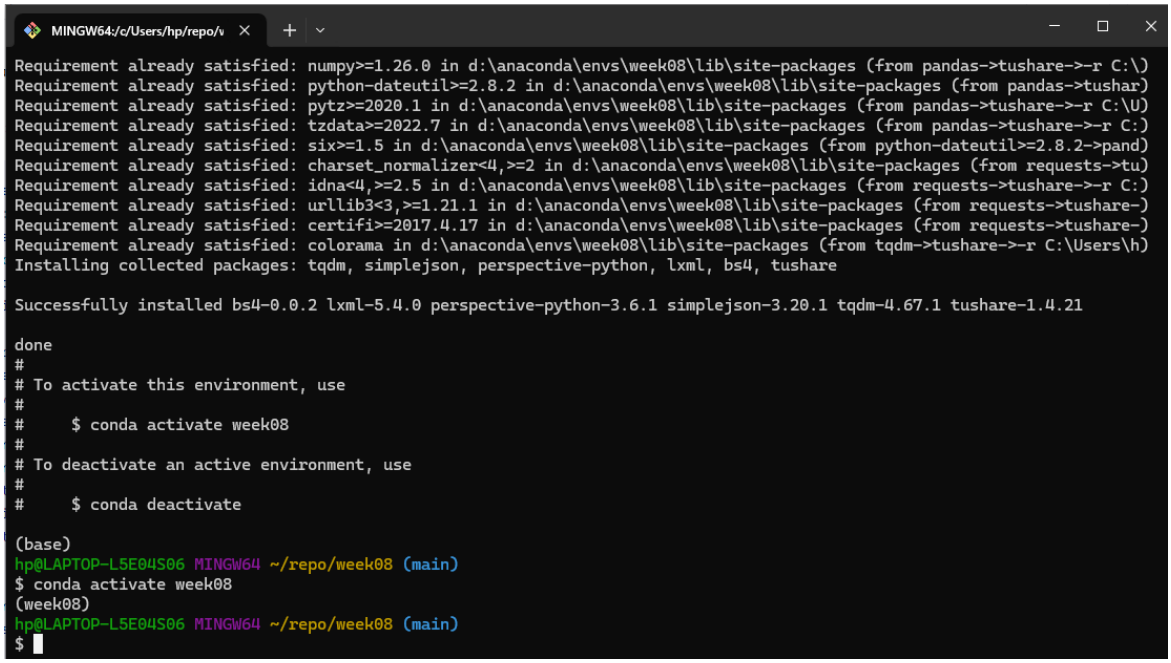
第八周学习报告

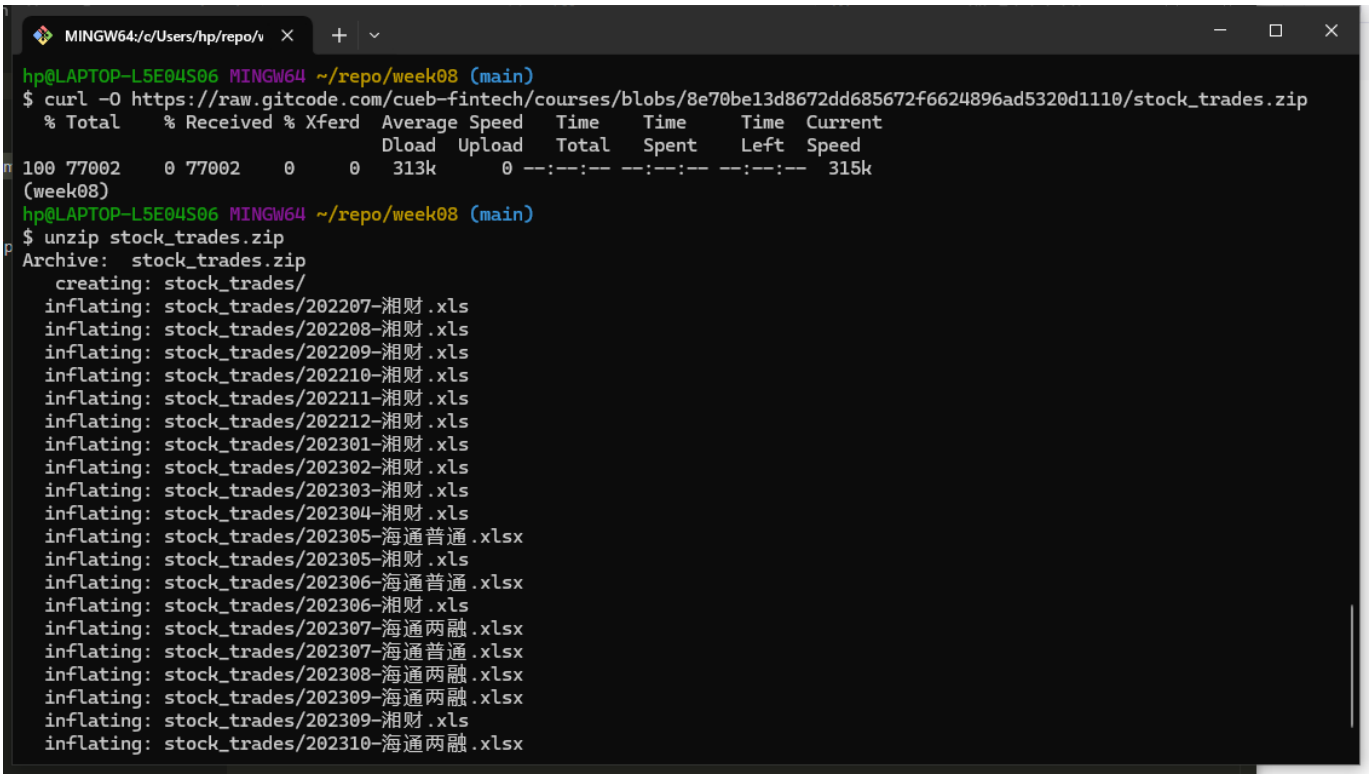
1.创建 Conda 环境

environment.yml

```yaml
name: week08
channels:
  - conda-forge
dependencies:
  - python=3.12
  - wat-inspector
  - xlrd
  - openpyxl
  - fastexcel
  - xlsxwriter
  - pandas
  - pyarrow
  - polars
  - jupyterlab
  - ipywidgets
  - jupyter-ruff
  - pip
  - pip:
    - perspective-python
    - tushare
```

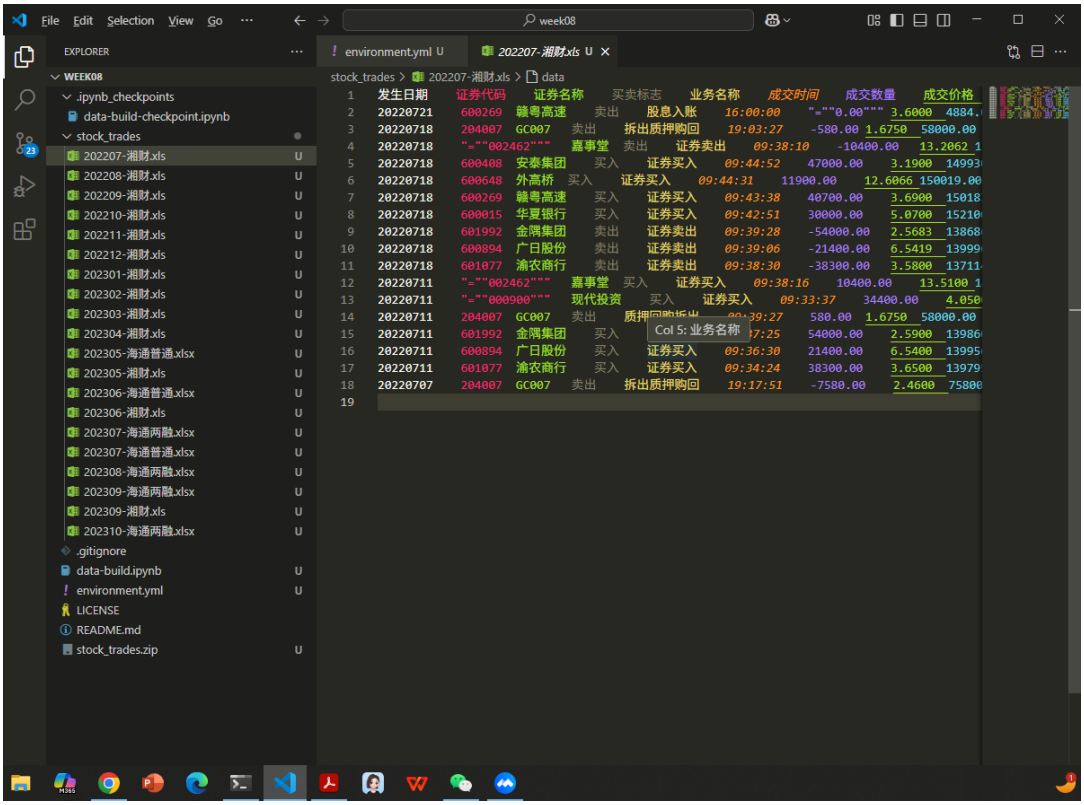


2.下载案例数据到你的本地，并解压出文件夹



3.在 JupyterLab 页面里完成 数据清洗 操作

- ❖ 选择 GB18030 编解码器



❖ 使用 polars.read_csv() 函数重新读取 202207-湘财.xls 文件，命名为 df

```
import polars  as pl
```

```
pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t")
```

shape: (17, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
i64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
20220721	"600269"	"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"="0.00""	3.6	4884.0	4884.0	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"股息入账:赣粤高速600269; 权益股数:40700;"	"人民币"
20220718	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	1.675	58000.0	58018.63	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:18.63实际占款天数：7-888880"	"人民币"
20220718	"="002462""	"嘉事堂"	"卖出"	"证券卖出"	"09:38:10"	"-10400.00"	13.2062	137344.0	137184.67	"21.98"	"137.35"	"1.38"	"="0.00""	"证券卖出"	"人民币"
20220718	"600408"	"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"	3.19	149930.0	-149955.5	"23.99"	"="0.00""	"1.51"	"="0.00""	"证券买入"	"人民币"
20220718	"600648"	"外高桥"	"买入"	"证券买入"	"09:44:31"	"11900.00"	12.6066	150019.0	-150044.49	"24.00"	"="0.00""	"1.49"	"="0.00""	"证券买入"	"人民币"
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
20220711	"204007"	"GC007"	"卖出"	"质押回购拆出"	"09:39:27"	"580.00"	1.675	58000.0	-58002.9	"2.90"	"="0.00""	"="0.00""	"="0.00""	"融券回购购回日:20220718预计利息:18.63参考占款...	"人民币"
20220711	"601992"	"金隅集团"	"买入"	"证券买入"	"09:37:25"	"54000.00"	2.59	139860.0	-139883.8	"22.38"	"="0.00""	"1.42"	"="0.00""	"证券买入"	"人民币"
20220711	"600894"	"广日股份"	"买入"	"证券买入"	"09:36:30"	"21400.00"	6.54	139956.0	-139979.8	"22.39"	"="0.00""	"1.41"	"="0.00""	"证券买入"	"人民币"
20220711	"601077"	"渝农商行"	"买入"	"证券买入"	"09:34:24"	"38300.00"	3.65	139795.0	-139818.75	"22.37"	"="0.00""	"1.38"	"="0.00""	"证券买入"	"人民币"
20220707	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:17:51"	"-7580.00"	2.46	758000.0	758357.61	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:357.61实际占款天数：7-888880"	"人民币"

```
df=pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t")
```

❖ 掌握检查 polars.DataFrame 对象时常用的属性 (attributes) / 方法 (methods)

➢ 形状/维度：df.shape、df.height（行）、df.width（列）、df.is_empty()

```
[9]: df.shape
[9]: (17, 16)
[10]: df.height
[10]: 17
[11]: df.width
[11]: 16
[12]: df.is_empty()
[12]: False
```

➢ 数据模式/架构/类型：df.schema、df.columns、df.dtypes

```
[13]: df.schema
[13]: Schema([('发生日期', Int64),
              ('证券代码', String),
              ('证券名称', String),
              ('买卖标志', String),
              ('业务名称', String),
              ('成交时间', String),
              ('成交数量', String),
              ('成交价格', Float64),
              ('成交金额', Float64),
              ('发生金额', Float64),
              ('手续费', String),
              ('印花税', String),
              ('过户费', String),
              ('其他费', String),
              ('备注', String),
              ('币种', String)])
```

返回的是有序字典

```
[14]: df.columns
```

```
[14]: ['发生日期',
       '证券代码',
       '证券名称',
       '买卖标志',
       '业务名称',
       '成交时间',
       '成交数量',
       '成交价格',
       '成交金额',
       '发生金额',
       '手续费',
       '印花税',
       '过户费',
       '其他费',
       '备注',
       '币种']
```

返回的是列表

```
[15]: df.dtypes
```

```
[15]: [Int64,
       String,
       String,
       String,
       String,
       String,
       String,
       Float64,
       Float64,
       Float64,
       String,
       String,
       String,
       String,
       String,
       String]
```

加行不改变架构，加列改变架构

- 数据提取/切片
  - df[...] (取行、取多行)

提取第一行

```
[16]: df[0]
```

```
[16]: shape: (1, 16)
```

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
i64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
20220721	"600269"	"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"="0.00""	3.6	4884.0	4884.0	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"股息入账:赣粤高速600269; 权益股数:40700;"	"人民币"

提取前四行

```
[18]: df[:4]
```

```
[18]: shape: (4, 16)
```

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
i64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
20220721	"600269"	"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"="0.00""	3.6	4884.0	4884.0	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"股息入账:赣粤高速600269; 权益股数:40700;"	"人民币"
20220718	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	1.675	58000.0	58018.63	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:18.63实际占款天数：7-888880"	"人民币"
20220718	"="002462""	"嘉事堂"	"卖出"	"证券卖出"	"09:38:10"	"-10400.00"	13.2062	137344.0	137184.67	"21.98"	"137.35"	"1.38"	"="0.00""	"证券卖出"	"人民币"
20220718	"600408"	"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"	3.19	149930.0	-149955.5	"23.99"	"="0.00""	"1.51"	"="0.00""	"证券买入"	"人民币"

提取5-6行

```
[19]: df[4:6]
```

```
[19]: shape: (2, 16)
```

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
i64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
20220718	"600648"	"外高桥"	"买入"	"证券买入"	"09:44:31"	"11900.00"	12.6066	150019.0	-150044.49	"24.00"	"="0.00""	"1.49"	"="0.00""	"证券买入"	"人民币"
20220718	"600269"	"赣粤高速"	"买入"	"证券买入"	"09:43:38"	"40700.00"	3.69	150183.0	-150208.53	"24.03"	"="0.00""	"1.50"	"="0.00""	"证券买入"	"人民币"

- df[..., ...] (行列双向限制)

提取第4列

```
]: df[:,3]
```

]: shape: (17,)

买卖标志
str
"卖出"
"卖出"
"卖出"
"买入"
"买入"
...
"卖出"
"买入"
"买入"
"买入"
"卖出"

提取3-5列

```
[21]: df[:,2:5]
```

[21]: shape: (17, 3)

证券名称	买卖标志	业务名称
str	str	str
"赣粤高速"	"卖出"	"股息入账"
"GC007"	"卖出"	"拆出质押购回"
"嘉事堂"	"卖出"	"证券卖出"
"安泰集团"	"买入"	"证券买入"
"外高桥"	"买入"	"证券买入"
...	...	...
"GC007"	"卖出"	"质押回购拆出"
"金隅集团"	"买入"	"证券买入"
"广日股份"	"买入"	"证券买入"
"渝农商行"	"买入"	"证券买入"
"GC007"	"卖出"	"拆出质押购回"

提取特定列

```
[22]: df[:,["业务名称","发生金额"]]
```

[22]: shape: (17, 2)

业务名称	发生金额
str	f64
"股息入账"	4884.0
"拆出质押购回"	58018.63
"证券卖出"	137184.67
"证券买入"	-149955.5
"证券买入"	-150044.49
...	...
"质押回购拆出"	-58002.9
"证券买入"	-139883.8
"证券买入"	-139979.8
"证券买入"	-139818.75
"拆出质押购回"	758357.61

提取特定行、列

```
[23]: df[[3,5,7],["业务名称","发生金额"]]
```

[23]: shape: (3, 2)

业务名称	发生金额
str	f64
"证券买入"	-149955.5
"证券买入"	-150208.53
"证券卖出"	138523.74

- df.row()返回的是元组
- df.rows()返回的是列表

➢ 数据概览/描述

- df.glimpse()

```
24]: df.glimpse()
```

```
Rows: 17
Columns: 16
$ 发生日期 <i64> 20220721, 20220718, 20220718, 20220718, 20220718, 20220718, 20220718, 20220718, 20220718, 20220718
$ 证券代码 <str> '600269', '204007', '="002462"', '600408', '600648', '600269', '600015', '601992', '600894', '601077'
$ 证券名称 <str> '赣粤高速', 'GC007', '嘉事堂', '安泰集团', '外高桥', '赣粤高速', '华夏银行', '金隅集团', '广日股份', '渝农商行'
$ 买卖标志 <str> '卖出', '卖出', '卖出', '买入', '买入', '买入', '买入', '卖出', '卖出', '卖出'
$ 业务名称 <str> '股息入账', '拆出质押购回', '证券卖出', '证券买入', '证券买入', '证券买入', '证券买入', '证券卖出', '证券卖出', '证券卖出'
$ 成交时间 <str> '16:00:00', '19:03:27', '09:38:10', '09:44:52', '09:44:31', '09:43:38', '09:42:51', '09:39:28', '09:39:06', '09:38:30'
$ 成交数量 <str> '="0.00"', '-580.00', '-10400.00', '47000.00', '11900.00', '40700.00', '30000.00', '-54000.00', '-21400.00', '-38300.00'
$ 成交价格 <f64> 3.6, 1.675, 13.2062, 3.19, 12.6066, 3.69, 5.07, 2.5683, 6.5419, 3.58
$ 成交金额 <f64> 4884.0, 58000.0, 137344.0, 149930.0, 150019.0, 150183.0, 152100.0, 138686.0, 139996.0, 137114.0
$ 发生金额 <f64> 4884.0, 58018.63, 137184.67, -149955.5, -150044.49, -150208.53, -152125.86, 138523.74, 139832.12, 136953.55
$ 手续费  <str> '="0.00"', '="0.00"', '21.98', '23.99', '24.00', '24.03', '24.34', '22.19', '22.40', '21.94'
$ 印花税  <str> '="0.00"', '="0.00"', '137.35', '="0.00"', '="0.00"', '="0.00"', '="0.00"', '138.69', '140.02', '137.13'
$ 过户费  <str> '="0.00"', '="0.00"', '1.38', '1.51', '1.49', '1.50', '1.52', '1.38', '1.46', '1.38'
$ 其他费  <str> '="0.00"', '="0.00"', '="0.00"', '="0.00"', '="0.00"', '="0.00"', '="0.00"', '="0.00"', '="0.00"', '="0.00"'
$ 备注   <str> '股息入账:赣粤高速600269; 权益股数:40700;', '融券购回:18.63实际占款天数: 7-888880', '证券卖出', '证券买入', '证券买入', '证券买入', '证券买入', '证券卖出', '证券卖出', '证券卖出'
$ 币种   <str> '人民币', '人民币', '人民币', '人民币', '人民币', '人民币', '人民币', '人民币', '人民币', '人民币'
```

- df.head()　提取前五行

```
: df.head()
```

shape: (5, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
i64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
20220721	"600269"	"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"="0.00""	3.6	4884.0	4884.0	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"股息入账:赣粤高速600269; 权益股数:40700;"	"人民币"
20220718	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	1.675	58000.0	58018.63	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:18.63实际占款天数: 7-888880"	"人民币"
20220718	"="002462""	"嘉事堂"	"卖出"	"证券卖出"	"09:38:10"	"-10400.00"	13.2062	137344.0	137184.67	"21.98"	"137.35"	"1.38"	"="0.00""	"证券卖出"	"人民币"
20220718	"600408"	"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"	3.19	149930.0	-149955.5	"23.99"	"="0.00""	"1.51"	"="0.00""	"证券买入"	"人民币"
20220718	"600648"	"外高桥"	"买入"	"证券买入"	"09:44:31"	"11900.00"	12.6066	150019.0	-150044.49	"24.00"	"="0.00""	"1.49"	"="0.00""	"证券买入"	"人民币"

- df.tail()　提取后五行

```
[26]: df.tail()
```

[26]: shape: (5, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
i64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
20220711	"204007"	"GC007"	"卖出"	"质押回购拆出"	"09:39:27"	"580.00"	1.675	58000.0	-58002.9	"2.90"	"="0.00""	"="0.00""	"="0.00""	"融券回购购回日:20220718预计利息:18.63参考占款...	"人民币"
20220711	"601992"	"金隅集团"	"买入"	"证券买入"	"09:37:25"	"54000.00"	2.59	139860.0	-139883.8	"22.38"	"="0.00""	"1.42"	"="0.00""	"证券买入"	"人民币"
20220711	"600894"	"广日股份"	"买入"	"证券买入"	"09:36:30"	"21400.00"	6.54	139956.0	-139979.8	"22.39"	"="0.00""	"1.41"	"="0.00""	"证券买入"	"人民币"
20220711	"601077"	"渝农商行"	"买入"	"证券买入"	"09:34:24"	"38300.00"	3.65	139795.0	-139818.75	"22.37"	"="0.00""	"1.38"	"="0.00""	"证券买入"	"人民币"
20220707	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:17:51"	"-7580.00"	2.46	758000.0	758357.61	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:357.61实际占款天数: 7-888880"	"人民币"

- df.sample() 随机提取一行　df.sample(5)随机提取五行

```
[27]: df.sample()
```

[27]: shape: (1, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
i64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
20220718	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	1.675	58000.0	58018.63	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:18.63实际占款天数: 7-888880"	"人民币"

- df.describe()描述性统计

```
[28]: df.describe()
```

[28]: shape: (9, 17)

statistic	发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
str	f64	str	str	str	str	str	str	f64	f64	f64	str	str	str	str	str	str
"count"	17.0	"17"	"17"	"17"	"17"	"17"	"17"	17.0	17.0	17.0	"17"	"17"	"17"	"17"	"17"	"17"
"null_count"	0.0	"0"	"0"	"0"	"0"	"0"	"0"	0.0	0.0	0.0	"0"	"0"	"0"	"0"	"0"	"0"
"mean"	2.0221e7	null	null	null	null	null	null	5.306059	160805.352941	815.642353	null	null	null	null	null	null
"std"	4.160387	null	null	null	null	null	null	3.974618	159439.770376	230047.143813	null	null	null	null	null	null
"min"	2.0220707e7	"204007"	"GC007"	"买入"	"拆出质押购回"	"09:33:37"	"-10400.00"	1.675	4884.0	-152125.86	"2.90"	"137.13"	"1.38"	"="0.00""	"股息入账:赣粤高速600269; 权益股数:40700;"	"人民币"
"25%"	2.0220711e7	null	null	null	null	null	null	2.59	137344.0	-140526.48	null	null	null	null	null	null
"50%"	2.0220718e7	null	null	null	null	null	null	3.65	139860.0	-139342.29	null	null	null	null	null	null
"75%"	2.0220718e7	null	null	null	null	null	null	6.54	149930.0	136953.55	null	null	null	null	null	null
"max"	2.0220721e7	"="002462""	"金隅集团"	"卖出"	"质押回购拆出"	"19:17:51"	"="0.00""	13.51	758000.0	758357.61	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"证券卖出"	"人民币"

- df.null_count() 显示每列的缺失值

```
[29]: df.null_count()
```

[29]: shape: (1, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
u32	u32	u32	u32	u32	u32	u32	u32	u32	u32	u32	u32	u32	u32	u32	u32
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

➢ 转换/导出

- df.to_pandas()

```
[30]: df.to_pandas()
```

[30]:

	发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
0	20220721	600269	赣粤高速	卖出	股息入账	16:00:00	="0.00"	3.6000	4884.0	4884.00	="0.00"	="0.00"	="0.00"	="0.00"	股息入账:赣粤高速600269; 权益股数:40700;	人民币
1	20220718	204007	GC007	卖出	拆出质押购回	19:03:27	-580.00	1.6750	58000.0	58018.63	="0.00"	="0.00"	="0.00"	="0.00"	融券购回:18.63实际占款天数: 7-888880	人民币
2	20220718	="002462"	嘉事堂	卖出	证券卖出	09:38:10	-10400.00	13.2062	137344.0	137184.67	21.98	137.35	1.38	="0.00"	证券卖出	人民币
3	20220718	600408	安泰集团	买入	证券买入	09:44:52	47000.00	3.1900	149930.0	-149955.50	23.99	="0.00"	1.51	="0.00"	证券买入	人民币
4	20220718	600648	外高桥	买入	证券买入	09:44:31	11900.00	12.6066	150019.0	-150044.49	24.00	="0.00"	1.49	="0.00"	证券买入	人民币
5	20220718	600269	赣粤高速	买入	证券买入	09:43:38	40700.00	3.6900	150183.0	-150208.53	24.03	="0.00"	1.50	="0.00"	证券买入	人民币
6	20220718	600015	华夏银行	买入	证券买入	09:42:51	30000.00	5.0700	152100.0	-152125.86	24.34	="0.00"	1.52	="0.00"	证券买入	人民币
7	20220718	601992	金隅集团	卖出	证券卖出	09:39:28	-54000.00	2.5683	138686.0	138523.74	22.19	138.69	1.38	="0.00"	证券卖出	人民币
8	20220718	600894	广日股份	卖出	证券卖出	09:39:06	-21400.00	6.5419	139996.0	139832.12	22.40	140.02	1.46	="0.00"	证券卖出	人民币
9	20220718	601077	渝农商行	卖出	证券卖出	09:38:30	-38300.00	3.5800	137114.0	136953.55	21.94	137.13	1.38	="0.00"	证券卖出	人民币
10	20220711	="002462"	嘉事堂	买入	证券买入	09:38:16	10400.00	13.5100	140504.0	-140526.48	22.48	="0.00"	1.41	="0.00"	证券买入	人民币
11	20220711	="000900"	现代投资	买入	证券买入	09:33:37	34400.00	4.0500	139320.0	-139342.29	22.29	="0.00"	1.39	="0.00"	证券买入	人民币
12	20220711	204007	GC007	卖出	质押回购拆出	09:39:27	580.00	1.6750	58000.0	-58002.90	2.90	="0.00"	="0.00"	="0.00"	融券回购购回日:20220718预计利息:18.63参考占款天数: 7-888880	人民币
13	20220711	601992	金隅集团	买入	证券买入	09:37:25	54000.00	2.5900	139860.0	-139883.80	22.38	="0.00"	1.42	="0.00"	证券买入	人民币
14	20220711	600894	广日股份	买入	证券买入	09:36:30	21400.00	6.5400	139956.0	-139979.80	22.39	="0.00"	1.41	="0.00"	证券买入	人民币
15	20220711	601077	渝农商行	买入	证券买入	09:34:24	38300.00	3.6500	139795.0	-139818.75	22.37	="0.00"	1.38	="0.00"	证券买入	人民币
16	20220707	204007	GC007	卖出	拆出质押购回	19:17:51	-7580.00	2.4600	758000.0	758357.61	="0.00"	="0.00"	="0.00"	="0.00"	融券购回:357.61实际占款天数: 7-888880	人民币

- df.to_arrow()

```
[31]: df.to_arrow()
```

```
[31]: pyarrow.Table
发生日期: int64
证券代码: large_string
证券名称: large_string
买卖标志: large_string
业务名称: large_string
成交时间: large_string
成交数量: large_string
成交价格: double
成交金额: double
发生金额: double
手续费: large_string
印花税: large_string
过户费: large_string
其他费: large_string
备注: large_string
币种: large_string
----
发生日期: [[20220721,20220718,20220718,20220718,20220718,...,20220711,20220711,20220711,20220711,20220707]]
证券代码: [["600269","204007","="002462"","600408","600648",...,"204007","601992","600894","601077","204007"]]
证券名称: [["赣粤高速","GC007","嘉事堂","安泰集团","外高桥",...,"GC007","金隅集团","广日股份","渝农商行","GC007"]]
买卖标志: [["卖出","卖出","卖出","买入","买入",...,"卖出","买入","买入","买入","卖出"]]
业务名称: [["股息入账","拆出质押购回","证券卖出","证券买入","证券买入",...,"质押回购拆出","证券买入","证券买入","证券买入","拆出质押购回"]]
成交时间: [["16:00:00","19:03:27","09:38:10","09:44:52","09:44:31",...,"09:39:27","09:37:25","09:36:30","09:34:24","19:17:51"]]
成交数量: [["="0.00"","-580.00","-10400.00","47000.00","11900.00",...,"580.00","54000.00","21400.00","38300.00","-7580.00"]]
成交价格: [[3.6,1.675,13.2062,3.19,12.6066,...,1.675,2.59,6.54,3.65,2.46]]
成交金额: [[4884,58000,137344,149930,150019,...,58000,139860,139956,139795,758000]]
发生金额: [[4884,58018.63,137184.67,-149955.5,-150044.49,...,-58002.9,-139883.8,-139979.8,-139818.75,758357.61]]
...
```

❖ 检查 polars.Series 对象 (命名为 s) 常用的属性 (attributes) / 方法 (methods)

➢ 基本属性：s.name、s.dtype、s.shape、s.len()

```
[32]: s=df.to_series()
[33]: s.name
[33]: '发生日期'
[34]: s.dtype
[34]: Int64
[35]: s.shape
[35]: (17,)
[36]: s.len()
[36]: 17
```

➢ 数据提取/切片：s[...] (取单值/取多值)

```
[37]: s[2,6,9]
```

[37]: shape: (3,)

发生日期
i64
20220718
20220718
20220718

➢ 数据概览/描述

- s.unique() 查看不同值

```
[38]: df.to_series(2).unique()
```

[38]: shape: (10,)

证券名称
str
"现代投资"
"外高桥"
"GC007"
"华夏银行"
"安泰集团"
"广日股份"
"金隅集团"
"嘉事堂"
"赣粤高速"
"渝农商行"

- s.value_counts() 查看重复值的次数

```
39]: df.to_series(2).value_counts()
```

39]: shape: (10, 2)

证券名称	count
str	u32
"嘉事堂"	2
"华夏银行"	1
"广日股份"	2
"现代投资"	1
"GC007"	3
"安泰集团"	1
"外高桥"	1
"赣粤高速"	2
"金隅集团"	2
"渝农商行"	2

- s.describe() 描述性统计

```
[40]: df[:,"发生金额"].describe()
```

[40]: shape: (9, 2)

statistic	value
str	f64
"count"	17.0
"null_count"	0.0
"mean"	815.642353
"std"	230047.143813
"min"	-152125.86
"25%"	-140526.48
"50%"	-139342.29
"75%"	136953.55
"max"	758357.61

- 转换/导出：s.to_list()

```
[41]: df[:,"发生金额"].to_list()
```

```
[41]: [4884.0,
       58018.63,
       137184.67,
       -149955.5,
       -150044.49,
       -150208.53,
       -152125.86,
       138523.74,
       139832.12,
       136953.55,
       -140526.48,
       -139342.29,
       -58002.9,
       -139883.8,
       -139979.8,
       -139818.75,
       758357.61]
```

- 指定参数 infer_schema=False，将所有列都先读取为字符串类型

```
[43]: pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t",infer_schema=False)
```

[43]: shape: (17, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
str	str	str	str	str	str	str	str	str	str	str	str	str	str	str	str
"20220721"	"600269"	"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"="0.00""	"3.6000"	"4884.00"	"4884.00"	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"股息入账:赣粤高速600269; 权益股数:40700;"	"人民币"
"20220718"	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	"1.6750"	"58000.00"	"58018.63"	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:18.63实际占款天数：7-888880"	"人民币"
"20220718"	"="002462""	"嘉事堂"	"卖出"	"证券卖出"	"09:38:10"	"-10400.00"	"13.2062"	"137344.00"	"137184.67"	"21.98"	"137.35"	"1.38"	"="0.00""	"证券卖出"	"人民币"
"20220718"	"600408"	"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"	"3.1900"	"149930.00"	"-149955.50"	"23.99"	"="0.00""	"1.51"	"="0.00""	"证券买入"	"人民币"
"20220718"	"600648"	"外高桥"	"买入"	"证券买入"	"09:44:31"	"11900.00"	"12.6066"	"150019.00"	"-150044.49"	"24.00"	"="0.00""	"1.49"	"="0.00""	"证券买入"	"人民币"
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
"20220711"	"204007"	"GC007"	"卖出"	"质押回购拆出"	"09:39:27"	"580.00"	"1.6750"	"58000.00"	"-58002.90"	"2.90"	"="0.00""	"="0.00""	"="0.00""	"融券回购购回日:20220718预计利息:18.63参考占款...	"人民币"
"20220711"	"601992"	"金隅集团"	"买入"	"证券买入"	"09:37:25"	"54000.00"	"2.5900"	"139860.00"	"-139883.80"	"22.38"	"="0.00""	"1.42"	"="0.00""	"证券买入"	"人民币"
"20220711"	"600894"	"广日股份"	"买入"	"证券买入"	"09:36:30"	"21400.00"	"6.5400"	"139956.00"	"-139979.80"	"22.39"	"="0.00""	"1.41"	"="0.00""	"证券买入"	"人民币"
"20220711"	"601077"	"渝农商行"	"买入"	"证券买入"	"09:34:24"	"38300.00"	"3.6500"	"139795.00"	"-139818.75"	"22.37"	"="0.00""	"1.38"	"="0.00""	"证券买入"	"人民币"
"20220707"	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:17:51"	"-7580.00"	"2.4600"	"758000.00"	"758357.61"	"="0.00""	"="0.00""	"="0.00""	"="0.00""	"融券购回:357.61实际占款天数：7-888880"	"人民币"

- 数据清洗

  - DataFrame.with_columns() 方法用来添加/修改列
  - DataFrame.select() 方法用来挑选/计算列
  - DataFrame.filter() 方法用来过滤行 (计算为 True 的行将被保留)

- 把 发生日期 列转换为 polars.Date 类型

```
[54]: df=pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t",infer_schema=False)
```

```
[55]: df=df.with_columns(
          pl.col("发生日期").str.to_date("%Y%m%d"),
      )
```

```
[56]: df
```

[56]: shape: (17, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
date	str	str	str	str	str	str	str	str	str	str	str	str	str	str	str
2022-07-21	"600269"	"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"=0.00""	"3.6000"	"4884.00"	"4884.00"	"=0.00""	"=0.00""	"=0.00""	"=0.00""	"股息入账:赣粤高速600269; 权益股数:40700;"	"人民币"
2022-07-18	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	"1.6750"	"58000.00"	"58018.63"	"=0.00""	"=0.00""	"=0.00""	"=0.00""	"融券购回:18.63实际占款天数：7-888880"	"人民币"
2022-07-18	"=002462""	"嘉事堂"	"卖出"	"证券卖出"	"09:38:10"	"-10400.00"	"13.2062"	"137344.00"	"137184.67"	"21.98"	"137.35"	"1.38"	"=0.00""	"证券卖出"	"人民币"
2022-07-18	"600408"	"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"	"3.1900"	"149930.00"	"-149955.50"	"23.99"	"=0.00""	"1.51"	"=0.00""	"证券买入"	"人民币"
2022-07-18	"600648"	"外高桥"	"买入"	"证券买入"	"09:44:31"	"11900.00"	"12.6066"	"150019.00"	"-150044.49"	"24.00"	"=0.00""	"1.49"	"=0.00""	"证券买入"	"人民币"
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2022-07-11	"204007"	"GC007"	"卖出"	"质押回购拆出"	"09:39:27"	"580.00"	"1.6750"	"58000.00"	"-58002.90"	"2.90"	"=0.00""	"=0.00""	"=0.00""	"融券回购购回日:20220718预计利息:18.63参考占款...	"人民币"
2022-07-11	"601992"	"金隅集团"	"买入"	"证券买入"	"09:37:25"	"54000.00"	"2.5900"	"139860.00"	"-139883.80"	"22.38"	"=0.00""	"1.42"	"=0.00""	"证券买入"	"人民币"
2022-07-11	"600894"	"广日股份"	"买入"	"证券买入"	"09:36:30"	"21400.00"	"6.5400"	"139956.00"	"-139979.80"	"22.39"	"=0.00""	"1.41"	"=0.00""	"证券买入"	"人民币"
2022-07-11	"601077"	"渝农商行"	"买入"	"证券买入"	"09:34:24"	"38300.00"	"3.6500"	"139795.00"	"-139818.75"	"22.37"	"=0.00""	"1.38"	"=0.00""	"证券买入"	"人民币"
2022-07-07	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:17:51"	"-7580.00"	"2.4600"	"758000.00"	"758357.61"	"=0.00""	"=0.00""	"=0.00""	"=0.00""	"融券购回:357.61实际占款天数：7-888880"	"人民币"

- 把这些多余的字符去掉

```
[61]: df=pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t",infer_schema=False)
      df=df.with_columns(
          pl.col("发生日期").str.to_date("%Y%m%d"),
          pl.col("证券代码").str.strip_prefix("=").str.strip_chars('"'),
       )
      df[:,"证券代码"].unique().to_list()
```

```
[61]: ['600408',
       '600894',
       '600015',
       '600648',
       '002462',
       '601077',
       '601992',
       '600269',
       '204007',
       '000900']
```

- 在 业务名称 列里把其他业务的行全部删除

```
[63]: df=pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t",infer_schema=False)
      df=df.with_columns(
          pl.col("发生日期").str.to_date("%Y%m%d"),
          pl.col("证券代码").str.strip_prefix("=").str.strip_chars('"'),
       )
      df=df.filter(
          pl.col("业务名称").is_in(["证券买入", "证券卖出"]),
      )
      df
```

[63]: shape: (13, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
date	str	str	str	str	str	str	str	str	str	str	str	str	str	str	str
2022-07-18	"002462"	"嘉事堂"	"卖出"	"证券卖出"	"09:38:10"	"-10400.00"	"13.2062"	"137344.00"	"137184.67"	"21.98"	"137.35"	"1.38"	"=0.00""	"证券卖出"	"人民币"
2022-07-18	"600408"	"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"	"3.1900"	"149930.00"	"-149955.50"	"23.99"	"=0.00""	"1.51"	"=0.00""	"证券买入"	"人民币"
2022-07-18	"600648"	"外高桥"	"买入"	"证券买入"	"09:44:31"	"11900.00"	"12.6066"	"150019.00"	"-150044.49"	"24.00"	"=0.00""	"1.49"	"=0.00""	"证券买入"	"人民币"
2022-07-18	"600269"	"赣粤高速"	"买入"	"证券买入"	"09:43:38"	"40700.00"	"3.6900"	"150183.00"	"-150208.53"	"24.03"	"=0.00""	"1.50"	"=0.00""	"证券买入"	"人民币"
2022-07-18	"600015"	"华夏银行"	"买入"	"证券买入"	"09:42:51"	"30000.00"	"5.0700"	"152100.00"	"-152125.86"	"24.34"	"=0.00""	"1.52"	"=0.00""	"证券买入"	"人民币"
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2022-07-11	"002462"	"嘉事堂"	"买入"	"证券买入"	"09:38:16"	"10400.00"	"13.5100"	"140504.00"	"-140526.48"	"22.48"	"=0.00""	"1.41"	"=0.00""	"证券买入"	"人民币"
2022-07-11	"000900"	"现代投资"	"买入"	"证券买入"	"09:33:37"	"34400.00"	"4.0500"	"139320.00"	"-139342.29"	"22.29"	"=0.00""	"1.39"	"=0.00""	"证券买入"	"人民币"
2022-07-11	"601992"	"金隅集团"	"买入"	"证券买入"	"09:37:25"	"54000.00"	"2.5900"	"139860.00"	"-139883.80"	"22.38"	"=0.00""	"1.42"	"=0.00""	"证券买入"	"人民币"
2022-07-11	"600894"	"广日股份"	"买入"	"证券买入"	"09:36:30"	"21400.00"	"6.5400"	"139956.00"	"-139979.80"	"22.39"	"=0.00""	"1.41"	"=0.00""	"证券买入"	"人民币"
2022-07-11	"601077"	"渝农商行"	"买入"	"证券买入"	"09:34:24"	"38300.00"	"3.6500"	"139795.00"	"-139818.75"	"22.37"	"=0.00""	"1.38"	"=0.00""	"证券买入"	"人民币"

➢ 把 成交时间 列转换为 polars.Time 类型

[65]:
```
df=pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t",infer_schema=False)
df=df.with_columns(
    pl.col("发生日期").str.to_date("%Y%m%d"),
    pl.col("证券代码").str.strip_prefix("=").str.strip_chars('"'),
    pl.col("成交时间").str.to_time(),
 )
df=df.filter(
    pl.col("业务名称").is_in(["证券买入", "证券卖出"]),
)
df[:,"成交时间"].sort()
```

[65]: shape: (13,)

成交时间
time
09:33:37
09:34:24
09:36:30
09:37:25
09:38:10
…
09:39:28
09:42:51
09:43:38
09:44:31
09:44:52

➢ 把 成交数量、成交价格 等几个数值类型的列都转换为 polars.Float64 类型

[71]:
```
df=pl.read_csv("stock_trades/202207-湘财.xls",encoding="gb18030",separator="\t",infer_schema=False)
df=df.with_columns(
    pl.selectors.all().str.strip_prefix("=").str.strip_chars('"'),
).with_columns(
    pl.col("发生日期").str.to_date("%Y%m%d"),
    pl.col("证券代码").str.strip_prefix("=").str.strip_chars('"'),
    pl.col("成交时间").str.to_time(),
    pl.col("成交数量","成交价格","成交金额","发生金额","手续费","印花税","过户费","其他费").cast(pl.Float64),
)
df=df.filter(
    pl.col("业务名称").is_in(["证券买入", "证券卖出"]),
)
df
```

[71]: shape: (13, 16)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种
date	str	str	str	str	time	f64	f64	f64	f64	f64	f64	f64	f64	str	str
2022-07-18	"002462"	"嘉事堂"	"卖出"	"证券卖出"	09:38:10	-10400.0	13.2062	137344.0	137184.67	21.98	137.35	1.38	0.0	"证券卖出"	"人民币"
2022-07-18	"600408"	"安泰集团"	"买入"	"证券买入"	09:44:52	47000.0	3.19	149930.0	-149955.5	23.99	0.0	1.51	0.0	"证券买入"	"人民币"
2022-07-18	"600648"	"外高桥"	"买入"	"证券买入"	09:44:31	11900.0	12.6066	150019.0	-150044.49	24.0	0.0	1.49	0.0	"证券买入"	"人民币"
2022-07-18	"600269"	"赣粤高速"	"买入"	"证券买入"	09:43:38	40700.0	3.69	150183.0	-150208.53	24.03	0.0	1.5	0.0	"证券买入"	"人民币"
2022-07-18	"600015"	"华夏银行"	"买入"	"证券买入"	09:42:51	30000.0	5.07	152100.0	-152125.86	24.34	0.0	1.52	0.0	"证券买入"	"人民币"
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
2022-07-11	"002462"	"嘉事堂"	"买入"	"证券买入"	09:38:16	10400.0	13.51	140504.0	-140526.48	22.48	0.0	1.41	0.0	"证券买入"	"人民币"
2022-07-11	"000900"	"现代投资"	"买入"	"证券买入"	09:33:37	34400.0	4.05	139320.0	-139342.29	22.29	0.0	1.39	0.0	"证券买入"	"人民币"
2022-07-11	"601992"	"金隅集团"	"买入"	"证券买入"	09:37:25	54000.0	2.59	139860.0	-139883.8	22.38	0.0	1.42	0.0	"证券买入"	"人民币"
2022-07-11	"600894"	"广日股份"	"买入"	"证券买入"	09:36:30	21400.0	6.54	139956.0	-139979.8	22.39	0.0	1.41	0.0	"证券买入"	"人民币"
2022-07-11	"601077"	"渝农商行"	"买入"	"证券买入"	09:34:24	38300.0	3.65	139795.0	-139818.75	22.37	0.0	1.38	0.0	"证券买入"	"人民币"

❖ 利用 pathlib.Path.glob() 遍历所有 *-湘财.xls 文件， 添加一个新列券商

[62]:
```
def read_df_湘财 (f:str |Path)-> pl.DataFrame:
    df=pl.read_csv(
       f,
       encoding="gb18030",
       separator="\t",
       infer_schema=False
    )
    df=df.with_columns(
    pl.selectors.all().str.strip_prefix("=").str.strip_chars('"'),
    ).with_columns(
       pl.col("发生日期").str.to_date("%Y%m%d"),
       pl.col("证券代码").str.strip_prefix("=").str.strip_chars('"'),
       pl.col("成交时间").str.to_time(),
       pl.col("成交数量","成交价格","成交金额","发生金额","手续费","印花税","过户费","其他费").cast(pl.Float64),
    )
    df=df.filter(
    pl.col("业务名称").is_in(["证券买入", "证券卖出"]),
    )
    return df
```

[52]:
```
from pathlib import Path
```

[67]:
```
df =[read_df_湘财(f) for f in (Path("stock_trades/").glob("*-湘财.xls"))]
```

[59]:
```
len(df)
```

[59]:
```
13
```

[70]:
```
d1=pl.concat(df)
```

[73]:
```
d1.with_columns(券商=pl.lit("湘财"),)
```

[73]: shape: (257, 17)

发生日期	证券代码	证券名称	买卖标志	业务名称	成交时间	成交数量	成交价格	成交金额	发生金额	手续费	印花税	过户费	其他费	备注	币种	券商
date	str	str	str	str	time	f64	f64	f64	f64	f64	f64	f64	f64	str	str	str
2022-07-18	"002462"	"嘉事堂"	"卖出"	"证券卖出"	09:38:10	-10400.0	13.2062	137344.0	137184.67	21.98	137.35	1.38	0.0	"证券卖出"	"人民币"	"湘财"
2022-07-18	"600408"	"安泰集团"	"买入"	"证券买入"	09:44:52	47000.0	3.19	149930.0	-149955.5	23.99	0.0	1.51	0.0	"证券买入"	"人民币"	"湘财"
2022-07-18	"600648"	"外高桥"	"买入"	"证券买入"	09:44:31	11900.0	12.6066	150019.0	-150044.49	24.0	0.0	1.49	0.0	"证券买入"	"人民币"	"湘财"
2022-07-18	"600269"	"赣粤高速"	"买入"	"证券买入"	09:43:38	40700.0	3.69	150183.0	-150208.53	24.03	0.0	1.5	0.0	"证券买入"	"人民币"	"湘财"

❖ 清洗202305-海通普通.xlsx数据

- ➢ 将成交日期列转换类型

[85]:
```
df=pl.read_excel(
    "stock_trades/202305-海通普通.xlsx",
    schema_overrides={
        "成交日期":pl.String,
        },
)
df.with_columns(
    pl.col("成交日期").str.to_date("%Y%m%d"),
)
```

[85]: shape: (10, 14)

证券代码	证券名称	成交日期	成交时间	成交数量	成交价格	成交金额	发生金额	操作	手续费	印花税	过户费	其他费	备注
str	str	date	str	i64	f64	f64	f64	str	f64	i64	f64	i64	str
"300107"	"建新股份"	2023-05-29	""	0	0.0	600.0	600.0	"卖"	0.0	0	0.0	0	"建新股份,深圳现金红利"
"131810"	"R-001"	2023-05-24	""	1500	100.005	150007.66	150007.66	"买"	0.0	0	0.0	0	"R-001,拆出购回"
"799999"	"指定登记"	2023-05-24	"15:00:00"	0	0.0	0.0	0.0	"指"	0.0	0	0.0	0	"指定登记指定交易"
"600626"	"申达股份"	2023-05-24	"10:06:14"	14800	3.37	49876.0	-49881.48	"买"	4.99	0	0.49	0	"申达股份证券买入"
"600178"	"东安动力"	2023-05-24	"09:59:17"	16400	6.07	99548.0	-99558.97	"买"	9.95	0	1.02	0	"东安动力证券买入"
"603002"	"宏昌电子"	2023-05-24	"09:54:48"	9800	5.06	49588.0	-49593.46	"买"	4.96	0	0.5	0	"宏昌电子证券买入"
"131810"	"R-001"	2023-05-23	"10:05:49"	1500	1.865	150000.0	-150001.5	"卖"	1.5	0	0.0	0	"到期日[20230524]，利息[7.66]，金额[1500...
"300107"	"建新股份"	2023-05-23	"10:01:57"	10000	5.0	50000.0	-50005.0	"买"	5.0	0	0.0	0	"建新股份证券买入"
"002224"	"三力士"	2023-05-23	"09:58:07"	10800	4.59	49572.0	-49576.96	"买"	4.96	0	0.0	0	"三力士证券买入"
"799998"	"指定撤销"	2023-05-22	"15:00:00"	0	0.0	0.0	0.0	"撤"	0.0	0	0.0	0	"指定撤销撤消指定"

- 把这些多余的字符去掉

[86]:
```
df=pl.read_excel(
    "stock_trades/202305-海通普通.xlsx",
    schema_overrides={
        "成交日期":pl.String,
        "成交时间":pl.String,
        },
)
df.with_columns(
    pl.col("成交日期").str.to_date("%Y%m%d"),
    pl.col("成交时间").replace({"":None}).str.to_time("%H:%M:%S"),
)
```

[86]: shape: (10, 14)

证券代码	证券名称	成交日期	成交时间	成交数量	成交价格	成交金额	发生金额	操作	手续费	印花税	过户费	其他费	备注
str	str	date	time	i64	f64	f64	f64	str	f64	i64	f64	i64	str
"300107"	"建新股份"	2023-05-29	null	0	0.0	600.0	600.0	"卖"	0.0	0	0.0	0	"建新股份,深圳现金红利"
"131810"	"R-001"	2023-05-24	null	1500	100.005	150007.66	150007.66	"买"	0.0	0	0.0	0	"R-001,拆出购回"
"799999"	"指定登记"	2023-05-24	15:00:00	0	0.0	0.0	0.0	"指"	0.0	0	0.0	0	"指定登记指定交易"
"600626"	"申达股份"	2023-05-24	10:06:14	14800	3.37	49876.0	-49881.48	"买"	4.99	0	0.49	0	"申达股份证券买入"
"600178"	"东安动力"	2023-05-24	09:59:17	16400	6.07	99548.0	-99558.97	"买"	9.95	0	1.02	0	"东安动力证券买入"
"603002"	"宏昌电子"	2023-05-24	09:54:48	9800	5.06	49588.0	-49593.46	"买"	4.96	0	0.5	0	"宏昌电子证券买入"
"131810"	"R-001"	2023-05-23	10:05:49	1500	1.865	150000.0	-150001.5	"卖"	1.5	0	0.0	0	"到期日[20230524]，利息[7.66]，金额[1500...
"300107"	"建新股份"	2023-05-23	10:01:57	10000	5.0	50000.0	-50005.0	"买"	5.0	0	0.0	0	"建新股份证券买入"
"002224"	"三 力 士"	2023-05-23	09:58:07	10800	4.59	49572.0	-49576.96	"买"	4.96	0	0.0	0	"三 力 士证券买入"
"799998"	"指定撤销"	2023-05-22	15:00:00	0	0.0	0.0	0.0	"撤"	0.0	0	0.0	0	"指定撤销撤消指定"

- 把成交时间列里值是空字符串的行全部删除

:
```
df=pl.read_excel(
    "stock_trades/202305-海通普通.xlsx",
    schema_overrides={
        "成交日期":pl.String,
        "成交时间":pl.String,
        },
)
df.filter(pl.col("成交时间")!="").with_columns(
    pl.col("成交日期").str.to_date("%Y%m%d"),
    pl.col("成交时间").str.to_time(),
)
```

: shape: (8, 14)

证券代码	证券名称	成交日期	成交时间	成交数量	成交价格	成交金额	发生金额	操作	手续费	印花税	过户费	其他费	备注
str	str	date	time	i64	f64	f64	f64	str	f64	i64	f64	i64	str
"799999"	"指定登记"	2023-05-24	15:00:00	0	0.0	0.0	0.0	"指"	0.0	0	0.0	0	"指定登记指定交易"
"600626"	"申达股份"	2023-05-24	10:06:14	14800	3.37	49876.0	-49881.48	"买"	4.99	0	0.49	0	"申达股份证券买入"
"600178"	"东安动力"	2023-05-24	09:59:17	16400	6.07	99548.0	-99558.97	"买"	9.95	0	1.02	0	"东安动力证券买入"
"603002"	"宏昌电子"	2023-05-24	09:54:48	9800	5.06	49588.0	-49593.46	"买"	4.96	0	0.5	0	"宏昌电子证券买入"
"131810"	"R-001"	2023-05-23	10:05:49	1500	1.865	150000.0	-150001.5	"卖"	1.5	0	0.0	0	"到期日[20230524]，利息[7.66]，金额[1500...
"300107"	"建新股份"	2023-05-23	10:01:57	10000	5.0	50000.0	-50005.0	"买"	5.0	0	0.0	0	"建新股份证券买入"
"002224"	"三 力 士"	2023-05-23	09:58:07	10800	4.59	49572.0	-49576.96	"买"	4.96	0	0.0	0	"三 力 士证券买入"
"799998"	"指定撤销"	2023-05-22	15:00:00	0	0.0	0.0	0.0	"撤"	0.0	0	0.0	0	"指定撤销撤消指定"

- 把操作列里，除了买和卖外的行全部删除

```
df=pl.read_excel(
    "stock_trades/202305-海通普通.xlsx",
    schema_overrides={
        "成交日期":pl.String,
        "成交时间":pl.String,
        },
)
df.filter(pl.col("成交时间")!="").filter(
    pl.col("操作").is_in(["买","卖"])
 ).with_columns(
    pl.col("成交日期").str.to_date("%Y%m%d"),
    pl.col("成交时间").str.to_time(),
)
```

shape: (6, 14)

证券代码	证券名称	成交日期	成交时间	成交数量	成交价格	成交金额	发生金额	操作	手续费	印花税	过户费	其他费	备注
str	str	date	time	i64	f64	f64	f64	str	f64	i64	f64	i64	str
"600626"	"申达股份"	2023-05-24	10:06:14	14800	3.37	49876.0	-49881.48	"买"	4.99	0	0.49	0	"申达股份证券买入"
"600178"	"东安动力"	2023-05-24	09:59:17	16400	6.07	99548.0	-99558.97	"买"	9.95	0	1.02	0	"东安动力证券买入"
"603002"	"宏昌电子"	2023-05-24	09:54:48	9800	5.06	49588.0	-49593.46	"买"	4.96	0	0.5	0	"宏昌电子证券买入"
"131810"	"R-001"	2023-05-23	10:05:49	1500	1.865	150000.0	-150001.5	"卖"	1.5	0	0.0	0	"到期日[20230524]，利息[7.66]，金额[1500...
"300107"	"建新股份"	2023-05-23	10:01:57	10000	5.0	50000.0	-50005.0	"买"	5.0	0	0.0	0	"建新股份证券买入"
"002224"	"三力士"	2023-05-23	09:58:07	10800	4.59	49572.0	-49576.96	"买"	4.96	0	0.0	0	"三力士证券买入"

- 证券名称列里存在 R-001 之类的国债逆回购的行都删除

[95]:
```
df=pl.read_excel(
    "stock_trades/202305-海通普通.xlsx",
    schema_overrides={
        "成交日期":pl.String,
        "成交时间":pl.String,
        },
)
df.filter(
    (pl.col("成交时间")!="")
    & (pl.col("操作").is_in(["买","卖"]))
    & (~pl.col("证券代码").str.starts_with("204"))
    & (~pl.col("证券代码").str.starts_with("1318"))
).with_columns(
    pl.col("成交日期").str.to_date("%Y%m%d"),
    pl.col("成交时间").str.to_time(),
)
```

[95]: shape: (5, 14)

证券代码	证券名称	成交日期	成交时间	成交数量	成交价格	成交金额	发生金额	操作	手续费	印花税	过户费	其他费	备注
str	str	date	time	i64	f64	f64	f64	str	f64	i64	f64	i64	str
"600626"	"申达股份"	2023-05-24	10:06:14	14800	3.37	49876.0	-49881.48	"买"	4.99	0	0.49	0	"申达股份证券买入"
"600178"	"东安动力"	2023-05-24	09:59:17	16400	6.07	99548.0	-99558.97	"买"	9.95	0	1.02	0	"东安动力证券买入"
"603002"	"宏昌电子"	2023-05-24	09:54:48	9800	5.06	49588.0	-49593.46	"买"	4.96	0	0.5	0	"宏昌电子证券买入"
"300107"	"建新股份"	2023-05-23	10:01:57	10000	5.0	50000.0	-50005.0	"买"	5.0	0	0.0	0	"建新股份证券买入"
"002224"	"三力士"	2023-05-23	09:58:07	10800	4.59	49572.0	-49576.96	"买"	4.96	0	0.0	0	"三力士证券买入"

❖ 遍历所有 *-海通普通.xlsx 文件，都进行以上处理，再添加一个新列券商，每行的值都填 "海通普通"

[182]:
```
df = [read_df_海通普通(p) for p in Path("stock_trades/").glob("*-海通普通.xlsx")]
df=pl.concat(df)
d2=df.with_columns(券商=pl.lit("海通普通"),)
d2
```

[182]: shape: (53, 15)

证券代码	证券名称	成交日期	成交时间	成交数量	成交价格	成交金额	发生金额	操作	手续费	印花税	过户费	其他费	备注	券商
str	str	date	time	f64	f64	f64	f64	str	f64	f64	f64	f64	str	str
"600626"	"申达股份"	2023-05-24	10:06:14	14800.0	3.37	49876.0	-49881.48	"买"	4.99	0.0	0.49	0.0	"申达股份证券买入"	"海通普通"
"600178"	"东安动力"	2023-05-24	09:59:17	16400.0	6.07	99548.0	-99558.97	"买"	9.95	0.0	1.02	0.0	"东安动力证券买入"	"海通普通"
"603002"	"宏昌电子"	2023-05-24	09:54:48	9800.0	5.06	49588.0	-49593.46	"买"	4.96	0.0	0.5	0.0	"宏昌电子证券买入"	"海通普通"
"300107"	"建新股份"	2023-05-23	10:01:57	10000.0	5.0	50000.0	-50005.0	"买"	5.0	0.0	0.0	0.0	"建新股份证券买入"	"海通普通"
"002224"	"三 力 士"	2023-05-23	09:58:07	10800.0	4.59	49572.0	-49576.96	"买"	4.96	0.0	0.0	0.0	"三 力 士证券买入"	"海通普通"
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
"605196"	"华通线缆"	2023-07-17	10:03:09	7200.0	7.81	56232.0	56169.59	"卖"	5.62	56.23	0.56	0.0	"华通线缆证券卖出"	"海通普通"
"002842"	"翔鹭钨业"	2023-07-11	10:04:39	5900.0	8.68	51212.0	51155.66	"卖"	5.12	51.22	0.0	0.0	"翔鹭钨业证券卖出"	"海通普通"
"002331"	"皖通科技"	2023-07-11	10:03:52	6900.0	7.33	50577.0	50521.36	"卖"	5.06	50.58	0.0	0.0	"皖通科技证券卖出"	"海通普通"
"600300"	"维维股份"	2023-07-11	10:13:04	16400.0	3.05	50020.0	-50025.5	"买"	5.0	0.0	0.5	0.0	"维维股份证券买入"	"海通普通"
"600287"	"江苏舜天"	2023-07-11	10:08:50	9900.0	5.06	50094.0	-50099.51	"买"	5.01	0.0	0.5	0.0	"江苏舜天证券买入"	"海通普通"

❖ 遍历所有 *-海通两融.xlsx 文件，进行以上处理，再添加一个新列券商，每行的值都填 "海通两融"

[183]:
```
df = [read_df_海通普通(p) for p in Path("stock_trades/").glob("*-海通两融.xlsx")]
df=pl.concat(df)
d3=df.with_columns(券商=pl.lit("海通两融"),)
d3
```

[183]: shape: (53, 15)

证券代码	证券名称	成交日期	成交时间	成交数量	成交价格	成交金额	发生金额	操作	手续费	印花税	过户费	其他费	备注	券商
str	str	date	time	f64	f64	f64	f64	str	f64	f64	f64	f64	str	str
"600638"	"新黄浦"	2023-07-26	09:33:56	9300.0	6.526	60696.0	60628.01	"卖"	6.68	60.7	0.61	0.0	"新黄浦证券卖出"	"海通两融"
"002492"	"恒基达鑫"	2023-07-27	09:30:42	8700.0	6.12	53244.0	53184.91	"卖"	5.86	53.23	0.0	0.0	"恒基达鑫证券卖出"	"海通两融"
"002136"	"安 纳 达"	2023-07-27	09:32:52	4400.0	11.881	52277.0	52218.97	"卖"	5.75	52.28	0.0	0.0	"安 纳 达证券卖出"	"海通两融"
"603967"	"中创物流"	2023-08-22	09:36:13	10600.0	9.36	99216.0	99104.85	"卖"	10.91	99.24	1.0	0.0	"中创物流证券卖出"	"海通两融"
"603967"	"中创物流"	2023-08-04	10:07:37	10600.0	9.43	99958.0	-99970.01	"买"	11.0	0.0	1.01	0.0	"中创物流证券买入"	"海通两融"
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
"300464"	"星徽股份"	2023-10-18	09:46:15	16100.0	5.74	92414.0	92358.97	"卖"	8.82	46.21	0.0	0.0	"星徽股份证券卖出"	"海通两融"
"002661"	"克明食品"	2023-10-18	09:55:41	8500.0	9.42	80072.0	-80079.64	"买"	7.64	0.0	0.0	0.0	"克明食品证券买入"	"海通两融"
"002753"	"永东股份"	2023-10-09	09:48:02	14200.0	7.02	99684.0	-99693.51	"买"	9.51	0.0	0.0	0.0	"永东股份证券买入"	"海通两融"
"000698"	"沈阳化工"	2023-10-09	09:45:18	23800.0	4.053	96460.0	96402.56	"卖"	9.2	48.24	0.0	0.0	"沈阳化工证券卖出"	"海通两融"
"605288"	"凯迪股份"	2023-10-09	09:44:40	2500.0	40.838	102094.0	102032.21	"卖"	9.74	51.05	1.0	0.0	"凯迪股份证券卖出"	"海通两融"

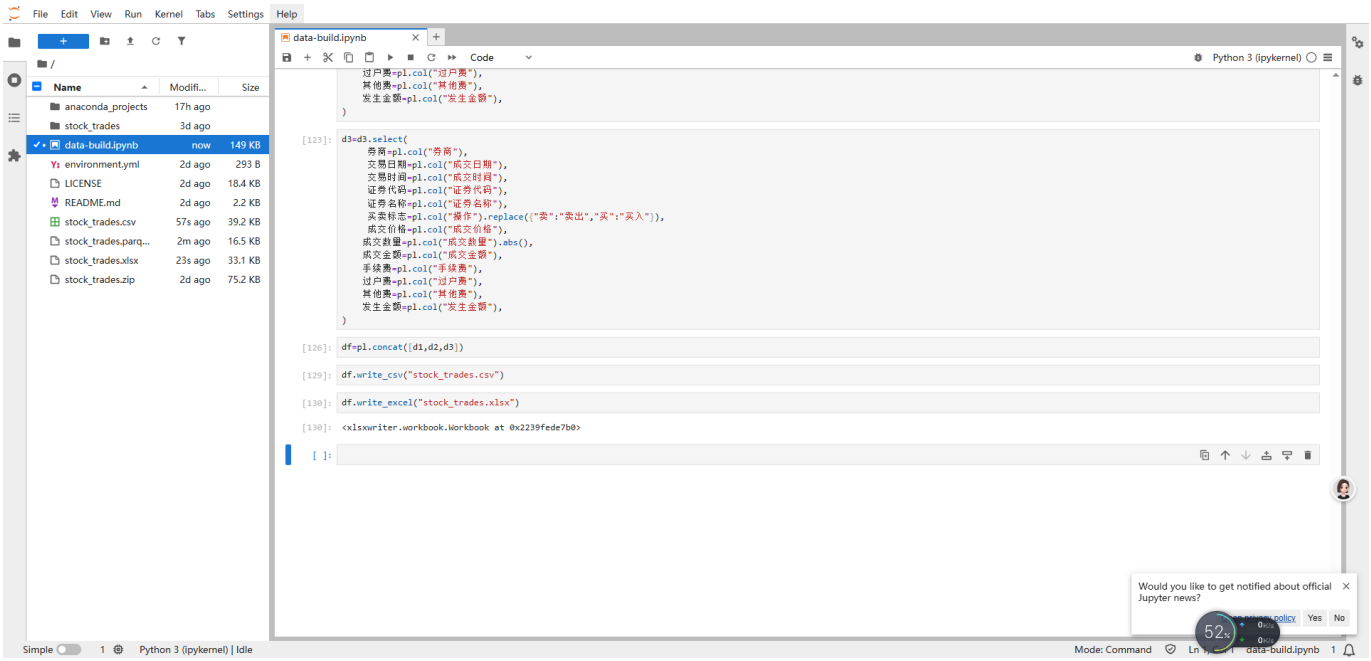
❖ 把 d1、d2、d3 合并在一起，保存为 stock_trades.parquet 文件

[123]:
```
d3=d3.select(
    券商=pl.col("券商"),
    交易日期=pl.col("成交日期"),
    交易时间=pl.col("成交时间"),
    证券代码=pl.col("证券代码"),
    证券名称=pl.col("证券名称"),
    买卖标志=pl.col("操作").replace({"卖":"卖出","买":"买入"}),
    成交价格=pl.col("成交价格"),
    成交数量=pl.col("成交数量").abs(),
    成交金额=pl.col("成交金额"),
    手续费=pl.col("手续费"),
    过户费=pl.col("过户费"),
    其他费=pl.col("其他费"),
    发生金额=pl.col("发生金额"),
)
```

[124]:
```
pl.concat([d1,d2,d3])
```

[124]: shape: (363, 13)

券商	交易日期	交易时间	证券代码	证券名称	买卖标志	成交价格	成交数量	成交金额	手续费	过户费	其他费	发生金额
str	date	time	str	str	str	f64	f64	f64	f64	f64	f64	f64
"湘财"	2022-07-18	09:38:10	"002462"	"嘉事堂"	"卖出"	13.2062	10400.0	137344.0	21.98	1.38	0.0	137184.67
"湘财"	2022-07-18	09:44:52	"600408"	"安泰集团"	"买入"	3.19	47000.0	149930.0	23.99	1.51	0.0	-149955.5
"湘财"	2022-07-18	09:44:31	"600648"	"外高桥"	"买入"	12.6066	11900.0	150019.0	24.0	1.49	0.0	-150044.49
"湘财"	2022-07-18	09:43:38	"600269"	"赣粤高速"	"买入"	3.69	40700.0	150183.0	24.03	1.5	0.0	-150208.53
"湘财"	2022-07-18	09:42:51	"600015"	"华夏银行"	"买入"	5.07	30000.0	152100.0	24.34	1.52	0.0	-152125.86
…	…	…	…	…	…	…	…	…	…	…	…	…
"海通两融"	2023-10-18	09:46:15	"300464"	"星徽股份"	"卖出"	5.74	16100.0	92414.0	8.82	0.0	0.0	92358.97
"海通两融"	2023-10-18	09:55:41	"002661"	"克明食品"	"买入"	9.42	8500.0	80072.0	7.64	0.0	0.0	-80079.64
"海通两融"	2023-10-09	09:48:02	"002753"	"永东股份"	"买入"	7.02	14200.0	99684.0	9.51	0.0	0.0	-99693.51
"海通两融"	2023-10-09	09:45:18	"000698"	"沈阳化工"	"卖出"	4.053	23800.0	96460.0	9.2	0.0	0.0	96402.56
"海通两融"	2023-10-09	09:44:40	"605288"	"凯迪股份"	"卖出"	40.838	2500.0	102094.0	9.74	1.0	0.0	102032.21



4.新建 Notebook，完成 数据计算操作

- ❖ 检查每一笔交易的费率
  - ➢ 分别计算每笔交易的 手续费率、印花税率、过户费率

[22]:
```
import polars as pl
from perspective.widget import PerspectiveWidget
```

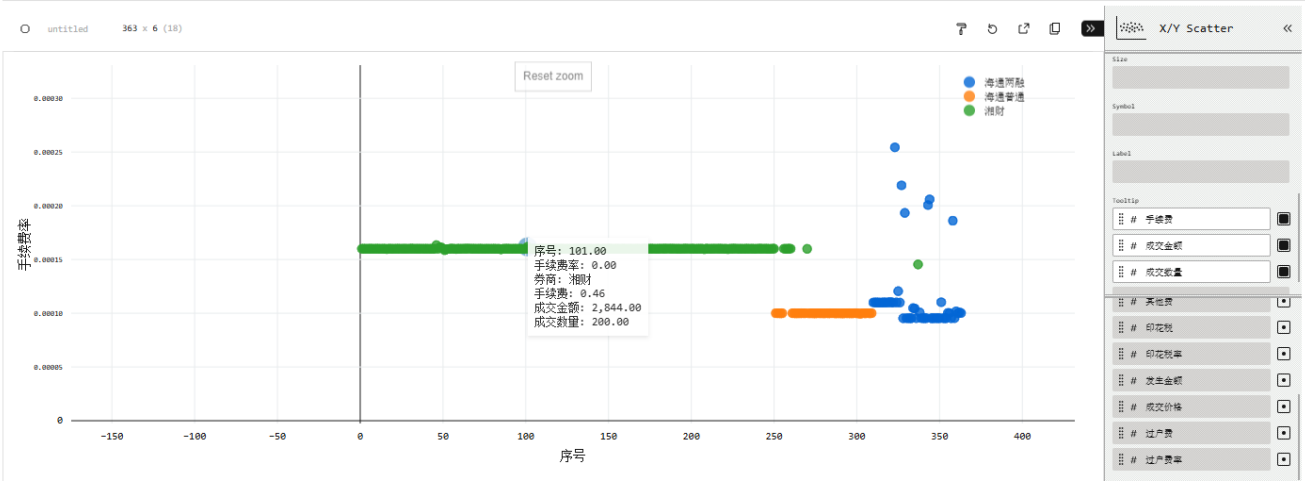
[23]:
```
df=pl.read_parquet("stock_trades.parquet")
df=df.with_columns(
    pl.col("交易时间").cast(pl.String),
    手续费率=pl.col("手续费")/pl.col("成交金额"),
    印花税率=pl.col("印花税")/pl.col("成交金额"),
    过户费率=pl.col("过户费")/pl.col("成交金额"),
)
df=df.with_row_index("序号",1)
df
```

[23]: shape: (363, 18)

序号	券商	交易日期	交易时间	证券代码	证券名称	买卖标志	成交价格	成交数量	成交金额	手续费	印花税	过户费	其他费	发生金额	手续费率	印花税率	过户费率
u32	str	date	str	str	str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64
1	"湘财"	2022-07-18	"09:38:10"	"002462"	"嘉事堂"	"卖出"	13.2062	10400.0	137344.0	21.98	137.35	1.38	0.0	137184.67	0.00016	0.001	0.00001
2	"湘财"	2022-07-18	"09:44:52"	"600408"	"安泰集团"	"买入"	3.19	47000.0	149930.0	23.99	0.0	1.51	0.0	-149955.5	0.00016	0.0	0.00001
3	"湘财"	2022-07-18	"09:44:31"	"600648"	"外高桥"	"买入"	12.6066	11900.0	150019.0	24.0	0.0	1.49	0.0	-150044.49	0.00016	0.0	0.00001
4	"湘财"	2022-07-18	"09:43:38"	"600269"	"赣粤高速"	"买入"	3.69	40700.0	150183.0	24.03	0.0	1.5	0.0	-150208.53	0.00016	0.0	0.00001
5	"湘财"	2022-07-18	"09:42:51"	"600015"	"华夏银行"	"买入"	5.07	30000.0	152100.0	24.34	0.0	1.52	0.0	-152125.86	0.00016	0.0	0.00001
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
359	"海通两融"	2023-10-18	"09:46:15"	"300464"	"星徽股份"	"卖出"	5.74	16100.0	92414.0	8.82	46.21	0.0	0.0	92358.97	0.000095	0.0005	0.0
360	"海通两融"	2023-10-18	"09:55:41"	"002661"	"克明食品"	"买入"	9.42	8500.0	80072.0	7.64	0.0	0.0	0.0	-80079.64	0.000095	0.0	0.0
361	"海通两融"	2023-10-09	"09:48:02"	"002753"	"永东股份"	"买入"	7.02	14200.0	99684.0	9.51	0.0	0.0	0.0	-99693.51	0.000095	0.0	0.0
362	"海通两融"	2023-10-09	"09:45:18"	"000698"	"沈阳化工"	"卖出"	4.053	23800.0	96460.0	9.2	48.24	0.0	0.0	96402.56	0.000095	0.0005	0.0
363	"海通两融"	2023-10-09	"09:44:40"	"605288"	"凯迪股份"	"卖出"	40.838	2500.0	102094.0	9.74	51.05	1.0	0.0	102032.21	0.000095	0.0005	0.00001

[27]:
```
PerspectiveWidget(df)
```

[27]:



- ❖ 计算每支股票是否都已完成平仓

[28]:
```
df.group_by("证券代码", "证券名称").agg(pl.len(), pl.col("买卖标志"))
```

[28]: shape: (152, 4)

证券代码	证券名称	len	买卖标志
str	str	u32	list[str]
"600231"	"凌钢股份"	2	["买入", "卖出"]
"603113"	"金能科技"	2	["买入", "卖出"]
"300132"	"青松股份"	1	["买入"]
"600261"	"阳光照明"	4	["买入", "卖出", … "卖出"]
"300701"	"森霸传感"	3	["买入", "卖出", "卖出"]
…	…	…	…
"002956"	"西麦食品"	4	["买入", "买入", … "卖出"]
"600894"	"广日股份"	2	["买入", "卖出"]
"300214"	"日科化学"	2	["买入", "卖出"]
"600015"	"华夏银行"	3	["买入", "卖出", "卖出"]
"002753"	"永东股份"	2	["买入", "卖出"]

[30]:
```
df.group_by("证券代码", "证券名称").agg(
    结余数量=(
        pl.when(pl.col("买卖标志") == "卖出")
        .then(-pl.col("成交数量"))
        .when(pl.col("买卖标志") == "买入")
        .then(pl.col("成交数量"))
        .sum()
    ),
).sort("结余数量")
```

[30]: shape: (152, 3)

证券代码	证券名称	结余数量
str	str	f64
"603167"	"渤海轮渡"	-13600.0
"300889"	"爱克股份"	-7000.0
"600333"	"长春燃气"	-3900.0
"600292"	"远达环保"	0.0
"002023"	"海特高新"	0.0
…	…	…
"600525"	"长园集团"	8500.0
"300132"	"青松股份"	9600.0
"300022"	"吉峰科技"	10800.0
"300215"	"电科院"	20000.0
"688660"	"电气风电"	20200.0

[32]:
```
df.join(
    df.group_by("证券代码", "证券名称")
    .agg(
        结余数量=(
            pl.when(pl.col("买卖标志") == "卖出")
            .then(-pl.col("成交数量"))
            .when(pl.col("买卖标志") == "买入")
            .then(pl.col("成交数量"))
            .sum()
        ),
    )
    .filter(pl.col("结余数量") < 0),
    on="证券代码",
    how="anti",
)
```

[32]: shape: (358, 18)

序号	券商	交易日期	交易时间	证券代码	证券名称	买卖标志	成交价格	成交数量	成交金额	手续费	印花税	过户费	其他费	发生金额	手续费率	印花税率	过户费率
u32	str	date	str	str	str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64
1	"湘财"	2022-07-11	"09:33:37"	"000900"	"现代投资"	"买入"	4.05	34400.0	139320.0	22.29	0.0	1.39	0.0	-139342.29	0.00016	0.0	0.00001
2	"湘财"	2022-07-11	"09:34:24"	"601077"	"渝农商行"	"买入"	3.65	38300.0	139795.0	22.37	0.0	1.38	0.0	-139818.75	0.00016	0.0	0.00001
3	"湘财"	2022-07-11	"09:36:30"	"600894"	"广日股份"	"买入"	6.54	21400.0	139956.0	22.39	0.0	1.41	0.0	-139979.8	0.00016	0.0	0.00001
4	"湘财"	2022-07-11	"09:37:25"	"601992"	"金隅集团"	"买入"	2.59	54000.0	139860.0	22.38	0.0	1.42	0.0	-139883.8	0.00016	0.0	0.00001
5	"湘财"	2022-07-11	"09:38:16"	"002462"	"嘉事堂"	"买入"	13.51	10400.0	140504.0	22.48	0.0	1.41	0.0	-140526.48	0.00016	0.0	0.00001
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
359	"海通两融"	2023-10-31	"09:31:53"	"002956"	"西麦食品"	"卖出"	14.13	5000.0	70650.0	6.74	35.35	0.0	0.0	70607.91	0.000095	0.0005	0.0
360	"海通两融"	2023-10-31	"09:39:57"	"603214"	"爱婴室"	"买入"	15.84	3100.0	49104.0	5.0	0.0	0.51	0.0	-49109.51	0.000102	0.0	0.00001
361	"海通两融"	2023-10-31	"09:40:55"	"300132"	"青松股份"	"买入"	5.21	9600.0	50016.0	5.0	0.0	0.0	0.0	-50021.0	0.0001	0.0	0.0
362	"海通两融"	2023-10-31	"09:43:13"	"002492"	"恒基达鑫"	"买入"	5.91	8400.0	49644.0	5.0	0.0	0.0	0.0	-49649.0	0.000101	0.0	0.0
363	"海通两融"	2023-10-31	"09:44:45"	"002111"	"威海广泰"	"买入"	9.24	5400.0	49896.0	5.0	0.0	0.0	0.0	-49901.0	0.0001	0.0	0.0

❖ 计算和做图观察这段期间累计的股票持仓数量变化情况

[41]:
```
start_date = df["交易日期"].min()
start_date
```

[41]: `datetime.date(2022, 7, 11)`

[42]:
```
end_date = df["交易日期"].max()
end_date
```

[42]: `datetime.date(2023, 10, 31)`

[43]:
```
k1 = pl.select(日期=pl.date_range(start_date, end_date))
k1
```

[43]: shape: (478, 1)

日期
date
2022-07-11
2022-07-12
2022-07-13
2022-07-14
2022-07-15
…
2023-10-27
2023-10-28
2023-10-29
2023-10-30
2023-10-31

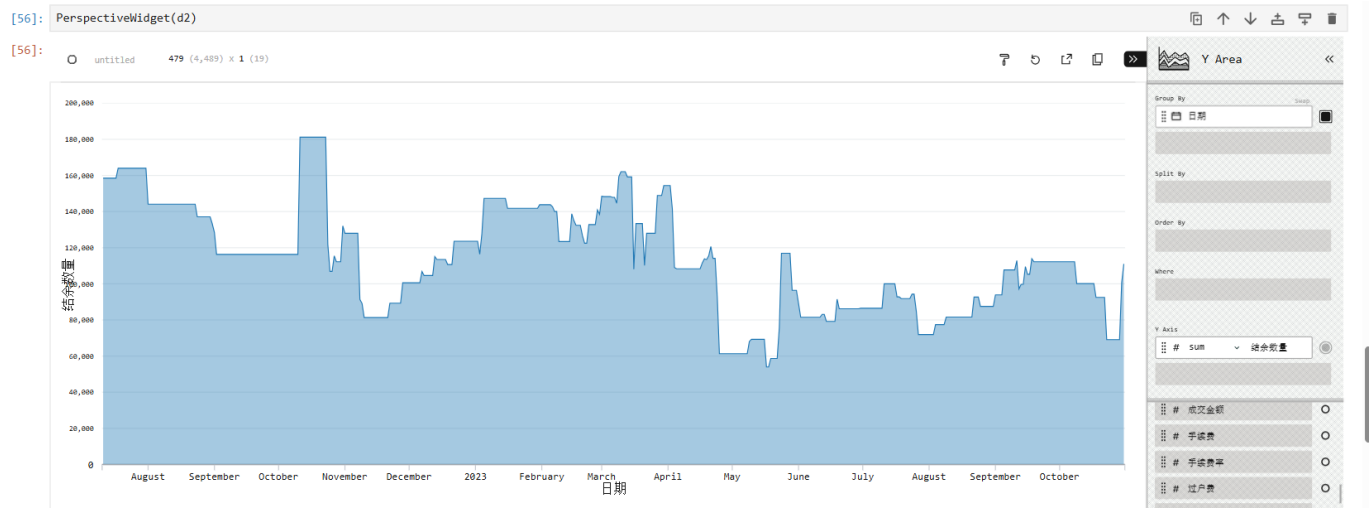
[45]:
```
k2 = df["证券代码"].unique().sort().to_frame()
k2
```

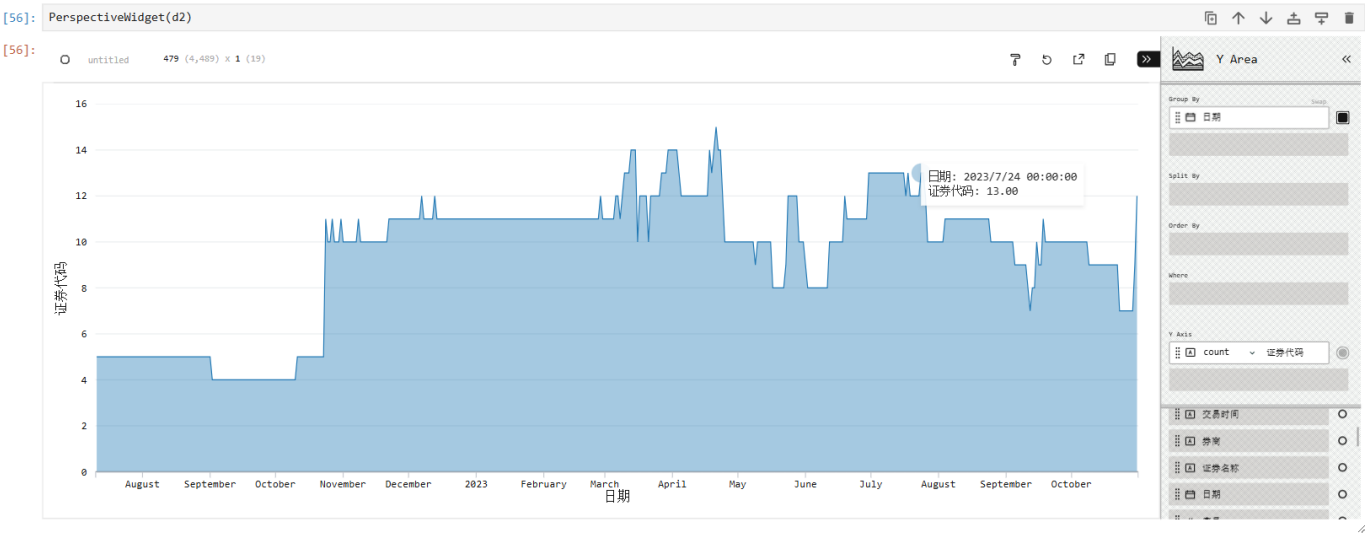
[45]: shape: (152, 1)

证券代码
str
"000096"
"000532"
"000559"
"000599"
"000655"
…
"688299"
"688321"
"688360"
"688393"
"688660"

[47]:
```
k=k1.join(k2, how="cross")
```

[55]:
```
d2 = (
    k.join(
        d1, left_on=["日期", "证券代码"], right_on=["交易日期", "证券代码"], how="left"
    )
    .sort("日期", "证券代码")
    .with_columns(
        结余数量=(
            pl.when(pl.col("买卖标志") == "买入")
            .then(pl.col("成交数量"))
            .when(pl.col("买卖标志") == "卖出")
            .then(-pl.col("成交数量"))
            .otherwise(0)
            .cum_sum()
            .over("证券代码")
        ),
    )
    .filter(pl.col.结余数量 > 0)
)
```





- 计算每日持股的市值的动态变化，并能够由此计算每日投资收益率，并与股市指数的每日收益率 (基准收益) 相对照
  - 调用 Tushare 的 daily 接口，转换出含交易所代码的证券代码

```
[15]: import tushare as ts
[16]: pro = ts.pro_api()
[20]: hq = pro.daily(
          ts_code="002462.SZ",
          start_date=format(start_date, "%Y%m%d"),
          end_date=format(end_date, "%Y%m%d"),
      )
      hq = pl.from_pandas(hq)
      hq
```

[20]: shape: (318, 11)

ts_code	trade_date	open	high	low	close	pre_close	change	pct_chg	vol	amount
str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64
"002462.SZ"	"20231031"	14.75	14.9	14.59	14.7	14.75	-0.05	-0.339	65859.96	96984.271
"002462.SZ"	"20231030"	13.88	14.89	13.88	14.75	13.93	0.82	5.8866	123932.16	180119.372
"002462.SZ"	"20231027"	13.7	13.98	13.51	13.93	13.64	0.29	2.1261	35782.0	49386.168
"002462.SZ"	"20231026"	13.49	13.68	13.4	13.64	13.62	0.02	0.1468	19215.0	26005.866
"002462.SZ"	"20231025"	13.65	13.77	13.58	13.62	13.67	-0.05	-0.3658	18484.0	25274.163
…	…	…	…	…	…	…	…	…	…	…
"002462.SZ"	"20220715"	13.61	13.66	13.12	13.13	13.59	-0.46	-3.3848	32967.65	44114.064
"002462.SZ"	"20220714"	13.54	13.75	13.5	13.59	13.54	0.05	0.3693	21967.0	29851.164
"002462.SZ"	"20220713"	13.55	13.63	13.39	13.54	13.61	-0.07	-0.5143	22793.0	30714.624
"002462.SZ"	"20220712"	13.65	13.69	13.41	13.61	13.65	-0.04	-0.293	29679.0	40146.31
"002462.SZ"	"20220711"	13.2	13.96	13.07	13.65	13.18	0.47	3.566	62827.0	85869.11

[26]:
```
ts_codes = (
    d1.select(
        证券代码=(
            pl.when(pl.col("证券代码").str.head(1).is_in(["0", "3"]))
            .then(pl.format("{}.SZ", pl.col("证券代码")))
            .when(pl.col("证券代码").str.head(1) == "6")
            .then(pl.format("{}.SH", pl.col("证券代码")))
        ),
    )
    .to_series()
    .unique()
    .sort()
    .to_list()
)
```

[27]:
```
from tqdm.notebook import tqdm
```

[32]:
```
hq=[
    pl.from_pandas(
      pro.daily(
         ts_code=ts_code,
    start_date=format(start_date, "%Y%m%d"),
    end_date=format(end_date, "%Y%m%d"),
     )
    )
    for ts_code in tqdm(ts_codes)
]
```

100% 149/149 [00:11<00:00, 16.63it/s]

[33]:
```
len(hq)
```

[33]:
```
149
```

[34]:
```
hq=pl.concat(hq)
```

[35]:
```
hq
```

- 使用 tqdm 软件包显示进度条，全部获取后合并，保存为 daily.parquet 文件

[36]:
```
hq.write_parquet("daily.parquet")
```

[40]:
```
hq = pl.read_parquet("daily.parquet")
hq = hq.with_columns(
    pl.col("ts_code").str.head(6),
    pl.col("trade_date").str.to_date("%Y%m%d"),
)
```

[42]:
```
d1.join(
    hq, left_on=["交易日期", "证券代码"], right_on=["trade_date", "ts_code"], how="left"
).with_columns(
    vratio=pl.col("成交数量")/100/pl.col("vol"),
)
```

- 检查每一行的交割单 成交数量 占股票 成交量 的比例

[42]: shape: (358, 28)

序号	券商	交易日期	交易时间	证券代码	证券名称	买卖标志	成交价格	成交数量	成交金额	手续费	印花税	过户费	其他费	发生金额	手续费率	印花税率	过户费率	open	high	low	close	pre_close
u32	str	date	str	str	str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64
1	"湘财"	2022-07-11	"09:33:37"	"000900"	"现代投资"	"买入"	4.05	34400.0	139320.0	22.29	0.0	1.39	0.0	-139342.29	0.00016	0.0	0.00001	4.08	4.13	4.04	4.12	4.06
2	"湘财"	2022-07-11	"09:34:24"	"601077"	"渝农商行"	"买入"	3.65	38300.0	139795.0	22.37	0.0	1.38	0.0	-139818.75	0.00016	0.0	0.00001	3.65	3.68	3.64	3.66	3.65
3	"湘财"	2022-07-11	"09:36:30"	"600894"	"广日股份"	"买入"	6.54	21400.0	139956.0	22.39	0.0	1.41	0.0	-139979.8	0.00016	0.0	0.00001	6.57	6.57	6.49	6.51	6.57