



Spotify Charts:

Une étude sur les marchés émergents et leur culture musicale



TFE 2024
BI developer - Data Visualization
by Martin Dornic





Spotify en chiffres - 2024

- . Lancé en 2008
- . Leader mondial en service de streaming audio
- . **~600M d'utilisateurs** (mensuellement actif) dont **226M d'abonnés Premium**
- . 28% en Europe, 19% aux USA + Canada (mais la plus haute moyenne journalière)
- . 29% ont 25-34 ans, 26% 18-24 ans
- . Présence dans **184 pays**
- . Plus de **100M de tracks**, 22M de plus d'ici fin 2024 (quasi 1 track/seconde)
- . **11M d'artistes et créateurs**
- . A payé **9Md \$** à ses artistes en 2023 soit **62.8%** de ses revenus annuels



Source de données:



- . Classement journalier ou hebdomadaire (top 200) en nombre de streams (1 stream = min 30s d'écoute)
- . En global (sur les utilisateurs du monde entier) ou par utilisateurs propres à un pays (plus de 70)

Choix du cadre d'étude:

. 15 pays + Global

USA

Europe (France, Turquie)

Afrique (Nigeria, Egypte)

Asie (Pakistan, Inde, Indonésie, Philippines, Vietnam, Taïwan, Japon)

Amérique latine (Brésil, Mexique, Argentine)

. sur les 3 dernières années (156 semaines jusqu'au 22/02/2024)

. 1 ligne Fact table = 1 position dans un chart (à la semaine et au pays correspondant)



Script Python pour le scrapping



```
token = "Bearer BQBm2C7MvU58q9SksRQdzQOHxNzn8FsapWm ... "
agent = 'Mozilla/5.0 (Windows NT 10.0; Win64; x64; rv:123.0) Gecko/20100101 Firefox/123.0'

current_date = date(2024, 2, 22)
for i in range(156):
    url = f'https://charts-spotify-com-service.spotify.com/auth/v0/charts/regional-global-weekly/{str(current_date)}'

    with rq.Session() as s:
        s.headers = {
            'authorization': token,
            'User-Agent': agent
        }

        r = s.get(url)
        print(r.status_code)
        data = r.json()
        sleep(random.randint(2,5))

        df = pd.json_normalize(data.get('entries'))
        artist_meta = pd.json_normalize(pd.json_normalize(df['trackMetadata.artists'])[0])
        df = pd.concat([df, artist_meta], axis=1)
        df['week'] = current_date
        df['region'] = 'Global'

        if i == 0:
            df.to_csv('Global_charts.csv', mode = 'w', index=False)
        else:
            df.to_csv('Global_charts.csv', mode = 'a', header=False, index=False)

    current_date -= timedelta(7)
```



Données récupérées: _charts.csv



Colonnes récupérées pour un pays (après nettoyage et renommage):

rank	prevRank	peakRank	peakDate	weeksOnChart	streak	streams	trend	entryRank	entryDate	trackName	track_ID	image	releaseDate	artist	artist_ID	week	
0	1	1	1	2024-02-15	5	5	41483503	NO_CHANGE	24	2024-01-25	Beautiful Things	6tNQ70jh4OwmPGpYy6R2o9	https://i.scdn.co/image/ab67616d00001e02f2a919...	2024-01-18	Benson Boone	22wbnEMDvgVIAGdFeeK6ET	2024-02-22
1	2	31	2	2024-02-22	2	2	36879384	MOVED_UP	31	2024-02-15	TEXAS HOLD 'EM	7AYmToZ5y67fBjq4RLFbuq	https://i.scdn.co/image/ab67616d00001e026c1836...	2024-02-10	Beyoncé	6vWDO969PvNqNYHIOW5v0m	2024-02-22
2	3	4	3	2024-02-22	2	2	34127981	MOVED_UP	4	2024-02-15	CARNIVAL	3w0w2T288dec0mgeZZqoNN	https://i.scdn.co/image/ab67616d00001e020a31b4...	2024-02-09	¥\$	4xPQFgDA5M2xa0ZGo5ilsv	2024-02-22
3	4	5	2	2023-07-06	55	49	33526124	MOVED_UP	16	2019-08-29	Cruel Summer	1BxfuPKGuaTgP7aM08bdwr	https://i.scdn.co/image/ab67616d00001e02e787cf...	2019-08-23	Taylor Swift	06HL4z0CvFAxyc27GXpf02	2024-02-22
4	5	3	1	2023-11-16	23	23	33039462	MOVED_DOWN	14	2023-09-21	greedy	3rUGC1vUpkDG9CZFHMr1t	https://i.scdn.co/image/ab67616d00001e0222fd80...	2023-09-15	Tate McRae	45dkTj5sMRSjrmBSBeiHym	2024-02-22
...	
31195	196	197	5	2020-02-20	54	2	4994138	MOVED_UP	7	2020-02-13	Intentions (feat. Quavo)	4umIPjkehX1r7uhmGvXiSV	https://i.scdn.co/image/ab67616d00001e027fe4a8...	2020-02-14	Justin Bieber	1uNFoZAHBGtllmzznpCI3s	2021-03-04
31196	197	192	49	2020-11-19	22	22	4982892	MOVED_DOWN	133	2020-10-08	Sofia	7B3z0ySL9Rr0XvZEAjWZzM	https://i.scdn.co/image/ab67616d00001e0233ccb6...	2019-08-02	Clairo	3lOCmX0FuQjFxr8SK7Vqag	2021-03-04
31197	198	196	41	2017-01-12	149	3	4974880	MOVED_DOWN	123	2016-12-29	Photograph	6fxVffaTuwjgEk5h9QyRjy	https://i.scdn.co/image/ab67616d00001e0213b3e3...	2014-06-21	Ed Sheeran	6eUKZXaKkcviiH0Ku9w2n3V	2021-03-04
31198	199	-1	2	2017-03-02	161	1	4962976	RE_ENTRY	132	2017-02-23	Something Just Like This	6RUKPb4LETWmmr3iAEQktW	https://i.scdn.co/image/ab67616d00001e020c13d3...	2017-04-07	The Chainsmokers	69GG8xA162ITqCwzJG5jLp	2021-03-04
31199	200	-1	6	2020-03-19	59	1	4955196	RE_ENTRY	127	2020-01-02	Say So	3Dv1eDb0MEgF93GpLXlucZ	https://i.scdn.co/image/ab67616d00001e02f14aa8...	2019-11-07	Doja Cat	5cj0IJcoR7YOSnhnXOPo5	2021-03-04
31200 rows x 17 columns																	

Rank: classement dans le chart (1 à 200)

PeakRank: plus haut classement atteint pour cette track

PeakDate: date à laquelle ce pic est atteint la première fois

WeeksOnChart: Total des semaines de présence dans le chart

Streak: Total des semaines consécutives

(si égal à WeeksOnChart, la track n'est jamais sortie des charts)

Streams: valeur cible à agrégérer

Trend: 'moved up', 'moved down', 'no change', 'new entry', 're-entry'

entryRank: rank à la première entrée dans le chart

entryDate: date de sa première entrée

Week: date de la semaine correspondante du chart



Spotify

Données additionnelles avec l'API Spotify



Spotify® for Developers

API credentials and access token

```
# Your Spotify API Credentials
client_id = 'my client ID'
client_secret = 'my client secret'

# Function to get Spotify access token
def get_token():
    auth_string = client_id + ":" + client_secret
    auth_bytes = auth_string.encode("utf-8")
    auth_base64 = str(base64.b64encode(auth_bytes), "utf-8")

    url = "https://accounts.spotify.com/api/token"
    headers = {
        "Authorization": "Basic " + auth_base64,
        "Content-Type": "application/x-www-form-urlencoded"
    }
    data = {"grant_type": "client_credentials"}
    result = post(url, headers=headers, data=data)
    json_result = json.loads(result.content)
    token = json_result['access_token']
    return token

def get_auth_header(token):
    return {"Authorization": "Bearer " + token}
```

Fonctions avec l'API Endpoint (url)

```
def get_list_features(token, list):
    url = f'https://api.spotify.com/v1/audio-features?ids={list}'
    headers = get_auth_header(token)
    result = get(url, headers=headers)

    json_result = json.loads(result.content)
    return json_result
```

```
def get_list_artist(token, list_artist_id):
    url = f'https://api.spotify.com/v1/artists?ids={list_artist_id}'
    headers = get_auth_header(token)
    result = get(url, headers=headers)
    json_result = json.loads(result.content)
    return json_result
```



Données de track: _tracks.csv



```
tracks = df[['track_ID']].drop_duplicates(subset=['track_ID'])
tracks
```

	track_ID
0	6tNQ70jh4OwmPGpYy6R2o9
1	7AYmToZ5y67fBjq4RLFbuq
2	3w0w2T288dec0mgeZZqoNN
3	1BxfuPKGuaTgP7aM0Bbdwr
4	3rUGC1vUpkDG9CZFHMu1t
...	...
31184	5gEUDNQvoQjdjklrwPdGwD
31192	6FZDfxM3a3UCqtzo5pxSLZ
31194	26w9NTiE9NGjW1ZvI0d1So
31195	4umIPjkehX1r7uhmGvXiSV
31196	783z0ySL9Rr0XvZEAjWZzM

2677 rows × 1 columns

```
token = get_token()

for i in range(0,2700,100):
    list_id = tracks.iloc[i:i+100, 0].tolist()
    list_id = ','.join(list_id)
    result = get_list_features(token, list_id)
    df_tracks = pd.json_normalize(result['audio_features'])
    sleep(random.randint(5,10))

    if i == 0:
        df_tracks.to_csv('Global_tracks.csv', mode = 'w', index=False)
    else:
        df_tracks.to_csv('Global_tracks.csv', mode = 'a', header=False, index=False)
```

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	type	id	uri	track_href	analysis_url	duration_ms	time_signature
0	0.472	0.471	10	-5.692	1	0.0603	0.151	0.000000	0.1400	0.219	105.029	audio_features	6tNQ70jh4OwmPGpYy6R2o9	spotify:track:6tNQ70jh4OwmPGpYy6R2o9	https://api.spotify.com/v1/tracks/6tNQ70jh4Owm...	https://api.spotify.com/v1/audio-analysis/6tNQ...	180304	3
1	0.725	0.709	2	-6.514	1	0.0720	0.588	0.000000	0.1350	0.353	110.024	audio_features	7AYmToZ5y67fBjq4RLFbuq	spotify:track:7AYmToZ5y67fBjq4RLFbuq	https://api.spotify.com/v1/tracks/7AYmToZ5y67f...	https://api.spotify.com/v1/audio-analysis/7AYm...	235636	4
2	0.594	0.811	1	-5.746	1	0.1590	0.189	0.000000	0.3390	0.311	148.144	audio_features	3w0w2T288dec0mgeZZqoNN	spotify:track:3w0w2T288dec0mgeZZqoNN	https://api.spotify.com/v1/tracks/3w0w2T288dec...	https://api.spotify.com/v1/audio-analysis/3w0w...	264324	4
3	0.552	0.702	9	-5.707	1	0.1570	0.117	0.000021	0.1050	0.564	169.994	audio_features	1BxfuPKGuaTgP7aM0Bbdwr	spotify:track:1BxfuPKGuaTgP7aM0Bbdwr	https://api.spotify.com/v1/tracks/1BxfuPKGuaTg...	https://api.spotify.com/v1/audio-analysis/1Bxf...	178427	4
4	0.750	0.733	6	-3.180	0	0.0319	0.256	0.000000	0.1140	0.844	111.018	audio_features	3rUGC1vUpkDG9CZFHMu1t	spotify:track:3rUGC1vUpkDG9CZFHMu1t	https://api.spotify.com/v1/tracks/3rUGC1vUpkDG...	https://api.spotify.com/v1/audio-analysis/3rUG...	131872	1
...
2672	0.761	0.296	0	-10.220	1	0.0484	0.551	0.000492	0.1340	0.579	127.938	audio_features	5gEUDNQvoQjdjklrwPdGwD	spotify:track:5gEUDNQvoQjdjklrwPdGwD	https://api.spotify.com/v1/tracks/5gEUDNQvoQjd...	https://api.spotify.com/v1/audio-analysis/5gEU...	167867	4
2673	0.752	0.488	6	-7.050	1	0.0705	0.297	0.000009	0.0936	0.533	136.041	audio_features	6FZDfxM3a3UCqtzo5pxSLZ	spotify:track:6FZDfxM3a3UCqtzo5pxSLZ	https://api.spotify.com/v1/tracks/6FZDfxM3a3UC...	https://api.spotify.com/v1/audio-analysis/6FZD...	201661	4
2674	0.883	0.600	7	-7.188	1	0.3700	0.114	0.000008	0.0931	0.691	113.903	audio_features	26w9NTiE9NGjW1ZvI0d1So	spotify:track:26w9NTiE9NGjW1ZvI0d1So	https://api.spotify.com/v1/tracks/26w9NTiE9NGj...	https://api.spotify.com/v1/audio-analysis/26w9...	156430	4
2675	0.806	0.546	9	-6.637	1	0.0575	0.300	0.000000	0.1020	0.874	147.986	audio_features	4umIPjkehX1r7uhmGvXiSV	spotify:track:4umIPjkehX1r7uhmGvXiSV	https://api.spotify.com/v1/tracks/4umIPjkehX1r...	https://api.spotify.com/v1/audio-analysis/4uml...	212867	4
2676	0.744	0.619	0	-9.805	1	0.0390	0.598	0.003720	0.2310	0.641	112.997	audio_features	783z0ySL9Rr0XvZEAjWZzM	spotify:track:783z0ySL9Rr0XvZEAjWZzM	https://api.spotify.com/v1/tracks/783z0ySL9Rr0...	https://api.spotify.com/v1/audio-analysis/783z...	188387	4

2677 rows × 18 columns



Spotify

Artiste et Genre: `_artists.csv`, `genres.csv`



Spotify® for Developers

```
artist = df[['artist_ID', 'artist']].drop_duplicates(subset=['artist_ID'])
artist
```

	artist_ID	artist
0	22wbnEMDvgVIAGdFeek6ET	Benson Boone
1	6vWDO969PvNqNYHIOW5v0m	Beyoncé
2	4xPQFgDA5M2xa0ZGo5ilsv	¥\$
3	06HL4z0CvFAxyc27GXpf02	Taylor Swift
4	45dkTj5sMRSjrmBSBeiHym	Tate McRae
...
30981	6i3DxIIAqnDkwELLw4aVrx	Lele Pons
30994	2kCcBybjI3SATlcwdWpUe3	Lil Peep
30995	1KP6TWI40m7p3QBTU6u2xo	BØRNS
31180	0vRvGUQVUjytro0xbpb26bs	Arizona Zervas
31196	3l0CmX0FuQjFxr8SK7Vqag	Clairo

```
token = get_token()

for i in range(0,850,50):
    list_artist_id = ','.join(artist.iloc[i:i+50, 0].tolist())
    result = get_list_artist(token, list_artist_id)

    df_artist = pd.json_normalize(result['artists'])
    df_artist = df_artist.explode('genres')
    sleep(random.randint(5,10))

    if i == 0:
        df_artist.to_csv('Global_artists.csv', mode = 'w', index=False)
    else:
        df_artist.to_csv('Global_artists.csv', mode = 'a', header=False, index=False)
```

```
dim_artist = pd.read_csv('Global_artists.csv')
dim_artist.groupby('id').first().to_csv('Global_genres.csv', index=True)
pd.read_csv('Global_genres.csv')
```

	id	genres	href	images	name	popularity	type	uri	external_urls.spotify	followers.href	followers.total
0	00FQb4jTyendYWaN8pK0wa	art pop	https://api.spotify.com/v1/artists/00FQb4jTyen...	 [{"height": 640, "url": "https://i.scdn.co/ima..."}]	Lana Del Rey	90	artist	spotify:artist:00FQb4jTyendYWaN8pK0wa	https://open.spotify.com/artist/00FQb4jTyendYWaN8pK0wa	NaN	32720361
1	01XYiBYaoMJcNhPokrg0I0	k-pop	https://api.spotify.com/v1/artists/01XYiBYaoMJ...	 [{"height": 640, "url": "https://i.scdn.co/ima..."}]	STAYC	62	artist	spotify:artist:01XYiBYaoMJcNhPokrg0I0	https://open.spotify.com/artist/01XYiBYaoMJcNhPokrg0I0	NaN	1745424
2	01veRCzqGIkUMo7L5PyBAD	NaN	https://api.spotify.com/v1/artists/01veRCzqGIk...	 [{"height": 640, "url": "https://i.scdn.co/ima..."}]	Kauai45	48	artist	spotify:artist:01veRCzqGIkUMo7L5PyBAD	https://open.spotify.com/artist/01veRCzqGIkUMo7L5PyBAD	NaN	27338
3	027TpXKGwdXP7iwbjUSpV8	chicago indie	https://api.spotify.com/v1/artists/027TpXKGwdX...	 [{"height": 640, "url": "https://i.scdn.co/ima..."}]	The Walters	71	artist	spotify:artist:027TpXKGwdXP7iwbjUSpV8	https://open.spotify.com/artist/027TpXKGwdXP7iwbjUSpV8	NaN	646376
4	048LktY5zMnakWq7PTtFrz	afro r&b	https://api.spotify.com/v1/artists/048LktY5zMn...	 [{"height": 640, "url": "https://i.scdn.co/ima..."}]	CKay	67	artist	spotify:artist:048LktY5zMnakWq7PTtFrz	https://open.spotify.com/artist/048LktY5zMnakWq7PTtFrz	NaN	1419563



Modélisation: table FACT_CHART (temp)

Réunification des _charts.csv en une seule table Fact:

```
merged_fact = pd.concat(map(pd.read_csv, glob.glob('*_charts.csv'))), ignore_index=True)
```

	rank	prevRank	peakRank	peakDate	weeksOnChart	streak	streams	trend	entryRank	entryDate	trackName	track_ID	image	releaseDate	artist	artist_ID	week	region
0	1	1	1	2024-01-25	7	7	4530846	NO_CHANGE	7	2024-01-11	HOLA PERDIDA	7JKptWv4YXmievld0HajID	https://iscdn.co/image/ab67616d00001e02546d66...	2024-02-02	Luck Ra	4kcQWQDK0u9AftVSpdrAgk	2024-02-22	Argentina
1	2	2	1	2024-01-18	7	7	3912379	NO_CHANGE	2	2024-01-11	Una Foto Remix (feat. Emilia)	1kqH58eGh2ZTOHwqBIB2tM	https://iscdn.co/image/ab67616d00001e02d7e325...	2024-01-05	Mesita	2IKdK6PbitvCjXt1t2bPU6	2024-02-22	Argentina
2	3	20	3	2024-02-22	3	3	2958914	MOVED_UP	63	2024-02-08	Gata Only	6XjDF6nds4DE2BBbagZol6	https://iscdn.co/image/ab67616d00001e021d0777...	2024-02-02	FloyyMenor	7CvTknweLr9feJtRGrpDBy	2024-02-22	Argentina
3	4	5	3	2023-12-28	17	17	2776856	MOVED_UP	52	2023-11-02	QUE ME FALTE TODO	1TyBUtMxdmRfKhW3MmbxtQ	https://iscdn.co/image/ab67616d00001e02546d66...	2024-02-02	Luck Ra	4kcQWQDK0u9AftVSpdrAgk	2024-02-22	Argentina
4	5	4	3	2024-01-18	7	7	2667555	MOVED_DOWN	83	2024-01-11	Young Miko: Bzrp Music Sessions, Vol. 58	6VciS3dRMpZhjDxT8ArFi	https://iscdn.co/image/ab67616d00001e027b5840...	2024-01-10	Bizzarrap	716NhGYqD1jl2wl1Qkgq36	2024-02-22	Argentina
...	
476795	196	179	11	2020-11-19	17	17	26352	MOVED_DOWN	24	2020-11-12	The Right Journey (feat. Gducky)	2h1KRcol4TvqCl1Lf8RWlo	https://iscdn.co/image/ab67616d00001e02f97a37...	2020-11-08	RAP VIỆT	1cPpd989kgzlZqfpP4KwL	2021-03-04	Vietnam
476796	197	-1	151	2021-01-21	7	1	26334	RE_ENTRY	179	2021-01-07	Cưới Đì (feat. ChangC)	1CPruipLCJH06r0prq4AL	https://iscdn.co/image/ab67616d00001e0240f0fb...	2020-08-28	2T	5a9t9Ace3c3WLhwIYYsTGH	2021-03-04	Vietnam
476797	198	177	1	2020-01-23	49	2	26297	MOVED_DOWN	1	2020-01-23	Black Swan	4TaJICXAkv1WKYODXmq75b	https://iscdn.co/image/ab67616d00001e02505190...	2020-02-21	BTS	3Nrpe0tUJi4K4DXYWgMUX	2021-03-04	Vietnam
476798	199	156	93	2021-01-28	6	6	26261	MOVED_DOWN	93	2021-01-28	Chương 2 Của Tương Lai	5sDfLKhaLMbVffAmmnlkXc	https://iscdn.co/image/ab67616d00001e025ed379...	2021-01-08	WEAN	30eFAXoU2kTjPf2cq80B8	2021-03-04	Vietnam
476799	200	-1	21	2018-05-17	137	1	26143	RE_ENTRY	61	2018-03-15	Attention	5cF0dROIMOK5uNZtivgu50	https://iscdn.co/image/ab67616d00001e02897f73...	2018-05-11	Charlie Puth	6VuMaDnrHyPL1p4EHjYLi7	2021-03-04	Vietnam

476800 rows x 18 columns



Modélisation: DIM Region + ajout 'region_ID' à la Fact

```
dim_region = df_fact[['region']].drop_duplicates()  
dim_region['index'] = dim_region.reset_index().index+1  
print(dim_region)  
dim_region.to_csv('DIM_REGION.csv', index=False)
```

region	index
Argentina	1
Brazil	2
Egypt	3
France	4
Global	5
India	6
Indonesia	7
Japan	8
Mexico	9
Nigeria	10
Pakistan	11
Philippines	12
Taiwan	13
Turkey	14
USA	15
Vietnam	16

```
# Créer un dictionnaire pour mapper les régions aux index  
region_index_mapping = dict(zip(dim_region['region'], dim_region['index']))  
  
# Ajouter une colonne 'regionID' à df_fact  
df_fact['region_ID'] = df_fact['region'].map(region_index_mapping)
```



Modélisation: DIM Track

Réunification des _tracks.csv en une seule table DIM track:

```
merged_tracks = pd.concat(map(pd.read_csv, glob.glob('*_tracks.csv'))), ignore_index=True

dim_track = merged_tracks.drop_duplicates(subset=['id']).drop(columns=['type', 'uri', 'track_href'])
```

Jointure avec la Fact pour récupérer la 'track name' (et la 'release date'):

```
dim_track = dim_track.merge(df_fact.drop_duplicates(subset=['track_ID'])[['trackName', 'track_ID', 'releaseDate']], how='left',
                            left_on='id', right_on='track_ID')
```

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	analysis_url	duration_ms	time_signature	trackName	track_ID	releaseDate
0	0.482	0.792	2	-5.861	0	0.0859	0.1820	0.000000	0.7440	0.572	154.932	https://api.spotify.com/v1/audio-analysis/7JKp...	166537	4	HOLA PERDIDA	7JKptWv4YXmievid0HajID	2024-02-02
1	0.760	0.703	6	-3.667	1	0.0631	0.3130	0.000000	0.0784	0.739	96.033	https://api.spotify.com/v1/audio-analysis/1kqH...	243899	4	Una Foto Remix (feat. Emilia)	1kqH58eGh2ZTOHwqBIB2tM	2024-01-05
2	0.791	0.499	8	-8.472	0	0.0509	0.4460	0.000024	0.0899	0.669	99.986	https://api.spotify.com/v1/audio-analysis/6XjD...	222000	4	Gata Only	6XjDF6nds4DE2BBbagZoI6	2024-02-02
3	0.731	0.855	3	-5.115	0	0.0513	0.0935	0.000000	0.6880	0.966	148.042	https://api.spotify.com/v1/audio-analysis/1TyB...	165919	4	QUE ME FALTE TODO	1TyBUtMxdmRfXhW3MmbxtQ	2024-02-02
4	0.704	0.571	7	-8.212	0	0.1560	0.5880	0.000000	0.1170	0.443	96.016	https://api.spotify.com/v1/audio-analysis/6Vci...	190000	4	Young Miko: Bzrp Music Sessions, Vol. 58	6VciS3dRMpZhJiDxT8ArFi	2024-01-10
...	
22249	0.895	0.341	8	-16.053	1	0.1640	0.0241	0.000000	0.0973	0.352	92.011	https://api.spotify.com/v1/audio-analysis/7GQM...	139276	4	Okeokeoke	7GQM4prnPn4Oes1z2nyanm	2020-11-09
22250	0.494	0.491	7	-6.527	1	0.0268	0.5460	0.000000	0.1050	0.308	143.915	https://api.spotify.com/v1/audio-analysis/12Ha...	272944	4	Cũng Đành Thôi	12HaeiaS5UJgoxvnLgU5ro	NaN
22251	0.548	0.481	2	-10.049	1	0.0408	0.7580	0.000004	0.1030	0.244	139.911	https://api.spotify.com/v1/audio-analysis/4Xke...	224589	4	Her Summer	4XkeDeEoSxS7V1B4cjDshU	2019-07-18
22252	0.777	0.729	1	-4.577	0	0.0495	0.3750	0.000000	0.3470	0.239	112.001	https://api.spotify.com/v1/audio-analysis/2h1K...	192857	4	The Right Journey (feat. Gducky)	2h1KRcoI4TvqCI1Lf8RWio	2020-11-08
22253	0.716	0.799	3	-4.659	1	0.1710	0.5460	0.000000	0.1480	0.671	80.066	https://api.spotify.com/v1/audio-analysis/5sDf...	243000	4	Chương 2 Của Tương Lai	5sDfLKhaLMbVffAmmnlkXc	2021-01-08

22254 rows x 17 columns



Modélisation: DIM Artist et DIM Genre

```
merged_artists = pd.concat(map(pd.read_csv, glob.glob('*_artists.csv')), ignore_index=True)
```

```
merged_genres = pd.concat(map(pd.read_csv, glob.glob('*_genres.csv')), ignore_index=True)
```

	artistID	artist	popularity	followers
0	4kcQWQDK0u9AftVSpdrAgk	Luck Ra	76	918452
1	2IKdK6PbitvCiXt1t2bPU6	Mesita	72	338404
2	7CvTknweLr9feJtRGrpDBy	FloyyMenor	70	164596
3	716NhGYqD1jl2wl1Qkgq36	Bizarrap	84	14384925
4	6C2mFGFh8dz1vKp9I5Ce8P	TWENTY SIX	66	14218
...
5155	4R3mugkUqCALXgkwSptTbg	Đạt G	47	309375
5156	37M5pPGs6V1fchFJSgCguX	FINNEAS	67	1919950
5157	0ZYv8jRFp5YAC29bfjaP9X	Finn Askew	38	37145
5158	2CYz62cthPQ03xzOFpq683	Cheung	2	28331
5159	2LX2SKJhvZjyH4Z7XMptjd	Quang Đăng Trần	31	11701

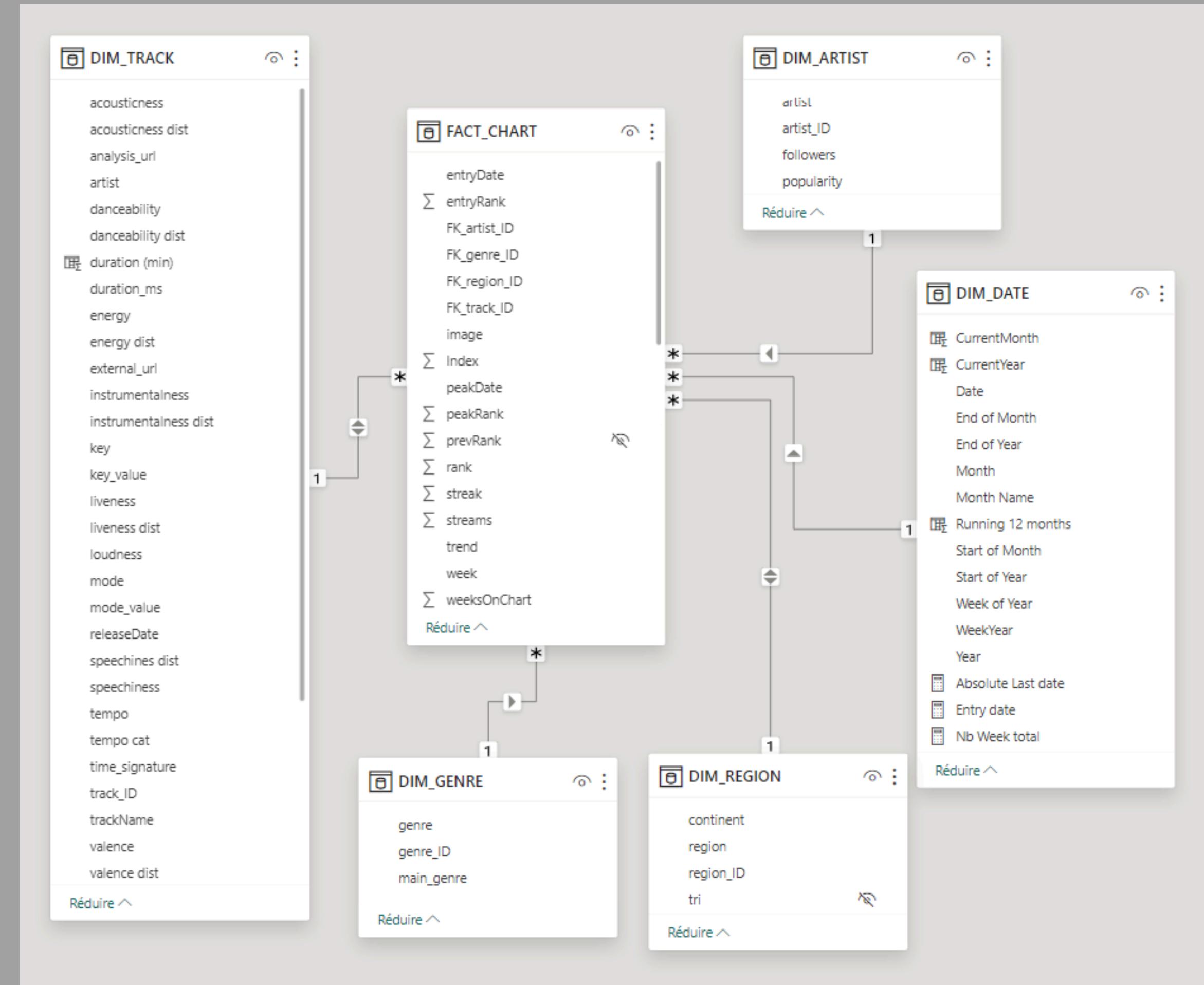
5160 rows × 4 columns

	genres	main_genre	genre_ID
0	latin hip hop	hip hop - rap	1
1	afro r&b	r-n-b - soul	2
2	chicha	world latin	3
3	pop	pop	4
4	cumbia pop	pop latin	5
...
800	vietnamese melodic rap	hip hop - rap	801
801	vietnamese bolero	world asian	802
802	t-pop	pop oriental	803
803	viet remix	pop oriental	804
804	rock viet	alternative rock	805
805

805 rows × 3 columns



Modélisation: Schéma dans Power BI





Overall Streams

Data Reporting Period :
from 04/03/2021 to 22/02/2024

Region

Tout

2021

2022

2023

2024



region	first date	Total Streams	Year 3Yago	Year 2Yago	Previous Year	% over LY
USA	2021-03	91.780M	26.989M	29.383M	35.408M	20,50 %
Mexico	2021-03	58.449M	12.505M	18.747M	27.197M	45,07 %
Brazil	2021-03	56.607M	16.704M	17.251M	22.652M	31,31 %
India	2021-03	37.360M	6.233M	12.435M	18.692M	50,31 %
Indonesia	2021-03	30.764M	5.379M	8.913M	16.472M	84,80 %
Turkey	2021-03	21.665M	4.757M	7.455M	9.453M	26,81 %
Philippines	2021-03	21.116M	4.580M	6.334M	10.202M	61,05 %
Argentina	2021-03	20.500M	5.411M	6.578M	8.510M	29,36 %
France	2021-03	19.099M	6.027M	6.205M	6.867M	10,66 %
Japan	2021-03	15.537M	4.607M	4.965M	5.966M	20,16 %
Vietnam	2021-03	3.906M	702M	1.027M	2.177M	112,02 %
Taiwan	2021-03	3.208M	883M	1.004M	1.321M	31,57 %
Nigeria	2022-02	1.938M	11M	471M	1.457M	209,48 %
Egypt	2021-03	1.592M	235M	394M	964M	144,49 %
Pakistan	2022-05	839M	0M	281M	558M	98,50 %
Total	2021-03	384.361M	95.024M	121.443M	167.894M	38,25 %

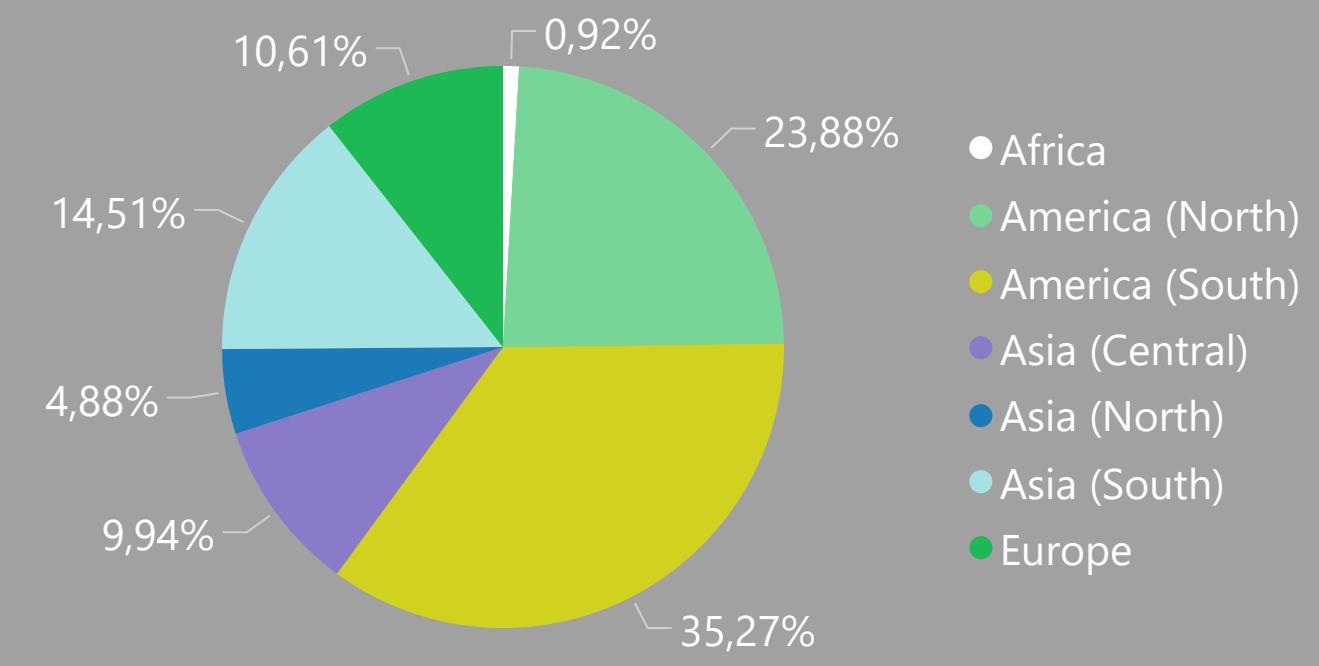
Global Streams

339Md

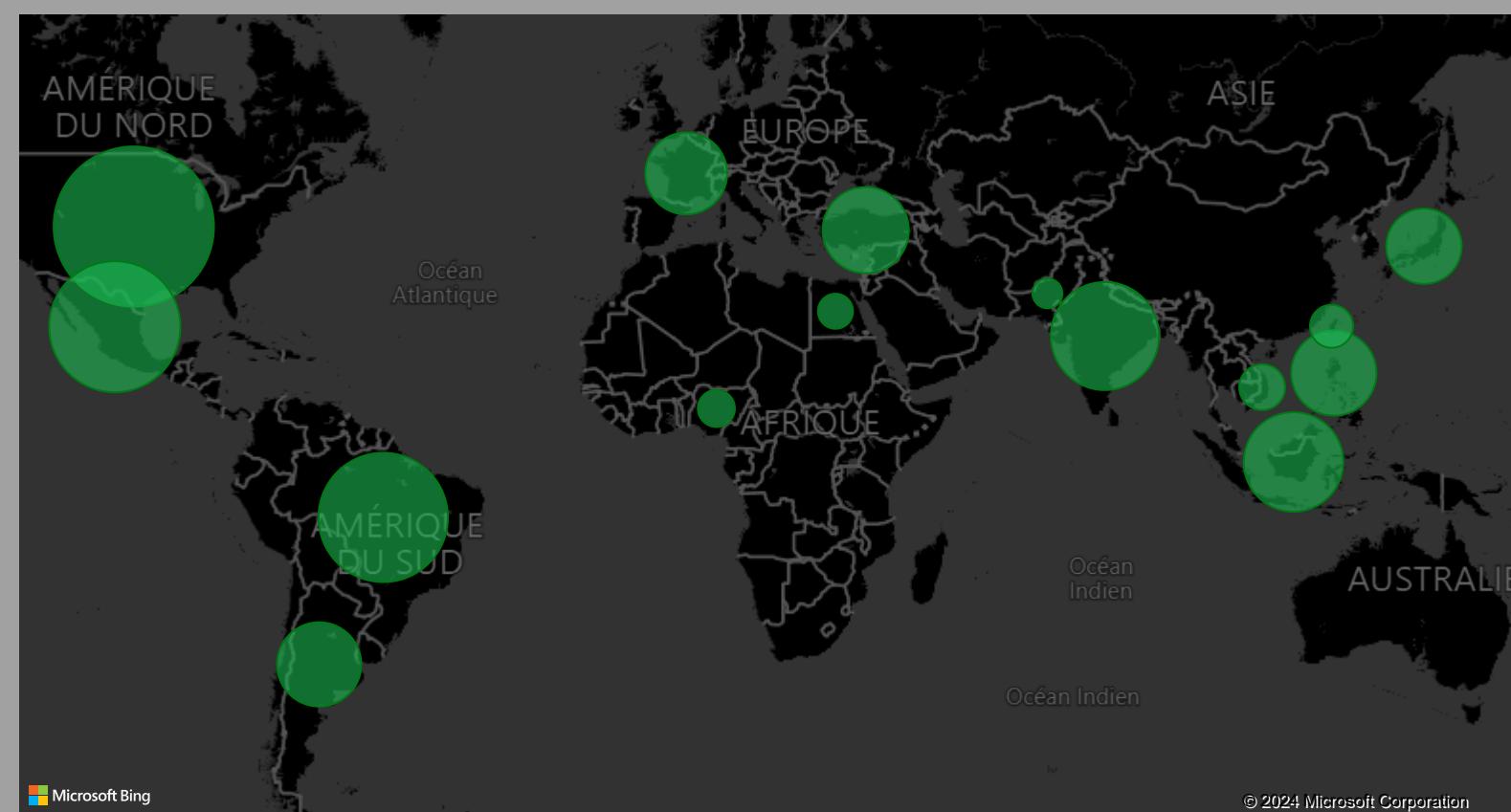
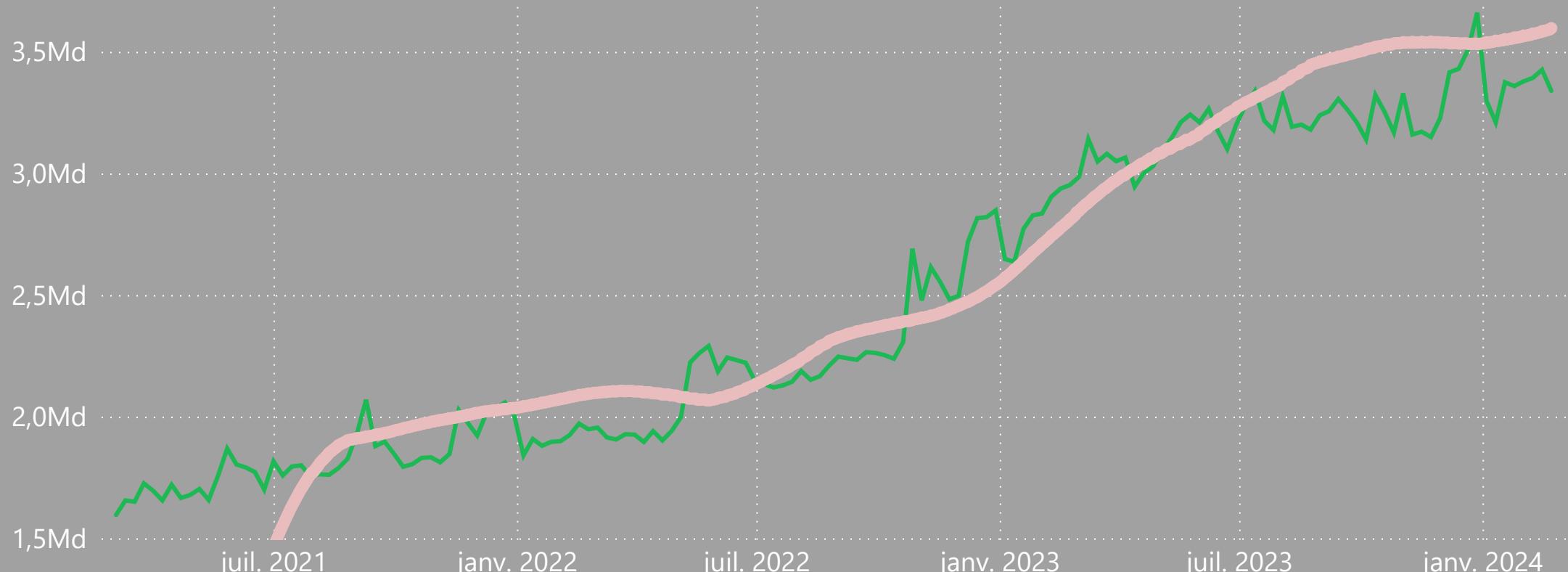
Total Region Streams

384Md

Avg Stream per year

128Md

● TotalStreams ● Rolling Avg 12 last weeks



Microsoft Bing

© 2024 Microsoft Corporation



Track Charts Performance

Region

Tout

23/12/2016

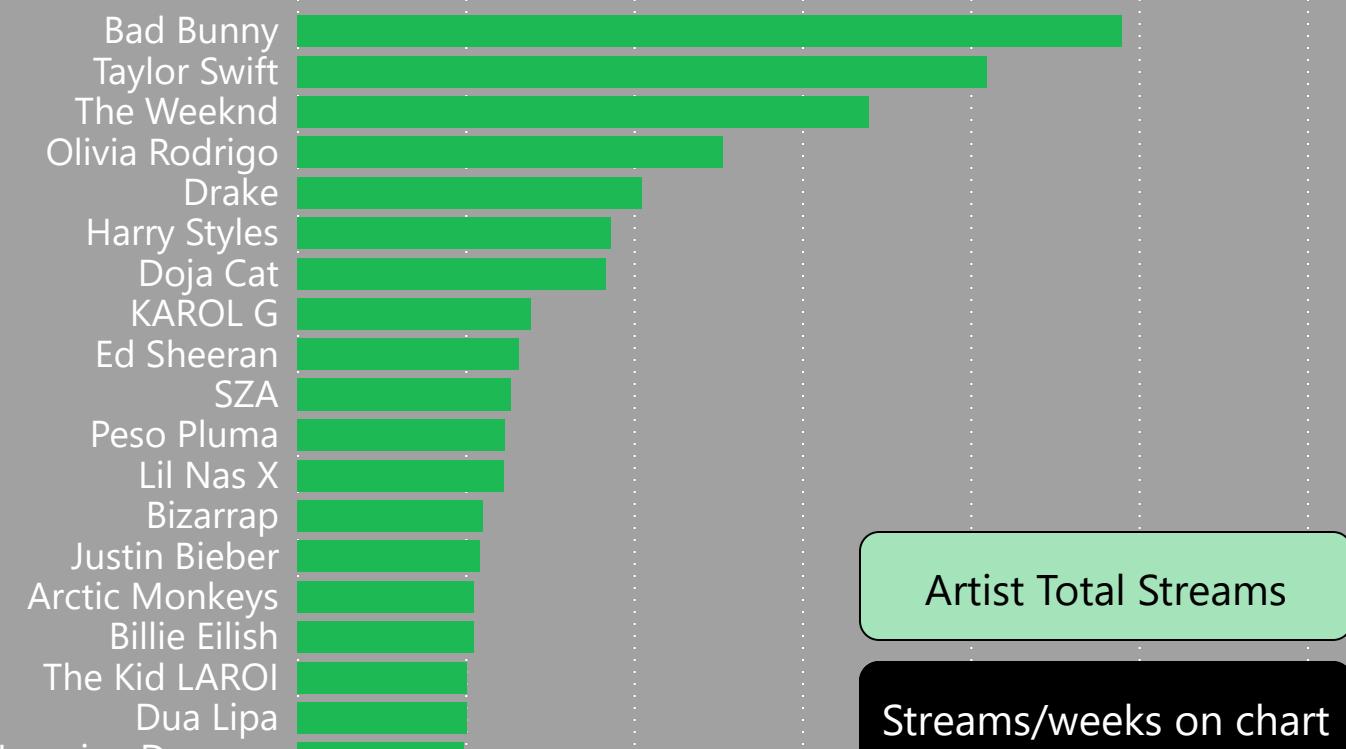
22/02/2024



Search and Select artist, track

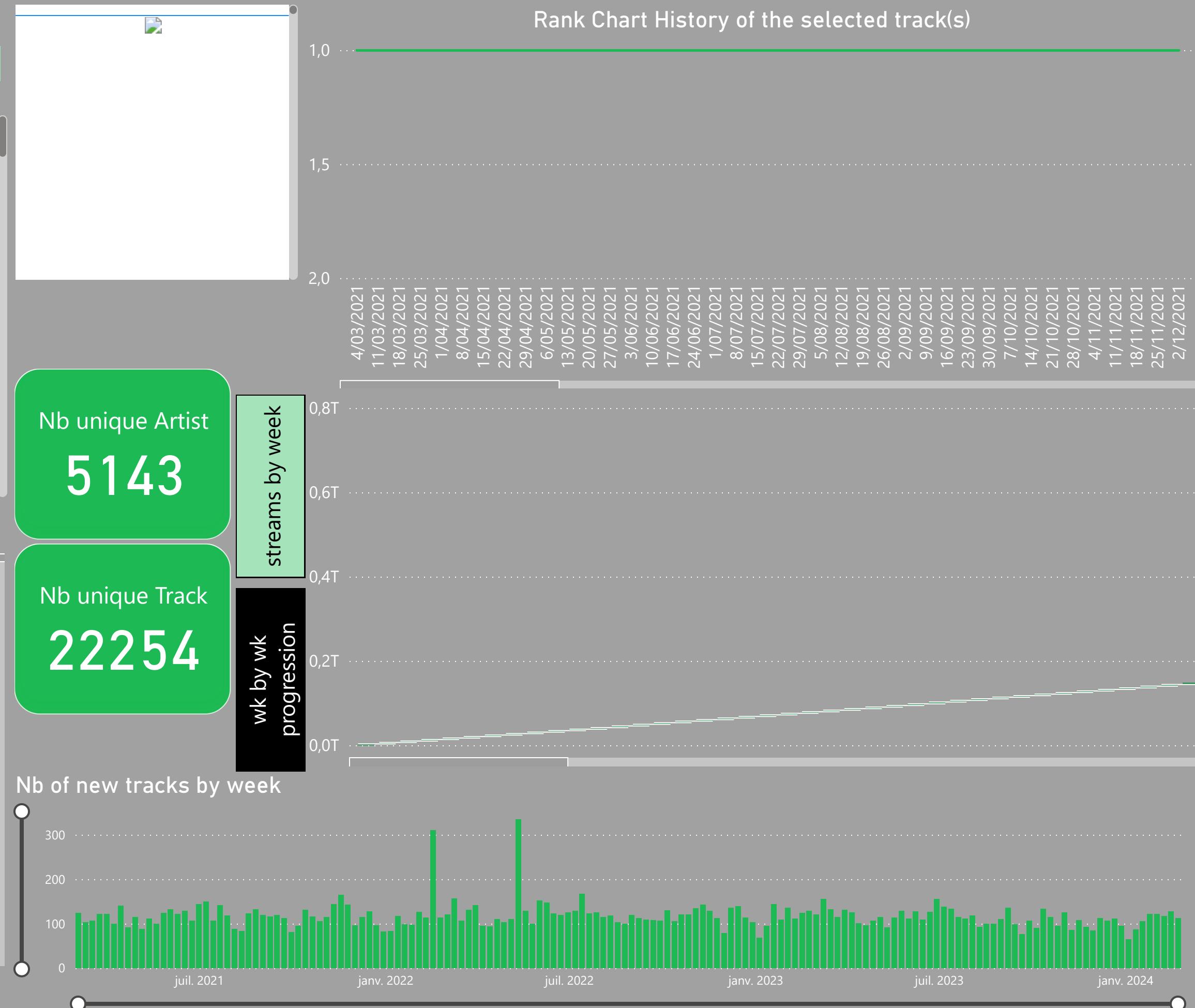
Tout

track	weeks	T10	pk	pk streams	streams	entryDate
As It Was	1184	157	1	78,5M	4.519M	7/04/2022
STAY (with Justin B...)	1328	262	1	70,5M	4.137M	15/07/2021
Heat Waves	1037	104	1	31,7M	3.655M	13/08/2020
Blinding Lights	1160	6	1	24,3M	2.812M	5/12/2019
Kill Bill	553	114	1	55,2M	2.567M	15/12/2022
Flowers	500	83	1	115,2M	2.560M	21/04/2022
Sweater Weather	590	13		16,9M	2.471M	29/03/2018
Me Porto Bonito	332	108	1	39,8M	2.470M	12/05/2022
INDUSTRY BABY (f...)	802	118	1	53,1M	2.434M	29/07/2021
good 4 u	645	69	1	84,1M	2.356M	20/05/2021
Die For You	826	13	4	23,0M	2.339M	29/12/2016



Artist Total Streams

Streams/weeks on chart





Audio Distribution

Region

Tout

23/12/2016

22/02/2024



Search and Select artist, track

Tout

Median duration (min)

3,25

by track

by artist

Top 10 genre by streams

pop

dance pop

canadi...

k-pop

reggaeton

corrido

hip hop

col...

agronejo

afrobeats

agronejo

arab pop

arrocha

corrido

desi pop

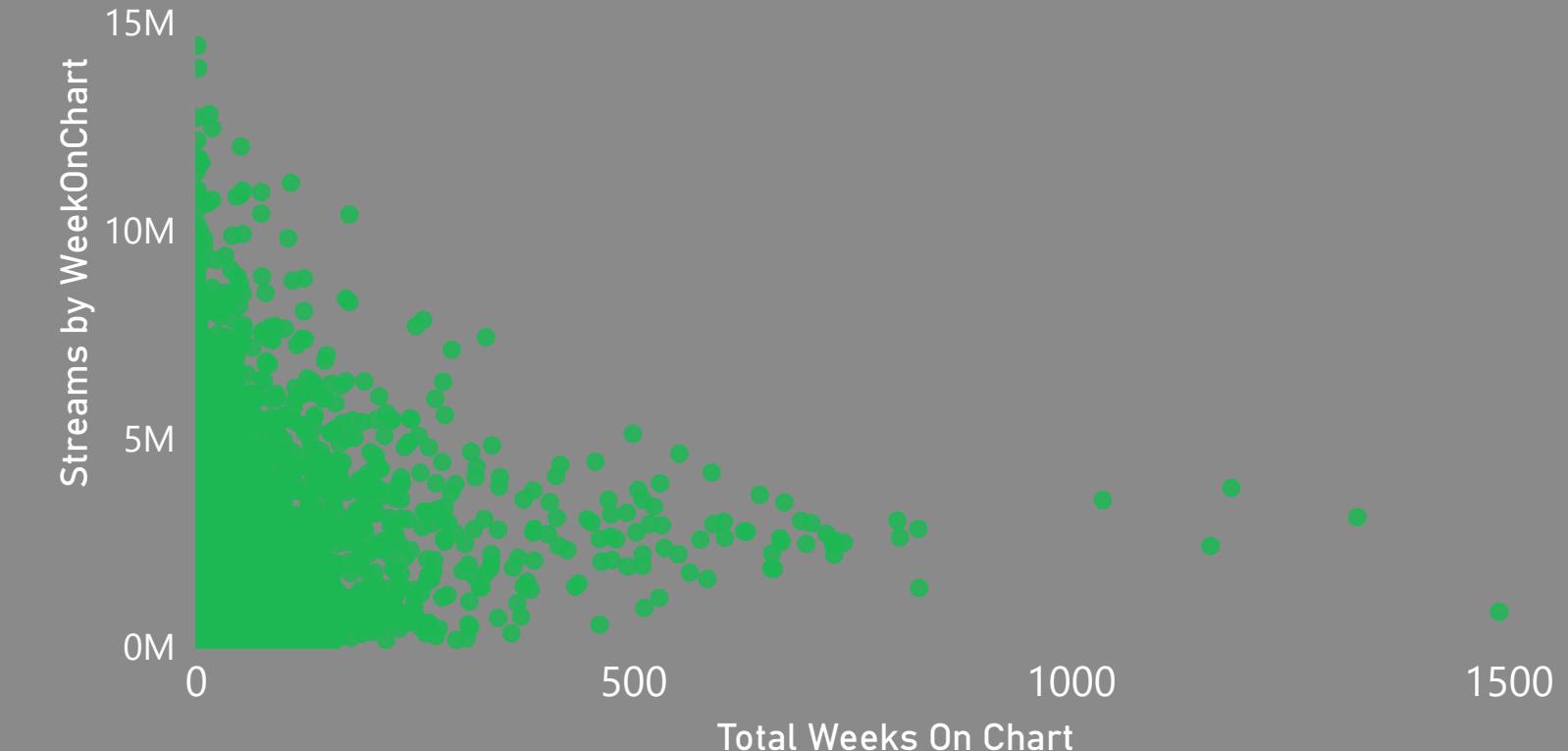
egyptian pop

filmi

french hip hop

funk carioca

indonesian pop



E minor

C#/Db major

C major

F major

D major

F minor

G#/#A major

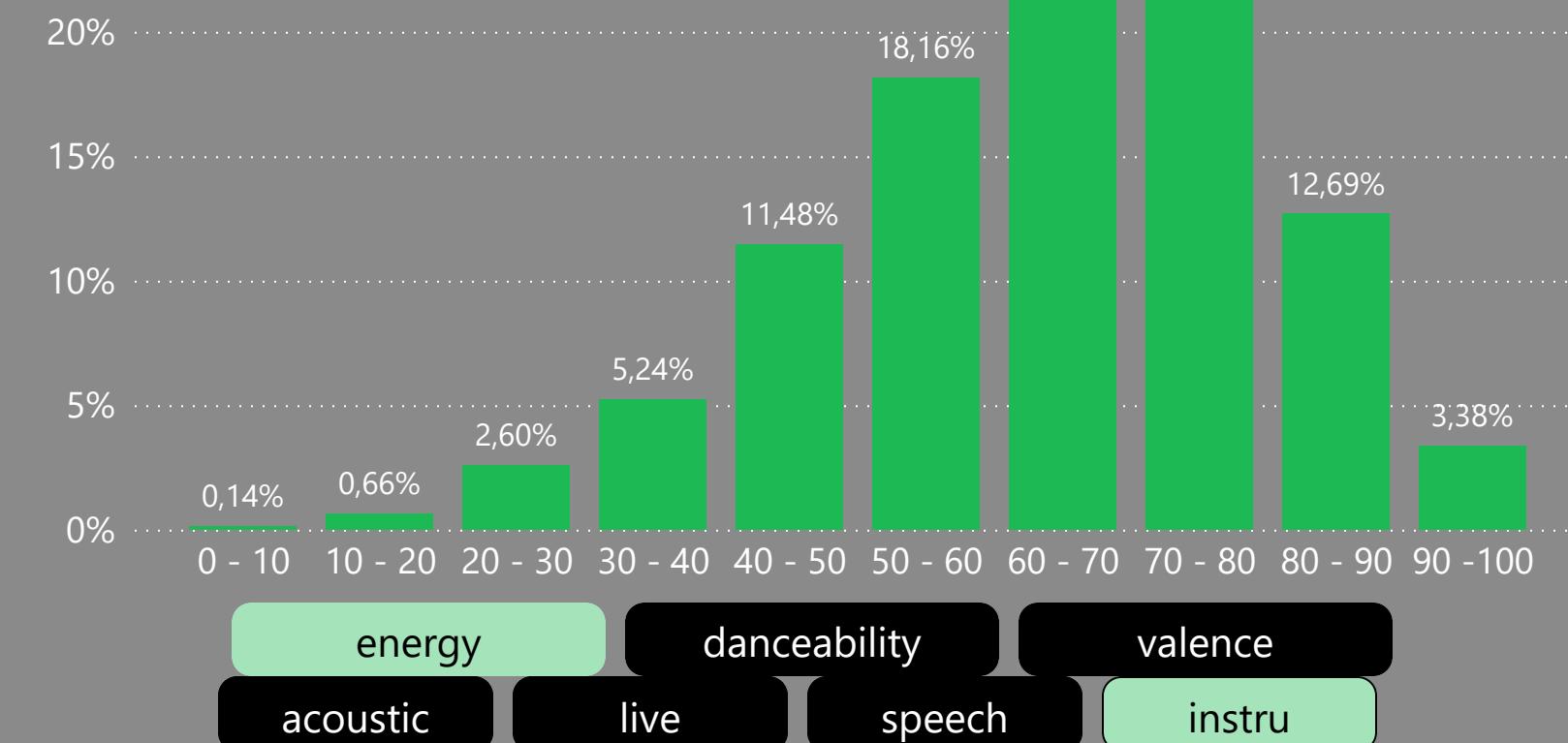
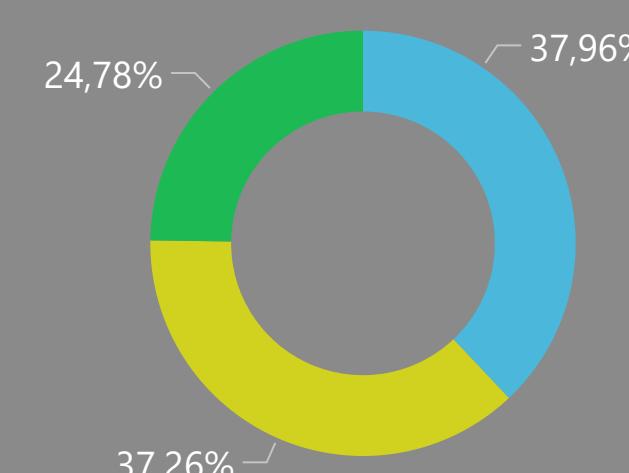
F#/Gb minor

G major

C#/Db minor

B minor

A major

● high tempo
 ● low tempo
 ● mid tempo
 

Pour aller plus loin

- . Récupération des datas jusqu'à 10 ans en arrière
- . Mise en place d'un flux ETL (SSIS) pour actualiser les données chaque semaine dans Power BI

Limites de l'étude

- . Données de tracks 'mainstream' pas forcément représentatives de la musique dans son ensemble
- . Pas d'accès aux données utilisateurs Spotify (nombre par pays, profil-type, ...)

Merci pour votre attention !