

RELATÓRIO TÉCNICO

Dashboard de Análise de Custos Variáveis Unitários - SIN

1. INTRODUÇÃO

Este relatório descreve o desenvolvimento de um dashboard interativo para análise dos custos variáveis unitários das usinas térmicas do Sistema Interligado Nacional (SIN) brasileiro. A ferramenta foi construída em Python, utilizando as bibliotecas Plotly e Dash, permitindo a exploração dinâmica dos dados de custos operacionais.

2. DESCRIÇÃO DO DATASET E OBJETIVOS

2.1 CONJUNTO DE DADOS

O dataset contém registros dos custos variáveis unitários das usinas térmicas do SIN, com as seguintes variáveis principais:

- data_inicio / data_fim: Período de análise
- ano / mes / ano_pmo / mes_pmo: Informações temporais
- numero_revisao / semana_operativa: Detalhes operacionais
- id_modelo_usina / id_subsistema / subsistema / usina: Identificação das usinas
- custo_variavel_unitario: Valor do custo variável unitário (R\$)

Esses dados permitem análises por região, período e usina específica.

2.2 OBJETIVOS DA ANÁLISE

1. Analisar a distribuição geral dos custos variáveis unitários das usinas térmicas
2. Visualizar a evolução temporal dos custos operacionais
3. Comparar padrões de custos entre subsistemas do SIN
4. Identificar usinas com maiores custos médios para priorização de intervenções
5. Fornecer análises estatísticas detalhadas para suporte à tomada de decisões

3. TRATAMENTO E TRANSFORMAÇÃO DOS DADOS

As transformações foram implementadas no módulo 'data_processing.py':

3.1 CARREGAMENTO E LIMPEZA

- Conversão de datas para o formato datetime (Pandas)
- Preenchimento de valores nulos em custo variável unitário (substituídos por zero)
- Remoção de registros duplicados

3.2 TRANSFORMAÇÕES APLICADAS

- Criação da coluna 'semana_do_ano' a partir da data de início
- Categorização dos custos em faixas:
"Baixo" (0-100), "Médio" (100-300), "Alto" (300-600), "Muito Alto" (>600)
- Criação da coluna 'ano_mes' para filtros temporais
- Ordenação dos dados por data

3.3 CÁLCULO DE ESTATÍSTICAS

- Média, mediana, mínimo, máximo
- Desvio padrão e variância
- Segmentação das estatísticas por subsistema ou período

4. VISUALIZAÇÕES E INTERATIVIDADE

O dashboard apresenta as seguintes visualizações principais:

4.1 HISTOGRAMA DE DISTRIBUIÇÃO DE CUSTOS

- Objetivo: Analisar a frequência dos diferentes níveis de custos
- Implementação: Histograma com bins configuráveis
- Insights: Identificação de faixas de concentração e outliers

4.2 SÉRIE TEMPORAL DE EVOLUÇÃO DOS CUSTOS

- Objetivo: Visualizar tendências e sazonalidade
- Implementação: Gráfico de linha com filtro por usina
- Insights: Identificação de períodos de alta/baixa e padrões sazonais

4.3 BOXPLOT PARA COMPARAÇÃO ENTRE GRUPOS

- Objetivo: Comparar a distribuição dos custos entre subsistemas/categorias
- Implementação: Boxplot agrupável
- Insights: Disparidades regionais e outliers

4.4 GRÁFICO DE BARRAS PARA RANKING DE USINAS

- Objetivo: Identificar usinas com maiores custos médios
- Implementação: Barras ordenadas (top N)
- Insights: Priorização de intervenções

4.5 INTERATIVIDADE

- Filtros: Subsistema, período (ano-mês), usina, agrupamento, top N
- Visualização dinâmica: Atualização automática das visualizações e estatísticas

5. INTERPRETAÇÃO E INSIGHTS

5.1 DISTRIBUIÇÃO DE CUSTOS

- Maioria das usinas opera na faixa média (100-300 R\$)
- Pequeno grupo com custos > 600 R\$ impacta a média geral

5.2 VARIAÇÕES REGIONAIS

- Norte/Nordeste: custos mais elevados
- Sul/Sudeste/Centro-Oeste: maior consistência e menor variabilidade

5.3 TENDÊNCIAS TEMPORAIS

- Custos aumentam em períodos de escassez hídrica
- Sazonalidade clara, com picos em meses específicos
- Tendência de aumento gradual ao longo dos anos

5.4 OPORTUNIDADES DE OTIMIZAÇÃO

- Ranking identifica usinas prioritárias para intervenções de eficiência

6. CONCLUSÃO

O dashboard facilita a identificação de padrões, anomalias e oportunidades de otimização nos custos variáveis das usinas térmicas do SIN. As visualizações e estatísticas permitem decisões mais informadas, com possibilidade de filtros específicos.

A escolha dos tipos de gráficos foi fundamentada na natureza dos dados e nos objetivos de análise:

- O histograma foi utilizado para analisar a distribuição dos custos variáveis unitários, pois permite visualizar rapidamente a frequência de diferentes faixas de valores e identificar concentrações ou outliers.
- O gráfico de linha (série temporal) foi escolhido para acompanhar a evolução dos custos ao longo do tempo, facilitando a identificação de tendências, sazonalidades e variações pontuais.
- O boxplot foi selecionado para comparar a distribuição estatística dos custos entre diferentes grupos (subsistemas ou categorias), pois evidencia a dispersão, a mediana e a presença de valores atípicos.
- O gráfico de barras foi empregado para o ranking das usinas, permitindo destacar de forma clara as usinas com maiores custos médios e facilitar a priorização de intervenções.