PLAYING WITH DATA

short line

# Dataset

Dataset que contiene datos de animales

# Preguntas

1. Cuántos datos tenemos:

filas y columnas? Filas: 26729 Rows: 10

Hay datos nulos? YES

En caso afirmativo, en qué porcentaje? Alrededor de un 8% de los datos son nulos.

1. El dataset se compone de algunas variables:

¿cuáles son numéricas, categóricas, fechas, ...?Hay una fecha (DateTime), una numerica (AgeUponOPutcome) y el resto son categoricas.

En el caso de que sean categóricas, de qué tipo? De las categoricas son todas politómica excepto AnimalType que es binaria.

1. ¿Existen duplicados? NO
2. Las variables tiene diferentes valores. Calcula cuántos tienen cada una y cuales son.

AnimalID: Valores unicos

Name: 6375 valores.

DateTime:

OutcomeType:

OutcomeSubtype:

AnimalType:

SexuponOutcome:

AgeuponOutcome:

Breed:

Color:

1. La variable objetivo es 'OutcomeType': qué valores tiene y en qué probabilidad?
2. 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?
3. Además de la variable 'OutcomeType' existe otra 'OutcomeSubtype'. Estudia su distribución. Haz lo mismo con la variable 'AgeuponOutcome'.
4. ¿Tiene algo que ver el género del animal con la variable objetivo?
5. ¿Hay algún momento del año en el que se produzcan más adopciones?
6. 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.
7. ¿Qué otros datos se pueden extraer? Describe al menos 2.

short dash