

## Taller 1. Análisis y procesamiento de datos en R y en Python.

Machine Learning para economistas. 21 de abril de 2022.

Entrega individual.

Fecha límite de entrega: 29 de abril de 2022 11:00pm.

1. Elegir una base de datos que no se haya trabajado en las sesiones de clase, pueden utilizar cualquier base de datos o las que aparecen en los siguientes links:
  - a. <https://github.com/jeffheaton/data>
  - b. <https://www.kaggle.com/datasets>
  - c. <https://archive.ics.uci.edu/ml/datasets.php>
2. Crear un archivo en R y otro en Python para hacer el análisis y procesamiento de la base de datos. Con comentarios.
3. En cada archivo de R y Python, realizar el procesamiento de los datos de la base de datos elegida: Cargar datos, explorar los datos, estadísticas, limpieza de datos, datos faltantes, reemplazo de datos, eliminación de datos y etiquetas, correlaciones, gráficos.
4. En cada archivo de R y Python, realizar la separación de los datos en un conjunto de entrenamiento y otro para test, con alguna de las funciones vistas en clase.
5. En un pdf mostrar los resultados del análisis y procesamiento de los datos, pueden usar datos cualitativos, numéricos, tablas y figuras. El análisis debe incluir:
  - a. Descripción de las variables/etiquetas de la base de datos elegida. Qué representan estos datos.
  - b. Definir por qué hacer un modelo de Machine Learning para estos datos. Objetivo de predicción, grupos, etc.
  - c. Breve análisis de lo que encuentran en los datos: tipo de datos, datos incompletos, datos faltantes.
  - d. Qué cambios le hacen a la base de datos y por qué. Si eliminan una columna, si eliminan datos, si reemplazan datos, etc.
  - e. Mostrar las correlaciones entre las variables que soporten las decisiones del punto anterior.
  - f. Hacer su propio análisis de Machine Learning “humano”. Después de depurar y analizar los datos. Concluir qué podría encontrar un modelo de Machine Learning. Por ejemplo, si es de predicción, hacer alguna conjetura sobre un posible resultado, teniendo en cuenta la base de datos.
  - g. Concluir muy brevemente sobre las diferencias o similitudes entre R y Python.
6. Entregas: PDF, archivo en R, archivo en Python.