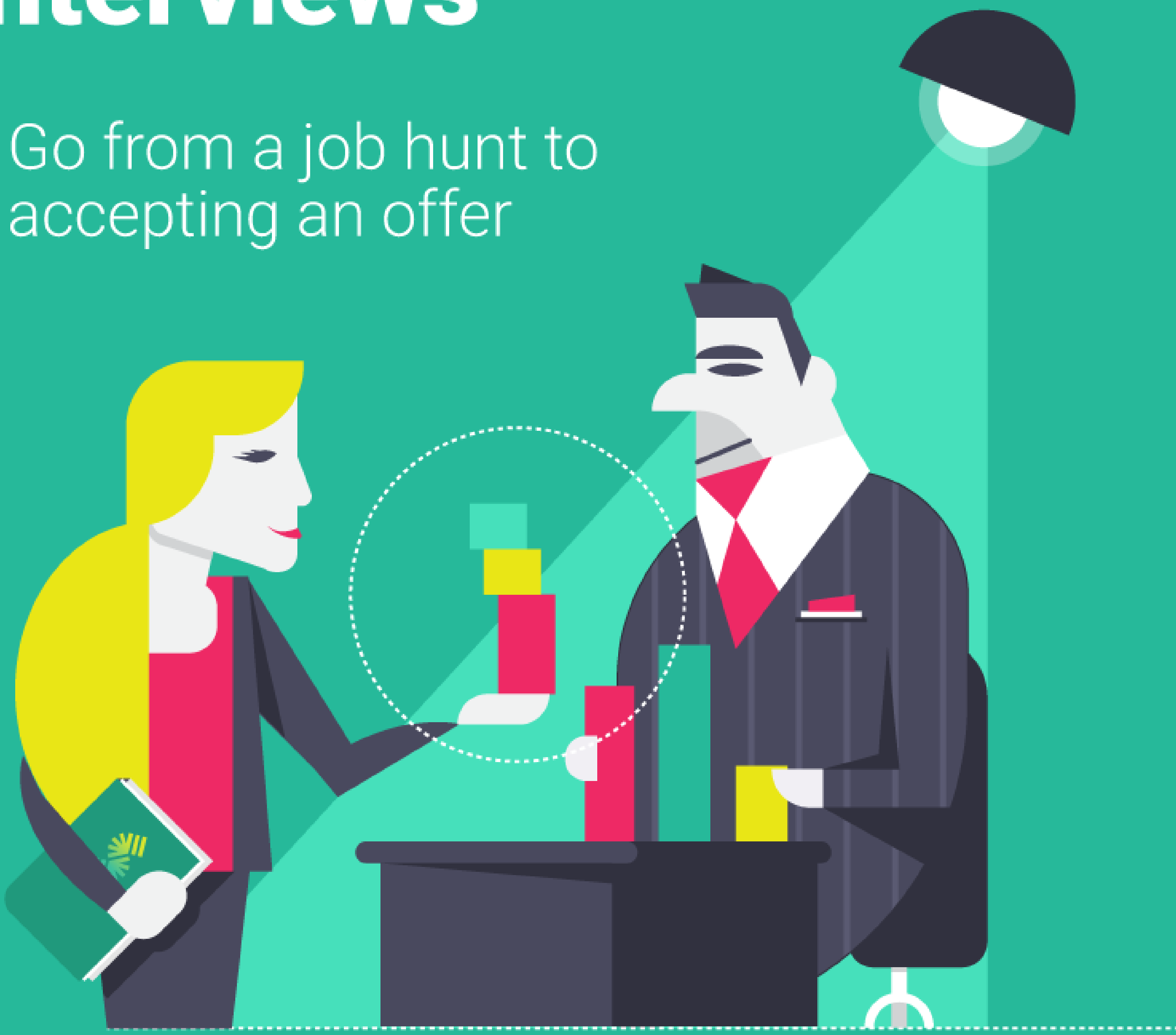
Ultimate Guide to Data Science Interviews

Go from a job hunt to accepting an offer



Springboard

Table of Contents

Introduction

When we first wrote the Springboard Careers Guide to Data Science, we didn't expect the engagement it'd garner. Thousands of people signed up in a few days, confirming our belief that there was a scarcity of great advice on what is an exciting but nebulous field.

In speaking with more and more people, we found only a few great resources that explained how to break into a data science career. There were individual stories and collections of interview questions, but we couldn't find a full guide to cover everything about the data science interview process--from how to get an interview in the first place to how to deal with any offered positions.

I wanted a guide collecting perspectives from people on both sides of the table. I wanted to talk to recruiters who refer candidates, hiring managers who table offers, and candidates who had successfully made it through the data science interview to demystify the data science interview process with insights from people who had previously gone through the process. **I co-authored this book with Sri Kanajan a senior data scientist in New York City at a major investment bank.**

At Springboard, we've taught thousands of data science aspirants through our mentored workshops. We built large, engaged communities of mentors and alumni, which afford us a unique vantage point to deliver real-life perspectives on the data science interview process.

It was difficult collecting everything here, like it was difficult for many of the candidates who made it through the process. Some of the leaders in data science, including the Chief Data Scientist of the United States, had to go through six months of waiting before they got an offer! Most companies' data science interview processes are designed to weed out all but the most determined and skilled candidates. It can seem, at times, like a hurdle preventing any sane job-seeker from entering. Yet, while the investment can seem immense, the return can be even greater.

Data science has been called the sexiest job of the 21st century. Data scientists don't just make good money; they drive significant social impact -- from mapping world poverty to stopping pandemics before they even happen. Data scientists unearthed the identity of Banksy, and they mastered the art of predicting basketball scores in March Madness. Working in data science isn't about just having a good salary and good work-life balance; it's about solving **big problems that matter**.

We wrote this guide because we wanted to you to go from being curious about data science to actively trying to get a job in the field. We wanted to unearth what it takes for you to make it through the data science interview process. We wrote this guide because **we want you to rock your data science interview.**

What is Data Science?

Before you look for data science interviews, you should know what the term means and what you're getting yourself into.

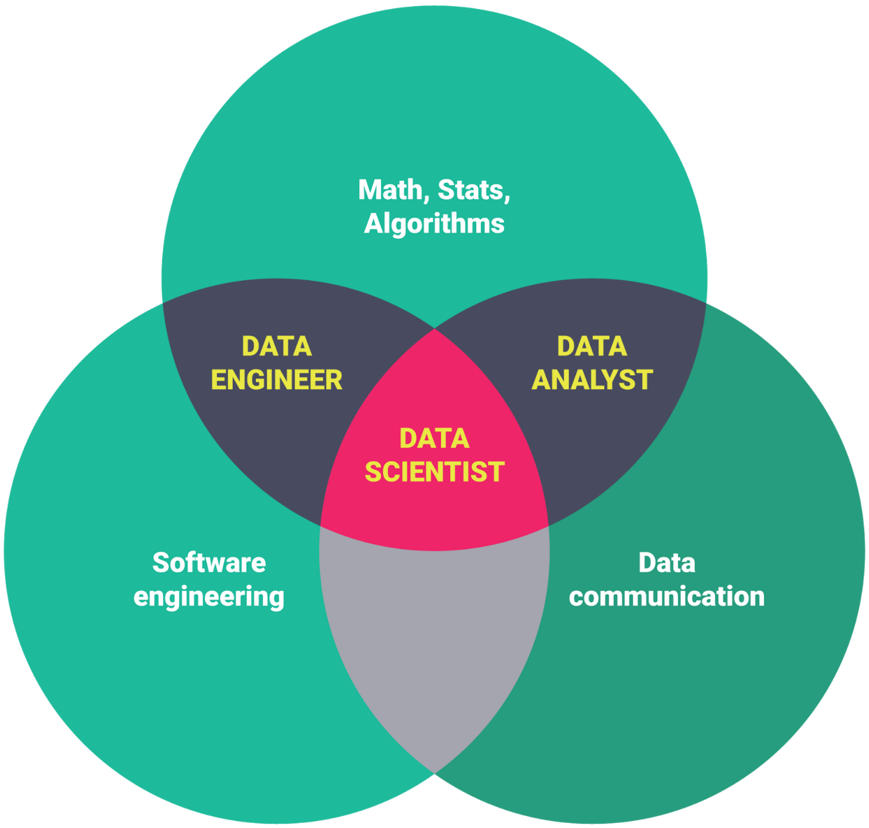
DJ Patil, the current Chief Data Scientist of the United States, first coined the term data science.

A decade after it was first used, the term remains contested. There is a lot of debate among practitioners and academics about what data science means, and whether or not it's different from the data analytics companies have always used. When people talk about big data and using machine learning to solve data problems, they are venturing into a whole new field whose terms are being defined right now.

Different companies have differing definitions of what data science means. Individual hiring managers may differ about exactly what they're looking for; they will hire and interview accordingly.

This confusion makes the data science interview process difficult for a lot of candidates. Data science can have vastly different definitions depending on what role you're applying for and the company you're interviewing with.

Different Roles within Data Science



Let's go through a sample data science project to elaborate on the different roles you'll see in data science. A data science team might be assigned to use deep learning to classify images like Yelp's team did.

Millions of photos are uploaded on Yelp every single day, but it can be hard to get images you want for each restaurant. Sometimes, the photos uploaded are all of the same category–maybe they're all photos of the food or the outside of the restaurant. A holistic evaluation of a restaurant requires images of different kinds.

You can use machine learning to automatically categorize which images fall into what category. Computers can, with the help of a training set, tell you whether or not an image is the outside of the restaurant or of food.

Data scientists create the model to help machines create those distinctions. They would be able to think through the types of data they need, from manually tagged photos to keywords in image captions. This tends to be a more senior-level role, as they often manage data products from end-to-end and deal with all facets of data science problems, from algorithm selection to engineering design.

Data engineers create systems to source all of the image data and store it, as well as implement some of the algorithms determined by data scientists at scale. This tends to be a role for people with strong technical chops but might not know as much about the theory of the algorithms they're implementing at scale.

Data analysts query and present the business implications of the change. Did it please users? How much more traffic did Yelp generate due to the recent change? These are questions data analysts would ask. Then, they communicate the insights they found. This role tends to be filled by more entry-level people and people in business-facing roles learning to apply their insights on a technical basis.

There are more roles we'll cover in detail later. For now, you should know that the data science interview process for all three of these general roles can be vastly different from one another--and in fact, they often are!

How Different Companies Think About Data Science

Not only are there different roles in data science, there are also different companies with vastly different interview processes!

In general, these roles can be split into four rough categories.

1- Early-stage startups (200 employees or fewer) looking to build a data product

Welcome to the beating heartland of Silicon Valley. The early-stage startup is a romantic notion, but one seeing a staggering amount of success in a rapid amount of time. If you join an early-stage startup, be prepared to wear a lot of hats and potentially take on all three data science roles at the same time. You will never have the resources you need in full, so be prepared to be scrappy and tough.

The bar will be especially high if the startup in question deals with data as its product. A platform optimizing other people's data or applies machine learning to different datasets will have much higher standards for how they think about data than companies trying to learn from their own data. The co-founders will likely be pioneers in the field of data science or have led large-scale data science teams. They will be looking for A-players who have significant experience in the field or tons of potential and drive. If you join an organization like this, be prepared for the learning experience of a lifetime, and be prepared to be held to the highest standard possible when it comes to data science.

Examples of this company type: Looker, Mode Analytics, RJMetrics

Sample job postings: Data Analyst (Looker), Senior Analyst (Mode Analytics)

MODE

Apply Now

Senior Analyst

at Mode (View all jobs)

San Francisco

Mode is a company built for analysts. In addition to building a product to help analysts get their job done, Mode aims to be a source of education and inspiration for analysts of all experience levels.

Mode's own analytics team is a critical part of this mission. The team has a dual mandate. First, Mode's analytics team serves internal customers on our product, marketing, and sales teams to help guide decision-making. Second, the team empowers and inspires other companies and analysts to be data-driven by "open-sourcing" the resources and analyses we create internally.

What you'll do

- Work with others at Mode to help them make data-driven decisions about product development, marketing, and sales
- Define and track key metrics
- Find opportunities where Mode could benefit from additional data insights and look into them
- Share your work with the Mode community

What we look for

- A great communicator—you don't just focus on producing great charts, you can also explain what they mean and their implications
- A creative analytical thinker—you think about the right questions and what you need to measure in addition to producing results
- Strong technical ability—we're believers in SQL but scripting languages like R and Python, visualization tools like D3 and experience developing data tools and pipelines are great bonuses as well
- A professional background in data analysis

Size of the company: 143 associated on LinkedIn (11-50 company size)

How to read this job description: Focus on communication and scripting languages for querying and visualizing data indicates this is a business-facing role where insights must be communicated to relevant teams.

2- Early-stage startups (200 employees or fewer) looking to take advantage of their data

The bar will be lower if a startup is merely looking to take advantage of its data rather than selling a data product to other companies, but since the smart use of data is essential to the competitive advantage of a startup, you should still expect a relatively high bar.

Startups in the tech industry contain a lot of technical talent, but they need somebody to bridge the business and tech teams, especially if there are communication issues between the different teams on how data is used. Be prepared to work hard for the company to embrace being data-driven at all levels, and be prepared to be the one who brings in new tools and processes for collecting and using data at all levels of the organization.

Working for a company that deals with its own data but doesn't think about data at scale may be an unique challenge as you'll be called upon to enforce and spread a data-driven culture throughout the organization. Be prepared to exercise your leadership and communication skills.

Lastly, B2B startups and B2C startups differentiate in the data they get. B2B startups are business-to-business; they sell software directly to large companies. Think Salesforce. B2C startups cater to many individual customers. Think Amazon. When you're dealing with B2B startups, you're likely going to be faced with data challenges that are small in volume but high in detail and features; startups that sell directly to businesses don't have many customers, but they focus maniacally on the ones they do have since each individual customer will bring in lots of revenue. B2C startups will have more data problems dealing with volume and scale as they will have many more customers, but the focus on individual customers will be diluted to focus on groups of them. A B2B startup may deal with 1,000 customers, all of whom pay $1,000 a month. A B2C startup may deal with 100,000 users, but each user may only generate $1 in revenue a month!

Be familiar with the company you're applying for and the unique data challenges it faces. Research thoroughly, and make sure you're only applying for companies that fit your passions and skills.

Examples of this company type: Springboard, Branch, Rocksbox, Masterclass, Sprig

Sample job postings: Lead Data Scientist at Branch, Data Scientist (Research) at Rocksbox, , Data Scientist at Masterclass

Data Scientist (Decision Scientist)

MasterClass

Job Description

Who we are:

MasterClass is transforming online education by enabling anyone in the world to learn from the very best. We are deconstructing what makes an actor able to cry on demand, how an athlete defies gravity, and what it takes to write a bestseller. Our online learning content is available to students anywhere anytime. We are democratizing access to genius, one class at a time.

We are a fast growing VC-funded startup based in San Francisco and have created online classes taught by famous masters of their craft -- Serena Williams, Dustin Hoffman, Kevin Spacey, James Patterson, Annie Leibovitz, Usher, Christina Aguilera, and many more to come.

Since launching in 2015, we have been growing our team. Apply now to find out more about what we are doing behind the scenes.

What we are looking for:

- Enthusiasm for working on projects across a broad range of analytic disciplines – from statistical analysis and predictive modeling, to user surveys and quantitative market research, to business intelligence and analytics.
- A pragmatist, who is deliverable-oriented and a self-starter. You like moving fast and are not afraid to get in the weeds, but keep overarching objective and priorities in mind, which enables you to deliver a 'good enough' solution in a short timeframe when required.
- Scientific mindset and desire to go beneath the surface and distill a problem into a clear set of hypotheses that can be tested.
- Strong interpersonal and communication skills, including the ability to describe the logic and implications of a complex model to all types of business partners.
- Comfortable with learning new tools or skills to remove bottlenecks and keep a project moving.
- Good business acumen and product sense, and a penchant for systems thinking.

Size of the company: 37 associated on LinkedIn (11-50 company size)

How to read this job description: Looking for a generalist who can dive deeper and still communicate different insights indicates this is a data scientist role that will be very broad in terms of skillsets demanded. This role is going to be proactive and entrepreneurial.

3- Mid-size and large Fortune 500 companies who are looking to take advantage of their data

The largest companies in the world know that taking advantage of their data is a top priority. Some will have established data science teams that are well-funded, robust, and fed with lots of data. Some will have startup-like teams within the organization to help them translate their data into business insights. There are a lot of companies hiring data science teams upon realizing how important data is to remaining competitive. Use this to your advantage; it can be easier passing the data science interview for a large, prestigious brand.

While a lot of these companies will have established corporate cultures and bureaucracies that make it harder to innovate, they will also have data on millions of people. Imagine processing logistics data for Walmart--you will have millions of data points, and your insights will make a difference in the lives of millions of people.

While these companies are not traditionally seen as the ones building cutting-edge data science solutions, there is still a lot of good work available for those who want to work on challenging datasets with talented teammates.

Examples of this company type: Walmart, JPMorgan, Morgan Stanley, Coca Cola, Capital One

Sample job postings: Data Scientist, Modeler at Morgan Stanley, Data Engineer at Capital One

Data Engineer

Apply now

Job ID: R5046
Date posted: 06/16/2016
Locations:
McLean, Virginia

1750 Tysons (12023), United States of America, McLean, Virginia
At Capital One, we're building a leading information-based technology company. Still founder-led by Chairman and Chief Executive Officer Richard Fairbank, Capital One is on a mission to help our customers succeed by bringing ingenuity, simplicity, and humanity to banking. We measure our efforts by the success our customers enjoy and the advocacy they exhibit. We are succeeding because they are succeeding.

Guided by our shared values, we thrive in an environment where collaboration and openness are valued. We believe that innovation is powered by perspective and that teamwork and respect for each other lead to superior results. We elevate each other and obsess about doing the right thing. Our associates serve with humility and a deep respect for their responsibility in helping our customers achieve their goals and realize their dreams. Together, we are on a quest to change banking for good.

Data Engineer

Do you think about using data to unleash the power of applications? Do you build applications you are proud of and want to tell your friends about? Do you appreciate elegant solutions but also know when to build simple first and iterate? Capital One is seeking a Sr. Data Engineer to build extremely elegant and scalable data solutions to deliver a game changing user experiences insights across both internal and external customer touch points.

As part of the Consumer Banking team that's leading the next wave of disruption on a whole new scale, you will play an integral part in advancing Capital One's ecosystem and culture of technical excellence. Your areas of responsibility will range from message queues such as Kafka, Big Data solutions such as Hadoop, Dynamo, Redshift, Cassandra, Mongo, APIs, MicroServices, Distributed Processing and beyond.

Size of the company: ~30,000 associated on LinkedIn (10,000+ company size)
How to read this job description: Focus on Big Data tools indicates that this is going to be a fairly specialized role that looks into handling the immense amounts of data Capital One is holding.

4- Large technology companies with well-established data teams

Large technology companies are a breed in and of themselves. They're the continuation of the startup obsession with data, except now they have scaled to a point dealing with millions of data points or more. Think of the Ubers, the Airbnbs, the Facebooks, and the Googles of the world. With large technical teams led by some of the most brilliant minds in the industry, data science roles here are heavily specialized, and you'll work on cutting-edge problems with data that requires ferociously innovative thinking.

Come here if you crave a challenge and if you want to learn a lot with a lot of data points. The upside isn't as good as the earlier stage startups, but you'll get good perks, good salary, and great teammates--and a great CV job description in case you ever want to move on.

Examples of this company type: Facebook, Google, Airbnb
Sample job postings: Data Scientist, Oculus, Data Scientist Airbnb - Machine Learning

Data & Analytics

Data Scientist, Analytics (Oculus)

(Menlo Park, CA)

Facebook was built to help people connect and share, and over the last decade our tools have played a critical part in changing how people around the world communicate with one another. With over a billion people using the service and more than fifty offices around the globe, a career at Facebook offers countless ways to make an impact in a fast growing organization.

We're looking for data scientists to work on our core and business products at Oculus with a passion for virtual reality to help drive informed business decisions. You will enjoy working with one of the richest data sets in the world, cutting edge technology, and the ability to see your insights turned into real products on a regular basis. The perfect candidate will have a background in a quantitative or technical field, will have experience working with large data sets, and will have some experience in data-driven decision making. You are scrappy, focused on results, a self-starter, and have demonstrated success in using analytics to drive the understanding, growth, and success of a product.

Size of the company: ~16,715 associated on LinkedIn (10,000+ company size)
How to read this job description: Focus on multi faceted, innovative skillset shows this is going to be an open-ended data science role that will be expected to think of new projects and lead them from end-to-end.

Industries that employ Data Scientists

Data science also varies depending on the industry. Industries have certain areas of knowledge specific to the industry itself, and they involve different types of data. A school will be focused on different metrics than a bank.

If you happen to have a passion for a certain industry, make sure it comes off with keywords on your CV and LinkedIn. Demonstrating why you love a certain industry and deep knowledge of the industry itself positively differentiates you as a candidate.

The three largest hiring industries for data science in O'Reilly's survey of the field are software companies, consulting companies, and banking/finance companies. Those three industries also tend to pay the most for data science professionals.

Different industries also vary in the types of roles they hire for. Software, medicine and telecommunications companies tend to be the largest hirers of data scientists. Software, aerospace, and information technology companies hire more data engineers. Lastly, data analysts tend to be hired by healthcare companies and consulting/banking organizations.

Be aware of the industry your potential employer is in, and infer what their data science needs are.

You have to be aware of the different roles, companies, and industries within data science to understand exactly how your data science interview process will go.

To do data science, you must be able to find and process large datasets. You'll often need to understand and use programming, math, and technical communication skills. You'll also need to tailor your skillset and how you present yourself to the different roles and hiring companies within the world of data science.

Most importantly, you need to have a **sense of determination to understand the world through data and not be deterred easily by obstacles.**

The data science interview process is designed to test for those skills and resilience. Be prepared to be challenged on every dimension.

Getting a Data Science Interview

The first step in the data science interview process isn't dealing with the interview; it's finding it in the first place, a process that in and of itself can take months of effort!

We surveyed about twenty people about the hardest parts of the data science interview process as part of the research for this book. The answer we got back had little to do with the technical questions we thought were the hardest. While technical questions ranked second with *68% of respondents* selecting it as one of the hardest parts of the interview process, a whooping *80% of respondents* selected getting a data science interview!

Literature was scarce out there about how to get an interview, especially for people transitioning from different careers. We dived deeper, and looked through real-life case studies in addition to different resources we've curated for you.

Nine Paths to a Data Science Interview

We found traditional paths to job interviews that could work to a certain degree in data science. We also found new, proactive approaches, especially with emerging startups, where non-traditional tactics could get candidates to the forefront of the hiring race.

Traditional Paths to Job Interviews

If it ain't broke, don't fix it. While a lot of the new, proactive tactics we discuss can have a lot more efficacy, it's always good to know the basics.

1- Data Science Job Boards and Standard Job Applications

You can submit your resumes and cover letters to company careers sites. Then, you can wait and hope. We're not saying to avoid this route, but it shouldn't be the one you rely on.

Use Indeed and Careerbuilder to search for different data science postings. Then, there are specific job boards for the data science space, such as the Kaggle Jobs Board.

2- Work with a Recruiter

You can contact recruiters who can help put you in touch with the right employers. There are recruiters who specialize in data science and technology spaces. They are gatekeepers to jobs never listed in public outlets. A quick search on LinkedIn for data science recruiters near you will help you find the most relevant matches.

3- Go to Job Fairs

Job fairs in data science are far and few in between, though Harvard and Stanford do host computer science job fairs that have plenty of data science jobs for their students. You're better off attending either events or meetups with the local data science community rather than looking around for your traditional job fair.

Proactive Paths to Job Interviews

We've covered the traditional paths to job interviews, the options that have been the default of job-seeking. These days, getting an offer sometimes requires hustle and grit outside of tried-and-true tactics. Startups provide a large number of new data science jobs. Their culture and hiring tactics trickled up to large companies that a decade ago were just startups as well. The result is a new hiring environment where oftentimes, one has to be proactive to reach decision-makers who have known nothing but grit when they built their own companies.

4- Attend or Organize a Data Science Event

You need to find people interested in the data science community to find hidden opportunities and become proactive at integrating into the community. There are several events where you can do this, from larger conferences to smaller community meetups.

Conferences

Strata Conference

The Strata Conference is a big data science conference that takes place worldwide in different cities. Speakers come from academia and private industry; the themes orient around cutting-edge data science trends in action. The conference allows you to learn the technology behind data science, and there are plenty of networking events.

KDD (Knowledge Discovery in Data Science)

KDD or Knowledge Discovery in Data Science is another large data science conference. It's also an organization that seeks to lead discussion and teaching of the science behind data science. Membership and attendance at these conferences offers a marvelous way to contribute to growing trends in data science.

NIPS (Neural Information Processing Systems)

NIPS, or Neural Information Processing Systems, is a largely academic data science conference focused on evaluating cutting-edge science papers in the field. Attending will give you a sneak preview of what will shape data science in the future.

Meetups

We've listed the major conferences where the data science community assembles, but there are often smaller meetups that serve to connect the local data science community.

The San Francisco Bay Area tends to have the most data meetups, though every major city in America usually has one. You can look up data science meetups near you with Meetup.com. Some of the largest data science meetups, with more than 4,000 members, are SF Data Mining, Data Science DC, Data Science London, and the Bay Area R User Group.

You'll want to join the events, or create a meetup yourself if you cannot find a nearby event. Our director of data science education, Raj, got a job by becoming known as a data science connector. He hosted a local meetup in Atlanta and invited distinguished speakers in data science. Soon, he was known as a data science influencer, and as soon as there were open data science positions, he was tapped to apply.

5- Freelance and Build a Portfolio

Sundeep Pattem is a data innovation leader at the California Department of Justice. He's also mentored for several data science courses, and as a data scientist, he works on creating end-to-end solutions that extract value from data. He has personal websites with different data science projects.

His breakthrough into data science came when he found an unsolved problem in energy sustainability and worked to solve it. He was soon a published author at a prestigious academic conference, and shortly thereafter, he was hired to become a practicing data scientist.

If you're unsure of what data you want to analyze, we have a list of 19 free, open source public datasets you can explore.

If you freelance around data problems you love and build incredible solutions, keep a record of everything you do in an accessible portfolio that tells stories around your passions.

6- Get Involved in Open Source and Open Data

The most interesting projects in the world don't necessarily reside in secretive company databases anymore. They are often in open source repositories on Github. This includes the Natural Language Toolkit project, which helps deal with human language as a data source and the various libraries that make up the Python data science and machine learning toolkit. The R community also hosts many of its packages on a consolidated public website.

Many leading CTOs will hire based on your contribution to open source projects, and may even find you through that route. It's easy to tell if somebody is able to work in a team and build marvelous things through the transparent glass of open source. Make sure you take advantage.

7- Participate in Data Science Competitions

If the broad confines of open source projects aren't your type of projects and your creativity thrives best in more confined situations, consider joining a data science competition.

Data science competition platforms like Kaggle, Datakind and Datadriven allow you to work with real corporate or social problems. By using your data science skills, you can show your ability to make a difference and create the strongest interview asset of all: a demonstrated bias to action.

One of our Springboard mentors, Sinan Ozdemir, competed his way to a data science job based on his work on problems on Kaggle. You can do the same.

8- Ask People for Coffees, Do Informational Interviews

At the end of the day, your network will get you the best chance at a new job. You should seek to know more people in the field you want to work in, if only to get an idea of the problems they have and which you can solve.

Legendary entrepreneur and strategist Steve Blank has a great framework for getting coffees with people too busy to see you, as most data scientists will be. You have to find a way to provide value of some kind and look to give them a fresh perspective on the problems they face.

This can culminate in an informational interview where you seek advice and information from data scientists in the field. If you do this right, you'll constantly grow your network of data science opportunities, and you'll understand more about how data science works in industry.

9- Data Hackathons

In line with the trend of seeing work in action, data hackathons offer you an unique opportunity to learn data skills with a motivated team. You will have to solve a data problem in a couple of days.

An example of this kind of hackathon is the DataWeek hackathon in San Francisco. By teaming up with others to deliver real solutions, you'll differentiate yourself from other job candidates. Many employers lie in wait at hackathons as well, some companies going as far as to sponsor hackathon prizes in the hopes of finding their next data scientist!

Working with Recruiters

For this section we worked with **Andy Musick**, **an Atlanta-based recruiter:** contact him at **andy.musick@hotmail.com** if you were looking for an Atlanta-area job. We also worked with **Anna Meyer,** a data science recruiter at **Robert Walters**, a recruitment agency specialized in data science. Feel free to contact her at anna.meyer@robertwalters.com.

How to Apply

CV vs LinkedIn

A lot of people out there may have a traditional view on what makes for a good job application. They're already missing a larger point: the traditional view is out.

There is a **fundamental difference between academia and working in an industry,** and it starts in how you present yourself.

We talked with recruiters, students, and hiring managers, and they all agreed that LinkedIn was the golden standard of recruitment. Having a well-optimized LinkedIn profile allows employers to size you up and recruiters to find you the right opportunity.

If you're not making sure you shine on LinkedIn, you're already losing out to candidates who are.

While a resume may be required to go through the process, it isn't the main draw that will get you in the door anymore. Recruiters will only look through your resume once it's presented in front of them, while a great LinkedIn could lead to inbound work opportunities on a constant basis.

Unlike in academia, where an impressive array of papers and academic work will win over everything else, applying for industry jobs involves being as succinct as possible and listing the impact you drive with your accomplishments. resumes are not so much read as scanned. Keep that in mind if you're going to build one. A recruiter spends an average of thirty seconds on a resume.

Key Advice on Resumes

1) Keep them short, preferably under a page. Remember that people are scanning your resume for signs of interest before they ever consider doing a deep dive.
2) Make sure your skills stand out and are highlighted (consider bolding relevant skills). Recruiters and hiring managers will look to see if you're a technical fit before looking further.
3) Have clear job headings, and at most, three one-line points in each one of your job descriptions. You want to clearly mark how your experience ties in with the job requirements you've applied for.
4) Demonstrate your impact with numbers! Don't say that you "did X." Tell the hiring manager what effects X had. You want to say you discovered something that helped thousands of people save hours of time--not that you simply discovered something. Write "created an automated sales email software that generated $400,000" not "created an automated sales email software."

Key Advice on LinkedIn

1) Don't be shy. Fill out as many details as you can; it makes a difference. Most hiring managers will want to see your LinkedIn before they ever interview you.
2) Make sure your job titles are clear and consistent with search terms that recruiters would use. Saying that you worked as a data scientist or as a data analyst is preferred to coming up with your own job title.
3) One way you can differentiate from others is adding some personal flavor to your profile. Add some of your interests and passions, and make sure they are evident in your LinkedIn. Hiring managers like evaluating candidates for technical skills and cultural fit. Being able to show that you have your own unique take on the world will only add value to your job search and help you stand out.

4) While you might not want to tailor your profile for certain jobs or industries, make sure you know what you're looking for, and make sure that comes out on your LinkedIn. You want to be very deliberate at constructing your profile so that it gets you the position you want. Avoid listing data entry if you don't want to get entry-level offers. Mention specific industries if your heart is set on working on a particular type of problem.

Make sure you know what roles you're applying for, and apply industry keywords and skill keywords that match. Interested in a data science job in finance? Don't hesitate to put industry terminology all over your CV and LinkedIn. If you have a skill that you researched is in demand for the role you're looking for, add it liberally! You can research what technologies a company uses; companies like Yelp and AirBnB will often blog about their data projects. If you see that the role in question demands Python and R skills, make sure that your CV and LinkedIn marks those skills. Endorsements also play a positive role in this regard when it comes to LinkedIn, so don't be shy at asking people who have worked with you to endorse your skills and give recommendations.

More recruiters and hiring managers look through LinkedIn than resumes today. A recruiter will look at a CV for an average of 30 seconds before discarding it. Make sure the impact you've driven is fleshed out with strong action verbs, you've formatted your resume and LinkedIn to stand out, and you've filled them with the right keywords.

Keep in mind that this is a first step and that applying with just a CV or LinkedIn will get you considered at most places, but not with any particular enthusiasm. You'll have joined the queue of thousands of others applying the same way, and you'll probably need to do more to get your dream job. Regardless, make sure you optimize every step of your application, including the CV or LinkedIn that employers will inevitably look over.

Cover Letter vs Email

A cover letter was always the standard for academic advancement. Nowadays, the recruiters we talked to confirmed that companies seldom read them. If you want to differentiate who you are, you'll have to do it on your CV or your LinkedIn.

If you're going to be proactive, send a brief summary of what you've done in an email to a hiring manager. This serves as more of an explainer they can share with other people in the company. You'll want to keep it brief--no more than a few paragraphs at best--and you'll want to keep this email focused on the top three points that define the impact you've driven.

How to get References and Your Network to Work for You

Most people don't realize how critical it is to build and maintain your network to get your feet in the door with the data science interview process. The strongest signal hiring companies look for is strong referrals, especially from internal sources. If you have somebody advocating for you inside the organization you're applying for, that can ensure that your CV will be looked over, and it can even get you to skip steps in the interview process!

We informally surveyed some of our alums going through the hiring process. It turns out that a referral from an insider within the company led to a **85% chance** of getting an interview with that particular application, while those who reached out cold and only applied with their CV or LinkedIn or through the standard format only had around a **10% chance** of getting an interview. Pursuing the former can improve your job-hunting process by an order of magnitude. Our alums also said that the referral doesn't even have to come from a friend, *the fact that an application has been referred by an existing employee often guarantees at least a phone interview.*

Take a **long-term view** on this by adding value to different people in your network, whether that's being generous with advice once that's asked of you or being generous with introductions to other people in your network. Hopefully, by the time you're looking for a job, you'll have built up a strong network of people also interested in data science that can make the right introductions and give you the right referrals.

If that isn't the case, and you're looking to get those referrals right now, you can use what is called the informational interview technique. This entails reaching out to people who are working in the field to get a sense of what's going on and what their problems are. People, even complete strangers, can be very generous with their time if you show that you're genuinely interested in what they're doing and you offer to help as well.

Look for people at meetups, or specifically target people on networks such as LinkedIn, Angellist and FounderDating. Present your intentions honestly, but indicate that you're very interested in the company and data science in general. Ask for a coffee where you can add perspective to a problem they're solving or learn about their company.

A sample script might go as follows (where you can add somebody on LinkedIn as a friend or message them directly on FounderDating or Angellist):

Hi [name],

I was super interested in the problems Airbnb is facing in data science. I've been aspiring to break into the field, and being a passionate follower of the Airbnb Nerds blog, I noticed that building trust with data is an important part of what drives Airbnb. Based on my background in psychology and statistics, I might be able to help come up with some creative ideas on how to foster trust.

I'd love to take you out to coffee and get a greater sense of what problems Airbnb has--perhaps I can help! Would you have some time in the coming weeks?

Cheers,

[your name]
Links to your LinkedIn, resume, portfolio and/or a recent project

If you reach out to enough people and seek introductions to people through your network, you'll be able to find people in any company to talk with. Check out your second connections on LinkedIn and how they are connected to you, which you can easily do through any LinkedIn company page. Here's an example of a company page for Airbnb.

Once you're set for an informational interview, make sure you've researched the company and the person you've talked with by looking at the company website and any other resources you find.

You should have a pretty good sense of what problems the company encounters on a day-to-day basis.

These informational interviews are a great chance to know exactly what is happening at a company and what their priorities are, which is greatly beneficial knowledge in an actual job interview. If you come in well-prepared and position yourself as someone who can help the company, the person you're having coffee with could become a strong internal advocate and help you jump through the usual recruiting hoops to get your first round interview.

Preparing for the Interview

Hopefully all the work you put into getting the data science interview pays off, and you get the email that signifies the start of the interview process for you: a company representative beckoning for an initial phone call. Here's what will happen and how you should prepare.

What to Expect

The data science interview is a complex beast, with behavioral questions mixed with a bunch of technical questions. You've gotten pretty far if you're able to get an interview in the first place, but you still have further to go.

Let's start from the beginning--a data science interview will be vastly different depending on the position you're applying for and the hiring organization. Certain organizations will be very rigorous and make you go through several technical challenges. Others will look more at culture fit and, especially if you have strong references, get you straight through to the final round.

The most rigorous process possible looks something like this:

1- The Phone Screen

This will typically be done by somebody in HR and acts as a filter to save hiring managers time. Sometimes, there will be basic technical questions to screen out candidates who are grossly unqualified, but most of the time, this phone screen involves establishing the beginnings of culture fit and making sure that the candidate has good enough communication skills to come off well in the interview.

In this call, you'll want to get a sense of what problems the data team is facing and the organizational structure of the team you're applying to. Come prepared with thoughtful questions that demonstrate a deep understanding of the business and the space they operate in, and be prepared to ask them at the end.

2- Take-home Assignment

After the phone screen, companies often send a prepared assignment for candidates, with some time pressure being applied. This is a good way to screen out candidates who may be technically weak, or who may not be committed enough to invest a lot of time in the recruitment process. Some companies dispense of this altogether, but those that do embrace the take-home assignment often use it as a testing bar to save their hiring managers time.

An example of a take-home assignment is doing a deep analysis on a specific dataset provided for you. When the assignment is designed well, the assignment is also an opportunity to learn more about the types of problems you would work on if you were to get the job. Here, you'd be expected to storytell around insights you'd find in the data. Another example would be having a dataset with significant errors in it that you'd be expected to clean. A final example would involve working with a specific problem relevant to the business, such as building a job recommendation system for applicants based on data from job descriptions.

Only those that pass the bar of having good assignments will talk to a hiring manager face-to-face. You'll get weeded out quickly if you refuse to do it all together.

Take the time to do the assignment, and try to see how it relates to what problems the company is undergoing. Using the assignment as a way to see what kind of skills you'll be tested on and how the company in question is thinking about your role ensures that you maximize your time. This is where you can shine in a hiring process and show how you are different from other candidates.

3- Phone Call with a Hiring Manager

You may receive another phone call screen that will be focused on either mathematics and statistics questions or coding questions. This will be done by a hiring manager or a technical person. This will likely be the final evaluation before a company invites you to an on-site interview. The phone call will typically be split into three components. Sometimes, this is done in one long call; other times, it is done in three short phone calls of about thirty minutes each.

Mathematical/Statistical Phone Call

You'll be evaluated on core mathematical and statistical concepts here, which will depend somewhat on what role and what company you're applying for. Web companies will tend to focus on your knowledge of A/B split testing, your understanding of how p-values are calculated, and what statistical significance means. Energy companies might test you more heavily on regression and linear algebra. No matter what type of interviewer you're talking with, you'll want to sketch out the entire thought process behind your problem solving.

If you're asked about A/B split tests, describe the A/B split test process in detail, fleshing out what pitfalls to watch out for and leaning on any experience you might have in the field. Treat the question like a mathematical proof and a test of your ability to statistically reason, but don't hesitate to turn your finer eyes to detail and a coherent story about why this matters to the company at hand.

Coding Phone Call

This part of the interview process is fairly typical and is also the closest to other technical interviews. You'll be evaluated on your ability, over the phone, to solve coding challenges by presenting either pseudo code, or in harder interviews, compile ready code. If you're applying for a data analyst position, this will swing more to asking you how you'd think about querying data with SQL. Otherwise, you'll be asked questions in the programming and scripting languages you've claimed experience in, from Java to Python.

Your interviewer may also use tools like [HackerRank] or [Collabedit] to evaluate you live online. In this case, your hiring manager will watch you as you type out your solution: be ready for approaches such as this, and train with those tools if you can!

There are plenty of great resources out there for coding interviews, from Cracking the Coding Interview to InterviewCake. Use them to your advantage.

Practice makes perfect here. Make sure you have a comfortable space and natural environment for you to code. Be prepared to jot down code on a paper and explain it on a phone call, or be prepared to type in the code on a laptop.

You will often be asked about data structures more than anything else. Know hashmaps, trees, stacks, and queues very well. Prepare for this phone call like how software engineers would prepare for a coding interview, and you'll pass with flying colors.

Call with the Hiring Manager

Finally, you'll be patched through to the hiring manager, who is now evaluating you on how well you communicate and if you'd fit well on the team. This may be on a separate call from the technical phone screens, or it can be the last part of a mega-call that encompasses all three. In this call, the hiring manager is trying to get a feel for your character, your motivation, your fit with their team, and your raw intelligence. Most hiring managers have a mental model for who they are looking for. The closer you fit to it, the more likely you will pass to onsite interviews.

This is where your work with the recruiter beforehand will shine. The more you know about the problems the hiring manager is facing and the kind of person they're looking for, the better you'll be prepared to present yourself as the perfect fit. Tailor your communications to that goal, and be confident and clear, and you'll make it to the next round. Try to pass the "airplane" test as well; imagine the hiring manager evaluating whether they'd like to spend hours of time with you. The workplace will force you to work together closely and spend a lot of time together. Make sure you show that you can get along with your manager!

4- On-site Interview with a Hiring Manager

Finally, if you've made it through the earlier calls, you'll meet your hiring manager face-to-face. They'll be evaluating you from both a technical and non-technical perspective. They're looking to ascertain if you're a fit, and they may test you on your technical chops by having you whiteboard different scenarios.

5- Technical Challenge

If this doesn't happen to you during the on-site interview, prepare to be challenged on your technical skills in one form or another, especially for roles that lean more towards data engineering. You'll often find that this is similar to a [software engineering interview]() where you will be asked to whiteboard and write down how you'd implement certain algorithms or solve certain problems.

Here is where strong knowledge of software engineering concepts such as time complexity/Big O notation and a strong grasp of the mathematics and statistics behind data algorithms can truly shine.

6- Interview with an Executive

If you pass the bar for your hiring manager, you'll often do a final interview with a senior executive. In a startup, this will often be the co-founder or the CEO themselves.

If you've made it this far, congratulations! Don't take it for granted, but this is a sign that a company is leaning to an offer for you. Normally, only candidates who have passed the technical bar will get here, so now you need to emphasize exactly how you can drive impact with your knowledge of the business itself, and the specific problems it faces. At this point, you're not looking to prove yourself so much as to avoid glaring errors.

What a data scientist is being evaluated on

Position Title	Mathematics/Statistics (e.g. P-value analysis, AB testing)	Database Querying (SQL)	Algorithms (e.g. Supervised learning, Entity Resolution)	Software Engineering (e.g. Python, Java, Object Oriented)	Big Data/Systems Engineering[1] (e.g. Spark, HBase, Hadoop)	Soft Skills/Domain Expertise (E.g. public speaking, presentation skills)
Product Data Scientists[2]	Medium	Medium	Medium	High	High	Medium
Data Engineering	Low	Medium	Low	High	High	Low
Data Scientist	High	Medium	High	Low	Low	High
Business Intelligence Data Scientists	Medium	High	Medium	Low	Low	High
Data Analyst	Low	High	Low	Low	Low	High

Different data science roles will have vastly different expectations on different skillsets. While a data engineer might not be expected to have many business presentation skills, they are expected to dominate all types of programming challenges. Conversely, a data analyst will lean more on their SQL skills and not be exposed to heavy technical problems, but they will be expected to be top-notch presenters.

This table implies the industry demand and difficulty of the positions from top to bottom, with Product Data Scientists being the most in-demand for their specialized, difficult to acquire skills.

Know what role you're applying for. Seek to scout out exactly what needs a company is looking for and what role they are trying to fit you in; it will help you navigate and predict their data science interview process.

[1] This is more inline with dealing with setting up large scale data engineering platforms and integrating various technologies together.
[2] These data scientists typically build the algorithm and productionize it through the data engineering infrastructure. E.g. They would build the recommendation system algorithm and productionize the recommendation system live on the platform.

Here's a high level overview of the specific roles:

Product Data Scientist: End to end data scientist with data engineering skills. Product data scientists lead teams to build a data product. They tweak algorithms and have a strong say in how the data is served to end users. They will often have the engineering ability to deliver on those ideas.

Data Scientist: The unicorn mix of technical skills, business skills, and mathematical knowledge. A data scientist understands how to create and optimize data algorithms, and how to explain their findings. They may need to know less programming than their data engineer peers, but they'll nevertheless need to understand enough to deal with data at scale.

Business Intelligence Data Scientists: Business Intelligence Data Scientists are focused on getting business insights out of data. They will understand enough about statistical methods and different machine learning algorithms to differentiate themselves from data analysts. They build dashboards and complete various analytical studies to help the various teams make better decisions.

Data Engineering: A data engineer isn't often counted on to have advanced knowledge of the statistics and mathematics, but they will have to ace every technical challenge out there to prove they can deal with implementing algorithms on massive amounts of data.

Data Analyst: An entry-level role that relies heavily on making one-off reports by looking through data and interpreting the results. This role typically requires a strong knowledge of SQL and Excel.

The Categories of Data Science Questions

Behavioral Questions

The data science interview process involves a lot of behavioral questions, similar to any other interview. The interviewer intends to test for your soft skills and see if you fit in culturally with the company.

1. Tell me about a data science project you have done in the past?

   Intent: The intent of the question is to understand the depth of knowledge and contributions you have from your past experiences. It tests your ability to tell a story around your work and whether you can tie it to impact on the company you worked with.

   How to Answer the Question:

   - Try to describe a project that demonstrates both product and engineering experience, i.e. the project provided the analytical insight and productionised the insight to make it actionable. For example, if you identified key topics in a text data set through topic extraction techniques, you should explain how these topics furthered company growth in a data product.
   - Go into detail about your specific contribution and the outcome from a business goal perspective. The interviewer wants to know what you specifically did while trying to understand the overall goal of the project.
   - Rehearse your experiences many times. This is a very common question, so have 2-3 go-to projects you can go into extreme detail about eloquently.

2. What have you liked and disliked about your previous position?

   Intent: The intent of the question is to identify whether the role you're interviewing for is suitable for you, and to identify why you're moving on from a previous position.

   How to Answer the Question:

   - Understand the role well. Use the HR contact to get insider information about the role and its challenges. The HR person can be a treasure trove of information about the role, team, history, and key immediate business goals.

- Avoid talking about issues with specific people, and be professional when talking about what you disliked. Introspect carefully and talk to what makes you passionate. For example, talk about deriving insights from data and conveying them to management in an actionable way as something you enjoy. You could also talk about learning new technologies that make data science more actionable through the organization. You could dislike how the organization is not placing data science at the center of its strategy or that the company has had significant attrition at top management level and the direction of the team is unclear. Keep it positive, points-oriented, and away from personal situations.
    - Bad: *I hated that data scientists were always put below the engineers and that management didn't have a clue what the company direction was!*
    - Good: *I realized I wanted to work in a company where data science is part of its core strategy and the company has a clear direction.*

3. Tell me about a situation in the past where you had to convince others about your position on a specific matter. What was the outcome?

   Intent: The intent is to find out how good are you at defending your position and your ability to engender change within a team.

   How to Answer the Question: Try to find an example where you were successful at making the change and that the change is quantifiable in its impact. If possible, use a data science type example if you have one. It's important that you demonstrate your communication and leadership skills here.

Mathematics Questions

Questions about the mathematics angle will come for data scientist roles where you are expected not only to implement algorithms, but also tweak them for specific purposes.

1- How does the Linear Regression algorithm figure out what are the best coefficient values? (This was a question asked in C3 Energy's Data Scientist interview)

Rationale: The intent of the question is to see how deeply you understand linear regression, which is critical because in many data science roles, you won't just work with algorithms in a black box; you'll implement them in some way. This category of question (and you could see it from any type of algorithm) tests how much you know about what is actually happening beyond the surface.

How to Answer the Question: Trace out every step of your thinking and write down the equations. Describe your thought process as you're writing out the solution.

The Answer: At the highest level, the coefficients are a function of minimizing the sum of square of the residuals. Next, write down these equations while paying careful attention to what is a residual. To go further, consider the following:

1. Write the minimization goal (ideally in linear algebraic (matrix) notation) of minimizing the sum of squares of the residuals given a linear regression model. .
2. Solve the minimization equation by illustrating that the sum of square of the residuals is a convex function, which can be differentiated and the coefficients can be derived by setting the differentiation to 0 and solving that equation.
3. Describe that the complexity of solving the linear algebra based solution in #2 is of polynomial time and a more common solution is by observing that the equation is convex and hence numerical algorithms such as gradient descent may be much more efficient.

Statistics Questions

A grasp of statistics is important for solving different data science problems. You'll be tested on your ability to reason statistically and your knowledge of the theory of statistics. Be prepared to recite your knowledge about statistical concepts like Type I error and Type II error flawlessly, and be prepared to demonstrate your grasp of different probability distributions.

1- What is the difference between Type I error and Type II error? (Our alumnus Niraj encountered this question).

Rationale: Companies will want to test your grasp of different basic statistical concepts to test how good you are with the fundamentals of statistics and see how you communicate different ideas you may not often apply with the sometimes technocratic language embedded in statistics.

How to Approach your Answer: Be no-nonsense, and communicate clearly whatever you are asked to define.

The Answer: Type I error is what is referred to as a “false positive,” or the incorrect rejection of the null hypothesis. Type II error is what is referred to as a “false negative,” or the incorrect acceptance of the null hypothesis. You may want to communicate your grasp of the concepts with an example and how it might be relevant to the business at hand. Type I error or a false positive would be telling a man they were pregnant, while Type II error would be telling a pregnant woman they weren’t. If you were running a fraud detection business, you might have a very high tolerance for false positives (a client will not fuss about an email on the potential of fraud), but a false negative (not detecting fraud when it is happening) could be disastrous to you.

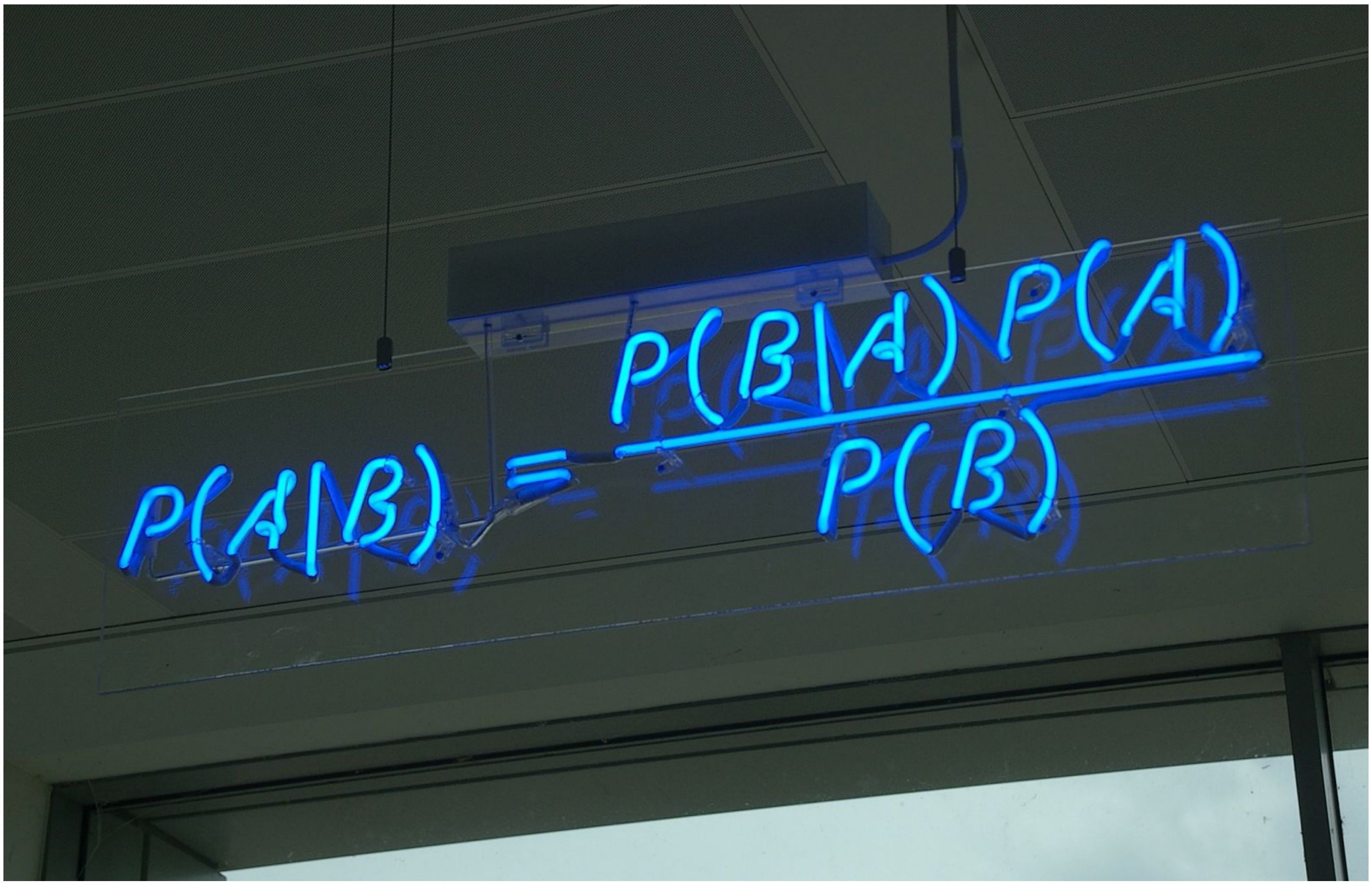
2- This was a question for a data scientist position at a big insurance company. Suppose a population is divided into two groups: aggressive drivers and non-aggressive drivers. 40% of the population are aggressive drivers while 60% are non-aggressive drivers. The probability of an aggressive driver getting into 3 accidents in one year is 15%. The probability of a non-aggressive driver getting into 3 accidents in one year is 5%. John is known to have 3 accidents in the past year. What is the probability that he is (a) an aggressive driver, and (b) a non-aggressive driver?

Rationale: A lot of companies will test your Bayesian inference skills as a primer for how you think statistically. Bayesian probability contrasts with frequentist interpretations of statistics, and your ability to reason through any Bayesian problem will show you have a quick grasp of statistical concepts and the mental math needed for it. If you need a refresher, one of Springboard’s mentors Will Kurt runs a blog called CountBayesie, and he has a wonderful guide to Bayesian statistics.

How to Approach your Answer:
The intent of the question is to see your level of understanding Bayesian probability. Sketch out all of your assumptions and the calculations you’re doing for your interviewer in a logical and organized fashion.

The Answer: Write out what you know.
Probability of aggressive drivers in the population= 40% or 0.4
Probability of non-aggressive drivers in the population = 60% or 0.6
Probability of aggressive drivers getting into three accidents a year = 15% or 0.15
Probability of non-aggressive drivers getting into three accidents a year= 5% or 0.05

You'll want to understand the concept of priors and posteriors for Bayesian equations. A prior is what you are given before the problem, data that you receive. The probability that somebody is an aggressive driver in the population is a prior assumption given to you that you cannot change. The posterior is the probability you derive from using the Bayes Rule on these assumptions (P(A/B)).



Bayes Rule

The first question is "what is the chance John is an aggressive driver if he's been in 3 accidents a year?"

Visually, you're really [trying to draw a Venn diagram]() of probabilities: of all of the people who have been in 3 accidents a year, how many are aggressive drivers? How many are not?

```python
#Your priors
aggDriver = 0.4
nonAggDriver = 0.6
threeAccAggDriver = 0.15
threeAccNonAggDriver= 0.05

#What is the probability that somebody who gets into 3 accidents a year is an aggressive driver?

#probability that an aggressive driver gets into 3 accidents a year times the number of aggressive drivers in the pop.
A = (aggDriver * threeAccAggDriver)

"""our numerator and the probability that a nonaggressive driver gets into three accidents a year times # of
nonaggressive drivers."""
B = ((aggDriver * threeAccAggDriver) + (nonAggDriver * threeAccNonAggDriver))

"""deriving the posterior by dividing 3 accident a year people who are aggressive against the sample of 3 accident a year
people who are both aggressive and nonaggressive"""
posterior = (A / B)

#converting the decimal to percent
print ("{:.0f}%".format(posterior * 100))
```

```
67%
```

There is a 67% probability (really 66.66% repeating) that somebody who gets into a 3 accidents a year is an aggressive driver. This is now your posterior.

The probability that somebody who gets into 3 accidents a year is non-aggressive is just the flipside of that. 1 - 0.6666 = 0.33333 repeating, or 33% probability.

3- What is probability distribution type (or show the derivation of the pdf) you would use to describe the following random variables?

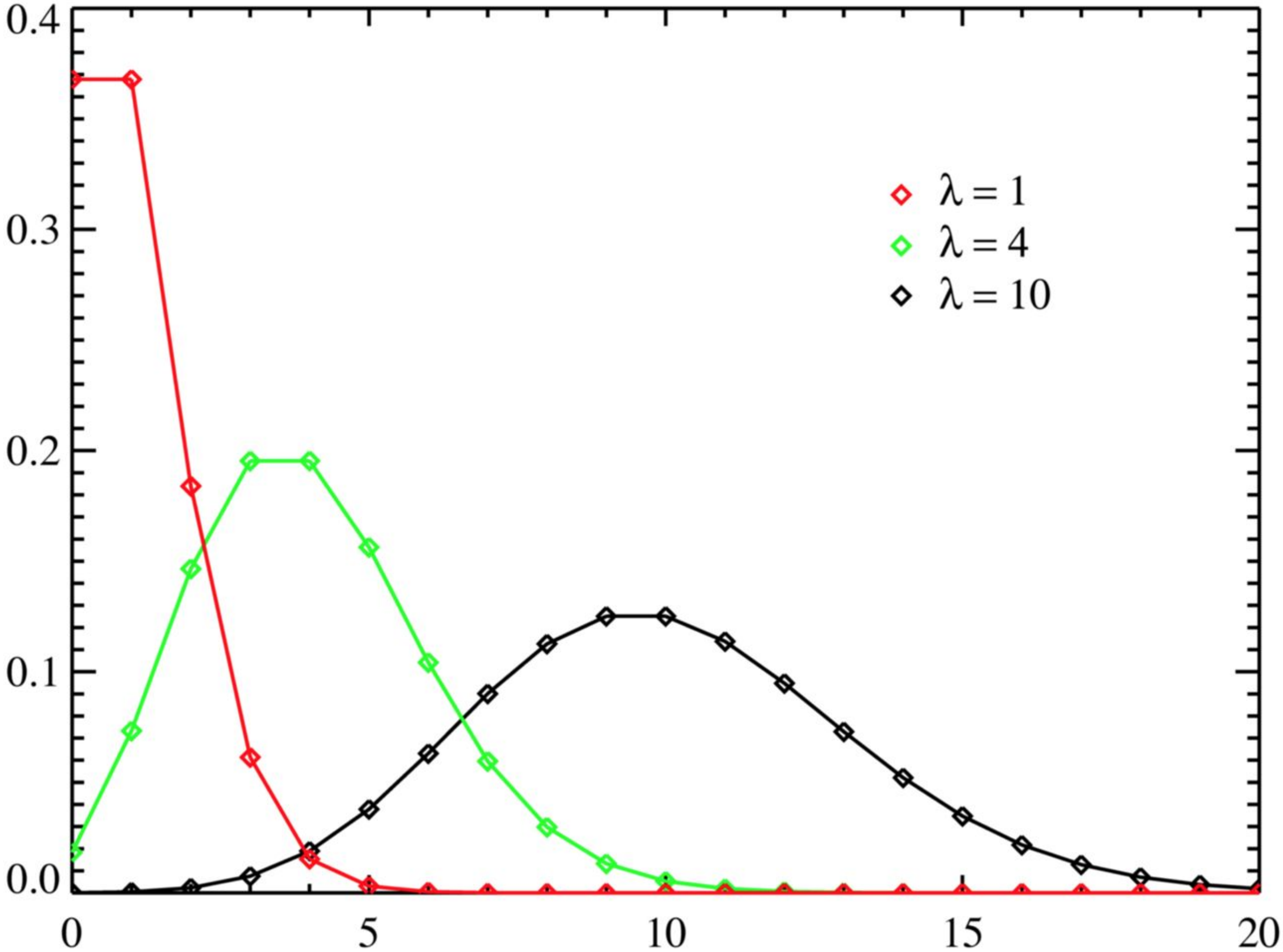
a. Probability of k customers arriving to a restaurant within a duration of t minutes
b. The probability of the height of a person in a crowd being at least X inches
c. The probability of the sum of two 6-sided fair dices being Y
d. The probability of having k heads thrown out of N coin throws

Rationale: This question tests your knowledge of probability distributions and tests whether or not you know what models to use given how your data is organized.

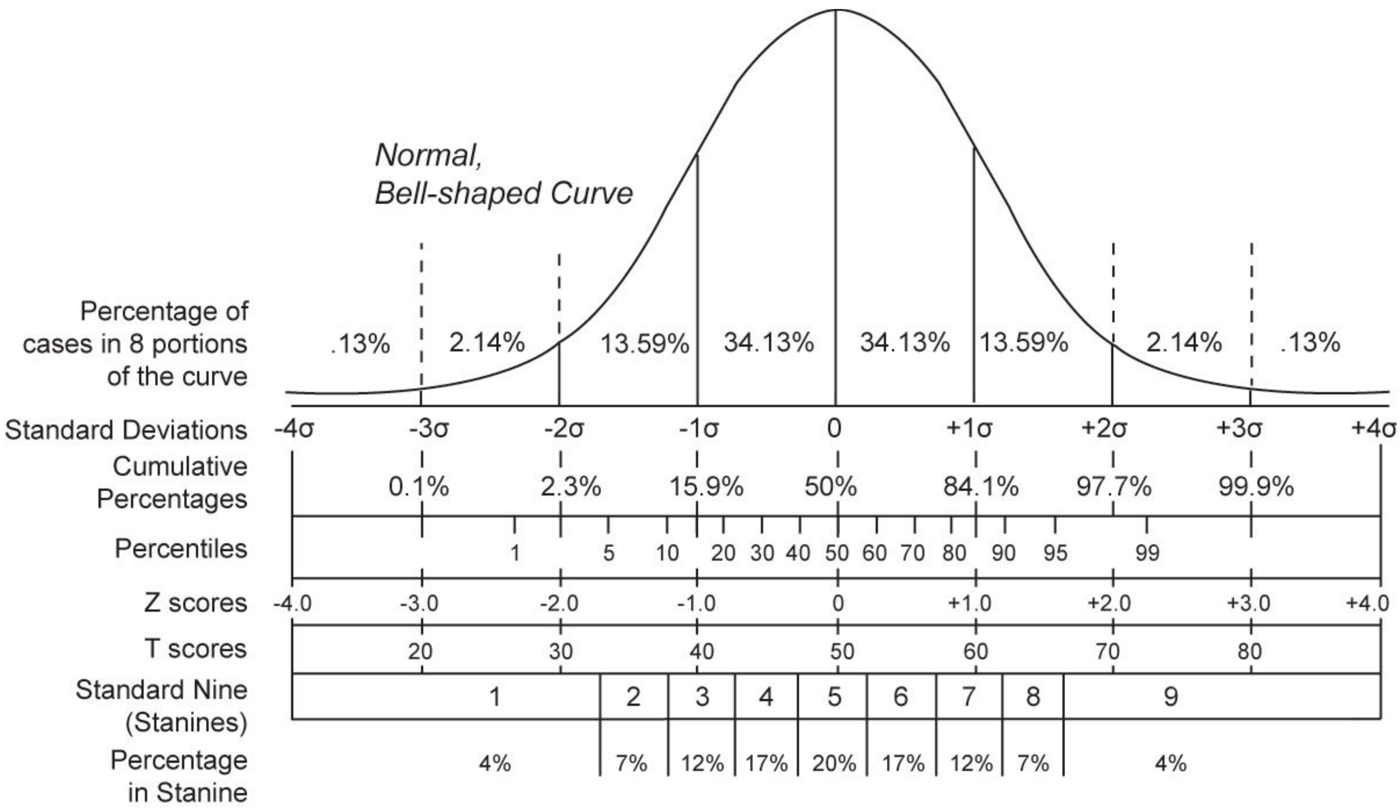
How to Answer the Question: Explain your assumptions about the data and the details of how the distribution in question fits the model. Be able to visualize distributions and explain to the interviewer why the distribution you visualize follows the model.

Answer:

a. Poisson distribution. This is assuming that customer arrivals are entirely independent from each other.



b. Normal distribution. Note that in a continuous distribution the likelihood of being exactly X inches is zero.



c. P(sum(x1+x2) = {0,1,(2,12),(3,11),(4,10)...36}) = {0,0,1/36,2/36,3/36,...}. You can plot this out where the x axis is the sum, and the y-axis is the probability. Illustrate that this is a probability mass function vs a continuous probability distribution function.

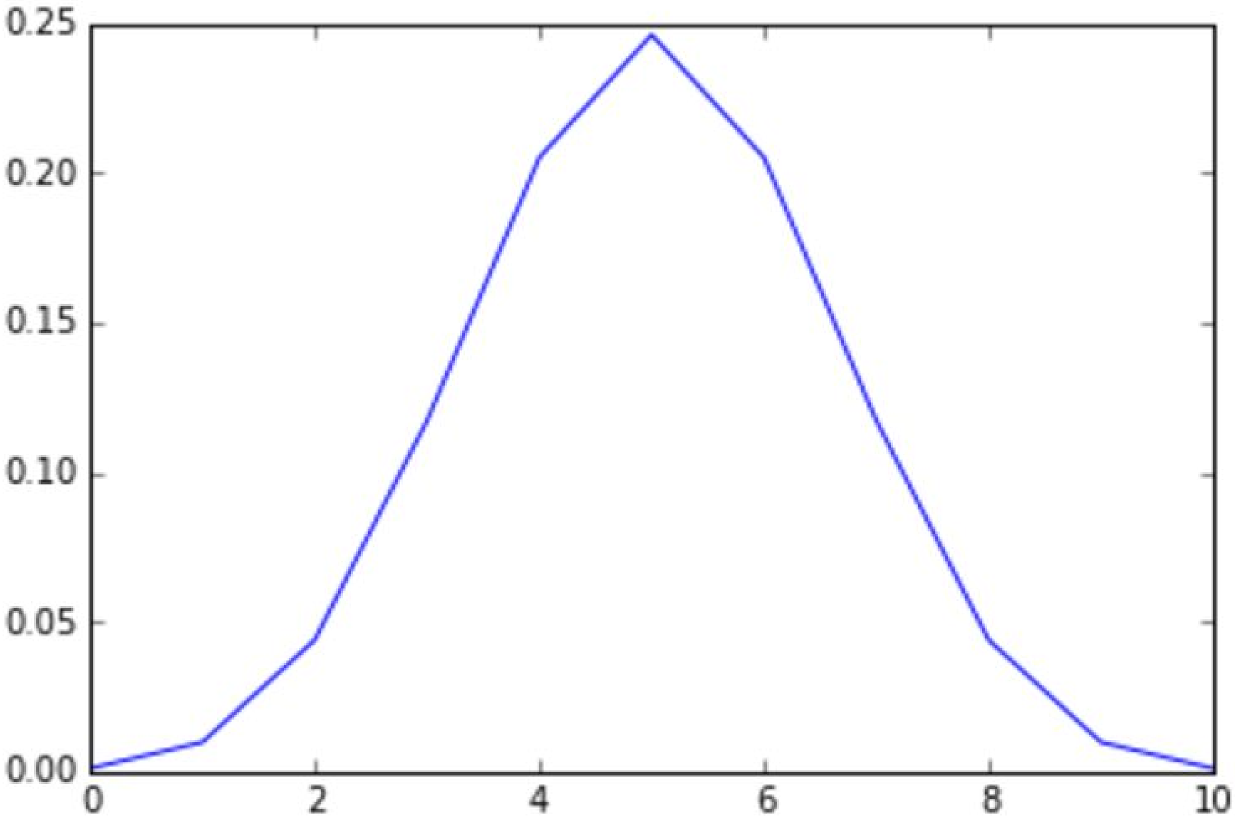
d. Binomial distribution. P(k is the number of heads in N throws):

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

```
%matplotlib inline

import scipy, scipy.stats
import matplotlib as plot
x = scipy.linspace(0,10,11)
pmf = scipy.stats.binom.pmf(x,10,0.5)
import pylab
pylab.plot(x,pmf)
```

```
[<matplotlib.lines.Line2D at 0x109f02860>]
```



Note that this visualization says there is a 25% chance you will get 5 coins out of 10 to be heads.

Coding Questions

A large part of a data science role, especially if it is more focused to data engineering, is programming to implement algorithms at scale. Be prepared to face something similar to a software engineering interview where you'll be tested on your experience with the technical tools a company uses and your overall knowledge of programming theory.

1- *SQL - Given a table of transactions (Transaction_ID, Item_ID, quantity, purchase_date (MM/DD/YY)) and another table of prices (item_ID, price), give the following information:*

1. *Total revenue*
2. *Total number/average/standard deviation of purchase quantities for the set of weekdays (Monday-Friday) ordered by descending number of purchases.*
3. *Number of item_ID's that were NOT purchased in the weekdays.*

Example table of transactions (defined as transactions):

Transaction_ID	Item_ID	Quantity	Purchase_Date
1	1	5	06/28/2016
2	2	3	06/27/2016
3	2	5	06/27/2016
4	2	1	06/26/2016

Example table of prices (defined as prices):

item_ID	Price
1	$2
2	$3

Rationale: The use of SQL to query databases is prevalent in larger startups and established companies looking to leverage their company. If you are a data analyst, your technical interview may exclusively be SQL questions. Understanding how to get data the right way can make the difference between getting a job and not.

How to Answer the Question: You will often be asked to sketch out your code on paper or work with a collaborative coding tool like [HackerEarth]() where you will be coding in the interpreter and your code is seen live by your interviewer. Make sure you try for the most efficient solution with as few errors as possible given a short time constraint. Use something like [SQLFiddle]() if you want to practice your SQL querying skills!

Answer:

1.
```
SELECT sum(a.quantity*b.price)
FROM transactions AS a
JOIN prices AS b ON a.item_ID=b.item_ID
```

This will join the price column from the prices table onto the transactions table, allowing you to multiply the quantity of each item with its price and then to sum up that multiplication. This will yield an answer of $37 for our two example tables.

2.
```
SELECT DAYOFWEEK(purchase_date),
       sum(quantity),
       avg(quantity),
       std(quantity)
FROM transactions
WHERE DAYOFWEEK(purchase_date) BETWEEN 2 AND 6
GROUP BY DAYOFWEEK(purchase_date)
ORDER BY 2 DESC
```

This query will use the [DAYOFWEEK]() function in MySQL, which returns a number index of which day a calendar day is, and returns a value from 1 and 7, with 1 corresponding to Sunday, and 7 corresponding to Saturday. Filtering, selecting, and then ordering by descending quantities satisfies the question of table 2.

If you ran the query on the sample table, you'd get the following output, with 2 corresponding to Monday (June 27th, 2016):

DAYOFWEEK(purchase_date)	sum(quantity)	avg(quantity)	std(quantity)
2	8	4	1
3	5	5	0

3. Two approaches (using Left Join vs. GroupBy):
   a.
```
SELECT COUNT(DISTINCT A.item_ID)
FROM transactions A
LEFT JOIN
  ( SELECT purchase_date
   FROM transactions
   WHERE day_of_week(purchase_date) IN (Monday,
                                        Tuesday,
                                        Wednesday,
                                        Thursday,
                                        Friday)) AS B ON
A.Transaction_ID=B.Transaction_ID
WHERE B.purchase_date=NULL
```

   b.
```
SELECT COUNT *
FROM
  (SELECT item_ID
   FROM transactions
   WHERE IsWeekDay(purchase_date) != TRUE groupby
item_ID)
```

Either approach will narrow down a table of items that were not purchased on the weekend, then apply a special count to it.

Tips for SQL Questions:

1. *Do small queries first instead of going to the subqueries. Break the problem down to specific intermediate tables, and do the queries for those intermediate tables first.*
2. *Be careful of the column you do the join on. Ask whether you want to keep rows where there wasn't a match (i.e. left join if needed).*
3. *If you don't know the exact transformation function, assume the existence of one, state the input/output to the interviewer, and move on.*

2- Develop a K Nearest Neighbors algorithm from scratch (**algorithm coding**)

Rationale: Showing you can write out the thinking behind an algorithm and deploy it efficiently in a given time constraint will be a critical way to evaluate data engineering skills. This kind of question will be asked of data scientists who have knowledge of both algorithms and their technical implementation, or data engineers who are given context on what is the algorithm. This questions can be asked of any algorithm, but most of the time interviewers will user K-nearest neighbours, as it's relatively easy to come up with code that can work.

How to Answer the Question: First, clarify the question. Given a feature vector, find the euclidean distance from that vector to every other known vector, and take the class that is the majority within the closest K vectors. This particular question tests your understanding of matrix computation and how to deal with vectors and matrices. Start by going through a sample set of inputs and outputs, and manually derive the answer. Also, keep an eye on the time/space complexity. In the solution below, each prediction is of O(2N + N log N) time complexity where N is the number of rows of training data.

You will want to write down your solution. Syntax counts, and so do various faults that will stop your code from compiling properly, but it doesn't count as much as expressing the logic behind the algorithm, and showing how you can apply algorithmic thinking to the plane of computer science.

Solution:

```python
import collections, numpy

class NearNeighbor(object):

    def __init__():
        pass

    def fit(X, y):
        self.X = X # This is a numpy matrix of NxM
        self.y = y  # Numpy vector of Nx1
        return None

    def predict(X_test,k): # X_test is a 1xM vector input
        X_test_matrix = array([X_test,]*self.X.shape[0]) # This just copies X_test into N rows
        Distance_vector=np.sqrt(np.sum((abs(X_test_matrix - self.X))**2,axis=1))
        # Here we get the distance to each of the training vectors. Note the use of **2 and
        # np.sum to get the distance
        Sort_index = np.argsort(Distance_vector)
        Closest_k_Classes = y[Sort_index][0:k] # take the class of the k closest vectors
        return collections.Counter(Closest_k_Classes).most_common()[0][0]
        # return the most common class
```

Other coding questions can be more big data specific. For example, asking about mapreduce is a typical question in the case that the position requires analysis of very large data sets. Questions here ask how to take the average of a large data set or find the most frequent event in an event stream.

3 - How does wordcount mapreduce work on Hadoop?

Rationale: You will get questions about Hadoop and big data tools if you indicate on your CV that you have experience with them, or if the company in question deals with massive data sets. Larger Fortune 500 companies and tech startups that have scaled beyond millions of users are likely to challenge you on your use of big data. You should demonstrate a knowledge of mapreduce, which can come from work experience or playing around with massive datasets on your own. Horton has resources dedicated to helping people learn MapReduce if you need to brush up.

How to Approach the Answer: This question sees how deeply you understand the mapreduce framework on Hadoop. This is typically done using Java. Although the word count problem is an extremely commonly understood one, knowing how it's implemented within the Java Hadoop framework is the important piece here.

Answer: The driver code would set up the job and configuration. If the data comes from HDFS and output is written to HDFS, add the input/output path to the job to those directories. Then the mapper job would take each line in the file and emit a value of 1 for each word as the key. Note that the data passed between mapper and reducer must use the Hadoop data structures such as Text and IntWritables since these are more efficient for byte array serialization vs. primitive types such as Strings and Ints. The mapper output would then be collected in each executor, and then the combiner task would be executed. The combiner is a local aggregator that is optionally set to reduce the amount of data sent between the mappers and reducers.

Once all the mappers are complete, only then can the shuffle phase begin. You might observe your jobs stuck at 33% reducer, which implies that the shuffle phase is waiting on the mappers to complete. Once all the keys are sent to the reducers based on this shuffle, the sort phase begins on each reducer. After that, the reduce logic is executed, and the output can be written to another HDFS file.

Common follow-up questions in interviews would be to estimate the time complexity of this algorithm, and the amount of data the system writes or communicates between machines. Don't forget to take redundancy into account, i.e. a Hadoop system usually makes multiple copies of data in case a machine goes down.

Scenario Questions

1- If you were a data scientist at a web company that sells shoes, how would you build a system that recommends shoes to a visitor? (Question asked in Verizon Data Scientist Interview)

Rationale: This question tests how you think about your work in terms of delivering products from end-to-end. Scenario questions don't test for knowledge in every field; they are set to explore a product from beginning to delivery and see what limits the candidate would have. While also evaluating for holistic knowledge of what it takes to manage a team to deliver a final product, this question is to see how the candidate would fit into a team situation.

Typically, data scientists will be asked this question, while data engineers or analysts might be asked for specific parts of the scenario relevant to them. Data engineers might be asked to think of how to implement a certain algorithm at scale without having to think of the algorithm itself, while data analysts might be asked what data they'd query to determine users' historical preferences for shoes.

How to Answer this Question: Be very honest as to where you can add a lot of value (emphasize what parts you've had experience in), but don't be shy about where you expect to get a little bit of help. Try to relate how your technical knowledge can help with business outcomes, and always enumerate the thought process behind your choices and the assumptions that guide them. Don't hesitate to ask questions that can better tailor-fit your answer.

Answer: Break the answer to two components: Data science and Data engineering

Let's discuss the data science element first. If it is a new company that does not have much historical user data, go with item-item similarity. If the number of different items/shoes is extremely large, consider using matrix factorization techniques to reduce the dimensions.

If you have historical data around user preferences (e.g. ratings of shoes), you can use a collaborative filter type approach. Mention specifically the rows and columns of the matrix you generate with either approach. Then discuss what kind of similarity metrics you would try. E.g. euclidean distance, Jaccard similarity, cosine distance.

After explaining the algorithmic aspect, you would discuss the data engineering side. Propose an engineering infrastructure that scales to millions of users/shoes where recommendations are generated in real time. As an example, you can stream the user data to a S3 bucket. You can perform the matrix analysis on a nightly basis, pre-compute the entire set of recommendations on a per user basis, and store this in a in-memory database such as Redis. Then you could build a REST API that would query the database and respond with the recommendations given a user id.

Question 2. How much is the monetary value of a share of a Change.org petition on facebook? (Change.org interview question)

Rationale: The intent of this question is to see how much you understand about the business and how well you can break a fairly complex problem down to basic concepts and then convert these concepts to analyzable chunks based on the available data. This is a good test to see how well you can absorb a company's framework for the data and how well you can communicate business insights derived from your data analysis.

How to Answer the Question: Make sure you research the companies you interview for thoroughly, especially their revenue model. Get a sense for what important metrics the company would use to track its performance, and get used to thinking about what actions a company must drive to make revenue. Ask questions and state any assumptions you might have, which sketch out how you're thinking about this problem, then answer with force and conviction as if you're presenting to your supervisor.

The Answer: This question requires some basic understanding of the Change.org business. A share of a petition can result in revenue generation in two different ways -

1. Another user clicking on an advertisement (i.e. signing a paid petition)
2. A new user signing up on the system who then goes on to click on a set of advertisements during that user's lifetime

The first step is figuring out a methodology that would allow you to derive a value of both of these ways. The trick is to start simple. You can simplify the value equation to the following:

Value of a share = Expected revenue from clicking an ad + Average number of new signups per share event * Lifetime Value of a new signup
Expected revenue from clicking an ad = Likelihood of an advertisement click * Average cost per click charged to publishers

Likelihood of an advertisement click can be derived by just looking at the historical data and finding the average conversion rate over the course of a time window such as a month or year. A similar value can be derived for the cost per click.

For the LTV, it's a little tricky. You need to look at users over the course of similar lifetimes and derive their total revenue generated. One common method of doing this is called the cohort analysis or retention analysis. You can group users that signed up on a specific month and look at how many of them clicked how many ads over the course of the next twelve months. Do this over twelve different cohort months, and then take the average revenue over the lifetime. Now, the lifetime to analyze can be set to be however long it takes that cause and effect relationship to be considered negligible, i.e. the user that signed up due to the initial share would have signed up anyway beyond that time window, hence the revenue generated cannot be solely attributed to the share.

Once you have the LTV, plug it into the original equation, and you have the value of a petition share. There are deeper elements you can go into, such as the revenue generated by the newly joining users sharing themselves which causes other users to join. Make sure that if you are going to include additional elements to your answer that it doesn't dilute your main message. Stay laser-focused on answering the original question. If you have assorted thoughts on the situation, leave them to the end.

3- Given a set of historical news articles that have been classified as specific categories such as Sports, Politics, World, how would you classify a new article?

Rationale: This question looks at how deeply you understand the data science methodology and your experience with dealing with unstructured text data, an important test for how comfortable you are with data formats that might be difficult to deal with.

How to Answer the Question: Specify how you would organize the text and how you think of classification systems.

Sample solution:

1. Explore the data and understand key elements of the data.
    a. Plot the distribution of various categories in your training set to determine if there is label imbalance.

    b. Look at the text to identify anything strange, such as non english text, heavy abbreviations, or misspellings.
    c. Do topic extraction to identify keywords for specific latent topics and find correlation to the labelled categories. This may give you a hint as to whether there are latent topics (keywords) that may correlate better than just using all the words.
2. Derive the training set by cleaning up the text. Remove lesser informative elements such as punctuation, abbreviations, and unicode characters. Do further cleaning by taking the lower case of words and lemmatization/stemming.
3. Use a TFIDF vectorizer to convert the data to a bag of words model with TFIDF metric. et lower and upper bounds to TFIDF to reduce the vocabulary size.
4. Build a pipeline where you can train various models and compare their performance relative to metrics such as AUC, F1 score, precision, and recall. You can do gridsearch to automate the cross-validation aspect as well.
5. Once you get the optimal model, you can publish this model to production using a pickled model (in python) or POJO (in java). This model can then be queried by using the exact same process of cleaning as done in #2 and #3 for the new articles.

4- Design an experiment to figure out which web design alternative to use. Assume there have been no other experiments done and there is no knowledge of the user behavior. Discuss potential issues that can occur with the conclusions and how to avoid them.

Rationale: Many web companies ask this question because it is their bread and butter to optimize their website for better business results. Think of Facebook constantly changing their homepage to get you to post more. The data scientist's role is often in helping the product manager setup the experiment or interpret the experiment results. The goal of the question is to see the depth of the knowledge of the interviewee in this topic.

Solution: Identify the nature of the change and the metric to consider to decide which version of the site to choose. For example, click through rate and average number of Facebook shares.

Next, decide the number of samples/visits necessary to hit the necessary statistical significance (e.g. 95%). This can be done by using a chi-squared test (if we are using a binomial random variable of clicking vs. not clicking) or a z-test (if we are using a normally distributed random variable). You can then evaluate the p-value to identify whether the metric of the B test is statistically significantly different than the metric of the baseline A test. If it is and the metric is better than the baseline, then the alternative site is the better way to go.

Some other issues you should consider in this answer:

1) Identify potential biases due to interactions across pages. Talk to the product manager and see if there are ways that a random sampling may not work to test the nature of the change you're proposing for a web page.
2) Perform a A/A test which implies testing two random samples of visitors, and check if the distribution and metric of choice does not have a statistically significant difference. This will ensure the fairness of the A/B test. An A/A test ensures that your audience doesn't have a particular skew or bias and a randomized selection for an A/B test will be statistically relevant.
3) What if the metric that we are evaluating has significant outliers that may cause the average to be a poor metric? The distribution may be highly skewed. We assume the average is a good metric of comparison since central limit theorem holds. This may not be true. Hence, check the distribution of the metric to ensure that taking an average (e.g. conversion rate or average number of shares per user) is a reasonable metric when comparing between alternatives. If one user has thousands of shares attributed to their account, for example, using share rate per user may not be the best performance metric.

In summary, case questions are designed to test for your experience and your knowledge in different fields of data science. They are designed to see if you have any limits to your ability. Demonstrate your knowledge thoroughly, and you'll come off well in any case analysis.

Tackling the Interview

1) Dressed accordingly. If it's an interview for a startup, a dress shirt will suffice. If it's an interview with a bank, wear suit and tie. If you're unsure of what to wear, ask.

2) Before you come into the interview, research your interviewer and the company. Come up with good questions to ask.
3) Be at the top of your game mentally. Eat well, be hydrated, exercise well, and do whatever you can to make sure you're prepared to handle an interview.
4) Answer questions in detail, and sketch out your thought process.
5) Smile, and be confident. Don't come in stressed. Meditate, stretch, or read--do whatever it takes to get you to your peak.

Conclusion

The data science interview process is a multi-faceted beast. You'll be challenged to program and come up with technical algorithms on the spot. You'll be challenged about your statistical and mathematical knowledge. You'll be challenged on your ability to lead teams, communicate, persuade, and influence.

It can be hard to see how to pass this beast of an interview process. Thankfully, we condensed actionable insights from successful applicants and the hiring managers on the other side of the table.

What Hiring Managers Are Looking For

Interview with Will Kurt (Quick Sprout)

Bio: Will Kurt is a Data Scientist with Quick Sprout. His main interests are probability, writing, and Haskell. He blogs at CountBayesie.com and can be found on Twitter as @willkurt

What do you look for when you're hiring candidates?

The biggest thing for me has always been a combination of creativity and genuine curiosity. In a startup environment, new problems come up everyday in a wide range of areas. One month you may be helping the product team add new features. The next month, you'll help sales improve their process, and the month after, you'll be helping marketing restructure their testing setup. The most valuable candidates are the ones interested in all of the company's data related problems and always thinking of new and interesting ways to solve them.

What's the best piece of advice you can give to people going through the data science interview process?

In my experience, all small companies and startups worth working for are excited about the idea of adding a new data scientist to the team. They hope your skills and experience will help them solve a range of problems they've been struggling with. Show up to the interview ready to listen to what they're trying to solve and get them excited about solving problems together. Every chance you get ask people what they're working on and get them brainstorming with you about ways you could make their day better. There are thousands of candidates out there with superb quantitative skills, but candidates who care and are excited are very rare. Leave the interview with everyone wanting to work with you on a project, and they'll be the ones hoping you say "yes."

What kind of interview questions do you like to ask? What are you trying to test?

All I care about is how your mind works once it's fixed itself on an interesting problem. At Kissmetrics, I gave out an open ended "homework" assignment. There was an obvious approach to the problem (build a classifier), but I mentioned this and cautioned that part of the test was to see if you could come up with something interesting. The results of the assignment didn't have to be long or complicated. What mattered is that they started a conversation and showed that the candidate had genuine curiosity in finding something worth talking about. Given that a candidate can code and is comfortable with linear algebra, calculus, and probability, they have the basics to learn everything else. It is very hard to teach someone to think creatively or become passionate about problems.

What is different about how Kissmetrics and Quick Sprout hire data scientists?

Right now, Quicksprout is a very small team in the early stages of product development, so we're not hiring new data scientists at the moment. One thing that aspiring data scientists should know is that many startups and small companies are looking for a data scientist but may have given up on finding one as the search process can be exhausting. One of our best candidates at Kissmetrics showed up at our door and said, "I want to work here!" People coming from academia or other large organizations might not be aware of how flexible startups and small companies can be when it comes to hiring. If you think a company is doing cool work, connect with them. It's hard to make a better impression on a group of people excited about their work than telling them you love what they're doing and want to be a part of it. Even if that company isn't hiring, you'll be at the top of the list if/when they do start looking.

Interview with Matt Fornito (OpsVision Solutions)

Bio: Matt Fornito is a Data Scientist and Leader with over ten years of experience in the research, analytics, and management domains. A passion for learning and devout work ethic continues to help him grow. This interview is transcribed from notes taken on a phone call with Matt.

What does you look for when you're hiring candidates?

I feel most comfortable hiring people with a strong quantitative background who can learn programming rather than the other way around. A Masters or a PH.D is very important to me, as I feel that undergrad is not a strong signal of success; it's a relative breeze for most people. I prefer hiring people able to pick up programming and effective communication-- knowing and understanding what the technical problems are to implementing a solution and being able to communicate those concepts is key. What differentiates data scientists and data analysts is the ability of data scientists to deeply understand data problems and how to solve for them.

I like recruiting Masters and PhDs from math and statistics, chemistry, physics, and bioinformatics and engineering. There are a small handful of people in MBAs that have worked out great for me. I am actually a PhD in organizational psychology, so though I tend to try to hire people with STEM backgrounds, it isn't a strict limitation.

What's the best piece of advice you can give to people going through the data science interview process?

Recruiters look at education level and the last two jobs on the CV and their pedigree. HRs only take a very quick glance at CVs, so you have to stand out in a matter of seconds. One piece of advice: get yourself into a big company that has a pedigree like Facebook, or go into a startup and take a high position so that you can stand out easily for advanced data science roles.

“Walk me through a project” questions where a hiring manager will ask exactly how you built something in the past are huge--everything from what data was used, what tools were used, what the outcomes were are important to recount clearly. successful interviewees have a comfortable grasp on what they’ve worked on and are ready to storytell on that element and relate how their work impacted the business they were working for.

What are you testing for?

Questions I ask involve working around a project to test problem solving and communication skills across the interview. I am also assessing a candidate’s passion for the company and data science. A drive for continuous learning and love of problem solving are key differentiators. Then on the technical side, I am interested in seeing candidates work on how to optimize data with Hadoop and Spark and working on the tradeoffs between different data science solutions. Do they think like a data scientist? Have they done data science work? These are important questions I am looking to uncover with my interview process.

I will then go into math questions such as asking how gradient descent, statistical techniques, and random forest work. A couple of situational questions where the candidate is put through a hypothetical client situation are deployed to see how the candidate would handle interfacing with clients. I have a strict requirement of ability to program in Python or R, but I am flexible with C++ and Java. I don’t believe in HackerRank like testing situations where you are expected to trace out a solution; I would rather test for adaption to new programming languages and an ability to learn skills rapidly. Anybody hired is going to have to have the latent skill of **adaptability, and that is the key thing I am testing for.**

Interview with Andrew Maguire (PMC/Google/Accenture)

Bio: Andrew has been working in Analytics/Data Science for 7 years in various roles across many different industries. He is a Data Scientist at Penske Media Corporation focusing on both data engineering infrastructure as well as applied business analytics. Prior to this position, he worked at Google (marketing analytics, then local data quality), Accenture's Analytics Innovation Centre (consultancy), and Aon's Center for Innovation and Analytics (product development team).

What do you look for when you're hiring candidates?

Beyond meeting the basic requirements from a technical and experience point of view, I'd say enthusiasm, willingness and ability to continually learn new things are key..

A good attitude is super important, so someone who is able to also tell me about their weaknesses as well as strengths is a good way to draw this out (sometimes selling too hard is a bit off putting; humility is much better).

Being approachable, open and honest is something that's key on the 'team fit' side. You don't have to know the answer for everything but being able to work with others to come up with a decent solution is crucial.

What's the best piece of advice you can give to people going through the data science interview process?

On the technical stuff, take your time, write stuff down, and ask clarifying questions. Also don't be afraid to tell them if it's an area you've not worked on before or an algorithm you're not that familiar with. Being able to admit when your knowledge is limited is super important as a data scientist; continually learning is one of the most important skills required.

Make sure you have two or three data science 'stories' you can chat about with an interviewer that touch on problem formulation, data wrangling, analysis and insights, visualization and stakeholder communication. Try to get the balance right between cool nerdy technical stuff and showing business understanding and insights. These 'stories' can be projects from your previous roles, college assignments, or projects you did on your own time. Get good at spotting openings from interviewer questions to use your stories to show concrete examples and experience. I find that chatting (in detail) about projects the candidate has done in the past is the best way to get a proper feel for them (and best place to probe deeper from), so make sure you make it easy for the interviewer to be interested and excited to ask you about some project's or example's from your CV.

What kind of interview questions do you like to ask? What are you trying to test?

What's the biggest or most complex dataset you have ever worked with? What problems did it create? (Trying to begin a discussion here that can lead into judging data wrangling skills and experience)

Give me an example of a time when you analysed a dataset, and communicate your findings back to the business. What was the problem faced? What did you find? How did this affect the business? (Touch on the extracting business insights and communicating back to stakeholders aspects)

I ask questions very related to what is on the CV, so if it's a project from a previous role for example, I want you to explain what the problem was, what sort of data you used, how you used it, what the insights were, and how this all fits into the wider business. Choose what you put on your CV **very deliberately**. If you find it hard to get all on two pages then maybe have different 'types' of cv's you might use for different types of roles.

Finally, I ask candidates to give me an example of a time when they failed, then add what they think went wrong and what they would do differently in future. This is something that comes out of HR 101, but I like to hear what they have to say :)

What is different about how Google hires data scientists from the rest of the industry?

I'm not sure there is too much of a difference anymore. Generally it depends on the specific role. For very specialized positions that are often more like research or fellowship positions, you would get much more detailed technical questions and problems to drive into the relevant area of expertise in very fine detail. For more generalist or business related roles, the focus is more on the right mix of technical skills, business understanding, working in teams, and communicating results to stakeholders.

The main difference in Google is that you have a lot more interviews and meet more people, so behind the scenes there are around 6+ people who have all met you and probed you from their own different angles. These people all have a different view of you and your strengths and weaknesses and must come to a decision and conclusion together that typically involves trade offs. Being able to show decent level of competence across the board as opposed to being a rockstar in one area but letting yourself down in others will generally serve you well. This is where attitude and being easy to get along with can be most important; even if you fall a little short on one of the competencies, if they like you and feel you could easily get up to speed in that area in a few months, then it's less likely to be a deal breaker.

Interview with Hristo Gyoshev (MasterClass)

Bio: Hristo Gyoshev is the Head of Business Operations & Strategy at MasterClass, a fast-growing startup that is democratizing access to genius and re-imagining online education. He previously worked on corporate strategy, business operations, and product strategy at both consumer web (e.g. Yahoo!) and enterprise SaaS companies. MasterClass is looking to hire a Data Scientist & many other positions. Check out the details at careers.masterclass.com/

What do you look for when you're hiring candidates?

One of the main assets we look for is a desire to work on projects across a very broad range of analytic disciplines – from quantitative market research and/or designing, conducting, and analyzing user surveys, to statistical analysis, to business intelligence and analytics. We also look for candidates who are comfortable learning something new to remove bottlenecks and keep a project moving, when necessary.

In terms of educational background and experience, we're looking for an analytical background that combines 1. sufficient knowledge of statistics to determine what is or isn't a valid statistical inference, recognize & prevent biases, etc.; and 2. the desire and ability to obtain and work with real-world data (which is always imperfect) and derive actionable insights.

Someone who has a very strong quantitative background and ability to process and analyze data using Excel, SQL, and Python or R; who also has experience in social science research or market/user research (through either academic or industry work); and who has experience with business reporting / analytics, could be an ideal candidate for us.

What's the best piece of advice you can give to people going through the data science interview process?

Strive to understand and keep in mind the broader context of the problem you are being asked to solve -- or the problem behind the question you are being asked. Whenever you are asked to perform a certain analysis, or build a model, someone at the company believes that this would help them solve a particular problem. Sometimes you can tell in advance that it won't, and sometimes you can suggest a better approach. Your analysis/model/other work-product will always be better if you start from a good understanding of the core objectives of the 'clients' of your analysis. (This applies as much to questions you are asked during the interview as it does to projects you are asked to work on once you get the job.)

What kind of interview questions do you like to ask? What are you trying to test?
We like to understand a candidate's previous experience with various types of work that we expect will be relevant to their role. Thus, we may ask for examples of specific types of projects they have worked on, and then ask them to walk us through their approach and thinking, the tools they used, the major challenges they encountered, and how they resolved them.

We may also ask candidates to complete a short project to see how they approach some specific problem – and yes, to be able to see the quality of a deliverable they produce.

What is different about how MasterClass hires data scientists?
Compared to most Data Science roles, the job with us involves very little machine learning or algorithms, and only minimal data wrangling, but a very wide variety of analyses that would inform a broad range of decisions about the products, business, and operations of the company. The work would, of course, involve some exporting, processing, and analyzing data from various systems, but would also involve building various predictive models; designing, conducting, and analyzing surveys or experiments; helping to define and setup reporting & metrics; and conducting one-off analyses related to various aspects of our business operations.

Correspondingly, we don't need candidates to be proficient in machine learning or algorithms, but we do need them to be highly versatile and familiar with a number of other aspects of data analysis. We also need them to be willing and able to learn tools or methods they may not have previously used.

Conclusion

Hiring managers across the board love it when you demonstrate:

1) Passion for the company and data science in general
2) An ability to get along well with everybody, which may even help you with weaknesses in your technical ability
3) Strong willingness to learn and demonstrated ability to rapidly do so
4) A strong record of previous projects and the ability to relate previous projects with impact driven
5) Strong analytical ability

Now let's talk about the other side of the table: successful applicants who now work as data scientists.

How Successful Interviewees Made It

Sara Weinstein

Data Scientist at Boeing Canada-AeroInfo, Springboard graduate

What is advice you'd have for how to ace the data science interview process?

In terms of preparation, I wish I spent more time thinking about analytics strategy. I prepped hard on stats, probability, ML, python/R...all the technical stuff, but was nearly caught off guard by a straightforward question about how I'd approach a particular problem given a specified data set. My answer wasn't as confident as I would have liked. I'd been so focused on the "hard" stuff that I hadn't thought that much about higher-level analytics methods & strategies.

What surprised me and what I found difficult:

How long the process took. I knew to expect several interviews, and in fact had three. With nearly a week between each, plus waiting for my background check to clear, the process from first contact to firm offer took a month. It was stressful to say the least. Staying positive, confident, and prepared for a whole month was challenging. It would have been much easier to bear if I'd known in advance that it would take that long. For others facing a lengthy multi-interview hiring process: meditation is your friend. It helped me sleep at night, and I used the techniques right before interviews to channel calm and confidence. ■

Niraj Sheth

Data Analyst at Reddit, Springboard graduate

What is advice you'd have for how to ace the data science interview process?

I wish I had studied more fundamental statistics before interviewing. It's silly, but people often look for whether you are familiar with terms like Type I and Type II errors. Depending on the time you have, I suggest getting a statistics textbook and at least becoming familiar with the terms out there.

I should have probably expected this, but I was surprised how poor we are as an industry in evaluating projects. When I talked about past projects, everyone just cared about interest value (does the analysis say something interesting?) -- nobody questioned deeply the methods I used.

You didn't ask this, but there were also some things I did that I think worked out well. One is to have a live project up somewhere with a neat visualization (i.e. more than a github repo with a readme). It doesn't have to be fancy--just prove you can build something that works (mine was a fog prediction map, for example). It definitely helps get your foot in the door.

The other thing is to ask for a take-home data set. I don't know about you, but I've found that for myself and other people who don't have a formal data background, it can be intimidating to work on a data set on the spot; I just hadn't developed the muscle memory for it yet. However, I knew the right questions to ask, and I could figure out how to answer them if I had a little time, so getting a take-home set let me show what I could do that way. ■

Sdrjan Santic

Data Scientist at Feedzai | Data Science Mentor at Springboard

What is some advice you'd have for how people can ace the data science interview process? What were some of the toughest questions?

The most important thing, in my opinion, is understanding how the major supervised and unsupervised algorithms work and being able to explain them in an intuitive way. A good command of Data Science terminology is crucial. Candidates should also have a thorough knowledge of relevant accuracy metrics, as well as the various approaches to evaluation (train/test, ROC curves, cross-validation). The tougher questions would relate to these same affairs, but with having to break out the math on a whiteboard.

How did your interview process go?

Luckily, very smoothly! Most of my interviews had a feeling of being a conversation between peers, so I didn't find them very stressful. The companies I interviewed with moved very quickly (one round a week), which helped streamlined the process. I was also very impressed as to how most companies that turned me down gave me very honest feedback as to why!

What were some of the factors for you in choosing your current job?

Primarily, it was the opportunity to use a technical toolset and solve problems I hadn't solved before. My previous role was very focused on just building models. The data was already completely cleaned and pre-processed, and the exploratory work was done using a commercial GUI-based tool. I felt that my data-wrangling and command line edge was being dulled slowly and jumped at the opportunity to work in an environment where I'll be able to "get my hands dirty" once more! ■

Conclusion

The common points for success these data scientists bring to the forefront are as follows:

1) Don't think questions about basic material won't be covered. Read up on statistical fundamentals before you go through the interview process.
2) Be prepared to do well on non-technical dimensions. Companies are testing you on your communication skills and your ability to get along with future co-workers as much as they are testing you on your statistical and programming knowledge.
3) Be prepared to storytell about who you are and why your passions and skills are uniquely valuable for the company at hand. Having relevant projects and being very clear about what you contributed to those works will mark you as a candidate worthy of passing to the next round.
4) Be patient. An interview process can take a long time. You'll want to be prepared to wait.

We've provided you all that we have on the actual data science interview process. Now we have to look at what happens after you've finished interviewing.

7 Things to Do After The Interview

After you've finished your data science interview, you might think your work is finished. That's not necessarily the case. Here are a list of things you can do after the interview to ensure, as best as possible, that you maximize your chances of making the best lasting impression on your potential employers.

1- Send a follow-up thank you note

It is now customary to send a follow-up thank you note. Most recruiters now agree that it is mandatory to do so. With each office worker receiving an average of 110 emails a day, you won't want to just stick with a boilerplate "Thank you for the opportunity" email. How you follow up on an interview can make the difference between internal advocates fighting to get you in, and apathy.

Make sure you're remembered. You'll want to send an email at the very least. Candidates who take the extra step of sending handwritten notes or a list of thoughts after the interview will stand out from the rest of the average 109 emails.

2- Send them thoughts on something they brought up in the interview

One easy way to differentiate yourself is to go beyond saying thanks. Remember what has happened in the interview and make a conscious effort to tease out exactly what pain points the employer is trying to solve. If sample problems within the interview are oriented towards a technical direction, or a question notes a disconnect between different teams, you'll want to make a note of it and send in-depth thoughts on any company problems that may have surfaced during your discussion.

After all, an interview isn't just a test; it's a discussion. If you listen carefully to the questions presented and ask the right questions yourself, you will know exactly what problems the company is facing. Make sure you send them thoughts on what solutions you'd pursue.

3- Send relevant work/homework to the employer

It can be difficult seeing how your different skills apply to the office, especially for somebody who has just met you. The sharpest hiring organizations will often give you a sample problem to solve that is sourced from some real issue they are facing right now. This gives you the chance to demonstrate how your efforts can impact the business in a positive manner.

Organizations that don't do that will hesitate to hire the right candidate because they haven't sufficiently demonstrated how they'd drive impact for the company in question. However, you can be proactive and use what you learned in the interview to follow up. You don't have to stop at sending them thoughts that show you listened carefully; you can give them actual, tangible solutions

The author of this post on Forbes was told that they didn't have enough of a portfolio to get a job as a freelance copywriter. After the interview, the hiring manager told them that they liked the spirit the candidate had, but were hesitant due to a lack of a portfolio. Having listened carefully throughout the interview, the candidate knew that a major project (the re-design of a website) was just over the horizon.

Instead of accepting defeat, the candidate sent ten proposed headlines for the website banner, free of charge. This burst of initiative got her the job of doing the rest of the writing for the website--and the attention of a very busy employer.

You need to have a portfolio that shows the impact you can make, but sometimes that isn't enough. If you're astute and you ask the right questions, you can find a major data problem for the company. There always is something--that's why they're hiring for the first place! There's a data project out there that everybody would love to see done or a thorny problem that no one can figure out.

Send them a plan for what you'd do or play with some of the data they've divulged, and give some solid insights into how you work. Proactive initiative will go a long way to getting you an offer.

4- Keep in touch, the right way

One of the most awkward parts of the post-interview process is waiting for a response. You don't want to come off as desperate by following up too many times, but companies take their time if you don't engage with them proactively.

It is possible to effect the post-interview decision from outside of the company, but you should keep in mind the appropriate channel to reach somebody. Make sure to ask before the interview ends how best to reach your interviewer. Everybody has a preferred mode of communication; if they specify short emails or to check in once in awhile in person, follow that rule and dispel some of the post-interview awkwardness.

5- Leverage connections

You should have come in with strong references both from external and internal sources. If you had been building your network and providing value to them, you should have strong advocates that can support your candidacy. Check in with people who have referred you internally every once in awhile, and if needed, get them to advance how excited you would be to work at the company and how lucky the company would be to hire you.

Hiring is often network-driven, and the strongest signal you can send to a potential employer is a strong network of people who are willing to go to bat for you.

6- Accept any rejection with professionalism

No matter what, you're often going to get rejected. Sometimes, you're not right for the role, or they might have found somebody who is a better fit. It's important at this point to maintain your composure, thank the employer for their time, and move on.

People in the industry talk amongst each other, and being unprofessional at this point will only be bad karma and might get you ignored at other companies. Being professional ensures the health of your network.More importantly, a no isn't always a no. Sometimes, companies do keep your profile on file--and they will reach out for a job that is the perfect fit for you.

Perhaps Winston Churchill put it best when he said "Success is the ability to go from one failure to another with no loss of enthusiasm."

J.K Rowling, the author of the popular Harry Potter series, shared her rejection letters from publishers. Brian Chesky, the founder of AirBnB (now valued at more than 10 billion dollars) published seven rejection letters from potential investors. In order to achieve greatness, you will have to endure rejection. Everybody successful already has.

7- Keep up hope

The interview process can be one of great anxiety. Your future can be mapped out by deciding what company you can work for. An interview can mean the beginning of a career change. It can mean moving cities. It is a period in our lives where other people have a disproportionate control over our destinies.

Nevertheless, as seen in the previous steps, you control a lot more than you think. It's important to keep your head up and do what you can. The most important thing you can do during the interview process **is to keep up hope**. Interviews are lengthy. Companies take time to get back to you. There are lengthy internal checks and processes before a candidate gets accepted. You may go through multiple rounds of interviews with the same company and not seem any closer to a final offer.

You have to set expectations. DJ Patil, the Chief Data Scientist of the United States (a position created for him by President Obama) took six months to transition out of academia to a job in the industry. **You should never be disheartened during your own journey.**

The Offer Process

Your goal is to get as many interesting offers as possible that you can evaluate and negotiate. While the process itself is difficult, and may take longer than you could expect, once you start getting offers, you'll have earned them.

It's key to emphasize how important it is to manage your expectations and keep your hope up. Several of the data scientists we interviewed talked about months to half-a-year of waiting to transfer from an advanced degree from a prestigious school to a secure job. A lot of them had to take entry-level positions to get their foot in the door.

You might have heard a lot of great things about data science, but you'll only experience that with a lot of hard work and waiting.

Make sure you weigh what is presented to you and choose the future you deserve once you've spent all the hard work earning it.

Handling Offers

If you finish a process successfully, you might have one offer or multiple offers. Congratulations!

Accepting an offer is a commitment of significant amounts of your time to the company in question. Always keep that in consideration. There are several factors you can use to ascertain whether or not an offer is the right one for you.

Company Culture

This might be one of the most important factors in determining when an offer is one you should accept. Make sure you ask about the kind of company culture you're going to be a part of. Look for signs that the company has individuals that genuinely enjoy spending time with one another, and run away from generic descriptors and companies that struggle to define their culture or wave that question away. Great companies invest tons of time and effort into making sure they have awesome people who love what they do. That'll come off in your questioning.

You should also check external and objective sources such as company reviews on Glassdoor. Approach current employees as well as former ones that you can find on LinkedIn to get their side of the story. You'll often find candid tales that can give you a good preview of working at your new job would be like.

Team

Company culture is an extension of the team that inhabits it, but you should be excited about coming to work every day and working with everybody else. Make sure that you're working with a team that you can learn from. You are the combination of the five people you spend the most time with, and you're going to be spending a lot of time with your office team.

Location

Make sure that you're comfortable where the company is located, especially if you're moving significant distances. You can't move without great difficulty, and it's important that you feel at ease with where you live. Matters like the weather and the transit system matter to a certain degree, especially if you're going to live with those conditions for years.

Negotiating Your Salary

An astonishing 18% of people never negotiate their salary, despite the fact that those who do typically see their salary raised by 7%.

When you first get your offer, you're at an unique leverage point that you might not see again for several years. This is the time to test what you're worth. Reach out with an offer--a company won't fire you or cancel a contract offer because you were asserting your worth. Initial offers are sent with a buffer for slight negotiation. Take advantage.

During a salary negotiation,

1) Come with a well-researched number for what you want. Look to industry averages, and get a sense from people working in the field what you should expect. Never come into a negotiation without k
2) Knowing what you want out of it.
3) Stay positive and don't push hard for what you "deserve." Instead, use this as a positive experience to assert your worth and the value you can create.
4) Negotiate a little bit higher than what you think you'll actually get. Anybody experienced at negotiation will come back to you with a counter-offer, and you'd best be prepared for it.
5) Most importantly, don't fear rejection! So long as you keep the process moving forward civilly and professionally, a company will appreciate you being frank and positive at what is often the most difficult part of the recruitment process for them.

Before you accept the offer, make sure you know how committed you are to the company, team, and money.

Facts and Figures

Negotiation is always easier if you have some average salaries to ground you. If you have specific offers to propose, you'll be stronger at the negotiation table.

Here are some facts and figures that can start your research.

Indeed.com cites an average salary of $65,000 for data analysts, an average salary of $100,000 for data engineers, and an average salary of $115,000 for data scientists. This varies from region to region, with the highest salaries tending to cluster in the tech-heavy Bay Area. California has the highest range and median of all regions when it comes data science according to O'Reilly Media. Globally, the United States has the highest median and range of data science salaries, while the UK, New Zealand, Australia, and Canada aren't far behind. Asia and Africa tend to have the lowest medians.

The highest paying industries tend to be technology and social networking companies, while the lowest paying ones tend to be education and non-for-profit sectors.

This salary also varies based on skills and tools used. O'Reilly has a definitive survey of hundreds of respondents in the industry. An open study, the results indicate a variety of different factors that lead to different average salaries. Just as an example, people who use the Scala language extensively, a specialized type of programming, receive above $100,000 in median salary, while those who use SPSS, a proprietary tool, earn significantly less.

Taking the Offer to the Best First Day

If you've accepted an offer, congratulations! You've accomplished the goal of this whole process and broken into the job you've sought, a job that promises good compensation and the ability to drive significant social impact.

You'll have to keep that momentum going forward if you want to learn as much as you can. Be aware that companies will work to make you as comfortable as possible. You should reach out to future teammates and figure out who they are and how you can help with their problems at work. Take the time to socialize and meet as many people as you can.

More importantly, if you have time between when you accepted the offer and when you start, relax and enjoy! Make sure you catch up with as many people as you can in your life, take the chance to rest, and be completely refreshed for your first day at work.

Conclusion

The data science interview process is one of the hardest recruitment processes to crack, and it's one of the most competitive. Your fellow interviewees will be advanced degree holders, and some of them will have extensive experience in data science.

While the field is attracting many talented people, remember that it has a slew of different industries, challenges, and teams to work with. If you think outside of the box and apply a few battle-tested tactics, you'll be able to get an interview and take it all the way to an offer you love.

Split the process into its composite steps, and remember what it takes to succeed. Don't search for jobs like everybody else by applying to the standard job posts and sending out forlorn cover letters. Be innovative and solve company problems proactively. Reach out to people within the organization for information interviews. Do something different from the hundreds of other candidates, and stand out as a great technical thinker and, above all else, a proficient communicator.

Go through the technical and non-technical parts of the data science interview. Once you've mastered the thinking behind the questions and what hiring managers are looking for, you'll have a good sense of how to excel throughout the process.

Finally, when you have an offer (or several) on the table, take the time to evaluate them with good judgement. Take the time after you accept an offer to relax, skill up, and bring the momentum forward to **your first day at a data science job.**

Final Thoughts

"Most of the world will make decisions by either guessing or using their gut. They will be either lucky or wrong."- *Suhail Doshi, CEO, Mixpanel*

"The whole enterprise of teaching managers is steeped in the ethic of data-driven analytical support. The problem is, the data is only available about the past. So the way we've taught managers to make decisions and consultants to analyze problems condemns them to taking action when it's too late."- *Clayton M. Christensen, management professor at Harvard*

"We're entering a new world in which data may be more important than software."- *Tim O'Reilly, Founder, O'Reilly Media*

"Web users ultimately want to get at data quickly and easily. They don't care as much about attractive sites and pretty design."- *Tim Berners-Lee*

"Data scientists are involved with gathering data, massaging it into a tractable form, making it tell its story, and presenting that story to others." – *Mike Loukides, VP, O'Reilly Media*

Checklist

1) Map out the role your skills fit
2) Map out the industries and types of companies you want to work for
3) Prepare your LinkedIn, CV, and email templates
4) Research each company and role you want to aim for thoroughly
5) Reach out proactively to individuals within companies with informational interviews
6) Build strong networks and referrals
7) Tackle the data science interview
8) Keep up hope

9) Negotiate your offer

Templates

Getting an informational interview

Hi [first name],

I was super interested in the problems AirBnB is facing in data science. I've been aspiring to break into the field, and being a passionate follower of the AirBnB Nerds blog, I noticed that building trust with data is an important part of what drives AirBnB. Based on my background in psychology and statistics, I might be able to help come up with some creative ideas on how to foster trust. I'd love to take you out to coffee and get a greater sense of what problems AirBnB has--perhaps I can help!

Cheers,

[your name]

[Greeting],

[Why are you interested in the company], [something the company has done that you love], [how you can help].

Reaching out to get a referral

Hi [first name],

It was great seeing you at the potluck! I've been looking around, and I'm interested in the problems Uber is facing, specifically the ones faced by data scientists on the growth team. Would you mind introducing me to the hiring manager or somebody on the team so I could see if I could help?

Cheers,

[your name]

[Greeting],

[Talk about last point of contact], [talk about interest in company and problems faced by a specific role], [ask to be introduced to hiring manager to help solve those problems]

Following up after an interview

Hi [ask how your interviewer prefers to be addressed],

It was a pleasure talking with you about Google's data science problems. I think I can help with some of the problems you've enumerated, and I look forward to the next steps in the process!

Hello [Ask your interviewer how they prefer to be addressed during the interview],

[Talk about problems you can help solve], [State that you're looking forward to next steps]

Glossary

A/B split test - An A/B split test is the golden standard of experiment design for web companies, where two groups of users are subjected to different treatments and measured to see their conversion rate to a certain goal. Optimizely, a web company dedicated to helping run A/B split tests has a good guide on the concept.

Feature - A nugget of information about an object, usually stored as a column in tabular data. If you measure and store the height, weight, and gender of an individual, you are storing three features about them.

Lifetime Value - The expected amount of revenue a customer is expected to generate over the time they spend with a company. A software as a service startup that sells software by the month can expect to calculate this by multiplying the monthly price with the number of months spent.

MapReduce - A set of algorithms that act to abstract away the difficulty of storing massive data sets by treating data split into multiple servers in a way as intuitive as handling it from one. MapReduce uses parallel, distributed logic to deal with massive datasets.

Overfitting - The tendency of a model to fit onto past data, overgeneralizing from those insights to make inaccurate predictions in the future, dragged down by the weight of the past.

Type I Error - A false positive is the incorrect acceptance that something is happening akin to telling a man that he is pregnant. In technical terms, it is the incorrect rejection of the null hypothesis.

Type II Error - A false negative is the incorrect acceptance that something isn't happening. It is akin to telling a pregnant woman she isn't pregnant. In technical terms, it is the incorrect acceptance of the null hypothesis.

For more glossary terms, consult this data science glossary.

Resources

A parody of the interview process that examines some hard truths from KDNuggets.

This book, called Data Science Interviews Exposed, offers more sample questions that you can tackle with your interview practice.

The Data Science Handbook offers real-life advice from data scientists, including some smart analysis on what makes for a great data scientist and what happens during the interview process to find those individuals. Its companion, the Data Science Interview Guide, offers 120 questions you might see in a data science interview.

Cracking the Coding Interview is a definitive resource for going through software engineering interviews and will help with the programming parts of the data science interview.

This Quora thread goes into how AirBnB hires for data scientists, an insightful look at the data science interview process from an established data science leader.

This expose by Trey Causey explains how to ace the data science interview process and offers a critical and unvarnished look on how one should approach the data science interview. Erin Shellman also talks about her experience getting a job in data science.

"As I've gotten older and more experienced, I push back in interviews. I ask questions about what the purpose of a problem is or state that I don't think this is a good evaluation of my skills or abilities. Some people probably see this as me thinking I'm "too good" to answer the questions everyone else has to answer, but I see it as doing my part to be a critical thinker about evaluation, prediction, and hiring. Hopefully you'll do this too and as more of us are in a position where we are building teams and hiring, we'll think more carefully about what we're trying to accomplish and how we can get there instead of copying the same patterns that have been around for years."

This article is an insightful read about how data science at Twitter works and offers the inside perspective of somebody who is a data scientist in the industry.

If you find yourself thinking about probability, refer to this cheatsheet to make sure you're on top of any problem. This Quora thread will help as well.

Ellen Chisa writes about things she has screwed up on when it comes to technical interviews; you should make sure to avoid those mistakes!

Finally, First Round Review has a primer on how to hire exceptional data scientists; read it to know how the people on the other side of the table think.

About the Co-Authors

Roger has always been inspired to learn more. He broke into a career in data by analyzing $700m worth of sales for a major pharmaceutical company. He has written for Entrepreneur, TechCrunch, The Next Web, VentureBeat, and Techvibes.

For this guide, he compiled insight from Springboard's network of hundreds of data science experts, including Sri Kanajan, his co-author.

Sri Kanajan is currently a senior data scientist in New York City at a major investment bank. He has 14 years of experience in various engineering and management capacities and made a career transition to be a data scientist in 2013. He completed a full time data science bootcamp in San Francisco and progressed to become a data scientist at two startups and eventually a data science manager at Change.org before taking on his current role. Sri also teaches part time as a lead instructor in General Assembly's Data Science course. He is passionate about helping others make the transition into data science.