

Evaluaciones Agropecuarias de los municipios de Cundinamarca y Boyacá

Andrea Nataly Hernández Fuentes, Edgar Felipe Torres Ortegón,
Helmy Andrea Moreno Navarro, y Samuel Mantilla Velásquez

Universidad Sergio Arboleda

Talento Tech

Objetivos

Objetivo General

Realizar un análisis integral de la biodiversidad, conservación ambiental y los ecosistemas en las regiones de Cundinamarca y Boyacá, utilizando técnicas avanzadas de análisis de datos y modelos predictivos. A través de la exploración, análisis y modelado de datos, buscamos identificar patrones significativos, áreas de interés prioritarias y proponer recomendaciones fundamentadas para la gestión sostenible y la conservación eficaz de la biodiversidad en estas regiones.

Objetivos Específicos

1. Realizar análisis estadísticos descriptivos para comprender la distribución y variabilidad de las Features.
2. Crear visualizaciones efectivas, como mapas, gráficos y diagramas, para comunicar de manera clara y persuasiva los resultados del análisis.
3. Identificar posibles áreas de conservación prioritarias o problemas ambientales que requieran atención.

Índice

Objetivos	1
Objetivo General	1
Objetivos Específicos	1
Análisis Descriptivo	3
Tratamiento de variables Categóricas	7
Visualizaciones	9
¿Qué tipo de cultivo se da por periodo del año?	9
Tendencias anuales o semestrales en la producción	12
Conclusiones	13
Bibliografía	15

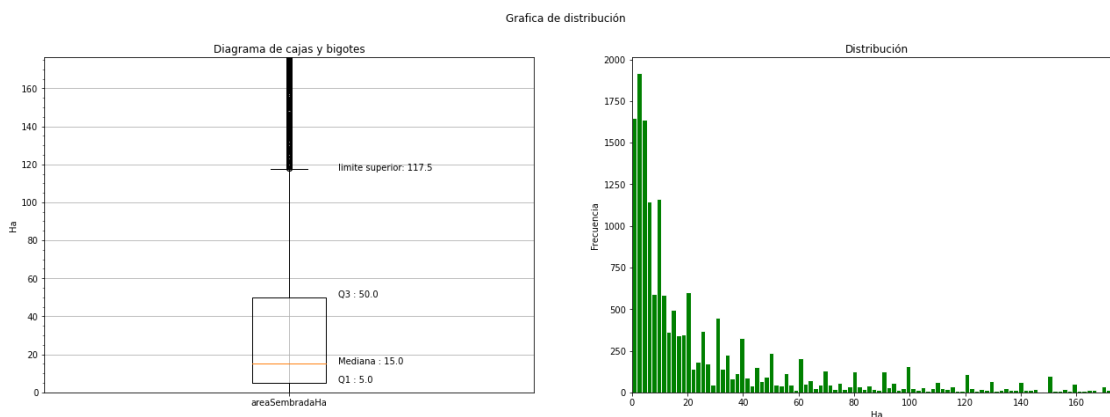
Análisis Descriptivo

A continuación, se expondrán los pasos considerados en el *Análisis descriptivo para las Evaluaciones agropecuarias de los municipios de Cundinamarca y Boyacá*. El objetivo principal es comprender la distribución de los datos y explorar posibles relaciones.

Partimos de un conjunto de datos previamente documentado y con formatos adecuados. Al analizar las variables numéricas, como ‘areaSembradaHa, areaCosechadaHa, produccionTon y rendimientoTonHa’, observamos que presentan una distribución sesgada hacia la izquierda. Esto se evidenció al construir un diagrama de cajas y bigotes para identificar posibles valores atípicos.

Figura 1

Distribución de los datos de ‘areaSembradaHa’



Como primera medida correctiva para los datos atípicos, se determinó la cantidad de datos típicos considerando aquellos que estuvieran por encima del límite superior del bigote. Se calculó el porcentaje de estos datos en la muestra, y dado que corresponde a un bajo porcentaje (13.8%), se decidió eliminar estos datos del conjunto de datos.

Como forma de validar este cambio, se elaboró un mapa de correlación entre las variables con el fin de observar si la correlación entre ellas había mejorado o no. Esto se puede ver en la figura 5 para más detalles.

Dicha técnica utilizada anteriormente no fue muy óptima ya que se ha perdido correlación entre las variables.

Figura 2

Distribución de los datos de 'areaCosechadaHa'

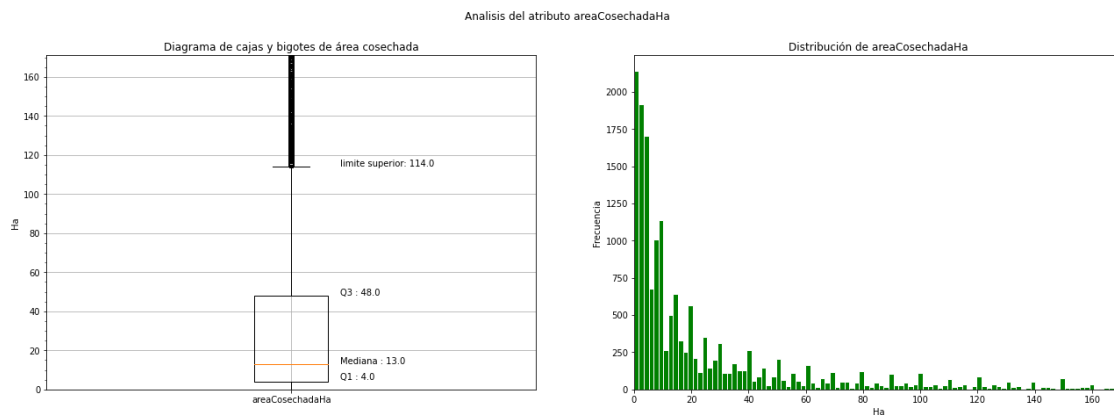


Figura 3

Distribución de los datos de 'produccionTon'

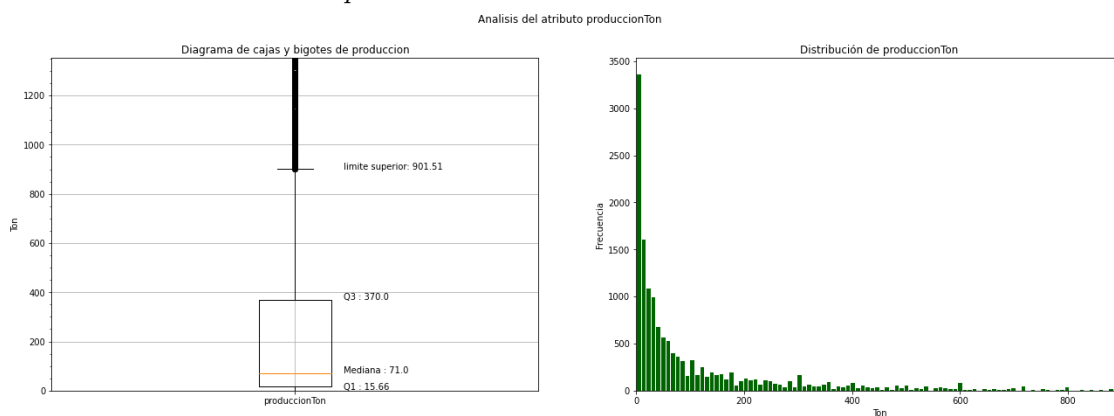


Figura 4

Distribución de los datos de 'rendimientoTonHa'

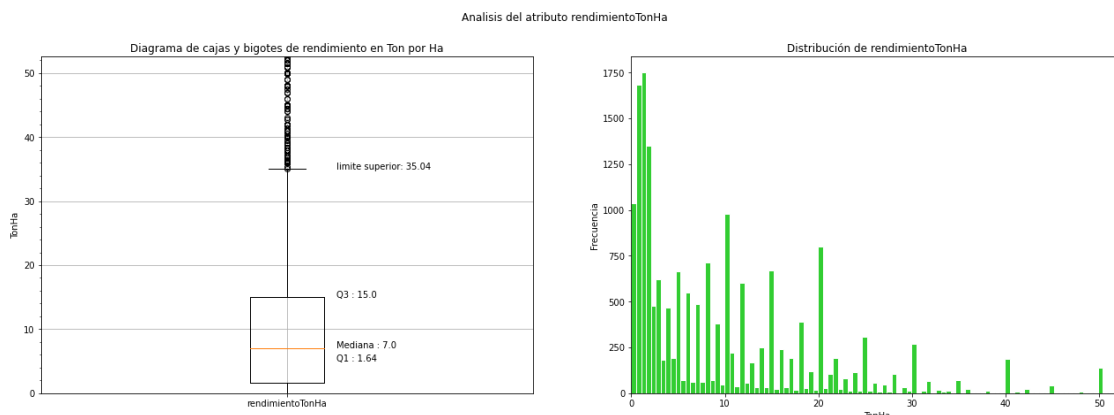
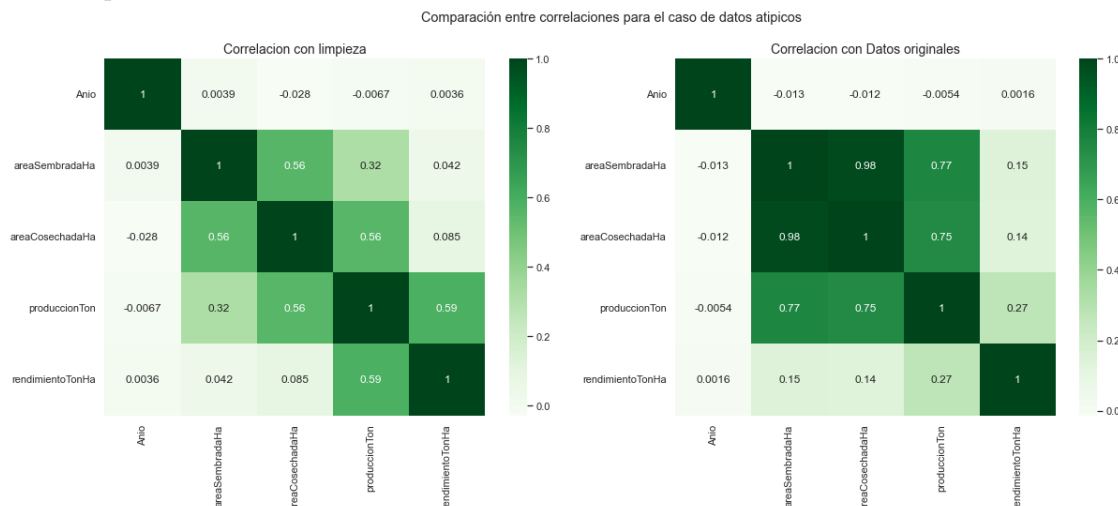


Figura 5

Gráfico comparativo entre Tablas de correlación Numéricas



Note. Las correlaciones son generalmente más bajas. Por ejemplo, la correlación entre *areaSembradaHa* y *produccionTon* es 0.32, mientras que en los datos originales es 0.77.

Las correlaciones son más altas y más consistentes. La correlación entre *areaSembradaHa* y *areaCosechadaHa* es 0.98, indicando una fuerte relación directa.

En la gráfica de correlación entre los datos originales y datos limpios el porcentaje dado entre las variables *produccionTon* y *rendimientoTonHa* se evidencia una mayor correlación en los datos que fueron previamente tratados.

Considerando los resultado de las tablas de correlaciones, el autor nos describe lo siguiente: “ La correlación no siempre indica causalidad. Incluso si la correlación disminuye, no necesariamente significa que los datos no eran atípicos”, y entendiendo lo que dice el autor consideramos que las variables pueden ser tratadas con otros métodos para identificación de valores atípicos <https://es.statisticseasily.com/detecci%C3%B3n-y-tratamiento-de-valores-at%C3%ADpicos/>.

****Es importante evaluar el contexto y comprender la naturaleza de los datos antes de sacar conclusiones definitivas****

De este análisis, podemos extraer dos conclusiones importantes. En primer lugar, debemos abordar los datos atípicos desde una perspectiva más amplia que involucre el contexto en el que se generaron. En segundo lugar, es crucial reconocer que algunos de los datos que inicialmente consideramos atípicos no lo son realmente; de hecho, desempeñan un papel

fundamental en nuestro modelo.

Basándonos en el contexto de los datos, encontramos una relación lógica entre dos variables: $areaSembradaHa$ y $areaCosechadaHa$, (donde Ha corresponde a hectárea). La definición de estas variables nos indica que el área sembrada corresponde al terreno inicial donde se planta un cultivo específico, mientras que el área cosechada representa la porción de ese terreno que ha sido productiva y ha dado lugar a la cosecha.

Por lo tanto, es fundamental que se cumpla la siguiente relación

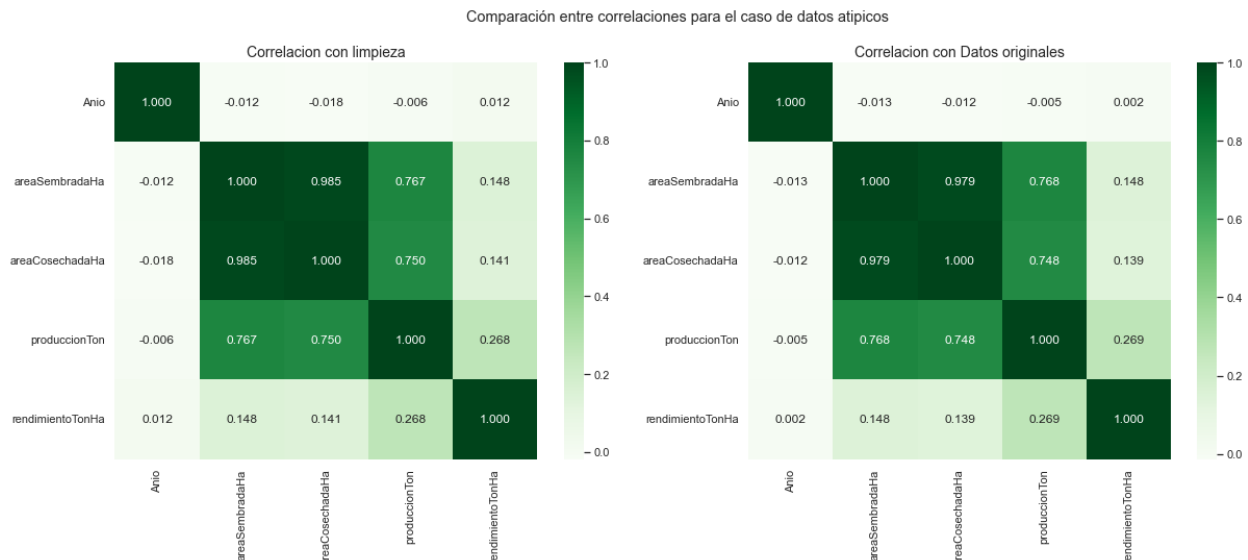
$$areaSembradaHa \geq areaCosechadaHa$$

si esta desigualdad se incumple ese registro se considerara como un valor atípico, ya que esto sería incoherente desde una perspectiva agrícola.

Después de aplicar el filtro correspondiente, observamos una mejora en la correlación entre las variables. Los datos eliminados, que no cumplieran con la condición establecida, representaron solo el 3.2 % del total.

Figura 6

Gráfico comparativo entre Tablas de correlación Numéricas Método 2



Al establecer una comparación entre los gráficos de mapa de calor 6 es evidente que

dicho proceso de limpieza unido a las relaciones lógicas entre las variables, son una propuesta útil para el tratamiento de datos atípicos.

Tratamiento de variables Categóricas

Con el objetivo de identificar variables útiles para la construcción de un modelo de predicción, se implementó un método de codificación mediante una matriz de 0 y 1. Este método se conoce como ***One Hot Encoding***.

One Hot Encoding es un proceso utilizado en el aprendizaje automático para convertir variables categóricas en una forma que pueda ser utilizada por algoritmos de predicción.

</> Definición:

One Hot Encoding (también conocido como one-of-K encoding) es una técnica que transforma variables categóricas en una representación binaria.

</> Cómo funciona:

Para cada categoría en una variable categórica, se crea una columna binaria (bit) en la matriz de datos. Si un dato pertenece a una categoría específica, se coloca un 1 en la columna correspondiente y 0 en las demás columnas.

</> Caso de estudio:

En nuestro caso de estudio, al analizar las gráficas de barras, seleccionamos los atributos categóricos que consideramos útiles para el modelo. Además, nos aseguramos de que estos atributos tuvieran pocos valores únicos para evitar la alta dimensionalidad en el conjunto de datos resultante.

Estas variables fueron [*‘Periodo’*, *‘cicloDelCultivo’*, *‘grupoCultivo’*, *‘Dpto’*, *‘estadoFisicoCultivo’*]

Una vez aplicado este método se elabora una tabla de correlación donde se incluye estas nuevas variables y las variables numéricas iniciales.

Figura 7

Tabla de correlación para nuevas columnas de periodo y ciclo de cultivo.

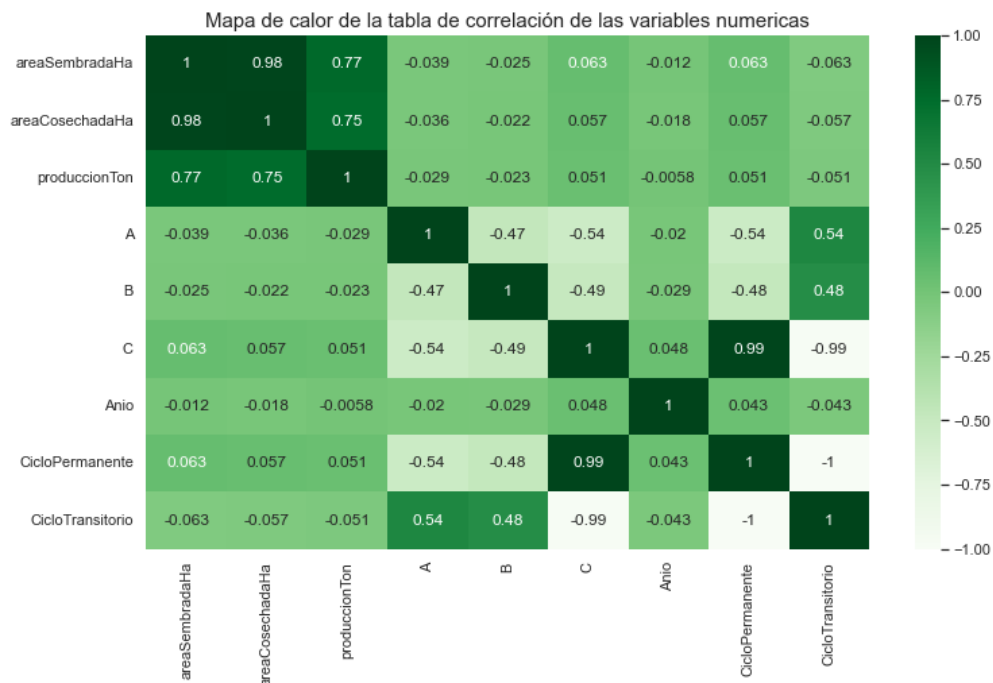


Figura 8

Tabla de correlación para las nuevas variables de grupo de cultivo

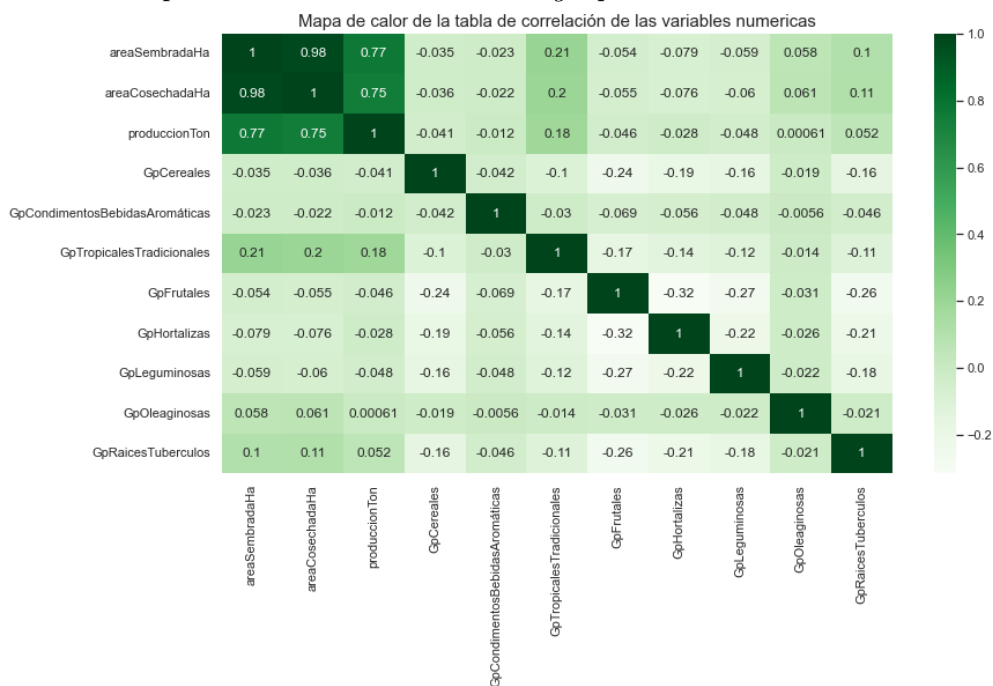
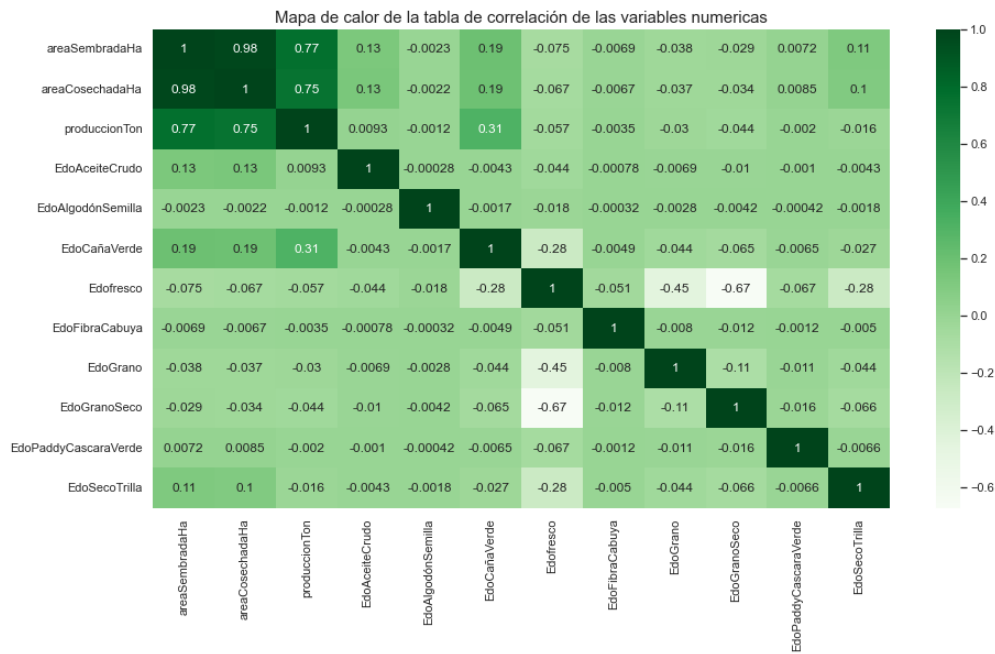


Figura 9

Tabla de correlación para las nuevas variables del estado final del cultivo



Visualizaciones

A partir de las visualizaciones, se busca responder algunas preguntas con el objetivo de comprender completamente la naturaleza de los datos, tales como

- </> ¿Cuál es la distribución de áreas sembradas y cosechadas por año y departamento?
- </> ¿Cuáles cultivos tienen el mayor rendimiento?
- </> ¿Existen tendencias anuales o semestrales en la producción?

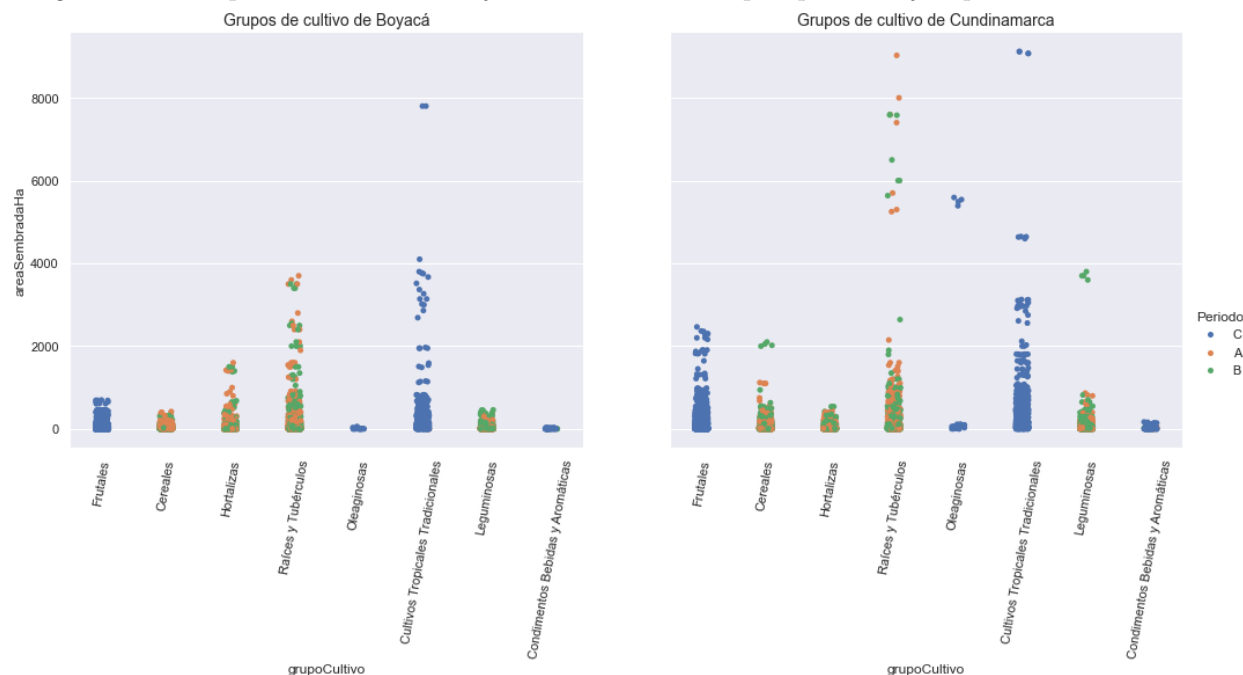
¿Qué tipo de cultivo se da por periodo del año?

Se elaboro un diagrama de dispersión según el grupo del cultivo por periodo del año y discretizado por departamento, en relación a la cantidad de Hectáreas sembradas (*areaSembradaHa*).

Al analizar la gráfica anterior 10 , se observan posibles tendencias en los grupos de cultivos tropicales tradicionales. Estos cultivos parecen ser sembrados de manera continua

Figura 10

Diagrama de dispersión de cultivo y areaSembradaHa por periodo y departamento



a lo largo de ciclos permanentes en ambos departamentos. Por otro lado, se evidencia que los cultivos de oleaginosas, condimentos, bebidas y aromáticas presentan un bajo número de áreas sembradas para ambos departamentos.

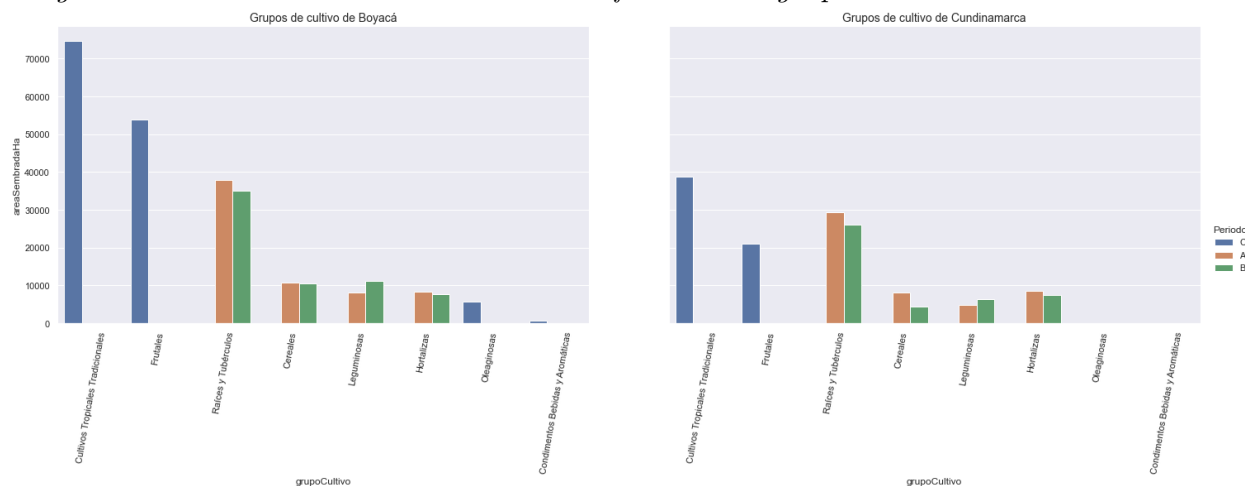
Se tiene la intención de crear un diagrama de barras que muestre la suma del área sembrada en función del grupo de cultivo, considerando el período del año y el departamento. El objetivo es observar cuáles son los cultivos más relevantes en términos de área sembrada. Ver figura 11. Además, se ha realizado un proceso similar para la producción y el área cosechada.

Al analizar la gráfica (Figura 11), se observan patrones de tendencia en la cantidad de área sembrada por grupo de cultivo. Es notable que la mayor cantidad de hectáreas sembradas se encuentra en el departamento de Boyacá.

Además, debido a la fuerte correlación entre las variables *areaSembradaHa* y *areaCosechadaHa*, se encontró un comportamiento similar, como se describe en la Figura 12.

Figura 11

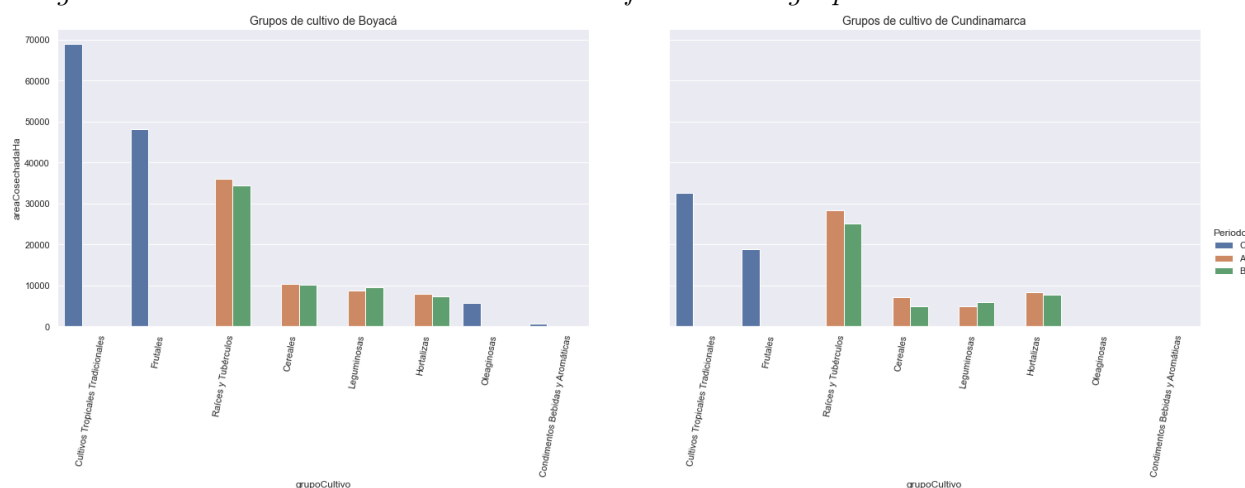
Diagrama de barras del areaSembradaHa en función del grupoCultivo



Note. Los datos visualizados corresponde a una suma agrupada del área sembrada según el grupo de cultivo, el año ,el periodo y el departamento.

Figura 12

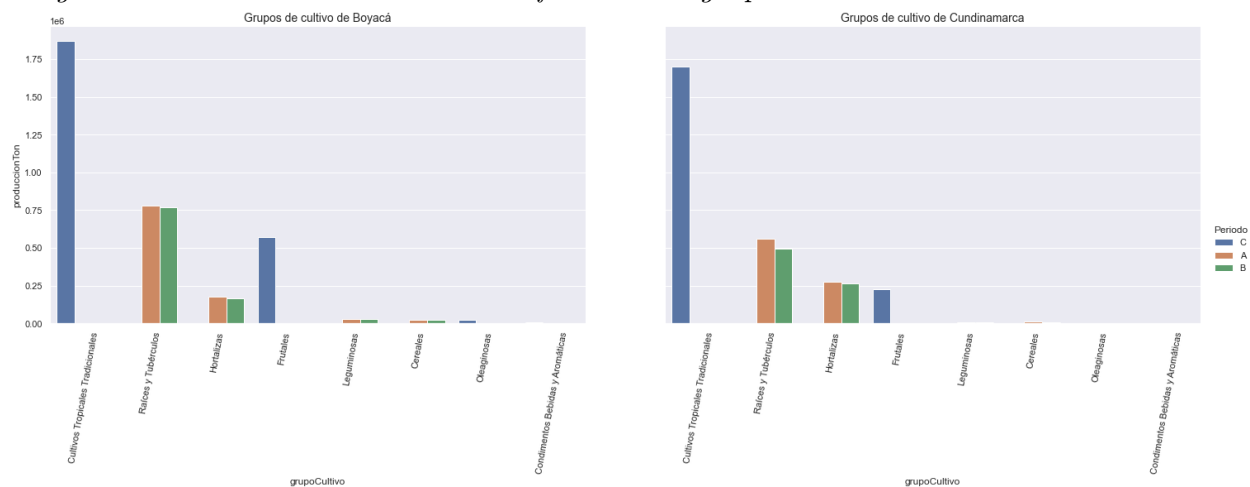
Diagrama de barras del areaCosechadaHa en función del grupoCultivo



Note. Los datos visualizados corresponde a una suma agrupada del área cosechada según el grupo de cultivo, el año ,el periodo y el departamento.

Figura 13

Diagrama de barras del Producción en función del grupoCultivo



Note. Los datos visualizados corresponde a una suma agrupada del la producción según el grupo de cultivo, el año ,el periodo y el departamento.

Tendencias anuales o semestrales en la producción

A continuación se presenta un diagrama de barras de datos agrupados bajo el criterio de la suma para los distintos tipos de cultivo en el año, el cual muestra la producción en toneladas agrupada.

Para una mejor visualización de la tendencia de la producción según el tipo de cultivo por año se elaboró un mapa de calor.

En Boyacá, la producción de Cultivos Tropicales Tradicionales ha aumentado de 1.422 MTon en 2019 a 1.914 MTon en 2022, mientras que raíces y tubérculos han disminuido ligeramente. La producción de hortalizas ha mostrado un crecimiento gradual. En Cundinamarca, los cultivos tropicales han mantenido una producción alta pero estable, mientras que raíces y tubérculos han disminuido. Frutales han crecido significativamente y hortalizas se han mantenido constantes. Otros cultivos muestran una producción baja y estable en ambos departamentos.

Figura 14

Producción en toneladas en función del grupo de cultivo en los años

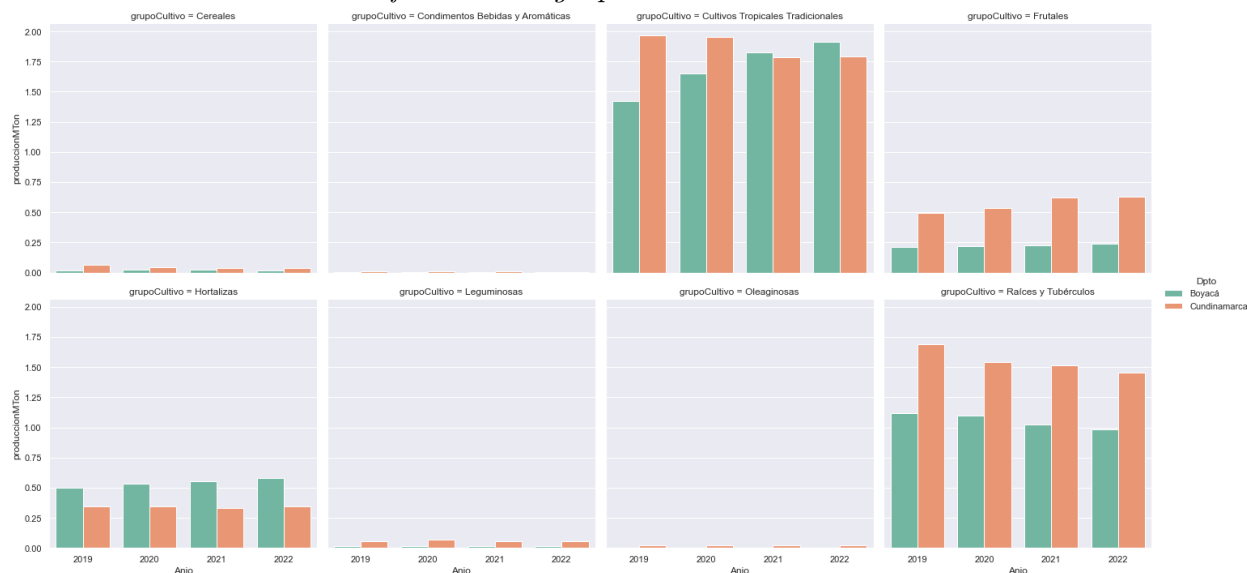
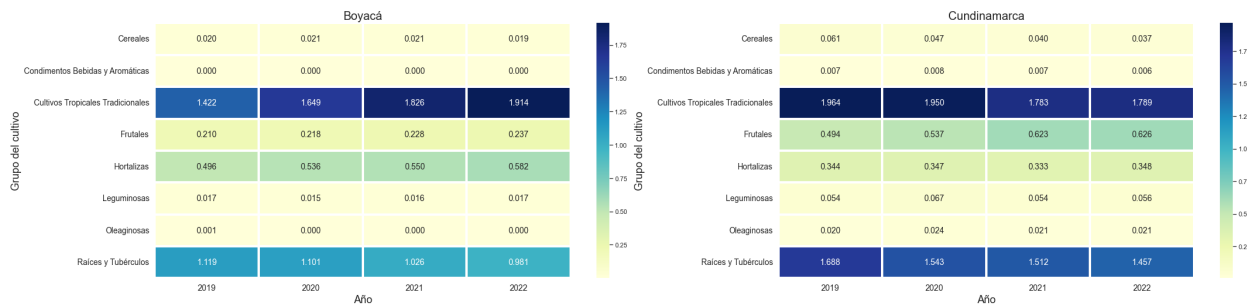


Figura 15

Mapa de calor de producción por Mton (mega toneladas) en función de grupo de cultivo y el año



Conclusiones

- Los datos atípicos introducen ruido en los datos, lo que lleva a una subestimación o sobreestimación de las correlaciones verdaderas. Este ruido puede ser perjudicial para cualquier análisis o modelo predictivo, ya que puede llevar a conclusiones incorrectas o poco fiables.
- Se identificaron y eliminaron valores atípicos que no cumplieran con la condición $areaSembradaHa \geq areaCosechadaHa$, mejorando así la calidad del dataset. Sin

embargo, es importante considerar el contexto de los datos antes de eliminar valores, ya que algunos datos atípicos pueden ser significativos y no simplemente errores.

- Las correlaciones en los datos originales son más coherentes con la expectativa. Por ejemplo, *areaSembradaHa* y *areaCosechadaHa* deben estar altamente correlacionadas (0.98 en datos originales), lo cual es lógico, ya que las áreas sembradas y cosechadas están intrínsecamente relacionadas.
- Las visualizaciones muestran que la distribución de áreas sembradas y cosechadas varía significativamente entre los municipios y a lo largo de los años, lo que puede estar influenciado por factores climáticos, económicos y de gestión
- En ambos departamentos, los “Cultivos Tropicales Tradicionales” tienen la mayor área sembrada, especialmente en Boyacá, donde alcanzan más de 70,000 hectáreas en el periodo C. Esto indica que estos cultivos son fundamentales para la economía agrícola de la región
- Boyacá muestra una mayor diversidad en términos de área sembrada en comparación con Cundinamarca. Además de los “Cultivos Tropicales Tradicionales”, se observa una considerable área sembrada de “Frutales” y “Raíces y Tubérculos”, con áreas sembradas que superan las 30,000 hectáreas.
- Otros grupos de cultivos, como Cereales, Leguminosas, Hortalizas, Oleaginosas y Condimentos Bebidas y Aromáticas, tienen áreas sembradas relativamente pequeñas en ambos departamentos. Esto podría indicar que estos cultivos son menos prioritarios o tienen menor demanda en estas regiones.
- La gráfica 11 compara las áreas cosechadas en Boyacá y Cundinamarca entre 2019 y 2022, clasificadas por grupos de cultivos y períodos de tiempo. En Boyacá, los Cultivos Tropicales Tradicionales dominan la producción anual, superando significativamente a frutales y raíces/tubérculos. Los frutales, aunque en segundo lugar anual, disminuyen

notablemente en los períodos semestrales, mientras que raíces y tubérculos mantienen una producción equilibrada durante ambos semestres. Otros cultivos como cereales y leguminosas tienen una producción mucho menor.

- En Cundinamarca, los Cultivos Tropicales Tradicionales también son predominantes, aunque con una diferencia menos marcada. La producción de frutales es considerable, especialmente anualmente, pero inferior a los cultivos tropicales. Raíces y tubérculos tienen una alta producción en el primer semestre, seguida por el segundo, mostrando consistencia a lo largo del año. Otros cultivos presentan una baja producción, similar a la observada en Boyacá.
- En resumen, los Cultivos Tropicales Tradicionales y raíces/tubérculos son los más productivos, aunque Boyacá muestra una tendencia de crecimiento más notable en los cultivos tropicales. Los frutales y hortalizas presentan un crecimiento positivo, sugiriendo un potencial de expansión. La disminución en la producción de raíces y tubérculos es preocupante y requiere investigación. Se recomienda diversificar cultivos y mejorar la producción de los grupos menos predominantes para fortalecer la resiliencia agrícola.

Bibliografía

- <https://agronet.gov.co/estadistica/Paginas/home.aspx?cod=1>
- <https://www.agronet.gov.co/estadistica/Paginas/home.aspx?cod=59>
- <https://es.statisticseasily.com/detecci%C3%B3n-y-tratamiento-de-valores-at%C3%ADpicos/>