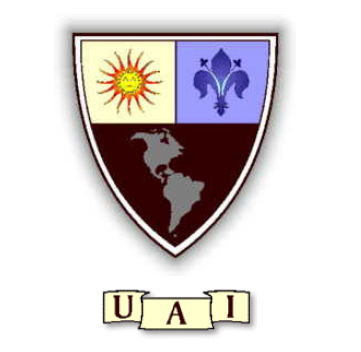
UNIVERSIDAD ABIERTA INTERAMERICANA



FACULTAD DE TECNOLOGÍA INFORMÁTICA

CARRERA: **Ingeniería en Sistemas Informáticos**

ALUMNO: **Perez Demonty; Torres; Gerardi; Destito Altieri**

MATERIA: Programación 2

Trabajo Práctico **Data Science**

AÑO: **2020**

Contenido

[**Introducción** 5](#_Toc55326891)

[**¿De dónde proviene el término Data Science?** 6](#_Toc55326892)

[**Conceptos clave en Data Science** 9](#_Toc55326893)

[¿Qué proceso sigue un Data scientist? 9](#_Toc55326894)

[**¿Qué se espera de un Data Scientist?** 10](#_Toc55326895)

[**Diferencias entre Data scientist y Big Data** 10](#_Toc55326896)

[**Proceso de recopilación de datos** 11](#_Toc55326897)

[Procesamiento y organización de los datos 12](#_Toc55326898)

[Análisis de los datos 13](#_Toc55326899)

[Desarrollo de modelos y algoritmos 13](#_Toc55326900)

[Visualización de los datos 15](#_Toc55326901)

[Toma de decisión 15](#_Toc55326902)

[**Python en Data Science** 16](#_Toc55326903)

[**10 ejemplos de aplicación de Big Data** 18](#_Toc55326904)

[1. Entendiendo y segmentando a los clientes 18](#_Toc55326905)

[2. Entendiendo y optimizando los procesos de negocio 19](#_Toc55326906)

[3. Cuantificación y optimización de rendimiento personal 20](#_Toc55326907)

[4. Mejorando la Salud Pública 21](#_Toc55326908)

[5. Mejorando el rendimiento deportivo 22](#_Toc55326909)

[6. Mejorando la Ciencia y la Investigación 24](#_Toc55326910)

[7. Optimizando el rendimiento de máquinas y dispositivos 25](#_Toc55326911)

[8. Mejorando la seguridad y el cumplimiento de la ley 26](#_Toc55326912)

[9. Marketing 27](#_Toc55326913)

[**10. Gobernanza**. 28](#_Toc55326917)

[**Resumen** 28](#_Toc55326918)

[**BIBLIOGRAFIA** 30](#_Toc55326919)

# **Introducción**

Data Science es la ciencia que se centra en el estudio de los datos. Se encarga de extraer información de grandes cantidades de datos.

[Data Science](https://www.neoland.es/data-science-bootcamp) combina la estadística, las matemáticas y la informática para interpretar datos. El objetivo de esta es lograr depurar los datos para transformarlos en información y utilizarlos para la tomar decisiones.

Estos datos se obtienen a través de diferentes canales. Los teléfonos móviles, las redes sociales, los e-commerce o las encuestas son solo algunas de las fuentes utilizadas. Nuestros gustos, rutinas o movimientos generan datos de gran valor para las empresas que quieren conocer a sus clientes al detalle. Sin embargo, la interpretación de los datos no estructurados no aporta valor a las compañías. De ahí surge la necesidad de contar con científicos de datos en sus equipos. Gracias al Data Science las empresas pueden anticiparse a la hora de tomar decisiones.

# **¿De dónde proviene el término Data Science?**

En 1962, [John W. Tukey](https://es.wikipedia.org/wiki/John_W._Tukey) precedió al término “Ciencia de Datos” en su artículo “The Future of Data Analysis” al explicar una evolución de la [estadística matemática](https://es.wikipedia.org/wiki/Estad%C3%ADstica_matem%C3%A1tica). En este, definió por primera vez el análisis de datos como: “Procedimientos para analizar datos, técnicas para interpretar los resultados de dichos procedimientos, formas de planificar la recopilación de datos para hacer su análisis más fácil, más preciso o acertado, y toda la maquinaria y los resultados de las estadísticas matemáticas que se aplican al análisis de datos.”​ En 1977 publicó “Exploratory Data Analysis”, argumentando que era necesario poner más énfasis en el uso de datos para sugerir hipótesis que probar en modelos estadísticos.

La ciencia de datos ha resultado para muchos una disciplina de reciente creación, pero en la realidad este concepto lo utilizó por primera vez el científico danés [Peter Naur](https://es.wikipedia.org/wiki/Peter_Naur) en la década de los sesenta como sustituto de las [ciencias computacionales](https://es.wikipedia.org/wiki/Ciencias_de_la_computaci%C3%B3n). En 1974 publicó el libro Concise Survey of Computer Methods donde utiliza ampliamente el concepto ciencia de datos, lo que permitió la utilización más libre en el mundo académico.

En 1977, el International Association for Statistical Computing (IASC) es establecido como una sección del [International Statistical Institute](https://es.wikipedia.org/wiki/International_Statistical_Institute) (ISI). “Es la misión de la IASC relacionar la metodología estadística tradicional, tecnología computacional moderna, y el conocimiento de expertos del tema, para convertir datos en información y conocimiento".

En 1996 el término ‘Ciencia de Datos’ fue utilizado por primera vez en una conferencia llamada "Ciencia de datos, clasificación y métodos relacionados", que tuvo lugar en una reunión de miembros de la ‘International Federation of Classification Societies’ (IFCS) con sede en [Kobe](https://es.wikipedia.org/wiki/K%C5%8Dbe), [Japón](https://es.wikipedia.org/wiki/Jap%C3%B3n).​ En 1997, [C.F. Jeff Wu](https://es.wikipedia.org/w/index.php?title=C.F._Jeff_Wu&action=edit&redlink=1) dio una charla llamada "Statistics = Data Science?", donde describió al trabajo estadístico como una trilogía conformada por recolección de datos, análisis y modelado de datos, y la toma de decisiones, haciendo la petición de que la estadística fuese renombrada como ciencia de datos, y los estadísticos como científicos de datos.

En 2001, [William S. Cleveland](https://es.wikipedia.org/w/index.php?title=William_S._Cleveland&action=edit&redlink=1) introdujo a la ciencia de datos como una disciplina independiente, extendiendo el campo de la estadística para incluir los avances en computación con datos en su artículo "Data science: an action plan for expanding the technical areas of the field of statistics". Cleveland estableció seis áreas técnicas que en su opinión conformarían al campo de la ciencia de datos: investigaciones multidisciplinarias, modelos y métodos para datos, computación con datos, pedagogía, evaluación de herramientas, y teoría.

En abril del 2002, el ‘International Council for Science: Committee on Data for Science and Technology’ ([CODATA](https://es.wikipedia.org/wiki/CODATA)) empezó la publicación del [Data Science Journal](https://es.wikipedia.org/w/index.php?title=Data_Science_Journal&action=edit&redlink=1),​ enfocada en problemas como la descripción de sistemas de datos, su publicación en Internet, sus aplicaciones, y sus problemas legales. Poco después, en enero del 2003, la [Universidad de Columbia](https://es.wikipedia.org/wiki/Universidad_de_Columbia) empezó a publicar [The Journal of Data Science](https://es.wikipedia.org/w/index.php?title=The_Journal_of_Data_Science&action=edit&redlink=1), la cual ofreció una plataforma para que todos los profesionales de datos presentaran sus perspectivas e intercambiaran ideas.

En 2005, [The National Science Board](https://es.wikipedia.org/w/index.php?title=The_National_Science_Board&action=edit&redlink=1) publicó "Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century", definiendo a los científicos de datos como "científicos de computación e información, programadores de bases de datos y software, y expertos disciplinarios, [...] que son cruciales para la gestión exitosa de una colección digital de datos, cuya actividad primaria es realizar investigación creativa y análisis".

Fue en el 2008 que Jeff Hammerbacher y DJ Patil lo reutilizaron para definir sus propios trabajos realizados en Facebook y LinkedIn, respectivamente,

En 2009, los investigadores Yangyong Zhu y Yun Xiong del ‘Research Center for Dataology and Data Science’, publicaron “Introduction to Dataology and Data Science”, en donde manifiestan que “a diferencia de las ciencias naturales y las ciencias sociales, Datología y Ciencia de Datos toman datos en la red y su objeto de estudio”.

En 2013 fue lanzado el ‘IEEE Task Force on Data Science and Advanced Analytics’,mientras que la primera conferencia internacional de ‘IEEE International Conference on Data Science and Advanced Analytics’ fue lanzada en el 2014.​ En 2015, el [International Journal on Data Science and Analytics](https://www.springer.com/computer/database%2Bmanagement%2B%26%2Binformation%2Bretrieval/journal/41060) fue lanzado por [Springer](https://es.wikipedia.org/wiki/Springer_Publishing) para publicar trabajos originales en ciencia de datos y analítica de [big data](https://es.wikipedia.org/wiki/Big_data)

# **Conceptos clave en Data Science**

Existen una serie de conceptos básicos que forman de la ciencia de los datos y que a continuación te explicamos brevemente en qué consisten.

## ¿Qué proceso sigue un Data scientist?

El proceso que sigue un Data Scientist para responder a las cuestiones que se le plantean se pueden resumir en estos 5 pasos:

* Extraer los datos, independientemente de su fuente (webs, csv, logs, APIs, etc.) y de su volumen (Big Data o Small Data).
* Limpiar los datos, para eliminar lo que distorsiona las mismas.
* Procesar los datos usando diferentes métodos estadísticos (inferencia estadística, modelos de regresión, pruebas de hipótesis, etc.).
* Diseñar nuevos tests o experimentos en caso necesario.
* Visualizar y presentar gráficamente los datos.

# **¿Qué se espera de un Data Scientist?**

Lo que se espera de un Data Scientist es que no sólo sea capaz de abordar un problema de explotación de datos desde el punto de vista de análisis, sino que también tenga las aptitudes necesarias para cubrir la etapa de gestión de datos. Así, el objetivo de un perfil de este tipo es acercar dos mundos (el de gestión y análisis de datos), que hasta ahora habían podido existir separados, pero que debido a los nuevos requisitos de volumen, de variedad de datos y de velocidad en la explotación de estas (ie, las tres V's de la definición estándar del término Big Data), se ha vuelto imprescindible llevar a cabo esta explotación a través de un perfil combinado, y que además, también entienda el negocio para dirigir esta explotación hacia resultados que puedan ser de interés para la compañía.

# **Diferencias entre Data scientist y Big Data**

Cuando hablamos de Big Data nos referimos a conjuntos de datos de gran volumen. Esto dificulta su almacenaje, gestión, procesamiento y análisis mediante tecnologías y herramientas convencionales. En definitiva, el Big Data se encarga de resolver los problemas de gestión y almacenamiento de datos. Esto permite dibujar patrones y obtener una visión más completa de los clientes.  Por otro lado, Data Science se centra en las herramientas que transforman los datos en información de valor.

# **Proceso de recopilación de datos**

A continuación, es necesario buscar datos que puedan ayudarnos a responder a las preguntas. Estos datos pueden provenir de diferentes fuentes, por ejemplo:

* Sistemas y aplicaciones,
* Búsquedas en Internet:
* Datos de organizaciones y empresas,
* Investigaciones.

La verdad es que son innumerables fuentes de datos, por eso es muy importante encontrar una fuente que presente la información de manera confiable y estructurada. Veamos algunos ejemplos:

* Para prever tu facturación en los próximos meses, puede ser interesante analizar cómo viene siendo la misma hasta el momento. Para ello, necesitarás extraer tu historial de ventas. Por otro lado, si entiendes que el resultado de tus ventas está directamente relacionado con el volumen de tráfico en tu sitio, es posible que encuentres los datos necesarios para responder a esta pregunta en una herramienta de análisis, como [Hotmart Analytics](https://blog.hotmart.com/es/hotmart-analytics/).
* Una de las maneras de descubrir el momento ideal de abordar al cliente y ofrecerle tu producto es analizar todas las acciones que ya has hecho con los clientes antiguos. Esta tarea la puedes hacer tanto con los que compraron tu producto como con los que no lo hicieron, tratando de comprender los motivos que llevaron a los clientes a comprar o no.
* Si tu duda tiene relación con el formato de contenido que te proporciona los mejores resultados, posts o vídeos, necesitarás cruzar los datos referentes al *engagement* de los leads que leyeron tus posts, los cuales se pueden obtener con Google Analytics, por ejemplo, utilizando los datos de *engagement* de los leads que vieron tus vídeos, obtenidos en [Youtube Analytics](https://blog.hotmart.com/es/youtube-analytics/).

## Procesamiento y organización de los datos

Con los datos ya recolectados, es muy importante hacer todo el proceso de limpieza, estandarización, procesamiento y organización de las informaciones. Esto sucede porque, casi siempre, los datos generados cuentan con una serie de inconsistencias que pueden perjudicar el análisis y llevar a la toma de decisiones erróneas.

Cuando los datos están organizados y procesados, ya es posible comenzar el análisis.

## Análisis de los datos

Hay varios tipos de análisis, desde los más simples hasta los sumamente complejos. Pero es importante recordar que, en la mayoría de los casos, un análisis básico ya puede reflejar un resultado de mucho valor para el negocio.

El motivo por el cual esto ocurre es muy sencillo: como muchas personas y empresas todavía no tienen el hábito de darle atención a los números, quien comienza a hacer análisis de datos (incluso análisis simples), generalmente está algunos pasos por delante de sus competidores.

## Desarrollo de modelos y algoritmos

En los casos en que el análisis de datos resulta muy complejo, o plantea más preguntas nuevas que respuestas a la pregunta inicial, quizá sea el momento de desarrollar modelos estadísticos y algoritmos para encontrar la solución que va a traerle un valor superior al negocio.

Estos modelos y algoritmos, por lo general, son necesarios cuando la “mente humana” ya no puede encontrar los mejores estándares para solucionar el problema, o cuando encontrarle una solución al problema podría tardar demasiado tiempo.

Un algoritmo puede ser utilizado para encontrar estándares que escapan de la percepción humana, o incluso para analizar millones de escenarios en pocos minutos, llevando a una decisión más asertiva en un corto intervalo de tiempo.

En el siguiente ejemplo podemos entender mejor cómo funciona.

Si tengo una base con 5.000 leads, y les envié 7 mensajes de correo electrónico por mes en los últimos 4 años, son más de 1,6 millones de eventos que tendría que analizar para tratar de encontrar un estándar de comportamiento de mis leads.

Por ese motivo, aunque concentrara todos mis esfuerzos en dicho análisis, probablemente me llevaría mucho tiempo encontrar estándares que un algoritmo podría identificar en segundos.

En caso de que yo quisiera entender la mejor forma de [anunciar en Facebook](https://blog.hotmart.com/es/como-vender-en-facebook/), podría analizar más de 50 indicadores diferentes por cada uno de los anuncios que ya he hecho.

Pero ¿cómo averiguar cuáles de esos indicadores realmente son relevantes para mi audiencia?

Si deseo probar qué detalles del diseño de mi página aumentan las posibilidades de que el visitante compre mi producto, yo necesitaría:

* Generar varias páginas diferentes,
* Separar grupos de usuarios que ingresarían a cada una de estas páginas y,
* Encontrar una manera de garantizar que el usuario de un grupo determinado solo tenga acceso a una sola página hasta que termine de evaluar los resultados de las diferentes páginas.

Por supuesto, se trata de algo aparentemente muy complejo, pero que se puede hacer de manera sencilla con algunas aplicaciones.

## Visualización de los datos

Después del uso de modelos y algoritmos, será necesario analizar visualmente los resultados para garantizar que las conclusiones del análisis están alineadas al objetivo de estudio.

Este análisis visual es realizado por gráficos que facilitan la detección de estándares y la toma de decisiones.

## Toma de decisión

Con los datos listos para ser analizados, llega el momento más importante de todos: la toma de decisiones estratégicas para tu negocio.

Al verificar los estándares encontrados podrá notar lo que está funcionando y lo que necesita ser mejorado. Esto te permitirá poner en práctica nuevas acciones y realizar pruebas para aprovechar los resultados obtenidos.

Estas decisiones, por supuesto, dependerán de cuál es tu negocio y de qué aspecto deseas optimizar.

Lo importante es que, antes de definir alguna acción a ser tomada, analices los datos que ya posees, a fin de optar por la decisión más asertiva.

# **Python en Data Science**

El Data Science se encarga de analizar, transformar datos y extraer información de utilidad para la toma de decisiones. Gracias a Python estas tareas se pueden llevar a cabo sin la necesidad de tener conocimientos avanzados de programación, con pocas líneas de código y en entornos de programación amigables que facilitan la programación y visualización de resultados.

Lo que nos lleva a preguntarnos: ¿Qué herramientas computacionales se pueden utilizar para desarrollar cada uno de los bloques?, la respuesta siempre será Python, porque cuenta con una gran cantidad de bibliotecas disponibles para cada uno, además de la gran comunidad de programadores que dan soporte en diferentes foros de programación.

Algunos ejemplos de estas bibliotecas son:

|  |  |
| --- | --- |
| **Nombre** | **Funcionalidad principal** |
| **Tensorflow** | Creación de redes neuronales artificiales. |
| **Seaborn** | Visualización de datos. |
| **Pandas** | Manipulación de conjuntos de datos |
| **NLTK** | Resolver problemas en el área del Procesamiento del Lenguaje Natural. |
| **NumPy** | Computación de datos en forma de matrices multidimensionales. |
| **Scikit-learn** | Procesamiento de datos y algoritmos de machine learning. |

# **10 ejemplos de aplicación de Big Data**

Sin embargo, hay unas áreas donde el big data está marcando la diferencia. Hemos resumido a continuación las aplicaciones de big data en 10 áreas donde actualmente se están concentrando las instalaciones y obteniendo los mejores resultados.

## 1. Entendiendo y segmentando a los clientes

Marketing y ventas son quizá las áreas de mayor aplicación de big data en la actualidad. Los datos se utilizan para comprender mejor a los clientes, sus comportamientos y preferencias. Las empresas están dispuestas a ampliar los centros de datos tradicionales con los de redes sociales, logs de navegación, análisis de textos y datos de sensores para obtener una imagen completa de su cliente. El objetivo principal en la mayoría de los casos es crear modelos predictivos. Como vimos en un artículo anterior, tuvo bastante repercusión [la noticia de que la cadena de distribución Target](http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#2e8af27934c6) fue capaz de detectar cuándo sus clientes esperaban un bebé. Las [empresas de telecomunicaciones](http://www.baoss.es/casos-de-uso-por-sectores/) pueden gracias al big data predecir mejor el churn de clientes. Los hipermercados pueden predecir mejor qué productos se venderán mejor, y las aseguradoras de coches pueden comprender mejor cómo conducen sus clientes. Incluso las campañas electorales pueden optimizarse gracias a big data analytics. Hay quienes sostienen que las recientes elecciones presidenciales en varios países han sido ganadas por los equipos con mayor capacidad de entender y aplicar el análisis de datos para buscar las preferencias de votantes y llegar a ellos por sus canales preferidos.

## 2. Entendiendo y optimizando los procesos de negocio

El big data se está utilizando cada vez más para optimizar los procesos de negocio en las empresas. En el sector de retail los negocios están optimizando su stock basándose en predicciones generadas gracias a datos de redes sociales, tendencias de búsquedas en la web y predicciones meteorológicas. Un proceso que se está transformando particularmente gracias al big data es el de la cadena de suministro y la optimización de rutas de reparto. Gracias al posicionamiento geográfico y sensores de identificación por radiofrecuencia se puede realizar un seguimiento de las mercancías y vehículos de reparto, optimizando las rutas, integrando datos de tráfico en tiempo real. Los procesos de recursos humanos también están siendo mejorados gracias al análisis del big data. Desde la detección y adquisición de talento, como en la [película Moneyball](http://www.baoss.es/big-data-y-business-analytics-en-el-deporte/), hasta la medición de la cultura empresarial y la involucración de la plantilla gracias a herramientas de big data.



## 3. Cuantificación y optimización de rendimiento personal

El big data no sólo es para empresas y para instituciones públicas o grandes organizaciones. Todos podemos beneficiarnos de los datos generados de [dispositivos wearables](http://www.pcmag.com/article2/0,2817,2404445,00.asp) como smart watches o pulseras. Estos dispositivos registran automáticamente datos de consumos de calorías (Fitbit), niveles de actividad y condición física (Google Fit, Apple Watch), o patrones de sueño. Aunque a nivel particular ya nos desvelan información interesante, el verdadero valor reside en analizar el conjunto de datos de todo el colectivo. Una de estas empresas, Jawbone, recoge en torno a 60 años de datos de sueño cada noche. Analizando estos volúmenes de datos desvelarán insights que beneficiarán a todos los usuarios.

Otro ejemplo donde la gente se beneficia del análisis del big data es para buscar a cupido. A la hora de encontrar a la mejor pareja compatible hacerlo sin la ayuda de algoritmos y técnicas de big data sería prácticamente imposible.

## 4. Mejorando la Salud Pública

Otra área de uso de datos masivos colectivos es el de la codificación de material genético. Cuantos más usuarios participan más beneficios se obtienen, bien para saber más sobre nuestros ancestros, qué dieta o alimentos son más adecuados para nuestro genotipo, o para descubrir cómo o porqué se activan determinados genes que pueden derivar en enfermedades crónicas. La capacidad de procesamiento de plataformas de análisis de big data nos permite ya descodificar cadenas enteras de ADN en cuestión de minutos y permitirá encontrar nuevos tratamientos y comprender mejor las enfermedades, sus desencadenantes y los patrones de propagación. Piense qué sucederá cuando todos los dispositivos y sensores que llevamos encima (y cada vez llevaremos más) con marcadores de nuestro cuerpo se apliquen a millones de otras personas. Los ensayos clínicos del futuro no tendrán que estar limitados a muestras pequeñas, sino que todo el mundo podrá formar parte.



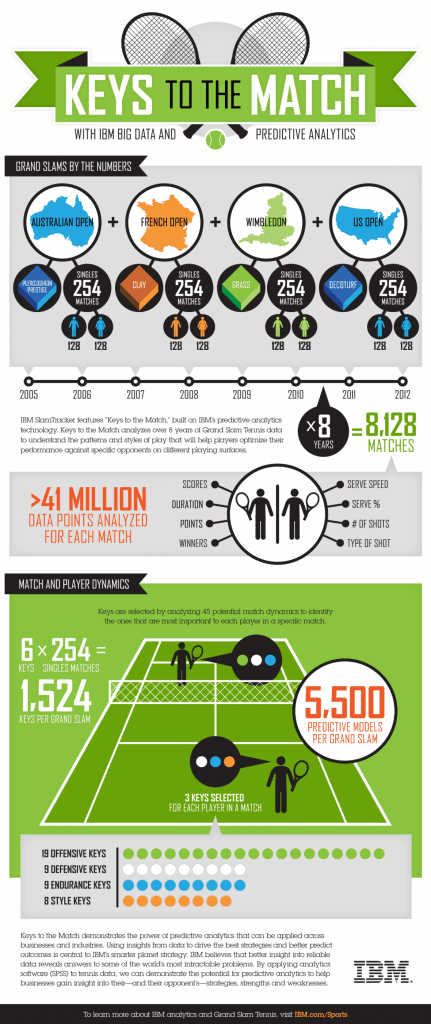
Las técnicas de big data ya están empleando por ejemplo para [monitorizar bebés](http://www.cnbc.com/2013/09/13/big-datas-powerful-effect-on-tiny-babies.html) en la unidad de neonatos de un hospital en Toronto. Grabando y analizando latidos y el patrón de respiración de cada bebé, la unidad ha desarrollado unos algoritmos que pueden predecir infecciones 24 horas antes de que los primeros síntomas aparezcan. De esta manera, el equipo médico puede intervenir y salvar vidas en un entorno en el que cada hora cuenta.

El análisis de datos masivos también se utiliza a la hora de controlar y predecir la evolución de las epidemias y brotes de enfermedades. Integrando datos de historiales clínicos con análisis de datos de redes sociales pueden [detectar brotes de gripe](http://www.baoss.es/convertir-el-big-data-en-valor/) en tiempo real simplemente escuchando lo que la gente publica en sus perfiles públicos.

Otras aplicaciones científicas prácticas llevan más tiempo: reconocimiento del habla, procesamiento imágenes en el cerebro para ciegos, pero es con el big data cuando los resultados están consiguiendo avances significativos.

## 5. Mejorando el rendimiento deportivo

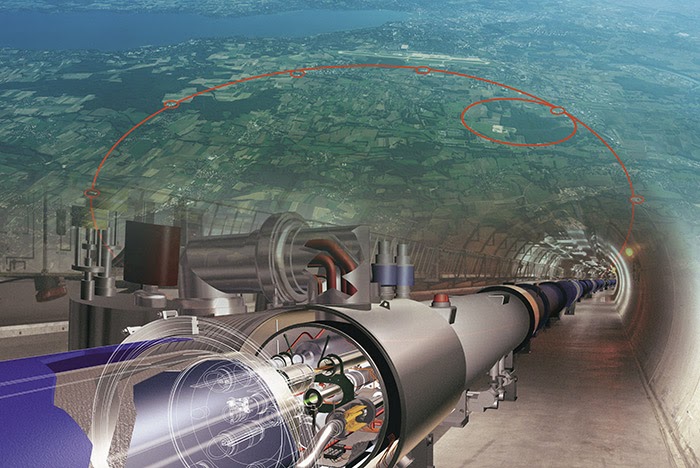
La mayor parte de deportistas de élite están ya adoptando técnicas de análisis de grandes volúmenes de datos. En tenis se lleva mucho tiempo utilizando la herramienta [SlamTracker](https://www-03.ibm.com/press/us/en/pressrelease/47578.wss) (basada en la tecnología de [IBM SPSS](http://www.baoss.es/content-analytics/ibm-spss-modeler/) de análisis predictivo) en los torneos más prestigiosos del mundo (Wimbledon, Roland Garros, Open de Australia). La plataforma lleva registrados más de 8 años de datos de Grand Slams (unos 41 millones de *data points*) para determinar patrones y estilos de jugadores ganadores.



Aplicando análisis a las grabaciones de vídeo de cada jugador de fútbol y tecnología de sensores en equipamiento deportivo como balones o palos de golf, podemos obtener retroalimentación de los datos generados y mostrar los resultados para mejorar el rendimiento o para mostrar una visión enriquecida de la retransmisión a los espectadores. Muchos equipos de élite realizan ya seguimiento de sus atletas fuera del entorno de competición, usando dispositivos inteligentes para monitorizar desde la nutrición, la preparación física y el sueño, hasta las conversaciones en redes sociales en las que participan para controlar el bienestar emocional.

## 6. Mejorando la Ciencia y la Investigación

La investigación científica se está viendo transformada por las nuevas posibilidades que ofrece el big data.



El CERN (laboratorio suizo de física nuclear con su gran colisionador de hadrones), uno de los mayores generadores de datos, intenta descubrir los secretos del universo gracias a los datos del acelerador de partículas. Aunque el centro de datos del CERN cuenta con 65.000 procesadores para analizar los 30 petabytes de datos, no es suficiente. Por ello distribuyen la capacidad de computación entre miles de ordenadores repartidos entre otros 150 centros de datos por todo el mundo para analizar los datos. Esta capacidad de computación distribuida que de otra manera sería imposible de procesar también se emplea en muchas otras áreas de la ciencia.

## 7. Optimizando el rendimiento de máquinas y dispositivos

El análisis de big data está ayudando a máquinas y dispositivos a ser más inteligentes y autónomos. Un ejemplo que ya es una realidad, el coche auto-pilotado de Google.



Los coches que usan para el proyecto están equipados con cámaras, GPS, conexión a internet, y un abanico de computadoras y sensores que permiten al vehículo circular de forma segura por la vía pública sin necesidad de intervención humana.

También se usan herramientas de análisis de big data para optimizar las redes de energía a partir de datos de los medidores inteligentes. podemos también aprovechar estas tecnologías para optimizar el rendimiento de servidores y datawarehouses.

## 8. Mejorando la seguridad y el cumplimiento de la ley

El análisis de big data se está empleando de forma intensiva en la mejora de la seguridad y en los cuerpos de aplicación de la ley. La noticia que se filtró via Wikileaks de que la NSA ha estado espiando en todas las comunicaciones de todos los ciudadanos. El objetivo es la protección de ataques terroristas.

## 9. Marketing

En septiembre de 1994, [BusinessWeek](https://es.wikipedia.org/wiki/Bloomberg_Businessweek) publicó el artículo “Marketing de base de datos”, manifestando que las empresas recopilan una gran cantidad de información sobre los clientes, la cual es analizada para predecir la [probabilidad](https://es.wikipedia.org/wiki/Probabilidad) de que compre un producto. Afirman que se utiliza ese conocimiento para elaborar un mensaje de [marketing](https://es.wikipedia.org/wiki/Marketing) calibrado con precisión para que el individuo busque conseguirlo. Asimismo, explican que, en los ochentas, un entusiasmo provocado por la propagación de los [lectores de códigos de barras](https://es.wikipedia.org/wiki/Lector_de_c%C3%B3digo_de_barras) terminó en una decepción generalizada pues muchas empresas fueron abrumadas por la gran cantidad de datos para lograr hacer algo útil con la información de sus clientes. Sin embargo, muchas empresas creen que no hay más remedio que desafiar la frontera marketing y bases de datos para desarrollar más las tecnologías necesarias.

En 2014 la empresa sueca de música en [streaming](https://es.wikipedia.org/wiki/Streaming) [Spotify](https://es.wikipedia.org/wiki/Spotify) compra [The Echo Nest](https://es.wikipedia.org/wiki/The_Echo_Nest), una compañía especializada en ciencia de datos musicales. Esta, es ahora la encargada de almacenar y analizar la información de sus 170 millones de usuarios. Con ayuda de dicha empresa, en 2015 Spotify lanzó un servicio de música personalizada llamado Discover Weekly que semanalmente recomienda a sus usuarios una selección de canciones que podría interesarles por medio de algoritmos y análisis de los datos de la música escuchada y el historial de búsqueda de la semana pasada. El servicio recibió una buena recepción generalizada y actualmente figura un fuerte punto de venta ante la competencia de la empresa.

[Netflix](https://es.wikipedia.org/wiki/Netflix), la empresa norteamericana de contenido multimedia en streaming ofrece a sus más de 120 millones de usuarios una plataforma capaz de analizar, mediante algoritmos, las costumbres de consumo de los usuarios para diferenciar los contenidos que estos buscan y lograr determinar qué nuevos contenidos les pueden interesar. [Todd Yellin](https://es.wikipedia.org/w/index.php?title=Todd_Yellin&action=edit&redlink=1), vicepresidente de producto en Netflix, explicó que algunos de los datos almacenados pueden extenderse desde la hora del día se conectan sus usuarios, cuánto tiempo pasan dentro de la plataforma, su lista de contenidos recientemente vistos (para analizar incluso el orden específico de estos). Toda la información que se almacena es utilizada específicamente para ser analizada, aprender del usuario y poder darle recomendaciones acertadas.

**10. Gobernanza**

En [América Latina](https://es.wikipedia.org/wiki/Am%C3%A9rica_Latina) el [Banco Interamericano de Desarrollo](https://es.wikipedia.org/wiki/Banco_Interamericano_de_Desarrollo) (BID) ha desarrollado estudios exploratorios en los que se analiza la ciencia de datos en la implementación y diseño de políticas públicas en la región, tomando casos en países como [Argentina](https://es.wikipedia.org/wiki/Argentina) y [Brasil](https://es.wikipedia.org/wiki/Brasil), presentando recomendaciones para su implementación y mantenimiento. Estas van desde temas como movilidad urbana sostenible, ciudades inteligentes, seguridad, propiedad de datos y privacidad. Entre las sugerencias presentadas en las investigaciones está la de lograr una “inteligencia del valor público, la cual “tiene la potencialidad de ser un componente estratégico para la toma de decisiones y el diseño, implementación y evaluación de políticas públicas”. Otra de ellas es la capacidad para lograr desde este campo una mejora de rendición de cuentas de los gobiernos ante la ciudadanía y promover un avance en cuanto a la curaduría de datos en las instituciones pública.

# **Resumen**

Entonces podemos decir que Data Science es la ciencia centrada en el estudio de los datos. Se encarga de extraer información de grandes cantidades de datos. Combinando la estadística, las matemáticas y la informática para interpretar datos.

Para poder llevar a cabo dicho cometido hay que aplicar los siguientes procesos:

* Extraer los datos, independientemente de su fuente (webs, csv, logs, APIs, etc.) y de su volumen (Big Data o Small Data).
* Limpiar los datos, para eliminar lo que distorsiona las mismas.
* Procesar los datos usando diferentes métodos estadísticos (inferencia estadística, modelos de regresión, pruebas de hipótesis, etc.).
* Diseñar nuevos tests o experimentos en caso necesario.
* Visualizar y presentar gráficamente los datos.

El objetivo de esta es lograr depurar los datos para transformarlos en información y utilizarlos para la tomar decisiones.

Si los datos aun no llegan a ser comprensibles para un ojo no entrenado, se aplicarán distintos algoritmos para poder filtrar la información de la manera deseada, así el cliente puede comprenderlo.

Es ideal para esto utilizar Python porque se pueden llevar a cabo sin la necesidad de tener conocimientos avanzados de programación, con pocas líneas de código y en entornos de programación amigables que facilitan la programación y visualización de resultados. Además de contar con una gran cantidad de bibliotecas disponibles para cada uno, además de la gran comunidad de programadores que dan soporte en diferentes foros de programación.

Data science revoluciono la manera de analizar los negocios y distintos aspectos de la vida, como, por ejemplo:

Para entender y segmentar a los clientes, también optimizar los procesos de negocio, analizando las ventas, cuando sean mayores o menores, productos mas vendidos y menos, etc. Para la Cuantificación y optimización de rendimiento personal y deportivos grupales, para la realización de campañas publicitarias, teniendo en cuenta el publico al cual va dirigido, para mejorar la seguridad informática, entre muchas otras aplicaciones posibles.

# **BIBLIOGRAFIA**

<https://es.wikipedia.org/wiki/Ciencia_de_datos>

<https://inlab.fib.upc.edu/es/blog/que-es-un-data-scientist>

<https://www.neoland.es/blog/que-es-data-science>

<https://blog.hotmart.com/es/que-es-data-science/#:~:text=%20Algunos%20ejemplos%20comunes%20del%20uso%20de%20la,de%20los%20seguidores%20en%20las%20redes...%20More%20>