

Puntos Por Desarrollar

Primer Punto

Ustedes cuentan con varias bases de datos que cuenta con información del precio del Petróleo Brent, del Gas Natural, Carbón, Gasolina Motor Corriente y finalmente del IPC para Colombia desde el 2000 hasta el 2024 con una periodicidad mensual. Cada una de estas bases de datos toma el nombre del bien al que hacen referencia junto con el sufijo .csv. Por ejemplo, cuentan con "Carbón.csv" y así con los demás bienes. Cada una de estas bases cuenta con tres columnas, una primera columna con la fecha mensualizada, una segunda columna con el precio del bien para Colombia y una tercera columna con el precio del bien promedio en el mundo. En su trabajo como analistas de información se les solicita realizar un reporte sobre estas series. Para ello, su jefa les solicita realizar las siguientes tareas:

- 1.1. Establezcan un código que les permita cargar en el programa R las bases de datos de los diferentes precios de combustibles.

Solución:

Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

Sin embargo, nos parece importante mencionar dos aspectos relevantes:

- Se utilizó la función `read_csv` de la librería `readr` para importar los datos de precios de los bienes. Esta función fue seleccionada porque permite importar las columnas de fecha como objetos de tipo `Date` de manera eficiente."
- Usamos una referencia de internet para importar datos de un archivo ".zip". El link de la referencia es el siguiente: Link: <https://www.statology.org/read-zip-file-in-r/>

- 1.2. Realicen una exploración inicial de la base de datos. Para esto, para cada una de las bases de datos describan la misma. Es decir, redacten un breve párrafo comentando qué variables contiene la base de datos, el número de observaciones, el periodo de tiempo que comprende, el tipo de las variables, entre otros aspectos relevantes.

Solución:

Conjunto de datos de petróleo:

El conjunto de datos de carbón contiene 8,667 observaciones y 2 columnas: una correspondiente al precio del petróleo, y otra a la fecha, con frecuencia diaria, que abarca el periodo desde el 1 de enero de 2000 hasta el 1 de enero de 2024. No obstante, el conjunto presenta fechas faltantes, lo que afecta la precisión de la frecuencia diaria, y tiene 88 valores

faltantes en la columna de precios. La columna de fecha es de tipo date, mientras que la columna de precio es de tipo numérico. El precio promedio es de 108.09 y la desviación estándar del precio es 16.26.

Conjunto de datos de carbón:

El conjunto de datos de carbón contiene 8,667 observaciones y 2 columnas: una correspondiente al precio del carbón y otra a la fecha, con frecuencia diaria, que abarca el periodo desde el 1 de enero de 2000 hasta el 1 de enero de 2024. No obstante, el conjunto presenta fechas faltantes, lo que afecta la precisión de la frecuencia diaria, y tiene 88 valores faltantes en la columna de precios. La columna de fecha es de tipo date, mientras que la columna de precio es de tipo numérico. El precio promedio es de 96.96 y la desviación estándar del precio es 19.07.

Conjunto de datos gas natural:

El conjunto de datos de carbón contiene 8,667 observaciones y 2 columnas: una correspondiente al precio del gas natural y otra a la fecha, con frecuencia diaria, que abarca el periodo desde el 1 de enero de 2000 hasta el 1 de enero de 2024. No obstante, el conjunto presenta fechas faltantes, lo que afecta la precisión de la frecuencia diaria, y tiene 88 valores faltantes en la columna de precios. La columna de fecha es de tipo date, mientras que la columna de precio es de tipo numérico. El precio promedio es de 118.31 y la desviación estándar del precio es 21.61.

Conjunto de datos gasolina:

El conjunto de datos de carbón contiene 8,667 observaciones y 2 columnas: una correspondiente al precio de la gasolina y otra a la fecha, con frecuencia diaria, que abarca el periodo desde el 1 de enero de 2000 hasta el 1 de enero de 2024. No obstante, el conjunto presenta fechas faltantes, lo que afecta la precisión de la frecuencia diaria, y tiene 88 valores faltantes en la columna de precios. La columna de fecha es de tipo date, mientras que la columna de precio es de tipo numérico. El precio promedio es de 108.94 y la desviación estándar del precio es 29.22.

Conjunto de datos IPC:

El conjunto de datos del IPC contiene 289 observaciones y 5 columnas: una para el valor del índice, otra para la variación anual, otra para la variación año corrido y una más para la fecha. La fecha abarca desde enero de 2000 hasta enero de 2024 con frecuencia mensual y no tiene fechas faltantes en dentro del periodo de tiempo que abarca. Este conjunto no presenta valores faltantes, y todas las columnas son de tipo numérico, ya que la fecha está representada como un número entero. Las estadísticas descriptivas de las variables en esta base de datos se encuentran en la siguiente tabla:

Tabla 1: Estadísticas descriptivas de la base de datos del IPC

Variable	Observaciones	Promedio	Desviación estándar	Mínimo	Máximo
Índice de Precios al Consumidor	289	79.165	24.132	40.300	138.980
Variación anual	289	5.345	2.677	1.490	13.340
Variación mensual	289	0.435	0.425	-0.380	2.300
Variación año corrido	289	3.651	2.211	0.300	13.120

Fuente: Elaboración propia con Datos_taller.zip, 2024.

Explorando las bases de datos de los precios de forma individual identifican que no todas tienen el mismo número de observaciones. Un hecho muy común cuando se trabaja con datos en formato panel es que algunas fechas para algunos productos no estén en la base de datos, es decir, hay fechas faltantes. A continuación, ustedes quieren “explicitar” las fechas faltantes.¹

1.3. Creen una función que tenga como parámetros un dataframe, una columna de fecha y una columna de precio, posteriormente, la función tiene que agregar las fechas faltantes en la columna de fecha y añadir un missing value en su valor correspondiente en el precio para esta fecha explicitada. Finalmente, tiene que retornar el dataframe con las fechas completas.

Solución:

Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

Sin embargo, nos parece relevante mencionar que para crear un vector que contenga todos los días desde 1 de enero de 2000 hasta el 1 de enero de 2024 se utilizó la siguiente referencia de internet:

Enlace:<https://www.tutorialspoint.com/how-to-create-a-vector-with-all-dates-in-a-particular-year-in-r>

También queremos resaltar que decidimos excluir el parámetro de “columna de precio” de la función, debido a que se la función que creamos realiza la operación de agregar las fechas faltantes sin necesidad de llamar a la columna de precios por su nombre. Por

¹ Este punto se apoyará y desarrollará en la tarea 3. Una forma de hacerlo es generando las fechas faltantes y posteriormente realizando un *merge* de esas fechas con la base de datos original.

lo tanto, la función va a cumplir su propósito sin importar cuál sea el nombre de la columna de precio.

1.4. Apliquen la función anterior a cada uno de los precios en la base de datos.

Solución:

Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

1.5. Realicen una unión de cada una de los datasets cargados que les permita juntar los precios y fechas en una misma base de datos.

Solución:

Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

Sin embargo, nos parece relevante mencionar que se utilizó una referencia de internet para unir varios conjuntos de datos con el operador “pipe” (`%>%`).

Enlace: <https://stackoverflow.com/questions/32066402/how-to-perform-multiple-left-joins-using-dplyr-in-r>

1.6. Utilizando las funciones de la librería *lubridate*, creen respectivamente columnas que correspondan al mes y año de las observaciones (revisen la documentación de: [Make Dealing with Dates a Little Easier • lubridate \(tidyverse.org\)](#) y [t](#)).

Solución:

Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

1.7. Revisen y presenten sus resultados en una tabla: ¿Qué columnas y en qué porcentajes de la base de datos cuenta con valores faltantes en el precio de algún bien? Sustituyan estos valores faltantes con el precio promedio de ese bien.

Solución:**Tabla 2: Valores faltantes en las bases de datos**

	Porcentaje de la muestra	Observaciones
Precio del petróleo	2.14%	188
Precio del carbón	2.14%	188
Precio del gas natural	2.14%	188
Precio de la gasolina	2.14%	188

Fuente: Elaboración propia con Datos_taller.zip, 2024.

La tabla anterior muestra que cada una de las bases de datos de precios de hidrocarburos tiene valores faltantes en el 2.1% de las observaciones. Esto corresponde a 188 días entre el primero de enero del 2001 y el primero de enero del 2024 para los que hay valores faltantes en los precios de los hidrocarburos. Vale la pena mencionar que las fechas de los días que tienen valores faltantes no son las mismas en todas las bases de datos.

1.8. Agrupen la información de los precios diarios a los valores promedios por mes y año.**Solución:**

Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

Por otra parte, para poder realizar comparaciones correctas entre precios es importante convertir los precios nominales en precios reales (constantes de un año base), pues permite controlar por el efecto inflacionario. Para transformarlo, se normaliza por la diferencia entre los Índices de Precios al Consumidor de la forma:

$$(1) \quad P_{real,t} = \frac{P_{nominal,t} * IPC_{base}}{IPC_t}$$

Donde IPC_{base} hace referencia al IPC de un año-mes específico. IPC_t al IPC del año-mes t sobre el cual se quiere convertir el precio nominal del año-mes t ($P_{nominal,t}$).

1.9. Creen una función que tenga como parámetros: una columna de una serie de un precio de la base de datos, un año, un mes y la columna del IPC de Colombia. Posteriormente, la función debe crear una nueva columna cuyo nombre sea: el “nombre_del_bien” + “ano_base” + “mes_base” + el sufijo “_transformada”. Esta columna debe ser la transformación de valores

nominales de la serie, a valores reales con base en el año-mes que toma como parámetro la función. Es decir, utilizando la ecuación (1) tendrán que consolidar una función que transforme los valores nominales en precios constantes con base en cualquier mes-año.

Solución: Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

Sin embargo, nos parece relevante mencionar que se utilizó una referencia de internet para invertir el orden de un vector con la función `rev()`.

Enlace: <https://www.geeksforgeeks.org/how-to-reverse-the-order-of-a-dataframe-in-r/>

- 1.10. Apliquen la función que desarrolló en el literal anterior para todos los bienes que definió en la base de datos, pueden usar cualquier año-mes base, lo importante es que sean explícitos.**

Solución:

Utilizamos el mes de enero del año 2000 como año base para calcular los precios reales de los combustibles.

- 1.11. Exporte la base de datos consolidada en un archivo .csv**

Las instrucciones de este punto no requieren argumentación ni interpretación de resultados, ni tampoco responder a una pregunta en forma de texto. Por lo tanto, la solución se limita únicamente al script de R asociado al taller.

Segundo punto

2. A partir de la base de datos consolidada en el punto anterior, este punto busca explorar y analizar los datos consolidados. Para esto, realicen lo siguiente:

2.1. Presenten una tabla de estadísticas descriptivas donde evidencien el número de observaciones, promedio, min, max, desviación estándar de los precios nominales y reales de los combustibles. Expliquen e interpreten en el documento los valores más importantes de la tabla.

Solución:

Tabla 3: Estadísticas descriptivas de los precios de los combustibles

	Observaciones	Promedio	Desviación estándar	Mínimo	Máximo
Precios nominales					
Precio nominal del Carbón	289	96.980	12.437	74.211	123.163
Precio nominal del Gas Natural	289	118.245	8.164	98.711	137.059
Precio nominal del Petróleo	289	108.079	7.801	91.248	128.404
Precio nominal de la Gasolina	289	109.136	25.226	61.648	166.097
Precios reales (base: enero de 2001)					
Precio real del Petróleo	289	59.221	14.529	30.65	95.936
Precio real del Carbón	289	52.196	9.753	29.544	79.565
Precio real de la Gasolina	289	56.791	4.440	43.948	71.006
Precio real del Gas Natural	289	64.943	16.420	28.623	107.789

Fuente: Elaboración propia con Datos_taller.zip, 2024.

Todos los precios cuentan con 289 observaciones, que corresponden a los meses entre enero del año 2000 y enero del año 2024.

Promedio: Representa el valor promedio de los precios en entre enero del 2000 y enero del 2024.

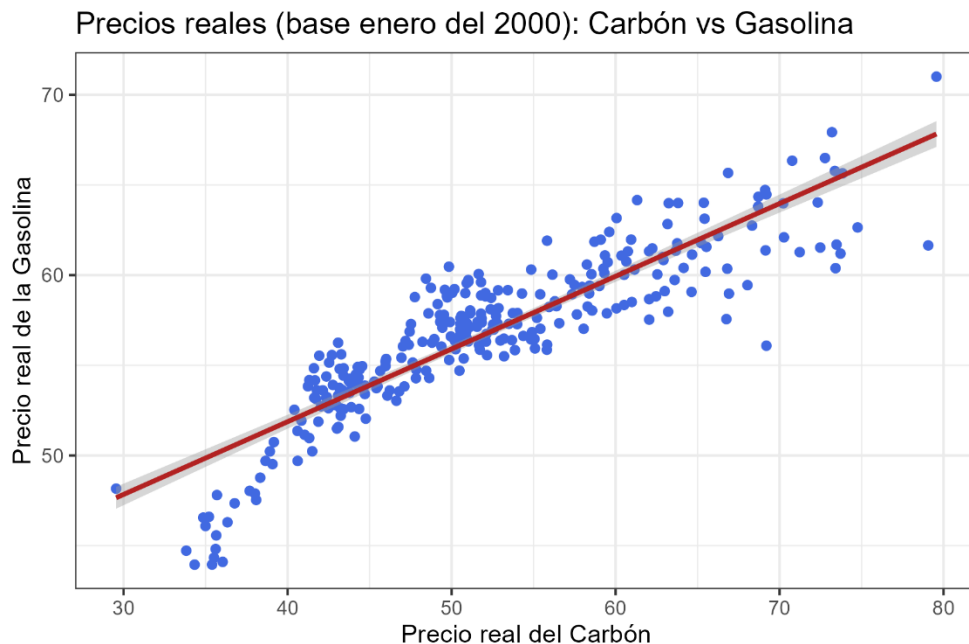
- El precio promedio nominal (sin ajustar por inflación) del carbón es 96.980, mientras que el del gas natural es 118.245, el del petróleo es 108.079 y el de la gasolina es 109.136.
- En términos reales (ajustados por inflación con base en enero de 2001), el promedio del precio del petróleo es 59.221, el del carbón es 52.196, el de la gasolina es 56.791 y el del gas natural es 64.943.

Desviación estándar: Mide la variabilidad o dispersión de los precios respecto al promedio. Un valor mayor indica mayor fluctuación.

- Por ejemplo, la gasolina tiene una desviación estándar de 25.226, lo que indica que sus precios nominales han tenido una mayor variabilidad que los del petróleo (7.801). En términos reales, el gas natural tiene una alta variabilidad (16.420), mientras que la gasolina tiene menos variabilidad (4.440).

2.2. Presenten una gráfica de dispersión de puntos (scatter) donde en el eje X esté el precio del carbón y en el eje Y el precio de la gasolina, ambos reales, para un año determinado. Añadan una línea de ajuste lineal, título, labels y demás elementos que permitan una gráfica autocontenida. Expliquen e interpreten en el documento los elementos más importantes de la base de datos.

Solución:



Fuente: Elaboración propia con Datos_taller.zip, 2024.

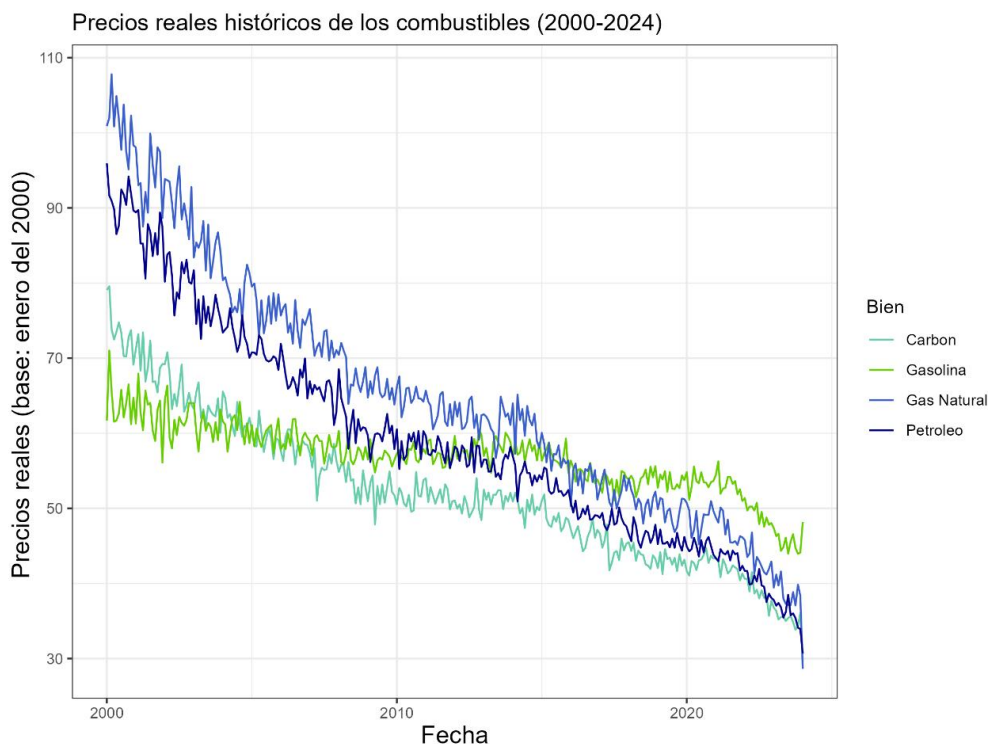
La gráfica muestra la relación entre los precios reales del carbón y los precios reales de la gasolina, ambos con base en enero del 2000. Los puntos azules representan pares de observaciones de precios en diferentes momentos del tiempo, mientras que la línea roja muestra una regresión lineal ajustada a los datos.

En primer lugar, la gráfica muestra una correlación positiva entre el precio real del carbón y el precio real de la gasolina. Esto indica a medida que el precio real del carbón aumenta, el precio real de la gasolina también tiende a aumentar. Esta tendencia muestra cómo los mercados de energía están interrelacionados. El carbón y la gasolina, ambos combustibles fósiles, pueden experimentar variaciones en sus precios debido a factores comunes como los costos de extracción, las políticas energéticas, o cambios en la demanda de energía.

Adicionalmente, se puede observar que la línea de tendencia no se ajusta perfectamente a los a datos (antes de que el carbón tome precio de 40) lo que indica que la relación entre estas variables no es lineal. Finalmente, se observa que la dispersión de los puntos aumenta a medida que aumenta el precio de ambos bienes.

2.3. Presenten una gráfica de serie de tiempo con el precio real del Carbón, Gas Natural, Petróleo y Gasolina Corriente para un mismo año, asegúrense de que tenga el título, labels, colores y demás elementos que permitan una gráfica autocontenida. Expliquen e interpreten en el documento los elementos más importantes de la base de datos.

Solución:



Fuente: Elaboración propia con Datos_taller.zip, 2024.

La gráfica presenta los precios reales históricos de varios combustibles entre los años 2000 y 2024, ajustados a la base de enero de 2000. Los combustibles analizados son el carbón, la gasolina, el gas natural y el petróleo, cada uno representado por una línea de diferente color.

En primer lugar, se observa que todas las series presentan una tendencia negativa, lo que indica que los precios de los combustibles han disminuido en términos reales. Esta tendencia puede deberse a una variedad de factores como mejoras en las tecnologías de producción de combustibles fósiles y el crecimiento de la demanda de fuentes de energía renovables (como la energía eólica, solar e hidráulica).

Además, en el año 2000, el gas natural registraba el precio real más alto, mientras que para el año 2024, la gasolina se posiciona como el bien con el precio más elevado. Asimismo, las series muestran variaciones en la intensidad de sus tendencias. El gas natural es el combustible con la mayor tendencia a la baja, mientras que la gasolina presenta la menor disminución en su precio real. Por último, la volatilidad también difiere entre las series: el gas natural exhibe mayores fluctuaciones, mientras que el carbón y la gasolina presentan cambios más moderados y aparentemente menos pronunciados.