

## **Analytics and Data Science**

### *Overview*

How does LinkedIn know who your 2<sup>nd</sup> and 3<sup>rd</sup> degree connections are and exactly how you are connected to them? How does Facebook target advertisements to your newsfeed based on websites you visited that day? How does Google know how to autocomplete the terms you enter into the search field? We are living now in the age of Big Data, and businesses need Big Minds to code and collect, store, analyze, interpret, report and utilize those data in ways that can improve their products and services.

The Institute for Operations Research and the Management Sciences (Informs), a professional organization for those working in operations research (O.R.), management science, and analytics, defines analytics as “the scientific process of transforming data into insight for making better decisions.” Analytics uses techniques such as mathematical modeling, data simulation, and probability and statistics to mine complex data sets for answers to complicated business questions. The emerging field of data science is not as clearly defined, but can be thought of as a special case of analytics. Data science requires mathematical, engineering, technological, communications, analytical, and business skills—which together are used to link multiple datasets from disparate sources and identify patterns and solutions across them.

The successful data scientist is one part programmer, one part statistician, one part creative, and thinks both like a CEO seeking solutions to business problems and a research scientist experimenting with the cyber world.

### *Why graduate skills are necessary*

Academics are trained to ponder the *What If* questions: *What if we manipulate variable X? What changes would we then observe in Y? What if event X had occurred before event Y, historically? Would the outcome of event Y have been different?* This is precisely the sort of thinking that is required in analytics and data science, in order to give businesses a competitive edge or discover solutions to pressing business problems.

Thus, not only are the hard skills learned in computer science and mathematics programs necessary for data scientists, but skills in experimental design and hypothesis-testing are also highly pertinent. Importantly, graduate students are taught to think outside the box, to be innovative and persistent in making sense of the universe. These skills are highly valued in analytics and data science.

### *Key players*

- Data and computing companies such as Cloudera and EMC<sup>2</sup>
- Social media platforms such as Google, Facebook, Twitter, LinkedIn, and Instagram
- Tech companies such as Apple, Microsoft, IBM
- Hospitals and hospital systems
- Insurance companies
- Government research

### *Resume best practices*

When seeking a position in analytics or data science, this is the resume where the section on “skills” will be important. Successful candidates will demonstrate experience with at least one programming language (e.g., Java Script, Python or R), applied statistics, and mathematics. The resume should also indicate that the candidate has had experience applying those skills creatively and effectively.

### *Cover letter best practices*

The cover letter is the space to demonstrate the human side of the machine. That is, having demonstrated the technological, computational, and statistical skills on the resume, candidates have the opportunity to show in the cover letter how they approach solving real problems with those skills. Use this space to describe how you think about data, and especially how they speak to you about consumer behavior. Show that you have done your research on the company by describing a challenge or a problem it might face in its operations, and what sorts of questions might be asked of the data in working to identify solutions to those problems. Let your creativity shine!

### *Final five things to know*

- Hadoop (Don't know what it is? Learn it, now!)
- “Big data’ typically refers to data on the scale of terabytes (10 to the 12th power) and petabytes (10 to the 15th power). A petabyte is a million gigabytes.”  
(<http://datascience.berkeley.edu/about/what-is-data-science/>)
- The explosion of Big Data will require people to fill roles other than data scientist, so it is worthwhile to get to know this emerging field and all its emergent options.
- Technological skills can be obtained through specialized courses and certification programs, and thus are easier to add onto a degree in another field than vice versa.
- Starting salaries are high, and jobs are proliferating rapidly.

### *Links to web resources to learn more*

- <http://informs.org>
- <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>
- <http://datascience.nyu.edu/what-is-data-science/>
- <https://medium.com/@hannawallach/big-data-machine-learning-and-the-social-sciences-927a8e20460d>
- <http://dataconomy.com/seven-vs-big-data/>
- <http://www.forbes.com/sites/gilpress/2012/08/28/the-data-science-interview-mingsheng-hong-hadapt/>
- <http://research.microsoft.com/apps/video/default.aspx?id=174224&r=1>
- <http://www.technologyreview.com/featuredstory/428150/what-facebook-knows/>

### *Links to web resources to find positions*

- <http://analyticaltalent.com>
- <https://www.kaggle.com/forums/f/145/data-science-jobs>

- <http://www.cybercoders.com/jobs/data-scientist-jobs/>