Chris Lindner
Product Science Manager at Indeed
"Where do Data Scientists come from?"
Data science (DS) is a rapidly growing and maturing field, and what it means to be a data scientist is constantly changing. Using job market data from Indeed, the world's #1 job site, we find that DS roles seem to fall into several logical groupings based on skills requirements and educational backgrounds. Furthermore, we discovered that data scientists come from a wide diversity of educational backgrounds. Most data scientists hold advanced degrees in non-DS fields, although a growing number are originating from newer data science and data analytics degree programs.

Christopher Davis
Applied Scientist at Descartes Labs
"Descartes Labs: Building a Data Refinery to Understand our Planet"
Our planet teems with activity, and we collect unseemly amounts of data in our quest to understand that activity. How can we effectively leverage petabytes of geospatial imagery to make good business and environmental decisions? In this talk, I will describe how Descartes Labs uses computer vision and machine learning techniques in a scalable cloud-based environment to turn pixels into decisions.

Destry Saul
Blockchain Engineer at Unchained Capital
"I'm a Blockchain Engineer, AMA"
'Blockchain' was one of 2018's hottest buzzwords. But similar to data science, big data, and AI, the reality of blockchain tech is more nuanced that the hype. I've had the privilege of being a blockchain engineer for two years now, and I'll give a whirlwind tour of what that means. What are blockchains and why are people so excited about them? What are the current limitations? What is true and what is false marketing? Exactly what sort of data is in a blockchain? How do you learn to be a blockchain engineer? And I'll leave plenty of time for questions.

Donald Lee-Brown
Program Director, Data Scientist at Insight Data Science
"Transitioning from Academia to Industry: Careers in Data Science"
In the United States, more than 150,000 currently open data science and data engineering positions cannot be filled due to lack of qualified candidates, and this talent gap is only expected to widen over the next few years. However, it is often surprisingly difficult for individuals with PhDs in quantitative fields to successfully move from academia to careers in data science. The problem is two-sided: PhD holders seeking to transition often lack industry-relevant skills, while on the hiring side it is challenging to identify the candidates who will be able to successfully leverage their advanced quantitative backgrounds. At Insight, an education startup founded in 2012, we've helped over 1500 individuals successfully transition to data science in industry through our Fellows programs, including nearly 500 alumni with backgrounds in physics or astronomy. In this talk, I'll highlight some lessons we've

learned at Insight about what makes a successful data scientist, and share some general advice for preparing for a data science career in industry.

Elizabeth (Betsy) Barton
Researcher at Renaissance Technologies LLC
"Data Science and Statistical Arbitrage"
Quantitative finance is an extremely data-intensive field. The data range from price and volume counts of financial instruments, some of which change hands millions of times per day, to vast volumes of government reports, news stories, and even weather data. I will describe in very broad terms how we sift through these data and look for signals that predict price moves.

Joshua Miller
Data Science Manager at Samba TV
"Creating Synthetic Control Groups to Understand Incremental Lift from TV Commercials"
To understand the effect of an advertisement, measuring a conversion rate alone does not paint a complete picture - we want to measure the incremental lift an advertisement has on an audience after taking into account the targeted audience's propensity to convert without being exposed to the advertisement. In digital advertising, this effect is quantified by selecting some of the people that meet the targeting criteria (location, demographics) that would have been exposed to the digital ad and instead serving them a "placebo ad," placing them into a control group. In general, most broadcast TV commercials are an all-or-nothing scenario, i.e., everyone watching the program gets the same commercial. Samba TV's proprietary Synthetic Control Group (SCG) process allows us to overcome this hurdle by deriving a control group using a household's TV viewership data. We design the control group to be, based on our data, the target group that was likely to see a specific ad campaign spot but that did not. We assign these households "synthetic" television exposures so that they can be compared to the same attribution window used to measure the ad campaign, allowing an apples-to-apples comparison. Comparing the exposed group's conversion rate to the conversion rate of this control group then reveals a measurable, attributable lift.

Julie Hollek
Sr. Data Scientist at Twitter
"Product(ive) Data Science"
Data scientists are the people behind the scenes, helping others deliver better, smarter results in their daily work. This is especially true for product data scientists who must hone their craft to determine what things are working for a given product, where do we want to take it next, and how can we make product decisions aligned with company is trying to build? Cross-functional communication is critical to success in this role; you need to be able to craft a message that is born out of math to make compelling arguments that are digestible by stakeholders across the business. In this session, we'll define Product Data Science and discuss contributing factors to success.

Matt George
Sr. Machine Learning Engineer at Planet
"Small Synoptic Survey Telescopes"
Satellites and drones are performing large and detailed imaging surveys of Earth for commercial, governmental, and humanitarian applications. These constellations of downward-pointing telescopes offer many parallels to astronomical imaging surveys and their own set of data challenges. I will outline the current capabilities of these time-series surveyors, with a focus on how machine learning and computer vision are being used to extract meaningful information from the data.

Michael Busha
Director of Research at Foundri
"It's OK Not to Know"
Adversarial attacks have recently gained significant attention as they have exposed unexpected weaknesses in deep neural networks that otherwise achieve state-of-the-art performance. Despite qualitative evidence that deep nets can generalize high-level features from data, adversarial attacks using imperceptible modifications indicate a challenge in training that is not yet understood. In this work, we investigate the question: can off-manifold data obtained from existing datasets and simple data-generation techniques improve resistance to certain adversarial attacks? Our results show that modest data augmentation techniques for training a network with off-manifold data can, indeed, provide robust defense against adversarial attacks without require updates to the training methods or algorithms.

Roozbeh Davari
Senior Data Scientist at Aisera
"Leveraging Natural Language Processing and Understanding for Adding Value to Businesses"
Since 1950, tremendous amount of progress have been achieved in Natural Language Processing (NLP). Over time, it has moved from being mainly rule-based toward more statistical solutions. This transition was not solely for achieving higher accuracy, but tackling problems never done before. Recent research has heavily relied on Machine learning, which can provide a useful solution when reaching an exact solution is not feasible. In this talk, we will discuss some of classic applications of NLP (e.g., sentiment analysis, and grouping of relevant document), and its more modern use cases (e.g., reading comprehension, and machine translation).

Taka Tanaka
Manager of Data Science; Diversity and Inclusion Committee Lead at Schireson Associates
"Client-facing Data Science: Strategies and Pitfalls"
The most successful bespoke data science solutions for businesses almost never look like what's taught in textbooks, or even what wins Kaggle competitions. In this talk, I'll share what I've learned as a data consultant for some of the largest businesses in the world: when the best models aren't the most helpful models, when to ditch model performance scores, how to arrive at the right solutions, how to

choose the right benchmarks, defining and exceeding deliverables, and building trust in your work.

Uma Vijh
Data Scientist at Kidaptive Inc.
"Empowering Tutors with Big Data"
Analytics Online education has been growing over the past few years and massive amounts of learning data are being generated. We will report on our efforts to use learning analytics to empower teachers to help all learners reach their full potential. We implemented reports to teachers providing general behavior and achievement insights of their students on a weekly basis and supplement these with summary monthly reports with the students' study patterns and trends to evaluate growth. The talk will provide more detailed descriptions of these reports and also includes results from a preliminary efficacy study we conducted to evaluate this intervention which shows positive effects on mean test scores of the students.