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 uer

"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 dang 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 soudclound(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": Uỷ 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 |