Upload Excel File Data To MS SQL Server Using Python (Source Code In Description)

<https://youtu.be/8o-6wq2LABA>



建立python與mssql的連接橋梁，create\_engine用來create connection



從這裡import URL class用來建構connection string



用於接到mssql



從excel file load data，也可以把excel變成csv之後再用numpy等等來操作owo

(參考drive上的ML作業

接著準備一些參數

一張含有 文字, 字型, 螢幕擷取畫面, 行 的圖片

自動產生的描述

接著創一個connection string

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

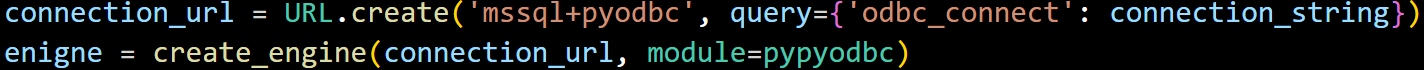
再來是創SQL Alchemy engine

使用URL.create創建url

第一個參數指定DB type(mssql)加上使用的module(pypyodbc)

第二個參數提供connection string，需要以dictionary傳入

用create\_engine建立SQL Alchemy engine傳入剛create好的URL並指定module



最後要Load excel file進來

用pd.read\_excel去讀檔案。確保每個工作表都會被load進來，so設定sheet\_name=None



一張含有 文字, 螢幕擷取畫面, 數字, 文件 的圖片

自動產生的描述到這邊跑一下(用jupiter or colab whatever)

然後print excel\_file就會看到他會以dictionary形式顯示

一張含有 文字, 字型, 螢幕擷取畫面, 設計 的圖片

自動產生的描述

這個dictionary的key是工作表名稱



Type of value is datafram



用以下的for迴圈 sheet\_name就是key

Df\_data就是value



對df object用tosql，在if\_exists有三個選項1.fail 可以raise an exception，2. Replace去取代現有的，3.也可以使用append去添加更多data進去

然後因為他的範例excel沒有index所以=false



For this demo他把sheet 1 sheet 2合併到一個single table

一張含有 文字, 螢幕擷取畫面, 字型, 資訊 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

先安裝SQLAlchemy跟pypyodbc跟panda

注意，似乎要在python檔案所在位置pip install才有用嗎？(發現好像不是

還有一種可能是python要使用跟anaconda同版本的去跑，比如我跑python3 tutorial.py會出錯但是改用python tutorial.py就成功跑起來了。

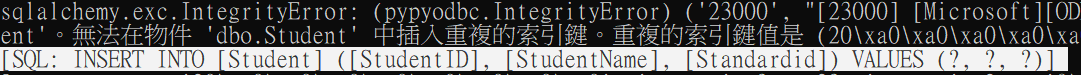
一張含有 文字, 字型, 螢幕擷取畫面 的圖片

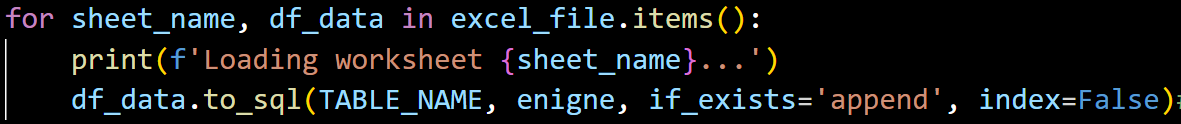
自動產生的描述

一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述

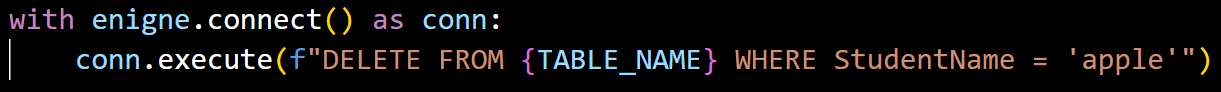
實驗1：把excel清空後執行會報錯，因為會insert 空的東西





實驗2：在python清除SQL的資料

當屬性填錯(Student表中的屬性例如StudentName)或符合該屬性的資料不存在例如已經刪除apple再刪除一次都會報錯



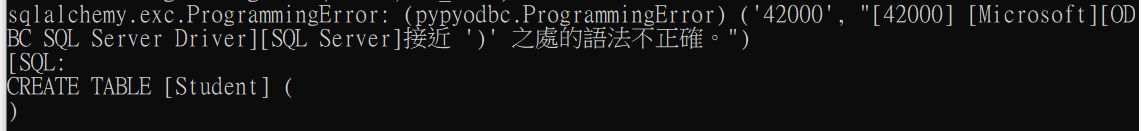
實驗3：append fail replace換掉會影響嗎？

首先創好空的Student，開始測試。

用append：資料好好地進了table 無論該table存在與否

用replace，因為該table存在所以創一個新table取代之，但不知為何失敗了

經一系列針對replace的測試 包含Student初始存在or不存在等，都無法執行成功，故不建議使用



用fail，因為該table存在所以回報錯誤而已

如果該table不存在 他會在資料庫創一個這個table然後不放資料就說 該表格存在了嗚嗚

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

實驗4：Excel的屬性順序錯誤會不會影響？

可以 只要標題跟著你的資料一起就沒問題

實驗5：Excel的屬性有多或有少會不會影響？

會

實驗中將SheetA增加一個屬性StudentBLA然後下面寫一些字當資料，SheetB維持原本屬性數量

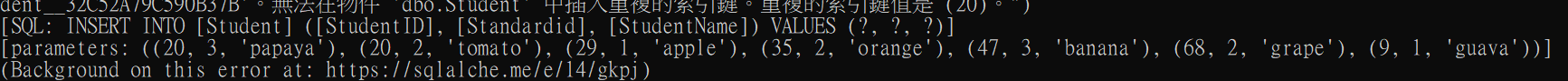
SheetA無法放入報錯，SheetB成功放入。

實驗6：Excel內容如果是數字的話有空格會錯？

發現excel會自動把數字周圍沒用的空格刪掉，他就不會錯OAO!!

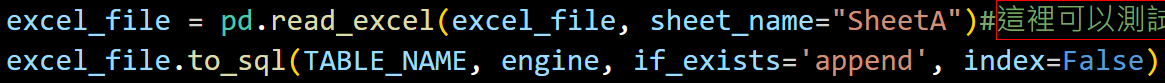
但是之前某次實驗中意外讓數字格子中有空格，就不行了

實驗7：Primary key如果我在匯入的時候重複了會報啥錯？



實驗8：Sheet name不是None的話？

就不用使用for迴圈來拆解excel了



實驗9：對excel檔案預處理

發現如果有換行，以read\_excel然後放入sql來處理，基本上不會有事

不過沒關係 我們還是處理他。

注意數字如果是primary key就不能直接用這個處哩，這個只能把數字寫成Null

換行符處理，將換行符替換成空格。

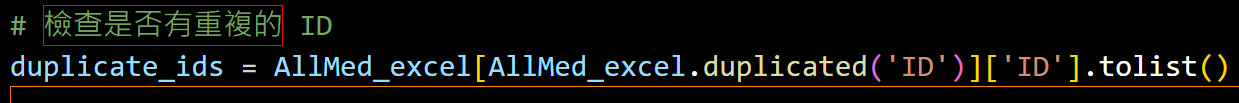
一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

Primary key 獨立



一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

還要檢查當前table的primary key有沒有跟現在excel中的重複，有的話就把那row殺掉

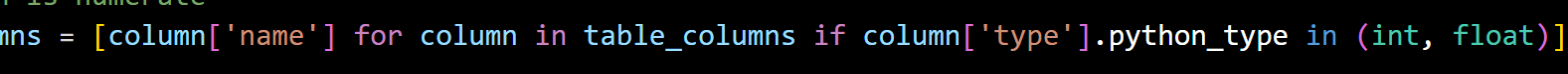
不允許Null的檢查

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

數字不能是文字

將該是數字但不是數字的變成NULL，注意要放在check not null之前跑，免得出錯



一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

Foreign key檢查

…待補充

長度限制

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

換行換空格

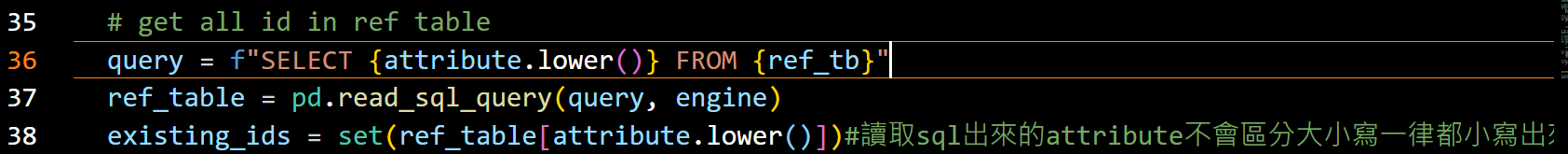
一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

表四 需要check foreign key是否存在於另外三個表，還需要check由med\_id source\_id sample\_id組合出來的primary key在excel中以及在mssql中 當前有無重複。



用這樣取得在mssql那些表中有那些id存在，然後跟現在excel中對比不存在的踢出去

valid\_rows = [row for \_, row in excel\_sheet.iterrows() if tuple(row[['Med\_id', 'Source\_id', 'Sample\_id']]) not in invalid\_rows\_set and tuple(row[['Med\_id', 'Source\_id', 'Sample\_id']]) not in duplicate\_set]

用這個處理primary key把重複的或不符合的row刪除

Jupiter：

按run之前記得先儲存

實測如下才算成功，否則要用prompt跑.py檔debug

一張含有 文字, 字型, 行, 螢幕擷取畫面 的圖片

自動產生的描述