# Technical Interviews & Assessments for Data Scientists, Analysts, and Engineers

***Check out these articles from our Blog, they're all about Data and might be interesting to you😎***

- What Data Scientists Do at takealot.com, Zoona, JUMO, and Curately
- How Luno uses Data to make Product Decisions
- How I got started with Data Visualisation
- DIY: Growing Chilli-Plants with an Augmented Reality System
- How I use Jupyter Notebooks as a Sandbox for building visualisations

***We've also gathered some tips from another source that we think are pretty cool! 💻***

## The Technical Assessment

The most common step in the Data Science interview process after an initial phone call is an assessment of your technical skills. This can be a dataset for you to analyze, a coding assessment, or a project that they would like you to present on. Companies may test you on your ability to work with ambiguity (e.g. here's a dataset, find some insights and pitch to business

stakeholders) or focused on a more concrete deliverable (e.g. Here's some data, build a classifier).

While all assessments may differ in their objectives, the common denominator is that you'll be receiving data from the company. So regardless of what they've asked you to do, the first step will always be Exploratory Data Analysis. Luckily there are some automated EDA solutions, such as SpeedML.

Primarily what you want to do here is investigate peculiarities in the data. More often than not, the company will have synthetically generated the data, leaving specific easter eggs for you to find (e.g. A power law distribution with customer revenue).

**These are Ken Jee from Towards Data Science's Top Tips**

- If they don't explicitly say you can't bring in data from outside sources, consider appending other useful data.
- Using feature engineering or feature reduction (and being able to explain why) can improve the quality of your analysis greatly.
- Putting some time to making your work interpretable to business stakeholders (detailed visuals) can show that you understand how to drive business value.
- Pro-Tip: Make sure to save every challenge you do. You never know when you may need to reuse a component in future challenges.

## Technical Interview

Expect to be asked about the projects on your resume and the logic behind your methodology. You will likely be asked to explain some of the math behind the algorithms that you used so make sure you're prepared with some revision of the theory. If they have any questions about the take-home test, they will generally ask them here as well. Since this is a project that they are familiar with (they may have taken the assessment themselves), be prepared for questions about the data cleaning, model choice, etc. Be prepared to reflect on what you did in detail and what result was achieved. Prepare to explain how you would improve on what you did.

The most common question asked in a Data Science interview is one where the interviewer asks you to pick an algorithm and explain how it works, what it tries to achieve and the pros and cons of using that particular algorithm. You can prepare an answer to this question ahead of the interview and adapt it to the question you get posed. Understand the performance metrics you would use to measure whether the algorithm is working and at what level.

Review general data science interview questions, these are generally statistics based. Without an advanced knowledge of statistics, it is difficult to succeed as a data scientist.

Be prepared to answer some fundamental statistics questions as part of your data science interview. For some common examples use the links under the resources heading below.

There will likely also be a programming section of the interview- language-wise, Python, SQL and R are typically the three that you could be asked to write in, however, this can differ based on the role and company. You may be asked to solve a problem on a whiteboard or using a computer so practice solving questions on a piece of paper or on a whiteboard with a partner present.
For more possible interview questions see the resources below.

## Commonly Asked Questions

- How do you select features for a model? What do you look for?
- What's the difference between logistic regression and linear regression?
- Explain decision trees.
- How would you test whether a new credit risk scoring model works?
- Explain K-means clustering and when it's useful.
- If you have more than one trained model, how do you assess which is best?
- Explain the bias-variance tradeoff and how you navigate it.
- What is the Central Limit Theorem?
- What are the assumptions of a linear model (or any other type of model)?
- What's the difference between K Nearest Neighbor and K-means Clustering?
- How do you address overfitting?
- Explain Naive Bayes algorithms.
- How do you find and correct biases in your data?
- What is cross-validation?
- What are confounding variables?

## Additional Resources

- If you're pushed for time here are the top 30 interview questions with answers and explanatory graphs
- Access commonly asked technical questions which can be filtered by difficulty
- A sheet created by Towards Data Science which covers various aspects of Data Science in detail with links to relevant articles and videos on each topic
- A free 1 week course with the aim to have you fully prepped for a technical interview
- To practice SQL questions- you can use resources such as SQL Zoo and Mode Analytics

## Technical Interview Nuances for Data Analysts

As a Data Analyst, you are the bridge between the business and the analytics. It is important to prepare for statistics questions as this role demands a high-level understanding of statistics as well as Business Intelligence tools. This is less of a technical role and more of a business role within the company and as such, there is no emphasis on programming skills.

**Common Questions for a Data Analyst**

- Explain the data analysis process.
- Why is data cleaning important?
- What kinds of problems would you look for when cleaning a data set?
- How would you get a data table from a web page into your code for analysis?
- How would you combine these two tables using SQL/Python/R?
- How would you estimate the number of windows in San Francisco? (There is an infinite variety of “logic” questions like this that are meant to test your logic and statistics skills).
- How would you sort the rows of this table numerically using SQL/Python/R?
- What kind of data would you want to collect to solve a specific business problem?
- What methods would you use to analyze the comparative performance of two different product search engines?
- In SQL, what’s the difference between Union vs. Union All? Union vs Join? Having vs Where?
- Explain random sampling, stratified sampling, and cluster sampling.
- Talk about a time you’ve worked with a large database or data set
- What are Z-scores and how are they useful?
- What would you do to analyze the best way for us to improve conversion rates for our users?
- What’s the best way to visualize this data and how would you do that using Python/R?
- If you were going to analyze our user engagement, what data would you collect and how would you analyze it?
- What’s the difference between structured and unstructured data?
- What is a p-value?
- How do you handle missing values in a data set?
- If an important metric for our company stopped appearing in our data source, how would you investigate the causes?

## Technical Interview Nuances for Data Engineers

You need to be able to understand and can track data at every point while making sure everything works after every layer of data that you add. You need to be able to ask whether each step is the most efficient way to add the layers and continue to think this way while transporting the data and know why you are making each decision.

The Data Engineer will use similar prep to the Data Scientist with the following differences:

- Some familiarity with statistics but a more basic understanding is required as opposed to a Data Scientist.
- Can you do some data manipulation, transform data.
- Wouldn't be asked so much about models. Lot less modeling and more what happens before it.
- SQL is heavier for engineers as it is their primary tool.

- Familiarity with databases is important, which have you worked with and what are the differences between databases. Why you would go with one over another? Knowing how they sort data and why that is beneficial over another one is important**.**

---

Sources:
https://towardsdatascience.com/mastering-the-data-science-interview-15f9c0a558a7
https://towardsdatascience.com/the-data-science-interview-what-to-expect-15deecea35cc
https://towardsdatascience.com/the-data-science-interview-what-to-expect-15deecea35cc
https://towardsdatascience.com/preparing-for-data-science-interviews-49500c51fec4