Data Science Interview Questions for Top Tech Companies 2022

- ❖ Facebook Data Science Interview Questions
- ❖ Insight Data Science Interview Questions
- ❖ Twitter Data Scientist Interview Questions
- ❖ Airbnb Data Science Interview Questions
- ❖ Google Data Science Interview Questions
- ❖ LinkedIn Data Science Interview Questions
- ❖ Mu Sigma Data Science Interview Questions
- ❖ Amazon Data Science Interview Questions
- ❖ Uber Data Science Interview Questions
- ❖ Netflix Data Science Interview Questions
- ❖ Microsoft Data Science Interview Questions
- ❖ Apple Data Science Interview Questions
- ❖ Adobe Data Scientist Interview Questions
- ❖ American Express Data Scientist Interview Questions
- ❖ Quora Data Scientist Interview Questions
- ❖ Goldman Sachs Data Scientist Interview Questions
- ❖ Walmart Data Science Interview Questions
- ❖ IBM Data Science Interview Questions
- ❖ Yammer Data Science Interview Questions
- ❖ Citi Bank Data Science Interview Questions

➢Facebook Data Science Interview Questions

1) A building has 100 floors. Given 2 identical eggs, how can you use them to find the threshold floor? The egg will break from any particular floor above floor N, including floor N itself.

2) In a given day, how many birthday posts occur on Facebook?

3) You are at a Casino. You have two dices to play with. You win $10 every time you roll a 5. If you play till you win and then stop, what is the expected pay-out?

4) How many big Macs does McDonald's sell every year in the US?

5) You are about to get on a plane to Seattle, you want to know whether you have to bring an umbrella or not. You call three of your random friends and as each one of them if it's raining. The probability that your friend is telling the truth is 2/3 and the probability that they are playing a prank on you by lying is 1/3. If all 3 of them tell that it is raining, then what is the probability that it is actually raining in Seattle.

6) You can roll a dice three times. You will be given $X where X is the highest roll you get. You can choose to stop rolling at any time (for example, if you roll a 6 on the first roll, you can stop). What is your expected payout?

7) How can bogus Facebook accounts be detected?

8) You have been given the data on Facebook users' friending or defriending each other. How will you determine whether a given pair of Facebook users are friends or not?

9) How many dentists are there in US?

10) You have 2 dices. What is the probability of getting at least one 4? Also find out the probability of getting at least one 4 if you have n dices.

11) Pick up a coin C1 given C1+C2 with probability of trials p (h1) =.7, p (h2) =.6 and doing 10 trials. And what is the probability that the given coin you picked is C1 given you have 7 heads and 3 tails?

12) You are given two tables- friend_request and request_accepted. Friend_request contains requester_id, time and sent_to_id and request_accepted table contains time, acceptor_id and requestor_id. How will you determine the overall acceptance rate of requests?

13) How would add new Facebook members to the database of members, and code their relationships to others in the database?

14) What would you add to Facebook and how would you pitch it and measure its success?

15) How will you test that there is increased probability of a user to stay active after 6 months given that a user has more friends now?

16) You have two tables-the first table has data about the users and their friends, the second table has data about the users and the pages they have liked. Write an SQL query to make recommendations using pages that your friends liked. The query result should not recommend the pages that have already been liked by a user.

17) What is the probability of pulling a different shape or a different colour card from a deck of 52 cards?

18) Which technique will you use to compare the performance of two back-end engines that generate automatic friend recommendations on Facebook?

19) Implement a sorting algorithm for a numerical dataset in Python.

20) How many people are using Facebook in California at 1.30 PM on Monday?

21) You are given 50 cards with five different colors- 10 Green cards, 10 Red Cards, 10 Orange Cards, 10 Blue cards, and 10 Yellow cards. The cards of each colors are numbered from one to ten. Two cards are picked at random. Find out the probability that the cards picked are not of same number and same color.

➢Insight Data Science Interview Questions

1) Which companies participating in Insight would you be interested in working for?

2) Create a program in a language of your choice to read a text file with various tweets. The output should be 2 text files-one that contains the list of all unique words among all tweets along with the count for repeated words and the second file should contain the medium number of unique words for all tweets.

3) What motivates you to transition from academia to data science?

➢Twitter Data Scientist Interview Questions

1) How can you measure engagement with given Twitter data?

2) Give a large dataset, find the median.

3) What is the good measure of influence of a Twitter user?

➢Airbnb Data Science Interview Questions

1) Do you have some knowledge of R - analyse a given dataset in R?

2) What will you do if removing missing values from a dataset cause bias?

3) How can you reduce bias in a given data set?

4) How will you impute missing information in a dataset?

➢Google Data Science Interview Questions

1) Explain about string parsing in R language

2) A disc is spinning on a spindle and you don't know the direction in which way the disc is spinning. You are provided with a set of pins.How will you use the pins to describe in which way the disc is spinning?

3) Describe the data analysis process.

4) How will you cut a circular cake into 8 equal pieces?

➢LinkedIn Data Science Interview Questions

1) Find out K most frequent numbers from a given stream of numbers on the fly.

2) Given 2 vectors, how will you generate a sorted vector?

3) Implementing pow function

4) What kind of product you want to build at LinkedIn?

5) How will you design a recommendation engine for jobs?

6) Write a program to segment a long string into a group of valid words using Dictionary. The result should return false if the string cannot be segmented. Also explain about the complexity of the devised solution.

7) Define an algorithm to discover when a person is starting to search for new job.

8) What are the factors used to produce "People You May Know" data product on LinkedIn?

9) How will you find the second largest element in a Binary Search tree ? (Asked for a Data Scientist Intern job role)

➢Mu Sigma Data Science Interview Questions

1) Explain the difference between Supervised and Unsupervised Learning through examples.

2) How would you add value to the company through your projects?

3) Case Study-based questions – Cars are implanted with speed tracker so that the insurance companies can track about our driving state. Based on this new scheme what kind of business questions can be answered?

4) Define standard deviation, mean, mode and median.

5) What is a joke that people say about you and how would you rate the joke on a scale of 1 to 10?

6) You own a clothing enterprise and want to improve your place in the market. How will you do it from the ground level?

7) How will you customize the menu for Cafe Coffee Day?

➢Amazon Data Science Interview Questions

1) Estimate the probability of a disease in a particular city given that the probability of the disease on a national level is low.

2) How will inspect missing data and when are they important for your analysis?

3) How will you decide whether a customer will buy a product today or not given the income of the customer, location where the customer lives, profession and gender? Define a machine learning algorithm for this.

4) From a long sorted list and a short 4 element sorted list, which algorithm will you use to search the long sorted list for 4 elements.

5) How can you compare a neural network that has one layer, one input and output to a logistic regression model?

6) How do you treat colinearity?

7) How will you deal with unbalanced data where the ratio of negative and positive is huge?

8) What is the difference between -

i) Stack and Queue

ii) Linkedin and Array

➢Uber Data Science Interview Questions

1) Will Uber cause city congestion?

2) What are the metrics you will use to track if Uber's paid advertising strategies to acquire customers work? How will you figure out the acceptable cost of customer acquisition?

3) Explain principal components analysis with equations.

4) Explain the various time series forecasting techniques.

5) Which machine learning algorithm will you use to solve a Uber driver accepting request?

6)How will you compare the results of various machine learning algorithms?

7) How to solve multicollinearity?

8) How will you design the heatmap for Uber drivers to provide recommendations on where to wait for passengers? How would you approach this?

9) If we added one rider to the current SF market, how would that affect the existing riders and drivers?

10) What are the different performance metrics for evaluating Uber services?

11) How will you decide which version (Version 1 or Version 2) of the Surge Pricing Algorithms is working better for Uber?

12) How will you explain JOIN function in SQL to a 10-year-old?

➢Netflix Data Science Interview Questions

1) How can you build and test a metric to compare ranked list of TV shows or Movies for two Netflix users?

2) How can you decide if one algorithm is better than the other?

➢Microsoft Data Science Interview Questions

1) Write a function to check whether a particular word is a palindrome or not.

2) How can you compute an inverse matrix faster by playing with some computation tricks?

3) You have a bag with 6 marbles. One marble is white. You reach the bag 100 times. After taking out a marble, it is placed back in the bag. What is the probability of drawing a white marble at least once?

➢Apple Data Science Interview Questions

1) How do you take millions of users with 100's of transactions each, amongst hundreds of products and group the users together in meaningful segments?

➢Adobe Data Scientist Interview Questions

1) Check whether a given integer is a palindrome or not without converting it to a string.

2) What is the degree of freedom for lasso?

3) You have two sorted array of integers, write a program to find a number from each array such that the sum of the two numbers is closest to an integer i.

➢American Express Data Scientist Interview Questions

1) Suppose that American Express has 1 million card members along with their transaction details. They also have 10,000 restaurants and 1000 food coupons. Suggest a method which can be used to pass the food coupons to users given that some users have already received the food coupons so far.

2) You are given a training dataset of users that contain their demographic details, the pages on Facebook they have liked so far and results of psychology test based on their personality i.e. their openness to like FB pages or not. How will you predict the age, gender and other demographics of unseen data?

➢Quora Data Scientist Interview Questions

1) How will you test a machine learning model for accuracy?

2) Print the elements of a matrix in zig-zag manner.

3) How will you overcome overfitting in predictive models?

4) Develop an algorithm to sort two lists of sorted integers into a single list.

➢Goldman Sachs Data Scientist Interview Questions

1) Count the total number of trees in United States.

2) Estimate the number of square feet pizza's eaten in US each year.

3) A box has 12 red cards and 12 black cards. Another box has 24 red cards and 24 black cards. You want to draw two cards at random from one of the two boxes, which box has a higher probability of getting cards of same colour and why?

4) How will you prove that the square root of 2 is irrational?

5) What is the probability of getting a HTT combination before getting a TTH combination?

6) There are 8 identical balls and only one of the ball is slightly heavier than the others. You are given a balance scale to find the heavier ball. What is the least number of times you have to use the balance scale to find the heavier ball?

➢Walmart Data Science Interview Questions

1) Write the code to reverse a Linked list.

2) What assumptions does linear regression machine learning algorithm make?

3) A stranger uses a search engine to find something and you do not know anything about the person. How will you design an algorithm to determine what the stranger is looking for just after he/she types few characters in the search box?

4) How will you fix multi-collinearity in a regression model?

5) What data structures are available in the Pandas package in Python programming language?

6) State some use cases where Hadoop MapReduce works well and where it does not.

7) What is the difference between an iterator, generator and list comprehension in Python?

8) What is the difference between a bagged model and a boosted model?

9) What do you understand by parametric and non-parametric methods? Explain with examples.

10) Have you used sampling? What are the various types of sampling have you worked with?

11) Explain about cross-entropy ?

12) What are the assumptions you make for linear regression?

13) Differentiate between gradient boosting and random forest.

14) What is the significance of log odds?

➢IBM Data Science Interview Questions

1) How will you handle missing data?

➢Yammer Data Science Interview Questions

1) How can you solve a problem that has no solution?

2) On rolling a dice if you get $1 per dot on the upturned face,what are your expected earnings from rolling a dice?

3) In continuation with question #2, if you have 2 chances to roll the dice and you are given the opportunity to decide when to stop rolling the dice (in the first roll or in the second roll). What will be your rolling strategy to get maximum earnings?

4) What will be your expected earnings with the two roll strategy?

5) You are creating a report for user content uploads every month and observe a sudden increase in the number of upload for the month of November. The increase in uploads is particularly in image uploads. What do you think will be the cause for this and how will you test this sudden spike?

➢Citi Bank Data Science Interview Questions

1) A dice is rolled twice, what is the probability that on the second chance it will be a 6?

2) What are Type 1 and Type 2 errors?

3) Burn two ropes, one needs 60 minutes of time to burn and the other needs 30 minutes of time. How will you achieve this in 45 minutes of time?

➢Data Science Interview Questions Asked at Other Top Tech Companies

1) R programming language cannot handle large amounts of data. What are the other ways of handling it without using Hadoop infrastructure? (Asked at Pyro Networks)

2) Explain the working of a Random Forest Machine Learning Algorithm (Asked at Cyient)

3) Describe K-Means Clustering. (Asked at Symphony Teleca)

4) What is the difference between logistic and linear regression? (Asked at Symphony Teleca)

5) What kind of distribution does logistic regression follow? (Asked at Symphony Teleca)

6) How do you parallelize machine learning algorithms? (Asked at Vodafone)

7) When required data is not available for analysis, how do you go about collecting it? (Asked at Vodafone)

8) What do you understand by heteroscedasticity (Asked at Vodafone)

9) What do you understand by confidence interval? (Asked at Vodafone)

10) Difference between adjusted r and r square. (Asked at Vodafone)

11) How does Facebook recommend items to the newsfeed? (Asked at Finomena)

12) What do you understand by ROC curve and how is it used? (Asked at MachinePulse)

13) How will you identify the top K queries from a file? (Asked at BloomReach)

14) Given a set of web pages and changes on the website, how will you test the new website feature to determine if the change works positively? (Asked at BloomReach)

15) There are N pieces of rope in a bucket. You put your hand into the bucket, take one end piece of the rope . Again you put your hand into the bucket and take another end piece of a rope. You tie both the end pieces together. What is the expected value of the number of loops within the bucket? (Asked at Natera)

16) How will you test if a chosen credit scoring model works or not? What data will you look at? (Asked at Square)

17) There are 10 bottles where each contains coins of 1 gram each. There is one bottle of that contains 1.1 gram coins. How will you identify that bottle after only one measurement? (Data Science Puzzle asked at Latent View Analytics)

18) How will you measure a cylindrical glass filled with water whether it is exactly half-filled or not? You cannot measure the water, you cannot measure the height of the glass nor can you dip anything into the glass. (Data Science Puzzle asked at Latent View Analytics)

19) What would you do if you were a traffic sign? (Data Science Interview Question asked at Latent View Analytics)

20) If you could get the dataset on any topic of interest, irrespective of the collection methods or resources then how would the dataset look like and what will you do with it. (Data Scientist Interview Question asked at CKM Advisors)

21) Given n samples from a uniform distribution [0,d], how will you estimate the value of d? (Data Scientist Interview Question asked at Spotify)

22) How will you tune a Random Forest? (Data Science Interview Question asked at Instacart).

23) Tell us about a project where you have extracted useful information from a large dataset. Which machine learning algorithm did you use for this and why? (Data Scientist Interview Question asked at Greenplum)

24) What is the difference between Z test and T-test? (Data Scientist Interview Questions asked at Antuit)

25) What are the different models you have used for analysis and what were your inferences? (Data Scientist Interview Questions asked at Cognizant)

26) Given the title of a product, identify the category and sub-category of the product. (Data Scientist interview question asked at Delhivery)

27) What is the difference between machine learning and deep learning? (Data Scientist Interview Question asked at InfoObjects)

28) What are the different parameters in ARIMA models? (Data Science Interview Question asked at Morgan Stanley)

29) What are the optimizations you would consider when computing the similarity matrix for a large dataset? (Data Science Interview questions asked at MakeMyTrip)

30) Use Python programming language to implement a toolbox with specific image processing tasks.(Data Science Interview Question asked at Intuitive Surgical)

31) Why do you use Random Forest instead of a simple classifier for one of the classification problems? (Data Science Interview Question asked at Audi)

32) What is an n-gram? (Data Science Interview Question asked at Yelp)

33) What are the problems related to Overfitting and Underfitting and how will you deal with these ? (Data Science Interview Question asked at Tiger Analytics)

34) Given an MxN dimension matrix with each cell containing an alphabet, find if a string is contained in it or not. (Data Science Interview Question asked at Tiger Analytics)

35) How do you "Group By" in R programming language without making use of any package? (Data Scientist Interview Question asked at OLX)

36) List 15 features that you will make use of to build a classifier for OLX website. (Data Scientist Interview Question asked at OLX)

37) How will you build a caching system using an advanced data structure like hashmap? (Data Scientist Interview Question asked at OLX)

38) How to reverse strings that have changing positions? (Data Scientist Interview Question asked at Tiger Analytics)

39) How do you select a cricket team? (Data Scientist Interview Question asked at Quantiphi)

40) What is the difference between trees and random forest? (Data Scientist Interview Question asked at Salesforce)