How to Survive the Data Science Internship Search Grind

By: Austin Chen and Kelly Du

As the majority of Data Science students know, there are a plethora of internship "guides" on the internet, ranging from summarizing the key skills required to succeed at various stages of the interview to where to actually seek out these applications. We are not attempting to replace these existing guides, and we are certainly not claiming to have made a "better mousetrap"; in fact, we encourage all Data Science Internship seekers to make use of all of the resources available online, since in most cases, more information is better. However, what we *will* do, is consolidate the recruiting experiences of two current students in this field (Austin and Kelly), and create an up-to-date, personalized, and battle-tested survival guide filled with tips and tricks we found useful during our own internship searches.

Places to Look:

Overview:

There are countless companies that hire Data Science Interns for either the Fall, Winter, Spring, or Summer sessions. However, despite this, most students struggle to find the actual locations of these application forms, and will therefore, potentially end the internship cycle without getting any responses (primarily just due to the lack of quantity). This section attempts to provide a starting point for anyone looking for a Data Science Internship (recruitment for Summer 2023 is still underway!).

*Disclaimer: Applying for internships is time consuming, and the application forms are oftentimes frustrating, but it will all be worth it in the end!

University of Pittsburgh - Computer Science Club (CSC) Github
https://github.com/pittcsc/Summer2023-Internships

From personal experience, this internship list maintained by the Pitt Computer Science Club is extremely top notch, with technology, software, computer science, quantitative, and data science internships updated regularly (anyone can contribute by submitting a pull request!). In fact, the Pitt CSC creates a new list for every single internship cycle (Summer 2022, Summer 2023 and beyond). The repository has information such as the company name, location of the position, other notes about the internship, and for the majority of the time, a direct link that takes you to the application website. Even for the companies without the direct link, at the very least, it can serve as a concrete list for you to conduct a "Company ABC internships" Google search and find the positions yourself. In conclusion, this list of 150+ positions is a great place to start applying, and if you apply to every single company on this repository, you will be in a fantastic spot for the year (in terms of quantity)!

Levels.fyi
https://www.levels.fyi/internships/

On a daily basis, this is probably the website I consult the most, as it not only contains the different companies that offer internships (ranging from Data Scientist to Technical Program Manager), but also the hourly salary for each position and for the majority of the time, a direct link to the application form. Like the Pitt CSC Github, even if the direct link does not exist, it can still be extremely useful as a concrete list to work off of. It is important to note, however, that all of the hourly salaries for these internships should be taken with a grain of salt, as they have been known to be inaccurate at times. Besides providing useful information for the internship search, levels.fyi also has valuable metrics on each company's full time offers, with different salaries grouped by position, years of experience, and location. This provides a means of evaluating potential internship offers (taking into account return offers), and even for new grad positions during our next cycle.

Linkedin
https://www.linkedin.com/jobs/

*If you do not yet have a Linkedin account, we advise you to create one now!

At this stage of your academic careers, I am sure the majority of you have heard of (or even used) Linkedin. Linkedin is not just a networking tool, it can be a wonderful tool to find different jobs to apply to as well. In fact, I make it a daily habit to check Linkedin in order to find out (as quickly as possible), if there are any new opportunities available. If you navigate to "Jobs", you are able to search for positions based on title, skill, company, or even location; and after doing this frequently, the Linkedin algorithm will actually recommend new positions based on your previous searches!

Glassdoor
https://www.glassdoor.com/

While Glassdoor is certainly a great resource to search and apply for jobs online, I find that the most useful feature of Glassdoor is not its ability to find positions of interest, but its information on past interview processes, questions, and reviews of each company. Before I begin the process for each company, I always conduct a "Company ABC Data Science Intern Interview Questions" Google search, and usually I will find a Glassdoor page on the number of rounds, the type of interview (behavioral, case, or technical), and sometimes even the actual questions asked. Not knowing what to expect is typically the largest stress factor for interviewees, so by knowing a little more about the interview process, hopefully you will perform better!

I have listed four different resources that I use frequently during my internship search, and like I have mentioned before, this is by no means an exhaustive list. However, I do hope that by utilizing one or more of these links, finding these positions becomes less daunting.

* You now know everything that I know about finding internships. Good luck!

Interview Process:

Overview:

It is incredibly difficult to generalize the interview process in a couple of sentences, as every company has their own way of evaluating their candidates. However, despite some differences in the number, the type, the duration, or the content of the interviews, some commonalities across all companies can still be realized. In this section, we will attempt to describe all common types of interviews (and any from personal experience), as well as provide resources to prepare for each one.

Behavioral Interviews:

During the interview process, behavioral questions are a very effective and common way for recruiters to gauge the competency of a candidate, which primarily focuses on the candidates' past experiences. This is a great chance to show recruiters, team members, and hiring managers specific strengths related to your skills, abilities, and knowledge. Here are some useful techniques to master these types of interviews.

Elevator Pitch

You will always be asked "tell me about yourself" in a behavioral interview (it is actually how most interviews start!). Talking about who you are and what you do may seem simple, but without practice or advanced thought, it can cause much anxiety. If possible, you should always prepare some talking points about yourself, including but not limited to, your past experiences and your future goals. An elevator pitch is particularly helpful to make your introductions simple and effective. Here is a quick Youtube tutorial that might help you understand how to create the perfect elevator pitch.

https://www.youtube.com/watch?v=Qncmc-yx3gI

This video walks you through all the knowledge you need to create a powerful story about who you are and why you are interested in this job in under two minutes. In this video, you will learn that there are four steps to give a compelling elevator pitch.

1. Start by introducing yourself. For instance, we can say something like "Hi, I'm xxx, a first-year graduate student in data science at Columbia University."

2. Provide a brief summary of what you do. This may incorporate your recent internship, and how you could add value to the company you are interviewing for.
3. Follow with an explanation of your goals and needs. This may include the reason you are interested in this job and what you expect to gain from it.
4. Finish with a call to action. End the conversation by asking what you want to do in the future to keep the conversation going.

S.T.A.R

The S.T.A.R method is the easiest and most effective way to answer behavioral questions.

- **Situation**: Describe the situation you are in or the work you need to complete. A specific event or circumstance needs to be described, instead of a broad overview of your prior experiences. The situation can be drawn from a former position, volunteer work, or any other relevant experience. Make sure to provide sufficient detail.
- **Task**: Describe your responsibilities toward the situation.
- **Action**: Describe the actions you took to complete the task. You should keep this part focused on yourself.
- **Result**: Describe the result or outcome of your actions and make sure to be as detailed as possible.

Now you have a general idea of what the STAR method is. Here are some useful tips to successfully apply this method to interviews:

- Always have succinct summaries of each situation ready, and be prepared to elaborate.
- Focus on yourself when describing any relevant experiences, even though you might work in teams or engage in collaborations.
- Make sure to give multiple positive and concrete outcomes. These outcomes could come from your recent projects or research. If you are asked about a challenging situation you have had at work, it's your time to demonstrate your capacity to perform under pressure. Tell the recruiter what you have discovered and how you overcame and would avoid a repeat of the situation.
- Prepare specific anecdotes that you can use to illustrate how you overcame a challenge or how you collaborated with a team.
- Know yourself and your resume. Think about your personal strengths and weaknesses, and construct answers that reflect your competency as a candidate.

* Although the S.T.A.R method is a recommended blueprint for answering behavioral style questions, it is important to not be robotic in your answers. In my own personal experience, making the behavioral portion as colloquial as possible, and just engaging as if it were a natural conversation with the interviewer has done wonders for me!

Case Interviews:
"A case interview is a job interview where the candidate is asked to solve a business problem." This is a great chance to show your critical and analytical thinking. In data science interviews, most case questions focus on specific products and companies. Thus, researching current projects across several departments about the industry can be helpful (although most interviewers will provide all the information you need during the interview).

There are four main types of data science case studies:

1. Product case studies: These types of questions often tend to test your understanding of a product and the capacity to recognize the metric that ought to be suggested to comprehend a product.
2. Data analytics case studies: These types of questions often ask you to explore analytical problems in different metrics. Sometimes you have to write SQL queries in addition to proposing the metrics.
3. Modeling and machine learning case studies: These types of questions often test your ability to use machine learning to build models to solve problems.
4. Business case questions: These types of questions are often associated with the company that is interviewing you. You might be asked to address problems that relate to the business. For example, a banking company might ask about FICO credit scores.

Here are some resources to help you prepare for case interviews:
https://www.interviewquery.com/p/data-science-case-study-interview-questions

Interview query gives you an overview of the case study and also 19 sample questions with solutions. I personally found it extremely useful to follow the four-step framework that was mentioned in the website to form answers to a case question.

https://stellarpeers.com/
https://www.interviews.tech/

Stellarpeers and interview.tech are platforms to practice product management mock interviews. You can get interview preparation and coaching from leaders. It's a good way to practice before your actual interview.

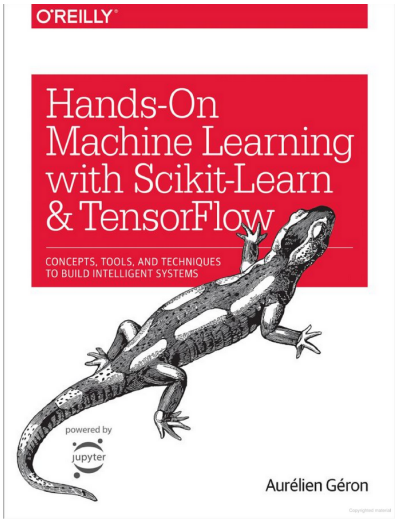
* Most case study interviews will not have a concrete right or wrong answer, instead, the interviewer is looking to gauge how a candidate approaches a new problem, how a candidate logically breaks the issue down into workable components, and finally, if a candidate can apply data science concepts to a real world problem. Do not be nervous! Always explain your thought process, and feel free to think a bit outside the box too! And as always, treat the interview as nothing more than an intellectual conversation between two data science enthusiasts.

Technical Interviews:
For most (at least for me), this is the most dreaded stage of the interview process, as you will have to demonstrate your technical competence in regards to the position that you are applying for. However, if you are able to master this stage (assuming all other rounds go smoothly), you will be on your way to your first internship offer! Here is a list of topics you should be familiar with and the resources to prepare for each one.

Machine learning:
There are many topics in machine learning, but here are 10 major ones that we should master.

a. Linear Regression
b. Logistic Regression
c. Regularization
d. K Nearest Neighbors
e. Gradient Descent
f. Decision Trees
g. Bagging and Boosting
h. Model Assessment and Model Selection
i. Feature selection or Feature reduction
j. Clustering

To study these topics, I personally recommend a book called *"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow"* by Geron Aurelien.
https://millengustavo.github.io/handson-ml/



At this stage, I assume we all have experience with Python (hopefully!). This book provides further guidance on comprehending the theories and resources about machine learning and intelligence systems.

Statistics

Data science uses computation to express mathematical concepts. You will need to have a firm understanding of the basic mathematical and statistical concepts. Here are some major concepts that are frequently asked during interviews.

a. Probability, random variables, and expectations
b. Key probability distributions
c. Parameter estimation and confidence intervals
d. Hypothesis testing
e. Bayesian statistics

https://github.com/kojino/120-Data-Science-Interview-Questions

This link above provides frequently asked questions in data science interviews, including communication, data analysis, modeling, probability, product metrics, programming, and statistical inference. If time permits, going through the 120 questions section by section before your interview will help you refresh and develop your statistical skills for data science.

Live Coding

Not all jobs will require a programming background, but it is still crucial to brush up on some key concepts, when needed. For a data science job, three languages are commonly used: Python, R, and SQL.

- Python

There are various libraries in python, but these four are the most used and important in the field of data science: Pandas, Numpy, Sci-kit Learn, Matplotlib.

- Pandas provides data structures for Python. It's an effective tool to analyze mid-size data sets.
    - Resources to learn pandas
        - This link provides you a short introduction of pandas: https://pandas.pydata.org/docs/user_guide/10min.html
        - This link gives you a more detailed tutorial on pandas: http://www.gregreda.com/2013/10/26/intro-to-pandas-data-structures/
        - In addition to understanding pandas, this link teaches you how to plot with pandas: https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html
- Numpy provides a simple data structure for data analysis. If you are still not quite familiar with numpy, here is a link to a detailed tutorial: https://realpython.com/numpy-tutorial/

- Sci-kit Learn is the most useful library for machine learning in python. As mentioned above, The book *"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow"* has everything you need to know about machine learning as well as scikit-learn in detail. I personally strongly recommend you walk through this book and do the exercises.
- Matplotlib is a plotting library for visualization in python. Here is a link to a basic tutorial for matplotlib that might help you refresh your memory a little bit. https://github.com/rougier/matplotlib-tutorial

- R

In exploratory data analysis, R is often used for data visualization and transformation. I assume that we all have basic knowledge about R (as we should), here is a link to the cheat sheets about common used packages in R:
https://posit.co/resources/cheatsheets/

- SQL

Structured Query Language, is a specific language used for managing data held in relational database management systems. Here are some links to get you started with SQL.

a. https://www.w3schools.com/sql/
b. https://www.sqltutorial.org/sql-cheat-sheet/

* The above three languages encompass what I like to call the "data sciencey" coding questions, as the problems will likely be actual issues we have encountered in our curriculum. Contrary to what you might expect, you know more than you think! Stay calm, stay composed, and ace this section!

Data Structures and Algorithms

Unlike the previous section on "data sciencey" coding problems, this section is what I struggle with the most (and a lot of other data scientists echo this same sentiment). As data scientists, we do not encounter DS&A problems on a daily basis, and many different data structures like linked lists or hashmaps have been forgotten. But unfortunately, these "LeetCode style" questions are still quite common for technical interviews (especially for tech companies), so it is best we embrace the LC (LeetCode) grind!

LeetCode
https://leetcode.com/

LeetCode is a free platform that anyone can use to practice and enhance their DS&A skills. There are literally thousands of LC questions on the platform, including problems on arrays, linked lists, trees, strings, dynamic programming, and graphs. It is almost impossible to complete all of them, that is why I personally recommend the "Blind 75" list. This list includes

the most common 75 questions that are asked in LC style interviews. If you are looking to get more practice, that is certainly a great place to start!

* LeetCode questions are divided into three levels of difficulty (Easy, Medium, and Hard), and luckily, for data science internships, LC Hards are rarely asked.

LeetCode is a great platform to study and prepare for interviews. I have been using it for several months, and I think it is the most effective tool to improve my DS&A skills. If you are new to LeetCode, I recommend that you start with the easiest question (skip it if it is too simple). I certainly advise against starting with questions that are too challenging, as you may waste too much time trying to come up with a solution or simply give up. Remember, consistency is the key! Doing it everyday for a few hours is better than doing it once a week. When coming up with solutions, feel free to start with brute force methods, and try to pinpoint the inefficiencies and optimize them.

* If you can't solve the problem or have already spent too much time, it is perfectly acceptable to look at the solution. Just make sure you take the time to understand it!

Online Assessments (OAs)

Online assessments are a great way for a company to gauge a candidate's competency, without ever having to spend the time to interview them. There are commonly three different types of OAs: time coding assessments through platforms like HackerRank and CodeSignal, recorded interviews through HireVue or SparkHire, or a take home data science project

HackerRank/CodeSignal

The problems a candidate may encounter on these platforms will differ significantly from company to company, and even from position to position. However, here are some general facts:

- These assessments will be timed (ranging from 60 minutes to 180 minutes)
- They often include LeetCode style questions and can include all aspects of coding (Exploratory Data Analysis, Machine Learning, …etc)
- Only the successful completion of the assessment (the criteria for "passing" will differ) can move a candidate to the next round (live interviews)
- Use the same resources outlined in the Case and Technical sections of this survival guide to prepare for this!

HireVue/SparkHire

These talent experience platforms are designed to make the hiring process easier for employers, and can also provide a fair opportunity for every candidate (as the same questions are usually asked). Here are some general facts:

- Behavioral style questions are most commonly asked on these platforms, but the occasional coding/multiple choice conceptual question does show up.
- The candidate will be presented with a series of questions, and will have a certain time limit (ranging from seconds to minutes) to think about the response before the recording begins. Then, they will again have a time limit to answer the question before the recording ends.
- Some companies will allow one or two additional attempts, but most of the time, candidates will have one chance to record something that they are satisfied with.
- The key here is to formulate the response before the actual recording begins (write some talking points on paper if necessary).
- Talking to a webcam can be uncomfortable (practice!), but try to maintain eye contact as best as possible, and be as conversational as you can.
- Once you get used to this format of interview, it will be an absolute breeze!

Take Home Project

A take-home project is a coding task that is given to the candidates to complete in a given time frame (ranging from hours to days). Think of this as a case study question that candidates actually code up. Here are some general facts:

- A company will typically send candidates a packet containing the instructions for completing the project, as well as datasets that are to be utilized.
- For a data science internship, the most common projects will consist of exploring a dataset (conducting missing value analysis, EDAV …etc), building models to generate a prediction (machine learning), and using the results of the model to address a business objective (make a recommendation).
- To give a simple example, a banking company might send a dataset of FICO scores and the accompanying metadata for each score (customer demographics, salary, occupation …etc). The objective might be for a candidate to build a model to predict the FICO score of new customers, and assess whether or not the bank should do business with the customer (based on credit risk). In this case, the candidate would use a language of their choice and start to explore the data, build, tune, and test the models, and finally make the required recommendations.
- These projects are typically open-internet, and each candidate is encouraged to use any resource available to them (as long as the work is individually done).
- Some companies might even require a powerpoint to be created, so the candidate may present their findings at a subsequent interview round.

Different Types of Interviewers

This will obvious differ from company to company, but in my personal experience, interviewers are divided into three different categories (Human Resources/Recruiter, Team Member/Actual Data Scientist, and Hiring Manager)

Human Resources/Recruiter:

- A candidate can expect to encounter HR during the first round interview/phone screen
- These interviews are typically quite low stress, with HR asking some bookkeeping questions, followed by common behavioral questions (refer to our section on Behavioral Interviews!)
- This is also a great chance for the candidate to learn more about the company, the position, or the team that they are applying for.

Team Member/Actual Data Scientist

- A candidate can expect to encounter an actual member of the team that they are interviewing for in the second or the final round.
- These interviews are oftentimes the most technically rigorous (Case and Technical), with some behavioral questions sprinkled throughout.
- Since these interviewers are likely from the same background as the candidate, it certainly will not be as "easy" as the first round, but do not worry, you have all the tools to succeed!

Hiring Manager

- A candidate can typically expect to encounter the hiring manager at the final stage of the interview process (congratulations on getting this far!)
- A hiring manager will typically not ask any rigorous technical questions (the occasional case study is certainly possible), and the interview will focus on behavioral style questions. The manager is attempting to gauge the "fit" of the candidate with the team that they manage.

* Like I have mentioned before, these will differ depending on the company, but what I have outlined above have held true in all of the processes I have personally gone through.

Other Resources:

Reddit
https://www.reddit.com/

Reddit gets a bad reputation (and to be honest, the bulk of it may be true); however, I find that for certain computer science/data science oriented subreddits, the information can actually be extremely helpful. I will list some of them below:

r/csMajors
https://www.reddit.com/r/csMajors/

r/cscareerquestions
https://www.reddit.com/r/cscareerquestions/

r/datascience
https://www.reddit.com/r/datascience/

By simply searching the company name in any of these subreddits, a plethora of information regarding the interview process (rounds, difficulty, type …etc) can be found, and other "bonus" facts such as work-life-balance, salary (intern and full time), and a general feel for where a company may be in the intern selection process can all be realized. Even if the information you seek is not present, you are always able to ask the question yourself, and more often than not, someone will have already gone through the same process, and can provide any assistance you require. Generally, I find that if nothing else, Reddit can be a great additional resource to gain more insights on the internship search process. But like all open-contribution websites, take everything you find with a grain of salt!

Compilation of Other Tips

1. Do not be afraid to ask your recruiter for clarification on the interview format (behavioral, case, or technical). They will be more than happy to provide you with all the information they can.
2. If you have a competing offer deadline, feel free to let your recruiter know, they may be able to expedite the interview process.
3. Ask around your network for any tips on an interview, more likely than not, some of your friends will have gone through the process with the same company.
4. Conduct a mock interview with your peers or career specialists, as it will help you build your confidence and perform well.
5. Avoid dead air time. If you are not showing progress or getting an answer to a problem, try to at least walk the interviewer through your thought process. Remember to keep the conversation going. If the concept seems strange to you, try to talk about the concepts that you are familiar with and are adjacent to that concept. Interviewers will oftentimes give you hints (pay close attention!)