**India Agriculture Crop Production**

Andres padilla,

Sergio Soler,

Alejandro Rios

Introducción a la ciencia de datos

Santiago Toledo

2024

**Introduccion**

En el contexto de la agricultura y la planificación rural, el análisis de datos se ha vuelto crucial para comprender las complejidades de la producción agrícola y diseñar estrategias efectivas de desarrollo agrícola. La base de datos "Cultivos India" representa un recurso invaluable en este sentido, al ofrecer una amplia gama de información detallada sobre la producción agrícola en India. Con más de 340,000 registros que abarcan múltiples estados, distritos, cultivos y años, esta base de datos proporciona una oportunidad única para explorar y comprender en profundidad la dinámica de la producción agrícola en el país. Desde la estimación de rendimientos por unidad de área hasta la identificación de cultivos dominantes por región y temporada, esta base de datos ofrece insights cruciales para informar políticas agrícolas, mejorar la eficiencia y promover el desarrollo sostenible en el sector agrícola de India.

**Justificacion**

Abordar el tema de la producción agrícola en India a través del análisis de datos es fundamental debido a su importancia económica y su impacto en la seguridad alimentaria y el desarrollo rural. Comprender los patrones de producción agrícola y su evolución a lo largo del tiempo es esencial para diseñar estrategias efectivas de planificación agrícola y promover un crecimiento sostenible en el sector agrícola. La base de datos "Cultivos India" ofrece una oportunidad única para investigar estos fenómenos, ya que proporciona una amplia gama de información detallada sobre la producción agrícola en el país. Al analizar esta base de datos, podemos identificar áreas con altos rendimientos agrícolas, cultivos dominantes en diferentes regiones y temporadas, así como evaluar la eficiencia en la producción agrícola. Estos hallazgos son fundamentales para informar políticas agrícolas, mejorar la distribución de recursos y promover un desarrollo equitativo en las zonas rurales de India. En el contexto del curso de Introducción a la Ciencia de Datos, analizar la base de datos "Cultivos India" proporciona una oportunidad valiosa para aplicar los conceptos teóricos y las técnicas prácticas aprendidas en el curso. La ciencia de datos se centra en la extracción de información significativa a partir de conjuntos de datos complejos, y la base de datos "Cultivos India" presenta un desafío relevante para desarrollar y poner a prueba estas habilidades analíticas en un contexto agrícola real.

**Objetivos**

**Objetivo general**

Utilizar la base de datos "Cultivos India" para comprender en profundidad la dinámica de la producción agrícola en el país, identificando patrones y tendencias clave que puedan informar estrategias de desarrollo agrícola y políticas públicas efectivas.

**Objetivos de Negocio**

* Identificar los distritos con mayor producción con el fin de invertir de mejor forma sin generar pérdidas
* Definir cuáles características debe tener el cultivo en cuanto a área y estación del año para poder tener mayor producción de un producto evitando que se vean afectadas por el cambio de temporadas
* Identificación de áreas con potencial de crecimiento y la implementación de prácticas agrícolas más efectivas.
* Establecer indicadores clave de rendimiento para monitorear y evaluar el los cultivos en términos de aumento de productividad, reducción de costos y mejora del bienestar de los agricultores
* Establecer colaboraciones con empresas de tecnología agrícola, instituciones académicas, agencias gubernamentales y organizaciones sin fines de lucro para apoyar la implementación exitosa de las inversiones y maximizar su impacto a largo plazo

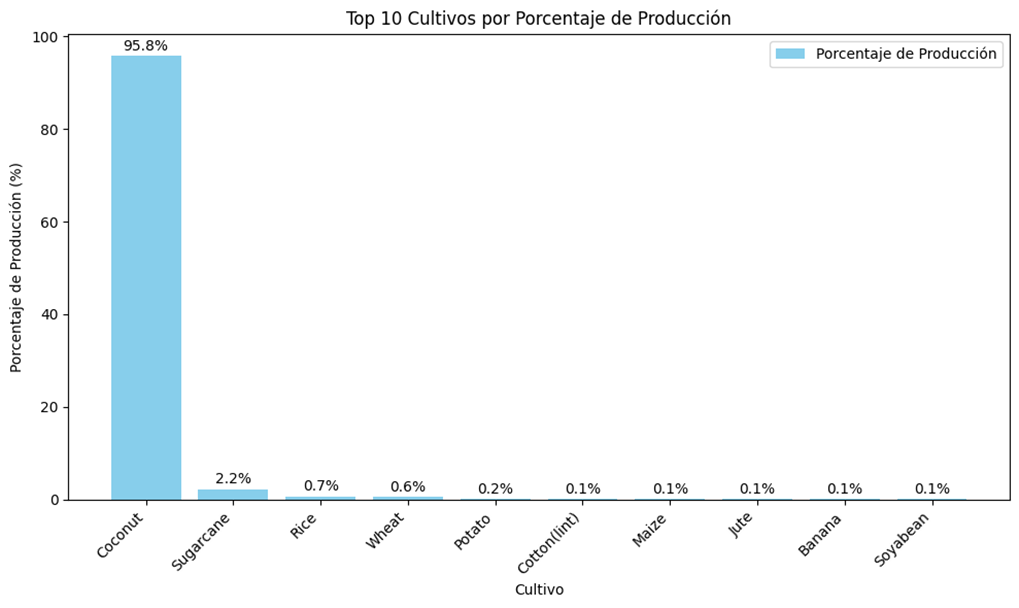
**Objetivos tecnicos**

* Analizar la variabilidad temporal y geográfica de la producción agrícola en India a partir de la base de datos "Cultivos India", con el fin de identificar tendencias estacionales y regionales en la producción de cultivos.
* Estimar el rendimiento agrícola por unidad de área para diferentes cultivos y regiones, utilizando la información detallada sobre área cultivada y producción proporcionada por la base de datos, con el propósito de identificar áreas con altos rendimientos y oportunidades de mejora en la eficiencia agrícola.

**Análisis de Datos**

Cultivos con mayor producción:

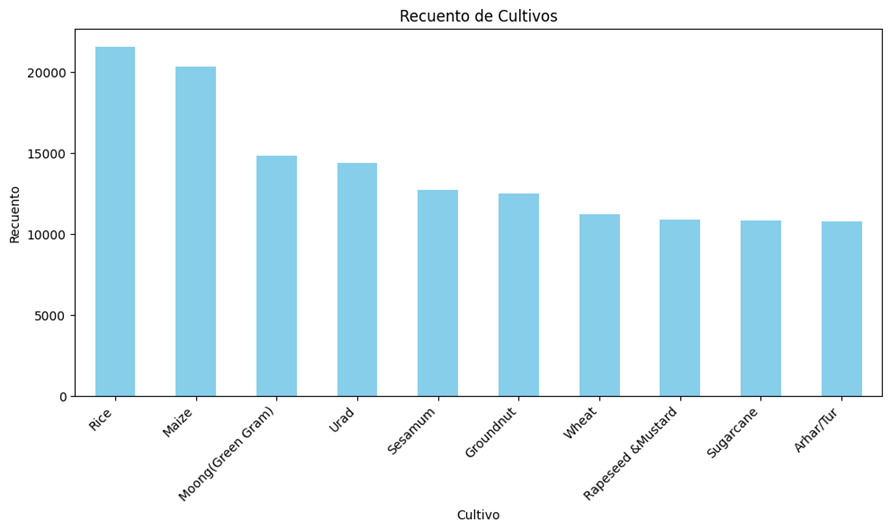
Se realizó un agrupamiento de todos los datos por tipo de cultivo para identificar los cultivos más importantes



Vemos que el más importante es el de coco por lo que sería pertinente centrarse en el para llevar a cabo los objetivos

Sin embargo, esto se debe a que el coco es significativamente más pesado que los demás cultivos y al estar la producción en toneladas no es una variable relativa a la importancia

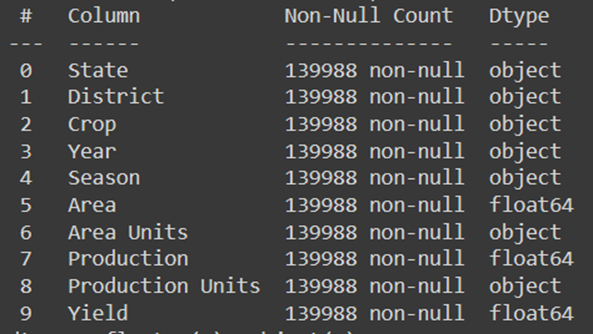
Para esto se realizó el conteo de datos registrados por cultivo



Esta si es una variable relativa a la importancia por lo que se realizará el estudio sobre estos datos.

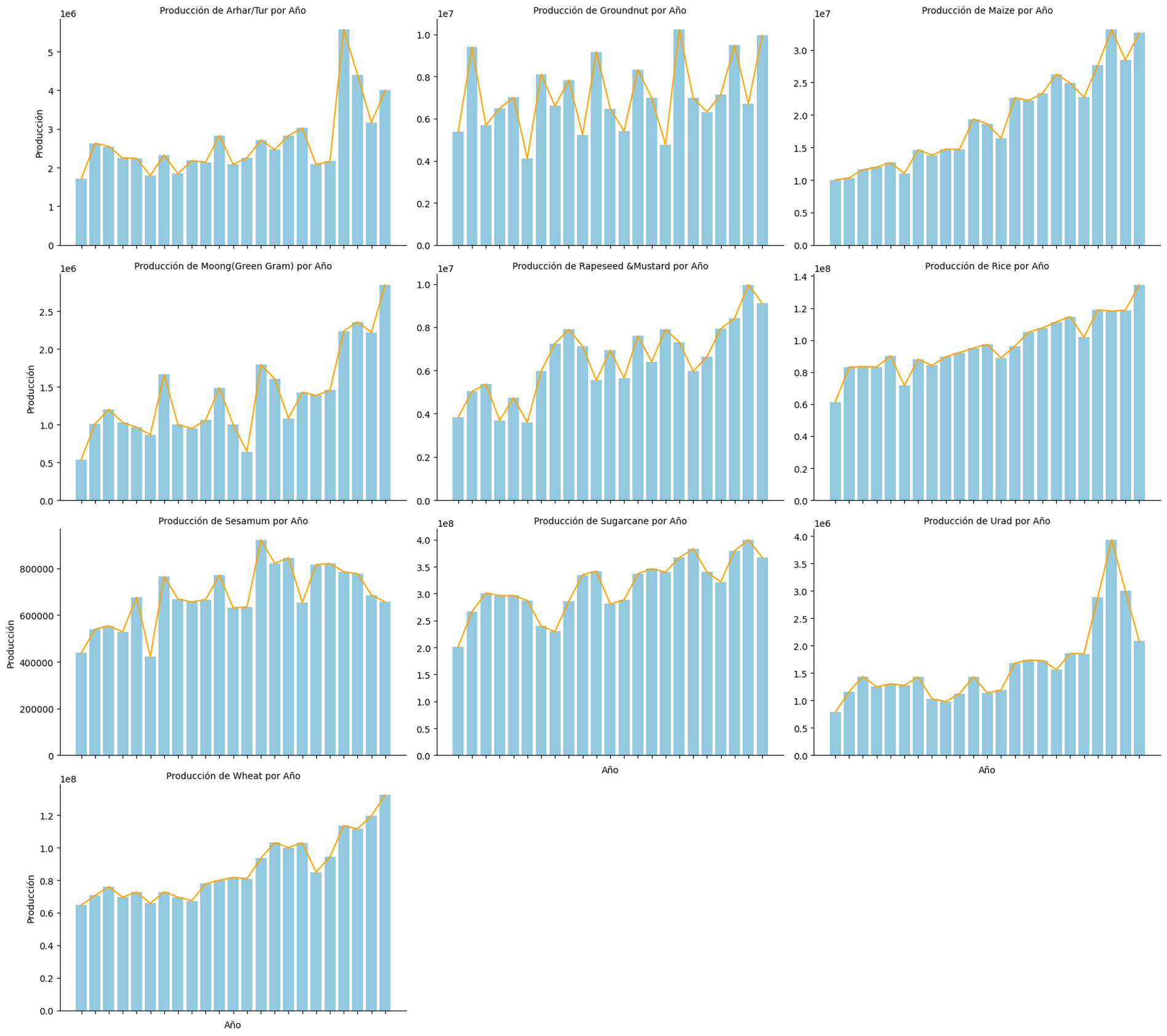
Sería pertinente limitar la base de datos eliminando los registros del último periodo ya que no toma el año completo

Con esto nos queda una base de datos final de 140101 registros a lo largo de esos 24 años



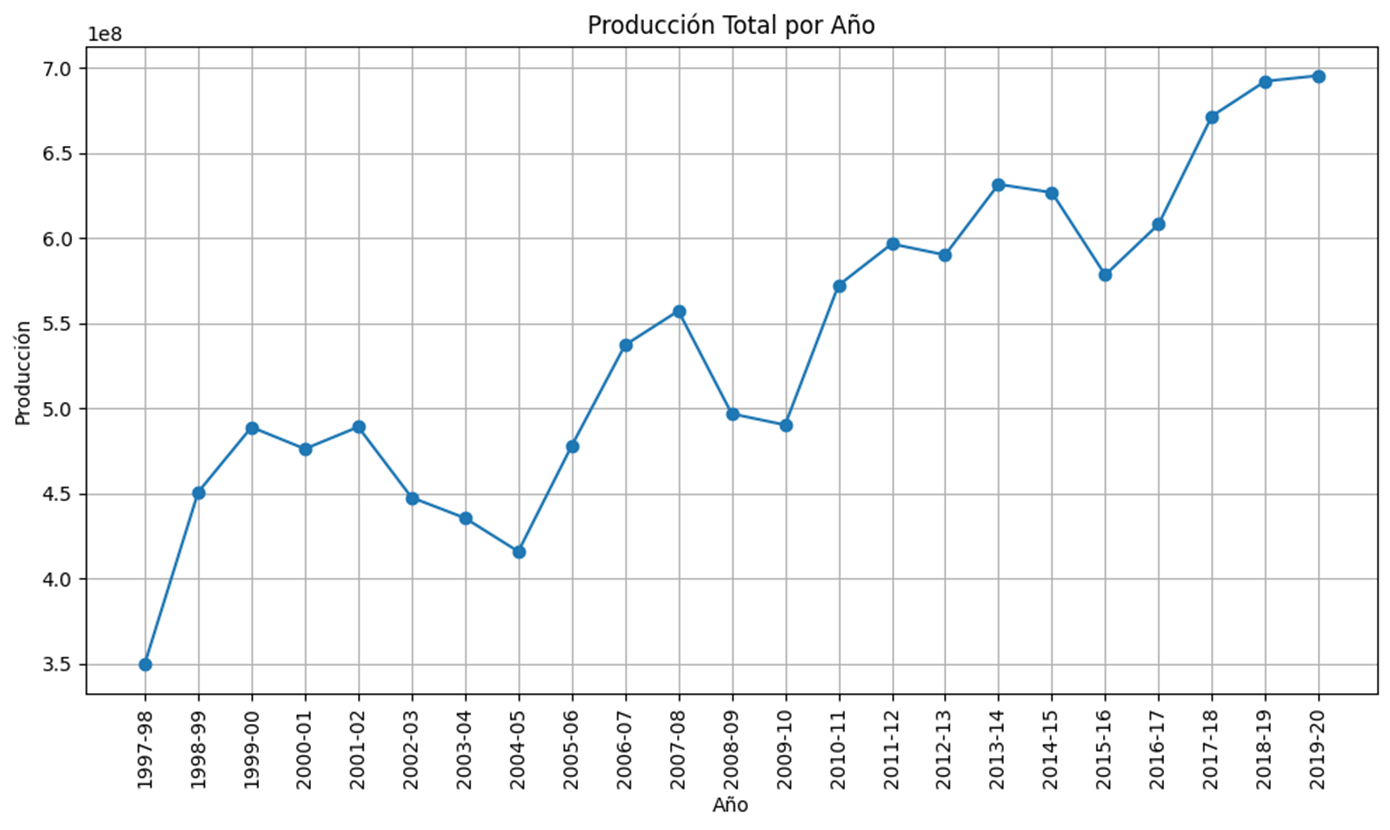
Producción por año por cultivo:

Se realizó una agrupación por año y cultivo y la agregación de su producción asociada

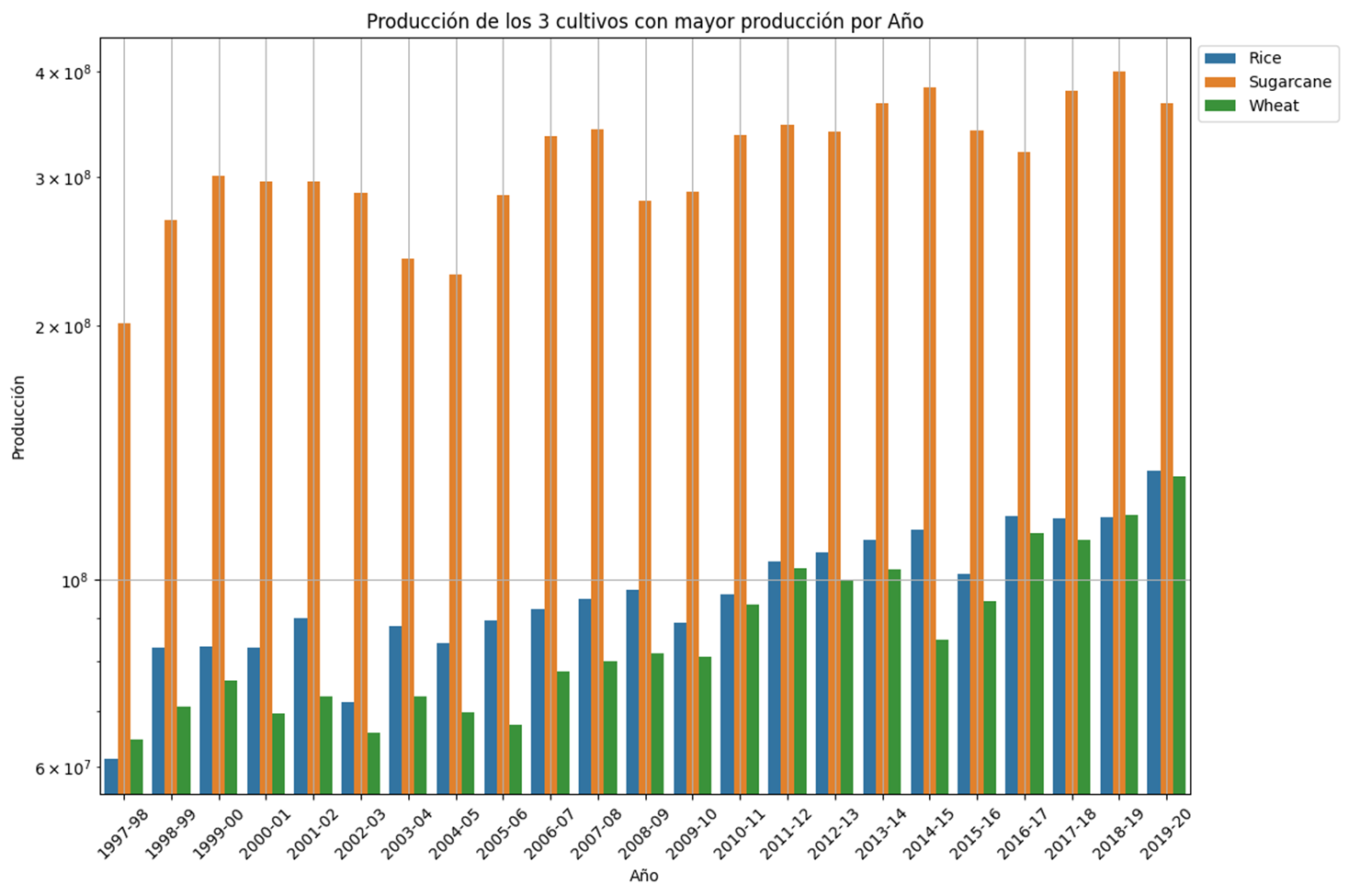


Vemos que en general estos 10 cultivos tuvieron una tendencia positiva, destacando groundnut con una estacionalidad o comportamiento cíclico lo cual es interesante para un modelo de predicción.

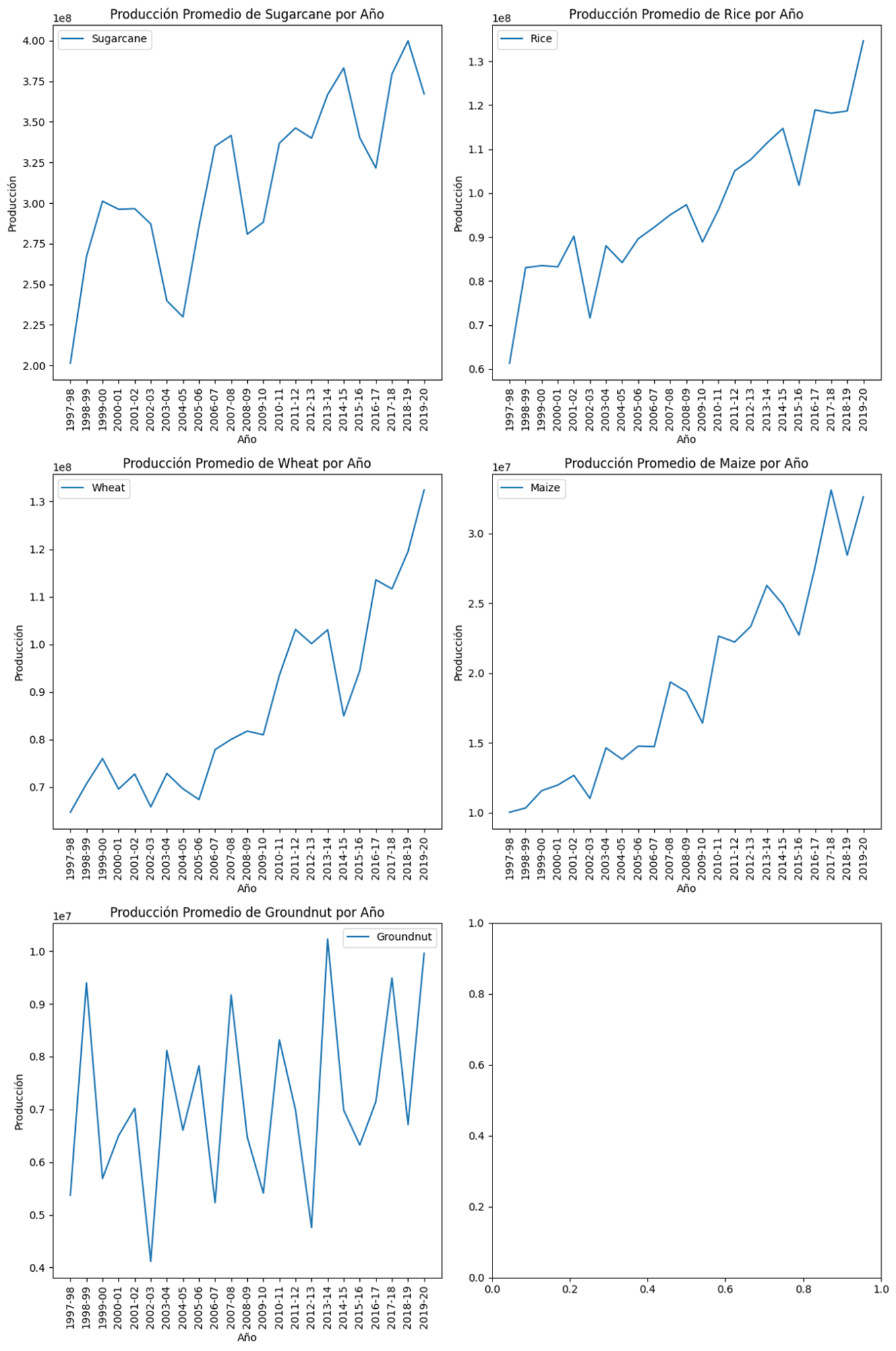
Esto también lo señala la producción total por año

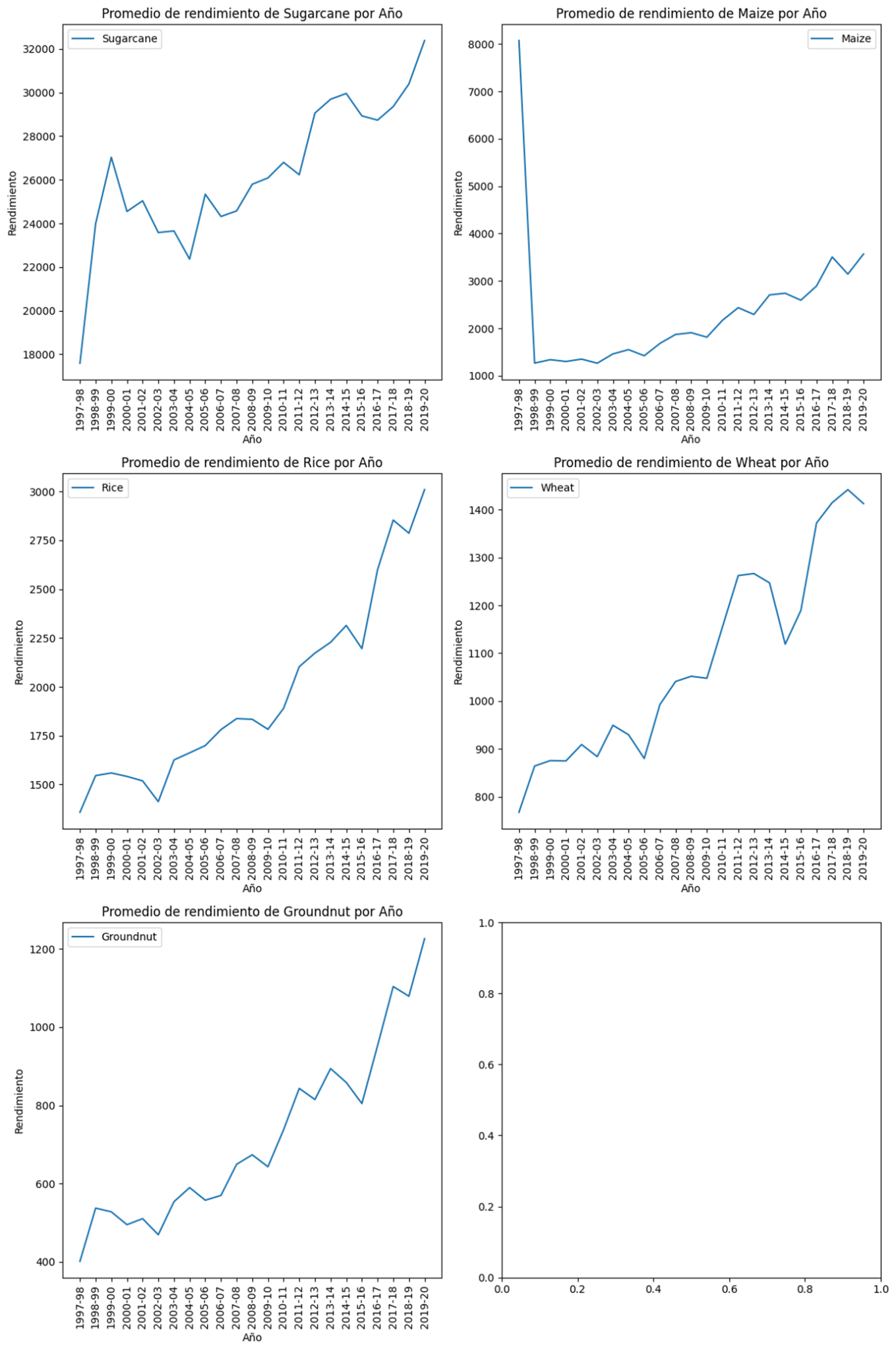


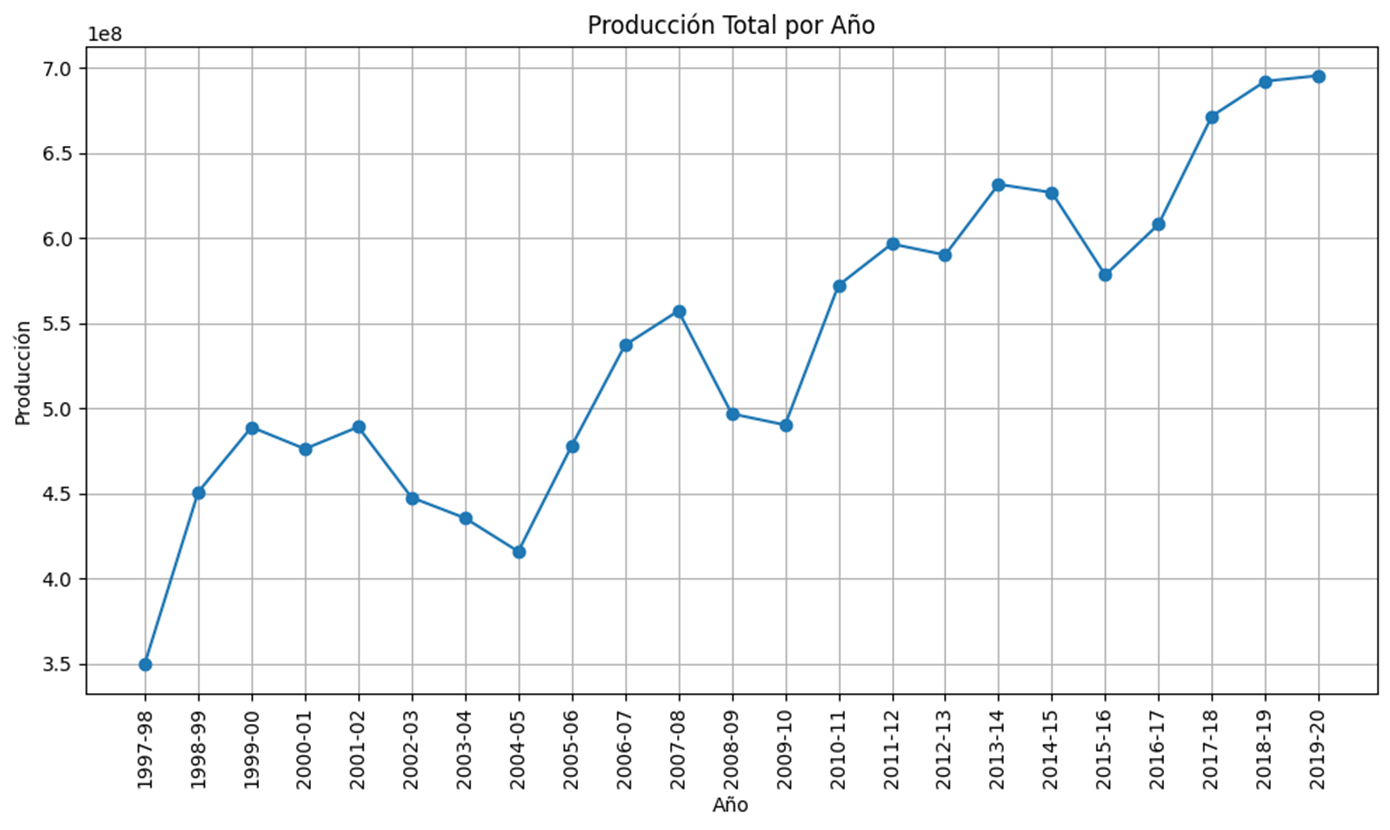
Estos son los 3 cultivos con mayor producción por año



Producción y rendimiento promedio por año:



**Pruebas de hipótesis**



**Primera prueba de hipótesis:** En el año 1997-98 el nivel de producción total (medida en el recuento de veces que aparece el cultivo) fue de menos de 3.5^8 (22518.75390625)

**Segunda prueba de hipótesis:** El porcentaje de crecimiento de la producción promedio por año ha sido mayor al 3%  
**Tercera prueba de hipótesis**: El porcentaje de crecimiento del rendimiento promedio por año (yield) ha sido mayor al x%

**Cuarta prueba de hipótesis**: El arroz tiene una producción mayor a 20.000

**Quinta prueba de hipótesis:** La producción promedio Andaman and Nicobar Islands es mayor a la producción en Tamil Nadu

**Sexta prueba de hipótesis:** La producción promedio de 2015-16 varía es menor a la producción promedio de 2019

**Séptima prueba de hipótesis:** Existe una relación significativa entre el área sembrada y la producción de cultivos.

**Octava prueba de hipótesis:** Existe al menos una diferencia significativa en la producción promedio de cultivos entre las temporadas.