

Carrera Data Scientist 2024

Data Analytics



Análisis de la Producción de Petróleo y Gas No Convencionales

ANTONINI GABRIELA PAOLA

Proyecto Final



Contenido

Descripción de la temática de los datos.....3

Objetivo de análisis.....3

Hipótesis.....3

Datasets4

Definición del Objetivo del Proyecto5

Alcance del Proyecto5

Usuario Final y Nivel de Aplicación del Análisis5

Bases de Datos6

Transformaciones en el DataSet.....11

Implementación del Dashboard14

Solapas del Dashboard19

Diagrama de Entidad Relación en Power BI23

Futuras Líneas.....24



Descripción de la temática de los datos

El análisis se centrará en identificar las características que distinguen a los pozos más productivos en la extracción de petróleo y gas no convencionales. Para ello, se explorarán los datos publicados en la web Datos Argentina.

El conjunto de datos cuenta con variables que describen la producción de pozos desde 2006 hasta 2023, incluyendo la cantidad de petróleo y gas producido en m³, métodos de extracción, estado del pozo, la inyección de diversos elementos en los pozos, formación geológica, ubicación geográfica, entre otros.

Se busca entender las tendencias históricas en la producción de petróleo y gas no convencional y cómo diversas variables pueden influir en el rendimiento de la extracción, ya sea para su crecimiento, estabilidad o declive a lo largo de los años.

Objetivo de análisis

El objetivo principal del análisis de datos es realizar una exploración detallada para identificar las características más importantes asociadas con la alta productividad de los pozos de petróleo y gas no convencionales. Se busca evaluar si la utilización de ciertos métodos de extracción, junto con la ubicación geográfica, el yacimiento, área de permiso, formación productiva, la profundidad y estado del pozo, influyen significativamente en la productividad de los mismos a lo largo del tiempo.

Hipótesis

Se espera que los métodos de extracción, la ubicación geográfica, el yacimiento, área de permiso, formación productiva, la profundidad y estado del pozo, ejerzan un efecto significativo en la productividad de los pozos petroleros a lo largo del tiempo.



Datasets

- Link: [Datos Argentina - Producción de Pozos de Gas y Petróleo No Convencional](#)

Las modificaciones que se realizaron al dataset fueron:

- Convertir de csv a Excel utilizando la función texto en columnas, delimitado por coma.
- Reemplazar el punto por la coma en los campos numéricos para los decimales
- Reemplazar el idempresa ENE1 por ENE_1 ya que lo tomaba como fecha ENE-01
- Dar formato personalizado d/m/yyyy hh:mm al campo fechaingreso



Definición del Objetivo del Proyecto

El proyecto tiene como objetivo analizar datos de producción de petróleo y gas no convencionales en Argentina para identificar las variables que influyen en la productividad de los pozos. Se busca comprender tendencias históricas y determinar cómo factores como métodos de extracción, ubicación geográfica, yacimiento, área de permiso, formación productiva y profundidad del pozo afectan el rendimiento de la extracción, con el fin de proporcionar conocimientos clave para mejorar la eficiencia y rentabilidad en la industria petrolera.

Alcance del Proyecto

El alcance del proyecto abarcará la limpieza, análisis y visualización de datos de producción de pozos no convencionales en Argentina durante los últimos 10 años (2013-2023). Se examinarán variables como cantidad de petróleo y gas producido, métodos de extracción, estado del pozo, inyección de elementos, formación geológica y ubicación geográfica para identificar tendencias históricas y evaluar la influencia de diversas variables en el rendimiento de los pozos.

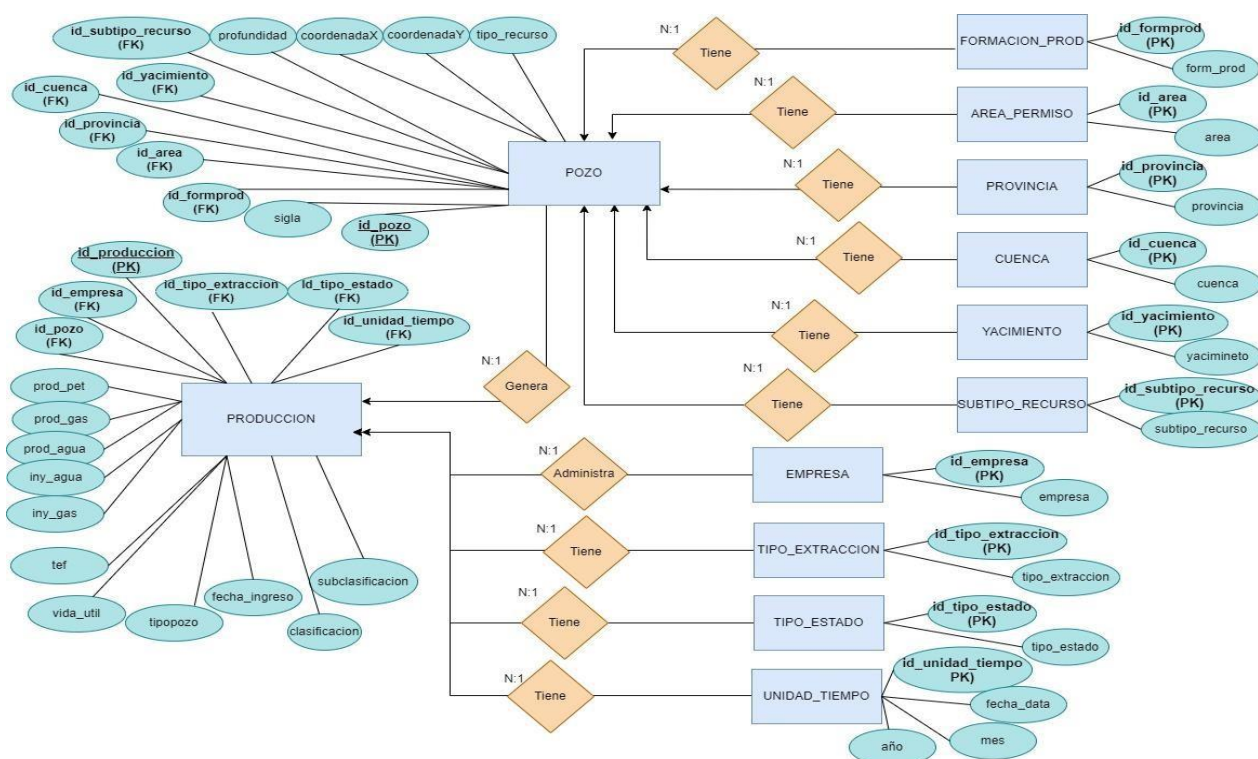
Usuario Final y Nivel de Aplicación del Análisis

El usuario final incluirá empresas del sector energético, agencias gubernamentales, consultoras especializadas, inversores y académicos.

El nivel de aplicación del análisis será estratégico, dirigido a líderes y tomadores de decisiones en la industria petrolera, así como a expertos en geología y extracción de hidrocarburos. Sin embargo, los hallazgos también serán útiles para ingenieros y operadores a nivel táctico, buscando mejorar la eficiencia operativa y productividad en las operaciones diarias de extracción.

Bases de Datos

Diagrama de entidad-relación



Listado de las tablas

1. PRODUCCION
2. EMPRESA
3. TIPO_EXTRACCION
4. TIPO_ESTADO
5. UNIDAD_TIEMPO
6. POZO
7. FORMACION_PROD
8. AREA_PERMISO
9. PROVINCIA
10. CUENCA
11. YACIMIENTO
12. SUBTIPO_RECURSO



1. PRODUCCION

Esta tabla contiene datos relacionados con la producción de pozos no convencionales de petróleo y gas. Incluye información de cantidad de petróleo, gas, agua producido por cada pozo, la fecha de producción, entre otras.

PRODUCCION		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_produccion	INT(11)
FK	id_empresa	VARCHAR(5)
FK	id_pozo	INT(11)
FK	id_tipo_extraccion	INT(11)
FK	id_tipo_estado	INT(11)
FK	id_unidad_tiempo	INT(11)
	prod_pet	FLOAT
	prod_gas	FLOAT
	prod_agua	FLOAT
	Anio	
	Mes	
	Id_cuenca	INT(11)
	Id_yacimiento	INT(11)
	Tipopozo	VARCHAR(30)
	fecha_data	DATETIME
	Clasificación	VARCHAR(30)
	Subclasificación	VARCHAR(30)

2. EMPRESA

En esta tabla se almacena información sobre las empresas que operan los pozos de petróleo y gas. Sólo se cuenta con el nombre y código de empresa.

EMPRESA		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_empresa	VARCHAR(5)
	empresa	VARCHAR(100)

3. TIPO_EXTRACCION

Esta tabla contiene los diferentes métodos para extraer hidrocarburos de formaciones Shale y Tight que son utilizados en la producción de petróleo y gas no convencionales.

TIPO_EXTRACCION		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_tipo_extraccion	INT(11)
	tipo_extraccion	VARCHAR(100)

4. TIPO_ESTADO

Aquí se registran los diferentes estados en los que pueden encontrarse los pozos de petróleo y gas. Esto podría incluir estados como "Extracción Efectiva", "Abandonado", "En Reparación", entre otros, que indican el estado operativo o la condición del pozo en un momento dado.

TIPO_ESTADO		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_tipo_estado	INT(11)
	tipo_estado	VARCHAR(100)

5. UNIDAD_TIEMPO

Esta tabla contiene registros relacionados con el tiempo en el que se cargaron los datos. Estas unidades son importantes para estandarizar y comparar la producción en diferentes períodos de tiempo.

UNIDAD_TIEMPO		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_unidad_tiempo	INT(11)
	fecha_carga	DATE
	año	INT
	mes	INT

6. POZO

Contiene información detallada sobre cada pozo individual, como su identificador único, ubicación geográfica, profundidad, latitud (coordenadaX) y longitud (coordenadaY).

POZO		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_pozo	INT(11)
	sigla	VARCHAR(100)
FK	id_formprod	VARCHAR(5)
FK	id_area	VARCHAR(5)
FK	id_provincia	VARCHAR(5)
FK	id_cuenca	VARCHAR(5)
FK	id_yacimiento	VARCHAR(5)
FK	id_subtipo_recurso	INT
	profundidad	FLOAT
	coordenadaX	FLOAT
	coordenadaY	FLOAT
	tipo_recurso	VARCHAR(30)

7. FORMACION_PROD

Esta tabla contiene datos de las formaciones geológicas de las que se extraen petróleo y gas. Sólo se cuenta con el nombre y código de formación.

FORMACION_PROD		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_formprod	VARCHAR(5)
	formprod	VARCHAR(100)

8. AREA_PERMISO

Aquí se registran las áreas geográficas para las cuales las empresas tienen permisos de exploración y producción de petróleo y gas. Sólo se cuenta con el nombre y código de área de permiso/concesión.

9. PROVINCIA

Contiene información sobre las provincias donde se encuentran ubicados los pozos de petróleo y gas. Sólo se cuenta con el nombre y código de provincia.

10. CUENCA

Describe las cuencas sedimentarias donde se encuentran ubicados los yacimientos de petróleo y gas. Sólo se cuenta con el nombre y código de Cuenca.

11. YACIMIENTO

Esta tabla proporciona información sobre los yacimientos de petróleo y gas donde se encuentran ubicados los pozos. Sólo se cuenta con el nombre y código de Yacimiento

12. SUBTIPO_RECURSO

Esta tabla contiene información de los subtipos de recursos Shale y Tight ya que el tipo de recurso para todos los pozos es NO CONVENCIONAL.

SUBTIPO_RECURSO		
Tipo Clave	Campo	Tipo de Campo
PK-Index	id_subtipo_recurso	INT
	subtipo_recurso	VARCHAR(100)

La tablas que luego van a ser importadas a POWER BI se encuentran en el siguiente archivo Excel [TABLAS PRODUCCION POZOS.xlsx](#)

Transformaciones en el DataSet

Tabla CUENCAS

```
= Table.TransformColumnTypes(CUENCAS_Table,{{"c_cuenca", type text}, {"id_cuenca", type text}})
```

Tabla PROVINCIAS

```
= Table.TransformColumnTypes("#Encabezados promovidos",{{"c_provincia", type text}, {"id_provincia", type text}})
```

Tabla TIPO_ESTADO

```
= Table.PromoteHeaders(TIPO_ESTADO_Sheet, [PromoteAllScalars=true])
```

```
= Table.TransformColumnTypes("#Encabezados promovidos",{{"c_tipoestado", type text}, {"Id_tipo_estado", Int64.Type}})
```

```
= Table.ReorderColumns("#Tipo cambiado",{"Id_tipo_estado", "c_tipoestado"})
```



Tabla TIPO_EXTRACCION

```
= Table.PromoteHeaders(TIPO_EXTRACCION_Sheet, [PromoteAllScalars=true])  
  
= Table.TransformColumnTypes("#Encabezados promovidos",{{"c_tipoextraccion", type text}, {"id_tipo_extraccion",  
    Int64.Type}})
```

Tabla UNIDAD_TIEMPO

```
= Table.PromoteHeaders(UNIDAD_TIEMPO_Sheet, [PromoteAllScalars=true])  
  
= Table.TransformColumnTypes("#Encabezados promovidos",{{"fecha_data", type date}, {"id_fecha_data", Int64.Type}})
```

Tabla YACIMIENTOS

```
= Table.TransformColumnTypes(YACIMIENTOS_Sheet,{{"Column1", type text}, {"Column2", type text}})  
  
= Table.PromoteHeaders("#Tipo cambiado", [PromoteAllScalars=true])  
  
= Table.TransformColumnTypes("#Encabezados promovidos",{{"id_areayacimiento", type text}, {"c_areayacimiento", type  
    text}})
```

Tabla POZO

```
= Table.TransformColumnTypes(POZO_Table,{{"idpozo", Int64.Type}, {"sigla", type text}, {"id_formprod", type text},  
    {"profundidad", type number}, {"id_areapermisoconcesion", type text}, {"id_areayacimiento", type text},  
    {"id_cuenca", type text}, {"id_provincia", type text}, {"coordenadax", type number}, {"coordenaday", type number},  
    {"id_tipo_recurso", Int64.Type}})  
  
= Table.SelectRows("#Tipo cambiado", each true)  
  
= Table.ReplaceValue("#Filas filtradas",-38.590069999999997,-69.41577,Replacer.ReplaceValue,{"coordenadax"})  
  
= Table.ReplaceValue("#Valor reemplazado Lat",-69.415769999999995,-38.59007,Replacer.ReplaceValue,{"coordenaday"})  
  
= Table.RenameColumns("#Valor reemplazado Lon",{{"idpozo", "id_pozo"}})
```



Tabla POZO

```
= Table.TransformColumnTypes(FORMACION_PROD_Table,{{"id_formprod", type text}, {"c_formacion", type text}})

= Table.TransformColumns(#"Tipo cambiado",{{"c_formacion", Text.Upper, type text}})
```

Tabla PRODUCCION

```
= Table.TransformColumnTypes(PRODUCCION_Table,{{"idproduccion", Int64.Type}, {"anio", Int64.Type}, {"mes", Int64.Type},
{"id_empresa", type text}, {"id_pozo", Int64.Type}, {"id_cuenca", type text}, {"id_yacimiento", type text},
{"id_areapermisiconcesion", type text}, {"prod_pet", type number}, {"prod_gas", type number}, {"prod_agua", type
number}, {"id_tipo_extraccion", Int64.Type}, {"id_estado", Int64.Type}, {"fecha_data ", type date},
{"id_subtipo_recurso", Int64.Type}, {"id_fecha_data", Int64.Type}})

= Table.SelectRows(#"Tipo cambiado", each ["fecha_data "] >= #date(2013, 1, 1) and ["fecha_data "] <= #date(2023, 12, 31))
```

Tabla TIPO_RECURSO

```
= Table.TransformColumnTypes(SUB_TIPO_RECURSO_Table,{{"subtipo_recurso", type text}, {"id_subtipo_recurso", Int64.Type}})
```

Tabla EMPRESAS

```
= Table.TransformColumnTypes(EMPRESAS_Table,{{"id_empresa", type text}, {"empresa", type text}})
```



Implementación del Dashboard

MOCKUP

En las siguientes páginas presentamos la Maqueta en Power BI de nuestro modelo de dashboard que vamos a desarrollar para la visualización del dataset de PETROAR.





PALETA DE COLORES



color theme_AdobeStock_7031635...
TEMA DE COLOR



color theme_Logo_pao_3
TEMA DE COLOR

FÓRMULAS

Se aplicaron fórmulas sólo en la tabla Pozo

TABLA POZO

Fecha de Inicio de producción de un pozo

```
Fecha_Inicio_Produccion =  
CALCULATE(  
    MIN(PRODUCCION[fecha_data ]),  
    FILTER(  
        PRODUCCION,  
        PRODUCCION[id_pozo] = pozo[id_pozo]  
    ) )
```

Fecha de la última carga de producción de un pozo

```
Fecha_Ultima_Produccion =  
CALCULATE(  
    MAX(PRODUCCION[fecha_data ]),  
    FILTER(  
        PRODUCCION,  
        PRODUCCION[id_pozo] = pozo[id_pozo]  
    ) )
```

Duración en meses de la producción de un pozo

```
Duracion_Produccion =  
DATEDIFF(  
    POZO[Fecha_Inicio_Produccion],  
    POZO[Fecha_Ultima_Produccion],  
    MONTH)
```



Producción de Gas Acumulada al inicio de la explotación de un pozo

```
Prod_Gas_Acumulado_Inicio_Pozo = CALCULATE(  
    SUM(PRODUCCION[prod_gas]),  
    FILTER(  
        ALL(Produccion),  
        Produccion[id_pozo] = pozo[id_pozo] &&  
        YEAR(PRODUCCION[fecha_data ]) = YEAR(POZO[Fecha_Inicio_Produccion])  
    )  
)
```

Producción de Gas Acumulada al final de la explotación de un pozo

```
Prod_Gas_Acumulada_Ultima_Fecha = CALCULATE(  
    SUM(PRODUCCION[prod_gas]),  
    FILTER(  
        ALL(Produccion),  
        Produccion[id_pozo] = pozo[id_pozo] &&  
        YEAR(PRODUCCION[fecha_data ]) = YEAR(POZO[Fecha_Ultima_Produccion])  
    ) )
```

Producción de Petróleo Acumulada al inicio de la explotación de un pozo

```
Prod_Pet_Acumulada_Inicio_Pozo =  
CALCULATE(  
    SUM(Produccion[prod_pet]),  
    FILTER(  
        ALL(Produccion),  
        Produccion[id_pozo] = pozo[id_pozo] &&  
        YEAR(PRODUCCION[fecha_data ]) = YEAR(POZO[Fecha_Inicio_Produccion])  
    ) )
```

Producción de Petróleo Acumulada al final de la explotación de un pozo

```
Prod_Pet_Acumulada_Ultima_Fecha = CALCULATE(  
    SUM(PRODUCCION[prod_pet]),  
    FILTER(  
        ALL(Produccion),  
        Produccion[id_pozo] = pozo[id_pozo] &&  
        YEAR(PRODUCCION[fecha_data ]) = YEAR(POZO[Fecha_Ultima_Produccion])  
    ) )
```

Rango de meses de duración de la producción de un pozo

```
Rango_Meses =
```

```

SWITCH(
    TRUE(),
    POZO[Duracion_Produccion] <= 6, "Menos de 6 meses",
    pozo[Duracion_Produccion] <= 18, "6-18 meses", pozo[Duracion_Produccion]
    <= 36, "18 - 36 meses",
    "Más de 36 meses" )

```

Último estado de un pozo

```

UltimoEstado =
VAR MaxFecha =
    CALCULATE (
        MAX ( PRODUCCION[fecha_data ] ),
        FILTER ( PRODUCCION, PRODUCCION[id_pozo] = EARLIER ( POZO[id_pozo] ) )
    )
RETURN
    MAXX (
        FILTER (
            PRODUCCION,
            PRODUCCION[id_pozo] = EARLIER ( POZO[id_pozo] ) && PRODUCCION[fecha_data
        ] = MaxFecha
        ),
        RELATED ( TIPO_ESTADO[c_tipoestado] )
    )

```

Variación de la producción de Gas de un pozo

```

Variacion_prod_gas =
POZO[Prod_Gas_Acumulada_Ultima_Fecha]POZO[Prod_Gas_Acumulado_Inicio_Pozo]

```

Variación de la producción de Petróleo de un pozo

```

Variacion_prod_pet = POZO[Prod_Pet_Acumulada_Ultima_Fecha]-
POZO[Prod_Pet_Acumulada_Inicio_Pozo]

```

MEDIDAS

```

Cantidad Empresas = COUNT(EMPRESAS[id_empresa])

```

```

Cantidad_Pozos = COUNT(POZO[id_pozo])

```

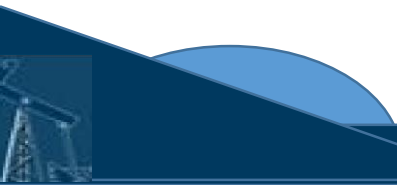
```

Mediana Profundidad_Pozo = MEDIAN(POZO[profundidad])

```



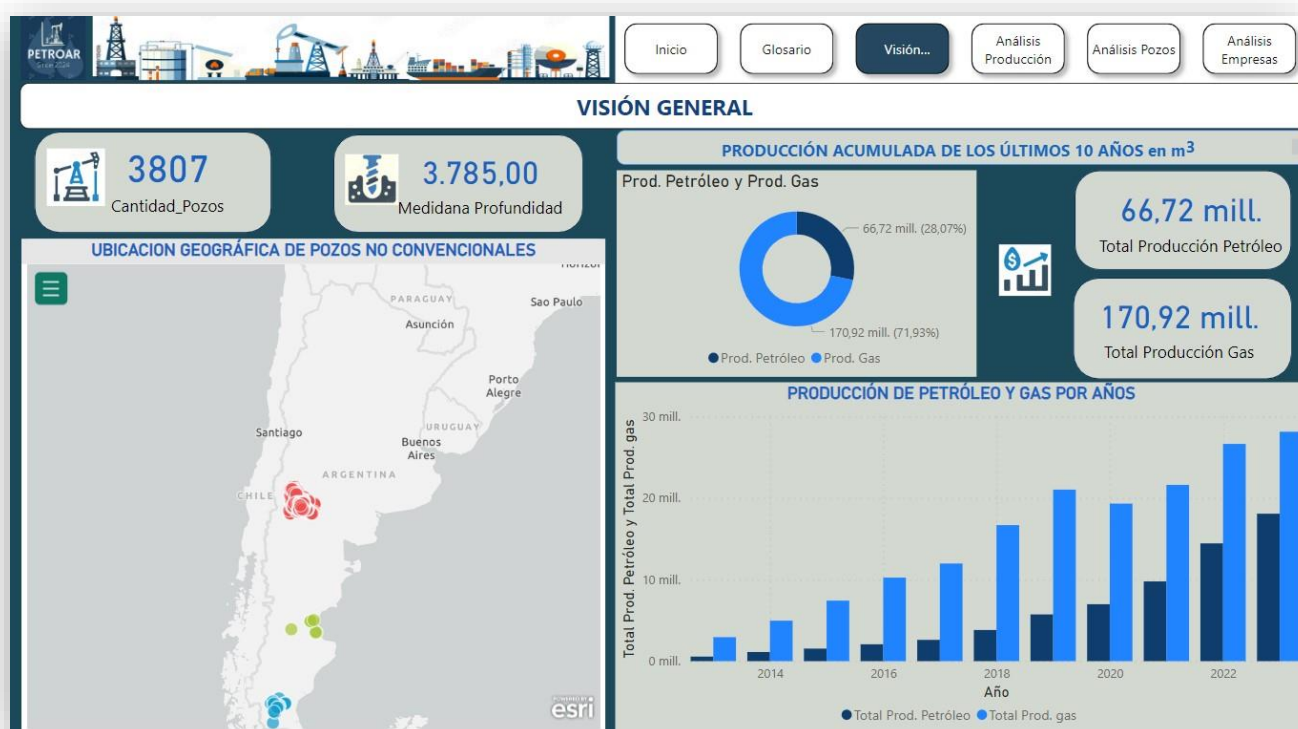
```
Pozos_Gas_Y_Petroleo =  
CALCULATE(  
    DISTINCTCOUNT(POZO[id_pozo]),  
    FILTER(  
        ALL(POZO),  
        POZO[Prod_Gas_Acumulada_Ultima_Fecha] > 0 &&  
        POZO[Prod_Pet_Acumulada_Ultima_Fecha] > 0  
    )  
)  
  
Pozos_Solo_Gas =  
CALCULATE(  
    DISTINCTCOUNT(POZO[id_pozo]),  
    FILTER(  
        ALL(POZO),  
        POZO[Prod_Gas_Acumulada_Ultima_Fecha] > 0 &&  
        POZO[Prod_Pet_Acumulada_Ultima_Fecha] = 0  
    )  
)  
  
Pozos_Solo_Petroleo =  
CALCULATE(  
    DISTINCTCOUNT(POZO[id_pozo]),  
    FILTER(  
        ALL(POZO),  
        POZO[Prod_Pet_Acumulada_Ultima_Fecha] > 0 &&  
        POZO[Prod_Gas_Acumulada_Ultima_Fecha] = 0  
    )  
)  
  
Total_Prod_Gas = SUM(PRODUCCION[prod_gas])  
  
Total_Prod_Petroleo = SUM(PRODUCCION[prod_pet])
```



Solapas del Dashboard

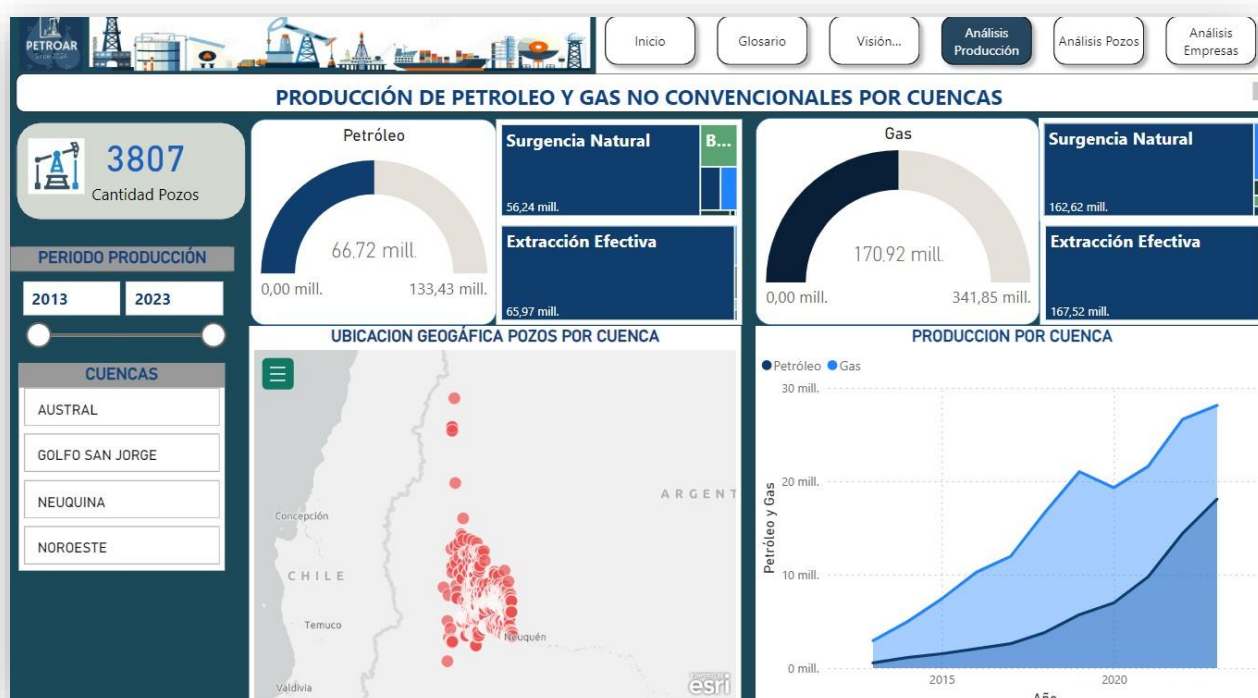
Visión General

Esta solapa nos proporciona información relevante sobre la cantidad de pozos, la mediana de profundidad de los mismos y su ubicación geográfica. Además, se ofrece una perspectiva histórica mediante la visualización de la producción acumulada de petróleo y gas en los últimos 10 años, expresada en metros cúbicos (m³). Destacando la evolución temporal de la producción, se incluye un histograma que permite apreciar el crecimiento tanto del petróleo como del gas a lo largo del período analizado.



Análisis de Producción

En esta solapa se muestra la cantidad de pozos, la producción de Petróleo y Gas, el tipo de Extracción Estado, la ubicación geográfica y la producción por Cuenca. El análisis que permite realizar es crucial para determinar qué cuenca es más productiva y cómo varían la cantidad de pozos y la producción en función de la cuenca seleccionada en la segmentación. Del mismo modo, permite explorar cómo cambia la producción según el período seleccionado para el análisis.



Análisis por Pozos

En esta solapa se examina el crecimiento en cantidad de pozos durante los últimos 10 años, así como la duración de la producción de petróleo y gas de cada pozo. Se utilizan gráficos de dispersión para visualizar estos datos, lo que permite identificar tendencias significativas en la actividad de perforación y producción.

Al observar el crecimiento en la cantidad de pozos a lo largo del tiempo, se puede apreciar la evolución de la industria petrolera y gasífera no convencional, así como el ritmo de desarrollo de nuevos yacimientos.



Análisis por Empresa

En esta solapa se proporciona la cantidad de empresas que operan en la explotación de pozos de petróleo y gas no convencionales. Se puede observar cómo varía la producción en función del período seleccionado, lo que permite comprender mejor las tendencias temporales en la actividad de las empresas. Además, se ofrece la posibilidad de explorar qué empresas operan en las distintas cuencas, lo que proporciona información valiosa sobre la distribución geográfica de la actividad empresarial en la industria.

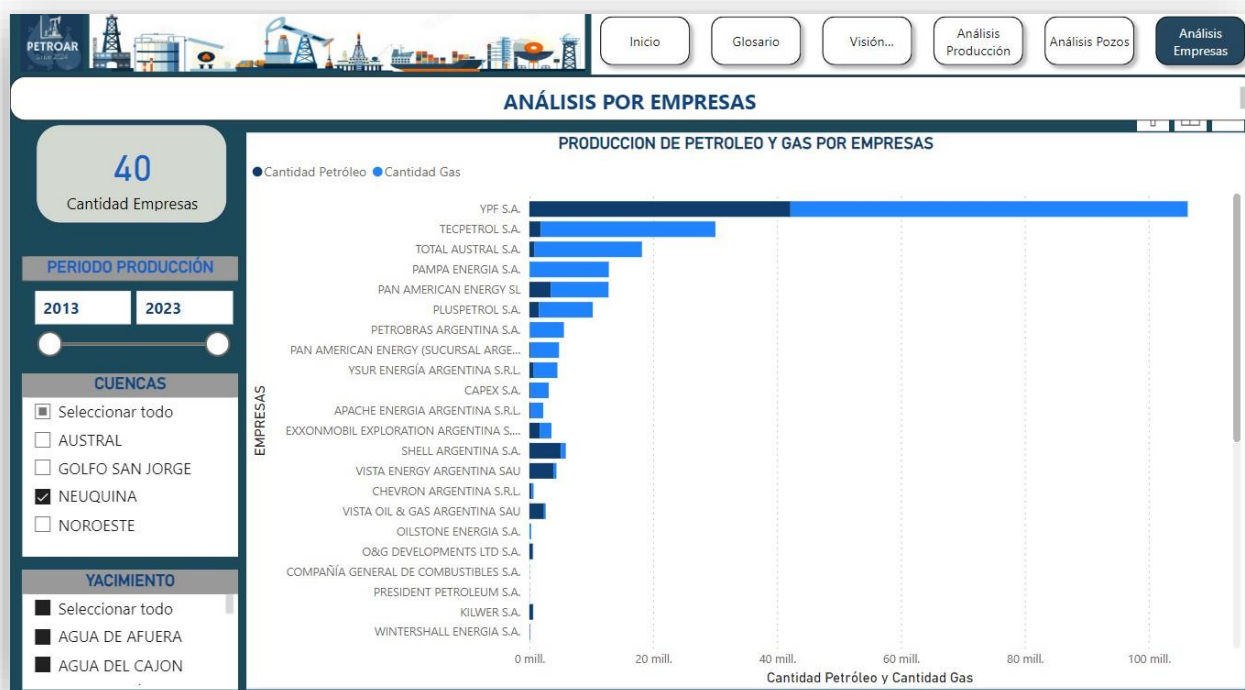
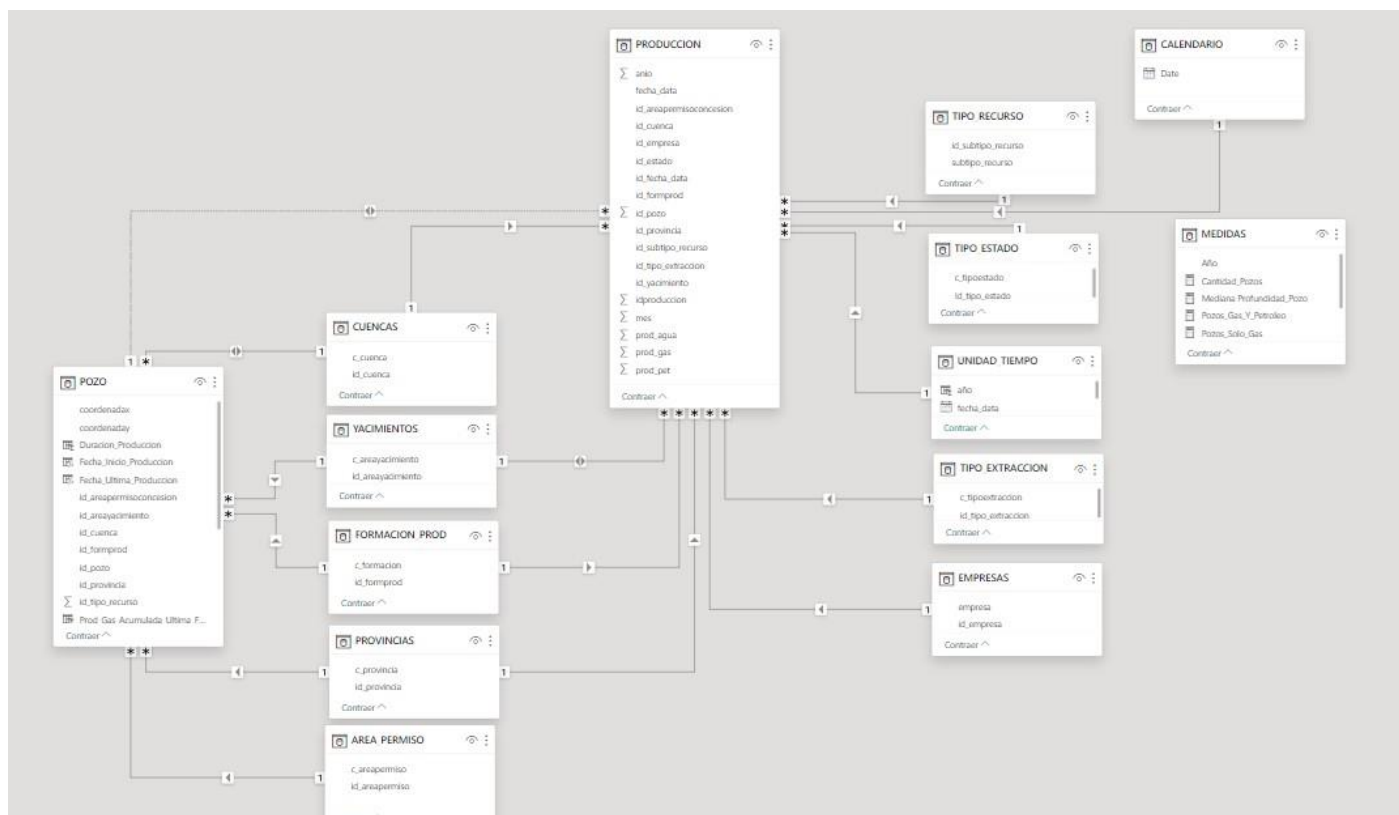


Diagrama de Entidad Relación en Power BI





Futuras Líneas

A la luz de la confirmación de nuestra hipótesis sobre los factores que afectan la productividad de los pozos petroleros y gasíferos, existen varias oportunidades para profundizar en el análisis y mejorar la comprensión de estos efectos. Algunas iniciativas sugeridas son:

- Realizar un análisis detallado de los métodos de extracción utilizados en los pozos, incluyendo la evaluación de la efectividad y eficiencia de cada técnica en diferentes condiciones geológicas y operativas.
- Investigar el impacto de la ubicación geográfica de los pozos, considerando factores como la geología local, la disponibilidad de infraestructura y la proximidad a los mercados de consumo, en la productividad y rentabilidad de los pozos.
- Estudiar cómo diferentes yacimientos y áreas de permiso afectan la productividad de los pozos, identificando patrones geológicos y operativos que puedan influir en los resultados de producción.
- Evaluar el papel de la formación productiva en la productividad de los pozos, incluyendo la caracterización de las formaciones geológicas y la identificación de factores que puedan mejorar o limitar la extracción de hidrocarburos.
- Investigar la relación entre la profundidad de los pozos y su productividad, examinando cómo la profundidad afecta la accesibilidad a los recursos, los costos de perforación y la eficiencia de extracción.
- Analizar cómo el estado del pozo, incluyendo su edad, mantenimiento y estado de operación, influye en su productividad y vida útil, y desarrollar estrategias para optimizar la gestión y mantenimiento de los pozos en función de estos factores.