

DSC540-T301_2237-1_Samanta_Rajib_week_7&8

July 30, 2023

```
[101]: # Class : DSC540-T301 Data Preparation (2237-1)
# Name : Rajib Samanta
# Assignment : Week 7 & 8 Excercises
## Assignment 1: Data Wrangling with Python: Activity 9 & 10
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
[102]: # Data Wrangling with Python: Activity 9, page 294
# Extracting the Top 100 eBooks from Gutenberg
# 1. Import the necessary libraries, including regex and BeautifulSoup:
import urllib.request, urllib.parse, urllib.error
import requests
from bs4 import BeautifulSoup
import ssl
import re
```

```
[103]: # 2. Check the SSL certificate
ctx = ssl.create_default_context()
ctx.check_hostname = False
ctx.verify_mode = ssl.CERT_NONE
```

```
[104]: # 3 Read the HTML from the URL:# Read the HTML from the URL and pass on to
↳BeautifulSoup
top100url = 'https://www.gutenberg.org/browse/scores/top'
response = requests.get(top100url)
response
```

```
[104]: <Response [200]>
```

```
[105]: # 4 Write a small function to check the status of the web request:
def status_check(r):
    if r.status_code==200:
        print("Success!")
        return 1
    else:
        print("Failed!")
```

```

        return -1
    # Check the status of response:
    print('The output of URL call Status : ' )
    status_check(response)

```

The output of URL call Status :
Success!

[105]: 1

```

[106]: # 5 Decode the response and pass it on to BeautifulSoup for HTML parsing:
        contents = response.content.decode(response.encoding) # get the content from
        ↳ the reposne
        soup = BeautifulSoup(contents, 'html.parser') # parse the content with html
        ↳ parser as this is html web based page
        # Check the data after HTML parsing
        soup

```

[106]: <!DOCTYPE html>

```

<html class="client-nojs" dir="ltr" lang="en">
<head>
<meta charset="utf-8"/>
<title>Top 100 | Project Gutenberg</title>
<link href="/gutenberg/style.css?v=1.1" rel="stylesheet"/>
<link href="/gutenberg/collapsible.css?1.1" rel="stylesheet"/>
<link href="/gutenberg/new_nav.css?v=1.321231" rel="stylesheet"/>
<link href="/gutenberg/pg-desktop-one.css" rel="stylesheet"/>
<meta content="width=device-width, initial-scale=1" name="viewport"/>
<meta content="books, ebooks, free, kindle, android, iphone, ipad"
name="keywords">
<meta content="wucOEvsNj5kP3Ts_360fP64laakK-1mVTg-ptrGC9io" name="google-site-
verification"/>
<meta content="4WNaCljsE-A82vP_ih2H_UqXZvM" name="alexaVerifyID"/>
<link href="https://www.gnu.org/copyleft/fdl.html" rel="copyright">
<link href="/gutenberg/favicon.ico?v=1.1" rel="shortcut icon">
<meta content="Project Gutenberg" property="og:title"/>
<meta content="website" property="og:type"/>
<meta content="https://www.gutenberg.org/" property="og:url"/>
<meta content="Project Gutenberg is a library of free eBooks."
property="og:description"/>
<meta content="615269807" property="fb:admins"/>
<meta content="115319388529183" property="fb:app_id"/>
<meta content="Project Gutenberg" property="og:site_name"/>
<meta content="https://www.gutenberg.org/gutenberg/pg-logo-144x144.png"
property="og:image"/>
</link></link></meta></head>
<body>

```

```

<div class="container"><!-- start body --><nav>
<!--<div id="main_logo"> -->
<a class="no-hover" href="/" id="main_logo">

</a>
<!--      </div>-->
<div id="menu">
<label for="tm" id="toggle-menu">Menu<span class="drop-icon"> </span></label>
<input id="tm" type="checkbox"/>
<ul class="main-menu cf">
<li>
<a href="/about/">About
      <span class="drop-icon"> </span>
</a>
<label class="drop-icon" for="sm0" title="Toggle Drop-down"> </label>
<input id="sm0" type="checkbox"/>
<ul class="sub-menu">
<li><a href="/about/">About Project Gutenberg</a></li>
<li><a href="/policy/collection_development.html">Collection
Development</a></li>
<li><a href="/about/contact_information.html">Contact Us</a></li>
<li><a href="/about/background/">History & Philosophy</a></li>
<li><a href="/policy/permission.html">Permissions & License</a></li>
<li><a href="/policy/privacy_policy.html">Privacy Policy</a></li>
<li><a href="/policy/terms_of_use.html">Terms of Use</a></li>
</ul>
</li>
<li>
<a href="/ebooks/">Search and Browse
      <span class="drop-icon"> </span>
</a>
<label class="drop-icon" for="sm8" title="Toggle Drop-down"> </label>
<input id="sm8" type="checkbox"/>
<ul class="sub-menu">
<li><a href="/ebooks/">Book Search</a></li>
<li><a href="/ebooks/bookshelf/">Bookshelves</a></li>
<li><a href="/browse/scores/top">Frequently Downloaded</a></li>
<li><a href="/ebooks/offline_catalogs.html">Offline Catalogs</a></li>
</ul>
</li>
<li>
<a href="/help/">Help
      <span class="drop-icon"> </span>
</a>
<label class="drop-icon" for="sm3" title="Toggle Drop-down"> </label>
<input id="sm3" type="checkbox"/>

```

```

<ul class="sub-menu">
<li><a href="/help/">All help topics </a></li>
<li><a href="/help/copyright.html">Copyright How-To</a></li>
<li><a href="/help/errata.html">Errata, Fixes and Bug Reports</a></li>
<li><a href="/help/file_formats.html">File Formats</a></li>
<li><a href="/help/faq.html">Frequently Asked Questions</a></li>
<li><a href="/policy/">Policies </a></li>
<li><a href="/help/public_domain_ebook_submission.html">Public Domain eBook
Submission</a></li>
<li><a href="/help/submitting_your_own_work.html">Submitting Your Own
Work</a></li>
<li><a href="/help/mobile.html">Tablets, Phones and eReaders</a></li>
<li><a href="/attic/">The Attic </a></li>
</ul>
</li>
<li><a href="/donate/">Donate</a></li>
</ul>
</div>
<div class="donate">
<div class="searchbox">
<form accept-charset="utf-8" action="/ebooks/search/" class="searchbox"
enctype="multipart/form-data" method="get">
<input class="searchInput" id="menu-book-search" maxlength="80" name="query"
placeholder=" Quick search" size="20" tabindex="20" title="Quick search"
type="text" value=""/>
<input name="submit_search" style="vertical-align:middle;" type="submit"
value="Go!"/>
</form>
</div>
<form action="https://www.paypal.com/cgi-bin/webscr" class="donatelink"
method="post" target="new">
<p><a href="/donate/">Donation</a></p>
<input name="cmd" type="hidden" value="_s-xclick"/>
<input name="hosted_button_id" type="hidden" value="XKAL6BZL3YPSN"/>
<input alt="Donate via PayPal" class="donbtn" name="submit"
src="/pics/en_US.gif" type="image"/>
</form>
</div>
</nav>
<div class="page_content"><!-- start content -->
<h1>Frequently Viewed or Downloaded</h1>
<p>These listings are based on the number of times each eBook gets downloaded.
Multiple downloads from the same Internet address on the same day count as
one download, and addresses that download more than 100 eBooks in a day are
considered robots and are not counted.</p>
<table>
<caption>Downloaded Books</caption>

```

2023-07-28	224074
last 7 days	1501104
last 30 days	5913305

- [Top 100 EBooks yesterday](#)
- [Top 100 Authors yesterday](#)
- [Top 100 EBooks last 7 days](#)
- [Top 100 Authors last 7 days](#)
- [Top 100 EBooks last 30 days](#)
- [Top 100 Authors last 30 days](#)

Top 100 EBooks yesterday

- [Romeo and Juliet by William Shakespeare \(2605\)](/ebooks/1513)
- [Moby Dick; Or, The Whale by Herman Melville \(2464\)](/ebooks/2701)
- [A Room with a View by E. M. Forster \(2237\)](/ebooks/2641)
- [Middlemarch by George Eliot \(2198\)](/ebooks/145)
- [Little Women; Or, Meg, Jo, Beth, and Amy by Louisa May Alcott \(2102\)](/ebooks/37106)
- [The Complete Works of William Shakespeare by William Shakespeare \(2007\)](/ebooks/100)
- [The Blue Castle: a novel by L. M. Montgomery \(1984\)](/ebooks/67979)
- [The Enchanted April by Elizabeth Von Arnim \(1979\)](/ebooks/16389)
- [Cranford by Elizabeth Cleghorn Gaskell \(1849\)](/ebooks/394)
- [The Adventures of Ferdinand Count Fathom - Complete by T. Smollett \(1841\)](/ebooks/6761)
- [The Expedition of Humphry Clinker by T. Smollett \(1807\)](/ebooks/2160)
- [The Adventures of Roderick Random by T. Smollett \(1785\)](/ebooks/4085)
- [Anna Karenina by graf Leo Tolstoy \(1783\)](/ebooks/1399)
- [Twenty Years After by Alexandre Dumas \(1750\)](/ebooks/1259)
- [History of Tom Jones, a Foundling by Henry Fielding \(1736\)](/ebooks/6593)
- [My Life - Volume 1 by Richard Wagner \(1705\)](/ebooks/5197)
- [Pride and Prejudice by Jane Austen \(1519\)](/ebooks/1342)
- [Alice's Adventures in Wonderland by Lewis Carroll \(878\)](/ebooks/11)
- [Dead letters by Maurice Baring \(831\)](/ebooks/71285)
- [The Count of Monte Cristo, Illustrated by Alexandre](/ebooks/1184)

Dumas (810)
 Frankenstein; Or, The Modern Prometheus by Mary Wollstonecraft Shelley (694)
 Richard the Lion Heart by Kate Norgate (634)
 Oliver Ellis : by James Grant (620)
 A Tale of Two Cities by Charles Dickens (610)
 Dracula by Bram Stoker (570)
 The Adventures of Sherlock Holmes by Arthur Conan Doyle (552)
 The House of Egremont : by Molly Elliot Seawell (536)
 The Picture of Dorian Gray by Oscar Wilde (519)
 The Report on Unidentified Flying Objects by Edward J. Ruppelt (457)
 The Brothers Karamazov by Fyodor Dostoyevsky (443)
 One Hundred Proofs That the Earth Is Not a Globe by William Carpenter (438)
 The Adventures of Tom Sawyer, Complete by Mark Twain (394)
 Criminality and economic conditions by Willem Adriaan Bongers (393)
 Grimms' Fairy Tales by Jacob Grimm and Wilhelm Grimm (389)
 A Child's History of the World by V. M. Hillyer (380)
 Metamorphosis by Franz Kafka (379)
 War and Peace by graf Leo Tolstoy (375)
 The Great Gatsby by F. Scott Fitzgerald (370)
 The Kama Sutra of Vatsyayana by Vatsyayana (369)
 Crime and Punishment by Fyodor Dostoyevsky (366)
 Ulysses by James Joyce (366)
 The Prince by Niccolò Machiavelli (364)
 Adventures of Huckleberry Finn by Mark Twain (348)
 The Romance of Lust: A classic Victorian erotic novel by Anonymous (345)
 The Prophet by Kahlil Gibran (343)
 Thus Spake Zarathustra: A Book for All and None by Friedrich Wilhelm Nietzsche (335)
 A Modest Proposal by Jonathan Swift (329)
 Moby Multiple Language Lists of Common Words by Grady

Ward (321)
 Meditations by Emperor of Rome Marcus Aurelius
 (319)
 Jane Eyre: An Autobiography by Charlotte Brontë
 (305)
 Beyond Good and Evil by Friedrich Wilhelm Nietzsche
 (299)
 Dombey and Son by Charles Dickens (297)
 A Study in Scarlet by Arthur Conan Doyle
 (295)
 Great Expectations by Charles Dickens (290)
 The Importance of Being Earnest: A Trivial Comedy for
 Serious People by Oscar Wilde (290)
 The Yellow Wallpaper by Charlotte Perkins Gilman
 (289)
 The Iliad by Homer (283)
 The Strange Case of Dr. Jekyll and Mr. Hyde by Robert
 Louis Stevenson (282)
 The Odyssey by Homer (281)
 Tractatus Logico-Philosophicus by Ludwig Wittgenstein
 (278)
 Calculus Made Easy by Silvanus P. Thompson
 (275)
 A Christmas Carol in Prose; Being a Ghost Story of
 Christmas by Charles Dickens (267)
 A Doll's House : a play by Henrik Ibsen
 (262)
 Don Quixote by Miguel de Cervantes Saavedra
 (262)
 Anne of Green Gables by L. M. Montgomery
 (252)
 Wuthering Heights by Emily Brontë (245)
 Walden, and On The Duty Of Civil Disobedience by Henry
 David Thoreau (244)
 The Republic by Plato (238)
 The Problems of Philosophy by Bertrand Russell
 (237)
 The Enchiridion by Epictetus (237)
 The slang dictionary : by John Camden Hotten
 (236)
 Emma by Jane Austen (236)
 Notes from the Underground by Fyodor Dostoyevsky
 (236)
 The Wonderful Wizard of Oz by L. Frank Baum
 (236)
 Lexicon Latinum : by Franz Wagner (230)
 Treasure Island by Robert Louis Stevenson

(227)
 The divine comedy by Dante Alighieri (224)
 Oliver Twist by Charles Dickens (222)
 Little Women by Louisa May Alcott (219)
 On the Duty of Civil Disobedience by Henry David Thoreau (219)
 Les Misérables by Victor Hugo (218)
 The War of the Worlds by H. G. Wells (214)
 Winnie-the-Pooh by A. A. Milne (213)
 The Rámáyan of Válmiki, translated into English verse by Valmiki (213)
 The King James Version of the Bible (207)
 The King in Yellow by Robert W. Chambers (207)
 David Copperfield by Charles Dickens (202)
 Peter Pan by J. M. Barrie (201)
 The Scarlet Letter by Nathaniel Hawthorne (193)
 The Art of War by active 6th century B.C. Sunzi (190)
 Sense and Sensibility by Jane Austen (188)
 Heart of Darkness by Joseph Conrad (185)
 Dubliners by James Joyce (176)
 Demonology and Devil-lore by Moncure Daniel Conway (173)
 Carmilla by Joseph Sheridan Le Fanu (170)
 Josefine Mutzenbacher by Felix Salten (167)
 The murder of Roger Ackroyd by Agatha Christie (165)
 The Jungle Book by Rudyard Kipling (163)
 The Mysterious Affair at Styles by Agatha Christie (162)
 Essays of Michel de Montaigne - Complete by Michel de Montaigne (161)

 <div class="padded">

 Top 100 EBooks yesterday
 Top 100 Authors yesterday
 Top 100 EBooks last 7 days
 Top 100 Authors last 7 days
 Top 100 EBooks last 30 days
 Top 100 Authors last 30 days

 </div>
 <h2 id="authors-last1">Top 100 Authors yesterday</h2>

Shakespeare, William (5889)
 Smollett, T. (Tobias) (5499)
 Dumas, Alexandre (3091)
 Garnett, Constance (3037)
 Tolstoy, Leo, graf (2833)
 Dickens, Charles (2812)
 Montgomery, L. M. (Lucy Maud) (2677)
 Melville, Herman (2631)
 Austen, Jane (2594)
 Alcott, Louisa May (2562)
 Eliot, George (2395)
 Forster, E. M. (Edward Morgan) (2390)
 Merrill, Frank T. (2172)
 Twain, Mark (2124)
 Gaskell, Elizabeth Cleghorn (2021)
 Von Arnim, Elizabeth (1995)
 Brock, C. E. (Charles Edmund) (1870)
 Doyle, Arthur Conan (1860)
 Fielding, Henry (1785)
 Wagner, Richard (1745)
 Dostoyevsky, Fyodor (1720)
 Wilde, Oscar (1500)
 Carroll, Lewis (1462)
 Nietzsche, Friedrich Wilhelm (1404)
 Stevenson, Robert Louis (1055)
 Baring, Maurice (1000)
 Wells, H. G. (Herbert George) (980)
 Shelley, Mary Wollstonecraft (910)
 Homer (870)
 Baum, L. Frank (Lyman Frank) (830)
 Verne, Jules (824)
 Christie, Agatha (818)
 Plato (803)
 Lang, Andrew (782)
 Widger, David (778)
 Grant, James (776)
 Kipling, Rudyard (764)
 Jowett, Benjamin (747)
 Norgate, Kate (705)
 Stoker, Bram (700)
 Doré, Gustave (694)
 Seawell, Molly Elliot (680)
 Russell, Bertrand (671)

Joyce, James (654)
 Chesterton, G. K. (Gilbert Keith) (638)
 Balzac, Honoré de (617)
 Kafka, Franz (609)
 Relyea, C. M. (Charles M.) (599)
 Dante Alighieri (597)
 Poe, Edgar Allan (595)
 Conrad, Joseph (589)
 Thoreau, Henry David (588)
 Goethe, Johann Wolfgang von (571)
 Swift, Jonathan (566)
 Hugo, Victor (566)
 London, Jack (566)
 Hawthorne, Nathaniel (560)
 Burton, Richard Francis, Sir (560)
 Grimm, Wilhelm (547)
 Grimm, Jacob (547)
 Maude, Aylmer (535)
 Chekhov, Anton Pavlovich (533)
 Cervantes Saavedra, Miguel de (525)
 Scott, Walter (515)
 Burroughs, Edgar Rice (509)
 James, Henry (506)
 Maude, Louise (503)
 Fitzgerald, F. Scott (Francis Scott) (497)
 Wyllie, David (Translator) (495)
 Shaw, Bernard (491)
 Rizal, José (470)
 Butler, Samuel (462)
 Ibsen, Henrik (459)
 Ruppelt, Edward J. (457)
 Pope, Alexander (441)
 Carpenter, William (438)
 Potter, Beatrix (429)
 Defoe, Daniel (424)
 Barrie, J. M. (James Matthew) (419)
 Howard, Robert E. (Robert Ervin) (416)
 Machiavelli, Niccolò (415)
 Wodehouse, P. G. (Pelham Grenville) (413)
 Maupassant, Guy de (413)
 Morley, Henry (410)
 Brontë, Charlotte (402)

Common, Thomas (393)
 Horton, Henry P. (Henry Pomeroy)
 (393)
 Lindsey, Edward (393)
 Bonger, Willem Adriaan (393)
 Norcross, Frank H. (Frank Herbert)
 (393)
 Levy, Oscar (390)
 Marcus Aurelius, Emperor of Rome
 (381)
 Boog, Carle Michel (380)
 Jones, Mary Sherwood Wright
 (380)
 Hillyer, V. M. (Virgil Mores)
 (380)
 Kemble, E. W. (Edward Windsor)
 (378)
 Ward, Grady (375)
 Vatsyayana (373)
 Indrajit, Bhagavanlal (369)
 Bhide, Shivaram Parashuram (369)

 <div class="padded">

 Top 100 EBooks yesterday
 Top 100 Authors yesterday
 Top 100 EBooks last 7 days
 Top 100 Authors last 7 days
 Top 100 EBooks last 30 days
 Top 100 Authors last 30 days

 </div>
 <h2 id="books-last7">Top 100 EBooks last 7 days</h2>

 Down and Out in the Magic Kingdom by Cory Doctorow
 (29350)
 Romeo and Juliet by William Shakespeare
 (18449)
 Moby Dick; Or, The Whale by Herman Melville
 (17100)
 A Room with a View by E. M. Forster (15378)
 Middlemarch by George Eliot (14963)
 Little Women; Or, Meg, Jo, Beth, and Amy by Louisa
 May Alcott (14512)
 The Complete Works of William Shakespeare by William
 Shakespeare (14171)
 The Blue Castle: a novel by L. M. Montgomery

(13815)
 The Enchanted April by Elizabeth Von Arnim
 (13687)
 The Adventures of Ferdinand Count Fathom - Complete
 by T. Smollett (12896)
 Cranford by Elizabeth Cleghorn Gaskell
 (12835)
 The Expedition of Humphry Clinker by T. Smollett
 (12587)
 The Adventures of Roderick Random by T. Smollett
 (12526)
 Twenty Years After by Alexandre Dumas
 (12331)
 History of Tom Jones, a Foundling by Henry Fielding
 (12315)
 My Life - Volume 1 by Richard Wagner (11968)
 Pride and Prejudice by Jane Austen (10942)
 A Child's History of the World by V. M. Hillyer
 (6157)
 Alice's Adventures in Wonderland by Lewis Carroll
 (5837)
 Spoon River Anthology by Edgar Lee Masters
 (5517)
 Frankenstein; Or, The Modern Prometheus by Mary
 Wollstonecraft Shelley (5109)
 Dracula by Bram Stoker (4272)
 Anna Karenina by graf Leo Tolstoy (4097)
 The Adventures of Sherlock Holmes by Arthur Conan
 Doyle (4020)
 The Picture of Dorian Gray by Oscar Wilde
 (3884)
 A Tale of Two Cities by Charles Dickens (3761)
 The Count of Monte Cristo, Illustrated by Alexandre
 Dumas (3578)
 Mark Twain's Speeches by Mark Twain (3526)
 The Brothers Karamazov by Fyodor Dostoyevsky
 (3096)
 The Prince by Niccolò Machiavelli (2880)
 Metamorphosis by Franz Kafka (2873)
 War and Peace by graf Leo Tolstoy (2819)
 The Great Gatsby by F. Scott Fitzgerald
 (2803)
 Ulysses by James Joyce (2744)
 A Modest Proposal by Jonathan Swift (2682)
 The Truth about the Titanic by Archibald Gracie
 (2569)
 Crime and Punishment by Fyodor Dostoyevsky

(2541)

The Yellow Wallpaper by Charlotte Perkins Gilman (2429)

Grimms' Fairy Tales by Jacob Grimm and Wilhelm Grimm (2406)

Thus Spake Zarathustra: A Book for All and None by Friedrich Wilhelm Nietzsche (2400)

Dombey and Son by Charles Dickens (2393)

The Romance of Lust: A classic Victorian erotic novel by Anonymous (2382)

The Kama Sutra of Vatsyayana by Vatsyayana (2346)

The Odyssey by Homer (2221)

The Adventures of Tom Sawyer, Complete by Mark Twain (2211)

A Doll's House : a play by Henrik Ibsen (2183)

Great Expectations by Charles Dickens (2161)

The Importance of Being Earnest: A Trivial Comedy for Serious People by Oscar Wilde (2143)

The Iliad by Homer (2127)

Meditations by Emperor of Rome Marcus Aurelius (2120)

The Prophet by Kahlil Gibran (2106)

Moby Multiple Language Lists of Common Words by Grady Ward (2051)

Jane Eyre: An Autobiography by Charlotte Brontë (2024)

Beyond Good and Evil by Friedrich Wilhelm Nietzsche (2002)

Adventures of Huckleberry Finn by Mark Twain (1977)

Simple Sabotage Field Manual by United States. Office of Strategic Services (1959)

The green girl by Jack Williamson (1952)

The Strange Case of Dr. Jekyll and Mr. Hyde by Robert Louis Stevenson (1898)

Tractatus Logico-Philosophicus by Ludwig Wittgenstein (1878)

Don Quixote by Miguel de Cervantes Saavedra (1875)

Anne of Green Gables by L. M. Montgomery (1868)

The Wonderful Wizard of Oz by L. Frank Baum (1784)

The slang dictionary : by John Camden Hotten (1769)

A Study in Scarlet by Arthur Conan Doyle
(1755)
 The Republic by Plato (1753)
 Emma by Jane Austen (1750)
 Walden, and On The Duty Of Civil Disobedience by Henry
David Thoreau (1742)
 Calculus Made Easy by Silvanus P. Thompson
(1723)
 Wuthering Heights by Emily Brontë (1711)
 Treasure Island by Robert Louis Stevenson
(1701)
 The Rámáyan of Válmiki, translated into English
verse by Valmiki (1664)
 Little Women by Louisa May Alcott (1639)
 The divine comedy by Dante Alighieri (1624)
 On the Duty of Civil Disobedience by Henry David
Thoreau (1606)
 Lexicon Latinum : by Franz Wagner (1582)
 Skull-face by Robert E. Howard (1567)
 The Scarlet Letter by Nathaniel Hawthorne
(1566)
 Winnie-the-Pooh by A. A. Milne (1558)
 The Problems of Philosophy by Bertrand Russell
(1539)
 Ukraine, the land and its people : by Stepan
Rudnytskyi (1510)
 Les Misérables by Victor Hugo (1479)
 The War of the Worlds by H. G. Wells (1477)
 Demonology and Devil-lore by Moncure Daniel Conway
(1469)
 Peter Pan by J. M. Barrie (1444)
 The King in Yellow by Robert W. Chambers
(1425)
 A Plea for Captain John Brown by Henry David Thoreau
(1403)
 Notes from the Underground by Fyodor Dostoyevsky
(1401)
 Dubliners by James Joyce (1389)
 The King James Version of the Bible (1385)
 The Gnome King of Oz by Ruth Plumly Thompson
(1374)
 Sense and Sensibility by Jane Austen (1370)
 The White Mail by Cy Warman (1366)
 Heart of Darkness by Joseph Conrad (1345)
 David Copperfield by Charles Dickens (1327)
 Occult science in medicine by Franz Hartmann
(1318)

Carmilla by Joseph Sheridan Le Fanu (1314)
 The Art of War by active 6th century B.C. Sunzi
 (1312)
 Waverley; Or, 'Tis Sixty Years Since by Walter Scott
 (1293)
 A Christmas Carol in Prose; Being a Ghost Story of
 Christmas by Charles Dickens (1278)
 The Jungle Book by Rudyard Kipling (1261)

 <div class="padded">

 Top 100 EBooks yesterday
 Top 100 Authors yesterday
 Top 100 EBooks last 7 days
 Top 100 Authors last 7 days
 Top 100 EBooks last 30 days
 Top 100 Authors last 30 days

 </div>
 <h2 id="authors-last7">Top 100 Authors last 7 days</h2>

 Shakespeare, William (41622)
 Smollett, T. (Tobias) (38420)
 Doctorow, Cory (29661)
 Dumas, Alexandre (19719)
 Dickens, Charles (19177)
 Montgomery, L. M. (Lucy Maud)
 (18667)
 Melville, Herman (18439)
 Austen, Jane (18336)
 Alcott, Louisa May (17772)
 Forster, E. M. (Edward Morgan)
 (16482)
 Eliot, George (16462)
 Twain, Mark (15753)
 Merrill, Frank T. (14851)
 Gaskell, Elizabeth Cleghorn
 (13931)
 Von Arnim, Elizabeth (13875)
 Brock, C. E. (Charles Edmund)
 (12964)
 Doyle, Arthur Conan (12734)
 Fielding, Henry (12650)
 Garnett, Constance (12571)
 Wagner, Richard (12253)
 Dostoyevsky, Fyodor (11488)
 Tolstoy, Leo, graf (11439)

Wilde, Oscar (10446)
 Nietzsche, Friedrich Wilhelm
 (9231)
 Carroll, Lewis (9031)
 Wells, H. G. (Herbert George)
 (7331)
 Stevenson, Robert Louis (7306)
 Baum, L. Frank (Lyman Frank) (6700)
 Shelley, Mary Wollstonecraft (6496)
 Homer (6281)
 Boog, Carle Michel (6157)
 Jones, Mary Sherwood Wright
 (6157)
 Hillyer, V. M. (Virgil Mores)
 (6157)
 Plato (5846)
 Masters, Edgar Lee (5793)
 Thoreau, Henry David (5617)
 Jowett, Benjamin (5489)
 Verne, Jules (5395)
 Christie, Agatha (5337)
 Stoker, Bram (5278)
 Lang, Andrew (5267)
 Joyce, James (4920)
 Widger, David (4869)
 Kipling, Rudyard (4849)
 Swift, Jonathan (4712)
 Russell, Bertrand (4633)
 Chesterton, G. K. (Gilbert Keith)
 (4608)
 Poe, Edgar Allan (4410)
 Kafka, Franz (4409)
 Doré, Gustave (4350)
 Scott, Walter (4192)
 Rizal, José (4168)
 Conrad, Joseph (4109)
 Hawthorne, Nathaniel (4092)
 Maude, Aylmer (4089)
 Hugo, Victor (4087)
 Dante Alighieri (4083)
 London, Jack (3868)
 Howard, Robert E. (Robert Ervin)
 (3860)
 Maude, Louise (3774)
 Fitzgerald, F. Scott (Francis Scott)
 (3773)
 Wyllie, David (Translator) (3697)

Goethe, Johann Wolfgang von (3656)
 James, Henry (3633)
 Burton, Richard Francis, Sir (3630)
 Chekhov, Anton Pavlovich (3628)
 Ibsen, Henrik (3615)
 Barrie, J. M. (James Matthew) (3559)
 Shaw, Bernard (3469)
 Grimm, Wilhelm (3422)
 Grimm, Jacob (3422)
 Cervantes Saavedra, Miguel de (3330)
 Machiavelli, Niccolò (3288)
 Balzac, Honoré de (3274)
 Butler, Samuel (3264)
 Pope, Alexander (3264)
 Burroughs, Edgar Rice (3086)
 Wodehouse, P. G. (Pelham Grenville) (3051)
 Marriott, W. K. (William Kenaz) (2914)
 Valmiki (2899)
 Neill, John R. (John Rea) (2891)
 Morley, Henry (2806)
 Brontë, Charlotte (2737)
 Gilman, Charlotte Perkins (2710)
 Common, Thomas (2681)
 Defoe, Daniel (2680)
 Burnett, Frances Hodgson (2602)
 Gracie, Archibald (2569)
 Marcus Aurelius, Emperor of Rome (2495)
 Levy, Oscar (2450)
 Schopenhauer, Arthur (2404)
 Milne, A. A. (Alan Alexander) (2391)
 Ormsby, John (2390)
 Vatsyayana (2379)
 Derbyshire, Charles E. (2363)
 Potter, Beatrix (2351)
 Maupassant, Guy de (2349)
 Indrajit, Bhagavanlal (2346)
 Bhide, Shivaram Parashuram (2346)
 Ward, Grady (2328)


```

<div class="padded">
<ul>
<li><a href="#books-last1">Top 100 EBooks yesterday</a></li>
<li><a href="#authors-last1">Top 100 Authors yesterday</a></li>
<li><a href="#books-last7">Top 100 EBooks last 7 days</a></li>
<li><a href="#authors-last7">Top 100 Authors last 7 days</a></li>
<li><a href="#books-last30">Top 100 EBooks last 30 days</a></li>
<li><a href="#authors-last30">Top 100 Authors last 30 days</a></li>
</ul>
</div>
<h2 id="books-last30">Top 100 EBooks last 30 days</h2>
<ol>
<li><a href="/ebooks/1513">Romeo and Juliet by William Shakespeare
(78748)</a></li>
<li><a href="/ebooks/2701">Moby Dick; Or, The Whale by Herman Melville
(72956)</a></li>
<li><a href="/ebooks/2641">A Room with a View by E. M. Forster (66500)</a></li>
<li><a href="/ebooks/145">Middlemarch by George Eliot (65326)</a></li>
<li><a href="/ebooks/37106">Little Women; Or, Meg, Jo, Beth, and Amy by Louisa
May Alcott (63129)</a></li>
<li><a href="/ebooks/100">The Complete Works of William Shakespeare by William
Shakespeare (62279)</a></li>
<li><a href="/ebooks/16389">The Enchanted April by Elizabeth Von Arnim
(59886)</a></li>
<li><a href="/ebooks/67979">The Blue Castle: a novel by L. M. Montgomery
(59852)</a></li>
<li><a href="/ebooks/6761">The Adventures of Ferdinand Count Fathom - Complete
by T. Smollett (56596)</a></li>
<li><a href="/ebooks/394">Cranford by Elizabeth Cleghorn Gaskell
(56327)</a></li>
<li><a href="/ebooks/2160">The Expedition of Humphry Clinker by T. Smollett
(55504)</a></li>
<li><a href="/ebooks/4085">The Adventures of Roderick Random by T. Smollett
(55234)</a></li>
<li><a href="/ebooks/6593">History of Tom Jones, a Foundling by Henry Fielding
(54191)</a></li>
<li><a href="/ebooks/1259">Twenty Years After by Alexandre Dumas
(53760)</a></li>
<li><a href="/ebooks/5197">My Life - Volume 1 by Richard Wagner (52695)</a></li>
<li><a href="/ebooks/1342">Pride and Prejudice by Jane Austen (42038)</a></li>
<li><a href="/ebooks/8086">Down and Out in the Magic Kingdom by Cory Doctorow
(29639)</a></li>
<li><a href="/ebooks/1727">The Odyssey by Homer (23447)</a></li>
<li><a href="/ebooks/11">Alice's Adventures in Wonderland by Lewis Carroll
(23265)</a></li>
<li><a href="/ebooks/84">Frankenstein; Or, The Modern Prometheus by Mary
Wollstonecraft Shelley (21197)</a></li>

```

On the Duty of Civil Disobedience by Henry David Thoreau (19593)
 A Tale of Two Cities by Charles Dickens (16627)
 Dracula by Bram Stoker (16085)
 The Picture of Dorian Gray by Oscar Wilde (15554)
 The Adventures of Sherlock Holmes by Arthur Conan Doyle (15521)
 The Count of Monte Cristo, Illustrated by Alexandre Dumas (13637)
 War and Peace by graf Leo Tolstoy (12608)
 Metamorphosis by Franz Kafka (12523)
 The Great Gatsby by F. Scott Fitzgerald (12261)
 The Brothers Karamazov by Fyodor Dostoyevsky (12120)
 How to Pick a Mate: The Guide to a Happy Marriage by Clifford R. Adams and Vance Packard (12028)
 Ulysses by James Joyce (11895)
 The Prince by Niccolò Machiavelli (10831)
 The Yellow Wallpaper by Charlotte Perkins Gilman (10802)
 Crime and Punishment by Fyodor Dostoyevsky (10065)
 Anna Karenina by graf Leo Tolstoy (9807)
 The Romance of Lust: A classic Victorian erotic novel by Anonymous (9723)
 The Kama Sutra of Vatsyayana by Vatsyayana (9582)
 Grimms' Fairy Tales by Jacob Grimm and Wilhelm Grimm (9457)
 Great Expectations by Charles Dickens (9115)
 The Iliad by Homer (8918)
 A Modest Proposal by Jonathan Swift (8908)
 The Wonderful Wizard of Oz by L. Frank Baum (8868)
 The Adventures of Tom Sawyer, Complete by Mark Twain (8754)
 Thus Spake Zarathustra: A Book for All and None by Friedrich Wilhelm Nietzsche (8519)
 The slang dictionary : by John Camden Hotten (8375)
 Adventures of Huckleberry Finn by Mark Twain (8325)
 Moby Multiple Language Lists of Common Words by Grady Ward (8172)

Beyond Good and Evil by Friedrich Wilhelm Nietzsche (8102)
 The Importance of Being Earnest: A Trivial Comedy for Serious People by Oscar Wilde (8084)
 A Doll's House : a play by Henrik Ibsen (7992)
 Jane Eyre: An Autobiography by Charlotte Brontë (7968)
 The Prophet by Kahlil Gibran (7797)
 Anne of Green Gables by L. M. Montgomery (7783)
 The Rámáyan of Válmiki, translated into English verse by Valmiki (7752)
 Meditations by Emperor of Rome Marcus Aurelius (7750)
 Tractatus Logico-Philosophicus by Ludwig Wittgenstein (7608)
 The Strange Case of Dr. Jekyll and Mr. Hyde by Robert Louis Stevenson (7599)
 Don Quixote by Miguel de Cervantes Saavedra (7510)
 Calculus Made Easy by Silvanus P. Thompson (7184)
 Mark Twain's Speeches by Mark Twain (7151)
 Treasure Island by Robert Louis Stevenson (7079)
 A Study in Scarlet by Arthur Conan Doyle (7014)
 Walden, and On The Duty Of Civil Disobedience by Henry David Thoreau (6857)
 The Man That Corrupted Hadleyburg by Mark Twain (6853)
 A Child's History of the World by V. M. Hillyer (6739)
 Wuthering Heights by Emily Brontë (6656)
 Winnie-the-Pooh by A. A. Milne (6647)
 The Republic by Plato (6612)
 The divine comedy by Dante Alighieri (6556)
 The Scarlet Letter by Nathaniel Hawthorne (6407)
 Emma by Jane Austen (6356)
 Demonology and Devil-lore by Moncure Daniel Conway (6310)
 Little Women by Louisa May Alcott (6232)
 The King in Yellow by Robert W. Chambers (6196)
 Spoon River Anthology by Edgar Lee Masters

(6125)

- Les Misérables by Victor Hugo (6091)
- Peter Pan by J. M. Barrie (5866)
- The War of the Worlds by H. G. Wells (5730)
- The King James Version of the Bible (5587)
- A Christmas Carol in Prose; Being a Ghost Story of Christmas by Charles Dickens (5563)
- The Time Machine by H. G. Wells (5508)
- Notes from the Underground by Fyodor Dostoyevsky (5442)
- The Problems of Philosophy by Bertrand Russell (5422)
- Carmilla by Joseph Sheridan Le Fanu (5406)
- Heart of Darkness by Joseph Conrad (5384)
- Through the Looking-Glass by Lewis Carroll (5369)
- Dubliners by James Joyce (5249)
- David Copperfield by Charles Dickens (5209)
- Sense and Sensibility by Jane Austen (5189)
- The Happy Prince, and Other Tales by Oscar Wilde (5125)
- The Jungle Book by Rudyard Kipling (4960)
- Noli Me Tangere by José Rizal (4858)
- The Philippines a Century Hence by José Rizal (4765)
- Simple Sabotage Field Manual by United States. Office of Strategic Services (4750)
- The Art of War by active 6th century B.C. Sunzi (4726)
- Oliver Twist by Charles Dickens (4709)
- The murder of Roger Ackroyd by Agatha Christie (4696)
- The Souls of Black Folk by W. E. B. Du Bois (4593)
- Struwwelpeter: Merry Stories and Funny Pictures by Heinrich Hoffmann (4504)

- Top 100 EBooks yesterday
- Top 100 Authors yesterday
- Top 100 EBooks last 7 days
- Top 100 Authors last 7 days
- Top 100 EBooks last 30 days
- Top 100 Authors last 30 days

</div>

Top 100 Authors last 30 days

- [Shakespeare, William \(176985\)](/browse/authors/s#a65)
- [Smollett, T. \(Tobias\) \(168998\)](/browse/authors/s#a838)
- [Dumas, Alexandre \(81922\)](/browse/authors/d#a492)
- [Montgomery, L. M. \(Lucy Maud\) \(80282\)](/browse/authors/m#a36)
- [Melville, Herman \(79069\)](/browse/authors/m#a9)
- [Alcott, Louisa May \(76612\)](/browse/authors/a#a102)
- [Dickens, Charles \(73972\)](/browse/authors/d#a37)
- [Eliot, George \(71769\)](/browse/authors/e#a90)
- [Forster, E. M. \(Edward Morgan\) \(71393\)](/browse/authors/f#a975)
- [Austen, Jane \(71154\)](/browse/authors/a#a68)
- [Merrill, Frank T. \(64373\)](/browse/authors/m#a6202)
- [Twain, Mark \(64184\)](/browse/authors/t#a53)
- [Gaskell, Elizabeth Cleghorn \(61059\)](/browse/authors/g#a220)
- [Von Arnim, Elizabeth \(60714\)](/browse/authors/v#a603)
- [Brock, C. E. \(Charles Edmund\) \(56936\)](/browse/authors/b#a26815)
- [Fielding, Henry \(55632\)](/browse/authors/f#a537)
- [Wagner, Richard \(53905\)](/browse/authors/w#a1325)
- [Doyle, Arthur Conan \(50718\)](/browse/authors/d#a69)
- [Dostoyevsky, Fyodor \(45896\)](/browse/authors/d#a314)
- [Garnett, Constance \(43427\)](/browse/authors/g#a2858)
- [Wilde, Oscar \(41745\)](/browse/authors/w#a111)
- [Homer \(39900\)](/browse/authors/h#a705)
- [Tolstoy, Leo, graf \(38797\)](/browse/authors/t#a136)
- [Carroll, Lewis \(38125\)](/browse/authors/c#a7)
- [Nietzsche, Friedrich Wilhelm \(36697\)](/browse/authors/n#a779)
- [Thoreau, Henry David \(31369\)](/browse/authors/t#a54)
- [Wells, H. G. \(Herbert George\) \(30964\)](/browse/authors/w#a30)
- [Doctorow, Cory \(30527\)](/browse/authors/d#a3826)
- [Stevenson, Robert Louis \(30209\)](/browse/authors/s#a35)
- [Butler, Samuel \(27694\)](/browse/authors/b#a748)
- [Shelley, Mary Wollstonecraft \(26459\)](/browse/authors/s#a61)
- [Baum, L. Frank \(Lyman Frank\) \(23969\)](/browse/authors/b#a42)
- [Plato \(22269\)](/browse/authors/p#a93)
- [Howard, Robert E. \(Robert Ervin\) \(21305\)](/browse/authors/h#a36031)
- [Christie, Agatha \(21246\)](/browse/authors/c#a451)
- [Verne, Jules \(21192\)](/browse/authors/v#a60)

Jowett, Benjamin (20822)
 Chesterton, G. K. (Gilbert Keith) (20765)
 Rizal, José (20618)
 Joyce, James (20489)
 Stoker, Bram (19514)
 Kipling, Rudyard (19465)
 Poe, Edgar Allan (19313)
 Kafka, Franz (18976)
 Widger, David (18648)
 Lang, Andrew (17992)
 Doré, Gustave (17882)
 Maude, Aylmer (17629)
 Russell, Bertrand (17582)
 Swift, Jonathan (16886)
 Dante Alighieri (16869)
 Hugo, Victor (16455)
 Maude, Louise (16415)
 Conrad, Joseph (16318)
 Hawthorne, Nathaniel (16127)
 Fitzgerald, F. Scott (Francis Scott) (16108)
 Wyllie, David (Translator) (15958)
 Scott, Walter (15615)
 London, Jack (15448)
 Burton, Richard Francis, Sir (15257)
 James, Henry (14792)
 Chekhov, Anton Pavlovich (14564)
 Goethe, Johann Wolfgang von (14492)
 Ibsen, Henrik (14122)
 Grimm, Jacob (13798)
 Grimm, Wilhelm (13798)
 Burroughs, Edgar Rice (13665)
 Cervantes Saavedra, Miguel de (13661)
 Pope, Alexander (13625)
 Shaw, Bernard (12847)
 Barrie, J. M. (James Matthew) (12628)
 Balzac, Honoré de (12598)
 Machiavelli, Niccolò (12513)
 Gilman, Charlotte Perkins (12183)
 Packard, Vance (12028)
 Adams, Clifford R. (Clifford Rose)

(12028)
 Wodehouse, P. G. (Pelham Grenville)
 (11701)
 Morley, Henry (11477)
 Defoe, Daniel (11252)
 Valmiki (11200)
 Brontë, Charlotte (11111)
 Marriott, W. K. (William Kenaz)
 (10995)
 Levy, Oscar (10669)
 Burnett, Frances Hodgson (10608)
 Derbyshire, Charles E. (10351)
 Common, Thomas (9864)
 Milne, A. A. (Alan Alexander)
 (9862)
 Schopenhauer, Arthur (9848)
 Vatsyayana (9711)
 Maupassant, Guy de (9671)
 Ormsby, John (9670)
 Bhide, Shivaram Parashuram
 (9582)
 Indrajit, Bhagavanlal (9582)
 Buckley, Theodore Alois (9538)
 Potter, Beatrix (9531)
 Ward, Grady (9496)
 Hardy, Thomas (9390)
 Hotten, John Camden (9377)
 Hapgood, Isabel Florence (9334)
 Aesop (9319)

 <div class="padded">

 Top 100 EBooks yesterday
 Top 100 Authors yesterday
 Top 100 EBooks last 7 days
 Top 100 Authors last 7 days
 Top 100 EBooks last 30 days
 Top 100 Authors last 30 days

 </div>
 </div><!--content ending-->
 </div>
 <div class="footer"><!-- start footer -->

 Privacy policy


```

<a href="/about/" title="About Project Gutenberg">About Project Gutenberg</a>
</li>
<li>
<a href="/policy/terms_of_use.html" title="Terms of Use">Terms of Use</a>
</li>
<li>
<a href="/about/contact_information.html" title="Contact Information">Contact
Information</a>
</li>
<li><a href="/help/" title="Get Help">Get Help</a></li>
</ul>
<a href="https://www.ibiblio.org/" title="Project Gutenberg is hosted by
ibiblio">

</a>
</div><!-- footer ending-->
<!-- body ending -->
</body>
</html>

```

```

[107]: #6 Find all the href tags and store them in the list of links. Check what the
      ↪list looks like - print the first 30 elements:
# Define Empty list to hold all the http links in the HTML page
lst_links=[]
# Find all the href tags and store them in the list of links
# a href is the tag name for link
for link in soup.find_all('a'):
    lst_links.append(link.get('href')) #print(link.get('href'))
# Print the 30 links :
lst_links[:30]

```

```

[107]: ['/',
'/about/',
'/about/',
'/policy/collection_development.html',
'/about/contact_information.html',
'/about/background/',
'/policy/permission.html',
'/policy/privacy_policy.html',
'/policy/terms_of_use.html',
'/ebooks/',
'/ebooks/',
'/ebooks/bookshelf/',
'/browse/scores/top',
'/ebooks/offline_catalogs.html',
'/help/',
'/help/',

```

```

'/help/copyright.html',
'/help/errata.html',
'/help/file_formats.html',
'/help/faq.html',
'/policy/',
'/help/public_domain_ebook_submission.html',
'/help/submitting_your_own_work.html',
'/help/mobile.html',
'/attic/',
'/donate/',
'/donate/',
'#books-last1',
'#authors-last1',
'#books-last7']

```

```

[108]: # 7 & 8 Use a regular expression to find the numeric digits in these links.
↳ These are the file numbers for the top 100 books. Initialize the empty list
↳ to hold the file numbers:
booknum=[] # define blank list variable to store book info
# Numbers 19 to 118 in the original list of links have the top 100 eBooks'
↳ numbers. Loop over the appropriate range and use a regex to find the numeric
↳ digits in the link (href) string. Use the findall() method:
for i in range(19,119):
    link=lst_links[i]
    link=link.strip() # Regular expression to find the numeric digits in the
↳ link (href) string
    n=re.findall('[0-9]+',link)
    if len(n)==1: # Append the filename casted as integer
        booknum.append(int(n[0]))
# Print the file numbers:
print ("\n\nThe file numbers for the top 100 ebooks on Gutenberg are shown
↳ below\n"+"-"*70)
print(booknum)

```

The file numbers for the top 100 ebooks on Gutenberg are shown below

```

-----
[1, 1, 7, 7, 30, 30, 1513, 2701, 2641, 145, 37106, 100, 67979, 16389, 394, 6761,
2160, 4085, 1399, 1259, 6593, 5197, 1342, 11, 71285, 1184, 84, 71283, 71286, 98,
345, 1661, 71284, 174, 17346, 28054, 55387, 74, 71282, 2591, 67149, 5200, 2600,
64317, 27827, 2554, 4300, 1232, 76, 30254, 58585, 1998, 1080, 3206, 2680, 1260,
4363, 821, 244, 1400, 844, 1952, 6130, 43, 1727, 5740, 33283, 46, 2542, 996, 45,
768, 205, 1497, 5827, 45109, 42108, 158, 600, 55, 71280, 120, 8800, 730, 514,
71, 135, 36, 67098, 24869, 10, 8492]

```

```

[109]: # 9 What does the soup object's text look like? Use the .text method and print
↳ only the first 2,000 characters

```

```

#. (do not print the whole thing as it is too long).

#soup.text[:2000])
## it gives extra spaces
# Remove extra spaces and print the first 2000 characters
cleaned_text = re.sub(r'\s+', ' ', soup.text[:2000].strip())
print(cleaned_text)

```

Top 100 | Project Gutenberg Menu About About Project Gutenberg Collection Development Contact Us History & Philosophy Permissions & License Privacy Policy Terms of Use Search and Browse Book Search Bookshelves Frequently Downloaded Offline Catalogs Help All help topics → Copyright How-To Errata, Fixes and Bug Reports File Formats Frequently Asked Questions Policies → Public Domain eBook Submission Submitting Your Own Work Tablets, Phones and eReaders The Attic → Donate Donation Frequently Viewed or Downloaded These listings are based on the number of times each eBook gets downloaded. Multiple downloads from the same Internet address on the same day count as one download, and addresses that download more than 100 eBooks in a day are considered robots and are not counted. Downloaded Books 2023-07-28224074 last 7 days1501104 last 30 days5913305 Top 100 eBooks yesterday Top 100 Authors yesterday Top 100 eBooks last 7 days Top 100 Authors last 7 days Top 100 eBooks last 30 days Top 100 Authors last 30 days Top 100 eBooks yesterday Romeo and Juliet by William Shakespeare (2605) Moby Dick; Or, The Whale by Herman Melville (2464) A Room with a View by E. M. Forster (2237) Middlemarch by George Eliot (2198) Little Women; Or, Meg, Jo, Beth, and Amy by Louisa May Alcott (2102) The Complete Works of William Shakespeare by William Shakespeare (2007) The Blue Castle: a novel by L. M. Montgomery (1984) The Enchanted April by Elizabeth Von Arnim (1979) Cranford by Elizabeth Cleghorn Gaskell (1849) The Adventures of Ferdinand Count Fathom - Complete by T. Smollett (1841) The Expedition of Humphry Clinker by T. Smollett (1807) The Adventures of Roderick Random by T. Smollett (1785) Anna Karenina by graf Leo Tolstoy (1783) Twenty Years After by Alexandre Dumas (1750) History of Tom Jones, a Foundling by Henry Fielding (1736) My Life - Volume 1 by Richard Wagn

```

[110]: # 10,11,12,13 Search the extracted text (using regular expression) from the
        ↳soup object to find the names of top 100 eBooks (yesterday's rank)
#. and print them:
lst_titles_temp=[] # Temp empty list of Ebook
# Index search only for the string 'Top 100 eBooks yesterday'
target_text= 'Top 100 eBooks yesterday'
start_idx=soup.text.splitlines().index('Top 100 eBooks yesterday')
if start_idx != -1:
    print(f"Found '{target_text}' at index {start_idx}.")
else:
    print(f"'{target_text}' not found in the HTML content.")

for i in range(100):

```

```

lst_titles_temp.append(soup.text.splitlines()[start_idx+2+i])

lst_titles=[]
for i in range(100): # looking for first 100 lines
    id1,id2=re.match('^[a-zA-Z ;,\'.:]*',lst_titles_temp[i]).span()
    lst_titles.append(lst_titles_temp[i][id1:id2])
# Remove blank lines using filter()
filtered_list = list(filter(lambda x: x.strip(), lst_titles))
#pr#--> Still list has Top line which are not book title, lets remove
# Print the list of titles:

# Remove lines starting with 'TOP ' using list comprehension
filtered_list = [item for item in filtered_list if not item.startswith('Top ')]
#print(filtered_list)
print('Print the list of titles : ')
print('----- ')
for title in filtered_list:
    print(title)

```

Found 'Top 100 EBooks yesterday' at index 113.

Print the list of titles :

Romeo and Juliet by William Shakespeare
 Moby Dick; Or, The Whale by Herman Melville
 A Room with a View by E. M. Forster
 Middlemarch by George Eliot
 Little Women; Or, Meg, Jo, Beth, and Amy by Louisa May Alcott
 The Complete Works of William Shakespeare by William Shakespeare
 The Blue Castle: a novel by L. M. Montgomery
 The Enchanted April by Elizabeth Von Arnim
 Cranford by Elizabeth Cleghorn Gaskell
 The Adventures of Ferdinand Count Fathom
 The Expedition of Humphry Clinker by T. Smollett
 The Adventures of Roderick Random by T. Smollett
 Anna Karenina by graf Leo Tolstoy
 Twenty Years After by Alexandre Dumas
 History of Tom Jones, a Foundling by Henry Fielding
 My Life
 Pride and Prejudice by Jane Austen
 Alice's Adventures in Wonderland by Lewis Carroll
 Dead letters by Maurice Baring
 The Count of Monte Cristo, Illustrated by Alexandre Dumas
 Frankenstein; Or, The Modern Prometheus by Mary Wollstonecraft Shelley
 Richard the Lion Heart by Kate Norgate
 Oliver Ellis : by James Grant
 A Tale of Two Cities by Charles Dickens
 Dracula by Bram Stoker

The Adventures of Sherlock Holmes by Arthur Conan Doyle
 The House of Egremont : by Molly Elliot Seawell
 The Picture of Dorian Gray by Oscar Wilde
 The Report on Unidentified Flying Objects by Edward J. Ruppelt
 The Brothers Karamazov by Fyodor Dostoyevsky
 One Hundred Proofs That the Earth Is Not a Globe by William Carpenter
 The Adventures of Tom Sawyer, Complete by Mark Twain
 Criminality and economic conditions by Willem Adriaan Bongers
 Grimms' Fairy Tales by Jacob Grimm and Wilhelm Grimm
 A Child's History of the World by V. M. Hillyer
 Metamorphosis by Franz Kafka
 War and Peace by graf Leo Tolstoy
 The Great Gatsby by F. Scott Fitzgerald
 The Kama Sutra of Vatsyayana by Vatsyayana
 Crime and Punishment by Fyodor Dostoyevsky
 Ulysses by James Joyce
 The Prince by Niccolò Machiavelli
 Adventures of Huckleberry Finn by Mark Twain
 The Romance of Lust: A classic Victorian erotic novel by Anonymous
 The Prophet by Kahlil Gibran
 Thus Spake Zarathustra: A Book for All and None by Friedrich Wilhelm Nietzsche
 A Modest Proposal by Jonathan Swift
 Moby Multiple Language Lists of Common Words by Grady Ward
 Meditations by Emperor of Rome Marcus Aurelius
 Jane Eyre: An Autobiography by Charlotte Brontë
 Beyond Good and Evil by Friedrich Wilhelm Nietzsche
 Dombey and Son by Charles Dickens
 A Study in Scarlet by Arthur Conan Doyle
 Great Expectations by Charles Dickens
 The Importance of Being Earnest: A Trivial Comedy for Serious People by Oscar Wilde
 The Yellow Wallpaper by Charlotte Perkins Gilman
 The Iliad by Homer
 The Strange Case of Dr. Jekyll and Mr. Hyde by Robert Louis Stevenson
 The Odyssey by Homer
 Tractatus Logico-Philosophicus by Ludwig Wittgenstein
 Calculus Made Easy by Silvanus P. Thompson
 A Christmas Carol in Prose; Being a Ghost Story of Christmas by Charles Dickens
 A Doll's House : a play by Henrik Ibsen
 Don Quixote by Miguel de Cervantes Saavedra
 Anne of Green Gables by L. M. Montgomery
 Wuthering Heights by Emily Brontë
 Walden, and On The Duty Of Civil Disobedience by Henry David Thoreau
 The Republic by Plato
 The Problems of Philosophy by Bertrand Russell
 The Enchiridion by Epictetus
 The slang dictionary : by John Camden Hotten
 Emma by Jane Austen

Notes from the Underground by Fyodor Dostoyevsky
 The Wonderful Wizard of Oz by L. Frank Baum
 Lexicon Latinum : by Franz Wagner
 Treasure Island by Robert Louis Stevenson
 The divine comedy by Dante Alighieri
 Oliver Twist by Charles Dickens
 Little Women by Louisa May Alcott
 On the Duty of Civil Disobedience by Henry David Thoreau
 Les Mis
 The War of the Worlds by H. G. Wells
 Winnie
 The R
 The King James Version of the Bible
 The King in Yellow by Robert W. Chambers
 David Copperfield by Charles Dickens
 Peter Pan by J. M. Barrie
 The Scarlet Letter by Nathaniel Hawthorne
 The Art of War by active
 Sense and Sensibility by Jane Austen
 Heart of Darkness by Joseph Conrad

```
[111]: # Activity 10: Extracting the top 100 eBooks from Gutenberg.org
# Register email address to generate API Key
# Open the omdb api key website (omdbapi.com/apikey.aspx)
#--> OMDb API: http://www.omdbapi.com/?i=tt3896198&apikey=a306ad67
{
  "OMDBapi" : "a306ad67"
}
```

```
[111]: {'OMDBapi': 'a306ad67'}
```

```
[112]: # Import urllib.request, urllib.parse, urllib.error, and json:
import urllib.request, urllib.parse, urllib.error
import json
```

```
[130]: # Open the APIkeys.json
with open('APIkeys.json') as f:
    keys = json.load(f)
    omdbapi = keys['OMDBapi']

#apikey = 'apikey='+omdbapi # store in local variable
apikey = omdbapi # store in local variable
print(apikey)

# The final URL to be passed with the APIkey
serviceurl = 'http://www.omdbapi.com/?' ## OMDb base url

# final url for title : Guardians of the Galaxy Vol. 2
```

```
title='Guardians of the Galaxy Vol. 2'
#http://www.omdbapi.com/?t='Guardians of the Galaxy Vol. 2'&apikey=secretapikey
a306ad67
```

```
[132]: def fetch_movie_details(api_key, title):
    url = f"http://www.omdbapi.com/?t={title}&apikey={api_key}"
    print(url)
    #url='http://www.omdbapi.com/?i=tt3896198&apikey=a306ad67'

    response = requests.get(url)
    if response.status_code == 200:
        return response.json()
    else:
        print(f"Failed to fetch movie details. Status Code: {response.
↪status_code}")
        return None

if __name__ == "__main__":
    api_key = "a306ad67" # Replace with OMDb API key
    title='Guardians of the Galaxy Vol. 2' # Replace with the title we want to
↪fetch details for
    print(title)

    movie_details = fetch_movie_details(api_key, title)
    if movie_details:
        print("Movie Details of -- Guardians of the Galaxy Vol. 2")
        print('-----')
        for key, value in movie_details.items():
            print(f"{key}: {value}")
```

```
Guardians of the Galaxy Vol. 2
http://www.omdbapi.com/?t=Guardians of the Galaxy Vol. 2&apikey=a306ad67
Movie Details of -- Guardians of the Galaxy Vol. 2
-----
Title: Guardians of the Galaxy Vol. 2
Year: 2017
Rated: PG-13
Released: 05 May 2017
Runtime: 136 min
Genre: Action, Adventure, Comedy
Director: James Gunn
Writer: James Gunn, Dan Abnett, Andy Lanning
Actors: Chris Pratt, Zoe Saldana, Dave Bautista
Plot: The Guardians struggle to keep together as a team while dealing with their
personal family issues, notably Star-Lord's encounter with his father, the
ambitious celestial being Ego.
Language: English
```

Country: United States
Awards: Nominated for 1 Oscar. 15 wins & 60 nominations total
Poster: https://m.media-amazon.com/images/M/MV5BNjMONTcONzItM2FlYS00YzEwLWE0YmUtNTA2ZWlzM2OTgxXkEyXkFqcGdeQXVyNTgwNzIyNzg@._V1_SX300.jpg
Ratings: [{'Source': 'Internet Movie Database', 'Value': '7.6/10'}, {'Source': 'Rotten Tomatoes', 'Value': '85%'}, {'Source': 'Metacritic', 'Value': '67/100'}]
Metascore: 67
imdbRating: 7.6
imdbVotes: 725,527
imdbID: tt3896198
Type: movie
DVD: 22 Aug 2017
BoxOffice: \$389,813,101
Production: N/A
Website: N/A
Response: True

```
[133]: # Utility function to download poster from movie information
def download_movie_poster(movie_details, save_path):
    if 'Poster' in movie_details and movie_details['Poster'] != 'N/A':
        poster_url = movie_details['Poster']

        response = requests.get(poster_url)
        if response.status_code == 200:
            with open(save_path, 'wb') as f:
                f.write(response.content)
            print(f"Poster downloaded and saved to: {save_path}")
        else:
            print(f"Failed to download poster. Status Code: {response.
↳ status_code}")
        else:
            print("No poster available for this movie.")

#--> Call the download function

poster_path = "movie_poster.jpg" # path and filename
download_movie_poster(movie_details, poster_path)
```

Poster downloaded and saved to: movie_poster.jpg

```
[134]: ## Call movie function by entering 'Titanic'
title="Titanic"
movie_details = fetch_movie_details(api_key, title)
if movie_details:
    print("Movie Details of -- Guardians of the Galaxy Vol. 2")
    print('-----')
    for key, value in movie_details.items():
        print(f"{key}: {value}")
```


<http://www.omdbapi.com/?t=Titanic&apikey=a306ad67>
Movie Details of -- Guardians of the Galaxy Vol. 2

Title: Titanic
Year: 1997
Rated: PG-13
Released: 19 Dec 1997
Runtime: 194 min
Genre: Drama, Romance
Director: James Cameron
Writer: James Cameron
Actors: Leonardo DiCaprio, Kate Winslet, Billy Zane
Plot: A seventeen-year-old aristocrat falls in love with a kind but poor artist aboard the luxurious, ill-fated R.M.S. Titanic.
Language: English, Swedish, Italian, French
Country: United States, Mexico
Awards: Won 11 Oscars. 126 wins & 83 nominations total
Poster: https://m.media-amazon.com/images/M/MV5BMDdmZGU3NDQtY2E5My00ZTliLWIzOTUtMTY4ZGI1YjdiNjk3XkEyXkFqcGdeQXVyNTA4NzY1MzY@._V1_SX300.jpg
Ratings: [{'Source': 'Internet Movie Database', 'Value': '7.9/10'}, {'Source': 'Rotten Tomatoes', 'Value': '88%'}, {'Source': 'Metacritic', 'Value': '75/100'}]
Metascore: 75
imdbRating: 7.9
imdbVotes: 1,228,124
imdbID: tt0120338
Type: movie
DVD: 08 Jan 2002
BoxOffice: \$674,292,608
Production: N/A
Website: N/A
Response: True

```
[135]: ## Call movie function by entering 'Random_error'
title="Random_error"
movie_details = fetch_movie_details(api_key, title)
if movie_details:
    print("Movie Details of -- Guardians of the Galaxy Vol. 2")
    print('-----')
    for key, value in movie_details.items():
        print(f"{key}: {value}")
```

http://www.omdbapi.com/?t=Random_error&apikey=a306ad67
Movie Details of -- Guardians of the Galaxy Vol. 2

Response: False
Error: Movie not found!

```
[136]: # 3. Connect to an API of your choice and do a simple data pull - you can use
        ↪ any API - except the API you have selected for your project.
## a. In previous versions of this course we have always used Twitter, but with
        ↪ recent organizational changes at Twitter, it has become increasingly
        ↪ difficult to access the free APIs available at Twitter. You are more than
        ↪ welcome to try to use Twitter's API for this portion of the assignment, but
        ↪ please note, there has been some inconsistency experienced when following
        ↪ along with their documentation posted.
## b. Connect to the API and do a "Get" call/operation on the API to return a
        ↪ subset of data from the API
def get_weather_data(api_key, city_name):
    url = f"http://api.openweathermap.org/data/2.5/weather?
    ↪ q={city_name}&appid={api_key}"

    response = requests.get(url)
    if response.status_code == 200:
        return response.json()
    else:
        print(f"Failed to fetch weather data. Status Code: {response.
        ↪ status_code}")
        return None

if __name__ == "__main__":
    api_key = "eef7222cf1fdaeae6fe076a9c5205bd3" # Replace with
    ↪ OpenWeatherMap API key
    city_name = "London" # Replace with the city name for which you want
    ↪ to fetch weather data
    weather_data = get_weather_data(api_key, city_name)
    if weather_data:
        print(f"Weather data for {city_name}:")
        print("Temperature:", weather_data["main"]["temp"], "K")
        print("Weather:", weather_data["weather"][0]["main"])
        print("Description:", weather_data["weather"][0]["description"])
        print("Humidity:", weather_data["main"]["humidity"], "%")
        print("Wind Speed:", weather_data["wind"]["speed"], "m/s")
    else:
        print("Failed to fetch weather data.")
```

```
Weather data for London:
Temperature: 287.91 K
Weather: Clouds
Description: scattered clouds
Humidity: 82 %
Wind Speed: 0.89 m/s
```

[138]:

4. Using one of the datasets provided, or a dataset of your own, choose 3 of the following visualizations to complete. You must submit via PDF along with your code. You are free to use Matplotlib, Seaborn or another package if you prefer.

```
## a. Line
## b. Scatter
## c. Bar
## d. Histogram
## e. Density Plot
## f. Pie Chart
```

```
# The data set about mall customer segmentation data file : Mall_Customers.csv
data = pd.read_csv('Mall_Customers.csv')
data.head(10)
```

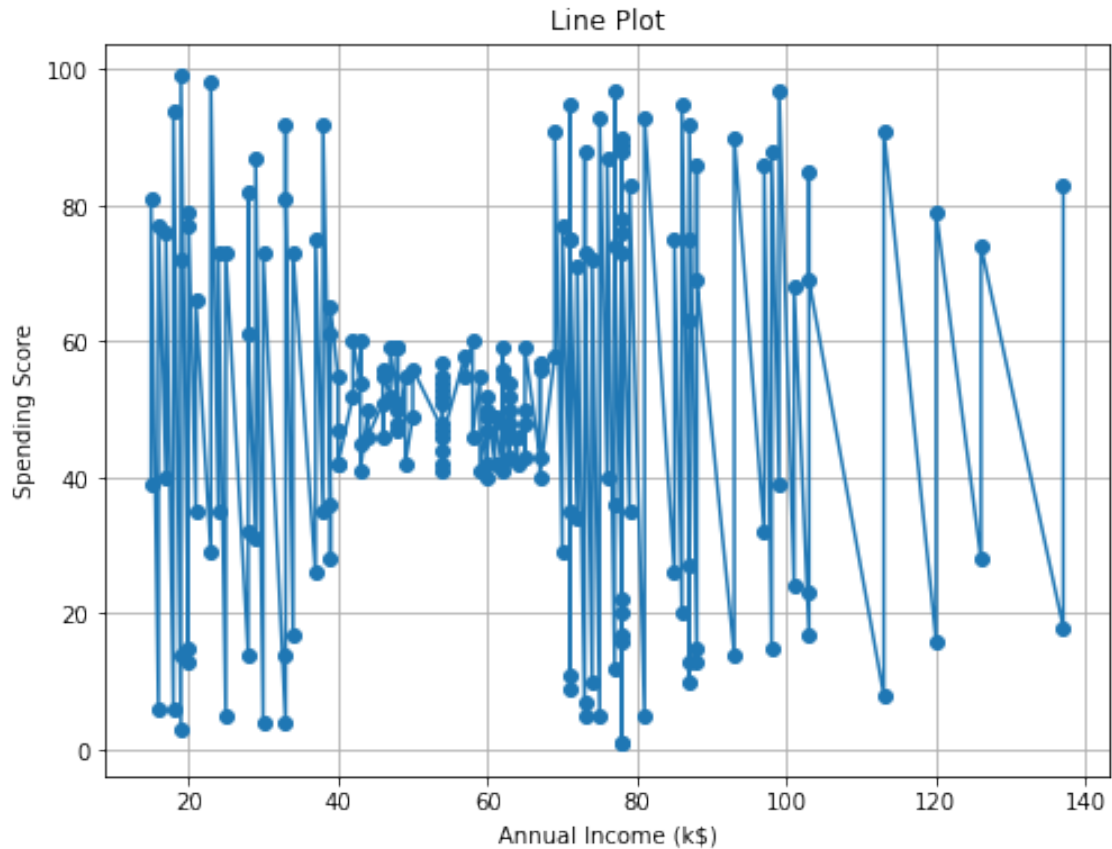
```
[138]:
```

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
5	6	Female	22	17	76
6	7	Female	35	18	6
7	8	Female	23	18	94
8	9	Male	64	19	3
9	10	Female	30	19	72

```
[145]: # Step 4: Choose three visualizations (Line, Scatter, and Bar)
## Plotting the Relation between Age , Annual Income and Spending Score
# Line Plot
plt.figure(figsize=(8, 6))
plt.plot(data['Annual Income (k$)'], data['Spending Score (1-100)'], marker='o')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score')

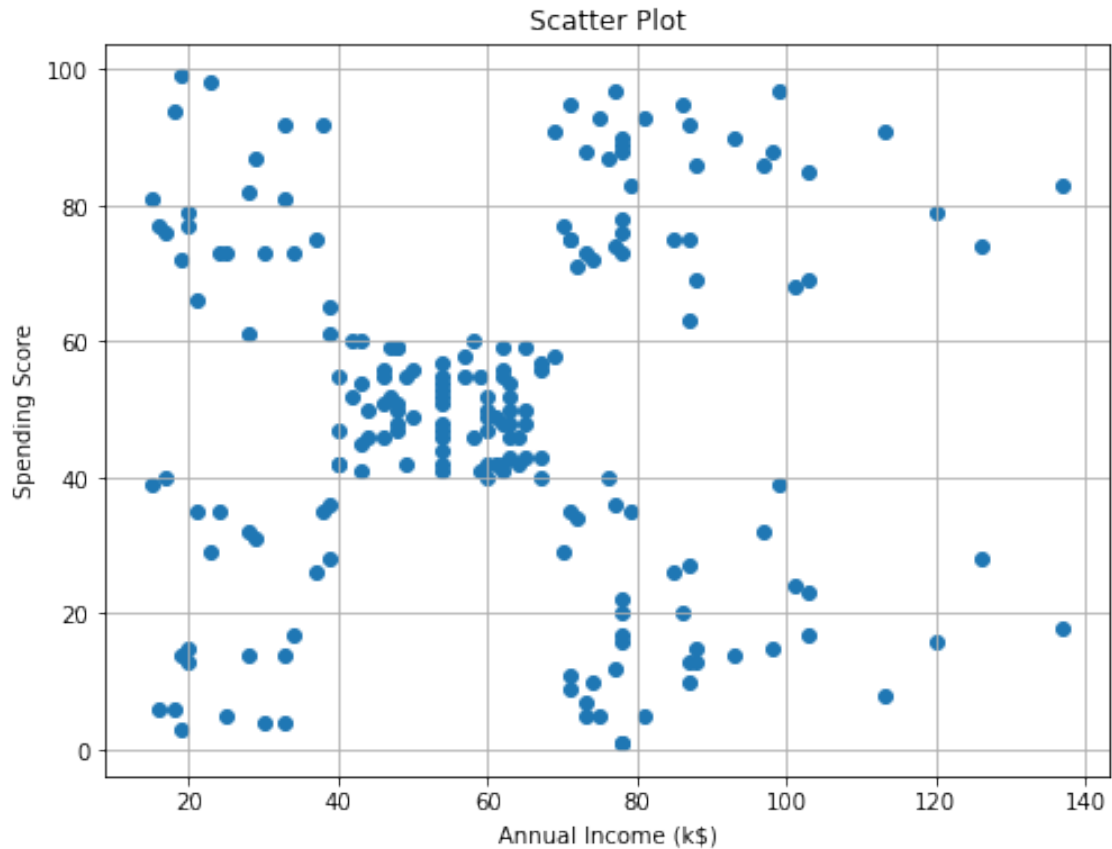
plt.title('Line Plot')
plt.grid(True)
plt.savefig('line_plot.png')
plt.show()

##--> Line plot show that Spending Score does not depend on the income
```

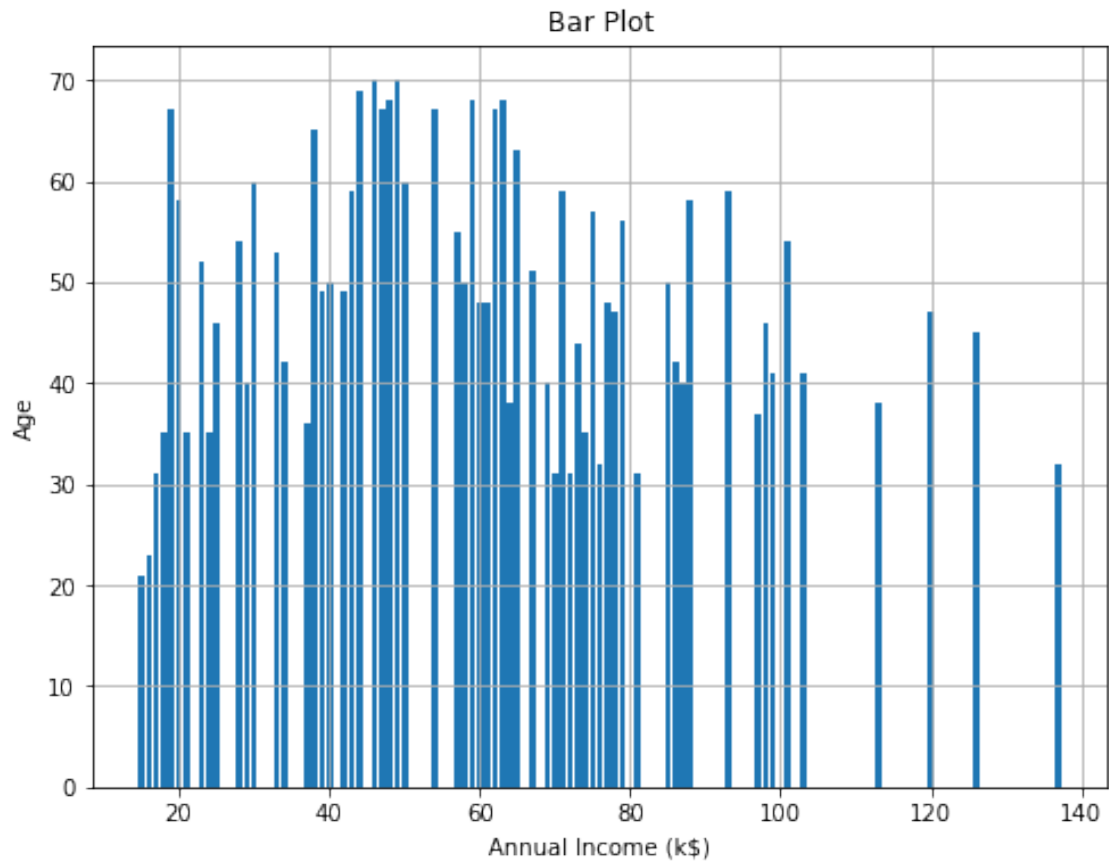


```
[146]: # Scatter Plot
plt.figure(figsize=(8, 6))
plt.scatter(data['Annual Income (k$)'], data['Spending Score (1-100)'])
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score')
plt.title('Scatter Plot')
plt.grid(True)
plt.savefig('scatter_plot.png')
plt.show()

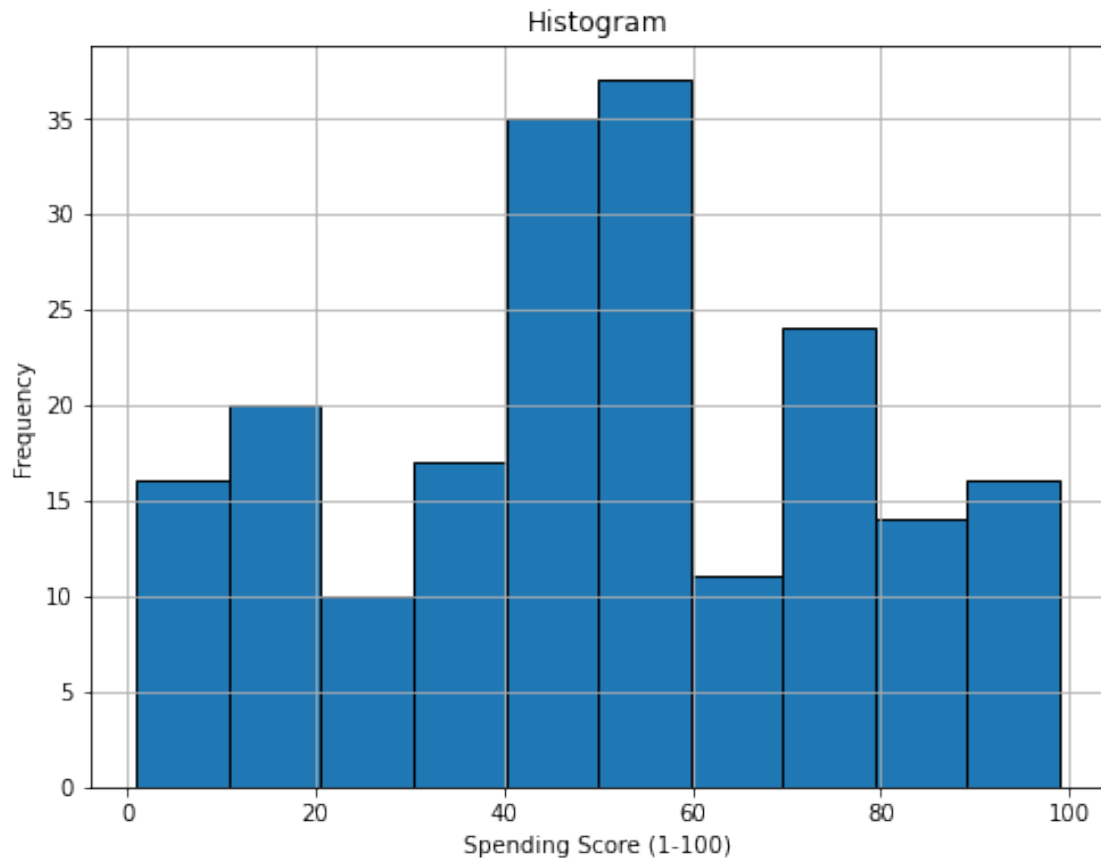
##--> Scatter plot show that Spending Score does not depend on the income
```



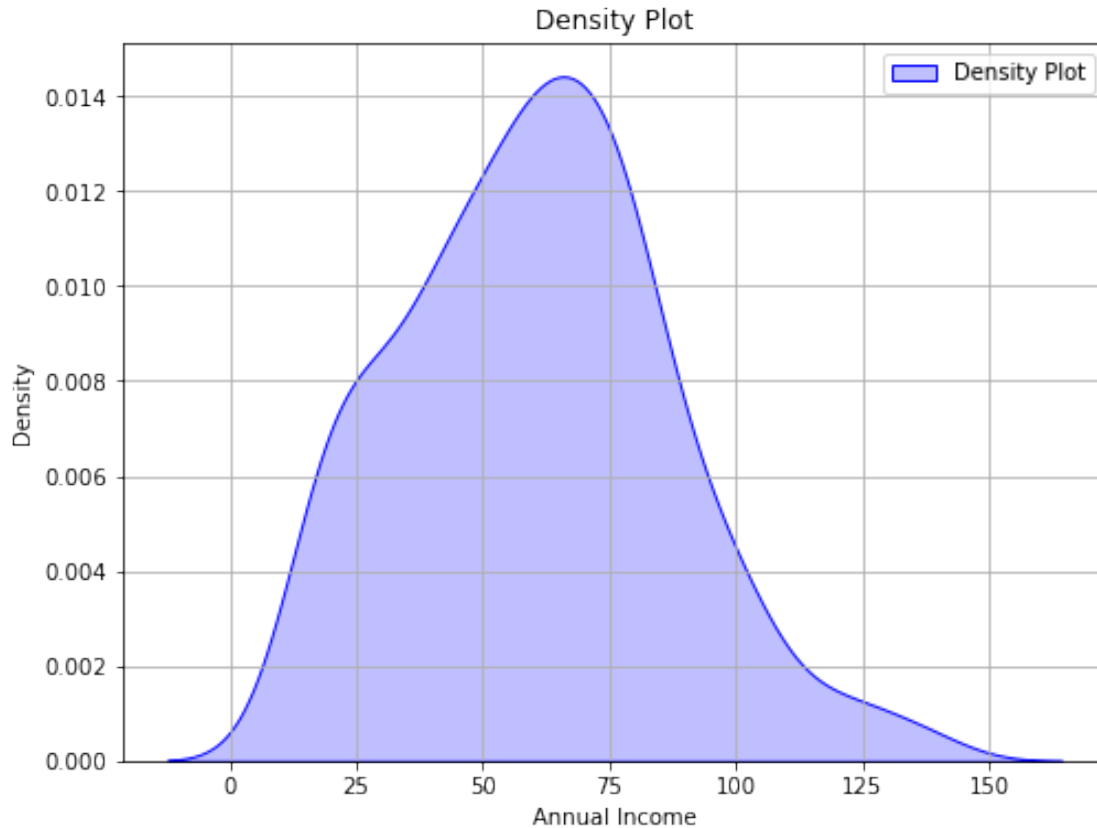
```
[147]: # Bar Plot
plt.figure(figsize=(8, 6))
plt.bar(data['Annual Income (k$)'], data['Age'])
plt.xlabel('Annual Income (k$)')
plt.ylabel('Age')
plt.title('Bar Plot')
plt.grid(True)
plt.savefig('bar_plot.png')
plt.show()
##--> Bar plot shows that between 30 to 45 age group do more income
```



```
[148]: # Histogram for Spending Score
plt.figure(figsize=(8, 6))
plt.hist(data['Spending Score (1-100)'], bins=10, edgecolor='black')
plt.xlabel('Spending Score (1-100)')
plt.ylabel('Frequency')
plt.title('Histogram')
plt.grid(True)
plt.savefig('histogram.png')
plt.show()
##--> Histogram of pending Score shows that maximum people spending score
      ↳ between 40 to 60
```



```
[149]: # Density Plot
import seaborn as sns
plt.figure(figsize=(8, 6))
sns.kdeplot(data['Annual Income (k$)'], fill=True, color='blue', label='Density_
↳Plot')
plt.xlabel('Annual Income')
plt.ylabel('Density')
plt.title('Density Plot')
plt.grid(True)
plt.legend()
plt.savefig('density_plot.png')
plt.show()
##--> Density Plot of Annual Income shows that maximum people annual income is_
↳between 50 to 75k
```



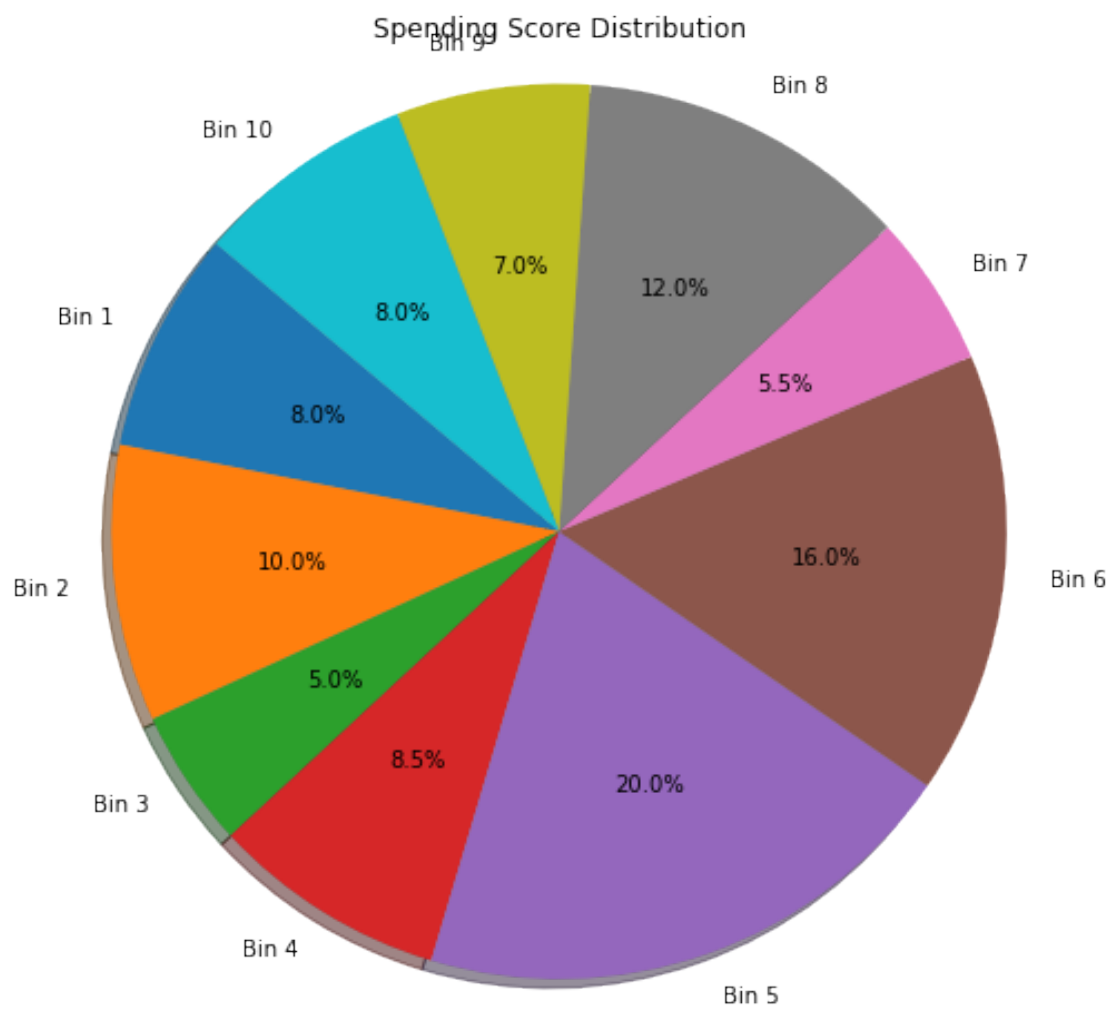
```
[144]: # Create a DataFrame from the sample data
df = pd.DataFrame(data)

# Create 10 bins with equal width for the Spending Score
# The 'cut' function will categorize the data into bins
num_bins = 10
df['Score Category'] = pd.cut(data['Spending Score (1-100)'], bins=num_bins,
    ↳ labels=[f'Bin {i+1}' for i in range(num_bins)])

# Count the number of occurrences in each bin
bin_counts = df['Score Category'].value_counts().sort_index()

# Create a Pie Chart
plt.figure(figsize=(8, 8))
plt.pie(bin_counts, labels=bin_counts.index, autopct='%1.1f%%', shadow=True,
    ↳ startangle=140)
plt.axis('equal')
plt.title('Spending Score Distribution')
plt.savefig('spending_score_pie_chart.png')
plt.show()
```


##--> Pie Chart of pending Score shows that maximum people spending score
↪ between 50 to 60 and is around 20% of the population.



[]: