

Readme4

数据合并

请编写Python程序将其中相同类别的Excel表合并到一起，有利于今后管理和分析数据

数据类型

源文件为aa.zip，包含12个月的图书销售记录，每月分别保存在一个*.xlsx中，每个sheet包含以下列：

时间 商品名称 SKU 品牌 一级类目 二级类目 三级类目 浏览量 访客数 人均浏览量 平均停留时长
成交商品件数 ISBN 首次入库时间 成交码洋 加购人数

各个xlsx表格直接进行轴向连接，按照时间进行排序

输出存储在result.xlsx的sort_by_time sheet中

```
filenames = os.listdir(dirpath)
# 调整文件的读取顺序，按照月份进行排序
filenames.sort(key=lambda x: int(re.sub('[^0-9]', '', x.split(".")[0])))
for filename in filenames:
    data = pd.read_excel(dirpath + filename, names=['时间', '商品名称', 'SKU', '品牌', '一级类目',
                                                    '二级类目', '三级类目', '浏览量', '访客数',
                                                    '人均浏览量', '平均停留时长', '成交商品件数',
                                                    'ISBN', '首次入库时间', '成交码洋',
                                                    '加购人数']) # [xxx rows x 16 columns]
    data.sort_values(by='时间', inplace=True) # sort by time
    data_total.append(data)

# Concatenate pandas objects in all xlsx files
df = pd.concat(data_total)
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	时间	商品名称	SKU	品牌	一级类目	二级类目	三级类目	浏览量	访客数	人均浏览量	均停留时	成交商品件	ISBN	次入库时	成交码洋	加购人数
2	2019-01-01 00:00:00	Android项目	12163105	其他品牌	图书	计算机与	编程语言	21	19	1	80	2	9.79E+12	#####	119.6	1
3	2019-01-01 00:00:00	Python从入门到精通	12451724	其他品牌	图书	计算机与	编程语言	810	355	2	95	36	9.79E+12	#####	3592.8	75
4	2019-01-01 00:00:00	零基础学Python	12353915	其他品牌	图书	计算机与	编程语言	519	258	2	74	28	9.79E+12	#####	2234.4	49
5	2019-01-01 00:00:00	零基础学Java	12250414	其他品牌	图书	计算机与	编程语言	396	167	2	78	20	9.79E+12	#####	1396	35
6	2019-01-01 00:00:00	零基础学C++	12185501	其他品牌	图书	计算机与	编程语言	278	104	3	112	26	9.79E+12	#####	1814.8	36
7	2019-01-01 00:00:00	SQL即查即学	12359944	其他品牌	图书	计算机与	数据库	210	109	2	51	12	9.79E+12	#####	597.6	26
8	2019-01-01 00:00:00	零基础学MySQL	12271986	其他品牌	图书	计算机与	编程语言	198	64	3	118	10	9.79E+12	#####	798	15
9	2019-01-01 00:00:00	Java项目实战	12163091	其他品牌	图书	计算机与	编程语言	164	62	3	82	12	9.79E+12	#####	717.6	14
10	2019-01-01 00:00:00	JavaWeb项目	12163151	其他品牌	图书	计算机与	编程语言	162	59	3	119	8	9.79E+12	#####	558.4	16
11	2019-01-01 00:00:00	C++项目实战	12163147	其他品牌	图书	计算机与	编程语言	154	72	2	111	7	9.79E+12	#####	488.6	10
12	2019-01-01 00:00:00	零基础学C#	12458548	其他品牌	图书	计算机与	编程语言	135	67	2	83	12	9.79E+12	#####	957.6	18
13	2019-01-01 00:00:00	零基础学C#	12271813	其他品牌	图书	计算机与	编程语言	134	51	3	82	8	9.79E+12	#####	638.4	11
14	2019-01-01 00:00:00	C#项目开发	12163129	其他品牌	图书	计算机与	编程语言	132	66	2	136	8	9.79E+12	#####	558.4	20
15	2019-01-01 00:00:00	Java精彩案例	12185937	其他品牌	图书	计算机与	编程语言	125	71	2	53	16	9.79E+12	#####	1276.8	11
16	2019-01-01 00:00:00	案例学Web	12449512	其他品牌	图书	计算机与	编程语言	107	55	2	69	3	9.79E+12	#####	149.4	11
17	2019-01-01 00:00:00	Android项目开发	12237622	其他品牌	图书	计算机与	编程语言	23	16	1	80	2	9.79E+12	#####	179.6	2
18	2019-01-01 00:00:00	C#精彩案例	12235813	其他品牌	图书	计算机与	编程语言	92	33	3	117	6	9.79E+12	#####	538.8	7
19	2019-01-01 00:00:00	JSP项目开发	12233850	其他品牌	图书	计算机与	编程语言	25	16	2	26	1	9.79E+12	#####	69.8	2
20	2019-01-01 00:00:00	Android项目开发	12402753	其他品牌	图书	计算机与	编程语言	30	19	2	93	3	9.79E+12	#####	357	3
21	2019-01-01 00:00:00	零基础学C#	12382901	其他品牌	图书	计算机与	编程语言	30	12	3	73	4	9.79E+12	#####	319.2	5
22	2019-01-01 00:00:00	玩转C语言	12418636	其他品牌	图书	计算机与	编程语言	33	19	2	60	2	9.79E+12	#####	99.6	4
23	2019-01-01 00:00:00	Java开发案例	12403477	其他品牌	图书	计算机与	编程语言	37	15	2	85	1	9.79E+12	#####	119	3
24	2019-01-01 00:00:00	ASP.NET项目开发	12163145	其他品牌	图书	计算机与	编程语言	40	21	2	115	1	9.79E+12	#####	79.8	3
25	2019-01-01 00:00:00	零基础学C#	12275164	其他品牌	图书	计算机与	编程语言	93	52	2	51	7	9.79E+12	#####	558.6	9
26	2019-01-01 00:00:00	PHP项目开发	12163093	其他品牌	图书	计算机与	网页制作	43	17	3	132	2	9.79E+12	#####	139.6	0
27	2019-01-01 00:00:00	零基础学C#	12206201	其他品牌	图书	计算机与	编程语言	47	21	2	120	2	9.79E+12	#####	159.6	3
28	2019-01-01 00:00:00	零基础学C#	12199033	其他品牌	图书	计算机与	编程语言	54	28	2	71	5	9.79E+12	#####	449	7
29	2019-01-01 00:00:00	ASP.NET项目开发	12163089	其他品牌	图书	计算机与	编程语言	59	33	2	77	3	9.79E+12	#####	209.4	6
30	2019-01-01 00:00:00	C语言项目	12163145	其他品牌	图书	计算机与	编程语言	81	43	2	98	5	9.79E+12	#####	299	7
31	2019-01-01 00:00:00	C语言项目	12199075	其他品牌	图书	计算机与	编程语言	82	40	2	111	7	9.79E+12	#####	558.6	8
32	2019-01-01 00:00:00	零基础学C#	12271911	其他品牌	图书	计算机与	编程语言	43	19	2	62	8	9.79E+12	#####	638.4	5
33	2019-01-02 00:00:00	C语言项目	12199075	其他品牌	图书	计算机与	编程语言	106	63	2	56	6	9.79E+12	#####	478.8	12
34	2019-01-02 00:00:00	JavaWeb项目开发	12163151	其他品牌	图书	计算机与	编程语言	122	62	2	89	3	9.79E+12	#####	209.4	11
35	2019-01-02 00:00:00	C#项目开发	12163129	其他品牌	图书	计算机与	编程语言	126	73	2	96	9	9.79E+12	#####	628.2	13

类目和名称做分层索引，按照时间进行排序

以一级类目、二级类目、三级类目、商品名称、时间做分层索引，结果样例如图所示，该样例输出存储在result.xlsx的sort_by_name sheet中

一级类目：[图书]

二级类目：[计算机与互联网]

三级类目：[编程语言与程序设计，数据库，网页制作/Web技术，移动开发]

```
df2 = df.set_index(['一级类目', '二级类目', '三级类目', '商品名称', '时间'])
df2.sort_values(by=['商品名称', '时间'], inplace=True)
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	一级类目	二级类目	三级类目	商品名称	时间	SKU	品牌	浏览量	访客数	人均浏览量	均停留时	成交商品件	ISBN	次入库时	成交码洋	加购人数		
2	图书	计算机与互联网	编程语言与程序设计	项目开发实战入门（全）	#####	12163089	其他品牌	59	33	2	77	3	9.79E+12	#####	209.4	6		
3					#####	12163089	其他品牌	52	31	2	161	8	9.79E+12	#####	558.4	10		
4					#####	12163089	其他品牌	104	35	3	153	5	9.79E+12	#####	349	7		
5					#####	12163089	其他品牌	71	36	2	140	3	9.79E+12	#####	209.4	5		
6					#####	12163089	其他品牌	70	30	2	166	3	9.79E+12	#####	209.4	8		
7					#####	12163089	其他品牌	67	33	2	144	33	9.79E+12	#####	2303.4	6		
8					#####	12163089	其他品牌	58	34	2	103	1	9.79E+12	#####	69.8	6		
9					#####	12163089	其他品牌	63	43	1	123	3	9.79E+12	#####	209.4	4		
10					#####	12163089	其他品牌	102	43	2	159	5	9.79E+12	#####	349	7		
11					#####	12163089	其他品牌	79	38	2	105	6	9.79E+12	#####	418.8	10		
12					#####	12163089	其他品牌	71	32	2	128	2	9.79E+12	#####	139.6	7		
13					#####	12163089	其他品牌	48	28	2	78	1	9.79E+12	#####	69.8	2		
14					#####	12163089	其他品牌	135	37	4	115	4	9.79E+12	#####	279.2	8		
15					#####	12163089	其他品牌	89	41	2	67	9	9.79E+12	#####	628.2	14		
16					#####	12163089	其他品牌	102	50	2	115	9	9.79E+12	#####	628.2	14		
17					#####	12163089	其他品牌	127	48	3	160	12	9.79E+12	#####	837.6	16		
18					#####	12163089	其他品牌	75	40	2	93	5	9.79E+12	#####	349	10		
19					#####	12163089	其他品牌	60	30	2	119	4	9.79E+12	#####	279.2	8		
20					#####	12163089	其他品牌	89	39	2	50	9	9.79E+12	#####	628.2	16		
21					#####	12163089	其他品牌	110	39	3	147	2	9.79E+12	#####	139.6	4		
22					#####	12163089	其他品牌	58	41	1	79	5	9.79E+12	#####	349	7		
23					#####	12163089	其他品牌	78	39	2	102	4	9.79E+12	#####	279.2	7		
24					#####	12163089	其他品牌	101	44	2	124	4	9.79E+12	#####	279.2	7		
25					#####	12163089	其他品牌	70	42	2	71	5	9.79E+12	#####	349	9		
26					#####	12163089	其他品牌	84	39	2	95	5	9.79E+12	#####	349	9		
27					#####	12163089	其他品牌	87	39	2	117	9	9.79E+12	#####	628.2	12		
28					#####	12163089	其他品牌	66	32	2	68	4	9.79E+12	#####	279.2	9		
29					#####	12163089	其他品牌	61	32	2	106	1	9.79E+12	#####	69.8	4		
30					#####	12163089	其他品牌	102	36	3	82	4	9.79E+12	#####	279.2	6		
31					#####	12163089	其他品牌	75	36	2	97	5	9.79E+12	#####	349	11		
32					#####	12163089	其他品牌	39	26	2	158	4	9.79E+12	#####	279.2	9		
33					#####	12163089	其他品牌	52	26	2	91	1	9.79E+12	#####	69.8	4		
34					#####	12163089	其他品牌	29	16	2	118	0	9.79E+12	#####	0	2		
35					#####	12163089	其他品牌	12	7	2	31	0	9.79E+12	#####	0	0		

进行初步数据统计

比如按照书名，对每本书的浏览量、访客数、成交商品件数等进行求和，求所有某书籍购买者的平均停留时长，结果样例如图所示，该样例输出存储在result.xlsx的statistical_data sheet中

```
df4 = pd.pivot_table(df, index=['name'],
                      values=['view', 'visitor', 'time_on_page',
                              'number_of_transactions'],
                      aggfunc={'view': np.sum, 'visitor': np.sum, 'time_on_page':
                               np.mean,
                              'number_of_transactions': np.sum}
                      # margins=True)
                      )
```

	A	B	C	D	E	F
1	name	number_of_transactions	time_on_page	view	visitor	
2	ASP.NET项目开发实战入门（全彩版）	1292	97.71584699	24279	12046	
3	Android开发详解（全彩版）	752	87.73569482	18404	9788	
4	Android精彩编程200例（全彩版）	1324	77.07629428	16658	8754	
5	Android项目开发实战入门（全彩版）	1719	72.60490463	28953	15217	
6	C#精彩编程200例（全彩版）	2773	76.6420765	26880	14801	
7	C#项目开发实战入门（全彩版）	3915	90.24043716	53949	26123	
8	C++项目开发实战入门（全彩版）	4054	73.86065574	100472	48943	
9	C语言精彩编程200例（全彩版）	5070	62.28142077	48495	24300	
10	C语言项目开发实战入门（全彩版）	4691	68.94262295	51571	25235	
11	JSP项目开发实战入门（全彩版）	618	82.29508197	12313	6217	
12	JavaWeb项目开发实战入门（全彩版）	2658	88.88797814	56003	25890	
13	Java开发详解（全彩版）	411	76.99453552	11830	6373	
14	Java精彩编程200例（全彩版）	4344	64.33606557	48783	24399	
15	Java项目开发实战入门（全彩版）	4809	65.93715847	71460	34363	
16	PHP项目开发实战入门（全彩版）	1345	87.98637602	25745	13044	
17	Python从入门到项目实践（全彩版）	33016	72.68032787	640860	291576	
18	Python编程锦囊（全彩版）	7297	63.25698324	67026	36363	
19	Python项目开发案例集锦（全彩版）	12474	78.8113879	163241	79802	
20	SQL即查即用（全彩版）	12436	58.69672131	185771	93529	
21	Visual Basic精彩编程200例（全彩版）	718	81.24043716	16271	8529	
22	案例学WEB前端开发（全彩版）	1926	65.35694823	41018	22811	
23	玩转C语言程序设计（全彩版）	437	59.2431694	11425	6780	
24	零基础学ASP.NET（全彩版）	993	92.7568306	15925	8455	
25	零基础学Android（全彩版）	2621	79.16076294	41119	19180	
26	零基础学C#（全彩版）	5038	79.43715847	62827	28212	
27	零基础学C++（全彩版）	7179	70.08469945	119216	53464	
28	零基础学C语言（全彩版）	16223	72.88797814	262088	110398	
29	零基础学HTML5+CSS3（全彩版）	4599	71.67302452	80066	37632	
30	零基础学JavaScript（全彩版）	4315	68.29427793	66111	31781	
31	零基础学Java（全彩版）	8522	70.91256831	120282	52145	
32	零基础学Oracle（全彩版）	2042	70.95095368	31541	16025	
33	零基础学PHP（全彩版）	1891	79.58469945	31609	14718	
34	零基础学Python（全彩版）	34495	68.06830601	598548	282610	
35						

sort_by_time

sort_by_name

statistical_data

test_data

+

7

◀