

# 第七课作业 1

在第 7 课的资料中，中有 sh600000 到 sh600114 这 100 只股票的历史日数据。以下对这些数据进行操作，需要完成如下三个任务。

## 任务一

先读取所有股票 2006 年 1 月 1 日开始到最后的的历史数据，接着按新浪行业进行分组，统计每个行业的股票数量（直接在程序中输出），如下图所示：

房地产	9
电力行业	7
交通运输	7
金融行业	6
公路桥梁	5
汽车制造	5
电子信息	5
生物制药	4
钢铁行业	4
化工行业	3
酿酒行业	3
开发区	2
服装鞋类	2
化纤行业	2
传媒娱乐	2
农林牧渔	2
造纸行业	2
建筑建材	2
机械行业	2
家电行业	1
船舶制造	1
电子器件	1

## 任务二

计算每个行业在每个交易日的涨跌幅（行业内股票涨跌幅的平均），最终输出如下图所示的一个 csv 文件，文件名为 output.csv。

交易日期	交通运输	传媒娱乐	公路桥梁	农林牧渔	农药化肥	化工行业	化纤行业	医疗器械	发电设备	商业百货	家电行业	建筑建材	开发区	房地产	摩托车	有色金属	服装
2006/1/4	0.003244	0.027999	0.000952	0.036115	0.017442	0.016803	0.014739	0	0.057772	0.029985	0.037707	0.017742	0.010716	0.013503	0.016598	0.019694	0.01
2006/1/5	0.040241	0.035324	0.024111	0.002142	0.012571	0.012911	0.010318	0.016	0.011704	-0.00291	0.008721	0.005551	0.038944	0.025733	0.02449	0.021459	0.01
2006/1/6	0.022135	-0.00036	0.017897	0.007937	0.030474	0.001648	0.008698	0.009843	0.015424	-0.00292	0.001441	0.005515	-0.00081	0.011776	0	-0.0063	-0.0
2006/1/9	-0.00017	0.012279	0.00124	0.008475	0	-0.00913	0.014611	-0.0039	0.002532	0.021962	0.035971	0.013292	0.017114	0.015478	0.011952	0.027484	0.01
2006/1/10	0.006127	-0.00942	-0.00951	-0.00054	0.004381	0.002417	-0.00205	-0.01761	0.020202	-0.01146	-0.00278	-0.00661	0.0079	0.007397	0.003937	0.024691	0.01
2006/1/11	-0.01123	0.032027	-0.00949	-0.01439	-0.00763	-0.01898	-0.01703	0.033865	-0.01114	-0.01884	0.02507	-0.01208	-0.00416	-0.01035	-0.01961	-0.00402	-0.0
2006/1/12	0.004682	0.014857	0.016519	0.023632	0.053846	0.021195	0.018217	0.00578	0.022528	0.031019	0.012228	0.023517	0.01462	0.022954	0.036	0.006048	0.01
2006/1/13	-0.013	0.005134	-0.00756	0.003337	0.008342	-0.00398	0.021511	-0.01533	-0.00979	0.010029	-0.00537	-0.00091	0.003597	0.004268	0.003861	-0.02204	0.03
2006/1/16	-0.01283	-0.04023	-0.02099	-0.00389	-0.04654	-0.01048	-0.01612	-0.01362	-0.04944	-0.02695	-0.04453	-0.02289	-0.03186	-0.00655	-0.01923	0.038934	-0.0
2006/1/17	-0.00092	-0.00025	0.011036	0.002943	-0.00542	-0.00056	-0.016	-0.00197	-0.0052	0.048105	0.001412	-0.00031	0.038005	0.006104	0.003922	-0.01381	
2006/1/18	0.01356	0.024868	0.018027	0.012087	0.038168	0.016251	0.032552	0.017787	0.023529	0.020862	0.03103	0.027077	0.023499	0.016382	0.023438	0.076	0.01
2006/1/19	-0.01148	0.016736	0.013583	0.005495	0.015756	0.022631	0.020197	0.009709	0.011494	0.002725	-0.00821	0.034859	0.024911	0.002512	0	0.009294	-0.0
2006/1/20	-0.02195	0.027735	-0.0133	-0.00535	0.005171	0.023871	-0.01346	0	0.045455	0.008152	-0.00138	0.011746	-0.01558	-0.00735	0.003817	0.033149	-0.0
2006/1/23	-0.00179	0.003876	0.003221	-0.00544	0.01749	-0.00258	0.001016	-0.00769	0.007246	-0.00674	0.041436	0.009542	0.001037	-0.00278	0.015209	0.065954	-0.0
2006/1/24	-0.00681	0.000965	0.001715	0.015385	-0.01213	0.015985	0.005023	0.003876	0.028777	-0.01628	0.027851	-0.00365	-0.01777	-0.01197	-0.01124	-0.04515	0.00
2006/1/25	0.00333	-0.00772	-0.01103	-0.0214	0.002047	0.006909	-0.00145	-0.01351	0.018648	-0.01103	0.012903	-0.00513	0.021212	-0.00857	0.003788	-0.02627	-0.0
2006/2/6	0.018074	0.029155	0.031347	0.014805	0.008172	0.01642	0.021356	0.013699	0	0.036262	0	0.015472	0.024876	0.028165	0.056604	0.100719	0.06
2006/2/7	0.004875	0.036827	-0.01255	0.042041	0.01925	0.020027	0.000976	0.03861	-0.04806	-0.03096	-0.01274	0.0324	0.002861	-0.0004	-0.01071	0.058824	-0.0
2006/2/8	0.018453	-0.00455	-0.00188	0.003871	0.022863	-0.01409	0.013788	-0.00558	-0.01442	-0.01111	-0.01677	0.020162	0.012261	0.03111	0.01083	-0.0571	0.00
2006/2/9	-0.00924	-0.03568	-0.0154	-0.0291	-0.02916	-0.01765	-0.0192	-0.02617	-0.03293	-0.00562	-0.02231	-0.02873	-0.01222	-0.00124	-0.025	0.006547	0.02
2006/2/10	0.005643	0.01518	0.028417	0.010249	0.016016	0.03367	0.01237	0.007678	0.016393	0.009897	0.020134	0.014905	0.019939	0.012584	0.003663	-0.01301	0.01

### 任务三

统计每个月涨跌幅最大行业，输出行业间的相关系数矩阵，选出相关性最高的两个行业（直接在程序中输出），如下图所示：

交易日期	
2006-01-31	有色金属
2006-02-28	酿酒行业
2006-03-31	酿酒行业
2006-04-30	有色金属
2006-05-31	服装鞋类
2006-06-30	船舶制造
2006-07-31	服装鞋类
2006-08-31	飞机制造
2006-09-30	电子器件
2006-10-31	钢铁行业
2006-11-30	钢铁行业
2006-12-31	有色金属
2007-01-31	农林牧渔
2007-02-28	船舶制造
2007-03-31	服装鞋类
2007-04-30	建筑建材
2007-05-31	医疗器械
2007-06-30	发电设备
2007-07-31	食品行业
2007-08-31	船舶制造

新浪行业	交通运输	传媒娱乐	公路桥梁	农林牧渔	农药化肥	化工行业	化纤行业			
新浪行业										
交通运输	1.000000	0.637453	0.833165	0.671671	0.476295	0.741106	0.770169	0.592919	0.541358	0.674408
传媒娱乐	0.637453	1.000000	0.651648	0.509452	0.392222	0.646489	0.563334	0.680742	0.552574	0.500004
公路桥梁	0.833165	0.651648	1.000000	0.665703	0.460285	0.747058	0.752948	0.653398	0.544669	0.610539
农林牧渔	0.671671	0.509452	0.665703	1.000000	0.393455	0.666586	0.667844	0.548059	0.480807	0.555122
农药化肥	0.476295	0.392222	0.460285	0.393455	1.000000	0.454586	0.485363	0.218316	0.301260	0.417526
化工行业	0.741106	0.646489	0.747058	0.666586	0.454586	1.000000	0.719229	0.681599	0.475036	0.661527
化纤行业	0.770169	0.563334	0.752948	0.667844	0.485363	0.719229	1.000000	0.566885	0.447292	0.641033
医疗器械	0.592919	0.680742	0.653398	0.548059	0.218316	0.681599	0.566885	1.000000	0.394219	0.475290
发电设备	0.541358	0.552574	0.544669	0.480807	0.301260	0.475036	0.447292	0.394219	1.000000	0.523425
商业百货	0.674408	0.500004	0.610539	0.555122	0.417526	0.661527	0.641033	0.475290	0.523425	1.000000
家电行业	0.575689	0.592709	0.630609	0.480834	0.408881	0.656623	0.555248	0.531554	0.448245	0.591551
建筑建材	0.736731	0.507151	0.771533	0.741508	0.353069	0.671323	0.731161	0.598810	0.555467	0.632443
开发区	0.778297	0.643814	0.756289	0.692080	0.448214	0.664000	0.702613	0.602580	0.485114	0.562861
房地产	0.758802	0.634076	0.737400	0.640371	0.406933	0.768566	0.748810	0.617135	0.495261	0.628536
摩托车	0.668622	0.524147	0.670062	0.721707	0.307831	0.697438	0.627888	0.582858	0.427356	0.604247
有色金属	0.567511	0.447685	0.514953	0.482924	0.474430	0.526577	0.490551	0.418736	0.403577	0.577964
服装鞋类	0.650817	0.520461	0.577235	0.582435	0.320295	0.693408	0.593909	0.532476	0.444354	0.615160
机械行业	0.739118	0.617579	0.715229	0.635446	0.476698	0.689416	0.706142	0.539212	0.568435	0.647112
汽车制造	0.828500	0.668667	0.768550	0.729143	0.426042	0.799357	0.738559	0.606408	0.592901	0.712519

相关性最高的两个行业：  
新浪行业  
化工行业 电子信息

## 作业中可能可以使用到的操作：

1. 数据的导入、导出
2. append
3. groupby
4. resample
5. sort\_values()

## 作业答案

完成作业后，请微信私信我索要作业答案，算作完成作业。