

## Python-Win32com-Excel



星空流

读书/数据分析/摄影爱好者

4 人赞同了该文章

工作中要经常使用Excel，文件少的时候，使用Excel进行数据分析还是很方便的。

但是如果动辄几百M的文件，再用Excel就显得力不从心了，于是就想到了Python。

环境：Windows10 + Python(python-3.7.5-embed-amd64)

python-3.7.5-embed-amd64 是一个最简单的Python环境，在哪儿都可以使用，免安装，缺点就是太简陋了，只有一些最最基本的库，简单到连pip都没有。

为什么用python-3.7.5-embed-amd64，因为办公环境不允许外网下载。

绞尽脑汁、千辛万苦终于安装了pip，再通过pip安装了numpy、pandas、win32com。

为什么使用win32com，而不使用xlwt等库，因为能力有限、实在是没安装成功。（哭脸）

本文的目的是：使用win32com调用excel API，实现对Excel文件的打开、数据读取、数据简单分析、将分析结果写入Excel、最后保存文件、关闭文件。

### 导入必要的库

```
from win32com.client import Dispatch
import pandas as pd
import numpy as np
```

### 前期准备工作

```
xlapp = Dispatch('Excel.Application')
file_path = r'gun_deaths_in_america.xlsx'

# 打开excel
xlbook = xlapp.Workbooks.Open(file_path, False)
```

### 简单操作

```
# 新建一个名为AanLysis的worksheet
xlbook.Worksheets.Add().Name = 'Analysis'
sht_names = [sht.Name for sht in xlbook.Worksheets]
xlbook.Worksheets(2).Name = 'Details'

# 输出worksheets个数
print(xlbook.Worksheets.Count)

# 创建details worksheet对象
details = xlbook.Worksheets('Details')

# 获取Excel Data的范围
row = details.UsedRange.Rows.Count
col = details.UsedRange.Columns.Count

# 获取单元格的值
print(details.Cells(2,1).Value)

# 输出连续单元格的值
print(details.Range('A1:K1').Value)
```

登录即可查看 超5亿 专业优质内容

超 5 千万创作者的优质提问、专业回答、深度文章和精彩视频尽在知乎。

立即登录/注册

## 知乎

```
analysis = xlbook.Worksheets('Analysis')
# 向Excel单元格中写入内容
analysis.Range('A1').Value = 'Win32com On Excel'
values_write = list('abcDeF')
analysis.Range('A2:F2').Value = values_write
analysis.Range(analysis.Cells(3,1),analysis.Cells(3,6)).Value = tuple(values_write)

# 清除内容
analysis.UsedRange.Clear()
```

### 获取数据并创建DataFrame

```
# 获取所有数据
details_value = list(details.Range(details.Cells(1,2),details.Cells(row,col)).Value)

# 将details_value传入Pandas中
details_df = pd.DataFrame(data=details_value[1:],columns=details_value[0])
```

### 使用Pandas进行分析

这里就简单使用describe()方法得到数据的大致描述。

```
# 简单分析
desc = details_df.describe(include=['object','number']).reset_index().T.reset_index()
r = desc.shape[0]
c = desc.shape[1]
```

为什么使用了两次reset\_index和一次转置，这个是为了把describe分析结果的行和列索引都写入到DataFrame的内容中。

### 获取DataFrame的值并进行简单处理

```
# 获取值
desc_values = desc.values.tolist()

# A list can be inserted only if all of the numpy data types are the same
# and there is a built-in data type in the list as well
for i in range(len(desc_values)):
    for j in range(len(desc_values[i])):
        # print(type(desc_values[i][j]))
        if type(desc_values[i][j]) in [np.int32,np.int64,np.float32,np.float64]:
            desc_values[i][j] = np.float64(desc_values[i][j])
```

上面代码是非常重要的，特别是这句英文描述，完美解决了一个困扰我很长时间的Error，“**MemoryError: Creating SafeArray**”，这个error不解决，数据是不能写入Excel的。

一个数据集里面的numpy Data 数据类型必须要一致。

### 写入Excel

```
analysis.Range(analysis.Cells(1,1),analysis.Cells(r,c)).Value = details_df
xlbook.Save()
xlbook.Close()
xlapp.Quit()
```

发布于 2020-03-12 07:31

登录即可查看 **超5亿** 专业优质内容

超 5 千万创作者的优质提问、专业回答、深度文章和精彩视频尽在知乎。

立即登录/注册

知乎

▲ 赞同 4 ▼

● 4 条评论

🚩 分享

♥ 喜欢

★ 收藏


📄 申请转载

...

写下你的评论...

4 条评论


默认 最新

- 

poxcc

为什么不用pandas直接读


2020-03-13

👍 1
- 

madmars

pandas读取不了带密码的excel，这个可以。


2021-06-16

👍 赞
- 

星空流 作者

pandas有pd.read\_excel()方法，但是这个方法也是调用另外一个库，那个库python embed版安装不了，所以目前为止只能想到这个方法。。。

2020-03-13

👍 赞
- 

张伟

excel的这些接口的方法和属性在哪里能找到啊？

2021-06-24

👍 赞

推荐阅读

Python: 读写Excel (openpyxl /...

项目周报汇报的时候要和数据汇总，总是要从不同的excel文件中去获取数据最后汇总到一个excel表里面，所以决定用python直接写个自动化脚本来自动执行。用python来读写excel，目前找了2个方...  
至成心境

python win32com 操作excel和ppt记录

记录学习win32com操作excel和ppt，怕我忘记了 excel操作有许多库，参考 Python Excel，之前用的openpyxl较多，但要去获取导出excel中已经存在的chart没找到好方法，在Stack Overflow上也没...  
辟暑大王

Pytho Excel的

前言数据应用场景，数据处理数据相关的打交道。(xls/xl  
韩诺

登录即可查看 超5亿 专业优质内容

超 5 千万创作者的优质提问、专业回答、深度文章和精彩视频尽在知乎。

立即登录/注册