

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

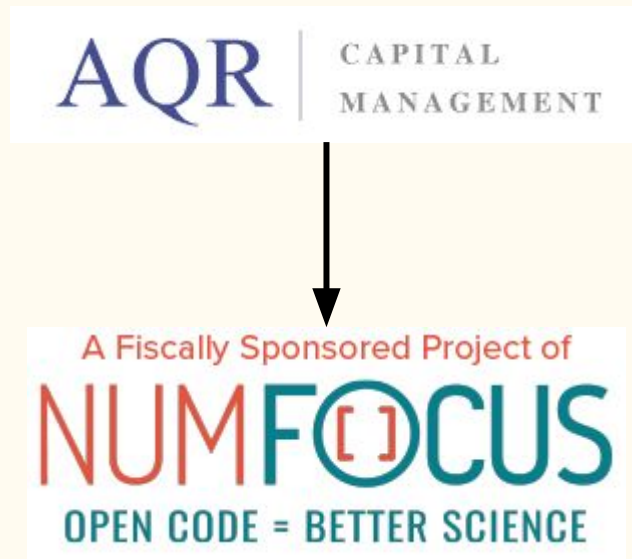


DataFrames with Pandas

—
By Will & Tori

A Brief History

- pandas originally began in 2008 at AQR Capital Management
- In 2009 it became an open source project
- Since 2015, pandas has been a NumFOCUS sponsored project
- Today, it is developed and supported by a community of developers and others who wish to contribute, following the community guidelines



What is a
DataFrame?

DataFrame

- 2 Dimensional data structure
 - Used to store relational data
 - Container that contains arrays which are the data
 - Aims to make working with relational data easy and intuitive
 - All values are mutable
 - Built on top of NumPy
-

When to use it?

- Tabular data
 - Ie: Excel and SQL like data with labels
 - Ordered and unordered time series
 - Arbitrary matrix data
 - Homogenous and heterogenous data
 - Unlabeled data and differently indexed data
-

Why DataFrames?

- Easily handles missing data
 - Drops missing or NaN data, replaces missing or NaN data, and interpolates data
 - Columns are mutable
 - Automatically aligns data
 - Can convert many other data structures into data frames easily
 - Ie: Numpy Arrays
 - Efficient data set operations like slicing, merging, and group by
-

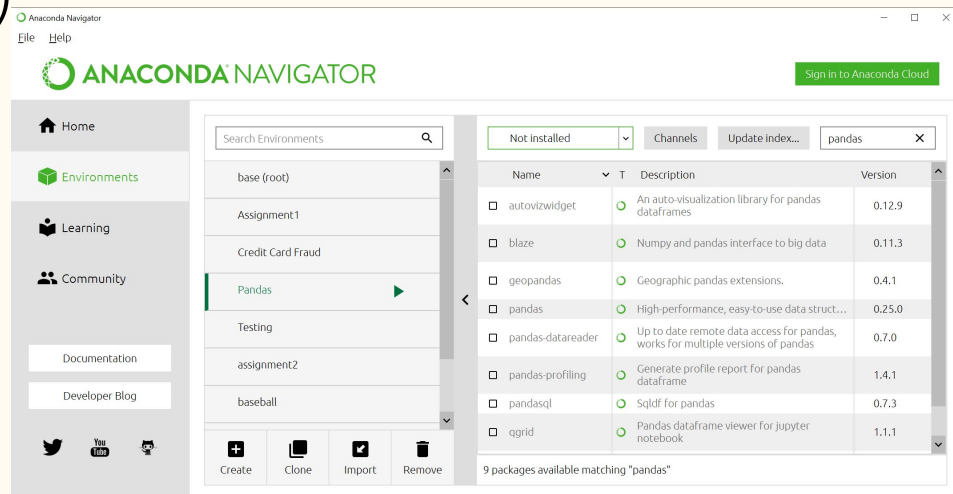
How to Install

—

Installation (require python 3.x)

Via Anaconda Navigator:

1. Select the Environment Tab
2. Select an existing environment or create a new one
3. Select *Not installed* from the drop down menu
4. Search for *pandas*, select the package, and hit apply



Via Anaconda Prompt

1. Navigate to the environment you want to use the package in
2. Run the command `$conda install pandas`
3. When asked to proceed with installation, enter *y*



Code Demo!

More than just DataFrames!

Series

- Series are the 1 dimensional versions of data frames
- Allow for the same functionalities as DataFrames but with 1 dimensional data
- It can be broken down even more as a container for scalars

Other pandas Offerings

- Allow for fast and efficient I/O of csv and excel like files
- Offers I/O support for the HDF5 format
- Plotting and Graphing tools for dataframes
- Various statistical computations such as variance, covariance, and correlation

Learn More

More information on dataframes and pandas, as well as more information on community contributions and related packages can be found at the pandas website:

<https://pandas.pydata.org/>

DataFrame documentation link:

<https://pandas.pydata.org/pandas-docs/stable/reference/frame.html#missing-data-handling>

numFOCUS

<https://numfocus.org/>

Questions?



https://miro.medium.com/max/3006/1*i1vVm3EqqDIkyucD0079wg.jpeg