



HO CHI MINH UNIVERSITY OF SCIENCE
FACULTY OF INFORMATION TECHNOLOGY

Thu thập dữ liệu

Đồ án 1

Người thực hiện:

1. Ngô Đăng Khoa - 19127444
2. Huỳnh Cao Nhật Hiếu – 19127399
3. Nguyễn Gia Hân – 19127134
4. Nguyễn Khắc Vỹ - 19127637





Contents	
I. HTML.....	2
1. Tóm tắt phương pháp thu thập.....	2
2. Kết quả thu thập dữ liệu.....	2
3. Thống kê đơn giản các tập dữ liệu tìm được.....	3
3.1 Users	3
3.2 Tracks	4
3.3 Playlists.....	5
II. API.....	7
1. Nội dung:	7
2. Đối tượng:	7
3. Ý tưởng:	7
4. Tính năng, đặc điểm:.....	7
4.1 URL:	7
4.2 Playlist:	8
4.3 Tracks:	9
4.4 User:.....	11
4.5 FILE CSV:	13
III. Các thư viện trong đồ án	14
1. Selenium	14
2. BeautifulSoup	14
3. Time	14
4. Pandas	14

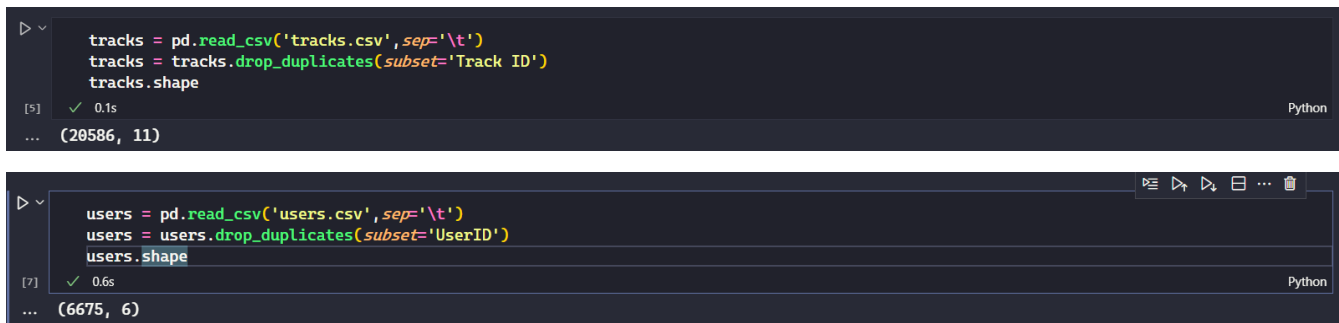
I. HTML

1. Tóm tắt phương pháp thu thập

- Tạo ra 3 hàm để crawl dữ liệu lần lượt cho Playlist, Track và User theo url của mỗi loại.
- Đầu tiên truy cập vào trang soundcloud.com/discover. Vì trang này dễ truy cập cho bước đầu cào dữ liệu, chứa nhiều các playlist được cập nhật liên tục. Ở trang này ta có thể crawl được các url của mỗi playlist.
- Từ những playlist url đã thu thập ở bước đầu, ta sẽ có được thông tin của playlist cũng như track và user trong mỗi playlist thông qua hàm crawl data cho playlist.
- Tiếp tục thu thập dữ liệu cho User và Track theo cách trên thông qua các hàm đã tạo cho mỗi entity.
- Riêng hàm của User có chút đặc biệt bởi vì có thể get thêm url của track, playlist do chính user đó tạo ra và những user mới trong danh sách following users.
- Chúng ta sẽ liên tục chạy lần lượt các bước trên ứng với mỗi cell trong Jupyter Notebook đến khi nào đủ dữ liệu thì dừng. Ta cũng có thể lưu lại danh sách các Url chưa crawl để cho những lần tiếp theo mà không cần quay lại trang Discover như ban đầu.
- Bước cuối cùng ta cần lọc lại cái phần tử duplicate theo ID và lưu lại thành các file csv.

2. Kết quả thu thập dữ liệu

- Đối với quá trình parse HTML, nhóm đã thu được 1194 playlists, 20856 tracks và 6675 users.



```
tracks = pd.read_csv('tracks.csv', sep='\t')
tracks = tracks.drop_duplicates(subset='Track ID')
tracks.shape
```

[5] ✓ 0.1s
... (20586, 11)

```
users = pd.read_csv('users.csv', sep='\t')
users = users.drop_duplicates(subset='UserID')
users.shape
```

[7] ✓ 0.6s
... (6675, 6)

- Dữ liệu track có tổng cộng 11 features, playlist cũng có 11 features và user có 6 features.
- Mặc dù parse được nhiều feature hơn nhưng những feature đó đa số có giá trị là rỗng hoặc rác hoặc không có gì nên nhóm đã không để vào trong danh sách feature tìm kiếm.

3. Thống kê đơn giản các tập dữ liệu tìm được

3.1 Users

- Thu thập được tổng cộng là 6675 users với 6 features.

```
users = pd.read_csv('users.csv', sep='\t')
users = users.drop_duplicates(subset='UserID')
users = users.reset_index(drop=True)
users.shape
```

[4] ✓ 0.6s Python

... (6675, 6)

- Số followers trung bình của user là 99128 và trung vị số followers của user là 2847.

```
print("The average amount of followers of the user: ", int(np.mean(users['followers_count'])))
print("The median of followers of the user: ", int(np.median(users['followers_count'])))
```

[6] ✓ 0.3s Python

... The amount of followers of the user: 99128
The median of followers of the user: 2847

- Số Playlist trung bình của user là 99128 và trung vị số Playlist của user là 2847.

```
print("The average amount of Playlist count of the user: ", int(np.mean(users['Playlist count'])))
print("The median of Playlist count of the user: ", int(np.median(users['Playlist count'])))
```

[13] ✓ 0.7s Python

... The amount of Playlist count of the user: 16
The median of Playlist count of the user: 2

- Số Track trung bình của user là 178 và trung vị số Track của user là 29.

```
print("The average amount of track count of the user: ", int(np.mean(users['track count'])))
print("The median of track count of the user: ", int(np.median(users['track count'])))
```

[17] ✓ 0.2s Python

... The amount of track count of the user: 178
The median of track count of the user: 29

- Thông tin user có số lượng followers lớn nhất:

```
index = users[users['followers_count'] == np.max(users['followers_count'])].index.values
print('The users have the the largest followers: \n')
for i in index:
    print(users.iloc[i])
```

[47] ✓ 0.3s Python

... The users have the the largest followers:

UserID	7436630
username	Future
permalink_url	https://soundcloud.com/futureisnow
followers_count	9973283
Playlist count	50
track count	586
Name: 3609, dtype: object	

Đồ án 1 - Thu thập dữ liệu



- Thông tin user có số lượng track lớn nhất:

```
index1 = users[users['track count'] == np.max(users['track count'])].index.values
print('The users have the the largest track: \n')
for i in index1:
    print(users.iloc[i])
```

[48] ✓ 0.8s Python

... The users have the the largest track:

UserID	2117244
username	Primedia Broadcasting
permalink_url	https://soundcloud.com/primediabroadcasting
followers_count	5596
Playlist count	159
track count	40410

Name: 3301, dtype: object

- Thông tin user có số lượng playlist lớn nhất:

```
index2 = users[users['Playlist count'] == np.max(users['Playlist count'])].index.values
print('The users have the the largest Playlist: \n')
for i in index2:
    print(users.iloc[i])
```

[15] ✓ 0.4s Python

... The users have the the largest Playlist:

UserID	35449466
username	VOAChina
permalink_url	https://soundcloud.com/voachina
followers_count	4815
Playlist count	5759
track count	29721

Name: 2987, dtype: object

3.2 Tracks

- Thu thập được tổng cộng 20586 tracks với 11 features.

```
tracks = pd.read_csv('tracks.csv', sep='\t')
tracks = tracks.drop_duplicates(subset='Track ID')
tracks.shape
```

[2] ✓ 0.2s Python

... (20586, 11)

+ Code + Markdown

- Số like trung bình và trung vị của track lần lượt là 21899 và 1638.

```
print("The average amount of like of the track: ", int(np.nanmean(tracks['Likes count'])))
print("The median of like of the tracks: ", int(np.nanmedian(tracks['Likes count'])))
```

[46] ✓ 0.4s Python

... The average amount of like of the track: 21899
The median of like of the tracks: 1638

Đồ án 1 - Thu thập dữ liệu



- Thông tin track có số lượt like nhiều nhất:

```
index2 = tracks[tracks['Likes count'] == np.max(tracks['Likes count'])].index.values
print('The track has the the greatest Likes: \n')
for i in index2:
    print(tracks.iloc[i])
```

[38] ✓ 0.4s Python

... The track has the the greatest Likes:

Track ID	247036982
Title	Miguel - Sure Thing (Andrey Azizov Remix)
Genre	Remix
Track url	https://soundcloud.com/andrey-azizov/miguel-su...
Created at	2016-02-14T18:57:49Z
Likes count	55876.0
Reposts count	6738
playback_count	3469992.0
Creator	Andrey Azizov 2
Creator ID	3738399
Creator url	https://soundcloud.com/andrey-azizov
Name: 547, dtype: object	

- Top 5 tracks có số lượt like nhiều nhất:

```
tracks.sort_values(by='Likes count', ascending=False).head(5)
```

[40] ✓ 0.5s Python

	Track ID	Title	Genre	Track url	Created at	Likes count	Reposts count	playback_count	Creator
545	193781466	Major Lazer & DJ Snake - Lean On (feat. MØ)	Major Lazer	https://soundcloud.com/majorlazer/major-lazer-...	2015-03-02T04:31:20Z	3056127.0	296098	221048937.0	Major Lazer
22357	237603952	Designer- "Panda" (Prod. By: Menace)	Designer	https://soundcloud.com/lifeofdesigner/designer-...	2015-12-15T02:55:22Z	2905986.0	235378	204490328.0	LifeOfDesigner
13916	239793212	Wonderful Ft.The Weeknd	travisscott	https://soundcloud.com/travisscott-2/wonderful...	2015-12-31T05:13:42Z	2056506.0	217877	139844519.0	Travis Scott
2115	250711755	The Chainsmokers - Don't Let Me Down (Illenium...	Dance & EDM	https://soundcloud.com/illeniumofficial/the-ch...	2016-03-07T20:25:55Z	2054966.0	161497	129517905.0	ILLENIUUM
12545	501876756	Lil Peep & XXXTENTACION - Falling Down	Alternative Rock	https://soundcloud.com/lil_peep/lil-peep-ft-xx...	2018-09-19T00:03:50Z	2022430.0	69066	137151135.0	☆Lil PEEP☆

3.3 Playlists

- Dữ liệu thu thập được có tổng cộng 3537 playlists với 11 features.

```
playlists = pd.read_csv('playlists.csv', sep='\t')
playlists.shape
```

[44] ✓ 0.1s Python

... (3537, 11)

Đồ án 1 - Thu thập dữ liệu



- Trung bình và trung vị của tổng số track trong mỗi playlist lần lượt là 9 và 4

```
print("The average amount of track count of the playlist: ", int(np.mean(playlists['Track Count'])))
print("The median of track count of the playlist: ", int(np.median(playlists['Track Count'])))
```

[49] ✓ 0.3s Python

... The average amount of track count of the playlist: 9
The median of track count of the playlist: 4

- Trung bình và trung vị của tổng số lượt like của mỗi playlist lần lượt là 3331 và 104.

```
print("The average amount of like count of the playlist: ", int(np.mean(playlists['Likes count'])))
print("The median of like count of the playlist: ", int(np.median(playlists['Likes count'])))
```

[51] ✓ 0.4s Python

... The average amount of like count of the playlist: 3331
The median of like count of the playlist: 104

- Thông tin playlist có số track nhiều nhất:

```
index4 = playlists[playlists['Track Count'] == np.max(playlists['Track Count']).index.values]
print('The playlist have the the largest Track: \n')
for i in index4:
    print(playlists.iloc[i])
```

[54] ✓ 0.4s Python

... The playlist have the the largest Track:

Playlist ID	21509293
Tiêu đề	Productions / remixes / bootlegs / mashups
Playlist url	https://soundcloud.com/ferry-corsten/sets/prod...
Genre	Trance
Likes count	443
Repost count	131.0
Track Count	226
TrackIDs	540077118 511702599 489436047 464136909 464124 ...
Creator ID	3787492
Creator name	ferry-corsten
user url	https://soundcloud.com/ferry-corsten
Name: 2442, dtype: object	

- Thông tin playlist có số lượt like nhiều nhất:

```
index4 = playlists[playlists['Likes count'] == np.max(playlists['Likes count']).index.values]
print('The playlist have the the largest likes: \n')
for i in index4:
    print(playlists.iloc[i])
```

[55] ✓ 0.3s Python

... The playlist have the the largest likes:

Playlist ID	228667052
Tiêu đề	Coloring Book
Playlist url	https://soundcloud.com/chancetherapper/sets/co...
Genre	Hip-hop & Rap
Likes count	360030
Repost count	63663.0
Track Count	14
TrackIDs	266129731 266129708 266129639 266131961 266129 ...
Creator ID	6969243
Creator name	"Chance The Rapper"
user url	https://soundcloud.com/chancetherapper
Name: 966, dtype: object	

II. API

1. Nội dung:

- Thu nhập các thông tin từ trang <https://soundcloud.com/> bằng API.

2. Đối tượng:

- Cụ thể là sẽ lấy dữ liệu về **playlist id**, **user id** và **track id** trên trang discovery (trang chủ của soundcloud).

3. Ý tưởng:

- Trước tiên, *scroll page* để web tải hết nội dung.
- Sau đó rồi dùng *driver.pagesource* bỏ vào *beautifulsoup* để parse dữ liệu playlist bằng html.
 - ➔ Mục đích để lấy đường link của tất cả các playlist trong trang discover của SoundCloud, từ đó lấy được id của playlist và một số id của những tracks nằm trong playlist do SoundCloud tạo (không có id playlist).
- Sau khi có id playlist, dùng API để lấy dữ liệu của playlist, trong dữ liệu đó có id của tracks và users.
- Tiếp tục dùng API lấy dữ liệu của tracks, trong dữ liệu đó có id của users.
- Cuối cùng, dùng API lấy dữ liệu của tất cả users mà lấy được id ở phía trên.
- Do đó, các file csv đều sẽ có mối liên hệ với nhau:
 - o File Playlist có id user và id những tracks trong playlist.
 - o File Track có id user tạo ra track đó.
 - o File User sẽ có số lượng tracks, playlist.

4. Tính năng, đặc điểm:

4.1 URL:

- Các đường dẫn get được.

```
['https://soundcloud.com/discover/sets/charts-top:indie:vn',  
'https://soundcloud.com/discover/sets/charts-trending:reggaeton:vn',  
'https://soundcloud.com/gravitation17/sets/study',  
'https://soundcloud.com/discover/sets/charts-top:religionspirituality:vn',  
'https://soundcloud.com/yourparadis/sets/x9uhnyawjkmx',  
'https://soundcloud.com/scott-asher/sets/eat-sleep-deep-repeat',  
'https://soundcloud.com/discover/sets/charts-top:comedy:vn',  
'https://soundcloud.com/hanriverlove/sets/winter-night-jazz-2',  
'https://soundcloud.com/user-221702713/sets/music-for-concentration',  
'https://soundcloud.com/discover/sets/charts-trending:dancehall:vn',  
'https://soundcloud.com/thenoisetier/sets/last-days-of-spring',  
'https://soundcloud.com/yeonkkot/sets/s12',  
'https://soundcloud.com/mohammad-shuvo-910800399/sets/english',  
'https://soundcloud.com/discover/sets/charts-top:dubstep:vn',  
'https://soundcloud.com/ilyanaazman/sets/anime-trap',  
'https://soundcloud.com/matt-wilkinson-4/sets/greek-week-1',  
'https://soundcloud.com/discover/sets/charts-trending:folksingersongwriter:vn',  
'https://soundcloud.com/lowestends/sets/weekend-workout-vol-4',  
'https://soundcloud.com/discover/sets/charts-top:metal:vn',
```


Đồ án 1 - Thu thập dữ liệu



4.2 Playlist:

- Thông tin playlists được lấy về từ trang discovery.
- Bao gồm 211 playlists và 32 thuộc tính.
- Thời gian chạy là khoảng 30 phút.
- Danh sách sẽ bao gồm url, uri, playlist id, thể loại, lượt nghe và cụ thể hơn là những track nằm trong playlist đó, ngày khởi tạo và hashtag,...
- Tất cả đặc tính sẽ được hiển thị trong notebook và file csv.

	artwork_url	created_at	description	duration	embeddable_by	genre	id	kind	label_name	last_modified	...
0	https://i1.sndcdn.com/artworks-000203205231-is...	2017-01-16T14:08:37Z	NaN	9136912	all	study	292217076	playlist	NaN	2018-03-01T07:52:48Z	...
1	https://i1.sndcdn.com/artworks-000390803682-1o...	2018-08-18T07:52:15Z	NaN	7133523	all	NaN	584936301	playlist	NaN	2019-02-15T19:22:52Z	...
2	https://i1.sndcdn.com/artworks-000109962512-uw...	2015-03-14T05:36:14Z	deep house x electronic x g house	6295891	all	deep house	89199012	playlist	NaN	2015-03-14T05:38:43Z	...
3	https://i1.sndcdn.com/artworks-000628460320-d8...	2019-11-01T02:40:33Z	NaN	6824079	all	Jazz & Blues	912910993	playlist	NaN	2019-12-05T10:48:11Z	...
4	https://i1.sndcdn.com/artworks-000179849277-6c...	2016-08-26T13:18:35Z	Declutter your mind and let the creativity flo...	4649239	all	NaN	253737934	playlist	NaN	2016-08-31T21:38:41Z	...

by_feeds	permalink	permalink_url	public	purchase_title	purchase_url	release_date	reposts_count	secret_token	sharing	tag_list
False	study	https://soundcloud.com/gravitation17/sets/study	True	NaN	NaN	2017-01-16T00:00:00Z	584	NaN	public	NaN
False	x9uhnyawjkmx	https://soundcloud.com/yourparadis/sets/x9uhny...	True	NaN	NaN	NaN	404	NaN	public	Sleep Relax
False	eat-sleep-deep-repeat	https://soundcloud.com/scott-asher/sets/eat-sl...	True	NaN	NaN	NaN	3547	NaN	public	electronic deep house ghouser deeper "kanye wes...
False	winter-night-jazz-2	https://soundcloud.com/hanriverlove/sets/winte...	True	NaN	NaN	NaN	963	NaN	public	Jazz mood winter read study 재즈 독서

title	uri	user_id	set_type	is_album	published_at	display_date	track_count	trackids
STUDYQ	https://api.soundcloud.com/playlists/292217076	282160952	NaN	False	2017-01-16T14:08:37Z	2017-01-16T14:08:37Z	13	134679527,135549659,130781881,61164815,1097136...
paradise	https://api.soundcloud.com/playlists/584936301	283347384	NaN	False	2018-08-18T08:12:33Z	2018-08-18T08:12:33Z	33	458988819,423958416,372337427,381831449,399075...
Repeat.	https://api.soundcloud.com/playlists/89199012	18041387	NaN	False	NaN	2015-03-14T05:36:14Z	23	194192478,196102820,196219909,195191484,196347...
Jazz #2	https://api.soundcloud.com/playlists/912910993	552289491	NaN	False	2019-11-02T14:54:03Z	2019-11-02T14:54:03Z	25	128157554,34522281,98219350,146026169,47708432...
Music For entration	https://api.soundcloud.com/playlists/253737934	249812102	NaN	False	2016-08-26T13:18:35Z	2016-08-26T13:18:35Z	20	263153662,43047740,256140609,172884842,2602404...



4.3 Tracks:

- Thông tin của tracks nằm trong các playlist trên và các playlist do SoundCloud tạo.
- Bao gồm 11447 tracks và 44 thuộc tính.
- Thời gian chạy là khoảng 5 tiếng.
- Danh sách sẽ bao gồm url, uri, track id, thể loại, lượt nghe, lượt bình luận, số lần repost,... và cụ thể hơn là những lượt tải, mô tả, user tạo track, playlist của track, ngày khởi tạo và hashtag,...
- Tất cả đặc tính sẽ được hiển thị trong notebook và file csv.

	artwork_url	caption	commentable	comment_count	created_at	description	downloadable	download_count
0	https://i1.sndcdn.com/artworks-000079395193-br...	NaN	True	702.0	2014-05-14T08:52:11Z	Hi I'm OOHYO and this is my first EP album, Gi...	False	0.0
1	https://i1.sndcdn.com/artworks-000219793663-mq...	NaN	True	3852.0	2017-04-26T15:43:01Z	Spotify: https://open.spotify.com/track/7wOqBO...	False	0.0
2	https://i1.sndcdn.com/artworks-000193629987-23...	NaN	True	2339.0	2016-11-15T05:15:51Z	From Cigarettes After Sex LP out now:\nhttps:/...	False	0.0
3	https://i1.sndcdn.com/artworks-000149324123-9w...	NaN	True	5965.0	2016-03-02T05:34:25Z	A slowed version of The Neighbourhood's Sweate...	False	0.0
4	https://i1.sndcdn.com/artworks-000228331152-m3...	NaN	True	1793.0	2017-06-14T15:28:57Z	Spotify: https://open.spotify.com/track/61RsA1...	False	0.0
...
11466	https://i1.sndcdn.com/artworks-000245051435-n2...	NaN	True	608.0	2017-04-19T15:59:04Z	Out now via Deadbeats: http://smarturl.it/DNMO...	False	0.0
11467	https://i1.sndcdn.com/artworks-000327043275-xx...	NaN	True	143.0	2018-03-29T17:00:33Z	when the opportunity came up to remix a record...	False	0.0
11468	NaN	NaN	True	19.0	2014-05-20T12:07:37Z	Trap Monster Official is proud to bring you th...	False	0.0

Đồ án 1 - Thu thập dữ liệu



t	id	kind	label_name	last_modified	license	likes_count	permalink	permalink_url	playback_count	public	purcha
3	149461930	track	XMG	2021-08-24T06:43:34Z	all-rights-reserved	21050.0	06-teddy-bear-rises	https://soundcloud.com/excellentmg/06-teddy-be...	1410568.0	True	
3	319583871	track	NaN	2021-11-11T13:15:43Z	all-rights-reserved	139652.0	rumors	https://soundcloud.com/neffexmusic/rumors	10828490.0	True	Free Dc
3	293095092	track	NaN	2021-11-11T17:55:56Z	all-rights-reserved	160167.0	k-3	https://soundcloud.com/cigarettesaftersex/k-3	8150918.0	True	
3	249782235	track	NaN	2021-11-11T18:37:54Z	all-rights-reserved	144213.0	the-neighbourhood-sweater-weather-slowed	https://soundcloud.com/user-578957262/the-neig...	8240705.0	True	
3	328026317	track	NaN	2021-10-17T14:24:22Z	all-rights-reserved	72294.0	careless	https://soundcloud.com/neffexmusic/careless	5527423.0	True	Free Dc
...
3	318478158	track	Deadbeats	2021-10-21T03:28:28Z	all-rights-reserved	56033.0	broken	https://soundcloud.com/iamdnmo/broken	3117116.0	True	Stream/Dc
3	421911600	track	NaN	2021-04-08T18:29:21Z	all-rights-reserved	19628.0	feelitremix	https://soundcloud.com/vincent_official/feelit...	850307.0	True	s p
3	150401647	track	NaN	2021-04-08T18:29:21Z	all-rights-reserved	840.0	nero-satisfy-simeon-trap-remix	https://soundcloud.com/trapmonsterofficial/ner...	61126.0	True	

state	streamable	tag_list	title	track_format	uri	urn	user_id
shed	True	"Electronic Pop" K-Pop Indie "Indie Pop" Korea...	OOHYO(우효) - 06 Teddy Bear Rises	single-track	https://api.soundcloud.com/tracks/149461930	soundcloud:tracks:149461930	90722996 https://wave.sndcd
shed	True	"Alternative Rock" Rock Trap Bass EDM Epic "No...	Rumors [Available on Spotify and Apple Music]	single-track	https://api.soundcloud.com/tracks/319583871	soundcloud:tracks:319583871	157193072 https://wave.sndcdn
shed	True	NaN	K.	single-track	https://api.soundcloud.com/tracks/293095092	soundcloud:tracks:293095092	19710656 https://wave.sndcd
shed	True	"the neighbourhood" "the neighborhood" "sweate...	The Neighbourhood - Sweater Weather - Slowed	single-track	https://api.soundcloud.com/tracks/249782235	soundcloud:tracks:249782235	208608248 https://wave.sndc
shed	True	Electronic Alternative Indie "Indie Rock" "Ind...	Careless [Available on Spotify and Apple Music...	single-track	https://api.soundcloud.com/tracks/328026317	soundcloud:tracks:328026317	157193072 https://wave.sndcdn
...
shed	True	NaN	DNMO & Sub Urban - Broken	single-track	https://api.soundcloud.com/tracks/318478158	soundcloud:tracks:318478158	83660731 https://wave.sndcdn
shed	True	NaN	KRANE ~ Feel It (feat. CXLOE & Khamsin) [Vince...	single-track	https://api.soundcloud.com/tracks/421911600	soundcloud:tracks:421911600	22944258 https://wave.sndcd
shed	True	NERO SIMEON TRAPREMIX REMIX TRAP	Nero - Satisfy (Simeon Trap Remix)	single-track	https://api.soundcloud.com/tracks/150401647	soundcloud:tracks:150401647	94217660 https://wave.sndcdr

Flilie Goulding



4.4 User:

- Thông tin các users tạo ra các tracks và playlist trên
- Bao gồm 6815 users và 32 thuộc tính.
- Thời gian chạy là khoảng 3 tiếng.
- Danh sách sẽ bao gồm url, uri, user id, ảnh đại diện, họ và tên,... và cụ thể hơn là những lượt tải, mô tả, ngày khởi tạo,... của các tracks, playlists do user tạo ra.
- Tất cả đặc tính sẽ được hiển thị trong notebook và file csv.

	avatar_url	city	comments_count	country_code	created_at	description	follow
0	https://i1.sndcdn.com/avatars-000332582368-xi6...	NaN	1	NaN	2017-03-28T21:35:35Z	NaN	
1	https://i1.sndcdn.com/avatars-2VdOHXNpTLn7Pww2...	chicago	18	US	2012-11-17T16:26:17Z	Raised on the South Side of Chicago, Vic rose	...
2	https://i1.sndcdn.com/avatars-1FH0LLD3P5Dkiabh...	Toronto	190	CA	2015-06-14T17:41:19Z	Dream State EP out now on all streaming Platfo...	
3	https://a1.sndcdn.com/images/default_avatar_la...	NaN	0	NaN	2019-12-02T11:16:41Z	NaN	
4	https://i1.sndcdn.com/avatars-yLSykC9ixLmgDfwN...	NaN	8	NL	2017-03-27T18:10:59Z	Calundé is the project of The Netherlands base...	
...
6823	https://i1.sndcdn.com/avatars-IGb20NzhIPOeyzeg...	ATL	94	NaN	2015-09-16T22:40:20Z	swag lol.	
6824	https://i1.sndcdn.com/avatars-000503672100-xtj...	Philadelphia, PA	56	US	2010-12-16T01:57:15Z	BOOKING: greg@datatouring.com\n\nPlatinum Reco...	
6825	https://i1.sndcdn.com/avatars-000206800148-bn6...	NaN	4	NaN	2015-06-06T14:26:18Z	NaN	
6826	https://i1.sndcdn.com/avatars-el17dJwKg3v2S72e...	Detroit/Los Angeles	277	US	2013-04-13T18:23:26Z	Follow me on IG @nolanvanlith \n\nCopyright fr...	
6827	https://i1.sndcdn.com/avatars-000357561461-zuj...	Newcastle Upon Tyne	340	NaN	2012-08-01T12:38:31Z	https://twitter.com/Cristophmusic\n\nhttps://www...	

Đồ án 1 - Thu thập dữ liệu



urn	username	verified	station_urn	station_permalink	visuals.urn	visuals.enabled	visuals.visuals
soundcloud:users:298254346	Sdvar26	False	soundcloud:system-playlists:artist-stations:29...	artist-stations:298254346	soundcloud:users:298254346	True	[[{"urn": "soundcloud:visuals:34195245", "entry...
soundcloud:users:28508171	Vic Mensa	True	soundcloud:system-playlists:artist-stations:28...	artist-stations:28508171	soundcloud:users:28508171	True	[[{"urn": "soundcloud:visuals:146210187", "entr...
soundcloud:users:158105618	Eric Ryan 🌞	True	soundcloud:system-playlists:artist-stations:15...	artist-stations:158105618	soundcloud:users:158105618	True	[[{"urn": "soundcloud:visuals:103502376", "entr...
soundcloud:users:742883347	Maty DeeJay	False	NaN	soundcloud:system-playlists:artist-stations:74...	artist-stations:742883347	NaN	NaN
soundcloud:users:297992211	CALUNDÉ	False	soundcloud:system-playlists:artist-stations:29...	artist-stations:297992211	soundcloud:users:297992211	True	[[{"urn": "soundcloud:visuals:145586145", "entr...
...
soundcloud:users:174817271	Lil Nas X	True	soundcloud:system-playlists:artist-stations:17...	artist-stations:174817271	soundcloud:users:174817271	True	[[{"urn": "soundcloud:visuals:61557448", "entry...
soundcloud:users:2424825	Wes Walker	True	soundcloud:system-playlists:artist-stations:24...	artist-stations:2424825	soundcloud:users:2424825	True	[[{"urn": "soundcloud:visuals:53677408", "entry...
soundcloud:users:156794876	Edlameth	False	soundcloud:system-playlists:artist-stations:15...	artist-stations:156794876	soundcloud:users:156794876	True	[[{"urn": "soundcloud:visuals:4978320", "entry_...
soundcloud:users:41779197	Nolan van Lith	True	soundcloud:system-playlists:artist-stations:41...	artist-stations:41779197	soundcloud:users:41779197	True	[[{"urn": "soundcloud:visuals:148430592", "entr...
soundcloud:users:21299199	Cristophmusic	True	soundcloud:system-playlists:artist-stations:21...	artist-stations:21299199	soundcloud:users:21299199	True	[[{"urn": "soundcloud:visuals:136502745", "entr...

ngs_count	first_name	full_name	groups_count	id	kind	last_modified	last_name	likes_count	playlist_likes_count	permalink
2	NaN	NaN	0	298254346	user	2017-09-01T16:16:00Z	NaN	4	0	user-424428524 https://soundcloud.com/sdvar26
9	Vic	Vic Mensa	0	28508171	user	2021-11-02T00:00:01Z	Mensa	16	1	vicsavemoney https://soundcloud.com/vicsavemoney
42	Eric	Eric Ryan	0	158105618	user	2021-05-29T20:12:26Z	Ryan	40	2	ericryanpascual https://soundcloud.com/ericryanpascual
0	NaN	NaN	0	742883347	user	2019-12-02T11:16:41Z	NaN	0	0	matydeejay-music https://soundcloud.com/matydeejay-music
30	NaN	NaN	0	297992211	user	2021-09-15T08:44:35Z	NaN	9	2	calunde https://soundcloud.com/calunde
...
19	Lil Nas	Lil Nas X	0	174817271	user	2021-09-17T04:01:57Z	X	15	1	secret-service-862007284 https://soundcloud.com/secret-service-862007284
551	wes	wes walker	0	2424825	user	2020-11-19T16:37:07Z	walker	332	15	weswalker https://soundcloud.com/weswalker
12	Edouard	Edouard Dion	0	156794876	user	2019-11-22T18:00:02Z	Dion	28	6	edouard-dion https://soundcloud.com/edouard-dion
76	Nolan	Nolan van Lith	0	41779197	user	2021-10-18T23:43:16Z	van Lith	116	16	nolanvanlith https://soundcloud.com/nolanvanlith
136	Cristoph	Cristoph	0	21299199	user	2021-08-23T16:51:53Z	NaN	44	5	cristophmusic https://soundcloud.com/cristophmusic

● ● ●

- **Playlist.csv:**

- **Track.csv:**

13

● ● ●

[illegible]

1. Selenium

- ## 2. BeautifulSoup

- ### 3. Time

- ## 4. Pandas

- 14