

LUAN SOUSA

luansousa@protonmail.com · [Portfolio](#) · [Github](#) · [LinkedIn](#)

Cientista de Dados e Estatístico com experiência em modelagem preditiva, machine learning e visualização de dados. Especializado em Python, R e SQL, com histórico na implementação de soluções de BI (Tableau, Apache Superset) e automação de pipelines de dados. Posuo background acadêmico em Modelos Lineares Generalizados e Aprendizado Supervisionado e Não Supervisionado, além de publicações em conferências de Estatística.

EDUCAÇÃO

Universidade Federal do Ceará

Fortaleza, CE

Departamento de Estatística e Ciência de Dados

Bacharelado - Estatística

SKILLS

Linguagens de Programação - Python (Pandas, Scikit-learn, BeautifulSoup, Scrapy), R (tidyverse, ggplot2), SQL (Postgresql)

Coursework Selecionados - Análise de Componentes Principais, Modelos Lineares Generalizados, Planejamento de Experimentos, Comparações de Modelos de Bossting, Análise de Cluster, Aprendizado de Maquina Supervisionado e Não Supervisionado, Diagnósticos e seleção de modelos, Selação de Variáveis, Visualização de dados, Mineração de Dados

Ferramentas de BI/Visualização - Tableau, Power BI, Apache Superset, Looker Studio, Matplotlib, Seaborn, Observable **Habili-**

dades e Competências - Comunicação Técnica para Não-Técnicos, Trabalho em Equipe Multidisciplinar, Capacidade Analítica e Pensamento Crítico, Resolução de Problemas Complexos, Adaptabilidade a Novas Tecnologias

Idiomas - Inglês(Intermediário - B1), - Espanhol(Intermediário - B1)

EXPERIÊNCIA

Perícia Forense do Estado do Ceará

Ago 2024-Ago 2025

Ciência de Dados, Estágio

Fortaleza, CE

- Liderei a migração para o Apache Superset como plataforma de BI em nuvem, substituindo soluções pagas (Power BI/Tableau), resultando em uma redução de custos de aproximadamente 30% substituindo soluções pagas, eliminando custos de licenças e proporcionando maior autonomia na criação de dashboards.”
- Projetei e implementei uma solução de engenharia de dados para criar um ambiente analístico isolado, replicando dados do banco de produção via scripts Python e SQL. Esta iniciativa reduziu o risco de lentidão ou indisponibilidade do sistema crítico em 20% e capacitou a equipe de dados a executar consultas complexas sem restrições, acelerando a geração de informações e na democratização dos dados.
- Desenvolvi e implantei pipelines de dados automatizados para atualizações incrementais, assegurando que os dashboards no Superset exibissem dados near real-time (quase em tempo real). Esta otimização reduziu a janela de atualização de dias para horas e eliminou a carga desnecessária de reprocessamentos completos, mantendo a integridade e a performance do sistema.
- Garanti a integridade e confiabilidade dos insights conduzindo uma Análise Exploratória de Dados (EDA) rigorosa em R (tidyverse, ggplot2). Identifiquei e solucionei vieses de amostragem, outliers críticos, correlações espúrias através de testes estatísticos (qui-quadrado, correlação) e análise de distribuições, assegurando a qualidade dos dados que alimentavam os dashboards e modelos decisórios

Sicredi

Abr 2023-Abr 2024

Análise de Dados, Estágio

Fortaleza, CE.

- Para combater a inadimplência e agilizar a concessão de crédito, ajudei a desenvolver um modelo preditivo usando séries temporais e PyCaret para comparar automaticamente os melhores algoritmos.
- Traduzi dados de RH em insights estratégicos através da análise de turnover, desempenho e clima organizacional. Desenvolvi modelos preditivos para identificar os principais drivers de rotatividade e criei projeções de crescimento de headcount e simulações de alocação de recursos, fornecendo a base analítica para decisões de planejamento estratégico e retenção de talentos.

Gauss - Empresa JR

Jun 2024-Jun 2025

Estatístico jr

Fortaleza, CE.

- Lead de Projetos de Ciência de Dados: Liderei a entrega de consultorias para clientes do setor público e privado, gerando insights a partir de dados não estruturados. Minhas responsabilidades incluíam desde a coleta automatizada de dados via web scraping (Scrapy, Selenium) até a modelagem estatística avançada e a apresentação dos resultados.
- Construí pipelines de dados robustos para coletar, limpar e estruturar informações de fontes públicas, realizando feature engineering para criar variáveis preditivas mais relevantes para a modelagem.

- Apliquei Modelos Lineares Generalizados (GLM), séries temporais e análise multivariada para identificar padrões comportamentais e socioeconômicos em projetos clientes.

PESQUISAS E PUBLICAÇÕES

Modelo De Regressão Gama Inversa Com Zeros Ajustados - 25th SINAPE

Ago 2024

- Neste trabalho, propomos um avanço metodológico ao estender a distribuição **Gamma Inversa** com ajuste zero para um modelo de regressão. Essa extensão permite a análise de variáveis dependentes com fatores explicativos associados. O modelo de regressão com ajuste zero proposto baseia-se na distribuição de Vitorino (2024), mas incorpora uma estrutura de regressão para avaliar o efeito de covariáveis na variável dependente. Assim, oferece uma ferramenta estatística mais abrangente e flexível para áreas que trabalham com esses tipos de dados.

Modelagem Preditiva de Partidas de League of Legends Usando Aprendizado Supervisionado - EMR

Out 2025

- Este estudo explora a aplicação de modelos de aprendizado supervisionado para prever os resultados de partidas de League of Legends, um dos esportes eletrônicos mais populares do mundo. Diversas abordagens foram analisadas, incluindo K-Nearest Neighbors (KNN), Regressão Logística, Gradient Boosting e Árvores de Decisão.