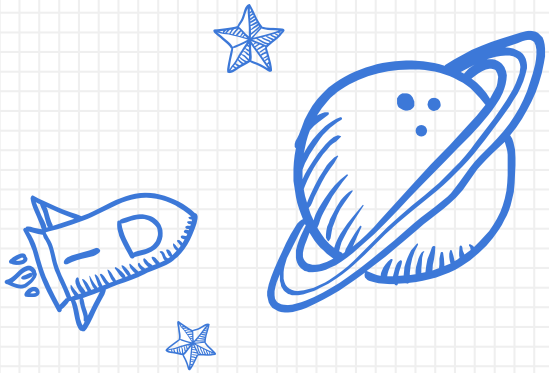
Data Analysis with Polars and Python

Nice to meet you!

Boris Paskhaver

- Software Engineer/Consultant
- Teacher
- Author of Pandas in Action



Data Analysis

The process of deriving an insight from a dataset to drive a decision

What is Polars?

- ✘ A fast tool for importing, sorting, filtering, grouping, joining data, and more
- ✘ A data analysis library written in Rust with bindings for Python
- ✘ A side project by developer Ritchie Vink
- ✘ 35,000+ stars on GitHub

Why is it called Polars?

- ✘ Some consider Polars the successor to the popular Pandas library
- ✘ Pandas is short for “panel data structures” (3-dimensional datasets)
- ✘ Polar bears are more powerful than panda bears
- ✘ Rust files have a “**.rs**” extension. Polars conveniently ends in “rs”.

What sets Polars apart?

- ✘ Built for speed and efficiency from the ground up, taking advantages of Rust's features
- ✘ Parallelizes computations across multiple threads and CPU cores
- ✘ Uses the columnar Apache Arrow data storage format to optimize column-based operations
- ✘ Supports lazy evaluation with query optimizations

The Plan

- ✘ Download the course materials from GitHub. They include both the datasets and complete / incomplete Jupyter Notebooks.
- ✘ Use the **uv** command-line tool to install Python, Polars, and the JupyterLab coding environment
- ✘ [**OPTIONAL**] Complete the Python bootcamp to learn the fundamentals of the language
- ✘ Dive into the Polars library.