

BÁI CÁO ĐỒ ÁN 1

1/ Bảng phân công công việc:

Tên thành viên	MSSV	Công việc	Tỉ lệ
Lê Tâm Anh	19127330	Parse HTML để lấy thông tin Users Lấy thông tin Playlists bằng API	33%
Trần Minh Thiện	19127281	Parse HTML để lấy thông tin Tracks Lấy thông tin Users bằng API	33%
Vương Thành An	19127326	Parse HTML để lấy thông tin Playlists Lấy thông tin Tracks bằng API	34%

2/ Giải thích các dữ liệu thu thập được:

API:

	Số mẫu	Số cột dữ liệu
User.csv	2000	38
Track.csv	2004	64
Playlist.csv	1982	31

GIẢI THÍCH THUỘC TÍNH :

User:

"avatar_url": link đến ảnh đại diện user

"city": nơi ở

"comments_count": số lượng bình luận

"country_code": mã quốc gia

"created_at": ngày tài khoản được tạo

"creator_subscriptions": [(loại tài khoản được tạo)

{

"product": {

"id": "creator-pro-unlimited"

}

}

],

"creator_subscription": {

"product": {

"id": "creator-pro-unlimited"

}

},

"description": chú thích thêm về user

"followers_count": số lượng người theo dõi user

"followings_count": số lượng người user theo dõi

"first_name": tên user

"full_name": tên đầy đủ

"groups_count": số lượng nhóm của user

"id": id của user

"kind": thể loại (user or track or playlist)
"last_modified": lần cuối chỉnh sửa tài khoản
"last_name": họ user
"likes_count": lượt like,
"playlist_likes_count": lượt like playlist,
"permalink": tên của liên kết tĩnh
"permalink_url": liên kết tĩnh
"playlist_count": số lượng playlist
"reposts_count": số lượng bài đăng lại
"track_count": số lượng track
"uri": uri của user
"urn": urn của user
"username": tên của user
"verified": xác minh
"visuals": {
 "urn": "soundcloud:users:1123",
 "enabled": true,
 "visuals": [(ảnh bìa)
 {
 "urn": "soundcloud:visuals:27143126" (urn của ảnh bìa),
 "entry_time": 0,
 "visual_url": <https://i1.sndcdn.com/visuals-000000001123-UXqYsN-original.jpg>
(link tới ảnh bìa)
 }
],
 "tracking": null

```
},  
"badges": { (các loại danh hiệu mà tài khoản này sở hữu)  
"pro": false,  
"pro_unlimited": true,  
"verified": xác nhận user  
},  
"station_permalink": "artist-stations:1123" (tên cố định station của user)
```

User attribute	Mean	Median
comments_count	193.51	2.0
followers_count	3426.15	1006.5
followings_count	170.11	7.0
groups_count	0.0	0.0
likes_count	348.46	10.0
playlist_count	23.73	3.0
playlist_likes_count	21.16	0.0
track_count	123.66	45.0
reposts_count	0.0	0.0

Track

"artwork_url": Ảnh đại diện của track
"caption": chú thích thêm
"commentable": có thể bình luận hay không
"comment_count": số bình luận trên track này
"created_at": ngày giờ track được tạo,

"description": chú thích thêm về track (người hát, sáng tác....)

"downloadable": có thể tải về hay không

"download_count": số lượt tải về

"duration": độ dài của một nốt nhạc tùy thuộc vào thời gian tồn tại của sóng âm trong không khí

"full_duration": Tổng thời lượng của bản track

"embeddable_by": track có thể nhúng bởi,

"genre": thể loại của bản track

"has_downloads_left": lượt tải về còn lại của track

"id": id của track

"kind": thuộc loại nào (track or user or playlist....)

"label_name": Tên nhãn

"last_modified": track được sửa đổi lần cuối vào ngày giờ năm nào

"license": đã đk bản quyền hay chưa

"likes_count": lượt thích

"permalink": tên của đường dẫn tĩnh đến track

"permalink_url": link tĩnh đến track

"playback_count": số lần được nghe

"public": được đăng công khai hay không

"publisher_metadata": siêu dữ liệu của người đăng

"purchase_title": tên của web bán track

"purchase_url": đường link để mua track

"release_date": ngày phát hành,

"reposts_count": số lần đăng lại

"secret_token": mã thông báo bí mật

"sharing": chế độ chia sẻ (public , private...)

"state": Trạng thái (đã hoàn thành hay chưa hoàn thành...)

"streamable": có phát trực tiếp được không

"tag_list": danh sách các thẻ list chứa track

"title": tên của track,

"track_format": định dạng track,

"uri": uri của track

"urn": urn của track

"user_id": id của người đăng track

"visuals" :trực quan

"waveform_url": url dạng sóng

"display_date": ngày hiển thị,

"media": {

"transcodings": [

{

"url": "https://api-v2.soundcloud.com/media/soundcloud:tracks:12456/53eacc68-06db-4e11-8b59-10fd84802e41/stream/hls",

"preset": định dạng

"duration": 473678,

"snipped": false,

"format": {

"protocol": loại protocol,

"mime_type": phân loại (định dạng) tập tin

},

"quality": chất lượng âm thanh

},

{

"url": "https://api-v2.soundcloud.com/media/soundcloud:tracks:12456/53eacc68-06db-4e11-8b59-10fd84802e41/stream/progressive",

"preset": "mp3_0_0",

"duration": 473678,

"snipped": false,

"format": {

"protocol": "progressive",

"mime_type": "audio/mpeg"

},

"quality": "sq"

}

]

},

"station_urn": định danh tài nguyên station của track trên soundcloud(station là một cách để người nghe tìm ra nhạc mình chưa nghe bao giờ),

"station_permalink": tên tĩnh (cố định) station của track

"track_authorization": Ủy quyền của track

"monetization_model": có áp dụng mô hình kiếm tiền không

"policy": chính sách bảo mật

Track attribute	Mean	Median
duration	1226629.59	715977
likes_count	8.33	1
commentable	0.9	1
downloadable	0.21	0
download_count	1.28	0

reposts_count	1.03	0
---------------	------	---

Play list:

"artwork_url": link đến ảnh đại diện playlist,

"created_at": ngày list được tạo,

"description": chú thích thêm

"duration": độ dài của một nốt nhạc tùy thuộc vào thời gian tồn tại của sóng âm trong không khí.

"embeddable_by": có thể nhúng bởi

"genre": thể loại

"id": id của playlist

"kind": thuộc loại nào (track or user or playlist....)

"label_name": tên nhãn

"last_modified": track được sửa đổi lần cuối vào ngày

"license": đăng ký bản quyền,

"likes_count": lượt thích

"managed_by_feeds": được quản lý bởi nguồn cấp dữ liệu hay không

"permalink": tên cố định trên đường dẫn tĩnh của playlist

"permalink_url": đường dẫn tĩnh của playlist

"public": công khai hay không

"purchase_title": ": tên của web bán list

"purchase_url": đường link để mua list

"release_date": ngày phát hành

"reposts_count": số lần đăng lại

"secret_token": mã thông báo bí mật

"sharing": chế độ chia sẻ

"tag_list": các thẻ được gắn của playlist

"title": tên của playlist

"uri": uri của playlist

"user_id": id của user đăng playlist

"set_type": đặt loại

"is_album": có phải là album không

"published_at": được xuất bản tại

"display_date": ngày hiển thị

"track_count": số lượng track trong playlist

Playlist attribute	Mean	Median
track_count	1.20	0
duration	811648.31	232798
likes_count	1.16	0
reposts_count	0	0

HTML:

	Số mẫu	Số cột dữ liệu
User.csv	1001	8
Track.csv	1011	12
Playlist.csv	1102	10

GIẢI THÍCH THUỘC TÍNH :

User:

Username: tên hiển thị của user

User_link: link đến User

Fullname: Họ Tên của User

Where: Vị trí của user

Followers: số người theo dõi user

Following: số Người user theo dõi

Tracks: số lượng track của user

Playlists: số lượng playlist của user

User attribute	Mean	Median
Followers	71074.17	3764
Following	97.25	4
Tracks	1266.35	112
Playlists	18.69	3

Track

Track url : link của track

Track_name: tên của track

User_name: tên user của track

Comments_count: số lượng comments của track

Play_count: số lượng plays của track

Like_count: số lượng likes của track

Reposts_count: số lượng reposts của track

Date_release: ngày đăng của track

Descript: ghi chú thêm của track

Related: những bài track liên quan

Duration: thời lượng của track

Tag: tag của track

Track attribute	Mean	Median
Comments_count	2.38	0
Play_count	1467.72	25
Like_count	42.05	0
Reposts_count	5.72	0

Play list:

Playlist_url: link đến playlist

Playlist_name: Tên của Playlist

User_name: Tên của User

Track_counts: Số lượng track trong playlist

Duration: thời lượng của playlist

Like counts: số lượng like của playlist

Reposts_count: số lượng reposts của playlist

Date_release: ngày đăng của playlist

Tag: tag của playlist

tracks_name: tên của track trong playlist

Playlist attribute	Mean	Median
Track_count	10.78	3
Like_count	30.65	1
Reposts_count	4.79	0