

Data Science: Getting Value out of Big Data

After this video you will be able to..

- Describe what modern data science is and its connections to big data
- Explain why data science is the key to getting value out of big data
- List the right set of skills for a data scientist to fit your organization



Insight → **Data Product**

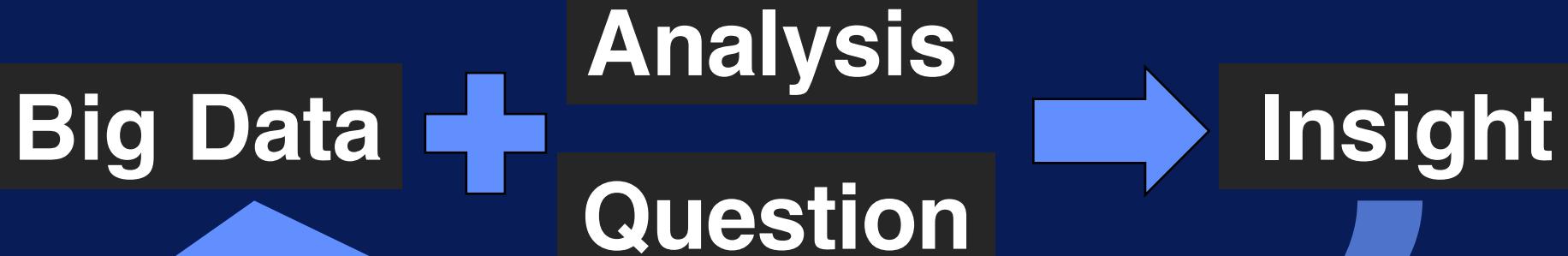
Big Data +

Analysis

Question

→ **Insight**

Insight → **Data Product**



Book Recommendations

Customer
Demographic

Previous
Purchases

Book reviews



What kind of
books does this
customer like?



Book
recommendations

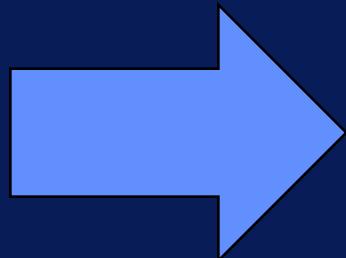


Find Potential Audience for a Book

Model of
customer's book
preferences



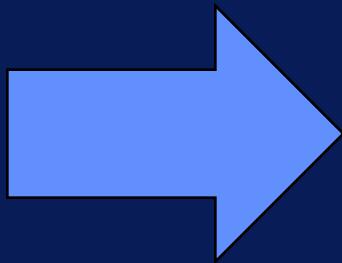
New book
information



Who is likely to
like this book?

Market a New Book

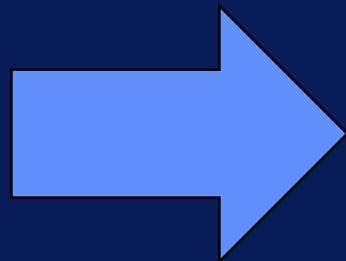
Who is likely to
like this book?



Action to market
the book to the
right audience

Market a New Book

Who is likely to like this book?



Action to market the book to the right audience

Insight



Action

Actionable Information

Historical data

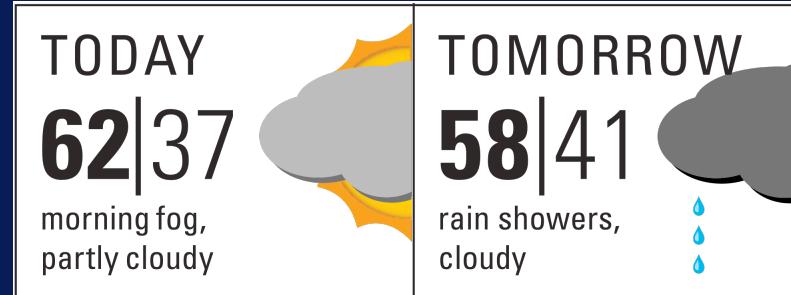


Near real-time data

Prediction

A large, light blue downward-pointing arrow, positioned centrally below the input boxes, pointing towards the output box.

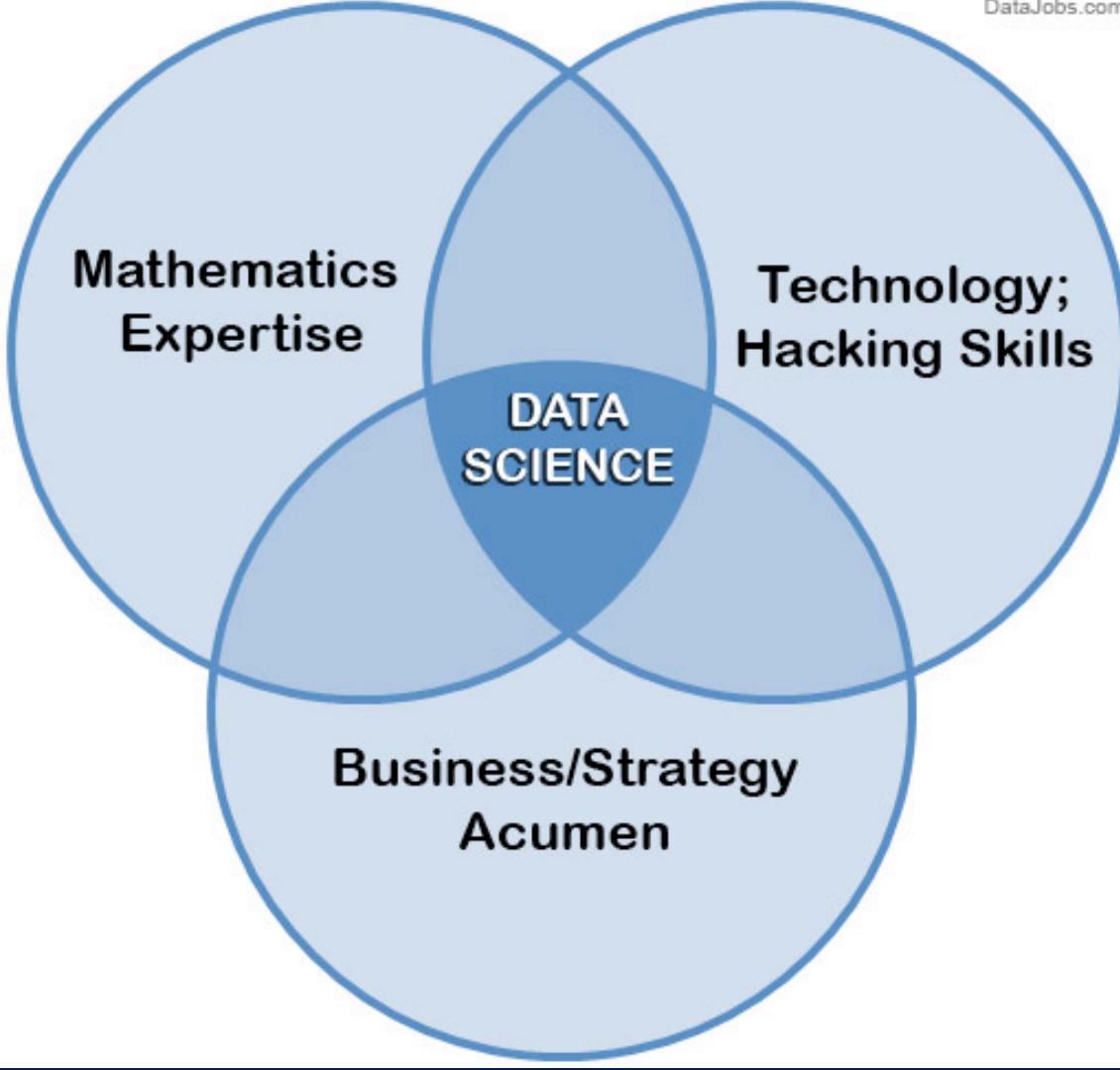
Prediction

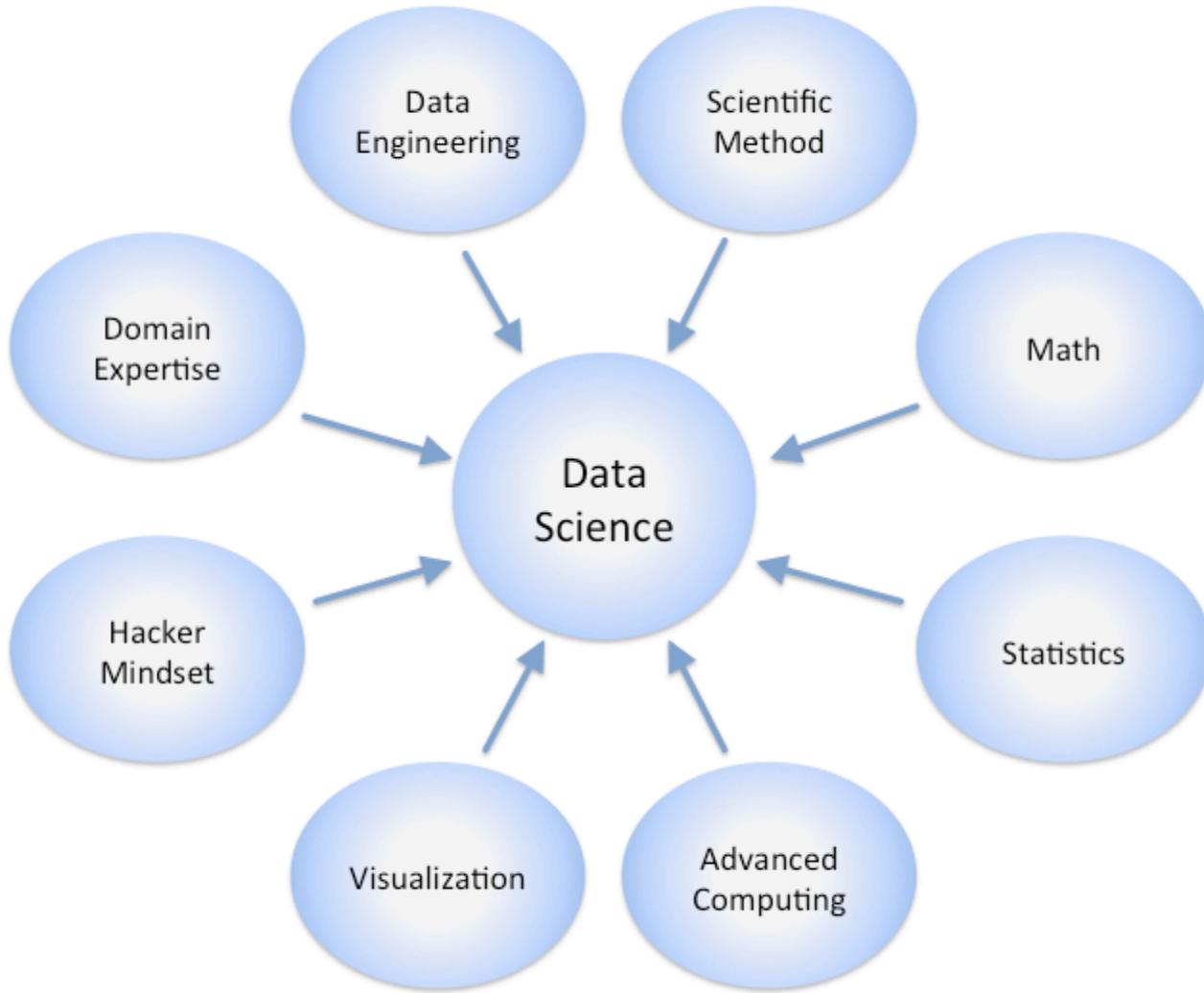


Action



Data Science is Team Work!





MODERN DATA S

Data Scientist, the sexiest job of 21th century requires a mixture of intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist is.

MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants



MATH & STATISTICS

- ★ Machine learning
- ★ Statistical modeling
- ★ Experiment design
- ★ Bayesian inference
- ★ Supervised learning: decision trees, random forests, logistic regression
- ★ Unsupervised learning: clustering, dimensionality reduction
- ★ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ★ Passionate about the business
- ★ Curious about data
- ★ Influence without authority
- ★ Hacker mindset
- ★ Problem solver
- ★ Strategic, proactive, creative, innovative and collaborative

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

PROGRAMMING & DATABASE

- ★ Computer science fundamentals
- ★ Scripting language e.g. Python
- ★ Statistical computing packages, e.g., R
- ★ Databases: SQL and NoSQL
- ★ Relational algebra
- ★ Parallel databases and parallel query processing
- ★ MapReduce concepts
- ★ Hadoop and Hive/Pig
- ★ Custom reducers
- ★ Experience with xaaS like AWS



COMMUNICATION & VISUALIZATION

- ★ Able to engage with senior management
- ★ Story telling skills
- ★ Translate data-driven insights into decisions and actions
- ★ Visual art design
- ★ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



Are data scientists unicorns?

"ALL FOR ONE-
ONE FOR ALL"



DOUGLAS
FAIRBANKS
presents
"THE THREE
MUSKETEERS"

*Adaption, Costuming
Research under*
EDWARD KNOBLOCK

Direction under
FRED NIBLO

Photography under
ARTHUR EDESON



Have passion for data

"ALL FOR ONE-
ONE FOR ALL"

DOUGLAS
FAIRBANKS

presents

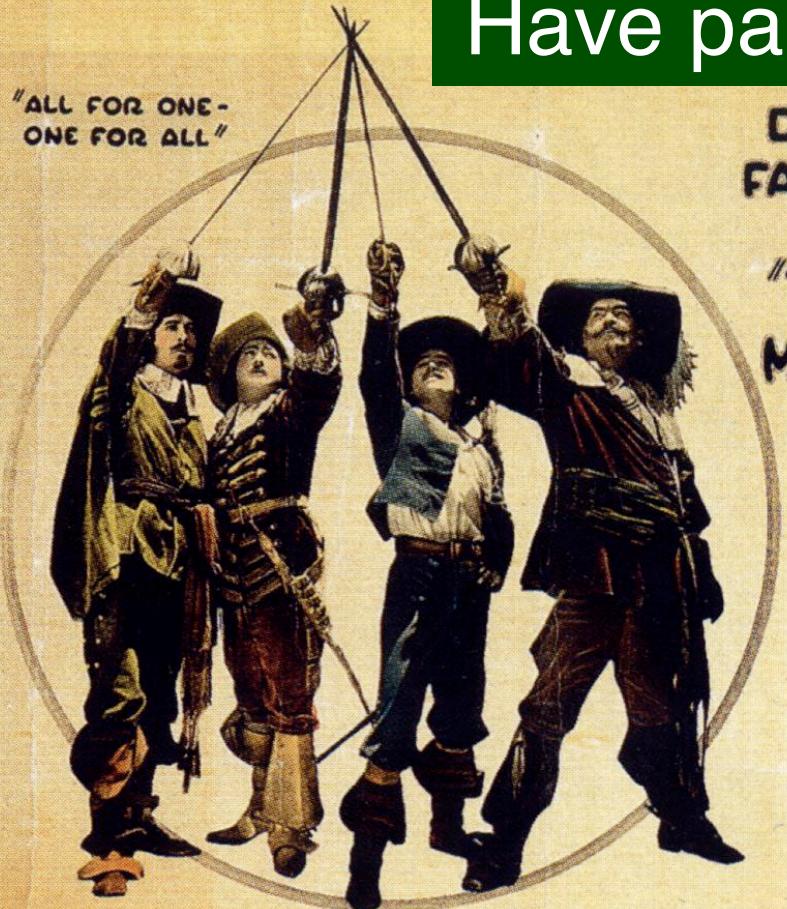
"THE THREE MUSKETEERS"

Adaption, Costuming
Research under

EDWARD KNOBLOCK

Direction under
FRED NIBLO

Photography under
ARTHUR EDESON



Have passion for data

Relate problems to analytics





Have passion for data

Relate problems to analytics

Care about engineering solutions



Have passion for data

Relate problems to analytics

Care about engineering solutions

Exhibit curiosity

Research under
EDWARD KNOBLOCK

Direction under
FRED NIBLO

Photography under
ARTHUR EDESON





Have passion for data

Relate problems to analytics

Care about engineering solutions

Exhibit curiosity

Communicate with teammates

Photography under
ARTHUR EDESON





Data Science

The sum is bigger than
the parts!

Big Data



Actionable Insight

Data
Scientist

Technical Skills

Business Skills

Soft Skills