce the amount of biryani prepared.**

**You are the manager and you are supposed to come up with a methodology for the same**  
**655. TVF recently went to app/website only mode. How do you think it makes money to survive and pay the actors ?**

**656. How to increase the marketing of online games?**

**657. Given a set of webpages and changes on the website, how will you test the new website feature to determine if the change works positively?**

**658. What is root cause analysis? How to identify a cause vs. a correlation? Give examples.**

**659. Can you choose a strategy for increasing the number of songs listened by the user on an online application? How will you decide the types of playlists to suggest him?**

**660. How would you measure the success of acquiring new users through a 30-day free trial at Netflix?**

**661. How would you measure the success of Facebook Groups?**

**662. Describe how you would build a model to predict Uber ETAs after a rider requests a ride.**

**Common machine learning case study problems like this are designed to explain how you would build a model. Many times this can be scoped down to specific parts of the model building process. Examining the example above, we could break it up into:**

**How would you evaluate the predictions of an Uber ETA model?**

**OR**

**What features would you use to predict the Uber ETA for ride requests?**

**663. You're a Data Scientist / Business Analyst working for a new eCommerce company called A&B Co. (similar to Amazon) and you've been asked to prepare a presentation for the Vice President of Sales and the Vice President of Operations that summarizes sales and operations thus far. The summary should include (at a minimum) a summary of current state the business, current customer satisfaction, and a proposal of 2-3 areas where the company can improve. Here are some facts:**

**It's currently September 2018 (e.g., you can ignore all data after September 2018)**

**The company's inception was January 2017 (so you can ignore all data before January 2017)**

**Company is US-based, but launched in Brazil (which is why some information is in Portuguese)**

**You can assume all orders are delivered (so ignore the order state field)**

**Your presentation should not have more than 10 slides of content, and the presentation itself should only take ~15 minutes.**

**664. Profit of a company selling mobile back cover is declining. List out all the possible reasons**

**665. How would you construct a feed to show relevant content for a site that involves user interactions with items?**

**666. How would you design the people you may know feature on LinkedIn or Facebook?**

**667. How would you predict who someone may want to send a Snapchat or Gmail to?**

**668. How would you suggest to a franchise where to open a new store?**

**669. In a search engine, given partial data on what the user has typed, how would you predict the user's eventual search query?**

**670. Given a database of all previous alumni donations to your university, how would you predict which recent alumni are most likely to donate?**

**671. You're Uber and you want to design a heat map to recommend to drivers where to wait for a passenger. How would you approach this?**

**672. How will you change the UI of the game to increase the number of people buying coins from the shop section (Clash Of Clans)?**

**673. Taj Group of Hotels is planning to start a new branch, What are the parameters it should consider to find the appropriate place?**

**674. List the e-commerce metrics**

**675. The case study was to increase the number of conversions of freemium customers to premium customers for a telecom company in India.**



**676. Flipkart has recently observed that people are falling out of it's buy funnel i.e. a list of pages following which a customer makes a payment.**

**The pages are – Listing page, Description Page, Delivery option page, Payment Page and Redirection page.**

**How would you identify the problem faced by the customers**

**677. For a particular e-commerce company, there has been a decline in the number and value of items stored in the cart. Tell all the possible reasons for this decline?**

**678. KFC wants to open its first branch in India, what data points should be considered before it entering in a new market?**

**679. For a particular e-commerce business the number of supplier and user both have grown by 5% week over week, but the revenue has declined by 5 %, what could be the reason?**

**680. What are the factors to consider if you work in the sales department of Samsung and you want to start a store in one of the most crowded malls of Bangalore?**

**Think through all the points like**

**-Foot fall**

**-Customers**

**-Supply and demand issue (connect with warehouse)**

**-Break even point of profit**

**-Exit points if things fail**

**680. Customer Obsession**

**681. Invent and Simplify**

**682. Are Right, A lot**

**683. Hire and Develop the Best**

**684. Think Big**

**685. Frugality**

**686. Earn Trust**

**687. Deliver Results**

**688. What are the KPIs for Nykaa?**

**689. What are the KPIs for Zomato?**

**690. What are the KPIs for Uber and Ola?**

**691. What are the KPIs of Amazon or e-commerce company?**

**692. What are the KPIs of Google?**

**693. What are the KPIs for Apple?**

**694. What are the KPIs of Tesla?**

**695. What are the KPIs for Walmart?**

**696. What are the KPIs for Microsoft?**

**697. What are the KPIs for ITC?**

**698. What are the KPIs of Pfizer?**

**699. How would you go about designing a recommendation system for an e-commerce website, and what data sources and algorithms would you consider?**

**700. Imagine you've been given a dataset on customer churn in a telecom company. What kind of analysis would you conduct to identify the key factors that contribute to customer churn, and what recommendations would you make to reduce it?**

**701. Suppose you're working on a project to improve customer engagement for a mobile app. What metrics would you use to measure engagement, and what techniques would you employ to increase it?**

**702. You're tasked with building a fraud detection system for a bank. What features and data sources would you consider, and what models and techniques would you use to identify fraudulent activity?**

**703. How would you go about developing a predictive maintenance system for a fleet of trucks, and what data sources and algorithms would you use to anticipate maintenance needs?**

**704. You've been asked to build a chatbot for a customer service team. What kind of data would you need to train the chatbot, and what technologies and techniques would you use to ensure it provides accurate and helpful responses?**

**705. You're working on a project to optimize the pricing strategy for a ridesharing company. What data sources would you use to inform your analysis, and what techniques would you employ to set optimal prices in real time?**

**706. You have been given a dataset containing millions of online transactions for an e-commerce platform. How would you analyze the data to identify customer behavior patterns and preferences, and what insights would you extract to optimize the platform's user experience and increase sales?**

**707. You are working on a project to build a recommendation engine for a streaming media platform. How would you design and implement the recommendation engine, and what algorithms and data sources would you use to provide personalized recommendations to users?**

**708. You are tasked with developing a predictive model for a healthcare provider to identify patients who are at risk of developing chronic conditions. What data sources and features would you use to train the model, and what algorithms and techniques would you employ to improve the accuracy of the predictions and inform targeted interventions?**

**709. You are working on a project to optimize a supply chain for a consumer goods company. What data sources and analytics tools would you use to identify bottlenecks and inefficiencies in the supply chain, and what recommendations would you make to improve performance and reduce costs?**

**710. You have been given a large dataset of sensor data from a fleet of industrial equipment. How would you analyze the data to identify patterns of performance and maintenance needs, and what techniques and models would you use to inform a predictive maintenance strategy?**

**711. You are tasked with developing a fraud detection system for a financial services company. What features and data sources would you use to train the model, and what algorithms and techniques would you employ to improve the accuracy of fraud identification and reduce false positives?**

**712. You are working on a project to optimize marketing spend for a consumer packaged goods company. What data sources and analytics tools would you use to inform decisions around channel allocation, audience targeting, and ad creative, and how would you measure the impact of the marketing campaigns on sales and customer engagement?**

**713. You have been given a dataset of social media activity for a large retailer. How would you analyze the data to identify trends and sentiment among customers, and what recommendations would you make to improve the retailer's social media presence and customer engagement?**

**714. You work for an e-commerce company that sells a variety of products online. The company has a website, mobile app, and social media accounts, and it uses various online marketing channels to drive traffic to its website and app. The company has been experiencing declining sales over the past few months and has brought you in to help identify the problem and recommend solutions.**

**Question:**

**How would you use digital analytics to help the e-commerce company identify the problem and recommend solutions to increase sales?**

**715. KPIs for Telecommunication companies**

**716. KPIs for Human Resource department**

**717. KPIs for Data Platform of any Technology company**

**718. KPIs for Finance Department**

**719. KPIs for Marketing Department**

**720. KPIs for Pharmaceutical client**

**721. KPIs for Insurance client**

**722. KPIs for Football Match**

**723. KPIs for a Chess Match**

**725. KPIs for Mobile Game**

**726. KPIs for Electric Vehicle**

**727. KPIs for stores**

**728. How OTT platform earns money?**

**729. KPI for Meesho**

**730. Number of Maggi sold in a day in India**

**731. How many t-shirts e-commerce companies selling in India per day?**

**732. What is the number of laptops sold in Bangalore on an average routine day?**

**733. What are the number of smart phones sold in India per year?**

**734. What is the total number of people who get a new job in India in a year?**

**735. How many red colored Swift cars are there in Delhi?**

**736. How many paan shops are there in India?**

**737. What is the maximum number of human beings that can survive on the earth?**

**738. How to calculate the number of taxis operating in Mumbai per day?**

**739. How to calculate the number of weddings in India per year?**

**740. How many KGs of paint are used in the USA annually?**

**741. How many street lights are there in India?**

**742. How many cups of tea were consumed in Mumbai last month?**

**743. How many iPhones are currently being used in India?**

**744. What is the size of the laptop market in the USA?**

**745. What number of tennis balls can fit inside a room?**

**746. How many cups of tea were consumed in Delhi in a month?**

**747. How much paint will be required for painting a 20 m x 20 m wall?**

**748. How many refrigerators are sold in India every year?**

**749. How many cups of coffee do Americans drink in New York City per month?**

## **Linear Regression**

**750. What is Linear Regression?**

**751. Why is linear regression important?**

**752. What are some common places where Linear Regression is used?**

**753. What is the Equation of Linear Regression?**

**754. What are Independent(X) and Dependent variables(Y)?**

**755. What is  $\beta_0$  OR Y Intercept?**

**756. What is  $\beta_1$  OR Slope coefficient?**

**757. What is  $\epsilon$  OR Error Term OR Residual?**

- 758. What is correlation? (V.V.I.)
- 759. What is Positive Correlation?
- 760. What is Negative Correlation?
- 761. What is a Simple linear regression?
- 762. What is a Multiple linear regression?
- 763. What are the advantages of Linear Regression?
- 764. What are the disadvantages of Linear Regression?
- 765. What is Overfitting?
- 766. How to deal with overfitting in Linear Regression?
- 767. What is Underfitting?
- 768. What is regularisation?
- 769. What is Feature Selection?
- 770. What are the type of feature selection?
- 771. What are the assumptions of Linear Regression?
- 772. When to drop an outlier and when not to drop an outlier?
- 773. What is the assumption of Linearity?
- 774. What is the assumption of multicollinearity?
- 775. What are the functions used to check multicollinearity in python?
- 776. What is corr() function?
- 777. What is vif?
- 778. How to deal with the problem of multicollinearity?
- 779. What is R square?
- 780. What is Adjusted R square?
- 781. Difference between R<sup>2</sup> and Adjusted R<sup>2</sup>?
- 782. What is RMSE?
- 783. Difference between Correlation and Regression?
- 784. Why do we split our data into training and testing?
- 785. Steps for performing Linear Regression in Python.
- 786. How does multicollinearity affect the linear regression?
- 787. Can you name a possible method of improving the accuracy of a linear regression model?
- 788. What are outliers? How do you detect and treat them?
- 789. How do you interpret a Q-Q plot in a linear regression model?
- 790. What is the importance of the F-test in a linear model?
- 791. What are the disadvantages of the linear regression model?
- 792. What is the curse of dimensionality? Can you give an example?
- 793. Compare Linear Regression and Decision Tree
- 794. Name a disadvantage of R-squared and explain how would you address it?
- 794. Does correlation imply causation? Why or why not?
- 795. Is linear regression suitable for time series analysis?

- 796. How do you determine if a linear regression model is a good fit for your data?
- 797. What are some common techniques for dealing with multicollinearity in multiple linear regression models?
- 798. How do you check for heteroscedasticity in a linear regression model, and what are some techniques for addressing it?
- 799. How do you evaluate the performance of a linear regression model, and what metrics do you use?
- 800. How do you handle outliers and influential points in a linear regression model?

## **Logistic Regression**

- 801. What is Classification?
- 802. Why is logistic regression called regression if it does the job of classification?
- 803. What is Logistic Regression?
- 804. What is the similarity between linear regression and logistic regression?
- 805. Explain the mechanism of Logistic Regression
- 806. What are the applications of Logistic Regression?
- 807. What are between Logistic Regression and Linear Regression?
- 808. What is a Sigmoid Function?
- 809. What is the difference between Sigmoid function and SoftMax function?
- 810. In a nutshell, Explain the advantages and disadvantages of Logistic Regression?
- 811. What are the assumptions of Logistic Regression?
- 812. Why does the response variable in the data should be binary in nature when using the logistic regression algorithm?
- 813. What is Hypothesis test?
- 814. What is Log Transformation?
- 815. What is the general workflow of Logistic Regression Algorithm?
- 816. What are the libraries required for implementing Logistic Regression?
- 817. How to import the dataset into the python Environment?
- 818. The dataset does not has headers, how to define the headers on the dataset?
- 819. How can you revert the original data frame in case any failure in your data analysis?
- 820. How to drop some of the variables in the dataset?
- 821. How to have look at the dataset?
- 822. How to check the null values in the dataset?
- 823. How to impute the missing values?
- 824. How to know the count of occurrences of the variables?
- 826. How to convert the categorical data into numerical data?
- 827. On what basis should we decide that outliers should be eliminated or not ??
- 828. How to check for outliers in the data?

- 829. Explain Label Encoder
- 830. Explain Manual Mapping
- 831. What are Dummy Variables?
- 832. Explain One Hot Label Encoding
- 833. Explain the steps of Label Encoding?
- 834. What is Scaling?
- 835. Explain the syntax of Scaling and which scaling technique we will be using in this algorithm?
- 836. How will we decide how to train the model and how to test the model on the data which is available to us?
- 837. What is the threshold for splitting the data?
- 838. What do you mean by feature splitting?
- 839. What do you mean by feature selection?
- 840. What is loc and iloc in python and what is the difference between them?
- 841. What is the code for building the Logistic Regression Model?
- 842. Can we create a custom function of Confusion Matrix so that we can picturized it beautifully?
- 843. How to check whether our model is performing well or not?
- 844. What is Accuracy of model?
- 845. What is Precision of model?
- 846. What is Recall Factor ?
- 847. Does Logistic Regression provide any feature in these condition where the recall factor is not satisfying?
- 848. What is the difference between SVM and Logistic Regression?
- 849. What is ROC?
- 850. What is AUC?
- 851. What is the code for plotting ROC curve?
- 852. How to calculate the AUC scores?
- 853. What is SGD Stochastic Gradient Descent Classifier?
  
- 854. What is the difference between (SGD)Stochastic Gradient Descent Classifier and Logistic Regression?

## **NLP**

- 855. What is NLP?
- 856. What are the uses of NLP?
- 857. What are the different algorithms in NLP?
- 858. What problems can NLP solve?
- 859. What is Regular Expression?
- 860. What are the packages in Python to help in Regular Expression

**861. What is the difference between match and search function?**

**862. Guess the output of the following**

```
import re  
re.split('\s','The Data Monk is cool')
```

**863. Work in finding the output of the following**

```
regx = r"\w+"  
strx = "This isn't my pen"  
re.findall(regx,strx)
```

**864. How to write a regular expression to match some specific set of characters in a string?**

**865. Write a regular expression to split a paragraph every time it finds an exclamation mark**

**867. Find the output of the following code?**

**868. What is tokenization?**

**869. What is NLTK?**

**870. What are the important nltk tokenizer?**

**871. What is the use of the function set() ?**

**872. Now get the unique words from the above paragraph**

**873. What is the use of .start() and .end() function?**

**874. What is the OR method?**

**875. What are the advance tokenization techniques?**

**876. How to write a regex to match spaces or commas?**

**877. How to include special characters in a regex?**

**878. What is the difference between (a-z) and [A-Z]?**

**879. Once again go through the difference between search() and match() function.**

**880. What is topic modeling?**

**881. What is bag-of-words?**

**882. How to counter the case sensitive nature of bag-of-words?**

**883. What is counter?**

**884. How to import Counter in Python?**

**885. Use the same paragraph used above and print the top 3 most common words**

**886. What is text preprocessing?**

**887. What are the commonly used methods of text preprocessing?**

**888. How to tokenize only words from a paragraph while ignoring the numbers and other special**

**Character?**

**889. What are stop words?**

**890. How to remove stop words from my text?**

**891. What is Lemmatization?**

**892. Give an example of Lemmatization in Python**

**893. How to lemmatize the texts in your paragraph?**



- 894. What is gensim?
- 895. What is a word vector?
- 896. What is LDA?
- 897. What is gensim corpus?
- 898. What is stemming?
- 899. Give an example of stemming in Python
- 900. What is tf-idf?
- 901. How to create a tf-idf model using gensim?
- 902. What is Named Entity Recognition?
- 903. What is POS?
- 904. What is the difference between lemmatization and stemming?
- 905. What is spacy package?
- 906. How to initiate the English module in spacy?
  
- 907. Why should one prefer spacy over nltk for named entity recognition?
- 908. What are the different packages which uses word vectors?
- 909. What if your text is in various different languages? Which package can help you in Named Entity Recognition for most of the largely spoken languages?
- 910. What is supervised learning?
- 911. How can you use Supervised Learning in NLP?
- 912. What is Naïve-Bayes model?
- 913. What is the flow of creating a Naïve Bayes model?
- 914. What is TF-IDF?
- 915. What is POS?
- 916. Take an example to take a sentence and break it into tokens i.e. each word
- 917.. Take the same sentence and get the POS tags
- 918. Take the following line and break it into tokens and tag POS using function
  
- 919. What is NER?
- 920. What are some of the common tags in POS.
- 921. Implement NER on the tokenized and POS tagged sentence used above.
- 922. What are n-grams?
- 923. Create a 3-gram of the sentence below  
"The Data Monk was started in Bangalore in 2018"
- 924. What is the right order for a text classification model components?
- 925. What is CountVectorizer?
- 926. How to create a dataset? What to write in it?
- 927. What all packages do I need to import for this project?
- 928. How to import a csv file in Python?
- 929. Let's view the top and bottom 5 lines of the file to make sure we are good to go with the Analysis

930. Now we will clean the dataset. Will start with removing numbers and punctuations. Write a regular expression for removing special characters and numbers review is the name of the data set and Review is the name of the column
931. Now we want to stem the words. Do you remember the definition of stemming?
932. What does the above snippet do?
933. Create the final dataset with only stemmed words.
934. How to use the CountVectorizer() function? Explain using an example
935. Now let's apply CountVectorizer on our dataset
936. How to separate the dependent variable?
937. Now we need to split the complete data set into train and test
938. Random forest is one of the best model to work on supervised learning. By the way, what is Random forest?
939. What is Random Forest?
940. Let's create our Random forest model here
941. Define n\_estimator
942. Define criterion. Why did you use entropy and not gini?
943. What is model.fit()?
944. Let's predict the output for the testing dataset
945. Now let's check the confusion matrix to see how many of our outputs were correct
946. Lastly, what is confusion matrix and how to know the accuracy of the model?
947. What is sub() method?
948. Convert all the text into lower case and split the words

## Decision Tree

949. What is a decision tree, and how does it work?
950. What are the advantages and disadvantages of using decision trees as a machine learning model?
951. What are the different types of decision trees, and how do they differ from each other?
952. What is entropy, and how is it used in decision tree algorithms?
953. What is the difference between the Gini impurity and entropy as measures of impurity in decision trees?
954. How are decision trees used in regression problems, and what is the difference between regression trees and classification trees?
955. What is overfitting, and how can it be addressed in decision tree models?

956. How are missing values handled in decision tree algorithms?
957. What is pruning, and how is it used to prevent overfitting in decision tree models?
958. What are some common algorithms used to build decision trees, and how do they differ from each other?
959. How can decision trees be used in combination with other machine learning models, such as random forests or gradient boosting machines?
960. How can the performance of a decision tree model be evaluated, and what metrics are commonly used?
961. What is the bias-variance tradeoff in machine learning, and how does it apply to decision trees?
962. How can decision trees be used for feature selection, and what advantages does this approach offer?
963. What are some limitations of decision trees, and when might they not be the best choice of model for a given problem?
964. List down some popular algorithms used for deriving Decision Trees along with their attribute selection measures.
965. Explain the CART Algorithm for Decision Trees.
966. List down the attribute selection measures used by the ID3 algorithm to construct a Decision Tree.
967. Briefly explain the properties of Gini Impurity.
968. Briefly explain the properties of Gini Impurity.
969. Measurement of the performance of decision tree algorithm.
970. What is F1 Score?
971. What is Recall?
972. What AUC?
973. What is a confusion matrix?
974. Can decision trees handle categorical variables? If so, how are they treated?
975. How does the depth of a decision tree affect its performance and complexity?
976. How can decision trees be used to perform feature selection, and what advantages does this approach offer over other feature selection methods?
977. What is the difference between a greedy algorithm and an exhaustive search algorithm for building decision trees?
978. Can decision trees be used for time-series forecasting or sequence prediction problems? If so, how are they adapted?
979. What is Pre-Pruning and Post-Pruning?
980. How to solve the problem of overfitting in Decision Tree?
981. What is Pre-Pruning and Post-Pruning?
982. What are the ways of evaluation of the Continuous Variable Decision Tree?

- 983. What is Accuracy Score?
- 984. What is classification report?
- 985. F1\_score is the harmonic mean of the precision and the recall.
- 986. What are the Advantages of Decision Tree?
- 987. What are the Disadvantages of Decision Tree?
- 988. Which is better Linear model or Tree-Based-Model?
- 989. What is hyperparameter tuning in Machine Learning?
- 990. What are the various hyperparameters used in Decision Tree for tuning?
- 991. Steps for performing Decision Tree in Python.

## **Random Forest**

- 992. What is a random forest?
- 993. What is the difference between a decision tree and a random forest?
- 994. How does a random forest work?
- 995. What are the advantages of using a random forest algorithm?
- 996. What are some of the parameters that can be tuned in a random forest algorithm?
- 997. What is bagging in the context of a random forest?
- 998. How does the random forest algorithm handle missing data?
- 999. What are some of the techniques used to evaluate the performance of a random forest model?
- 1000. What is the feature importance in a random forest model?
- 1001. What are the limitations of the random forest algorithm?
- 1002. Does Random Forest need Pruning? Why or why not?
- 1003.. Explain how the Random Forests give output for Classification, and Regression problems?
- 1004. How is a Random Forest related to Decision Trees?
- 1005. How would you find the optimal size of the Bootstrapped Dataset?
- 1006. What are Ensemble Methods?
- 1007. Explain the advantages of using Random Forest
- 1008. How does Random Forest handle missing values?
- 1009. How is it possible to perform Unsupervised Learning with Random Forest?
- 1010. How would you improve the performance of Random Forest?
- 1011. What are proximities in Random Forests?
- 1012. What does Random refer to in Random Forest?
- 1013. What is Entropy?
- 1014. Why Random Forest models are considered not interpretable?
- 1015. Why is the training efficiency of Random Forest better than Bagging?
- 1016. Implement Random Forest in Python
- 1017. What is n\_estimators in Random forest?
- 1018. What is random\_state in random forest?
- 1019. What are the hyperparameters in Random Forest?

**1020. What is max\_depth in Random Forest ?**

**1021. What is min\_samples\_split in Random Forest?**

**1022. What is min\_smamples\_leaf in Random Forest?**

**1023. What is max\_features in Random Forest?**

**1024. What is the criterion in Random Forest?**

**1025. What is bootstrap in Random Forest?**

**1026. Implement Random Forest in R.**

**1027. Working of Random Forest Classifier?**

**1028. Working of Random Forest Regresor**

**1029. Grid Search in Hyperparameter tuning?**

**1030. Random Search in hyperparameter tuning?**

**1031. What is the impact of correlated features on a Random Forest model, and how can this issue be addressed?**

**1032. Can a Random Forest model suffer from overfitting? If so, how can overfitting be avoided or mitigated?**

**1033. How can the importance of individual trees in a Random Forest model be measured and used to improve the overall model?**

**1034. How does the out-of-bag (OOB) error estimate work in Random Forest, and what are its limitations?**

**1035. What is the difference between a feature selection technique and a feature importance measure in the context of Random Forest?**

**1036. How can the performance of a Random Forest model be improved by reducing the correlation among the trees?**

**1037. How can imbalanced class distributions affect the performance of a Random Forest model, and what techniques can be used to address this issue?**

**1038. How can the computational efficiency of a Random Forest model be improved for large datasets or high-dimensional feature spaces?**

**1039. How can the interpretability of a Random Forest model be improved, and what techniques can be used to extract insights from the model?**

**1040. What are some of the limitations and drawbacks of Random Forest, and when might it not be the best choice of model for a given problem?**

**Kmean**

**1041.What is K-means?**

**1042. What is clustering?**

- 1043. How does k-means clustering works?
- 1044. What does k represents in k-means clustering?
- 1045. What is a cluster?
- 1046. What is a centroid?
- 1047. What does k-means term denotes?
- 1048. How to find the optimal value of k in k-means?
- 1049. What is Elbow method?
- 1050. What is the Within-Cluster Sum of Squared Error?
  
- 1051. Implement WCSS Error in Python
  
- 1052. What is the drawback of K-means clustering algorithm?
- 1053. What are the different ways to solve the problem of initialization sensitivity in k-means algorithm?
- 1054. What is K-means++ algorithm?
- 1055. What do you mean by intracluster distance?
- 1056. What do you mean by intercluster distance?
- 1057. Difference between K-means and K-means++ algorithm?
- 1058. Difference between Classification and Clustering?
- 1059. What are the advantages of k-means clustering algorithm?
- 1060. What are the Disadvantages of k-means clustering algorithm?
- 1061. Difference between KNN and K-means algorithm?
- 1062. What are the different types of distance metrics used in K-means algorithm?
- 1063. What are the different types of clustering?
- 1064. What are the applications of K-means Clustering Algorithm?
- 1065. Steps for performing K-means Clustering Algorithm in Python.
- 1066. What are the libraries to apply K-Mean cluster?
- 1067. How to import dataset as Data Frame?
- 1068. How to check the missing values in the dataset?
- 1069. Plot using scatter plot in Python
- 1070. How to save the result in the excel file?

1071. What is the KNN Algorithm?
1072. Why is KNN a non-parametric Algorithm?
1073. What is "K" in the KNN Algorithm?
1074. How does the KNN algorithm make the predictions on the unseen dataset?
1075. Is Feature Scaling required for the KNN Algorithm? Explain with proper justification.
1076. What is space and time complexity of the KNN Algorithm?
1077. Can the KNN algorithm be used for regression problem statements?
1078. Why is the KNN Algorithm known as Lazy Learner?
1079. Why is it recommended not to use the KNN Algorithm for large datasets?
1080. How to handle categorical variables in the KNN Algorithm?
1081. How to choose the optimal value of K in the KNN Algorithm?
1082. How can you relate KNN Algorithm to the Bias-Variance tradeoff?
1083. Which algorithm can be used for value imputation in both categorical and continuous categories of data?
1084. Explain the statement- "The KNN algorithm does more computation on test time rather than train time".
1085. What are the things which should be kept in our mind while choosing the value of k in the KNN Algorithm?
1086. What are the advantages of the KNN Algorithm?
1087. What are the disadvantages of the KNN Algorithm?
1088. Is it possible to use the KNN algorithm for Image processing?
1089. What are the real-life applications of KNN Algorithms?
1090. Explain the Difference between K-Means and KNN

**1091. Show an implementation of KNN Algorithm**

## **Power BI**

**1100. What are the parts of Microsoft self-service business intelligence solution?**

**1101. What is self-service business intelligence?**

**1102. What is Power BI?**

**1103. What is Power BI Desktop?**

**1104. What data sources can Power BI connect to?**

**1105. What are Building Blocks in Power BI?**

**1106. What are the different types of filters in Power BI Reports?**

**1107. What are content packs in Power BI?**

**1108. What is DAX?**

**1109. What are some of the DAX functions?**

**1110. How is the FILTER function used?**

**1111. What are the functions and limitations of DAX?**

**1112. What is SUMMARIZE() and SUMMARIZECOLUMNS() DAX?**

**1113. What are some benefits of using Variables in DAX ?**

**1114. How would you create trailing X month metrics via DAX against a non- standard calendar?**

**1115. What are the different BI add-in to Excel ?**

**1116. What is Power Pivot?**

**1117. What is Power Pivot Data Model?**

**1118. What is xVelocity in-memory analytics engine used in Power Pivot?**

**1119. What are some of differences in data modeling between Power BI Desktop and Power Pivot for Excel?**

**1120. Can we have more than one active relationship between two tables in data model of power pivot?**

**1121. What is Power Query?**

**1122. What are the data destinations for Power Queries?**

**1123. What is query folding in Power Query?**

**1124. What are some common Power Query/Editor Transforms?**

**1125. Can SQL and Power Query/Query Editor be used together?**

**1126. What are query parameters and Power BI templates?**

**1127. Which language is used in Power Query?**



**1128. Why do we need Power Query when Power Pivot can import data from mostly used sources?**

**1129. What is Power Map?**

**130. What are the primary requirement for a table to be used in Power Map?**

**1131. What are the data sources for Power Map?**

**1132. What is Power View?**

**1134. Can we refresh our Power BI reports once uploaded to cloud (Share point or Powebi.com)?**

**1135. What are the different types of refreshing data for our published reports?**

**1136. Is Power BI available on-premises?**

**1137. What is data management gateway and Power BI personal gateway?**

**1138. What is Power BI Q&A?**

**1139. What are some ways that Excel experience can be leveraged with Power BI?**

**1140. What is a calculated column in Power BI and why would you use them?**

**1141. Why might you have a table in the model without any relationships to other tables?**

**1142. What is the Power BI Publisher for Excel?**

**1143. What are the differences between a Power BI Dataset, a Report, and a Dashboard?**

**1144. What are the three Edit Interactions options of a visual tile in Power BI Desktop?**

**1145. What are some of the differences in report authoring capabilities between using a live or direct query connection such as to an Analysis Services model, relative to working with a data model local to the Power BI Desktop file?**

**1146. How does SSRS integrate with Power BI?**

**1147. What is the general business need for Power BI?**

**1148. Which data sources are often associated with Power BI?**

**1149. Can you please tell me what the distinctive channels in the Power BI reports.**

**1150. Can you please tell me what some of the key competitive differences are between Power BI and Power BI Pro?**

**1151. Do you know how much Power BI costs?**

**1152. Can you please tell me about the Power BI desktop application?**

**1153. Can you please tell me what DAX is?**

**1154. What are some of the core parts of Microsoft self service business intelligence solutions?**

**1155. Can you please tell me what the self-service business intelligence offering is?**

**1156. What are the building blocks when it comes to Power BI?**

**1157. What are the content packs when it comes to Power BI?**

**1158. Can you please tell me how you use the FILTER function?**

**1159. What are the various names of Power BI?**

**1160. Name some of the filters that you can utilize with the Power BI Reports.**

1161. Can you tell me how well you understand the Power BI designer?
1162. What are the main differences between the Power BI gateway and Data management gateways?
1163. Can you please tell me what the selection pane is inside of Power BI?
1164. What is your opinion of the dynamic filtering option within Power BI?
1165. What data sets are generally used for creating a visual dashboard within the streaming data tiles?
1166. What would you consider to be the normal amount of table capacity for gathering of importing data in Power BI?
1167. Can you tell me what some of the advantages are of utilizing the variables within DAX?
1168. Can you tell me what Power Pivot is?
1169. Tell me what some of the key contrasts in the display of data between Power BI Desktop and Power Pivot for Excel?
1170. Can you tell me what the Power Pivot Data Model is?
1171. When is the x-Velocity examination tool utilized for Power Pivot?
1172. Can you tell me if you would have access of one dynamic connection between two tables in the data model in the intensity rotate process?
1173. Can you please describe to me what Power Query is ?
1174. What are data goals when it comes to Power Queries?
1175. Can you tell me what question collapsing within Power Query is?
1176. Can you tell me if Power BI available for on-prem (or on premises)?
1177. Does Power BI support general mobile devices like Android and iOS?
1178. The Power BI Desktop's Software License Terms show that one may install and use one copy of the software on the premises. True or false?
1179. Can you tell me what the query folding tool is in Power Query?
1180. What are some of the most commonly used Power Query Data Transforms?
- 1181 . Can the SQL and Power Query Editor be used together?
1182. What are the main query parameters and Power BI Templates you have access to?
1183. What is the main language, which is used in Power Query?
1184. Can you tell me why you would need Power Query when Power Pivot can do the same?
1185. Can you tell me what the Power Map is within Power BI?

**1186. What are some of the main requirements for a table to be used within Power Map for Power BI?**

**1187. What are the data hotspots when it comes to Power Map?**

**1188. Is it possible to invigorate the Power BI reports once they are uploaded to the cloud?**

**1189. What are some of the unique data invigorations used for generating the distributed reports within Power BI?**

**1190. Explain the Power BI Designer.**

**1191. Is there any process for refreshing Power BI reports one uploaded to the cloud?**

**1192. What is the major difference between Power BI personal Gateway and Data Management Gateway?**

**1193. What is the use of split function?**

**1194. Name all the platforms for which the Power BI app is available.**

**1195. Differentiate between older and newer Power BI.**

**1196. Is it possible in the power pivot data model to have more than one active relationship between two tables?**

**1197. What is the purpose of the 'Get Data' icon in Power BI?**

**1198. What is Row-level Security?**

**1199. What are the general data shaping techniques?**

**1200. Which data sets can be used to create dashboards with streaming data tiles?**

**1201. What are the KPIs in Power BI?**

**1202. What could be the difference between Distinct() and Values() in DAX?**

**1203. State the advantages of the Direct query method.**

**1204. What is What if the parameter in power BI?**

**1205. What is the incremental refresh?**

**1206. What are the three main tabs in Reports development Window?**

**1206. How is data security implemented in Power BI ?**

**1207. What are many-to-many relationships and how can they be addressed in Power BI ?**

**1208. What is a time series?**

**1209. What is time series analysis?**

**1210. How is time series analysis performed?**

**1211. What is a trend in time series?**

**1212. What is a seasonality in time series?**

**1213. What is stationarity in time series?**

**1214. What is time-series forecasting, and how is it used in data analysis?**

**1215. Can you explain the differences between time-series forecasting and other types of data analysis techniques?**

**1216. What is autoregression, and how is it used in time-series forecasting?**

**1217. How can you evaluate the accuracy of a time-series forecasting model?**

**1218. Can you explain the differences between moving average and exponential smoothing methods for time-series forecasting?**

**1219. How would you use ARIMA models for time-series forecasting in Python?**

**1220. Can you explain how to use Prophet library for time-series forecasting in Python?**

**1221. How can you handle missing values in a time-series data set?**

**1222. Can you explain the differences between additive and multiplicative seasonality in time-series forecasting?**

**1223. How can you use machine learning algorithms like random forests and neural networks for time-series forecasting?**

**1224. Can you explain how to use cross-validation to evaluate time-series forecasting models?**

**1225. How would you deal with non-stationarity in time-series data?**

**1226. Can you explain how to use state space models for time-series forecasting in Python?**

**1227. How can you use time-series forecasting to predict demand for a product or service?**

**1228. How can you use time-series forecasting to identify trends and patterns in financial data?**

**1229. What is an ARIMA model, and how is it used for time-series forecasting?**

**1230. Can you explain the differences between AR, MA, and ARMA models, and when would you use each?**

**1231. How would you determine the order of an ARIMA model for a given time-series data set?**

**1232. Can you explain the differences between stationary and non-stationary time-series data, and how does it impact the use of ARIMA models?**

**1233. How can you use ACF and PACF plots to select the parameters for an ARIMA model?**

**1234. Can you explain the Box-Jenkins method for building ARIMA models?**

**1235. How would you handle seasonality in a time-series data set when using ARIMA models?**

**1236. Can you explain how to fit an ARIMA model in Python using the Statsmodels library?**

**1237. How would you evaluate the accuracy of an ARIMA model, and what metrics would you use?**

**1238. How can you use ARIMA models for anomaly detection in time-series data?**

**1239. Can you explain how to use SARIMA models to incorporate seasonal effects in ARIMA modeling?**

**1240. How would you use ARIMA models for forecasting multiple time-series data sets simultaneously?**

**1241. Can you explain the differences between ARIMA models and exponential smoothing methods for time-series forecasting?**

**1242. How would you use ARIMA models for forecasting in the presence of missing data in a time-series data set?**

**1243. Can you explain the limitations of ARIMA models and when it may not be appropriate to use them for time-series forecasting?**

**1245. What are the assumptions underlying an ARIMA model?**

**1246. How do you determine the appropriate order of differencing for an ARIMA model?**

**1247. How do you determine the appropriate order of autoregression and moving average terms for an ARIMA model?**

**1248. What is the difference between an ARIMA model and an ARMA model?**

**1249. What is the difference between a stationary and non-stationary time series?**

**1250. How do you test for stationarity in a time series?**

**1251. What are some common techniques for transforming a non-stationary time series into a stationary one?**

**1252. How do you evaluate the performance of an ARIMA model?**

**1253. How do you forecast future values using an ARIMA model?**

**1254. What are some limitations of using an ARIMA model for time series analysis?**

**1255. What are the conditions for a time series to be stationary?**

- 1256. What is white noise?
- 1257. Is white noise a type of stationary time series?
- 1258. What is correlation?
- 1259. How is correlation calculated?
- 1260. What does ACF stand for?
- 1261. What is Autocorrelation Function?
- 1262. What does the Autocorrelation function signify?
- 1263. What does PACF stand for?
- 1264. What is Partial Autocorrelation Function?
- 1265. How is Partial Autocorrelation calculated?
- 1266. What does the Partial Autocorrelation Function signify?
- 1267. What is a persistence model in time series?
- 1268. How does a persistence model work?
  
- 1269. Why is a persistence model used?
- 1270. What is regression?
- 1271. What are different types of regression?
- 1272. How does regression work?
- 1273. What is Auto-Regression?
  
- 1274. How is ACF and PACF used in Auto regression in time series?
- 1275. What will happen if you don't use ACF or PACF to filter the parameters to be used in the auto regression model?
- 1276. What is a moving average?
- 1277. What is the use of using moving average on a time series?
- 1278. What is an exponential moving average (EMA)?
  
- 1279. What is a moving average model in time series forecasting?
- 1280. What does 'ARMA' stand for?
- 1281. What is an Auto Regressive Moving Average (ARMA) model?
- 1282. What is the notation used for an ARMA model?
- 1283. How does an ARMA model work?
- 1284. What does ARIMA stand for?
- 1285. What is the notation used for ARIMA models?
- 1286. When is an ARIMA model usually used in time series forecasting?
- 1287. How does an ARIMA model work?
- 1288. What is the use of using the differences in ARIMA model?
- 1289. What are some variations of the ARIMA model?
- 1290. What does 'SARIMA' stand for?
- 1291. When is the "SARIMA" model used in time series forecasting?
- 1292. What is the notation used for SARIMA?
  
- 1293. What does the seasonality factor 'm' in the SARIMA model signify?
- 1294. How does a SARIMA model work?

1295. What does ARIMAX stand for?
1296. What is ARIMAX model ?
1297. What are two types of time series classified according to the number of variables?
1298. When is an ARIMAX model used in time series forecasting?
1299. How does an ARIMAX model work?
1300. Libraries for ARIMAX model in python?
1301. Load and create an ARIMAX model
1302. Fit the model and generate prediction
1303. What is MAE?
1304. What is RMSE?
1305. What is MAPE?
1306. What is SMAPE?
1307. Difference between ARIMA, ARIMAX and SARIMA model?
1308. Application of ARIMA model?
1309. What is the difference between qualitative and quantitative forecasting methods, and what are some common techniques used in each?
1310. How do you evaluate the accuracy and performance of a forecasting model, and what are some common metrics used to do so?
1311. What are some common challenges and limitations associated with time series forecasting, and how do you address them?
1312. How do you handle missing data or outliers in a time series data set, and how do they affect your forecasting results?
1313. How do you handle seasonality in time series data, and what are some common techniques used to remove or adjust for it?
1314. What is the difference between stationary and non-stationary time series data, and how do you handle each when building a forecasting model?
1315. How do you select the appropriate forecasting model for a given data set, and what factors should be considered in making this decision?
1316. What is the difference between univariate and multivariate time series analysis, and when would you choose to use one over the other?
1317. How do you handle trend and drift in time series data, and what are some common techniques used to remove or adjust for them?
1318. What are some common software tools used for time series forecasting, and how do you choose the appropriate tool for a given application?

**1319. What is Data??**

**1320. What is the difference between data and information?**

**1321. Why is data important?**

**1322. What is Data Pre-processing?**

**1323. Why data pre-processing is so important?**

**1324. What is raw data?**

**1325. What are the types of raw data?**

**1326. What is types of data?**

**1327. What are the 5 Major Steps in Data Pre-processing?**

**1328. What type of dataset does the machine learning algorithms work?**

**1329. What are Libraries and why do we need them?**

**1330. What are the popular libraries in Python we use for data pre-processing?**

**1331. How to set the working directory in Spyder?**

**1332. How to import the dataset which we want to work on?**

**1333. How to pass the address of the dataset on the function?**



**1334. What are the additional parameters of the read\_csv() function ?**

**1335. How to have a brief look at the data ?**

**1336. Which commands will give you a descriptive look at the data?**

**1337. What is missing data and why it is a big problem?**

**1338. When should be consider deleting the missing data?**

**1339. How can we impute the categorical missing values?**

**1340. How can we impute the continuous missing variable?**

**1341. How to find out the missing numbers present in the data?**

**1342. How to find out the total number of missing values in the data?**

**1343. What are Outliers??**

**1344. How should you handle Outliers in the dataset?**

**1345. How can Outliers occur in the dataset?**

**1346. How to Detect Outliers??**

**1347. What is BoxPlot?**

**1348. How to Treat Outliers?**

**1349. What is the syntax of the boxplot for visualizing the outliers?**

**1349. How to generate a pairplot ?**

**1350. What does the 'kind' parameter stand for?**

**1351. What is Histogram and how to generate the histogram of the variables?**

**1352. What is Normal Distribution?**

**1353. Why do we need Log Transformation?**

**1354. What is Skewness?**

**1355. What is Kurtosis?**

**1356. How can we convert the variables data type ?**

**1357. What is correlation?**

**1358. How are covariance and correlation different from one another?**

**1359. What is Multicollinearity and how we find it in Python?**

**1360. What is VIF?**

**1361. What is a confusion matrix and why do you need it?**

**1362. How do we check the normality of a data set or a feature?**

**1363. What is the idea behind Splitting the data?**

**1364. What is the threshold for splitting the data?**

**1365. What is Scaling?**

**1366. What are the different types of Scalers available?**

**1367. Explain in Min Max Scaler**

**1368. Explain Max abs Scaler**

**1369. Explain Robust Scaler**

**1370. Explain Standard Scaler**

**1371. Why do we need to convert the categorical variables into numerical ones?**

**1372. What to do when you have a variable with no missing values but has no variance?**

**1373. If your dataset is suffering from high variance, how would you handle it?**

**1374. What do you mean by feature engineering?**

**1375. What do you mean by feature splitting?**

1376. How do you select the important features in your data?
1377. What are the different ways by Which you can convert the categorical variables into numerical ones.?
1378. Explain Label Encoder
1379. Explain Manual Mapping
1380. What are Dummy Variables?
1381. Explain One Hot Label Encoding
1382. What are the Different Types of Feature Selection Techniques?
1383. What is a filter method and what are the types of filter method?
1384. What is a wrapper method and what are the types of a wrapper method?
1385. What is Statistics?
1386. What is the importance of Statistics in Machine Learning?
1387. How can Statistics be classified?
1388. What are the measures of Central Tendency?
1389. What is mean of a sample?
1390. What is the median of a sample?
1391. What steps should be followed in order to calculate the median of a Sample?
1392. What is the Mode of a sample?
1393. What is train-test split?
1394. How can we implement train test split in our python notebooks?
1395. How can one analyze a Regression Model?
1396. What is a loss function?
1397. What are the various types of loss function in order to analyze a regression model?
1398. What is MAE?
1399. How can you implement MAE in your python notebook and analyze the regression model?
1400. What is RMSE and MSE?
1401. How can you implement MSE and RMSE in your python notebooks and analyze the regression model?
1402. What is the relation between the two main loss functions, RMSE and MAE?
1403. How can one analyze a classification model?

1404. What do you mean by accuracy of a classification model?
1405. What is a confusion Matrix?
1406. What is True Positive?
1407. What is False Positive?
1408. What is True Negative?
1409. What is False Negative?
1410. What is the relation between accuracy and other confusion matrix attributes?
1411. What is precision?
1412. What is the desired value of precision and why?
1413. What is recall?
1414. What is the desired value of recall and why?
1415. What do you mean by f1 score?
1416. What is the need of f1 score for evaluating a model and what should be its desired value?
1417. Explain Confusion Matrix and its attributes with the help of an example.
1418. How can we implement the confusion matrix along with all its attributes in python notebook?
1419. What do you mean by threshold and optimizing the threshold value?
1420. What is ROC Curve?
1421. What steps should be followed in order to create a ROC Curve?
1422. What is AUC?
1423. What do you mean by a classification model?
1424. What do you mean by a regression model?
1425. What are parameters and hyperparameters?
1426. What is Hypothesis Testing?
1427. What are Null and Alternate Hypothesis?
1428. What are the steps involved in order to perform Hypothesis Testing?
1429. Explain or illustrate hypothesis testing with the help of an example.
1430. What do you understand by p-value?
1431. What are statistical tests?
1432. What are the various types of parametric tests and when do we use them?
1433. What is the need of all the parametric tests after applying Hypothesis Testing?
1434. What is Ensemble Model?
1435. What are the various types of ensemble techniques?
1436. What do you mean by AIC and BIC?
1437. On what basis are AIC and BIC values calculated?
1438. How does the above-mentioned model parameters affect the model?
1439. How are AIC and BIC values calculated and what is the desired value for the selection of a better model?
1440. What do you understand by Correlation?
1441. What are the different types of Correlation?
1442. What is Multicollinearity?

1443. What are the methods to remove Multicollinearity?
1444. What is the statistical significance of degree of freedom?
1445. What does one mean by Degree of Freedom in Machine Learning?

## **MS Excel**

1446. Explain MS Excel in brief.
1447. What is a spreadsheet?
1448. What is a cell address in Excel?
1449. How to freeze columns in Excel?
1450. What is formula in Excel?
1451. What is function in Excel?
1452. Mention the order of operations used in Excel while evaluating formulas.
1453. What is count function?
1454. What is counta function?
1455. Can you protect workbooks in Excel?
1456. Does each cell have unique address?
1457. How can you add cells, rows or columns in Excel?
1458. How would you format a cell? What are the options?
1459. What is the use of comment? How to add comments to a cell?
1460. What does the red triangle indicate at the top right hand corner of the cell?
1461. How would you add comments to a cell?
1462. How can you use the INDEX and MATCH functions together in a formula in Excel?
1463. What can you use instead of VLOOKUP to get same result?
1464. What is the difference between relative and absolute cell references in MS Excel?
1465. How to check two columns are equal in excel?
1466. How can you use VLOOKUP in MS Excel to search for a specific value in a table?
1467. How would you use the SUMIF function in MS Excel to sum values based on a specific condition?
1468. Which are the different workbook protection types in Excel?
1469. How is a Formula different from a Function in Excel?
1470. How will you write the formula for the following? -  
Multiply the value in cell A1 by 10, add the result by 5, and divide it by 2.
1471. How do you create a hyperlink in Excel?
1482. How can we merge multiple cells text strings in a cell?
1483. What is the syntax of Vlookup?
1484. Which function is used to determine the day of the week for a date?
1485. What is the shortcut to add a filter to a table?

1486. Write a VBA function to calculate the area of a rectangle.
1487. Calculate your age in years from the current date.
1488. How many report formats are available in Excel and what are their names?
1489. How would you calculate the number of occurrences of a specific text in a range of data in Excel?
1489. How would you find the second largest value in a range of data in Excel?
1490. How would you calculate the product of the largest 3 values in a range of data in Excel?
1491. How would you calculate the average of only positive numbers in a range of data in Excel?
1492. How would you create a drop-down list in Excel?
1493. How would you calculate the difference between the largest and smallest values in a range of data in Excel?
1494. . What is the difference between absolute and relative cell references?
1495. How would you calculate the number of days between two dates in Excel?
1496. How would you calculate the percent change between two values in Excel?
1497. How would you calculate the correlation coefficient between two sets of data in Excel?
1498. How can we fix #N/A error when using VLOOKUP function?
1499. What are other alternate functions which can be used in place of VLOOKUP?
1500. I want output as Hello/Hola in one cell and both the Hello, Hola are entered in 2 different cells now how can I achieve my desired output?
1501. How can one find outliers using Ms- Excel?
1502. We have limited number of cells present in one sheet in excel but I want to work on more dataset on the same sheet using more rows and columns and do the required analyses then in this situation how can I do it?
1503. What all quality checks a data analyst needs to do and how can it be done in Ms-Excel?
1504. In a column, values are entered as name, city\_name. For my analyses I want to have both the values in different columns. How can it be done?
1505. What is the date format followed in Ms-Excel?
1506. How can I fix date issues in excel something like

**03/12/2022 where 12 should be my month and 03 as date of that day?**

**1507. I want to know count, sum and maximum of 5 different product categories of a brand. Then how can we do it quickly?**

**1507. How can I filter data in Microsoft Excel?**

**1508. How to use the IF function in MS Excel?**

**1509. How can I apply conditional formatting in a range of cells?**

**1510. How can you create a dynamic named range in Microsoft Excel?**

**1511. How can you use the Solver tool in Microsoft Excel to solve optimization problems?**

**1512. How can I format a cell as a percentage in Microsoft Excel?**

**1513. How can I sort data in Microsoft Excel?**

**1514. How can you use the VBA programming language in Microsoft Excel to automate tasks and create custom solutions?**

**1515. How can you use the data analysis tools in Microsoft Excel, such as regression analysis, to perform statistical analysis?**

**1516. How can you use the INDEX and MATCH functions in Microsoft Excel to perform advanced lookups?**

**1517. is it possible to hide or show the ribbon?**

**1518. Apply vlookup formula to find the marks of "amreez" in Math subject given the column**

**is C and cell number is 6.**

**1519. . How to remove duplicates in the excel sheet?**

**1520. How to sum values of marks only for ""RAMESH"" in the name column?**

**You can Assume the range of cells**

**1521. How to extract First Name from a full name?**

**1522. How can you restrict someone from copying a cell from your worksheet?**

**1523. How will you write the formula for the following? - Multiply the value in cell A1 by 10, add the result by 5, and divide it by 2.**

**1524. How to highlight those cells where total sales > \$5000**

**1525.. Count the number of blank cells in column ""SALES"" from cells A5:A20?**

**1526. What are left, right, fill and distributed alignments?**

**1527. How is VLOOKUP different from the LOOKUP function?**

**1528. Calculate your age in years from the current date.**

**1529. Write a VBA function to calculate the area of a rectangle.**

**1530. Write a VBA code to create a bar chart with the given data.**

**1531. How do you create a pivot chart in Excel?**



**1532. How would you work around the VLOOKUP limitation of working only in left to right direction?**

**1533. - Which function will you use to count values containing a specific range containing data of any type?**

**1534. When and how does a pivot table gets Refreshed?**

**1535. How would you find the top 10 products sold during a year by revenue if you are given Product ID, Product Name, Unit Price, Unit Sold and Billing Date ?**

**1536. A Vlookup formula contains \$A\$3 in its Vlookup value, what pupose does \$ play here?**

**1537. Can you protect your data in Excel? If so, how?**

**1538. How do you apply a single format to all the sheets present in a workbook?**

**1539. There are multiple blank cells present in your dataset. How to ensure they do not interfere in your analysis?**

**1540. Is It Possible To Use Multiple Data Sources To Render PivotTables?**

**1541. Numerology Sum of the Digits aka Sum the Digits till the result is a single digit**

**1542. Get File Name through Formula**

**1543. Number of times a character appears in a string**

**1544. Calculate Age from Given Birthday**

**1545. Number of Days in a Month**

**1546. Reverse a String**

**1547. Convert English Alphabets to Numbers**

**1548. Roman Representation of Numbers**

**1548. Extract Domain Name from an E Mail ID**

**1549. Write the function to solve this problem: Divide the value of cell H3 by square root of value of cell A2 and multiply this entire term by 5.**

**1550. What is the order of precedence while executing operations in MS Excel?**

**1551. What is the result of INT(H2) where the data in H2 cell = 115.89?**

**1552. Calculate your current age using MS Excel functions when the present date is provided.**

**1553. What will be the result of the function \*ch if the data in various cells are ""March, Chair, Chess""?**

**1554. Which function in Excel rounds a given number up, to the nearest multiple of significance?**

**1554. Write a function to find random numbers between 5 and 90.**

**1555. Which function should be entered to calculate all products marked with ""YES"" in the expired column and ""Refrigerator"" in the store column?**

**1556. How can you update the values of formula cells if Auto Calculate mode of Excel is disabled?**

**1557. Which of the following formulas will Excel not calculate?**

- A. =SUM(Sales)-A3**
- B. =SUM(A1:A5)\*.5**
- C. =SUM(A1:A5)/(10-10)**
- D. =SUM(A1:A5)-10**

**Big Data Technologies**

**1558. What is PostgreSQL?**

**1559. What is Debezium used for?**

**1560. What are CI and CD?**

**1561. What is Hadoop used for?**

**1562. What is Data Warehouse?**

**1563. Difference between data lake and data warehouse.**

**1564. What is orchestration?**

**1565. What do you mean by data pipeline?**

**1566. What is the difference between Hive and Presto?**

**1567. Define HDFS**

**1568. Difference between Apache Spark and Hadoop?**

**1569. What are the 5 types of database?**

**1570. Does Presto stores data?**

**1571. What is Hive used for?**

**1572. What is the difference between Hive and Presto?**

**1573. What is the importance of Hadoop?**

**1574. What are the components of Hadoop?**

**1575. What are the three layers of Big Data Architecture?**

**1576. What is the use of Debezium?**

**1577. What is the difference between Kafka Connect and Debezium?**

**1578. What is CDC?**

**1579. How is Hadoop and Big Data related?**

**1580. How Map Reduce works?**

**1581. How to deploy a Big Data Model? Mention the key steps involved.**

**1582. What is the use of Cassandra?**

**1583. Is Cassandra a SQL database?**

**1584. Is MongoDB better than Cassandra?**

**1585. S3 vs HDFS**

**586. Is S3 slower than HDFS?**

**1587.. What are the different layers of Data platform architecture?**

**1588. What is meant by data cube?**

**1589. What is the difference between OLAP and OLTP?**

**1590. Which is better OLAP or OLTP?**

**1591. Is Hadoop OLTP or OLAP?**

**1592. What is the ACID in database?**

**1593. SQL order of execution**

**1594. What is a Zookeeper? What are the benefits of using a zookeeper?**

**1595. What are the benefits of Zookeeper?**

**1596. What is the default replication factor in HDFS?**

**1597. What are the features of Apache Sqoop?**

**1598. What is partitioning in Hive?**

**1599. Explain the steps of big data project life cycle**

**1600. What Are Big Data Infrastructure Solutions?**

- 1601. Different file formats to store and process input data using Apache Hadoop
- 1602. What is zookeeper?
- 1603. What is sqoop?
- 1604. What is cdc in data engineering?
- 1605. What is Hudi and what is it used for?
- 1606. What is CRUD in Data Engineering?
- 1607. What is the difference between delete, upsert and incremental pipeline?
- 1608. How is data partitioned in Hadoop, and why is this important for performance?
- 1609. What is MapReduce, and how is it used in Hadoop?
- 1610. What is Spark, and how does it differ from Hadoop? What are some common use cases for Spark?
- 1611. What is a distributed file system, and how does it differ from a traditional file system?
- 1612. What is the CAP theorem, and how does it relate to distributed systems?
- 1613. What is NoSQL, and how does it differ from traditional relational databases?
  
- 1614. What is a data warehouse, and how is it used for business intelligence?
  
  
- 1615. How do you ensure the security and privacy of big data, especially when dealing with sensitive information?
  
  
- 1616. What are some of the most common techniques used to extract insights from big data?
- 1617. How do you validate the accuracy and completeness of big data sets?
- 1618. What are some of the most common tools and technologies used in big data analytics?
- 1619. How do you handle data quality issues in big data sets?
- 1620. What are some of the most common use cases for big data processing, and how do they provide value to organizations?
- 1621. In what all modes can Hadoop be run?
- 1622. What are the real-time industry applications of Hadoop?
- 1623. What is HBase?
- 1624. What is a Combiner?
- 1625. Name some of the important tools useful for Big Data analytics.
- 1626. What are the five 'V's of Big Data?
  
  
- 1627. What are HDFS and YARN? What are their respective components?
- 1628. What is FSCK?
- 1629. What is the goal of A/B Testing?
- 1630. What is a Distributed Cache?
- 1631. Explain the steps to deploy a Big Data solution.

1632. What is Hive Metastore?
1633. What kind of Dataware house application is suitable?
1634. It's true that HDFS is to be used for applications that have large data sets. Why is it not the correct tool to use when there are many small files?
1635. What are the main distinctions between NAS and HDFS?
1636. Is it possible to create multiple tables in the hive for the same data?
1637. What is Hadoop Big Data Testing?
1638. What do we test in Hadoop Big Data?
1639. How do we validate Big Data?
1640. How is data quality being tested?
1641. What do you understand by Data Staging?
1642. What is Architecture Testing?
1643. What is Performance Testing?
1644. What is Data ingestion?
1645. What is Data Processing in Hadoop Big data testing?
1646. What are the tools applied in these scenarios of testing?
1647. What is Query Surge?
1648. What Benefits do Query Surge provides?
1649. What is Query Surge's architecture?
1650. Can you explain the difference between batch processing and stream processing, and when would you use one over the other?
1651. How do you measure the performance of a big data processing system, and what metrics are commonly used for this purpose?
1652. Can you describe the Lambda Architecture, and how it can be used to build a scalable and fault-tolerant big data processing system?
1653. What are some of the common data integration challenges faced in big data projects, and how can they be addressed?
1654. Can you explain how data sharding works, and how it can be used to improve the performance of a distributed data processing system?
1655. What are the key considerations when designing a data model for a NoSQL database, and how does this differ from designing a data model for a relational database?
1656. Can you describe some of the common approaches for data governance in big data projects, and how they can be used to ensure compliance with data privacy and security regulations?
1657. How can you optimize the performance of a big data processing system, and what techniques are commonly used for this purpose?

## **AWS**

- 1658. Explain what is AWS ?**
- 1659. Mention what the key components of AWS are?**
- 1660. How can you send a request to Amazon S3?**
- 1661. Explain how the buffer is used in Amazon web services(AWS)?**
- 1662. What is AWS Lambda?**
- 1663. Name the AWS service that exists only to redundantly cache data and images?**
- 1664. What are the different types of Load Balancers in AWS services?**
  
- 1665. Name some of the DB engines which can be used in AWS RDS?**
- 1666. What are the important features of Amazon cloud search?**
- 1667. What is the role of AWS CloudTrail?**
- 1668. Explain how Amazon SNS can be used for push messaging and event-driven architectures?**
- 1669. How can you configure an Amazon S3 bucket to serve static content with custom domain names and SSL certificates?**
- 1670. How can you migrate a complex, multi-tier application to AWS with minimal downtime?**
- 1671. how Amazon Kinesis can be used to process real-time streaming data?**
- 1672. What is the difference between Amazon S3 Intelligent-Tiering and Amazon S3 Standard-IA?**
- 1673. A shipping company has to process requests received from different sources in batches at a periodic intervals in order to receipt of requests.**
- 1674. What is SQS?**
  
- 1675. The sales team of a company needs to see its sales data in form of visual dashboards.**
- 1676. What is Amazon Quicksight?**
- 1677. A company wants to track the usage of AWS services and resources by its departments in real time? What service can help them monitor the usage?**
  
- 1678. What is Cloudwatch?**
- 1679. A company wants to deploy its new application without using a virtual machine how can it do it?**
- 1680. What is Amazon's frigate?**
  
- 1681. How can we host only one function which can execute task on dedicated event triggers?**
- 1682. What is the AWS service used for delivering push notifications to mobile devices?**
- 1683. What is the AWS service used for executing, managing, and scaling background jobs?**
- 1684. What is the AWS service used for storing and processing large amounts of data?**

1685. What is the AWS service used for disaster recovery and backing up data?
1686. What is the AWS service used for centralized management of AWS resources?
1687. How many S3 buckets can be created?
1688. What is the maximum limit of elastic IPs anyone can produce?
1689. Mention 2 major issues that are not supported by the AWS support?
1690. With specified private IP addresses, can an Amazon Elastic Compute Cloud (EC2) instance be launched? If so, which Amazon service makes it possible?
1691. Will your standby RDS be launched in the same availability zone as your primary?
- 1692.. Your business prefers to use its email address and domain to send and receive compliance emails. What service do you recommend to implement it easily and budget-friendly?
1693. Give one instance where you would prefer Provisioned IOPS over Standard RDS storage?
1694. Can you change the Private IP Address of an EC2 instance while it is running or in a stopped state?
1695. How many Subnets can you have per VPC?
1696. How can you send a request to Amazon S3?
- 1697.. You plan to design an application by encrypting all the data in an Amazon Redshift cluster. How will you encrypt the data at rest?
1698. An organization decides to build an Amazon Redshift cluster to host sensitive data in their shared services VPC. What control does the organization implement for networks accessing the cluster?
1699. An application saves the logs to an S3 bucket. A user needs to keep the logs for one month for troubleshooting purposes and then clear the logs. What action will enable this?
1700. A website experiences inconstant traffic, and the database cannot keep up with the write requests during peak traffic times. What AWS Service helps to decouple the web application from the database?

## **AWS**

1701. A solution architect is designing a new web application on AWS. To make the application very popular, the architect focuses on software development and new features without managing or provisioning instances. Which solution is best suited for that?
1702. What are the different types of instances?
1703. The types of AMI provided by AWS are:
1704. Amazon Web Services supports which Type II Audits?
1705. Amazon S3 is which type of storage service?
1706. How many buckets can you create in AWS by default?
1707. How would you design a cost-effective solution for processing large amounts of data in real-time using AWS technologies?

1708. How would you design an architecture to support a multi-region, highly available and disaster recovery setup for a web application?
1709. How would you design a cost-efficient solution for archiving infrequently accessed data for compliance purposes, taking into consideration retrieval times and cost?
1710. What would you do if your AWS S3 bucket suddenly became public and sensitive data was being downloaded by unauthorized users?
1711. How would you design a solution for migrating an on-premises application to AWS?
1712. How would you implement a solution to process big data in real-time using AWS technologies, such as Kinesis, Lambda, and S3?
1713. How can you configure an Amazon S3 bucket to serve static content with custom domain names and SSL certificates?
1714. How can you migrate a complex, multi-tier application to AWS with minimal downtime?
1715. How Amazon Kinesis can be used to process real-time streaming data?
1716. How disaster recovery can be implemented for AWS infrastructure?
1717. What is the difference between Amazon S3 Intelligent-Tiering and Amazon S3 Standard-IA?
1718. How would you implement an auto-scaling solution for an application that experiences sudden spikes in traffic?
1719. How do you differentiate between terminating and stopping an instance from each other?
1720. How can you spot the instant difference between on-demand instances and reserved instances?
1721. Do you think one elastic IP address is enough for all instances of running?
1722. What are the differences between Amazon RDS, Dynamo DB, and Redshift?
1723. What are the important components of IAM?
1724. What is Lambda@Edge in AWS?
1725. What are the security mechanisms available in Amazon S3?
1727. If an organization splits its workload between public cloud and private servers, what do you call this approach?
1728. In simple terms, explain the difference between vertical and horizontal scaling.
1729. What is an Elastic IP Address?
1730. Can you explain the difference between Amazon S3 and Amazon EC2?
1731. How do you handle disaster recovery in AWS?
1732. How would you optimize an Amazon RDS database for improved performance?
1733. What is the difference between an Amazon VPC and a public subnet?
1734. Can you explain the security measures available in Amazon S3?



**1735. How do you migrate an on-premise database to Amazon RDS?**

**1736. How would you troubleshoot a high network latency issue in Amazon EC2?**

**1737. Can you explain the benefits and limitations of Amazon Lambda?**

**1738. How do you handle storage management in Amazon S3?**

**1739. Can you explain the difference between Amazon Route 53 and Amazon CloudFront?**

**Amazon Route 53 and Amazon CloudFront are two different services offered by Amazon Web Services (AWS).**

**1740. Suppose you have an application that serves content over a network to millions of users and is running on an EC2 instance. The application has become increasingly popular, and the EC2 instance is now overutilized. How would you design a scalable solution to handle this increased demand?**

**1741. How would you design a highly available and disaster-recovery solution for a multi-tier web application on AWS?**

**1742. What strategies would you use to secure sensitive data stored in AWS?**

**1743. How would you implement a cost-effective disaster recovery solution for a large number of EC2 instances in different regions?**

**1744. Can you explain the difference between EC2 placement groups and EC2 auto-scaling groups and when you would use each?**

**1745. How would you design a scalable and cost-effective storage solution for a big data analytics application on AWS?**

**1746. Can you explain how to optimize network performance for an application running on multiple EC2 instances in different availability zones?**

**1747. How would you set up a multi-tier application in AWS using EC2, RDS, and S3, and ensure high availability and reliability?**

**1748. Can you describe how you would implement a serverless architecture for a web application on AWS using Lambda, API Gateway, and DynamoDB?**

**1749. How would you design a solution to monitor and alert on critical issues in an AWS infrastructure?**

**1750. What is the difference between an Amazon ELB and an Amazon ALB?**

**1751. What is Amazon SNS and how do you use it?**

**1752. What is the difference between an Amazon VPC and an Amazon VPN?**

**1753. How do you monitor AWS costs and optimize resources for cost savings?**

**1754. What are the different types of load balancers in AWS and when would you use each type?**

**1755. What is AWS CloudFormation and how do you use it?**

**1756. What is AWS Identity and Access Management (IAM) and how do you use it?**

**1757. How do you ensure high availability and fault tolerance in your AWS architecture?**

**1758. What are some of the key considerations for designing a scalable and highly available architecture on AWS?**

**Regular expression**

**1759. What is the most common field or application where regular expressions are used?**

**1760. How can you use backreferences in a regular expression to match a string that contains a specific pattern that is repeated multiple times, with each repetition having a unique variation?**

**1761. How would you use a regular expression to match and extract all the phone numbers from a large block of text, where the phone numbers can be in different formats explain with code?**

**1762. How would you use a regular expression to match and extract all the URLs from a large block of text?**

**1763:**

**What the mean of different symbols like ^ \$ in Regular Expression?**

**1764. How do you Improve the performance of regular expressions?**

**1765. How do you match starting and ending characters in a string using a regular expression?**

**1766. How do you match a specific number of occurrences of a pattern?**

**1767. How do you match overlapping patterns in a string?**

**1768. How do you perform a negative lookahead or look behind match in a regular expression?**

**1769. How do you match a pattern that is separated by a specific character or string?**

**1770. How do you match a pattern that is optionally present in a string?**

**1771. How do you extract a specific portion of a matched pattern?**

**1772. What is the simple regular expression for a two-digit decimal?**

**1773.**

**Write a regex to split String by a new line.**

**1774. What is Zero width assertions in Regular expressions?**

**1774.**

**How to replace all non-alphanumeric characters with empty strings?**

**1775. How can you match a string that does not contain a specific pattern in a regular expression?**

**1776. Can you explain how backreferences work in regular expressions?**

**1777. Mention one way how can regular expressions be optimized for extracting a substring from non-consistent format of string?**

**1778. How can you identify slow or inefficient regular expressions in SQL?**

**1779. Why are REGEXPs' are considered to computationally expensive?**

**1780. How do you use the tilde operator (~) or equivalent operator in SQL to match regular expressions against text data?**

**1781. What are the factors that affect the performance of regular expressions in SQL?**

**1782. Can you explain the difference between greedy and non-greedy matching in regular expressions?**

**1783. Can non-greedy matching in regular expressions reduce the computational expense of the SQL query?**

**1784. What is the simple regular expression for a two-digit decimal or a decimal with a precision of 2 ?**

**1785. How to search for all text files from a system ?**

**1786. Can we give regular expression range in descending order like [30-5] ?**

**1787. What is the best way to match the term TheDataMonk at the start of a line?**

**1787. Write simple regex expression to match the date pattern (like 23.08.2023, 07/04/2016, 12-08-2002) ?**

1788. Try naming more than one animal whose name consists of 3 letters and the middle letter is the vowel a in regex form ?
1788. Write regex expression for 10 digit mobile number starting with 8 or 9 ?
1789. Write a regex expression with first character in uppercase and rest all characters in lower case with only one digit allowed in between ?
1790. Write a simple regular expression to capture the content of each line, without the extra whitespace ?
1791. Regex expression for trying to extract the protocol, host and port from any kind of general URL ?
1792. Write a Python program to check that a string contains only a certain set of characters (in this case a-z, A-Z and 0-9).
1793. Write a Python program that matches a string that has an a followed by zero or more b's.
1794. Write a Python program that matches a string that has an a followed by one or more b's.
1795. Write a Python program to remove leading zeros from an IP address.
1796. Write a Python program to search the numbers (0-9) of length between 1 to 3 in a given string.
1797. Write a Python program to search some literals strings in a string.
1798. Write a Python program to search a literals string in a string and also find the location within the original string where the pattern occurs.
1799. Write a Python program to find the substrings within a string.
1800. Write a Python program to replace whitespaces with an underscore and vice versa.
1801. Write a Python program to match if two words from a list of words starting with letter 'P'.
1802. What are the most commonly used metacharacters in regular expressions?
1803. How do you match a specific character or sequence of characters using regular expressions?
1804. How do you match a range of characters using regular expressions?
1805. How do you match repeating characters using regular expressions?
1806. How do you match multiple patterns using regular expressions?
1807. How do you match substrings using regular expressions?
1808. What is the difference between greedy and non-greedy matching in regular expressions?
1809. How do you perform case-insensitive matching using regular expressions?
1810. How do you capture and extract parts of a string using regular expressions?
1811. How do you validate and parse data using regular expressions, such as email addresses or phone numbers?
1812. How do you use regular expressions with programming languages such as Python, Java, or JavaScript?

## **CLTV**

- 1813. What is CLTV and why is it important for businesses?**
- 1814. How do you calculate CLTV, and what factors do you need to consider?**
- 1814. What is the difference between CLTV and customer acquisition cost (CAC)?**
- 1815. How do you measure and track CLTV over time, and what tools or techniques do you use?**
- 1816. How do you use CLTV to inform your marketing and customer retention strategies?**
  
- 1817. How do you use CLTV to prioritize and allocate resources, such as sales and marketing budget?**
- 1818. How do you use CLTV to evaluate the impact of customer churn, upsells, and cross-sells on the business?**
- 1819. How do you use CLTV to identify and target high-value customers, and what are the best practices for doing so?**
- 1820. How do you use CLTV to forecast future revenue and growth potential, and what challenges do you face in doing so?**
- 1821. How do you use CLTV to measure the return on investment (ROI) of your customer engagement and retention initiatives?**
- 1822. How do you use CLTV to measure the impact of customer experience and customer satisfaction on the business?**
- 1823. How do you use CLTV to evaluate the effectiveness of your customer loyalty and retention programs, and what metrics do you use to do so?**
- 1824. What is the formula for calculating CLTV**
- 1825. How to Maximize Your CLTV?**
- 1826. What are the Benefits of Maximizing Your CLTV?**
- 1827. Applications of CLV**
- 1828. Strategies to Increase CLV**
- 1829. Types of Customer Lifetime Value**
- 1830. How to Improve Your Purchase Frequency?**
- 1831. How to Improve Your Customer Lifetime Value (CLTV)**
  
- 1832. How do you calculate Customer Lifetime Value (CLTV) in a B2B scenario where the customer acquisition cost is high and the customer relationship spans over several years?**
- 1833. How do you take into account customer churn when calculating CLTV?**
- 1834. Is it fair to use CLTV at aggregate level or individual level? Why?**
- 1835. Is CLTV historic or predictive?**
- 1836. How is CLTV helpful for sales management decisions?**
- 1837. What are the methods to improve CLTV?**
- 1838. What are the components of CLTV?**
- 1839. What is the difference between cost to serve and customer lifetime value ?**
- 1840. What is a cohort model in CLTV?**

**1841. Techniques to Raise CLV?**

**1842. How much of the CLV has already been seen compared to the expected future revenue/profit?**

**1843. Is CLV historical, predictive, or both?**

**1844. Can the CLV take into consideration the different client tenures?**

**1845. Is it reasonable to include or exclude extremely elderly customers?**

**1846. When to use the simple CLV formula?**

**1847. How does profit depend on customer lifetime value?**

**1848. Difference between LTV AND CLTV?**

**1849. If the projected CLTV value fluctuates frequently, how can you follow those changes?**

**1850. What type of people are most profitable through CLTV analysis?**

**1851. When does the profit margin normally start to fall?**

**1852. Difference between LTV AND CLTV?**

**1853. Is the CLV at the aggregate level or individual level? If it is aggregate, is it the mean or median?**

**1854. How much of CLV is already observed vs predicted future revenue/profit?**

**1855. If the CLV is predictive, how far into the future does that prediction extend?**

**1856. Can the CLV account for varying tenure of customers?**

**1857. Does the predictive CLV number change often and, if so, how can you track changes to CLV?**

**1858. What do you think about the current maturity level around the Customer Lifetime Value metric?**

**1859. What are the biggest challenges in using the Customer Lifetime Value (CLV) metric to design business strategy?**

**1860. What kind of external data can be leveraged to improve the confidence in CLV metric values?**

**1861. In order to calculate Customer Lifetime Value (CLV), one needs to have the correct attribution model. In today's world of several consumer touch-points and multi-platform online experience, what are your recommendations to develop a reliable attribution model?**

**1862. This is the shell script:**

**(A) A file with a list of commands**

**(B) A file with special symbols**

**(C) Group of instructions**

**(D) Group of operations**

**1863. How do you add a new user to your Linux system?**

**(A) Using useradd**

**(B) Using adduser**

(C) Using Linuxconf

(D) the entire list

1864. Which shell is the most common and best to use?

A. Korn shell

B. C shell

C. Bourne shell

D. Bash Shell

1865. With \_\_\_\_ first line of every shell script begins.

(A). &

(B). #

(C). \$

(D). !

1866. What command allows the creation of environment variables?

(A) Export

(B) Set

(C) Read and Export

(D) None of the Above

1867. The hash command:-

(A) Manages an internal hash table

(B) Finds and remembers the full path name of the specified command

(C) Shows the number of hits and used command names

(D) Performs all of the aforementioned functions.

1868. Which option of the kill command sends the given signal name to the specified process?

a) -l

b) -n

c) -s

d) -a

1869. Which command identifies the resource of a command?

a) type

b) typeset

c) select

d) source

1870. Which command prints the accumulated user and system times for processes run from the shell?

a) time

b) times

c) both time and times

d) none of the mentioned

1871. Which command executes ""command,"" a built-in shell command, with the specified argument?

(A) built-in

(B) caller

(C) No command is available for this purpose

(D) None of the aforementioned are true

1872. What does the following command do?

1873. What is the difference between `""&""` and `"" ; ""` in shell scripts?

1874. What is the difference between a shell script and an executable file?

1875. How do you redirect the error output of a command to a file?

1876. What does the following command do?

1877. What does the following command do?

1878. What does the following command do?

1879. What does the following command do?

1880. What does the following command do?

1881. What is the difference between a single and double square bracket in shell scripts?

1882. What is a Shell Scripting?

1883. What is the difference between `$*` and `$@`?

1884. I have a file called 'mypaper.txt' or 'mypaper.tex' or some such somewhere, but I don't remember where I put it. How do I find it?

1885. I have a file containing a list of words. I want to sort it in reverse order and print it.

886. In a typical UNIX environment how many kernels and shells are available

1887. What does the. (dot) indicate at the beginning of a file name and how should it be listed?

1888. Shell programs are stored in which file?

1889. How to check whether a link is a hard one or a soft link?

1889. Write the difference between `$*` and `$@`

1890. What are different types of variables mostly used in shell scripting?

1891. Explain the term positional parameters.

1892. How can you redirect the standard output and standard error to separate files in shell scripting?

1893. What is the difference between `&&` and `||` in shell scripting?

1894. Write the command that is used to execute a shell file.

1895. Write the difference between `$*` and `$@`

1896. What is the best way to debug the shell script/program problems?

1897. What is the difference between single quotes ( ' ') and double quotes ( "" "" ) in shell scripting?

1898. How do you run a shell script in the background?

1899. How to check whether a link is a hard one or a soft link?

1900. What is the difference between `[[ $string == ""efg*"" ]]` and `[[ $string == efg* ]]` ?

1901. What command needs to be used to take the backup?

1902. What can be used to modify File permissions?



- 1903. What is the alternative command available to echo and what does it do?
- 1904. How to find out the number of arguments passed to the script?
- 1905. How many fields are present in a crontab file and what does each field specify?
- 1906. What is the significance of the Shebang line in Shell Scripting?
- 1907. What are the two files of crontab command?
- 1908. What is a shell and what is shell scripting?
  
- 1909. What is the difference between a shell and a terminal?
  
- 1910. What are some common shell environments, such as bash, zsh, or csh?
  
- 1911. How do you navigate the file system and manage files and directories in shell?
  
- 1912. How do you execute and manage processes in shell?
- 1913. How do you manipulate and process strings, numbers, and arrays in shell?
- 1914. How do you use variables, conditionals, loops, and functions in shell scripts?
- 1915. How do you perform input/output redirection and pipeline processing in shell?
  
- 1916. How do you use regular expressions in shell scripts?
- 1917. How do you handle errors and exceptions in shell scripts?
- 1918. How do you debug and troubleshoot shell scripts?
- 1919. How do you use shell scripts to automate tasks and run commands on multiple systems?
  
- 1920. How do you integrate shell scripts with other tools, such as sed, awk, or grep?
- 1921. What are the best practices for writing and testing shell scripts?
- 1922. How do you optimize their performance?

## **Machine Learning Topics**

- 1923. What is the importance of data cleaning?
- 1924. What are the basic checks for data cleaning?
- 1925. What is the difference between forecasting & prediction.
- 1926. Explain Missing Value Treatment by mean, mode, median, and KNN Imputation
  
- 1927. Define Loss function.
- 1928. What are the metrics to measure the performance of your Linear Regression model?
  
- 1929. What is the need to remove multicollinearity?
- 1930. Define ROC in layman terms.
- 1931. What is the best fit line in Linear Regression?

1932. When do we use Linear and when do we use Logistic regression?
1933. How is Ridge Regression different from Linear Regression?
1934. Explain precision in the simplest terms.
1935. Explain recall in simple terms.
1936. How does the value of R squared and adjusted R Squared error change when you add new variable to your model?
1937. How to select the number of trees in a random forest?
1938. How to choose k in k-means?
1939. What are the different use cases where machine learning algorithms can be used?
1940. Differentiate between KNN and K-means.
1941. What is the meaning of “Overfitting”?
1942. How to do dimension reduction?
1943. Explain Confusion Matrix Machine Learning
1944. What is Auto Regression?
1945. How to handle an imbalanced dataset?
1946. Given that you have wifi data in your office, how would you determine which rooms and areas are underutilized and overutilized?
1947. What evaluation approaches would you work to gauge the effectiveness of a machine learning model?
1948. How can we implement different word2vec methods?
1949. What is VIF?
1950. How do linear and logistic regression differ in their error minimization techniques?
1951. What is topic modeling?
1952. What is pruning in case of decision trees?
1953. You are given a data set based on the cancer detection results. You’ve built a classification model and achieved an accuracy of 95%.
1954. What are the assumptions required for linear regression?
1955. How to improve accuracy of linear regression model? Can you name a possible method of improving the accuracy of a linear regression model?
1956. Explain about the box cox transformation in regression models.
1957. How can we use the Naive Bayes classifier for categorical features?
1958. When can parallelism make your algorithms run faster?
1959. Suppose you are building a spam classifier. Which among the following two metrics would you consider while implementing the same? Precision or Recall ?

1960. How to cluster unsupervised data where all the attributes and its values are categorical?
1961. Is it beneficial to perform dimensionality reduction before fitting an SVM?
1962. What is the need of feature selection?
1963. When are the linear regression lines perpendicular?
1964. Is rotation necessary in PCA?
1965. Prior to building any kind of model, why do we need to complete the feature selection step?
1966. Name and describe three different kernel functions and in what situation you would use each.
1967. What to do if one of my columns with integer value is having more than 30% missing values?
1968. You are given a train data set having 2000 rows and 1.2 Million columns. The data set is based on a classification problem. You are asked to reduce the dimension of this data so that model computation time can become manageable. What will be your suggestion?
1969. Which one would likely perform better- Linear Regression or Random Forest Regression? Why?
1970. Do gradient descent methods at all times converge to a similar point?
1971. How to make the model free from underfitting?
1972. How would you evaluate a logistic regression model?
1973. Do you think 50 small decision trees are better than a large one?
1974. What metrics we should use to evaluate a binary classification model?
1975. What are your favorite use cases of machine learning models?
1976. If through training all the features in the dataset, an accuracy of 100% is obtained but with the validation set, the accuracy score is 75%. What should be looked out for?
1977. What are the three components of time series data?
1978. What should be the value of a good variable which we should include in our model?
1979. What will you do if removing missing values from a dataset causes bias?
1980. What is bag-of-words?
1981. How to identify the important variable for my Linear regression model ?
1982. What is the different regression algorithm? Explain the trend in forecasting.
1983. How can you handle an imbalanced dataset?

**1984. What are the steps for wrangling and cleaning data before applying machine learning algorithms?**

**1985. What happens to our linear regression model if the column z in the data is a sum of columns x and y and some random noise?**

**1986. Time series regression model got higher accuracy than decision tree model. Can this happen? Why?**

**1987. Is it better to spend five days developing a 90-percent accurate solution or 10 days for 100-percent accuracy?**

**1988. While working on model, what among them is more important- Model Accuracy or Model Performance?**

**1989. Is it necessary to perform resampling in your dataset? How would you initiate with this process?**

**1990. Give an example of outlier values and how can they be treated?**

**1991. Do having more outlier values a good thing or a bad thing?**

**1992. What is tokenization and lemmatization in NLP?**

**1993. How much data will you allocate for your training, validation and test sets?**

**1994. What is p-value?**

**1995. Is the mean imputation of missing data acceptable practice? Why or why not?**

**1996. What is multivariate normality in assumptions of Linear Regression?**

**1997. If two predictors are highly correlated, what is the effect on the coefficients in the logistic regression?**

**1998. While working on a data set, how do you select important feature?**

**1999. What could be some issues if the distribution of the test data is significantly different than the distribution of the training data?**

**2000. What is the use of stringAsFactors set as False?**

**2001. State one situation where the set-based solution is advantageous over the cursor-based solution.**

**2002. What is Multicollinearity ?**

**2003. We know that one hot encoding increases the dimensionality of a dataset, but label encoding doesn't. How?**

**2004. What are the various aspects of a Machine Learning process?**

**2005. How would you perform feature selection on the dataset?**

**2006. What is the difference between machine learning and deep learning?**

**2007. What is CNN?**

**2008. Is it better to have too many false negatives or too many false positives?**

**2009. Executing a binary classification tree algorithm is a simple task. But, how does a tree splitting take place?**

**2010. How to read PACF graph?**

**2011. What cross-validation technique would you use on a time series dataset?**

**2012. When modifying an algorithm, how do you know that your changes are an improvement over not doing anything?**

**2013. How do you select the appropriate algorithm for a given problem?**

**2014. What is overfitting in machine learning and how can you avoid it?**

**2015. What is cross-validation and why is it important?**

**2016. How do you handle missing data in a dataset?**

**2017. What is the difference between bias and variance in machine learning models?**

**2018. Can you explain the bias-variance tradeoff in machine learning?**

**2019. What is regularization and why is it important?**

**2020. How do you evaluate the performance of a machine learning model?**

**2021. What are some common techniques used for feature selection in machine learning?**

**2022. Can you cite some examples where both false positive and false negatives are equally important?**

**2023. How do you come up with an algorithm that will predict what the user needs after they type only a few letters?**

**2024. If you're attempting to predict a customer's gender, and you only have 100 data points, what problems could arise?**

**2025. Give examples where a false negative is more important than a false positive, and vice versa.**

**2026. If the model isn't perfect, how would you like to select the threshold so that the model outputs 1 or 0 for label?**

**2027. List out the difference between linear and logistic regression**

**2028. You have built a multiple regression model**

**2029. Why is mean square error a bad measure of model performance? What would you suggest instead?**

**2030. Can we use the logistic regression algorithm for a regression problem?**

**2031. Why is "Naive Bayes" naive?**

**2032. Give some problems or scenarios where map-reduce concept works well and where it doesn't work.**

**2033. When it comes to Evaluation of Linear Regression which Evaluation Metrics would be the best when we have redundant Variables in our dataset?**

**2034. What is maximum likelihood estimation? Could there be any case where it doesn't exist?**

**2035. How do predict "y" at time t+1?**

**2036. What is the role of trial and error in data analysis?**

**2037. You notice a system uses a lot of triggers to enforce foreign key constraints, and the triggers are error-prone and difficult to debug. What changes can you recommend to reduce the use of triggers?**

**2038. What is the difference between Boosting and Bagging?**

**2039. What could be some issues if the distribution of the test data is significantly different than the distribution of the training data?**

**2040. Treating a categorical variable as a continuous variable would result in a better predictive model? How?**

**2041. Explain Long Short Term Memory Algorithm in brief Question With code snippet in R or Python**

**2042. What is the main difference between K-Means and K-Means++ algorithm?**

**2043. What is the difference between cyclicity and seasonality?**

**2044. What is the acceptable value range for p,d and q in ARIMA?**

**2045. What is Homoscedasticity in assumptions of Linear Regression?**

**2046. If you're attempting to predict a customer's gender, and you only have 100 data points, what problems could arise?**

**2047. Differentiate between classification and regression in Machine Learning.**

**2048. What are the types of Machine Learning?**

**2049. Why learning rate should be less in gradient descent?**

**2050. How to identify given data is structured or unstructured?**

**2051. How would you validate a model you created to generate a predictive model of a quantitative outcome variable using multiple regression?**

**2052. How would you optimize a web crawler to run much faster, extract better information, and better summarize data to produce cleaner databases?**

**2053. For tuning hyperparameters of your machine learning model, what will be the ideal seed?**

**While working on a data set, how do you select important variables?**

**2054. We know that one hot encoding increases the dimensionality of a dataset, but label encoding doesn't. How?**

**2055. What are Loss Function and Cost Functions? Explain the key Difference Between them?**

**2056. What is Ensemble learning?**

**2057. How do you make sure which Machine Learning Algorithm to use?**

**2058. How to Handle Outlier Values?**

**2059. What are Recommender Systems?**

2060. How do check the Normality of a dataset?
2061. Explain Correlation and Covariance?
2062. What are Parametric and Non-Parametric Models?
2063. What is Reinforcement Learning?
2064. Difference Between Sigmoid and Softmax functions?
2065. How should outlier values be handled?
2066. Explain How a System Can Play a Game of Chess Using Reinforcement Learning.
2067. When Will You Use Classification over Regression?
2068. How Do You Design an Email Spam Filter?
2069. What is the Trade-off Between Bias and Variance?
2070. Briefly Explain Logistic Regression.
2071. What is the difference between Lasso and Ridge regression?
2072. What is the Bayes' Theorem? Why do we use it?
2073. Explain difference between Type I and Type II error.
2074. What is the difference between a discriminative and a generative model?
2075. What are parametric models? Provide an example.
2076. Explain LDA for unsupervised learning.
2077. Explain how you would develop a data pipeline.
2078. How do you fix high variance in a model?
2079. What are hyperparameters? How do they differ from model parameters?
2080. What is the default method for splitting in decision trees?
2081. You are told that your regression model is suffering from multicollinearity. How do verify this is true and build a better model?