

《Python金融应用实验》作业（五）

四个行业指数的数据获取、清洗、统计分析与可视化

选定的4个行业指数：

- 399986.SZ — 中证银行指数（银行）
- 399976.SZ — 中证新能源指数（新能源）
- 399967.SZ — 中证军工指数（军工）
- 000933.SH — 中证医药卫生指数（医药）

数据区间： 2023年1月1日 ~ 2025年12月31日

任务1：数据获取

tushare 初始化完成

正在获取 中证银行(399986.SZ) 的数据...

正在获取 中证新能源(399976.SZ) 的数据...

正在获取 中证军工(399967.SZ) 的数据...

正在获取 中证医药(000933.SH) 的数据...

共获取 2181 条记录

	行业名称	ts_code	trade_date	close	open	high	low	pre_close	change	p
0	中证新能源	399976.SZ	20251231	4465.9923	4507.3229	4521.0382	4446.9742	4491.0767	-25.0844	-
1	中证新能源	399976.SZ	20251230	4491.0767	4358.0418	4516.4444	4343.9768	4416.4201	74.6566	
2	中证新能源	399976.SZ	20251229	4416.4201	4507.1664	4507.1664	4406.4147	4499.7021	-83.2820	-
3	中证新能源	399976.SZ	20251226	4499.7021	4455.6210	4543.0341	4438.1196	4400.5194	99.1827	
4	中证新能源	399976.SZ	20251225	4400.5194	4350.8618	4412.7509	4329.8751	4384.0416	16.4778	

数据已保存至 行业指数日线数据.csv

任务2：数据读取与初步探索

数据已导入，日期已设置为索引

日期	行业名称	ts_code	close	open	high	low	pre_close	change	pct_chg
2025-12-31	中证新能源	399976.SZ	4465.9923	4507.3229	4521.0382	4446.9742	4491.0767	-25.0844	-0.5585
2025-12-30	中证新能源	399976.SZ	4491.0767	4358.0418	4516.4444	4343.9768	4416.4201	74.6566	1.6904
2025-12-29	中证新能源	399976.SZ	4416.4201	4507.1664	4507.1664	4406.4147	4499.7021	-83.2820	-1.8508
2025-12-26	中证新能源	399976.SZ	4499.7021	4455.6210	4543.0341	4438.1196	4400.5194	99.1827	2.2539
2025-12-25	中证新能源	399976.SZ	4400.5194	4350.8618	4412.7509	4329.8751	4384.0416	16.4778	0.3759

===== 前5行 =====

日期	行业名称	ts_code	close	open	high	low	pre_close	change	pct_chg
2025-12-31	中证新能源	399976.SZ	4465.9923	4507.3229	4521.0382	4446.9742	4491.0767	-25.0844	-0.5585
2025-12-30	中证新能源	399976.SZ	4491.0767	4358.0418	4516.4444	4343.9768	4416.4201	74.6566	1.6904
2025-12-29	中证新能源	399976.SZ	4416.4201	4507.1664	4507.1664	4406.4147	4499.7021	-83.2820	-1.8508
2025-12-26	中证新能源	399976.SZ	4499.7021	4455.6210	4543.0341	4438.1196	4400.5194	99.1827	2.2539
2025-12-25	中证新能源	399976.SZ	4400.5194	4350.8618	4412.7509	4329.8751	4384.0416	16.4778	0.3759

===== 后5行 =====

日期	行业名称	ts_code	close	open	high	low	pre_close	change	pct
2023-01-09	中证医药	000933.SH	10631.7960	10551.1105	10669.1731	10526.0215	10550.4279	81.3681	0.
2023-01-06	中证医药	000933.SH	10550.4279	10630.4851	10630.4942	10517.6806	10643.9992	-93.5713	-0.
2023-01-05	中证医药	000933.SH	10643.9992	10429.7631	10694.2725	10399.3016	10398.0806	245.9186	2.
2023-01-04	中证医药	000933.SH	10398.0806	10335.0387	10443.1471	10306.8527	10330.4228	67.6578	0.
2023-01-03	中证医药	000933.SH	10330.4228	10296.6228	10388.1177	10235.8970	10280.0950	50.3278	0.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 2181 entries, 2025-12-31 to 2023-01-03
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  -
0   行业名称      2181 non-null   object
1   ts_code      2181 non-null   object
2   close        2181 non-null   float64
3   open         2181 non-null   float64
4   high         2181 non-null   float64
5   low          2181 non-null   float64
6   pre_close    2181 non-null   float64
7   change       2181 non-null   float64
8   pct_chg      2181 non-null   float64
9   vol          2181 non-null   float64
10  amount       2181 non-null   float64
dtypes: float64(9), object(2)
memory usage: 204.5+ KB
```

各行业指数交易日天数：
行业名称
中证军工 727
中证医药 727
中证新能源 727
dtype: int64

各行业指数重复日期数量：

中证新能源：0 条重复

中证军工：0 条重复

中证医药：0 条重复

删除重复前：2181 行，删除重复后：2181 行，共删除：0 行

任务3：数据清洗与预处理

各行业收盘价(close)缺失值统计：

中证新能源：0 个缺失值

中证军工：0 个缺失值

中证医药：0 个缺失值

前向填充(fill)完成

填充后缺失值统计：

中证医药：0 个缺失值

中证新能源：0 个缺失值

中证军工：0 个缺失值

日收益率计算完成，示例：

	ts_code	行业名称	close	日收益率
日期				
2023-01-03	000933.SH	中证医药	10330.4228	NaN
2023-01-03	399976.SZ	中证新能源	4155.5851	NaN
2023-01-03	399967.SZ	中证军工	11352.5432	NaN
2023-01-04	399967.SZ	中证军工	11298.5051	-0.004760
2023-01-04	399976.SZ	中证新能源	4070.4384	-0.020490
2023-01-04	000933.SH	中证医药	10398.0806	0.006549
2023-01-05	399976.SZ	中证新能源	4206.8658	0.033517
2023-01-05	000933.SH	中证医药	10643.9992	0.023650
2023-01-05	399967.SZ	中证军工	11331.1334	0.002888
2023-01-06	399976.SZ	中证新能源	4279.4100	0.017244

删除日收益率缺失值：2181 -> 2178，共删除 3 行

日收益率绝对值大于20%的记录共 0 条：

未发现绝对值大于20%的异常收益率。

任务4：描述性统计分析

===== 各行业描述性统计 =====

	平均日收益率	日收益率标准差	年化收益率	年化波动率	夏普比率
行业名称					
中证军工	0.000434	0.017010	0.109290	0.270022	0.404744
中证医药	-0.000238	0.014082	-0.060051	0.223547	-0.268630
中证新能源	0.000288	0.019502	0.072549	0.309581	0.234346

===== 四个行业日收益率相关系数矩阵 =====

行业名称	中证军工	中证医药	中证新能源
行业名称			
中证军工	1.0000	0.5996	0.6579
中证医药	0.5996	1.0000	0.7225
中证新能源	0.6579	0.7225	1.0000

年化收益率最高的行业：中证军工，年化收益率 = 0.1093

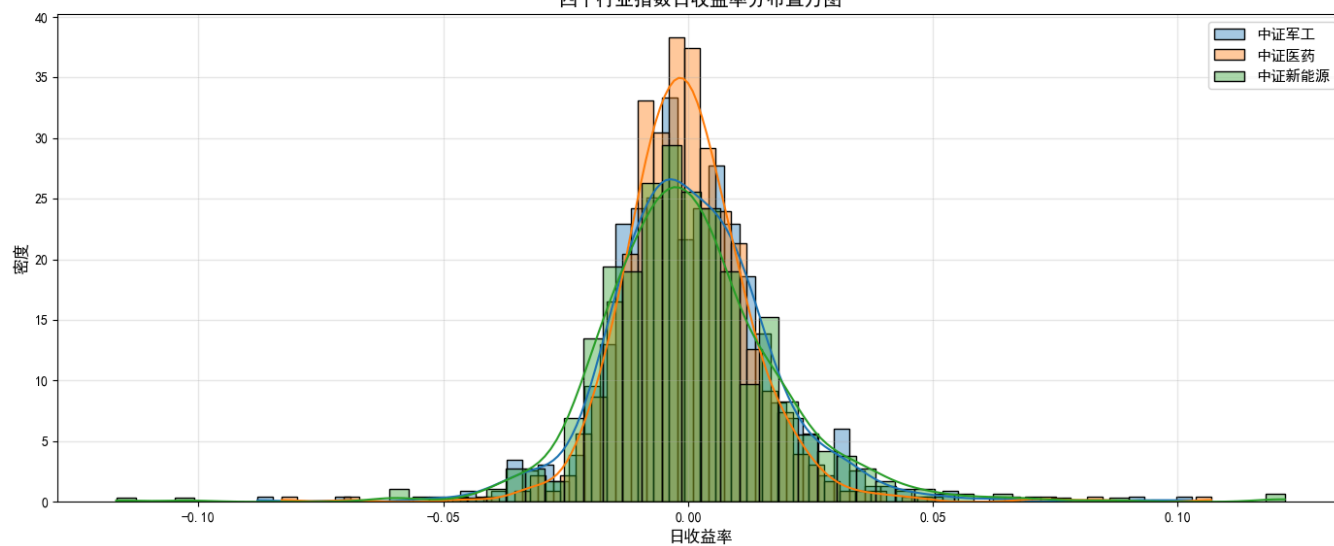
波动最大的行业：中证新能源，年化波动率 = 0.3096

夏普比率最优的行业：中证军工，夏普比率 = 0.4047

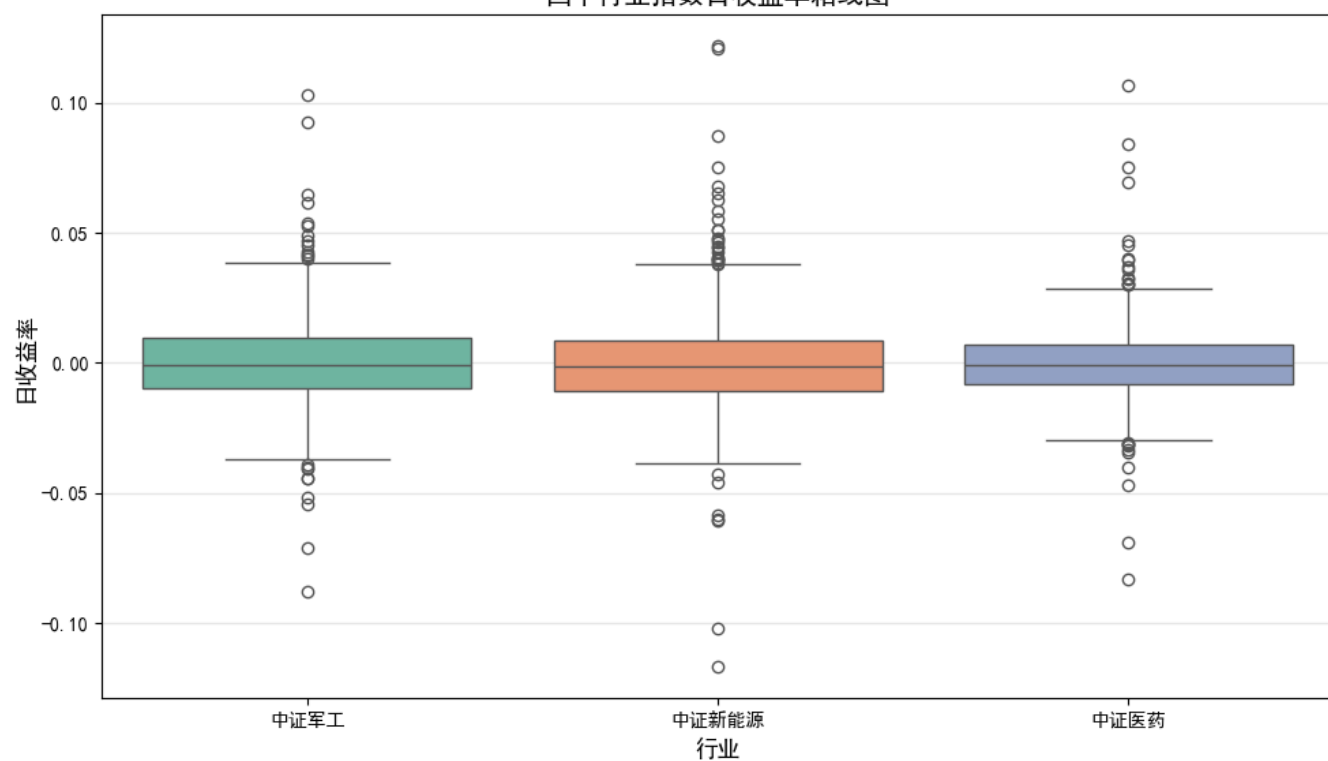
任务4（续）：数据可视化分析\n所有图形包含必要的标题、轴标签

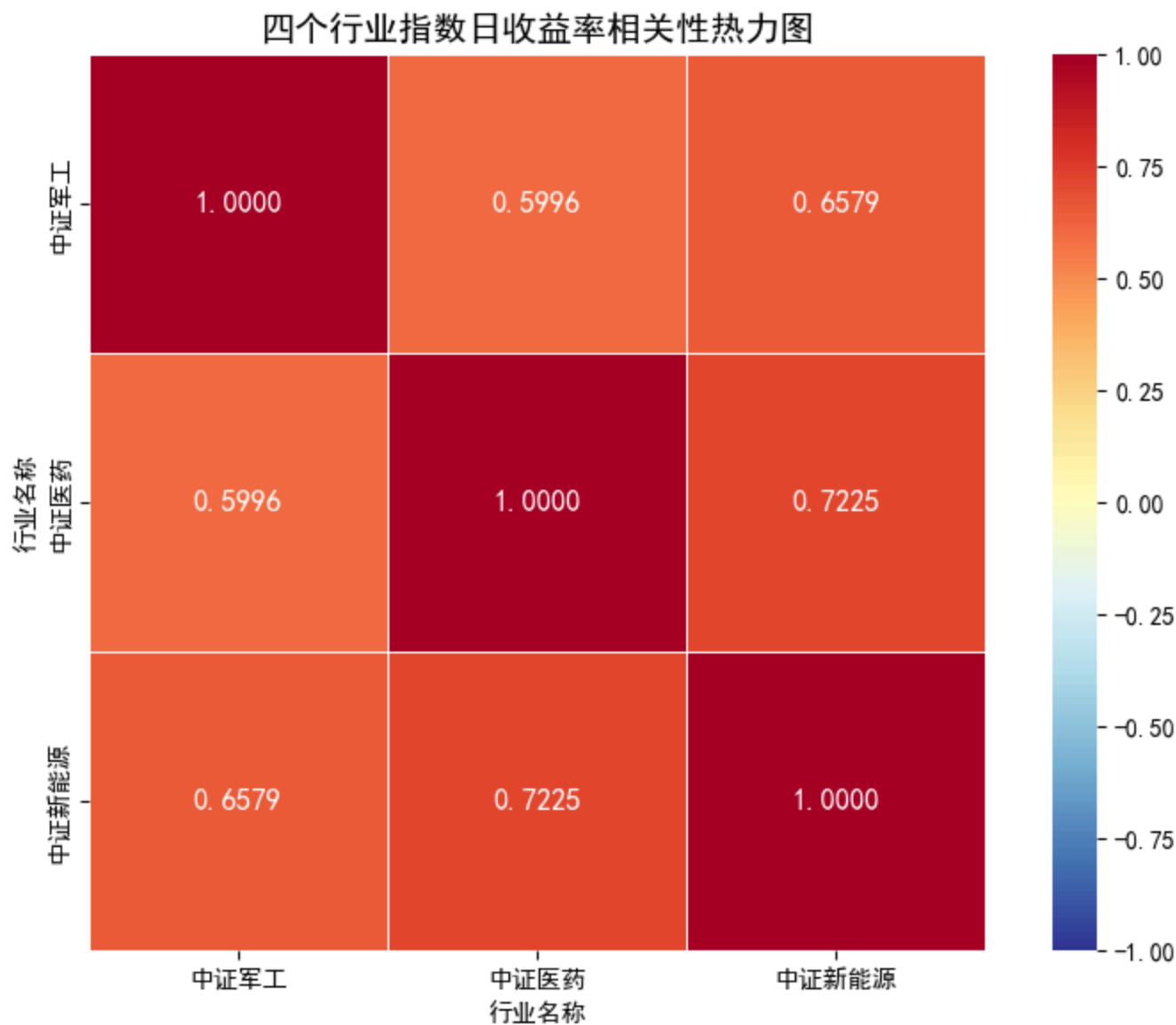


四个行业指数日收益率分布直方图



四个行业指数日收益率箱线图

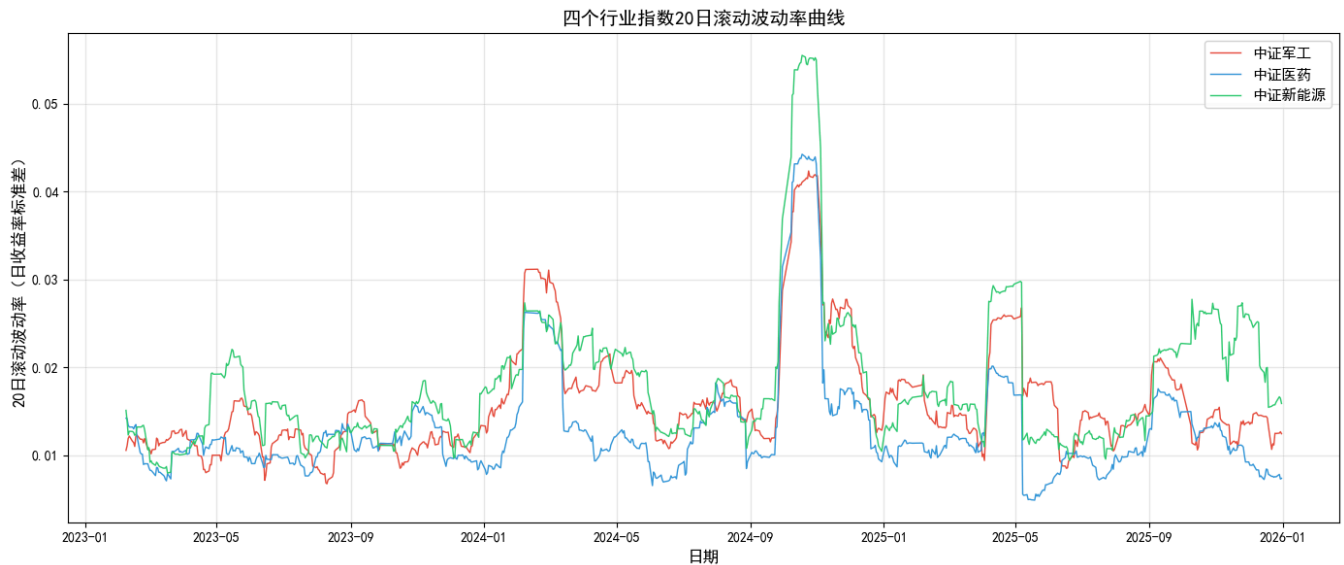




任务5：滚动波动率与风险分析

20日滚动波动率计算完成，示例：

	中证军工_20日滚动波动率	中证医药_20日滚动波动率	中证新能源_20日滚动波动率
日期			
2023-02-07	0.010552	0.014207	0.015103
2023-02-08	0.011064	0.014172	0.013930
2023-02-09	0.011901	0.013406	0.012304
2023-02-10	0.012185	0.013208	0.012723
2023-02-13	0.011568	0.013204	0.012694



===== 各行业每年滚动波动率最高的日期及波动率值 =====

【中证军工】

2023年：最高波动率日期 = 2023-05-24，波动率 = 0.016496

2024年：最高波动率日期 = 2024-10-24，波动率 = 0.042328

2025年：最高波动率日期 = 2025-05-07，波动率 = 0.026779

【中证医药】

2023年：最高波动率日期 = 2023-10-30，波动率 = 0.015781

2024年：最高波动率日期 = 2024-10-18，波动率 = 0.044228

2025年：最高波动率日期 = 2025-04-10，波动率 = 0.020137

【中证新能源】

2023年：最高波动率日期 = 2023-05-15，波动率 = 0.022046

2024年：最高波动率日期 = 2024-10-18，波动率 = 0.055476

2025年：最高波动率日期 = 2025-05-06，波动率 = 0.029776

===== 风险分析总结 =====

各行业平均20日滚动波动率：

中证军工：0.015889

中证医药：0.012686

中证新能源：0.017813

波动率整体最高的行业：中证新能源（平均20日滚动波动率 = 0.017813）

----- 多个行业波动率同时升高的时段 -----

共有 37 个交易日出现3个及以上行业同时高波动，分布时段：

2024-02-07 ~ 2024-02-08 （共2个交易日）

2024-02-19 ~ 2024-02-21 （共3个交易日）

2024-02-29 ~ 2024-03-01 （共2个交易日）

2024-09-30 ~ 2024-09-30 （共1个交易日）

2024-10-08 ~ 2024-11-07 （共22个交易日）

2025-04-09 ~ 2025-04-17 （共7个交易日）