

Mongo and Python

Keeton Martin

What is Python?

- Python is a nice programming language, especially for data science
- “Python is an interpreted, object-oriented, high-level programming language with dynamic semantics”
- Known for readability
- “Python supports modules and packages, which encourages program modularity and code reuse”
 - These are important for today



What is Pandas?

- It's a data wrangling python package
- Built on top of NumPy, which allows support for n-dimensional arrays
- Pandas allows easy I/o, Subsetting, Stats, Data cleaning, and more
- Many Data Science packages is written for use with Pandas DataFrames
- DataFrames are like Relations

Learn more:

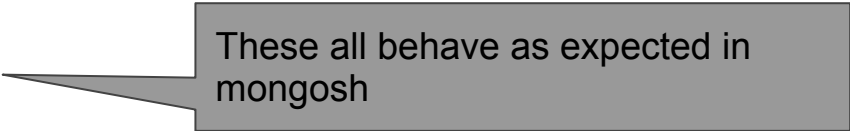
https://pandas.pydata.org/docs/getting_started/index.html

The diagram shows a table representing a DataFrame. The columns are labeled 'Name', 'Team', 'Number', 'Position', and 'Age'. The rows are indexed from 0 to 6. Annotations include: 'Columns' pointing to the header row; 'Rows' pointing to the first three rows; and 'Data' pointing to a sub-region of the table containing the 'Number', 'Position', and 'Age' columns for rows 2, 3, 4, and 5. The 'Data' region is highlighted with a pink box, and individual cells within it are also highlighted with pink boxes.

	<i>Name</i>	<i>Team</i>	<i>Number</i>	<i>Position</i>	<i>Age</i>
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0

The Basics of PyMongo

- Lazy Creation
 - Collections and Databases
- Python Dictionaries are used to represent documents
- Once you instantiate a collection in python, you can do lots of simple things
 - `my_collection.find_one()`
 - `my_collection.insert_one()`
 - `my_collection.insert_many()`
 - `my_collection.count_documents({})`
- `my_collection.find()` returns a Cursor object, which is iterable, and contains documents



These all behave as expected in mongosh

Learn more about PyMongo: <https://pymongo.readthedocs.io/en/stable/tutorial.html>

Pandas and PyMongo

- Cursor objects can just be turned into Python Lists
- Python Lists can be turned into DataFrames when they have dictionaries in them
- Once you have a DataFrame in Python you can:
 - Make plots
 - Calculate stats
 - Do some machine learning (build some models)
 - Save your data as a CSV
- You can also do everything in reverse
 - CSV → Pandas DF → Dictionary → `.insert_many()` → Into Mongo!





Demo Time!

Sources

- <https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/>
- <https://www.python.org/doc/essays/blurb/>
- https://pandas.pydata.org/docs/getting_started/index.html
- <https://pymongo.readthedocs.io/en/stable/tutorial.html>
- https://scikit-learn.org/stable/auto_examples/index.html
- <https://matplotlib.org/>
- <https://seaborn.pydata.org/>

