



# KHANACADEMY

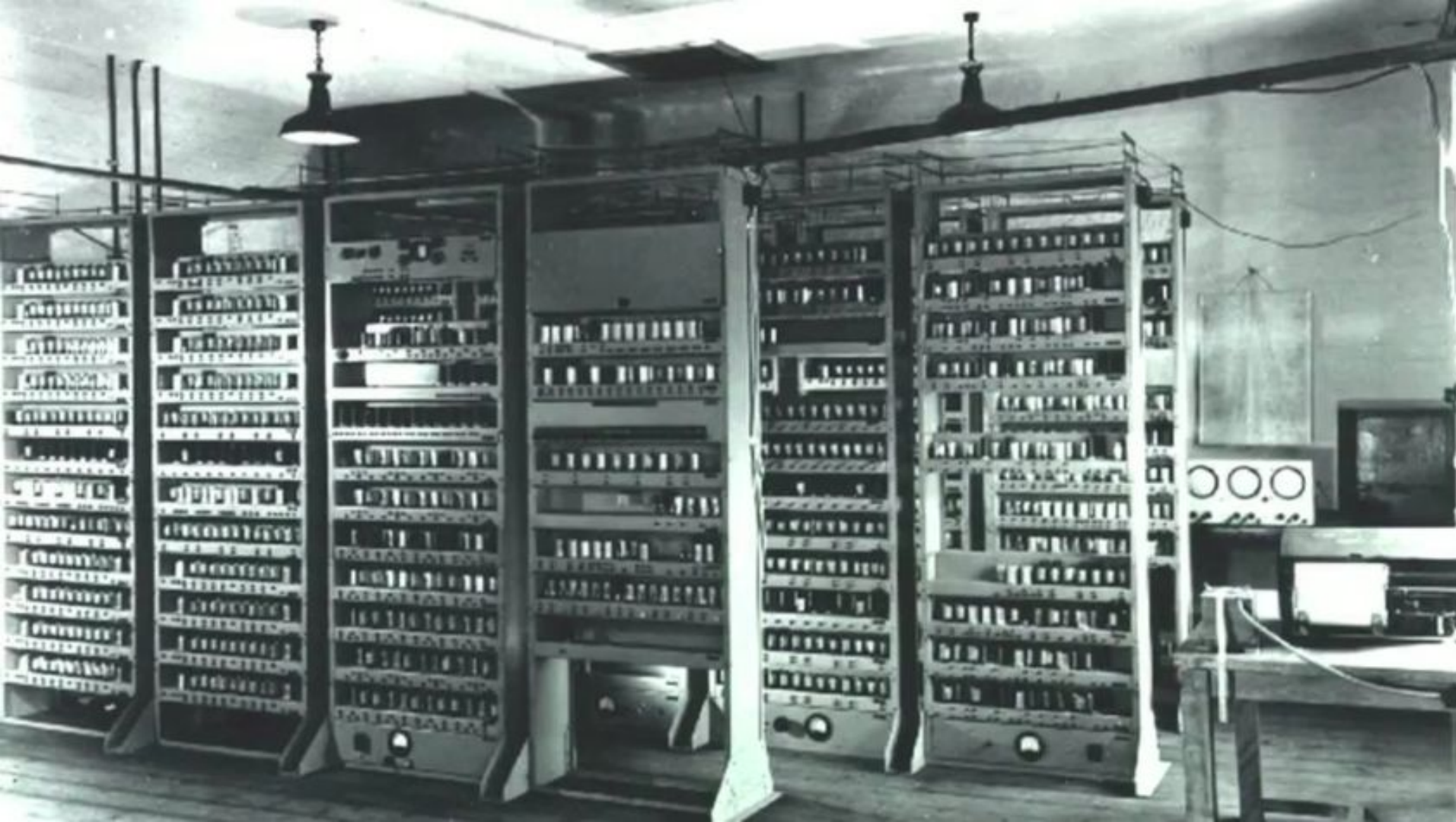


0:05 Nobody's born smart

1:08 Because the most beautiful, complex concepts in the whole universe are built on basic ideas

1:13 that anyone can learn, anywhere can understand. Whoever you are, wherever you are

1:18 You only have to know one thing: You can learn anything



May 11th, 1997  
**Computer won world champion of chess**  
(Deep Blue) (Garry Kasparov)



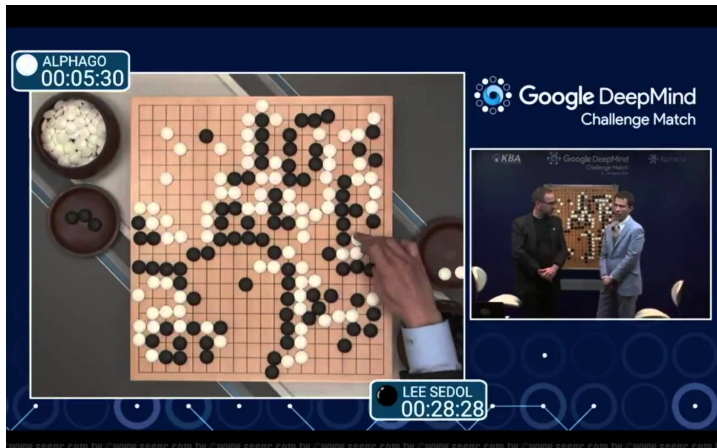
(Reuters = Kyodo News)



## Data Science

- 1952 - Tic Tac Toe  $\Rightarrow$  Human vs Computer
- 1997 - Deep Blue - Chess  $\Rightarrow$  Exploring Solution Space
- 2011 - Watson - Jeopardy  $\Rightarrow$  Constructive Reasoning
- 2017 - AlphaGo - Go  $\Rightarrow$  Developing Intuition

In AlphaGo, no. of possibilities  $>$  total no. atoms in this universe.







# Introduction to Data Science

June, 2017

# About Me

- Education

- 2012 Pass out, M.Sc. Information system - Bits, Pilani Rajasthan.
- Trained in RHEL 6, AIX Business Communications
- Certified Data Modelling Engineer.

- Software Engineer

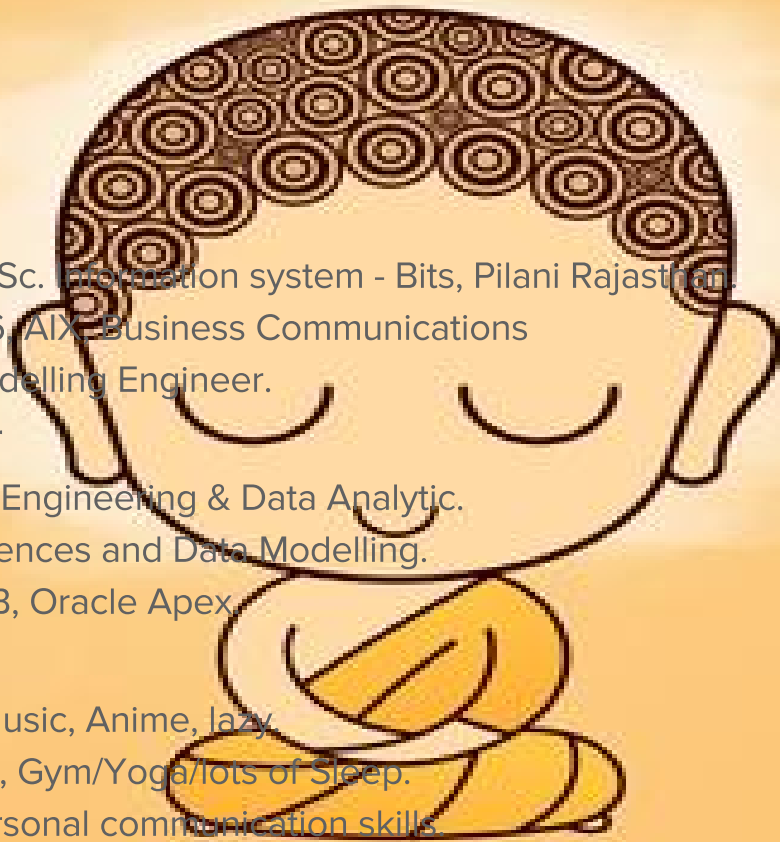
- 4.5 Years in Data Engineering & Data Analytic.
- 1 Year in Data Sciences and Data Modelling.
- Python, Oracle DB, Oracle Apex.

- Personal Life

- Teaching(blog), Music, Anime, lazy.
- Health Conscious, Gym/Yoga/lots of Sleep.
- Technology & Personal communication skills.

- Motivation:

- Bridge the gap between Technology and People. Lead a R&D Team.



# Plan

## Introduction

- **Definitions [ Data Science ]**
- **What, Why and How**
- **Examples**

## Data Science - In Action

- **Stages [DG, DC, DM, ME]**
- **Regression & Clustering Models**
- **Basics [ LR, GD ]**

## Real Life Application

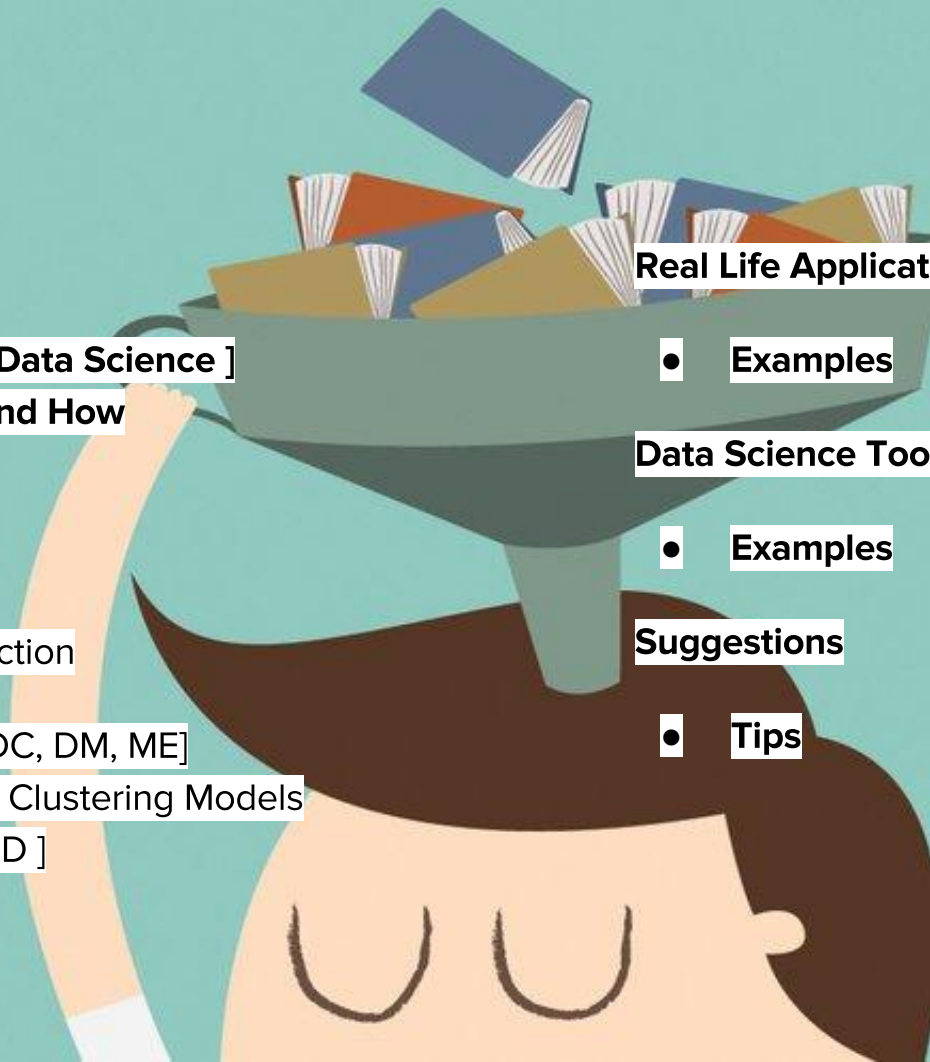
- **Examples**

## Data Science Tools

- **Examples**

## Suggestions

- **Tips**



# What is Data Science?

---

# What is Data Science?

da•ta

Factual information, especially information organized for analysis or used to reason or make decisions.

Computer Science Numerical or other information represented in a form suitable for processing by computer.

Values derived from scientific experiments.

sci·ence (sī'əns)

The **observation, identification, description, experimental investigation, and theoretical explanation of phenomena**. Ex. New advances in science and technology.

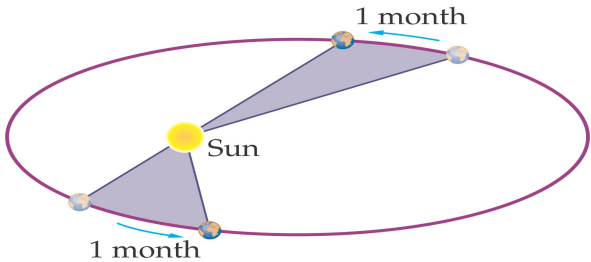
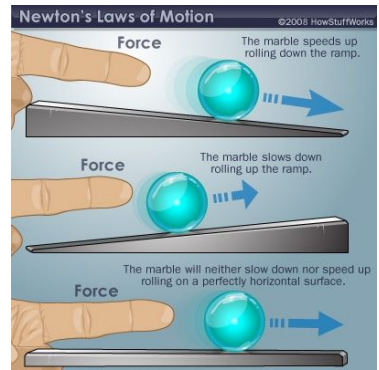
Such activities restricted to a class of natural phenomena. Ex. The science of astronomy.

A systematic method or **body of knowledge in a given area**. Ex. The science of marketing.

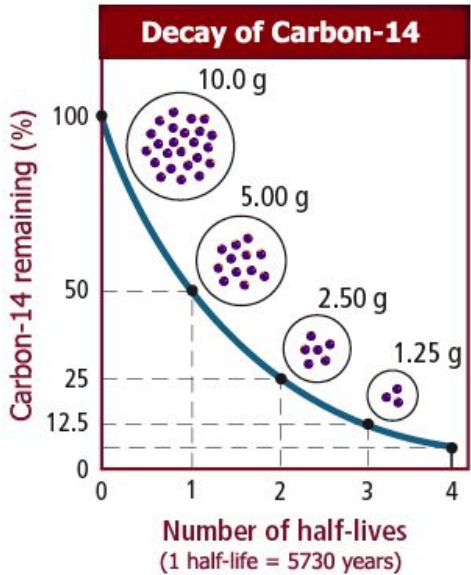
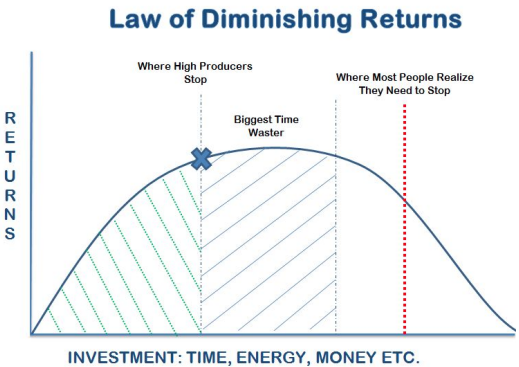
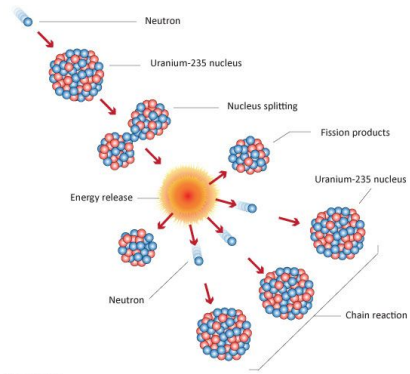
Archaic Knowledge, especially that gained through **experience**.



# Data Science Examples



(a)



# Why Data Science?

---

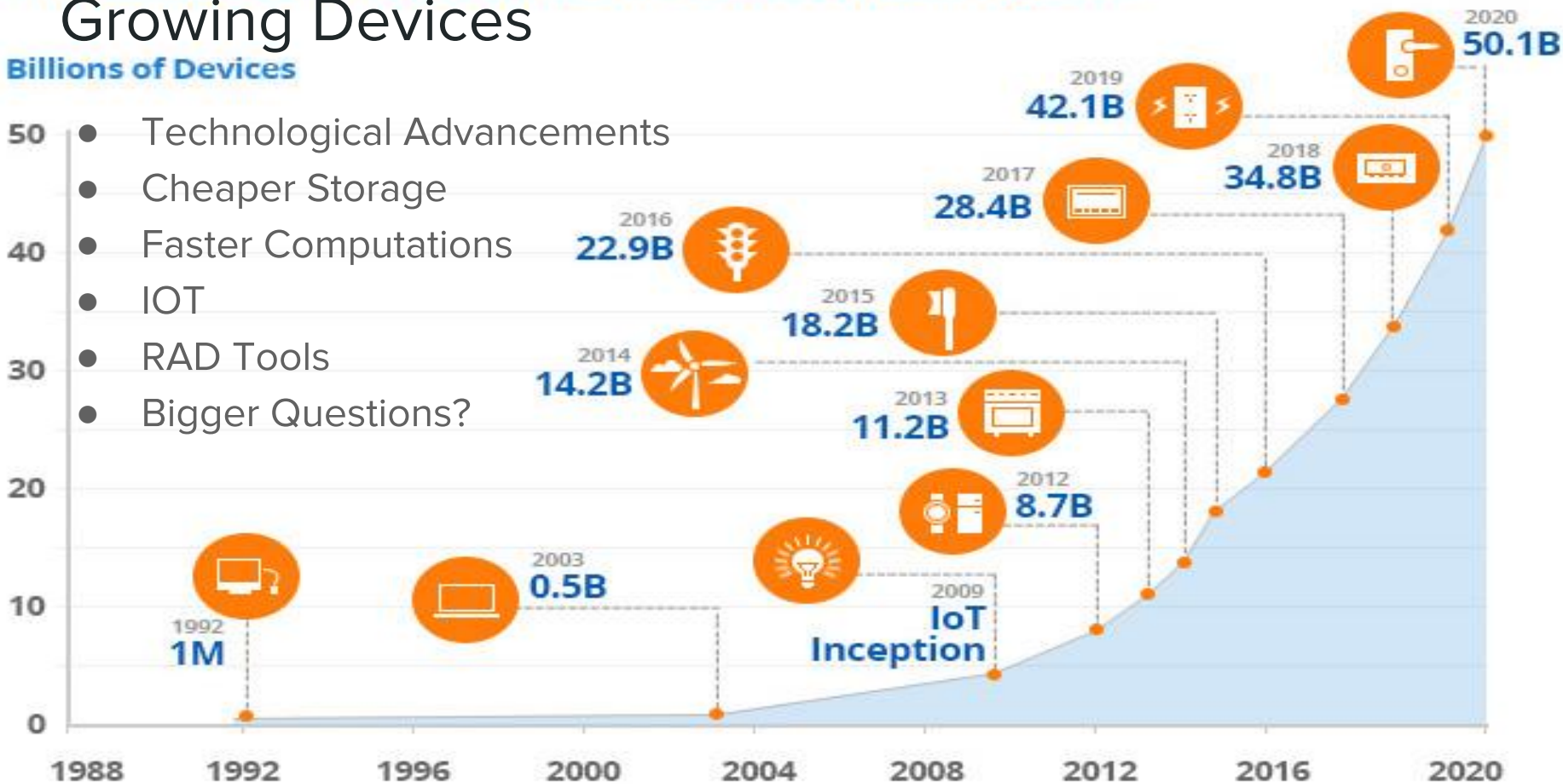
# Growth in the internet of things

The number of connected devices will exceed 50 billion by 2020

## Growing Devices

Billions of Devices

- Technological Advancements
- Cheaper Storage
- Faster Computations
- IOT
- RAD Tools
- Bigger Questions?



# Information Explosion & Doubling Processing Power

**Metcalfe's law** states that the value of a telecommunications network is proportional to the square of the number of connected users of the system ( $n^2$ ).

**Moore's law** is the observation that the number of transistors in a dense integrated circuit doubles approximately every two years.

(Population - Thanks to Advanced Medical Sciences & Improving Health Care.)

Sources: Wikipedia

# How to do Data Science?

---



# How to Data Science? - AI, ML



Rosey, Spacely, Jetson



MIT Cheetah Robot

# How to do Data Science

You can use lots of sophisticated analytical & Business Intelligent tools and come to a simple understandable explanations.

(or)

You can also use, simple tools like calculators or excel sheet to generate simple and simple results.

# Plan

## Introduction

- **Definitions [ Data Science ]**
- **What, Why and How**
- **Examples**

## Data Science - In Action

- **Stages [DG, DC, DM, ME]**
- **Regression & Clustering Models**
- **Basics [ LR, GD ]**

## Real Life Application

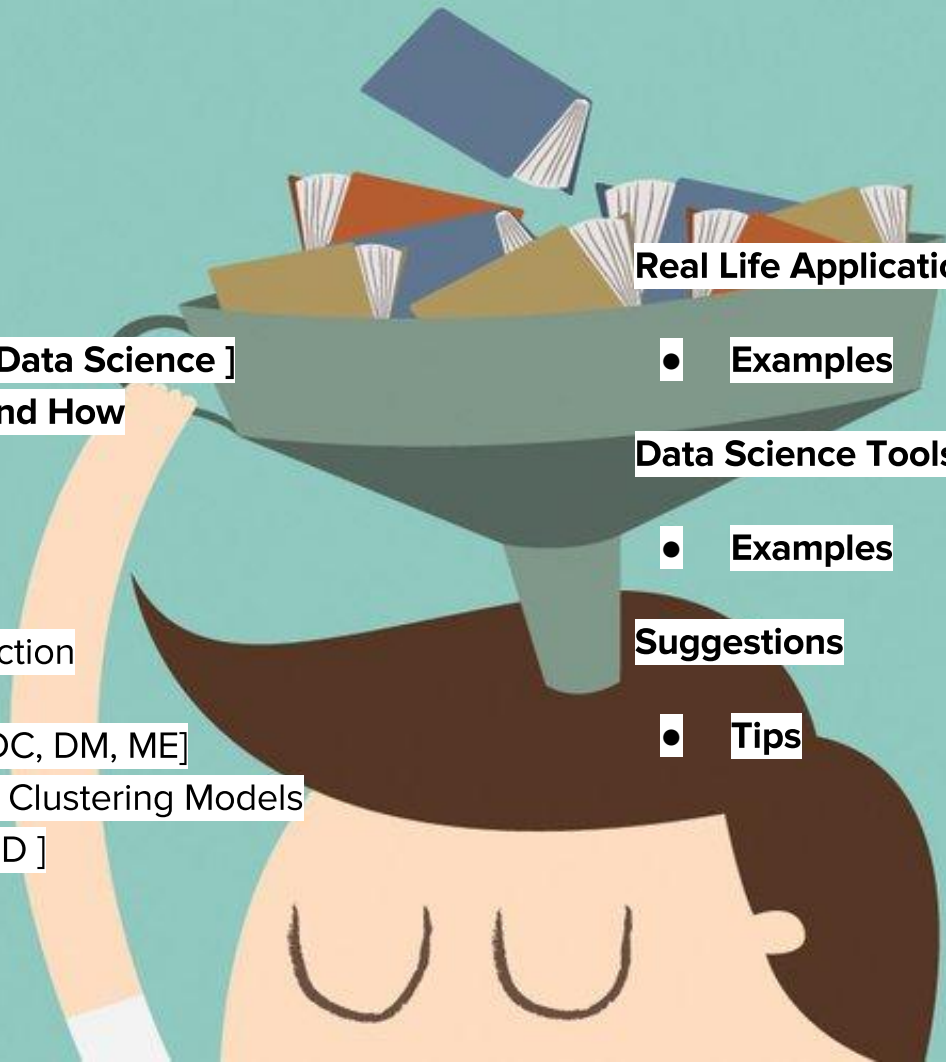
- **Examples**

## Data Science Tools

- **Examples**

## Suggestions

- **Tips**



# Data Science - In Action

---

Battles behind the scenes

# Stages of Data Science

- Purpose
- Relevant Data Collection
- Wrangling(cleansing)\*
- Data Analytics
- Feature Engg.\*
- Data Modelling\*
- Data Prediction\*
- Evaluation\*
- Reportings
- Finalising Report
- Data Product Building (software development)
  - Architecture
  - Development
  - Testing
  - Deployment

(\*) ⇒ Repetitive stages



# Data Model

- Random Forest Model
  - Bagging
- SVM
  - Linear Equation

# Iris Dataset - Goal

<< Ipython Notebook >>

# Plan

## Introduction

- **Definitions [ Data Science ]**
- **What, Why and How**
- **Examples**

## Data Science - In Action

- **Stages [DG, DC, DM, ME]**
- **Regression & Clustering Models**
- **Basics [ LR, GD ]**

## Real Life Application

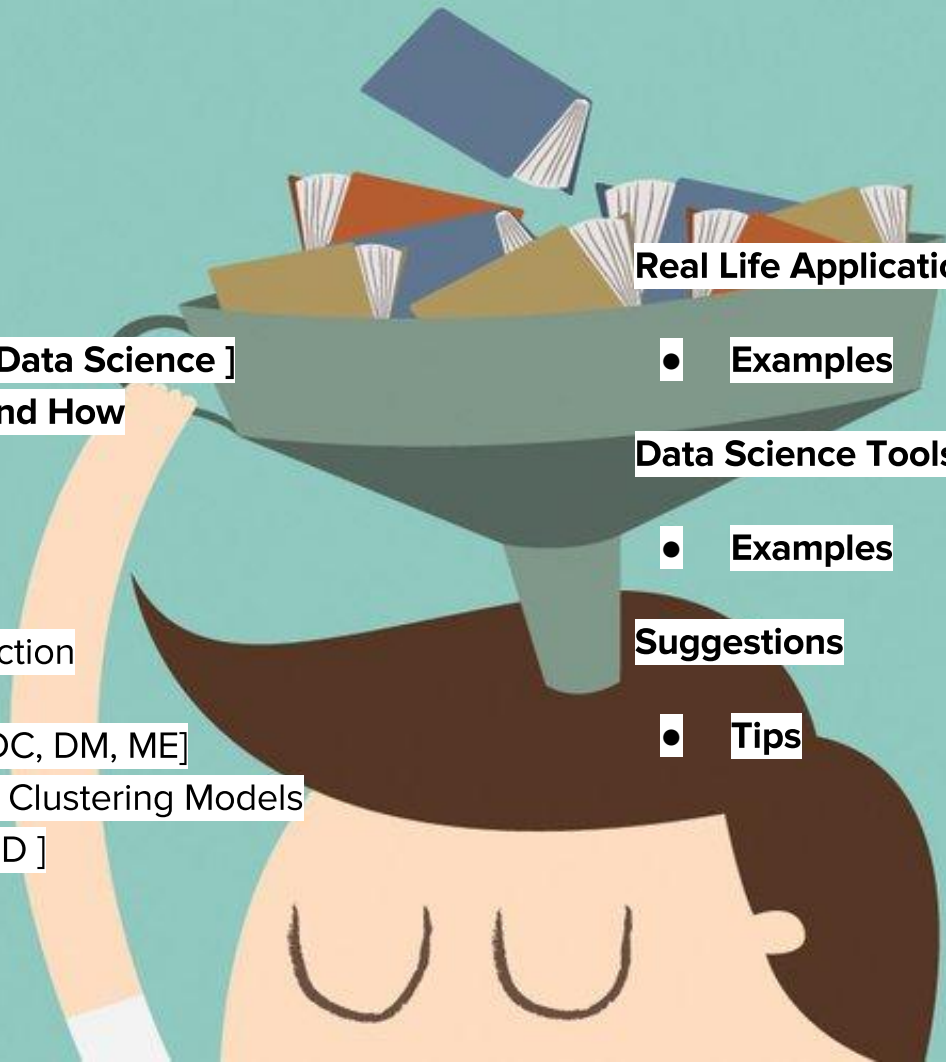
- **Examples**

## Data Science Tools

- **Examples**

## Suggestions

- **Tips**



# Data Science - Real Life App

---

Few applications that inspired me

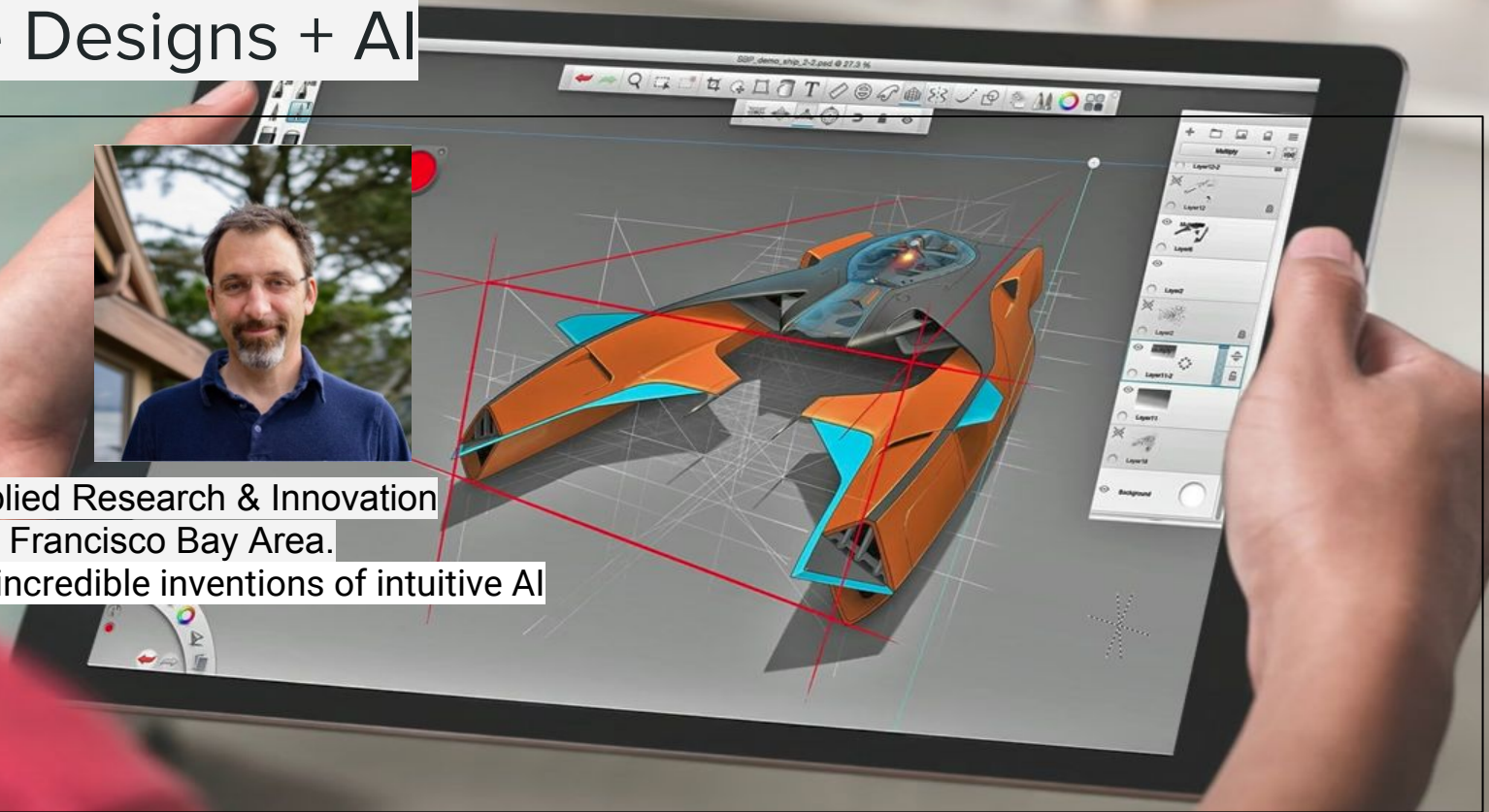
# Passive Designs + AI



**Maurice Cont**

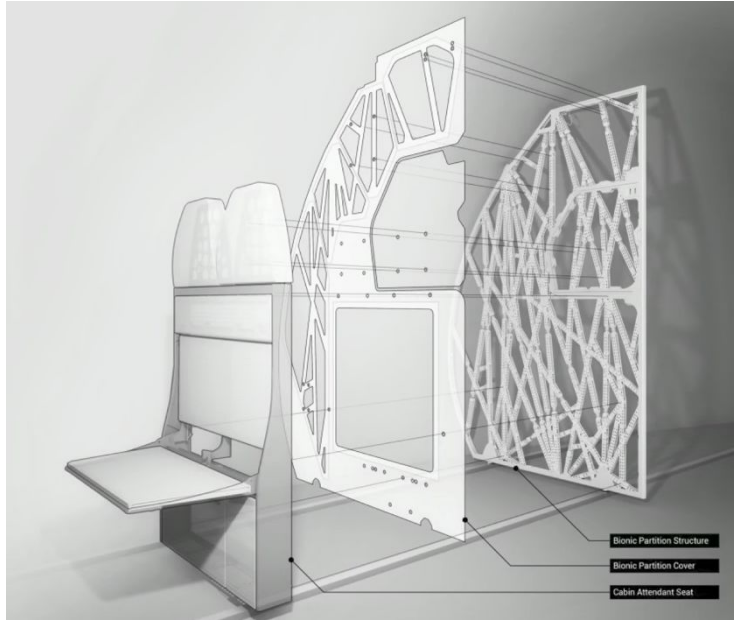
Director of Applied Research & Innovation  
Autodesk, San Francisco Bay Area.

TED Talk: The incredible inventions of intuitive AI





# Generative Designs > Passive Designs



AI Designed Lightweight Cabin Partition  
Airbus - A320



AI Designed Lightweight Drone Chassis

# Generative Designs



# Generative Designs

## AI Designed Car Chassis



# Music XRay

- Jimmy Lloyd Songwriter Showcase
- Popular songs share Melody & Rhythm
- Genere - 70
- Cluster 60
- Singer & Song Writer NY
- <http://www.heidimerrill.com/epk/index.html>



# Pred Pole

- 2011 Santa Cruz Pred Pole
- Crime, Location & Date-Time
- <https://www.predpol.com/>

## Results:

- 50% Crime Rate control
- 20% reduction in Crime Rate





# Generative Designs

## Project - Interlace



# Plan

## Introduction

- Definitions [ Data Science ]
- What, Why and How
- Examples

## Data Science - In Action

- Stages [DG, DC, DM, ME]
- Regression & Clustering Models
- Basics [ LR, GD ]

## Real Life Application

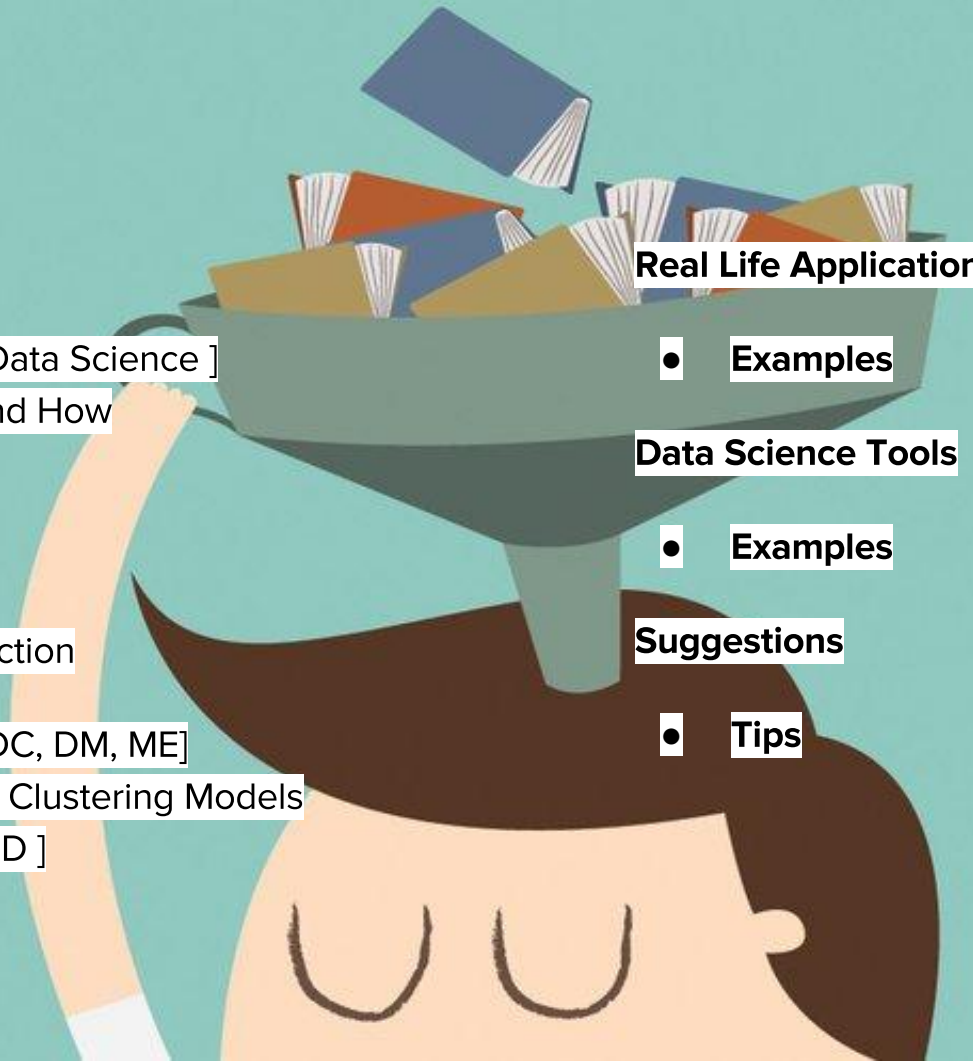
- Examples

## Data Science Tools

- Examples

## Suggestions

- Tips



# Data Science - Tools

---

Too many to name, but none of them are close perfection.

# Data Science Tools

- Languages: Scala, R, **Python**, Java, C#
- Lib: **Scikit**, DeepNet, **Tensor flow**, Theano, H2O
- Frameworks: Apache Spark

These are some used by used us (Imaginea Labs - Data Sciences - 4th Floor, Hyd).

# Suggestions

---

Challenges in DS & Tips to who want to start.

# Suggestions?

- Data Preparation
  - **“Give me six hours to chop down a tree and I will spend the first four sharpening the axe”.**  
Abraham Lincoln
  - Python, Scala, Excel, Databases(regex).
- Data Analytics
  - **“Seeing is believing”**
  - Python(Matplotlib, Seaborn), D3.Js, Excel.
- Data Models
  - **“There are no perfect solutions, but some work better”**
  - Learn 2-3 types of Clustering, Regression Models(LR,RF,SVM,KNN,XGB)
- Evaluation
  - **“A product not tested is broken by default”**
  - Accuracy, RMSE, Precision-Recall, F1 Score

# Questions?

---

Sampath - Desk 4F 072. Imaginea Labs - Data Sciences.

Sachin, Keerat, Bipul, Kavi, Mageshwaran.



Thank you

