

Data Science

for the automotive industry



Me



Algeciras

2011
Aerospace
Engineering

Internship
Image Processing

2013
M. Sc. in
Electronics, signal
Treatment and
Communications

Internship
Mechanical
testing

2018
PhD in Structural
Health Monitoring

R&D Engineer in
Material and
Processes

Reverse Engineering
Consultant

2018
Telemetry systems
specialist

2020
Data
Scientist



What about you?

Name, e-mail and brief description of background (Coding?).

Course roadmap

1) Data Science (1d)

What

Who

Why

When

How

2) Practical sessions with Python (3d)

Machine Learning (ML)



Deep Learning (DP)

3) Hackathon Python (1d)

Objective

Data Science

Introduce students to the concept

Reduce entry barrier

Pre- definitions

Data Science

What

Who

Why

When

How

Pre- definitions

Data Science

What

Who

Why

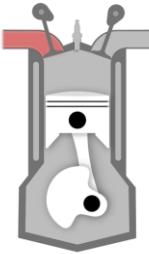
When

How

Low vs High level

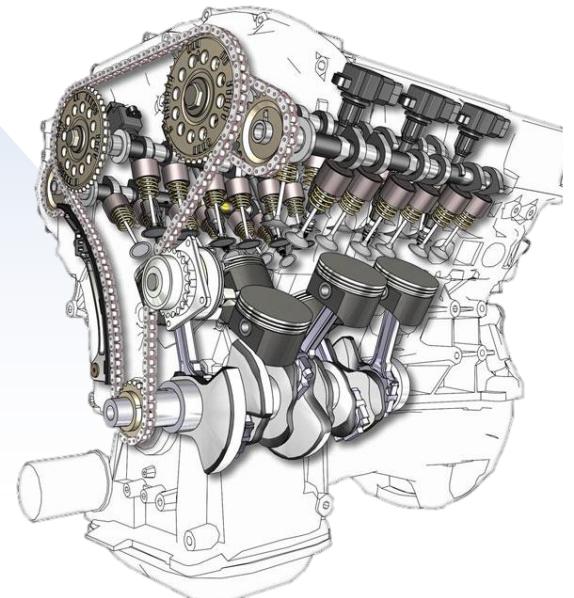
High Level

Petrol
Engine

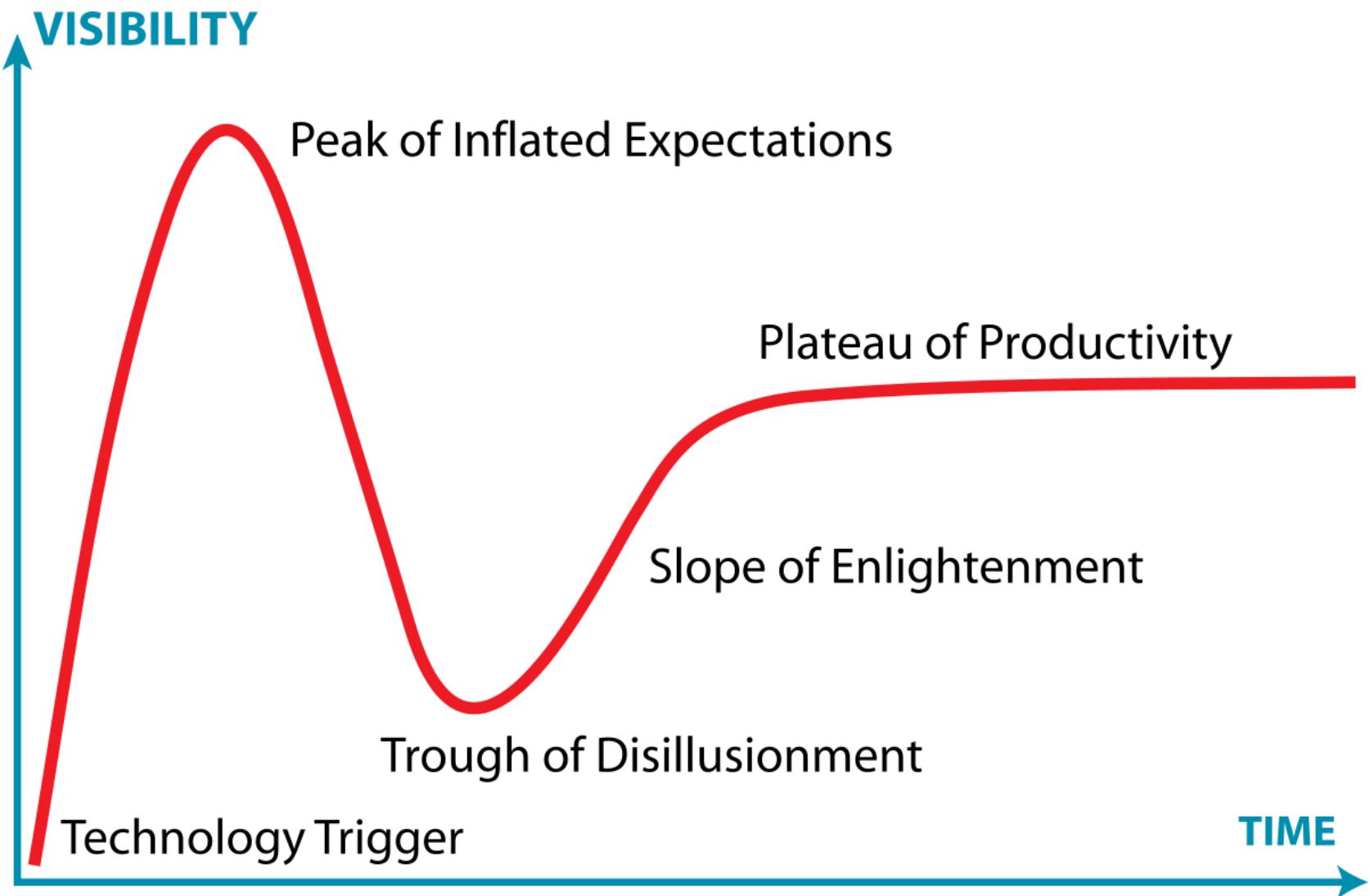


1. Intake
2. Compression
3. Power
4. Exhaust

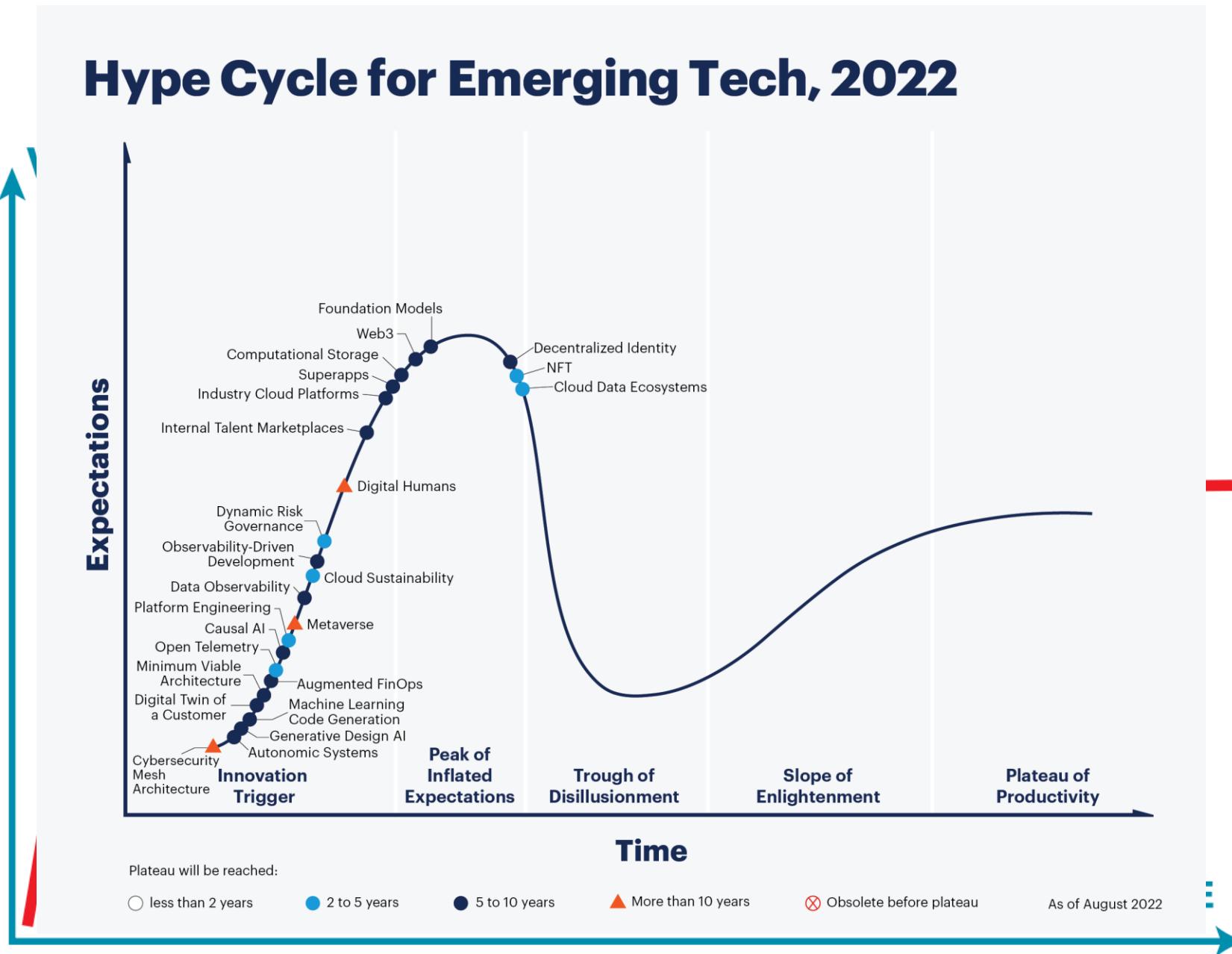
Low Level



Hype Cycle



Hype Cycle

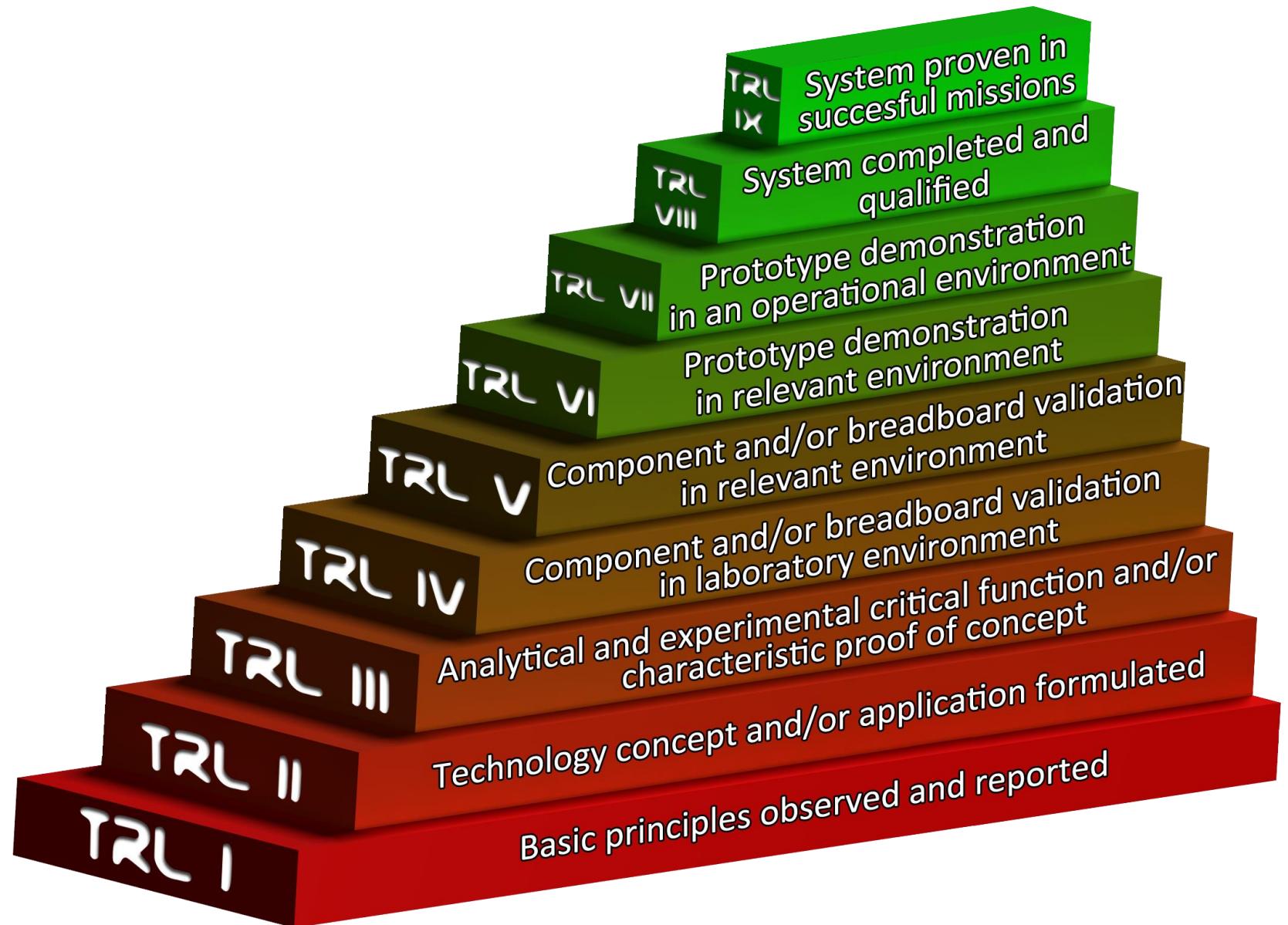


Technology Readiness Level

Deployment - Innovation
(Private companies)

Development
(Technology centres)

Basic Science - Research
(Universities)



Technology Readiness Level

Deployment - Innovation
(Private companies)

Development
(Technology centres)

Basic Science - Research
(Universities)

Technology Readiness Level (TRL) Process

NASA's quest to make jet engines quieter led to the development of chevrons, which moved relatively quickly through the TRL process to be deployed into the commercial marketplace.



TRL 8-9 (2005-now)

- Certification by the Federal Aviation Administration
- Deployed into market



TRL 7 (2001-2005)

- Validation of concept in flight
- Flight tests, final design



TRL 6 (1998-2000)

- Full scale tests for acoustics and aerodynamics
- Static engine tests



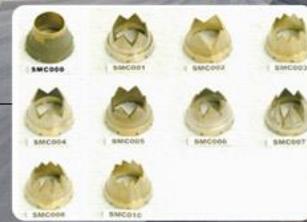
TRL 4-5 (1995-1997)

- Model tests for acoustics and aerodynamics
- Sub-scale model tests



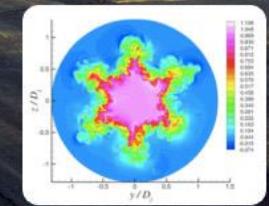
TRL 3 (Early 1990s)

- Applications to small nozzles and airfoils
- Lab tests, concept on paper



TRL 1-2 (1980s)

- Fundamental investigations of air-mixing devices (tabs, chevrons, etc.)
- No specific application, basic research in fluid physics



Pre-
definitions

Data Science

What

Who

Why

When

How

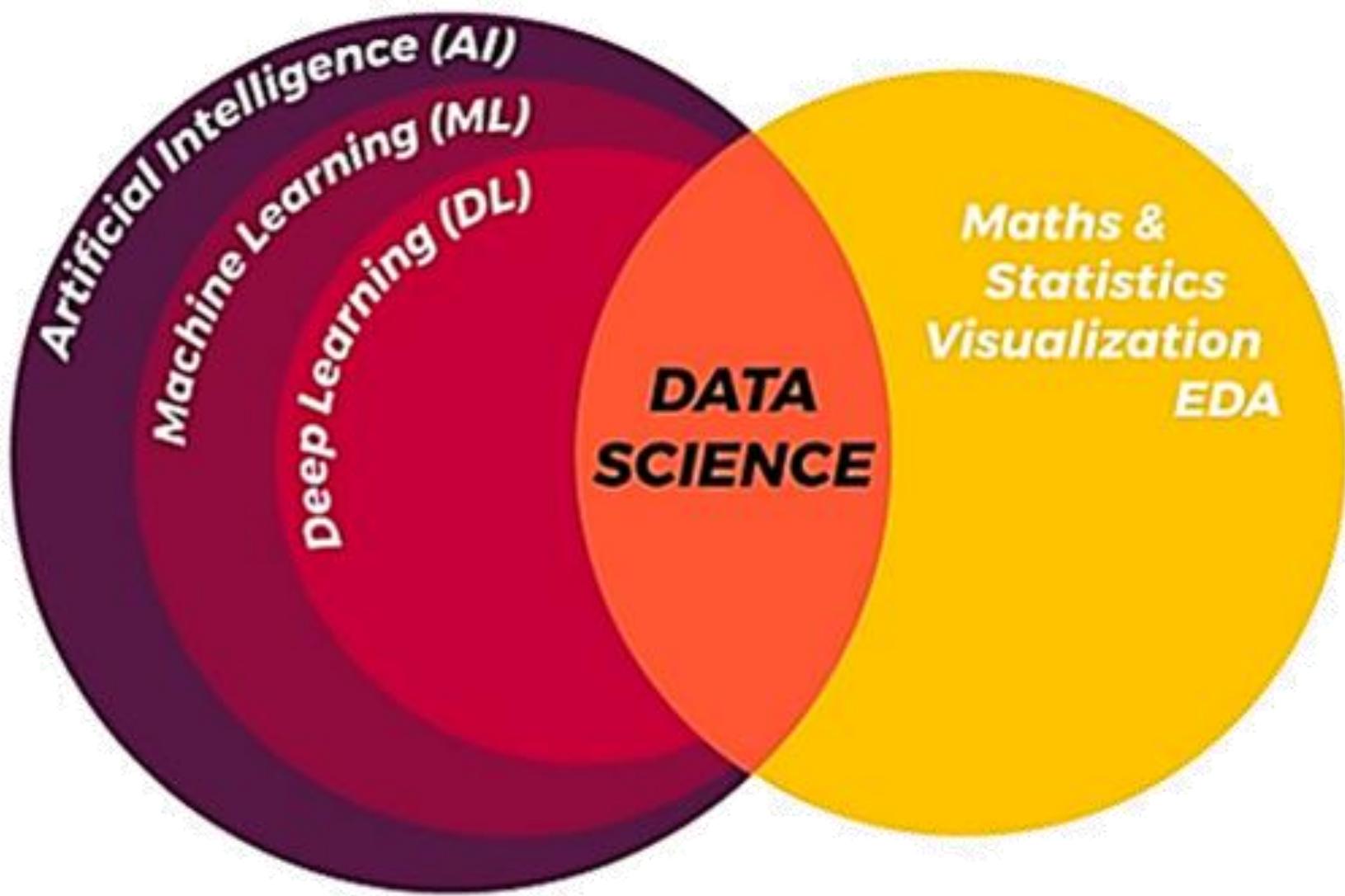
What is Data Science?

Interdisciplinary field focused on extracting knowledge and insights buried in data and develop advanced tasks.



MATHEMATICS STATISTICS PROGRAMMING
DATA ANALYSIS INFORMATICS
STORYTELLING

What is Data Science?

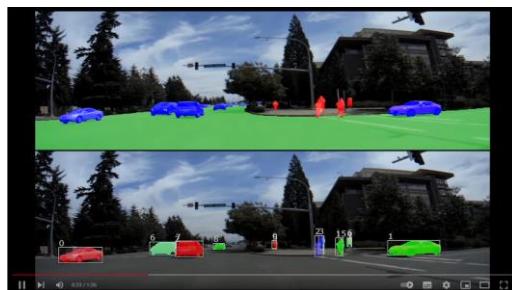
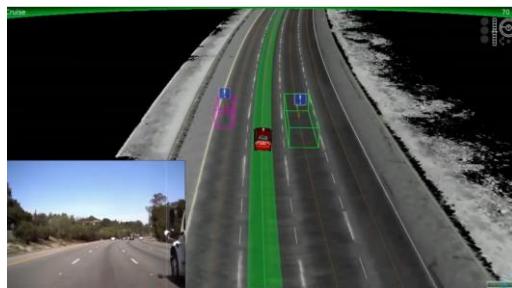


Examples

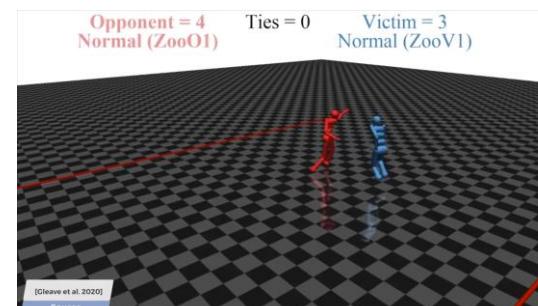
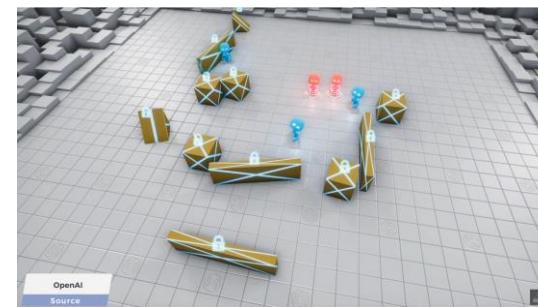
Deep fake



Autonomous driving

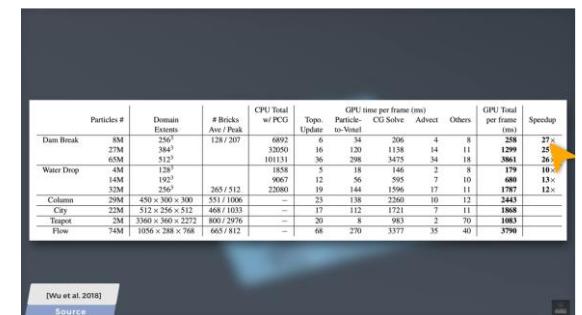
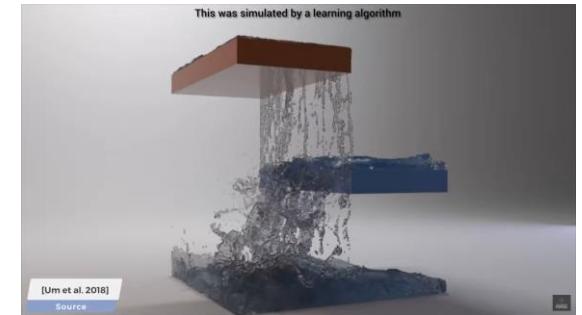


Reinforcement Learning



SOPHY in GT

Physics Simulation



[Wu et al. 2018]
Source

**Pre-
definitions**

Data Science

What

Who

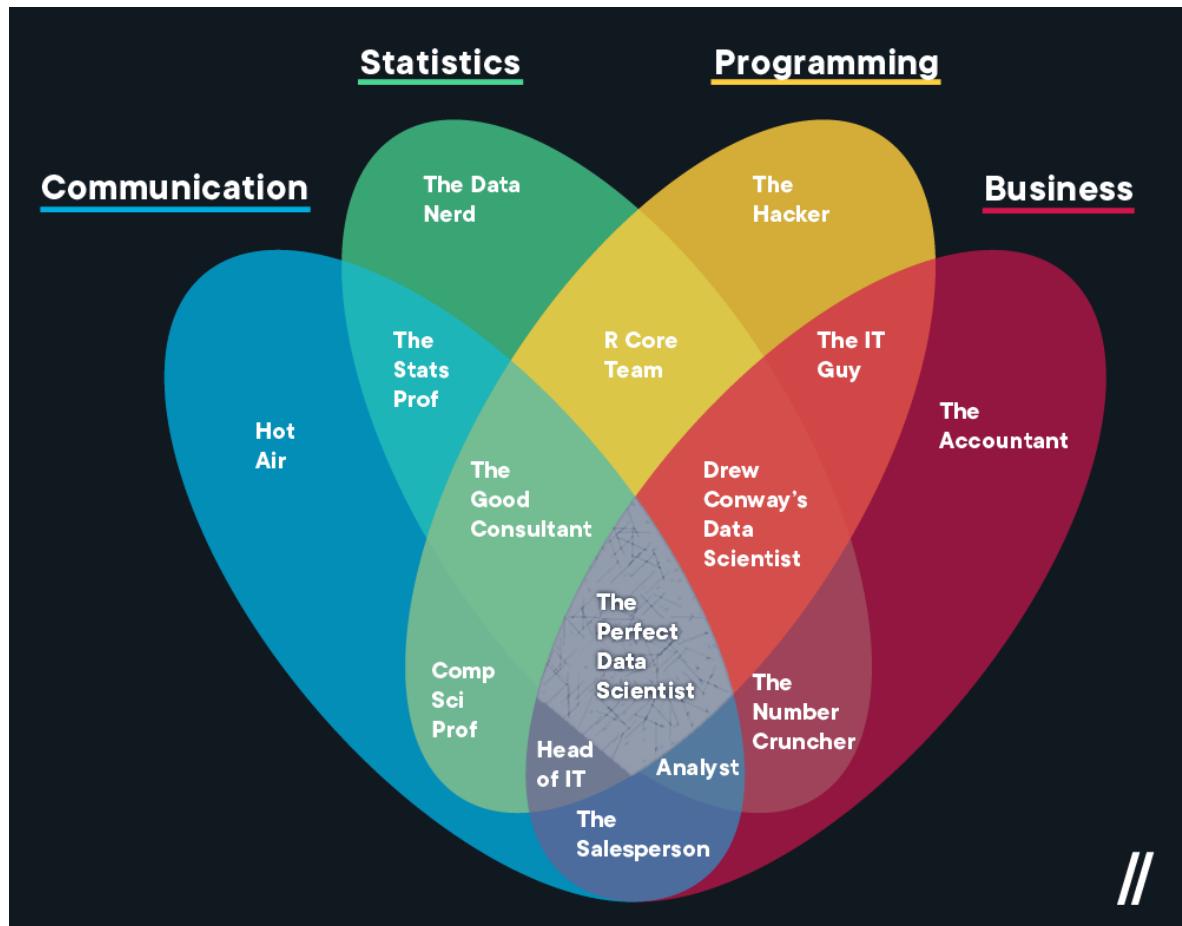
Why

When

How

What is a Data Scientist?

Data scientists analyse, process, and model data to find trends and uncover solutions.
Then they communicate the results.



Hard Skills

Maths
Statistics
Programming

Soft Skills

Curiosity
Story telling
Teamwork
Humbleness

Field knowledge
New technologies
Avid learner

//

What is a Data Scientist?

Data Science is not owned by Data Scientists!

Data Science is a very big and diverse field.

Humbleness is a must in Data Science as many people may know more than you in specific methods.

What is a Data Scientist?

Data Scientists create *data models*

What is a Data Scientist?

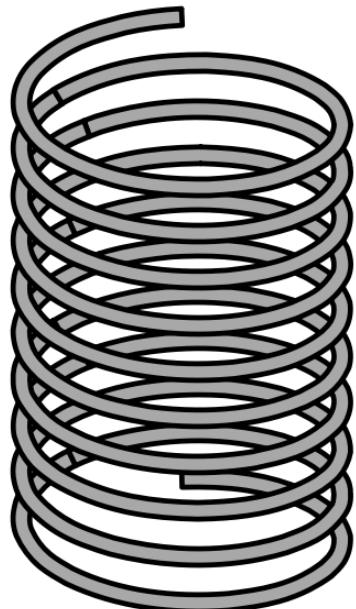
Data Scientists create *data models*

A model is a virtual representation of the behaviour of a system in a subdomain of the variables space. It maps inputs and outputs.

What is a Data Scientist?

Data Scientists create *data models*

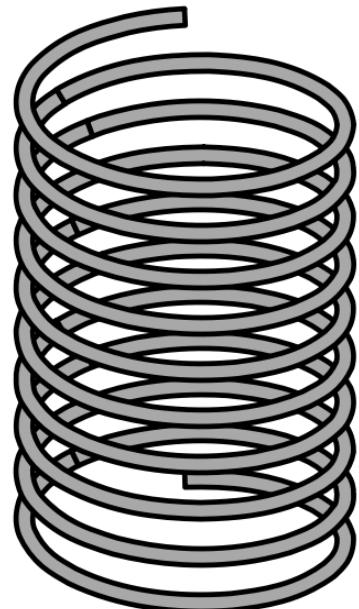
A model is a virtual representation of the behaviour of a system in a subdomain of the variables space. It maps inputs and outputs.



What is a Data Scientist?

Data Scientists create *data models*

A model is a virtual representation of the behaviour of a system in a subdomain of the variables space. It maps inputs and outputs.



Hooke Law

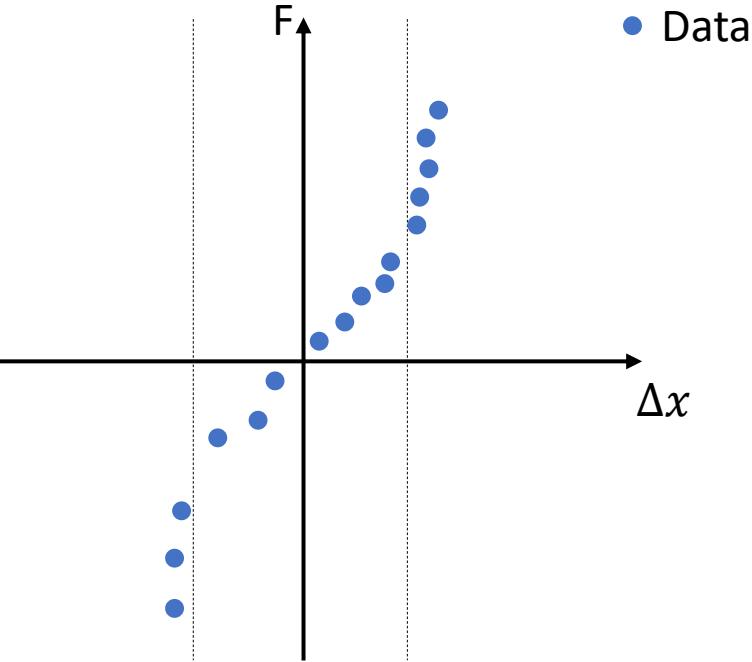
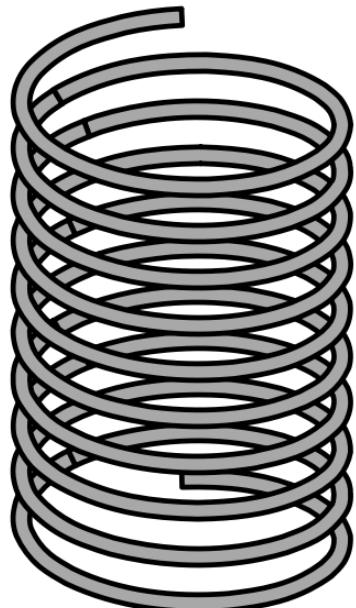
$$F = k\Delta x$$

(in the linear range)

What is a Data Scientist?

Data Scientists create *data models*

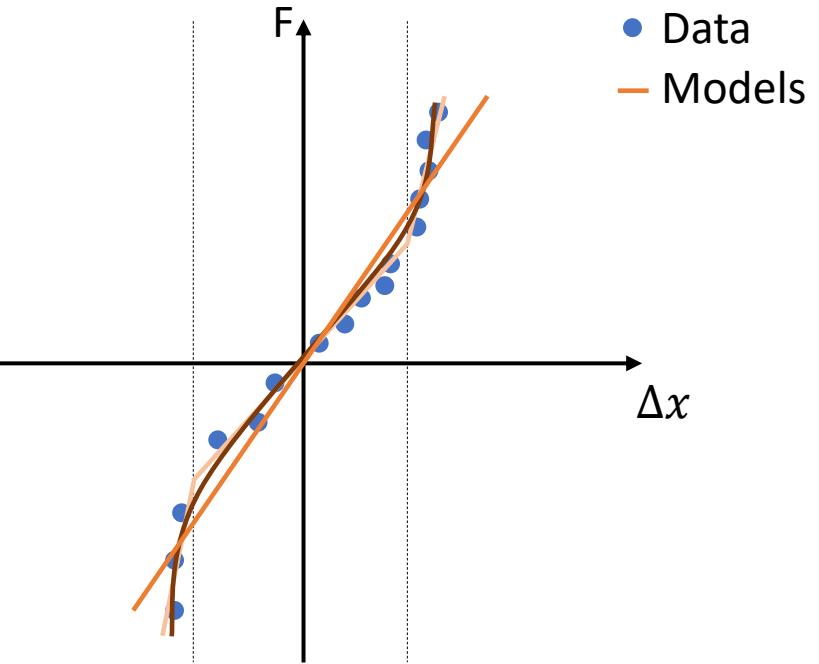
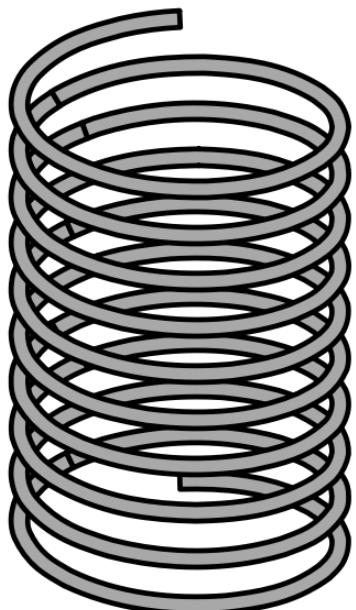
A model is a virtual representation of the behaviour of a system in a subdomain of the variables space. It maps inputs and outputs.



What is a Data Scientist?

Data Scientists create *data models*

A model is a virtual representation of the behaviour of a system in a subdomain of the variables space. It maps inputs and outputs.

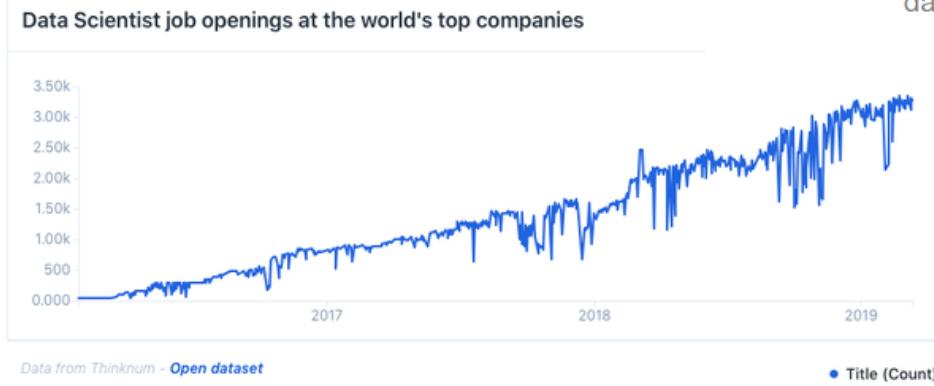


What is a Data Scientist?

Analytics And Data Science

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil



Pre-
definitions

Data Science

What

Who

Why

When

How

Why data, Why now?

Sensors are everywhere

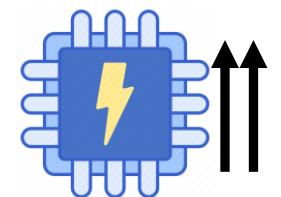
(Volume of data doubles every three years)



**Data storage and computing power has increased,
cost has plummeted**



Algorithms are advancing

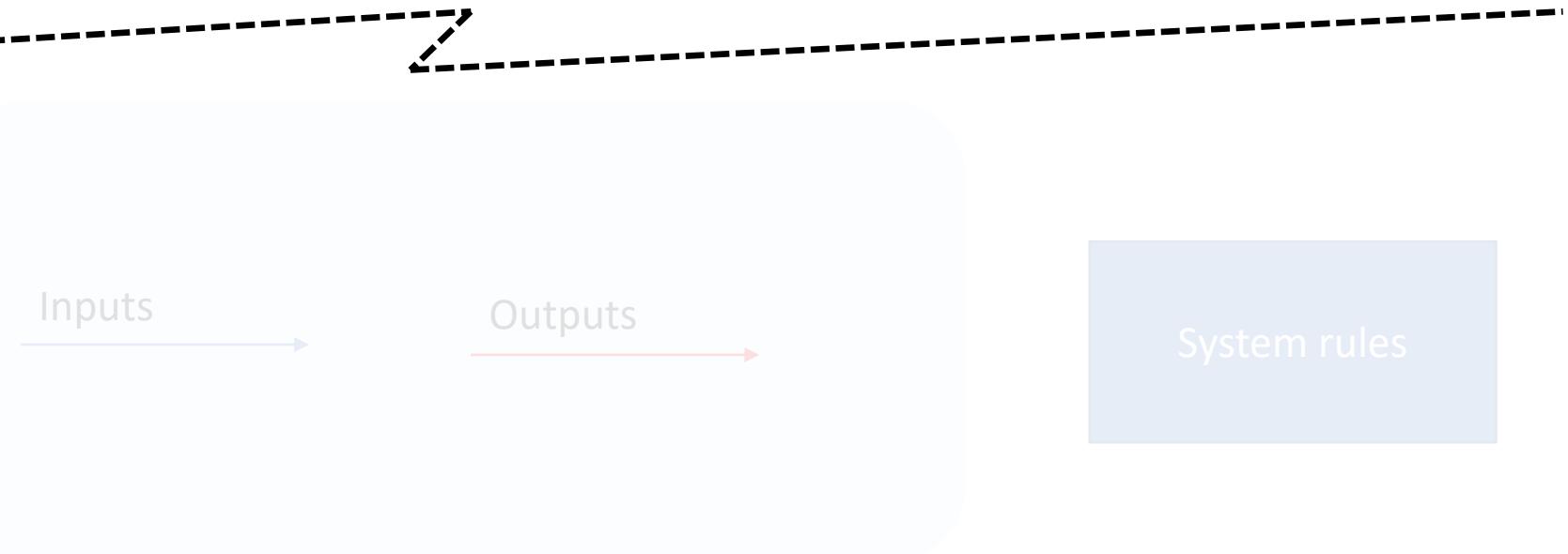


Data Paradigm

**Classic
models**

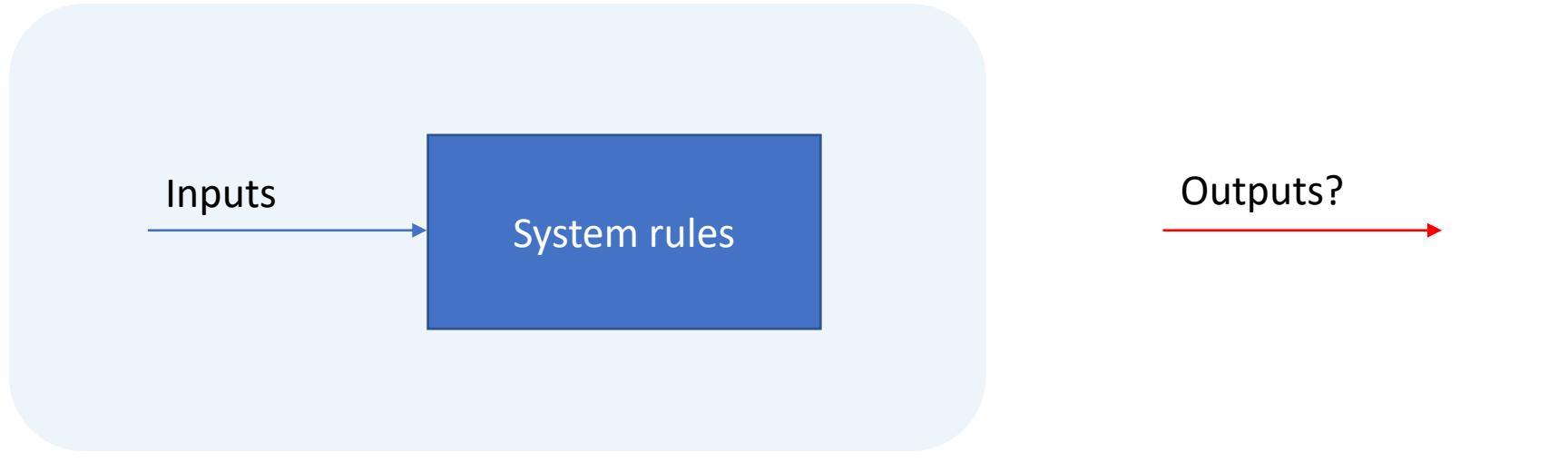


**Data
models**

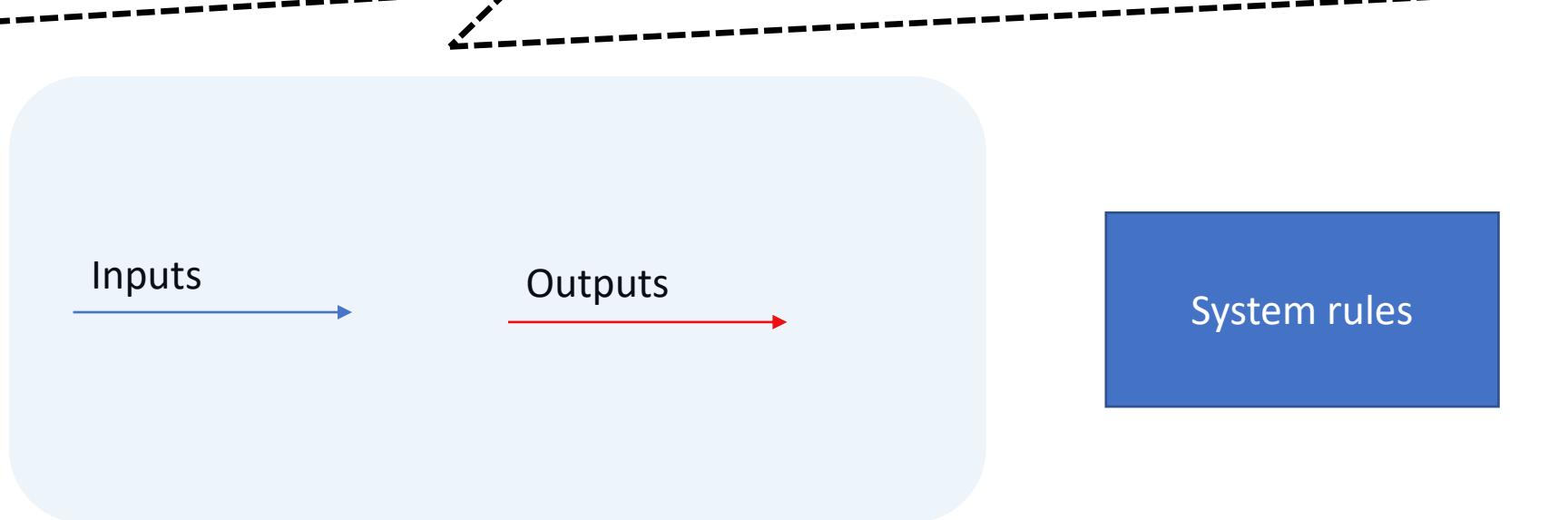


Data Paradigm

**Classic
models**



**Data
models**



**Pre-
definitions**

Data Science

What

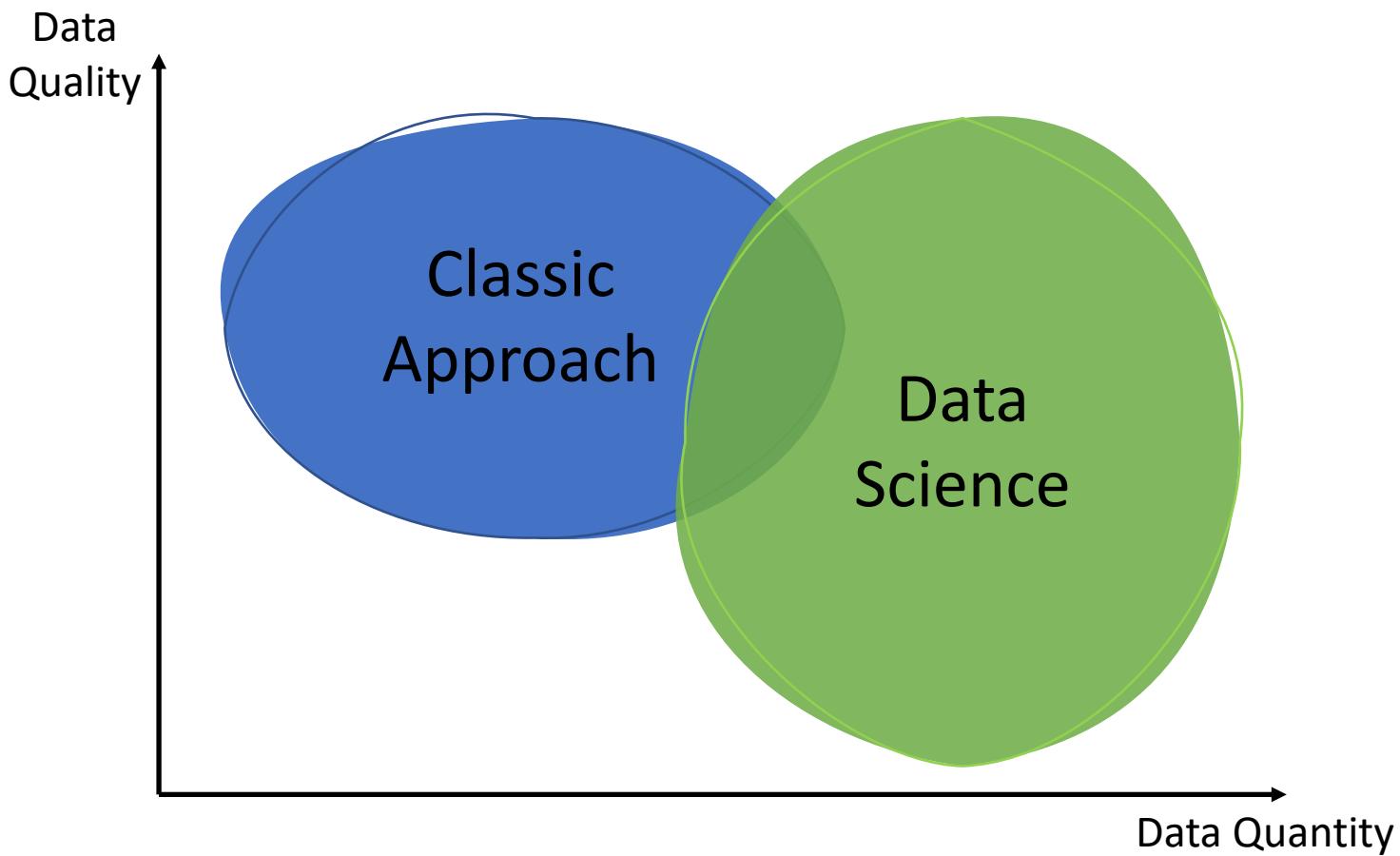
Who

Why

When

How

When Data Science?



Rules of ML (by Google)

Before ML	Rule 01: Go for simple heuristics first. Rule 02: Design and implement metrics. Rule 03: Choose ML over complex heuristics. Rule 04: Keep the first model simple and get the infrastructure right.
First Pipeline	Rule 07: Turn heuristics into features. Rule 13: Choose a simple metric for your first objective Rule 16: Plan to launch and iterate.
Feature Engineering	Rule 19: Use very specific features when you can. Rule 23: You are not a typical end user. Rule 24: Measure the delta between models.
Growth	Rule 43: ...

Pre- definitions

Data Science

What

Who

Why

When

How

Context

Toolkit

Project

Data in F1

Programming

Pre- definitions

Data Science

What

Who

Why

When

How

Context

Toolkit

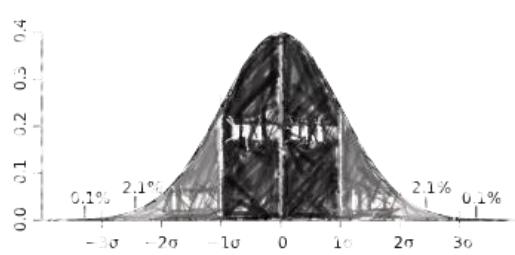
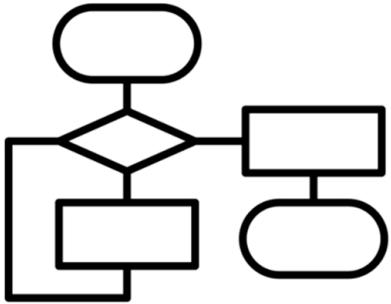
Project

Data in F1

Programming

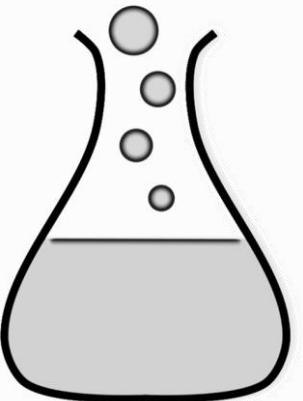
Data Analyst

Overview/pre-analysis



Data Scientist

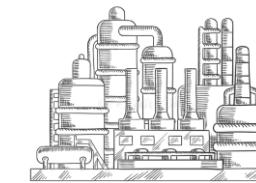
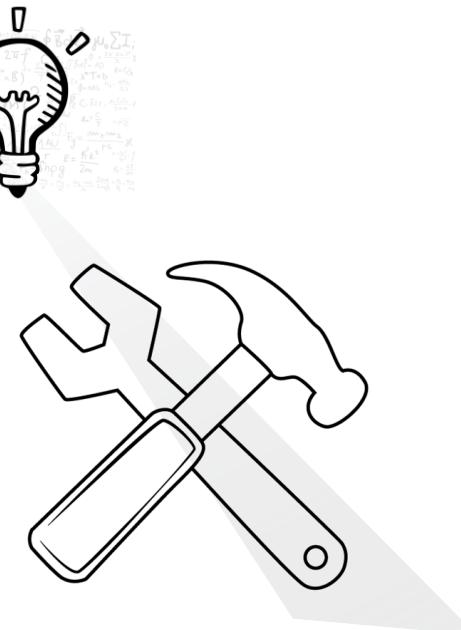
Algorithms/models



$$\begin{aligned} Q &= mc\Delta t \quad R = \frac{U}{I} \quad k = \pm \sqrt{\frac{2m}{h^2}}(E - h) \quad \oint \vec{B} d\ell = \mu_0 \sum I; \\ \beta &= \frac{\Delta I_c}{\Delta I_B} \quad E = \frac{1}{2}h\nu_m \quad \omega = 2\pi f \quad C_p = \frac{2\pi^2 \sin 2\lambda}{\lambda} g \\ f_0 &= \frac{1}{2\pi\sqrt{KL}} \quad S = \frac{1}{g\mu_0} (\vec{E} \times \vec{B}) \quad \hat{J} = \frac{eN_2}{T} \quad \vec{P} = \iint \vec{B} dS = AD \quad V = C/A \\ R &= \rho \frac{L}{S} \quad \vec{F}_v = S \frac{F_n}{R} \quad E = mc^2 \quad \vec{J} = \frac{eN_2}{T} \quad \Phi = NBS \quad H_A = \frac{\Delta M_e}{\Delta A} \\ k &= \frac{1}{4\pi G r^2} \quad E = \hbar\omega \quad R_m = \frac{C}{T} \quad \vec{F}_n = \frac{1}{2\pi} \int \vec{P} \cdot \vec{B} d\ell = \frac{\mu_0 I_1 I_2}{2\pi d} \cdot \ell \\ V &= \frac{1}{2\pi r m_e} \quad \sigma = \frac{Q}{T} \quad M_e = \sigma T^4 \quad \vec{F}_m = \frac{C}{T} \quad \nu = \frac{1}{2\pi r m_e} \quad \Omega = \frac{Q}{S} \quad I_m = \int_{\text{in}}^{\text{out}} \left[\frac{1}{r^2} + \left(\frac{1}{x_c} - \frac{1}{x_l} \right)^2 \right] F_g = \frac{m_1 m_2}{r^2} \cdot \delta \\ M &= \vec{F}_d \cos \alpha \quad T = \frac{4\pi n_e}{(m_1 m_2)^2} \quad F_h = S \rho g \quad E = \frac{\hbar^2 k^2}{2m} \cdot \delta \quad \phi_e = \frac{\Delta E}{\Delta t} \\ \oint \vec{D} dS &= Q \quad F_x = \frac{1}{2} C_p \rho g^2 \cdot \frac{m_1}{x} + \frac{m_2}{x_l} = \frac{m_2 - m_1}{r} \quad \sin \alpha = \frac{V}{L} \cdot \frac{m_2}{m_1} \quad \sin \beta = \frac{V}{L} \cdot \frac{m_1}{m_2} \\ S &= N_m = \frac{Q}{\rho e} \cdot \frac{M_m}{N_A} \quad F_y = \frac{F_x}{\sqrt{1 - \frac{F_x^2}{F_z^2}}} \quad F_z = \frac{F_x}{\sqrt{1 - \frac{F_x^2}{F_z^2}}} \end{aligned}$$

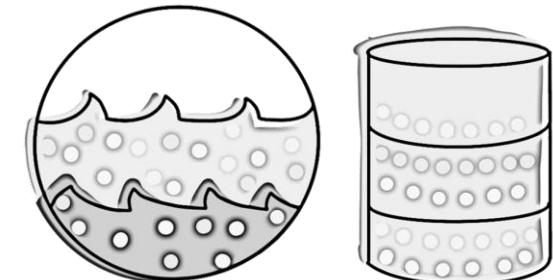
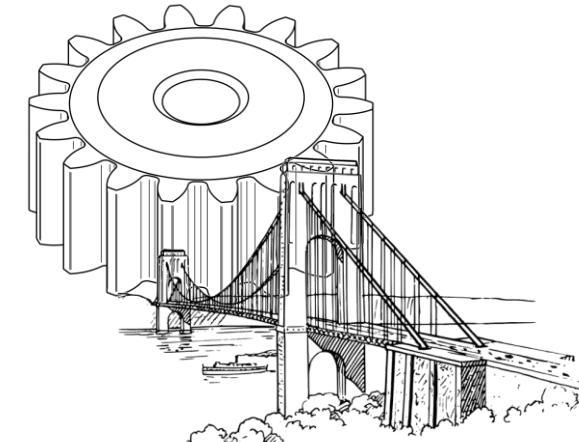
ML Engineer

Implementation



Data Engineer

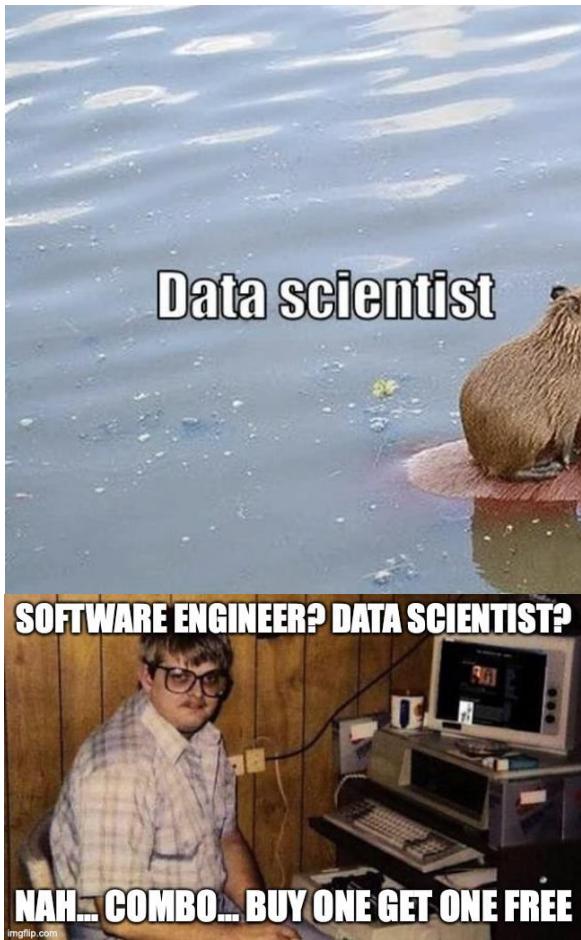
Infrastructure



Technology stack

Data Analyst

Overview/pre-analysis



Data Scientist

Algorithms/models

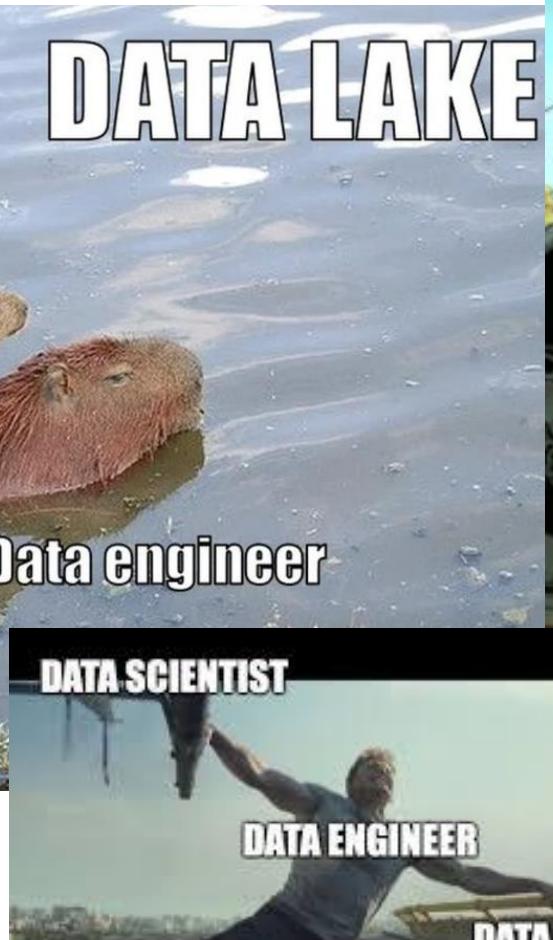
DATA LAKE

Data engineer

DATA SCIENTIST

DATA ENGINEER

DATA PLATFORM



ML Engineer

Implementation



Data Engineer

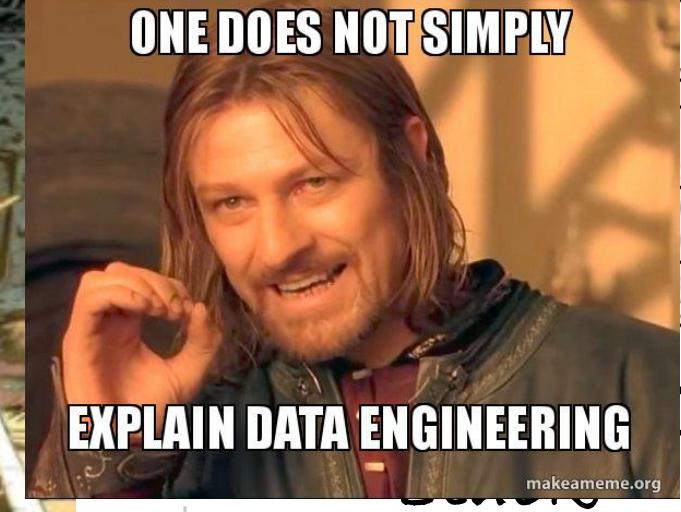
Infrastructure

I HAVE DATA, I HAVE AN

ENGINEER

パンハナッポーアップオープン
UHH...DATA ENGINEER?

ONE DOES NOT SIMPLY



Structure



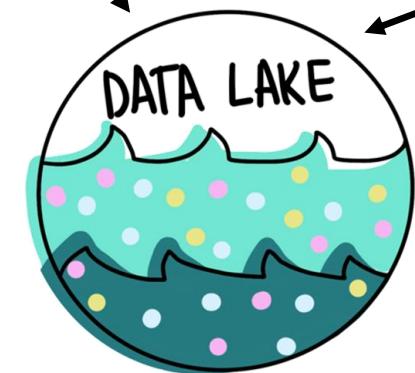
Unstructured Data

→
Data Mining



Structured Data

Storage



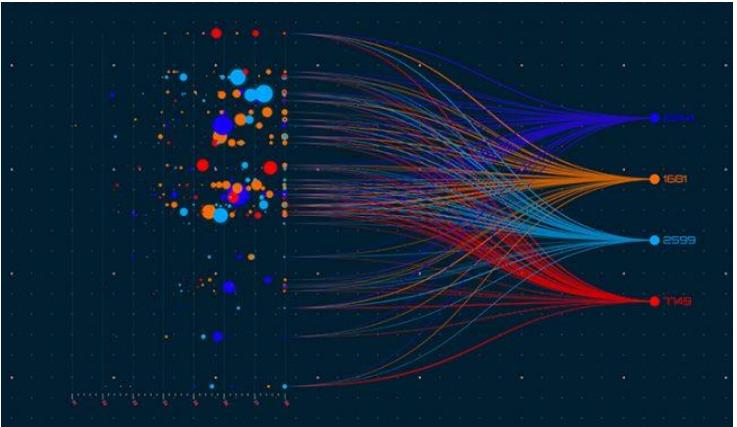
Size

BIG DATA

SMALL DATA

Data Science Work Flow

DATA



Data Science Work Flow

DATA

- Structured**
 - Databases (SQL, non SQL ...)
 - CSV, JSON files
 - Proprietary file formats
- Unstructured**
 - Websites
 - Videos, pictures, audio...



Data Science Work Flow

DATA

- Structured**
 - Databases (SQL, non SQL ...)
 - CSV, JSON files
 - Proprietary file formats
- Unstructured**
 - Websites
 - Videos, pictures, audio...



Data Science Work Flow

DATA

- Structured**
 - Databases (SQL, non SQL ...)
 - CSV, JSON files
- Proprietary file formats**
- Unstructured**
 - Websites
 - Videos, pictures, audio...



Exploratory
Data
Analysis



Data Science Work Flow

DATA

- **Structured**
Databases (SQL, non SQL ...)
CSV, JSON files
Proprietary file formats
- **Unstructured**
Websites
Videos, pictures, audio...



Exploratory
Data
Analysis

Feature
Engineering



Data Science Work Flow

DATA

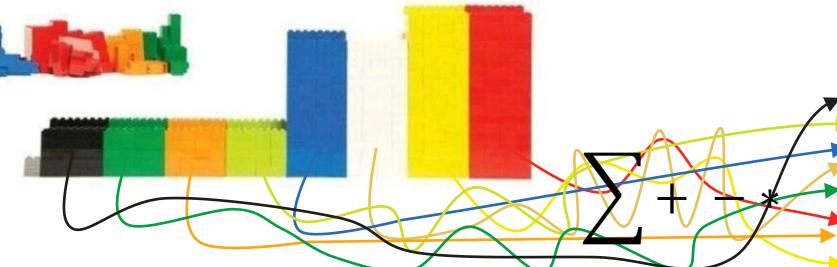
- Structured**
Databases (SQL, non SQL ...)
CSV, JSON files
- Unstructured**
Proprietary file formats
Websites
Videos, pictures, audio...



Exploratory
Data
Analysis

Feature
Engineering

Modelling



Data Science Work Flow

DATA

- **Structured**
Databases (SQL, non SQL ...)
CSV, JSON files
- **Proprietary file formats**
- **Unstructured**
Websites
Videos, pictures, audio...

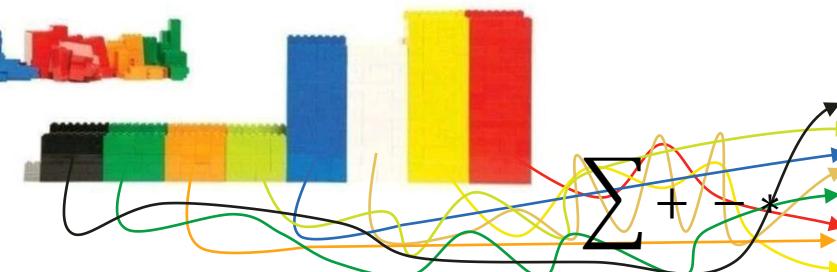


Exploratory
Data
Analysis

Feature
Engineering

Modelling

Deploying



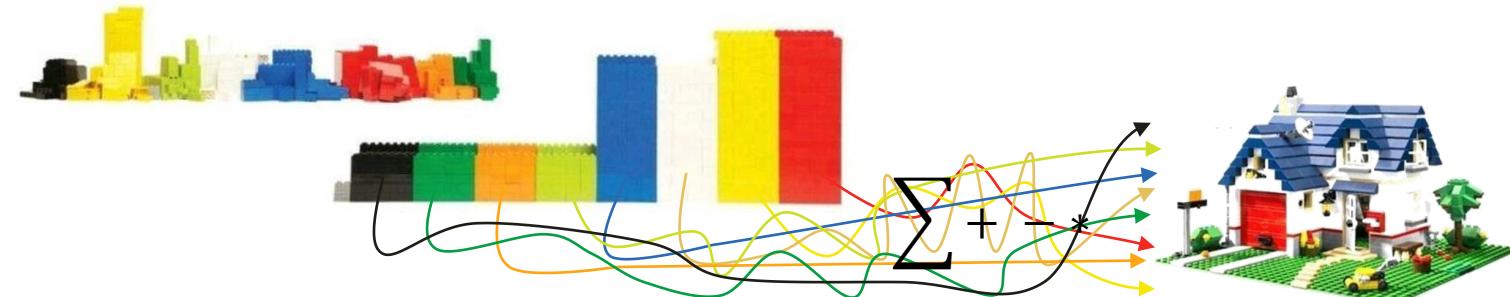
Data Science Work Flow

DATA

- Structured**
Databases (SQL, non SQL ...)
CSV, JSON files
- Unstructured**
Proprietary file formats
Websites
Videos, pictures, audio...



Exploratory
Data
Analysis



Feature
Engineering

Modelling

Deploying

Pre- definitions

Data Science

What

Who

Why

When

How

Context

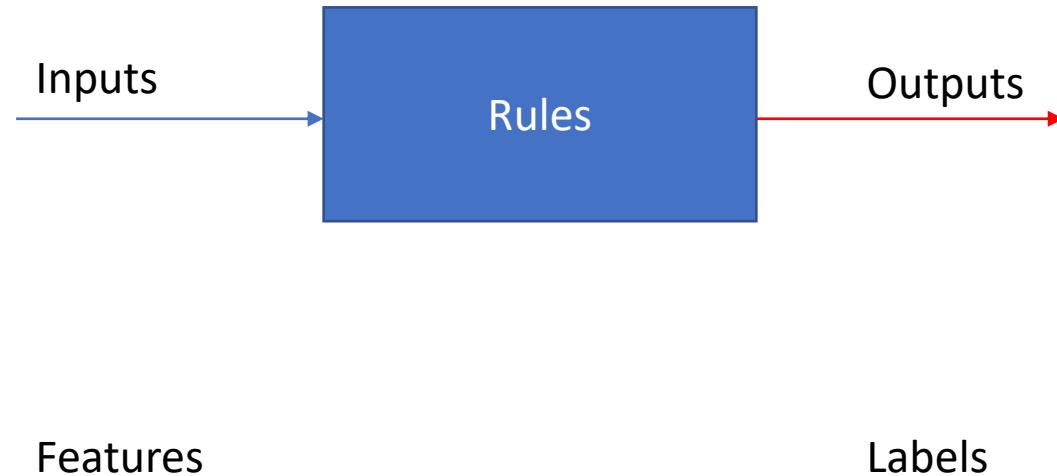
Toolkit

Project

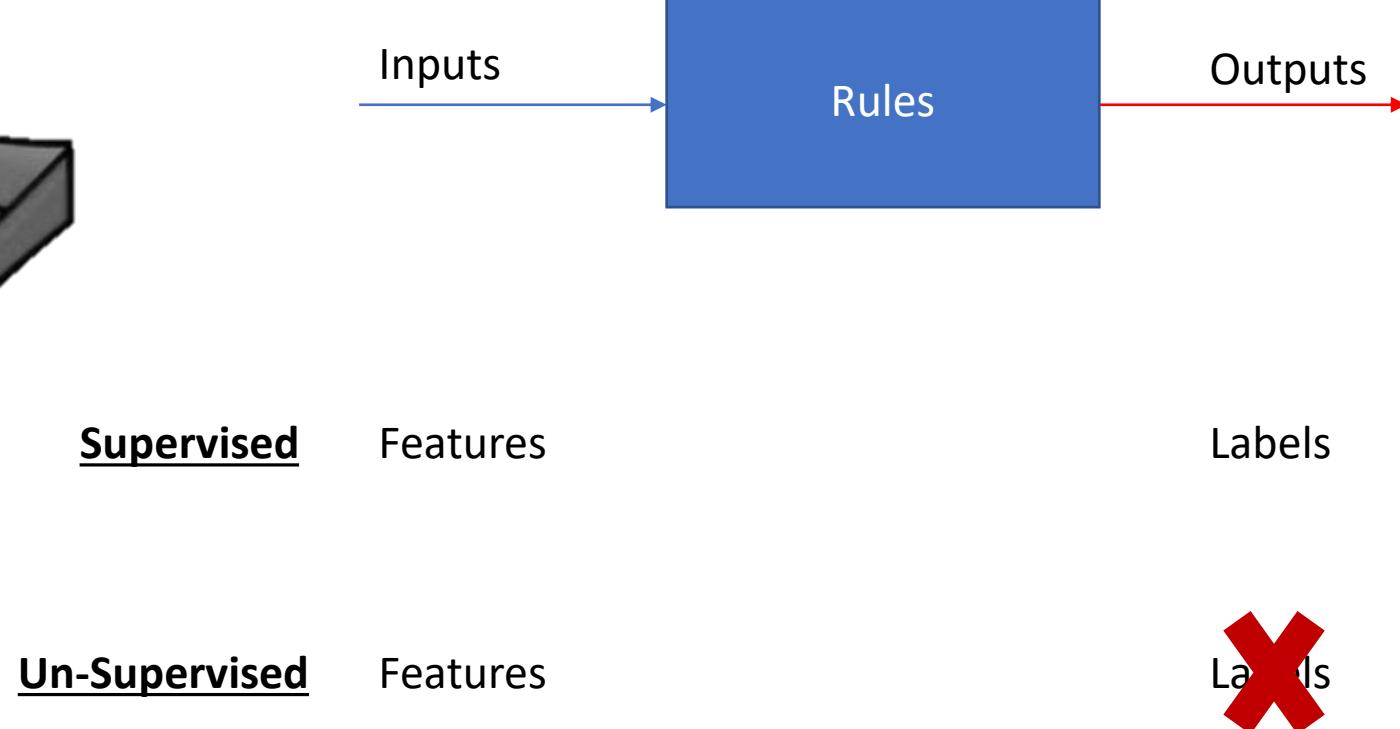
Data in F1

Programming

Modelling Toolkit



Modelling Toolkit



Modelling Toolkit

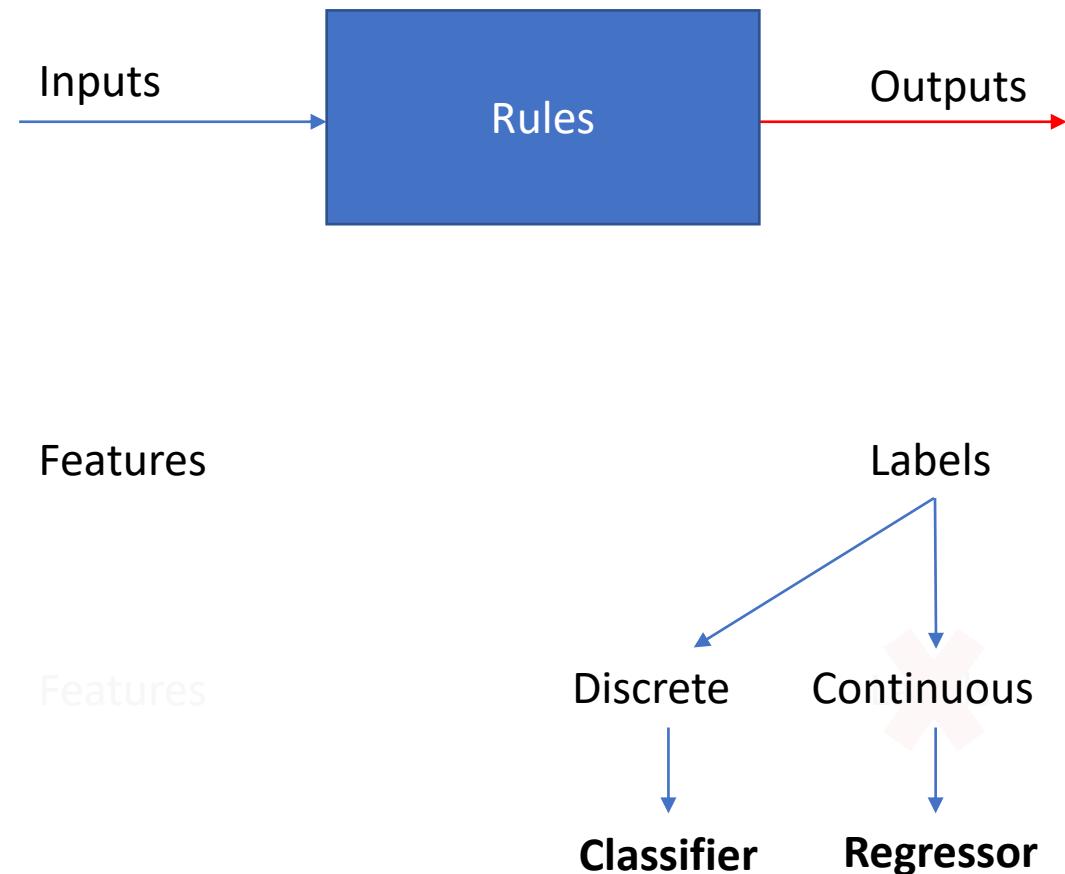


Supervised

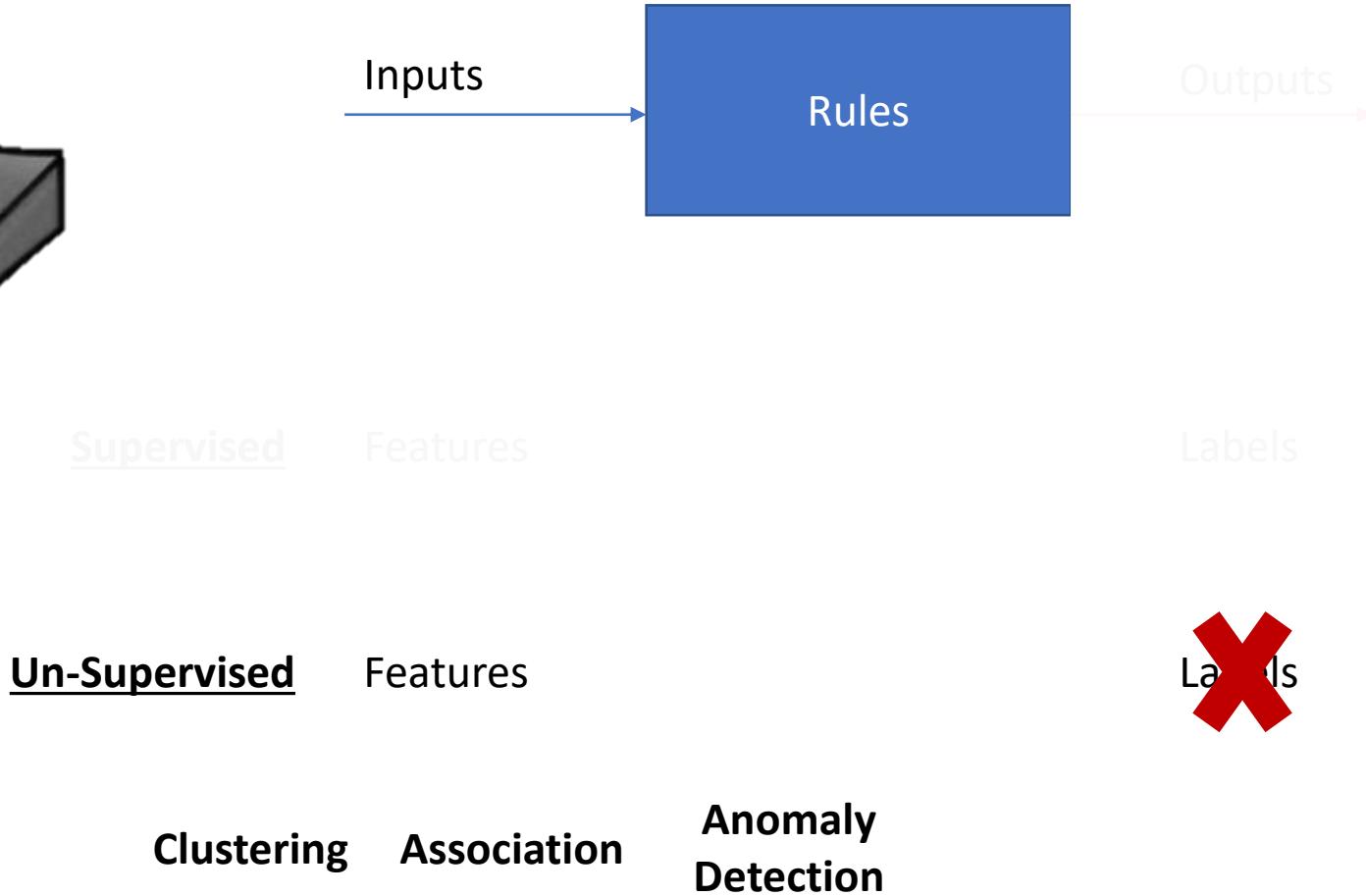
Un-Supervised

Features

Features



Modelling Toolkit



MACHINE LEARNING

SUPERVISED LEARNING

REGRESSION

LINEAR

RIDGE/
LASSO

SVR

MULTI

POLY

DECISION
TREE

CLASSIFICATION

LOGISTIC
REGRESSION

DECISION TREE

NAÏVE BAYES

SVM

UNSUPERVISED LEARNING

CLUSTERING

K-MEANS

HIERERCHICAL

ASSOCIATION

APRIORI

ECLAT

ANOMALY DETECTION

DENSITY-
BASED

SVM

CLUSTERING

SVM

REINFORCEMENT LEARNING

MARKOV DECISION
PROCESS

UPPER BOUND
CONFIDENCE

THOMSON
SAMPLING

DQN

DEEP LEARNING

MACHINE LEARNING

SUPERVISED LEARNING

REGRESSION

LINEAR

RIDGE/
LASSO

SVR

MULTI

POLY

DECISION
TREE

CLASSIFICATION

LOGISTIC
REGRESSION

DECISION TREE

NAÏVE BAYES

SVM

• • •

UNSUPERVISED LEARNING

CLUSTERING

K-MEANS

HIERERCHICAL

• • •

ASSOCIATION

APRIORI

ECLAT

• • •

ANOMALY DETECTION

DENSITY-
BASED

SVM

CLUSTERING

SVM

• • •

REINFORCEMENT LEARNING

MARKOV DECISION
PROCESS

UPPER BOUND
CONFIDENCE

THOMSON
SAMPLING

DQN

• • • Course

DEEP LEARNING

MACHINE LEARNING

SUPERVISED LEARNING

REGRESSION

LINEAR

RIDGE/
LASSO

NN

POLE
BALANCER

TREE

CLASSIFICATION

LOGISTIC
REGRESSION

CNN

SVM

...

...

UNSUPERVISED LEARNING

CLUSTERING

K-MEANS

HIERARCHICAL

...

ASSOCIATION

APRIORI

ECLAT

...

ANOMALY DETECTION

DENSITY-
BASED

SVM

CLUSTERING

SVM

...

...

REINFORCEMENT LEARNING

MARKOV DECISION
PROCESS

UPPER BOUND
CONFIDENCE

THOMSON
SAMPLING

DQN

...

...

DEEP LEARNING

Course

Pre- definitions

Data Science

What

Who

Why

When

How

Context

Toolkit

Project

Data in F1

Programming

System to study

Sensors study

Installation of
sensors

Interface

Storing the data

Data Science
Project

System to study

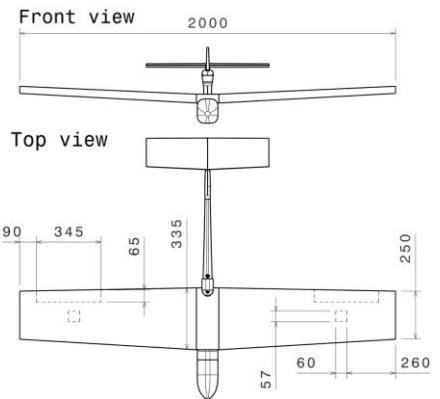
Sensors study

Installation of
sensors

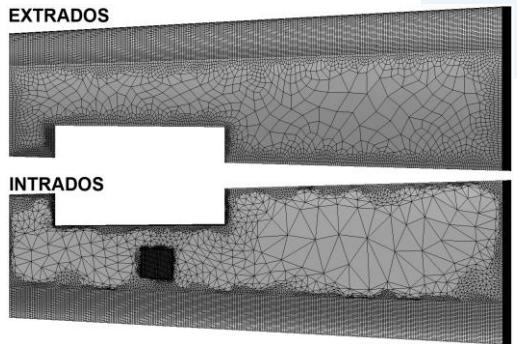
Interface

Storing the data

Data Science
Project

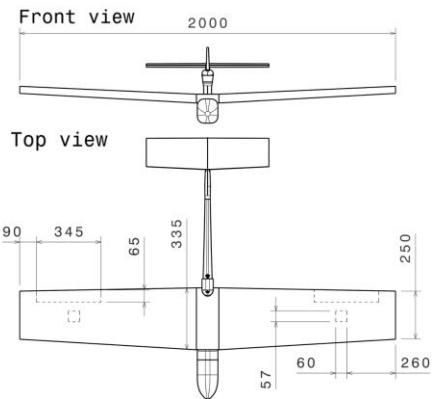


System to study



Sensors study

Installation of
sensors

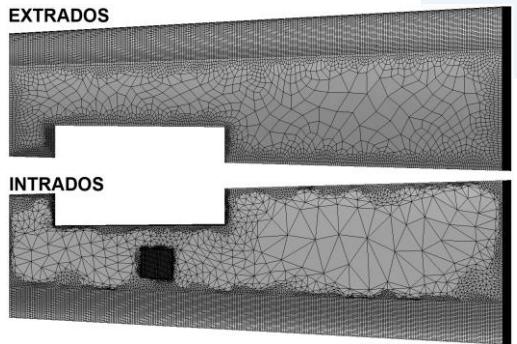


Interface

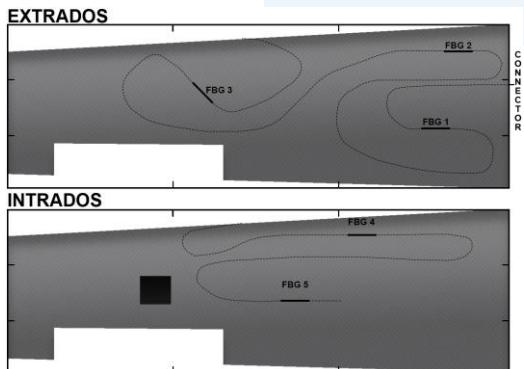
Storing the data

Data Science
Project

System to study

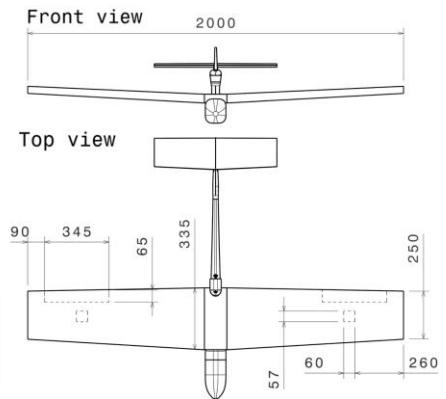


Sensors study



Load cell

Installation of
sensors

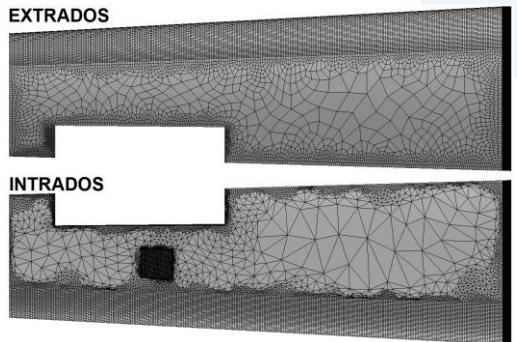


Interface

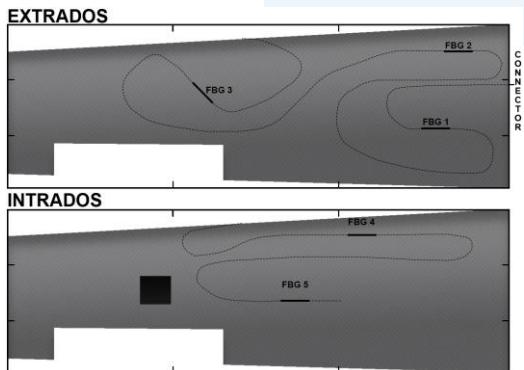
Storing the data

Data Science
Project

System to study

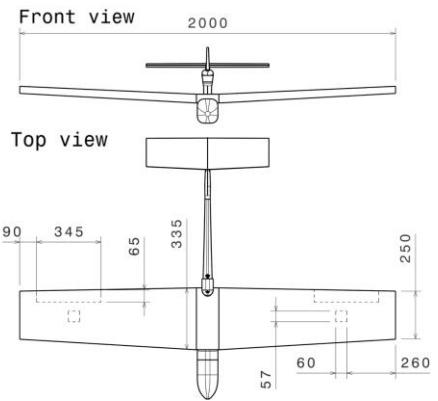


Sensors study



Load cell

Installation of
sensors



Interface



Storing the data

Data Science
Project



DATA IS THE NEW OIL

ROB CROOKE

Senior Vice President and General Manager,
Non-Volatile Memory (NVM) Solutions Group

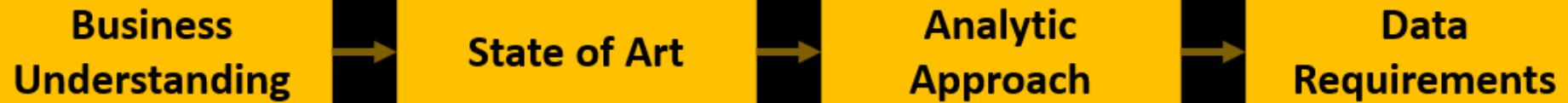


Standard
Tape 33\$ TB
Disk 45\$ TB

Storing the data

Data Science
Project

ARQUITECTURE



DEPLOYMENT & MAINTENANCE

Retrospective

Feedback

Deployment

Evaluation

IMPLEMENTATION

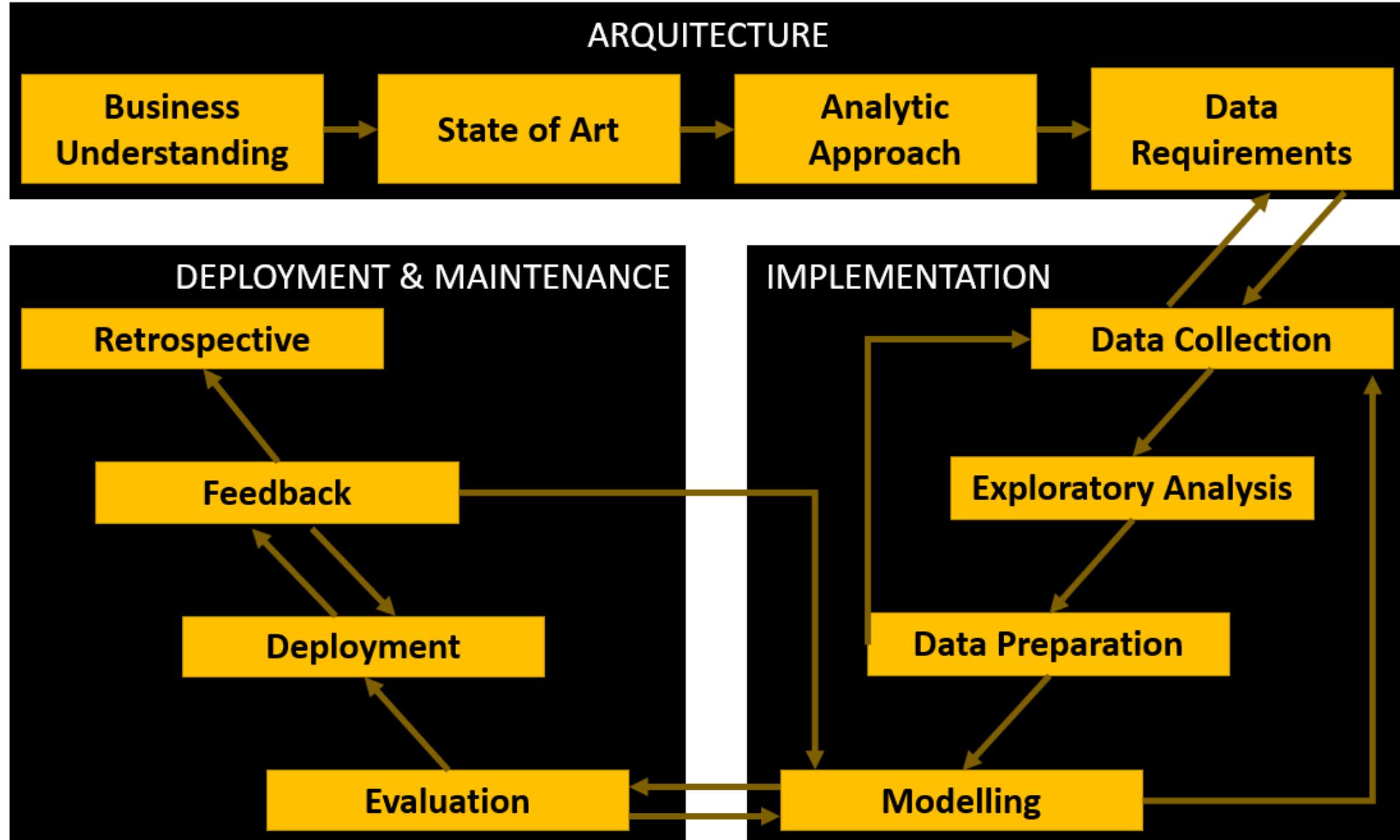
Data Collection

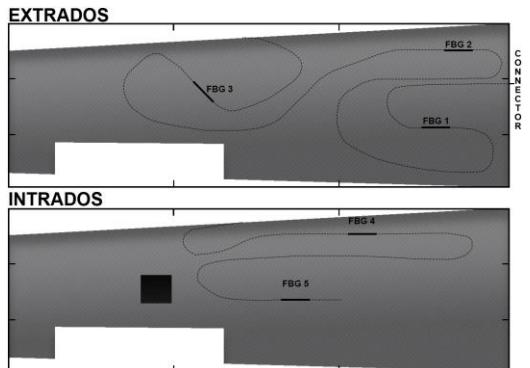
Exploratory Analysis

Data Preparation

Modelling

Data Science
Project





Storing the data



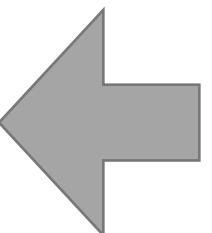
Load cell

- Alarms (thresholds) for load and strains
- Fatigue analysis
- Impact detection
- Dynamic analysis (aeroelasticity?)
- Distribution of the loads

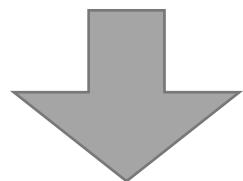
- Regression models load-strain.
- Clustering of manoeuvres.
- Patterns repeated in time.
- Anomaly detection (Sensor error or dangerous manoeuvres)
- Classification of structural state.

- ...

Improvement of the product



Refinement of design and calculation,
reduce of costs and improvement of
performances



Pre- definitions

Data Science

What

Who

Why

When

How

Context

Toolkit

Project

Data in F1

Programming

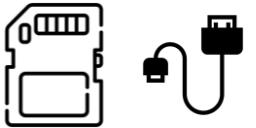
Data in F1



\approx (O) 100 sensors

\approx (O) 10Gb per weekend

Data in F1

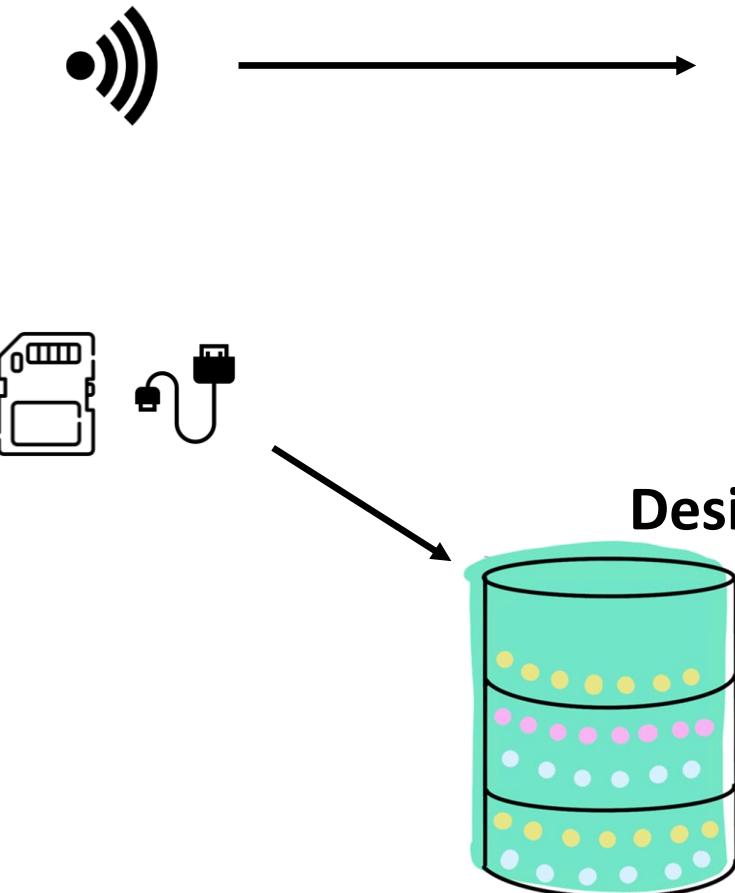


≈ (O) 100 sensors
≈ (O) 10Gb per weekend

Data in F1



≈ (O) 100 sensors
≈ (O) 10Gb per weekend



Trackside
engineers

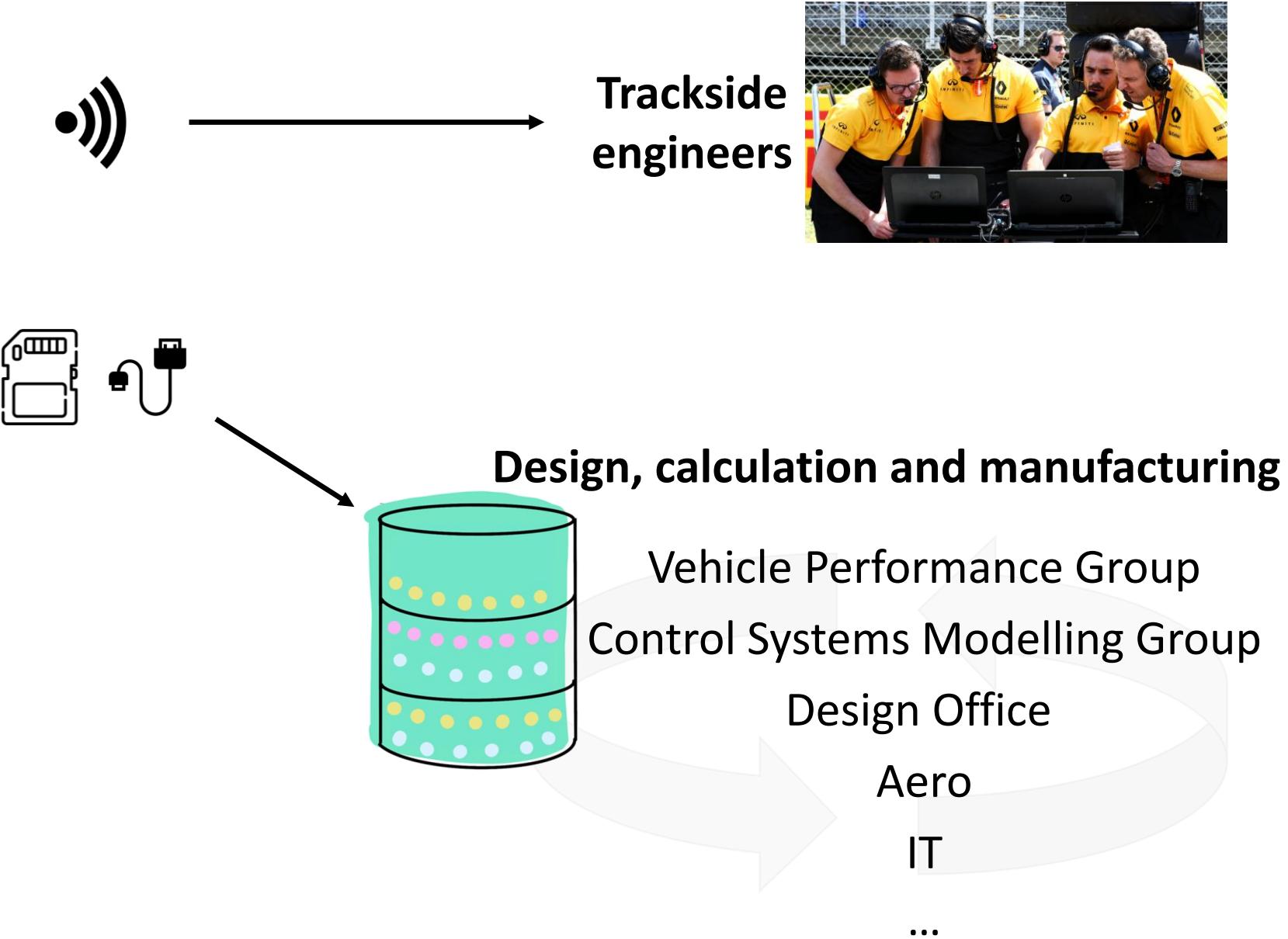


Design, calculation and manufacturing

Data in F1



≈ (O) 100 sensors
≈ (O) 10Gb per weekend



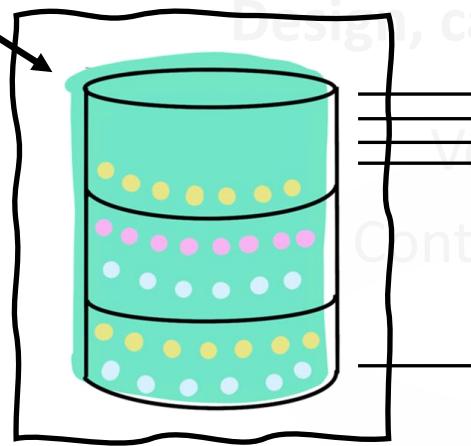
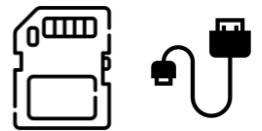
Data in F1



≈ (O) 100 sensors
≈ (O) 10Gb per weekend



Trackside
engineers



Design, calculation and manufacturing

Classic Models

Control Systems Modelling Group

Design Office

Data Models

Aero

IT

...

Data in F1

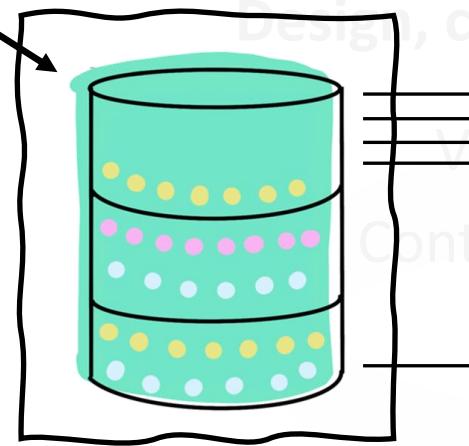
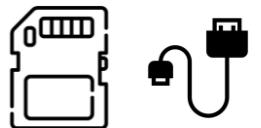


Trackside
engineers



≈ (O) 100 sensors
≈ (O) 10Gb per weekend

**Data Science is exploding in F1,
stay tuned if you are
interested!!**



Design, calculation and manufacturing

Classic Models

Control Systems Modelling Group

Design Office

Data Models

Aero

IT

...

Pre- definitions

Data Science

What

Who

Why

When

How

Context

Toolkit

Project

Data in F1

Programming

Python for Data Science



Bare Python

Python from python.org

Select an IDE

Link the IDE to the Python interpreter



Great for Data Science
GUI for environments
Easy import-export of env
Easy integration with IDEs
It can be used for Jupyter as well



Nice interface to program and share.
Need Python first (Or docker)



Virtual environments



Python from python.org

pipx
pipenv

- Env per project_1
- Env per project_2
- Env per project_3

Python for Data Science



Bare Python

Python from python.org

Select an IDE

Link the IDE to the Python interpreter



ANACONDA®

Great for Data Science
GUI for environments
Easy import-export of env
Easy integration with IDEs
It can be used for Jupyter as well



QUESTIONS?

