

2025

PROYECTO

Inteligencia Artificial ➔

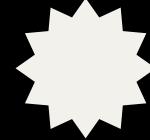
<https://colab.research.google.com/drive/1Xm5SOlaP7-jCJBoMB1q26Gwbq4t1Q8xD>

Integrantes :

Rios Duran Israel

Alvarado Ramirez Marco

Levano Talla Diego



2025

REPUBLICA DEL PERU

ENCUESTA NACIONAL AGROPECUARIA 2025

INFORMACIÓN CONFIDENCIAL AMPARADA POR DECRETO SUPREMO N° 043-2001-PCM DEL SECRETO ESTADÍSTICO

INEI INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA

DOC.01.03

Nº DE LA UNIDAD AGROPECUARIA

CUESTIONARIO N° Cuestionario Adicional
1

UBICACIÓN GEOGRÁFICA Y MUESTRAL DE LA UNIDAD AGROPECUARIA

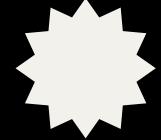
1. DEPARTAMENTO:	4. SEGMENTO N°:	4A. TOTAL DE PRODUCTORES AGROPECUARIOS EN EL SEGMENTO			
2. PROVINCIA:	5. CUADRANTE ALEATORIO N°:	5A. USO: Sí.....1 No.....2			
3. DISTRITO:	6. UNIDAD AGROPECUARIA N°:	5B. TOTAL DE PARTICIONES:			
7. CENTRO POBLADO MÁS CERCANO A LA UNIDAD AGROPECUARIA:					
7A. DETERMINACIÓN DE UNIDAD AGROPECUARIA: Sí.....1 No.....2					
7B. DETERMINACIÓN DE PRODUCTOR PECUARIO SIN TIERRA DENTRO DEL DISTRITO: Sí.....1 No.....2					
9. ENTREVISTA					
VISITA	ENTREVISTADOR				(*) CÓDIGOS DE RESULTADO
	FECHA	HORA	PRÓXIMA VISITA	RESULTADO DE LA VISITA (*)	
DE	A	FECHA	HORA		
Primera					

Encuesta Nacional Agropecuaria

Este dataset contiene información recolectada por la Encuesta Nacional Agropecuaria (ENA), elaborada por el Instituto Nacional de Estadística e Informática (INEI) del Perú.

Su objetivo es describir las condiciones del sector agrícola y ganadero del país, incluyendo características de los productores, cultivos, tierras, animales, uso de tecnología, y prácticas agrícolas.

[Encuesta Nacional Agropecuaria - ENA 2025 - Informes y publicaciones - Instituto Nacional de Estadística e Informática - Plataforma del Estado Peruano](#)



2025

Propósito Dataset

Análisis Ambiental

Evaluar prácticas agrícolas y uso de recursos (agua, suelo, pastos) desde una perspectiva de sostenibilidad.

Manejo de Cultivos y Pastos

Analizar frecuencia de cortes, tipo de cosecha y densidad de cultivos en relación con la conservación del suelo.

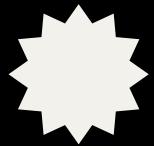
Gestión del Agua

Identificar fuentes de riego y su impacto ambiental en la agricultura regional.

Sostenibilidad Rural

Generar información confiable sobre el equilibrio entre producción agrícola y cuidado ambiental.

VARIABLE OBJETIVO



2025

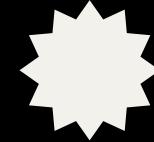
■ P213

- Fuente de agua
- Río, pozo, manantial, reservorio
- Gestión hídrica
- Sostenibilidad
- Impacto ambiental

213.	¿QUÉ SISTEMA DE RIEGO UTILIZÓ?
Exudación ..	1
Goteo.....	2
Micro- aspersión....	3
Aspersión... .	4
Multicom- puertas	5
Mangas	6
Gravedad ...	7
Otro _____	8
(Especifique)	
CÓDIGO	



ANALISIS AMBIENTAL



2025

P218

- Densidad de cultivo
 - Uso del suelo
 - Conservación ecosistémica

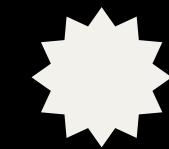
P218 A

- Pastoreo intensivo → erosión
 - Corte controlado → sostenible
 - Regeneración de pastos

P218 B

- Frecuencia = presión suelo
 - Muchos cortes = sobreexplotación
 - Pocos cortes = equilibrio
 - Sostenibilidad forrajera

Inputacion de Datos Faltantes



2025

■ Moda

Es el valor o categoría que más veces se repite en un conjunto de datos.

■ Imputación condicionada

Es el reemplazo de datos faltantes utilizando valores estimados dentro de subgrupos o condiciones específicas

Ejemplo: si conoces la región, el tamaño de parcela o el tipo de cultivo, puedes usar la moda dentro de ese grupo.

Así, en zonas costeras podríamos imputar con “canal de regadío” y en zonas altoandinas con “manantial” o “agua de lluvia almacenada”.



LECTURA DE DATASET:



```
from google.colab import drive
drive.mount('/content/drive')
import pandas as pd
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt
entrenamiento = pd.read_csv('/content/drive/MyDrive/2025-II/Inteligencia
Artificial/EncuestaNacionalAgropecuaria.csv',
sep=',',
encoding='latin-1',
low_memory=False,
on_bad_lines='skip',
na_values=['', ' ', ' ', ' ', ' ', 'NA', 'N/A', 'n/a', 'null', 'NULL', 'Null', '-', '--', '---', 'nan', 'NaN', '#N/A',
'#NA'],
keep_default_na=True)
```



DATASET: PRIMERAS 5 FILAS



2025

```
14     entrenamiento.head() #Devuelve las 5 primeras filas del dataset
```

	ANIO	CCDD	NOMBREDD	CCPP	NOMBREPV	CCDI	NOMBREDI	CONGLOMERADO	NSELUA	UA	...	P223A	P223B_1	P223B_2	P223B_3	P223B_4	P223B_5	P223B_6	P223B
0	2022	24	TUMBES	2	CONTRALMIRANTE VILLAR	3	CANOAS DE PUNTA SAL		1	3	1	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0
1	2022	24	TUMBES	2	CONTRALMIRANTE VILLAR	3	CANOAS DE PUNTA SAL		1	3	1	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0
2	2022	24	TUMBES	2	CONTRALMIRANTE VILLAR	3	CANOAS DE PUNTA SAL		1	3	1	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0
3	2022	24	TUMBES	2	CONTRALMIRANTE VILLAR	3	CANOAS DE PUNTA SAL		1	3	1	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0
4	2022	24	TUMBES	2	CONTRALMIRANTE VILLAR	3	CANOAS DE PUNTA SAL		1	3	1	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0

5 rows × 122 columns

DATASET: ULTIMAS 5 FILAS



2025

```
15  entrenamiento.tail() #Devuelve las 5 ultimas filas del dataset
```

	ANIO	CCDD	NOMBREDD	CCPP	NOMBREPV	CCDI	NOMBREDI	CONGLOMERADO	NSE
105608	2022	23	TACNA	1	TACNA	1	TACNA		19264
105609	2022	23	TACNA	1	TACNA	1	TACNA		19264
105610	2022	23	TACNA	1	TACNA	1	TACNA		19264
105611	2022	23	TACNA	1	TACNA	1	TACNA		19264
105612	2022	23	TACNA	1	TACNA	1	TACNA		19264

5 rows × 122 columns

DATASET: INFORMACIÓN DE COLUMNAS



2025

```
16   entrenamiento.info() #Devuelve información de las columnas del dataset
```

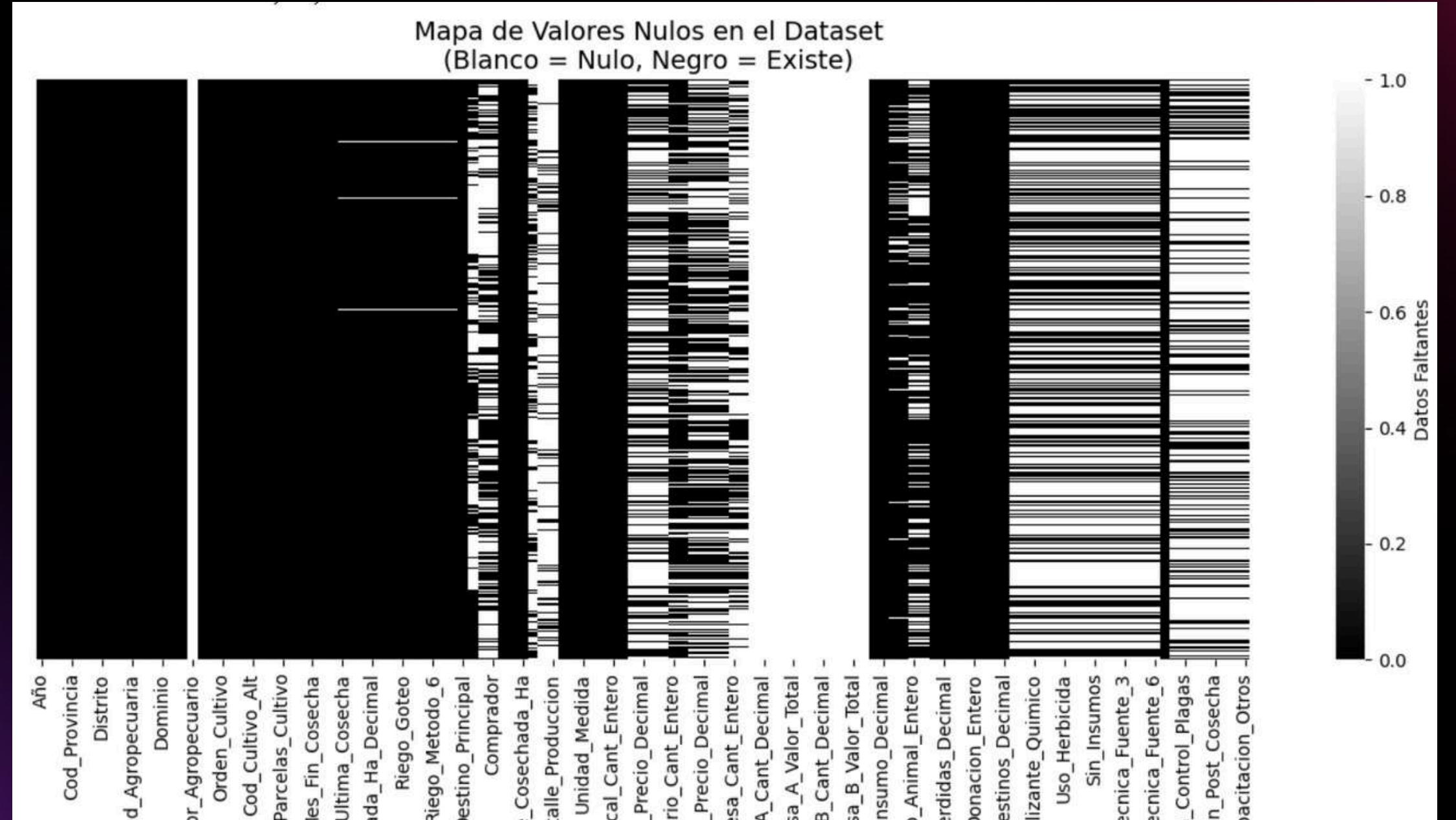
<class 'pandas.core.frame.DataFrame'>	39	P211_6	104997	non-null	float64	97	P221_1	51068	non-null	float64
RangeIndex: 105613 entries, 0 to 105612	40	P211_7	104997	non-null	float64	98	P221_2	51068	non-null	float64
Data columns (total 122 columns):	41	P211_8	105003	non-null	float64	99	P222_1	51068	non-null	float64
# Column	42	P212	105613	non-null	int64	100	P222_2	51068	non-null	float64
Column	43	P213	49378	non-null	float64	101	P222_3	51068	non-null	float64
Non-Null Count Dtype	44	P214	54621	non-null	float64	102	P222_4	51068	non-null	float64
-----	45	P215	54620	non-null	float64	103	P222_5	51068	non-null	float64
0 ANIO	46	P217_SUP_1	105613	non-null	int64	104	P222_6	51068	non-null	float64
1 CCDD	47	P217_SUP_2	105613	non-null	int64	105	P222_7	51068	non-null	float64
2 NOMBREDD	48	P217_SUP_ha	105613	non-null	float64	106	P223_1	51068	non-null	float64
3 CCP	49	P218	35840	non-null	float64	107	P223_2	51068	non-null	float64
4 NOMBREPV	50	P218A	14248	non-null	float64	108	P223_3	51068	non-null	float64
5 CCDI	51	P218B	14248	non-null	float64	109	P223_4	51068	non-null	float64
6 NOMBREDI	52	P219_CANT_1	105613	non-null	int64	110	P223_5	51068	non-null	float64
7 CONGLOMERADO	53	P219_CANT_2	105613	non-null	int64	111	P223_6	51068	non-null	float64
8 NSELUA	54	P219_UM	105613	non-null	object	112	P223A	105612	non-null	float64
9 UA	55	P219_UM_COD	105613	non-null	int64	113	P223B_1	36622	non-null	float64
10 RESFIN	56	P219_EQUIV_KG	105613	non-null	float64	114	P223B_2	36622	non-null	float64
11 REGION	57	P220_1_CANT_1	105613	non-null	int64	115	P223B_3	36622	non-null	float64
12 DOMINIO	58	P220_1_CANT_2	105613	non-null	int64	116	P223B_4	36622	non-null	float64
13 FACTOR	59	P220_1_PREC_1	51068	non-null	float64	117	P223B_5	36622	non-null	float64
14 CODIGO	60	P220_1_PREC_2	51082	non-null	float64	118	P223B_6	36622	non-null	float64
15 P101A	61	P220_1_VAL	51068	non-null	float64	119	P223B_7	36622	non-null	float64
16 P102_1	62	P220_1_PRE_KG	51068	non-null	float64	120	P223B_8	36622	non-null	float64
17 P102_2	63	P220_2_ENT	87434	non-null	float64	121	OMICAP200	0	non-null	float64
18 P105_N	64	P220_2_DEC	87434	non-null	float64	dtypes: float64(75), int64(42), object(5)				
19 P203	65	P220_2_PREC_1	62328	non-null	float64	memory usage: 98.3+ MB				
20 P204_NOM	66	P220_2_PREC_2	62328	non-null	float64					
21 P204_COD										
22 P204_TIPO										
23 P205_N										

Dataset:



2025

NaNs X Columnas



NaN X Enfoque

