

# OLS y Predicción

## Ciencia de Datos y Economía Aplicada

Ignacio Sarmiento-Barbieri

Universidad de los Andes

# Agenda

## 1 ¿Qué es la Ciencia de Datos?

- ¿Que es Big Data?
- ¿Que es Aprendizaje de Máquinas?

## 2 Predicción vs Causalidad

# ¿Qué es la Ciencia de Datos?

**Ciencia de Datos** es el proceso de usar datos para **entender el mundo, predecir comportamientos y tomar decisiones.**

Analytics And Data Science

## **Data Scientist: The Sexiest Job of the 21st Century**

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

Source: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

# ¿Qué es la Ciencia de Datos?

¿Qué implica?

- ▶ Recolección y análisis de "big data"
- ▶ Herramientas de múltiples disciplinas:
  - ▶ Estadística y **econometría** para inferencia y causalidad
  - ▶ **Aprendizaje de Máquinas** (*Machine learning*) para predicción
  - ▶ Programación y bases de datos para manipulación y escalabilidad
  - ▶ Visualización para comunicar hallazgos

# Agenda

## 1 ¿Qué es la Ciencia de Datos?

- ¿Que es Big Data?
- ¿Que es Aprendizaje de Máquinas?

## 2 Predicción vs Causalidad

# ¿Que es Big Data?

- ▶ Big n, es solo parte de la historia
- ▶ Big también es big k, muchos covariates, a veces  $n \ll k$
- ▶ Vamos a entender Big también como datos que no surgen de fuentes tradicionales. Donde enfrentamos las 4 V:
  - ▶ **Volumen:** grandes cantidades de datos
  - ▶ **Variedad:** múltiples tipos y fuentes
  - ▶ **Velocidad:** generación continua
  - ▶ **Veracidad:** calidad y ruido

# Agenda

## 1 ¿Qué es la Ciencia de Datos?

- ¿Que es Big Data?
- ¿Que es Aprendizaje de Máquinas?

## 2 Predicción vs Causalidad

# ¿Qué es Aprendizaje de Máquinas?

El Aprendizaje de Máquinas trata sobre predicción

- ▶ El aprendizaje de máquinas es una rama de la informática y la estadística, dedicada a desarrollar algoritmos para predecir resultados *y* a partir de variables observables  $X$ .
- ▶ Esto se deja como un problema empírico que la computadora puede “aprender”.
- ▶ En general, esto significa que abstraemos del modelo subyacente, el enfoque es pragmático.



# ¿Qué es Aprendizaje de Máquinas?

El Aprendizaje de Máquinas trata sobre predicción

- ▶ El aprendizaje de máquinas es una rama de la informática y la estadística, dedicada a desarrollar algoritmos para predecir resultados *y* a partir de variables observables  $X$ .
- ▶ Esto se deja como un problema empírico que la computadora puede “aprender”.
- ▶ En general, esto significa que abstraemos del modelo subyacente, el enfoque es pragmático.

**“Lo que funciona, funciona...”**

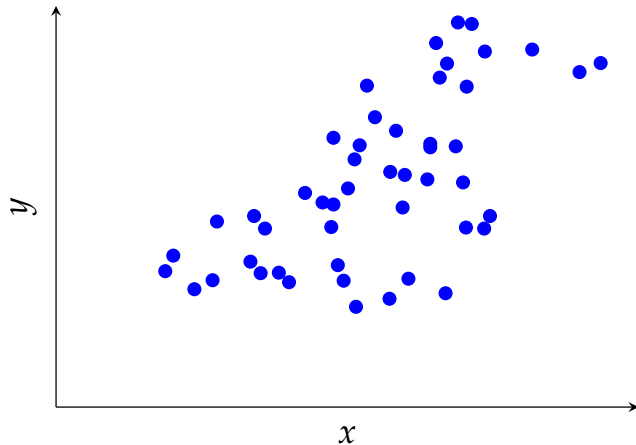
# Agenda

## 1 ¿Qué es la Ciencia de Datos?

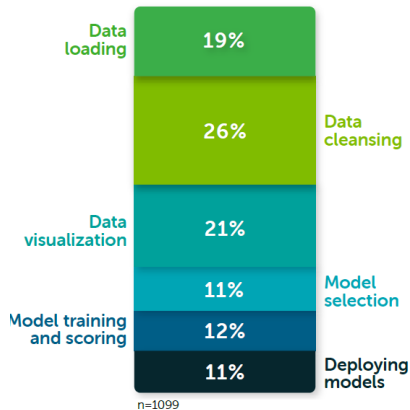
- ¿Que es Big Data?
- ¿Que es Aprendizaje de Máquinas?

## 2 Predicción vs Causalidad

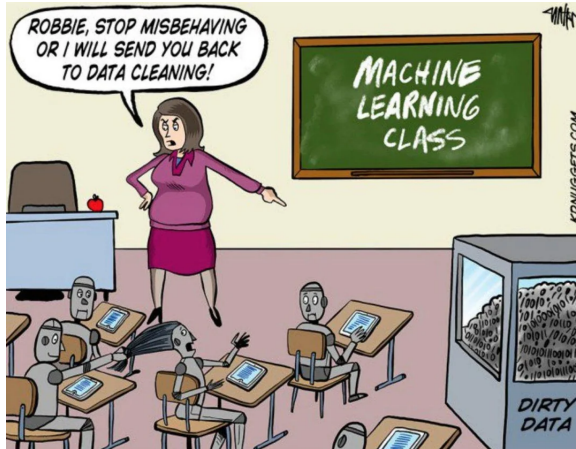
# Predicción vs Causalidad



# Caso de Estudio: Predicción del precio de autos usados



# Caso de Estudio: Predicción del precio de autos usados



# Caso de Estudio: Predicción del precio de autos usados



photo from <https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/>