

DATOS DE PRODUCCION DE CULTIVOS INDIA

Leslie Johana Galviz

Cristian Machado

2024

****Docente:** Yennyfer Portilla Yela**

Universidad del Valle - Facultad de Ingeniería

Probabilidad y Estadística

Santiago de Cali

Tabla de Contenido

- Introducción
- Objetivos
- Informe sobre Variables de Datos de Cultivo en India
- Métodos
- Resultados De La Investigación
- Conclusiones
- Referencias

Introducción

En un mundo donde la seguridad alimentaria y la sostenibilidad agrícola son cada vez más cruciales, comprender las dinámicas de la producción de cultivos se convierte en un pilar fundamental para la formulación de políticas agrícolas y la toma de decisiones estratégicas en el sector agroalimentario.

India es uno de los grandes productores de alimentos a nivel mundial. Tiene cultivos clave, como arroz, trigo, caña de azúcar, algodón, frutas y verduras. Específicamente, India produce alrededor del 10.9% de las frutas y el 8.6% de las verduras a nivel mundial. También es el mayor productor de legumbres, cubriendo aproximadamente el 25% de la producción global y satisfaciendo el 27% de la demanda mundial de esta importante fuente de proteínas. Además, India lidera la producción de leche y contribuye significativamente a la producción de especias, té y ganado.

Este estudio se centra en un extenso conjunto de datos sobre la producción de cultivos en India, abarcando más de 246,000 registros a lo largo de varios años. La información incluye diversos tipos de cultivos, métricas de producción y datos regionales que reflejan la compleja realidad agrícola del país. Analizaremos las tendencias temporales y regionales en la producción de cultivos, identificando patrones significativos y variaciones regionales que influyen en la eficiencia y productividad agrícola.

Objetivos

Objetivo Principal

Desarrollar una metodología de análisis predictivo de producción agrícola en India, integrando datos históricos y factores externos relevantes, con el propósito de anticipar los niveles de producción a mediano plazo y generar información útil para la toma de decisiones en la gestión de riesgos en el sector agrícola.

Objetivos Específicos

- Recopilar y analizar datos históricos de producción agrícola en India, así como identificar los factores externos (clima, políticas, mercado) que impactan en el rendimiento de los cultivos.
- Aplicar técnicas de análisis estadístico y de series de tiempo para establecer patrones y tendencias en los datos de producción, identificando los principales indicadores que afectan su comportamiento.
- Diseñar y evaluar una metodología que permita generar predicciones sobre la producción agrícola a mediano plazo, y elaborar recomendaciones estratégicas para la gestión de riesgos en el sector agrícola en función de estos resultados.

Informe sobre Variables de Datos de Cultivo en India

1. State_Name (Nombre del Estado)

- **Descripción:** Representa el nombre del estado de la India donde se registraron los datos del cultivo.
- **Importancia:** Permite analizar variaciones en la producción agrícola entre diferentes estados, ayudando a identificar áreas con mayor potencial de producción y aquellas que podrían requerir apoyo adicional.
- **Clasificación:** Cualitativa.

2. District_Name (Nombre del Distrito)

- **Descripción:** Especifica el distrito dentro del estado donde se recopilieron los datos del cultivo.
- **Importancia:** Facilita un análisis más detallado a nivel regional, permitiendo la identificación de patrones locales y el diseño de políticas específicas para mejorar la producción en distritos con bajo rendimiento.
- **Clasificación:** Cualitativa.

3. Crop_Year (Año de Cultivo)

- **Descripción:** Indica el año en el que se cosechó el cultivo.
- **Importancia:** Es esencial para el análisis de tendencias temporales en la producción agrícola, permitiendo comparar datos de diferentes años y evaluar el impacto de factores externos como cambios climáticos o políticas agrícolas.
- **Clasificación:** Cuantitativa.

4. Temporada (Temporada Agrícola)

- **Descripción:** Indica la temporada agrícola (por ejemplo, Kharif, Rabi) durante la cual se cultivó el cultivo.
- **Importancia:** Comprender las temporadas de cultivo ayuda a anticipar períodos de producción y escasez, así como a planificar mejor el uso de recursos hídricos y otros insumos agrícolas.
- **Clasificación:** Cualitativa.

5. Cultivo (Tipo de Cultivo)

- **Descripción:** Identifica el tipo de cultivo que se cultivó (por ejemplo, arroz, trigo, maíz).
- **Importancia:** Permite segmentar la producción agrícola y estudiar las preferencias del mercado, así como las necesidades de los agricultores en función de los cultivos específicos.
- **Clasificación:** Cualitativa.

6. Área (Superficie)

- **Descripción:** Representa la superficie total de tierra (en hectáreas o acres) utilizada para el cultivo.
- **Importancia:** Proporciona datos sobre la escala de producción y la disponibilidad de tierras para cultivos, lo que es crucial para estimar la capacidad de producción y la eficiencia del uso del suelo.
- **Clasificación:** Cuantitativa.

7. Producción (Cantidad Total)

- **Descripción:** Indica la cantidad total (en toneladas métricas) del cultivo producido en el área especificada.
- **Importancia:** Esta variable es clave para el análisis de la productividad agrícola y la planificación de la oferta en el mercado. Ayuda a prever el comportamiento futuro de la producción y su impacto en la economía global.
- **Clasificación:** Cuantitativa.

Métodos

La limpieza de datos es un paso esencial para garantizar la calidad y fiabilidad de los análisis realizados. En este estudio, se busca optimizar el conjunto de datos sobre la producción de cultivos en India para obtener conclusiones más precisas.

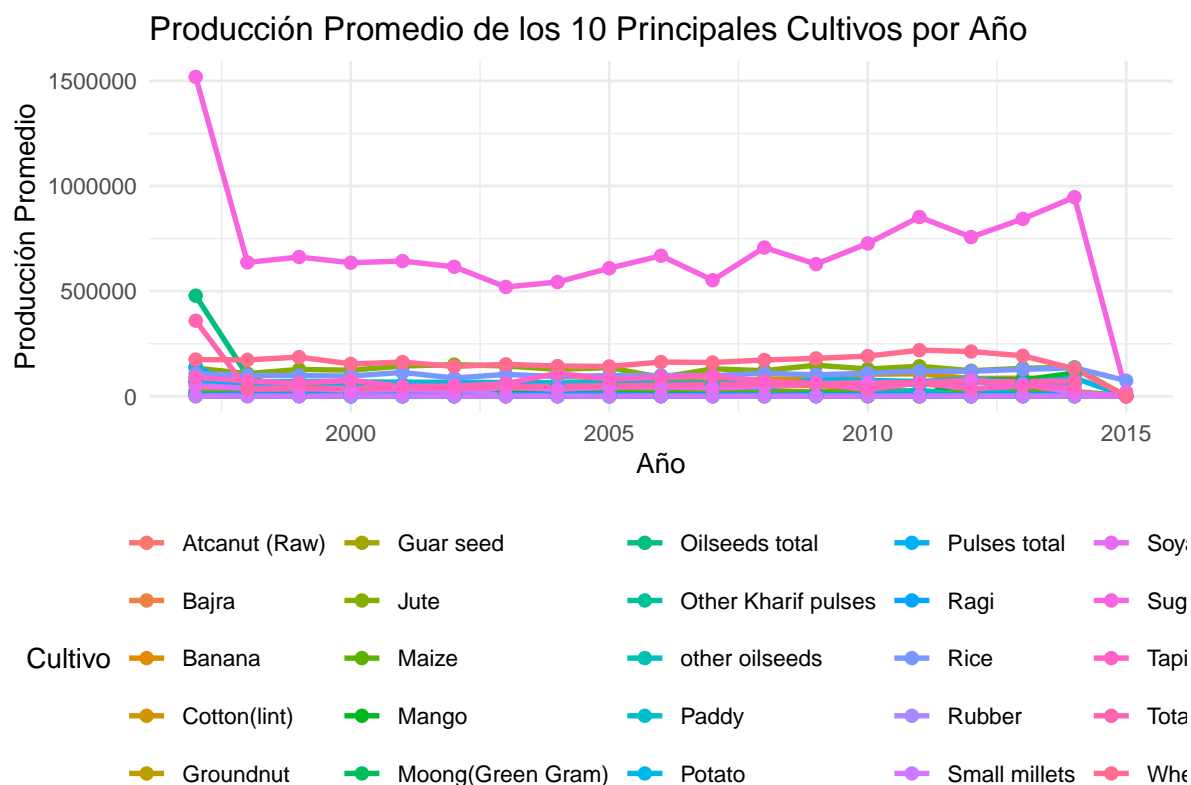
Proceso de Limpieza

1. **Carga de Datos:** Se cargaron múltiples archivos CSV desde un directorio y se combinaron en un único conjunto de datos.
2. **Separación de Columnas:** Inicialmente, todos los datos estaban en una sola columna. Se separaron en varias columnas representativas como: Estado, Distrito, Año de Cultivo, Temporada, Cultivo, Área y Producción.
3. **Detección de Datos Faltantes:** Se realizó un análisis de los datos faltantes, identificando registros incompletos mediante un mapa de datos faltantes.
4. **Conversión de Tipos de Datos:** Las columnas se convirtieron a sus tipos adecuados, asegurando que los valores numéricos se trataran correctamente para análisis futuros.
5. **Eliminación de Datos Faltantes y Duplicados:** Se eliminaron filas con datos faltantes y registros duplicados para mantener la integridad del conjunto de datos.
6. **Identificación y Eliminación de Valores Atípicos:** Se revisaron los datos en busca de valores atípicos utilizando gráficos, y se eliminaron aquellos que no cumplían con criterios razonables.
7. **Normalización de Texto y Filtrado:** Se normalizó el texto de las columnas para estandarizar la nomenclatura y se filtraron registros no válidos, asegurando que solo se mantuvieran entradas con valores positivos y años dentro de un rango razonable.

Proceso de Procesamiento de Datos

Ahora contamos con un conjunto de datos más preciso, lo que nos permite incorporar nuevas variables para la formulación de conclusiones. Utilizando diversas técnicas estadísticas, podremos llegar a conclusiones de alta calidad, minimizando la dispersión de datos y garantizando una interpretación más clara y relevante para poder culminar el objetivo.

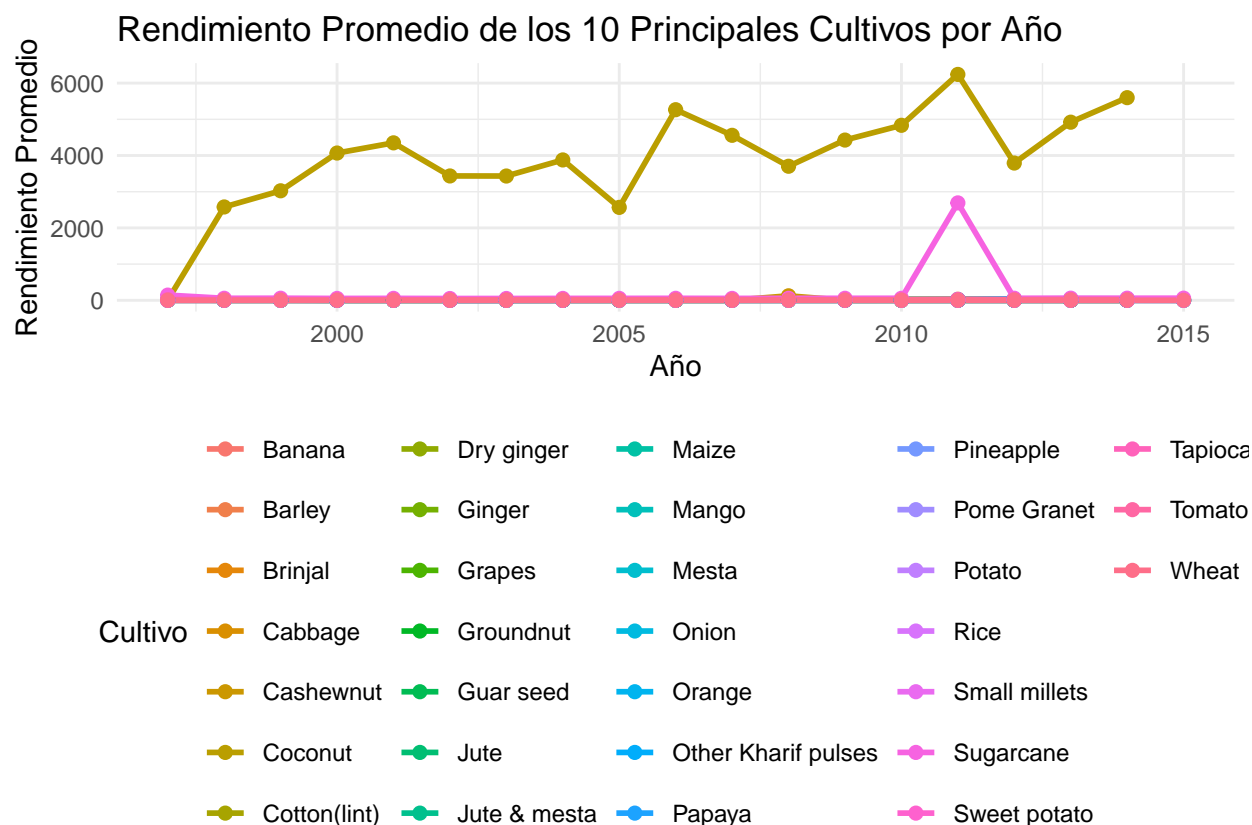
Resultados De La Investigación



En este gráfico se observa la producción promedio anual de los principales cultivos en India desde el año 2000 hasta 2015. La caña de azúcar destaca como uno de los cultivos de mayor producción en la última década. Aunque se observa una caída pronunciada de su producción en el inicio del período estudiado, mantuvo una tendencia estable en los años posteriores, incluso logrando incrementos en ciertos puntos. Este desempeño ha permitido a India consolidarse entre los cinco principales productores de caña de azúcar a nivel mundial.

El trigo y el arroz también figuran como cultivos esenciales, con niveles de producción constantes a lo largo del tiempo, contribuyendo significativamente a la producción de cereales en el país. La producción de trigo en India ha permitido que en 2022 el país se posicione como el décimo exportador más grande de trigo y otros cereales mezclados, alcanzando exportaciones por un valor de \$2.32 mil millones.

Cabe mencionar que, para mejorar la claridad del análisis, se excluyó el cultivo de coco, ya que su alta producción distorsionaba los datos, dificultando una interpretación detallada y equilibrada de los otros cultivos incluidos en el estudio.

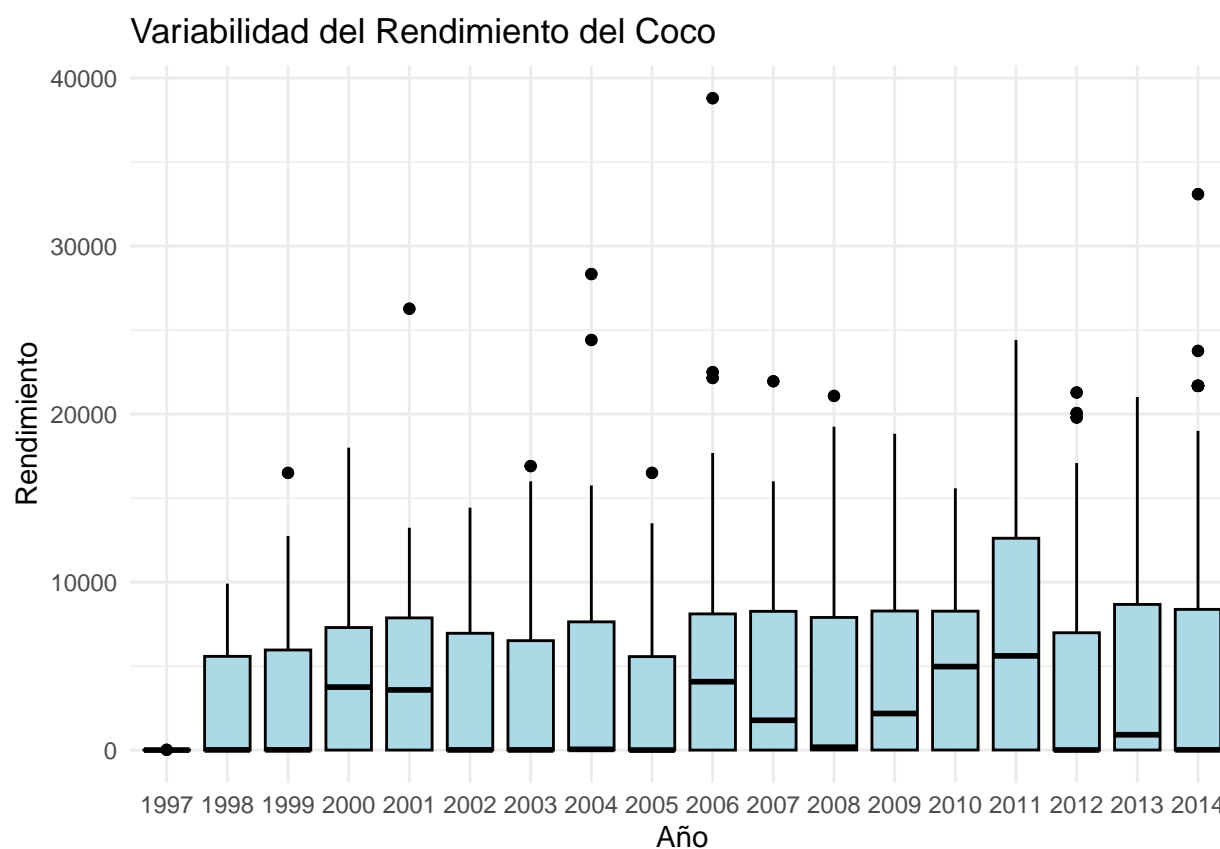


Este gráfico muestra el “Rendimiento Promedio de los 10 Principales Cultivos por Año” en India, donde el rendimiento se calcula como la relación entre el área cultivada y la producción total (rendimiento = producción / área). Esta métrica nos da una idea de la productividad de cada cultivo, es decir, qué tan eficiente es cada cultivo en términos de producción por unidad de área.

En contraste, otros cultivos muestran rendimientos considerablemente más bajos en comparación con el coco, lo que destaca la especialización de India en su producción, apoyada en condiciones agroclimáticas favorables y técnicas de cultivo eficientes. India es, de hecho, el mayor productor de coco a nivel mundial, con una producción de 19.247 millones de frutos en 2021-22, lo que representa el 31,45% de la producción global. Esta capacidad productiva le permite mantener un mercado interno robusto y, además, comenzar a ganar terreno en mercados internacionales.

Es importante señalar que es probable que los datos consideren cada coco como una unidad individual, mientras que otros cultivos, como el trigo, podrían medirse en bultos o toneladas. Esta diferencia en las unidades de medición podría contribuir a la alta cifra reportada para el coco, lo que sugiere una posible discrepancia en el volumen aparente de producción en comparación con otros cultivos.

Análisis de la Variabilidad del Rendimiento del coco

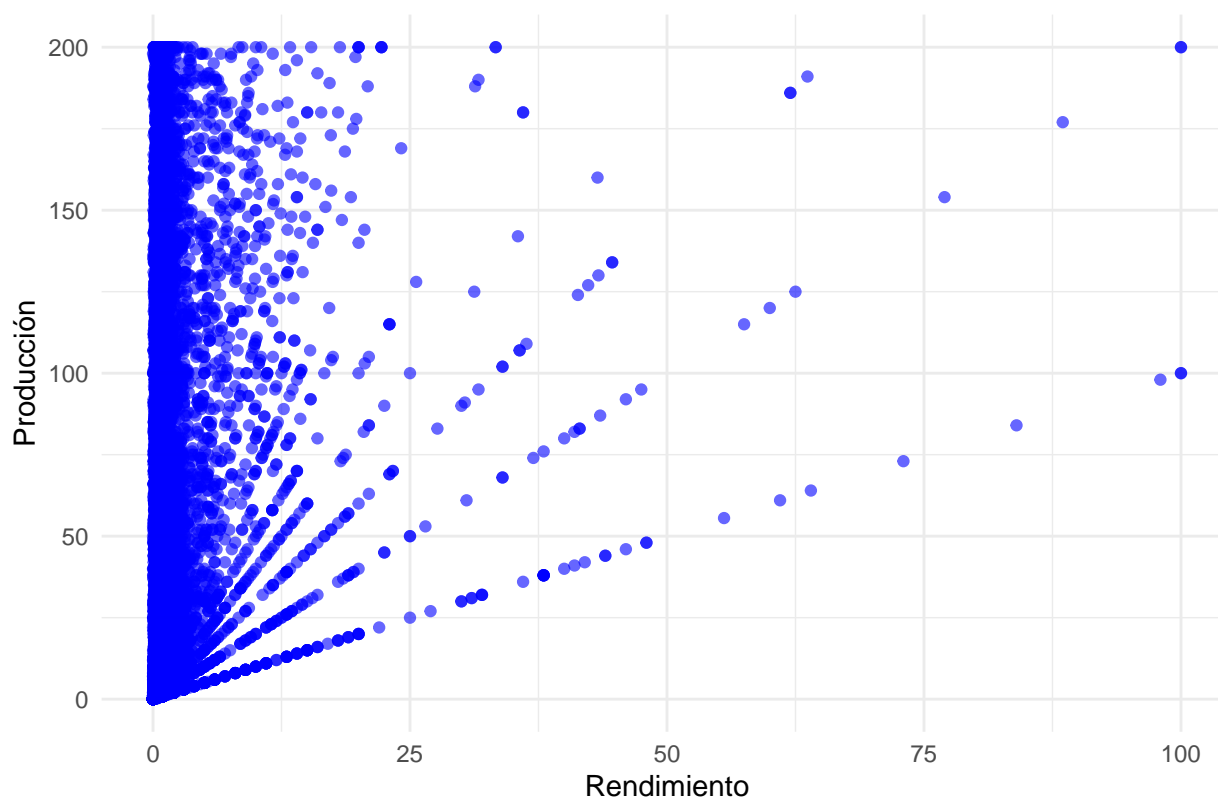


El gráfico de caja muestra la distribución del rendimiento del coco para cada año. La caja de cada año representa la mediana, el rango intercuartílico y los valores atípicos, los cuales corresponden a puntos de rendimiento inusualmente altos o bajos.

- Año 2006: La mediana del rendimiento está notablemente por debajo de las 10,000 unidades, indicando que la mayoría de los rendimientos fueron bajos. Sin embargo, hay varios valores atípicos que alcanzan cerca de 40,000 unidades, lo que sugiere que algunos cultivos o regiones tuvieron un rendimiento excepcionalmente alto. La distribución es aproximadamente simétrica, mostrando un equilibrio en ambos lados de la mediana.
- Año 2011: Este año presenta un rendimiento promedio alto, lo que indica que, en general, la producción de coco fue buena. Sin embargo, la alta variabilidad en los rendimientos, junto con una asimetría positiva, sugiere que algunos rendimientos fueron mucho más altos que otros, lo que puede reflejar una mezcla de factores que afectaron la consistencia en los resultados de la cosecha..
- Año 2014: Este año presenta una gran variabilidad en los rendimientos, con la mayor parte de los

valores en niveles altos. La distribución es asimétrica y muestra una tendencia hacia rendimientos elevados. Podemos observar que el 75% de los valores de rendimiento alcanzan o superan los 8,378.576, lo que sugiere que hubo algunos meses con una producción notablemente alta de coco. La desviación estándar para este año es de 7,835.711, lo cual indica que los datos están considerablemente dispersos alrededor de la media, reflejando una variabilidad significativa en la producción mensual.

Relación entre Producción y Rendimiento del Coco



En este análisis observamos una gran concentración de datos cercanos a 0 en el eje X, lo que indica un rendimiento bajo independientemente de la producción del cultivo. A pesar de este rendimiento bajo, la producción presenta una mayor variabilidad, alcanzando valores significativamente altos en algunos casos.

Es importante mencionar que esta gráfica muestra solo una parte de la totalidad de los datos debido a la gran cantidad de observaciones, por lo que estamos viendo una muestra limitada para facilitar el análisis visual.

Conclusion

La investigación sobre la producción de cultivos en India muestra una perspectiva compleja y dinámica de la agricultura en uno de los principales productores agrícolas del mundo. Los datos indican que cultivos como la caña de azúcar, el trigo, el arroz y el coco destacan por su alta producción y su impacto económico, tanto a nivel nacional como en mercados globales.

La caña de azúcar ha demostrado ser un cultivo fundamental en la agricultura india, con una producción estable y una importancia económica considerable. Esta estabilidad en la producción posiciona a India entre los cinco principales productores de caña de azúcar a nivel mundial, con un sector que sigue creciendo y adaptándose a las demandas tanto internas como de exportación. Sin embargo, este crecimiento presenta desafíos, ya que la industria enfrenta problemas de sostenibilidad y bienestar de los trabajadores, lo que sugiere la necesidad de políticas que equilibren el éxito económico con el desarrollo social y la equidad laboral.

En cuanto al coco, India se destaca como el mayor productor mundial, con un rendimiento notablemente alto en comparación con otros cultivos. Esto se debe a condiciones agroclimáticas favorables y a técnicas de cultivo optimizadas. Sin embargo, el análisis también revela posibles inconsistencias en la interpretación de los datos de producción, ya que las unidades de medición pueden variar entre cultivos, afectando la percepción de volumen de producción.

En conclusión, India posee una capacidad agrícola robusta y diversificada, con cultivos que aportan significativamente a la economía y a la seguridad alimentaria mundial. No obstante, el país enfrenta desafíos que requieren una gestión sostenible de los recursos y políticas inclusivas para asegurar que el sector agrícola pueda seguir siendo un pilar de crecimiento y estabilidad en el largo plazo.

Referencias

- “Crop Production Data India” en Kaggle. <https://www.kaggle.com/datasets/iamtapendu/crop-production-data-india>
- “Sugarcane Area Poised to Rise by 8-10% in 2011-12” en Business Standard. https://www.business-standard.com/article/markets/sugarcane-area-poised-to-rise-by-8-10-in-2011-12-111052600014_1.html
- “Sugar Annual Report - New Delhi, India” del USDA Foreign Agricultural Service. https://apps.fas.usda.gov/newgainapi/api/report/downloadreportbyfilename?filename=Sugar%20Annual_New%20Delhi_India_4-14-2011.pdf
- “India: El éxito de la industria de la caña de azúcar sacrifica el bienestar de los trabajadores” en France24. <https://www.france24.com/es/programas/en-foco/20230707-india-el-%C3%A9xito-de-la-industria-de-la-ca%C3%B1a-de-az%C3%BAcar-sacrifica-el-bienestar-de-los-trabajadores>
- “Top 10 de países productores de caña de azúcar en 2023-2024” en Tecnicana. <https://tecnicana.org/2024/04/26/mercados/top-10-de-paises-productores-de-cana-de-azucar-en-2023-2024/?v=42983b05e2f2>