

# Análisis de distribución de petróleo

CODERHOUSE  
DATA ANALYTICS  
COMISION #44725  
ENTREGA FINAL

AUTOR:  
SIDONI, OCTAVIO

## Contenido del proyecto

Introducción .....	2
Objetivos .....	3
Tipos de análisis .....	3
Nuestra meta es .....	3
Pasos a seguir para desarrollar un análisis descriptivo .....	3
Alcance .....	3
Usuario Final.....	3
Base de datos .....	4
Informe del Dataset .....	4
Diagrama de Entidad Relación .....	4
Listado de tablas.....	5
Tabla N°1 .....	5
Tabla N°2 .....	5
Tabla N°3 .....	5
Tabla N°4: .....	5
Detalle de columnas por tablas.....	6
Visualización .....	7
Solapas .....	7
1. General .....	7
2. Análisis por Provincia .....	8
3. Análisis por Empresa .....	9
Transformaciones Realizadas .....	11
Modelo .....	14
Medidas Calculadas generadas y sus formulas .....	14
Columnas Calculadas .....	14
Tabla Puente .....	15
Tabla Calendario.....	15
Conclusión .....	17

## Introducción

El destino final del petróleo y sus derivados es el consumidor final. En el proceso intervienen distribuidores mayoristas y minoristas, y se emplean todos los medios posibles para el transporte y venta. Se agrupan bajo esos vocablos las operaciones finales, pero no las más delicadas de la industria petrolífera, que consiste en transportar los productos salidos de la refinería, almacenarlos en depósitos y puntos de ventas y, por último, expenderlos a los clientes gracias a una red de comercialización que cubra el conjunto del territorio nacional.

Para el presente trabajo se extrajo información del Ministerio de Energía y Minería. Secretaría de Planeamiento Energético Estratégico acerca de la distribución nacional del petróleo.

## Objetivos

- Evolución de la cantidad distribuida por provincia y por región de los últimos 10 años
- Evolución mensual comparando con los últimos 3 años, ya que la base se actualiza todos los meses.
- Clasificación/tipo de distribución (esto hace referencia cual es el destino de la distribución)

Los objetivos han sido planteados siguiendo la metodología (SMART – Específicos, Medibles, Alcanzables, Relevantes, Tiempo)

## Tipos de análisis

El tipo de análisis a implementar es el “Descriptivo”. Por lo cual podemos sintetizar los datos históricos para obtener información útil acerca de la distribución nacional del petróleo.

Nuestra meta es:

- Recolectar y ordenar la información por medio de gráficos y medios visuales.
- Extraer las características más representativas de la base de datos.
- Describir tendencias.
- Filtrar los datos menos significativos

Pasos a seguir para desarrollar un análisis descriptivo

- Obtener el dataset
- Limpiar datos
- Calcular medidas simples de composición y distribución de variables

## Alcance

- Realizamos este trabajo para poder observar cómo es el comportamiento de la distribución del petróleo en Argentina, ya que es un recurso importantísimo a nivel país y de uso diario.
- Nos encontramos con limitaciones, ya que la base de datos que obtuvimos solo tiene información de los datos históricos de la distribución del petróleo desde el año 2009 a la actualidad, por lo tanto, no pudimos hacer un análisis más profundo acerca del destino final del petróleo (tratamiento nacional, exportación, comercialización interna, etc.).
- Este proyecto sirve como partida si se quiere realizar un análisis más profundo en términos económicos, ambiental, social.

## Usuario Final

Toda persona interesada y/o relacionada con el tema de estudio.

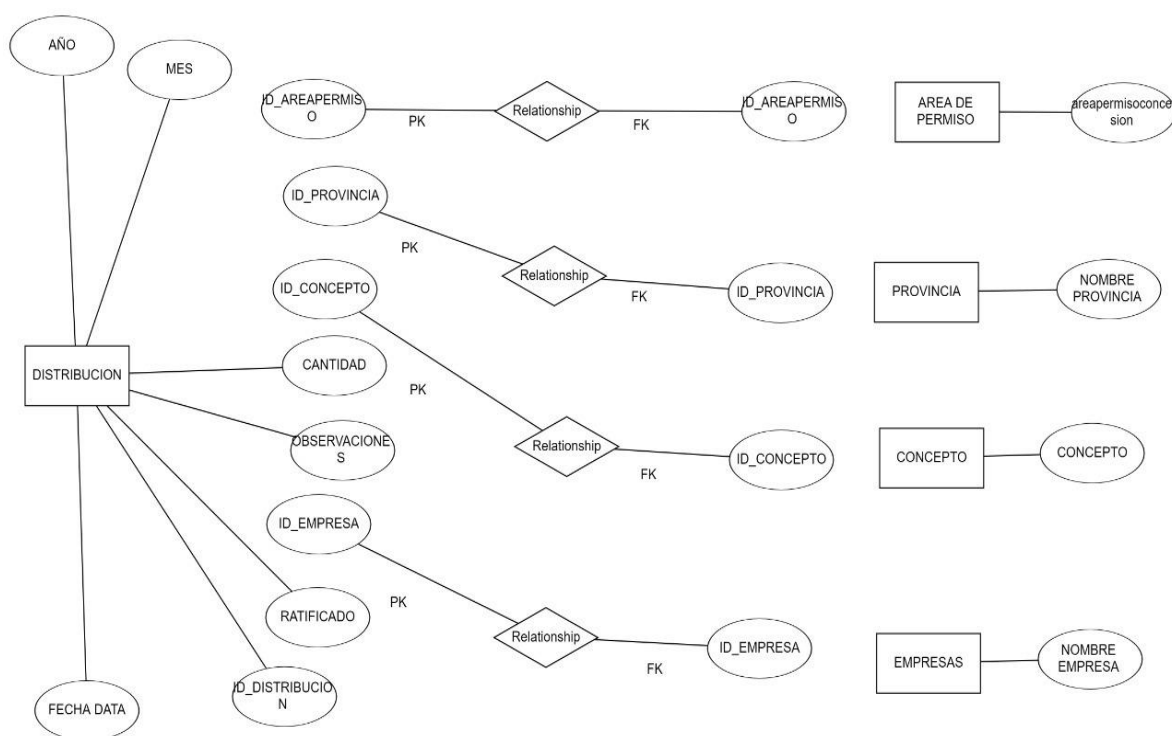
## Base de datos

El dataset seleccionado se puede consultar en la página <https://datos.gob.ar/dataset/energia-distribucion-petroleo> bajo el nombre de "Distribución del petróleo"

### Informe del Dataset

<b>Nombre</b>	<i>Distribución del petróleo</i>
<b>Frecuencia de actualización</b>	<i>Mensual</i>
<b>Fecha de Publicación</b>	<i>20 de enero de 2017</i>
<b>Fecha de Actualización</b>	<i>20 de agosto de 2022</i>
<b>Contenido</b>	<i>Datos históricos de la distribución del petróleo desde el año 2009 a la actualidad</i>
<b>Página de Descarga</b>	<a href="https://datos.gob.ar/dataset/energia-distribucion-petroleo">https://datos.gob.ar/dataset/energia-distribucion-petroleo</a>

### Diagrama de Entidad Relación



## Listado de tablas

**Tabla N°1: DistribucionPetroleo:** esta tabla contiene la fecha (detallado por mes y año) y la cantidad de petróleo distribuido a lo largo del país, detallado por empresa, provincia, área de distribución y si el producto ha sido rectificado o no.

- año
- mes
- PK: idareapermisococoncesion
- PK: idprovincia
- cantidad
- observaciones
- PK: idempresa
- rectificado
- iddistribucion
- fecha\_data

**Tabla N°2: EMPRESA:** esta tabla contiene el id y el nombre de cada empresa detallada.

- FK: ID\_EMPRESA
- NombreEmpresa

**Tabla N°3: Provincia:** esta tabla contiene el id y el nombre de provincia detallada.

- FK: ID\_PROVINCIA
- Provincia

**Tabla N°4: CONCEPTO:** esta tabla contiene el id y concepto del destino detallado de la distribución del petróleo.

- FK: ID\_CONCPETO
- Concepto

**Tabla N°5: AREAPERMISO:** esta tabla contiene el id y el nombre del área de permiso de cada distribución del petróleo.

- FK: ID\_AREAPERMISO
- AreaPermiso

## Detalle de columnas por tablas

**DistribucionPetroleo**

Campo	Tipo de Campo	Tipo de Clave
año	INT	-
mes	INT	-
ID_AREAPERMISO	VARCHAR(10)	PK
ID_PROVINCIA	CHAR(5)	PK
ID_CONCPETO	INT	PK
cantidad	INT	-
observaciones	TEXT(150)	-
ID_EMPRESA	INT	PK
rectificado	CHAR(5)	-
iddistribucion	INT	-
fecha_data	DATE	-

**Empresa**

Campo	Tipo de Campo	Tipo de Clave
ID_EMPRESA	INT	FK
NombreEmpresa	TEXT(150)	-

**Provincia**

Campo	Tipo de Campo	Tipo de Clave
ID_PROVINCIA	CHAR(5)	FK
Provincia	TEXT(150)	-

**Concepto**

Campo	Tipo de Campo	Tipo de Clave
ID_CONCEPTO	INT	FK
Concepto	VARCHAR(40)	-

**AreaPermiso**

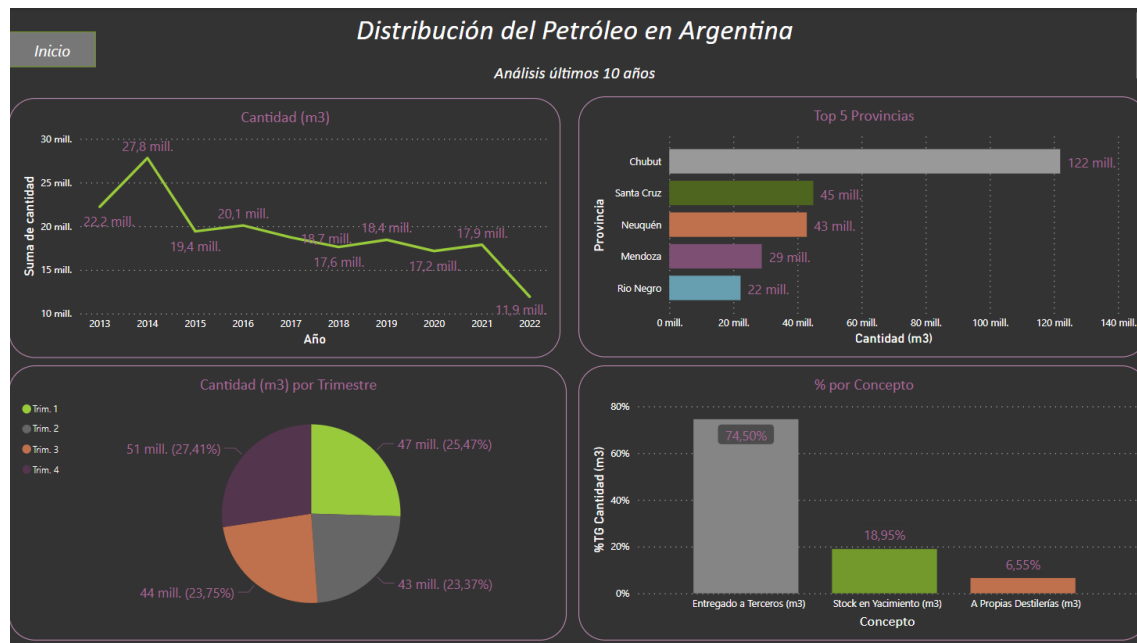
Campo	Tipo de Campo	Tipo de Clave
ID_AREAPERMISO	VARCHAR(10)	FK
AreaPermiso	VARCHAR(50)	-

## Visualización

### Solapas

#### 1. General

En la solapa General se analiza cómo es la variación de distribución de petróleo, estudiando los últimos 10 años. La unidad de medida de los valores expresados es el m3 (metro cubico).



#### Reseña por grafico

- Gráfico 1: Gráfico de Líneas; muestra la cantidad de m3 distribuidos por año.
- Gráfico 2: Gráfico de barra apiladas; muestra las 5 provincias con mayor cantidad de petróleo distribuido en los últimos 10 años.
- Gráfico 3: Gráfico circular; muestra la cantidad de m3 distribuidos (expresada en cantidad y porcentaje de participación.) por trimestre en los últimos 10 años
- Gráfico 4: Gráfico de columnas apiladas y líneas; muestra el porcentaje de participación de cada concepto de distribución (Entregados a terceros, stock en yacimientos y a propias destilerías.)

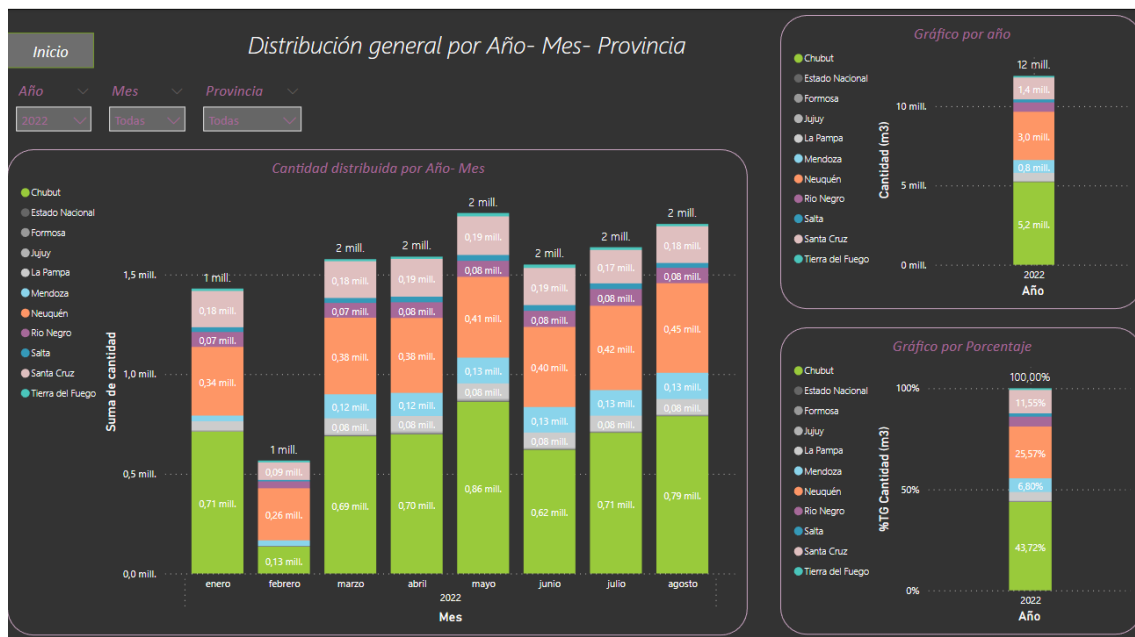
#### Segmentaciones Solapa General

- En esta hoja utilizamos una segmentación general donde todos los gráficos contiene datos de los últimos 10 años de análisis.
- El Gráfico que contiene la información de las provincias usamos una segmentación para mostrar el Top 5 de las empresas con mayor cantidad de distribución por m3.



## 2. Análisis por Provincia

En la solapa “Análisis por Provincia” se analiza cómo es la variación de distribución de petróleo, analizado por año-mes-provincia. La unidad de medida de los valores expresados es el m3 (metro cubico).



### Reseña por gráfico

- Gráfico 1: Gráfico de columnas apiladas; muestra la cantidad de m3 distribuidos por cada provincia en cada mes del año.
- Gráfico 2: Gráfico de columnas apiladas; muestra la cantidad de m3 distribuidos por cada provincia en el año de análisis.
- Gráfico 3: Gráfico de columnas apiladas; muestra el porcentaje de participación de cada provincia en el año de análisis.

### Segmentaciones Solapa “Análisis por Provincia”

**Año**  **Mes**  **Provincia**

- Segmentación 1 “AÑOS”

Con esta segmentación se puede elegir el año que se desea analizar, permite una sola selección.

- Segmentación 2 “Mes”

Con esta segmentación se puede elegir el mes que se desea analizar, permite una sola selección.

- Segmentación 3 “Provincia”

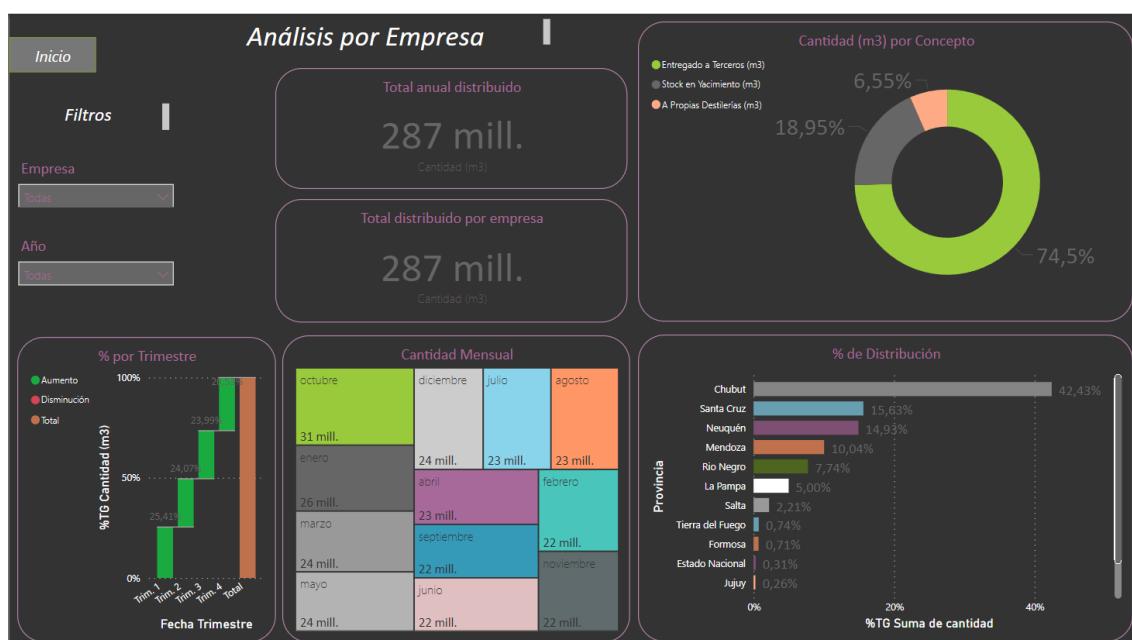
Con esta segmentación se puede elegir la provincia que se desea analizar, permite una sola selección.

Estas 3 segmentaciones se pueden combinar para un análisis más detallado.

Las segmentaciones están relacionadas con todas las visualizaciones de la hoja.

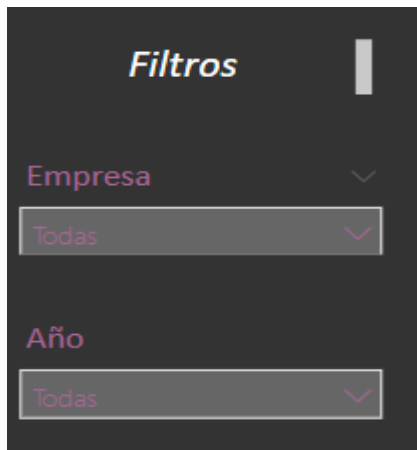
### 3. Análisis por Empresa

En la solapa “Análisis por Empresa” se analiza cómo es la variación de distribución de petróleo analizado por empresa, a su vez se puede analizar el desempeño de la empresa a tanto a nivel histórico como nivel de año específico. La unidad de medida de los valores expresados es el m3 (metro cubico).



#### Reseña por Grafico

- Gráfico 1: Tarjeta; indica la cantidad anual distribuida.
- Gráfico 2: Tarjeta; indica la cantidad anual distribuida por empresa.
- Gráfico 3: Gráfico de anillos; muestra el porcentaje de participación de cada concepto de distribución.
- Gráfico 4: Gráfico de cascada; muestra el crecimiento de distribución por trimestre.
- Gráfico 5: Treemap; muestra la cantidad distribuida por meses.
- Gráfico 6: Gráfico de barras apiladas, muestra la participación de cada provincia en la distribución.



- Segmentación 1 “Empresa”

Con esta segmentación se puede elegir la empresa que se desea analizar, permite una sola selección.

- Segmentación 2 “Años”

Con esta segmentación se puede elegir el año que se desea analizar, permite una sola selección.

La segmentación empresa se relaciona con todas las visualizaciones de la hoja.

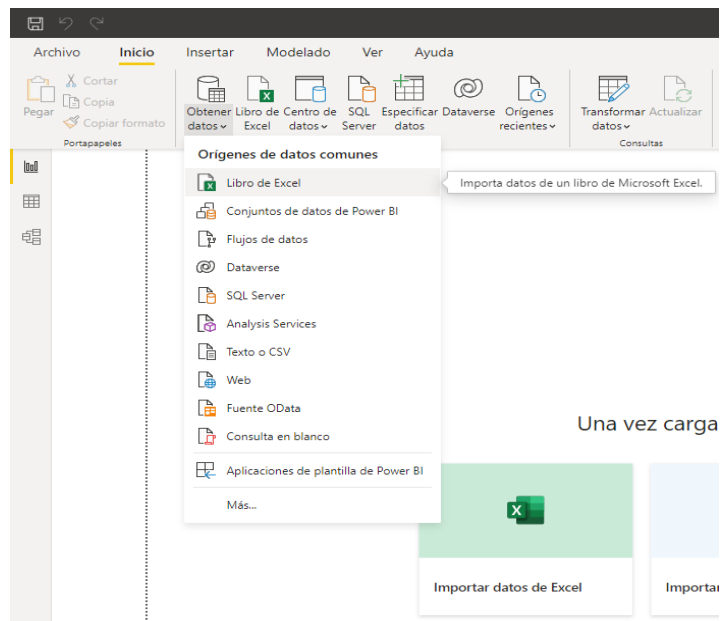
La segmentación año se relaciona con todas las visualizaciones de la hoja, excepto la tarjeta de visualización llamada “Total anual distribuido.”

Estas dos segmentaciones se pueden combinar para un análisis más detallado.

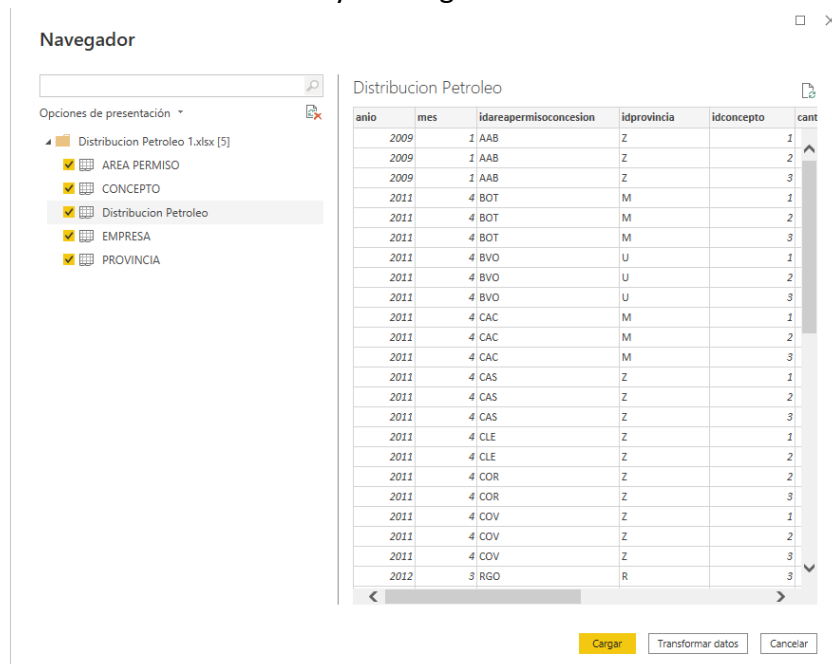
## Transformaciones Realizadas

La particularidad de este trabajo es que nosotros obtuvimos los datos para la realización del tablero desde un archivo de Excel.

### 1. Obtención de datos a través de un libro de Excel



### 2. Seleccionamos los datos y los cargamos en Power Bi



### 3. Datos cargados

**Campos** >>

🔍 Buscar

> 📊 AREA PERMISO

> 📊 CONCEPTO

> 📊 Distribucion Petroleo

> 📊 EMPRESA

> 📊 PROVINCIA

- Ahora necesitamos relacionar las tablas, para lograr esto ejecutamos consultadas combinadas dentro de Power Query.
- Realizamos un total de cuatro consultas combinadas que vamos a detallar a continuación

Nos dirigimos a nuestra tabla principal llamada “Distribucion Petroleo” (hacemos las transformaciones de formatos de datos necesarias, Por ej, promover encabezados, quitar errores, filtrar columnas vacías, etc.)  
Elegimos la opción combinar consultas.

**Combinar** ✕

Seleccione una tabla y las columnas coincidentes para crear una tabla combinada.

Distribucion Petroleo

anio	mes	idareapermisoconcesion	idprovincia	idconcepto	cantidad	observaciones	idempresa
2009	1	AAB	Z	1	0	null	167
2009	1	AAB	Z	2	89	null	167
2009	1	AAB	Z	3	0	null	167
2011	4	BOT	M	1	0	null	1274

PROVINCIA

ID_PROVNCIA	PROVINCIA
Z	Santa Cruz
Z	Santa Cruz
Z	Santa Cruz
M	Mendoza
M	Mendoza

Tipo de combinación

Interna (todas las filas coincidentes)

☐ Use las coincidencias aproximadas para comparar la combinación.

▷ Opciones de coincidencia aproximada

✓ La selección coincide con 83406 de 83406 filas de la primera tabla y con 8...

Aceptar

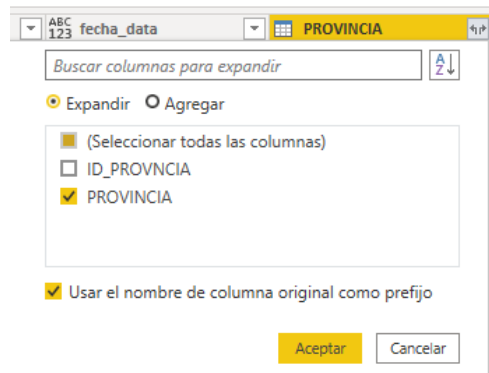
Cancelar

En la imagen podemos observar que combinamos la columna “idprovincia” de la tabla “Distribucion Petroleo” con la columna “ID\_PROVINCIA” de la tabla “PROVINCIA”, elegimos el tipo de combinación interna para que solo nos muestre los valores que se relacionan por igual en las dos tablas (se asemeja a la función Inner Join en SQL).

6. Aceptamos la combinación.

Para que la combinación se ejecute sin problemas las dos columnas relacionadas tienen que estar con el mismo formato de datos. Esto logra una combinación de 1 (tabla izquierda) a \* (tabla derecha)

7. Una vez realizada la combinación, expandimos la misma para elegir que columna de la tabla derecha queremos mostrar, en este caso mostraremos la columna “PROVINCIA”



Estos pasos se repiten para las tres combinaciones restantes.

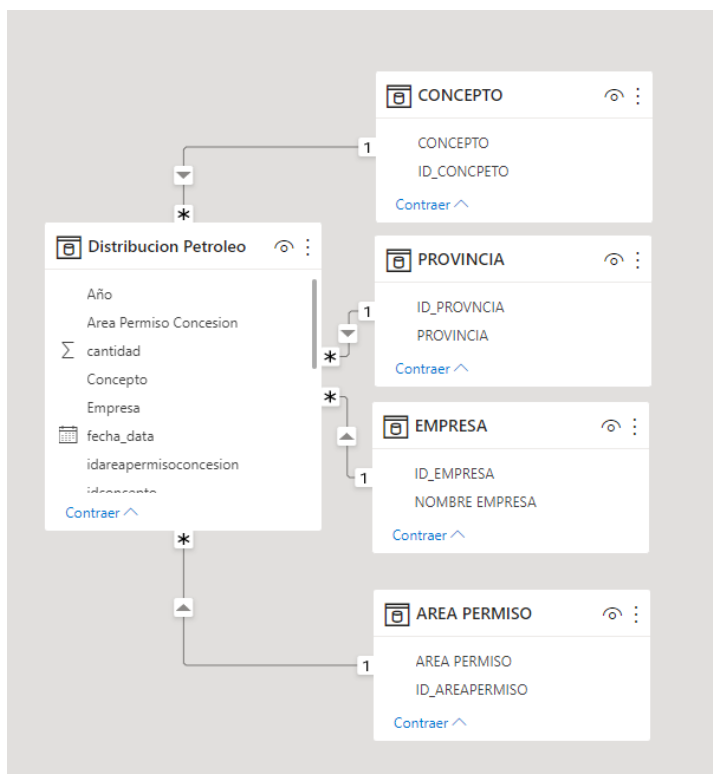
Columna “idconcepto” de la tabla “Distribucion Petroleo” con la columna “ID\_CONCEPTO” de la tabla “CONCEPTO”

Columna “idempresa” de la tabla “Distribucion Petroleo” con la columna “ID\_EMPRESA” de la tabla “EMPRESA”

Columna “idareapermisiconcesion” de la tabla “Distribucion Petroleo” con la columna “ID\_AREAPERMISO” de la tabla “AREA PERMISO”

Gracias a las consultas combinadas podemos relacionar las tablas secundarias con la principal

## Modelo



## Medidas Calculadas generadas y sus formulas

- **Cantidad (m3):** medida utilizada para obtener la sumatoria de los m3 de petróleo distribuidos. Esta medida nos sirvió para analizar la distribución general, por provincia y por empresas  
Formula Dax :  $\text{SUM('Distribucion Petroleo'[cantidad])}$
- **Cantidad de Distribuciones:** medida utilizada para obtener el número de distribuciones realizadas, es decir, cuenta la cantidad de veces que se distribuyó el petróleo.  
Formula Dax:  $\text{Cantidad de Distribuciones} = \text{COUNTA('Distribucion Petroleo'[Año])}$

## Columnas Calculadas

Realizamos varias columnas calculadas para poder generar dos tablas adicionales, tabla "Puente" y tabla "Calendario"

### Tabla Puente

Para obtener esta tabla realizamos los pasos siguientes:

1. Transformar datos
2. Duplicar consulta Distribución petróleo
3. Seleccionamos la columna que necesitamos (fecha data) y eliminamos las restantes.
4. Renombramos la columna fecha data ---> fecha
5. Transformamos la columna de formato general a fecha corta
6. Eliminamos los duplicados de la columna fecha

En la tabla puente calculamos 3 columnas utilizando funciones Dax

- Año = YEAR('Tabla Puente'[Fecha])
- Mes = MONTH('Tabla Puente'[Fecha])
- Trimestre = QUARTER('Tabla Puente'[Fecha])

### Tabla Calendario

Para obtener esta tabla realizamos los pasos siguientes:

1. Transformar datos
2. Duplicar consulta Distribución petróleo
3. Seleccionamos la columna que necesitamos (fecha data) y eliminamos las restantes.
4. Transformamos la columna de formato general a fecha corta
5. Eliminamos los duplicados de la columna fecha

En la tabla Calendario calculamos 3 columnas utilizando funciones Dax

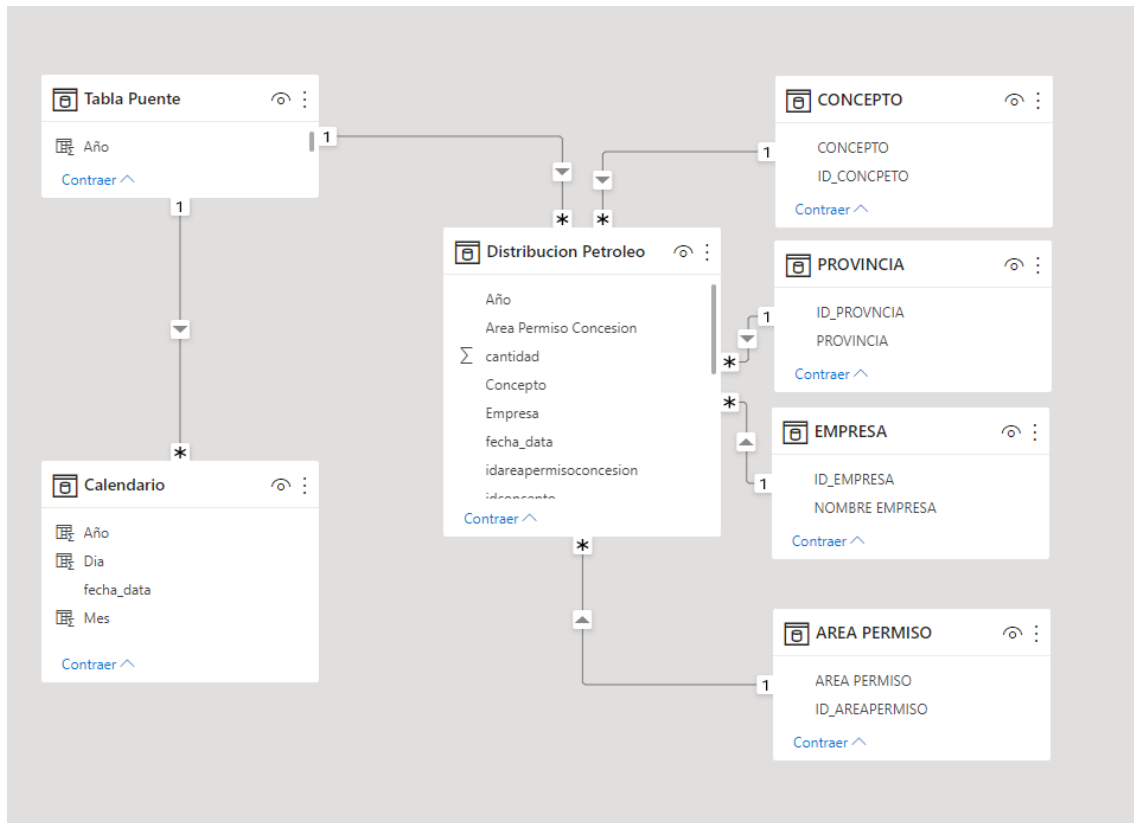
- Año = YEAR('Calendario'[fecha\_data])
- Mes = MONTH('Calendario'[fecha\_data])
- Día = DAY('Calendario'[fecha\_data])

El objetivo de crear las tablas mencionadas anteriormente es para poder relacionar nuestro conjunto de fechas de cualquier tabla con una tabla puente.

Para lograr esto eliminamos los duplicados de la tabla puente así podíamos lograr una relación de 1 a \* (muchos) con cualquier tabla que contenga fecha.

A continuación, adjuntamos una imagen donde se ve reflejado lo detallado anteriormente





## Conclusión

Concluido el proyecto podemos observar cómo se comporta la distribución del petróleo a nivel nacional.

Identificamos parámetros que se repiten a través de los años:

- Poca variación de la cantidad distribuida por años
- Poca variación de la cantidad distribuida si se analiza por trimestre.
- Top 5 provincias que más distribuyen se mantienen constante.
- El destino final (Entregados a terceros) con mayor repetición se mantiene constante
- Empresas con mayor participación se mantienen constante.