

Twitter Thread by PyQuant News



PyQuant News

[@pyquantnews](#)



College completely failed to teach me data analysis.

So I spent over 10,000 hours learning Python.

Then, I picked the 13 best libraries for machine learning and data analysis.

But unlike college, these won't cost you \$120,000.

Here they are for free:

AutoViz

AutoViz performs automatic visualization of any dataset with a single line of Python code. Give it any input file (CSV, txt or json) of any size and AutoViz will visualize it.

<https://t.co/bywRcRK6xD>

Numba

Numba translates Python functions to optimized machine code at runtime using the industry-standard LLVM compiler library. Numba-compiled numerical algorithms in Python can approach the speeds of C or FORTRAN.

<https://t.co/ZwldyKWfeb>

NetworkX

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

<https://t.co/wE5NyoN66G>

pandas

pandas is a fast, powerful, flexible, and easy-to-use open source data analysis and manipulation tool, built on top of the Python programming language.

<https://t.co/u92fVCtMtJ>

Vaex

Vaex is a high-performance Python library for lazy Out-of-Core DataFrames (similar to Pandas), to visualize and explore big tabular datasets.

<https://t.co/Vm9slvG8pm>

PyMC

PyMC is a Python package for Bayesian statistical modeling focusing on advanced Markov chain Monte Carlo (MCMC) and variational inference (VI) algorithms. Its flexibility and extensibility make it applicable to a large suite of problems.

<https://t.co/52rGR8w2SF>

statsmodels

statsmodels is a Python package that provides a complement to scipy for statistical computations including descriptive statistics and estimation and inference for statistical models.

<https://t.co/uui7vu0abv>

bokeh

Bokeh is an interactive visualization library for modern web browsers. It provides elegant, concise construction of versatile graphics, and affords high-performance interactivity over large or streaming datasets.

<https://t.co/09unmJ9ecq>

Blaze

Blaze translates a subset of modified NumPy and Pandas-like syntax to databases and other computing systems. Blaze allows Python users a familiar interface to query data living in other data storage systems.

<https://t.co/walSiZp230>

SparklingPandas

SparklingPandas aims to make it easy to use the distributed computing power of PySpark to scale your data analysis with Pandas. SparklingPandas builds on Spark's DataFrame class to give you a polished, pythonic, and Pandas-like API.

<https://t.co/wMqx1FR83a>

Superset

Superset is a modern data exploration and data visualization platform. Superset can replace or augment proprietary business intelligence tools for many teams. Superset integrates well with a variety of data sources.

<https://t.co/EhyiJEfLzr>

PyCM

PyCM is a multi-class confusion matrix library written in Python that supports both input data vectors and direct matrix, and a proper tool for post-classification model evaluation that supports most classes and overall statistics parameters.

<https://t.co/m29qsogDTt>

Plotly Dash

Built on top of Plotly.js, React, and Flask, Dash ties modern UI elements like dropdowns, sliders, and graphs directly to your analytical Python code.

<https://t.co/xVke1X4O62>

Keep your \$120,000.

Learn Python:

- Vaex
- Blaze
- PyMC
- bokeh
- PyCM
- Numba
- AutoViz
- pandas
- Superset
- NetworkX
- Plotly Dash
- statsmodels
- SparklingPandas

That's a wrap!

If you enjoyed this thread:

1. Follow me [@pyquantnews](#) for more of these

2. RT the tweet below to share this thread with your audience <https://t.co/l31UHm05Hn>

College completely failed to teach me data analysis.

So I spent over 10,000 hours learning Python.

Then, I picked the 13 best libraries for machine learning and data analysis.

But unlike college, these won't cost you \$120,000.

Here they are for free:

— PyQuant News (@pyquantnews) [July 14, 2022](#)