



# Music & Art Festivals

IST 736 Text Mining Project, Winter 2019

Adil Gokturk, Scott Snow



Emeric Fohlen/NurPhoto via Getty Images

The music festival industry is currently (2018) worth \$ 3bn globally and continues to grow, according to research by *Festicket*.

The “Burning Man” is one of the best known festivals, globally.



Emeric Fohlen/NurPhoto via Getty Images

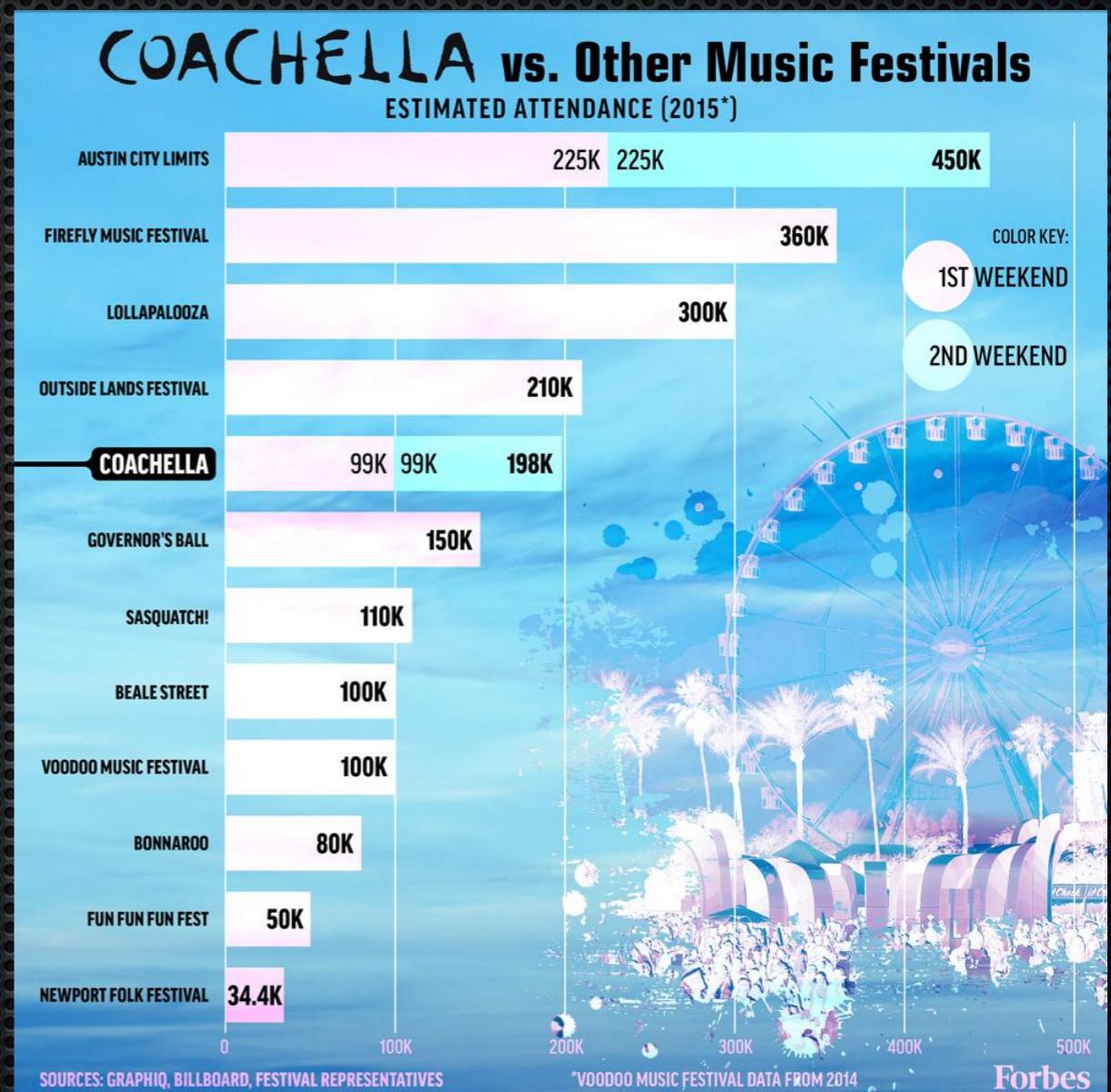
# Glastonbury

Set in 900 acres of countryside in England's south-west

- ◆ 1970 attendance 1500 people and ticket cost app \$1.5
- ◆ 2017 attendance 200.000 people—each day and ticket cost was about \$296

# Music & Art Festival Numbers

- According to Forbes, Coachella is the fifth-most popular music festival by attendance in 2015.
- In 2015, The festival sold out both of its weekends for a cumulative attendance of 198,000 people.
- About 450,000 people went to Austin City Limits in 2015.





## Research Question:

- In what ways can we utilize data-driven text mining algorithms of Twitter data—tweets to gain valuable insights about music festivals?

# Research Goals:

- A Twitter text mining project
- We download tweets as data set to investigate the phenomena of the Music Festivals
- Categories:
  - Sentimental analysis
  - Musical Trends and Genres
  - Specific language structures
    - Festival slang!
  - Sponsors
  - Consumptions
    - Soft/hard liquor
    - Costume/fashion trends
    - Others (could be personal and restrictive!)



# Python Library and Data: Tweets

The screenshot shows the official Tweepy Python library documentation. At the top, it says "tweepy v3.5.0". Below that is a search bar labeled "Search docs". A sidebar on the left contains a "Getting started" section with links to "Introduction" and "Hello Tweepy", and an "API" section.

## Hello Tweepy

```
import tweepy

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)

api = tweepy.API(auth)

public_tweets = api.home_timeline()
for tweet in public_tweets:
    print tweet.text
```

- Need to set up a twitter developer account
- Get an access to use API: application programming interface
- Python libraries
  - nltk, tweeps, textblob

- Hash-tag: **#musicfestival**
  - n=500
  - Cleaning and preparing Data
  - Applying Text Mining Applications
  - Finding applicable working model

## ■ Cleaning and prep of the data: tweets

### #musicfestival

RT @cassidyaspinall: Can't wait to walk into festivals with the music banging, sun shining and my best girls with me! Summer hurry up alrea...  
The Premier's war on music has to stop. We need to protect our arts and culture in NSW, not put more barriers in th... <https://t.co/bENpmZa5pF>  
RT @ElPasoTXGov: El Paso is not only amongst the safest cities in the nation, we are also thriving. From new restaurants in downtown, cultu...  
Can't wait to walk into festivals with the music banging, sun shining and my best girls with me! Summer hurry up already 😊  
I think I need to start going to music festivals. People seem to find real love there <https://t.co/01PCAJj2tJ>  
To help you choose where to get your groove on during your #holiday, we have compiled a list for you of the Top 5 M... <https://t.co/0yP6vbQfnw>  
RT @UVMLondon: Sun, Sea, Cider, Great Live Music– sounds like heaven – we're there on 26th May. #festivals #campingbeciderseaside #devon #t...

# Unstructured Data—a sample tweet

```
{ "created_at": "Mon Jan 15 01:14:44 +0000 2018", "id": 952710610915090432, "id_str": "952710610915090432", "text": "RT @jamesMetville: Donald Trump\u2019s State of the Union speech is on January 30th. \n\nHe\u2019s obsessed by TV ratings.\n\nAmerica - don\u2019t watch his\u2026", "source": "\u003ca href=\"http://\\twitter.com\\download\\iphone\" rel=\"nofollow\"\u003eTwitter for iPhone\u003c\\a\u003e", "truncated": false, "in_reply_to_status_id": null, "in_reply_to_status_id_str": null, "in_reply_to_user_id": null, "in_reply_to_user_id_str": null, "in_reply_to_screen_name": null, "user": { "id": 743212353546817540, "id_str": "743212353546817540", "name": "Mariah", "screen_name": "mariahtrash003", "location": null, "url": null, "description": null, "translator_type": "none", "protected": false, "verified": false, "followers_count": 75, "friends_count": 81, "listed_count": 0, "favourites_count": 4195, "statuses_count": 1206, "created_at": "Wed Jun 15 22:43:28 +0000 2016", "utc_offset": null, "time_zone": null, "geo_enabled": false, "lang": "en", "contributors_enabled": false, "is_translator": false, "profile_background_color": "F5F8FA", "profile_background_image_url": "", "profile_background_image_url_https": "", "profile_background_tile": false, "profile_link_color": "1DA1F2", "profile_sidebar_border_color": "C0DDEED", "profile_sidebar_fill_color": "DDEEF6", "profile_text_color": "333333", "profile_use_background_image": true, "profile_image_url": "http://\\pbs.twimg.com\\profile_images\\943920250780901377\\e8_hLGL1_normal.jpg", "profile_image_url_https": "https://\\pbs.twimg.com\\profile_images\\943920250780901377\\e8_hLGL1_normal.jpg" }, "coordinates": null, "place": null, "quoted_status_id": null, "quoted_status_id_str": null, "geo": null }
```

## » Tweet Data Dictionary

- » **created\_at**
- » **id**
- » **id\_str**
- » **text**
- » **source**
- » **truncated**
- » **in\_reply\_to\_status\_id**
- » **in\_reply\_to\_status\_id\_str**
- » **in\_reply\_to\_user\_id**
- » **in\_reply\_to\_user\_id\_str**
- » **in\_reply\_to\_screen\_name**
- » **user**
- » **coordinates**
- » **place**
- » **quoted\_status\_id**
- » **quoted\_status\_id\_str**
- » **geo**

# ♦ Bag of Words

♦#musicfestivals

```
In [14]: len(BagOfWords)  
Out[14]: 4149
```

```
'June', 'magia', 'existe', 'Solo', 'tienes', 'que', 'abrir', 'los', 'ojos',  
'annual', 'Calle', 'Ocho', 'event', 'Miami', 'Blu', 'New', 'Single', 'CAR',  
'AWFUL', 'shredded', 'Fringe', 'Vest', 'Festival', 'vest', 'gypsy', 'boho',  
'Tassel', 'vest', 'WHITE', 'Earthy', 'retro', 'inspired', 'vest', 'made',  
'elegant', 'fearsome', 'from', 'grand', 'Just', 'click', 'here', 'Show',  
'organizer', 'and', 'post', 'your', 'show', 'with', 'all', 'the', 'details',  
'and', 'images', 'tienes', 'que', 'disfrutar', 'que', 'musica', 'guie', 'sure  
'and', 'get', 'your', 'tickets', 'for', 'They', 'are', 'running', 'out', 'fas  
'Amo', 'Amas', 'June', "we'll", 'celebrate', 'years', 'family', 'pride',  
'tradition', 'and', 'puro', 'mariachi', 'Will', 'you', 'there', 'Best', 'seat  
'Colorful', 'Austin', 'THE', 'FOO', 'FIGHTERS', 'Contiki', 'Sounds', 'returns  
'with', 'trip', 'the', "UK's", 'Reading', 'Festival', 'SCARF', 'beach', 'cove  
'gypsy', 'sarong', 'scarf', 'festival', 'clothing', 'NAVY', 'blue', 'boho',  
'chic', 'hippie', 'beach', 'skirt', 'the', 'building', 'performing', 'live',  
'Official', 'Hats', 'MOHAWK', 'BENDS', 'killed', 'travels', 'April',  
'documentary', 'about', 'toda', 'energía', 'música', 'year', 'benefiting',  
'support', 'vital', 'mental', 'health', 'services', 'for', 'youth', 'and',  
'families', 'Robot', 'work', 'progress', 'Wave', "I'm", 'Surfing', 'Right',  
'Now', 'Dude', 'TUNE']
```

# ♦ Bag of Hash-Tags

♦#musicfestivals

```
In [16]: len(BagOfHashes)  
Out[16]: 7950
```

```
'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album', 'newsong', 'live',  
'np', 'music', 'productor', 'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic',  
'NewMusic', 'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'productor',  
'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019',  
'album', 'newsong', 'live', 'np', 'music', 'productor', 'TOPTUNE',  
'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album',  
'newsong', 'live', 'np', 'music', 'productor', 'TOPTUNE', 'musicfestival',  
'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album', 'newsong', 'live',  
'np', 'music', 'productor', 'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic',  
'NewMusic', 'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'productor',  
'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019',  
'album', 'newsong', 'live', 'np', 'music', 'productor', 'TOPTUNE',  
'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album',  
'newsong', 'live', 'np', 'music', 'productor', 'TOPTUNE', 'musicfestival',  
'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album', 'newsong', 'live',  
'np', 'music', 'productor', 'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic',  
'NewMusic', 'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'productor',  
'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019',  
'album', 'newsong', 'live', 'np', 'music', 'show', 'edclasvegas', 'lasvegas',  
'electro', 'musicismylife', 'rappers', 'goodmusic', 'pop', 'studio', 'livemusic',
```

# ♦ Bag of Links

## ♦#musicfestivals

```
In [18]: len(BagOfLinks)  
Out[18]: 823
```

```
9G2vhHH#DJ', 'https://t.co/ohF108A34r#DJ', 'https://t.co/QgWLiRGajv#DJ',  
tps://t.co/WDf6fEM7lb#DJ', 'https://t.co/Q8lrUnuwN6#DJ', 'https://t.co/  
ERoq5sC#DJ', 'https://t.co/QPZM1h8AHZ#DJ', 'https://t.co/sMCY27n0fX#DJ',  
tps://t.co/xJs8QJtas7#DJ', 'https://t.co/RyelbKqFQt#DJ', 'https://t.co/  
8j4yax0#DJ', 'https://t.co/qDAvEMCI7oRegister', 'https://t.co/9sGRPa1ttL',  
tps://t.co/nGgSo6zBRC😍⚡️Solo', 'https://t.co/jTAfiGi756🔥🔥喇叭#instagood'  
tps://t.co/Uj0sDwccE1Make', 'https://t.co/bGZX591H3💕 kup Yo', 'https://t.  
GMrbDoXOn', 'https://t.co/Q6A7otvmzehttps://t.co/grgFkr3hhW', 'https://t.cc  
AD3H60ART', 'https://t.co/JU0FWjXswkSEE', 'https://t.co/SGPusa26bP...',  
tps://t.co/wLKPTd6DrFCROCHET', 'https://t.co/TY9JABL4dNRT', 'https://t.co/  
z27UJER.#AfroRootsFest', 'https://t.co/prTeA6h7VP😍 liberar', 'https://t.co/  
Hok1nM3#bohochic', 'https://t.co/5ouy9TNCg0This', 'https://t.co/  
1IfmG5NEnvious', 'https://t.co/T4bbFTgx2XCatch', 'https://t.co/37bMEtqPx2
```

## • BigBag = BagOfWords + BagOfHashes

• **#musicfestivals**

```
In [27]: len(BigBag)
Out[27]: 12099
```

```
'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'show',
'edclasvegas', 'lasvegas', 'electro', 'musicismylife', 'rappers',
'goodmusic', 'pop', 'studio', 'livemusic', 'band', 'like', 'rock',
'musicfestival', 'coachella2019', 'coachella', 'coachellastyle',
'ravecouples', 'electro', 'DJ', 'productor', 'musicfestival',
'musicians', 'album', 'newsong', 'music', 'EDM', 'production',
'SXSW2019', 'Texas', 'austintexas', 'city', 'SouthbySouthWest',
'musicfestival', 'filmfestival', 'musicfestival', 'SXSW2019',
'blaqpyrates', 'Aust', 'artist', 'musician', 'artistandrepertoire',
'magmilesshows', '6thstreetaustin', 'independent', 'KeyWest',
'burningman', 'blackrockcity', 'ravebooty', 'electro', 'bohemian',
'festival', 'festivalfashion', 'music', 'musicfestival', 'sunday',
'sundayfunday', 'bohostyle', 'craft', 'music', 'instafun',
'hiphopblog', 'hiphopnation', 'hiphoplife', 'bhfyp', 'hiphophead',
'instagood', 'musicismylife', 'rappers', 'goodmusic', 'pop']
```

# Ignored Words

# #musicfestivals

```
In [28]: len(IgnoreThese)  
Out[28]: 264
```

ティバル", "ラジオ", "フェスティバル", "ラジオ", "フェスティバル", "ラジオ", "フェ  
", "フェスティバル", "ラジオ", "صب\_غرة", "自動運転中", "おし  
", "제1회빛이나예술제", "제1회빛이나예술제", "حفل", "الحظات", "\_\_\_",  
"Mae", "ختام", "ختام", "حفل", "الحظات", "\_\_\_", "لحظات", "قضاء", "عيشها", "Muslim", "Holi  
ティバル", "ラジオ", "フェスティバル", "قضاء", "بإمكانكم", "Muslim", "Holi  
"amsNagpur", "Haldirams", "제1회빛이나예술제", "जगह", "حفل", "السينمائي", "Holi  
ي", "السينمائي", "ختام", "حفل", "السينمائي", "Holi  
ちゃんと間違われる", "川崎東田商店街その", "بإمكانكم", "زال", "الورقية", "الطائرات", "الثالث", "الدولي", "مهرجان", "يوم", "آخر", "ممتعة",  
丑商店街その", "リカちゃんと間違われる", "川崎東田商店街その", "제1회빛이나예술제", "着  
みでお祭り", "リカちゃんと間違われる", "川崎東田商店街その", "제1회빛이나예술제", "着  
ちゃんと間違われる", "川崎東田商店街その", "の開催が決定", "クラブミュージック", "現  
つりをお楽しみください", "برای", "بیشتری", "سود", "زمستان", "فصل", "گردشگری", "دلت", "لولت",  
"chanchan", "ようこそ女子大エリア", "一年一度的節日", "今天剛好早下班", "就難得出門  
霍打給我", "跟她一起看煙火", "覺得很開心滿足", "偕樂園", "제1회빛이나예술제", "偕樂園  
", "Nagpur", "BOTH", "menunjukkan", "HOLIKA", "BLACK", "SLAB", "PDF", "Ul

# ✿ Words Dictionary

✿#musicfestivals

```
In [43]: len(WordDict)
Out[43]: 1871
```

```
stival2019': 6, 'WeLiveTheBeachLi': 3, 'BeachLifeFestival2': 3, 'Bea': 3,
: 3, 'JimLindberg': 3, 'BRATSONTHEBEAT': 3, 'RAMONESFORKIDS': 3, 'stall': 3
y': 3, 'oughtringtonprimaryschool': 3, 'lymm': 3, 'helpneeded': 3, 'oughtyf
': 3, 'RootsPicnic': 6, 'SingerSongwriter': 6, 'JUN0': 6, 'CFMA': 6, 'BC': 6
3, 'BeachLifeFestLA': 3, 'LeylandFe': 6, 'MarlayPark': 6, 'psymusic': 6,
': 6, 'musi': 9, 'breakoutwest': 3, 'canadianmusic': 3, 'coachellaoutfit': 3
shion': 3, 'Franklin': 3, 'talent': 6, 'feisliverpool': 9, 'Etsy': 3, 'etsy'
: 6, 'festivalstyle': 3, 'model': 3, 'studio': 6, 'light': 3, 'bed': 3, 'wa
musicfestivalRT': 6, 'fdnwhour': 6, 'manchestermusicscene': 3, 'musicphotog
3, 'girlzrawk': 3, 'girlzrawkmanchester': 3, 'PushitChallenge': 3, 'Beatblo
Tone': 9, '80s': 9, 'MUSICFestival': 9, 'musicman': 3, 'musicianlife': 3,
': 3, 'HereWeGoLo': 3, 'cuzeventsrock': 3, 'rootspicnic': 3, 'philly': 3, 't
sBTQ': 3, 'qotd': 3, 'quote': 3, 'wild': 3, 'flowerchild': 3, 'LeylandFesti
ure': 3, 'dayfornight': 3, 'shades': 3, 'eprom': 3, 'alixperez': 3, 'PalNor
: 6, 'FreddieMercury': 3, 'goodmusic': 9, 'familyfriendly': 3, 'pr': 9, 'mu
list': 6, 'spotify': 6, 'Spotifyplays': 6, 'SpotifyMusicPromotion': 6,
opid': 3, 'reggae': 3, 'iration': 3, 'ska': 3, 'rebelution': 3, 'pepper': 3
```

# ◆ Word Cloud - top 50 words

## ◆ *#musicfestivals*



# Results: Sentiment Analysis

- Python textblob library

Positive tweets percentage: 41.935483870967744 %  
Negative tweets percentage: 3.225806451612903 %



# Frequency Distribution

- Used raw tweets.
- tokenized and lowered the tweeds
- 34886 words**
- Created Frequency distribution

```
('musicfestival', 1965)
('music', 1848)
('musicians', 1599)
('album', 1545)
('newsong', 1545)
('productor', 1536)
('live', 1464)
('edm', 1449)
('toptune', 1413)
('np', 1413)
('ff', 1095)
('deephouse', 687)
('onairnow', 654)
('lovemusic', 648)
('newmusic', 615)
('dj', 522)
('festival', 339)
('top', 219)
('production', 153)
('tickets', 120)
('crochet', 120)
('djaxer', 96)
('newmusicalert', 93)
('electronicmusic', 90)
'summer', 75)
'day', 69)
'show', 60)
'get', 60)
'april', 60)
'clothing', 57)
'halter', 57)
'check', 57)
'stage', 57)
'airplay', 57)
'see', 54)
'art', 54)
'electro', 51)
'ultra2019', 51)
'june', 48)
'wave', 45)
'boho', 45)
'love', 45)
'days', 45)
'one', 45)
'musician', 45)
'party', 42)
'event', 42)
'fringe', 42)
'march', 39)
'may', 39)
'weekend', 39)
```

# Bigram

- Used raw tweets.
- tokenized and lowered the tweeds
- 3329 bigrams**
- List of top 50 scored

```
(('album', 'newsong'), 0.04428710657570372)
((('musicfestival', 'musicians'), 0.04428710657570372)
((('live', 'np'), 0.04050335378088631)
((('newsong', 'live'), 0.04050335378088631)
((('productor', 'toptune'), 0.04050335378088631)
((('np', 'music'), 0.04041735939918592)
((('music', 'edm'), 0.039901393108983545)
((('toptune', 'musicfestival'), 0.036203634695866536)
((('edm', 'ff'), 0.031387949320644384)
((('musicians', 'album'), 0.025712320128418276)
((('deephouse', 'productor'), 0.019606719027690192)
((('musicians', 'lovemusic'), 0.018574786447285444)
((('ff', 'onairnow'), 0.018488792065585048)
((('lovemusic', 'newmusic'), 0.017628848248581092)
((('newmusic', 'album'), 0.015221005560970016)
((('dj', 'deephouse'), 0.00894341569684114)
((('ff', 'productor'), 0.008599438170039558)
((('onairnow', 'deephouse'), 0.006965544917732041)
((('dj', 'productor'), 0.005417646047124921)
((('edm', 'production'), 0.003783752794817405)
((('newsong', 'music'), 0.003783752794817405)
((('productor', 'musicfestival'), 0.0035257696497162185)
((('onairnow', 'dj'), 0.0029238089778134495)
((('production', 'djsxer'), 0.0027518202144126585)
((('music', 'festival'), 0.002665825832712263)
((('newmusicalert', 'musicfestival'), 0.002665825832712263)
((('toptune', 'newmusicalert'), 0.002665825832712263)
((('djsxer', 'electronicmusic'), 0.0024938370693114715)
((('edm', 'productor'), 0.0024938370693114715)
((('onairnow', 'productor'), 0.002235853924210285)
((('music', 'productor'), 0.0017198876340079115)
((('ff', 'dj'), 0.001633893252307516)
((('toptune', 'airplay'), 0.001633893252307516)
((('halter', 'top'), 0.0015478988706071203)
((('top', 'crochet'), 0.0014619044889067248)
((('music', 'deephouse'), 0.0013759101072063293)
((('newmusic', 'ultra2019'), 0.0013759101072063293)
((('ultra2019', 'album'), 0.0013759101072063293)
((('airplay', 'musicfestival'), 0.0011179269621051425)
((('coub', 'album'), 0.001031932580404747)
((('festival', 'top'), 0.001031932580404747)
((('newmusic', 'coub'), 0.001031932580404747)
((('top', 'festival'), 0.001031932580404747)
((('check', 'out'), 0.0009459381987043513)
((('festival', 'clothing'), 0.0009459381987043513)
((('lovemusic', 'album'), 0.0009459381987043513)
((('music', 'wave'), 0.0009459381987043513)
((('musicfestival', 'art'), 0.0009459381987043513)
((('onairnow', 'prod'), 0.0009459381987043513)
((('wave', 'musicfestival'), 0.0009459381987043513)
```

# Bigram - PMI

## Pointwise Mutual Information

- tf-idf vectorizer
- A measure of how often two events x and y occur, compared with what we would expect if they were independent
- Frequency filter = 3
- 3329 bigrams**
- List of top 50 scored

```
(('able', 'look'), 13.505398067619186)
((('ableton', 'logicpro'), 13.505398067619186)
((('abril', 'formali'), 13.505398067619186)
((('abrir', 'los'), 13.505398067619186)
((('acid', 'psytribe'), 13.505398067619186)
((('addition', 'comes'), 13.505398067619186)
((('adrian', 'kosky'), 13.505398067619186)
((('alto', 'mare'), 13.505398067619186)
((('american', 'icon'), 13.505398067619186)
((('amo', 'amas'), 13.505398067619186)
((('amped', 'spend'), 13.505398067619186)
((('américacantat9', 'panamá'), 13.505398067619186)
((('andorra', 'five'), 13.505398067619186)
((('andy', 'track-a-day'), 13.505398067619186)
((('anniversary', 'design'), 13.505398067619186)
((('arizona', 'final'), 13.505398067619186)
((('arrive', 'trust'), 13.505398067619186)
((('avoid', 'disappointment'), 13.505398067619186)
((('awesome', 'session'), 13.505398067619186)
((('awful', 'shredded'), 13.505398067619186)
((('bald', 'eagle'), 13.505398067619186)
((('bamboo', 'head'), 13.505398067619186)
((('bastard', 'drops'), 13.505398067619186)
((('bathing', 'swim'), 13.505398067619186)
((('bea', 'nowplaying'), 13.505398067619186)
((('beard', 'entered'), 13.505398067619186)
((('bed', 'wall'), 13.505398067619186)
((('bends', 'killed'), 13.505398067619186)
((('better', 'heading'), 13.505398067619186)
((('beyondwonderland', 'insomniac'), 13.505398067619186)
((('bfhyp', 'hiphophead'), 13.505398067619186)
((('bigtakeoverband', 'thebigtakeover'), 13.505398067619186)
((('blackrockcity', 'ravebooty'), 13.505398067619186)
((('blaqpyrates', 'aust'), 13.505398067619186)
((('blossom', 'jamboree'), 13.505398067619186)
((('booking', 'fort'), 13.505398067619186)
((('bourbonandbeyond', 'bourbon'), 13.505398067619186)
((('brampton', 'mississauga'), 13.505398067619186)
((('bratsonthethebeat', 'ramonesforkids'), 13.505398067619186)
((('breakout', 'west'), 13.505398067619186)
((('breakoutwest', 'canadianmusic'), 13.505398067619186)
((('brimmed', 'beanie'), 13.505398067619186)
((('build-your-own', 'subscriptions'), 13.505398067619186)
((('burning', 'man'), 13.505398067619186)
((('busy', 'tones'), 13.505398067619186)
((('caledon', 'orangeville'), 13.505398067619186)
((('calle', 'ocho'), 13.505398067619186)
((('cap', 'brimmed'), 13.505398067619186)
((('cardrossestate', 'dtrhfest'), 13.505398067619186)
((('carla', 'maxwell'), 13.505398067619186)
```

# Bigram - PMI

## Pointwise Mutual Information - II

- Frequency filter = 50
- Significantly different than freq\_filter = 3
- 3329 bigrams**
- List of top 50 scored

```
(('djsxer', 'electronicmusic'), 8.456488467138236)
((('production', 'djsxer'), 7.8329727256476875)
((('halter', 'top'), 7.23757099673789)
((('top', 'crochet'), 6.0811082551021425)
((('lovelmusic', 'newmusic'), 5.750510565455716)
((('ff', 'onairnow'), 4.9736539386643415)
((('dj', 'deephouse'), 4.923690501814605)
((('toptune', 'airplay'), 4.625814818006404)
((('toptune', 'newmusicalert'), 4.6258148180064005)
((('edm', 'ff'), 4.589518688783411)
((('live', 'np'), 4.574660730056298)
((('productor', 'toptune'), 4.505398067619183)
((('deephouse', 'productor'), 4.499084293686977)
((('album', 'newsong'), 4.496969445548601)
((('musicians', 'lovelmusic'), 4.44740634486006)
((('newsong', 'live'), 4.4458153575985015)
((('edm', 'production'), 4.37652496544921)
((('newmusic', 'album'), 4.28509489612615)
((('onairnow', 'deephouse'), 4.23785995762994)
((('np', 'music'), 4.2355452238765)
((('music', 'edm'), 4.180713143216078)
((('newmusicalert', 'musicfestival'), 4.150046971194371)
((('musicfestival', 'musicians'), 4.100483870505773)
((('toptune', 'musicfestival'), 3.988140144647982)
((('musicians', 'album'), 3.6629793969875273)
((('onairnow', 'dj'), 3.381733088243866)
((('dj', 'productor'), 3.0397344952703733)
((('ff', 'productor'), 2.6375016036265286)
((('music', 'festival'), 2.3726288748959696)
((('ff', 'dj'), 1.7986294314466633)
((('onairnow', 'productor'), 1.437653460983352)
((('newsong', 'music'), 0.6896145234909952)
((('productor', 'musicfestival'), 0.5075989758124528)
((('edm', 'productor'), 0.4474996839109835)
((('music', 'productor'), -0.43946037818835393)
```

# Bigram - PMI

## Pointwise Mutual Information - III

- Frequency filter = 100
- Significantly different than freq\_filter = 3
- 3329 bigrams**
- List of top 50 scored

```
('lovemusic', 'newmusic'), 5.750510565455716)
('ff', 'onairnow'), 4.9736539386643415)
('dj', 'deephouse'), 4.923690501814605)
('edm', 'ff'), 4.589518688783411)
('live', 'np'), 4.574660730056298)
('productor', 'toptune'), 4.505398067619183)
('deephouse', 'productor'), 4.499084293686977)
('album', 'newsong'), 4.496969445548601)
('musicians', 'lovemusic'), 4.447406344860006)
('newsong', 'live'), 4.4458153575985015)
('edm', 'production'), 4.37652496544921)
('newmusic', 'album'), 4.28509489612615)
('onairnow', 'deephouse'), 4.23785995762994)
('np', 'music'), 4.2355452238765)
('music', 'edm'), 4.180713143216078)
('musicfestival', 'musicians'), 4.100483870505773)
('toptune', 'musicfestival'), 3.988140144647982)
('musicians', 'album'), 3.6629793969875273)
('onairnow', 'dj'), 3.381733088243866)
('dj', 'productor'), 3.0397344952703733)
('ff', 'productor'), 2.6375016036265286)
('newsong', 'music'), 0.6896145234909952)
('productor', 'musicfestival'), 0.5075989758124528)
```

```
djxexer', 'electronicmusic'), 8.456488467138236)
production', 'djxexer'), 7.8329727256476875)
halter', 'top'), 7.23757099673789)
top', 'crochet'), 6.0811082551021425)
lovemusic', 'newmusic'), 5.750510565455716)
ff', 'onairnow'), 4.9736539386643415)
dj', 'deephouse'), 4.923690501814605)
toptune', 'airplay'), 4.625814818006404)
toptune', 'newmusicalert'), 4.6258148180064005)
edm', 'ff'), 4.589518688783411)
live', 'np'), 4.574660730056298)
productor', 'toptune'), 4.505398067619183)
deephouse', 'productor'), 4.499084293686977)
album', 'newsong'), 4.496969445548601)
musicians', 'lovemusic'), 4.447406344860006)
newsong', 'live'), 4.4458153575985015)
edm', 'production'), 4.37652496544921)
newmusic', 'album'), 4.28509489612615)
onairnow', 'deephouse'), 4.23785995762994)
np', 'music'), 4.2355452238765)
music', 'edm'), 4.180713143216078)
newmusicalert', 'musicfestival'), 4.1500469711943)
musicfestival', 'musicians'), 4.100483870505773)
toptune', 'musicfestival'), 3.988140144647982)
musicians', 'album'), 3.6629793969875273)
onairnow', 'dj'), 3.381733088243866)
dj', 'productor'), 3.0397344952703733)
ff', 'productor'), 2.6375016036265286)
music', 'festival'), 2.3726288748959696)
ff', 'dj'), 1.7986294314466633)
onairnow', 'productor'), 1.437653460983352)
newsong', 'music'), 0.6896145234909952)
productor', 'musicfestival'), 0.5075989758124528)
edm', 'productor'), 0.4474996839109835)
music', 'productor'), -0.43946037818835393)
```

# Results:

- Musical Trends and Genres
- EDM

(('music', 'edm'),  
0.039901393108983545)

- DeepHouse

(('music', 'deephouse'),  
0.0013759101072063293)



```
djxexer', 'electronicmusic'), 8.456488467138236)
production', 'djxexer'), 7.8329727256476875)
halter', 'top'), 7.23757099673789)
top', 'crochet'), 6.0811082551021425)
lovemusic', 'newmusic'), 5.750510565455716)
ff', 'onairnow'), 4.9736539386643415)
dj', 'deephouse'), 4.923690501814605)
toptune', 'airplay'), 4.625814818006404)
toptune', 'newmusicalert'), 4.6258148180064005)
edm', 'ff'), 4.589518688783411)
live', 'np'), 4.574660730056298)
productor', 'toptune'), 4.505398067619183)
deephouse', 'productor'), 4.499084293686977)
album', 'newsong'), 4.496969445548601)
musicians', 'lovemusic'), 4.447406344860006)
newsong', 'live'), 4.4458153575985015)
edm', 'production'), 4.37652496544921)
newmusic', 'album'), 4.28509489612615)
onairnow', 'deephouse'), 4.23785995762994)
np', 'music'), 4.2355452238765)
music', 'edm'), 4.180713143216078)
newmusicalert', 'musicfestival'), 4.15004697119437)
musicfestival', 'musicians'), 4.100483870505773)
toptune', 'musicfestival'), 3.988140144647982)
musicians', 'album'), 3.6629793969875273)
onairnow', 'dj'), 3.381733088243866)
dj', 'productor'), 3.0397344952703733)
ff', 'productor'), 2.6375016036265286)
music', 'festival'), 2.3726288748959696)
ff', 'dj'), 1.7986294314466633)
onairnow', 'productor'), 1.437653460983352)
newsong', 'music'), 0.6896145234909952)
productor', 'musicfestival'), 0.5075989758124528)
edm', 'productor'), 0.4474996839109835)
music', 'productor'), -0.43946037818835393)
```

# Results:

- ❖ Specific language structures, Slang
- ❖ np: no problem, now playing
- ❖ ff: fully functional
- ❖ crochet: body piercing



```
djxexer', 'electronicmusic'), 8.456488467138236)  
production', 'djxexer'), 7.8329727256476875)  
halter', 'top'), 7.23757099673789)  
top', 'crochet'), 6.0811082551021425)  
lovemusic', 'newmusic'), 5.750510565455716)  
ff', 'onairnow'), 4.9736539386643415)  
dj', 'deephouse'), 4.923690501814605)  
toptune', 'airplay'), 4.625814818006404)  
toptune', 'newmusicalert'), 4.6258148180064005)  
edm', 'ff'), 4.589518688783411)  
live', 'np'), 4.574660730056298)  
productor', 'toptune'), 4.505398067619183)  
deephouse', 'productor'), 4.499084293686977)  
album', 'newsong'), 4.496969445548601)  
musicians', 'lovemusic'), 4.447406344860006)  
newsong', 'live'), 4.4458153575985015)  
edm', 'production'), 4.37652496544921)  
newmusic', 'album'), 4.28509489612615)  
onairnow', 'deephouse'), 4.23785995762994)  
np', 'music'), 4.2355452238765)  
music', 'edm'), 4.180713143216078)  
newmusicalert', 'musicfestival'), 4.15004697119437  
musicfestival', 'musicians'), 4.100483870505773)  
toptune', 'musicfestival'), 3.988140144647982)  
musicians', 'album'), 3.6629793969875273)  
onairnow', 'dj'), 3.381733088243866)  
dj', 'productor'), 3.0397344952703733)  
ff', 'productor'), 2.6375016036265286)  
music', 'festival'), 2.3726288748959696)  
ff', 'dj'), 1.7986294314466633)  
onairnow', 'productor'), 1.437653460983352)  
newsong', 'music'), 0.6896145234909952)  
productor', 'musicfestival'), 0.5075989758124528)  
edm', 'productor'), 0.4474996839109835)  
music', 'productor'), -0.43946037818835393)
```

# Results:

- Festivals
  - burning man
  - Austin Music
  - ultra2019
  - coachella
  - halsburywest
  - microfest2018



```
djxexer', 'electronicmusic'), 8.456488467138236)
production', 'djxexer'), 7.8329727256476875)
halter', 'top'), 7.23757099673789)
top', 'crochet'), 6.0811082551021425)
lovemusic', 'newmusic'), 5.750510565455716)
ff', 'onairnow'), 4.9736539386643415)
dj', 'deephouse'), 4.923690501814605)
toptune', 'airplay'), 4.625814818006404)
toptune', 'newmusicalert'), 4.6258148180064005)
edm', 'ff'), 4.589518688783411)
live', 'np'), 4.574660730056298)
productor', 'toptune'), 4.505398067619183)
deephouse', 'productor'), 4.499084293686977)
album', 'newsong'), 4.496969445548601)
musicians', 'lovemusic'), 4.447406344860006)
newsong', 'live'), 4.4458153575985015)
edm', 'production'), 4.37652496544921)
newmusic', 'album'), 4.28509489612615)
onairnow', 'deephouse'), 4.23785995762994)
np', 'music'), 4.2355452238765)
music', 'edm'), 4.180713143216078)
newmusicalert', 'musicfestival'), 4.1500469711943)
musicfestival', 'musicians'), 4.100483870505773)
toptune', 'musicfestival'), 3.988140144647982)
musicians', 'album'), 3.6629793969875273)
onairnow', 'dj'), 3.381733088243866)
dj', 'productor'), 3.0397344952703733)
ff', 'productor'), 2.6375016036265286)
music', 'festival'), 2.3726288748959696)
ff', 'dj'), 1.7986294314466633)
onairnow', 'productor'), 1.437653460983352)
newsong', 'music'), 0.6896145234909952)
productor', 'musicfestival'), 0.5075989758124528)
edm', 'productor'), 0.4474996839109835)
music', 'productor'), -0.43946037818835393)
```

# Results:

- ◆ Costume/  
fashion trends
- ◆ bikini
- ◆ boho
- ◆ bohemian



# Conclusion

- The Music Festivals are the rising stars in entertainment industry, worth almost \$ 3billion globally.
- While the concept of a music festival hasn't changed in decades, technology and social media are making them more data-driven, more efficient and more accountable.
- Twitter text mining is one of the most crucial tools helping decision makers to reach their goals in the industry.

Thanks!

*– Questions?*



