# Amazon WebScraping Project

```
#Import Libraries
from bs4 import BeautifulSoup
import requests
import time
import datetime
import smtplib

# Connect to Website and pull in data

URL =
'https://www.amazon.com/Funny-Data-Systems-Business-Analyst/dp/B07FNW9
FGJ/ref=sr_1_3?dchild=1&keywords=data%2Banalyst
%2Btshirt&qid=1626655184&sr=8-3&customId=B0752XJYNL&th=1'

headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0
Safari/537.36", "Accept-Encoding":"gzip, deflate",
"Accept":"text/html,application/xhtml+xml,application/xml;q=0.9,*/
*;q=0.8", "DNT":"1","Connection":"close", "Upgrade-Insecure-
Requests":"1"}
#we will get from website
page = requests.get(URL, headers=headers)#getting data

soup1 = BeautifulSoup(page.content, "html.parser")
soup2 = BeautifulSoup(soup1.prettify(),"html.parser")
#remember id which u want to pull
title=soup2.find(id='productTitle').get_text()
#pulling price
price_symbol = soup2.find(class_='a-price-
symbol').get_text(strip=True)
price_whole = soup2.find(class_='a-price-whole').get_text(strip=True)
price_fraction = soup2.find(class_='a-price-
fraction').get_text(strip=True)
price = f'{price_symbol}{price_whole}{price_fraction}'

print(title)
print(price)


                    Got Data Funny Business Data Analyst T-Shirt

$19.99

title=title.strip()
price=price.strip()[1:]
```

```
print(title)
print(price)
```

```
Got Data Funny Business Data Analyst T-Shirt
19.99
```

```
import datetime

today=datetime.date.today()
print(today)
```

```
2024-12-07
```

```
# If uou want to try sending yourself an email (just for fun) when a
price hits below a certain level you can try it
# out with this script

def send_mail():
    server = smtplib.SMTP_SSL('smtp.gmail.com',465)
    server.ehlo()
    #server.starttls()
    server.ehlo()
    server.login('AlexTheAnalyst95@gmail.com','xxxxxxxxxxxxxx')

    subject = "The Shirt you want is below $15! Now is your chance to
buy!"
    body = "Alex, This is the moment we have been waiting for. Now is
your chance to pick up the shirt of your dreams. Don't mess it up!
Link here: https://www.amazon.com/Funny-Data-Systems-Business-
Analyst/dp/B07FNW9FGJ/ref=sr_1_3?
dchild=1&keywords=data+analyst+tshirt&qid=1626655184&sr=8-3"

    msg = f"Subject: {subject}\n\n{body}"

    server.sendmail(
        'AlexTheAnalyst95@gmail.com',
        msg

    )

import csv

header=['Title','Price','Date']
data=[title,price,today]
type(data) #list type

with
open('AmazonWebScraperDataset.csv','w',newline='',encoding='UTF8') as
f:
    writer=csv.writer(f)
    writer.writerow(header)
```

```
    writer.writerow(data)


#file will come to C:\Users\User


import pandas as pd
df=pd.read_csv(r"C:\Users\User\AmazonWebScraperDataset.csv")
print(df)
```

```
                                              Title  Price        Date
0  Got Data Funny Business Data Analyst T-Shirt  19.99  2024-12-07
1  Got Data Funny Business Data Analyst T-Shirt  19.99  2024-12-07
2  Got Data Funny Business Data Analyst T-Shirt  19.99  2024-12-07
```

```
#we are appeding the data

with
open('AmazonWebScraperDataset.csv','a+',newline='',encoding='UTF8') as
f:
    writer=csv.writer(f)
    #writer.writerow(header)
    writer.writerow(data)

def check_price():
        URL = 'https://www.amazon.com/Funny-Data-Systems-Business-
Analyst/dp/B07FNW9FGJ/ref=sr_1_3?dchild=1&keywords=data%2Banalyst
%2Btshirt&qid=1626655184&sr=8-3&customId=B0752XJYNL&th=1'

        headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64;
x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0
Safari/537.36", "Accept-Encoding":"gzip, deflate",
"Accept":"text/html,application/xhtml+xml,application/xml;q=0.9,*/
*;q=0.8", "DNT":"1","Connection":"close", "Upgrade-Insecure-
Requests":"1"}
        #we will get from website
        page = requests.get(URL, headers=headers)#getting data

        soup1 = BeautifulSoup(page.content, "html.parser")
        soup2 = BeautifulSoup(soup1.prettify(),"html.parser")
        #remember id which u want to pull
        title=soup2.find(id='productTitle').get_text()
        #pulling price
        price_symbol = soup2.find(class_='a-price-
symbol').get_text(strip=True)
        price_whole = soup2.find(class_='a-price-
whole').get_text(strip=True)
        price_fraction = soup2.find(class_='a-price-
fraction').get_text(strip=True)
        price = f'{price_symbol}{price_whole}{price_fraction}'
        title=title.strip()
```

```
        price=price.strip()[1:]

        import datetime
        today=datetime.date.today()

        import csv

        header=['Title','Price','Date']
        data=[title,price,today]

        with
open('AmazonWebScraperDataset.csv','a+',newline='',encoding='UTF8') as
f:
                writer=csv.writer(f)
                #writer.writerow(header)
                writer.writerow(data)

        if(price<14):
            send_mail()


while(True):
    check_price()
    time.sleep(5)
```

```
---------------------------------------------------------------------------
-----
KeyboardInterrupt                         Traceback (most recent call
last)
Cell In[109], line 3
      1 while(True):
      2     check_price()
----> 3     time.sleep(5)

KeyboardInterrupt:
```

```
import pandas as pd
df=pd.read_csv(r"C:\Users\User\AmazonWebScraperDataset.csv")
print(df)
```

```
                                              Title  Price        Date
0  Got Data Funny Business Data Analyst T-Shirt  19.99  2024-12-07
1  Got Data Funny Business Data Analyst T-Shirt  19.99  2024-12-07
2  Got Data Funny Business Data Analyst T-Shirt  19.99  2024-12-07
3  Got Data Funny Business Data Analyst T-Shirt  14.97  2024-12-07
4  Got Data Funny Business Data Analyst T-Shirt  14.97  2024-12-07
5  Got Data Funny Business Data Analyst T-Shirt  14.97  2024-12-07
6  Got Data Funny Business Data Analyst T-Shirt  14.97  2024-12-07
7  Got Data Funny Business Data Analyst T-Shirt  14.97  2024-12-07
```

```
8  Got Data Funny Business Data Analyst T-Shirt  14.97  2024-12-07
9  Got Data Funny Business Data Analyst T-Shirt  14.97  2024-12-07
```