

INTRODUCTION TO DATA SCIENCE

CHRIS QI

PH.D. UNIVERSITY OF MARYLAND



OVERVIEW







ENGINE = COMPUTER

FUEL = DATA





The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades, ... because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Hal Varian, Google's Chief Economist
The McKinsey Quarterly, Jan 2009

The Economist

Gordon Brown's pitch
What went wrong at RBS
Genetically modified crops blossom
The EU woos Russia
The right to eat cats and dogs.

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT



The New York Times

For Today's Graduate, Just One Word: Statistics

By STEVE LOMER
Published: August 5, 2009

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls "all the computer and math stuff" that was part of the job.

[TWITTER](#)
[LINKEDIN](#)
[COMMENTS \(8\)](#)
[SIGN IN TO EMAIL](#)

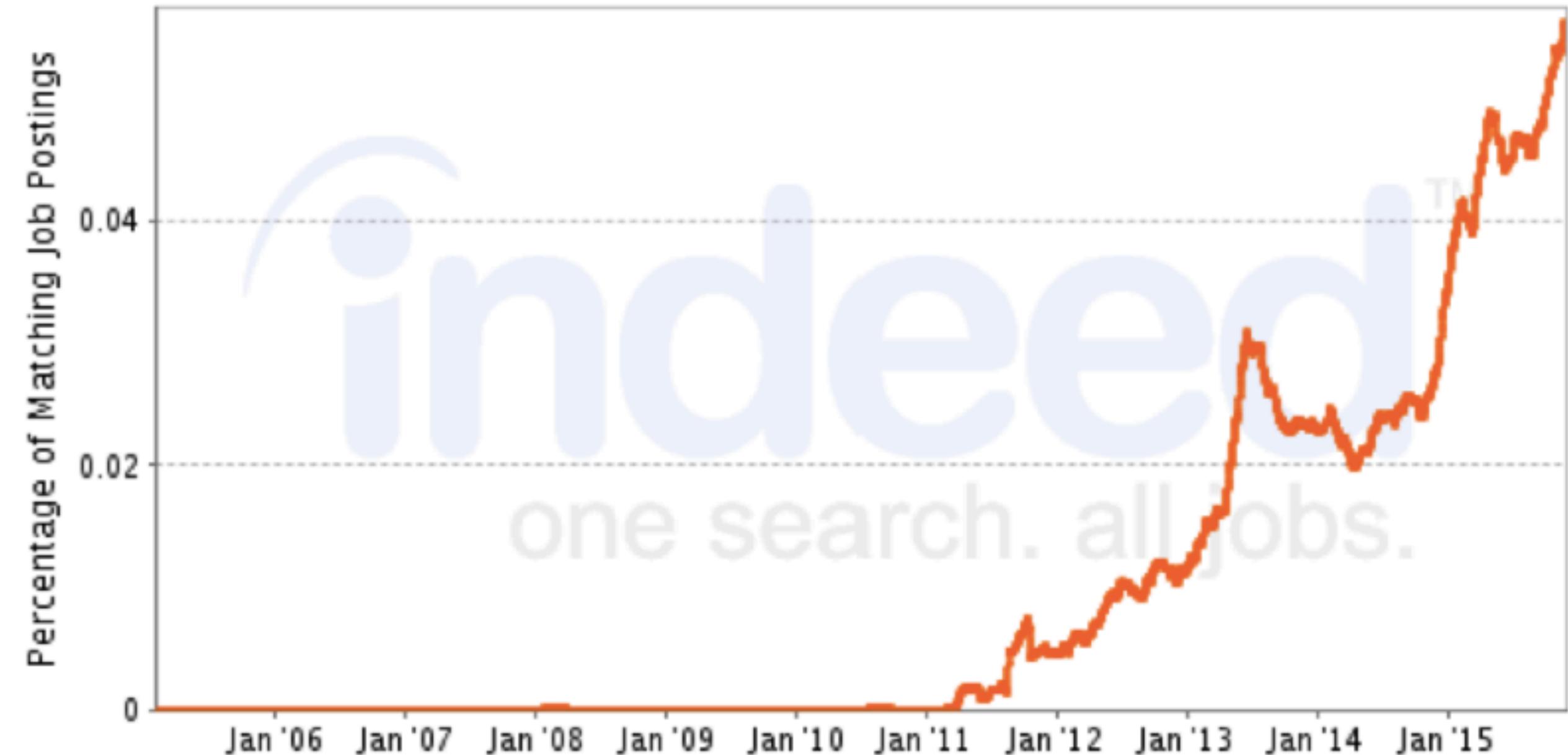
Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you

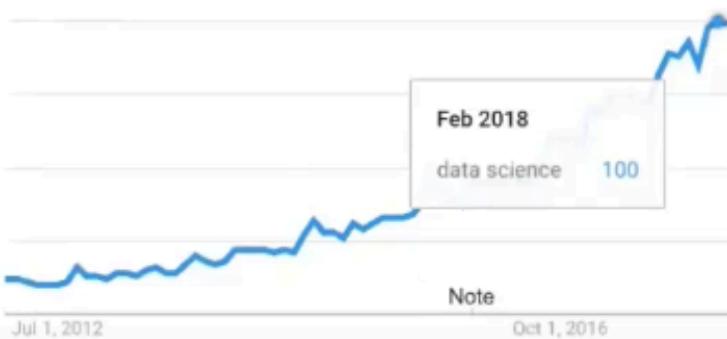
Job Trends from Indeed.com

— "Data Scientist"



职位热度

热度随时间变化的趋势



未来5年, Data Scientist一直会是就业市场中**紧缺**的人才

By 2018, the US alone could face a **shortage** of **1.5 million** new data scientists.

[McKinsey & Company](#)

薪资水平

Data Scientist整体薪资水平逐年递增
并且远高于National平均水平

Software Engineer

Average Base Pay

\$109,006 /yr

Data Scientist Salaries in San Francisco, CA Area

About This Data ?

607 Salaries Updated Mar 3, 2018

All Industries

All Company Sizes

0-1 Years

Average Base Pay

\$119,981 /yr

19% above national average

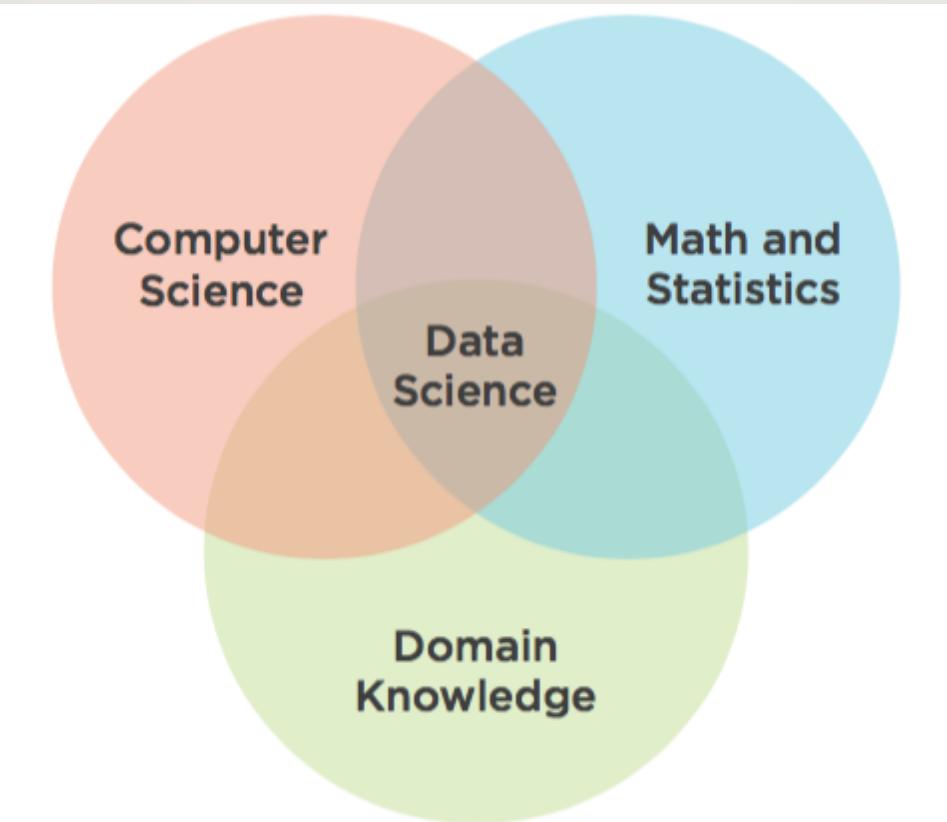


Salaries for Related Job Titles

Data Analyst \$86K

Data Scientist Intern \$107K

WHAT IS DATA SCIENCE



APPLICATIONS

数据科学的业界应用

Application	Companies
House Price Prediction	Zillow
Movie Recommendation	Netflix
Object Detection	Tesla
Fraud Detection	Chase



PREREQUISITES

- Basic programming experience
- Basic statistics experience

DATA SCIENCE SKILLS

- Programming
- Working with data
- Descriptive statistics
- Data visualization
- Statistical modeling
- Machine learning

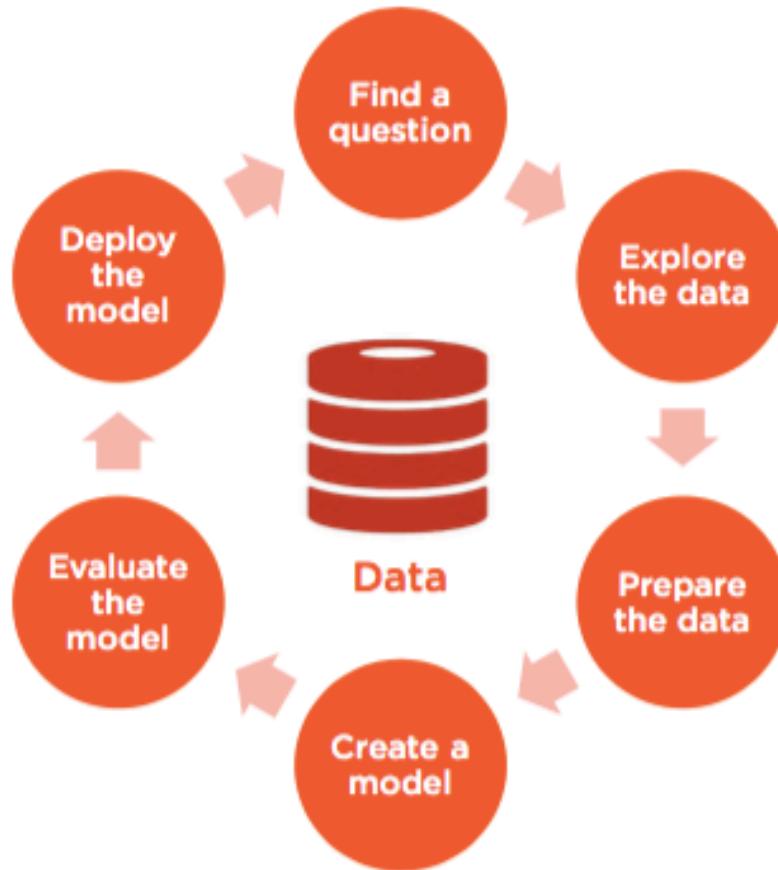
WHY IS DATA SCIENCE IMPORTANT?

- Internet of Things
- Big Data
- Machine learning
- Fully autonomous systems

WHY IS DATA SCIENCE IMPORTANT?

- Driven by economy
- Possible by technology
- Cost is decreasing
- Value is increasing

The Data Science Process



PREVIEW

- Basic Statistics
- R basic commands
- Exploratory analysis
- Data visualization
- Linear Modeling
- Prediction
- Machine Learning

COURSE OBJECTIVE

Provide Solid Foundation	Real World Examples	What & Why ?
Follow Along	Not To Cover Every Aspect	Take The First Step