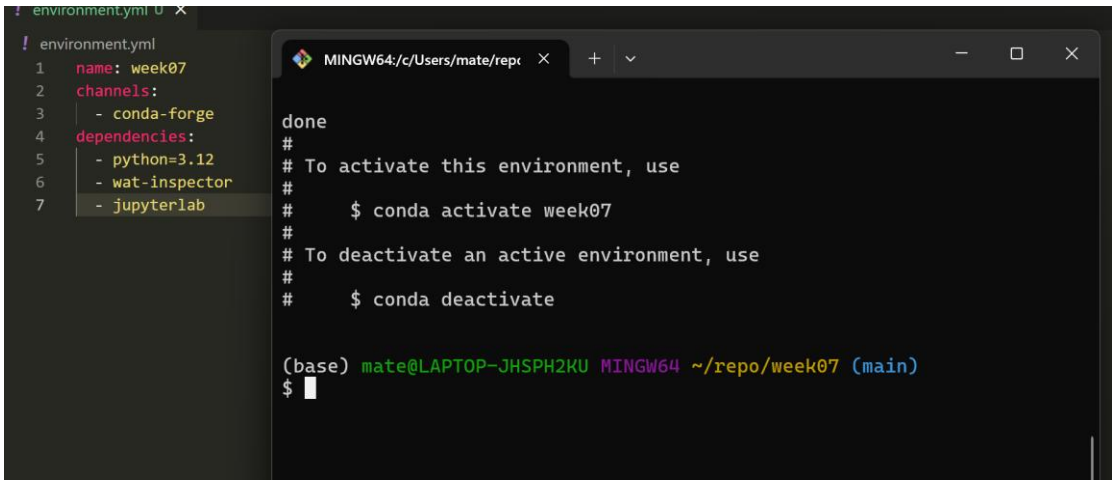
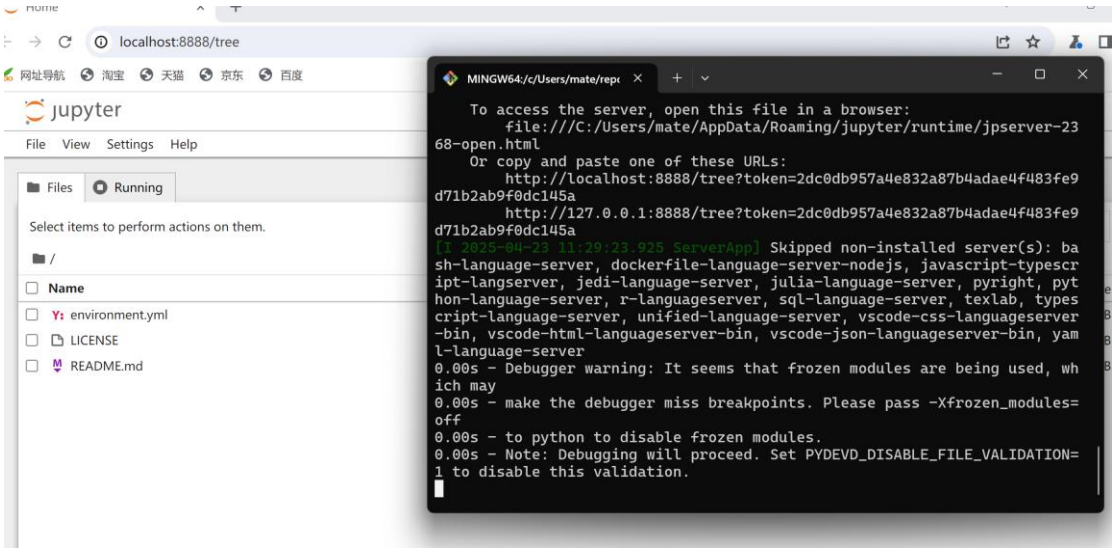
第七周学习报告-数据可视化与交互



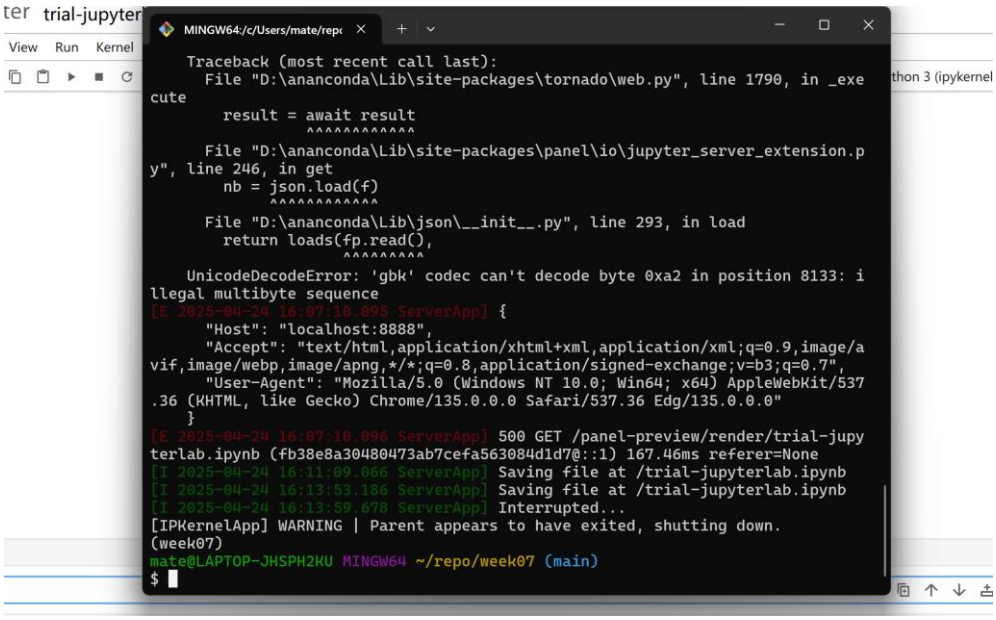


浏览器前端 终端后端

```
[ ]: 1
```

```
[15]: 1 import time
      2
      3 while True:
      4     print('hello')
      5     time.sleep(3)
```

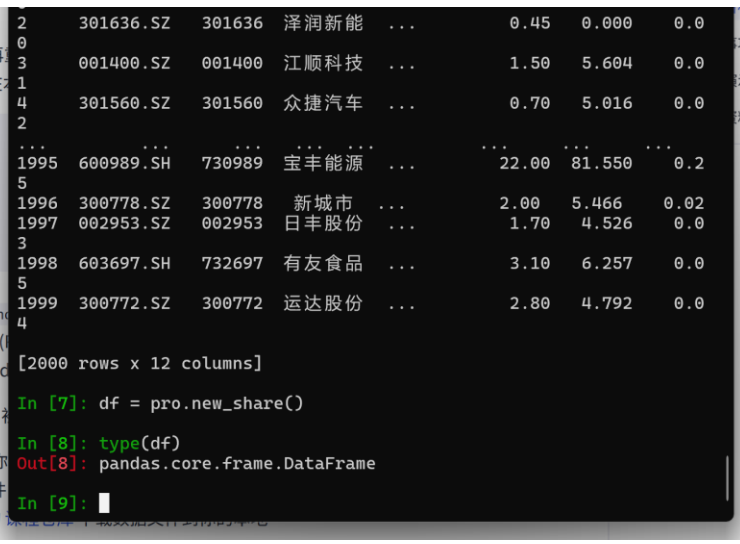
```
hello
hello
hello
hello
hello
hello
hello
hello
hello
hello
---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
Cell In[15], line 5
      3 while True:
      4     print('hello')
----> 5     time.sleep(3)

KeyboardInterrupt:
```



复设置。

按 Ctrl+D 结束前面的 IPython 进程，再
服务器请求 IPO新股列表 数据，并保存在

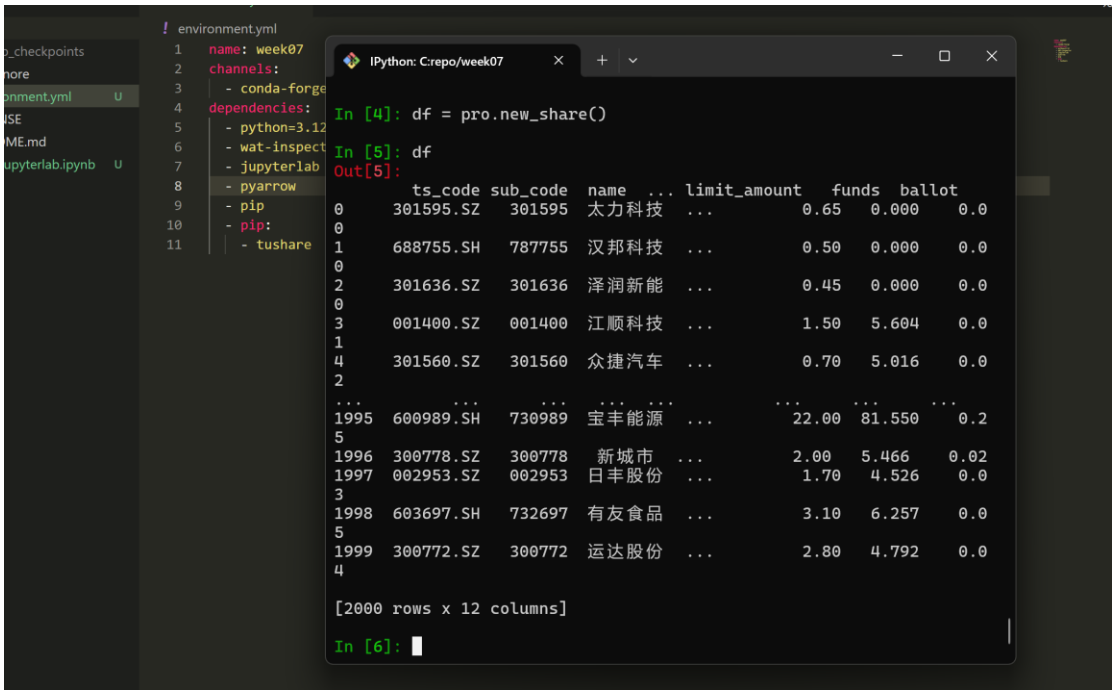
```
import tushare as ts

pro = ts.pro_api()
df = pro.new_share()
df.to_parquet("new_share.parquet")
```

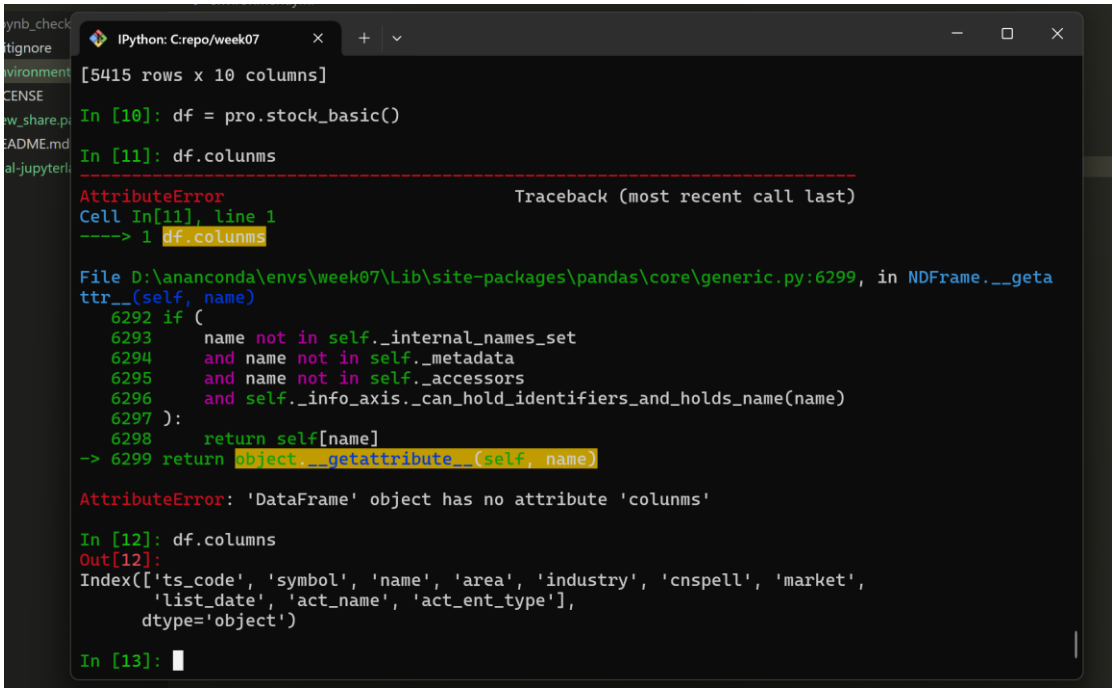


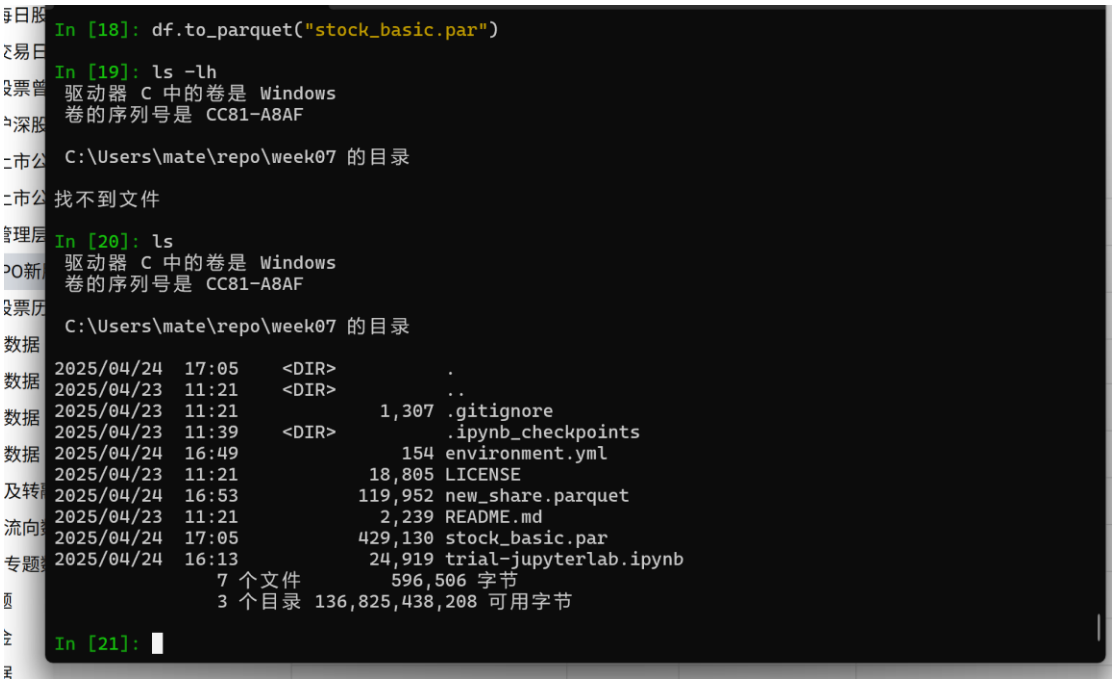
其中请求数据函数返回的对象 df 是 pan
中的 DataFrame 数据按照 Parquet 格式 (
(serialize) 为字节串 (bytes) 保存到磁盘 (d

询问豆包 (或 DeepSeek 等任何大模型)，

new_share 接口只需要 120 积分，如果你
将数据保存为 stock_basic.parquet 文件
以在终端运行下面的命令，从我们开源的







1. **Parquet 格式特点**

列式存储

Parquet 是一种列式存储文件格式。在存储数据时，它会将同一列的数据存储在一起。例如，在一个包含用户信息（如用户 ID、姓名、年龄、地址等字段）的表格中，Parquet 会把所有用户 ID 存放在一个区域，姓名存放在另一个区域，依此类推。

这种存储方式的优势在于，当进行查询操作时，如果只需要查询某一列或某几列的数据，例如统计所有用户的年龄分布情况，系统可以直接读取存储年龄的那一列数据，而无需读取整个行的数据。这大大减少了数据读取量，提高了查询效率，尤其是在处理大数据场景下，数据量庞大时，这种优势更为明显。

* **高效的压缩和编码**

* Parquet 支持多种压缩和编码算法，如 Snappy、Gzip、LZO 等。通过压缩，可以有效减少数据的存储空间。例如，对于文本类型的列数据，使用合适的压缩算法可以使其存储空间大幅减少。

* 同时，编码算法（如字典编码、游程编码等）可以进一步优化数据存储。字典编码是将列中重复出现的值用一个较短的代码来代替，从而节省存储空间。例如，一个列中有大量重复的字符串值“apple”，在字典编码后，可以用一个代码（如 1）来代替所有“apple”，在解码时再将其还原。

* **数据类型丰富**

* 它支持多种数据类型，包括基本数据类型（如整数、浮点数、字符串等）、复杂数据类型（如嵌套数据结构、列表、映射等）。这使得 Parquet 格式能够很好地存储和表达各种复杂的数据模型。

* **与大数据生态系统高度集成**

* Parquet 是为大数据处理而设计的，与 Hadoop、Spark、Presto 等大数据处理工具和框架无缝集成。在 Spark 中，可以很方便地将数据读取为 Parquet 格式的数据集，并进行各种分布式计算操作。例如，在 Spark SQL 中，可以通过简单的 SQL 语句查询 Parquet 格式的数据，而无需进行繁琐的数据转换。

2. **CSV 格式特点**

* **简单的文本格式**

* CSV（Comma - Separated Values）格式是一种简单的文本文件格式。其文件内容是以行为单位，每行表示一条记录，行内的各个字段之间用逗号（,）分隔。例如，“1,John,25,New York”表示一条包含用户 ID、姓名、年龄和地址四个字段的记录。

* **易于阅读和编辑**

* 由于是纯文本格式，不需要特殊的软件就可以阅读和编辑。在简单的文本编辑器（如 Notepad、Notepad++ 等）或者电子表格软件（如 Excel）中都可以方便地打开和修改 CSV 文件。这使得它在数据交换和简单数据处理场景中非常方便。

* **存储效率低**

* 相比 Parquet 的列式存储，CSV 是行式存储。在存储大量数据时，尤其是包含很多列的数据时，存储效率较低。而且它没有像 Parquet 那样高效的压缩机制，存储空间占用较大。

* **数据类型支持有限**

* CSV 本身不直接存储数据类型信息。在读取 CSV 文件时，需要根据实际数据内容来推断数据类型，这可能会导致一些数据类型转换的错误或者不确定性。例如，一个字段可能既包含数字字符串（如“123”）又包含文本字符串（如“abc”），在读取时很难准确判断其数据类型。

3. **Parquet 格式适用领域**

* **大数据分析和处理**

* 在处理海量数据时，如数据仓库中的数据存储和查询，Parquet 的列式存储和高效压缩编码能够大大提升性能。例如，在电信行业，对用户的通话记录、流量使用记录等海量数据进行分析时，Parquet 格式可以快速地过滤和查询特定列的数据，如统计某个地区用户的流量使用高峰时段。

* **数据湖存储**

* 数据湖是一种存储大量原始数据的存储库，Parquet 格式是数据湖中常用的数据存储格式之一。它可以存储不同类型的数据源（如日志数据、传感器数据等）生成的结构化和半结构化数据，并且方便后续的数据分析和挖掘。例如，在物联网场景下，存储来自各种传感器的大量时间序列数据，Parquet 格式可以有效地组织和存储这些数据，便于后续的时序分析。

* **分布式计算框架中的数据存储**

* 在 Spark 等分布式计算框架中，Parquet 格式的数据可以被高效地分布式存储和处理。因为 Parquet 的文件格式支持数据的分片存储，能够很好地适应分布式文件系统（如 HDFS）的存储架构，使得数据可以在多个计算节点上并行处理。

4. **CSV 格式适用领域**

* **数据交换和共享**

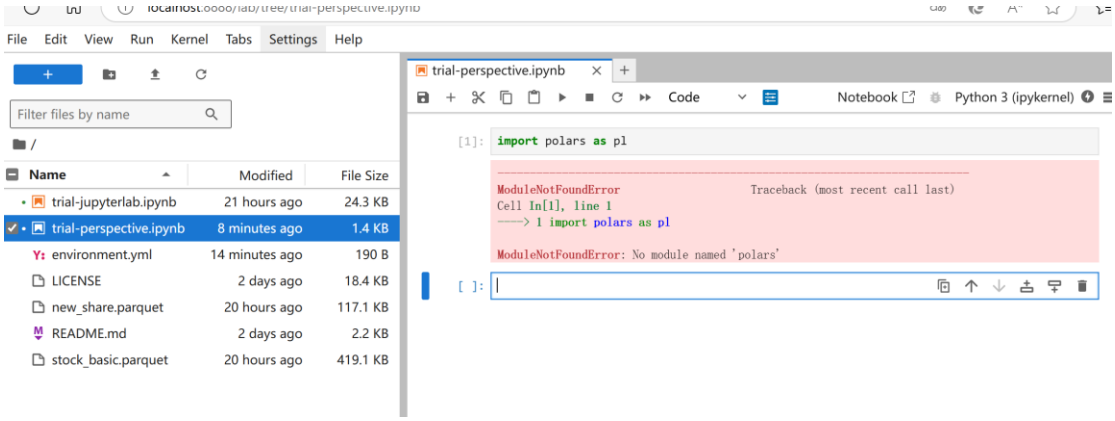
* CSV 格式常用于不同系统之间进行数据交换。例如，在一个企业的不同部门之间，或者企业与合作伙伴之间，当需要共享一些简单的数据（如客户名单、产品列表等）时，CSV 格式是一种简单易用的选择。因为大多数数据库系统和数据分析软件都支持导入和导出 CSV 文件。

* **简单的数据处理和分析**

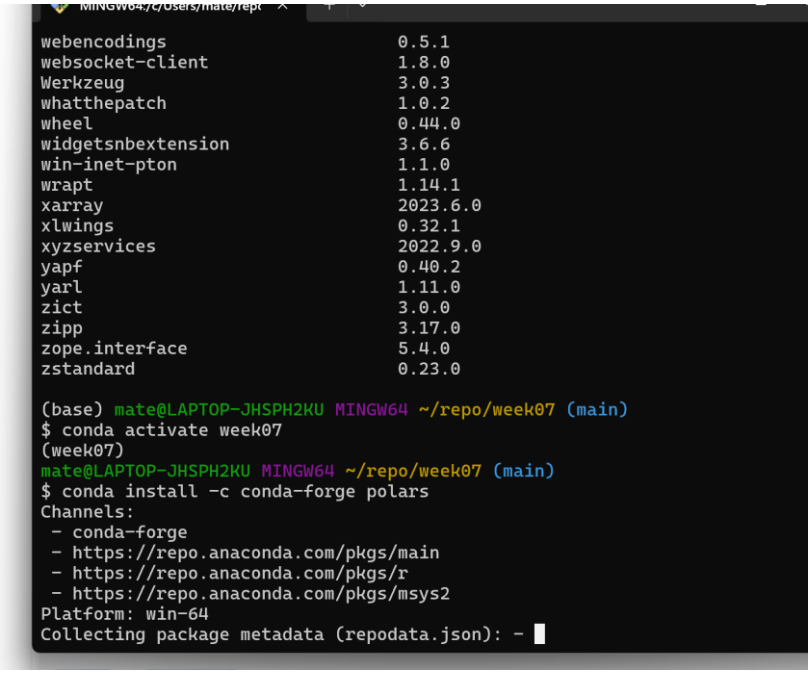
* 对于一些小型的、简单的数据处理任务，如在 Excel 中进行数据的排序、筛选、简单的统计分析等，CSV 格式可以直接在电子表格软件中打开和使用。例如，一个小商店可以使用 CSV 格式记录每天的销售数据，并在 Excel 中进行数据的初步分析，如计算每日销售额、统计销售量等。

* **数据导入导出的中间格式**

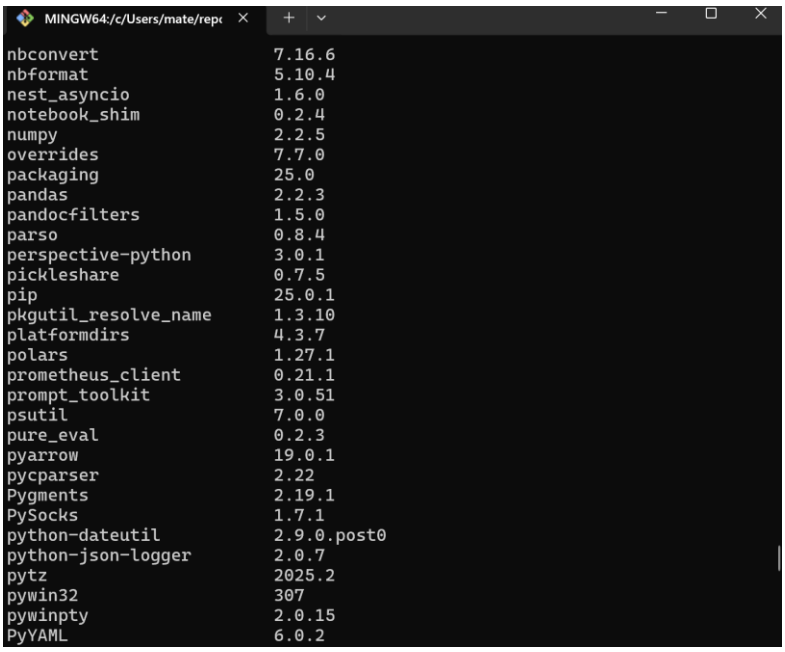
* 在数据从一个系统迁移到另一个系统，或者从一个软件工具导入到另一个软件工具时，CSV 格式常作为中间格式。例如，将一个旧的数据库系统中的数据导出为 CSV 文件，然后将其导入到新的数据库系统或者数据分析软件中。

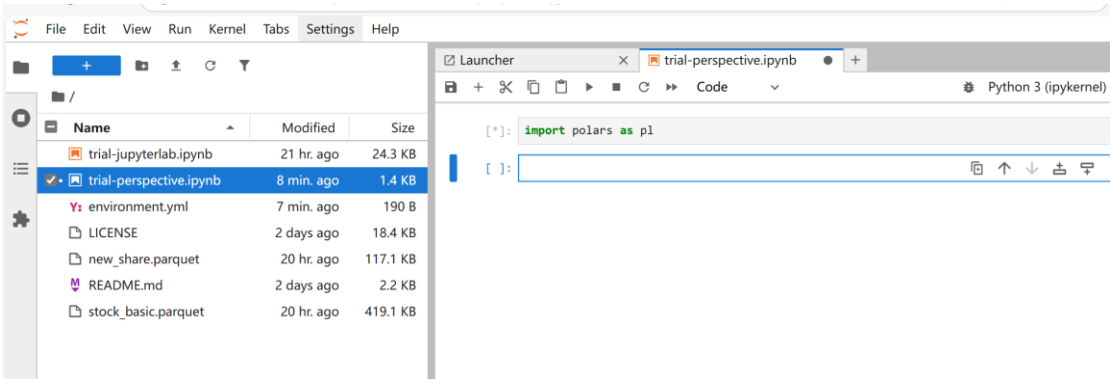


遇到问题，没有安装 polars，但是我已经在 environment.yml 里面输入了 polars

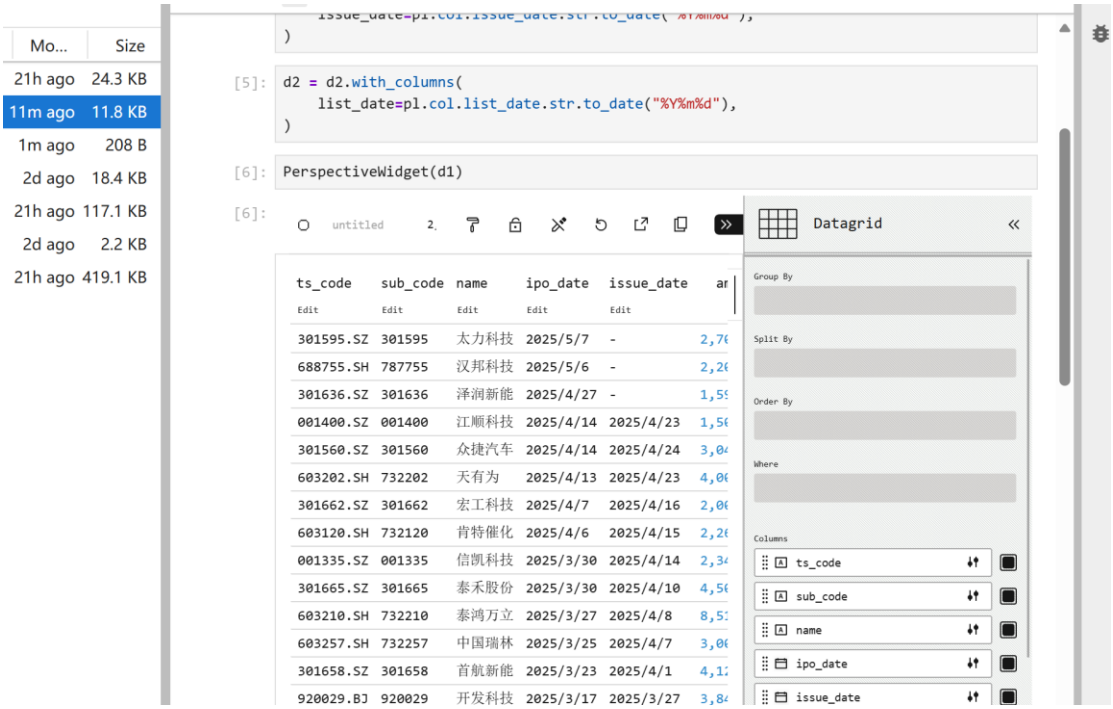


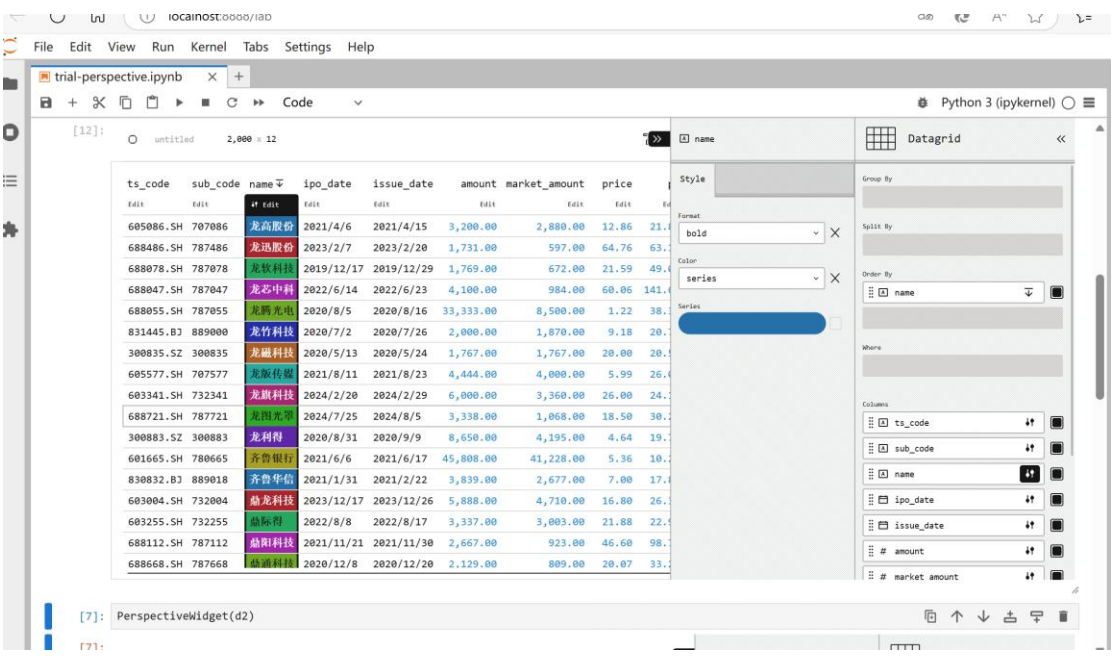
检查是否安装 polars 发现未安装 重新安装 polars

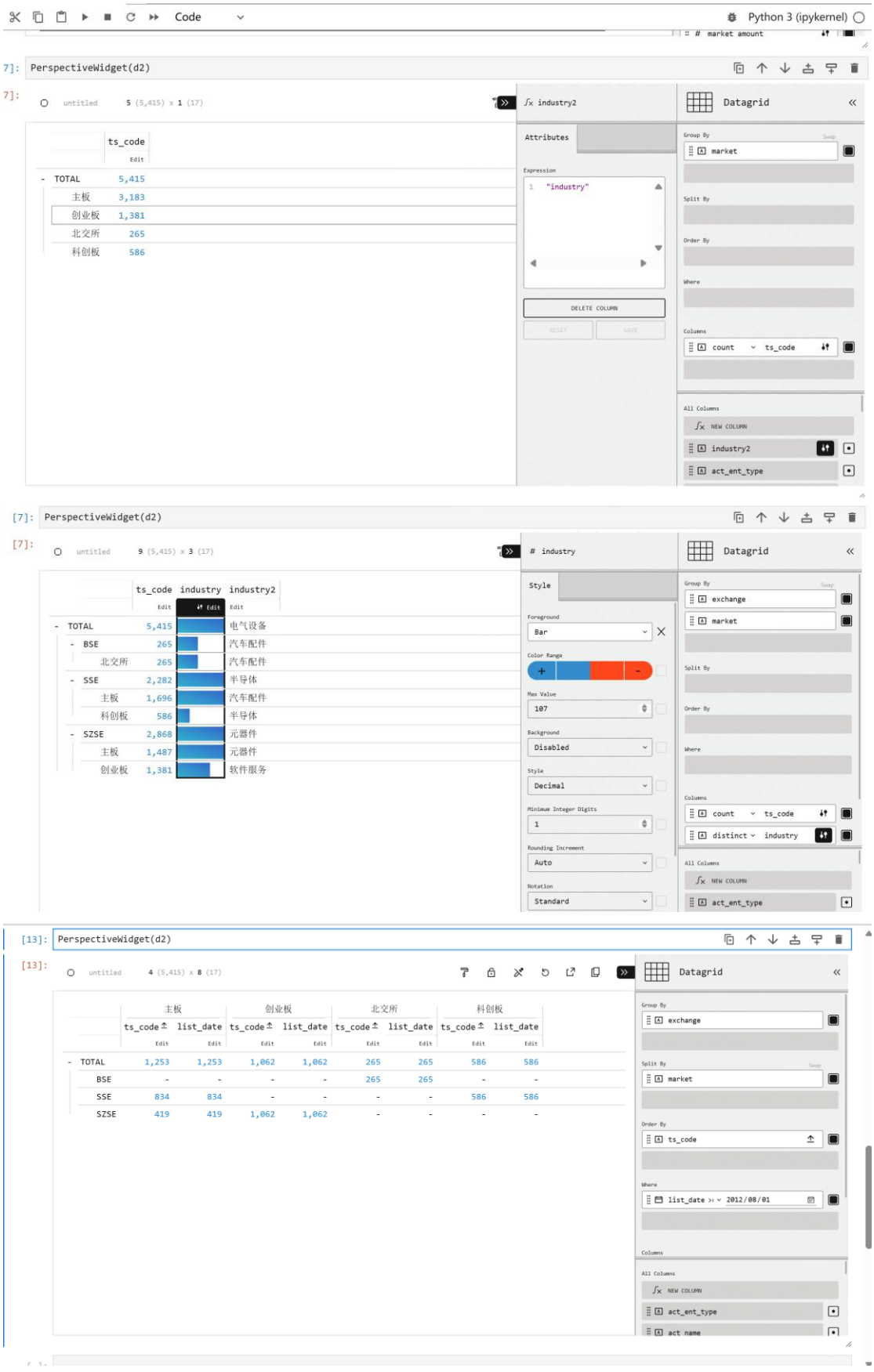


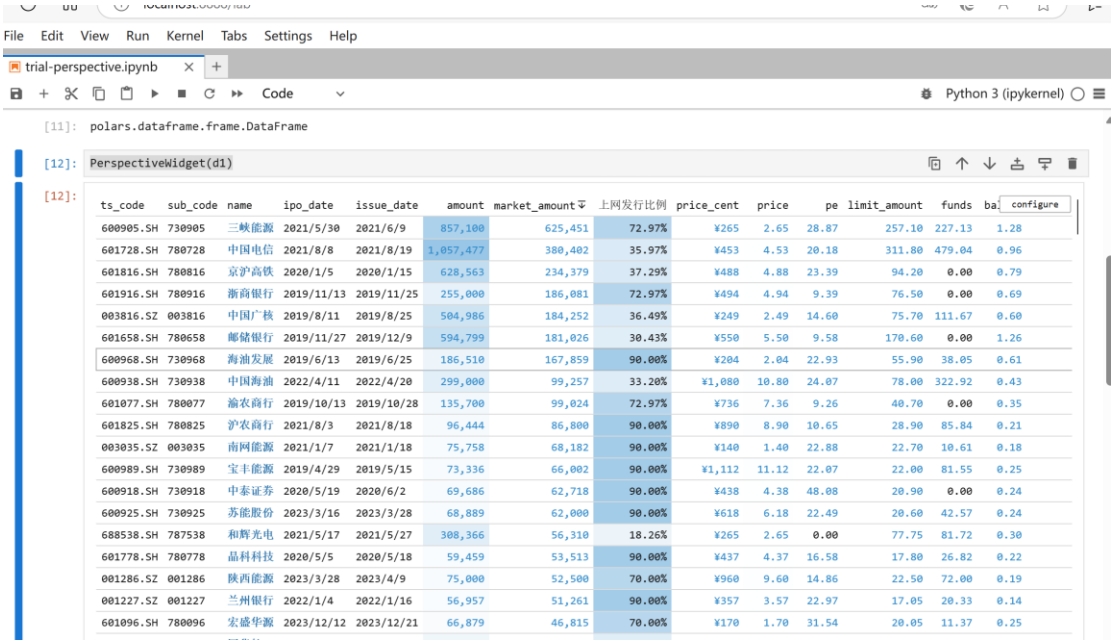

File Edit View Run Kernel Tabs Settings Help
Launcher
trial-perspective.ipynb
Code
Python 3 (ipykernel)
Name
Modified
Size
trial-jupyterlab.ipynb 21 hr. ago 24.3 KB
trial-perspective.ipynb 8 min. ago 1.4 KB
environment.yml 7 min. ago 190 B
LICENSE 2 days ago 18.4 KB
new_share.parquet 20 hr. ago 117.1 KB
README.md 2 days ago 2.2 KB
stock_basic.parquet 20 hr. ago 419.1 KB
[*]: import polars as pl
[ ]:

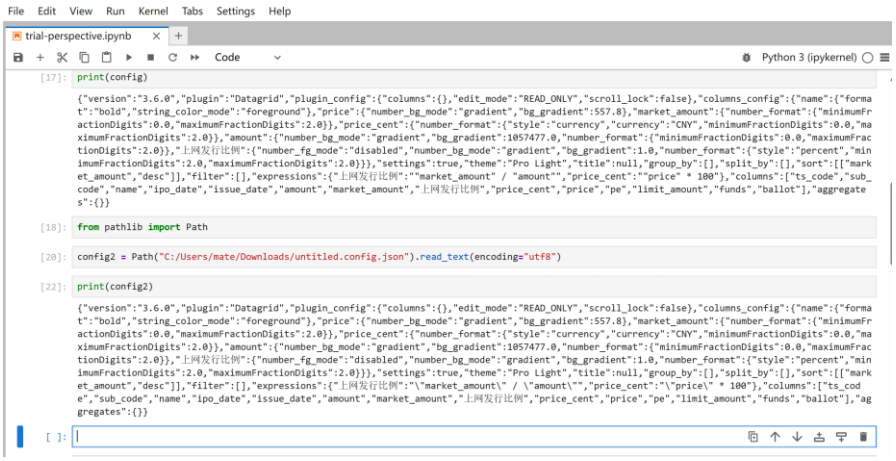
成功


Mo... Size
21h ago 24.3 KB
11m ago 11.8 KB
1m ago 208 B
2d ago 18.4 KB
21h ago 117.1 KB
2d ago 2.2 KB
21h ago 419.1 KB
[5]: d2 = d2.with_columns(
list_date=pl.col.list_date.str.to_date("%Y%m%d"),
)
[6]: PerspectiveWidget(d1)
[6]: untitled
Datagrid
ts_code sub_code name ipo_date issue_date
301595.SZ 301595 太力科技 2025/5/7 -
688755.SH 787755 汉邦科技 2025/5/6 -
301636.SZ 301636 泽润新能 2025/4/27 -
001400.SZ 001400 江顺科技 2025/4/14 2025/4/23
301560.SZ 301560 众捷汽车 2025/4/14 2025/4/24
603202.SH 732202 天有为 2025/4/13 2025/4/23
301662.SZ 301662 宏工科技 2025/4/7 2025/4/16
603120.SH 732120 肯特催化 2025/4/6 2025/4/15
001335.SZ 001335 信凯科技 2025/3/30 2025/4/14
301665.SZ 301665 泰禾股份 2025/3/30 2025/4/10
603210.SH 732210 泰鸿万立 2025/3/27 2025/4/8
603257.SH 732257 中国瑞林 2025/3/25 2025/4/7
301658.SZ 301658 首航新能 2025/3/23 2025/4/1
920029.BJ 920029 开发科技 2025/3/17 2025/3/27
Group By
Split By
Order By
Where
Columns
ts_code
sub_code
name
ipo_date
issue_date

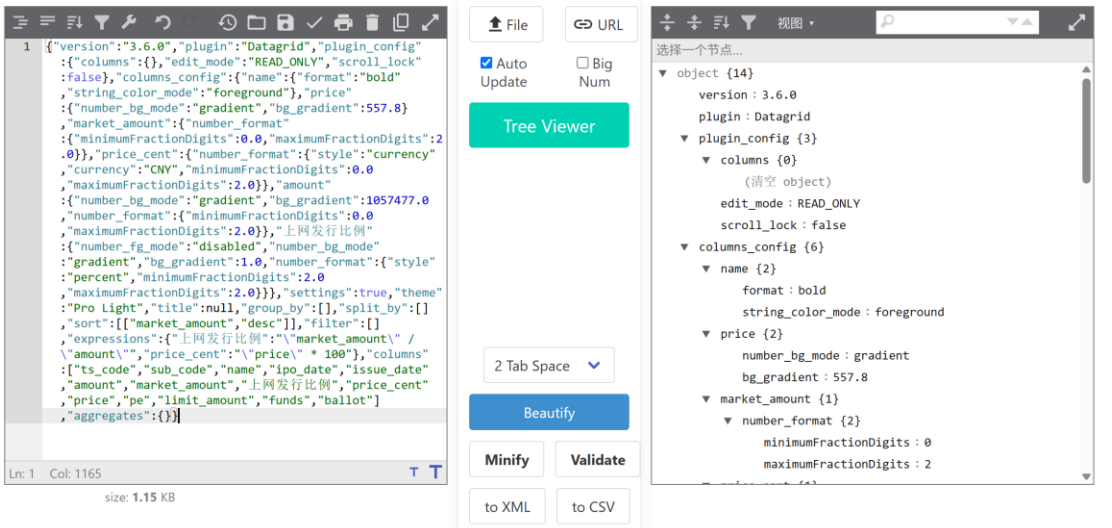

File Edit View Run Kernel Tabs Settings Help
trial-perspective.ipynb
Code
Python 3 (ipykernel)
[12]: untitled 2,000 × 12
Datagrid
ts_code sub_code name ipo_date issue_date amount market_amount price
605086.SH 707086 龙高股份 2021/4/6 2021/4/15 3,200.00 2,880.00 12.86
688486.SH 787486 龙迅股份 2023/2/7 2023/2/20 1,731.00 597.00 64.76
688078.SH 787078 龙软科技 2019/12/17 2019/12/29 1,769.00 672.00 21.59
688047.SH 787047 龙芯中科 2022/6/14 2022/6/23 4,100.00 984.00 60.06
688055.SH 787055 龙腾光电 2020/8/5 2020/8/16 33,333.00 8,500.00 1.22
831445.BJ 889000 龙竹科技 2020/7/2 2020/7/26 2,000.00 1,870.00 9.18
300835.SZ 300835 龙磁科技 2020/5/13 2020/5/24 1,767.00 1,767.00 20.00
605577.SH 707577 龙版传媒 2021/8/11 2021/8/23 4,444.00 4,000.00 5.99
603341.SH 732341 龙旗科技 2024/2/20 2024/2/29 6,000.00 3,360.00 26.00
688721.SH 787721 龙图光罩 2024/7/25 2024/8/5 3,338.00 1,068.00 18.50
300883.SZ 300883 龙利得 2020/8/31 2020/9/9 8,650.00 4,195.00 4.64
601665.SH 780665 齐鲁银行 2021/6/6 2021/6/17 45,808.00 41,228.00 5.36
830832.BJ 889018 齐鲁华信 2021/1/31 2021/2/22 3,839.00 2,677.00 7.00
603004.SH 732004 鼎龙科技 2023/12/17 2023/12/26 5,888.00 4,710.00 16.80
603255.SH 732255 鼎际得 2022/8/8 2022/8/17 3,337.00 3,003.00 21.88
688112.SH 787112 鼎阳科技 2021/11/21 2021/11/30 2,667.00 923.00 46.60
688668.SH 787668 鼎通科技 2020/12/8 2020/12/20 2,129.00 809.00 20.07
Style
Format
bold
Color
series
Series
Group By
Split By
Order By
name
Where
Columns
ts_code
sub_code
name
ipo_date
issue_date
amount
market amount
[7]: PerspectiveWidget(d2)

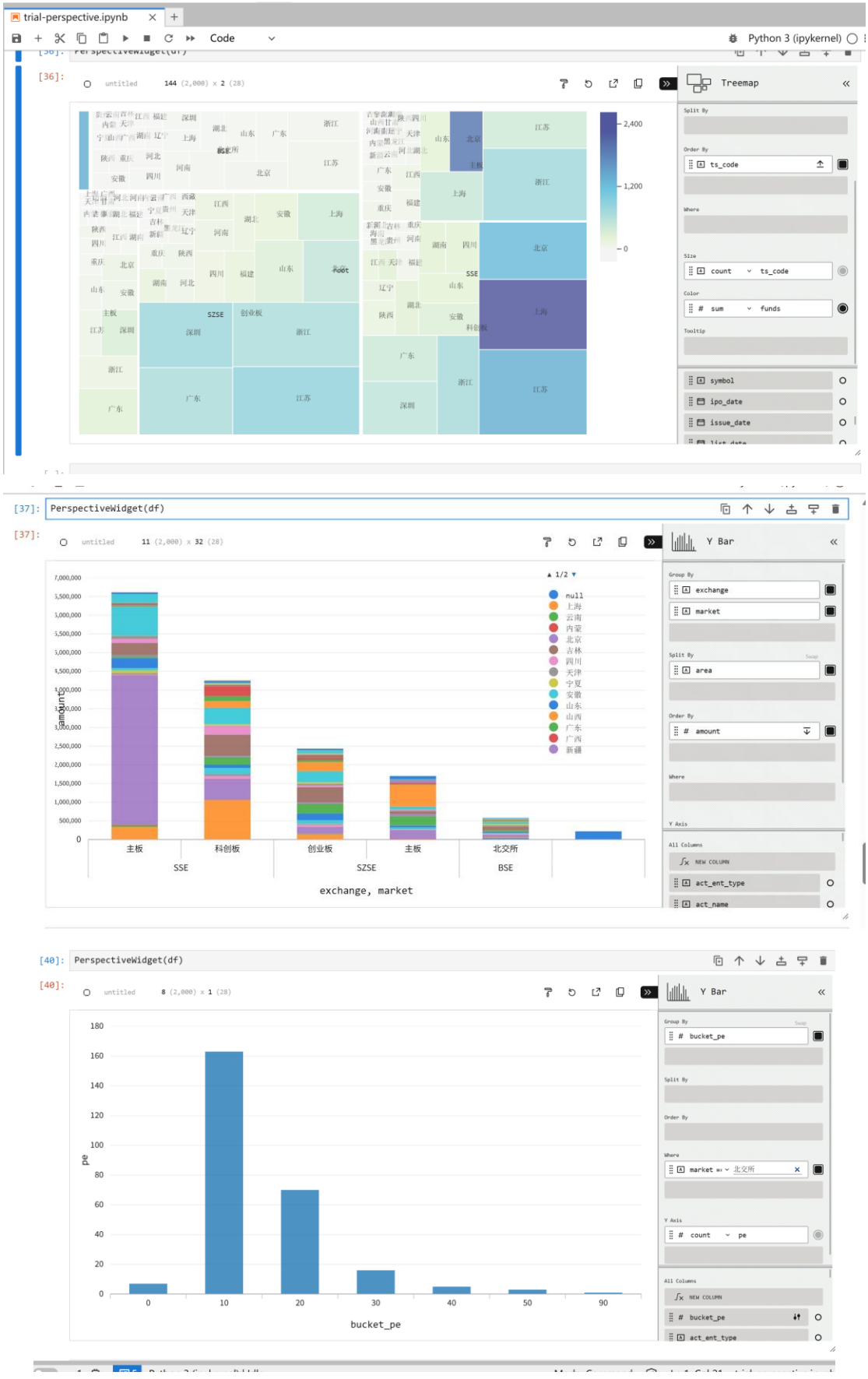

Code
Python 3 (ipykernel)
7]: PerspectiveWidget(d2)
7]:
untitled 5 (5,415) x 1 (17)
ts_code
TOTAL 5,415
主板 3,183
创业板 1,381
北交所 265
科创板 586
industry2
Attributes
Expression
"industry"
DELETE COLUMN
RESET
SAVE
Datagrid
Group By
market
Split By
Order By
Where
Columns
count ts_code
All Columns
NEW COLUMN
industry2
act_ent_type
[7]: PerspectiveWidget(d2)
[7]:
untitled 9 (5,415) x 3 (17)
ts_code industry industry2
TOTAL 5,415 电气设备
BSE 265 汽车配件
北交所 265 汽车配件
SSE 2,282 半导体
主板 1,696 汽车配件
科创板 586 半导体
SZSE 2,868 元器件
主板 1,487 元器件
创业板 1,381 软件服务
industry
Style
Foreground
Bar
Color Range
Max Value
107
Background
Disabled
Style
Decimal
Minimum Integer Digits
1
Rounding Increment
Auto
Notation
Standard
Datagrid
Group By
exchange
market
Split By
Order By
Where
Columns
count ts_code
distinct industry
All Columns
NEW COLUMN
act_ent_type
[13]: PerspectiveWidget(d2)
[13]:
untitled 4 (5,415) x 8 (17)
主板 创业板 北交所 科创板
ts_code list_date ts_code list_date ts_code list_date ts_code list_date
TOTAL 1,253 1,253 1,062 1,062 265 265 586 586
BSE - - - - 265 265 - -
SSE 834 834 - - - - 586 586
SZSE 419 419 1,062 1,062 - - - -
Datagrid
Group By
exchange
Split By
market
Order By
ts_code
Where
list_date >= 2012/08/01
Columns
All Columns
NEW COLUMN
act_ent_type
act_name


File Edit View Run Kernel Tabs Settings Help
trial-perspective.ipynb
Code
Python 3 (ipykernel)
[11]: polars.dataframe.frame.DataFrame
[12]: PerspectiveWidget(d1)
[12]:
configure
ts_code sub_code name ipo_date issue_date amount market_amount 上网发行比例 price_cent price pe limit_amount funds ba
600905.SH 730905 三峡能源 2021/5/30 2021/6/9 857,100 625,451 72.97% ¥265 2.65 28.87 257.10 227.13 1.28
601728.SH 780728 中国电信 2021/8/8 2021/8/19 1,057,477 380,402 35.97% ¥453 4.53 20.18 311.80 479.04 0.96
601816.SH 780816 京沪高铁 2020/1/5 2020/1/15 628,563 234,379 37.29% ¥488 4.88 23.39 94.20 0.00 0.79
601916.SH 780916 浙商银行 2019/11/13 2019/11/25 255,000 186,081 72.97% ¥494 4.94 9.39 76.50 0.00 0.69
003816.SZ 003816 中国广核 2019/8/11 2019/8/25 504,986 184,252 36.49% ¥249 2.49 14.60 75.70 111.67 0.60
601658.SH 780658 邮储银行 2019/11/27 2019/12/9 594,799 181,026 30.43% ¥550 5.50 9.58 170.60 0.00 1.26
600968.SH 730968 海油发展 2019/6/13 2019/6/25 186,510 167,859 90.00% ¥204 2.04 22.93 55.90 38.05 0.61
600938.SH 730938 中国海油 2022/4/11 2022/4/20 299,000 99,257 33.20% ¥1,080 10.80 24.07 78.00 322.92 0.43
601077.SH 780077 渝农商行 2019/10/13 2019/10/28 135,700 99,024 72.97% ¥736 7.36 9.26 40.70 0.00 0.35
601825.SH 780825 沪农商行 2021/8/3 2021/8/18 96,444 86,800 90.00% ¥890 8.90 10.65 28.90 85.84 0.21
003035.SZ 003035 南网能源 2021/1/7 2021/1/18 75,758 68,182 90.00% ¥140 1.40 22.88 22.70 10.61 0.18
600989.SH 730989 宝丰能源 2019/4/29 2019/5/15 73,336 66,002 90.00% ¥1,112 11.12 22.07 22.00 81.55 0.25
600918.SH 730918 中泰证券 2020/5/19 2020/6/2 69,686 62,718 90.00% ¥438 4.38 48.08 20.90 0.00 0.24
600925.SH 730925 苏能股份 2023/3/16 2023/3/28 68,889 62,000 90.00% ¥618 6.18 22.49 20.60 42.57 0.24
688538.SH 787538 和辉光电 2021/5/17 2021/5/27 308,366 56,310 18.26% ¥265 2.65 0.00 77.75 81.72 0.30
601778.SH 780778 晶科科技 2020/5/5 2020/5/18 59,459 53,513 90.00% ¥437 4.37 16.58 17.80 26.82 0.22
001286.SZ 001286 陕西能源 2023/3/28 2023/4/9 75,000 52,500 70.00% ¥960 9.60 14.86 22.50 72.00 0.19
001227.SZ 001227 兰州银行 2022/1/4 2022/1/16 56,957 51,261 90.00% ¥357 3.57 22.97 17.05 20.33 0.14
601096.SH 780096 宏盛华源 2023/12/12 2023/12/21 66,879 46,815 70.00% ¥170 1.70 31.54 20.05 11.37 0.25


File Edit View Run Kernel Tabs Settings Help
trial-perspective.ipynb
Code
Python 3 (ipykernel)
[17]: print(config)
{"version":"3.6.0","plugin":"Datagrid","plugin_config":{"columns":{},"edit_mode":"READ_ONLY","scroll_lock":false},"columns_config":{"name":{"format":"bold","string_color_mode":"foreground"},"price":{"number_bg_mode":"gradient","bg_gradient":557.8},"market_amount":{"number_format":{"minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"price_cent":{"number_format":{"style":"currency","currency":"CNY","minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"amount":{"number_bg_mode":"gradient","bg_gradient":1057477.0,"number_format":{"minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"上网发行比例":{"number_fg_mode":"disabled","number_bg_mode":"gradient","bg_gradient":1.0,"number_format":{"style":"percent","minimumFractionDigits":2.0,"maximumFractionDigits":2.0}}},"settings":true,"theme":"Pro Light","title":null,"group_by":[],"split_by":[],"sort":[["market_amount","desc"]],"filter":[],"expressions":{"上网发行比例":"\"market_amount" / "amount"","price_cent":"\"price" * 100"},"columns":["ts_code","sub_code","name","ipo_date","issue_date","amount","market_amount","上网发行比例","price_cent","price","pe","limit_amount","funds","ballot"],"aggregates":{}}
[18]: from pathlib import Path
[20]: config2 = Path("C:/Users/mate/Downloads/untitled.config.json").read_text(encoding="utf8")
[22]: print(config2)
{"version":"3.6.0","plugin":"Datagrid","plugin_config":{"columns":{},"edit_mode":"READ_ONLY","scroll_lock":false},"columns_config":{"name":{"format":"bold","string_color_mode":"foreground"},"price":{"number_bg_mode":"gradient","bg_gradient":557.8},"market_amount":{"number_format":{"minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"price_cent":{"number_format":{"style":"currency","currency":"CNY","minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"amount":{"number_bg_mode":"gradient","bg_gradient":1057477.0,"number_format":{"minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"上网发行比例":{"number_fg_mode":"disabled","number_bg_mode":"gradient","bg_gradient":1.0,"number_format":{"style":"percent","minimumFractionDigits":2.0,"maximumFractionDigits":2.0}},"settings":true,"theme":"Pro Light","title":null,"group_by":[],"split_by":[],"sort":[["market_amount","desc"]],"filter":[],"expressions":{"上网发行比例":"\"market_amount\" / \"amount\"","price_cent":"\"price\" * 100"},"columns":["ts_code","sub_code","name","ipo_date","issue_date","amount","market_amount","上网发行比例","price_cent","price","pe","limit_amount","funds","ballot"],"aggregates":{}}
[ ]:


首经贸金融科技实验室 - GitCo
项目文件预览 - 金融AI编程实
Best JSON Formatter and JSC
Best JSON Viewer and JSON
codebeautify.org/jsonviewer


{"version":"3.6.0","plugin":"Datagrid","plugin_config":{"columns":{},"edit_mode":"READ_ONLY","scroll_lock":false},"columns_config":{"name":{"format":"bold","string_color_mode":"foreground"},"price":{"number_bg_mode":"gradient","bg_gradient":557.8},"market_amount":{"number_format":{"minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"price_cent":{"number_format":{"style":"currency","currency":"CNY","minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"amount":{"number_bg_mode":"gradient","bg_gradient":1057477.0,"number_format":{"minimumFractionDigits":0.0,"maximumFractionDigits":2.0}},"上网发行比例":{"number_fg_mode":"disabled","number_bg_mode":"gradient","bg_gradient":1.0,"number_format":{"style":"percent","minimumFractionDigits":2.0,"maximumFractionDigits":2.0}},"settings":true,"theme":"Pro Light","title":null,"group_by":[],"split_by":[],"sort":[["market_amount","desc"]],"filter":[],"expressions":{"上网发行比例":"\"market_amount\" / \"amount\"","price_cent":"\"price\" * 100"},"columns":["ts_code","sub_code","name","ipo_date","issue_date","amount","market_amount","上网发行比例","price_cent","price","pe","limit_amount","funds","ballot"],"aggregates":{}}
Ln: 1 Col: 1165
size: 1.15 KB
File
URL
Auto Update
Big Num
Tree Viewer
2 Tab Space
Beautify
Minify
Validate
to XML
to CSV
视图
选择一个节点...
object {14}
version : 3.6.0
plugin : Datagrid
plugin_config {3}
columns {0}
(清空 object)
edit_mode : READ_ONLY
scroll_lock : false
columns_config {6}
name {2}
format : bold
string_color_mode : foreground
price {2}
number_bg_mode : gradient
bg_gradient : 557.8
market_amount {1}
number_format {2}
minimumFractionDigits : 0
maximumFractionDigits : 2


trial-perspective.ipynb
Code
Python 3 (ipykernel)
[36]:
Treemap
Split By
Order By
ts_code
Where
Size
count
ts_code
Color
sum
funds
Tooltip
symbol
ipo_date
issue_date
2,400
1,200
0
[37]:
PerspectiveWidget(df)
Y Bar
Group By
exchange
market
Split By
area
Order By
amount
Where
Y Axis
All Columns
NEW COLUMN
act_ent_type
act_name
funds
主板
科创板
创业板
主板
北交所
SSE
SZSE
BSE
exchange, market
[40]:
PerspectiveWidget(df)
Y Bar
Group By
bucket_pe
Split By
Order By
Where
market
北交所
Y Axis
count
pe
All Columns
NEW COLUMN
bucket_pe
act_ent_type
pe
bucket_pe

直方图分桶