

## 1. 数据整理

练习1: melt/pivot 函数使用

练习2: stack/unstack 函数的使用

练习3: wide\_to\_long 函数的使用

## 2. 数据透视表

练习1: 统计不同种类产品在不同城市的销量(分别使用分组聚合和透视表实现)

练习2: 统计不同销售渠道中男性和女性客户的总销售额(分别使用分组聚合和透视表实现)

练习3: 统计不同销售渠道销售的订单数和总销售额

## 3. 数据的导入和导出

练习1: pandas 从 winfunc 数据库导入 movie 表和 review 表的数据

练习2: pandas 加载之后, 将左侧和右侧数据按照电影 ID 进行关联

练习3: 统计不同电影的用户平均分, 并将统计结果导出到 movie\_rating 数据表中

## 1. 数据整理

### 练习1: melt/pivot 函数使用

注意: 使用 work1.csv 数据集

	day	lucknow	delhi	kolkata
0	Monday	65	65	56
1	Tuesday	45	66	58
2	Wednesday	21	68	60
3	Thursday	45	75	80
4	Friday	36	78	83
5	Saturday	44	82	85
6	Sunday	35	48	26

**需求1:** 使用 melt 方法将 work1.csv 的数据转换为如下长数据

	day	city	count
0	Monday	lucknow	65
1	Tuesday	lucknow	45
2	Wednesday	lucknow	21
3	Thursday	lucknow	45
4	Friday	lucknow	36
5	Saturday	lucknow	44
6	Sunday	lucknow	35
7	Monday	delhi	65
8	Tuesday	delhi	66
9	Wednesday	delhi	68
10	Thursday	delhi	75
11	Friday	delhi	78
12	Saturday	delhi	82
13	Sunday	delhi	48
14	Monday	kolkata	56
15	Tuesday	kolkata	58
16	Wednesday	kolkata	60
17	Thursday	kolkata	80
18	Friday	kolkata	83
19	Saturday	kolkata	85
20	Sunday	kolkata	26

需求2：使用 pivot 方法将需求1的结果再转换为如下宽数据

city	day	delhi	kolkata	lucknow
0	Friday	78	83	36
1	Monday	65	56	65
2	Saturday	82	85	44
3	Sunday	48	26	35
4	Thursday	75	80	45
5	Tuesday	66	58	45
6	Wednesday	68	60	21

## 练习2：stack/unstack 函数的使用

注意：使用 work1.csv 数据集

	day	lucknow	delhi	kolkata
0	Monday	65	65	56
1	Tuesday	45	66	58
2	Wednesday	21	68	60
3	Thursday	45	75	80
4	Friday	36	78	83
5	Saturday	44	82	85
6	Sunday	35	48	26

需求1：使用 stack 实现练习1中需求1的效果如下：

	day	city	count
0	Monday	lucknow	65
1	Tuesday	lucknow	45
2	Wednesday	lucknow	21
3	Thursday	lucknow	45
4	Friday	lucknow	36
5	Saturday	lucknow	44
6	Sunday	lucknow	35
7	Monday	delhi	65
8	Tuesday	delhi	66
9	Wednesday	delhi	68
10	Thursday	delhi	75
11	Friday	delhi	78
12	Saturday	delhi	82
13	Sunday	delhi	48
14	Monday	kolkata	56
15	Tuesday	kolkata	58
16	Wednesday	kolkata	60
17	Thursday	kolkata	80
18	Friday	kolkata	83
19	Saturday	kolkata	85
20	Sunday	kolkata	26

### 练习3: wide\_to\_long 函数的使用

注意：使用 work2.csv 数据集

	famid	birth	ht_1	ht_2
0	1	1	2.8	3.4
1	1	2	2.9	3.8
2	1	3	2.2	2.9
3	2	1	2.0	3.2
4	2	2	1.8	2.8
5	2	3	1.9	2.4
6	3	1	2.2	3.3
7	3	2	2.3	3.4
8	3	3	2.1	2.9

**需求：**使用 wide\_to\_long 将 work2.csv 数据转换为如下长数据

	famid	birth	age	ht
0	1	1	1	2.8
1	1	1	2	3.4
2	1	2	1	2.9
3	1	2	2	3.8
4	1	3	1	2.2
5	1	3	2	2.9
6	2	1	1	2.0
7	2	1	2	3.2
8	2	2	1	1.8
9	2	2	2	2.8
10	2	3	1	1.9
11	2	3	2	2.4
12	3	1	1	2.2
13	3	1	2	3.3
14	3	2	1	2.3
15	3	2	2	3.4
16	3	3	1	2.1
17	3	3	2	2.9

## 2. 数据透视表

注意：使用 uniqlo.csv 数据集

**练习1：统计不同种类产品在不同城市的销量(分别使用分组聚合和透视表实现)**

product	T恤	当季新品	毛衣	牛仔裤	短裤	袜子	裙子	运动	配件
channel									
线上	3254	866	248	415	626	736	158	295	1054
线下	15171	4472	1108	2017	2195	2903	837	1499	3568

分组聚合实现：

pivot\_table 透视表实现：

**练习2：统计不同销售渠道中男性和女性客户的总销售额(分别使用分组聚合和透视表实现)**

gender_group	Female	Male	Unkown
channel			
线上	449266.64	201011.84	2892.00
线下	2078492.70	817558.81	7210.86

分组聚合实现：

pivot\_table 透视表实现：

### 练习3：统计不同销售渠道销售的订单数和总销售额

	order	revenue
channel		
线上	6823	653170.48
线下	30005	2903262.37

分组聚合实现：

pivot\_table 透视表实现：

## 3. 数据的导入和导出

练习1：pandas 从 winfunc 数据库中导入 movie 表和 review 表的数据

movie @winfunc (localhost) - 表				
电影表				
id	title	release_year	genre	editor_rating
1	阿凡达	2009	科幻	8
2	泰坦尼克号	1997	剧情	10
3	教父	1972	剧情	8
4	艳舞女郎	1995	纪录片	3
5	外太空计划9	1959	科幻	2

review @winfunc (localhost) - 表			
用户评分表			
id	rating	customer_id	movie_id
1	7	2	1
2	8	6	1
3	10	1	2
4	10	2	2
5	8	3	2
6	9	1	2
7	9	3	3
8	9	7	3
9	8	2	3
10	6	5	3
11	3	5	4
12	3	3	4
13	5	6	4
14	1	2	4
15	1	1	5
16	1	3	5
17	3	4	5
18	2	7	5

练习2：pandas 加载之后，将左侧和右侧数据按照电影 ID 进行关联

	id_x	title	release_year	genre	editor_rating	id_y	rating	customer_id	movie_id
0	1	阿凡达	2009	科幻	8	1	7	2	1
1	1	阿凡达	2009	科幻	8	2	8	6	1
2	2	泰坦尼克号	1997	剧情	10	3	10	1	2
3	2	泰坦尼克号	1997	剧情	10	4	10	2	2
4	2	泰坦尼克号	1997	剧情	10	5	8	3	2
5	2	泰坦尼克号	1997	剧情	10	6	9	1	2
6	3	教父	1972	剧情	8	7	9	3	3
7	3	教父	1972	剧情	8	8	9	7	3
8	3	教父	1972	剧情	8	9	8	2	3
9	3	教父	1972	剧情	8	10	6	5	3
10	4	艳舞女郎	1995	纪录片	3	11	3	5	4
11	4	艳舞女郎	1995	纪录片	3	12	3	3	4
12	4	艳舞女郎	1995	纪录片	3	13	5	6	4
13	4	艳舞女郎	1995	纪录片	3	14	1	2	4
14	5	外太空计划9	1959	科幻	2	15	1	1	5
15	5	外太空计划9	1959	科幻	2	16	1	3	5
16	5	外太空计划9	1959	科幻	2	17	3	4	5
17	5	外太空计划9	1959	科幻	2	18	2	7	5

### 练习3：统计不同电影的用户平均分，并将统计结果导出到 movie\_rating 数据表中

	movie_id	title	rating
0	1	阿凡达	7.50
1	2	泰坦尼克号	9.25
2	3	教父	8.00
3	4	艳舞女郎	3.00
4	5	外太空计划9	1.75