
Oleksandr Romanko, Ph.D.

Senior Research Analyst, Risk Analytics, Watson Financial Services, IBM Canada
Adjunct Professor, University of Toronto

MIE1624H – Introduction to Data Science and Analytics Lecture 1 – Introduction

About me

Dr. Oleksandr Romanko

- Senior Research Analyst, Quantitative Research at Risk Analytics, Business Analytics, IBM, with the company since 2010
- Ph.D. in Computer Science from McMaster University
- Author of over 20 papers and reports
- Adjunct professor at University of Toronto and lecturer at McMaster University
- Research areas:
 - business analytics, operational research, optimization, finance
 - portfolio optimization, multi-objective optimization
 - market and credit risk modeling and optimization
 - numerical methods for risk management
 - design of numerical algorithms and their software implementation

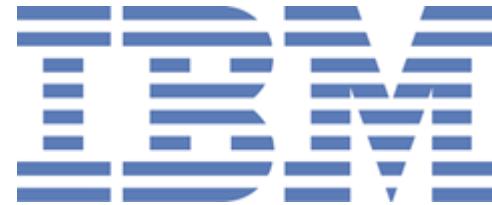




Being an IBMer



IBM Centennial: A Century of Progress



2011



1911

Incorporated on June 16, 1911 in US as the
Computing Tabulating Recording Company

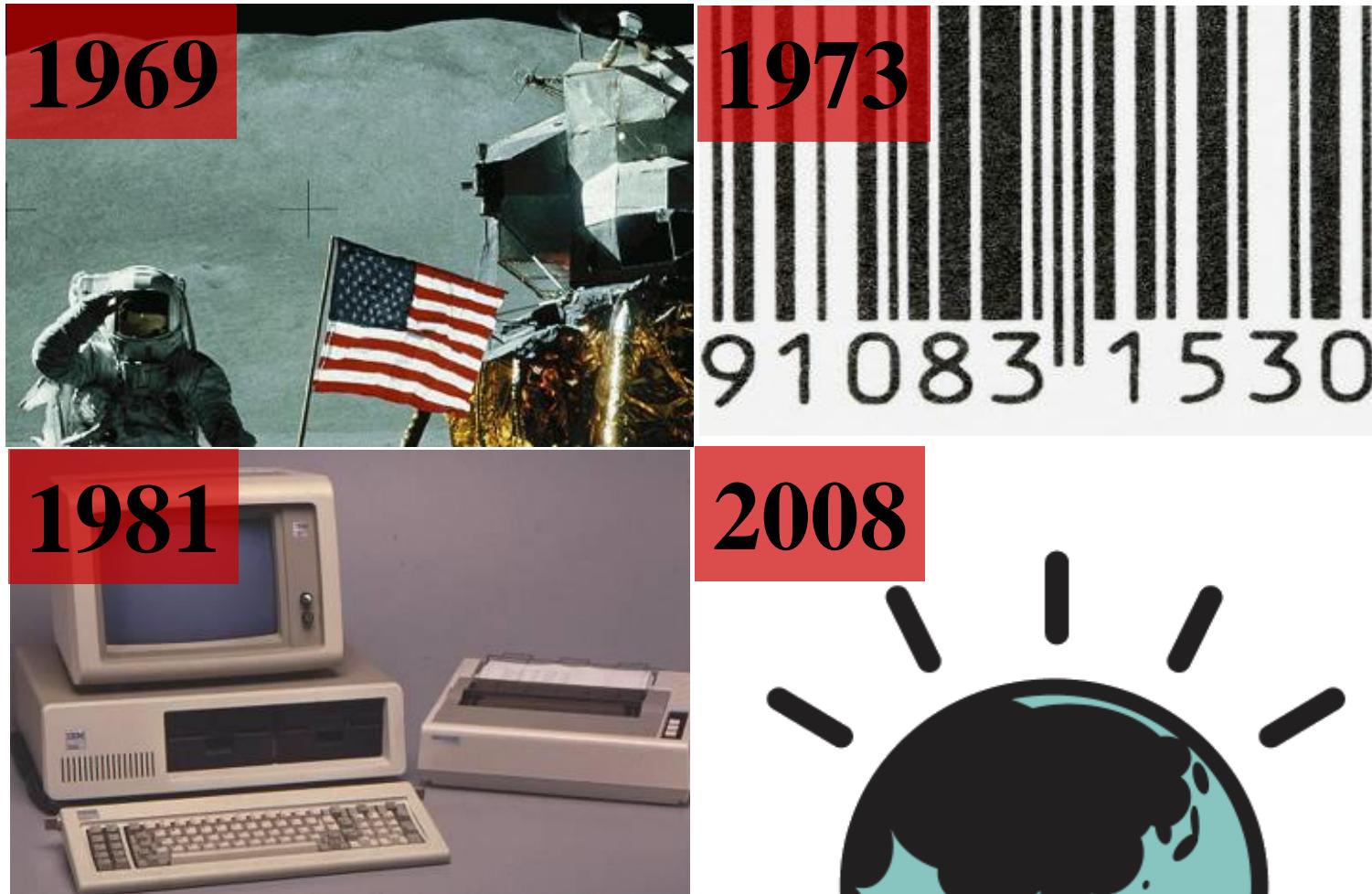
CTR changed its name to International
Business Machines Corporation
globally in 1924

CTR changes name in Canada to International
Business Machines Company in 1917



COMPUTING-TABULATING-RECORDING CO.

Making the world work better – pioneering the science



IBM Centennial: 100 Years of Innovation



A Snapshot of IBM Milestones

This June, IBM will reach a significant milestone—its 100th anniversary. Join us in taking a look back at just a few of the historical breakthroughs that have helped IBM shape the century and the company.



1900



IBM was formed as the Computing Tabulating and Recording Company, or C-T-R, specializing in punch cards, commercial scales and clocks.



With global ambitions, the company renamed itself International Business Machines – today, it operates in 170 countries.



IBM worked with the U.S. government to start Social Security – the largest accounting project of its time.

1923



IBM's punched card technology helped tackle large-scale projects like the U.S. Census.

1935



IBM pioneered training courses for women so they could work in technical positions traditionally filled by men.

1956



RAMAC (Random Access Method of Accounting and Control), the first magnetic hard disk drive, created the data storage industry.

1961



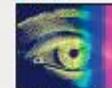
The Selectric Typewriter was an instant design sensation, delighting typists for 26 years.

1969



IBM technology guided the Apollo mission to the moon – the company has played a part in the U.S. space program since the 1960s.

1980



IBM was granted the first patent for LASIK surgery technology – and continues to hold more U.S. patents than any company.

1997



IBM's Deep Blue supercomputer defeated the best chess player in the world.

2011



Watson supercomputer can detect nuances in words, irony and wit – and inspire new realms of search queries and artificial intelligence.

1900

1930

1960

1980

2011

1924



With global ambitions, the company renamed itself International Business Machines – today, it operates in 170 countries.

1944



IBM's 604 Electronic Sequence Controlled Calculator was the first machine to handle long calculations automatically.

1964



The company made a big bet with the System/360 – this early mainframe ushered in the era of computer compatibility.

1973



Supermarkets started scanning UPC bar codes, invented by IBM. Today, they track everything from clothing to dairy cows.

1986



IBM scientists won the Nobel Prize for the scanning tunnel microscope – which would eventually manipulate atoms to spell I-B-M.

1997



IBM launched "eBusiness," turning the Internet into a tool for business and ushering in the future of electronic commerce.

1962



IBM and American Airlines launched the world's first computer-driven airline reservation system, SABRE – which paved the way for online banking technology.

1969



IBM labs developed the magnetic strips on credit cards – still ubiquitous on ID cards, drivers' licenses and ATM bank cards.

1981



The IBM Personal Computer launched the PC revolution, helping computers go mainstream beyond hobbyists and geeks.

2008



Smarter Planet launched to improve how the world works – now smart buoys in Ireland's Galway Bay detect pollution, protect fish stock.



Best Jobs

Profession



«Choose a **job you love**,
and you will never have to
work a day in your life.»

Confucius

«The only way to do great work is to
love what you do. If you haven't found
it yet, **keep looking**. Don't settle.»

Steve Jobs



Best jobs

Forbes / Tech / #BigData

FEB 25, 2016 @ 11:26 PM 14,811 VIEWS

Is Being A Data Scientist Really The Best Job In America?



Bernard Marr, CONTRIBUTOR

I write about big data, analytics and enterprise performance [FULL BIO](#)

Opinions expressed by Forbes Contributors are their own.

It's official – data scientist is the best job in America, according to users of online employment analysts Glassdoor.

Glassdoor's service allows employees to anonymously rate their jobs and their employers, awarding scores for how well they are paid, treated, and helped to advance in their careers.

 glassdoor

Jobs Companies Salaries
Interviews

Search Jobs or Companies...

Employers | Try Free Job Postings

2.5K Shares     

25 Best Jobs in America

Want a new job? Glassdoor is here to help, identifying the 25 Best Jobs in America for 2016. The jobs that make this list have the highest overall Glassdoor Job Score, determined by combining three key factors – number of job openings, salary and career opportunities rating. These jobs stand out across all three categories.

United States  2016 

1		Data Scientist	1,736
		Job Openings	\$116,840
		Median Base Salary	4.1
		Career Opportunity	4.7
		Job Score	
2		Tax Manager	1,574
		Job Openings	\$108,000
		Median Base Salary	3.9
		Career Opportunity	4.7
		Job Score	

Big Data's Big Problem: Little Talent



Forbes
TECH | 1/18/2013 @ 10:18AM | 9,232 views
Combating the Big Data skills shortage

THE CHRONICLE OF HIGHER EDUCATION

November 14, 2013

[Home](#) [News](#) [Global](#) [Opinion & Ideas](#) [Facts & Figures](#) [Blogs](#) [Advice](#) [Forums](#) [Jobs](#) [Search The Ch](#)

Administration

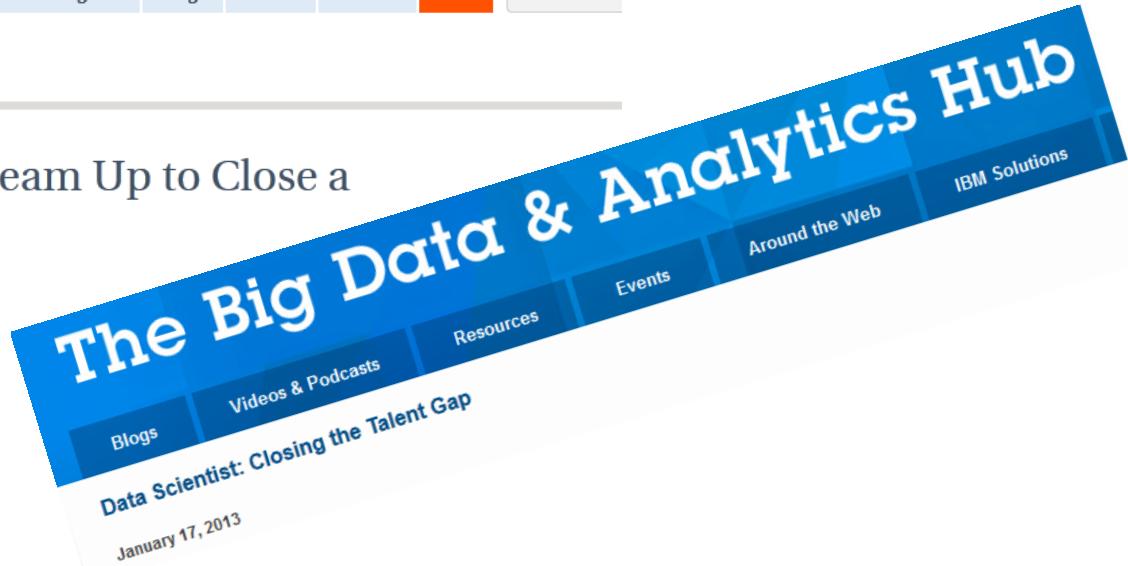
August 14, 2013

IBM and Universities Team Up to Close a 'Big Data' Skills Gap



THE MAGAZINE
October 2012

Data Scientist: The Sexiest Job of the 21st Century



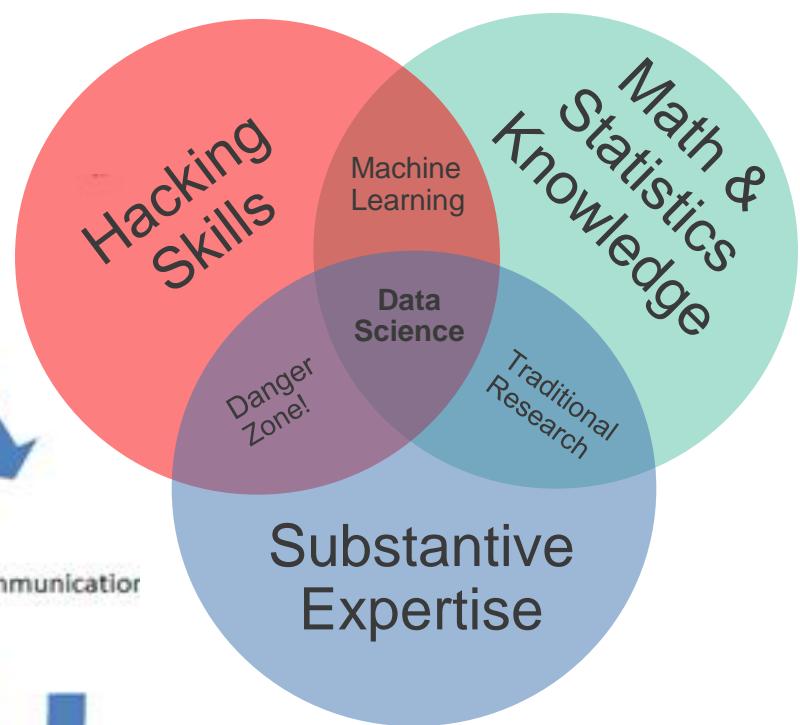
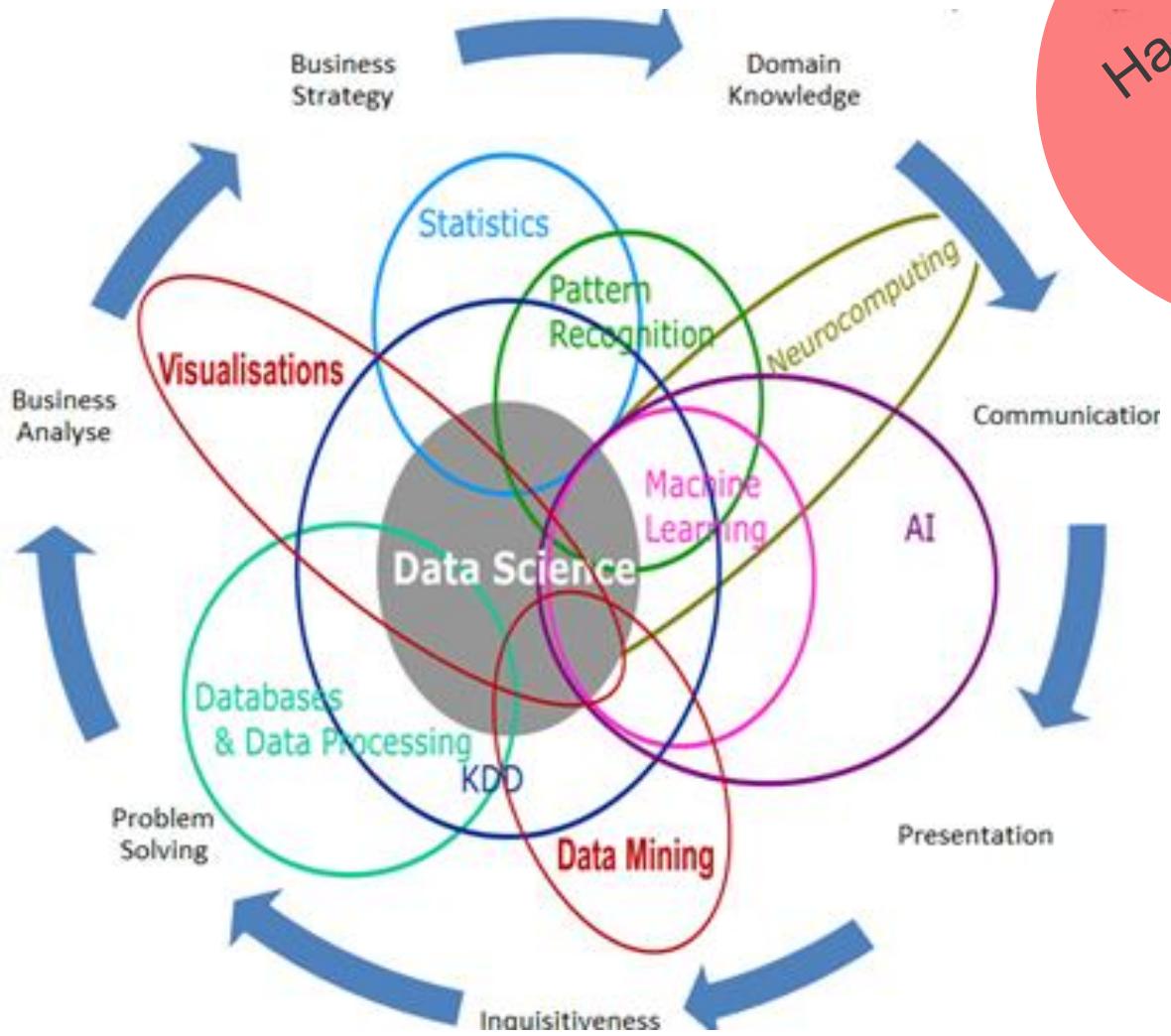
The Big Data & Analytics Hub

Blogs Videos & Podcasts Resources Events Around the Web IBM Solutions

Data Scientist: Closing the Talent Gap

January 17, 2013

Data Science





Analytics

What is analytics?

Analytics is the scientific process of deriving **insights** from **data** in order to make **decisions**



Descriptive Analytics
What has happened?

Predictive Analytics
What will happen?

Prescriptive Analytics and Artificial Intelligence
What should we do?

Business Value

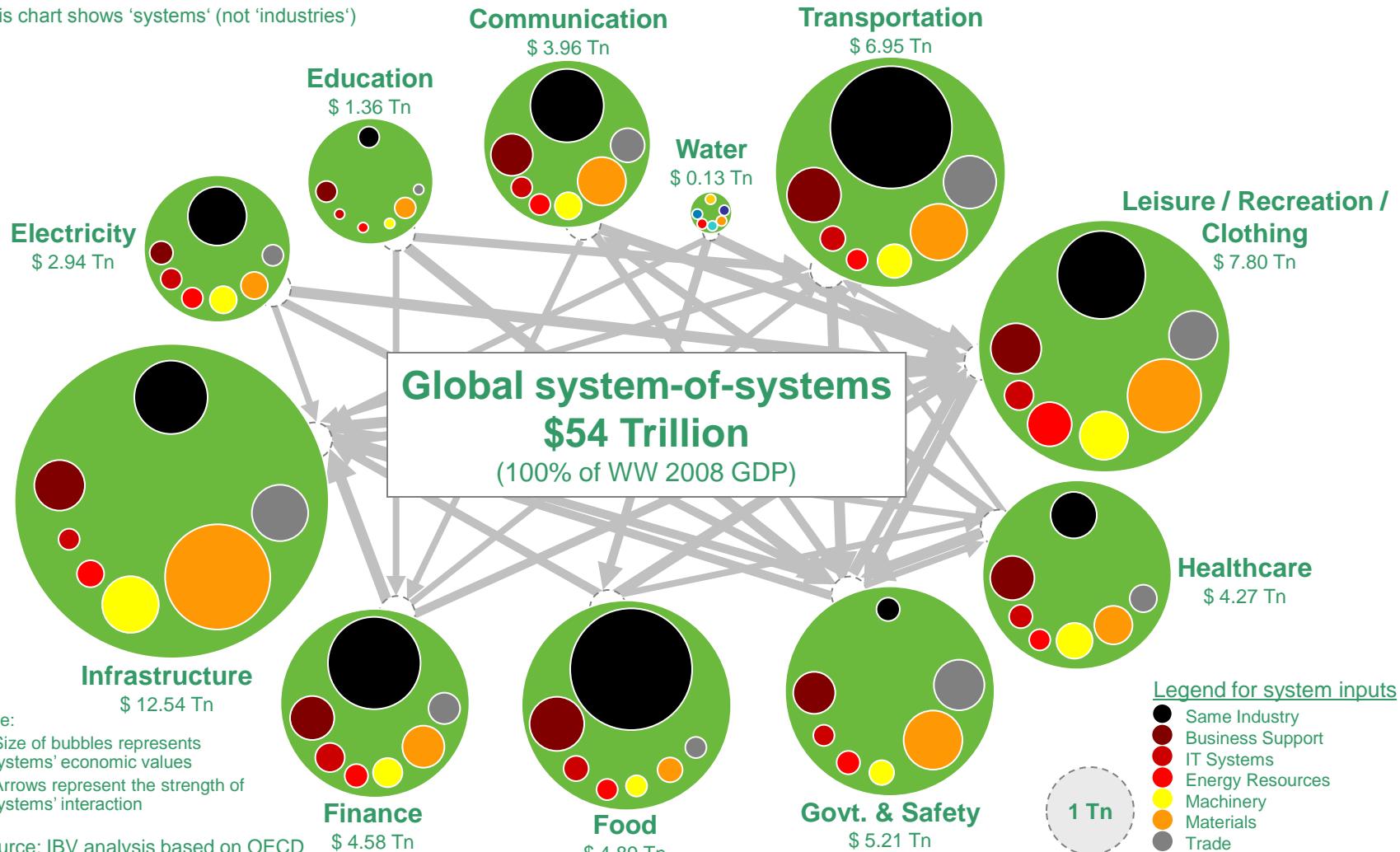
Operations research

- **Operations Research** (O.R.) is the discipline of applying advanced analytical methods to help make better decisions
- **Analytical techniques:**
 - Simulation – giving you the ability to try out approaches and test ideas for improvement
 - Optimization – narrowing your choices to the very best when there are virtually innumerable feasible options and comparing them is difficult
 - Probability and Statistics – helping you measure risk, mine data to find valuable connections and insights, test conclusions, and make reliable forecasts
 - Mathematical Modeling – algorithms and software



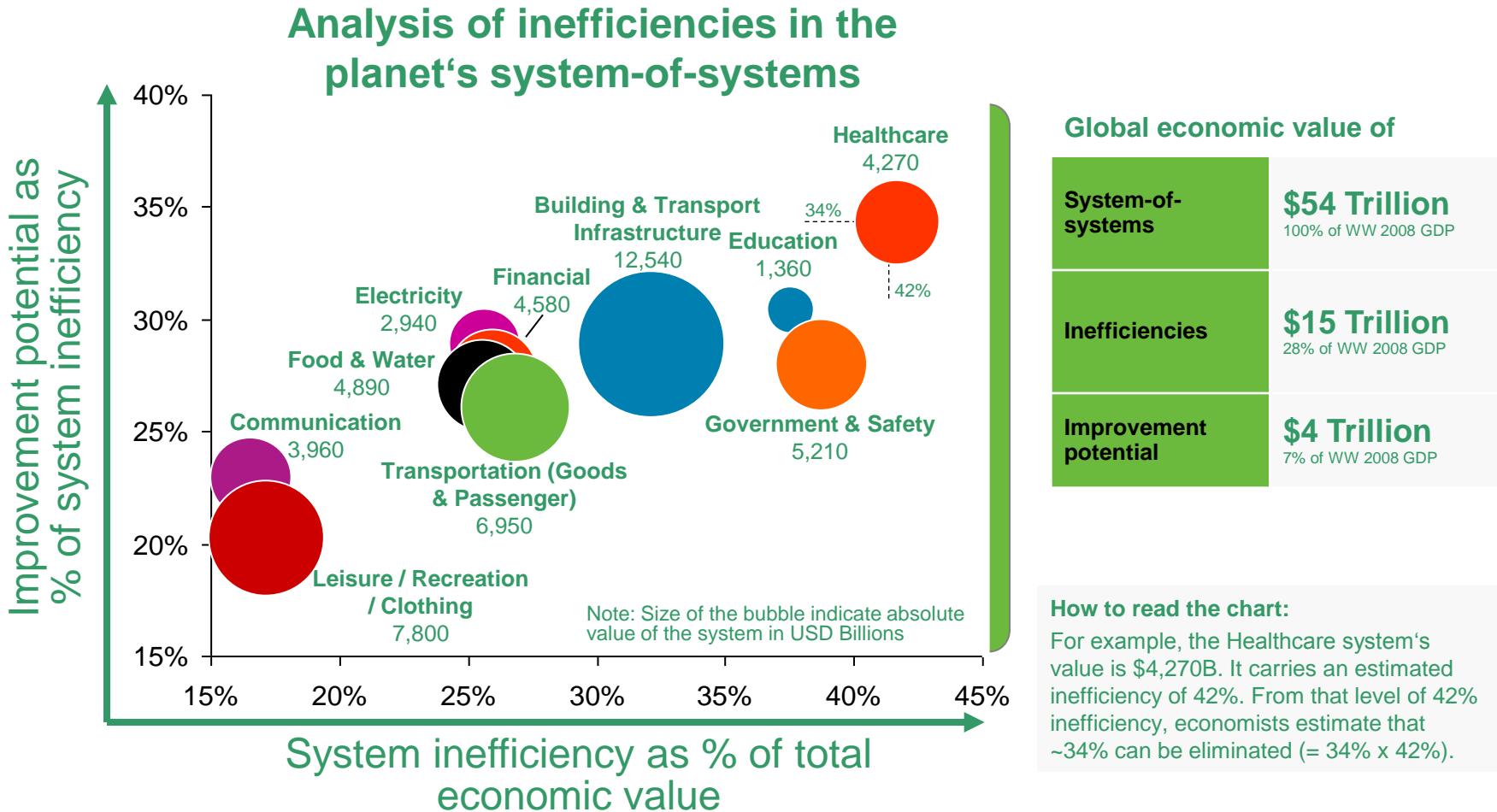
Our planet is a complex, dynamic, highly interconnected \$54 Trillion system-of-systems (OECD-based analysis)

This chart shows 'systems' (not 'industries')



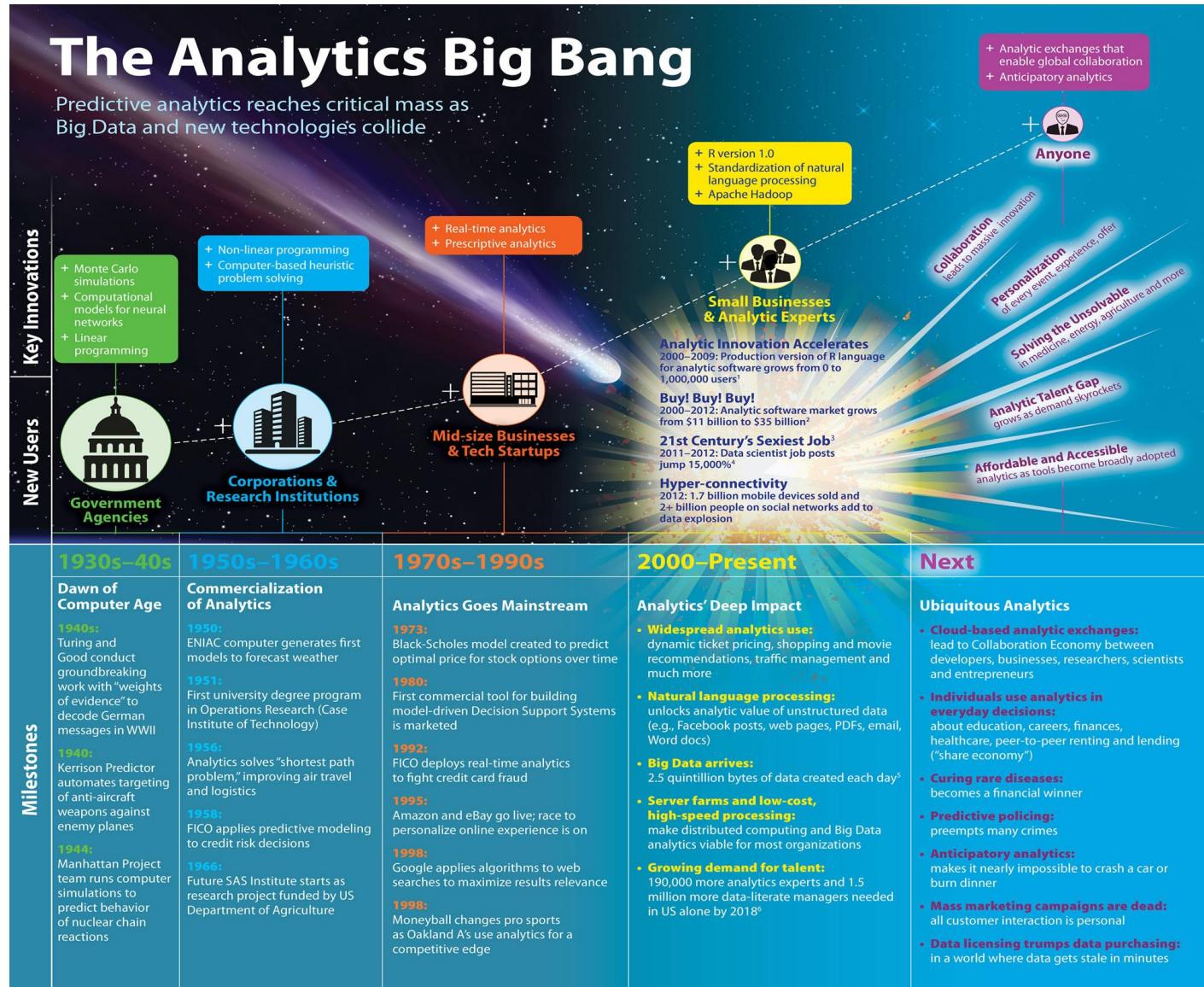
Economists estimate, that all systems carry inefficiencies of up to \$15 Tn, of which \$4 Tn could be eliminated

This chart shows 'systems' (not 'industries')



Source: IBM economists survey 2009; n= 480

History of analytics



History of business analytics

Milestones	Key Innovations	New Users	Timeline of Business Analytics Evolution		
			1930s–40s	1950s–1960s	1970s–1990s
	<ul style="list-style-type: none"> + Monte Carlo simulations + Computational models for neural networks + Linear programming  <p>Government Agencies</p>	 <p>Corporations & Research Institutions</p> <ul style="list-style-type: none"> + Non-linear programming + Computer-based heuristic problem solving 	 <p>Mid-size Businesses & Tech Startups</p> <ul style="list-style-type: none"> + Real-time analytics + Prescriptive analytics 		
			<p>1940s: Turing and Good conduct groundbreaking work with “weights of evidence” to decode German messages in WWII</p> <p>1940: Kerrison Predictor automates targeting of anti-aircraft weapons against enemy planes</p> <p>1944: Manhattan Project team runs computer simulations to predict behavior of nuclear chain reactions</p>	<p>1950: ENIAC computer generates first models to forecast weather</p> <p>1951: First university degree program in Operations Research (Case Institute of Technology)</p> <p>1956: Analytics solves “shortest path problem,” improving air travel and logistics</p> <p>1958: FICO applies predictive modeling to credit risk decisions</p> <p>1966: Future SAS Institute starts as research project funded by US Department of Agriculture</p>	<p>1973: Black-Scholes model created to predict optimal price for stock options over time</p> <p>1980: First commercial tool for building model-driven Decision Support Systems is marketed</p> <p>1992: FICO deploys real-time analytics to fight credit card fraud</p> <p>1995: Amazon and eBay go live; race to personalize online experience is on</p> <p>1998: Google applies algorithms to web searches to maximize results relevance</p> <p>1998: Moneyball changes pro sports as Oakland A's use analytics for a competitive edge</p>
			<p>2000–Present</p> <p>Analytics' Deep Impact</p> <ul style="list-style-type: none"> • Widespread analytics use: dynamic ticket pricing, shopping and movie recommendations, traffic management and much more • Natural language processing: unlocks analytic value of unstructured data (e.g., Facebook posts, web pages, PDFs, email, Word docs) • Big Data arrives: 2.5 quintillion bytes of data created each day⁵ • Server farms and low-cost, high-speed processing: make distributed computing and Big Data analytics viable for most organizations • Growing demand for talent: 190,000 more analytics experts and 1.5 million more data-literate managers needed in US alone by 2018⁶ 	<p>Next</p> <p>Ubiquitous Analytics</p> <ul style="list-style-type: none"> • Cloud-based analytic exchanges: lead to Collaboration Economy between developers, businesses, researchers, scientists and entrepreneurs • Individuals use analytics in everyday decisions: about education, careers, finances, healthcare, peer-to-peer renting and lending (“share economy”) • Curing rare diseases: becomes a financial winner • Predictive policing: preempts many crimes • Anticipatory analytics: makes it nearly impossible to crash a car or burn dinner • Mass marketing campaigns are dead: all customer interaction is personal • Data licensing trumps data purchasing: in a world where data gets stale in minutes 	



Course Outline

Course summary

- **Course title:** Introduction to Data Science and Analytics
- **Course summary:** The objective of the course is to learn analytical models and overview quantitative algorithms for solving engineering and business problems. Data science or analytics is the process of deriving insights from data in order to make optimal decisions. It allows hundreds of companies and governments to save lives, increase profits and minimize resource usage. Considerable attention in the course is devoted to applications of computational and modeling algorithms to finance, risk management, marketing, health care, smart city projects, crime prevention, predictive maintenance, web and social media analytics, personal analytics, etc. We will show how various data science and analytics techniques such as basic statistics, regressions, uncertainty modeling, simulation and optimization modeling, data mining and machine learning, text analytics, artificial intelligence and visualizations can be implemented and applied using Python. Python and IBM Watson Analytics are modeling and visualization software used in this course. Practical aspects of computational models and case studies in Interactive Python are emphasized.

Course outline

Introduction to data science and analytics

- Data science concepts
- Application areas of quantitative modeling

Python programming, data science software

- Introduction to Python
- Comparison of Python, R and Matlab usage in data science

Basic statistics

- Random variables, sampling
- Distributions and statistical measures
- Hypothesis testing
- Statistics case studies in Ipython

Overview of linear algebra

- Linear algebra and matrix computations
- Functions, derivatives, convexity

Course outline

Modeling techniques, regression

- Mathematical modeling process
- Linear regression
- Logistic regression
- Regression case studies in IPython

Data visualization and visual analytics

- Visual analytics
- IBM Watson Analytics

Simulation modeling

- Random number generation
- Monte Carlo simulations
- Simulation case studies in Ipython

Optimization

- Overview of optimization algorithms
- Optimization case studies in IPython

Course outline

Data mining and machine learning

- Classification (decision trees)
- Clustering (K-means, Fuzzy C-means, Hierarchical Clustering, DBSCAN)
- Association rules
- Data mining case studies in IPython

Cognitive computing and artificial intelligence

- Text analytics
- Social media analytics
- Neural networks
- Spatio-temporal analytics
- Cognitive computing case studies in IPython

Storytelling based on analytics, analytical decision making

- Validating analytics
- Storytelling based on analytics
- Decision-making based on analytics

Assignments, exams and grading (tentative)

Assignment #1 – Solving an analytics problem in Python (**15%**)

- Individual assignment.

Assignment #2 – Solving an analytics problem in Python (**15%**)

- Individual assignment.

Course Project (Assignment #3 and Assignment #4) – Analytics of news, social media and web-sites via machine learning and cognitive computing in Python (**30%**)

- Group project (groups of 6 students), the same groups as for In-Class Presentations.

In-Class Group Presentation (15%)

- Group presentations of up to 10-12 minutes are required to cover topics related to additional course materials and the course project.

Final Exam (25%)

- For the final exam you may be responsible for analyzing, computing and writing up a solution to case problems. Each solution must be completed individually.

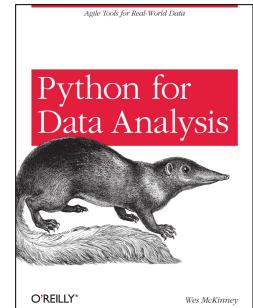
Notes

- If a student gets less than 50% mark at the Final Exam, her/his course mark will be reduced one letter grade down. E.g., a student got 14 pts (Assg 1) + 13 pts (Assg 2) + 28 pts (Course Project) + 14 pts (In-Class Presentation) + 12 pts (Final Exam) = 81 pts that corresponds to A- course mark, but because a student got 12 pts out of 25 pts at the Final Exam (less than 50%), the course mark will be reduced from A- to B+.

Course materials and readings

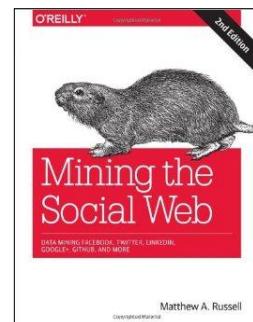
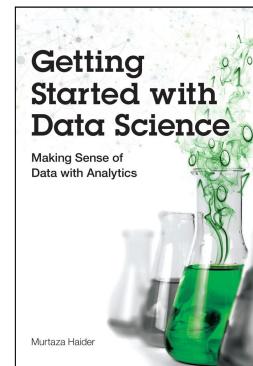
Required

- **Course slides** by O. Romanko and D. Rosu, 2017
[Blackboard](#)



Optional

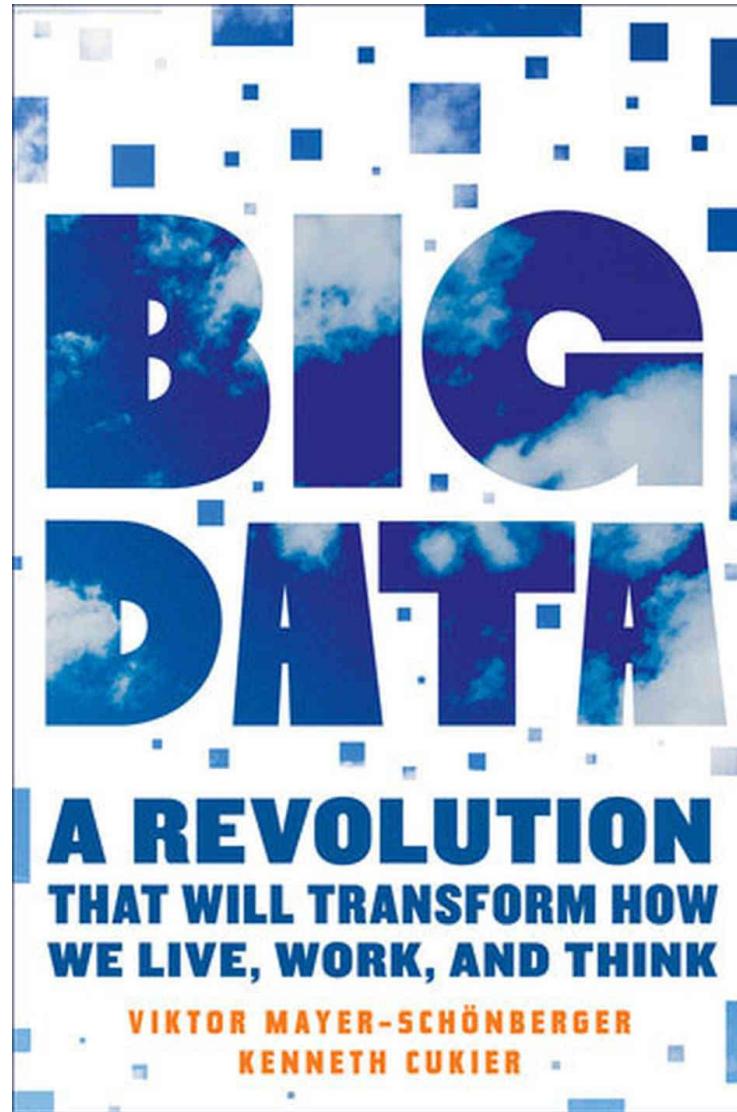
- **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython** by W. McKinney, 2012
<https://www.amazon.com/Python-Data-Analysis-Wrangling-IPython/dp/1449319793/>
- **Getting Started with Data Science: Making Sense of Data with Analytics** by M. Haider, 2015
<https://www.amazon.com/Getting-Started-Data-Science-Analytics/dp/0133991024/>
- **Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More** by M. Russell, 2013
<https://www.amazon.com/Mining-Social-Web-Facebook-LinkedIn/dp/1449367615/>



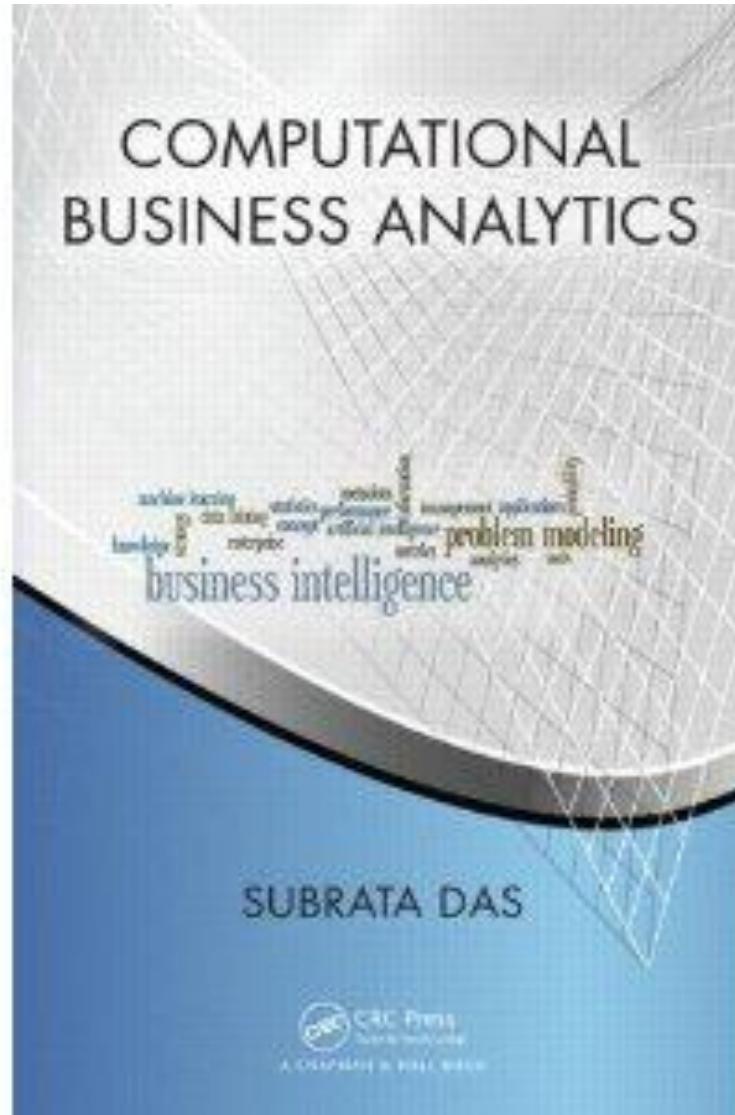
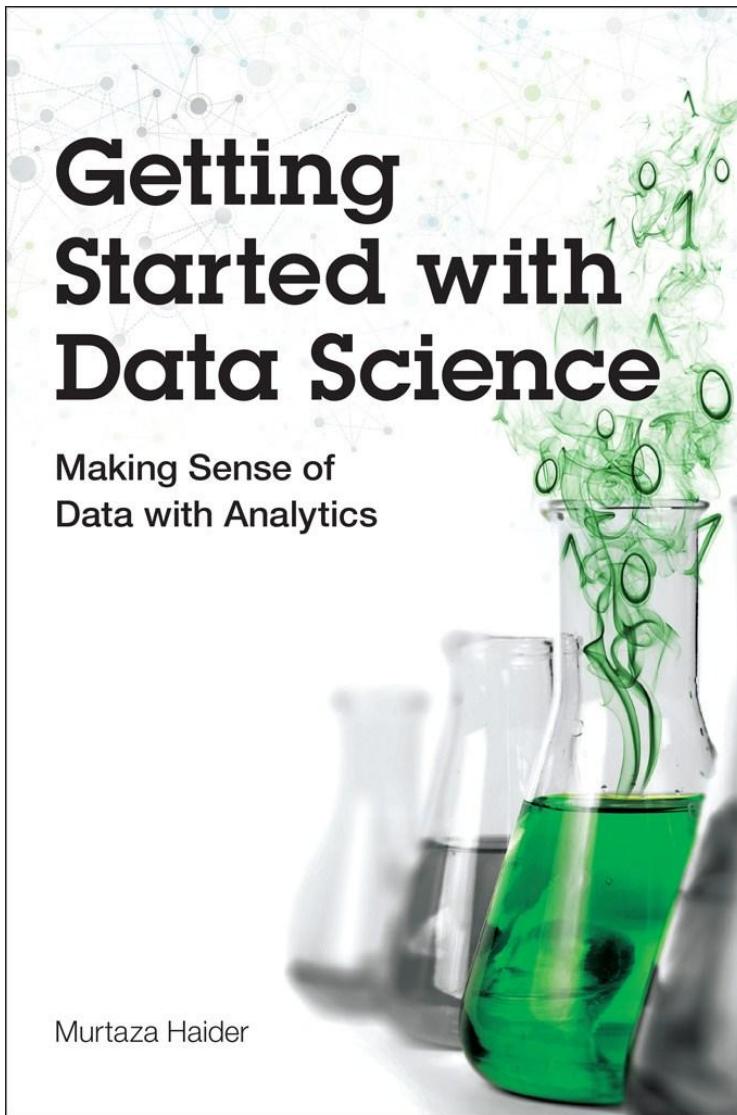


Recommended Literature

Literature



Literature



Literature

Agile Tools for Real-World Data

Python for Data Analysis



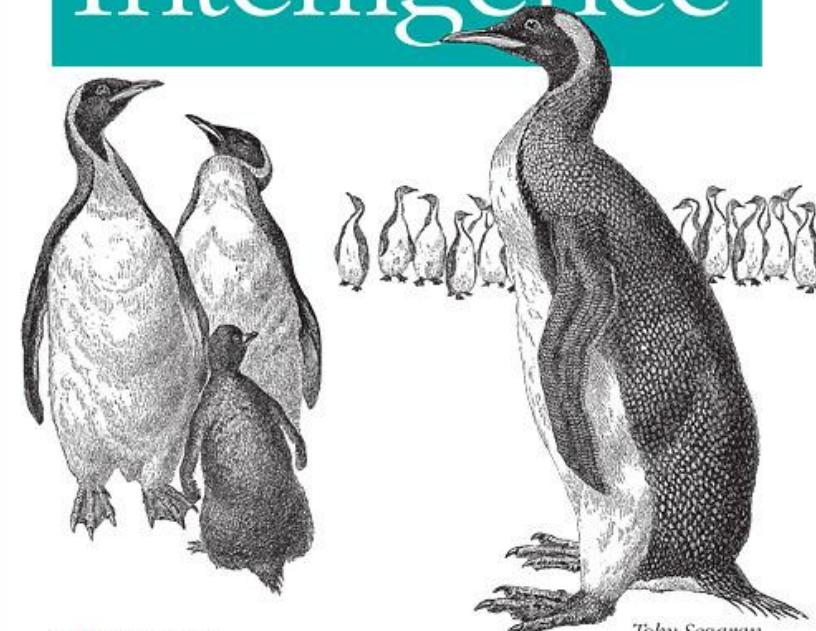
O'REILLY®

Wes McKinney

Building Smart Web 2.0 Applications

Programming

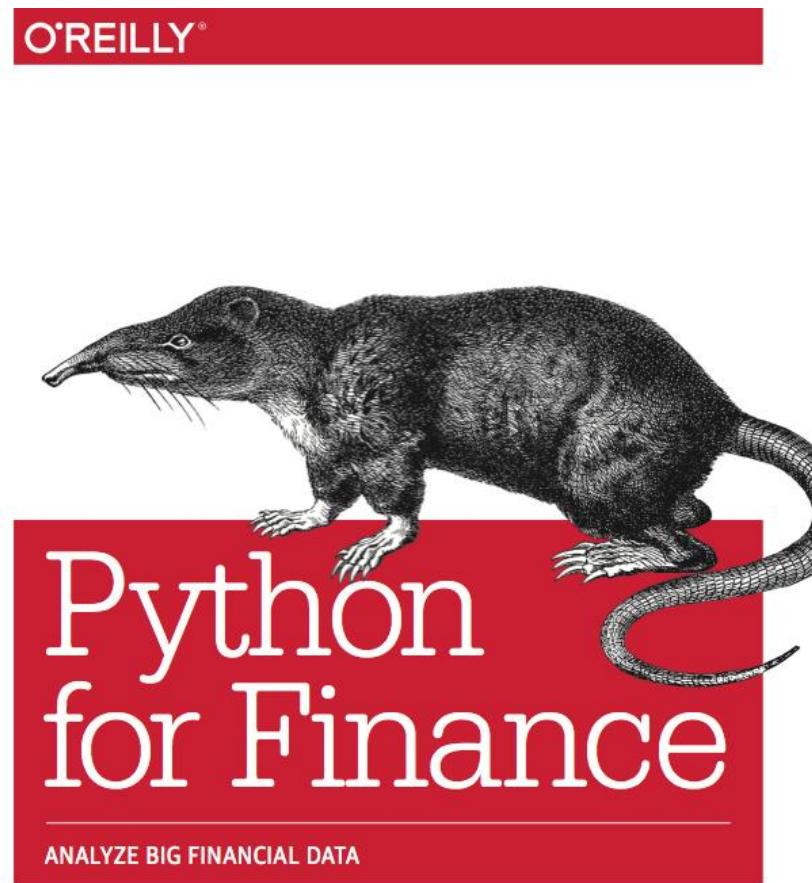
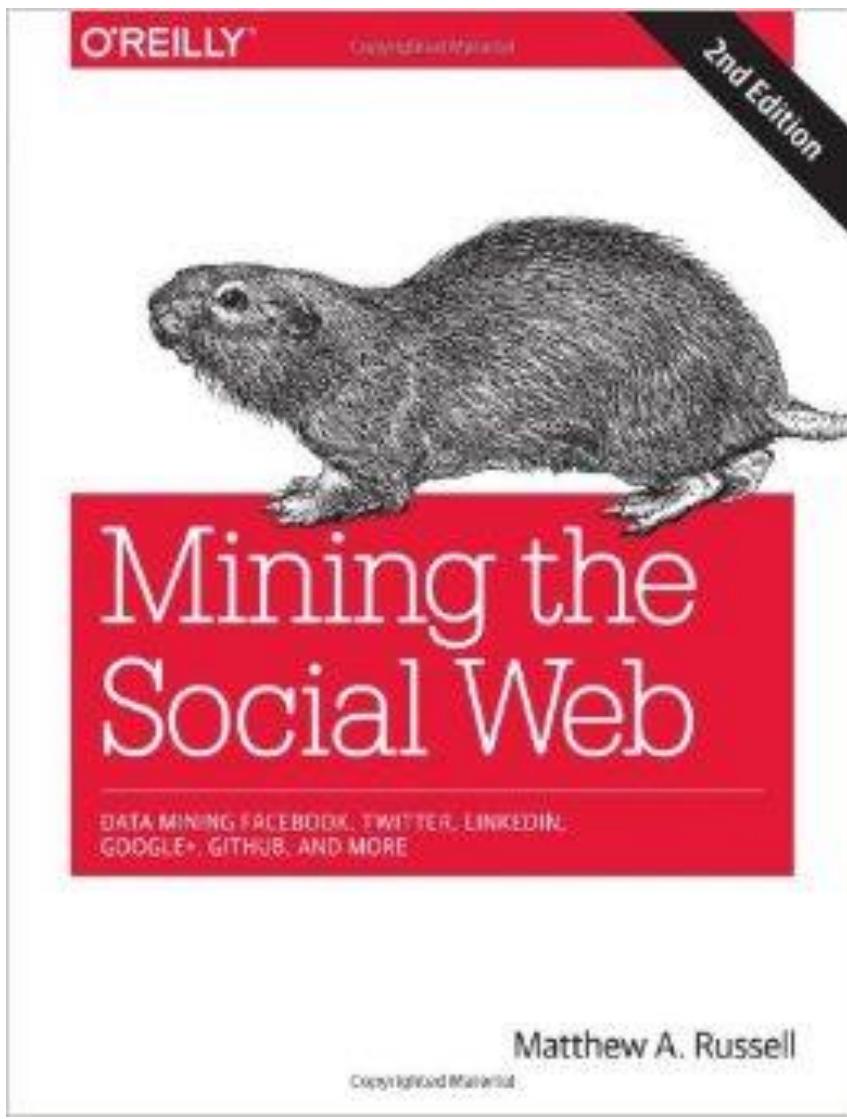
Collective Intelligence



O'REILLY®

*Toby Segaran
Foreword by Tim O'Reilly*

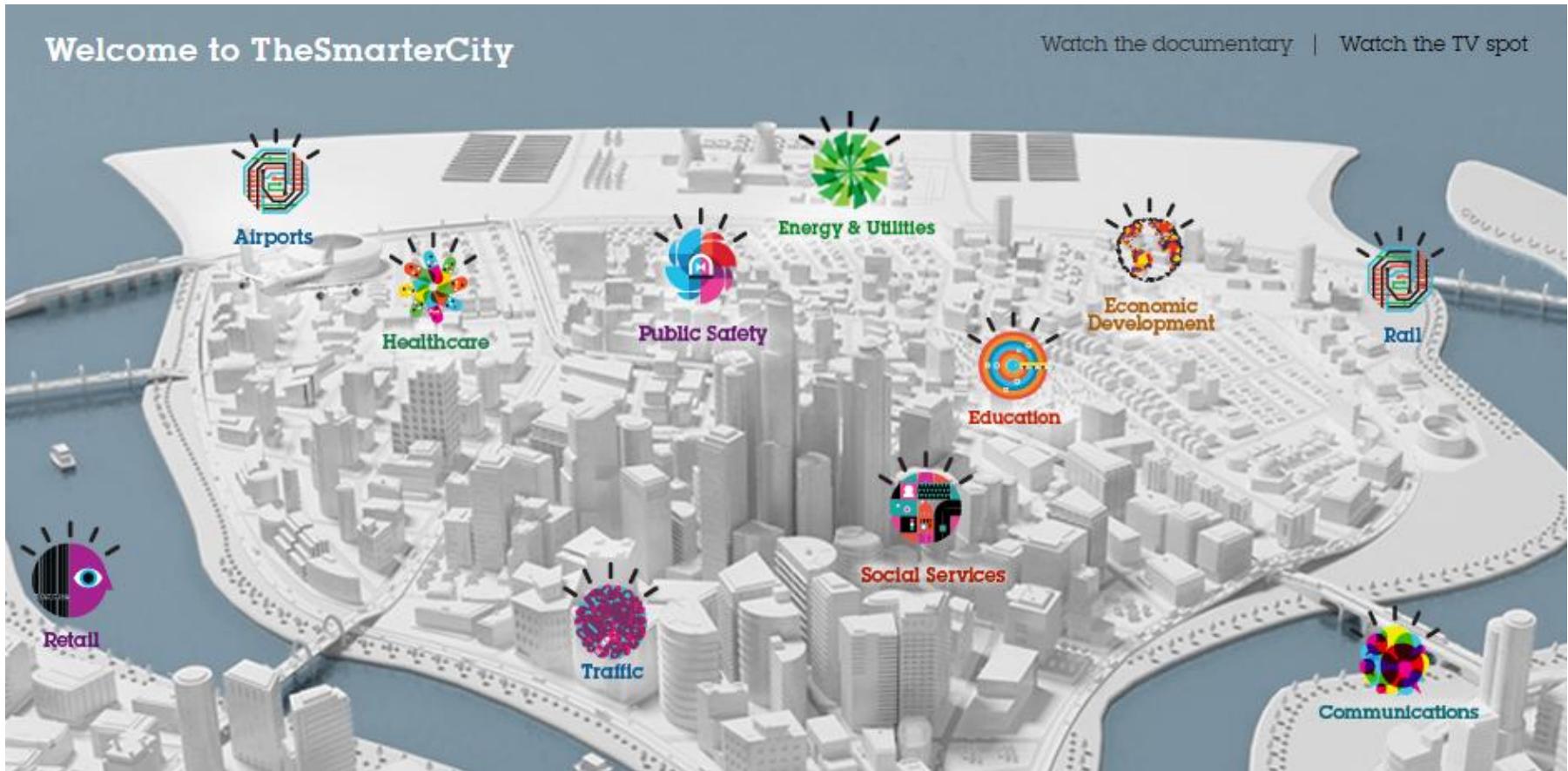
Literature





Analytics Examples

Smarter Cities



Use of camera phones at the Papal inauguration in 2005 and 2013



Data reveals hidden city dynamics



We can collect information from almost everything to make better decisions

30 billion

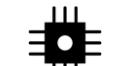
RFID tags
embedded into our
world and across
entire ecosystems

1 billion

Camera phones in
existence able to
document accidents,
damage, and crimes

85%

Of new automobiles
will contain event data
recorders collecting
travel information



Instrumented



Interconnected



Intelligent



What is big data?

Big data are datasets that grow so large that they become awkward to work with using on-hand database management tools.

Difficulties include capture, storage, search, sharing, analytics, and visualizing.

Source: Wikipedia

Big social data



Applications of big data analytics

Smarter Healthcare



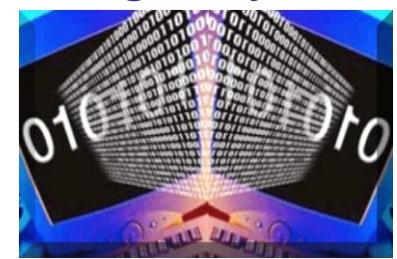
Multi-channel



Finance



Log Analysis



Homeland Security



Traffic Control



Telecom



Search Quality



Manufacturing



Trading Analytics



Fraud and Risk

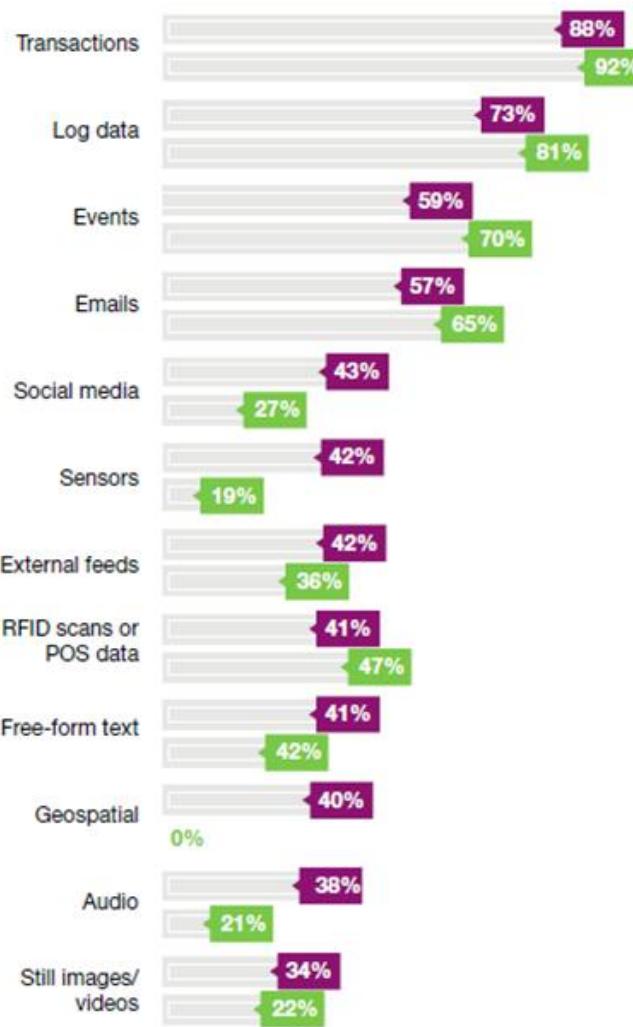


Retail: Churn, NBO

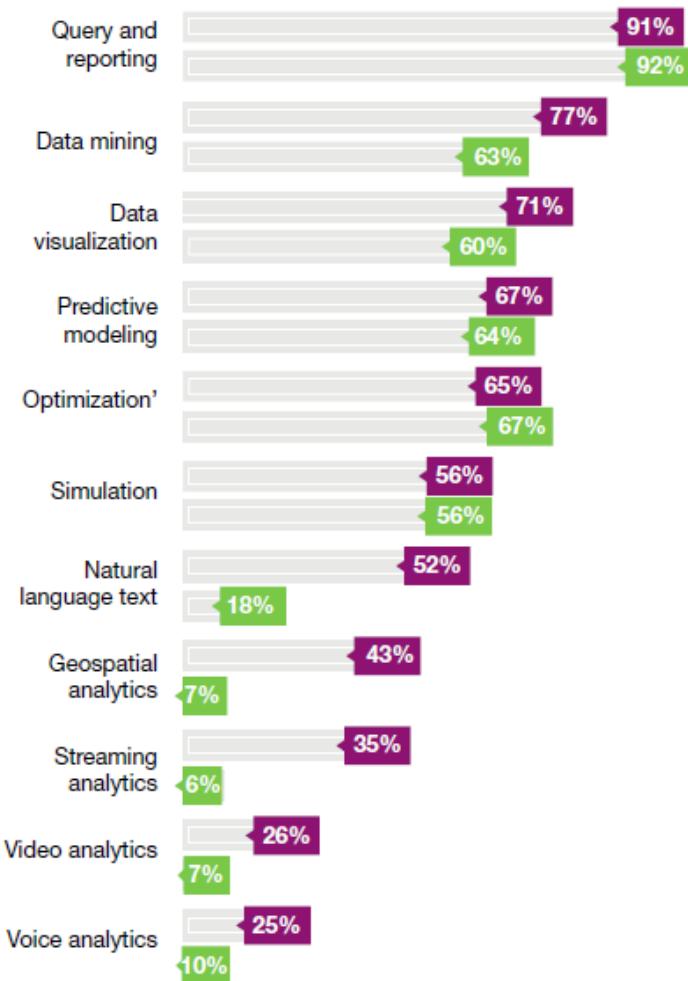


Use of Big Data globally and in the financial sector

Big data sources



Analytics capabilities



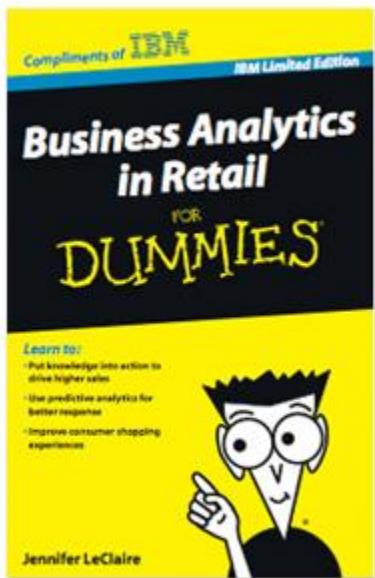
■ Global

■ Banking and Financial Markets

Global

Banking and Financial Markets

Multiple responses accepted



IBM Predictive Analytics in

RETAIL

What will your customers want next?



Fitting room analytics

Good



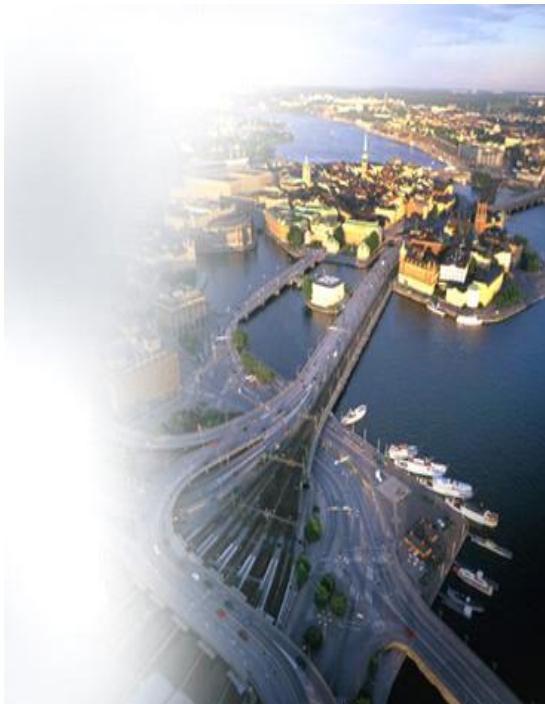
Bad



Source: Adme

Intelligent transport systems

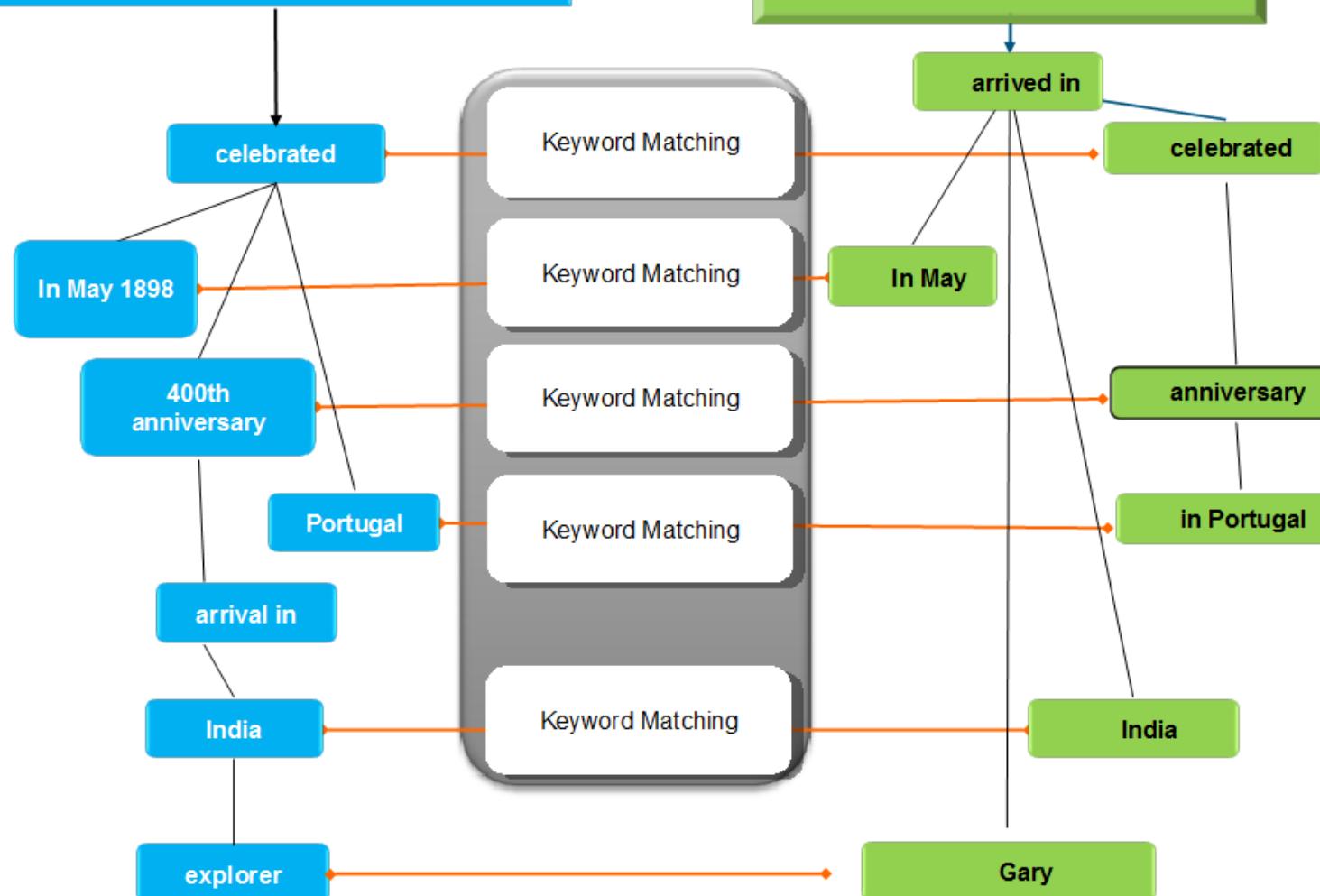
- **Real time monitoring & forecasting of congestion in cities enables real time action to reduce traffic and emissions**
 - Can charge drivers at point of use for access to city centers
- **Stockholm Congestion Tax Project**
 - Involves 18 barrier-free control points
 - Allows differentiated pricing by time of day, congestion level, and potentially emissions level
 - Results:
 - Traffic reduced by 100,000 vehicle passages per day (25%)
 - Public transportation passengers increased by 40,000 / day
 - Congestion during peak hours and CO₂ emissions were dramatically reduced



Artificial intelligence

In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India.

In May, Gary arrived in India after he celebrated his anniversary in Portugal.



Artificial intelligence

In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India.

celebrated

Portugal

May 1898

400th anniversary

arrival in

India

explorer

- Search Far and Wide
- Explore many hypotheses
- Find Judge Evidence
- Many inference algorithms

Temporal Reasoning

Statistical Paraphrasing

GeoSpatial Reasoning

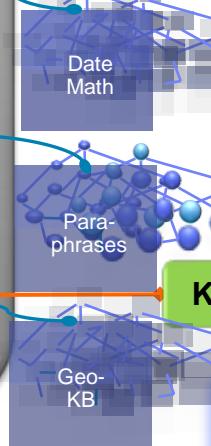
On the 27th of May 1498, Vasco da Gama landed in Kappad Beach

landed in

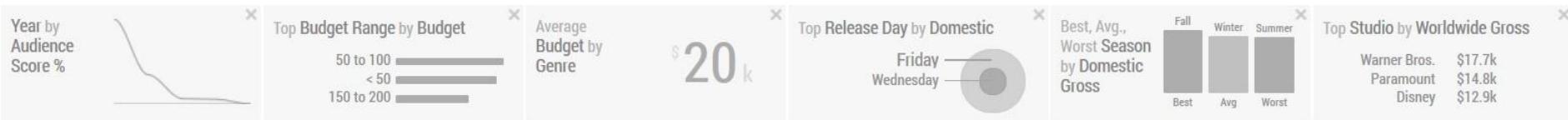
27th May 1498

Kappad Beach

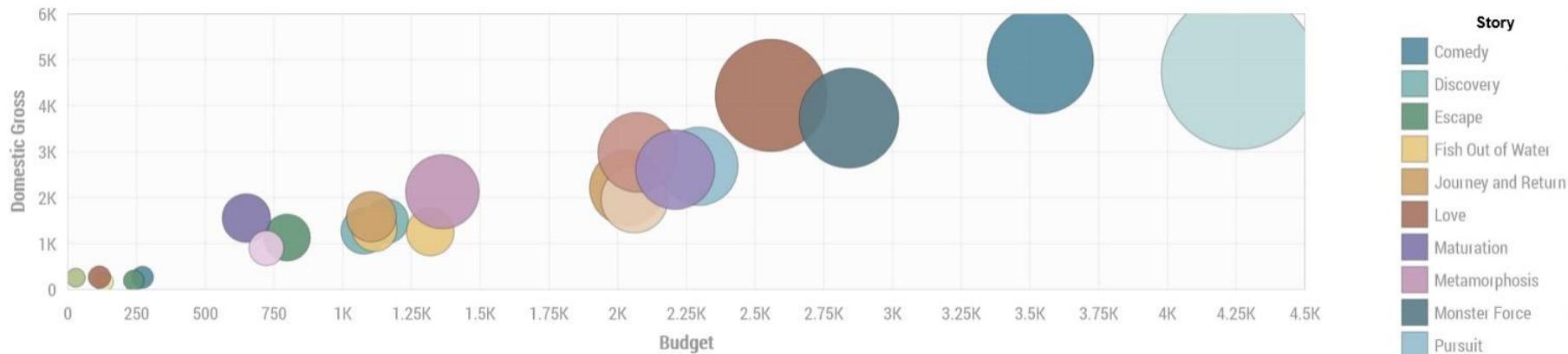
Vasco da Gama



IBM Watson Analytics



What is the relationship between **Budget** and **Domestic Gross** by **Story** ?



IBM Watson Analytics





Text Analytics

Text analytics and sentiment analysis

 ShareThis



Sentiment analysis of tweets



Cloud

IBM Cloud – Data & Analytics services

Docs 104 Trial Days Remaining ▾ University of Toronto | US South : oleksandr.ro...o@utoronto.ca : dev 

IBM Bluemix Catalog Catalog Support Manage

Data & Analytics

All Categories >

Essential data services; limitless possibilities.

Infrastructure Compute Storage Network Security Containers VMware	Apache Spark IBM Analytics for Apache Spark for Bluemix. 	BigInsights for Apache Hadoop Provision managed Apache Hadoop and Spark clusters within minutes. 	BigInsights for Apache Hadoop (Subscription) Provision managed bare metal Apache Hadoop clusters for 
Platform Boilerplates APIs Application Services Blockchain Cloud Foundry Apps Data & Analytics DevOps Finance Functions Integrate Internet of Things Mobile Network Security Watson	Cloudant NoSQL DB Cloudant NoSQL DB is a fully managed data layer designed for  	Compose Enterprise IBM Compose Enterprise is a service which provides a private 	Compose for Elasticsearch Elasticsearch combines the power of a full text search engine with the index 
	Compose for etcd etcd is a key/value store developers can use to hold the always-correct c  	Compose for JanusGraph JanusGraph is a scalable graph database optimized for storing an  	Compose for MongoDB MongoDB with its powerful indexing and querying, aggregation and wide 
	Compose for MySQL MySQL is probably the most popular open source relational database in t  	Compose for PostgreSQL Postgres is a powerful, open source object-relational database that is h 	Compose for RabbitMQ RabbitMQ asynchronously handles the messages between your application 
	Compose for Redis Redis is an open-source, blazingly fast, key/value low maintenance store. 	Compose for RethinkDB RethinkDB is a JSON document based, distributed database with an integrat 	Compose for ScyllaDB ScyllaDB is a highly performant, in-place replacement for the Cassandra  

IBM Cloud – Watson services

Docs 104 Trial Days Remaining ▾ University of Toronto | US South : oleksandr.ro...o@utoronto.ca : dev 

IBM Bluemix Catalog Catalog Support Manage

All Categories >

Watson

Build cognitive apps that help enhance, scale, and accelerate human expertise.

Infrastructure

- Compute
- Storage
- Network
- Security
- Containers
- VMware

Platform

- Boilerplates
- APIs
- Application Services
- Blockchain
- Cloud Foundry Apps
- Data & Analytics
- DevOps
- Finance
- Functions
- Integrate
- Internet of Things
- Mobile
- Network
- Security
- Watson

Conversation
Add a natural language interface to your application to automate
 Lite IBM

Discovery
Add a cognitive search and content analytics engine to applications.
 Lite IBM

Document Conversion
Converts a HTML, PDF, or Microsoft Word™ document into a normalized
 IBM

Language Translator
Translate text from one language to another for specific domains.
 Lite IBM

Natural Language Classifier
Natural Language Classifier performs natural language
 IBM

Natural Language Understanding
Analyze text to extract meta-data from content such as concepts,
 Lite IBM

Personality Insights
The Watson Personality Insights derives insights from transactions
 Lite IBM

Retrieve and Rank
Add machine learning enhanced search capabilities to your
 IBM

Speech to Text
Low-latency, streaming transcription
 IBM

Text to Speech
Synthesizes natural-sounding speech from text.
 IBM

Tone Analyzer
Tone Analyzer uses linguistic analysis to detect three types of
 Lite IBM

Visual Recognition
Find meaning in visual content!
Analyze images for scenes, objects
 IBM

IBM Cloud – Visual Recognition service

Docs 104 Trial Days Remaining ▾ University of Toronto | US South : oleksandr.romanko@utoronto.ca : dev ⓘ

IBM Bluemix Catalog Catalog Support Manage

Visual Recognition

Find meaning in visual content! Analyze images for scenes, objects, faces, and other content. Choose a default model off the shelf, or create your own custom classifier. Find similar images within a collection. Develop smart applications that analyze the visual content of images or video frames to understand what is happening in a scene.

IBM

[View Docs](#)

AUTHOR IBM
PUBLISHED 09/18/2017
TYPE Service
LOCATION US South, Sydney, United Kingdom

Service name: Visual Recognition-yh

Credential name: Credentials-1

Select region to deploy in: US South Choose an organization: oleksandr.romanko@utoronto.ca Choose a space: dev

Connect to: Leave unbound

Features

- **General Classification**
Generate class keywords that describe the image. Use your own images, or extract relevant image URLs from publicly accessible webpages for analysis.
- **Visual Training**
Create custom, unique visual classifiers. Use the service to recognize custom visual concepts that are not available with general classification.
- **Face Detection**
Detect human faces in the image. This service also provides a general indication of age range and gender of faces.
- **Similar Image Search (BETA)**
Upload and search through image collections to find visually similar images.

Need Help?
[Contact Bluemix Sales](#)

Estimate Monthly Cost
[Cost Calculator](#)

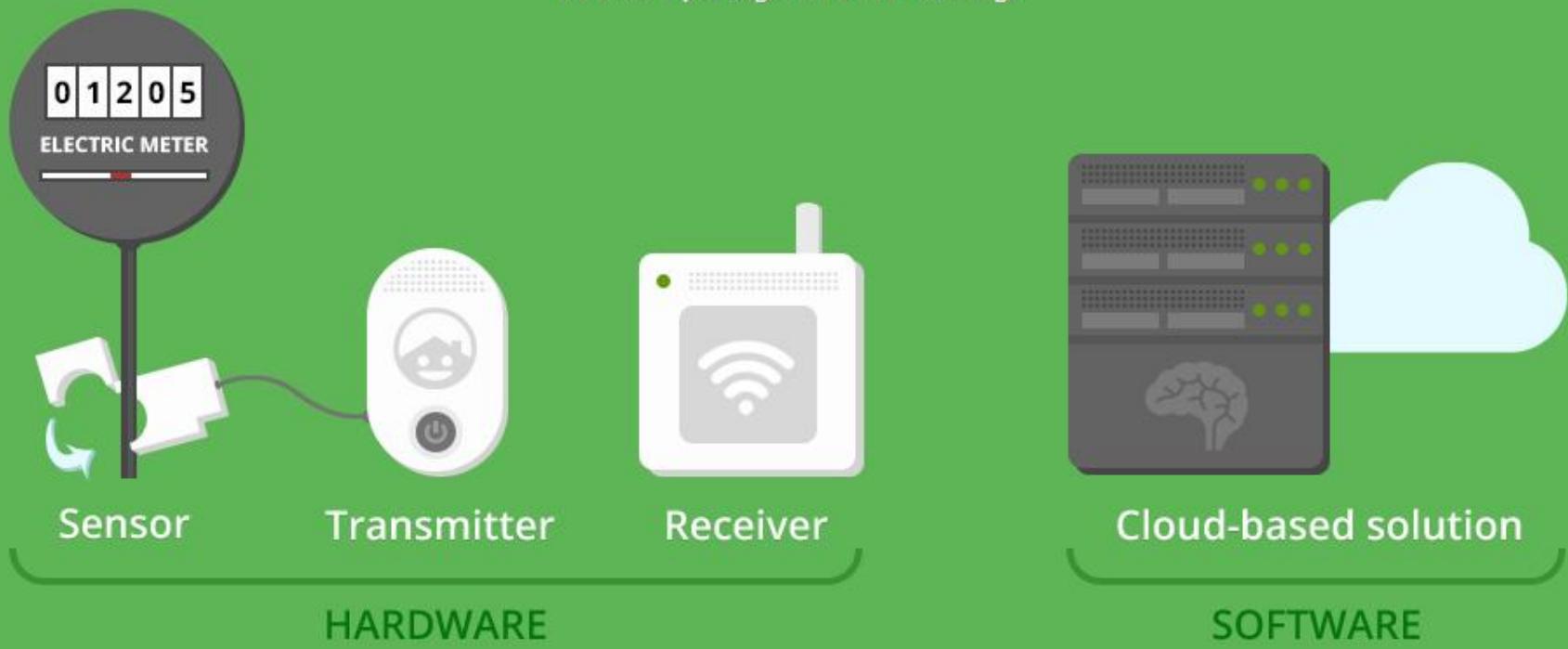
Create

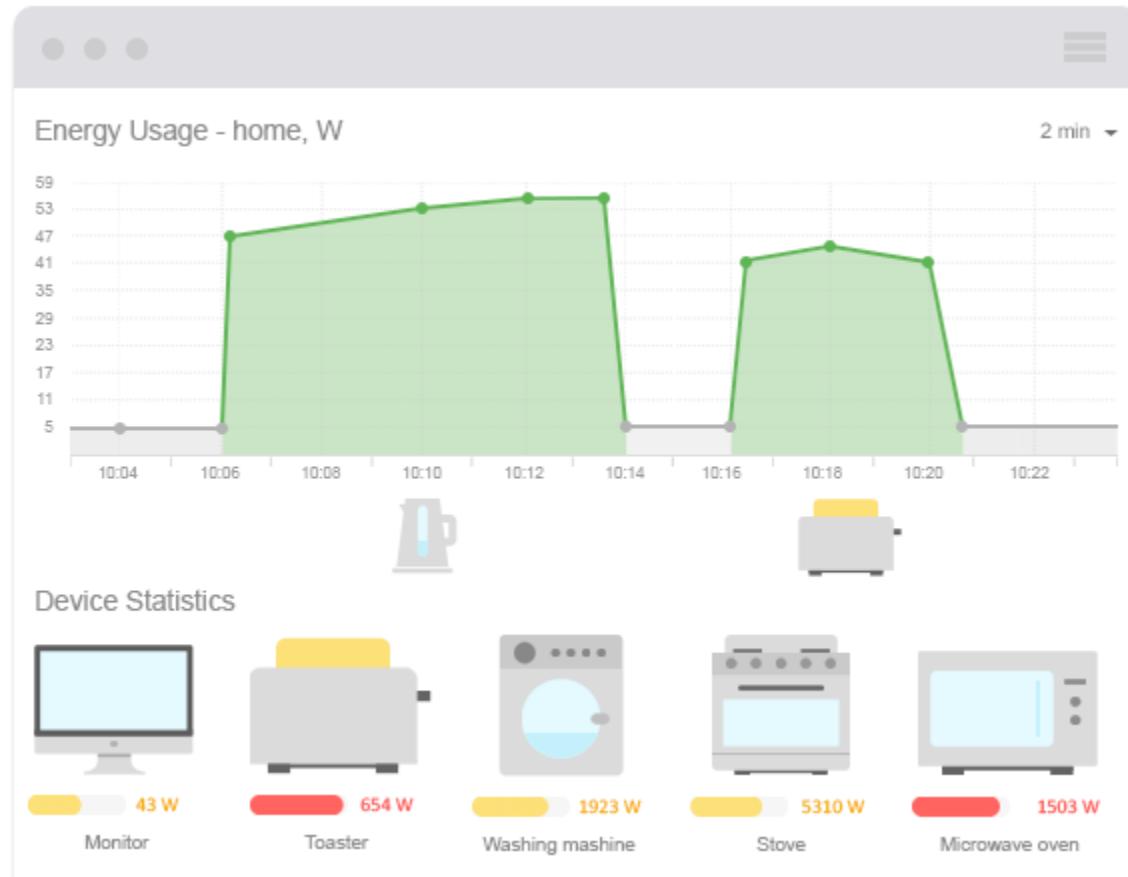


Startups

IT IS SIMPLE — JUST CLICK AND SIGN IN!

Basically, Ecois.me is a combination of hardware and software-based service which makes analysis of energy spending. Remember, you need only SINGLE Ecois.me hardware package for its functioning!





Total consumption for the month:

The best - 100 kWh
You - 140 kWh
Average - 170 kWh

Philips HD4656 Kettle 2984 W

In The Kitchen Roland Edvarlum

Buy a teapot. If you are making tea for more than one person you can save on teabags and water. Use a tea cosy or use a woolly hat instead to keep the tea warm. Buy a teapot. If you are making tea for more than one

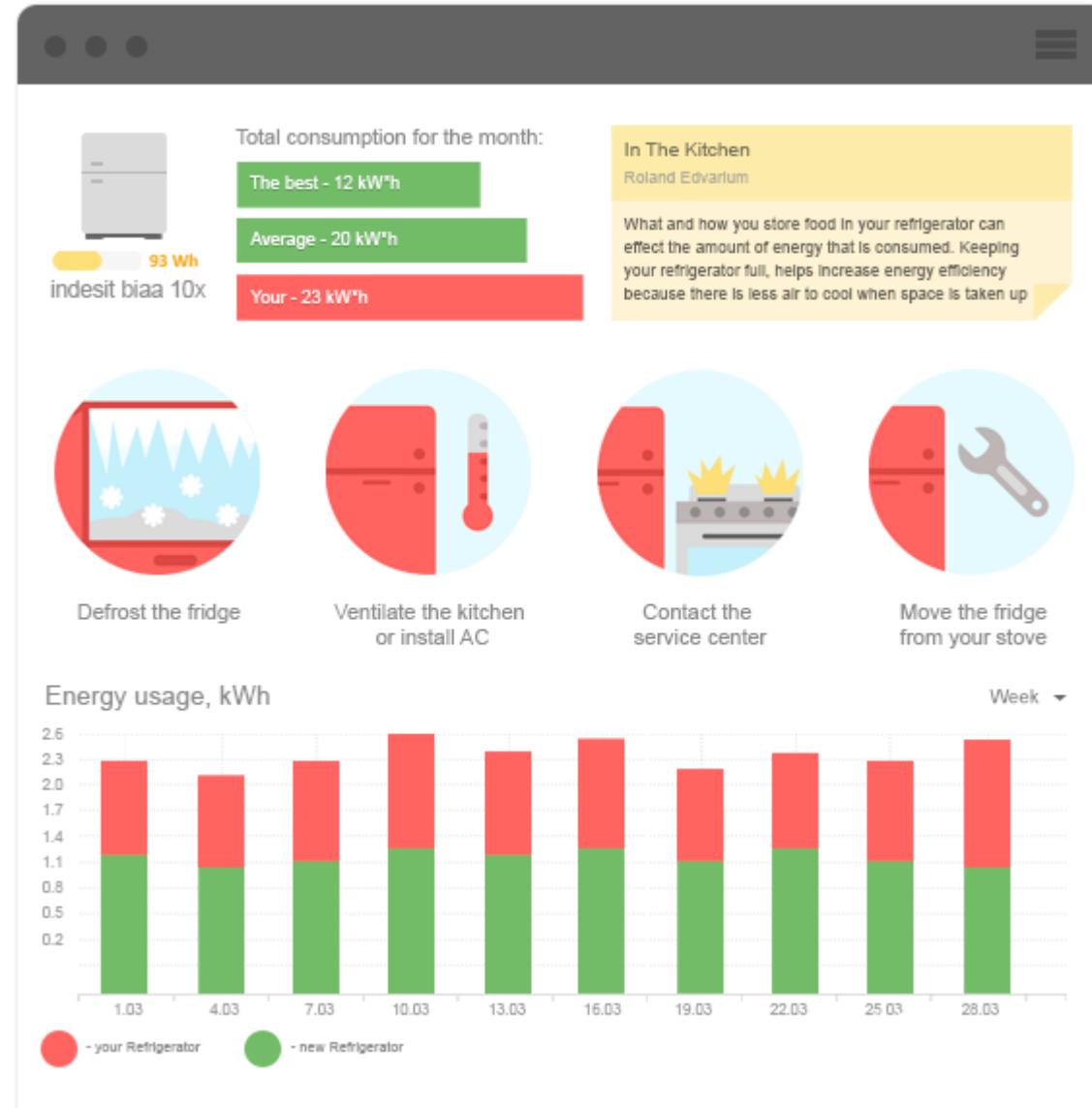
Energy usage, kW

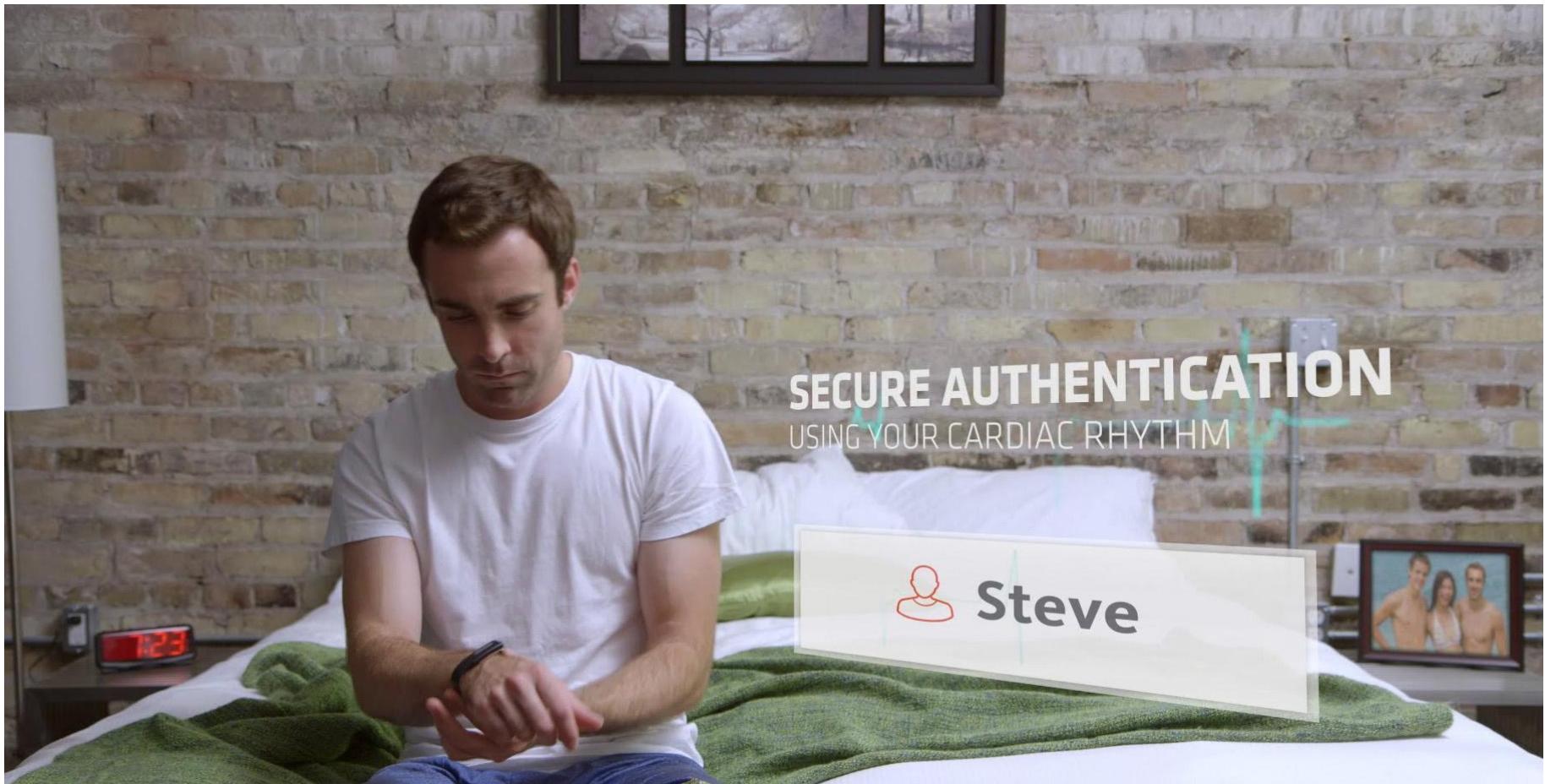
Week ▾

Date	new Kettle (kW)	your Kettle (kW)	Total (kW)
1.03	~2.8	~3.2	~6.0
4.03	~1.5	~3.0	~4.5
7.03	~1.5	~3.0	~4.5
10.03	~2.8	~4.4	~7.2
13.03	~2.8	~4.4	~7.2
16.03	~2.8	~3.5	~6.3
19.03	~2.8	~3.5	~6.3
22.03	~3.0	~4.2	~7.2
25.03	~2.0	~2.4	~4.4
28.03	~3.0	~4.2	~7.2

- your Kettle - new Kettle

Kettle BRAUN WK 500 White save 50 EUR/year [Buy it now for 60 EUR](#)







Online and In-Person Education

Coursera (coursera.org)

coursera

Courses Partners About ▾ | Oleksandr... ▾

YOUR COURSES



An Introduction to Interactive Programming in Python

Apr 15th (9 weeks long)

Rice University

Jun 17th

[Go to class](#)

[View course info](#) | [Un-enroll](#)



Introduction to Computational Finance and Financial Econometrics

Ended 7 months ago

University of Washington

Sep 4th (10 weeks long)

[View class archive](#)

[View course info](#) | [Un-enroll](#)



Computational Investing, Part I

Ended 6 months ago

Get statement of accomplishment. View grade.

Georgia Institute of Technology

Oct 22nd (9 weeks long)

[View class archive](#)

[View course info](#)



Computational Methods for Data Analysis

Ended 3 months ago

University of Washington

Jan 7th (10 weeks long)

[View class archive](#)

[View course info](#) | [Un-enroll](#)



Financial Engineering and Risk Management

Ended a month ago

You did not earn a statement of accomplishment.

Columbia University

Feb 24th (10 weeks long)

[View class archive](#)

[View course info](#) | [Un-enroll](#)

EdX (edx.org)



Home > All Subjects > Data Analysis & Statistics > The Analytics Edge



The Analytics Edge

Through inspiring examples and stories, discover the power of data and use analytics to provide an edge to your career and your life.



Massachusetts
Institute of
Technology

Archived

Future Dates To Be Announced

Enroll Now

I would like to receive email from MIT and learn about its other programs.

About this course

128 Reviews 5/5 ★★★★☆

This is an Archived Course

EdX keeps courses open for enrollment after they end to allow learners to explore content and continue learning. All features and materials may not be available, and course content will not be updated. Check back often to see when new course start dates are announced.

See more

What you'll learn

- An applied understanding of many different analytics methods, including linear regression, logistic regression, CART, clustering, and data visualization
- How to implement all of these methods in R
- An applied understanding of mathematical optimization and how to solve optimization models in spreadsheet software

	Length:	12 weeks
	Effort:	10 - 15 hours/week
	Price:	FREE Verified Certificate option closed
	Institution:	MITx
	Subject:	Data Analysis & Statistics
	Level:	Intermediate
	Languages:	English
	Video Transcripts:	English

CognitiveClass MOOC

<http://CognitiveClass.ai>

The screenshot shows the homepage of CognitiveClass.ai. At the top, there's a navigation bar with links for Learning Paths, Courses, Badges, Business, Competitions, a search bar, and user options like Login and Sign Up. The main header features a close-up image of a person's eye and the text "Data Science and Cognitive Computing Courses". Below this, a call-to-action button says "Free Courses Sign Up". The next section, titled "What are the benefits?", lists three items: "IT'S FREE" (with a price tag icon), "EARN BADGES" (with a sunburst icon), and "EXPAND YOUR KNOWLEDGE" (with a computer monitor icon). The bottom section, "Follow learning paths to maximize your potential", includes four numbered steps: 1) Select a Learning Path (with a QR code icon), 2) Complete Courses (with a target icon), 3) Earn Badges (with a sunburst icon), and 4) Show off your Badges (with a globe icon).

Learning Paths Courses Badges Business Competitions Q Explore new learning opportunities

Login Sign Up

Data Science and Cognitive Computing Courses

Build Data Science and Cognitive Computing skills for free today.

Free Courses Sign Up

What are the benefits?

- IT'S FREE Our courses are free so you have nothing to lose!
- EARN BADGES Earn badges for your portfolio
- EXPAND YOUR KNOWLEDGE We have courses for all skill levels

Follow learning paths to maximize your potential

- 1) Select a Learning Path
- 2) Complete Courses
- 3) Earn Badges
- 4) Show off your Badges

CognitiveClass MOOC (<http://CognitiveClass.ai>)



Learning Paths Courses ▾ Badges ▾ Business Competitions

Explore new learning opportunities

Login Sign Up

- Free courses, free study materials
- Cloud-based sandbox for exercises
- 800000+ registered students
- 50+ courses

COURSES



**Build Your Own Chatbot
(Beta)**

Cognitive Class CB0103EN

Beginner



Big Data 101

Cognitive Class / Fireside Analytics Inc.

BD0101EN

Beginner



Data Science 101

Cognitive Class / Fireside Analytics Inc.

DS0101EN

Beginner



Hadoop 101

Cognitive Class BD0111EN

Beginner



Python 101

Cognitive Class PY0101EN

Beginner



R 101

Cognitive Class RP0101EN

Beginner



Scala 101

LightBend SC0101EN

Beginner



Deep Learning 101

DeepLearningTV ML0115EN

Intermediate



**Deep Learning with
TensorFlow**

Cognitive Class ML0120EN

Advanced



Text Analytics 101

Cognitive Class TA0105

Beginner



Watson Analytics 101

Cognitive Class WA0101EN

Beginner



Spark Fundamentals I

Cognitive Class BD0211EN

Beginner

TED and TEDx



Talks	TED Conferences	TED Conversations	About TED
Speakers	TEDx Events	TED Community	TED Blog
Playlists <small>NEW</small>	TED Prize	TED-Ed <small>ED</small>	TED Initiatives
Translations	TED Fellows		

 Search

Riveting talks by remarkable people, free to the world



How to live before you die

Watch Steve Jobs' inspiring call to dream and create.



New: Ryan Coogler's favorite talks

Watch 6 talks that inspired the director of *Fruitvale Station*.



New TED Book

Learning from the Voices in my Head, by TED speaker Eleanor Longden.

Resize by:

- Newest releases
- Date filmed
- Most viewed
- Most emailed this week
- Most comments this week
- Most popular this month
- Rated jaw-dropping
- ... persuasive
- ... courageous
- ... ingenious
- ... fascinating
- ... inspiring
- ... beautiful
- ... funny
- ... informative

Show talks related to:

- all
- technology
- entertainment
- design
- business
- science
- global issues

[View all topics »](#)

 NEW TODAY Michael Sandel: Why we shouldn't trust markets with our civic life	 NEW TODAY Amy Webb: How I hacked online dating	 NEW TODAY Malcolm Gladwell: The unheard story of David and Goliath	 NEW TODAY Michael Porter: Why business can be good at solving social problems	 Kelli Swazey: Life that doesn't end with death
 James Flynn: Why our IQ levels are higher than our grandparents'	 Jason Pontin: Can technology solve our big problems?	 Fabian Oefner: Psychedelic science	 Kevin Breel: Confessions of a depressed comic	 Onora O'Neill: What we don't understand about trust
 Stuart Firestein: The pursuit of ignorance	 Elizabeth Loftus: The fiction of memory	 Benjamin Barber: Why mayors should rule the world		



To Do before Lecture 2

Run IPython example provided in class

■ Register for Data Scientist Workbench

- You can use Python on cloud via <https://datascientistworkbench.com>

■ Install Python

- Recommended to use Python version 3.X
- You may use your own Python distribution, e.g., Anaconda that can be downloaded from <https://www.continuum.io/downloads>

■ Get access to IBM Cloud

- Register at UofT software portal <http://uoft.onthehub.com> and register for access to IBM Cloud from IBM – Cloud – Cloud Access – IBM Cloud section

■ Check class web-page on Q