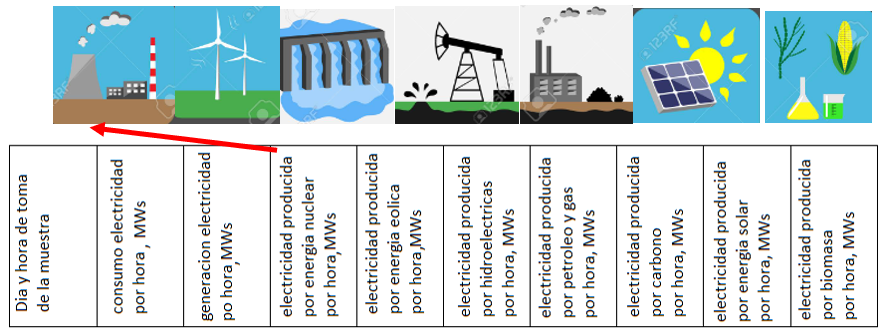
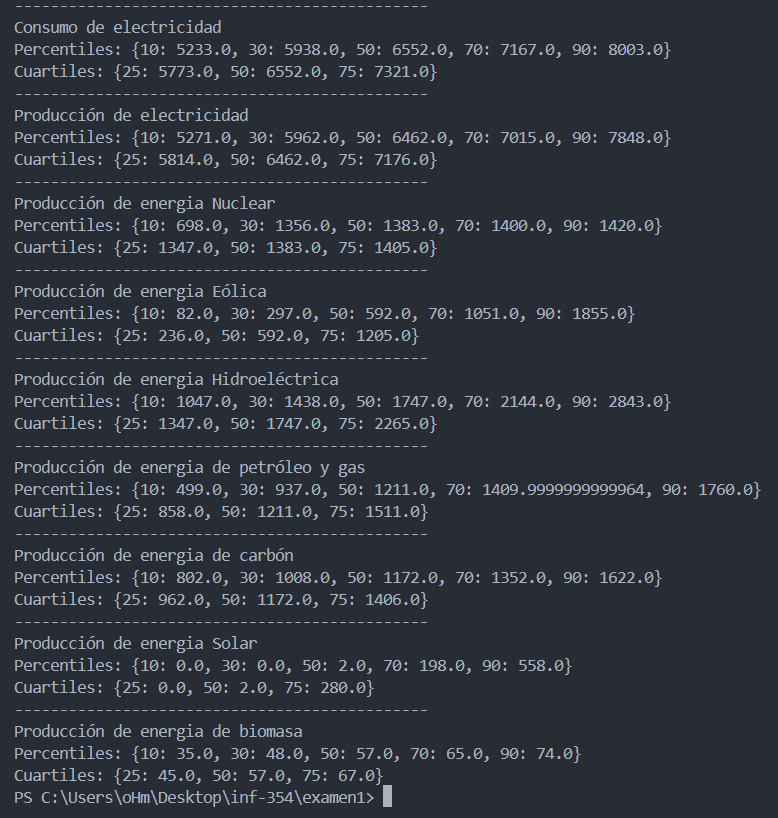
Dataset Electricidad en Rumania 10 columnas, 46012 filas

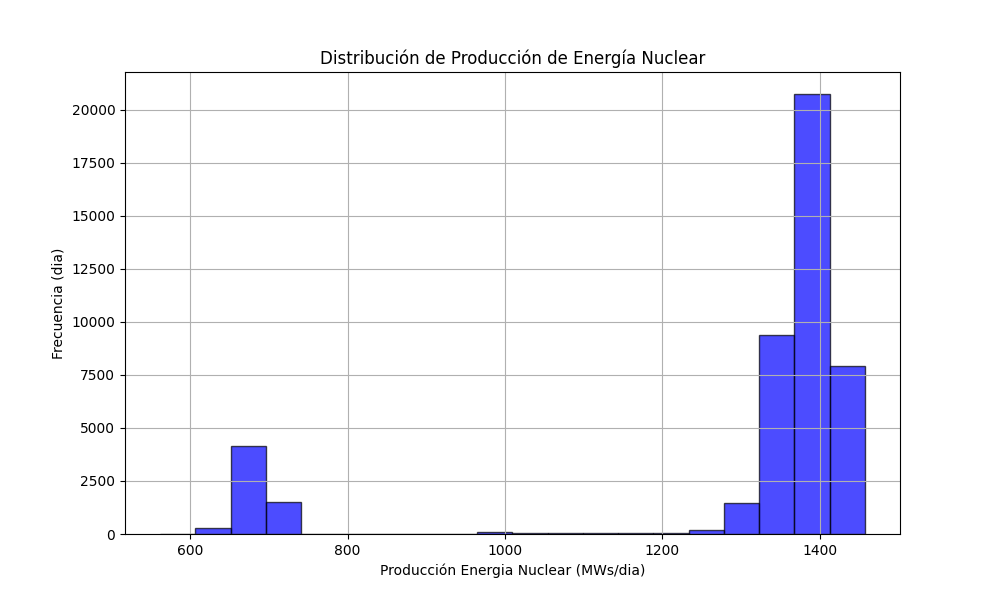
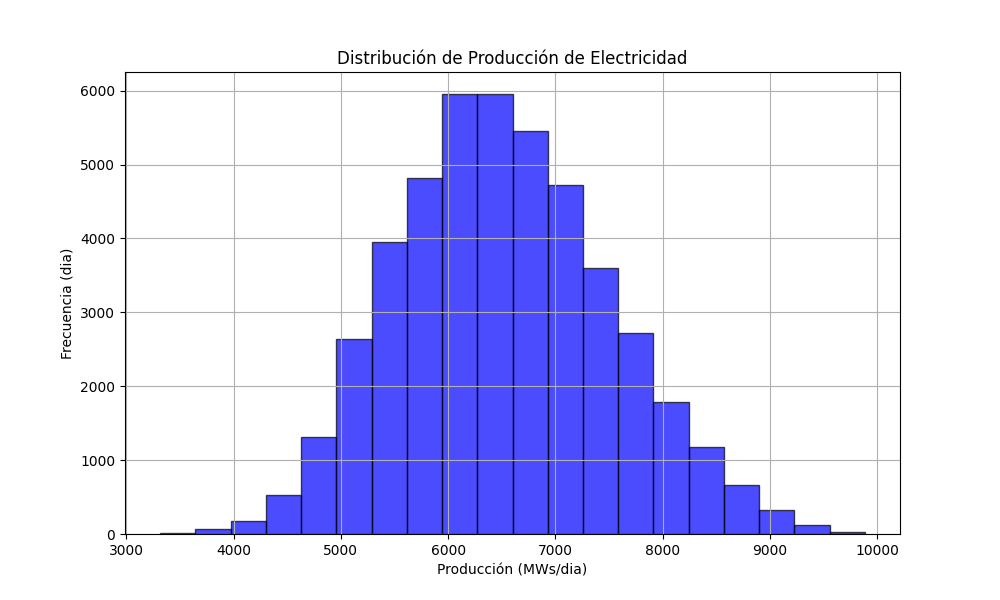
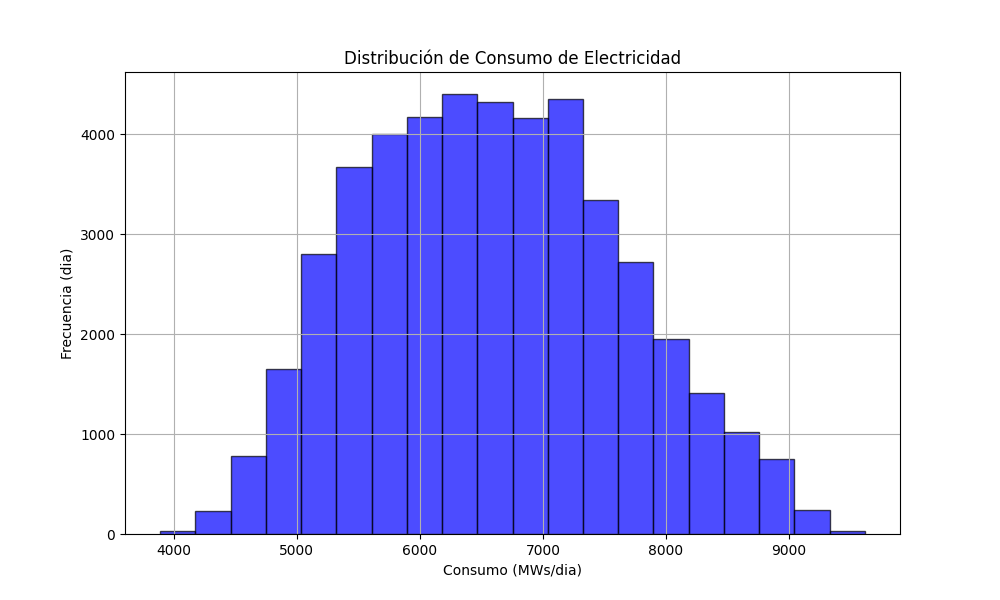


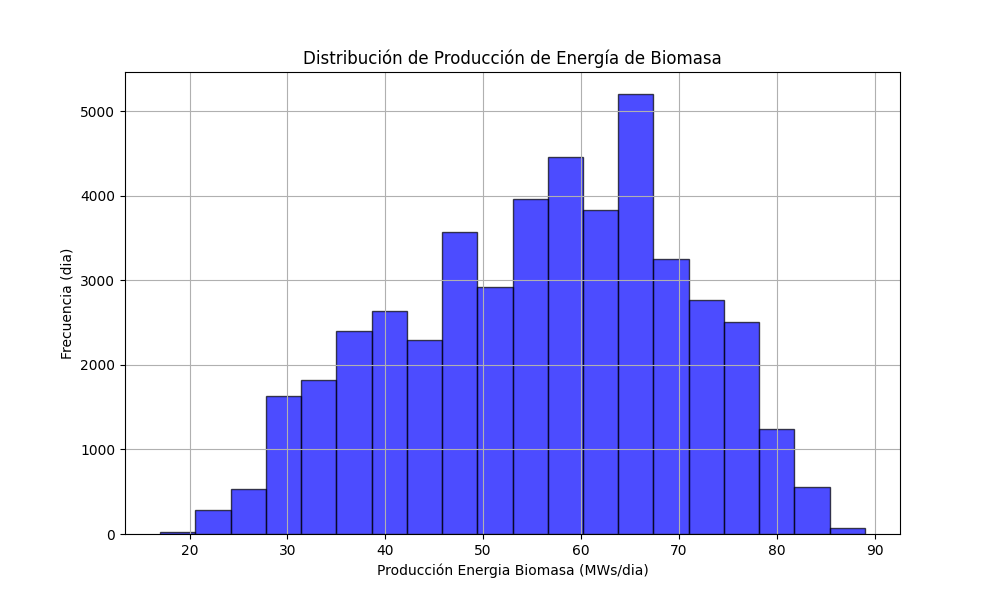
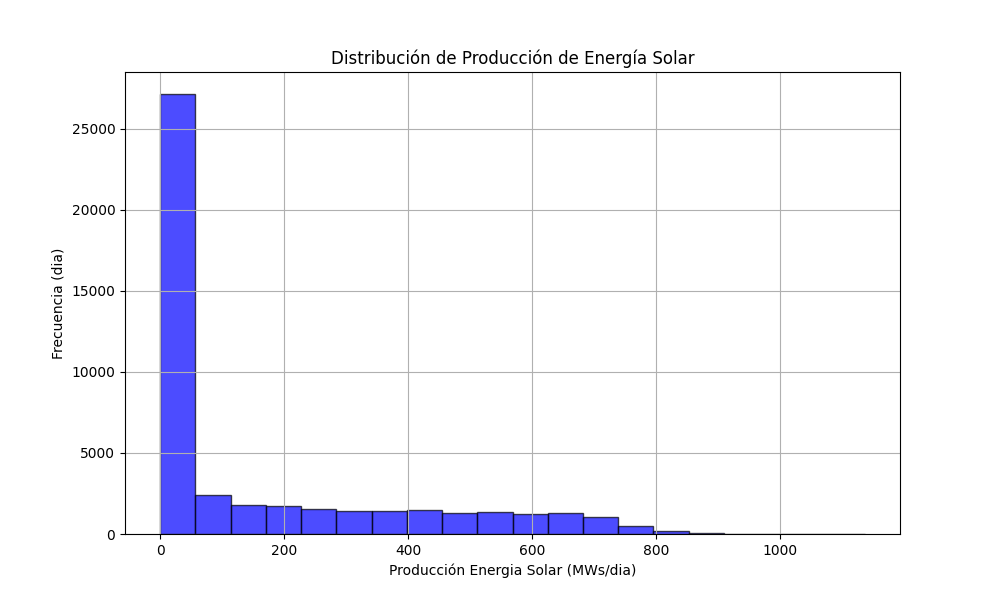
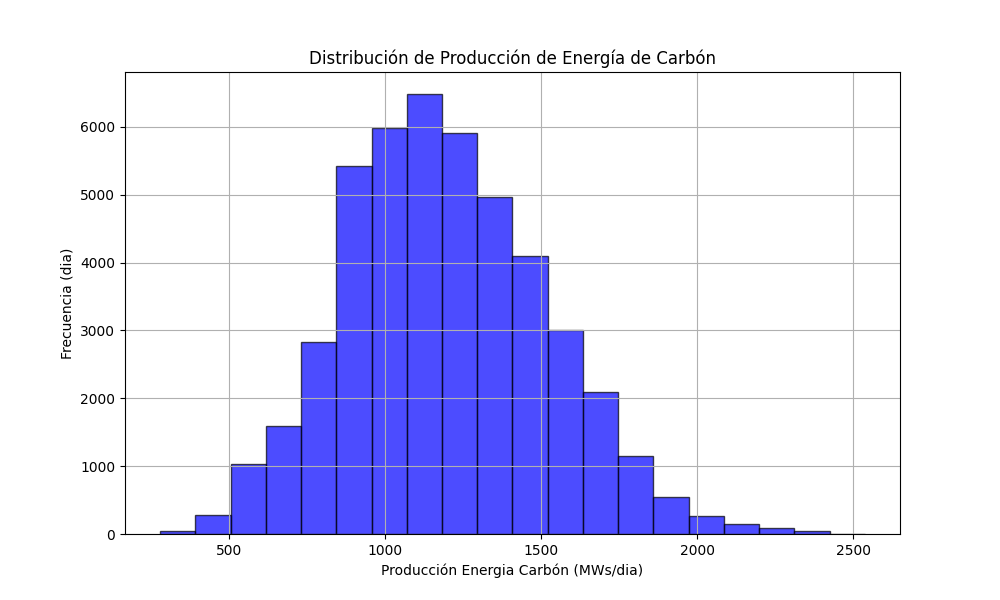
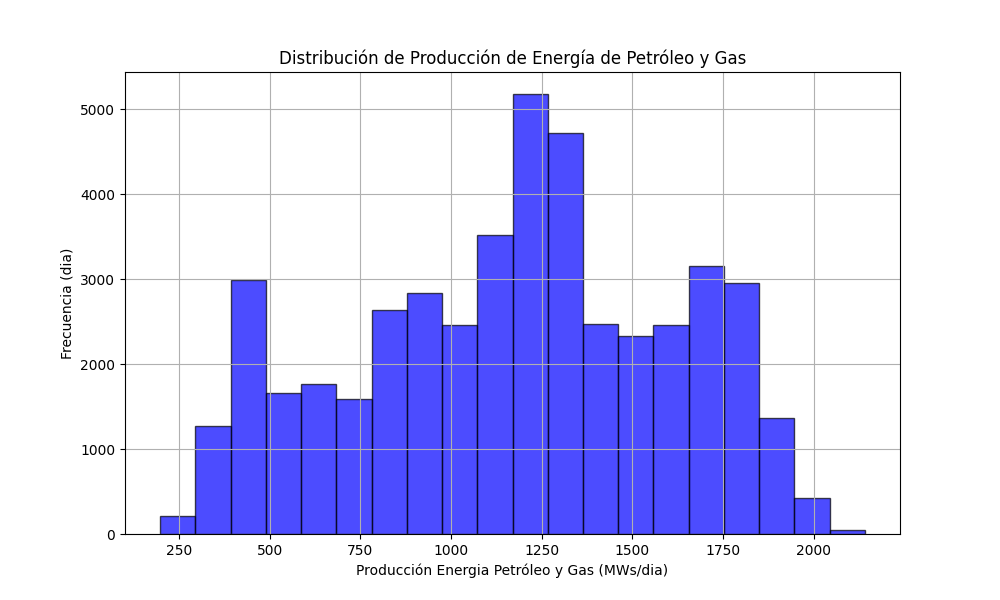
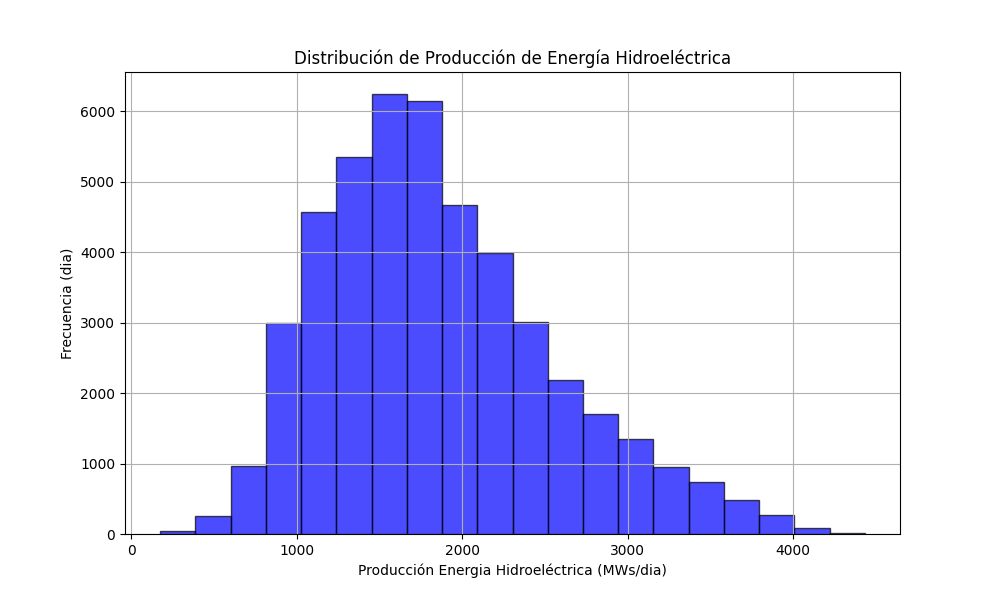
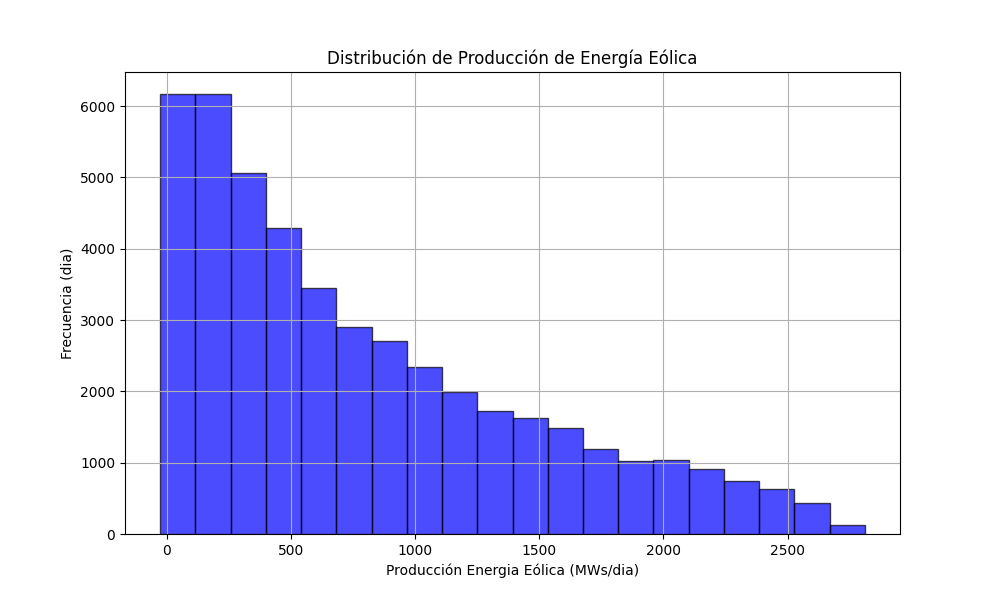
1. Sin el uso de librerías en Python programe el percentil y cuartil de cada columna. Que distribución se puede aplicar en su caso normal, Bernoulli, gaussiana, poisson, otros. Indique la razón de su uso graficando con matplotlib.

Cálculo del percentil ( 10 , 30, 50, 70 y 90)%, cuartiles (25, 50, 75)% de cada columna, excepto día y hora de toma de muestra



-**Normal (Gaussiana):** se usará esta distribución normal. Esto para verificar visualmente con un histograma o un gráfico de densidad el comportamiento de la energía eléctrica tanto de generación como de consumo en funcion de la frecuencia.





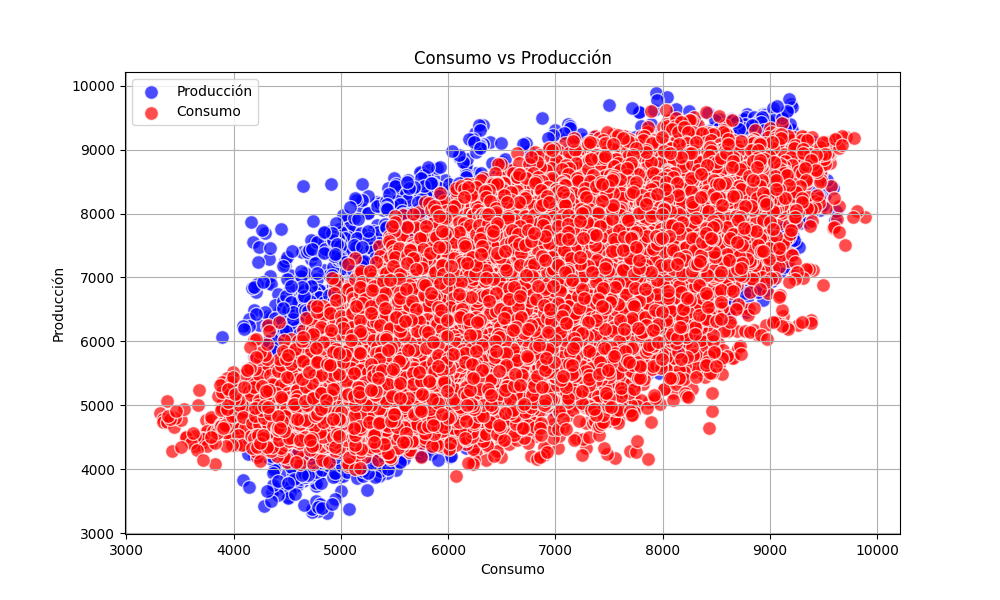
b.      De al menos tres columnas seleccionadas por usted indique que datos son relevantes de estas, grafique la misma (puede ser dispersión o mapa de calor, otros), indique al menos 4 características por columna seleccionada.

Seleccionare 4 columnas :

-Consumo de electricidad (MWs/hora)

-Generación de electricidad (MWs/hora)

Entre estas primeras 2 , are una comparación de consumo vs producción de electricidad por dispersión:



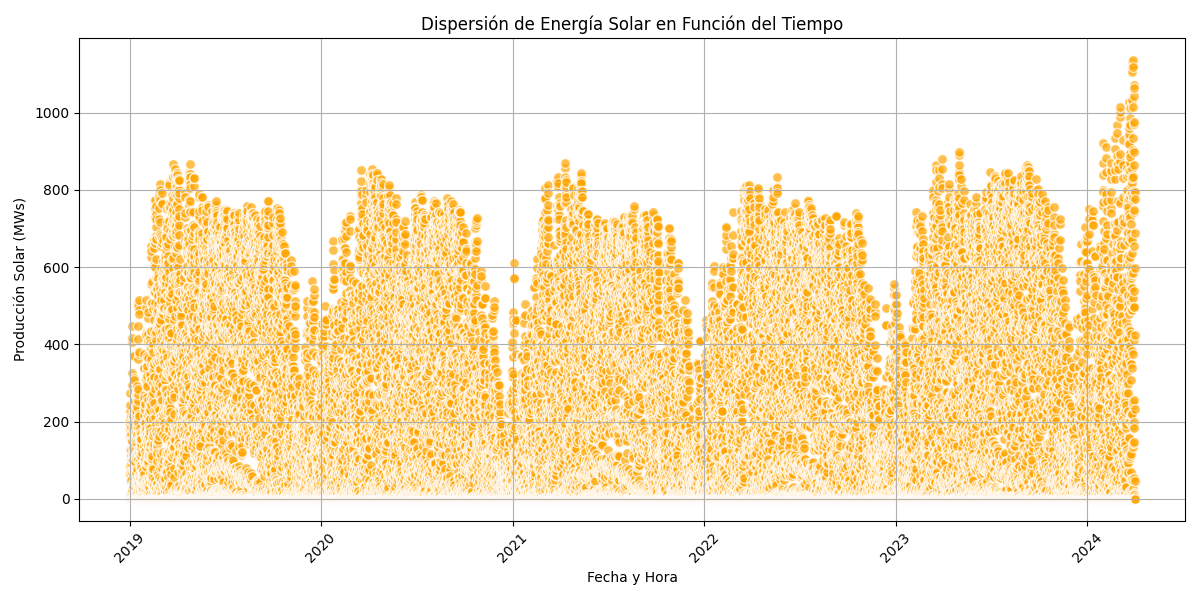
Características:

* **Tendencia lineal positiva**: Existe una relación lineal entre el consumo y la producción de electricidad. A medida que el consumo aumenta, también lo hace la producción, lo que sugiere que la generación está ajustándose a la demanda eléctrica.
* **Distribución más amplia en consumo**: La nube de puntos correspondiente al consumo (color rojo) parece estar más dispersa en comparación con la de producción. Esto puede indicar que el consumo es más variable, probablemente debido a fluctuaciones en la demanda de electricidad durante diferentes momentos del día o del año.
* **Producción menor que el consumo en ciertos puntos**: Hay una región donde el consumo supera la producción. Aunque no es la mayoría, en ciertos puntos de alto consumo, la producción parece no cubrir completamente la demanda, lo que podría ser una señal de deficiencia temporal o variaciones en la capacidad de generación.
* **Acumulación en rangos medios**: Tanto el consumo como la producción están más concentrados entre los rangos de 5000 a 8000 MWs/hora, sugiriendo que este rango es el más común tanto para la generación como para el consumo durante el periodo analizado.

- Día y hora de toma de muestra

-Generación de electricidad Solar (MWs/hora)

are la gráfica de producción de electricidad solar por dispersión, en función del tiempo (5 años DateTime):



Características

 **Patrón estacional anual**: Se observa una clara repetición anual en la producción de energía solar. Hay picos y valles regulares que coinciden con los cambios de estaciones, lo que sugiere que los niveles de generación solar son significativamente más altos durante ciertas épocas del año, como el verano, cuando la exposición al sol es mayor.

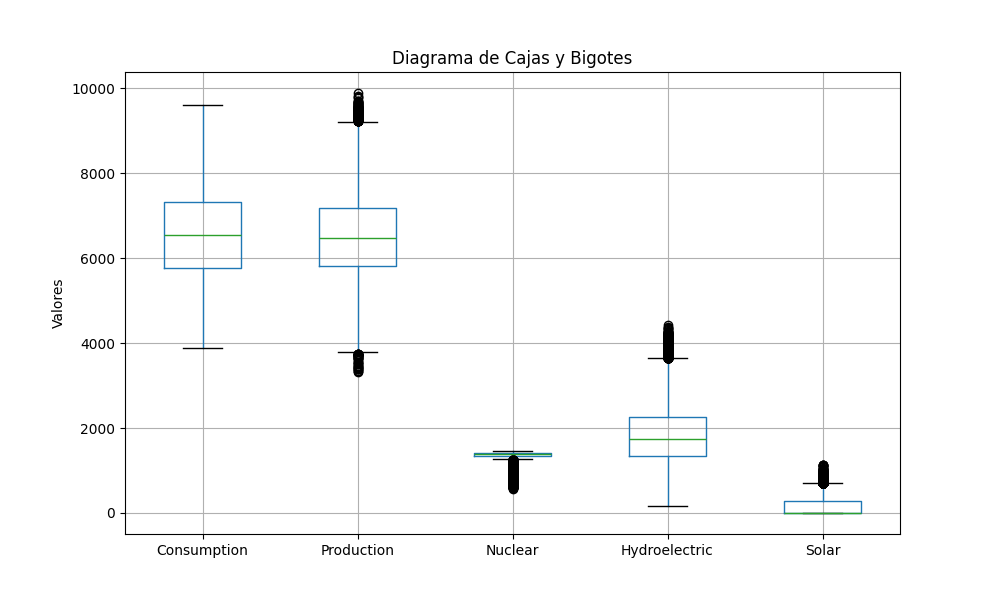
 **Disminución notable en invierno**: Entre cada año, se puede ver una caída en la generación de energía solar durante ciertos meses, que probablemente corresponden a los meses de invierno, cuando los días son más cortos y hay menos luz solar disponible. Este patrón indica una dependencia estacional de la energía solar.

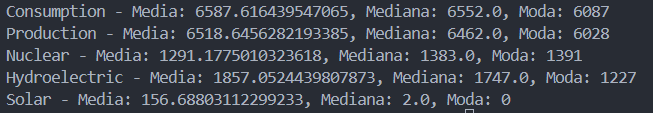
 **Fluctuación en el máximo de producción**: Aunque los picos de producción son regulares, hay una ligera variabilidad en los máximos de cada año, lo que puede deberse a condiciones climáticas (nublados o tormentas) que afectan la eficiencia de la generación solar. Esto sugiere que la producción solar, aunque cíclica, no es completamente predecible debido a factores meteorológicos.

 **Incremento progresivo en el tiempo**: Se nota que a medida que pasan los años, especialmente a partir de 2023 y entrando en 2024, la producción solar parece aumentar en general, con más puntos concentrados hacia los niveles superiores de generación. Esto podría deberse a un aumento en la capacidad instalada o mejoras en la tecnología de generación solar.

c.       Obteniendo la media, mediana, moda con el uso de librerías, grafique un diagrama de cajas-bigote de al menos 3 columnas. Explique el resultado.

Las columnas serán : consumo, producción general , producción nuclear, producción Hidroeléctrica y producción solar de energía eléctrica





1. **Consumo**:

El rango intercuartílico (dentro de la caja) es amplio, lo que indica una variabilidad considerable en los datos de consumo.

Los bigotes (líneas que salen de la caja) son largos, mostrando que los valores extremos están distribuidos de manera relativamente amplia, aunque no se observan valores atípicos claros.

La mediana (la línea dentro de la caja) parece estar centrada, lo que indica una distribución simétrica de los datos de consumo.

2**. Producción General:**

El comportamiento es similar al del consumo, con un rango intercuartílico considerable, pero parece haber una mayor variabilidad.

Los bigotes también muestran una amplitud grande, sugiriendo que la producción de energía varía significativamente.

Hay algunos valores atípicos (los puntos fuera de los bigotes), lo que indica que en algunos momentos la producción ha sido anormalmente alta o baja.

3. **Producción Nuclear:**

La mediana está hacia la parte inferior de la caja, lo que sugiere una distribución sesgada hacia valores más bajos.

Hay varios valores atípicos por encima y por debajo del rango esperado, indicando que la producción nuclear puede fluctuar bastante en algunos momentos.

4. **Producción Hidroeléctrica:**

La caja es más pequeña en comparación con el consumo y la producción general, lo que sugiere una menor variabilidad en la producción hidroeléctrica.

Los bigotes también son cortos, lo que implica que la mayoría de los datos están concentrados cerca de la mediana.

No hay muchos valores atípicos, lo que sugiere una producción más estable.

5. **Producción Solar:**

Es la categoría con la menor variabilidad (caja pequeña y bigotes cortos).

La mediana está cerca del límite inferior de la caja, lo que indica que la mayoría de los datos están en la parte baja del rango de producción.

Algunos puntos atípicos sugieren que, ocasionalmente, la producción solar es mucho mayor que lo usual.