Análisis univariado

El dataset no presenta valores faltantes.

En relación al valor “cero”, son valores posibles para las variables indicando la ausencia del atributo que se está midiendo.

**Variables continuas:**

Edad:

Se presenta una amplia variabilidad en las edades registradas. El rango de edades va desde un mínimo de 0 meses hasta un máximo de 255 meses, lo que indica que se incluyen mascotas de todas las edades.

La mediana de la edad es de 3 meses, lo que sugiere que la mitad de las mascotas en el conjunto de datos son menores de 3 meses, mientras que la otra mitad son mayores de 3 meses.

El rango intercuartil (RIQ) de 10 meses indica que el 50% intermedio de las edades se encuentra en un rango de 2 meses a 12 meses. La desviación estándar de 18.16 indica una dispersión significativa en las edades.

Se presentan edades extremas en este dataset.

Cantidad: número de perfiles de las mascotas

Se evidencia que la mayoría de los valores se concentran en torno a un número bajo, ya que tanto la mediana como los cuartiles (Q1 y Q3) tienen un valor de 1. Además, el rango intercuartil (RIQ) es igual a 0, lo que indica que el 50% intermedio de los datos está en el mismo valor (1).

Sin embargo, la variable muestra tambien una amplia dispersión en los datos. Esto sugiere que, aunque la mayoría de los perfiles de mascotas tienen una cantidad baja, existen algunos valores atípicos o extremos que contribuyen a esta dispersión. El coeficiente de variación (CV) es menor que 1, lo que indica una variabilidad relativamente baja en relación con la media.

Tarifa de adopción

Es evidente que una gran proporción de las adopciones se ofrecen de forma gratuita, ya que el valor mínimo, los percentiles 5, 25, 50 y 75, todos tienen un valor de 0. Esto indica que muchas mascotas se ofrecen sin costo alguno para los adoptantes.

Sin embargo, también se observa una amplia variación en los datos, ya que el rango es de 3000, lo que significa que la tarifa máxima de adopción es 3000 veces mayor que la tarifa mínima. Esto se refleja en la alta desviación estándar y el coeficiente de variación (CV) mayor a 1, lo que indica una dispersión significativa en los datos en relación con la media.

Por lo tanto, aunque la mayoría de las adopciones son gratuitas, existen casos excepcionales donde se cobra una tarifa sustancial por la adopción.

Se podría categorizar antes de hacer el bivariado con la target

0: gratuito (84.5%)

1: tiene costo

Cantidad de videos

La variable muestra que la gran mayoría de los perfiles de mascotas no incluyen videos, pero también señala una pequeña cantidad de casos excepcionales con un pequeño número de videos. La distribución de datos es altamente sesgada hacia la falta de videos.

Se podría categorizar antes de hacer el bivariado con la target

0: No presenta video (96.2%)

1: Presenta video

Cantidad de fotos

La variable tiene un rango que va desde un mínimo de 0 (indicando que no hay fotos) hasta un máximo de 30 (indicando un valor extremo de fotos). Esto refleja la variedad en la cantidad de fotos que se incluyen en los perfiles de mascotas.

Los percentiles y cuartiles muestran que la mayoría de los perfiles de mascotas tienen al menos algunas fotos, ya que el percentil 5, el primer cuartil (Q1) y la mediana son mayores que 0. El tercer cuartil (Q3) indica que el 75% de los perfiles tienen 5 fotos o menos, y el percentil 95 indica que la gran mayoría de los perfiles tienen 10 fotos o menos.

La distribución de datos tiende a concentrarse en valores bajos, pero permite un número variable de fotos en algunos perfiles.

**Vbles Categóricas**

Raza1: predominio de la raza 307 (39,5%) y 206 (24.2%). En menor medida la raza 265 (8.4%).

El resto de las razas tienen frecuencias menores al 2.5%.

Agrupamos?

307

206

265

resto

Raza2: predominio de la raza 0 (71.8%). En menor medida la raza 307 (11.5%).

El resto de las razas tienen frecuencias menores al 4%.

Agrupamos?

0

307

resto

Color1: predominio colo1 1 (49.5%). En menor medida colo1 2 (25%). El resto menor al 6.5%

Agrupamos?

1

2

Resto

Color2. Predominio color2 0 (29.8%), en segundo lugar color2 7 (22.9%) y tercer lugar color2 2 (22.1%)

Agrupamos?

0

7

2

Resto

Color3: predomina el color 0 (70.7%), color3 7 (21.5%). Resto inferior 2.5%

Agrupamos?

0

3

resto