**Predicción de Picos de Demanda Eléctrica en Argentina mediante Ciencia de Datos**

*Trabajo Práctico – Ciencia de Datos (UADE, Junio 2025)*

*Grupo 5: Martina Degui, Jazmín Martínez, Ramiro Fiol, Dolores Núñez Trasar, Pedro Stella*

# Introducción

* **Contexto del problema:** En el mercado eléctrico argentino, anticipar correctamente los momentos de mayor consumo de energía es clave para la estabilidad del sistema. Evitar picos imprevistos ayuda a prevenir cortes de luz y optimizar los costos de generación.
* **Importancia:** La toma de decisiones basada en datos permite a las áreas técnicas y comerciales planificar con precisión, prevenir desvíos y mejorar la eficiencia operativa. CAMMESA (operador del Mercado Eléctrico Mayorista) busca herramientas analíticas para pronosticar la demanda y así garantizar un suministro confiable al menor costo posible.

# Objetivos e Hipótesis

* **Objetivo General:** Desarrollar una solución de *Data Science* que permita predecir los picos de demanda eléctrica en Argentina, combinando análisis exploratorio, modelos predictivos y visualizaciones interactivas para apoyar las decisiones de la Gerencia Comercial y Técnica.
* **Hipótesis:** *“Los picos de demanda energética no son aleatorios, sino que siguen patrones estacionales o históricos que pueden anticiparse con modelos adecuados”* . Esto implica que, analizando datos históricos de consumo, seremos capaces de identificar tendencias repetitivas (por ejemplo, aumentos en verano e invierno) y predecir con antelación cuándo ocurrirán los próximos picos de demanda.

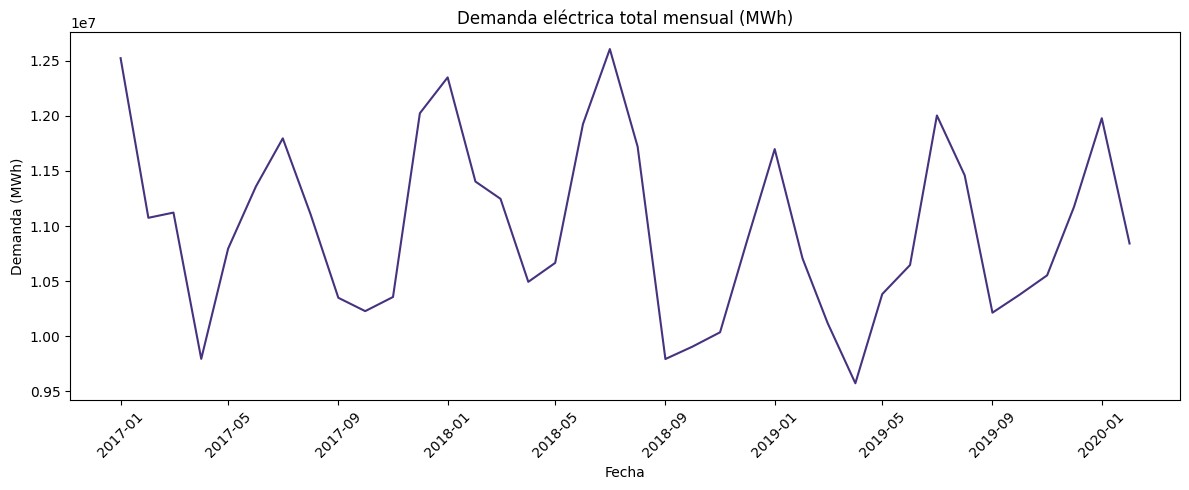
# Datos Utilizados

* **Fuente de Datos:** Se utilizaron datos históricos reales de demanda eléctrica mensual en Argentina, provistos por CAMMESA (Mercado Eléctrico Mayorista). Incluyen la demanda de energía en MWh por mes y por región del país en los últimos años.
* **Alcance Temporal:** El dataset abarca varios años recientes (e.g. 2017–2020), con frecuencia mensual. Esto nos permitió analizar patrones a lo largo de múltiples ciclos estacionales (veranos e inviernos).
* **Preprocesamiento:** Se realizó limpieza de datos (p. ej., manejo de caracteres especiales en encabezados) y se agregaron columnas de fecha a partir de los índices de tiempo (año-mes) para facilitar el análisis temporal. Asimismo, se agregaron datos de *balance* energético para enriquecer el análisis (por ejemplo, capacidad de generación vs. demanda), aunque el foco principal estuvo en la demanda.

# Metodología de Trabajo

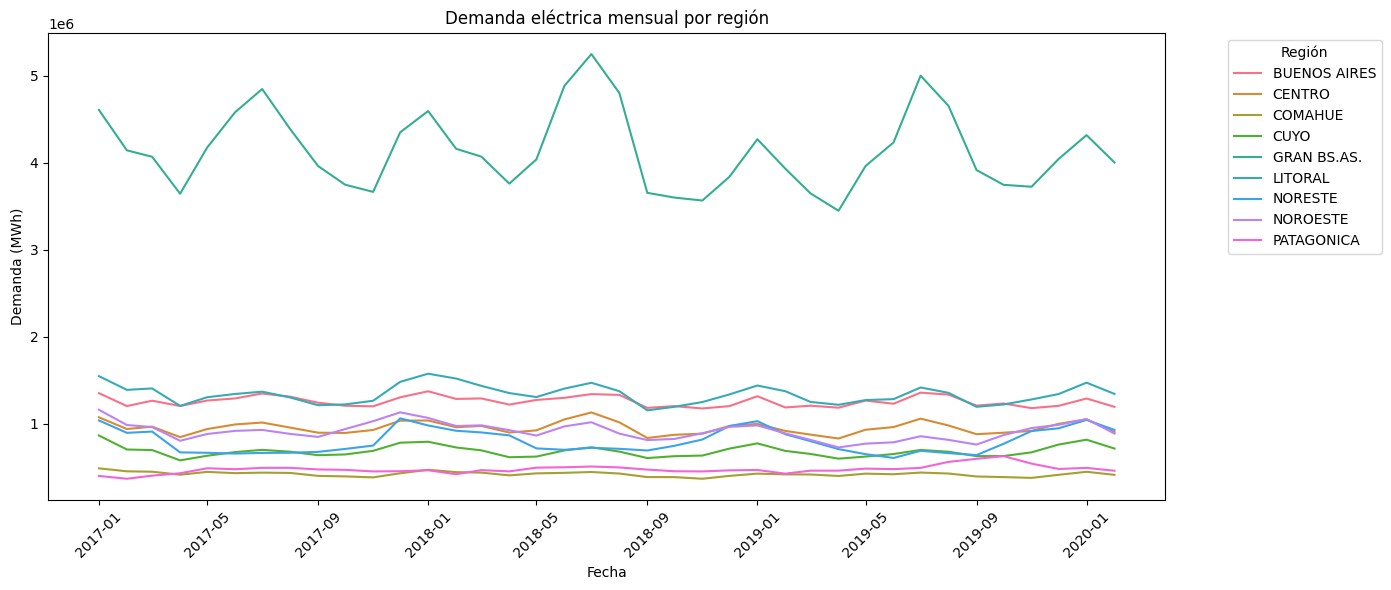
* **Enfoque Ágil (Scrum):** El proyecto se organizó con una metodología ágil. Definimos roles dentro del equipo: un *Product Owner* (definió la visión y la hipótesis), una *Scrum Master* (coordinó las tareas y el cronograma) y roles técnicos (Data Engineers y Data Scientist). Esto facilitó la colaboración en un grupo de cinco integrantes.
* **Cronograma:** El desarrollo se dividió en 5 sprints semanales:
* *Sprint 1:* Definición del dominio, formulación de hipótesis y recopilación de fuentes de datos.
* *Sprint 2:* Limpieza de datos, análisis exploratorio inicial y primeras visualizaciones.
* *Sprint 3:* Implementación del modelo de pronóstico (Facebook Prophet) y generación de predicciones.
* *Sprint 4:* Análisis por región y segmentación mediante clustering (K-Means).
* *Sprint 5:* Integración de hallazgos, storytelling y preparación de la presentación final.
* **Herramientas:** Se utilizó Python en Google Colab para el análisis, con librerías de ciencia de datos (*pandas, NumPy, matplotlib, seaborn, plotly, Prophet, scikit-learn*). La visualización interactiva se logró con Plotly y se generaron gráficos estáticos para el informe.

# Análisis Exploratorio – Tendencia General de la Demanda



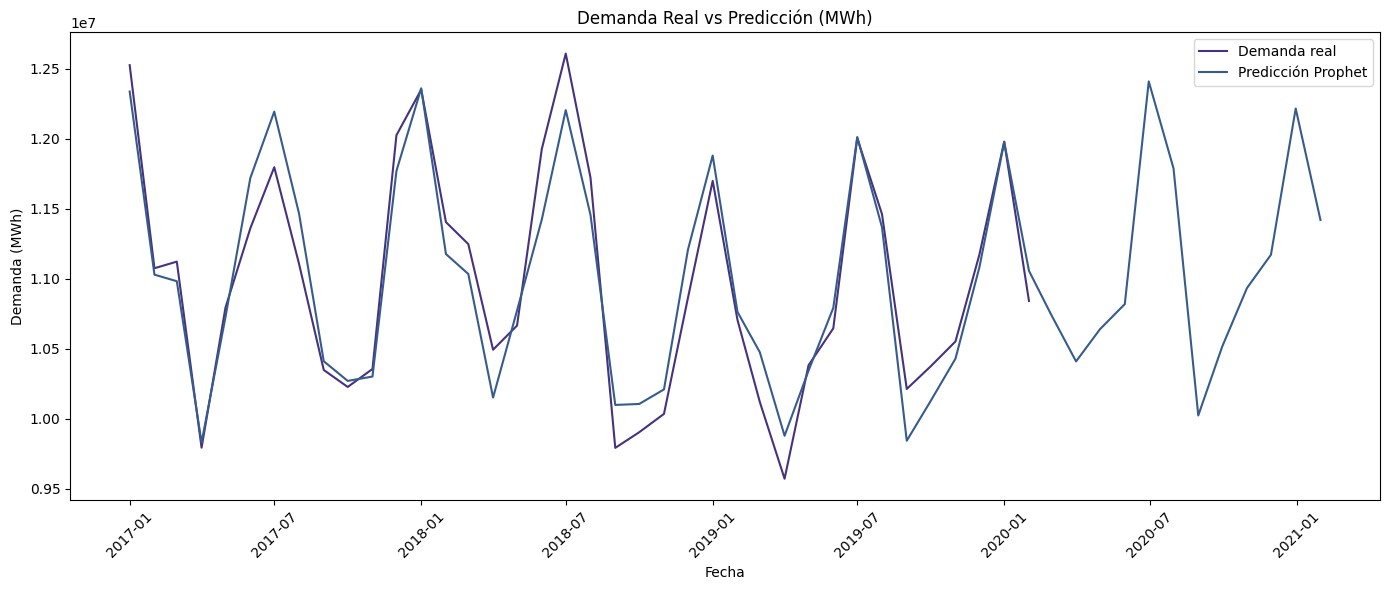
*Gráfico 1: Evolución de la demanda eléctrica total mensual en Argentina (MWh). Se observan claros patrones estacionales:* los meses de invierno y verano presentan picos de consumo más elevados, mientras que en primavera y otoño la demanda es menor. En el gráfico, por ejemplo, se notan máximos recurrentes alrededor de julio-agosto (invierno austral) y también aumentos en enero (pleno verano), seguidos de descensos posteriores. Estos ciclos anuales evidencian que la demanda eléctrica tiene un comportamiento periódico, probablemente asociado al uso de calefacción en invierno y aire acondicionado en verano. Además, a lo largo de los años analizados se aprecia una tendencia general ligeramente ascendente en el consumo máximo, indicando un crecimiento sostenido de la demanda energética en el país.

# Análisis Exploratorio – Demanda por Región



*Gráfico 2: Demanda eléctrica mensual desagregada por región (MWh). Cada línea de color representa una región del país.* Se advierte que la región del **Gran Buenos Aires** (línea verde, que sobresale por encima de 4–5×10^6 MWh) lidera el consumo en todo el período, reflejando su alta concentración poblacional e industrial. Otras regiones (Buenos Aires, Centro, Litoral, etc.) muestran demandas menores (entre 0.5 y 1.5×10^6 MWh) pero siguen patrones estacionales similares: todas presentan picos en los mismos meses de invierno y verano. Esto sugiere que si bien el nivel absoluto de demanda difiere por región (por factores como clima, población y actividad económica), **el fenómeno de picos estacionales es coherente a nivel nacional**. Esta sincronía en los picos regionales refuerza la idea de que existen factores estacionales comunes (temperatura, estacionales) impulsando la demanda en todo el país.

# Modelo Predictivo – Forecast con Prophet



*Gráfico 3: Comparación entre la demanda real histórica y la predicción del modelo Prophet.* La línea violeta representa la demanda real observada mes a mes, y la línea azul muestra la predicción generada por Prophet . Podemos ver que **ambas curvas siguen de cerca el mismo patrón**, indicando que el modelo logró capturar adecuadamente tanto la tendencia general como las variaciones estacionales de la serie. Para entrenar este modelo se usaron exclusivamente los datos históricos de demanda total mensual, sin variables externas. Prophet proyectó la demanda 12 meses hacia el futuro, mostrando picos pronosticados en las mismas épocas del año que los picos históricos (lo cual concuerda con la estacionalidad detectada). Las pocas desviaciones entre la predicción y la realidad son pequeñas y ocurren en algunos picos muy pronunciados, lo que sugiere oportunidades de mejora (por ejemplo, incluyendo variables como temperatura extrema, eventos atípicos, etc.). No obstante, **la similitud entre las líneas real y predicha sugiere un buen ajuste del modelo**, suficiente para fines de planificación.

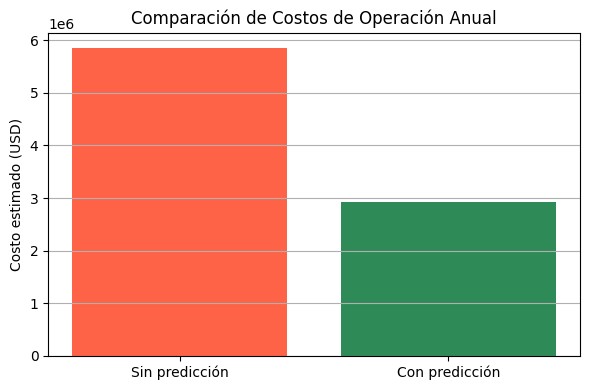
# Desempeño del Modelo (Métricas de Precisión)

* **Error Absoluto Medio (MAE):** 193.166 MWh por mes [10](file://file-1Rn5EiqfUsPX56gQC8wCfZ#:~:text=,1%2C6/) . En promedio, la predicción mensual difiere del valor real en unos 193 mil MWh. Dado que la demanda mensual ronda entre 10 y 12 millones de MWh, este error representa apenas 1.6% del total [10](file://file-1Rn5EiqfUsPX56gQC8wCfZ#:~:text=,1%2C6/) – un nivel de error bastante aceptable para planeamiento agregado.
* **Raíz del Error Cuadrático Medio (RMSE):** 235.256 MWh . Es ligeramente mayor que el MAE, lo que indica que no hubo errores extremos muy fuera de lo común (de lo contrario el RMSE crecería mucho más respecto al MAE). Esto sugiere que el modelo es consistente y sus errores están distribuidos de forma relativamente uniforme sin grandes outliers.
* **Coeficiente de Determinación (R²):** 0.9139 (91.4%). Implica que el modelo explica alrededor del **91% de la variabilidad** de la demanda mensual histórica, lo cual denota un ajuste muy bueno para un modelo de esta naturaleza. Un R² tan alto confirma que Prophet, pese a su simplicidad, logró modelar eficazmente el comportamiento de la serie temporal (tendencias + estacionalidades).
* *Interpretación:* Estas métricas validan la capacidad predictiva del modelo: un error relativo 1.6% significa que podemos anticipar la demanda con un alto grado de precisión, suficiente para tomar decisiones operativas (como preparar oferta de generación). Aún así, queda 9% de variabilidad sin explicar, atribuible a factores externos no incluidos en el modelo (por ejemplo, condiciones climáticas extremas, variaciones económicas, feriados largos, etc.), lo que abre posibilidad a mejorar incorporando más variables en futuros modelos.

# Análisis de Clustering – Segmentación de Meses por Perfil de Consumo

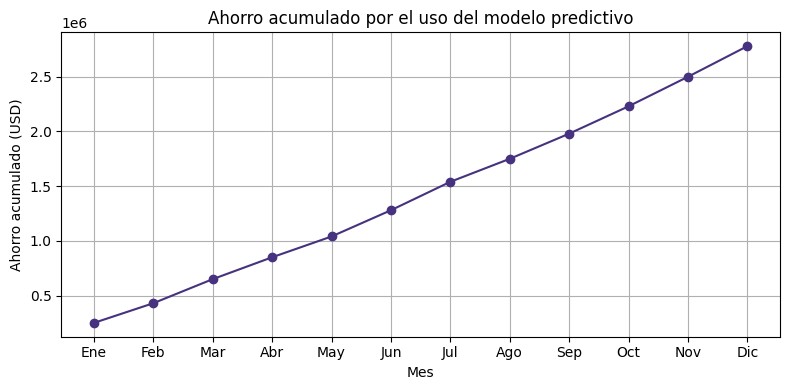
* **Motivación:** Aplicamos *clustering* (K-Means) para ver si los meses pueden agruparse según similitudes en sus niveles de demanda regional. Esto busca descubrir patrones no supervisados que validen la hipótesis de estacionalidad desde otra perspectiva.
* **Procedimiento:** Se preparó una matriz con los meses como filas y la demanda de cada región como columnas. Tras escalar los datos (normalización), se ejecutó K-Means con *k=3* clusters, buscando clasificar meses en categorías de **baja**, **media** y **alta** demanda.
* **Resultados:** El algoritmo efectivamente agrupó los meses por perfil: por ejemplo, **los meses de demanda muy alta (cluster “picos”) correspondieron principalmente a julio y agosto** (invierno), mientras que **los meses de baja demanda cayeron en otro cluster** distinto (generalmente meses templados de primavera/otoño). Un tercer cluster representó demanda intermedia. Estos grupos muestran patrones estacionales claros: prácticamente todos los julios y agostos quedaron juntos en el cluster de máxima demanda, confirmando que cada año esos meses son similares entre sí en su alto consumo.
* **Conclusión del Clustering:** La segmentación no supervisada respaldó nuestra hipótesis. Encontramos un cluster específico para los picos estacionales, lo que evidencia que **existe un comportamiento recurrente anual**. Esto significa que conociendo en qué cluster cae un mes (según sus características de consumo regional), podemos inferir su nivel de demanda esperado. En resumen, **los picos de consumo pueden anticiparse** porque los meses críticos comparten atributos y se repiten cada año, en línea con patrones estacionales.

# Análisis Económico – Beneficio de la Predicción



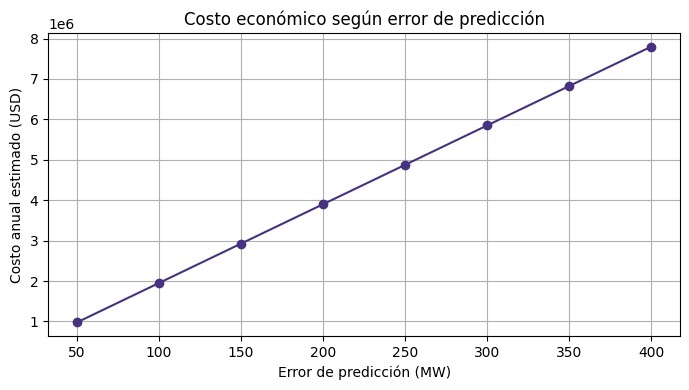
*Gráfico 4: Comparación de costos de operación anual con y sin modelo predictivo.* Se modeló un escenario simplificado de costos para CAMMESA: suponemos un costo de generación eficiente de 70 USD/MWh (p.ej., usando hidroeléctrica) vs. un costo de generación de respaldo de 200 USD/MWh (p.ej., usando centrales térmicas costosas). Cuando hay errores en la predicción de demanda, el sistema puede incurrir en costos extra al tener que echar mano de generación de emergencia (más cara) o comprar energía de último momento. En el caso **“Sin predicción”**, asumimos un error promedio de 300 MW no pronosticados, ocurriendo durante 150 horas al año; esto implicaría costos adicionales de alrededor de **5.85 millones de USD** anuales. En cambio, en el escenario **“Con predicción”** (usando nuestro modelo), el error se reduce a 150 MW (la mitad), con el mismo número de horas de desvío. El costo anual estimado baja a unos **2.93 millones de USD** . La diferencia es un **ahorro de aproximadamente 2.9 millones de USD por año** gracias al modelo. En el gráfico, la barra roja representa el costo sin predicción y la verde con predicción: la reducción del 50% en el error se traduce en una reducción del 50% en costos operativos. **Esto evidencia el impacto económico directo de mejorar el pronóstico de demanda.**

# Análisis Económico – Ahorro Acumulado Mensual



*Gráfico 5: Ahorro acumulado mes a mes al utilizar el modelo predictivo.* Este gráfico ilustra cómo se van acumulando los ahorros a lo largo de un año si CAMMESA implementa la herramienta de predicción. Comenzando en enero, cada mes se evitarían costos (por comprar o generar energía de más) del orden de 180k–270k USD, sumando progresivamente. Para mitad de año ya se habría ahorrado alrededor de 1 millón de USD, y la proyección cierra en **2.8 millones de USD al final de diciembre** . La línea morada asciende casi linealmente, indicando que **mientras antes se comience a usar el modelo en el año, mayor será el beneficio acumulado**. Este resultado cuantifica el valor financiero de nuestra solución: más allá de la estabilidad del sistema, hay un incentivo económico claro para mejorar la previsión (millones en ahorro anual). El mensaje a la empresa es que invertir en analítica predictiva se justifica por su **retorno económico tangible**.

# Sensibilidad: Error de Predicción vs. Costo Operativo



*Gráfico 6: Relación entre el error de pronóstico y el costo anual estimado.* Aquí analizamos distintos escenarios de error medio en la demanda no pronosticada (en MW) y su impacto en costos adicionales. Observamos una **relación prácticamente lineal:** por cada incremento de 50 MW en el error de predicción, el costo anual aumenta en aproximadamente **0.98 millones de USD** . Por ejemplo, un error de 400 MW implicaría cerca de 7.8 MM USD en costos (activando mucha generación cara), mientras que un error de apenas 50 MW costaría ~0.98 MM USD. Esto subraya que **mejorar la precisión del modelo tiene un efecto directo en reducir costos**: incluso reducir unos pocos megawatts de error puede ahorrar cientos de miles de dólares. Desde la perspectiva de CAMMESA, vale la pena refinar los modelos predictivos (o incorporar más datos como clima, economía) ya que cada punto porcentual de mejora en la precisión se traduce en dinero ahorrado.

• **Costo de Generación por Fuente:** Cabe destacar la diferencia de costos entre tipos de generación, lo cual explica la pendiente de la relación anterior. Fuentes como hidroeléctrica o gas natural tienen costos unitarios bajos (~55–70 USD/MWh), mientras que combustibles de respaldo como fuel oil o gasoil pueden costar **hasta 200 USD/MWh**. Esta brecha de casi 3x en el precio es la razón por la que ante faltantes imprevistos (errores de predicción) el operador debe recurrir a energía muy costosa. Nuestro modelo ayuda a minimizar el uso de esas fuentes caras, favoreciendo una planificación con mayor porcentaje de energía barata (eficiente) y menor dependencia de picos de generación de emergencia.

# Conclusiones

* **Validación de la Hipótesis:** Los análisis confirmaron que la demanda eléctrica presenta patrones estacionales pronunciados. Identificamos picos recurrentes en invierno y verano cada año, lo que demuestra que *sí es posible anticipar los picos de consumo* observando el historial.

El clustering también corroboró estos grupos estacionales de alta demanda. En síntesis, la hipótesis se cumple: los picos no son eventos aleatorios, sino que siguen un patrón periódico predecible.

* **Eficacia del Modelo Predictivo:** La herramienta desarrollada con Prophet resultó efectiva para pronosticar la demanda con alta precisión (R² 0.91). Capturó tendencia y estacionalidad de la serie, alcanzando errores medios bajos (1.6% del total). Esto significa que, aún con un modelo relativamente simple y datos limitados (solo históricos de demanda), pudimos **explicar la mayor parte de la variación** y obtener predicciones útiles para la toma de decisiones. Se sugiere que incorporando variables externas (temperatura, actividad económica, etc.) se podría mejorar aún más el desempeño.
* **Impacto Económico:** Más allá de las métricas técnicas, cuantificamos un beneficio económico significativo. Con nuestro modelo, CAMMESA podría ahorrar del orden de **2.8 millones de USD por año** en costos de operación, al reducir la sobrecontratación de energía de respaldo costosa. Visualizamos cómo este ahorro se acumula mes a mes, evidenciando que la predicción aporta valor financiero tangible, no solo precisión técnica. Además, analizamos cómo el error de pronóstico impacta linealmente en los costos: cada mejora en la precisión reduce gastos operativos, liberando recursos que la empresa puede destinar a otras inversiones.
* **Recomendaciones:** Se recomienda implementar esta solución de pronóstico en la planificación energética de CAMMESA y otras empresas del sector. Herramientas de ciencia de datos como ésta permiten una **planificación proactiva**: anticipar picos de demanda con tiempo, optimizar la programación de generación, negociar contratos de abastecimiento con mejor información y evitar compras de energía de última hora a precios elevados. También se sugiere mantener un proceso de mejora continua del modelo (agregar nuevas fuentes de datos, retraining frecuente) para adaptarse a cambios en los patrones de consumo.
* **Conclusión Final:** El proyecto demostró el poder del análisis de datos en un entorno real de negocio. Con datos históricos convertidos en conocimiento accionable, proporcionamos una herramienta que mejora la **seguridad del suministro eléctrico**, **reduce costos operativos** y apoya la toma de decisiones estratégicas basadas en evidencia. Esto genera valor tanto técnico como financiero, validando la inversión en ciencia de datos dentro del sector energético.

# Preguntas & Gracias

*¿Alguna pregunta? ¡Gracias por su atención!*