ASSIGNMENT-12

REDDIVARI SAI SARAN

G01142501

1 )Write Python code using BeautifulSoup package to generate the CSV output file:

# import matplotlib.pyplot as plt

# import urllib

# from bs4 import BeautifulSoup

# import requests

# import pandas as pd

# import csv

# import xlrd

# url = requests.get("https://repo.vse.gmu.edu/ait/AIT580/580books.html")

# soup = BeautifulSoup(url.content, 'html.parser')

# listitems = soup.find\_all('li')

# filename = "listofbooks\_saran.csv"

# f = open(filename, "w")

# headers = "title, author, publisher, release\n"

# f.write(headers)

# for entry in listitems:

# 

# title = entry.a.booktitle.text

# title = title.replace(",", "|")

# author = entry.author.text

# author = author.replace(",", "|")

# publisher = entry.publisher.text

# publisher = publisher.replace(",", "|")

# release = entry.release.text

# release = release.replace(",", "|")

# 

# f.write(title + "," + author + "," + publisher + "," + release + "\n")

# f.close()

# for record in soup.findAll('tr'):

# tbltxt = ""

# for data in record.findAll('td'):

# tbltxt = tbltxt + data.text + ","

# print(tbltxt)

# print(tbltxt[0:-1])

**output:**

Title

,Author(S)

,Publisher

,PubDate

,

Title

,Author(S)

,Publisher

,PubDate

Learning Python,Mark Lutz,O'Reilly Media,June 2013,

Learning Python,Mark Lutz,O'Reilly Media,June 2013

Web Scraping with Python,Ryan Mitchell,O'Reilly Media,November 2017,

Web Scraping with Python,Ryan Mitchell,O'Reilly Media,November 2017

Python Coding for Kids,Don Wilcher,Maker Media, Inc,September 2017,

Python Coding for Kids,Don Wilcher,Maker Media, Inc,September 2017

Python Machine Learning By Example,Yuxi Liu,Packt Publishing,May 2017,

Python Machine Learning By Example,Yuxi Liu,Packt Publishing,May 2017

Python Data Structures and Algorithms,Benjamin Baka,Packt Publishing,May 2017,

Python Data Structures and Algorithms,Benjamin Baka,Packt Publishing,May 2017

Python in a Nutshell,Alex Martelli, Anna Ravenscroft,O'Reilly Media,April 2017,

Python in a Nutshell,Alex Martelli, Anna Ravenscroft,O'Reilly Media,April 2017

Invent Your Own Computer Games with Python,Al Sweigart,No Starch Press,December 2016,

Invent Your Own Computer Games with Python,Al Sweigart,No Starch Press,December 2016

Head First Python,Paul Barry,O'Reilly Media,November 2016,

Head First Python,Paul Barry,O'Reilly Media,November 2016

Integrating Python with Leading Computer Forensics Platforms,Chet Hosmer,Elsevier / Syngress,September 2016,

Integrating Python with Leading Computer Forensics Platforms,Chet Hosmer,Elsevier / Syngress,September 2016

Data Science Essentials in Python,Dmitry Zinoviev,Pragmatic Bookshelf,August 2016,

Data Science Essentials in Python,Dmitry Zinoviev,Pragmatic Bookshelf,August 2016

1. Read the output csv file from part 1 and write python code to answer the following question. You can use Pandas Package in python. (20 points)

a) Print the frequency count for publishers (that is, how many books for each publisher).

# data=pd.read\_csv('C:/Users/saran/OneDrive/Desktop/listofbooksait580.csv')

# data1=data[' publisher'].value\_counts()

# print(data1)

**output:**

O'Reilly Media 4

Packt Publishing 2

No Starch Press 1

Pragmatic Bookshelf 1

Maker Media| Inc 1

Elsevier / Syngress 1

Name: publisher, dtype: int64

b) Visualize the output of (a) using a chart of your own choice.

# data1.plot()

# data1.plot(kind='bar')

# plt.xlabel('PUBLISHER')

# plt.ylabel('FREQUENCY')

# plt.title('visulization for publisher')

# plt.show()

**output:**

**A screenshot of a cell phone

Description generated with very high confidence**

c) Print the frequency count for year of publication.

# df=data.iloc[:,3].str.split(' ', expand=True)

# print(df)

# data2=df[1].value\_counts()

# print(data2)

**OUTPUT:**

0 1

0 June 2013

1 November 2017

2 September 2017

3 May 2017

4 May 2017

5 April 2017

6 December 2016

7 November 2016

8 September 2016

9 August 2016

2017 5

2016 4

2013 1

Name: 1, dtype: int64

d) Visualize the output of (b) using a chart of your own choice.

# data2.plot(kind='bar')

# plt.ylabel('Frequency')

# plt.xlabel('year of Publication')

# plt.title('Visualize for year of Publication')

# plt.show()

OUTPUT:

**A screenshot of a cell phone

Description generated with very high confidence**