Introduction to Data Science Programming in Python

J. Hathaway - Data Science Program Chair

Disclaimers

I am focusing on modern tools for data science in Python - Polars over Pandas, Plotly over Matplotlib, and Streamlit over Dash. These modern tools reflect the best of

- declarative programming (task focused programming)
- clean grammar (language abstraction that allows intuitive but complex actions)
- industry respect (all three tools are very popular for the quality and have rapid

Agenda

1. Introduction and Set-up (30 minutes)
2. Polars for data munging (45 minutes)
3. Break (10 minutes)
4. Plotly for data visualization (45 minutes)
5. Streamlit for dashboards (45 minutes)

J. Hathaway

Data Scientists with ~20 years of industry experience and 8 years in Academia. Undergraduate degree in Economics (University of Utah) and a graduate degree in Statistics (BYU).

Checking our installation

1. Python Installed
2. VS Code Installed
3. Python VS Code Extension Installed
4. Python packages installed.

```
pip install polars[all] plotly streamlit
```

Introduction to Polars and Data Munging

Polars is a lightning fast DataFrame library/in-memory query engine. Its embarrassingly parallel execution, cache efficient algorithms and expressive API makes it perfect for efficient data wrangling, data pipelines, snappy APIs and so much more. Polars is about as fast as it gets, see the results in the H2O.ai benchmark. Polars Website

Polars programming

Introduction to Plotly and Data Visualization