# Informe del Proyecto - Análisis de Producción CNH

Este documento resume los hallazgos clave del proyecto de ingeniería de datos realizado sobre la producción de petróleo y gas en México, utilizando los datasets abiertos de la CNH. El procesamiento fue realizado con PySpark en Databricks y la información se estructuró siguiendo las capas Bronze, Silver y Gold.

## Hallazgos Clave

🛢️ Los campos con mayor producción de petróleo son liderados por el Campo Akal, seguido por otros campos relevantes en la región marina de la Sonda de Campeche.

🔥 En cuanto a gas, el Campo Ku encabeza la lista con una producción sustancial, seguido por campos de las mismas cuencas marinas.

🏢 PEMEX continúa siendo el principal operador tanto en producción de gas como de petróleo, aunque en años recientes se observa participación creciente de operadores privados como Eni y Hokchi.

📉 La tendencia general de producción ha mostrado una disminución en los volúmenes totales año tras año.

## Reflexión Estratégica

Aunque los esfuerzos de apertura energética han traído nuevos jugadores al sector, los datos reflejan que aún no han logrado superar o igualar la producción histórica de PEMEX. Esto plantea la necesidad de analizar si existen barreras técnicas, económicas o regulatorias que están limitando el potencial de estos nuevos operadores.

## Próximos Pasos con Machine Learning

✅ Lo que sí se puede hacer:

- Predicción de producción futura por campo, cuenca u operador.

- Detección de anomalías o caídas inesperadas en producción.

- Análisis de estacionalidad o patrones ocultos en la producción mensual.

- Clustering de campos similares en comportamiento o geografía.

❌ Lo que no se puede hacer directamente (con los datos actuales):

- Evaluación económica detallada por operador (falta de variables financieras).

- Predicción de descubrimientos de nuevos campos (no hay variables geológicas o de exploración).

- Modelos precisos de producción sin un historial de mantenimiento o variables operativas internas.

Este proyecto constituye un punto de partida sólido para la exploración de técnicas avanzadas de análisis de datos y modelado predictivo en el sector energético mexicano. El siguiente paso será desarrollar modelos locales con Python e implementar gráficos y Machine Learning con mayor profundidad.