

```
(base) 12460@LAPTOP-T1VTLFCA MINGW64 /d/gitcode/week08 (main)
$ conda activate week08
(week08)
12460@LAPTOP-T1VTLFCA MINGW64 /d/gitcode/week08 (main)
$ curl -O https://raw.gitcode.com/cueb-fintech/courses/blobs/8e70be13d8672dd685672f6
% Total % Received % Xferd Average Speed Time Time Time Current
Dload Upload Total Spent Left Speed
100 77002 0 77002 0 0 14262 0 --:--:-- 0:00:05 --:--:-- 14446
(week08)
12460@LAPTOP-T1VTLFCA MINGW64 /d/gitcode/week08 (main)
$ unzip stock_trades.zip
Archive: stock_trades.zip
creating: stock_trades/
inflating: stock_trades/202207-湘财.xls
inflating: stock_trades/202208-湘财.xls
inflating: stock_trades/202209-湘财.xls
inflating: stock_trades/202210-湘财.xls
inflating: stock_trades/202211-湘财.xls
inflating: stock_trades/202212-湘财.xls
inflating: stock_trades/202301-湘财.xls
```

```
[1]: import polars as pl
```

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

```
[5]: 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
i64	str	str	str	str	str	str	f64	f64	f64
20220721	"600269"	"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"=0.00"	3.6	4884.0	4884.0
20220718	"204007"	"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	1.675	58000.0	58018.63
				"证...					

```
[15]: df[:,0]
```

```
[15]: shape: (17,)
```

```
发生日期
i64
20220721
20220718
20220718
20220718
20220718
...
20220711
20220711
20220711
20220711
20220707
```

```
[21]: 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'
$ 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'
```

```
[26]: df[:, "证券名称":"成交数量"].sample(5)
```

```
[26]: shape: (5, 5)
```

证券名称	买卖标志	业务名称	成交时间	成交数量
str	str	str	str	str
"广日股份"	"买入"	"证券买入"	"09:36:30"	"21400.00"
"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"
"GC007"	"卖出"	"拆出质押购回"	"19:17:51"	"-7580.00"
"华夏银行"	"买入"	"证券买入"	"09:42:51"	"30000.00"
"赣粤高速"	"买入"	"证券买入"	"09:43:38"	"40700.00"

```
[27]: df.describe
```

```
[27]: <bound method DataFrame.describe of shape: (17, 16)>
```

发生日期	证券代码	证券名称	买卖标志	...	过户费	其他费	备注	币种
---	---	---	---	...	---	---	---	---
164	str	str	str	...	str	str	str	str
20220721	600269	赣粤高速	卖出	...	"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	...	"0.00"	"0.00"	融券购回:18.63实际占款天数: 7-8 88880	人民币
20220718	"002462"	嘉事堂	卖出	...	1.38	"0.00"	证券卖出	人民币
20220718	600408	安泰集团	买入	...	1.51	"0.00"	证券买入	人民币
20220718	600648	外高桥	买入	...	1.49	"0.00"	证券买入	人民币
...	...	...	...	...	...	...	...	...
20220711	204007	GC007	卖出	...	"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	...	1.42	"0.00"	证券买入	人民币
20220711	600894	广日股份	买入	...	1.41	"0.00"	证券买入	人民币
20220711	601077	渝农商行	买入	...	1.38	"0.00"	证券买入	人民币
20220707	204007	GC007	卖出	...	"0.00"	"0.00"	融券购回:357.61实际占款天数: 7- 888880	人民币

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

	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
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
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
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

```
[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, 2.461 ]]
```

```
[32]: df.to_dicts()
```

```
[32]: [{ '发生日期': 20220721,
      '证券代码': '600269',
      '证券名称': '赣粤高速',
      '买卖标志': '卖出',
      '业务名称': '股息入账',
      '成交时间': '16:00:00',
      '成交数量': '=0.00',
      '成交价格': 3.6,
      '成交金额': 4884.0,
      '发生金额': 4884.0,
      '手续费': '=0.00',
      '印花税': '=0.00',
      '过户费': '=0.00',
      '其他费': '=0.00',
      '备注': '股息入账:赣粤高速600269; 权益股数:40700;',
      '币种': '人民币'},
    { '发生日期': 20220718,
      '证券代码': '204007',
      '证券名称': 'GC007',
      '买卖标志': '卖出',
      '业务名称': '拆出质押购回',
      '成交时间': '19:03:27',
      '成交数量': '-580.00',
      '成交价格': 1.675,
      '成交金额': 58000.0,
```

[34]: s = df.to\_series()

[35]: s.name

[35]: '发生日期'

[36]: s.dtype

[36]: Int64

[37]: s.shape

[37]: (17,)

[38]: s[:5]

[38]: shape: (5,)

发生日期

i64

20220721

20220718

[60]: 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"),  
)  
df

[60]: 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
date	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"
2022-07-18	"204007"	"GC007"	"卖出"	"拆出质押"	"19:03:27"	"-580.00"	"1.6750"	"58000.00"	"58018.63"	"=-0.00"	"=-0.00"	"=-0.00"

[76]: 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

[76]: 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"

[84]: 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").cast(pl.Utf8).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

[84]: 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"



```
[87]: 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(''),
)
df = df.with_columns(
    pl.col("发生日期").cast(pl.Utf8).str.to_date("%Y%m%d"),
    pl.col("成交时间").str.to_time(),
    pl.col("成交数量").cast(pl.Float64),
)
df = df.filter(
    pl.col("业务名称").is_in(["证券买入", "证券卖出"]),
)
df
```

[87]: 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	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	-10400.0	"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	47000.0	"3.1900"	"149930.00"	"-149955.50"	"23.99"	"0.00"	"1.51"	"0.00"	"证券买入"	"人民币"

```
[94]: [f.name for f in Path("stock_trades/").glob("*-湘财.xls")]
```

```
[94]: ['202207-湘财.xls',
'202208-湘财.xls',
'202209-湘财.xls',
'202210-湘财.xls',
'202211-湘财.xls',
'202212-湘财.xls',
'202301-湘财.xls',
'202302-湘财.xls',
'202303-湘财.xls',
'202304-湘财.xls',
'202305-湘财.xls',
'202306-湘财.xls',
'202309-湘财.xls']
```

```
: 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"),
)
```

: 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"

```
[140]: df = pl.concat(df)
```

```
[141]: d2 = df.with_columns(
    券商=pl.lit("海通普通"),
)
d2
```

[141]: shape: (30, 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	str	str	i64	f64	f64	f64	str	f64	i64	f64	i64	str	str
"300107"	"建新股份"	"20230529"	" "	0	0.0	600.0	600.0	"卖"	0.0	0	0.0	0	"建新股份,深圳现金红利"	"海通普通"
"131810"	"R-001"	"20230524"	" "	1500	100.005	150007.66	150007.66	"买"	0.0	0	0.0	0	"R-001,拆出购回"	"海通普通"
"799999"	"指定登记"	"20230524"	"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"	"海通普通"
"600626"	"申达股份"	"20230524"	"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"	"海通普通"
"600178"	"东安动力"	"20230524"	"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"	"海通普通"

```
[183]: d3 = d3.select(
    券商=p1.col("券商"),
    交易日期=p1.col("成交日期"),
    交易时间=p1.col("成交时间"),
    证券代码=p1.col("证券代码"),
    证券名称=p1.col("证券名称"),
    买卖标志=p1.col("操作").replace(("卖": "卖出", "买": "买入")),
    成交价格=p1.col("成交价格"),
    成交数量=p1.col("成交数量").abs(),
    成交金额=p1.col("成交金额"),
    手续费=p1.col("手续费"),
    印花费=p1.col("印花税"),
    过户费=p1.col("过户费"),
    其他费=p1.col("其他费"),
    发生金额=p1.col("发生金额"),
)

[186]: df = p1.concat([d1, d2, d3])
```

```
[201]: p1.read_parquet("stock_trades.parquet")
```

[201]: shape: (363, 14)

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	印花费	过户费	其他费	发生金额
str	date	time	str	str	str	f64	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	137.35	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	0.0	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	0.0	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	0.0	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	0.0	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	46.21	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	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	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	48.24	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	51.05	1.0	0.0	102032.21

第四节

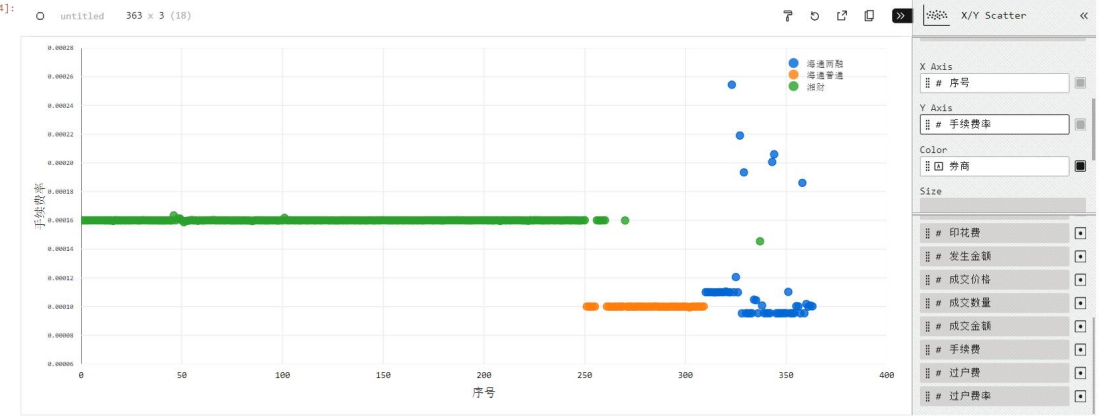
```
[211]: import polars as pl
from perspective.widget import PerspectiveWidget
```

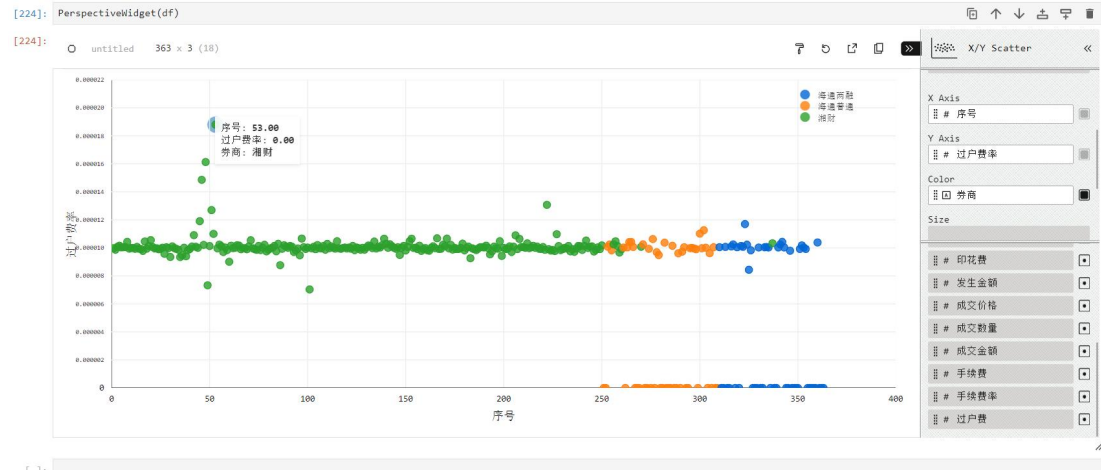
```
[214]: df = p1.read_parquet("stock_trades.parquet")
df = df.with_columns(
    手续费率=p1.col("手续费") / p1.col("成交金额"),
    印花税率=p1.col("印花费") / p1.col("成交金额"),
    过户费率=p1.col("过户费") / p1.col("成交金额"),
)
df = df.with_row_index("序号", 1)
df
```

[214]: 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	印花费	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	time	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

```
4]: PerspectiveWidget(df)
```



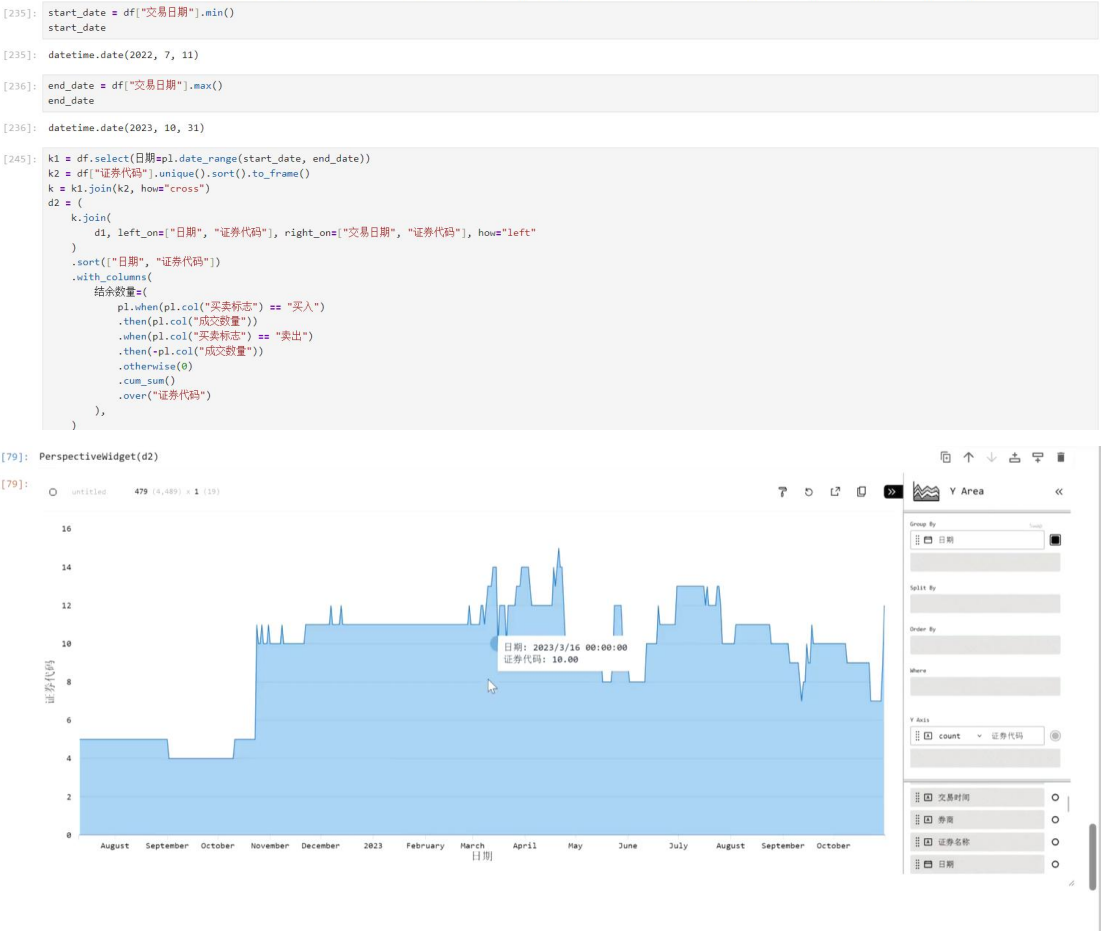


```
[231]: d1 = df.join(
df.group_by("证券代码", "证券名称")
    .agg(
        结余数量=(
            pl.when(pl.col("买卖标志") == "买入")
            .then(pl.col("成交数量"))
            .when(pl.col("买卖标志") == "卖出")
            .then(-pl.col("成交数量"))
            .sum()
        ),
    )
    .filter(pl.col("结余数量") < 0),
    on="证券代码",
    how="anti"
)

[232]: d1
```

```
[232]: shape: (358, 18)
```

[illegible]



## 第四部分

```
[10]: import tushare as ts

[11]: pro = ts.pro_api()

[13]: hq = pro.daily(ts_codes="002462.sz", start_dates="20220711", end_dates="20230731")

[14]: hq
```

	ts_code	trade_date	open	high	low	close	pre_close	change	pct_chg	vol	amount
0	002462.SZ	20230731	14.09	14.18	13.92	13.94	14.09	-0.15	-1.0646	50048.00	70042.798
1	002462.SZ	20230728	14.10	14.16	13.97	14.09	14.09	0.00	0.0000	22203.00	31262.160
2	002462.SZ	20230727	14.07	14.23	14.01	14.09	14.07	0.02	0.1421	24770.00	34947.762
3	002462.SZ	20230726	14.04	14.12	14.00	14.07	14.05	0.02	0.1423	17943.00	25209.741
4	002462.SZ	20230725	13.96	14.05	13.90	14.05	13.96	0.09	0.6447	26352.00	36828.767
...	...	...	...	...	...	...	...	...	...	...	...
253	002462.SZ	20220715	13.61	13.66	13.12	13.13	13.59	-0.46	-3.3848	32967.65	44114.064
254	002462.SZ	20220714	13.54	13.75	13.50	13.59	13.54	0.05	0.3693	21967.00	29851.164
255	002462.SZ	20220713	13.55	13.63	13.39	13.54	13.61	-0.07	-0.5143	22793.00	30714.624
256	002462.SZ	20220712	13.65	13.69	13.41	13.61	13.65	-0.04	-0.2930	29679.00	40146.310
257	002462.SZ	20220711	13.20	13.96	13.07	13.65	13.18	0.47	3.5660	62827.00	85869.110



```
[97]: 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()
)
```

```
[98]: ts_codes
```

```
[82]: len(ts_codes)
```

```
[82]: 149
```

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

```
•[103]: hq = []
for ts_code in tqdm(ts_codes):
    df = pro.daily(
        ts_codes=ts_code,
        start_date=format(start_date, "%Y%m%d"),
        end_date=format(end_date, "%Y%m%d"),
    )
    if df is not None and not df.empty:
        polars_df = pl.from_pandas(df)
        hq.append(polars_df)

# 后续如果需要合并所有获取到的数据
if hq:
    hq = pl.concat(hq)
else:
    hq = pl.DataFrame()
```

100% 149/149 [00:24<00:00, 5.98it/s]

```
[104]: hq
```

```
[104]: shape: (47_355, 11)
```

```
6]: hq.write_parquet("daily.parquet")
```

```
0]: 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"),
)
```

```
7]: d1.join(
    hq, left_on=["交易日期", "证券代码"], right_on=["trade_date", "ts_code"], how="left"
).filter(
    ~pl.col("成交价格").is_between(pl.col("low"), pl.col("high")),
)
```

```
7]: shape: (0, 27)
```

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	印花费	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	change	pct_chg	vol	amount
u32	str	date	str	str	str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	

```
]:
```

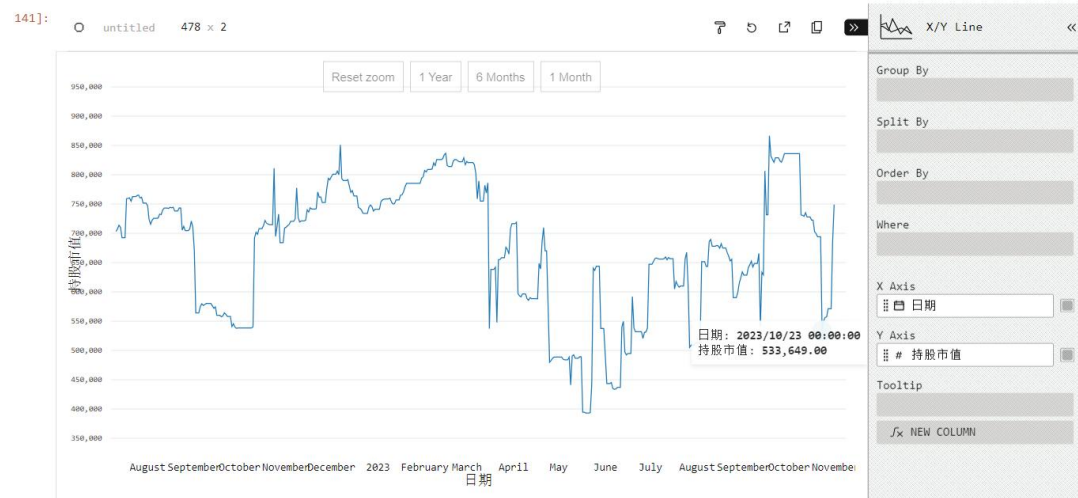


```
[129]: d4 = (
    k.join(
        d1, left_on=["日期", "证券代码"], right_on=["交易日期", "证券代码"], how="left"
    )
    .sort(["日期", "证券代码"])
    .with_columns(
        结余数量=(
            pl.when(pl.col("买卖标志") == "买入")
            .then(pl.col("成交数量"))
            .when(pl.col("买卖标志") == "卖出")
            .then(-pl.col("成交数量"))
            .otherwise(0)
            .cum_sum()
            .over("证券代码")
        ),
    )
)
d4
```

[129]: shape: (72\_671, 19)

```
.40]: d5 = (
    d4.join(
        hq, left_on=["日期", "证券代码"], right_on=["trade_date", "ts_code"], how="left"
    )
    .sort(["证券代码", "日期"])
    .with_columns(close=pl.col("close").fill_null(strategy="forward").over("证券代码"))
    .with_columns(持股市值=pl.col("结余数量") * pl.col("close"))
    .group_by("日期")
    .agg(pl.col("持股市值").sum())
    .sort("日期")
)
```

.41]: PerspectiveWidget(d5)



```

]: d5 = (
    d4.join(
        hq, left_on=["日期", "证券代码"], right_on=["trade_date", "ts_code"], how="left"
    )
    .sort(["证券代码", "日期"])
    .with_columns(close=p1.col("close").fill_null(strategy="forward").over("证券代码"))
    .with_columns(持股市值=p1.col("结余数量") * p1.col("close"))
    .group_by("日期")
    .agg(p1.col("持股市值").sum(), p1.col("发生金额").sum())
    .sort("日期")
    .with_columns(
        转账金额=p1.when(p1.int_range(0, p1.len()) == 0).then(100_0000).otherwise(0)
    )
    .with_columns(
        现金余额=(p1.col("转账金额") + p1.col("发生金额")).cum_sum(),
    )
    .with_columns(
        总资产=(p1.col("持股市值") + p1.col("现金余额"))
    )
)
d5

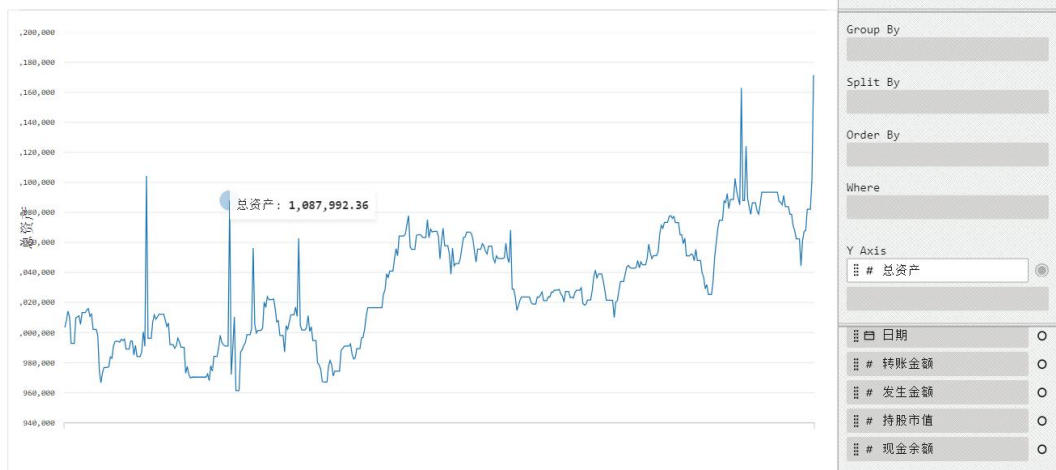
```

```
]: shape: (478, 6)
```

日期	持股市值	发生金额	转账金额	现金余额	总资产
date	f64	f64	i32	f64	f64

```
152]: PerspectiveWidget(d5)
```

```
152]: O untitled 478 x 1 (6)
```



```

[162]: ihq = pro.index_daily(
    ts_code="000300.SH",
    start_date=format(start_date, "%Y%m%d"),
    end_date=format(end_date, "%Y%m%d"),
    fields="ts_code,trade_date,pct_chg",
)

```

```

[163]: pl.from_pandas(ihq).write_parquet("index_daily.parquet")

```

```

[172]: ihq = pl.read_parquet("index_daily.parquet")
ihq = (
    ihq.with_columns(pl.col("pct_chg") / 100 + 1)
    .sort("trade_date")
    .with_columns(car=pl.col("pct_chg").cum_prod())
    .with_columns(沪深300=pl.col("car") * 100_000)
)
ihq

```

```

[172]: shape: (318, 5)

```



```
[188]: d6 = d5.join(ihq, left_on="日期", right_on="trade_date")
d6 = d6.unpivot(
    on=["总资产", "沪深300"], index="日期", variable_name="资产类型", value_name="财富"
)
```

[188]: shape: (318, 10)

日期	持股市值	发生金额	转账金额	现金余额	总资产	ts_code	pct_chg	car	沪深300
date	f64	f64	i32	f64	f64	str	f64	f64	f64
2022-07-11	703040.0	-699551.12	1000000	300448.88	1.0035e6	"000300.SH"	0.983254	0.983254	98325.4
2022-07-12	707714.0	0.0	0	300448.88	1.0082e6	"000300.SH"	0.990585	0.973997	97399.666359
2022-07-13	713855.0	0.0	0	300448.88	1.0143e6	"000300.SH"	1.001818	0.975767	97576.738952
2022-07-14	710079.0	0.0	0	300448.88	1.0105e6	"000300.SH"	1.000142	0.975906	97590.594849
2022-07-15	692377.0	0.0	0	300448.88	992825.88	"000300.SH"	0.982983	0.959299	95929.895697
...	...	...	...	...	...	...	...	...	...
2023-10-25	556522.0	0.0	0	510845.96	1.0674e6	"000300.SH"	1.004969	0.791288	79128.845378
2023-10-26	557109.0	0.0	0	510845.96	1.0680e6	"000300.SH"	1.002764	0.793476	79347.557506
2023-10-27	571195.0	0.0	0	510845.96	1.0820e6	"000300.SH"	1.013727	0.804368	80436.761428
2023-10-30	686345.0	-94884.93	0	415961.03	1.1023e6	"000300.SH"	1.006003	0.809196	80919.623307
2023-10-31	748947.0	6490.78	0	422451.81	1.1714e6	"000300.SH"	0.996856	0.806652	80665.212012

[187]: PerspectiveWidget(d6)