

第八周 数据清洗与计算 (初级)

1. 读取使用 `polars.read_excel()` 函数读取名称为 202207-湘财.xls 的文件

```
[6]: pl.read_csv("stock_trades/202207-湘财.xls", encoding="GB18030", separator="\t")  
[6]: 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	拆出质押回购	19:03:27	-580.00	1.675	58000.0	58018.63
20220718	002462	嘉事堂	卖出	证券卖出	09:38:10	-10400.00	13.2062	137344.0	137184.67
20220718	600408	安泰集团	买入	证券买入	09:44:52	47000.00	3.19	149930.0	-149955.5
20220718	600648	外高桥	买入	证券买入	09:44:31	11900.00	12.6066	150019.0	-150044.49
...
...	质押

```
has_header: 'bool' = True,  
columns: 'Sequence[int] | Sequence[str] | None'  
new_columns: 'Sequence[str] | None' = None,  
separator: 'str' = ',',  
comment_prefix: 'str | None' = None,  
quote_char: 'str | None' = '',  
skip_rows: 'int' = 0,  
skip_lines: 'int' = 0,  
schema: 'SchemaDict | None' = None,  
schema_overrides: 'Mapping[str, PolarsDataType] | None' = None,  
null_values: 'str | Sequence[str] | dict[str, str] | None' = None,  
missing_utf8_is_empty_string: 'bool' = False,  
ignore_errors: 'bool' = False,  
try_parse_dates: 'bool' = False,  
n_threads: 'int | None' = None,  
infer_schema: 'bool' = True,  
infer_schema_length: 'int | None' = 100,  
batch_size: 'int' = 8192,  
n_rows: 'int | None' = None,  
encoding: 'CsvEncoding | str' = 'utf8',  
low_memory: 'bool' = False,  
rechunk: 'bool' = False,  
use_pyarrow: 'bool' = False,  
storage_options: 'dict[str, Any] | None' = None,  
skip_rows_after_header: 'int' = 0,  
row_index_name: 'str | None' = None,  
row_index_offset: 'int' = 0,  
sample_size: 'int' = 1024,  
eol_char: 'str' = '\n',  
raise_if_empty: 'bool' = True,  
truncate_ragged_lines: 'bool' = False,  
decimal_comma: 'bool' = False,  
glob: 'bool' = True,  
) -> 'DataFrame'  
Source:  
@deprecate_renamed_parameter("dtypes", "schema_overr  
@deprecate_renamed_parameter("row_count_name", "row_  
version="0.20.4")
```

1	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
2	20220721	600269	赣粤高速	卖出	股息入账	16:00:00	0.00	3.6000	4884.00	4884.00
3	20220718	204007	GC007	卖出	拆出质押回购	19:03:27	-580.00	1.6750	58000.00	58018.63
4	20220718	002462	嘉事堂	卖出	证券卖出	09:38:10	-10400.00	13.2062	137344.00	137184.67
5	20220718	600408	安泰集团	买入	证券买入	09:44:52	47000.00	3.1900	149930.00	-149955.50
6	20220718	600648	外高桥	买入	证券买入	09:44:31	11900.00	12.6066	150019.00	-150044.49
7	20220718	600269	赣粤高速	买入	证券买入	09:43:38	40700.00	3.6900	150183.00	-150208.53
8	20220718	600015	华夏银行	买入	证券买入	09:42:51	30000.00	5.0700	152100.00	-152125.86
9	20220718	601992	金隅集团	卖出	证券卖出	09:39:28	-54000.00	2.5683	138686.00	138523.74
10	20220718	600894	广日股份	卖出	证券卖出	09:39:06	-21400.00	6.5419	139996.00	139832.12
11	20220718	601077	渝农商行	卖出	证券卖出	09:38:30	-38300.00	3.5800	137114.00	136953.55
12	20220711	002462	嘉事堂	买入	证券买入	09:38:16	10400.00	13.5100	140504.00	-140526.00
13	20220711	000900	现代投资	买入	证券买入	09:33:37	34400.00	4.0500	139320.00	-139320.00
14	20220711	204007	GC007	卖出	质押回购拆出	09:39:27	580.00	1.6750	58000.00	-58002.90
15	20220711	601992	金隅集团	买入	证券买入	09:37:25	54000.00	2.5900	139860.00	-139883.80
16	20220711	600894	广日股份	买入	证券买入	09:36:30	21400.00	6.5400	139956.00	-139979.80
17	20220711	601077	渝农商行	买入	证券买入	09:34:24	38300.00	3.6500	139795.00	-139818.75
18	20220707	204007	GC007	卖出	拆出质押回购	19:17:51	-7580.00	2.4600	758000.00	758357.61
19										

2. 掌握以下几个检查 `polars.DataFrame` 对象时常用的属性(attributes)/方法(methods): 以下为部分截图:

```
|: df =pl.read_csv("stock_trades/202207-湘财.xls",encoding="GB18030",separator="\t")
```

```
|: df.shape
```

```
|: (17, 16)
```

```
|: df.height
```

```
|: 17
```

```
|: df.width
```

```
|: 16
```

```
|: df.is_empty()
```

```
|: False
```

```
»]: df.dtypes
```

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

```
»]: list(df.schema.keys())== df.columns
```

```
»]: True
```

```
l]: list(df.schema.values())== df.columns
```

```
l]: False
```

```
] |
```

3]: df[0]

3]: 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"

4]: df[1]

4]: 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"

5]: df[2]

5]: 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	"=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"	

6]: df[-1]

6]: 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
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"

[27]: df[3:5]

[27]: 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	"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"	

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

[28]: 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"

```
[33]: df[:,["证券名称","成交金额"]]
```

[33]: shape: (17, 2)

证券名称	成交金额
str	f64
"赣粤高速"	4884.0
"GC007"	58000.0
"嘉事堂"	137344.0
"安泰集团"	149930.0
"外高桥"	150019.0
...	...
"GC007"	58000.0
"金隅集团"	139860.0
"广日股份"	139956.0
"渝农商行"	139795.0
"GC007"	758000.0

```
[35]: df[:5, "证券名称": "成交金额"]
```

```
[35]: shape: (5, 7)
```

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
str	str	str	str	str	f64	f64
"赣粤高速"	"卖出"	"股息入账"	"16:00:00"	"=0.00"	3.6	4884.0
"GC007"	"卖出"	"拆出质押购回"	"19:03:27"	"-580.00"	1.675	58000.0
"嘉事堂"	"卖出"	"证券卖出"	"09:38:10"	"-10400.00"	13.2062	137344.0
"安泰集团"	"买入"	"证券买入"	"09:44:52"	"47000.00"	3.19	149930.0
"外高桥"	"买入"	"证券买入"	"09:44:31"	"11900.00"	12.6066	150019.0

```
[ ]:
```

```
[36]: df.row(0)
```

```
[36]: (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',
'=0.00',
'股息入账:赣粤高速600269; 权益股数:40700;',
'人民币')
```

```
[ ]:
```

```
'人民币')
```

```
[38]: r=df.rows()
```

```
[39]: type(r)
```

```
[39]: list
```

```
[40]: type(r[0])
```

```
[40]: tuple
```

```
[41]: len(r)
```

```
[41]: 17
```

```
[42]: len(r[0])
```

```
[42]: 16
```

```
[ ]: 5
```

[45]: df.to_series(2)

[45]: shape: (17,)

证券名称

str

"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"

[]:

[47]: df.head(5)

[47]: 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"

[48]: df.tail(3)

[48]: shape: (3, 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	"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"

[49]: df.sample(3)

[49]: shape: (3, 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	"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"
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"
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"

[51]: df.describe()

[51]: 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"

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

```
[52]: 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

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

```
[54]:
```

	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
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;
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
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
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
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
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
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
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
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
9	20220718	601077	浦农商行	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
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

```
[55]: s= df.to_series()
```

```
[56]: s.name
```

```
[56]: '发生日期'
```

```
[57]: s.dtype
```

```
[57]: Int64
```

```
[58]: s.shape
```

```
[58]: (17,)
```

```
[59]: s.len()
```

```
[59]: 17
```

```
[60]: s[:5]
```

```
[60]: shape: (5,)
      发生日期
      i64
      -----
      20220721
      20220718
      20220718
      20220718
      20220718
```

```
[61]: s[0,5,8]
```

```
[61]: shape: (3,)
      发生日期
      i64
      -----
      20220721
      20220718
      20220718
```

```
[65]: df.to_series(4).unique
```

```
[65]: <bound method Series.unique of shape: (17,)
      Series: '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"
        ...
        "质押回购拆出"
        "证券买入"
        "证券买入"
        "证券买入"
        "拆出质押购回"
      ]>
```

```
1:
```

```
[67]: df.to_series(4).value_counts()
```

```
[67]: shape: (5, 2)
      业务名称  count
      str      u32
      -----
      "拆出质押购回"  2
      "质押回购拆出"  1
      "证券卖出"      4
      "证券买入"      9
      "股息入账"      1
```



```
[68]: df.to_series(8).value_counts()
```

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

成交金额	count
f64	u32
150019.0	1
139956.0	1
138686.0	1
4884.0	1
137114.0	1
...	...
140504.0	1
139320.0	1
137344.0	1
139795.0	1
152100.0	1

```
[69]: 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

```
[72]: s.describe()
```

```
[72]: shape: (9, 2)
```

statistic	value
str	f64
"count"	17.0
"null_count"	0.0
"mean"	2.0221e7
"std"	4.160387
"min"	2.0220707e7
"25%"	2.0220711e7
"50%"	2.0220718e7
"75%"	2.0220718e7
"max"	2.0220721e7

```
[73]: s.null_count()
```

```
[73]: 0
```

```
[70]: s.value_counts()

[70]: shape: (4, 2)
      发生日期  count
      i64    u32
20220707      1
20220721      1
20220718      9
20220711      6

[71]: s.unique()

[71]: shape: (4,)
      发生日期
      i64
20220707
20220711
20220718
20220721

[75]: df[:, "发生金额"].to_list()

[75]: [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]

76]: 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', '0002462', '600408', '600648', '600269', '600015', '601992', '600894', '601077'
$ 证券名称 <str> '赣粤高速', '国泰君安', '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, 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
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'
```

3.浏览 Expressions [文档](#)，寻找并组合恰当的 Expressions，对 df 进行以下清洗操作：

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

[85]: 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"	"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"	"人民币"

```
)
df[:, "证券代码"].unique().to_list()
```

```
88]: ['600648',
      '="000900"',
      '="002462"',
      '204007',
      '600015',
      '600894',
      '601077',
      '601992',
      '600408',
      '600269']
```

```
[95]: df =pl.DataFrame(
      {
        "foo":["asldj1","jjjask","hasna"]
      }
    )
df.with_columns(foo2=pl.col("foo").str.strip_chars("a"))
```

[95]: shape: (3, 2)

foo	foo2
str	str
"asldj1"	"sldj1"
"jjjask"	"jjjask"
"hasna"	"hasn"

```
[96]: df =pl.DataFrame(
      {
        "foo":["asldj1","jjjask","hasna"]
      }
    )
df.with_columns(foo2=pl.col("foo").str.strip_chars("a1"))
```

[96]: shape: (3, 2)

foo	foo2
str	str
"asldj1"	"sldj1"
"jjjask"	"jjjask"
"hasna"	"hasn"

```
[188]: 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("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
```

[188]: 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"

]: 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"
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"	"湘财"
...
2023-06-19	"603967"	"中创物流"	"卖出"	"证券卖出"	10:18:46	-5000.0	9.13	45650.0	45596.59	7.3	45.65	0.46	0.0	"证券卖出"	"人民币"	"湘财"
2023-06-12	"300641"	"正丹股份"	"卖出"	"证券卖出"	13:22:32	-9600.0	5.05	48480.0	48423.76	7.76	48.48	0.48	0.0	"证券卖出"	"人民币"	"湘财"
2023-06-02	"000655"	"金岭矿业"	"卖出"	"证券卖出"	10:26:35	-7500.0	6.48	48600.0	48543.61	7.78	48.61	0.47	0.0	"证券卖出"	"人民币"	"湘财"
2023-06-01	"300214"	"日科化学"	"卖出"	"证券卖出"	10:45:35	-7400.0	6.8	50320.0	50261.63	8.05	50.32	0.51	0.0	"证券卖出"	"人民币"	"湘财"
2023-09-11	"603577"	"汇金通"	"卖出"	"证券卖出"	09:46:10	-5100.0	9.3024	47442.0	47410.89	6.9	23.72	0.49	0.0	"证券卖出"	"人民币"	"湘财"

[146]: d2

[146]: 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"	"翔鹭铝业"	2023-07-11	10:04:39	5900.0	8.68	51212.0	51155.66	"卖"	5.12	51.22	0.0	0.0	"翔鹭铝业证券卖出"	"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"

[147]: d3

[147]: 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"

l]: 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	印花税	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

```
] : df.with_columns(  
    成交金额2=p1.col("成交价格") * p1.col("成交金额"),  
).with_columns(  
    成交金额D=p1.col("成交金额") - p1.col("成交金额2"),  
).sort("成交金额D")
```

] : shape: (363, 16)

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	成交金额2	成交金额D
str	date	time	str	str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64
"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	4.1693e6	-4.0672e6
"海通两融"	2023-09-13	09:52:16	"605288"	"凯迪股份"	"买入"	39.029	2500.0	97573.0	9.31	0.0	1.0	0.0	-97583.31	3.8082e6	-3.7106e6
"湘财"	2022-10-11	09:36:33	"600729"	"重庆百货"	"买入"	20.4053	6800.0	138756.0	22.2	0.0	1.39	0.0	-138779.59	2.8314e6	-2.6926e6
"湘财"	2022-10-24	09:31:49	"600729"	"重庆百货"	"卖出"	20.25	6800.0	137700.0	22.03	137.83	1.36	0.0	137538.78	2.788425e6	-2.650725e6
"湘财"	2023-02-17	10:32:28	"688069"	"德林海环保"	"买入"	28.64	2600.0	74464.0	11.91	0.0	0.75	0.0	-74476.66	2.1326e6	-2.0582e6
...
"湘财"	2022-10-25	09:34:50	"600829"	"人民同泰"	"买入"	6.06	300.0	1818.0	0.29	0.0	0.02	0.0	-1818.31	11017.08	-9199.08
"湘财"	2022-10-25	09:34:17	"300283"	"温州宏丰"	"买入"	5.25	300.0	1575.0	0.25	0.0	0.02	0.0	-1575.25	8268.75	-6693.75
"湘财"	2022-10-25	13:00:01	"300641"	"正丹股份"	"卖出"	5.32	200.0	1064.0	0.17	1.06	0.02	0.0	1062.77	5660.48	-4596.48
"湘财"	2022-10-25	09:25:00	"002998"	"优彩资源"	"卖出"	6.73	100.0	673.0	0.11	0.67	0.01	0.0	672.22	4529.29	-3856.29
"湘财"	2022-10-25	09:30:19	"300368"	"汇金股份"	"卖出"	6.2	100.0	620.0	0.1	0.62	0.01	0.0	619.28	3844.0	-3224.0

```
1]: df = pl.concat([d1, d2, d3])

2]: df.with_columns(
    成交金额2=pl.col("成交价格") * pl.col("成交金额"),
).with_columns(
    成交金额D=pl.col("成交金额") - pl.col("成交金额2"),
).with_columns(
    发生金额D=pl.col("发生金额")
)
```

```
3]: shape (363, 17)
```

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	成交金额2	成交金额D	发生金额D
str	date	time	str	str	str	f64	f64	f64	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	1.8138e6	-1.6764e6	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	478276.7	-328346.7	-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	1.8912e6	-1.7412e6	-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	554175.27	-403992.27	-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	771147.0	-619047.0	-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	530456.36	-438042.36	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	754278.24	-674206.24	-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	699781.68	-600097.68	-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	390952.38	-294492.38	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	4,1693e6	-4,0672e6	102032.21

```
2]: df.with_columns(
    成交金额2=pl.col("成交价格") * pl.col("成交金额"),
).with_columns(
    成交金额D=pl.col("成交金额") - pl.col("成交金额2"),
).with_columns(
    发生金额D=(
        pl.col("发生金额")
        - (
            pl.when(pl.col("买卖标志") == "买入")
            .then(-pl.col("成交金额"))
            .when(pl.col("买卖标志") == "卖出")
            .then(pl.col("成交金额"))
            - pl.col("手续费")
            - pl.col("印花税")
            - pl.col("过户费")
            - pl.col("其他费")
        )
    ).round(4)
).sort("发生金额D")
```

```
3]: shape (363, 17)
```

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	成交金额2	成交金额D	发生金额D
str	date	time	str	str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64
"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	478276.7	-328346.7	0.0
"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	1.8912e6	-1.7412e6	0.0
"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	554175.27	-403992.27	0.0
"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	771147.0	-619047.0	0.0
"湘财"	2022-07-18	09:39:28	"601992"	"金隅集团"	"卖出"	2.5683	54000.0	138686.0	22.19	138.69	1.38	0.0	138523.74	356187.2538	-217501.2538	0.0
...
"湘财"	2022-10-24	09:37:27	"300283"	"温州志丰"	"买入"	5.4038	12900.0	69709.0	11.15	0.0	0.83	0.0	-69720.15	376693.4942	-306984.4942	0.83
"湘财"	2022-10-11	09:32:03	"000900"	"现代投资"	"卖出"	3.8465	34400.0	132321.0	21.17	132.35	1.31	0.0	132167.48	508972.7265	-376651.7265	1.31
"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	1.8138e6	-1.6764e6	1.38
"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	564246.0	-424926.0	1.39
"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	1.8982e6	-1.7577e6	1.41

📄	stock_trades.csv	now	41 KB
📄	stock_trades.par...	1m ago	17.8 KB
📄	stock_trades.xlsx	7s ago	35 KB

4.在 JupyterLab 页面里，新建 Notebook，改名为 data-query.ipynb，在里面按照下面的要求完成数据计算操作：

```
[4]: df.with_columns(  
    手续费率=p1.col("手续费")/p1.col("成交金额"),  
    印花税率=p1.col("印花税")/p1.col("成交金额"),  
    过户费率=p1.col("过户费")/p1.col("成交金额"),  
)
```

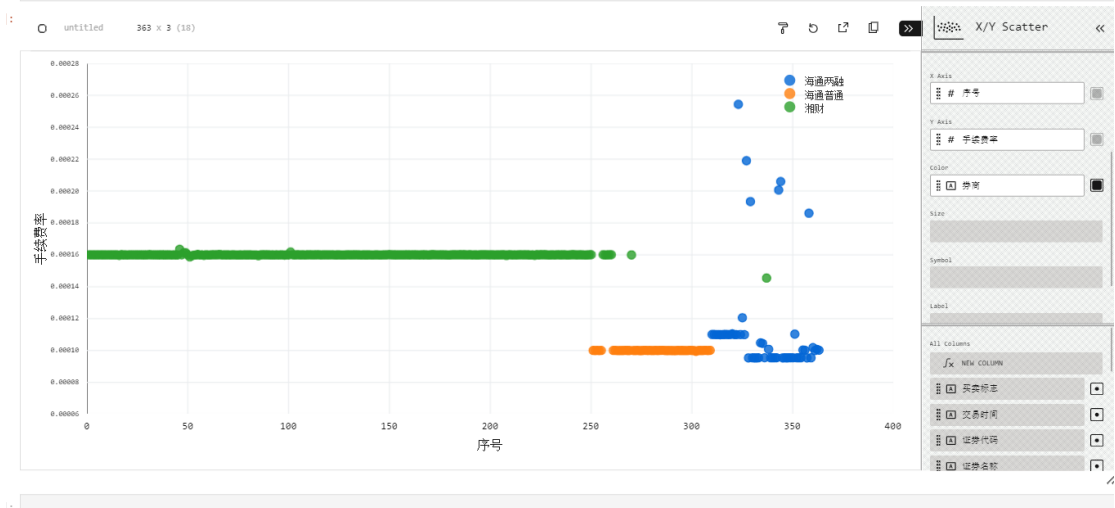
```
[4]: shape: (363, 17)
```

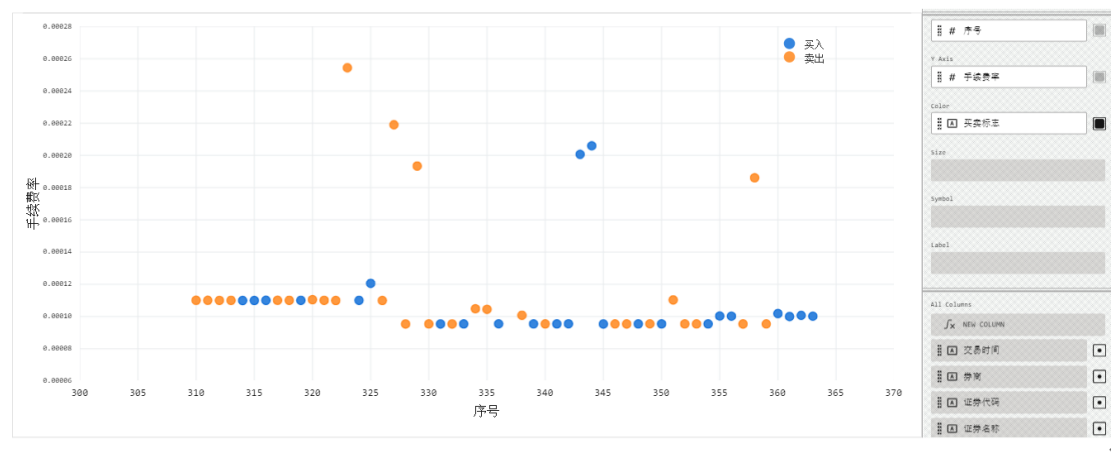
券商	交易日期	交易时间	证券代码	证券名称	买卖标志	成交价格	成交数量	成交金额	手续费	印花税	过户费	其他费	发生金额	手续费率	印花税率	过户费率
str	date	time	str	str	str	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64	f64
"湘财"	2022-07-18	09:38:10	"002462"	"嘉事堂"	"卖出"	13.2062	10400.0	137344.0	21.98	137.35	1.38	0.0	137184.67	0.00016	0.001	0.00001
"湘财"	2022-07-18	09:44:52	"600408"	"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
"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
"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
"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
...
"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
"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
"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
"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
"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

```
[10]: df.with_row_index("序号",1)
```

```
[10]: 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	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
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





```
df.groupby(["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",
)
```

[31]: 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

3]: df

3]: 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"	2022-07-11	"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

d1

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

```
: k1=p1.select(  
    日期=p1.date_range(start_date,end_date)  
)  
k1
```

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

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

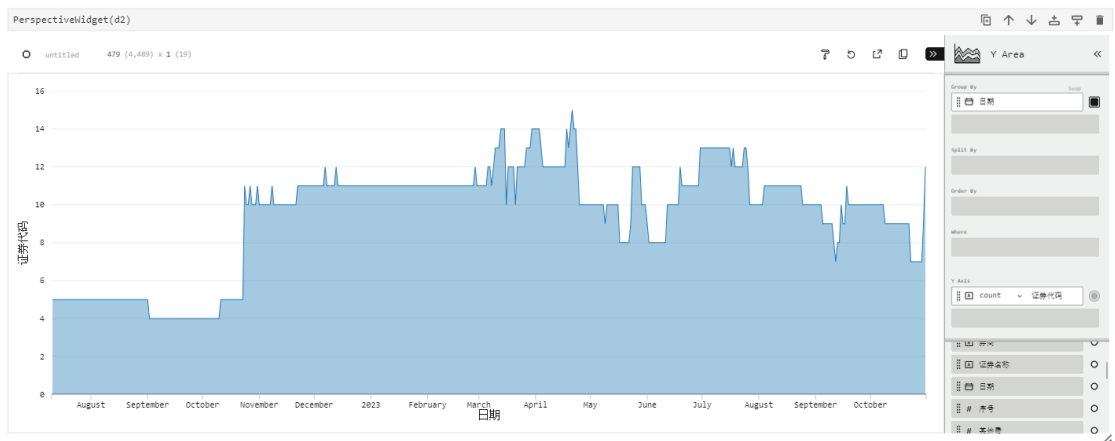
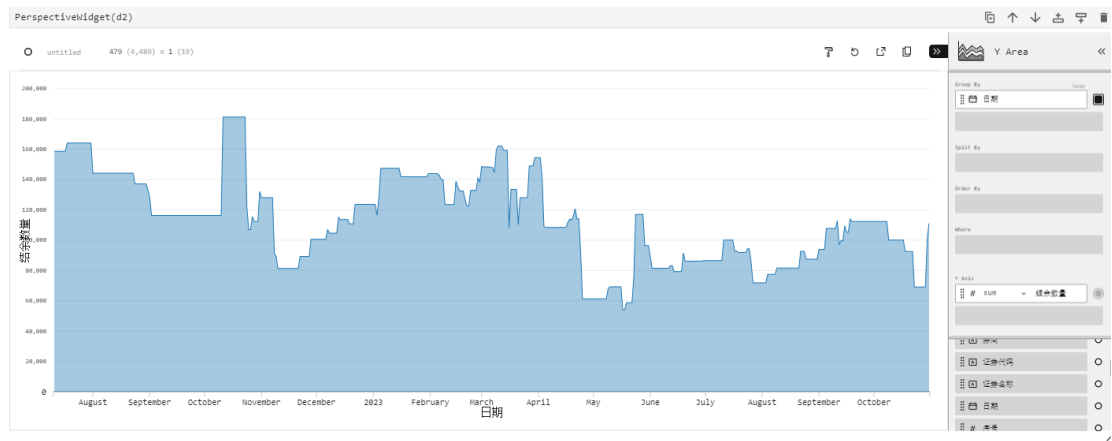
shape: (152, 1)

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

```
k1.join(k2,how="cross")
```

shape: (72_656, 2)

日期	证券代码
date	str
2022-07-11	"000096"
2022-07-11	"000532"
2022-07-11	"000559"
2022-07-11	"000599"
2022-07-11	"000655"
...	...
2023-10-31	"688299"
2023-10-31	"688321"
2023-10-31	"688360"
2023-10-31	"688393"
2023-10-31	"688660"



```
] : 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
```

```
] : 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

```
] :
```

```
: 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()
)
```

```
: from tqdm.notebook import tqdm
```

```
: 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 [00:10<00:00, 18.10it/s]

```
: len(hq)
```

```
: 149
```

```
: hq=pl.concat(hq)
```

```
: hq
```

```
shape: (47_355, 11)
```

[illegible]

```
shape: (358, 28)
```

u3	券种	交易日期	交割时间	证券代码	证券名称	买卖	成交均价	成交量	成交金额	手续费	印花税	过户费	其他费	发生盈亏	手续费	印花税率	过户费率	open	high	low	close	pre_close	change	pet_change	val	amount	varatio
u3	date	date		str	str	str	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4		fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4	fsd4
237	"沪"	2023-05-09	"09:33:16"	"601166"	"兴业银行"	"卖出"	17.75	3000.00	53250.00	8.52	353.25	0.53	0.00	53187.7	0.00016	0.001	0.00001	17.78	17.94	17.5	17.56	17.63	-0.07	-0.3971	1627946	2881046	0.000018
48	"沪"	2023-05-10	"09:30:19"	"300368"	"汇川技术"	"卖出"	6.2	100.00	620.00	0.1	0.62	0.01	0.00	619.28	0.00016	0.00001	0.00001	6.178	6.15	5.95	6.06	6.2	-0.14	-2.2581	4292159	2991193	0.000002
49	"沪"	2023-10-25	"09:30:39"	"002996"	"博瑞通信"	"卖出"	13.64	100.00	1364.02	0.22	1.36	0.01	0.00	1362.42	0.00016	0.000997	0.000007	13.57	13.93	13.26	13.82	13.64	0.18	1.3196	36358.22	53212.19	0.000026
235	"沪"	2023-04-25	"13:35:07"	"601166"	"兴业银行"	"买入"	17.15	3000.00	51450.0	8.23	0.0	0.51	0.00	-51458.74	0.00016	0.00001	0.00001	16.96	17.29	16.96	17.27	16.93	0.34	2.0083	959324.22	1707165	0.000003
46	"沪"	2023-05-10	"09:25:00"	"002998"	"兴发集团"	"卖出"	6.73	100.00	673.0	0.11	0.67	0.01	0.00	672.22	0.00163	0.000996	0.000015	6.73	6.73	6.56	6.64	6.72	-0.08	-1.1905	161450.4	10687.814	0.000082
30	"沪"	2022-10-11	"09:41:06"	"600231"	"海药股份"	"买入"	2.2	62700.00	137940.0	22.07	0.0	1.29	0.00	-137963.36	0.00016	0.0000009	0.0000009	2.21	2.22	2.15	2.2	2.21	-0.01	-0.4525	61371.2	13366.328	0.010217
3	"沪"	2022-10-11	"09:36:30"	"600884"	"广汇能源"	"买入"	6.14	21400.00	139956.0	22.09	0.0	1.41	0.00	-139978.9	0.00016	0.0000001	0.0000001	6.57	6.57	6.49	6.51	6.57	-0.06	-0.9132	20773.5	13537.134	0.010321
142	"沪"	2023-02-21	"09:37:16"	"300385"	"普华永道"	"买入"	6.5293	11500.00	75500.7	12.01	0.0	0.77	0.00	-75099.01	0.00016	0.000001	0.000001	6.53	6.59	6.45	6.52	6.5	0.02	0.3077	11117.6	7206.686	0.010327
152	"沪"	2023-03-01	"09:31:46"	"688069"	"德林药业"	"买入"	28.2828	260.00	73583.8	11.77	73.58	0.07	0.00	73451.88	0.00016	0.001001	0.000001	28.52	28.86	28.27	28.67	28.54	0.13	0.4555	2416.06	6987.976	0.010761
27	"沪"	2022-10-11	"09:37:35"	"600817"	"杭州热电"	"买入"	3.11	41660.00	138857.0	22.22	0.0	1.33	0.00	-138880.55	0.00016	0.0000001	0.0000001	3.1	3.14	3.06	3.13	3.11	0.02	0.6431	61419.3	12872.499	0.010765

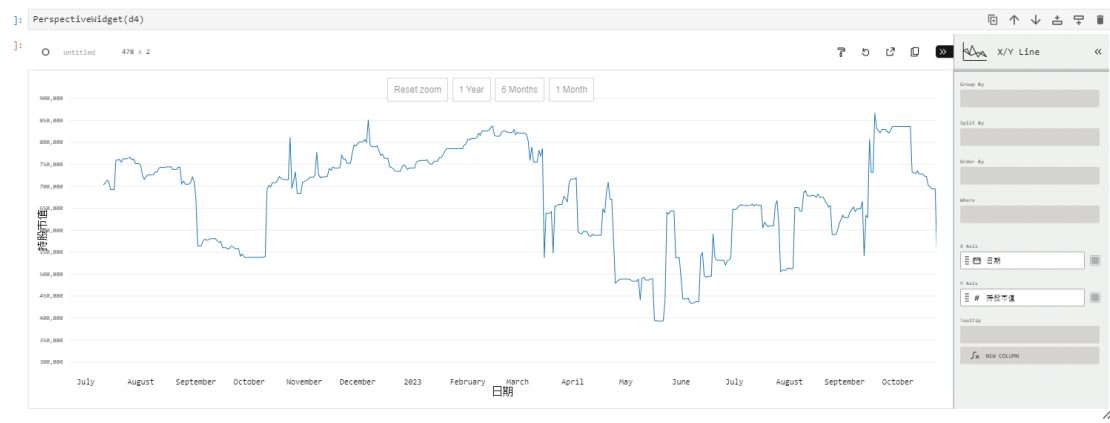
```

•[110]: d3.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(
    close2=pl.col("close").fill_null(strategy="forward").over("证券代码")
)[:10, ["日期", "证券代码", "结余数量", "close", "close2"]]

```

[110]: shape: (10, 5)

日期	证券代码	结余数量	close	close2
date	str	f64	f64	f64
2022-07-11	"000096"	0.0	9.29	9.29
2022-07-12	"000096"	0.0	9.2	9.2
2022-07-13	"000096"	0.0	9.23	9.23
2022-07-14	"000096"	0.0	9.14	9.14
2022-07-15	"000096"	0.0	8.94	8.94
2022-07-16	"000096"	0.0	null	8.94
2022-07-17	"000096"	0.0	null	8.94
2022-07-18	"000096"	0.0	9.16	9.16
2022-07-19	"000096"	0.0	9.21	9.21
2022-07-20	"000096"	0.0	9.25	9.25



```
] : pro.index_daily(ts_code="000300.SH",start_date=format(start_date,"%Y%m%d"),end_date=format(end_date,"%Y%m%d"))
```

	ts_code	trade_date	close	open	high	low	pre_close	change	pct_chg	vol	amount
0	000300.SH	20231031	3572.5056	3581.6759	3581.6759	3554.5822	3583.7718	-11.2662	-0.3144	112587267.0	1.963177e+08
1	000300.SH	20231030	3583.7718	3554.1055	3592.7188	3554.0238	3562.3851	21.3867	0.6003	144010426.0	2.429733e+08
2	000300.SH	20231027	3562.3851	3501.1710	3576.5861	3494.5488	3514.1448	48.2403	1.3727	120225108.0	2.243020e+08
3	000300.SH	20231026	3514.1448	3484.7807	3515.5716	3478.0503	3504.4576	9.6872	0.2764	105143566.0	1.760885e+08
4	000300.SH	20231025	3504.4576	3524.2965	3531.8314	3500.9154	3487.1302	17.3274	0.4969	128009617.0	2.036917e+08
...
313	000300.SH	20220715	4248.5273	4299.4714	4344.8572	4248.5273	4322.0742	-73.5469	-1.7017	150214453.0	3.167932e+08
314	000300.SH	20220714	4322.0742	4308.8229	4352.0978	4292.7595	4321.4617	0.6125	0.0142	145250926.0	3.015699e+08
315	000300.SH	20220713	4321.4617	4313.3589	4341.4663	4284.9419	4313.6184	7.8433	0.1818	131113179.0	2.860755e+08
316	000300.SH	20220712	4313.6184	4352.6456	4370.3411	4296.6215	4354.6167	-40.9983	-0.9415	130187838.0	2.910723e+08
317	000300.SH	20220711	4354.6167	4406.1476	4406.1476	4330.5819	4428.7810	-74.1643	-1.6746	134448578.0	2.991952e+08

318 rows × 11 columns

```
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",  
)
```

	ts_code	trade_date	pct_chg
0	000300.SH	20231031	-0.3144
1	000300.SH	20231030	0.6003
2	000300.SH	20231027	1.3727
3	000300.SH	20231026	0.2764
4	000300.SH	20231025	0.4969
...
313	000300.SH	20220715	-1.7017
314	000300.SH	20220714	0.0142
315	000300.SH	20220713	0.1818
316	000300.SH	20220712	-0.9415
317	000300.SH	20220711	-1.6746

318 rows × 3 columns

