

O'REILLY®

Python Polars

The Definitive Guide

Transforming, Analyzing, and Visualizing Data
with a Fast and Expressive DataFrame API



Jeroen Janssens
& Thijs Nieuwdorp
Foreword by Ritchie Vink,
Creator of Polars

Praise for *Python Polars: The Definitive Guide*

Jeroen and Thijs have done an excellent job—not only teaching you the ins and outs of Polars but also helping you unlearn habits from other tools like pandas. They really bring out the power of expressions, which are key to using Polars effectively, guiding you toward a more declarative, functional approach to data processing. As you work through this book, I’m sure you’ll gain a deep understanding of Polars and discover fresh ways to approach data processing.

—Ritchie Vink, Creator of Polars (excerpt from
the Foreword)

Polars has become a rising star in the Python data ecosystem, showing what’s possible in a next-generation data frame library. Jeroen and Thijs have written a timely and essential resource to help you take advantage of everything Polars has to offer.

—Wes McKinney, Creator of pandas, Principal
Architect, Posit PBC

Polars has brought a ton of much-needed innovation to the data frame world with its much more streamlined API and efficient implementation. As a result, the capabilities of data analysis in Python are pushed to new heights. We also greatly enjoy Ritchie and team as a part of the Amsterdam data ecosystem.

I greatly respect Jeroen's commitment to teaching data science in an accessible way, whether it be on the command line or elsewhere. His and Thijs's book is a testament to this commitment, and I recommend it to the data science community.

—Hannes Mühleisen, Cocreator of DuckDB

As a client working closely with Thijs and Jeroen on migrating a data pipeline to Polars, we were initially skeptical, but we soon experienced the speed and intuitiveness of Polars and its API. While Jeroen and Thijs worked late hours to make progress with their book, we directly benefited from the improvements in our pipeline. We hope this book helps you along the way and that you find all the little gems Polars has to offer—while being lazy, of course!

—Marnix van Lieshout and Bram Timmers,
Data Scientists at Alliander

This book will change how you think about data analysis. Jeroen and Thijs have done a phenomenal job including all kinds of comparisons, diagrams, and examples. Polars has an incredible amount of functionality, and it's clear they've put great care into organizing and breaking all the pieces down. I appreciate their focus on data visualization throughout the book and the inclusion of table styling!

—Michael Chow, Principal Software Engineer
at Posit PBC, Co-maintainer of Great Tables

This book cleverly demystifies Polars' powerful ecosystem. Thijs and Jeroen seamlessly guide you through the theoretical foundations and hands-on examples, making complex concepts accessible without sacrificing depth. Whether you're migrating from pandas or starting fresh with Polars, this guide provides the roadmap you need to realize your data workflows could be running at a far more bearable speed.

—Hella Haanstra, Machine Learning Engineer
at Xomnia

When I first interacted with Polars, it was so early days that I made the PR for the `.pipe()` method on DataFrames. I was pleased with the speedups that Polars gave me, but I was massively impressed by the API. It just felt like such good taste and a great direction for the future. Fast-forward a few years, and today Polars has become an established tool with so many features that the ecosystem was in dire need of a guide. This book gives us just that. It is a guide, but also a reference!

—Vincent D. Warmerdam, Data Person,
Cofounder of Calmcode

Python Polars: The Definitive Guide manages to offer a comprehensive overview of everything Polars has to offer while also providing a great learning experience in the form of excellent code examples. Truly a great resource!

—Stijn de Gooijer, Core Contributor to Polars

Polars is emerging as one of the leading data frameworks in Python, especially for time series analysis and forecasting. It is now fully integrated with libraries like Nixtla's MLForecast and StatsForecast, allowing for the creation of forecasts at scale with high performance.

Jeroen and Thijs have done an excellent job of establishing a solid foundation for both new practitioners who wish to learn how to process data with Python and experienced users looking to transition from pandas to Polars.

—Rami Krispin, Senior Manager, Data Science
and Engineering at Apple

The depth that Jeroen and Thijs went into in order to produce this phenomenally good book is impressive. There are some pretty good Polars books out there, but this is the best one. They don't just repeat what's in the user guide—they go above and beyond: in-depth explanations of expressions, (friendly) comparisons with other tools, an example of how to go beyond what Polars offers out-of-the-box with a geocoding plugin! Whether you're new to Polars or want to improve your understanding of it, I wholeheartedly recommend this book.

—Marco Gorelli, Senior Software Engineer at
Quansight, Core Contributor to Polars and
pandas, Creator of Narwhals

Python Polars: The Definitive Guide

Transforming, Analyzing, and Visualizing Data with a Fast and Expressive DataFrame API

Jeroen Janssens and Thijs Nieuwdorp

O'REILLY®

Python Polars: The Definitive Guide

by Jeroen Janssens and Thijs Nieuwdorp

Copyright © 2025 Jeroen Janssens and Thijs Nieuwdorp. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://oreilly.com>). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

- Acquisitions Editor: Aaron Black
- Development Editor: Sarah Grey
- Production Editor: Jonathon Owen
- Copyeditor: Sonia Saruba
- Proofreader: Miah Sandvik
- Indexer: WordCo Indexing Services, Inc.

- Interior Designer: David Futato
- Cover Designer: Karen Montgomery
- Illustrator: Kate Dullea

- February 2025: First Edition

Revision History for the First Edition

- 2025-02-19: First Release

See <http://oreilly.com/catalog/errata.csp?isbn=9781098156084> for release details.

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *Python Polars: The Definitive Guide*, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

The views expressed in this work are those of the authors and do not represent the publisher's views. While the publisher and the authors have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the authors disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of

or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-098-15608-4

[LSI]

Foreword

It was never meant to be this serious.

In December 2019, I became a dad, and then the pandemic hit. This left me dazed; juggling life with a newborn and trying to hold on to a sense of “me.” In a burst of new-dad sleep deprivation, I figured I’d take on a new project. Honestly, it started as nothing more than a hobby.

At work, I had to join two CSV files while programming in Rust, and it felt like way more hassle than it should have been. I wondered if there were easier ways to get it done than setting up SQLite. So, as one often does in software development, I decided to try my hand at creating my own join algorithm.

At the time, I was pretty new to Rust and didn’t know much about optimizing performance. So as a proud writer of my first join algorithm, I learned that my implementation was much slower than pandas. This unsatisfying result planted the seed of what would later become Polars.

This led me down a path of researching database engines, learning Rust, and a lot of trial and error over the next year and a half. As I learned more about databases, algorithms,

performance, memory, unsafe code, etc., my goals shifted. I went from just wanting to make a faster join than pandas to building a DataFrame package for Rust, and eventually, a high-performance query engine that could rival the state of the art in the Python landscape.

I drew inspiration from pandas' strengths and weaknesses, the declarative approaches of SQL and PySpark, functional programming principles, and the rigor of Rust's type system. At first, I even thought I'd model it directly on the pandas API. But I quickly realized that limiting myself that way hurt my creative drive and would only lead to a less effective tool. When I decided to let go of this constraint and merged the lazy and eager APIs into one expression-based API, the project was molding into something people might now recognize as Polars.

On March 15, 2021, I released this as a research project on PyPI. I was able to evolve this pet project into a full-fledged DataFrame processing package called Polars. Eventually, Polars' success enabled me to secure funding and start my own company, Polars Inc.

Today, Polars is its own company—something I'm proud of—and it's a constant source of energy for me. Finally, a lot of the ideas I have can come to fruition (not all of them, but hey, I'm

ambitious). With Polars, my hope is to create one DataFrame API that works across the board, whether you're dealing with small or big data, all on the same tech stack.

Together with a dedicated team at Polars Inc. and a growing, enthusiastic community of contributors, we're tackling many interesting technical challenges. Each scale comes with its own set of challenges, and finding solutions for these tough problems so that millions of users have a better experience is incredibly rewarding. One challenge involves accurately handling time zones. This may sound trivial, but dealing with time zone conversions and ensuring consistency across different systems is notoriously difficult. A big technical challenge we're currently facing is translating the DataFrame API into a full streaming model. This transition requires rethinking how operations—such as computing the mean or accessing the last row—will work in a streaming context, as opposed to in-memory batch processing. I'm confident that we'll rise to these challenges.

Over time, Polars has grown beyond my initial expectations. It's been incredibly rewarding to see it used in surprising scenarios, such as finite element method simulations involving 1,500 joins. Users have also employed Polars for

metaprogramming tasks, generating complex queries that would be impractical to write by hand.

Looking ahead, the future of Polars is exciting. We're focused on extending its streaming capabilities, allowing you to process huge datasets on your own laptop. We're also working on creating a distributed cloud environment to do fast distributed computing on massive datasets. Another key goal is to support extension types, paving the way, for example, for geospatial data types. These developments aim to make Polars a go-to tool for data processing—from tiny datasets that fit in memory to massive datasets that require distributed computing. On top of that, we'll keep investing in the user experience through improvements like error messages that inform you up front when a query will eventually fail.

I'm really glad you're holding *Python Polars: The Definitive Guide* by Jeroen Janssens and Thijs Nieuwdorp. I know Jeroen and Thijs well from our days at Xomnia, where we shared a lot of training sessions, company trips, drinks, ideas—the whole mix. When Jeroen first suggested I write a book, I had to thank him—my energy was all going into Polars—but I wholeheartedly supported him when he decided to take up the challenge with Thijs. Then, on a company trip to Jordan, both Jeroen and Thijs held me to it, and I ended up spending hours

on the bus explaining—and occasionally defending—my design choices for Polars.

Jeroen and Thijs have done an excellent job—not only teaching you the ins and outs of Polars but also helping you unlearn habits from other tools like pandas. They really bring out the power of expressions, which are key to using Polars effectively, guiding you toward a more declarative, functional approach to data processing. As you work through this book, I'm sure you'll gain a deep understanding of Polars and discover fresh ways to approach data processing.

Enjoy the read, follow along with the examples, and take your data analysis skills to a new level.

Ritchie Vink, Creator of Polars

Amsterdam, November 2024

Preface

Polars has quickly emerged as one of the most exciting innovations in the Python data ecosystem. With its blazingly-fast performance, expressive API, and ability to handle massive datasets, it's become a powerful alternative to traditional tools like pandas. Built with efficiency in mind, Polars leverages Rust for its backend to offer great speeds, especially when it comes to multi-core and GPU acceleration. Whether you're working with large-scale data wrangling, complex transformations, or real-time analysis, Polars has shown that it can handle the toughest challenges with ease.

We first learned about Polars through our colleague, Ritchie Vink, the creator of Polars, at Xomnia in Amsterdam. This was back in 2022, when Polars wasn't as well-known as it is now. Intrigued by Ritchie's enthusiasm and explanation of some of its features, we decided to give it a try. We were hooked. The elegant syntax and the impressive speed stood out immediately. It was a stark contrast to the complexity and performance limitations we had encountered with other tools. Polars felt like a breath of fresh air—simple, yet powerful. We could see that this was something special.

Soon after, we had the perfect opportunity to put Polars to the test. We were working together on the same team at Alliander, the largest utility company in the Netherlands. A particular data pipeline needed to be scaled up so that it would process the entire electrical grid on a weekly basis. Unfortunately, because it was implemented in a combination of R and Python with pandas, this was infeasible. Thanks to a minimal viable benchmark with Polars, we managed to convince the team to reimplement the entire pipeline using Polars. As a result, we were able to drastically improve the performance of the pipeline while reducing memory consumption. The impact was immediate and transformative, not only saving the project but also cementing our belief in Polars' potential.

This success convinced us that Polars deserved a book so that this package and our hard-won knowledge could be shared with a larger audience. Fast-forward a year and a half, and here's the result: *Python Polars: The Definitive Guide*. We hope that by reading this book, you'll be equipped to achieve similar success in your own projects.

Who This Book Is For

This book is designed for anyone looking to leverage the power of Polars in Python to transform, analyze, and visualize data more efficiently and effectively. Whether you're a seasoned data analyst, a data engineer, or even someone new to the world of data science, you'll find valuable insights and practical examples that can be applied directly to real-world challenges. To illustrate the diverse ways in which Polars can benefit different users, let's take a look at two key personas: Hanna, a seasoned data analyst, and Kosjo, an experienced data engineer.

Hanna: The Data Analyst

Hanna is a seasoned data analyst. She's comfortable with Python and has a good grasp of pandas but occasionally struggles with its syntax and feels there must be a more elegant way to perform certain operations. Like many analysts, she regularly tackles exploratory data analysis (EDA) tasks that involve cleaning, transforming, and summarizing large datasets. However, she often finds herself battling with pandas' sometimes complex and unintuitive syntax, especially when it comes to performing more advanced data manipulations or scaling her work to larger datasets.

For someone like Hanna, this book offers a streamlined, more intuitive alternative to pandas, with the added benefit of being able to handle data at a larger scale without sacrificing speed or readability. Polars provides a more Pythonic and performant way to perform the types of analyses Hanna does daily. By learning Polars, Hanna can simplify her workflow, write more elegant code, and unlock greater performance in her exploratory data analysis tasks.

Kosjo: The Data Engineer

Kosjo is an experienced data engineer, tasked with processing large volumes of data and building pipelines that support complex data workflows. They are highly skilled in Python and work with various technologies to ensure smooth data movement and processing. As part of their role, Kosjo is often responsible for optimizing processes to reduce infrastructure costs, especially when working with big data. This means reducing the time and resources required for heavy transformations without having to manage a distributed computing cluster.

Polars can help Kosjo achieve these goals. It is designed for speed and performance, especially when dealing with large datasets or intensive transformations. Its parallel execution