UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE COMERCIO Y TURISMO

MASTER EN BIG DATA & DATA SCIENCE

APLICACIONES AL COMERCIO, EMPRESA Y FINANZAS



TRABAJO FINAL DE MÁSTER

2022 – 2023

**MODELO DE ESTIMACIÓN DE LA DEMANDA DE ENERGÍA O ELECTRICIDAD SECTORIAL DE UN PAÍS**

Autores: (GRUPO 1)

* Ruth Fernández Padilla
* Ángel Martínez Barrial
* Alejandro Minguez Bonache
* Eduardo Urrutia Rivas
* Nam Nguyen Thi
* Antonio Villamayor Delgado

Supervisor:

* Tutor

Septiembre, 2023

**Declaration of Authorship**

MRC Consultants and Transaction Advisers

Signed:

Date: September, 2023

*“Los datos no son solo información, son la clave para el conocimiento y la toma de decisiones informadas en la era digital”*

Peter Drucker

**Abstract**

The purpose of this project was to predict energy consumption in Brazil up to year 2040 using data from matrices and balances published on the official website of OLADE. The objective of this project was to test whether a machine learning model could produce sufficiently accurate results for a complex forecasting problem. We explored various machine learning techniques and developed a data-driven model for energy consumption forecasting. The dataset included a 52-year record (1970-2021), and we used an ARIMA model to train the data. The model's performance was evaluated using the Mean Absolute Percentage Error (MAPE) to ensure direct comparability with energy readings in the dataset. The results indicate that machine learning algorithms can effectively predict energy consumption. These findings can be applied to other countries in South America for energy consumption forecasting.

**Resumen**

El objetivo de este proyecto fue predecir el consumo de energía en Brasil hasta el año 2040 utilizando datos de matrices y balances publicados en el sitio web oficial de OLADE. El objetivo de este proyecto era probar si un modelo de aprendizaje automático podría producir resultados suficientemente precisos para un problema de pronóstico complejo. Exploramos varias técnicas de aprendizaje automático y desarrollamos un modelo basado en datos para la previsión del consumo de energía. El conjunto de datos incluyó un registro de 52 años (1970-2021) y utilizamos un modelo ARIMA para entrenar los datos. El rendimiento del modelo se evaluó utilizando el error porcentual absoluto medio (MAPE: en sus siglas en inglés Mean Absolute Percentage Error) para garantizar la comparabilidad directa con las lecturas de energía en el conjunto de datos. Los resultados indican que los algoritmos de aprendizaje automático pueden predecir eficazmente el consumo de energía. Estos hallazgos se pueden aplicar a otros países de América del Sur para pronosticar el consumo de energía.

**INDICE**

[**1. Introducción 1**](#_heading=h.gjdgxs)

[**2. Fundamentos 2**](#_heading=h.30j0zll)

[**2.1 Series de tiempo** 2](#_heading=h.1fob9te)

[**2.2 Características de las series temporales** 2](#_heading=h.3znysh7)

[**2.3 Métodos de pronóstico de series temporales** 2](#_heading=h.2et92p0)

[**2.4 ARIMA (p,d,q)** 2](#_heading=h.tyjcwt)

[**3. Procesamiento de datos 3**](#_heading=h.3dy6vkm)

[**3.1. Datos** 3](#_heading=h.1t3h5sf)

[**3.2 Preprocesamiento de datos** 5](#_heading=h.4d34og8)

[**3.3 Visualización de los datos originales** 6](#_heading=h.2s8eyo1)

[**3.4 Exploración de datos** 7](#_heading=h.17dp8vu)

[**4. Machine Learning 8**](#_heading=h.3rdcrjn)

[**4.1 Modelo ARIMA (Univariado)** 8](#_heading=h.26in1rg)

[**4.4.1 ARIMA para todos los sectores** 8](#_heading=h.lnxbz9)

[**4.4.2 Entrenamiento, validación y predicción del modelo** 9](#_heading=h.35nkun2)

[**4.4.3 Escenarios** 11](#_heading=h.1ksv4uv)

[**4. Análisis y visualización de los resultados 15**](#_heading=h.44sinio)

[Estimaciones y predicciones de la demanda de energía sectorial 15](#_heading=h.2jxsxqh)

[Visualización y aplicación interactiva 15](#_heading=h.z337ya)

[**5. Conclusiones 15**](#_heading=h.3j2qqm3)

[Conclusiones obtenidas 15](#_heading=h.1y810tw)

[Limitaciones y futuras líneas investigaciones 15](#_heading=h.4i7ojhp)

[**6. Recomendaciones . 15**](#_heading=h.2xcytpi)

[**Referencias bibliográficas/ Anexos 15**](#_heading=h.1ci93xb)

# **1. Introducción**

La demanda mundial de energía para el consumo de energía aumenta día a día, y parece un desafío para la mayoría de los países satisfacer la demanda de energía con la producción total de energía.

En un mundo cada vez más tenso, los países BRICS (Brasil, Rusia, India, China y Sudáfrica) son muy complementarios y son los principales productores de energía del mundo, así como los principales consumidores de combustible.

El sector energético brasileño ha sido testigo de numerosos cambios tecnológicos y ha evolucionado hasta convertirse en líder mundial en ventas de tecnologías limpias, tanto en el mercado interno como externo. Muchos factores contribuyeron a las actividades innovadoras en su sector eléctrico, que incluyen la contribución tanto del gobierno como de la IED.

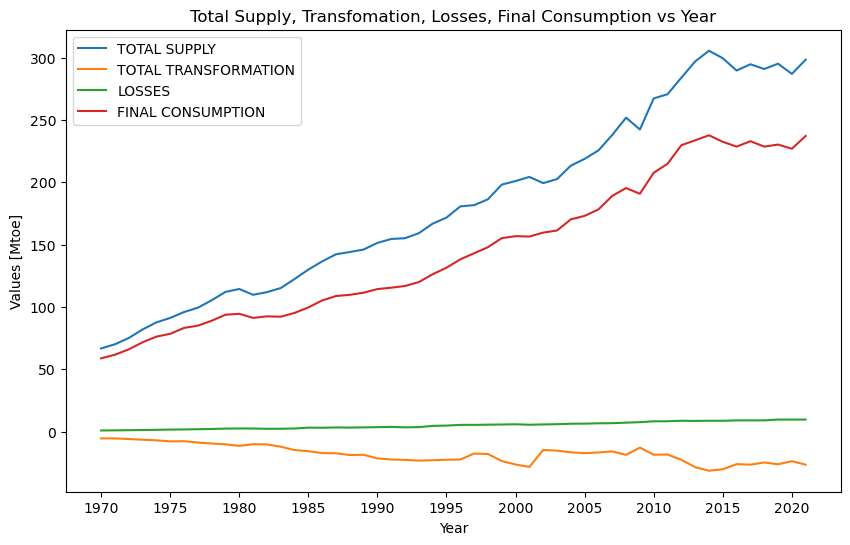


Gráfico 1. Total Supply, Transformation, Losses, Final Consumption por año

En los últimos años, la inteligencia artificial en general y el aprendizaje automático en concreto presentan soluciones a una enorme cantidad de problemas. Una de las aplicaciones prometedoras es predecir el consumo energético futuro utilizando series temporales históricas adquiriendo su modelo de imitación.

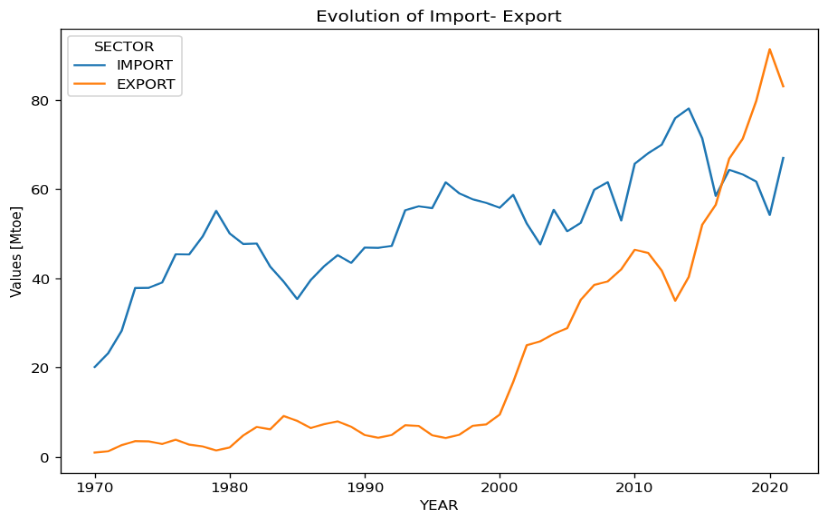


Gráfico 2. Evolución de importación y exportación de energía

Este proyecto tiene como objetivo probar si es posible aplicar el modelo ARIMA a los datos de series temporales. El objetivo es obtener resultados suficientemente buenos en la previsión del consumo de energía.

# **2. Fundamentos**

## **2.1 Series de tiempo**

Los datos de series de tiempo rastrean los valores durante un período específico, generalmente con el tiempo como variable independiente, a menudo utilizada para pronósticos futuros. Se puede clasificar en series de tiempo continuas y series de tiempo discretas según los intervalos de tiempo. Además, los datos de series de tiempo se pueden dividir en univariados (seguimiento de una sola variable a lo largo del tiempo) y multivariados (seguimiento de múltiples variables a lo largo del tiempo) según la cantidad de variables dependientes.

## **2.2 Características de las series temporales**

Las series temporales tienen tres características claves y cruciales para realizar pronósticos precisos:

* **Autocorrelación:** Mide la relación entre el valor actual de una variable y sus valores pasados.
* **Estacionalidad:** Se refiere a variaciones regulares y de corto plazo que ocurren, como patrones semanales, mensuales o trimestrales.
* **Estacionariedad:** Implica que las propiedades estadísticas de una serie temporal permanecen constantes a lo largo del tiempo, con una media y varianza consistentes y una covarianza independiente del tiempo.

## **2.3 Métodos de pronóstico de series temporales**

A lo largo de los años, numerosos estudios han evaluado el rendimiento de los métodos clásicos y de aprendizaje automático. Algunos métodos clásicos bien conocidos incluyen:

* Auto-Regressive Moving Average (ARMA)
* Auto-Regressive Integrated Moving Average (ARIMA)
* Seasonal Autoregressive Integrated Moving-Average (SARIMA)

Mientras que algoritmos modernos como:

* Multi-Layer Perceptron (MLP)
* Bayesian Neural Networks (BNN)
* Generalized Regression Neural Networks (GRNN).

Otros dos algoritmos modernos son:

* Recurrent Neural Networks (RNN)
* Long Short- Term Memory (LSTM)-Brownlee, 2016

## **2.4 ARIMA (p,d,q)**

ARIMA significa media móvil integrada autorregresiva en sus siglas en inglés Auto-Regressive Integrated Moving Average. Es una forma de modelar datos de series de tiempo para realizar pronósticos (es decir, para predecir puntos futuros en la serie), de tal manera que:

• **AR**: Autoregresión. Modelo que utiliza la relación de dependencia entre una observación y un cierto número de observaciones rezagadas.

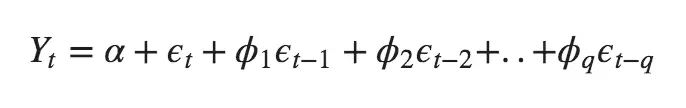
• **I**: Integrado. El uso de diferenciación de observaciones sin procesar (por ejemplo, restar una observación de una observación en el paso de tiempo anterior) para hacer que la serie de tiempo sea estacionaria.

• **MA**: Media Móvil o en sus siglas en inglés *Moving Average*. Un modelo que utiliza la dependencia entre una observación y un error residual de un modelo de promedio móvil aplicado a observaciones rezagadas.

Los parámetros del modelo ARIMA se definen de la siguiente manera:

* **p:** El número de observaciones de retraso incluidas en el modelo, también llamado orden de retraso.
* **d:** El número de veces que se diferencian las observaciones sin procesar, también llamado grado de diferenciación.
* **q:** El tamaño de la ventana de media móvil, también llamado orden de media móvil.

Un modelo ARIMA es aquel en el que la serie temporal se diferencia al menos una vez para hacerla estacionaria y se combinan los términos AR y MA. Entonces la ecuación queda:



Para aplicar el modelo ARIMA, siga estos pasos:

| 1. **Preparación de datos:** obtenga los datos de series temporales en orden. 2. **Estacionalidad de los datos:** eliminar tendencias y estacionalidades. 3. **Identificar parámetros:** encontrar el orden correcto del modelo (p, d, q). 4. **Ajustar ARIMA:** aplique el modelo ARIMA a sus datos. 5. **Verifique los residuos:** asegúrese de que los residuos se parezcan al ruido blanco. 6. **Pronóstico:** utilice el modelo para predicciones futuras. 7. **Evaluar:** evaluar la precisión del pronóstico y ajustarlo si es necesario. 8. **Validar:** probar nuevos datos para confirmar la confiabilidad. 9. **Visualizar:** Comunique los resultados visualmente. 10. **Implementar:** ponga el modelo en acción si cumple con sus requisitos. |
| --- |

# **3. Procesamiento de datos**

## **3.1. Datos**

El conjunto de datos utilizado para este estudio abarca el balance de la matriz energética de Brasil desde 1970 hasta 2021, obtenido del sitio web de OLADE (<https://www.olade.org/>). Este conjunto completo de datos está estructurado en un archivo Excel con varias hojas. Importamos este archivo de Excel directamente desde un repositorio en [GitHub](https://github.com/NamNguyen2015/TFM/blob/main/datas/Option_B/Brazil_Energy%20balance%20matrix.xlsx) como un diccionario, y cada clave corresponde a un año específico dentro del período de 1970 a 2021, después de una limpieza de datos.

Cada año, los datos se organizan como un DataFrame, donde las columnas representan las fuentes de energía y las filas representan los sectores donde se distribuyen las energías.

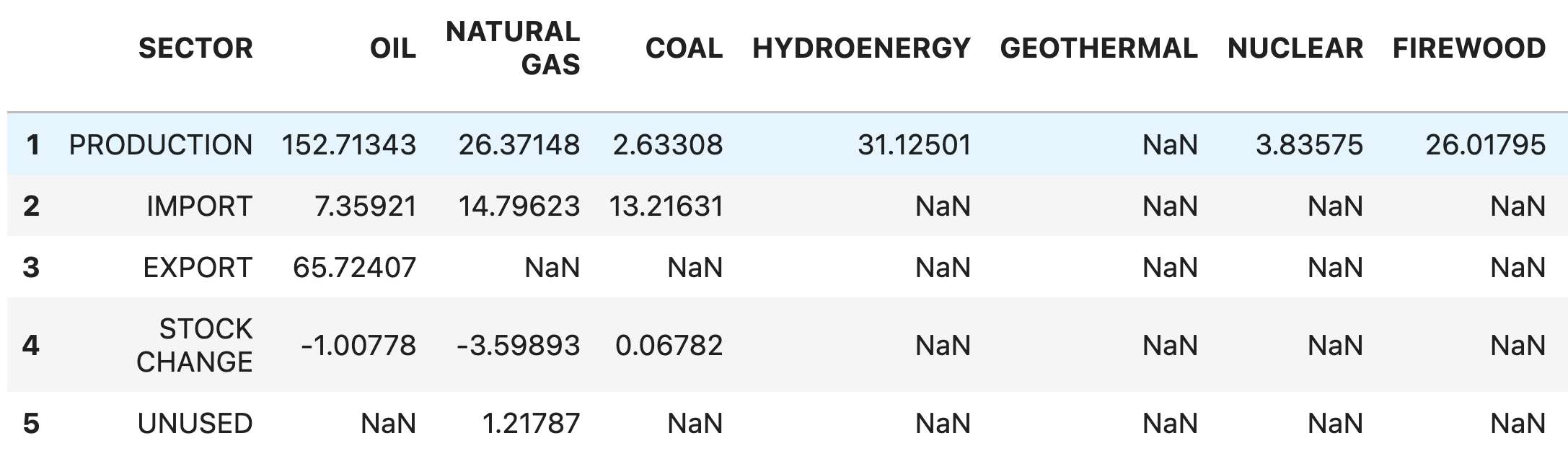


Gráfico 3. Primeras filas y columnas del DataFrame del año 2021

Todas las columnas del conjunto de datos son importantes para nuestro estudio debido a su función de proporcionar una visión integral del equilibrio de la matriz energética de Brasil de 1970 a 2021. Cada columna representa un aspecto específico de la producción y el consumo de energía y factores relacionados, incluidas diversas fuentes de energía y sectores.

Para conocer mejor la información que manejamos, vamos a explicar brevemente cada una de las variables que componen el set de datos:

| * **SECTOR:** Indica el sector económico al que pertenecen los datos, como Agricultura, Industria, Transporte, Residencial, Comercial, etc. Cada fila corresponde a un sector y un año específico. * **OIL:** La cantidad de energía consumida o producida a partir del petróleo en megatones equivalentes de petróleo (Mtoe). * **NATURAL GAS:** La cantidad de energía consumida o producida a partir del gas natural en megatones equivalentes de petróleo (Mtoe). * **COAL:** La cantidad de energía consumida o producida a partir del carbón en megatones equivalentes de petróleo (Mtoe). * **HYDROENERGY:** La cantidad de energía hidroeléctrica consumida o producida en megatones equivalentes de petróleo (Mtoe). * **GEOTHERMAL:** La cantidad de energía geotérmica consumida o producida en megatones equivalentes de petróleo (Mtoe). * **NUCLEAR:** La cantidad de energía nuclear consumida o producida en megatones equivalentes de petróleo (Mtoe). * **FIREWOOD:** La cantidad de energía consumida o producida a partir de leña en megatones equivalentes de petróleo (Mtoe). * **SUGARCANE AND PRODUCTS:** La cantidad de energía consumida o producida a partir de caña de azúcar y sus productos en megatones equivalentes de petróleo (Mtoe). * **OTHER PRIMARY:** La cantidad de energía consumida o producida a partir de otras fuentes primarias en megatones equivalentes de petróleo (Mtoe). * **TOTAL PRIMARIES:** La suma total de todas las fuentes primarias de energía consumidas en el sector en un año específico. * **ELECTRICITY:** La cantidad de energía consumida en forma de electricidad en megatones equivalentes de petróleo (Mtoe). * **LPG:** La cantidad de energía consumida en forma de gas licuado de petróleo en megatones equivalentes de petróleo (Mtoe). * **GASOLINE/ALCOHOL:** La cantidad de energía consumida en forma de gasolina y alcohol en megatones equivalentes de petróleo (Mtoe). * **KEROSENE/JET FUEL:** La cantidad de energía consumida en forma de queroseno y combustible para aviones en megatones equivalentes de petróleo (Mtoe). * **DIESEL OIL:** La cantidad de energía consumida en forma de aceite diesel en megatones equivalentes de petróleo (Mtoe). * **FUEL OIL:** La cantidad de energía consumida en forma de aceite combustible en megatones equivalentes de petróleo (Mtoe). * **COKE:** La cantidad de energía consumida en forma de coque en megatones equivalentes de petróleo (Mtoe). * **CHARCOAL:** La cantidad de energía consumida en forma de carbón vegetal en megatones equivalentes de petróleo (Mtoe). * **GASES:** La cantidad de energía consumida en forma de gases en megatones equivalentes de petróleo (Mtoe). * **OTHER SECONDARY:** La cantidad de energía consumida en otras formas secundarias en megatones equivalentes de petróleo (Mtoe). * **NON-ENERGY:** La cantidad de energía utilizada para fines no relacionados con la generación de energía en megatones equivalentes de petróleo (Mtoe). * **TOTAL SECUNDARIES:** La suma total de todas las fuentes secundarias de energía consumidas en el sector en un año específico. * **NON-ENERGY:** La cantidad de energía utilizada para fines no relacionados con la generación de energía en megatones equivalentes de petróleo (Mtoe). * **TOTAL:** La suma total de todas las formas de energía consumidas en el sector en un año específico. * **YEAR:** Indica el año al que corresponde la información de consumo de energía para ese sector en particular. |
| --- |

Si bien existen valores atípicos que pueden influir en los resultados del modelo, se han conservado para el análisis a fin de capturar posibles ideas valiosas o anomalías en los datos energéticos. Los valores máximo y mínimo generalmente se alinean con las tendencias centrales del conjunto de datos, lo que sugiere que están dentro de límites razonables. Comprender y considerar estos aspectos es crucial para un análisis de datos y un desarrollo de modelos sólidos.

La presencia de valores NaN (Not-a-Number) en la base de datos se debe principalmente a información faltante o no disponible. En este conjunto de datos, los valores de NaN son particularmente abundantes porque representan casos en los que los valores reales no están registrados o no están disponibles. Por lo tanto, estos valores faltantes se han representado como NaN para indicar la ausencia de datos.

## **3.2 Preprocesamiento de datos**

Realizamos diferentes acciones como:

* Redondear valores, cambiar el nombre de columnas, eliminar espacios y corregir nombres. Además, agregamos el año como una nueva columna.
* Convertimos unidades de medida de Ktoe (kilotones de equivalente de petróleo en sus siglas en inglés Kilo-Tonnes of Oil Equivalent) a Mtoe (megatones de equivalente de petróleo en sus siglas en inglés Mega-Tonnes of Oil Equivalent) en todas las columnas numéricas.
* Los valores numéricos se redondearon a dos decimales.
* Se eliminaron los encabezados de fila que contenían unidades de medida y las columnas "Sin nombre: 0" pasaron a llamarse "SECTOR" y "OTHER PRIMARY\_x000d\_" pasó a llamarse "OTHER PRIMARY".
* Se corrigió el nombre del sector 'COKE PLANTS AND BLAST FURNACES\_x000d\_' fue corregido a 'COKE PLANTS AND BLAST FURNACES.'.

Estas transformaciones garantizan que los datos estén en un formato coherente y listos para un análisis posterior.

* Los datos procesados ​​se almacenan en un nuevo diccionario llamado 'modified\_dict', que reemplaza al original.
* Finalmente, el diccionario limpio se guarda en un nuevo archivo Excel denominado “*Brazil\_Energy balance matrix\_cleaned.xlsx*” preservando las hojas originales para su posterior análisis y visualización.

Este proceso garantiza la calidad y coherencia de los datos antes de su utilización en análisis más profundos.

## **3.3 Visualización de los datos originales**

Este proyecto intenta responder a la pregunta de si es posible crear un modelo ARIMA para predecir el consumo energético futuro.

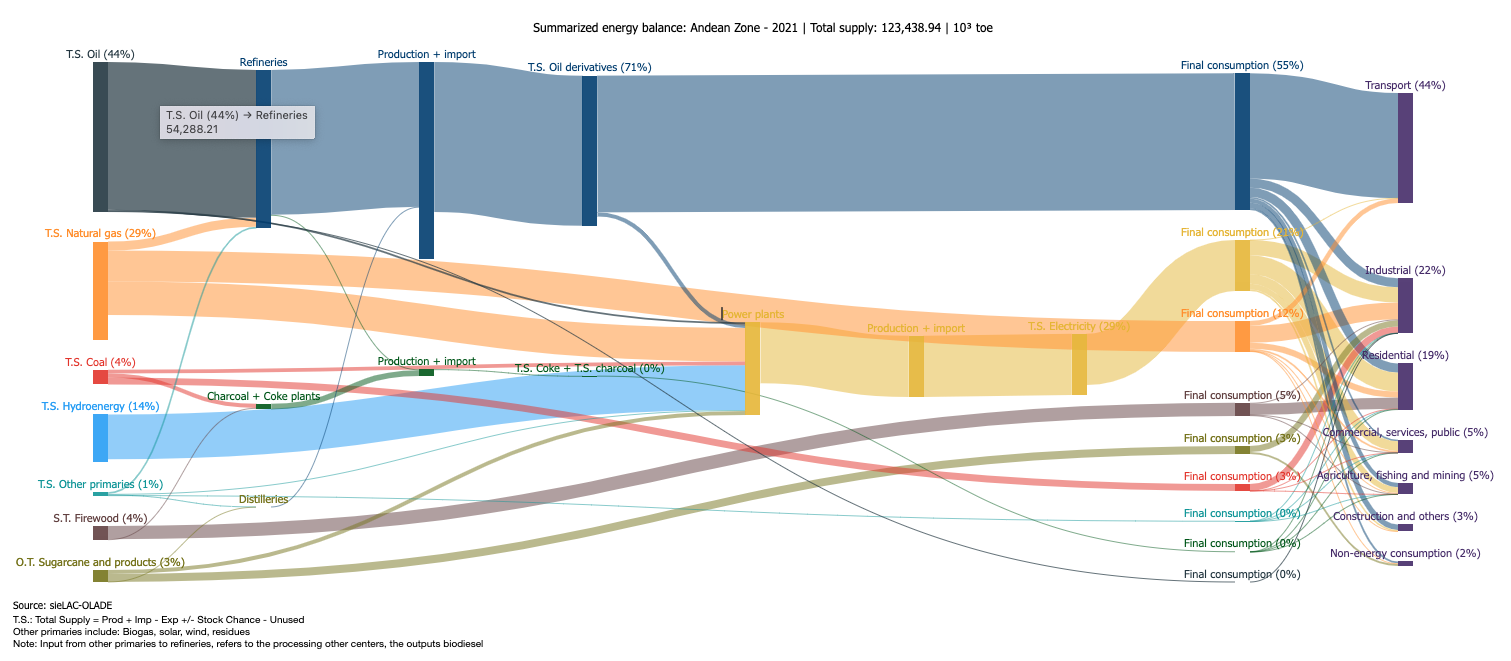


Gráfico 4. Diagrama de Sankey de la Matriz Energética de Brasil, 2021.

El suministro de energía depende de fuentes de energía tanto primarias como secundarias.

Gráfico

Descripción generada automáticamenteGráfico

Descripción generada automáticamente

Gráfico 5. Fuentes de Energía Primaria y Secundaria por año.

El consumo de energía final se distribuye entre varios sectores del país, incluidos el transporte, la industria, el residencial, los servicios públicos comerciales, la agricultura, la pesca y la minería, la construcción y el consumo no energético.

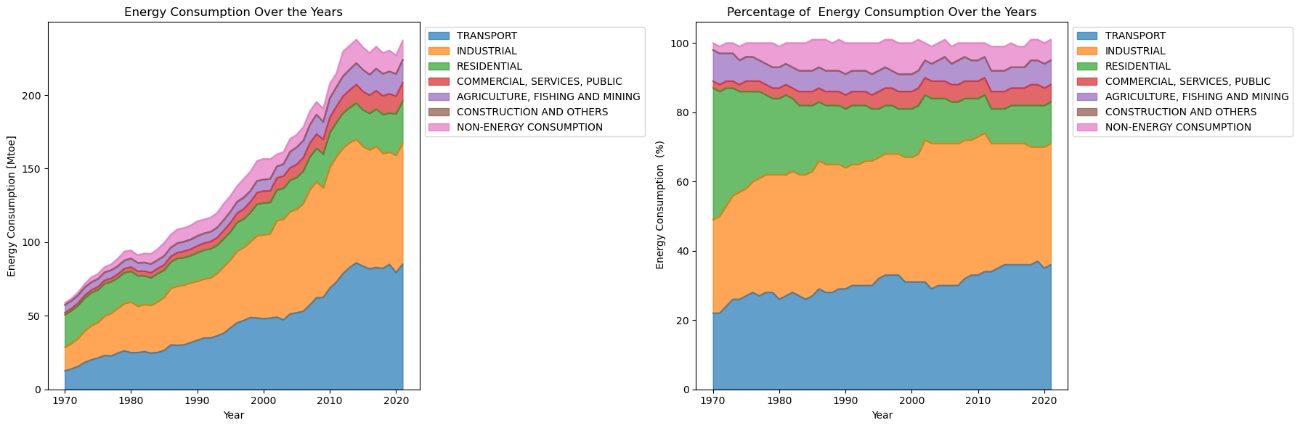


Gráfico 6. Tendencias del consumo de energía a lo largo del tiempo (Volumen - a, Porcentaje - b).

## **3.4 Exploración de datos**

La exploración de datos se considera un proceso importante para comprender los datos con métodos estadísticos y de visualización. Una de las ventajas de este proceso es identificar patrones y problemas en este conjunto de datos.

Tabla

Descripción generada automáticamente Tabla

Descripción generada automáticamente con confianza media

Gráfico 6. Datos de la primera y las últimas 5 filas.

Todas las columnas del conjunto de datos son importantes para nuestro estudio debido a su función de proporcionar una visión integral del equilibrio de la matriz energética de Brasil de 1970 a 2021. Cada columna representa un aspecto específico de la producción y el consumo de energía y factores relacionados, incluidas diversas fuentes de energía y sectores.

Si bien existen valores atípicos que pueden influir en los resultados del modelo, se han conservado para su análisis a fin de capturar posibles ideas valiosas o anomalías en los datos energéticos. Los valores máximo y mínimo generalmente se alinean con las tendencias centrales del conjunto de datos, lo que sugiere que están dentro de límites razonables. Comprender y considerar estos aspectos es crucial para un análisis de datos y un desarrollo de modelos sólidos.

La presencia de valores NaN (en sus siglas en inglés Not-a-Number) en la base de datos se debe principalmente a la información faltante o no disponible. En este conjunto de datos, los valores de NaN son particularmente abundantes porque representan casos en los que los valores reales son cero, pero esta información no se registró o no está disponible. Por lo tanto, estos valores faltantes se han representado como NaN para indicar la ausencia de datos.

# **4. Machine Learning**

## **4.1 Modelo ARIMA (Univariado)**

Aplicamos el modelo ARIMA a todos los sectores, pero primero debemos realizar algunas comprobaciones para seleccionar los mejores parámetros (p, d, q).

En el análisis de series de tiempo, si necesitamos escalar los datos (es decir, estandarizarlos o normalizarlos) depende de las características específicas de los datos y de las técnicas de modelado que planeamos utilizar. El escalado puede ser necesario o no.

En esta sección, nos centraremos en modelar el problema univariado.

## **4.4.1 ARIMA para todos los sectores**

* **Paso 1:** Verifique manualmente para elegir d:

En este paso, primero verificamos la estacionariedad de una serie temporal realizando una prueba estadística, como la prueba Dickey-Fuller aumentada (ADF en sus siglas en inglés Augmented Dickey-Fuller), y examinando el valor p asociado con la prueba. Si el valor p es menor que 0,05, podemos considerar la serie como estacionaria; de lo contrario, hacemos que la serie sea estacionaria aplicando diferenciación. El orden mínimo de diferenciación es el valor de d.

Además, trazamos la función de autocorrelación (ACF en sus siglas en inglés Auto-Correlation Function) y la función de autocorrelación parcial (PACF en sus siglas en inglés Partial Auto-Correlation Function) en cada orden de diferenciación. Este paso nos permite seleccionar los valores apropiados para 'p' y 'q' según las características de los gráficos ACF y PACF.

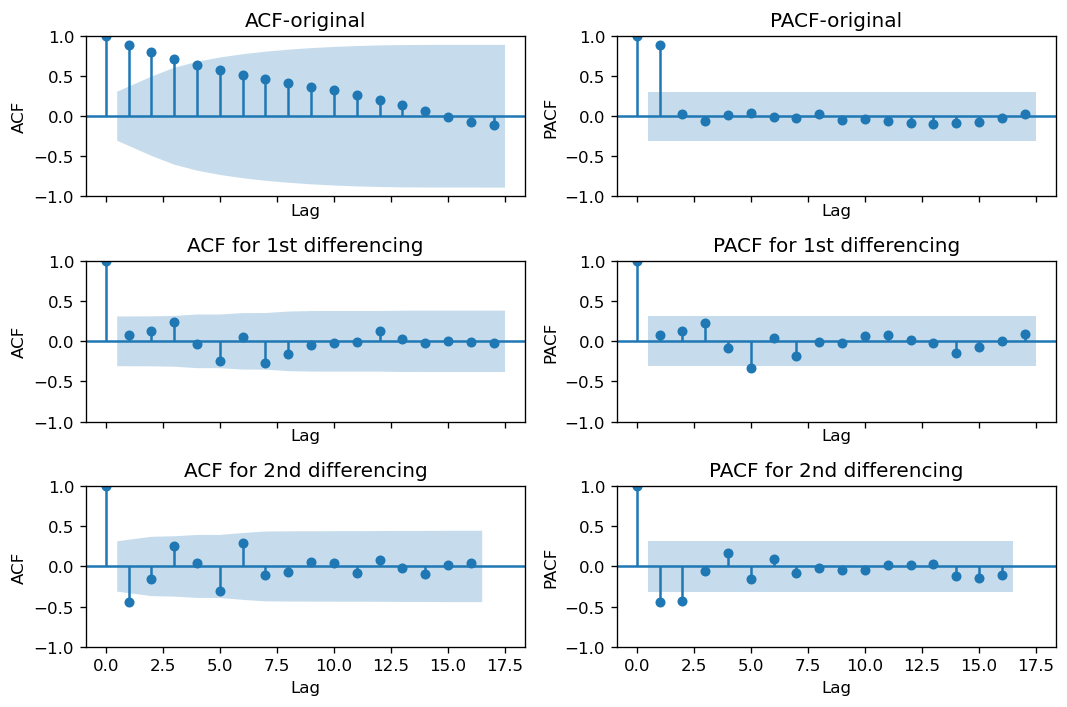


Gráfico 7. ACF y PACF

Para todos los sectores, la segunda diferenciación (d = 2) hace que la serie temporal sea estacionaria, lo que indica que se aplicará una diferenciación de segundo orden en el modelado ARIMA de cada sector para lograr la estacionariedad requerida. *(CONSULTAR ANEXOS).*

* **Paso 2**: Encuentra el mejor (p,q) manualmente:

Este paso es opcional, ya que “p” y “q” también se pueden seleccionar encontrando los valores mínimos de AIC (Criterio de información de Akaike en sus siglas en inglés Akaike Information Criterion) y BIC (Criterio de información bayesiano en sus siglas en inglés Bayesian Information Criterion) si “d” ya está determinado.

Los valores óptimos de “p” y “q” varían entre sectores, y la elección de los criterios (AIC o BIC) puede influir en la selección. Estas variaciones indican la complejidad de modelar el consumo de energía en diferentes sectores, y la selección del mejor modelo ARIMA puede depender del contexto específico y los objetivos del análisis.

## **4.4.2 Entrenamiento, validación y predicción del modelo**

En nuestro proceso de modelado, asignamos una parte importante del conjunto de datos para fines de *training* o entrenamiento y *test* o validación. Específicamente, utilizamos el 90% de los datos para *training*-entrenamiento y el 10% para *test* - validación. Esta división nos permitió entrenar eficazmente nuestros modelos ARIMA y evaluar su rendimiento.

Para la fase de validación, utilizamos el conjunto de datos de validación para evaluar la precisión de las predicciones de nuestros modelos. Este proceso implicó comparar los valores predichos generados por nuestros modelos ARIMA con los datos reales en el conjunto de validación. Luego, calculamos el Error Porcentual Absoluto Medio (MAPE) para cuantificar el nivel de precisión alcanzado. El MAPE proporcionó información valiosa sobre el rendimiento de los modelos, ayudándonos a comprender cuán bien pueden generalizar hacia datos no vistos. Si el valor del MAPE para cada sector es menor al 10%, podemos considerar que el modelo se ajusta bien. Sin embargo, en algunos sectores (como construcción y consumo no energético), carecemos de datos e información suficientes para emitir un juicio concluyente sobre el ajuste del modelo.

* TRANSPORT: 9.37%
* INDUSTRIAL: 9.76%
* RESIDENTIAL: 14.27%
* COMMERCIAL, SERVICES, PUBLIC: 9.69%
* AGRICULTURE, FISHING AND MINING: 9.03%
* CONSTRUCTION AND OTHERS: 9165359077.49% (an outlier)
* NON-ENERGY CONSUMPTION: 22.68%
* FINAL CONSUMPTION: 8.57%

Estos valores MAPE proporcionaron información valiosa sobre el rendimiento de los modelos, ayudando a comprender qué tan bien podrían generalizarse a datos invisibles. Si bien la mayoría de los sectores alcanzaron valores MAPE razonables, es esencial abordar el valor atípico en el sector “CONSTRUCTION AND OTHERS”, donde MAPE parece excepcionalmente alto, lo que indica problemas potenciales con el desempeño del modelo para este sector.

En la fase de predicción, proyectamos el consumo futuro de energía para cada sector.

Específicamente, pronosticamos las tendencias del consumo de energía para los próximos 20 años, lo que nos permite hacer predicciones informadas sobre el consumo de energía en los años venideros.

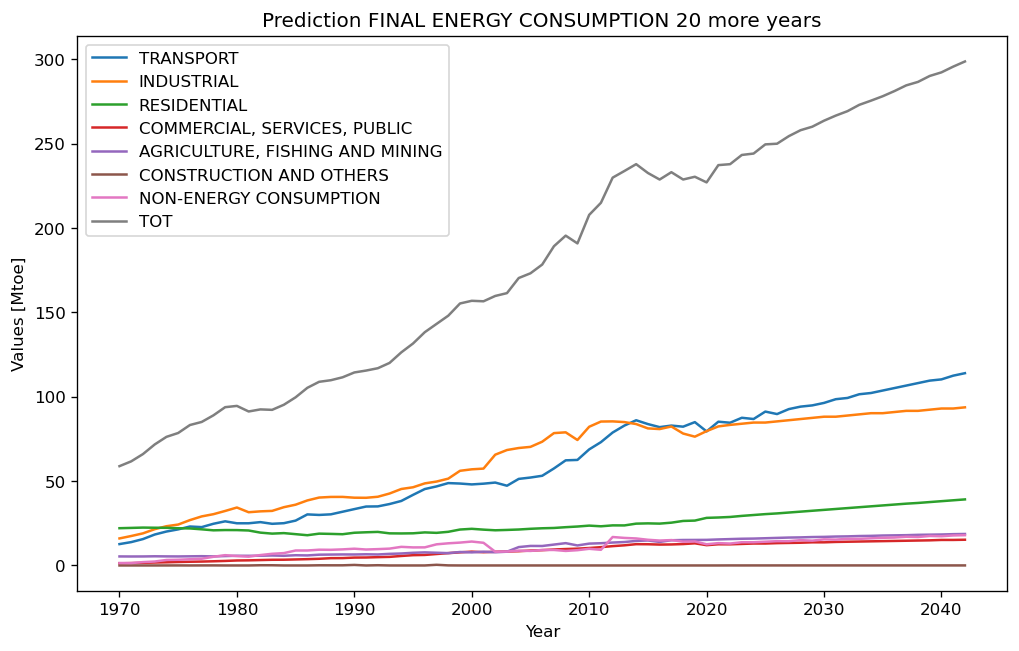


Gráfico 8. Predicción de “FINAL ENERGY CONSUMPTION” +20 años.

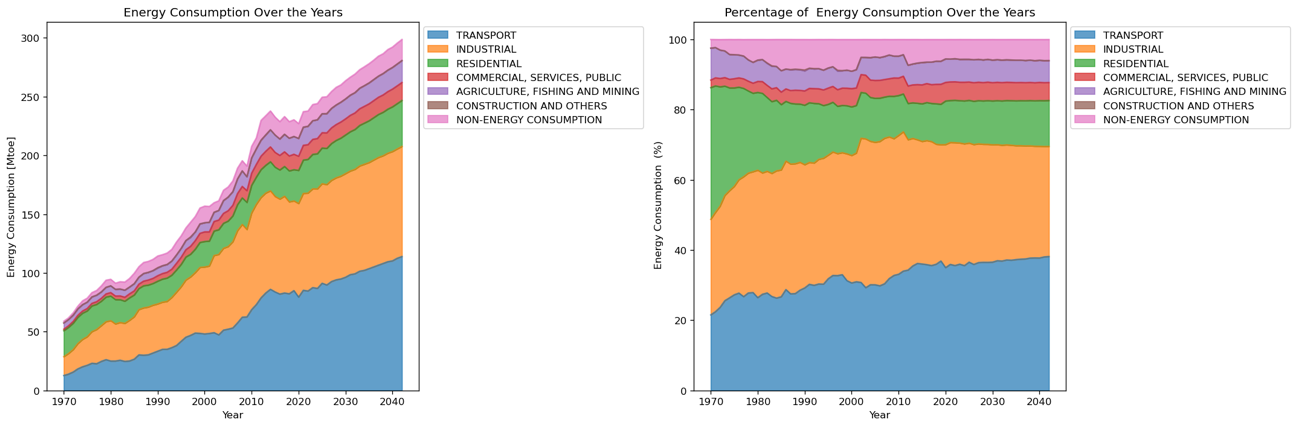


Gráfico 9. Predicción “ENERGY CONSUMPTION BY SECTOR” +20 años.

Pronosticamos las tendencias del consumo de energía para los próximos 20 años. Para hacer esto, utilizamos los modelos ARIMA entrenados y especificamos la cantidad de períodos que queríamos pronosticar, lo que nos permite hacer predicciones informadas sobre el consumo de energía en los años venideros.

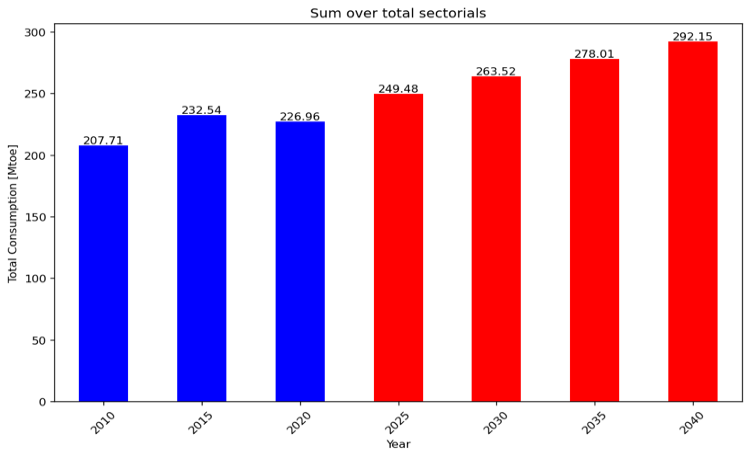


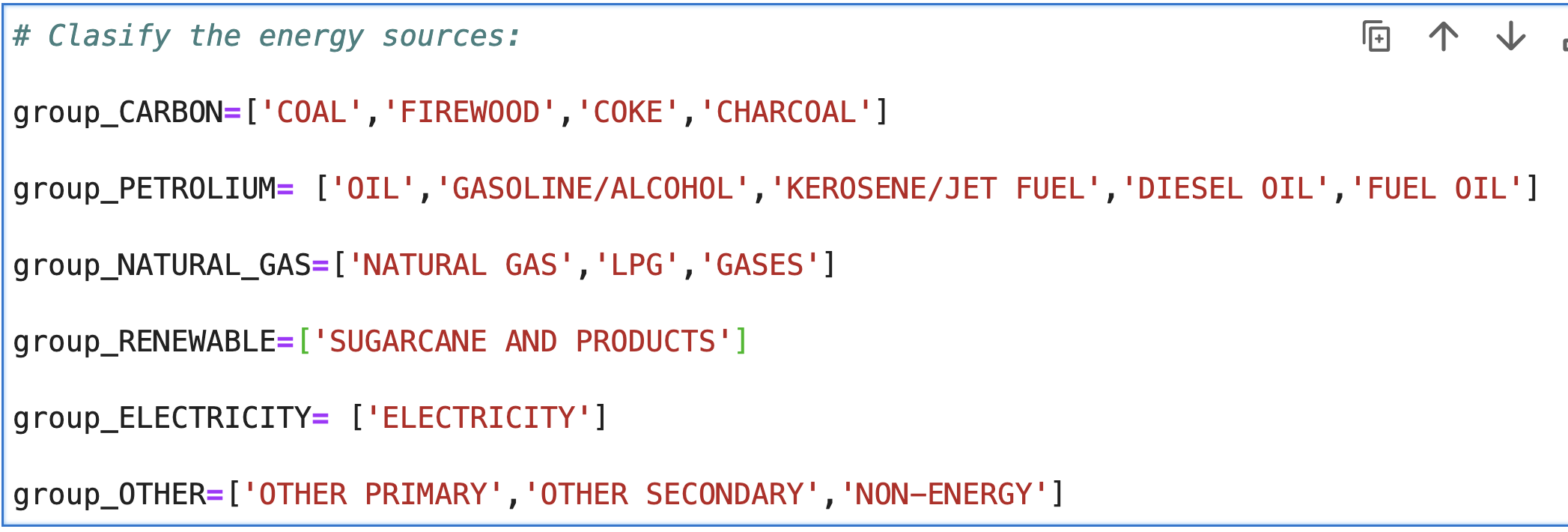
Gráfico 10. Predicción de “SUM OVER TOTAL SECTORIALS” +20 años

Siguiendo este enfoque de *training*-entrenamiento, *test*-validación y predicción, nos aseguramos de que nuestros modelos fueran sólidos y capaces de realizar pronósticos confiables para diversos sectores de consumo de energía. Estos pasos fueron esenciales en la preparación de nuestros datos para análisis posteriores y procesos de toma de decisiones.

## **4.4.3 Escenarios**

En esta sección, hemos realizado una serie de pasos de preparación y análisis de datos para crear escenarios para el consumo futuro de energía en diferentes grupos de fuentes de energía. Los principales objetivos de este análisis fueron comprender los patrones históricos de consumo de energía, clasificar las fuentes de energía en grupos significativos y estimar cómo estos grupos podrían contribuir al consumo de energía futuro.

Comenzamos seleccionando el sector “FINAL CONSUMPTION” de nuestro conjunto de datos y limpiando los datos para eliminar columnas innecesarias. A continuación, categorizamos las fuentes de energía en distintos grupos, como CARBON, PETROLEUM, NATURAL GAS, RENEWABLE, ELECTRICITY y OTHER. Esta agrupación nos permitió agregar el consumo de energía dentro de cada categoría.



Tengamos en cuenta que desde 1990, Brasil ha estado desarrollando su programa nuclear como una fuente de energía relativamente joven. Ha realizado una contribución significativa a la generación de electricidad en el país.

Para comprender la distribución histórica de las fuentes de energía, calculamos la distribución porcentual de cada grupo en relación con el consumo total de energía para cada año. Estos porcentajes se redondearon a números enteros por simplicidad.

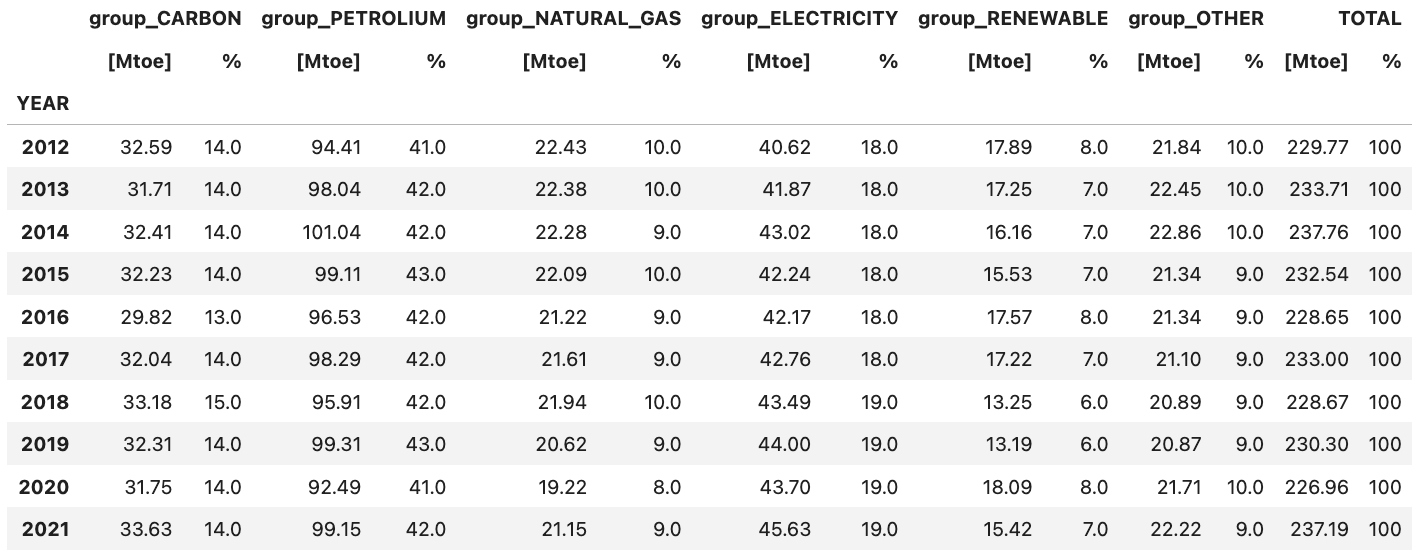


Tabla 1. Fuentes de energía en el consumo final durante los últimos 10 años.

Luego, los datos resultantes se combinaron en un DataFrame completo, *final\_df*, que incluye valores históricos de consumo de energía (en Mtoe) y la distribución porcentual de las fuentes de energía para cada año. Este DataFrame proporciona una descripción clara de las tendencias del consumo de energía y la distribución de fuentes hasta el año actual.

Además, creamos escenarios para el consumo de energía futuro estimando cómo los grupos de fuentes de energía podrían contribuir al consumo total de energía en los próximos años. Estos escenarios se basan en el supuesto de que la distribución de las fuentes de energía seguirá siendo coherente con las tendencias históricas recientes.

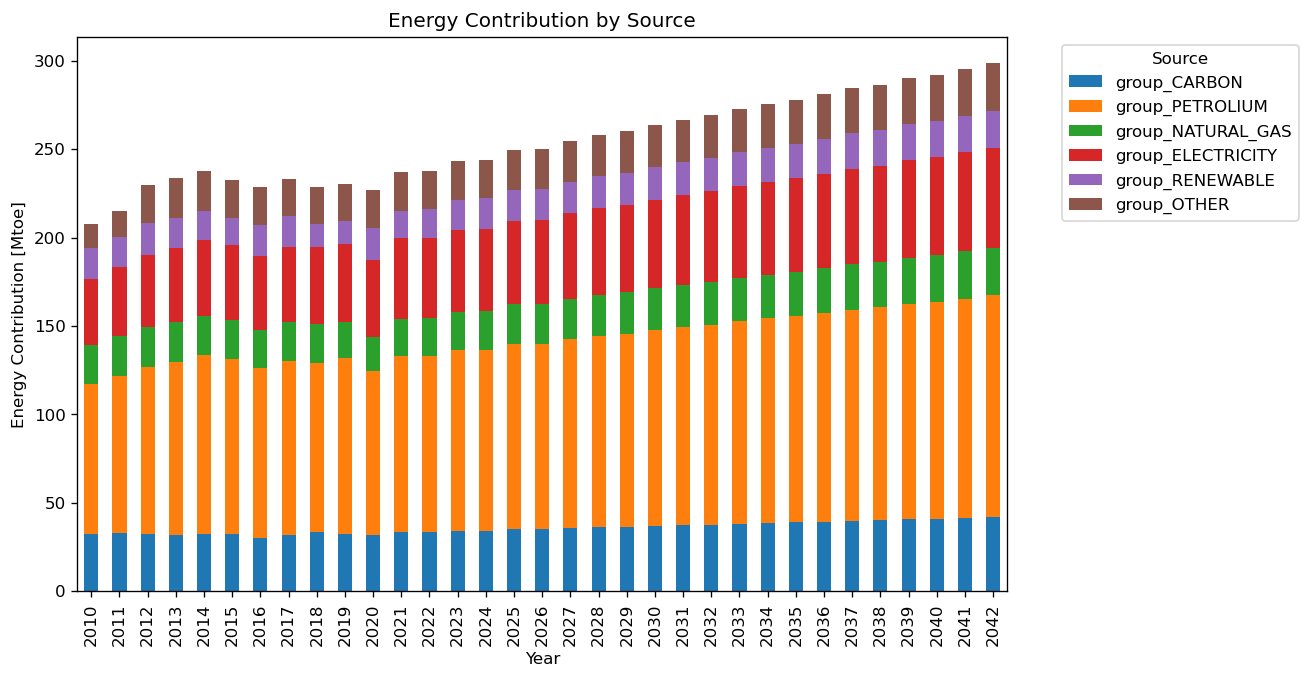


Gráfico 11. Predicción del consumo de energía por sector durante los próximos 20 años

Para escenarios futuros de consumo de energía (de 2022 a 2042), observamos varias tendencias clave:

* Fuentes basadas en carbono: El consumo de fuentes de energía basadas en carbono (CARBON) se mantiene relativamente estable a lo largo de los años, con una ligera tendencia al alza.
* Fuentes basadas en petróleo: El consumo de energía procedente de fuentes basadas en petróleo (PETROLEUM) también muestra un aumento gradual, lo que refleja una demanda continua.
* Gas Natural: El consumo de gas natural (NATURAL GAS) mantiene un patrón de crecimiento consistente, contribuyendo significativamente al consumo total de energía.
* Electricidad: El consumo de electricidad (ELECTRICITY) aumenta constantemente durante el período de pronóstico, lo que indica su importancia en varios sectores.
* Energías renovables: El consumo de fuentes de energía renovables (RENEWABLE) experimenta un aumento continuo, aunque moderado, lo que pone de relieve el creciente énfasis en la energía sostenible.
* Otras fuentes: La energía de otras fuentes (OTHER) muestra una trayectoria ascendente constante, probablemente influenciada por diversas alternativas energéticas.

Estos escenarios pueden servir como herramientas valiosas para las partes interesadas involucradas en la planificación, formulación de políticas y toma de decisiones energéticas.

Proporcionan información sobre posibles cambios en los patrones de consumo de energía, lo que permite estrategias y acciones informadas para abordar los desafíos energéticos futuros de manera efectiva.

# 4. Análisis y visualización de los resultados

## Estimaciones y predicciones de la demanda de energía sectorial

## Visualización y aplicación interactiva

# 5. Conclusiones

## Conclusiones obtenidas

## Limitaciones y futuras líneas investigaciones

# 6. Recomendaciones .

# Referencias bibliográficas/ Anexos