

Pedro Duarte Faria

Data Platform Engineer

Belo Horizonte, Brasil | Telefone: +55 (31) 98512-4294 | Email: pedropark99@gmail.com

[Website pessoal](#) | [GitHub](#) | [LinkedIn](#) | [X \(Twitter\)](#) | [Mastodon](#)

1 Por que sou um candidato forte ?

- ajudei a criar um resultado de 400%¹ em aumento de vendas em um dos nossos clientes-chave. Também desenvolvi um novo serviço de dados com receita potencial na casa de milhões de dólares.
- mais de 5 anos de experiência na indústria de dados. Utilizando Python, SQL, R, Apache Spark, Databricks e Azure Data Factory para criar e manter pipelines de dados e processos de ETL complexos. Trabalhei em grandes projetos de dados com algumas das maiores empresas do mercado financeiro brasileiro (Sodexo, Banco Pan, Banco BMG), utilizando bancos de dados que ultrapassam a marca de um bilhão de linhas.
- Tenho bastante experiência com programação em Python, R e SQL, e também possuo conhecimento em C, Rust e Zig.
- Durante 2 anos eu criei e mantive 1 mil pipelines de ingestão de dados com vários tipos de “*targets*” ou “destinos” de ingestão diferentes, como REST API, SFTP, Azure Blob Storage e Amazon S3. Conectando empresas a seus parceiros por meio de dados.
- Sou um engenheiro de dados certificado em Apache Spark pela empresa Databricks (*Associate Developer for Apache Spark 3.0*), e tenho experiência trabalhando com Data Lakes hospedados em ambiente cloud Azure.
- Usuário experiente de Git. Eu mantenho alguns projetos open-source em meu GitHub, e recentemente fiz algumas contribuições para o pacote `knitr`², o qual é um grande projeto open-source da comunidade internacional de R³.
- Possuo experiência em ferramentas de DataViz, especialmente Power BI.

2 Educação

Universidade Federal de Ouro Preto - UFOP

B.A. Economia, Fevereiro 2022

Universidade Federal de Minas Gerais - UFMG

Intercâmbio acadêmico - Economia, Fevereiro 2019 - Novembro 2020.

¹Você pode ler mais a respeito desse caso em: <https://pedro-faria.netlify.app/posts/2023/2023-08-31-data-decisions/en/>.

²<https://github.com/yihui/knitr/pull/2140> ; <https://github.com/yihui/knitr/pull/2180>

³Na verdade, eu tenho uma [página que lista todas as minhas contribuições open-source](#), caso você queira olhar esse assunto de forma mais profunda.

3 Experiências profissionais

Data Platform Engineer - [Blip](#), Abril 2024 - *presente*

Desenvolvendo uma nova geração de produtos e serviços de dados para os nossos clientes, além de uma nova plataforma de dados *self-service* para a empresa, com o objetivo de empoderar os nossos colegas com dados. Este trabalho envolve uma boa quantidade de desenvolvimento *back-end*, integrando diferentes serviços e APIs, e usando principalmente Delta Sharing, Python, Databricks, Grafana e serviços da Azure Cloud.

Engenheiro de Dados - [Blip](#), Fevereiro 2023 - Março 2024

Desenvolvi e mantive mais de 1 mil pipelines de dados com o objetivo de fornecer infraestrutura de dados para o departamento de Customer Success da empresa. Para isso, utilizo principalmente Python, SQL, R, Apache Spark, Databricks e Azure Data Factory. Eu também desenvolvi novas soluções de dados para a empresa que aprimoraram processos internos.

Analista de dados - [Blip](#), Maio 2021 - Fevereiro 2023

Eu ajudo clientes a tomarem melhores decisões com dados, e a desenvolverem os seus negócios com o nosso produto (plataforma Blip). Eu também desenvolvo muitas pipelines de dados (com Python e SQL) e dashboards (com Power BI). Meus clientes incluem a [Sodexo](#), [Banco BMG](#) e o [Banco Pan](#);

Estágio no Departamento de Estatísticas e Informações - [Fundação João Pinheiro](#), Junho 2019 - Março 2021

Eu ajudei vários pesquisadores a automatizarem suas rotinas e pipelines de dados com a linguagem R. Esse estágio teve um impacto muito grande sobre o meu livro e meus outros projetos pessoais.

4 Projetos pessoais

4.1 Projetos de livro

- [Introduction to Zig: a project based book](#). Primeira edição. Um livro introdutório sobre a linguagem de programação Zig. Este livro introduz a linguagem através de pequenos projetos, que são implementados em conjunto com o leitor.
- [Introduction to pyspark](#). Primeira edição. Um livro introdutório (em inglês) sobre o pyspark, que é a API em Python para o Apache Spark.
- [Introdução à Linguagem R: seus fundamentos e sua prática](#). Quinta edição. Um livro introdutório (em português) sobre a linguagem de programação R.
- Estou coordenando um projeto de tradução do livro *rOpenSci Packages: Development, Maintenance and Peer Review* para o Português, como parte do projeto [Multilingual Publishing Project](#) da rOpenSci.

4.2 Projetos open-source mais recentes

- [figma](#): um pacote de R que fornece um cliente para a API da plataforma Figma.
- [spark_map](#): um pacote de Python que te ajuda a aplicar funções sobre várias colunas de um Spark DataFrame.
- [ggfunnel](#): um pacote de R que te ajuda a produzir gráficos de funil semelhantes aos disponíveis no Power BI.

5 Línguas

- Português, nativo.
- Inglês, intermediário para avançado.
- Francês, atualmente no processo de aprendizado.