

What does exploratory vs. explanatory analytics mean to you?

What characteristics differentiate a good vs. a great analyst / visualizer / etc? (label depends on role, but IMHO, answers generally do not)

Describe your analytics workflow

What is your favourite function/plugin/package in R?

I tried a role-play scenario once where I encouraged them to ask me questions about data, it was a good way to see their thinking process.

Me "you get an email from a colleague with an attachment. It says 'could you have a look at this for me?' What do you do?"

Practically there were three useful conversations that could result - Data and IT security awareness, approach to data problems, or work management. All were useful to me.

I avoid asking technical questions even for technical roles. I want to get an understanding of their attitude and personality. Any data technique, skill, or other are easy to acquire in the job. Come equipped with real world examples and what you have learned from them.

Explain what a sampling distribution is.

I've asked people to explain what checks they would do on their data before starting any analysis (eg is it survey data vs full population, is it based on measurement or human judgment) it helps me understand people's thinking around data quality (rubbish in = rubbish out)

What programming languages do you know?

What ML frameworks do you use? What is your process for determining which ML algorithm is appropriate for a project?

Provide an example of a project that failed due to bad data--what did you learn?

It's the communication ones that are key. Summarising info, focusing on answering the question they are asking, thinking carefully about visualisation choices.

SQL: What gives the longest table, inner or left join?

What have you been doing to keep my skills up to date (I had that at most of my interviews)

What are the three most important things you would advise data managers do with their data to make your job as a data analyst easier?

Personally by the interview stage I'm trying to find out more about the person rather than their knowledge and skills. Motivation, attitude, empathy, reflection, curiosity, focus,

communication. Give examples from work, study or personal life. You can teach skills and knowledge

[@isd\\_scotland](#) we ask competency based questions around areas such as experience of working with [#data](#) and [#analytical](#) skills, teamwork, communication and data interpretation.

What are the three most important things you would advise data managers do with their data to make your job as a data analyst easier?

For my Scottish government interview I got a lot of questions about SQL, version control and my working style, but that wasn't a data analysis role so not sure if it's relevant.

A few questions I've found in job applications:

1. What was an interesting data problem you worked on within the last year? How did you identify and address it?
2. Describe a situation where you did not have access to all of the data needed to triage a problem or analyze a situation, and how you adapted to it.
3. How would you measure customer lifetime value

One interview was very structured, with scripted competency-based questions in the format "describe a time when you..." The other was more informal, but both covered similar topics, and were a lot less technical than I was expecting.

One major theme was how would I deal with introducing new techniques and initiating/leading change within the team. Another was how do I feel about learning new things/how do I keep my skillset up to date.

Prior to one of the interviews I was asked to analyse some data that would be used for service planning, and produce a one-page report. During the interview they asked how I checked whether the data was clean and reliable, and we discussed some of the findings. It had involved some descriptive analysis, but there wasn't enough data to build a good predictive model. They asked what additional data would be needed and how I would obtain it.

During the interview I was also asked to interpret a fairly simple time series graph. There were a couple of obvious outliers and I was asked to suggest reasons why these might have occurred, and ways in which I might investigate them further. That required some familiarity with the context (hospital admissions data in this case).

In the other interview I was shown a spreadsheet and asked how I would improve it. Basically a question on data cleaning, to see if I would spot badly formatted dates, missing data, typos etc. I was also asked how I'd perform certain tasks in Excel—my response was that I'd use Python or R instead, which was apparently the right answer!

There were also quite a few questions on management and soft skills. In particular, about dealing with “difficult” co-workers.

#### Technical:

1. Which do you feel are the important skills to have in R for data science?
2. How would you go about removing personal information from a dataset before publishing it?
3. Give examples of some of the issues you might face when dealing with a large dataset. How would you address them?
4. What's your favourite R package? Why?
5. You're managing a cinema and have a database with tables storing details of the films showing and customers visiting. They have a many-to-many relationship with tickets acting as a join table between them. How would you find the total number of customers who have seen a given film? Would you use R or SQL?
6. What would be some of the challenges in cleaning a dataset containing personal information gathered from web forms?
7. Explain the concept of a confidence interval.
8. What is the difference between explanatory and predictive analysis?
9. What is overfitting? Why is it a problem?
10. What is the difference between sample and population?
11. How do you create a relationship between tables in a SQL database? Give an example of such a relationship.
12. What is a p-value?
13. What do you currently do or intend to do to keep your technical skills up-to-date?
14. What is the difference between trend and seasonality in a time series model?
15. What gives the longest table: an inner join or a left join?
16. What is imputation for missing values?
17. What checks would you perform before analysis?

#### Business:

1. How might you develop a system to show trends in students' exam performance?
2. How might privacy considerations impact the data gathering process of a major customer-facing company?
3. A courier company is tracking the routes taken by its drivers each day. They know the location of each delivery, size of the package(s) and which vehicle is being used. They are using metrics such as average time per delivery and total number of deliveries completed to measure performance. How would you go about visualising that data to help them optimise their delivery schedule? What additional information (if any) would you need to help you?

4. What data would you need to create a dashboard showing purchasing trends for a specific product on Amazon? There are many factors to consider: season (especially for clothing and leisure equipment); sale periods; similar products appearing in searches