PANDAS Tutorial

A method of data analysis in python

2017

Why is this package useful?

PANDAS is used for data modeling and analyzing in the Python space

In the lab....

PANDAS will allow us to handle missing data more efficiently and has more analytical tools than num.py

What other tools does it work with?

i.e. python, num.py, timecorr, stock prophet

Organization and Logic behind the package

Organization

The two main types of tools in PANDAS are the DataFrame and the series. The series is a 1D scalar made of arrays while the DataFrame is a 2D array made of series.

Logic of package

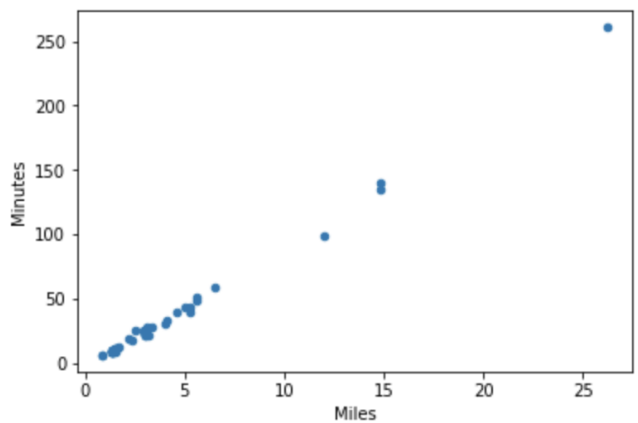
This distinction of series and DataFrames allows for faster regrouping and indexing.

Example Code

	Date	Miles	Minutes	Minutes_per_Mile	Day_of_week
0	2014-08-01	NaN	NaN	NaN	Friday
1	2014-08-02	1.70	NaN	NaN	Saturday
2	2014-08-03	1.40	NaN	NaN	Sunday
3	2014-08-04	2.10	NaN	NaN	Monday
4	2014-08-05	NaN	NaN	NaN	Tuesday
5	2014-08-06	2.38	NaN	NaN	Wednesday
6	2014-08-07	2.30	NaN	NaN	Thursday
7	2014-08-08	3.40	NaN	NaN	Friday
8	2014-08-09	2.50	NaN	NaN	Saturday
9	2014-08-10	3.36	28.616667	8.516865	Sunday

```
df.plot(x='Miles', y='Minutes', kind='scatter')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x10dda3780>
```



Common Commands

Rolling Apply

- The rolling apply function allows one function to be applied to an entire dataset

Describe

- provides general statistical functions of a dataset (mean, mode, number of unique points ...)

Quick data manipulation

- whether using the sort, rename, or splitting the data it is fast to organize and reorganize data and your work environment in PANDAS

Main takeaways from the package

- PANDAS will allows to analyze and manipulate data faster
- PANDAS also works well with many of the lab's already existing packages

- Next Steps
  - Tutorial: https://github.com/ContextLab/Tutorials/tree/master/PANDAS
  - Useful website/ reading on the package:
    - https://www.udemy.com/data-analysis-in-python-with-pandas/learn/v4/t/lecture/2339796?start=150
    - https://pandas.pydata.org

- Sources: udemy.com