



Edgard Leonardo Castañeda García
Johan José Castañeda Vega
John Alexander Cárdenas

Problema 04

Precio de la energía

Índice

1. Análisis del proyecto	2
2. Metodología	3
3. Datos disponibles	3
4. Análisis de los datos	4
4.1. Descripción de la información	4
4.2. Análisis descriptivo mediante Power Bi	10
4.2.1. Promedio de precios por año	11
4.2.2. Promedio de precios por cuarto de año	11
4.2.3. Promedio de precios por mes	12
4.2.4. Promedio de precios por día	13
4.3. Análisis del precio mediante regresiones	13
4.3.1. Regresión lineal múltiple	13
4.3.2. Regresión lineal	14
4.4. Régimen de Agua mayor a \$400	16
4.5. Análisis del precio mediante Deep Learning con el método LSTM	17
4.5.1. Variables utilizadas	17
4.5.2. Topología de la red LSTM	17
4.5.3. Resultados del entrenamiento	18
4.5.4. Análisis de precios usando red LSTM	21
5. Conclusiones	22



1. Análisis del proyecto

En el mercado de energía mayorista se negocia la compraventa de energía en donde cada día los agentes generadores deben hacer una oferta, que consiste en declarar cuanta energía puede vender a cada hora del día siguiente, y a qué precio. Dependiendo de la forma de obtener la electricidad o de la fuente primaria como agua, carbón, gas o fuel oil, influye en los precios de oferta.

El operador del mercado decide cuanta energía debe comprarse a cada agente en cada franja horaria, en donde el precio más elevado de compra determina el precio de bolsa para esa franja. La Unidad de Planeación Minero Energética del Ministerio de Minas evalúa permanentemente escenarios energéticos factibles a futuro. Uno de los aspectos que debe considerarse es el efecto de los precios de los recursos energéticos en el precio de la energía eléctrica.

A continuación, se expone el análisis del efecto de los recursos energéticos en el precio de bolsa diarios de la energía eléctrica, para esto se utilizan los datos del precio ponderado de oferta para cada recurso energético día a día desde enero del 2015 hasta marzo del 2022 publicados por XM.

Para tener una vista más amplia del problema desde la perspectiva de la energía y la informática se muestra el análisis de las siguientes dimensiones

1. *Dominio*

El problema que se trata es la explicación de los precios de bolsa diarios a partir de los precios diarios ponderados de los recursos de generación, esto a partir de los datos suministrados por XM. Este dilema se ubica en la parte del mercado mayorista en la generación de la cadena de valor con perspectiva en los agentes y los recursos ya que visualizar el precio ponderado de oferta para cada recurso energético día a día, mostrara una relación entre el precio de la energía ofertaos por cada agente y el recurso que utiliza este.

2. *Impacto efectivo*

El análisis del precio de la energía en la búsqueda de la explicación y relación con los precios diarios ponderados de los recursos de generación ayudara a la toma de decisiones al operador de mercado en la compra de energía para cada franja horaria, también ver el comportamiento del precio de la energía con los recursos en ese tiempo determinado mostrara la tendencia del uso de recursos más baratos y que ocasionen un menor impacto ambiental.

3. *Ciclo de vida*

El mercado mayorista de energía eléctrica en Colombia se encuentra en la etapa de operación, como objetivo principal garantizar la eficiencia, transparencia y competencia en la comercialización de energía eléctrica.

4. *Flujo de datos*

Los datos son provenientes de XM la cual es una entidad encargada de la operación y administración del SIN y del Mercado de Energía Mayorista (MEM), en donde se encuentran los precios



diarios de Agua, Carbón, Gas, Líquidos, y los precios de bolsa máximo, mínimo, ponderado y el precio promedio mensual de bolsa desde enero del 2015 hasta marzo del 2022. Se observa que el dataframe está ordenado por fecha, donde no hay valores faltantes o en cero.

5. *Categoría del problema de datos*

Este problema de datos tiene el objetivo de encontrar una explicación de los precios de bolsa diarios, a partir de los precios diarios ponderados de los recursos de generación, lo que encasilla el dilema en la categoría de análisis de asociaciones para comprender y describir mejor las interdependencias de todos los atributos para buscar las relaciones entre estos.

6. *Método de solución*

El método de solución que se emplea para resolver este problema de precio de la energía es el descubrimiento de explicaciones ya que se tiene interés en ver la relación del precio de la energía con el precio del recurso.

2. Metodología

3. Datos disponibles

Para el desarrollo de este proyecto se dispone de un archivo con los siguientes datos:

- Precio diario del agua.
- Precio diario del carbón.
- Precio diario del gas.
- Precio diario de los líquidos.
- Precio de bolsa máximo diario.
- Precio de bolsa ponderado diario.
- Precio de bolsa mínimo diario.
- Precio promedio mensual de bolsa.

El archivo toma estos valores en un lapso comprendido entre enero de 2015 y marzo de 2022.



4. Análisis de los datos

4.1. Descripción de la información

En primera instancia se realiza un análisis preliminar graficando los precios diarios ponderados de los recursos de generación desde enero del 2015 a marzo del 2022 sin tener en cuenta los precios de bolsa diarios (mínimo, máximo y ponderado).

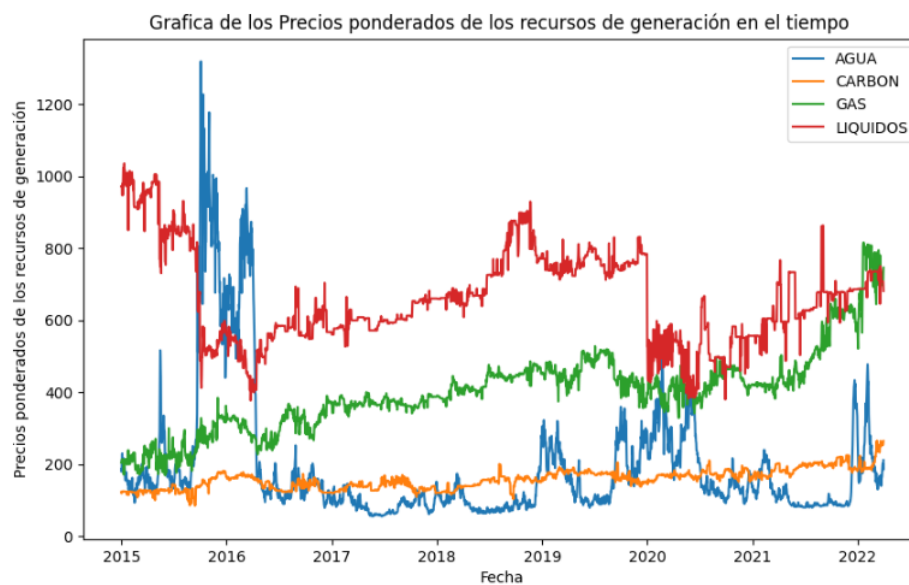


Figura 1: Precios ponderados de los recursos de generación Enero 2015 – Marzo 2022

En la figura 1 se observa que el recurso de generación con precio más bajo es el agua, el segundo el carbón, el tercero el gas y el más costoso los líquidos de petróleo. El valor del carbón para producir energía eléctrica es muy constante, en cambio el precio del agua tuvo un aumento entre la mitad del 2015 hasta la mitad del 2016 esto puede ser debido a en Colombia se experimentó en esa época un fuerte fenómeno del niño lo que redujo el nivel de agua en los embalses y disminuyó en la disponibilidad de agua para el almacenamiento de energía. También se ve que el precio del gas en Colombia para generación eléctrica aumenta cada año y valor de los líquidos es muy variable pero siempre el más costoso.

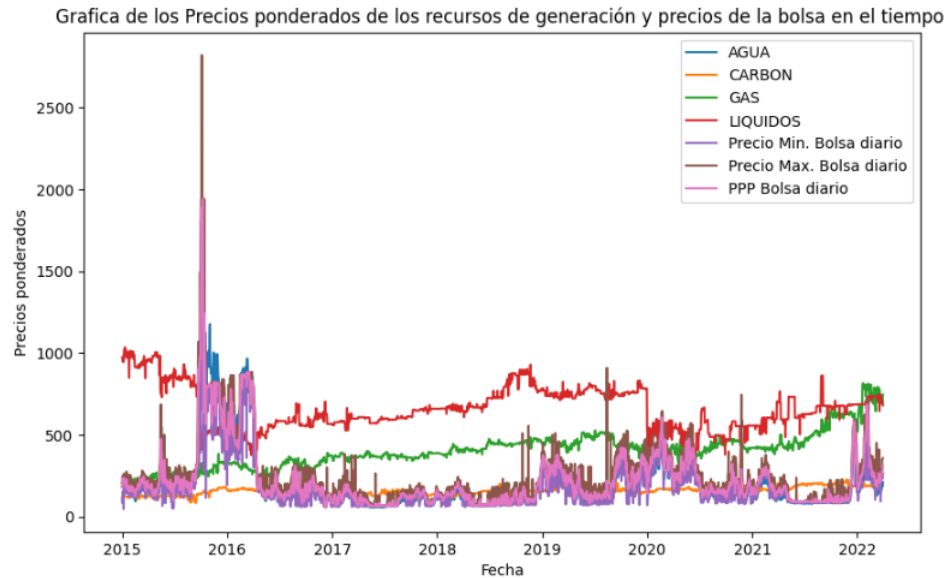


Figura 2: Precios ponderados de los recursos de generación y precios de bolsa Enero 2015 – Marzo 2022

También graficando los precios de bolsa diarios (Mínimo, máximo y ponderado) con los precios diarios ponderados de los recursos de generación en la figura 2 se observa que el precio de bolsa tiende a tener el mismo comportamiento del precio diario del agua. Para ver mejor este comportamiento solo se graficará el precio del agua y los precios de bolsa en la figura 3. Y el precio del agua y el precio ponderado de bolsa en la figura 4.

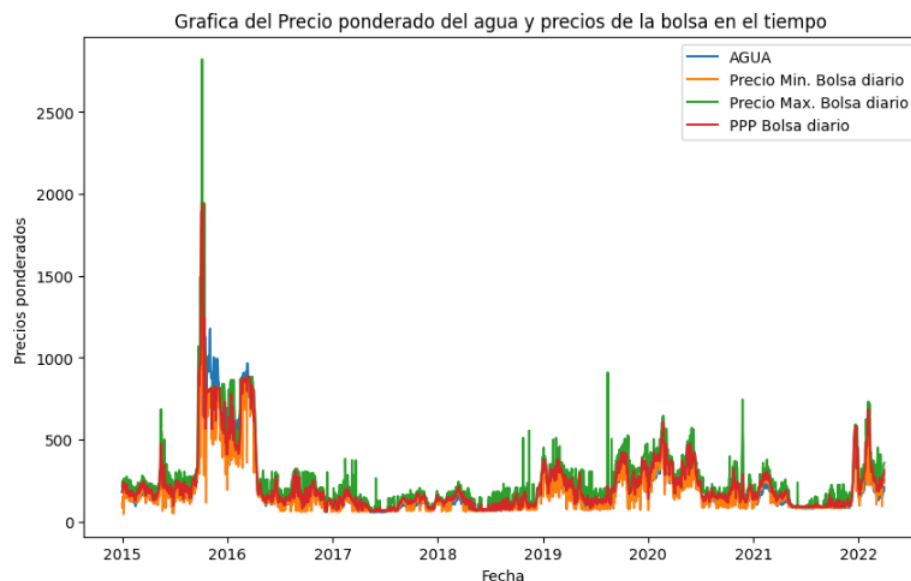


Figura 3: Precios ponderados del agua y precios de bolsa Enero 2015 – Marzo 2022

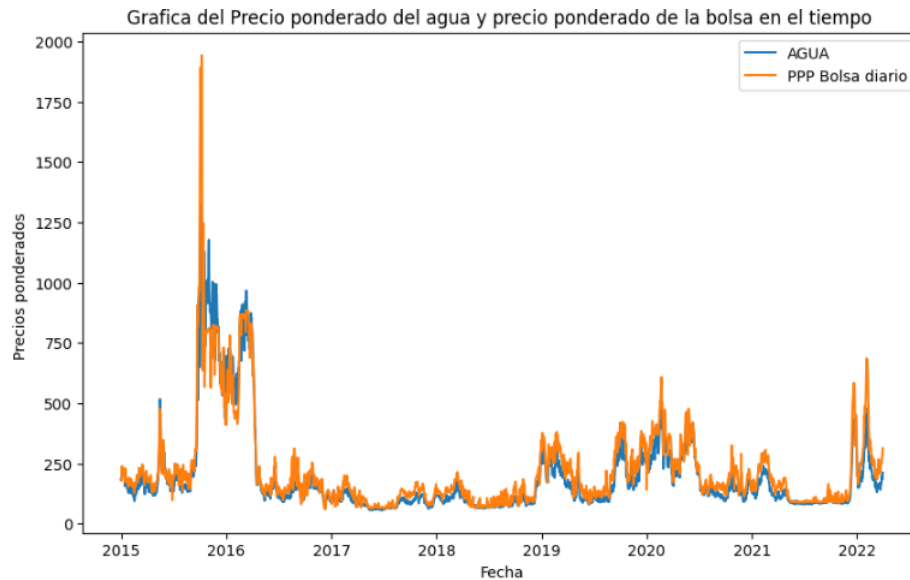


Figura 4: Precios ponderados del agua y el precio ponderado de bolsa Enero 2015 – Marzo 2022

Según las dos gráficas anteriores se observa que efectivamente el precio ponderado del agua y los precios ponderados de la bolsa tienden a ser muy parecidos, y pareciera que el recurso que influye más en el precio de bolsa es el agua, esto a que la mayor parte de la generación eléctrica en Colombia es de las hidroeléctricas. Para comprobar esto se realiza correlaciones entre el precio de cada recurso y precio de bolsa.

A continuación, se realizaron gráficas para determinar la correlación entre el precio de cada recurso de generación y el precio diario ponderado de bolsa. Estos datos se presentan en la figura 5, donde se representan los precios diarios de bolsa en función de cada uno de los recursos de generación:

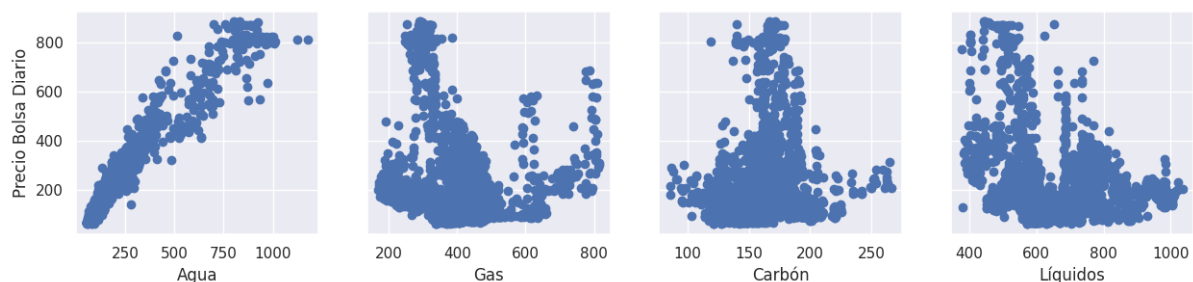


Figura 5: Comparación del precio ponderado de bolsa diario y el precio de los recursos de generación.

Se aprecia una marcada correlación entre el precio del agua y el precio diario ponderado de energía, ya que ambos aumentan o disminuyen en conjunto. Sin embargo, no se observa una correlación significativa



entre los demás recursos de generación y el precio de bolsa. Los mismo ocurre con el precio mínimo de bolsa diario como se observa en la figura 6

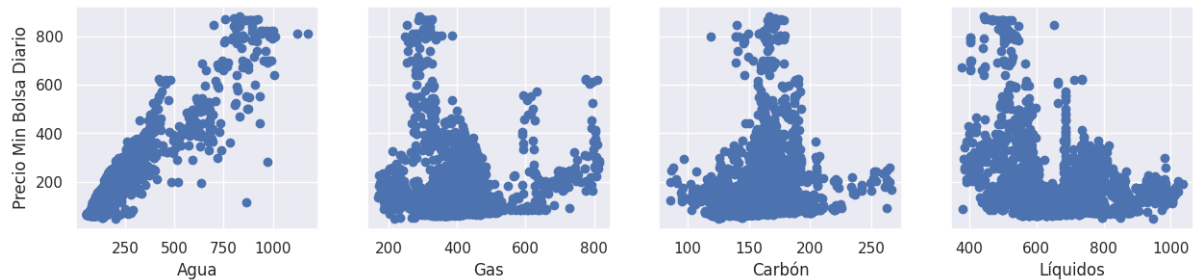


Figura 6: Comparación del precio mínimo de bolsa y el precio de los recursos de generación.

Este patrón se mantiene incluso al considerar el precio máximo de la energía, como se aprecia en la figura 7:

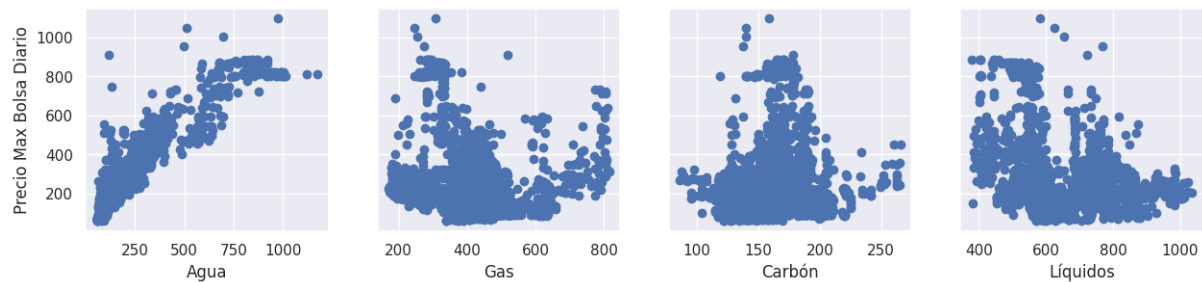


Figura 7: Comparación del precio máximo de bolsa y el precio de los recursos de generación.

Al analizar la relación entre el precio mensual de bolsa y el precio de los recursos de generación, se confirma la persistencia de una correlación significativa entre el precio del agua y el precio mensual de bolsa:

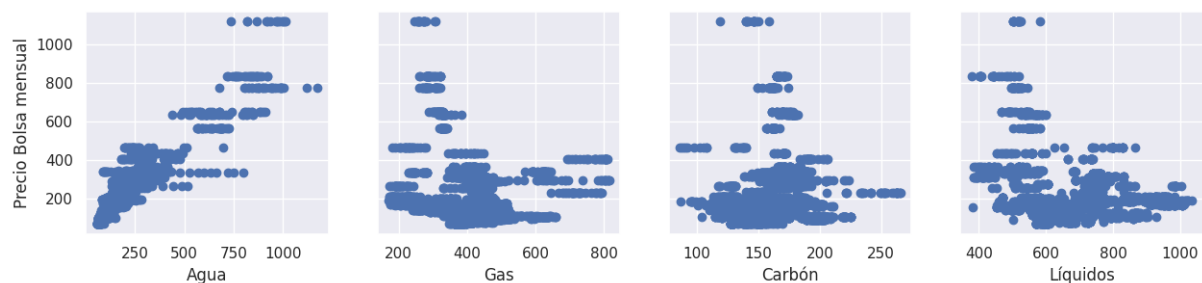


Figura 8: Comparación del precio mensual de bolsa y el precio de los recursos de generación.



Para confirmar de manera concluyente la fuerte correlación entre el precio del agua y los precios diarios de energía mínimo, ponderado y máximo, también se ha llevado a cabo un análisis de correlación mediante una matriz de correlación de los datos disponibles:

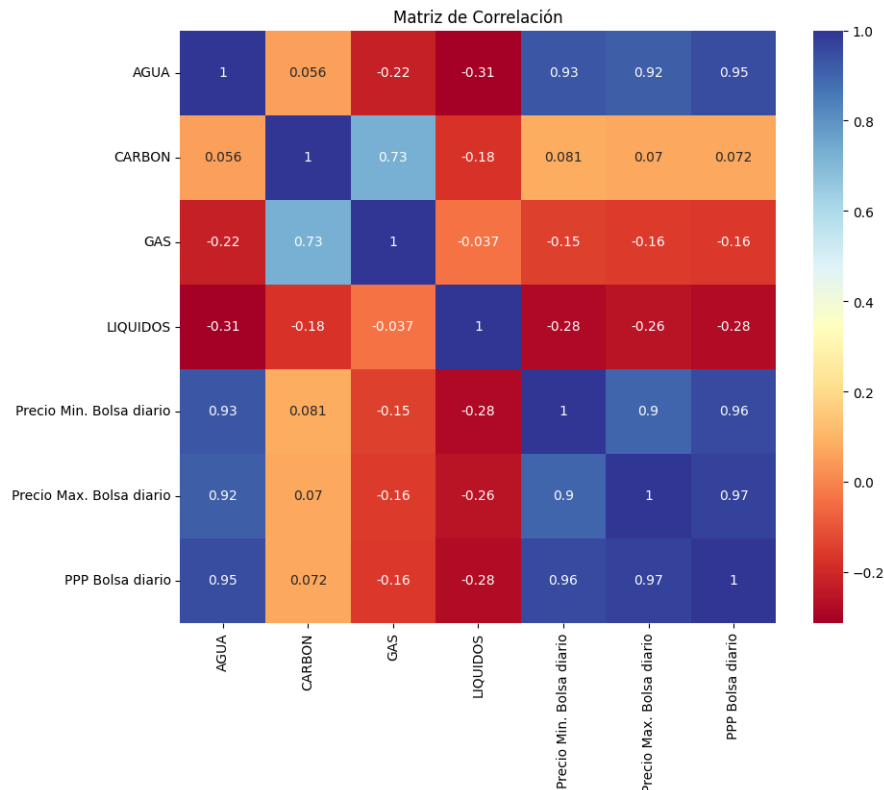


Figura 9: Matriz de correlación de los datos.

Se observa una correlación muy fuerte entre el precio del agua y el precio de bolsa de la energía (un valor de correlación cercano a 1), lo cual es esperado considerando que más del 60 % de la generación eléctrica en Colombia se realiza a partir de recursos hídricos [1]. No se evidencia una correlación significativa entre los demás recursos de generación y el precio de bolsa. Esto refuerza la influencia dominante que tiene el precio del agua en la determinación del precio de bolsa de la energía, dada la considerable participación de la generación hidroeléctrica en la matriz energética del país.

Posteriormente, se realizó el mismo ejercicio para hallar posibles correlaciones entre el precio de los recursos de generación disponibles. A continuación, se muestra la comparación entre los precios de los recursos disponibles y el precio de los líquidos:

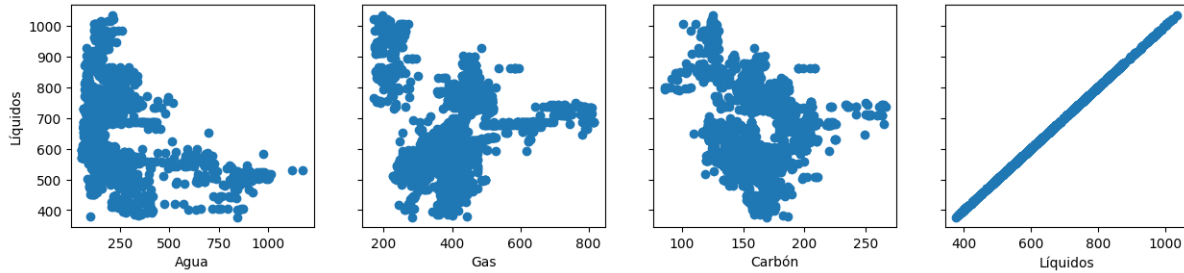


Figura 10: Comparación del precio de los líquidos y el precio de los demás recursos.

En la figura 10 se puede apreciar que no se evidencia una correlación clara entre el precio de los líquidos y el precio de los demás recursos de generación.

En el caso del precio del carbón se obtuvieron las siguientes gráficas:

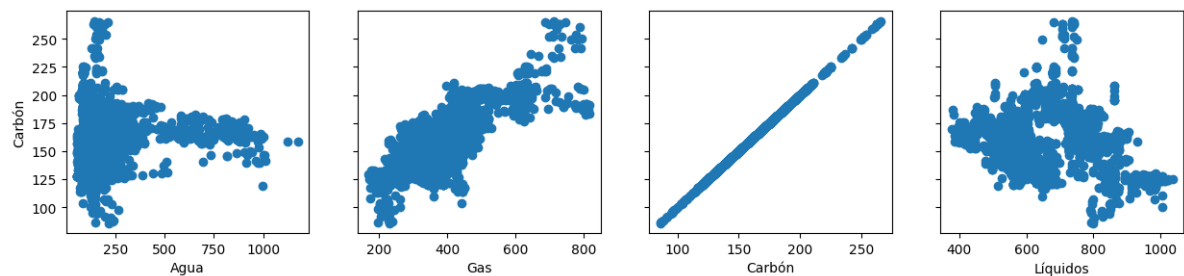


Figura 11: Comparación del precio del carbón y el precio de los demás recursos de generación.

Se aprecia que no existe una correlación significativa entre el precio del carbón y el precio de los líquidos y el agua. No obstante, es relevante destacar una ligera correlación entre el precio del gas y el precio del carbón. Esto también se evidencia en el valor de correlación entre el gas y el carbón en la figura 9 (0,73) valor relativamente alto en comparación con los valores de correlación que se presentan entre los demás recursos de generación. Una posible explicación para este fenómeno radica en que durante las épocas de sequías en Colombia, las centrales térmicas (que operan principalmente con carbón y gas) incrementan su generación [2], lo que provoca cambios en los precios de estos recursos.

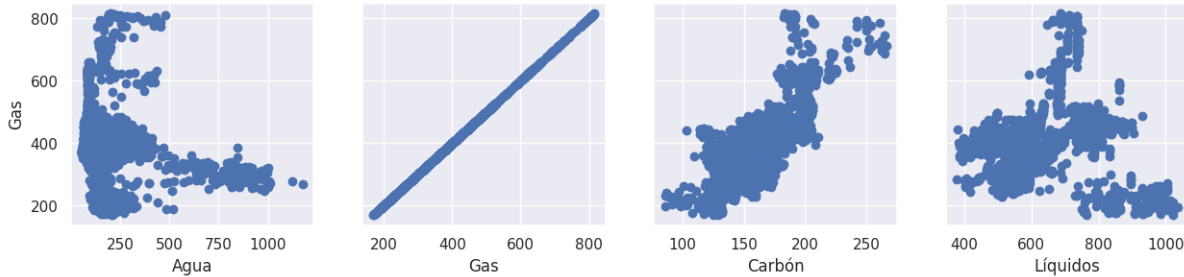


Figura 12: Comparación del precio del gas y el precio de los demás recursos de generación.

En la figura 12 se reafirma la presencia de una ligera correlación entre el precio del gas y el precio del carbón. Sin embargo, se confirma una vez más que el precio del gas no presenta una correlación significativa con los demás recursos de generación.

En el caso del precio del agua y su correlación con el precio de los demás recursos se obtienen las siguientes gráficas:

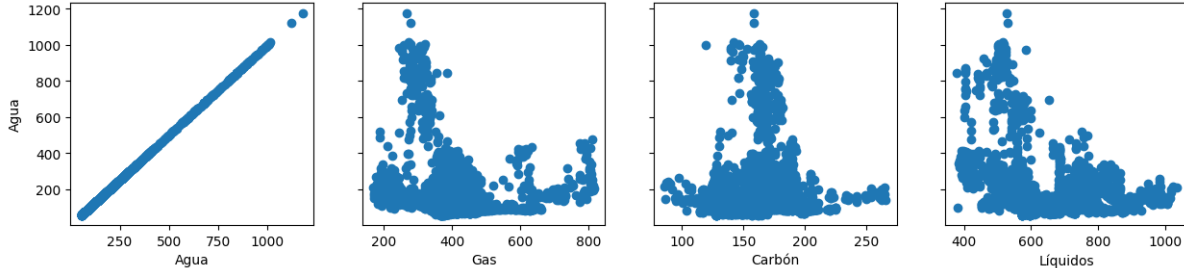


Figura 13: Comparación del precio del agua y el precio de los demás recursos de generación.

En la figura 13 tampoco se observa una correlación significativa entre el precio del agua y el precio de los demás recursos.

4.2. Análisis descriptivo mediante Power Bi

Usando las capacidades de Power Bi para la manipulación de los datos, se pueden obtener conclusiones relevantes para los posteriores análisis.



4.2.1. Promedio de precios por año

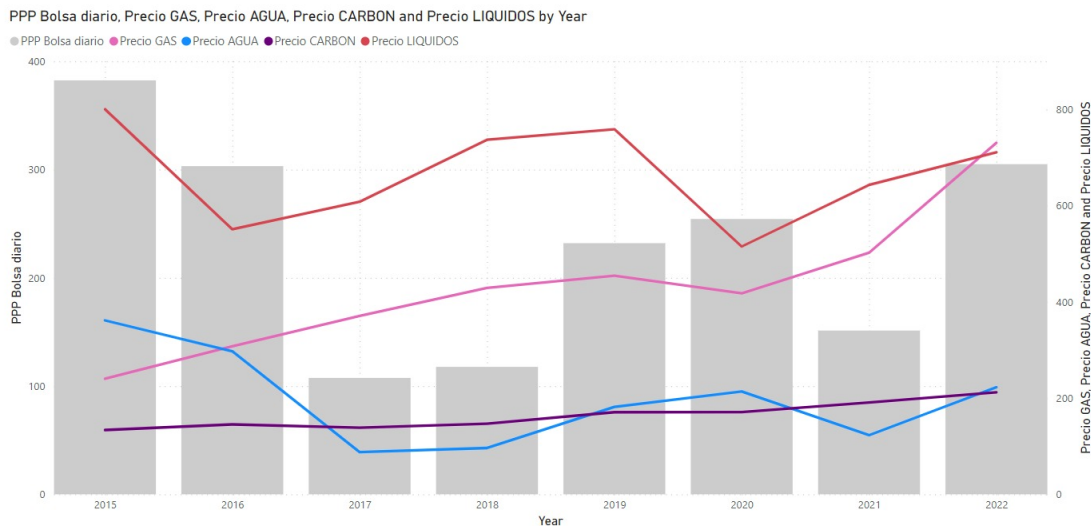


Figura 14: Promedio de los precios para cada año durante el periodo de 8 años.

Se observa de la figura 14 que, salvo por los líquidos, el precio de los combustibles fósiles ha ido al alza en los últimos 8 años.

Adicionalmente, se encuentra de nuevo que el precio promedio en bolsa sigue la tendencia del precio promedio del agua, aún en los promedios anuales. Esta tendencia se repite en todas las segregaciones por periodo de tiempo.

4.2.2. Promedio de precios por cuarto de año

Cuarto calendario:

- Qtr1: Enero, Febrero, Marzo.
- Qtr1: Abril, Mayo, Junio.
- Qtr1: Julio, Agosto, Septiembre.
- Qtr1: Octubre, Noviembre, Diciembre.

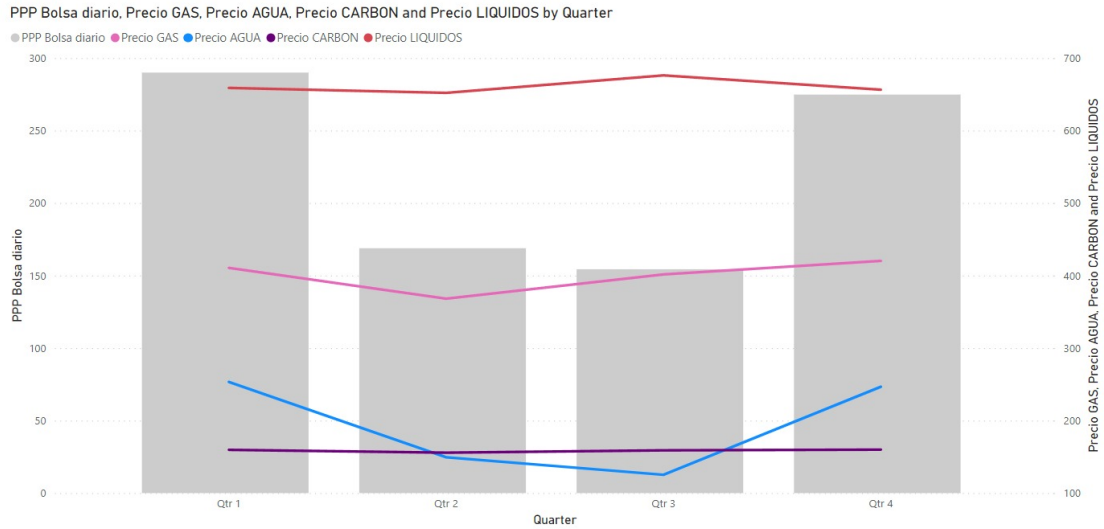


Figura 15: Promedio de los precios para cada cuarto de año durante el periodo de 8 años.

En la figura 15 se evidencia la estacionalidad en los precios del agua y por ende de la bolsa. Siendo los periodos de menor precio promedio las temporadas de invierno y las jornadas de mayor precio aquellas en las cuales sobreviene el verano y el fenómeno del niño.

4.2.3. Promedio de precios por mes

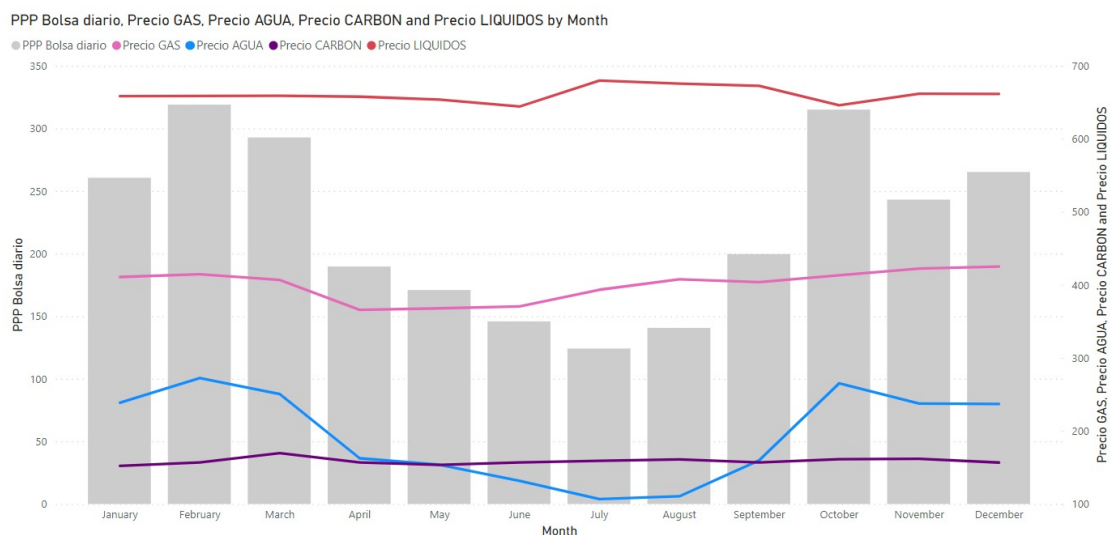


Figura 16: Promedio de los precios para cada mes durante el periodo de 8 años.



La figura 17, permite observar una mayor granularidad en los datos de precio respecto a la estacionalidad. Resulta interesante que los precios promedio de los combustibles fósiles han comenzado a estabilizarse, siendo casi constante a lo largo de todos los meses durante los 8 años de datos.

4.2.4. Promedio de precios por día

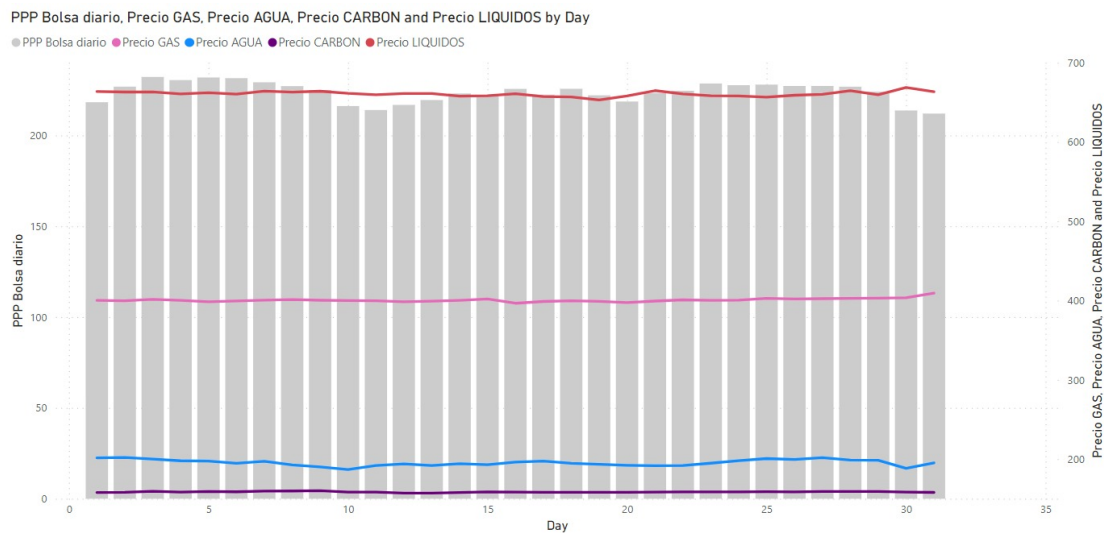


Figura 17: Promedio de los precios para cada día durante el periodo de 8 años.

En esta figura ocurre un fenómeno inesperado en el precio promedio de los combustibles debido a que estos prácticamente se mantienen constantes en el agregado para cada día durante el periodo de 8 años.

Esta tendencia a la estabilidad del promedio implica una autocorrelación entre el precio de bolsa de un día respecto al mismo día de los otros años, de manera que se estabiliza el promedio global.

Dicha autocorrelación, sumada a las conclusiones de la matriz 21, serán de utilidad para encontrar las variables que mejor representen el valor actual de los precios en un análisis de series de tiempo.

4.3. Análisis del precio mediante regresiones

Para los siguientes análisis se intentará explicar el Precio de bolsa promedio diario, pues es el dato que más variabilidad contiene y está altamente correlacionado con los otros índices de bolsa (Figura 9).

4.3.1. Regresión lineal múltiple

Con base en las correlaciones de la sección 4.1, es evidente que el precio de los combustibles Gas, Carbón y Líquidos, no explican la variabilidad en el precio diario de bolsa. Por lo cual un análisis de regresión múltiple es, a priori, inviable.



Adicionalmente, una de las hipótesis para poder utilizar un conjunto de variables en una regresión lineal múltiple es que no exista una alta colinealidad entre ellas. Este parámetro puede evaluarse mediante la clase `variance_inflation_factor` de la librería `statsmodels`.

vif	Variable
2.6	Agua
34.0	Gas
58.9	Carbón
13.8	Líquidos

Mediante esta librería se puede evaluar a multicolinealidad entre elementos, la cual se considera significativa por encima de 10. Esto implica que sólo es posible usar la variable Agua para hacer la regresión.

4.3.2. Regresión lineal

Para realizar la regresión lineal se busca la transformación de variables que maximice la correlación de estas.

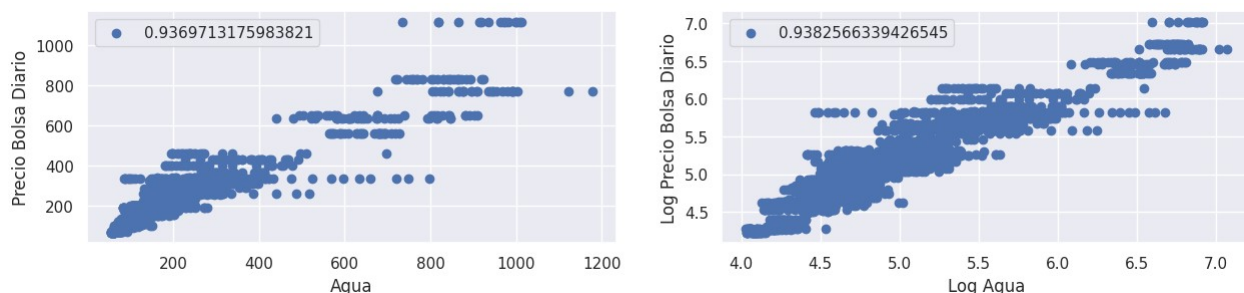


Figura 18: Comparación de la correlación para dos variables sujetas a transformaciones logarítmicas.

Puede observarse que la forma Log-Log maximiza la correlación, por lo cual se hará la regresión lineal con base en estas nuevas variables.

La regresión se hizo con la clase `LinearRegression` de la librería `sklearn.linear_model`.

Usando el método `score`, la regresión obtenida mediante estas variables explica el 93.534% de la variabilidad de los datos.

Puede observarse de la distribución de los residuos $y_{real} - \hat{y}$, que el error tiene una distribución aproximadamente normal al rededor de 0. Lo cual hace que la regresión sea satisfactoria y sea posible llegar a conclusiones usándola como referencia.

Se puede observar también de la figura 19, que los residuos tienen una cola hacia la parte negativa de la gráfica. Por lo que la regresión está sobreestimando el valor de una parte de los datos.

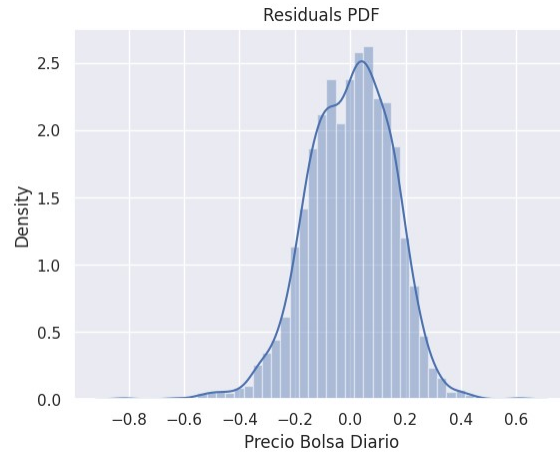


Figura 19: Distribución de los residuos $y_{real} - \hat{y}$

Para poder visualizar la regresión sobre los datos reales se convierte la ecuación lineal en su forma no lineal despejando la variable \hat{y} : "Precio de bolsa Diario Estimado".

$$\hat{y} = e^{0.931132 \cdot \ln(a) + 0.513626} \quad (1)$$

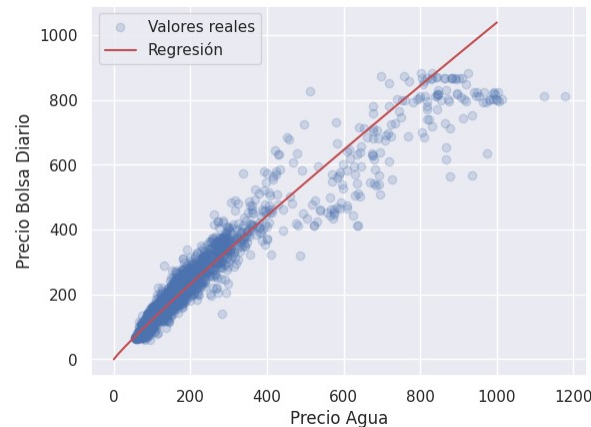


Figura 20: Comparación de la regresión 1 y los datos reales. La mayor concentración de valores se representa en un color más oscuro.

De la figura 20, se pueden observar dos regímenes:

1. Precio del Agua menor a \$400: En este régimen el precio de la bolsa y del agua están altamente correlacionados. La regresión representa correctamente la mayoría de los datos y se puede considerar que el Precio promedio de bolsa diario "sigue" al precio del agua.



2. Precio del Agua mayor a \$400: La regresión lineal ya no explica correctamente el precio de la bolsa, y sobrestima el valor de esta, como era esperable del análisis de la distribución de residuos. Puede observarse también una suerte de "techo", en el precio de la bolsa, el cual solo podría ser representado con una regresión no lineal.

4.4. Régimen de Agua mayor a \$400

De la evaluación del conjunto de datos para el régimen superior, se tiene la siguiente matriz de correlación:

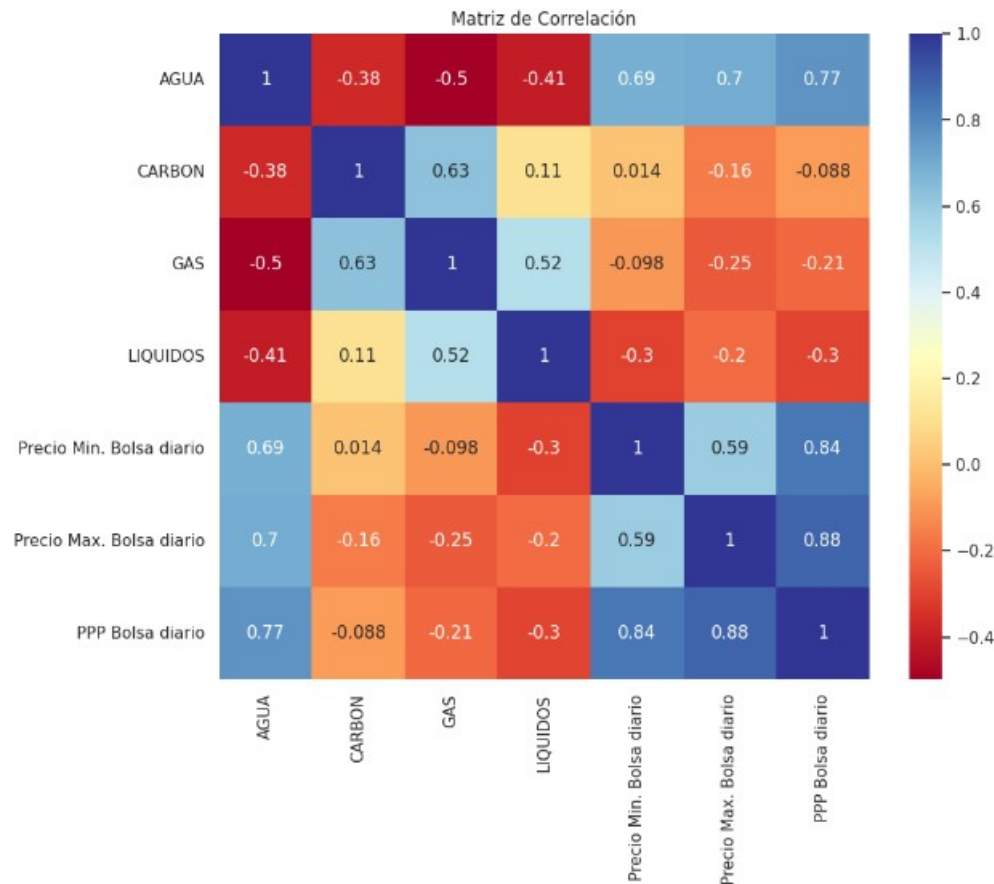


Figura 21: Matriz de correlación cuando el precio del agua es mayor a \$400.

Se observa un cambio severo en los valores de bolsa diario. Aunque el precio del agua sigue siendo el valor más influyente, ahora los combustibles fósiles tienen una importancia mayor.

Por otro lado los precios de bolsa diario máximo, y mínimo ya no tienen una correlación tan elevada con el precio diario promedio como podría esperarse. Aunque la correlación del valor mínimo con el valor diario es significativamente mayor que la correlación del valor máximo con el valor diario. Esta conclusión será estudiada más a fondo en el siguiente acápite.



4.5. Análisis del precio mediante Deep Learning con el método LSTM

El método LSTM, del inglés Long Short-Term Memory, es una topología recurrente de redes neuronales que permite evaluar la información de un instante t , relacionándola con una "memoria" de corto plazo (los eventos más recientes) y una "memoria" de largo plazo (los eventos más antiguos). Adicional a ello, puede integrar otras variables para predecir los valores futuros.

La popularidad del método LSTM se debe a la capacidad que tiene de predecir valores futuros con mucha precisión. Sin embargo, al ser una metodología de Deep Learning, tiene como desventaja que se comporta como una caja negra, por lo que tiene una muy baja interpretabilidad, tanto para sintonización como para evaluación de los resultados.

Para sintonizar el método LSTM, se probaron diferentes topologías y esquemas de entrenamiento. Resulta especialmente interesante que una vez encontrada una topología óptima o subóptima, los resultados del entrenamiento tienden a converger. En tanto que cuando la topología no es óptima, estos divergen sustancialmente de un entrenamiento a otro.

4.5.1. Variables utilizadas

De la correlación de Régimen 2 se observó que todos los combustibles tienen una influencia en el precio diario de bolsa, pero con una alta multicolinealidad. Por lo que una relación no lineal es esperable y serán incluidos como variables para el desarrollo del LSTM.

De la subsección 4.2.4, se observó que hay una autocorrelación entre los precios diarios en diferentes años. De la misma forma se observa en la matriz de correlación 21. De esta forma, se integrarán como variables de entrada el Precio diario de Bolsa, Precio máximo diario de Bolsa, y Precio mínimo diario de Bolsa.

La variable de salida será el Precio diario de Bolsa, pues es la variable de salida más representativa.

4.5.2. Topología de la red LSTM

Por su naturaleza altamente no lineal, la red neuronal profunda es caótica y muy sensible a cambios en alguno de los parámetros de la red. A continuación se describirá la topología:

La red neuronal contiene dos capas de 64 neuronas cada una, con activación ReLU, Dropout de 0.2 y una neurona de salida.

Se busca pronosticar el valor de un día (1) con base en los treinta (30) días anteriores a la fecha en cuestión.

Para el entrenamiento se usaron 15 epochs con lotes (batches) de 16 y un 10 % de los datos en cada epoch como validación.

El algoritmo de optimización será ADAM y la función de pérdidas será el Error Cuadrático Medio.

```
model = Sequential()  
model.add(LSTM(64, activation='relu',  
              input_shape=(trainX.shape[1], trainX.shape[2]),  
              return_sequences=True))
```



```
model.add(LSTM(64, activation='relu',  
              return_sequences=False))  
model.add(Dropout(0.2))  
model.add(Dense(trainY.shape[1]))  
  
model.compile(optimizer='adam', loss='mse')  
model.summary()
```

4.5.3. Resultados del entrenamiento

El modelo resultante tiene la siguiente estructura:

Model: test		
Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 30, 64)	18432
lstm_3 (LSTM)	(None, 64)	33024
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65
=====		
Total params: 51,521		
Trainable params: 51,521		
Non-trainable params: 0		

Tabla 1: Descripción del modelo LSTM.

El error en el entrenamiento es considerablemente bajo, por debajo de 0.09 para los conjuntos de validación. Para confirmar que el modelo es convergente, este entrenamiento se repitió 30 veces obteniendo valores similares en cada ocasión.

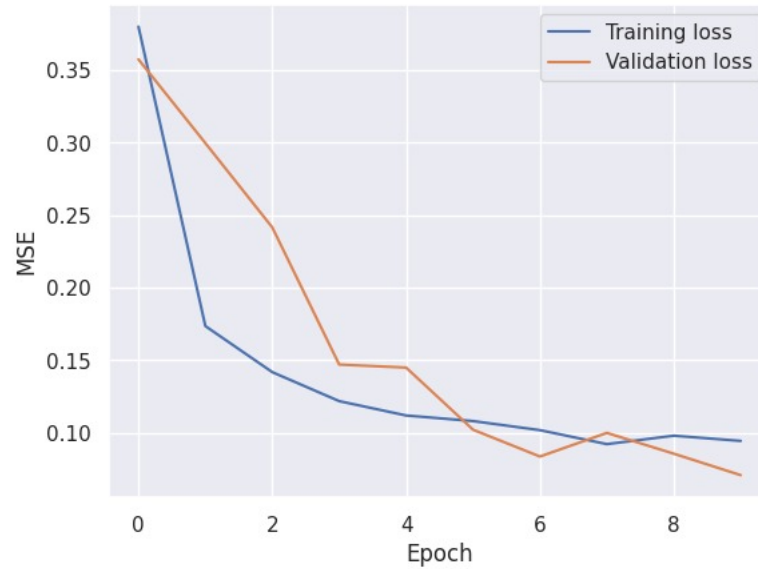


Figura 22: Gráfica de función de pérdidas típica para un entrenamiento.

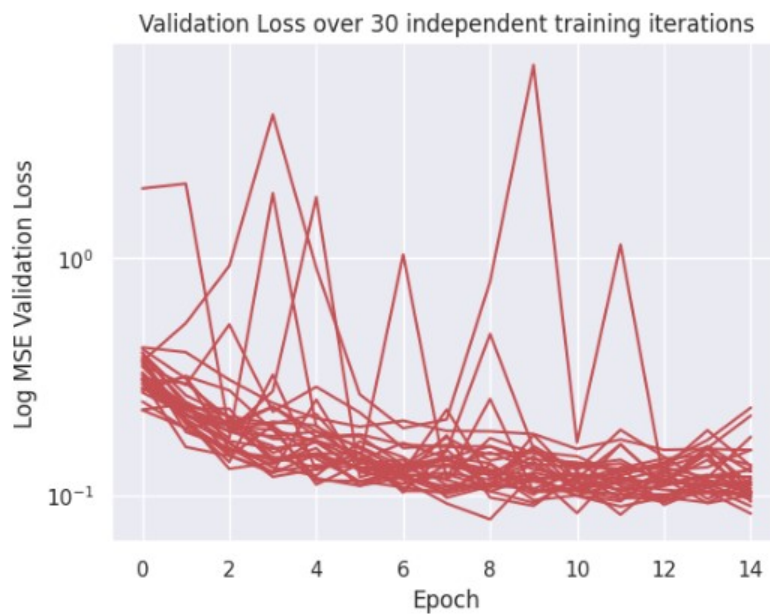


Figura 23: Gráfica de función de pérdidas en validación para 30 entrenamientos independientes.

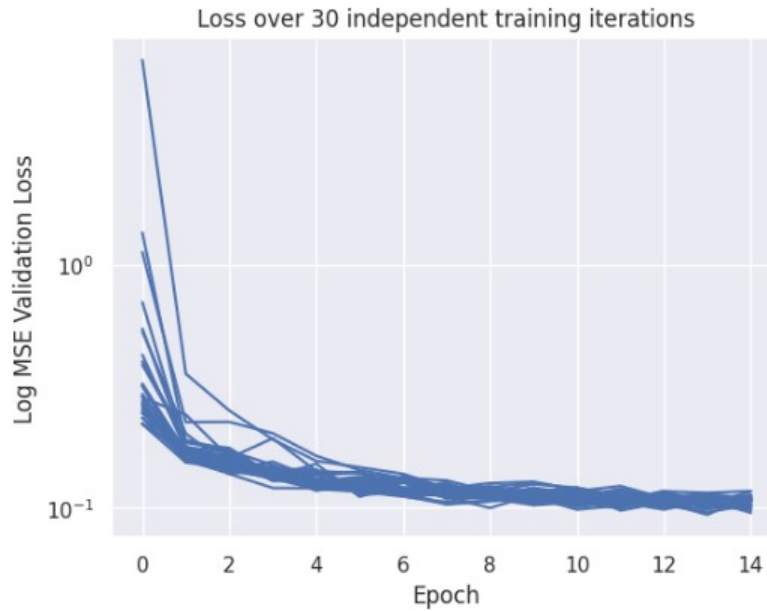


Figura 24: Gráfica de función de pérdidas para 30 entrenamientos independientes.

El desempeño de la red neuronal se puede observar en la siguiente gráfica, en la que se están pronosticando los últimos 500 días de datos (Línea naranja) y se alimenta a la red con los restantes.

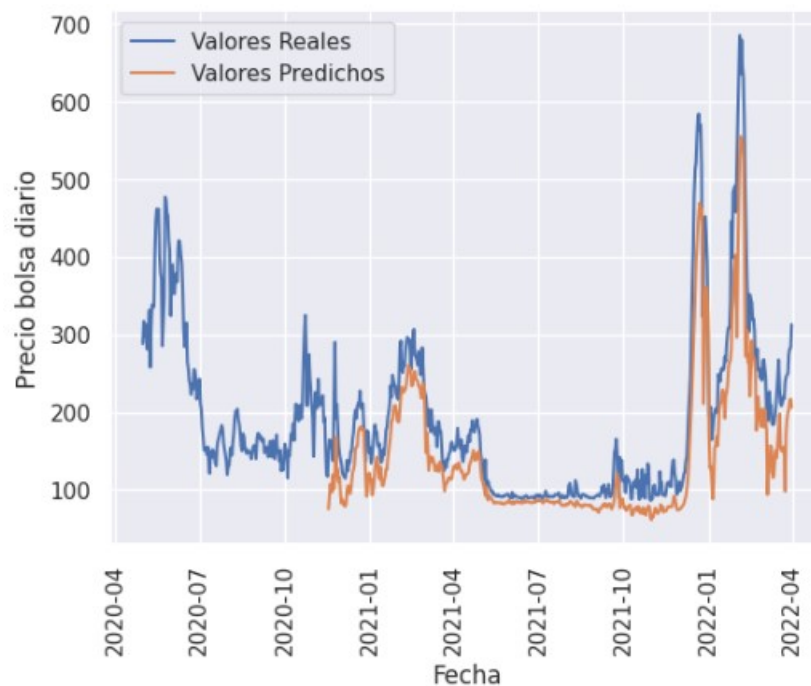


Figura 25: Comparación de los datos reales con los datos pronosticados por la red neuronal.



4.5.4. Análisis de precios usando red LSTM

Se observa que las variables implicadas en el modelamiento de la red explican apropiadamente el comportamiento de los precios diarios. Sin embargo, la predicción tiende a subestimar el valor estimado del valor real.

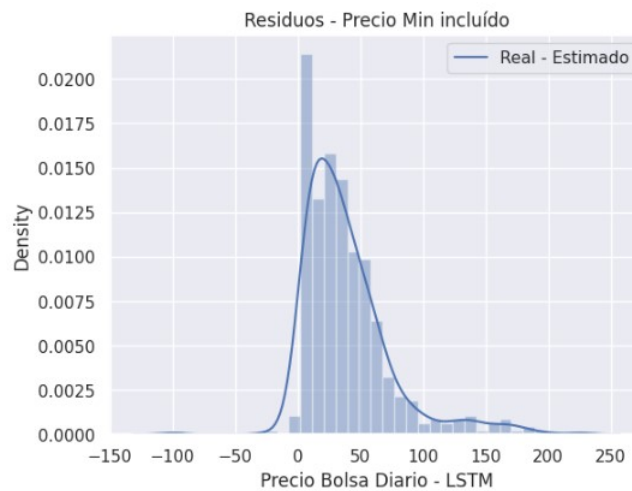


Figura 26: Distribución de los residuos del método LSTM con la variable Precio Mínimo incluída.

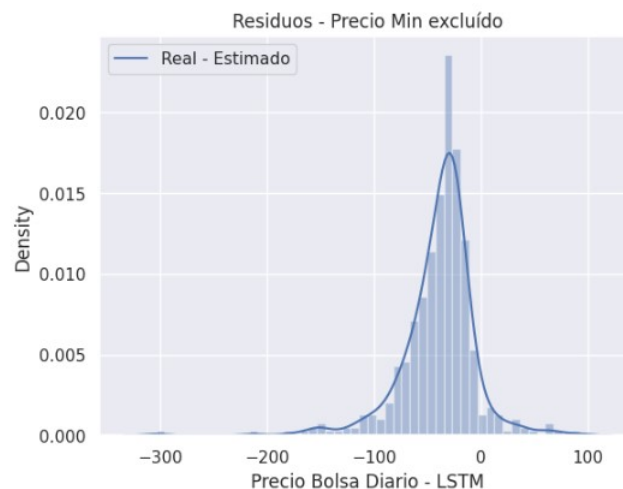


Figura 27: Distribución de los residuos método LSTM con la variable Precio Mínimo excluída.

Para explicar este comportamiento se entrenó otro modelo de las mismas características pero dejando solo el valor máximo Figuras 27 y 28. De allí se obtuvo que la red comenzó a sobreestimar los precios. Por lo que se puede considerar que el valor mínimo tiene una influencia fuerte en las estimaciones. Esta afirmación es soportada por la correlación mayor que existe entre el precio mínimo y el precio diario



(0.84), con respecto a la correlación que hay entre el precio máximo y el precio diario (0.59) (Figura 21).

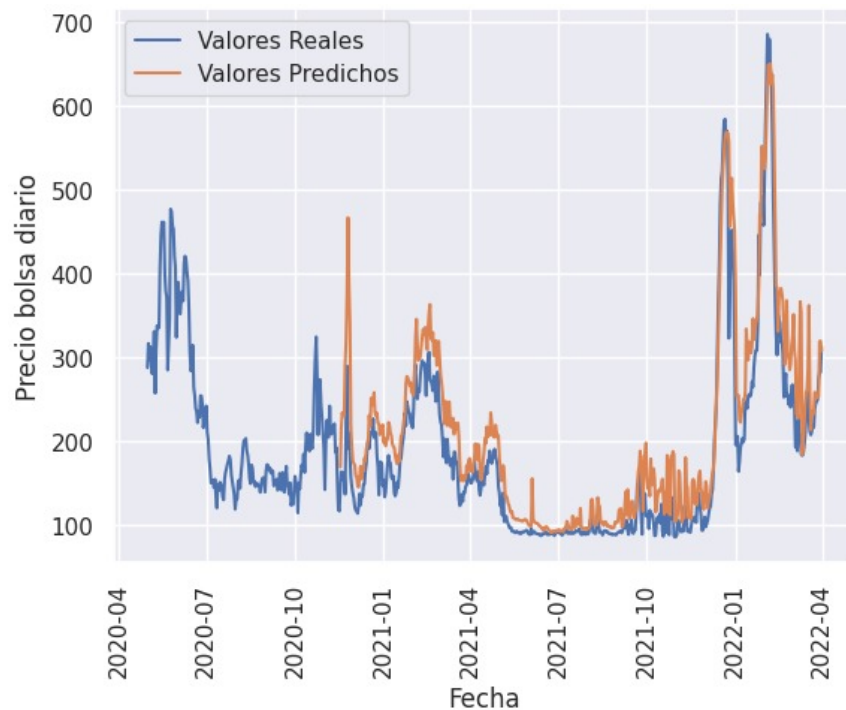


Figura 28: Comparación de los datos reales con los datos pronosticados por la red neuronal.

5. Conclusiones

- El desarrollo del proyecto ha involucrado varios métodos que confirman consistentemente que el recurso con mayor impacto en el precio de la energía en Colombia es el agua. Esto se debe a la amplia presencia y contribución de la generación hidroeléctrica en la matriz energética nacional.
- No existen correlaciones significativas entre el precio de los diferentes recursos de generación. Sin embargo, se aprecia una leve correlación entre el precio del gas y el precio del carbón.
- El significativo impacto del precio del agua en el precio de la energía se refleja en una marcada variabilidad en el precio de la energía, especialmente durante eventos naturales como El Niño, que tienen un impacto considerable en la disponibilidad de este recurso de generación. Durante estos períodos, la disminución de la disponibilidad de agua para la generación hidroeléctrica puede generar un aumento en el precio de la energía, debido a la mayor dependencia de otras fuentes de generación más costosas.
- El precio del agua influye fuertemente el precio de bolsa cuando se encuentra por debajo de \$400. Para valores superiores el precio del mercado no tiene relación lineal con el agua sino que debe



ser representado como una función del costo de las otras variables, precio de combustible y precio de bolsa.

- Los métodos de regresión lineal pueden dar explicaciones de las tendencias dominantes en los datos y son muy apropiados para hacer análisis cuantitativo de los datos. Sin embargo para hacer predicciones tienen debilidades y que puede ocurrir que no todo el conjunto de estos se represente correctamente con las regresiones, de forma que se requieren modelos más complejos.
-

Referencias

- [1] Enel, "Descripción del mercado", disponible en: <https://www.enel.com.co/es/empresas/enel-generacion/como-se-genera-la-electricidad.html>
- [2] El Colombiano, " Alistan 30 térmicas para frenar riesgo de un apagón por el arribo de El Niño", disponible en <https://www.elcolombiano.com/inicio/alistan-30-termicas-para-frenar-riesgo-de-un-apagon-por-el-arribo-de-el-nino-GH21405344>