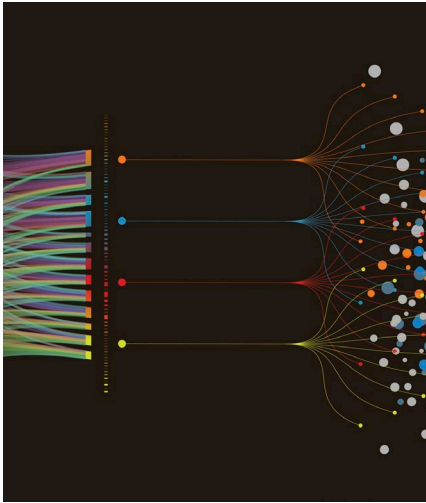


Bloque 1.

Tema 1:
Introducción
a data science

¿Qué es data science?

Es una disciplina que utiliza métodos y procesos científicos para obtener información de diversas fuentes de datos. Asimismo, extrae, prepara, limpia, presenta, codifica, analiza, visualiza y brinda mantenimiento a la información.



Data science o **ciencia de datos** es una combinación multidisciplinaria de procesamiento de datos, desarrollo de algoritmos y tecnología para resolver problemas analíticamente complejos. Por consiguiente, el actor principal son los **datos**, que son grandes cantidades de información sin procesar transmitida y almacenada en *data warehouses* empresariales.

La ciencia de datos se trata en última instancia de utilizar estos datos de manera creativa para generar valor comercial. Por lo tanto, es una combinación de habilidades en tres áreas principales (Lo, s.f.):

- Experiencia matemática.
- Tecnología y piratería informática.
- Fuerte perspicacia para los negocios.

Para saber más sobre **qué es data science**, te invitamos a ver el siguiente video:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.



Simplilearn. (2018, 4 de diciembre). *Data Science in 5 Minutes | Data Science For Beginners | What Is Data Science? | Simplilearn* [Archivo de video]. Recuperado de <https://bit.ly/2Of4hyL>

¿Qué es un data scientist?

Los científicos de datos son expertos en datos analíticos y poseen habilidades técnicas para resolver problemas complejos, así como la curiosidad de explorar aquellos que necesitan resolverse. No existe una descripción definitiva del puesto, pero los científicos de datos desempeñan las siguientes funciones (SAS, s.f.):

- Manejar diversos lenguajes de programación.
- Tener amplio dominio de la estadística.
- Recopilar grandes cantidades de datos para transformarlos en productos o formatos utilizables.
- Resolver problemas relacionados con los negocios empleando técnicas basadas en datos.

Para saber más sobre **qué es un científico de datos**, te invitamos a ver los siguientes videos:



- Borne, K. (s.f.). *Why Data Science is a Business Function* [Archivo de video]. Recuperado de <https://bit.ly/2CP4VOd>
- 365 Data Science. (2018, 14 de agosto). *Can You Become a Data Scientist?* [Archivo de video]. Recuperado de <https://bit.ly/2OPIKy6>

Los científicos de datos utilizan algunas de las siguientes herramientas y lenguajes de programación de acuerdo con sus necesidades (DataFlair Team, 2019):



Es un lenguaje de programación de secuencias de comandos y fue específicamente diseñado para operaciones estadísticas, además de ser orientado a objetos. Asimismo, es ampliamente utilizado para análisis de datos, modelado estadístico, pronóstico de series de tiempo, agrupación, etcétera.

Es un lenguaje versátil de programación de alto nivel que se utiliza principalmente para la ciencia de datos y el desarrollo de software. Asimismo, ha ganado popularidad debido a su facilidad de uso y su legibilidad de código. Como resultado, es ampliamente utilizado para análisis de datos, procesamiento de lenguaje natural y visión por computadora.



Se utiliza para extraer, administrar, manipular y consultar datos que están almacenados en bases de datos.

Es un software de código abierto que regula el almacenamiento de conjuntos de datos masivos. De la misma manera, emplea un sistema de almacenamiento distribuido mediante el modelo *MapReduce*. Por consiguiente, se ha convertido en el software más popular para big data debido a su rápida capacidad de procesar datos colosales, su arquitectura escalable y su implementación de bajo costo.





Es un software de visualización de datos especializado en análisis gráfico de datos, permitiendo a sus usuarios crear visualizaciones interactivas y paneles. Por consiguiente, es una opción ideal para mostrar diversas tendencias de datos en forma de gráficos interactivos, histogramas, diagramas de cajas, etc. Además, una característica importante de Tableau es su capacidad para conectarse con hojas de cálculo, bases de datos relacionales y plataformas en la nube.

Es usado generalmente para la minería de datos, pero también consta de varias herramientas necesarias para las operaciones de machine learning. Además, es un software de código abierto que utiliza la interfaz GUI, lo que facilita la interacción de los usuarios, ya que no requiere ninguna línea de código.




¿Cómo ayuda data science a los individuos y a las empresas en la toma de decisiones?

La respuesta es muy sencilla, ya que en vez de tomar en cuenta predicciones o supuestos, se toman decisiones con base en datos, hechos, información y análisis. A continuación, te presentamos las siguientes ventajas:


- Veracidad, confiabilidad y fiabilidad.
- Oportunidad.
- Manejo y análisis de gran cantidad de datos a la vez.
- Presentación del análisis en herramientas y productos entendibles.
- Ventaja sobre la competencia.
- Ahorro tiempos y costos.
- Elaboración de pronósticos y administración de inventarios.
- Exposición de necesidades, áreas de oportunidad y situación actual de una empresa, departamento o circunstancia a evaluar.

A pesar de que existe una gran variedad de herramientas muy amigables, libres y gratuitas para realizar el análisis de nuestros datos e información, deberás contar con asesoría especializada, ya sea de manera interna o contratando a expertos externos, para asegurarte de aprovechar correctamente todas las bondades y posibilidades que te brinda cada herramienta.

La buena noticia es que, por medio de este certificado, tú puedes ser el científico de datos que tu empresa o proyecto necesita.



Según Glassdoor, que es un portal especializado en la búsqueda de empleo, **el profesional en data science fue el mejor remunerado y con mayor índice de satisfacción en el 2019**. Además, el salario anual promedio que reciben estos profesionistas en los Estados Unidos es de 110 mil dólares (Guijosa, 2019).





IBM Data Science platform

Los equipos de ciencia de datos pueden ayudar a varios departamentos utilizando un conjunto diverso de herramientas y técnicas disponibles en IBM Data Science platform. Asimismo, puedes descubrir que la ciencia de datos ayuda a escalar modelos predictivos, prediciendo la implementación de los ajustadores y las reclamaciones fraudulentas. Te invitamos a que veas el siguiente video para que conozcas la amplia variedad de soluciones que proporciona IBM:



IBM Data and AI. (2018, 20 de junio). *Solving Business Issues with Data Science. Learn from this Insurance Use Case* [Archivo de video]. Recuperado de <https://bit.ly/2rLCka9>

Referencias

- DataFlair Team. (2019). *What is Data Science? A Complete Data Science Tutorial for Beginners*. Recuperado de <https://data-flair.training/blogs/what-is-data-science/>
- Guijosa, C. (2019). *Nueva academia de Ciencia de Datos en Coursera*. Recuperado de <https://bit.ly/2O06Buw>
- Lo, F. (s.f.). *What is Data Science?* Recuperado de <https://bit.ly/2rQf4YR>
- SAS. (s.f.). *What is a Data Scientist? Who they are, what they do and why you want to be one*. Recuperado de <https://bit.ly/2CP4VOd>

Recursos adicionales que te recomendamos conocer para expandir tu conocimiento:

Los siguientes enlaces son externos a la Universidad Tecmilenio, al acceder a estos considera que debes apegarte a sus términos y condiciones.

Podcasts

Para conocer más sobre **la toma de decisiones y data science**, te recomendamos escuchar los siguientes podcasts:

- Cutrell, J. (18 de diciembre de 2017). *Data Science w/ Elena Grewal (Part 1)* [Audio podcast]. Recuperado de <https://spec.fm/podcasts/developer-tea/101053>
- Cutrell, J. (20 de diciembre de 2017). *Data Science w/ Elena Grewal (Part 2)* [Audio podcast]. Recuperado de <https://spec.fm/podcasts/developer-tea/101055>
- Cutrell, J. (26 de abril de 2017). *Interview with Chris Albon (Part 1 of 3)* [Audio podcast]. Recuperado de <https://spec.fm/podcasts/developer-tea/66408>
- Cutrell, J. (28 de abril de 2017). *Interview with Chris Albon (Part 2 of 3)* [Audio podcast]. Recuperado de <https://spec.fm/podcasts/developer-tea/66410>
- Cutrell, J. (1 de mayo de 2017). *Interview with Chris Albon (Part 3 of 3)* [Audio podcast]. Recuperado de <https://spec.fm/podcasts/developer-tea/66418>
- Data Latam. (10 de marzo de 2019). *Ciencia de datos y su impacto en políticas públicas* [Audio podcast]. Recuperado de <http://www.datalatam.com/46/>

MOOC

Te recomendamos cursar el siguiente MOOC gratuito para principiantes:

- IBM. (s.f.). *Programa especializado Introduction to Data Science*. Recuperado de <https://es.coursera.org/specializations/introduction-data-science>

Te recomendamos cursar el siguiente MOOC gratuito para adquirir herramientas y habilidades propias del tema:

- IBM. (s.f.). *Certificado profesional de Ciencia de datos de IBM*. Recuperado de <https://www.coursera.org/professional-certificates/ibm-data-science?action=enroll>

La obra presentada es propiedad de ENSEÑANZA E INVESTIGACIÓN SUPERIOR A.C. (UNIVERSIDAD TECNILENIO), protegida por la Ley Federal de Derecho de Autor; la alteración o deformación de una obra, así como su reproducción, exhibición o ejecución pública sin el consentimiento de su autor y titular de los derechos correspondientes es constitutivo de un delito tipificado en la Ley Federal de Derechos de Autor, así como en las Leyes Internacionales de Derecho de Autor.

El uso de imágenes, fragmentos de videos, fragmentos de eventos culturales, programas y demás material que sea objeto de protección de los derechos de autor, es exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por UNIVERSIDAD TECNILENIO.

Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de esta obra sin la autorización previa por escrito de UNIVERSIDAD TECNILENIO. Sin embargo, usted podrá bajar material a su computadora personal para uso exclusivamente personal o educacional y no comercial limitado a una copia por página. No se podrá remover o alterar de la copia ninguna leyenda de Derechos de Autor o la que manifieste la autoría del material.