

# Lecture 1: Introducción

## Big Data and Machine Learning en el Mercado Inmobiliario Educación Continua

Ignacio Sarmiento-Barbieri

Universidad de los Andes

April 25, 2023

# Agenda

## 1 Motivación

- El mercado de viviendas
- Economía Urbana y este curso
- ¿Qué es Big Data? y Machine Learning?

## 2 Organización del Curso

## 3 Para seguir leyendo

## 4 Break

# Motivación: El mercado de viviendas

- ▶ La vivienda es una parte importante de los bienes de los hogares.
- ▶ La elección de este activo como un activo relevante dentro de las finanzas personales se debe a su estabilidad, rendimiento y seguridad
- ▶ Por ejemplo, en los países latinoamericanos, los inversionistas perciben los bienes raíces como un activo más seguro en términos de mayor resiliencia de sus precios con respecto a las condiciones del mercado. Especialmente cuando se pierde confianza en la macroeconomía
- ▶ Los precios de las propiedades en Colombia han aumentado significativamente desde 2005

# Motivación: Precios Propiedades



Fuente: Banco de la República, con información proveniente de las entidades financieras Davivienda, BBVA, AV Villas, Bancolombia, Colmena BCSC y Colpatria.


# Motivación: El mercado de viviendas

- ▶ La vivienda es una parte importante de los bienes de los hogares.
- ▶ La elección de este activo como un activo relevante dentro de las finanzas personales se debe a su estabilidad, rendimiento y seguridad
- ▶ Por ejemplo, en los países latinoamericanos, los inversionistas perciben los bienes raíces como un activo más seguro en términos de mayor resiliencia de sus precios con respecto a las condiciones del mercado. Especialmente cuando se pierde confianza en la macroeconomía
- ▶ Los precios de las propiedades en Colombia han aumentado significativamente desde 2005
- ▶ Cada casa tiene su propio conjunto único de características que afectan el valor y que generan desafíos a la hora de valuarlas.

# Motivación: Precios Propiedades

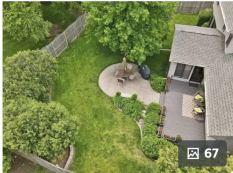
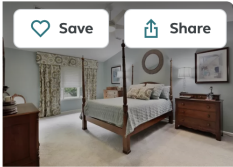
FOR SALE

NEW



Save

Share



67

2701 Worthington Dr

Champaign, IL 61822

4 Beds

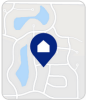
3 Baths

3,340 sqft

\$329,900

Est. Mortgage \$1,765/mo\*

Get Pre-Qualified



Schedule A Tour

Request Info

Tour Type ?

In-Person

Video Chat

Fuente: [www.zillow.com](http://www.zillow.com)

# Economía Urbana y este curso

- ▶ El economista urbano busca formular de una explicación económica rigurosa para entender regularidades observadas en las estructuras espaciales de las ciudades del mundo real.
- ▶ El más obvio entre ellos es heterogeneidad en las disposiciones a pagar por viviendas y la variación espacial en la intensidad del uso del suelo urbano.
- ▶ Nuestro 'swiss army knife' va a ser el modelo de valuación hedónico y el concepto de equilibrio espacial.
- ▶ En este curso la teoría y la evidencia internacional nos va a guiar, y usaremos herramientas Big Data y Machine Learning para entender los precio de las propiedades.

# Big Data and Machine Learning

- ▶ ¿Qué es Big Data (las 3 V's) ?
  - ▶ Volumen (n y k)
  - ▶ Variedad
  - ▶ Velocidad
- ▶ ¿Machine Learning?
  - ▶ Predicción Robusta fuera de muestra
  - ▶  $\neq$  Estadística Clásica (Small Data?, Inferencia)



# Ejemplo Abstracto

$$y_i \approx \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (1)$$

- Estadística/Econometría  $E(\hat{\beta}_j) = \beta_j$

# Ejemplo Abstracto

$$y_i \approx \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} \quad (1)$$

- ▶ Estadística/Econometría  $E(\hat{\beta}_j) = \beta_j$
- ▶ Machine learning:

$$y_{n+1} \approx \beta_1 x_{1n+1} + \beta_2 x_{2n+1} + \dots + \beta_k x_{kn+1} \quad (2)$$

- ▶ hacer que  $\hat{y}_{n+1}$  sea lo mas cercano posible a  $y_{n+1}$

# Ejemplo Concreto

## La primera victoria y derrota del Big Data y Machine Learning

- ▶ Contexto ¿similar? al de hoy: Epidemia de la gripe A en 2009
- ▶ En EEUU la forma de monitorear es a través de reportes de la CDC
- ▶ La CDC agrega a nivel de ciudad, condado, estado, región y a nivel nacional
- ▶ Todo esto llevaba aproximadamente 10 días → demasiado tiempo para una epidemia

# Ejemplo Concreto

Google se ha unido a la conversación

- ▶ Google propuso un mecanismo ingenioso: **Google Flu Trends**
- ▶ Punto de partida:
  - ▶ Proporción de visitas semanales por Gripe A en hospitales
  - ▶ 9 regiones  $\times$  5 años (2003-2007) = 2,340 datos
  - ▶ Estos son los datos que tomaban 10 días en elaborarse (comparemos con la Colombia de 2009)
- ▶ Google cruzó estos datos con las búsquedas sobre la gripe A
- ▶ Con estos datos, construyeron un modelo para predecir intensidad de gripe A

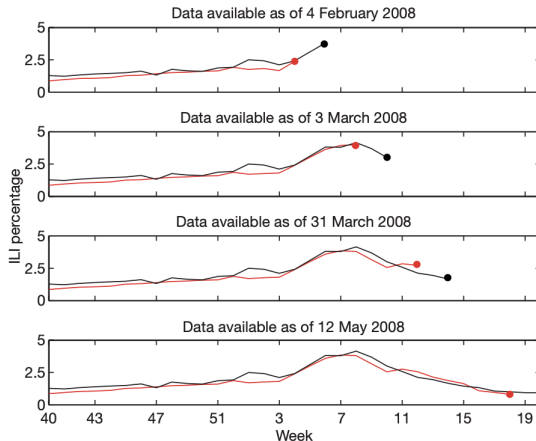
# Ejemplo Concreto

Google se ha unido a la conversación

- ▶ Un solo modelo?
- ▶ Los investigadores de Google estimaron **450 millones** de modelos
- ▶ Eligieron el que mejor predice sobre la intensidad de búsqueda
- ▶ Les permite tener información diaria, semanal o mensual para cualquier punto de EEUU y el mundo
- ▶ A Google le toma 1 día lo que a la CDC 10!

# Ejemplo Concreto

Google se ha unido a la conversación



**Figure 3 | ILI percentages estimated by our model (black) and provided by the CDC (red) in the mid-Atlantic region, showing data available at four points in the 2007-2008 influenza season.** During week 5 we detected a sharply increasing ILI percentage in the mid-Atlantic region; similarly, on 3

# Ejemplo Concreto

El rey ha muerto, larga vida al rey

- ▶ Qué tienen en común Google Flu y Elvis?
  - ▶ Abanderados de la revolución
  - ▶ Definió y redefinió las reglas sistemáticas para hallar la solución a un problema
  - ▶ Éxito rotundo → Publicación en Nature!  
<https://www.nature.com/articles/nature07634>
  - ▶ Pero como a Elvis el éxito fue efímero
  - ▶ La predicciones comenzaron a sobre-estimar
  - ▶ Google Flu esta ahora archivado (disponible al publico)
  - ▶ Continúa recolectando datos pero solo algunas instituciones científicas tienen acceso

# Ejemplo Concreto: El debacle de Zillow



## Zillow's home-buying debacle shows how hard it is to use AI to value real estate

By [Rachel Metz](#), CNN Business

Published 7:32 AM EST, Tue November 9, 2021



# Organización del curso

- ▶ Clases: teoría + práctica en R
- ▶ Primera mitad del curso (5 clases) enfocado en Big Data, obtener datos de distintas fuentes
  - ▶ Manejo de datos espaciales
  - ▶ Obtención y procesamiento automáticos de datos en línea
- ▶ Segunda mitad del curso (5 clases) enfocado en Machine Learning
  - ▶ Modelos de Machine Learning destinados a la predicción.
- ▶ Certificado de participación a los estudiantes que cursen como mínimo el 85% de las sesiones (9/10)

## Para seguir leyendo

- ▶ Glaeser, E. L. (2008). Cities, agglomeration, and spatial equilibrium. Oxford University Press.
- ▶ Lucas, Robert Jr, (1976). Econometric policy evaluation: A critique. Carnegie-Rochester Conference Series on Public Policy, Elsevier, vol. 1(1), pages 19-46, January.
- ▶ O'Sullivan, A. Urban Economics. 8va Edición.
- ▶ Taddy, M. (2019). Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions. McGraw Hill Professional
- ▶ Tom Shaffer The 42 V's of Big Data and Data Science.  
<https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>

Volvemos en 10 min con R