

CIENCIA DE DATOS

ANÁLISIS DE DATOS

- Por su definición:

El análisis de datos es un proceso que consiste en inspeccionar, limpiar y transformar datos con el objetivo de resaltar información útil, para sugerir conclusiones y apoyo en la toma de decisiones.



CIENCIA DE DATOS

- **Por definición:**

La ciencia de datos es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático, y la analítica predictiva.



CIENTIFICO DE DATOS

Se les define como una mezcla de estadísticos, computólogos y pensadores creativos, con las siguientes habilidades:

- Recopilar, procesar y extraer valor de las diversas y extensas bases de datos.**
- Imaginación para comprender, visualizar y comunicar sus conclusiones a los no científicos de datos.**
- Capacidad para crear soluciones basadas en datos que aumentan los beneficios, reducen los costos.**
- Los científicos de datos trabajan en todas las industrias y hacen frente a los grandes proyectos de datos en todos los niveles.**



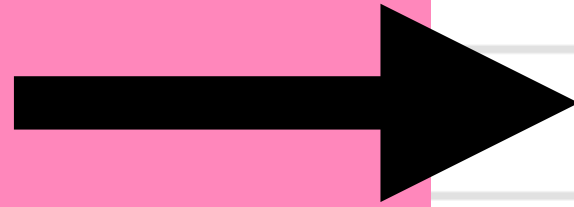
EL PROCESO QUE SIGUE UN CIENTÍFICO DE DATOS:

1. **Extraer datos**, independientemente de la fuente y de su volumen.
2. **Limpiar los datos**, para eliminar lo que pueda sesgar los resultados.
3. **Procesar los datos** usando métodos estadísticos como inferencia estadística, modelos de regresión, pruebas de hipótesis, etc.
4. **Diseñar experimentos** adicionales en caso de ser necesario.
5. **Crear visualizaciones gráficas** de los datos relevantes de la investigación



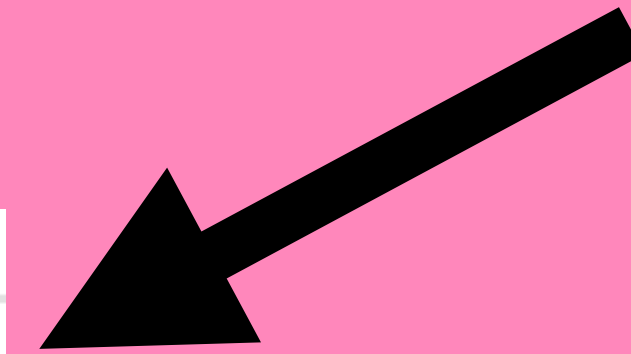
Terminos más usados

Variables

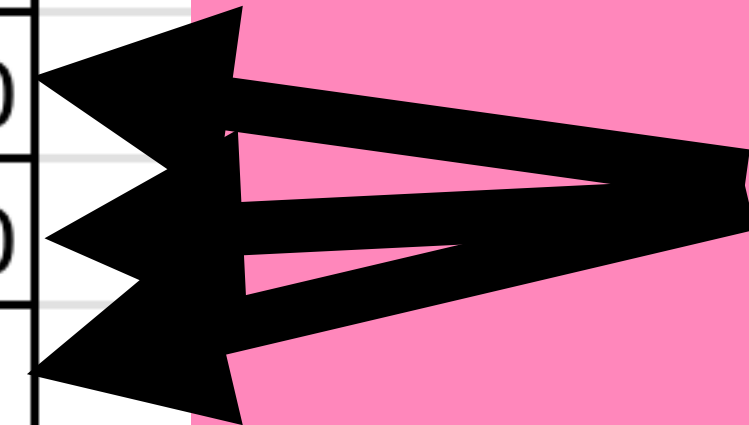


Precios Casas		
M2	#cuartos	Precio
100	1	10000
200	2	20000
300	3	X

Dataset



Datapoints



MODELOS

- En ciencias aplicadas y en tecnología, un modelo matemático es uno de los tipos de modelos científicos que emplea algún tipo de formulismo matemático para expresar relaciones, proposiciones sustantivas de hechos, variables, parámetros, entidades y relaciones entre variables de las operaciones, para estudiar comportamientos de sistemas complejos ante situaciones difíciles de observar en la realidad.

