
Oleksandr Romanko, Ph.D.

Associate Director, Financial Risk Quantitative Research, SS&C Algorithmics
Adjunct Professor, University of Toronto

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

About me

Dr. Oleksandr Romanko

- Associate Director, Financial Risk Quantitative Research at SS&C Algorithmics, formerly with Watson Financial Services, IBM
- Ph.D. in Computer Science from McMaster University
- Author of over 25 papers and reports
- Adjunct professor at University of Toronto and American University Kyiv
- Research areas:
 - AI, 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





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

Billionaires

Innovation

Leadership

Money

Business

15,267 views | Jan 23, 2019, 12:10am

Data Scientist Leads 50 Best Jobs In America For 2019 According To Glassdoor



Louis Columbus Contributor
Enterprise Tech



Today In: Innovation

- Data Scientist is the best job in America for the 4th year in a row (4.7 job score, 4.3 job satisfaction rating) with 6,510 open positions paying a median base salary of \$108,000.

glassdoor



Job Title, Keywords, or Company

Featured, Jobs

The Best Jobs in America 2019

January 22, 2019 | Posted by Glassdoor Team

1. Data Scientist

- Job Score: 4.7
- Job Satisfaction Rating: 4.3
- Number of Job Openings: 6,510
- Median Base Salary: \$108,000
- [View Jobs](#)

2. Nursing Manager

- Job Score: 4.6
- Job Satisfaction Rating: 4.0
- Number of Job Openings: 13,931
- Median Base Salary: \$83,000
- [View Jobs](#)

3. Marketing Manager

- Job Score: 4.6
- Job Satisfaction Rating: 4.2

Best jobs

Forbes

4,534 views | Jan 15, 2020, 12:02am EST

Glassdoor's Best Jobs In 2020 Dominated By Tech



Louis Columbus Senior Contributor Enterprise & Cloud



GETTY IMAGES/ISTOCKPHOTO

- The top seven spots on Glassdoor's best jobs in 2020 are tech roles, including [Front End Engineer](#) (#1), [Java Developer](#) (#2), [Data Scientist](#) (#3), [Product Manager](#) (#4), [DevOps Engineer](#) (#5), [Data Engineer](#) (#6) and [Software Engineer](#) (#7).
- [Front End Engineer](#) is the best job in America, replacing [Data Scientist](#) which held the number one spot for the past four years.

glassdoor



Job Title, Keywords, or Company

Featured, Jobs

The Best Jobs in America 2020

January 14, 2020 | Posted by [Glassdoor Team](#)

Rank	Job Title	Median Base Salary	Job Satisfaction	Job Openings
1	Front End Engineer	\$105,240	3.9	13,122
2	Java Developer	\$83,589	3.9	16,136
3	Data Scientist	\$107,801	4.0	6,542
4	Product Manager	\$117,713	3.8	12,173
5	Devops Engineer	\$107,310	3.9	6,603
6	Data Engineer	\$102,472	3.9	6,941
7	Software Engineer	\$105,563	3.6	50,438
8	Speech Language Pathologist	\$71,867	3.8	29,167
9	Strategy Manager	\$133,067	4.3	3,515
10	Business Development Manager	\$78,480	4.0	6,560

Best jobs

Glassdoor's 50 Best Jobs In America For 2021 Dominated By Tech Careers



Louis Columbus Former Contributor +
Enterprise Tech



glassdoor Jobs Company Reviews Salaries Interviews Salary Calculator For Employers Post Jobs

Job Title, Keywords, or Company Jobs Location

50 Best Jobs in America for 2021

Best Jobs 2021 United States Share

Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1 Java Developer	\$90,830	4.2/5	10,103	View Jobs
#2 Data Scientist	\$113,736	4.1/5	5,971	View Jobs
#3 Product Manager	\$121,107	3.9/5	14,515	View Jobs
#4 Enterprise Architect	\$131,361	4.0/5	10,069	View Jobs
#5 Devops Engineer	\$110,003	4.0/5	6,904	View Jobs
#6 Information Security Engineer	\$110,000	4.0/5	5,621	View Jobs
#7 Business Development Manager	\$82,182	4.1/5	8,827	View Jobs
#8 Mobile Engineer	\$94,301	4.1/5	4,631	View Jobs

Best jobs

≡ Forbes Subscribe Sign In Search

LEADERSHIP • EDITORS' PICK

The 50 Best Jobs In America For 2022, According To Glassdoor

Jair Hilburn Forbes Staff
I am the assistant editor for Communities and Leadership.

Feb 2, 2022, 06:23pm EST

f  in

Young factory worker holding presentation about production development to company managers and his coworkers. GETTY

Two years into the coronavirus pandemic, the workplace has changed dramatically. With millions of Americans working from home and reevaluating what's most important to them, the jobs considered most desirable have changed—and may even come as a surprise.

glassdoor Search for job titles, companies, or keywords Location Search

Jobs Companies Salaries Careers For Employers Post Jobs

50 Best Jobs in America for 2022

	Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1	Enterprise Architect	\$144,997	4.1/5	14,021	View Jobs
#2	Full Stack Engineer	\$101,794	4.3/5	11,252	View Jobs
#3	Data Scientist	\$120,000	4.1/5	10,071	View Jobs
#4	Devops Engineer	\$120,095	4.2/5	8,548	View Jobs
#5	Strategy Manager	\$140,000	4.2/5	6,977	View Jobs
#6	Machine Learning Engineer	\$130,489	4.3/5	6,801	View Jobs
#7	Data Engineer	\$113,960	4.0/5	11,821	View Jobs
#8	Software Engineer	\$116,638	3.9/5	64,155	View Jobs
#9	Java Developer	\$107,099	4.1/5	10,201	View Jobs
#10	Product Manager	\$125,317	4.0/5	17,725	View Jobs
#11	Back End Engineer	\$112,384	4.2/5	6,221	View Jobs

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 Big Data & Analytics Hub

Blogs Videos & Podcasts Resources Events Around the Web IBM Solutions

Data Scientist: Closing the Talent Gap

January 17, 2013



THE MAGAZINE

October 2012

Data Scientist: The Sexiest Job of the 21st Century

Future of Jobs report 2023

Fastest growing vs. fastest declining jobs



Top 10 fastest growing jobs

1. AI and Machine Learning Specialists
2. Sustainability Specialists
3. Business Intelligence Analysts
4. Information Security Analysts
5. Fintech Engineers
6. Data Analysts and Scientists
7. Robotics Engineers
8. Big Data Specialists
9. Agricultural Equipment Operators
10. Digital Transformation Specialists

Top 10 fastest declining jobs

1. Bank Tellers and Related Clerks
2. Postal Service Clerks
3. Cashiers and ticket Clerks
4. Data Entry Clerks
5. Administrative and Executive Secretaries
6. Material-Recording and Stock-Keeping Clerks
7. Accounting, Bookkeeping and Payroll Clerks
8. Legislators and Officials
9. Statistical, Finance and Insurance Clerks
10. Door-To-Door Sales Workers, News and Street Vendors, and Related Workers

Source

World Economic Forum, Future of Jobs Report 2023,

Note

The jobs which survey respondents expect to grow most quickly from 2023 to 2027 as a fraction of present employment figures

Future of Jobs report 2023

Top 10 skills of 2023



- | | |
|---|---|
| 1. Analytical thinking | 6. Technological literacy |
| 2. Creative thinking | 7. Dependability and attention to detail |
| 3. Resilience, flexibility and agility | 8. Empathy and active listening |
| 4. Motivation and self-awareness | 9. Leadership and social influence |
| 5. Curiosity and lifelong learning | 10. Quality control |

Type of skill

Cognitive skills Self-efficacy Management skills Technology skills Working with others

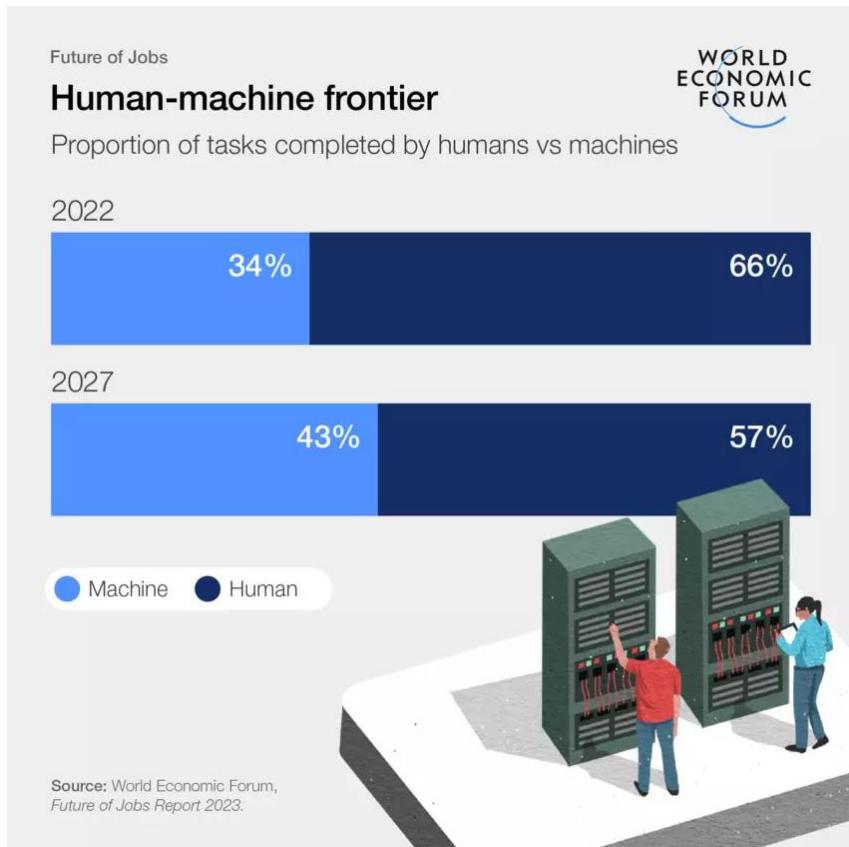
Source

World Economic Forum, Future of Jobs Report 2023.

Note

The skills judged to be of greatest importance to workers at the time of the survey

Future of Jobs report 2023

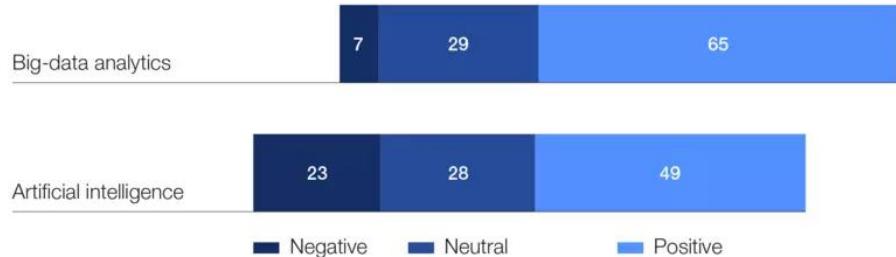


Future of Jobs

Businesses expect Big Data and AI to drive job growth

WORLD ECONOMIC FORUM

Expected impact of technologies on jobs:



Including jobs such as



AI and machine learning specialists,



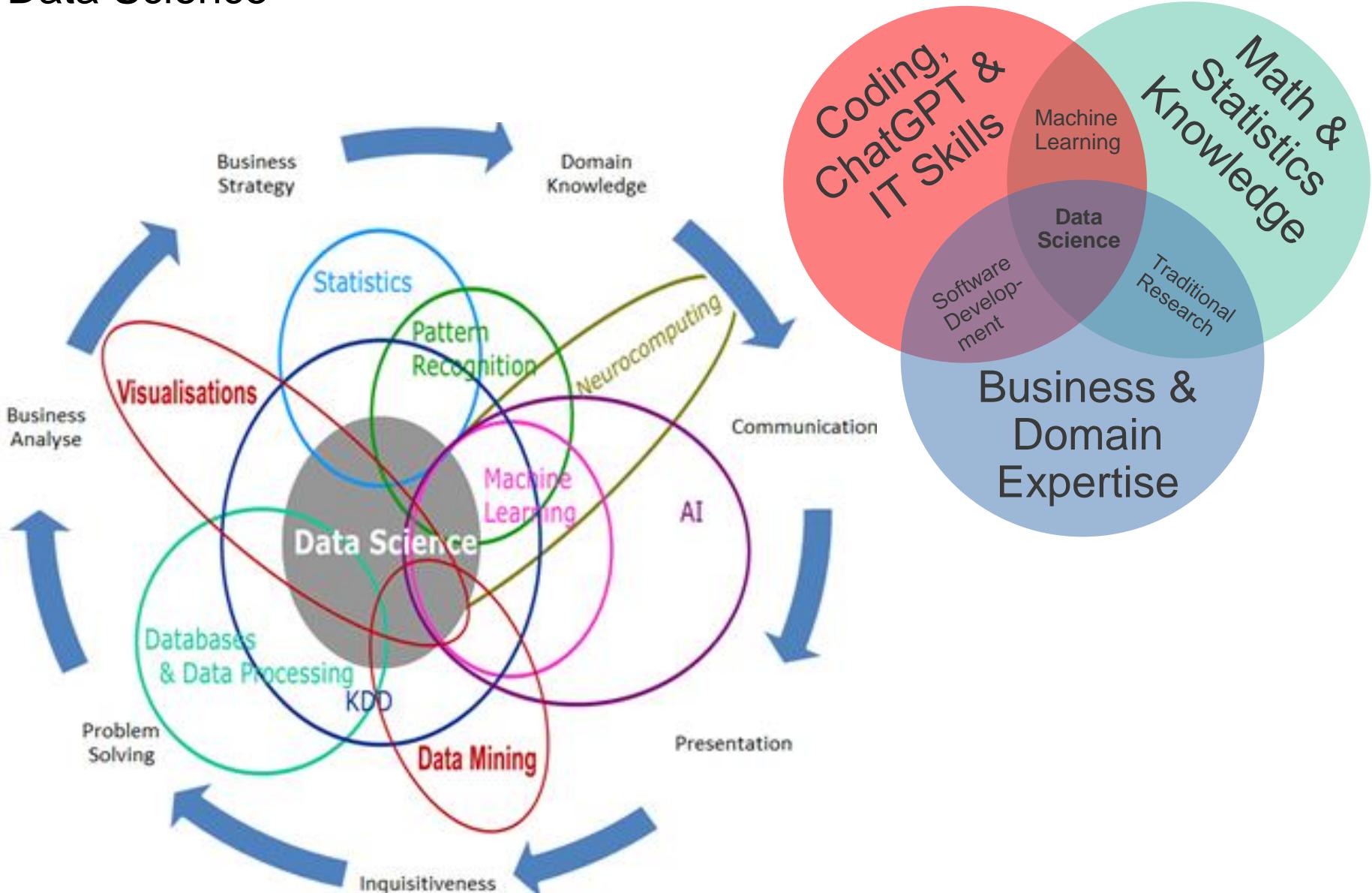
Data analysts and scientists, and



Big data specialists.

Source: Future of Jobs, World Economic Forum, April 2023.

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

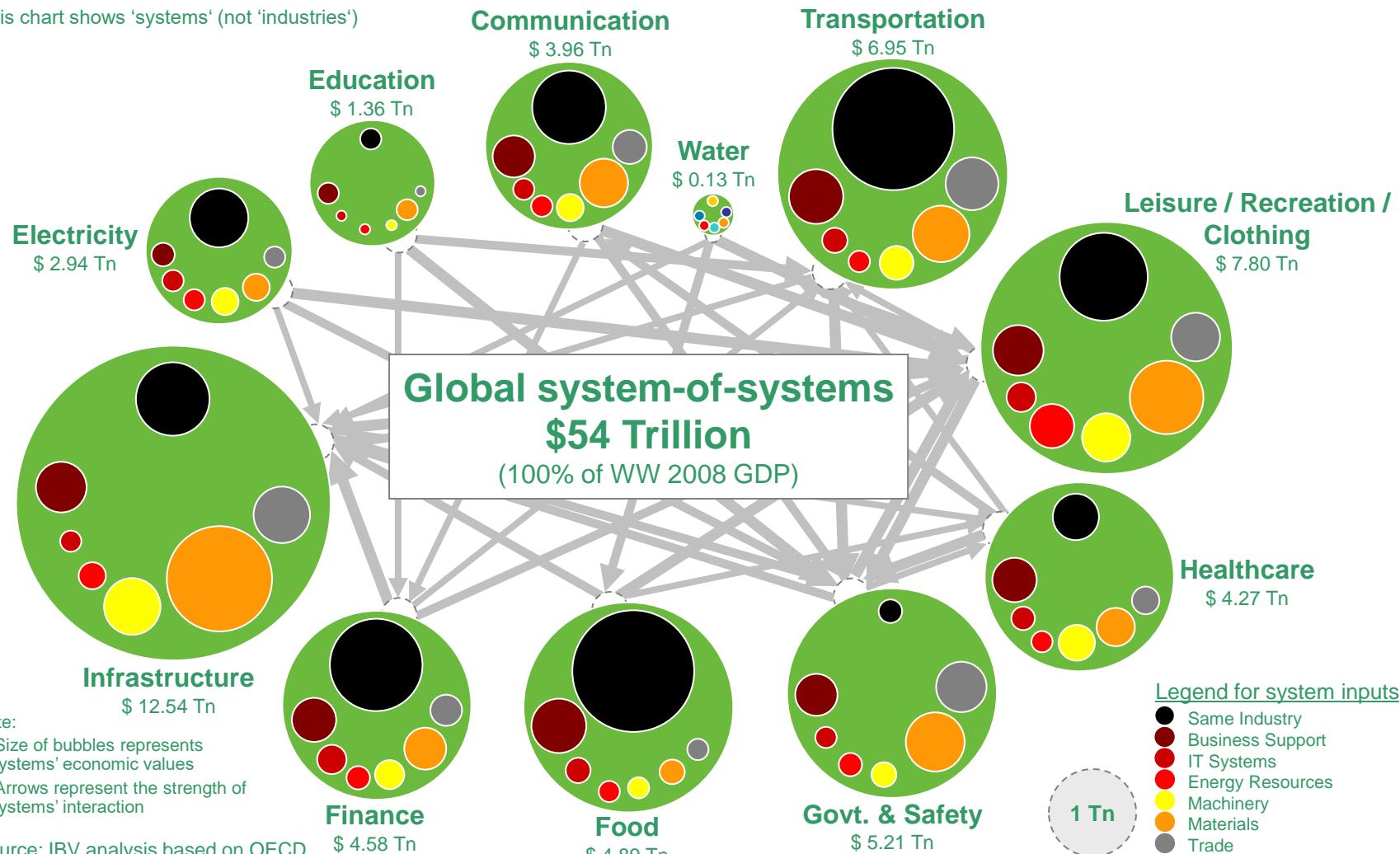
Operations research

- **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



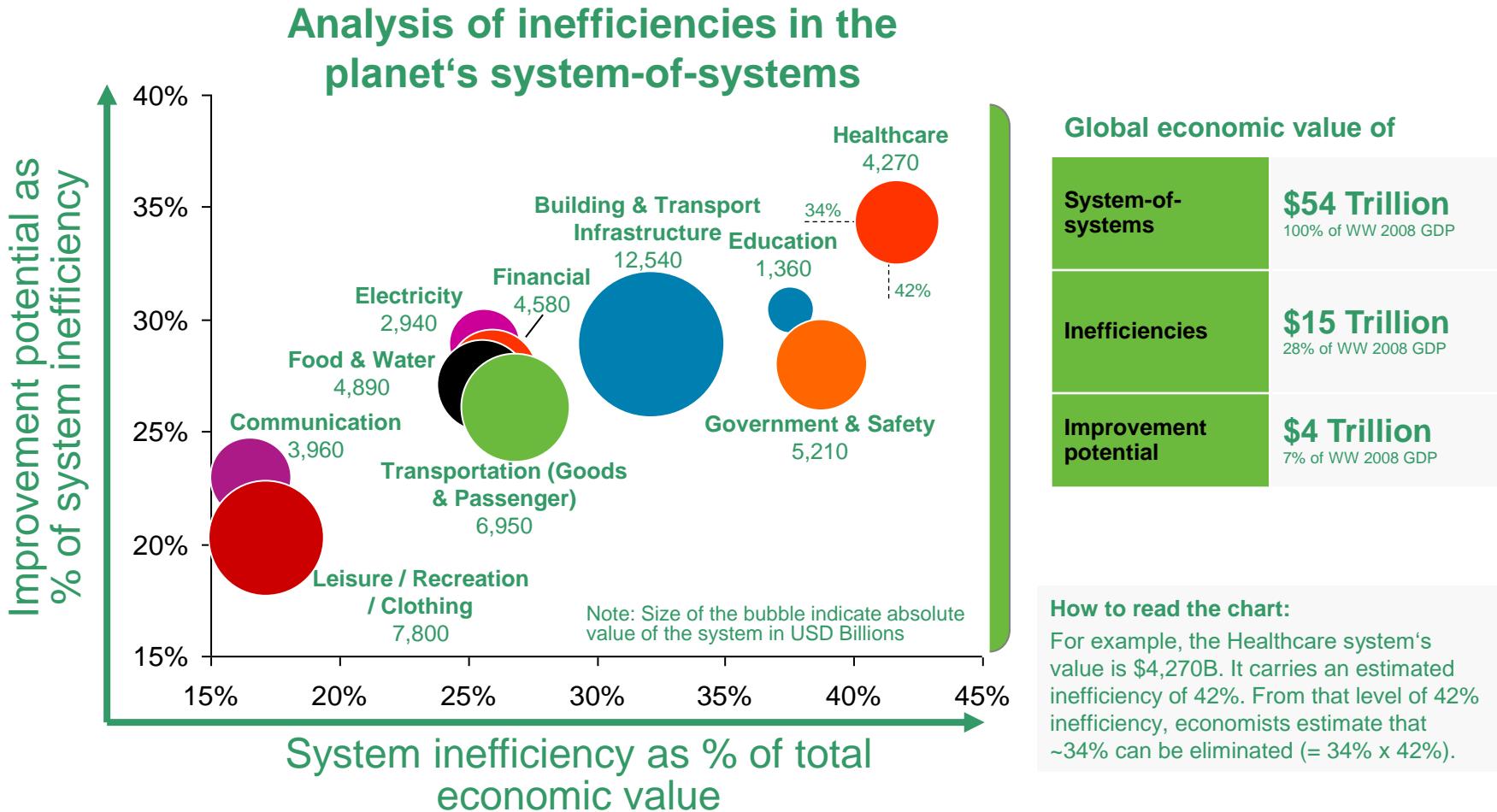
Our planet is a complex, dynamic, highly interconnected \$54 Trillion system-of-systems (OECD-based analysis)

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



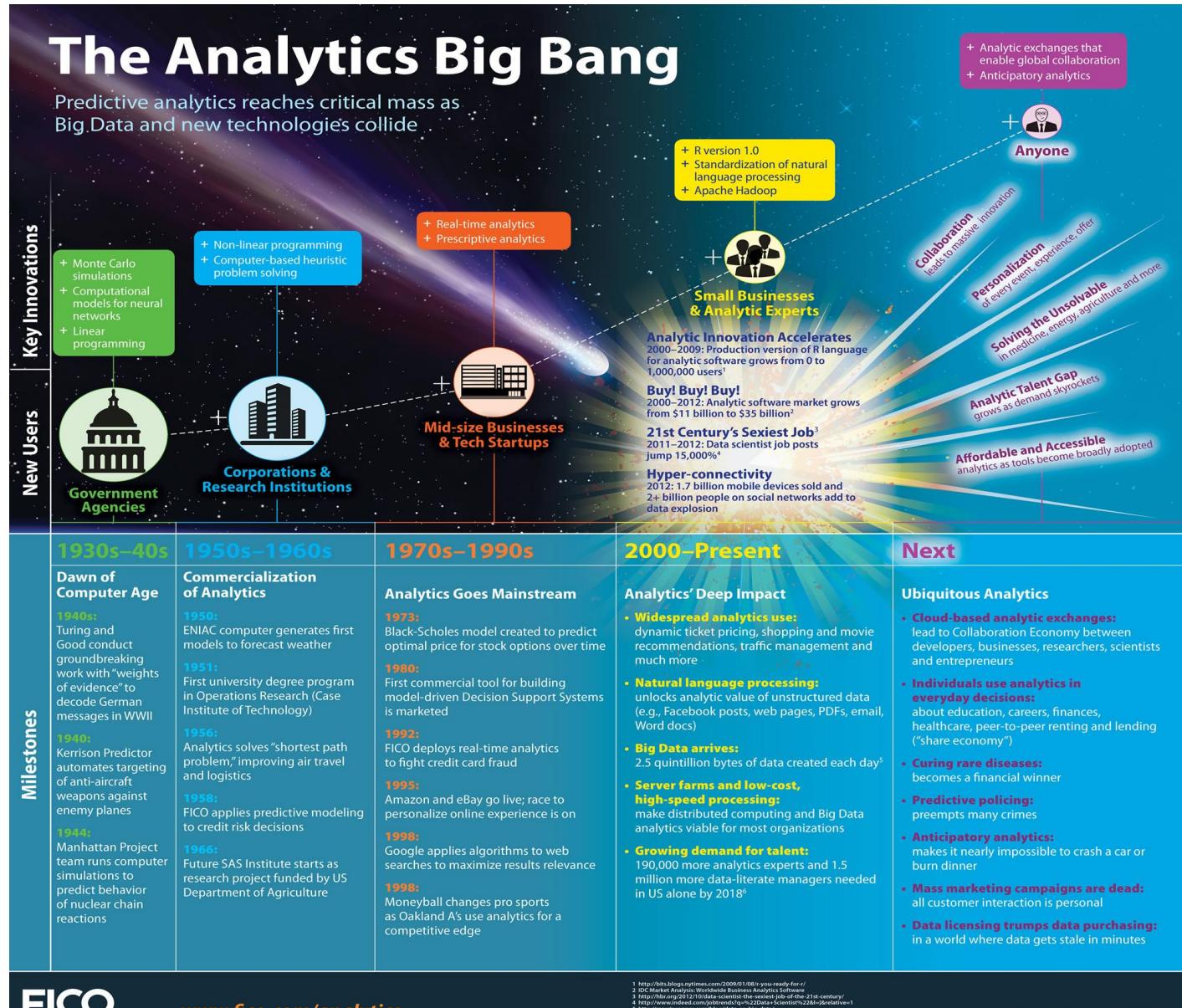
Economists estimate, that all systems carry inefficiencies of up to \$15 Tn, of which \$4 Tn could be eliminated

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



Source: IBM economists survey 2009; n= 480

History of analytics



History of analytics

Milestones	Key Innovations		
	New Users	Corporations & Research Institutions	Mid-size Businesses & Tech Startups
1930s–40s	Government Agencies <ul style="list-style-type: none">+ Monte Carlo simulations+ Computational models for neural networks+ Linear programming	Corporations & Research Institutions <ul style="list-style-type: none">+ Non-linear programming+ Computer-based heuristic problem solving	Mid-size Businesses & Tech Startups <ul style="list-style-type: none">+ Real-time analytics+ Prescriptive analytics
1950s–1960s	Commercialization of Analytics <ul style="list-style-type: none">1950: ENIAC computer generates first models to forecast weather1951: First university degree program in Operations Research (Case Institute of Technology)1956: Analytics solves “shortest path problem,” improving air travel and logistics1958: FICO applies predictive modeling to credit risk decisions1966: Future SAS Institute starts as research project funded by US Department of Agriculture	Analytics Goes Mainstream <ul style="list-style-type: none">1973: Black-Scholes model created to predict optimal price for stock options over time1980: First commercial tool for building model-driven Decision Support Systems is marketed1992: FICO deploys real-time analytics to fight credit card fraud1995: Amazon and eBay go live; race to personalize online experience is on1998: Google applies algorithms to web searches to maximize results relevance1998: Moneyball changes pro sports as Oakland A's use analytics for a competitive edge	Analytic Innovation Accelerates <ul style="list-style-type: none">2000–2009: Production version of R language for analytic software grows from 0 to 1,000,000 users¹Buy! Buy! Buy!<ul style="list-style-type: none">2000–2012: Analytic software market grows from \$1 billion to \$35 billion²21st Century's Sexiest Job³<ul style="list-style-type: none">2011–2012: Data scientist job posts jump 15,000%⁴Hyper-connectivity<ul style="list-style-type: none">2012: 1.7 billion mobile devices sold and 2+ billion people on social networks add to data explosion
1970s–1990s			2000–Present <ul style="list-style-type: none">Analytics' Deep Impact<ul style="list-style-type: none">Widespread analytics use: dynamic ticket pricing, shopping and movie recommendations, traffic management and much moreNatural 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 day⁵Server farms and low-cost, high-speed processing: make distributed computing and Big Data analytics viable for most organizationsGrowing demand for talent: 190,000 more analytics experts and 1.5 million more data-literate managers needed in US alone by 2018⁶Next<ul style="list-style-type: none">Ubiquitous Analytics<ul style="list-style-type: none">Cloud-based analytic exchanges: lead to Collaboration Economy between developers, businesses, researchers, scientists and entrepreneursIndividuals use analytics in everyday decisions: about education, careers, finances, healthcare, peer-to-peer renting and lending (“share economy”)Curing rare diseases: becomes a financial winnerPredictive policing: preempts many crimesAnticipatory analytics: makes it nearly impossible to crash a car or burn dinnerMass marketing campaigns are dead: all customer interaction is personalData licensing trumps data purchasing: in a world where data gets stale in minutes

Visual history of AI

1943 Electronic Brain by McCulloch & Pitts	1950 Turing Test by Alan Turing	1952 Checkers Program by Arthur Samuel	1956 Symbolic AI: The first automated reasoning program.	1956 Dartmouth Summer Research Project organized by John McCarthy	1957 Perceptron by Frank Rosenblat	1958 LISP by John McCarthy	1959 General Problem Solver by Newell, Simon, Shaw	1960 ADALINE by Widrow & Hoff	1964 ELIZA by Joseph Weizenbaum
First mathematical model of a neuron. $\Phi(x) = \begin{cases} 1 & \text{if } x \geq \theta \\ -1 & \text{otherwise} \end{cases}$	Turing test is proposed.	One of the first computer board game.	Logic Theorist by Newell, Simon, Shaw	The conference that started it all.	Weights automatically learned.	One of the first AI programming languages. (PRINT (*EVALQT X Y)) (GO A)) (SETQ Y (READ))	Reasoning as search.	The first cost function. $E(w) = \frac{1}{2} \sum_i (y_i - \Phi(z_i))^2$	An early natural language processing conversation program.
1964 Universal Bayesian Methods by Ray Solomonoff	1964 Solving algebra word problem, an early NLP accomplishment. $P_M(x) = \sum_{i=1}^{\infty} 2^{- s_i(x) }$	1964 Fuzzy Logic by Lotfi Zadeh	1965 The term Fuzzy Logic is introduced. Cold — Warm	1965 DENDRAL by Feigenbaum, Buchanan, Lederberg, Djerassi	1966 SRI International	1966 ALPAC report shuts down NLP research.	1968 SHRDLU by Terry Winograd	1969 XOR Problem by Minsky & Papert	1970 Automatic differentiation by Seppo Linnainmaa
Algorithmic probability & formal induction is introduced.	STUDENT by Daniel G. Bobrow	Fuzzy Logic by Lotfi Zadeh	One of the first Expert System: organic chemistry.	1966 Shakey the Robot by SRI International	1966 ALPAC Report by ALPAC	An early natural language understanding computer program, in the world of blocks.	Perceptron cannot learn XOR.	1970 INTERNIST-I by Myers, Miller, Poplie	Backpropagation & automatic differentiation. $\frac{\partial y}{\partial z} = \frac{\partial y}{\partial w_1} \frac{\partial w_1}{\partial x}$
1972 PROLOG by Colmerauer & Kowalski	1973 Lighthill Report by James Lighthill	1974 MYCIN by Shortliffe, Buchanan, Cohen	1979 Neocognitron by Kunihiko Fukushima	1982 Bayesian Networks by Judea Pearl	1983 SOAR by Laird, Newell, Rosenbloom	1986 Backpropagation in MLP by Rumelhart, Hinton, Williams	1987 Subsumption by Rodney Brooks	1989 LeNet by Yann LeCun	1992 TD-Gammon by Gerald Tesauro
One of the first logic programming languages. <code>animal(X) :- cat(X).</code>	UK ended support for AI research.	1974 RULE-BASED EXPERT SYSTEMS	An early backward chaining expert system for medical diagnosis.	Foundation of graphical probabilistic models.	A cognitive architecture for general intelligence.	Backpropagation is popularized.	A reactive robotic architecture.	Convolutional neural networks (CNN) used for recognizing handwriting.	Almost champion-level backgammon, using reinforcement learning.
1994 Chinook by a Team led by Jonathan Schaeffer	1995 Support Vector Machines by S. Vapnik & Cortes	1995 NIST	1995 MNIST is born.	1996 DeepBlue by IBM	1997 Long Short-Term Memory (LSTM) by Hochreiter & Schmidhuber	2006 Deep Boltzmann Machine by Salakhutdinov & Hinton	2009 ImageNet by Fei-Fei Li	2012 AlexNet by Krizhevsky & Hinton	2013 Word2Vec by Tomas Mikolov
Chinook, draughts player: the first program to win the world champion title against humans.	Soft-margin SVM is introduced.	NIST	MNIST is born.	DeepBlue beats Kasparov in chess.	LSTM for addressing vanishing gradients.	Modern deep learning is born.	ImageNet, a large-scale image dataset is introduced.	First significant results in deep learning.	Generative Adversarial Network (GAN) introduced.
2016 AlphaGo by DeepMind	2017 Google	2018 Ethical AI takes off: accuracy disparities in commercial gender classification	2019 Grandmaster level in Starcraft II.	2020 AI Activism by Ethical AI Community	2021 AlphaFold by DeepMind	2022 OpenAI	2023 Microsoft	2023 VALL-E	AI Winter
The first program to beat a professional Go player. 	Transformer network is introduced, paving the way for pretrained language models.	Gender Shades by Buolamwini & Gebru	AlphaStar by DeepMind	IBM, Amazon, and Microsoft ban sale of facial recognition products to law enforcement.	Predicting a protein's structure from amino-acid sequence	DALL-E 2, ChatGPT	Zero-Shot Text to Speech Synthesis	VALL-E	Events Leading to AI Winter

Parisa Rashidi, January 2023. CC BY 4.0



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 Tableau, Power BI 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
- Visualizations in Python

Simulation modeling

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

Optimization

- Unconstrained non-linear optimization algorithms
- Overview of constrained optimization algorithms
- Optimization case studies in IPython

Course outline

Advanced machine learning

- Decision trees
- Advanced supervised machine learning algorithms (Naive Bayes, k-NN, SVM)
- Intro to ensemble learning algorithms (Random Forests, Gradient Boosting)
- Intro to neural networks
- Text analytics and natural language processing
- Generative AI
- Clustering (K-means, Fuzzy C-means, Hierarchical Clustering, DBSCAN)
- Dimensionality reduction
- Association rules
- Overview of reinforcement learning
- Machine learning case studies in IPython

Introduction to Deep Learning

- Mathematics of neural networks
- Introduction to Deep Learning
- Convolutional Neural Networks (CNN)

Assignments, exams and grading (tentative)

Assignment #1 – Solving an analytics problem in Python (12%)

- Individual assignment.

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

- Individual assignment.

Assignment #3 – Solving an analytics problem in Python (15%)

- Individual assignment.

Course Project – Business consulting project using machine learning and data analytics in Python (20%)

- Group project (groups of 7 students), the same groups as for In-Class Presentations.

In-Class Group Presentation (12%)

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

Final Exam (26%)

- 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.

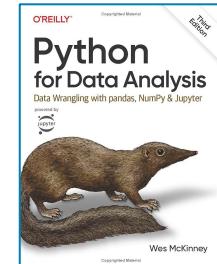
Notes

- 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 11 pts (Assg 1) + 14 pts (Assg 2) + 14 pts (Assg 3) + 19 pts (Course Project) + 11 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 26 pts at the Final Exam (less than 50%), the course mark will be reduced from A- to B+.

Course materials and readings

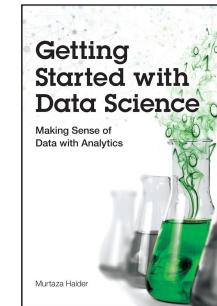
Required

- ❑ Course slides by O. Romanko, 2023
[Quercus](#)

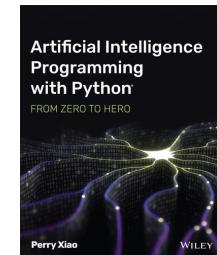


Optional

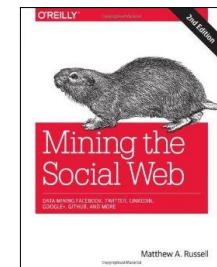
- ❑ Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter
by W. McKinney, 2022
<https://www.amazon.ca/Python-Data-Analysis-Wrangling-Jupyter/dp/109810403X/>



- ❑ Getting Started with Data Science: Making Sense of Data with Analytics
by M. Haider, 2015
<https://www.amazon.ca/Getting-Started-Data-Science-Analytics/dp/0133991024/>



- ❑ Artificial Intelligence Programming with Python by P. Xiao, 2022
<https://www.amazon.ca/Artificial-Intelligence-Programming-Python-Zero/dp/1119820863/>

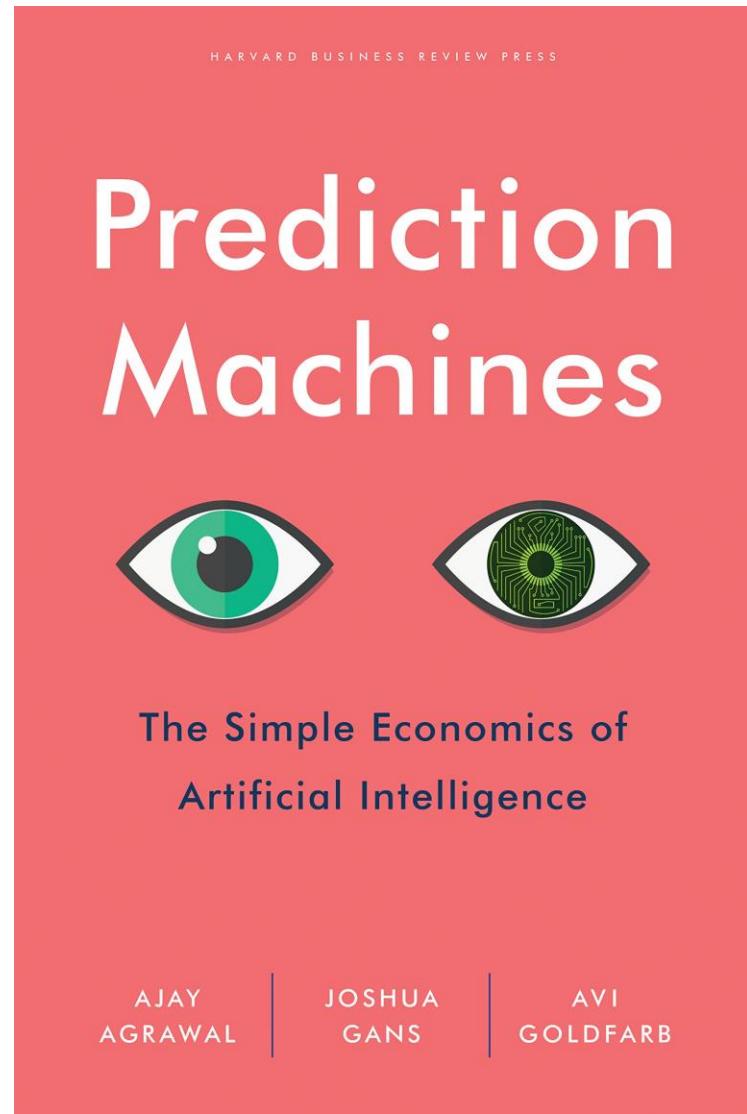
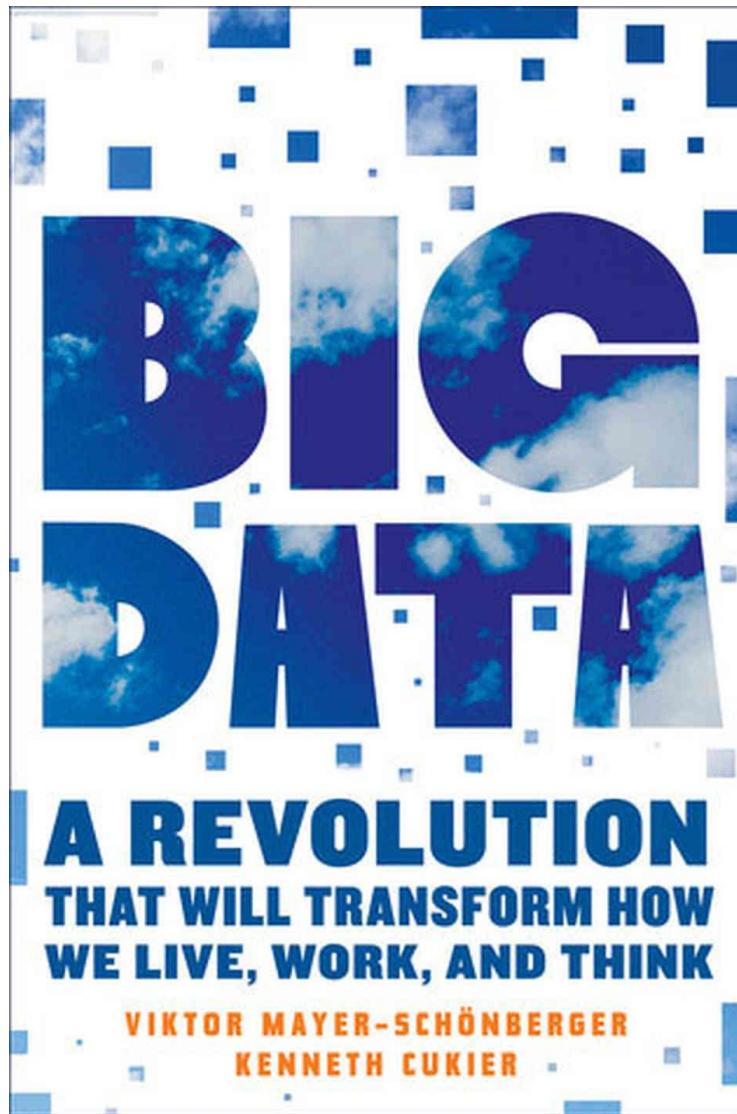


- ❑ Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More by M. Russell and M. Klassen, 2019
<https://www.amazon.ca/Mining-Social-Web-Facebook-Instagram/dp/1491985046/>

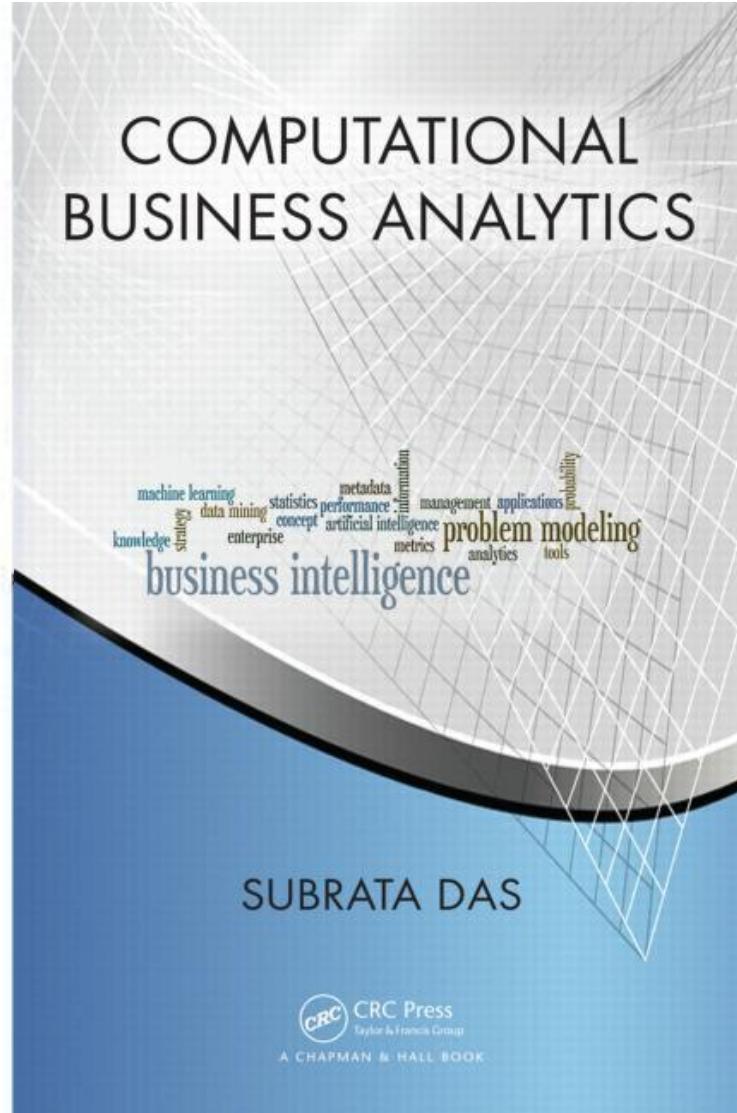
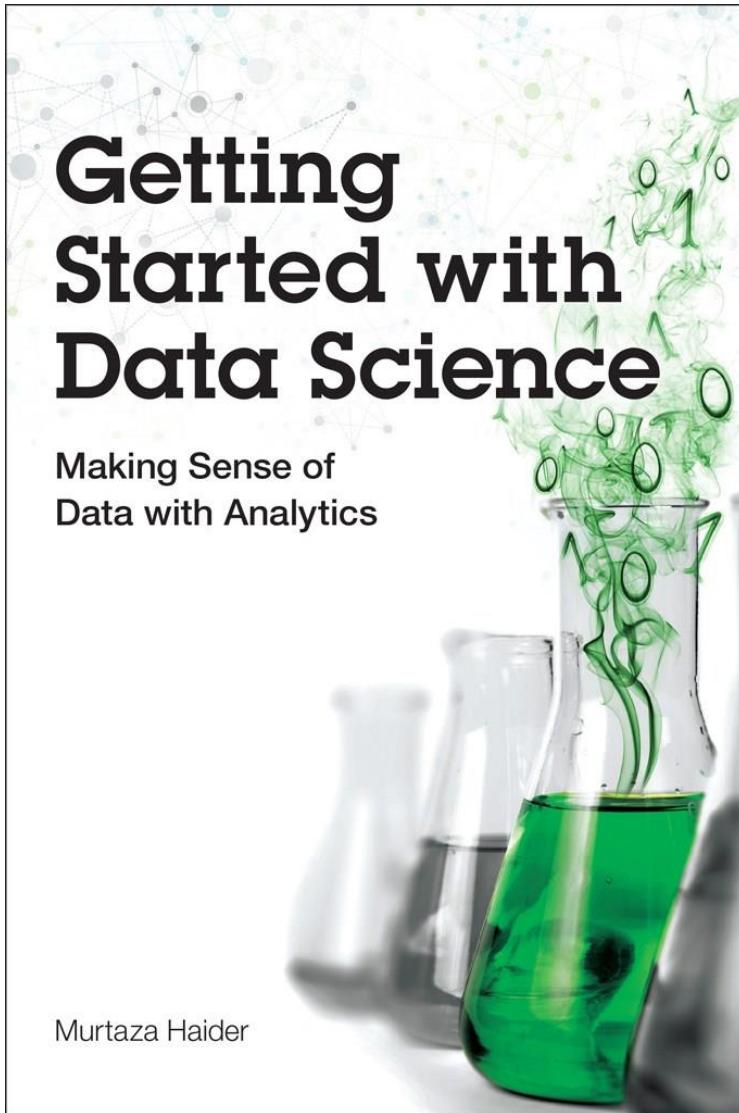


Recommended Literature

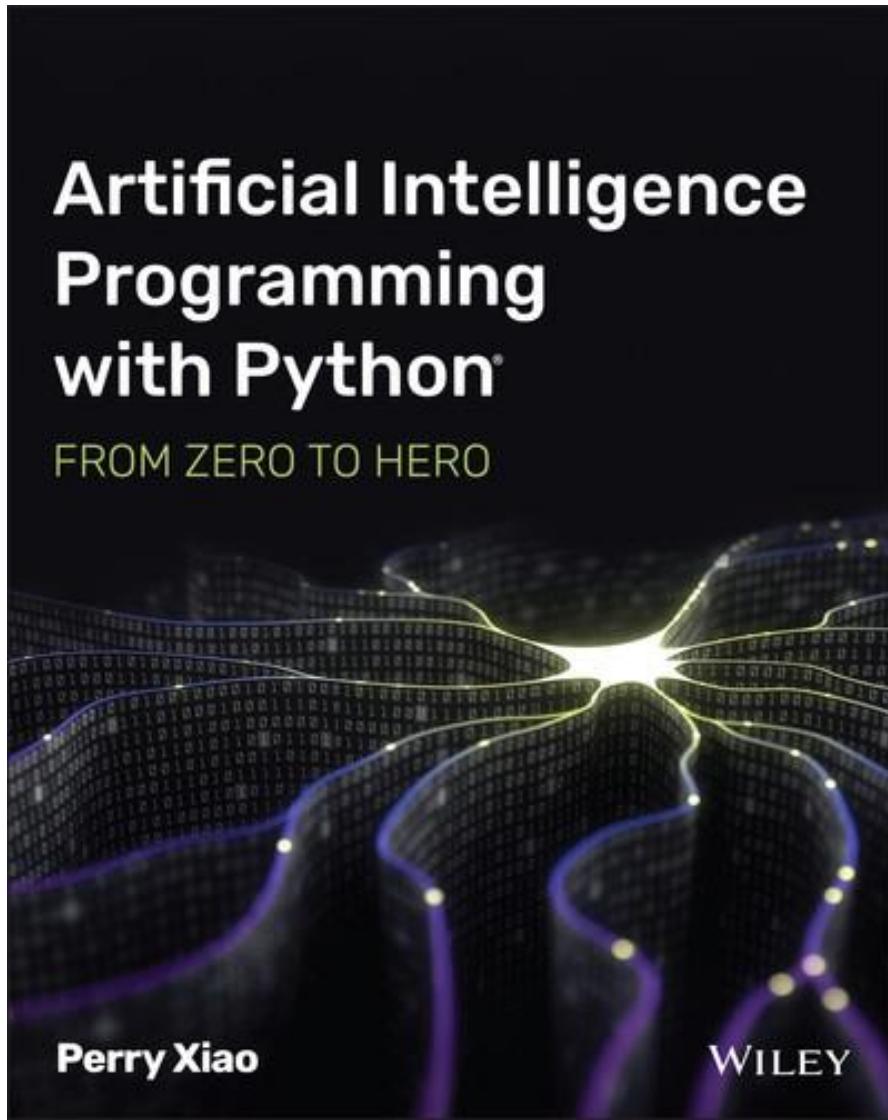
Literature



Literature



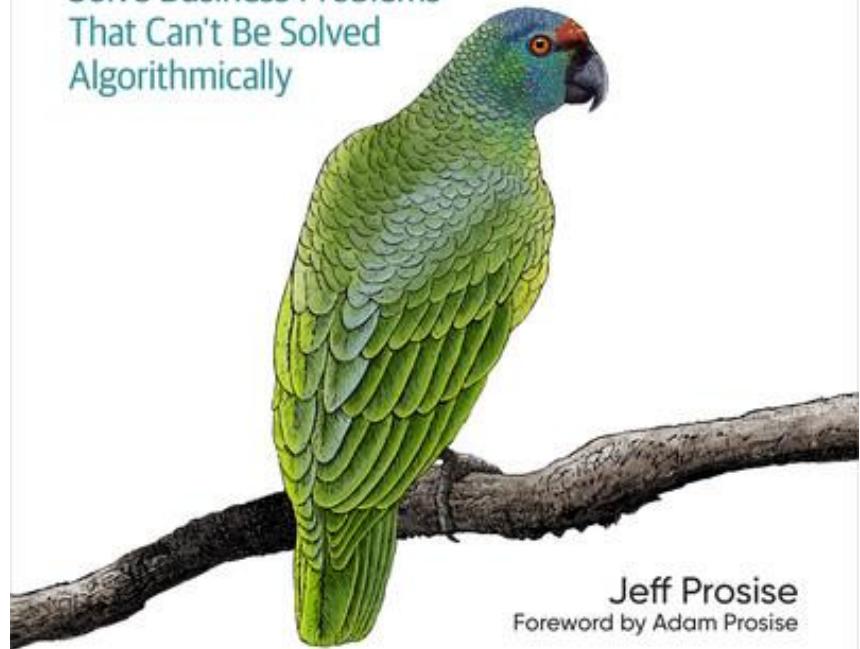
Literature



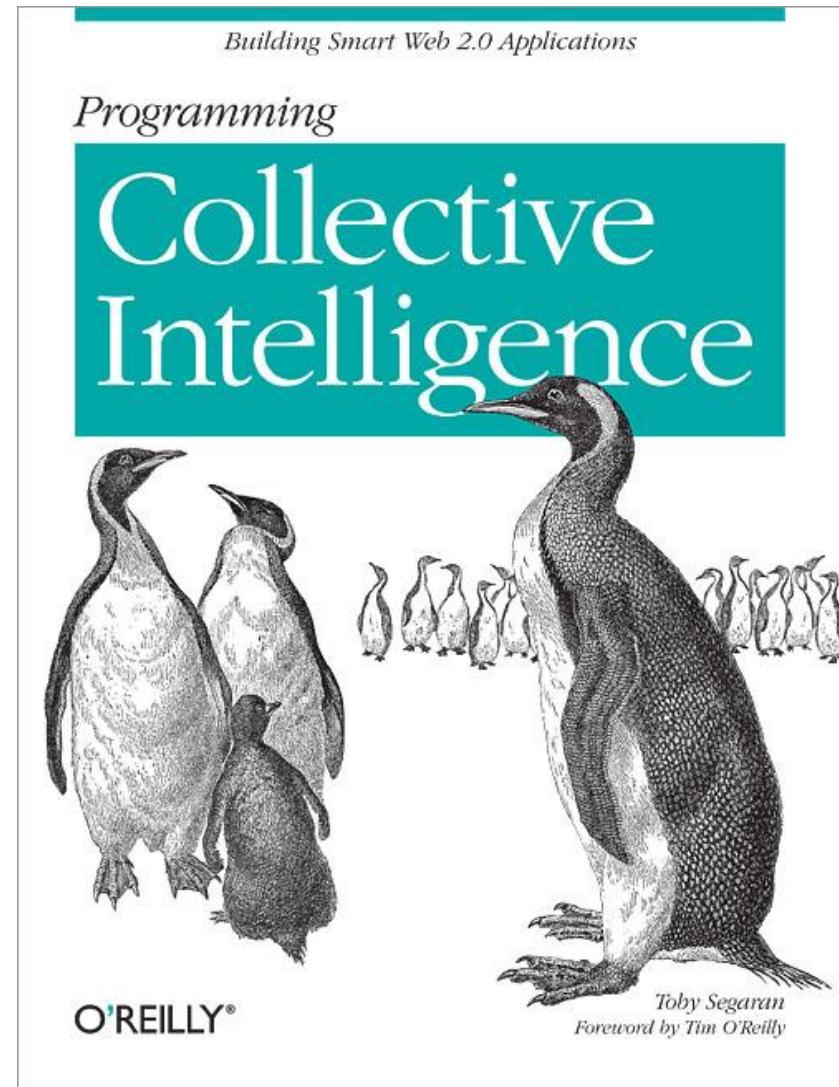
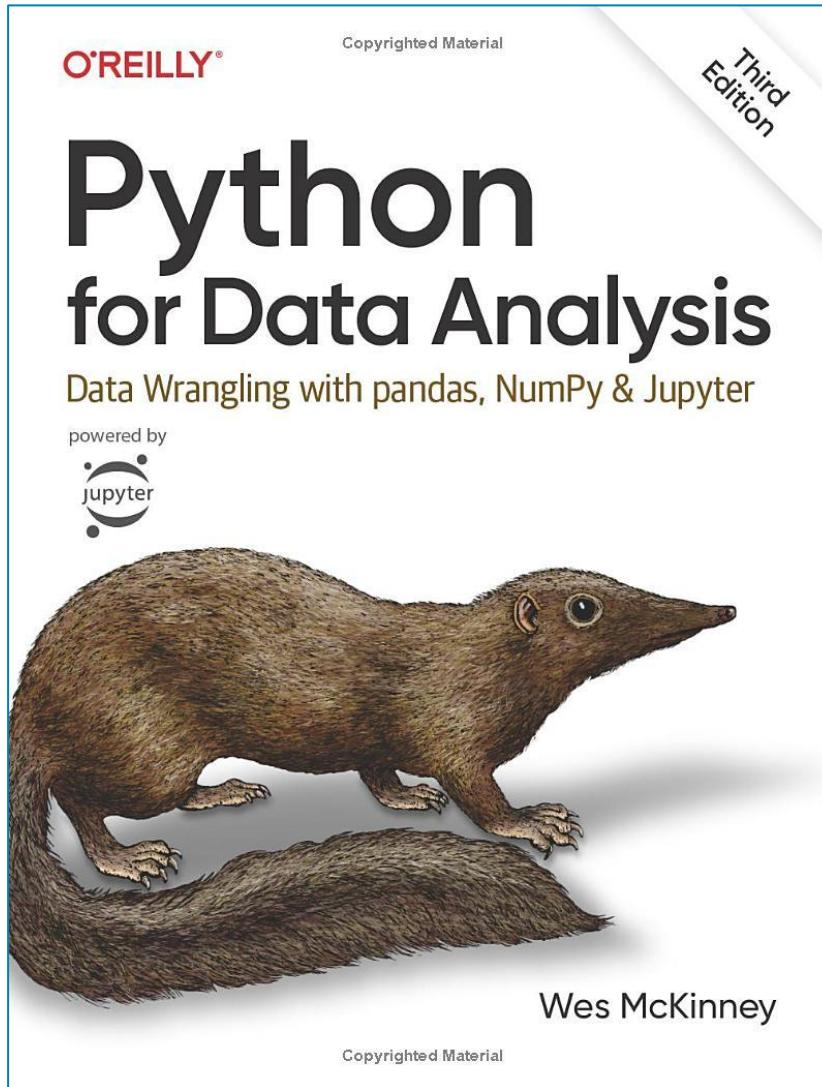
O'REILLY®

Applied Machine Learning and AI for Engineers

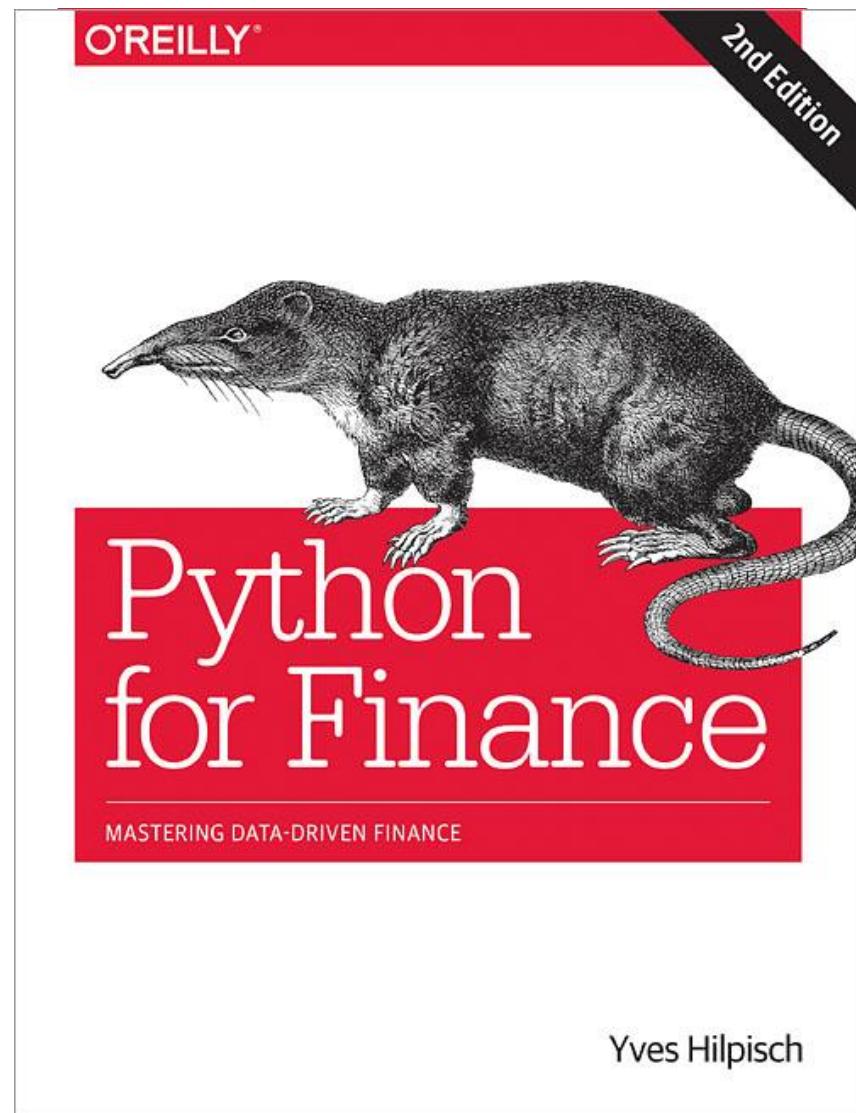
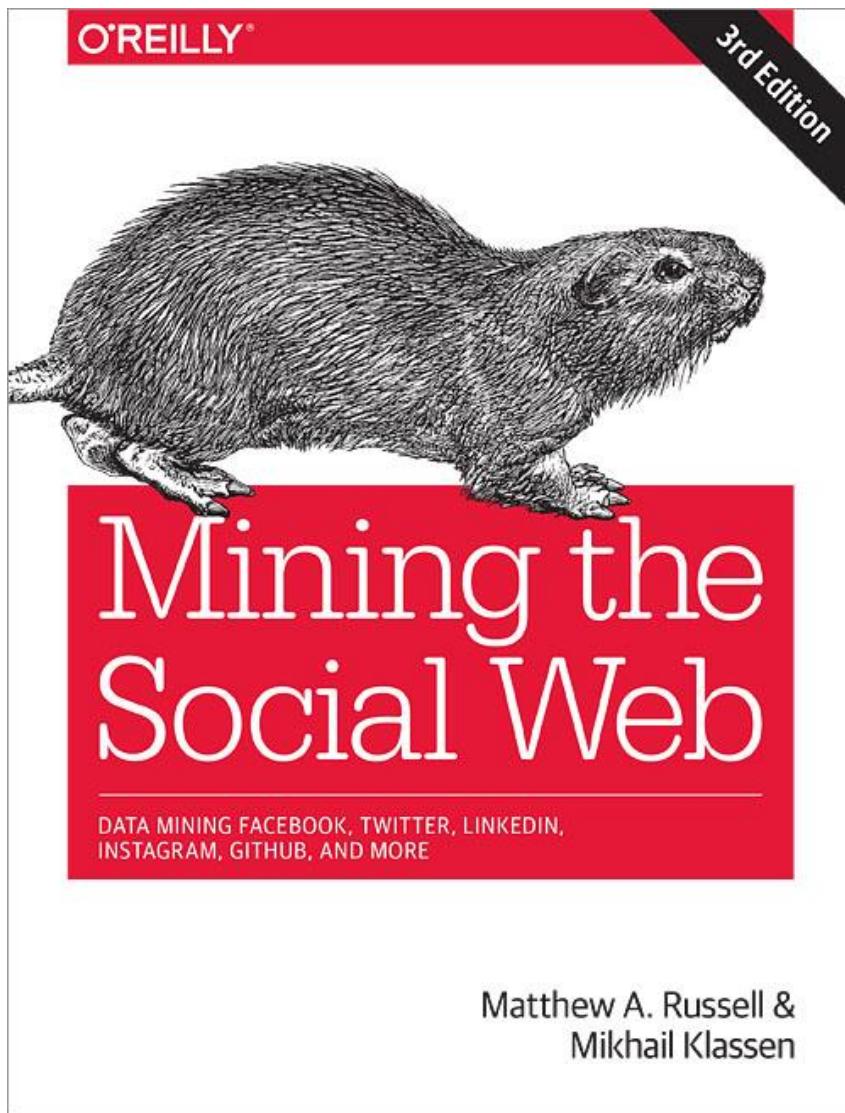
Solve Business Problems
That Can't Be Solved
Algorithmically



Literature



Literature





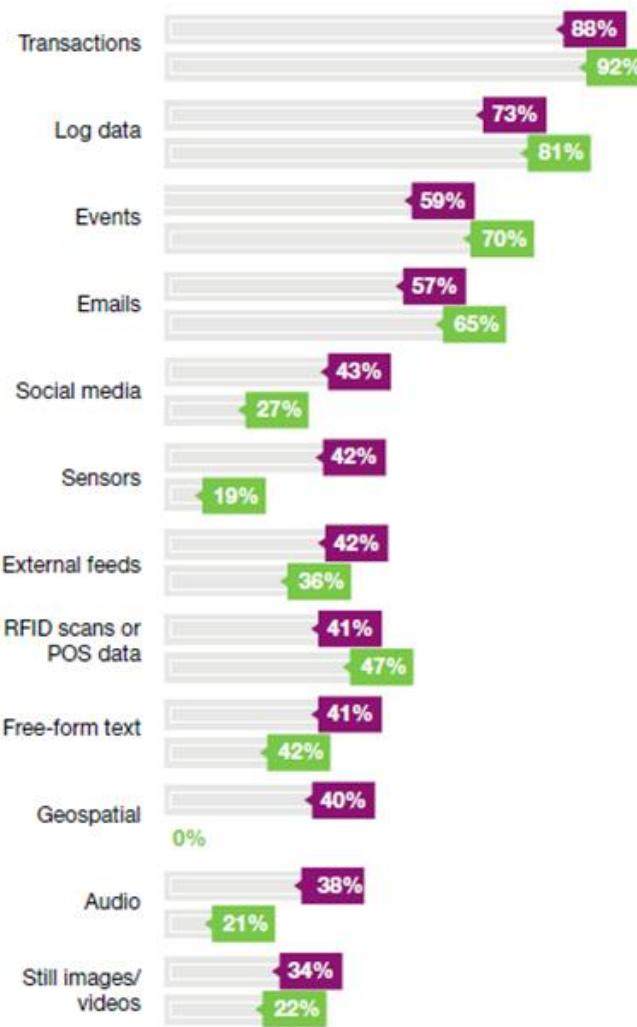
Sources of Data

Data sources

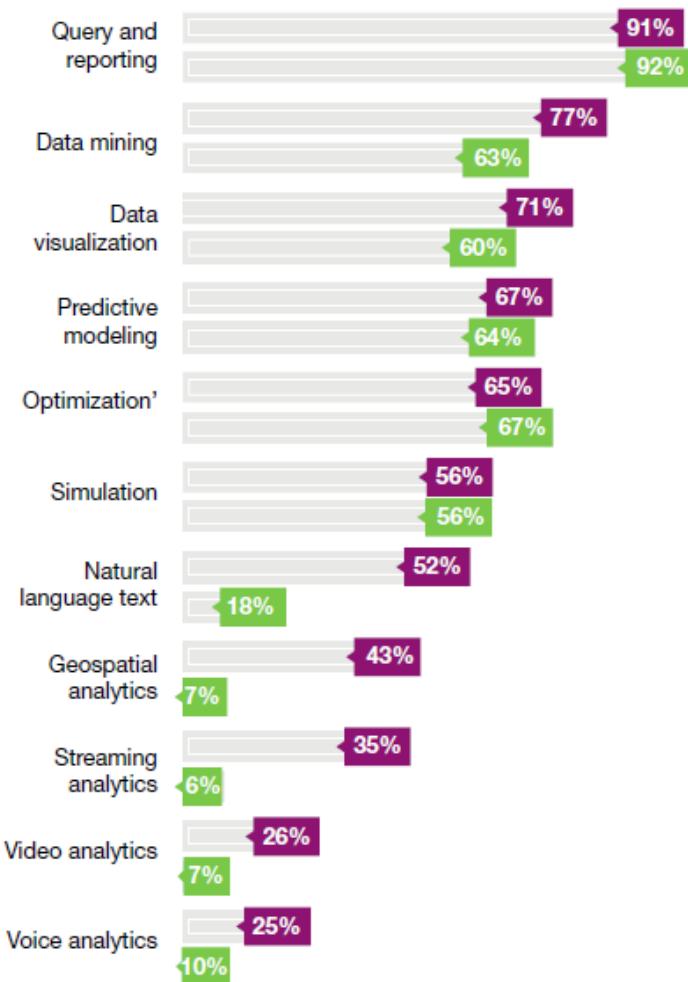
- **Data files** – demo in Python
 - ✓ CSV (comma separated value) files
 - ✓ Spreadsheet files, e.g., Excel or Google Spreadsheet
- **Databases**
 - ✓ SQL
 - ✓ NoSQL
- **Internet** – demo in Python
 - ✓ Web scraping
 - ✓ APIs
- **Big Data platforms and Cloud**
 - ✓ Hadoop
 - ✓ Cloud (AWS, Google Cloud, Microsoft Azure, IBM Cloud)

Use of 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

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



We can collect information from almost everything to make better decisions

50 billion

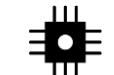
RFID tags
embedded into our
world and across
entire ecosystems

3 billion

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

95%

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





Analytics Examples

Data reveals hidden city dynamics



Data and Signals Reveal Hidden City Dynamics

Applications of big data analytics

Smarter Healthcare



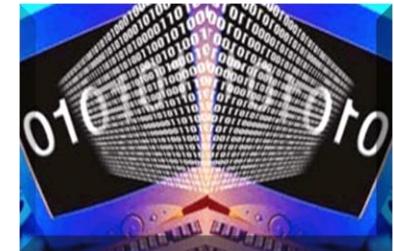
Multi-channel



Finance



Log Analysis



Homeland Security



Traffic Control



Telecom



Search Quality



Manufacturing



Trading Analytics



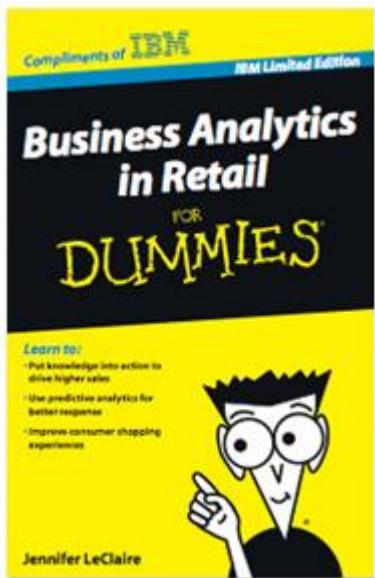
Fraud and Risk



Retail: Churn, NBO



Marketing analytics



IBM Predictive Analytics in RETAIL

What will your customers want next?



Fitting room analytics

Good



Mango



SELA



New Yorker



Bad



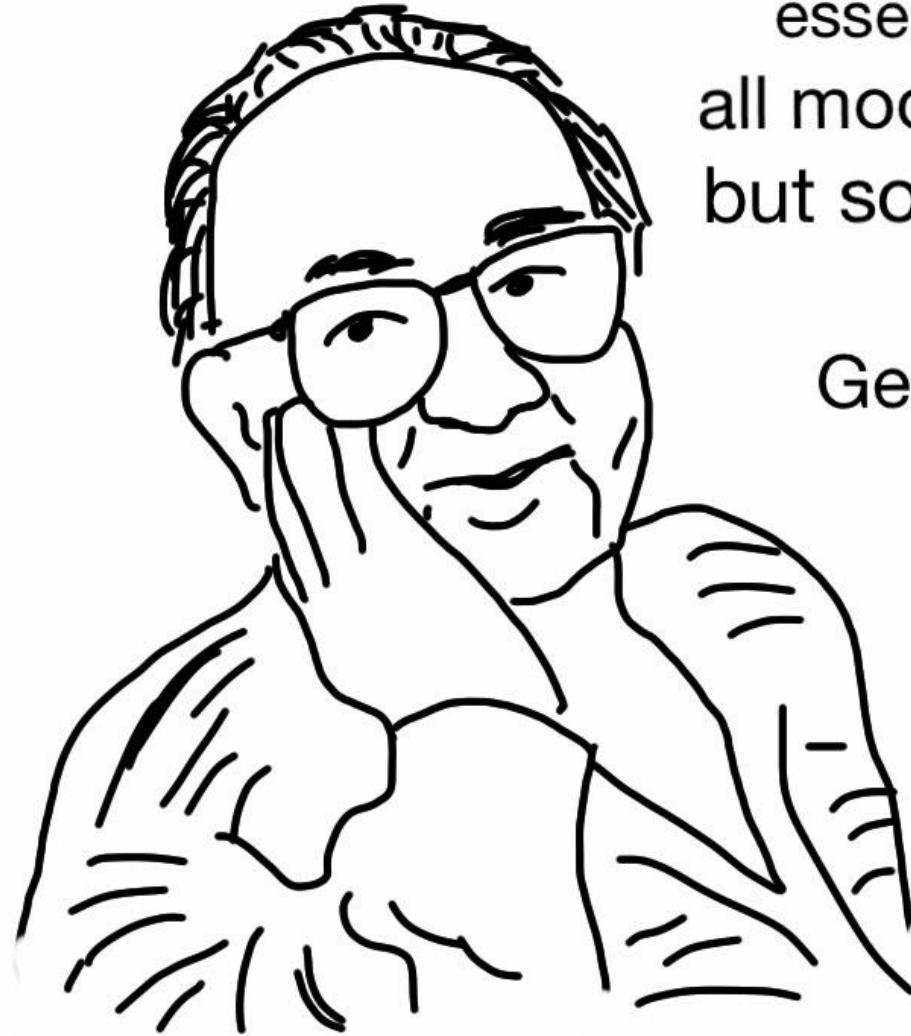


Modeling

Questions that we can try to answer with models

- **Statistics** – exploratory analysis and hypothesis testing
 - ✓ Decisions made from samples
 - ✓ Hypothesis testing
- **Machine learning** – learning from examples
 - ✓ Supervised learning (prediction, classification)
 - ✓ Unsupervised learning (clustering, dimensionality reduction, associations)
- **Artificial intelligence** – advanced analytics
 - ✓ Text analytics, social media analytics, NLP
 - ✓ Spatio-temporal analytics
 - ✓ Image and visual recognition (deep learning)
 - ✓ Reinforcement learning and autonomous systems
- **Modeling uncertainty** – what would happen in the future?
 - ✓ Monte Carlo simulations
- **Optimizing decisions** – what's best?
 - ✓ Optimization
- **Finding connections** – is FB related to Cambridge Analytica?
 - ✓ Graph/network models

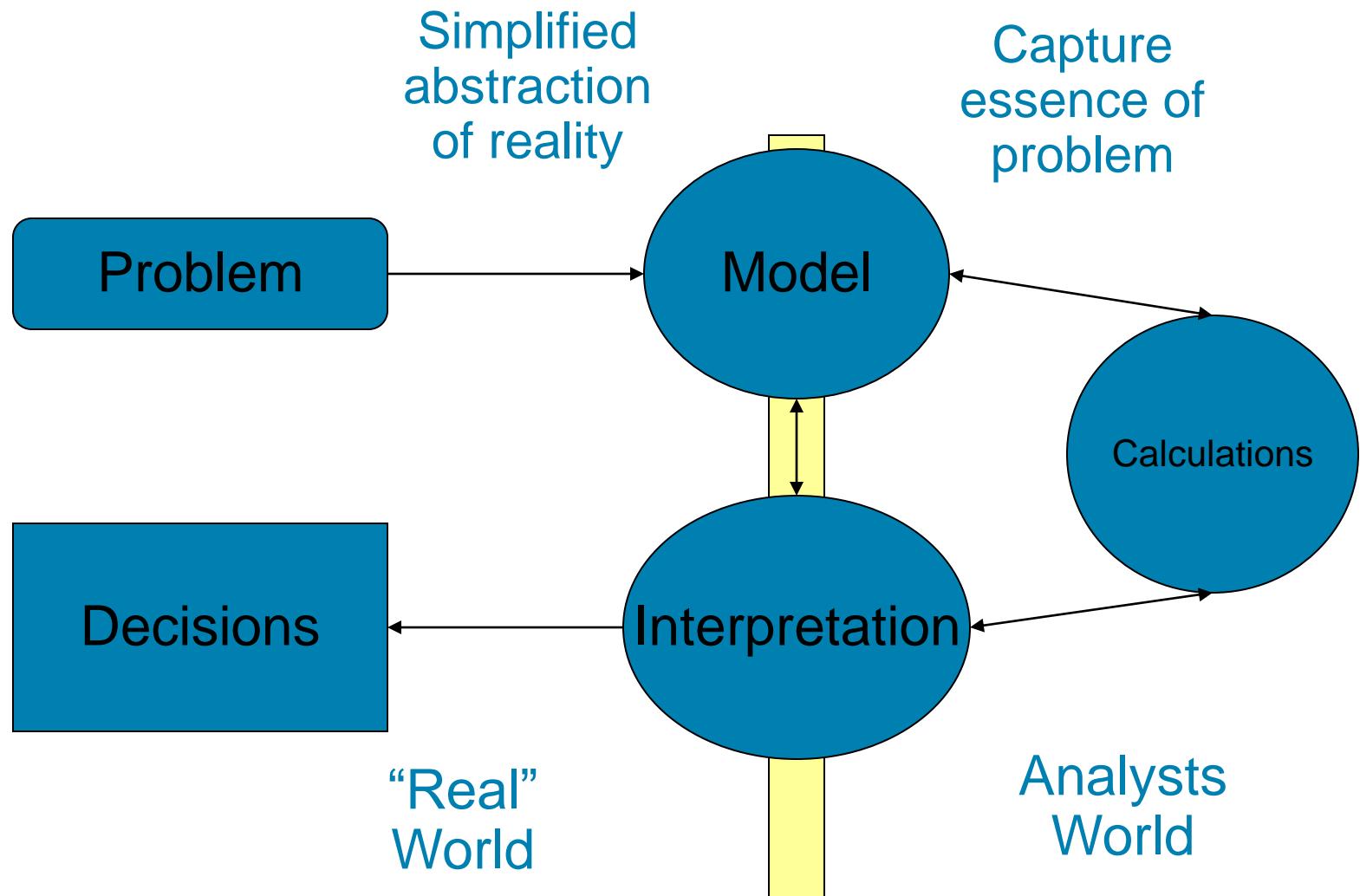
Models



essentially,
all models are wrong,
but some are useful

George E. P. Box

Models and reality

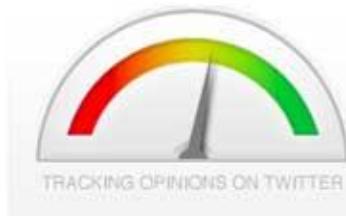




Artificial Intelligence

Text analytics and sentiment analysis

 ShareThis



twitrratr

SEARCH

SEARCHED TERM	POSITIVE TWEETS	NEUTRAL TWEETS	NEGATIVE TWEETS	TOTAL TWEETS
starbucks	708	4495	234	5437

13.02% POSITIVE

-  k i feel dumb.... apparently i was meant to 'dm' for the starbucks competition! i guess its late :/ i would have won too! ([view](#))
-  sleep so i can do a ton of darkroom tomorrow i have to resist the starbucks though if i want enough money for the bus ([view](#))
-  honestly, i enjoyed the best buy/starbucks one. we should do that again. or have it here at my house tomorrow night. we can watch dr. horibl ([view](#))
-  i'm at starbucks coffee (1978 contra costa blvd, pleasant hill, ca 94523, usa) - <http://bkite.com/01yvd> ([view](#))
-  i found an excellent deal on an apartment.and it's walking distance from a starbucks i could transfer to. did i mention free rent

82.67% NEUTRAL

-  I like how that girl @ starbucks tonight let me stand in line for 10 mins w/ another dude in front of me, before saying "oh. I'm closed.." ([view](#))
-  Tweets on 2008-10-23: Sitting in Starbucks, drinking Verona, and writing a sermon about the pure in heart.. <http://tinyurl.com/57zx2d> ([view](#))
-  @samanthaprince Why are you boycotting Starbucks?? ([view](#))
-  If you haven't heard about the Starbucks Gold card program, read about it here: <http://bit.ly/3ljXJO> Worth playing for? ([view](#))
-  @MyStarbucksIdea How do I enter the Starbucks Gold giveaway? ([view](#))

4.30% NEGATIVE

-  @macoy sore throat from the dark roast cheesecake? @rom have you tried the dark roast cheesecake at starbucks? its my addiction for the week ([view](#))
-  ...i'm really really thinking about not showing up for work tomorrow...or ever again...god i'm so pissed... **i hate starbucks** ([view](#))
-  **i hate** adium it **won't** let me change my icon. im still a starbucks drink from freshman year with startrek sounds. arg ([view](#))
-  omfg in a starbucks... shaft theme song! i hear that shaft is a **bad** mutha... ([view](#))
-  going for a frappucino @ starbucks. i'm hot and **tired** from rehearsals! :-p ([view](#))
- going for a frappucino @

Sentiment analysis of tweets

Natural Language Processing: features and target variable in sentiment analysis

examples (news articles)

	features (words)	target
All bears are lovely	bear tea love	sentim 56%
Our tea was bad	bad drink	-35%
That bear drinks with bear		-5%
The bear drinks tea		4%
We love bears		63%
	$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5$	y

Stop words that were removed:

- all
- are, was
- our
- that
- with
- the

Natural Language Processing: ‘bag of words’ based on Word Frequency (WF) and sentiment analysis

examples (news articles)

All bears are lovely
 Our tea was bad
 That bear drinks with bear
 The bear drinks tea
 We love bears

	features (WF)					target
	bear	tea	love	bad	drink	sentim
All bears are lovely	1	0	1	0	0	56%
Our tea was bad	0	1	0	1	0	-35%
That bear drinks with bear	2	0	0	0	1	-5%
The bear drinks tea	1	1	0	0	1	4%
We love bears	1	0	1	0	0	63%

bag of words

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad y$

Supervised machine learning algorithm:

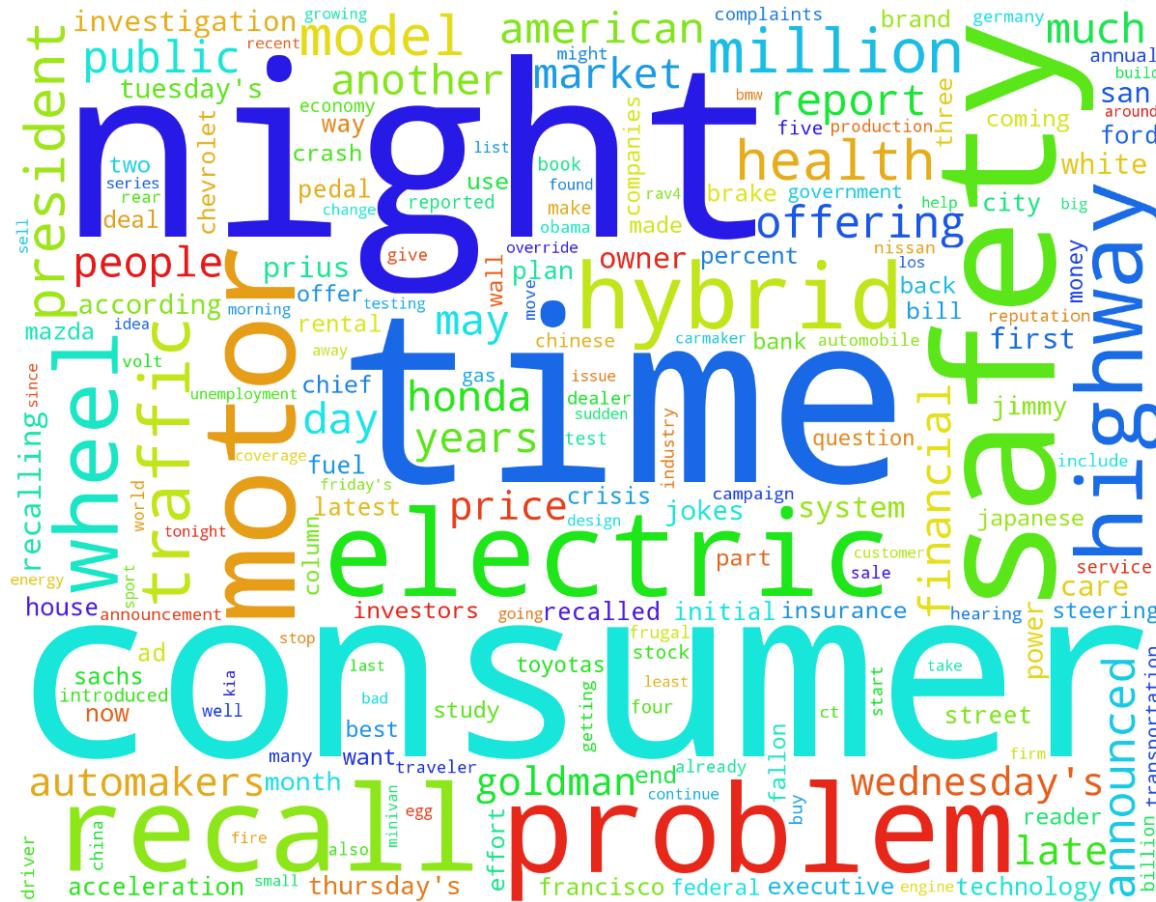
- Linear regression
- Decision trees
- SVM regression
- k-NN regression
- Ensembles (random forests, XGBoost)
- Artificial neural nets (deep learning)

$$y = f_{\theta}(x) = f_{\theta}(x_1, \dots, x_5)$$

$$y = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \dots + \theta_5 \cdot x_5 + \epsilon$$

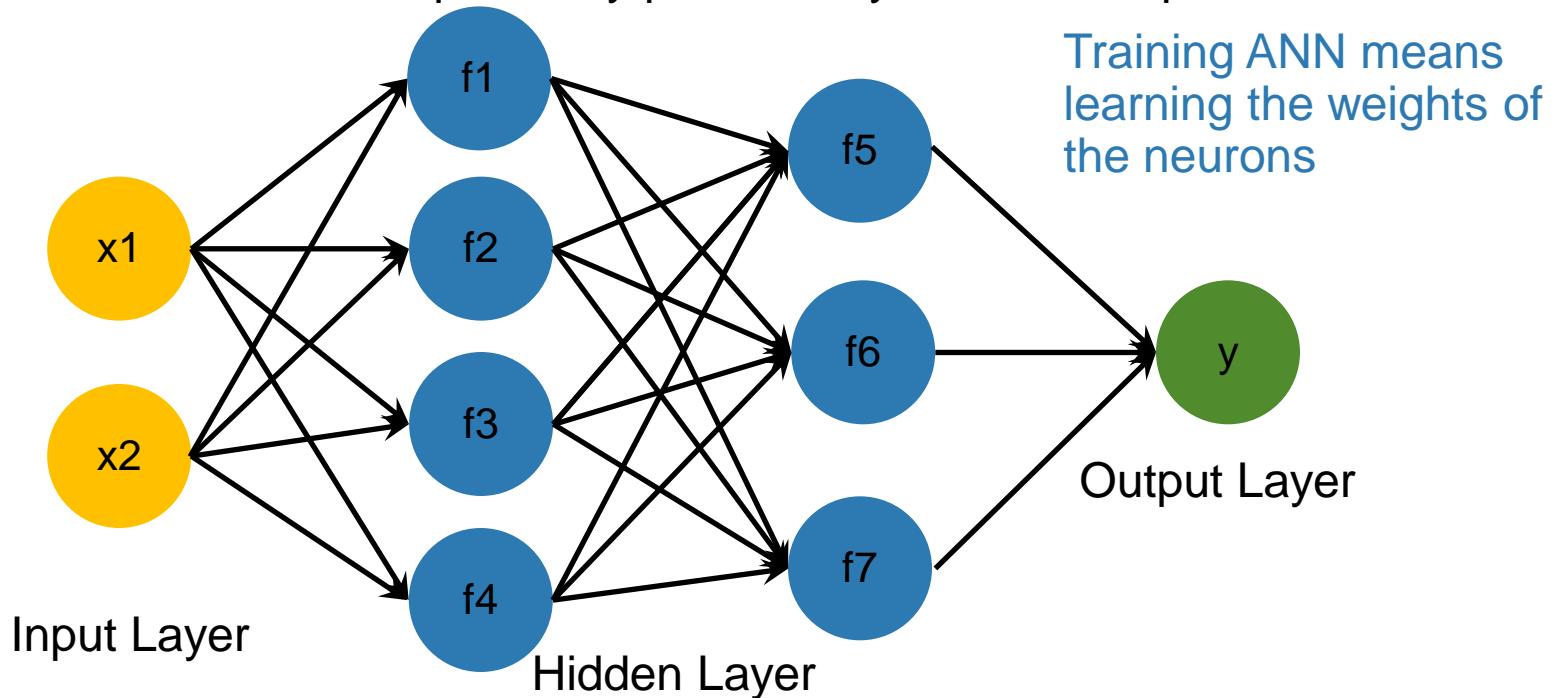
Natural Language Processing: word frequency (Word Cloud)

Word Cloud about Toyota

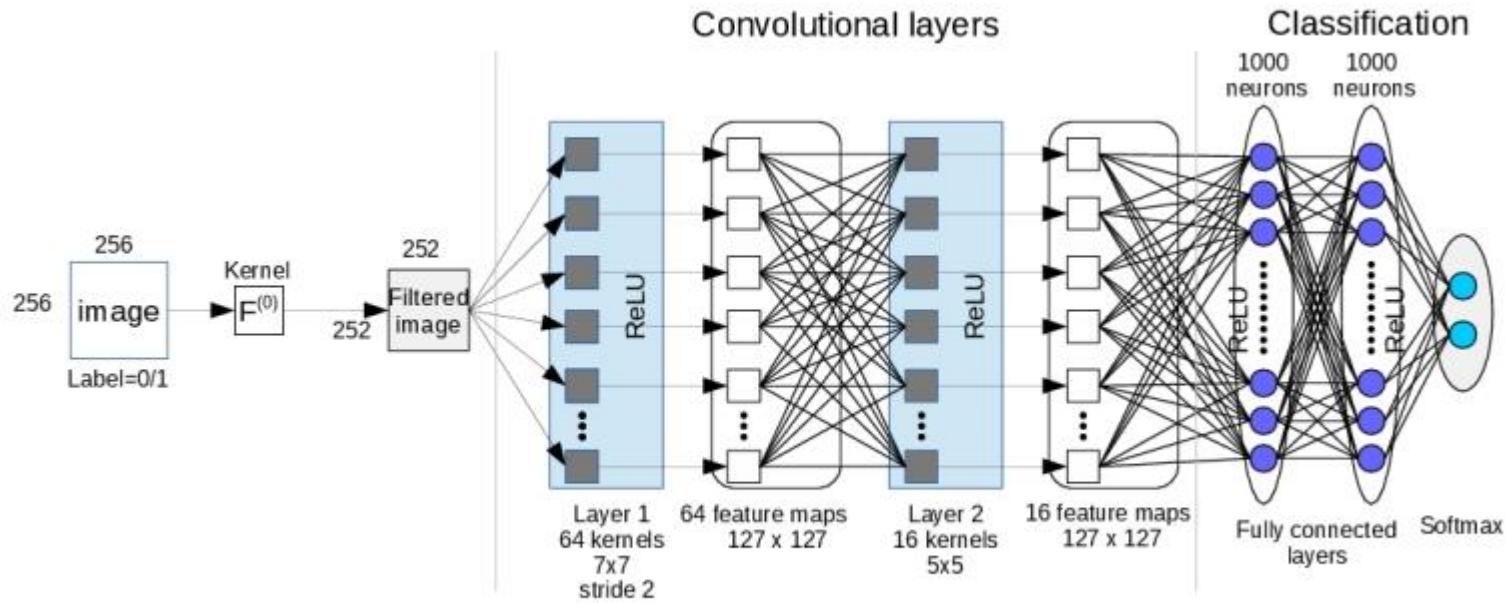
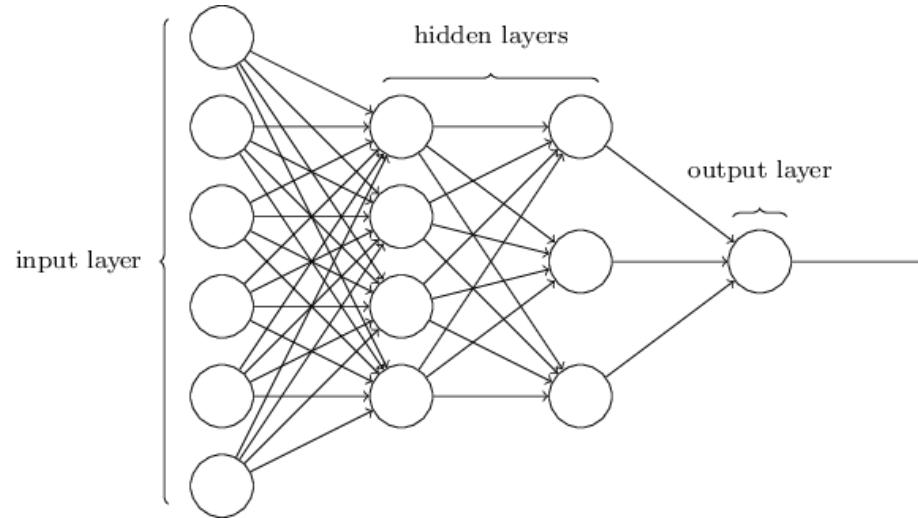


Neural networks and deep learning

- Based loosely on computer models of how brains work
- Model is an assembly of inter-connected neurons (nodes) and weighted links
- Each neuron applies a nonlinear function to its inputs to produce an output
- Output node sums up each of its input value according to the weights of its links
- Used for classification, pattern recognition, speech recognition
- “Black Box” model – no explanatory power, very hard to interpret the results

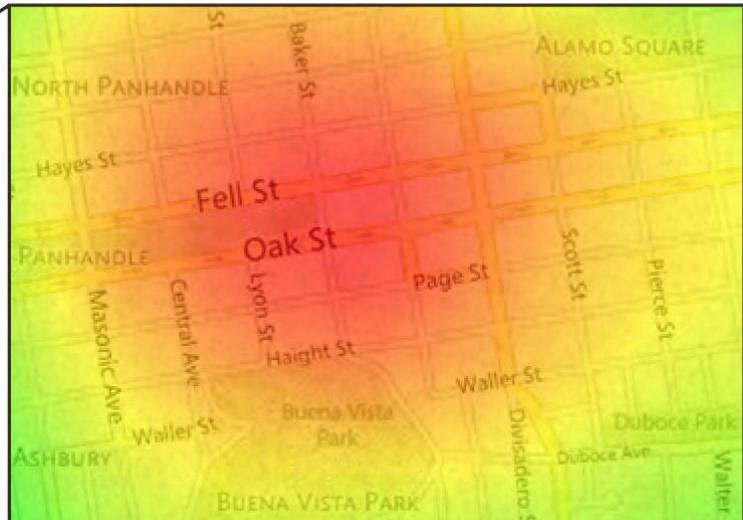
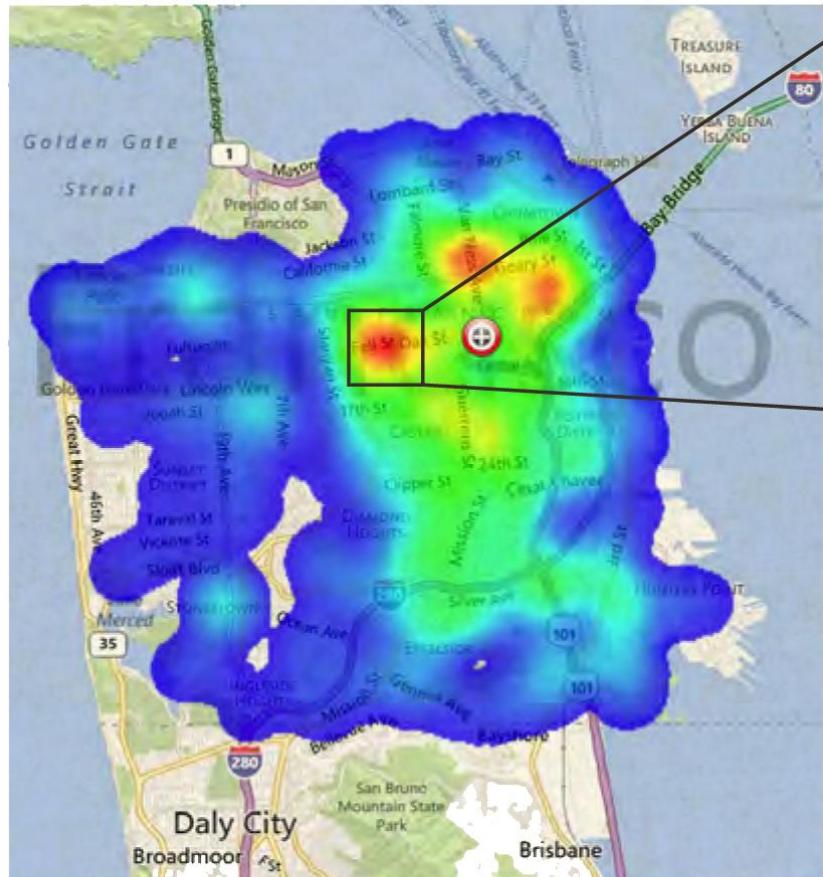


Neural networks and deep learning

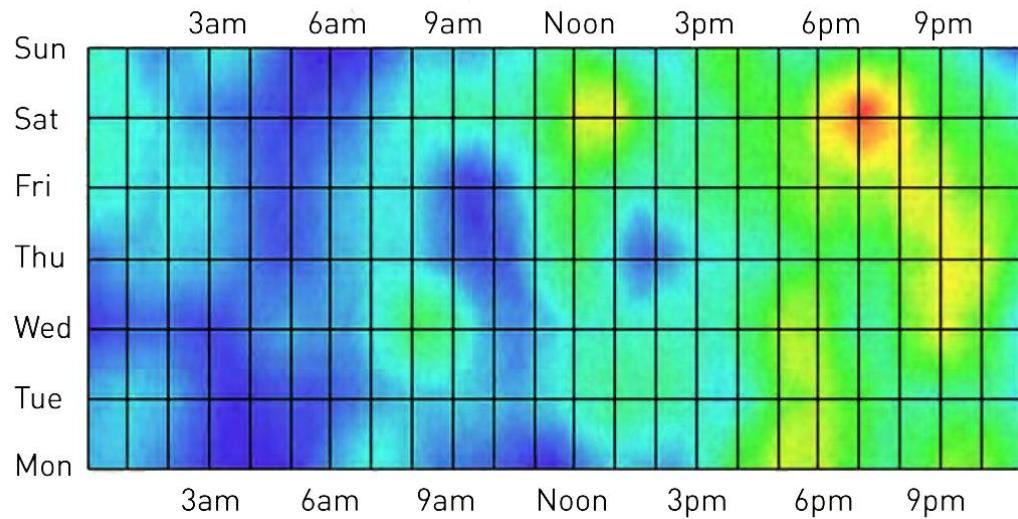


Spatio-temporal analytics – car theft hotspots

GEOSPATIAL HOTSPOTS



TEMPORAL HOTSPOT





ChatGPT API

Case study – ChatGPT API for YouTube video summarization and text generation

```
In [17]: # OpenAI API parameters  
model = "gpt-3.5-turbo-16k"  
# model = "gpt-3.5-turbo" # 4K tokens  
# model = "gpt-4"  
max_tokens = 1024  
n = 1  
stop = None  
temperature = 0.5
```

Prompt 1 - ChatGPT generating text for a Facebook post ¶

```
In [18]: prompt_1 = 'Please write a three sentence Facebook post in French language about how Canada can innovate based on the followi
```

```
In [19]: prompt1 = prompt_1.format(input=text)  
response1 = openai.ChatCompletion.create(  
    model=model,  
    messages=[  
        {"role": "system", "content": "You are a helpful assistant."},  
        {"role": "user", "content": prompt1},  
    ],  
    max_tokens=max_tokens,  
    n=n,  
    stop=stop,  
    temperature=temperature,  
)
```

```
In [20]: coutput1 = response1['choices'][0]['message']['content']  
print(coutput1)
```

"Le Canada a la possibilité d'innover en s'inspirant de l'expérience de la Silicon Valley aux États-Unis. Comme le montre cette vidéo, la clé du succès de la Silicon Valley réside dans la flexibilité, la collaboration entre l'université et l'industrie, ainsi que le financement par le capital-risque. Le Canada peut adopter ces facteurs clés pour encourager l'innovation et stimuler la croissance économique dans le pays. #Innovation #SiliconValley #Canada"

<https://youtu.be/WNTtWYFhWus?t=20>



Cloud

IBM Cloud – all services

The screenshot shows the IBM Cloud Catalog interface. At the top, there's a navigation bar with 'IBM Cloud', a search bar, 'Catalog', 'Manage', and user information. Below the navigation is a decorative banner with people interacting with servers and databases. The main area is titled 'Catalog' and has a search bar. On the left, a sidebar lists categories: Compute (30), Containers (11), Networking (29), Storage (21), Converged infrastructure (3), Enterprise applications (3), AI / Machine Learning (22), Analytics (14), Blockchain (1), Databases (22), Developer tools (18), Logging and monitoring (3), Migration (9), Integration (12), Internet of Things (1), Security (28), and Mobile (2). The central area displays 'Viewing 212 products' and a grid of service cards. Each card includes a logo, name, provider, description, and status. The cards shown are:

- Analytics Engine** By IBM: Submit your Apache Spark applications as needed and customize the Spark runtimes to satisfy the requirements of... IAM-enabled • IBM supported
- AnonTech ViziVault Platform** By Anon Technology, Inc.: Manage personal information as-a-service safely, securely, and in compliance with data privacy regulation... Lite • Free • HIPAA Enabled • IAM-enabled • Third party supported
- Anycloud Backup for 365** By any.cloud A/S: Anycloud Backup for 365 is a partner-ready SaaS offering of data protection and is created to safely backup data an... IAM-enabled • Third party supported
- API Connect** By IBM: An enterprise-grade platform for creating, securing, managing, sharing, monetizing, and analyzing custom APIs located on-... Lite • Free • IAM-enabled • IBM supported
- App Configuration** By IBM: Centralized, in-flight configuration for web and mobile applications and distributed environments. Lite • Free • Financial Services Validated • IAM-enabled • Service Endpoint Supported • IBM supported
- App Connect** By IBM: Connect your applications, automate tasks, and improve productivity Lite • Free • Deprecated • IBM supported
- App ID** By IBM: User Authentication and User Profiles for your apps. Lite • Free • EU Supported • Financial Services Validated • HIPAA Enabled • IAM-enabled • IBM supported
- Bare Metal Servers for Classic** By IBM: IBM Cloud Bare Metal Servers provide performance, flexibility, on-demand provisioning, and control. SAP Certified • IBM supported

<http://cloud.ibm.com>

Click "Catalog" at the top of the dashboard

IBM Cloud – Watson AI / ML services

Screenshot of the IBM Cloud Catalog interface showing the AI / Machine Learning section.

Header: IBM Cloud, Search resources and products..., Catalog, Manage, 2256482 - Oleksandr R..., Help, Notifications, User profile.

Search Bar: Search the catalog...

Filter Bar: AI / Machine Learning (selected), Alphabetically, Sort icon.

Product Grid: Viewing 12 products.

Type	Provider	Pricing plan	Icon	Name	Description	Details
All	IBM (12)	Lite		Watson Assistant	By IBM	Watson Assistant lets you build conversational interfaces into any application, device, or channel. Lite • Free • EU Supported • HIPAA Enabled • IAM-enabled • IBM supported
Services	Bespoken (1)			Watson Studio	By IBM	Develop sophisticated machine learning models using Notebooks and code-free tools to infuse AI throughout your... Lite • Free • HIPAA Enabled • IAM-enabled • IBM supported
Software	Cerebral Blue LLC (2)			IBM Match 360 with Watson	By IBM	IBM Match 360 with Watson (Match 360) improves trust in AI pipelines by identifying duplicate records and... Lite • Free • IAM-enabled • IBM supported
Deployable architectures	Cognitive View (1)			Knowledge Studio	By IBM	Teach Watson the language of your domain. Lite • Free • EU Supported • HIPAA Enabled • IAM-enabled • IBM supported
Professional services	Dubber Pty Ltd (1)			Language Translator	By IBM	Translate text, documents, and websites from one language to another. Create industry or region-specific translations... Lite • Free • EU Supported • HIPAA Enabled • IAM-enabled • IBM supported
	Insight Specialist Ltd. (4)			Natural Language Understanding	By IBM	Analyze text to extract meta-data from content such as concepts, entities, emotion, relations, sentiment and more. Lite • Free • EU Supported • IAM-enabled • IBM supported
	Show more			Speech to Text	By IBM	Low-latency, streaming transcription Lite • Free • EU Supported • HIPAA Enabled • IAM-enabled • IBM supported
				Text to Speech	By IBM	Synthesizes natural-sounding speech from text. Lite • Free • EU Supported • HIPAA Enabled • IAM-enabled • IBM supported

<http://cloud.ibm.com>

Click "Catalog" at the top of the dashboard

IBM Cloud – Speech to Text service

The screenshot shows the IBM Cloud interface for the Speech to Text service. At the top, there's a navigation bar with 'IBM Cloud' and a search bar. Below the search bar, the 'Catalog' tab is selected. The main content area displays the 'Speech to Text' service details.

Catalog / Speech to Text
Low-latency, streaming transcription

Create **About**

Type: Service **Provider**: IBM **Updated on**: 08/19/2022 **Category**: AI / Machine Learning **Compliance**: EU Supported, HIPAA Enabled, IAM-enabled **Location**: Sydney, Frankfurt, London, Tokyo, Washington DC, Dallas, Seoul

Select a location: Dallas (us-south)

Select a pricing plan: Displayed prices do not include tax. Monthly prices shown are for country or location: Canada

Plan	Features	Pricing
Lite	500 Minutes per Month	Free
Plus - NEW!	Minutes Per Month Simple Volume Tiers	Click to view tiers and pricing detail
Premium - NEW!	Everything in Plus Plan, plus... The Premium Plan provides the same features and benefits of using the Plus Plan, but with significantly greater capacity for concurrent transcriptions, enhanced security features to ensure that your data is isolated and encrypted end-to-end while in transit and at rest, and HIPAA readiness. Up to 500 concurrent transcriptions with the option to add more, and 150,000 free minutes to start.	

Summary

Speech to Text **Location**: Dallas **Plan**: Lite **Service name**: Speech to Text-al **Resource group**: Default

I have read and agree to the following license agreements:
[Terms](#)

Create

IBM Cloud – Speech to Text service

The screenshot shows the IBM Cloud interface for managing a "Speech to Text-hy" service instance. The left sidebar has a "Service credentials" tab selected. The main area displays a table of service credentials, with one entry shown in detail:

Key name	Date created
Auto-generated service credentials	2021-08-20 12:20 AM

The detailed view shows the following JSON data:

```
{  
  "apikey": "REDACTED",  
  "iam_apikey_description": "Auto-generated for key 8960cbf9-484f-4d9c-a2db-b4f5b4d73581",  
  "iam_apikey_name": "Auto-generated service credentials",  
  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Manager",  
  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity:a/4cbf213887dbd668ecd75f8c909624f5::serviceid:ServiceId-61ab65ac-d959-45f4-a5b0-5215df61a4d2",  
  "url": "https://api.us-south.speech-to-text.watson.cloud.ibm.com/instances/b926f502-060d-4edf-8147-e58c4bb1d54"  
}
```

Copy “API Key” to your Python code

Case study – Watson Speech-to-Text, Natural Language Understanding and Text-to-Speech prototype on IBM Cloud

```
In [5]: ⏎ import IPython  
IPython.display.Audio(filename)
```

Out[5]:



Speech-to-Text

```
In [7]: ⏎ from ibm_watson import SpeechToTextV1  
from ibm_cloud_sdk_core.authenticators import IAMAuthenticator  
  
# Instantiate the service using your credentials  
speech_to_text = SpeechToTextV1(  
    authenticator=IAMAuthenticator('fda4hUJkXhu3eN4IJKOWiqygFrMJOpJjb2NYRG184f7y')  
)  
  
audio_file = open(filename, "rb")  
  
result = speech_to_text.recognize(audio=audio_file, content_type='audio/wav', timestamps=True, word_confidence=True).get_result()  
  
result
```

Out[7]:

```
{'result_index': 0,  
 'results': [{final: True,  
 'alternatives': [{transcript: 'thunderstorms could produce large hail isolated tornadoes and heavy rain ',  
 'confidence': 1.0,  
 'timestamps': [[['thunderstorms', 1.45, 2.32],  
 ['could', 2.32, 2.55],  
 ['produce', 2.55, 3.01],  
 ['large', 3.01, 3.35],  
 ['hail', 3.35, 3.65],  
 ['isolated', 3.65, 3.95],  
 ['tornadoes', 3.95, 4.25],  
 ['and', 4.25, 4.55],  
 ['heavy', 4.55, 4.85],  
 ['rain', 4.85, 5.15]]},  
 'score': 1.0}],  
 'score': 1.0}]}
```



Crowd-Sourced Analytics

Bellingcat – open source investigations



За Воинскую доблесть II степени...

Added 27 January 2015 | Like 15

Show 12 previous comments



Пётр Борис

██████████, спасибо спасибо))

28 Jan at 5:17 pm to ██████████ | Reply



Пётр Борис

██████████, комбриг же сказал что нельзя выбрасывать фото с медалями!!!!

7 Feb at 12:56 pm | Reply



Пётр Борис

██████████, мне лично он ничего не говорил, так что как то пох))

7 Feb at 1:37 pm to ██████████ | Reply



Пётр Борис
██████████

Share

Save to my photos

Open original

Report ▾

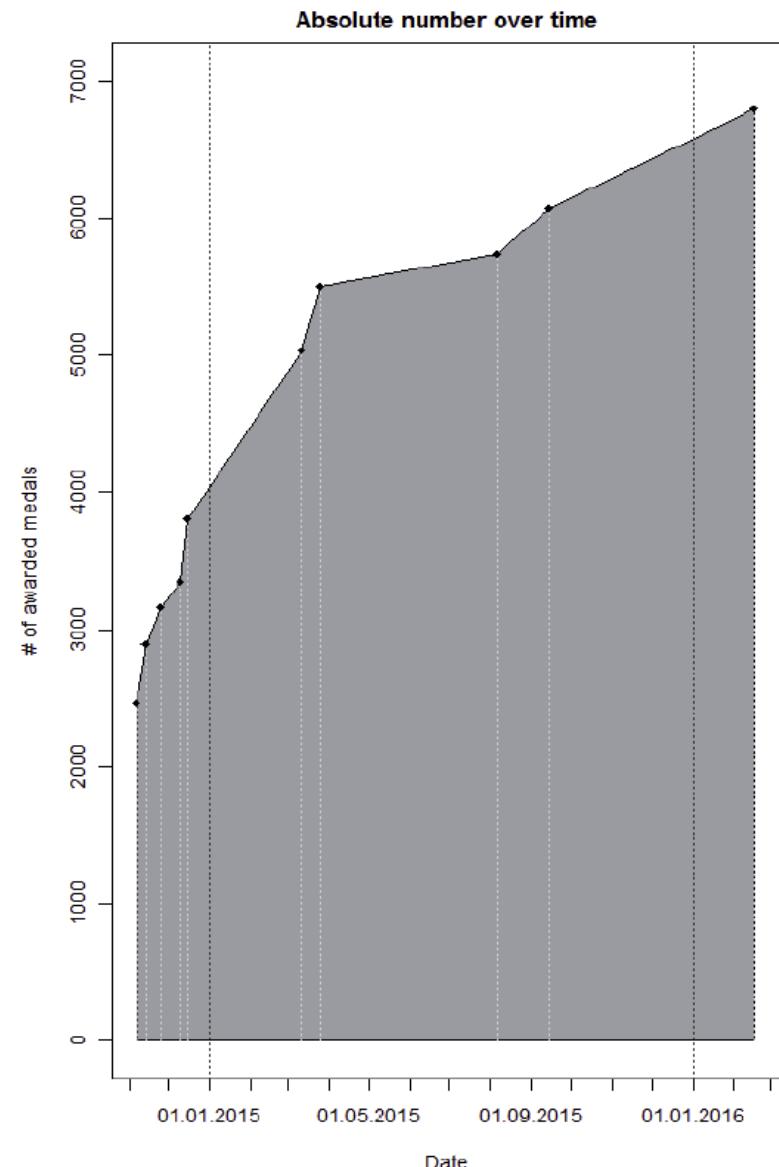
Bellingcat – open source investigations



Bellingcat – open source investigations

A number of awarded medals “For Distinction in Combat” is 4300 between 07.11.2014 and 18.02.2016, strongly suggests larger combat operations with active Russian military involvement in this period. In sum, the data suggests that **more than 10000** medals of all four considered types were awarded in the considered period.

Date	Number	Comment
07.11.2014	2464	
14.11.2014	2889	
25.11.2014	2984	
25.11.2014	3077	
25.11.2014	3102	
25.11.2014	3164	
XX.12.2014	3261	Day not readable
10.12.2014	3346	
XX.XX.2014	3380	Date not readable, likely 2014
16.12.2014	3479	
16.12.2014	3804	
12.03.2015	5033	“Medal of Suvorov” 41799
15.04.2015	5387	Date does not fit medal number
26.03.2015	5462	
26.03.2015	5502	
26.03.2015	5507	
06.08.2015	5741	
XX.09.2015	6076	Day not readable
18.02.2016	6802	





Online and In-Person Education

Coursera (coursera.org)

coursera Explore 

Online Degrees Find your New Career For Enterprise For Universities Log In Join for Free

Filter by

Subject

Arts and Humanities
 Business
 Computer Science
 Data Science
[Show more](#)

Skills

3d Modeling
 Account Management
 Accounting
 Accounting Software
[Show more](#)

Level

Beginner
 Intermediate
 Advanced
 Mixed

Duration

3-6 Months
 1-4 Years

2787 results for "data science"



IBM Skills Network

IBM Data Science

Skills you'll gain: Algebra, Algorithms, Business Analysis, Communication, Computational Logic...

4.6 (97.5k reviews)

Beginner · Professional Certificate · 3-6 Months



Google

Google Data Analytics

Skills you'll gain: Algorithms, Analysis, Application Development, Big Data, Budget...

4.8 (75.1k reviews)

Beginner · Professional Certificate · 3-6 Months



IBM Skills Network

Introduction to Data Science

Skills you'll gain: Analysis, Communication, Computer Programming, Data Analysis, Data...

4.6 (71.5k reviews)

Beginner · Specialization · 3-6 Months



University of Michigan

Applied Data Science with Python

Skills you'll gain: Algorithms, Analysis, Applied Machine Learning, Artificial Neural Networks,...

4.5 (32.2k reviews)

Intermediate · Specialization · 3-6 Months



Johns Hopkins University

Data Science

Skills you'll gain: Algorithms, Application Development, Applied Machine Learning,...

4.5 (49.3k reviews)

Beginner · Specialization · 3-6 Months



DeepLearning.AI, Stanford University

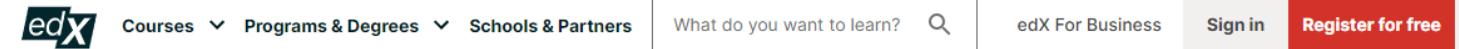
Machine Learning

Skills you'll gain: Accounting, Algorithms, Applied Machine Learning, Artificial Neural...

4.9 (3.2k reviews)

Beginner · Specialization · 1-3 Months

EdX (edx.org)



Catalog > Data Analysis & Statistics Courses



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.



13 weeks

10–15 hours per week



Instructor-paced

Instructor-led on a course schedule



Free

Optional upgrade available

This course is archived

[View course materials](#)

I would like to receive email from MITx and learn about other offerings related to The Analytics Edge.

[About](#)

[What you'll learn](#)

[Instructors](#)

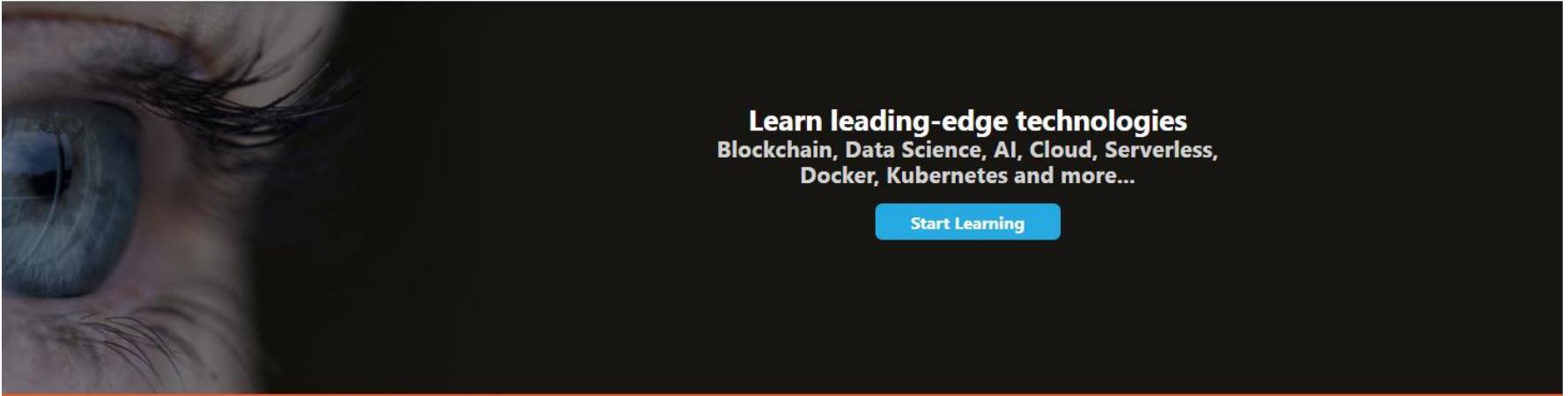
[FAQs](#)

[Ways to enroll](#)

About this course

In the last decade, the amount of data available to organizations has reached unprecedented levels. Data is transforming business, social interactions, and the future of our society. In this course, you

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



The website features a large, high-resolution close-up of a person's eye in the background, occupying the top half of the page. The eye is light-colored with dark eyelashes.

COGNITIVE CLASS.ai

Courses & Projects Badges Learning Paths Business Sign in **Register**

Learn leading-edge technologies
Blockchain, Data Science, AI, Cloud, Serverless,
Docker, Kubernetes and more...

Start Learning

Why learn with us?

EARN CERTIFICATES AND BADGES
Showcase your newly-acquired skills.

IT'S FREE
Our courses are free of charge!

TAKE IT TO THE NEXT LEVEL
Build your career with skills employers seek.

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>)

- Free courses, free study materials
- Cloud-based sandbox for exercises
- 2000000+ registered students
- 80+ courses

Screenshot of the CognitiveClass MOOC website showing the course catalog.

The page features a navigation bar with links for Courses & Projects, Badges, Learning Paths, Business, a search bar, and sign-in/register buttons.

The main content area is titled "Explore Content" and includes a "Reset Filters" button and a "Tier" filter dropdown set to "Free".

The results section displays 211 courses in a grid format:

- Blockchain Essentials**: Beginner Course, 3 hours, 4.5 stars (405 reviews)
- Data Science Methodology**: Beginner Course, 3 hours, 4.5 stars (1.35k+ reviews)
- Docker Essentials: A Developer Introduction**: Beginner Course, 3 hours, 4.5 stars (1.69k+ reviews)
- Deep Learning Fundamentals**: Beginner Course, 3 hours, 5 stars (0 reviews)
- Python for Data Science**: Beginner Course, 20 hours, 4.5 stars (6.36k+ reviews)
- Introduction to Cloud**: Beginner Course, 6 hours, 4.5 stars (2.05k+ reviews)
- Hadoop**: Beginner Course, 3 hours, 5 stars (0 reviews)
- Machine Learning**: Beginner Course, 20 hours, 4.5 stars (6.36k+ reviews)
- Quantum Computing**: Beginner Course, 3 hours, 5 stars (0 reviews)

TED and TEDx

TED Ideas worth spreading

WATCH DISCOVER ATTEND PARTICIPATE ABOUT SIGN IN

MEMBERSHIP



LATOYA RUBY FRAZIER

A creative solution for the water crisis in Flint, Michigan



ANICKA YI

Intelligent floating machines inspired by nature



CATHERINE PRICE

Why having fun is the secret to a healthier life

• • •

Newest talks

[See all newest talks >](#)



SUTU
Everyone can participate in building the metaverse

Posted Sep 2022



VAITEA COWAN
How green hydrogen could end the fossil fuel era

Posted Sep 2022



HOLLY HERNDON
What if you could sing in your favorite musician's voice?

Posted Sep 2022



DEAN R. LOMAX
A giant Jurassic sea dragon, unearthed

Posted Sep 2022



KATHERINE MANGU-WARD
What capitalism gets right -- and governments get wrong

Posted Sep 2022



LEO LANNA AND LVCAS FIAT
The colorful, shapeshifting wonder of the Amazon's praying mantises

Posted Sep 2022

TED and TEDx

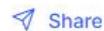


WATCH DISCOVER ATTEND



How AI could empower any business

2,134,502 views | Andrew Ng | TED2022 • April 2022



Share



Add



Like (64K)

Read transcript

Expensive to build and often needing highly skilled engineers to maintain, artificial intelligence systems generally only pay off for large tech companies with vast amounts of data. But what if your local pizza shop could use AI to predict which flavor would sell best each day of the week? Andrew Ng shares a vision for democratizing access to AI, empowering any business to make decisions that will increase their profit and productivity. Learn how we could build a richer society – all with just a few self-provided data points.

Technology, Engineering, Business, Entrepreneur, Software, Innovation, AI, Data, Machine Learning, Manufacturing

https://www.ted.com/talks/andrew_ng_how_ai_could_empower_any_business

Data Camp

The screenshot shows the DataCamp Groups interface for the 'MIE1624HF - Introduction ...' group. The left sidebar includes links for Invite Members, Dashboard, Members, Teams, Settings, LEARN (Custom Tracks, TRY, Assignments, Leaderboard, Insights & Analytics, Reporting, Custom Reports, Skill Matrix), and a search bar. The main content area displays 'Assignments / Everyone' with tabs for ACTIVE (selected), PAST DUE, and ARCHIVED. A green 'Create Assignment' button is in the top right. The 'Active Assignments' section has a search bar and a 'Filter By Type' dropdown. The assignment table lists five entries:

TITLE	ASSIGNEES	STATUS	DUE BY	C	A	CR	DETAILS
Introduction to Python Course	Organization	Active	Sep 20, 23:00 EDT	0	0	0%	<button>View</button>
Introduction to Data Science in Python Course	Organization	Active	Sep 27, 23:00 EDT	0	0	0%	<button>View</button>
Intermediate Python Course	Organization	Active	Oct 4, 23:00 EDT	0	0	0%	<button>View</button>
Python Data Science Toolbox (Part 2) Course	Organization	Active	Oct 11, 23:00 EDT	0	0	0%	<button>View</button>
Python Data Science Toolbox (Part 1) Course	Organization	Active	Oct 11, 23:00 EDT	0	0	0%	<button>View</button>

Register for free 6-month access with your email @mail.utoronto.ca

https://www.datacamp.com/groups/shared_links/3ddc04d58bf769882257745925a12c8665e9e932378dec021c9dccd5a1cfbf40



To Do before Lecture 2

Run IPython examples provided in class

■ Install Python on your laptop

- Recommended to use Python version 3.9, 3.10 or 3.11
- You may use your own Python distribution, Anaconda distribution is recommended to install <https://www.anaconda.com/download>

■ Use Python on cloud via Google Colab

- You can use Python on Google cloud via <https://colab.research.google.com>

■ Get access to ChatGPT and ChatGPT API

- Sign-in for ChatGPT at <https://chat.openai.com> (free version 3.5) and register for API access to OpenAI models at <https://platform.openai.com>

■ Get access to IBM Cloud (optional)

- Sign-in for [IBM Academic Initiative](#) and register for access to IBM Cloud, or get free lite access to IBM Cloud directly at <https://cloud.ibm.com/registration>

■ Check class web-page on Quercus