

computer_bases thinking

姚雨薇 2020103730

task1 :store part.tbl as tuple

在计算云上运行 载入包(conda下载)

```
import numpy as np
import pandas as pd
import os
import shutil
```

获得本地路径: `os.getcwd()`

输出: `'/home/u2020103730/jupyterlab'`

移动tbl文件: `shutil.copy('/home/u2020103730/part.tbl', '/home/u2020103730/jupyterlab')`

输出: `'/home/u2020103730/jupyterlab/part.tbl'`

不要怪我弄个路径还整这么麻烦…我用的学校计算云，刚进来根本不知道我是谁我在哪我要干什么
读取文件:

```
f=open('part.tbl', "r")
text=f.read()
```

了解一下文件类型:

`type(text)`

结果: `str`

按行划分:

```
tuple_new=tuple(text.split('\n')) # 对str进行划分，以换行符/n进行划分，每行元素作为一个str，存入tuple中
print(type(tuple_new))
print(type(tuple_new[2]))
```

结果: `<class 'tuple'> <class 'str'>`

看一下长度: `len(tuple_new)` 结果: `200001`

按列分割:

```
# 将字符串进行分割，分割成9个元素
tuple_final=[]
for i in range(len(tuple_new)-1):
    tuple_new_new=list(tuple_new[i].split(sep='|',maxsplit=8)) # 因为tuple没有替换元素的功能，座椅转换成list
    try:
        tuple_new_new[0]=int(tuple_new_new[0])
        tuple_new_new[5]=int(tuple_new_new[5])
    except :
        print(i)
        tuple_new_new[7]=float(tuple_new_new[7])
        tuple_new_new[8]=tuple_new_new[8].strip('|')
    tuple_final.append(tuple(tuple_new_new))
```

我觉得这样挺笨的，因为要单个指定每个非字符串型元素的类型，but我实在想不出别的思路

看看成功没:

```
print(len(tuple_final))
tuple_final[len(tuple_final)-1]
```

结果:

```
200000
(200000,
 'peach royal cornsilk sky sandy',
 'Manufacturer#5',
 'Brand#53',
 'MEDIUM ANODIZED TIN',
 22,
 'LG CAN',
 1100.0,
 'xes sleep quick')
```

明确一下类型:

```
type(tuple_final[i])
# 所以tuple_final是个元素是tuple的list
```

task 2:some calculations

回顾一下part.tbl表的内容:

```
CREATE TABLE PART (
  P_PARTKEY integer not null,
  P_NAME varchar(55),
  P_MFGR char(25),
  P_BRAND char(10),
  P_TYPE varchar(25),
  P_SIZE integer,
  P_CONTAINER char(10),
  P_RETAILPRICE float,
  P_COMMENT varchar(23),
```

```
align varchar(1) null );
```

再明确一下自己的任务：统计P_SIZE<15并且P_CONTAINER='WRAP BOX'的P_RETAILPRICE累加和

我们就知道自己该咋做了：

```
sum_reta=0
for i in range(len(tuple_final)):
    if (tuple_final[i][5] <15) & (tuple_final[i][6]=='WRAP BOX'):
        sum_reta=sum_reta+tuple_final[i][7]
sum_reta
```

结果是：2129994.780000001

领悟：

- 多条件时，每个条件都要加()
- 瞎想思路的时候想到一个函数很久没用了：`collections.Counter`