

# ¿Qué es ciencia de datos?

Yess Espinosa

# ¡bienvenidos!

Un poco de mí ...



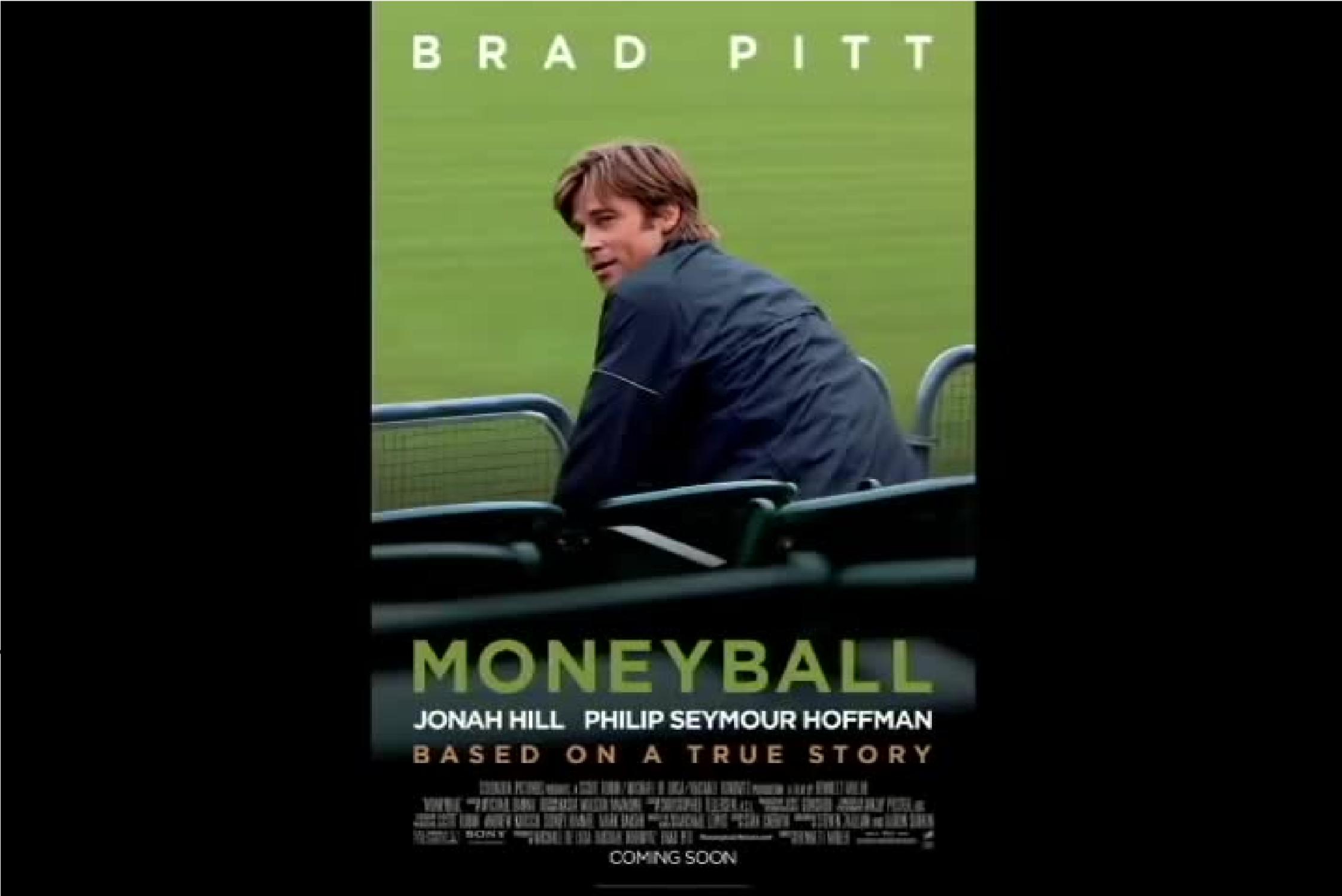
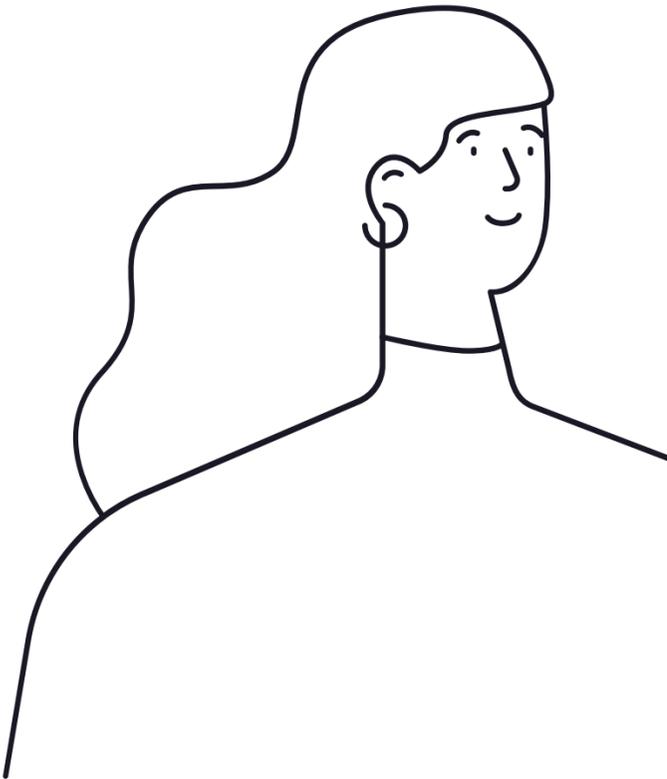
Sección de preguntas y respuestas al final



## ¿Qué es ciencia de datos?

Yess Espinosa

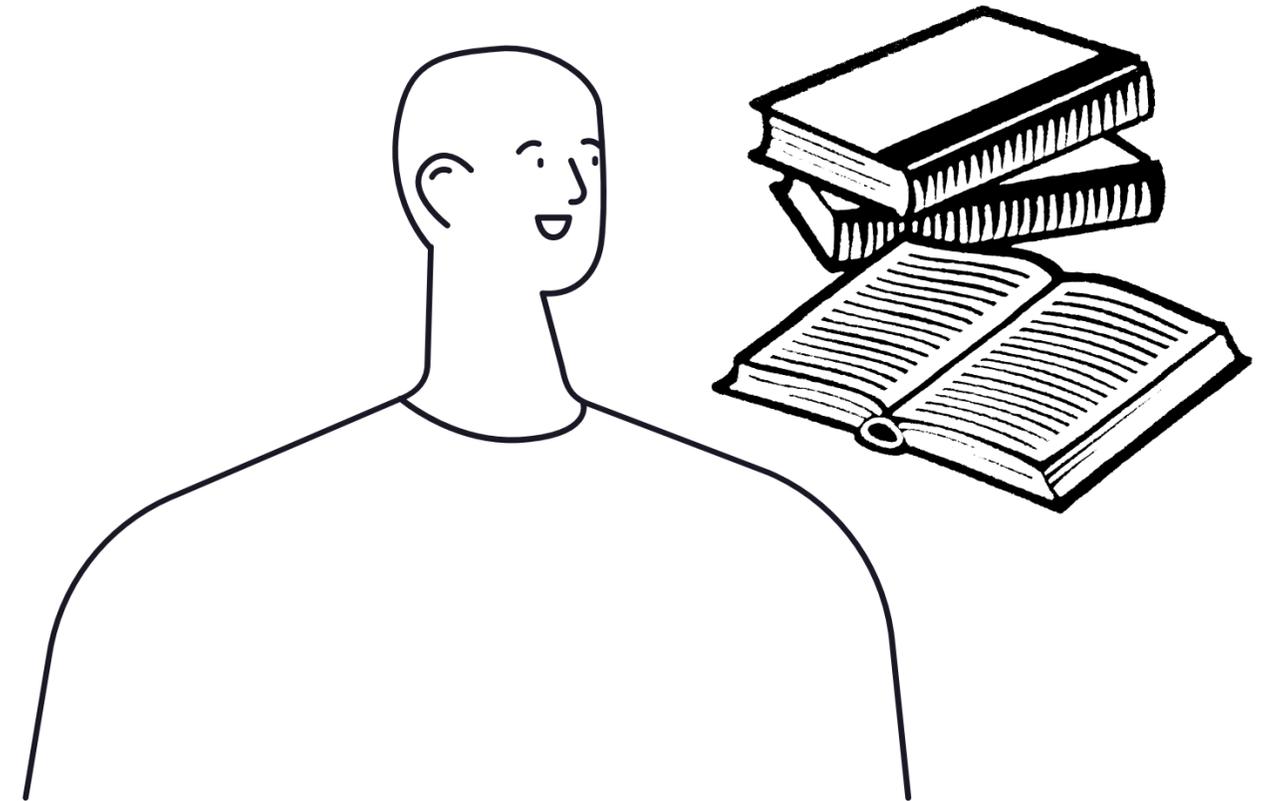
# Presentación de video



CONCEPTOS  
PRINCIPALES

**ESTRUCTURA**

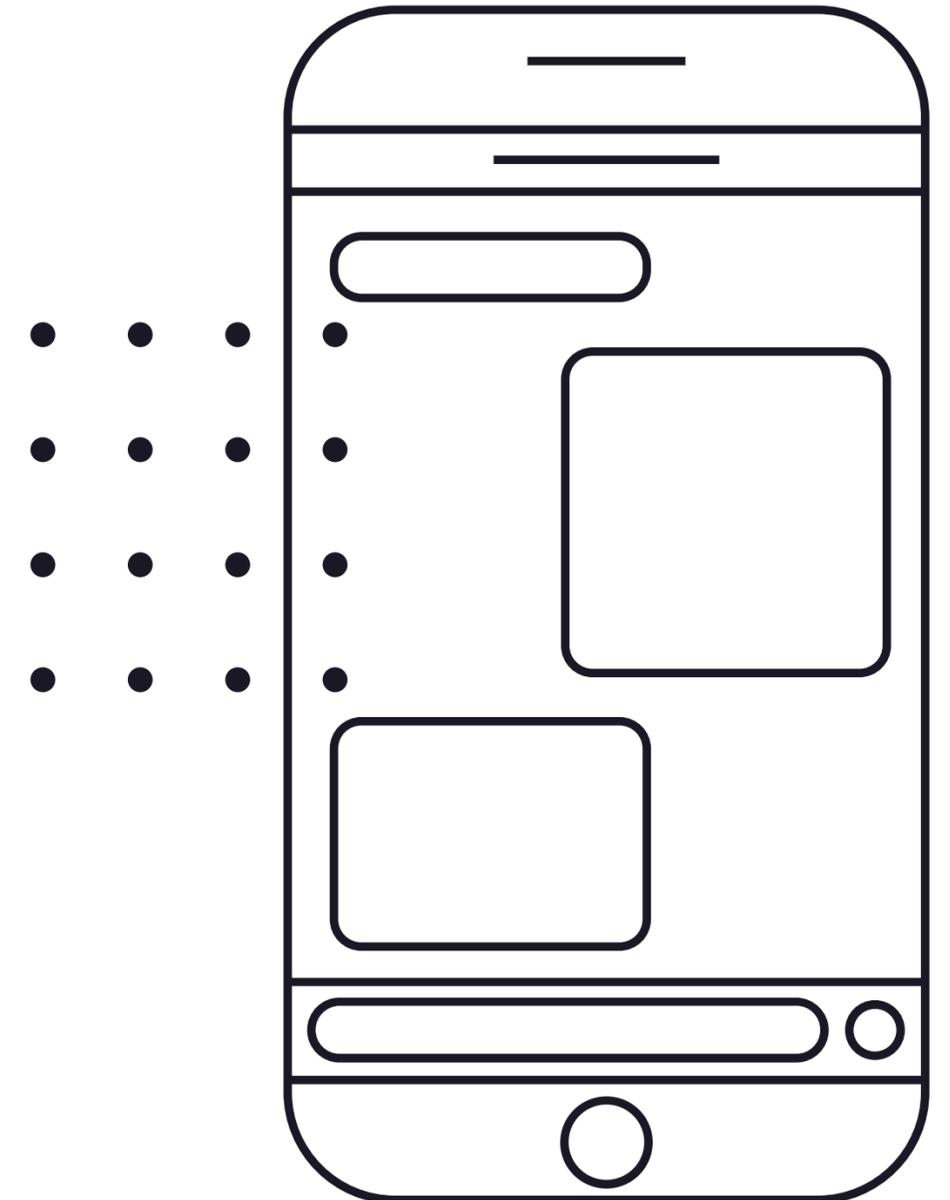
APRENDIZAJE



**DATOS**



# FUENTES CRECIENTES DE DATOS



# FUENTES DE DATOS

- Dispositivos de almacenamiento de datos
- Datos generados por la actividad
- Perfiles de redes sociales
- Bases de datos
- Medios de comunicación
- Nube
- Web
- Internet de las cosas



# BIG DATA

volumen

velocidad

variedad

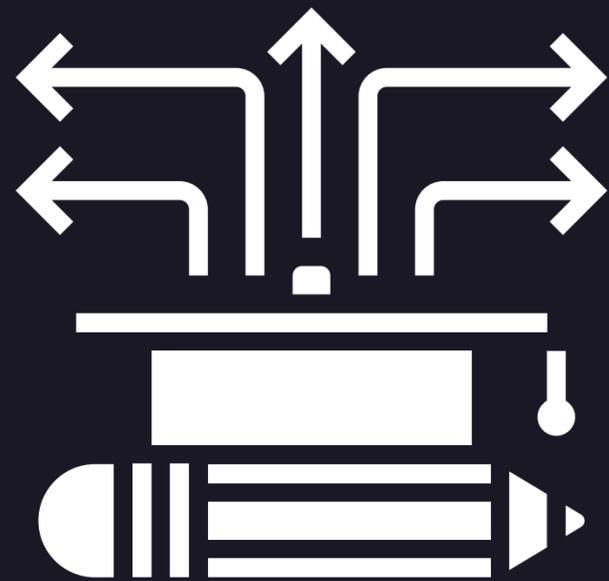


veracidad

valor

# CIENCIA DE DATOS

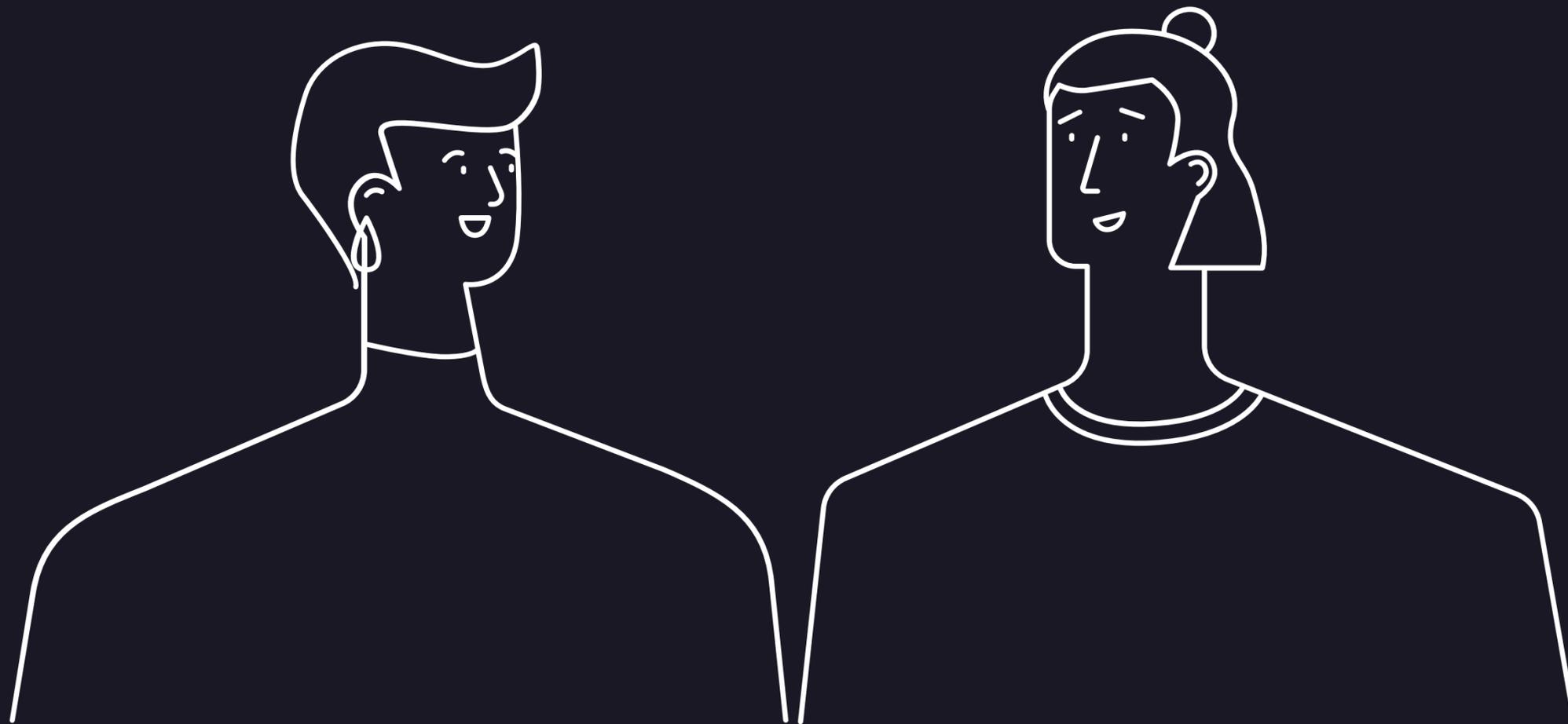
Abstracto



Uso de métodos estadísticos y algoritmos matemáticos apoyado con tecnologías para la toma de decisiones.

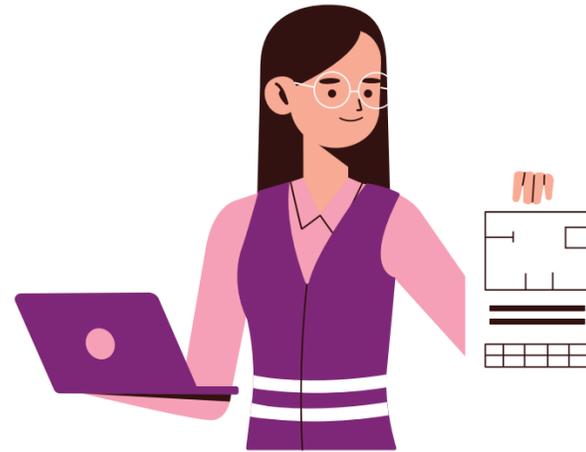
# CIENTÍFICOS DE DATOS

Cómo contar historias con los datos



# DISCIPLINAS

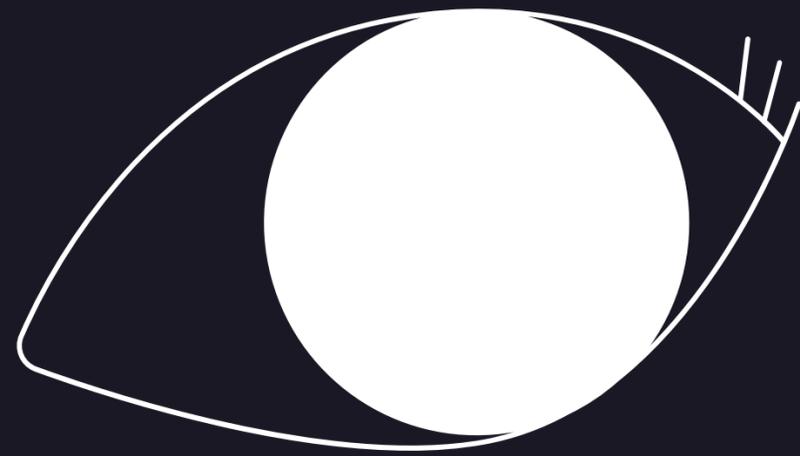
Matemáticas  
Y ESTADÍSTICA



**programación**

**NEGOCIO**

COMUNICACIÓN



**ALGUNOS CASOS DE USO**

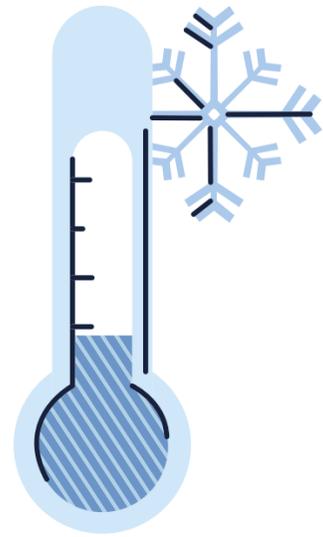
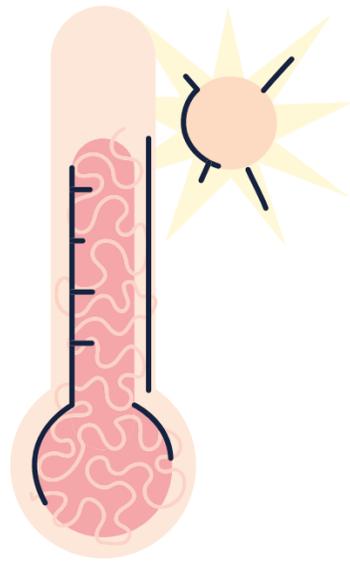
# ANÁLISIS DE NEGOCIO



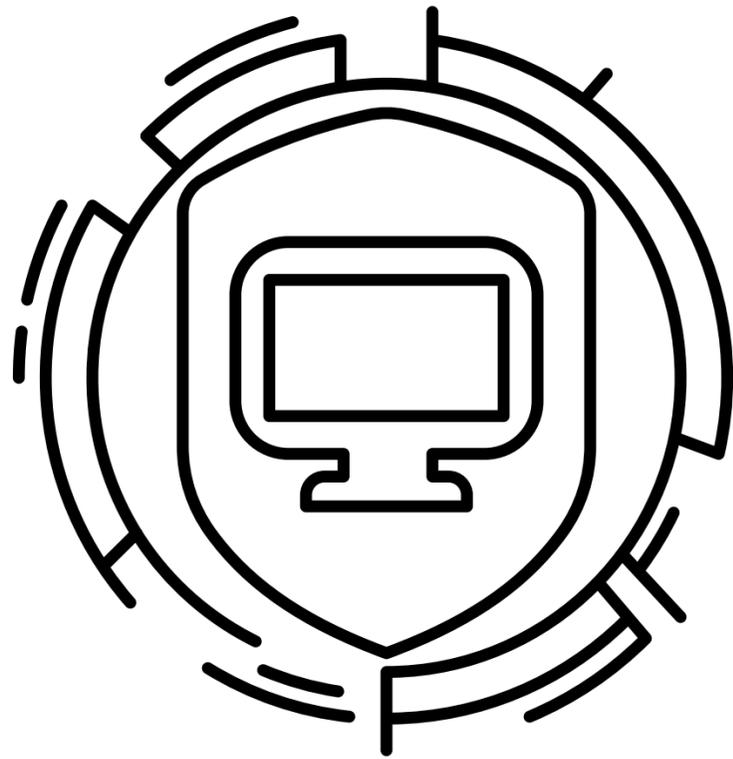
# PREDECIR TENDENCIAS



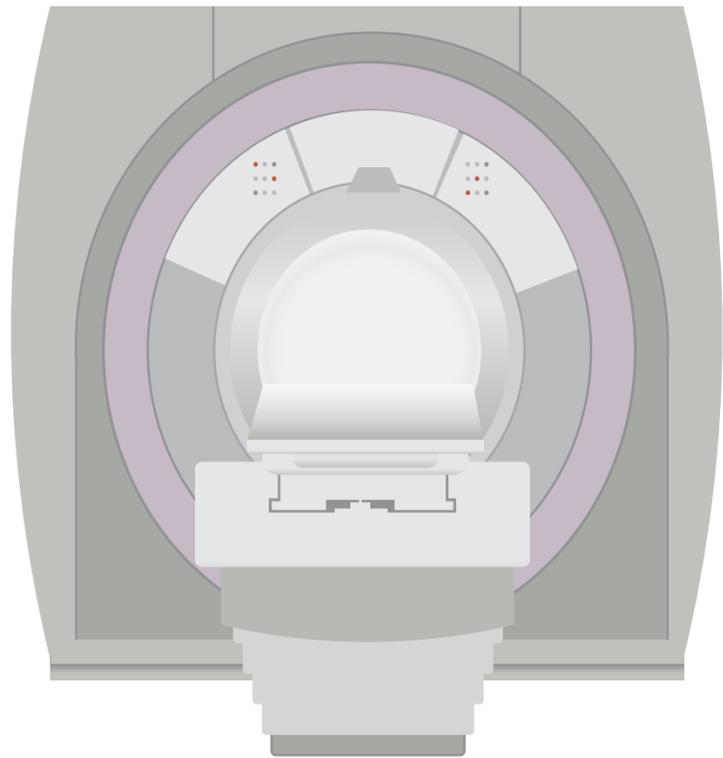
# PREVENIR CATÁSTROFES



# BANCA O FINANZAS



# MEDICINA

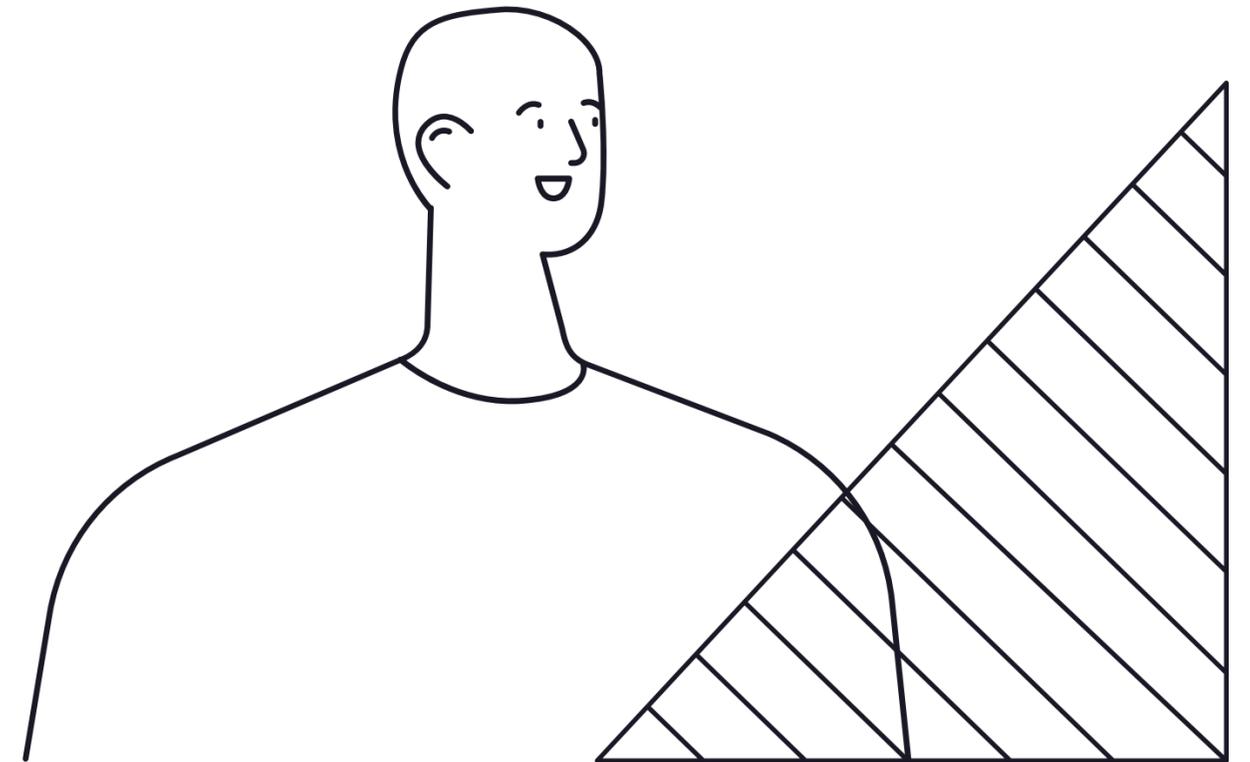


# SEGMENTACIÓN DE CLIENTES





# PRINCIPALES HERRAMIENTAS





# Aprendizaje

Disciplinas

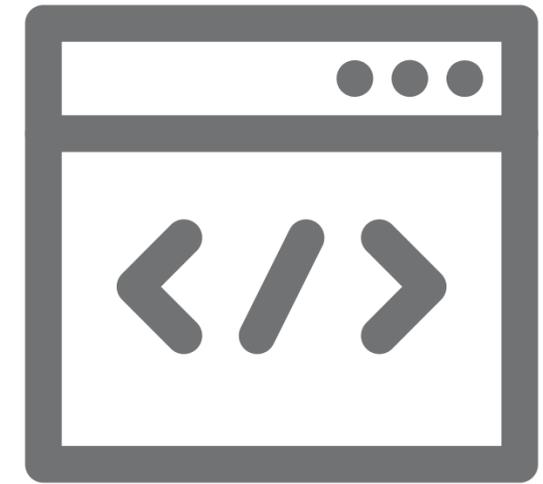
# DISCIPLINAS PRINCIPALES



Probabilidad y  
estadística



Aprendizaje  
automático



Algoritmos

# KIT DE HERRAMIENTAS



*DATOS*

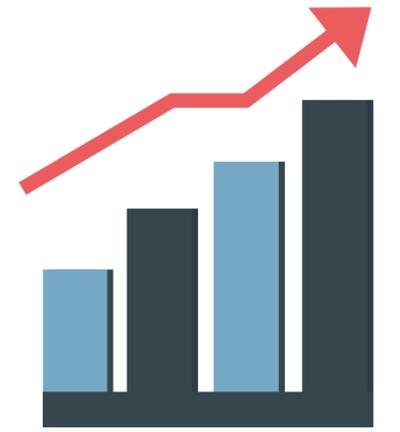
JSON  
CSV  
Excel  
Base de datos  
API

## ***PREPARACIÓN DE DATOS***

Selección de datos  
Subconjuntos  
Ordenamiento  
Agrupación  
Datos faltantes

Filtrado  
Unión de los  
datos  
Reducción de  
dimensionalidad

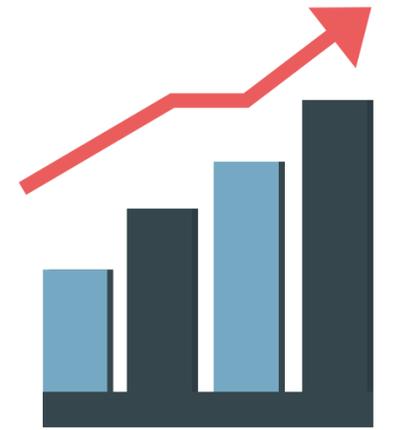
# PROBABILIDAD Y ESTADÍSTICA



## Estadística descriptiva

- Medidas de tendencia central
  - Media
  - Mediana
  - Moda
- Medidas de dispersión
  - Rango, varianza, desviación estándar
  - Rango intercuartílico
  - Covarianza
  - Coeficiente de variación de Pearson
  - Coeficiente de correlación de Spearman
- Medidas de posición no central
  - Cuartiles y percentiles
- Asimetría y curtosis
- Frecuencias
  - Absoluta
  - Absoluta acumulada
  - Relativa
- Gráficos
  - Lineal, de barras, circular,
  - polígono de frecuencias, histograma,
  - pareto, diagramas de caja

# PROBABILIDAD Y ESTADÍSTICA



## Estadística inferencial

- Estimaciones puntuales y de intervalo
- Pruebas de hipótesis
  - P-values
  - z-test, t-test, chi-square, f-test
  
- ANOVA (Análisis de la variable)
- MANOVA (Análisis multivariante de la variable)
- ANCOVA (Análisis de la covarianza)

# APRENDIZAJE AUTOMÁTICO

## Conceptos

- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje reforzado
- Aprendizaje profundo



# APRENDIZAJE AUTOMÁTICO

## Regresión

- Regresión lineal
- Logística

## Árboles de decisión

- ID3
- C4.5
- CART
- Random forest

## Redes Bayesianas

- Naive bayes

## Clustering

- K-means
- Clustering jerárquico

## Máquinas de soporte vectorial

## SVM



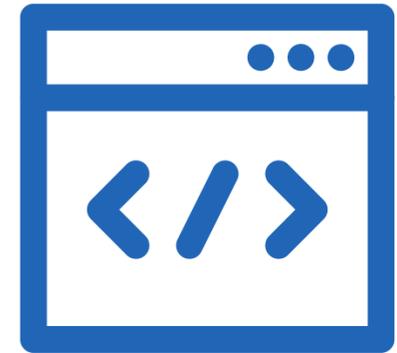
# APRENDIZAJE AUTOMÁTICO

## Desafíos

- Tipos de datos
- Datos faltantes
- Outliers
- Feature selection
- Sobreajuste (overfitting) subajuste (underfitting)
- Conjunto de entrenamiento y prueba
- Reducción de dimensiones



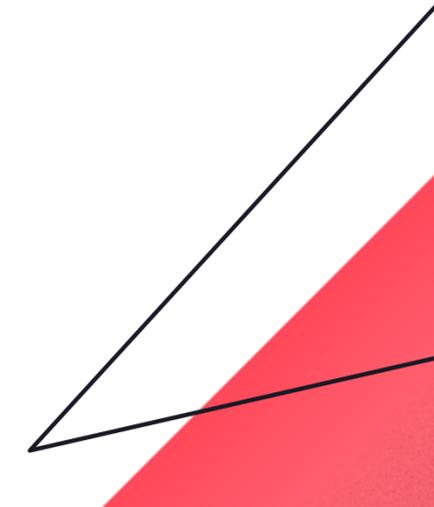
# ALGORITMOS



Referencia: <https://ds-path.netlify.app/sections/ml-path>



**No te rindas, aprende de tus  
errores y nunca dejes de  
aprender.**

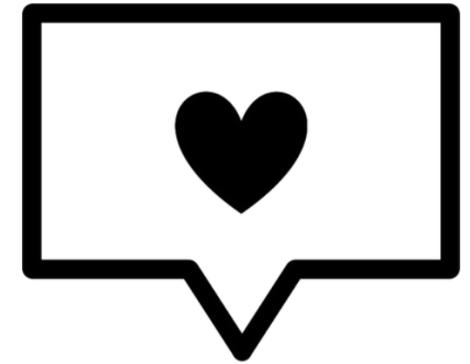




# PREGUNTAS

¿?

# GRACIAS



<https://www.linkedin.com/in/yess-espinoza/>



<https://www.facebook.com/yess.espinoza.mx/>



@yessespinosamx



yessmetal@gmail.com

