

# Exploration Report (Ficha Vaca)

- **Primeros descubrimientos**

- Algunos atributos son **descriptivos o identificadores del evento**
  - **Tipo de evento**
  - **Fecha del evento**
  - **Descripción**
  - **Usuario**
- Uno de los atributos está completamente vacío y **no aporta información útil**
  - **Evento** (todos los valores son nulos)
- Uno de los atributos puede considerarse **de observación complementaria o comentario libre**
  - **Comentario** (más del 80% de los valores son nulos)
- El dataset contiene **múltiples categorías de eventos (≈12 tipos distintos)**, lo que sugiere que puede usarse para **clasificar o agrupar** actividades veterinarias y operativas.
- Muchos de los atributos presentan **valores nulos superiores al 50 %**, especialmente en las columnas **Evento** y **Comentario**.
- No hay una variable que funcione directamente como **variable dependiente (Y)**, ya que los datos reflejan **registros cronológicos de acciones**, no resultados medibles o continuos.
- Los atributos clave para análisis exploratorio o de patrones serían:
  - **Tipo de evento**: permite clasificar la naturaleza de las acciones.
  - **Fecha del evento**: posibilita análisis temporales o de frecuencia.

- **Hipótesis inicial y su impacto en el proyecto**

- Los eventos registrados en este dataset reflejan actividades operativas y veterinarias del rancho, como diagnósticos, tratamientos o cambios en la alimentación. Estos sucesos podrían influir en el bienestar y productividad de los animales, sin embargo, por sí solos no permiten establecer una relación directa con indicadores productivos.

- **Gráficas y figuras**

## **Tipos de datos de todas las features**

---

Lo primero que se realizó es juntar todos los datasets, pues son muy similares y con eso llegamos a tener 3512 entradas, por lo que teniendo esto hicimos una evaluación inicial para ver cuales son los elementos que los incluyen, obteniendo la siguiente información.

```

RangeIndex: 3512 entries, 0 to 3511
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Evento                 1 non-null      object
1   Tipo de evento         3511 non-null   object
2   Fecha del evento       3511 non-null   datetime64[ns]
3   Descripción            3507 non-null   object
4   Usuario                3434 non-null   object
5   Comentario             711 non-null    object
6   ID_Vaca                3512 non-null   object
dtypes: datetime64[ns](1), object(6)

```

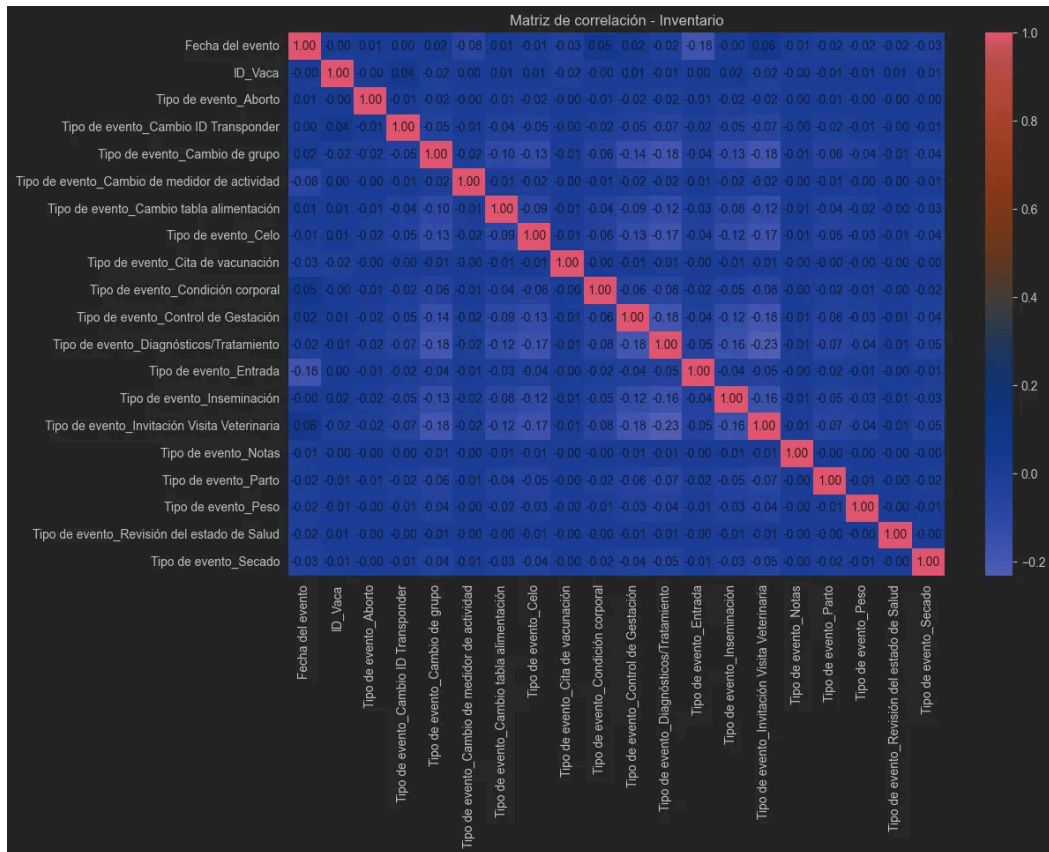
Podemos observar que la mayoría de las variables de evento y comentario vienen vacías, por lo que sería mejor eliminarlas, además de que “Comentario” no da información relevante, a menos que se quiera saber las enfermedades en concreto.

Viendo las demás variables podemos notar que la única que nos podrían funcionar son el tipo de evento, fecha de evento y el ID de la vaca pues el usuario es una variable que no nos ayudaría en nuestro análisis actual, así como la descripción que sólo es información muy específica la cual no se puede categorizar.

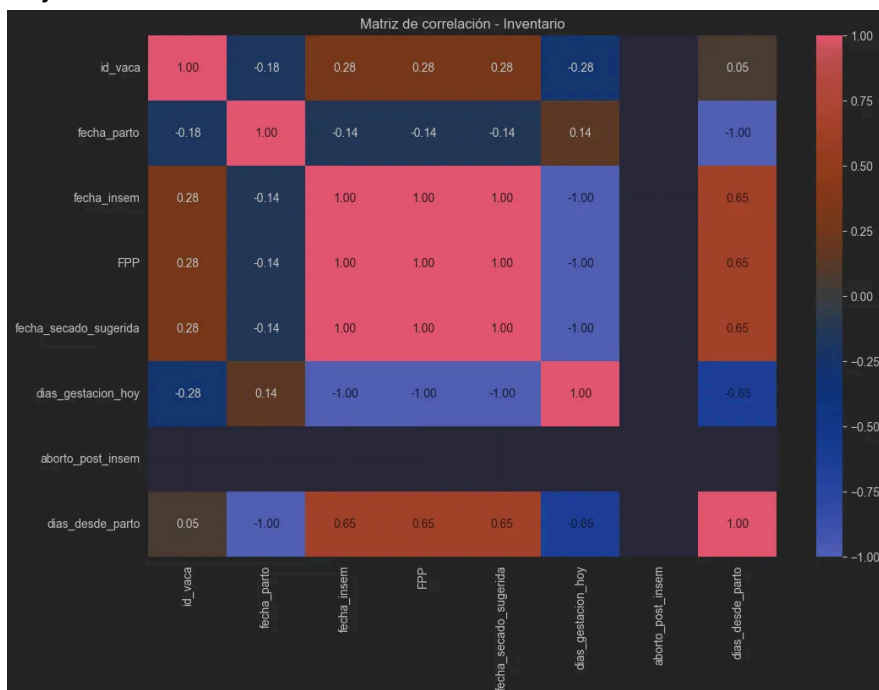
## Mapa de correlaciones

---

Realizando el One Hot Encoding en la variable de tipo de evento, hicimos una correlación con la fecha de evento y el ID dando la siguiente matriz de correlación:



Como podemos ver no hay ninguna de las correlaciones, relevantes en este dataset, por lo que se decidió hacer unas modificaciones en el dataset para que quede de mejor forma, siendo la nueva correlación de esta forma.



Como podemos ver esto tiene unas relaciones muy ligadas, algunas son muy relacionadas pero otras no tanto, pudiendo tomar como pivote la fecha de secado para ver cómo es que se podría correlacionar con otras.