

# Rafaella Duarte

 [rafaelladuarte](#) |  [rafaella-duarte-carvalho](#) |  [rafaelladuarte.github.io/](#) |  [rafaella.d.d.carvalho@gmail.com](mailto:rafaella.d.d.carvalho@gmail.com)

## RESUMO

Graduada em Ciências Econômicas, com mais de 3 anos de experiência em engenharia e análise de dados utilizando Python. Atuo em todo o ciclo de dados, incluindo coleta via web scraping, rotinas ETL/ELT, criação de APIs REST e integração com bancos de dados relacionais e não relacionais. Atualmente, trabalho como engenheira de dados júnior CLT na Legal Data da E-xyon e desenvolvo projetos pessoais focados em Big Data, explorando tecnologias como Spark, Airflow, ecossistema Hadoop e computação em nuvem. Tenho também familiaridade com RabbitMQ, Google Cloud Storage, BigQuery e Amazon S3.

## SKILLS

Linguagens	Python, Bash, SQL
Banco de Dados	PostgreSQL, SQL Server, MongoDB, Elasticsearch, Redis e Neo4j
Cloud	Google Cloud Storage, Google Cloud BigQuery, Amazon S3
Ferramentas	Git, RabbitMQ, Docker, Airflow, FastAPI, Selenium
Metodologias	Web Scraping, ETL/ELT, API Rest

## EDUCAÇÃO

2019 - 2025 Bacharelado em Ciências Econômicas **Universidade Federal de Uberlândia**

## EXPERIÊNCIAS

**Legal Data, E-xyon - Engenharia de Dados Junior** Jan 2025 - atual  
CLT Tempo Integral - Remoto

**Flow2Go - Engenharia de Dados Junior** Jan 2024 - Dez 2024  
Freelancer - Remoto

- Desenvolveu Web Scraping com resolução de captchas para coleta de dados jurídicos (CNA e CNSA).
- Criou scripts Python para integração e saneamento de dados no MongoDB, garantindo qualidade e consistência.
- Realizou integração e enriquecimento de dados com bases públicas (CNPJ, IBGE, CEP, SFN).
- Desenvolveu API Python para sumarização e criação de dashboards de análise.

**Linx, StoneCo - Engenharia de Dados Junior** Nov 2022 - Out 2023  
CLT Tempo Integral - Remoto

- Automatizou e expandiu rotinas de Web Scraping para múltiplos marketplaces (Amazon, Mercado Livre, Magalu, etc.).
- Desenvolveu bibliotecas internas em Python para otimização de processos. - Replicou dados do Elasticsearch para BigQuery, viabilizando análises de catálogo de produtos.
- Monitorou filas no RabbitMQ e pipelines em Kubernetes (Cron Jobs).
- Implementou ingestão de dados no Google Cloud Storage e Amazon S3.

**Nodis, StoneCo - Analista de Dados** Jan 2021 - Nov 2022  
Estágio Meio Período - Remoto

- Criou e manteve rotinas de Web Scraping em Python para marketplaces.

- Desenvolveu pipelines ETL para PostgreSQL, MongoDB e Elasticsearch.
- Integrou dados com a plataforma VTEX via API.
- Produziu dashboards no Google Data Studio para análise de KPIs internos.

## PROJETOS

---

### **Woman A3Data Challenge - Hackathon Engenharia de Dados - [Link da Solução](#)**

Projeto de solução para hackathon promovido pela A3 Data, que consistiu na criação de um pipeline ELT para extração, transformação e disponibilização de dados da Base CNPJ (2010–2021). Realizei limpeza, organização e estruturação dos dados, além de criar visualizações para responder perguntas-chave do desafio. Técnicas utilizadas: web scraping, automação de processos e integração com bancos de dados. Tecnologias: Python, ETL, visualização de dados e pipeline orquestrado. Resultado: 2º lugar.

### **Web Scraping de Imobiliárias Orquestrado com Apache Airflow - [Link da Solução](#)**

Projeto pessoal para automação da coleta e análise de dados de imóveis à venda em sites de imobiliárias de Uberlândia-MG. Desenvolvi pipeline ETL que realiza extração, limpeza, transformação e carga dos dados em PostgreSQL e MongoDB. Os dados são apresentados em dashboard interativo no Power BI para análise de tendências do mercado imobiliário. Tecnologias usadas: Python, web scraping, ETL, bancos relacionais e não relacionais, visualização de dados.

### **Assistente IA com RAG para Identificação de Aves - [Link da Solução](#)**

Projeto final desenvolvido para o conclusão do curso LLM Zoomcamp da DataTalks.Club. Utiliza RAG, Streamlit e MinSearch para identificar aves a partir de descrições fornecidas pelo usuário.