Common Questions:

Why do you want to change from academia to industry?

As a student, I have many opportunities in academia because there are a lot of school-affiliated research institutes. However, I have talked to many Data Scientists through LinkedIn as well as my alumni who have rich industry experience. I asked them about their career life in the industry. Then I got to know that working in the industry is what I really want.

I learned that in the industry there are opportunities to work on various projects that have a faster implementation period so people can quickly see the impact of their work. This style attracts me a lot. I want to do something useful and practical. I like to solve problems and generate enduring value. And I like to work in a fast-paced environment where I can stay focused and always learn new things.

What do you see yourself in 5 years? What's your career path? Is there anything changed in the past few years?

Well, I'm definitely really excited about the Data Scientist position at XXX, and I can see myself growing professionally in this role. I think within the next five years I would seek to make a significant impact at XXX. I'm also looking forward to taking on additional responsibilities and possibly taking the lead on some projects.

(How) I will make the best of my skills and knowledge, learn everything that's helpful to my work, and do my best on projects. Although I am a new graduate with limited experience, I believe in my potential.

(What) I'm so eager to be a professional senior data scientist, you know, professional senior data scientists are those who have expertise in analytics, tools, and techniques, take the lead on projects and are able to solve valuable business problems in a way that will generate enduring value. And I believe one day I can become one of them.

Why are you a good fit for this role? Why should we hire you over other candidates? What makes you best fit the role?

What can you bring to the role? (50%)	I have the exact mix of xx and xx skills you are looking for. And I have experience of xxx which is transferable to the new role. Beyond my technical skills, I'm a fast learner and open to learning new things from the internet or team members. And I have the ability and interest to dive deep and figure out things, I'm good at getting interesting insights from data or finding solutions to complex problems.
What you are trying to get out of it? (30%)	Most importantly, this company cares deeply about xxx, you promotes xxx. That is what I find exciting about this role. I think I would be a good match. And I believe I can grow fast in your company and make an impact.

What is your greatest strength?

I'm good at working with data and I truly love it. I've been working with data for more than 2 years since I was an undergraduate. I love to do statistical analysis or build machine learning models. I'm always feeling excited and never get bored when analyzing data and getting insights from data. Because I pay attention to details and I have the ability and interest to dive deep and figure out things, I can always get interesting findings or find solutions to complex problems.

I believe the other strength of mine is I'm eager to learn new things either from the internet, school or by talking to friends and team members. And I'm a fast learner and would like to learn new things to get the work done perfectly.

What is your greatest weakness?

S	My greatest weakness is public speaking. I'm ok with speaking in front of a small group of people but afraid of speaking in front of a large group of people, like 100 students in a big classroom.
T	I realized my weakness a few years ago when I gave a presentation to my whole class. I was nervous and not confident.
A	But I wanted to be a more confident public speaker like the speakers on the Ted Talks. So I took every opportunity to give presentations. I find that the more I practice, the more confident I will be.

R	Now I have made a big improvement. The most recent presentation in front of a big group of people was also a course project presentation. I did a great job. And I think in the long term, public speaking would not be a problem for me anymore.

Tell me about a time you made a mistake.

S	I was doing a simulation study to assess the impact of ascertainment on the cancer risk estimation in my previous internship.
T	After writing and testing the simulation code, I started to simulate the population. My goal was to simulate 2 million families. That's a large number. Each family has about 20 members, so there can be 40 million rows.
A	My strategy was to divide the task into four rounds, each round generating 0.5 million families. And I could do this in parallel which was time-saving. After generating the population and I made some statistical summaries such as the mean, std, or count of the features in the population. And I also presented the result to the team. After my presentation, I suddenly realized that I used the same seed when running the four rounds. This was incorrect because in this way the whole population would simply be a combination of four same sub-population. The diversity would be reduced. As I realized my mistake, I immediately let the team know what I did wrong with the simulation, and let them know my next step which was to rerun the code and redo the summary.
R	This mistake taught me to always pay attention to details, go through all possibilities of mistakes, and double-check the numbers and the code. It's normal to make mistakes, but I need to be careful enough to minimize the possibility.

Describe a time when you worked in a team.

S	I have worked in a large research lab where there are data scientists, statisticians, and medical doctors.
T	I have a research goal in mind and there was a new dataset that can be used to answer my research questions. My first step was to summarize and analyze it, get insight, and present new findings to my team.
A	The data were collected and cleaned by a data scientist in the first place. I set up several meetings with him asking questions about the meanings of variables and values. After getting familiar with the data, I started to do some visualization and statistical testing. I gave regular updates to the team, presenting what I found and asking questions. My mentor is a research data scientist. She gave me very useful suggestions and feedback. When I got insight from the data, I present my work to the medical doctors. I also did additional work when the medical doctors have other questions that can be answered by using this new dataset.
R	In the end the insights from the data were very helpful to inform medical doctors when they are making clinical decisions. This is when I have a strong feeling of teamwork. Everyone is working on his or her own project, but we are eager to help each other, so we make progress on projects quickly. I was feeling so good in such a team.

Describe a time when you help someone.

S	Once there was a new member coming into my team.
T	During the regular meeting she talked about her project idea and I realized the documented simulation code might be helpful for her project. I was familiar with the code but it was hard for a new member to understand and use it to do simulation.
A	So I offered to set up a meeting with her and walked her through the code. I also offered to answer any questions she had about the code and simulated data.
R	She was grateful for my help. She said my help sped up her progress on her project. I was very happy to hear that.

Give me an example of a time when you motivated others.

S	In my most recent group project, we had a fixed deadline for delivery.
T	As the delivery date was very close to the final, everyone was busy studying for the exam. The project was making slow progress.

A	I was worried about we wouldn't be able to complete the project in time, so I came up with an idea. I set up a meeting with my group members to decide on a whole day when everyone was available to do the project together. And I promised a team pizza party after completing the project.
R	On that day we worked together very efficiently and quickly completed the project. And we enjoyed the pizza together.

Show an example that you pay attention to details.

S	I remember when I did data cleaning, preprocessing, and statistics summary on a new dataset, it really required attention to detail to get it done without error.
T	First I needed to extract the subset of data for a particular project. I filtered by some conditions on one dataset and get the IDs. Then extracted the metadata from the other dataset by these IDs. I needed to make sure all the valid data were selected. The new dataset was messy in both values and structure. There were missing values and error entries. To make a statistics summary, I needed to remove duplicates, take care of error entries, collapse some of the rows, etc. Sometimes, there were abnormal values, so I need to record the IDs and look into the raw data and figure out what is going on.
A	Every step required a double check. I had all my code documented and clearly commented, so it's easy to go back and check.
R	Eventually, I completed the cleaning and summary task without mistake.

Tell me about a time you failed & what you have learned from it.

S	I was doing a research project in a lab this summer.
T	I got a dataset from the team and I told them I could finish the exploratory analysis and model building part within 3 weeks. I thought this was doable but I underestimated how messy the data was, so it ended up taking 4 weeks.
A	I offered an apology to my team. I realized that I should have been more conservative with my timeline estimation, and my team would be upset if I don't keep my promise since this would also affect their schedule.
R	I learned from this experience. So in the next few projects and also the projects I will be working on in the future, I always set a more conservative expectation of timeline, and I take less amount of time to finish my tasks than my expectation. Also, I keep more frequent updates to my team, letting them know my progress. The same failure never happens again and I will also try to avoid it in the future.

Tell me about a time when you disagreed with your boss.

S	I did have a little disagreement with my supervisor. That was when I was working in a research group on my thesis project.
T	I was summarizing a new dataset. During a weekly meeting, I talked about my statistical summaries. My supervisor looked at the numbers and found they did not match the numbers summarized when doing the data cleaning. I gave my explanations and reasons that it's possible that one subject had two events. So when counting the number of events, the count can be larger than the count of subjects. But my supervisor still didn't agree with me saying that the possibility is too small.
A	In order to show that I was correct, I pull the raw data and found the actual subjects with more than 1 event. I showed these subjects to my supervisor to convince her that the summarized numbers were correct.
R	Eventually, my supervisor and I were in agreement with the results.

What are the common themes for senior data scientists?

• [Technical skills, analytical, and critical thinking] They have strong technical skills and can take full ownership of a significant feature in the backlog. They can work with key stakeholders and subject matter experts and perform any research necessary to implement their work. They have analytical and critical thinking skills.
• [Problem-solving, either by troubleshooting/investigating the problem, or by researching/learning how to resolve it, or by collaborating with people] They can troubleshoot problem areas and investigate a solution that will fix the issue. If they are unsure of the problem, they can research and learn how to resolve the problem or collaborate with others to aid them in the effort.
• [Mentoring] They can mentor younger or less experienced individuals and teach them what they have learned. They can lead other members who are contributing to their projects and work effectively with them.
• [See the big picture, understand the value for business; see the needs of the business, design solutions for the needs, provide positive value back to the company] They can look at the "big picture" and understand where their value is for the business. They can analyze user needs, understand where the gaps are, and lead change across the company to adopt data science practices. They see the needs of the business, design solutions for those needs, and provide positive value back to the company.
• [Communication and collaboration skills; pay attention to details; good documentation for the work; work effectively with others to drive projects to closure] They demonstrate excellent communication and collaboration skills and can explain their work to both technical and non-technical audiences. They have good documentation for their work and attention to detail. They can work well with others from different teams to drive a project to closure.
• [Take responsibility for projects they lead; work with others to stay on schedule] They take responsibility for their projects and can complete the work. They are held accountable for the success or failure of the projects they lead and can work with others to stay on schedule.
• Skills: Data Mining, Python / R, Machine Learning, Leadership, Communication, Collaboration, Big Data Analytics, Cloud, Predictive Analytics, Know How to Learn, Mentor

Questions for Recruiters

1. How would you describe the company's culture?
2. Who will be interviewing me?
3. What types of interview questions shall I expect?
4. What are the next steps? How long the interview process is expected to take?

Culture Fit

Tell me about a time in which your colleagues weren't convinced by your approach you believed strongly in. How did you change your approach to influencing them?

S	In a previous data science project, we had a goal of predicting future COVID-19 cases using Google Search data.
T	One of my teammates suggested first defining outbreak points and then building a binary classification model. I suggested considering the problem as a time series and I proposed building a recurrent model to predict the cases. He didn't agree with my method. He said it won't work because there was no reasonable relationship between the feature and the number of cases.
A	I understood why he didn't agree. I explained to him the differences between time series forecasting and traditional ML prediction and how RNN works. I also found a relevant research paper that had a similar problem as ours and addressed it by time-series modeling. And my explanation showed that the method I proposed is feasible in theory. I also built a simple RNN and visualized the forecasting results to show that the model is learnable. And this demonstration showed the method I proposed is feasible.
R	As a result, he was convinced and we eventually successfully built an advanced RNN to forecast future cases.

Tell me about a time when your role in a project was unclear. How did you resolve the issue?

S	This usually happens when I was working with team members on coursework or group project. We work together but don't have a specific role. This can be resolved after the first group meeting.
T	For instance, I worked with 2 students to complete a data analysis project.
A	In the first meeting I introduced myself regarding my background and my strength. I showed that I'm confident in my data analysis skills and would like to take responsibility for processing and analyzing data. I brainstormed various approaches that I know well to solve the problem. And I would like to try on those approaches and present the results to the rest of the team.
R	So this is how my role was defined in a project.

Tell me about a time when you disagreed with the direction of a project. What did you do?

S	In a previous data science project for a competition.
T	We have already done data preprocessing, exploratory data analyses, feature engineering, and model building. And we wanted to achieve a higher accuracy score.
A	My teammate suggested trying more advanced models. However, I disagreed with his idea and thought this was not the correct direction at that stage because there was not much room to improve performance by changing model type or structure. Instead, I thought it was feature engineering to find good features that would help break through the bottleneck.
R	After my teammate quickly optimized the model, he joined me in finding useful features. Eventually, we found several good features that boost the model's performance.

Why should we hire you over other candidates? What makes you best fit the role?

What can you bring to the role? (50%)	I have the exact mix of xx and xx skills you are looking for. And I have experience of xxx which are transferable to the new role. Beyond my technical skills, I'm a fast learner and open to learning new things from the internet or team members. And I have the ability and interest to dive deep and figure out things, I'm good at getting interesting insights from data or finding solutions to complex problems.

What you are trying to get out of it? (30%)	Most importantly, this company cares deeply about xxx, you promote xxx. That is what I find exciting about this role. I think I would be a good match. And I believe I can grow fast in your company and make an impact.

Disagree and Commit

Describe a situation where others you were working with on a project disagreed with your ideas. What did you do?

S	In a previous data science project, we had a goal of predicting future COVID-19 cases using Google Search data.
T	One of my teammates suggested first defining outbreak points and then building a binary classification model. I suggested considering the problem as a time-series problem and proposed building a recurrent model to predict the cases. He didn't agree with my method. He said it won't work because there was no reasonable relationship between the feature and the number of cases.
A	I understood why he didn't agree. I explained to him what the differences between time series forecasting and traditional ML prediction are and how RNN works. I also found a relevant research paper that's similar to the problem we were facing with, where time series forecasting was performed. And my explanation showed that the method I proposed was feasible in theory. I also built a simple RNN and visualized the forecasting results to show that the model was learnable. And this demonstration showed the method I proposed is feasible in practice.
R	As a result, he was convinced and we eventually successfully built an advanced RNN to forecast future cases.

How to deal with team conflicts? What if people have different opinions? Tell me about a time you had a conflict at work and share example.

S	Last year, I was working with a team to do a group project for a regression analysis course.
T	We were able to come up with our own project idea as long as we used some regression techniques to get some insights from the data. After we decided on our project topic, we found many data sources but we cannot decide on which data we should pick. One of my teammates said she had a friend who might be able to provide an unpublished dataset for us. So she wanted to wait for her friend's reply. I realized that the deadline was approaching and there were many things to do after choosing a dataset. So I said we could download a dataset from a public database and get started on doing analysis before it's too late. However, she insisted to wait for her friends' dataset. I was a little bit angry about this disagreement because of the stress of the approaching deadline.
A	I understood she wanted to do a unique work with an unpublished dataset. But the problem was if we waited for too long we might not able to finish the project in time. So I let her know my concern about the time limitation. We had a good conversation and tried to find a good solution to this problem. We both agreed to work on a published dataset first, while waiting for the new dataset at the same time. When the new dataset was available we then switched the data and redo the analysis and interpretation. In this way, we didn't waste too much time.
R	As a result, we completed a high-quality project with that new dataset in time.

Adaptability

Describe an experience in which you had moving deadlines for several projects, but felt that one key project element needed extra attention. How did you balance your daily responsibilities with this new element?

S	I always plan ahead. I create a schedule for each week of all the tasks I have to do, and also a schedule for each day. I do the tasks with higher priority first.
T	One time, I had to deliver code for an assignment by the end of the week, but I felt that the code needed additional 5 hours to be fully optimized.
A	Each day I spent one more hour on code optimization and fewer minutes on other tasks.
R	Eventually I was able to deliver the code by the end of the week, and other tasks were not affected.

Tell me about a data project you have worked on where you encountered a challenging problem. How did you respond?

S	It was in my research internship where the goal is to conduct a simulation study to assess the impact of ascertainment bias on cancer risk estimation.
T	I had to simulate a population with 2 million families with demographic features like sex, current age, death age, as well as clinical features like cancer and gene mutation. I needed to include 11 different types of cancer and 9 different gene mutations. And the prevalence of cancer and the probability of getting mutations should be the same as what's in the real world. The family relationship and demographic should make sense. So it's a very large and complex dataset that required a complicated simulation code structure.
A	What I have done is to first understand what the real data was like by doing research and asking my teammates. Then I started to build from the very simple code structure and functions, and make it more and more complex to include more features. Every time I added one feature, I tested the code and checked whether there was any bug. If there was a bug, I had to fix it.
R	Eventually, I successfully completed the simulation code and the code was able to generate 2 million families that allow me to do more analyses.

Tell me about a time you failed. What did you learn from it? (Provide an example of a goal you did not meet and how you handled it.)

S	A few years ago, I was doing a research project.
T	I got a dataset from the team and I told them I could finish exploratory analysis and model building part within 3 weeks. I thought this was doable but I underestimated how messy the data was. Unfortunately, it ended up taking 4 weeks.
A	I offered an apology to my team and I learned from this experience. I think about various strategies to be more efficient and prevent delays. I learned to keep track of the project progress, identify potential problems and act quickly, keep open communication with the team, etc.
R	I learned from that experience. After that, I can always finish my task in time in the next few projects.

Invent and Simplify

Tell me about a time you invent something.

S	It was in my research internship where the goal is to conduct a simulation study to assess the impact of ascertainment bias on cancer risk estimation.
T	I had to simulate a population with 2 million families with demographic features like sex, current age, death age, as well as clinical features like cancer and gene mutation. I needed to include 11 different types of cancer and 9 different gene mutations. And the prevalence of cancer and the probability of getting mutations should be the same as what's in the real world. The family relationship and demographic should make sense. So it's a huge and complex dataset that required a complicated simulation code structure.
A	What I have done is to first understand what the real data was like by doing research and asking my teammates. Then I started to build from the very simple code structure and functions, and make it more and more complex to include more features. Every time I added one feature, I tested the code and checked whether there was any bug. If there was a bug, I had to fix it.
R	Eventually, I successfully completed the simulation code and the code was able to generate 2 million families which allows me to do more analyses.

Are Right, A Lot

The mistake/failure questions: failure, error in judgment, bad decision, or regret.

S	A few years ago, I was doing a research project.
T	I got a dataset from the team and I told them I could finish the exploratory analysis and model building part within 3 weeks. I thought this was doable but I underestimated how messy the data was. Unfortunately, it ended up taking 4 weeks.
A	I offered an apology to my team and I learned from this experience. I think about various strategies to be more efficient and prevent delays. I learned to keep track of the project progress, identify potential problems and act quickly, keep open communication with the team and ask for help with difficulties, etc.
R	I learned from that experience. After that, I can always finish my task in time in the next few projects.

The interpersonal conflict questions

S	Last year, I was working with a team to do a group project for a regression analysis course.
T	We were able to come up with our own project idea as long as we used some regression techniques to get some insights from the data. After we decided on our project topic, we found many data sources but we had a disagreement about which data we should pick. One of my teammates said she had a friend who might be able to provide an unpublished dataset for us. So she wanted to wait for her friend's reply. I realized that the deadline was approaching and there were many things to do after choosing a dataset. So I said we could download a published dataset online and get started to do analysis before it's too late. However, she insisted to wait for her friends' dataset. I was a little bit angry about this disagreement because of the stress of approaching the deadline.
A	But this disagreement didn't develop into a conflict. I understood she wanted to do a novel project with new data. I also talked to her about my concern about time limitations. We had a good conversation and tried to find a good solution to this problem. Eventually decided to talk to the lecturer to extend the deadline for us.
R	As a result, we completed a high-quality project with that new dataset.

How do you collect data, do you have good judgment, do you have good intuition, do you have any experience to guide you?

S	Feature engineering is an important part to boost model performance. Usually I have to make decisions about which feature to create so that the model could be improved.
T	In a pervious data science project, I did feature engineering by the insight came from exploratory data analysis as well as doing research.

A	There was a feature called availability 365 meaning the number of days when the house is available for booking. After plotting this feature vs the outcome, I saw that there is no linear relationship between them. And by plotting the distribution of the availability 365 I realized there were two peaks at < 12 available days and more than 350 available days. So I decided to segment the feature into three parts: less than 12 available days, more than 350 available days, and the rest. With longitude and latitude data, I was able to locate the house. And by plotting the location on the map I knew which area was more expensive, so I defined geographic areas by the overall price of houses and I created a new categorical feature indicating which area that house was located in. This is how I created new features based on EDA and I also did some research to confirm that. Doing research can not only confirm what we found in EDA, but also give some insights for other possible feature engineering ideas that we cannot come up with by EDA. For example, I found that houses close to scenery spots or famous buildings like museums were more expensive, so I can create a new feature that is the distance between the house and the closest spots or buildings.
R	By doing EDA and research in this field I was able to create many useful features and eventually the model performance has been improved a lot.

S	Because I have experience of data science projects and have knowledge learnt from courses. I know how to deal with missing data.
T	Based on the data I have, usually I make several decisions about what to do to better deal with the data missingness, either through deletion or imputation.
A	At first, I'm gonna try to check whether the missing data is MCAR, MAR, or MNAR. Although we might not be able to figure out exactly the true type, it worths trying. If the missing is because of entry errors at random, it's not a problem. If the missing data is MNAR, that's a problem because it would bias the estimate of coefficients. If I have time I would definitely look into the reason why there is missing data. And I might need sensitivity analysis to quantify the impact of missing data on the estimation. That's for statistical inference. When it comes to machine learning, if a column has too many missing values like 80%. The whole column would be useless and I might not want to keep it. But if it's just a few missing data, I can impute the data by mean or median or mode, or using multiple imputation or kNN imputation.
R	So when I deal with missing data I always follow this guideline in mind.

Dive Deep

Tell me about a time you performed an analysis that resulted in process improvements.

S	Feature engineering is an important part to boost model performance. Usually, I have to make decisions about which feature to create so that the model could be improved.
T	In a previous data science project, I did feature engineering by the insight that came from exploratory data analysis as well as doing research.
A	There was a feature called availability 365 meaning the number of days when the house is available for booking. After plotting this feature vs the outcome, I saw that there is no linear relationship between them. And by plotting the distribution of the availability 365 I realized there were two peaks at < 12 available days and more than 350 available days. So I decided to segment the feature into three parts: less than 12 available days, more than 350 available days, and the rest. With longitude and latitude data, I was able to locate the house. And by plotting the location on the map I knew which area was more expensive, so I defined geographic areas by the overall price of houses and I created a new categorical feature indicating which area that house was located in. This is how I created new features based on EDA and I also did some research to confirm that. Doing research can not only confirm what we found in EDA, but also give some insights for other possible feature engineering ideas that we cannot come up with by EDA. For example, I found that houses close to scenery spots or famous buildings like museums were more expensive, so I can create a new feature that is the distance between the house and the closest spots or buildings.
R	By doing EDA and research in this field I was able to create many useful features and eventually the model performance has been improved a lot.

Learn and be Curious

How do you stay inspired, acquire new knowledge, or innovate in your work?

So far as a student I acquire knowledge from courses. When doing a research project in the research institute, I always read papers to get novel ideas to solve problems.

But if I work in a company I will definitely talk to colleagues often so I can get insights in the field I will be working in, and I will also read articles and blogs more actively and get new novel ideas for the project I would be working on.
I believe if I work in a company, I would be exposed to the company business, so it's easier to identify and understand the problem. I would gather all the information and learn all the things from various sources to solve the problem.

Tell me about a data professional you have worked with whom you really admired. What specifically do you admire about them?

My mentor is a research data scientist at Dana-Farber Cancer Institute.
I really admired her because she is always skeptical of team members' hypotheses, models, and findings. She is able to notice any small abnormal numbers or patterns and asks us to figure out the reason or double-check whether there is any error.
When I go back to find reasons or double-check the data and code, sometimes I can get more interesting findings, and sometimes I find there is a bug in the code. Sometimes I notice there is an unsolvable problem which then becomes a limitation of my work. So this gives me opportunities to be creative to find solutions.
So I realized that it's important to be skeptical about the data I'm using and the work I'm doing, I should be used to double-check my work, and should constantly ask "how am I being fooled by the result and how can I test that".
This is why I admire my mentor.

Good candidates for jobs are those who balance creativity and skepticism.

A good candidate can learn on the job to fill in knowledge deficiencies and there is a large gap between knowing facts and the ability to apply these effectively. A better characterization of a data scientist's innate potential is the way they balance creativity and skepticism.

Creative individuals are capable of looking at problems in novel ways and envisioning solutions where others have failed. Data science is the art of looking at data, hearing a business objective, and then envisioning all the steps, features, and validations needed to get from data to success. For example, feature engineering constantly requires researchers to creatively design novel representations of the data that bring out key insights and patterns.
Creativity not only allows data scientists to explore multiple possible solutions, but to envision pitfalls and risks of each, thereby preventing costly tangents and dead-end solutions.

Skepticism, seeing all the confounders, conflicting hypotheses, and potential sources of error within the pipeline and code base. Trust is the biggest trap of modern data science; data scientists should constantly ask, "How am I being fooled by this result and how can I test that?"

Insist on the Highest Standards / Dive Deep

Tell me about a time you wouldn't compromise on achieving a great outcome when others felt something was already good enough.

S	In my previous research project I wanted to make the project more complete so after I finished the initial task I continued the analyses and found more interesting findings.
T	Before the project started, the plan was to report associations between mutation and cancer and quantify the impact of ascertainment bias on risk estimation.
A	After finishing this part, I read some research papers in this field and found that often the risks are estimated by age group. So I extended the project and did an additional regression analysis grouped by age. I found an interesting finding that the associations are different by age as well as the impact of ascertainment.
R	I present these interesting findings to the team and got positive feedback.

Think Big

Give an example of how you set goals

S	It was in my research internship where the goal is to conduct a simulation study to assess the impact of ascertainment bias on cancer risk estimation.
T	I had to simulate a population with 2 million families with demographic features like sex, current age, death age, and clinical features like cancer and gene mutation. I needed to include 11 different types of cancer and 9 different gene mutations. And the prevalence of cancer and the probability of getting mutations should be the same as what's in the real world. The family relationship and demographic should make sense. So it's a huge and complex dataset that required a complicated simulation code structure.
A	What I have done is to first understand what the real data was like by doing research and asking my teammates. Then I started to build from the very simple code structure and functions, and make it more and more complex to include more features. Every time I added one feature, I tested the code and checked whether there was any bug. If there was a bug, I had to fix it.
R	Eventually, I successfully completed the simulation code and the code was able to generate 2 million families which allows me to do more analyses.

Ownership

Have you gone above and beyond the call of duty? If so, how?

S	I once worked on a DS team project. We planned to go through every step in the DS project lifecycle. So we divided the work and we worked on different tasks.
T	I noticed that one of my teammates was struggling with his tasks. To some extent, my work was dependent on his work. If he delivered late, I would have less time on my part and probably also delivered late.
A	I approached him discreetly over lunch and offered to stay late and help him out. We worked through the night.
R	Eventually, he was able to deliver in time and so was I.