

Essential Machine Learning with Python

Luis Chacón

https://www.facebook.com/qslearningperu/?ref=page_internal



QS Learning



+51 937 012 707 / +51 915 111 457



quants.admission@gmail.com

Agenda

01 Data science

Ciencia de datos, crecimiento y Skills

02 Machine Learning

Introducción a la inteligencia artificial y Machine learning

03 Python introduction

Python, crecimiento y fundamentos

04 Anaconda & Colaboratory

Get a modern PowerPoint Presentation that is beautifully designed. I hope and I believe that this Template will your Time.



Ciencia de datos

**Ciencia emocionante que
convierte los datos en
información y
progresivamente en
conocimiento.**

La importancia, y quizá también lo
atractivo, de la CD radica en el
crecimiento exponencial del volumen de
datos.

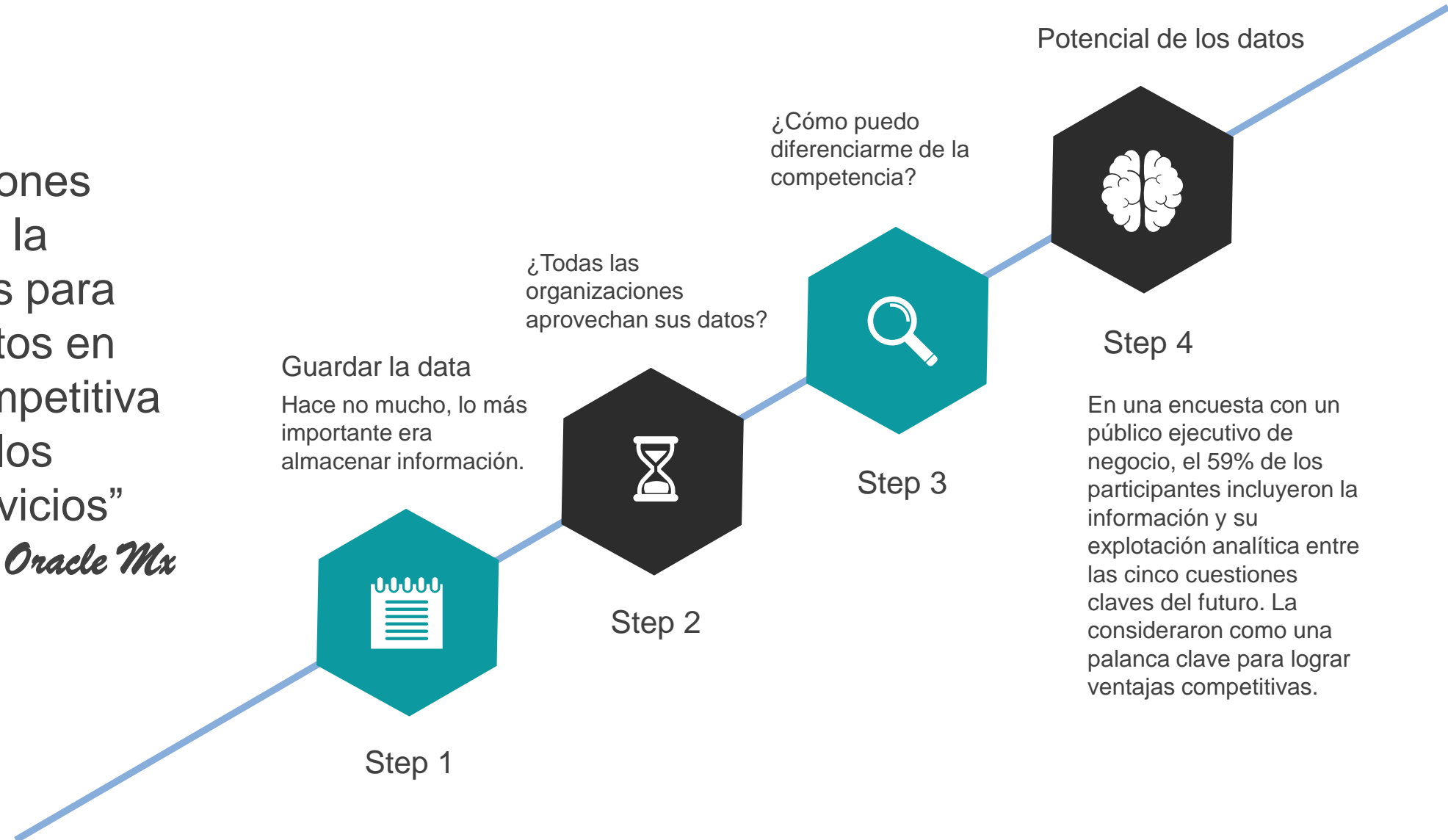
Este crecimiento “Desmesurado” y los
continuos avances de la computación han
sido el ecosistema propicio para una de
las disciplinas en las cuales se juntan
diferentes perfiles con “skills”
característicos.



Ciencia de Datos

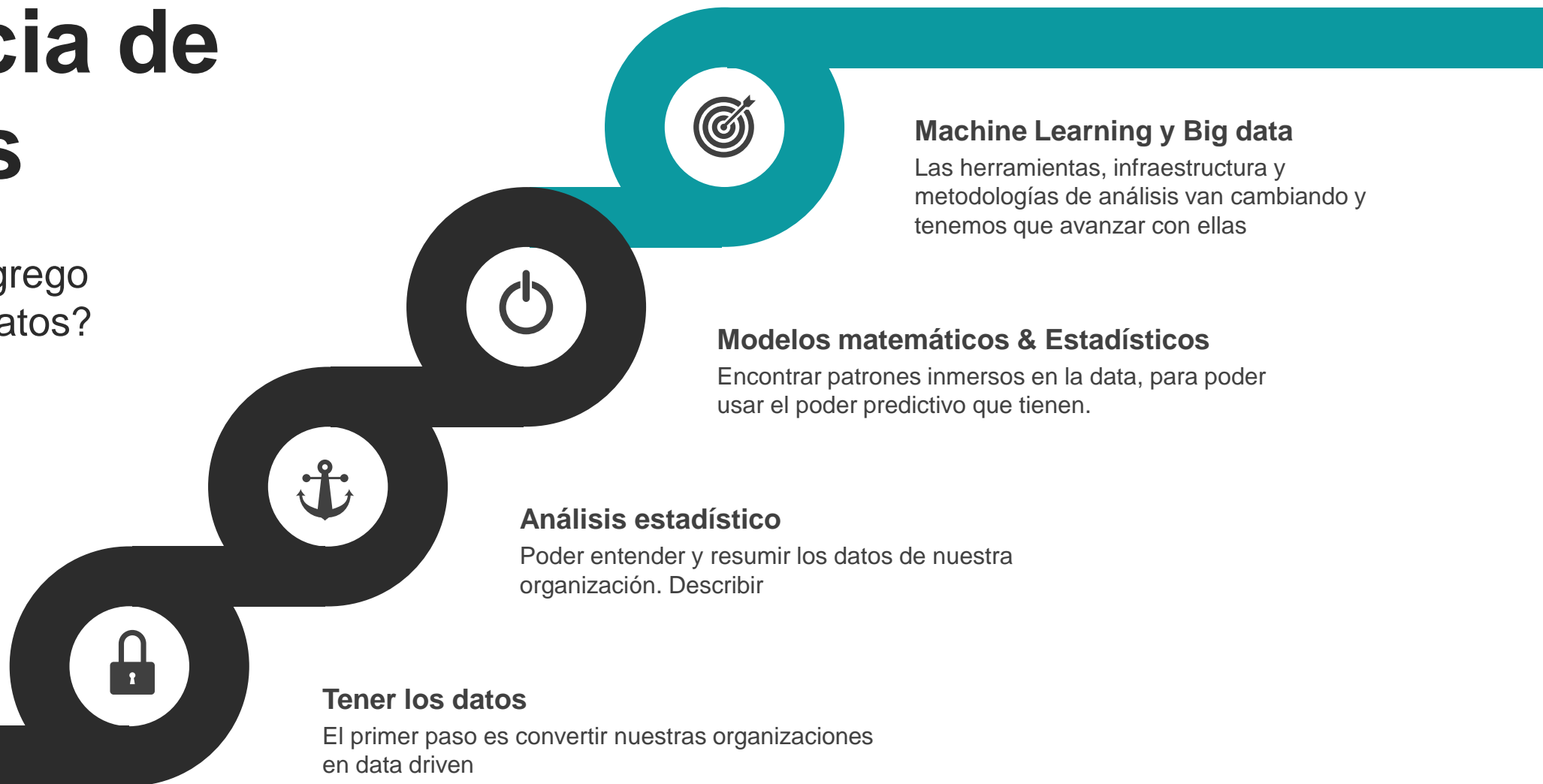
“Las organizaciones están utilizando la ciencia de datos para convertir los datos en una ventaja competitiva al perfeccionar los productos y servicios”

Oracle Mx



Ciencia de Datos

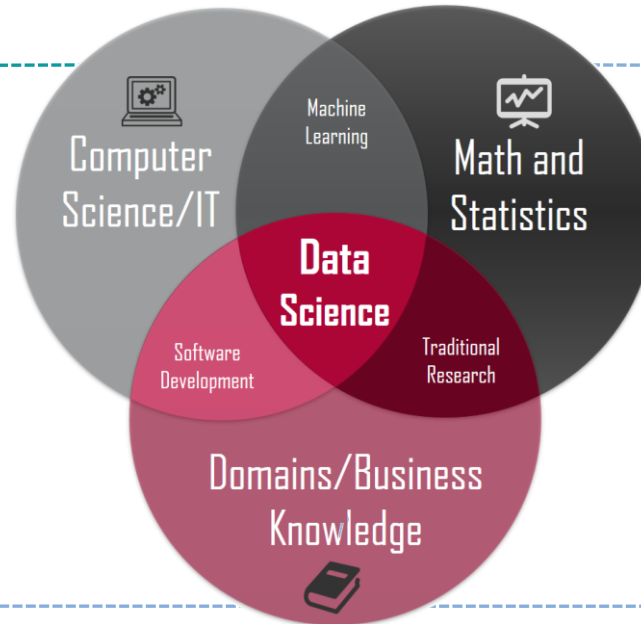
¿Cómo le agrego valor a los datos?



Perfil de un Científico de Datos

Data Scientist

Los científicos de datos combinan una variedad de conocimientos para analizar los datos recopilados de la web, de teléfonos inteligentes, de clientes, sensores y otras fuentes



Computer science

Manejo de diferentes de lenguaje de programación y conocimiento de IT



Math and Statistics

Manejo de las diferentes técnicas estadísticas y conocimiento de matemática avanzada



Business Knowledge

Conocimiento de la estructura y modelo de negocio, para que las soluciones sean las que cubran las necesidades de la organización



Inteligencia Artificial

La Inteligencia Artificial (IA) es la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano

Historia de la IA

1943-1956

Los inicios

- McCulloch y Pitts: primer modelo de neurona, demostración de que cualquier función es computable por una red de neuronas.
- Hebb: regla de aprendizaje Hebbiano o de Hebb
- Minsky, Edmonds: primer simulador de red neuronal (40 neuronas)
- Taller de Dartmouth (bautizo de la IA)
- Newell, Simon: Teórico Lógico (TL)

1966-1973

La dura realidad

- Muchos sistemas de IA actuaban de manera prometedora en problemas sencillos y fallaban ante problemas reales.
- Muchos problemas eran intratables (NP); el tamaño del problema es relevante.
- La búsqueda exhaustiva es impracticable.
- Minsky, Papert: en Perceptrons se demuestra la escasa capacidad de representación de los perceptrones.

1980-Actualidad

Contents

- El éxito de los sistemas expertos lleva a su uso comercial: R1, XCON (sistema de configuración en Digital Equipment Corporation).
- Japoneses: quinta generación de ordenadores
- Estadounidenses: MCC
- Software para el desarrollo de la IA
- Máquinas de arquitectura dedicada



1952-1969

Grandes expectativas

- SRGP: sistema de resolución general de problemas, imitación de las demostraciones humanas
- Samuel: programa jugador de damas, aprendizaje
- Sistemas planificadores, representación interna del mundo
- Aplicaciones en dominios reducidos: integración simbólica, mundo de bloques
- Redes neuronales: perceptrones, aprendizaje

1969-1979

Sistemas basados en conocimiento

- Los mecanismos generales de resolución de problemas no son practicables (métodos débiles).
- DENDRAL: el conocimiento del dominio del problema ayuda a resolverlo mejor.
- Sistemas Expertos (MYCIN, PROSPECTOR): incertidumbre, factores de certeza
- Representación del conocimiento y razonamiento (redes semánticas, marcos, Prolog, ontologías)

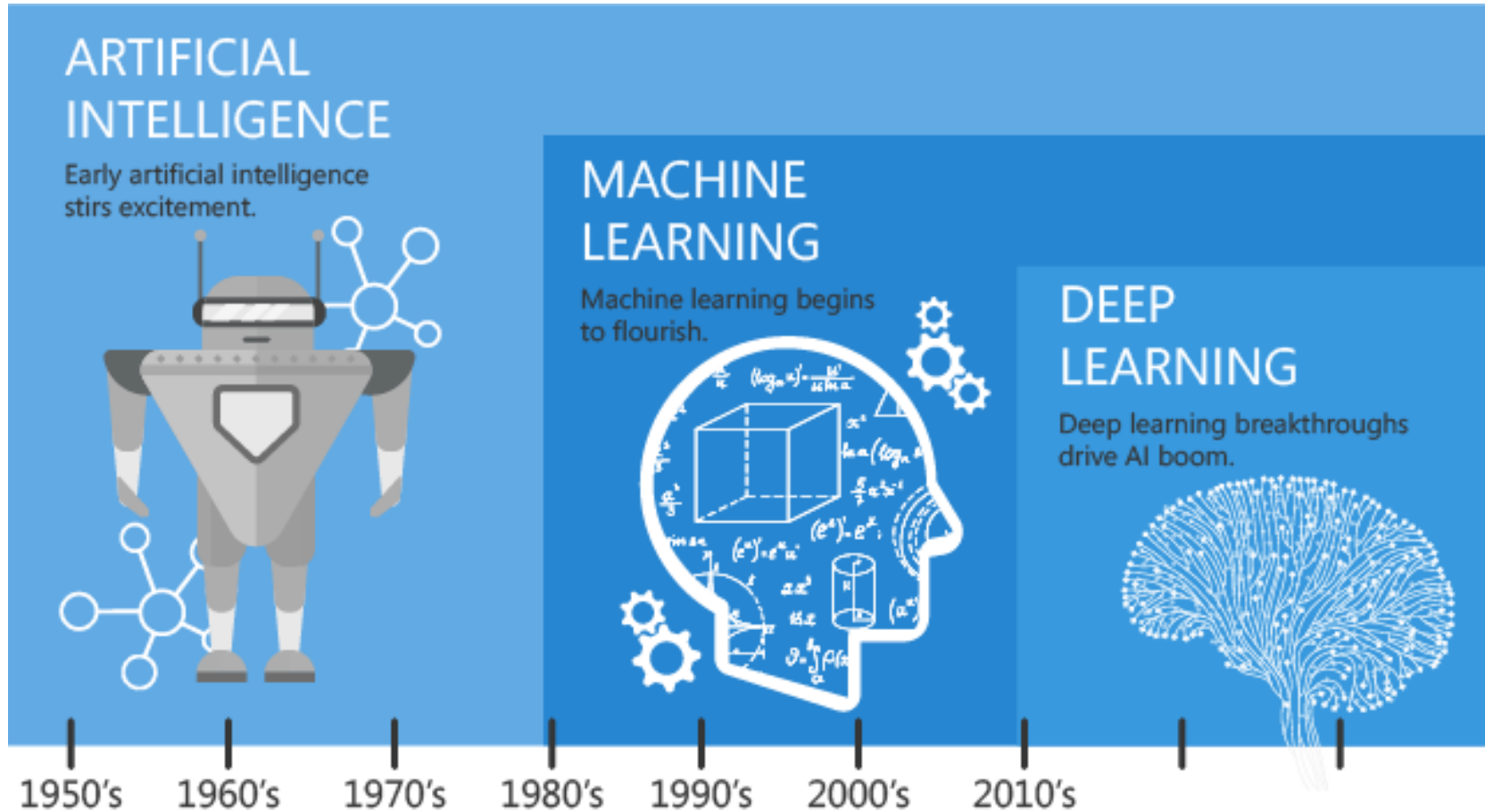
“La IA está transformando rápidamente la vida y los negocios americanos, mejorando la forma en que diagnosticamos y tratamos las enfermedades, cultivamos nuestros alimentos, fabricamos y entregamos nuevos productos, administramos nuestras finanzas, alimentamos nuestros hogares y viajamos del punto A al punto B.”
(1)

*“Los expertos prevén que continuarán los rápidos progresos en el campo de la inteligencia artificial especializada. Aunque **es muy poco probable que las máquinas exhiban una inteligencia ampliamente aplicable comparable o superior a la de los humanos en los próximos 20 años**, es de esperar que las máquinas alcancen y superen el rendimiento humano en más y más tareas.”* (2)

(1) <https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/05/Summary-Report-of-White-House-AI-Summit.pdf>

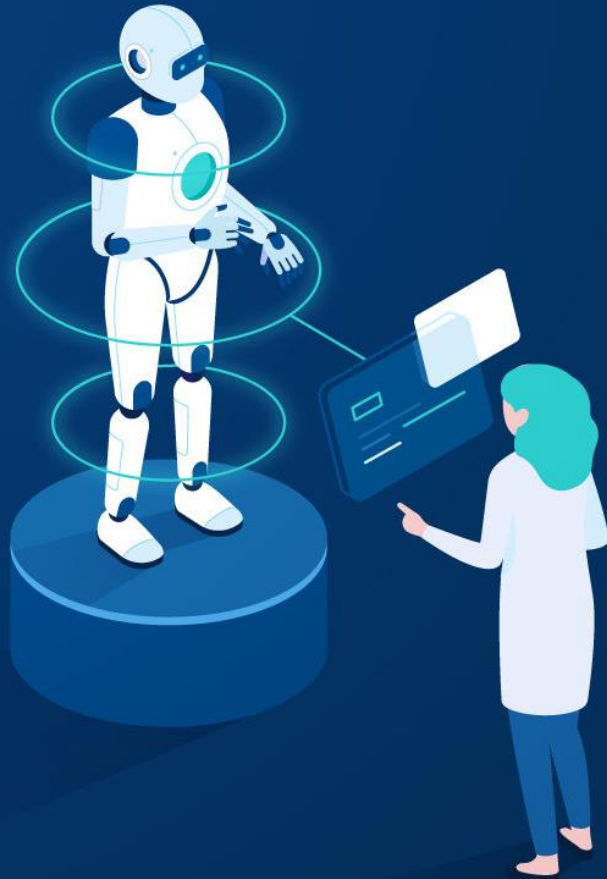
(2) https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Introduction to Machine Learning



IBM WATSON

SAS ML



Machine Learning

*“El aprendizaje automático es un subdominio de la IA que proporciona a los sistemas la capacidad de aprender y mejorar automáticamente a partir de la experiencia **sin ser programados explícitamente**. Se basa en la hipótesis de crear el modelo y trata de mejorarlo ajustando más datos en el modelo a lo largo del tiempo.”*

Introduction to Machine Learning



¿Qué es?

- El **machine learning** es un campo que se deriva de la IA. Consiste en desarrollar procesos que permitan a las máquinas aprender por sí solas a partir de los datos.
- No aprenden por sí mismas, sino que están programadas para adaptar su algoritmo conforme reciben información.



¿Cómo?

- No se ingresa manualmente las ordenes.
- Los algoritmos ML resuelven por su cuenta las adversidades que tienen al hacer inferencias a partir de los datos, y mientras más datos tengan, mejores serán los resultados.
- La clave está en que puede reencausar los resultados parciales que va obteniendo



¿Hacia donde vamos?

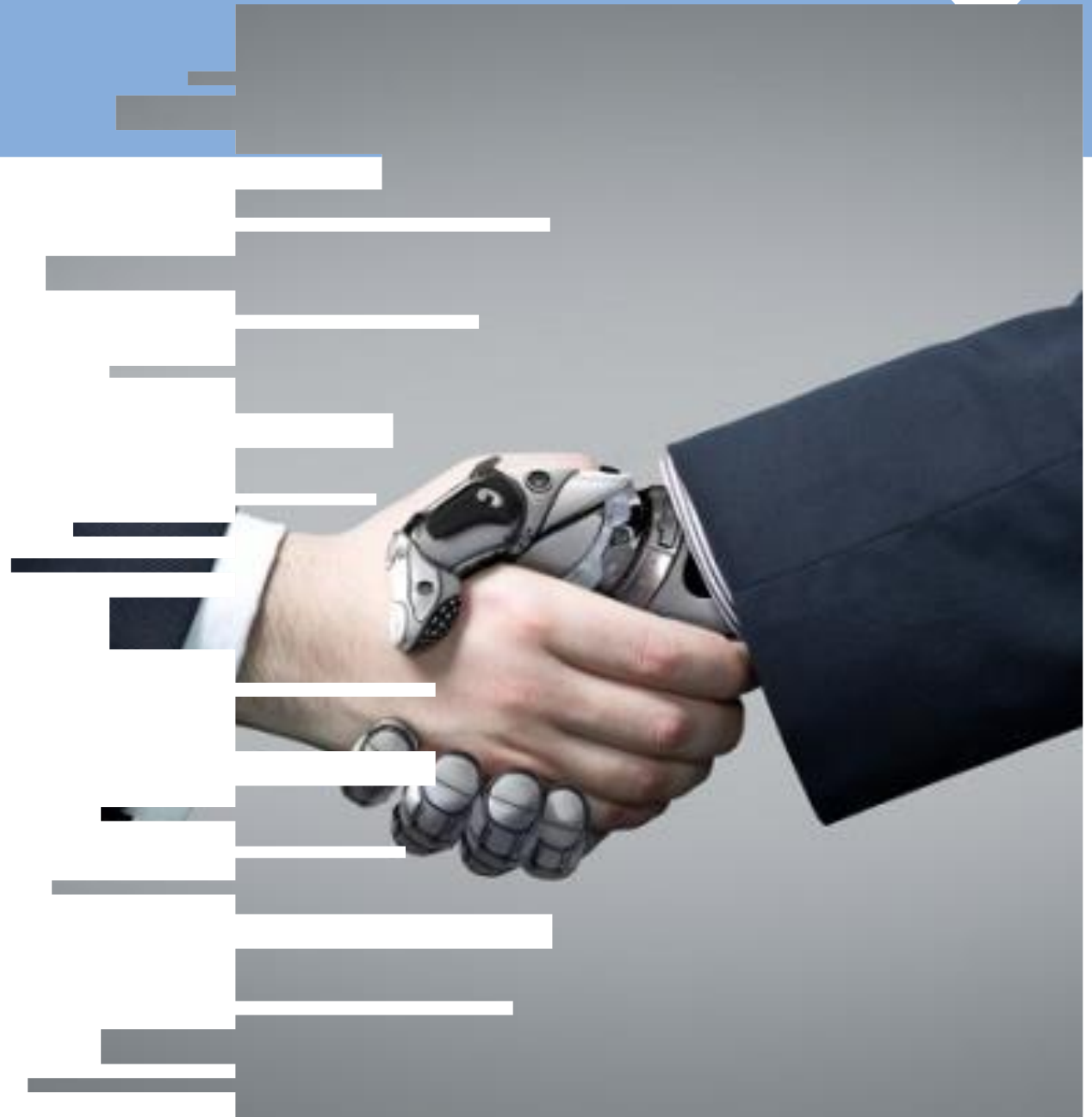
- Vamos en camino de complementar el trabajo humano con el trabajo de la máquina.
- Especializar las aplicaciones según rubro o actividades específicas. (Sistema experto).
- Las aplicaciones en el trabajo abarcan todo, desde mejorar las experiencias minoristas en las tiendas con IoT hasta aumentar la seguridad con datos biométricos hasta predecir y diagnosticar enfermedades



Aplicaciones de AA

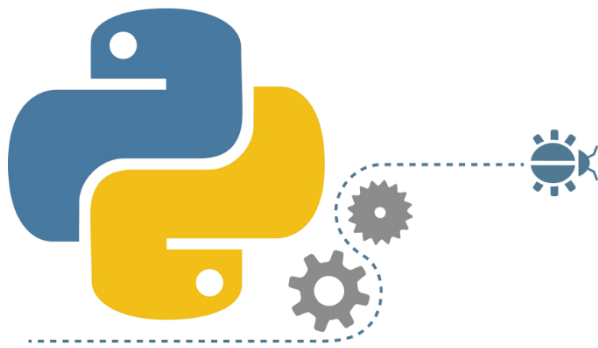
Según la revista Forbes:

1. Seguridad de los datos
2. Seguridad personal
3. Comercio financiero
4. Cuidado de la salud
5. Marketing personalizado
6. Detección de fraudes
7. Recomendaciones
8. Búsqueda online
9. Procesamiento de lenguaje natural
10. Coches inteligentes



A Introducción a Python

Python es un lenguaje de programación interpretado de tipado dinámico cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Se trata de un lenguaje de programación multiparadigma y disponible en varias plataformas.



Desarrollado por la
[Python Software Foundation](#)

Interpretado

Multiparadigma

Multiplataforma

Gratuito

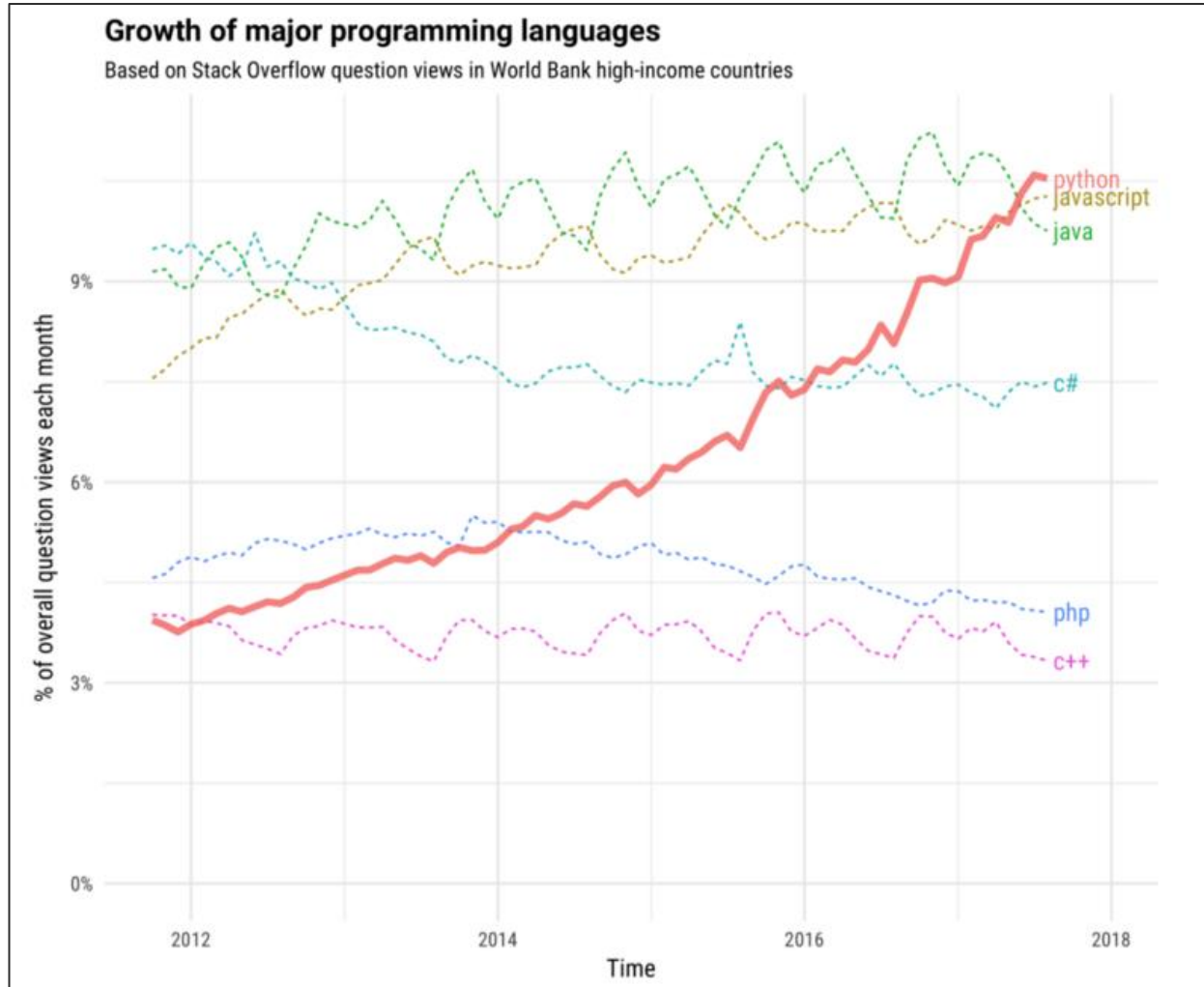
“Pilas incluidas”

```
def __init__(self, settings):
    self.file = None
    self.fingerprints = set()
    self.logdupes = True
    self.debug = debug
    self.logger = logging.getLogger(__name__)
    if path:
        self.file = open(os.path.join(path, 'requests.txt'), 'a')
        self.file.seek(0)
        self.fingerprints.update(s.request() for s in requests)

    @classmethod
    def from_settings(cls, settings):
        debug = settings.getbool('SUPERFINGER_DEBUG')
        return cls(job_dir(settings), debug)

    def request_seen(self, request):
        fp = self.request_fingerprint(request)
        if fp in self.fingerprints:
            return True
        self.fingerprints.add(fp)
        if self.file:
            self.file.write(fp + os.linesep)
```

A Python en Ciencia de Datos

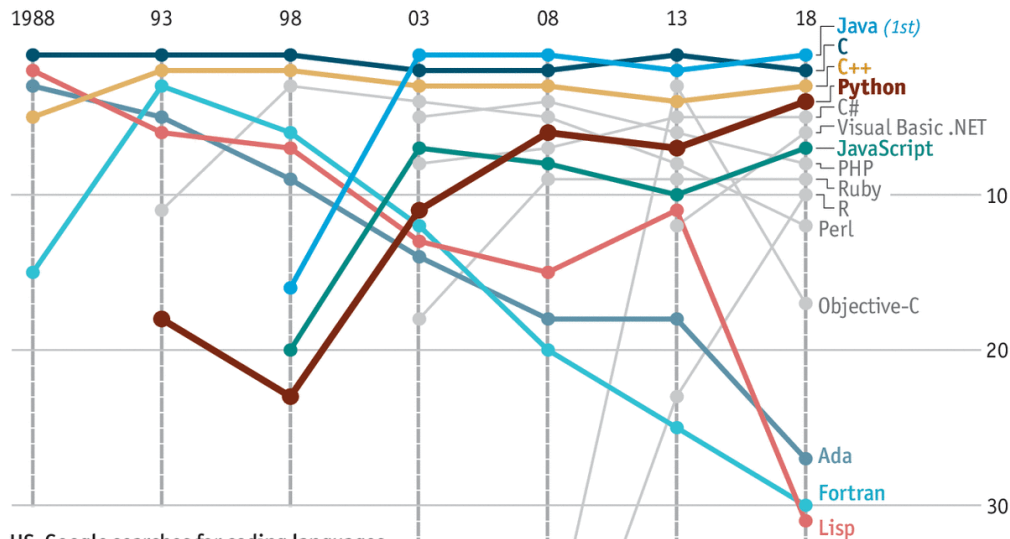


- ✓ Junio de 2017 fue el primer mes en el que **Python fue la etiqueta más visitada en Stack Overflow dentro de las naciones de altos ingresos**.
- ✓ Esto incluía ser la etiqueta más visitada dentro de los EE. UU. Y el Reino Unido, y en el top 2 en casi todos los demás países con ingresos altos (junto a Java o JavaScript).
- ✓ Esto es especialmente impresionante porque en 2012, fue menos visitado que cualquiera de los otros 5 idiomas, y se ha multiplicado por 2.5 en ese tiempo.

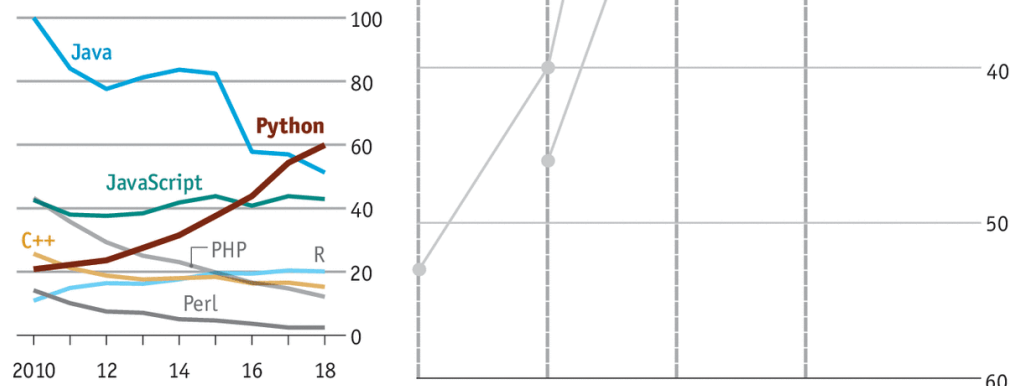
A Python en ciencia de datos

Code of conduct

Ranking of programming languages*



US, Google searches for coding languages
100 = highest annual traffic for any language



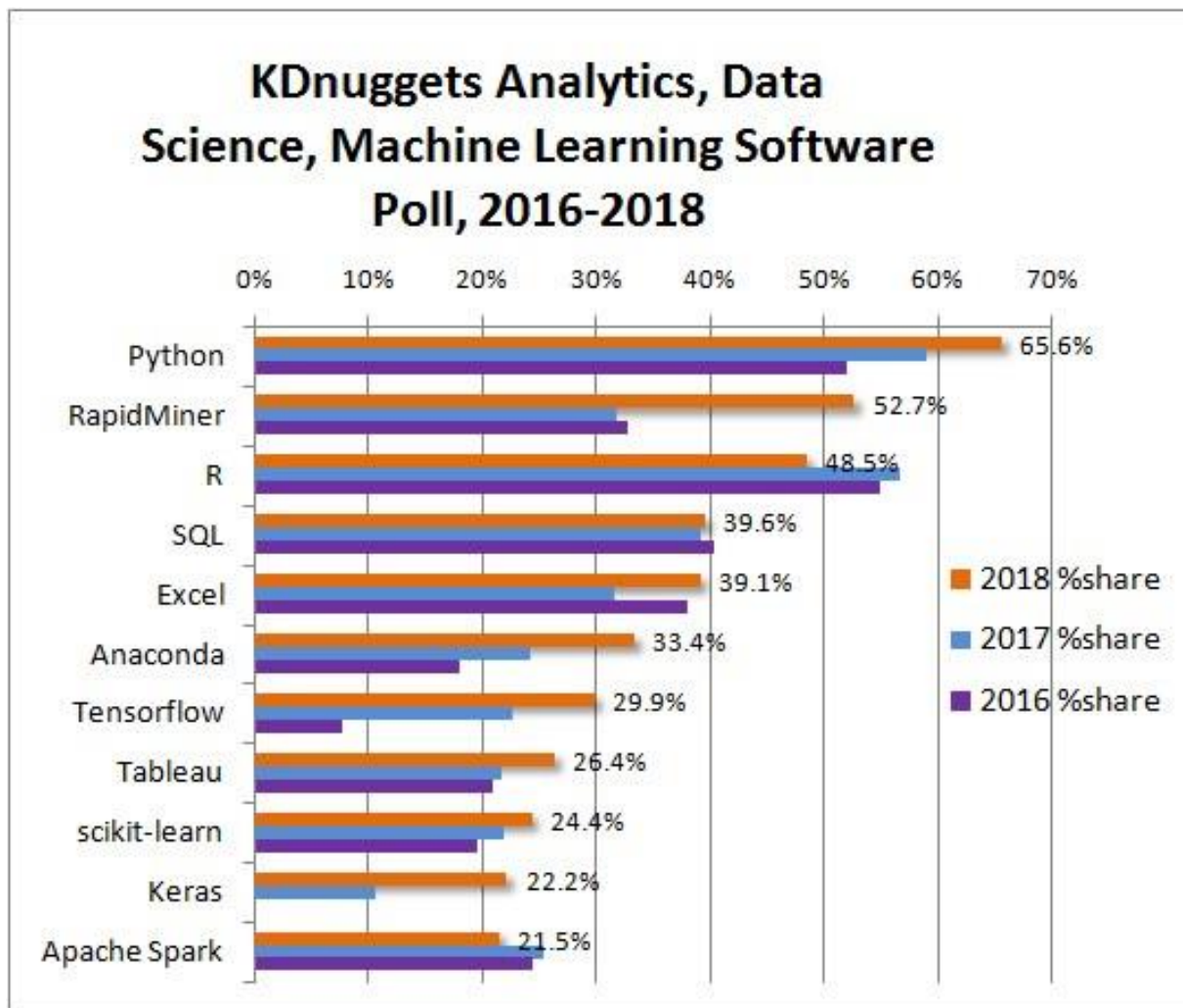
Source: TIOBE, Google Trends

*Ranked by global search-engine popularity

“No me propuse crear un lenguaje destinado al consumo masivo”, dice Guido van Rossum, un científico informático holandés que ideó Python, un lenguaje de programación, en 1989. Pero casi tres décadas después, su invento ha superado casi todos sus rivales y llevó la codificación a la punta de los dedos de las personas que una vez fueron desconcertados por ella.

En los últimos 12 meses, los estadounidenses han buscado Python en Google con más frecuencia que Kim Kardashian, una estrella de la televisión de realidad. El número de consultas se ha triplicado desde 2010, mientras que los de otros lenguajes de programación importantes han sido planos o en disminución.

A Python en ciencia de datos

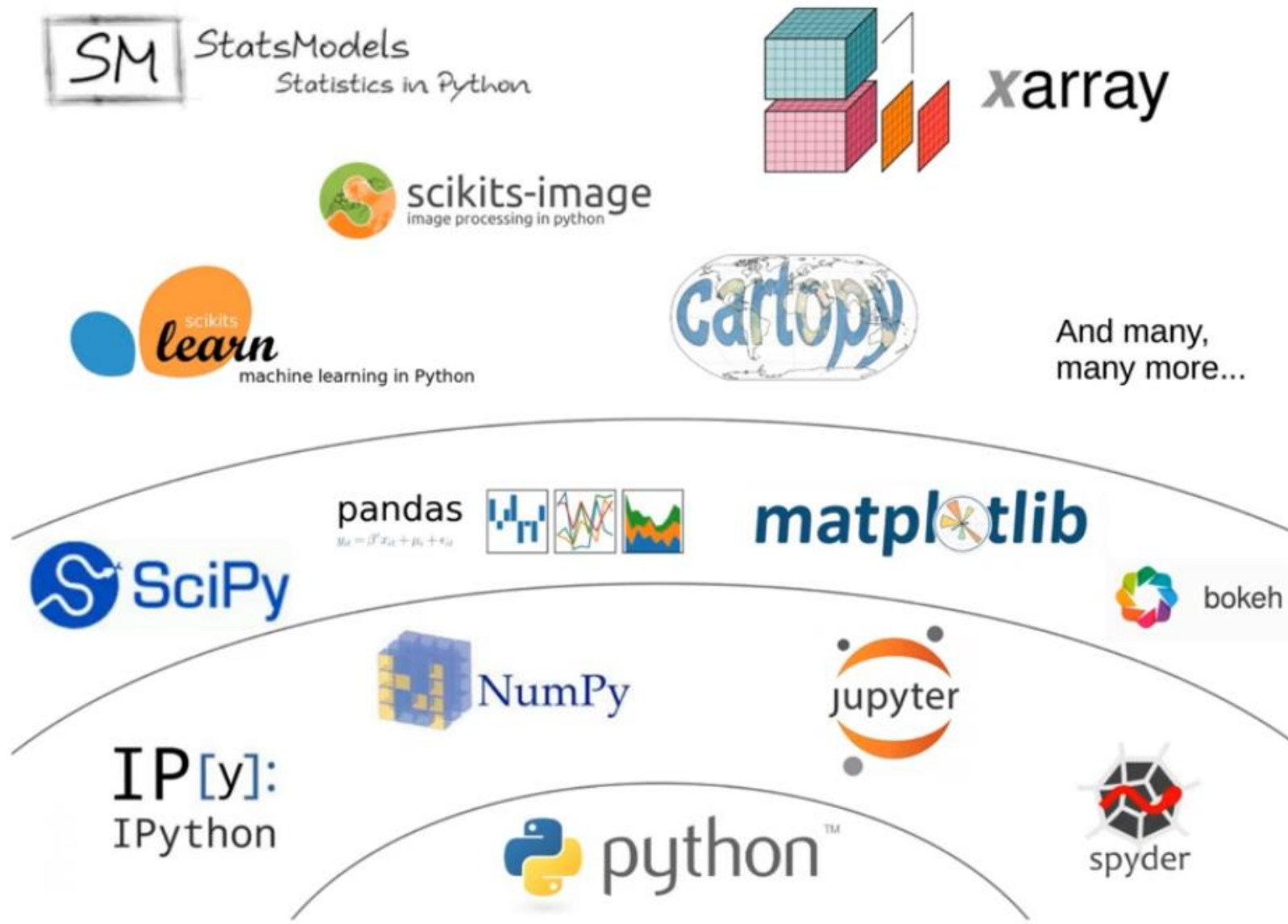


La 19ª encuesta anual de KDnuggets Software tuvo más de 2,300 votantes, algo menos que en 2017, tal vez porque solo un proveedor, RapidMiner, tuvo una campaña muy activa para votar en la encuesta de KDnuggets. En promedio, un participante seleccionó aproximadamente 7 herramientas diferentes utilizadas, por lo que los votos con solo una herramienta seleccionada se destacaron.

Eliminamos alrededor de 260 de estos votos "solitarios" (que principalmente fueron de RapidMiner), porque incluso si representaban a usuarios legítimos de esa herramienta, su experiencia era muy atípica y distorsionaría los resultados.



Python en ciencia de datos



¿Dónde aprender Python?



Introducción a jupyter notebooks

- Project Jupyter existe para desarrollar software de código abierto, estándares abiertos y servicios para computación interactiva en docenas de lenguajes de programación.
- Jupyter Notebook es una aplicación web de código abierto que le permite crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Los usos incluyen: limpieza y transformación de datos, simulación numérica, modelado estadístico, visualización de datos, aprendizaje automático y mucho más.



Idioma de elección

El portátil tiene soporte para más de 40 lenguajes de programación, incluidos Python, R, Julia y Scala.



Compartir cuadernos

Las computadoras portátiles se pueden compartir con otras personas mediante el correo electrónico, Dropbox, GitHub y el [visor de computadoras portátiles Jupyter](#).



Salida interactiva

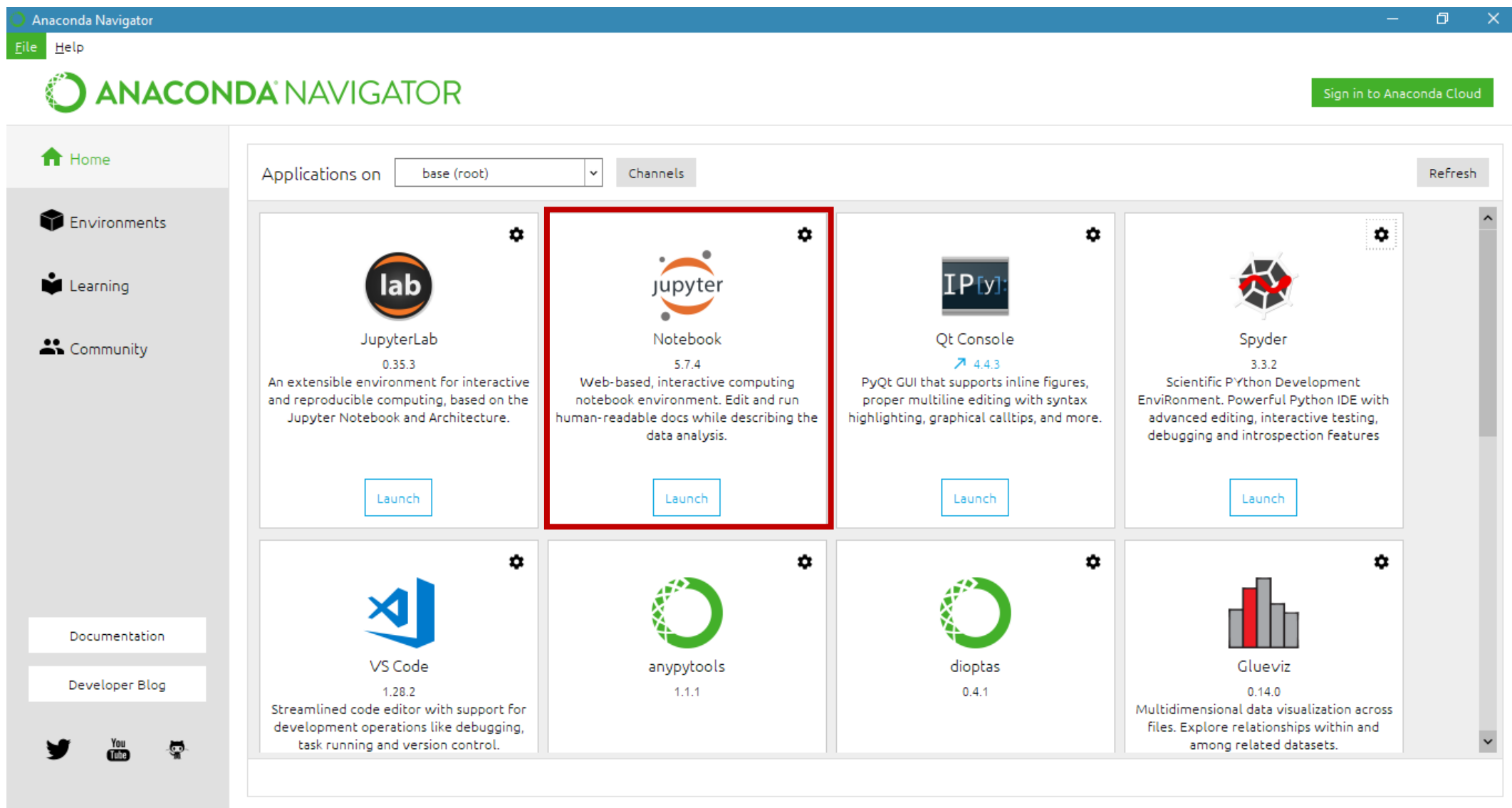
Su código puede producir una salida rica e interactiva: HTML, imágenes, videos, LaTeX y tipos MIME personalizados.



Integración de big data

Aproveche las herramientas de big data, como Apache Spark, de Python, R y Scala. Explora los mismos datos con pandas, scikit-learn, ggplot2, TensorFlow.

Jupyter desde anaconda



```
In [2]: print("Hola Mundo!")
Hola Mundo!!

In [4]: for i in range(10):
        print(i)
0
1
2
3
4
5
6
7
8
9
```

usamos negrita

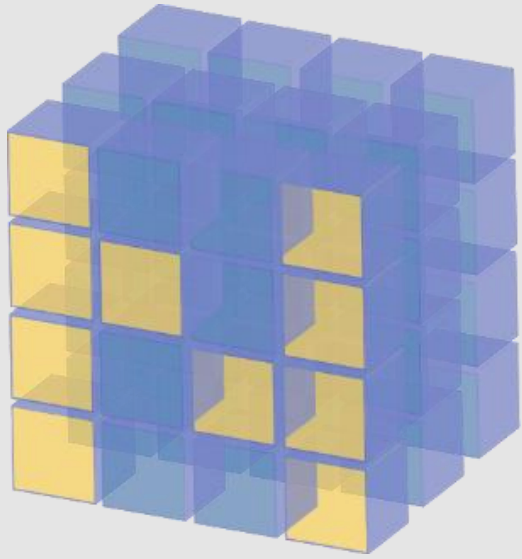
Usamos cursiva

Titulo

Subtitulo

- Ahora te toca a ti!!!

Jupyter Notebook



NumPy

Numpy

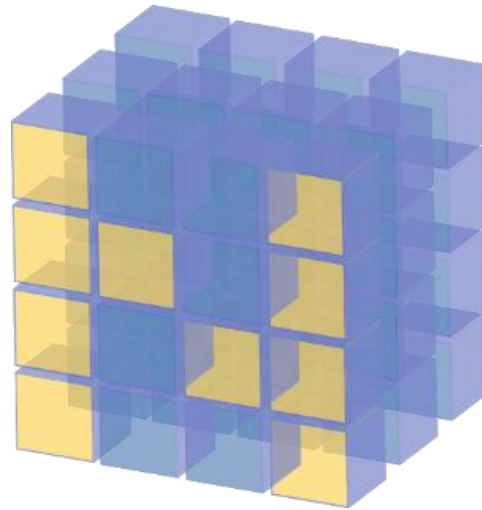
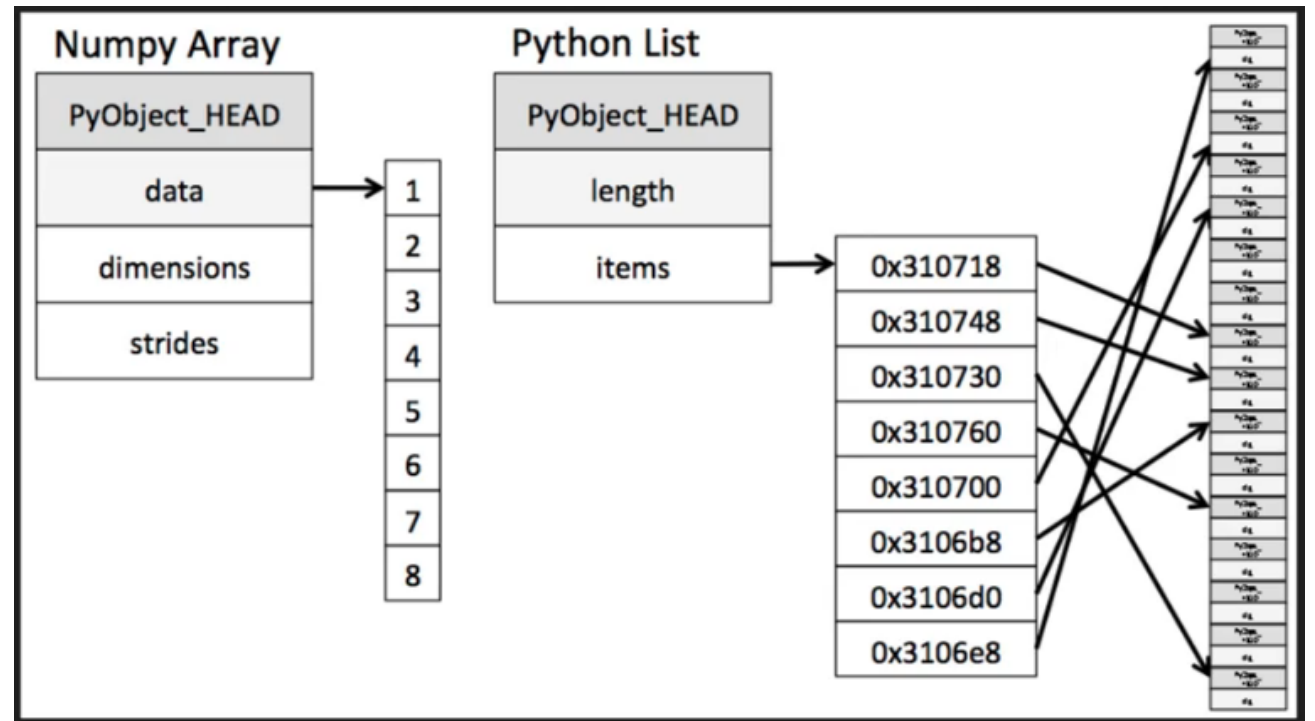
NumPy es el paquete fundamental para la computación científica con Python. Contiene entre otras cosas:

- un poderoso objeto de matriz N-dimensional
- Funciones sofisticadas (difusión)
- Herramientas para la integración de código C / C ++ y Fortran.
- Álgebra lineal útil, transformada de Fourier y capacidades de números aleatorios.

Además de sus obvios usos científicos, NumPy también se puede usar como un eficiente contenedor multidimensional de datos genéricos. Se pueden definir tipos de datos arbitrarios. Esto permite que NumPy se integre a la perfección con una amplia variedad de bases de datos.

NumPy está bajo la licencia BSD , lo que permite su reutilización con pocas restricciones.

Numpy

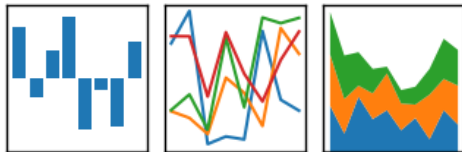


NumPy

Pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



¿Qué problema resuelven los *pandas* ?

- Python siempre ha sido excelente para la recopilación y preparación de datos, pero no tanto para el análisis y el modelado de datos. *pandas* lo ayuda a llenar este vacío, permitiéndole llevar a cabo todo su flujo de trabajo de análisis de datos en Python sin tener que cambiar a un idioma más específico del dominio como R.
- Combinado con el excelente kit de herramientas de IPython y otras bibliotecas, el entorno para realizar análisis de datos en Python es excelente en rendimiento, productividad y capacidad de colaboración.
- *pandas* no implementa una funcionalidad de modelado significativa fuera de la regresión lineal y de panel; Para esto, mira *statsmodels* y *scikit-learn* . Todavía se necesita más trabajo para hacer de Python un entorno de modelado estadístico de primera clase, pero esta bien encaminados hacia ese objetivo.

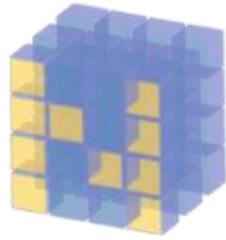
Pandas

Pandas



- Pandas es un paquete de Python que proporciona estructuras de datos rápidas, flexibles y expresivas diseñadas para hacer que el trabajo con datos y de series de tiempo sea fácil e intuitivo.
- Pretende ser el elemento fundamental de alto nivel para realizar análisis de datos prácticos y del mundo real en Python.
- Además, tiene el objetivo más amplio de convertirse en la herramienta de análisis / manipulación de datos de código abierto más potente y flexible disponible en cualquier idioma .
- Las dos estructuras de datos principales de pandas, Series y DataFrame, manejan la gran mayoría de los casos de uso típicos en finanzas, estadísticas, ciencias sociales y muchas áreas de ingeniería.
- Para los usuarios de R, **DataFrame** proporciona todo lo que el **data.frame** de R proporciona y mucho más. pandas está construido sobre **NumPy** y está diseñado para integrarse bien en un entorno de **computación científica** con muchas otras bibliotecas de terceros.

Pandas



NumPy

 PyTorch



plotly

matplotlib

theano



TensorFlow



Keras



python™

Natural Language Analyses
with NLTK