

# Exploration Report (Global HATO)

- **Primeros descubrimientos**
  - Dataset limitado a los datos básicos por vaca.
  - Contiene identificadores y algunos atributos generales (estatus, grupo, fecha de entrada).
  - Baja cantidad de registros y no aporta series temporales relevantes.
- **Hipótesis inicial y su impacto en el proyecto**
  - Este archivo se puede usar como **test** debido a su baja profundidad temporal.
  - No aporta suficiente información para entrenar un modelo pero sí para validar consistencia con otros archivos más completos.
- **Gráficas y figuras**

## Tipos de datos de todas las features

---

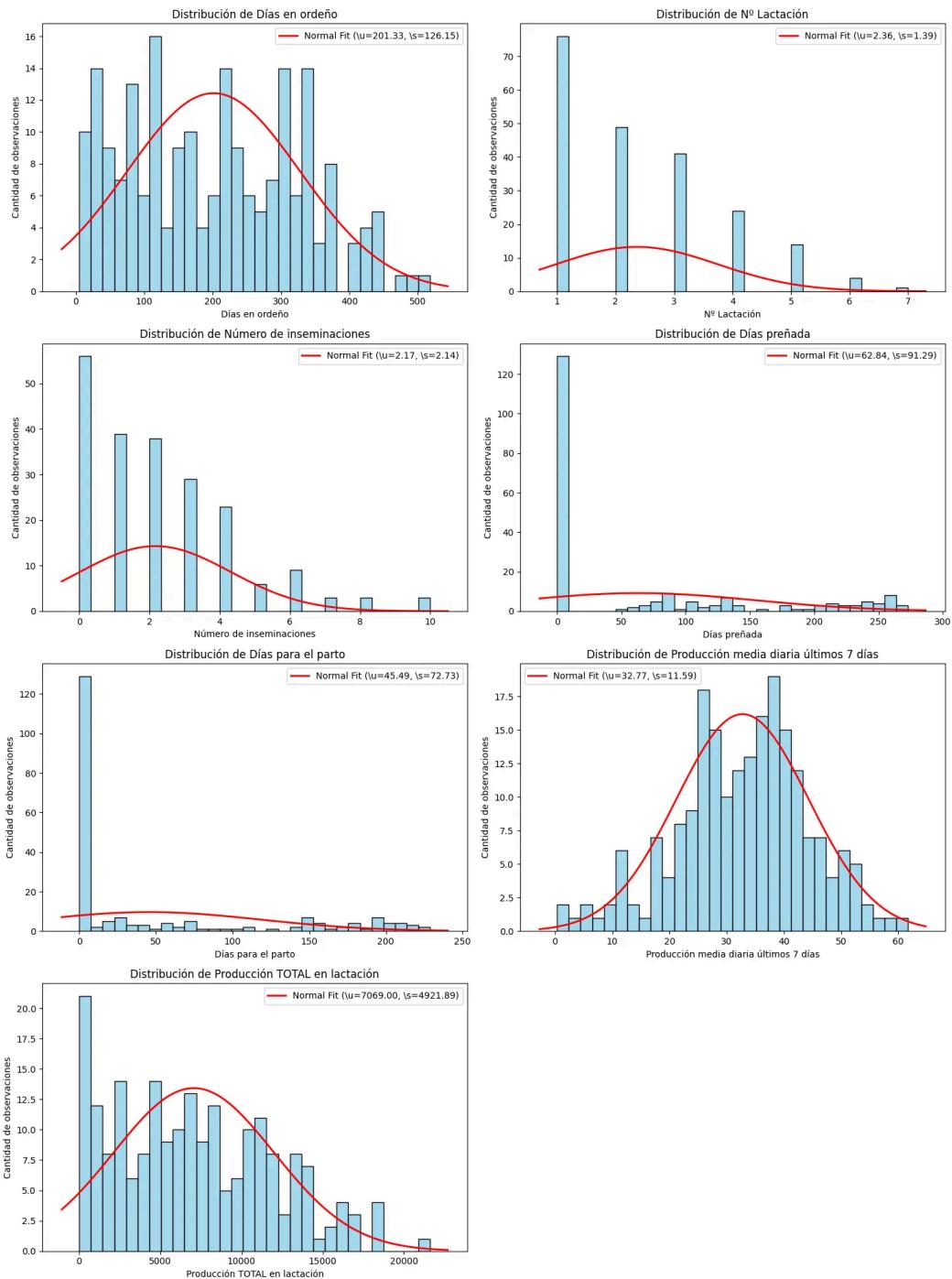
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 13 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Número del animal    583 non-null   int64  
 1   Nombre del grupo     583 non-null   object  
 2   Número(s) de selección de animal  210 non-null   object  
 3   Estado de la reproducción  413 non-null   object  
 4   Días en ordeño       211 non-null   float64 
 5   Nº Lactación        413 non-null   float64 
 6   Número de inseminaciones 198 non-null   float64 
 7   Días preñada        108 non-null   float64 
 8   Días para el parto   108 non-null   float64 
 9   Fecha de parto esperada 108 non-null   object  
 10  Producción de leche ayer  160 non-null   float64 
 11  Producción media diaria últimos 7 días 163 non-null   float64 
 12  Producción TOTAL en lactación 209 non-null   float64 
dtypes: float64(8), int64(1), object(4)
memory usage: 59.3+ KB
```

Como se puede observar los datos se encuentran en diferentes formatos como:

- Números int
- Números float
- Objects

De igual forma podemos observar que algunos de los atributos cuentan con valores faltantes, esto puede ser debido a que no se han podido registrar o en su caso que aún no entran en ese periodo de días.

## Histograma de todas las columnas para la visualización de distribuciones



- En el histograma de días de ordeño podemos observar que la distribución es muy variada, es decir que las vacas que se tienen están dispersas en sus lactancias.
- Para número de lactancias se puede observar que para este rancho la mayoría de las vacas van en su primera lactancia apenas, es decir que aun no van a ser grandes productoras.

- En número de inseminaciones se puede observar que para este momento la mayoría aún no han sido inseminadas, por otro lado también se tienen vacas que cuentan con 10 inseminaciones lo que significa que son vacas con poca fertilidad o son más difíciles de gestar.
- Para la producción media diaria de los últimos 7 días se observa que existe una distribución parecida a una normal donde la mayoría de las vacas tiene una producción entre los 25lts y los 40lts.
- En la distribución de la producción total se observa que la mayoría de las vacas están desplazadas hacia la izquierda lo que puede significar que apenas están empezando su lactancia o que su producción está siendo baja.