



LA IMPORTANCIA DE (pensar acerca de) ANÁLISIS DE DATOS

M.Sc. Sergio Morales Esquivel
Universidad Cenfotec

ACERCA DE MÍ:

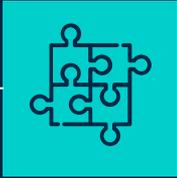
- Profesor Posgrado de Análisis de Datos @**U Cenfotec**
- Global Technology Strategist @**Growth Acceleration Partners**
- Data Scientist, Software Engineer.



ANÁLISIS DE DATOS

- Procesamiento de **Grandes Cantidades de Datos** (Big Data).
- Datos **Estructurados** (tablas, transacciones) y **No Estructurados** (imágenes, videos).
- Hallazgos **no implícitos** en la **Estructura** o **Semántica** original.
- Computación, Estadística, Diseño Experimental, Sociología, Economía, etc.

ANÁLISIS DE DATOS



DESCRIPTIVO

¿Qué está pasando?

Contexto de Datos.

Estadística, Semántica.



PREDICTIVO

¿Qué va a pasar?

Modas y tendencias.

Necesita datos históricos.



PRESCRIPTIVO

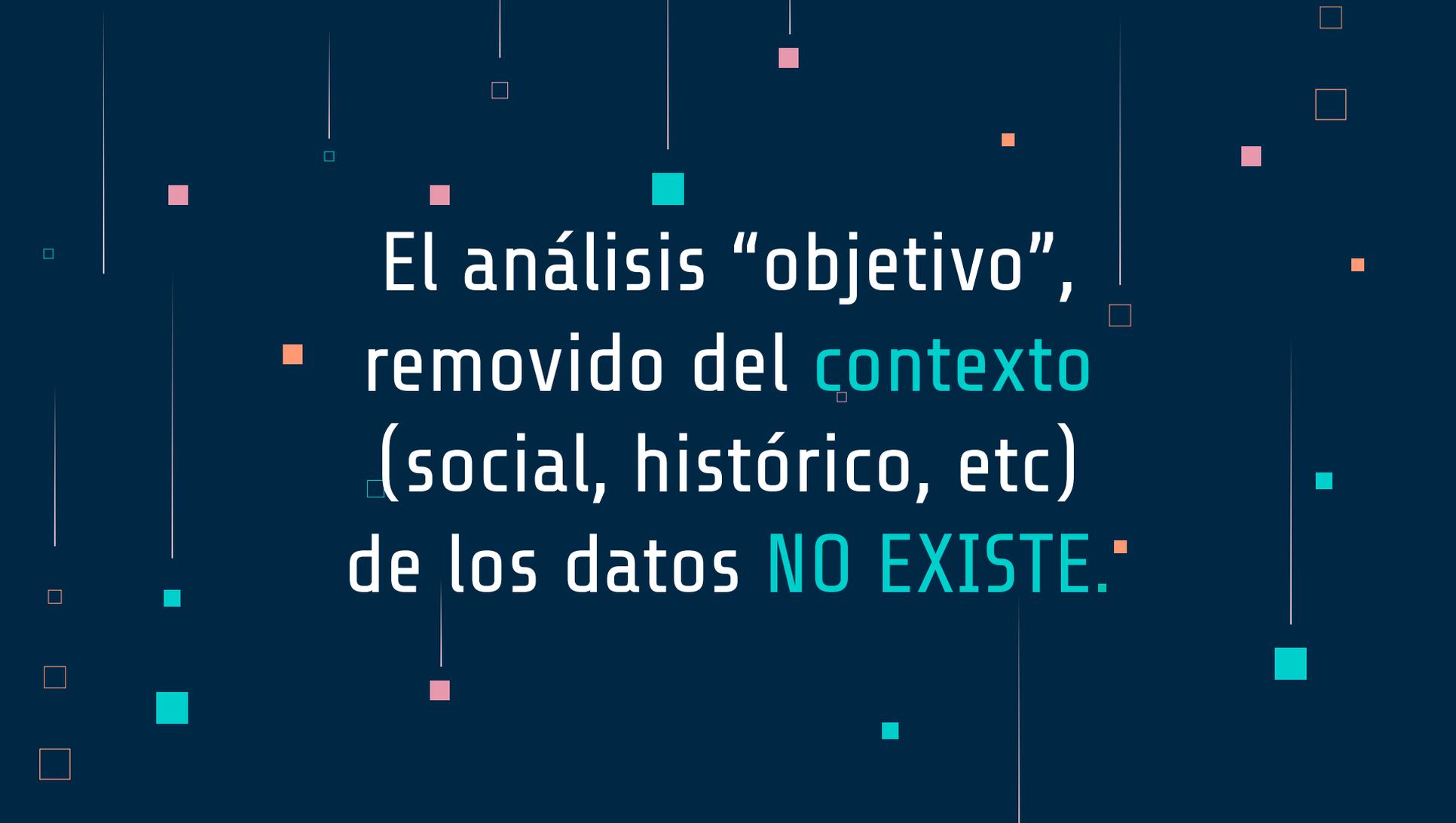
¿Qué debo hacer?

Basado en metás.

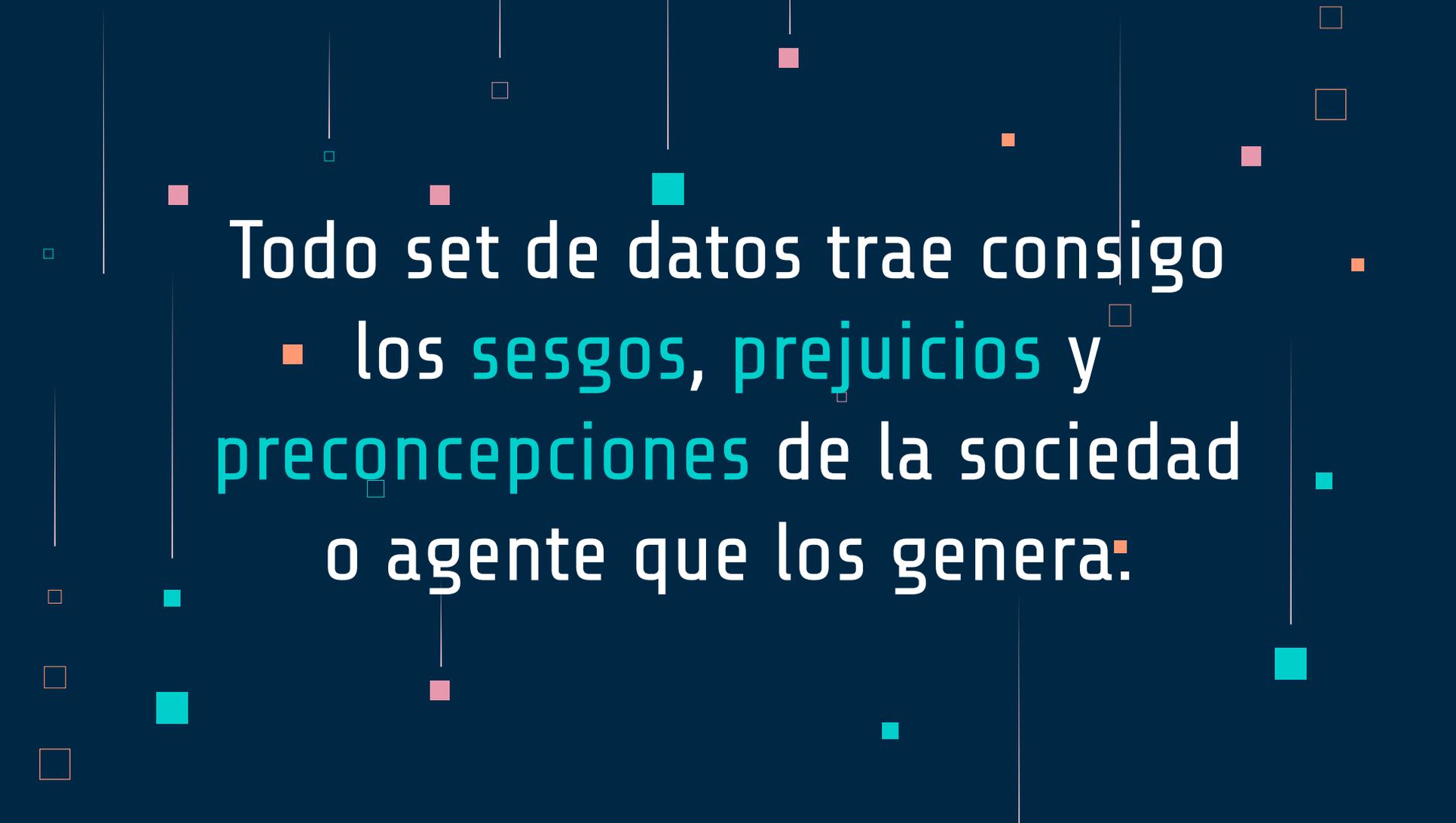
Busca optimizar métricas.

ESTRUCTURA vs SEMÁNTICA

- **Estructura:** **Tipos** de datos (números, caracteres, verdad/falso), **Dimensiones** (observaciones, columnas), **Costo** (almacenamiento, velocidad).
- **Semántica:** El **significado** de los datos (edades, años, regiones, niveles económicos, etc). Su origen **contextual** e **histórico**. ¿De dónde vienen? ¿Qué intereses sirven?



El análisis “objetivo”,
removido del contexto
(social, histórico, etc)
de los datos **NO EXISTE.**

The background is a dark blue gradient. It features several vertical white lines of varying lengths. Scattered throughout are small squares in teal, pink, and orange. Some squares are solid, while others are hollow outlines. The text is centered and consists of a main title and a bulleted point.

Todo set de datos trae consigo

- los sesgos, prejuicios y preconcepciones de la sociedad o agente que los genera.



Runaway Feedback Loops in Predictive Policing

Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, Suresh Venkatasubramanian

Academics Confirm Major Predictive Policing Algorithm is Fundamentally Flawed

Feedback loops can occur when **your model is controlling the next round of data you get**. The data that is returned quickly becomes flawed by the software itself.

“Predictive policing is aptly named: it is predicting future policing, not future crime.” -- Suresh [Venkatasubramanian](#)

PROCESO DE ANÁLISIS

Recolección de
Datos



Análisis
Exploratorio

Limpieza y
Preparación



Modelado/
Reportería

RECOLECCIÓN

Decisiones tomadas al inicio del proyecto **son difíciles de cambiar:**

- ¿Qué recolectar?
- ¿Donde almacenar?
- Controles de Acceso
- Medios de Acceso
- Seguridad, Compliance

EXPLORACIÓN

Análisis Exploratorio:

Añadir semántica a la estructura

- Storytelling
- Verificación de Supuestos/Hipótesis
- Identificación de Sesgos

PREPARACIÓN

Decisiones influenciadas por **hallazgos previos** y **objetivos** del análisis.

- **Limpieza**: Valores Nulos, Duplicados, Outliers, Typos, etc.
- **Transformación**: Normalización, Codificación, Balanceo de clases.

VISUALIZACIÓN Y REPORTERÍA

Ayudan tanto al análisis como a la divulgación.

- Considerar la audiencia.
- Es fácil “mentir” con visualizaciones, aún “por accidente”.
- Aprovechar datos geográficos y temporales.

MODELADO:

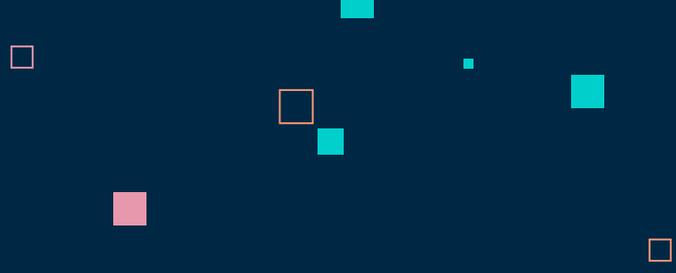
Los “últimos 100 metros” de la carrera.

- Un proceso de modelado perfecto no arregla errores de recolección, análisis o preparación.
- Modelos de **caja blanca** vs. **caja negra**, xAI.
- Puede amplificar sesgos y prejuicios.
- Modelar comportamiento a través de líneas culturales es difícil.

Está bien decidir no hacer un modelo

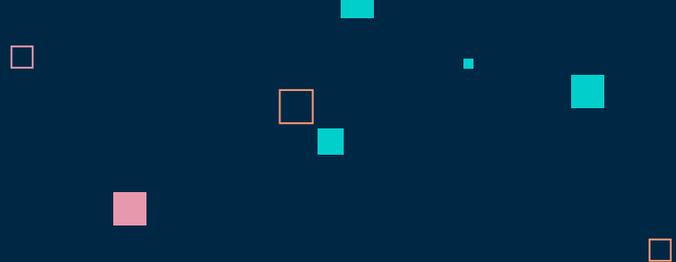
OTROS CONSEJOS



A collection of small, semi-transparent squares in various colors (cyan, pink, orange) scattered in the top right corner of the slide.

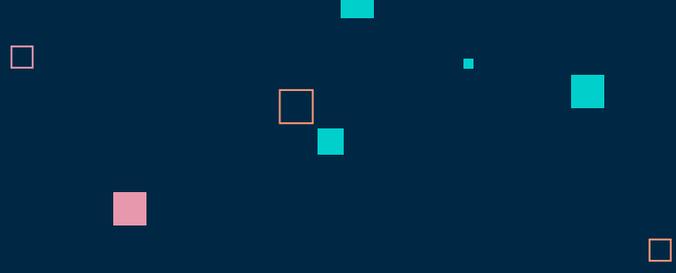
Elegir muy bien **objetivos y métricas** de cualquier proceso de análisis:

- Evitar feedback loops
 - Es difícil cuantificar el bien de una comunidad
- 
- A small cluster of semi-transparent squares in cyan and orange colors located in the bottom left corner of the slide.

A collection of small squares in various colors (cyan, pink, orange) scattered in the top right corner of the slide.

El paradigma “Silicon Valley” es atractivo pero sirve un **conjunto de intereses** particular.

- Evitar el **colonialismo algorítmico**
 - Decidir qué importar de historias “exitosas”
- 
- A small cluster of squares in the bottom left corner, including a cyan square and an orange square.

A collection of small squares in various colors (cyan, pink, orange) scattered in the top right corner of the slide.

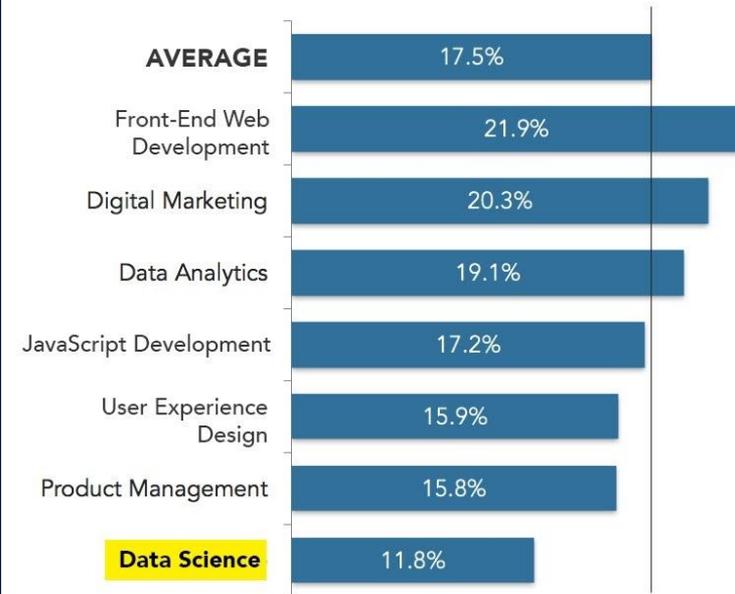
La toma de decisiones “automáticas” hace difícil tener **responsabilidad**.

- Usar modelos transparentes/xAI.
 - Diseñar y habilitar medios formales de “reclamo”.
- 
- A few small squares in the bottom left corner, including a cyan one and an orange one.

Ensamblar y dar voz a un equipo **diverso** en el diseño, implementación y gobernanza de datos.

La diversidad evita sesgos y promueve el pensamiento analítico.

Percentage of Hispanic/Latino and African-American Students Enrolled in Part-Time GA Courses



Source: General Assembly part-time student data (09/2016-01/2017)

*Average = the courses listed above



Datos de Contacto

smorales@ucenfotec.ac.cr
+506 8885 4764
www.fireblend.com

GRACIAS!

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#)
Please keep this slide for attribution

Recursos Recomendados:

Colonialismo Algorítmico:

- <https://reallifemag.com/the-algorithmic-colonization-of-africa/>
- https://docs.google.com/document/d/1wyes_kCE_WlctUtEQw5yuuJplddkSQtuGHBMxsjycGA/

La Política de los Artefactos:

- <https://www.cc.gatech.edu/~beki/cs4001/Winner.pdf>

Toolkit Ético para la Práctica de Ingeniería y Diseño:

- <https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/>