Handling Time Series

Paweł Kordek

SOFTWARE ENGINEER

@pawel_kordek   https://kordek.github.io

Summary

Why another data type(s)?

Pandas to rule them all

Demo

Time series data is everywhere

The Problem with Time Series

playerID	birth Year	birth Mon	birth Day
hlushto01	1970	2	7
cimetro01	1970	2	15
pankegr01	1970	11	6

playerID	birthYear	birthMon	birthDay
hlushto01	1970	2	7
cimetro01	1970	2	15
pankegr01	1970	11	6

birthYear	birthMon	birthDay
1970	2	2
	11	1

```
master.groupby(['birthYear', 'birthMon']).count()
```

birthYear	birthMon	birthDay	playerID
1970	2	7	hlushto01
		15	cimetro01
	11	6	pankegr01

birthYear	birthMon	birthDay	playerID
1970	2	7	hlushto01
		15	cimetro01
	11	6	pankegr01

```
master.loc[(1970, 2, 15), :]
```

	EUR/USD
24-03-2017 01:00:15	1.789
24-03-2017 01:00:30	1.790
24-03-2017 01:00:45	1.791

Time Series in Pandas

pandas.Timestamp

pandas.Timestamp

Represents point in time

pandas.Timestamp

Represents point in time

Many time-specific methods and attributes

pandas.Timestamp

Represents point in time

Many time-specific methods and attributes

Practically like Python's datetime.datetime

```
ts = pd.Timestamp("1997-03-04 14:21:54")
```

pandas.Timestamp

Fully time-aware data type

```
ts = pd.Timestamp("1997-03-04 14:21:54")

ts.year # Gives 1997
```

pandas.Timestamp
Fully time-aware data type

```python
ts = pd.Timestamp("1997-03-04 14:21:54")

ts.year # Gives 1997

ts.day_name() # Gives Tuesday
```

pandas.Timestamp
Fully time-aware data type

```
td = pd.Timedelta('3 days 02:13:10')
```

pandas.Timedelta

Difference between two dates/times

```python
p = pd.Period('1999-03', 'M')

p.start_time

p.end_time

p.start_time < ts < p.end_time
```

pandas.Period

Represents period of time

Data type	Index type

Data type	Index type
Timestamp	DatetimeIndex

Data type	Index type
Timestamp	DatetimeIndex
Timedelta	TimedeltaIndex

Data type	Index type
Timestamp	DatetimeIndex
Timedelta	TimedeltaIndex
Period	PeriodIndex

Other Data Types

Advanced Strategies

Summary

You are ready to work with time series

Domain is complex

Remember about the docs