

# ***Informe de Análisis Exploratorio de Datos (EDA): Censo Nacional Agropecuario – Ganado***

## **1 Introducción**

Dado que el foco del presente estudio es sobre las Unidades Productoras Agropecuarias (UPA) en la zona rural dispersa de Colombia durante el período de noviembre de 2013 a noviembre de 2014, surgen un par de preguntas

¿Cuáles son los factores que afectan la productividad agrícola de las Unidades Productoras Agropecuarias (UPA) en la zona rural dispersa de Colombia entre noviembre de 2013 y noviembre de 2014?

¿Cómo influye la ubicación geográfica y las condiciones socioeconómicas en los niveles de productividad de las Unidades Productoras Agropecuarias en las zonas rurales de Colombia durante este periodo?

### **1.1 Objetivo del informe:**

Analizar cómo las Unidades Productoras Agropecuarias se relacionan con la cantidad de bovinos disperso colombiano entre noviembre de 2013 y noviembre de 2014.

### **1.2 Objetivos Específicos:**

- Realizar un análisis descriptivo de las características de las UPA.
- Identificar patrones y correlaciones entre las variables sociales, económicas y geográficas con la productividad de las UPA.
- Realizar un análisis geoespacial para observar para ver cómo se correlacionan con factores específicos.
- Evaluar si la ubicación geográfica (departamentos) tienen un impacto en los niveles de producción.

### **1.3 Descripción del Dataset:**

El dataset contiene información sobre las Unidades Productoras Agropecuarias (UPA) en las zonas rurales dispersas de todo el país. Entre las variables incluidas están el número de UPA con ganado bovino y el número de cabezas de ganado censado en estas unidades productivas.

## **2 Descripción del Problema**

En esta sección, se detallan las variables del dataset, las preguntas de investigación y el contexto del problema a analizar.

### **2.1 Variables:**

**Código Departamento:** Código de la división Político Administrativa del país

**Departamento:** Nombre del departamento

**Código Municipio:** Código de la división Político Administrativa del país

**Municipio:** Nombre del Municipio

**Número de UPA con ganado bovino:** El número total de Unidades Productoras Agropecuarias que con ganado bovino en el área rural dispersa censada.

**Número de cabezas de ganado:** La cantidad de cabezas de ganado bovino registradas en el área rural dispersa censada.

## 2.2 Preguntas del trabajo:

¿Cuál es la distribución geográfica de las UPA con ganado bovino en el área rural dispersa?

¿Cómo se correlacionan las variables de número de UPA y número de cabezas de ganado?

¿Cuáles son los factores que influyen en la concentración del ganado bovino en ciertas regiones?

# 3 Proceso de Análisis

## 3.1 Análisis Exploratorio de Datos (EDA)

**Objetivo:** El objetivo de esta fase es comprender la estructura, distribución y características del conjunto de datos para obtener insights iniciales y posibles áreas de interés para un análisis posterior.

## 3.2 Pasos realizados en el EDA:

**Carga y limpieza de los datos:** Inspección de los datos para verificar la existencia de valores nulos, duplicados, inconsistencias, y tratamiento de los mismos.

### 3.2.1 Carga de los datos:

```
ganado = pd.read_csv('../Censo_Nacional_Agropecuario_-_Ganado.csv')
```

### 3.2.2 Resumen de los datos:

```
ganado.head(3)
ganado.info()
ganado.describe()
```

### 3.2.3 Identificar Datos Faltantes y Outliers: Verificar valores nulos y anomalías en los datos

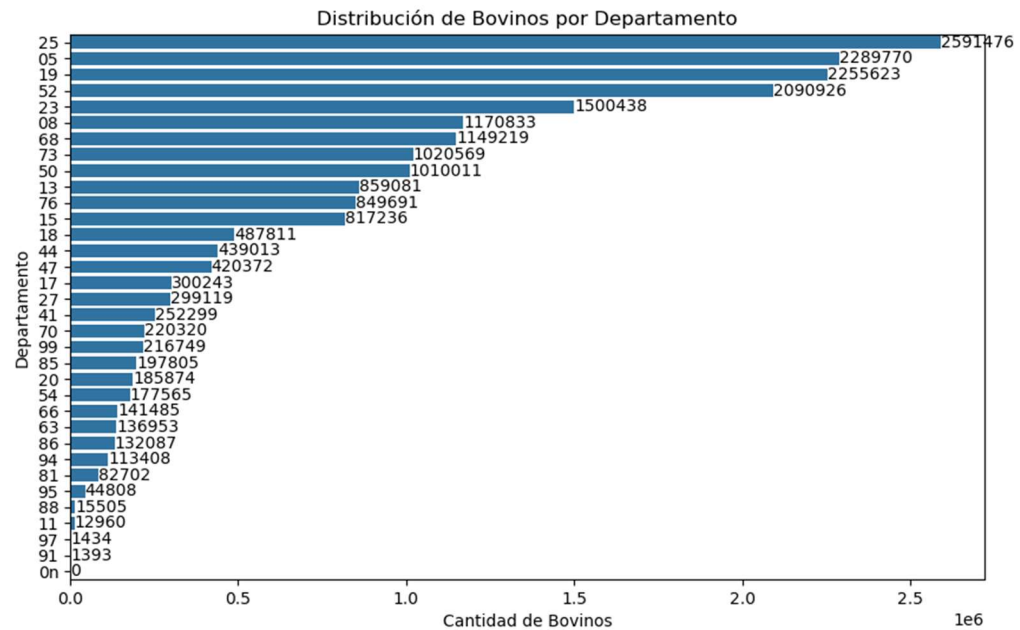
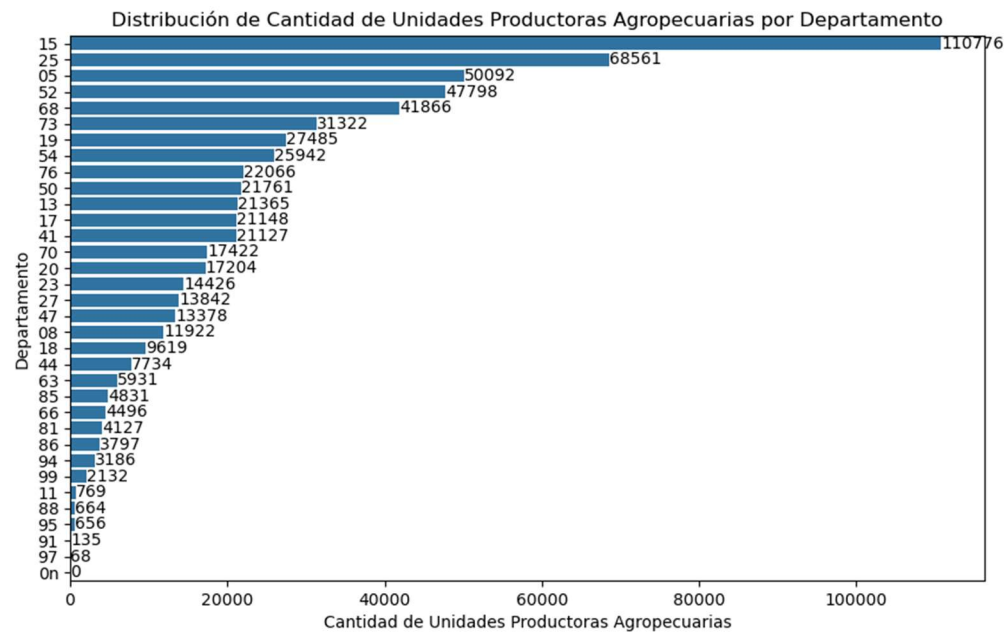
```
registros_nulos = resumen_ganado[resumen_ganado.isnull().any(axis=1)]
```

### 3.2.4 Distribución de Variables Categóricas: Utilizar gráficos de barras para ver la distribución de las variables categóricas

```
sns.countplot(x='COD_DEP', data=df)
```

### 3.2.5 Visualización de Variables Numéricas: Usar histogramas para explorar las distribuciones de las variables numéricas

```
sns.histplot(df['UPAGanad'])
```



### 3.3 Análisis Univariado:

#### 3.3.1 Variables Numéricas:

```
resumen_ganado.describe()
```

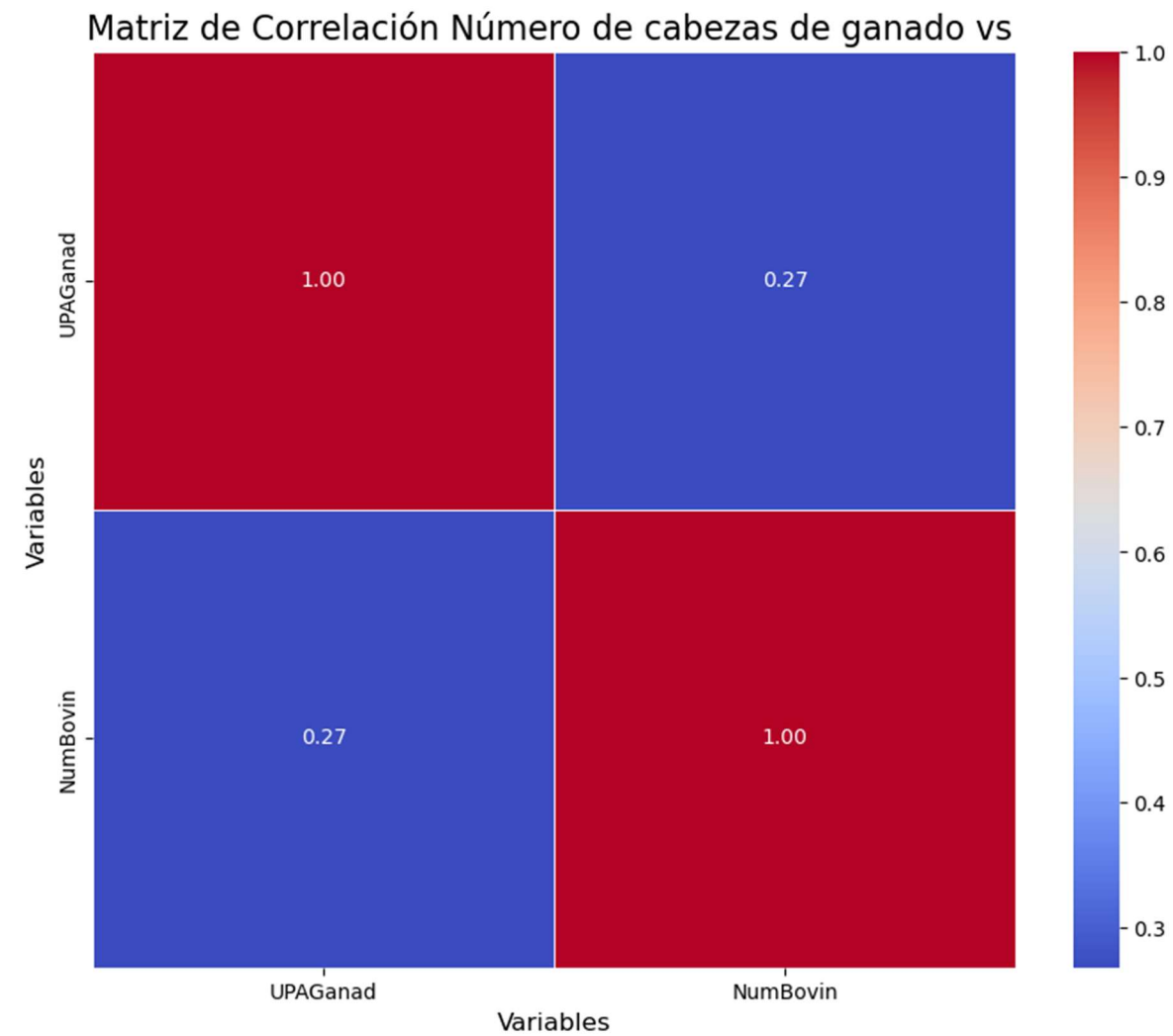
#### 3.3.2 Variables Categóricas:

```
frecuencias = ganado['COD_DEP'].value_counts()
```

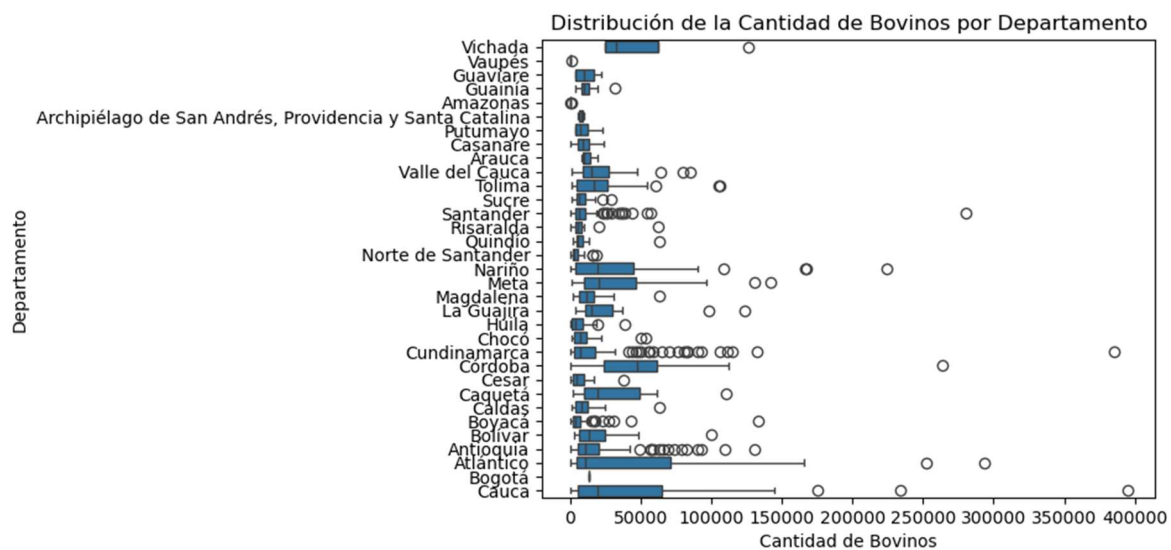
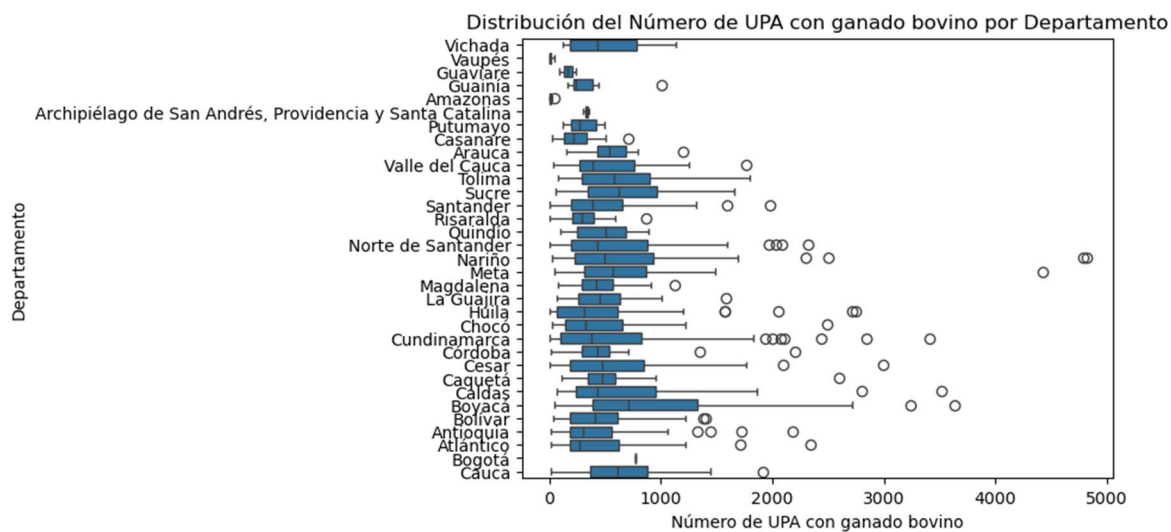
```
print(frecuencias)
```

### 3.4 Análisis Bivariado: Análisis entre dos variables

#### 3.4.1 Relación entre variables numéricas



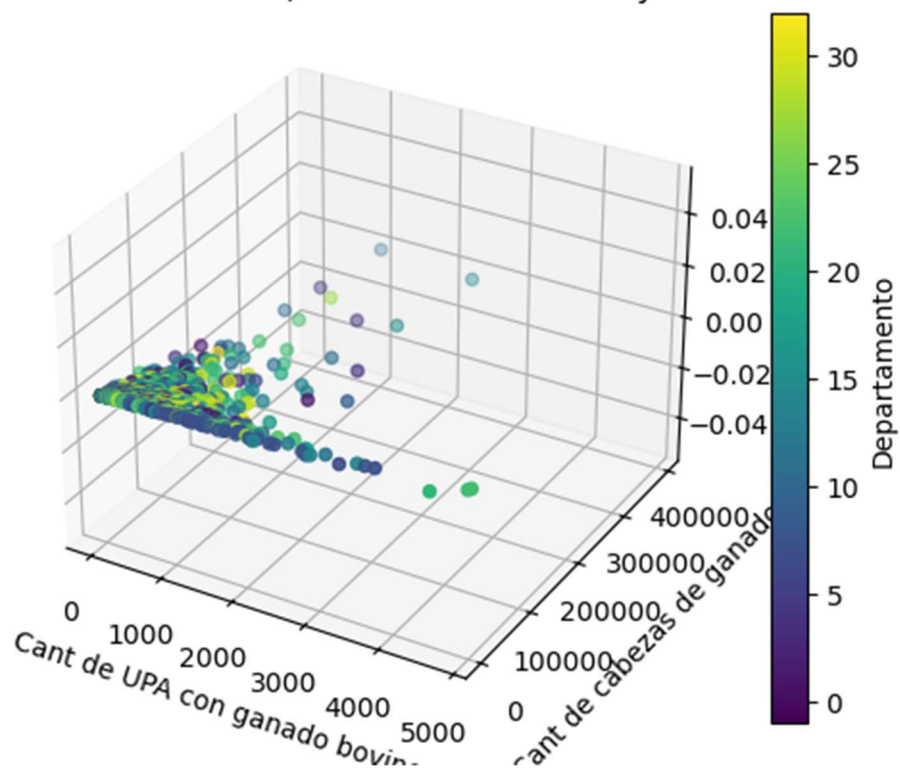
### 3.4.2 Relación entre variable numérica y categórica



### 3.5 Análisis Multivariado:

#### 3.5.1 Gráficos de dispersión 3D

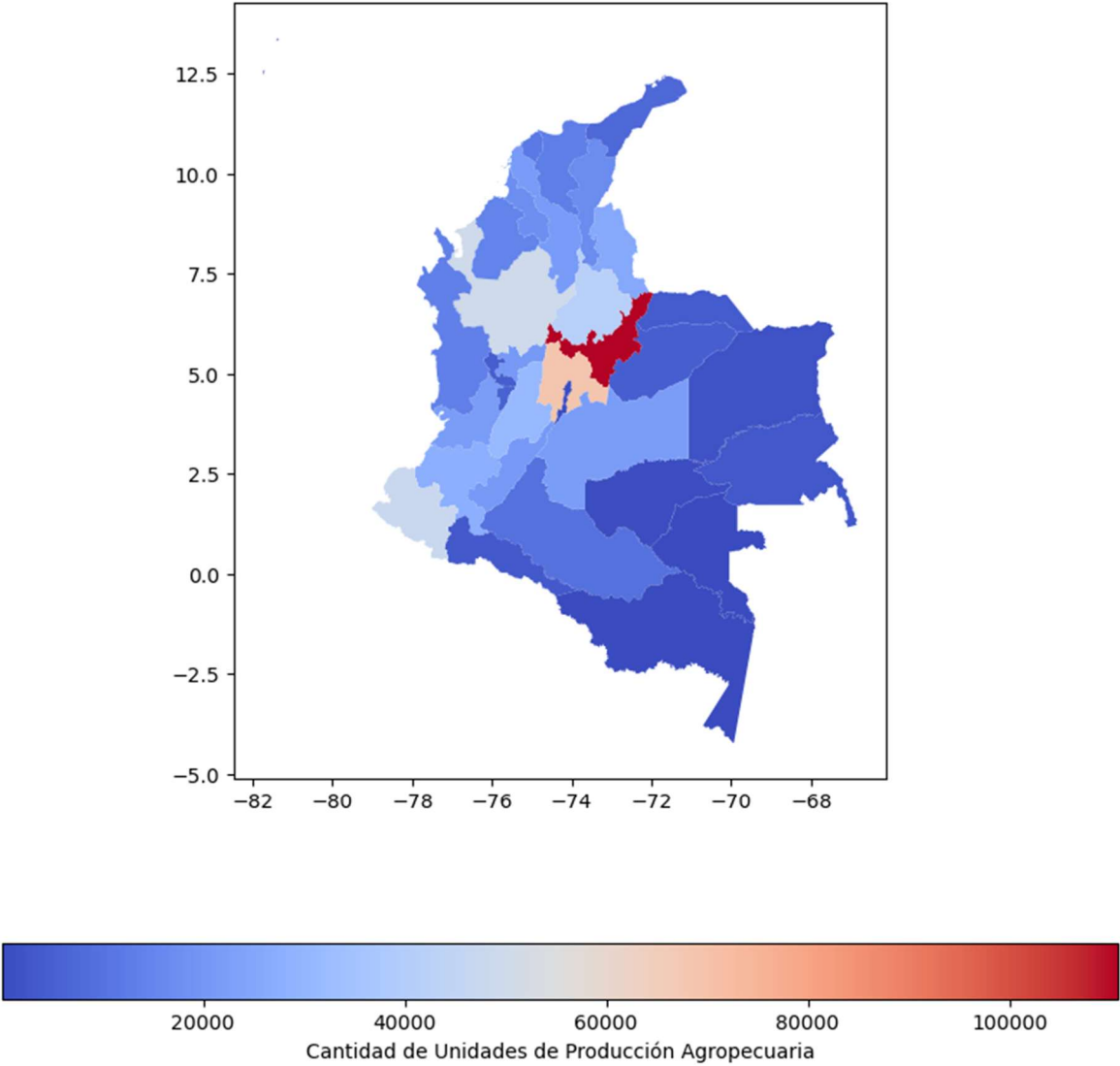
Relación entre UPA, Cabezas de Ganado y DPTO

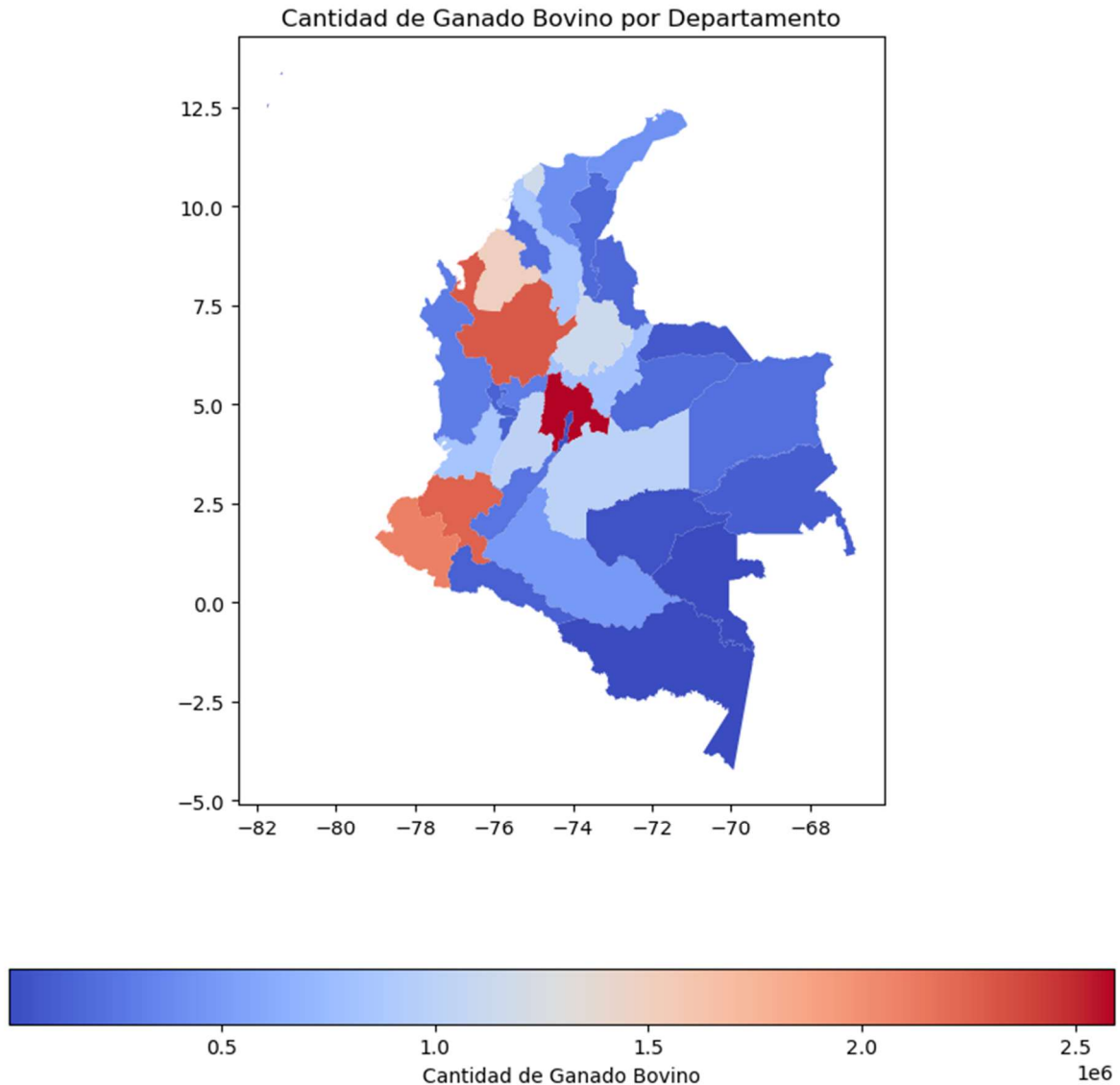


### 3.6 Análisis Geoespacial

#### 3.6.1 Visualización en Mapa

Cantidad de Unidades de Producción Agropecuaria por Departamento





## 4 Resultados

- 4.1 **Descripción del dataset:** El conjunto de datos "Censo\_Nacional\_Agropecuario\_-\_Ganado.csv" contiene un total de 1122 registros, correspondientes a 1121 municipios, con un registro en blanco.
- 4.2 **Resumen de la variable "sum\_UPAGanad" (Unidades Productoras Agropecuarias):**

Número de observaciones (n) = 34  
Media (M) = 19,048.47  
Desviación estándar (SD) = 23,028.87



Valor mínimo = 0.00  
Percentil 25 = 3,879.50  
Mediana (Percentil 50) = 13,610.00  
Percentil 75 = 21,989.75  
Valor máximo = 110,776.00

#### 4.3 Resumen de la variable "sum\_bovin" (Número de bovinos):

Número de observaciones (n) = 34  
Media (M) = 631,905.24  
Desviación estándar (SD) = 740,688.15  
Valor mínimo = 0.00  
Percentil 25 = 133,303.50  
Mediana (Percentil 50) = 275,709.00  
Percentil 75 = 972,278.50  
Valor máximo = 2,591,476.00

#### 4.4 Análisis de las Unidades Productoras Agropecuarias (UPA):

- En promedio, los departamentos de Colombia tienen 19,048.47 Unidades Productoras Agropecuarias (UPA), con una desviación estándar de 23,028.87 y una mediana de 13,610.
- El departamento con la menor cantidad de UPA es Vaupés, con un valor de 0, mientras que el departamento con la mayor cantidad de UPA es Boyacá, con 110,776.

#### 4.5 Análisis del número de bovinos:

- En promedio, los departamentos de Colombia reportan 631,905.24 bovinos, con una desviación estándar de 740,688.15 y una mediana de 275,709.
- El departamento con la menor cantidad de bovinos es Amazonas, con 0, mientras que el departamento con la mayor cantidad de bovinos es Cundinamarca, con 2,591,476.

#### 4.6 Relación entre las Unidades Productoras Agropecuarias y el número de bovinos:

El coeficiente de correlación entre el número de Unidades Productoras Agropecuarias (UPA) y el número de bovinos es de 0.27, lo que indica una correlación positiva baja entre ambas variables.