

Seminar

# Data Science and Machine Learning

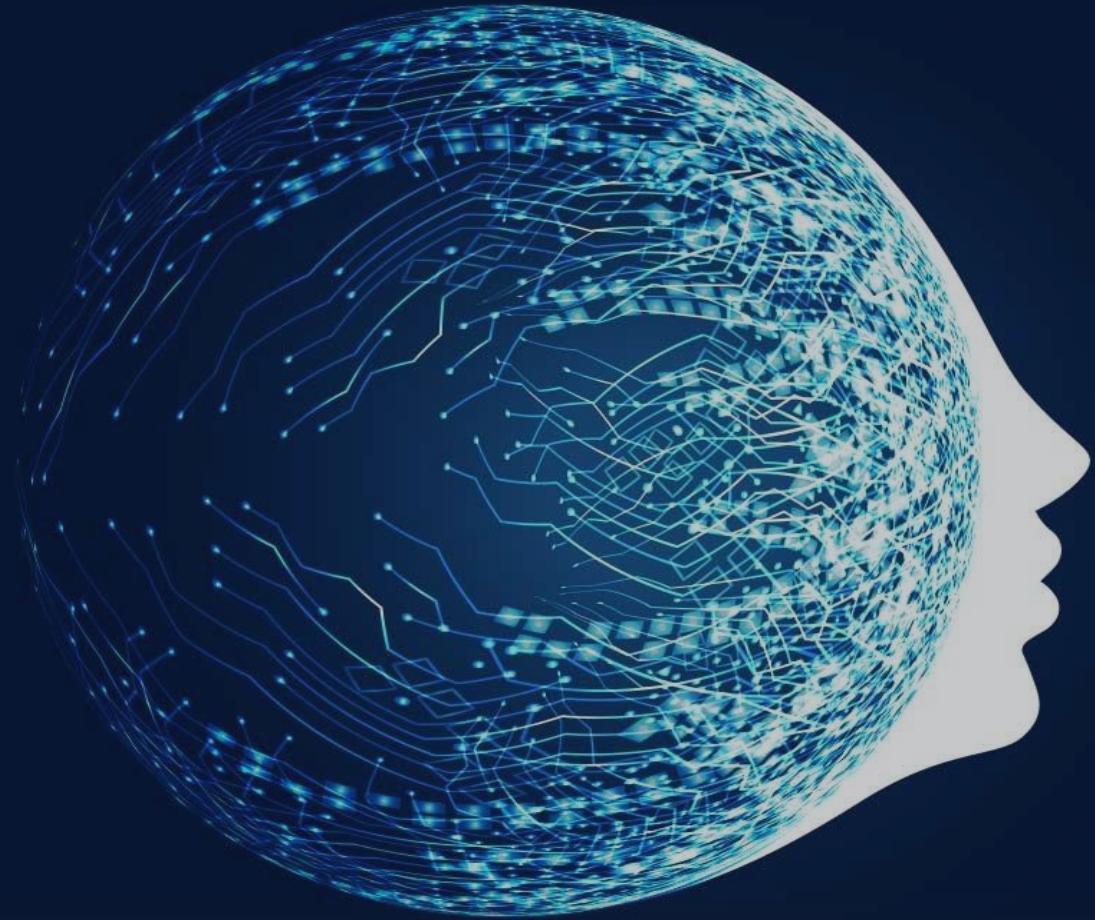
Start to Success



Lecturer:

**Sajad Heydari**

Aspiring Data Scientist  
MSc. Industrial Engineering





## ❖ MY JOURNEY

MSc. Industrial Engineering, Tarbiat Modares University  
Data Science Researcher especially in Intelligent Transportation Systems  
Data Science Courses Instructor in Iranian top-notch Universities  
Data Scientist at Snapp!Food  
Data Science Consultant in Industries: Healthcare, Supply Chain, Transportation,...  
Marketing Data Analyst at Pasargad Insurance Co.

# Importance of Data

5 Exabyte Data generated every minutes **NOW!**

ONLY **0.5%** of generated Data are analyzed

## Data Explosion



**64,140**  
Instagram stories  
posted



**336,480**  
Skype calls



**567,360**  
Tweets sent



**5,365,260**  
YouTube videos  
watched



**5,500,560**  
Google searches  
conducted



**181,331,340**  
Emails sent

Picture source: SAS

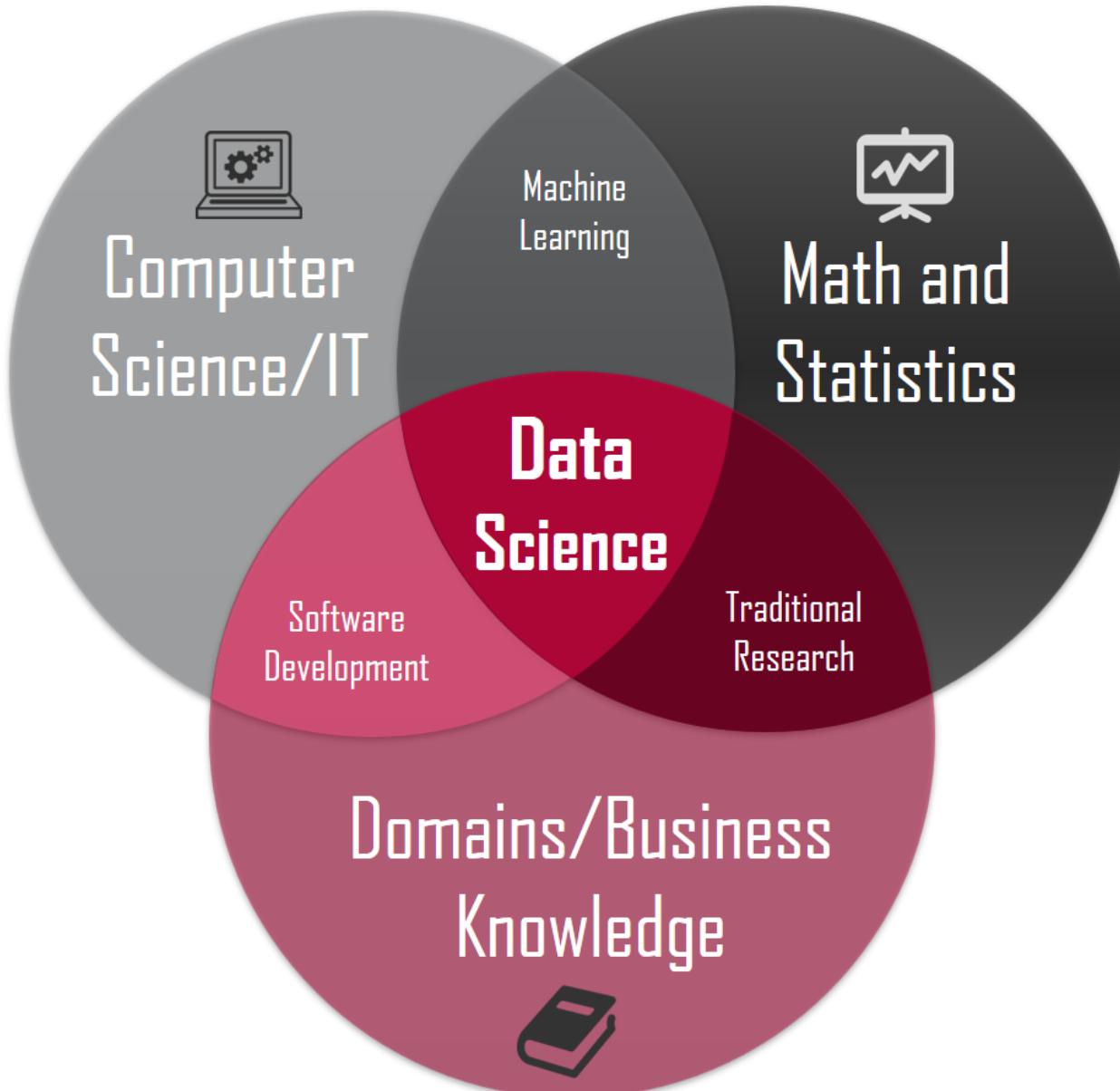
# What is Data Science?



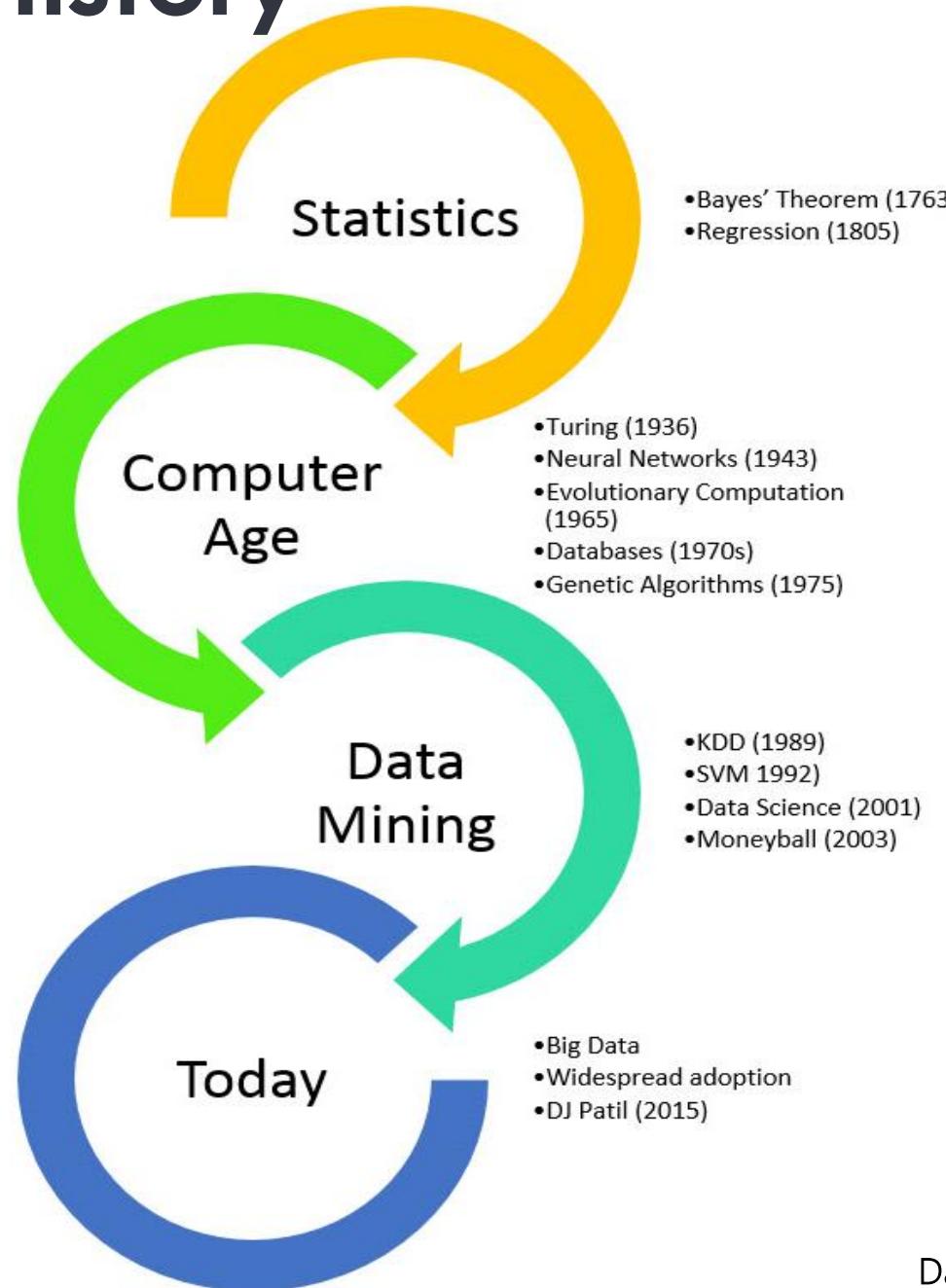
Data science is a  
**Multidisciplinary** approach to  
extracting **actionable insights**.



# What is Data Science?



# Data Science History



# Handle unforeseen situations with DS

Business Analysis  
for Smarter  
Decisions

Making Better  
Products

Managing  
Business  
Efficiently

Predictive  
Analysis to  
Predict  
Outcomes

Leveraging Data  
for Business  
Decisions

Assessing  
Business  
Decisions

Automating  
Recruiting  
Process

# Data Mining Tasks

- **Predictive**

Use some variables to predict unknown or values of other variables.

Classification, Regression

- **Descriptive**

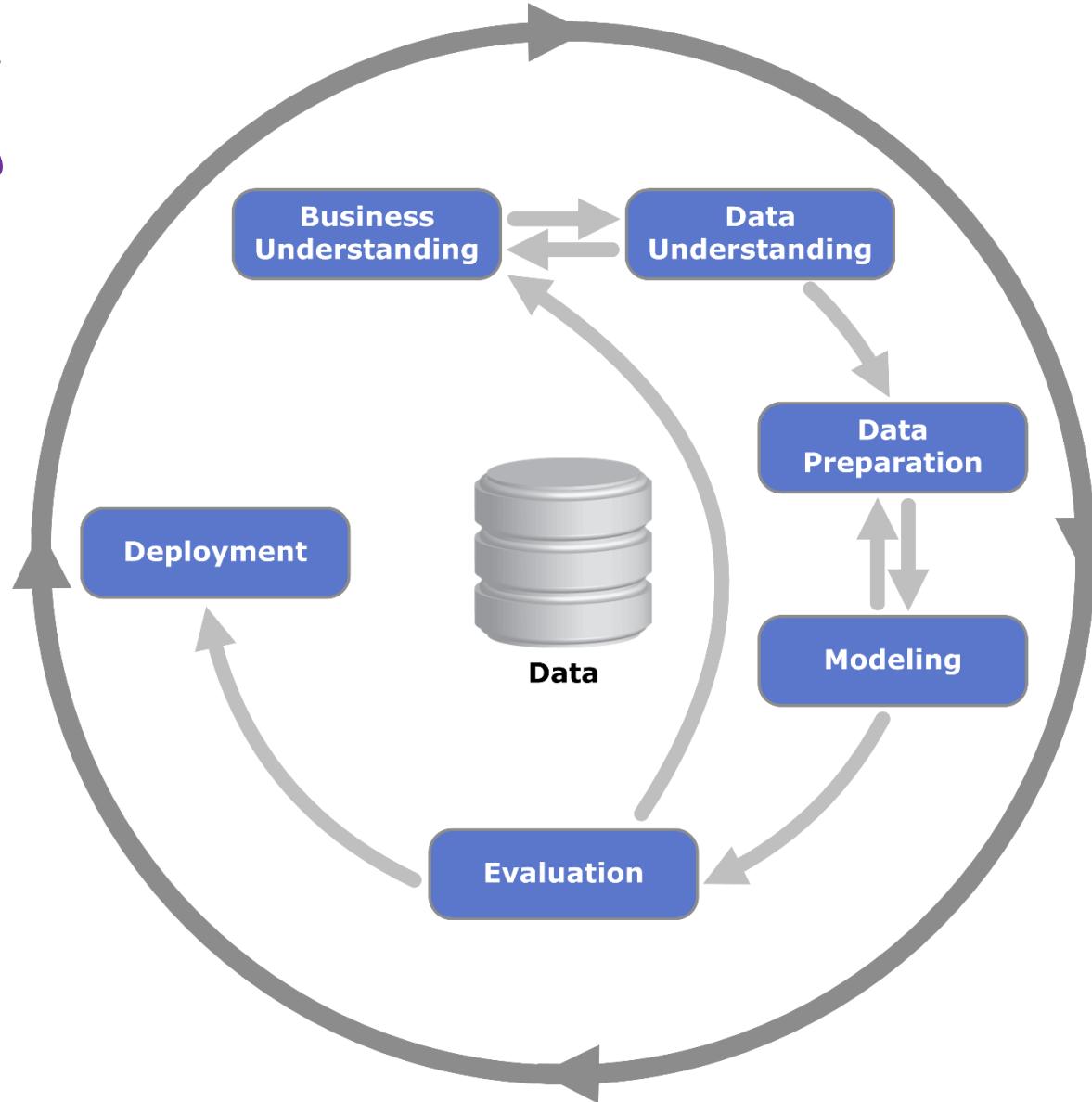
Find human-Interpretable patterns that describe the data

Clustering

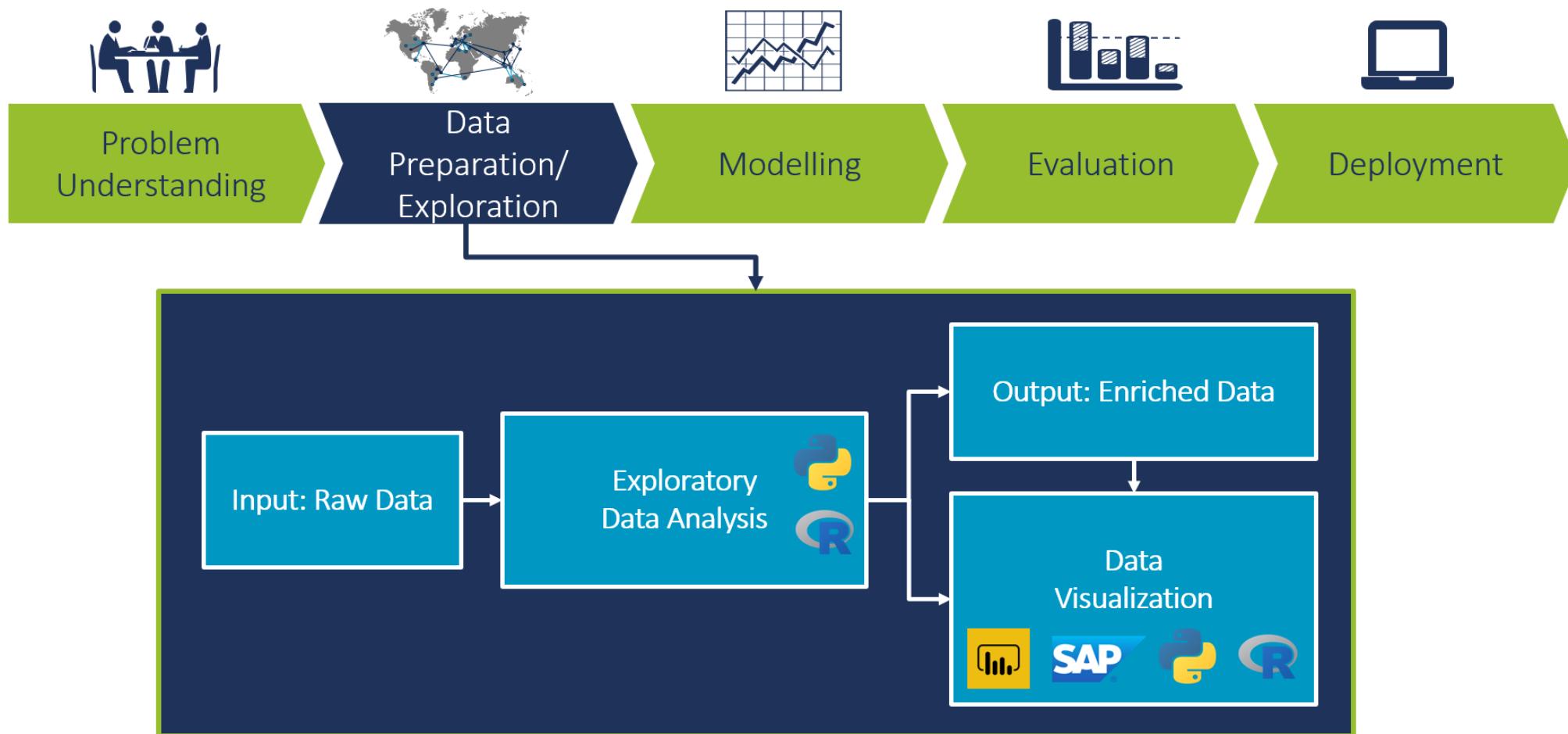


# Crisp-DM Methodology

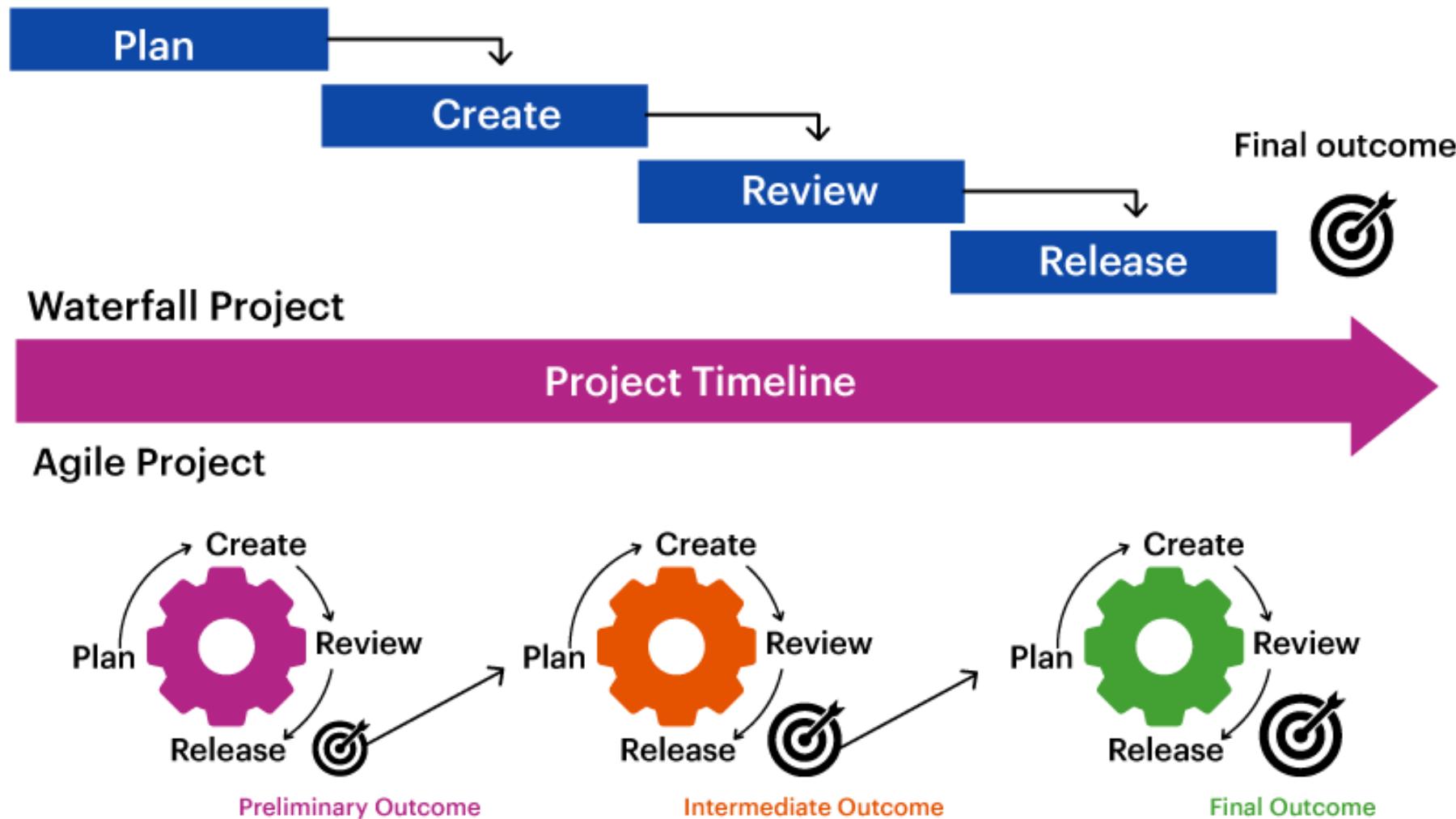
Cross-Industry Standard  
Process for Data Mining



# What is EDA?

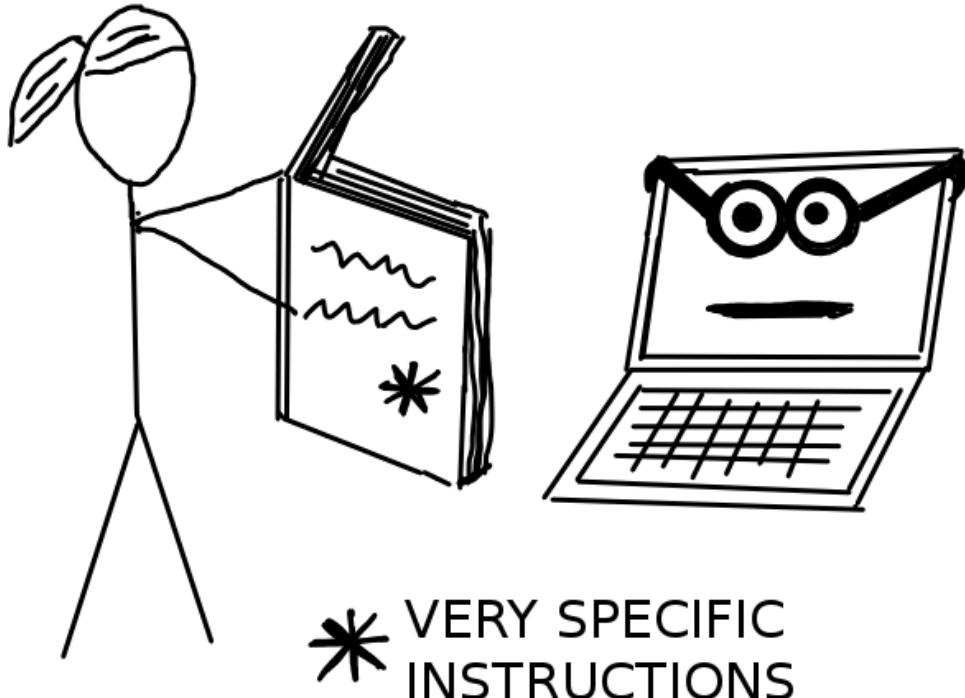


# Agile + Data Science

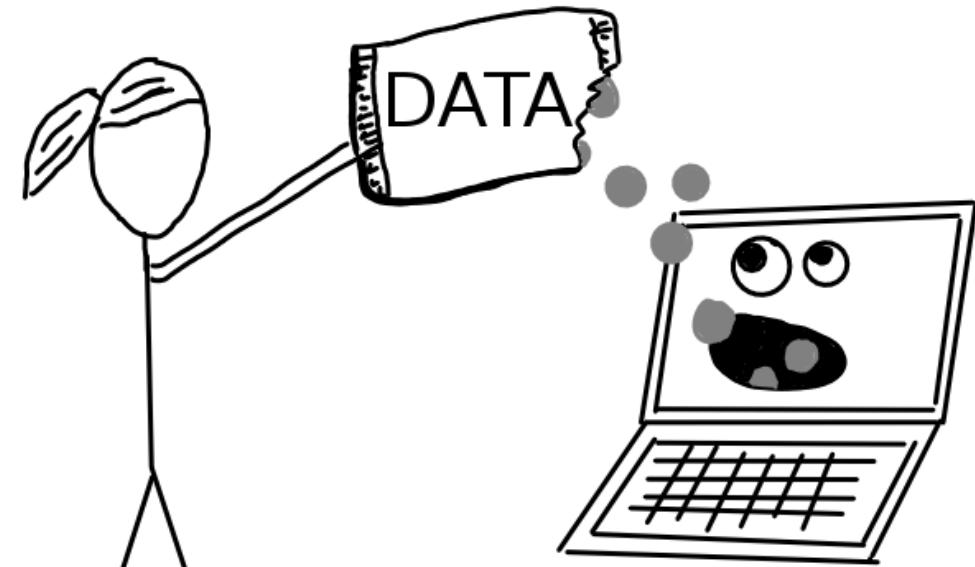


# What is Machine Learning?

## Without Machine Learning



## With Machine Learning



# ML Sections



Supervised Learning

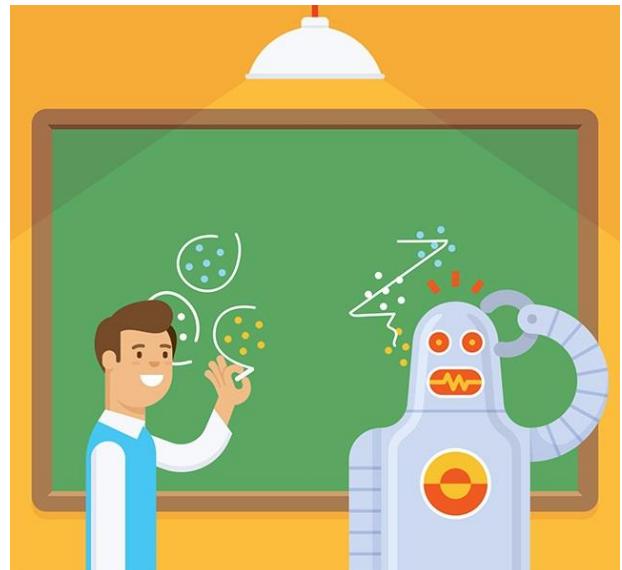
agent



Reinforcement Learning

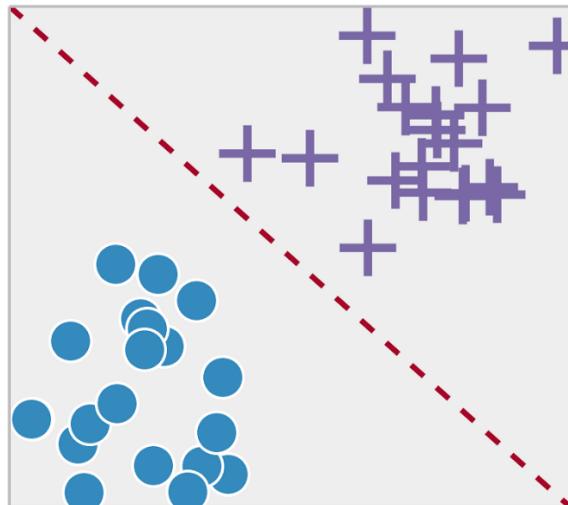
Unsupervised Learning

Semi-Supervised Learning

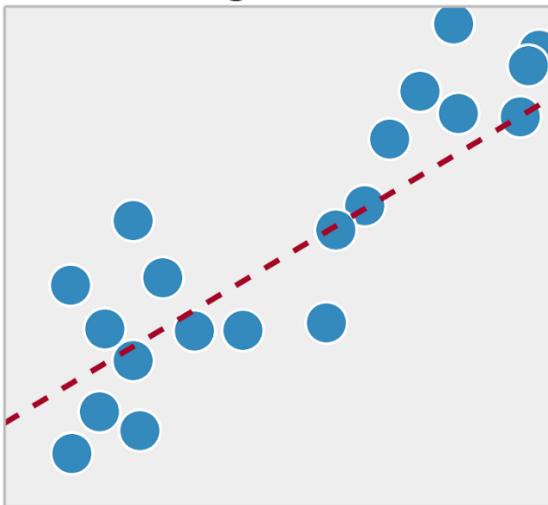


# Supervised Learning Sections

Classification

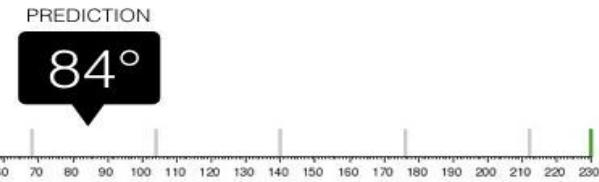


Regression



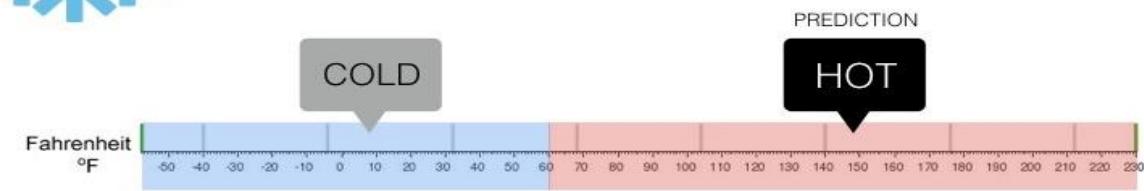
## Regression

What is the temperature going to be tomorrow?



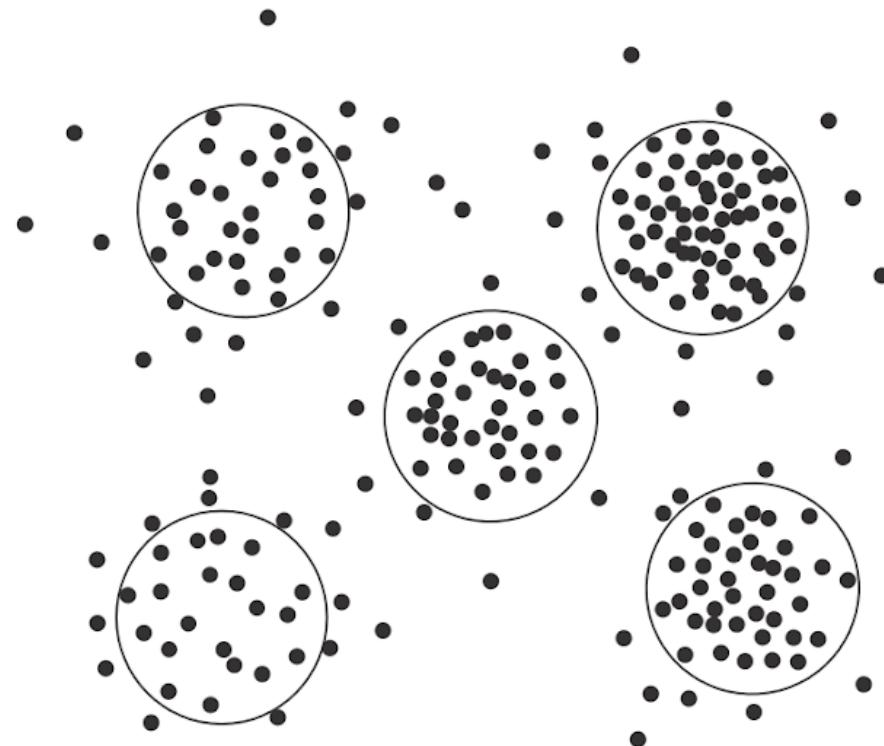
## Classification

Will it be Cold or Hot tomorrow?

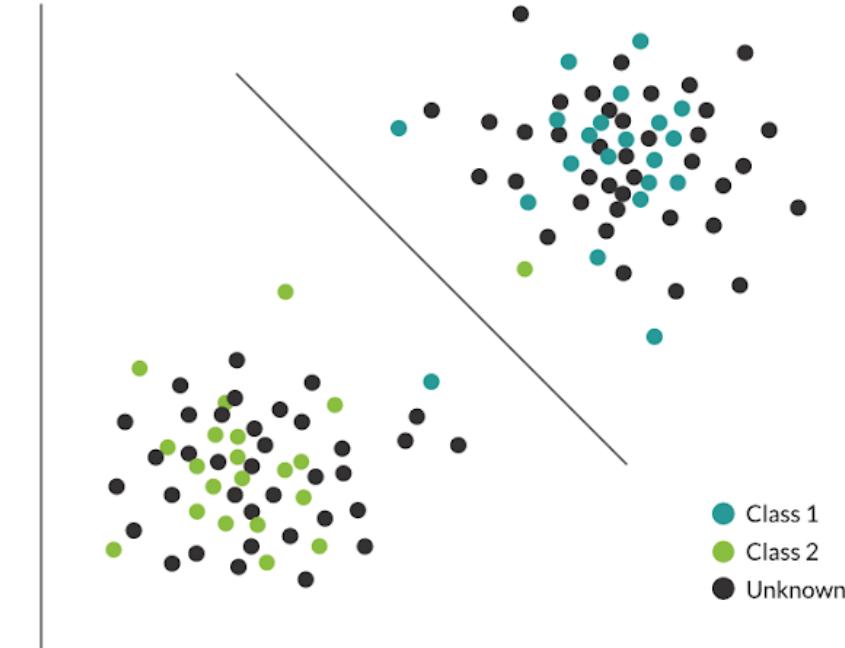


# What is Unsupervised Learning?

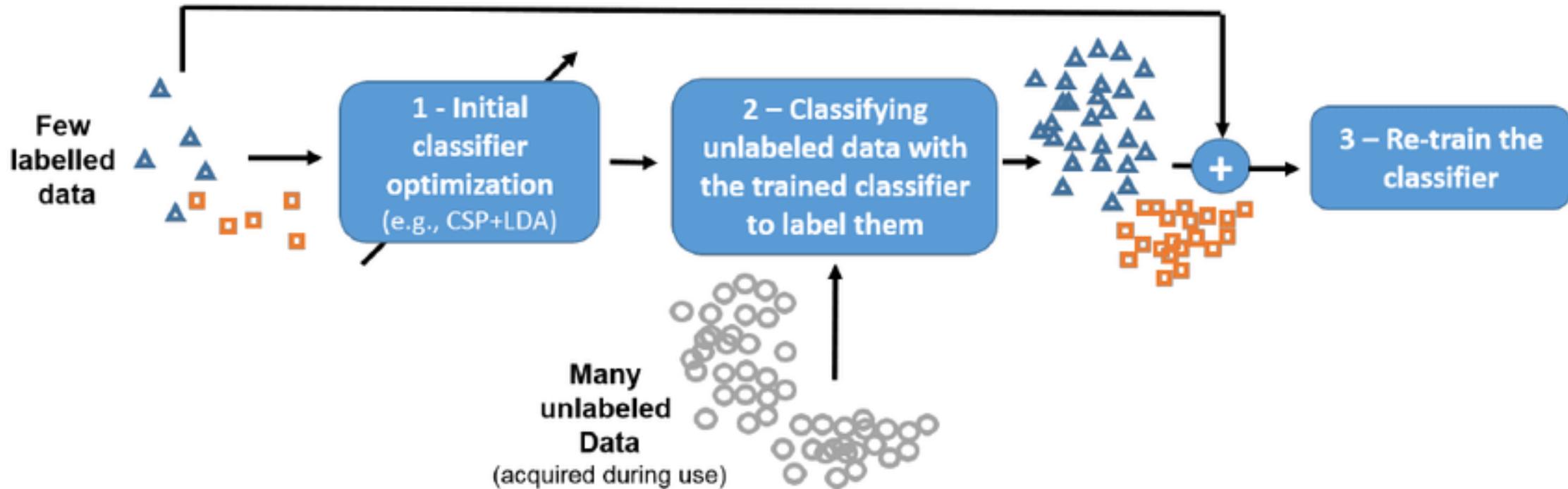
Unsupervised



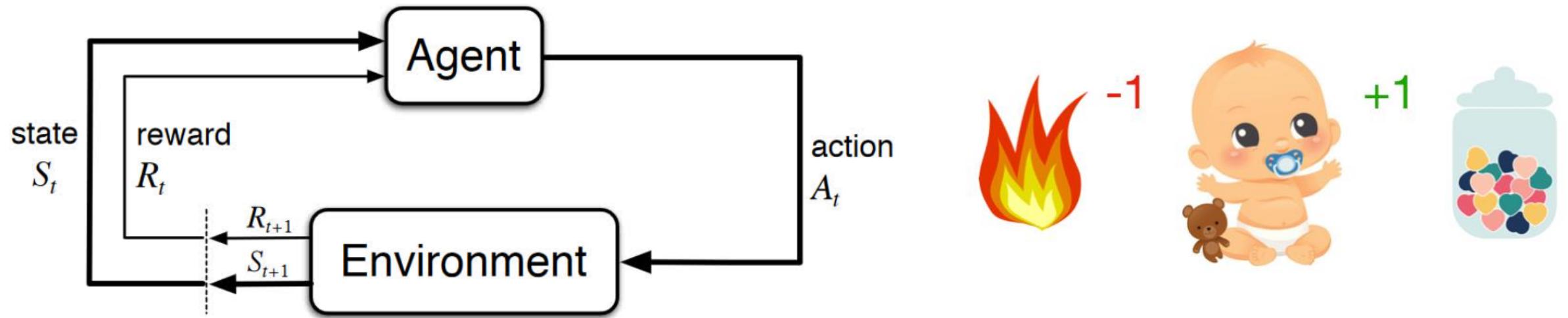
Supervised



# What is Semi-Supervised Learning?

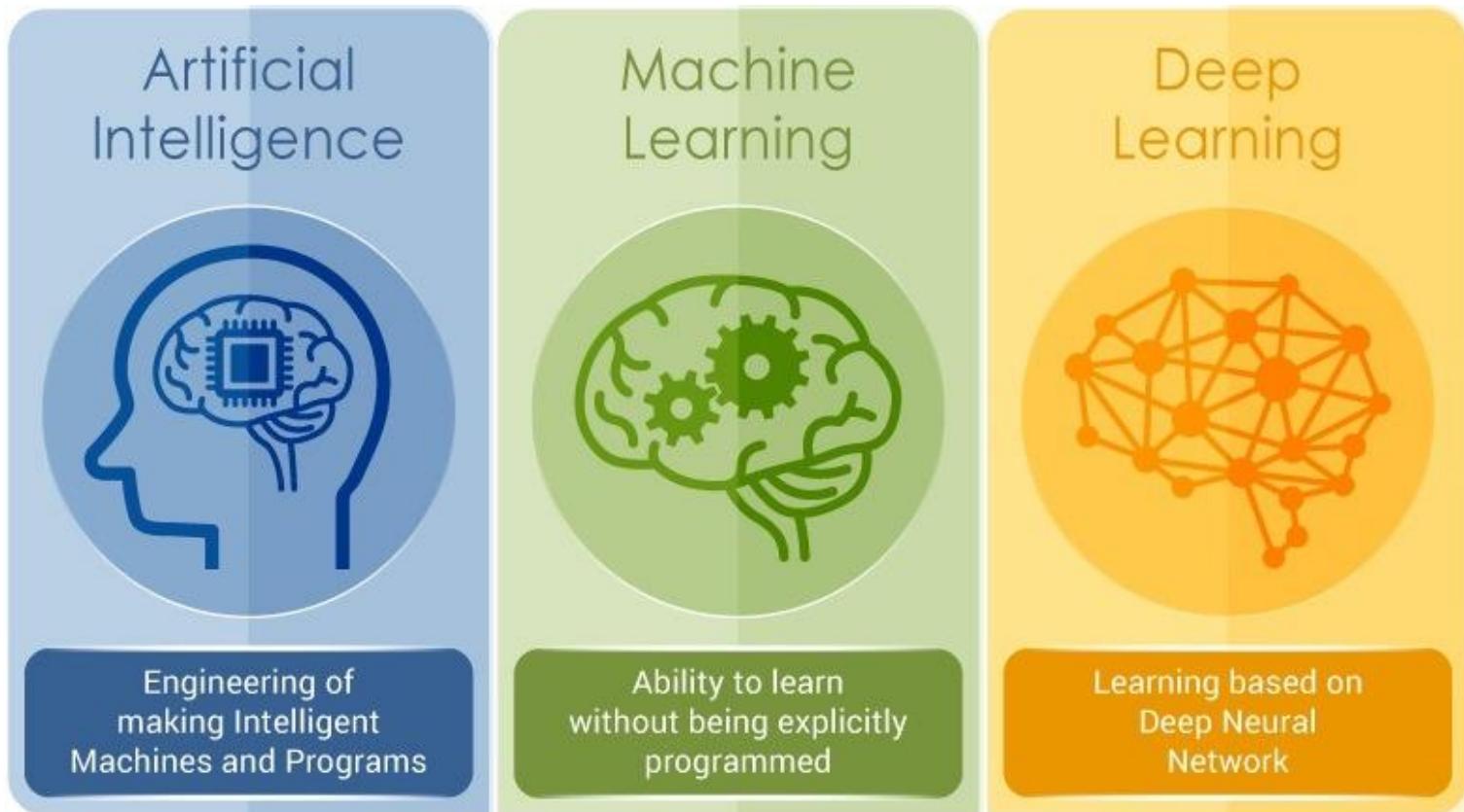
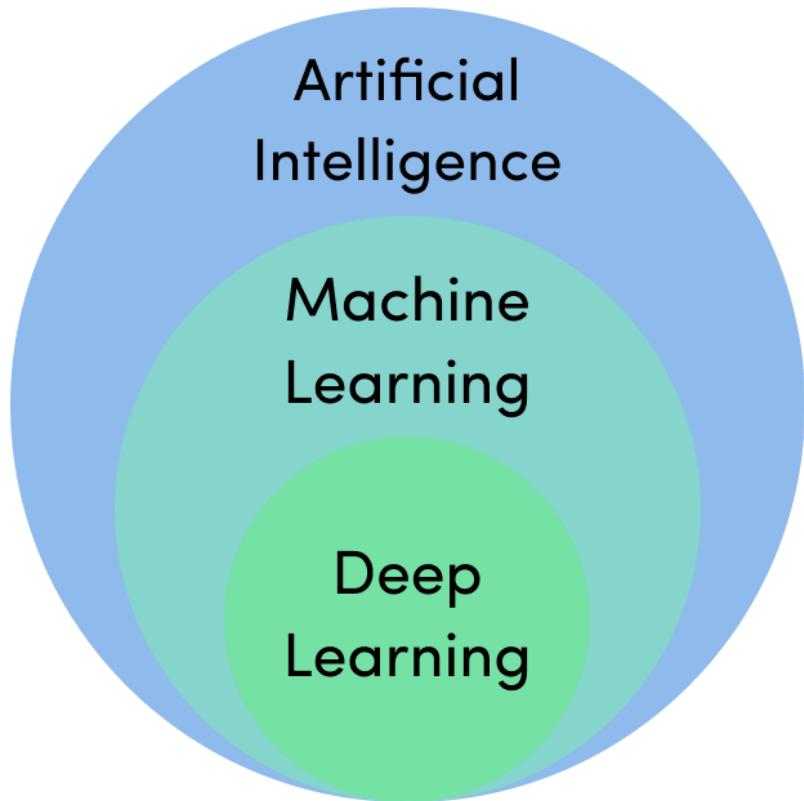


# What is Reinforcement Learning?

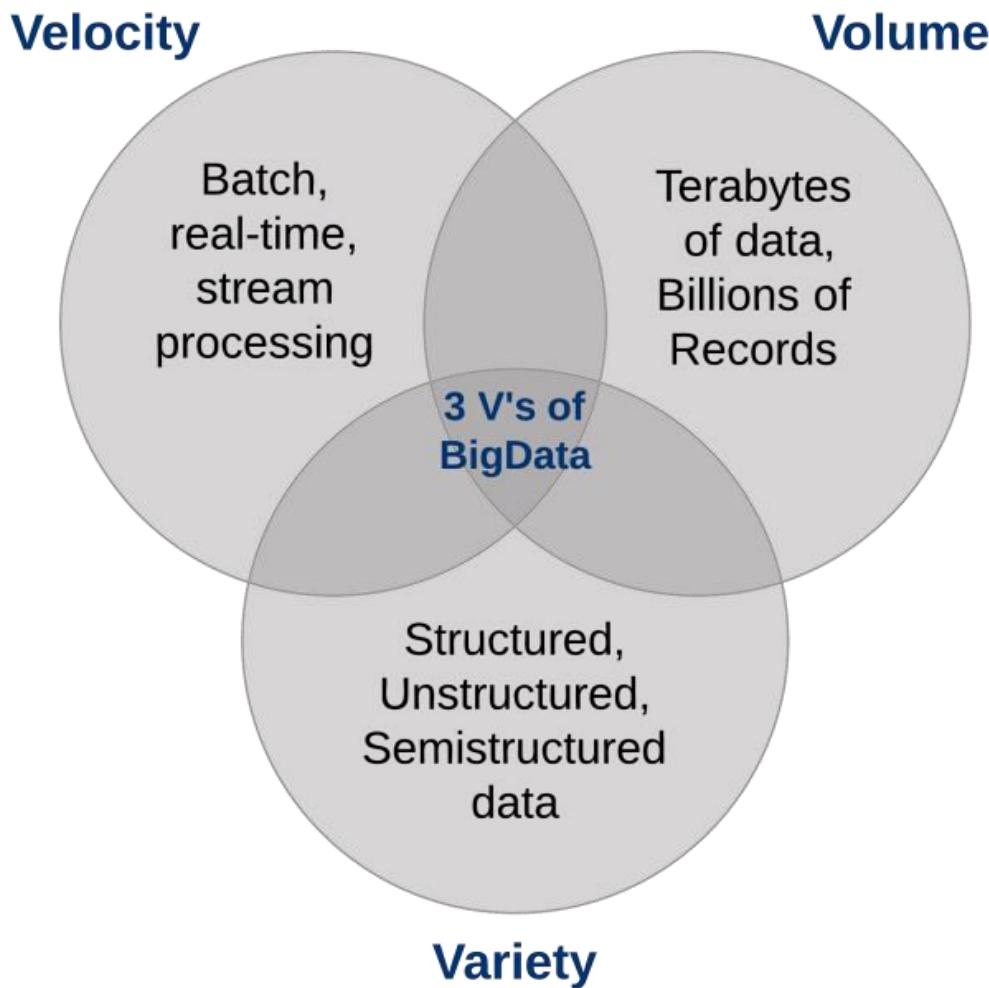


**Reinforcement learning is the science of decision-making.**

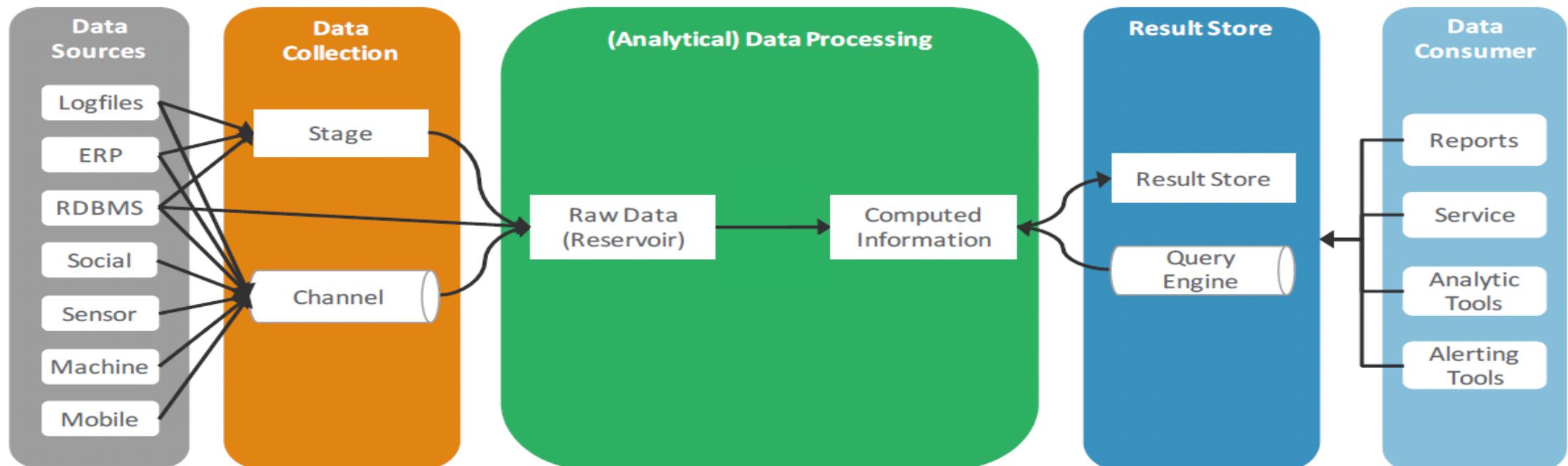
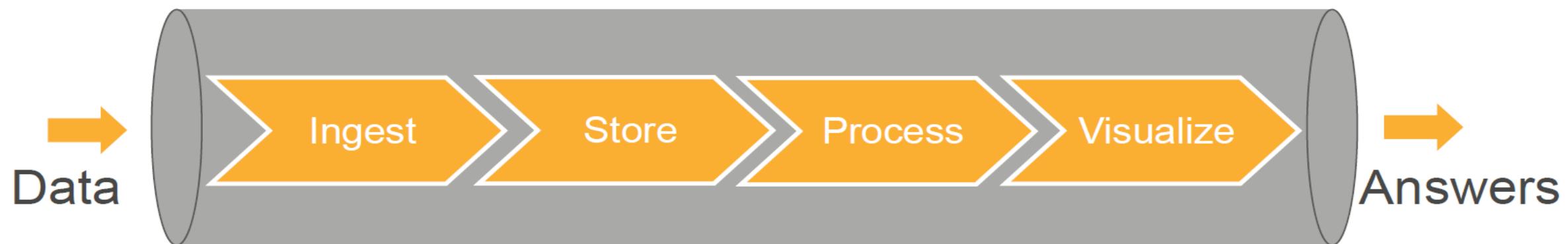
# Deep Learning



# What is Big Data?

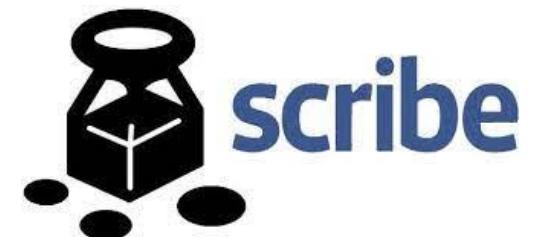


# How Big Data Works?



# Big Data Layers: Data Ingestion

- Scalable, Extensible to capture streaming and batch data
- Provide capability to business logic, filters, validation, data quality, routing, etc. business requirements
- Technology Stack:
  - Apache Kafka
  - Apache Sqoop
  - Apache NiFi
  - Logstash
  - Fluentd
  - Facebook Scribe
  - Amazon Kinesis
  - ...



# Big Data Layers: Data Storage

- Depending on the requirements data can placed into Distributed File System, Object Storage, NoSQL Databases, etc.
- Technology Stack:
  - HDFS, Hive
  - Redis, MongoDB, Hbase, Cassandra, ElasticSearch,...
  - RDBMS



# Big Data Layers: Data Processing

- Processing is provided for batch, streaming and near real time cases
- Scale-Out Instead of Scale-Up
- Fault-Tolerant based methods
- Technology Stack:
  - MapReduce
  - Spark
  - Strom

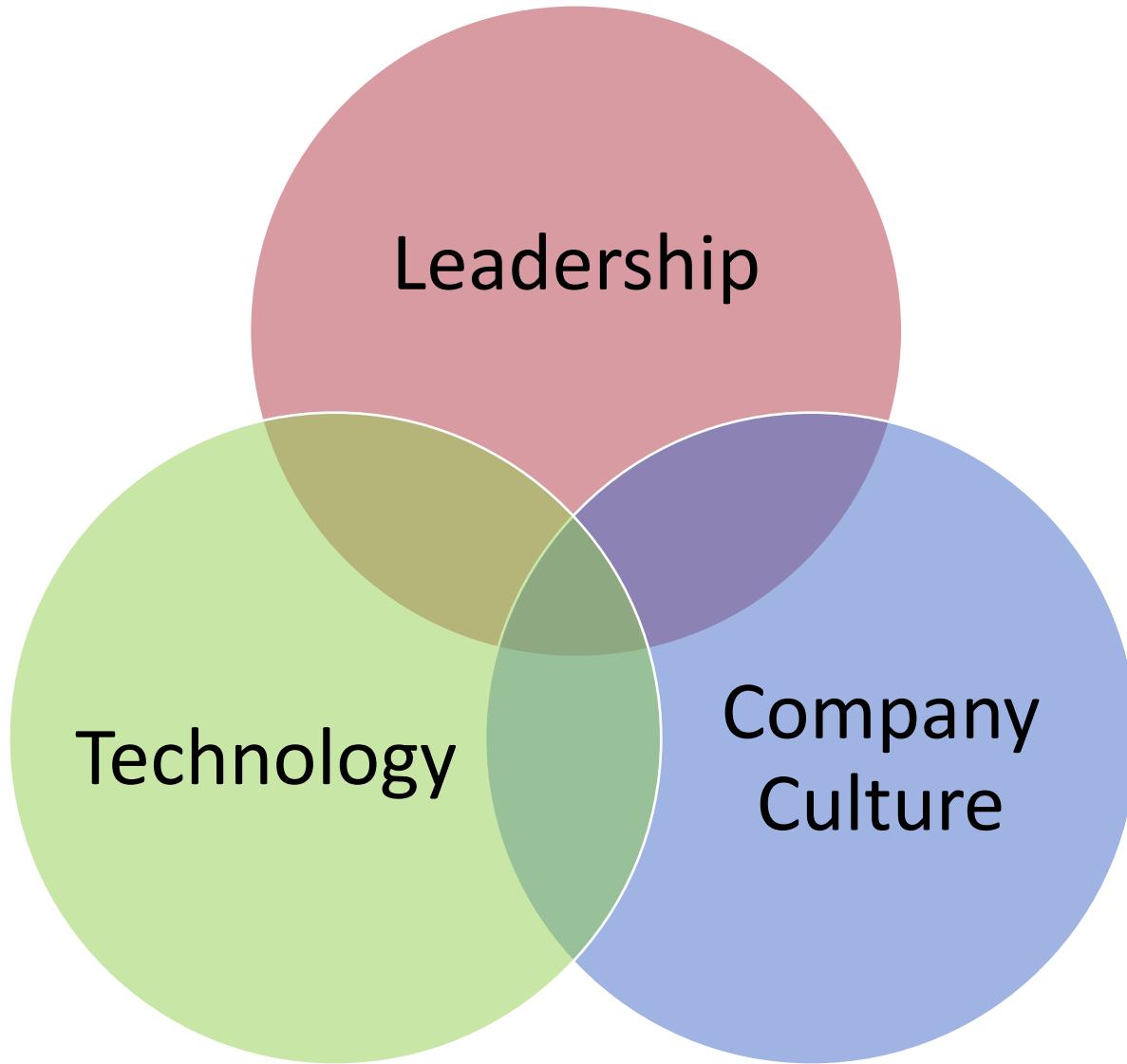


# Big Data Layers: Data Visualization

- Dashboard and applications that provides valuable business insights
- Technology Stack:
  - Qlik
  - Tableau
  - Google Data Studio



# Business Challenges with Big Data



# ML Cases



Machine Learning Project on Netflix  
Recommendation System

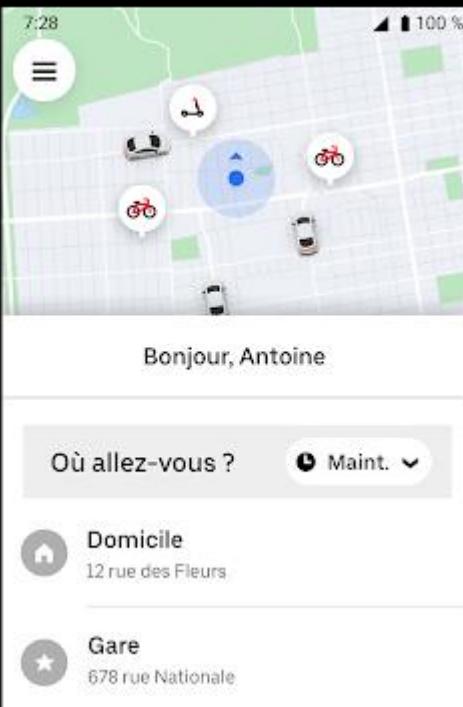


# ML Cases

Find Best Price



Optimize Waiting Time and  
Idle Vehicle



U B E R

# ML Cases



A screenshot of the Uber Eats app showing the profile for Lucia Ristorante. The screen includes a photo of two pasta dishes, the restaurant's name, its rating (4.5), and delivery/pickup information. Below this, there are sections for "Picked for you" featuring Fettuccine Alfredo, Mediterranean Pizza, and Giambotta.

A screenshot of the Uber Eats app showing search results. It displays a breakfast cafe option with a photo of a dish, delivery details, and a rating of 4.8. Below this, it shows "Today's offers" for Lucia Ristorante and Louis Pipi, each with a photo of their respective dishes.

A screenshot of the Uber Eats app showing reviews for Lucia Ristorante. It displays a 4.5 rating, a breakdown of review counts by star, and a specific review from a user named Christina. The review photo shows a Margherita pizza.

# ML Cases

# amazon



# Big Data Cases

Analyze genetic variations and potential treatment effectiveness



# Big Data Cases

## Content and user viewing behavior



# Big Data Cases



Monitor various activities of players

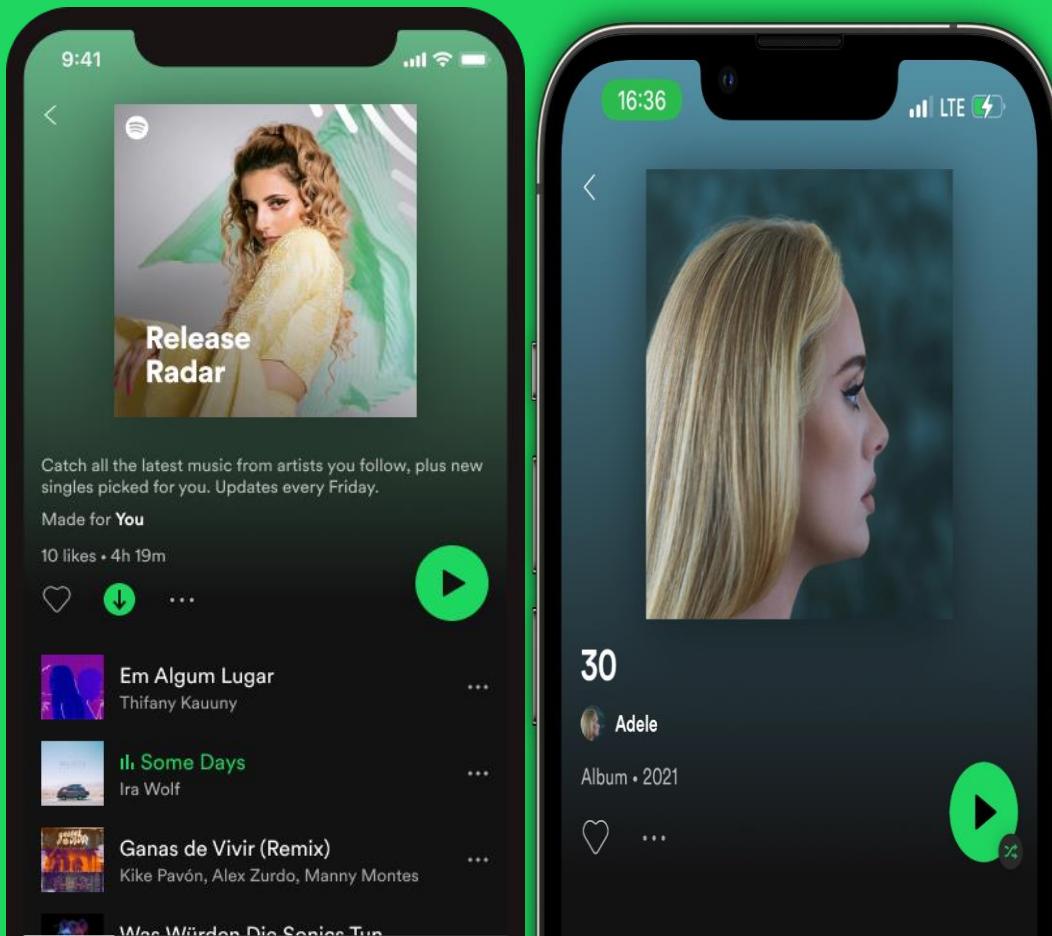


# Applied Data Science



Sound Modelling by Complex  
Deep Neural Networks

Recommend to users

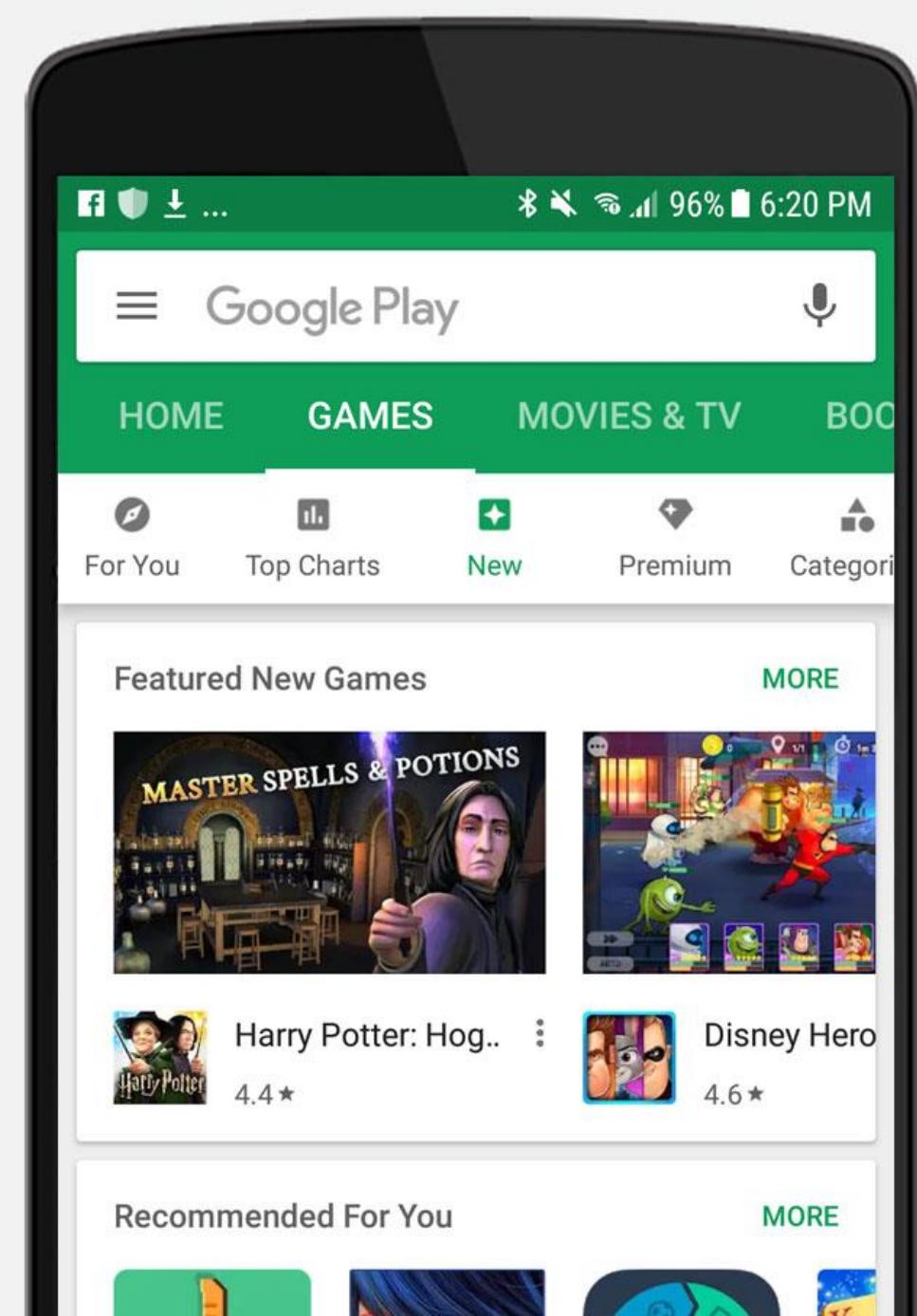


# Applied Data Science



# Google Play

## Classify Applications by Rating



# Applied Data Science



## Dynamic Inventory Management



# Applied Data Science



Customer behavior segmentation  
in Internet Packages



## Real-time analytics

combines the data related to customer profiles, network, location, traffic, and usage to create a 360-degree user-centric view of the product or service.

# Learning Path: Relational Databases



PostgreSQL



Microsoft®  
SQL Server®

# Learning Path: NoSQL Databases



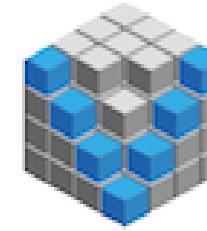
cassandra



Flock DB



Dynamo



Voldemort



# Learning Path: Programming

Which is Better to Choose



# Learning Path: Visualization Tool



# Learning Path: Statistics and Probability

Probability distributions



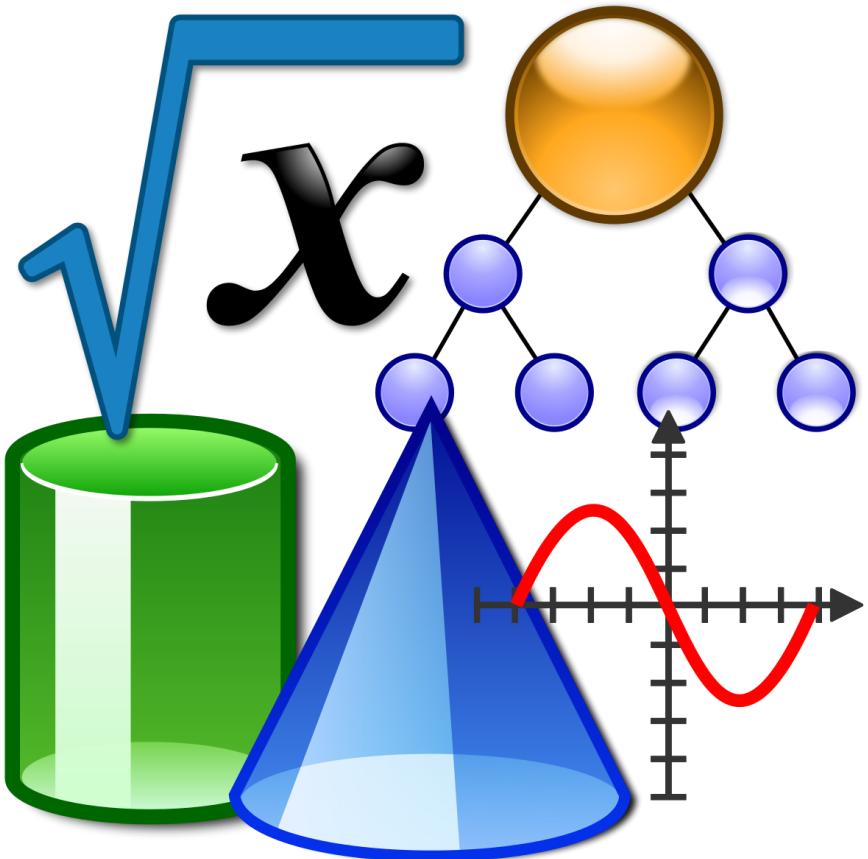
Kinds of Data

Sampling methods

Descriptive Statistics

Inferential Statistics

# Learning Path: Math



Distances

Sigma relations

Specific Functions (sigmoid,...)

Convex and non-Convex Functions

Gradients

# Learning Path: Linear Algebra

$$A = [1 \ 2 \ 3]$$

Matrices

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

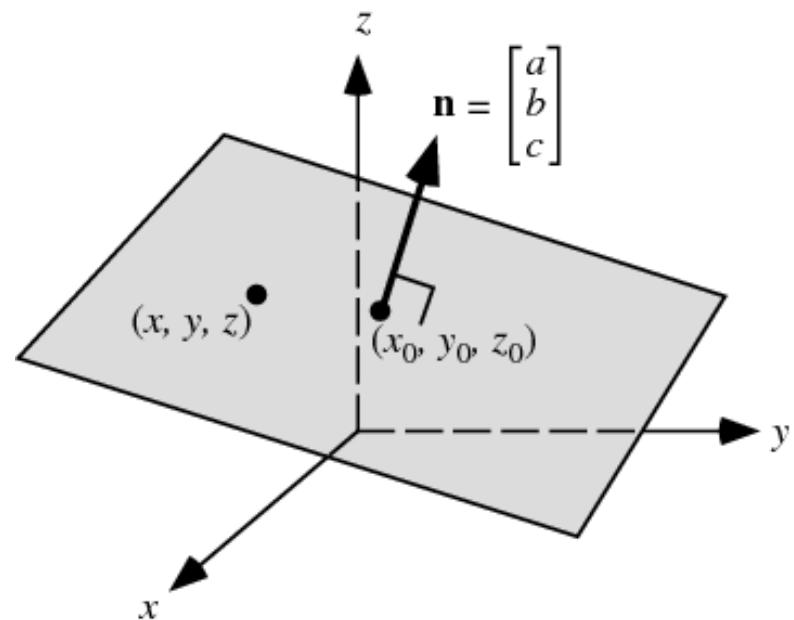
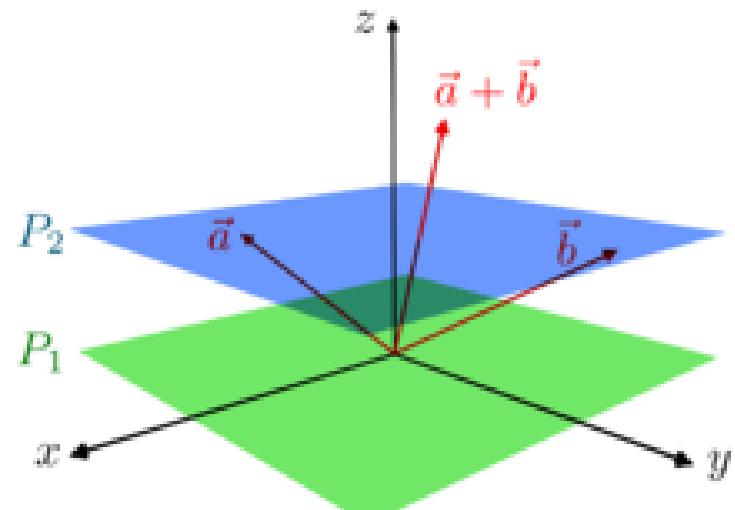
Eigen values

$$B = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Matrix Relations

Sets Relations

# Learning Path: Geometry

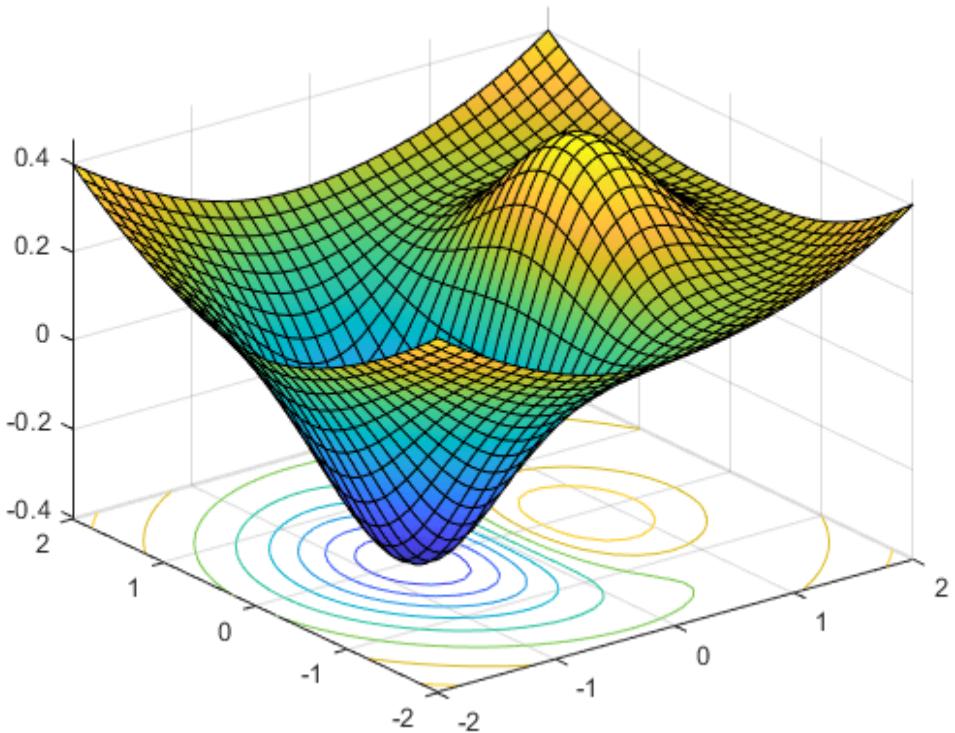


Hyperplane

Spaces

Vectors

# Learning Path: Optimization



Linear Programming

Quadratic Programming

Stochastic Programming

Dynamic Programming

# Successful Career in Data Science

## Completing High-Quality Courses



### Newest courses in Data Science



100 Days Data Science Bootcamp: Build 100 Real Life...

Pianalytix .

4.7 ★★★★★ (14)

\$19.99

New



Build 75 Powerful Data Science & Machine Learning Projects

Pianalytix .

4.9 ★★★★★ (16)

\$19.99

New



Data Science Mega-Course: #Build {120-Projects In 120-...

Pianalytix .

4.4 ★★★★★ (18)

\$19.99

New



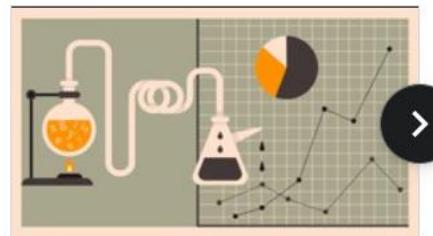
Mining and Analyzing LinkedIn Data

Jones Granatyr, IA Expert Academy

5.0 ★★★★★ (5)

\$19.99

Bestseller



Be Aware of Data Science

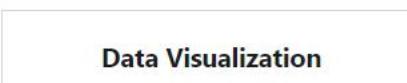
Robert Barcik, Jana Gecelovska, Patrik Zatko

4.8 ★★★★★ (39)

\$49.99

New

### Topics recommended for you



# Successful Career in Data Science

Completing High-Quality Courses



coursera

Explore ▾

What do you want to learn?



Online Degrees

Find your New Career

For Enterprise

For Universities

Log In

Join for Free

Courses and Specializations

## Courses You Can Complete in a Day

Over 4,000 courses in topics like business analytics, graphic design, Python, and more.



Foundations of  
Mindfulness  
Rice University

COURSE



Innovation Through  
Design: Think, Make,  
Break, Repeat  
The University of Sydney

COURSE



Cloud Computing Basics  
(Cloud 101)  
LearnQuest

COURSE



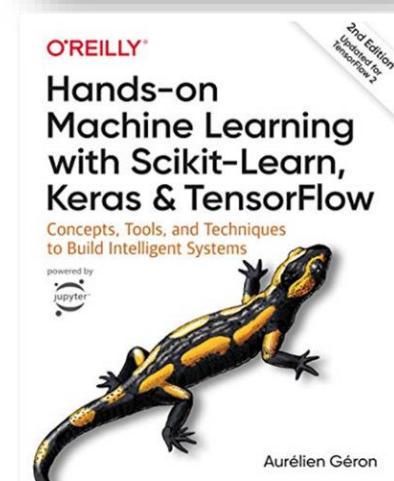
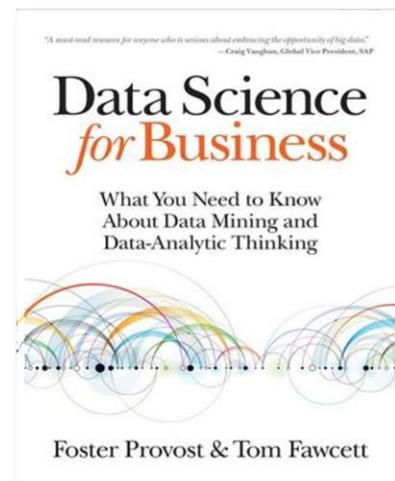
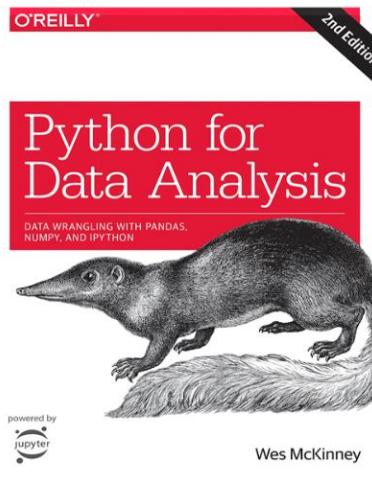
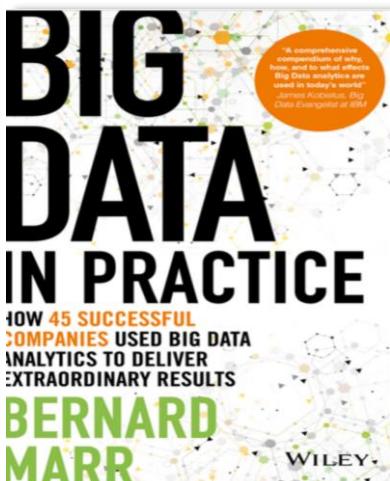
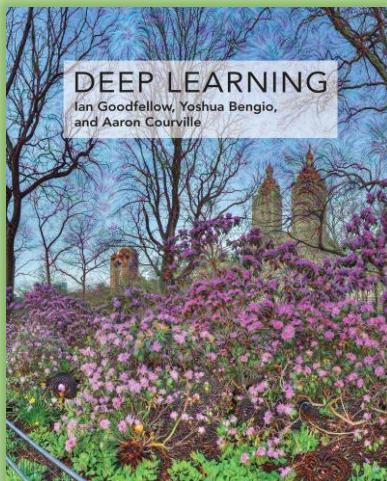
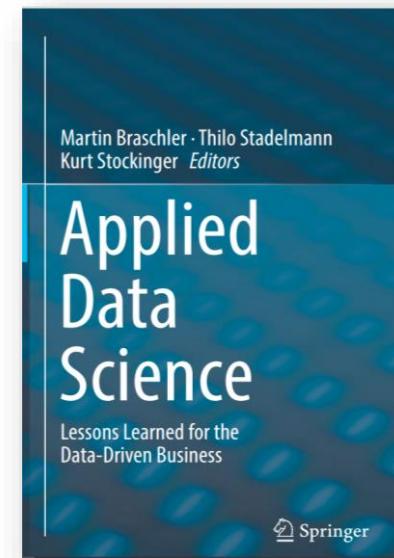
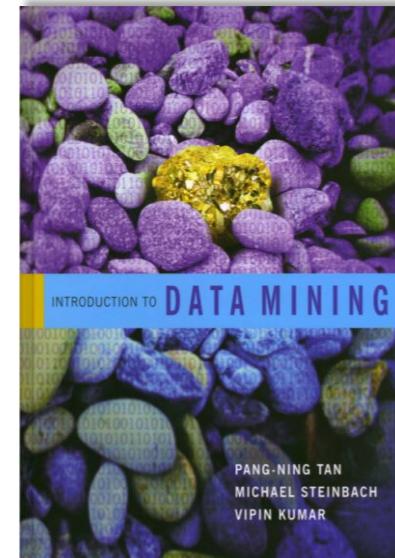
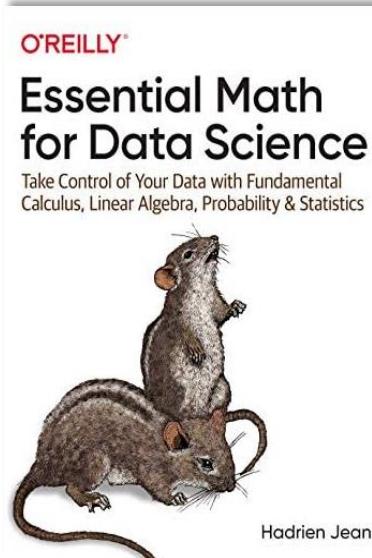
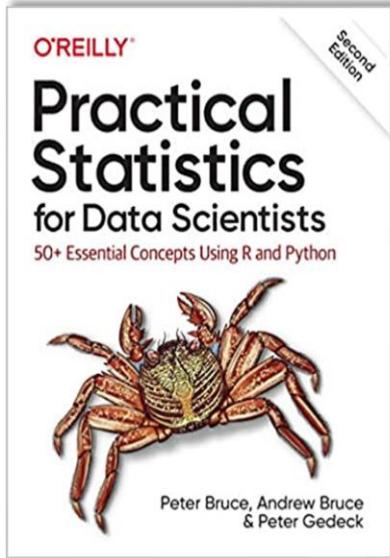
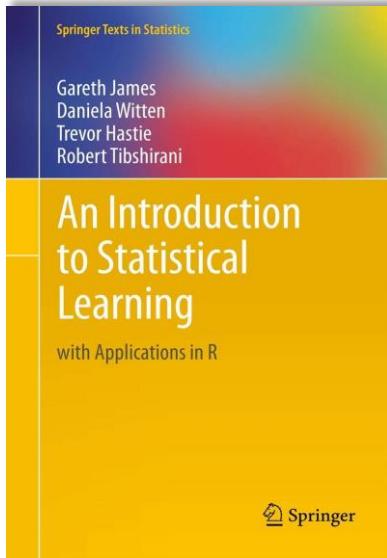
Statistics with SAS  
SAS

COURSE



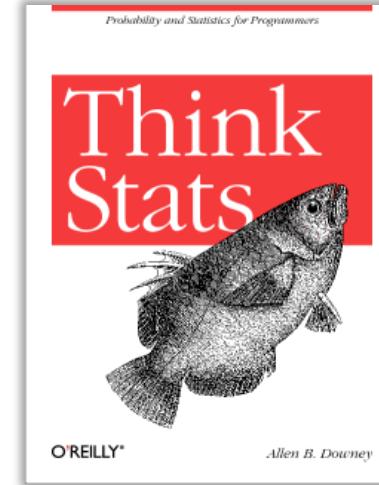
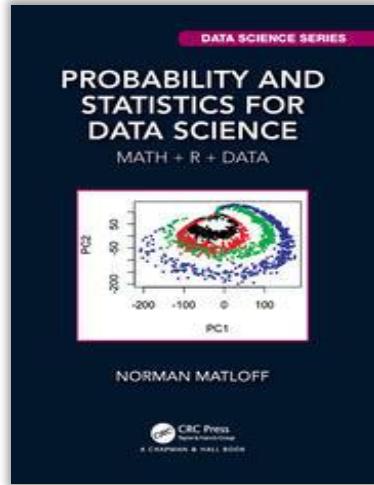
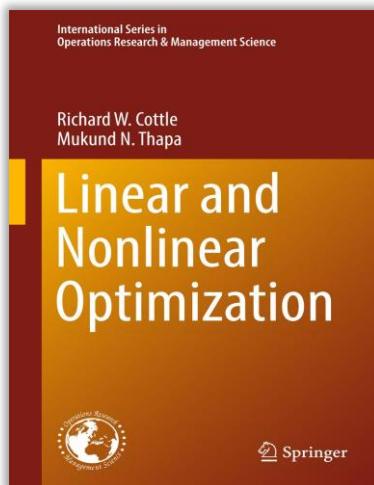
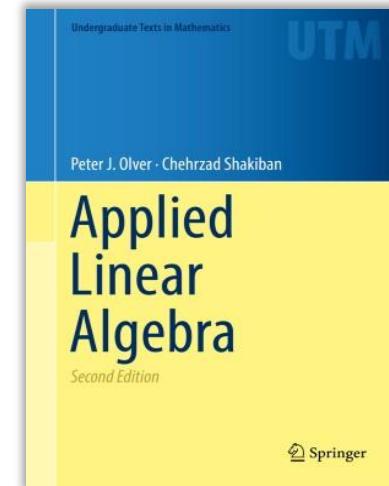
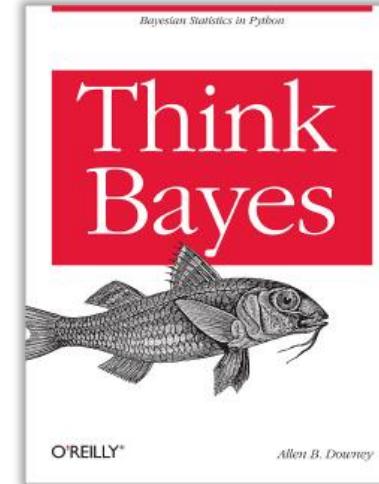
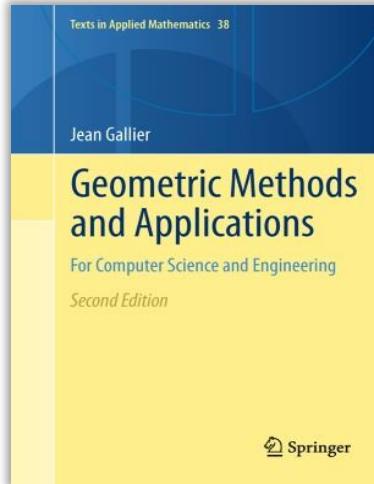
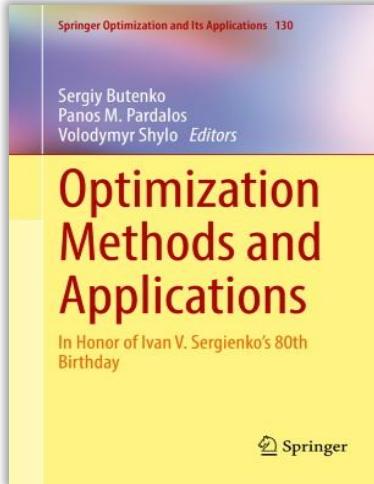
# Successful Career in Data Science

## Completing High-Quality Books



# Successful Career in Data Science

## Completing basic High-Quality Books



# Successful Career in Data Science

Completing High-Quality Articles



ScienceDirect

ELSEVIER



Springer



Google Scholar



# Successful Career in Data Science

Competitions:



DataCamp

Articles:



# Useful Advices!

Strong Relationship with Programming

Good Command of English

Do NOT Learn Everything!

Learn from Everywhere!

Be Patient



# Conclusion and Questions?

# THANK YOU!

---



@ScienceWave



Sciencewave Academy



ScienceWave

