How to Find the Data Scientists You Need

1. Focus recruiting at the “usual suspect” universities (Stanford, MIT, Berkeley, Harvard, Carnegie Mellon) and also at a few others with proven strengths: North Carolina State, UC Santa Cruz, the University of Maryland, the University of Washington, and UT Austin.

2. Scan the membership rolls of user groups devoted to data science tools. The R User Groups (for an open-source statistical tool favored by data scientists) and Python Interest Groups (for PIGgies) are good places to start.

3. Search for data scientists on LinkedIn—they’re almost all on there, and you can see if they have the skills you want.

4. Hang out with data scientists at the Strata, Structure:Data, and Hadoop World conferences and similar gatherings (there is almost one a week now) or at informal data scientist “meet-ups” in the Bay Area; Boston; New York; Washington, DC; London; Singapore; and Sydney.

5. Make friends with a local venture capitalist, who is likely to have gotten a variety of big data proposals over the past year.

6. Host a competition on Kaggle or TopCoder, the analytics and coding competition sites. Follow up with the most-creative entrants.

7. Don’t bother with any candidate who can’t code. Coding skills don’t have to be at a world-class level but should be good enough to get by. Look for evidence, too, that candidates learn rapidly about new technologies and methods.

8. Make sure a candidate can find a story in a data set and provide a coherent narrative about a key data insight. Test whether he or she can communicate with numbers, visually and verbally.

9. Be wary of candidates who are too detached from the business world. When you ask how their work might apply to your management challenges, are they stuck for answers?

10. Ask candidates about their favorite analysis or insight and how they are keeping their skills sharp. Have they gotten a certificate in the advanced track of Stanford’s online Machine Learning course, contributed to open-source projects, or built an online repository of code to share (for example, on GitHub)?

2012年，Tim O’Reilly 列出了世界上排名前7位的数据科学家。

* Larry Page，谷歌CEO。
* Jeff Hammerbacher，Cloudera的首席科学家和DJ Patil，Greylock风险投资公司企业家。
* Sebastian Thrun，斯坦福大学教授和Peter Norvig，谷歌数据科学家。
* Elizabeth Warren，Massachusetts州美国参议院候选人。
* Todd Park，人类健康服务部门首席技术官。
* Sandy Pentland，麻省理工学院教授。
* Hod Lipson and Michael Schmidt，康奈尔大学计算机科学家。

[D.J. Patil](https://hbr.org/search?term=d.j.+patil) is the data scientist in residence at Greylock Partners, was formerly the head of data products at LinkedIn, and is the author of *Data Jujitsu: The Art of Turning Data into Product* (O’Reilly Media, 2012).

Kirk D. Borne is a leading Data Scientist, Professor of Astrophysics and Computational Science at George Mason University, Big Data Consultant and Advisor.

* Anjul Bhambhri，IBM的大数据产品副总裁。
* Jonathan Goldman，LinkedIn数据科学家。

2006年的6月份进入商务社交网站LinkedIn，当时LinkedIn只有不到800万用户。高德曼在之后的研究中创造出新的模型，利用数据预测注册用户的人际网络。具体来讲，他以用户在LinkedIn的个人资料，来找到和这些信息最匹配的三个人，并以推荐的形式显示在用户的使用页面上——这也就是我们熟悉的”你可能认识的人（People you may know）”。这个小小的功能让LinkedIn增加了数百万的新的页面点击量(数据挖掘的应用典型之一推荐系统）。

* John Rauser, 亚马逊大数据科学家。
* Steven Hillion, EMC Greenplum数据分析副总裁。
* Monica Rogati, LinkedIn资深数据科学家。
* Daniel Tunkelang，LinkedIn首席数据科学家。
* Michael Rappa，北卡罗莱纳州立大学教授。
* 林仕鼎，百度大数据首席架构师。