

# Trabajo práctico 2: Estadística Descriptiva

## LABORATORIO DE DATOS

Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires

Verano 2022

Este trabajo práctico debe entregarse en un **notebook** de **R**. Intercale texto, código y gráficos. Utilice el texto para describir lo que usted realiza en las celdas de código, describir los gráficos y mencionar las magnitudes que calcula y las conclusiones que saca. Pienselo como una narración en la que le cuenta a otra persona lo que aprende del dataset. Asegurese de revisar las distintas columnas y entender qué información tiene una fila del dataset.

Tomando los datos de propiedades del dataset Properati (`ar_properties.csv`), construya un dataset más pequeño que retenga las propiedades de *Boedo*, *Colegiales*, *Centro / Microcentro*, *Mataderos* y *Puerto Madero*. ¿Cuántas propiedades hay en cada barrio?

### Tamaño de las propiedades

1. Revise los datos y responda: ¿existen propiedades con datos faltantes en términos de superficie total, superficie cubierta o número de ambientes? Elimine del dataframe todas las observaciones que tengan datos faltantes de superficie total o cubierta o ambientes.
2. Visualice la superficie total y número de habitaciones en cada barrio. *Sugerencia: Construya histogramas para la superficie, y barplots para el número de habitaciones.*
3. Caracterice la superficie de una propiedad típica de este dataset usando promedio, desvío estandar, mediana y cuartiles. Calcule estas magnitudes para cada barrio y presentelas en una tabla que permita compararlas. A partir de esta tabla, responda:
  - a) ¿Cómo es la media comparada a la mediana?
  - b) ¿Entre qué valores varían las superficies?
  - c) ¿Dónde está la propiedad con superficie más grande?
  - d) ¿Dónde varía más el tamaño de las propiedades?
4. Construya un boxplot para la superficie y el número de habitaciones, desagregando por barrio.
5. Explore el tamaño de una habitación promedio en los distintos barrios. Construya esta variable como: `surface_covered / rooms`.

### Relación entre el precio y las características de una propiedad

Retenga únicamente las propiedades cuyo precio esté en dolares.

1. Construya la variable `fondo = surface_total - surface_covered`.
2. Grafique el precio de la propiedad en función de las distintas variables:
  - a) Superficie total.
  - b) Superficie cubierta.
  - c) Fondo.

d) Tamaño promedio de habitación.

Pruebe usar una escala natural y una logarítmica. *¿Qué observa de la relación entre cada variable y el precio? ¿Le permiten identificar distintos comportamientos? ¿Observa alguna tendencia general?*

3. Explore el precio en función del tipo de propiedad. Visualice como se reparten estas propiedades por los distintos barrios, y qué precios tienen.
4. Incorpore la información del tipo de propiedad al gráfico de superficie total vs precio. *¿Considerar el tipo de propiedad permite diferenciar comportamientos en el precio?*

## Conclusiones

Resuma en un párrafo o dos lo que aprendió del dataset estudiado.