Jhonnes Amaral Soares Toledo

Cientista de Dados Júnior

Juiz de Fora, MG · Brasil | jas_toledo@hotmail.com | linkedin.com/in/jhastoledo | github.com/jhastoledo

Mestre em Física pela UFJF e especialista em Ciência de Dados, com 12 projetos publicados que vão do modelo relacional em SQL ao aplicativo em produção. Trabalho com dados tabulares, séries temporais e texto em português, escolhendo a métrica pelo problema — PR-AUC sob desbalanceamento, QWK em alvo ordinal, RMSPE em previsão — e usando SHAP para explicar o que o modelo aprendeu.

COMPETÊNCIAS TÉCNICAS

Linguagens	Python, SQL (CTEs, window functions, subqueries correlacionadas, plano de execução)
Modelagem	scikit-learn, XGBoost, LightGBM, statsmodels, Prophet, PyTorch, Transformers (BERTimbau)
Interpretab. e tuning	SHAP, Optuna, Ray Tune, testes de hipótese e tamanho de efeito
Visualização e BI	Power BI, Matplotlib, Seaborn, Plotly, GeoPandas, Folium
Entrega	Streamlit, Streamlit Cloud, Git/GitHub, JupyterLab, conda, SQLite
Familiaridade	R (análise estatística) · AWS (fundamentos de nuvem, curso concluído)

PROJETOS SELECIONADOS

Rossmann — previsão de vendas github.com/jhastoledo/rossmann_forecast

Previsão de 42 dias para **1.115 lojas** com um único modelo global: **RMSPE 0,1321** e **MAPE 9,7%** no teste, vencendo o SARIMAX especializado em 3 das 4 séries. Documentei também onde o clássico ainda ganha — as 17 lojas de regime atípico.

Validação walk-forward por data, guarda anti-vazamento, LightGBM, Ray Tune + Optuna, SHAP, app Streamlit

Detecção de fraude em cartão de crédito github.com/jhastoledo/credit_card_fraud

Classificação com desbalanceamento de **578:1** em 284.807 transações: **PR-AUC 0,826**, recall 0,811 e precisão 0,951 no limiar 0,264 — escolhido pela curva, não pelo padrão. PR-AUC no lugar do ROC-AUC, que engana sob desbalanceamento extremo.

Pipeline sklearn (RobustScaler → XGBoost), Optuna, seleção de limiar, SHAP, app Streamlit

Olist — PLN de avaliações em português github.com/jhastoledo/olist_reviews_nlp

Previsão de nota de 1 a 5 estrelas a partir do texto, com alvo ordinal medido por QWK. BERTimbau fine-tunado alcançou **QWK 0,8640** contra 0,8169 do TF-IDF + SVM — mas o SVM é ~147x mais rápido em inferência e dispensa GPU; registrei o trade-off, não só o vencedor.

PyTorch, BERTimbau, TF-IDF, análise comparada de erro, app Streamlit

Olist — analytics em SQL github.com/jhastoledo/olist_sql_analytics

Modelagem relacional de 9 tabelas com PK/FK sobre ~100 mil pedidos, com 30 consultas versionadas. Achados: **17,6% dos vendedores concentram 80% da receita** e o atraso de entrega só derruba a avaliação acima de 7 dias (de 4,29 para 1,71).

SQL puro, SQLite, EXPLAIN QUERY PLAN, análise RFM, relatório final publicado

EXPERIÊNCIA

Cientista de Dados — projetos independentes abr/2026 – atual

Portfólio próprio · Remoto

- Portfólio de **12 projetos completos**, do dado cru ao aplicativo publicado em Streamlit Cloud.
- Arquitetura repetida em todos: Pipeline do scikit-learn como artefato único, pacote **src/** modular e notebooks numerados do EDA ao relatório, com baseline explícito antes de qualquer modelo complexo.
- Cobertura de regressão, classificação, clustering, séries temporais, dados desbalanceados e PLN.

Pesquisador em Física Aplicada abr/2022 – set/2024

Universidade Federal de Juiz de Fora · Juiz de Fora, MG

- Modelagem de sistemas complexos e análise quantitativa de dados experimentais em Python.
- Propagação de incerteza, validação de modelos contra medida e comunicação técnica de resultados.

FORMAÇÃO

Especialização em Ciência de Dados — UNINASSAU 2025

Mestrado em Física — Universidade Federal de Juiz de Fora 2022 – 2024

Bacharelado em Física — Universidade Federal de Juiz de Fora 2019 – 2022

Certificações: Ciência de Dados Impressionador (Hashtag Treinamentos, 2026) · Algoritmos e Lógica de Programação (Hashtag Treinamentos, 2026)