

A DATA SCIENTIST'S GUIDE TO START-UPS

Foster Provost,¹ Geoffrey I. Webb,²
Ron Bekkerman,³ Oren Etzioni,⁴
Usama Fayyad,⁵⁻⁷ and Claudia Perlich⁸



Abstract

In August 2013, we held a panel discussion at the KDD 2013 conference in Chicago on the subject of data science, data scientists, and start-ups. KDD is the premier conference on data science research and practice. The panel discussed the pros and cons for top-notch data scientists of the hot data science start-up scene. In this article, we first present background on our panelists. Our four panelists have unquestionable pedigrees in data science and substantial experience with start-ups from multiple perspectives (founders, employees, chief scientists, venture capitalists). For the casual reader, we next present a brief summary of the experts' opinions on eight of the issues the panel discussed. The rest of the article presents a lightly edited transcription of the entire panel discussion.

Background

INTRODUCED IN ALPHABETICAL ORDER, participant **RON BEKKERMAN** (currently at the University of Haifa) was a venture capitalist with Carmel Ventures in Israel at the time of the panel discussion, and before that was an early data scientist at LinkedIn. Ron is also cofounder of a stealth-stage start-up. **Oren Etzioni** (co)founded MetaCrawler (bought by Infospace), Netbot (bought by Excite), ClearForest (bought by Reuters), Farecast (bought by Microsoft), and Decide (bought by Ebay), and also is a venture partner in the Madrona Venture Group. **Usama Fayyad** (currently chief data officer at Barclay's Bank) cofounded ChoozOn Corp., Open Insights, Audience Science, and DMX Group (bought by Yahoo!), and is executive chairman at Oasis 500, the first early stage/seed investment company in Jordan, with a vision of starting 500 companies in 5 years. **Claudia Perlich** is chief

scientist at fast-growing Dstillery, after notable success as a data scientist at IBM.

The panel organizers/moderators were **Foster Provost** of NYU, cofounder of Dstillery, EveryScreen Media, and Integral Ad Science, coauthor of the book *Data Science for Business*, and former editor-in-chief of the journal *Machine Learning*, and **Geoff Webb** of Monash University, founder of GI Webb & Associates, a data science consultancy, and editor-in-chief of the journal *Data Mining and Knowledge Discovery*.

Summary of the Experts' Opinions

Here is a quick summary of the main issues that arose and were discussed by the panelists.

1. The motivations for being involved in start-ups are not all about money. They include the excitement caused

¹NYU Stern School of Business, New York, New York.

²Monash University, Clayton, Australia.

³University of Haifa (formerly of Carmel Ventures), Haifa, Israel.

⁴CEO, Allen Institute for AI, Seattle, Washington.

⁵Barclay's Bank, London, United Kingdom.

⁶Oasis 500, Amman, Jordan.

⁷Open Insights, Bellevue, Washington.

⁸Dstillery, New York, New York.



by the speed and unpredictability of events; the opportunity for real-world impact; the benefits of working in a small, focused team with a “can-do” attitude; the rewards of being an integral component of something big, interesting, and worthwhile; the thrill of creating something big from nothing, and, of course, the potential of substantial financial reward.

2. The risks are low because current demand for data scientists is so high and, no matter what happens, you will gain valuable data science experience. Also, you can negotiate remuneration to balance equity (i.e., potential long-term profit) against salary (i.e., certain current income). It is critical to negotiate a good deal when you join any company.
3. The financial rewards are arguably greatest for the founders. Once a start-up is reasonably established, joining it may be no more beneficial *financially* than joining a very established company. On the other hand, an established start-up can provide many of the same nonfinancial rewards (see point 1), as well as better work–life balance.
4. The greatest critical success factor is the team. A great team can make something from very little. A poor team is unlikely to succeed no matter how good the vision. The most critical member of the team is the chief executive officer (CEO). The team must be coupled with an idea that addresses some real pain or major opportunity. To get major venture capital funding, the business plan should be for a \$1 billion-plus business.
5. If you want to assess the success prospects of an established start-up, an excellent indicator is who is funding it. If it is funded by a top venture-capital firm, then you know that it has been assessed as a good bet by an informed and likely competent team.
6. Now is a very challenging time to hire data scientists to start-ups (and elsewhere). One strategy for companies is to make yourself publically visible as a top data-science company, as top-notch data scientists benefit considerably from working with other top-notch data scientists. When assessing potential staff, look for passion, vision, and excitement.
7. On the question of whether data scientists need a PhD, it is clear that it is not necessary to have one in order to get a rewarding data science position and to succeed at it. A PhD definitely adds substantial value, but it is not clear that on average this is any greater than the value of 5 years of focused industry data-science experience. One of the key factors either way is the mentorship—a great PhD with a great advisor is hard to beat in terms of skill set, critical thinking, and independence; these also can be developed in an industry position with a great

“THE GREATEST CRITICAL SUCCESS FACTOR IS THE TEAM. A GREAT TEAM CAN MAKE SOMETHING FROM VERY LITTLE.”

mentor. On the other hand, you are unlikely to gain great skills if you go straight from a master’s degree to leading a data science project. Data science is a craft, and, as with most complex crafts, one learns best by working with top-notch, experienced practitioners.

8. With respect to founding a data science start-up, it is important to have top-notch technical and business leadership. If you want to be a technical founder, then it is essential to partner with a great business-savvy cofounder.

The Panel Discussion

Geoff Webb: I’m going to give the first question to Usama. Usama, you were a very successful researcher, and you were also very successful at Microsoft in industry. What on earth made you leave all of that in order to launch your first start-up?

Usama Fayyad: I don’t know. If you asked my family at the time, they would’ve said, “He’s crazy. He lost it.” No. The real driver is—the biggest factor is possibly curiosity of seeing start-ups happen and wondering what that world is about if

you’re just doing research and just pure data science. But then there are the second and third factors—which are just as important—number two is you really want to see the impact of your work. You really want to see your work spread faster, and you always get to a place where you could only have so much influence from a platform, whatever that platform is. And for me Microsoft was a wonderful huge platform, but it has

its limitations. And, of course, the third one is unlimited upside, so if you really do a good job, if you really have a product that people want out there, then the sky is the limit and there’s nothing as exciting as that.

Geoff Webb: So what were the biggest surprises?

Usama Fayyad: Well—now that I’ve done a few start-ups, it’s no longer a surprise—but at the time...look, nothing goes like you planned. It’s always going to need more money, take longer time; you’ll go through some very, very, very dark times; you have to—every good company has to—go through a cycle where you have to lay off employees and you face zero cash balance and the world looks really dark and then somehow things work out and everything comes together through sheer willpower. But I think the biggest surprise now—having done a few start-ups, it still surprises me—is no matter how many times you do it and how many times you reflect on it—every month I give a lecture to about 60 entrepreneurs and one of the lectures is about mistakes to avoid when you do a start-up, and I’ll just share

with everyone—with every start-up I end up making the same mistakes over and over again. So that's surprising.

Foster Provost: I have a question for Oren that's sort of along the same lines. We know that there are lots of academics in the audience and you had a nice cushy job as a professor and doing your research and all that. Why in the world would you go off and do this? And what does it add to your life as an academic to be involved in founding companies. Presumably, it's a lot of work?

Oren Etzioni: Let me quickly give two answers. I think one is that in academia, once we've figured out how to survive, we realize that we want to have an impact, and engines like Google Scholar help to quantify just how little impact we're having. I saw my articles cited maybe a few times, my favorite ones maybe not at all, and then when you click on the citation and actually see what they said, it becomes clear they didn't read the article. It's like they had to put an article on X and a list of articles on Y, so they threw mine in if I was lucky.

So for me I felt like: how do I have an impact? Teaching has an impact, research has an impact, but I found I wanted to create something that people love. The first thing that I created where I had that kind of relationship with customers was MetaCrawler, a Meta search engine. Anybody vaguely remember MetaCrawler?

Geoff Webb: I do!

Oren Etzioni: Thank you. Most of you are too young. This is before there was Google and such. Anyway, so I think, having an impact was a big deal for me. And then doing it for a while I realized that academia is kind of like playing bridge: it's cerebral, it's challenging, it's fun. But start-ups are like playing poker: more immediate satisfaction, more at stake like Usama was saying, although I don't know I've ever believed in unlimited upside, but it's definitely higher stakes, and so I actually loved the ability to do both.

Foster Provost: And you feel like—from a perspective of an academic—there are particular costs that ought to be taken into account before somebody just jumps into doing it?

Oren Etzioni: So you're asking is there a cost to you as a professor and an academic to do start-ups?

Foster Provost: Yes.

Oren Etzioni: Look, there is a cost and there is a benefit. I mean, I think I'm a much better teacher, for example. I teach senior capstone course where people write projects and cutting-edge software, and I'm able to bring things from industry to the classroom, but there's clearly a cost. I mean, at the times where it's most intense, you are torn in multiple

directions and feel like you're not doing either job as well you'd like, yes, so there's a cost, yes.

Geoff Webb: Claudia, along the same lines. You were very successful at IBM Research. You were contributing to the bottom line. You were winning serial KDD Cups. What drove you to leave a very comfortable research position to join a very small company?

Claudia Perlich: So the job kind of found me. I wasn't in any which way looking. I was actually quite content. As you pointed out, my life really wasn't that bad. But I just felt that if I were to say no to this opportunity, I would regret it for the rest of my life. And that's a very personal position. So it is not even the upside. I'm actually quite risk averse. I'm not really looking for the upside potential or facing the downside, but what fascinated me was just the kind of energy level that you found when you started talking to people in start-ups, the pace at which things happened. It was a completely different world, and I felt that it was time for me to dip a foot in the water and see how it suits me, and I've never regretted it.

Geoff Webb: What were the big surprises for you?

Claudia Perlich: Big surprises? I had no idea what I was walking into initially. You realize very quickly that delivery is all done within hours or days, and if something breaks, then you get a call at midnight saying, "Look, whatever this is you built is broke, so fix it."

It wasn't really a surprise...but it was a notable change. What I was amazed by is the turnaround. If I said I needed this data stream fixed, it was done. Tomorrow. There were no questions asked and that was really amazing to me.

Geoff Webb: Ron,....

Oren Etzioni: Excuse me, first can I interrupt for a second just to ...

Geoff Webb: Absolutely.

Oren Etzioni:...liven things up a little bit. Claudia brings up one of my pet themes here, which is this notion of risk. So are people in the audience thinking when this question has come up well, do you take a risk when you do a start-up?

And I'm actually—it's interesting that Claudia says she's not a risk taker. I'm not a risk taker either in the sense that my first car was a Volvo, and I like to say that the only roller coasters I like to get on are start-ups, but that's what I'm talking about, physical risk, okay. To me with start-ups, the only risk is not doing it. Staying in some cushy job of whatever kind and then looking back 20 years later when you've got a mortgage payment and you're too old to say, "Oh you know, I could've taken a chance." That's a real risk.



If you do a start-up, and you work hard and you give it a shot, if it fails, okay you can move on to the next thing, but that's not a risk, that's an adventure, an intellectual adventure and, like Claudia said, it's high energy, it's exciting, it's one of the great things that work in this country. We always hear about how Congress is not working and our economy is not working, so many things are not working. Well, the start-up system is working. It's working here due to people like Usama. It's also starting to work more and more globally. It's just such an exciting thing, so I just want to say, what is really the risk?

Usama Fayyad: Yes. I just want to emphasize this one like a plus, plus. If any of you in this room—all of you understand what data science means? All of you understand big data? All of you understand data mining, analysis, algorithms?

If any of you, including the academics, and I underline it 10 times, think that you are taking a risk by making a career shift and you're giving up 10 years' safety and all that, you're dreaming. Guys, you are in a field that has the hottest job demand that I can ever remember in my life. Companies are dying to get their hands on big data people, on data analysis people, on data mining people, and this includes banks, telcos, insurance companies, you name it. So I think a start-up now is zero risk and I agree completely with Oren. The risk is in not trying it.

Foster Provost: OK. Ron, you have two perspectives: working for a start-up in the past and as a venture capitalist (VC) now. Let's talk about the first of the two. Tell us about getting hired at LinkedIn. I mean, it was a pretty big start-up but a start-up nevertheless?

Ron Bekkerman: Yes. My LinkedIn story started in the beginning of 2009. One day I was approached by a small company named Facebook. They told me they were building a data team and they were looking for someone of my profile to be a senior person on that team. My response was no. I mean it was such a small company. It was practically run by teenagers. Besides, I wasn't on Facebook at that time—I didn't connect myself with their value proposition. So I said no but started thinking. Back then, it was already obvious that social networks were the future. I wasn't deeply into Facebook, but I actually was really attracted by the value proposition of another social network, the value proposition of keeping in touch with your professional connections and knowing about what's going on in your field and finding a job when you need it.

So I wanted to work for LinkedIn. Despite the fact that LinkedIn people are generally well connected, I didn't know anyone at LinkedIn at that point, so I basically went to their careers webpage and applied. Surprisingly, a LinkedIn recruiter came back to me the next day. And yes, it did feel like a start-up—intense and dynamic.

Foster Provost: What do you think about joining a company that is still pre-IPO and still very much a start-up, but is established, versus founding your own start-up? Do you have an opinion which one of those is preferable?

Ron Bekkerman: Yes, I do actually. I think that we can take it as an optimization problem. Obviously, it's very hard to optimize our lives, but let's just do an exercise. Say we want to have a very interesting job, so we choose to work for a start-up, but we also want to make some money after all. An objective function that we would probably optimize is our monetary reward multiplied by the probability of getting this reward. First, let's decouple this problem by saying that we're maximizing the amount of the reward.

“COMPANIES ARE DYING TO GET THEIR HANDS ON BIG DATA PEOPLE, ON DATA ANALYSIS PEOPLE, ON DATA MINING PEOPLE, AND THIS INCLUDES BANKS, TELCOS, INSURANCE COMPANIES, YOU NAME IT.”

If we want to get the biggest reward, presumably at a lower probability of getting it, the right approach will definitely be to start a company. If you found a company and it's successful, you can make something like \$10 to \$20 million on an exit, but the probability of getting this money is very low. Let's say that your start-up does data science and that's why there is a decent probability of actually making this money. I'd say 10% is a very high probability to succeed for a company that practically doesn't yet exist.

So, on one side, we end up with \$20 million times 10% success rate, which equals \$2 million. On the other side, we can maximize the probability of success and then we're likely to get a low reward—that's what I did at LinkedIn. I joined as employee number 400-something. The company was already very prosperous, so the probability of success was pretty high. The uplift was relatively low. Let's say a person can make a million dollars if he or she joins a very successful company like LinkedIn—with the probability of success being pretty much 100%.

By the way, the real story is that shortly after I joined LinkedIn, I had two conversations with LinkedIn executives. Both of them had left Google to work for LinkedIn. I asked each of them pretty much the same question: “How could you leave a very strong, successful company and join a start-up like LinkedIn? It's probably very risky, isn't it?” To my

surprise, they gave me exactly the same answer: “You know what, Ron? There’s just no risk at all.”

So, if you start a small company, our objective function gives you about \$2 million. On the other extreme, if you are a data scientist in a large, very successful pre-IPO company, the objective function gives you \$1 million. Between those two points, the objective function goes down. If you’re a number one employee (but not a founder) at an early-stage start-up, the probability of getting your money is the same 10% as if you were a founder, but your reward is much lower, probably 10 times lower. So our objective gives you \$2 million times 10% success rate, that is, \$200K. Not worth the effort.

Let’s take another scenario: say you joined a relatively established company that is not yet near the IPO stage. Say you’re employee number 50, or something like that. In this case, your probability of success is much below 100% (let’s say it is 25%), but your reward would be about the same regardless of whether you are number 50 or number 400. That is, the objective function gives you \$1 million times 25% equals \$250K. Still not very interesting.

Foster Provost: Usama, do you think the estimate of about 2 million bucks is the expected value for a founder of a data science start-up?

And I mean the probability of success times the value-given-success. Let’s say you’re the data science guy on a three-person founding team?

Usama Fayyad: OK. So I think Ron may have done a more careful analysis than I have but here’s my rough, high-level reaction to that analysis. If that were true, given the salaries I’m familiar with for data scientists these days, then that equation would say you should never leave your job. So my guess is, look, this is a new area that we know is in high demand, that we know is going to be crucial for quite an upcoming time, so this is not like your average start-up addressing an average consumer play in a very crowded space. This is a pretty rare skill set here, so your chances of getting competition, there are so many barriers that have to do with knowledge and have to do with intuition, all of that’s working for you, and my suspicion, if Ron were to condition his numbers on data science and big data, I would imagine the returns to be higher. My rough guess is that you could easily stand to make, in expectation, \$10 million in 5 years, so that would be my guess.

Claudia Perlich: All right guys, if I hear you right then forget it. I’m not going to chair KDD 2014. I’m going to look for a different start-up.

All right, I take it back. Do you have health insurance?

Ron Bekkerman: Yes.

Claudia Perlich: You do? Why?

In expectation, that’s a stupid move!

The insurance company is making money, so in expectation you know that you will spend less on your healthcare than if you pay the insurance; otherwise, they would be out of business. My point is, from a very personal perspective, this may all be true. Maybe this is the female voice here: so I’m running the single-mom gig with a nine-year-old. It’s great that there’s the expectation that I make a \$10 million. I still need to feed my son tomorrow, and I can’t afford working 14 hours a day at a start-up that may fail. I think you’re right that the downside potential actually isn’t as big. It’s just lost time.

The good news is, if it goes belly up, well, there are so many other things you can do. It’s not like you’re stuck without a job for the rest of your life, so the risk is not that bad. It’s just at the point when I entered the game, I had certain different expectations on life and that came with some time for myself and a nice salary that I knew I would be easily making. So I negotiated my terms joining a 20-person start-up. I’m saying, “Look, equity is great, good for you. I’m in for somewhat more security” and that’s actually an argument you can have. It’s not like you have to work for free and hope for the big bang; there are ways in between.

Foster Provost: Let’s move to one of the questions from our special audience question management system! There are at least a couple questions that have gotten a lot of up-votes, although they’ve gotten a couple down-votes too.

Geoff Webb: Yes. This is a question for everyone. I might actually put it to Usama and then Oren: How do you know when an idea is start-up-worthy?

Usama Fayyad: If it’s personal, if it’s you doing it, it’s a completely different objective function than if you’re an investor. So these days I do the latter a lot. I evaluate probably hundreds of ideas to try to decide which dozens are we going to invest in, and there are many, many, many factors. It’s not—my first answer to that question—it’s not the idea, it’s the team ...

[Agreements from the panel members]

Usama Fayyad:...it’s team, team, team, because that idea most likely is going to change and change multiple times so—and by the way when I say team, I will underline team many times. Single founder, bad. Many founders, good.

The question I ask myself is: Do I believe this team, with this idea, in this market, can double the valuation of the company



in 6 months? That's really what I look for as a start-up. We're talking seed-stage. So that's a final decision I go through—do I put in money, do I sign on the check or not—is do I really believe that? And if I don't hit that degree of belief, into that comes a lot of factors; I just don't invest, but again that's the objective function of an investor.

An entrepreneur, on the other hand, including myself, is way more optimistic, loves the idea so much, blind to so many other issues, often doesn't listen. What I find is: having a team with you as an entrepreneur, if that team—if you can get that team to believe with you—then you probably have a good idea. So I'm a strong believer in the team dynamic.

Oren Etzioni: Let's see a couple of points. First of all to pick up on the theme, I very much agree it's a social enterprise. So one of the questions that people ask themselves is: Should I really be secretive about this or is somebody going to steal my idea if I talk about it?

You should be so lucky that somebody wants to steal your idea. That means it's a really good idea and most ideas statistically are not that good. About 50% are below average just for starters, et cetera, et cetera, so I definitely think you should be talking to people about it and see how they react. And talk to people who have some experience.

Another thing I always tell people: The best thing to do is to go immediately to your second start-up, because you learn so much with your first; if you can, then start with your second start-up. One approximation to that is to talk to people who have experience.

Let me give you two quick dichotomies that are helpful that people overlook, and all these things are matter of judgment. It's like buy low, sell high. Who wouldn't do that? But what's low and what's high? That's where the art is, but one thing is: nice-to-have versus must-have. OK, so people—I'll give you an example. Somebody came to me with what might be a great idea—it's really hard to manage your to-do list; everybody kind of keeps stuff in their inbox—"I'm going to build the greatest tool for managing to-do lists."

And by the way there are some such tools out there, and I said, "That's kind of neat, but the truth is that's a nice-to-have, because we kind of get by managing stuff on note pads and in our inbox and various things." I'm not going to go and spend a thousand dollars to get a great to-do list manager. Maybe if it's open source, I'll download it for free, or maybe I'll spend 99 cents, you know what I'm saying? So that's an example of a nice-to-have, not a must-have.

What's a thing that's a really huge pain-point where you say, "People really are not going to be able to live without this. They need this like oxygen?" If you feel that way about your idea, you're starting to head in the right direction, because again there is a discount factor for what other people think.

"THE QUESTION I ASK MYSELF IS: DO I BELIEVE THIS TEAM, WITH THIS IDEA, IN THIS MARKET, CAN DOUBLE THE VALUATION OF THE COMPANY IN 6 MONTHS?"

A second dichotomy is the distinction between a feature and a product. Again, so is what you're designing, is that an actual product that people will buy and gravitate to it? Or is it really kind of a small feature? And again often what people have is just a feature, and then

you have to figure out, okay, then how am I going to make sense of that? So again I could go on and on but just some little tidbits to think about.

Foster Provost: Here's an analogy—for those of you who are researchers in the audience—I think it's interesting to think about what kind of a researcher you are. There are some researchers, let's just say I'm going to caricature them, who basically can write articles and just get them accepted easily and those articles get a few citations. And there are other researchers who have the hardest time actually getting their articles accepted, but when they get accepted they get massive numbers of citations.

These both could be very good researchers. Those are similar to different kinds of ideas, ideas that may be low impact but high probability, maybe because they are incremental or just small "features," versus ones that you may have to talk to an awful lot of VCs before you actually get somebody to bite and give you funding. But if you can get a big VC to give you funding, the idea likely has much more upside potential. And so I think we each have to be sort of ruthlessly self-critical about what kind of ideas we have.

Ron Bekkerman: Can I give a VC perspective on investing, following what Usama was just saying? A VC comes into play after angel investors, so they usually invest more money, say, \$5 million to \$10 million, and they look for bigger returns. VCs don't invest in ideas, and they actually don't invest in teams. It's kind of given that it should be a good idea, and it should be a good team.

VCs invest in technology, but not because they really care about technology. Obviously enough, VCs care about a return on their money. The main question that needs to be answered when you ask a VC for an early-stage investment is: *Are you going to be a billion-dollar company?* This is the most important question—and there is a very, very simple math behind why VCs actually want you to become a billion-dollar company.

Say a VC firm has a hundred million dollars to invest. They don't invest their own money—they invest money of their investors and they promise a return of, say, a hundred percent in 8–10 years. This means that they need to make \$200 million over that period. A typical VC invests in about 10 companies, and they usually own about 20% of each company in their portfolio. If one of their investees becomes a billion-dollar company, the VC gets 20%, which is \$200 million. It's exactly what the VC owes their investors.

The probability is very low that there will be two very successful companies on the portfolio of 10 companies overall, but all the VC needs is *one* company that is a billion-dollar company, and they want you to be that company. They don't care about whether you are successful or not. They care about whether you are successful *big time*.

Geoff Webb: Claudia, do you have a counterpoint to anything's that's been said?

Claudia Perlich: I think the question for most of us isn't whether our idea is brilliant. I guess I speak for myself, but I consider myself kind of a geek, and if I wanted to be an entrepreneur I probably would've done that the last 20 years and not gotten into this now, so I think the more interesting question is what start-up do you want to join? Is it a start-up worth joining?

And the one observation aside from it being a good idea, and I actually cannot trust my judgment: I would never have invested in Facebook. To me, it is really: *how influential and how valuable, how big a part are you going to be?*

Are you going to be on the sidelines writing reporting for a product that has nothing really core to do with analytics? Or are you going to be a core part of the product?

And I'm just not even willing to consider the first type of job. Either I'm a core part and I can move things and then I can contribute, or I'm not interested because if you're on the sidelines, as we said, the business ideas change 15 times, and soon you'll find yourself just doing coding.

Foster Provost: Stepping up one level, you are advising data scientists to work for companies where the data science is central to the product or service?

Claudia Perlich: That's the most—by far—rewarding thing to do, yes.

Oren Etzioni: If you think about this question—which is a great one for people—should I join this particular start-up? How do I decide? I would say there are two things that are rough measures.

One is look at the people. It's hard to assess if the company is going to be successful or not if you don't have experience. But if the people are really good, that already mitigates your risk. You'll learn from them. You'll enjoy working with them. So both your peers and your boss, and to the extent that you can assess people across the aisle on the business side, are these really top-notch people both in terms of your interaction with them and in terms of what they've done in the past?

And then the second thing is you see who is funding them, and again it can be difficult in the beginning if it's angel funding and so on. On the other hand, if one of the top VCs in the world is funding a company, you know at least that some folks who are extremely savvy have decided to make this bet. So those are a kind of two quick shorthands that help you tell whether it's worth considering joining.

Foster Provost: Yes. The next question up now actually turns out to be this question turned around—so let me address Claudia, because I know that she's been working on this, but I think probably everyone here can then contribute.

Part 1: What can you do to recruit smart people to your company?

Part 2: How can a start-up compete for talent against well-established companies, today?

“ARE YOU GOING TO BE ON THE SIDELINES WRITING REPORTING FOR A PRODUCT THAT HAS NOTHING REALLY CORE TO DO WITH ANALYTICS? OR ARE YOU GOING TO BE A CORE PART OF THE PRODUCT?”

Claudia Perlich: We recently hired two additional data scientists (for a total of six now) and that's an incredibly hard position to be in, which I'm sure all the industry people can testify to. What seems to have worked best for us is actually putting ourselves out there—so I'm going and giving talks at meet-ups and conference and so on, so I'll

share the excitement of the work and I'll meet people who are interested and want to know what good data science places are. That has worked very well on my side.

Now if you don't already have a data scientist, that's a tough proposal. If you're trying to hire your first one, you probably want to find somebody else who has a good data scientist to advise you on how to do this, because most companies couldn't tell a good data scientist if they saw one.



Foster Provost: You guys have any counter-opinions or things to add?

Usama Fayyad: Yes, look, I think the formula for hiring good people is the same, and I've done it from the companies and I've done it from start-ups; it's passion, vision, and excitement. You couple that with the upside, including the risk, people really, really will surprise you—the recruits will surprise you positively. I've done it with start-ups where I've taken them away from Microsoft and Amazon and Google. I've done it at Yahoo!, and we had to build the data group. I've done it on a bigger scale when we have had to build Yahoo! Research, where we actually convinced a whole bunch of people to leave university careers and join a newly formed lab, and it was really based on the scope of the vision, the ability to do stuff that you couldn't do anywhere else.

And every start-up will have a unique niche, something they are focused on, where you can't do it at the big company, and what you want to do is really attract the people who resonate with that, so I think passion and excitement and the big vision are very, very big in attracting people.

Geoff Webb: OK, let's move on. We've got another audience question, and I'm going to direct this one to Usama because you are involved in start-ups on opposite sides of the world, in both the West and the Middle East. So the question is, how much does the location of a data mining start-up matter?

Usama Fayyad: Oh, excellent question. I would argue that for a data mining or even a big data company, the location doesn't matter, and here's why. Maybe the norm is that you want to be where the talent is and all that kind of stuff. I'm going to share something with you. I mean, I've been out of Yahoo! now long enough that I can share something internal with you. In my last year at Yahoo!, I think I made a lot of enemies; as an EVP and chief data officer I came up with a rule and I had two pretty big groups reporting to me. Essentially no one could hire any engineer in the Bay Area without my approval as EVP and officer of the company. People complained and people made noises and all of that, and the reason for that is yes, Silicon Valley is the heart-bed. Silicon Valley is where a lot of talent is and where a lot of excitement is. But my ultimate judgment was: overinflated titles, overinflated salaries, overinflated talent claims, and no ability to retain whatsoever.

So we went to very strange places to find talent. We went and acquired a whole company in Urbana-Champaign, Illinois, to get talent. When I did my own start-ups, I had teams in places like Chile, Jordan, and Syria when it was okay to do that. You can find talent if you train them and you can retain them. The retention has a lot of value in deep training. At some point, India was a good place to do it; now India is almost like Silicon Valley or at least Bangalore is, so location

matters from the sense of watching your cost and being able to retain people, as opposed to going to the hotspots. That's for data science.

Now it's a completely different discussion when it comes to marketing and where you have to have your salespeople and all of that. That's different; design, sales, marketing is a different ballgame. But here we're talking about the data science talent.

Foster Provost: I'd like to move on because I really like the question that just bubbled up to the top here. Is it still worth pursuing a PhD as a data scientist? Oren?

Oren Etzioni: Yes. I've been advising PhDs now for more than 20 years. At a certain point for every one of my students, I take them aside and I try to persuade them that they should not get a PhD. It's a major investment of time as you very narrowly focus. What's the point? You could be out there involved in these ventures, making a lot of money.

All the students who have graduated with their PhDs have passed that gauntlet and decided to stay. So why? One reason is if you want to be a professor, a PhD is one of the things you still need. Also, in certain companies it gives you access to certain kinds of problems.

But often it does not make sense to get a PhD. I'll just give you one quick example. I had a student I worked with, an incredibly, incredibly talented hacker, deep understanding of data science, the math, the statistics, just not a great researcher. I said to him: go do what you're good at. And he for his own personal reasons insisted on getting the PhD. We decided not to work together. He went and did a PhD in robotics. He got his PhD and then he went to Google. And he went to Google about three or four years later than he would have otherwise, and that was several million dollars.

So basically, if you want to get a PhD, the burden of proof is on you. Why? It's not like God decreed you should have a PhD if you're smart. Why should you get a PhD?

Foster Provost: Usama do you agree?

Usama Fayyad: OK. I know, Foster, you probably should answer this as a professor but—I'm not a professor, I play one on TV. Here is what I will say. My advice I think is not going to be very popular with academia, but I think in the long-term it's in the benefit of academia. I think if you've just finished your master's and you're thinking of doing a PhD, especially in the field we're in, my strong advice is: don't do it. Go get a job. Go get a job for five years and then come back if you really want to and do the PhD, because here's what happens.

You have no idea what real problems are about. I think 80–90% of thesis topics are totally useless, and it's not because

the people are bad. It's just that they don't know better. They just didn't see what's happening. They didn't understand what the deep problems are. I actually think if people work for 5–10 years and then have it in them to go do a PhD, they will do better research. The professors will get better articles; the industry will benefit a lot more; even theory will be more relevant.

I'm a huge believer that the most fundamental theories lie in the most mundane of little details of applications and building systems and things like that, and great discoveries in science probably attest to that, but you will do yourself a favor and it will be a much better investment of your valuable, valuable time that Oren talked about if you do that PhD just a bit later.

Don't wait too long because then you'll lose your brains but ...

Oren Etzioni: I just wanted to strongly disagree with that. Finally, we have a good disagreement!

The problem with what Usama said is that people don't come back. I mean, it's a statistical fact. After 5 years, even after three years, people just don't come back. It's not to say that you don't learn a huge amount like he said, but if you leave the program you don't come back. So again I just said, you need to think through why you want to do it. But don't leave and say "Oh, I can always come back." Nobody comes back after 5 years.

Ron Bekkerman: I think that doing a PhD in the area of data mining is very helpful for being a good data scientist—just for one reason. When you're doing your PhD, you are pretty much on your own. You need to do good research and you need to have your research published. You're reaching a certain level of excellence so that you can have your articles published at KDD and other top venues, and you have much more confidence in yourself after you have done this. Each article is a small start-up of itself—after all, you invest about half a year of your life, you do all your experimentation, and then you "sell" it, so you're doing everything from prototyping to marketing.

Once you have done this 5–10 times and your articles are getting accepted, you have a lot of confidence to come to the industry and say, "You know what, guys? I can do it."

Claudia Perlich: I spent 6 years getting a PhD, and I sure don't regret it. I probably would've been happy doing

something else with my life too, but I really love the time I had. And from the perspective that Ron brought up, good data scientists take seasoning. It's a craft. You have to go through the potholes. You have to make all the mistakes. I can't teach you all the mistakes—I can't prevent it. If you're straight out of your master's and have taken two data science courses or maybe five, you're still going to have to learn an awful lot of hard lessons.

From a hiring perspective, I like people with a PhD exactly for that fact. They did this stuff for 5 years, and they have learned certain things.

Of course, somebody who was in industry doing data science for 5 years may be even better, but you have to start somewhere, so if you don't have that experience, for me PhD is a value proposition as I look at it.

Foster Provost: Actually, let me jump in as well because I think there's a distinction that's generally not made when we talk about data scientists these days. If you look in companies, there are two very different sorts—well, there's probably more than two but at least two—very different sorts of data scientists. There are people who are essentially what I would call data science engineers. Their primary job is building things. Then there are people who we might say are the data scientists proper—essentially researchers, whose job is largely coming up with new ideas, evaluating them, and so on.

For the former, getting a PhD is a "nice to have."

For the latter, as Claudia points out, doing data science research is a craft and, just like most of the mature crafts, the best way to learn it is via an apprenticeship. One way to get that apprenticeship is by apprenticing with a really good professor. So, just getting a PhD where somebody tells you what to do and you go and do it isn't going to help you out very much. But having a really good professor who oversees a good apprenticeship can lead you to go on and be a great data scientist.

And so it's not really just getting a PhD. It's, do you get a good apprenticeship? You could also get a good apprenticeship not through a PhD but by going into a company where there's a fantastic chief scientist who really mentors her people. So I don't think we should think about it as the label PhD. It's whether you get a good apprenticeship such that you can come up with good ideas, design the right systems, put together the right evaluations, and so on.

Geoff Webb: OK. We've got more questions, and we're running out of time. To save time, I want to ask you each to address this in point form, so just give me a list of the

"I ACTUALLY THINK IF PEOPLE WORK FOR 5–10 YEARS AND THEN HAVE IT IN THEM TO GO DO A PHD, THEY WILL DO BETTER RESEARCH."



points. So, we'll start with Claudia; what are the critical success factors for data science-based start-ups?

Claudia Perlich: The first one is: make sure data science is core to the product. The second, you need to have an amazing team and you need developers who have covered the data science space when they built things. Your excellent Java programmer doesn't necessarily get you what you need when building a good data science environment. There is a skill that even the programmers need around handling data, so you need a team that has that experience and preferably has done that before.

Geoff Webb: Point form, Ron?

Ron Bekkerman: Again bringing the VC perspective, there are three very important points. The first one is the size of the addressable market. If, say, you have a start-up that sells to bookstores in Chicago, how many bookstores in Chicago do you have? One hundred? Probably not 1 million, and each bookstore is unlikely to pay you \$1 million annually, so your addressable market is not that big. Another point is your go-to-market strategy—how you actually start selling and whom you are selling to. And the third point is your pricing: If you sell to a billion people, but each person pays you one penny a year, you will be making \$10 million a year, which in the VC start-up world is not considered a great success.

Geoff Webb: Point form, Oren?

Oren Etzioni: So we talked about team, but I want to emphasize this is not all equal. Who is the most important player in the Miami Heat? LeBron James.

Who is the most important person on the team in a start-up? It's the CEO. And by orders of magnitude. So look very carefully at the CEO.

Geoff Webb: Point form, Usama?

Usama Fayyad: I will add to what was said by Ron—execution. So many people have great ideas and great business models; the market size is there and ideas are so cheap and so easy to have. Execution is really the distinguishing factor. I think it was Warren Buffett who said, "I'll invest in a boring idea with good execution any time over a good idea with no execution."

Foster Provost: All right. Thank you.

Next question from the audience: For a graduating student, which path is better toward starting a company in the future?

(1) Go to a research lab and then start the company later. Or (2), go to a big company and then start the company later.

Oren Etzioni: If you're graduating soon, send me your resume. The competition is intense. I mean, for trying to find people.

Ron Bekkerman: I guess the advice is, just don't start your company right away. Get some experience elsewhere—as much as you can—and then you will be in good shape to start a company.

Claudia Perlich: So exactly what Foster brought up earlier, and he phrased it much better than I did: Where do you get good mentorship? Where are you going to learn the most and make the most of these next few years? And it's not exactly clear whether this is a research lab or it's an industry job. It could be either, but take a very close look at who you are going to work with and what you can hope to learn.

Usama Fayyad: I mean, excellent points. I would add, I'm a great believer in gradient descent, so there are two major directions here in learning. Learnings you get from a start-up are almost orthogonal, sociologically and so forth, workwise than what you learn from a big company, and both are extremely important. So if you have the choice—I completely agree with Ron—don't do your start-up right out of school. Have a little bit of experience beforehand. With a little bit of experience, you'll do the start-up faster, better, cheaper, and ...

Foster Provost: And remember, in your career it's stochastic gradient descent!

Geoff Webb: OK. The top question remaining is, can you give any tips to a technical founder who wants to get early-stage funding? Let's go from Usama.

“SO MANY PEOPLE HAVE GREAT IDEAS AND GREAT BUSINESS MODELS; THE MARKET SIZE IS THERE AND IDEAS ARE SO CHEAP AND SO EASY TO HAVE. EXECUTION IS REALLY THE DISTINGUISHING FACTOR.”

Usama Fayyad: Yes, quick tips. If you're a technical founder, then get yourself a business cofounder, preferably a very strong CEO. And we say this all the time, as we get a lot of business people who want to do a technology company and the first thing we tell them is, "We won't go talk to you until you go get yourself a technical cofounder," and they usually just go and get themselves an employee and we say, "No. That doesn't count." Cofounder means they own a good chunk of

the company. So I think it's the team. Get that right team together. It will help you build what's important, which is the idea and the execution. Money will come. If you have a good

idea and the execution and the team, I think money is not a problem.

Geoff Webb: OK. Do any of the other panelists want to give a counterpoint to that? Otherwise, we'll move on.

Oren Etzioni: The thing that I would add is it's not always easy for technical people to find that great nontechnical co-founder, so sometimes the whole investment community and network is less about money but more about connections and good questions. So you'll go to them and they'll say—we're not ready to fund you, but why don't you talk to these three people, who will lead to these nine people, who will lead to the team. So I'm actually a great fan of talking to angels, talking to investors early. If they don't give you the money, they can give you helpful connections.

Ron Bekkerman: And I second it. A venture capital company is not going to invest in you, because they invest in maybe two companies a year, so just work for an outreach, talk to all the venture capital companies and all the angels and to all the venture capital and angel networks.

Foster Provost: A question from audience member Steve is: Should a start-up patent ideas or worry about patent lawsuits?

Oren Etzioni: They would not be my primary focus. In software, there are so many ways to get around patents; there's so much money and time you can spend writing meaningless patents; it's a 5-percent activity. Unless you've really got the intermittent windshield wiper, that would not be my focus.

Ron Bekkerman: Patents are on VC's checklist, so you have to have them. Check.

Geoff Webb: I think the next one's for Ron. Can you provide tips to a data scientist who joins a big well-established start-up company?

Ron Bekkerman: You need to negotiate your deal, and it's quite crucial. You're getting salary, and you're getting stock options. Here I disagree a bit with Claudia. Your salary—well, you have to have it, but the actual amount of money doesn't matter that much. Say it's \$150K nowadays—that's what a data scientist is getting as far as I understand. It doesn't matter if it's \$120K or \$160K, because over the start-up's lifecycle, those differences would sum up into \$100K at most. On the other hand, the amount of stock options that you are getting is really crucial. Even if we're talking about a big start-

up that is probably planning to go IPO in the near future, and you are unlikely to get a lot of stock options, they still matter. Every stock option is translated into a certain amount of money, say, \$10 to \$100. If you own 1,000 stock options, you are just not making that much money, but if you own 1 million options, this can make you rich. Since stock options are not cash; companies are generally more flexible in negotiating your stock option plan, and you don't want to lose it. After all, it might be your only opportunity to make real money.

“WALK UP TO THE BOSS AND SAY, LOOK, I’LL GET YOU X IN WHATEVER—REVENUE INCREASE, PROFITABILITY INCREASE, WHATEVER—IF YOU SHARE WITH ME Y.”

Usama Fayyad: I actually want to emphasize what Ron said about negotiating the package. That's one of the things that I think data scientists don't do enough. I've done it, and I've seen people do it, and I've rewarded them for doing it—walk up to the boss and say, look, I'll get you X in whatever—revenue increase, profitability increase, whatever—if you share with me Y. You'd be surprised. Even the stingiest bosses,

if they believe the number you're promising is not possible, they'll do it, and that's how you'll generate wealth.

Foster Provost: So share with us some: what Ys are reasonable? What can people actually?

Usama Fayyad: Well, I'm not going to name the company, but I was in a situation where somebody walked up to me and said, “We will make an extra \$35 million in one year and in return if I make you this 35 million, I want 10 million.”

And you know what I said to this guy?

“OK, make me 35 million. I'll give you 2 million.”

And he said, “Yes.”

Oren Etzioni: Here's the thing again, particularly for this audience about negotiation. Negotiations are really important. Coming out of academia, we're often not used to negotiation and that puts us at a disadvantage. Let me give you one tidbit, which is midlevel advice—actually, let me give you two.

One thing Ron mentioned, he mentioned a number of stock options. Number of stock options is meaningless. You need to understand what it is as a percentage of the company. That's the only number that has meaning. Even if it's 0.001%, you need to understand it as a percentage of the company and what it's worth. For a data scientist, the key to negotiation is data. So what does that mean?



If I'm negotiating a stock option package, and like Ron said, it's incredibly important, first of all, what have other people in the company gotten? That's a piece of data that makes a big difference. And what's the comparable thing at the place across the street?

Those pieces of data are invaluable in your negotiation. And negotiation is not only about blustering and blah, blah; a lot of it is about data. If you can present that data, that really enhances your negotiation position.

Foster Provost: So if we put up a site where, anonymously, everybody can put in their salary and the number of stock options and so on, will you all contribute to it?

Oren Etzioni: There are, of course, sites, so PayScale.com—very useful site. Glassdoor.com is a very useful site if you're looking for jobs. These are two Seattle start-ups that help a lot with this.

Foster Provost: We're getting to the end of our time. So I wanted to jump down to a question about the future.

Where's the money coming from in the future in terms of sort of industry sector? Still from advertising? We have healthcare; we have other big segments that are very interesting from the point of view of doing data science. Where do you guys see the money coming from for start-ups in the next 10 years?

Oren Etzioni: The enterprise.

Ron Bekkerman: Just like in the previous 50 years.

Usama Fayyad: Oren, are you saying there's no money in consumer Internet?

Oren Etzioni: I see a lot of money coming from the enterprise. There's not just one answer, but I was trying to be brief.

Usama Fayyad: The golden rule for money is the following. If you understand the customer pain and that customer can be a consumer, that customer can be an enterprise, that customer can be a government, whatever you like. If you un-

derstand the customer pain and you have something that solves it, this is what we keep telling our companies, we invest in companies that make painkillers, not vitamins. Vitamins are good for you; it's optional, you can take them or not. If you're in pain, baby, you're going to take that painkiller and you're going to pay for it, so ...

Oren Etzioni: No pain, no gain.

Foster Provost: All right, I think in order that we wrap things up on time, let me turn it over to Geoff.

Geoff Webb: All things come to an end, even great panels, and so I'm going to have to ask the final question. We're almost out of time, so I'm going to ask you all to answer this in less than 30 seconds, so one or two sentences, and the question is: *Are start-ups a good way to monetize data science expertise?*

Claudia Perlich: Absolutely.

Ron Bekkerman: It depends.

Oren Etzioni: Yes.

Usama Fayyad: Monetization, guaranteed; you can make a lot of money from big companies if you're up on risk and you understand customer pain; you can make a lot more money with a start-up. So either way the good news is, in data science, you're going to make money.

Foster Provost: Thank you all for your participation.

Address correspondence to:

*Foster Provost
Leonard N. Stern School of Business
New York University
44 West Fourth Street, 8th Floor
New York, NY 10012*

E-mail: fprovost@stern.nyu.edu