Data Scientist: A Career for 2015 and Beyond

Richard F. Eng
PRINCE2, PMP, CSQE, CRE, CQE
reng@mitre.org
703-201-9112
20 March 2015

The author’s affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE’s concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author.
Outline

- Hype or Reality?
- Data Scientist
- Opportunities
- Unicorns
- Acquiring Unicorn Skills
- Challenges
- Conclusions
- Next Steps
Hype or Reality?

- “Data Scientist: The Sexiest Job of the 21st Century”
  – Thomas H. Davenport and D.J. Patil

- “Analytics is defined as the scientific process of transforming data into insight for making better decisions.”
  – The Institute for Operations Research and the Management Sciences (INFORMS)

- “With more and more companies using big data, the demand for data analytic specialists,—sometimes called data scientists, who know how to manage the tsunami of information, spot patterns within it and draw conclusions and insights—is nearing a frenzy.”
  – Chris Morris, CNBC
"A person who is better at statistics than any software engineer and better at software engineering than any statistician."

Josh Wills
Director of Data Science at Cloudera

"Data scientists are inquisitive: exploring, asking questions, doing “what if” analysis, questioning existing assumptions and processes. Armed with data and analytical results, a top-tier data scientist will then communicate informed conclusions and recommendations across an organization’s leadership structure."

Anjul Bhambhri, IBM
“Of the almost 2 million openings he expects over the next three years in the U.S. alone (4 million worldwide), Peter Sondergaard (Gartner) predicts that only about one-third will be filled, making analytics software whizzes “a scarce, valuable commodity” that employers will have to fight to hire and retain.”

What Do Data Scientist Do?

- Interface with analytics, product management, and operations teams.
- Perform large-scale data analysis and develop effective statistical models for segmentation, classification, optimization, time series, etc.
- Design and implement reporting dashboards that track key business metrics and provide actionable insights.
- Identify actionable insights, suggest recommendations and influence the direction of the business by effectively communicating results to cross functional groups.
- Work closely with Product or Engineering & Operations teams to proactively create rule and manage decisions.
- Suggest improvements in the tools and techniques to help scale the team.
What Do Data Scientists Do? (cont.)

- Apply data-mining, machine learning and/or graph analysis techniques for a variety of modeling and relevance problems involving users, their relationships, their tweets and their interests.
- Design and evaluate novel approaches for handling high-volume real-time data streams.
- Code using primarily Java, Scala, and scripting languages such as Python or Ruby.
- Conduct design and code reviews.
- Work with large unstructured and structured data sets (multi-terabyte+, 100MM+ daily transaction volumes).
- Utilize data science and quantitative methodologies to help shape clinical care and long-term planning.
Finding Real Insights …

Data Sources
• Sensor data
• Log data
• Internet of Things
• Etc.

Big data growth

Big data market is estimated to grow 45% annually to reach $25 billion by 2015

Growth of Global data - Zettabytes
Zettabyte = one million petabytes

2010 Stored data* - Petabytes
Petabyte - one quadrillion (short scale) bytes

*greater than
Sources: Nasscom -CRISIL GR&A analysis

Do we really need to collect & mine all the data?

Reuters graphic/Catherine Trevethan 05/10/12

MITRE
Data Analytics

Descriptive Analytics
- What happened?
- How many, how often, where?
- Where exactly is the problem?
- Why is this happening?

Predictive Analytics
- What actions are needed?
- What if these trends continue?
- What will happen next?

Prescriptive Analytics
- What’s the best that can happen?
- Optimization

Optimization helps you determine the best that can happen, so you can take action in ways that will deliver significant performance improvements. Advances in technology have made this process easier and more powerful.

Unicorn Skills

- Communications
- Leadership
- Domain
- Business
- Tools
  - R
  - Python
  - Java
  - Tableau
  - SQL
  - Hadoop
  - SAS/SAS Enterprise Miner

Acquiring Unicorn Skills

- Self-study
- On the job training
- Websites
  - Data Science Central [http://www.datasciencecentral.com/](http://www.datasciencecentral.com/)
- Journals
  - IEEE Transactions on Computational Intelligence and AI in Games
  - IEEE Computational Intelligence
- Conferences
  - IEEE International Conference on Big Data 2014
  - IEEE 2014 International Conference on Data Science & Engineering (ICDSE)
  - Predictive Analytics World
- Courses and Degree programs
  [http://www.kdnuggets.com/education/online.html](http://www.kdnuggets.com/education/online.html)
Salaries

Note: Salary data varies among surveys

Source: http://cdn.edureka.co/blog/wp-content/uploads/2015/01/salaryjobtitles.png
Challenges

- **Lots of data NOT ALWAYS EQUAL TO Insights**
  - Need to carefully select attributes
  - Need to understand the context of the problem and results
  - Do the model results make sense!
  - Realize that models degrade with time!

- **Gaining**
  - Data analytics experience
  - Knowledge and skills
  - Experience that employers care about

- **Employers**
  - Finding skilled staff
  - Staff in Centralized Data Analytics Center versus Business Units
  - Retaining staff
  - Training and growing staff
Conclusions

- Electrical Engineers, Computer Science, and Math Professionals can transition to Data Science
- Already know and use various aspects of
  - Descriptive analytics
  - Prescriptive analytics
  - Predictive analytics
- Resources exist to allow information technology professionals to fill-in their knowledge gaps
- Data Science is a growth area
Next Steps

- Take courses and read to fill in the knowledge gaps
- Add a data science, data analytics, or predictive analytics degree if time permits
- Find small projects where you are to
  - Leverage your current skills and strengths
  - Apply data analytics as a value added benefit
- Talk with people interested in data science and data analytics

“Continuous effort – not strength or intelligence – is the key to unlocking our potential.”

Winston Churchill
Back-up Slides
Resources and Further Reading

- Kdnuggets – Data Mining Community
- Siegel, E. (2013). The power to predict who will click, lie, buy, or die. Wiley.
- The Data Warehousing Institute (TWDI) http://tdwi.org/Home.aspx
Resources and Further Reading (cont.)

- **Journals**
  - IEEE Transactions on Computational Intelligence and AI in Games
  - IEEE Computational Intelligence

- **Conferences**
  - IEEE International Conference on Big Data 2014
  - IEEE 2014 International Conference on Data Science & Engineering (ICDSE)
  - Predictive Analytics World
Courses and Certification Programs (Sample)

- CalTech, Learning from Data, http://work.caltech.edu/telecourse.html
- Coursera https://www.coursera.org/specialization/jhudatascience/1/overview
- New Jersey Institute of Technology Certificate in Data Mining http://online.njit.edu/programs/certs/datamining-cert.php
Degrees Programs (Sample)

- Berkeley Masters of Information and Data Science [http://datascience.berkeley.edu/](http://datascience.berkeley.edu/)
- DePaul University, Master of Science Predictive Analytics [http://www.cdm.depaul.edu/academics/Pages/MS-In-Predictive-Analytics.aspx](http://www.cdm.depaul.edu/academics/Pages/MS-In-Predictive-Analytics.aspx)
- University of Maryland, Masters of Science in Data Analytics [http://www.umuc.edu/analytics/index.cfm](http://www.umuc.edu/analytics/index.cfm)
Common Job Posting Terms
Salaries by State

Factors Influencing Salaries

Source: 2014 TDWI Salary, Roles, and Responsibilities Report

- People who moved tended to make more money.

- Having 5+ years of experience pays off.

- Having IT certifications increases salaries.
  Most certifications require at least 5 years of professional experience.
Impact of Education

Salary Trends

Source: http://cdn.edureka.co/blog/wp-content/uploads/2015/01/salary-1.png
Salaries
## Opportunities

### Average Salary by Region

<table>
<thead>
<tr>
<th>Region</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid-Atlantic</td>
<td>$116,597</td>
<td>$135,568</td>
<td>$122,258</td>
<td>6%</td>
</tr>
<tr>
<td>Southwest</td>
<td>$110,116</td>
<td>$118,487</td>
<td>$118,661</td>
<td>6%</td>
</tr>
<tr>
<td>Pacific</td>
<td>$113,552</td>
<td>$117,649</td>
<td>$118,396</td>
<td>14%</td>
</tr>
<tr>
<td>Northeast</td>
<td>$111,659</td>
<td>$110,867</td>
<td>$114,333</td>
<td>15%</td>
</tr>
<tr>
<td>Southeast</td>
<td>$101,633</td>
<td>$109,630</td>
<td>$106,108</td>
<td>11%</td>
</tr>
<tr>
<td>South</td>
<td>$104,038</td>
<td>$104,692</td>
<td>$106,063</td>
<td>4%</td>
</tr>
<tr>
<td>Midwest</td>
<td>$100,939</td>
<td>$99,006</td>
<td>$102,823</td>
<td>21%</td>
</tr>
<tr>
<td>Rocky Mountains</td>
<td>$105,294</td>
<td>$106,308</td>
<td>$100,985</td>
<td>4%</td>
</tr>
<tr>
<td>Central Plains</td>
<td>$106,786</td>
<td>$92,446</td>
<td>$91,761</td>
<td>2%</td>
</tr>
<tr>
<td>Canada</td>
<td>$93,898</td>
<td>$88,251</td>
<td>$88,599</td>
<td>16%</td>
</tr>
</tbody>
</table>

*2013 data*

### Average Salary by Industry

<table>
<thead>
<tr>
<th>Industry</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Change*</th>
<th>Respondents**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media/entertainment/publishing</td>
<td>$116,363</td>
<td>$110,357</td>
<td>$129,114</td>
<td>+17.0%</td>
<td>2%</td>
</tr>
<tr>
<td>Consulting/professional services</td>
<td>$115,237</td>
<td>$120,395</td>
<td>$115,726</td>
<td>-3.9%</td>
<td>13%</td>
</tr>
<tr>
<td>Software/Internet</td>
<td>$112,238</td>
<td>$113,953</td>
<td>$110,791</td>
<td>-2.8%</td>
<td>7%</td>
</tr>
<tr>
<td>Manufacturing (computers and non-computers)</td>
<td>$97,276</td>
<td>$109,223</td>
<td>$109,995</td>
<td>&lt;1.0%</td>
<td>6%</td>
</tr>
<tr>
<td>Food/beverage</td>
<td>$118,738</td>
<td>$103,300</td>
<td>$109,989</td>
<td>+6.5%</td>
<td>2%</td>
</tr>
<tr>
<td>Hospitality/travel</td>
<td>$113,213</td>
<td>$111,534</td>
<td>$108,367</td>
<td>-2.8%</td>
<td>2%</td>
</tr>
<tr>
<td>Government (federal)</td>
<td>$102,229</td>
<td>$119,215</td>
<td>$108,178</td>
<td>-9.3%</td>
<td>2%</td>
</tr>
<tr>
<td>Financial services</td>
<td>$113,076</td>
<td>$109,056</td>
<td>$106,942</td>
<td>-1.9%</td>
<td>14%</td>
</tr>
<tr>
<td>Retail/wholesale/distribution</td>
<td>$112,337</td>
<td>$106,592</td>
<td>$106,545</td>
<td>&lt;1.0%</td>
<td>5%</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>$98,880</td>
<td>$101,531</td>
<td>$106,006</td>
<td>+4.4%</td>
<td>3%</td>
</tr>
<tr>
<td>Healthcare</td>
<td>$99,633</td>
<td>$102,412</td>
<td>$105,034</td>
<td>+2.6%</td>
<td>13%</td>
</tr>
<tr>
<td>Insurance</td>
<td>$99,333</td>
<td>$100,345</td>
<td>$104,544</td>
<td>+4.2%</td>
<td>10%</td>
</tr>
<tr>
<td>Education</td>
<td>$85,113</td>
<td>$92,768</td>
<td>$95,170</td>
<td>+2.6%</td>
<td>5%</td>
</tr>
<tr>
<td>Transportation/logistics</td>
<td>$99,180</td>
<td>$102,531</td>
<td>$92,672</td>
<td>-9.6%</td>
<td>3%</td>
</tr>
<tr>
<td>Government (state/local)</td>
<td>$89,201</td>
<td>$87,173</td>
<td>$83,695</td>
<td>-4.0%</td>
<td>3%</td>
</tr>
</tbody>
</table>

*Y-Y 2012–13

**2013 data. Column does not total 100% because industries with lower representation were excluded.
Salaries by Industry