Insight Data Science: Bridging the gap between academia and industry

Josiah Walton
Physics Careers Seminar
UIUC Department of Physics
April 23, 2015
Why data science?

- Chance to positively impact people’s lives
- Variety: learn many new skills and areas
- Pace of work
- Location
- Compensation
- Many opportunities
Update: I’ve accepted a data scientist position at Biogen Idec
The Rise of Data Science
# Big Data

## Stored data in the United States, 2009

<table>
<thead>
<tr>
<th>Industry</th>
<th>Stored data (Petabytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete manufacturing</td>
<td>966</td>
</tr>
<tr>
<td>Government</td>
<td>848</td>
</tr>
<tr>
<td>Communications and media</td>
<td>715</td>
</tr>
<tr>
<td>Process manufacturing</td>
<td>694</td>
</tr>
<tr>
<td>Banking</td>
<td>619</td>
</tr>
<tr>
<td>Health care providers</td>
<td>434</td>
</tr>
<tr>
<td>Securities and investment services</td>
<td>429</td>
</tr>
<tr>
<td>Professional services</td>
<td>411</td>
</tr>
<tr>
<td>Retail</td>
<td>364</td>
</tr>
<tr>
<td>Education</td>
<td>269</td>
</tr>
<tr>
<td>Insurance</td>
<td>243</td>
</tr>
<tr>
<td>Transportation</td>
<td>227</td>
</tr>
<tr>
<td>Wholesale</td>
<td>202</td>
</tr>
<tr>
<td>Utilities</td>
<td>194</td>
</tr>
<tr>
<td>Resource industries</td>
<td>116</td>
</tr>
<tr>
<td>Consumer &amp; recreational services</td>
<td>106</td>
</tr>
<tr>
<td>Construction</td>
<td>51</td>
</tr>
</tbody>
</table>

## Number of firms with >1,000 employees

<table>
<thead>
<tr>
<th>Industry</th>
<th>Number of firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete manufacturing</td>
<td>1,000</td>
</tr>
<tr>
<td>Government</td>
<td>647</td>
</tr>
<tr>
<td>Communications and media</td>
<td>399</td>
</tr>
<tr>
<td>Process manufacturing</td>
<td>835</td>
</tr>
<tr>
<td>Banking</td>
<td>321</td>
</tr>
<tr>
<td>Health care providers</td>
<td>1,172</td>
</tr>
<tr>
<td>Securities and investment services</td>
<td>111</td>
</tr>
<tr>
<td>Professional services</td>
<td>1,478</td>
</tr>
<tr>
<td>Retail</td>
<td>522</td>
</tr>
<tr>
<td>Education</td>
<td>843</td>
</tr>
<tr>
<td>Insurance</td>
<td>280</td>
</tr>
<tr>
<td>Transportation</td>
<td>283</td>
</tr>
<tr>
<td>Wholesale</td>
<td>376</td>
</tr>
<tr>
<td>Utilities</td>
<td>129</td>
</tr>
<tr>
<td>Resource industries</td>
<td>140</td>
</tr>
<tr>
<td>Consumer &amp; recreational services</td>
<td>708</td>
</tr>
<tr>
<td>Construction</td>
<td>222</td>
</tr>
</tbody>
</table>

## Stored data per firm (>1,000 employees), 2009

<table>
<thead>
<tr>
<th>Industry</th>
<th>Stored data (Terabytes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete manufacturing</td>
<td>967^2</td>
</tr>
<tr>
<td>Government</td>
<td>1,312</td>
</tr>
<tr>
<td>Communications and media</td>
<td>1,792</td>
</tr>
<tr>
<td>Process manufacturing</td>
<td>831^2</td>
</tr>
<tr>
<td>Banking</td>
<td>1,931</td>
</tr>
<tr>
<td>Health care providers</td>
<td>370</td>
</tr>
<tr>
<td>Securities and investment services</td>
<td>3,866</td>
</tr>
<tr>
<td>Professional services</td>
<td>278</td>
</tr>
<tr>
<td>Retail</td>
<td>697</td>
</tr>
<tr>
<td>Education</td>
<td>319</td>
</tr>
<tr>
<td>Insurance</td>
<td>870</td>
</tr>
<tr>
<td>Transportation</td>
<td>801</td>
</tr>
<tr>
<td>Wholesale</td>
<td>536</td>
</tr>
<tr>
<td>Utilities</td>
<td>1,507</td>
</tr>
<tr>
<td>Resource industries</td>
<td>825</td>
</tr>
<tr>
<td>Consumer &amp; recreational services</td>
<td>150</td>
</tr>
<tr>
<td>Construction</td>
<td>231</td>
</tr>
</tbody>
</table>

Estimated demand by 2018

140,000–190,000
more deep analytical talent positions, and

1.5 million
more data-savvy managers
needed to take full advantage
of big data in the United States

What is Data Science?

“Nobody knows what it means, but it’s provocative”
What is Data Science?

Field still nascent
Role continually defined
Varies from company to company

Drew Conway
What do data scientists do?
Leverage critical thinking and analytical expertise to ask piercing questions

Extract “actionable insights” with real business value from complex data sets
Many Questions in Data Science

- **Khan Academy**: Millions of math problems are attempted online each day.
  - When is someone proficient?
  - What is the best problem order?

- **The New York Times and Wall Street Journal**:
  - What is the right payment model for journalism?
  - How do we prevent customer churn?

- **LinkedIn**:
  - Does this LinkedIn account correspond to a real person?

- **Netflix**:
  - What new shows should we make?

- **Everywhere**:
  - What are some efficient ways in which the data can be stored, accessed, and streamed?
Where do data scientists come from?
“The best data scientists tend to be ‘hard scientists’, rather than computer science majors.”

- DJ Patil

- Chief Data Scientist at U.S. OSTP
- Former head of data science at LinkedIn
- Insight mentor
- PhD mathematician
- Coined “data science”
Founded 2012

NYC
Palo Alto, CA
insightdatascience.com

Palo Alto, CA
insightdataengineering.com

Boston, MA
insightthehealthdata.com
Applying to Insight

- **Initial Stage** - Formal application:
  - CV
  - Experience with programming, machine learning, etc.
  - Why you’d like to participate

- **Final stage**: Interview

- **Competitive**: ~5% acceptance (growing though!)

### Upcoming Sessions

- **Data Science**
  - SILICON VALLEY & NYC
  - Sept. 2015 (Deadline: June 29, 2015)

- **Data Engineering**
  - SILICON VALLEY
  - Sept. 2015 (Deadline: June 29, 2015)

- **Health Data Science**
  - BOSTON
  - July 2015 (Deadline: May 4, 2015)
Program Timeline

7 Weeks

Week 1  Weeks 2-4  Weeks 5-7

Intro Week  Data Project  Company Visits

4-8 Weeks

Job Interviews
The Bridge

1. Tools
2. Context
3. Pace
4. Priorities

You (~90% of skills)

Insight (last ~10% of skills)

Data Project

Data Science Career

Personal Portfolio

- Machine Learning
- New languages/tools
- Short time scale
- Concrete Deliverable
- Acceptable Performance
Outcomes

• Individual data product, effectively walks you into interviews

• Extensive job application and mock interview preparation

• Outstanding professional network (cohort and leading data scientists)

• Interest from many recruiters

• 100% job placement (usually within 2 months)

• Starting salary: $110,000 $120,000 $150,000
200+ Insight Fellows working as Data Scientists and Data Engineers
Insight Companies

Memorial Sloan-Kettering Cancer Center
NBC
Netflix
LinkedIn
Facebook
Twitter
Microsoft
Airbnb
Palantir
Apple
Square
MTV
Khan Academy
Yelp
Spotify
The New York Times
+Many Startups…
Doing an internship while in grad school
Difficult

- Loss of research time
- Potentially delayed graduation
- Incurred debt

Rewarding

- Experience
- Networking
Important Questions

- Applicability of new skills?
- Teaching/research responsibilities?
- Pay?
- Work while away?
- Time spent away?
- Current lease/renting situation?
Ideally light at end of the tunnel
It’s your life...
Don’t Fret!
You don’t have to get into an internship like Insight to find an awesome career in data science!
What you do possess are a very particular set of skills...

- Analytical mindset
- Innate curiosity and passion
- Communication skills
- Team work mentality
- Adaptability
- Technical expertise (math, stats, complex/messy data. etc.)
- ...

...
Tips and Tactics

• Start early!

• Learn open source software (python, R, etc.)

• Build knowledge of machine learning, statistics, comp sci fundamentals, etc., using library, free blogs, coursera, youtube, etc…

• Accumulate evidence of your skills (learn by doing!):
  • Develop real projects
  • Post work on Github
  • Write personal blog

• Connect (network!): LinkedIn, Twitter, meetups

• Informational interviews: Vocabulary, substantive expertise, tools

• Learn to sell yourself and accomplishments!
Data Science Job Interviews

• Interview prep: practice, practice, practice (mock interviews)

• **Be confident and charismatic!** Hiring process is much about human nature as intellectual ability

• Typical interview route (assuming your resume makes it):
  1. Phone screen (HR and or technical questions)
  2. *Data challenge*
  3. On-site presentation and meetings
Ending thoughts

• Get in touch with me if you’re interested in Insight or want to talk about data science in general!

• Connect with me on LinkedIn!

• Thanks for listening!