

Business Analytics: Strategic Issues and Solutions

Professor **Venky Shankar**

Coleman Chair in Marketing

Director, Center for Retailing Studies

Mays Business School

Texas A&M University

<http://www.venkyshankar.com>

venky@venkyshankar.com

I am grateful to Tarun Kushwaha for sharing his slides from which many of the slides are adapted .

Venky Shankar

My Research in Data Science

- Mixture of Hidden Markov Models
- Topic Modeling
- NLP
- Causal inference and machine learning
- Deep learning
- AI and IoT

Venky@venkyshankar.com



Agenda

- Introduction to data science & business analytics
- Business analytics framework
- Asking the right questions
- Value of data & insights

Agenda (Cont'd)

- Descriptive, Predictive analytics
- Machine learning models
- Causality trap
- Analytics pitfalls & questions

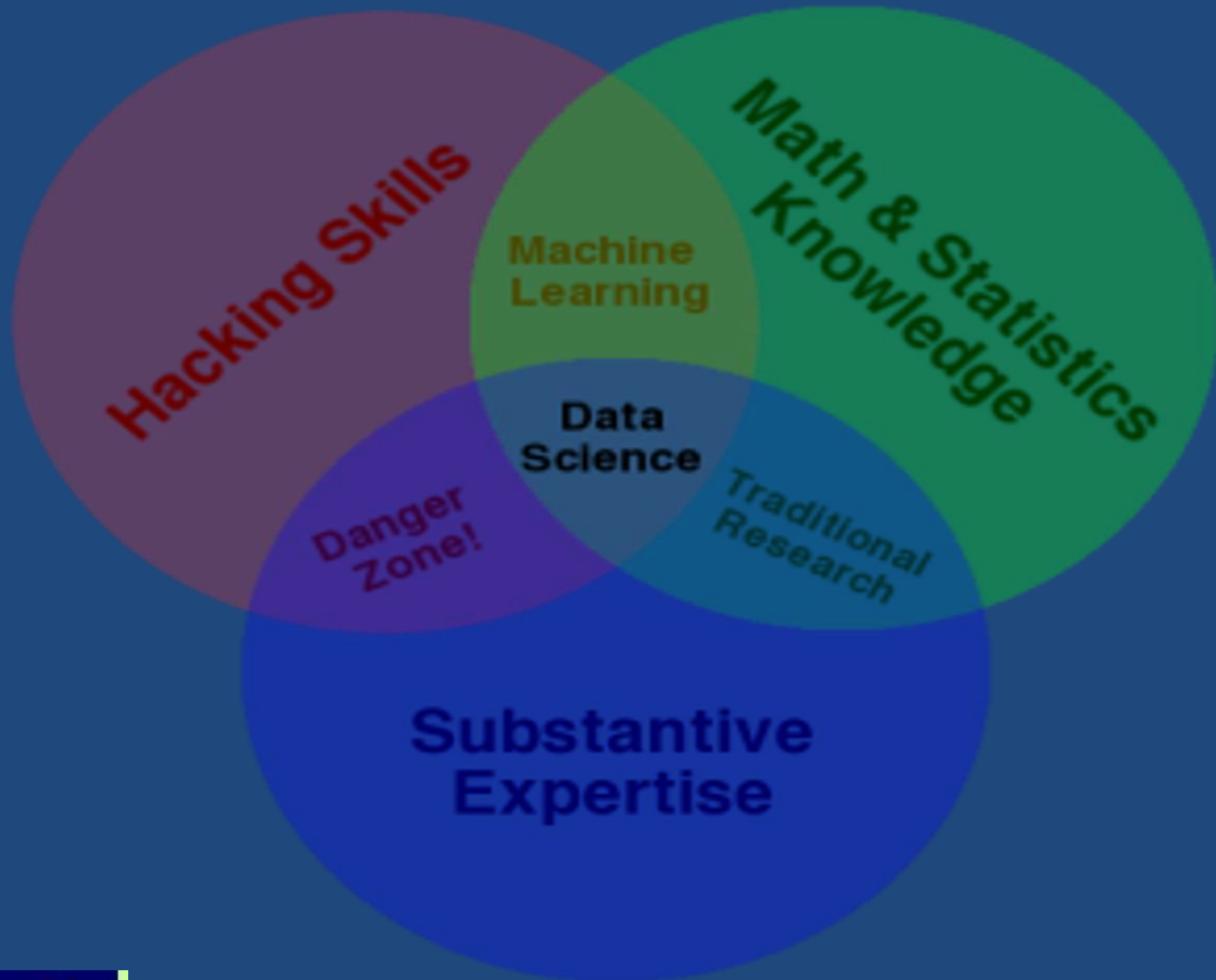
Introduction to Data Science and Analytics

[Analytics in Movies](#)

What is Data Science?

- Interdisciplinary field comprising scientific methods, processes, and systems to collect, organize, analyze and extract knowledge from all types of data, structured and unstructured regardless of the domain
- Integrates mathematics, statistics, computer science, and information science

Data Science



The Data Scientist

Based on 500+ conversations with CFOs from around the world

82%

of CFOs see the value of integrating enterprise-wide data



But only

24%

think their team is up to the task

(Source:IBM)

Domain Expert

Statistics Expert

**Visualization &
Communication Expert**

**Programming
Expert**

**Database Technology
Expert**

[5 Dimensions of the So-Called Data Scientist](#)

Venky Shankar

What are Analytics & Business Analytics

- **Analytics:**
 - “The scientific process of transforming data into insight for making better decisions.”
 - Institute for Operations Research and Management Science (INFORMS)
 - “The discovery and communication of meaningful patterns in data.”
 - Wikipedia entry “Analytics”
- **Business Analytics:**
 - “The broad use of data and quantitative analysis for decision-making within organizations.”
 - Tom Davenport, author of *Competing on Analytics*

Related Terms

- **Business Intelligence:** Umbrella term encompassing knowledge, querying, reporting, analytics
- **Predictive Analytics:** a set of analytic techniques/tools for predicting the future using data from the past
- **Statistics:**
 - E.g., linear & logistic regression, hypothesis testing
 - Classically assumes limited data and computation
 - Asks “what can we learn from the data?”
- **Machine Learning:**
 - Set of methods comprising models & algorithms to perform a task by learning from data
- **Data Mining:**
 - E.g., neural networks, classification trees
 - “Statistics at scale and speed.” (Pregibon, 1999)
 - Asks “what can we do with the data?”

Big Data

- Large and complex data with challenges to collect, curate, store, search, transfer, analyze, and visualize
- Structured, unstructured; text, audio, video
- Requires massively parallel software running on thousands of servers
- Characterized by 5Vs: Volume, Velocity, Variety, Veracity, Value
- Propelled by the rise of SMACIT (Social, Mobile, Analytics, Cloud, Internet of Things)
- Analysis requires models to accommodate 5Vs

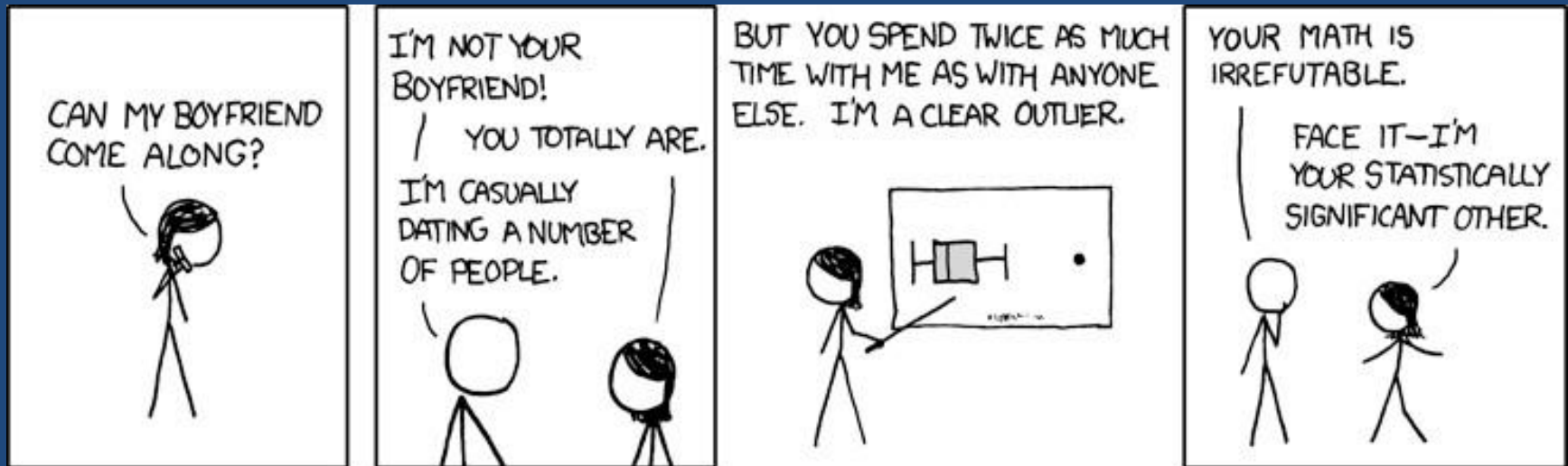
The Big Deal about Big Data

- Business data double every 1.2 years
- 500,000 data centers
- By 2020, 1/3 data through cloud, 35ZB of total data
- Market for big data applications
 - \$139B in 2020 (Marketsandmarkets report)
- Companies investing in Big Data will outperform others by 20% (Gartner)
- Volume, Velocity (Walmart: 1M transactions/hr-2.5K terabytes)
- Emergence of data science/data scientist: 2.72M by 2020 (IBM)
- 40% of data science tasks will be automated by 2020 (Gartner)

Big Data Joke



Analytics Social Application



Data Analytics Framework

Analytics Types



Hindsight

Historical
Performance

Descriptive
Analytics

Mean, Median,
Variance, % Change,
Basic Visualization

Insight

Drivers of
Performance

Predictive
Analytics

Correlation, Association,
Bivariate Graphs, Regression,
Machine Learning, Trees

Foresight

Altering
Performance

Prescriptive
Analytics

Regression, Experiments,
Forecasting, Simulations,
Machine Learning

Asking the Right Questions

What is a Crunchy Question?

- “Crunchy questions are practical, detailed inquiries into tough business issues—roll-up-your-sleeves questions for people who don’t have time to mess around with fluff. Crunchy questions are designed to lay the groundwork for action.” – ([Deloitte](#))
- “They are capable of forming the connective tissue between tactical and senior [strategic] level objectives in an organization.” – John Lucker ([Deloitte](#))
- “Crunchy questions are aligned with strategic goals, relate to key performance indicators (KPIs), are designed to be actionable as well as informational, and provide foresight rather than hindsight.” – Nicole Laskowski ([TechTarget](#))

Examples of Crunchy Questions

- Which prospects should we target for acquisition? Why? What marketing message or offer will enhance the likelihood of acquisition?
- What factors drive customer retention? Why? What can we do to enhance retention rate?
- Which customers will we likely lose next? Why? What can we do to prevent defection?
- What mobile app features are associated with higher customer lifetime value (CLV)? Why? What innovations will enhance CLV?
- What factors influence medication compliance by patients? Why? What marketing communication can we send to patients to improve compliance?

Value of Data & Insights

Data Visualization

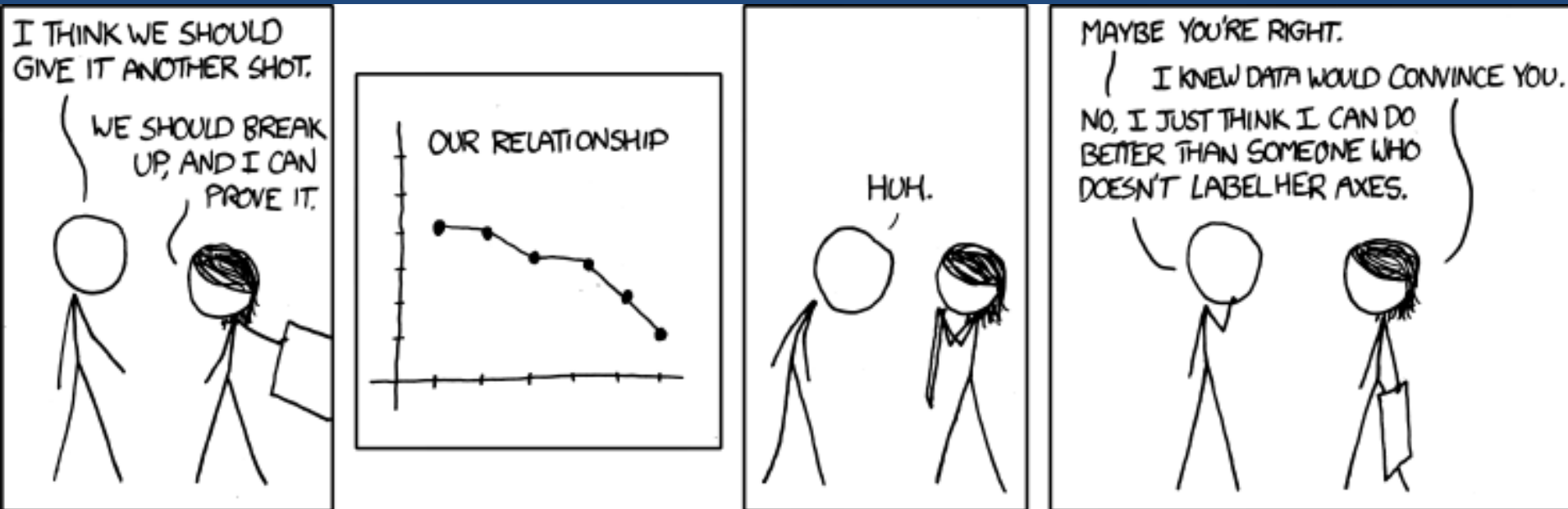
<https://www.youtube.com/watch?v=jbkSRLYSoj0>

Descriptive and Predictive Analytics

Venky Shankar

Copyright © Venkatesh Shankar. All rights reserved.

Predictive Analytics



Machine Learning

- Science of getting computers to learn without being explicitly programmed
- A scientific discipline that explores the construction and study of models and algorithms to perform a task by learning from data
- Arthur Samuels of IBM coined it in 1959; a branch of computer science; now combined with statistics
- Supervised, Unsupervised (Data mining), Reinforcement, Deep Learning

Supervised Learning

- Class of methods which learn a general rule that maps inputs to outputs based on pre-fed inputs and desired outputs
- Classification (logit, discriminant analysis, SVM, decision trees)
- Regression
- Collaborative filtering (Nearest neighbors)

Unsupervised Learning

- Class of methods which find structure in input and produces outputs which can the goal (discovering hidden patterns in data) or a means towards an end (feature learning) (data mining/exploratory statistics)
- Clustering (K-means, Hierarchical, mixture models)
- Anomaly detection
- Neural network (Hebbian learning, GAN)
- Blind signal separation (principal component analysis, factor analysis)
- LDA, CTM

Logistic Regression

- Utility theory
- Logistic regression
- Model fit
- Interpretation of coefficients
- Interpretation of elasticities
- Amazon's book retail business

Classification: Logit vs SVM

Logit

Statistical properties

Works well for structured data

More vulnerable to overfitting

No. of features is much smaller than training samples

More interpretable

Less generalizable, scalable

SVM

Geometric properties

Works well for unstructured and semi-structured data

Less susceptible to overfitting

No. of features is relatively smaller than training examples

Less interpretable

More generalizable, scalable

Predictive Analytics



Analytics Application

\$300 Billion Savings in Healthcare

Source: McKinsey & Co

Thank You!

Venky@venkyshankar.com

Venky Shankar