

Group Operations



Paweł Kordek

SOFTWARE ENGINEER

@pawel_kordek

<https://kordek.github.io>



Overview



Motivation

Iterating through groups

Built-in methods



Motivation



Min Values

	artist	...	acquisitionYear
15819	Walker, John	...	1975
15820	Walker, John	...	1975
15851	Walker, John	...	1979
672	Bacon, Francis	...	1950
673	Bacon, Francis	...	1952



Min Values

artist

...

acquisitionYear

15819	Walker, John	...	1975
15820	Walker, John	...	1975
15851	Walker, John	...	1979
672	Bacon, Francis	...	1950
673	Bacon, Francis	...	1952



Aggregation



Missing Values

	artist	...	medium
15819	Walker, John	...	Screenprint on paper
15820	Walker, John	...	Screenprint on paper
15851	Walker, John	...	nan
672	Bacon, Francis	...	Oil paint on canvas
673	Bacon, Francis	...	Oil paint on canvas



Missing Values

artist

...

medium

15819	Walker, John	...	Screenprint on paper
15820	Walker, John	...	Screenprint on paper
15851	Walker, John	...	Screenprint on paper
672	Bacon, Francis	...	Oil paint on canvas
673	Bacon, Francis	...	Oil paint on canvas



Transformation



Dropping Groups

	artist	...	title
92330	Hughes, Arthur	...	The Singer
4504	Ford, Edward Onslow	...	The Singer

Filtering



Iterating Over Groups



for name, group in df.groupby('artist'):

	artist	...	acquisitionYear
15819	Walker, John	...	1975
15820	Walker, John	...	1975
15851	Walker, John	...	1979
672	Bacon, Francis	...	1950
673	Bacon, Francis	...	1952

name: 'Walker, John'

← group

name: 'Bacon, Francis'



Demo



We will:

- Iterate over artist groups
- Find the date of the first acquisition for each artist
- Fill medium missing values



Built-in Methods



```
grouped_medium = df.groupby('artist')['medium']  
df.loc[:, 'medium'] = grouped_medium.transform(fill_values)
```

Built-ins

Transformation




```
grouped_acq_year = df.groupby('artist')['acquisitionYear']  
min_acquisition_years = grouped_acq_year.agg(np.min)  
min_acquisition_years = grouped_acq_year.min()
```

Built-ins
Aggregation



```
grouped_titles = df.groupby('titles')  
condition = lambda x: len(x.index) > 1  
filtered_df = grouped_titles.filter(condition)
```

Built-ins

Filtering



Demo



We will:

- Perform aggregation, transformation and filtering using built-in methods

