Multi-level Indexes

Paweł Kordek
SOFTWARE ENGINEER

@pawel_kordek https://kordek.github.io

Summary

Why they exist

How it works

MultiIndex API

The Need for Multi-level Index

playerID	firstName	lastName	pos
aaltoan01	Antti	Aalto	C
abdelju01	Justin	Abdelkader	L
abidra01	Ramzi	Abid	L

playerID	firstName	lastName	pos
aaltoan01	Antti	Aalto	C
abdelju01	Justin	Abdelkader	L
abidra01	Ramzi	Abid	L

```
master.loc['aaltoan01', :]
```

playerID	year	tmID	goals
aaltoan01	1997	ANA	0
aaltoan01	1998	ANA	3
aaltoan01	1999	ANA	7

playerID	year	tmID	goals
aaltoan01	1997	ANA	0
aaltoan01	1998	ANA	3
aaltoan01	1999	ANA	7

```
scoring
    .loc[scoring['year'] == 1997, :]
    .loc['aaltoan01', :]
```

playerID	year	tmID	goals
aaltoan01	1997	ANA	0
	1998	ANA	3
	1999	ANA	7

level 0	level 1	tmID	goals
playerID	year		
aaltoan01	1997	ANA	0
	1998	ANA	3
	1999	ANA	7

Creating MultiIndex

Just pass a list of column names

```
scoring.set_index(['playerID', 'year'])
```

Creating MultiIndex

Just pass a list of column names

```
scoring.groupby(level=1)['goals'].max()
scoring.groupby(level='year')['goals'].max()
```

Grouping Using MultiIndex

Pass a label or position

```
idx = pd.IndexSlice
```

Flexible Indexing

Slices to the rescue

```
idx = pd.IndexSlice
scoring.loc[idx['aaltoan01', 1997:2000], :]
```

Flexible Indexing

Slices to the rescue

```
idx = pd.IndexSlice
scoring.loc[idx['aaltoan01', 1997:2000], :]
scoring.loc[idx[:, 1997:2000], :]
```

Flexible Indexing

Slices to the rescue

`scoring.sort_index()`

Slicing

Remember to order your index!

MultiIndex

It’s nothing scary

It simplifies life

It’s the Pandas way of thinking