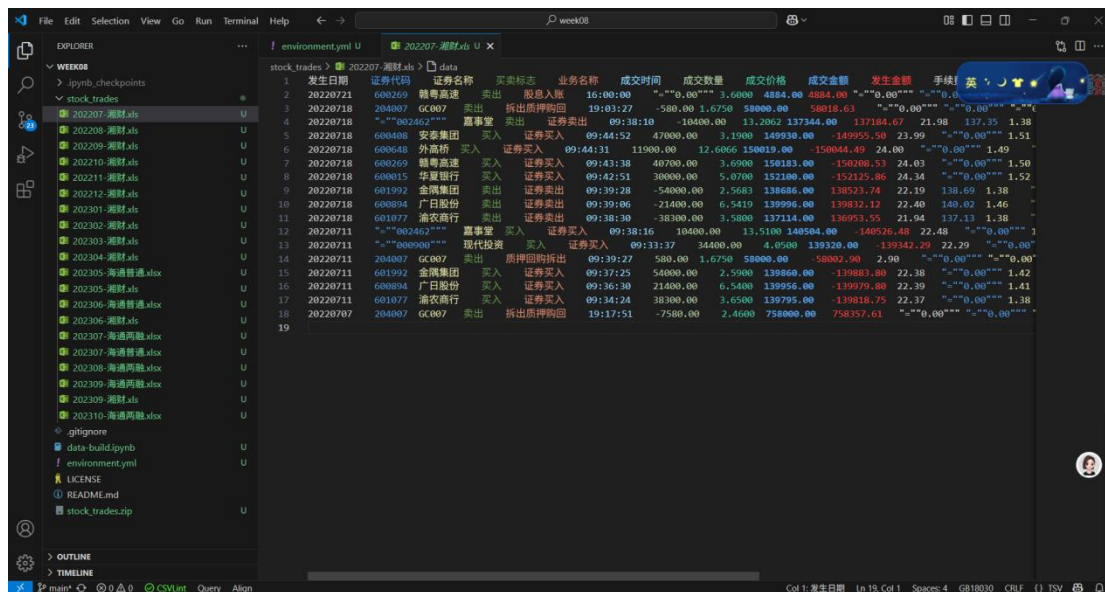


## 1. 解压文件

```
MINGW64/c/Users/13212/rep x + v
Dload Upload Total Spent Left Speed
100 77002 0 77002 0 0 257k 0 ---:--:-- ---:--:-- ---:--:-- 261k
(week08)
13212@LAPTOP-76SH9KP3 MINGW64 ~/repo/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
  inflating: stock_trades/202302-湘财.xls
  inflating: stock_trades/202303-湘财.xls
  inflating: stock_trades/202304-湘财.xls
  inflating: stock_trades/202305-海通普通.xlsx
  inflating: stock_trades/202305-湘财.xls
  inflating: stock_trades/202306-海通普通.xlsx
  inflating: stock_trades/202306-湘财.xls
  inflating: stock_trades/202307-海通两融.xlsx
  inflating: stock_trades/202307-海通普通.xlsx
  inflating: stock_trades/202308-海通两融.xlsx
  inflating: stock_trades/202309-海通两融.xlsx
  inflating: stock_trades/202309-湘财.xls
  inflating: stock_trades/202310-海通两融.xlsx
(week08)
13212@LAPTOP-76SH9KP3 MINGW64 ~/repo/week08 (main)
$
```

## 2. 成功变成中文



The screenshot shows a file explorer window with a directory structure. The files are named in Chinese, indicating successful localization. The files are organized into a hierarchy, with the main directory being 'stock\_trades'. The files are listed in a table format, showing the file name, its type, and its location.

File Name	Type	Location
202207-湘财.xls	U	stock_trades
202208-湘财.xls	U	stock_trades
202209-湘财.xls	U	stock_trades
202210-湘财.xls	U	stock_trades
202211-湘财.xls	U	stock_trades
202212-湘财.xls	U	stock_trades
202301-湘财.xls	U	stock_trades
202302-湘财.xls	U	stock_trades
202303-湘财.xls	U	stock_trades
202304-湘财.xls	U	stock_trades
202305-海通普通.xlsx	U	stock_trades
202305-湘财.xls	U	stock_trades
202306-海通普通.xlsx	U	stock_trades
202306-湘财.xls	U	stock_trades
202307-海通两融.xlsx	U	stock_trades
202307-海通普通.xlsx	U	stock_trades
202308-海通两融.xlsx	U	stock_trades
202309-海通两融.xlsx	U	stock_trades
202309-湘财.xls	U	stock_trades
202310-海通两融.xlsx	U	stock_trades
gitignore	U	stock_trades
data-build.py	U	stock_trades
environment.yml	U	stock_trades
LICENSE	U	stock_trades
README.md	U	stock_trades
stock_trades.zip	U	stock_trades

## 3. 数据清洗完毕

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

[13]: 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*
				*证											

## 处理海通的相关数据

```
data-build.ipynb
p1.col("成交日期").str.to_date("%Y%m%d"),
p1.col("成交时间").replace("=",None).str.to_time("%H:%M:%S"),
)
[114]: 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	*指*	0.0	0	0.0	0	*指定撤销:撤销指定*

data-build.ipynb

```
[118]: dfapl.read_excel(
        "stock_trades/202305-沪深普通.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"])
).filter(
    (~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("%H:%M:%S"),
)
```

[118]: 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*

## 准备合并生成新列

data-build.ipynb

```
[208]: d1.select(
        券商=pl.col("券商"),
        交易日期=pl.col("发生日期"),
        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("业务名称").replace({"证券卖出":"卖出","证券买入":"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"),
        印花税=pl.col("印花税"),
        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"),
    )
```

[208]: shape: (257, 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 至	date 日期	time 时间	str 至	str 至	str 至	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

成功拼出 363 行，14 列

data-build.ipynb

```
df = df1.concat([d1,d2,d3])
```

[229]: 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"	"震事堂"	"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

Simple Python 3 (ipykernel) | Idle Mode: Edit Ln 1, Col 4 data-build.ipynb

## 找出两者差额

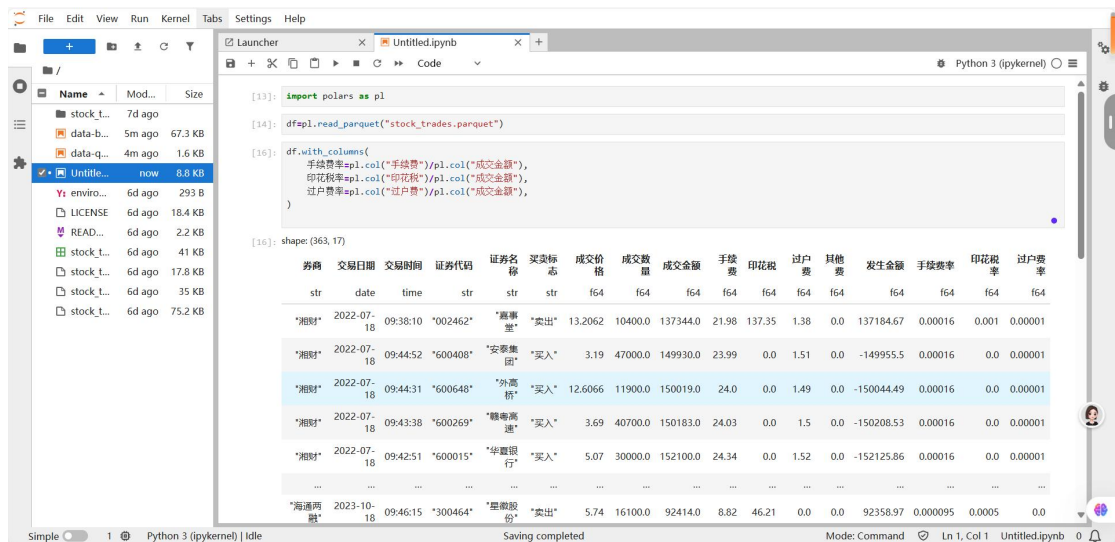
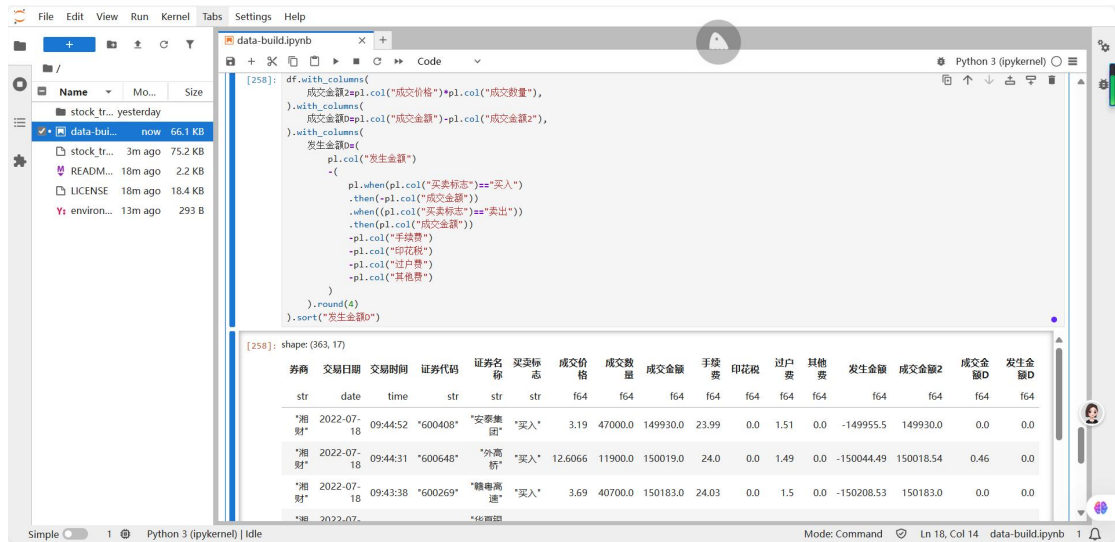
data-build.ipynb

```
), sort("成交金额")
```

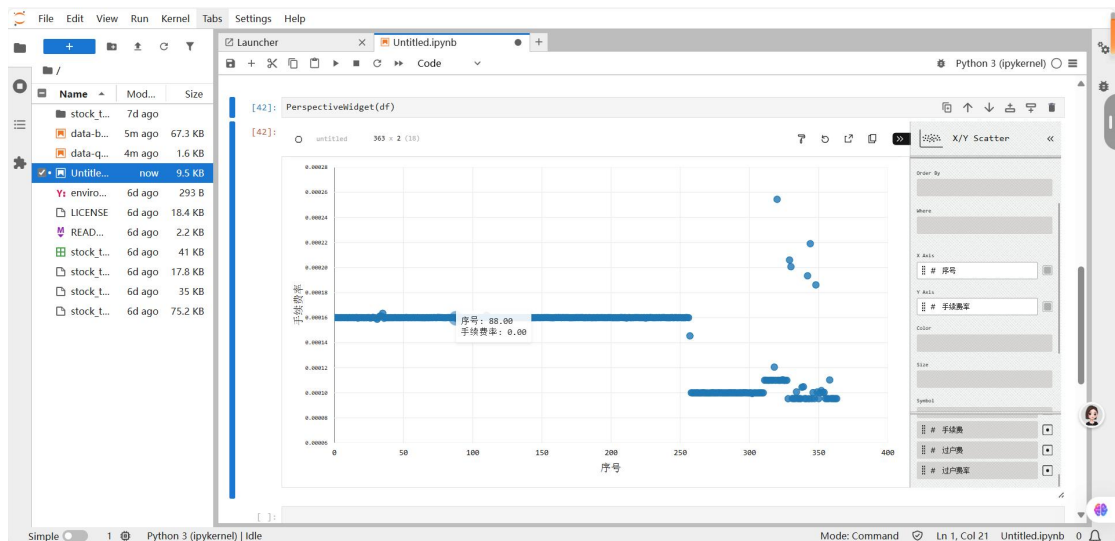
[233]: shape: (53, 16)

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	成交金额 2	成交金额 D
str	str	date	time	f64	f64	f64	f64	str	f64	f64	f64	f64	str	f64	f64
"300464"	"星徽股份"	2023-09-05	09:55:34	16100.0	6.176	99427.0	-99436.49	"买"	9.49	0.0	0.0	0.0	"星徽股份证券买入"	99433.6	-6.6
"300241"	"瑞丰光电"	2023-09-11	09:50:24	9700.0	5.116	49621.0	49591.18	"卖"	5.0	24.82	0.0	0.0	"瑞丰光电证券卖出"	49625.2	-4.2
"600336"	"澳柯玛"	2023-09-01	14:53:33	17400.0	6.362	110697.0	110629.98	"卖"	10.56	55.35	1.11	0.0	"澳柯玛证券卖出"	110698.8	-1.8
"002956"	"西麦食品"	2023-09-18	10:01:07	5200.0	14.34	74566.6	-74573.71	"买"	7.11	0.0	0.0	0.0	"西麦食品证券买入"	74568.0	-1.4
"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"	96461.4	-1.4
"600793"	"宜宾纸业"	2023-09-05	09:52:06	4000.0	11.966	47865.0	47835.58	"卖"	5.0	23.94	0.48	0.0	"宜宾纸业证券卖出"	47864.0	1.0
"600513"	"联环药业"	2023-09-05	09:48:44	5100.0	9.352	47697.0	47667.68	"卖"	5.0	23.84	0.48	0.0	"联环药业证券卖出"	47695.2	1.8
"002753"	"永东股份"	2023-10-31	09:31:06	14200.0	7.59	107780.0	107715.83	"卖"	10.28	53.89	0.0	0.0	"永东股份证券卖出"	107778.0	2.0
"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"	80070.0	2.0
"600638"	"新黄浦"	2023-07-27	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"	60691.8	4.2

Simple Python 3 (ipykernel) | Idle Mode: Command Ln 5, Col 16 data-build.ipynb

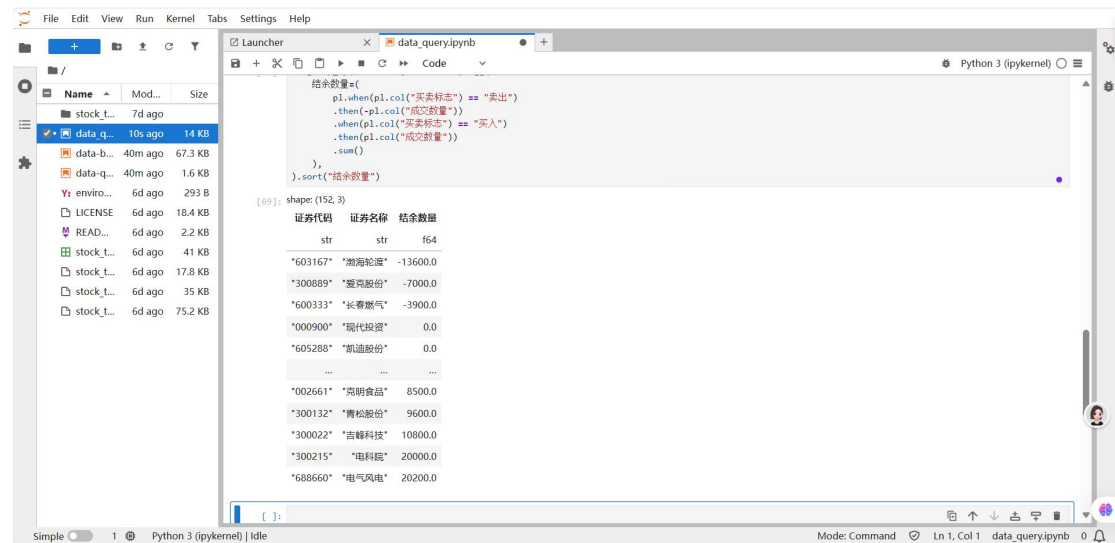


计算出三项税率，包括散点图

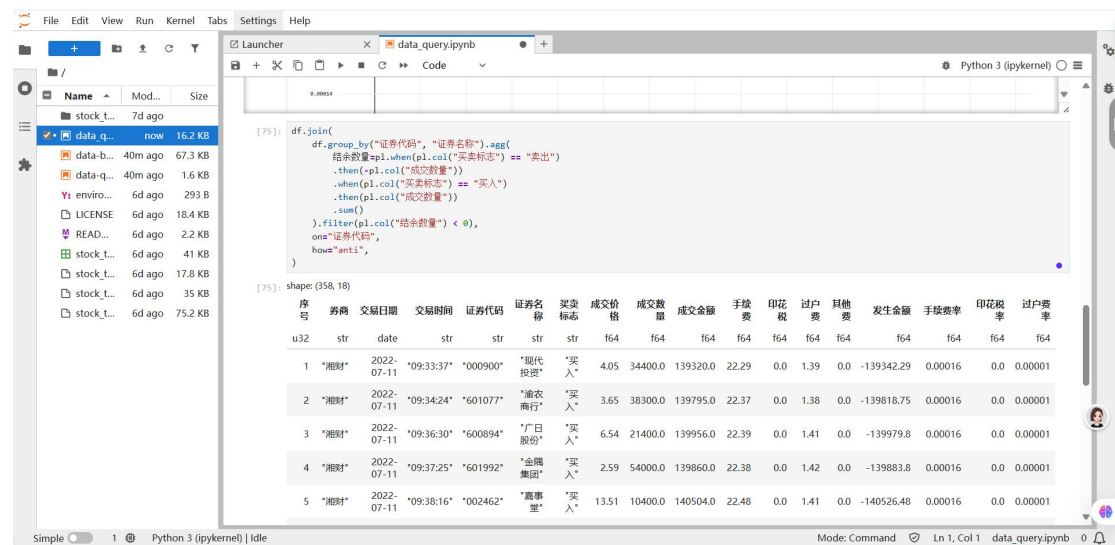




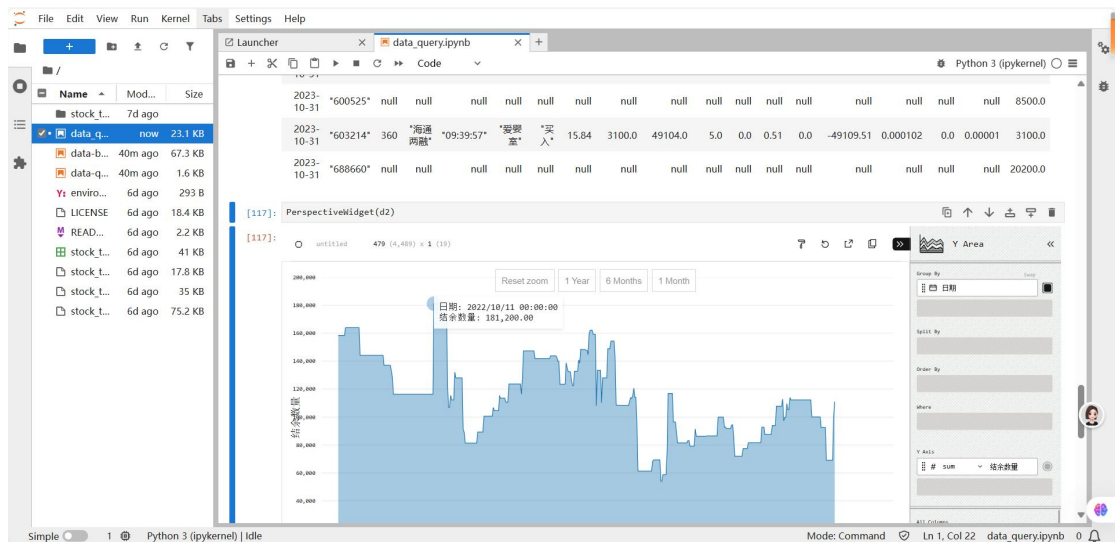
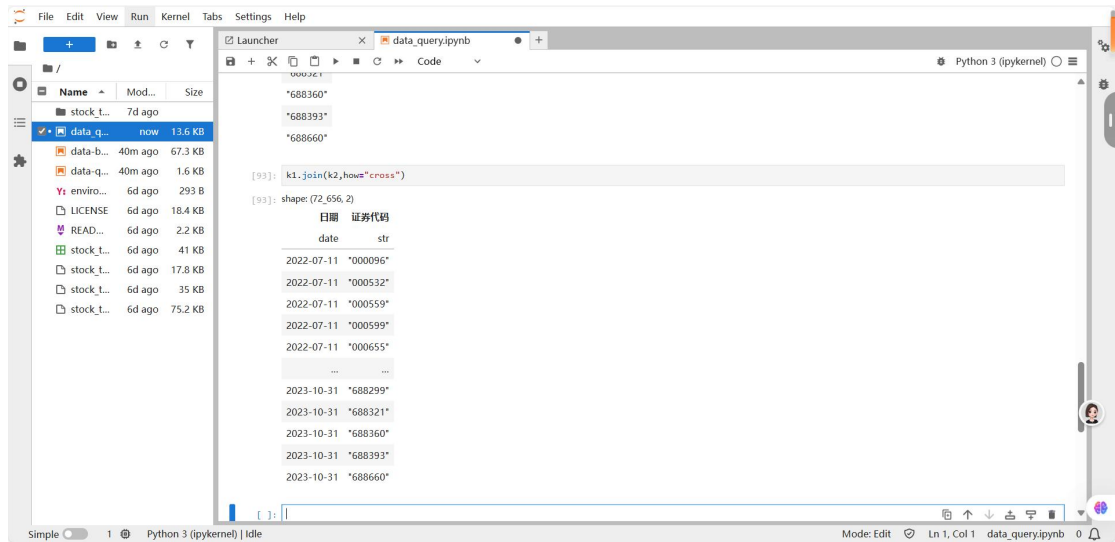
计算出结余数量



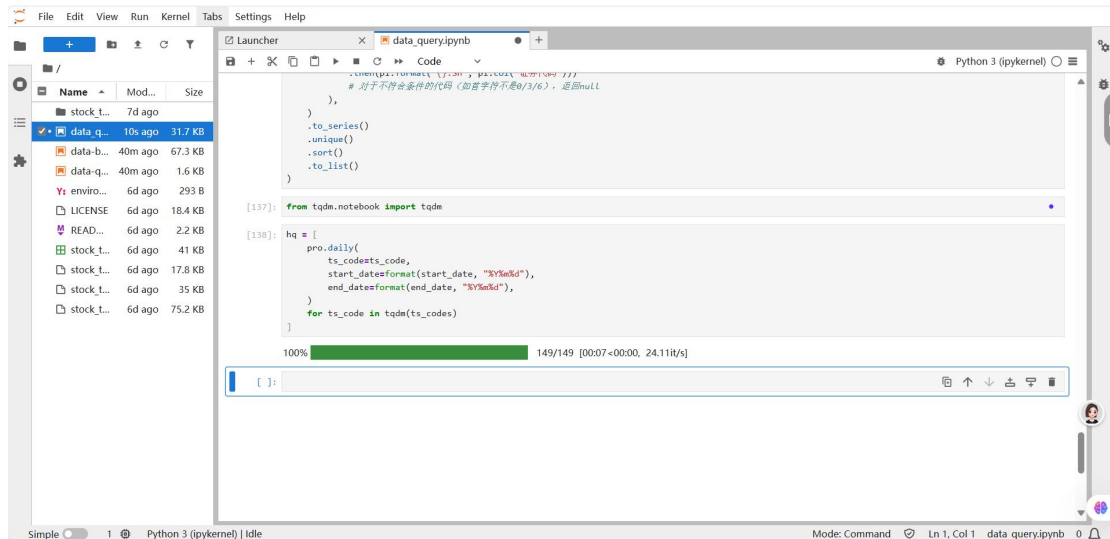
提出负值



## 笛卡尔计算出乘积



用 Perspective 的 Y Area 视图，横轴 Group By 设定为 交易日期，纵轴 Y Axis 可以查看每日总的结余数量



## 生成滚动条

将行情数据与每日持股数据匹配，将非交易日缺失的价格数据填充为最近数值 (`fill_null().over()`)，按 交易日期 分组，汇总计算每日总的 持股市值，用 **Perspective** 的 X/Y Line 视图观察每日 持股市值 的动态变化

