Careers in Data Science
Transitioning from Academia to Industry

Zach Gazak, Wafa Soofi
Insight Data Science

PIEP Workshop
October 8, 2015
Data Is Growing Exponentially
Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte

<table>
<thead>
<tr>
<th>Category</th>
<th>Stored data in the United States, 2009</th>
<th>Number of firms with &gt;1,000 employees</th>
<th>Stored data per firm (&gt;1,000 employees), 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete manufacturing</td>
<td>966</td>
<td>1,000</td>
<td>967</td>
</tr>
<tr>
<td>Government</td>
<td>648</td>
<td>647</td>
<td>1,312</td>
</tr>
<tr>
<td>Communications and media</td>
<td>715</td>
<td>399</td>
<td>1,792</td>
</tr>
<tr>
<td>Process manufacturing</td>
<td>694</td>
<td>835</td>
<td>831</td>
</tr>
<tr>
<td>Banking</td>
<td>619</td>
<td>321</td>
<td>1,931</td>
</tr>
<tr>
<td>Health care providers</td>
<td>434</td>
<td>1,172</td>
<td>370</td>
</tr>
<tr>
<td>Securities and investment services</td>
<td>429</td>
<td>111</td>
<td>1,478</td>
</tr>
<tr>
<td>Professional services</td>
<td>411</td>
<td>522</td>
<td>278</td>
</tr>
<tr>
<td>Retail</td>
<td>364</td>
<td>843</td>
<td>697</td>
</tr>
<tr>
<td>Education</td>
<td>269</td>
<td>280</td>
<td>319</td>
</tr>
<tr>
<td>Insurance</td>
<td>243</td>
<td>283</td>
<td>870</td>
</tr>
<tr>
<td>Transportation</td>
<td>227</td>
<td>376</td>
<td>801</td>
</tr>
<tr>
<td>Wholesale</td>
<td>202</td>
<td>129</td>
<td>536</td>
</tr>
<tr>
<td>Utilities</td>
<td>194</td>
<td>140</td>
<td>825</td>
</tr>
<tr>
<td>Resource industries</td>
<td>116</td>
<td></td>
<td>1,507</td>
</tr>
<tr>
<td>Consumer &amp; recreational services</td>
<td>106</td>
<td>708</td>
<td>825</td>
</tr>
<tr>
<td>Construction</td>
<td>51</td>
<td>222</td>
<td>231</td>
</tr>
</tbody>
</table>

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

“The best data scientists tend to be ‘hard scientists,’ rather than computer science majors.”

- DJ Patil
WHAT IS DATA SCIENCE?

Drew Conway
THE GAP

1. Tools
2. Context
3. Pace
4. Priorities
BRIDGING THE GAP

INSIGHT
EVIDENCE OF INTEREST

Build a data project.

Show off your:

- Initiative
- Technical Ability
- Product Sense
GATHER INTEL

Do some informational interviews.

Ask questions like:

What is your data science stack/toolset?

Do you work on production-level tools?

How big is the data science team at your company?
CONNECT

Stay informed.

Check out:

Meetup groups (or start your own)

Hacker News

Data Science Weekly

LinkedIn
EVIDENCE OF INTEREST
Build a data project

GATHER INTEL
Do some informational interviews

CONNECT
LinkedIn, Twitter, meetups
Insight Fellows working as Data Scientists and Data Engineers
Insight Fellows are Data Scientists and Data Engineers at:

- LinkedIn
- Facebook
- Twitter
- Microsoft
- Square
- Palantir
- NBC
- Airbnb
- Netflix
- Intuit
- Tripadvisor
insightdatascience.com

Upcoming Sessions:
SILICON VALLEY
NEW YORK
SEATTLE, LA, DC, CHICAGO, & Other Cities
January 2016, June 2016

INSIGHT DATA SCIENCE FELLOWS PROGRAM

An intensive seven week post-doctoral training fellowship bridging the gap between academia and data science

WHITE PAPER
APPLY NOW

Want to be notified of future dates? Click Here
Program Timeline

7 Weeks

Week 1
Intro Week

Weeks 2-4
Data Project

Weeks 5-7
Company Visits

4-8 Weeks
Job Interviews
AREAS OF FOCUS

Product Analytics
Predictive Modeling (e.g. Risk, Sales)
Data Products
MANY QUESTIONS IN DATA SCIENCE

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

**The New York Times and The Wall Street Journal:** What is the right payment model for journalism?

**LinkedIn:** Does this LinkedIn account correspond to a real person?
MANY QUESTIONS IN DATA SCIENCE

Netflix: What new shows should we make?

Memorial Sloan Kettering Cancer Center: Who is a good candidate for clinical trials?

Everywhere: What are some efficient ways in which the data can be stored, streamed, or accessed?
DATA PROJECTS

• lendinglounge.rodeo
• reviewskimmer.com
• hammerpricer.com
• beautifulcity.me
• routearoundcrime.com