Code:

from selenium import webdriver

from selenium.webdriver.firefox.options import Options

from bs4 import BeautifulSoup

import time

import os

import pandas as pd

import xlwt

from xlwt import Workbook

book\_ASIN\_list = []

options = Options()

options.headless = True

base\_path = os.path.realpath('.')

geckodriver\_path = base\_path + r'\geckodriver.exe'

options.binary\_location = r'C:\Program Files\Mozilla Firefox\firefox.exe'

driver = webdriver.Firefox(options=options, executable\_path = geckodriver\_path)

print("Web Driver Identified")

read\_asin = pd.read\_csv("numbers.csv")

print("CSV data read")

wb = Workbook()

print("WorkBook Added")

sheet1 = wb.add\_sheet('DESC')

print("Sheet Added")

print("Starting reading of desc")

for index, row in read\_asin.iterrows():

    print("[x] Index : " + str(index))

    ASIN = str(row["ASIN"])

    print("[x]Getting data for ASIN number" + ASIN)

    if len(ASIN) == 9:

        ASIN = "0" + ASIN

    URL = 'https://www.amazon.in/dp/' + ASIN

    start = time.time()

    try:

        print("[x]Visiting url for " + ASIN)

        driver.get(URL)

        driver.switch\_to.frame("bookDesc\_iframe")

        source\_code = driver.page\_source

        inspect\_code = driver.execute\_script("return document.documentElement.innerHTML;")

        soup = BeautifulSoup(inspect\_code, "html.parser")

        description\_received = soup.find("div" , {"id":"iframeContent"}).text

        sheet1.write(index,0,ASIN)

        sheet1.write(index,1,description\_received)

        print("[x]Description Writing done for " + ASIN)

    except:

        book\_ASIN\_list.append(ASIN)

        print("[E]ERROR RECEIVED FOR " + ASIN)

    end = time.time()

    print(f"Runtime of the program is {end - start}")

driver.quit()

print("Driver Closed")

print("Saving file")

wb.save('desc\_data.xls')

print("File Saved")

print("Saving ASIN to delete")

dict = {"ASIN" : book\_ASIN\_list}

df = pd.DataFrame(dict)

df.to\_csv('amz\_todel2.csv')

print("Saved file")

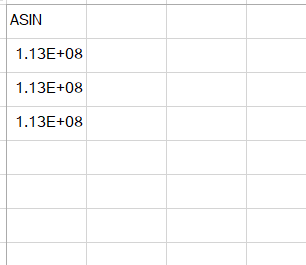
print("Closing Program")

Output:

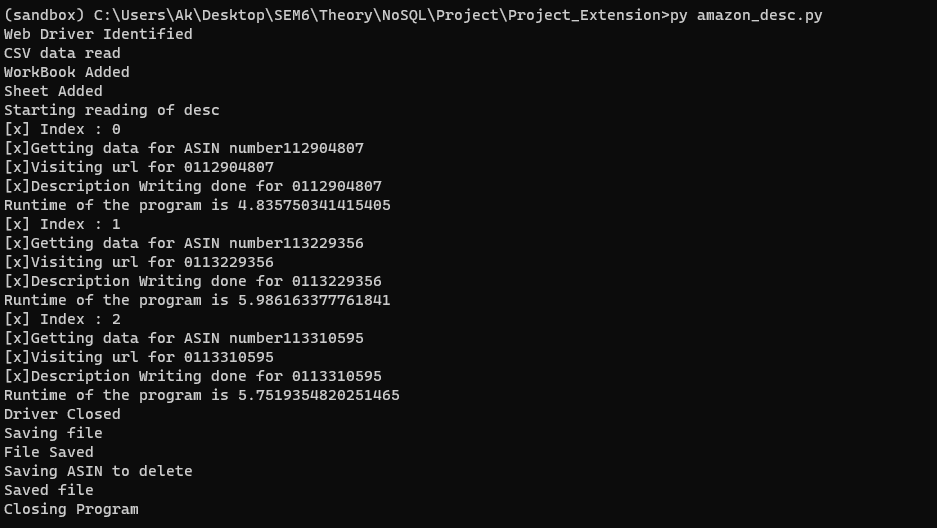
For output purpose we run the code for two ASIN numbers. Upon successful completion of the code two files are created. One containing the ASIN and description and the other ASIN for which the description data is not available on the Amazon webpage.

For the program to work the program needs a numbers.csv file which contains all the ASIN numbers for which description is needed to be found.

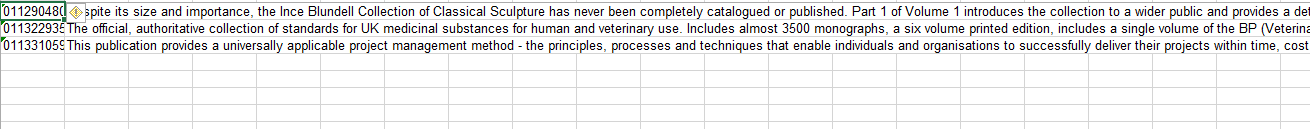
Numbers.csv



Code Output:



As seen from the output the running time for each ASIN number is near to 5 sec. Now we see the two files created.



We can see that it created a excel file with ASIN numbers and their corresponding descriptions. Since we found description for all the file containing ASIN numbers for which there was some error to be empty.

