

《python数据清洗》

-【美】迈克尔. 沃克尔

path

Pycharm

C:\dataClean

终端环境pandas1

环境准备

在vscode和pycharm中均存在这个问题：

```
ImportError: cannot import name 'ABCIndex' from 'pandas.core.dtypes.generic' (C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\dtypes\generic.py)
```

窗口搜索shell以管理员身份打开， notepad C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\indexers\utils.py， 将所有的ABCIndex改成ABCIndexClass

过于复杂， 重开

在Anaconda Prompt 中创建一个包含 Python 3.7.0 和 Pandas 1.2.0 的环境， 然后在 PyCharm 中选择这个环境作为项目的解释器。例如， 在 Anaconda Prompt 中， 你可以运行以下命令：

```
conda create -n myenv python=3.7 pandas=1.2.0 conda activate myenv
conda activate myenv
```

然后在 PyCharm 中选择这个名为 myenv 的环境即可：

settings >> interpreter >> add interpreter >> Conda Environment >> using existing environment
选择myenv即可

依旧有问题 找不到指定DDL， 在settings >> interpreter 里面查看包并降级 numpy 1.18.1, pandas 1.0.3

下载jupyter模块：先关闭代理； pip install jupyter即可； kernel选择当前默认的python3即可

CHAPTER01

在pip install pyreadstat 时报错： error: Microsoft Visual C++ 14.0 or greater is required. Get it with "Microsoft C++ Build Tools": <https://visualstudio.microsoft.com/visual-cpp-build-tools/>

因为pyreadstat是一个混合型的包

解决办法：（因为在conda环境中） conda install -c conda-forge pyreadstat 有遇到
`Solving environment: failed with initial frozen solve. Retrying with flexible solve.

不用管

import pyreadstat 报错： AttributeError: module 'pandas' has no attribute 'Float64Dtype':
原因： pandas版本太低， 升级到1.2.0, 重启pycharm即可

pip install pyreadr报错： error: command 'C:\Program Files (x86)\Microsoft Visual Studio\2022\BuildTools\VC\Tools\MSVC\14.40.33807\bin\HostX86\x64\cl.exe' failed with exit code 2

用conda install -c conda-forge方法

import request(2.26.0)报错： ImportError: cannot import name 'get_host' from 'urllib3.util.url' (C:\Users\0ly\.conda\envs\pandas1\lib\site-packages\urllib3\util\url.py)

检查： url.py文件包含get_host函数，有误。重启电脑再运行，即可。

urllib3(1.25.6)

CHAPTER04

双变量表可以这样看：

```
pd.crosstab(covidtotalonly.total_cases_q, covidtotalonly.total_deaths_q)
```

Executed at 2024.08.02 10:53:58 in 45ms

total_deaths...	123 very low	123 low	123 medium	123 high	123 very high
very low	34	7	1	0	0
low	12	19	10	1	0
medium	1	13	15	13	0
high	0	0	12	24	6
very high	0	0	2	4	36

安装sklearn 时候pip install , conda install 均报错，解决办法：

pip install scikit-learn -i https://pypi.tuna.tsinghua.edu.cn/simple/