

Laboratorio 2. Series de Tiempo.

INSTRUCCIONES:

Se trabajará en este laboratorio con los datos de importación, consumo y precios de gasolina super, gasolina regular, diesel y gas propano. En el caso de las importaciones, considere como las importaciones de diesel desde 2018, los datos del diesel con bajo contenido de azufre (dieselLS).

Esta hoja de trabajo se realizará en PAREJAS. Para que se pueda calificar su laboratorio debe estar inscrito en algún grupo de canvas.

DESCRIPCIÓN DEL DATASET

Descripción de los conjuntos de datos:

- Consumo de combustibles
 - o Contiene 300 observaciones desde enero del año 2000 hasta mayo del año 2024
 - o El volumen está dado en barriles de 42 galones
- Importación de combustibles
 - o Contiene 280 observaciones desde enero del año 2001 hasta mayo del año 2024
 - o El volumen está dado en barriles de 42 galones
- Precios Promedios a consumidor final en Ciudad Capital
 - o Contiene observaciones con los precios diarios de combustibles desde el 1 enero de 2021 hasta el 28 de julio de 2024.
 - o El precio está dado en Quetzal por galón en el caso de las gasolinas regular y super, y en quetzal por cilindro de 25 lbs en el caso del gas licuado.

Los datos los puede encontrar en el siguiente vínculo: [Estadísticas de comercialización de hidrocarburos – MEM](#), y los de precios en el siguiente vínculo: <https://mem.gob.gt/historico-precios-nacionales/>

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, diesel y gas licuado de todos los conjuntos), 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. Van a trabajar con 3 de las siguientes series, al menos 1 de ellas debe estimar precios:
 - a. Consumo mensual de gasolina super, gasolina regular, diésel y gas licuado
 - b. Importación mensual de gasolina super gasolina regular, diésel y gas licuado
 - c. Precios en ciudad capital de gasolina super, gasolina regular, diésel y gas licuado, puede usar las de autoservicio o las de servicio completo.
3. De cada serie de las seleccionadas:
 - 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. 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. Báse 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. Para esto analice los residuos y las métricas AIC y/ BIC.
 - h. Haga un modelo usando el algoritmo prophet de Facebook o redes neuronales. Compárelo con los modelos del inciso anterior. ¿Cuál funcionó mejor?
4. Haga una predicción de los valores de la serie para los últimos 3 años. Explique qué tan buena es su predicción. En el caso de la serie de precios, use los datos de 2024 como conjunto de prueba
5. Haga una predicción para todo este año 2024 determine qué tan apegada es a la realidad teniendo en cuenta los meses pasados del año actual.
6. Discuta cuál fue el comportamiento de las series durante la pandemia. ¿Qué tanto afectó las predicciones? ¿se recuperaron los valores habituales en el regreso a la “nueva normalidad”?

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, consumo y precio de diésel, gasolina superior, gasolina regular y gas licuado.
- Para cada una de las series se informa inicio, fin y frecuencia.
- Se explora el comportamiento de la serie durante y después de la pandemia.

(15 puntos) Análisis de las series de tiempo

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

(25 puntos) Determinación de Estacionariedad.

- Para cada una de las series creadas:
 - o Se analiza si es estacionaria en varianza y en caso de no serlo se aplica una transformación adecuada.
 - o 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:
 - o Se determinan los valores de los parámetros p, q, y d. Para esto se basa en las funciones de autocorrelación y autocorrelación parcial.
 - o Se explica la elección de los parámetros y de los modelos (al menos 2 por serie). Se deben explicar los parámetros, aunque sean propuestos de forma automática por R (en caso de usar este lenguaje).
 - o Se genera un modelo con el algoritmo prophet o redes neuronales.
 - o Se comparan los modelos de acuerdo con el comportamiento de los residuos y las métricas AIC y BIC.

(20 puntos) Predicción con los modelos generados.

- Para cada una de las series creadas:
 - o Se crean los conjuntos de entrenamiento y prueba siguiendo las instrucciones.
 - o Se explica que tan bueno es el modelo prediciendo los volúmenes de importación, consumo y precio de combustible para el conjunto de prueba.
 - o Se comparan los modelos generados con ambos algoritmos.
 - o Se predicen los datos de consumo, importación y precio para todo el 2023, se compara con lo que se conoce hasta el momento. ¿Qué tan bueno son los modelos?
 - o Se discute el impacto de la pandemia en la predicción.

MATERIAL A ENTREGAR

- Archivo .pdf con el informe que contenga, los resultados de los análisis y las explicaciones.
- Link de Google drive donde trabajó el grupo.
- Script de R (.r o .rmd) o de Python que utilizó para hacer su análisis exploratorio y predicciones.
- Link del repositorio usado para versionar el código.

FECHAS DE ENTREGA

- 1 de agosto de 2024 23:59:
 - .1. (55 puntos) Análisis Exploratorio, análisis de las series de tiempo, determinación de estacionariedad.
- 21 de julio de 2024 23:59
 - .1. (45 puntos) Generación de modelos, predicción y análisis de los resultados obtenidos.

NOTA: Para poder tener nota completa debe entregar las asignaciones en el tiempo adecuado. No se calificará el avance del laboratorio si no fue entregado en tiempo, aunque esté en el repositorio.