PLAYING WITH DATA

short line

# Dataset

Dataset que contiene datos de animales

# Preguntas

1. Cuántos datos tenemos: filas y columnas? 26729, 10 Hay datos nulos? si En caso afirmativo, en qué porcentaje? 7,98%
2. El dataset se compone de algunas variables: ¿cuáles son numéricas, categóricas, fechas, ...? En el caso de que sean categóricas, de qué tipo?

Numéricas: AgeuponOutcome

Categoricas: OutcomeType OutcomeSubtype AnimalType SexuponOutcome Breed Color

Cat. Descriptiva: AnimalID Name

Fechas: DateTime

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

len de AnimalID: 26729

len de Name: 6375

len de DateTime: 22918

len de OutcomeType: 5

['Return\_to\_owner' 'Euthanasia' 'Adoption' 'Transfer' 'Died']

len de OutcomeSubtype: 17

[nan 'Suffering' 'Foster' 'Partner' 'Offsite' 'SCRP' 'Aggressive'

'Behavior' 'Rabies Risk' 'Medical' 'In Kennel' 'In Foster' 'Barn'

'Court/Investigation' 'Enroute' 'At Vet' 'In Surgery']

len de AnimalType: 2

['Dog' 'Cat']

len de SexuponOutcome: 6

['Neutered Male' 'Spayed Female' 'Intact Male' 'Intact Female' 'Unknown'

nan]

len de AgeuponOutcome: 45

['1 year' '2 years' '3 weeks' '1 month' '5 months' '4 years' '3 months'

'2 weeks' '2 months' '10 months' '6 months' '5 years' '7 years' '3 years'

'4 months' '12 years' '9 years' '6 years' '1 weeks' '11 years' '4 weeks'

'7 months' '8 years' '11 months' '4 days' '9 months' '8 months'

'15 years' '10 years' '1 week' '0 years' '14 years' '3 days' '6 days'

'5 days' '5 weeks' '2 days' '16 years' '1 day' '13 years' nan '17 years'

'18 years' '19 years' '20 years']

len de Breed: 1380

len de Color: 366

1. La variable objetivo es 'OutcomeType': qué valores tiene y en qué probabilidad?

Adoption 0.402896

Transfer 0.352501

Return\_to\_owner 0.179056

Euthanasia 0.058177

Died 0.007370

1. 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 de perros y gatos?

Gatos:

Transfer 0.494431

Adoption 0.383690

Euthanasia 0.063769

Return\_to\_owner 0.044907

Died 0.013203

Perros:

Adoption 0.416608

Return\_to\_owner 0.274832

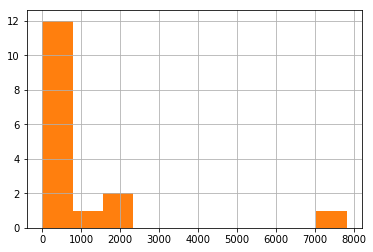
Transfer 0.251170

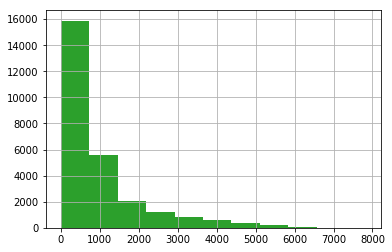
Euthanasia 0.054184

Died 0.003206

Los perros se devuelven mas a sus dueños y los gatos son mas transferidos. Los gatos mueren mas.

1. Además de la variable 'OutcomeType' existe otra 'OutcomeSubtype'. Estudia su distribución. Haz lo mismo con la variable 'AgeuponOutcome'.





1. ¿Tiene algo que ver el género del animal con la variable objetivo?

Male:

Adoption 0.404390

Transfer 0.321783

Return\_to\_owner 0.204750

Euthanasia 0.061711

Died 0.007366

Females:

Adoption 0.437029

Transfer 0.339470

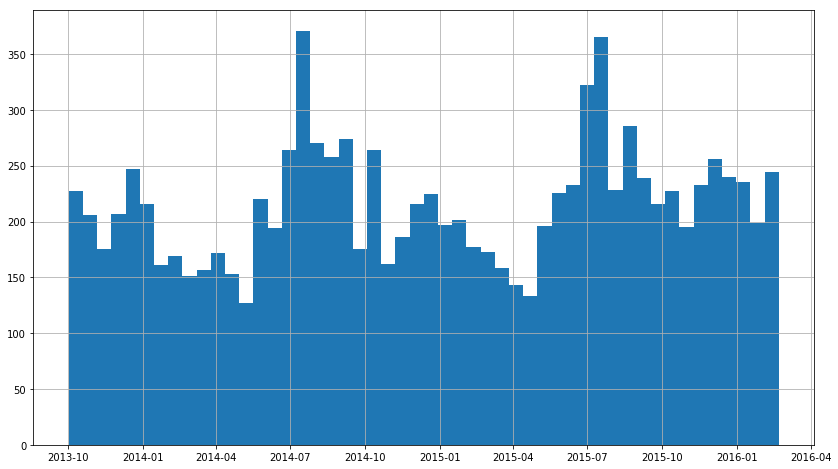
Return\_to\_owner 0.166167

Euthanasia 0.051334

Died 0.006001

Poca diferencia, se devuelven al dueño mas machos y se adoptan algo mas las hembras.

1. ¿Hay algún momento del año en el que se produzcan más adopciones?



En verano hay mas adopciones.

1. 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.

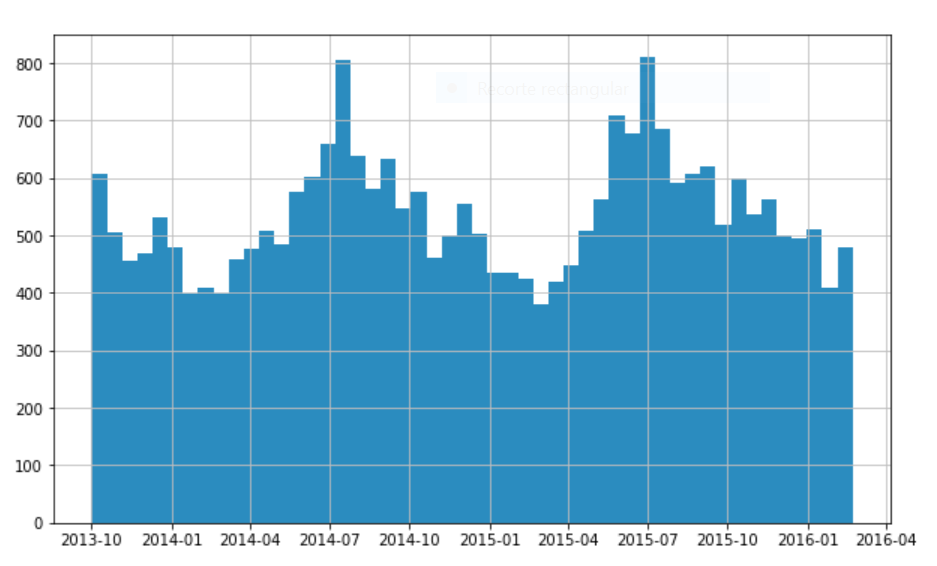
H0: medias son iguales

H1: medias son diferentes

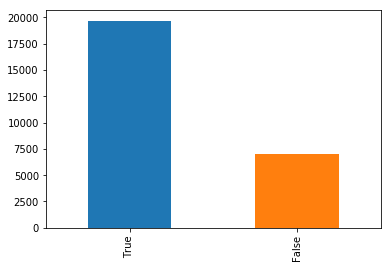
No podemos descartar H0, porque p-value =0.7338056018576609

1. ¿Qué otros datos se pueden extraer? Describe al menos 2.

Hay mas animales recuperados en verano, la gente abandona mas a sus animales en vacaciones.



Hay mas animales castrados que no castrados, por tanto provienen de dueños que los pierden



short dash