Data Science Part-Time Septiembre 23'

Introducción al Data Science

Instructors

Carlos Sevilla Barceló

Fernando Carrasco Díaz

Teacher Assistant

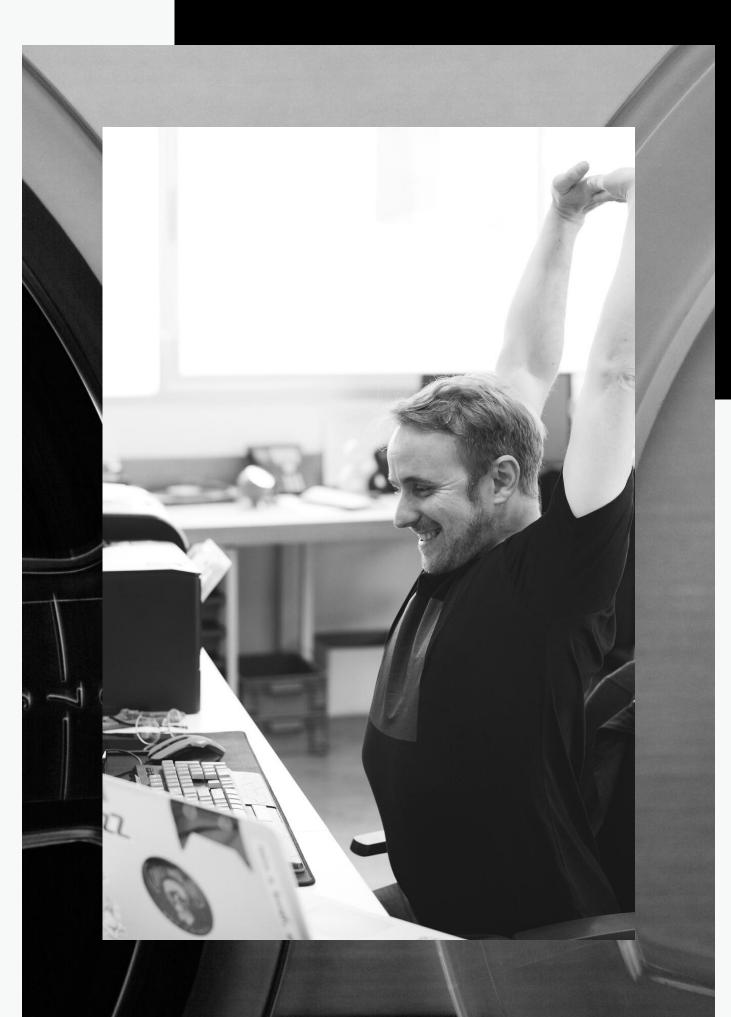
Alba Cruz Gálvez



Índice

- ☐ ¿Qué significa ser DS?
- ☐ Hablemos de aplicaciones
- ☐ Auge del DS
- Nuestro camino hacia ser DS
- ☐ ¿Why Python?



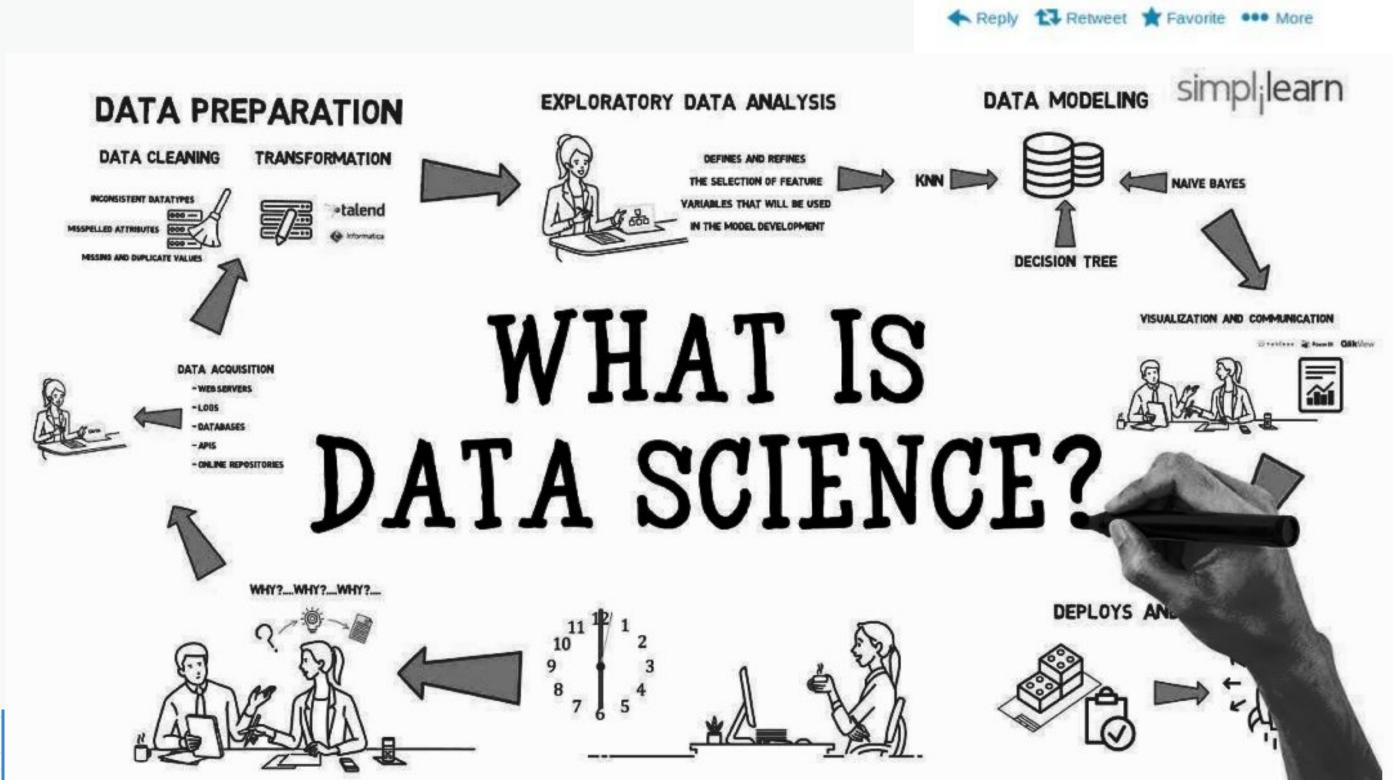


¿Qué significa serlo? I



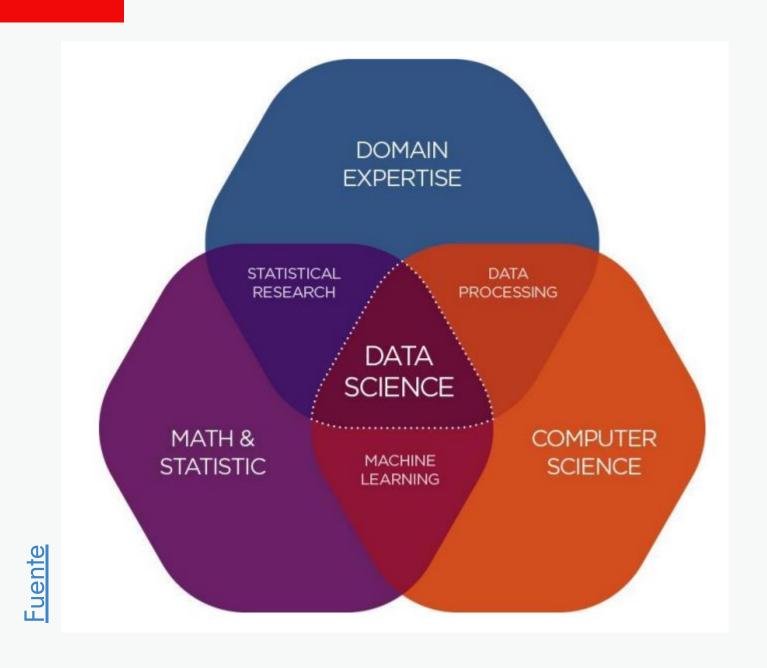


Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

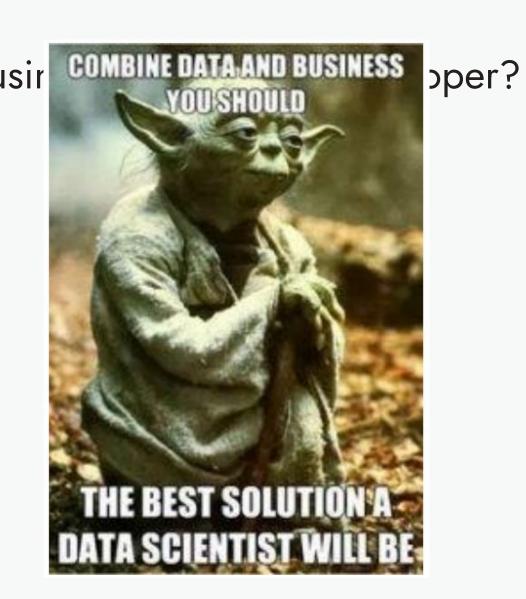


THE BRIDGE

¿Qué significa serlo? Il



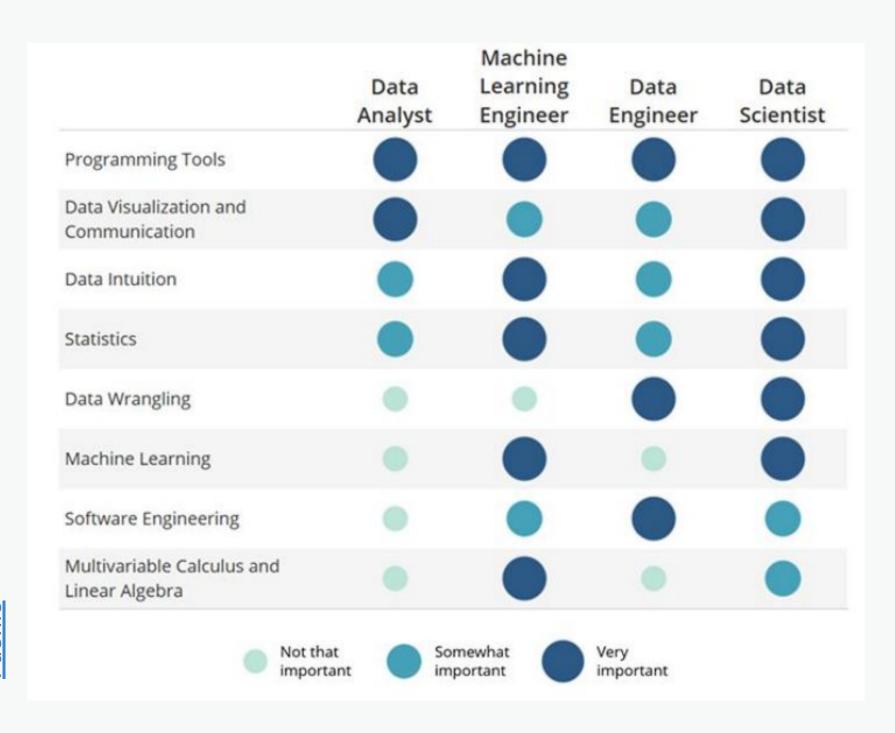
La trinidad del DS



Nuestro sitio

THE BRIDGE

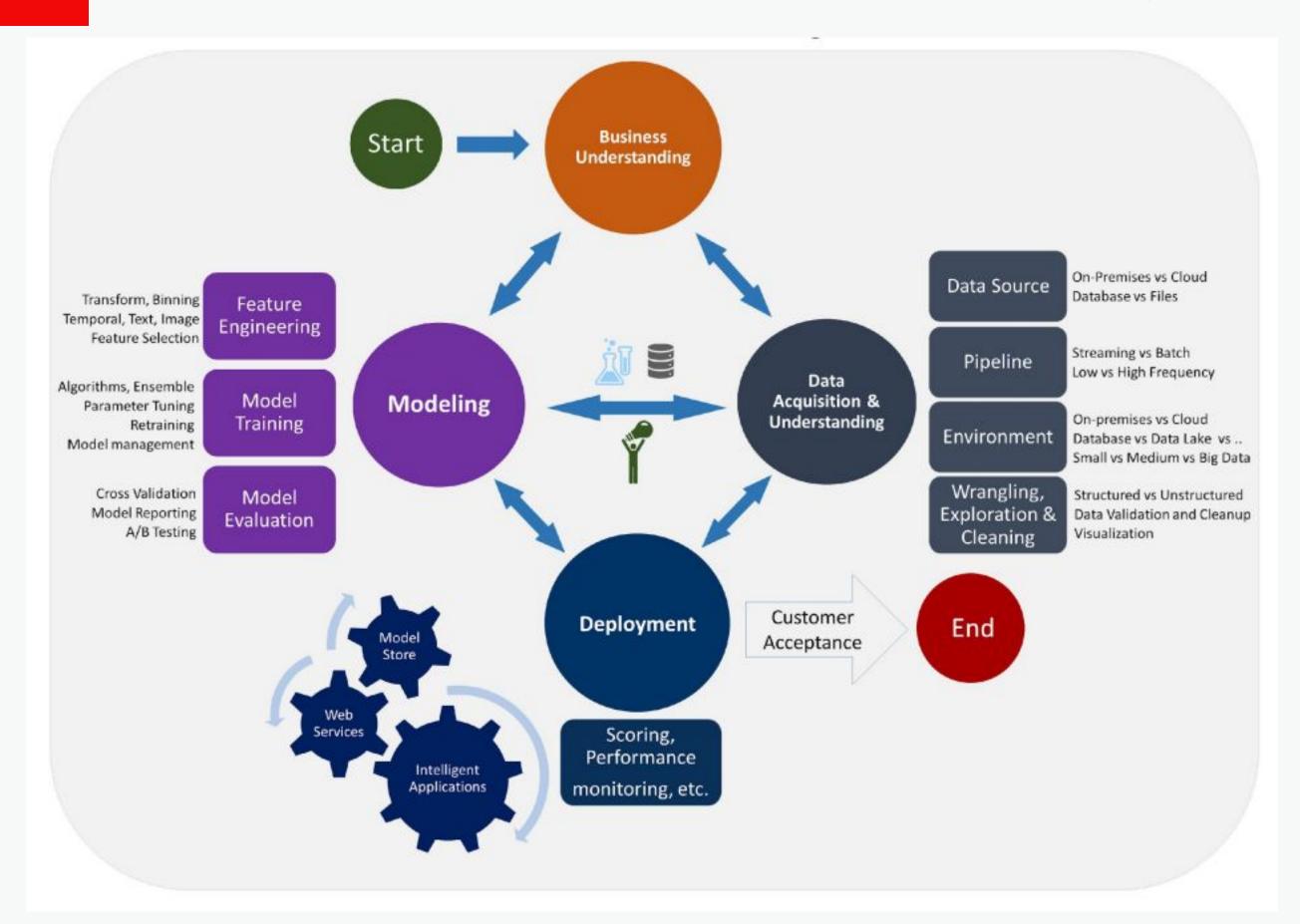
¿Qué significa serlo? II



Fuente

En qué dedicamos nuestro esfuerzo

¿Cómo abordamos nuestro trabajo?



THE BRIDGE DIGITAL TALENT ACCELERATOR

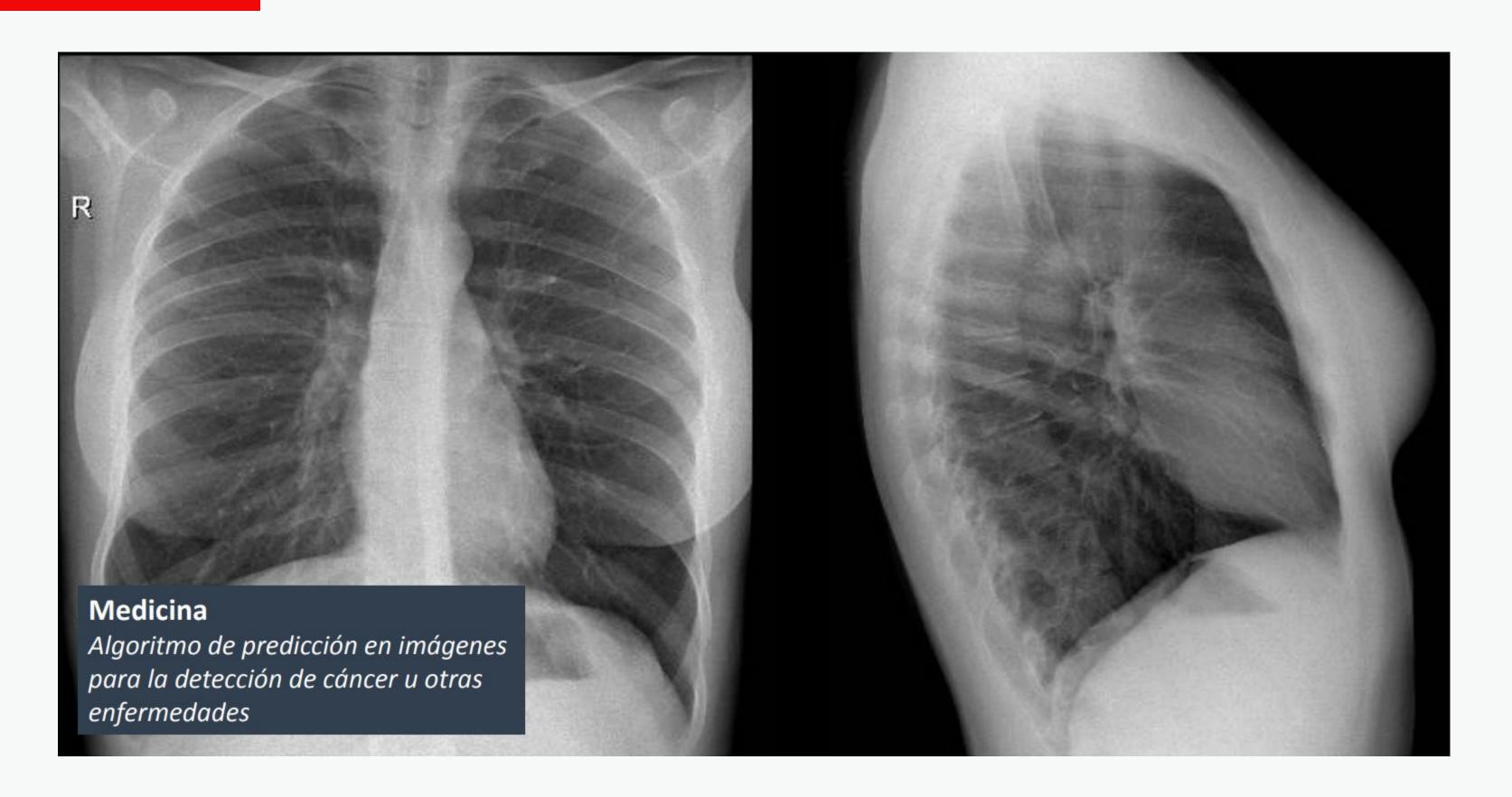
Hablemos de aplicaciones l



Hablemos de aplicaciones II



Hablemos de aplicaciones III



Hablemos de aplicaciones IV



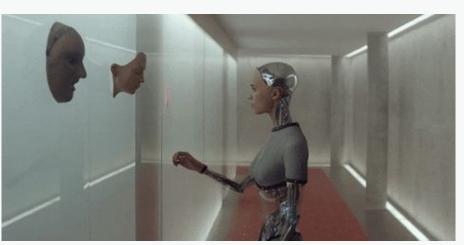
THE BRIDGE

Hablemos de aplicaciones V





Her (2013)



Ex-Machina (2014)

Alerta Spoiler: Esto no es nuevo, pero... I

Modelos Predictivos

History [edit]

Least squares linear regression, as a means of finding a good rough linear fit to a set of points was performed by Legendre (1805) and Gauss (1809) for the prediction of planetary movement. Quetelet was responsible for making the procedure well-known and for using it extensively in the social sciences. [26]

Distribuciones de población

History [edit]

Bayes' theorem was named after Thomas Bayes (1701-1761), who studied how to compute a distribution for the probability parameter of a binomial distribution (in modern terminology). Bayes's unpublished manuscript was significantly edited by Richard Price before it was posthumously read at the Royal Society. Price edited[11] Bayes's major work "An Essay towards solving a Problem in the Doctrine of Chances" (1763), which appeared in Philosophical Transactions, [12] and

Computación

History [edit]

See also: Algorithm and Church-Turing thesis

They were described in 1936 by Alan Turing.

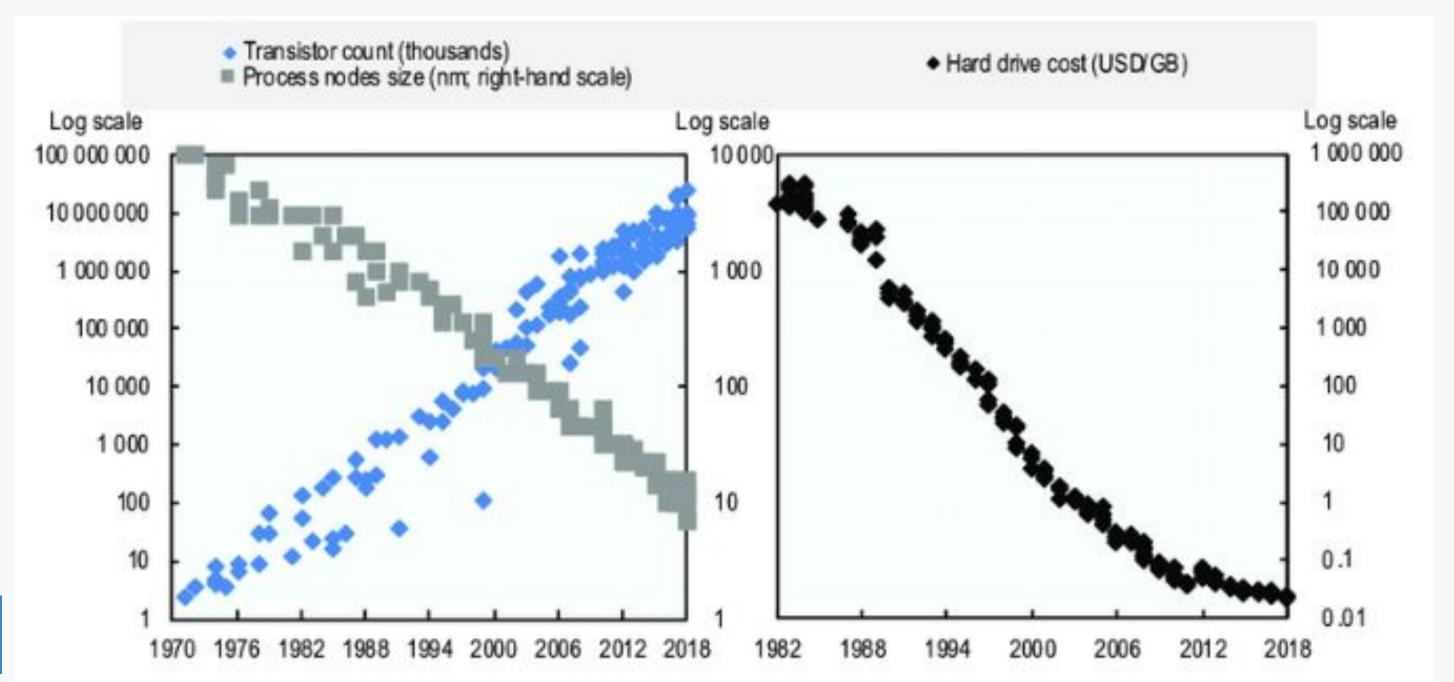
Historical background: computational machinery [edit]

Robin Gandy (1919–1995)—a student of Alan Turing (1912–1954), and his lifelong friend—traces the lineage of the notion of "calculating machine" back to Charles Babbage (circa 1834) and actually proposes "Babbage's Thesis":



Alerta Spoiler: Esto no es nuevo, pero... Il

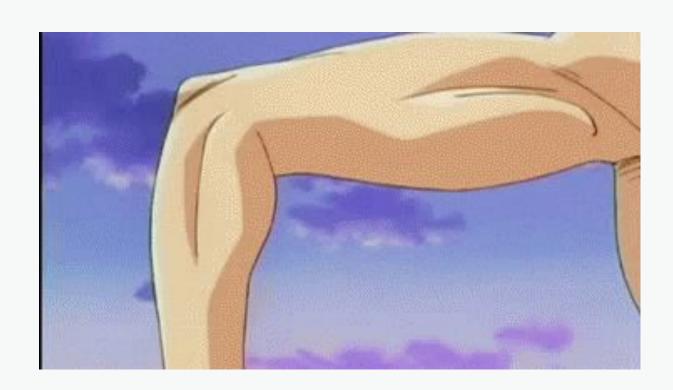
Procesado + Almacenamiento + Coste = Accesibilidad



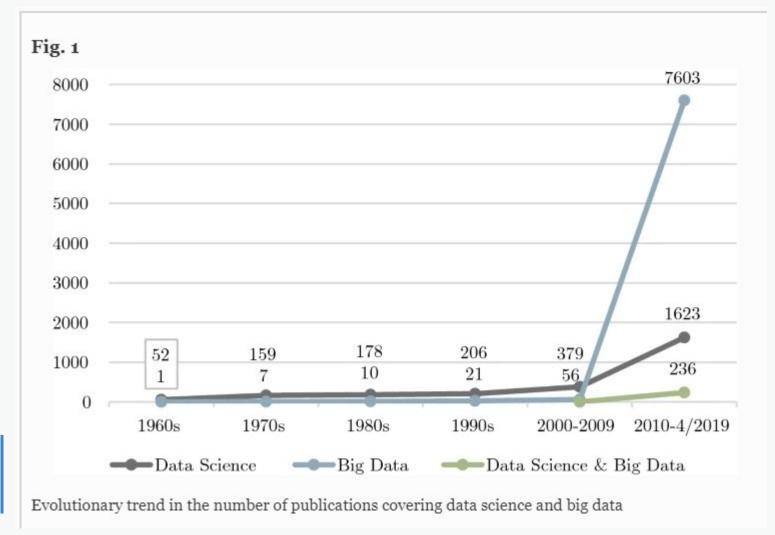
Alerta Spoiler: Esto no es nuevo, pero... III

Generación de datos

- 1.7MB of data is created every second by every person during 2020.
- In the last two years alone, the astonishing 90% of the world's data has been created.
- 2.5 quintillion bytes of data are produced by humans every day.
- 463 exabytes of data will be generated each day by humans as of 2025.
- 95 million photos and videos are shared every day on Instagram.
- By the end of 2020, 44 zettabytes will make up the entire digital universe.
- Every day, 306.4 billion emails are sent, and 5 million Tweets are made.



Conocimiento e investigación



Fuente

9. Data Munging Sampling (Using ETL How much Data? Principal Google OpenRefine 5. Text Mining / NLP 8. Data Ingestion xtraction 6. Visualization Data Exploration in R (Hist, Boxplot etc) Linear Regression Uni, Bi & Multivariate Viz Regression Ranking ggplot2 Logistic Regression Histogram & Pie (Uni) Tree & Tree Map 10. Toolbox Author: Swam classification drasekaran 4. Machine Learning Scatter Plot (Bi) MS Excel w/ Analysis ToolPak Line Charts (Bi) Prob Den Fn (PDF) 1. Fundamentals Java, Python Spatial Charts ANOVA (R, R-Studio, Rattle Survey Plot Skewness Matrices & Linear Algebra Fundamentals Weka, Knime, RapidMiner Continuos Distributions Hash Functions, Binary Tree, O(n) Normal, Poisson, Gaussian) Hadoop Dist of Choice Data Frames Lists Cumul Dist Fn (CDF) Spark, Storm Relational Algebra, DB Basics Factors Reading CSV Data Random Variables Flume, Scibe, Chukwa Inner, Outer, Cross, Theta Join Arrays Reading Raw Data Bayes Theorem CAP Theorem Nutch, Talend, Scraperwiki Matrices Subsetting Data Probability Theory Tabular Data Zookeeper Webscraper, Flume, Sqoop Data Frames & Series Vectors Manipulate Data Frames Avro Percentiles & Outliers Name & Data Nodes Storm: Hadoop () tm, RWeka, NLTK Sharding (Histograms Functions Variables Setup Hadoop (IBM / Cloudera / HortonWorks) RHIPE Rhadoop, OLAP (Exploratory Data Analysis Data Replication Principles D3.js, ggplot2, Shiny Expressions Multidimensional Data Model Descriptive Statistics HDFS nean, median, range, SD, Var) R Basics Cassandra ETL (Hadoop Components ick a Dataset R Setup R Studio MongoDB, Neo4 Map Reduce Fundamentals Reporting Vs BI Vs Analytics 7. Big Data 2. Statistics 3. Programming

Fuente

¿Qué vamos a aprender? En titulares l

- ☐ Aprender a programar: *Python B2*
- EDA (Exploratory Data Analysis): El Cacharreo
- Acceso a Bases de Datos: A ver qué lío tienen aquí montado
- Dashboards con BI Tools: Esto con colores y barras va a quedar genial
- Resolución de problemas: Que apañao/apañá eres hijo/hija
- ☐ Machine Learning/IA: ¿Esto no me va a quitar el trabajo no?
- Desafío tripulaciones: A mi se me da bien hacer powerpoints

¿Qué vamos a aprender? En titulares II

- ☐ Productivización: ¿Ah pero que esto lo ejecutas en tu ordenador?
- ☐ Big Data: Mi proceso es genial pero tarda 4 horas
- □ Data Storytelling: Lo que me estás contando me suena a chino
- ☐ Impacto en negocio: Que sí que muy bien, ¿pero a mi esto en qué me ayuda?



THE BRIDGE DIGITAL TALENT ACCELERATOR

¿Why Python?



Videojuegos

Battlefield, Civilization



Web

Django



Productivización

Google Cloud



Data Science

Mayor comunidad



Apps

Capacidad Front y Back

Orientado a objetos

Comunidad viva

Fácil de aprender

Open Source

Propósito general

Fuertemente tipado

Módulos/ Frameworks Pensado en producción

Multiplataforma

Mercado laboral



¡BIENVENIDOS DE NUEVO!

LinkedIn:

Alba Cruz

Carlos Sevilla

Fernando Carrasco

Emails:

alba.cruz@thebridgeschool.es

carlos.sevilla@thebridgeschool.es

fernando.carrasco@thebridgeschool.es



¿Preguntas?