PLAYING WITH DATA

short line

# Dataset

Dataset que contiene datos de animales

# Preguntas

1. Cuántos datos tenemos: filas y columnas? Hay datos nulos? En caso afirmativo, en qué porcentaje?
2. El dataset se compone de algunas variables: ¿cuáles son numéricas, categóricas, fechas, ...? En el caso de que sean categóricas, de qué tipo?
3. ¿Existen duplicados?
4. Las variables tiene diferentes valores. Calcula cuántos tienen cada una y cuales son.
5. La variable objetivo es 'OutcomeType': qué valores tiene y en qué probabilidad?
6. Ahora que sabes cuál es el posible futuro de estos animales ... ¿ocurre lo mismo con gatos y perros? Podrías decir que estos datos se deben a la proporción entre el número e perros y gatos?
7. Además de la variable 'OutcomeType' existe otra 'OutcomeSubtype'. Estudia su distribución. Haz lo mismo con la variable 'AgeuponOutcome'.
8. ¿Tiene algo que ver el género del animal con la variable objetivo?
9. ¿Hay algún momento del año en el que se produzcan más adopciones?
10. Toma una muestra de 10000 datos de manera aleatoria. Haz un test de hipótesis sobre la variable 'AgeuponOutcome' para ver si la media corresponde a la media total.
11. ¿Qué otros datos se pueden extraer? Describe al menos 2.

short dash