

# Dathaton santuario nueva amazonas

---

----- Este proyecto tiene fines educativos -----

El Santuario Nueva Amazonas ha registrado un alarmante incremento de una enfermedad viral letal que afecta a diversas especies. Ante la urgencia de contener su propagación, se nos ha encomendado la tarea de analizar minuciosamente los datos clínicos para identificar patrones en los síntomas. Con esta información, el equipo médico podrá diseñar estrategias precisas para erradicar el virus y, así, preservar la rica diversidad biológica que habita.

Nuestra hipótesis

Los animales que presentan una combinación específica de síntomas tienen una mayor probabilidad de ser clasificados como peligrosos, lo que indica una posible correlación entre ciertos síntomas y la gravedad de la enfermedad viral.

## PREGUNTAS DISPARADORAS ?

**¿Qué proporción de casos son considerados peligrosos vs no peligrosos?**

**Cuáles son los síntomas más frecuentes?**

**¿Cuáles son las especies que más sufren estos sintomas?**

**¿Cuáles son las especies que mas sufren el sintoma principal?**

**¿Cuáles son las especies que más han fallecido?**

**¿Cuáles son el conjunto de síntomas presentados por animales fallecidos?**

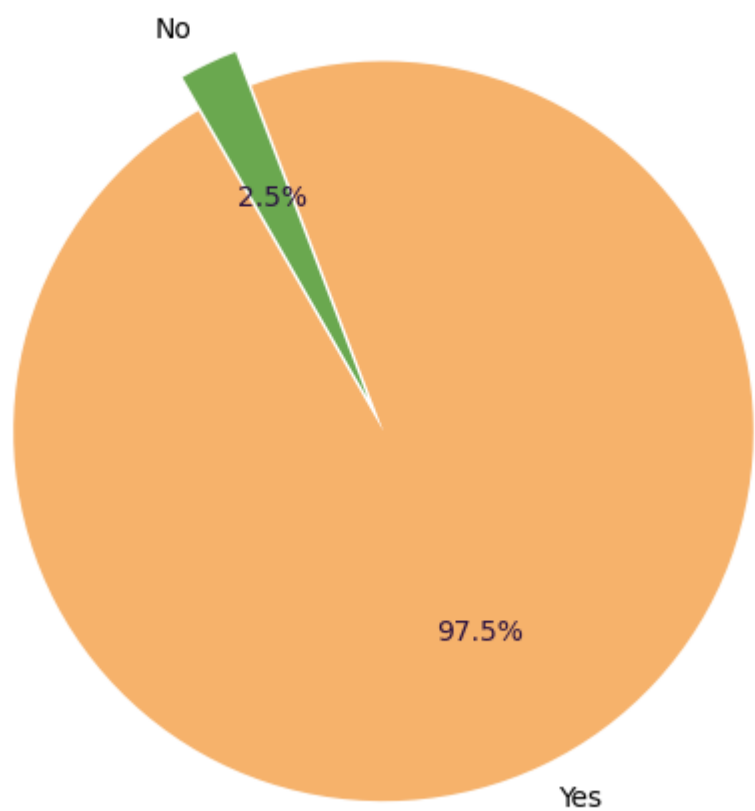
## Visualización de Datos

---

### Proporción de casos peligrosos vs no peligrosos

Se cuenta la cantidad de registros para cada clase de Dangerous y se visualiza con un gráfico de pastel ya que son datos que contrastan. Esta gráfica informa el rango de casos clínicos registrados como peligrosos, lo cual indica que la tasa de mortalidad es elevada.

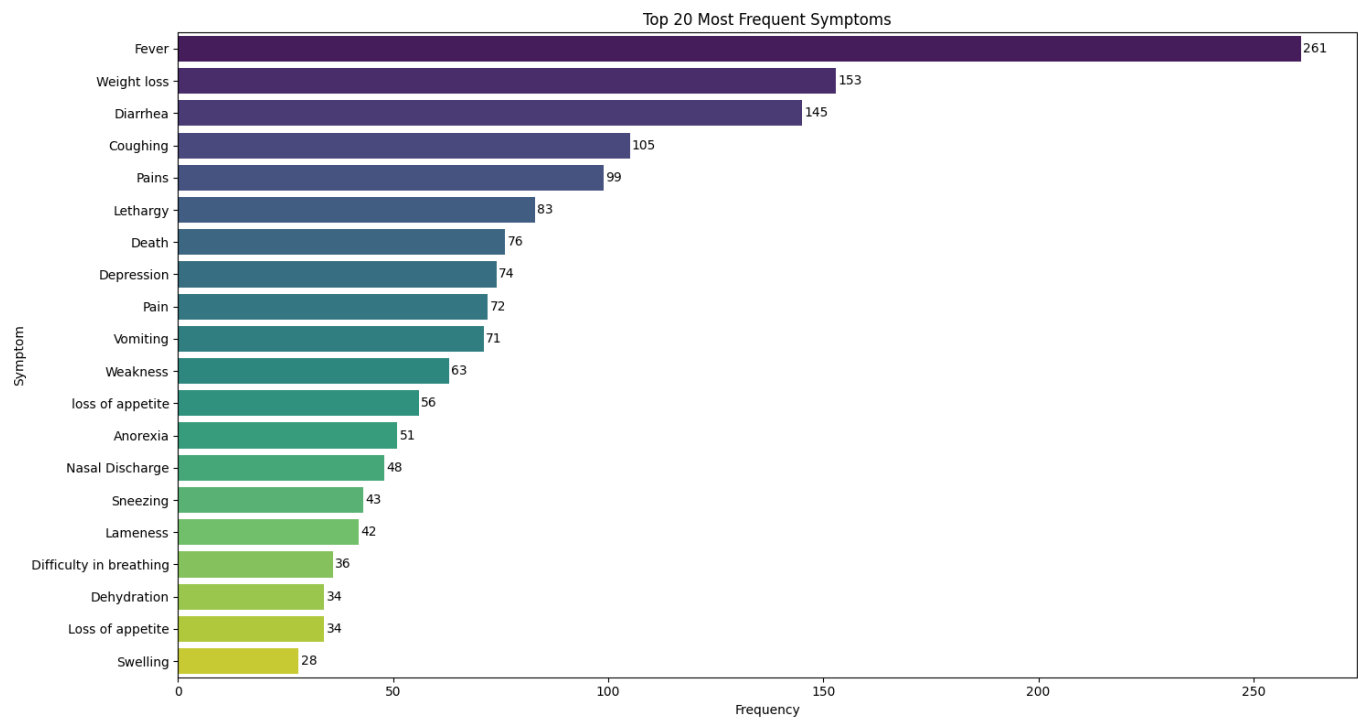
Proportion of dangerous vs. non-dangerous cases



Análisis de los síntomas
 4

¿Cuáles son los síntomas más frecuentes? Los síntomas se encuentran en las columnas symptoms1 - symptoms5. Así que unimos estos datos en una sola serie para contar su frecuencia. Usaremos un gráfico de barras (ya que nos permiten mostrar los rangos de valores de forma sencilla) y el conteo.

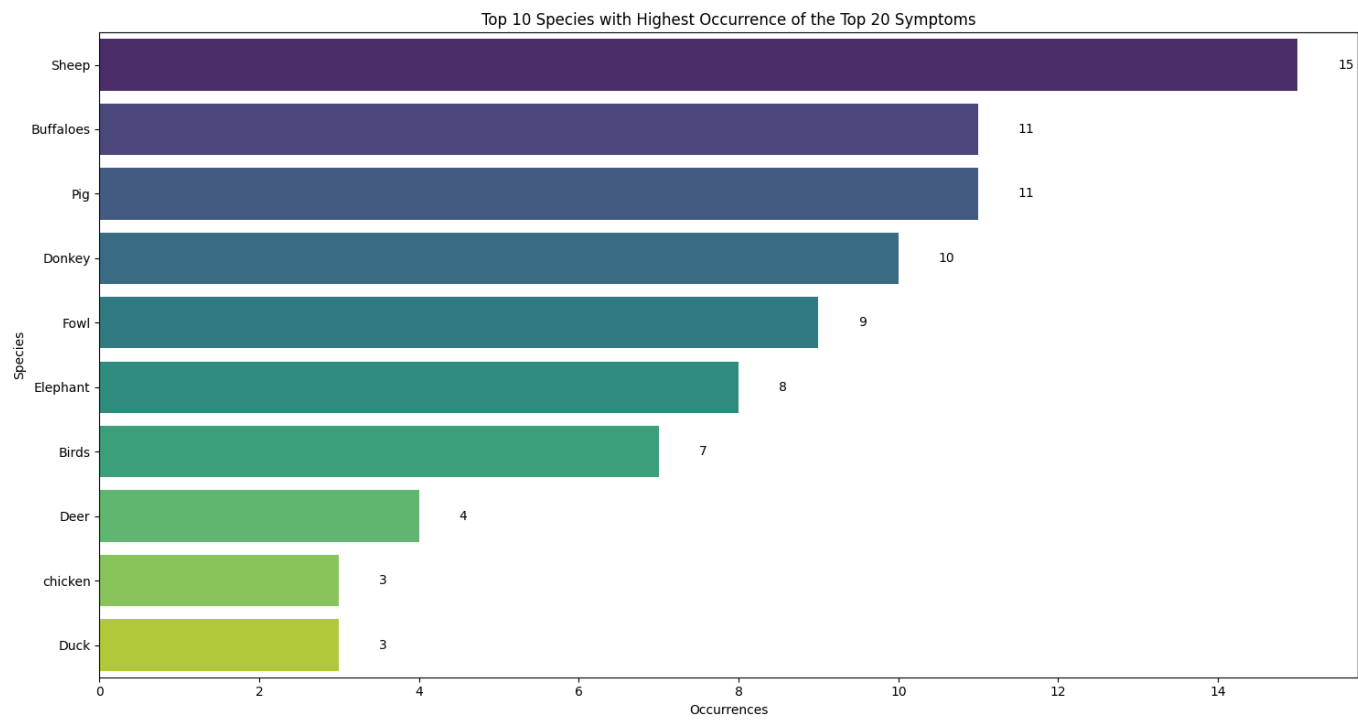
Los síntomas de "Fever", "Weight Loss" y "Diarrhea" son los síntomas más comunes. Los veterinarios deberán actuar de manera rápida cuando estos se manifiesten.



¿Cuál es la especie que más sufre esos síntomas? 🐏 🐐

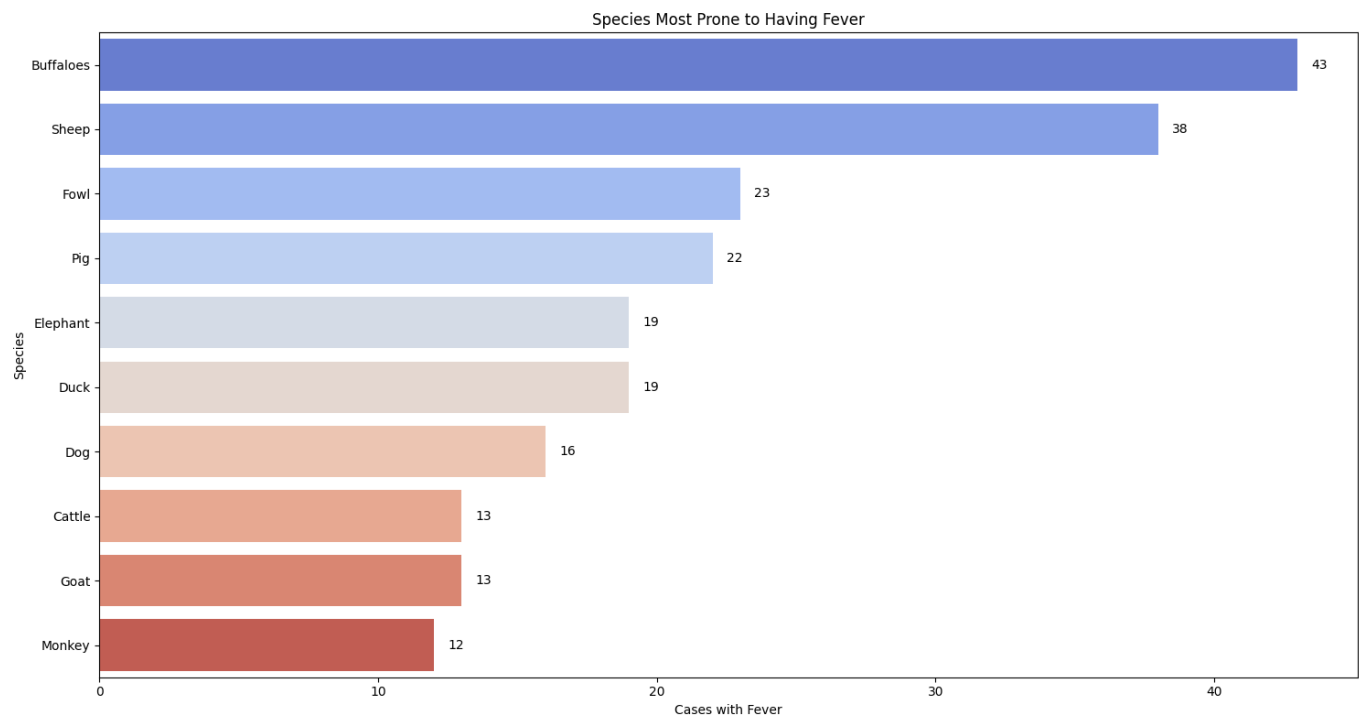
Aquí destacaría que de la lista de síntomas frecuentes las convertimos en un DataFrame largo ya que así estos síntomas van asociados a su especie y filtramos para quedarnos solo con los del top y así agruparlos.

Esta gráfica les ayuda a los veterinarios a conocer que animales deben tenerlos bajo vigilancia, eso hace que su trabajo sea menos fragmentado en esta crisis virica.



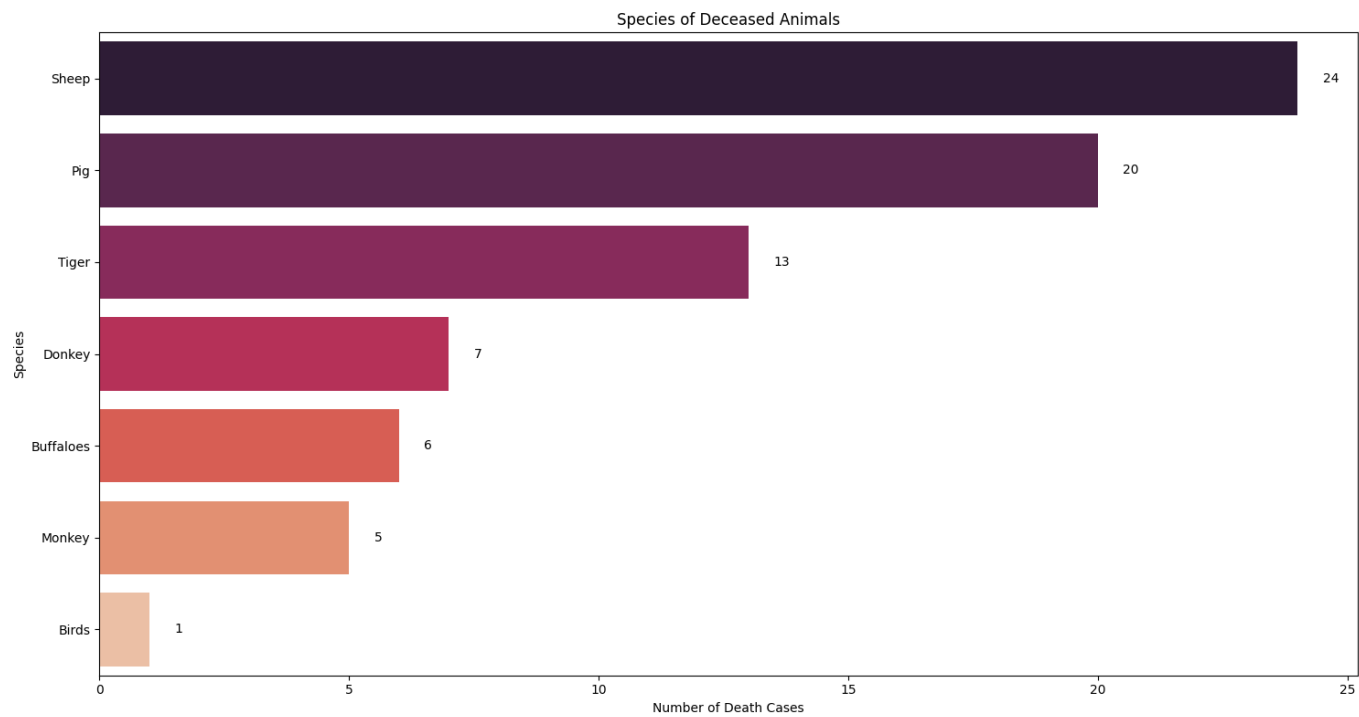
¿Cuáles son las especies que mas sufren el sintoma principal? 🦠

En este caso tener fiebre, parece ser un indicativo clave y que se repite con frecuencia en los casos sobretodo peligrosos, conocer este dato no solo refuerza la rapida ejecución, si no también conocer un top de especies puede permitir realizar mejores pruebas de diagnostico para así conseguir erradicar este virus.



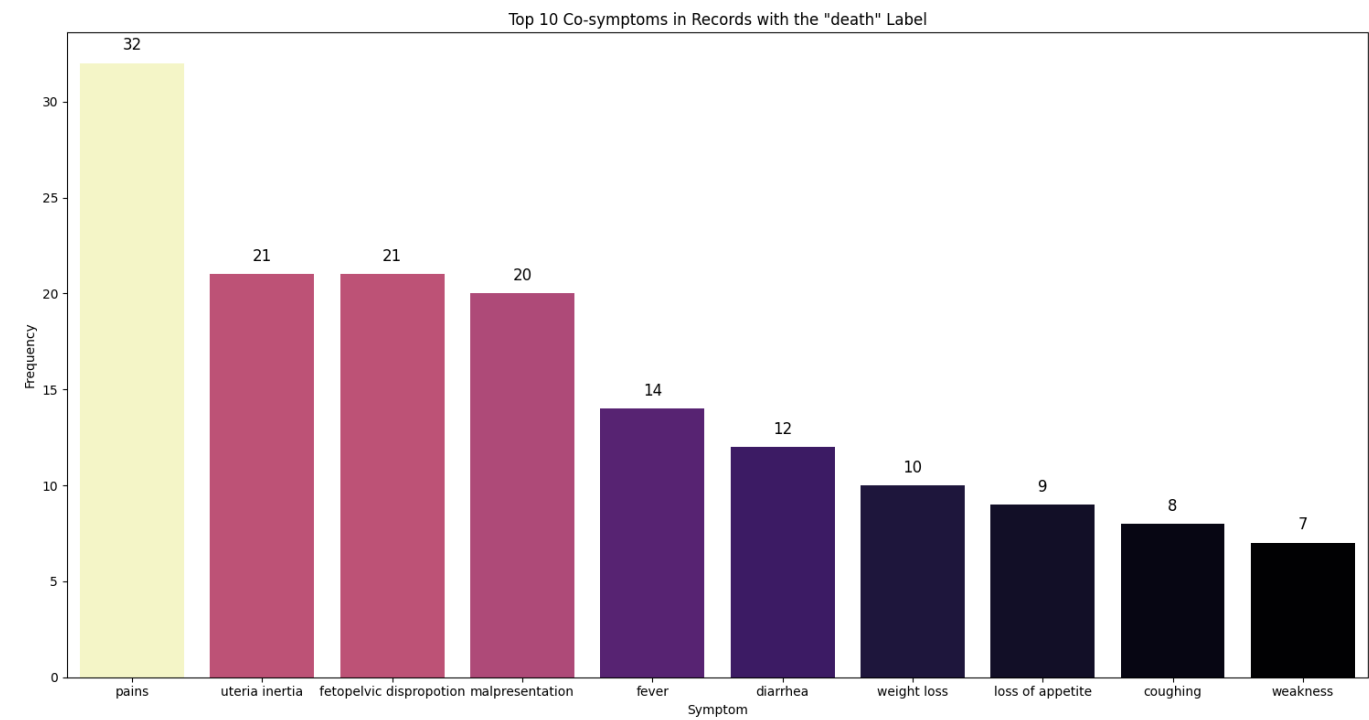
¿Cuáles son las especies que más han fallecido debido a este virus? 💀

Para ello vamos a filtrar los datos que tengan la etiqueta death, y mostrarlo en en un gráfico sencillo de lineas horizontales, esto ayudará a los médicos veterinarios a conocer que especie tiene una baja resistencia y reforzar sus defenzas en un futuro brote.

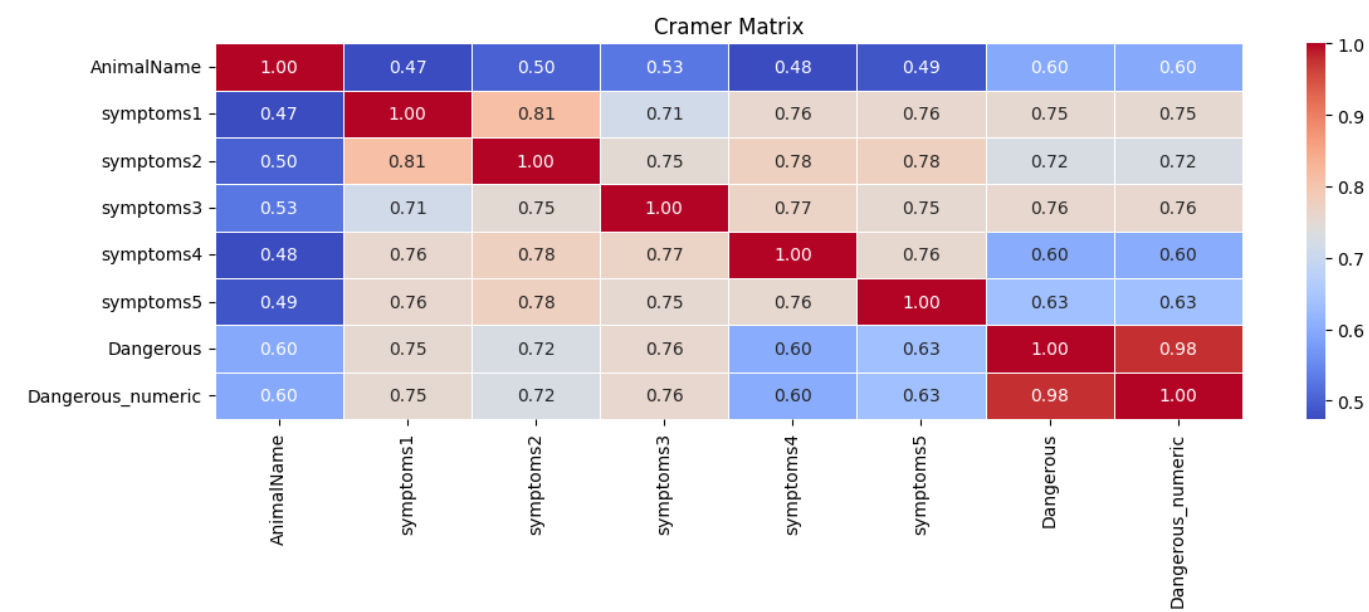


¿Cuáles son el conjunto de síntomas presentados por animales fallecidos? 🦠 🦠

Esta gráfica es clave ya que nos da un patrón de cuales han sido los síntomas que han acompañado a esos animales que tenian la etiqueta "death", esto ayuda a determinar el protocolo de actuación ya que con esto se confirman cuales son la mayoría de los síntomas letales que ocasiona el virus, si se actua de una forma rápida la tasa de mortalidad se reducirá.



Prueba de hipotesis 🔍



La matriz de V de Cramer mide la asociación entre variables categóricas. Vamos a interpretarla en relación con la hipótesis propuesta:

Hallazgos clave:

Fuerte asociación entre síntomas y "Dangerous"

Los valores de V de Cramer entre los síntomas (symptoms1 a symptoms5) y la variable Dangerous oscilan entre 0.61 y 0.78, lo que sugiere una relación moderada a fuerte.

Esto indica que ciertos síntomas tienen una influencia notable en la clasificación de los animales como en peligro.

Correlaciones altas entre síntomas

Los síntomas están altamente correlacionados entre sí (valores entre 0.74 y 0.82), lo que sugiere que los animales que presentan un síntoma tienden a mostrar otros también.

Esto podría indicar que la enfermedad progresa con un conjunto de síntomas específicos, lo que refuerza la idea de patrones sintomáticos asociados a la gravedad.

### Confirmación parcial de la hipótesis

Dado que existe una relación significativa entre los síntomas y su peligrosidad, se respalda la hipótesis de que ciertos síntomas pueden predecir qué animales es más probable que se encuentren en peligro.

Sin embargo, la alta correlación entre síntomas podría significar que un solo síntoma no es suficiente para hacer la clasificación, sino que se necesita un conjunto específico de síntomas.

## Conclusiones y próximos pasos 🐾

---

Esperar el feedback de departamento clínico y conocer si estos resultados han sido realmente fructíferos para conseguir erradicar el virus y preservar la calidad de vida de los animales.

En caso de que se necesite profundizar más y analizar más datos que pueden ayudar a una pronta detección del virus propongo:

- Hacer un análisis uno a uno de los 4 especies animales que más han fallecido a causa de este virus, con la finalidad de poder perfilar aun más los síntomas que lo causan.
- Dado que una de los principales síntomas co-relacionado con "death" es **uteria inertia** sería interesante analizar los casos que a habido así como la especie en la que más predomina y si las hembras con esta afección tienen mayor riesgo mortal al contraer el virus.
- Seguir analizando los demás síntomas del top 20 y hacer diagramas de relaciones.
- Realizar un modelo predictivo (como regresión logística o árbol de decisión) para identificar qué síntomas contribuyen más a la peligrosidad.
- Hacer un análisis de clusters para ver si hay combinaciones específicas de síntomas que se asocian con animales peligrosos.
- Aplicar pruebas estadísticas (como chi-cuadrado) para confirmar si la relación es significativa.
- En general, los datos sugieren que ciertos síntomas pueden ser utilizados para predecir la peligrosidad de los animales en el santuario.