

Article

The power of visual analytics and language processing to explore the underlying trend of highly popular song lyrics

Tanish Maheshwari^{1,*}, Tarpara Nisarg Bhaveshbhai¹ and Mitali Halder¹

¹ Presidency University, Bengaluru, Karnataka, India.

* Correspondence: tanishm72@gmail.com

Received: 1 March 2021; Accepted: 5 August 2021; Published: 30 September 2021.

Abstract: The number of songs are increasing at a very high rate around the globe. Out of the songs released every year, only the top few songs make it to the billboard hit charts. The lyrics of the songs place an important role in making them big hits combined with various other factors like loudness, liveness, speech ness, pop, etc. The artists are faced with the problem of finding the most desired topics to create song lyrics on. This problem is further amplified in selecting the most unique, catchy words which if added, could create more powerful lyrics for the songs. We propose a solution of finding the bag of unique evergreen words using the term frequency-inverse document frequency (TF-IDF) technique of natural language processing. The words from this bag of unique evergreen words could be added in the lyrics of the songs to create more powerful lyrics in the future.

Keywords: Lyrics extraction; Billboard hot 100 collections; Term frequency; inverse document frequency; Bag of unique evergreen words.

1. Introduction

Music influences a person on a neurological level. It impacts the temperament and conduct of a person [1]. It influences the function of the brain and human actions including, relieving stress, depression symptoms, as well as enhancing cognitive and motor functions, spatial-temporal learning and neurogenesis which is the capacity of the brain to create neurons [2]. The lyrics of the song play a major role in increasing this influence of music. For instance, songs with 'prosocial' lyrics may make you more empathic and could lead to long-term changes in attitudes and behavior for the better [3,4].

The problem arises in creating the lyrics of these songs. Artists face the problem of deciding topics to create songs on. The problem is further intensified in selecting the most unique, catch words which if added, could create more powerful lyrics for the songs. These powerful lyrics with the right combination of other music features (liveliness, loudness, speech ness, etc.) could create hit songs in the future.

A solution for this problem was proposed by this study which analyzed the lyrics of the billboard songs over the past 10 years. This study was carried out using the 'Hot 100 collection of Billboard songs' dataset from 'data world'. This dataset contains every weekly Hot 100 singles chart from 2009-2019. Each row of data represents a song and the corresponding position on that week's chart.

The lyrics of the songs in the dataset were analyzed to find the most important words in each song using TF-IDF technique of natural language processing.

The use of the different type of important words in songs captured the preference of users, being motivated to listen to the most popular hits of the time. This important word's analysis provided a deeper understanding on the choice of words used by the most popular artists.

Furthermore, a unique bag of evergreen words used in song's lyrics from this dataset for the period of 2009-2019 was also found. The lyrics made using these evergreen words could be used to create more powerful songs in the future. This hypothesis of including evergreen words in the song lyrics to create more powerful songs was validated by most desired genre 'Dance Pop' most beloved artist Katy Perry's songs lyrics.

2. Other Works

A study by Petrie *et al.*, [5] used LIWC (Lindquist Inquiry Word Count) for the analysis of the famous band's Beatles songs lyrics during its lifetime. It was observed that the effects of their lyrics were reduced in terms of emotions, focus, orientation, cognitive complexity. They became less social over time.

Another research named 'The Bob Dylan Encyclopedia, New York, Continuum', [6] was focused on repeated use of 'Ecstasy' in rap lyrics and its changing reference over the decade from 1996. The study found that the increase in the number of raps was directly proportionate to the increasing use of 'Ecstasy' among secondary school teenagers. It also considered changes in songs ciphered with positive, ambiguous, or negative messages about the use of this drug.

In a study Ronce [7], discussed the evolution of Dylan's song lyrics throughout 50 years of his career as an artist by using quantitative and qualitative analyses methods. It analyzed change in his word selection and theme selection explored in his songs in terms of qualitative and quantitative changes respectively. It was found that lyrics of the most popular songs could capture sociocultural changes across generations. A word-count method was used to compute the percentages of words belonging to various psychosocial categories including positive emotion, religion, social processes. Sociocultural studies captured word selection possibly indicating (Dylan's generation) generational changes in American culture. Further examination produced the correlations between the year of album release and the 10 LIWC categories of words. The correlations provided the fact that years were directly proportionate to the type of words used over time.

In another research conducted by Napier *et al.*, [8], the songs were grouped together by year in which they were produced and their tone scores were averaged for each year. The standard deviation and standard error were calculated for every averaged tone of every year. This was followed by conducting two tests to spot a linear relation between features, year and specific tone from the lyrics of the respective year. It was used to examine the relationship between two continuous variables. Furthermore, the linear dependency between these variables was validated using linear regression technique. This research concluded that some of Pearson's correlation coefficients (anger, disgust, fear, and conscientiousness) were considered of having strong positive correlations throughout. This analysis indicated the tonal change of popular music lyrics through generations. Over the years, an increment in some sentiments such as anger, disgust, fear, sadness, tentativeness and conscientiousness and decrement in sentiments such as joy, analytics, confidence and openness was found.

3. Methodology

The Billboard Dataset was used in carrying out the analysis of songs. An additional column of lyrics was added in the dataset to analyze the lyrics of the respective songs. Lyrics were extracted using 'Lyrics Extractor' API in python. The term frequency-inverse document frequency (TF-IDF) scores for the respective lyrics were calculated to find the important words in each song. This word feature selection was based upon selecting words having TF-IDF score greater than the minimum threshold of 0.2. The important words thus found were then visualized using the frequency distribution plot for the respective year in the dataset. This frequency distribution plot was also used in finding the evergreen important words from 2009-2019. These words if added in songs could be used to create more powerful lyrics, thereby helping in creating more hit songs in the future.

This hypothesis was validated by the most desired genre's most beloved artist. A bar chart was visualized to find the most desired genre, i.e., 'Dance Pop'. The most beloved artist, i.e., 'Katy Perry' according to the dataset was found using Pie chart visualized over the top 10 artists of the Dance Pop genre. The lyrics of the songs created by Katy Perry were visualized using word cloud. This word cloud gave an overview of words used in her songs. These words thus found were validated using the bag of unique evergreen words to prove this hypothesis.

Furthermore, a cluster map was plotted to find the correlation between different music features.

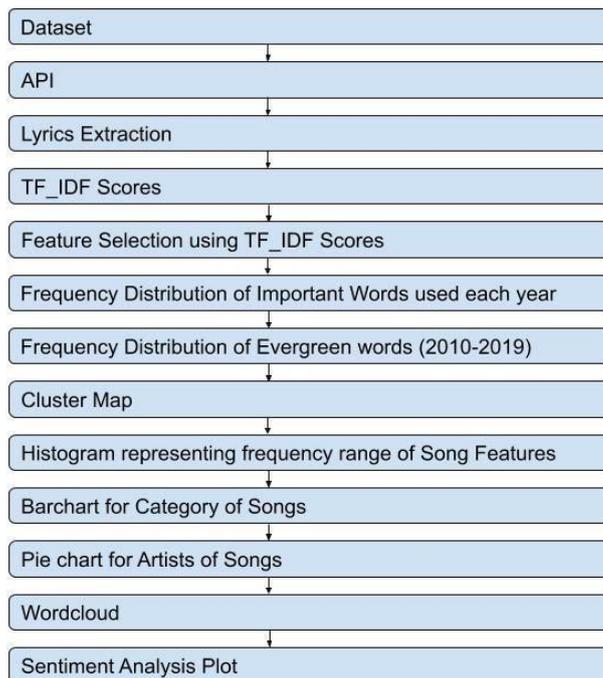


Figure 1. Hot 100 billboard songs dataset analysis flowchart

3.1. Lyrics extraction

The initial dataset was extended with an additional ‘lyrics’ parameter. The lyrics were required to perform sentiment analysis on songs. The lyrics for every song were fetched, extracted and added within the dataset using Python’s lyrics-extractor Library. This library extracts songs from numerous websites [9]. The library needs the title of the song for the extraction of the lyrics.genius.com web site [10] was used for extracting Billboard Top One Hundred Songs Collection. genius.com has the widest variety of song collections to look from. The GENIUS API (Application Performance Interface) was used for the aim of extracting songs from genius.com.

Genius API follows a collection of protocols, routines and tools for extraction of songs from genius.com web site. Lyrics extractor initially needed a Google API key and an Engine Id of Google custom Search JSON API. An API key is a distinctive selector that is used to validate a user, developer or calling application to an API. The custom search JSON API allows you to build websites and programs to fetch and show search results from Google custom search procedurally, it requires a unique authenticator key to initiate the search process (lyrics extraction). This API was integrated with GENIUS API as mentioned above to extract the lyrics of the specified song from the lyrics-extractor library.

title	artist	top genre	year	bpm	nrzy	dnce	dB	live	val	dur	acous	spch	pop	lyrics
Moment 4 Life - Album Version (Edited)	Nicki Minaj	dance pop	2011	130	88	50	-4	22	37	279	39	38	28.0	Nicki Minaj - Moment 4 Life Lyrics
Last Friday Night (T.G.I.F.)	Katy Perry	dance pop	2011	126	81	85	-4	67	72	231	0	4	27.0	Katy Perry - Last Friday Night (T.G.I.F.) Lyrics
Firework	Katy Perry	dance pop	2011	124	83	84	-6	11	85	228	14	6	26.0	Katy Perry - Firework Lyrics
Muny - Album Version (Edited)	Nicki Minaj	dance pop	2011	98	72	80	-5	19	53	227	56	33	7.0	Nicki Minaj - Muny Lyrics
Titanium (feat. Sia)	David Guetta	dance pop	2012	126	79	60	-4	13	30	245	7	10	80.0	David Guetta - Titanium Lyrics

Figure 2. Hot 100 Billboard Songs Dataset after extraction of Lyrics

3.2. Computation of term frequency-inverse document frequency scores of vocabulary of song

The term frequency inverse document frequency scores of word features of respective lyrics for each song were computed. This was followed by important word feature selection in the Dataset.

TF-IDF is defined as a statistical technique which evaluates the significance of the respective word in a document within a collection of documents (Dataset). Its applications include in the field of machine learning (Natural Language Processing) and automated text analysis for the computation of scores of word features [11].

It is computed by the multiplication of two metrics namely term frequency and inverse document frequency. The term frequency (TF) of a word in a document is the frequency of a particular word in the document.

The inverse document frequency (IDF) of the word across a set of documents is used to find the respective word's significance in the entire corpus (Dataset). It's closeness to zero is directly proportionate to the frequency of the word in the corpus. It is computed as the logarithmic value of the quotient of the total number of documents and the frequency of documents containing the word by taking the total number of documents.

This produces the TF-IDF score of a word in a document. The score is directly proportionate to the importance of the word in the particular document.

3.3. Extraction of most important features based on term frequency inverse document frequency scores (TF-IDF)

The sum of the TF-IDF score of each word was computed across the 'Lyrics' feature to find the relevance of the word in the dataset. The sum of the scores of word features possessing scores greater than the threshold of 0.2 were selected as the important word features for the respective lyrics of the song in the dataset. The value of 0.2 is conventionally taken as the threshold for selecting important word features based on the TF-IDF scores in the document dataset.

The TF-IDF features thus obtained was used for carrying out the analysis of song lyrics keywords used in lyrics of the songs present in the dataset.

4. Experimental setup

4.1. About dataset

This study uses the 'Hot 100 collection of Billboard songs' dataset from 'data.world' [12]. It contains 650 records divided based on 14 attributes. Some of the attributes included in the dataset are liveness, acoustics, energy, loudness, valence, song title, song artists, etc. Additional features lyrics, TF-IDF features were added to complement the analysis of songs present in the dataset.

The lyrics for each song were extracted using lyrics extractor API in python. an API is defined a set of functions and procedures which allows the creation of applications capable of accessing the features or data of an operating system, application, or other services.

Lyrics extraction was followed by finding the most important word of each song in the dataset for broadening analysis on the importance of words in creating lyrics of a song.

4.2. Data preprocessing

The lyrics found using lyrics extractor API were preprocessed for normalization of the 'Lyric' column. The lyrics were normalized using 're', 'nltk' (natural language toolkit) library in python.

Initially only words were selected using the 're' module of python, followed by normalizing the words into lowercase. The words were then converted into their base forms using stemming. Stemming is the process of decreasing a word to its word stem by affixing to suffixes and prefixes or to the roots of words known as a lemma.

This led to the stop words removal from dataset using natural language toolkit library in python. Useless words (data) are referred to as stop words in the dataset.

The corpus of lyrics thus obtained was used for calculating term frequency-inverse document frequency. The sum of the TF-IDF scores was used for word feature selection, selecting the important words in the lyrics in the dataset. The words features having sum greater than threshold value (0.2) computed across the dataset were selected to be used for further analysis and visualization processes.

5. Visualization and analysis of features

The word features obtained after the TF-IDF feature selection technique was used for finding the top 50 words used in songs for every year from 2009-2019. These word features were visualized as frequency distribution plots to support their importance in the dataset. Further analysis led to the discovery of the evergreen words in the past 10 years and the correlation of the features of the dataset using cluster maps and pair plots. Finally, our results for the evergreen words were validated by analyzing the features used in the songs of the most popular pop-dance category’s artist Katy Perry.

5.1. Frequency distribution plot

The most important word features (top 50 words) used in lyrics of songs for each year from 2009-2019 were visualized using ‘nlk’ and ‘matplotlib’ library in python. For the year 2010, the most important words were ‘like’, ‘wanna’, ‘go’, ‘yeah’, ‘oh’, ‘love’ etc. This was followed by ‘girl’, ‘run’, ‘yeah’, ‘baby’ etc in 2011 and ‘love’, ‘feeling’, ‘tonight’, ‘like’, ‘talk’ in 2019 (Figure 3) and so on. It was also found how the lyrics have changed in the short span of 10 years, as the result of the change in listening preferences of billboard chart listeners.

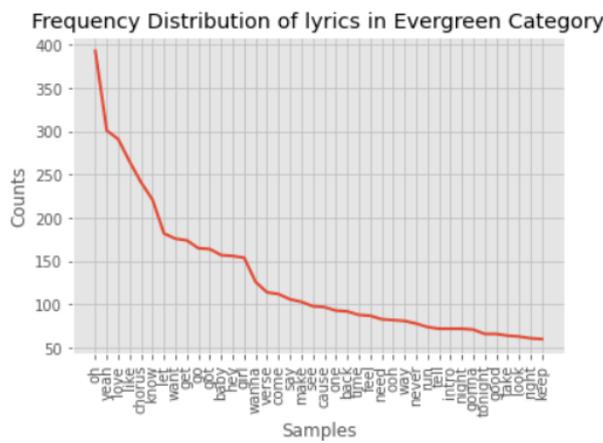


Figure 3. Frequency distribution of evergreen words used in lyrics (2009-2019)

Analysis on genres of songs and their respective artists revealed that the most popular genre according to the frequency of songs was found to be ‘Dance Pop’ (Figure 4).

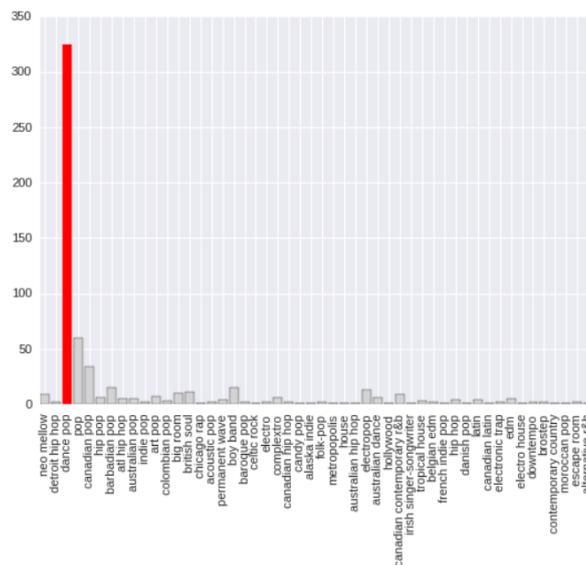


Figure 4. Genres of songs

Further it was found that the most beloved artist among the top ten artists was found to be Katy Perry producing 12.8% of the total billboard songs (Figure 5). On analysis of Katy Perry’s songs, it was found that maximum songs produced belonged to the ‘Dance Pop’ genre.

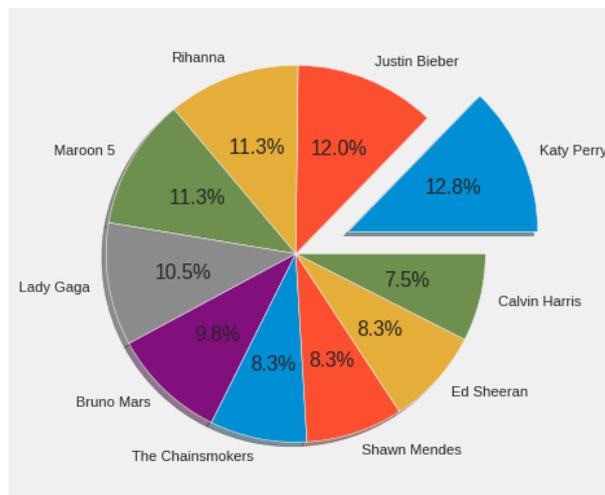


Figure 5. Pie chart representing the top 10 (2009-2019)

5.2. Histograms

The histograms of the most popular ‘Dance Pop’ genre artist Katy Perry songs features were visualized to find the range of frequencies of features like valence, acoustic ness, liveliness, etc.

Katy Perry’s songs revealed frequency range of (110-130), (80-85), (60-70), (80-85), (-5.5 - -4.5), (0-2), (4-11) of beats per minute, energy, dance, decibels, acoustic ness, speech ness features respectively (Figure 6).

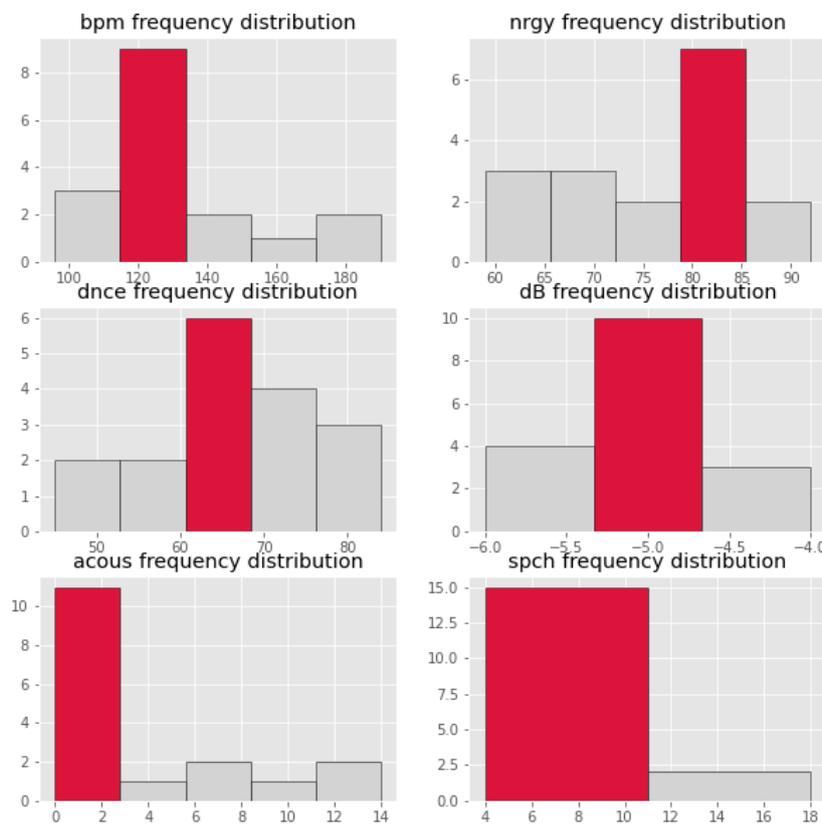


Figure 6. Histogram representing the frequency range of features used in Katy Perry’s songs

5.3. Cluster map

The features of the dataset excluding ‘TF-IDF-features’ (important words) attribute like pop, acoustic ness, liveliness, etc, were explored using cluster map (Figure 7) to gather insights on the correlation or dependence of different features of songs. The cluster map was created ‘seaborn’ Library in Python.

A cluster map is created using hierarchical clustering techniques to find the amount of similarity or correlation between two features. It was also found that the features energy and loudness (dB) are highly correlated with a positive score of +0.54. This correlation concluded the fact that people doing workouts listen to loud music to achieve efficiency in their workout.

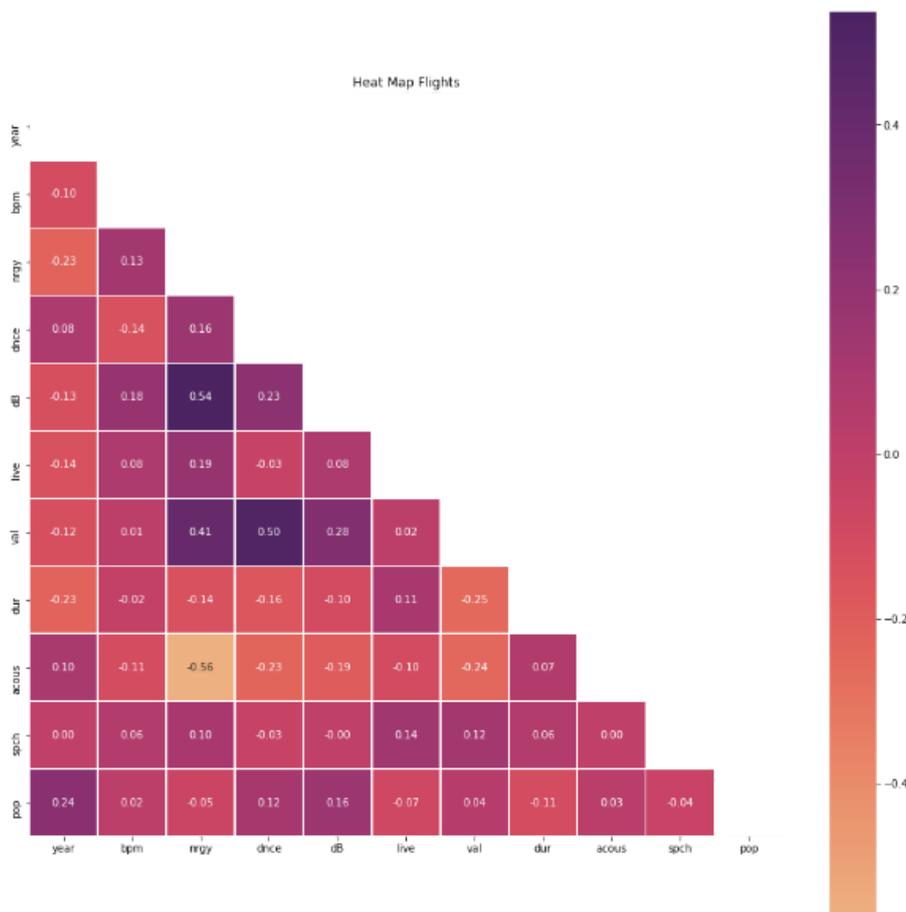


Figure 7. Cluster map representing the correlation between different features of the dataset

Acoustics (the characteristic of a space for determination of transmission of sound through it) are not correlated with energy, showcased with a value of -0.56 on the correlation matrix. Liveliness or the probability of live audience listening to the song is not at all correlated with the duration of the song $+0.11$. This means that liveliness or the moment a song tries to create does not depend on the length (duration) of the song. It can be felt within the song, during the entire song or not at all in the song.

5.4. Word cloud

A word cloud was made from the most important word features in the ‘Dance Pop’ star Katy Perry’s song’s lyrics.

Word cloud also known as a text cloud or a tag cloud is based on simple technique, i.e., the boldness and the size of the word is directly proportionate to the frequency of the word in the specific document. They are generally used for analyzing customer feedback/review and identifying new SEO (Search Engine Optimization) song lyrics keywords to target. The word cloud was created using the ‘wordcloud’ and ‘pandas’ library and was visualized using ‘matplotlib’ library in python. The ‘wordcloud’ is a library licensed by MIT. It also includes ‘DroidSansMono.ttf’ apache licensed, a true style font developed by Google [13].

The word cloud generated from Katy Perry’s, most beloved artist according to billboard (2009-2019) dataset was visualized to give overall view of the song lyrics keywords used in her songs lyrics as well as for validating the most frequent terms of the word cloud with the most frequent terms from the evergreen words distribution plot. It was found that words like ‘love’, ‘night’, ‘friday’ (also appearing in the top 50 evergreen words) used in Katy Perry’s songs were also the most catchy, loveable and desired words for the listeners during the period of 2009-2019 (Figure 8).



Figure 8. Word cloud representing the frequency of important words in the Katy Perry song’s lyrics

These words found in her songs motivated the audience to listen more to these songs. This also led to an increase in the audience for these ‘Dance Pop’ songs, thereby increasing the number of hits of her songs.

5.5. Sentiment analysis

The song lyrics of Katy Perry’s songs were further analyzed on the basis of sentiment, showcased in her songs.

Sentiment analysis is deciphering, analyzing and classifying emotions (like happy, sad, neutral) within textual data using analytical techniques. In methodological terms, sentiment analysis detects polarity within the text, documents, comments etc. It assists organizations in distinguishing an individual’s sentiments towards products or services like garments, music etc.

Understanding the emotions through people’s point of view is vital as people will communicate their conclusions even more uninhibitedly now in light of digitization. In this way, by analyzing the feedback given by the people, brands will tailor their products and services regarding the people’s necessities [14]. The ‘vaderSentiment’ library was used to find the polarity scores of the songs [15].

The data frame containing ‘positive’, ‘negative’, ‘neutral’, ‘compound’ scores was plotted as a bar plot using ‘pandas’ library to visualize the sentiment present in her song’s lyrics (Figure 9). This also helped in highlighting the type of sentiment favored by the audience in song’s lyrics

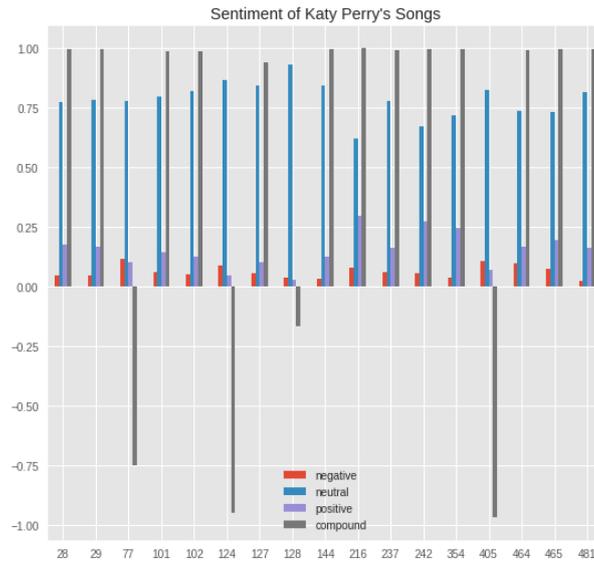


Figure 9. Sentiment plot highlighting the sentiment of Katy Perry’s songs lyrics

6. Results

The popularity of the songs based upon the no of unique evergreen words was plotted and compared with the popularity of songs based on the no of YouTube views (Figure 10). The plot showed that 12 out of 17 songs of the dataset’s most beloved artist Katy Perry followed a direct proportionality trend. This trend showed that the number of YouTube views were directly proportional to the no of important words present in the songs. These important words were also present in the unique bag of evergreen words. For instance, a song ‘Dark horse’ had 83 unique evergreen words in its lyrics and had a ‘very popular’ likeability (Figure 11). Similarly, another song ‘birthday’ had only 3 evergreen words and had a ‘popular’ likeability. This analysis found that the increase in the number of evergreen words made the lyrics more powerful and in turn increased the number of views of the song (popularity).

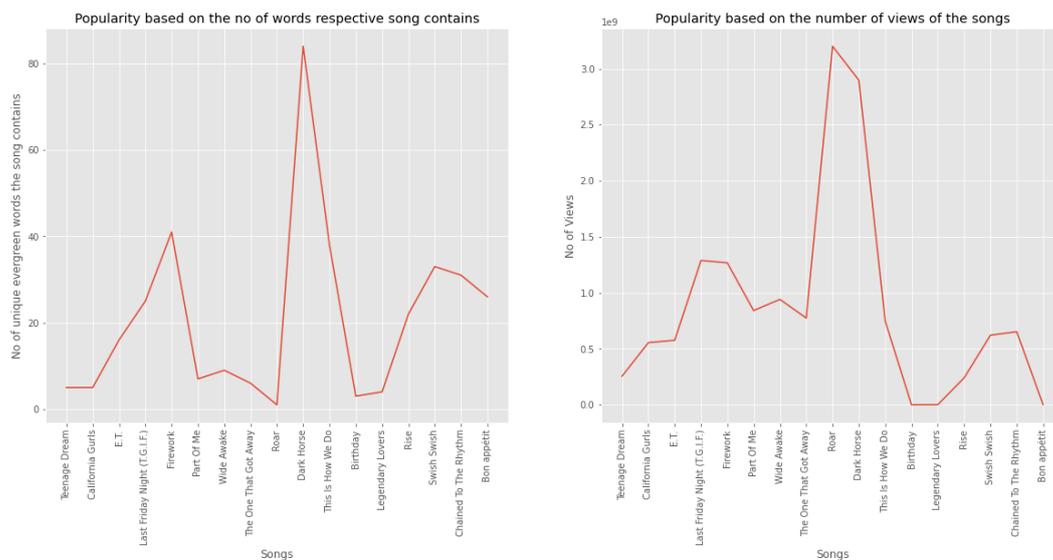


Figure 10. Line Plot comparing the popularity of Katy Perry’s Songs lyrics based on the unique evergreen words

	title	artist	top genre	year	bpm	energy	dance	db	live	val	dur	acous	spch	pop	lyrics	tfidf_features	views	likeability	No of Words it contains
1	Teenage Dream	Katy Perry	dance pop	2010	120	80	72	-5	13	59	228	2	4	63	[Verse 1] You think I'm pretty without any mak...	['back', 'ever', 'let', 'look', 'tonight]	256004650	Popular	5
2	California Gurls	Katy Perry	dance pop	2010	125	75	79	-4	18	40	235	0	5	62	[Intro: Snoop Dogg] Greetings, loved ones Let...	['oh', 'beach', 'love', 'california', 'oh]	555000735	Popular	5
3	E.T.	Katy Perry	dance pop	2011	152	87	62	-5	37	76	230	2	18	66	[Verse 1] You're so hypnotizing Could you be t...	['could', 'devil', 'hypnotizing', 'verse', 'an...]	576171600	Popular	16
4	Last Friday Night (T.G.I.F.)	Katy Perry	dance pop	2011	126	81	65	-4	67	72	231	0	4	27	[Verse 1] There's a stranger in my bed There's a...	['barbeque', 'barbies', 'bed', 'bruise', 'dj', ...]	1288239765	Very Popular	25
5	Firework	Katy Perry	dance pop	2011	124	83	64	-5	11	65	228	14	5	25	[Verse 1] Do you ever feel like a plastic bag ...	['bag', 'drifting', 'ever', 'leef', 'like', 'p...]	1267290164	Very Popular	41
6	Part Of Me	Katy Perry	dance pop	2012	130	92	68	-5	7	77	216	0	4	71	[Verse 1] Days like this, I want to drive away...	['away', 'ever', 'gonna', 'never', 'part', 'la...]	841432743	Popular	7
7	Wide Awake	Katy Perry	dance pop	2012	160	68	51	-5	39	57	221	7	4	68	[Intro] I'm wide awake I'm wide awake I'm wide...	['awake', 'wide', 'awake', 'chorus', 'cloud', ...]	940880306	Very Popular	9
8	The One That Got Away	Katy Perry	dance pop	2012	134	80	69	-4	16	88	227	0	4	67	[Verse 1] Summer after high school, when we fi...	['another', 'away', 'got', 'lifo', 'one', 'wou...]	774137177	Popular	6
9	Roar	Katy Perry	dance pop	2013	180	77	55	-5	35	46	224	0	4	78	[Verse 1] I used to bite my tongue and hold my...	['oh]	3200328952	Very Popular	1
10	Dark Horse	Katy Perry	dance pop	2014	132	59	65	-6	17	35	216	0	5	78	[Intro: Juicy J] Yeah, ya'll know what it is K...	['huh', 'intro', 'juicy', 'katy', 'know', 'let...]	2897126320	Very Popular	84
11	This Is How We Do	Katy Perry	dance pop	2014	96	64	69	-6	15	80	204	2	5	66	[Intro] This is how we do This is how we do (O...	['big', 'deal', 'oh', 'yeah', 'respect', 'back...]	748136940	Popular	38
12	Birthday	Katy Perry	dance pop	2014	126	67	72	-5	11	84	215	9	7	65	[Verse 1] I'm another year older I won't cry a...	['birthday', 'like', 'yeah]	22024	Popular	3
13	Legendary Lovers	Katy Perry	dance pop	2015	126	84	62	-5	8	39	224	2	8	41	[Verse 1] Under a silver moon, tropical temper...	['la', 'legendary', 'lover', 'oh]	1429883	Popular	4
14	Rise	Katy Perry	dance pop	2016	101	64	55	-6	11	44	203	5	5	66	[Verse 1: Mako] Welcome to the wild, no heroes...	['along', 'got', 'mako', 'rise', 'smith', 'tel...]	240447349	Popular	22
15	Swish Swish	Katy Perry	dance pop	2017	120	71	84	-5	10	57	243	2	4	73	[Refrain] They know what is what But they don't...	['fuck', 'know', 'refrain', 'strut', 'another'...]	621243570	Popular	33
16	Chained To The Rhythm	Katy Perry	dance pop	2017	190	80	45	-5	20	47	238	8	17	72	[Verse 1: Katy Perry] Are we crazy? Living our...	['crazy', 'katy', 'perry', 'verse', 'bubble', ...]	652294244	Popular	31
17	Bon appétit	Katy Perry	dance pop	2017	106	79	80	-6	22	52	228	13	6	68	[Intro: Quavo] Ayy, yeah Katy Perry, Migos Ayy...	['ayy', 'intro', 'katy', 'migos', 'perry', 'qu...]	1992348	Popular	26

Figure 11. Table summarizing the Katy Perry’s Songs popularity based on the lyrics

7. Conclusion

The analysis study found the bag of unique evergreen words using term frequency inverse document frequency and frequency distribution plot. This unique bag of evergreen words contained 2696 words for the songs analyzed from 2009-2019. The top 50 unique evergreen words included ‘love’, ‘like’, ‘yeah’, ‘baby’, ‘let’, ‘friday’, ‘night’, etc. The hypothesis that lyrics made using these unique evergreen words could be used to create more powerful lyrics in the future was validated using most popular artist Katy Perry’s song’s lyrics. This analysis also concluded that the no of views (popularity) were positively correlated to the number of words used in the song. The increment in the no of unique evergreen words used in the song also led to an increase in the number of views of the respective songs.

This analysis was also found useful in deciding the genre for creating songs on, which was ‘Dance Pop’. The words from the unique bag of evergreen words also could be used in deciding the title on which songs could be created to increase the no of hits of the songs. For example ‘last friday night’ song by Katy Perry was among the most popular songs. The song lyrics keywords ‘friday’ and ‘night’ are also present in bag of unique evergreen words. Similarly a song ‘love me like you do’ by Ellie Goulding also contains the evergreen words ‘like’, ‘love’, ‘you’ gained large no of hits and was loved by the audience. The title and the category could also be used to capture the sentiment loved by the people (love and like here). The title of the song could be further used to investigate the emotions special to that topic, for example, the ‘Friday’ could be used to capture the specialty of that day. Similarly, a word ‘like’ could be used to capture the feeling of being liked or liking something in the songs.

The analysis in future could deepen its understanding of songs using the tonal sequence while analyzing the popularity of the respective song. Furthermore, an increment in the number of songs in the dataset could help in improving the accuracy of this analysis to create more powerful song’s lyrics in the future.

Author Contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflicts of Interest: “The authors declare no conflict of interest.”

References

- [1] Storr, A. (2015). *Music and the Mind*. Simon and Schuster.
- [2] Louise Nornes (2016). *How Does Music at Events Affect People’s Behaviour?* IOP Publishing Eventbrite.
- [3] Frith, S. (2001). *The Cambridge companion to pop and rock*. Cambridge University Press.
- [4] Allen, S. (2017). *Five Ways Music Can Make You a Better Person*. IOP Publishing Greater Good Magazine.
- [5] Petrie, K. J., Pennebaker, J. W., & Sivertsen, B. (2008). Things we said today: A linguistic analysis of the Beatles. *Psychology of Aesthetics, Creativity, and the Arts*, 2(4), 197–202.
- [6] Shepherd, J. (2003). *Continuum Encyclopedia of Popular Music of the World: Performance and production. Volume II (Vol. 1)*. A&C Black.
- [7] Ronce, O. (2007). How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics*, 38, 231-253.
- [8] Napier, K., & Shamir, L. (2018). Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*, 30(4), 161-176.

- [9] <https://pypi.org/project/lyrics-extractor/>.
- [10] <https://genius.com/>.
- [11] Stecanella, B. (2019). *What-is-TF-IDF?* IOP Publishing monkey learn.
- [12] <https://data.world/kcmillersean/billboard-hot-100-1958-2017>.
- [13] <https://pypi.org/project/wordcloud/>.
- [14] Sentiment Analysis: A Definitive Guide. IOP Publishing monkey learn Everything There Is to Know about Sentiment Analysis (monkeylearn.com). Accessed 17 October 2020.
- [15] <https://pypi.org/project/vaderSentiment/>.



© 2021 by the authors; licensee PSRP, Lahore, Pakistan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).