TUTORIAL

Python pandas tutor ide for beginners

Data Science Python

Are you ready to begin your pandas journ to get started.

May 2023 15 min read

Vidhi Chugh
AI Strategist and Ethicist

pandas is arguably the most important Python package for data analysis. With over 100 million downloads per month, it is the de facto standard package for data manipulation and exploratory data analysis. Its ability to read from and write to an extensive list of formats makes it a versatile tool for data science practitioners. Its data manipulation functions make it a highly accessible and practical tool for aggregating, analyzing, and cleaning data.

In our blog post on **how to learn pandas**, we discussed the learning path you may take to master this package. This beginner-friendly tutorial will cover all the basic concepts and illustrate pandas' different functions. You can also check out our course on **pandas Foundations** for further details.

This article is aimed at beginners with basic knowledge of Python and no prior experience with pandas to help you get started.

What is pandas?

pandas is a data manipulation package in Python for tabular data. That is, data in the form of rows and columns, also known as DataFrames. Intuitively, you can think of a DataFrame as an Excel sheet.

pandas' functionality includes data transformations, like **sorting rows** and taking subsets, to calculating summary statistics such as the mean, reshaping DataFrames, and joining DataFrames together. pandas works well with other popular Python data science packages, often called the PyData ecosystem, including

- **NumPy** for numerical computing
- **Matplotlib**, **Seaborn**, **Plotly**, and other data visualization packages
- **scikit-learn** for machine learning

What is pandas used for?

pandas is used throughout the data analysis workflow. With pandas, you can:

- Import datasets from databases, spreadsheets, comma-separated values (CSV) files, and more.
- Clean datasets, for example, by dealing with missing values.
- Tidy datasets by reshaping their structure into a suitable format for analysis.
- Aggregate data by calculating summary statistics such as the mean of columns, correlation between them, and more.
- Visualize datasets and uncover insights.

pandas also contains functionality for time series analysis and analyzing text data.

Key benefits of the pandas package

Undoubtedly, pandas is a powerful data manipulation tool packaged with several benefits, including: