

DOCUMENTAÇÃO COMPLETA DO PROJETO — PIPELINE DE DADOS DE ACIDENTES RODOVIÁRIOS

1. Descrição do Problema

Órgãos de segurança, pesquisadores e analistas enfrentam dificuldade para consolidar, limpar e analisar grandes volumes de dados de acidentes rodoviários...

2. Objetivos do Sistema e Justificativa Técnica

Objetivos

- Automatizar ingestão de dados CSV de acidentes.
- Padronizar formatação e normalizar campos críticos.
- Validar consistência e identificar registros nulos ou inconsistentes.
- Criar base consolidada para análises estatísticas.
- Gerar saídas estruturadas.
- Garantir repetibilidade.

Justificativa Técnica

- Dados brutos apresentam formatos irregulares.
- Volumetria elevada impede processamento manual.
- Pipeline automatizado melhora governança.

3. Escopo da Solução

Incluído

- Ingestão, padronização, validação, limpeza, normalização, enriquecimento, relatórios.

Não incluído

- Machine learning, API pública, dashboards avançados, enriquecimento geoespacial.

4. Arquitetura Completa do Pipeline

Etapas: Ingestão → Validação → Limpeza → Normalização → Enriquecimento → Armazenamento → Relatórios.

5. Ferramentas e Tecnologias Utilizadas

Python, Pandas, Docker, GitHub, orquestradores opcionais.

6. Decisões Técnicas

CSV como fonte; Pandas para transformação; Parquet como destino analítico.

7. Guia de Execução

Instalar dependências → preparar CSVs → executar pipeline → revisar logs → analisar saídas.

8. Dependências

Python 3.x, Pandas, libs de validação, Docker (opcional), orquestrador (opcional).

9. Descrição dos Dados

(id, data_inversa, dia_semana, horario, uf, br, km, municipio, causa_acidente, tipo_acidente, classificacao_acidente, fase_dia, sentido_via, condicao_meteorologica, tipo_pista, tracado_via, uso_solo, pessoas, mortos, feridos_leves, feridos_graves, ilesos, ignorados, feridos, veiculos, latitude, longitude, regional, delegacia, uop)

10. Pontos de Falha e Limitações

Datas invertidas, coordenadas com vírgula, padronização complexa, layout variável do CSV.

11. Trabalho Individual

A: arquitetura e validações.

B: limpeza e normalização.

C: documentação e dicionário.

D: infraestrutura e orquestração.