



FUNDACIÓN YPF

Plan 4x4 de YPF

PRIMER PILAR: VACA MUERTA



Cyntia Nasabun - Antonella Fontanetto

Julio 2025

Contenido

- 01 Introducción
- 02 Sobre el Proyecto
- 03 Análisis Exploratorio
- 04 Transformación de variables
- 05 Modelo Supervisado
- 06 Modelo No Supervisado
- 07 Anexo: Redes neuronales
- 08 Conclusiones

Introducción - Plan 4x4

El Plan 4x4 de YPF es una iniciativa estratégica que busca cuadruplicar en cuatro años la producción no convencional de petróleo y gas en Vaca Muerta. Este plan apunta a consolidar la posición de YPF como líder en la Cuenca Neuquina, acelerando la inversión, el desarrollo de pozos, la infraestructura de midstream y la exportación de hidrocarburos. Con foco en eficiencia, escalabilidad y generación de divisas, el Proyecto 4x4 representa un pilar fundamental para el crecimiento energético y económico de la Argentina.

01

Primer Pilar: Vaca Muerta

Este es el pilar más central del Plan 4x4 de YPF y está enfocado en acelerar la producción de petróleo y gas no convencional en la Formación Vaca Muerta, principalmente en la Cuenca Neuquina (Neuquén, Río Negro, Mendoza).

¿Por qué es estratégico?

- Porque Vaca Muerta es una de las formaciones de shale más grandes del mundo.
- YPF ya tiene un know-how técnico y operativo en la región, lo que le permite escalar con rapidez.
- El petróleo de Vaca Muerta tiene un mercado claro: refinación interna y exportación por puertos patagónicos y atlánticos.



Sobre el Proyecto

En el marco del Plan 4x4 de YPF, desarrollamos un análisis predictivo basado en datos históricos de producción no convencional de petróleo y gas en Vaca Muerta. Aplicamos técnicas de Machine Learning, combinando el modelo Prophet para series temporales y un modelo no supervisado para identificar patrones y comportamientos ocultos en la producción. El objetivo es estimar la evolución productiva para los próximos 3 a 4 años, brindando una herramienta estratégica que acompañe el crecimiento proyectado por YPF en su plan de expansión.



Análisis Exploratorio

01

Dataset

Los datos utilizados provienen de la Secretaría de Energía de la Nación, que publica mensualmente información oficial sobre la producción de petróleo y gas en Vaca Muerta.

02

Limpieza y transformación

Limpieza y transformación del dataset (formato de fechas, renombrar columnas, valores faltantes)

03

Visualización

Visualización de la evolución histórica (líneas de tiempo, boxplots, heatmaps)

03

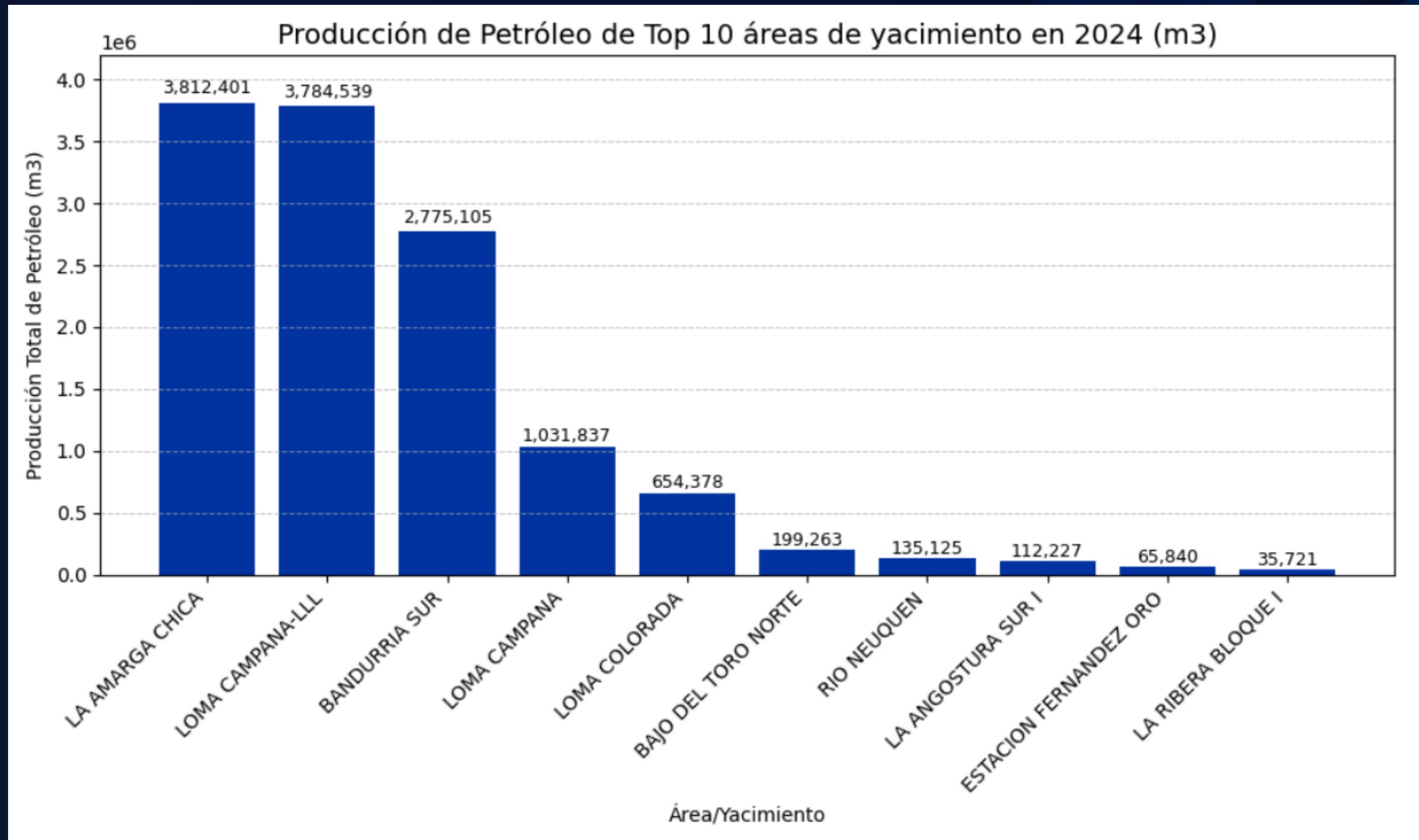
04

Outliers

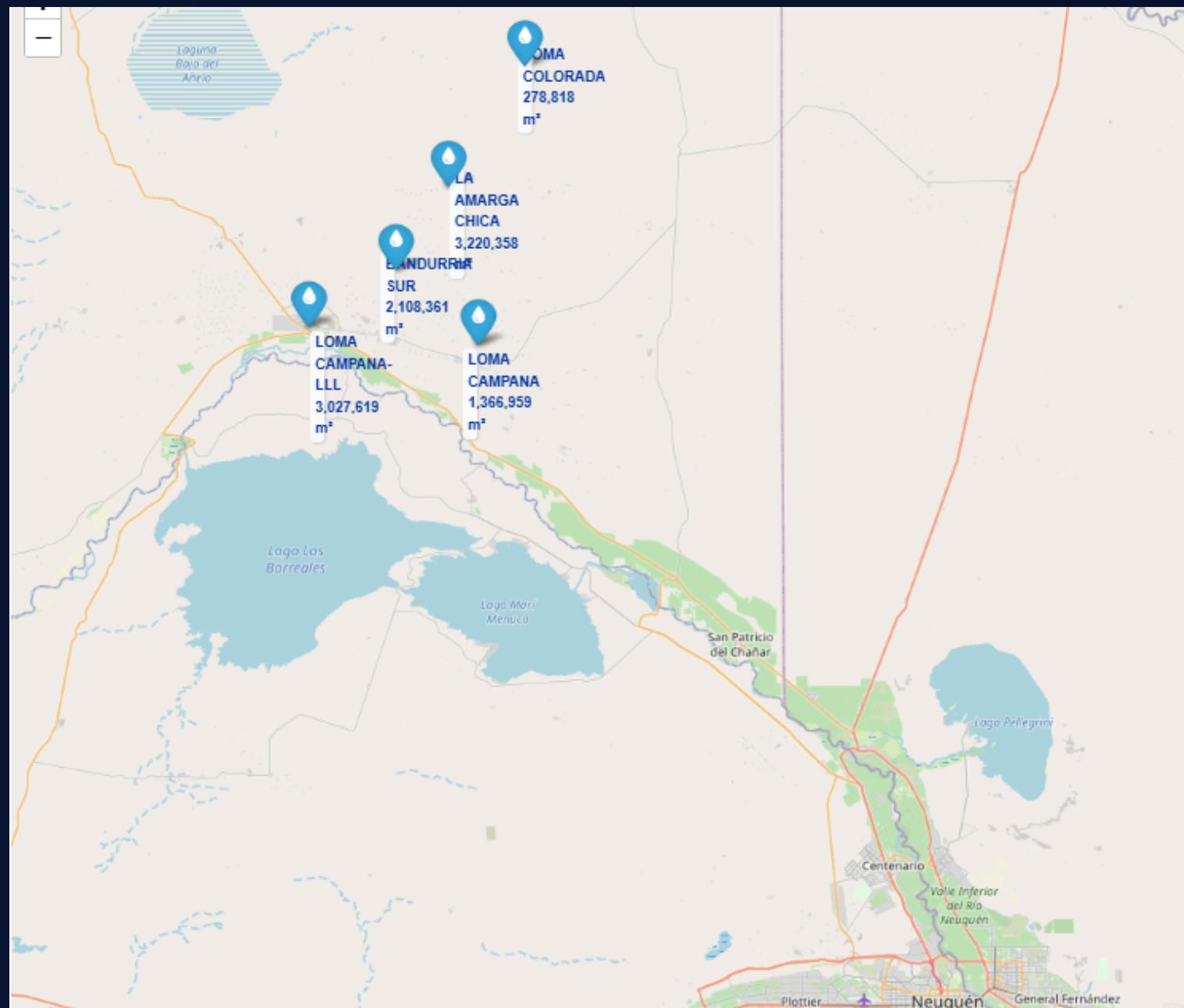
Identificación de outliers y eventos atípicos. Cálculo de rango intercuartílico.



Producción de Petróleo



Mapa: Producción de Petróleo

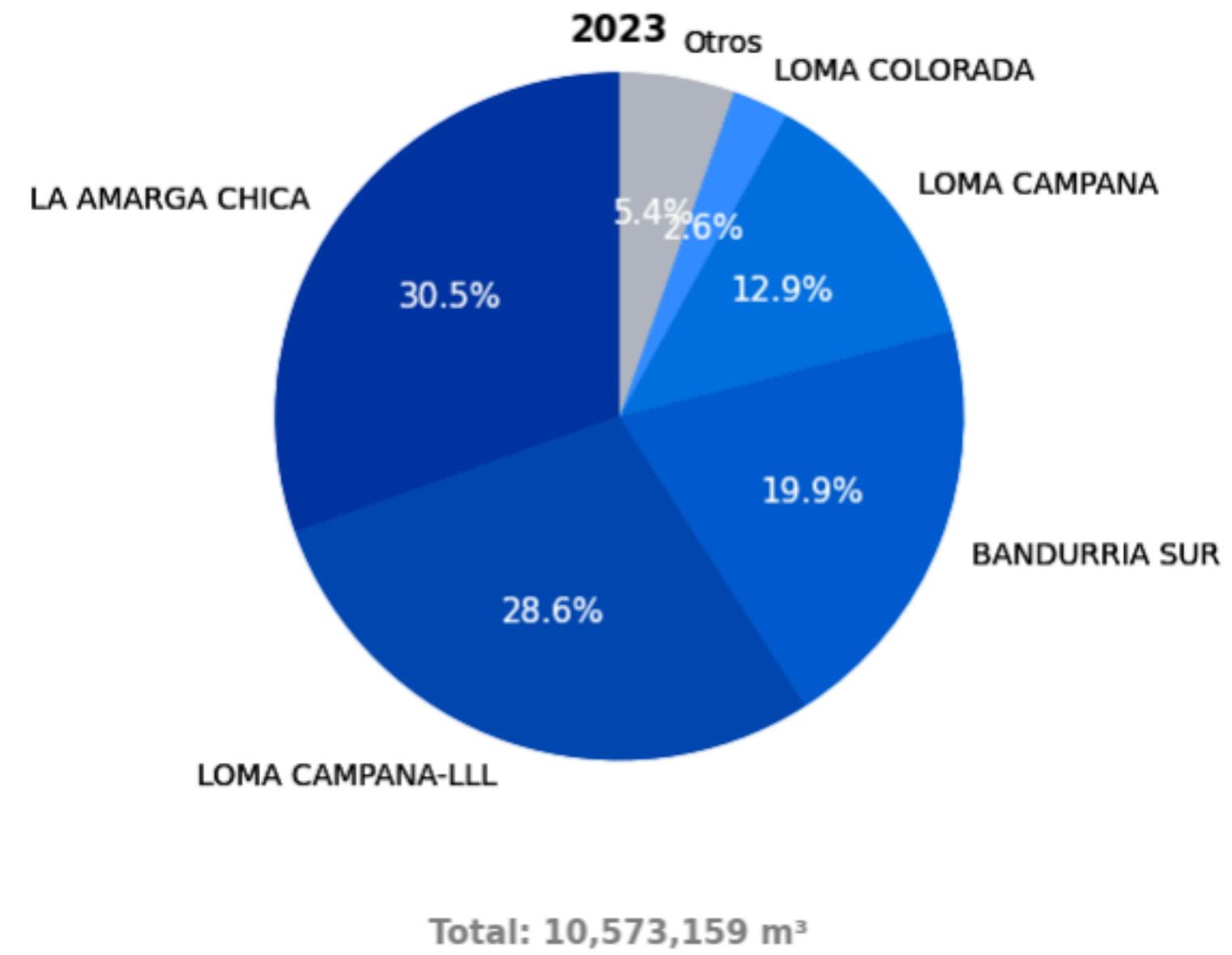


Mapa de las principales 5 áreas de yacimiento

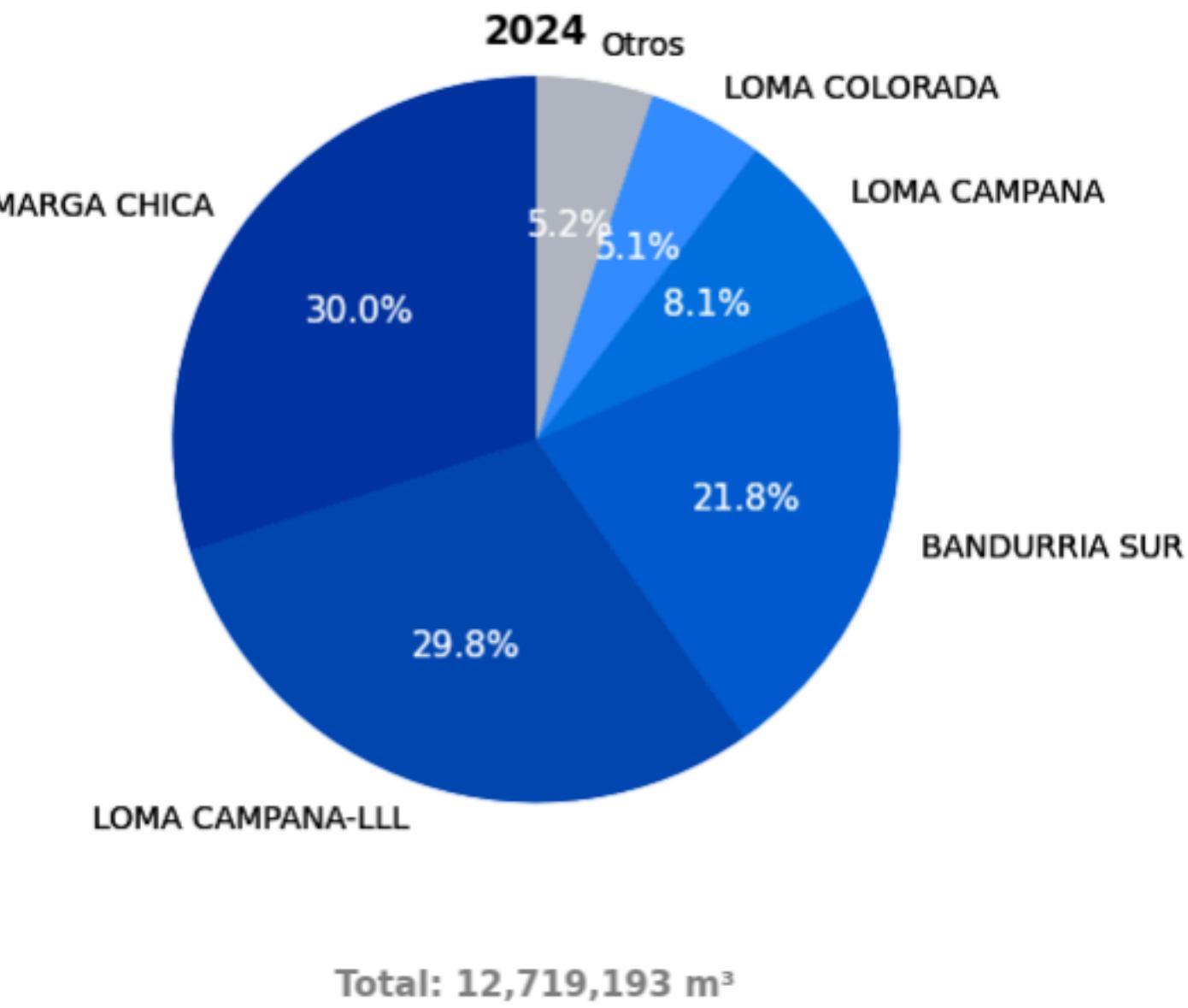
- Ubicación geográfica: Los yacimientos están ubicados en la cuenca neuquina.
- Principales áreas productoras: Las áreas más destacadas por volumen de producción son La Amarga Chica (3.220.358 m³) y Loma Campana – LLL (3.027.619 m³), ambas con una producción significativa de hidrocarburos.
- Otras áreas relevantes: También se destacan Bandurria Sur (2.108.361 m³) y Loma Campana (1.366.959 m³), que complementan la actividad productiva en esta región con altos volúmenes.
- Yacimiento de menor volumen: Loma Colorada es el yacimiento con menor producción visualizada en el mapa, con 278.818 m³.
- Concentración territorial: Los yacimientos están relativamente cercanos entre sí, lo cual favorece la logística, el transporte y la planificación integrada de operaciones en la zona de explotación.

Producción de Petróleo

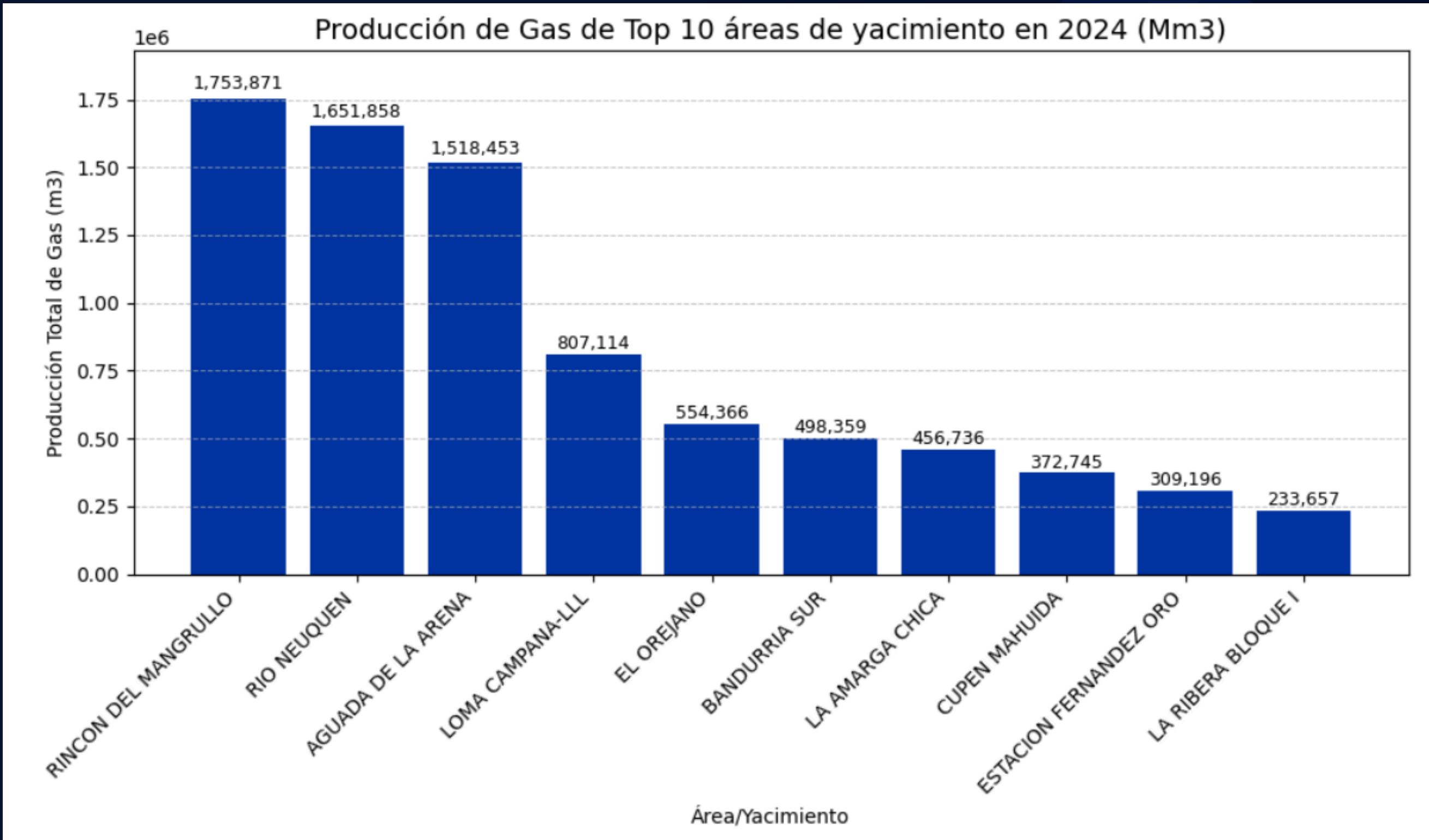
Participación de la producción de petróleo de las Top 5 áreas



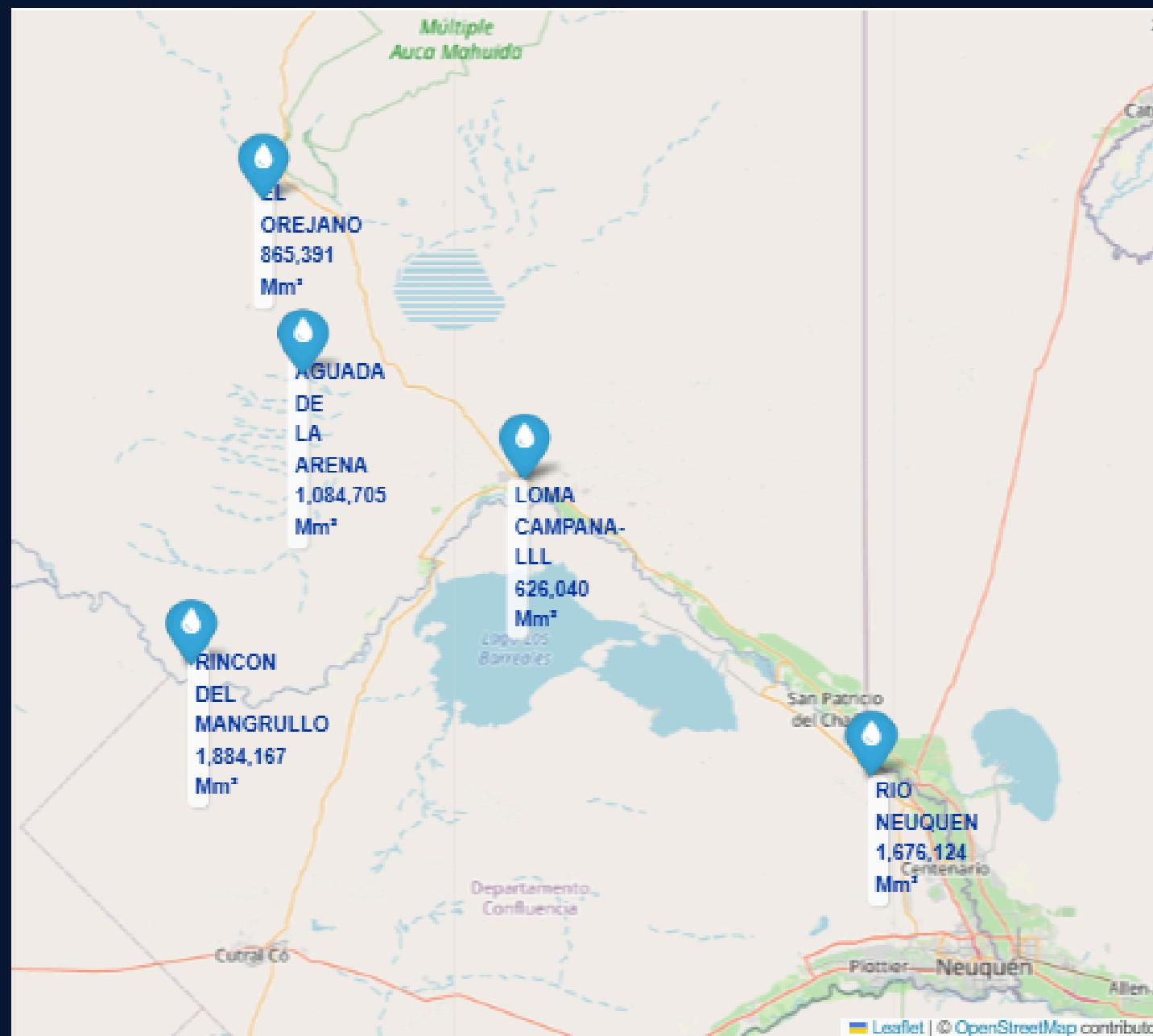
Participación de la producción de petróleo de las Top 5 áreas



Producción de Gas



Mapa: Producción de Gas

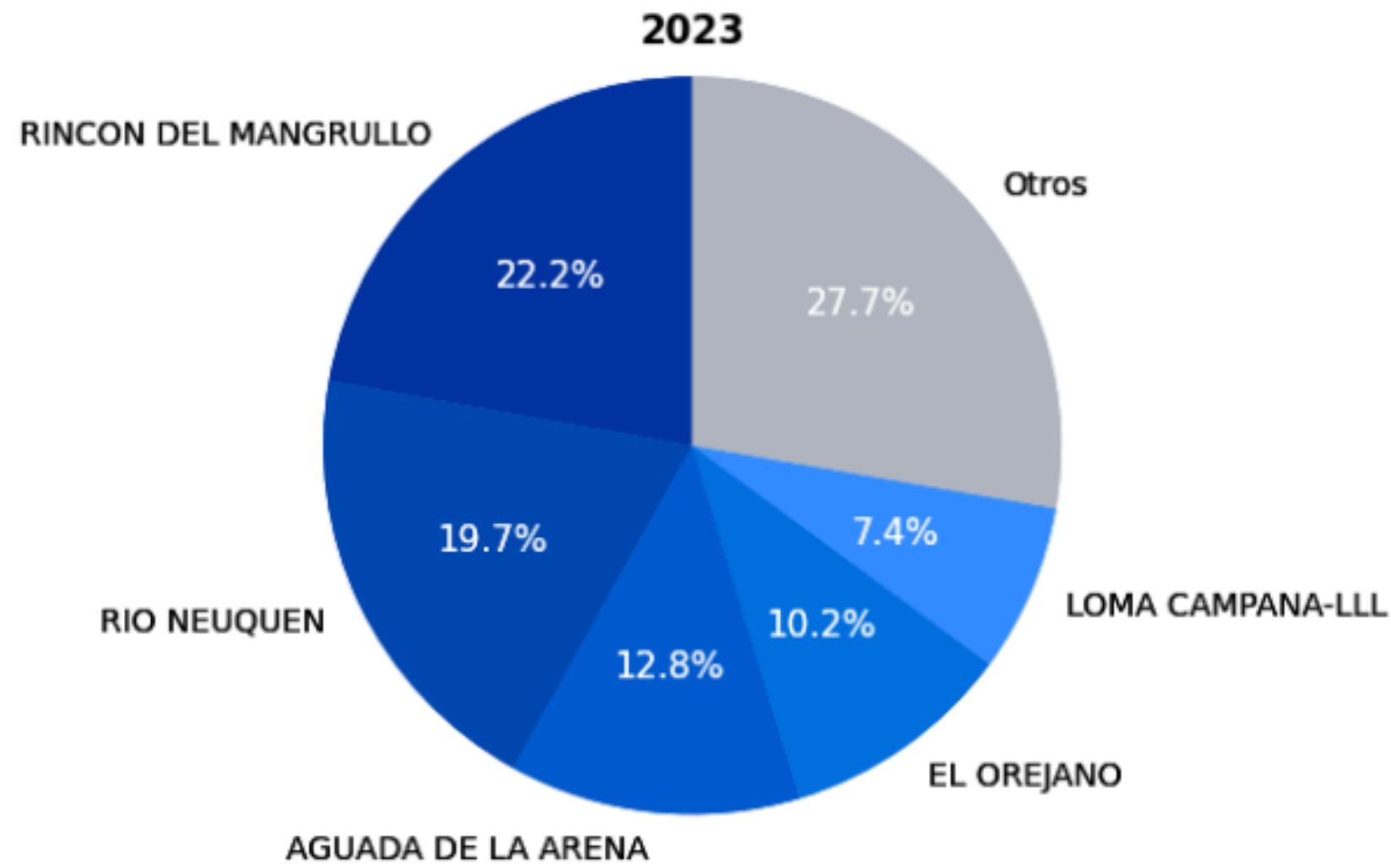


Mapa de las principales 5 áreas de yacimiento

- Áreas productivas distribuidas: El mapa muestra cinco yacimientos distribuidos principalmente al norte y oeste de la ciudad de Neuquén, reflejando la amplitud de la explotación en la Cuenca Neuquina.
- Yacimiento con mayor producción: Rincón del Mangrullo lidera en volumen con 1.884.167 Mm³, consolidándose como una de las áreas más productivas del mapa.
- Otros yacimientos destacados: Le siguen Río Neuquén con 1.676.124 Mm³ y Aguada de la Arena con 1.084.705 Mm³, también con importantes niveles de producción.

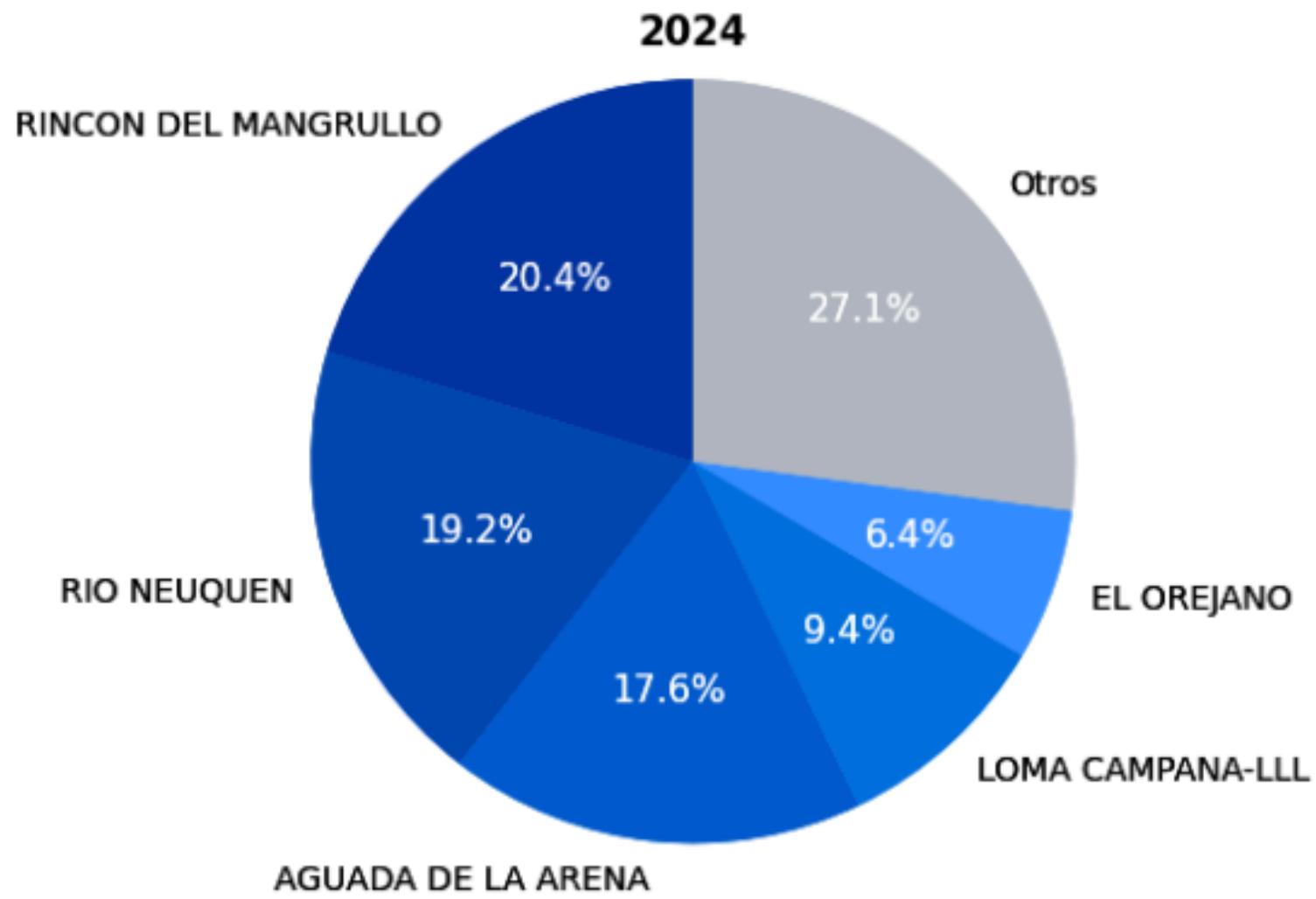
Producción de Gas

Participación de la producción de gas de YPF de las Top 5 áreas



Total: 8,490,367 Mm³

Participación de la producción de gas de YPF de las Top 5 áreas



Total: 8,616,679 Mm³



04

Transformación de variables



- **Estandarización de variables categóricas:** Utilizamos One Hot Encoding para transformar variables como la cuenca o el tipo de producto en columnas numéricas, facilitando su uso en modelos de machine learning.
- **Preparación para modelos supervisados:** Esta codificación permitió integrar variables categóricas al modelo Prophet y a otros algoritmos supervisados, sin violar sus requerimientos de entrada numérica.
- **Mejora de la interpretabilidad y segmentación:** Al separar las categorías en columnas independientes, fue posible analizar mejor el comportamiento de cada segmento productivo (por ejemplo, petróleo vs gas).



04 Transformación de variables

```
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
import pandas as pd

# Crear una copia y seleccionar solo las columnas necesarias
produccion_ypf = prod_ypf.copy()
produccion_ypf = produccion_ypf[
    ['Produccion de Petroleo', 'Produccion de Gas','Producción de agua', 'Profundidad' , 'Tipo de pozo', 'Tipo de recurso',
     'Subtipo de recurso', 'Area de permiso de concesion']
]

# Especificar las columnas categóricas
categorical_cols = ['Tipo de pozo',
                     'Tipo de recurso', 'Subtipo de recurso',
                     'Area de permiso de concesion']

# Crear el OneHotEncoder dentro de un ColumnTransformer
column_transformer = ColumnTransformer(
    transformers=[
        ('encoder', OneHotEncoder(sparse_output=False), categorical_cols)
    ],
    remainder='passthrough'
)

# Transformar los datos
prod_encoded_array = column_transformer.fit_transform(produccion_ypf)

# Obtener los nombres de las nuevas columnas
encoded_features_names = column_transformer.named_transformers_['encoder'].get_feature_names_out(categorical_cols)

# Crear el DataFrame con las columnas codificadas y numéricas
prod_encoded_df = pd.DataFrame(
    prod_encoded_array,
    columns=list(encoded_features_names) + ['Produccion de Petroleo', 'Produccion de Gas','Producción de agua', 'Profundidad'
])
```

categorical_cols

Column_Transformer

Transforma los datos

Nuevo dataframe

MODELO SUPERVISADO

MODELO PROPHET

Prophet es un modelo de series temporales desarrollado por Facebook, diseñado para realizar pronósticos de forma precisa y automática. Es especialmente útil en contextos con datos históricos con tendencias claras y estacionalidad, como la producción energética. Permite incorporar eventos externos y es robusto ante datos faltantes y cambios abruptos.



Modelado

Aplicamos el modelo Prophet por separado a las series temporales de producción de petróleo y de gas no convencional en Vaca Muerta. El modelo permitió capturar tendencias generales, patrones estacionales y variaciones mensuales en cada tipo de producción. A partir del entrenamiento con datos históricos, generamos pronósticos para los próximos 3 a 4 años, alineados con los objetivos del Plan 4x4 de YPF



PREPARACIÓN

Requiere crear dos columnas particulares para predecir, como producción de petróleo o gas).



ENTRENAMIENTO

Se ajustó Prophet con los datos históricos, permitiendo que detecte tendencias y estacionalidad.



PREDICCIÓN

Se creó un rango futuro de fechas (3–4 años) para proyectar la producción.



VISUALIZACIÓN

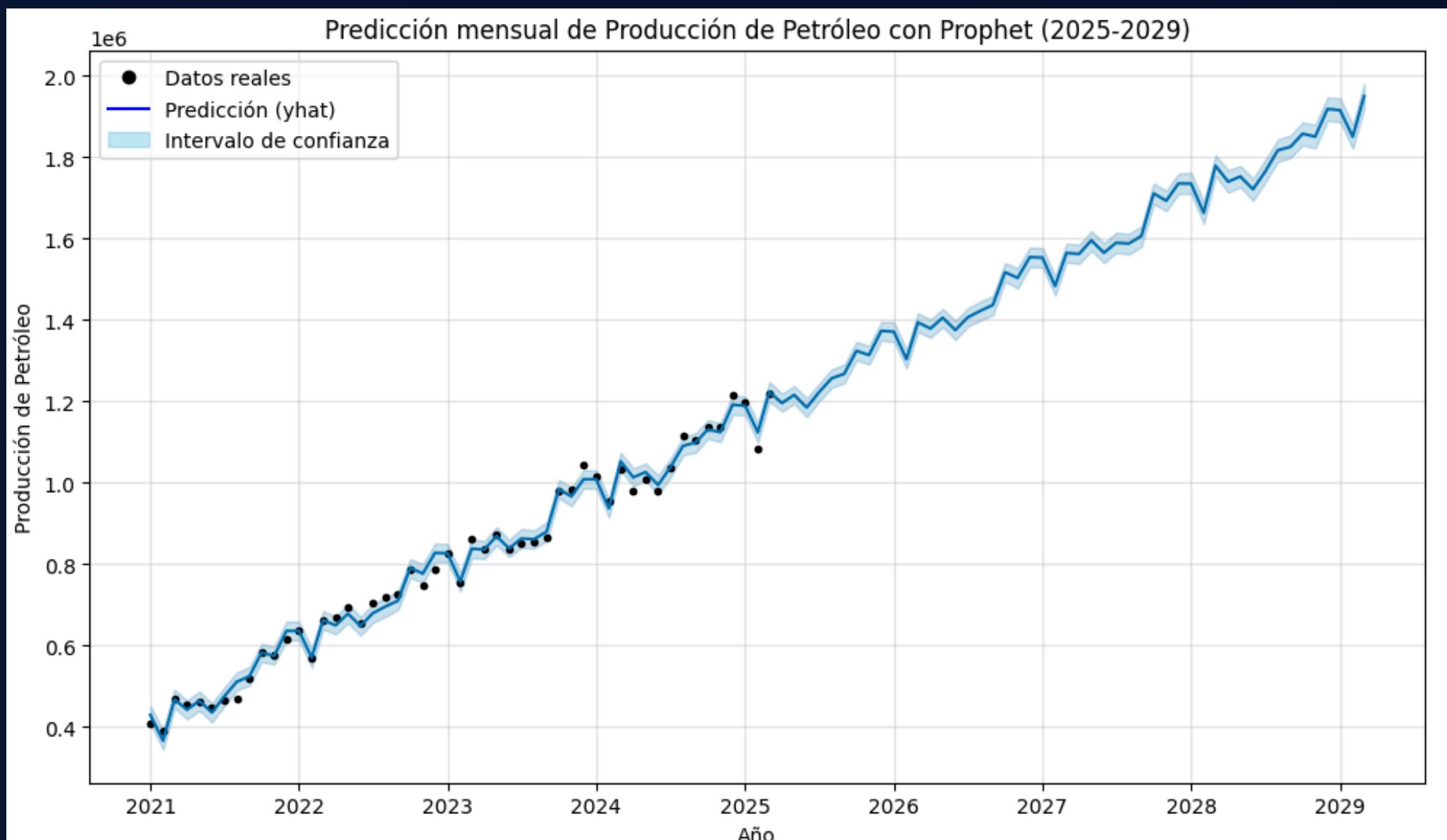
Se obtuvieron las proyecciones y se graficaron junto con los intervalos de confianza.



EVALUACIÓN

Se analizaron las curvas resultantes para entender el comportamiento esperado y su relación con el Plan 4x4.

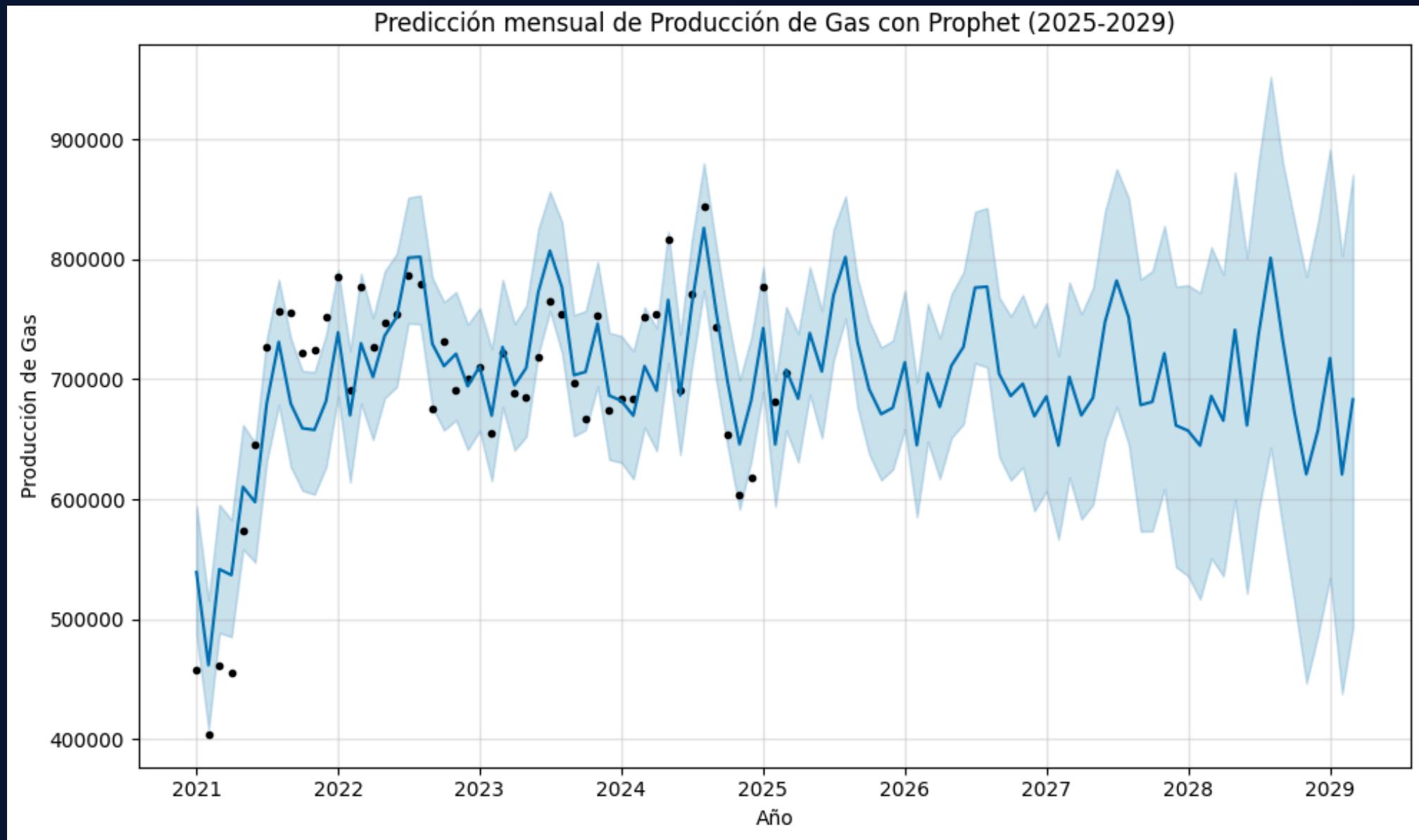
Modelado: Producción de Petróleo



Interpretación del Gráfico

El gráfico muestra la predicción mensual de la producción de petróleo en Vaca Muerta entre 2025 y 2029. La línea azul representa la proyección generada por el modelo Prophet, mientras que la banda celeste indica el intervalo de confianza. Se observa una tendencia creciente sostenida, en línea con los objetivos del Plan 4x4 de YPF. Los puntos negros corresponden a los datos reales, que validan el buen ajuste del modelo en la fase histórica.

Modelado: Producción de Gas



Interpretación del Gráfico

El gráfico muestra la predicción mensual de la producción de gas en Vaca Muerta entre 2025 y 2029 utilizando el modelo Prophet. La línea azul representa la proyección, mientras que la banda celeste muestra el intervalo de confianza. A diferencia del petróleo, la producción de gas presenta una tendencia más inestable, con mayor variabilidad y un intervalo de predicción más amplio hacia el futuro. Esto sugiere incertidumbre en el comportamiento a largo plazo, lo cual puede estar asociado a factores estacionales, regulatorios o de demanda.

Modelado: Métricas

```
from sklearn.metrics import mean_absolute_error, mean_squared_error

y_true = test['y'].values
y_pred = pred_test['yhat'].values

mae = mean_absolute_error(y_true, y_pred)
rmse = np.sqrt(mean_squared_error(y_true, y_pred))
mape = (abs((y_true - y_pred) / y_true).mean()) * 100

print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"MAPE: {mape:.2f}%)
```

MAE: 51743.27
RMSE: 64859.03
MAPE: 7.22%

El modelo Prophet alcanzó un MAPE del 7,22%, lo que indica una alta precisión en las predicciones. Este valor significa que, en promedio, el error porcentual entre los valores reales y estimados fue bajo. Un MAPE menor al 10% se considera un buen desempeño para series temporales en contextos reales.

MODELO NO SUPERVISADO

MODELO K-MEANS Y MEANSHIFT

K-Means es un algoritmo de clustering no supervisado que agrupa los datos en k clusters predefinidos, minimizando la distancia entre los puntos y el centroide de su grupo. En el contexto de producción de petróleo y gas, K-Means permite identificar perfiles productivos claramente diferenciados según patrones de volumen.

MeanShift, en cambio, no requiere especificar la cantidad de clusters de antemano. Utiliza estimaciones de densidad para encontrar automáticamente regiones de alta concentración de datos, lo que lo hace útil cuando no se conoce la estructura del conjunto.



Modelado

Aplicamos varios modelos, entre ellos el modelo KMeans y MeanShift aplicado a la producción de petróleo y de gas no convencional en Vaca Muerta. El modelo Kmeans permitió detectar 3 clusters de comportamiento frente a las variables mencionadas anteriormente.

PREPARACIÓN

Se utilizaron dos versiones de dataset, uno con datos agrupados mensualmente y otro con los outliers

ENTRENAMIENTO

Se ajustó KMeans y MeanShift con los datos de producción de petróleo y producción de gas.

PREDICCIÓN

Detección de la cantidad de clústers según el modelo, es decir, diferentes grupos de producción de petróleo y gas.

VISUALIZACIÓN

Visualización de la cantidad de clusters encontrados por los diferentes modelos no supervisados.

EVALUACIÓN

Evaluación a través de diferentes métricas como: Silhouette, Elbow, Inertia, Davis-Bouldin, Variance Ratio Criterion

Modelado: KMeans



Interpretación del Gráfico

El gráfico muestra la segmentación de los registros según sus patrones de producción de petróleo y gas, utilizando el algoritmo K-Means.

- Cluster 0 (verde): Registros con producción baja de ambos recursos.
- Cluster 1 (naranja): Producción de petróleo alta con gas moderado o bajo. Representa la mayoría de los casos.
- Cluster 2 (azul): Alta producción de gas, pero con producción de petróleo casi nula.

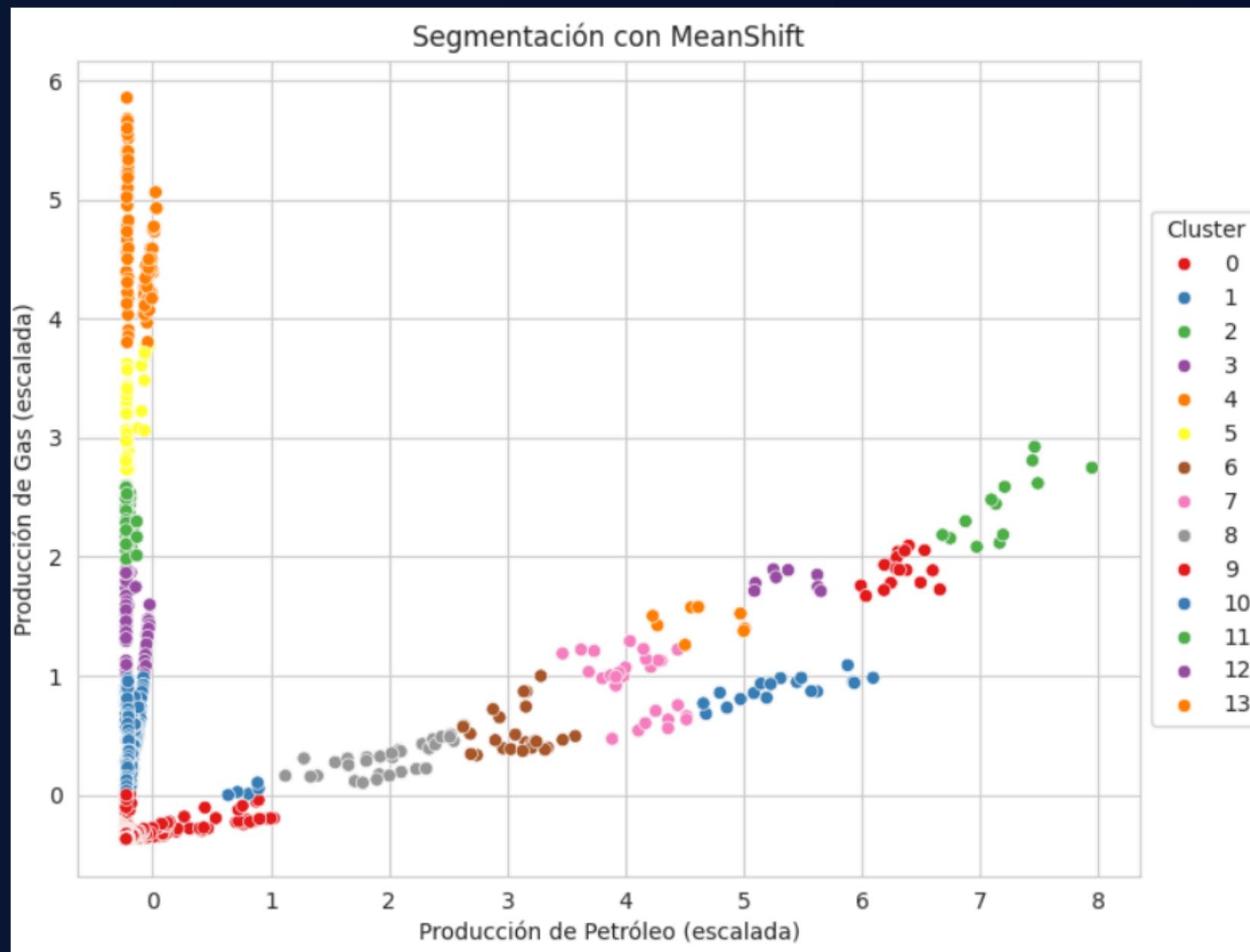
Esta segmentación permite identificar grupos de pozos o áreas con distintos perfiles productivos, útil para estrategias diferenciadas de análisis, inversión o gestión operativa.

Métrica: Silhouette

```
:> from sklearn.metrics import silhouette_score  
  
sil_score = silhouette_score(X_scaled, labels)  
print(f"Silhouette Score: {sil_score:.3f}")  
  
Silhouette Score: 0.900
```

El Silhouette Score obtenido fue de 0.900, lo que indica una excelente calidad de agrupamiento. Los datos fueron correctamente agrupados con alta cohesión interna y buena separación entre clústeres, validando la efectividad del modelo no supervisado aplicado.

Modelado: MeanShift



Este gráfico muestra cómo el algoritmo MeanShift agrupó los registros según sus patrones de producción de petróleo y gas sin necesidad de especificar la cantidad de clusters previamente.

- Se detectaron 14 clusters automáticamente, basados en la densidad de los datos.
- Se observan grupos verticales donde la producción de gas es alta pero la de petróleo es casi nula, y otros alineados diagonalmente donde ambas producciones aumentan de forma correlativa.
- Esto sugiere que hay distintos perfiles productivos: pozos muy gasíferos, pozos petroleros, y otros con producción combinada.

La segmentación es más granular que K-Means, permitiendo detectar subgrupos con comportamientos particulares.

Métrica: Silhouette

```
from sklearn.metrics import silhouette_score  
  
score = silhouette_score(X_scaled, labels_ms)  
print(f'Silhouette Score: {score:.3f}')
```

Silhouette Score: 0.864

MeanShift obtuvo un Silhouette Score de 0.864, también alto, pero levemente inferior, lo que sugiere una buena

REDES NEURONALES

KOHONEN SOM Y AUTOENCODER

Kohonen SOM (Self-Organizing Map):

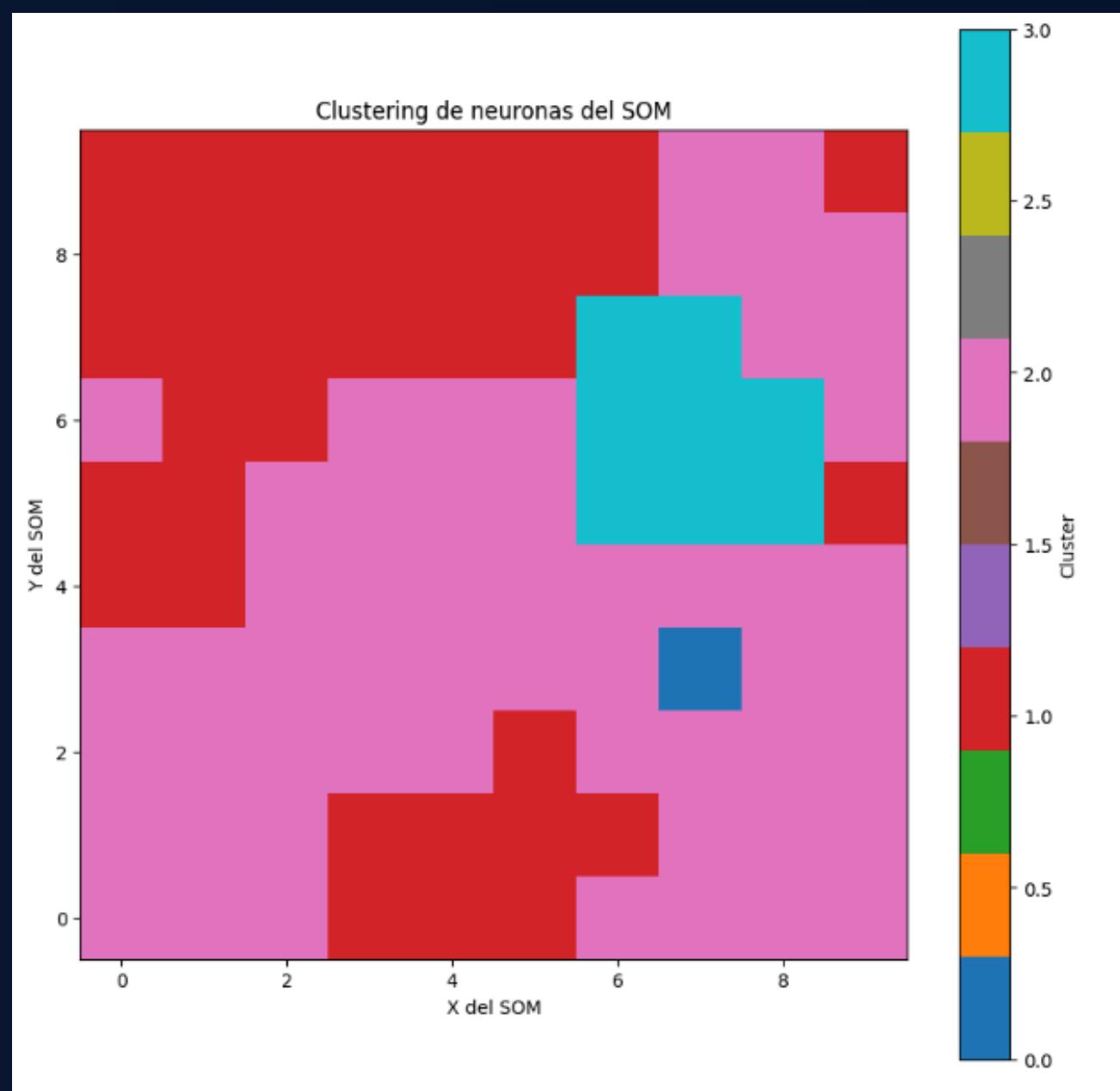
Es una red neuronal no supervisada que proyecta datos de alta dimensión en un mapa de menor dimensión (generalmente 2D). Se usa comúnmente para segmentación, detección de patrones y visualización de clusters.

Autoencoder:

Es una red neuronal que aprende a comprimir (codificar) y luego reconstruir los datos de entrada. Se usa para reducción de dimensionalidad, detección de anomalías y preentrenamiento de redes profundas.



Modelado: Kohonen SOM



Interpretación del Gráfico.

Se identifican tres clusters principales:

- ◆ El cluster 2 (color rosa) es el más extendido, ocupando gran parte del mapa, lo que indica que la mayoría de las neuronas presentan patrones similares.
 - ◆ El cluster 1 (color rojo) forma regiones bien delimitadas en la parte izquierda y centro-sur, lo que indica un grupo de observaciones con patrones diferenciados.
 - ◆ El cluster 0 (color turquesa) aparece en una región compacta del centro-derecha, señalando otro subconjunto con características comunes, pero distintas del resto.
- La visualización permite interpretar cómo el SOM ha segmentado el espacio de características en diferentes zonas estructuradas, útiles para tareas como segmentación, reducción de dimensionalidad o análisis exploratorio.

Modelado: Autoencoder



¿Qué muestra?

Cada punto es una muestra (puede ser por pozo, por día, por etapa de producción, etc.). Dim1 y Dim2 son las nuevas variables aprendidas por el autoencoder (las 2 dimensiones reducidas). Esta es una representación abstracta que resume toda la información original en un plano.

Se observa una estructura clara, en forma de “abanico” o “cuchillo de pez”. Esta forma indica que los datos tienen una estructura no lineal compleja, que fue capturada por el autoencoder.

Las ramas podrían representar distintos tipos de comportamiento, por ejemplo:

- Pozos con corte de agua creciente,
- Etapas distintas de producción (inicio, meseta, declinación),
- Anomalías o patrones operativos.

Conclusión

Estas herramientas nos brindan una base técnica robusta para proyectar escenarios futuros con mayor precisión, especialmente dentro del marco del Plan 4x4 de YPF, que busca cuadruplicar la producción y aumentar las exportaciones hacia 2030.

Este enfoque no solo potencia la toma de decisiones estratégicas, sino que también posiciona la analítica avanzada como un pilar clave para el desarrollo energético sostenible de la región. A futuro, integrar estos modelos con variables económicas, logísticas y ambientales permitirá construir un sistema predictivo integral que acompañe los objetivos de crecimiento e inversión del país.





FUNDACIÓN YPF

¡Gracias IngenIAs!

«MÁS MUJERES EN TECNOLOGÍA»

Cyntia Nasabun - Antonella Fontanetto