

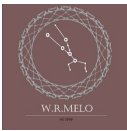
Primeira análise dos dados enviados

O Manifesto de Dados — Fase 1 Diagnóstico constitui o primeiro esforço estruturado de consolidação e interpretação do acervo operacional disponível dos fornos F1 a F5. Ele oferece uma visão abrangente sobre a extensão temporal, a granularidade e a coerência das medições históricas, evidenciando um conjunto de dados robusto, contínuo e diversificado, capaz de sustentar análises quantitativas em múltiplas escalas. Essa base, que cobre mais de seis anos de operação, permite não apenas reconstruir o comportamento metalúrgico e eletrotérmico dos fornos, mas também avaliar padrões de eficiência, variações de consumo e impactos de eventos operacionais. A partir deste diagnóstico inicial, torna-se possível delinear um plano de exploração analítica que vá desde a avaliação da qualidade e completude dos dados até a modelagem preditiva e o desenvolvimento de gêmeos digitais, definindo os caminhos para transformar o histórico operacional em conhecimento acionável e suporte à decisão técnica.

Manifesto de Dados — Fase 1 Diagnóstico

Este documento é crucial, pois define a **extensão e a granularidade** dos dados operacionais disponíveis para os fornos **F1 a F5**.

Abaixo está um mapa mental organizado por forno, detalhando a quantidade, a extensão temporal e a abrangência das informações, e sua relevância para uma análise inicial de projetos de metalurgia e eletrometalurgia.



Primeira análise dos dados enviados

Mapa Mental da Disponibilidade de Dados Operacionais por Forno

O conjunto de dados abrange um período extenso de **6 anos e 4 meses**, de **janeiro de 2018 a abril de 2025 (parcial)**, para a maioria das categorias, fornecendo uma base robusta para a análise de tendências de longo prazo, ciclos operacionais e eventos específicos.

Categorias de Informação Comuns (F1, F2, F3, F4, F5)

Variável Inferida	Extensão Temporal (Aprox.)	Quantidade (Média/Ano)	Abrangência e Relevância
Informações Diárias (Inf.Diario.csv)	2018 a 2025 (abril)	cerca de 365 linhas/ano/forno (45 a 52 colunas)	Contém dados agregados diários. Essencial para acompanhamento de performance de alto nível e identificação de tendências gerais.
Corridas (Corrida.csv)	2018 a 2025 (abril)	entre 2.500 e 4.100 linhas/ano/forno (34 a 35 colunas)	Registra cada corrida (vazamento) de metal e escória. Crucial para a análise metalúrgica (composição química, peso, taxa de produção, rendimento).
Consumo Fornos (Consumo.csv)	2018 a 2025 (abril)	cerca de 900.000 a 1.020.000 linhas/ano/forno (26 a 42 colunas)	Dados de alta frequência (o volume sugere leituras de minuto em minuto ou subminuto). Fundamental para análise de eficiência energética e modelagem dinâmica do processo eletrotérmico (por exemplo, flutuações de potência).
Medição Eletrodo	2021 a 2025 (abril)	761 medições totais (5 colunas)	Fornecer o histórico de desgaste e consumo do eletrodo (Comprimento_Eletrodo). Não está vinculado a um forno específico no manifesto, mas é relevante para a engenharia e modelagem elétrica do forno.



Primeira análise dos dados enviados

Com base no mapa mental do manifesto, estima-se que cada forno (F1 a F5) possua entre **950 mil e 1,1 milhão de registros anuais apenas nas séries de consumo**, somando-se a isso os dados de **corridas (~3 mil registros/ano)** e **informações diárias (~365 registros/ano)**.

Em um horizonte de **seis anos e quatro meses**, cada forno concentra cerca de **6 a 7 milhões de observações individuais**, levando o total consolidado do parque (cinco fornos) para aproximadamente **30 a 35 milhões de registros** — sem incluir os conjuntos supervisórios de alta frequência, que, sozinhos, acrescentam centenas de milhares de medições adicionais por segundo ou minuto nos fornos F4 e F5.

Trata-se, portanto, de um volume de dados expressivo, mas plenamente compatível com o ecossistema moderno de **Data Science aplicada à indústria**, onde técnicas de amostragem, particionamento e processamento paralelo em frameworks como **Pandas, Dask ou PySpark** permitem o tratamento eficiente de milhões de linhas.

Essa escala garante representatividade estatística suficiente para análises exploratórias, validação cruzada de modelos e desenvolvimento de algoritmos preditivos capazes de capturar tanto tendências de longo prazo quanto variações dinâmicas de processo.

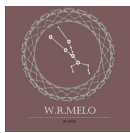
Detalhamento por Forno e Diferenciais

Forno F1 (Completo e Consistente)

- **Quantidade e Extensão:** Apresenta um volume completo de dados nas três categorias principais (Diário, Consumo, Corrida) para todos os anos, de 2018 a 2025.
- **Abrangência:** Os arquivos de Consumo para F1 possuem maior variação de colunas (26 a 42 colunas), especialmente em 2019 e 2020 (42 colunas), sugerindo um conjunto mais rico de variáveis de consumo.
- **Relevância:** Alto potencial para modelagem estável. A consistência do registro ao longo de mais de seis anos, sem grandes lacunas, o torna excelente candidato para construção de modelos preditivos robustos e análise de série temporal.

Forno F2 (Completo e Consistente)

- **Quantidade e Extensão:** Cobertura completa nas três categorias de 2018 a 2025.
- **Abrangência:** Estrutura de colunas nos arquivos de Consumo é mais constante (26 colunas).



MARINGÁ FERRO LIGAS



Primeira análise dos dados enviados

- **Relevância:** Permite validação cruzada de modelos desenvolvidos em F1, garantindo resultados não específicos de um único forno.

Forno F3 (Inconsistente e com Lacunas Operacionais)

- **Quantidade e Extensão:** F3 apresenta as maiores lacunas temporais.
 - Inf.Diário e Corrida têm registros faltantes em 2020, 2022 (312 dias), 2023 (230 dias) e 2024 (284 dias).
 - Dados de Consumo de 2023 começam apenas em 10/02 e os de 2024 em 26/03.
- **Abrangência:** Arquivos de Consumo têm 27 a 28 colunas, mais do que F2.
- **Relevância:** Lacunas sugerem paradas operacionais significativas (manutenção, reformas ou campanhas). Relevante para estudar impacto de startups e interrupções, mas desafiador para séries contínuas.

Forno F4 (Supervisório de Alto Volume e Detalhe)

- **Quantidade e Extensão:** Cobertura completa nas categorias base. O Inf.Diário de 2018 cobre 2018–março/2022 (1.551 linhas).
- **Abrangência — Diferencial Chave:** Único forno com dados Supervisórios de alta densidade:
 - Dois arquivos (F4_2024_1S.csv e F4_2024_2S.csv) somam aproximadamente 486 mil linhas.
 - Coluna de timestamp zerada (1970-01-01), mas arquivos de 505 MB e 366 MB com 127 colunas (tags como CMF4_PAC4200_FP_, PTR4_COR_SEC_, VALOR_UN_PV).
- **Relevância:** Essencial para modelagem eletrotérmica e controle. Permite correlacionar ações de controle (movimento de eletrodo, mudança de tap) com resposta elétrica do forno.

Forno F5 (Supervisório Detalhado em Minuto e Hora)

- **Quantidade e Extensão:** Cobertura completa nas categorias base.
- **Abrangência — Diferencial Chave:** Dados Supervisórios com granularidade diferenciada:
 - Supervisor_F5_Hora.csv → 26.266 linhas de dados por hora.
 - Supervisorio_F5_Minuto.csv → 215.833 linhas de dados por minuto (2022-12-30 a 2023-06-01).



MARINGÁ FERRO LIGAS

Primeira análise dos dados enviados

- **Relevância:** Dados de minuto são extremamente valiosos para análise cinética e otimização de curto prazo. O contraste temporal com F4 (2024 vs 2023) permite avaliar diferentes campanhas ou evoluções de controle.

Conclusão para a Primeira Análise – Possibilidades

1. Análise de Longo Prazo e Eficiência Global (F1–F5, Corrida e Diário)

- Calcular eficiência metalúrgica e consumo específico de energia (kWh/t) entre 2018–2025, combinando dados de Corridas (produção) e Consumo (energia).

2. Modelagem Eletrotérmica Detalhada (F4 e F5 - Supervisório)

- Analisar dados Supervisórios para desenvolver modelos de controle de potência e resistência.
- A alta frequência (1 milhão de linhas/ano) permite correlação de energia com variáveis elétricas e mecânicas (**posição do eletrodo**).

3. Análise de Descontinuidade (F3)

- Investigar causas e impactos das paradas operacionais no F3, correlacionando reinícios com mudanças de performance.

4. Integração com a Teoria de Ferroligas

- Comparar consumo de energia, rendimento de metal e composição das corridas com faixas teóricas esperadas para fornos SAF.

PARA CONFIRMAÇÃO: o foco atual será desenvolver a possibilidade 2?