

```
In [1]: import requests
import pandas as pd
from bs4 import BeautifulSoup
```

```
In [3]: url = 'https://github.com/trending'
response = requests.get(url)
response.status_code
```

Out[3]: 200

```
In [4]: page_content = response.text
page_content[:100]
```

Out[4]: '\n\n<!DOCTYPE html>\n<html lang="en" data-color-mode="auto" data-light-theme="light" data-dark-theme="d'

```
In [6]: with open('trending.html', 'w', encoding='utf-8') as f:
f.write(page_content)
```

Beautiful Soup

```
In [5]: data = BeautifulSoup(page_content, 'html.parser')
```

For username and repository name

```
In [16]: h1 = data.find_all('h1', class_='h3 lh-condensed')
```

```
In [25]: names = []
for i in h1:
    name = i.find('a').text
    # print(sp.strip())
    names.append(name.strip())
    # print(name)
print(names)
```

```
['bytedance /\n\n      btrace', 'sindresorhus /\n\n      awesome', 'google /\n\n      lo
g4jscanner', 'DataTalksClub /\n\n      data-engineering-zoomcamp', 'PaddlePaddle /\n\n
PaddleDetection', 'papers-we-love /\n\n      papers-we-love', 'google /\n\n      python-
fire', 'donnemartin /\n\n      system-design-primer', 'youngyangyang04 /\n\n      leetco
de-master', 'ja-netfilter /\n\n      ja-netfilter', 'angular /\n\n      angular', 'Micro
softLearning /\n\n      AZ-204-DevelopingSolutionsforMicrosoftAzure', 'dfir-iris /\n\n
iris-web', 'CleverRaven /\n\n      Cataclysm-DDA', 'ZLMediaKit /\n\n      ZLMediaKit',
'element-plus /\n\n      element-plus', 'Mas0nShi /\n\n      typoraCracker', 'gatsbyjs /
\n\n      gatsby', 'google /\n\n      gvisor', 'facebook /\n\n      react-native', 'hibi
ken /\n\n      asynq', 'alibaba /\n\n      formily', 'skyline-emu /\n\n      skyline',
'vnpv /\n\n      vnpv']
```

For language

```
In [38]: field = data.find_all('div',class_='f6 color-fg-muted mt-2')
categories = []
for i in field:
    category = i.find('span').text
    # print(category)
    categories.append(category.strip())
print(categories)

['Kotlin', 'Built by', 'Go', 'Built by', 'Python', 'Shell', 'Python', 'Python', 'Built by', 'Java', 'TypeScript', 'C#', 'JavaScript', 'C++', 'C++', 'Vue', 'Python', 'JavaScript', 'Go', 'JavaScript', 'Go', 'TypeScript', 'C++', 'Python']
```

For stars

```
In [58]: rating = data.find_all('div',class_='f6 color-fg-muted mt-2')
stars = []
for i in rating:
    s = i.find_all('a',class_='Link--muted d-inline-block mr-3')
    star = s[0].text
    stars.append(star.strip())
print(stars)

['654', '182,542', '561', '623', '6,029', '51,388', '21,467', '156,839', '20,676', '2,524', '78,583', '1,245', '151', '6,184', '5,377', '13,316', '645', '52,062', '12,233', '100,218', '2,237', '6,878', '2,531', '17,199']
```

For repos

```
In [60]: repo_counts = data.find_all('div',class_='f6 color-fg-muted mt-2')
repos = []
for i in repo_counts:
    r = i.find_all('a',class_='Link--muted d-inline-block mr-3')
    repo = r[1].text
    repos.append(repo.strip())
print(repos)

['86', '22,456', '45', '95', '1,512', '4,340', '1,255', '28,946', '5,291', '640', '20,632', '967', '11', '3,027', '1,492', '2,480', '230', '10,073', '1,010', '21,593', '175', '879', '475', '6,919']
```

Split name data to get username and repos name separately

```
In [71]: username = []
repo_name = []
for i in names:
    x = i.split('/')
    username.append(x[0])
    repo_name.append(x[1].strip())
print(username)
print(repo_name)
```

```
['bytedance ', 'sindresorhus ', 'google ', 'DataTalksClub ', 'PaddlePaddle ', 'papers-we-love ', 'google ', 'donnemartin ', 'youngyangyang04 ', 'ja-netfilter ', 'angular ', 'MicrosoftLearning ', 'dfir-iris ', 'CleverRaven ', 'ZLMediaKit ', 'element-plus ', 'Mas0nShi ', 'gatsbyjs ', 'google ', 'facebook ', 'hibiken ', 'alibaba ', 'skyline-emu ', 'vnpyp']
['btrace', 'awesome', 'log4jscanner', 'data-engineering-zoomcamp', 'PaddleDetection', 'papers-we-love', 'python-fire', 'system-design-primer', 'leetcode-master', 'ja-netfilter', 'angular', 'AZ-204-DevelopingSolutionsforMicrosoftAzure', 'iris-web', 'Cataclysm-DDA', 'ZLMediaKit', 'element-plus', 'typoraCracker', 'gatsby', 'gvisor', 'react-native', 'asynq', 'formily', 'skyline', 'vnpyp']
```

In [129...

```
desc = data.find_all('article', class_='Box-row')
repo_desc = []
for i in desc:
    x = i.find('p', class_='col-9 color-fg-muted my-1 pr-4')
    # print(x)
    repo_desc.append(x)

repo_detail = []
for j in repo_desc:
    if j == None:
        repo_detail.append(j)
    else:
        r = j.text
        repo_detail.append(r.strip())
print(repo_detail)
```

['🐛🐛 btrace(AKA RheaTrace) is a high performance Android trace tool which is based on Systrace, it support to define custom events automatically during building apk and using bhook to provider more native events like IO.', '😄 Awesome lists about all kinds of interesting topics', None, 'Code for Data Engineer Zoomcamp course', 'Object Detection toolkit based on PaddlePaddle. It supports object detection, instance segmentation, multiple object tracking and real-time multi-person keypoint detection.', 'Papers from the computer science community to read and discuss.', 'Python Fire is a library for automatically generating command line interfaces (CLIs) from absolutely any Python object.', 'Learn how to design large-scale systems. Prep for the system design interview. Includes Anki flashcards.', '《代码随想录》LeetCode 刷题攻略: 200道经典题目刷题顺序, 共60w字的详细图解, 视频难点剖析, 50余张思维导图, 支持C++, Java, Python, Go, JavaScript等多语言版本, 从此算法学习不再迷茫! 🐛🐛 来看看, 你会发现相见恨晚! 🐛', 'A javaagent framework', 'The modern web developer's platform', 'AZ-204: Developing solutions for Microsoft Azure', 'Incident Response collaborative platform', 'Cataclysm - Dark Days Ahead. A turn-based survival game set in a post-apocalyptic world.', 'WebRTC/RTSP/RTMP/HTTP/HLS/HTTP-FLV/WebSocket-FLV/HTTP-TS/HTTP-fMP4/WebSocket-TS/WebSocket-fMP4/GB28181 server and client framework based on C++11', '🐛 A Vue.js 3 UI Library made by Element team', 'A extract & decryption and pack & encryption tools for typora.', 'Build blazing fast, modern apps and websites with React', 'Application Kernel for Containers', 'A framework for building native applications using React', 'Asynq: simple, reliable, and efficient distributed task queue in Go', 'Alibaba Group Unified Form Solution -- Support React/ReactNative/Vue2/Vue3', 'Run Nintendo Switch homebrew & games on your Android device!', '基于Python的开源量化交易平台开发框架']

All basic info is extracted

Now creating dataframe using pandas




In []:

In [130...

```
trend_repo = pd.DataFrame({
    'user_name': username,
    'repo_name': repo_name,
    'repo_desc': repo_detail,
    'repo_stars': stars,
    'repo_count': repos
})
```

In [131... trend_repo

Out[131...

	user_name	repo_name	repo_desc	repo_stars
0	bytedance	btrace	 btrace(AKA RheaTrace) is a high performance...	654
1	sindresorhus	awesome	 Awesome lists about all kinds of interesting...	182,542
2	google	log4jscanner	None	561
3	DataTalksClub	data-engineering-zoomcamp	Code for Data Engineer Zoomcamp course	623
4	PaddlePaddle	PaddleDetection	Object Detection toolkit based on PaddlePaddle...	6,029
5	papers-we-love	papers-we-love	Papers from the computer science community to ...	51,388
6	google	python-fire	Python Fire is a library for automatically gen...	21,467
7	donnemartin	system-design-primer	Learn how to design large-scale systems. Prep ...	156,839
8	youngyangyang04	leetcode-master	《代码随想录》LeetCode 刷题攻略：200道经典题目刷题顺序，共60w字的详细图解，视...	20,676
9	ja-netfilter	ja-netfilter	A javaagent framework	2,524
10	angular	angular	The modern web developer's platform	78,583
11	MicrosoftLearning	AZ-204-DevelopingSolutionsforMicrosoftAzure	AZ-204: Developing solutions for Microsoft Azure	1,245
12	dfir-iris	iris-web	Incident Response collaborative platform	151
13	CleverRaven	Cataclysm-DDA	Cataclysm - Dark Days Ahead. A turn-based surv...	6,184
14	ZLMediaKit	ZLMediaKit	WebRTC/RTSP/RTMP/HTTP/HLS/HTTP-FLV/WebSocket-F...	5,377
15	element-plus	element-plus	 A Vue.js 3 UI Library made by Element team	13,316
16	Mas0nShi	typoraCracker	A extract & decryption and pack & encryption t...	645
17	gatsbyjs	gatsby	Build blazing fast, modern apps and websites w...	52,062

	user_name	repo_name	repo_desc	repo_stars
18	google	gvisor	Application Kernel for Containers	12,233
19	facebook	react-native	A framework for building native applications u...	100,218
20	hibiken	asynq	Asynq: simple, reliable, and efficient distrib...	2,237
21	alibaba	formily	Alibaba Group Unified Form Solution - Support...	6,878
22	skyline-emu	skyline	Run Nintendo Switch homebrew & games on your A...	2,531
23	vnpy	vnpy	基于Python的开源量化交易平台开发框架	17,199

Converting data into csv file

In [134...

```
trend_repo.to_csv('github_trending_repo.csv',index=False)
```

In []: