



Introduction to Data Science

2110531 Data Science and Data Engineering Tools (2025/1)

Assoc. Prof. Peerapon Vateekul, Ph.D.

Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University

Peerapon.v@chula.ac.th

www.cp.eng.chula.ac.th/~peerapon/



Outline: Understand definitions and terminologies

- Introduction
 - Data is important
 - Data Science Definition by Dr.Virote
 - Data Science Definition by Aj.Natawut
- Key Data Science Activities
 - Data Science Process
 - Types of Data Science Projects
 - AI/ML/DL/GenAI
 - Data Engineering + **Big Data Analytics**
 - MLOps
 - Cloud Technologies
- Conclusion
- Disclaimer

+

Introduction



Data is important (in 2017)

The Economist Topics ▾ Current edition More ▾

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules



Print edition | Leaders >
May 6th 2017

[Twitter](#) [Facebook](#) [LinkedIn](#) [Email](#) [Print](#)

- Alphabet (Google's parent company), Amazon, Apple, Facebook and Microsoft
- \$25bn in net profit in the first quarter of 2017
- Amazon captures half of all dollars spent online in America.
- Google and Facebook accounted for almost all the revenue growth in digital advertising in America last year

<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

+ Data is important (in 2018)! (cont.)

The New Oil

Jennifer Presley Executive Editor, E&P Magazine Hart Energy Thursday, November 1, 2018 - 6:40am



With a number of successful projects under its collective belt, the oil and gas industry is proving Big Data is more than just a buzzword. (Source: Makhnach_S/Shutterstock.com; Design by Felicia Hammons)

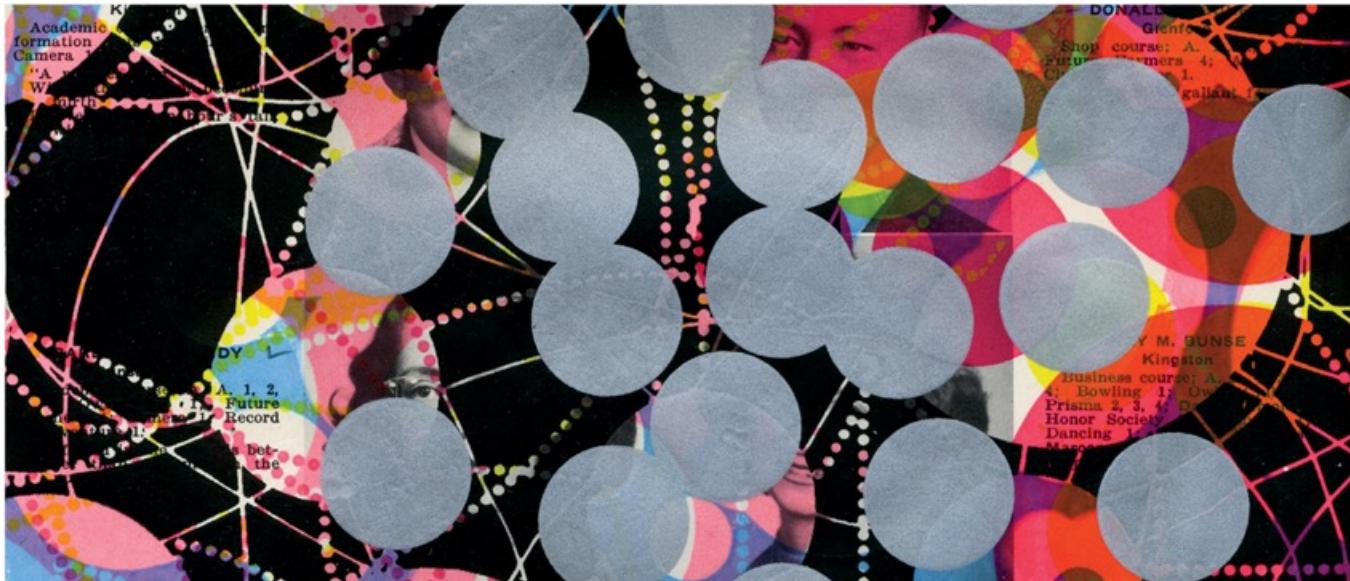
Data Science
(AI,ML,DM)
+
Big Data

<https://www.epmag.com/new-oil-1720651>



Who analyzes these data!

Harvard
Business
Review



DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

WHAT TO READ NEXT

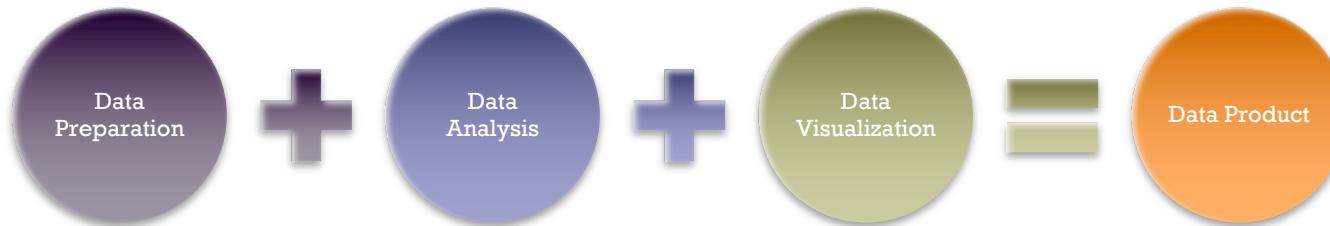


Competing on Analytics



What is data science? (aka. data analytics)

- Data
 - Facts and statistics collected for reference or analysis
- Science
 - A systematic study through observation and experiment
- Data Science
 - The scientific exploration of data to extract meaning or insight,
 - and the construction of software to utilize such insight in a business context.





What is data science? (cont.)

1. Transform data into **valuable insights**
2. Transform data into **data products**
3. Transform data into **interesting stories**



Ta Virot Chiraphadhanakul
Data Scientist, Facebook

Code Mania 2 (01), Jan-2015



1) Transform data into valuable insights

9

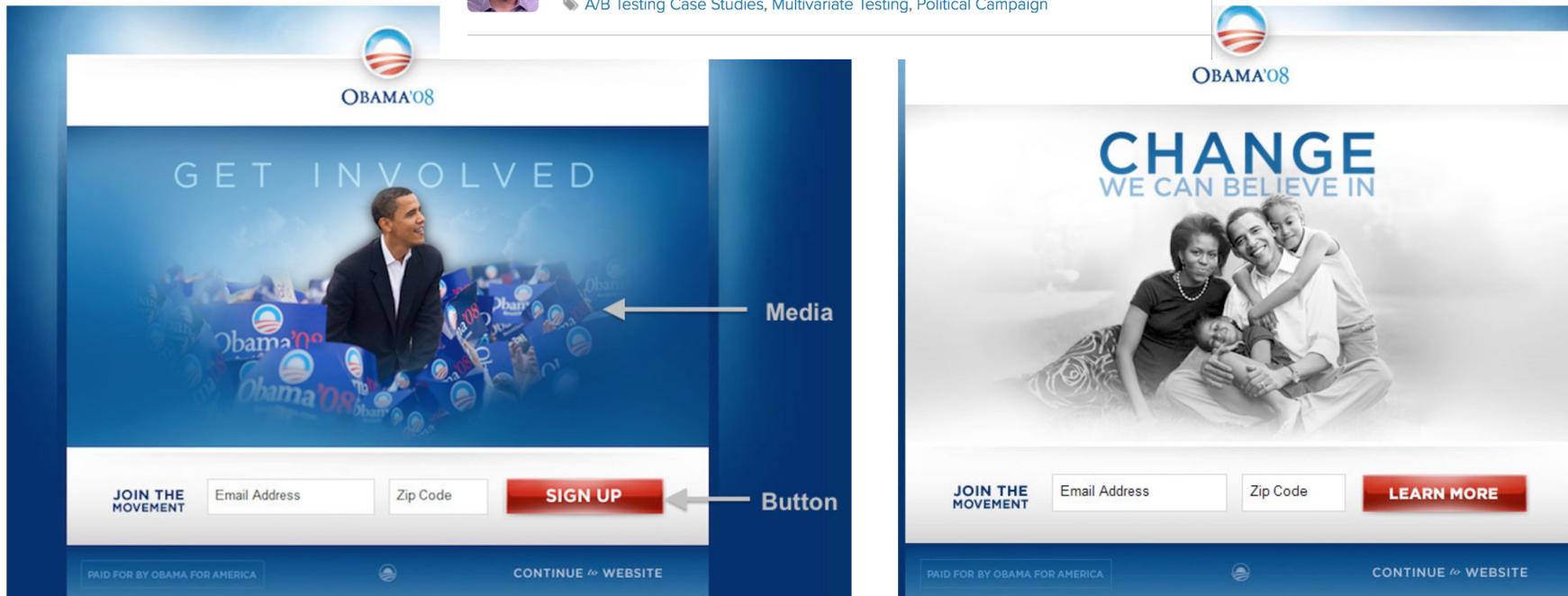
How Obama Raised \$60 Million by Running a Simple Experiment



By Dan Siroker

November 29, 2010

A/B Testing Case Studies, Multivariate Testing, Political Campaign





1) Transform data into valuable insights (cont.)



BUSINESS

Amazon introduces next major job killer to face Americans

By James Covert, Linda Massarella and Bruce Golding

December 5, 2016 | 9:59pm | Updated



The Amazon Go storefront
Amazon

<http://nypost.com/2016/12/05/amazon-introduces-next-major-job-killer-to-face-americans/>





2) Transform data into data products



Action required: Please confirm activity.



FRAUD PROTECTION SERVICES

Chase Sapphire
Account Ending: XXXX

We want to help keep your account secure so we continuously monitor it for possible fraudulent activity. We're writing to verify whether the transaction below was authorized by you or another Cardmember. Click YES below if you

The screenshot shows the Microsoft Outlook interface. The top navigation bar includes 'Outlook', a search bar, and various action buttons like 'New message', 'Empty folder', 'Mark all as read', and 'Undo'. On the left, a sidebar lists 'Favorites' and 'Folders'. The 'Inbox' folder is selected, showing 45 items. A red box highlights the 'Junk Email' folder, which contains 128 items. The main pane displays several email messages from spam sources, such as 'Work At Home Opportunities', 'NETFLIX SURVEY', and 'Thank You Costco'.

From	Subject	Date
W	Work At Home Opportunities New work from home progr...	1:47 PM
CS	Client service NETFLIX SURVEY	1:40 PM
TC	Thank You Costco Re: Costco Has a Surprise Fo...	12:01 PM
CS	Client service - Are you a friend of Amazo...	8:43 AM



3) Transform data into interesting stories Consumer Price Index (CPI) - Inflation

12

The Billion Prices Project

Home Our Public Data Our Research News

THE BILLION PRICES PROJECT

AN ACADEMIC INITIATIVE TO IMPROVE INFLATION MEASUREMENT

RESEARCH PAPERS DOWNLOAD DATA

<http://www.thebillionpricesproject.com/>





The Billion Prices Project: Using Online Prices for Measurement and Research *

14

Alberto Cavallo

MIT and NBER

Roberto Rigobon

MIT and NBER

This Version: April 8, 2016

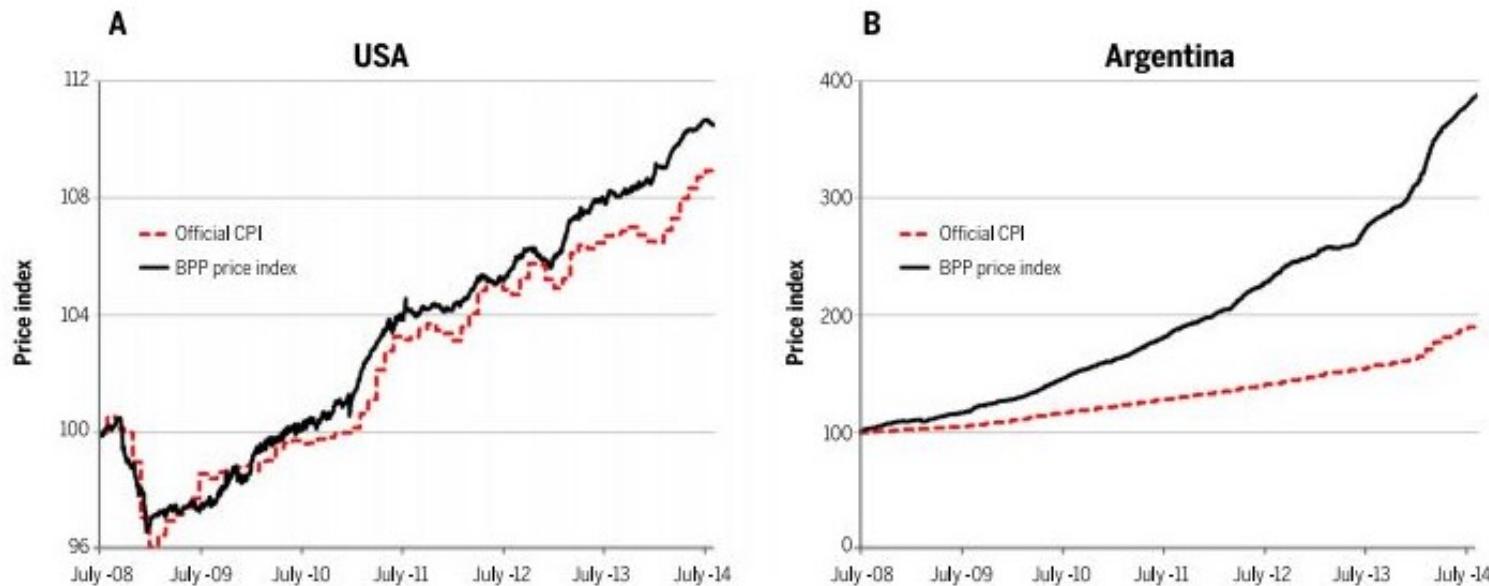


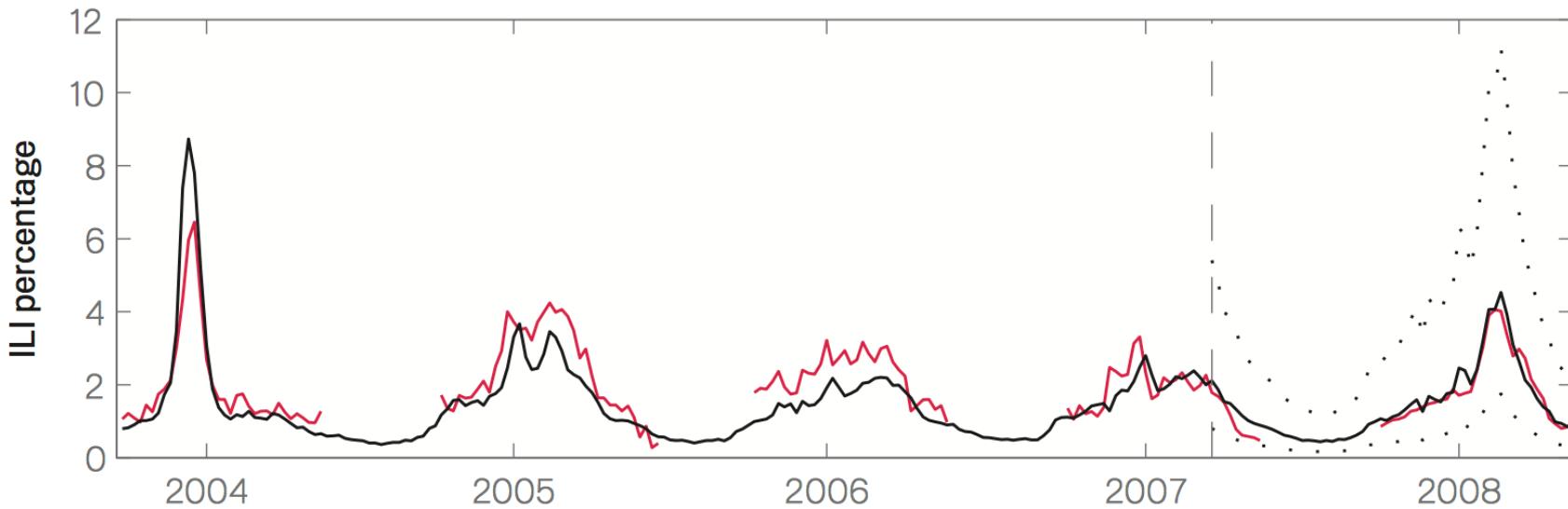
Fig. 2. BPP price index. Dashed red lines show the monthly series for the CPI in the United States (A) and Argentina (B), as published by the formal government statistics agencies. Solid black lines show the daily price index series, the "State Street's PriceStats Series" produced by the BPP, which uses scraped Internet data on thousands of retail items. All indices are normalized to 100 as of July 2008. In the U.S. context, the two series track

each other quite closely, although the BPP index is available in real time and at a more granular level (daily instead of monthly). In the plot for Argentina, the indices diverge considerably, with the BPP index growing at about twice the rate of the official CPI. [Updated version of figure 5 in (18), provided courtesy of Alberto Cavallo and Roberto Rigobon, principal investigators of the BPP]

https://www.hbs.edu/faculty/Publication%20Files/BPP_JEP_m_13b5e009-4162-4f2c-b507-593a9a98c082.pdf



Google Flu Trend



Ginsberg, Jeremy; Mohebbi, Matthew H.; Patel, Rajan S.; Brammer, Lynnette; Smolinski, Mark S.; Brilliant, Larry (19 February 2009). "Detecting influenza epidemics using search engine query data". *Nature*. **457** (7232): 1012–1014.



What are they using data science for?

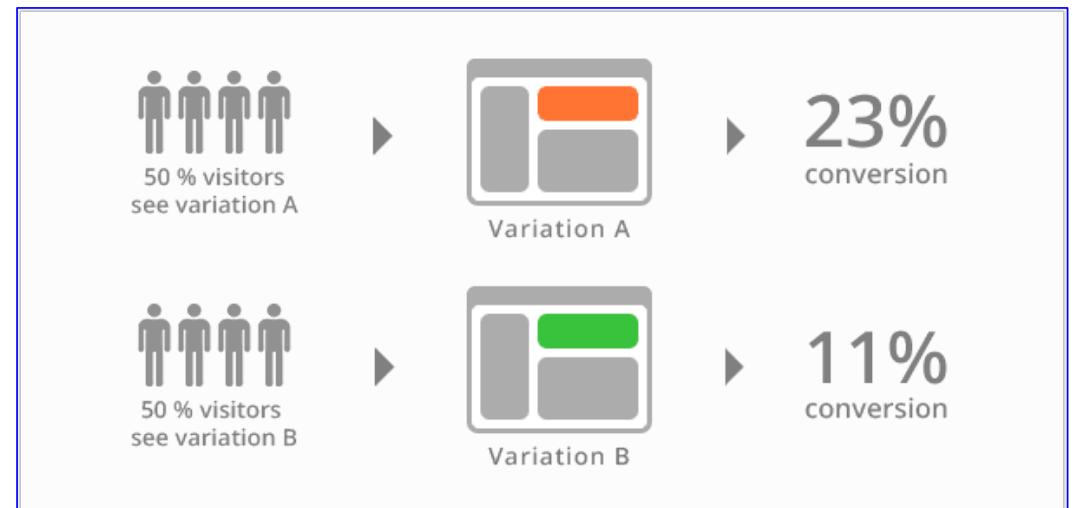
1. Measurement
2. Insights
3. Data Products





1) Measurement

- To **make a decision** based on data (aka. Benchmarking)
- Turning qualitative information into **quantitative values** (metrics or indicators)
- Support comparison:
 - Between options (e.g., which notebook to buy)
 - Before vs. after (e.g., tuning, upgrades)
 - **A/B Testing**



Source: <https://vwo.com/ab-testing/>



Example: SimCity

18

1. Remove product banner: SimCity sees 43% more conversions without hero banner image

Control

The screenshot shows the official SimCity website. At the top, there's a large hero banner with the text "PRE-ORDER AND GET \$20 OFF YOUR NEXT PURCHASE". Below this, there are two main product sections: "SIMCITY™" and "SIMCITY™ DIGITAL DELUXE EDITION". Each section includes a thumbnail image, the price (\$59.99 or \$79.99), and purchase options (PC Download or PC Physical). A red box highlights the hero banner.

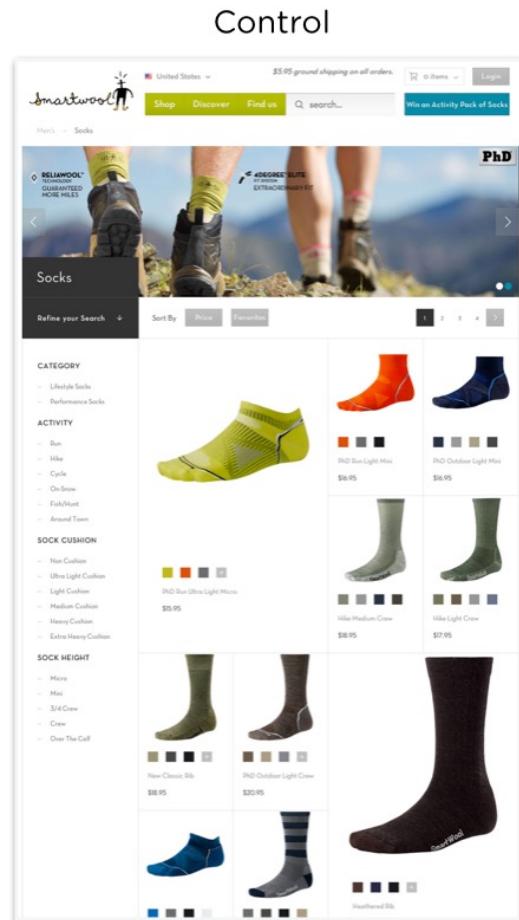
Variation

The screenshot shows the same SimCity website, but the hero banner at the top has been removed. Instead, there are two separate product sections: "SIMCITY™" and "SIMCITY™ DIGITAL DELUXE EDITION". To the right of the "DIGITAL DELUXE EDITION" section, a blue circular callout contains the text "43% increase in checkouts". Below the sections, there are "Key Features" and "WHAT IS SIMCITY?" sections.

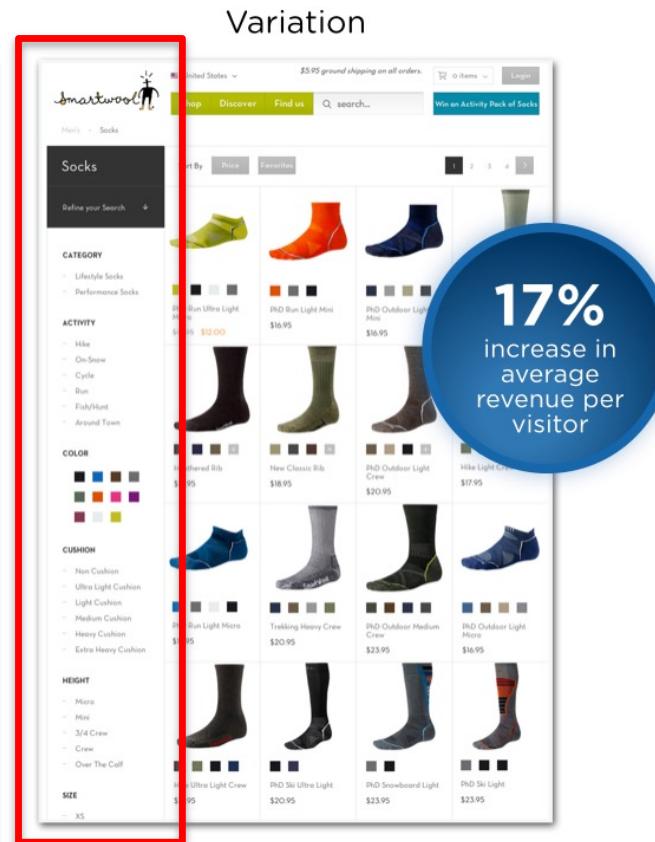
Source: <https://blog.optimizely.com/2015/06/04/ecommerce-conversion-optimization-case-studies/>



Example: SmartWool



3. Use a well-defined grid layout for your online shopping experience: Uniform product page images increase ARPV 17% for SmartWool



Source: <https://blog.optimizely.com/2015/06/04/ecommerce-conversion-optimization-case-studies/>



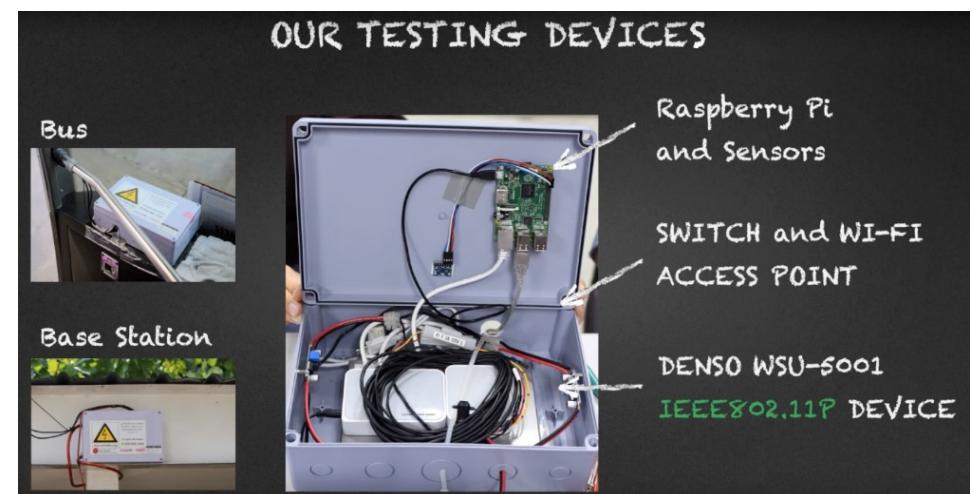
2) Insights

- **Good understanding of user behavior** can lead to new product development or improvements of the existing products
- **Financial startup** -- Typing with proper capitalization indicates creditworthiness
 - Online loan applicants who complete the application form with the correct case are more dependable debtors
- **Starbucks** use customer purchase information from My Starbucks Mobile Apps to figure out new products

<https://blogs.scientificamerican.com/quest-blog/9-bizarre-and-surprising-insights-from-data-science/>



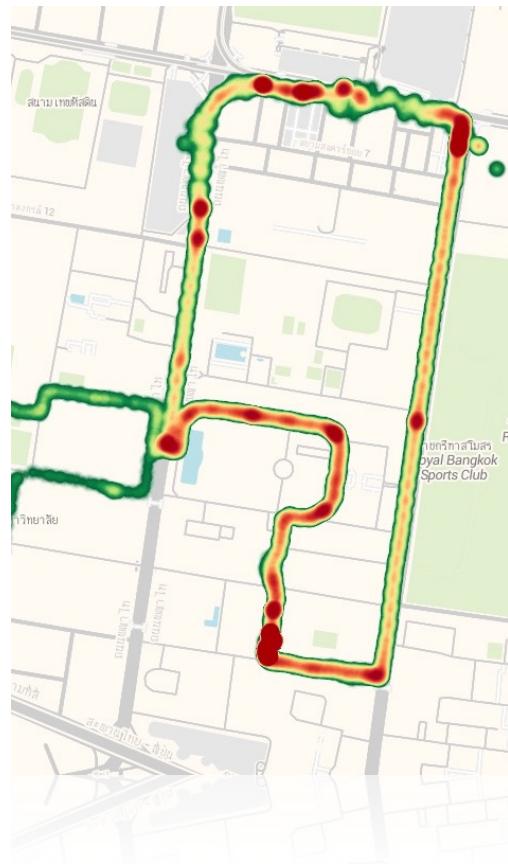
Example: Tracing Traffic



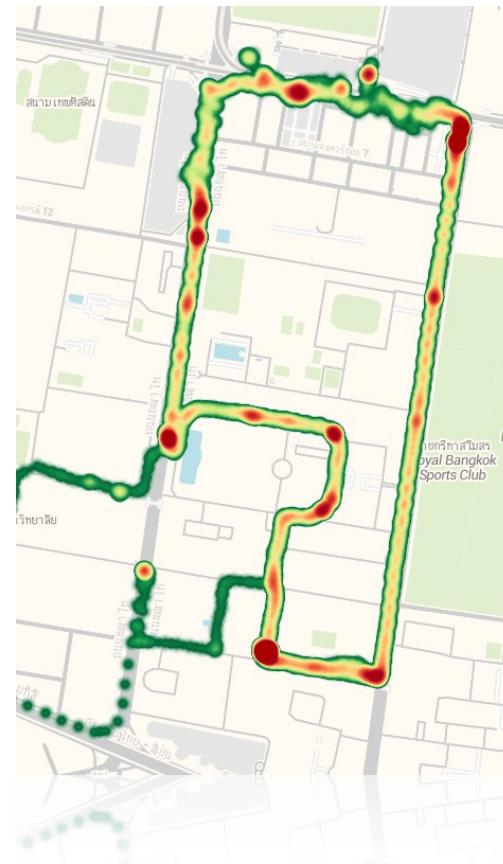


GPS Average Speed

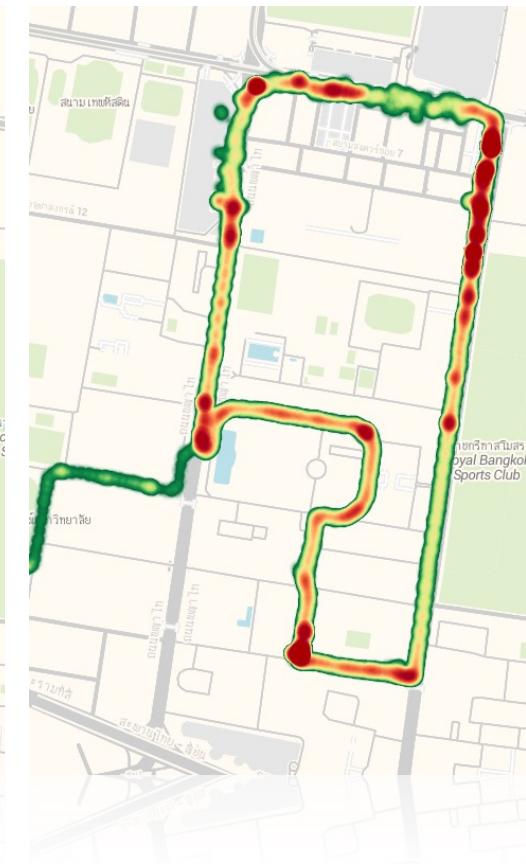
6:00-10:00



10:00-15:00



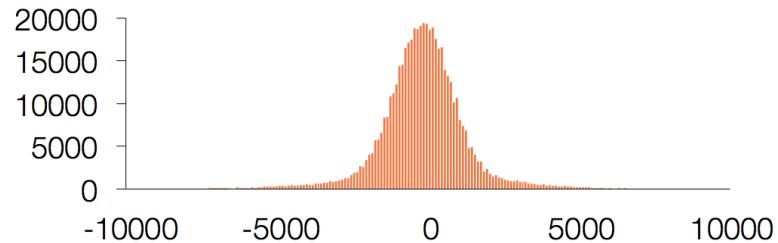
15:00-18:00



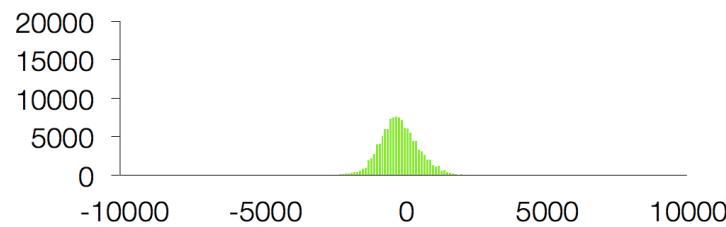
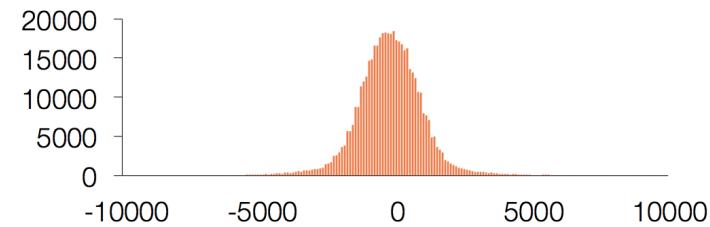
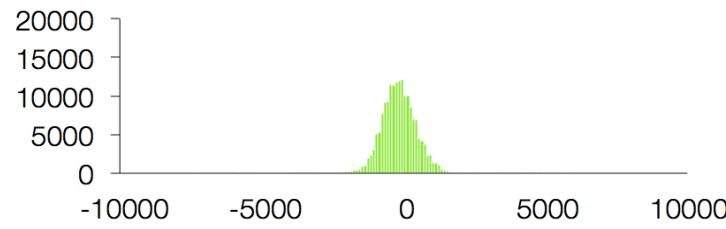
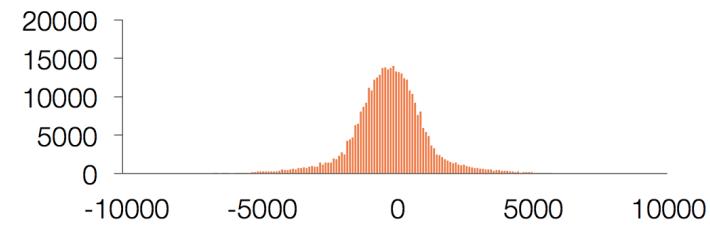
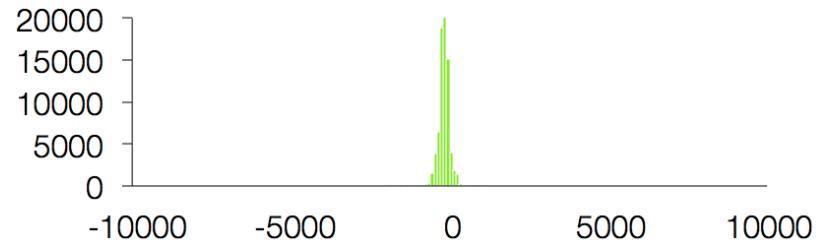


Bus Drivers' Behaviors

Bus A



Bus B





3) Data Products

- An application or system that uses data to provide “intelligent” products or services, which create more data that can be further used
- **Machine Learning** plays an important role in building great data products

Example: Amazon Recommendation

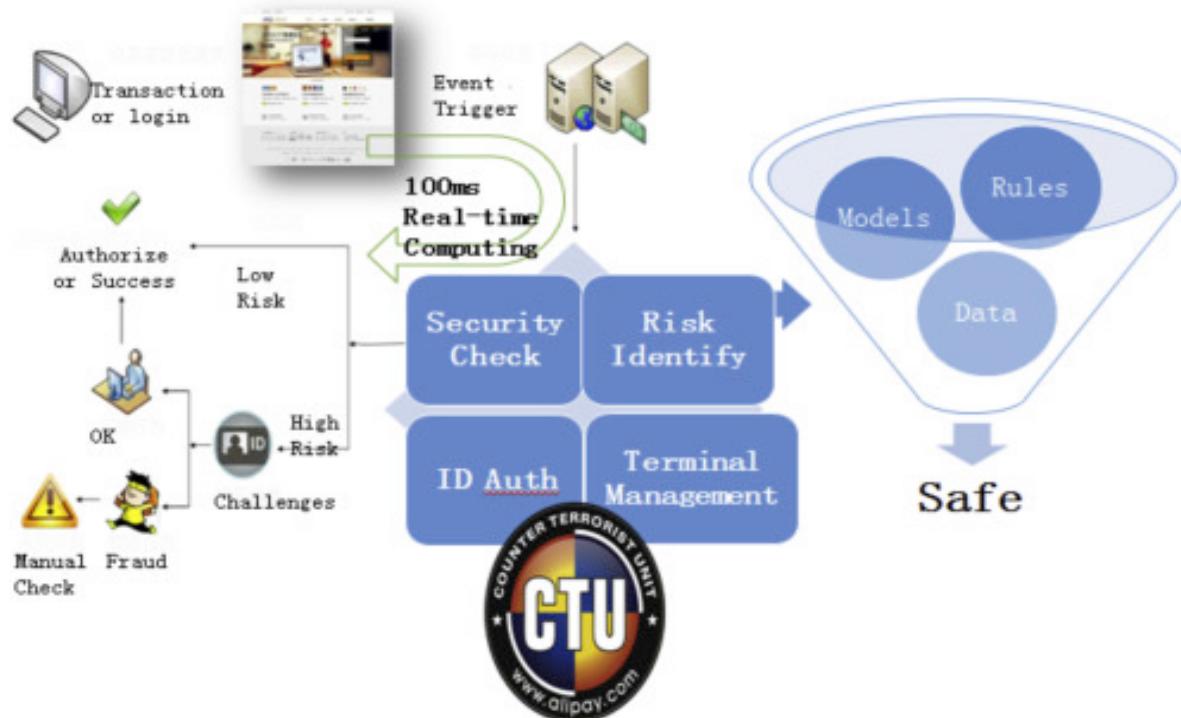
- Amazon sells 480M products (485k new products per day)
- Use recommendation systems to bring products to customers
- Analyze data from 300M customers
 - Purchase history
 - Reviews / Ratings
 - Search history
 - Views

The screenshot shows a portion of an Amazon website with a navigation bar at the top. Below the navigation, a banner for 'Natawut's Amazon' is displayed, along with a message encouraging users to sign in for order status and balances. The main content area features a grid of recommended products:

- Computer & Technology Books**: 92 items. Includes a large image of the book 'Hadoop Application Architectures' by Mark Grover, Ted Malaska, Jonathan Seidman, and Gwen Shapira.
- Science & Math Books**: 51 items. Includes a large image of the book 'Storytelling' by Robert J. Knell.
- Other recommendations**: Includes images of books like 'Own the Room' and 'Introductory Machine Learning'.



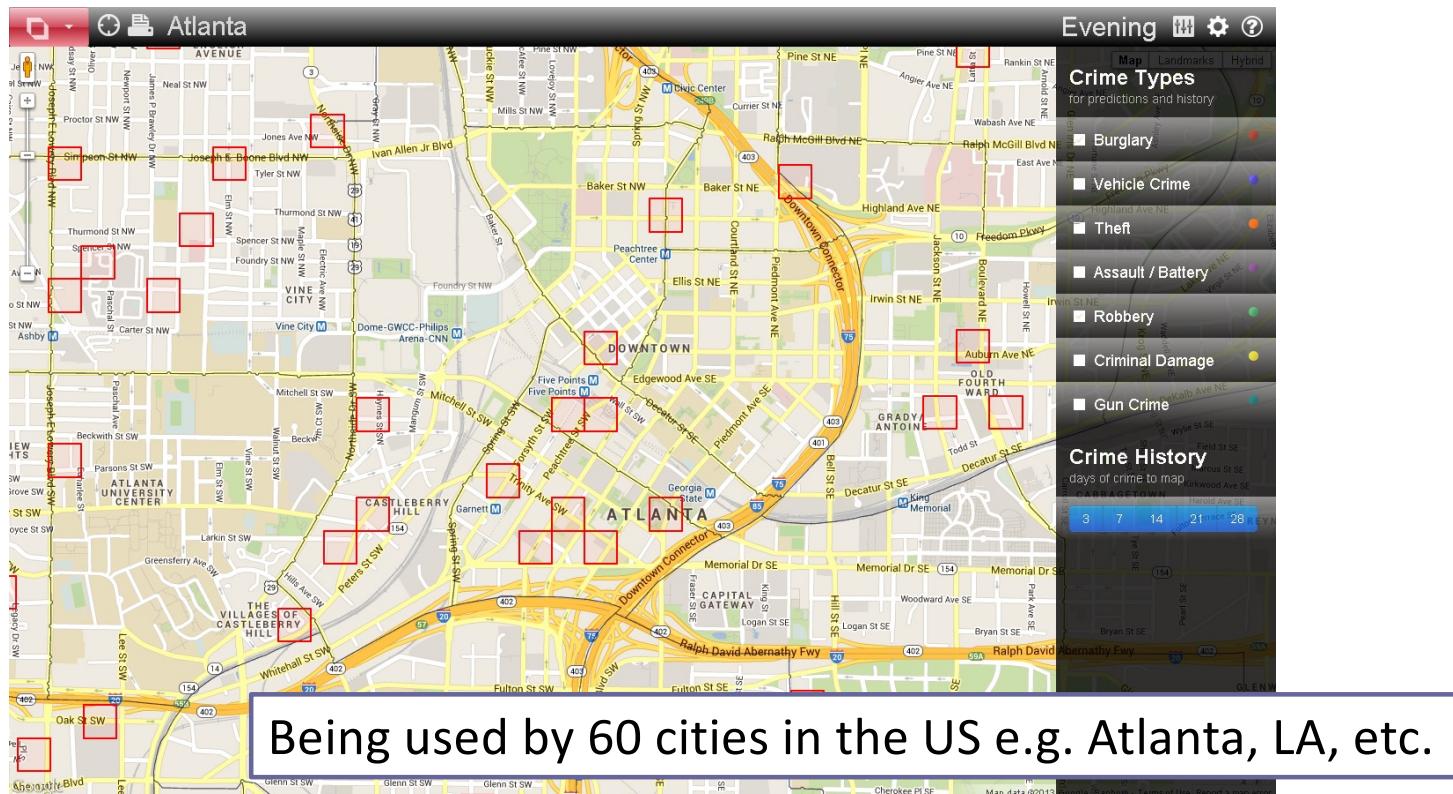
Case study: Alibaba Fraud Detection



Source: <http://www.sciencedirect.com/science/article/pii/S2405918815000021>



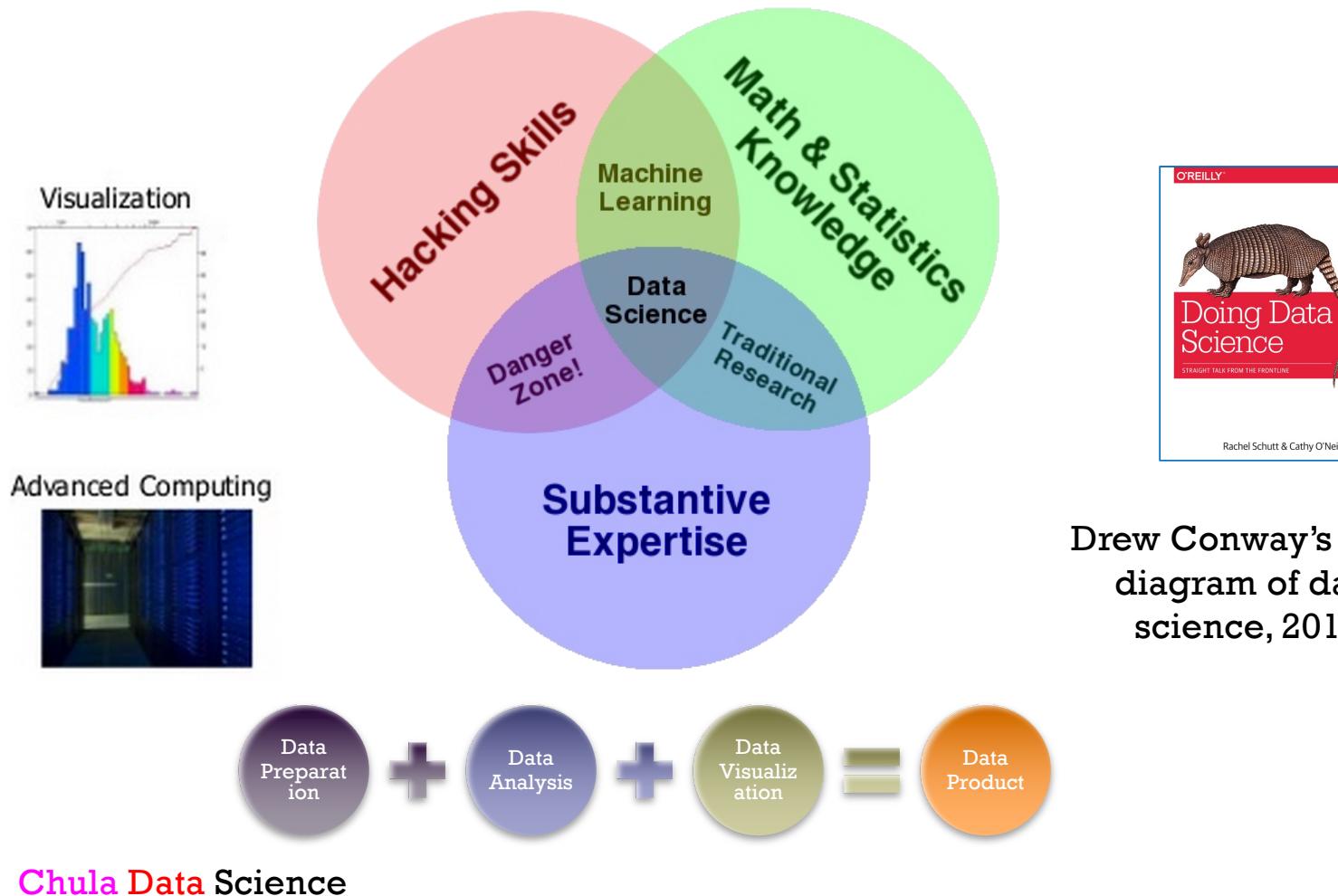
Case study: Predictive Policing



Source: <http://www.forbes.com/sites/ellenhuet/2015/02/11/predpol-predictive-police/>



Drew Conway's Data Science Venn diagram (Skills)



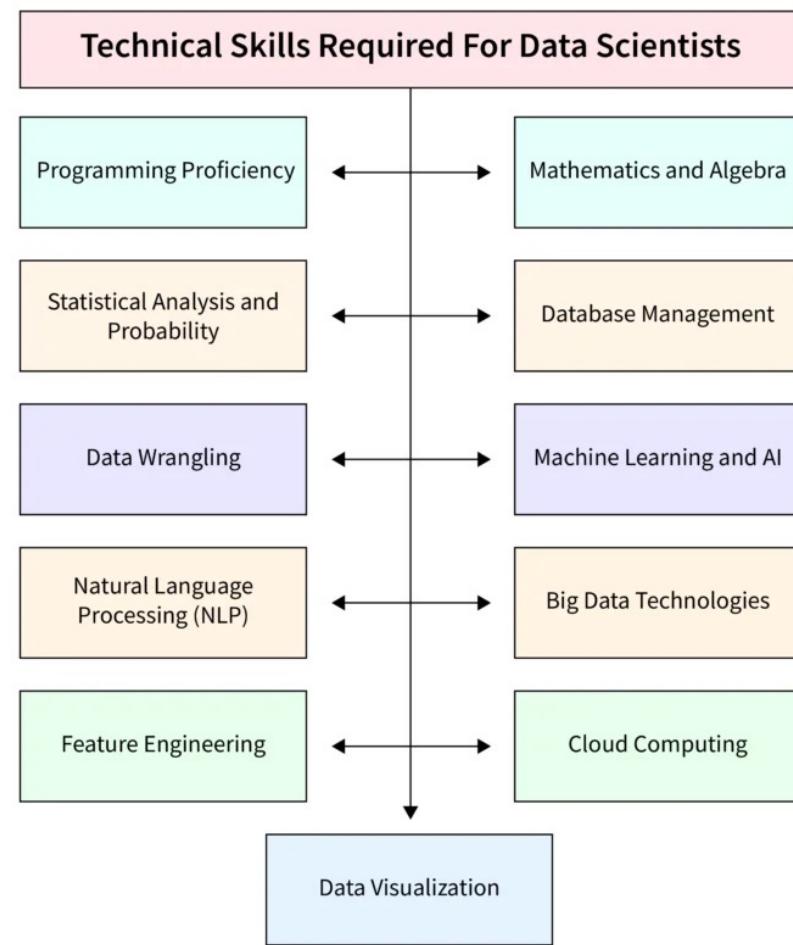
Top 20 Data Scientist Skills You Need in 2025



X in gs ⌂ ... 24 Min Read

Written by: **SCALER TEAM**

Last updated: December 18, 2024 7:49 pm



SCALER

<https://www.scaler.com/blog/data-scientist-skills/>



Key Data Science Activities

- Data Science Process
- Types of Data Science Projects
- AI/ML/DL/GenAI
- Data Engineering
- MLOps
- Cloud Technologies



Data Science Process

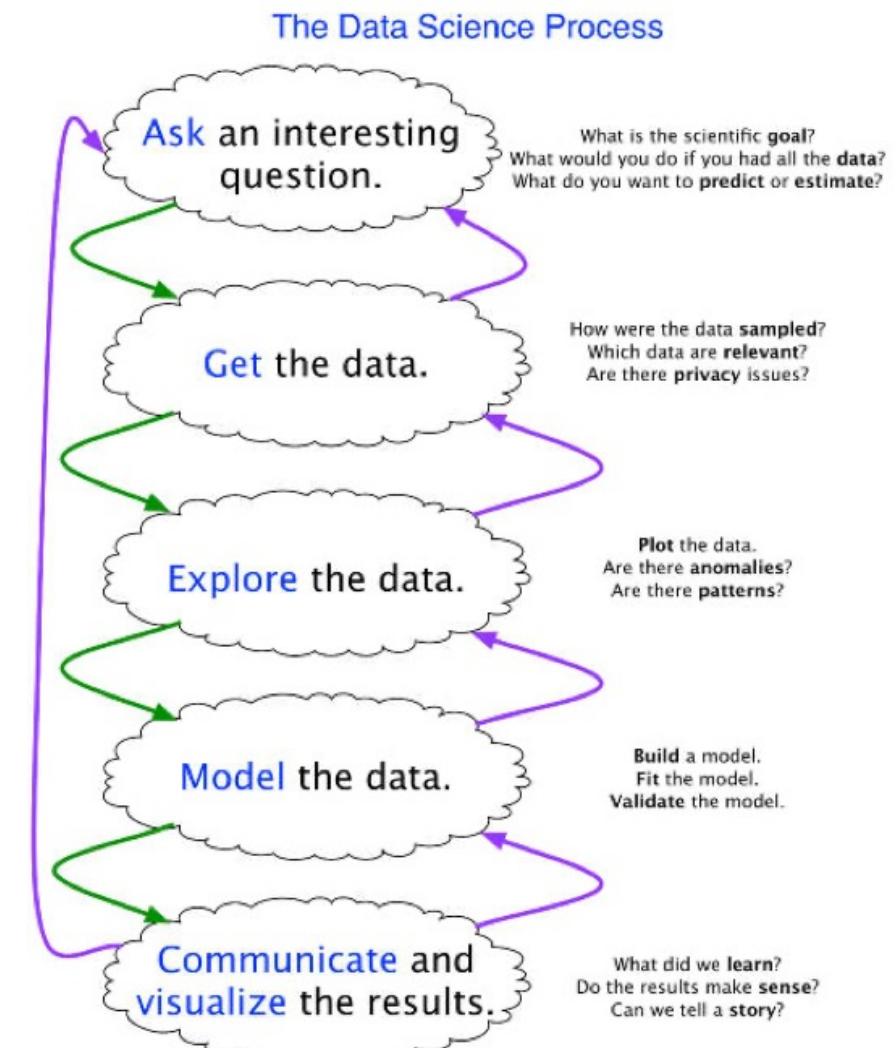


Dr. Virote

1. Transform data into **valuable insights**
2. Transform data into **data products**
3. Transform data into **interesting stories**

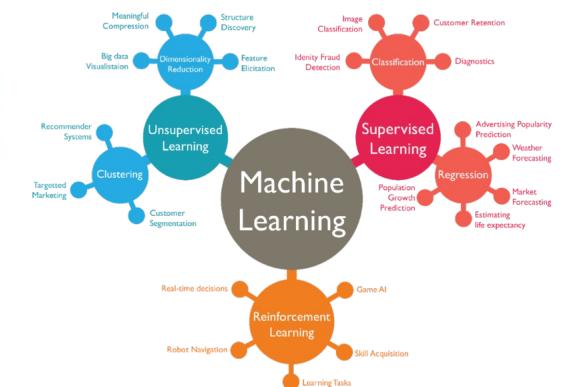
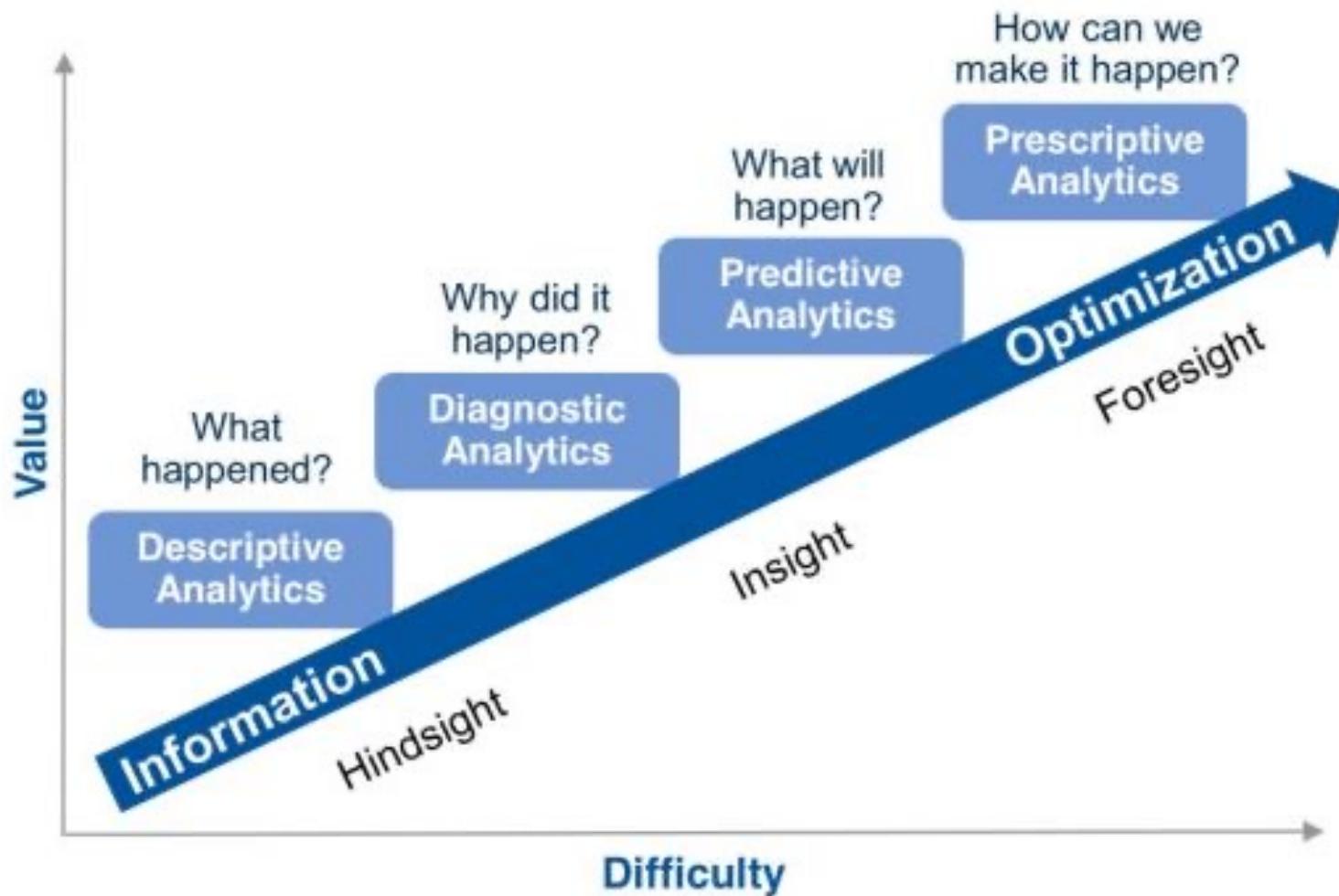
Aj.Natawut

1. Measurement (**decision**)
2. Insights (**knowledge**)
3. Data Products (**innovation, intelligent**)



Joe Blitzstein and Hanspeter Pfister, created for the Harvard data science course <http://cs109.org/>.

Data Analytics (Data Science)





Types of Data Science Projects

Valuable insights

- Data visualization
- Analytical skills & storytelling
- Infographic



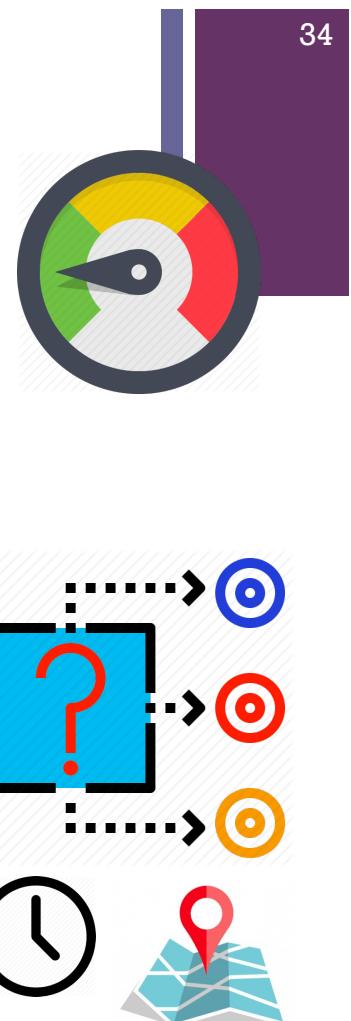
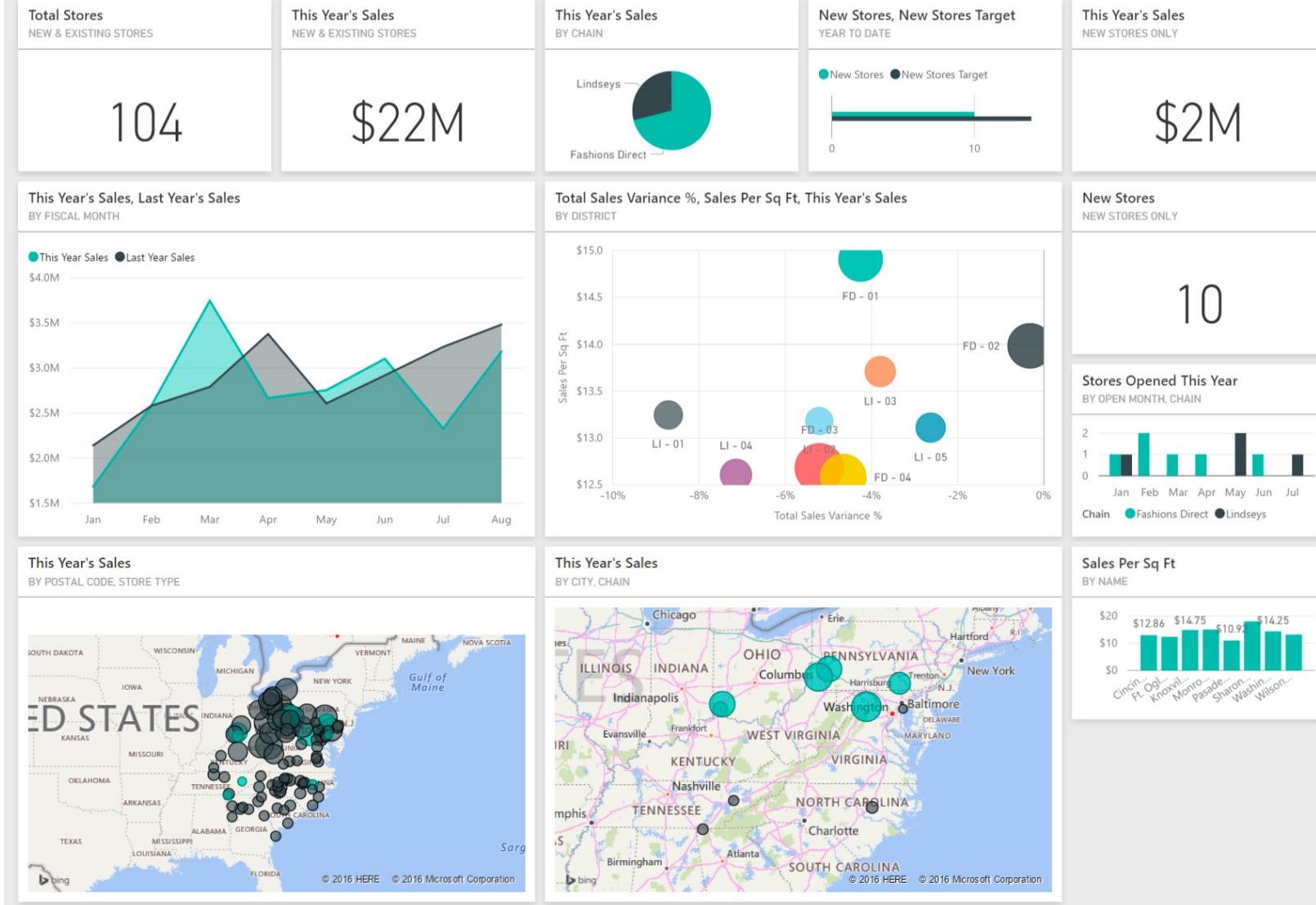
Advanced analytics

- AI/Machine Learning/Deep Learning
- Prediction, Forecasting, Clustering, etc.



Retail Analysis Sample

Ask a question about your data



2025 Gartner® Magic Quadrant™ for Analytics and Business Intelligence Platforms



<https://wwwqlik.com/us/gartner-magic-quadrant-business-intelligence>

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms

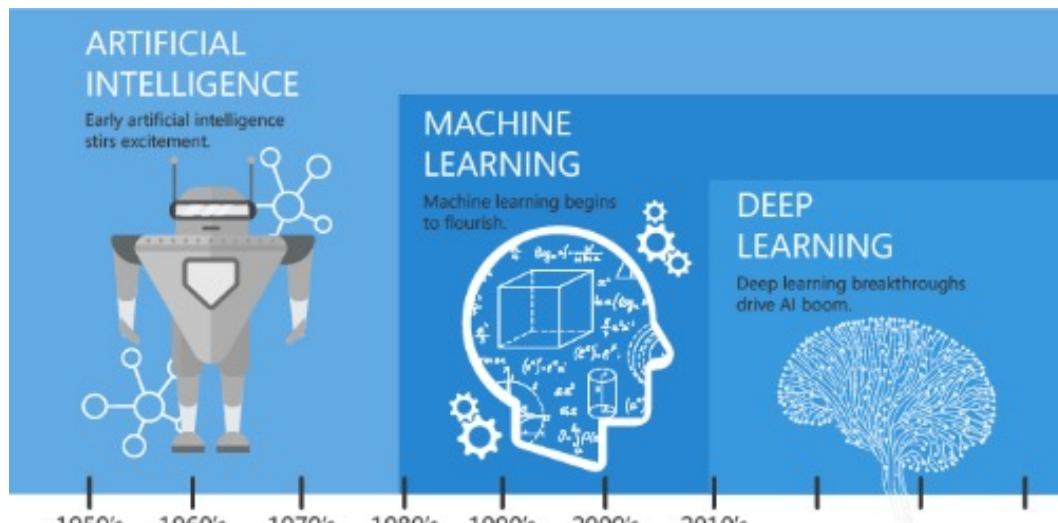


Gartner

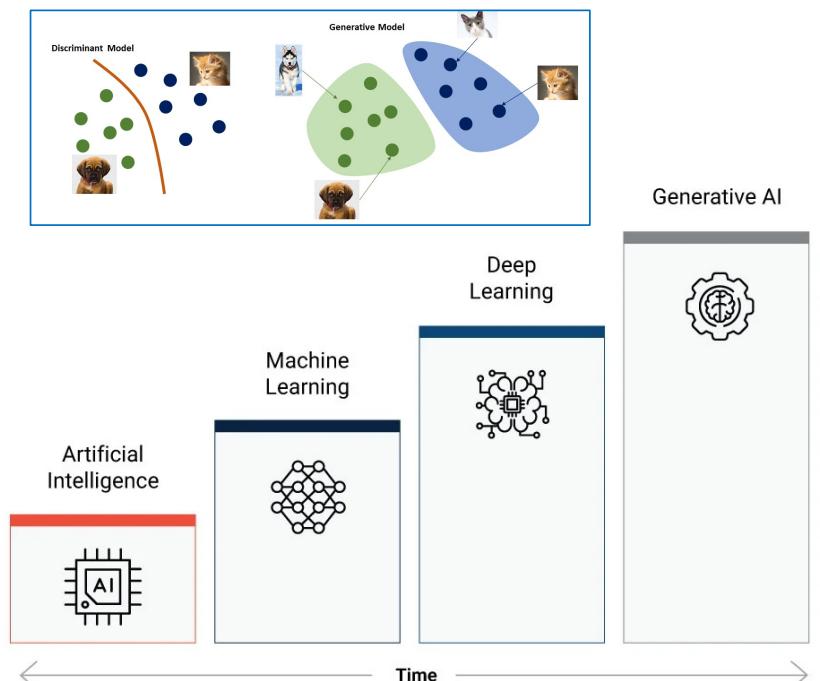
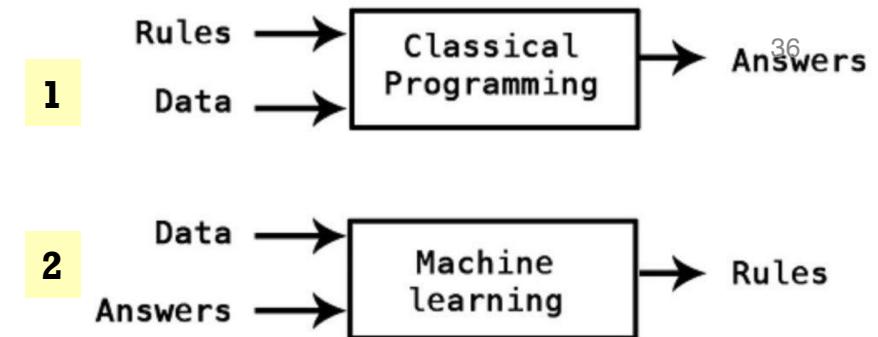


AI = Automation

- 1) Rule-based AI
- 2) Machine Learning (ML)



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.



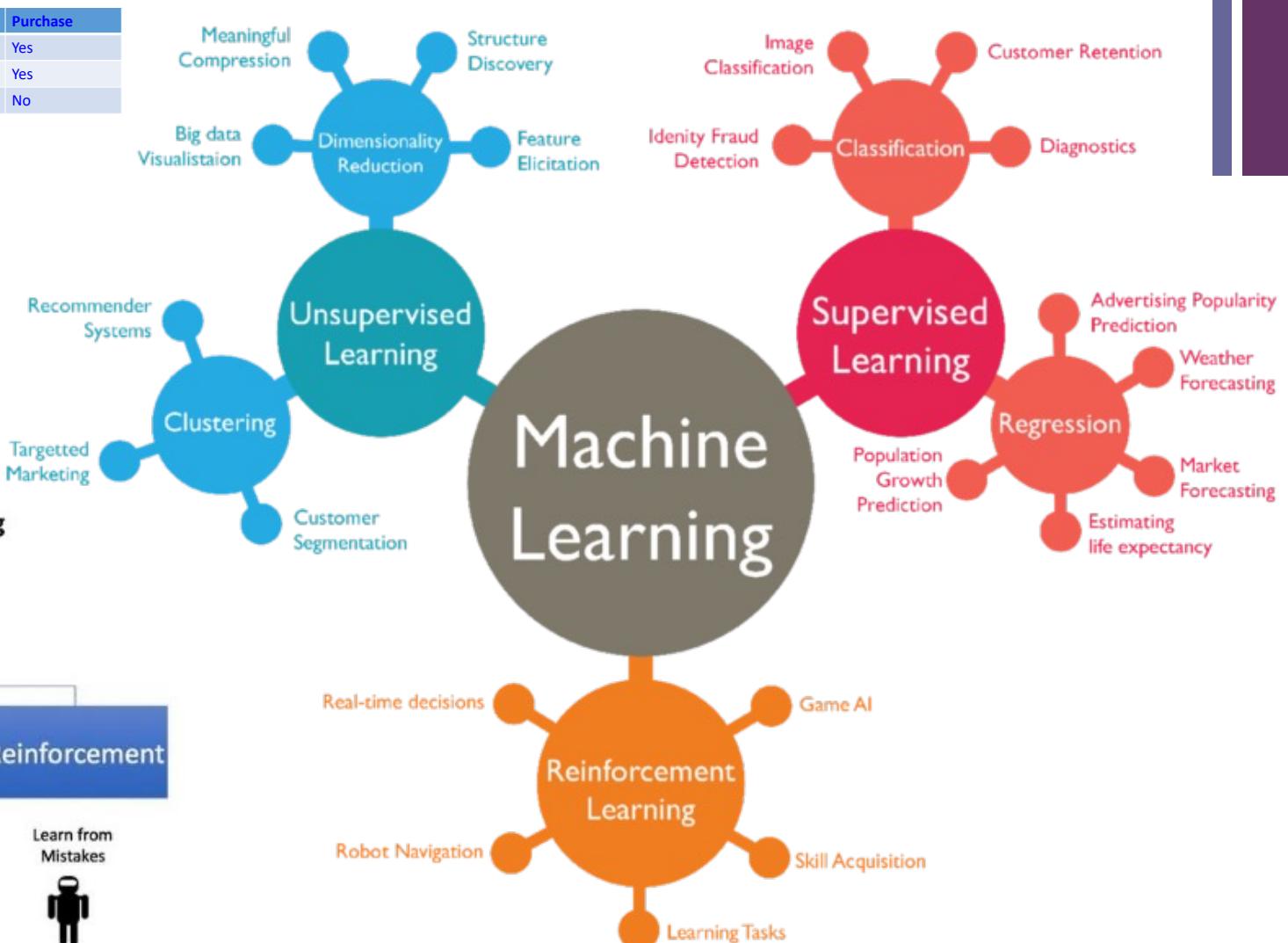
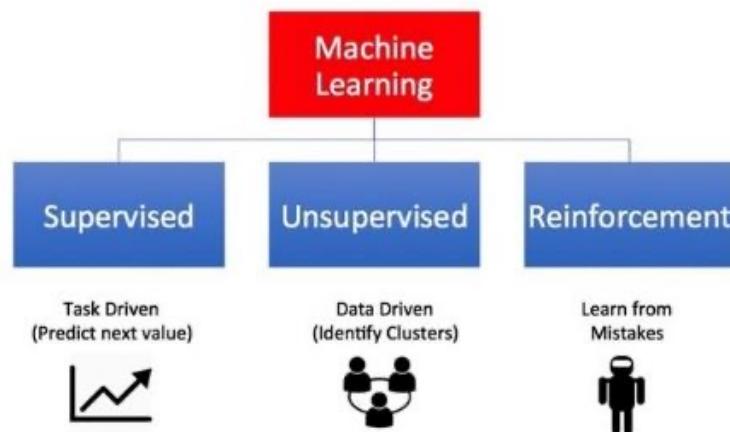
<https://mc.ai/machine-learning-basics-artificial-intelligence-machine-learning-and-deep-learning/>

+ Machine Learning (ML)

37

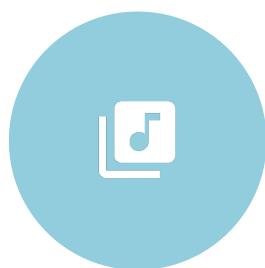
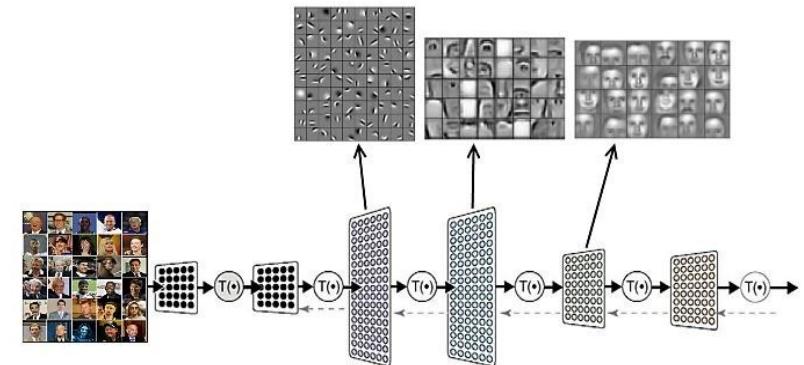
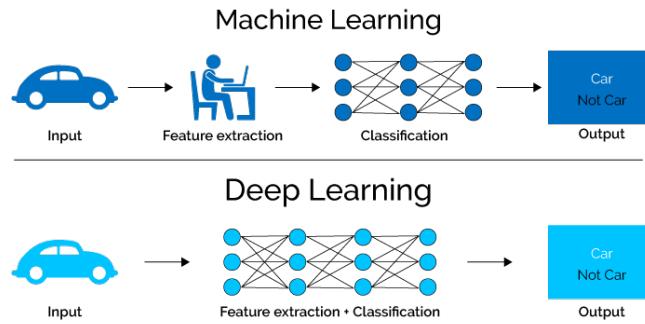
Age	Income	Gender	Province	Purchase
25	25,000	Female	Bangkok	Yes
35	50,000	Female	Nontaburi	Yes
32	35,000	Male	Bangkok	No

Types of Machine Learning





Deep Learning (DL)



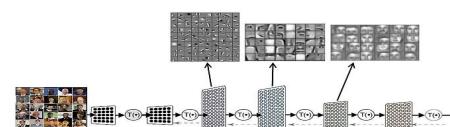
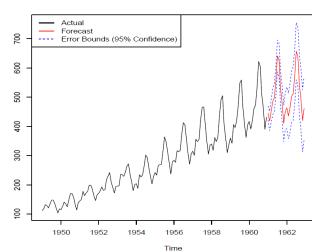
Speech
Recognition



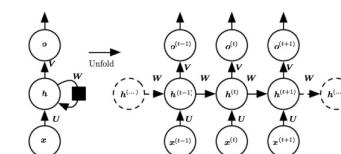
Computer
Vision



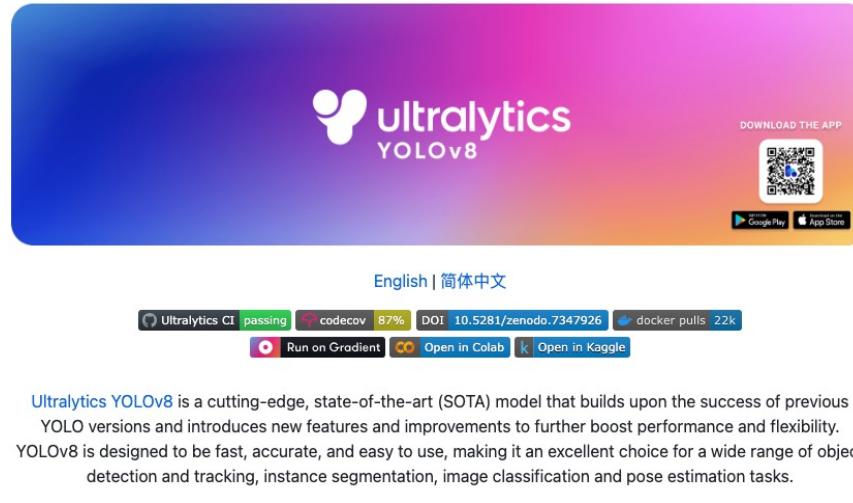
Natural
Language
Processing



CNN



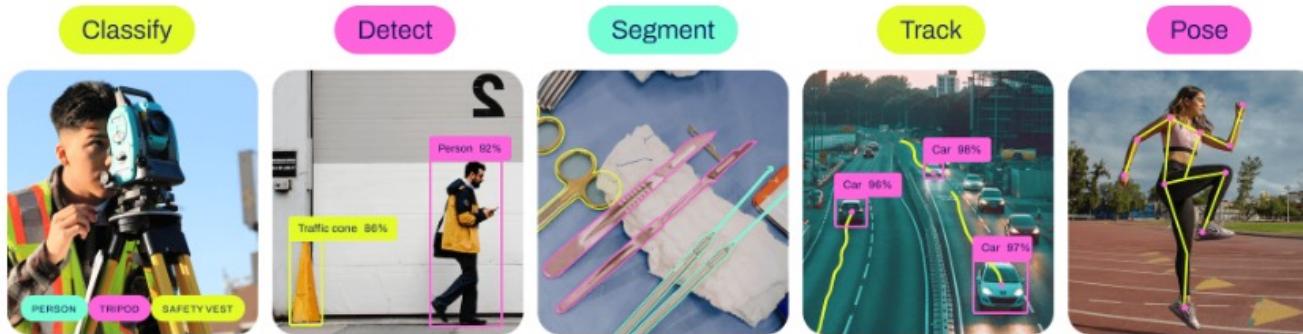
RNN (LSTM)



The screenshot shows the Ultralytics YOLOv8 homepage. At the top, there's a logo with three white shapes and the text "ultralytics YOLOv8". To the right, there's a QR code with the text "DOWNLOAD THE APP" above it, and links for "Google Play" and "App Store". Below the logo, there are language options "English | 简体中文". Underneath, there are several status indicators: "Ultralytics CI passing", "codecov 87%", "DOI 10.5281/zenodo.7347926", "docker pulls 22k", "Run on Gradient", "Open in Colab", and "Open in Kaggle". A descriptive text block below reads: "Ultralytics YOLOv8 is a cutting-edge, state-of-the-art (SOTA) model that builds upon the success of previous YOLO versions and introduces new features and improvements to further boost performance and flexibility. YOLOv8 is designed to be fast, accurate, and easy to use, making it an excellent choice for a wide range of object detection and tracking, instance segmentation, image classification and pose estimation tasks.".

Models

YOLOv8 [Detect](#), [Segment](#) and [Pose](#) models pretrained on the [COCO](#) dataset are available here, as well as YOLOv8 [Classify](#) models pretrained on the [ImageNet](#) dataset. [Track](#) mode is available for all Detect, Segment and Pose models.

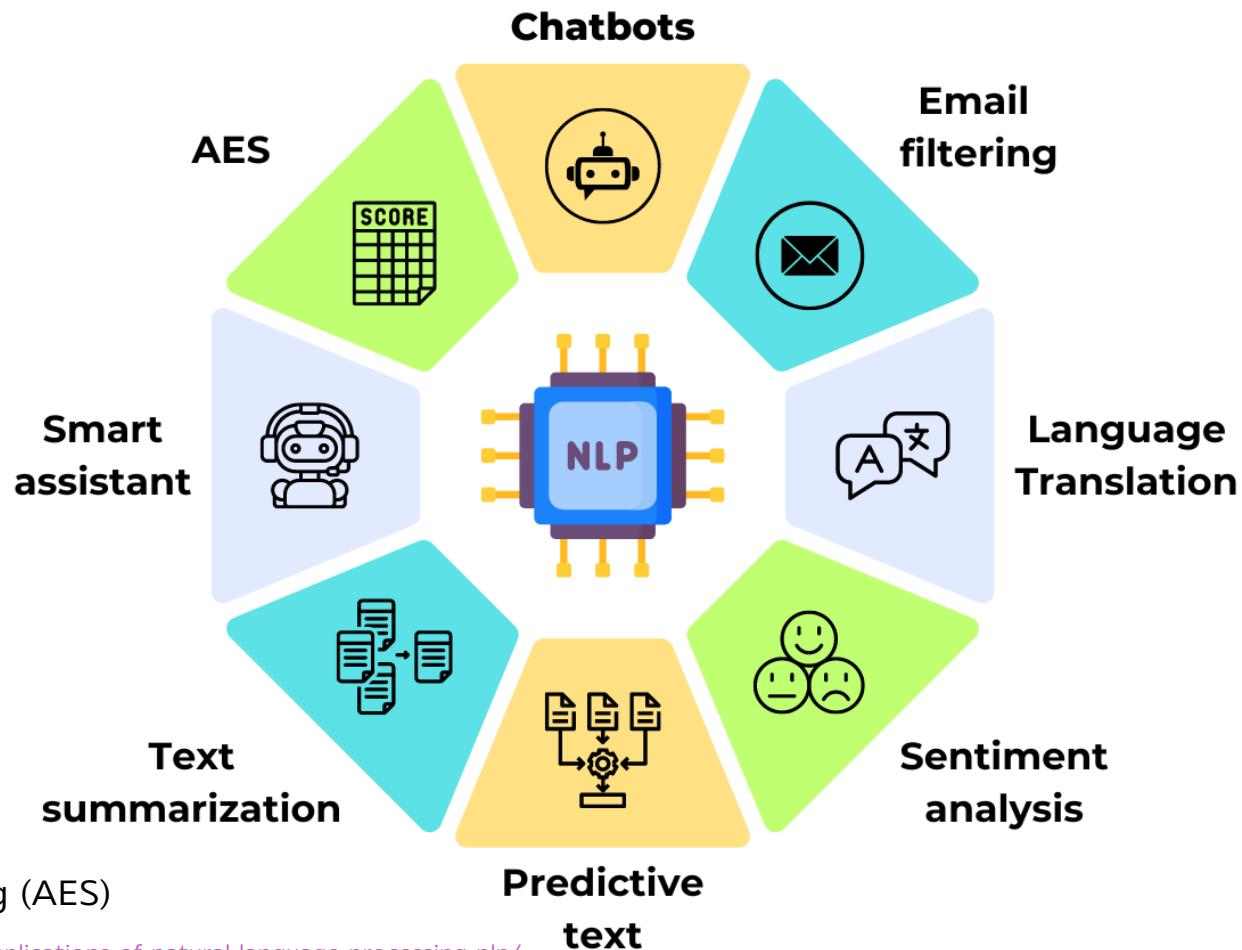


All [Models](#) download automatically from the latest Ultralytics [release](#) on first use.

<https://github.com/ultralytics/ultralytics>

Top 8 Applications of Natural Language Processing (NLP)

Applications of Natural Language Processing

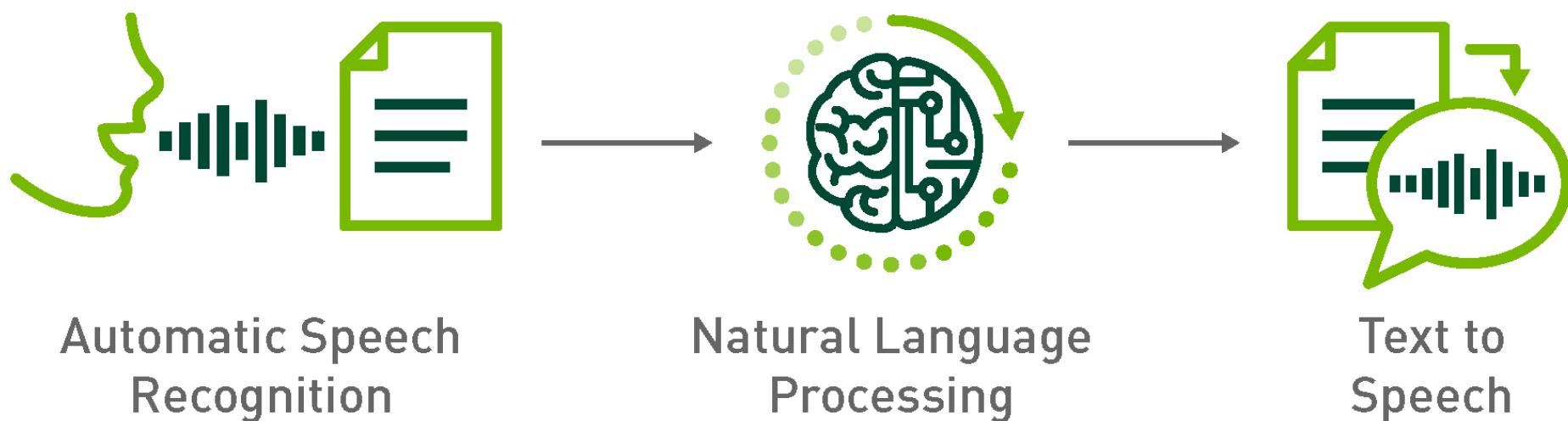


Automated Essay Scoring (AES)

<https://eastgate-software.com/top-8-applications-of-natural-language-processing-nlp/>



Automatic Speech Recognition (ASR) & Text-to-Speech (TTS)





ChatGPT Surges in Popularity: Now Boasting Over 200 Million Weekly Users

Onome · September 2, 2024

AI NEWS 3 MINS READ

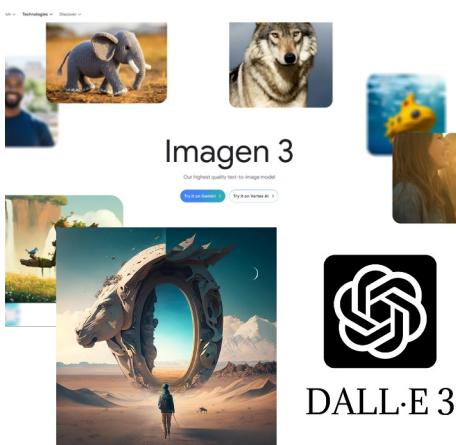


<https://autogpt.net/chatgpt-surges-in-popularity-now-boasting-over-200-million-weekly-users/>



Generative AI

1) Image

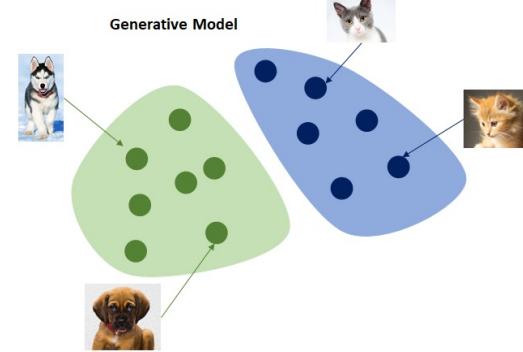
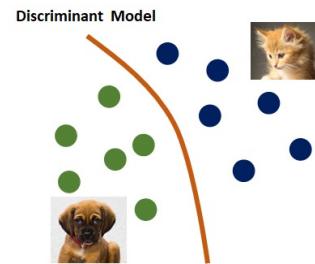


2) Text



 Claude

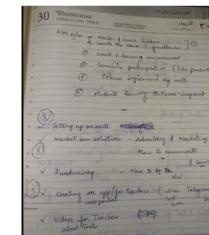
 Gemini
 deepseek



3) Video



4) Multimodal



Wednesday, 30th

• HM Roles:

1. Mentor & coach, provide feedback inside the class.
2. Create a learning environment.
3. Encourage community participation (like parents).
4. Ensure policies are implemented well.
5. Improve students' learning outcomes.

• Tasks/Goals:

- * Set up a website.
- * Market the solution – focus on advertising & marketing.
- * Understand how to communicate – possibly with the tool 'Canva'.
- * Fundraising: Explore ways to tap new resources.

Jobs of the Future: AI Prompt Engineer



Cody W Burns

Emerging Technology Visionary | Distributed Systems | Privacy | Executive Leadership

10 articles

+ Follow

October 19, 2022

JOB OF THE FUTURE: AI PROMPT ENGINEER

Cody Burns
44

<https://www.linkedin.com/pulse/jobs-future-ai-prompt-engineer-cody-w-burns/>

+ Types of **Data** Science Projects (recap)

If we don't have data, can we still perform these tasks?

Valuable insights

- Data visualization
- Analytical skills & storytelling
- Infographic



Advanced analytics

- AI/Machine Learning/Deep Learning
- Prediction, Forecasting, Clustering, etc.





Data engineering: data acquisition (e.g., SQL, web scraping), Cleansing, and Storing

Steps of The Data Refinement Pipeline

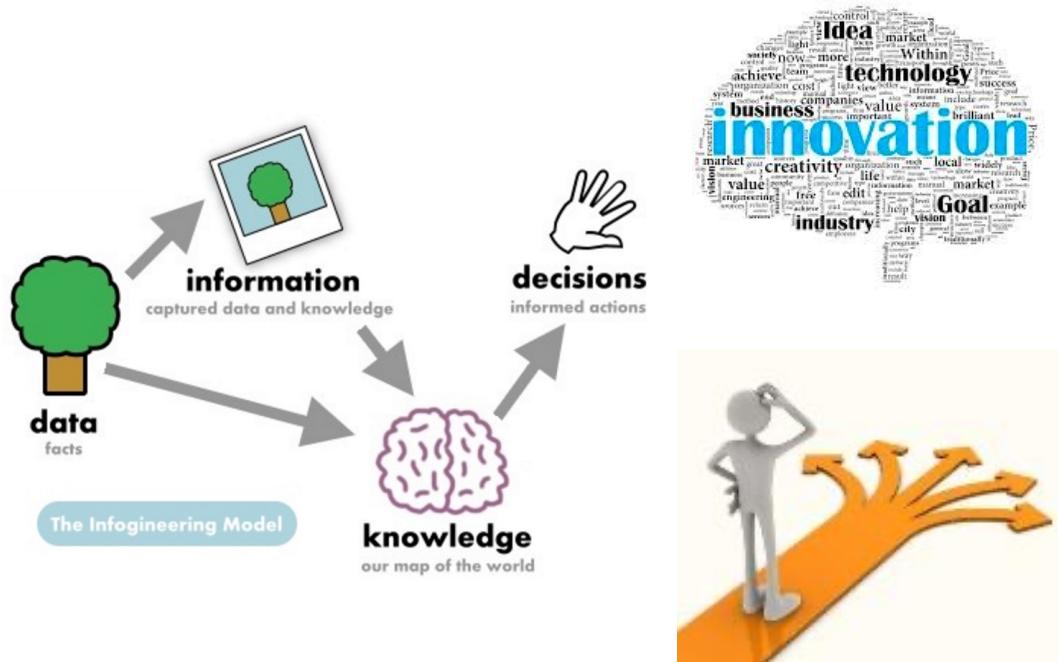


Data Pipeline

invgate.com

Big Data Analytics

- It is a process of examining **Big Data** to uncover useful information and knowledge.
- More data means better decision!



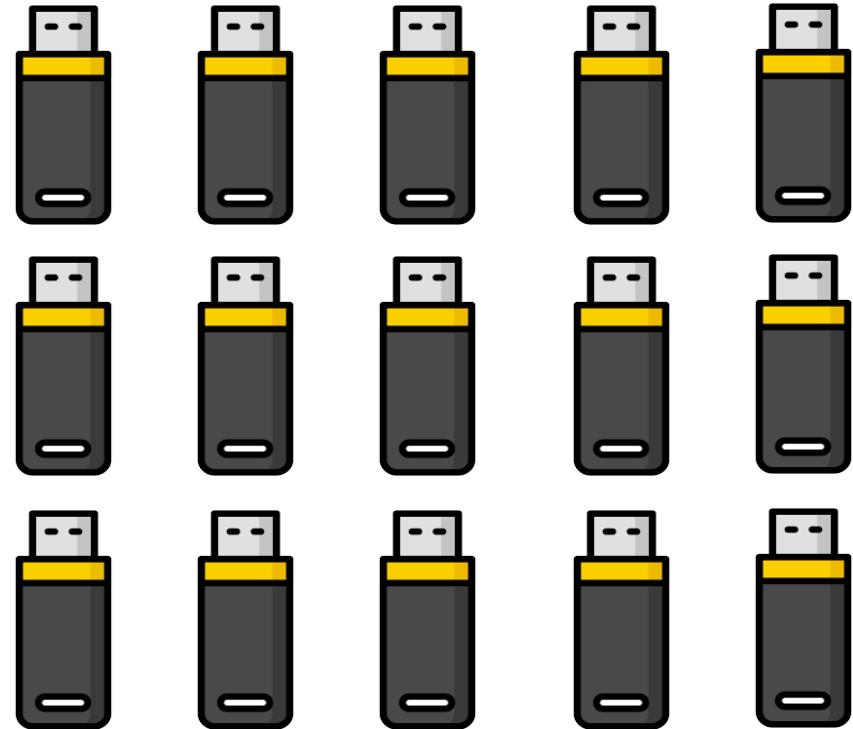
Big Data Challenges

Same tasks, but much more difficult!

2MB



200TB



Big Data Solution



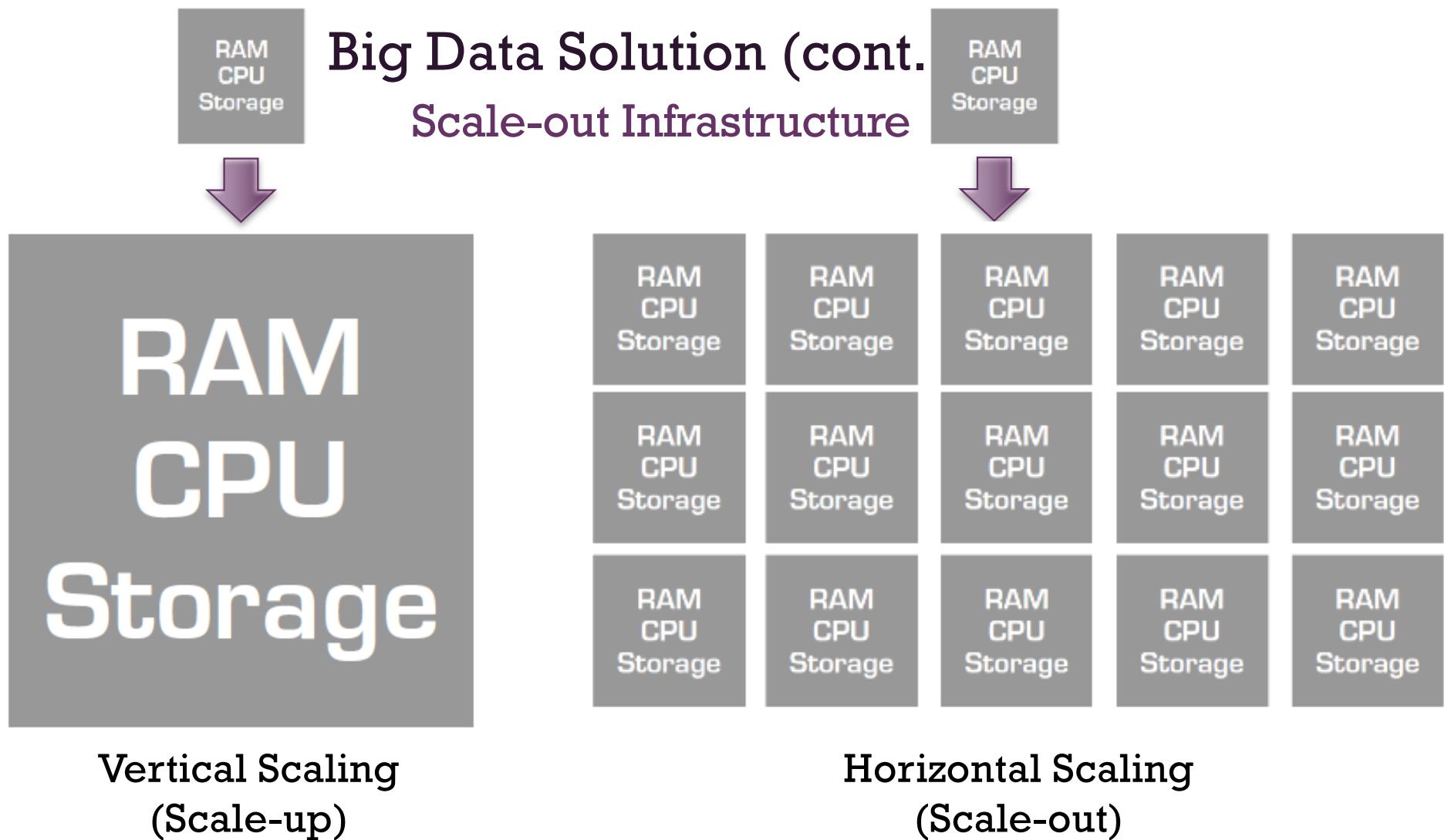
INFRASTRUCTURE



ALGORITHM

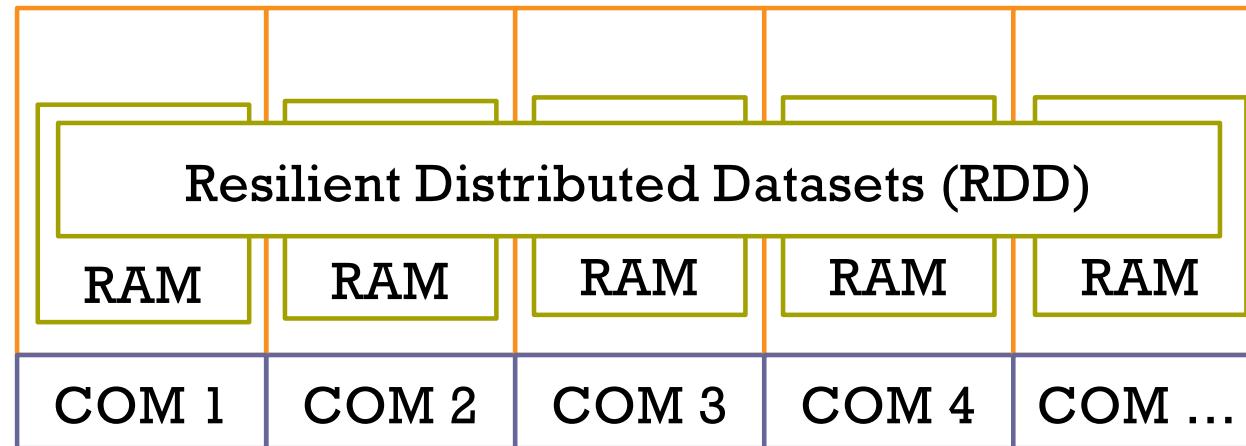
Big Data Solution (cont.)

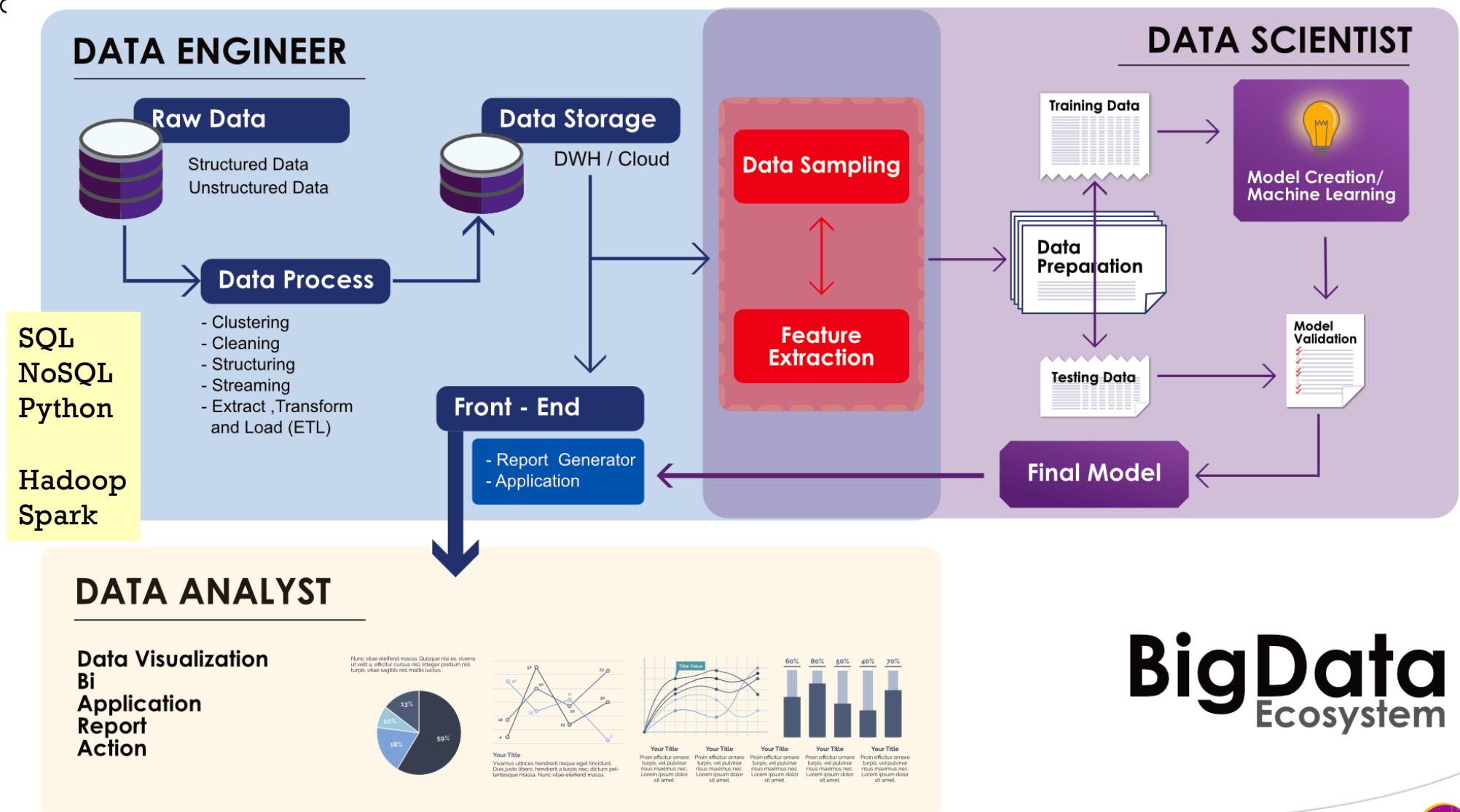
Scale-out Infrastructure



Big Data Solution (cont.)

In-memory & Distributed Computing





BigData Ecosystem

[LINK](#)



Business Intelligence By Coraline

<https://blog.datath.com/data-engineer-guide/>



Top Chef Thailand ตอนสุดท้าย ที่ผู้เข้าแข่งขันต้องช่วยกันทำงานเป็นทีม – ขอบคุณรูปจาก one31

Data Engineer ก็เหมือนกับผู้ช่วยเชฟ มีหน้าที่จัดเตรียมข้อมูลจากแหล่งต่าง ๆ มารวมกันไว้ในจุดเดียว โดยต้องทำให้ข้อมูลมีความถูกต้อง และดูแลระบบว่าทำงานได้ไม่เกิดปัญหาอะไร (ในชีวิตจริงนี่ต่อให้เราวางแผนมาดีแค่ไหน เจอข้อมูลเยอะ ๆ วันเดี๋ยวนี้ก็ล้มได้ครับ T_T)

+ We have the data and the ML model.

So, are we all set? Or is there more to consider?

Do you think that user can use this code to get the prediction result? NO!!!

CO 2_Linear-Regression-v2.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive

```
[ ] 1 coeff_df = pd.DataFrame(lm.coef_,lm.feature_names_in_,columns=['Coefficient'])  
2 coeff_df
```

Does this make sense? Probably not because I made up this data. If you want real data to repeat this sort of analysis, check out the [boston dataset](#):

```
from sklearn.datasets import load_boston  
boston = load_boston()  
print(boston.DESCR)  
boston_df = boston.data
```

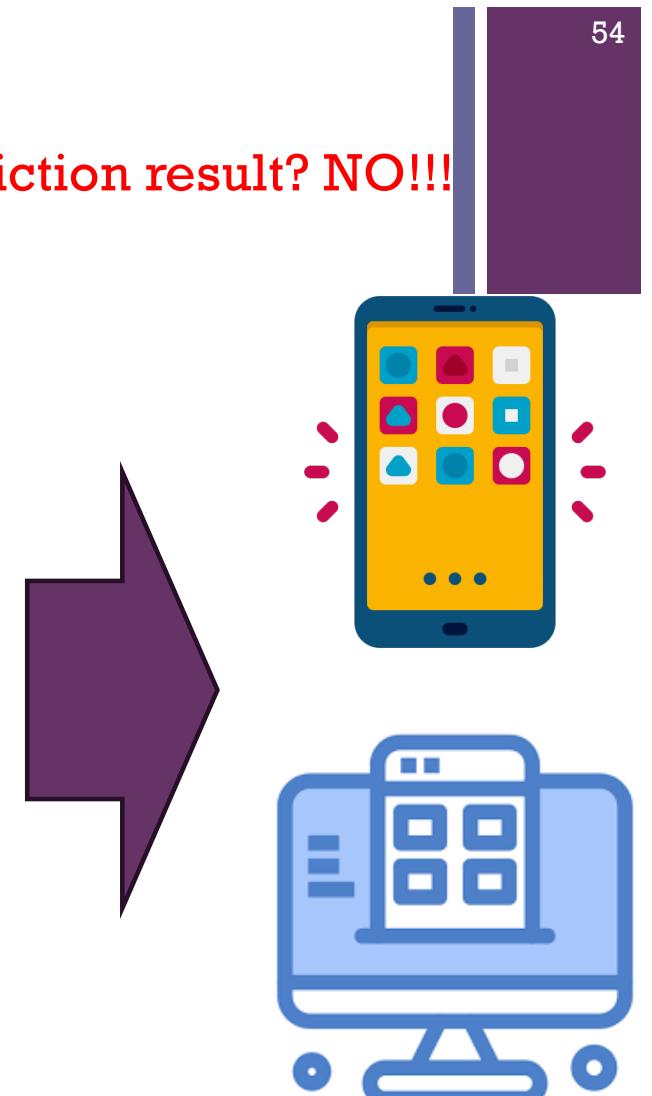
▼ Predictions from our Model

Let's grab predictions off our test set and see how well it did!

```
[ ] 1 predictions = lm.predict(X_test)  
[ ] 1 plt.scatter(y_test,predictions)
```

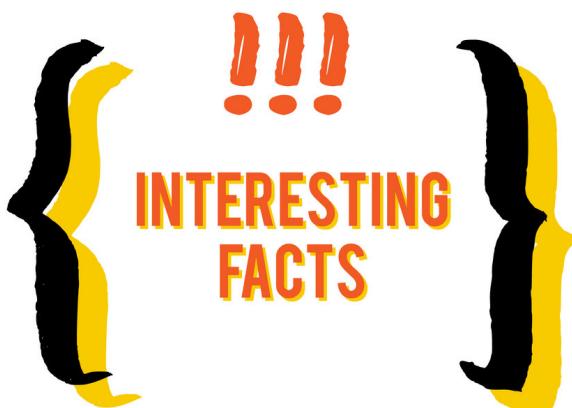
Residual Histogram

```
[ ] 1 sns.distplot(y_test-predictions,bins=50);
```





Interesting facts



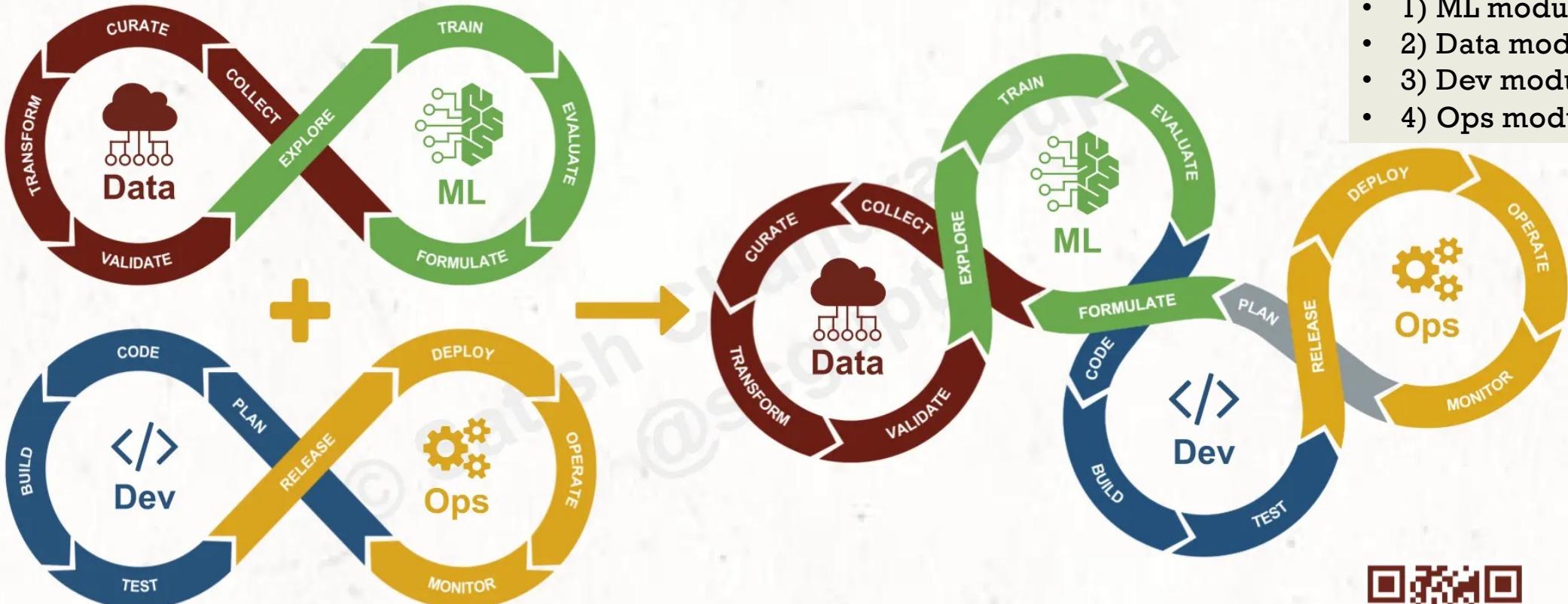
- More than 50% of AI projects were **FAILED** since they didn't plan about the deployment.
- | For the AI project, finish building the model is only 50% of the work.
- | The remaining work is about the deployment as a touchpoint to the target user.

MLOps = DataML + DevOps

ml4devs.com/mlops-lifecycle



- 1) ML module
- 2) Data module
- 3) Dev module
- 4) Ops module



© 2022 Satish Chandra Gupta



CC BY-NC-ND 4.0 International License

<https://www.ml4devs.com/Images/Illustrations/ml-lifecycle-fusing-model-and-software-development.webp>

scgupta.me
twitter.com/scgupta
linkedin.com/in/scgupta



Data Scientist + ML Engineer



Data Scientist

Datamites
Global Institute for Data Science

57

VS

Data Engineer



VS

ML Engineer



VS

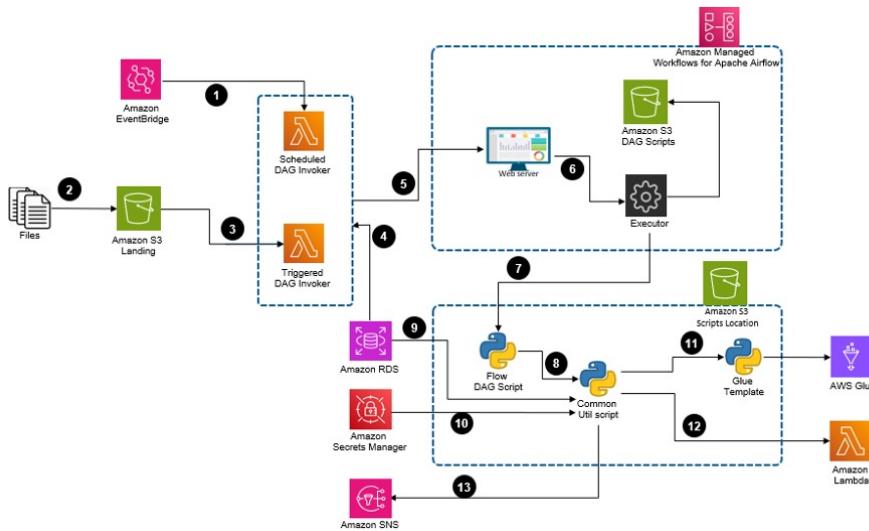
MLOps Engineer



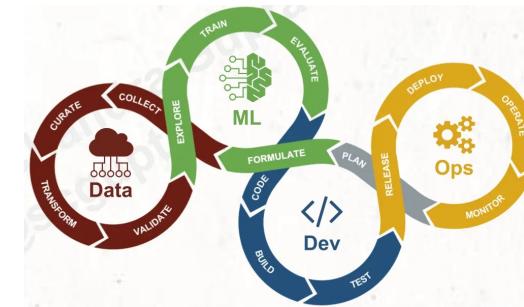
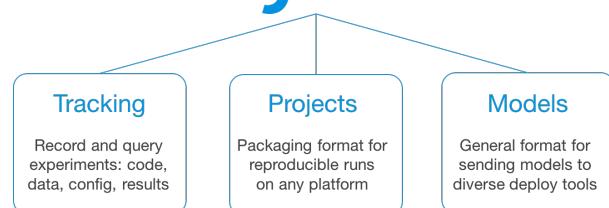
<https://vocal.media/education/data-scientist-vs-data-engineer-vs-ml-engineer-vs-ml-ops-engineer>
www.datamites.com



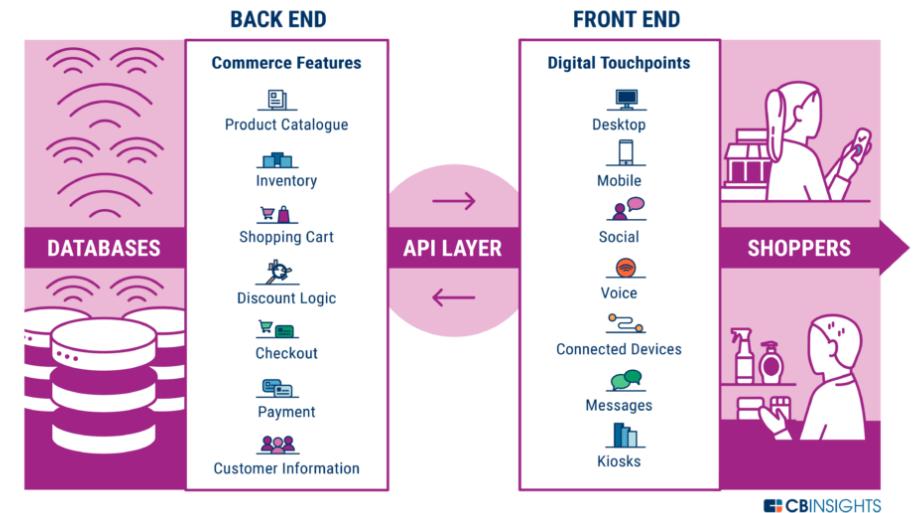
Many MLOps tasks: Deployment, Automation, Integration, Monitoring



mlflow

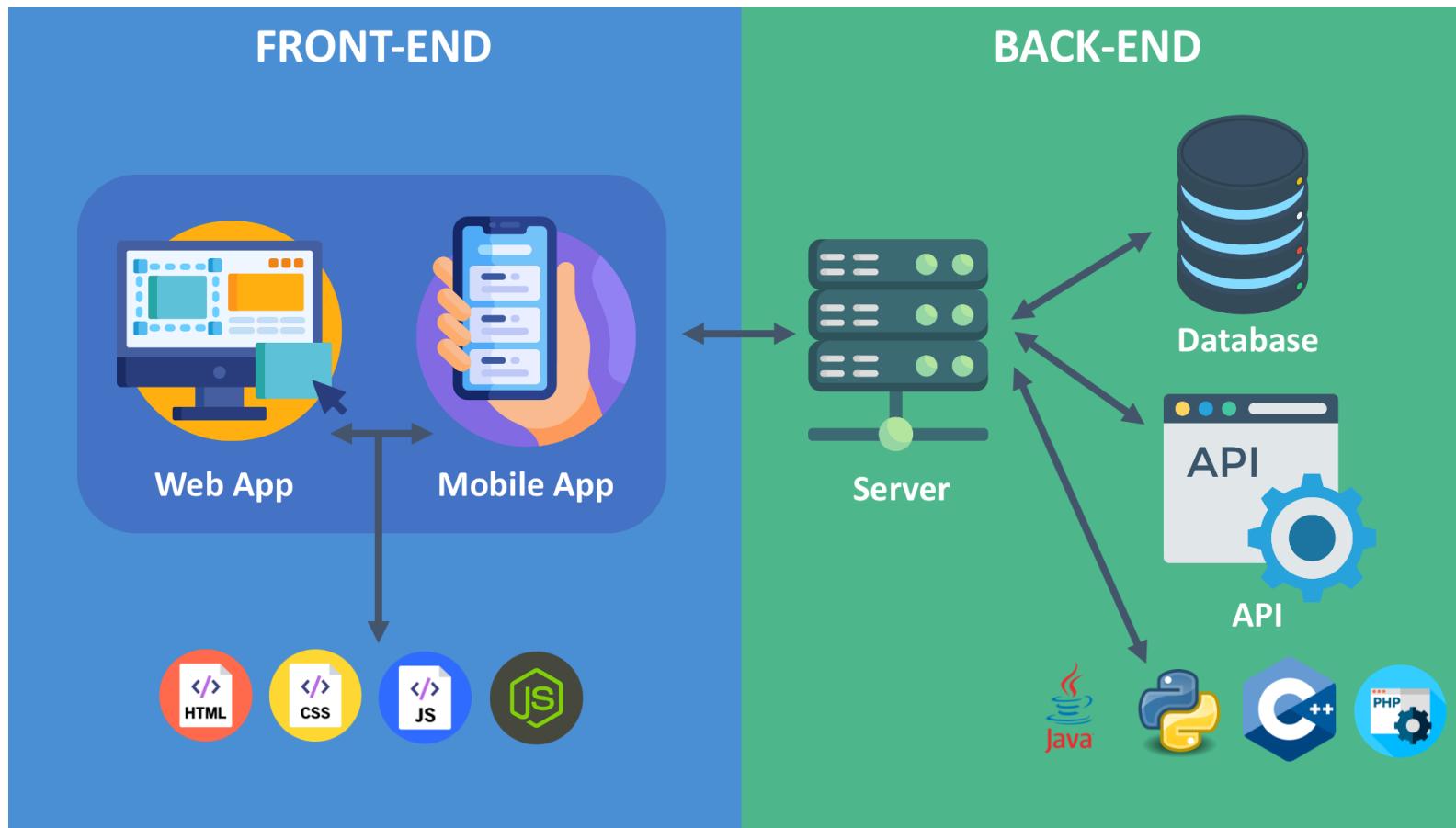


Components of a headless commerce architecture





Where should you develop and deploy your model or application? **Do you need to set up all the servers yourself?**



Google is a Leader in Gartner Magic Quadrant for Strategic Cloud Platform Services

October 24, 2024

Brad Calder
VP & GM, Google Cloud Platform

Amin Vahdat
VP/GM, ML, Systems, and Cloud AI



Figure 1: Magic Quadrant for Strategic Cloud Platform Services



Categories of AWS services



Vit Niennattrakul, Ph.D.



Analytics



Application
Integration



AR and VR



Blockchain



Business
Applications



Compute



Cost
Management



Customer
Engagement



Database



Developer Tools



End User
Computing



Game Tech



Internet
of Things



Machine
Learning



Management and
Governance



Media Services



Migration and
Transfer



Mobile



Networking and
Content Delivery
network



Robotics



Satellite



Security, Identity, and
Compliance



Storage

This document is protected under the copyright laws of Thailand and contains information that is proprietary and confidential to DailiTech, which shall not be disclosed outside the recipient's company or duplicated, used, or disclosed in whole or in part by the recipient for any purpose other than to the purpose of this document. Any other use or disclosures in whole or in part of this information without the express written permission of DailiTech are prohibited.

Data preparation for AI / ML and data science

Right



Vit Niennattrakul, Ph.D.



Amazon Forecast



Amazon Comprehend



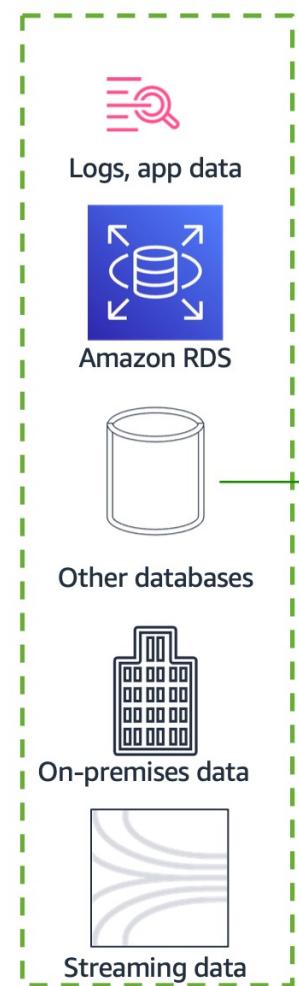
Amazon SageMaker



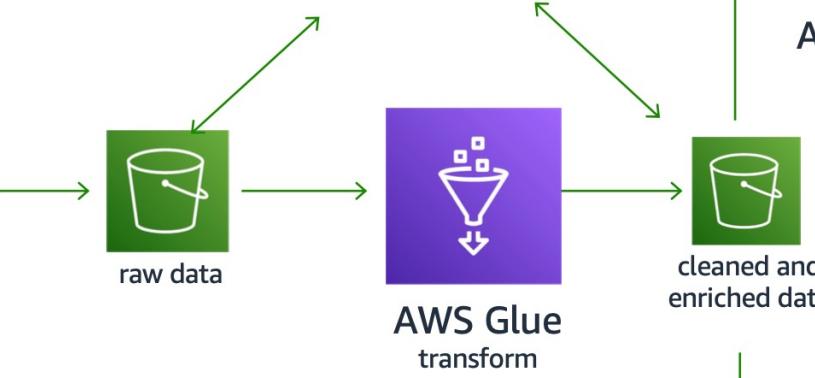
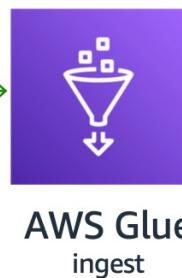
Amazon Rekognition



Amazon Lex



Notebooks:
data exploration,
experimentation



AWS Academy Service

AWS Academy Learner Lab

- Amazon API Gateway
- AWS App Mesh
- Application Auto Scaling
- AWS AppSync
- Amazon Athena
- Amazon Aurora
- AWS Backup
- AWS Certificate Manager (ACM)
- AWS Batch
- AWS Cloud9
- AWS CloudFormation
- Amazon CloudFront
- Amazon CloudSearch
- AWS CloudShell
- AWS CloudTrail
- Amazon CloudWatch
- AWS CodeCommit
- AWS CodeDeploy
- Amazon CodeWhisperer
- AWS Config
- AWS Systems Manager (SSM)
- Amazon Textract
- AWS Cost and Usage Report
- AWS Cost Explorer
- AWS Data Pipeline
- AWS DeepComposer
- AWS DeepLens
- AWS DeepRacer
- AWS Directory Service
- Amazon EC2 Auto Scaling
- AWS Elastic Beanstalk
- Amazon Elastic Block Store (EBS)
- Amazon Elastic Container Registry (ECR)
- Amazon Elastic Container Service (ECS)
- Amazon Elastic File System (EFS)
- Amazon Elastic Inference
- Amazon Elastic Kubernetes Service (EKS)
- Elastic Load Balancing (ELB)
- Amazon Elastic MapReduce (EMR)
- Amazon ElastiCache
- Amazon EventBridge
- AWS Fargate
- Amazon Timestream
- AWS Trusted Advisor
- Amazon Forecast
- AWS Glue
- AWS Glue DataBrew
- Amazon GuardDuty
- AWS Health
- AWS Identity and Access Management (IAM)
- AWS IAM Access Analyzer
- Amazon Inspector
- AWS IoT 1-Click
- AWS IoT Analytics
- AWS IoT Core
- AWS IoT Greengrass
- Amazon Kendra
- AWS Key Management Service (KMS)
- Amazon Kinesis
- Amazon Lex
- Amazon Machine Learning (Amazon ML)
- AWS Marketplace Subscriptions
- AWS Mobile Hub
- Amazon Neptune
- Amazon Virtual Private Cloud (Amazon VPC)
- AWS WAF - Web Application Firewall
- AWS OpsWorks
- Amazon Personalize
- Amazon QuickSight
- Amazon Redshift
- Amazon Relational Database Service (RDS)
- AWS Resource Groups & Tag Editor
- AWS RoboMaker
- Amazon Route 53
- AWS Secrets Manager
- AWS Security Hub
- AWS Security Token Service (STS)
- AWS Serverless Application Repository (SAR)
- AWS Service Catalog
- Amazon Simple Notification Service (SNS)
- Amazon Simple Queue Service (SQS)
- Amazon Simple Storage Service (S3)
- Amazon Simple Storage Service Glacier (S3 Glacier)
- Amazon Simple Workflow Service (SWF)
- AWS Step Functions
- AWS Storage Gateway
- AWS Well-Architected Tool
- AWS X-Ray

AWS Academy Lab Project - Cloud Data Pipeline Builder

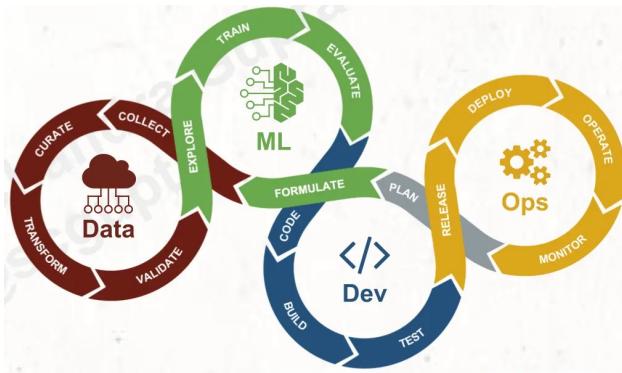
- Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Both Learner Lab & Lab Project - Cloud Data Pipeline Builder

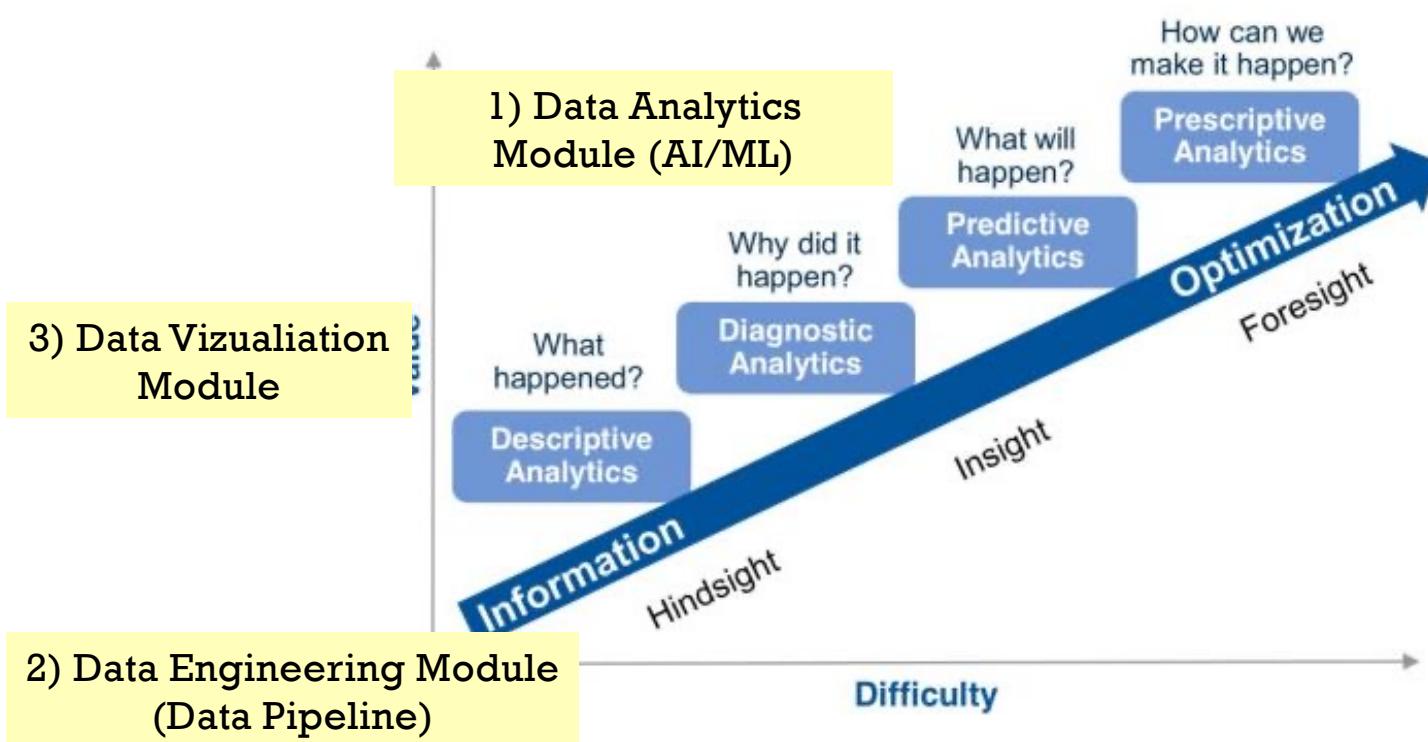
- Amazon SageMaker
- Amazon Elastic Compute Cloud (EC2)
- Amazon DynamoDB
- AWS Lambda
- Amazon Kinesis Video Streams
- Amazon Rekognition

+

Conclusion



Conclusion



4) Cloud technology



Blog > Career Advice > Data Analyst Job Outlook 2025 [Research On 1,000 Job Postings]

Data Analyst Job Outlook 2025 [Research on 1,000 Job Postings]

Join over 2 million students who advanced their careers with 365 Data Science. Learn from instructors who have worked at Meta, Spotify, Google, IKEA, Netflix, and Coca-Cola and master Python, SQL, Excel, machine learning, data analysis, AI fundamentals, and more.

[Start for Free](#)

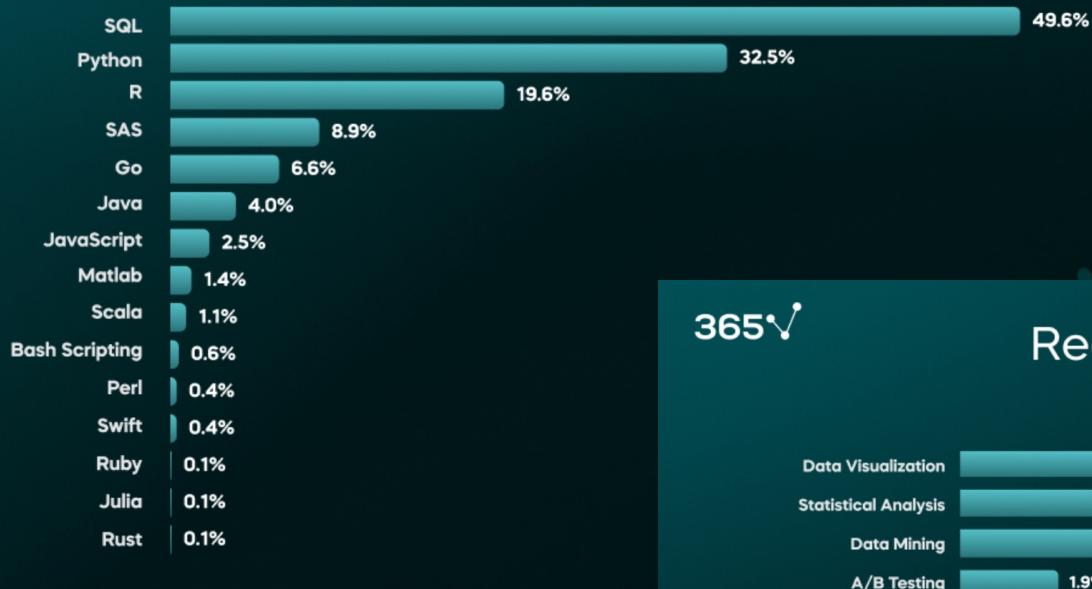
Sophie Magnet • 8 Apr 2025 • 21 min read

<https://365datasience.com/career-advice/data-analyst-job-outlook-2025/>



Required Programming Languages

Data Analysts 2025

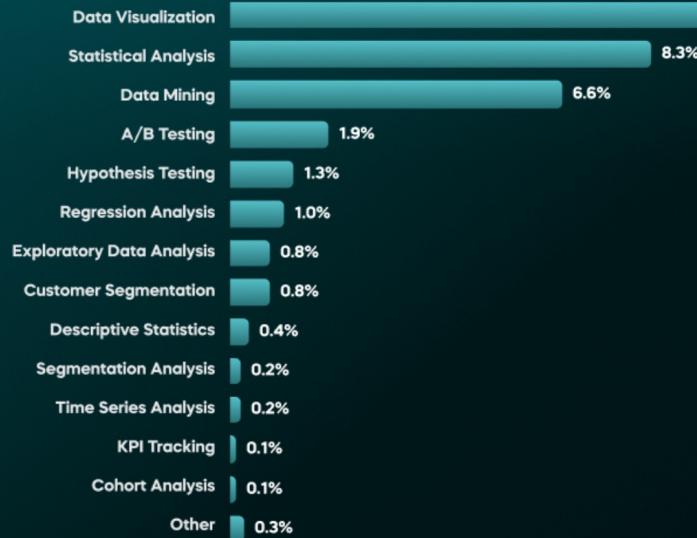


■ SQL

■ Python

Required Data Analysis Skills

Data Analysts 2025



■ Data visualization

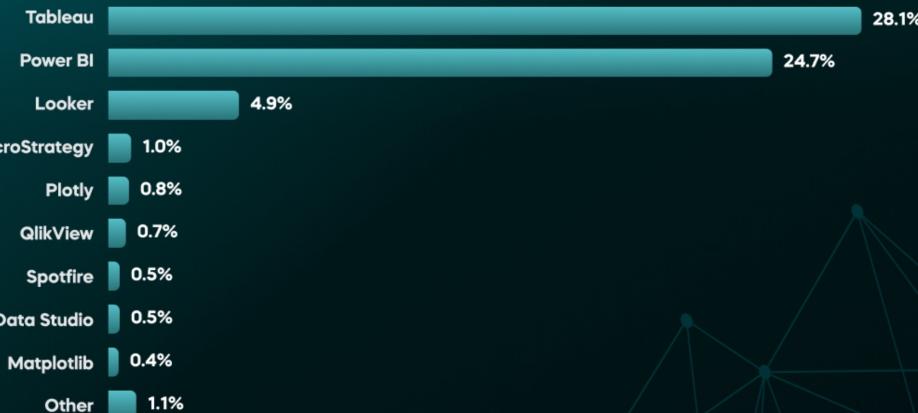
■ Statistical analysis

■ AI/ML

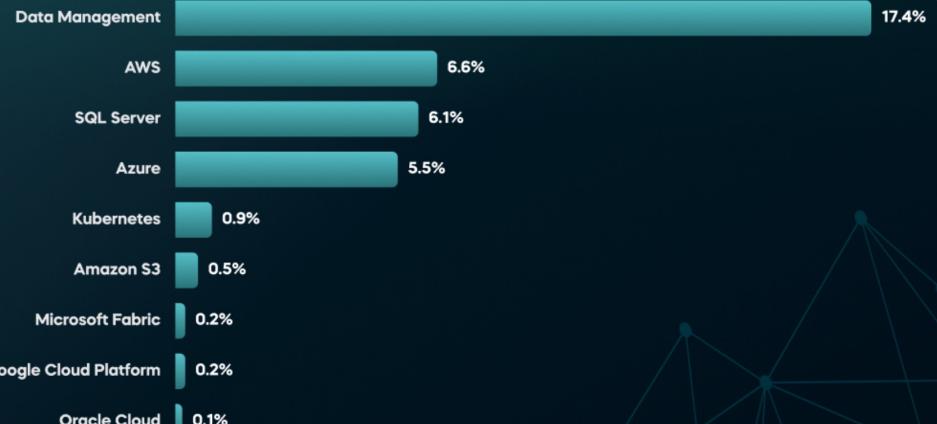
■ Exploratory Data Analysis



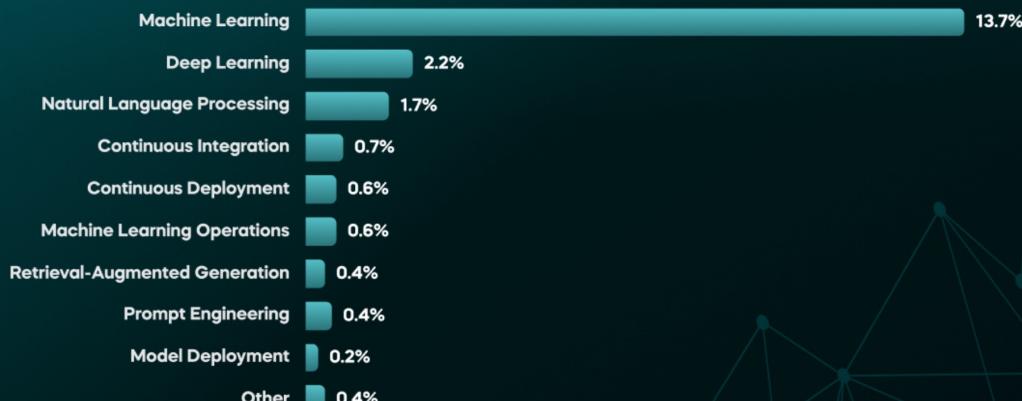
Required Data Visualization Tools Data Analysts 2025



Required Cloud Skills Data Analysts 2025



Required AI Skills Data Analysts 2025



+

Disclaimer



2110531: DS & DE Tools

Practical Data Analytics & ML Pipeline

- This course focuses on:
- Fundamental data processing
- ML pipeline
- Practical process (MLOps)



- **Remark1: DS != AI/ML**
- We cannot cover all algorithms (NN, DL, NLP, CV, etc.) in this course!
- **Remark2: DS != DE (aka. DW)**
- There are a lot more DE tasks (NoSQL, ETL, etc.)

+

Any questions? ☺