

Análisis de un país como prosumidor energético

1. *Resumen*
2. *Preprocesamiento de los datos*
3. *Interpretación de los Análisis*
 - a. *Análisis por temporadas*
 - b. *Análisis por COVID-19*
 - c. *Análisis por grandes festivos*
4. *Estacionariedad Dickey-Fuller*
5. *Regresión Autorregresiva Integrada de Media Móvil (ARIMA)*
6. *RNN*
 - a. *Por días*
 - b. *Por meses*
7. *Conclusiones*

1. Resumen

Para este proyecto contamos con dos dataset diferentes que muestran la demanda y el suministro(producción) de las provincias de Irak por horas. Este proyecto es una prueba de concepto, en adelante PoC ,de mi Trabajo de Fin de Máster, el cual comparto con un compañero y debido a esto mi estudio será del total del País y no a nivel de provincias.

En este estudio trataremos de ver el comportamiento de la demanda y producción del país a lo largo del tiempo, gracias a los datos verificados que constan desde el 01-01-2019 hasta el 02-01.2022. Analizar solamente dos variables puede ser una tarea muy complicada ya que se necesita ser muy fino a la hora de aplicar modelos que sean los correctos para obtener resultados interpretables. Por ello en esta PoC, en consecuencia de la naturaleza de los datos, vamos a investigar sobre series temporales y diferentes prácticas para poder analizar el comportamiento de la energía, además de buscar relaciones con elementos externos que puedan influir directamente. Un aspecto muy importante de este estudio ha sido la búsqueda de información y obtener datos veraces, debido a que la política exterior de este país es muy restrictiva en estos casos.

El planteamiento inicial era investigar diferentes modelos que aplicar a problemas de series temporales, pero debido a nuestro desconocimiento en series temporales y Deep Learning el alcance inicial tuvo que sufrir cambios de rumbos. Así que primero planteamos estudiar el marco social-económico del país sacando así información sobre eventos que afectan directamente a estos valores, con esto me refiero a estaciones del año, eventos extraordinarios como la COVID-19, días de grandes festivales en el país, guerras...

Una línea bastante importante que detectamos, una vez hecho un análisis inicial de los datos, fue que la producción no abastecía a la demanda. Por lo cual aquí se abrieron dos caminos, uno para ver las principales fuentes de energía del propio país y otro para la importación de esta. Como comentamos anteriormente, obtener datos de este país nos ha resultado bastante complejo por su política exterior.

Completados estos análisis, se decidió estudiar la estacionariedad para ver que su media y varianza permanecen constantes a lo largo del tiempo. Para ello aplicamos el test de Dickey-Fuller. Una vez obtenido los resultados de este test pasamos a construir un modelo autorregresivo integrado de media móvil, más bien conocido como ARIMA sobre la demanda, ya que el test de Dicker-fuller nos demostró que es una serie temporal estacionaria.

Una vez estudiados estos resultados, decidimos pasar a modelos de Deep Learning. Tiramos por la aplicación de redes neuronales recurrentes (RNN) debido a que son las redes especializada en procesar series temporales cuya arquitectura permite que la red obtenga memoria artificial, es decir se distinguen porque contiene un "memoria" que guardan los valores de entradas anteriores para los nuevos datos de entradas y sacar así resultados. En concreto hemos usado la arquitectura de memoria a largo-corto plazo, LSTM, abordando así dependencias a largo plazo.

2. Preprocesamiento de los datos

Para esta variedad de análisis y pruebas hemos tratado de diferentes formas nuestro dataset base. La mayoría de estos cambios se han ido a medida que se necesitaban para los diferentes modelos.

Partimos de que tenemos dos dataset uno para la demanda y otro para la producción, ambos con una columna para la hora, donde la primera fila es corresponde a las 1:00 del 01-01-2019 y la última a 24:00 del 02-01-2022, una columna con la demanda o producción por cada una de las provincias del país. Todos los tratamientos se han realizado en ambos dataset, aunque mostremos a expliquemos uno solo.

El primer tratamiento que realizamos consiste en sacar el total por cada una de las horas y sin las horas, producción así 4 datasets para diferentes análisis.

total		total Hours	
0	11479	0	11479 1
1	10595	1	10595 2
2	10132	2	10132 3
3	10437	3	10437 4
4	11576	4	11576 5

Ahora pasamos a obtener la demanda o producción por días, también mergueamos ambos dataset y lo indexamos según la fecha para obtener una serie temporal correctamente estructurada.

			Demanda Produccion		
			fecha		
			2019-01-01	379726	289005
			2019-01-02	383590	288843
			2019-01-03	378901	283581
			2019-01-04	378180	285621
			2019-01-05	385805	283991
		
			2021-12-29	411592	261122
			2021-12-30	378184	248905
			2021-12-31	414996	257533
			2022-01-01	437452	272710
			2022-01-02	417780	277922
Total	fecha				
0	379726	2019-01-01			
1	383590	2019-01-02			
2	378901	2019-01-03			
3	378180	2019-01-04			
4	385805	2019-01-05			

Para RNN se ha cambiado el index a meses, para de esta manera ver las diferencias entre días y meses.

Otro cambio realizado es eliminar los días de 2022 de cara a graficar y predecir valores, ya que solo contábamos con unos pocos días de ese año y esto podría afectar a los resultados.

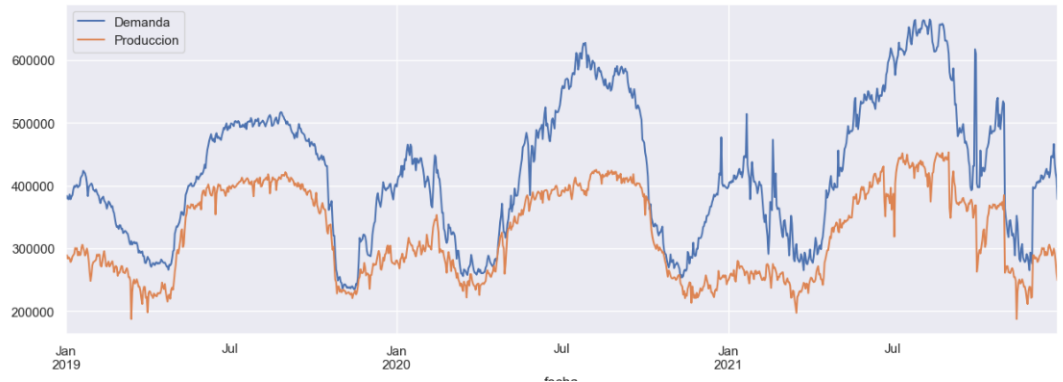
Para el análisis por temporadas se ha añadido una columna que nos informa de la temporada de ese día. Esta distribución se ha realizado según la temporalidad de ese país, la cual es peculiar debido a su ubicación geográfica.

3. Interpretación de los Análisis

Ahora pasamos a desarrollar diferentes análisis según eventos importantes del país

a. Análisis demanda frente produccion

En esta visualización vemos que la producción de electricidad en Iraq no llega siempre a satisfacer la demanda del consumidor.



En un país tan energético como Iraq según nuestra investigación la producción de la electricidad no es 100% local sino 69% aproximadamente pertenece generalmente a los tres siguientes recursos: Hydro %7.6, Oil %30.3 , y Gas natural %62.1.

b. Análisis por temporadas

Generalmente in Iraq hay 3 climas regionales:

- Clima mediterráneo o húmedo.
- Clima semi-desértico.
- Clima desértico.

En estas tres regiones las temperaturas difieren en principio y fin pero en general las temperaturas en Iraq están como el siguiente:

Invierno: Los meses de Diciembre, Enero, Febrero, y Marzo.

Primavera: Los meses de Abril, y Mayo.

Verano: Los meses de Junio, Julio, y Septiembre.

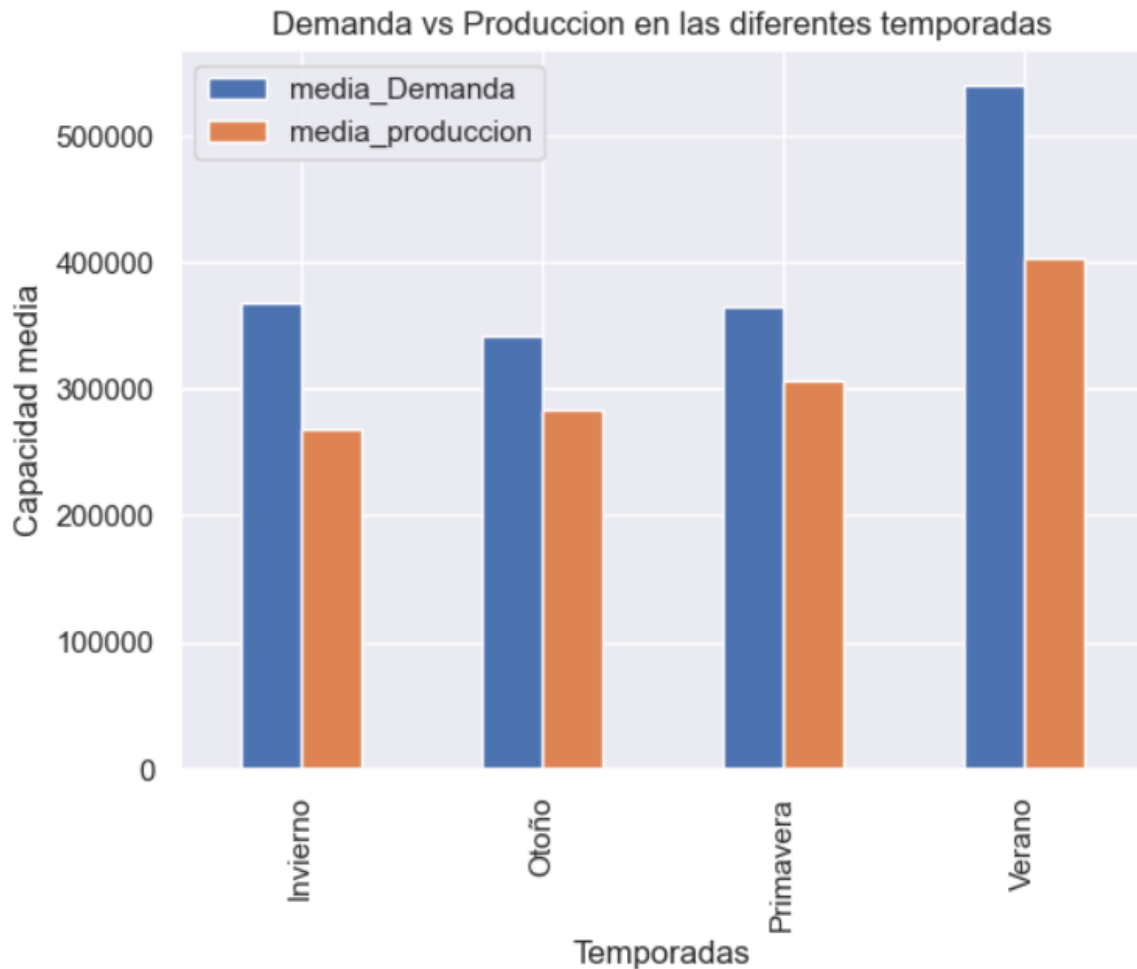
Otoño: Los meses de Octubre, y Noviembre.

En el siguiente análisis hemos creado otra dataset de la principal que tenemos donde presenta las medias de producción y demanda en cada temporada en los de arriba.

	media_Demanda	media_produccion
estacion		
Invierno	367675.931129	268132.415978
Otoño	342338.863388	283917.808743
Primavera	365714.278689	306742.420765
Verano	540844.855191	403776.814208

En la visualización siguiente generada de los datos de la dataset de los medios de producción y temporada, hemos sacado los puntos siguientes/

- Siempre en verano hay más media de demanda y más media producción.
- En el invierno aunque es la segunda temporada crítica de demanda, el estado ofrece la más baja producción, lo que necesita una revisión de equivalencia.



c. Análisis por COVID-19

Según hemos podido ver en diferentes fuentes de información el pico de contagios de covid-19 empezó en Iraq en la provincia de Najaf donde un estudiante religioso iraníano sido positivo el 24 de Febrero del 2020, la restricciones y cuarentena sido más intensa esta fecha hasta fines de abril del mismo año, entonces lo que hicimos es:

- Analizar la producción y la demanda en este periodo.
- Compararlos con los dos meses anteriores.

Este análisis nos permite tener una idea de cómo el covid ha afectado este sector.

Periodo Covid intenso:

Hemos sacado una dataset de periodo de covid intenso de la dataset principal.

fecha	Demanda	Produccion	estacion
2020-02-24	314394	274931	Invierno
2020-02-25	308539	261047	Invierno
2020-02-26	313477	282819	Invierno
2020-02-27	338249	294075	Invierno
2020-02-28	332861	286547	Invierno
...
2020-04-26	371136	325099	Primavera
2020-04-27	344921	304380	Primavera
2020-04-28	325709	300122	Primavera
2020-04-29	325771	259414	Primavera
2020-04-30	316859	274876	Primavera

67 rows × 3 columns

En la visualización siguiente se aparece como siempre en la producción es muy lejo de satisfacer demanda pero había siempre una lógica de paralelo de subida y baja, solamente hay los siguientes excepciones:

- En los días desde 12 hasta 23 de abril 2020 la producción ha podido satisfacer la demanda.
- En los días 17, 20 de Marzo, y 1,y 2 de abril no había esta lógica de paralelo en subida y baja, había una baja tremenda de producción mientras en los dos días de abril había subida de demanda.

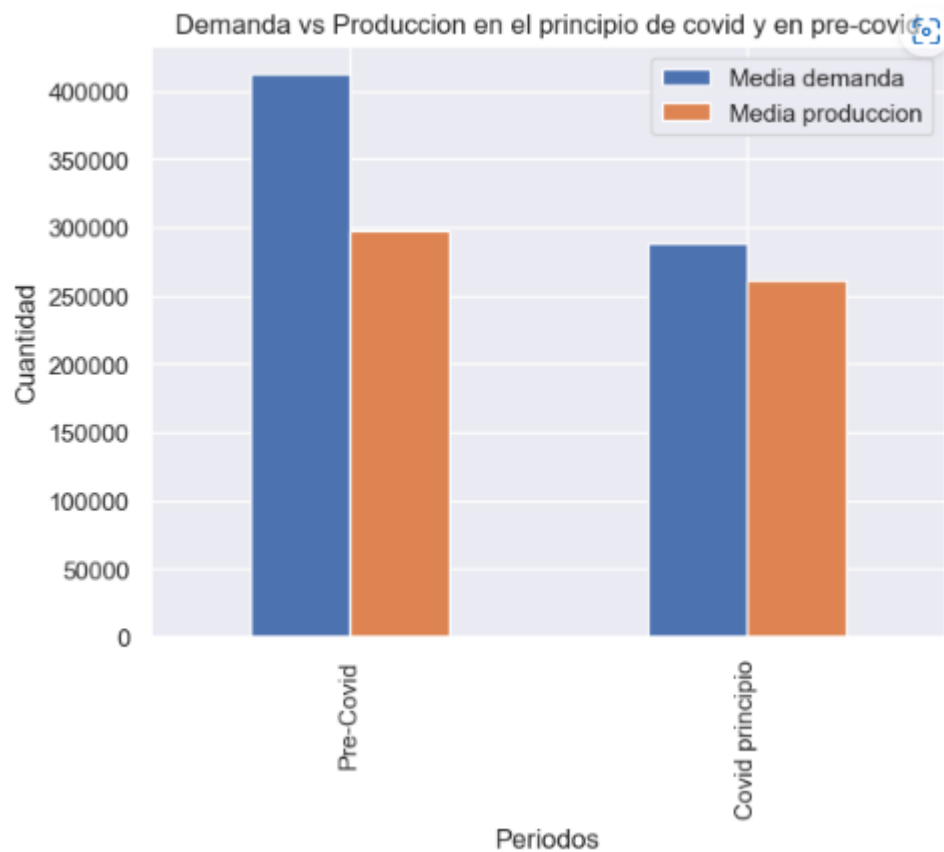


Dos meses Pre-Covid



Hemos notado que la demanda y la producción son más altos en el periodo de dos meses pre-covid que en los dos primeros meses de Covid, y eso está presentado en esta dataset y visualización de Bars siguientes que refleja las medias.

	Media demanda	Media produccion
Periodo		
Pre-Covid	412271.957746	297922.295775
Covid principio	288816.000000	261000.223881

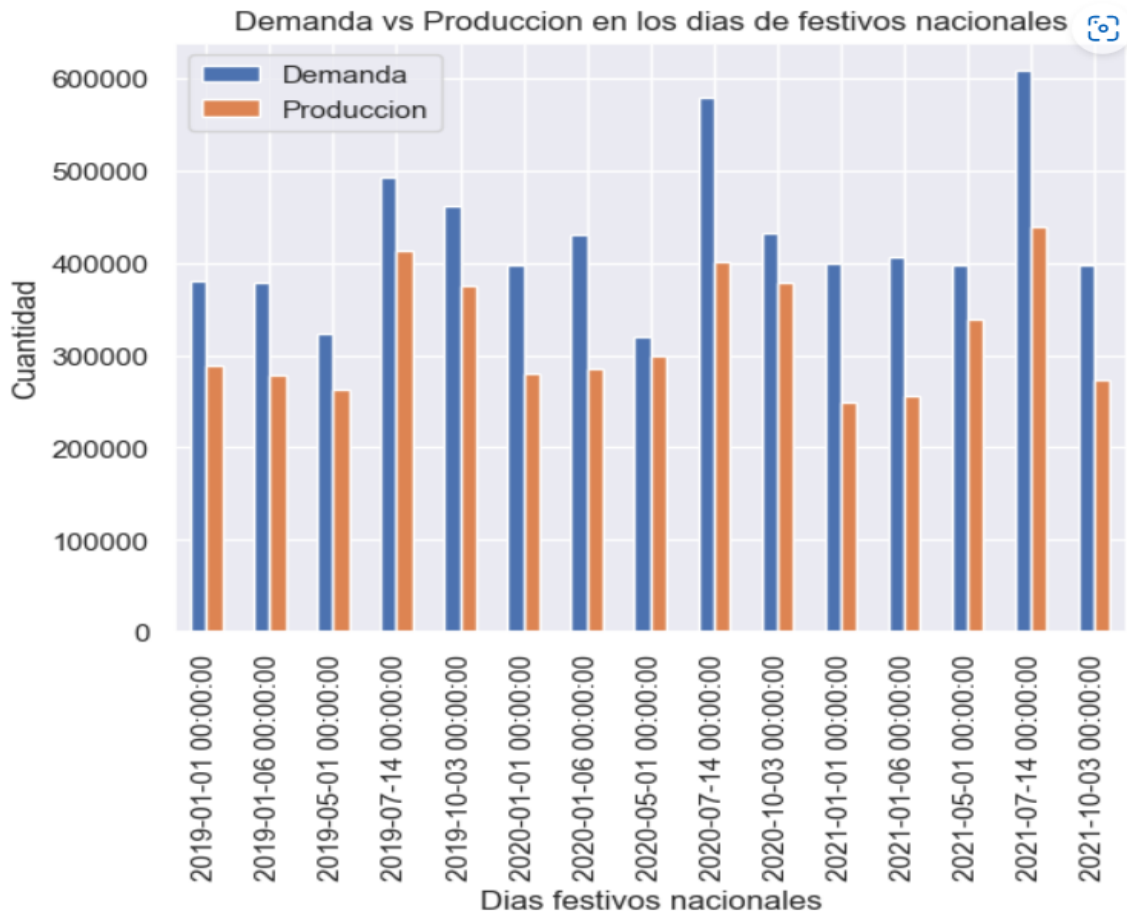


d. Análisis por grandes festivos

En esta primera parte hemos investigado en los diferentes **días nacionales de festivo**, la fecha siguiente presenta la fechas de días de festivo en Iraq en los tres años 2019, 2020, y 2021.

fecha	Demanda	Produccion	estacion
2019-01-01	379726	289005	Invierno
2019-01-06	378280	278314	Invierno
2019-05-01	323099	262720	Primavera
2019-07-14	493512	412650	Verano
2019-10-03	461922	375271	Otoño
2020-01-01	398283	281007	Invierno
2020-01-06	431218	285232	Invierno
2020-05-01	320559	299753	Primavera
2020-07-14	579399	400581	Verano
2020-10-03	431841	378241	Otoño
2021-01-01	398854	249832	Invierno
2021-01-06	406022	256305	Invierno
2021-05-01	397250	339273	Primavera
2021-07-14	608133	439941	Verano
2021-10-03	397250	273672	Otoño

En la visualización siguiente presenta la demanda y producción en los diferentes día de festivo, lo que hemos notado que el día de 14 de Julio que es el día de celebración de revolución es la más alta de demanda como en producción en los festivos, y eso es por la importancia del día y la celebración grande a nivel nacional.



Sha'ban vs Ramadán

En esta segunda parte queremos valorar la producción/consumo de electricidad en el mes sagrado de Ramadán que es un mes de ayuno y culto para los musulmanes, por eso hemos hecho esta comparación entre este mes y el mes precedente que es el mes de Sha'aban.

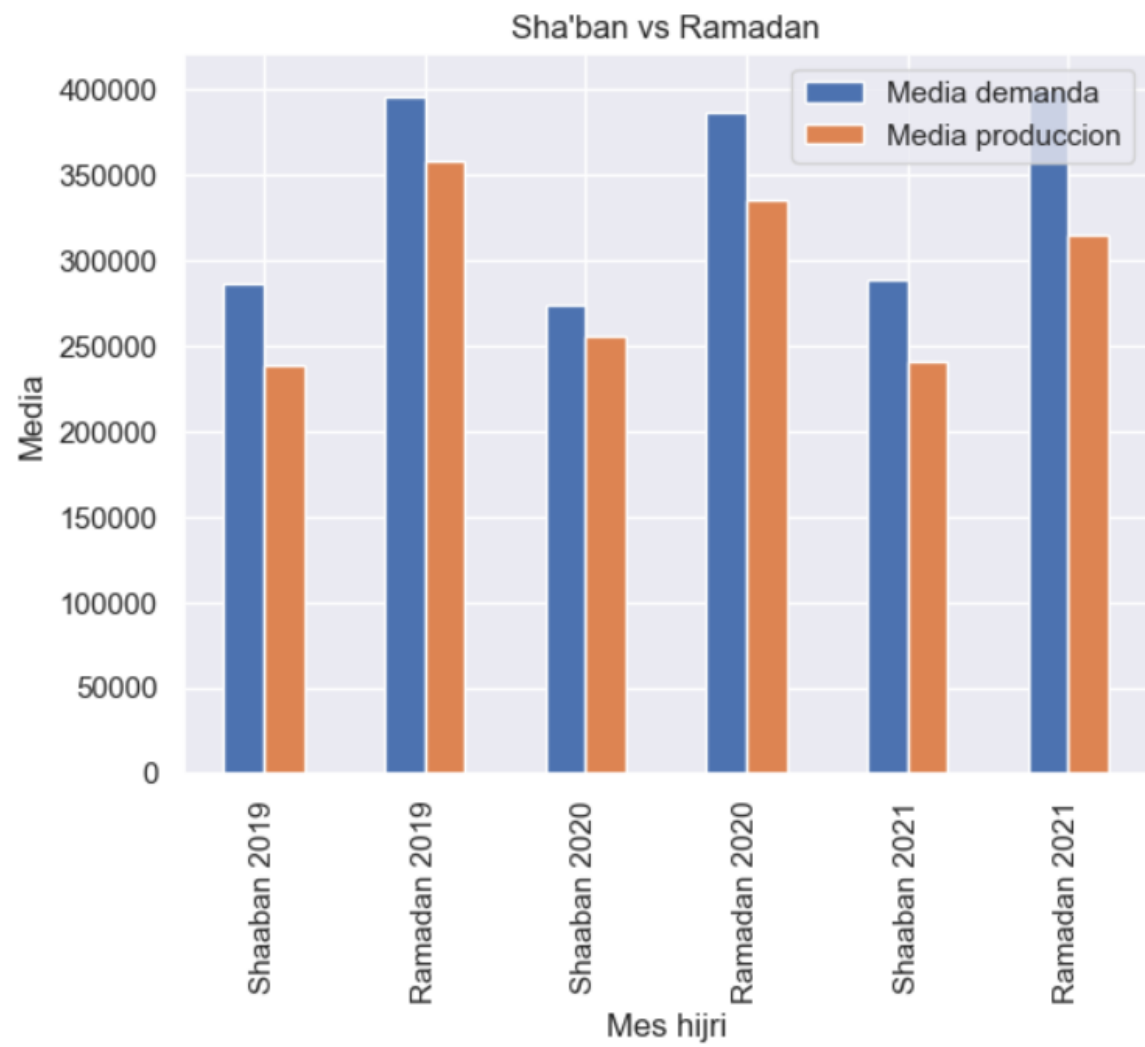
La comparación está basada en los medios de semana y producción de los dos meses en los tres años que están en la dataset general 2019, 2020, y 2021.

Por eso hemos generado una la dataset siguiente del dataset principal para tener idea sobre los medios y visualizarlos.

	Media demanda	Media produccion
Mes_hijri		
Shaaban 2019	286469.000000	238806.866667
Ramadan 2019	395727.033333	357752.066667
Shaaban 2020	273865.533333	255552.966667
Ramadan 2020	386320.000000	335347.600000
Shaaban 2021	288550.200000	240607.533333
Ramadan 2021	400432.066667	314738.500000

Los que hemos notado:

- La proconsumacion siempre sube cuando termina el mes de Sha'ban y empieza el mes de Ramadán, y eso por razón del ayuno y la sed que implica el uso intenso de aires acondicionados, y electrodomésticos en este mes sagrado.
- La producción de electricidad en el país durante el mes culto de Ramada sigue bajando en los años 2019, 2020, y 2021.

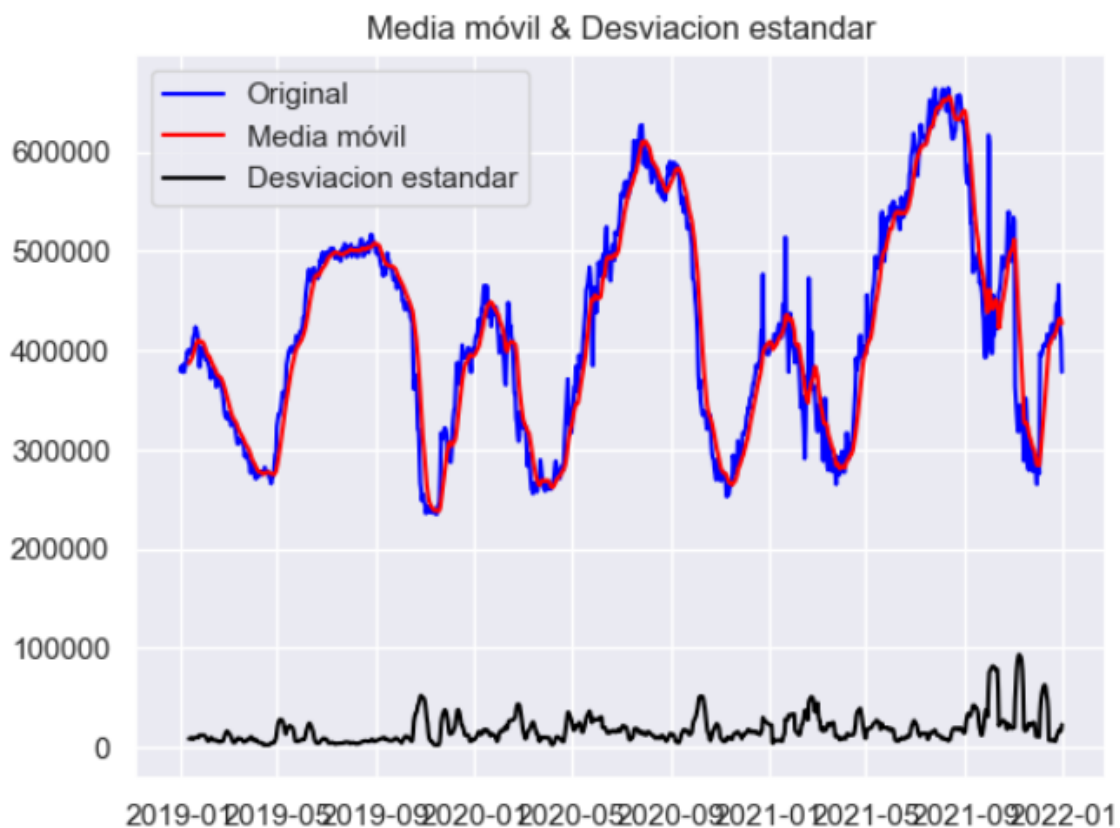


4. Estacionariedad Dickey-fuller

El test Dickey-fuller se realiza para conocer la existencia de raíces unitarias de las series temporales, siendo el caso nulo la existencia de una raíz unitaria. Con esto queremos llegar a ver si nuestros datos son estacionarios o no para usar modelos de regresión. La estacionariedad es determinada cuando su media y varianza no se alteran con el tiempo

Para ello hemos de ver los valores p-value y valores críticos para tres niveles de significancia: 0,01(1%), 0,05(5%) y 0,1(10%) ofrecidos por el test, además graficamos la media móvil y desviación estándar para ayudarnos a sacar conclusiones:

Para la demanda obtenemos los siguientes resultados:

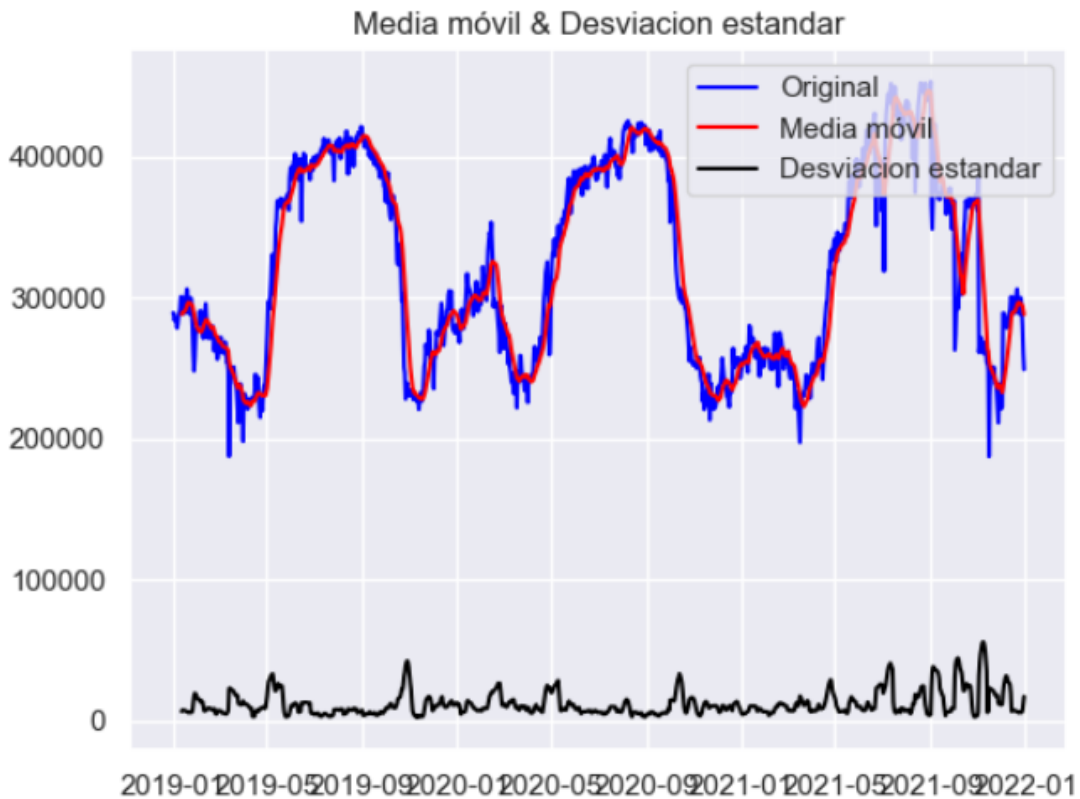


```
Resultados del Dickey-Fuller Test:
Estadístico de prueba      -3.080073
p-value                    0.028063
#Lags Used                 16.000000
Numero de observaciones   1078.000000
Valores criticos (1%)     -3.436431
Valores criticos (5%)     -2.864225
Valores criticos (10%)    -2.568200
dtype: float64
```

Según estos resultados podemos afirmar que el dataset según la demanda es estacionario por diversos motivos, el p-value nos indica que tenemos una probabilidad muy baja del

0,268% de que los resultados observados en la serie de tiempo se deban al azar lo que nos indica que podemos rechazar la hipótesis nula. Luego el estadístico de prueba es menor que los valores críticos para los niveles de significación de 1% y 5% y no difiere mucho del 10% por lo que volvemos a rechazar la hipótesis nula.

Para la producción obtenemos los siguientes resultados:



```
Resultados del Dickey-Fuller Test:
Estadístico de prueba      -2.022303
p-value                    0.276880
#Lags Used                 4.000000
Numero de observaciones   1090.000000
Valores criticos (1%)     -3.436364
Valores criticos (5%)     -2.864195
Valores criticos (10%)    -2.568184
dtype: float64
```

Con la producción vemos que ocurre todo lo contrario, el p-value es mayor que el valor de significancia del 0.05 y además nuestro estadístico de prueba es mayor que todos los valores críticos. Por lo cual podemos afirmar que la serie de tiempo no es estacionaria.

5. ARIMA

Una vez hemos confirmado que la demanda es estacionaria podemos aplicarle un modelo regresivo, en este caso hemos aplicado un modelo autorregresivo integrado de media móvil o ARIMA el cual usa la regresión y las variaciones para encontrar patrones y poder predecir. Para ello la serie de tiempo usada tiene que ser estacionaria.

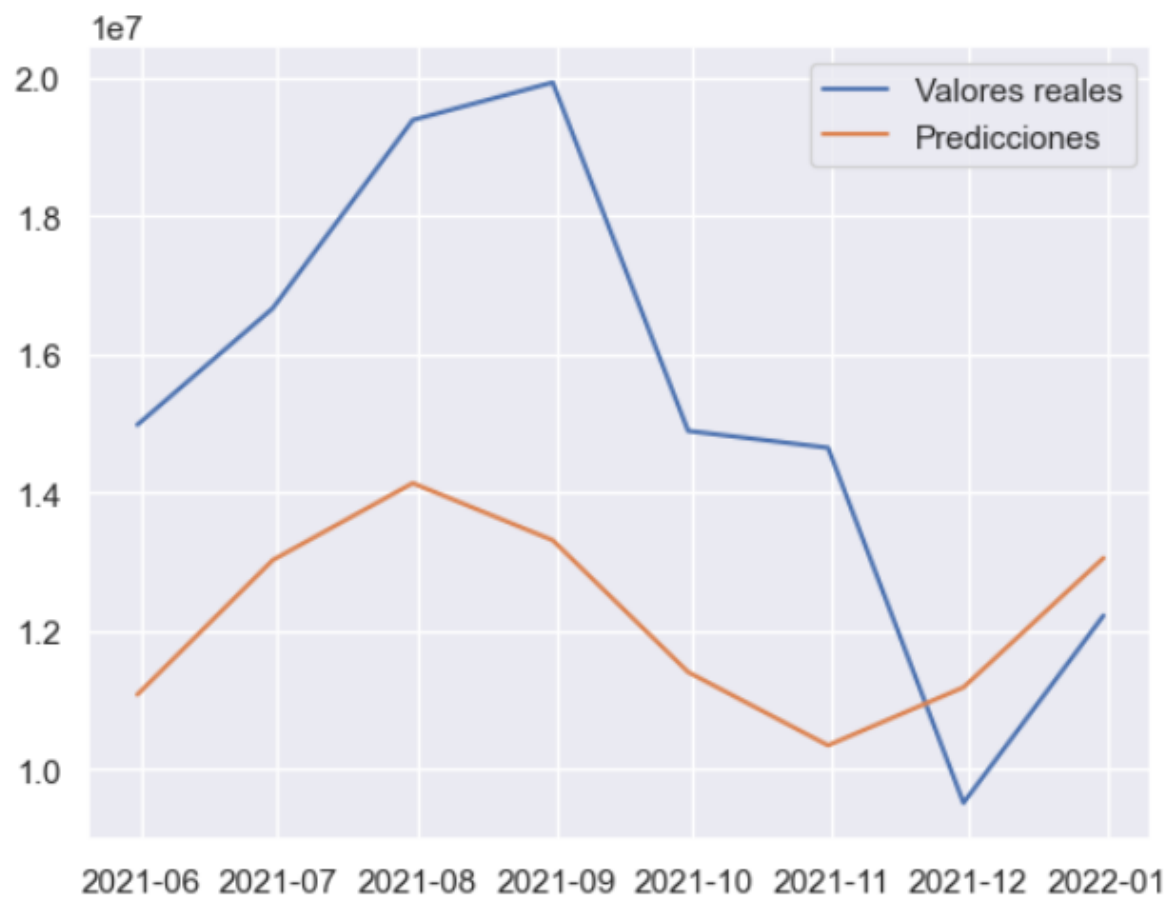
Después de hacer pruebas con diferentes valores, el mejor valor de AIC son para los valores $p = 2$, $d = 1$, $q = 2$

```
=====
Dep. Variable:          Demanda      No. Observations:          28
Model:                ARIMA(2, 1, 2)  Log Likelihood           -431.166
Date:                 Sun, 05 Feb 2023  AIC                        872.333
Time:                 13:19:59         BIC                        878.812
Sample:              01-31-2019       HQIC                       874.260
                  - 04-30-2021
Covariance Type:          opg
=====
```

Una vez elegido los parámetros de entrada pasamos a entrenar el modelo y hace predicciones para ver su acierto:

```
2021-05-31    1.108311e+07
2021-06-30    1.302891e+07
2021-07-31    1.414135e+07
2021-08-31    1.331579e+07
2021-09-30    1.140824e+07
2021-10-31    1.034849e+07
2021-11-30    1.118810e+07
2021-12-31    1.305750e+07
```

Visualizamos estas predicciones frente a los reales para ver la precisión del modelo:



Obteniendo una precisión de

```
Precision modelo ARIMA es: 76.83425210441922  
El error es: 23.16574789558078  
El error cuadrático mse: 16757269390052.26 kw
```

6. RNN

Las RNN son un tipo de red neuronal que se caracteriza por procesar series temporales y de esta manera obtener una “memoria”, de esta manera la tener los datos por días y meses del año podemos hacer que nuestra red recuerde que hay meses que se consume más o menos energía.

Esta red ha sido entrenada respecto a la producción del país, ya que la demanda está condicionada por la producción debido a la manera que están planteados los sistemas de abastecimiento eléctricos, de ahí que el país deba importar energía para suplir a la demanda. En resumen dependiendo de la cantidad de energía que se genere se podrá acotar el rango de la demanda.

Dentro de estas redes existe un tipo llamada LSTM o memoria a largo-corto plazo, esta red decidirá si cierta información es recordada u olvidada, así irá apoyándose en esta información.

Para indagar un poco más sobre las redes, se ha decidido entrenar una por días y otra por meses.

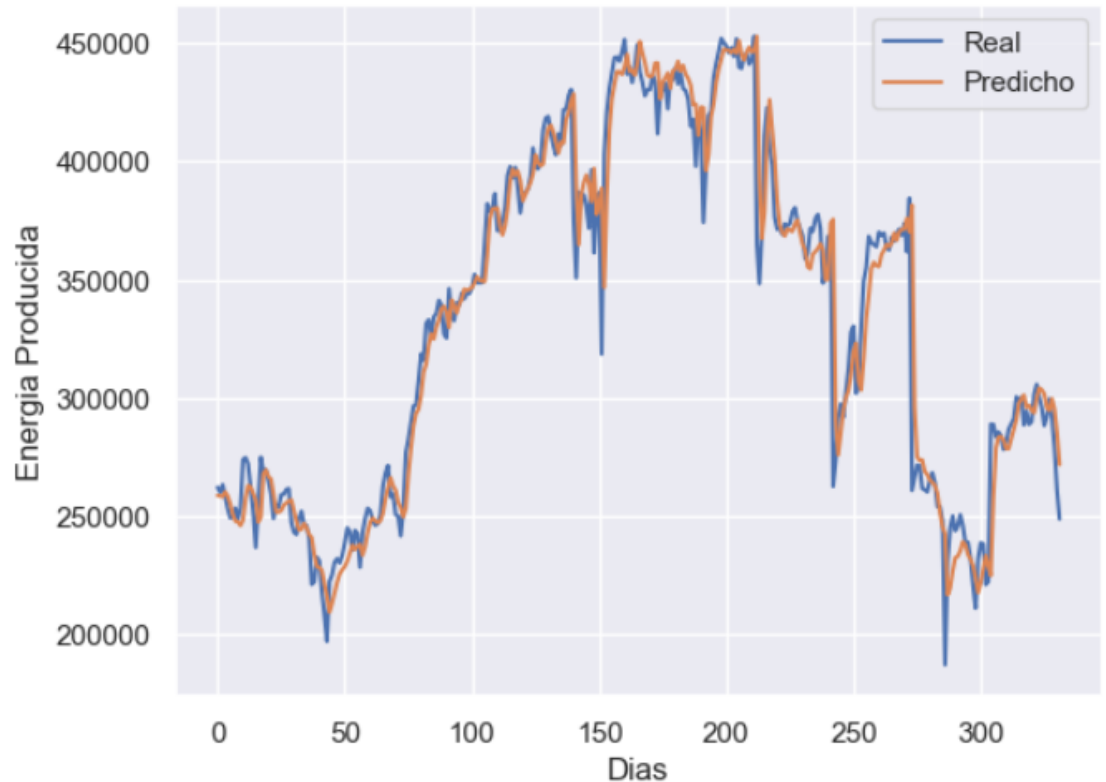
a. RNN por días

Lo primero que hemos realizado ha sido escalar los datos entre 0 y 1 para tratarlos mejor, posteriormente hemos preparado diferentes parámetros para la red y hemos dividido el dataset en entrenamiento y prueba.

Para esta red hemos usado dos capas LSTM con 50 unidades, y una Dense con una única salida, luego hemos optimizado con Adam el cual combina las bondades y para el error hemos usado el mse mean.

La precisión obtenido es la siguiente, también pasamos a visualizarla:

```
Precision de la RNN por días es: 96.67310225382668  
El error es: 3.3268977461733114  
El error cuadrático mse: 292739113.45247567 kw
```



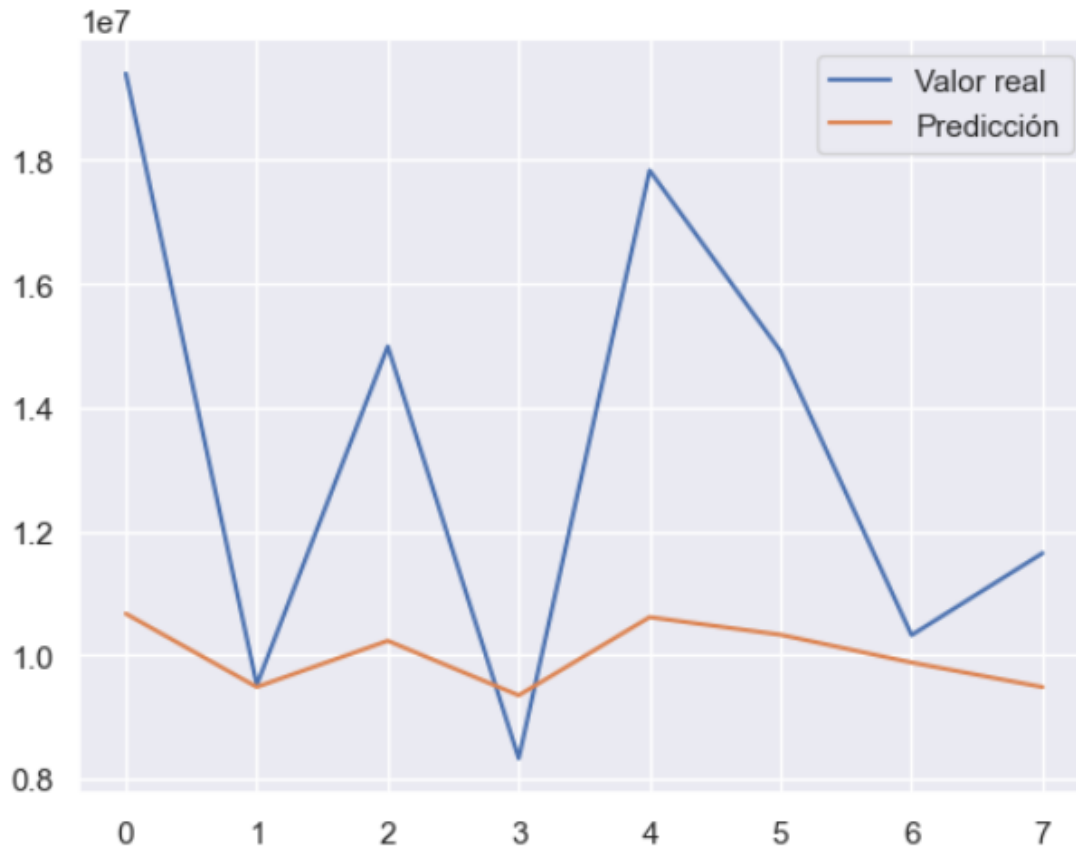
b. RNN por meses

Para el entrenamiento de por meses hemos realizado los mismos pasos que para la RNN por días pero cambiando algunos valores para ver el comportamiento.

En cambio para esta red solo se ha montado una capa LSTM con 30 unidades, para ver el comportamiento de estas redes con menos datos. También consta con una Dense con una única salida, error mse mean y optimizador adam.

Los valores arrojados son los siguientes:

```
Precision de la RNN por meses es: 77.0762748922426  
El error es: 22.923725107757402  
El error cuadrático mse: 22215323424061.492 kw
```



Vemos que por meses tenemos menos precisión y esto es debido a que cuenta con menos datos, y además la red tiene menos iteraciones y capas que por días.

7. Conclusiones

La análisis que hemos hecho en la dataset de demanda y producción de electricidad en Iraq en los años 2019, 2020, y 2021, además de usar otros datos externos como la temperaturas regionales, periodos críticos como Covid-19 y otros Ramadán además de los festivos nacionales, nos a permitido de conseguir unas conclusión visualizadas que justifica que la producción en de electricidad en Iraq es muy lejo de satisfacer la demanda. Hemos notado que en temporadas como verano o invierno se sube la demanda pero en el último se baja la producción complemente a los otras temporadas y eso es un punto que necesita revisar porque en la temporada en invierno presenta la segunda media de demanda, también el mes culto de Ramada se aumenta la demanda de electricidad por razón de ayuno y sed que exige el uso de aires acondicionados, y electrodomésticos, en otros días de festivo como el 14 de Julio que es el día de la revolución en Irak aumenta la demanda a un nivel importante por razón de los actividades de celebración. En otro punto importante hemos hecho una análisis que visualiza el impacto de los primeros meses de COVID en iraq que empezo el día 24 de Febrero 2020 hemos notado que en estos dos meses bajaba la demanda junto

con la producción complemente a la época pre-covid en iraq eso por razón del cierre inmediato de muchas industrias.

Después de todo el proceso hemos conseguido grandes hitos referentes tanto a las series temporales como al problema planteado. Respecto a las series hemos obtenido nociones super importantes de cómo plantear un problema de esta magnitud y además saber aplicar correctamente los modelos de Deep Learning correctos. Las redes neuronales han sido un campo nuevo en nuestro conocimiento dando una visión de cómo tratar datos en el tiempo correctamente. En general en la parte que corresponde a machine learning estamos muy contentos con los conocimientos adquiridos.

PD: No se ha extendido más las aplicaciones de ciertos modelos de cara a no alargar el documento porque para mi TFM pienso explicar más en profundidad estos temas

REFERENCIAS

Github con el código → <https://github.com/Josdelser/MLE-PROSUMER>

Series temporales → <https://www.tibco.com/es/reference-center/what-is-time-series-analysis>
<https://quantspace.es/2020/08/01/analisis-de-series-temporales-con-python-parte-2/>
http://repositorio.udec.cl/jspui/bitstream/11594/2011/3/Tesis_Proyeccion_de_series_de_tiempo_para_el_consumo_de_la_energia.Image.Marked.pdf
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1794-12372021000100077

Dickey fuller → <https://quantspace.es/2020/08/01/analisis-de-series-temporales-con-python-parte-2/>
<https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/methods-and-formulas/methods-and-formulas/>

ARIMA → <https://www.aitoralbertobaez.com/modelo-arima-pdq>
<https://machinelearningmastery.com/time-series-data-stationary-python/>

RNN → <https://abdatum.com/tecnologia/redes-neuronales-recurrentes>
<https://www.ibm.com/es-es/topics/recurrent-neural-networks>
<https://www.aprendemachinelearning.com/una-sencilla-red-neuronal-en-python-con-keras-y-tensorflow/>

Día Festivos nacionales en Iraq:
https://ar.wikipedia.org/wiki/%D9%82%D8%A7%D8%A6%D9%85%D8%A9_%D8%A7%D9

[%84%D8%B9%D8%B7%D9%84_%D8%A7%D9%84%D8%B1%D8%B3%D9%85%D9%8A%D8%A9_%D9%81%D9%8A_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82](#)

Importacion y produccion local de electricidad en Iraq:
<https://www.worldometers.info/electricity/iraq-electricity>