

# Prueba Técnica - Ing Datos

## Contexto

En el crecimiento de pollos de engorde es común encontrar todo tipo de lesiones, ya sea lesiones “físicas” o lesiones generadas por enfermedades. Los productores con los que trabajamos tienen interés en que respondamos una serie de preguntas sobre las lesiones que se están presentando sus animales. Con este objetivo en mente necesitamos desarrollar un proceso de ETL que nos permita extraer la información de un archivo de excel.

Por otra parte, nos interesa contrastar estas lesiones con el plan de vacunación de los animales de cada granja/ciclo por lo cual es de vital importancia extraer esta información.

## Parte I. ETL

Se adjunta el archivo excel donde se consigna la información de las lesiones de cada pollo. Cada pestaña contiene datos de pollos con ciertas lesiones que se presentaron en una semana particular. El objetivo de esta tarea es extraer la data y consignarla de forma tabular donde sea fácil identificar qué lesiones ha sufrido un pollo, de una granja en una fecha particular.

La data consignada en el excel contiene la información de las lesiones de cada animal. El recuadro siguiente es una muestra de la data.

Lesiones Generales*	Score**	1***	2	3
Pierna Palida	0-1			
Rasguños	0-1			
Quemadura Cojinete	0-3			
Lesion Cavity Oral	0-1			
Retención Saco Vitelino	0-1	1	1	2
Atrofia del Timo	0-2			
Bursometro	0-8	3	3	4

Lista de lesiones\*

Rango de valores posibles, cualquier dato que no cumpla esta condición se considera un mal input\*\*

Id del Animal\*\*\*

Al finalizar esta tarea se espera obtener una tabla de registros donde se consigne la información de interés, particularmente nos interesa la información de sexoAnimales,

bursómetro, condición de hígado e Integridad Intestinal. Estos 3 indicadores se encuentran dentro de la lista de lesiones, sin embargo nos interesa que el dato se replique por cada lesión encontrada. Adjunto se envía un archivo sample.xlsx con la estructura de los datos deseada.

Adicionalmente se adjunta un archivo json que contiene los nombres de las granjas y las lesiones estandarizados

Preguntas clave:

- ¿Cuántos pollos tienen datos erróneos?
- ¿Qué hacemos con datos erróneos?
- ¿Cuántos registros de la tabla final contienen datos erróneos?
- ¿Qué información clave podemos extraer de esta tabla que sea interesante para nuestros stakeholders?

## **Parte II Servir los datos.**

El cliente nos pide un sistema que permita acceder a la información a través de una aplicación web. Para esto definimos en nuestra arquitectura la utilidad de crear un paquete/scripts en Python o un sistema similar donde se exponga la información y se creen los siguientes filtros para obtener registros de la tabla creada en la parte I:

- Filtro por granja
- Filtro por fechas (Inicio, Fin)
- Filtro por lesiones

Se requiere un demo que evidencie el acceso a la información

## **Parte III Scale.**

Deseamos que el sistema escale para recibir y guardar datos de varios clientes. Diseñe un sistema de bases de datos que permita guardar estos datos. Justifique su diseño y selección. Puede utilizar una herramienta como [dbdiagram.io - Database Relationship Diagrams Design Tool](https://dbdiagram.io/)

No se requiere implementación del mismo

## **Entregables Partes I-II-III**

- Código con las soluciones propuestas
- Tabla resultado del pipeline de ETL (En archivo csv o.xlsx)
- Demo Parte II

- Diseño Parte III.

#### **Parte IV Presentación.**

Posterior al envío de los entregables el candidato presentará sus resultados en una reunión concertada con el evaluador.