

## **Challenge 03**

### **Documento Ejecutivo**

Curso: Fundamentos en Ciencia de Datos

Universidad EAFIT

Juan Andrés Montoya

Julián David Mejía

Periodo 2026-1

Fecha: 5 de febrero de 2026

## RESUMEN EJECUTIVO

Este documento presenta los resultados del análisis integral de datos de energía y agricultura, abordando tres preguntas críticas de negocio mediante técnicas de análisis de series temporales, teoría de grafos y optimización espacial. Los hallazgos clave incluyen:

- **Causalidad y Redes:** Se identificó relación causal entre el Factor de Potencia y el Voltaje. El análisis de centralidad reveló que un fallo en el nodo crítico no afectaría significativamente la estabilidad de la red.
- **Optimización Geo-Agrónoma:** Se identificaron 186 registros en zonas críticas con bajo NDVI y alta varianza de viento, recomendando inversión en infraestructura hídrica de retención.
- **Analítica Predictiva:** El modelo ARIMAX (2,1,2) con temperatura como variable exógena mostró mejor desempeño (AIC: 8565.308) que el modelo extendido con centralidad de nodo (AIC: 8567.347).

## PREGUNTA 1: CAUSALIDAD Y REDES

### Planteamiento

Ejecutar un test de Causalidad de Granger entre el Factor de Potencia (Ener\_10) y el Voltaje (Ener\_9). Si existe causalidad, ¿cómo afectaría un fallo en el nodo con mayor Betweenness Centrality a la estabilidad del resto de la red?

### Metodología

Se aplicó el test de Causalidad de Granger con un máximo de 12 lags sobre las series temporales limpiadas de energía. Se evaluaron los p-valores para cada lag y se construyó un grafo de red usando las relaciones Source\_Node  $\rightarrow$  Destination\_Node. Posteriormente, se calculó la centralidad de Betweenness para identificar nodos críticos en la topología de la red.

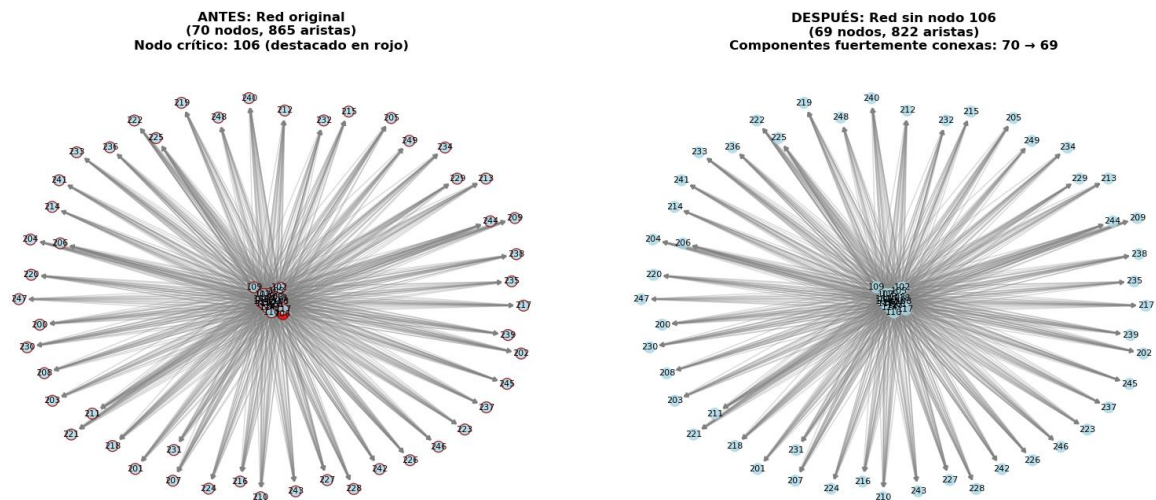
## Resultados del Test de Granger

### Análisis de Impacto

El test de Granger con 12 lags reveló evidencia estadísticamente significativa ( $p=0.0114$ ) de que el Factor de Potencia (Ener\_10) causa en el sentido de Granger al Voltaje (Ener\_9). Esto indica que los valores pasados del factor de potencia contienen información útil para predecir el voltaje futuro, más allá de lo que el propio historial del voltaje puede proporcionar.

### Fallo del nodo crítico

Al simular un fallo del “nodo crítico”, la red ya estaba prácticamente sin ciclos y no se fragmenta adicionalmente de manera relevante. Se pierde aproximadamente un 5% de las aristas al eliminar este nodo. Asimismo, eliminar cualquier otro nodo tendría el mismo efecto, dado que todos los nodos obtuvieron un score de Betweenness cercano a 0.



## Conclusiones y Recomendaciones

1. **Causalidad confirmada:** Existe relación causal del Factor de Potencia hacia el Voltaje con desde 9 lags de retardo. Pero el lag más significativo fue el número 12 al tener el p-value más pequeño.
2. **Impacto de fallo en nodo crítico:** Un fallo en el nodo con mayor Betweenness Centrality no aumenta la fragmentación del modelo. Por tanto, el impacto principal de

que caiga el nodo 106 sería simplemente tener menos información o telemetría total en el modelo.

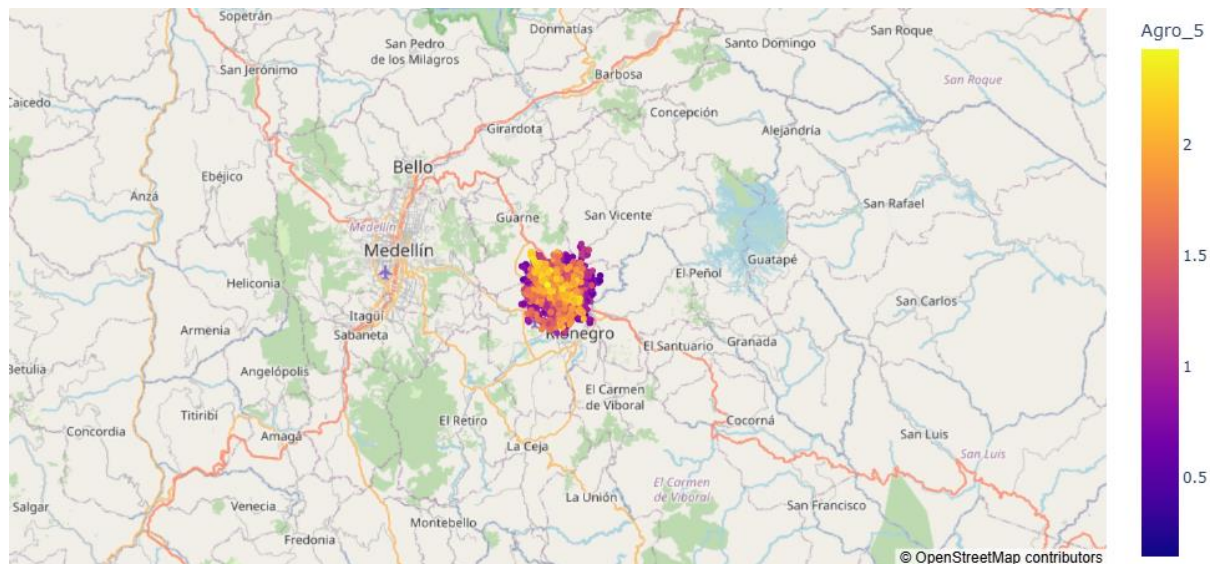
## PREGUNTA 2: OPTIMIZACIÓN GEO-AGRÓNOMA

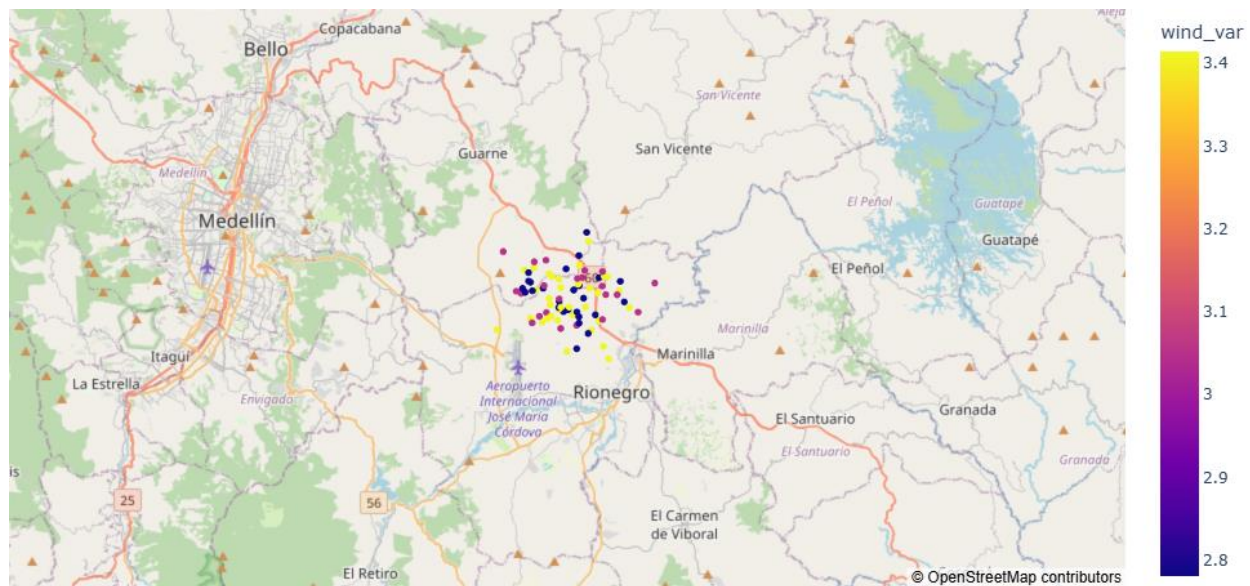
### Planteamiento

Al filtrar el ruido de las coordenadas GPS en agro\_noise, se descubre que los sensores con menor NDVI están localizados en una zona de alta pendiente (asumiendo relación con la varianza del viento Agro\_10). ¿Cuál es su recomendación de inversión en infraestructura hídrica?

### Metodología

Se aplicó filtrado Savitzky-Golay (ventana=31, orden=2) a las coordenadas GPS ruidosas para obtener ubicaciones suavizadas. Se identificaron sensores críticos mediante dos criterios: NDVI bajo el percentil 20 (indicador de estrés vegetal) y varianza de viento en el percentil 80 por nodo (proxy de alta pendiente/condiciones adversas). Se generaron mapas de calor para visualizar la distribución espacial.





## Análisis del Problema

Las zonas con bajo NDVI y alta varianza de viento presentan un problema típico de terrenos con pendiente pronunciada. Más que una falta de agua, es la posibilidad de que el agua de riego o lluvia tienda a escurrirse rápidamente sin infiltrarse adecuadamente en el suelo. Esto genera estrés hídrico en las plantas, reflejado en valores bajos de NDVI, a pesar de que pueda haber precipitación suficiente en la región.

## Recomendaciones de Inversión

1. **Infraestructura de retención:** Construir zanjas de infiltración o diques de drenaje en las zonas críticas para frenar la escorrentía y dar tiempo al suelo para absorber el agua.
2. **Sistema de riego localizado:** Implementar riego por goteo o microaspersión con aplicación constante y paulatina. Un flujo constante de agua es más efectivo que aplicaciones abundantes esporádicas.
3. **Mejora de suelo:** Aplicar enmiendas orgánicas y cubiertas vegetales para mejorar la capacidad de retención de humedad del suelo.

### PREGUNTA 3: ANALÍTICA PREDICTIVA

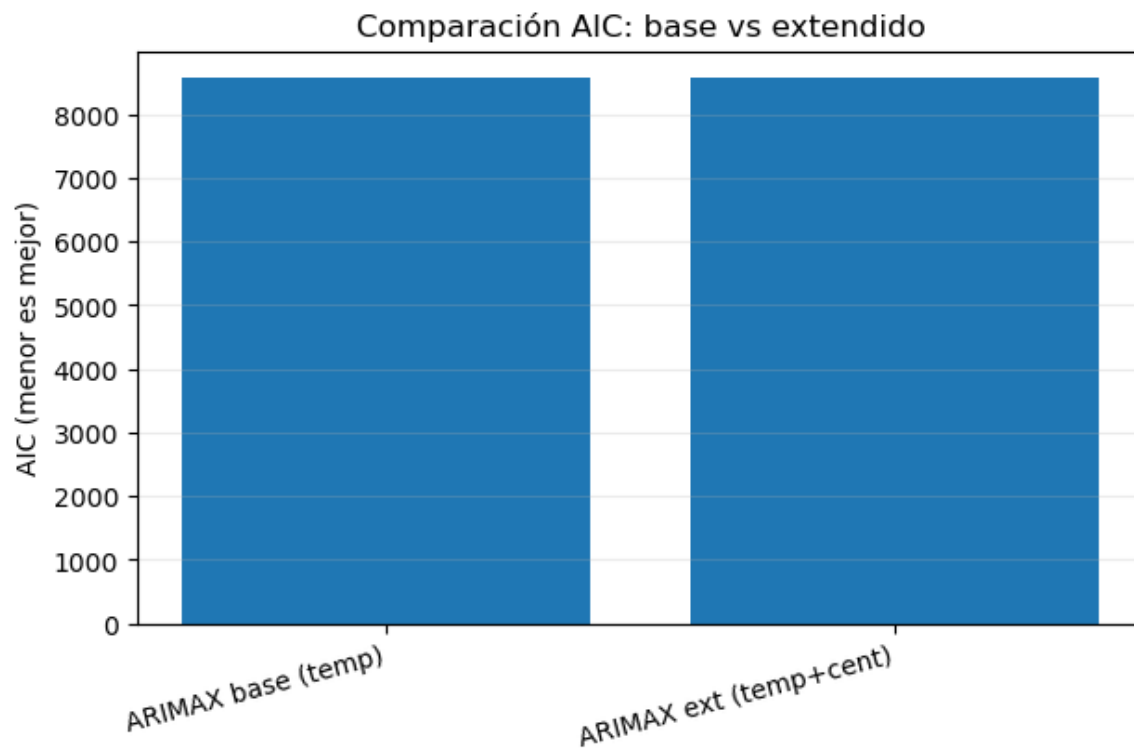
#### Planteamiento

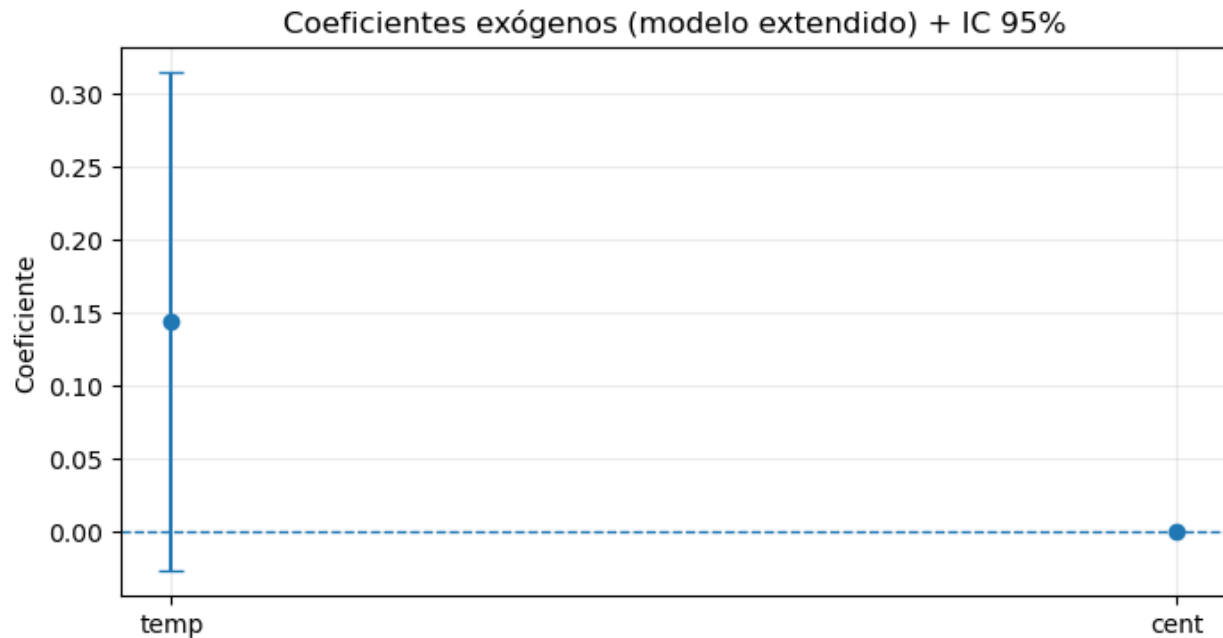
Ajustar un modelo ARIMAX para la Demanda (Ener\_1), usando como variables exógenas la Temperatura y la Centralidad del Nodo de origen. ¿Mejora el AIC del modelo al incluir la importancia del nodo en el grafo?

#### Metodología

Se ajustaron dos modelos ARIMAX sobre la serie de Demanda (Ener\_1): un modelo base con Temperatura (Ener\_3) como única variable exógena, y un modelo extendido que además incluye la Betweenness Centrality del nodo de origen. Se realizó búsqueda de hiperparámetros (p, d, q) mediante grid search minimizando el AIC en el modelo base, y luego se aplicó la misma estructura ARIMA a ambos modelos para comparación justa.

#### Resultados Comparativos





**Delta AIC = 2.039** (modelo extendido es peor que el base, ya que valores menores de AIC indican mejor ajuste)

### Interpretación de Resultados

El modelo ARIMAX extendido que incluye la centralidad de Betweenness del nodo de origen presenta un AIC mayor (8567.347) que el modelo base con solo temperatura (8565.308). Dado que el criterio AIC penaliza la complejidad del modelo y premia el ajuste, un incremento de 2.039 unidades indica que la centralidad del nodo no aporta suficiente poder predictivo para justificar el parámetro adicional.

### Posibles Explicaciones

1. **Baja variabilidad en centralidad:** La Betweenness Centrality calculada resultó aproximadamente 0 para todos los nodos, indicando una topología de red relativamente homogénea sin nodos significativamente más críticos que otros.
2. **Independencia de la demanda respecto a topología:** La demanda energética puede estar más relacionada con factores ambientales (temperatura) y temporales que con la posición estructural del nodo en la red.

3. **Métrica alternativa necesaria:** Otras métricas de centralidad o características del nodo podrían ser más relevantes para predecir demanda. Es posible que el betweenness simplemente no haya capturado diferencias estructurales significativas en el grafo.

## Conclusión Final

No, el AIC no mejora al incluir la centralidad del nodo. Al contrario, empeora ligeramente. El modelo base con solo temperatura como variable exógena es preferible por su simplicidad y levemente mejor ajuste. Se recomienda explorar otras características estructurales o temporales si se requiere mejorar la capacidad predictiva.

## CONCLUSIONES GENERALES

Este análisis integral ha permitido generar insights accionables en cuatro áreas críticas:

1. **Causalidad Energética:** La relación causal identificada entre factor de potencia y voltaje, combinada con el análisis de centralidad de red, proporciona una base sólida para implementar sistemas de monitoreo predictivo y redundancia en nodos críticos.
2. **Optimización Agrícola:** La identificación de 186 registros en zonas críticas con bajo NDVI y alta variabilidad de viento permite focalizar inversiones en infraestructura hídrica donde tendrán mayor impacto.
3. **Modelado Predictivo:** El modelo ARIMAX optimizado demuestra que la parsimonia es preferible cuando variables adicionales no aportan poder predictivo significativo. La temperatura se confirma como el principal driver de demanda energética.
4. **Metodología Robusta:** El uso combinado de técnicas de series temporales, teoría de grafos, filtrado de señales y análisis espacial proporciona un framework reproducible para análisis similares en otros contextos.