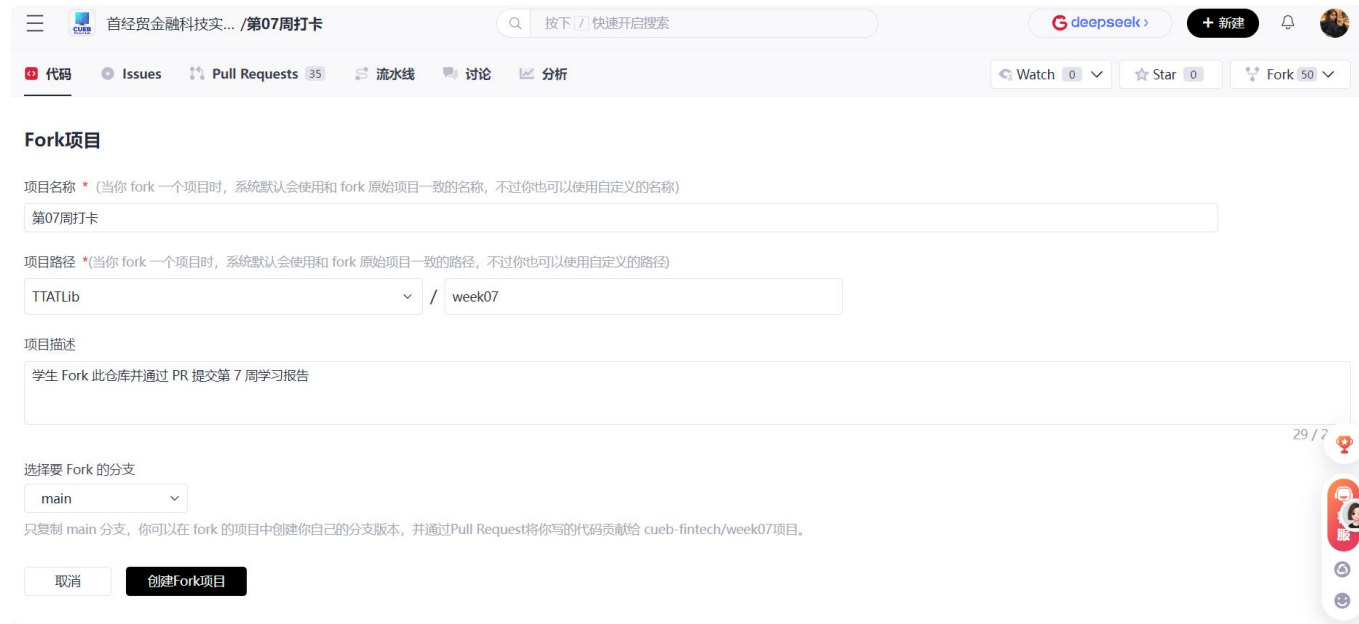


金融编程与计算

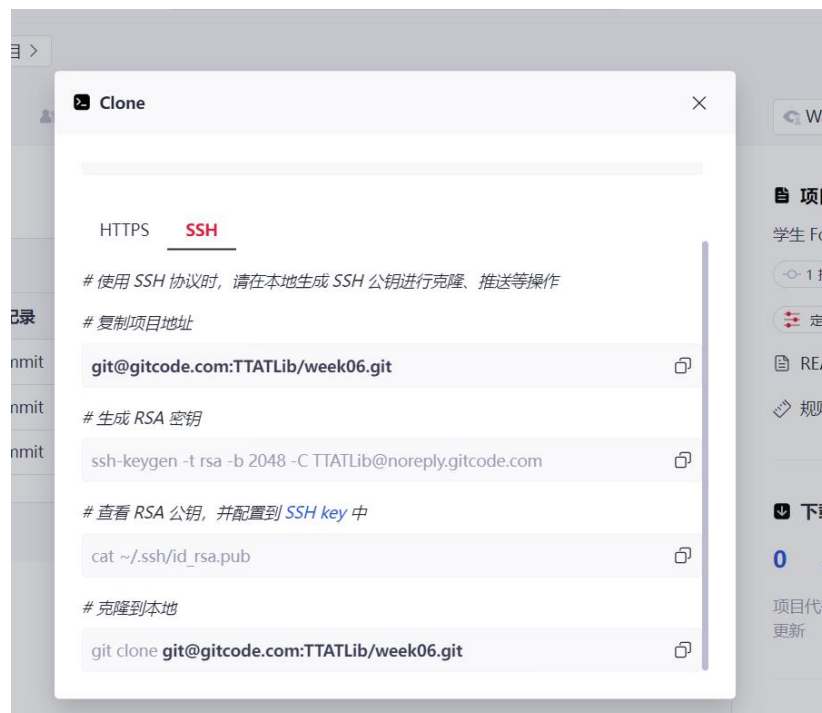
Week07——数据可视化与交互（初级）

一、Fork 第 07 周打卡 仓库至你的名下，然后将你名下的这个仓库 Clone 到你的本地计算机

（一）fork



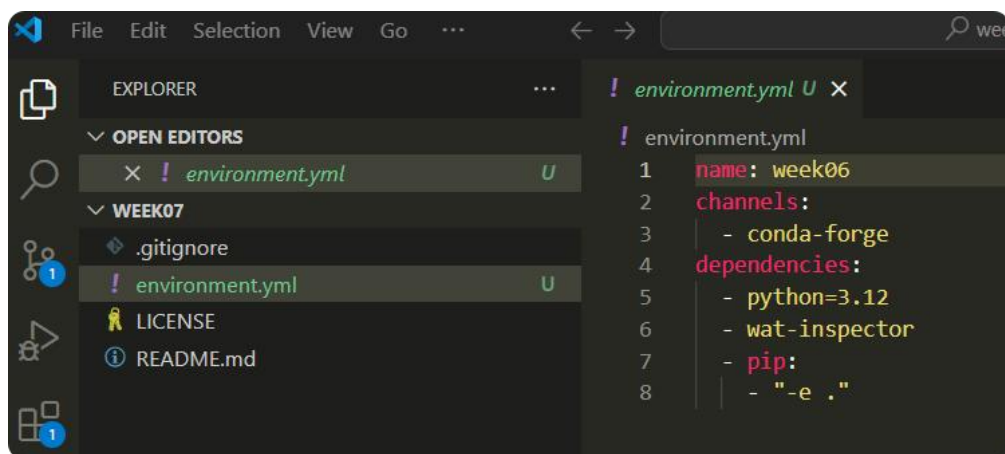
（二）Clone



```
(base) Administrator@PC-20220115W0JX MINGW64 ~/repo
$ git clone git@gitcode.com:TTATLib/week07.git
Cloning into 'week07'...
remote: Enumerating objects: 5, done.
remote: Counting objects: 100% (5/5), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 5 (delta 0), reused 5 (delta 0), pack-reused 0 (from 0)
Receiving objects: 100% (5/5), 8.45 KiB | 1.41 MiB/s, done.
```

二、用 VS Code 打开项目目录，新建一个 `environment.yml` 文件，指定安装 `Python 3.12` 和 `jupyterlab`，然后运行 `conda env create` 命令创建 Conda 环境

(一) 从 week06 里面 copy 过来



(二) 创建 conda 环境

```
(base) Administrator@PC-20220115W0JX MINGW64 ~/repo
$ cp week06/environment.yml week07/

(base) Administrator@PC-20220115W0JX MINGW64 ~/repo
$ cd week07

(base) Administrator@PC-20220115W0JX MINGW64 ~/repo/week07 (ma
$ conda env create
Retrieving notices: done
Channels:
- conda-forge
- https://repo.anaconda.com/pkgs/main
- https://repo.anaconda.com/pkgs/r
- https://repo.anaconda.com/pkgs/msys2
Platform: win-64
Collecting package metadata (repodata.json): \ |
```

```
! environment.yml U X
! environment.yml
1  name: week07
2  channels:
3    - conda-forge
4  dependencies:
5    - python=3.12
6    - wat-inspector
7    - jupyterlab
8
```

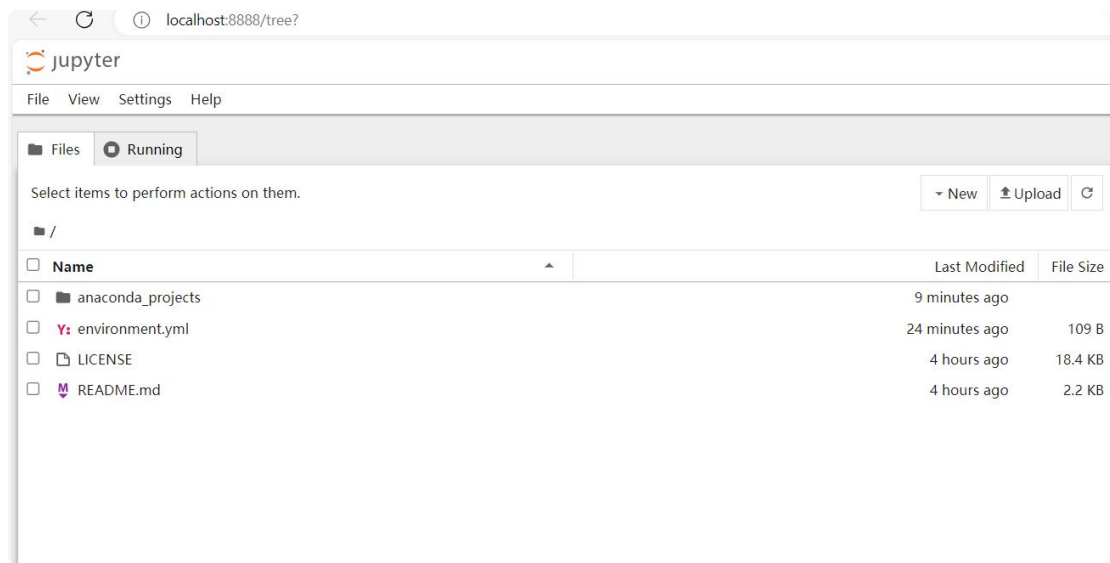
(三) 在项目目录下，运行 `jupyter lab` 命令，启动 后端 (Backend) 服务，在浏览器里粘贴地址访问 前端 (Frontend) 页面

```
(base) Administrator@PC-20220115WOJX MINGW64 ~/repo/week07 (main)
$ conda activate week07
(week07)
Administrator@PC-20220115WOJX MINGW64 ~/repo/week07 (main)
$ ipython
Python 3.12.7 | packaged by Anaconda, Inc. | (main, Oct 4 2024, 13:17:27) [
Type 'copyright', 'credits' or 'license' for more information
IPython 8.27.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: quit
```

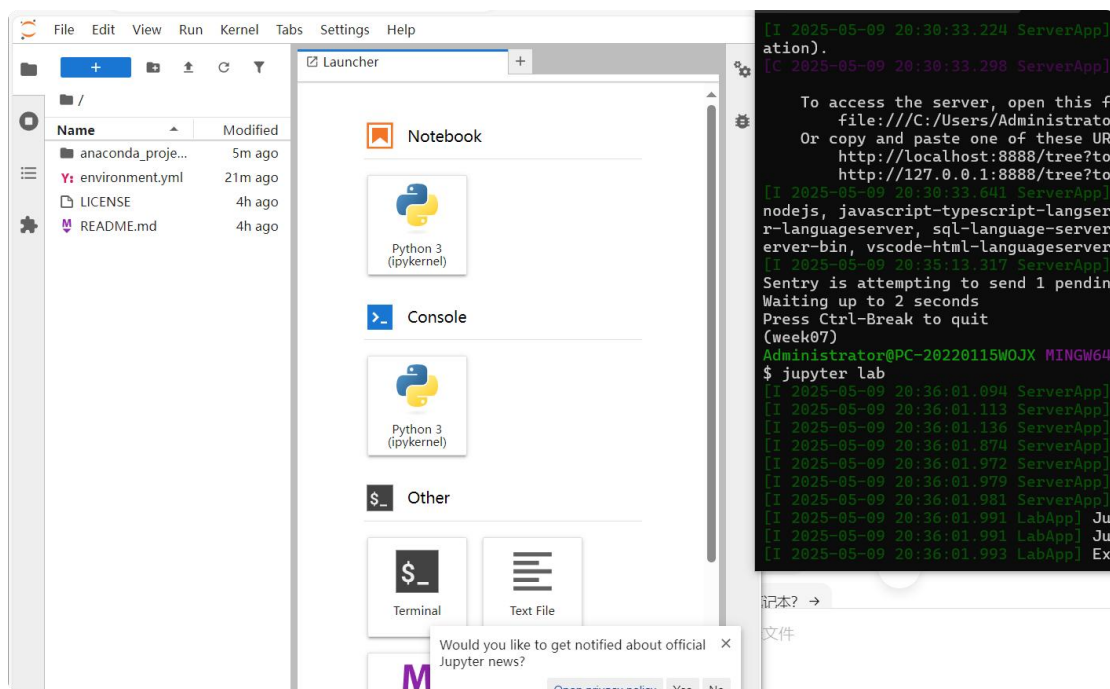
打开 jupyter notebook

```
Administrator@PC-20220115WOJX MINGW64 ~/repo/week07 (main)
$ jupyter notebook
[I 2025-05-09 20:30:19.746 ServerApp] Extension package aext_assistant took 4.0615s to import
[I 2025-05-09 20:30:19.820 ServerApp] ***** ENVIRONMENT [Panels] Environment.PR
[I 2025-05-09 20:30:19.827 ServerApp] ***** ENVIRONMENT [Panels] Environment.PR
[E 2025-05-09 20:30:20.260 ServerApp] Database not initialized by migration_upgrade(). Runni
Database not initialized by migration_upgrade(). Running it again
[I 2025-05-09 20:30:20.272 ServerApp] --> Alembic Config: {'db_conn_url': 'sqlite:///C:\\Use
k07\\anaconda_projects\\db\\project_filebrowser.db', 'db_path': 'C:\\Users\\Administrator\\r
ts\\db\\project_filebrowser.db'}
[I 2025-05-09 20:30:20.273 ServerApp] Default dir: C:\\Users\\Administrator\\repo\\week07. Runni
[I 2025-05-09 20:30:20.273 ServerApp] Checking for database file @ C:\\Users\\Administrator\\re
db\\project_filebrowser.db
[I 2025-05-09 20:30:20.275 ServerApp] Creating database file...
[I 2025-05-09 20:30:20.400 ServerApp] Migrations executed
[I 2025-05-09 20:30:20.475 ServerApp] --> Alembic Config: {'db_conn_url': 'sqlite:///C:\\Use
k07\\anaconda_projects\\db\\project_filebrowser.db', 'db_path': 'C:\\Users\\Administrator\\r
ts\\db\\project_filebrowser.db'}
[I 2025-05-09 20:30:20.476 ServerApp] Default dir: C:\\Users\\Administrator\\repo\\week07. Runni
[I 2025-05-09 20:30:20.476 ServerApp] Checking for database file @ C:\\Users\\Administrator\\re
db\\project_filebrowser.db
[I 2025-05-09 20:30:20.477 ServerApp] Database file already exists
[I 2025-05-09 20:30:20.494 ServerApp] Migrations executed
[I 2025-05-09 20:30:20.494 ServerApp] Checking project configuration: C:\\Users\\Administrator
```



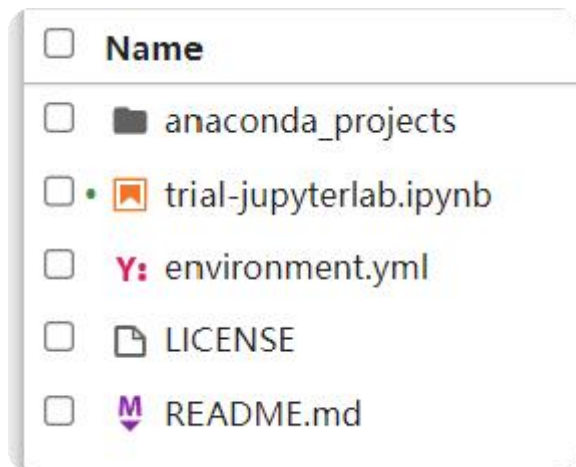
浏览器里的叫前端，终端里的叫后端

输入 jupyter lab



(四) 在 JupyterLab 页面里, 新建一个 Notebook, 改名为 **trial-jupyterlab.ipynb**

```
$ ls -l
total 29
drwxr-xr-x 1 Administrator 197121 0 5月 9 20:30 anaconda_projects/
-rw-r--r-- 1 Administrator 197121 109 5月 9 20:15 environment.yml
-rw-r--r-- 1 Administrator 197121 18805 5月 9 16:12 LICENSE
-rw-r--r-- 1 Administrator 197121 2239 5月 9 16:12 README.md
-rw-r--r-- 1 Administrator 197121 628 5月 9 20:50 trial-jupyterlab.ipynb
(week07)
```



1.在单元格 (Cell) 里编写 Python 代码, 按 Shift+Enter 运行 Cell 并下移

```
x = 5
x = "a" * x
print(x)
```

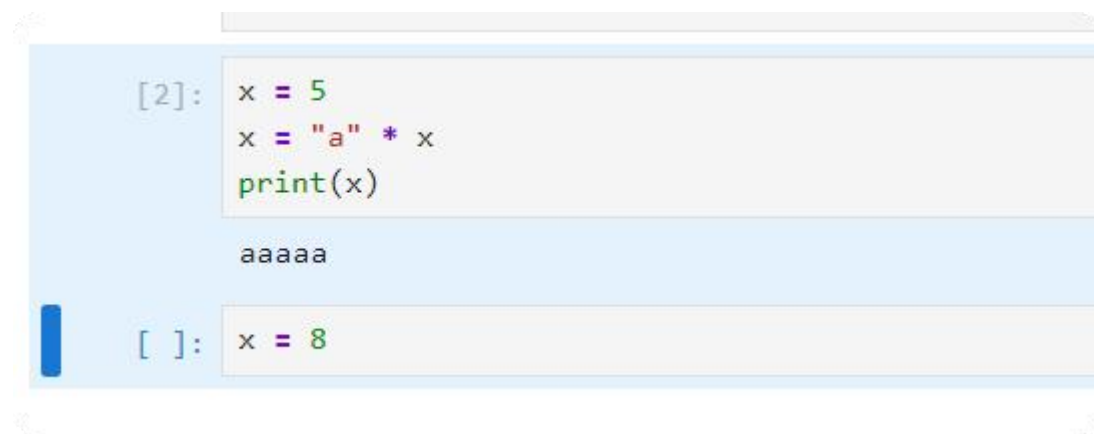
aaaaa

2.在单元格 (Cell) 上按 ESC 切换到 命令模式 (command mode), 按 Enter 切换到 编写模式 (edit mode)



A: above B: below

3.在单元格 (Cell) 的命令模式下, 按 j 选择下一个, 按 k 选择上一个, 按 a 在上方添加, 按 b 在下方添加, 按 dd 删除, 按住 Shift 多选, 按 x 剪切, 按 c 复制, 按 v 粘贴, 按 Shift+M 合并, 按 z 撤销, 按 Shift+Z 重做, 按 Shift+L 显示/隐藏代码行号



```
[2]: x = 5
      x = "a" * x
      print(x)

      aaaaa

[ ]: x = 8
```

Shift+下箭把这两个都选上

Shift+M 合并

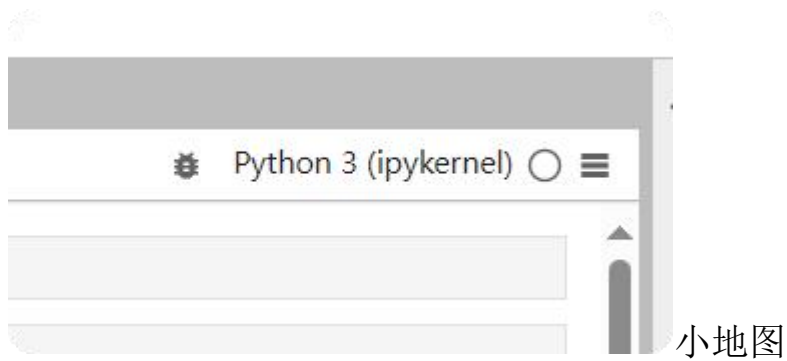


```
[ ]: x = 5
      x = "a" * x
      print(x)

      x = 8
```

Z: 撤销

Shift+Z: 撤销后再撤销



4. 单元格最后一行如果是 表达式 (expression) 且运行后返回的对象不是 `None`，则计输出 (Out)，否则只计输入 (In)，序号为 `i` 的输出，可以用 `_i` 变量来引用

```
x = 5
x = "a" * x
print(x)
x = 8
```

aaaaa 最后一行不是表达式

```
[3]: x = 5
      x = "a" * x
      print(x)
      x = 8
```

aaaaa

```
[4]: x = 5
      x = "a" * x
      print(x)
      x = 8
      x * 3
```

aaaaa

[4]: 24

```
[6]: a = [4, 2, 9]
a

[6]: [4, 2, 9]

[7]: a.pop()

[7]: 9

[8]: a

[8]: [4, 2]

[9]: print(a)

[4, 2]

[10]: a

[10]: [4, 2]

[13]: a.append(7)
a

[13]: [4, 2, 7, 7]

[12]: a

[12]: [4, 2, 7]

[14]: _8

[14]: [4, 2, 7, 7]
```

序号为 `i` 的输出，可以用 `_i` 变量来引用

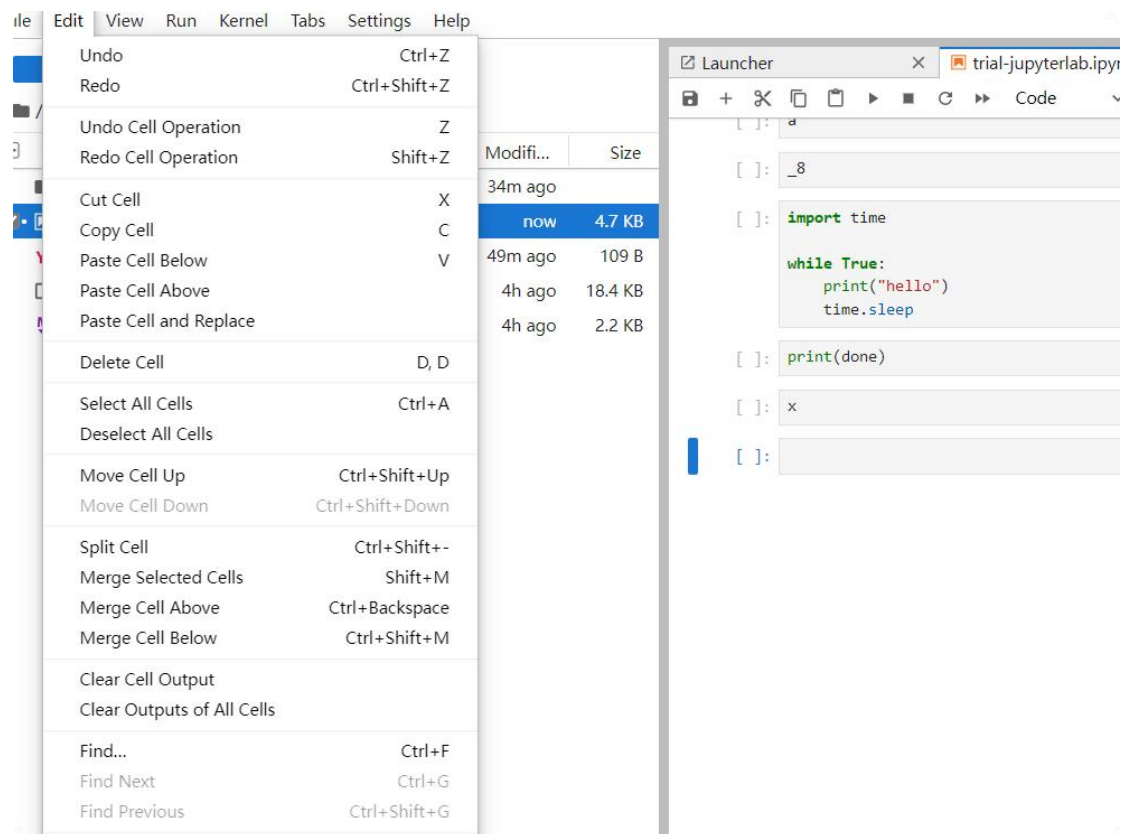

```
: import time

while True:
    print("hello")
    time.sleep(3)

hello
hello
hello
hello
```

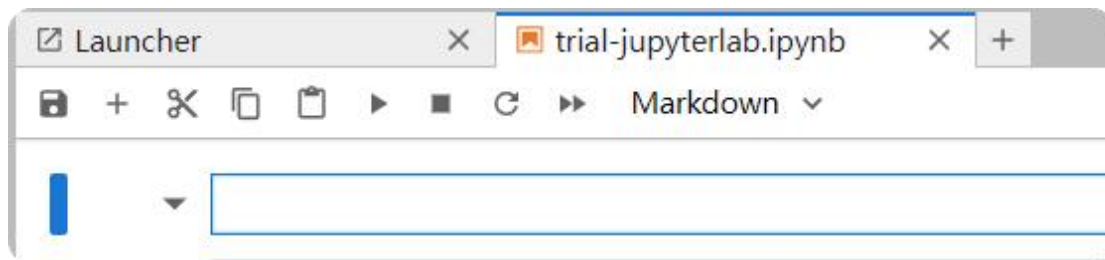
time.sleep (3) 每三秒出来一个 hello

清空所有输出: clear output of all cells



5. 在单元格 (Cell) 的命令模式下, 按 m 切换至 Markdown 模式, 按 y 切换至 Python 模式

(区别在于前面有没有方括号)



6. 用豆包 (或 DeepSeek 等任何大模型) 生成一段示例 Markdown 代码，复制粘贴进 **Markdown** 单元格，运行以呈现 (Render)

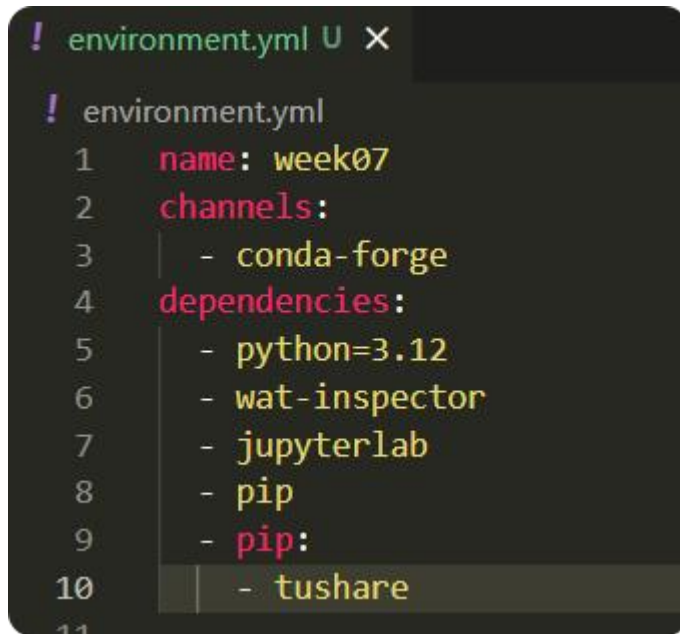


(五) 通过 **tushare** 软件包下载保存一些数据：

1. 在 Tushare 网站上 注册 并登陆，完善修改个人资料，浏览阅读 平台介绍 和 数据接口



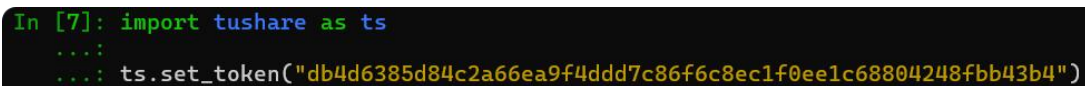
2.修改 environment.yml 文件, 添加 pip: tushare (注意, conda-forge 没有收录 tushare, 只能从 PyPI 安装, 参考) 依赖项, 运行 conda env update 更新 Conda 环境



```
! environment.yml
! environment.yml
1  name: week07
2  channels:
3    - conda-forge
4  dependencies:
5    - python=3.12
6    - wat-inspector
7    - jupyterlab
8    - pip
9    - pip:
10     - tushare
11
```

3. 在终端 (Terminal) 激活 week07 Conda 环境, 运行 ipython 命令启动 IPython 交互界面 (IPython 是 Jupyter 项目的一部份, ipython 是 jupyterlab 的依赖项之一)

4. 在 IPython 提示符下, 运行下面的 Python 代码设置 Tushare Token



```
In [7]: import tushare as ts
...:
...: ts.set_token("db4d6385d84c2a66ea9f4ddd7c86f6c8ec1f0ee1c68804248fbb43b4")
```

5. 按 Ctrl+D 结束前面的 IPython 进程, 再重新启动一个新的 IPython 进程, 运行下面的 Python 代码向 Tushare 服务器请求 IPO 新股列表 数据, 并保存在本地

```

In [2]: import tushare as ts
In [3]: pro = ts.pro_api()
In [4]: pro.new_share()
Out[4]:

```

	ts_code	sub_code	name	ipo_date	issue_date	amount	market_amount	price	pe	limit_amount	funds	ballot
0	001390.SZ	001390	古麒绒材	20250519	None	5000.0		0.0	0.00	0.00	2.00	0.000
1	603014.SH	732014	威高血净	20250508	None	4114.0	2685.0	26.50	24.82	1.10	10.902	0
2	301595.SZ	301595	太力科技	20250508	None	2707.0	1178.0	17.05	21.55	0.65	4.615	0
3	688755.SH	787755	汉邦科技	20250507	None	2200.0	704.0	22.77	26.35	0.50	5.009	0
4	301636.SZ	301636	泽润新能	20250428	None	1597.0	774.0	33.06	17.57	0.45	5.279	0
...
1995	300776.SZ	300776	帝尔激光	20190507	20190517	1654.0	1654.0	57.71	22.99	1.60	9.543	0
1996	300777.SZ	300777	中简科技	20190506	20190516	4001.0	3601.0	6.06	22.98	1.10	2.425	0
1997	603267.SH	732267	鸿远电子	20190430	20190515	4134.0	3721.0	20.24	16.50	1.60	8.367	0
1998	600989.SH	730989	宝丰能源	20190430	20190516	73336.0	66002.0	11.12	22.07	22.00	81.550	0
1999	300778.SZ	300778	新城市	20190425	20190510	2000.0	2000.0	27.33	22.99	2.00	5.466	0

```

[2000 rows x 12 columns]

```

安装 pyarrow

```

! environment.yml
! environment.yml
1  name: week07
2  channels:
3    - conda-forge
4  dependencies:
5    - python=3.12
6    - wat-inspector
7    - jupyterlab
8    - pyarrow
9    - pip
10   - pip:
11     - tushare
12
13

```

启动 ipython

```
In [2]: import tushare as ts
In [3]: pro = ts.pro_api()
In [4]: df = pro.new_share()
In [5]: df
Out[5]:
```

	ts_code	sub_code	name	ipo_date	issue_date	amount	market_amount	price	pe	limit_amount	funds	ballot	
0	001390.SZ	001390	古麒绒材	20250519	None	5000.0		0.0	0.00	0.00	2.00	0.000	0
1	603014.SH	732014	威高血净	20250508	None	4114.0	2685.0	26.50	24.82		1.10	10.902	0
2	301595.SZ	301595	太力科技	20250508	None	2707.0	1178.0	17.05	21.55		0.65	4.615	0
3	688755.SH	787755	汉邦科技	20250507	None	2200.0	704.0	22.77	26.35		0.50	5.009	0
4	301636.SZ	301636	泽润新能	20250428	None	1597.0	774.0	33.06	17.57		0.45	5.279	0
...
1995	300776.SZ	300776	帝尔激光	20190507	20190517	1654.0	1654.0	57.71	22.99		1.60	9.543	0
1996	300777.SZ	300777	中简科技	20190506	20190516	4001.0	3601.0	6.06	22.98		1.10	2.425	0
1997	603267.SH	732267	鸿远电子	20190430	20190515	4134.0	3721.0	20.24	16.50		1.60	8.367	0
1998	600989.SH	730989	宝丰能源	20190430	20190516	73336.0	66002.0	11.12	22.07		22.00	81.550	0
25													

```
In [7]: pro.stock_basic()
Out[7]:
```

	ts_code	symbol	name	area	industry	cnsPELL	market	list_date	act_name	act_ent_type
0	000001.SZ	000001	平安银行	深圳	银行	payh	主板	19910403		无实际控制人
1	000002.SZ	000002	万科A	深圳	全国地产	wka	主板	19910129	深圳市人民政府国有资产监督管理委员会	
2	000004.SZ	000004	*ST国华	深圳	软件服务	stgh	主板	19910114	李映彤	民营企业
3	000006.SZ	000006	深振业A	深圳	区域地产	szya	主板	19920427	深圳市人民政府国有资产监督管理委员会	
4	000007.SZ	000007	全新好	深圳	其他商业	qxh	主板	19920413	王玩虹	民营企业
...
5407	920489.BJ	920489	佳先股份	None	None	jxgf	北交所	20200727	None	None
5408	920682.BJ	920682	球冠电缆	None	None	qgdl	北交所	20200727	None	None
5409	920799.BJ	920799	艾融软件	None	None	arrj	北交所	20200727	None	None
5410	920819.BJ	920819	颖泰生物	None	None	ytsw	北交所	20200727	None	None
5411	689009.SH	689009	九号公司-WD	北京	摩托车	jhgs	科创板	20201029		None

[5412 rows x 10 columns]

```
SyntaxError: invalid character ' (' (U+FF08)

In [9]: df = pro.stock_basic()
In [10]: df.columns
AttributeError                                Traceback (most recent call last)
Cell In[10], line 1
----> 1 df.columns

File D:\anaconda\Lib\site-packages\pandas\core\generic.py:6299, in NDFrame.__getattr__(self, name)
    6292 if (
    6293     name not in self._internal_names_set
    6294     and name not in self._metadata
    6295     and name not in self._accessors
    6296     and self._info_axis._can_hold_identifiers_and_holds_name(name)
    6297 ):
    6298     return self[name]
-> 6299 return object.__getattr__(self, name)

AttributeError: 'DataFrame' object has no attribute 'columns'

In [11]: df.columns
Out[11]:
Index(['ts_code', 'symbol', 'name', 'area', 'industry', 'cnsPELL', 'market',
      'list_date', 'act_name', 'act_ent_type'],
      dtype='object')

In [12]:
```

TUSHARE

股票数据

基础数据

股票列表

每日股本 (盘前)

交易日历

股票曾用名

沪深股通成分股

上市公司基本信息

上市公司管理层

管理层薪酬和持股

北交所新旧代码对照

IPO新股上市

股票历史列表

行情数据

财务数据

参考数据


```
In [13]: df = pro.stock_basic(fields="ts_code,symbolname,area")
```

```
In [14]: df
```

```
Out[14]:
```

	ts_code	area
0	000001.SZ	深圳
1	000002.SZ	深圳
2	000004.SZ	深圳
3	000006.SZ	深圳
4	000007.SZ	深圳
...
5407	920489.BJ	None
5408	920682.BJ	None
5409	920799.BJ	None
5410	920819.BJ	None
5411	689009.SH	北京

```
[5412 rows x 2 columns]
```

```
In [15]: df = pro.stock_basic(fields="ts_code,symbol,name,area")
```

```
In [16]: df
```

```
Out[16]:
```

	ts_code	symbol	name	area
0	000001.SZ	000001	平安银行	深圳
1	000002.SZ	000002	万科A	深圳
2	000004.SZ	000004	*ST国华	深圳
3	000006.SZ	000006	深振业A	深圳
4	000007.SZ	000007	全新好	深圳
...
5407	920489.BJ	920489	佳先股份	None
5408	920682.BJ	920682	球冠电缆	None
5409	920799.BJ	920799	艾融软件	None
5410	920819.BJ	920819	颖泰生物	None
5411	689009.SH	689009	九号公司-WD	北京

```
[5412 rows x 4 columns]
```

```
In [17]: df = pro.stock_basic(fields="ts_code,symbol,name,area,industry")
```

```
In [18]: df
```

```
Out[18]:
```

	ts_code	symbol	name	area	industry
0	000001.SZ	000001	平安银行	深圳	银行
1	000002.SZ	000002	万科A	深圳	全国地产
2	000004.SZ	000004	*ST国华	深圳	软件服务
3	000006.SZ	000006	深振业A	深圳	区域地产
4	000007.SZ	000007	全新好	深圳	其他商业
...
5407	920489.BJ	920489	佳先股份	None	None
5408	920682.BJ	920682	球冠电缆	None	None
5409	920799.BJ	920799	艾融软件	None	None
5410	920819.BJ	920819	颖泰生物	None	None
5411	689009.SH	689009	九号公司-WD	北京	摩托车

```
[5412 rows x 5 columns]
```

```
In [19]: df = pro.stock_basic(fields="ts_code,symbol,name,area,industry,fullname")
In [20]: df
Out[20]:
```

	ts_code	symbol	name	area	industry	fullname
0	000001.SZ	000001	平安银行	深圳	银行	平安银行股份有限公司
1	000002.SZ	000002	万科A	深圳	全国地产	万科企业股份有限公司
2	000004.SZ	000004	*ST国华	深圳	软件服务	深圳国华网安科技股份有限公司
3	000006.SZ	000006	深振业A	深圳	区域地产	深圳市振业(集团)股份有限公司
4	000007.SZ	000007	全新好	深圳	其他商业	深圳市全新好股份有限公司
...
5407	920489.BJ	920489	佳先股份	None	None	安徽佳先功能助剂股份有限公司
5408	920682.BJ	920682	球冠电缆	None	None	宁波球冠电缆股份有限公司
5409	920799.BJ	920799	艾融软件	None	None	上海艾融软件股份有限公司
5410	920819.BJ	920819	颖泰生物	None	None	北京颖泰嘉和生物科技股份有限公司
5411	689009.SH	689009	九号公司-WD	北京	摩托车	九号有限公司

```
[5412 rows x 6 columns]
```

保存

```
In [21]: df = pro.stock_basic(fields="ts_code,symbol,name,area,industry,fullname,ename,cnspell,market,exchange,curr_ty
...: pe,list_status,list_sate,delist_date,is_hs,act_ent_type")
In [22]: df
Out[22]:
```

	ts_code	symbol	name	area	industry	...	curr_type	list_status	delist_date	is_hs	act_ent_type	
0	000001.SZ	000001	平安银行	深圳	银行	...	CNY	L	None	S	无	
1	000002.SZ	000002	万科A	深圳	全国地产	...	CNY	L	None	S	地方国企	
2	000004.SZ	000004	*ST国华	深圳	软件服务	...	CNY	L	None	N	民营企业	
3	000006.SZ	000006	深振业A	深圳	区域地产	...	CNY	L	None	S	地方国企	
4	000007.SZ	000007	全新好	深圳	其他商业	...	CNY	L	None	N	民营企业	
...	
5407	920489.BJ	920489	佳先股份	None	None	...	CNY	L	None	N	None	
5408	920682.BJ	920682	球冠电缆	None	None	...	CNY	L	None	N	None	
5409	920799.BJ	920799	艾融软件	None	None	...	CNY	L	None	N	None	
5410	920819.BJ	920819	颖泰生物	None	None	...	CNY	L	None	N	None	
5411	689009.SH	689009	九号公司-WD	北京	摩托车	...	CNY	L	None	None	None	

```
[5412 rows x 15 columns]
In [23]: df.to_parquet("stock_basic.parquet")
```

在终端查看


```

(base) Administrator@PC-20220115W0JX MINGW64 ~/repo/week07 (main)
$ ls -l
total 557
drwxr-xr-x 1 Administrator 197121 0 5月 9 20:30 anaconda_projects/
-rw-r--r-- 1 Administrator 197121 158 5月 10 12:01 environment.yml
-rw-r--r-- 1 Administrator 197121 18805 5月 9 16:12 LICENSE
-rw-r--r-- 1 Administrator 197121 121474 5月 10 12:08 new_share.parquet
-rw-r--r-- 1 Administrator 197121 2239 5月 9 16:12 README.md
-rw-r--r-- 1 Administrator 197121 387352 5月 10 12:19 stock_basic.parquet
-rw-r--r-- 1 Administrator 197121 31480 5月 9 21:53 trial-jupyterlab.ipynb

(base) Administrator@PC-20220115W0JX MINGW64 ~/repo/week07 (main)
$ ls -lh
total 557K
drwxr-xr-x 1 Administrator 197121 0 5月 9 20:30 anaconda_projects/
-rw-r--r-- 1 Administrator 197121 158 5月 10 12:01 environment.yml
-rw-r--r-- 1 Administrator 197121 19K 5月 9 16:12 LICENSE
-rw-r--r-- 1 Administrator 197121 119K 5月 10 12:08 new_share.parquet
-rw-r--r-- 1 Administrator 197121 2.2K 5月 9 16:12 README.md
-rw-r--r-- 1 Administrator 197121 379K 5月 10 12:19 stock_basic.parquet
-rw-r--r-- 1 Administrator 197121 31K 5月 9 21:53 trial-jupyterlab.ipynb

(base) Administrator@PC-20220115W0JX MINGW64 ~/repo/week07 (main)
$ ls -lh *.parquet
-rw-r--r-- 1 Administrator 197121 119K 5月 10 12:08 new_share.parquet
-rw-r--r-- 1 Administrator 197121 379K 5月 10 12:19 stock_basic.parquet

```

6. 询问豆包 (或 DeepSeek 等任何大模型), 初步了解 Parquet 格式和 CSV 格式的特点和适用领域

Parquet 格式的特点与适用领域

特点：

- **列式存储**：按列组织数据，查询时可仅读取所需列，减少数据扫描量，提升分析性能，尤其适合多字段表场景。
- **高效压缩**：采用复杂压缩技术，减少存储成本，如去除重复信息，以更节省空间的格式存储。
- **支持嵌套结构**：能有效管理如 JSON、物联网传感器数据中的嵌套数组等复杂数据结构。
- **兼容性强**：与 Apache Spark、Hive、Hadoop 等分布式系统和数据工具兼容，支持多语言（Python、Java 等），且不绑定特定框架。
- **二进制格式**：不可直接用文本编辑器阅读，需通过特定工具解析。

适用领域：

- **数据湖与数据仓库**：高效存储海量数据，支持快速查询与商业智能分析，如零售连锁店分析销售模式。
 - **机器学习管道**：处理大型训练数据集中的特定属性，提升模型训练效率。
 - **复杂数据分析**：金融领域处理海量交易数据、计算风险指标；医疗领域整合多系统患者记录并支持模式演进。
 - **分布式计算场景**：在 Hadoop 环境中性能优于传统格式，适合大规模数据的复杂处理。
-

CSV 格式的特点与适用领域

特点：

- **纯文本格式**：以逗号（或其他符号）分隔字段，可用任何文本编辑器打开编辑，直观易读。
- **跨平台兼容**：不受操作系统限制，Windows、macOS、Linux 均可读写。
- **简单通用**：支持多种分隔符（逗号、分号、制表符等），广泛用于程序间数据转移，如数据库与电子表格间交换数据。
- **易处理**：可直接导入 Excel、Python 的 Pandas 库、R 语言等工具，适合小规模数据的手动校对与初步处理。
- **无复杂约束**：不支持数据类型定义（如日期、二进制），无完整性约束或关系，结构简单。

适用领域：

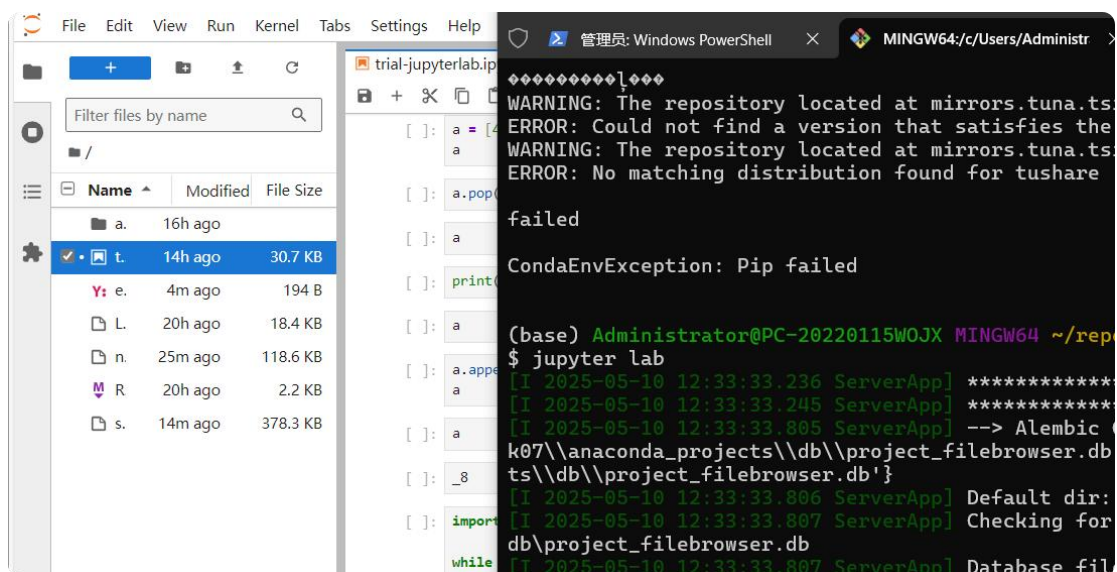
- **简单数据交换**：不同程序间转移表格数据（如私有格式数据库导出至电子表格）。
- **小规模数据处理**：科研实验数据记录、社会调查结果整理，便于手动编辑或轻量级分析。
- **通用输入 / 输出格式**：作为可选择的格式，适配不兼容的系统间数据传输，如企业交换业务数据。

（六）通过 **perspective-python** 软件包查看 **polars.DataFrame** 数据，实践交互式可视化：

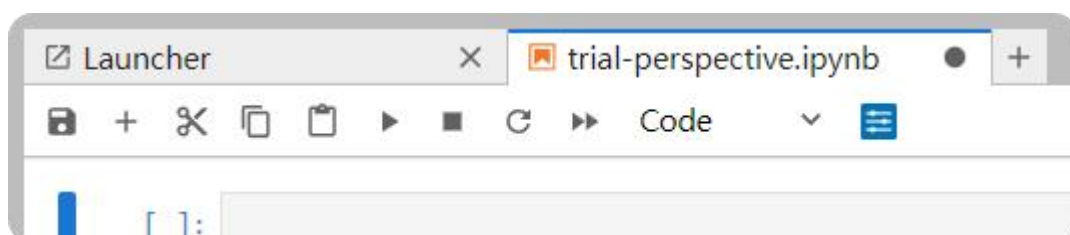
1. 修改 `environment.yml` 文件，添加 `perspective-python` 和 `polars` 依赖项，运行 `conda env update` 更新 Conda 环境

```
dependencies:
- python=3.12
- wat-inspector
- jupyterlab
- pyarrow
- perspective-python
- polars
- pip
- pip:
- tushare
```

2. 启动 JupyterLab，新建一个 Notebook，改名为
trial-perspective.ipynb



重命名



3. 调用 `polars.read_parquet` 函数，分别读取磁盘 (disk) 中的 `new_share.parquet` 文件和 `stock_basic.parquet` 文件，得到内存 (memory) 中的 `polars.DataFrame` 对象，命名为 `d1` 和 `d2`

```
import polars as pl

d1 = pl.read_parquet("new_share.parquet")
d1
```

shape: (2,000, 12)

ts_code	sub_code	name	ipo_date	issue_date	amount	market_amount	price	pe	limit_amount	funds	ballot
str	str	str	str	str	f64	f64	f64	f64	f64	f64	f64
"001390.SZ"	"001390"	"古麒新材"	"20250519"	null	5000.0	0.0	0.0	0.0	2.0	0.0	0.0
"603014.SH"	"732014"	"威高血净"	"20250508"	null	4114.0	2685.0	26.5	24.82	1.1	10.902	0.03
"301595.SZ"	"301595"	"太力科技"	"20250508"	null	2707.0	1178.0	17.05	21.55	0.65	4.615	0.02
"688755.SH"	"787755"	"汉邦科技"	"20250507"	null	2200.0	704.0	22.77	26.35	0.5	5.009	0.03
"301636.SZ"	"301636"	"泽润新能"	"20250428"	null	1597.0	774.0	33.06	17.57	0.45	5.279	0.02
...
"300776.SZ"	"300776"	"帝尔激光"	"20190507"	"20190517"	1654.0	1654.0	57.71	22.99	1.6	9.543	0.01
"300777.SZ"	"300777"	"中简科技"	"20190506"	"20190516"	4001.0	3601.0	6.06	22.98	1.1	2.425	0.04
"603267.SH"	"732267"	"鸿远电子"	"20190430"	"20190515"	4134.0	3721.0	20.24	16.5	1.6	8.367	0.03
"600989.SH"	"730989"	"宝丰能源"	"20190430"	"20190516"	73336.0	66002.0	11.12	22.07	22.0	81.55	0.25
"300778.SZ"	"300778"	"新城市"	"20190425"	"20190510"	2000.0	2000.0	27.33	22.99	2.0	5.466	0.02

```
d2 = pl.read_parquet("stock_basic.parquet")
d2
```

shape: (5,412, 15)

ts_code	symbol	name	area	industry	fullname	enname	cspell	market	exchange	curr_type	list_status	delist_date	is_hs	act_ent_type
str	str	str	str	str	str	str	str	str	str	str	str	str	null	str
"000001.SZ"	"000001"	"平安银行"	"深圳"	"银行"	"平安银行股份有限公司"	"Ping An Bank Co., Ltd."	"payh"	"主板"	"SZSE"	"CNY"	"L"	null	"S"	"无"
"000002.SZ"	"000002"	"万科A"	"深圳"	"全国地产"	"万科企业股份有限公司"	"China Vanke Co.,Ltd."	"wka"	"主板"	"SZSE"	"CNY"	"L"	null	"S"	"地方国企"
"000004.SZ"	"000004"	"ST国华"	"深圳"	"软件服务"	"深圳国华网安科技股份有限公司"	"Shenzhen Guohua Network Securi..."	"stgh"	"主板"	"SZSE"	"CNY"	"L"	null	"N"	"民营企业"
"000006.SZ"	"000006"	"深振业A"	"深圳"	"区域地产"	"深圳市振业(集团)股份有限公司"	"Shenzhen Zhenye(Group) Co., Lt..."	"szya"	"主板"	"SZSE"	"CNY"	"L"	null	"S"	"地方国企"
"000007.SZ"	"000007"	"全新好"	"深圳"	"其他商业"	"深圳市全新好股份有限公司"	"Shenzhen Quanzinhao Co.,Ltd."	"qxh"	"主板"	"SZSE"	"CNY"	"L"	null	"N"	"民营企业"
...
"920489.BJ"	"920489"	"佳先股份"	null	null	"安徽佳先功能助剂股份有限公司"	"Anhui Jiaxian Functional Auxil..."	"jxgf"	"上交所"	"BSE"	"CNY"	"L"	null	"N"	null
"920682.BJ"	"920682"	"球冠电缆"	null	null	"宁波球冠电缆股份有限公司"	"Ningbo Qunning Cable Co.,Ltd"	"qgdl"	"上交所"	"BSE"	"CNY"	"L"	null	"N"	null
"920799.BJ"	"920799"	"文融软件"	null	null	"上海文融软件股份有限公司"	"Shanghai I2finance Technology..."	"arj"	"上交所"	"BSE"	"CNY"	"L"	null	"N"	null
"920819.BJ"	"920819"	"颖泰生物"	null	null	"北京颖泰生物科技股份有限公司"	"Nutrichem Company Limited"	"ytsw"	"上交所"	"BSE"	"CNY"	"L"	null	"N"	null
"689009.SH"	"689009"	"九号公司-WD"	"北京"	"摩托车"	"九号有限公司"	"Ninebot Limited"	"jhgs"	"科创板"	"SSE"	"CNY"	"L"	null	null	null

4. 进行适当的列变换，尤其是要把实际上是日期类型的列，从 `polars.String()` 类型转换为 `polars.Date()` 类型

```
] d1.with_columns(  
    ipo_date=pl.col.ipo_date.str.to_date("%Y%m%d"),  
    issue_date=pl.col.issue_date.str.to_date("%Y%m%d"),  
)
```

```
] shape: (2_000, 12)
```

ts_code	sub_code	name	ipo_date	issue_date	amount	market_amount	price	pe	limit_amount	funds	ballot
str	str	str	date	date	f64	f64	f64	f64	f64	f64	f64
"001390.SZ"	"001390"	"古麒新材"	2025-05-19	null	5000.0	0.0	0.0	0.0	2.0	0.0	0.0
"603014.SH"	"732014"	"威高血净"	2025-05-08	null	4114.0	2685.0	26.5	24.82	1.1	10.902	0.03
"301595.SZ"	"301595"	"太力科技"	2025-05-08	null	2707.0	1178.0	17.05	21.55	0.65	4.615	0.02
"688755.SH"	"787755"	"汉邦科技"	2025-05-07	null	2200.0	704.0	22.77	26.35	0.5	5.009	0.03
"301636.SZ"	"301636"	"泽润新能"	2025-04-28	null	1597.0	774.0	33.06	17.57	0.45	5.279	0.02
...
"300776.SZ"	"300776"	"帝尔激光"	2019-05-07	2019-05-17	1654.0	1654.0	57.71	22.99	1.6	9.543	0.01
"300777.SZ"	"300777"	"中简科技"	2019-05-06	2019-05-16	4001.0	3601.0	6.06	22.98	1.1	2.425	0.04
"603267.SH"	"732267"	"鸿远电子"	2019-04-30	2019-05-15	4134.0	3721.0	20.24	16.5	1.6	8.367	0.03
"600989.SH"	"730989"	"宝丰能源"	2019-04-30	2019-05-16	73336.0	66002.0	11.12	22.07	22.0	81.55	0.25
"300778.SZ"	"300778"	"新城市"	2019-04-25	2019-05-10	2000.0	2000.0	27.33	22.99	2.0	5.466	0.02