

Laboratorio 01 Series de Tiempo

INTRODUCCIÓN:

Trabaje con los datos de **importación y consumo** de gasolina super, gasolina regular y diesel (alto azufre) de los últimos 24 años. En el caso de las importaciones, considere como las importaciones de diesel desde 2001, los datos del diesel con alto contenido de azufre.

Esta hoja de trabajo se realizará en grupos de 2 o 3 personas. Es obligatorio que los datos de todos los integrantes se incluyan de manera explícita en cada uno de los entregables.

DESCRIPCIÓN DEL DATASET

El conjunto de datos histórico de importación de combustibles contiene 288 observaciones. Representa el volumen de importaciones de distintos tipos de combustibles desde enero de 2001 hasta diciembre de 2024. Mientras que el conjunto de datos histórico sobre consumo de combustibles contiene 300 observaciones. El volumen, en ambos conjuntos, está dado en barriles de 42 galones. Los datos los puede encontrar en el siguiente vínculo: [Estadísticas de comercialización de hidrocarburos – MEM](#)

Algunos vínculos interesantes:

- <https://otexts.com/fpp2/>
- <https://otexts.com/fpp2/arima.html>
- <https://otexts.com/fpp2/accuracy.html>
- <https://robjhyndman.com/papers/mase.pdf>

EJERCICIOS

1. Haga un análisis exploratorio de los datos que se le presentan (recuerde utilizar solo las columnas de gasolinas regular, super y diesel de ambos conjuntos de datos), se sugiere explorar el comportamiento de las variables y si están distribuidas normalmente, en caso de ser continuas. Meses en los que más importaciones hay, picos en importaciones por año por tipo de combustibles, comportamiento en los últimos x años, comportamiento durante la pandemia, etc.
2. Haga una serie univariante por cada una de las variables (gasolina regular, gasolina super, diesel) de cada conjunto (6 series en total). De cada serie:
 - a. Especifique Inicio, fin, y frecuencia.
 - b. Haga un gráfico de la serie y explique qué información puede obtener a primera vista.

- c. Descomponga la serie en sus componentes. Teniendo en cuenta el diagrama de la serie y sus componentes discuta si es posible hablar de estacionariedad en media y en varianza.
 - d. Determine si es necesario transformar la serie. Explique.
 - e. Explique si no es estacionaria en media. Para esto:
 - i. Haga el gráfico de autocorrelación y úselo para explicar la no estacionariedad en media.
 - ii. Básese en los valores de estadísticos como la prueba de Dickey-Fuller aumentada para corroborar la no estacionariedad en media. ¿Qué es necesario hacer para hacerla estacionaria en media en caso de que no lo sea?
 - f. Una vez analizada la serie, elija los parámetros p , q y d del modelo ARMA o ARIMA que utilizará para predecir. Explique en qué se basó para darle valor a estos parámetros, basándose en las funciones de autocorrelación y autocorrelación parcial. Si usa la función `autoarima` de R, determine y explique si tiene sentido el modelo propuesto.
 - g. Haga varios modelos ARIMA, y diga cuál es el mejor de ellos para estimar los datos de la serie.
 - h. Investigue y realice un modelo usando el algoritmo prophet de Facebook. Compárelo con los modelos del inciso anterior. ¿Cuál funcionó mejor?
3. Haga una predicción de los valores de la serie para los últimos 3 años. Explique qué tan buena es su predicción.
 4. Diríjase nuevamente al [enlace](#) de descargas del dataset, pero ahora descargue los datos de este año (que van hasta Mayo 2025). Haga una predicción a partir del año 2024 y 2025 usando los años anteriores y determine qué tan apegada es a la realidad, tomando en cuenta los meses que van de este año.
 5. Discuta cuál fue el comportamiento de las series durante el fenómeno de la pandemia. ¿Qué tanto afectó las predicciones? ¿Se recuperaron los valores habituales en el regreso a la “nueva normalidad” hasta el mayo de 2025?

EVALUACIÓN

NOTA: La evaluación de cada integrante del grupo será de acuerdo con sus contribuciones al trabajo grupal

(15 puntos) Análisis exploratorio:

- Se elaboró un análisis exploratorio en el que se explican los cruces de variables, hay gráficos explicativos y análisis que permiten comprender el conjunto de datos.

- Se crearon las series de tiempo correspondientes a los datos de las variables que representan las importaciones y consumo de diesel, gasolina superior y gasolina regular.
- Para cada una de las series se informa inicio, fin y frecuencia.

(15 puntos) Análisis de las series de tiempo

- Para cada una de las series creadas se analiza:
 - El gráfico de la serie y su descomposición en componentes.
 - Si la serie presenta estacionalidad o no y que implica que sí tenga.
 - Si la serie presenta tendencia o no y esto que significa.

(25 puntos) Determinación de Estacionariedad

- Para cada una de las series creadas:
 - Se analiza si es estacionaria en varianza y en caso de no serlo se aplica una transformación adecuada.
 - Se analiza si es estacionaria en media, para esto se basa en la función de autocorrelación y en la prueba de Dickey-Fuller aumentada. Se determina la cantidad de diferenciaciones que hay que hacer en caso de que no sea estacionaria en media.

(25 puntos) Generación de modelos

- Para cada una de las series creadas:
 - Se determinan los valores de los parámetros p , q , y d . Para esto se basa en las funciones de autocorrelación y autocorrelación parcial.
 - Se explica la elección de los parámetros y de los modelos en caso de que se propongan varios modelos. Se deben explicar los parámetros, aunque sean propuestos de forma automática por R (en caso de usar este lenguaje).
 - Se incluye la generación de un algoritmo de prophet.

(20 puntos) Predicción con los modelos generados

- Para cada una de las series creadas:
 - Se crean los conjuntos de entrenamiento y prueba siguiendo las instrucciones.
 - Se explica que tan bueno es el modelo prediciendo los volúmenes de importación de combustible para el conjunto de prueba.
 - Se comparan los modelos generados con ambos algoritmos.
 - Se discute el impacto de la pandemia en las predicciones del último año y medio.

MATERIAL A ENTREGAR

- Archivo .r, .rmd, .ipynb, o .py, con el código de las acciones tomadas desde que se carga el conjunto de datos hasta que se termina de limpiar.
- Link del repositorio con el código, se debe evidenciar el trabajo colaborativo de todos los integrantes.
- Documento del Google docs donde se trabajó el libro de códigos
- Archivo pdf, con el libro de códigos

- Archivo csv con los datos limpios utilizados