365 DataScience

164 Data Science Interview Questions & Answers

Iliya Valchanov
Ned Krastev
Ellie Kaloyanova
Martin Ganchev

INTRODUCTION BY THE AUTHORS

Dear 365 Readers,

We wrote this book to help you master the art of interviewing for a data science position. From job-specific technical questions to tricky behavioral inquires and unexpected brainteasers and guesstimates, we will prepare you for any job candidacy in the field – data scientist, data analyst, BI analyst, data engineer or data architect.

Data Science Interview Questions and Answers is the result of our data science expertise, direct experience interviewing at companies, and countless conversations with job candidates.

Its goal is to teach by example - not only by giving you a list of interview questions and their answers, but also by sharing the techniques and thought processes behind each question and the expected answer. Once you read it, you'll have all the knowledge and tools to succeed during the data science interview.

How to Use This Book for Best Results? Award yourself with enough time to work through the questions. This way, you'll really understand what they are asking and what information you should highlight for the best response. If studied well, this book will enhance both your technical and communication skills.

We are excited for your data science journey to begin! Do your very best, practice, and best of luck!

Iliya Valchanov
Co-Founder and Instructor at 365 Data Science

Ellie Kaloyanova
Python, R, and ML Instructor

Ned Krastev
Co-Founder and Instructor at 365 Data Science

Martin Ganchev
Python and SQL Instructor

CONTENTS

How to prepare for data science interview questions? 4

Data Scientist Interview Questions 5

- General Data Scientist Interview Questions 6
- Technical Data Scientist Interview Questions 19
- Behavioral Questions 25
- Brainteasers 31
- Guesstimate 32

Data Analyst Interview Questions 33

- General Data Analyst Interview Questions 34
- Technical Data Analyst Interview Questions 41
- Behavioral Questions 45
- Brainteasers 51
- Guesstimate 52

BI Analyst Interview Questions 53

- General BI Analyst Interview Questions 54
- Technical BI Analyst Interview Questions 59
- Behavioral Questions 65
- Brainteasers 69
- Guesstimate 70

Data Engineer Interview Questions 71

- General Data Engineer Interview Questions 72
- Technical Data Engineer Interview Questions 79
- Behavioral Questions 84
- Brainteasers 88
- Guesstimate 89

Data Architect Interview Questions 90

- General Data Architect Interview Questions 91
- Technical Data Architect Interview Questions 96
- Behavioral Questions 102
- Brainteasers 107
- Guesstimate 108

About the Authors 110

365 Data Science Courses 111

HOW TO PREPARE FOR DATA SCIENCE INTERVIEW QUESTIONS?

If you want to successfully land a job in data science, knowing your stuff and putting it in a neat package with an impressive CV, an outstanding portfolio, and a flashy resume will only get you halfway through the door.

What will open it is understanding the whole data science interview process and how to navigate it smoothly – from seeing that job posting to closing the deal with a welcome-to-the-team handshake.

And with this in-depth resource, we're going to show you how to get there.

We've prepared a collection of 180 straight-to-the-point data science questions paired with their answers categorized by career paths – data scientist, data analyst, BI analyst, data engineer, and data architect.

We'll guide you all the way:

- from the general interview questions that you need to make a great first impression;
- through the fundamental technical part to
- to the behavioral questions, brainteasers, and guesstimates that will help you sign that contract.

Practically, everything you need to know about all levels of preparation. And those are the insights that will ultimately help you get the job you want and you're qualified for.

So let's dive right in.

DATA SCIENTIST
INTERVIEW QUESTIONS

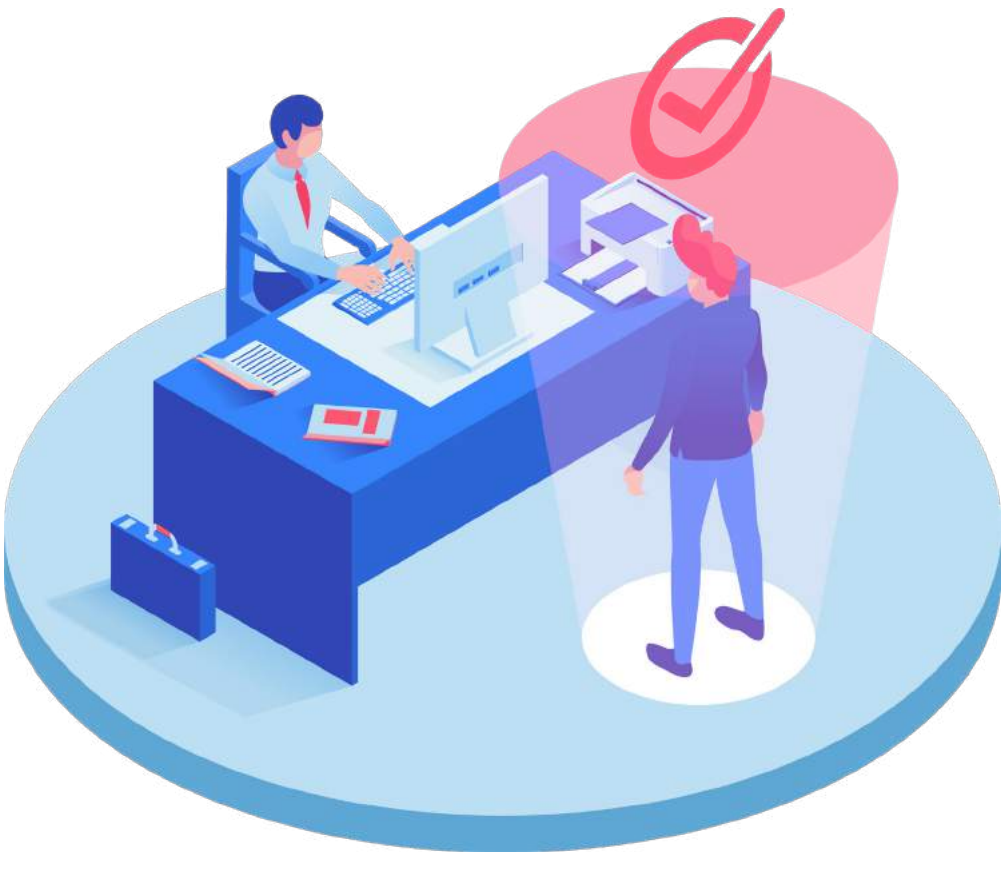
Data scientist interview questions cover a wide scope of multidisciplinary topics. That means you can never be quite sure what challenges the interviewer(s) might send your way.

That said, being familiar with the type of questions you can encounter is an important aspect of your preparation process.

Below you'll find examples of real-life questions and answers. Reviewing those should help you assess the areas you're confident in and where you should invest additional efforts to improve.

GENERAL DATA SCIENTIST INTERVIEW QUESTIONS

Here are a few examples of warm-up questions that will get you ready for the more in-depth inquiries ahead...



1. Tell me about yourself.

This will probably be the very first question of the interview. A very generic question, which is tougher than it sounds. You need to avoid telling the story of your life, but you don't want to pause after three sentences either. Given that it is the opening question of the interview, your answer becomes even more important, as it sets the tone for the rest of the conversation. The Hiring Manager is interested in seeing whether you will strongly structure the answer to a very broad question.

The secret for responding well to this question is scripting and practicing before every interview. What should you include in your response?

- Tell the interviewer only facts that you want him/her to know
- Give a hint about your personal life with one or two sentences (*"I was born and raised in the UK", "I moved to New York because it is a vibrant city and I like the dynamic environment."*)
- Show that you are a perfect fit for the job under consideration, you have the right education, and your previous work experiences will be a valuable asset to the firm;
- Conclude by explaining why you are excited about this possibility and how your strengths match with the profile that the company is looking for.

Prepare a script that addresses each of the points above and practice answering the "Tell me about yourself" question, as you know it's coming your way once an interview starts.

2. What relevant work experience do you have?

A straightforward question, which leaves little space for maneuvering. Make sure that you are well prepared before the interview. Carefully study the job description and identify how your work experience is going to be useful in handling the responsibilities at this new position. Try to be specific and point out, which are the activities that you learned to do in your previous jobs and how they would allow you to perform well at this new position.

3. Where do you see yourself in 5 years?

A potentially dangerous question. The interviewer wants to know whether the company would be able to count on you in the long run - whether you are looking for a job to tide you over or for a career. Besides hiring someone that is qualified and skilled, most firms want to choose a person that believes in a future with the company. They don't want to invest a great deal of time and money in order to recruit and train someone who will leave in two years.

The hiring manager wants to understand exactly how you think. Perhaps you intend to gain one or two years of practical work experience and join your family's business. Maybe you want to start your own company, or maybe you believe that one or two years at this job would allow you to pursue much more interesting opportunities with other companies. This is why you need to be prepared and have a good answer in mind.

Instead of replying where you will be in 5 years, which is kind of dangerous for the above-mentioned reasons, you can talk about exactly what you would like to learn in the next five years. You can say that you want to become very good at what you do, gain hands-on practical experience in managing people and that you always wanted to become a technical expert in the field for which you are interviewing.

As a closing statement, you can add that you are excited about this opportunity because you believe that it is a step in the right direction and would allow you to achieve your goals.

By spinning the question in this direction, you are able to achieve three things. First, you protect yourself from answering a potentially dangerous question. Second, you will be able to emphasize that the main driver in your career is professional growth and self-improvement. And third – you are able to affirm that you are excited about the job opportunity. Sounds good, right?

Similar versions of this question are *"What do you want to achieve in your career?", "Describe your ideal job", "What are your long-term career goals?"* The same logic applies to all of these questions too.

4. How do data scientists use statistics?

According to Iliya, co-founder of 365 Data Science, "An answer like 'Data scientists use statistics in almost everything they do' would be good enough for me if I was interviewing you. However, keep in mind that this is a very tricky question. Not because it is hard to answer – on the contrary. But sometimes the question is not asked for the answer itself, but rather for the way you structure your thought process and express an idea.

Therefore, assuming you got asked this question, you'd need to maintain your composure and structure a nice-sounding answer. One of the better ways to achieve that is to frame the question within a framework.

Here's one possible answer. Certainly, you can work on creating a smaller version of it as you also don't want to bore your interviewer.

'If we think about data science as a field, we can identify several pillars it is built upon: Mathematics, Probability, Statistics, Economics, Programming, Data visualization, Machine learning and modeling in general, etc. Now, we could simplify this framework by ignoring Mathematics as a pillar, as it is the basis of every science. Then we could assume probability is an integral part of statistics and continue simplifying further until reaching three fairly independent fields: Statistics, Economics, and Programming. Programming is just a tool for materializing ideas into solutions. Economics, on the other hand, is more about the 'business thinking' about a problem. Therefore, all of a data scientist's work boils down to statistics.

One could argue that Machine learning is a separate field, but it is actually an iterative, programmatically efficient application of statistics.

Models such as linear regression, logistic regression, decision trees, etc., are all developed by statisticians. Their predictions are nothing more than statistical inferences based on the original distributions of the data and making assumptions about the distribution of the future values.

Deep learning? Well, one of the most common methods for backpropagation is called: 'Stochastic gradient descent' and the word 'stochastic' is a probabilistic term, therefore, falling within the field of statistics.

Data visualizations also could fall under the umbrella of descriptive statistics. After all, a visualization usually aims to describe the distribution of a variable or the interconnection of several different variables.

One notable exception is data preprocessing. That's an activity which is mainly related to programming and often does not require statistical knowledge. That's why data engineers and data architects exist. They need not be proficient in statistics – that's the data scientist's job. Finally, there is an exception to the exception – statistical data preprocessing. Here we've got creation of dummy variables, feature scaling, regularization, and so on. While preprocessing tasks in their execution, they require solid statistical knowledge.'

5. What's the difference between SAS, R, And Python Programming?

SAS is one of the most popular analytics tools used by some of the biggest companies in the world. It has great statistical functions and graphical user interface. However, it is too pricey to be eagerly adopted by smaller enterprises or individuals.

R, on the other hand, is a robust tool for statistical computation, graphical representation, and reporting.

The best part about R is that it is an Open Source tool. As such, both academia and the research community use it generously and update it with the latest features for everybody to use.

In comparison, Python is a powerful open-source programming language. It's intuitive to learn and works well with most other tools and technologies. Python has a myriad of libraries and community created modules. Its functions include statistical operation, model building and many more. The best characteristic of Python is that it is a general-purpose programming language so it is not limited in any way.

6. What is the difference between WHERE and HAVING clause in SQL?

Adding a WHERE clause to a query allows you to set a condition which you can use to specify what part of the data you want to retrieve from the database.

HAVING is a clause frequently implemented with GROUP BY because it refines the output from records that do not satisfy a certain condition.

HAVING needs to be inserted between the GROUP BY and ORDER BY clauses. In a way, HAVING is like WHERE but applied to the GROUP BY block.

On some occasions, an identical result could be obtained by implementing the same condition, either with the WHERE or with the HAVING clause.

The main distinction between the two clauses is that HAVING can be applied for subsets of aggregated groups, while in the WHERE block, this is forbidden. In other words, after HAVING, you can have a condition with an aggregate function, while WHERE cannot use aggregate functions within its conditions.

7. What is a Normal distribution?

A distribution is a function that shows the possible values for a variable and how often they occur.

To answer this question, you are likely to need to first define what a distribution is.

So, in statistics, when we use the term distribution, we usually mean a probability distribution. Here's one definition of the term:

A Normal distribution, also known as Gaussian distribution, or The Bell Curve, is probably the most common distribution. There are several important reasons:

- It approximates a wide variety of random variables
- Distributions of sample means with large enough sample sizes could be approximated to Normal, following the Central Limit Theorem

- All computable statistics are elegant (they really are!!!)
- Decisions based on Normal distribution insights have a good track record.

What is very important is that the Normal distribution is symmetrical around its mean, with a concentration of observations around the mean. Moreover, its mean, median and mode are the same. Finally, you should get an extra point if you mention that 95% of the data points from a Normal distribution are located within 2 standard deviations from the mean, and 99.7% of the data points are located within 3 standard deviations from the mean.

Now, you may be also expected to give an example.

Since many biological phenomena are normally distributed it is going to be the easiest to turn to a biological example. Try to showcase all facts that you just mentioned about a Normal distribution.

Let focus on the height of people. You know a few people that are very short and a few people that are very tall. You also know a bit more people that are short but not too short, and approximately an equal amount that are tall, but not too tall. Most of your acquaintances, though have a very similar height, centered around the mean height of all the people in your area or country. There are some differences which are mainly geographical, but the overall pattern is such.

8. How are missing values and impossible values represented in R?

One of the main issues, when working with real data is handling missing values. These are represented by NA in R. Impossible values (division by 0, for example) are represented by NAN (not a number).

9. What is an example of a dataset with a non-Gaussian distribution?

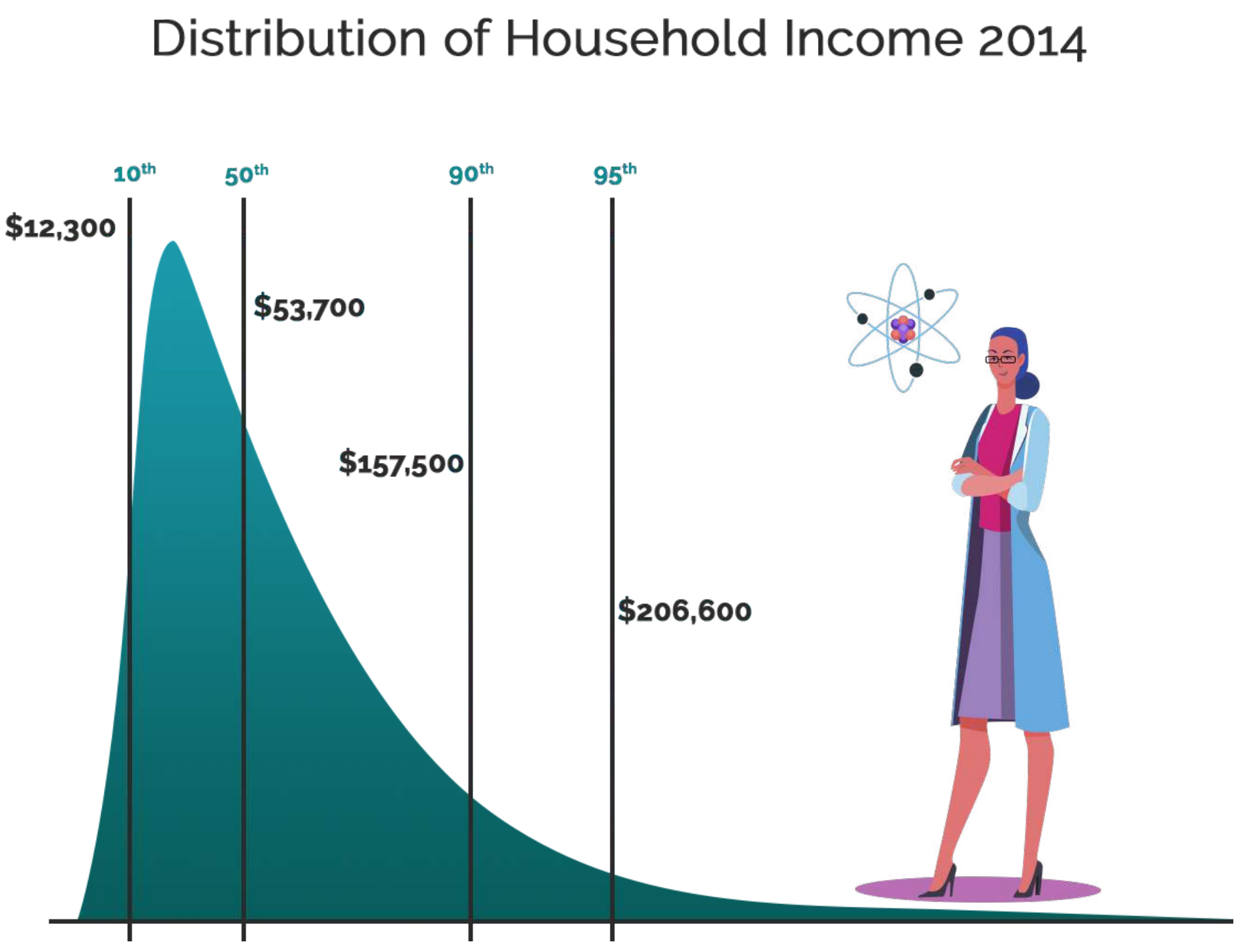
First, it may make sense to research what is a Gaussian distribution. In fact, it is also known as 'Normal distribution' or 'The Bell Curve'. For further information, please refer to the question 'What's Normal Distribution?'

Once you are sure that you know what a Gaussian distribution is, we can proceed to the question at hand.

We established that for a distribution to be non-Gaussian, it shouldn't follow the normal distribution. One of the main characteristics of the normal distribution is that it is symmetric around the mean, the median and the mode, which all fall on one point. Therefore, all we have to do is to select a distribution, which is not symmetrical, and we will have our counterexample.

One of the popular non-Gaussian instances is the distribution of the household income in the USA .

You can see where the 50th percent line is, but that is not where the mean is. While the graph is from 2014, this pattern of inequality still persists and even deepens in the United States. As such, household income in the US is one of the most commonly quoted non-Gaussian distributions in the world.



10. R has several packages for solving a particular problem.

How do you decide which one is best to use?

R has extensive documentation online. There is usually a comprehensive guide for the use of popular packages in R, including the analysis of concrete data sets. These can be useful to find out which approach is best suited to solve the problem at hand.

Just like with any other script language, it is the responsibility of the data scientist to choose the best approach to solve the problem at hand. The choice usually depends on the problem itself or the specific nature of the data (i.e., size of the data set, the type of values and so on).

Something to consider is the tradeoff between how much work the package is saving you, and how much of the functionality you are sacrificing.

It bears also mentioning that because packages come with limitations, as well as benefits, if you are working in a team and sharing your code, it might be wise to assimilate to a shared package culture.

11. What are interpolation and extrapolation?

Sometimes you could be asked a question that contains mathematical terms. This shows you the importance of knowing mathematics when getting into data science. Now, interpolation and extrapolation are two very similar concepts. They both refer to predicting or determining new values based on some sample information.

There is one subtle difference, though.

Say the range of values we've got is in the interval (a, b). If the values we are predicting are inside the interval (a, b), we are talking about interpolation (inter = between). If the values we are predicting are outside the interval (a, b), we are talking about extrapolation (extra = outside).

Here's one example.

Imagine you've got the number sequence: 2, 4, _, 8, 10, 12. What is the number in the blank spot? It is obviously 6. By solving this problem, you interpolated the value.

Now, with this knowledge, you know the sequence is 2, 4, 6, 8, 10, 12. What is the next value in line? 14, right? Well, we have extrapolated the next number in the sequence.

Finally, we must connect this question with data science a bit more. If they ask you this question, they are probably looking for you to elaborate on that.

Whenever we are doing predictive modeling you will be trying to predict values – that's no surprise. Interpolated values are generally considered reliable, while extrapolated ones – less reliable or sometimes invalid. For instance, in the sequence from above: 2, 4, 6, 8, 10, 12, you may want to extrapolate a number before 2. Normally, you'd go for '0'. However, the natural domain of your problem may be positive numbers. In that case, 0 would be an inadmissible answer.

In fact, often we are faced with issues where extrapolation may not be permitted because the pattern doesn't hold outside the observed range, or the domain of the event is … the observed domain. It is

extremely rare to find cases where interpolation is problematic. Please bear in mind that last bit and don't forget to mention it in the interview!

12. What is the difference between population and sample in data?

A population is the collection of all items of interest to our study and is usually denoted with an uppercase N. The numbers we've obtained when using a population are called parameters.

A sample is a subset of the population and is denoted with a lowercase n, and the numbers we've obtained when working with a sample are called statistics.

That's more or less what you are expected to say.

Further, you can spend some time exploring the peculiarities of observing a population. Conversely, it is likely that you'll be asked to dig deeper into why in statistics we work with samples and what types of samples are there.

In general, samples are much more efficient and much less expensive to work with. With the proper statistical tests, 30 sample observations may be enough for you to take a data-driven decision.

Finally, samples have two properties: randomness and representativeness. A sample can be one of those, both, or neither. To conduct statistical tests, which results you can use later on, your sample needs to be both random and representative.

Consider this simplified situation.

Say you work in a firm with 4 departments: IT, Marketing, HR, and Sales. There are 1000 people in each department, so a total of 4000 people. You want to evaluate the general attitude towards a decision to move to a new office, which is much better on the inside but is located on the other side of the city.

You decide you don't really want to ask 4000 people, but 100 is a nice sample. Now, we know that the 4 groups are exactly equal. So, we expect that in those 100 people, we would have 25 from each department.

1) We pick 100 people (out of the 4000) at random and realize that we have 30 IT, 30 Marketing, 30 HR, and 10 from Sales. Obviously, the opinion of the Sales department is underrepresented. We have a sample, which is random but not representative.

2) I've been working in this firm for quite a while now, so I have many friends all over it. I decide to ask the opinion of my friends from each department because I want them to feel comfortable in the workplace. I pick 25 people from each department. The sample is representative but is not random.

In the first case, we have underrepresented some group of people. In the second case, we've made a decision based on a specific circle of people and not the general 'public'.

If I want it to be random and representative, I will pick 25 people from IT at random, then 25 people from Marketing at random, same for HR and Sales.

In this way, all groups will be represented, and the sample will be random.

You can decide to skip that detailed explanation, or better – ask them if they want you to dive deeper into the topic and then impress them with your detailed understanding!

13. What are the steps in making a decision tree?

First, a decision tree is a flow-chart diagram. It is extremely easy to read, understand and apply to many different problems. There are 4 steps that are important when building a decision tree.

1) Start the tree. In other words, find the starting state – maybe a question or idea, depending on your context.

2) Add branches. Once you have a question or an idea, it branches out into 1,2, or many different branches.

3) Add the leaves. Each branch ends with a leaf. The leaf is the state which you will reach once you have followed a branch.

4) Repeat 2 and 3. We then repeat steps 2 and 3, where the starting points are the leaves, until we finish-off the tree. In other words, every question and possible outcome should be included.

Depending on the context you may be expected to add additional steps like: complete the tree, terminate a branch, verify with your team, code it, deploy it, etc.

However, these 4 steps are the main ones in creating a decision tree. Whether to include these extra steps really depends on the position you are applying for.

If you are applying for some data science project management position, you may be expected to say: 'Validate with all stakeholders to ensure the quality of the decision tree'.

If you are applying for a data scientist position, you may be expected to explain a bit more about the programming language and library you intend to use. This also includes the reason why you'd choose that library.

14. How is machine learning deployed in real-world scenarios?

This question is a bit tricky. Model deployment is a part of a data science job, but in fact, efficient model deployment is more often related to engineering, software development, cloud computing, etc. In other words, to make sure everything is right, you'd better turn to your IT department or hire a computer scientist in your team.

Now, there are 3 important steps:

1) Once you train a model, you should save it, or better – store it in a file. There are different ways in which this could be achieved. The general 'Pythonic' ways are through pickle or joblib. However, libraries such as TensorFlow deal with much more complicated model objects and thus they offer ad-hoc functions for deployment. Often they look like this: ***.save('filename').***

This part of the process is always done by the data scientist, ML engineer, or whoever is in charge of the model training.

2) ***Computing instance.*** AWS and Microsoft Azure offer computing instances or cloud-based environments that can run the model you've just created. Surely, you can share the file with your colleagues through email or Messenger, but more often, there will be some cloud that handles the deployment.

The computing instance should be set-up to communicate with all other systems that feed the inputs and/or require the outputs of the model.

3) ***Job scheduler.*** Having a model and a place to run it, you can specify when and how to run it. That could be once a week, once per day, or every time an event occurs (e.g. a transaction, new user registration, etc.). At the desired time, new data would be taken, loaded, cleaned, preprocessed, fed to the model, etc. until you reach the desired outcome.

Having completed these 3 steps, you are practically done.

You will have a model, running on some cloud at prescheduled times. Once you've got an output, you can return it to a Python notebook, or better connect it to yet another system (that could be considered a part of 2.). Depending on your needs, it could be a web app (e.g. a recommender system gives information about a particular customer and shows them relevant results), or some kind of visualization software such as Tableau or PowerBI which would analyze your data in real time.

Needless to stress, 2. and 3. would rarely be a data scientist's primary job. Still, in a smaller team, that may fall on them, too!

15. What is K-means clustering? How can you select K for K-means?

The main goal of clustering is to group individual observations so that the observations from one group are very similar to each other. In addition, we'd like them to be very different from the observations in other groups. There are two main types of clustering: flat and hierarchical. Hierarchical clustering is much more spectacular because of the dendrograms we can create, but flat clustering techniques are much more computationally efficient. Therefore, we usually opt for the latter.

K-means clustering is the most prominent example of flat clustering.

It consists of finding K clusters, given their mean distance from the centers of the clusters. K stands for the number of clusters we are trying to identify. This is a value, selected prior to the clustering.

Now, the optimal number of clusters is obviously what we are usually interested in.

There are several ways to approach that, but the most common one is called: 'The Elbow Method'.

There, we solve the clustering problem with 1, 2, 3, 4, 5, 6 and so on number of clusters. We then plot them on a graph where on the x-axis we've got the number of clusters, while on the y-axis, the WCSS (within cluster sum of squares). The resulting image resembles a human elbow. The place where the kink is signifies the optimal clustering solution. And that's how you choose the 'K' in K-means!

16. What are the disadvantages of a linear model?

This is one of the strangest questions you could be asked. It is like being asked: 'what are the disadvantages of playing tennis barefoot?' You don't need shoes to play tennis, but it is much better if you do.

Now, the most common linear models are the linear regression model and linear time series model. Therefore, let's answer the question in that context.

The single biggest advantage of a linear model is that it is simple. From there, there are mainly disadvantages and limitations.

Therefore, let's focus on the top 3 cons of using a linear model.

1) Linear model implies linear relationships.

A linear model assumes that the independent variables explain the dependent one(s) in a linear way, e.g. a = bx + c. No powers, exponents, logarithms, etc. are allowed.

Obviously, this is a great simplification – the real world is not linear. Using a linear model, would either disregard some patterns or force us to execute complicated transformations to reach a linear representation.

2) Data must be independent.

In the general case, that's not always true, but in 95+% of the linear models conducted in practice – it is. Most linear models assume that the variables in the model are not collinear. Alternatively, we observe multicollinearity or the math behind the model estimation 'brakes'. Assuming that the variables are independent is obviously a very brave statement especially because we are limited to a linear relationship (if we had exponents and logarithms, the probability that they are collinear would drop dramatically).

3) Outliers are a big, big issue.

Since linear models assume linearity, having values that are too big, or too small regarding any feature may be devastating for the model. All points are expected to be close to some line, which as you can imagine is rather unrealistic. To deal with that we often complicate the linear model in ways that practically make it behave like a non-linear one.

17. Describe a time when you were under pressure.

Do you know the saying "When the going gets tough, the tough get going?"

Every Hiring Manager wants to make sure you can handle the pressure of the job. Are you someone who is likely to abandon the boat when things get a little tough? Every firm needs people that are reliable. All jobs involve a certain element of pressure; some more than others, obviously. Your task here is to give an example of a stressful situation and show how you coped with it.

Here's an example of such a situation:

I was under significant pressure before taking my GMAT exam. I needed a really good grade in order to be admitted to the graduate school that I am now graduating from. A few weeks before the exam, I noticed that I was becoming nervous. Two things helped me handle the pressure much better; I started sleeping for at least 7 hours (going to bed earlier in the evening) and I dedicated at least one hour a day to sports activities. This had a hugely positive impact on my concentration and stress level.

18. How would you add value to our company?

Did you see "The Wolf of Wall Street"? Remember Jordan Belfort's famous quote "sell me this pen"? The same principle applies to this question as well, although instead of selling a pen, you need to sell the idea of you landing the job. This is what the recruiter is asking you to do. You need to convince him/her that you will add value to the company. But, how are you going to be able to tell how you would add value to the company before having worked for the company?

Most candidates will start by listing their qualifications, work experience, personal traits, achievements, and they will be hoping to push the right button, somewhere along the way.

Similarly, when facing the "sell me this pen" task, most people start describing the pen's attributes; it is a great pen, writes very well, it is very shiny and smooth, etc.

It is natural to focus on your qualities and qualifications when asked how are you going to add value to the company.

However, this is a trap.

Most people would do just that. They will explain that they are great and that they are qualified. But that fails to answer the question itself, right? How are you going to add value? Analogically, the person who is being sold a pen can ask "Why do I need this pen?" Instead of falling for this trap and responding like everybody else, you can instead show that you are different by using an alternative approach.

Turn this into a back and forth dialogue and figure out what value needs to be added to the team that you will be joining.

What does the company need? Are there any supplementary skills that are missing? Is there a particular area that they would like to reinforce? Learn more about the Interviewer's take on the current situation and understand precisely what is expected

from you. Don't be shy to ask about the company's mid-term strategy and the type of people that they will need in the future. Then you can nail the question by pointing out how your qualifications and motivation match with the needs that they have.

The whole dynamic of this type of question is driven by the fact that before you are able to sell a pen, you have to know more about the person who is going to buy it, what are his needs and what kind of pens is he usually writing with. Once you have positively identified a need, you can point out that your product is the right solution for that need.

TECHNICAL DATA SCIENTIST INTERVIEW QUESTIONS

Statistics, programming, machine learning – those are all bound to come up at a certain point in the data scientist interview process. Here's a list of technical questions you can use for practice.

19. How do you explain Random Forrest to a non-technical person?

Random Forest is a classification algorithm. Its main purpose is to match a specific observation with its observed outcome.

An important defining characteristic of a random forest is that it is simply a collection of decision trees. There are many terms involved, but in fact, the concept is rather simple and could be easily illustrated with an example.

Let's say you want to create a meeting. A decision tree for that meeting may be:

Monday

o No

o Yes

1PM to 2PM

o No

o Yes

Room 160

o No

o Yes

3PM to 4PM

o No

o Yes

Room 155

o No

o Yes

And so on..

Based on this tree, we would normally estimate probabilities to have the meeting in one place or another.

The main issue is that this is a very bad classifier. However, combining many such trees we reach a random forest. The underlying assumption is that many bad classifiers equal a good classifier. Each tree makes a prediction (which observation to put in what class) and then the class with the most "votes" across all trees will be our random forest prediction.

20. What's wrong with training and testing a machine learning model on the same data?

This is one of the more common questions. When we are training a model, we are exposing it to the 'training data'. This means it is learning the patterns from it. By the end of the training, it becomes very good at predicting this particular dataset. However, sometimes we may overfit. This is a situation where we keep improving the accuracy, but not because the model is good, but just because it has learned every little detail about the data it is given.

If we test on that data, we will be checking the accuracy of the training. This is not a test per se. That's simply a 'train accuracy check'. Our model will seem to be very accurate and working properly, but that is because we trained it on that same data. We are essentially asking the model to predict what was already predicted, which is not a hard task.

To truly test a model, we must expose it to data it has never seen before. This will reveal if it learned the patterns of the population, or simply the noise in the training data.

21. How to make sure you are not overfitting while training a model?

First, we need to clarify what overfitting is exactly. Usually, overfitting happens when your model fits the training data so well that it misses the point. In other words – it doesn't look for the general patterns, but for the noise in the data provided. If that happens, when provided with new data, the model behaves disastrously in a real-life setting.

Regularization - In the context of machine learning refers to the process of modifying a learning algorithm so as to make it simpler often to prevent overfitting or to solve a badly posed problem.

- Early stopping – early stopping is the most common type of regularization. It is designed precisely to prevent overfitting. It consists of techniques that interrupt the training process, once the model starts overfitting.

  * Here you may be expected to say 'validation' or 'cross-validation'. In fact, early stopping methods always use the outputs from the validation to determine whether to stop the training process.

- Feature selection – for some models, having useless input features leads to much worse performance. Therefore, you have to make sure to choose only the most relevant features for your problem otherwise this may affect (among other things) overfitting.

• Ensembling. Ensembles are methods to combine several base models in order to produce one optimal predictive model. A good example of the ensemble method is Random Forest (a collection of decision trees).

It is very important to realize that overfitting is an extremely important issue. Every model will overfit if no preventative techniques have been implemented. Therefore, you should always aim to apply one or more of these techniques in your model building efforts.

22. What is cross-validation? How to do it right?

Cross-validation refers to many model validation techniques that use the same dataset for both training and validation. Usually, it is on a rotational basis so that observations are not overexposed to the training process and thus can serve as better validation. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice.

Why do we even need to validate?

Well, when you use sample data (so most of the time), you need to make sure that your model is not overfitting the parameters.

So how do we validate? We take out like 10% of the data for later use and train on the remaining 90%. Once we are done, we validate on the 10% we set aside at the beginning. This is a pretty common practice but has one major drawback - some of the data (these 10% precisely) is not really utilized in the training process.

That is where cross-validation comes in.

Cross-validation does the same thing as simple validation, but it first divides the dataset into equal parts (5,10,20 depending on the size of data). To cross-validate, it sets aside the first part and trains on the remaining parts. Then it sets aside the 2nd part and trains on the remaining ones (this time, including the first part). We continue in that way, utilizing a different subset for each validation. In that way, the model gets exposed to all the data in contrast to conventional validation.

23. How do you create a table in R without using external files?

This is practical knowledge that can be tested with a coding task, but it's possible you are asked this question as a stand-alone. In that case, you can ask about the use case for the numbers you're generating.

First, if you need a data table for the sake of having data to test on, you can just use one of R's preloaded datasets. You can access the list by calling data().

If you'd like to still create a table from scratch, you can use any of the random generator functions in R to generate random numbers according to a distribution, and store them in a matrix or a data frame. The functions are:

- runif()
- rnorm()
- rbinom()
- rexp()

You can also use sampling with or without replacement to generate your data and populate a table.

If you need an empty table to be filled out later, you can initiate empty vectors and create your data frame.

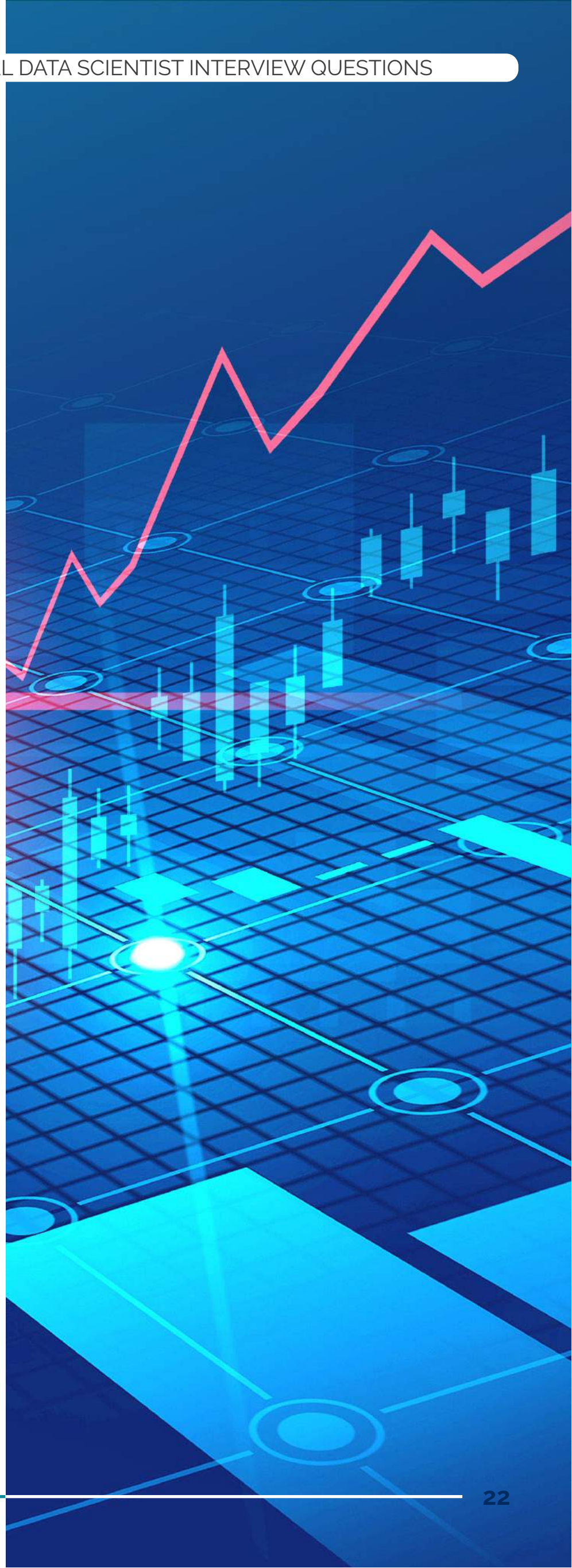
Finally, you could use an interactive method, for a quick solution to a small-scale problem. Creating an empty data frame and editing it with edit(df) will toggle an interactive spreadsheet view you can manually populate.

24. Explain the significance of Transpose in R.

Transpose is one of the simplest ways you can reshape a data structure in R. If you transpose a data frame or a matrix, you will essentially be rotating the data, so rows become columns, and vice versa.

In terms of use cases, transposing is sometimes needed to tidy data for analysis. If the raw format has observations recorded as columns, transposing the data structure would ensure the data is keeping to the convention whereby observations are organized in rows and variables are represented by columns. Transposing is also necessary for matrix multiplication used vastly in machine learning, deep learning, etc.

The t() function is the default way to transpose in R. If this simple function fits your needs, you don't need anything else. If the data you're trying to reshape is messier, {dplyr} and {tidyr} can provide a good set of functions to deal with it – grouping, mutating, pivoting...



25. Why would you use a Null as a data value?

To answer this question, it's important not to confuse a NULL value with the value of 0 or with a "NONE" response. Instead, think of a null value as a missing value. 0 or "NONE" could be values assigned by the user, while "NULL" is a value assigned by the computer if the user has provided no value for a given record.

Consider the Customers table below.

If you know John McKinley has filed 0 complaints, then in the "number of complaints" column in the "Customers" table, you could insert 0. This doesn't mean the value is null – not at all! You have a value of zero, and the information in this field for this first record is not null. It means John has filed no complaints.

If we have no information regarding the number of complaints John has filed, then the value would have been null.

By the same logic, imagine there was an additional column, called "Feedback", and that it is optional. If the first three customers have provided some feedback, while Catherine has said she didn't want to leave any, does that mean this value is null? No, because Catherine didn't want to provide any feedback, so we could mark her response as "NONE".

If she hasn't replied yet, only then our value would have been null.

Customers

Customer ID	First name	Last name	Email adress	Number of complaints
c_1	John	McKinley	john.mckinley @365datascience.com	0
c_2	Elizabeth	McFarlane	e.mcfarlane @365datascience.com	2
c_3	Kevin	Lawrence	kevin.lawrence @365datascience.com	1
c_4	Catherine	Winnfield	c.winnfield @365datascience.com	0

26. What is a primary key and a foreign key?

A primary key is a column (or a set of columns) whose value exists and is unique for every record in a table. It's important to know that each table can have one and only one primary key.

Therefore, you can think of a primary key as the field (or group of fields) that identifies the content of a table in a unique way. For this reason, the primary keys are also called the unique identifiers of a table.

Another crucial feature of primary keys is they cannot contain null values. This means, in an example with a single-column primary key, there must always be a value inserted in the rows under this column. You cannot leave it blank.

One last remark about primary keys - not all tables you work with will have a primary key, although almost all tables in any database will have a single-column or a multi-column primary key.

A foreign key, instead, is a column (or a set of columns) that references a column (most often the primary key) of another table. Foreign keys can be called identifiers, too, but they identify the relationships between tables, not the tables themselves.

In the relational schemas form of representation, relations between tables are expressed in the following way – the column name that designates the logical match is a foreign key in one table, and it is connected with a corresponding column from another table. Often, the relationship goes from a foreign key to a primary key, but in more advanced circumstances, this will not be the case. To catch the relations on which a database is built, we should always look for the foreign keys, as they show us where the relations are.

27. Describe a parent-child relationship in the context of a relational database.

Remember the function of a foreign key (see above)? It points to a column of another table and, thus, links the two tables. It is a field or collection of fields from one table, called the child table, and it refers to a column in another table, called the parent table. Usually, the column or the set of columns in the parent table is the primary key of that table. (The child table can also be called the referencing table, and the parent table can be called the referenced table.)

28. Given a table with duplicate data, how would you extract only specific rows based on business requirements provided?

In most cases, the tools form the Data Manipulation Language (DML) will allow you to do that. Usually, you could either use a SELECT DISTINCT statement to select distinct rows only or apply a GROUP BY clause to a join to filter the data in the desired way.

BEHAVIORAL QUESTIONS

This type of questions has become increasingly important in the hiring process. The reason is they help employers assess if your personality and motivations make you the right fit for the job. Most of them are centered around your behavior in similar past work situations.

29. What motivates you about this position?

By asking this question, the recruiter wants to understand whether you are excited about the new opportunity that lies ahead of you. Your enthusiasm, of course, is highly correlated with the amount of effort you will put once the job is offered.

A motivated person would try to be proactive and create a positive working environment, which is precisely what every company needs. The real question isn't whether you should say that you are motivated. Of course, you should. You need to think of a way that would best show that you are genuinely interested in the position under consideration.

There are a lot of different things that can motivate you:

- The learning opportunities that you will have on the job
- Future growth prospects
- You like the team that you will be inserted in (if you have met them)
- You share the company's values/mission
- The company operates in a dynamic, ever-changing industry
- The company's prestige

Of course, remuneration is one of the main motivators for almost all people. However, talking about money is not a good idea at this point in the selection

process. Instead, focus on some of the aspects that we listed above and customize them to the specific position that you are applying for.

What you say while answering this question is not the only important thing. Your interviewer will be eager to see that all signs point in the same direction. Try to show that you are excited through your voice, posture and body language. This can be the critical difference that will determine whether or not you will be selected.

30. Give me an example of a time when you had to go the extra mile?

"The only way to do great work is to love what you do"

Steve Jobs

Going the extra-mile is rarely a one-time act. More often, it is an ingrained habit. You need to properly explain to your recruiter that you love the idea of working that job. Also, explain how you want to be excellent at it. Your internal drive towards excellence is what motivates you to go the extra mile – to do the things that you are not expected to do:

- Study during the weekends
- Stay late in the office
- Striving for excellence constantly

If the job you are interviewing for is what you chose for your life, then you want to be excellent at it. Striving to achieve excellent performance is important. It means that you want to put quality in your work and create value for the company. Internal drive is probably the best reason to go the extra mile; you are willing to do what is necessary in order to be good at what you do.

An example of such a situation:

During your previous internship experience, you put in a lot of extra effort in order to show that your tutor who also recruited you did not make a mistake. You stayed late and studied during the weekend because you wanted to improve your skills and to do it faster. The positive impression that you left with your work led to an excellent valuation and very positive feedback about your willingness to learn.

31. Can you tell me a time when you were able to build motivation in your co-workers?

This question aims to assess whether you are a good leader and a positive influence at your workplace. Hiring managers look for people who are motivated themselves and are able to transmit their drive to their co-workers. Strong motivation makes for excellent results.

In order to be able to motivate someone, you have to fully understand the person that you are approaching. What is it that they currently need in order to be excited about a project?

Perhaps they need:

- One-on-one coaching
- Interesting tasks
- More complicated tasks
- Responsibility
- Autonomy
- Recognition
- A positive perspective

An example of such a situation:

During your previous internship within the Corporate Finance department of a large firm, you were asked to prepare a Valuation model. There was another intern who was assigned to work with you. Given that she had less experience with Financial Modeling, she could only help you with minor data entry and consistency checks. You noticed that this was not particularly stimulating for her, as this is something she already knew how to do and she really wanted to learn how to create the model itself. You realized that she would be more motivated to do her part if she was given the opportunity to learn as well. That is why you asked her whether she would like to sit next to you while you work on the model, so that the two of you can comment on what you are doing together. This greatly motivated her and she came up with some valuable suggestions when you had to prepare a presentation that summarizes the model that you prepared.

32. How do you handle a challenge?

First of all, you want to give the impression that you are someone who welcomes a challenge. You are a person who is willing to leave his/her comfort zone and embrace challenging situations. You learn the most when you are put in a difficult situation. And this is certainly something that the Hiring Manager is looking to hear from you. The second part of the question is how you actually handle a challenge. Do you have a structured approach? Are you a person who builds a plan of action and then sticks to it? It would be best if you could provide an example of your past experience. A story showing that:

- you understood the issue
- you created a plan of action
- you executed the plan of action successfully

An example of such a situation:

Let's say that you were admitted to a Master's in Economics. A really challenging situation arose because you knew that most of the people in the class had already studied Finance and Econometrics, while you concentrated on Leadership courses. There was a significant gap between your skills and those of others. You realized that. You also realized that the only way to address the issue was to start with the very basics and fill the knowledge gap step by step; a very long process that required significant efforts on your end. An encouraging sign was that the results at the end of the first semester showed that you reduced the gap significantly and were

heading in the right direction. By the end of the second semester, your GPA was slightly higher than the average for the class.

33. What is your greatest weakness?

The problem with this question is that you are being asked about your shortcomings, while you are doing an interview and you want to make a good impression. Make sure that you don't choose something that can impede you from being great at the job you are interviewing for. For example, if you are interviewing for a controller or a financial analyst, it is OK to say that you do not like to speak in public. However, if you are applying for a consulting or an investment bank job you should not say that, because public speaking can be essential for those professions.

Choose a weakness that you can turn into a positive. "I am usually not good at…but I am making an effort to improve that". Avoid cliché answers like "I work too hard" and "I am a perfectionist". No one is perfect – that is why you need to indicate a weakness when you are asked about one. This shows that you are self-aware and have listened to feedback.

An example of such a situation:

The tutor at my previous internship gave me some interesting feedback: "Don't try to do too much." I remembered that and had a chance to reflect on it, once the internship was over. He was right; I tried to do too much. I was eager to prove myself and implement everything that I learned in university so I could perform great. Trying to implement complex models and "doing too much" is something that I need to control in the future.

This experience allowed me to understand that greatness is a lot of small things done well.

Therefore, I decided that the next time when I am facing a similar situation, I will focus on my own duties and will make sure that I do everything that is expected of me well, instead of trying to invent the next formula of relativity.

34. How do you handle working with numbers/ clients multiple tasks/stress?

Each of these aspects can be really important for a given position and the Hiring Manager will want to make sure that you are the right person that he/she is looking for. Try to figure out the most important characteristics of the job that you are applying for. Are you expected to do multitasking? What part of your overall responsibilities would be related to financial figures? Are you going to interact with many people?

Based on your findings, you will know what to expect. Prepare good examples from your past that can serve as proof of your statements.

35. What would you do if the priorities of an important project you were working on suddenly changed?

It's a very broad question, isn't it? Try answering by asking some questions that can guide you to the right answer:

- Who changed the project's priorities? Your boss? Clients? Suppliers? External Factors?
- Why did they change priorities?

Try to understand the reason behind the decision and assess whether it is a valid one. Is there something that you can do about it?

If you believe that you can propose a solution, don't be shy about contacting the responsible manager and sharing your idea.

If you believe that the reason for shifting priorities is not valid, raise your concerns with Management.

If there is nothing to do about the decision (external factors that can't be changed are the reason or your Boss says that despite your concerns, the decision to change priorities remains), create a course of action and make sure that everybody on your team is aligned with the new priorities. Schedule a reasonable deadline and think of the best way that you can achieve the new goals.

36. What would you do if someone at work resisted your ideas?

Again, open communication is the best way to approach this problem. First of all, you need to make sure that you are fully explaining your ideas. Perhaps you can try an alternative approach? You can provide practical examples or make a list of the pros and cons of your suggestion. Then you should try to understand your colleague's point of view. What are the reasons behind his resistance? If his point is valid as well, think of an alternative approach together regarding the problem. Maybe you can create a hybrid solution that will include your ideas and will address his concerns.

37. Is there anything else that we should know about you?

Yes. The answer to this question is always "Yes". There are many things that they should know about you. This question typically comes at the end of the interview and it is an opportunity to close in a strong fashion. There is no need to pass up on this extra opportunity that you are given by the interviewer. Try to address some of the following points that did not come up during the interview:

- Skills that are relevant for the job under consideration
- Past experience that will help you to be successful at this job

- Motivation to work for the company in the particular role that you are interviewing for
- What is going to be your added value to the team that you will be placed in

One of the basic rules in sales is that you need to convince your client that he/she needs your product. This is a similar situation. Make a closing statement that will convince your interviewer that you are the right person that they are looking for.

BRAINTEASERS

Brainteasers give the interviewer an overall idea of your logic and math abilities, critical thinking and creativity. But, above all, they're intended to examine how you apply all of these under pressure. So, here are a few questions and their answers that will help you hone your problem-solving skills. Or at least give you an understanding of the schema around which your answer should be organized.

38. Imagine the following situation:

You're in a room with 3 light switches. In the next room, there are 3 light bulbs, each controlled by one of the switches. You have to find out which switch controls each bulb by checking the room just once. Keep in mind that all lights are initially off, and you can't see into 1 room from the other. So, how can you figure out which switch is connected to which light bulb?

And here's the answer:

Let's say we have switches 1, 2, and 3. What you can do is leave switch 1 off, turn switch 2 on for 5 minutes, and then turn it off. Then turn switch 3 on and leave it like that. Then you enter the room. Obviously, switch 3 controls the light bulb you left on. The bulb that is off but still warm, is controlled by switch 2. And switch one controls the light bulb you never turned on.

39. You want to buy a work of art that was $400 but is now sold at 25% off.

How much is the promotion price?

It's time for a quick calculation: What's 75% off $400? The answer is $300. Of course, if you're into numbers and like using shortcuts, don't hesitate to think out loud.

GUESSTIMATE

Guesstimate cases are a sort of a prelude to a full-blown business situation case. They show the data science interviewer how you approach problems and test both your judgment and numerical thinking.
Here are a few to get you started...

40. How many square feet of pizza are eaten in the United States each month?

Let's say there are roughly 300 million people in America, out of which 200 million eat pizza. Now, suppose the average pizza-eater has pizza twice a month and eats two slices at a time. That makes four slices per month. Let's say the usual slice of pizza is about six inches at the base and 10 inches long. That means the slice is 30 square inches of pizza. Consequently, four slices of pizza would amount to 120 square inches. We know that one square foot equals 144 square inches, we can say that each pizza-eater consumes one square foot per month. And, as there are 200 million pizza-eaters in America, we can conclude that 200 million square feet of pizza are consumed in the US each month.

DATA ANALYST INTERVIEW QUESTIONS

If you're aiming for a data analyst job, sooner or later, you'll reach the final stage of the application process - the data analyst job interview. So, how can you ace the interview with ease? By being well-familiar with the data analyst interview questions in advance. And that's exactly why we've prepared the following collection of actual data analyst interview questions and answers.

GENERAL DATA ANALYST INTERVIEW QUESTIONS

General data analyst interview questions are not just about your background and work experience. In fact, interviewers might surprise you with questions requiring details about the projects you've been involved, and how you approach complex data sets. So, let's take a look...

1. Can you share details about the largest data set you've worked with?

How many entries and variables did the data set comprise? What kind of data was included?

How to Answer

Working with large datasets and dealing with a substantial number of variables and columns is important for a lot of hiring managers. When answering the question, you don't have to reveal background information about the project or how you managed each stage. Focus on the size and type of data.

Answer Example

"I believe the largest data set I've worked with was within a joint software development project. The data set comprised more than a million records and 600-700 variables. My team and I had to work with Marketing data which we later loaded into an analytical tool to perform EDA."

2. In your role as a data analyst, have you ever recommended a switch to different processes or tools?

What was the result of your recommendation?

How to Answer

For hiring managers, it's important that they pick a data analyst who is not only knowledgeable but also confident enough to initiate a change that would improve the company's status quo. When talking about the recommendation you made, give as many details as possible, including your reasoning behind it. Even if the recommendation you made was not implemented, it still demonstrates that you're driven and you strive for improvement.

Answer Example

"Although data from non-technical departments is usually handled by data analysts, I've worked for a company where colleagues who were not on the data analysis side had access to data. This brought on many cases of misinterpreted data that caused significant damage to the overall company strategy. I gathered examples and pointed out that working with data dictionaries can actually do more harm than good. I recommended that my coworkers depend on data analysts for data access. Once we implemented my recommendation, the cases misinterpreted data dropped drastically."

3. How would you assess your writing skills?

When do you use written form of communication in your role as a data analyst?

How to Answer

Working with numbers is not the only aspect of a data analyst job. Data analysts also need strong writing skills, so they can present the results of their analysis to management and stakeholders efficiently. If you think you are not the greatest data "storyteller", make sure you're making efforts in that direction, e.g. through additional training.

Answer Example

"Over time, I've had plenty of opportunities to enhance my writing skills, be it through email communication with coworkers, or through writing analytical project summaries for the upper management. I believe I can interpret data in a clear and succinct manner. However, I'm constantly looking for ways to improve my writing skills even further."

4. Have you ever used both quantitative and qualitative data within the same project?

How to Answer

To conduct meaningful analysis, data analysts must use both the quantitative and qualitative data available to them. In surveys, there are both quantitative and qualitative questions, so merging those 2 types of data presents no challenge whatsoever. In other cases, though, a data analyst must use cre-

ativity to find matching qualitative data. That said, when answering this question, talk about the project where the most creative thinking was required.

Answer Example

"In my experience, I've performed a few analyses where I had qualitative survey data at my disposal. However, I realized I can actually enhance the validity of my recommendations by also implementing valuable data from external survey sources. So, for a product development project, I used qualitative data provided by our distributors, and it yielded great results."

5. What is your experience in conducting presentations to various audiences?

How to Answer

Strong presentation skills are extremely valuable for any data analyst. Employers are looking for candidates who not only possess brilliant analytical skills, but also have the confidence and eloquence to present their results to different audiences, including upper-level management and executives, and non-technical coworkers.

So, when talking about the audiences you've presented to, make sure you mention the following:

- Size of the audience;
- Whether it included executives;
- Departments and background of the audience;
- Whether the presentation was in person or remote, as the latter can be very challenging.

Answer Example

"In my role as a Data Analyst, I have presented to various audiences made up of coworkers and clients with differing backgrounds. I've given presentations to both small and larger groups. I believe the largest so far has been around 30 people, mostly colleagues from non-technical departments. All of these presentations were conducted in person, except for 1 which was remote via video conference call with senior management."

6. Have you worked in an industry similar to ours?

How to Answer

This is a pretty straightforward question, aiming to assess if you have industry-specific skills and experience. Even if you don't, make sure you've prepared an answer in advance where you explain how you can apply your background skills from a different field to the benefit of the company.

Answer Example

"As a data analyst with financial background, I can say there are a few similarities between this industry and healthcare. I think the most prominent one is data security. Both industries utilize highly sensitive personal data that must be kept secure and confidential. This leads to 2 things: more restricted access to data, and, consequently, more time to complete its analysis. This has taught me to be more time-efficient when it comes to passing through all the security. Moreover, I learned how important it is to clearly state the reasons for requiring certain data for my analysis."

7. Have you earned any certifications to boost your career opportunities as a Data Analyst?

How to Answer

Hiring managers appreciate a candidate who is serious about advancing their career options through additional qualifications. Certificates prove that you have put in the effort to master new skills and knowledge of the latest analytical tools and subjects. While answering the question, list the certificates you have acquired and briefly explain how they've helped you boost your data analyst career. If you haven't earned any certifications so far, make sure you mention the ones you'd like to work towards and why.

Answer Example

"I'm always looking for ways to upgrade my analytics skillset. This is why I recently earned a certification in Customer Analytics in Python. The training and requirements to finish it really helped me sharpen my skills in analyzing customer data and predicting the purchase behavior of clients."

8. Name a few libraries used in Python for data analysis.

These are the most important Python libraries you should mention.

Numpy is an essential library, as it used for matrices and arrays and includes methods for their manipulation.

Pandas is the second library, which is used in almost any data analysis performed in Python.

It includes data structures and operations for manipulating numerical tables and time series. It often uses numpy to produce linear math results and is, therefore, a lot faster than standard Python. So, a good knowledge of the Pandas library is a must if you're a data analyst using Python.

Scipy and Scikit Learn –are two of the main machine learning libraries.

Sci-py boasts an impressive number of mathematical algorithms and high-level commands and classes to help data scientists in their data analysis tasks. Scikit learn was originally developed during a "Google Summer of Code" project, as a third party extension for Scipy. Scikit learn includes various classification, regression, and clustering algorithms, designed to be incorporated with the Scipy and Numpy packages.

And once you're done with machine learning, you'll also need a good way of visualizing the results. Matplotlib\ Seaborn are-visualization libraries, which are great for that.

Tensorflow, Keras and Pytorch are libraries for deep learning. If you want to train neural networks, for example in the context of NLPs or Computer Vision, these are the way to go. Here knowing the difference between Tensorflow 1 and Tensorflow 2 could be a bonus during an interview.

9. What is a Logistic regression?

A logistic regression is one of the simplest classification models. It is widely used mainly due to its simplicity and ease of interpretation. Logistic regressions are well understood and studied throughout the years and thus are still a data scientist's preferred classification choice on many occasions.

A logistic regression could be used in 2 distinct ways that sound different yet are reached in the same way, methodologically speaking.

The first use case is whenever we've got a categorical outcome. Examples are: Yes/No, Will buy/Won't buy, and 0/1 situations. As any other classification method, a logistic regression would output the category it deems most probable to be the answer.

Speaking of probabilities, we reach ***the second use case.*** We could employ a logistic regression to determine the exact probability that an event is going to occur.

The mechanics of the two use cases follow the same path.

For instance, imagine a logistic regression predicts that a customer is 70% likely to buy and 30% likely to not buy. Under these conditions, the prediction will be classified as 'Will buy'. Depending on our needs we could use one the probabilistic representation or simply the output class.

Finally, it is useful to note that we were discussing a binary logistic regression.

Binary here stands for an outcome with only 2 possibilities. The logistic regression model could be generalized to many categories, in which case it would be called a multinomial logistic regression.

At this point, you may or may not decide to mention the multinomial logistic regression. In 99% of the cases where we use the term 'logistic regression', we mean binary logistic regression. Referring to the multinomial case could prompt the interviewer to ask you additional questions on multinomial logistic regression, which would definitely be much trickier for you, especially if you have never used it.

10. Have you worked with comparatively large data sets in a project?

How did you collect and prepare the data for analysis?

How to Answer

Working with large data sets can be challenging. So, with this question, the hiring manager wants to assess your ability to deal with the issues that might occur. If you have relevant experience, talk about the problems you have faced and how you managed to resolve them. In case you've never experienced any issues working with large data sets, describe the details of the project and all the stages of preparing the data for analysis.

Answer Example

"In the last company I was in, I often worked with large data sets from external suppliers. For example, survey responses for Customer Analytics projects. And that means a large data set with a huge sample size. So, to prepare the data for analysis, I'd go through the following steps. First, I'd run predetermined frequencies and queries to check the validity of the data. This helped me pin down various problems, such as missing data, problems with the data type, or skip-pattern errors in the survey data set. I'd check with the supplier, so we can implement the necessary corrections before we move forward with the analysis. Once done, I'd often consult with a Data Engineer to pick the most suitable analysis tool for a data set of this size. Finally, I'd load the data and start my analysis."

11. Which tools have you used in each stage of your previous data analysis projects?

How to Answer

A data analyst must be experienced in using a wide range of tools in the various phases of their analyses – from preparation, through exploration, to presenting the end results. Hiring managers know that a single tool can be utilized in multiple stages of the analytical process. So, if that's your experience, make sure you highlight it. This will demonstrate your expertise in working with that specific tool. However, if you have worked with multiple tools throughout your experience, share that, too. That's how you'll showcase the span of your skills.

Answer Example

"In my experience as a data analyst, I've used a variety of tools that have helped me build up a strong skillset. In the preparation and exploration stages, I've mostly used Microsoft Excel and Microsoft Access, depending on the complexity of the data set. While in the exploration phase, I've also used SAS and SPSS to extract insights from the data. Apart from these statistical programs, I've employed analytical tools, such as Tableau and Cognos Analytics. I find Tableau, together with Power BI to be great tools for creating powerful dashboard visualizations. And, of course, Excel and PowerPoint are classic tools for building in-company presentations."

12. In large companies, data is often stored in multiple data warehouses.

Have you ever worked on a complex analytical project, where you had to query multiple data warehouses in order to gather the required data?

How to Answer

The technical complexity of your work as a Data Analyst may vary depending on the size of the companies you have worked at in the past. Strong technical skills is an important attribute of a Data Analyst's background. Having experience retrieving data from multiple data warehouses demonstrates your understanding of databases, data structures, and programming languages.

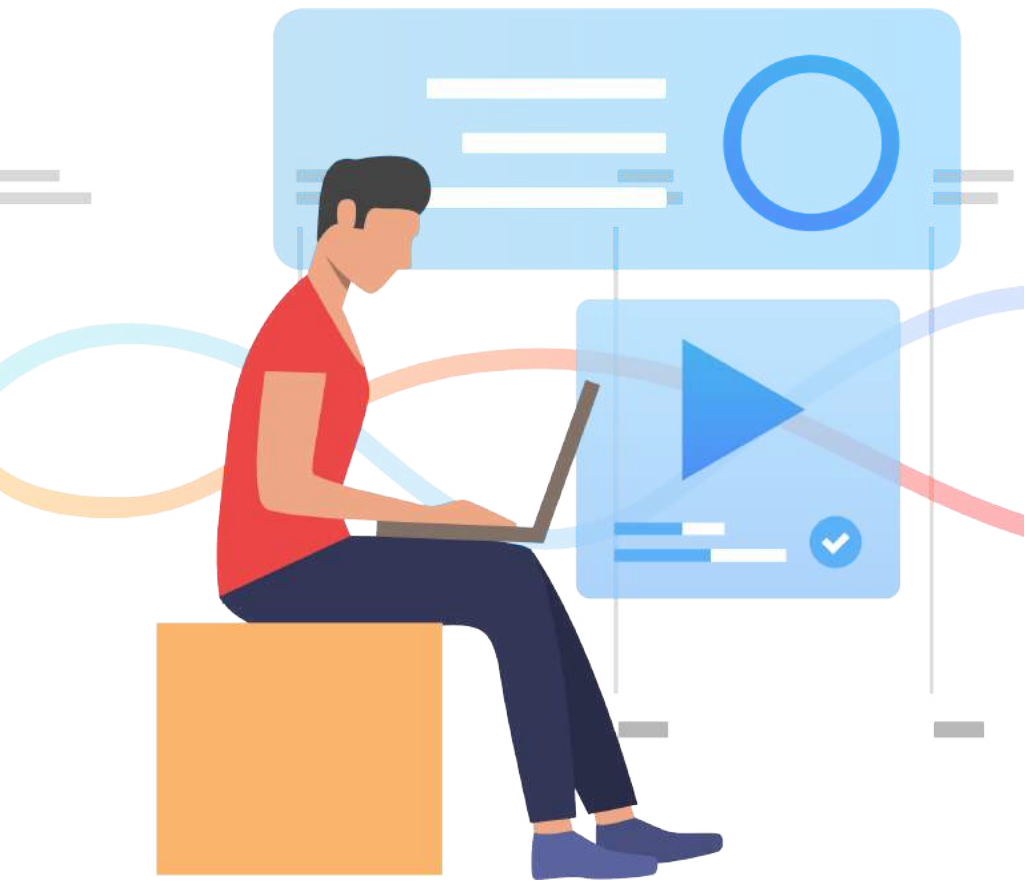
The size of the companies you've worked for can affect the technical complexity of your tasks as a data analyst. That said, a strong technical skillset is always a plus in the eyes of your future employer. So, having retrieved data from multiple data warehouses in your work on past projects will showcase your expertise in databases and data structures, as well as in programming languages.

Answer Example

"I've had the chance to work for a big corporation in the past. I can say my work there has been of great importance to developing my technical skillset. Once, I queried against 5 different data warehouses to retrieve the data for a large-scale company project. Once I had all the necessary records and variables, I built a dataset I later utilized in my analysis."

TECHNICAL DATA ANALYST INTERVIEW QUESTIONS

Technical data analyst interview questions are focused on assessing your proficiency in analytical software, visualization tools, and scripting languages, such as SQL and Python. Depending on the specifics of the job, you might be requested to answer some more advanced statistical questions, too. Here are some real-world examples...



13. What tools or software do you prefer using in the various phases of data analysis and why?

How to Answer

Although you might think you should have experience with as many tools as possible to ace this question, this is not the case. Each company uses specific data analysis tools, so it's normal that your expertise is limited to those. Of course, if you have worked for a large number of companies, you're bound to have exposure to a wider variety of analytical software. That said, the interviewer would like to know which tools you feel comfortable with, rather than the number of tools you've utilized.

Answer Example

"When it comes to data analysis tools, I can say I'm a traditionalist. That's why I find Microsoft Excel and Microsoft Access most useful. I feel truly comfortable working with those, and they're available in almost every company out there. Moreover, you can achieve great results with them with the right training."

14. Have you ever created or worked with statistical models?

If so, please describe how you've used it to solve a business task.

How to Answer

As a data analyst, you don't specifically need experience with statistical models, unless it's required for the job you're applying for. If you haven't been involved in building, using, or maintaining statistical models, be open about it and mention any knowledge or partial experience you may have.

Answer Example

"Being a data analyst, I can't say I've had direct experience building statistical models. However, I've helped the statistical department by making sure they have access to the proper data and analyzing it. The model in question was built with the purpose of identifying the customers who were most inclined to buy additional products and predicting when they were most likely to make that decision. My job was to establish the appropriate variables used in the model and assess its performance once it was ready."

15. Which step of a data analysis project do you enjoy the most?

How to Answer

It's normal for a data analyst to have preferences for certain tasks over others. However, you'll most probably be expected to deal with all steps of a project – from querying and cleaning, through analyzing, to communicating findings. So, make sure you don't show antipathy to any of the above. Instead, use this question to highlight your strengths. Just focus on the task you like performing the most and explain why it's your favorite.

Answer Example

"If I had to select one step as a favorite, it would be analyzing the data. I enjoy developing a variety of hypotheses and searching for evidence to support or refute them. Sometimes, while following my analytical plan, I have stumbled upon interesting and unexpected learnings from the data. I believe there is always something to be learned from the data, whether big or small, that will help me in future analytical projects."

16. What's your knowledge of statistics and how have you used it in your work as a data analyst?

How to Answer

Data analysts should have basic statistics knowledge and experience. That means you should be comfortable with calculating mean, median, and mode, as well as conducting significance testing. In addition, as a data analyst, you must be able to interpret the above in connection to the business. If a higher level of statistics is required, it will be listed in the job description.

Answer Example

"In my line of work, I've used basic statistics – mostly calculated the mean and standard variances, as well as significance testing. The latter helped me determine the statistical significance of measurement differences between two populations for a project. I've also determined the relationship between 2 variables in a data set, working with correlation coefficients."

17. What scripting languages have you used in your projects as a data analyst?

Which one you you'd say you like best?

How to Answer

Most large companies work with numerous scripting languages. So, a good command of more than one is definitely a plus. Nevertheless, if you aren't well familiar with the main language used by the company you apply at, you can still make a good impression. Demonstrate enthusiasm to expand your knowledge, and point out that your fluency in other scripting languages gives you a solid foundation for learning new ones.

Answer Example

"I'm most confident in using SQL since that's the language I've worked with throughout my Data Analyst experience. I also have a basic understanding of Python and have recently enrolled in a Python Programming course to sharpen my skills. So far, I've discovered that my expertise in SQL helps me advance in Python with ease."

18. How many years of SQL programming experience do you have?

In your latest job, how many of your analytical projects involved using SQL?

How to Answer

SQL is considered as one of the easiest scripting languages to learn. So, if you want to be competitive in the job market as a Data Analyst, you should be able to demonstrate an excellent command of SQL. Even if you don't have many years of experience, highlight how your skills have improved with each new project.

Answer Example

"I've used SQL in at least 80% of my projects over a period of 5 years. Of course, I've also turned to other programming languages for the different phases of my projects. But, all in all, it's SQL that I've utilized the most and consider the best for most of my data analyst tasks."

19. Which Excel functions have you used on a regular basis so far?

Can you describe in detail how you've used Excel as an analytical tool in your projects?

How to Answer

If you are an Excel expert, it would be difficult to list all the functions you have experience using. Instead, concentrate on highlighting the more difficult

ones, particularly statistical functions. If you have experience utilizing the more challenging functions, hiring managers will presume you have experience using the more basic ones. Be sure to highlight your pivot table skills, as well as your ability to create graphs in Excel. If you have not attained these skills yet, it is worthwhile to invest in training to learn them.

If you're an Excel pro, there is no need to recite each and every function you've used. Instead, highlight your advanced Excel skills, such as working with statistical functions, pivot tables, and graphs. Of course, if you lack the experience, it's worth considering a specialized Excel training that will help you build a competitive skillset.

Answer Example

"I think I've used Excel every day of my data analyst career in every single phase of my analytical projects. For example, I've checked, cleaned, and analyzed data sets using Pivot tables. I've also turned to statistical functions to calculate standard deviations, correlation coefficients, and others. Not to mention that the Excel graphing function is great for developing visual summaries of the data. As a case in point, I've worked with raw data from external vendors in many customer satisfaction surveys. First, I'd use sort functions and pivot tables to ensure the data was clean and loaded properly. In the analysis phase, I'd segment the data with pivot tables and the statistical functions, if necessary. Finally, I'd build tables and graphs for efficient visual representation."

20. What's your experience in creating dashboards?

Can you share what tools you've used for the purpose?

How to Answer

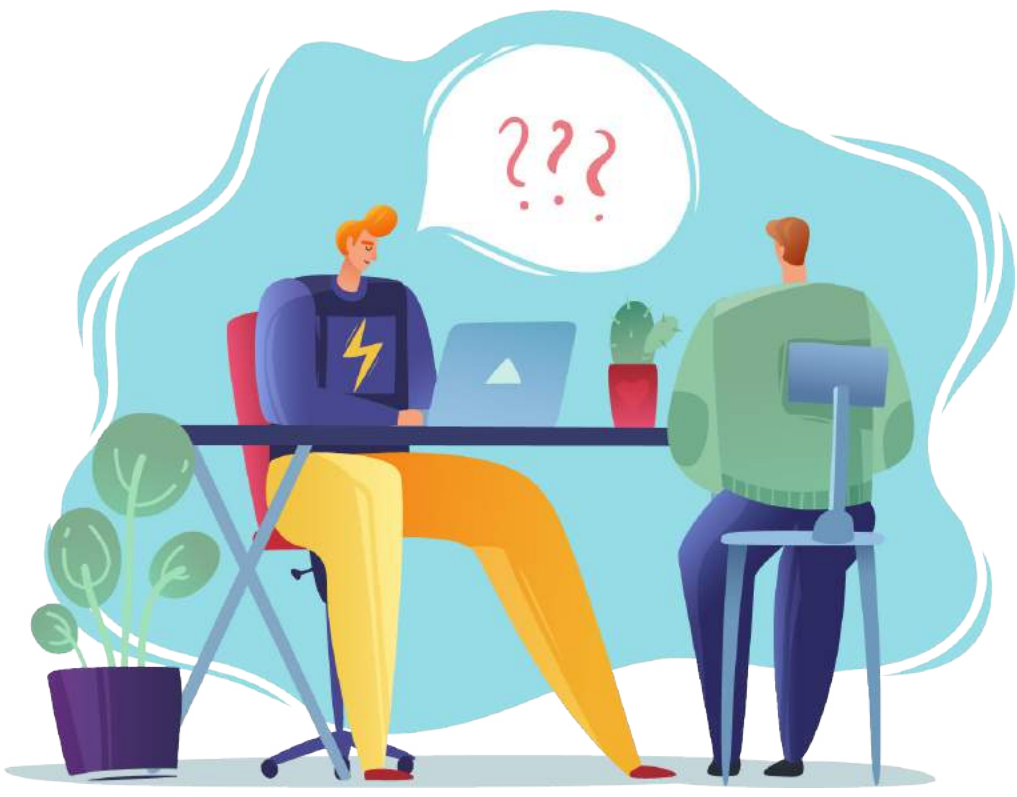
Dashboards are essential for managers, as they visually capture KPIs and metrics and help them track business goals. That said, data analysts are often involved in both building and updating dashboards. Some of the best tools for the purpose are Excel, Tableau, and Power BI (so make sure you've got a good command of those). When you talk about your experience, outline the types of data visualizations, and metrics you used in your dashboard.

Answer Example

"In my line of work. I've created dashboards related to customer analytics in both Power BI and Excel. That means I used marketing metrics, such as brand awareness, sales, and customer satisfaction. To visualize the data, I operated with pie charts, bar graphs, line graphs, and tables."

BEHAVIORAL QUESTIONS

To answer this type of interview questions with ease, you'll need to take a walk down memory lane and recall details about how you handled specific challenges in your work with stakeholders, coworkers, or clients. Here's what we have in mind:



21. Imagine the following situation:

As a data analyst, you'll often work with stakeholders who lack the technical background and deeper understanding of data and databases. Have you ever been in a situation like this and how did you handle this challenge?

How to Answer

Data analysts often face the challenge of communicating findings to coworkers from different departments or senior management with limited understanding of data. This requires excellent skills in interpreting specific terms using non-technical language. Moreover, it also requires extra patience to listen to your coworkers' questions and provide answers in an easy-to-digest way. Show the interviewer that you're capable of working efficiently with people from different types of background who don't speak your "language".

Answer Example

"In my work with stakeholders, it often comes down to the same challenge – facing a question I don't have the answer to, due to limitations of the gathered data or the structure of the database. In such cases, I analyze the available data to deliver answers to the most closely related questions. Then, I give the stakeholders a basic explanation of the current data limitations and propose the development of a project that would allow us to gather the

unavailable data in the future. This shows them that I care about their needs and I'm willing to go the extra mile to provide them with what they need."

22. Tell me about a time you and your team were surprised by the results of a project.

How to Answer

When starting an analysis, most data analysts have a rough prediction of the outcome rested on findings from previous projects. But there's always room for surprise, and sometimes the results are completely unexpected. This question gives you a chance to talk about the types of analytical projects you've been involved in. Plus, it allows you to demonstrate your excitement about drawing new learnings from your projects. And don't forget to mention the action you and the stakeholders took as a result of the unexpected outcome.

Answer Example

"While performing routine analysis of a customer database, I was completely surprised to discover a customer subsegment that the company could target with a new suitable product and a relevant message. That presented a great opportunity for additional revenue for the company by utilizing a subset of an existing customer base. Everyone on my team was pleasantly surprised and soon enough we began devising strategies with Product Development to address the needs of this newly discovered subsegment."

23. Why do you think creativity is important for a data analyst?

How have you used creative thinking in your work so far?

How to Answer

A data analyst is usually seen as a professional with a technical background and excellent math and statistical skills. However, even though creativity is not the first data analyst quality that comes to your mind, it's still important in developing analytical plans and data visualizations, and even finding unorthodox solutions to data issues. That said, provide an answer with an example of your out-of-the-box way of thinking.

Answer Example

"I can say creativity can make all the difference in a data analyst's work. In my personal experience, it has helped me find intriguing ways to present analysis results to clients. Moreover, it has helped me devise new data checks that identify issues resulting in anomalous results during data analysis."

24. What are the most important skills a data analyst should possess to work efficiently with team members with various backgrounds, roles, and duties?

How to Answer

When answering this question, keep in mind that the hiring manager would like to hear something different than "communication skills". Think of an approach you've used in your role as a data analyst to improve the quality of work in a cross-functional team.

Answer Example

"I think the role of a data analyst goes beyond explaining technical terms in a non-technical language. I always strive to gain a deeper understanding of the work of my colleagues, so I can bridge my explanation of statistical concepts to the specific parts of the business they deal with, and how these concepts relate to the tasks at hand they need to solve."

25. In your opinion, which soft skills are essential for a data analyst and why?

How to Answer

Soft skills, a.k.a. non-technical skills are important for working efficiently with others and maintaining a high level of performance. As with most professions, data analysts should be aware of how their behavior and work habits affect the members on their team. Therefore, here you should base your answer on past work experience and highlight an important soft skill you have developed.

Answer Example

"I believe leadership skills are one of the major soft skills a data analyst should develop. The way I understand it, leadership means taking action to guide and help the members on your team. And this doesn't necessarily mean you have to be in a managerial position. In my line of work, leadership would translate into providing expert insights regarding company data and its interpretation. That's a skill I've worked hard to develop over the years. I can say being confident in my abilities has now established me as a leading figure in my area, and my team members know they can rely on my expertise."

26. Tell us about a project where, due to data limitations, the stakeholders couldn't reach the answer they needed.

How did you resolve this issue?

How to Answer

The interviewer wants to be reassured that, as a data analyst, you can deal with all types of data challenges. That's particularly important when collaborating with stakeholders who may lack an in-depth understanding of data. This question is also ideal for showcasing your problem-solving skills.

Answer Example

"A few years back, I worked on a customer segmentation project initiated by the company executives. Unfortunately, they couldn't come up with a substantial customer segmentation plan, as the data in the customer data warehouse wasn't robust enough. To help with the progress of the project, I worked closely with the data warehouse team. Our collaboration resulted in outlining data initiatives and actionable steps which ultimately led the project to its final goal."

27. What web analytics tools have you used in your professional experience?

How to Answer

More and more data analyst job postings require web analytics experience (or list it as a preferred skill). And, while some companies separate the roles and their job descriptions, others prefer to hire a data analyst with an all-encompassing skillset. So, if you have relevant experience, it's a good idea to mention the metrics you were tracking and the field of their application."

Answer Example

"Using Google Analytics, I have used web analytics as part of a larger marketing campaign evaluation project. The web metrics I tracked included open rate, click-through rate, average time on page and conversion rate. In addition, I was able to build funnels within Google Analytics to measure where visitors were dropping off before converting. By tracking these web metrics in conjunction with non-web marketing efforts, I was able to recommend the best marketing channels to use to target specific segments."

"I have experience using Google Analytics for a Black Friday campaign evaluation project. For the purpose, I had to track the following metrics – open-rate, click-through rate, conversion rate, and average time on page. I also used Google Analytics to build funnels that measure at which part of their journey the visitors dropped off prior to converting. Tracking these web metrics helped me come up with recommendations about the best marketing channels for targeting specific audiences."

28. Give me an example of a time when you worked as a team.

"Coming together is a beginning.
Keeping together is progress.
Working together is success."

Henry Ford

One of the greatest virtues in the modern corporate world is the ability to work well as a team. Make sure that you are ready with a story that shows you are able to do exactly that. A team worker can be distinguished by his/her ability to:

- Put the team's needs first
- Communicate well with the other team members
- Want to succeed as a part of a group

- Listen actively
- Respect others
- Appreciate other work styles

Keep in mind these qualities when you think of a story when you were part of a team. The story should demonstrate not only the fact that you were part of the team, but also that you were a great one too.

Here's an example of such a situation:

A group assignment during the last year of my studies required me and four of my classmates to perform a detailed Company Valuation.

This was a pretty difficult task that included a significant amount of work. The deadline for submitting the complete work was in 2 weeks. At the time, I was busy filling out internship applications and had to prepare for some of my other exams. This was the case for the other team members as well.

Nevertheless, all of us concentrated full-time on the project, as I understood that this was the only way we could have respected the tight deadline imposed. Another interesting thing about the project was that we managed to work well together, despite the different styles that each group member had. We listened actively and were open to the ideas that the others had. Given that we came from a different background, each of us certainly added value to the project. Good communication helped us coordinate our responsibilities and integrate the separate pieces of work that we were assigned individually.

29. Describe a time when you failed to meet your goals.

Some failure in life is inevitable. Those who are brave and bold attempt many new things and thus fail much more often. Don't be afraid to explain a time when you wanted to achieve something, but you were not able to do it. Chances are that the interviewer is more interested in learning how you handled the failure that you experienced. He wants to know whether you learned from your mistakes and whether you are motivated to succeed in the future.

When you think of a story, don't pick a major failure and try to choose a story where external factors influenced your failure as well. Inexperience on your part is OK too, given that you are in the early stages of your career. Don't point out as a reason for your failure qualities that can have a negative impact on your work in the future (for example attention to detail, ability to handle pressure, etc.).

It is very important to show that you turned a negative situation into a valuable learning experience. This will make a great impression on the Interviewer.

Here's an example of such a situation:

Last year, I was eager to find a summer internship opportunity, but I wasn't able to do that. One of the main reasons behind this was the tough job market that we are currently facing. Along with that, I believe I was too inexperienced and did not realize how difficult it was to find a good opportunity.

This year I had a totally different approach. You could say I learned my lesson perfectly. So, I started preparing myself since November and created

a shortlist of opportunities that I wanted to pursue. Then I researched all potential employers and chose the ones that were really interesting. I had more time to work on my CV and Cover Letters and to prepare for interviews. Of course, I wasn't going to make the same mistake twice.

30. Why should we hire you?

This question is very similar to "How would you add value to our company". The Hiring Manager challenges you to sell him/her the idea of you being hired. Your profile is the product that needs to be sold. Remember the example that we gave with the pen?

Most people will start listing their qualities and qualifications, hoping that they will touch the right nerve along the way. But that is not the way to go.

The Hiring Manager has read your CV, he/she already knows about your credentials. What he/she wants to understand is whether you can handle a tough question and be persuasive while making a valid point. Try to open your answer with a question instead:

Manager: Let me ask you, with so many people applying for this job, why should we hire you?

Job-Seeker: A great question. But I would like to ask you something as well. Can I?

Manager: Sure, go ahead.

Job-Seeker: What makes a great Analyst with your firm?

Manager: We are looking for people who are very independent and are able to learn fast, even when they are under pressure. Does that make sense?

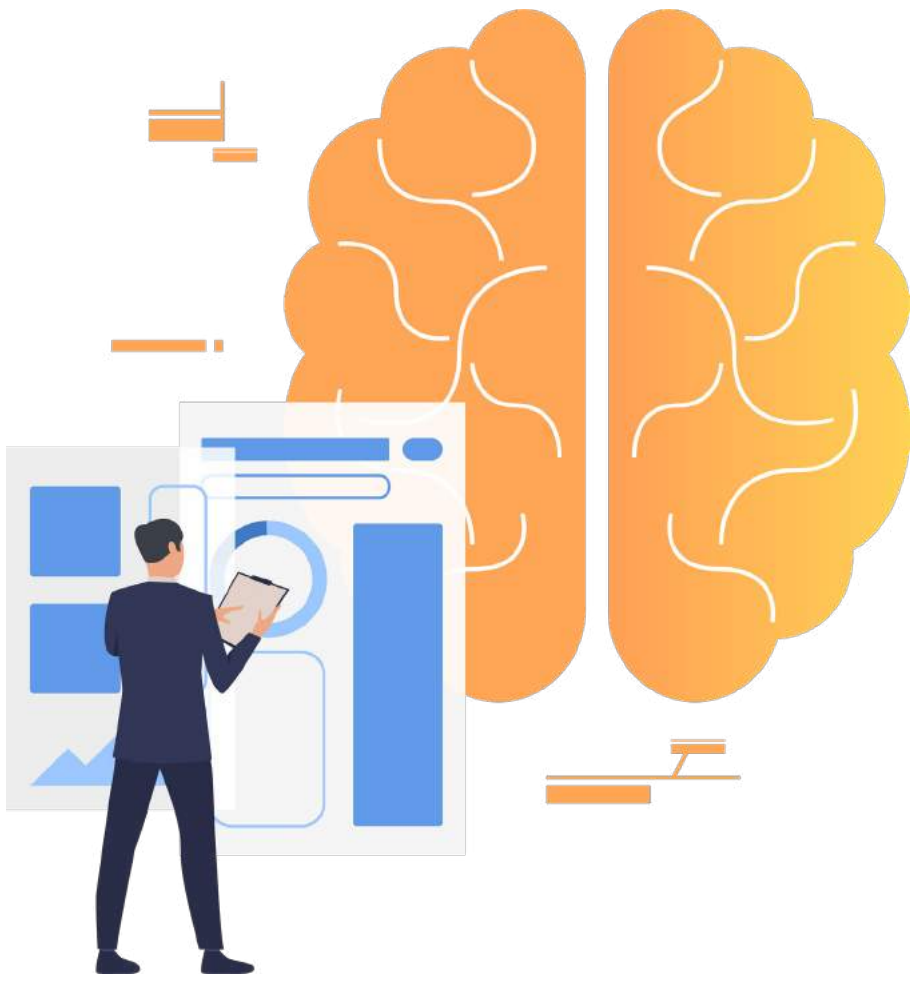
Job-Seeker: Sure, it does. I can imagine that the environment in which your firm operates requires such qualities. This is precisely what made me apply for this position in the first place. I want to be a part of your dynamic environment. I am able to learn fast and adapt to changing circumstances quite easily.

Sounds much better, right?

In order to respond successfully to this question, you need to communicate well with the interviewer and understand exactly what they are looking for. Otherwise, you simply don't know why they should hire you, leaving your answer to be a shot in the dark.

BRAINTEASERS

Interviews for analytical and technical positions often include brainteasers that aim to evaluate how you apply logic, critical thinking, and creativity under pressure. Here's an example...



31. Imagine the following situation:

A car travels a distance of 60 miles at an average speed of 30 miles per hour. How fast does the car need to travel on the way back (taking the same road) in order to average 40 miles per hour over the course of the entire trip?

You need to build the following equation:

The total distance that needs to be traveled both ways is 120 miles. The average speed that we need to obtain is 40 miles; therefore, the car must travel for 3 hours in order to achieve that:

120 miles/40 miles per hour = 3 hours

The car has already traveled for two hours:

60 miles/30 miles per hour = 2 hours

So, on the way back it needs to travel only 1 hour. The distance is 60 miles. Hence the car needs to travel at 60 miles per hour.

GUESSTIMATE

Guesstimates can be critical in picking the right candidate for a data analyst job, as they assess your problem-solving abilities, confidence with numbers, and how you handle different scenarios.

The restaurant is open 6 days of the week (they are closed on Monday), which means that is open 25 times during lunch and dinner time per month. It is a small family restaurant with around 60 places. On average 30 customers visit the restaurant at lunch and 40 people come to have dinner. The typical lunch menu costs 10 euro, while dinner at this restaurant costs twice that amount – 20 euro. Therefore, they are able to achieve revenues of:

25 (days) * 30 (customers) * 10 (EUR) =
= 7,500 EUR (lunch)
25 (days) * 40 (customers) * 20 (EUR) =
= 20,000 EUR (dinner)

32. What is the monthly profit of your favorite restaurant?

Pick a small family restaurant and not a chain of restaurants. This should make calculations much easier.

Then define the main parameters of the restaurant that we are talking about:

- Days of the week in which the restaurant is open
- Number of tables/seats
- Average number of visitors during lunchtime
- Average number of visitors at dinner
- Average expenditure per client during lunch
- Average expenditure per client during dinner

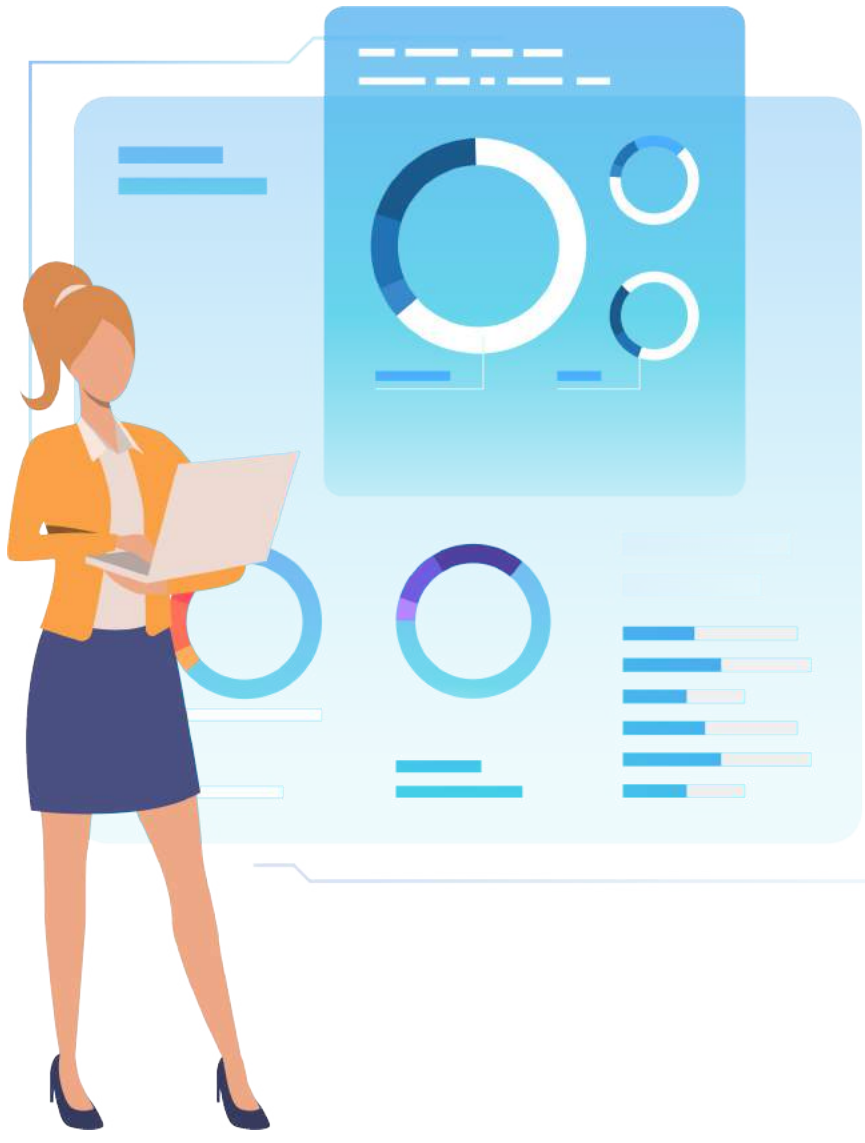
The restaurant is able to achieve 27,500 EUR of sales. Besides, the owner and his wife 4 people work there as well. Let's say that the 3 waiters make 2,000 EUR each and the chef makes 3,000 EUR (including social security contributions). So the cost of personnel is 9,000 EUR. Usually, food and drinks cost around one-third of the overall amount of sales. Therefore Cost of goods sold amounts to 9,125 EUR. Utility and other expenses are another 10% of Sales, so we will have an additional cost of 2,750 EUR. The owners do not pay rent, because they own the place. After the calculations that we made, it results in a monthly profit of (before taxes) 6,625 EUR.

BI ANALYST INTERVIEW QUESTIONS

So, you've got an interview for a BI analyst role coming soon. Congratulations! As businesses strive to maximize the benefits of their data, the demand for BI analysts is massive...But so is the competition. So, to make sure you're ahead of the rest, you should be prepared to answer a plethora of BI analyst interview questions like the ones listed below.

GENERAL BI ANALYST INTERVIEW QUESTIONS

Apart from general questions related to your background as a BI analyst, in this part of the interview, you can expect various industry-specific questions, such as the definition of particular terms, software tools for BI projects, and additional job-related competencies and qualifications.



1. Tell me about your educational background and the business intelligence analysis field you're experienced in.

How to Answer

A business intelligence analyst can concentrate on various industries, such as finance, economics, IT, statistics, manufacturing, and more. Share with the interviewer which area you specialized in while obtaining your university degree, and briefly outline where your career journey has taken you so far. Make sure to demonstrate a keen interest in the company's industry or business sphere.

Answer Example

"I'm a Finance Graduate specialized in Business Administration. My education has helped me greatly on my business intelligence career path, as my interest and expertise evolved in fields such as business law, microeconomics, and financial accounting."

2. What's your experience in SDLC and UAT?

How to Answer

A seasoned BI analyst will have exposure to systems development life cycle (SDLC) and user acceptance testing (UAT). When a company introduces a new software or application to their business, the transition must be well thought out, carefully tested,

and then effectively deployed within the organization. An experienced business intelligence analyst can facilitate this process, saving the company time and financial resources. Talk about your exposure to SDLC and UAT. In case you lack the experience, emphasize your interest in becoming familiar with these activities and learning.

Answer Example

"Although I have limited exposure to SDLC, I've been involved in the UAT phase of some projects. I enjoy analyzing which aspects of a new software program or application are the hardest to implement, which are the easiest to accommodate, and how to proceed from there."

3. Do you plan on continuing your education with an MBA?

How to Answer

A lot of accomplished BI analysts have a Bachelor's degree, while others hold an MBA. With this question, the hiring manager wants to assess your interest in further development that would result in greater career opportunities. As a Master of Business Administration, you'll have an in-depth understanding of enterprise business, the economy, and how various economic and social factors affect the business environment. That said, an MBA isn't a must-have for BI analysts. However, showing an interest could give you some competitive edge.

Answer Example

"I have certainly thought about earning an MBA parallel to advancing in my career. As a Business Intelligence analyst, I believe that having an MBA will undoubtedly expand my knowledge of business economics. And that will definitely be beneficial to my future employer and their clients."

4. What is your opinion about Agile software development for BI projects?

Do you support employing Agile methodologies with your company's clients?

How to Answer

Agile software development has received a warm welcome from companies worldwide since its onset. Agile stimulates collaboration with a team's clients and the end-users, enabling more cross-functional projects to run smoothly. However, there are still some who strongly prefer the structured development methodology of Waterfall, for example. So, before sharing your thoughts on Agile with the interviewer, make sure you know where the company stands on Agile.

Answer Example

"As far as I know, Agile software development is much more collaborative in comparison to other software development models. I believe Agile can be the best solution for many projects. However, maybe that's not always the case. That said, I'd love

to get familiar with the methodologies employed here. At the end of the day, it's the end results that matter most, and not the methodologies behind the projects."

5. In your opinion, what are the key strengths a business analyst should possess?

How to Answer

A great business analyst should have a strong analytical mind, an "out-of-the-box" approach to solving problems, and the ability to handle pressure. Those are just a few of the strengths that a business analyst must possess. However, to avoid searching for an answer on the spot, carefully review the job description for the role. Make a note of the key strengths listed by the employer, and base your response on that.

Answer Example

"When it comes to key strengths, I'd say business analysts should have a profound understanding of the business and its processes. They should also be able to collaborate efficiently with company executives, even if the latter lack technical or analytics background. Last but not least, attention to detail is crucial in this line of work. That said, I've worked hard to develop those skills, and I can't wait to put them into practice in your organization."

6. Do you have a B-plan when faced with a change of course on a moment's notice?

How to Answer

Every skilled business intelligence analyst knows how to pivot, adapt, and change when the plan suddenly falls apart. The ability to solve problems creatively in tense situations is one of the most valuable assets of a business intelligence analyst. So, don't be shy to go into detail about coming up with a number of alternative scenarios for your clients. Although you may never have to resort to them, the fact that you're prepared for emergencies is a great sign for the interviewer.

Answer Example

"Contingency plans are my favorite! As a business intelligence analyst, I know it's great if we can do "X", as planned. However, things aren't always perfect, and plans can change quickly. Especially, if there are a few decision-makers involved in a project. That's why I'm always ready to go with "Y" if the situation calls for it. Having a B-plan takes the edge off, and reassures the whole team that we have a go-to strategy in case we encounter any issues."

7. Have you worked with teams from various departments in a company?

How to Answer

Being able to work in a cross-functional environment is certainly a plus for larger companies. Hiring

managers are aware that you'll probably have to collaborate on projects with teams from other departments, such as HR, IT, or Marketing. Therefore, they want to know more about your exposure to the challenges that may arise in this line of work. That said, make sure you share how you've solved any issues you've faced in your experience.

Answer Example

"In my last job as a business intelligence analyst, I was often exposed to cross-functional teamwork. I've mostly worked with our HR and IT departments. In my experience, if the team is attuned to the needs of the company for that particular project, it can turn out to be a huge success. I do my best to communicate expectations clearly. In addition, I take into account that everyone has different work styles, strengths, and weaknesses. Usually, that largely depends on their expertise and job role."

8. Tell us about the last presentation you gave. In your opinion, how did it go?

How to Answer

As a business intelligence analyst, giving presentations to the executives of your company or the company's clients, will be an important part of your work. You'll often be expected to extract the insights from the data, prepare the presentation, along with compelling visuals and dashboards, and then deliver it – all by your own efforts. If you have plenty of experience, discuss the topic of your presentations and the feedback you received. If you're straight out of college, think of a presentation you had to prepare as a part of your education. Of course, it would be more than great if you have a sample of your best presentation on your phone or tablet to show to the hiring manager.

Answer Example

"One of the presentations I'm proud of was related to the launching of a client's new app. I had to share the results from the preliminary user testing. What I came up with was an engaging presentation with lots of eye-catching visuals. I believe the latter, together with intriguing content, is key to a well-received presentation. I highlighted both the areas of strength and the areas of improvement. After that, I shared some actionable tips for product improvement with the client. The feedback was positive, and I can actually show you a copy of my presentation on my tablet."

9. What does the acronym INVEST stand for?

How to Answer

As a business intelligence analyst, you should understand what the acronym INVEST means to technical teams and product managers. It stands for:

– Independent
– Negotiable
– Valuable
– Estimable
– Sized appropriately
– Testable

If you're familiar with the term, break down each word to show the interviewers you know what you're talking about. If not, make sure you show interest in understanding the concept and which industries mostly use it.

Answer Example

"I've mostly worked in the banking and telecommunications fields. My business analysis was mostly done on the strategic side, and I have limited exposure to this term. I know INVEST is mostly used by business intelligence analysts collaborating with IT and developers teams. As far as I remember, it stands for Independent, Negotiable, Valuable, Estimable, Sized appropriately, and Testable. I'll be happy to gain better knowledge about INVEST and how it is utilized in your company."

10. Are you Six Sigma certified? Do you think that's important and why?

How to Answer

A Six Sigma certification is not a must, but it's certainly a plus for a BI analyst. Six Sigma certifications have different levels, starting from white belt through yellow, green and black belts to master black belt and champion belt. If you have completed the training, talk about your experience, the skills you've acquired, and how you apply them in your job as a BI analyst. If not, share your perspective on why you would consider taking the training.

Answer Example

"Although I haven't started any Six Sigma training yet, I'm aware that expertise in lean management will certainly be helpful to my clients, as I build up my professional portfolio. So, earning a Six Sigma certification is definitely an option I intend to explore in the future."

11. What does the acronym PEST mean?

Have you used it in your business intelligence experience?

How to Answer

The acronym PEST stands for Political, Economic, Social, and Technological. A PEST analysis is a strategic business tool that allows BI analysts to discover, evaluate, organize, and track macro-economic factors that can influence their business and make them more competitive in the future. If you're experienced in the business intelligence field, you should have some knowledge of PEST and how it works.

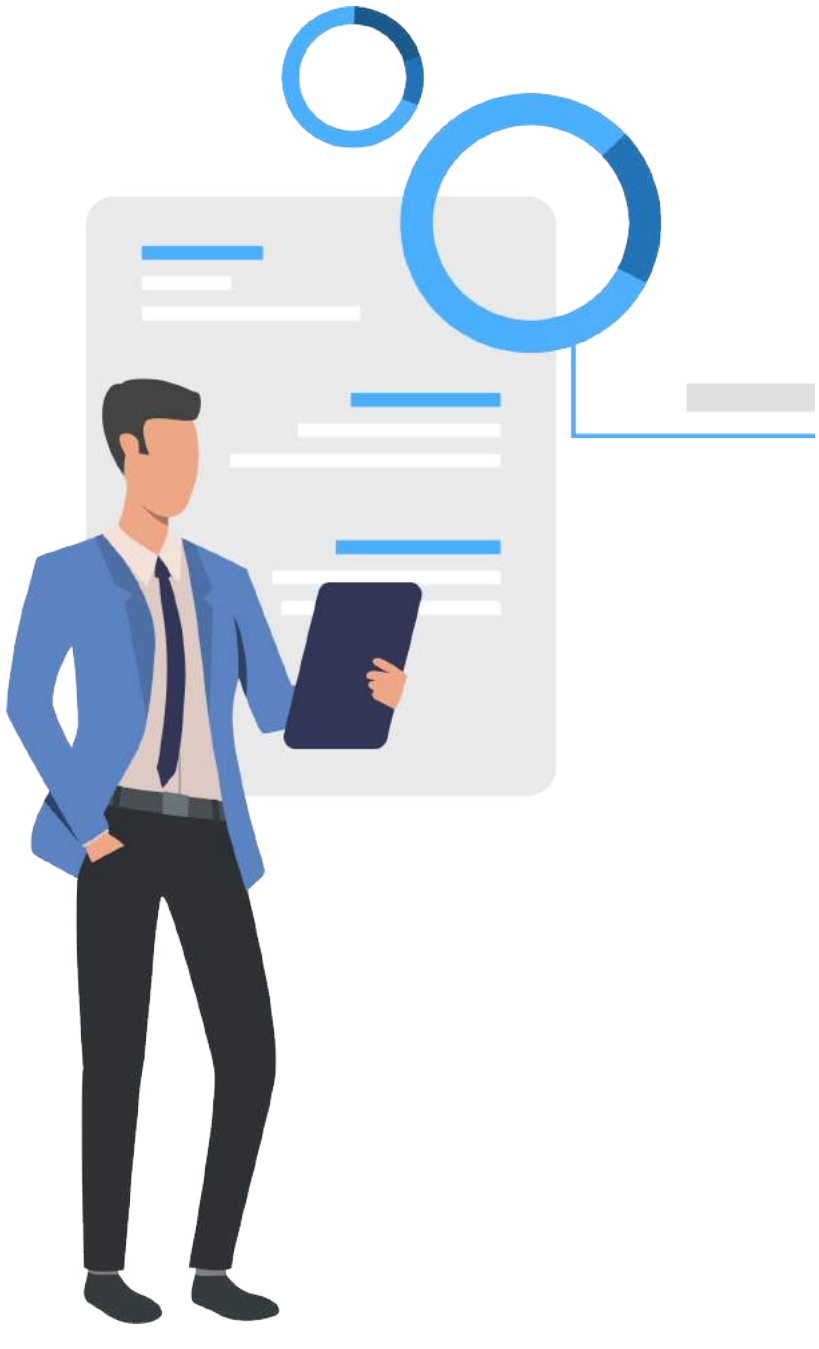
However, if you haven't had the chance to employ PEST in your work experience, show the hiring manager you have a basic idea of the concept and that you're more than willing to apply this form of analysis in your future job.

Answer Example

"I am just starting my career in business intelligence, so I haven't applied PEST analysis in my work just yet. Nevertheless, I've implemented PEST in a case study while in college. I had to discover the political, economic, social, and technological factors affecting the airline industry in recent years. I think it's a really efficient type of analysis and I'd be happy to become proficient in it in the future."

TECHNICAL BI ANALYST INTERVIEW QUESTIONS

The technical part of the BI analyst job interview comprises some questions on data modeling software, along with inquiries about statistics, programming, and decision-making techniques. Here's what you should have in mind:



12. Which data modeling software do you prefer to use?

How to Answer

BI analysts mostly use Microsoft Excel or Power BI for their data modeling needs. The required or preferred tools will be most probably listed in the company's job description, so it would be best to refer to those. If you have relevant experience, share your level of expertise with the interviewer. In case you lack exposure to their preferred software or programs, explain how you can incorporate your skills into their systems.

Answer Example

"I do most of my data modeling in Excel, as I find it most convenient for data mapping. I have some exposure to Power BI, as well. However, I believe I can benefit from sharpening my skills in that program. That's why I'm currently taking a Power BI online training."

13. What specific technical skills do you have as a BI analyst?

How to Answer

Your BI analyst experience and skillset are closely related to the focus of your career. Depending on whether you are a data BI analyst, an IT BI analyst, or

a strategic BI analyst, your answer to this question will be different.

If you're applying for a data-focused role, your technical skillset may include proficiency in data analysis software and visual presentation tools, such as Power BI. A BI analyst in the IT field would probably have exposure to some software development programs. While a strategic BI analyst would be well-familiar with business case analysis software and applications. Taking that into consideration, share with the interviewer the technical skills you will bring to the company.

Answer Example

"As a data BI analyst, I've been exposed to data mining and big data software, such as LIONsolver and Oracle. I'm highly skilled in Microsoft Excel which I use for data modeling and Power BI where I create rich visuals and client presentations."

14. Specify two important chart types in your BI analyst arsenal. Why do you find them important?

How to Answer

The hiring authority wants to see that you have basic knowledge when it comes to the diagrams and charts that you will be using during your business analyst career. Some examples include:

The interviewer wants to check your basic knowledge of charts you'll be using in your BI analyst's tasks.

Some examples include:

- area charts;
- bar charts;
- clustered column charts;
- combo charts;
- doughnut charts;
- funnel charts;
- gauge charts;
- line charts;
- pie charts;
- scatter plots;
- waterfall charts.

You are probably familiar with most of these, so just choose two options which you have experience with and can easily talk about.

Answer Example

"The two types of charts I use most often are area charts and bar charts. In my role as BI analyst, area charts have helped me display where a specific trend is headed in the future, which, in turn, makes planning easier. Bar charts, on the other hand, can show clearly which products are most popular among customers or display the number of unique visitors on a landing page based on various criteria."

If you want to learn more about the different chart types, read our article on visualizing data through charts. And, by the way, if you're finding this answer useful, consider sharing this article, so others can benefit from it, too. Helping fellow aspiring BI analysts reach their goals is one of the things that make the data science community special.

15. How would you define benchmarking and why do you consider it important?

How to Answer

Benchmarking is the practice of evaluating and comparing the business processes in a company with the best competitors' practices and use these insights to set standards and make improvements to your company's business performance. When a BI analyst is benchmarking, they study various metrics and processes, such as product development, manufacturing procedures, and more. Discuss with the interviewer how you use benchmarking to help your company achieve its goals.

Answer Example

"Benchmarking is an important practice of comparing your business against other businesses that are already very successful. It's like a smart, analytical comparison. I believe it's essential to benchmark when a company is looking at making a significant change, are seeing a loss of revenue, are anticipating the launch of a new product, or need to recalibrate their business operations in one way or another."

16. How do you differentiate between a risk and an issue?

How to Answer

If you're an experienced BI analyst, you know for sure there's a tremendous difference between real risk and an issue. The interviewer wants to check if you can be mindful of probability, while, at the same time, stay focused and hands-on when it comes to current issues.

Answer Example

"In my role as a business intelligence analyst, my focus is more on risk than issues. I view risk as a predicted problem that could come up in the future, so it's up to me to assess this risk and help my clients prevent it. An issue, on the other hand, is a risk that has already happened. In such cases, I can advise my clients on how to do damage control. But I'd strongly prefer helping them avoid the issue altogether."

17. What's your preferred decision-making technique?

How to Answer

The interviewer wants to see what you know about decision-making and what techniques you use to arrive at reliable conclusions in your projects. Some of the common decision-making techniques are T-Chart Analysis, Pareto Analysis, a.k.a. the 80/20 rule, etc. Discuss the techniques you utilize with the interviewer and the reasons for your preferences.

Answer Example

"I don't limit myself to one technique only. In decision-making, my choice depends largely on the stage of the project. That said, sometimes, I use a variety of techniques within the same project, such as Pareto Analysis, T-Chart Analysis, SWOT Analysis, or decision trees. All of these help me resolve certain issues and come to a decision."

18. Explain Selection Bias.

Selection bias is the bias introduced by the selection of individuals, groups or data during the sampling process when randomization is not achieved. This means that the sample we created does not represent the general population properly. It is called 'selection' because it refers to the 'sample selection'.

Selection bias is a very broad term which encompasses many different biases.

Here are some examples:

- **Sampling bias** – or also known as 'sample selection bias' occurs when not everyone in the general population has an equal chance to be in the sample. Let's say that we want to make a survey about students in one university. We can go to the university, enter random classrooms and ask all of them to participate in our survey. Great, right? Well, not exactly. There are two main issues:
  - » We assumed that everyone who is a student at the university will be present at the chosen time and date. And that's never the case, since students don't have lectures every day, work part-time, get sick or even go on holiday.
  - » We also expected that everyone who is present will want to answer the survey, which is a very optimistic assumption. Students can be forced to, but in that case, one shouldn't expect a great quality of answers.
- **Length time bias** – occurs when different observations in our sample have different development cycles. The most common case is when we are dealing with diseases like cancer. Some types of cancer develop faster than others, so examining 6 months of disease development for 2 different individuals may result in one having no change, while the other passing away. The reason is that they may be in different stages of cancer or different organisms react differently to the disease so time is extremely problematic in general for our sample.
- **Exposure** – This one is very common and problematic. Imagine you've got funding to explore everything about a group of your customers, for instance, female customers. You complete your study and everything is great. However, you are then asked to conduct another study – about the shopping habits of your customers overall. Problem is – you don't have data on the male customers. Using only the female data would lead you to some results, but they would definitely be problematic. If you are not provided with male data and are forced to complete the study, you will experience exposure bias. This is not uncommon when resources are limited – sampling has been done once and nobody wants to pay for another sampling process.
- **Data** – there are several popular examples, but one such case is removing outliers with correct data. Usually, we remove the outliers to get better results, but sometimes some important patterns may be contained there.

- **Studies** – occurs when we use only the studies, where we have reached a good result. Often, we have formed a hypothesis and we look for studies that support it. In this effort, one could be misled to reference only papers that support their claim and thus introduce a bias. Note that in general academia is extremely biased in that regard. There is research that suggests that papers that reach results are 4 times more likely to be published than papers that have non-satisfactory results. This is very problematic as we know that determining there is 'no effect' is still a valid result.
- **Attrition** – it is related to survivorship bias. The most common example that everyone uses is businesses (e.g. startups). All companies that could be studied are successful (profitable ones). Those that are non-successful seize to exist and cannot be studied.
- **Observer** – is the tendency to see what we expect to see – meaning we have already "decided" what outcome we want and we strive for them to be right. As you can sense this one is related to the 'studies bias' from above.

Okay. So, there are a lot of problems around selection bias, right?

The best way to get rid of selection bias… is to not introduce our sample to it. No joke, right? What we mean by that is that selection bias is formed in the sampling process. If we are not careful when we collect our data, our analysis will definitely be flawed.

20. What is Kano Model Analysis and why is it important?

How to Answer

The Kano Model Analysis taps into customer's emotions and needs to improve product development.

It helps a company add certain features to their product that would increase customer satisfaction without costing a fortune. According to the Kano Model Analysis, there are 3 types of attributes to products:

- Basic attributes;
- Performance attributes;
- Excitement attributes, a.k.a. Delighters.

When answering this question, demonstrate that you're not only familiar with the three points of satisfaction but you also know how they act together to help customer satisfaction analysis.

Answer Example

"Kano Analysis is of major importance to developing new products and services. It helps companies understand customer needs and make sure they have a competitive edge before launching them on the market. The threshold attributes are the basic features a customer expects from the product. The performance attributes also called "satisfiers", are additional features that increase customer satisfaction. And "delighters" are the elements of surprise that can really increase the product's competitive edge."

21. What are the most important SDLC models?

How to Answer

SDLC denotes Software Development Life Cycle. It's a concept in IT that is often employed by BI analysts in the field. In its essence, SDLC is a process that starts with the decision to launch a product and ends with the full removal of the software product from exploitation. There are various types of SDLC models, each predetermined by the software product type in development. The most popular are:

- Waterfall model;
- Iterative model;
- Spiral model;
- V-shaped model;
- Agile model.

Even if you have no experience in the field, show the interviewer that you understand the differences between the models by briefly outlining them.

Answer Example

"Although I don't have practical experience with ADLC models, I learned in college that there are 5 primary SDLC model types: Waterfall, Iterative, Spiral, V-shaped, and Agile. The Agile model is related to flexibility and adapting to change. The Iterative model refers to the "incremental build" approach in large development efforts. I'm less familiar with the rest but I would enjoy diving deeper and learning more."

22. How would you create a taxonomy to identify key customer trends in unstructured data?

First, you have to understand the company's objectives prior to categorizing the data. Once you've done this, it is always good to follow an iterative approach by pulling new data samples and improving the model accordingly. And you do this by validating it for accuracy through solicited feedback from the stakeholders of the business. This helps ensure that your model is producing actionable results and improving over time.

23. What is an SQL View?

A view is a virtual table whose contents are obtained from an existing table or tables, called base tables. The retrieval happens through an SQL statement, incorporated into the view. So, you can think of a view object as a view into the base table. The view itself doesn't contain any real data; the data is electronically stored in the base table. The view simply shows the data contained in the base table.

BEHAVIORAL QUESTIONS

Many of the behavioral questions for BI analysts are related to your ability to work efficiently with clients and upper management, and how you address complex issues that may arise during larger projects.

24. How do you demonstrate to your clients the importance of dialogue during a project?

How to Answer

When it comes to clear communication through every stage of a project, leading by example is key. As a business intelligence analyst, it is your job to establish the tone of the dialogue and the means of communication. Show the interviewer that you know how to set the foundations of proper communication with your clients and their teams. If possible, give examples of projects you've worked on.

Answer Example

"As a business intelligence analyst, I like to keep everyone in the loop about the development of a project. I often promote the use of project management apps that make collaboration easier and gives access to every detail of the project at any stage."

25. As a business analyst, when do you regard a project as complete?

How to Answer

A great business analyst knows that when a client signs off a project, it doesn't mean it's successful (or finished) yet. So, make sure you explain to the interviewer that you remain available to your clients and you support them until you're sure their expectations are met and they are happy with the results.

Answer Example

"As a BI analyst, I always make sure there are no unresolved issues when the client signs off a project. Nevertheless, I'm available in case their expectations aren't fully met and I need to make adjustments to deliver what has been promised. However, this rarely happens once there are no outstanding invoices and documentation is archived."

26. How often do you brainstorm new ideas with your coworkers?

How to Answer

Having regular discussions with other team members is of great importance when it comes to project plans and aligning ideas. Let the interviewer know that you're a team player who is open to others' views and opinions.

Answer Example

"I believe learning from each other's working styles and approaches is invaluable for any project. I support the collaborative spirit in my team and I'm sure we always come up with better ideas together rather than individually."

27. Is there a case in your experience when you broke a confidentiality agreement?

How to Answer

Confidentiality agreements ensure the protection of company trade secrets. This question gives you a chance to present yourself as a trustworthy individual that abides by their company's policy and respects their clients.

Answer Example

"I have signed NDAs on countless occasions in my career as a business intelligence analyst. When working on a project, confidentiality is one of my team's top priorities. None of us has broken the trust of our company and clients."

28. How do you respond when you're unhappy with the end result of a project?

How to Answer

Even the best BI analysts experience failure at times. Not all projects are perfect, and not all clients can be satisfied. What the interviewer would like to know is if you're capable of accepting disappointment and responding in a mature and productive way.

Answer Example

"I think business intelligence requires perfectionism at all times. When I'm not happy with my performance, or I make a mistake, I take a step back and take my time to fine-tune my work before submitting it."

29. How do you plan to improve yourself professionally this year?

How to Answer

Employers are seeking BI analysts who are constantly upgrading their skills and strive to stay relevant. You can set career development goals and accomplish them by attending conferences, earning online certificates, listening to podcasts, or even joining a mentoring program. When you mention some of these examples and the goals you've set for yourself this year, make sure you bridge the knowledge you'll gain with the benefits you'll bring to the company.

Answer Example

"This year, I've enrolled in a Power BI online course to refresh my expertise, and I've also signed up for a few TDWI seminars in Predictive Analytics and Data Modeling. I can't wait to take my skills to another level and, hopefully, apply what I've learned as a BI analyst in your company."

30. What do you do if you disagree with someone at work?

This one is part of the business analyst behavioral interview questions and answers. It is perfectly normal to disagree with someone at work. Similar situations occur all the time. When answering this question, do not speak about the person that you disagreed with. This is really important, as you do not want to come off as someone who bad mouths people; this is unprofessional. The Hiring Manager is not interested in learning saucy details about the bad habits of that other person. Instead, he/she wants to know more about your conflict management abilities. He/she is eager to learn whether you are an active listener and whether you are good at persuading people.

Every behavioral question comes together with a story that supports the answer given by the candidate.

When you answer this question, try to think of a disagreement that was not personal, but derived from different views regarding the execution of a certain task. It is much safer to have this type of disagreement, as it does not suggest you are someone that is difficult to work with.

There are a few key points which you should concentrate on:

- You listened actively
- looked for the best possible solution
- had at heart the team's success rather than showing muscles
- were persuasive

Here's an example of such a situation:

You and three of your classmates were asked to prepare a Business Plan.

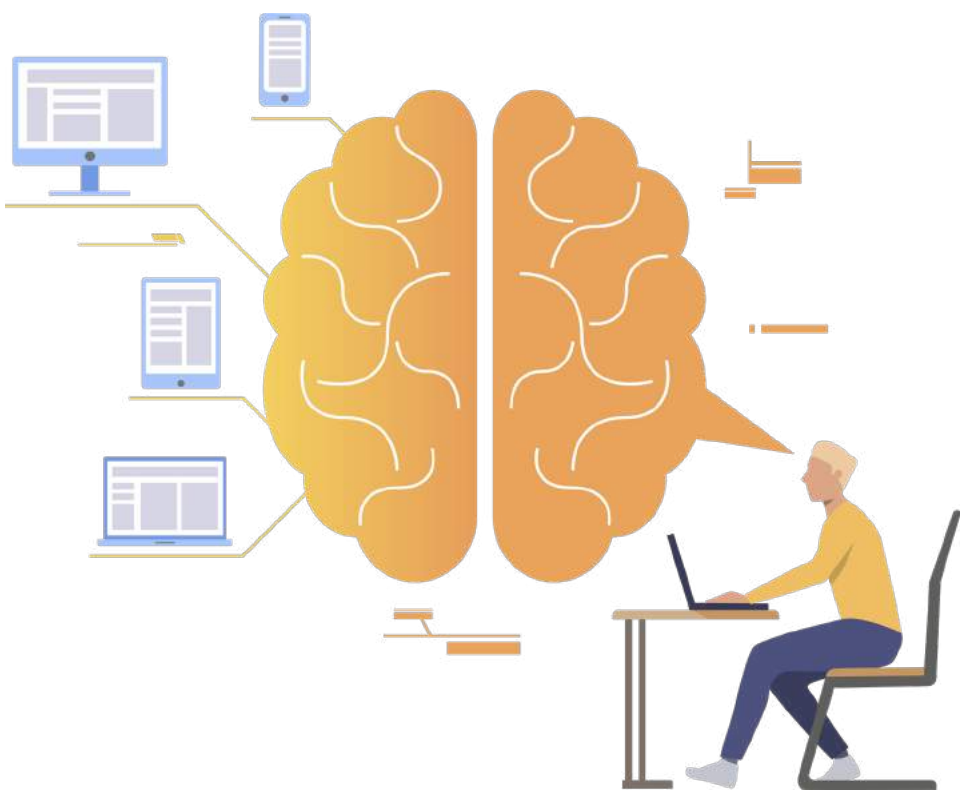
Probably the most important part of the whole Business Plan is the prediction of the top-line – revenues. You wanted to use a bottom-up approach and one of your classmates thought that the top-down approach would be more useful. Both you and your colleague were convinced that your own approach

was correct. The work could not continue before resolving this issue.

So, you asked your classmate to elaborate on his point and demonstrated that you are interested in his idea; he made a valid point. There was a recent market study that your team could use as a reference. It predicted the overall dimensions of the market for the next five years. This was a valuable piece of information, although it is difficult to predict the firm's market share. You explained that the advantage of the bottom-up approach is that you can base your growth assumptions on historical data and incorporate data that is specific for the firm under consideration. After each of you explained your points of view, you came to the conclusion that the best thing to do is to use both approaches and obtain a range that would indicate the company's revenues.

BRAINTEASERS

As a BI analyst, you should be able to apply logical thinking at all times, especially under stress. Here's an example of the type of question the interviewer is likely to challenge you with...



31. Imagine the following situation:

You have 100 balls (50 red balls and 50 blue balls) and 2 buckets. You can choose how to divide the balls into the two buckets so as to maximize the probability of selecting a blue ball if 1 ball is chosen from 1 of the buckets at random.

Put 1 blue ball in one of the buckets and put the rest of the balls in the other bucket. This way you will have 50% chance of selecting the bucket with only 1 ball and then, even if it is not selected and you have to draw a ball from the other bucket you would have almost 50% chance of selecting a blue ball (49 blue balls versus 50 red balls). The joint probability of the two events would equal almost 75%.

GUESSTIMATE

One way to prepare a business intelligence analyst for a real business situation case question is through a warm-up guesstimate. Here's one that will give you a good idea of what to expect...

32. How would you estimate the weight of the Chrysler building?

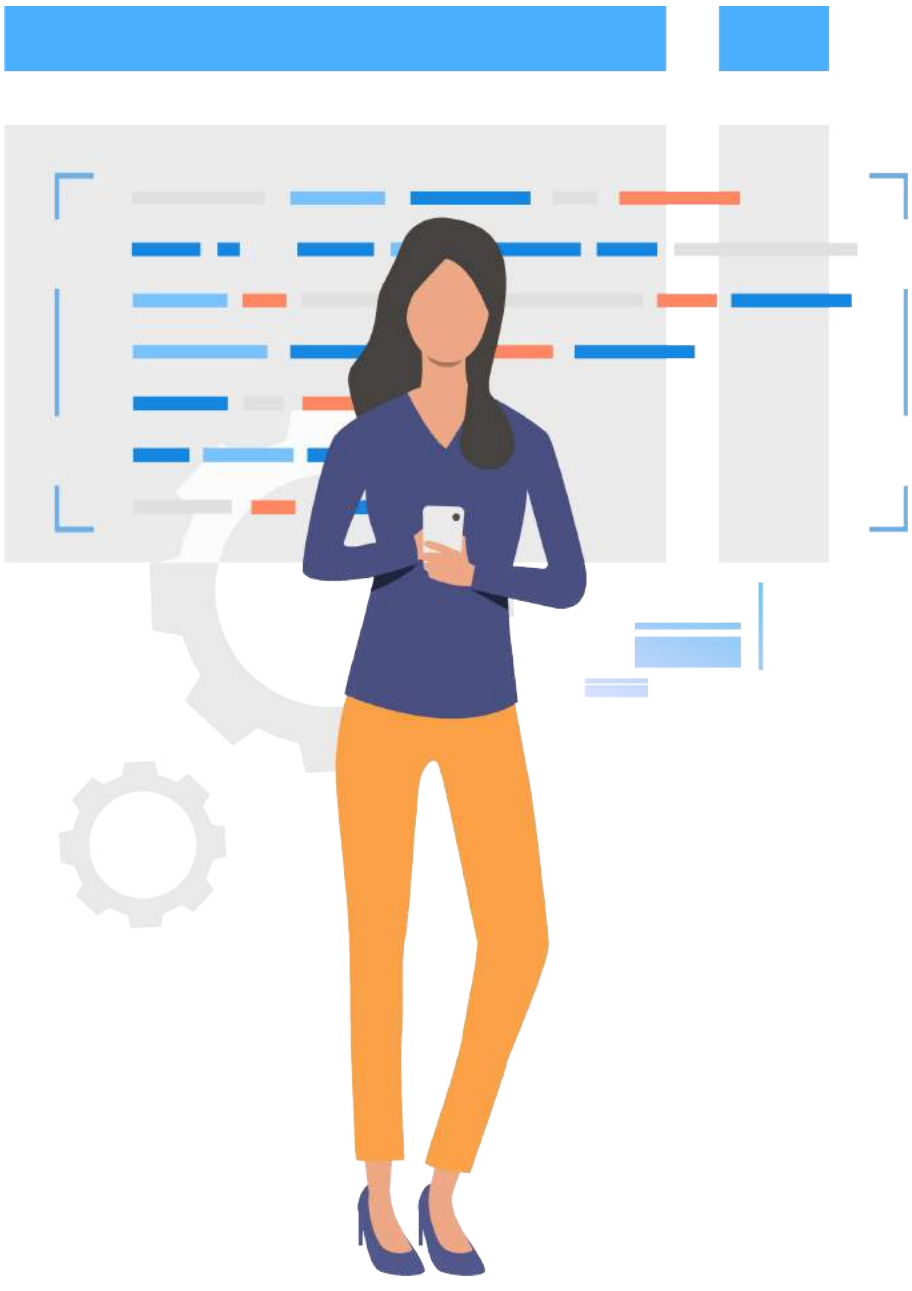
This is a process guesstimate - the interviewer wants to know if you know what questions to ask. First, you would find out the dimensions of the building (height, weight, depth). This will allow you to determine the volume of the building. Does it taper at the top? (Yes.) Then, you need to estimate the composition of the Chrysler building. Is it mostly steel? Concrete? How much would those components weigh per square inch? Remember the extra step - find out whether you're considering the building totally empty or with office furniture, people, etc? (If you're including the contents, you might have to add 20 percent or so to the building's weight.)

DATA ENGINEER
INTERVIEW QUESTIONS

Data engineer interview questions are a major component of your interview preparation process. So, if you want to maximize your chances of landing a data engineer job, you should be able to answer both the specific technical questions and the tricky behavioral ones with poise and confidence. Here's a comprehensive list of data engineer questions and answers that will help you achieve that.

GENERAL DATA ENGINEER INTERVIEW QUESTIONS

Usually, interviewers start the conversation with a few more general questions. Their aim is to take the edge off and prepare you for the more complex data engineering questions ahead. Here are a few that will help you get off to a flying start...



1. How did you choose a career in data engineering?

How to answer

The answer to this question helps the interviewer learn more about your education, background and work experience. You might have chosen the data engineering field as a natural continuation of your degree in Computer Science or Information Systems. Maybe you've had similar jobs before, or you're transitioning from an entirely different career field. In any case, don't shy away from sharing your story and highlighting the skills you've gained throughout your studies and professional path.

Answer Example

"Ever since I was a child, I have always had a keen interest in computers. When I reached senior year in high school, I already knew I wanted to pursue a degree in Information Systems. While in college, I took some math and statistics courses which helped me land my first job as a Data Analyst for a large healthcare company. However, as much as I liked applying my math and statistical knowledge, I wanted to develop more of my programming and data management skills. That's when I started looking into data engineering. I talked to experts in the field and took online courses to learn more about it. I discovered it was the ideal career path for my combination of interests and skills. Luckily, within a couple of months,

a data engineering position opened up in my company and I had the chance to transfer without a problem."

2. What do you think is the hardest aspect of being a data engineer?

How to answer

Smart hiring managers know not all aspects of a job are easy. So, don't be hesitate to answer this question honestly. You might think Its goal isn't to make you pinpoint a weakness. But, in fact, what the interviewer wants to know is how you managed to resolve something you struggled with.

Answer Example

"As a data engineer, I've mostly struggled with fulfilling the needs of all the departments within the company. Different departments often have conflicting demands. So, balancing them with the capabilities of the company's infrastructure has been quite challenging. Nevertheless, this has been a valuable learning experience for me, as it's given me the chance to learn how these departments work and their role in the overall structure of the company."

3. Can you think of a time where you experienced an unexpected problem with bringing together data from different sources?

How did you eventually solve it?

How to answer

This question gives you the perfect opportunity to demonstrate your problem-solving skills and how you respond to sudden changes in the plan. The question could be data-engineer specific or a more general one about handling challenges. Even if you don't have this particular experience, you can still give a satisfactory hypothetical answer.

Answer Example

"In my previous work experience, my team and I have always tried to be prepared for any issues that may arise during the ETL process. Nevertheless, every once in a while, a problem will occur completely out of the blue. I remember when that happened while I was working for a franchise company. Its system required for data to be collected from various systems and locations. So, when one of the franchises changed their system without prior notification, it created quite a few loading issues for their store's data. To deal with this issue, first I came up with a short-term solution to get the essential data into the company's corporate wide-reporting system. Once I took care of that, I started developing a long-term solution to prevent such complications from happening again."

4. Data engineers collaborate with Data Architects on a daily basis. What makes your job as a data engineer different?

How to Answer

With this question, the interviewer is most probably trying to see if you understand how job roles differ within a data warehouse team. However, there is no "right" or "wrong" answer to this question. The responsibilities of both data engineers and data architects vary (or overlap) depending on the requirements of the company/database maintenance department you work for.

Answer Example

"Based on my work experience, the differences between the two job roles vary from company to company. Yes, it's true that data engineers and data architects work closely together. Still, their general responsibilities differ. Data architects are in charge of building the data architecture of the company's data systems and managing the servers. They see the full picture when it comes to the dissemination of data throughout the company. In contrast, data engineers focus on testing and maintaining of the architecture, rather than on building it. Plus, they make sure that the data available to analysts within the organization is reliable and of the necessary high quality."

5. Can you tell us a bit more about the data engineer certifications you have earned?

How to Answer

Certifications prove to your future employer that you've invested time and effort to get formal training for a skill, rather than just pick it up on the job. The number of certificates under your belt also shows how dedicated you are to expanding your knowledge and skillset. Recency is also important, as technology in this field is rapidly evolving, and upgrading your skills on a regular basis is vital. However, if you haven't completed any courses or online certificate programs, you can mention the trainings provided by past employers or the current company you work for. This will indicate that you're up-to-date with the latest advancements in the data engineering sphere.

Answer Example

"Over the past couple of years, I've become a certified Google Professional Data Engineer, and I've also earned Cloudera Certified Professional credential as a Data Engineer. I'm always keeping up-to-date with new trainings in the field. I believe that's the only way to constantly increase my knowledge and upgrade my skillset. Right now, I'm preparing for the IBM Big Data Engineer Certificate Exam. In the meantime, I try to attend big data conferences with recognized speakers, whenever I have the chance.

6. Explain data import in R.

R reads data from a decent number of sources, like text, Excel, SPSS, SAS, Stata, systat… with text, and more specifically, CSV, being the most popular. Depending on the format of the data, you'd need to use different packages to import it into R.

In terms of syntax, there is nothing too shocking about the operations – a standard read call is used in most situations.

Importing text files is fairly straightforward.

The user can use the barebones read.table() function from the built-in {utils} package, and set all relevant arguments, or opt for using read.csv() which has default values for the arguments most often used in importing a CSV file. Both of these would result is you creating a data frame. You could also choose to use the read_csv() from the {tibble} package and import your data as a tibble. That's the method to be preferred if you're using R to do data science.

Importing Excel files happens with the {xlsx} package.

Importing SPSS and SAS data often requires the {Hmisc} package. For .sas7bdat files specifically, Hadley Wickham's {haven} package can be helpful.

Importing Stata and systat data typically happens with R's {foreign} package.

7. What is the difference between UNION and UNION ALL?

The UNION command is very similar to the JOIN command, as they are both used to select related information from multiple tables. However, the UNION command selects only columns of the same data type. Furthermore, UNION selects distinct values only, i.e. it combines the result set of two or more SELECT statements. In contrast, UNION ALL selects all values (without eliminating duplicate rows).

8. What programming/scripting languages have you used?

Which one are you most experienced with?

How to Answer

Generally, job descriptions list the required and preferred programming skills for the role. So, when you talk about the languages you're most experienced with, make sure you emphasize your work with the preferred/required ones in past projects. In case you lack experience in these, focus on the languages you're proficient in and list any similarities they may have with the required. And don't forget to point out that you're a fast learner that can easily grasp new concepts and languages. This will show the interviewer that you'll be committed to using the necessary tools, even if you have to complete additional training.

Answer Example

"I have worked with both Python and SQL. However, I'm most comfortable using Python, due to the

nature of the tasks in the previous company I worked for. I understand that SQL is preferred, and I can assure you I can advance my SQL skills quickly on the job. I'm a quick learner and learning new concepts has always come easy to me."

9. Have you ever found a new use for existing data that has brought a positive change in your employer's business?

How to Answer

A data engineer is often one of the few people who has the broadest view of the company's data. It's quite common for departments to work with a limited set of tables within the organization's databases and thus hinder the accuracy of their analyses. That said, a good data engineer should be familiar with the projects and initiatives of each department. This will allow them to provide other employees with valuable insight into what data is available and how they can utilize it to improve the quality of analyses throughout the organization."

Answer Example

"As a data engineer, it's important for me to be familiar with all initiatives taken up by the company's departments. I believe employers should have access to data from other departments in order to improve their work. In my previous job, I proposed to connect employee data with sales data. As it turned out, there was a correlation between the education and work experience of hired employees and high or low sales periods. The subsequent detailed analysis showed that certain employee profiles result in considerable increases in sales for a significant period of time. I take pride in this discovery, as HR data had never been cross-referenced with sales data for analytical purposes in this company before."

10. What is your experience in working with data scientists?

What do you think are the common skills you share?

How to Answer

It's true that data engineers and data scientists have some skills and qualifications in common. It's also possible that they have some overlap in responsibilities, depending on the requirements of the employer. But, in their essence, their roles are quite different.

Data engineers should be aware of the data scientists' ongoing projects. They handle the maintenance, architecture, and preparation of data for future analysis.

Data scientists, on the other hand, rely on the data engineers' work to extract insights from the data and present the results to management and executives.

That said, a hiring manager would like to know how well you understand the work of data scientists and what is your experience interacting with them.

Answer Example

"I've had the chance to work with data scientists on many projects and occasions and I can say it's been a very productive and rewarding experience. We both understand analytics and programming

languages which made it easy for me to help them with their projects. The fact that we have overlapping skills allowed the data scientists to grasp the limitations of our infrastructure and data availability. At the same time, I was able to easily understand their data needs."

11. What is your preferred field of work? Do you prefer Pipeline or Database, or a more Generalist role?

How to Answer

A data engineer's role heavily depends on the size of the company and the specific tasks they're assigned. Generalists employ a variety of skills, as they are responsible for many different tasks. If you're focused on Pipeline, this means you have experience in working closely with data scientists and have a better understanding of how to prepare data for analysis. Data engineers who have worked mostly in Database, have in-depth knowledge of the ETL process and table schemas. No matter which role/s you have been in, include all your experiences in your answer. You can also go into moderate detail in explaining why you prefer one type over the other.

Answer Example

"I've always worked in more of a Generalist role. I can say I like this one more than the other types because I like having a broader scope of expertise. I enjoy being in-the-know about the whole structure and process, as opposed to focusing on just one subset of skills I've acquired."

12. According to some Big Data professionals, data engineering is a non-analytical career path.

Do you consider this statement true or false? Why?

How to Answer

This statement can't be interpreted in a single way. Yes, it's true that compared to a data analyst, a data engineer's work is much less analytical in nature. However, this doesn't mean that data engineers lack analytical skills or that they don't implement them at all. When giving your answer, tell the hiring manager how you view your role as a data engineer and how you've used your analytics skills on the job."

Answer Example

"I'd have to say I firmly disagree with this statement. I've used my analytical skills on numerous occasions. As a data engineer, I've often performed analyses to ensure the high quality and integrity of the data. My analytical skills have also helped me immensely in my mutual projects with data scientists and data analysts. Thanks to my analytical mindset, I've been able to identify and help them with their data needs."

13. What trainings would you enroll in to advance your data engineering skills?

How to Answer

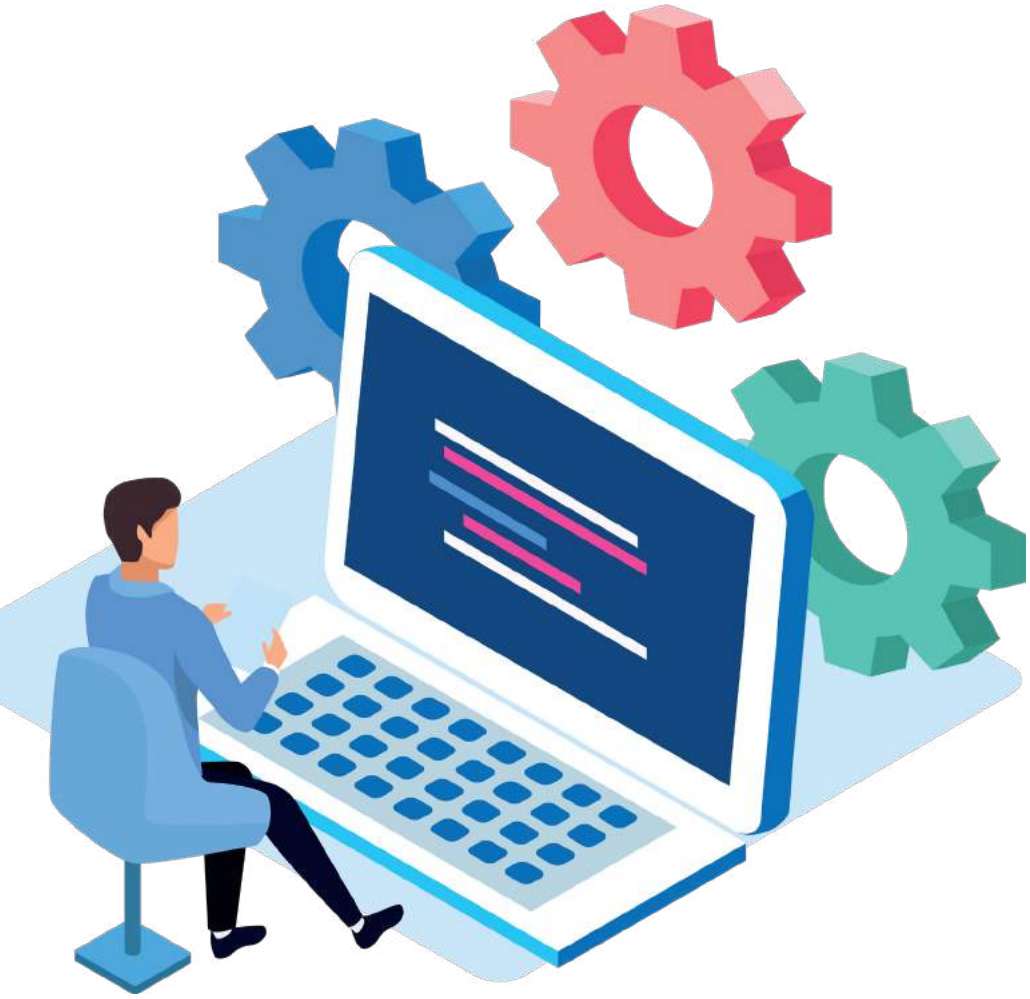
Technology's constantly changing, so, if you're setting high goals for yourself, this question may prompt you to list several trainings you'd like to fit in your schedule. However, make sure you convey that you'd like to complete these courses as they cover topics of interest and not to make up for weaknesses in your preparation. Balance your answer by mentioning your strengths and the skills you've already acquired.

Answer Example

"I think enrolling in trainings is crucial for any data engineer that wants to be up-to-date with the advancements in the industry. Personally, I'd like to expand my current expertise in ETL processes and the cloud environment. Although I have significant experience working with both, I believe my future work can only benefit from continuous learning."

TECHNICAL DATA ENGINEER INTERVIEW QUESTIONS

The technical data engineer questions help the interviewer assess 2 things: whether you have the skills necessary for the role; and if you're experienced with (or willing to advance in) the systems and programs utilized in the company. So, here's a list of technical questions you can practice with...



14. Which ETL tools have you worked with? Do you have a favorite one? If so, why?

How to Answer

The hiring manager needs to know that you're no stranger to the ETL process and you have some experience with different ETL tools. So, once you enumerate the tools you've worked with and point out the one you favor, make sure to substantiate your preference in a way that demonstrates your expertise in the ETL process.

Answer Example

"I have experience with various ETL tools, such as IBM Infosphere, SAS Data Management, and SAP Data Services. However, if I have to pick one as my favorite, that would be Informatica's PowerCenter. In my opinion, what makes it the best out there is its efficiency. PowerCenter has a very top performance rate and high flexibility which, I believe, are the most important properties of an ETL tool. They guarantee access to the data and smoothly running business data operations at all times, even if changes in the business or its structure take place.

15. Have you built data systems using the Hadoop framework?

If so, please describe a particular project you've worked on.

How to Answer

Hadoop is a tool that many hiring managers ask about during interviews. You should know that whenever there's a specific question like that, it's highly likely that you'll be required to use this particular tool on the job. So, to prepare, do your homework and make sure you're familiar with the languages and tools the company uses. More often than not, you can find that information in the job description. If you're experienced with the tool, give a detailed explanation of your project to highlight your skills and knowledge of the tool's capabilities. In case you haven't worked with this tool, the least you could do is do some research to demonstrate some basic familiarity with the tool's attributes.

Answer Example

"I've used the Hadoop framework while working on a team project focused on increasing data processing efficiency. We chose to implement it because of its ability to increase data processing speeds while, at the same time, preserving quality through its distributed processing. We also decided to implement Hadoop because of its scalability, as the company I worked for expected a considerable increase in its data processing needs over the next few months. In addition, Hadoop is an open-source network which made it the best option, keeping in mind the limited resources for the project. Not to mention that it's Java-based, so it was easy to use by everyone on the team and no additional training was required."

16. Do you have experience with a cloud computing environment?

What are the pros and cons of working in one?

How to Answer

Data engineers are well aware that there are pros and cons to cloud computing. That said, even if you lack prior experience working in cloud computing, you must be able to demonstrate a certain level of understanding of its advantages and shortcomings. This will show the hiring manager that you're aware of the present technological issues in the industry. Plus, if the position you're interviewing for requires using a cloud computing environment, the hiring manager will know that you've got a basic idea of the possible challenges you might face.

Answer Example

"I haven't had the chance to work in a cloud computing environment yet. However, I have a good overall idea of its pros and cons. On the plus side, cloud computing is more cost-effective and reliable. Most providers sign agreements that guarantee a high level of service availability which should decrease downtimes to a minimum. On the negative side, the cloud computing environment may compromise data security and privacy, as the data

is kept outside the company. Moreover, your control would be limited, as the infrastructure is managed by the service provider. All things considered, cloud computing could be both right or wrong choice for a company, depending on its IT department structure and the resources at hand."

17. In your line of work, have you introduced new data analytics applications?

If so, what challenges did you face while introducing and implementing them?

How to Answer

New data applications are high-priced, so introducing such within a company doesn't happen that often. Nevertheless, when a company decides to invest in new data analytics tools, this could turn into quite an ambitious project. The new tools must be connected to the current systems in the company, the employers who are going to use them should be formally trained. Moreover, maintenance of the tools should be administered and carried out on a regular basis. So, if you have prior experience, point out the obstacles you've overcome or list some scenarios of what could have gone wrong. In case you lack relevant experience, describe what you know about the process in detail. This will let the hiring manager know that, if a problem arises, you have the basic know-how that would help you through.

Answer Example

"In my last position as a data engineer, I took part in the roll-out of a new data analytics application in my company. Introducing new applications takes careful planning to attempt a smooth transition to the new tool. With my experience, I found that even with careful planning, unforeseen circumstances can occur. In the early stages of the tool's introduction, we had an unexpected high demand for user licenses. This required reallocation of resources to purchase additional licenses as well as reprioritization of training schedules to meet the needs of different departments. We also had to ensure that our infrastructure could support the significantly higher number of employees utilizing the tool."

"As a data engineer, I've taken part in the introduction of a brand-new data analytics application in the last company I've worked for. The whole process requires a well-thought-out plan to ensure the smoothest transition possible. However, even the most careful planning can't rule out unforeseen issues. One of them was the high demand for user licenses which went beyond our expectations. The company had to reallocate financial resources to obtain additional licenses. Furthermore, training schedules had to be set up in a way that doesn't interrupt the workflow in different departments. In addition, we had to optimize our infrastructure, so that it could support the considerably higher number of users."

18. What is your experience level with NoSQL databases?

Tell me about a situation where building a NoSQL database was a better solution than building a relational database.

How to Answer

There are certain pros and cons of using one type of database compared to another. To give the best possible answer, try to showcase your knowledge about each and back it up with an example situation that demonstrates how you have applied (or would apply) your know-how to a real-world project.

Answer Example

"Building a NoSQL database can be beneficial in some situations. Here's a situation from my experience that first comes to my mind. When the franchise system in the company I worked for was increasing in size exponentially, we had to be able to scale up quickly in order to make the most of all the sales and operational data we had on hand. But here's the thing. Scaling out is the better option, compared to scaling up with bigger servers, when it comes to handling increases data processing loads. Scaling out is also more cost-effective and it's easier to accomplish through NoSQL databases. The latter can deal with larger volumes of data. And that can be crucial when you need to respond quickly to considerable shifts in data loads in the future. Yes, it's true that relational databases have better connectivity to various analytics tools. However, as more of those are being developed, there's definitely a lot more coming from NoSQL databases in the future. That said, the additional training some developers might need is certainly worth it.

19. What's your experience with data modeling?

What data modeling tools have you used in your work experience?

How to Answer

As a data engineer, you probably have some experience with data modeling. In your answer, try not only to list the relevant tools you have worked with, but also mention their pros and cons. This question also gives you a chance to highlight your knowledge of data modeling in general.

Answer Example

"I've always done my best to be familiar with the data models in the companies I've worked for, regardless of my involvement with the data modeling process. This is one of the ways I gain a deeper understanding of the whole system. In my work experience, I've utilized Oracle SQL Developer Data Modeler to develop two types of models. Conceptual models for our work with stakeholders, and logical data models which make it possible to define data models, structures and relationships within the database."

20. Have you ever taken part in a data disaster recovery situation?

If so, describe what happened and how you solved the issue at hand.

How to Answer

Completing daily assignments is only part of the data engineer's job. Above all, hiring managers are looking for someone who can quickly respond to urgent situations and contribute to their remedy. Sometimes a data infrastructure may fail. Or data can become inaccessible, lost, or even destroyed. All of these can hurt the company's processes. So, when answering this question, present yourself as a decisive person with a hands-on approach to solving unforeseen issues.

Answer Example

"In my most recent data engineer job, I was part of a team project focused on developing a Disaster Recovery Strategy. This is how I got familiar with the actions that needed to be taken when we faced a real data disaster recovery situation. A corrupt file somehow got loaded into the company's system. This caused databases to lock up. As a result, a lot of the data was corrupted as well. What I did was immediately approach the IT team. Together, we made sure our data backups were loaded as quickly as possible, so that the operations in the company can continue to run smoothly."

21. Have you ever created custom analytics applications?

If so, please share details about the application you've built.

How to Answer

In order to build a custom analytics application, a data engineer should have an in-depth understanding of the analytic needs of all departments within the company. Creating such applications requires careful planning and teamwork. That said, you should answer in a way that highlights not only your programming expertise but also your excellent communication skills.

Answer Example

"The goal of the custom application I built was to marry primary marketing research data with sales data that was stored in the company's databases. The app helped the Marketing department avoid the tedious process of requesting data from the data warehouse and loading it into Excel. This resulted in a much quicker performance of specific analyses."

BEHAVIORAL DATA ENGINEER INTERVIEW QUESTIONS

Behavioral data engineer interview questions give the interviewer a chance to see how you have handled unforeseen data engineering issues or teamwork challenges in your experience. The answers you provide should reassure your future employer that you can deal with high-pressure situations and a variety of challenges. Here are a few examples to consider in your preparation.

22. Data maintenance is one of the routine responsibilities of a data engineer.

Describe a time when you encountered an unexpected data maintenance problem that made you search for an out-of-the-box solution".

How to Answer

Usually, data maintenance is scheduled and covers a particular task list. Therefore, when everything is operating according to plan, the tasks don't change as often. However, it's inevitable that an unexpected issue arises every once in a while. As this might cause uncertainty on your end, the hiring manager would like to know how you would deal with such high-pressure situations.

Answer Example

"It's true that data maintenance may come off as routine. But, in my opinion, it's always a good idea to closely monitor the specified tasks. And that includes making sure the scripts are executed successfully. Once, while I was conducting an integrity check, I located a corrupt index that could have caused some serious problems in the future. This prompted me to come up with a new maintenance task that prevents corrupt indexes from being added to the company's databases."

23. Data engineers generally work "backstage".

Do you feel comfortable with that or do you prefer being in the "spotlight"?

How to Answer

The reason why data engineers mostly work "backstage" is that making data available comes much earlier in the data analysis project timeline. That said, c-level executives in the company are usually more interested in the later stages of the work process. More specifically, their goal is to understand the insights that data scientists extract from the data via statistical and machine learning models. So, your answer to this question will tell the hiring manager if you're only able to work in the spotlight, or if you thrive in both situations.

Answer Example

"As a data engineer, I realize that I do most of my work away from the spotlight. But that has never been that important to me. I believe what matters is my expertise in the field and how it helps the company reach its goals. However, I'm pretty comfortable being in the spotlight whenever I need to be. For example, if there's a problem in my department which needs to be addressed by the company executives, I won't hesitate to bring their attention to it. I think that's how I can further improve my team's work and reach better results for the company."

24. Do you have experience as a trainer in software, applications, processes or architecture?

If so, what do you consider as the most challenging part?

How to Answer

As a data engineer, you may often be required to train your co-workers on the new processes or systems you've created. Or you may have to train new teammates on the already existing architectures and pipelines. As technology is constantly evolving, you might even have to perform recurring trainings to keep everyone on track. That said, when you talk about a challenge you've faced, make sure you let the interviewer know how you handled it.

Answer Example

"Yes, I have experience training both small and large groups of co-workers. I think the most challenging part is to train new employees who already have significant experience in another company. Usually, they're used to approaching data from an entirely different perspective. And that's a problem because they struggle to accept the way we handle projects in our company. They're often very opinionated and it takes time for them to realize there's more than one solution to a certain problem. However, what usually helps is emphasizing how successful our processes and architecture have proven to be so far. That helps them open their minds to the alternative possibilities out there."

25. Have you ever proposed changes to improve data reliability and quality?

Were they eventually implemented? If not, why not?

How to Answer

One of the things hiring managers value most is constant improvements of the existing environment, especially if you initiate those improvements yourself, as opposed to being assigned to do it. So, if you're a self-starter, definitely point this out. This will showcase your ability to think creatively and the importance you place on the overall company's success. If you lack such experience, explain what changes you would propose as a data engineer. In case your ideas were not implemented for reasons such as lack of financial resources, you can mention that. However, try to focus on your continuous efforts to find novel ways to improve data quality.

Answer Example

"Data quality and reliability have always been a top priority in my work. While working on a specific project, I discovered some discrepancies and outliers in the data stored in the company's database. Once I've identified several of those, I proposed to develop and implement a data quality process in my department's routine. This included bi-weekly meetups with coworkers from different departments where we would identify and troubleshoot data issues. At first, everyone was worried that this would take too much time off their current projects. However, in time, it turned out it was worth it. The new process prevented the occurrence of larger (and more costly) issues in the future.

26. Have you ever played an active role in solving a business problem through the innovative use of existing data?

How to Answer

Hiring managers are looking for self-motivated people who are eager to contribute to the success of a project. Try to give an example where you came up with a project idea or you took charge of a project. It's best if you point out what novel solution you proposed, instead of focusing on a detailed description of the problem you had to deal with.

Answer Example

"In the last company I worked for, I took an active part in a project that aimed to identify the reason's for the high employee turnover rate. I started by closely observing data from other areas of the company, such as Marketing, Finance, and Operations. This helped me find some high correlations of data in these key areas with employee turnover rates. Then, I collaborated with the analysts in those departments to gain a better understanding of the correlations in question. Ultimately, our efforts resulted in strategic changes that had a positive influence on employee turnover rates."

27. Which non-technical skills do you find most valuable in your role as a data engineer?

How to Answer

Although technical skills are of major importance if you want to advance your data engineer career, there are many non-engineering skills that could aid your success. In your answer, try to avoid the most obvious examples, such as communication or interpersonal skills.

Answer Example

"I'd say the most useful skills I've developed over the years are multitasking and prioritizing. As a data engineer, I have to prioritize or balance between various tasks daily. I work with many departments in the company, so I receive tons of different requests from my coworkers. To cope with those efficiently, I need to put fulfilling the most urgent company needs first without neglecting all the other requests. And strengthening the skills I mentioned has really helped me out."

BRAINTEASERS

Interviewers use brainteasers to test both your logical and creative thinking. These questions also help them assess how quickly you can resolve a task that requires an out-of-the-box approach.

28. You have the following situation:

You have eight balls of the same size. Seven of them weigh the same, and one of them weighs slightly more. How can you find the ball that is heavier by using a balance and only two attempts at weighing?

You can put six of the balls on the balance. If one of the sides is heavier you will know that the heavier ball is on that side. If not, the heavier ball is among the two that you did not measure and it will be really easy to determine precisely which ball is heavier with your second weighing.

After you determine which side is heavier, you will be left with 3 balls to choose from. You have another attempt at weighing left. You can put two of the balls on the balance and see if one of them is heavier. If it is, then you have found the heavier ball. If it is not, then the third ball is the one that is heavier.

29. You have the following situation:

A windowless room has three light bulbs. You are outside the room with 3 switches, each of them controlling one of the light bulbs. If you were told that you can enter the room only once, how are you going to tell which switch controls which light bulb?

You have to be creative in order to solve this one. You switch on two of the light bulbs and then wait for 30 minutes. Then you switch off one of them and enter the room. You will know which switch controls the light bulb that is on. Here is the tough part. How are you going to be able to determine which switch corresponds to the other two light bulbs? You will have to touch them. Yes. That's right. Touch them and feel, which one of them is heated. That will be the other bulb that you had turned on for 30 minutes.

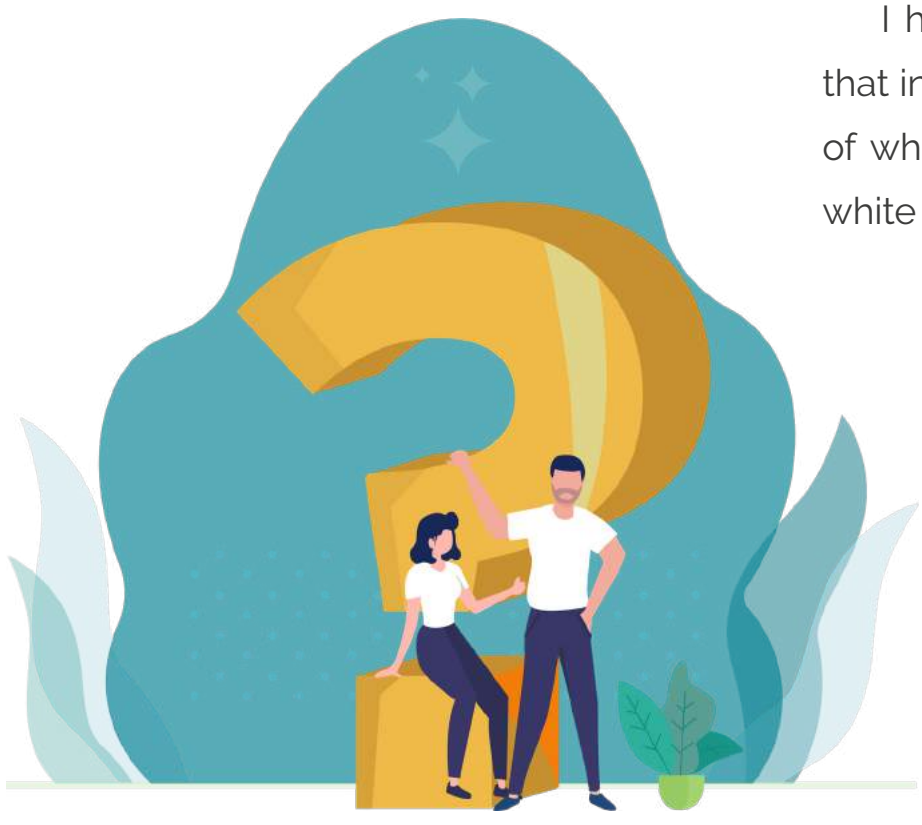
You will be in serious trouble if the interviewer says that the light bulbs are LED (given that they don't emit heat).

GUESSTIMATE

Although guesstimates aren't an obligatory part of the data engineer interview process, many interviewers would ask such a question to assess your quantitative reasoning and approach to solving complex problems. Here's a good example...

30. How many gallons of white house paint are sold in the US every year?

Find the number of homes in the US: Assuming that there are 300 million people in the US and the average household contains 2.5 people then we can conclude that there are 120 million homes in the US.



Number of houses: Many people live in apartments and other types of buildings different than houses. Let's assume that the percentage of people living in houses is 50%. Hence, there are 60 million houses.

Houses that are painted in white: Although white is the most popular color, many people choose different paint colors for their houses or do not need to paint them (using other types of techniques in order to cover the external surface of the house). Let's hypothesize that 30% of all houses are painted in white, which makes 18 million houses that are painted in white.

Repainting: People need to repaint their houses after a given amount of years. For the purposes of this exercise, let's hypothesize that people repaint their houses once every 9 years, which means that every year 2 million houses are repainted in white.

I have never painted a house, but let's assume that in order to repaint a house you need 30 gallons of white paint. This means the total US market for white house paint is 60 million gallons.

DATA ARCHITECT
INTERVIEW QUESTIONS

Data architect interview questions don't just revolve around role-specific topics, such as data warehouse solutions, ETL, and data modeling. In fact, interviewers will also challenge you with brainteasers, behavioral and situational questions. So, how do you prepare for a successful data architect interview?

GENERAL DATA ARCHITECT INTERVIEW QUESTIONS

The more general part of the interview is not necessarily focused solely on your resume. It could also include some questions regarding the projects you've worked on and how you manage your time and priorities.

1. Have you ever taken part in improving a company's existing data architecture?

Please describe your involvement in the process and the overall impact the changes had on the company.

How to Answer

Routine tasks and maintenance are an important part of a data architect's job. However, as a data architect, you should also be proactive and strive to improve the company's data processes and structures. Employers want to hire data architects with a critical mindset who are willing to take part in increasing the efficiency and productivity of current environments. So, do your best to show the interviewer you don't get preoccupied with routine tasks, and you don't lose sight of the bigger picture.

Answer Example

"In my work experience, marrying external data with internal data in corporate systems can pose a variety of threats to data integrity. That's why I launched a project where I established a step-by-step screening process for our 3-rd party purchased data. I also managed to further improve the relationship with our data supplier, who, in turn, agreed to run a few checks on their data before sending it to us. This initiative had a positive impact on the company's data reliability and decreased database errors by 29% within 1 year."

2. As a data architect, have you faced any challenges related to the company's data security?

How did you ensure the integrity of the data was not compromised?

How to Answer

Data security is a top priority for every company. That's why hiring managers would like to learn more about your experience with data security issues. When answering this question, emphasize that data security is an important aspect of your job, although your background isn't focused in that particular field.

Answer Example

"When working in a team, it's sometimes hard to agree on what could pose a security risk. I remember a situation when some colleagues of mine wanted to change the established process for uploading franchise data to our system. I was sure theses changes could result in security risks. So, in order to validate my point, I calculated the possible financial loss to the company in case security was compromised. This prompted the team members to modify their plan to strengthen data security measures."

3. As a data architect, you should be up to date with the latest technologies and developments in the field.

How do you keep yourself informed about the new trends in data architecture?

How to Answer

When working in a technical role, it's common to get absorbed in the company's current processes and miss out on the latest industry developments. Hiring managers will value your willingness to educate yourself despite your busy schedule. So, try to list news resources you're subscribed to, and mention some conferences or trainings, or industry events you attend when you have the chance.

Answer Example

"I do my best to stay informed about the latest industry trends and technology advancements. I believe this helps me learn things that can improve my work... Or inspire me to come up with an idea that will benefit the company's status quo. I'm subscribed to newsfeeds such as InformationWeek and TechNewsWorld. I also attend 2-3 conferences a year where I network with other professionals in the field. Whenever my schedule allows it, I attend specialized trainings and seminars.

4. What is referential integrity?

Referential integrity is a subset of data integrity that refers to the accuracy and consistency of data

linked between tables. Referential integrity is majorly important -if a database lacks referential integrity, this can result in return of incomplete data without any indication of an error.

For instance, we can say the foreign key in a certain child table maintains the referential integrity within the database by referencing a valid, existing primary key in the parent table.

A foreign key in SQL is defined through a foreign key constraint. This type of constraint verifies that the values in the child and parent tables match. Therefore, referential integrity doesn't allow us to add records to a related table unless there is an associated record in the primary table. It also prevents us from changing values in a primary table that would lead to orphaned records in a related table. Moreover, it makes it impossible to delete records from a primary table in case there are matching related records.

To visualize how the fields from the various tables within a database refer to each other, people usually use Entity-Relationship diagrams (ER diagrams), or, the simpler and handier tool – relational schemas.

5. Provide me an example of a time when you had to teach someone a new skill?

This is an easy one, right? The Hiring Manager wants you to demonstrate that you are a person that is willing to teach others. The fact that you are willing to teach means a few very important things:

- You are willing to share knowledge (very valu able for the company)
- You're a team player who is willing to help others
- You relate well to people

The second aspect that is important about this question is the method that you used when you were teaching. How did you share your knowledge? Did you have to use some special technique in order to explain a given concept? Did you have a strategy that helped to facilitate learning? Perhaps you provided valid practical examples?

Here's an example of such a situation:

You can say that you always wanted to teach your younger brother how to create good PowerPoint presentations. At first, it was difficult because it was very hard to get his attention. Then you proposed creating a presentation together – a presentation about his favorite motorbike company. He instantly agreed because it was something that he was interested in sharing with his friends and perhaps post in one of his favorite forums. At first, you were the one who was working with the mouse and the keyboard, but then you let him complete the second half of the presentation and you gave him your guidance throughout the process. The results were amazing your brother learned so much in such a short period of time. This was a very fulfilling experience for you and you realized that you enjoy teaching.

6. As a data architect, what steps have you taken to understand how different departments use the company's stored data?

How to Answer

Different departments have different data needs. And, as a data architect, you must have the ability to work with people from non-technical backgrounds to understand how they use the available data. When you answer this question, do your best to convey that you're willing to educate yourself to improve your job and better serve the company's data requirements.

Answer Example

"As a data architect, understanding the work of my colleagues in different departments has always been important to me. In my previous workplace, I'd regularly meet with reps from other teams to discuss their current and future projects. I would ask a series of questions, instead of making assumptions. This approach has allowed me to correctly identify and plan their data needs."

7. To effectively manage a company's data infrastructure, it is important for a data architect to have an in-depth understanding of the business and its strategic challenges.

How have you approached this requirement in your past position?

How to Answer

Missing the bigger picture is a common problem for data architects, due to the technical nature of their work. With your answer, you have to reassure the hiring manager that you're capable of taking proactive steps and stay on track with the overall business strategy and goals of the company.

Answer Example

"In my experience as a data architect, I've learned that in order to improve my performance, I have to be constantly aware of the company's short-term and long-term goals. This is why I've been proactive in my communication with management and c-level executives. I've also attended corporate trainings on a regular basis. This has given me a chance to ask the right questions to the right people."

8. What is/(are) your greatest strength/(s)?

A question that leaves a much more pleasant flavor than "What is your biggest weakness?" Nevertheless, you need to prepare to answer it, because it

is an important one and it comes up at almost all HR interviews.

Think of the role you are applying for. What are the greatest strengths that someone who wants to be successful in this position must have? Let's say that you are interviewing for the position of Project Manager.

A Project Manager needs to be a great:

- Communicator
- Motivator
- Team player
- Problem Solver

If the interviewer asks you for your greatest strength (singular) pick one of these qualities. The one that is, in fact, your greatest strength, so make sure that you have a great story illustrating that you are really good at this skill. If you are asked to list multiple strengths, you can pick up to three of these qualities. Don't list more than three strengths, as it will come off as though you are strong with everything, which will dilute the effect that you obtained in the first place.

Avoid vague words (such as maybe, probably, guess, usually) when you talk about your biggest strength/s.

TECHNICAL DATA ARCHITECT INTERVIEW QUESTIONS

The technical questions in a data architect interview concentrate on your work with certain programming languages, tools, and technologies, and your ability to use them to fulfill project goals or solve unforeseen issues.

9. A lot of companies use data from both internal and internal sources.

Have you faced any problems while trying to integrate a new external data source into the existing company's infrastructures? How did you solve these problems?

How to Answer

External data often comes from sources using different data formats and systems. Obviously, that may cause a series of issues when importing this data into the company's data systems. As a data architect, you have to make sure the data format is readable and ready-to-use, before storing it in the data warehouse. With this question, hiring managers want to assess your problem-solving skills when faced with external data integration challenges. So, try to provide an answer that will demonstrate how you address such issues.

Answer Example

"In my work experience, the cause for external data integration issues is usually a different system that creates the data in an incompatible format. Unfortunately, it isn't possible for all companies to use the same systems. So, I solved this problem by creating and running a script prior to uploading the data in my company's warehouse tables. The script not only changed the external data format but also ran tests to ensure the new format was compatible with our systems."

10. Have you worked with open source technology? Tell us about some issues you have comes across when using it.

How to Answer

When an interviewer asks a specific question like that, the company is either considering using open source technology in the future or is already utilizing it. If you have relevant experience, give some particular examples. And be sure you also highlight your ability to modify the open-source programming code. If you haven't encountered any problems using it, mention any possible disadvantages to open source technology you're aware of.

Answer Example

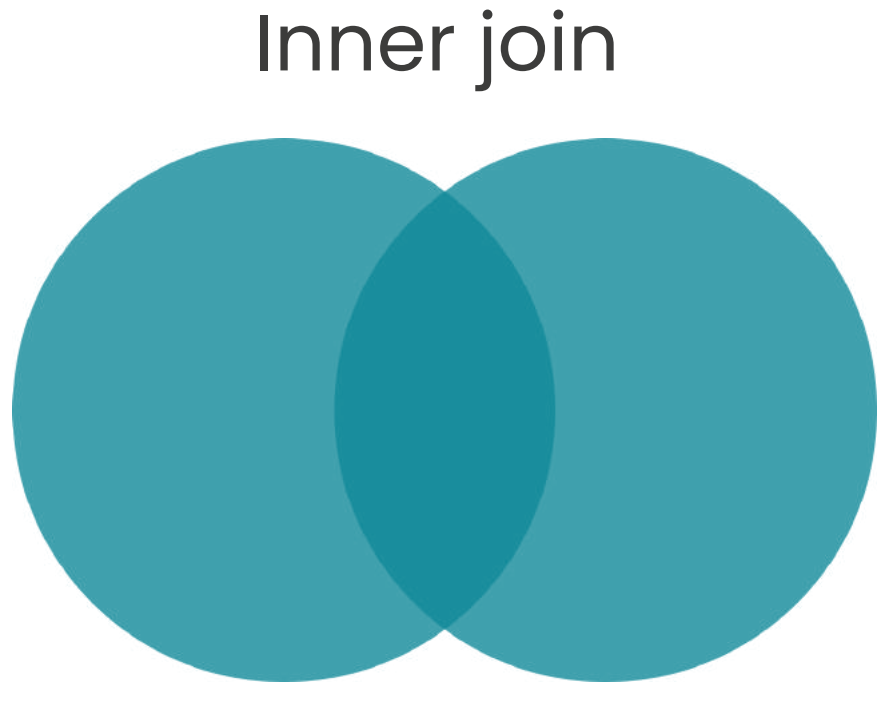
"I've worked with both Hadoop and MySQL without facing any major problems. Nevertheless, I realize that using open source databases or software utilities has its drawbacks. For example, you have to rely on advice from user forums, as there is no formal customer support to address your issue. Another thing is that developers don't spend a lot of time on their user interface, so you may lack the resources you need to get started."

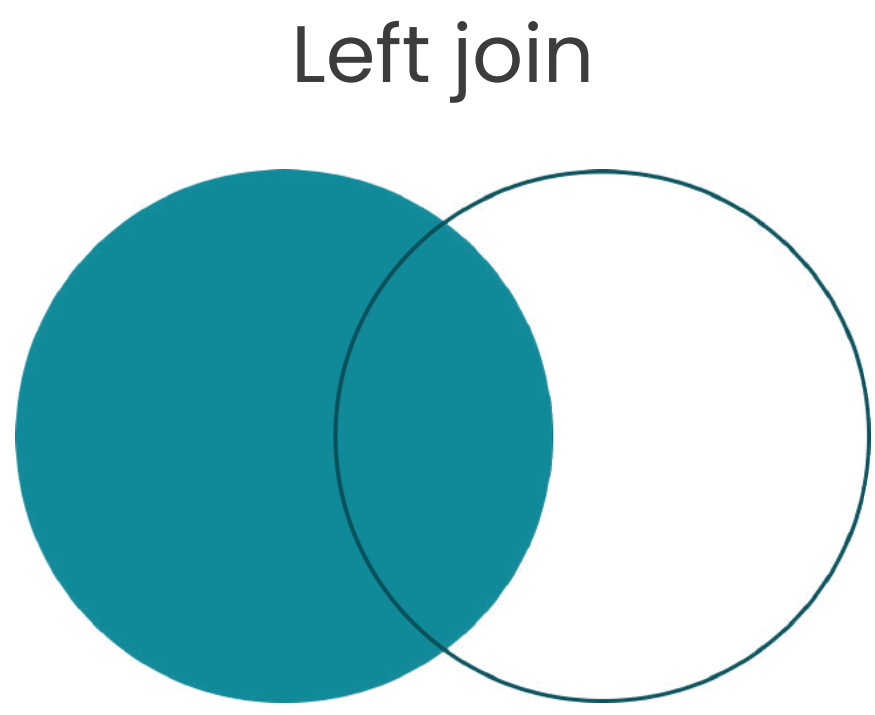
11. State and describe the different types of SQL joins.

How to Answer

The basic types of SQL joins are: inner, left, and right (in SQL theory, there is one more type of join – full. However, it is used very rarely today). The easiest and most intuitive way to explain the difference between the inner, left, and right joins is by using a Venn diagram, which shows all possible logical relations between data sets.

The INNER JOIN lets us select all records from Table A and Table B, as long as there is a match between the columns.





The SQL LEFT JOIN returns all records from the left table, plus the matched values from the right table. In case there are no matches, the left join still returns all rows from the left table and a NULL value from the right.

Regarding the functionality of the SQL RIGHT JOINS – it is identical to LEFT JOINS, but with the opposite direction of the operation.

12. What is a primary key and a foreign key?

How to Answer

A primary key is a column (or a set of columns) whose value exists and is unique for every record in a table. It's important to know that each table can have one and only one primary key.

Therefore, you can think of a primary key as the field (or group of fields) that identifies the content of a table in a unique way. For this reason, the primary keys are also called the unique identifiers of a table.

Another crucial feature of primary keys is they cannot contain null values. This means, in an example with a single-column primary key, there must always be a value inserted in the rows under this column. You cannot leave it blank.

One last remark about primary keys - not all tables you work with will have a primary key, although almost all tables in any database will have a single-column or a multi-column primary key.

A foreign key, instead, is a column (or a set of columns) that references a column (most often the primary key) of another table. Foreign keys can be called identifiers, too, but they identify the relationships between tables, not the tables themselves.

In the relational schemas form of representation, relations between tables are expressed in the following way – the column name that designates the logical match is a foreign key in one table, and it is connected with a corresponding column from another table. Often, the relationship goes from a foreign key to a primary key, but in more advanced circumstances, this will not be the case. To catch the relations on which a database is built, we should always look for the foreign keys, as they show us where the relations are.

13. How many types of data structures does R have?

How to Answer

This question is important because virtually everything you do in R involves data in some shape or form. The most commonly used data structures in R are these:

- Vectors (atomic and lists);
- Matrixes;
- Data frames;
- Factors.

14. What modeling tools have you used in your work so far? Which do you consider efficient or powerful?

How to Answer

Even if data modeling isn't one of your main responsibilities, your role as a data architect requires you to have an in-depth understanding of data modeling. If you lack the experience, demonstrate that you are well-informed on the topic and mention the data modeling tools you find most useful. The interviewer will value that you're at least familiar with the subject.

Answer Example

"I've used mainly both Oracle SQL Developer Data Modeler, and PowerDesigner. I can say that the Oracle Data Modeler has been more than sufficient for my needs with its dimensional modeling, and integrated source code control that supports collaborative development. However, PowerDesigner also boasts some wonderful technology-centric metadata management capabilities for data architects, and business-centric techniques for non-technical coworkers. Overall, I think both tools are worth the try, depending on the company's needs."

15. What's your experience with batch and real-time data processing?

How to Answer

Each of these data processing methods can be applied depending on the business case. If you have experience with only one of them, provide examples of situations where the other processing method would be a better fir. This will indicate you have a basic understanding of both batch and real-time data processing.

Answer Example

'I'm familiar with both types of data processing. However, I've had more exposure to batch processing. That's because one of my responsibilities was to write programs that captured, processed, and produced output for the company's billing department. As I mentioned, I've had less experience with real-time data processing. However, I know our company uses it to take immediate action on the data collected from our stores' POS systems."

16. In your role as a data architect, what metrics have you created or used to measure the quality of new and existing data?

How to Answer

Having established processes to ensure the quality of data is key to a company's data infrastructure. With this question, the hiring manager wants to as-

sess your relevant experience. Make sure you highlight the particular dimensions you've monitored to validate the data quality.

Answer Example

"I've always been involved in ensuring data quality in my job as a data architect. My team and I monitored some specific dimensions to validate the quality of data. These included completeness, uniqueness, timeliness, validity, accuracy, and consistency. Monitoring these dimensions helped us detect inconsistencies that could negatively affect the accuracy of data analysis."

17. Tell us about a situation when you made changes to a company's data management systems and the impact it made on the company.

The data needs of companies change and hiring managers want to make sure they hire an architect that will not only adapt to the new requirements but will also take up the initiative to implement these changes and introduce some new improvements. If you are just beginning your career as a Data Architect and you don't have experience in dealing with such changes, think of a hypothetical situation that will demonstrate your problem-solving skills and hands-on approach to challenges.

Here's an example of such a situation.

While working for my previous employer, I was part of a project aiming to make data more accessible to all company employees. Each department's data was siloed and team members in other departments couldn't access it. Acquiring data outside one's own department was a dull and tiresome process that prevented timely analyses. I actively took part in making data sharing among the company's departments easy without compromising data security. Thus, analysts were able to complete their projects in time using a much more robust dataset than before. This made it possible for senior management to make fast and better-informed strategic decisions.

18. What issues have you faced while leading teams tasked with data/database strategy development?

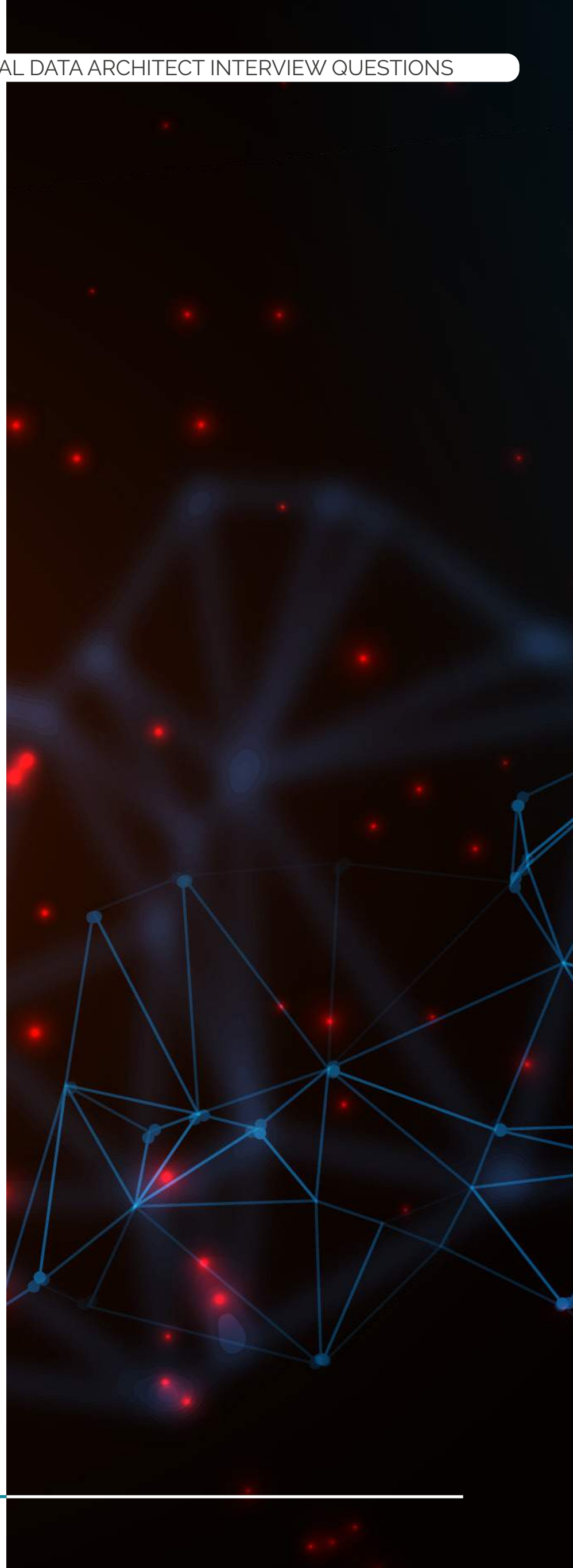
Tell us how you solved these issues.

How to Answer

You can approach this question in a more general way, or describe a real situation you and your team have faced when working on a specific task. Either way, make sure you point out your problem-solving skills and the ability to work in a team to reach a common goal.

Answer Example

"In my experience as a data architect, I've often worked with teams to develop changes in the data architecture of our company. Of course, people on a team come from different backgrounds and have varying opinions that affect their priorities. What I've discovered is that making compromises is crucial to the success of the task, along with staying open-minded to others' ideas. That said, once we've identified our common goals, a consensus has always been easy to reach."



BEHAVIORAL DATA ARCHITECT INTERVIEW QUESTIONS

Data architects often work with coworkers from various departments, backgrounds, and responsibilities. This is why you should be prepared to answer some behavioral questions focused on your work style and ability to handle conflict in cross-functional teams.

19. What challenges have you faced working with colleagues with no technical background?

How did you address and overcome these challenges?

How to Answer

Data architects often work with other departments within a company. That involves collaborating with people who lack technical background and understanding of the data processes. The interviewer would like to assess your communication style and your ability to reach common ground with your coworkers, in spite of your differences. Describe a specific situation to illustrate the issues you encountered and how you solved them.

Answer Example

"I believe a good data architect should understand the needs of the different departments across the company. That said, I've had to work with people who don't fully understand my role and responsibilities on numerous occasions. Some of my coworkers would pose requests that I had to reject due to our data architecture limitations. And that has lead to certain tensions. I'd say overcoming such challenges takes time. Gradually, we learned more about each other's work which helped us brainstorm possible solutions. All in all, making the extra step to educate myself and the others made has made all the difference."

20. How would you describe your work style?

How to Answer

This question is not so much about your personality, but more about how you approach your work to get things done. Talk about the way you handle tasks and projects, and how you communicate with coworkers and clients. Your work style might be: collaborative, well-structured, speedy, flexible, or independent. No matter what word you choose to describe it, keep the job description in mind and how your work style fits the profile.

Answer Example

"I'd describe my work style as collaborative. I like to work on full-team participation projects and co-create with my teammates. If I'm not sure of the direction I should take on a project, I always consult with my team. This way we can work toward consensus and align our ideas."

21. How would you resolve a conflict within your team?

How to Answer

The hiring manager wants to hear about your ability to professionally solve team issues when they occur. Think of an example where you had to use your communication skills to handle a conflict with your coworkers. Or when you managed to help 2 of your teammates find common ground as a mediator.

Answer Example

"I like to think I have excellent conflict management skills. As a data architect in a large company, I've worked in a high-stress environment. And that has sometimes caused tension to build up among team members. When this escalates to a conflict, I try to deal with it openly. Usually, I'd organize a group meeting where everyone can voice their concerns. This is how we can sort out the issue and move on with our work on the project."

22. What is the most critical factor for you when taking a job?

How to Answer

There are a lot of factors that may influence your decision to take on a new job. These include:

- career growth opportunity;
- compensation;
- work/life balance;
- travel required for the role;
- medical and health benefits;
- perks such as a gym membership, onsite kids center, spending account;
- paid vacation time;
- the company's location;
- the company's reputation and culture.

Share with the interviewer which factors are most important to you when you consider starting a new job. If you aren't sure about all the details regarding this position, this is a good time to get informed.

Answer Example

"The most important factors for me, as a data architect, are the company's industry and the workplace culture. The first one predefines the projects I'll be involved in. The second one indicates if the work environment will be positive and teamwork-oriented. To me, those are equally important to compensation and benefits."

23. Are you also interviewing with any of our close competitors?

How to Answer

If the interviewer wants to know if you're also applying for a job at a competitor's company, you can give a direct answer. However, you should refrain from giving away the name of the company or sharing too many details. Let the interviewer know you aren't putting all of your eggs in one basket. At the same time, try to leave the impression that you are critical when it comes to the companies you apply at.

Answer Example

"I wouldn't disclose the names of the competitors I'm currently interviewing with. However, I can tell you that I'm in the mid-interview stages with 3 other companies. That said, your company is my first choice and I'm happy that we've reached the final stage in the process."

24. How would you assess your performance in the interview so far?

How to Answer

This is a question you should answer openly. Generally, you would know if you performed well, or if your interview were a disaster. In fact, if you address the issues of your performance, you might get a chance to answer some additional questions that could give you extra points."

Answer Example

If you think that your performance in the interview is going great:

"I'm positive that the interview has been quite successful and I'm satisfied with my performance. Is there anything you'd like me to clarify from our talk?"

If you think that your performance in the interview is not satisfactory:

"I don't think I managed to portray myself in the best light possible in this interview. However, I'm always trying to do my best. So, if there's anything I could further clarify for you, I'd be more than happy to do so."

25. What would you do if a colleague was using a company phone for personal use?

These types of questions about the unethical behavior of one of your colleagues are difficult to answer. First of all, it is a very awkward situation. Most

people don't want to rat their co-workers but are not OK with unethical behavior neither. That means that they need to make a tough decision between two conflicting actions.

Depending on how serious and unethical the actions of your colleague are, you usually have two options:

- Talk with them before reporting to the manager and try to convince them to change their behavior
- Report them directly to the manager

In this case, given that a personal phone call from a company phone is not something that endangers the company and its reputation in the long run, you might try to fix the issue yourself by talking to your colleague and explaining to him that using the company phone for private conversations is not allowed. Strengthen your argument by saying that if everybody started doing such things, the company would eventually go bankrupt. Furthermore, he is setting a wrong example for the rest of your colleagues. Given that the company trusted you with this job, you need to repay that with solid work and consistently ethical behavior. If the pattern continues even after you talked to your colleague, you should contact Management.

Had the question involved a more serious violation (sexual harassment, stealing, disclosure of confidential information, etc.) you need to demonstrate your readiness to report the issue directly to your supervisor.

26. How would you deal with a significant mistake at work?

The best way to deal with a mistake at work is to own up to it. Otherwise, it will haunt you and will probably transform into something that cannot be fixed. A timely reaction could prevent the damage deriving from your mistake and shows strength of character.

One of the worst things that can happen to you is to have a manager who has lost trust in your work.

Hiding mistakes can cause that. It will be much better to confront your manager immediately and admit that you made a mistake. Then, once he knows about the situation, he will be able to take appropriate action in order to resolve the situation. It is more likely that he will know how to address the issue because he is more experienced than you.

The more subtle aspect of this question is about how you learn from significant mistakes. Are you going to remember that mistake and learn from it in the future? Are you going to do everything possible in order to avoid it in the future? What type of precautionary measures would you take? Everybody makes mistakes, yes. The important thing is that you show that you are determined to learn from yours.

27. What would you do if one of your colleagues was not performing well?

Open communication is the best way to address problems when you are working with people. Remember that. By openly sharing your concerns with your colleague and hearing his opinion, you will make sure that both of you are on the same page about the current situation. You need to fully understand what caused his weak performance.

It could be due to:

- Misunderstanding of his tasks
- Lack of experience in handling this type of tasks
- Personal problems
- Anxiousness to do too much

Then, once you have figured out what the problem is, the next step is to figure out a way to resolve the issue.

For example, you can propose the following solutions:

- Misunderstanding of his tasks – Go through his tasks together and tackle any problematic areas
- Lack of experience in handling this type of task – Depending on the knowledge gap and the deadline that you have you can i) propose to go through the unfamiliar topics together ii) propose to change his assignment with something that he is familiar with and where he can excel
- Personal problems – Offer flexible hours or suggest that he asks the manager for help and explain his personal situation; say that you are behind him and that everyone has difficulties at some point
- Anxiousness to do too much – Explain that the best employees are great at doing well the small things; assure him that he needs to focus on doing well his ordinary tasks without being distracted by issues that are outside of his current capabilities.

BRAINTEASERS

Brainteasers help the interviewer assess your logical thinking combined with your ability to come up with a creative problem solution on the spot.

28. What is the sum of the numbers from 1 to 100?

There's a little bit of history coming with this question. The math teacher of young Karl Gauss, the famous mathematician, asked the entire class to sum the numbers from 1 to 100. He expected that the task would require at least half an hour to his students, but was shocked when Gauss gave him the exact number within mere seconds. Anyway, here is how this question is solved.

There are precisely 50 pairs of numbers from 1 to 100, whose sum is 101.

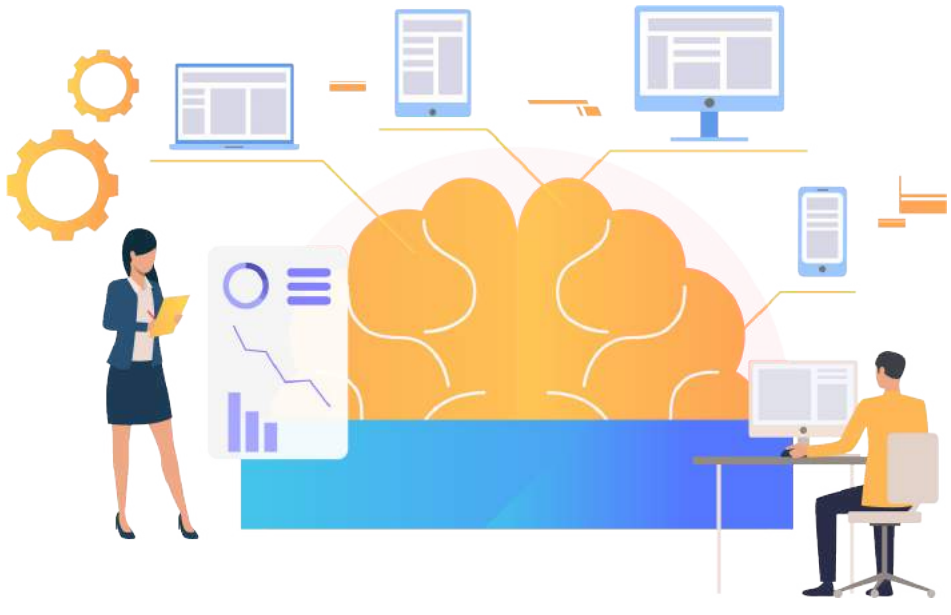
1 + 100 = 101, 2 + 99 = 101, 3 + 98 =101, etc.

50 * 101 = 5050

This trick will work for any series of numbers provided that they are evenly spaced. You need to find the sum of the first and the last number and then multiply by the number of pairs.

29. You are given two containers - one is 5 and the other one is 7 gallons. How do you use them to measure 4 gallons of water?

Fill the entire 7 gallon container with water. Then use the water in the 7 gallon container in order to fill the entire 5 gallon container. This would leave 2 gallons of water in the 7 gallon container. Dump the water in the 5 gallon container and then pour in it the 2 gallons of water that are in the 7 gallon container. Fill the entire 7 gallon container with water and then start pouring the water in the 5 gallon container. Given that it is already filled with 2 gallons of water, you will be able to pour only 3 gallons, which means that 4 gallons would remain within the 7 gallon container. This is how you are able to measure 4 gallons of water.



GUESSTIMATE

Guestimates are not necessarily a part of each data architect interview. However, if the interviewer decides to throw you a curveball, you should be prepared. Here's an example...

30. How many flat-screen TVs have been sold in Australia in the past 12 months?

The population of Australia is approximately 24 million. Let's assume that the average household comprises of2 people (there are a lot of families with 3 or 4 people, but this is balanced by those people who are living alone). So the number of homes is 12 million (provided that all people have a home).

Then we need to find out how many of the TVs in these 12 million homes will need to be replaced with new ones. Let's assume that people need to replace their old TV with a new TV every six years and that every home has 1.5 TVs. Nowadays, it is reasonable to expect that all new TVs that are purchased have a flat-screen. Therefore, the number of flat-screen TVs that are purchased in Australia in one year is equal to:

1/6 of the homes buy a new TV this year * 12 million homes * 1,5 TVs per home = 3 million flat-screen TVs.

This brings our list of 180 data science interview questions to an end.

We believe this concise guide will help you “expect the unexpected” and enter your first data science interview with confidence.

Of course, when it comes to preparing for a data science career, and data science interview questions in particular, more is more. So, make sure you check out our career resources, as they will help you on the path towards your professional data science goals.

ABOUT THE AUTHORS

Iliya
Co-founder, 365 Data Science

Iliya is the co-founder of 365 Data Science. He's a Finance graduate with a strong quantitative background. Iliya is the author of our online courses in Mathematics, Statistics, Machine Learning, and Deep Learning. He has won more than 90 national and international awards and competitions through the years. He first started teaching while at university; now, Iliya has authored and instructed 5 online courses and he's happy to share his expertise in data science with all aspiring professionals in the field.

Ned
Co-founder, 365 Data Science

Ned is the co-founder of 365 Data Science. He has a Bachelor's degree in Business Administration and Management, and a Master's in Finance at Bocconi University in Milan, Italy. Ned has rich experience in financial advisory, and has worked for renowned international companies, such as Pwc (Italy), Coca-Cola (United Kingdom), and Infineon Technologies (Germany). He is also the author of numerous career resources helping aspiring professionals reach their highest goals.

Ellie
Python and SQL Instructor

Ellie is a Computational Biologist, with expertise in the fields of algorithms and data structures, phylogenetics, and population genetics. She has a solid academic background in Bioinformatics with publications on constructing Phylogenetic Networks and Trees. Elitsa is one of the authors of the course Customer Analytics in Python in the 365 data science Program and is currently creating the upcoming 365 Data Visualization Course.

Martin
Python, R and ML Instructor

Martin is the author of the Python, SQL, and Integration courses in the 365 Data Science Program. He has an MSc in Economic and Social Sciences from Bocconi University in Milan, Italy. Martin's experience includes assisting an empirical research for Innocenzo Gasparini Institute of Economic Research, and working for DG Justice and Consumers at the European Commission.

365 DataScience

Ready to take the next step?

The 365 Data Science Program offers a complete data science training taught by expert instructors with a fun, interactive, and beginner-friendly approach. The courses start with the fundamentals, cover in-demand programming languages like Python, R, and SQL, visualization tools like Power BI and Tableau, and finish off with advanced specialized courses, including state-of-the-art Machine and Deep Learning.

You can try the course (12 hours of video instruction) **for free.**

Sign Up for Free Preview