How to Ace Your Data Science Interview

Author, Nicholas Deveney

A Data Science Foundation Blog

August 2017

\-

www.datascience.foundation

Copyright 2016 - 2017 Data Science Foundation

With dissertation deadlines glooming, Data Science students are gearing up to leave the academic world and find their feet in a data science role. We all know the demand for the skills and the shortage supply of experienced data scientists means there are opportunities everywhere and companies are looking to secure grad talent, so finding some data science jobs should not be too difficult.

But before you reach that commercial goldmine, your faced with the job interview. Not matter how much experience and exposure you have in previous interviews, public speaking or Data science discussions, this preparation is still hard.

Data Science interviews tend to cover a wide range of topics. From technical exposure, to statistical understanding, to solving and communicating complex business problems.

At Eden Smith we work with a number of business in hiring across the Data science spectrum and to support you ace your interview have curated a list of common Data Science interview questions. We have enriched this data with information from online and insight from our Data Science partners to help you prepare for the types of questions that can be thrown at you during your Data Science Interview.

Building Models

Building data models for machine learning or pure data transformation and analysis is one of the most common tasks of the modern data scientist and more and more businesses are developing teams, particularly with grads, that are modelling and coding heavy, this is resulting in more interviews covering the various modelling techniques and statistical theories.

Not all interviews will be technical, but below are some questions that will help you prepare and refamiliarize yourself with.

1. **How would you create a logistic regression model?**
2. **What is linear regression? What do the terms P-value, coefficient, R-Squared value mean? What is the significance of each of these components?**
3. **Why is Central Limit Theorem important?**
4. **Explain hash table collisions?**
5. **In your opinion, which is more important when designing a machine learning model: Model performance? Or model accuracy?**

6. **What are some situations where a general linear model fails?**
7. **Is it better to have too many false positives, or too many false negatives?**
8. **How would you validate a model you created to generate a predictive model of a quantitative outcome variable using multiple regression?**
9. **What is an example of a dataset with a non-Gaussian distribution?**
10. **Explain Bayes Theorem, when might you use Bayes Inference?**

Programming

Most Data Science teams are involved in both the ingestion of data for modelling and analysis and the production of models into the enterprise environment. Whether this is led by Data Engineering, software engineering or a database development team, you will be expected to have a strong understanding of various program languages, those directly involved in Data Science and those surrounding data integration and exportation. Be sure to brush up on your Python, R, SQL and relevant big data programming languages such as Scala.

1. **Python or R, which would you prefer for text analysis?**
2. **What modules/libraries/Packages are you most familiar with? What do you like or dislike about them?**
3. **What are the different types of sorting algorithms available in R language?**
4. **What is the difference between a tuple and a list in Python?**
5. **How do you split a continuous variable into different groups/ranks in R?**
6. **What is the purpose of the group functions in SQL? Give some examples of group functions.**
7. **Tell me the difference between an inner join, left join/right join, and union.**
8. **Describe a data science project in which you worked with a substantial programming component. What did you learn from that experience?**
9. **How would you clean a dataset in your programming preference?**
10. **What are the two main components of the Hadoop Framework?**

Data Science Process

Although being hands on with data and modelling and programming are the major aspects of any data science in today's world; often businesses are looking to understand how insights and results are created. Interviewers are looking for you to demonstrate a clear understanding and be able to explain various methods and processes throughout a data science project and their Pros. Cons and use cases to a non-technical audience. Practice articulating and giving clear simple explanations of various complex data science procedures.

Data Science Foundation

Data Science Foundation, Atlantic Business Centre, Atlantic Street, Altrincham, WA14 5NQ
Tel: 0161 926 3670   Email:admin@datascience.foundation Web: www.datascience.foundation
Registered in England and Wales 4th June 2015, Registered Number 9624670

1. **What are various steps involved in an analytics project?**
2. **What is the goal of A/B Testing?**
3. **Explain the use of Combinatorics in data science?**
4. **What is the difference between Cluster and Systematic Sampling?**
5. **What is logistic regression? Or State an example when you have used logistic regression recently.**
6. **Explain false negatives and false positives, which is better to have too many of?**
7. **What was the business impact of your last project?**
8. **Can you explain the difference between a Test Set and a Validation Set?**
9. **What makes a dataset gold standard?**
10. **What are outliers and inliers? What would you do if you find them in your dataset?**

General

Data Science is still a position that has great variety and a lack of standardisation across the market. Therefore, every Data Science position and company you interview for will take a slightly different approach and expect additional skills and awareness of the surrounding subjects. Be sure to explore the business your interviewing with, check current employees; data scientists and analysts and see what additional products, technologies and soft skills they have experience with. Some common general questions are:

1. **What visualisation tools are you familiar with?**
2. **Explain a time when you had to handle stakeholder's expectations?**
3. **Describe a time when you have been innovative and creative?**
4. **Which Cloud services have you used and how have you interacted with them?**
5. **What external data sources do you think could be interesting to our domain?**
6. **Present to us your last data science project.**
7. **What's a project you would want to work on at our company?**
8. **What data would you love to acquire if there were no limitations?**
9. **How important is the product in Data Science?**

Eden Smith

If you want more advice or support in how to land your dream data science opportunity or if you're a manager looking to scale a data science team *get in touch with us today*.

About the Data Science Foundation

The Data Science Foundation is a professional body representing the interests of the Data Science Industry. Its membership consists of suppliers who offer a range of big data analytical and technical services and companies and individuals with an interest in the commercial advantages that can be gained from big data. The organisation aims to raise the profile of this developing industry, to educate people about the benefits of knowledge based decision making and to encourage firms to start using big data techniques.

Contact Data Science Foundation

Email: admin@datascience.foundation
Telephone: 0161 926 3641
Atlantic Business Centre
Atlantic Street
Altrincham
WA14 5NQ
web: www.datascience.foundation