

Ciencia de datos con peludos

Análisis estadístico variado e implementación de pequeño modelo de machine learning sobre nuestros amigos peludos



Desde pequeño siempre me han gustado los perros, así que decidí utilizar las habilidades que adquirí sobre ciencia de datos en los últimos meses sobre ellos.

Estuve buscando datasets sobre perros que me llamaron la atención a lo que quería hacer hasta que encontré uno que me pareció muy interesante llamado [Best in Show: The Ultimate Data Dog](#) un dataset el cual contiene múltiples datos acerca del tamaño, entrenabilidad, costo de manutención, etc acerca de cada raza al final me quede con este y empecé a explorarlo. Me di cuenta que este tenía varias filas y columnas con información extra que explicaban más a detalle otras filas y columnas pero esto causaba problemas al momento de leer el dataset con pandas además de ello el dataset estaba en formato xlsx y dividido en varias hojas de excel por lo que se hacia aun mas difícil para pandas manejarlo.

Pensé en yo mismo limpiarlo y transformarlo a csv pero decidí buscarlo en kaggle para ver si alguien lo había transformado ya y ahorrarme un poco de trabajo y si precisamente alguien ya lo había hecho así que descargue este [nuevo dataset](#) en formato csv.

Pero para mi sorpresa aunque ya estaba en formato csv aún tenía varios problemas los cuales solucione con varias transformaciones que especificare en el pdf técnico.

Luego de tener el dataset listo, empecé haciendo un pequeño análisis exploratorio con la librería de pandas profiling y aunque me ayudó a ver en más detalle todas las variables que tenía el dataset también fue un poco confuso ver tantas variables y las relaciones entre ellas.

Después de esto empecé el análisis por mi mano y me hice varias preguntas:

Preguntas

Cual es el top 5 de perros más populares en USA?

Aqui encontre que el top 5 lo componen:

5.Yorkshire Terrier



4.Golden Retriever



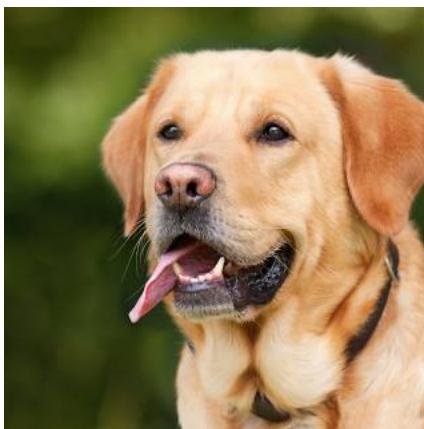
3.Beagle



2.German Shepherd



1.Labrador Retriever



El primer lugar se lo llevó el Labrador Retriever!

¿Cuál es el promedio de vida de los perros entre las razas?

¡Encontré que el promedio de vida está en aproximadamente **11 años!**

Al ver la distribución de los datos de longevidad se me ocurrió otra pregunta.

¿Hay alguna correlación entre el tamaño y longevidad de los perros?

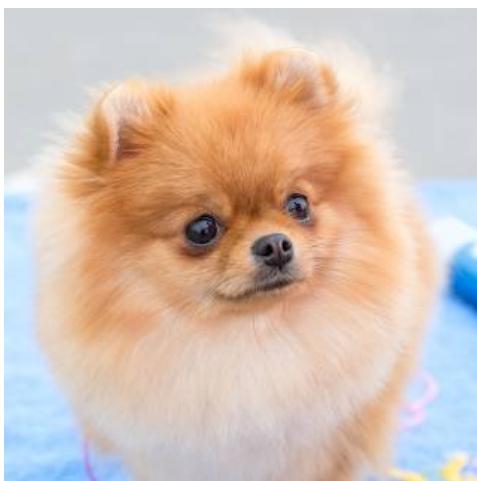
Al final encontré que aunque es leve si se logra ver una pequeña correlación negativa en cuanto al tamaño/peso de un perro con su longevidad, podría decir que es más probable que viva más un perro pequeño a uno grande.

¿Cuales son los perros más baratos de mantener?

10.Greyhound (Gasto promedio \$15.819 dólares)



9.Pomeranian (Gasto promedio \$15.792 dólares)



8. Boxer (Gasto promedio \$15.746 dólares)



7.Leonberger (Gasto promedio \$10.948 dólares)



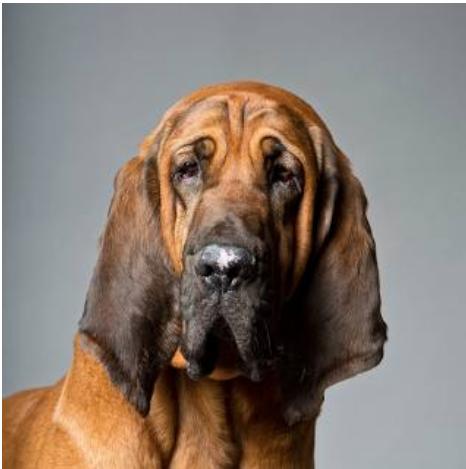
6.Great Dane (Gasto promedio \$14.662 dólares)



5.Bullmastiff (Gasto promedio \$13.936 dólares)



4. Bloodhound (Gasto promedio \$13.824 dólares)



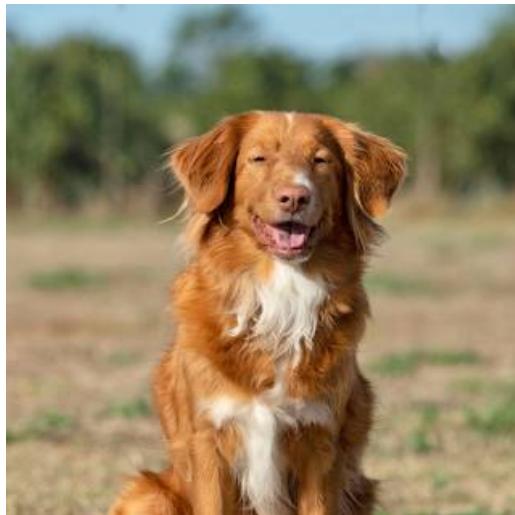
3. Mastiff (Gasto promedio \$13.581 dólares)



2. Bulldog (Gasto promedio \$13.479 dólares)



1.Nova Scotia Duck Tolling Retriever (Gasto promedio \$12.653 dólares)



Gracias a esta pregunta también pude contestar a su contraparte.

¿Cuáles son los perros más caros de mantener?

Estos son el top 10 de los perros mas caros de mantener, los ordenare en orden ascendente:

10.Afghan Hound (Gasto promedio \$24.077 dolares)



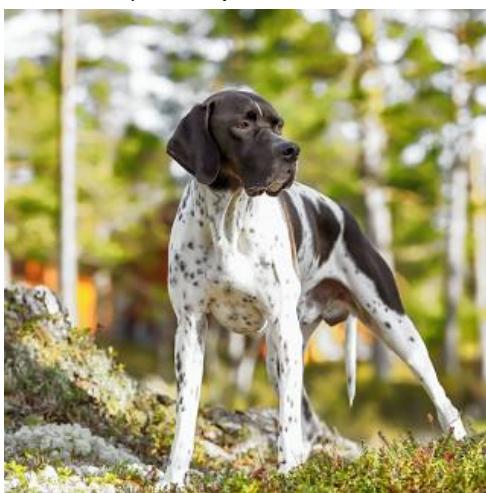
9.Norfolk Terrier (Gasto promedio \$24.308 dolares)



8.Cocker Spaniel (Gasto promedio \$24.330 dolares)



7. Pointer (Gasto promedio \$24.445 dólares)



6. Saluki (Gasto promedio \$24.866 dólares)



5. Samoyed (Gasto promedio \$25.352 dólares)



4. Tibetan Spaniel (Gasto promedio \$25.549 dólares)



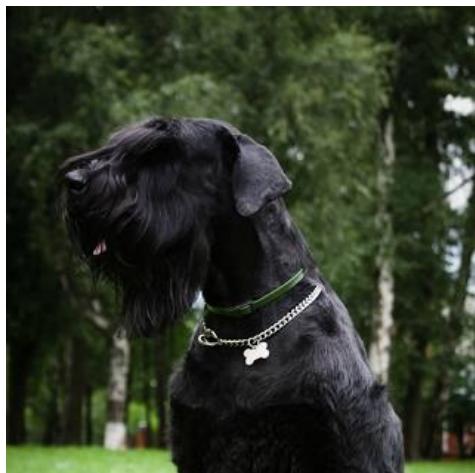
3. German Shorthaired Pointer (Gasto promedio \$25.842 dólares)



2. Chihuahua (Gasto promedio \$26.250 dólares)



1. Giant Schnauzer (Gasto promedio \$26.686 dólares)



Al final el chihuahua y el Giant Schnauzer se alzan como los perros con más gastos! Pero a qué se debe esto? Acaso es la percepción que se tiene de los chihuahuas y de los Giant Schnauzer como perros de lujo y debido a esto la gente gasta más en ellos? ¿O acaso estos tienen más problemas de salud con lo que sus gastos en medicina son superiores?

Es algo bastante peculiar ver que polos opuestos se llevan la corona un perro super pequeño y uno super grande.

Una pequeña curiosidad

Que encontré fue que en los perros de gran tamaño el gasto de comida es mucho mas elevado, me imagino que por su gran peso necesitan comer mas tiene sentido, los 3 perros mas grandes y que comen mas son:

1.Irish Wolfhound



2.Saint Bernard



3.Giant Schnauzer



Luego decidí ver la distribución de posibles enfermedades genéticas según el tamaño de un perro.

Descubrí que entre las posibles enfermedades que puede desarrollar un perro el 40% está en las razas de tamaño grande y el otro 60% está dividido entre las razas medianas y pequeñas equitativamente.

Lo que me lleva a mas preguntas las razas pequeñas y medianas tienen una probabilidad mayor de desarrollar mas enfermedades genéticas?

La correlacion es demasiado minima para poder afirmar algo.

Tambien descubri que entre los perros que mas posibles enfermedades genéticas pueden desarrollar están:

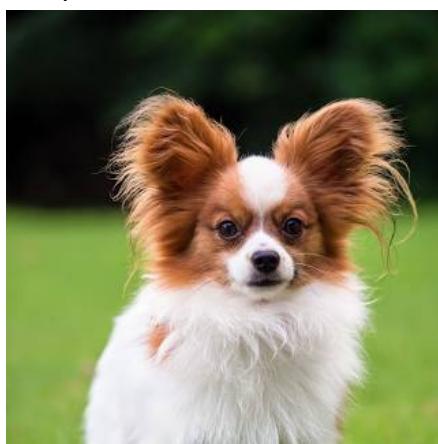
1. Pembroke Welsh Corgi



2. German Shepherd



3. Papillon



Con estos descubrimientos no estoy seguro si el tamaño tenga alguna correlación con las posibles enfermedades creo que debería hacer una análisis más profundo o conseguir más datos.

Tambien quise ver si el tamaño está correlacionado con otra variable y se me ocurrió la siguiente pregunta!

El tamaño de un perro influye en su inteligencia/entrenabilidad?

Al final no vi una correlación evidente entre el tamaño y la entrenabilidad en verdad, hay tanto perros grandes, medianos y pequeños que son faciles o difíciles de entrenar, podría decir que tal vez los perros extremadamente grandes son más tontos o más difíciles de entrenar que perros con tamaños medianos o pequeños. Y que los perros medianos dominan un poco la entrenabilidad por encima del promedio.

Como no encontre nada concluyente me hice una ultima pregunta.

El tamaño de un perro influye en su compatibilidad con los niños?

Encontre que los perros con mas afinidad hacia los niños estan dentro del rango de 1 a 22 kg, y a mas peso desde ahí parece haber una correlación negativa.

Con esta última pregunta termine mi análisis y empiezo con la parte de machine learning!

Modelo de Deep Learning Emociones Peludas

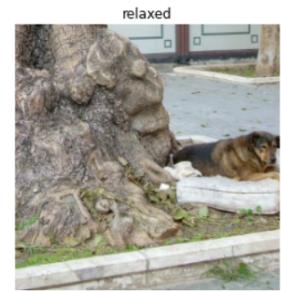
Para la parte de machine learning intente hacer algo o darme una idea con el mismo dataset con el que había hecho el análisis pero no se me vino a la mente una idea clara de que podía hacer con machine learning y los datos que tenía, por lo que empecé a buscar otro dataset y encontre uno que me llamó mucho la atención.

El [Dog Emotions Prediction](#) me pareció un dataset que encajaba perfecto con mi temática, por lo que me puse manos a la obra.

Utilice Tensor Flow/Keras para hacer el modelo de clasificación entre las 4 categorías del dataset:

- 1- happy
- 2- sad
- 3- angry
- 4- relaxed

Explicaré el funcionamiento de este más a detalle en el PDF técnico, al final obtuve un modelo el cual puede clasificar perros entre estas 4 emociones con un 98% de precisión.



Con esto doy por terminado mi proyecto y quiero expresar mis más grande gratitud al equipo de codigofacilito por organizar tan bien este programa sobre temas que son bastante densos, lo único que hubiera cambiado sería tal vez dividir el bootcamp en 2 ya que creo que aunque el machine learning se alimenta de datos, si encontré una gran división entre todo el análisis de datos y ya la parte de machine learning la cual se sintió como la punta del iceberg. De resto quiero felicitarlos por su gran trabajo.

Johan Sebastian Castillo Mazo