CADS Speaker Panel Responses

**Anu Miller:** Director of Data Science at Metlife

**General Questions**

What is an example of a high-impact Data Science project you've been involved in and why was it particularly meaningful?

The most impactful data science projects that I have been involved in include deep partnerships with business subject matter experts throughout the project. I believe it brings you closer to the real problem you are trying to solve as opposed to going heads down to find the most technically precise solution. Also, just listening and bouncing clarifying questions about the domain with an expert enables you to build more robust models with richer features. The most rewarding projects involve co-developing hypotheses to test and iterating model results together towards a solution.

What's a skill you find to be important for the workforce that you didn't learn from school? What did you do in college that helped you prepare for entering the job market?

As I've gained experience and grown in my career, I would say creativity and an ability to communicate your ideas effectively are skills that I didn't really learn or focus on in school. Of course, you get opportunities to do presentations, but there is a special art to connecting with your audience based on their goals and emotions that I've learned through experience and observation. If you are in a highly technical or quantitative field, this can be a differentiator in your career progress.

When I was in college, I had a few different internships which exposed me to day-to-day work environments across a few industries. This gave me the opportunity to observe work cultures and various roles that co-exist within a company. I gained invaluable skills from the hands-on work experience related to the internship itself.

I would recommend speaking with as many professionals tied to your field of study and across industries that you are interested in. This will give you a good perspective of what different careers look like to help align with your professional and personal interests. Organizations like CADS are a wonderful way to gain this exposure and I would have loved being a part of this group back when I was at UNC.  Data science is a diverse discipline, so learn from people applying it in academia and industry, in large companies and startups, those building products or doing R&D to help identify a good fit for you.

What trends in data scientist/research scientist roles do you predict will occur? Will these roles be offered more in the upcoming years/what areas of machine learning research will there be more focus on?

I think we will continue to see data science impacting all aspects of life. It will continue to be embedded in products and will help drive research solving the world's toughest problems. As a result, we may see more and more specialization in specific domain areas or in different parts of the machine learning engineering process.

I also think we will continue to see a democratization of machine learning where more people of different technical backgrounds get involved as newer technology reduces some of the technical barriers to entry. I believe we will also integrate more creative and design principles with machine learning as more and more technology products and solutions have AI and machine learning embedded, where models are features within products.

What are projects and other endeavors that we can take on this summer, with internships and other opportunities canceled/postponed due to COVID-19? There is more time available for personal projects, maybe even online courses that you could suggest for us.

Get involved in events like Kaggle contests or data hackathons. Find a personal interest (such as a sport or cooking) and design an analytics experiment that you can articulate for future employers. Aligning it to a personal interest will increase your passion for the work, its quality and this will come across in interviews. Be able to explain your approach and methodology from a technical perspective as well as how it solves a real-world problem for someone. If you haven't, set up a GitHub account to showcase your work but also practice explaining your project's results to people with strong data science backgrounds and those without a background. This will help strengthen your communication skills.

Take advantage of online learning opportunities (there is a lot of free and low-cost content out there!) to continue to refine your skills in areas that interest you or you want to strengthen. For example, if you feel strong in machine learning principles, this might be a good time to get stronger in natural language processing or time series or learn a new technology like TensorFlow. Continuously look for ways to round out your skills.

**Speaker-Specific Questions**

Advice for students deciding between a career in DS vs. Actuarial Science? What tools are commonly used in your work and how do you stay up to date with new technologies? How do you go about mentoring younger employees, and what qualities should one look for in a good mentor?

Data science and actuarial science are both strong career paths with overlapping skills such as statistical modeling and deep understanding of mathematics. The disciplines are coming closer together as actuaries learn more data science concepts and data scientists working in insurance companies learn about actuarial modeling. I've seen the two roles partner very well together on projects. My advice again is to speak to people in each career path to learn the best fit because the application of these skills can be different.

Data science tools are constantly evolving, and the discipline is extremely dynamic, which makes it very exciting and requires and attitude of lifelong learning. One way I support this is by encouraging my teams to learn new skills such as a new technology or an advanced analytics technique they picked up in a project, but also teach those back to others through hands-on learning. We host monthly lab sessions where data scientists take turns teaching a new concept with hands-on practical learning.

Mentoring is very important to me and I think everyone has the capacity to mentor and be mentored regardless of experience. I personally enjoy coaching younger employees because I learn a lot in return. I also like pairing up interns and junior data scientists with more experienced ones, so they can learn directly from them on projects. A good mentor will make the relationship a priority and be a sounding board for questions and professional guidance. They will listen closely and often can connect you with other people or resources that help you achieve your goals.

**Dr. Jorge Silva:** Adjunct prof at UNC CS and Principal Developer, Machine Learning at SAS

**General Questions**

What is an example of a high-impact Data Science project you've been involved in and why was it particularly meaningful?

Back in Europe, years ago, I participated in the development of a computer-vision system for recognition or car license plates in highways. This allowed for enforcement of tolls without requiring toll booths, thus making traffic flow much faster. It was implemented and made it into production, covering 1000s of miles of highways, and it is still in production now! Being able to say "I was involved in that" about something used by millions of people definitely gives plenty of bragging rights!

What's a skill you find to be important for the workforce that you didn't learn from school? What did you do in college that helped you prepare for entering the job market?

Most of all, curiosity. This is a driving force that will allow you to overcome obstacles and acquire any skills that the market requires in the future. The particular course materials you learn in college are far less important than the ability to be a life-long learner.

What trends in data scientist/research scientist roles do you predict will occur? Will these roles be offered more in the upcoming years/what areas of machine learning research will there be more focus on?

The honest answer is "I don't know". Above all, be flexible and master the mathematical foundations: optimization, algorithms, data structures, statistics, linear algebra, etc. Focus less on specific software packages, be ready to learn any programming language.

What are projects and other endeavors that we can take on this summer, with internships and other opportunities canceled/postponed due to COVID-19? There is more time available for personal projects, maybe even online courses that you could suggest for us.

Read everything you can get your hands on! Coursera is an obvious source of online courses, and many tech companies provide education free trials. Example: <https://www.sas.com/en_us/training/academy-data-science/trial.html>

In the spirit of curiosity, also play with Kaggle datasets. Learn by doing, as much as you can!

**Speaker-Specific Questions**

What challenges have you come across in bringing new research into production? Why did you switch from academia to industry?

Many challenges! Developing enterprise software is radically different than developing a research prototype. Specifically, you need to worry far more about validating inputs and gracefully handling error conditions. Testing is front and center, and we all must get used to test-driven development.

Other examples of things to worry about:

- Ensure that your outputs and messages (errors and otherwise) are localizable, meaning that they can be translated to other languages without changing the rest of your source code.

- Backward and forward compatibility. You are at far less liberty to introduce breaking changes to your API/syntax after your software has been released and is part of customer code. This is important in open-source as well.

- Documentation is source code too. Think about usability, if you want people to use your code.

As for why I changed from research to industry, it's simple: industry has bigger data :) Many of the most interesting challenges in Machine Learning and Data Science are related to business datasets. Look at Kaggle competitions, look at the Netflix challenge in recommender systems, and look at the Google ImageNet challenge. There are many exciting projects and opportunities for both PhD-level research and for non-PhD-level engineering in industry.

**Da'Esha McPhaul:** Assistant Head of Admissions, Institute for Advanced Analytics
**General Questions**

What's a skill you find to be important for the workforce that you didn't learn from school? What did you do in college that helped you prepare for entering the job market?

The MSA program focuses on preparing the "whole professional" to address concerns like these. Often times, educational programs see the student as the customer and the curriculum as the product, making sure to deliver a diverse menu of course options for the students to get the experience they believe is most interesting or important to them. The MSA program is far different. We believe the employers are our customers and our students are the product. We regularly engage industry professionals to explore what is happening in the real world, what kinds of problems they're facing, what skills our alumni believe they need but may not have encountered in our program, etc. We use that information to regularly update our curriculum to make sure our graduates are leaving with the most up-to-date/industry-current skills. Beyond that, we offer many support services to round out their technical skillset. This includes professional development coaching, communications training and support, in-house career services, and mental health counseling & workshops. Given this comprehensive experience, many of our students value the cohort format of the program that helped them build life-long relationships with peers in the field, and a network of alumni to support them throughout their careers. They also value our "learn by doing" approach. Many have told me that they "learned how to learn" while in the program. Things come at them fast in our 10-month program and there's no way anyone person is going to know everything they'll need to know (in fact, some industry experience has to be learned and earned on the job) but they learn how to seek out resources and collaborate with others to get the job done well.

What are projects and other endeavors that we can take on this summer, with internships and other opportunities canceled/postponed due to COVID-19? There is more time available for personal projects, maybe even online courses that you could suggest for us.

From an admissions standpoint, for those who may consider graduate studies in the future, take advantage of this time to:

1) Take care of yourself. Recognize that everything we know has been disrupted and it's okay if you're not firing on all cylinders at all times. Take the time to explore self-care techniques that help keep you centered and focused. And remember to have some fun. With that said...

2) Practice! Programming is like a musical instrument. You don't get better at it without practicing. Admissions committees love to see your strength in programming. In an accelerated program like ours, we like to see that applicants have put in the time to at least begin learning to code on their own (using statistical programs like SAS/R/Python/SQL). Side projects, online hackathons, and online coursework are great resources to learn and build your coding skills. Regardless of what type of role our students go into, it will involve some form of programming so demonstrating your ability in this area will help you stand out with employers and be prepared for the various types of interviews you might encounter. Datacamp.com has a ton of free resources on open source tools. SAS University edition may also be available to you as current students.

3) Build your statistical knowledge. Often times when people think of data science, they think of programming but a successful data scientist or analytics professional needs to understand the math. Explore some online coursework deepen your understanding of the basic concepts and advanced methods like regression analysis. We offer an online self-paced course at the institute for those who have already completed introductory topics but may need a bit more (feel free to reach out to our admissions team if you want to learn more about it). There are several coursera courses that are helpful. Duke has a statistics course using R and Johns Hopkins has a great course on Regression Modeling, both on coursera. There's also an introduction to statistics using R course on datacamp that we've seen applicants have success with.

**Speaker-Specific Questions**

What makes candidates stand out during admissions? What non-traditional backgrounds see the best results in your program? What are the characteristics of students who have the most success in recruiting? Tips for advising students on picking between different DS graduate programs?

We love candidates who know their stuff. We don't just mean those who aced every stats course they've ever taken. There's a difference between learning material to pass the test and truly understanding it. We value candidates who know the material well enough to explain it to others. A key part of our training at the Institute is helping our students learn how to translate highly technical concepts to non-technical or business audiences. The same thing applies for programming. At least knowing how to write code on your own would set a candidate up for success at the Institute, given the fast pace of our curriculum. Being able to talk us through a project you've done (either in class, at work, or thorugh a side project) adds a bit to the conversation, especially considering that our program conducts interviews with all finalists. Knowing your material and being able to confidently talk through it is a strength. With the application itself, the personal statement or statement of purpose may be your only opportunity to speak directly to the admissions commitee so thinking carefully about why you want to pursue a career in analytics and being able to clearly and concisely relay that is key.

If you are exploring many programs, think about what is most important to you. Is it going to a reputable program? Is it their track record for career placement? Is it the curriculum? Is it the student experience? Program duration? Support services? There are so many factors you can consider, take the time to figure out which is most important to you. When looking at the curriculum, take a look at the courses offered to see how they differ. Some data science programs focus solely on programming skills and not on data analysis techniques. We see tons of applicants who've worked in positions labeled data scientist, even at senior levels, but do not understand key statistical concepts. There are also programs that may focus more heavily on the business side (we see it more with those labeled business analytics) where students become familiar with what they may need to work in a more client-facing space, but their technical expertise may not be as strong. We train our students for both. Most enter the workforce in a position that combines some balance of technical and soft skills/communication skills. There are certainly those who have an affinity for one or the other but our most successful students have a solid grasp on both. So when you're considering programs, figure out which is going to provide the skillset you want, then narrow down using your other priorities. Feel free to reach out to admissions representatives for those programs to see what they are looking for specifically and what they offer that might not be obvious from their website. Graduate school is a huge investment of time and money, take the time to do your research.

I'm happy to answer any questions you might have, whether you're considering our program at NC State or any others. We offer admissions consultations to help students talk through what they may need to do to prepare for our program but I'm also happy to answer any questions about preparing for graduate school in general. Feel free to email me at daesha\_mcphaul@ncsu.edu if you have any questions.

**Dr. Mario Giacomazzo:**Teaching Assistant Professor at UNC STOR

**General Questions**

What is an example of a high-impact Data Science project you've been involved in and why was it particularly meaningful?

During my PhD, I interned at a company called Health Services Advisory Group (HSAG) where we did research in public health. This was my first time working with an analytics team. In this team, I was given the responsibility to analyze and compare different home healthcare agencies across the country. My team let me recommend certain methods, make a significant number of the visuals, and even write some of the report. I learned more in this internship than any course in my PhD.

What's a skill you find to be important for the workforce that you didn't learn from school? What did you do in college that helped you prepare for entering the job market?

From the workforce, I learned how to explain complex models and methods orally and in writing to people unfamiliar with statistics. From college, I learned about the many methods available to be used in data science and the mathematics behind those methods.

What trends in data scientist/research scientist roles do you predict will occur? Will these roles be offered more in the upcoming years/what areas of machine learning research will there be more focus on?

Given what I have seen in the past few months, more companies are going to advance their ability to do work remotely. They are going to expect their incoming data scientists to be comfortable accessing data and computing in the cloud. They will want to know that their analytics team can produce high quality work at a distance. I don’t think these job opportunities are going to decline in the upcoming years, but as more students start studying data science, companies will find creative ways to find the best applicant.

What are projects and other endeavors that we can take on this summer, with internships and other opportunities canceled/postponed due to COVID-19? There is more time available for personal projects, maybe even online courses that you could suggest for us.

You should always be learning something. The data science trend is not unique to UNC. At every university, there has been an explosion in data science courses. As more students start studying data science, the level of competition for jobs will increase. For good students who are unsure or worried about jumping into the workforce during this time, I highly recommend that you consider a Masters’ or Doctoral degree. The best time to incur educational debt is during a poor economic time. Continuing education allows you to defer paying your loans which is helpful when the job market is tight. Also, if you get accepted for a PhD program, you will go for free and get a stipend. For the summer I have four goals for all students, 1) Start a Github account for yourself, 2)Find an R or Python textbook to work through in a topic that you are interested in 3)As you work through the textbook, update a Github repository of your practice 4) Start a free personal website for yourself using Github pages and summarize what you did.

**Speaker-Specific Questions**

What are some topics that you wish that you could include in the Data Science/STOR curriculum at UNC? What made you decide to go into teaching Data Science specifically? What are some examples of some class projects that really stuck out to you?

I wish we had more coursework on Bayesian statistical methods and time series analysis. My dissertation was on Bayesian estimation of time series. I think it is important to see probability and inference from the Bayesian perspective. Time series methods are useful for forecasting applications in business, economics, energy companies, etc. Because of my research, I had a considerable amount of experience and passion for data science. Having such an applied dissertation provided me an opportunity to teach Data Science at UNC. The final project in my course is the most important assessment. Students have the opportunity to work in a team and perform data science on any dataset they want. If a student is part a successful group project, the final paper can be used by a company to assess the future applicant’s ability to work on a team, think creatively, produce stunning visuals, perform predictive modeling, and accurately summarize what was learned. For a sample of some outstanding group projects, check out <https://supermariogiacomazzo.github.io/STOR320_WEBSITE/> and look in the final project section. There are links to final papers involving Pokemon, Seattle housing, missing migrants, YouTube, and CDC data.