'''

# 单页面

import requests

from bs4 import BeautifulSoup

import pandas as pd

# 请求URL

url = 'http://www.12365auto.com/zlts/0-0-0-0-0-0\_0-0-0-0-0-0-0-1.shtml'

# 得到页面的内容

headers={'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/74.0.3729.131 Safari/537.36'}

html=requests.get(url,headers=headers,timeout=10)

content = html.text

# 通过content创建BeautifulSoup对象

soup = BeautifulSoup(content, 'html.parser', from\_encoding='utf-8')

def analysis(soup):

# 找到完整的投诉信息框

temp = soup.find('div',class\_="tslb\_b")

# 创建DataFrame

df = pd.DataFrame(columns = ['id', 'brand', 'car\_model', 'type', 'desc', 'problem', 'datetime', 'status'])

tr\_list = temp.find\_all('tr')

# ToDo：提取汽车投诉信息

for tr in tr\_list:

temp={}

td\_list=tr.find\_all('td')

if len(td\_list)>0:

id,brand,car\_model,type,desc,problem,datetime,status=td\_list[0].text,td\_list[1].text,td\_list[2].text,td\_list[3].text,td\_list[4].text,td\_list[5].text,td\_list[6].text,td\_list[7].text

temp['id'],temp['brand'], temp['car\_model'],temp['type'], temp['desc'], temp['problem'], temp['datetime'], temp['status']= id,brand,car\_model,type,desc,problem,datetime,status

df=df.append(temp,ignore\_index=True)

return df

df=analysis(soup)

print(df)

df.to\_csv('car\_complain.csv', index=False)

'''

#20 页

import requests

from bs4 import BeautifulSoup

import pandas as pd

# 请求URL

def get\_page\_content(request\_url):

#url = 'http://www.12365auto.com/zlts/0-0-0-0-0-0\_0-0-0-0-0-0-0-1.shtml'

# 得到页面的内容

headers={'user-agent': 'Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/74.0.3729.131 Safari/537.36'}

html=requests.get(request\_url,headers=headers,timeout=10)

content = html.text

# 通过content创建BeautifulSoup对象

soup = BeautifulSoup(content, 'html.parser', from\_encoding='utf-8')

return soup

def analysis(soup):

# 找到完整的投诉信息框

temp = soup.find('div',class\_="tslb\_b")

# 创建DataFrame

df = pd.DataFrame(columns = ['id', 'brand', 'car\_model', 'type', 'desc', 'problem', 'datetime', 'status'])

tr\_list = temp.find\_all('tr')

# ToDo：提取汽车投诉信息

for tr in tr\_list:

temp={}

td\_list=tr.find\_all('td')

if len(td\_list)>0:

id,brand,car\_model,type,desc,problem,datetime,status=td\_list[0].text,td\_list[1].text,td\_list[2].text,td\_list[3].text,td\_list[4].text,td\_list[5].text,td\_list[6].text,td\_list[7].text

temp['id'],temp['brand'], temp['car\_model'],temp['type'], temp['desc'], temp['problem'], temp['datetime'], temp['status']= id,brand,car\_model,type,desc,problem,datetime,status

df=df.append(temp,ignore\_index=True)

return df

page\_num=20

base\_url= 'http://www.12365auto.com/zlts/0-0-0-0-0-0\_0-0-0-0-0-0-0-'

result=pd.DataFrame(columns = ['id', 'brand', 'car\_model', 'type', 'desc', 'problem', 'datetime', 'status'])

for i in range(page\_num):

request\_url=base\_url+str(i+1)+'.shtml'

soup= get\_page\_content(request\_url)

df=analysis(soup)

print(df)

result=result.append(df)

result.to\_csv('result.csv', index=False)