

Pedro Duarte Faria

Data Platform Engineer

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

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

1 What makes me a strong candidate?

- Helped to improve sales by 400%¹ at one of our key clients. I have also developed a new data service with potential income of millions of dollars at launch.
- more than 5 years of experience in the data industry. Using Python, SQL, R, Apache Spark, Databricks, and Azure Data Factory to create and maintain complex data pipelines and ETL processes. I've worked on big data environments (more than 6 TB of data generated per day), developing data projects with some of the biggest companies in the Brazilian financial market (Sodexo, Pan Bank, BMG Bank).
- I'm certified in Apache Spark by the company Databricks (*Associate Developer for Apache Spark 3.0*), and I have experience working with Data Lakes hosted in an Azure cloud environment.
- I have a lot of experience with programming in Python, R, and SQL on my daily work. But I also have experience with programming in low-level languages, like C, Zig and Rust.
- I have built and maintained 1 thousand data ingestion pipelines for 2 years, with various types of “targets” or “endpoints”, such as REST API, SFTP, Azure Blob Storage, and Amazon S3. Connecting companies to their partners through data.
- Experienced Git user. I maintain a few open-source projects on my GitHub, and I have recently made some contributions to the knitr² package, which is a large open-source project of the international R community³.
- I have experience with DataViz tools, especially Power BI.

2 Education

Federal University of Ouro Preto - UFOP

B.A. Economics, completed at February 2022.

Federal University of Minas Gerais - UFMG

Economics Department, Academic exchange, February 2019 - November 2020.

¹You can read more about this case at: <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>

³I actually have a [webpage that lists all of my open-source contributions](#) , if you want to take a deeper look at it.

3 Professional Experiences

Data Platform Engineer - [Blip](#), Apr 2024 - *present*

Developing next-generation data products and services for our clients, and also a new self-service data platform to our company, with the objective of empowering our colleagues with data. This work involves a lot of back-end development, and also, integrating different services together, using mainly Azure Cloud Services, Delta Sharing, Python, Databricks and Grafana.

Data Engineer - [Blip](#), Feb 2023 - Mar 2024

Developed and maintained data pipelines for various purposes, aiming to provide data infrastructure for the company's Customer Success department, and the clients of Blip. For that, I mainly use Python, SQL, R, Apache Spark, Databricks, and Azure Data Factory. I also developed new data solutions for the company that have improved internal processes.

Data Analyst - [Blip](#), May 2021 - Jan 2023

I helped clients make better decisions with data and develop their businesses with our product (Blip platform). I developed a lot of data pipelines (using Python and Apache Spark) and dashboards (using Power BI), in addition to writing several presentations and reports using Data Storytelling as well. My clients included big companies from the Brazilian financial market, such as [Sodexo](#), [BMG bank](#) and [Pan bank](#).

Internship at Statistics and Information Department - [João Pinheiro Foundation](#), June 2019 - March 2021

I helped multiple researchers automate their tasks and data pipelines with R, and I got a lot of experience processing big volumes of raw data. This was a huge experience for me, and it had a great impact on my book and course projects.

4 Personal projects

4.1 Book projects

- [Introduction to Zig: a project-based book](#). First edition. An introductory book about the Zig programming language. This book introduces the language by implementing small projects together with the reader.
- [Introduction to pyspark](#). First edition. An introductory book about the Python package pyspark.
- [Introdução à Linguagem R: seus fundamentos e sua prática](#). Fourth edition. An introductory book (in portuguese) about the R programming language.
- I am working on the translation to Portuguese of **rOpenSci Packages: Development, Maintenance, and Peer Review** book, which teaches best practices on R package development, as a part of the [Multilingual Publishing project](#).

4.2 Most recent open-source projects

- [figma](#): R package that provides a R client to the Figma API.
- [spark_map](#): Python package to easily apply a function over multiple columns of a Spark DataFrame.
- [ggfunnel](#): R package to easily produce Power BI like funnel charts in R.

5 Languages

- Portuguese, native.
- English, intermediary to advanced.
- French, currently learning.