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

birthDay

Dirtifear	DIFTHMOH	
1970	2	2
	11	1

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



playerID

birthYear birthMon birthDay

hlushto01	7	2	1970
cimetro01	15		
pankegr01	6	11	



playerID

birthYear birthMon birthDay

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



The need for a better representation



Time Series in Pandas





Represents point in time



Represents point in time

Many timespecific methods and attributes



Represents point in time

Many timespecific 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
```

Fully time-aware data type



```
ts = pd.Timestamp("1997-03-04 14:21:54")
ts.year # Gives 1997
ts.day_name() # Gives Tuesday
```

Fully time-aware data type



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

pandas.Timedelta

Difference between two dates/times

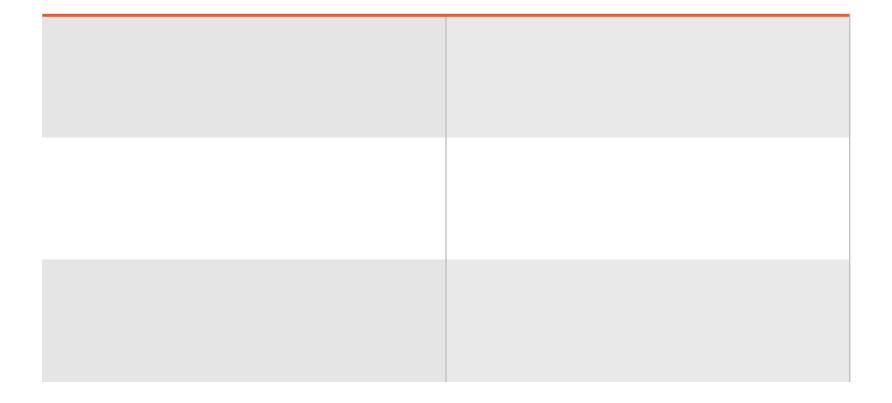


```
p = pd.Period('1999-03', 'M')
p.start_time
p.end_time
p.start_time < ts < p.end_time</pre>
```

pandas.Period

Represents period of time







Timestamp	DatetimeIndex



Timestamp	DatetimeIndex
Timedelta	TimedeltaIndex



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

