



Interview Q&A for Data Analyst Roles

A comprehensive list of technical & behavioral Qs asked in interviews for data analyst positions.

Ms. Anupama Rajaram
sunnyblogger@gmail.com
<http://www.journeyofanalytics.com/>

Contents

Data analyst interview Qs..... 5

- What is involved in typical data analysis?..... 5
- What are your communication strengths? 5
- What has been your most difficult analysis to date? 5
- What kind of data analysis software experience do you possess?..... 6
- You have just been assigned a new analytics project. Where do you begin and what are the steps that follow? 6
- How do you define big data? 6
- You develop a big data model, but your end user has difficulty understanding how the model works and the insights it can reveal. How do you communicate with the user to get your points across? 7
- Can you describe a big data project you worked on?..... 7
- What challenges have you encountered while working with big data?..... 7
- What are your technical competencies? 8
- According to you, which are the qualities that a data analyst must possess to be successful at this position?..... 8
- Describe the most difficult database problem you have faced? Why was it so difficult than other analytical problems you solved?..... 8
- What procedure do you follow to analyze the given data?..... 9
- What do you understand from the term data cleaning? 9
- Besides analyzing, have you ever participated in database designing and database development tasks? 9
- I have provided you here some data regarding the sales figures of last 5 quarters. It shows the expenses incurred as a part of the sales activities? How would you analyze this data?..... 9
- Which technical tools do you use often for analysis and presentation purpose?..... 9
- What is involved in typical data analysis?..... 10
- Do you have any experience with corporate websites?..... 10
- How do you handle database as a data analyst?..... 10
- Have you ever developed any programs or products relevant to your profession? 10
- How often should you retrain or refresh a model? 10
- How would you handle the QA process when you’re creating a predictive model to forecast customer churn? 11
- How would you create a taxonomy to identify key customer trends in unstructured data? 11
- Describe the abilities you have in order to work with us as intelligence analyst? 12

What are the knowledge elements you obtained from your education, training and work experience would support your intelligence analyst career?.....	12
How would you describe (analyst or your) work style?.....	12
Tell about yourself and why you think you are successful intelligence analyst?	12
Why do you like to work as intelligence analyst?	12
How would you create a taxonomy to identify key customer trends in unstructured data?	12
Which technique is used to predict categorical responses?.....	13
What is logistic regression? Or Explain a situation when you have used logistic regression recently.	13
What are Recommender Systems?.....	13
Why data cleaning plays a vital role in analysis?	13
Differentiate between univariate, bivariate and multivariate analysis.	14
What do you understand by the term Normal Distribution?	14
What is Linear Regression?	14
What is Interpolation and Extrapolation?.....	15
What is power analysis?	15
What is K-means? How can you select K for K-means?	15
What is Collaborative filtering?	15
What is the difference between Cluster and Systematic Sampling?	15
Are expected value and mean value different?.....	15
What does P-value signify about the statistical data?.....	15
Do gradient descent methods always converge to same point?.....	16
What are categorical variables?.....	16
A test has a true positive rate of 100% and false positive rate of 5%. There is a population with a 1/1000 rate of having the condition the test identifies. Considering a positive test, what is the probability of having that condition?	16
How you can make data normal using Box-Cox transformation?	16
What is the difference between Supervised Learning an Unsupervised Learning?	16
Why is vectorization considered a powerful method for optimizing numerical code?.....	16
What is the goal of A/B Testing?.....	16
What is an Eigenvalue and Eigenvector?	16
What is Gradient Descent?	17
How can outlier values be treated?	17
How can you assess a good logistic model?	17

How can you iterate over a list and also retrieve element indices at the same time? Ans.....	17
During analysis, how do you treat missing values?	17
Explain about the box cox transformation in regression models.	18
Can you use machine learning for time series analysis? - Yes, it can be used but it depends on the applications.....	18
Write a function that takes in two sorted lists and outputs a sorted list that is their union.	18
What is the difference between Bayesian Inference and Maximum Likelihood Estimation (MLE)?	18
What is Regularization and what kind of problems does regularization solve?.....	18
What is multi-collinearity and how you can overcome it?	18
What is the curse of dimensionality?.....	18
How do you decide whether your linear regression model fits the data?	18
What is the difference between squared error and absolute error?	18
Which is your favorite machine learning algorithm and why?	18
In which libraries for Data Science in Python and R, does your strength lie?	18
What kind of data is important for specific business requirements and how, as a data scientist will you go about collecting that data?.....	18
Tell us about the biggest data set you have processed till date and for what kind of analysis.....	18
Suppose you are given a data set, what will you do with it to find out if it suits the business needs of your project or not.....	18
What were the business outcomes or decisions for the projects you worked on?	18
What unique skills you think can you add on to our data science team?	18
Why do you want to pursue a career in data science?.....	19
What have you done to upgrade your skills in analytics?.....	19
What has been the most useful business insight or development you have found?.....	19
How will you explain an A/B test to an engineer who does not know statistics?	19
When does parallelism helps your algorithms run faster and when does it make them run slower?.....	19
How can you ensure that you don't analyze something that ends up producing meaningless results?.....	19
How would you explain to the senior management in your organization as to why a particular data set is important?	19
Is more data always better?.....	19
How do you go about collecting information to support your analyses?	19
What types of data have you researched and analyzed in the past?	19

What are your computer skills and experience using analytic software.....	19
Explain how you organize and create presentations to report analytical findings?.....	20
What is meant by Data Analytics?	20
What do you do in the current job? What is your education background? Describe your experience.....	20
Explain about big-data and its applications?	20
How can you handle databases as a data analyst?.....	20
Assume, you develop the big data model, but the customer feels difficult to use your model? How do you communicate with the customer at this point?	20
R can't handle large amounts of data so what are the others ways to handle data?	20
What is Multicollinearity? How can we solve it?	20
Explain about the clustering methods you are familiar with?.....	21
Define P-value?	21
What are your strongest technical skills?	21
Difference between statistical and machine learning?	21
Difference between recurrent neural networks and recursive neural networks?	21
What is your key strength in the business analyst?.....	21
Say about your project that you are working as a data analyst?.....	21
What is Collaborative filtering / Recommendation Engine?	21
Tell about most challenging project you have done so far.....	21
How will you design a Recommender (recommendation) System with more optimizations and variants?.....	21
Describe a recent use of logistic regression.....	21
Describe a recent Data science case study.	21

Data analyst interview Qs

What is involved in typical data analysis?

The interviewer is making certain that you have a basic understanding of the work you'll be doing. Your answer is extremely important, especially if this will be your first time in a data analyst position. "Typical data analysis involves the collection and organization of data. Then, finding correlations between that analyzed data and the rest of the company's and industry's data. It also entails the ability to spot problems and initiate preventative measures or problem solve creatively."

What are your communication strengths?

Communication is key in any position. Specifically, with a data analyst role, you will be expected to successfully present your findings and collaborate with the team.

Sample answer –

My greatest communication strength would have to be my ability to relay information. I'm good at speaking in a simple, yet effective manner so that even people who aren't familiar with the terms can grasp the overall concepts. I think communication is extremely valuable in a role like this, specifically when presenting my findings. This is even more important when those findings could be beneficial or detrimental to other departments within the company and you need to make sure that everyone understands the overall message."

What has been your most difficult analysis to date?

The interviewer wants to see if you are an effective problem solver. Be sure to include how you overcame the challenge.

Sample answer –

"My biggest challenge was making prediction sales during the recession period and estimating financial losses for the upcoming quarter. Interpreting the information was a seamless process. However, it was slightly difficult to forecast future trends when the market fluctuates frequently. Usually I analyze and report on data that has already occurred. In this case, I had to research how receding economic conditions impacted varying income groups and then make an inference on the purchasing capacity of each group."

What kind of data analysis software experience do you possess?

Data analysts are required to use software in their role. Additionally, analysts are sometimes responsible for database design and security. Share your relevant experience with the interviewer.

Sample answer –

"I have advanced data analysis software experience. A few examples include creating PivotTables in Excel, producing databases from scratch in Access, and developing data mining algorithms in ELKI. Also in my previous role, I was tasked with upgrading the database to meet the demands of the market and the company to ensure it ran smoothly."

You have just been assigned a new analytics project. Where do you begin and what are the steps that follow?

The interviewer wants to get a look inside your thought process to make sure you are thorough and organized.

Sample answer –

"The very first thing I would do is clearly define the problem or objective so I have a solid direction. Second, I would explore the data and become more familiar with it. This is extremely critical especially if I am working with a new set of data. Next, I would prepare the data for modeling. This entails data validation, detecting outliers, treating missing values, etc. With those steps completed, I would begin modeling the data until I discover the most significant or valuable results. Lastly, I would implement the model and track my results. As I'm sure you are aware, this process could vary slightly based upon the type of problem and the data and tools available."

How do you define big data?

It's likely that you'll be interviewed by an HR rep, an end business user, and an IT pro. Each person will probably ask you to explain [what big data is](#), and how the data analysis discipline works with big data to produce insights.

Sample answer –

You can start your answer with something fundamental, such as "big data analysis involves the collection and organization of data, and the ability to discover correlations between the data that provide revelations or insights that are actionable." You must be able to explain this in terms that resonate with each interviewer; the best way to do this is to illustrate the definition with an example.

The end business user wants to hear about a hypothetical case where a specific set of data relationships uncovers a business problem and offers a solution to the problem. An HR rep might be receptive to a more general answer, though the answer is more impressive if you can cite an HR issue, such as how to look for skills areas in the company where personnel need more training. The IT pro also wants to hear about an end business hypothetical where big data analysis yields results, but he also wants to know about the technical process of arriving at the data postulates and conclusions.

You develop a big data model, but your end user has difficulty understanding how the model works and the insights it can reveal. How do you communicate with the user to get your points across?

Many big data analysts come from statistics, engineering, and computer science disciplines; they're brilliant analysts, but [their people and communications skills lag](#). Businesses understand that to obtain results, you need both strong execution and strong communication. You can expect your HR, end business, and IT interviewers to focus on your communications skills, and to try to test them with a hypothetical situation.

Can you describe a big data project you worked on?

Companies understand that they have to train and orient you to their business and technical environments, but they also expect you to bring skills, experience, and fresh ideas to the job. The end business user and the IT interviewer will be especially interested in your project work. For the IT person, be sure to go into the data quality, analysis, publication, and actionalization processes, covering both the end business and the technical enablement details. For the end business person, review the project from a business results perspective, but avoid using technical jargon unless asked.

What challenges have you encountered while working with big data?

Big data doesn't always work as advertised, which is why your IT interviewer will likely probe you about big data setbacks or limits that you've encountered, and ask how you worked through them. Be prepared to answer this question in a straightforward, factual manner, and cap your answer with a discussion of what you gained from the experience and how it benefits you now.

What are your technical competencies?

Before the interview, do your homework on the analytics environment that the interviewing company uses. During the IT interview, you will be asked to review your technical competencies and skillsets. How well the company feels your technical skills fit with the data analytics approaches and tools they use in their environment can have a make-or-break effect on whether you get the job.

According to you, which are the qualities that a data analyst must possess to be successful at this position?

Analytical and problem solving skills are crucial to be successful at this position. Also, one needs to be skilled at organizing and formatting data, so that information is available in an easy-to-read manner. Technical proficiency is important too, especially if your organization depends largely on the software and data processing tools. (Talk about other skills that the recruiter expects in his candidates as per the given job description)

Describe the most difficult database problem you have faced? Why was it so difficult than other analytical problems you solved?

This is a personal interpretative question that must be answered as per your experience. Introduce the problem and highlight the exact nature of the problems faced. Give details of how you perceived the problem, analyzed the information, applied principles of analysis, managed information and found solutions to the objectives.

Sample answer –

My toughest challenge was to make prediction sales during the recession period and estimate the losses the organization may suffer in the coming quarter. It was difficult since I had to interpret information and forecast future trends. Till then I had analyzed the information that I already had and concluded on what had already happened. I had to evaluate the impact of receding economic conditions on different income groups and make an inference on the purchasing capacity of that group. I used different statistical methods and economic parameters to make this conclusion.

What procedure do you follow to analyze the given data?

The procedure I adopt to solve analytical problems depends upon the objective of the analysis. At first, I organize information by making categories. Secondly, I find relationship between two or more categories such that any changes in one category affects the results of the other. Further, the procedure includes data cleaning, defining structure of samples, determining quality verification measures, computing statistics and making reports on the analysis.

What do you understand from the term data cleaning?

Data cleaning refers to the task of inspecting and cleaning data. From the given data, it is important to sort out information that is valuable for analysis. Also, one needs to eliminate information that is incorrect, unnecessary or repetitive. However, the entire database should be retrievable. Data cleaning does not impose deleting information completely from the database.

Besides analyzing, have you ever participated in database designing and database development tasks?

Customize this answer as per your experience

Sample answer –

I have not directly participated in data designing and data development activities since we had a dedicated team to perform those functions. However, I have given valuable inputs to those teams during the discussions and helped them in research and data formatting.

I have provided you here some data regarding the sales figures of last 5 quarters. It shows the expenses incurred as a part of the sales activities? How would you analyze this data?

The answer to this question depends upon the nature of the problem that is given to you to solve. However, here are some steps that you should follow while solving the problem. Read the problem and the data carefully. Ask for an explanation for the terms that you could not understand. Extract the unusual and interesting trends that need attention. Apply the various statistical analysis methods and principles. Make a presentation or give a verbal account of your analysis as demanded.

Which technical tools do you use often for analysis and presentation purpose?

The tools to be used depend upon the nature of analysis being done.

Sample answer –

I am comfortable using the QDA (Qualitative Data Analysis) Miner, KNIME (Konstanz Information Miner), Root, PAW, etc. Also, most of my work is done using MS Excel features of shewhart control charts, stratification, histogram, scatter diagram, correlation and covariance, etc. I am keen on learning new tools as may be necessary to work at this organization.

What is involved in typical data analysis?

Data analysis involves collection and organization of data, correlation between analyzed data and the rest of the company and market, and the ability to then creatively think of solutions to existing problems, or point out problems and initiate preventive measures.

Do you have any experience with corporate websites?

Relevant experience might involve creating the rules for various corporate consumer websites, depending on market conditions and company goals. Analysts might also help design website features, so that they align with their analysis of company data and performance predictions.

How do you handle database as a data analyst?

Data analysts may be responsible for database design and security. They are responsible for upgrading the database on a regular basis so that it meets the demands of the market and company needs. And they make sure that the database runs smoothly in general.

Have you ever developed any programs or products relevant to your profession?

As a business analyst working in the software and technology industry, you might have experience actually developing information-related products for the company, which can be used for data analysis and data presentation.

How often should you retrain or refresh a model?

A top-notch data analyst understands how changing business dynamics affect the efficacy of a predictive model. She's not just a data cruncher: she's a valuable consultant who uses her analytical skills and business acumen to solve root-cause business problems.

Sample answer - I'll work with the business owner to establish an appropriate time period upfront. However, I would retrain a model immediately should the company expand into a new market, consummate an acquisition, or encounter emerging competition. Models must be

retrained quickly to adjust for changing customer behavior patterns or shifting market conditions.”

How would you handle the QA process when you're creating a predictive model to forecast customer churn?

To operationalize analytics, data analysts need a collaborative environment and input from the business owner. They also need a repeatable, effective, efficient process to create predictive models and reliable architecture to deploy predictive analytic models into production. Without feedback from the business owner, the model will probably sit on the shelf. And no self-respecting analyst wants to create a one-and-done model.

Sample answer –

“I'd partition the data into three sets: training, testing and validation. To eliminate bias in the first two sets, I'd show the results from the validation set to the business owner. I need input to gauge whether the model accurately predicts customer churn and provides actionable results.”

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

Why You Should Say It: You can analyze the data until the cows come home, but a model is meaningless unless it produces actionable results. An experienced analyst varies his or her strategy depending on the type of data being analyzed and the desired results. For instance, was a customer complaint retweeted on Twitter? If so, should that data be included or excluded? Plus, sensitive data has to be protected, so it's best to consult with the business owner to make sure you're following compliance regulations, disclosure laws and so forth.

Sample answer –

“First of all, I'd consult with the business owner from the outset to understand their objectives in categorizing this data. Then, I would use an iterative process, pulling new samples and modifying the model accordingly, evaluating for accuracy and inclusivity along the way. I'd follow the basic process of mapping the data, creating an algorithm and legend, mining the data, visualizing the data and so forth. However, I would tackle the project in segments in order to solicit feedback from the business stakeholder, and to continue enriching the model to ensure that I'm producing actionable results.”

Describe the abilities you have in order to work with us as intelligence analyst?

I have the ability to combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events), read and understand information and ideas presented in writing, apply general rules to specific problems to produce answers that make sense, listen to and understand information and ideas presented through spoken words and sentences, communicate information and ideas in speaking so others will understand.

What are the knowledge elements you obtained from your education, training and work experience would support your intelligence analyst career?

The Knowledge of the structure and content of the English language including the meaning and spelling of words, rules of composition, and grammar, laws, legal codes, court procedures, precedents, government regulations, executive orders, agency rules, and the democratic political process, relevant equipment, policies, procedures, and strategies to promote effective local, state, or national security operations for the protection of people, data, property, and institutions, media production, communication, and dissemination techniques and methods. This includes alternative ways to inform and entertain via written, oral, and visual media, circuit boards, processors, chips, electronic equipment, and computer hardware and software, including applications and programming.

How would you describe (analyst or your) work style?

My work style matching exactly what cashier job requires by: analyzing information and using logic to address work-related issues and problems, being careful about detail and thorough in completing work tasks, being honest and ethical, a willingness to take on responsibilities and challenges, being reliable, responsible, and dependable, and fulfilling obligations.

Tell about yourself and why you think you are successful intelligence analyst?

Why do you like to work as intelligence analyst?

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

The best way to approach this question is to mention that it is good to check with the business owner and understand their objectives before categorizing the data. Having done this, it is always good to follow an iterative approach by pulling new data samples and improving the model

accordingly by validating it for accuracy by soliciting feedback from the stakeholders of the business. This helps ensure that your model is producing actionable results and improving over the time.

Python or R – Which one would you prefer for text analytics?

This would depend on your competence with R or Python.

Sample answer –

I prefer Python because it has Pandas library that provides easy to use data structures and high performance data analysis tools.

Which technique is used to predict categorical responses?

Classification technique is used widely in mining for classifying data sets.

What is logistic regression? Or Explain a situation when you have used logistic regression recently.

Logistic Regression often referred as logit model is a technique to predict the binary outcome from a linear combination of predictor variables. For example, if you want to predict whether a particular political leader will win the election or not. In this case, the outcome of prediction is binary i.e. 0 or 1 (Win/Lose). The predictor variables here would be the amount of money spent for election campaigning of a particular candidate, the amount of time spent in campaigning, etc.

What are Recommender Systems?

A subclass of information filtering systems that are meant to predict the preferences or ratings that a user would give to a product. Recommender systems are widely used in movies, news, research articles, products, social tags, music, etc.

Why data cleaning plays a vital role in analysis?

Cleaning data from multiple sources to transform it into a format that data analysts or data scientists can work with is a cumbersome process because - as the number of data sources increases, the time take to clean the data increases exponentially due to the number of sources and the volume of data generated in these sources. It might take up to 80% of the time for just cleaning data making it a critical part of analysis task.

Differentiate between univariate, bivariate and multivariate analysis.

These are descriptive statistical analysis techniques which can be differentiated based on the number of variables involved at a given point of time. For example, the pie charts of sales based on territory involve only one variable and can be referred to as univariate analysis. If the analysis attempts to understand the difference between 2 variables at time as in a scatterplot, then it is referred to as bivariate analysis. For example, analyzing the volume of sale and a spending can be considered as an example of bivariate analysis. Analysis that deals with the study of more than two variables to understand the effect of variables on the responses is referred to as multivariate analysis.

What do you understand by the term Normal Distribution?

Data is usually distributed in different ways with a bias to the left or to the right or it can all be jumbled up. However, there are chances that data is distributed around a central value without any bias to the left or right and reaches normal distribution in the form of a bell shaped curve. The random variables are distributed in the form of an asymmetrical bell shaped curve.

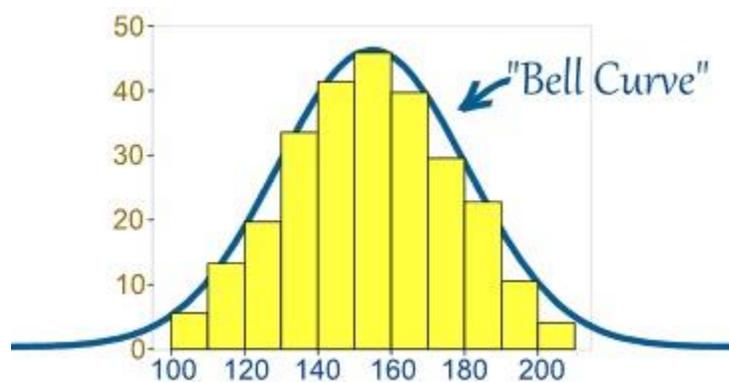


Image Credit : mathisfun.com

What is Linear Regression?

Linear regression is a statistical technique where the score of a variable Y is predicted from the score of a second variable X. X is referred to as the predictor variable and Y as the criterion variable.

What is Interpolation and Extrapolation?

Estimating a value from 2 unknown values from a list of values is Interpolation.

Extrapolation is approximating a value by extending a known set of values or facts.

What is power analysis?

An experimental design technique for determining the effect of a given sample size.

What is K-means? How can you select K for K-means?

What is Collaborative filtering?

The process of filtering used by most of the recommender systems to find patterns or information by collaborating viewpoints, various data sources and multiple agents.

What is the difference between Cluster and Systematic Sampling?

Cluster sampling is a technique used when it becomes difficult to study the target population spread across a wide area and simple random sampling cannot be applied. Cluster Sample is a probability sample where each sampling unit is a collection, or cluster of elements. Systematic sampling is a statistical technique where elements are selected from an ordered sampling frame. In systematic sampling, the list is progressed in a circular manner so once you reach the end of the list, it is progressed from the top again. The best example for systematic sampling is equal probability method.

Are expected value and mean value different?

They are not different but the terms are used in different contexts. Mean is generally referred when talking about a probability distribution or sample population whereas expected value is generally referred in a random variable context.

What does P-value signify about the statistical data?

P-value is used to determine the significance of results after a hypothesis test in statistics. P-value helps the readers to draw conclusions and is always between 0 and 1. P- Value > 0.05 denotes weak evidence against the null hypothesis which means the null hypothesis cannot be rejected. P-

value ≤ 0.05 denotes strong evidence against the null hypothesis which means the null hypothesis can be rejected. P-value=0.05 is the marginal value indicating it is possible to go either way.

Do gradient descent methods always converge to same point?

What are categorical variables?

A test has a true positive rate of 100% and false positive rate of 5%. There is a population with a 1/1000 rate of having the condition the test identifies. Considering a positive test, what is the probability of having that condition?

How you can make data normal using Box-Cox transformation?

What is the difference between Supervised Learning an Unsupervised Learning?

If an algorithm learns something from the training data so that the knowledge can be applied to the test data, then it is referred to as Supervised Learning. Classification is an example for Supervised Learning. If the algorithm does not learn anything beforehand because there is no response variable or any training data, then it is referred to as unsupervised learning. Clustering is an example for unsupervised learning.

Why is vectorization considered a powerful method for optimizing numerical code?

What is the goal of A/B Testing?

It is a statistical hypothesis testing for randomized experiment with two variables A and B. The goal of A/B Testing is to identify any changes to the web page to maximize or increase the outcome of an interest. An example for this could be identifying the click through rate for a banner ad.

What is an Eigenvalue and Eigenvector?

Eigenvectors are used for understanding linear transformations. In data analysis, we usually calculate the eigenvectors for a correlation or covariance matrix. Eigenvectors are the directions along which a particular linear transformation acts by flipping, compressing or stretching. Eigenvalue can be referred to as the strength of the transformation in the direction of eigenvector or the factor by which the compression occurs.

What is Gradient Descent?

How can outlier values be treated?

Outlier values can be identified by using univariate or any other graphical analysis method. If the number of outlier values is few then they can be assessed individually but for large number of outliers the values can be substituted with either the 99th or the 1st percentile values. All extreme values are not outlier values.

How can you assess a good logistic model?

There are various methods to assess the results of a logistic regression analysis- Using Classification Matrix to look at the true negatives and false positives. Concordance that helps identify the ability of the logistic model to differentiate between the event happening and not happening. Lift helps assess the logistic model by comparing it with random selection.

What are various steps involved in an analytics project?

Understand the business problem Explore the data and become familiar with it. Prepare the data for modeling by detecting outliers, treating missing values, transforming variables, etc. After data preparation, start running the model, analyze the result and tweak the approach. This is an iterative step till the best possible outcome is achieved. Validate the model using a new data set. Start implementing the model and track the result to analyze the performance of the model over the period of time.

How can you iterate over a list and also retrieve element indices at the same time? Ans

This can be done using the enumerate function which takes every element in a sequence just like in a list and adds its location just before it.

During analysis, how do you treat missing values?

The extent of the missing values is identified after identifying the variables with missing values. If any patterns are identified the analyst has to concentrate on them as it could lead to interesting and meaningful business insights. If there are no patterns identified, then the missing values can be substituted with mean or median values (imputation) or they can simply be ignored.

Explain about the box cox transformation in regression models.

Can you use machine learning for time series analysis? - Yes, it can be used but it depends on the applications.

Write a function that takes in two sorted lists and outputs a sorted list that is their union.

What is the difference between Bayesian Inference and Maximum Likelihood Estimation (MLE)?

What is Regularization and what kind of problems does regularization solve?

What is multi-collinearity and how you can overcome it?

What is the curse of dimensionality?

How do you decide whether your linear regression model fits the data?

What is the difference between squared error and absolute error?

Which is your favorite machine learning algorithm and why?

In which libraries for Data Science in Python and R, does your strength lie?

What kind of data is important for specific business requirements and how, as a data scientist will you go about collecting that data?

Tell us about the biggest data set you have processed till date and for what kind of analysis.

Suppose you are given a data set, what will you do with it to find out if it suits the business needs of your project or not.

What were the business outcomes or decisions for the projects you worked on?

What unique skills you think can you add on to our data science team?

Why do you want to pursue a career in data science?

What have you done to upgrade your skills in analytics?

What has been the most useful business insight or development you have found?

How will you explain an A/B test to an engineer who does not know statistics?

When does parallelism help your algorithms run faster and when does it make them run slower?

How can you ensure that you don't analyze something that ends up producing meaningless results?

How would you explain to the senior management in your organization as to why a particular data set is important?

Is more data always better?

How do you go about collecting information to support your analyses?

What types of data have you researched and analyzed in the past?

The employer might need data analyses to create new advertising strategies, prepare short and long-term finance budgets, or determine which company products are most profitable. Answer data-collecting questions with specific examples of how you successfully used group samples, conducted market research, reviewed financial reports or analyzed surveys to make fair and consistent assessments.

What are your computer skills and experience using analytic software?

Data analysts process collected data and reach conclusions with the help of computer software, according to the U.S. Bureau of Labor Statistics. Discuss any experience you've had with statistical software, such as Stata, RStudio, PSPP or GMDH Shell. If most of your previous work has been with inter-office spreadsheets or Microsoft Excel files, assure the interviewer that you are proficient with those types of data files and would be willing to learn any new software programs necessary for the job.

Explain how you organize and create presentations to report analytical findings?

Answer these questions with specific examples of presentations, reports and seminars you've created or hosted. The interviewer wants assurance that you have the people skills and interpersonal strengths to effectively relay your analyses and results.

What is meant by Data Analytics?

Data analytics (DA) is the science of examining raw data with the purpose of drawing conclusions about that information. A data warehouse is often built to enable Data Analytics

What do you do in the current job? What is your education background? Describe your experience.

Explain about big-data and its applications?

How can you handle databases as a data analyst?

Assume, you develop the big data model, but the customer feels difficult to use your model? How do you communicate with the customer at this point?

R can't handle large amounts of data so what are the others ways to handle data?

What is Multicollinearity? How can we solve it?

Explain about the clustering methods you are familiar with?

Define P-value?

What are your strongest technical skills?

Difference between statistical and machine learning?

Difference between recurrent neural networks and recursive neural networks?

What is your key strength in the business analyst?

Say about your project that you are working as a data analyst?

What is Collaborative filtering / Recommendation Engine?

Tell about most challenging project you have done so far.

How will you design a Recommender (recommendation) System with more optimizations and variants?

Describe a recent use of logistic regression.

Describe a recent Data science case study.