

Data Science Intro





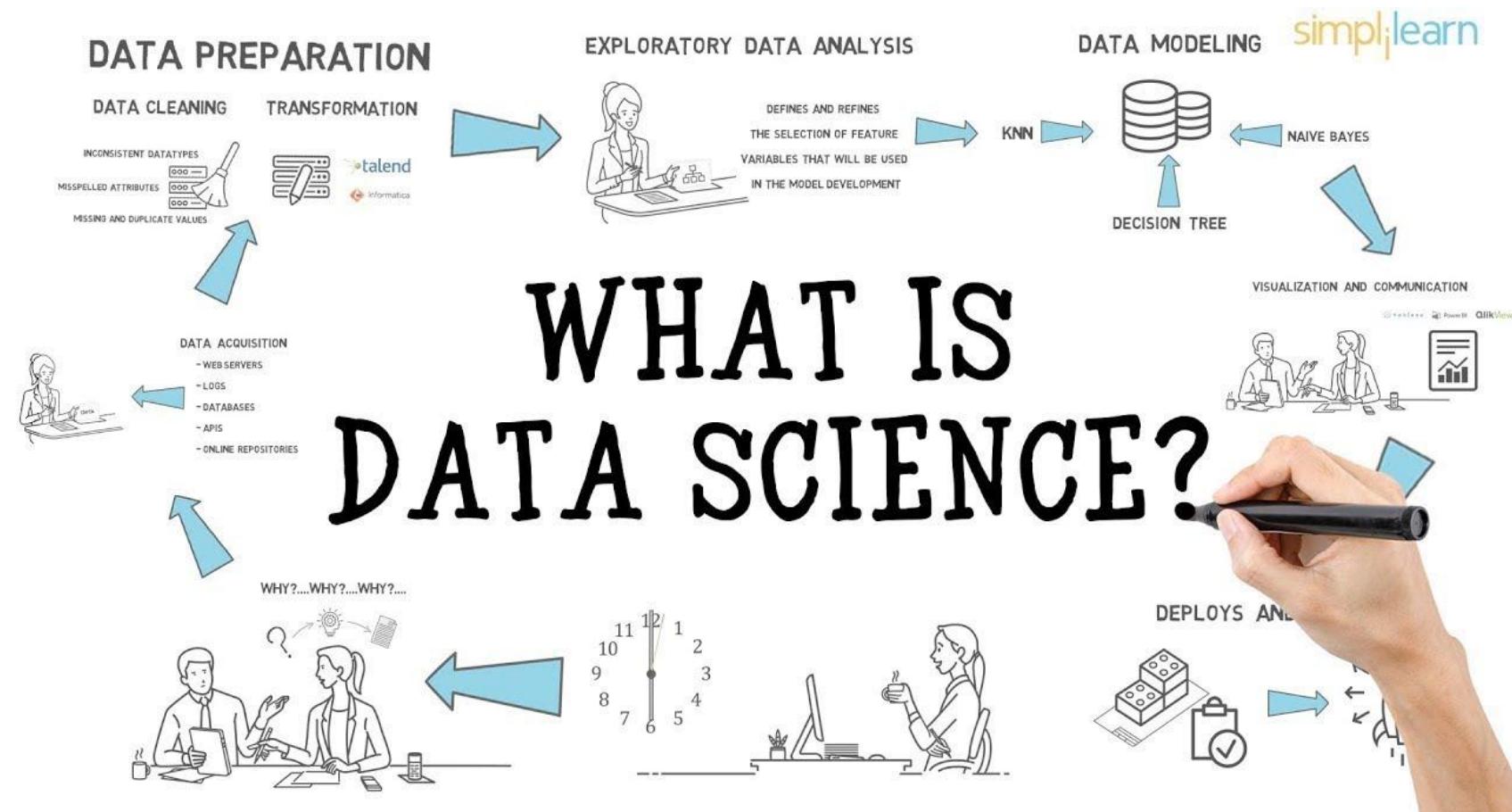
Josh Wills
@josh_wills

Follow

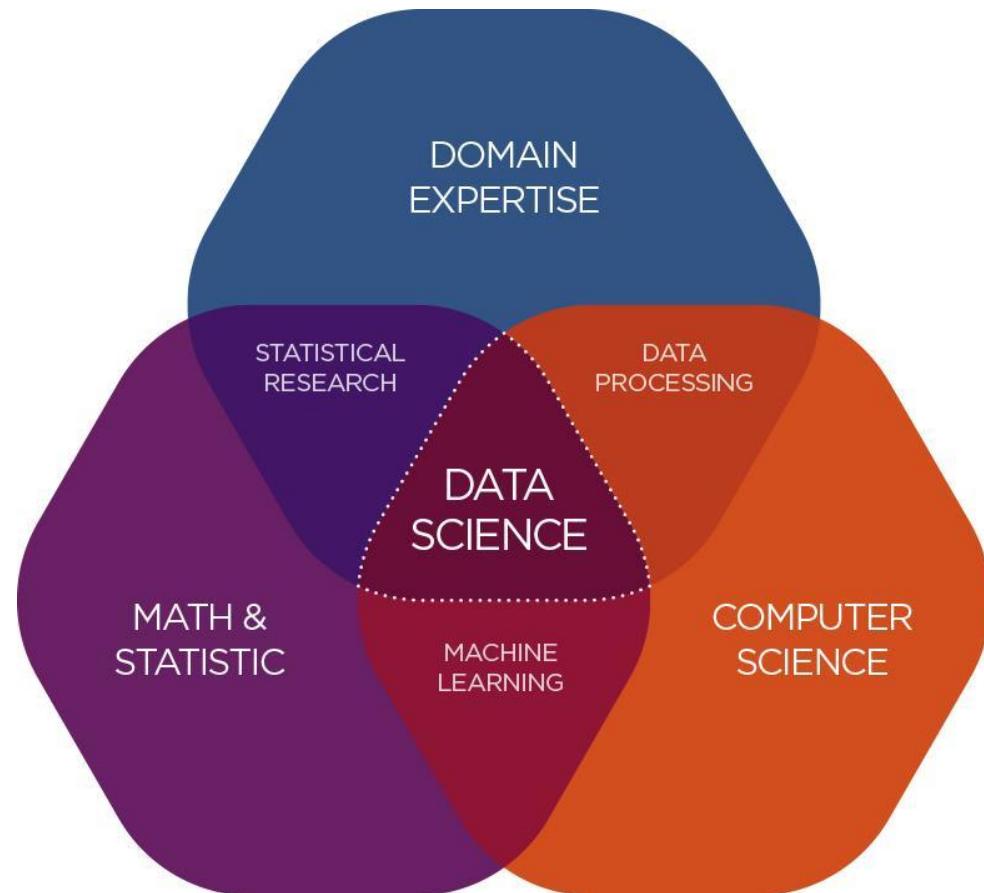
¿Qué es un científico de datos?

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

Reply Retweet Favorite More



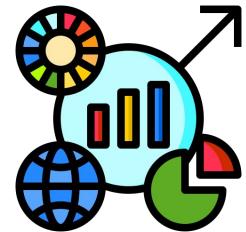
Skills de un data scientist



Sectores donde trabaja un Data Scientist

PRÁCTICAMENTE
TODOS

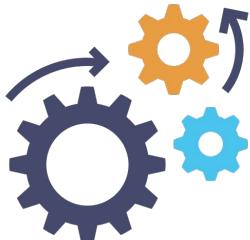
Análisis de datos para la toma de decisiones



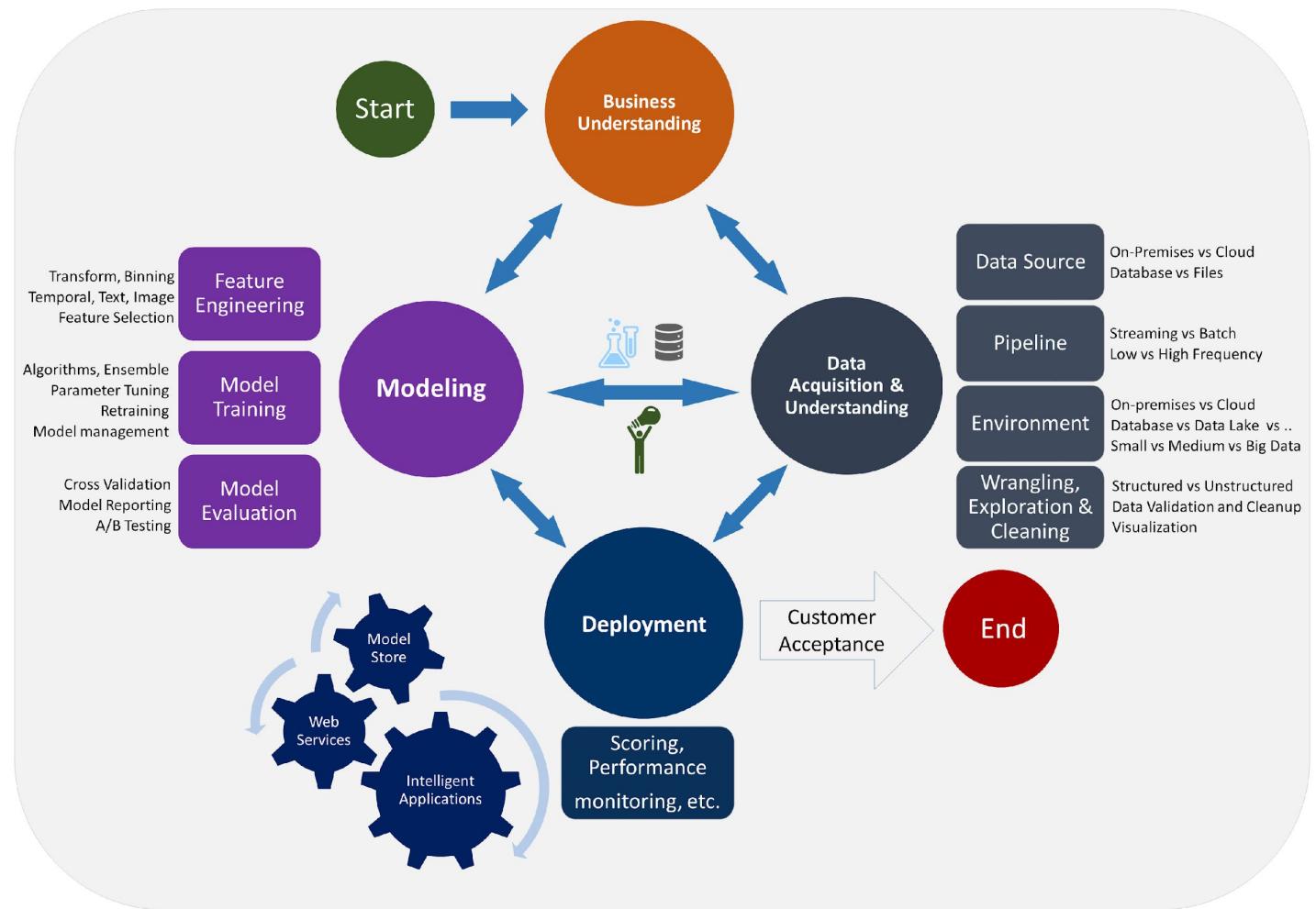
Algoritmos predictivos



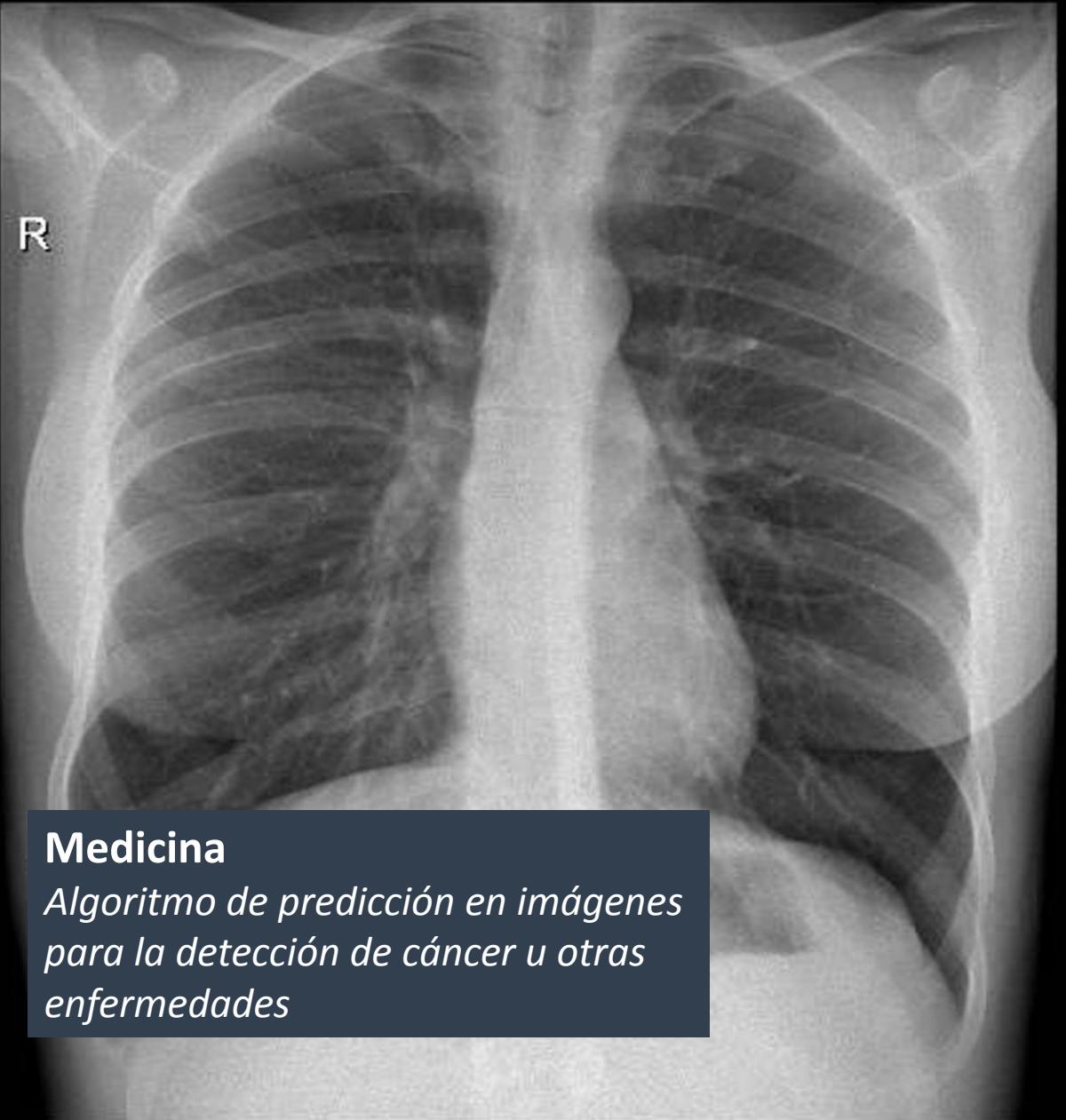
Automatizar procesos



Data Science Lifecycle



Ciclo de vida de
un proyecto de
Data Science



Medicina

*Algoritmo de predicción en imágenes
para la detección de cáncer u otras
enfermedades*

A close-up photograph of an aircraft's engine and fuselage. The engine is on the right, showing its metallic cowling and the complex fan and compressor blades. The fuselage of another plane is visible on the left, featuring blue and white stripes and several small circular windows.

Transportes y logística

*Algoritmos de predicción de averías en función
de los sensores del vehículo*



Traders y banca

Predicción de subidas y caídas del mercado, así como algoritmos de default

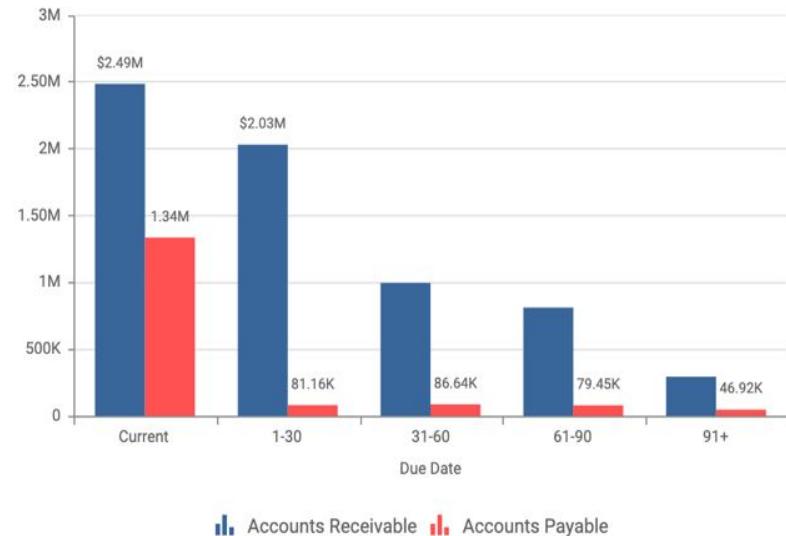
Total Accounts Receivable
\$6,621,280

Total Accounts Payable
\$1,630,270

Equity Ratio
75.38 %

Debt Equity
1.10 %

Total Accounts Receivable and Payable Aging



Current Ratio ⓘ



DSI
[Days Sales Inventory]



DSO
[Days Sales Outstanding]



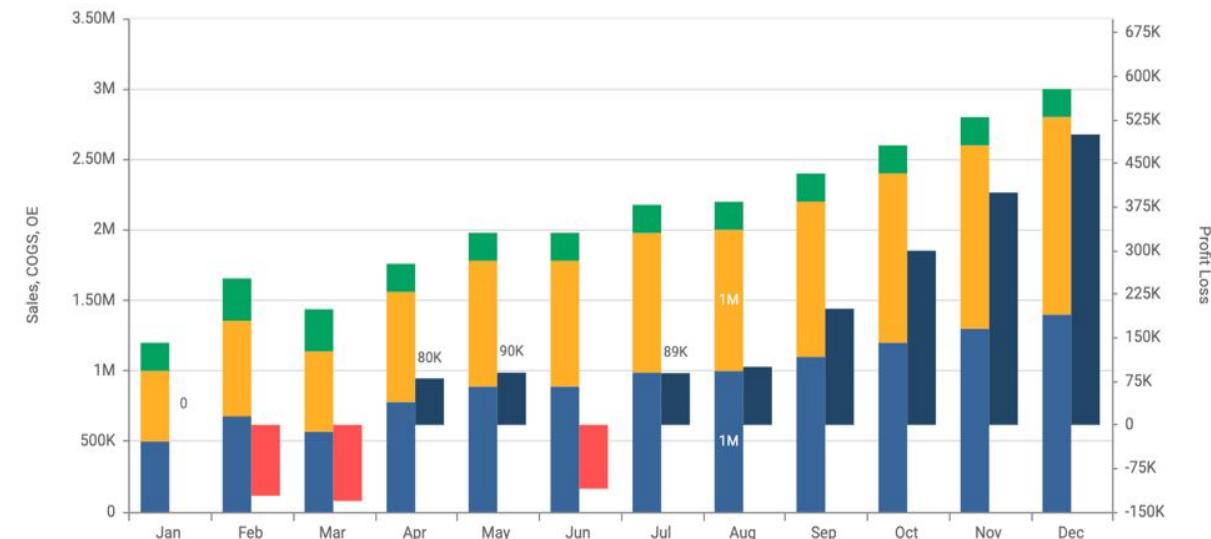
DPO
[Days Payable Outstanding]



Net Working Capital vs Gross Working Capital



Profit and Loss summary



Asistentes virtuales



 amazon alexa

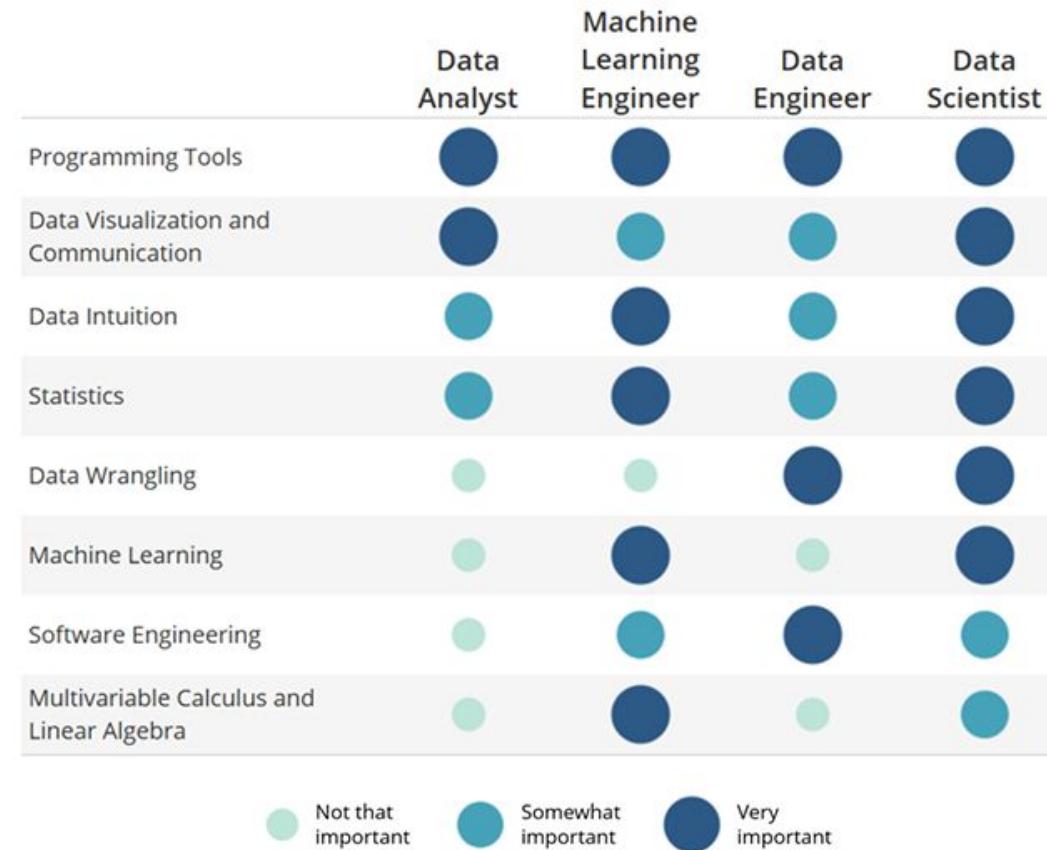
 Google
ASSISTANT

 Cortana

 Siri

 Bixby

Roles de Data Scientist



¿Por qué este boom ahora?

Linear Regression - 1805

History [edit]

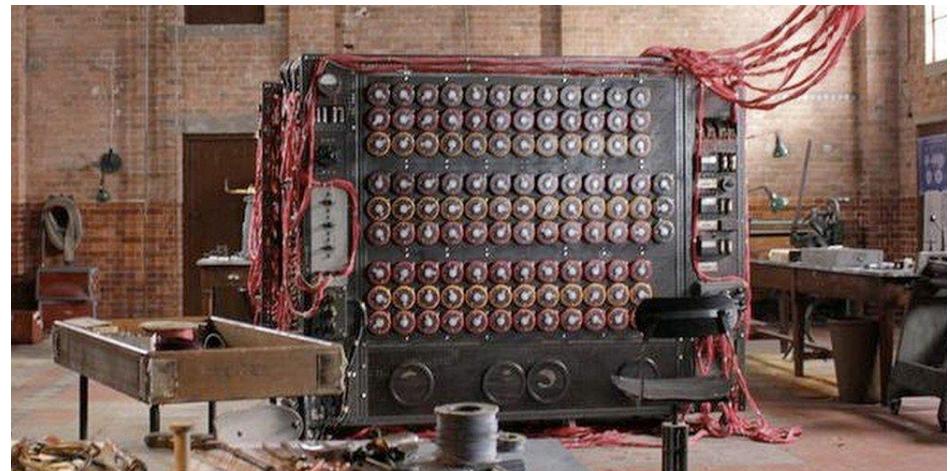
Least squares linear regression, as a means of finding a good rough linear fit to a set of points was performed by Legendre (1805) and Gauss (1809) for the prediction of planetary movement. Quetelet was responsible for making the procedure well-known and for using it extensively in the social sciences.^[26]

Teorema de Bayes - 1763

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

Máquina de Turing - 1936



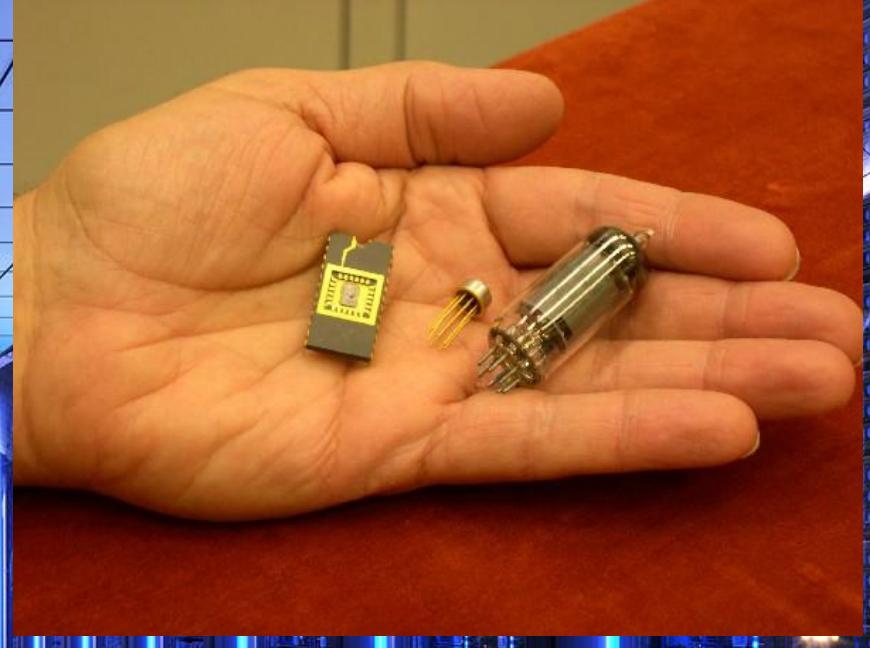


Capacidad de cómputo

Gran velocidad en la generación de los datos

Internet: red interconectada en todo el mundo

Diversidad de los datos: texto, imágenes, audio...



- **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.

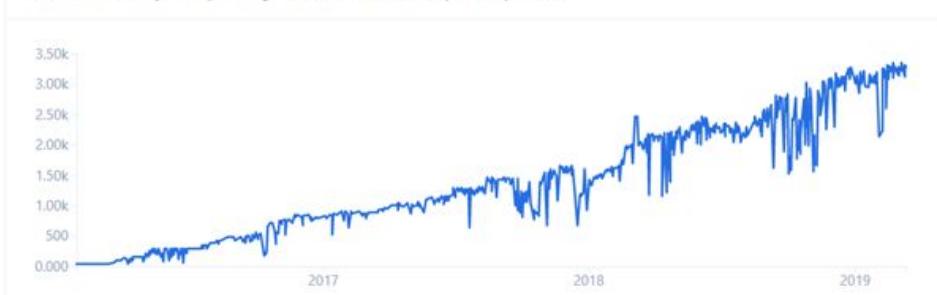
Hype del Data Scientist

Data scientists are in high demand

Data scientist job postings, per 1 million postings on Indeed

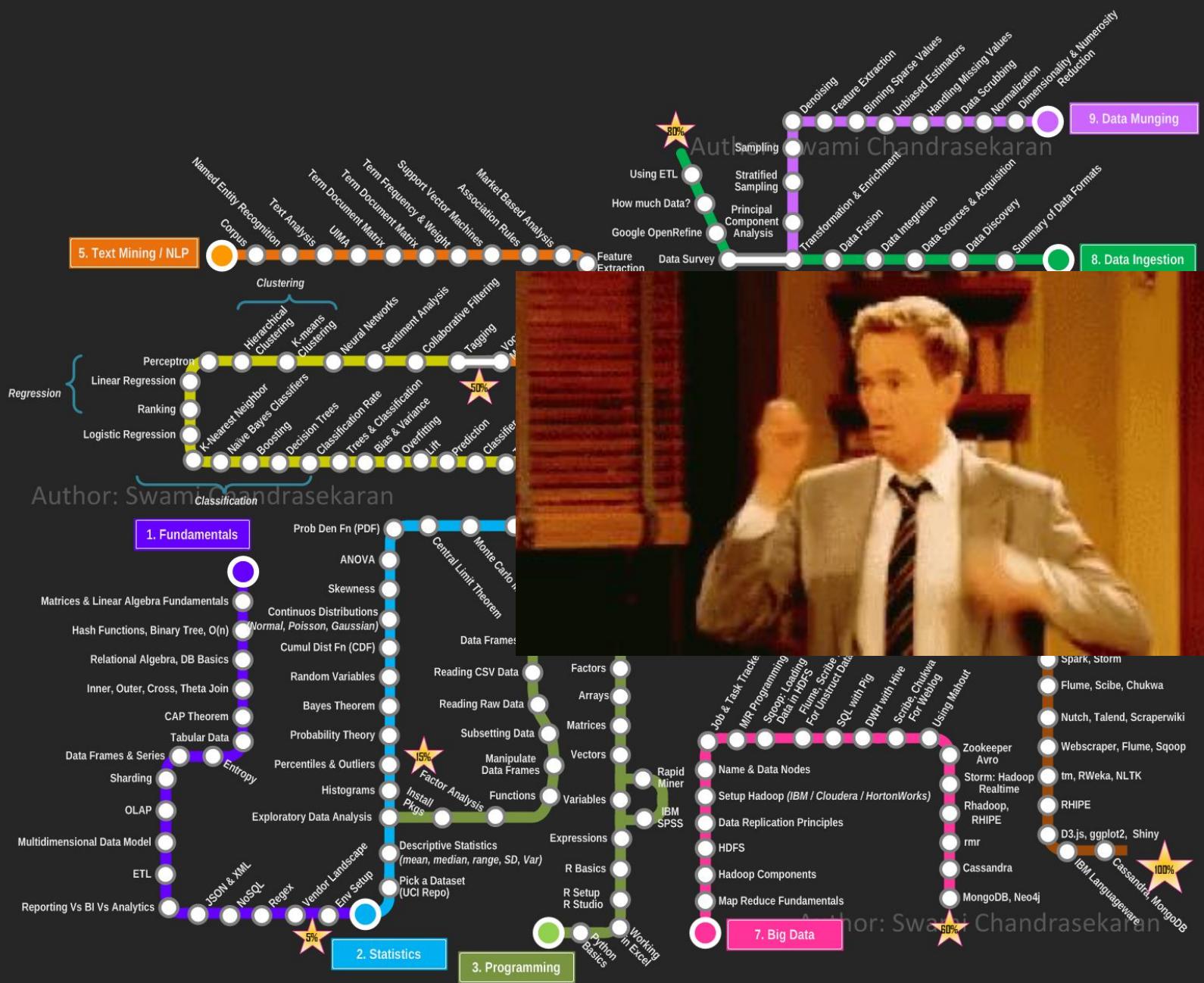


Data Scientist job openings at the world's top companies



Data from Thinknum - [Open dataset](#)

• Title (Count)



Ahora bien, ¿qué piden las empresas?

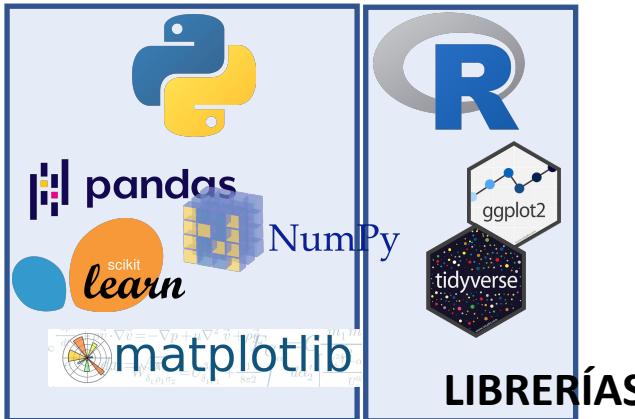
Fuente:
<http://nirvacana.com/thoughts/2013/07/08/becoming-a-data-scientist/>

¿Qué herramientas vamos a necesitar?



Lenguajes de programación

Con los lenguajes de programación haremos todo. Leer datos, limpiarlos, machine learning, visualizaciones



Visualización

Librerías de visualización de Python y herramientas de Business Intelligence



IDE (Entorno de Desarrollo Integrado)

Software donde programaremos con nuestros lenguajes



Otros frameworks

Software libre con APIs abiertas para Python



Control de versiones

Compartir archivos y llevar una traza de nuestros programas



¿Qué vas a aprender en el bootcamp?

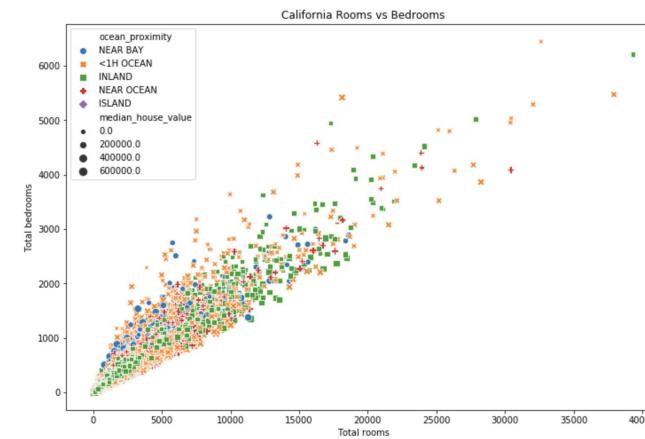
Aprender a programar en Python



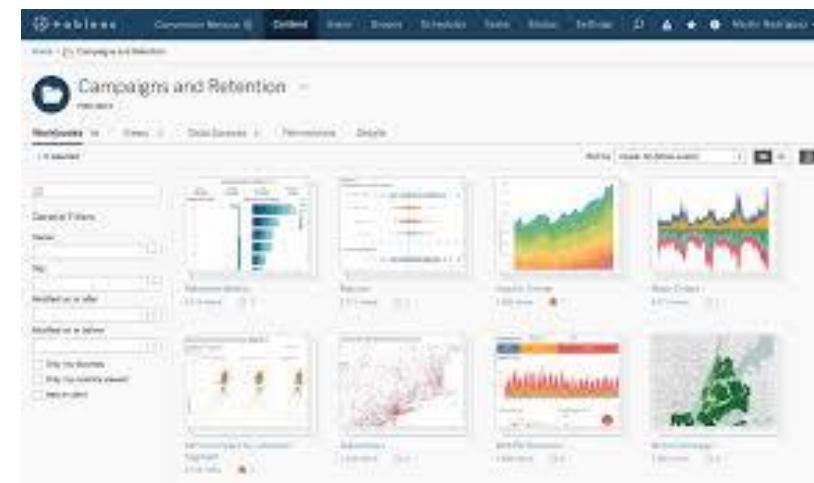
Acceso a Bases de Datos/Web



EDA (Exploratory Data Analysis)



Dashboards con herramienta de Business Intelligence



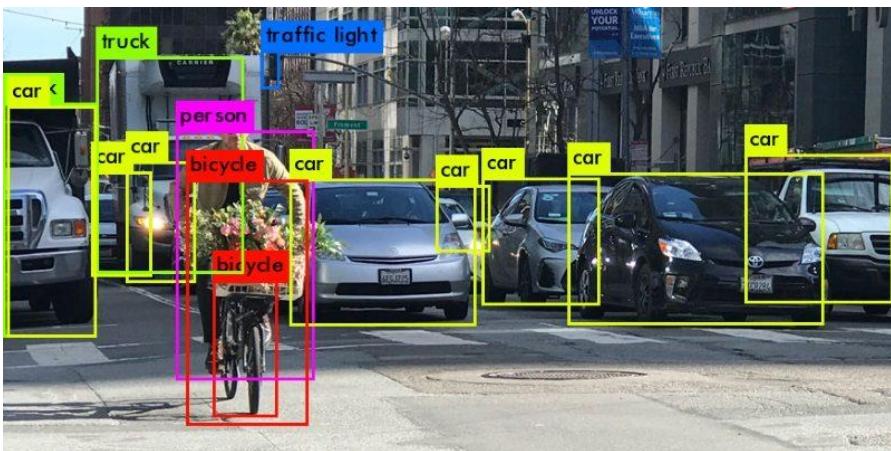
Resolución de problemas

Problem

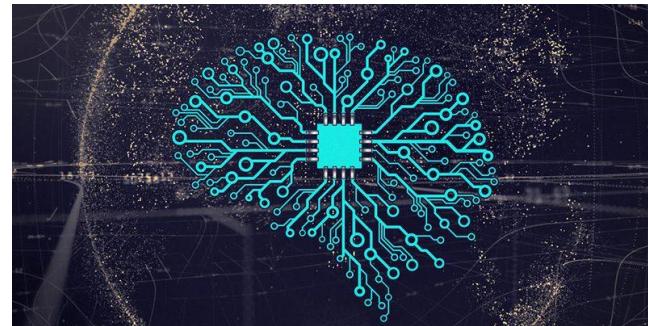


Solving

Inteligencia Artificial



Machine Learning



Trabajo en equipo Desafío de tripulaciones



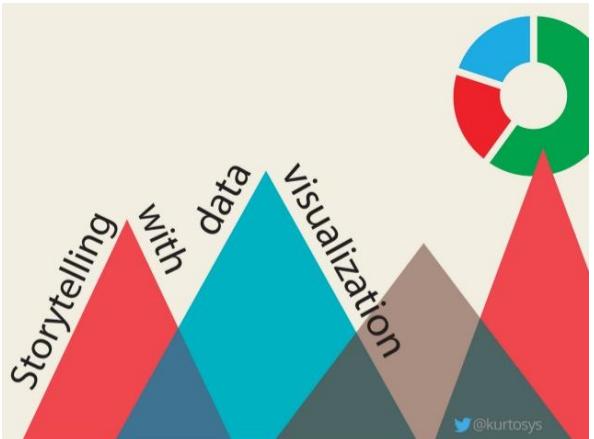
Productivización



Big Data



Data Storytelling



Impacto en negocio



¿Por qué Python?

¿Quién usa Python?



¿Para qué se usa Python?



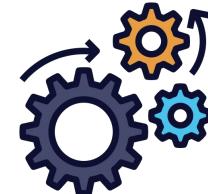
Videojuegos

Videojuegos como Battlefield, Civilization IV o Sims



Web

Webs como la de Instagram están desarrolladas en el framework de Python Django



Automatización

Simples procesos batch de datos, o despliegues de software en producción



Data Science

Análisis exploratorio de datos, visualización, modelos de Machine Learning.



APPs de escritorio

Aplicaciones sencillas como una calculadora, lista de to-do...

Python

Lenguaje de programación
orientado a objetos

Open source

Fácil de aprender

Comunidad

Fuertemente tipado

Módulos

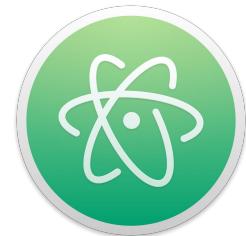
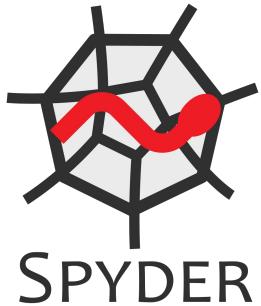
Propósito general

Pensado en producción

Multiplataforma

¿IDEs de Python?

IDE (Integrated Development Environment)



Atom



Visual Studio®

¡Bienvenidos!

