melt-pivot

October 4, 2023

[70]:
```
import pandas as pd
baby = pd.read_csv('https://raw.githubusercontent.com/ernbilen/Data200_Fall23/
 ↪main/data/babynames_melt.csv')
baby
```

[70]:

	sex	2015	2016	2017
0	Female	Emma	Emma	Emma
1	Male	Noah	Noah	Liam

[79]:
```
baby.melt(id_vars=['sex'], var_name=['year'],
                    value_name='name') #id_vars is the level of identifying␣
 ↪row, var_name is the var we will create
# name is the third col we create which will have baby names
```

[79]:

	sex	year	name
0	Female	2015	Emma
1	Male	2015	Noah
2	Female	2016	Emma
3	Male	2016	Noah
4	Female	2017	Emma
5	Male	2017	Liam

[1]:
```
# pivot
df = pd.read_csv('/users/bilene/downloads/ind_usa.csv')
```

[2]:
```
df
```

[2]:

	code	year	indicator	value
0	IND	2005	pop	1147.61
1	IND	2010	pop	1234.28
2	IND	2015	pop	1310.15
3	USA	2005	pop	295.52
4	USA	2010	pop	309.33
5	USA	2015	pop	320.74
6	IND	2005	gdp	820.38
7	IND	2010	gdp	1675.62
8	IND	2015	gdp	2103.59
9	USA	2005	gdp	13036.64

```
10  USA  2010       gdp  14992.05
11  USA  2015       gdp  18219.30
```

[12]:
```
df.pivot(index=['code','year'],columns='indicator') # index is the lvl of data,␣
 ↪columns is where your cols
# will come from
```

[12]:
```
              value
indicator       gdp      pop
code year
IND  2005    820.38  1147.61
     2010   1675.62  1234.28
     2015   2103.59  1310.15
USA  2005  13036.64   295.52
     2010  14992.05   309.33
     2015  18219.30   320.74
```

[36]:
```
# pivot_table
df2 = pd.read_csv('/users/bilene/downloads/ind_usa2.csv').
 ↪sort_values(['code','year','indicator'])
df2
```

[36]:
```
    code  year indicator     value
6    IND  2005       gdp    820.38
18   IND  2005       gdp     80.38
0    IND  2005       pop   1147.61
12   IND  2005       pop    147.61
7    IND  2010       gdp   1675.62
19   IND  2010       gdp    175.62
1    IND  2010       pop   1234.28
13   IND  2010       pop    134.28
8    IND  2015       gdp   2103.59
20   IND  2015       gdp    203.59
2    IND  2015       pop   1310.15
14   IND  2015       pop    110.15
9    USA  2005       gdp  13036.64
21   USA  2005       gdp   1036.64
3    USA  2005       pop    295.52
15   USA  2005       pop     25.52
10   USA  2010       gdp  14992.05
22   USA  2010       gdp   1992.05
4    USA  2010       pop    309.33
16   USA  2010       pop     39.33
11   USA  2015       gdp  18219.30
23   USA  2015       gdp   1219.30
5    USA  2015       pop    320.74
17   USA  2015       pop     30.74
```

[38]:
```
# df2.pivot(index=['code','year'],columns='indicator') # this throws error. ask␣
↪why?
```

[49]:
```
df2.pivot_table(index=['code','year'], columns=['indicator'], aggfunc='mean').
↪reset_index()
```

[49]:

	code	year	value	
indicator			gdp	pop
0	IND	2005	450.38	647.61
1	IND	2010	925.62	684.28
2	IND	2015	1153.59	710.15
3	USA	2005	7036.64	160.52
4	USA	2010	8492.05	174.33
5	USA	2015	9719.30	175.74

[82]:
```
# practice time
data = {
    'Student_ID': [1, 2, 3, 4, 5],
    'Name': ['John', 'Jane', 'Bob', 'Alice', 'Eve'],
    'Math': [85, 92, 78, 88, 90],
    'Science': [90, 88, 75, 82, 95],
    'English': [75, 80, 92, 88, 70]
}

students = pd.DataFrame(data)
students
```

[82]:

	Student_ID	Name	Math	Science	English
0	1	John	85	90	75
1	2	Jane	92	88	80
2	3	Bob	78	75	92
3	4	Alice	88	82	88
4	5	Eve	90	95	70

[142]:
```
data = {
    'Student_ID': [1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5],
    'Name': ['John', 'Jane', 'Bob', 'Alice', 'Eve', 'John', 'Jane', 'Bob',␣
↪'Alice', 'Eve', 'John', 'Jane', 'Bob', 'Alice', 'Eve'],
    'Subject': ['Math', 'Math', 'Math', 'Math', 'Math', 'Science', 'Science',␣
↪'Science', 'Science', 'Science', 'English', 'English', 'English', 'English',␣
↪'English'],
    'Score': [85, 92, 78, 88, 90, 90, 88, 75, 82, 95, 75, 80, 92, 88, 70]
}

students = pd.DataFrame(data)
students
```

[142]:

	Student_ID	Name	Subject	Score
0	1	John	Math	85
1	2	Jane	Math	92
2	3	Bob	Math	78
3	4	Alice	Math	88
4	5	Eve	Math	90
5	1	John	Science	90
6	2	Jane	Science	88
7	3	Bob	Science	75
8	4	Alice	Science	82
9	5	Eve	Science	95
10	1	John	English	75
11	2	Jane	English	80
12	3	Bob	English	92
13	4	Alice	English	88
14	5	Eve	English	70

[151]:
```
# Should you pivot or melt?
pivot_df = students.pivot(index='Student_ID', columns='Subject',␣
 ↪values='Score').reset_index()
pivot_df.columns.name = None # to remove index name
pivot_df
```

[151]:

	Student_ID	English	Math	Science
0	1	75	85	90
1	2	80	92	88
2	3	92	78	75
3	4	88	88	82
4	5	70	90	95

[154]:
```
# Creating a sample wide DataFrame
data = {
    'Country': ['USA', 'Canada', 'Mexico'],
    '2000': [100, 200, 300],
    '2005': [150, 250, 350],
    '2010': [200, 300, 400],
    '2015': [250, 350, 450]
}

countries = pd.DataFrame(data)
countries
```

[154]:

	Country	2000	2005	2010	2015
0	USA	100	150	200	250
1	Canada	200	250	300	350
2	Mexico	300	350	400	450

[162]:
```
# Should you melt or pivot?
melted_df = pd.melt(countries, id_vars=['Country'], var_name='Year',␣
 ↪value_name='Gdp').sort_values('Country')\
                                          .reset_index(drop=True)
melted_df
```

[162]:

	Country	Year	Gdp
0	Canada	2000	200
1	Canada	2005	250
2	Canada	2010	300
3	Canada	2015	350
4	Mexico	2000	300
5	Mexico	2005	350
6	Mexico	2010	400
7	Mexico	2015	450
8	USA	2000	100
9	USA	2005	150
10	USA	2010	200
11	USA	2015	250

[ ]:
```
# joining data with pandas, reshaping data with .melt()
# reshaping data with pandas,
```