

Ken Jee

Starting a Career in Data Science

Course Notes

365[✓]DataScience

Table of Contents

1. Job Description	4
1.1 Data science roles.....	5
1.2 The interview process	5
2. Portfolio	6
2.1 What is a data science project	7
2.2 How to differentiate your projects.....	7
2.3 Best practices on GitHub	8
2.4 How to build a Kaggle profile	8
3. Resume Best Practices	9
3.1 Sections.....	10
3.2 Storytelling	10
3.3 Additional Tips.....	11
4. Networking practices	12
5. Phone Interview Process.....	13
6. Take-home Assignments	14
6.1 Types of take-home assessments:	14
6.2 Take-home test tips	14
7. The In-Person Interview	16
7.1 Answering Questions	16
7.2 The Briefcase Method	17

Abstract

Data science jobs are hyper-competitive. For each position, there are multiple other highly qualified candidates eyeing the same role. It is like you are all competing for a \$100,000+ prize. If you frame it this way, wouldn't you want to go the extra mile?

By studying these course notes you will be doing just that. You will learn valuable information that can give you a much-needed edge over other candidates.

What better way to approach data science job hunting than learning from the experience of someone who is an actual data scientist and has recruited data scientists for his team?

You will learn how to:

- Land a job in data science
- Build your resume
- Succeed during the phone interview
- Ace the behavioral and technical questions
- Create your data science project portfolio
- Get an interview through networking
- Solve the take home test

Keywords: data science career, data science portfolio, take-home tests, phone interview

1. Job Description

A data scientist is someone who uses math, computer science, and business logic to solve problems. Usually, data scientists work on large datasets, but sometimes they work with small datasets too.

Data collection, data cleaning, data exploration, model building, explaining models, model deployment are all things that data scientists do on the job. These activities together are known as the data science lifecycle.

To work as a data scientist, you need 1 programming language (Python, R) + a query language like SQL.

In this way, you will be able to:

- Manipulate data
- Implement relevant algorithms
- Visualize outcomes

A data scientist needs solid quant skills. You should have a solid foundation in:

- Statistics
- Linear algebra
- Calculus
- Discrete math

To communicate your findings - Tableau and Power BI

Get your programming and math skills to an intermediate level before applying for jobs.

You may ask "How will I know when I have learned enough?"

You will be ready when you can implement the recommended projects on your own.

1.1 Data Science Roles

Within the broader umbrella of data science, we have several subcategories:

- Data analyst – data cleaning, data exploration
- Data engineer – data collection, data cleaning
- ML engineer – working on ML models
- Data scientist

In smaller companies, you can work on more areas of the business. In large companies, you specialize in one area.

If you have expertise in a specific area – you have good chances of landing a role when a company is looking to hire someone who will work in that area.

Evaluate your experience – aim for roles that fit your current expertise. If you don't have prior experience, you might want to start as a data analyst.

1.2 The Interview Process

Most interviews start with a phone screen. The goal is to hear about your past experiences, but sometimes you might also be asked a technical question.

Next, you will have a technical assessment. Often, data scientists are invited to analyse a dataset. Some companies might give you a test in which you have to answer data science-related questions. These are not too demanding.

If you pass these stages, you will be asked for an in-person interview. Expect to meet with a data scientist on the team, a data science manager, and a representative from HR. You can anticipate behavioural and technical questions.

1.3 The Perfect Candidate

Companies look for 3 things in a candidate:

- 1) Skills to excel at work - coding, model building, deployment, understanding of business concepts
- 2) Will get along with the team - show you communicate well
- 3) Excited about the company and will stay for a while - you need to show interest

Character traits you should aim to develop:

- 1) Autonomy - project work will help you do that
- 2) Growth mindset - it is difficult to do all of this, but you need to be mentally prepared
- 3) Always be learning
- 4) Be comfortable about your work - you need to be able to make your work understandable by others

2. Portfolio

Experience - the single largest factor in getting a data science job. It might seem surprising, but you don't need a job to get data science experience.

Projects can show you're a self-starter and that you can work autonomously. Doing projects is the best way to learn data science. While doing project work, you get yourself dirty and you realize your shortcomings.

2.1 What is a Data Science Project

Planning phase - Find a problem to be solved or data to be explored; Data collection (collect your own data or use Kaggle or Google Datasets); data cleaning (make the data usable); exploratory analysis (find trends and highlight them with visuals); find the best models suitable for your use case; model deployment (optional for GitHub);

Do a retrospective to evaluate how you could have done better.

You get to choose the projects that are interesting to you.

In data science, we have 3 main types of problems to solve:

- 1) Regression
- 2) Classification
- 3) Clustering

Do at least 1 project on each of these problem types. Do a project on advanced topic like Deep learning, Image classification, or NLP.

Try to have 4-5 projects in GitHub, Kaggle.

2.2 How to differentiate your projects

1. Unique projects - If you tackle a problem that hasn't been explored before
2. Project that provides value for someone - perhaps help a non-profit
3. Use unique data you have scraped
4. Having new features in your model improves performance; for ex. If you have latitude and longitude data, you can create a new feature that is the distance from a common location.

5. Solving a problem using different models

6. Deploy your model

7. Publish your work. Having academic credentials is a badge of honor.

Once you have a project, you should add it to: GitHub, Bitbucket, Kaggle, or your own website. Make sure your code is clean and the code is clearly explained

2.3 Best Practices on GitHub

- A high-quality photo is highly recommended.
- It is recommended to be relatively active on GitHub and show a certain level of engagement from time to time.
- The 'readme' is a high-level overview of the project.
- Make sure you delineate the project's goals and the outcomes.
- Make sure you communicate the value you're creating. Describe the packages you used.
- If some of the features in your dataset might be confusing you want to go in depth and elaborate on what each of them means.
- Make sure to include the exploratory data analysis and some pictures of your findings.
- The main goal is to build a reasonable readme that is easy to digest.

2.4 How to build a Kaggle profile

Make sure you have either a GitHub or a Kaggle profile (or both).

In your Kaggle profile, there are four main sections:

- Competitions
- Datasets
- Notebook Contributions – if other users upvote your analysis of datasets in Kaggle
- Discussion Contributions – an employer can see how active you're in the community

Similarly to GitHub, it's good to have a strong level of activity in Kaggle. This way employers will see that you're passionate about data science.

At the top of the description of your project make sure to include:

- Results
- Why you chose to work on this data
- Clearly outline the different stages of your work

Commenting out your code is a best practice, so that people are able to follow your thought process. In this way, they can understand your decision-making process.

It is best practice to have a summary table providing the results of the models you run. The takeaway is that you want to be clear about the decision and when in doubt, just go in more depth in your explanations.

3. Resume Best Practices

The resume isn't the only way employers find information about you. Your online presence is an extension of your resume. The resume isn't going to get you a job, but it can get you an interview. Optimize your resume for humans and for computers. Many companies scan resumes with Applicant Tracking Systems.

Use clear, legible fonts, and go for a modern look. Keep your resume to a single page.

3.1 Sections

Organize your resume in sections.

- 1) Technical aptitudes - Include programming abilities; Do not include general words and subjective qualities.
- 2) Education / Work experience - depending on which is your most recent experience; make sure you highlight your relevant work experience
- 3) Project work - use the same format as the one of your work experience section; be focused on the business value you've created
- 4) Add something short about yourself (ex. "5 books I read most recently")

3.2 Storytelling

Data scientists need to be able to tell a great story using data.

Makes sure your past job experiences feature 3 qualities:

1. Create value
2. Quantifiable
3. Action-oriented

Try to phrase what you did in a results-oriented manner. Instead of "Used xyz model to optimize growing conditions for a farm" try with "Increased crop yield by 12% over the 6-month growing season through the integration of XYZ model"

➔ 'Action verb' + 'Quantitative outcome' + 'method' (resonates best with recruiters)

3.3 Additional Tips

- Customize your resume for each application.
- Make sure to include the technical requirements mentioned in the job description.
- LinkedIn can be more prevalent than your traditional resume.
- Start with a personal statement at the top.
- For work experience - use an abbreviated version of your resume.
- Make sure to get some social proof by asking people for recommendations.
- Your resume can be longer than a page.
- A recruiter might reach out to you via your LinkedIn profile.

Resume checklist:

1. Add links to Kaggle and GitHub.
2. Customize for the position you're applying to.
3. Focus on outcomes instead of algorithms.
4. Use specific and quantifiable descriptions.
5. Show you have soft skills with experiences rather than with words.
6. Inject personality.
7. Don't have any grammatical errors.
8. Include your project experience.

Anything you do that other candidates won't do, gives you an advantage.

Don't spend the whole time talking about yourself, but how you helped the company. Try to answer the question: "How could I improve the company?" Start

your cover letter with something that you love about the company and explain how you would add value.

4. Networking Practices

Most positions are filled with networking or employee referrals. As many as 85% are filled in these manners. Referrals are preferred because of:

- Lower turnover
- Higher ROI
- Higher chance to accept the position

When you get a job, you are required to spend a significant amount of time with the people on your team, and when someone can speak to your character, this greatly reduces the risk of conflict within the team.

Networking works best when it isn't transactional.

3 best practices:

1. Ask good questions – keep the conversation flowing, encourage others to participate in the conversation
2. Connect on commonalities – can come in different shapes (data science, sports, etc.)
3. Telling interesting stories – have a few good stories in your back pocket

Make sure you follow up with people and look at these connections like planting seeds.

Connection building:

- LinkedIn – 1st and 2nd degree connections – easy introduction
- University alumni

- Hackathons and meetups
- Informational interview – get better understanding of a connection and their experience with work

Do research on the person. The goal is to learn from their experience.

Be more specific. Lower the risk and time loss for the person you reach out to.

Formula for a good intro: Introduction + Specific conversation topic + times and dates + value that you provide

Make sure to thank them for their time. Almost all recruiters have LinkedIn and read their InMail. The key is to make it sound personal. Some companies pay their recruiters based on their hires. Reaching out and asking a few specific questions can be great.

5. Phone Interview Process

Phone interviews are around 30 minutes that usually start with a general question about your background. Almost all interviews start with an introduction of the company. Sometimes questions are technical – in most cases related to statistics. At the end you might ask questions on your own. A good practice is to go on Glassdoor.com which can give you a good idea what you will be asked. Some other interview preparation methods include:

- Researching the company and check the news
- Throwing in some keywords from the job description
- Using stories for each of your previous work experiences – demonstrate what you achieved
- Preparing 2-3 well thought out questions on your own

- Doing as many mock interviews as possible
- Relaxing and treating the interview like a normal conversation

Show 3 things:

- 1) Genuine interest in the company
- 2) Technical skills
- 3) Your experience matches with what they're looking for

At the end of the phone interview do not forget to thank the recruiter for their time.

6. Take-home Assignments

6.1 Types of take-home assessments

- Take-home dataset. They might give you a target around which to build a model
- SQL or coding assessment - sometimes done live
- Written test - far less common

Take-home dataset problems

- 1) The data is open-ended, and you're expected to do an exploratory analysis
- 2) Build a specific type of model to solve a problem

6.2 Take-home Test Tips

Watch out for:

- Missing data
- No values
- Sparse data
- Data with different types of distributions

- Understanding data of different types

They want to see the tools you used and your business logic.

Exploratory analysis - Try to do the following:

- 1) Evaluate the data
- 2) Choose a few features that you want to examine
- 3) Build a model to predict or understand the values better

Determine whether the model that you're building is for understanding the data or for maximizing predictive power.

Build a specific type of model - Try to do the following:

- 4) Understand your data and create features
- 5) Try a few different models and explain why you chose them - make sure you can clearly explain the math behind the models you use
- 6) Use cross-validation properly
- 7) Tune your models and explore composites
- 8) Make your model easy to productionize

Data scientists, coding is easier (much less complex than for a software engineer)

SQL questions might be advanced

Glassdoor can help you find what questions you might get asked.

Simulate the real experience with a partner.

The interviewer must be able to assess your logic.

- 9) Focus on statistics and model-building questions. Understand the math behind models.

7. The In-Person Interview

The in-person interview will be done by:

- Technical recruiter
- Data science manager
- 1 or 2 other people: data scientists on the team, project manager, software engineer, or even CTO (depending on company size)

There are 3 types of in-person interviews:

- Behavioural interviews
- In-person assessments
- Technical interviews

Side note: You can expect 2-6 interviews.

7.1 Answering Questions

Interviewers are trying to understand whether you would be a good fit, so be prepared to answer, "Why our company?"

You'll also be asked questions regarding how you handled questions in the past.

Prepare 7-8 different stories from your past experience that can match almost any type of behavioural question you may get asked and use a clear structure when answering these questions.

Star methodology - Situation, Task, Action, Result

Rehearse these stories and be able to tell them. If you have a gap in your resume, you should be able to explain it. Outside of the actual content of the interview, try to convey that:

- 1) You're grateful for the opportunity

- 2) You're able to communicate well
- 3) You can speak towards the company mission
- 4) You can ask high quality questions

You will be either asked about coding and SQL or about data and math problems.

Therefore, success in this interview is about practice and pattern matching. You might get asked to solve the task on a whiteboard. When you do that, make sure to talk your way through your code. If an interviewer understands your logic and you make a mistake, they might help you out. It's normal to get stuck. So, if you talk your way out of a problem, there's a good chance they won't give you any negative marks. If you're stuck for a while, it's ok to ask for help.

Finally, following up can improve your chances of getting a job. Mention something specific about the interview with that person as it shows your individual touch, which shows you listened. You can also clarify an answer. Following up is your opportunity to do that.

7.2 The Briefcase Method

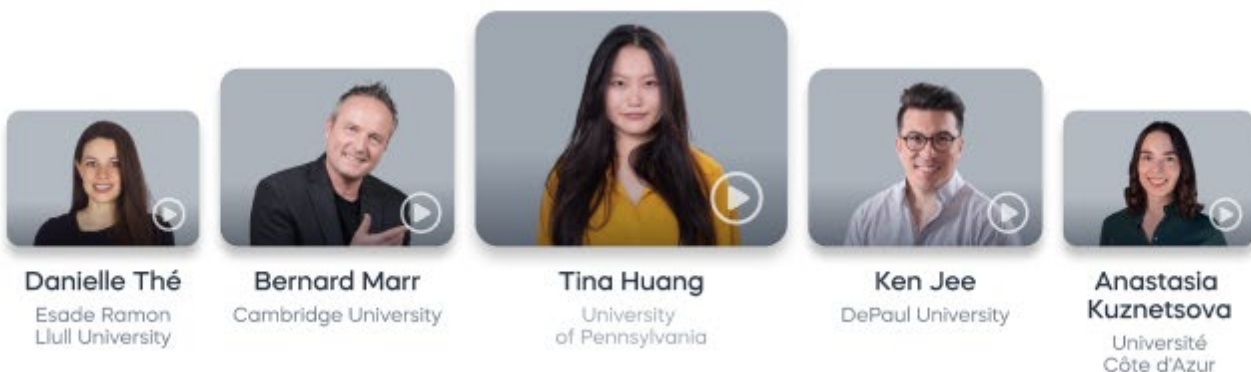
The briefcase method:

Before the interview you do very careful research about the company. You come to the interview prepared with a written document that contains a list of the projects you believe could add value to the company. You have an in-depth idea of the data needed, the feasibility, the timeline, and the outcomes. The idea is that when you show up, you would give them an idea of the value you will bring. Use this technique at the end of your interview. Make sure you leave a paper copy with them.

Learn DATA SCIENCE anytime, anywhere, at your own pace.

If you found this resource useful, check out our **e-learning program**. We have everything you need to succeed in data science.

Learn the most sought-after data science skills from **the best experts in the field!**
Earn a **verifiable certificate** of achievement trusted by employers worldwide and future proof your career.



Comprehensive training, exams, certificates.

- ✓ 162 hours of video
- ✓ 599+ Exercises
- ✓ Downloadables
- ✓ Exams & Certification
- ✓ Personalized support
- ✓ Resume Builder & Feedback
- ✓ Portfolio advice
- ✓ New content
- ✓ Career tracks

Join a global community of 1.8 M successful students with an annual subscription
at 60% OFF with coupon code **365RESOURCES**.

~~\$432~~ **\$172.80**/year



Start at 60% Off

VAT may be applied

A faint, light blue network diagram is visible in the background, consisting of several nodes connected by thin lines, creating a web-like structure.

Ken Jee

Email: team@365datascience.com

365[°]DataScience