



O uso da infraestrutura avançada de medição para transformação da CEMIG em uma empresa data-driven

Tema: Tecnologia da Informação

Autores: Natalie Costa Araujo

Co-Autores: Flavio Henrique Martins Vieira

Empresa: CEMIG Distribuição S.A

Resumo

Neste artigo, examinaremos o processo de transformação da CEMIG para uma abordagem centrada em dados. A partir de um exemplo prático, abordaremos as etapas, motivadores, desafios e decisões tomadas para converter os dados da infraestrutura avançada de medição (*AMI – Advanced Metering Infrastructure*) em descobertas importantes para a concessionária. Esse esforço incluiu definição de infraestrutura robusta, redefinição de processos, uso de novas tecnologias e desenvolvimento de novas habilidades. Os resultados obtidos reforçam como os benefícios do AMI podem ir muito além dos operacionais e como a criação de uma estrutura fundamentada em dados poderá contribuir para decisões mais objetivas e confiáveis.

1. Introdução

A implementação do AMI na CEMIG resultou em economia e melhorias operacionais para a concessionária e diversos benefícios para os mais de 370 mil clientes contemplados no projeto.

No cenário vigente, avançar no âmbito de análise de dados representa um grande potencial a ser explorado no AMI. Com uma média diária superior a 40 milhões de registros ao dia, a exploração dos dados exigiu a definição de uma estrutura robusta, escalável e flexível, sem a necessidade de grandes investimentos iniciais. Outro ponto fundamental foi a adoção de um modelo descentralizado que permitisse às áreas de negócio executarem suas análises com autonomia e agilidade.

Demonstramos, através de um exemplo prático, como a continuidade, detalhes e abrangência dos dados do AMI puderam contribuir para a aferição dos cálculos estatísticos de demanda em um transformador. Lidar com esse volume de dados demandou a adoção de ferramentas, tecnologias especializadas e habilidades específicas. Nesse estudo, tivemos a oportunidade de executar as principais etapas necessárias em processos de análises descritivas e diagnósticas. Os resultados destacaram a oportunidade de revisões e ajustes nos cálculos vigentes, além de potencial para avanços futuros em análises preditivas e prescritivas.

2. Desenvolvimento

2.1. Coleta de dados do AMI

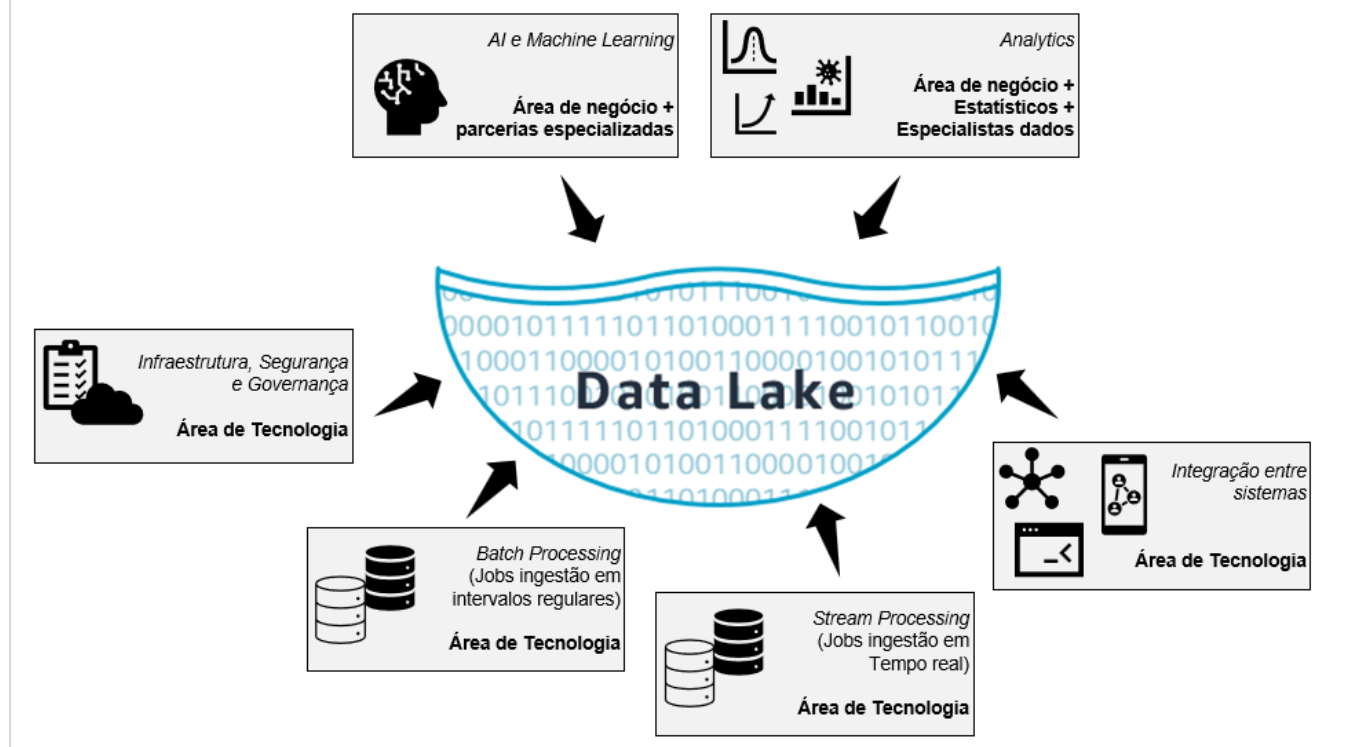
A CEMIG possui AMI implantado em mais de 370 mil clientes e uma expectativa de expansão de 1,25 milhão até o final de 2027. Os principais dados coletados são memória de massa, eventos e totalizadores diários. Entre eles, a memória de massa é a que mais contribui para o volume de dados coletados, com uma média diária superior a 35 milhões de registros. Em segundo lugar, os eventos representam 6 milhões de registros ao dia.

Esse volume de dados pode ser avassalador se não for gerenciado adequadamente. Nesse contexto, era primordial que a solução adotada pela CEMIG oferecesse recursos de armazenamento, processamento e escalabilidade de maneira econômica e flexível. A utilização de um *data lake* em nuvem pública se revelou com uma alternativa que permitiria ajustar rapidamente os recursos conforme a demanda, sem a necessidade de grandes investimentos iniciais. Além disso, a ampla gama de serviços e ferramentas disponíveis na nuvem possibilitaria alto desempenho no tratamento e entrega dos dados, sem impactos para as operações transacionais de alta prioridade nos sistemas do AMI.

2.2. Descentralização da análise de dados

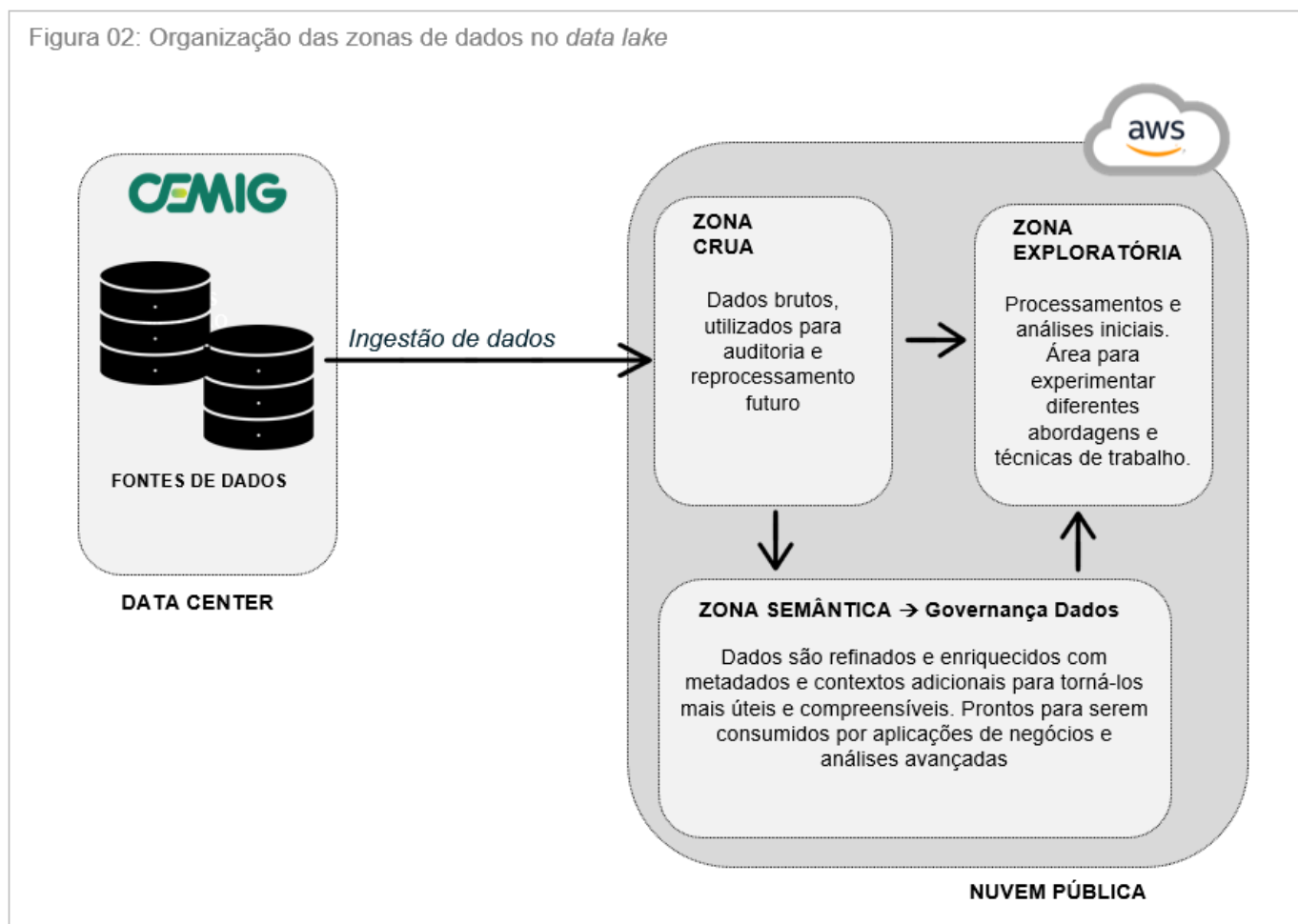
Organizar volumes grandes de dados dispersos não é algo trivial. A adoção de um modelo descentralizado e acessível permite que as análises sejam realizadas com mais velocidade, entregando a tão desejada rapidez na tomada de decisão. Outra vantagem é a autonomia para a experimentação. Ambientes em que as áreas de negócio podem experimentar os dados sem afetar os sistemas de produção, permitem uma exploração segura, viabiliza descobertas e promove a inovação dentro da organização. Além disso, a análise feita pelas áreas de negócio tende a ser mais alinhada com as necessidades e prioridades do que as realizadas por departamentos de TI, resultando em insights que são diretamente aplicáveis e acionáveis.

Figura 01: Descentralização da análise de dados



No data lake corporativo da CEMIG, os dados do AMI foram organizados em diferentes zonas, para melhorar a gestão, processamento e acesso. A zona crua é onde os dados estão armazenados em seu formato original, sem qualquer transformação ou processamento. A zona crua permite que diferentes equipes acessem os dados brutos para diversas finalidades. A zona exploratória é onde os usuários podem

manipular e explorar os dados de forma flexível e independente. Permite que os usuários façam ajustes rápidos em suas análises e modelos. Se uma hipótese inicial não se confirmar, eles podem rapidamente modificar suas abordagens e testar novas hipóteses. Ao final, as experimentações e análises irão resultar em dados tratados, acessíveis e com significado, disponibilizados na zona semântica.



2.3. Análises básicas e evolução gradual

Quando se trata de análise de dados, é comum haver grandes expectativas em relação ao uso imediato de tecnologias avançadas como ML (*Machine Learning*) e IA (Inteligência Artificial). No entanto, a adoção precipitada dessas tecnologias pode levar a investimentos elevados e resultados insatisfatórios. A CEMIG optou por começar com análises básicas e evoluir gradualmente. Essa abordagem reduz riscos e custos, promove o desenvolvimento de competências e habilidades das equipes e, principalmente, permite a construção de uma base sólida de compreensão dos dados disponíveis. ML e IA dependem de dados de alta qualidade. Dados inadequados ou tendenciosos podem levar a resultados imprecisos ou enganosos.

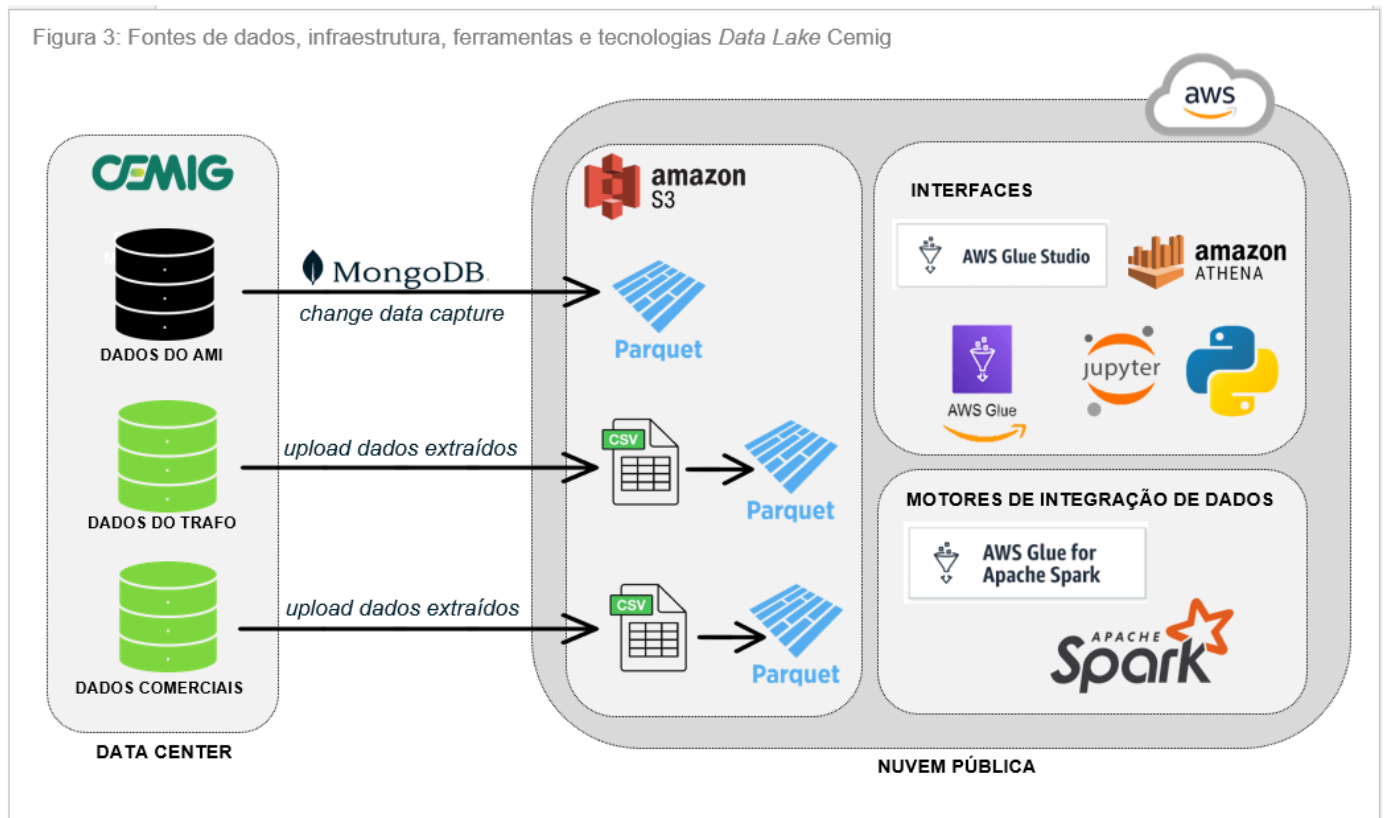
2.4. Data Lake: infraestrutura, fontes de dados, tecnologias e ferramentas

Manipular o grande volume de dados do AMI pode ser um grande desafio para os usuários das áreas de negócio. Computadores pessoais geralmente não tem a capacidade de memória e processamento necessárias, e ferramentas comuns de análise não são adequadas para lidar com grandes cargas de trabalho, resultando em lentidão ou até falhas. O armazenamento e a manipulação desses dados, exigiu a adoção e uso de ferramentas e tecnologias especializadas.

- **Fontes de dados:** Os dados do AMI são automaticamente extraídos do MongoDB por *jobs* e disponibilizados na zona crua. Os dados das outras fontes utilizadas nesse estudo foram carregadas através de arquivos CSV.

- **Armazenamento:** Utilizamos o *Amazon S3 (Simple Storage Service)*. Os dados extraídos e sumarizados foram convertidos para *Parquet*, o qual permite armazenamento eficiente de dados em colunas, otimiza a compactação e é ideal para consultas analíticas em grandes volumes de dados.
- **Interface:** Utilizamos *Notebooks* do *AWS Glue* para a codificação dos scripts de extração, transformação e carga dos dados na zona exploratória. Também foi utilizado o *Amazon Athena*, para executar consultas SQL diretamente nos dados armazenados no S3.
- **Motores de integração de dados:** Nesse estudo, foi adotado o *AWS Glue form Apache Spark*

Figura 3: Fontes de dados, infraestrutura, ferramentas e tecnologias *Data Lake Cemig*



2.5. Exemplo prático: Aferição dos cálculos estatísticos de demanda no tráfego a partir das medições do AMI

Nesse artigo, iremos explorar um exemplo prático para ilustrar o processo de análise de dados do AMI na CEMIG.

Objetivos e hipóteses

As informações de carregamento do tráfego são informações dinâmicas calculadas pelos Sistemas de Informação Geográfica (SIG) da CEMIG. Para o cálculo, são utilizadas informações técnicas do tráfego, curvas típicas e histórico de consumo do cliente. Nesse contexto, foram elaboradas as seguintes hipóteses:

- Hipótese 1: Os dados do AMI fornecem curvas típicas mais precisas do que as campanhas de medidas, devido à sua continuidade e abrangência de clientes.
- Hipótese 2: Os percentuais de carregamento calculados pelo SIG podem ser aferidos com precisão pelas medições dos transformadores que atendem exclusivamente clientes com medidores AMI, devido às informações detalhadas de consumo.

Coleta e tratamento dos dados

O ponto ótimo de volume de dados para uma análise exploratória depende de vários fatores. Nesse estudo, optamos pelo uso de um subconjunto de dados do AMI que mantivesse a representatividade sem comprometer o desempenho. O critério utilizado foi a extração de um mês completo de dados (abril/2024) de todos os medidores inteligentes, resultando num total superior a 1 bilhão de registros.

A primeira etapa envolveu a normalização dos dados. Essa etapa foi essencial para compreensão das unidades de medida utilizadas, do comportamento do medidor no registro das informações e dos métodos corretos para conversão das granularidades. A visualização gráfica das informações foi fundamental para a entendimento, identificação de erros e validação da qualidade dos dados. Os *outliers* (valores que desviam significativamente da tendência geral dos outros valores) podem distorcer e comprometer os resultados da análise. Sua identificação é um processo que exige julgamento e conhecimento do contexto, pois embora extremos, ainda assim podem ser válidos e representativos para o estudo. Nesse artigo foram considerados para a detecção de *outliers*:

- Modelagem temporal: as análises foram segmentadas por tempo, para respeitar o padrão horário de consumo e demais grandezas.
- Dados incompletos ou anômalos: interrupções ou oscilações de tensão podem levar a lacunas de dados ou valores anômalos que podem ser erroneamente identificados como outliers. Para mitigar esse risco, incorporamos às análises os dados referentes aos alertas de qualidade e falta de energia.
- Variações temporárias no consumo: A transição de um imóvel de ocupado para desocupado (ou vice-versa) também pode criar picos ou quedas abruptas no consumo. A integração de dados comerciais sobre a situação do imóvel ajudou na interpretação dessas variações.
- Geração distribuída: Os medidores do AMI registram apenas a diferença entre o total consumido e o total gerado pelo cliente. Esse comportamento oculta os padrões de consumo/geração do cliente e impacta fortemente na identificação de padrões ou anomalias. Além disso, quando ocorre a inversão do fluxo (a direção do fluxo muda de entrada para saída ou vice-versa), a presença de canais zerados ou próximos de zero na memória de massa reduzem a média geral. Diante dos impactos e obstáculos do uso da geração distribuída nas análises, optamos por excluí-la desse estudo.

• **Análise dos Resultados**

O SIG utiliza nos cálculos estatísticos as curvas típicas e de fator de potência. O caráter adimensional de uma curva típica permite estudar a forma sem se preocupar com os diferentes patamares de consumo dos clientes. Os clientes AMI estão predominantemente distribuídos nas categorias de tarifa B1 e B3. As demais categorias não tinham quantitativos suficientes e por esse motivo foram descartadas. As análises segmentaram as curvas típicas em dia útil, sábado e domingo.

Quando comparamos os resultados obtidos a partir das medições do AMI com os resultados das curvas obtidas através das campanhas de medidas, percebemos formatos similares. Observamos que as curvas do AMI possuem arestas mais arredondadas, indicando que os dados são mais contínuos e que os padrões mudam gradualmente. Constatamos também que os horários de picos não coincidem. Em relação ao fator de potência, os cálculos estatísticos utilizam apenas dois patamares de valores, enquanto o AMI apresenta valores mais contínuos e discretamente menores aos picos da campanha de medições.

Curva típica cliente residencial AMI



Curva típica cliente comercial AMI



Curva típica cliente residencial SGI



Curva típica cliente comercial SGI

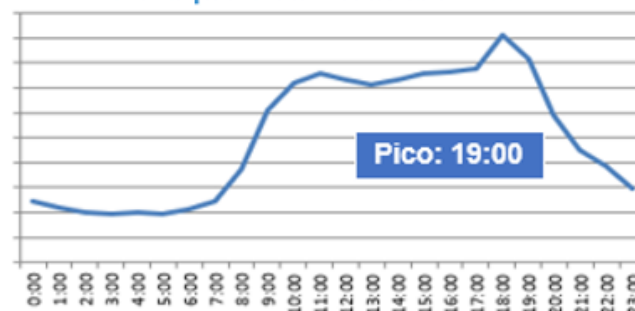


Figura 04: curvas típicas AMI x campanha medidas

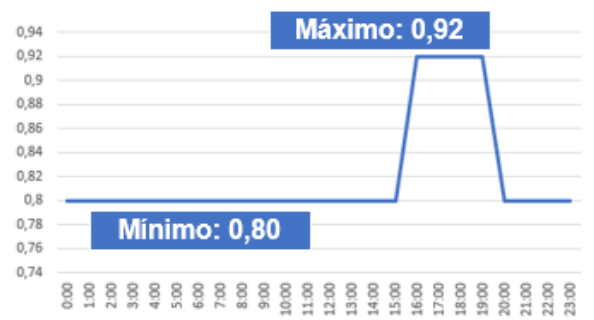
Fator de potência cliente residencial AMI



Fator de potência cliente comercial AMI



Fator de potência cliente residencial SGI



Fator de potência cliente comercial SGI

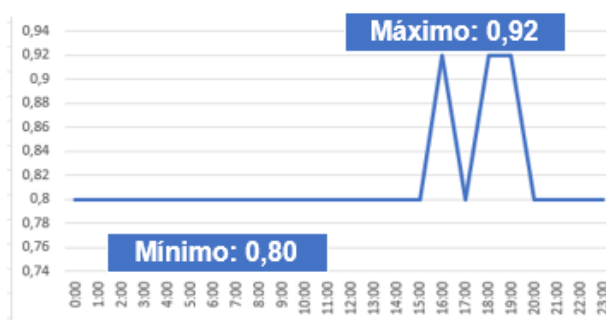


Figura 05: fator de potência AMI x campanha medidas

Embora os padrões de consumo e fator de potência resultantes tenham se apresentado satisfatórios, é importante ressaltar que a amostra de dados analisada do AMI apresenta uma predominância de características urbanas, o que pode limitar sua aplicação em regiões com características rurais. A segunda parte do estudo foi comparar as medições coletadas do AMI com os valores de demanda máxima calculados no SIG. Da mesma forma que as curvas típicas, as análises foram segmentadas em dia útil, sábado e domingo. Foram analisados 603 trafos cujos medidores são todos AMI. Para cada trafa, foram somadas as curvas diárias (mediana) de potência aparente dos respectivos medidores. A soma gráfica da curva foi confrontada com a potência nominal e a soma de demanda máxima do SIG.

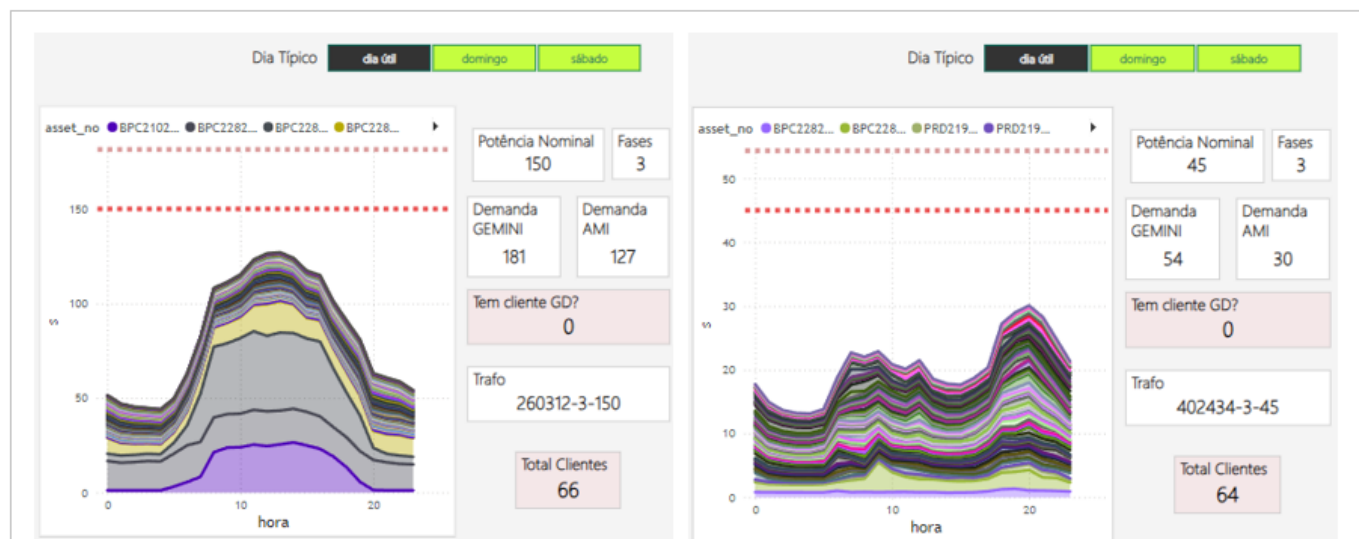


Figura 06: potência aparente total no trafa para trafos exclusivos AMI

Os resultados obtidos destacaram que análises individuais ou de grupos pequenos podem revelar variações significativas que não são aparentes em grandes grupos. Quando examinamos os transformadores cujo total de clientes é superior ou igual a 10, alcançamos resultados coerentes com os valores estimados. Porém, quando examinados os transformadores que possuem menos de 10 clientes, percebemos uma discrepância entre os cálculos estatísticos e as medições coletadas. Constatamos que para o cálculo de demanda em um transformador, as relações observadas para grupos não se aplicam, necessariamente, aos indivíduos dentro desse grupo. Por esse motivo, ao planejar modelos estatísticos é importante utilizar técnicas que considerem múltiplos níveis de dados (individual e coletivo), garantindo que as conclusões sejam válidas e aplicáveis ao nível correto.

Nas agregações superiores a 10 clientes por transformador, verificamos que as medições observadas estavam predominantemente abaixo das previsões do modelo estatístico do SIG. Em alguns contextos é aceitável e até estratégico superestimar valores como uma margem de segurança. Nas concessionárias, previsões conservadoras podem mitigar potenciais falhas. Porém em alguns casos, essa tática pode resultar em desperdício de recursos ou ineficiências operacionais.

Revisões são parte natural da evolução de um modelo estatístico e ajudam a garantir que ele continue atendendo aos objetivos à medida que os dados e as condições mudam. O uso dos dados do AMI se demonstrou como uma alternativa extremamente valiosa para um possível processo de revisão. O cálculo estatístico do SIG poderia ser enriquecido com variáveis adicionais sobre variabilidade temporal, sazonalidade e características do cliente e medidor. O fluxo constante de dados do AMI permitiria que o modelo fosse validado e atualizado continuamente com dados recentes, refletindo melhor o comportamento atual dos consumidores e das condições de operação. Por fim, a alta frequência dos dados do AMI, combinadas com dados de eventos e interrupções, ajudaria a distinguir entre ruído e variações reais, contribuindo para dados mais limpos e significativos.

3. Conclusão

Implementar uma cultura onde decisões baseadas em dados são incentivadas e valorizadas requer tempo, investimentos e alinhamento nos diversos níveis da organização. Para lidar com o grande volume de dados

do AMI, a CEMIG fez investimentos em infraestrutura robusta e escalável, incluindo ferramentas de análise, plataformas de armazenamento e integração de sistemas.

Iniciar com projetos pilotos provou ser uma estratégia eficaz para lidar com os dados do AMI, permitindo validar ideias rapidamente e aprender com os resultados antes de expandir para um escopo maior. O estudo apresentado neste artigo nos possibilitou experimentar as principais etapas de um processo de análise e exploração de dados. Além disso, os resultados identificaram a oportunidade de futuros projetos para revisão e ajuste dos cálculos estatísticos de carga nos transformadores, com foco em melhorar a precisão e considerar múltiplos níveis de dados, garantindo que as conclusões sejam válidas e aplicáveis a diferentes contextos.

Garantir que os dados se tornem o centro das decisões estratégicas envolve superar desafios que vão além da tecnologia, abrangendo mudanças culturais, desenvolvimento de habilidades e adaptação de perfis profissionais. Como próximos passos, a CEMIG está empenhada em estabelecer parcerias estratégicas com empresas especializadas para explorar soluções inovadoras e impulsionar a capacitação de seus profissionais. Essas iniciativas visam garantir inovações no setor, além de proporcionar benefícios significativos para seus clientes e partes interessadas.

4. Referências bibliográficas

- Kinslow, J., Ghayouri, H., & Hoss, F. (2022, June 1). AMI – advanced analytics opportunities. *Utility Analytics*. <https://utilityanalytics.com/2022/06/ami-advanced-analytics-opportunities/>
- White, D. (2021, November 30). *5 challenges that get in the way of becoming data-driven*. Techfunnel. <https://www.techfunnel.com/information-technology/challenges-that-get-in-the-way-of-becoming-data-driven/>
- *From research to action*. (n.d.). Electric Energy Online. Retrieved November 6, 2024, from <https://electricenergyonline.com/energy/magazine/917/article/from-research-to-action-utility-data-analytics-project-makes-molehills-out-of-mountains.htm>
- Wambaugh, O. *The Value of AMI - It's so much more than billing*. (n.d.). Electric Energy Online. Retrieved November 6, 2024, from <https://electricenergyonline.com/energy/magazine/727/article/the-value-of-ami.htm>
- *Acting on the deluge of newly created automation data: Using big data technology and analytics to solve real problems*. (n.d.). Electric Energy Online. Retrieved November 6, 2024, from <https://electricenergyonline.com/energy/magazine/684/article/acting-on-the-deluge-of-newly-created-automation-data-using-big-data-technology-and-analytics-to-solve-real-problems.htm>
- *4 types of data analytics to improve decision-making*. (2021, October 19). Business Insights Blog. <https://online.hbs.edu/blog/post/types-of-data-analysis>
- DAVIES, L. **What is the hierarchy of needs in data science?** Disponível em: <https://resources.noodle.com/articles/data-science-hierarchy-of-needs/>. Acesso em: 19 nov. 2024.
- BURLEY, J. *Practical AI: The Data Science hierarchy of needs*. Disponível em: <https://www.actif.ai/resource/practical-ai-the-data-science-hierarchy-of-needs/>. Acesso em: 19 nov. 2024.