

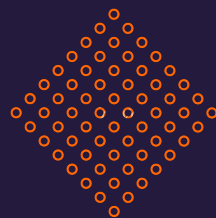
Data science

Bases de Datos II



63%

Adrián Amador
Jose Vargas
Pablo Mora



I semestre - 2023

Canva



¿Qué es Data Science?

- Estudio de datos con el fin de extraer información significativa.
- Es un campo multidisciplinario que combina áreas como la Matemática, Estadística, la Inteligencia Artificial, Machine Learning y la Ingeniería en Computación para analizar grandes cantidades de datos.
- ¿Qué pasó?, ¿Por qué pasó?, ¿Qué pasará? y ¿Qué se puede hacer con los resultados?



Definición

Importancia del Data Science



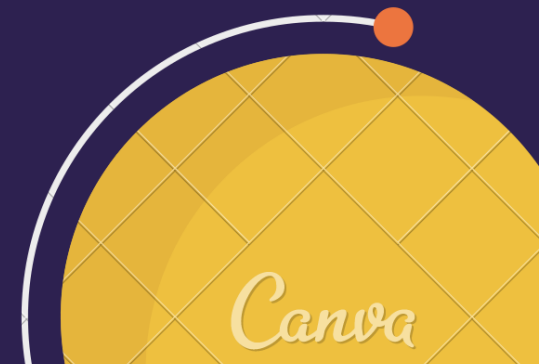
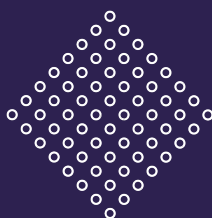
Organizaciones modernas están inundadas en datos



Sistemas en línea y portales de pago capturan datos en los campos del comercio electrónico, medicina, finanzas y cualquier otro aspecto de la vida humana.



Disposición de grandes cantidades de datos no estructurados



Ciclo de Vida del Data Science



Ingesta de datos
(data ingestion)



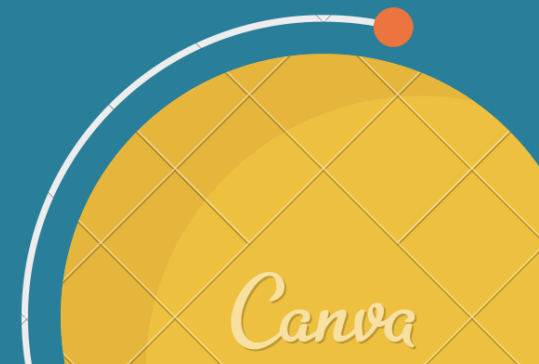
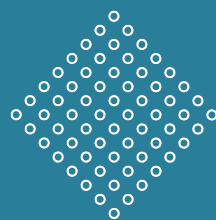
Almacenamiento y
procesamiento de los
datos



Análisis de datos



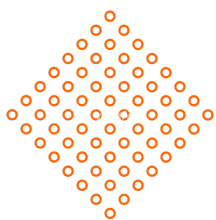
Comunicar



Responsabilidades

Data Scientist

- Conocer lo suficiente acerca del negocio
- Aplicar estadística y ciencias de la computación
- Usar una gran variedad de herramientas y técnicas para preparar y extraer datos
- Extraer conocimientos del Big Data
- Escribir programas que automatizan el procesamiento
- Expresar el significado de los resultados
- Explicar cómo los resultados pueden ser usados
- Colaborar con otros miembros del equipo



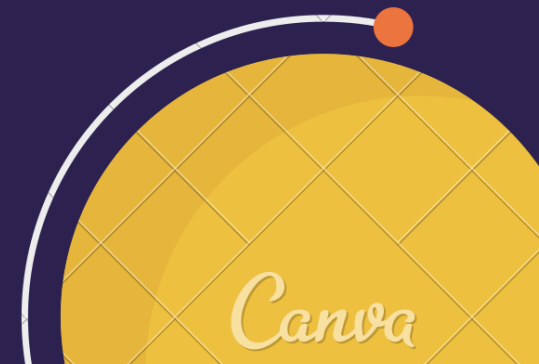
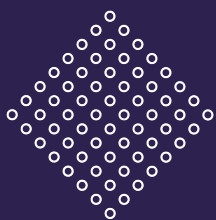


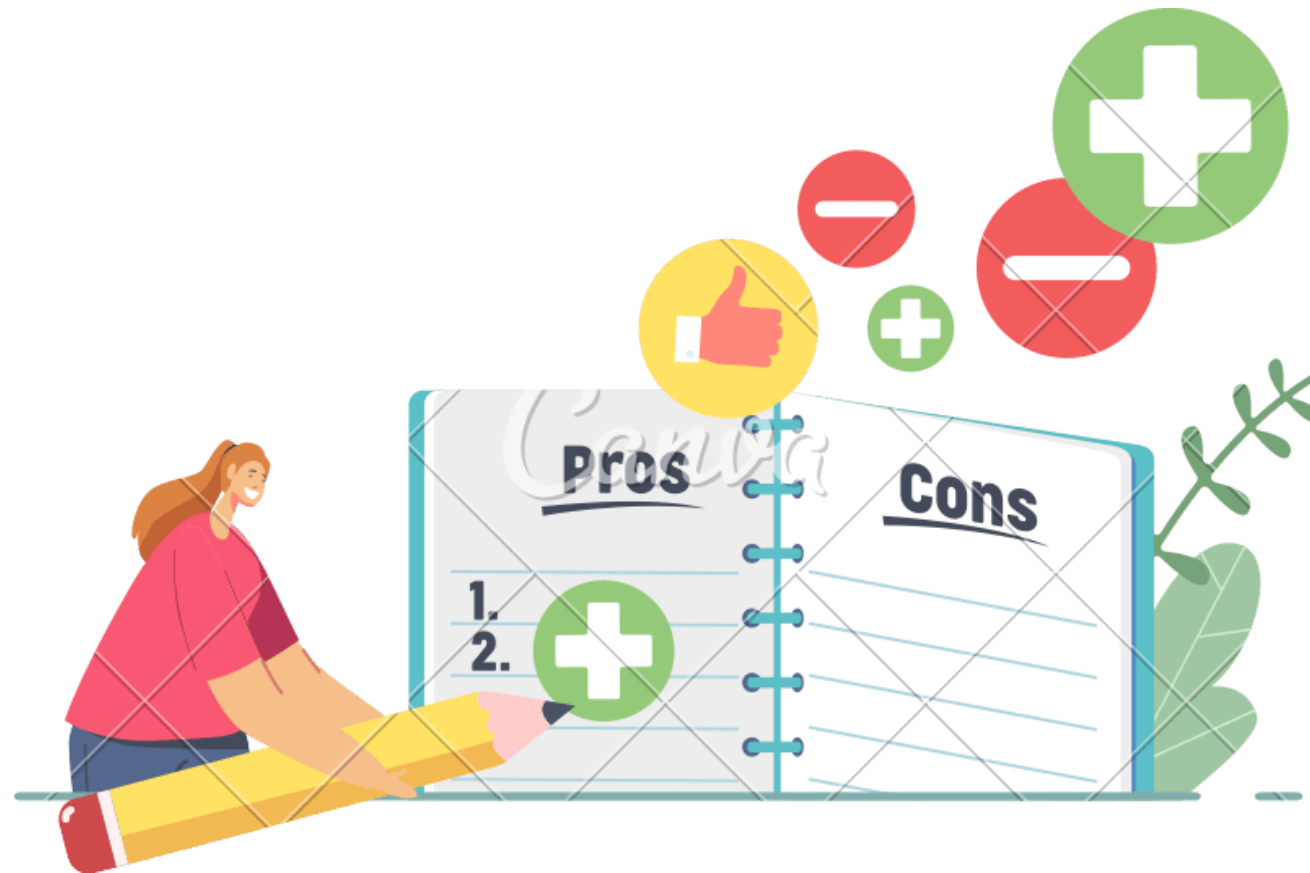
Ejemplos de usos/aplicaciones

Un banco internacional ofrece servicios de préstamos más rápidos con una aplicación móvil que utiliza modelos de riesgo crediticio basados en machine learning y una arquitectura híbrida de computación en la nube que es poderosa y segura.

Una empresa de tecnología de medios digitales creó una plataforma de análisis de audiencia que permite a sus clientes ver qué atrae a las audiencias de televisión a medida que se les ofrece una gama cada vez mayor de canales digitales. La solución emplea análisis profundos y machine learning para recopilar información en tiempo real sobre el comportamiento del espectador.

Un departamento de policía urbana creó herramientas de análisis de incidentes estadísticos para ayudar a los oficiales a comprender cuándo y dónde desplegar recursos para prevenir el crimen. La solución basada en datos crea informes y paneles para aumentar la conciencia situacional de los oficiales de campo.





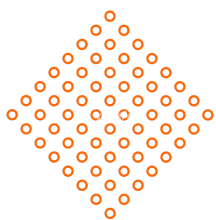
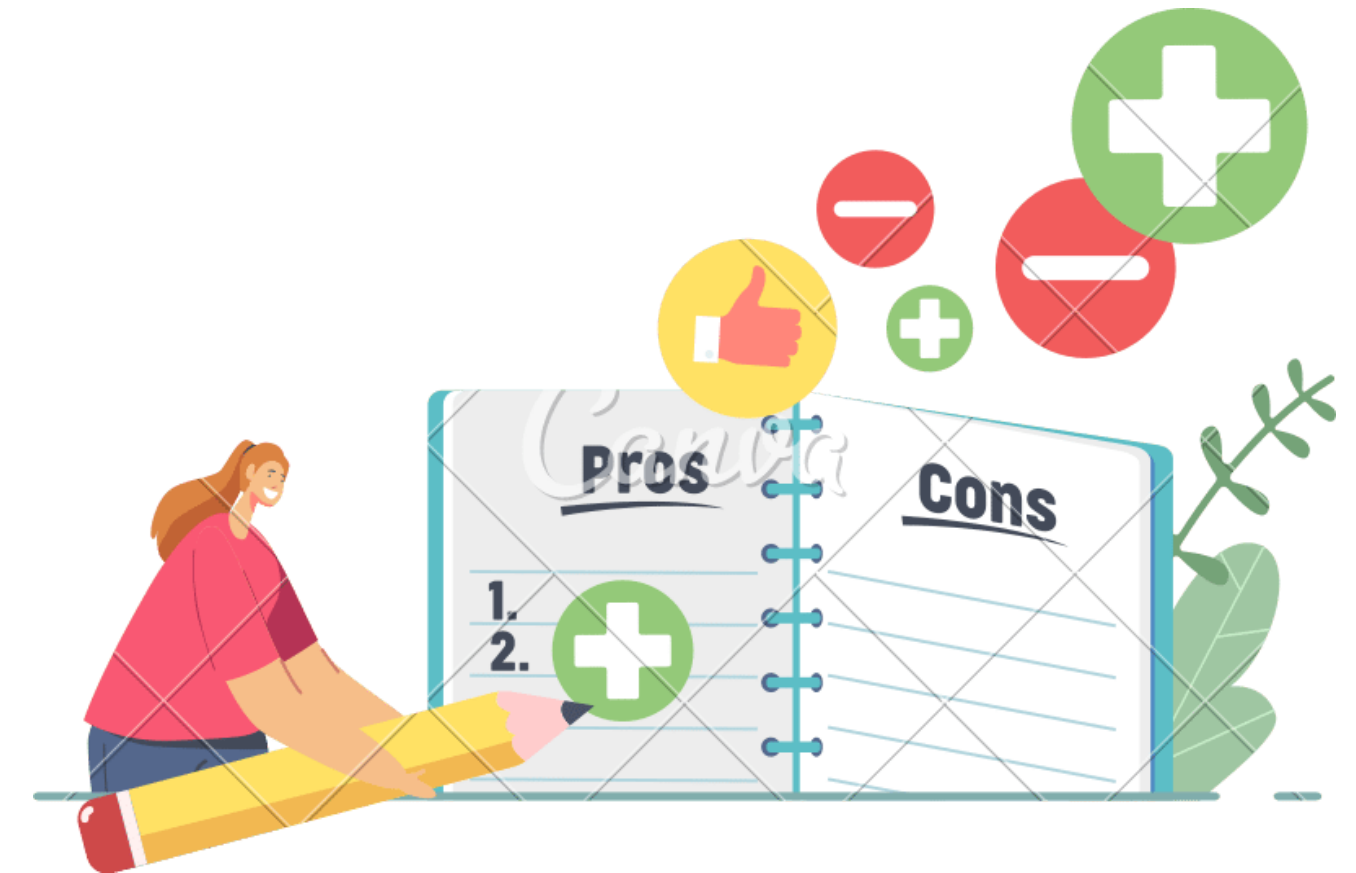
Ventajas

- Descubrir patrones desconocidos de transformación: Permite a las empresas descubrir nuevos patrones y relaciones con el potencial de transformar la organización
- Innovar con nuevos productos y soluciones: Puede revelar lagunas y problemas que de otro modo pasarían desapercibidos
- Optimización en tiempo real: puede ayudar a las empresas a predecir los cambios y reaccionar de forma óptima ante las distintas circunstancias



Desventajas

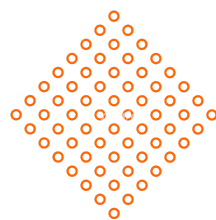
- Se necesita un amplio conocimiento de muchas áreas
- Los datos arbitrarios pueden producir resultados inesperados
- El problema de la privacidad de los datos (datos de los clientes por ejemplo)
- Es un proceso costoso





¡Gracias!

“If you can’t explain it simply, you don’t understand it well enough.”
— Albert Einstein





Referencias

- <https://www.ibm.com/topics/data-science>
- <https://aws.amazon.com/es/what-is/data-science/>
- <https://ischoolonline.berkeley.edu/data-science/what-is-data-science/>
- <https://www.oracle.com/what-is-data-science/>
- Vasant Dhar. 2013. Data science and prediction. Commun. ACM 56, 12 (December 2013), 64–73. <https://doi.org/10.1145/2500499>
- <https://www.imperva.com/learn/application-security/web-scraping-attack/>
- <https://www.devopsschool.com/blog/what-is-data-science-advantages-and-disadvantages-of-data-science/>
- <https://data-flair.training/blogs/pros-and-cons-of-data-science/>

