

MEMORIA

EDA **E**MISIONES **D**E **C**O2

INDICE

- 1. INTRODUCCIÓN**
 - 1.1. OBJETIVOS**
 - 1.2. SOBRE EL DATASET**
- 2. ENTENDIMIENTO DE LOS DATOS**
 - 2.1. REVISIÓN DE VARIABLES**
 - 2.1.1. HEAD, TAIL, SAMPLE**
 - 2.1.2. INFO**
 - 2.1.3. DESCRIBE**
- 3. LIMPIEZA DE LOS DATOS**
 - 3.1. MISSINGS**
 - 3.2. ELIMINACIÓN DE DATOS**
 - 3.3. OUTLIERS**
 - 3.4. TRANSFORMACIONES (nuevas variables)**
- 4. ANÁLISIS DE LOS DATOS**
 - 4.1. ANÁLISIS UNIVARIANTE**
 - 4.1.1. ESTADÍSTICOS DE CENTRALIDAD: MEDIA, MEDIANA, MODA, CUARTILES, PERCENTILES**
 - 4.1.2. ESTADÍSTICOS DE DISPERSIÓN: VARIANZA, DESVIACIÓN ESTÁNDAR, RANGO, MÍNIMO Y MÁXIMO**
 - 4.1.3. HISTOGRAMAS(PAIRPLOT), BOXPLOT**
 - 4.2. ANÁLISIS BIVARIANTE**
 - 4.2.1. SCATTERPLOTS (PAIRPLOT)**
 - 4.3. ANÁLISIS MULTIVARIANTE**
 - 4.3.1. HEATMAP**
- 5. RESULTADOS:**
 - 5.1. CONTRASTE DE HIPÓTESIS**
- 6. ANEXOS**
 - ANEXO 1, PRODUCCIÓN DE ENERGÍA VS PIB DE ESTADOS UNIDOS**
 - ANEXO 2, PRODUCCIÓN DE ENERGÍA VS PIB CHINA**
 - ANEXO 3, HEATMAP DE USA**
 - ANEXO 4, HEATMAP DE CHINA**
 - ANEXO 5, HEATMAP GENERAL**
 - ANEXO 6, MAYORES PRODUCTORES DE ENERGÍA DEL MUNDO**
 - ANEXO 7, EMISIONES DE CO2 TOTALES**
 - ANEXO 8, CRECIMIENTO DEL PIB POR PAÍSES (PPP)**
 - ANEXO 9, BOXPLOT DE USA**
 - ANEXO 10, BOXPLOT DE CHINA**
 - ANEXO 11, PROPORCIÓN DE ENERGÍA USADA POR PAÍSES**
 - ANEXO 12, PAIRPLOT DE CHINA**
 - ANEXO 13, PAIRPLOT DE USA**
 - ANEXO 14, EMISIONES DE CO2 POR CONTINENTES**
 - ANEXO 15, PRODUCCIÓN VS CONSUMO DE ENERGÍA POR CONTINENTES**
 - ANEXO 16, PIB POR CONTINENTES**
 - ANEXO 17, POBLACIÓN 40 MAYORES PAÍSES PRODUCTORES ENERGÍA**

ANEXO 18, BOXPLOT GENERAL

1. INTRODUCCIÓN

1.1. OBJETIVOS

Vamos a realizar un EDA sobre un data set de emisiones de co2 con el objetivo de poder corroborar o desmentir las siguientes hipótesis de partida:

- a. China crece tanto y tan rápido porque contamina más
- b. China contamina más porque usa energía que no respeta el medio ambiente
- c. La economía de China está basada en el carbón y la nuclear
- d. EE. UU. y Europa se basan en la misma producción de energía que China, pero están reconvirtiendo el sector

Y en general:

- a. Extraer información útil de las variables, desechando las que no aporten nada
- b. Identificar outliers, missing values o errores
- c. Entender la relación entre las variables
- d. Responder a las cuestiones anteriores

1.2. SOBRE EL DATASET

Los datos provienen originalmente del ministerio de energía de Estados Unidos, y alojados en kaggle, de donde lo hemos obtenido. En ellos encontramos una colección de grandes factores influyentes en las emisiones de co2 por cada uno de los países del mundo, desde el año 1980 hasta el año 2020.

Descripción de las variables:

- Country: el país en cuestión
- Energy_type: el tipo de fuente de energía usado
- Year: El año en el que el dato fue capturado
- Energy_consumption: Cantidad de consumo específico de una fuente de energía, medido en quad/btu (medida anglosajona de la energía).
- Energy_production: Cantidad de producción específica de una fuente de energía, medida en quad/btu.
- GDP: El PIB de cada país medido en paridades de poder de compra (ppp) en billones del año 2015 en la divisa dólar (\$).
- Population: población de cada país, medida en millones de personas
- Energyintensityper_capita: medida de la ineficiencia energética de una economía. Es calculado como unidades de energía per cápita (cápita = una persona individual) en millones de btu/persona.
- Energyintensityby_GDP: medida de la ineficiencia energética de una economía calculado en unidades de energía por unidad monetaria de PIB (1000 btu/2015\$ PIB ppp).
- CO2_emission: cantidad de co2 emitido, medido en millones de toneladas de co2

2. Entendimiento de los datos

2.1. Revisión de variables

2.1.1. Head, Tail, Sample:

Cargamos nuestros datos y realizamos un primer vistazo de ellos mediante las funciones de pandas head, tail y sample con el objetivo de tener una percepción general de lo que contiene nuestro data set. En este sentido, mirando el head, vemos una columna llamada "Unnamed: 0" que debemos eliminar, y que dentro de la variable "Country" tenemos como valor el dato World, que es la suma de todos los datos de cada país y que, además, en la columna "Energy_type", se incluye el tipo all_energy_types que es

la suma de todos los tipos de energía. Estas filas las vamos a eliminar del dataframe pues no nos aportan nada que no podamos obtener realizando la suma de los datos y nos van a molestar a la hora de graficar datos.

Mirando el tail, vemos que tenemos al país Zimbabwe en el que detectamos que tiene un nan en la columna "Energy_consumption", por lo que confirmamos que el data set tienen missing values y veremos más tarde cuántos y donde.

Por último, miramos un sample aleatorio de 10 filas y vemos más missing values por lo que pasamos a usar la función info de pandas con el objetivo de comprobar estos nan y el tipo de variables que tenemos.

2.1.2. INFO:

Comprobamos que la columna Year debemos convertirla en una variable de tipo datetime y que el resto tienen un tipo correcto, lo realizamos.

Por otra parte, vemos que tenemos missing values en las variables Energy_consumption, Energy_production, GDP, Population, Energy_intensity_per_capita, Energy_intensity_by_GDP y CO2_emisión.

2.1.3. DESCRIBE

Con el describe podemos ver algunos puntos que necesitamos estudiar, en este sentido tendremos que revisar lo siguiente:

1. vemos que el mínimo de "Energy_consumption" es negativo, ver si tiene sentido, y que el 25% de los datos se acumulan hasta el 0 ¿tiene sentido esto también?
2. En "energy production" pasa lo mismo que en "Energy_consumption", sin embargo, para los percentiles 50 y 75 puede tener sentido que el país tenga una producción negativa si estos no producen la suficiente energía y tienen que importarla del exterior y que su saldo sea negativo. En ese caso, la variable sería un saldo neto, pero no es el caso. En este sentido, consultado un informe sobre el sistema eléctrico español de 2020 elaborado por REE (Red Eléctrica Española) comprobamos que un dato negativo de producción a nivel nacional para un país significa que la energía consumida para la producción excede a la energía generada, por lo que los mantenemos.
3. El máximo de GDP es muy elevado, corroborar que no se trata de un outlier, hacer lo mismo para el mínimo. Comprobación de outliers.
4. Posibles outliers también en intensity per cápita y por GDP, comprobar los outliers de manera general.
5. Las emisiones de co2 vemos que la mínima es negativa, ¿tiene esto sentido? y el 50% de los valores se acumulan hasta 0 también, tiene sentido? Sin embargo, el máximo es muy elevado lo que nos da una desviación típica muy elevada también, ¿tiene sentido?

En las nuevas variables creadas, tengo países que, aunque su producción no es 0 es tan pequeña que se puede considerar cero y que me están entorpeciendo el análisis, puesto que en la ratio de eficiencia se me ponen en los puestos primeros, aunque su emisión de c02 sea tan pequeña, dado que su producción es igualmente pequeña, pero en una proporción muchísimo mayor a la emisión de co2 y me los indica como muy contaminantes, cuando en proporción al mundo no lo son. En este sentido esos números

tan cercanos a 0 los vamos a redondear a 3 decimales y se nos van a quedar en 0. Al tener un número de emisiones de CO₂ dividido por 0 nos quedará un número infinito, que más tarde vamos a sustituir por 0. Una vez hecho esto, vemos que seguimos obteniendo los mismos resultados porque, aunque la producción sea cercana a cero, en realidad hay producción y emisión y al ser una ratio, son realmente poco eficientes, aunque su aportación contaminante de manera relativa al total mundial sea inapreciable, entonces vamos a marcar un umbral de producción para estudiar realmente si merece la pena mantener todos los países, dados los problemas que nos generan. Esto último lo veremos en la parte de eliminación de datos

6. LIMPIEZA DE LOS DATOS

6.1. MISSINGS

Miramos que porcentaje de missing values tenemos en cada una de estas variables y las revisamos:

- Energy_consumption: Tenemos un 20,11% de missing values.
- Energy_production: Tenemos un 20,11% de missing values.
- GDP: Tenemos un 27,80% de missing values.
- Population: Tenemos un 17% de missing values.
- Energy_intensity_per_capita: Tenemos un 9,17% de missing values.
- Energy_intensity_by_GDP: Tenemos un 9,17% de missing values.
- CO₂_emission: Tenemos un 6,9% de missing values.

En los datos vemos que la mayoría vienen de países muy pequeños los cuales es posible que en esa época no proporcionaran datos, por diferentes motivos, no estar adheridos a algún sistema internacional de aportación de datos, ser países en desarrollo que no los recogen o proporcionan, estar en guerra, entre otros. También hay países desarrollados en la lista, pero en mucha menor medida, cuya falta de datos puede ser debida a algún tipo de error en la recolección. Para estar más seguros, vemos en qué años se han producido los missing values. En este sentido, los missing se reparten de una manera bastante homogénea a través de los diferentes años. Vemos, además que algunos países ya no existen como tal y acumulan grandes cantidades de missing values como pueda ser, Former Yugoslavia, Germany East, Germany West, Former Serbia and Montenegro, Former U.S.S.R., Former Czechoslovakia. Además, se confirma que muchos de los países son en vías de desarrollo al obtener los 20 países que más datos nulos tienen y, dada la cantidad elevada de datos nulos, pero también la cantidad elevada de datos que disponemos, la eliminación homogénea por años y el poco aporte al análisis que nos harían, decidimos eliminar los nulos de las columnas Energy_production y Energy_consumption.

Para los nulos de las columnas GDP, Population, Energy_intensity_per_capita, Energy_intensity_by_GDP y CO₂_emission podemos rellenarlos con algún método que decidamos, ya que son variables en las cuales pensamos que el rellenado es mejor que la eliminación, y no representan una proporción grande de datos sobre el conjunto. En este sentido, vamos a considerar que en la columna GDP, en los años en que nos falte el dato, el PIB se mantuvo constante he igual al año anterior; con la población vamos a seguir el mismo criterio. De este modo, para estas dos variables vamos a usar el método de relleno backfill.

Los índices de intensidad per cápita tienen aún menos nulos en relación con el total de datos y puesto que son un cociente que incluye a las anteriores variables vamos a mantener el mismo método para mantener la homogeneidad, por tanto, para:

- Energy_intensity_per_capita: usaremos backfill
- Energy_intensity_by_GDP: usaremos backfill

Por último, para las emisiones de CO2, dadas las decisiones tomadas con anterioridad, y por los mismos motivos, vamos a aplicar el mismo método, pues al asumir que se mantuvieron constantes las anteriores, también lo harán las emisiones de CO2.

Los datos que nos faltan de longitud y latitud de los 40 países que nos hemos quedado los buscamos y los rellenamos, así como los continentes de los países que nos faltan.

6.2. ELIMINACIÓN DE DATOS

Comprobamos la columna de Energy_consumption y vemos que todos los consumos negativos de Energía se deben al tipo de generación renovable, y tiene sentido, puesto que al ser renovables no consumen energía para la producción de esta, sino que la aportan al sistema y de ahí que los datos sean negativos. En este sentido, no eliminamos los datos negativos en el dataframe.

Por otra parte, revisamos los datos de Energy_production y vemos que los valores negativos corresponden todos a Qatar y para energías renovables, comprendidos en el periodo de 1984 a 1995. Podemos apreciar que no tenía consumo de energías renovables y que la producción se puede considerar también cero. Es decir, Qatar en estos periodos no produce ni consume energía renovable por lo que vamos a eliminarlos. Esto es debido, según hemos podido encontrar, a que Qatar es el mayor exportador de gas del mundo y nunca le ha hecho falta preocuparse por las renovables, así ha empezado a preocuparse por esto en torno al año 2019.

En relación con si nos debíamos quedar con todos los países del dataframe o solo con los más relevantes a nivel mundial para la producción de energía y emisión de co2 y en base a los problemas que estábamos teniendo en el análisis de estos pequeños países y el beneficio que nos podrían aportar, hemos obtenido los siguientes datos:

1. la producción mundial media es de 141.42 millones de btu
2. la producción media de los 40 países con más producción es 141.37 millones de btu
3. El porcentaje de producción media mundial de estos 40 países es 99.97%
4. El porcentaje de producción media mundial del resto de países es de 0.03%

Por tanto, decidimos que solo vamos a estudiar los 40 países con mayor producción y el resto los eliminamos.

Por otra parte, eliminamos todas las filas correspondientes con el país "World" ya que si nos hiciera falta solo tendríamos que sumar las variables de todos los países y obtendríamos el 99,97% del valor mundial al quedarnos solo con esos 40 países.

6.3. OUTLIERS

Dada la distribución de los datos vista antes en las cuales la mayoría de la producción y consumo de energía, así como las emisiones de co2 y la población se concentran en unos

pocos países, consideramos normales los boxplot que vemos y no podemos considerar los datos outliers (ver anexo 23). En Energy_intensity_by_GDP sí que podríamos ver que tenemos dos outliers pero que no merece la pena quitarlos. Aun así, con la representación de la población mundial podemos ver que esta está concentrada en China, Indonesia, India y Estados Unidos (ver anexo 17) y de ahí que no los pongamos como outliers porque al igual que la población la producción y el consumo de energía también tiende hacia China, Estados Unidos y Rusia.

6.4. TRANSFORMACIONES

Creemos nuevas variables a partir de los datos para nuestro estudio:

Balance = producción energética - consumo energético: nos dirá si la economía es productora o consumidora neta de energía y por tanto si exporta o importa la misma.

Eficiencia = emisiones de CO₂ / producción energética = nos dirá cuanto co₂ emite por tonelada de producción energética. Por tanto, nos dirá quién es más responsable con el medio ambiente o que economía es más limpia.

Dependencia energética = consumo energético / PIB, nos dirá cuanta energía consume la economía por cada euro de PIB generado, y por tanto cuanto depende esta de la energía.

Intensidad de uso = consumo energético / población, nos dirá la energía que consume cada persona en cada país y por tanto que tanto dependen las personas de esa economía de la energía, lo que redundará también en unas mayores emisiones de co₂.

co₂_pc = emisiones de co₂ / población, nos dirá cuanto co₂ emite cada una de las personas de la economía

latitude: añadimos la latitud de los países para representarlos en un gráfico

longitude: añadimos la longitud de los países para representarlos en un gráfico

geometry: añadimos la geometría de los países

CODE: añadimos el código país para representarlos en un gráfico

4 ANÁLISIS DE LOS DATOS

4.1. ANÁLISIS UNIVARIANTE

4.1.1. ESTADÍSTICOS DE CENTRALIDAD: MEDIA, MEDIANA, MODA

A través de las medias obtenidas con el describe de nuestras variables podemos observar que la media de producción de energía está un poco por encima que el consumo de esta, lo que nos indica que en media a nivel mundial existe exceso de oferta, sin embargo, está bastante ajustado, siendo sus desviaciones típicas muy parecidas. De ahí también, que la media del balance mundial nos arroje un valor positivo de 0.2222.

Por otra parte, vemos que en la moda de nuestras variables el valor más frecuente en energía producida y consumida es cero, lo que nos indica que la misma está muy concentrada en ciertos países, siendo la energía más usada el carbón a nivel general. Del país que tenemos más datos es de Alemania, y del continente que hay más datos es de Asia.

Que la media de la energía producida y consumida sea de 2,41 y 2,19 y el 50 por ciento de los valores se acumulen hasta unos valores de producción y consumo de 0,52 y 0,55 nos da una idea de la concentración que existe en unos pocos países de estas variables y reafirma el hecho de que nos centremos solo en unos pocos y pongamos el foco en China y Estados Unidos, siendo Rusia importante también, pero sobre la cual no vamos a realizar el estudio, pues no forma parte de las hipótesis iniciales.

4.1.2. ESTADÍSTICOS DE DISPERSIÓN: VARIANZA, DESVIACIÓN ESTÁNDAR, RANGO, MÍNIMO Y MÁXIMO

Vemos que el rango es bastante amplio en las variables GDP, Population y efficiency indicándonos que existirá una gran dispersión entre sus datos, y es debido a las grandes diferencias que existen entre países, puesto que solo unos pocos concentran la mayoría de la producción y emisiones de CO₂ de energía. En este sentido, por ejemplo, el valor mínimo y máximo para ellas es de 0,43 billones de dólares para el país con menor PIB y de 32289 billones de dólares para el de mayor pib; en población, el país que tiene una menor población tiene 50 millones de personas y el país con mayor población 1,4 miles de millones de personas(ver gráfico en anexo 17); y por último en eficiencia, es decir en la emisión de co₂ por tonelada de producción de co₂, también hay grandes diferencias, teniendo países con la menor eficiencia en 247029 toneladas de co₂ emitidas por btu de energía producida y los de mayor eficiencia en 0. Los países a los que corresponden estos datos, en estas variables son:

GDP: el mínimo corresponde a Arabia Saudí en el 82 y el máximo a Algeria en el 85

Population: el mínimo de población corresponde a Yugoslavia en 1980 y el máximo a China en el 2019

Efficiency: el mínimo corresponde a Venezuela, Iran, Russia, Former Czechoslovakia, Former Yugoslavia, Norway, Oman, Netherlands, Nigeria, Qatar, Poland, Malaysia, Libya, México, United States, United Kingdom, United Arab Emirates, Saudi Arabia, South Africa, Canada, Colombia, China, Argentina, Australia, Algeria, Angola, Brazil, Indonesia, India, Kuwait, Iraq, Japan, Egypt, Germany, France, Turkmenistan, Ukraine, Uzbekistan y Kazakhstan a lo largo de todos los años estudiados y el máximo corresponde a Francia en el 2018(un país que se basa en la nuclear en el 75,5%)(ver gráfico en anexo 11).

4.1.3. HISTOGRAMAS, BOXPLOT

Los histogramas y boxplot los hemos estudiado solo para china y estados unidos que son los países en los que nos vamos a centrar. Así para China, según los histogramas del pairplot (Véase anexo 12) vemos que ninguna de nuestras variables sigue una distribución normal y que por lo general los datos están bastante concentrados en unos pocos valores. Por otra parte, en los boxplot (véase gráfico anexo 10) vemos que en general no tenemos outliers más allá de uno en energy intensity y de otros dos en efficiency por arriba y por abajo que no merece la pena tratar.

En relación con Estados Unidos (véase gráfico en anexo 13), vemos con los histogramas que en general los datos se distribuyen bastante homogéneamente durante los diferentes años, siendo un poco la excepción las variables c02 per cápita y energy dependency, sin embargo, tampoco podemos afirmar que sigan una distribución normal.

Mirando los boxplot de Estados Unidos (véase gráfico 9) apreciamos que únicamente en Energy_production tenemos outliers correspondientes a la producción de energía a partir del año 2013 debido a que a partir de ese año se aumenta bastante la producción de energía procedente del petróleo y del gas natural. Buscamos información para saber el porqué de este aumento, solo podemos constatarlo a través de las noticias dadas en los periódicos, pero no se expone una causa explícita para este aumento en la producción energética tan fuerte.

4.2. ANÁLISIS BIVARIANTE

4.2.1. SCATTERPLOTS

Según vemos en los scatterplots de China (véase anexo 12) GDP tiene una relación negativa con la dependencia energética, la eficiencia, el balance y la intensidad de energía por GDP y una relación positiva con la producción per cápita, las emisiones de Co2, el consumo de energía y la producción de energía. Vemos que la dependencia energética y la intensidad de energía por GDP, nos representan lo mismo y no nos habíamos dado cuenta por lo que la eliminamos. Por otra parte, la producción de energía con la eficiencia está relacionado positivamente al igual que con las emisiones de co2, pero es porque la eficiencia es la ratio del co2 entre la producción.

En el caso de Estados Unidos (véase gráfico 13), vemos al igual que en China, pero de manera más marcada que en Estados Unidos las gráficas entre las variables se parecen mucho y las que tienen una mayor relación son: la intensidad de energía por GDP con el GDP y con la población; el consumo y la producción con el GDP y con la población; y por último la población con el GDP y la eficiencia con el balance. Todo esto lo comprobaremos con los heatmap.

4.3. ANÁLISIS MULTIVARIANTE

4.3.1. HEATMAP

Según los Heatmap para el caso de China (véase gráfico anexo 4) vemos que existe una correlación importante entre las variables:

GDP y efficiency (0.9), Energy_intensity_by_GDP (-0.8), population (0.8), per_capita_production (1), Co2_emission (1), Energy_consumption (1), Energy_production (1). En este sentido con las cuatro últimas tengo que revisar porque me preocupa que sea algo tan perfecto como 1, indicando una relación matemática entre las variables.

Population y Energy_intensity_by_GDP (-0.9), per capita production (0.8), co2_emission (0.8), energy_consumption (0.8), Energy_production (0.8).

Co2_emission y energy_production (1) y Energy_consumption con co2_emission (1)

Igualmente, los unos me resultan extraños y tendría que tratar de investigar más a fondo el porqué.

En el caso de los Estados Unidos las relaciones más importantes entre variables las vemos entre las siguientes (véase gráfico anexo 3):

- Producción y GDP (0.8)
- Consumo y GDP (0.9)
- Energy_dependency y GDP (0.9)
- Producción y población (0.8)
- Consumo y población (0.9)
- Energy_dependency y población (-0,9)
- Emisiones de co2 y consumo de energía (0,8)
- Eficiencia y emisiones de co2 (0,9)

5. RESULTADOS:

CONTRASTE DE HIPÓTESIS

Empezando por una visión general de los datos y por tanto dividida entre continentes podemos ver que Asia, Norte América y Europa tienen los mayores pib mundiales y que por detrás se encuentran América del Sur, África y Oceanía (véase gráfico anexo 16). En esta misma división, podemos ver que los mayores productores siguen siendo los tres anteriores, pero entre estos tres los más destacados con diferencia en pib y producción energética son en primer lugar Asia y en segundo lugar América del Norte. La diferencia fundamental aquí es que mientras Asia en su conjunto tiene superávit de energía, siendo una exportadora neta, América del Norte y Europa son deficitarias y necesitan comprar energía al exterior (véase gráfico anexo 15). Dados estos datos, es normal que Asia y norte América sean los continentes que más co2 emiten a la atmósfera, siendo el continente asiático con gran diferencia el mayor contaminante mundial con China a la cabeza y Estados Unidos en la parte norteamericana. Esto es así, porque el continente americano lo hemos subdividido en sus partes y por tanto Estados Unidos queda como representante de Norteamérica junto a Canadá y en Asia se concentra una mayor parte de productores mundiales como China, India y Rusia. En este último punto, clasificar a Rusia en Asia puede ser debatible dado que la mayor parte de su territorio, el 77% se encuentra en el continente asiático, pero la mayor parte de la población rusa vive en la parte que se encuentra en el continente europeo por lo que se podría considerar que los rusos son europeos no así la mayor parte del territorio y por tanto el lugar donde se encuentra la energía y decidimos que vamos a dejar la clasificación tal como está.

Hemos visto que China es el segundo mayor productor de energía y que el 75 por ciento de ella proviene del carbón, el 11,8% del petróleo y otros líquidos y el 8,1% de renovables (véase gráfico anexo 11). Además, vemos que su PIB ha crecido de manera muy importante durante el periodo (véase gráfico anexo 2), llegando a superar a los Estados Unidos que ha tenido un crecimiento lineal (véase gráfico anexo 1). En este sentido el pib de China en 1980 era de 684,62 millones de dólares de paridad de poder de compra y en 2019 llegó a ser de 23.128,34 dólares.

Estados Unidos tiene la producción de energía más diversificada que China, así produce el 28,4% de la energía con gas natural, el 28,1% con carbón, el 24,1% con petróleo y otros líquidos y solo el 9,7% con renovables y el 9,4% con nuclear (véase anexo gráfico 15).

En términos de pib, Estados Unidos en 1980 estaba por delante de China con una cifra de 7080,75 millones de dólares ppp, pero en 2019 las cosas cambiaron y terminó con un pib de 19.925,43 euros. Mirando las gráficas de evolución del pib y de la producción energética por tipos de fuente hemos podido ver que hay dos periodos diferenciados en ambas economías (véanse los anexos 1 y 2 de China y Usa), en las cuales Estados Unidos

realiza un cambio en el mix productivo, potenciando las renovables y el gas natural y reduciendo notablemente la producción con carbón, esto se produce a partir del año 2007. En China también hay un cambio, pero es la potenciación, con la producción mucho más fuerte, de la energía procedente del carbón a partir del año 1999.

Mirando la diferencia en las tasas de crecimiento entre estos dos periodos diferenciados en ambas economías y comparándolos entre ellos podemos apreciar que Estados Unidos venía creciendo a tasas anuales medias del 3,15% en el periodo inicial de 1980 a 2007, y China lo venía haciendo a tasas del 9,931% en el periodo de 1980 a 1993. Cuando se produce el cambio productivo de energía en ambas economías, estas tasas bajan en ambos países, pero en mayor proporción en los Estados Unidos que venía ya con cifras muy inferiores y manteniendo China cifras muy altas de crecimiento, lo que les ha permitido adelantar a Estados Unidos. En este sentido, después del cambio productivo China crecía al 8,986% de tasa media anual y Estados Unidos lo hacía en el 1,658%. Según estos datos podemos decir que, aunque el tipo de producción energética está relacionada con el Pib, no va a explicar todo el crecimiento de este y que a Estados Unidos le ha afectado el cambio de producción en la obtención de menores tasas de crecimiento, posiblemente por ser energías menos eficientes en el sentido del coste de producción, que el carbón, que es la energía que fue predominante en Estados Unidos hasta el 2007 y que China potenció en gran medida a partir de 1999.

Por otra parte, también hemos visto que China no es el país que emite más CO₂ por tonelada de BTU de energía producida en media (véase anexo 7) y esto no nos cuadra mucho con que la producción en su mayoría venga del carbón, estando Estados Unidos por delante de China, aunque muy parejos, teniendo este último un mix de generación más repartido entre fuentes, pero es lo que nos dicen los datos.

Vamos a ver a través de un contrastaste de hipótesis t-test pareado si efectivamente existe algo interesante que estudiar, o, dicho de otra manera, si las medias del periodo de 1980-1999, en el caso de China, y de 1980-2007, en el de Estados Unidos, son diferentes a sus respectivas medias en el periodo de 2000-2019 y 2008-2019 respectivamente:

Como el p-value ($pvalue=5.66926347973168e-12$) es menor que $\alpha=0.05$ podemos decir que sí ha ocurrido algo que merezca la pena estudiar en el pib de Estados Unidos a partir del año 2007. Vamos a realizar la misma prueba para la producción de energía. Como nuestro pvalue es igual a 0.0008798452929901086, podemos decir lo mismo que en la producción de energía para ambos periodos y por tanto es muy probable que el cambio en el mix productivo de la energía de Estados Unidos haya tenido algún efecto en el crecimiento del pib. Lo realizamos también para las emisiones de co₂, obteniendo un pvalue de 0.03361231406200629, e igualmente hay algo relevante entre los periodos que hace que las medias sean diferentes y por tanto confirma que el cambio en el mix ha tenido un efecto en las emisiones de co₂.

En el caso de China los resultados que hemos obtenido es que, en el pib sí que hay algo relevante que ha ocurrido entre los periodos de 1980-1999 y 2000-2019, al igual que con Estados Unidos. Por otra parte, con las otras dos variables, emisiones de co₂ y producción de energía, también aceptamos la hipótesis alternativa de que hay algo relevante a estudiar en ellas.

En definitiva, las respuestas a nuestras hipótesis iniciales serían:

a. China crece tanto y tan rápido porque contamina más

Como hemos visto a lo largo del estudio esta hipótesis no se confirma puesto que el crecimiento ha venido determinado por otros factores. Esto lo podemos corroborar en la comparación del crecimiento del pib de Estados Unidos y China en los periodos de 1980 a 1993 donde China producía con carbón, pero unos niveles muy inferiores a después de 1993 y aun así las emisiones de co2 entre China y Estados Unidos eran muy parejas siendo las de Estados Unidos superiores.

b. China contamina más porque usa energía no respeta el medio ambiente

En este punto debemos decir que China emite un co2 similar al de Estados Unidos que tiene un mix de producción energética mucho más diversificado, siendo el de Estados Unidos algo superior. Además, si vemos la ratio de eficiencia (co2/producción), ni China ni Estados Unidos son los más contaminantes, estando Estados Unidos por delante de China.

c. La economía de China está basada en el carbón y la nuclear

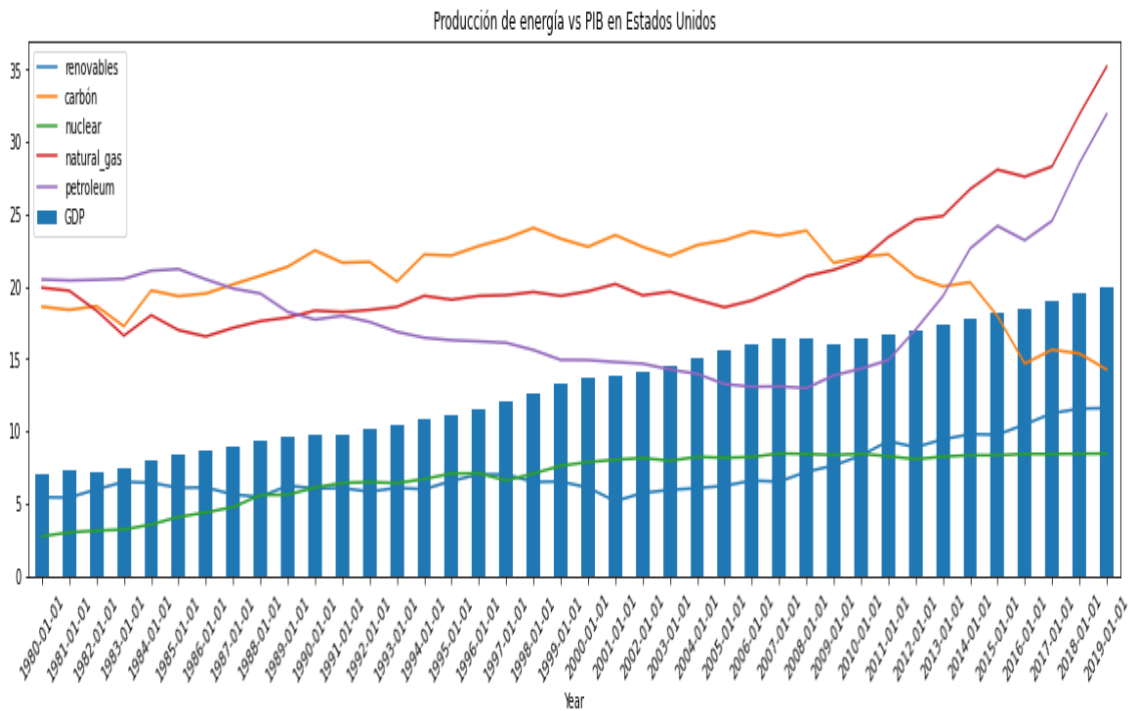
Hemos comprobado que prácticamente en su totalidad, un 75,4%, está basado en el carbón y sólo un 1,3 por ciento en la nuclear.

d. EE. UU. y Europa se basan en la misma producción de energía que China, pero están reconvirtiendo el sector

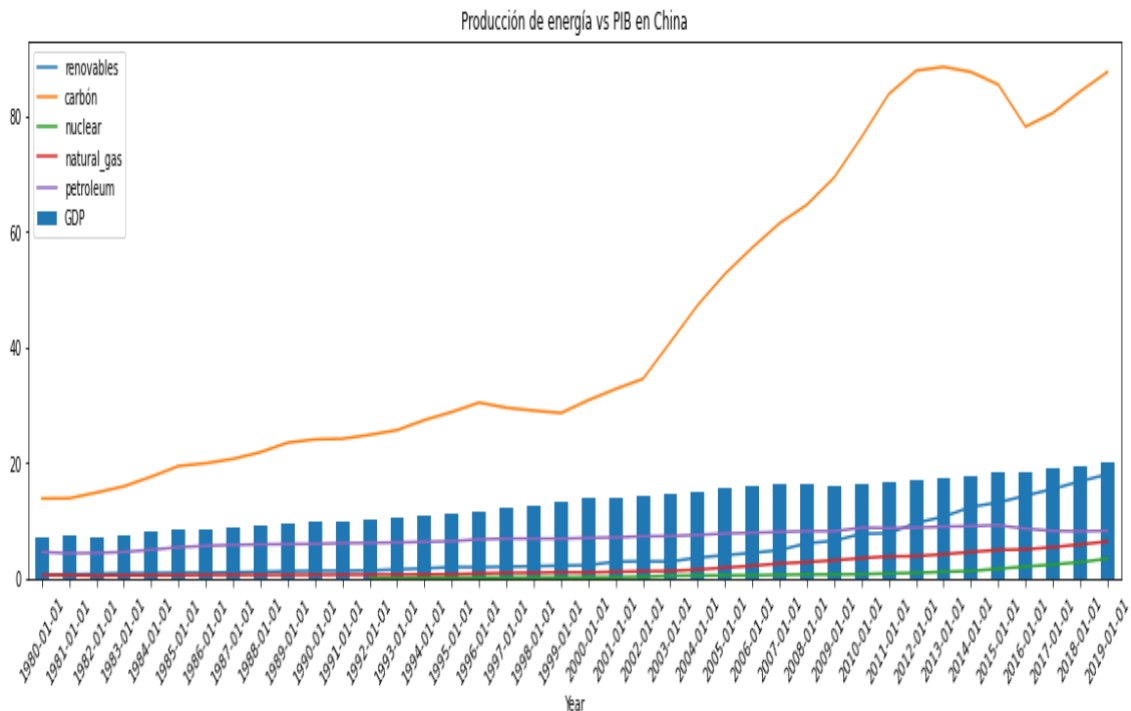
Vemos en el anexo 25 que en un inicio para Europa así era, pero desde 1987 aproximadamente, empezaron a eliminar drásticamente la producción de carbón para potenciar solo las renovables y reducir el resto, aunque a diferentes velocidades. Vemos que la nuclear y el gas se ha ido reduciendo en menor medida que el petróleo que también lo hizo fuertemente. Por tanto, podemos decir que en unos inicios todos se basaban en mayor o menor proporción en lo mismo, pero Europa y Estados Unidos han ido reconvirtiendo el sector para potenciar las renovables y China ha seguido un camino totalmente opuesto potenciando el carbón, en vez de reducirlo. Hay que decir en este sentido, que Europa es la que más ha apostado por las renovables ya que Estados Unidos, aunque ha reducido el carbón, ha potenciado más el gas natural y el petróleo.

ANEXOS

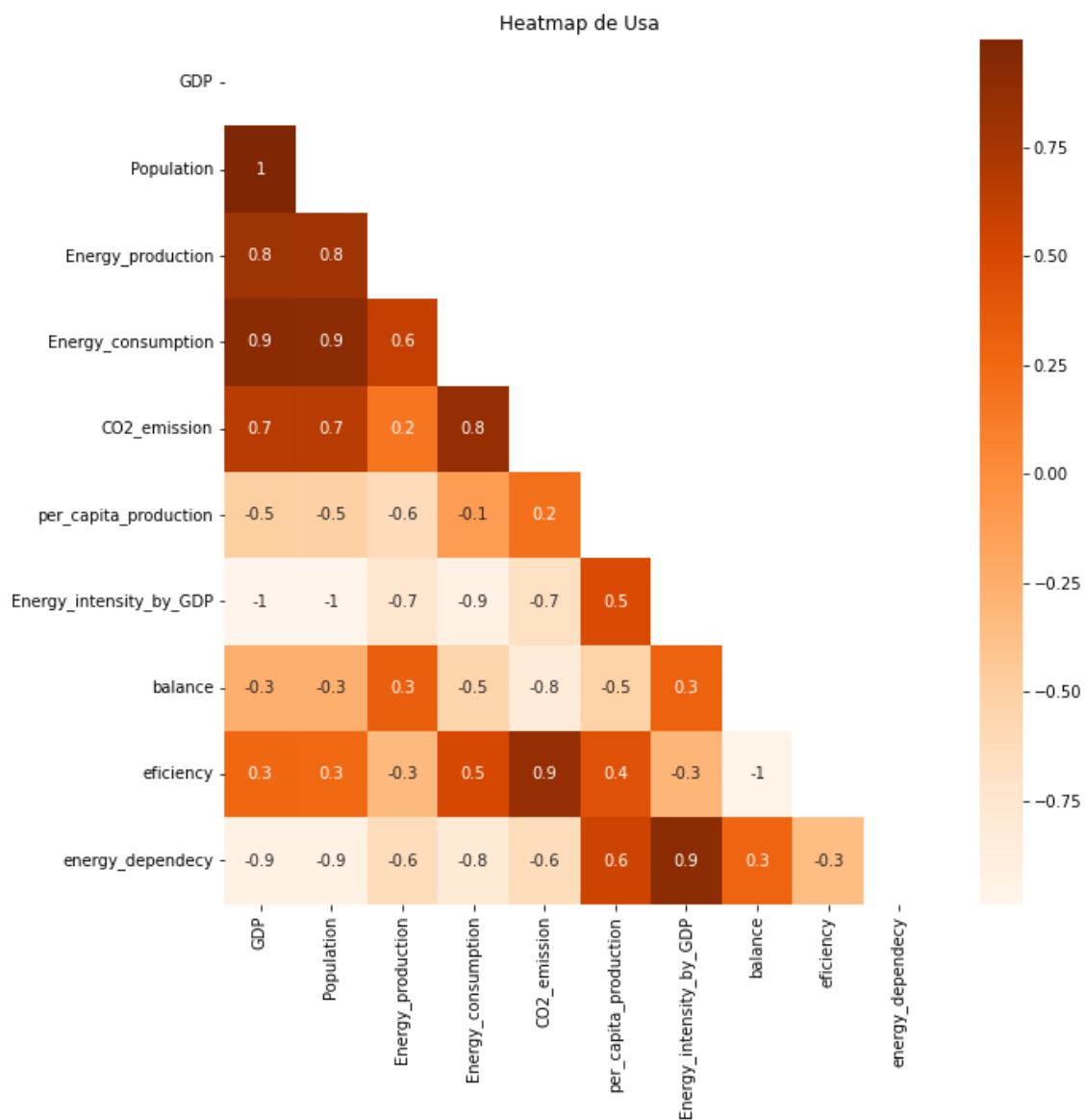
Anexo1, PRODUCCIÓN DE ENERGÍA VS PIB DE ESTADOS UNIDOS:



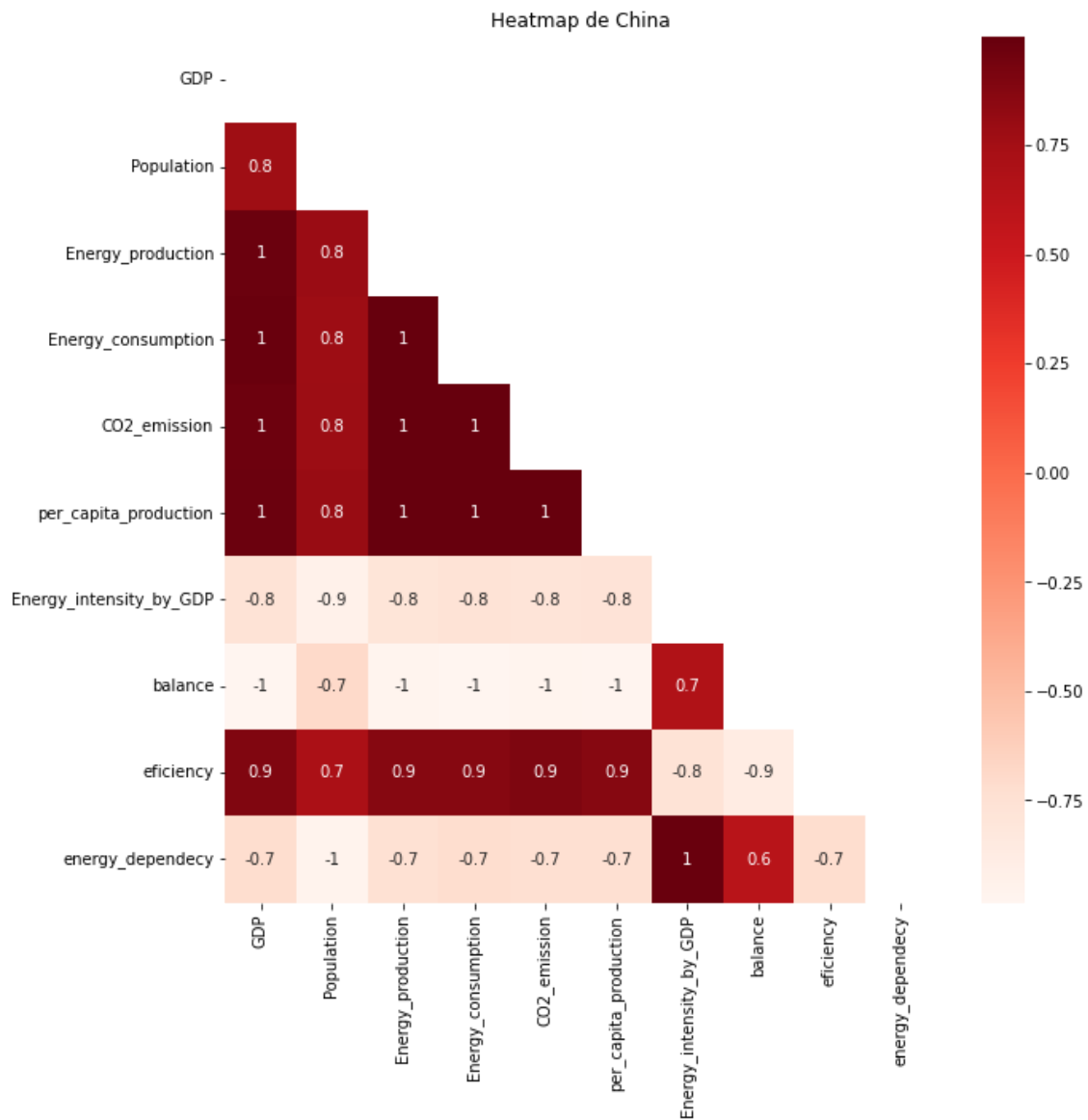
Anexo 2, PRODUCCIÓN DE ENERGÍA VS PIB CHINA:



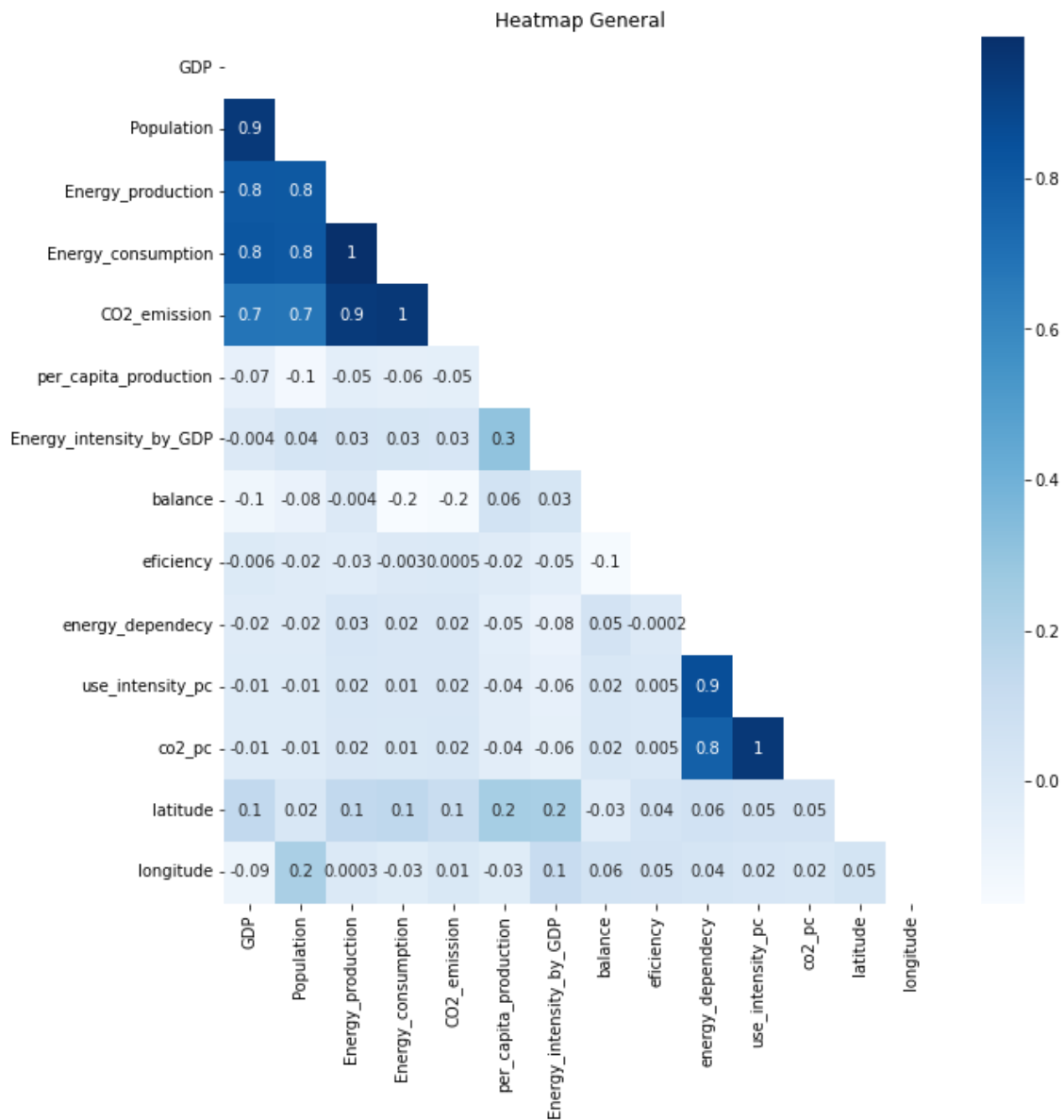
Anexo 3, HEATMAP DE USA:



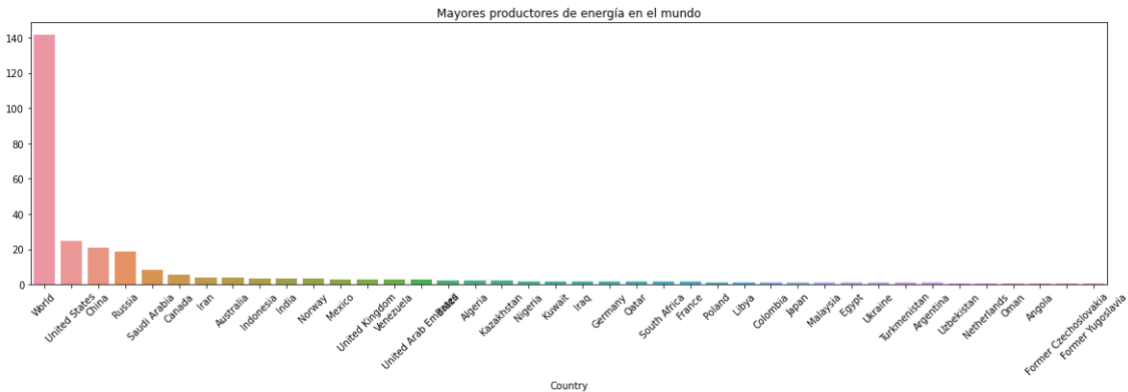
Anexo 4, HEATMAP DE CHINA:



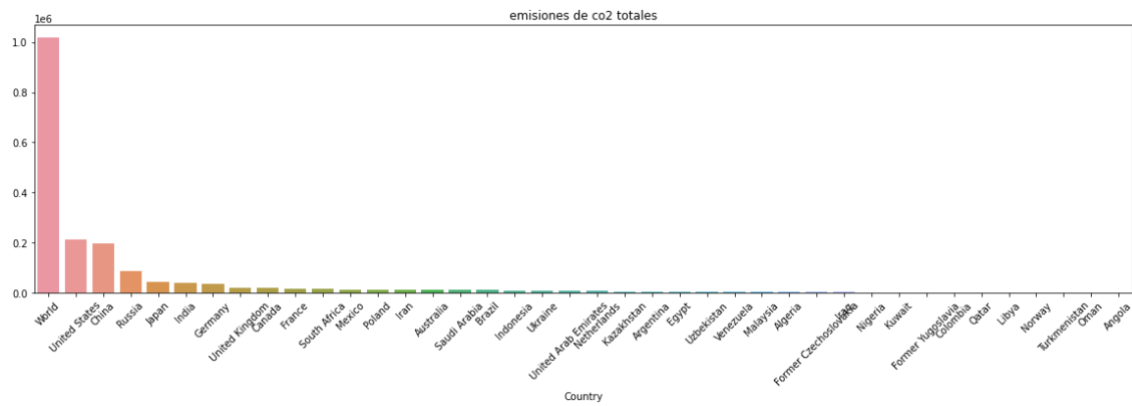
ANEXO 5, HEATMAP GENERAL:



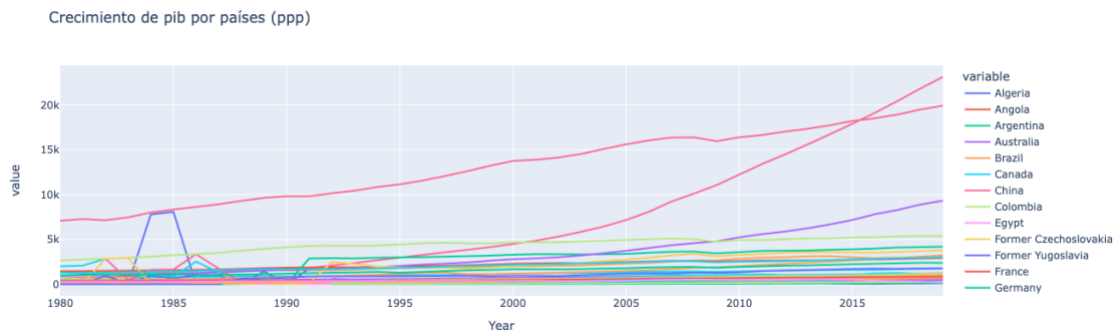
ANEXO 6, MAYORES PRODUCTORES DE ENERGÍA DEL MUNDO:



ANEXO 7, EMISIONES DE CO2 TOTALES:

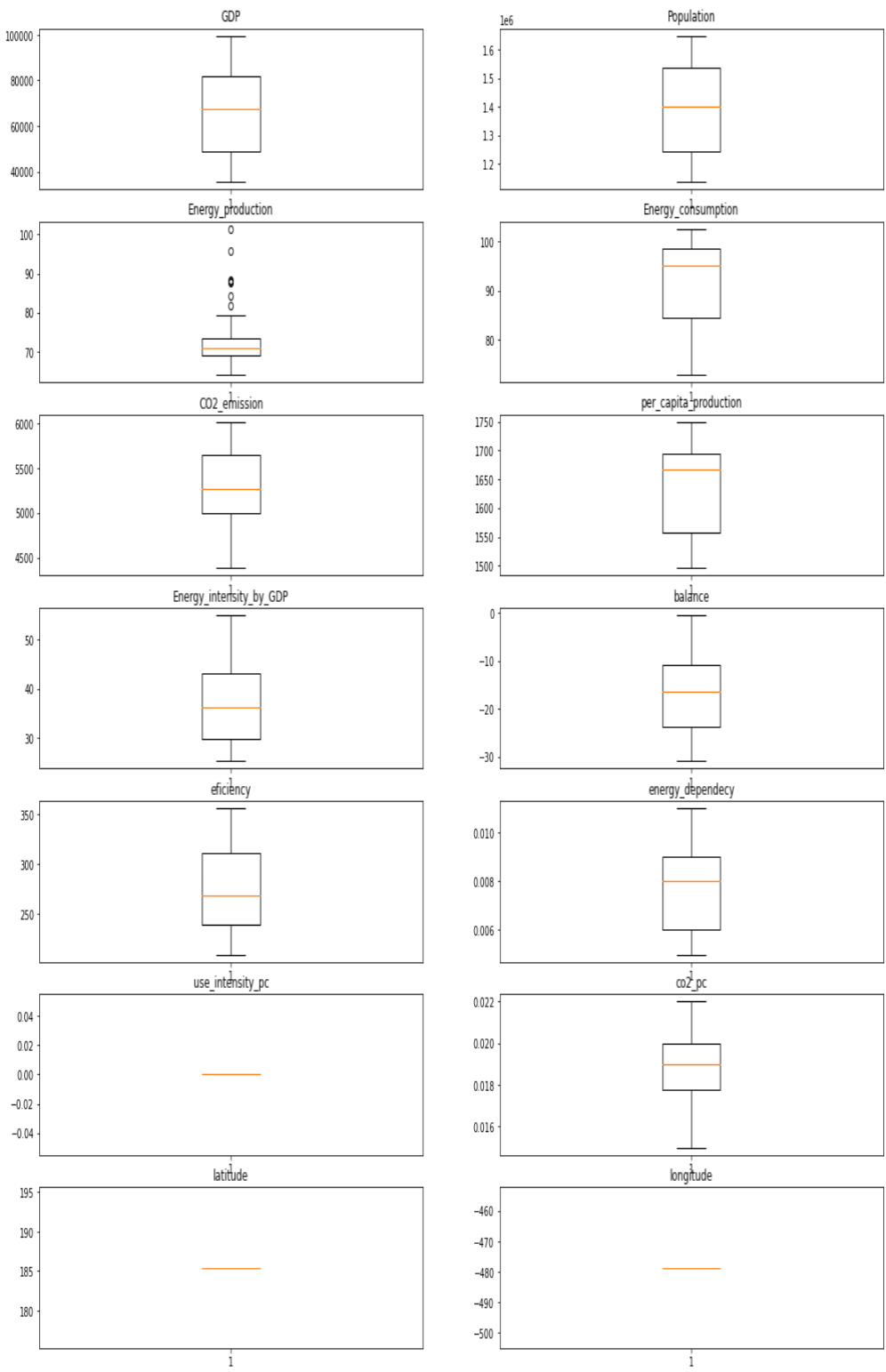


ANEXO 8, CRECIMIENTO DEL PIB POR PAÍSES (PPP):

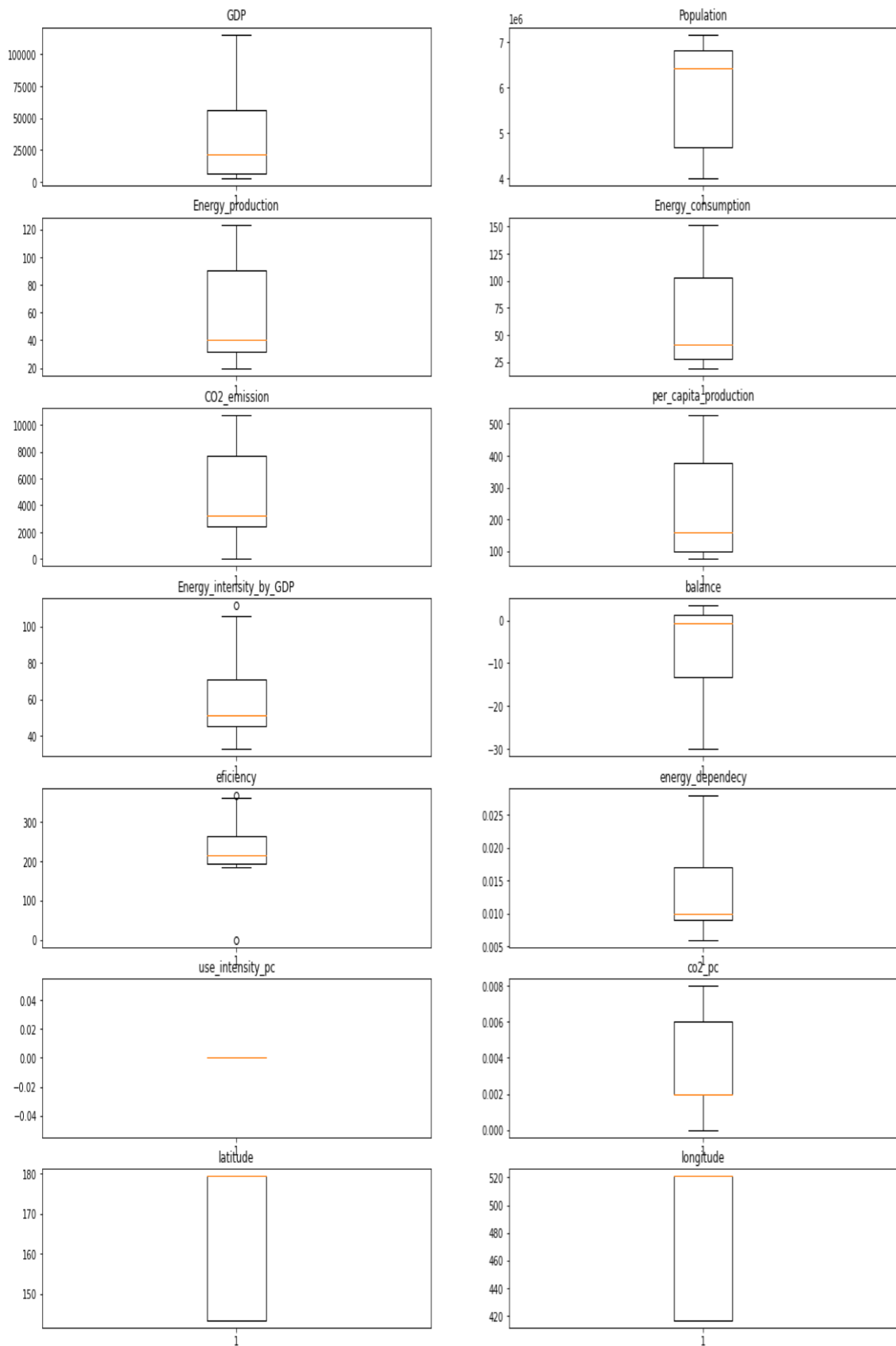


ANEXO 9:

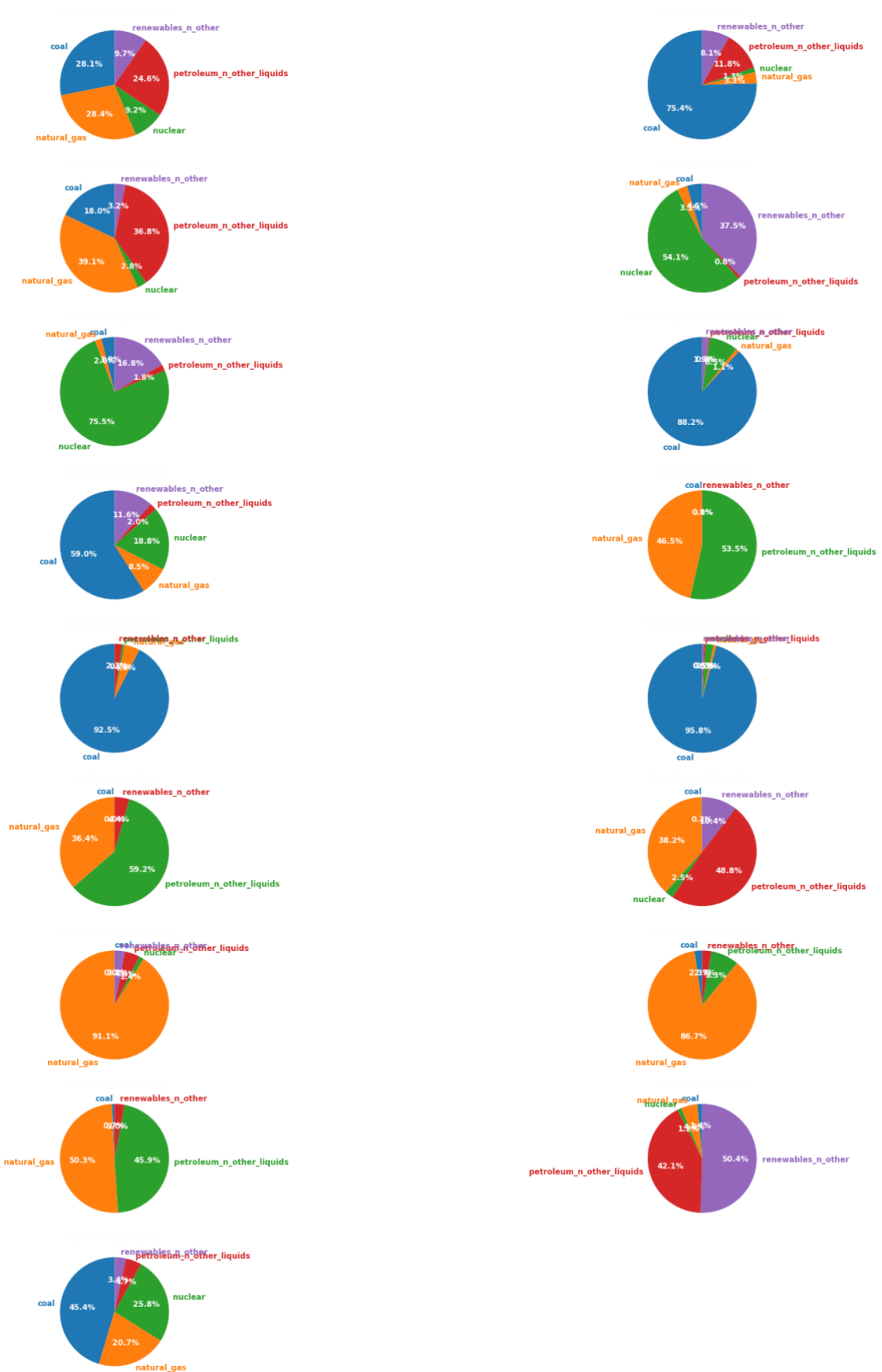
BOXPLOT DE USA:



ANEXO 10:
BOXPLOT CHINA:



ANEXO 11, PROPORCIÓN DE ENERGIA USADA POR PAÍSES:

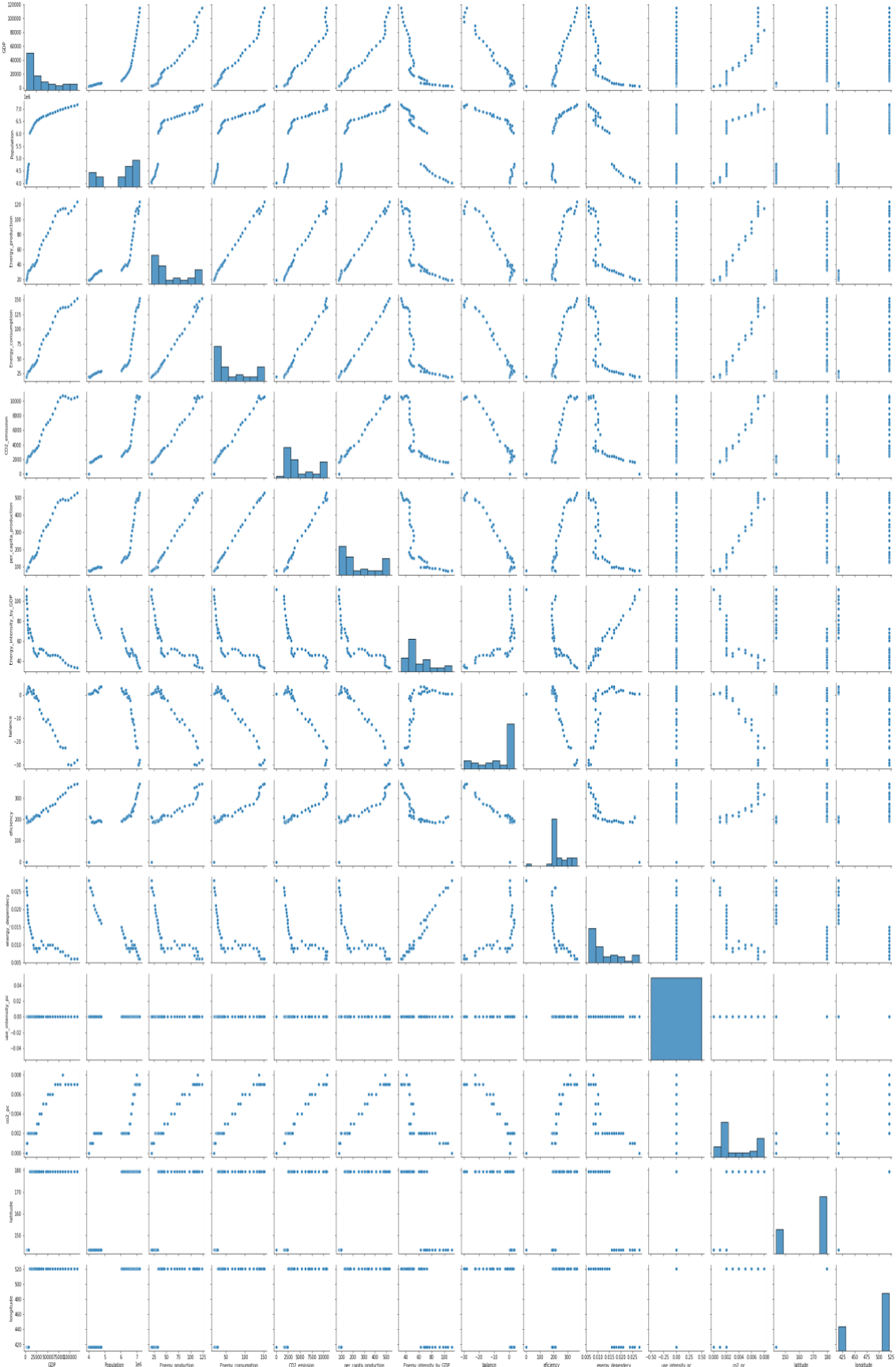


ANEXO 12:

MEMORIA: EMISIONES DE CO2

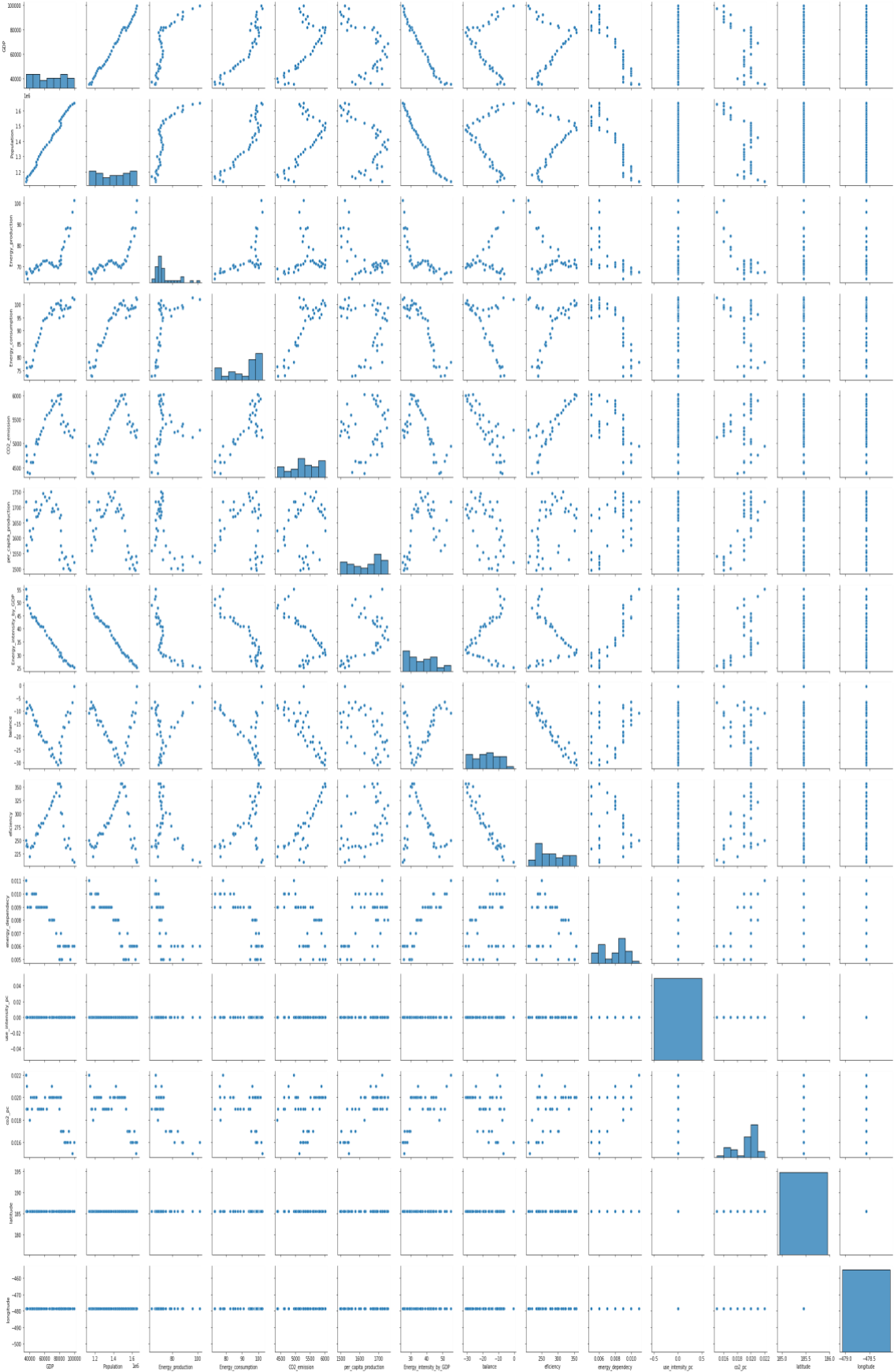
FERNANDO SÁNCHEZ OLMO

PAIRPLOT CHINA:

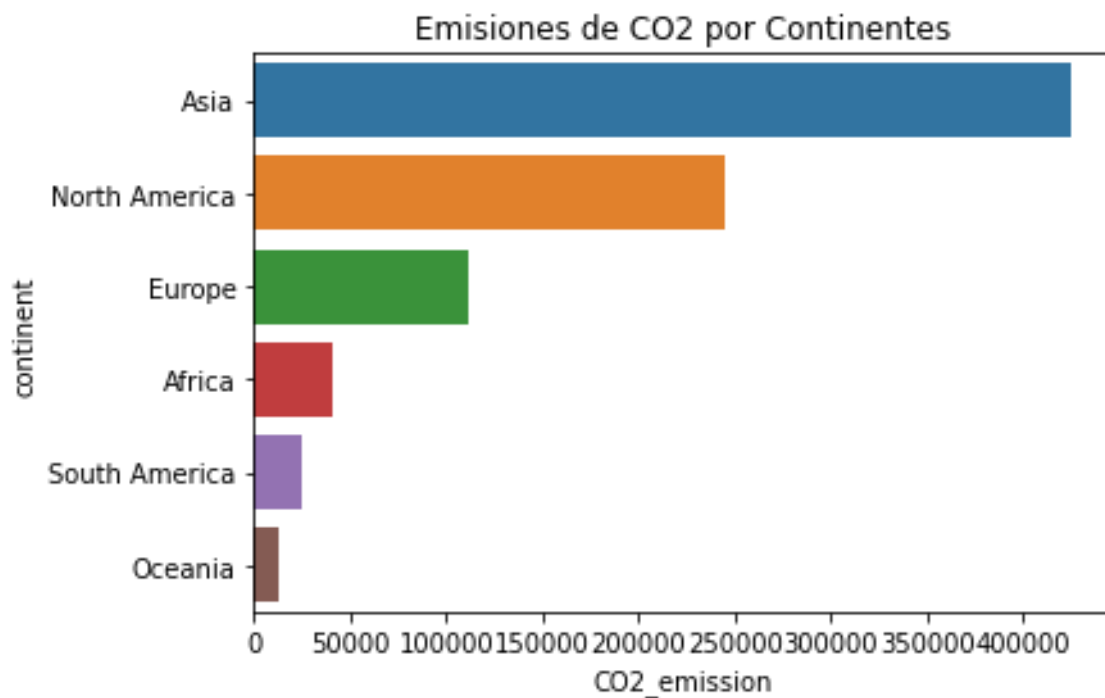


ANEXO 13:

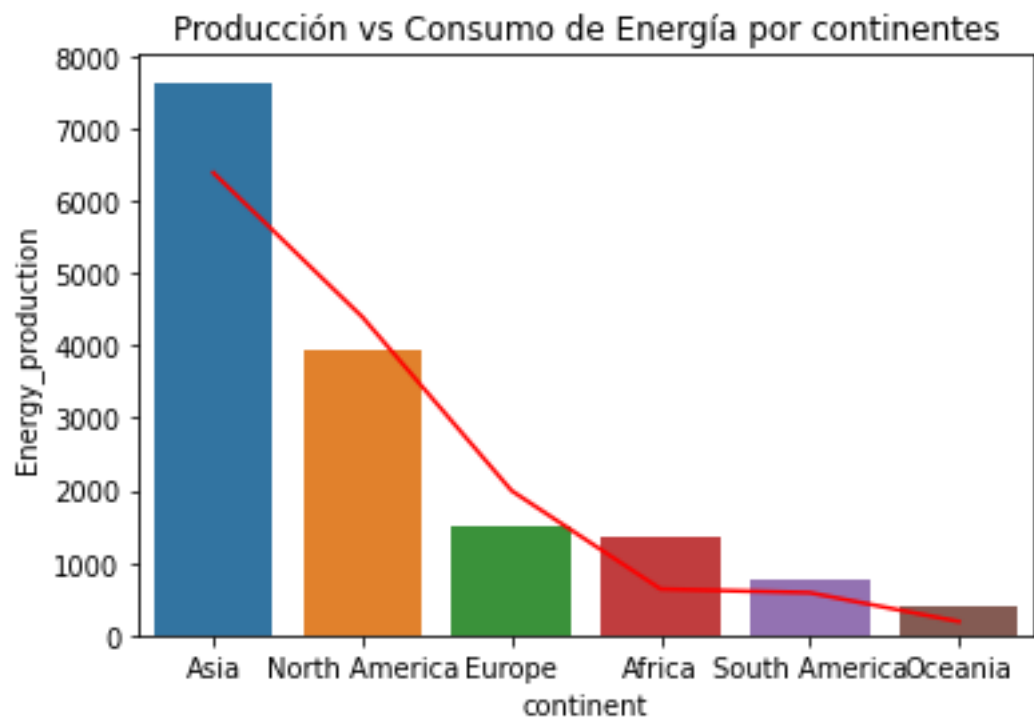
PAIRPLOT USA:



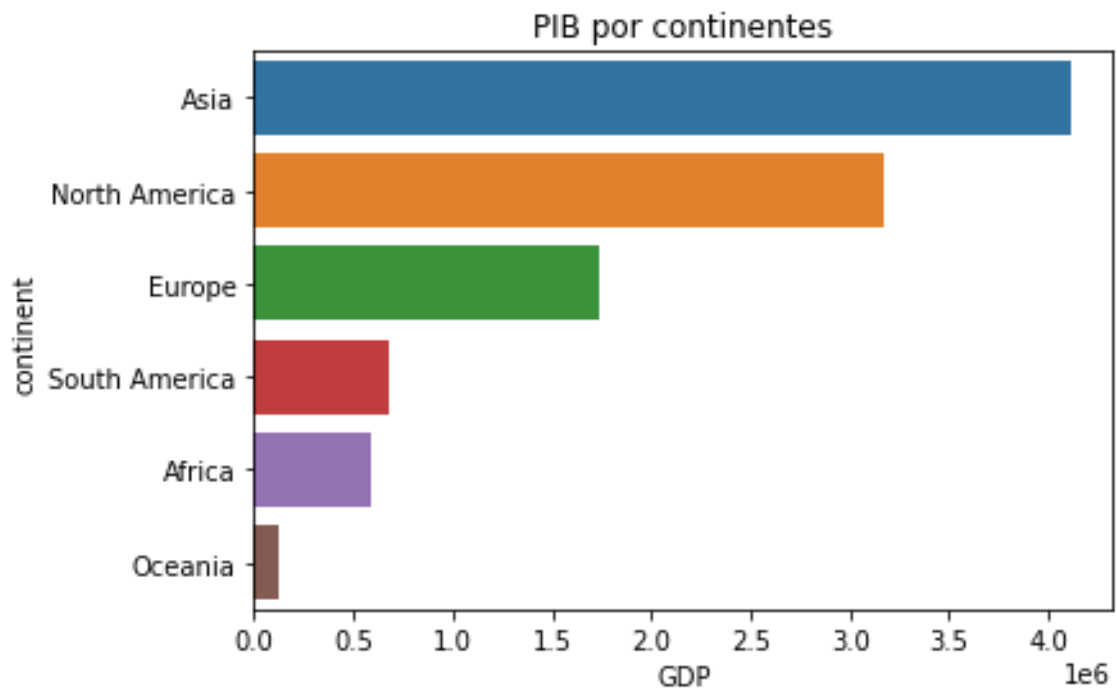
ANEXO 14, EMISIONES DE CO2 POR CONTINENTES:



ANEXO 15, PRODUCCIÓN VS CONSUMO DE ENERGÍA POR CONTINENTES:

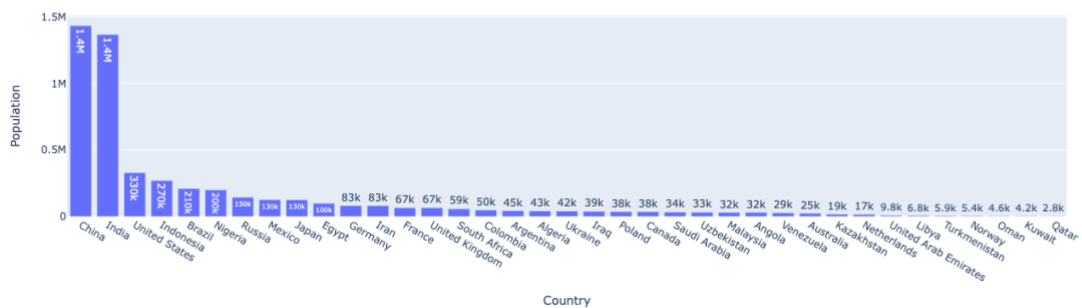


ANEXO 16, PIB POR CONTINENTES:



ANEXO 17

Población 40 Mayores Países Productores de Energía del Mundo



ANEXO 18

