

04_transform_data

March 21, 2022

1 Transform Data

Abid Ali

<https://github.com/abid-2362>

```
[ ]: import pandas as pd
data = pd.read_csv('./data/artwork_sample.csv')
```

```
[ ]: data.head()
```

```
[ ]:
      id accession_number      artist artistRole  artistId \
0  1035          A00001  Blake, Robert    artist        38
1  1036          A00002  Blake, Robert    artist        38
2  1037          A00003  Blake, Robert    artist        38
3  1038          A00004  Blake, Robert    artist        38
4  1039          A00005  Blake, William    artist        39

                                     title      dateText \
0  A Figure Bowing before a Seated Old Man with h...  date not known
1  Two Drawings of Frightened Figures, Probably f...  date not known
2  The Preaching of Warning. Verso: An Old Man En...    ?c.1785
3  Six Drawings of Figures with Outstretched Arms    date not known
4  The Circle of the Lustful: Francesca da Rimini... 1826-7, reprinted 1892

                                     medium \
0  Watercolour, ink, chalk and graphite on paper...
1                                     Graphite on paper
2  Graphite on paper. Verso: graphite on paper
3                                     Graphite on paper
4  Line engraving on paper

                                     creditLine  year  acquisitionYear \
0  Presented by Mrs John Richmond 1922    NaN        1922
1  Presented by Mrs John Richmond 1922    NaN        1922
2  Presented by Mrs John Richmond 1922  1785.0        1922
3  Presented by Mrs John Richmond 1922    NaN        1922
4  Purchased with the assistance of a special gra... 1826.0        1919
```

		dimensions	width	height	depth	units	inscription	\
0	support:	394 x 419 mm	394.0	419.0	NaN	mm	NaN	
1	support:	311 x 213 mm	311.0	213.0	NaN	mm	NaN	
2	support:	343 x 467 mm	343.0	467.0	NaN	mm	NaN	
3	support:	318 x 394 mm	318.0	394.0	NaN	mm	NaN	
4	image:	243 x 335 mm	243.0	335.0	NaN	mm	NaN	

	thumbnailCopyright	thumbnailUrl	\
0	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
1	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
2	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
3	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
4	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	

	url
0	http://www.tate.org.uk/art/artworks/blake-a-fi...
1	http://www.tate.org.uk/art/artworks/blake-two-...
2	http://www.tate.org.uk/art/artworks/blake-the-...
3	http://www.tate.org.uk/art/artworks/blake-six-...
4	http://www.tate.org.uk/art/artworks/blake-the-...

```
[ ]: data.height.transform(lambda x: print(x))
# because we did not return anything in lambda function, so return type is None
↳ by default
```

```
419.0
213.0
467.0
394.0
335.0
338.0
334.0
340.0
335.0
340.0
```

```
[ ]: 0    None
      1    None
      2    None
      3    None
      4    None
      5    None
      6    None
      7    None
      8    None
      9    None
      Name: height, dtype: object
```

```
[ ]: data.height.transform(lambda x: x)
```

```
[ ]: 0    419.0
      1    213.0
      2    467.0
      3    394.0
      4    335.0
      5    338.0
      6    334.0
      7    340.0
      8    335.0
      9    340.0
      Name: height, dtype: float64
```

```
[ ]: # we can transform the data by using this technique
      data.height.transform(lambda x: x/10) # convert height into something useful e.
      ↪g. mm or cm, or meters etc.
```

```
[ ]: 0    41.9
      1    21.3
      2    46.7
      3    39.4
      4    33.5
      5    33.8
      6    33.4
      7    34.0
      8    33.5
      9    34.0
      Name: height, dtype: float64
```

```
[ ]: # Again because we did not replace the value, so the original data has not been
      ↪modified
      data.height
```

```
[ ]: 0    419.0
      1    213.0
      2    467.0
      3    394.0
      4    335.0
      5    338.0
      6    334.0
      7    340.0
      8    335.0
      9    340.0
      Name: height, dtype: float64
```

```
[ ]: # transform is really helpful when used with groupby
data.groupby('artist')
```

```
[ ]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000017211FCDF00>
```

```
[ ]: # number of unique values per row and per column, grouped by artist
data.groupby('artist').transform('nunique')
```

```
[ ]:
```

	id	accession_number	artistRole	artistId	title	dateText	medium	\
0	4	4	1	1	4	2	3	
1	4	4	1	1	4	2	3	
2	4	4	1	1	4	2	3	
3	4	4	1	1	4	2	3	
4	6	6	1	1	6	1	1	
5	6	6	1	1	6	1	1	
6	6	6	1	1	6	1	1	
7	6	6	1	1	6	1	1	
8	6	6	1	1	6	1	1	
9	6	6	1	1	6	1	1	

	creditLine	year	acquisitionYear	dimensions	width	height	depth	units	\
0	1	1	1	4	4	4	0	1	
1	1	1	1	4	4	4	0	1	
2	1	1	1	4	4	4	0	1	
3	1	1	1	4	4	4	0	1	
4	1	1	1	6	5	4	0	1	
5	1	1	1	6	5	4	0	1	
6	1	1	1	6	5	4	0	1	
7	1	1	1	6	5	4	0	1	
8	1	1	1	6	5	4	0	1	
9	1	1	1	6	5	4	0	1	

	inscription	thumbnailCopyright	thumbnailUrl	url
0	0	0	4	4
1	0	0	4	4
2	0	0	4	4
3	0	0	4	4
4	0	0	6	6
5	0	0	6	6
6	0	0	6	6
7	0	0	6	6
8	0	0	6	6
9	0	0	6	6

```
[ ]: data.groupby('artist')['height']
```

```
[ ]: <pandas.core.groupby.generic.SeriesGroupBy object at 0x0000017212445660>
```

```
[ ]: data.groupby('artist')['height'].mean()
```

```
[ ]: artist
      Blake, Robert      373.25
      Blake, William    337.00
      Name: height, dtype: float64
```

```
[ ]: data.groupby('artist')['height'].transform('mean')
```

```
[ ]: 0    373.25
      1    373.25
      2    373.25
      3    373.25
      4    337.00
      5    337.00
      6    337.00
      7    337.00
      8    337.00
      9    337.00
      Name: height, dtype: float64
```

```
[ ]: data.artist
```

```
[ ]: 0    Blake, Robert
      1    Blake, Robert
      2    Blake, Robert
      3    Blake, Robert
      4    Blake, William
      5    Blake, William
      6    Blake, William
      7    Blake, William
      8    Blake, William
      9    Blake, William
      Name: artist, dtype: object
```

```
[ ]: data['mean_height_by_artist'] = data.groupby('artist')['height'].
      ↪transform('mean')
      data
```

```
[ ]:   id accession_number      artist artistRole  artistId \
0  1035      A00001  Blake, Robert    artist        38
1  1036      A00002  Blake, Robert    artist        38
2  1037      A00003  Blake, Robert    artist        38
3  1038      A00004  Blake, Robert    artist        38
4  1039      A00005  Blake, William    artist        39
5  1040      A00006  Blake, William    artist        39
6  1041      A00007  Blake, William    artist        39
7  1042      A00008  Blake, William    artist        39
```

8	1043	A00009	Blake, William	artist	39
9	1044	A00010	Blake, William	artist	39

	title	dateText	\
0	A Figure Bowing before a Seated Old Man with h...	date not known	
1	Two Drawings of Frightened Figures, Probably f...	date not known	
2	The Preaching of Warning. Verso: An Old Man En...	?c.1785	
3	Six Drawings of Figures with Outstretched Arms	date not known	
4	The Circle of the Lustful: Francesca da Rimini...	1826-7, reprinted 1892	
5	Ciampolo the Barrator Tormented by the Devils	1826-7, reprinted 1892	
6	The Baffled Devils Fighting	1826-7, reprinted 1892	
7	The Six-Footed Serpent Attacking Agnolo Brunel...	1826-7, reprinted 1892	
8	The Serpent Attacking Buoso Donati	1826-7, reprinted 1892	
9	The Pit of Disease: The Falsifiers	1826-7, reprinted 1892	

	medium	\
0	Watercolour, ink, chalk and graphite on paper...	
1	Graphite on paper	
2	Graphite on paper. Verso: graphite on paper	
3	Graphite on paper	
4	Line engraving on paper	
5	Line engraving on paper	
6	Line engraving on paper	
7	Line engraving on paper	
8	Line engraving on paper	
9	Line engraving on paper	

	creditLine	year	...	\
0	Presented by Mrs John Richmond 1922	NaN	...	
1	Presented by Mrs John Richmond 1922	NaN	...	
2	Presented by Mrs John Richmond 1922	1785.0	...	
3	Presented by Mrs John Richmond 1922	NaN	...	
4	Purchased with the assistance of a special gra...	1826.0	...	
5	Purchased with the assistance of a special gra...	1826.0	...	
6	Purchased with the assistance of a special gra...	1826.0	...	
7	Purchased with the assistance of a special gra...	1826.0	...	
8	Purchased with the assistance of a special gra...	1826.0	...	
9	Purchased with the assistance of a special gra...	1826.0	...	

	dimensions	width	height	depth	units	inscription	\
0	support: 394 x 419 mm	394.0	419.0	NaN	mm	NaN	
1	support: 311 x 213 mm	311.0	213.0	NaN	mm	NaN	
2	support: 343 x 467 mm	343.0	467.0	NaN	mm	NaN	
3	support: 318 x 394 mm	318.0	394.0	NaN	mm	NaN	
4	image: 243 x 335 mm	243.0	335.0	NaN	mm	NaN	
5	image: 240 x 338 mm	240.0	338.0	NaN	mm	NaN	
6	image: 242 x 334 mm	242.0	334.0	NaN	mm	NaN	

7	image:	246 x 340 mm	246.0	340.0	NaN	mm	NaN
8	image:	241 x 335 mm	241.0	335.0	NaN	mm	NaN
9	image:	243 x 340 mm	243.0	340.0	NaN	mm	NaN

	thumbnailCopyright	thumbnailUrl	\
0	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
1	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
2	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
3	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
4	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
5	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
6	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
7	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
8	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	
9	NaN	http://www.tate.org.uk/art/images/work/A/A00/A...	

	url	mean_height_by_artist
0	http://www.tate.org.uk/art/artworks/blake-a-fi...	373.25
1	http://www.tate.org.uk/art/artworks/blake-two-...	373.25
2	http://www.tate.org.uk/art/artworks/blake-the-...	373.25
3	http://www.tate.org.uk/art/artworks/blake-six-...	373.25
4	http://www.tate.org.uk/art/artworks/blake-the-...	337.00
5	http://www.tate.org.uk/art/artworks/blake-ciam...	337.00
6	http://www.tate.org.uk/art/artworks/blake-the-...	337.00
7	http://www.tate.org.uk/art/artworks/blake-the-...	337.00
8	http://www.tate.org.uk/art/artworks/blake-the-...	337.00
9	http://www.tate.org.uk/art/artworks/blake-the-...	337.00

[10 rows x 21 columns]

[]: