

Projeto final de Ciência de Dados

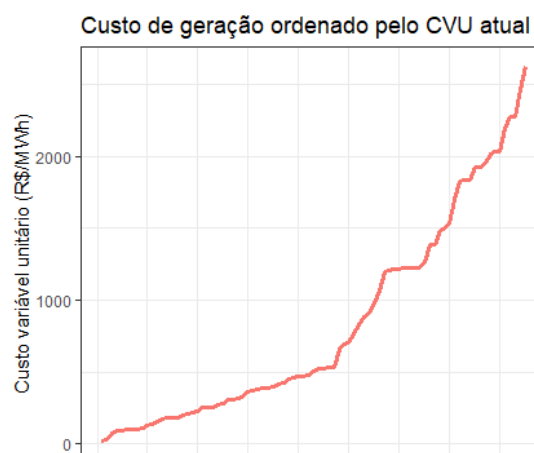
Tema: aumento de custo geração de energia
elétrica pelas usinas termelétricas

Justificativa

A escolha do tema ocorreu em razão do forte aumento no preço da energia elétrica em 2021. Esse aumento não pesa no bolso dos indivíduos apenas pelo aumento na conta de luz, mas também encarece o custo de muitos dos elos de diferentes cadeias produtivas de vários setores. Por exemplo, no caso dos alimentos, podemos citar os produtos que dependem de irrigação, o custo mais alto de manutenção da câmara fria de frigoríficos e até mesmo a geladeira dos supermercados. Já no caso da indústria, o aumento no custo da energia prejudica principalmente setores intensivos em energia, como metalurgia, com impacto indireto também em outros processos que utilizam esses produtos como insumos. Há até mesmo serviços que podem ser bastante afetados, como por salões de beleza (devido ao uso intensivo de secadores de cabelo) e outros em que o custo do ar-condicionado é relevante.

Apesar do desejo óbvio de correlacionar o custo da energia com a inflação de sentida pelo consumidor, essa não foi a abordagem escolhida devido às complexidades envolvidas. Em primeiro lugar, a energia elétrica é um “preço administrado”, ou seja, tem regras bem estabelecidas em contrato para reajuste de preços (periodicidade anual, aprovação pela Aneel são alguns exemplos). E diferentemente da gasolina, cuja política de preço da Petrobras tem sido a manutenção da paridade internacional, no caso da energia tem havido bastante interferência da Aneel para segurar os reajustes, desde o início da pandemia¹. Além disso, a volatilidade da bandeira tarifária atrapalha bastante a análise, porque apesar de haver um padrão esperado de sazonalidade, depende bastante de variáveis climáticas e varia muito de ano para ano.

A opção escolhida dada a limitação de tempo foi uma análise mais simplificada, com foco em duas razões para o aumento do custo de geração: o aumento da demanda de energia proveniente das termelétricas, em razão da escassez hídrica e o aumento do preço do petróleo, que leva a um aumento no preço de seus derivados, que constituem quase a totalidade dos combustíveis do parque termelétrico nacional no momento. O gráfico à direita apresenta o CVU (custo variável unitário) de todas as usinas térmicas que podem ser despachadas pelo ONS (Operador Nacional do Sistema), que pode ser visto como uma proxy do custo marginal de geração já que, em geral, as usinas mais caras tendem a ser acionadas apenas em caso de extrema necessidade². Como curiosidade, a usina térmica com maior CVU



¹ Em 2020, o reajuste da tarifa-base (desconsiderando o adicional de bandeira) de energia foi bastante baixo, sendo até em negativo em algumas localidades apesar do forte aumento do câmbio no período, que encarece o custo da energia de Itaipu (e também dos combustíveis das térmicas). Essa “manobra” foi possibilitada pelo empréstimo conhecido como “Conta-Covid”, que começou a ser pago em 2021. Em 2020, também não houve cobrança do adicional de bandeira até abril, “jogando a conta” para a frente. Em 2021, com a subida ainda maior do câmbio e também dos derivados de petróleo, o esforço da Aneel se concentrou em aprovar reajustes em um patamar aceitável (inferior a dois dígitos).

² O SIN (sistema interligado nacional) é muito mais complexo, já que o despacho da usina depende de sua localização, da proximidade com a demanda do mercado consumidor, se há “folga” nas linhas de transmissão, etc. Entretanto, geralmente a ordem de mérito é respeitada na maioria dos casos.

atualmente é a de Xavantes, localizada em Goiânia, cujo custo é cima de (R\$2.500/MWh)³.

Essa abordagem permitiu que fossem utilizadas ferramentas aprendidas em sala de aula para o desenvolvimento do trabalho, principalmente a obtenção de dados disponíveis na internet e a manipulação dos mesmos, de forma a transformá-los em informação. O processo foi feito de forma segmentada, de forma a garantir que o projeto fosse finalizado no prazo, mesmo que por algum motivo os dados não estivessem mais disponíveis ou que a instabilidade da rede prejudicasse a obtenção deles novamente. Sendo assim, a cada segmento concluído (script), foram salvos os dados gerados em formato .csv para aproveitamento posterior nas demais etapas.

Procedimento

O primeiro script criado foi getDataONS.R, acessa todas as páginas do “Boletim diário de Informação”, iterativamente. Por esse script foi gerada a tabela Ger_termica_diaria.csv⁴. Essa tabela contém a energia efetivamente gerada por cada usina, diariamente. A princípio pensei em utilizar dados desde 2019 (por isso esta tabela tem início em 01/01/2019, mas só consegui coletar as informações do custo variável unitário – CVU (etapa 2) a partir de outubro de 2019. A informação disponível do CVU consta em documentos semanais de programação mensal de operação (PMO), também do ONS.

Conseguir esses dados foi mais trabalhoso, já que os documentos estão em formato pdf. Tentei utilizar o pacote do tabulizer, mas depois de várias tentativas sem sucesso resolvi organizar os dados “na mão”, com um leitor simples de texto. Embora haja documentos anteriores à outubro de 2019 com as tabelas do CVU, o leitor de texto não consegue capturar (uma possibilidade é que as tabelas tenham sido incorporadas como imagens, mas não investiguei o porquê). Como consegui os dados desde o início da pandemia, prossegui para o restante da análise. O script desenvolvido para esse fim foi o getCVU.R, armazenando as informações na tabela Ger_termica_PMO.csv.

O passo seguinte foi completamente manual. Percebi que cruzar as informações das duas tabelas de forma automática seria muito mais difícil do que observar os dados para tentar fazer uma “limpeza” manual. Isso porque havia várias linhas com NA em quase todas as colunas sendo mas um ou outro valor numérico, sendo que o nome da usina é muito semelhante ao de outra (mas com um espaço a menos por exemplo). Então o que fiz foi copiar “na mão” essas informações “perdidas” em uma linha com nome quase igual e apagar as “linhas repetidas”, o que resultou na tabela Ger_termica_PMO_limpa.csv. Além disso, tentei montar uma tabela de correspondência entre os nomes (Corresp_nome_mod.csv), também de forma completamente manual.

Já de posse dos dados necessários, o desafio passou então a ser o cruzamento dos dados entre as duas tabelas criadas anteriormente. Pela limitação de tempo, não foi possível inspecionar a tabela de geração diária de energia, então a decisão foi de descartar os dados em que a correspondência entre os nomes da usina não tenha sido identificada. Houve casos em que mais de um valor foi

³ Citando alguns dados para comparação: o PLD (preço de liquidação das diferenças, que pode ser entendido como “preço spot de energia” é de R\$ 583,88 (estrutural) e R\$ 1.141,85/MWh (preço horário). No mercado regulado, a diferença de custo entre o PLD e o custo de operação é cobrada via encargos, sendo utilizado o valor arrecadado com o adicional de conta bandeiras ou, se não for suficiente, entra no próximo processo de reajuste tarifário.

⁴ Precisei restaurar uma versão antiga dessa tabela, pois o script para gerá-la demora bastante no meu computador. Como havia apenas um backup no formato do Excel, foi preciso certificar de que o separador decimal era ponto e não havia separação de milhar antes ler o arquivo em outro script. A diferença principal é que preciso importar as tabelas criadas via Excel com separador “;” (talvez seja configuração do meu Excel, mas apenas resolvi o problema e não me aprofundei).

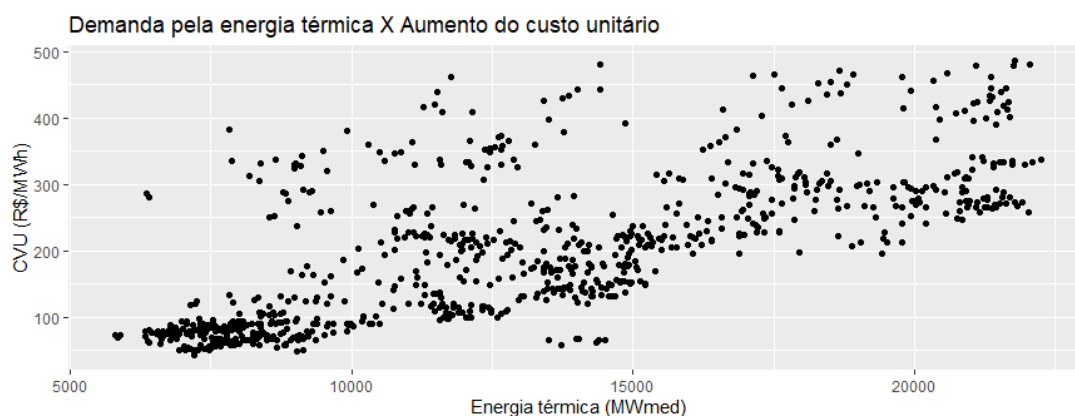
encontrado, então a opção foi utilizar a média. O nome do script desenvolvido para fazer esse cálculo é bem sugestivo (CalculaCustoPonderado.R). Como o dado do CVU é semanal, foi criada uma tabela de custo diário repetindo o custo semanal até que haja informação um CVU mais recente⁵. O peso de cada usina foi calculado com base no total de energia térmica gerada no dia⁶. Para facilitar análises posteriores, o script salva três tabelas .csv com dados diários: peso_ger_termica_diaria.csv, cvu_diario_pmo.csv e custo_diario_termicas.csv, além de outras duas com o total da geração térmica diária (geracao_diario_total.csv) e o custo ponderado diário (custo_diario_total.csv).

Em seguida, foi construído um script para plotar os principais gráficos que constam na seção de resultados: GeraGráficos.R, que inclusive gerou o gráfico da seção anterior. Aquele gráfico foi obtido considerando o CVU mais recente divulgado pelo ONS. Por fim, foi criado o script petroleo_prices_brl.R que contém as informações de petróleo (em US\$) e câmbio (em R\$) para comparar com a média de CVU.

A maioria dessas usinas tem como combustível ou o gás natural ou o diesel (algumas podem funcionar com ambos). Portanto, é de se esperar que haja correlação positiva entre o preço do petróleo (aqui foi utilizado o preço do Brent em reais, com o Brent em dólares baixado do site do Fred manualmente <https://alfred.stlouisfed.org/series/downloaddata?seid=DCOILBRETEU>), Houve ajuste manual, com a cópia da cotação do dia anterior para as datas em que o valor veio zerado. Para o câmbio, foi realizada consulta no site do BCB de forma manual. Essa opção foi feita pela viabilidade de cópia das tabelas em arquivo Excel e conversão para “.” em decimal (além de manter um formato de data na primeira coluna mais facilmente tratável), já que o período máximo de consulta de cotação é de 6 meses.

Análise dos dados

Os gráficos a seguir mostra tanto o custo de geração térmica quanto a quantidade de energia térmica gerada. No gráfico abaixo podemos notar que a nuvem de pontos sugere uma correlação positiva entre essas variáveis, como esperado. Além disso, os outliers de custo mais elevado são bem mais numerosos do que os outliers com custo mais baixo.

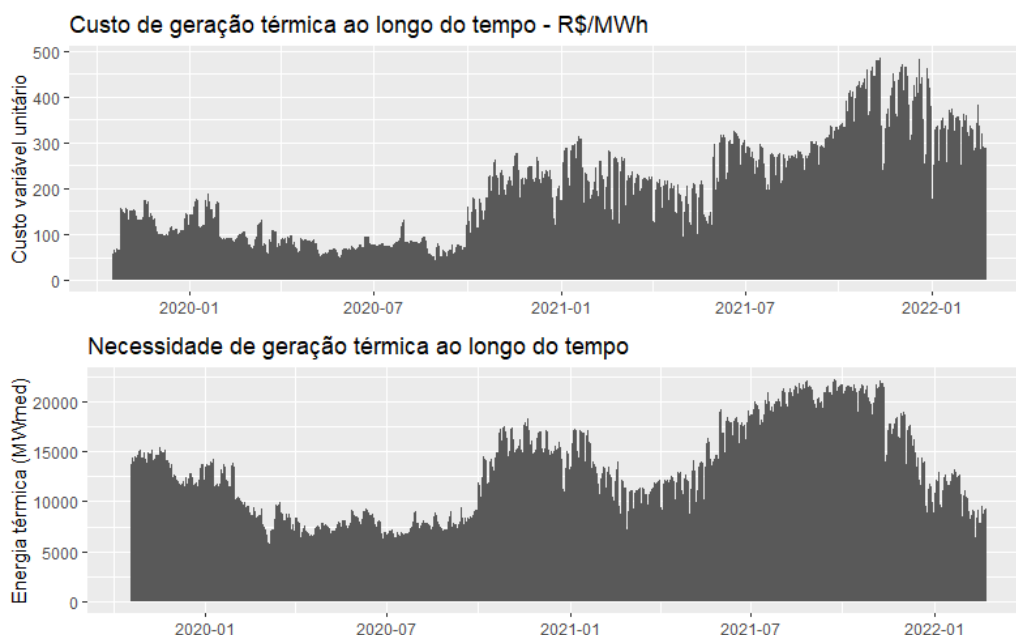


A próxima imagem é composta de dois gráficos, para comparação da demanda de energia térmica e do CVU ao longo do tempo. Para uma análise mais aprofundada referente a esses três gráficos sobre a possibilidade do aumento da demanda em geral (e impossibilidade de suprimento por

⁵ A programação é semanal, mas caso tenha havido falha na obtenção da informação para alguma semana, esse custo é repetido por um prazo maior.

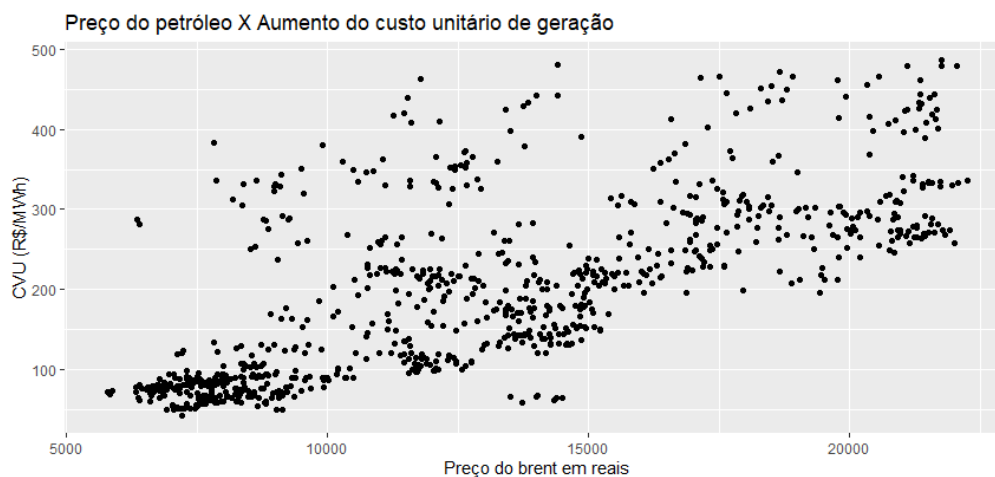
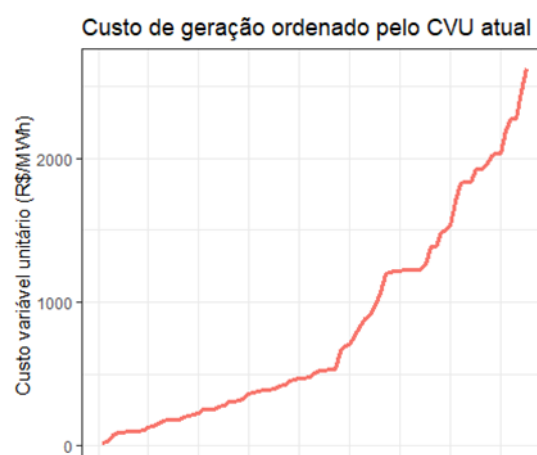
⁶ Como os itens sem correspondência foram descartados, seria necessário fazer um ajuste para garantir que a soma dos pesos seja 1. Entretanto, não houve tempo hábil para tal.

outras fontes) poderia ser feita deflacionando o CVU pelo IPCA, mas pelo fato de o período ser curto (pouco mais de 2 anos), é provável que não seja tão relevante.

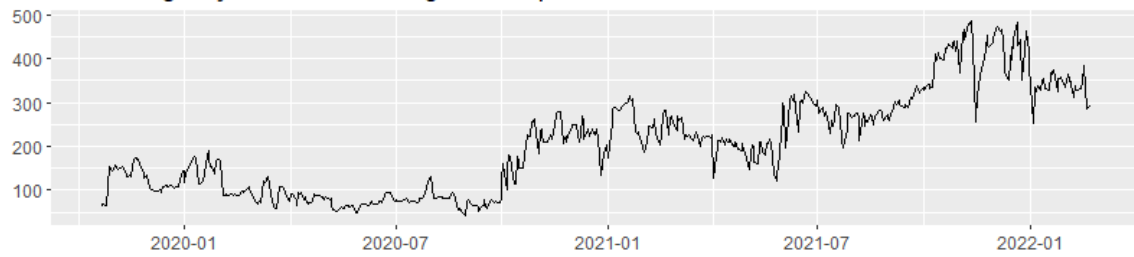


Entretanto, esse não é o único fator que influencia o custo da energia térmica. O custo de geração térmica é claramente influenciado pelo preço dos derivados de petróleo, já que os principais combustíveis utilizados são o gás natural e o diesel, o que pode ser visto nos próximos três gráficos. Também no gráfico abaixo é possível notar que os outliers com custo mais alto são bem mais numerosos do que aqueles com preço abaixo da nuvem de pontos.

Uma possível explicação para a existência desses outliers é a não linearidade do CVU (o primeiro gráfico, que copio novamente aqui ao lado). É possível observar que essa curva começa linear, mas torna-se cada vez mais inclinada na ponta.



Custo de geração térmica ao longo do tempo em R\$/MWh



Preço do petróleo - Brent em reais

