

# Data science: definición de la fuente de información

## Integrantes del grupo:

- Juan Martin Carini
- Lucia Hudovsky
- José Saint Germain
- Lucía Buzzeo

## Fuentes de información

### 1. Encuesta-anual-hogares-2019-Alquiler

Este dataset es nuestra primera preferencia para realizar el proyecto final. El dataset contiene datos de una encuesta que se realiza en la Ciudad Autónoma de Buenos Aires para relevar datos demográficos.

En la página de [datos abiertos de Buenos Aires](#) se describe brevemente el dataset así como sus variables. Link al [dataset en formato CSV](#)

- Target:  
Planteamos como potencial Target a **Nivel\_max\_educ**  
Actualmente hay 8 categorías, pero pensamos que podríamos fusionarlas a 5:
  - Sala de 5 terminada
  - Primario terminado
  - Secundario terminado
  - Terciario/universitario
  - No corresponde

### 2. Healthcare Data Stroka attack

Este dataset lo obtuvimos desde [Kaggle](#). El mismo contiene información de personas con diversas características físicas y de salud, teniendo como foco principal la predicción de derrames cerebrales. (Link al [dataset en formato CSV](#))

Como target elegimos la variable “stroke”, para así generar predicciones sobre la aparición de un derrame cerebral.