**DAILY ASSESSMENT FORMAT**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date:** | **4 June 2020** | **Name:** | **Shreya poojary** |
| **Course:** | **python** | **USN:** | **4al16ec074** |
| **Topic:** | **udemy** | **Semester & Section:** | **8-B** |
| **Github Repository:** | **Shreya-test** |  |  |

|  |
| --- |
| **FORENOON SESSION DETAILS** |
| **Image of session** |
|  |
| **REPORT**  The large property listing companies target an entire nation and work on millions of properties. But in case you are a real estate agent, or if you are setting up shop and targeting a specific state or region, it is better that instead of trying to gather data yourself, you scrape it from a major real estate listing website.  #!/usr/bin/python  # -\*- coding: utf-8 -\*-  import urllib.request  import urllib.parse  import urllib.error  from bs4 import BeautifulSoup  import ssl  import json  import ast  import os  from urllib.request import Request, urlopen  # For ignoring SSL certificate errors  ctx = ssl.create\_default\_context()  ctx.check\_hostname = False  ctx.verify\_mode = ssl.CERT\_NONE  # Input from user  url = input('Enter Zillow House Listing Url- ')  # Making the website believe that you are accessing it using a mozilla browser  req = Request(url, headers={'User-Agent': 'Mozilla/5.0'})  webpage = urlopen(req).read()  # Creating a BeautifulSoup object of the html page for easy extraction of data.  soup = BeautifulSoup(webpage, 'html.parser')  html = soup.prettify('utf-8')  property\_json = {}  property\_json['Details\_Broad'] = {}  property\_json['Address'] = {}  # Extract Title of the property listing  for title in soup.findAll('title'):  property\_json['Title'] = title.text.strip()  break  for meta in soup.findAll('meta', attrs={'name': 'description'}):  property\_json['Detail\_Short'] = meta['content'].strip()  for div in soup.findAll('div', attrs={'class': 'character-count-truncated'}):  property\_json['Details\_Broad']['Description'] = div.text.strip()  for (i, script) in enumerate(soup.findAll('script',  attrs={'type': 'application/ld+json'})):  if i == 0:  json\_data = json.loads(script.text)  property\_json['Details\_Broad']['Number of Rooms'] = json\_data['numberOfRooms']  property\_json['Details\_Broad']['Floor Size (in sqft)'] = json\_data['floorSize']['value']  property\_json['Address']['Street'] = json\_data['address']['streetAddress']  property\_json['Address']['Locality'] = json\_data['address']['addressLocality']  property\_json['Address']['Region'] = json\_data['address']['addressRegion']  property\_json['Address']['Postal Code'] = json\_data['address']['postalCode']  if i == 1:  json\_data = json.loads(script.text)  property\_json['Price in $'] = json\_data['offers']['price']  property\_json['Image'] = json\_data['image']  break  with open('data.json', 'w') as outfile:  json.dump(property\_json, outfile, indent=4)  with open('output\_file.html', 'wb') as file:  file.write(html)  print ('----------Extraction of data is complete. Check json file.----------') |
|  |