



Final Year Project

Colm Gallagher (T00170514)

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

Submitted To: Andrew Shields, Paul Collins, Helen Fitzgerald, Robert Sheehy,
Peter Given, & Padraig Moriarty

Submitted By: Colm Gallagher (T00170514)

B.Sc. (HONS) Computing with Media Development

Date Submitted: 12/05/2021

**'Declaration of Originality Form' - Institute of Technology, Tralee.**

This form **must** be completed and signed and submitted with all assignments.

Please complete the information below (using BLOCK CAPITALS).

Name: COLM GALLAGHER

T Number: T00170514

Class Group: CP400 COMPUTING WITH MEDIA DEVELOPMENT

Assignment Title: SENTIMENT ANALYSIS OF UK NUMBER ONES (1952-2020) TO DETECT CORRELATION BETWEEN LYRICS AND SOCIETY.....

Students are advised to inform themselves of the Institute Anti-Plagiarism Policy.

I confirm that this assignment is my own work and that I have:

Familiarised myself with the Institute's Anti-Plagiarism Policy

Used the Institute's approved referencing style throughout

Clearly referenced, in both the text and the bibliography or references, all sources used in the work

Not made use of the work of any other student(s) past or present without acknowledgement. This includes any of my own work, that has been previously, or concurrently, submitted for assessment, either at this or any other educational institution

Not sought or used the services of any professional agencies to produce this work

In addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with Institute regulations

Acknowledgements

I would like to thank my lecturers for their four years of continuous patience, support, and dedication to helping me to accomplish a goal set in 2017 to complete my Bachelor of Science (HONS) in Computing with Digital Media. Special thanks goes to my supervisor, Andrew Shields, who has always been so generous with his time in order to encourage and support me when it was sorely needed and requested from him. Intelligent, practical, and most importantly, unflappable; I could not have wished for a better mentor. Helen Fitzgerald is another who has given me lots of great guidance during the past four years. I think it only fair to acknowledge all of the lecturers and staff members who have played a role in my achievement. Without them, I am sure I would not be where I am, and I will forever appreciate their efforts:

Mairead O'Sullivan, Gene Farrell, David Crowley, Owen Mackessy, Claire Horgan, Catherine Woods, Cathryn Casey, Peter Given, Enda Grennan, Susan Motherway, Rob Sheehy, Aine O'Mahony, Therese Enright, Phillip Doyle, Bob Jackson, Fergus McLysaght, Padraic Moriarty, Linda O'Loughlin, Deirdre Boyle, Lisa McElligott, Liam McCarthy, and Jim Quinn.

As a mature student, entering education again after such a long absence has been one of the toughest challenges I have ever taken on. There have been numerous occasions where I have considered quitting due to stress and loss of confidence, amongst other personal issues. But I have persevered, and much of that was because of the enormous amount of kindness, listening ability, and confidence shown by the following people:

John Brosnan, Anne O'Brien, Paul Collins, John Walsh, Brigid Crowley, and Valerie Moore. I would also like to express my eternal appreciation, and admiration, to Denis O'Sullivan. I honestly believe that Denis changed my life for the better in more ways than I could mention. He is owed a huge debt of gratitude for his support to me, which I will never forget.

Sincere thanks also goes to my peers. All have played their part, but special mention must go to those who offered help and support no matter how busy they were with their own study; Tracey Brosnan, Paul Cafferkey, Daragh Carroll, Jason Dowling, James Mahoney, Peter McCafferty, Carraig McNulty, Darragh O'Brien, Ian O'Regan, and Caitie Pearce. Genuinely some of the nicest, and hardest working people I have had the pleasure of sharing college life

with. They are incredibly helpful, talented, and respectful. The industry is in great hands with them as its future.

My parents, Micheál and Kathleen, have never lost their faith in me, even at my lowest points. Nor, for that matter, have my siblings and friends.

Lastly, the two women in my life who this final project is dedicated to. I would like to thank my wonderful partner, Mary, for her loyalty, encouragement, love, and understanding. It has not been easy for her to witness my struggles throughout the course. I have not always been the easiest to live with and I am sure she will be relieved for me to eventually get to sleep at a reasonable hour. Talking of lack of sleep, to baby Caitilín...this is for you!

Abstract

Sentiment Analysis, also known as opinion mining, is the process of gathering data on the opinions and emotional feelings, in text form, emanating from human behaviour. This process is achieved with techniques such as Machine Learning and Natural Language Processing (NLP), using algorithms to understand the tone of the sentiment being expressed by the user.

The objective of this project was to establish whether there appeared to be an obvious correlation between music lyrics and society over the period of the UK singles charts (1952-2020). Each set of lyrics were given a polarity score of between -1 (negative) and 1 (positive).

A prototype was completed on the American Billboard 100 charts using Python code in conjunction with Jupyter Notebook. A web scrape of the official UK charts website was then executed using BeautifulSoup to obtain every UK number one single. This dataset was amended by acquiring the sentiment with TextBlob, and the associated lyrics to each song from Genius.com. Once the data was cleaned, the sentiment analysis was performed utilising Microsoft Power BI.

The resulting data indicated that the lyrical content of UK number ones has become gradually more negative across the history of the UK singles chart. Although, decreasing by an average of 0.67% annually, the average sentiment of UK number ones has remained at a consistently neutral level.

No definitive proof of lyrical content was identified as impacting on societal behaviour, or vice versa. However, on observing the most commonly used words of each year in the study, they would illustrate an unequivocal trend towards profanity, misogyny, racism, and so forth.

Several instances of erroneous lyrical data was supplied by Genius.com, something to consider if revisiting the process as time would need to be allocated to correcting the issue.

Keywords:

sentiment analysis; Text Mining; natural language processing (NLP); UK number ones; music lyrics; culture and society

Table of Contents

Acknowledgements.....	iii
Abstract	v
Chapter 1: Introduction	1
1.1 Overview	1
1.2 Problem Statement	2
1.3 Aims and Objectives	2
Chapter 2: Sentiment Analysis.....	3
2.1 Introduction	3
2.1.1 Positive	3
2.1.2 Neutral	3
2.1.3 Negative.....	4
2.2 Sentiment Analysis history.....	5
2.3 Types of Sentiment Analysis	8
2.3.1 Fine-Grained.....	8
2.3.2 Coarse-Grained.....	8
2.3.3 Aspect-Based.....	9
2.3.4 Emotion Detection	9
2.4 Sentiment Analysis Algorithms	11
2.5 Sentiment Analysis Challenges	11
2.5.1 Context.....	11
2.5.2 Sarcasm.....	12
2.5.3 Recognition.....	12
2.5.4 Modern Language.....	12
Chapter 3: Music Lyrics.....	14
3.1 Introduction	14

3.2	Stereotyping of Music Genres.....	15
3.2.1	Hip-Hop	17
3.2.2	Heavy Metal	17
	Chapter 4: Methodology	18
4.1	Key Research Findings.....	18
4.2	Research Question.....	19
4.2.1	Research sub-questions.....	20
4.3	Project Scope.....	20
4.4	Proposed Methodology	20
4.4.1	Functional Specifications	21
4.4.1.1	MoSCoW Method.....	21
4.4.1.2	User Stories	22
4.5	Risk Analysis.....	30
4.6	Prototype	32
	Chapter 5: Implementation	36
5.1	Introduction	36
5.2	Sprint Schedule	36
5.3	Sprint 1 – Understanding prototype and Python code.....	37
5.4	Sprint 2 – Web scraping and cleaning of UK chart data	43
5.5	Sprint 3 – Running UK data and performing sentiment analysis.....	47
5.6	Sprint 4 – Researching UK cultural and societal data.....	69
5.7	Sprint 5 – Bonus data	75
5.8	Sprint 6 –Answering research question, & preparing final submission.....	76
	Chapter 6: Results Analysis of Data	78
6.1	Introduction	78
6.2	Lyrics.....	78

Chapter 7: Discussion & Conclusions	86
7.1 Introduction	86
7.2 Discussion	86
7.3 Answering the Research Question	87
References.....	88
Bibliography	97
Appendices.....	124
Appendix A - Prototype screenshots.....	124
Appendix B – Lemmatisation vs Stemming	134
Appendix C – Web scraping and cleaning data in Power BI.....	137
Appendix D – Google Colab run of UK dataset	142
Appendix E – Sentiment Analysis in Power BI	148
End Notes.....	187

Chapter 1: Introduction

1.1 Overview

“a single unit of language that has meaning and can be spoken or written” (word, 2021)

Words are incredibly powerful tools. We tell our parents, our partners, or our children that we love them, and it sets off a ball of emotion for all concerned. Then, there are times that we are stuck in traffic, or hit our toe off the corner of the bed; again, we use words to express our intense frustration or pain (some of *these* words will be discussed later in this paper).

Some of the most powerful figures throughout history have used words to great effect.

Winston Churchill’s, “We Shall Fight on the Beaches”, roused the nation and reassured his people early on in World War II; The sense of a new dawn arising through Martin Luther King’s, “I Have a Dream”, brought black people hope that, at long last, their voice would finally be heard.

“a short poem that expresses the personal thoughts and feelings of the person who wrote it” (lyric, 2021)

Yes, words are incredibly powerful. But added to music as lyrics, they take that power to a whole new level. Many of us will have heard the single line, “We shall fight on the beaches”, but how many of us, in all honesty, could recite much more of the speech? But if we were asked to sing the chorus to “Mr Brightside”, by The Killers after a few drinks in a bar? I rest my case m’lud.

This thesis will explore the sentiment behind the lyrics, how positive, negative, or neutral they may be. The literature review chapter will look to explain both sentiment analysis, and music lyrics, in some more detail.

In the methodology chapter, the tools, and technologies of how the sentiment analysis will be produced are examined, as well as showcasing a fully functioning prototype.

The following chapter will give a detailed journey of the implementation process from the early period of gathering data, through to the performing of the sentiment analysis.

My results and findings will then be discussed as I conclude the thesis with, *hopefully*, a sufficient answer to the research question. Hopefully, you will enjoy this paper, but if not...

“The record shows I took the blows, and did it...my way” (Frank Sinatra – My Way, 2021)

1.2 Problem Statement

Research question

Can sentiment analysis of UK number ones (1952-2020) detect correlation between lyrics and society?

1.3 Aims and Objectives

Aim

To discover whether there exists a tangible connection between music lyrics and trends in society of the same era.

Objectives

- Produce a fully functioning prototype
- Find a comprehensive UK chart dataset
- Perform sentiment analysis on the data
- Present results and findings in an articulate and reasoned manner

Chapter 2: Sentiment Analysis

2.1 Introduction

Sentiment Analysis, also known as *Opinion Mining*, is the process of gathering data on the opinions and emotional feelings, in text form, emanating from human behaviour. This process is achieved by using techniques such as *Machine Learning* and *Natural Language Processing* (NLP), using algorithms in order to understand the tone of the sentiment being expressed by an individual user, or *users* (Bannister, Kristian, 2020).

The purpose of the algorithms being used is to separate these sentiments into three specific categories of polarity: *Positive (+1)*, *Neutral (0)*, or *Negative (-1)*. These categories are explained in better detail below with the use of text.

2.1.1 Positive

An example of a *positive* customer review where the adjective “Fantastic” clearly shows the satisfaction of the customer, therefore giving a highly positive score.

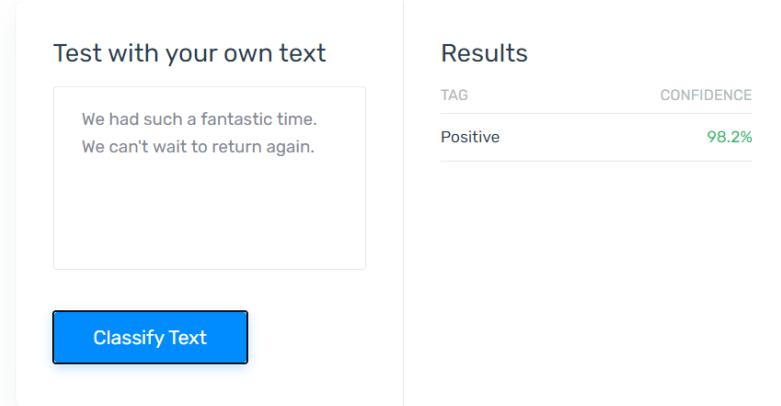


Figure 1- Example of Positive analysis (Everything There Is to Know about Sentiment Analysis, 2020)

2.1.2 Neutral

The following is an example of a neutral sentence. A *positive* is the baby arriving but the issue of having to wait is construed as a *negative*.

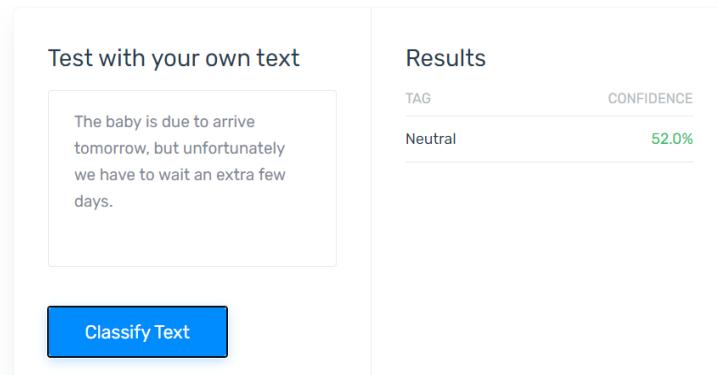


Figure 2 - Example of Neutral analysis (Everything There Is to Know about Sentiment Analysis, 2020)

2.1.3 Negative

This is clearly proposed as a *negative* review by the customer. The use of the word “Terrible” and the fact that they will not be going back ensures a very high *negative* score.

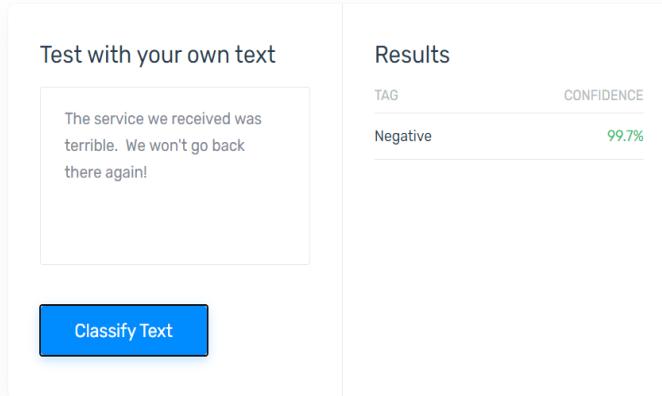


Figure 3 - Example of Negative analysis (Everything There Is to Know about Sentiment Analysis, 2020)

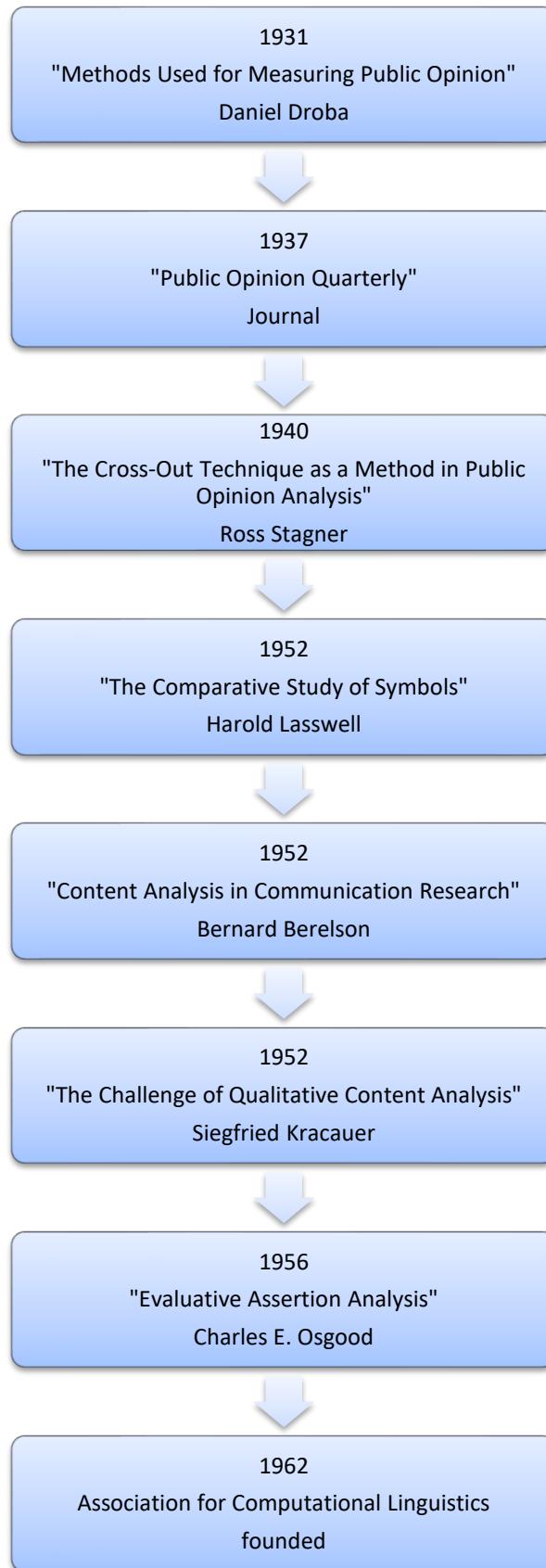
Sentiment Analysis is becoming ever more important to businesses, in particular, who use the process as a tool to better understand the feelings of their customer base. This gives them the information they require in order to improve certain services, drop others, or be satisfied that some are working perfectly well. With social media becoming more prominent each year, the opportunity for customers to express their opinions is more accessible than ever, and so companies will actively monitor these feelings on a wider scale, making sentiment analysis a vital component moving forwards.

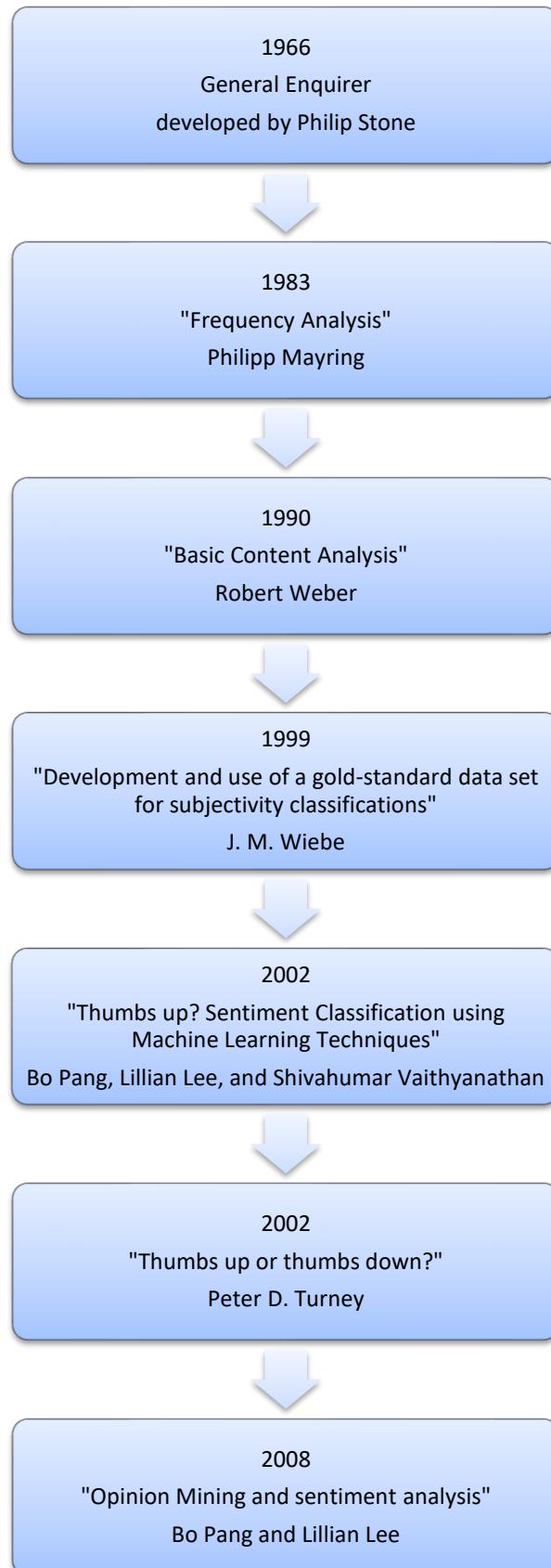
It is not just the commercial world that benefits from Sentiment Analysis. Politicians, for example, are increasingly using it to gauge the opinions of voters in order to help them prepare policies which would likely receive positive reaction from the public. During the 2012 U.S. election campaign, Barack Obama used sentiment analysis as a tool to aid with his

aspiration of becoming President (Bannister, Kristian, 2020). And since 2014, the field of social media has eclipsed market research as the most commonly produced papers on the topic of sentiment analysis (Mäntylä, Graziotin and Kuutila, 2018).

2.2 Sentiment Analysis history

Sentiment Analysis is one of many branches which descend from *Content Analysis*. To understand the origins of sentiment analysis, one must delve deeper into its predecessors. (Kuckartz, 2019) and (Prasad, 2008) note that “Content Analysis” was a well-known method used by the American’s during the Second World War to analyse enemy propaganda under the directorship of Harold Lasswell. But the concept of *Content Analysis* can be traced back to at least the time of the Ancient Greek’s (Mäntylä, Graziotin and Kuutila, 2018), and latterly, to the reaction of 17th century churches in relation to the printing press (Shelley and Krippendorff, 1984). Some of the more noteworthy milestones follow:





2.3 Types of Sentiment Analysis

2.3.1 Fine-Grained

This could be seen as the most precise of the sentiment analysis types due to it having a five-tier score system. As well as the more commonly used, *Positive, Neutral, and Negative*, *Very Positive* and *Very Negative* are also added on for more minute detail into the customers feelings.

Whilst this process can be seen to provide greater polarity, it can also prove to be more costly to implement. Ironically, it can hit both ends of its own scale (Goyal, Kechit, 2020)

2.3.2 Coarse-Grained

This type of sentiment analysis is performed at, both, document level and sentence level. At document level the process focuses on a piece as a whole. The example below could be a simple review on Amazon.

“My laptop is superb. It provides me with exactly what I need in relation to functionality.”

This sentiment is wholly focussed on the laptop in general terms and does not delve into the specific pros and cons of the machine. This means that a positive review would be understood from the sentence, which is acceptable as the customer is displaying clear satisfaction with the product.

At sentence level, however, the user may express further, more in-depth, opinions on the machine. The first task of coarse-grained sentiment analysis is to determine whether a sentence is *subjective* or *objective* (Sentiment Analysis: Types, Tools, and Use Cases, n.d.). An *objective* example in relation to the above user would be as follows.

“The computer has 256 gigabytes of memory and 8 gigabytes of RAM.”

What is being stated above is simply fact, not opinion. The second task of coarse-grained sentiment analysis is to decide the emotion of a *subjective* sentence. All of the below quoted examples would be classified as *subjective* as they express a personal opinion. If they were to produce this in a blog, for example, each sentence would be analysed. This could be positive, as below.

“The SSD drive is very large which is great. It allows me plenty of room to store my music with loads of space left over.”

Conversely, they may express a negative opinion in a sentence elsewhere in the blog, as such.

“The speakers are very poor quality and produce quite a ‘Tinny’ sound. I wouldn’t recommend this machine for pure music lovers.”

And, finally, a neutral point of view.

“The speed of the laptop is very impressive, but the graphics are below average.”

2.3.3 Aspect-Based

This form of sentiment analysis focuses on a specific feature of an entity.

“The Guinness at Mike the Pies is exquisite!”

The entity here being the public house, *Mike the Pies*, and the feature being the *Guinness* drink, itself. This would be particularly helpful for *Mike the Pies* review ratings on a website such as *Trip Advisor*, for example.

It is important to note that the sole focus is on the Guinness only. Other features of the pub, space, toilets, atmosphere, are irrelevant as the quote pinpoints that one aspect of the establishment (Goyal, Kechit, 2020).

2.3.4 Emotion Detection

This method aims to detect sentiment from a range of words classified under the following common emotions:

- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise
- Love
- Joy

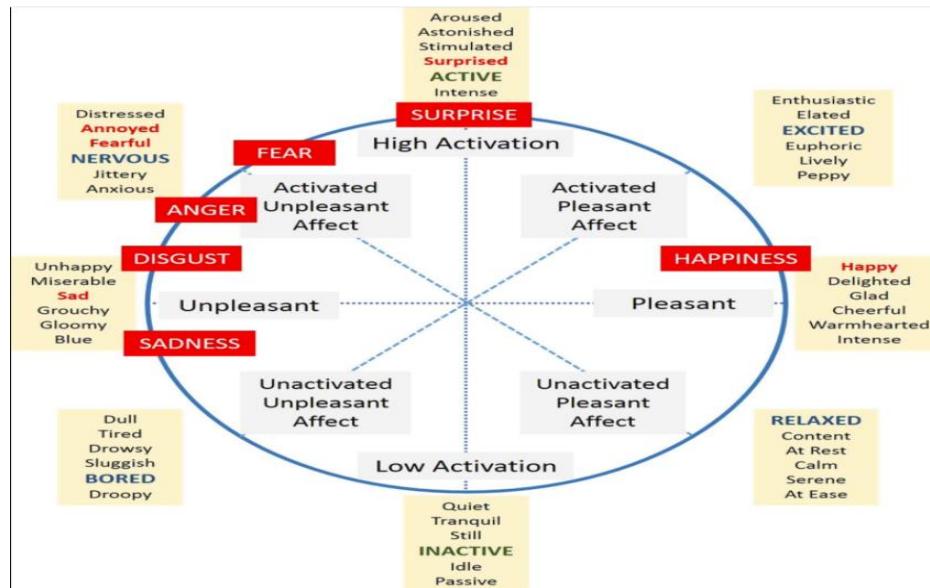


Figure 4 – Ekman’s words associated with emotion detection (Fernández-Caballero et al., 2016)

American psychologist, Paul Ekman, proposed the first six of the above as the basic forms of emotion (Fernández-Caballero et al., 2016). In 2001, Professor of Psychology at Georgetown University, W. Gerrod Parrot, elaborated on this as he put forward his theory on emotion detection (H and el, 2011).

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
Surprise	Surprise	Amazement, surprise, astonishment
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envoy	Envy, jealousy
Sadness	Torment	Torment
	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
Fear	Sympathy	Pity, sympathy
	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

Figure 5 - Parrot's emotion detection theory (H and el, 2011)

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

There are varying opinions on what method is best practice in the gathering of emotion detection. The three main algorithms used are outlined in the next section of this paper.

2.4 Sentiment Analysis Algorithms

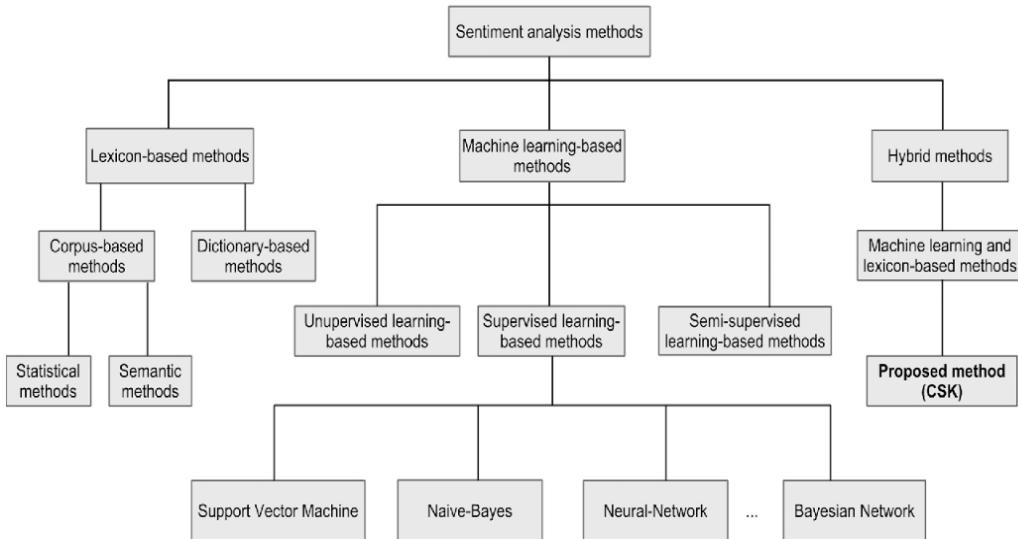


Figure 6 - Sentiment Analysis tree diagram (Pandey, Rajpoot and Saraswat, 2017)

2.5 Sentiment Analysis Challenges

2.5.1 Context

A major challenge to the accuracy of sentiment analysis is the difficulty in understanding the context of a piece of text.

“Yesterday happened to be a very *challenging* day.”

The word “Challenging” could be perceived as being either positive or negative in this sentence. The person could have thrived upon completing the challenge or found it to have been extremely tough.

“Jim is a really *laid-back* kind of guy. You’ll find this when you get to work with him.”

Is Jim “Laid-back” in a positively likeable way? Or is the person suggesting that they are lazy and will be a pain to work with? (Positive and Negative Connotations: Example Sentences, 2021)

“Well, I’ve never tasted chicken like that before!”

Is this person suggesting that the chicken tasted good or bad? (Examples of Ambiguity, 2021)

Without proper context, the text can be completely ambiguous, and this can prove to be very problematic for anybody attempting to understand the sentiment behind the text. The sentences are subjective as opposed to objective.

2.5.2 Sarcasm

Another common challenge is the usage of sarcasm or irony within text. A good example of this is the following sentence.

“My flight has been delayed. Brilliant!” (Bannister, Kristian, 2020)

If this were spoken in person, an average human could acknowledge that this would be said in a sarcastic manner. The person is clearly unhappy that their flight has been delayed. But in a machine learning environment, the computer will not necessarily understand the connotation behind the sentiment. It would probably just read the word “Brilliant” and process it as a completely positive experience for the user, therefore producing an incorrect evaluation of the feeling behind the sentence as a whole.

2.5.3 Recognition

There are many occasions when a word can have more than one meaning or relate to multiple entities.

“As a youngster, I really enjoyed squash.”

The use of the word *squash* here is ambiguous. Did the person enjoy drinking squash or playing squash?

“The Aldi in County Kerry has very impressive customer service.”

There are a number of branches of *Aldi* spread over the county. It is impossible to identify which store the customer is referring to.

2.5.4 Modern Language

Language naturally evolves over time but, to some, it would appear to seem even more rapidly in the current technological age. The internet, mobile devices, and modern slang have all brought a new challenge to the task of sentiment analysis in the form of emojis, emoticons, acronyms, abbreviations, etc.

“U woz bare funny last nite 😊. Bro u had me ROFL...fyi da mandem r meetin nxt tues 4 footy...b dare @8”.

The sentence above could easily be viewed as a modern-day text message. A machine attempting to decipher this text would face a substantial challenge.

Chapter 3: Music Lyrics

3.1 Introduction

While beats, riffs, tempo, and other aspects give the musicality to songs, it is the lyrics which tell the story of a hit. Many would argue that there have been no greater storytellers than the main songwriters behind *The Beatles*, John Lennon, and Paul McCartney. Whissell (1996) used particular data to surmise that the lyrics became less cheerful over the course of the eight year chart history of The Beatles (1962-1970) (Pettijohn and Sacco, 2009).

Words on their own can of course be powerful but placed with music they can bring one to an altogether different emotional space. For example, “Lager, lager, lager” written on paper may not automatically resonate, but added to the pounding beat in Underworld’s 1996 dance anthem, “Born Slippy”, it would undoubtedly evoke a far greater adrenaline rush and sense of euphoria to one who has fond memories of that particular era.

A look at the UK singles chart over the 21st century would appear to show that writing a good Christmas song has become a lost art. But in 1973, two of the greatest festive songs were battling for supremacy as Wizzard’s “I Wish It Could Be Christmas Everyday” lost out to “Merry Xmas Everybody” by Slade (Hayes, 2021).

“So here it is Merry Christmas (it's Christmas)
Everybody's having fun
Look to the future now
It's only just begun”.

To some, these lyrics would appear to be a simple jolly rhyme to celebrate Christmas. But songwriter, Noddy Holder, explained the deeper meaning behind the words:

“The whole economic situation in the UK was that there was miners' strikes, there were power cuts, TV was going off at ten o'clock at night, the whole industry and everything was in turmoil.

And the Christmas time was a very, very, sombre time for people. And what I wanted to do in the chorus to this was say - 'Look to the future now, it's only just begun'. It's optimism, the New Year, it'll all be better in the future!” (Singer and the Song | Learning English | BBC World Service, 2020).

The sentiment and emotion behind the lyrics is vital to this study.

3.2 Stereotyping of Music Genres

Stereotyping of music has existed, seemingly, forever. Some of the more common accusations include the following:

- Country music is performed by cowboys and only listened to by white people (Country music reckons with racial stereotypes and its future, 2020).
- Rap music is performed by criminals and only listened to by black people (How Hip-Hop Holds Blacks Back, 2015).
- Rock and Heavy Metal fans all have long hair and bad hygiene (Urban Dictionary: metalhead, 2021).
- Gay people only listen to flamboyant female artists (Madonna, Kylie, Britney, Gaga) (GayBoiBlue, 2019) or musical tunes (What's with Gay Men and Musical Theater? | by James Finn | Th-Ink Queerly | Medium, 2021).
- Only females listen to love songs (Why do girls like listening to love songs? - Quora, 2021).

An Australian study by Dr Felicity Baker in 2008 (Baker and Bor, 2008) was summarised by newspaper, *The Age* with a simple graphic:

WHAT STUDIES SAY ABOUT YOUR SOUNDS:

POP: Conformists, overly responsible, role-conscious, struggling with sexuality or peer acceptance.

HEAVY METAL: Higher levels of suicidal ideation, depression, drug use, self-harm, shoplifting, vandalism, unprotected sex.

DANCE: Higher levels of drug use regardless of socio-economic background.

JAZZ/RHYTHM & BLUES: Introverted misfits, loners.

RAP: Higher levels of theft, violence, anger, street gang membership, drug use and misogyny.

Figure 7 - Australian study on teen music listening stereotypes.

This study seemed to be dismissed as stereotypical nonsense by other writers (Stereotyping musical genres, 2008), (Diagnosing Teenager's Problems: Listen to Their Music, n.d.).

But stereotypes become stereotypes for a reason, surely? There must be a common theme and a pattern for which they are hoisted upon differing musical genres. A 2009 paper (Rentfrow, McDonald and Oldmeadow, 2009) would appear to back up some of these claims. For example, Rap music was found to be listened to by black, lower-class people, whilst Classical music is listened to by an upper-class, white demographic.

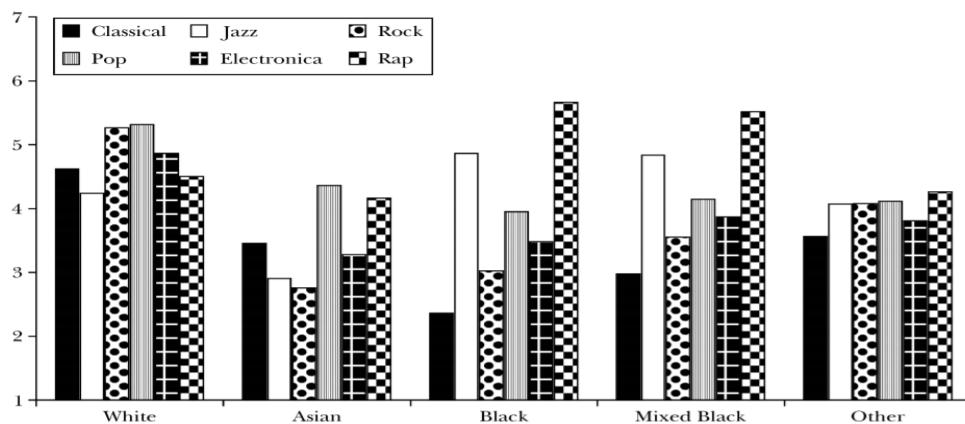


Figure 8 - UK study on musical taste by ethnicity

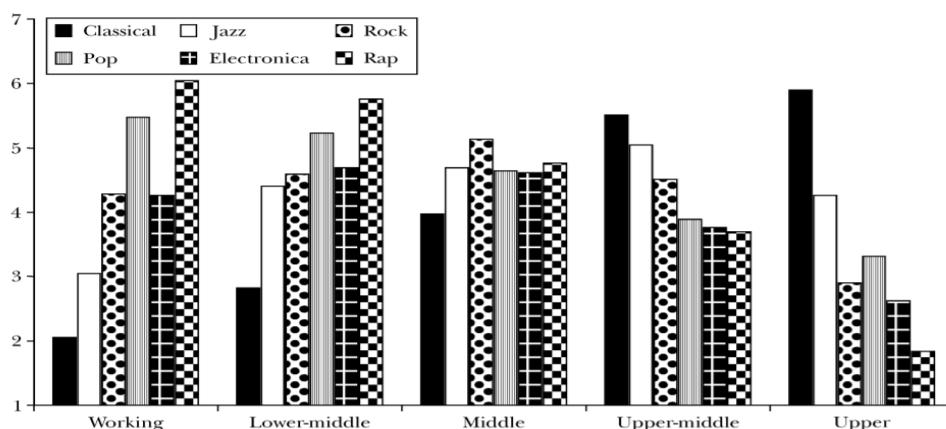


Figure 9 - UK study on musical taste by social class

In this thesis, this author will attempt to clarify just how accurate certain stereotypes are in relation to music, specifically lyrically. As expressed in the methodology chapter, a larger detail of focus will be aimed at the genres below.

3.2.1 Hip-Hop

Hip-Hop culture, and in particular its representation in the music industry in the form of Rap, has been tagged by many commentators as having a negative impact on society. This paper will seek to find out whether or not these claims are justified by looking at lyrics in the genre.

In 2014, a survey was carried out on over 1000 adults and found that Rap music had become the most popular genre of music by a significant majority of half of those polled (Chakraborty et al., 2018).

There would appear to be a general consensus that this genre of music has been stereotyped as glamorising gang culture, violence, drugs, and offensive language. “Rap has been called ‘musical graffiti’ (vandalism to some, art to others)” (Diffusion Of Innovations (4th edition), Rogers, 1995, n.d.). Another claim is that it seems to objectify women on a regular basis.

“Young women who identify with or choose to emulate the video-vixens or groupies they observe in mainstream commercial Hip Hop risk performing what Muhammad (2007) refers to as “twenty first century Venus Hottentots.”ⁱ (Bates, Talbot and Tobias, 2014)

But filmmaker, Ava Duvernay, suggests that it could be seen in a positive light as female artists involved in the genre taking control of their own narrative. Duvernay hypothesises that the women use their sex appeal cleverly by marketing themselves in a way which will sell more records, therefore bringing them further wealth (Bates, Talbot and Tobias, 2014).

3.2.2 Heavy Metal

As with Hip-Hop / Rap, Heavy Metal music has also gained a somewhat negative reputation with regards to its effect on the behaviour of youths. A number of studies conducted between 1988-1996 conclude that teens who listened to Heavy Metal were more likely to partake in unsafe sexual activities and drug use (Arnett, 1991;1992), perform poorly academically (Singer, Levine, & Jou, 1993), and contract problems psychiatrically (King 1988).

In the same paper (Fried, 2003), some reported that it is not necessarily the music that was the cause of these behaviours, but instead that the teens who shared these troubling issues simply gravitated towards the Heavy Metal genre. For example, research implied that anti-social behaviours influence musical taste, rather than the musical taste, itself, being the cause of said anti-social behaviour (Verden, Dunleavy, and Powers, 1989).

Chapter 4: Methodology

4.1 Key Research Findings

The previous three chapters in this paper outline the key research findings on which this thesis will be based upon.

In chapter two, the area of sentiment analysis was described in great detail as it is an essential process which must be performed meticulously on the data gathered in order to give the most accurate results upon which any hypotheses can be proven. As the purpose of this thesis is focussed on words, lyrics to be precise, the plan to analyse the sentiment in the songs is an obvious choice, so as to achieve the aim of answering the research questions stated further on in this chapter.

As there are numerous methods of performing sentiment analysis, more research will need to be done in an attempt to procure the most suitable technique which will best suit the needs of this paper. An agile approach will therefore be followed.

Music lyrics were covered in chapter three, and this is the key to the whole thesis as the aim is to show correlation between changes in lyrics and changes in culture and society over the past 70 years in the United Kingdom.

To emphasise the importance of lyrics to this paper, many hours, days, and weeks have been spent on researching previous studies on music lyrics and the various impacts and hypotheses which have been put forward in those studies. A wide number of sub-topics have been included to give a broad scope of issues to dissect in further detail such as race, crime, death, substance abuse, gender, misogamy, sex, and more.

The genre of Hip-Hop / Rap, in particular, has been studied extensively. The battle against misogyny, for example, is the subject matter of *Flipping the Misogynist Script: Gender, Agency, Hip Hop and Music Education* (Bates, Talbot and Tobias, 2014). The authors examine the role of females in a male dominated genre.

“Women in Hip Hop, particularly those who do not conform to expected normative roles, face marginalization by a patriarchal male-dominated system within Hip Hop culture and the music industry (Neal 2004).”

The perception of Rap / Hip-Hop as casting a widely negative portrayal of young black culture is also addressed.

“References in mainstream or commercial rap music to places such as violent streets, dance clubs, strip clubs (Miller 2004, Sharpley-Whiting 2007) or actions such as inflicting violence, sex acts, consuming alcohol, and indulging in drugs complicate the process of addressing identity and meaning making through Hip Hop in the context of K–12 schools. Furthermore, women in mainstream Hip Hop have been overwhelmingly portrayed as sexual objects in lyrics and videos.”

The valuable information provided in the above paper, as well as the many others mentioned, will hopefully help to achieve some insightful answers to the questions this study raises.

Finally, chapter four looked at the bigger cultural issues of each respective decade, and more specifically explored the impact of the music of each era upon society.

Another paper researched, *What has America been singing about? Trends in themes in the U.S. top-40 songs: 1960-2010* (Christenson et al., 2019), looks at the effects on society through various forms including;

- Alcohol/Drugs. Generic (“Let’s go get drunk”) as well as specific (“Let’s go chug some Scotch”) content related to either alcohol or drugs.
- Wealth/Status. Content related to money or the trappings of wealth as avenues to social status— e.g., reference to expensive champagne, Cuban cigars, luxury cars.
- Social/Political Issues. Content that seemed consciously to involve social, political, and/or controversial “issues.” The classic example would be the protest song.
- Race/Ethnicity. References to race and race issues, belonging to a distinct ethnic group, relations between ethnic groups.
- Religion/God. Content related to institutionalized religion, religious feelings, or a personal relation to God or another Supreme being.

The objective of this thesis, as previously stated, is to seek whether there would appear to be an obvious correlation between music lyrics and society over the decades from the 1950s to the present day. The three chapters explained above provide the key research areas in order for this objective to be met.

4.2 Research Question

Can sentiment analysis of UK number ones (1952-2020) detect correlation between lyrics and society?

4.2.1 Research sub-questions

- Has there proved to be a tangible correlation between the evolution of lyrics and cultural changes in society?
- What are the common traits or differences between love songs in each decade?
- When did misogynistic lyrics become more prevalent?
- How did the introduction of MTV influence the sexual connotations of music lyrics?
- Why are Rap & Hip-Hop hits perceived to be aggressive, and misogynistic, all whilst glamorising crime? Is this accusation justified?
- Who are the people in society mostly represented by the lyrics of each respective decade? E.g., Gender? Class? Age? Race? Etc.
- Are there cases of particular lyrics causing concern in wider society?
- Is Heavy Metal directly linked to young suicide rates?

4.3 Project Scope

The aim of this project is to measure the correlation, if any, between music lyrics from UK chart hits and changes in culture and society between 1952-2020. It will do this by scraping data from the official UK charts website and the Genius.com API before assigning positive, neutral, or negative sentiments to the lyrics of all of the songs which charted during this period. Datasets relating to cultural and societal change during the same time frame will also be analysed to ascertain if there is a connection between events and the lyrics of each respective era.

There would appear to be scope for more detail to be added to the project, time permitting. This would be chart based with ideas such as comparing certain acts to one another. For example, *The Beatles* and *One Direction*, as they were the two biggest “boy bands” of their generations. Comparing the lyrics of 1950s love songs to modern day love songs would be another example.

4.4 Proposed Methodology

The project will be attempted in conjunction with the following technologies / programs / websites:

- Python programming language
- Python libraries and packages
- Google Colab

- Jupyter Notebook
- Microsoft Excel
- Beautiful Soup
- Lyrics Genius
- Genius API
- Data.World
- Microsoft Power BI
- GitHub
- Billboard Charts
- Official UK Charts

4.4.1 Functional Specifications

As arranged in conjunction with my supervisor and other lecturers, the *MoSCoW method* will be used to facilitate the prioritisation of tasks to be worked on over the course of the project. This technique is commonly used for agile approaches where deadlines (or *timeboxes*) are set within the planning of a project. It was founded by former *Oracle* software developer, Dai Clegg in 1994 (What is the MoSCoW Method?, 2021).

4.4.1.1 MoSCoW Method

MoSCoW Method	
1	Must Have
2	Should Have
3	Could Have
4	Will not Have

Unique ID	Description	MoSCoW priority
001	Access to either <i>Google Colab</i> or <i>Jupyter Notebook</i>	1
002	Explore datasets to produce a prototype	1
003	Run <i>Python</i> script on data from <i>American Billboard Charts</i> to produce a prototype	1
004	Test interaction with <i>Lyrics Genius</i> website API	1

005	Display findings of prototype	1
006	Write <i>Python</i> script to perform a web scrape on <i>Official UK Charts</i> website	1
007	Test interaction with <i>Beautiful Soup</i> for web scraping	1
008	Save web scrape data to .csv file	1
009	Perform data pre-processing and cleaning data	1
010	Upload .csv file to <i>data.world</i>	2
011	Create <i>Python</i> script to be run on UK Charts	1
012	Import essential packages into application	1
013	Gather data from <i>Official UK Charts</i> to produce statistics to be analysed	1
014	Research academic papers, websites, and articles to gather information on UK Culture and Society	2
015	Perform Sentiment Analysis on lyrical data	1
016	Display relevant findings and correlations between UK Culture and Society, in relation to Sentiment Analysis results	2
017	Integration with <i>Power BI</i> to visualise data	3
018	Include any other data findings of interest	3
019	Explain my findings and end conclusions	1

4.4.1.2 User Stories

User Story ID	001
Unique Feature ID	001
Title	Access to either Google Colab or Jupyter Notebook
Description	As a user, I want to find a suitable environment, so that I can create and produce Python language code, in order to proceed with this project.
MoSCoW priority	1
Estimated time (hours)	1

Acceptance Criteria	Once a piece of Python code is ran successfully, the criteria will have been met.
----------------------------	---

User Story ID	002
Unique Feature ID	002
Title	Explore datasets to produce a prototype
Description	As a user, I want to explore a wide range of datasets, so that I can produce a detailed prototype.
MoSCoW priority	1
Estimated time (hours)	8
Acceptance Criteria	When a dataset has been found which closely resembles the planned end goal of the project.

User Story ID	003
Unique Feature ID	003
Title	Run Python script on data from American Billboard Charts to produce a prototype
Description	As a user, I want to run a Python script, so that the prototype produces sufficient information.
MoSCoW priority	1
Estimated time (hours)	6
Acceptance Criteria	Upon the script delivering statistics which provide a proof of concept.

User Story ID	004
----------------------	-----

Unique Feature ID	004
Title	Test interaction with Lyrics Genius website API
Description	As a user, I want to connect to the relevant API, so that the correct lyrics can be obtained.
MoSCoW priority	1
Estimated time (hours)	<1
Acceptance Criteria	Once the correct lyrics to a randomly selected song are displayed correctly in full.

User Story ID	005
Unique Feature ID	005
Title	Display findings of prototype
Description	As a user, I want to present the findings of my work, so that it is proven that the project is achievable.
MoSCoW priority	1
Estimated time (hours)	4
Acceptance Criteria	Providing the prototype put forward to my supervisor is acceptable enough to allow me to proceed.

User Story ID	006
Unique Feature ID	006
Title	Write Python script to perform a web scrape on Official UK Charts website
Description	As a user, I want to write a Python script, so that I can produce my own dataset.
MoSCoW priority	1

Estimated time (hours)	6
Acceptance Criteria	Upon producing a full dataset of the information required to move forward with the project.

User Story ID	007
Unique Feature ID	007
Title	Test interaction with Beautiful Soup for web scraping
Description	As a user, I want, so that
MoSCoW priority	1
Estimated time (hours)	<1
Acceptance Criteria	As long as no errors arise from the Python code run.

User Story ID	008
Unique Feature ID	008
Title	Save web scrape data to .csv file
Description	As a user, I want to save the dataset as a csv file, so that I can train said data in future.
MoSCoW priority	1
Estimated time (hours)	<1
Acceptance Criteria	The csv file displays fully in excel.

User Story ID	009
Unique Feature ID	009
Title	Perform data pre-processing and cleaning data

Description	As a user, I want perform pre-processing on the data, so that any potential obstacles to gaining clean data, or any irrelevant data are removed.
MoSCoW priority	1
Estimated time (hours)	4
Acceptance Criteria	When any empty fields, duplicates, incomplete records, or any other dirty data is believed to have been erased to a satisfactory level.

User Story ID	010
Unique Feature ID	010
Title	Upload .csv file to data.world
Description	As a user, I want to upload my dataset online, so that it can be attained from within the Python environment.
MoSCoW priority	2
Estimated time (hours)	2
Acceptance Criteria	The file is successfully uploaded to the website.

User Story ID	011
Unique Feature ID	011
Title	Create Python script to be run on UK Charts
Description	As a user, I want to create a Python script, so that suitable and relevant data can be obtained.
MoSCoW priority	1
Estimated time (hours)	8

Acceptance Criteria	Once the script is running with no errors and producing high quality results.
----------------------------	---

User Story ID	012
Unique Feature ID	012
Title	Import essential packages into application
Description	<p>As a user,</p> <p>I want to import the correct packages into the Python environment,</p> <p>so that all bases are covered in order to achieve the data required.</p>
MoSCoW priority	1
Estimated time (hours)	<1
Acceptance Criteria	Correct and essential data is output to the terminal.

User Story ID	013
Unique Feature ID	013
Title	Gather data from Official UK Charts dataset to produce statistics to be analysed
Description	<p>As a user,</p> <p>I want to produce only relevant data,</p> <p>so that it can be analysed within the boundaries set out for the project.</p>
MoSCoW priority	1
Estimated time (hours)	8
Acceptance Criteria	When the data and statistics produced are substantial enough to perform a thorough sentiment analysis.

User Story ID	014
Unique Feature ID	014

Title	Research academic papers, websites, and articles to gather information on UK Culture and Society
Description	As a user, I want to research a sufficient number of resources, so that I can amass detailed information on UK Culture and Society relating to the proposed research question.
MoSCoW priority	2
Estimated time (hours)	8 (ongoing)
Acceptance Criteria	Upon assembling the right amount of data relevant to the research topic.

User Story ID	015
Unique Feature ID	015
Title	Perform Sentiment Analysis on lyrical data
Description	As a user, I want to perform a sentiment analysis on lyrics, so that the first half of the project is completed.
MoSCoW priority	1
Estimated time (hours)	2
Acceptance Criteria	Once the sentiment analysis provides sufficient detail about the lyrical content provided.

User Story ID	016
Unique Feature ID	016
Title	Display relevant findings and correlations between UK Culture and Society, in relation to Sentiment Analysis results
Description	As a user, I want to display any relevant correlations between the sets of data studied,

	so that I can attempt to answer the research question to my best ability.
MoSCoW priority	2
Estimated time (hours)	8
Acceptance Criteria	When it is agreed with my supervisor that the research question has been answered to a sufficiently high level.

User Story ID	017
Unique Feature ID	017
Title	Integration with Power BI to visualise data
Description	As a user, I want to connect to the Power BI service, so that the results of the data and analysis can be displayed visually to enhance the project.
MoSCoW priority	3
Estimated time (hours)	6
Acceptance Criteria	The data displayed in Power BI is of a standard befitting the project.

User Story ID	018
Unique Feature ID	018
Title	Include any other data findings of interest
Description	As a user, I want to include any other data of interest, so that the project can be further enriched.
MoSCoW priority	3
Estimated time (hours)	4
Acceptance Criteria	Only information deemed worthy of being seen to enhance the project would be considered for inclusion.

User Story ID	019
----------------------	-----

Unique Feature ID	019
Title	Explain my findings and end conclusions
Description	As a user, I want detail my findings, so that I can sufficiently bring the project to a proper conclusion.
MoSCoW priority	1
Estimated time (hours)	4
Acceptance Criteria	The findings are relevant enough to be included and they provide a good understanding of the project undertaken.

4.5 Risk Analysis

With any project, there comes a certain amount of risk involved. The following table details any issues of concern which may arise, and how they can be addressed.

Risk	Probability	Appropriate/Corrective Action
Time constraints	High	To adhere to the sprint schedule as rigorously as possible in order to avoid a backlog at the end.
Loss of work	Medium	The project will be saved to two 4TB <i>Seagate</i> storage drives.
Inexperience with Sentiment Analysis	High	The whole of semester 1 and much of semester 2 will be spent researching the concept of Sentiment Analysis with the aim of gaining a good level of knowledge on the subject. My supervisor will also be assisting me with any concerns.
Acquiring relevant datasets	Medium	While there would appear to be many datasets available on the American singles charts, none on

		the UK charts has been located. This means web scraping of those charts will be essential.
Coding abilities	Medium	I have struggled with coding in previous projects and will need to develop a better understanding of the <i>Python</i> language. This will be attained through watching and reading tutorials online.
Hardware	Low	A <i>Dell Inspiron 7566</i> laptop with i7 processor, 4 dual cores, 16GB RAM, using <i>Microsoft Windows</i> 10 x64. Although now 5 years-old, the system should be good enough to deal with the work required.
Software	Low	Using <i>Google Colab / Jupyter Notebook</i> IDEs to run <i>Python</i> language code, storing to and uploading datasets from <i>data.world</i> online, retrieving music lyrics from <i>lyricsgenius.com</i> , parsing code with <i>Beautiful Soup</i> package, and researching relevant information from academic papers found online.
Completing project	Medium	With a hectic schedule, including other modules to work on, I hope to answer the research question to the best of my abilities.

4.6 Prototype

A tutorial on the steps taken through a previous example by Salim Zubair on the American Billboard charts (salimzubair/lyric-sentiment: Sentiment analysis of all songs that entered the Billboard Hot 100 charts from 1958-2019, n.d.).

Prototype	Start Date	Finish Date
1	23/11/2020	17/12/2020

Task ID	Description	Status
1	Researched example projects	Complete
2	Saved necessary files from GitHub and data.world	Complete
3	Created notebook in Google Colab	Complete
4	Import required modules and packages for Python code language: <ul style="list-style-type: none"> • lyricsgenius • textblob • spacy • nltk • pandas 	Complete
5	First few runs of code	Incomplete
6	Modified code snippets to eradicate errors and improve process, such as: <ul style="list-style-type: none"> • Genius API key • Smaller subset of data run • Indentation issues • Erroneous comments 	Incomplete (Completed enough for most of prototype to be sufficiently demonstrated)
7	Created notebook in Jupyter notebook to test full lyricsgenius database	Incomplete (Full dataset crashes so smaller subset will be demonstrated)

8	Modified further code in order to extract ‘keywords’ data columns	Complete (More accuracy desired once full data set is run)
----------	---	--

Once research on example projects had been completed, Zubair’s *GitHub* repository was pulled down to the host machine and backed up to an external drive. The required dataset was also downloaded and backed up from *data.world.com*.

Google Colab was chosen as the program in which to implement the python code and the necessary packages were installed at the beginning.

```

:
1  from textblob import TextBlob
2  #install modules
3  !pip install lyricsgenius
4  !pip install -U gensim
5  !pip install -U spacy
6  !pip install textblob
7  !python -m textblob.download_corpora
8  !pip install rake-nltk
9
10 #import packages
11 from gensim.summarization import keywords
12 import spacy
13 import nltk
14 from nltk.corpus import stopwords
15 nltk.download('stopwords')
16 stopwords = set(stopwords.words('english'))
17 from spacy.lang.en import English
18 nlp = English()
19 nlp.max_length = 10000000
20 import lyricsgenius
21 import pandas as pd
22 from textblob import TextBlob

```

Figure 10 - Installing libraries and packages.

It was during this initial period that issues were being raised with regards to the suitability of *Google Colab*. Due to the large data of songs, the code was having to be run overnight, and this led to numerous instances of the program disconnecting, causing the prototype to eventually crash. It was then decided to switch to *Jupyter notebook* to see if the code could be run successfully through that program. Similar problems occurred and so it was concluded that a smaller subset of the data set would be ran to provide a functioning prototype.

The next step was connecting to the *Genius.com* API in order to collect the lyrics to songs in the dataset. A simple piece of test code was written to ascertain that the connection was

successful. The example shown below was to display some Taylor Swift lyrics, but any song could be hard coded in to the cell to be tested.

```

1 #define Genius API authentication
2 genius = lyricsgenius.Genius('_g20PhiAv3u_FpHXcs_vPjwUflainNCXpbfiYgaRSwpLt4FiDFUdpesNSOIq8ppA')
3
4 lyric_test = genius.search_song('Shake It Off', 'Taylor Swift').lyrics
5 print(lyric_test)

```

```

[Post-Chorus]
Shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off
Shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off (Yeah)
Shake it off, I shake it off
I, I, I shake it off, I shake it off (You got to)
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off

```

[Video directed by Mark Romanek]

Figure 11 - Testing connection between lyricsgenius and Genius API (a), and Taylor Swift's "Shake It Off" proving successful connection between lyricsgenius and Genius API (b).

Once the test had been proven to be successful, the aforementioned smaller sample of 1000 songs was imported.

```

1 #import billboard hot100 dataset
2 hot100_df = pd.read_csv('https://query.data.world/s/qf6et5c7dh23kglnvjcoztlmom62it')
3 hot100_df.drop_duplicates(subset='SongID', inplace = True) #remove duplicate occurrences of songs
4 hot100_df = hot100_df.head(1000)
5 hot100_df.reset_index()

```

		index	url	WeekID	Week Position	Song	Performer	SongID	Instance	Previous Week Position	Peak Position	Weeks on Chart
0	0	0	http://www.billboard.com/charts/hot-100/1958-0...	8/2/1958	1	Poor Little Fool	Ricky Nelson	Poor Little Fool	Ricky Nelson	1	NaN	1
1	1	1	http://www.billboard.com/charts/hot-100/1995-1...	12/2/1995	1	One Sweet Day	Mariah Carey & Boyz II Men	One Sweet Day	Mariah Carey & Boyz II Men	1	NaN	1
2	2	2	http://www.billboard.com/charts/hot-100/1997-1...	10/11/1997	1	Candle In The Wind 1997/Something About The Wa...	Elton John	Candle In The Wind 1997/Something About The Wa...		1	NaN	1
3	3	3	http://www.billboard.com/charts/hot-100/2006-0...	7/1/2006	1	Do I Make You Proud	Taylor Hicks	Do I Make You Proud	Taylor Hicks	1	NaN	1
4	4	4	http://www.billboard.com/charts/hot-100/2009-1...	10/24/2009	1		3	3	Britney Spears	1	NaN	1

Figure 12 - Importing csv file from data.world (a), then displayed in tabular form in notebook (b).

Zubair seemingly had some errors in his code, and so some had to be modified in order for it to be implemented successfully. An example, shown below, is the code needed to extract ‘keywords’ from song lyrics.

```

1 #Use return_keywords function on Lyrics_resample to get the top 20 keywords for each year
2 lyric_keywords = [return_keywords(x[1]) for x in lyrics_resample.iteritems()]
3
4 #lyric_keywords = ""
5 #for x in lyrics_resample:
6 #    #print(x)
7 #    if x != "0":
8 #        #print(x)
9 #        lyric_keywords = lyric_keywords + str(x)
10 #lyrics_resample
11
12 lyric_keywords

```

[['love', 'like', 'lovely things', 'hearts', 'heart', 'humpty', 'biz', 'kiss', 'kissed', 'got'], ['great', 'near', 'come', 'comes', 'kings', 'loves', 'place', 'people', 'blood'], ...]	['knows', 'yeah'], ['tell', 'let', 'night like', 'baby', 'chorus', 'dance', 'dancing', 'yeah', 'damn time', 'know'], ['come', 'comes', 'like', 'baby', 'liked loved', 'feeling', 'chorus talk', ...]	'likes', 'yeah', 'shits', 'got', 'nigga', 'niggas', 'gon hit like', 'chorus', 'know'], ['yeah', 'likes', 'shit like', 'chorus', 'verse', 'know', 'knows', 'money got', 'ooh', 'shits']]
--	---	---

Figure 13 - Calling the return_keywords function (a) to display an array of the top keywords for each year (b), (c), (d).

There were some issues which needed to be fixed throughout the process, but the prototype appears to be sufficiently working enough to display the intended point of use. It is shown in full detail in Appendix A - Prototype screenshots.

Chapter 5: Implementation

5.1 Introduction

Although the prototype worked on a smaller subset of the data than desired, it still sufficiently proves that the process can be successful if followed to a similar pattern. More understanding of the code behind the prototype is needed and that will be the aim of the first sprint. The remainder of the sprint schedule is laid out below and an agile approach will be used to ensure that any impediments can be bypassed temporarily as to not disrupt the flow of the project.

5.2 Sprint Schedule

Sprint ID	Description	Start Date	End Date
1	Understanding Prototype and Python code	25/01/2021	07/02/2021
2	Web Scraping and cleaning of UK chart data	08/02/2021	28/02/2021
3	Running UK chart data and performing sentiment analysis	01/03/2021	21/03/2021
4	Researching UK cultural and societal data	22/03/2021	11/04/2021
5	Answering the thesis question; Can sentiment analysis of UK number ones (1952-2020) detect correlation between lyrics and society?	12/04/2021	02/05/2021
6	Bonus data, Presentation, and final submission	03/05/2021	13/05/2021

5.3 Sprint 1 – Understanding prototype and Python code

Sprint ID	Description	Start Date	Finish Date
1	Understanding Prototype and Python	25/01/2021	07/02/2021

Task ID	Description	Status
1	Libraries	Completed
2	API connection	Completed
3	Functions	Completed

Task 1

Beginning with looking at the official Python documentation (Installing Packages — Python Packaging User Guide, 2021) I needed to ensure that I had the correct version of the language installed (version 3). Salim Zubair had used pip which is the most popular tool for installing Python libraries and packages. He first used pip to install TextBlob, a library which specialises in textual data. This will help to obtain the level of polarity and subjectivity emanating from the loaded dataset (Sentiment Analysis in Python With TextBlob, 2021). Gensim and Rake-nltk are two powerful libraries for extracting keywords from text, to aid with semantic modelling (Gensim: topic modelling for humans, 2021); (Sharma, n.d.). Finally, spaCy, a library which excels in Natural Language Processing (NLP). It uses a wide number of features, including those listed below, which are beneficial to the field of linguistics (spaCy 101: Everything you need to know · spaCy Usage Documentation, 2021).

- Text Classification
- Entity Linking (EL)
- Lemmatization
- Part-of-speech (POS) Tagging
- Tokenization

One of the common problems when working with text in sentiment analysis is collecting irrelevant data. This is where the Natural Language Toolkit (nltk) proves to be conducive. It eliminates what are known as ‘Stop Words’; commonly used words in our everyday

language, but which are not necessary in the conducting of a sentiment analysis (Removing stop words with NLTK in Python, 2017).

```
import nltk
from nltk.corpus import stopwords
print(stopwords.words('english'))
```

{'ourselves', 'hers', 'between', 'yourself', 'but', 'again', 'there', 'about', 'once', 'during', 'out', 'very', 'having', 'with', 'they', 'own', 'an', 'be', 'some', 'for', 'do', 'its', 'yours', 'such', 'into', 'of', 'most', 'itself', 'other', 'off', 'is', 's', 'am', 'or', 'who', 'as', 'from', 'him', 'each', 'the', 'themselves', 'until', 'below', 'are', 'we', 'these', 'your', 'his', 'through', 'don', 'nor', 'me', 'were', 'her', 'more', 'himself', 'this', 'down', 'should', 'our', 'their', 'while', 'above', 'both', 'up', 'to', 'ours', 'had', 'she', 'all', 'no', 'when', 'at', 'any', 'before', 'them', 'same', 'and', 'been', 'have', 'in', 'will', 'on', 'does', 'yourselves', 'then', 'that', 'because', 'what', 'over', 'why', 'so', 'can', 'did', 'not', 'now', 'under', 'he', 'you', 'herself', 'has', 'just', 'where', 'too', 'only', 'myself', 'which', 'those', 'i', 'after', 'few', 'whom', 't', 'being', 'if', 'theirs', 'my', 'against', 'a', 'by', 'doing', 'it', 'how', 'further', 'was', 'here', 'than'}

Figure 14 - Example of nltk 'Stop Words' in the English language

To give an example of this, let us look at a sentence.

“The goalkeeper is an embarrassment to his team. The way he pushed the ball in off of the post was comical.”

After importing the nltk package, ‘stop words’ will be removed from this sentence leaving only the relevant data remaining, as below.

“goalkeeper”, “embarrassment”, “team”, “way”, “pushed”, “ball”, “post”, “comical”.

This will save valuable time and system resource from being wasted and lead to a far more accurate picture of the sentiment throughout the data.

Another Python library, Pandas, will be used to read files into a dataframe which is a 2D table object. Pandas is a popular tool used within the data analysis realm (Pandas vs NumPy - javatpoint, 2021).

Task 2

Genius.com is a music website which acts as an encyclopaedia, of sorts, for lyrics. It is very similar to Wikipedia in that it relies on the general audience (they’re called scholars on the site) to input the information about the lyrics by way of annotations (Genius – How Genius Works, 2021). Fortunately, it provides an application programming interface (API) which will be used in this project to retrieve the lyrics correlating to the songs located in the UK chart data. To do this I will need a way to connect the Genius API to my Python environment, and that is where lyricsgenius will be incorporated.

A Python client, lyricsgenius calls upon the web-scraping tool, BeautifulSoup, to obtain the lyrics of the songs on genius.com which have been requested in the Python environment (How It Works — lyricsgenius documentation, 2021). I set up an account on the Genius API and gained an access key which entailed me to link the API with the lyricsgenius package.

API Clients

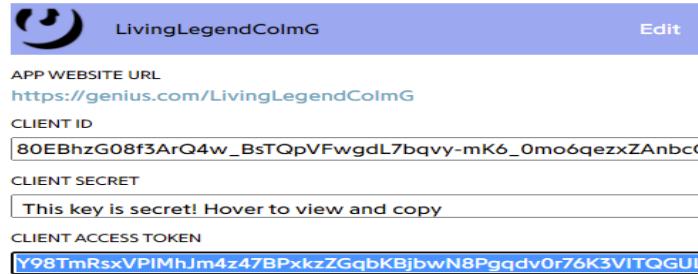


Figure 15 - Obtaining the client access token from genius.com API.

The final step here was to choose a random song and determine if the connection between the two tools would work successfully. I decided to choose a song that was not necessarily a ‘mega-hit’ as I was curious to see just how comprehensive the genius.com database was. Adele’s “Hometown Glory” happened to be one her more modest ranking hits on the UK chart, reaching number 19, and this seemed to be a suitable selection to test.

```

1 #define Genius API authentication
2 genius = lyricsgenius.Genius('Y98TmRsxVPIMhJm4z47BPxkzZGqbKBjbwN8PgqvOr76K3VITQGULjm8pnTSOHIB')
3 lyric_test = genius.search_song('Hometown Glory', 'Adele').lyrics
4 print(lyric_test)

Searching for "Hometown Glory" by Adele...
Done.
[Verse 1]
I've been walking in the same way as I did
Missing out the cracks in the pavement
And tutting my heel and strutting my feet
"Is there anything I can do for you dear?"
"Is there anyone I could call?"
"No and thank you, please madame
I ain't lost, just wandering"

[Chorus]
'Round my hometown
Memories are fresh
'Round my hometown
Ooh, the people I've met
Are the wonders of my world
Are the wonders of my world
Are the wonders of this world
Are the wonders now

```

Figure 16 - Testing Genius API with Adele's "Hometown Glory".

The code ran and found the song very quickly confirming that the connection between lyricsgenius and genius.com was fully operational.

Task 3

Once Zubair had used the pandas library to create a dataframe containing the Billboard 100 dataset, he moved onto creating some functions to act as a foundation for the remainder of the analysis.

- Def get_lyrics(title, artist)
- Def get_lyric_sentiment(lyrics)
- Def preprocess(text)
- Def return_keywords(texts)

‘Def’ is a Python keyword which is shorthand for ‘Define’. In the first instance here (Figure 17), the ‘get_lyrics’ function is being set up to fetch the lyrics to the songs in the dataframe. It achieves this by linking up to the Genius API to search its database and retrieving any songs and artists which match.

```

1 #Function to return lyrics of each song using Genius API
2 def get_lyrics(title, artist):
3     try:
4         return genius.search_song(title, artist).lyrics
5     except:
6         return 'not found'

```

Figure 17 - Defining 'get_lyrics' function.

If a match is not found, it will return a ‘not found’ message before a further line of code drops those instances from the dataframe (Figure 18). As shown in the example below, the lyrics to Eddie Fisher and Sally Sweetland’s “I’m Walking Behind You” is the first of these instances and this row will be dropped from the newly created ‘Lyrics’ column in the dataframe moving forwards, to ensure accuracy with the keywords search at the end of the notebook.

```

1 #Use get_lyrics funcion to get lyrics for every song in dataset
2 lyrics = UK1s_df.apply(lambda row: get_lyrics(row['Song title'], row['Artist(s')]), axis =1)
3 UK1s_df['Lyrics'] = lyrics
4 UK1s_df = UK1s_df.drop(UK1s_df[UK1s_df['Lyrics'] == 'not found'].index) #drop rows where lyrics are not found on Genius
5 UK1s_df

```

Searching for "HERE IN MY HEART" by AL MARTINO...
Done.
Searching for "OUTSIDE OF HEAVEN" by EDDIE FISHER...
Done.
Searching for "I'M WALKING BEHIND YOU" by EDDIE FISHER WITH SALLY SWEETLAND...
No results found for: 'I'M WALKING BEHIND YOU EDDIE FISHER WITH SALLY SWEETLAND'
Searching for "ANSWER ME" by FRANKIE LAINE...
Done.

Figure 18 - code implemented to drop any songs where lyrics cannot be found.

The next function defined is ‘get_lyric_sentiment’. This will work in conjunction with the TextBlob library, mentioned earlier on in this sprint, to apply a sentiment polarity rating on each set of lyrics in the dataframe.

```
8  def get_lyric_sentiment(lyrics):
9      analysis = TextBlob(lyrics)
10     return analysis.sentiment.polarity
```

Figure 19 - Defining ‘get_lyric_sentiment’ function.

Another library mentioned earlier, spaCy, is engaged in the function of pre-processing the text. A doc is created with the purpose of *enabling* NLP but the *disabling* of ‘ner’ and ‘parser’. This means that the text will not be parsed or contain any named entities (Language Processing Pipelines · spaCy Usage Documentation, 2021).

The generating of ‘Lemmas’ is to convert words, or lyrics in this study, to their base form and to contextualise them correctly (Prabhakaran, 2018). Context is vital to the sentiment analysis process as we need to understand the meaning of the lyrics in the dataframe. For example, one of the songs in the UK dataset is 1988 number 1, “Nothing’s Gonna Change My Love For You” by Glenn Medeiros. The word ‘change’ could denote money, as an often-used term for coins. But we as humans will understand that the word, in relation to the song, is synonymous with such words as ‘alter’, ‘modify’, or ‘adjust’. Computers will not necessarily have the attributes to detect such context, which is why lemmatisation is so key to the piece of code seen in Figure 20. It has the ability to read the words before and after ‘change’ and determine the context in which the word should be taken. Prabhakaran (2018) also provides a good example as to why lemmatisation could be more advantageous than stemming in this situation and there are other examples to support this notion (Appendix B – Lemmatisation vs Stemming). An ‘if’ statement then removes any ‘stop words’ and non-alphabetic characters. If the condition has been met, the lemma tokens will be returned upon the function being called.

```
12 # Function to preprocess text
13 def preprocess(text):
14     # Create Doc object
15     doc = nlp(text, disable=['ner', 'parser'])
16     # Generate Lemmas
17     lemmas = [token.lemma_ for token in doc]
18     # Remove stopwords and non-alphabetic characters
19     a_lemmas = [lemma for lemma in lemmas
20                 if lemma.isalpha() and lemma not in stopwords]
21
22     return ' '.join(a_lemmas)
```

Figure 20 – Defining ‘preprocess’ function.

The final function defined is ‘return _keywords’; possibly the most important piece of code in the notebook as it is those ‘keywords’ that will significantly impact the answer to the research question of this paper. Firstly, a variable is declared, ‘xkeywords’, which will store an array of keywords. The pre-processed text will be scored, and a new line created for each word in the array. This is stored in the variable, ‘values’. Then a ‘for’ loop is run on ‘values’ using a ‘slicer’ to set the array to take the first 10 key values and appending any newer values to index zero of the list (Schoell, 2017). If the argument is satisfied it will return the keywords, otherwise a “no content” display.

```

24  """Extract Keywords from text"""
25  def return_keywords(texts):
26      xkeywords = []
27      values = keywords(text=preprocess(texts), split='\n', scores=True)
28      for x in values[:10]:
29          xkeywords.append(x[0])
30      try:
31          return xkeywords
32      except:
33          return "no content"

```

Figure 21 - Defining 'return_keywords' function.

The first two of the aforementioned functions, ‘get_lyrics’ and ‘get_lyric_sentiment’ are called next in the notebook followed by a cell which uses some code to set the index of the dataframe to be sorted by the ‘WeekID’ column. This is then ‘resampled’ so that the sentiment data can be categorised by the year, which will help to separate the values by decade for the purposes of this project.

5.4 Sprint 2 – Web scraping and cleaning of UK chart data

Sprint ID	Description	Start Date	Finish Date
2	Web Scraping and cleaning of UK chart data	08/02/2021	28/02/2021

Task ID	Description	Status
1	Official UK charts data search	Completed
2	Web scraping with Beautiful Soup	Completed
3	Scraping chart data and saving as csv file	Completed
4	Cleaning data	Completed

Task 1

Having been a long-time member of many music chart forums, I was confident that I would be able to find some sufficient data on the UK singles chart. These included:

- Haven forum (Chart Archive | Haven, 2021)
- Polyhex (polyhex | UK singles chartruns, 2021)
- Buzzjack Music Forum (Forums, 2021)
- UKMIX Forums (Chart Histories, 2021)
- everyHit.com (everyHit.com - UK Top 40 Chart Archive, British Singles & Album Charts, 2021)

But, I knew that the only real guarantee of gathering the data I sought, would be from the official UK charts website (Official Charts - Home of the Official UK Top 40 Charts, 2021).

The UK Singles Chart was introduced on November 14th 1952 when the inaugural top 12 was published in famed music magazine, the New Musical Express (NME). It was the brainchild of publisher, Percy Dickens, who originally compiled the chart by calling approximately twenty record shops and requesting their sales data (Williams, 2002). Today, more than 6,500 sources are aggregated through physical form, an electronic point of sale (EPOS) system which scans the physical data, or by calculating streaming data from services such as Spotify and iTunes. This data is collected by market research company, Kantar, on behalf of

the Official Charts Company which was set up in 1998 to manage this UK institution (Who We Are - Our charts & data, 2021).

The singles chart has acquired a cult following over the past seven decades forming the inspiration for the renowned television show, Top of the Pops (1964-2006), and being announced on BBC Radio 1 from October 1967 to the present day (Who We Are - History of the Official Charts - the Sixties, 2020).

Task 2

As opposed to their American alternative, acquiring the UK chart dataset was proving to be incredibly difficult to find. This meant that I would need to create my own dataset, and with no experience whatsoever in undertaking such a task, I set upon researching what it was that I would have to do to address this issue.

My supervisor, Andrew Shields, had mentioned performing a web scrape and so I focussed on exploring what exactly this would entail. I found a few examples online of the UK charts being scraped but some appeared to be too complex (QuickGraph #4: UK Official Singles Chart 2019 · Mark Needham, 2021); some not specific enough to my aim (Charting the Rise of Song Collaborations with Scrapy and Pandas, 2017); and others produced a smaller subset of data than I had desired (Scrap UK top 40 singles chart from BBC - DZone, 2021).

Eventually, I discovered an almost perfect example to help me achieve my goal. Cain Osborne's *Analysing UK Chart history – 1956 to 2017* focusses on the UK Album Charts rather than the singles chart which I would be covering, and the data he was looking for was slightly different to that which I required, but nonetheless, his code appeared to be something I could sample with some modifications of my own (Osborne, 2017).

I noticed that as well as using pandas, which I had now become accustomed to in Python, he mentioned Beautiful Soup in his code.

```
soup = bs4.BeautifulSoup(req.text, "lxml")
```

Figure 22 - Cain Osborne's code snippet referring to Beautiful Soup.

Beautiful Soup was something that I had, fleetingly, heard mentioned previously, but I wanted to ascertain how it would aid me with my own objective. The text data which I required from the official charts website was embedded in tables within a HyperText Markup Language (HTML) structure. Beautiful Soup is a library in Python which sits on top of other

parsers and enables the scraping and extracting of data from HTML and Extensible Markup Language (XML) (What is Beautiful Soup?, 2021). This seemed like a perfect match for my project.

Task 3

I copied Osborne's code (Osborne, 2017) into Google Colab to use as a foundation because I knew it would need only minor tweaking by switching out 'albums' for 'singles'.

```

import requests, bs4, pandas as pd
import csv

def getalbums(url):
    allalbums = []
    print('Getting Page %s ' %url)
    req = requests.get(url)
    req.raise_for_status()

    #Exit loop if status code is not 200
    if req.status_code != 200:
        return None

    soup = bs4.BeautifulSoup(req.text, "lxml")

    #Retrieve chart dates and tidy the format
    sdate = soup.find_all("p", class_="article-date")
    date = sdate[0].text.split('-')[0]

    #retrieve album position, artist and album name
    positions = soup.find_all("span", class_="position")
    albums = soup.find_all("div", class_="title")
    artists = soup.find_all("div", class_="artist")

    #create a list of each album, tidying the format
    for i in range(0,len(positions)):
        album = []
        album.append(date.strip('\n').strip('\n').strip(' '))
        album.append(positions[i].text)
        album.append(artists[i].text.strip('\n').strip('r'))
        album.append(albums[i].text.strip('\n').strip('r'))

    #append each album list to the weeks list
    allalbums.append(album)

    #find previous weeks information and create link, exit loop if link
    #can't be found
    prelink = soup.find("a",text="prev")
    if prelink == None:
        return None
    link = (prelink['href'])
    link = 'http://www.officialcharts.com/' + link

    #write weekly albums to CSV, appending to existing file
    with open("output.csv",'a',newline='') as resultFile:
        wr = csv.writer(resultFile)
        wr.writerow(allalbums)
        resultFile.close()

    #clear out the weekly list and proceed to next weeks file
    allalbums = []
    getalbums(link)

    #Enter start page to start the loop
    getalbums('http://www.officialcharts.com/charts/albums-
chart/19610702/7502/')

1 import requests, bs4, pandas as pd
2 import csv
3
4 def getsingles(url):
5     allsingles = []
6     print('Getting Page %s ' %url)
7     req = requests.get(url)
8     req.raise_for_status()
9
10    #Exit loop if status code is not 200
11    if req.status_code != 200:
12        | return None
13
14    soup = bs4.BeautifulSoup(req.text,"lxml")
15
16    #Retrieve chart dates and tidy the format
17    sdate = soup.find_all("p", class_="article-date")
18    date = sdate[0].text.split('-')[0]
19
20
21    #retrieve single position, artist and single name
22    positions = soup.find_all("span", class_="position")
23    singles = soup.find_all("div", class_="title")
24    artists = soup.find_all("div", class_="artist")
25    label = soup.find_all("span", class_="label")
26
27
28    #create a list of each single, tidying the format
29    for i in range(0,len(positions)):
30        single = []
31        single.append(date.strip('\n').strip('\n').strip(' '))
32        single.append(positions[i].text)
33        single.append(artists[i].text.strip('\n').strip('r'))
34        single.append(singles[i].text.strip('\n').strip('r'))
35        single.append(label[i].text.strip('\n').strip('r'))
36
37    #append each single list to the weeks list
38    allsingles.append(single)
39
40
41    #find previous weeks information and create link, exit loop if link can't be found
42    prelink = soup.find("a",text="prev")
43    if prelink == None:
44        | return None
45    link = (prelink['href'])
46    link = 'http://www.officialcharts.com/' + link
47
48    #write weekly singles to CSV, appending to existing file
49    with open("output.csv",'a',newline='') as resultFile:
50        wr = csv.writer(resultFile)
51        wr.writerow(allsingles)
52        resultFile.close()
53
54    #clear out the weekly list and proceed to next weeks file
55    allsingles = []
56    getsingles(link)
57
58
59    #Enter start page to start the loop
60    getsingles('https://www.officialcharts.com/charts/singles-chart/19521114/7501/')

Getting Page https://www.officialcharts.com/charts/singles-chart/19521114/7501/

```

Figure 23 - Osborne's code for scraping UK albums chart (a), (c) & my modifications for scraping UK singles chart (b), (d).

The code ran comprehensively from the last published chart in 2019, all the way back to the very first chart in 1952. One small bug was that the code would intermittently stop running at points meaning that I would have to re-run it from whichever place it had stalled at. This was

a minor issue in the grand scheme of things as I was able to observe it happening whilst I worked on other tasks.

Task 4

Due to the issue mentioned, it resulted in having to save each run of the code as a separate csv file; there were fourteen in all. This necessitated having to amalgamate the files into one completed dataset to use subsequently.

The only other cleaning of the data conducted at this point was to implement suitable headers to each column.

5.5 Sprint 3 – Running UK data and performing sentiment analysis

Sprint ID	Description	Start Date	Finish Date
3	Running UK data and performing sentiment analysis	01/03/2021	21/03/2021

Task ID	Description	Status
1	Running UK chart dataset	Completed
2	Choosing a smaller UK chart dataset	Completed
3	Scraping and cleaning UK chart dataset	Completed
4	Re-running UK chart dataset	Completed
5	Performing Sentiment Analysis	Completed

Task 1

As mentioned earlier in this paper, there had been problems with running the full prototype dataset as it was extremely large. On many occasions the data was left to run in Google Colab overnight, only for me to find it had crashed at various points. I believe this is down to a RAM issue, but it could just as conceivably be my inexperience with working with these programs. This left me concerned that I may run into the same predicament when it was time to run my newly created UK dataset, and unfortunately, those fears were confirmed when the crash error repeated itself. The scraped dataset happened to contain 263,983 rows which was proving too much for the notebook to run smoothly. Rather than waste any more time attempting to resolve the issue, I decided, as I had done with the prototype, to proceed with a different method.

Task 2

As I had finally been able to get the prototype working with a smaller amount of data, approximately 1000 songs, I determined that a similar sized subset would be more practical for my project to proceed with. As I have a very keen interest in the charts as a hobbyist, it was not too difficult to devise a new strategy to an area which I would consider myself to be knowledgeable. I settled on choosing to create a dataset comprising all of the UK number ones between 1952-2020. This was important for three key reasons:

- Ensuring that I would be able to persist with keeping the broad scope of the project.
- Matching the smaller subset requirement as there were 1360 distinctive number one songs in the chosen period.
- Number one songs typically sell more than other hits (depending on the era). This means that the lyrics should give a fair reflection of society during that specific era.

Task 3

In an effort to demonstrate the adaptability which I have acquired over the duration of my degree course, I elected to perform the next scraping and cleaning of data in a different program. Microsoft Power BI is best known as a data visualisation tool enabling the creation and presenting of dashboard based reports (mihart, 2021b). But it is also an excellent resource in relation to the process of web scraping and transforming data.

Power BI has a simple method for collecting data from websites which is both quick and simple to process (Figure 24). It is just a matter of clicking on ‘Get data’, selecting ‘web’, and entering the URL of the page which stores the data. I wanted to collect every number one UK single from officialcharts.com (All The Official Singles Chart Number 1s, 2021).

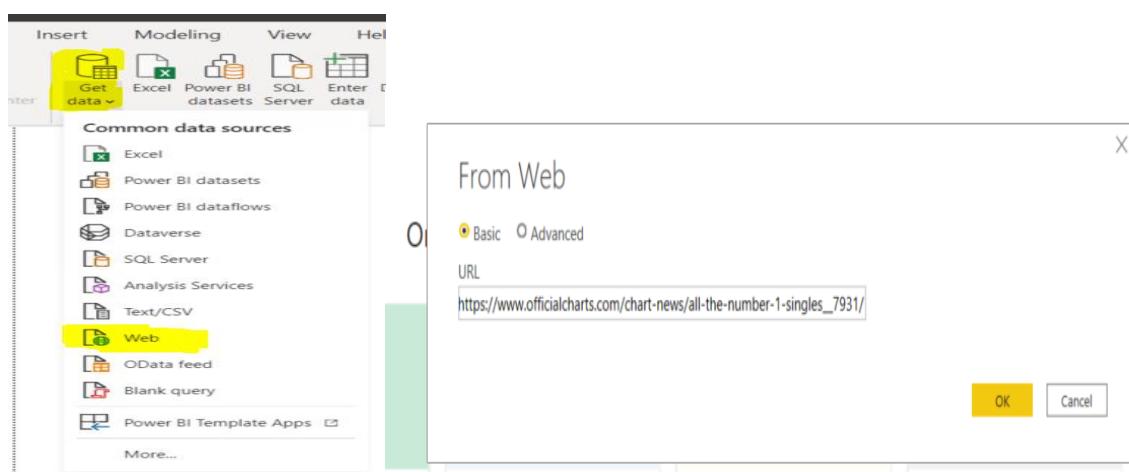


Figure 24 - Power BI's simple method to retrieve data (a) from a website (b).

The issue with collecting the data from the official charts website was that it would be bringing the data in by the year. This, I believed, would be quite a slow process and I discovered that by using Wikipedia instead, I would be able to pull the data over by the decade which was far more logical to the work flow. Although I would usually advise against using Wikipedia as a reference tool due to its reputation for often supplying false

information, I know that in this particular case, Wikipedia took its material from the official charts site anyway, so in my view, this was verification for my own method.

The next step was to determine which parts of the data needed cleaning, yet another thing which Power BI performs with relative ease. The full process is detailed in Appendix C – Web scraping and cleaning data in Power BI but I noticed five main areas which needed attention.

- I would need to give meaningful titles to the headings.
- Remove the rows with year values.
- Eliminate any duplicate values which would impede the index.
- Extract any symbols appended to song titles.
- Amend floating-point values in integer-based fields.

The screenshot shows the Microsoft Power BI desktop interface. In the center is a table titled 'Table 3' with six columns. The columns are labeled 'Column1' through 'Column6'. The data in the table is as follows:

No.	Artist	Song	Record Label	Week ending date	Weeks at number one
1		Single			
2	Al Martino	"Here In My Heart"	Capitol	14 November 1952	9
3		"You Belong to Me"	Columbia	16 January 1953	1
4		"Comes A-Long A-Love"	Capitol	23 January 1953	1
5	Eddie Fisher	"Outside of Heaven"	HMV	30 January 1953	1
6	Perry Como	"Don't Let the Stars Get in Your Eyes"	HMV	6 February 1953	5
7	Guy Mitchell	"She Wears Red Feathers"	Columbia	13 March 1953	4
8		"Broken Wings"	Decca	10 April 1953	1
9	The Stargazers	"How Much Is That Doggie in the Window?"	Decca	17 April 1953	1
10		"I Believe"	Philips	24 April 1953	9
11	Lita Roza	"Look At That Girl!"	Philips	11 September 1953	6
12	Frankie Laine	"Hey Joe"	Philips	23 October 1953	2
13	Eddie Fisher	"I'm Walking Behind You"	HMV	26 June 1953	1
14	Frankie Laine	"I Believe"	Philips	3 July 1953	6
15	Mantovani	"The Song from Moulin Rouge"	Decca	14 August 1953	1
16	Frankie Laine	"I Believe"	Philips	21 August 1953	3
17	Guy Mitchell	"Look At That Girl!"	Philips		
18	Frankie Laine	"Hey Joe"	Philips		
19	David Whitfield	"Answer Me"	Decca		
20	Frankie Laine	"Answer Me"	Philips		
21	David Whitfield	"Answer Me"	Decca		
22	David Whitfield	"Answer Me"	Decca	11 December 1953	7.5
23	Eddie Calvert	"Oh Mein Papa"	Columbia	8 January 1954	9
24	The Stargazers	"I See the Moon"	Decca	12 March 1954	5
25	Doris Day	"Secret Love"	Philips	16 April 1954	1
26	The Stargazers	"I See the Moon"	Decca	23 April 1954	1
27	Johnnie Ray	"Such a Night"	Philips	30 April 1954	1
28	Doris Day	"Secret Love"	Philips	7 May 1954	8
29	David Whitfield	"Cara Mia"	Decca	2 July 1954	9.10
30	Kitty Kallen	"Little Things Mean a Lot"	Brunswick	10 September 1954	1

Figure 25 - Wikipedia data brought into Power BI before cleaning.

Once the data from each decade had been transformed to a satisfactory state, each dataset was saved as a distinct csv file (mihart, 2021a). It was then a case of copying and pasting each decade into a larger Excel sheet containing all 1360 number ones which occurred between 1952-2020. Two more columns were added at this point; 'Year' and 'Country'. I had previously noted issues with the time date format in one piece of code from the prototype (Figure 26) when I had tried a test run of my UK code. This is why I added the 'Year' column, as a backup of sorts to use instead of the 'WeekID' example in Zubair's version

(salimzubair/lyric-sentiment: Sentiment analysis of all songs that entered the Billboard Hot 100 charts from 1958-2019, n.d.).

```

1 #Set the index of the dataframe to the WeekID. This sets us up to resample dataframe based on time
2 hot100_df['WeekID'] = pd.to_datetime(hot100_df['WeekID'],infer_datetime_format=True)
3 hot100_df = hot100_df.sort_values(by='WeekID')
4 hot100_df = hot100_df.reset_index(drop=True)
5 hot100_df = hot100_df.set_index('WeekID')

```

Figure 26 - Setting the index of the dataframe to be sorted by 'WeekID' column. Although this worked for Zubair, it had caused errors on the UK code test run.

The ‘Country’ column was added because I had an interest in discovering how the sentiment of the lyrics differed throughout the nationality of the artists. This happened to be one of the more time-consuming aspects to the project as I needed to manually input the countries of all 1360 number ones. The majority of this knowledge was known to me but I had to research a substantial portion of this information online with sites such as allmusic.com (AllMusic | Record Reviews, Streaming Songs, Genres & Bands, 2021). I finally had my own creation of a UK dataset with which to implement in Google Colab.

A	B	C	D	E	F	G	H
Index	Artist	Song Title	Record Label	Week ID	Year	Country	Weeks at number one
1	1 Al Martino	"Here in My Heart"	Capitol	14/11/1952	1952	AMERICA	9
2	2 Jo Stafford	"You Belong to Me"	Columbia	15/01/1953	1953	UNITED KINGDOM	1
3	3 Kay Starr	"Comes A-Long A-Love"	Capitol	23/01/1953	1953	AMERICA	1
4	4 Eddie Fisher	"Outside of Heaven"	HMV	30/01/1953	1953	AMERICA	1
5	5 Perry Como	"Don't Let the Stars Get in Your Eyes"	HMV	06/02/1953	1953	AMERICA	5
1357	1356 Shawn Mendes and Camila Cabello	"Señorita"	EMI, Syco	18/07/2019	2019	CANADA/CUBA	1
1358	1357 Ed Sheeran featuring Khalid	"Beautiful People"	Asylum	25/07/2019	2019	UNITED KINGDOM	1
1359	1358 Ed Sheeran featuring Stormzy	"Take Me Back to London"	Asylum	05/09/2019	2019	UNITED KINGDOM	5
1360	1359 Tones and I	"Dance Monkey"	Bad Batch	10/10/2019	2019	AUSTRALIA	11
1361	1360 LadBaby	"I Love Sausage Rolls"	FiftyFive	26/12/2019	2019	UNITED KINGDOM	1

Figure 27 - UK dataset created with added columns. First five number 1s (a) and last five number 1s (b)

Task 4

Modifications to Zubair’s code needed to be made in order for my newly constructed UK dataset. An update had been made to genism which necessitated altering the pip install of that library.

```

4 #!pip install -U gensim
5 !pip install gensim==3.8.3

```

Figure 28 - Altering genism install code.

The prototype had been loaded in from data.world, and although I stored my csv up on the hosting site too, I decided to show that it could also be uploaded from the local machine easily.

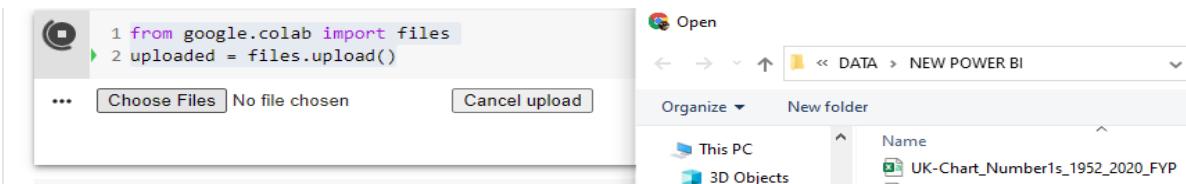


Figure 29 - Importing csv file from local drive.

Using the previously demonstrated ‘get_lyrics’ and ‘get_lyric_sentiment’ functions, my UK dataset was now amended with the Python code to display all available lyrics from genius.com and also the sentiment polarity score for each number one in the table. 1248 number ones were displayed, suggesting that lyricsgenius could not locate the lyrics to 112 of the songs in my originally uploaded dataset.

	Index	Artist	Song_Title	Record_label	Week_ID	Year	Country	Weeks_at_number_one	Lyrics	Sentiment
0	1	Al Martino	'Here In My Heart'	Capitol	14/11/1952	1952	AMERICA	9	Here in my heart I'm alone, I'm so lonely\nHer...	0.055000
1	2	Jo Stafford	'You Belong to Me'	Columbia	16/01/1953	1953	UNITED KINGDOM	1	See the pyramids along the Nile\nWatch the sun...	0.214286
2	3	Kay Starr	"Comes A-Long A-Love"	Capitol	23/01/1953	1953	AMERICA	1	The last tall son of Lot and Bellicent,\nAnd t...	0.153604
3	4	Eddie Fisher	"Outside of Heaven"	HMV	30/01/1953	1953	AMERICA	1	I pass your house with misty eyes\nThere stand...	0.195382
4	5	Perry Como	"Don't Let the Stars Get in Your Eyes"	HMV	06/02/1953	1953	AMERICA	5	Don't let the stars get in your eyes\nDon't le...	0.489784

Figure 30 - 'Lyrics' and 'Sentiment' columns added to UK dataframe..

An example plot graph was produced to show the sentiment values of the initial years of the chart. This would prove to be helpful later on when running the sentiment analysis on the whole dataset.

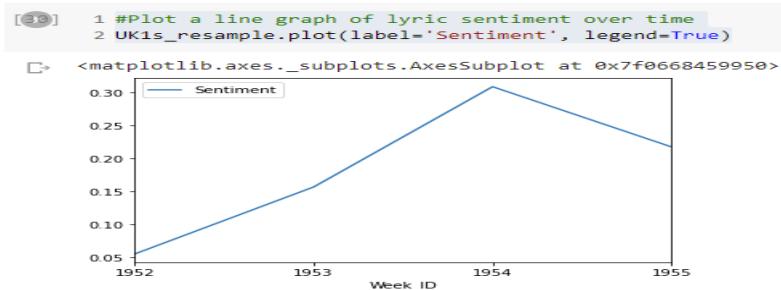


Figure 31 - Plot graph of UK chart (1952-1955).

Lastly, a new piece of code was added to save the amended dataset back into a csv file for further modifications. The entire code run of the UK dataset can be viewed in Appendix D – Google Colab run of UK dataset.

```
[35] 1 from google.colab import files
2 UK1s_df.to_csv('UK_Number1s_Lyrics_FYP.csv')
3 files.download('UK_Number1s_Lyrics_FYP.csv')
```

Figure 32 - Amended dataset saved as a csv file.

When opening the new csv file, it was instantly clear that there were numerous errors within the data. Firstly, there appeared to be random sequences of text throughout the document. Secondly, the sheet was well over 51,000 lines in length.

A	B	C	D	E	F	G	H	I	J
Week ID	Index	Artist	Song Title	Record Label	Year	Country	Weeks at number one	Lyrics	Sentiment
2	14/11/1952	1 Al Martino	"Here in My Heart"	Capitol	1952	AMERICA	9	Here in my heart I'm alone, I'm so lonely	0.055
3	15/01/1953	2 Jo Stafford	"You Belong to Me"	Columbia	1953	UNITED KINGDOM	1	See the pyramids along the Nile	0.214285714
4	23/01/1953	3 Kay Starr	"Comes A-Long A-Love"	Capitol	1953	AMERICA	1	The last tall son of Lot and Bellicent,	
5		f her good horse							
6		Was nigh to burst with violence of the beat							
7		Perforce she stay and overtaken spoke,							
8		What doest thou scullion in my fellowship?							
9		Deem'st thou that I accept thee aught the more							
51608	18/07/2019	1356 Shawn Mendes and Camila Cabello	"Señorita"	EMI, Syco	2019	CANADA/CUBA	1	[Chorus: Camila Cabello]	0.24553571
51609	25/07/2019	1357 Ed Sheeran featuring Khalid	"Beautiful People"	Asylum	2019	UNITED KINGDOM	1	[Intro: Ed Sheeran]	0.211
51610	05/09/2019	1358 Ed Sheeran featuring Stormzy	"Take Me Back to London"	Asylum	2019	UNITED KINGDOM	5	[Chorus: Ed Sheeran & Stormzy]	0.02725029
51611	10/10/2019	1359 Tones and I	"Dance Monkey"	Bad Batch	2019	AUSTRALIA	11	[Verse 1]	0.27857143
51612	26/12/2019	1360 LadBaby	"I Love Sausage Rolls"	Fiftyfive	2019	UNITED KINGDOM	1	[Verse 1: Roxanne]	0.35692641

Figure 33 - error-laden csv file.

Upon further investigation, I found that some songs had been presented with completely unrelated, and many obscene, lyrics; Hale & Pace's 1991 charity single, "Stomp", contained lyrics from a 2012 song by Odd Future, called "P"; 1953 hit, "Look At That Girl", by Guy Mitchell was supplied with the lyrics from, what Google describes as a 'rap battle' between HFK vs Charron, "King of the Dot"; and another charity song, Band Aid 20's, "Do They Know It's Christmas" displayed a peculiar mixture of 2012's, "Jet Skis in Hawaii", by LiveRaw GBT, and, bizarrely, Ernest Hemmingway's 1927 book of short stories, "Men Without Women".

<p>[Intro] Broke dick, I'm looking for a drug lord</p> <p>[Verse 1: Hodgy Beats] Okay, welcome to my 12 bar (One, um..) The beat, wears it like a kevlar As I smoke my tree, medlar, South African Coughing 'til there's pains in my fucking abdomen I spit negative just like a halogen My breeze more of a fucking masculine The rascals win, killing 'em off with a javelin Snakes, I'm just rattling, suit tailor never tattling You niggas a bunch of squares, Madison As far as real niggas, bitch nigga my battle's sick On top of being talented, bitch I'm fucking passionate 1 I'm a golden curse, call me</p>	<p>[Round 1: Charron] Over the years, we've partied a lot I've gotta thank you, you're my dude I picked up that Scotian chick in the club You let me bang her in your room But two hours later, you went down on her Don't say that isn't true You gave them kids a licking like Jamaican parents do! I'm the type to bang a bitch, and tell the broad I'll see her later You ate out that pussy like the Octomom in labor! And what about that girl at your work party you harassed for a bit See was passing out drunk and you started grabbing her tits Your boss found out and fired your ass for that shit ; You told me not to bring it</p>	<p>I say Jet skis in Hawaii, the appearance what she liking Tweeting bout the weather while I'm...on...the beach skypin' I...bet Hawaiian chicks be dykin' Chilling on...shore, no fightin' Up in Hawaii for my birthday Got a lot of cake, no icing She throw it at me, imma stroke her butt Sip ciroc coconut out of coconuts Popping Rose at the Luau So if the girl throw it at me, then I know to fuck Imma call her up, imma pick her up, imma smash her out Imma lift her up On the ocean view, she can get the nut Cuz imma be a king with the</p>
--	---	--

Figure 34 - Erroneous lyrics assigned to Hale & Pace (a), Guy Mitchell (b), and Band Aid 2020 (c)

I decided to manually remove the large chunks of rows with text which were not needed. Although this was, again, time-consuming, it was not difficult to execute. Resolving the misleading lyrics situation, though, was not a challenge that would be accomplished with such ease. There was simply not a sufficient amount of time to sift through all 1248 song lyrics to unearth every error. I was comfortable enough to proceed with performing the sentiment analysis on the trust that the majority of the data would have been successfully pulled down from genius.com.

Task 5

The revised dataset was uploaded to Power BI to and was ready for the sentiment analysis to be performed on it. The full report is in Appendix E – Sentiment Analysis in Power BI, it can be used interactively at this link [Sentiment Analysis in Power BI](#), or a pdf version can be viewed here [Power BI pdf report](#).

Firstly, three stacked column charts and three area charts were created (Figure 123-81) in order to display the following relevant statistics per year.

- Average sentiment
- Standard of deviation
- Variance
- Median

- Minimum
- Maximum

These charts provided an early indication of what to expect at the outcome of the sentiment analysis. The two key ones for this observer were the average sentiment and the median charts. To take the year, 1969, as an example; the average sentiment was very much on the lower end of the scale, at 0.03. However, the median of 1969 was 0.09, a not insignificant increase.

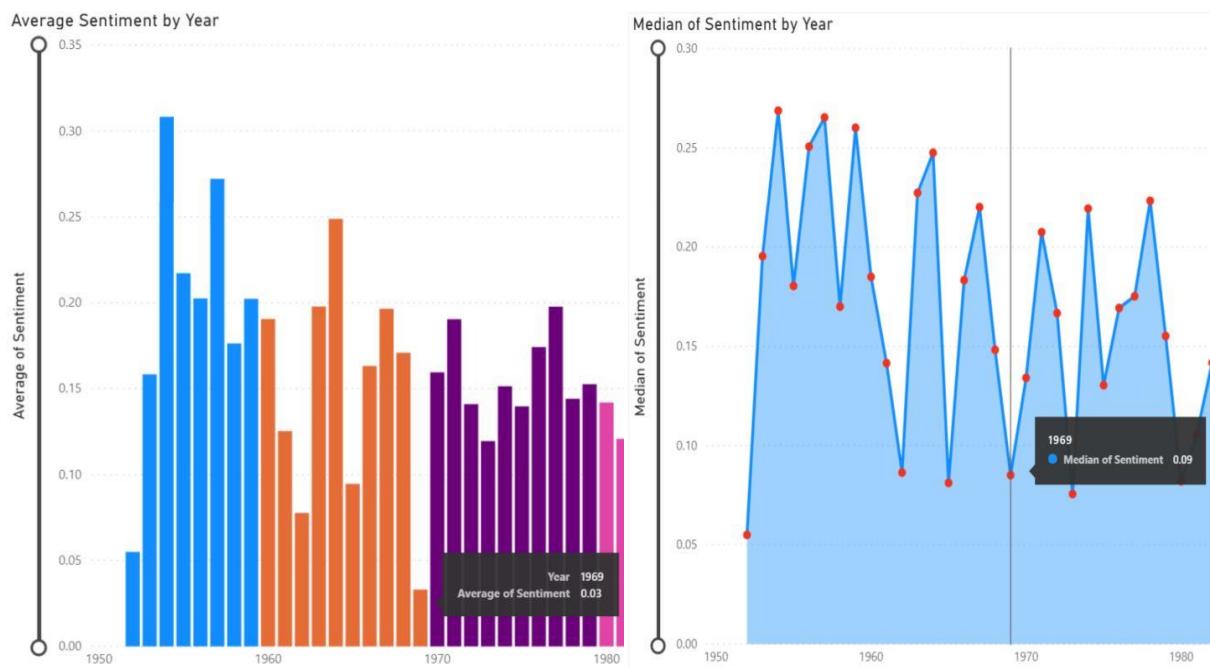


Figure 35 - Chart (a) showing 0.03 as the average sentiment of 1969, while the median of 0.09 is displayed in (b).

Why the disparity? Some people find it quite difficult to understand the difference between average (mean) and median so let us take a look at the 1969 data to try and explain it better.

Artist	Average of Sentiment	Song Title	Weeks at number 1	Nationality	Year
Creedence Clearwater Revival	-0.68	"Bad Moon Rising"	3	AMERICA	1969
Jane Birkin and Serge Gainsbourg	-0.50	"Je t'aime... moi non plus"	1	FRANCE	1969
Desmond Dekker & The Aces	-0.28	"Israelites"	1	JAMAICA	1969
The Move	-0.01	"Blackberry Way"	1	UNITED KINGDOM	1969
Tommy Roe	0.01	"Dizzy"	1	UNITED KINGDOM	1969
Rolf Harris	0.03	"Two Little Boys"	6	AUSTRALIA	1969
Marvin Gaye	0.04	"I Heard It Through the Grapevine"	3	AMERICA	1969
The Beatles	0.05	"The Ballad of John and Yoko"	3	UNITED KINGDOM	1969
The Beatles with Billy Preston	0.12	"Get Back"	6	UNITED KINGDOM	1969
The Archies	0.12	"Sugar, Sugar"	8	AMERICA	1969
Peter Sarstedt	0.16	"Where Do You Go To (My Lovely)?"	4	UNITED KINGDOM	1969
Thunderclap Newman	0.19	"Something in the Air"	3	UNITED KINGDOM	1969
Amen Corner	0.24	"(If Paradise Is) Half as Nice"	2	UNITED KINGDOM	1969
Bobbie Gentry	0.30	"I'll Never Fall in Love Again"	1	AMERICA	1969
The Rolling Stones	0.34	"Honky Tonk Women"	5	UNITED KINGDOM	1969
Marmalade	0.39	"Ob-La-Di, Ob-La-Da"	1	UNITED KINGDOM	1969
Overall Average Sentiment	0.03		49		

Figure 36 - 1969 dataframe in Power BI.

Figure 36 exhibits all of the data relating to 1969 in the dataset. The ‘Average of Sentiment’ column has been highlighted to display the polarity values. There are sixteen number ones in the table and to get the median value, it is just a matter of arranging the column in an orderly sequence and determining the middle value. As there are an even number of songs, we have two middle values, which will need to be added together and divided by two (LLC, 2021a).

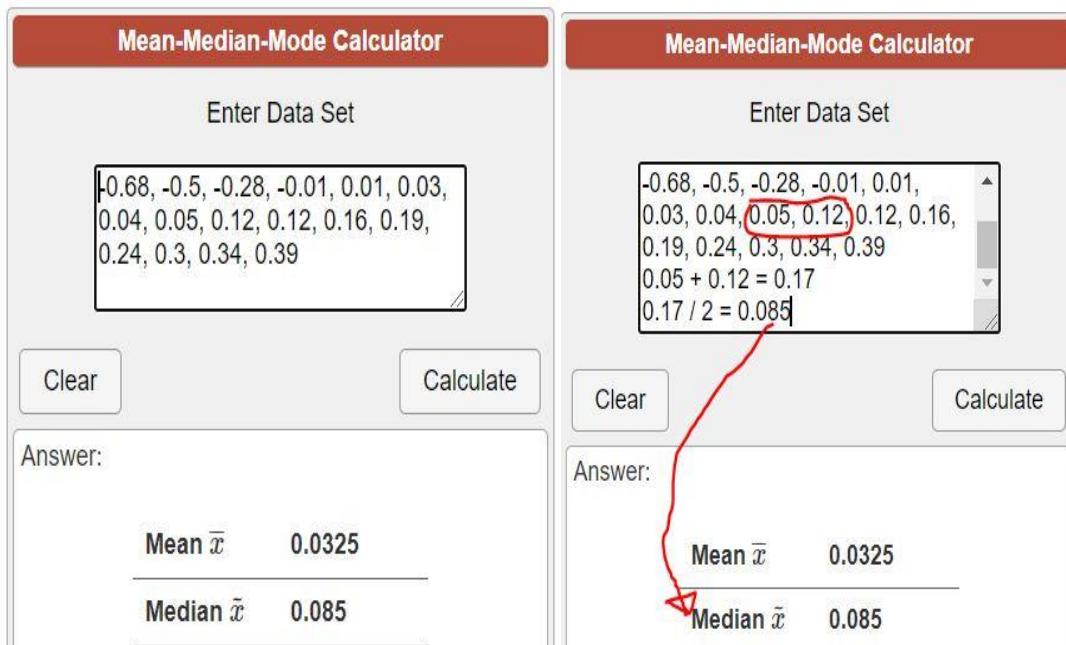


Figure 37 - Arranging the values in order (a) before calculating the median value (b).

To obtain the average, meanwhile, all of the values will be added up before being divided by sixteen (number of songs in the dataframe).

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

Mean-Median-Mode Calculator

Enter Data Set

```
-0.68, -0.5, -0.28, -0.01, 0.01, 0.03, 0.04,
0.05, 0.12, 0.12, 0.16, 0.19, 0.24, 0.3,
0.34, 0.39
= 0.52
0.52 / 16 = 0.0325
```

Answer:

Mean \bar{x}	0.0325
Median \tilde{x}	0.085

Figure 38 - Example of how to calculate average.

The difference between the two stats suggests a high variance of sentiment in 1969 and the evidence shows that it does, in fact, have the highest level of variance in the entire table. This means that there is a larger spread of values in 1969 than in any other year.

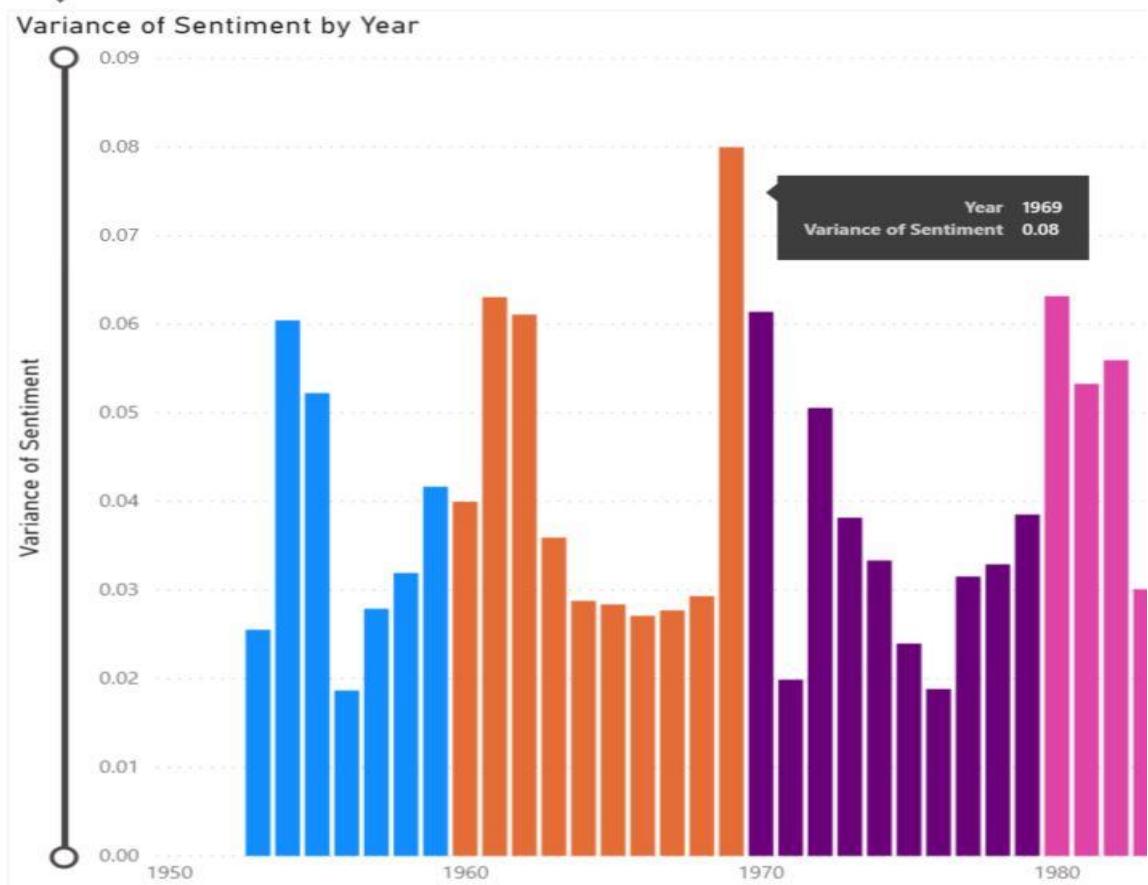
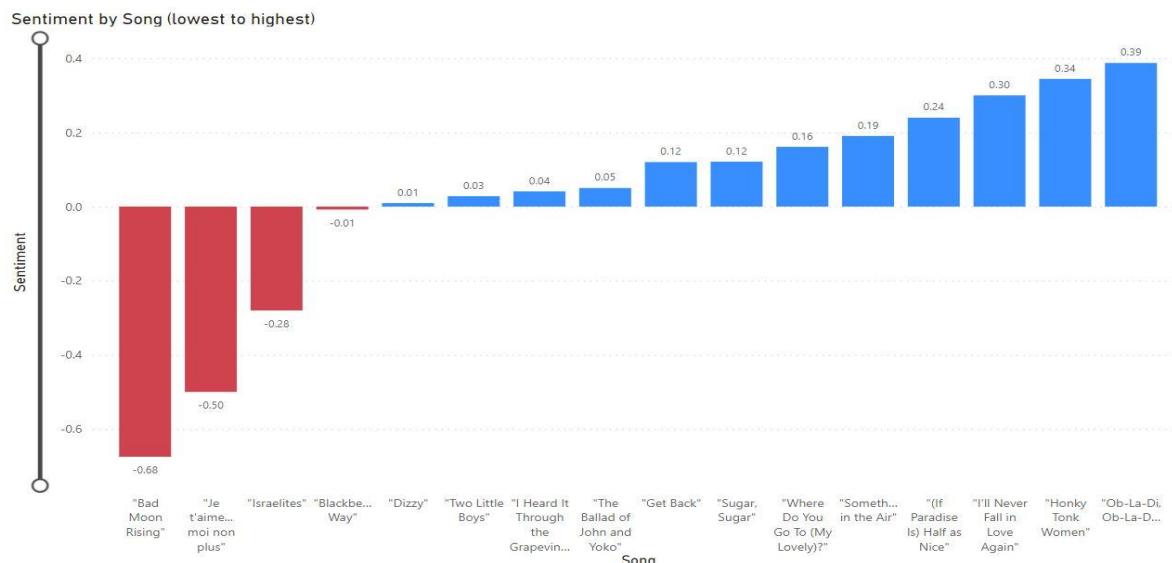


Figure 39 - Variance of sentiment in 1969.

Looking at this in more detail, we can see that 1969 contains three major outliers; Creedence Clearwater Revival's, "Bad Moon Rising"; "Je t'aime...moi non plus" by Jane Birkin and Serge Gainsbourg; and "The Israelites" from Desmond Dekker & The Aces.



Average of Sentiment by Artist (lowest to highest)

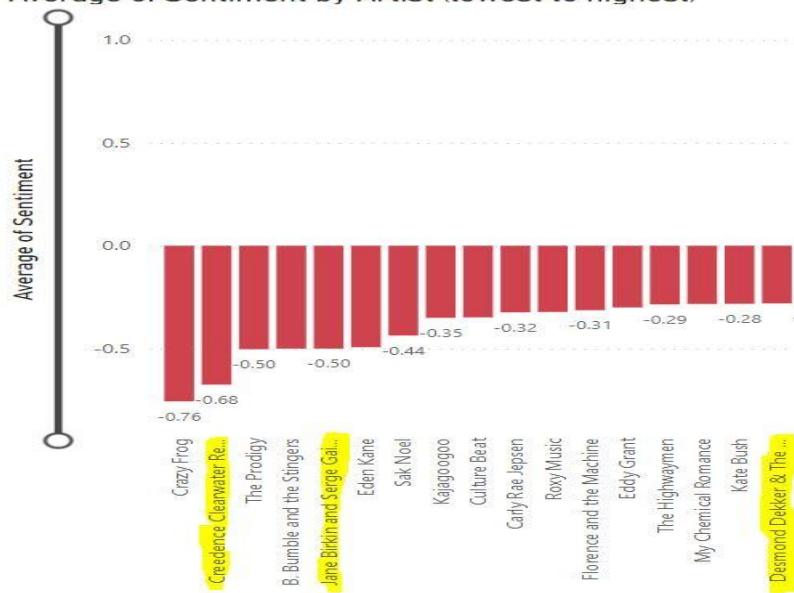


Figure 40 - 1969 outliers (a). How they compare to the full dataset (b).

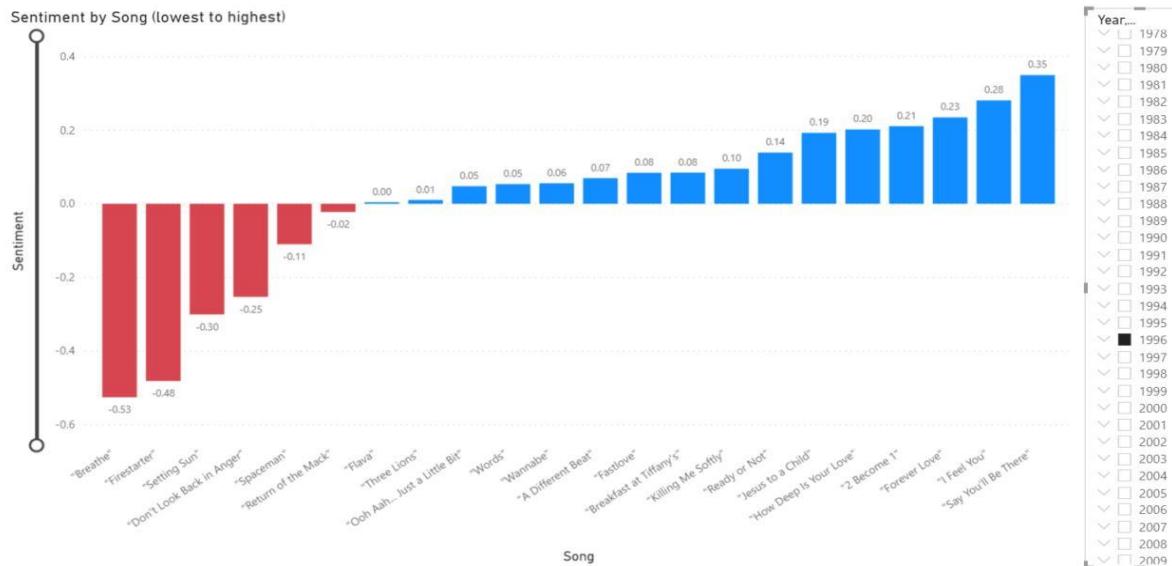
Not only are these songs low in sentiment value for the year of 1969, but of all 1248 number ones in the dataset, these three songs happen to be in the bottom 17 places for sentiment. Obviously, this has a profound impact on the average sentiment of 1969 as the data is skewed towards the negative elements of this year. There would be a case here for using the median value instead of the average, to show a truer reflection of the year's sentiment.

Another year which suffers a similar fate to the one above is, 1996. Officially, it is the year with the lowest average sentiment of all, at 0.02 (Figure 41). However, it is comparable to 1969 in that the disproportionately large number of negatively rated songs provide a distorted view. The four main outliers on this occasion; “Breathe” and “Firestarter”, both by The Prodigy; “Setting Sun” from The Chemical Brothers; and Oasis’, “Don’t Look Back in Anger”. All of these songs occupy a place in the bottom 27 of the 1248 total for sentiment polarity (Figure 42 (a) & (b)). As stated in the previous example, when outliers are involved, it could be suggested that the median value, 0.05, should be used as a more accurate representation in this case, too.

←
1996 - Year with the lowest average sentiment (0.02)

1996

Figure 41 - 1996, the lowest ranked year in the dataset with an average sentiment of just, 0.02.



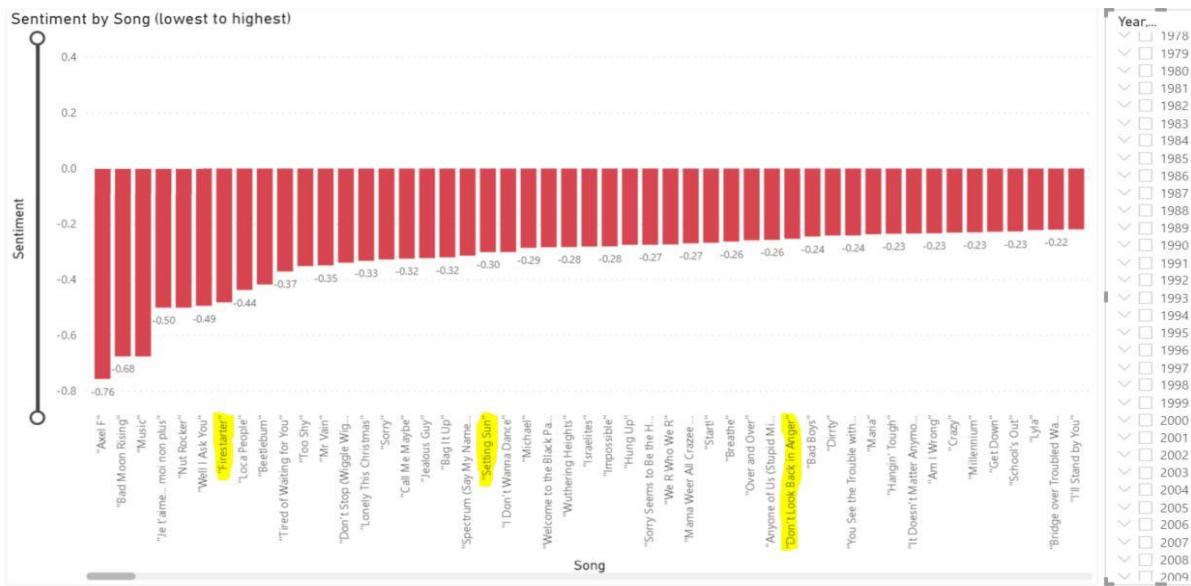


Figure 42 - 1996 data showing songs with a significantly low polarity (a), and those songs at the lowest end of the sentiment scale overall (b).

1969 and 1996 are included as two of the four years with the lowest average of sentiment; the other two being, the more recent, 2015, and 2016, respectively.

The four years with the lowest average sentiment (between 0.02–0.04)



Figure 43 - The four lowest scoring years over the history of the charts, of which two have occurred very recently.

The big difference, in my opinion, between the former two and the latter two years is that the variance of the latter two (2015, 0.02 and 2016, 0.01) is much smaller (Figure 44). This suggests to me that there are no outliers in either of these years, and so, their average sentiment is far more likely to be an accurate representation of central tendency than those of 1969 or 1996. Figure 45 shows a word cloud of common lyrics emanating from the four years mentioned and the words that sprung out to me were, “Fucking”, and “Piss”. Both words came from ZAYN’s, 2016 hit, “Pillowtalk”, reaffirming my belief that, although the earlier years may match the low polarity, the modern years certainly contain more profanities.

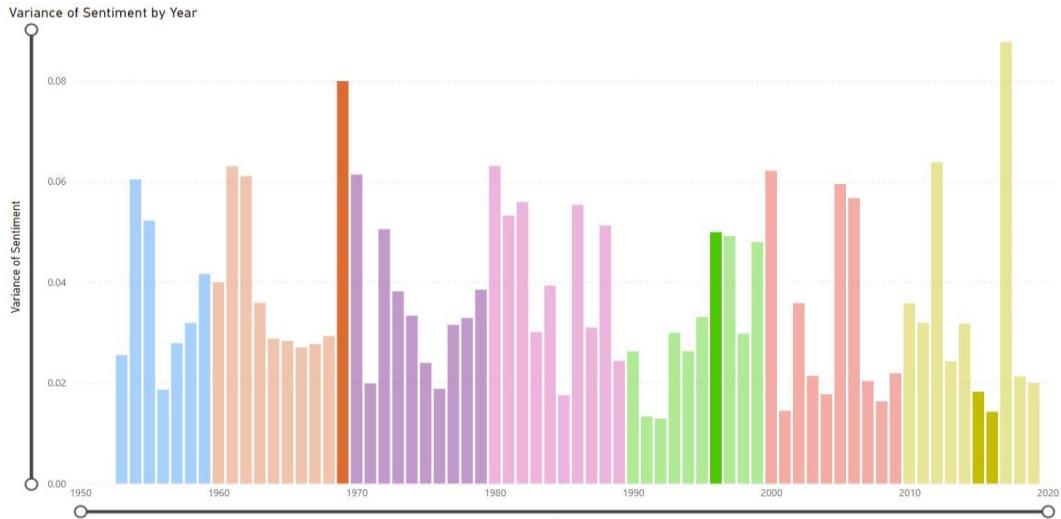


Figure 44 - Variances of the four lowest rated years; 1969, 1996, 2015, and 2016, are bolded.

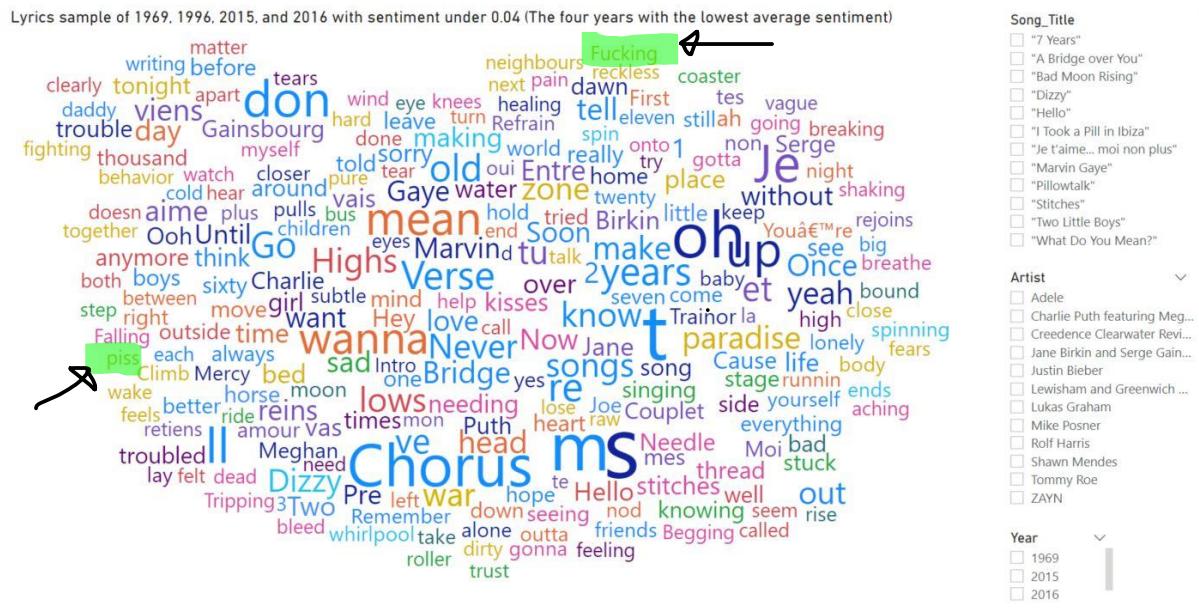


Figure 45 - Song lyrics from the four lowest sentiment rating years (word cloud).

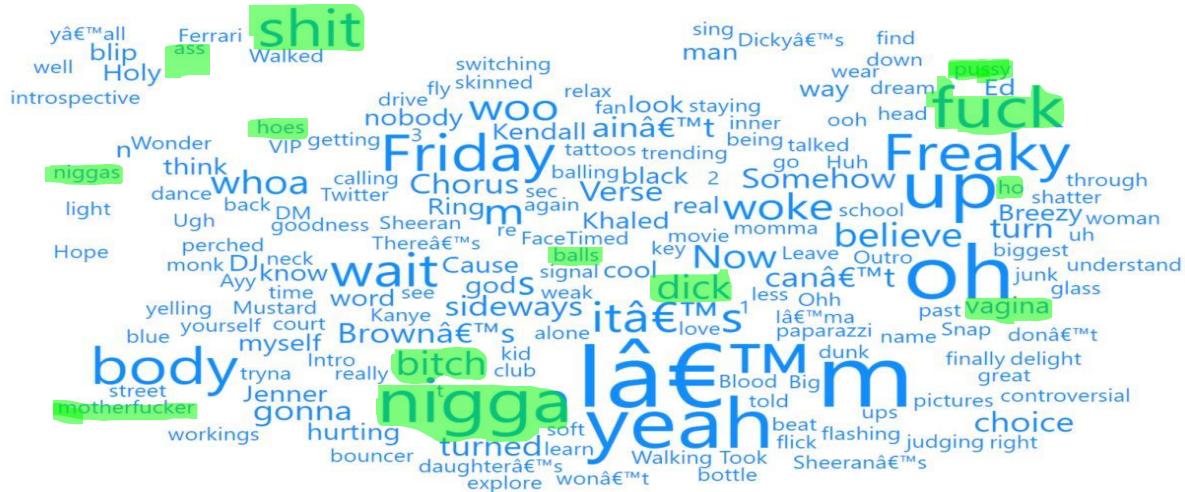
Just a couple of years later, in 2018, more than just obscenities could be found with Lil' Dicky and Chris Brown's, "Freaky Friday", and Drake's, "Nice For What" (Figure 46).

- Obscenities: "Shit"; "Fuck"; "Motherfucker"; "Motherfuckin"
- Misogyny / sex: "Ass"; "Hoes"; "Pussy"; "Ho"; "Vagina"; "Dick"; "Balls"; "Bitch"
- Racism: "Niggas"; "Nigga"

Amazingly, to me at least, these were not even the lowest polarity rated songs of that year as Clean Bandit's, "Solo", received a -0.12 value with no offensive words of which I could

detect (Figure 47). This is, maybe, where subjectivity needs to be applied though; I will discuss that further on in the paper.

2018 - Another song containing "Shit" and "Fuck" ("Freaky Friday" by Lil Dicky ft Chris Brown is full of expletives)



2018 - Another number 1 from this year containing "Shit" and "Fuck" (2018s "Nice For What" by Drake is full of expletives)



Figure 46 - "Shit and "Fuck" were amongst many obscene words littered throughout 2018 number ones, "Freaky Friday", by Lil' Dicky and Chris Brown.(a) and Drake's, "Nice For What" (b).

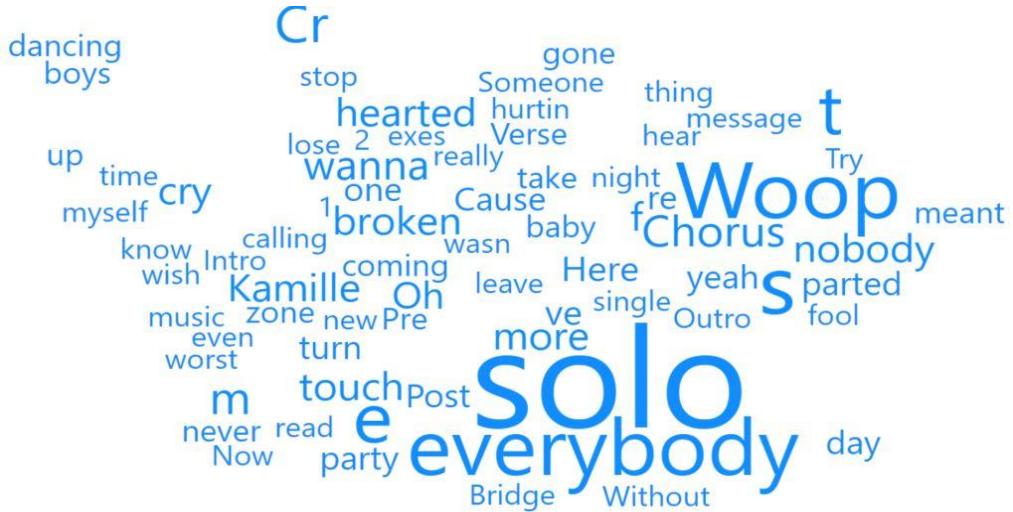


Figure 47 - Clean Bandit ft Demi Lovato's, "Solo" received the lowest polarity of 2018; -0.12.

(Figure 150) provides us with a word cloud snapshot of the aforementioned, most negative, years of 1996, 2015, 2016, and 2018; but also included is the year with the highest average sentiment rating, 1954. The difference in the language and tone, in comparison, is remarkable. With a polarity of 0.31, 1954's word cloud is full of lyrics (Figure 48) which would underpin its score; "Smile"; "Funny"; "Lovable"; "Adore"; "Joy"; "Beloved"; "Devine"; "Best"; "Precious"; "Gentle"; "Wonderful"; "Happiness"; "Laughter".

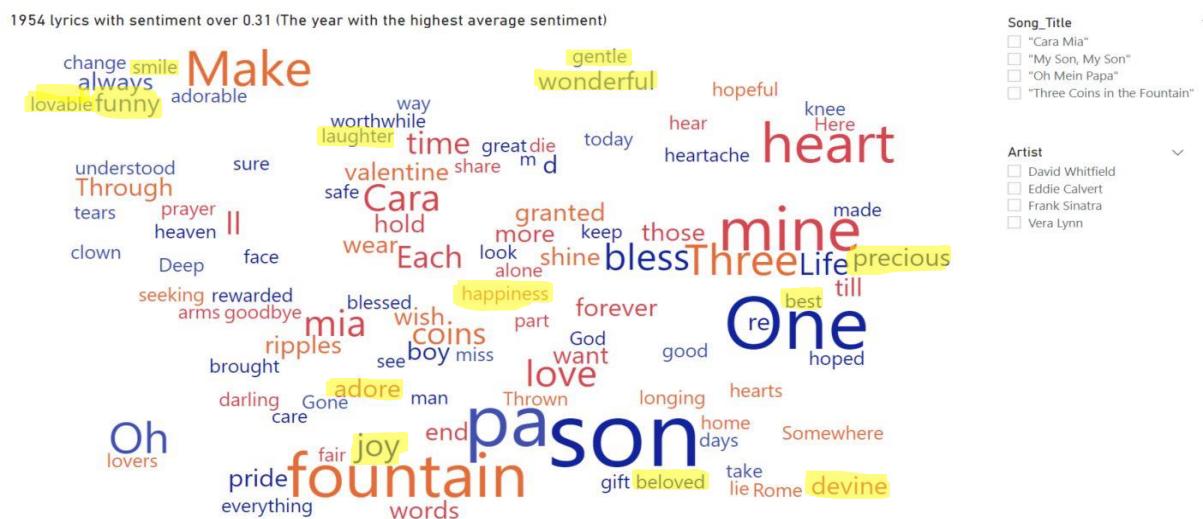


Figure 48 - Songs and their lyrics from 1954 which at least matched the 0.31 sentiment score.

I had earlier indicated the issue with outliers in relation to the more negatively oriented years. To bring a sense of balance, I should also highlight that this can be noticed just as prevalently in the years with the highest average sentiment. The top four years in this regard are.

- 1954
- 1957
- 1988
- 1994

The first three of these also include, what could be deemed as, outliers. Of the 1248 total number ones, 1954 has 4 songs in the top 50 of highest average sentiment; “Three Coins in the Fountain”, by Frank Sinatra; Vera Lynn’s, “My Son, My Son”; David Whitfield with, “Cara Mia”; and “Oh Mein Papa”, from Eddie Calvert (Figure 49).

There is one outlier in each of the other two years which appear in the overall top 50 for average sentiment; Frankie Vaughan’s, “The Garden of Eden”, 1957; and 1988’s (Figure 50), “Doctorin’ the Tardis”, by The Timelords (Figure 51).

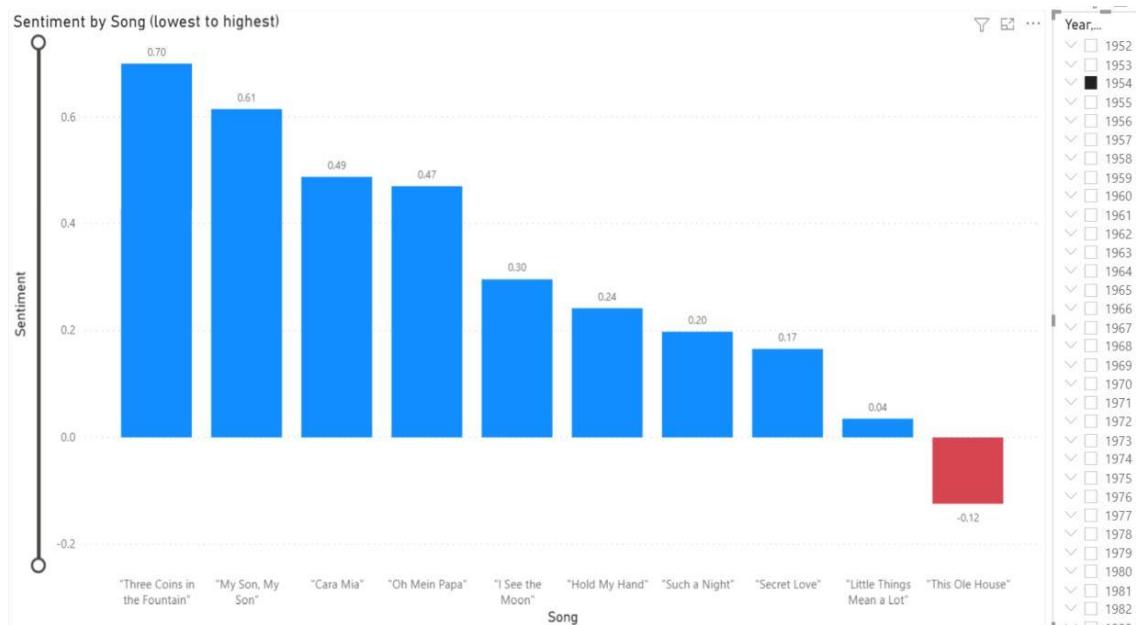


Figure 49 - 1954 outliers.

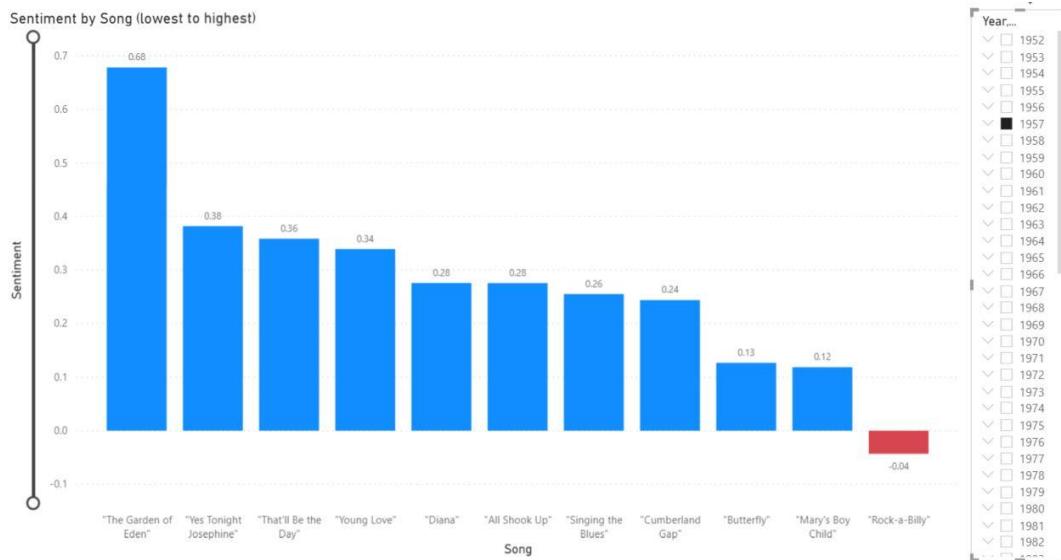


Figure 50 - 1957 outlier.

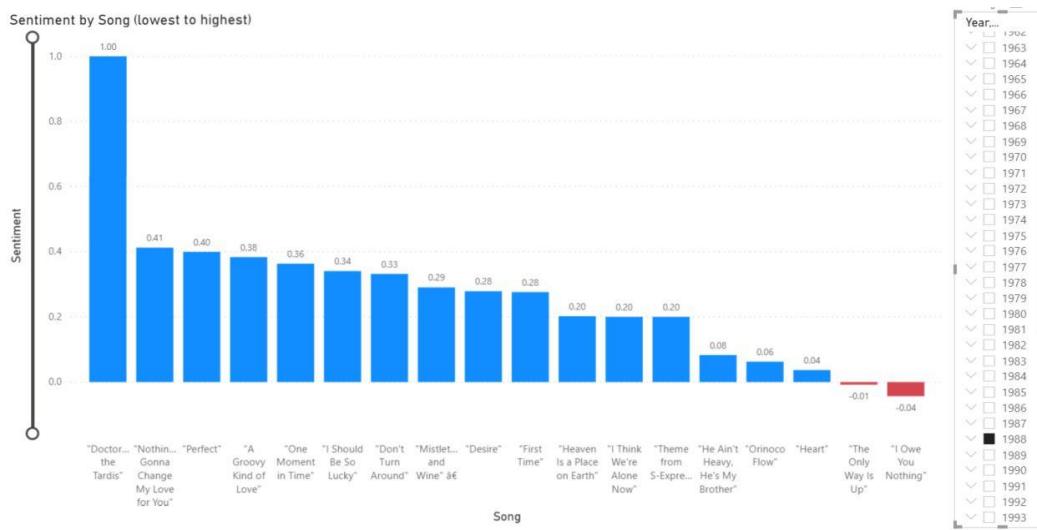


Figure 51 - 1988 outlier.

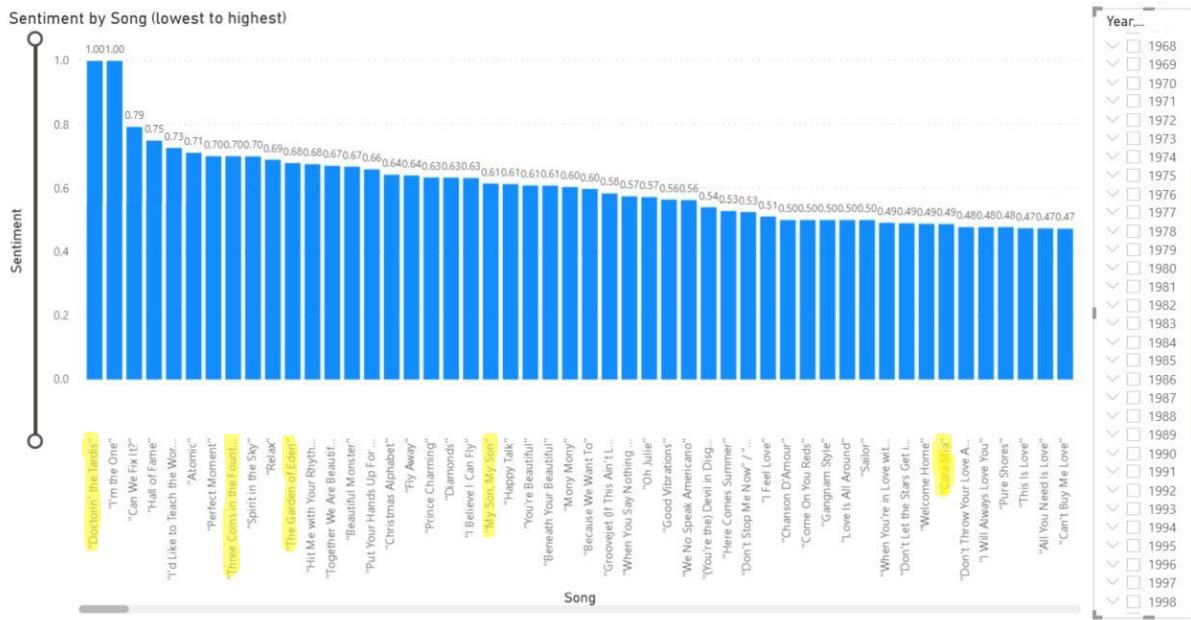


Figure 52 - The 50 songs with the highest average sentiment.

Yet if we take a look at the 4th year mentioned earlier, 1994, there appears to be no clear outliers (Figure 53). Yes, it can be said that the top 2 songs of 1994 (Manchester United's, "Come On You Reds", and "Love Is All Around", from Wet Wet) appear in the top 50 overall, but the smaller level of variance shows there is a much more consistent flow to the year. 1994 is another example of, possibly, showing the best use of an average rating over the median value, whereas with the previous three years, they could be more warranting of being judged by their median.

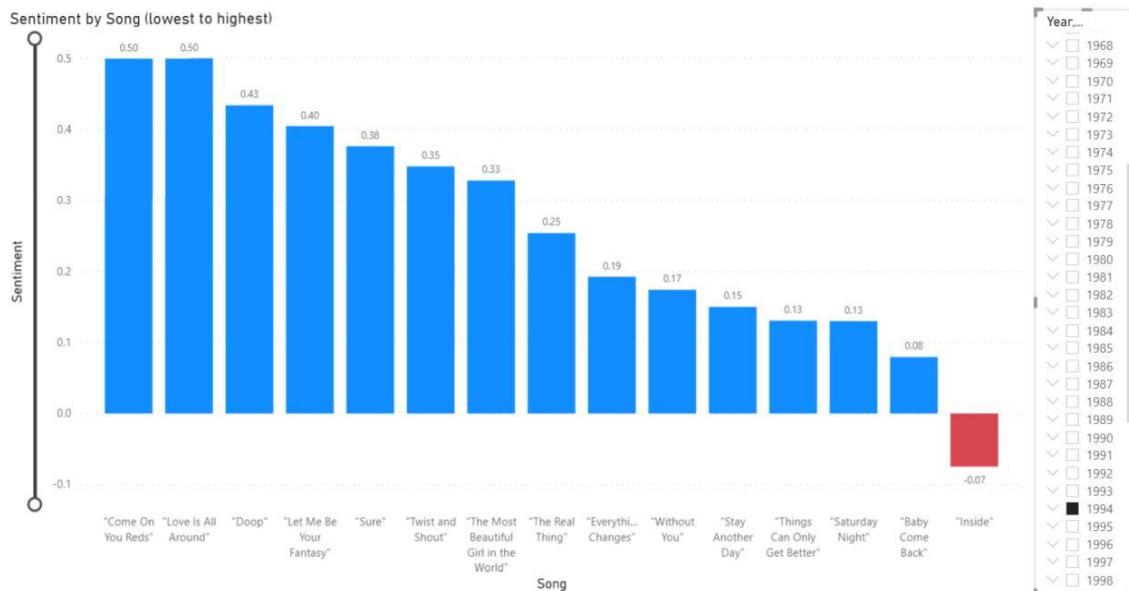


Figure 53 - 1994's lack of outliers.

So why, you may ask, are outliers so important to this author? Well, I would definitely not consider myself as a master of explaining such a science, but as a soccer fan, I will try to do so with a football analogy which, at the time of writing this paper, was quite topical.

I am a life-long supporter of Tottenham Hotspur (for my sins). Our star player is a striker by the name of, Harry Kane. He is a *positive*, brilliant player but there is talk this summer of Kane leaving ‘Spurs’ which would, of course, have a devastating impact on the team. Serge Aurier is a right-back and, in my opinion, a *negative*, terrible footballer. He also may be leaving ‘Spurs’, along with Lucas Moura, a decent footballer who has his plusses and minuses; a *neutral* player, if you like. So, if we break this down.

- Harry Kane = “Doctorin’ The Tardis”; Take either away from ‘Spurs’ or 1988 and their dynamic shifts significantly as they are instantly not as positive.
- Serge Aurier = “Bad Moon Rising”; Take either away from ‘Spurs’ or 1969 and their dynamic shifts significantly as they are instantly not as negative.
- Lucas Moura = “Want You To Want Me”; Take either away from ‘Spurs’ or 2015 and their dynamic remains mainly the same.

To summarise, Kane and Aurier are outliers whose absence would have a major impact, positively or negatively, on Spurs as a whole. If Moura were taken away, the impact would be minuscule. Judging by this, I would say that, on reflection, there is an argument for suggesting that 2015, and 2016, are truer negative years than 1969 or 1996; and 1994 is probably the most positive year over 1954, 1957, and 1988.

Looking at the 24 songs which have had the longest reigns at number 1, 8 weeks or more, we can see that 4 of the top 7 in terms of average sentiment charted in the 1950s; “Cara Mia”, from David Whitfield; “Oh Mein Papa”, by Eddie Calvert; Slim Whitman’s, “Rose Marie”; and Perry Como’s, “Magic Moments”.

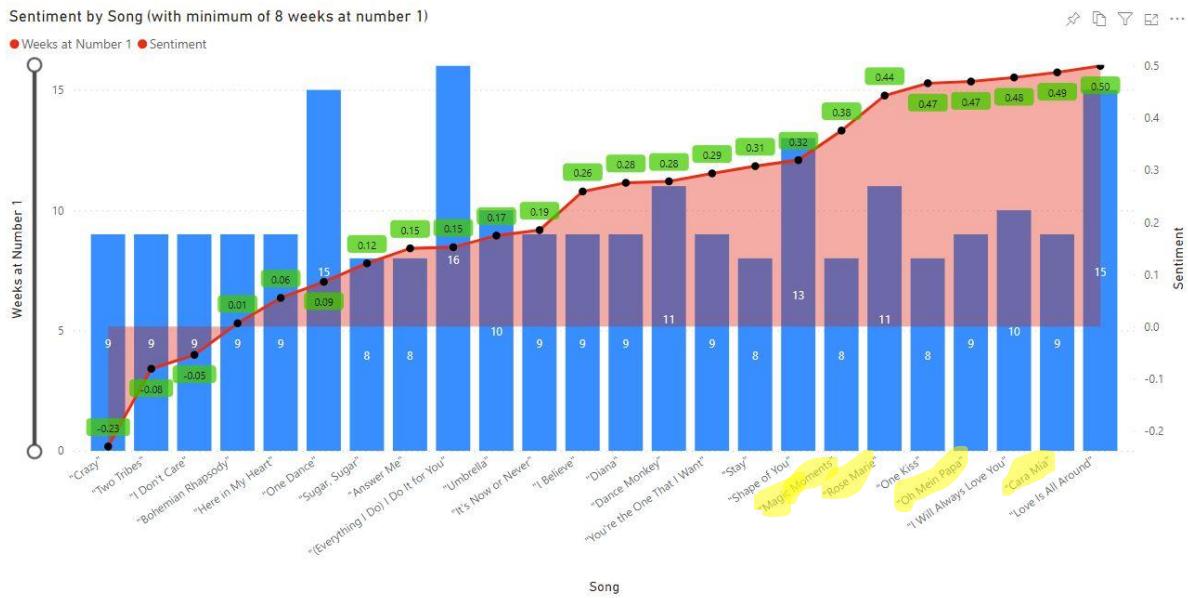


Figure 54 - Sentiment of songs which topped the UK charts for at least eight weeks. The 1950s perform particularly well.

Figure 55 is a donut chart which provides an alternate view of the yearly average sentiment. Interesting to note is that of all 68 years studied, the 21st century is last to enter the picture with 2012 and 2008 appearing 17th and 18th, respectively. The statistics illustrated in (Figure 56) would corroborate the notion that the sentiment of lyrics has fallen over the history of the UK charts.

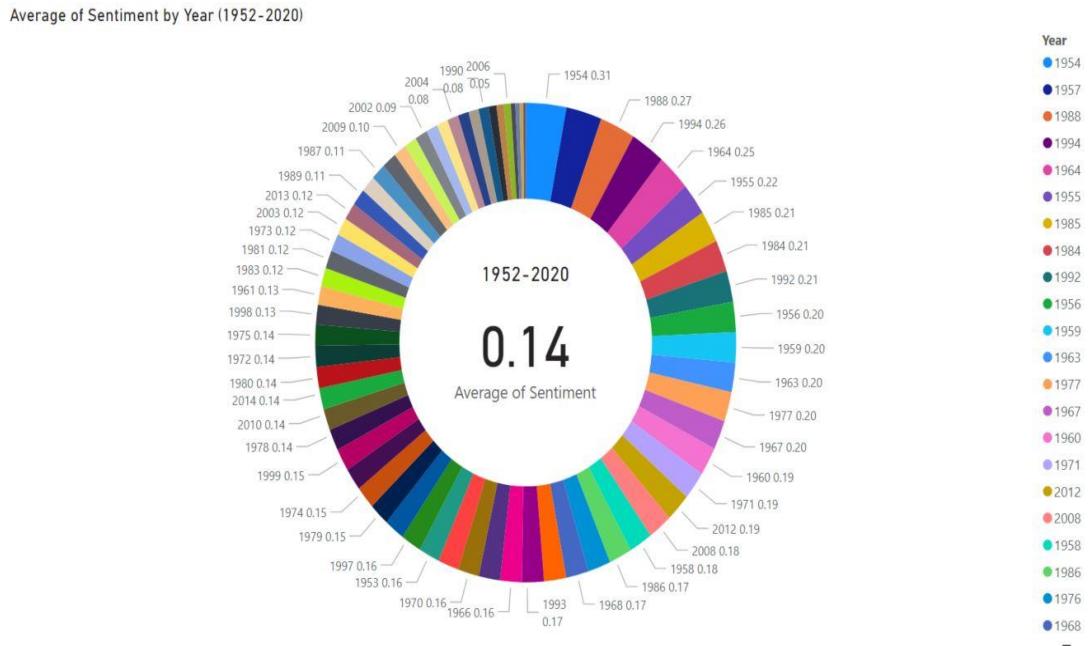


Figure 55 - The average sentiment of every year of the UK chart, 1952-2020

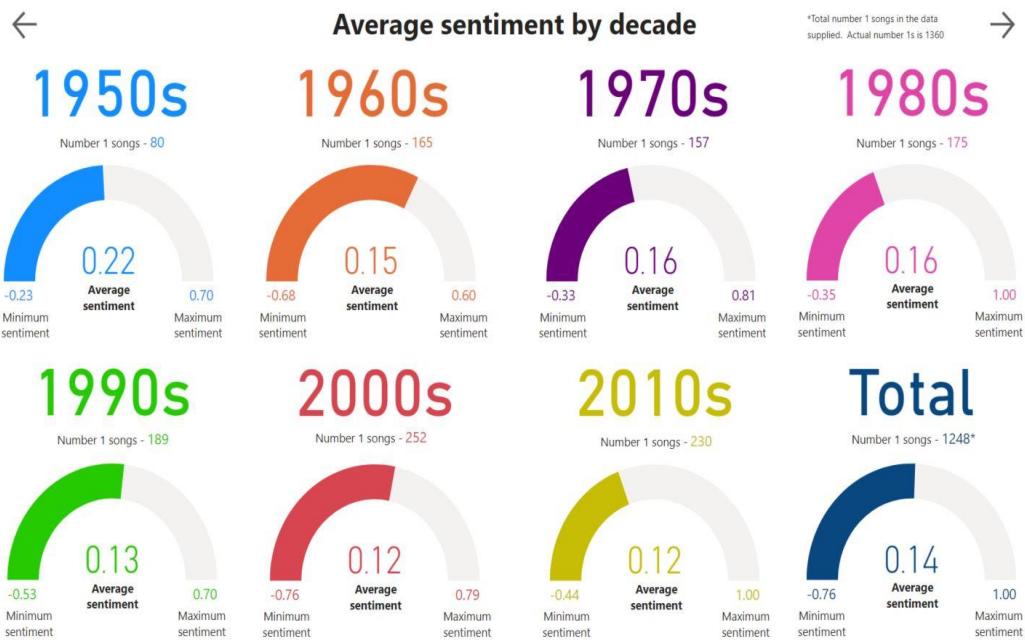


Figure 56 - The average sentiment for each decade of the UK charts.

That concludes my look at average sentiment, in general, but I will delve deeper into specific language in the results section of this paper (Chapter 6: Results Analysis of Data).

5.6 Sprint 4 – Researching UK cultural and societal data

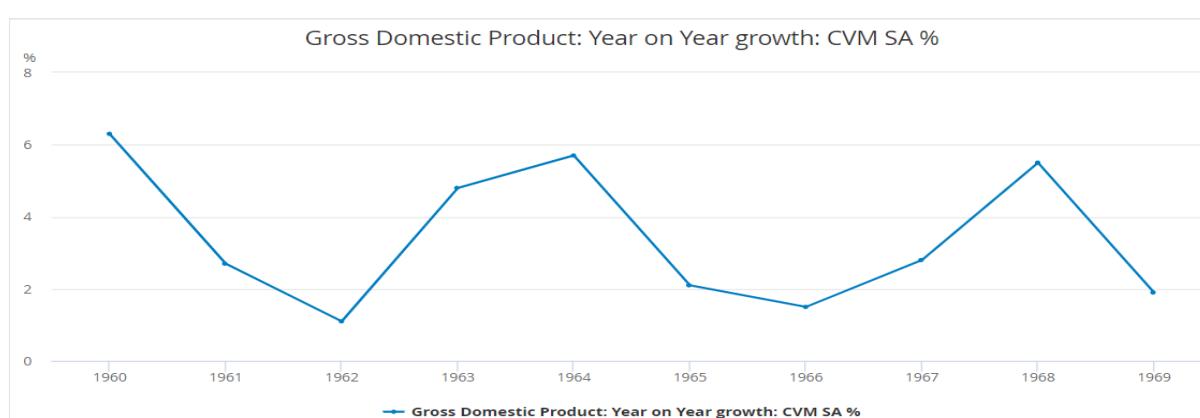
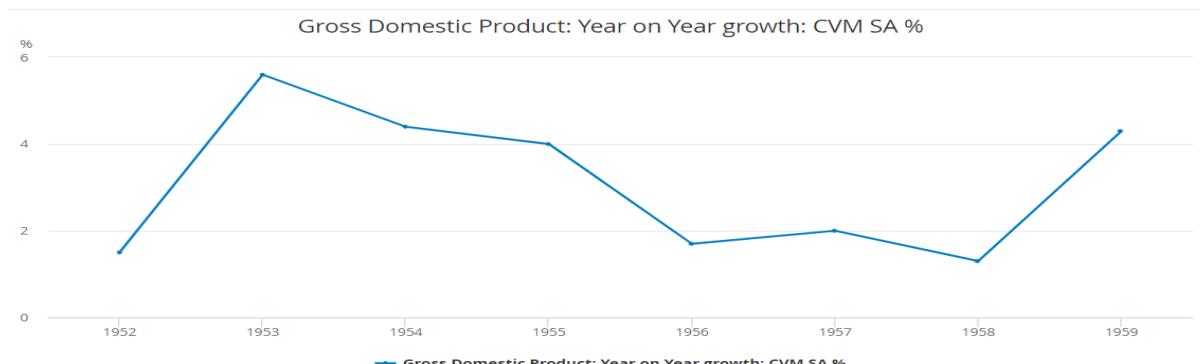
Sprint ID	Description	Start Date	Finish Date
4	Researching UK cultural and societal data	22/03/2021	11/04/2021

Task ID	Description	Status
1	Researching UK cultural and societal data	Completed

Task 1

My aim was to research and find data about culture and society, whilst keeping it relevant enough to be able to determine any correlation with the UK number ones data. This was possibly the most difficult aspect of the thesis as time was very limited during this period. I am hoping that some of the information gathered will be of use for my concluding chapter.

Beginning with some line graphs exhibiting the annual gross domestic product (GDP) of the UK between 1952-2020 (Gross Domestic Product: Year on Year growth: CVM SA % - Office for National Statistics, 2021).



Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

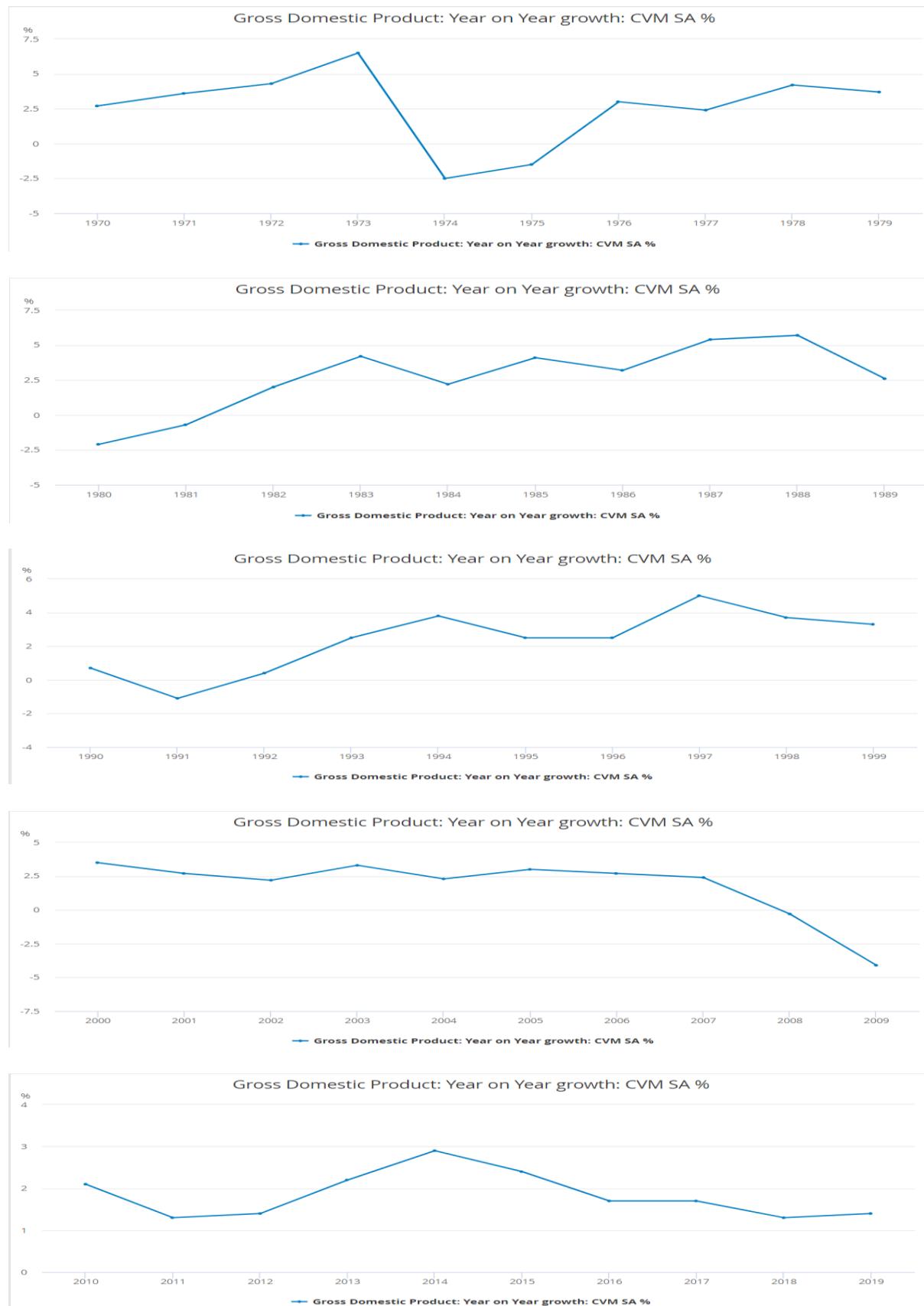


Figure 57 - Line graphs detailing the annual UK GDP of the 1950s (a), 1960s (b), 1970s (c), 1980s (d), 1990s (e), 2000s (f), and 2010s (g)

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

“The only two certainties in life are death and taxes”.

A famous quotation from Benjamin Franklin which still resonates today (Investor Brief - Death And Taxes, 2021). I am curious to see if there is any link between Income Tax and the sentiment of the decades (Public Spending Chart for United Kingdom 1950-2019 - Central Government Local Authorities, 2021).

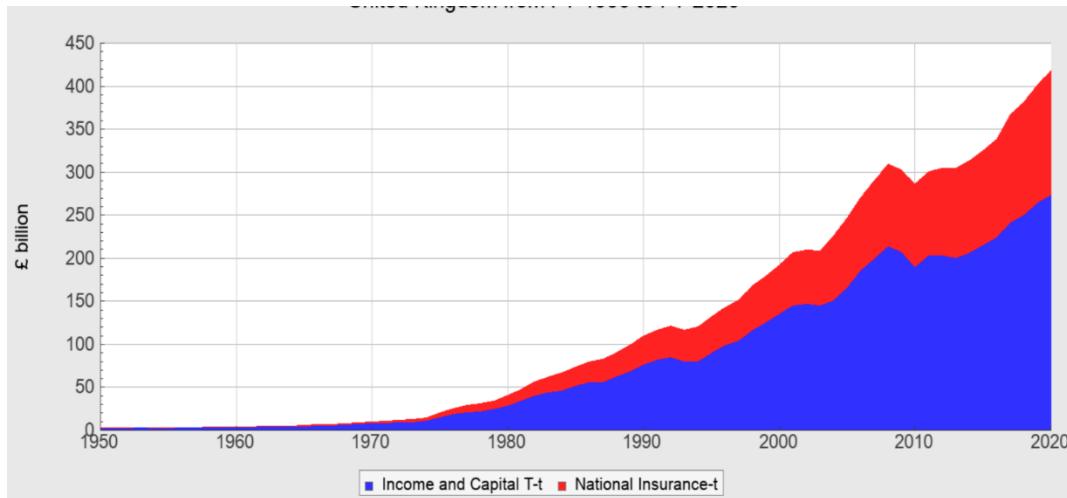
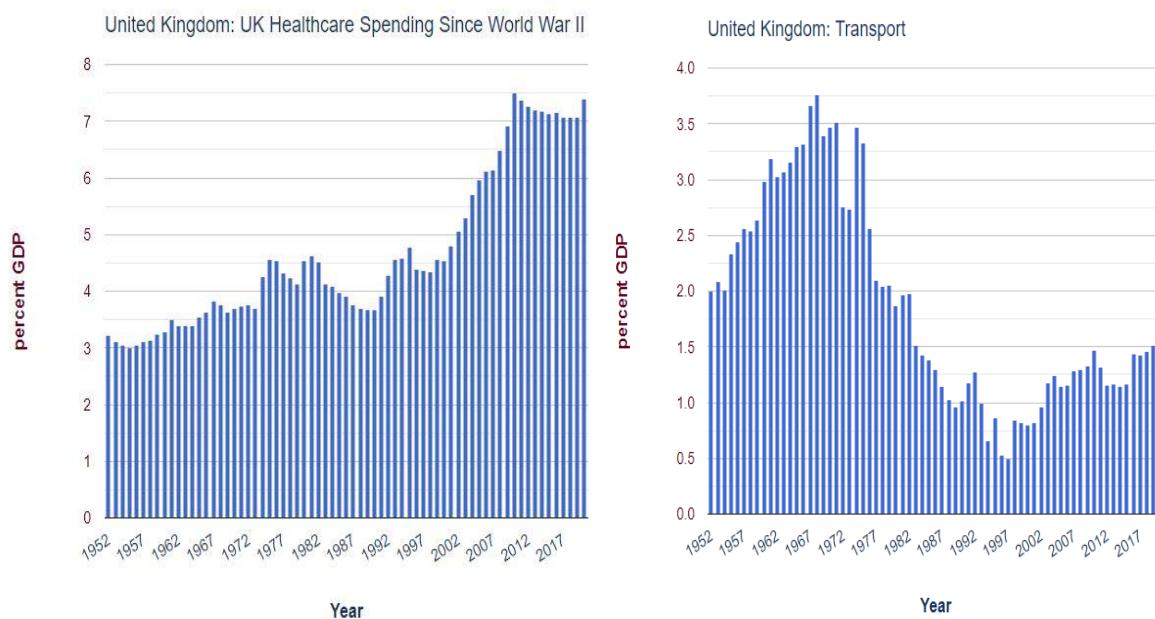


Figure 58 - Chart displaying significant growth relating to Income Tax and National Insurance.

There is a definite possibility that the sentiment across society would be impacted by public spending on; Healthcare; Transport; Education; Welfare.

(UK Healthcare Spending Since World War II for United Kingdom 1952-2020 - Central Government Local Authorities, 2021)



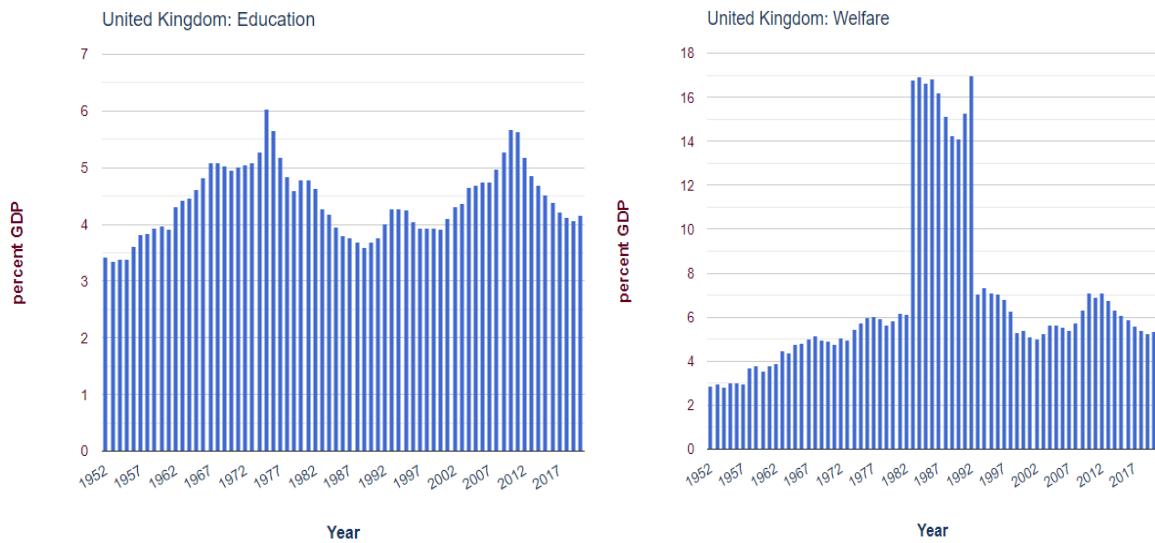


Figure 59 - Public spending figures for healthcare (a), transport (b), education (c), and welfare (d).

As with public spending, unemployment rates are certain to effect the mood of the nation
 (United Kingdom Long Term Unemployment Rate | 1992-2020 Data | 2021-2023 Forecast, 2021).

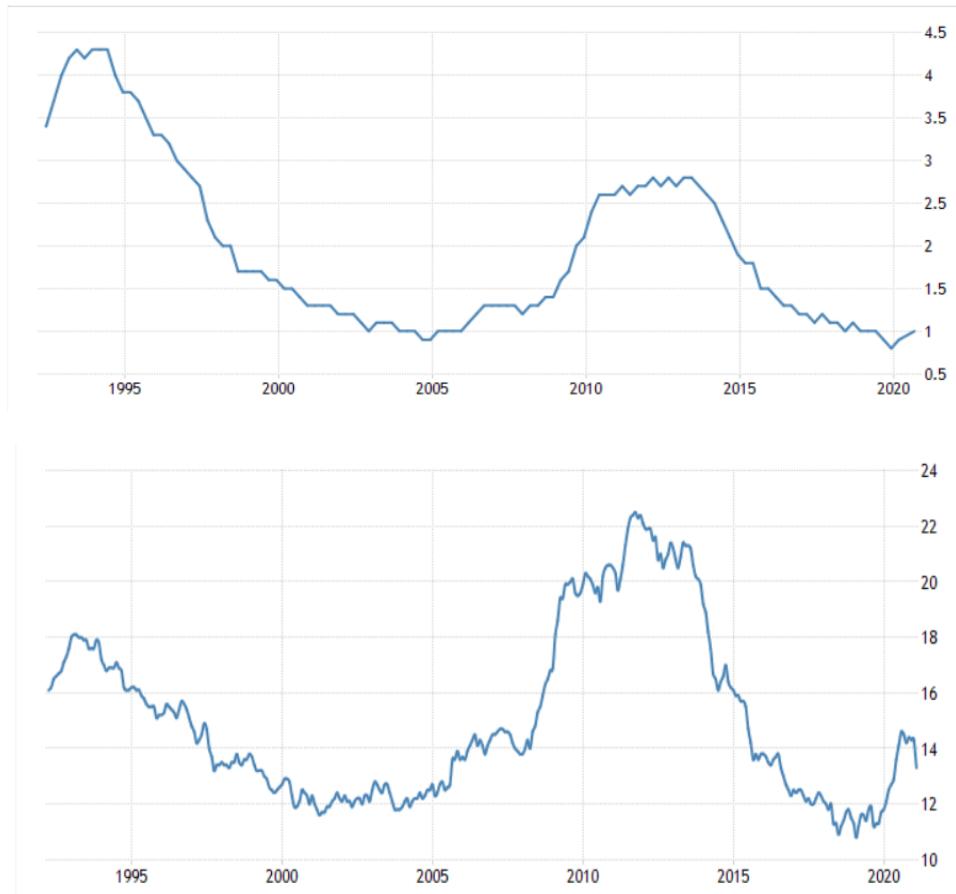


Figure 60 - Unemployment rates in general terms (a), and youth unemployment (b).

Crime data was the one thing I strived to find more than any other. It is intriguing to discover if there is a correlation which could be established between crime rates, and the increase in profanities in lyrics (Crimes of the century, 2021).

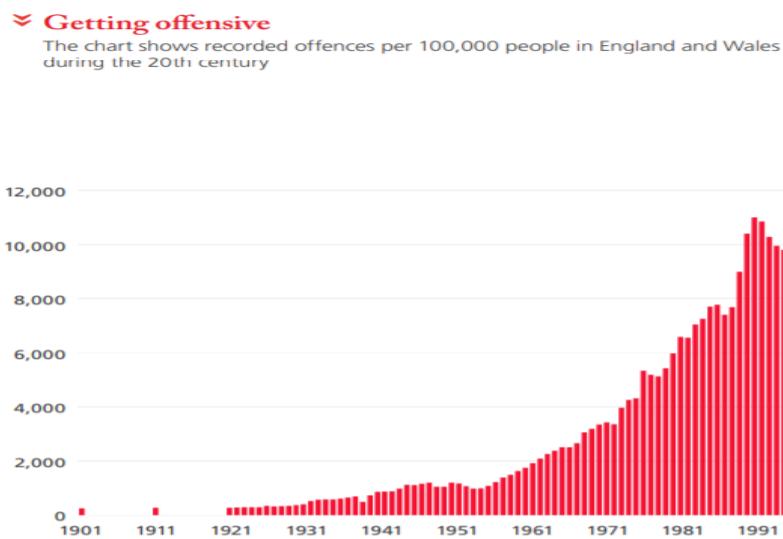


Figure 61 - Crime figures across the 20th century.

Whereas the figures in (Figure 61) show a broader range of offences, (Figure 62) focuses on homicides (Morgan et al., n.d.).

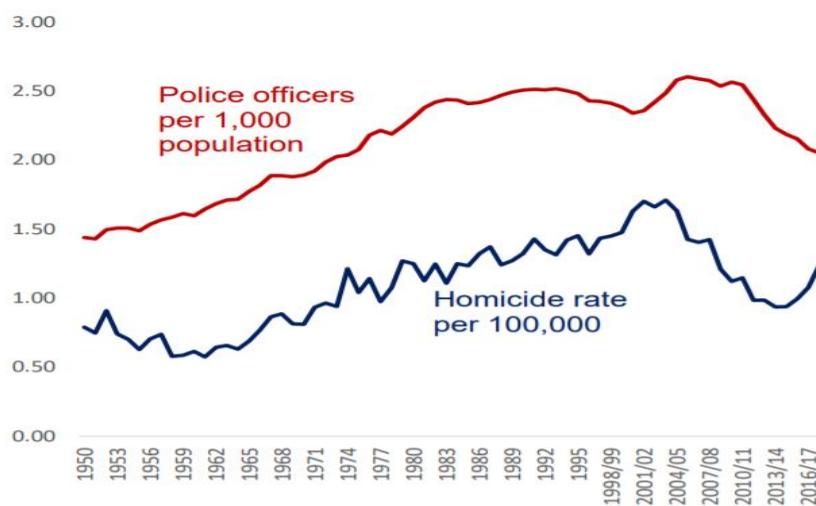


Figure 62 - Homicide rates, 1950s-2017.

Does the level of alcohol consumption across the decades tally with the sentiment of the music charts? It is something worth observing (Morgan et al., n.d.).

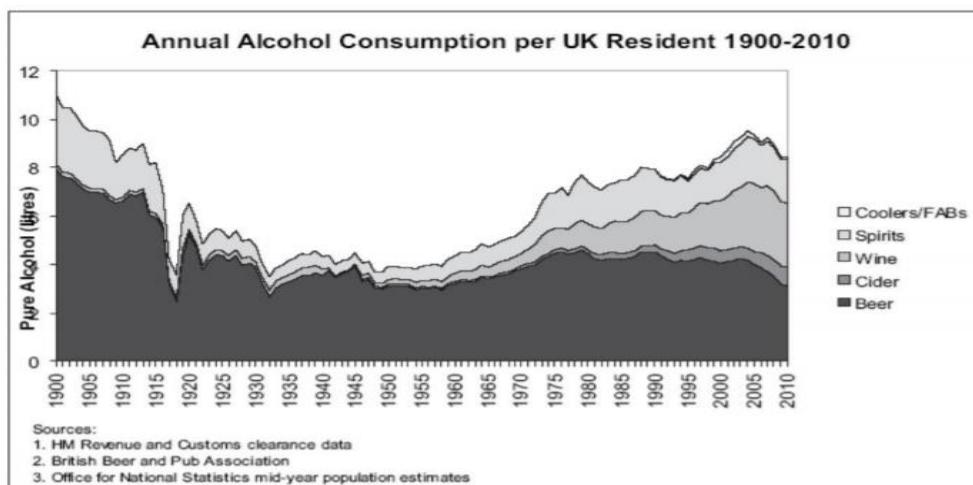


Figure 63 - Alcohol consumption rates 1900-2010.

Relationships are such a vital component to music lyrics, so I am interested to learn what impact marriage and divorce rates have on sentiment (Marriages in England and Wales - Office for National Statistics, 2021).

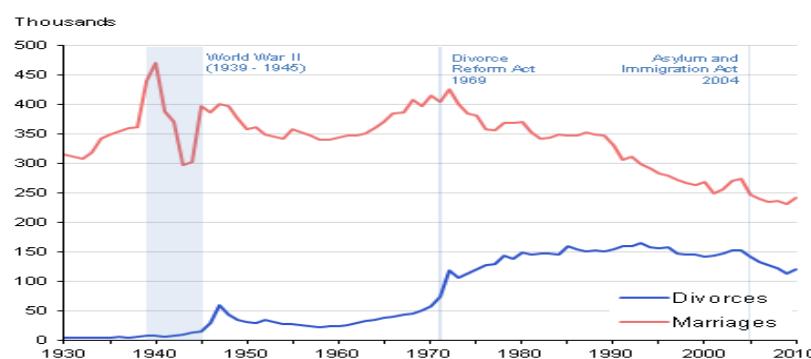


Figure 64 - Marriage and divorce rates 1930-2010.

Is there a link between the sex of the divorcee and number ones? (Figure 65) narrows the amount of years to give a better look (Divorce Rates Continue To Decline As Couples Look To Cohabitation, 2017).



Figure 65 - Closer look at the period of 1950-2014 for divorce rates.

5.7 Sprint 5 – Bonus data

Sprint ID	Description	Start Date	Finish Date
5	Bonus data	12/04/2021	02/05/2021

Task ID	Description	Status
1	Further analysis of the Power BI data	Completed

Task 1

Some of the Power BI data does not, necessarily, have as big an impact on the main research question as other information cited in earlier pages; but it is still of enough interest to cover. The figures referenced in relation to the subsequent text can be found in (Appendix E – Sentiment Analysis in Power BI).

- Average sentiment shown by nationality of artist(s) (Figure 133)
- Average sentiment of ‘Teen heartthrobs’ (Figure 152)
- Song sentiment polarity of ‘Teen heartthrobs’ (Figure 153) & (Figure 154)
- Average sentiment of ‘Girl groups’ (Figure 155)
- Song sentiment polarity of ‘Girl groups’ (Figure 156) & (Figure 157)
- Average sentiment of ‘Era-defining’ artists (Figure 158) & (Figure 159)
- Song sentiment polarity of ‘Era-defining’ artists (Figure 160) & (Figure 161)
- Average sentiment of male solo artists (Figure 162) & (Figure 163)
- Song sentiment polarity of male solo artists (Figure 164) & (Figure 165)
- Average sentiment of female solo artists (Figure 166)
- Song sentiment polarity of female solo artists (Figure 167) & (Figure 168)
- Average sentiment of groups (Figure 169)
- Song sentiment polarity of groups (Figure 170) & (Figure 171)
- Interactive tool allowing the user to search any number 1 they desire (Figure 181)

5.8 Sprint 6 –Answering research question, & preparing final submission

Sprint ID	Description	Start Date	Finish Date
6	Answering research question - Can sentiment analysis of UK number ones (1952-2020) detect correlation between lyrics and society?	03/05/2021	13/05/2021

Task ID	Description	Status
1	Answering thesis question	Completed
2	Presentation	Completed
3	Final submission	Completed

Task 1

“Can sentiment analysis of UK number ones (1952-2020) detect correlation between lyrics and society?”

This was my research question, and I hope to provide a convincing answer in (Chapter 7: Discussion & Conclusions).

Task 2

As part of the module, I was required to present my findings to a team of lecturers on May 13th, 2021. In the 15 minutes allocated, I spoke as confidently, and honestly, as was possible. I used the time to give a virtual PowerPoint presentation, outlining some key aspects of my project.

- Prototype
- Web scraping
- Data cleaning
- Running of UK dataset

To end, I pinpointed the tasks remaining to be completed before final submission. The full PowerPoint document can be viewed here [FYP PowerPoint presentation](#).

Task 3

The final task of all, submitting my project to my supervisor, Andrew Shields, and module facilitator, Helen Fitzgerald. The folder includes.

- Thesis (in both Word, and PDF, format)
- Development & Test files
- Installation instructions
- PowerPoint presentation
- Video presentation
- Poster presentation
- Declaration of original work

Chapter 6: Results Analysis of Data

6.1 Introduction

Previously in this thesis (**5.5 Sprint 3 – Running UK data and performing sentiment analysis** -Task 5), the field of average sentiment was explored in great detail. I will return to that in the discussion chapter, but for this results section, I will focus more on language in the lyrics.

6.2 Lyrics

When presenting my findings on the aforementioned average sentiment, I suggested that there was an argument to be made for 1994 being the most accurate candidate for the title of most positive year. But the two years following that year are noticeable for more objectionable milestones.

In 1995, “Don’t Stop”, by The Outhere Brothers became the first UK number one to contain the words, “Ass”, and “Pussy”. The sexist, and offensive, tone continued; Dick”; “Suck”; “Cum”; “Booty”; “Biatch”; “Motherfuckers”. The song ended up with, by far, the most negative polarity; almost three times less than the song closest to it. Curiously, I am not even certain that ‘biatch’ is classified as a negative word, due to its misspelling, so the sentiment score could have been worse!

1995 - Of songs containing the words, "Ass" and "Pussy", the first number 1 with a negative sentiment is "Don't Stop" by The Outhere Brothers



Figure 66 - "Don't Stop" by The Outhere Brothers, from 1995, was the first number one to contain both of the words, "Ass" and "Pussy".

1996 was first brought to attention during the sentiment analysis performance (**5.5 Sprint 3 – Running UK data and performing sentiment analysis** -Task 5) as it was noted for being ranked as the most negative of all 68 years in the study. I opined that this may be a false narrative due to the outliers involved, but that is not the only scenario which makes 1996 such an intriguing year to examine.

It was a breakthrough year for The Fugees who had two number ones, following up, "Killing Me Softly" with, "Ready or Not". The latter of which became the first UK number one to contain the word, "Nigga(s)".

1996 - Of songs containing the word, "nigga", the first number 1 with a POSITIVE sentiment value is "Ready or Not" by The Fugees

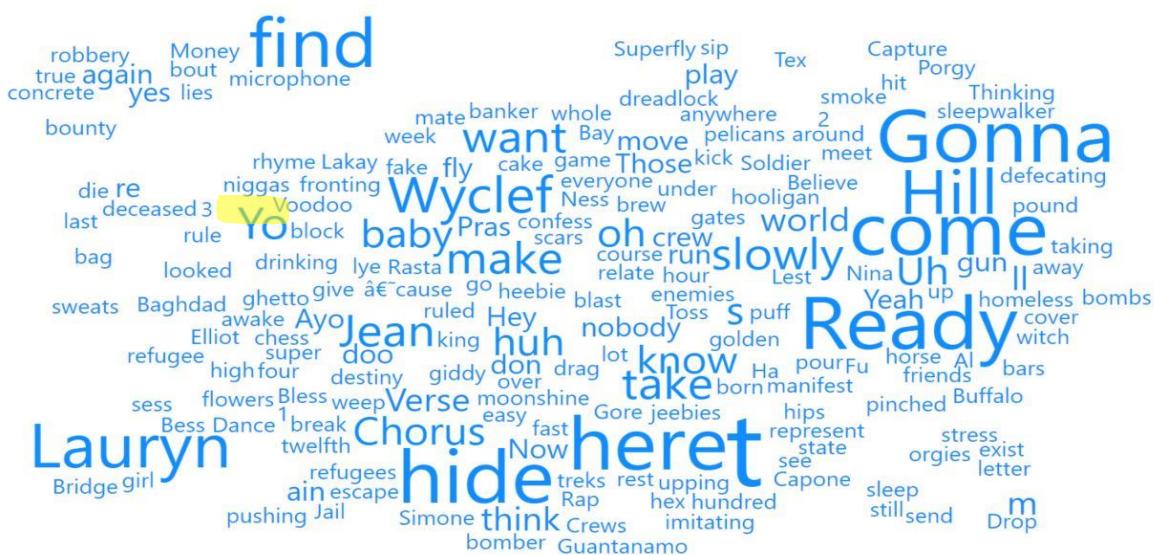


Figure 67 - 1996 included the first case of the word, "nigga", appearing in a number one of, surprisingly, high polarity.

However, the most interesting thing about this, for me, is that the song has a positive sentiment polarity. In fact, it is a median based song, as it shares the 0.14 rating which happens to be the total average for all 1248 number ones.

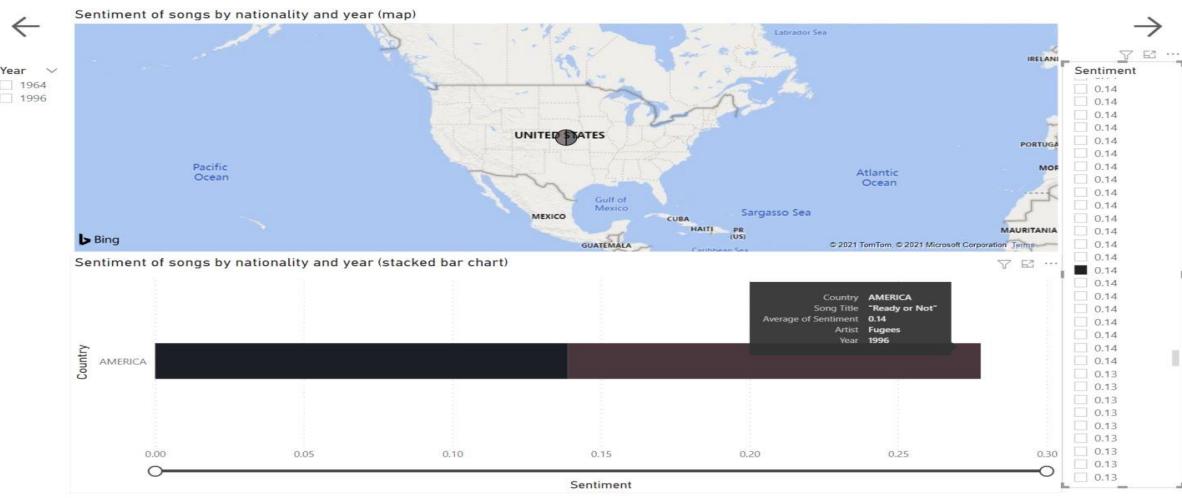


Figure 68 - Sentiment of songs by artist, year, and nationality. 1996 hit, "Ready or Not" receiving a sentiment rating of 0.14.

A decade later, 2006 produced the first number one containing “Nigga” to receive a negative polarity. The song was, “Nasty Girl”, by Notorious B.I.G., and looking at some of the other lyrics, the score would certainly appear to be justified.

- “Titties” (mentioned 10 times)
- “Butt”
- “Bitch”
- “Bitches”
- “Nasty”
- “Cum”
- “Fuckin”
- “Naked”
- “Grind”
- “Strip”
- “Swallow”
- “Bonin”
- “Sex”
- “Chick”
- “Squirt”
- “thong”

It should be noted that all of the words from ‘naked’ down would not necessarily be deemed as negative words, but this is where tools such as NLP and lemmatisation demonstrate their Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

strengths. Context is key, and the tools will observe the other words in “Nasty Girl” and judge the context in which the ambiguous words may be taken.

2006 - Of songs containing, "nigga", the first number 1 with a NEGATIVE sentiment is Notorious B.I.G with "Nasty Girl". 2006

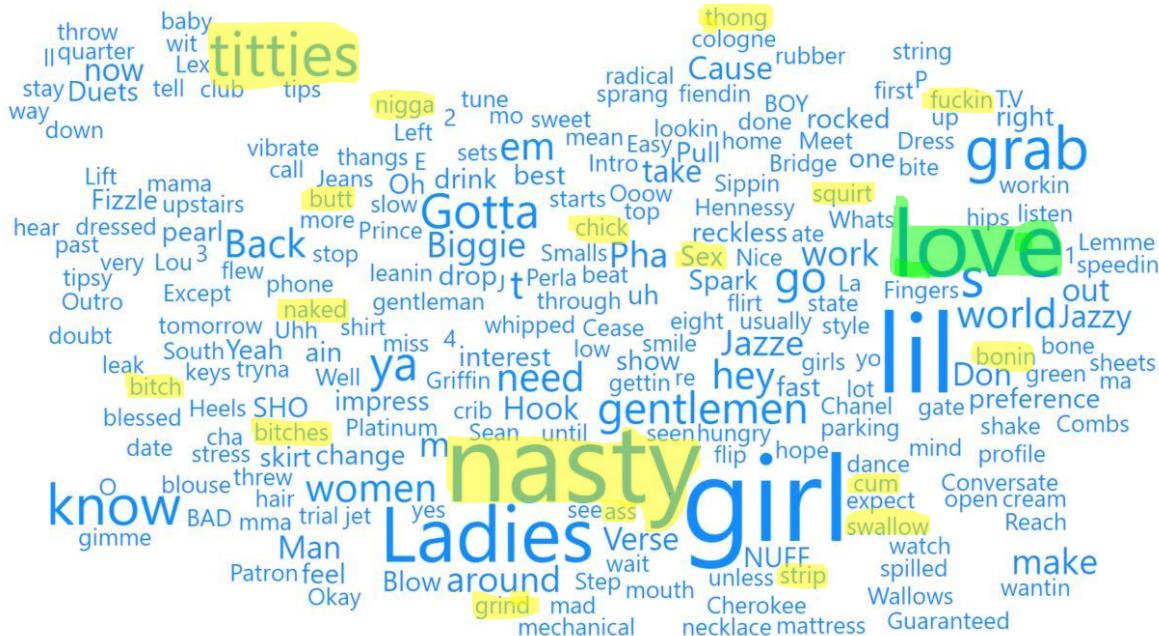


Figure 69 - "Nasty Girl", by Notorious B.I.G., became the first number one to use, "nigga", which had a negative polarity score.

Another hip-hop artist, Eminem, became the first to reach number one with both the words, “Shit” and “Fuck”. He did this in the year 2000 with not one, but two songs: “Stan” and “The Real Slim Shady”.

2000 - Of songs containing the words, "Shit" and "Fuck", the first case of both words appearing in a number 1 was in the year 2000. Eminem achieved this feat twice, with "Stan" and "The Real Slim Shady.



Figure 70 - Eminem's year 2000 included two number ones containing the words, "Shit" and "Fuck".

Unsurprisingly, there are many obscenities in the song. One word which I have also highlighted, and also in “Nasty Girl”, is ‘love’. It goes to show that there *are* positive elements to many of these songs too.

Speaking of love, only one song in the past decade (with an average sentiment of above 0.2) has included both “Love” and “Happy” in its lyrics; Gotye’s, “Somebody That I Used to Know”, from 2012.

2012 - Of songs containing the words, "Love" and "Happy", the only number 1 in the past decade with a sentiment of at least 0.2 is Gotye's "Somebody That I used to Know" from 2012



Figure 71 - Gotye is the highest scoring artist with a song containing, "Love" and "Happy" in the past decade.

This song went some way to helping 2012 becoming one of just two decades in this century, the other being 2008, to score a sentiment average of at least 0.18. Figure 72 - 2012 is also the only year of the past decade to reach an average sentiment of at least 0.18. illustrates the positivity of the 20th century in comparison.

2012 - The only year in the past decade with an average sentiment of at least 0.18 (*Word Cloud includes the other years which match this average)

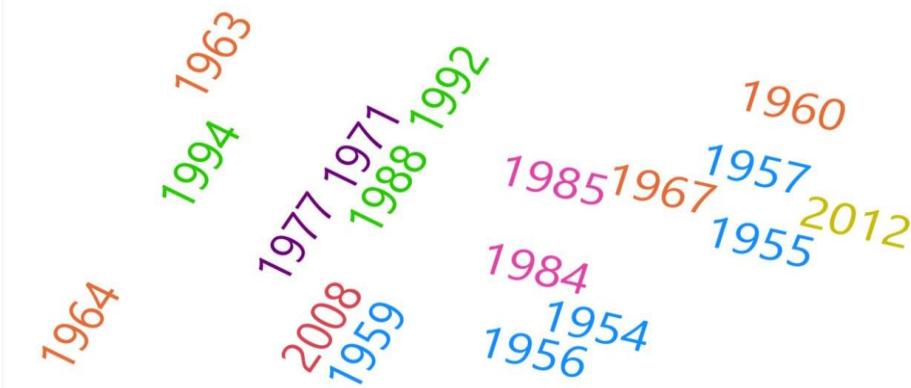


Figure 72 - 2012 is also the only year of the past decade to reach an average sentiment of at least 0.18.

Two of the years seen in Figure 72 - 2012 is also the only year of the past decade to reach an average sentiment of at least 0.18. are 1963, and 1964. These years, akin to 2012, also had

songs which contained the words, “Love” and “Happy”. Both were by The Beatles; 1963’s, “I Want to Hold Your Hand”, and “I Feel Fine”, from 1964.

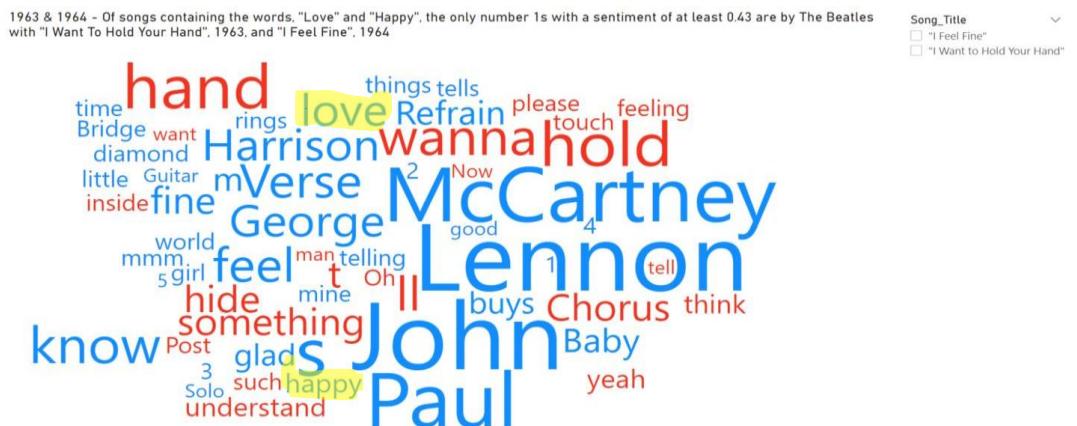
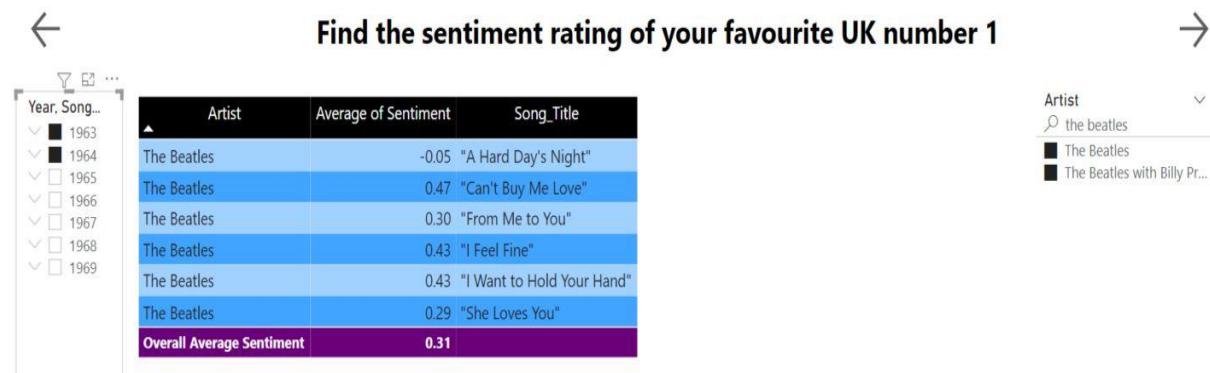


Figure 73 - The highest sentiment scoring songs containing the lyrics, 'Love' and 'Happy' both came from The Beatles in 1963, and 1964, respectively.

These two songs actually have a sentiment of almost double that of Gotye's 2012 number one. This does not come as a surprise when we take a more in-depth look at the two years as a whole for The Beatles. They scored five very positively rated songs, with just, "A Hard Day's Night", being an outlier for the period.

However, when we look at The Beatles' following five years, 1965-1969, the contrast is significant. Only one of the ten releases, "All You Need is Love", matches the positivity of the songs in the two earlier years (Figure 74 - Proof of how a sentiment score can be notably varied over time. The first two years of Beatles' number ones scored highly (a) which is in stark contrast to their number ones which followed over the next five years (b). A considerable dip in polarity.).



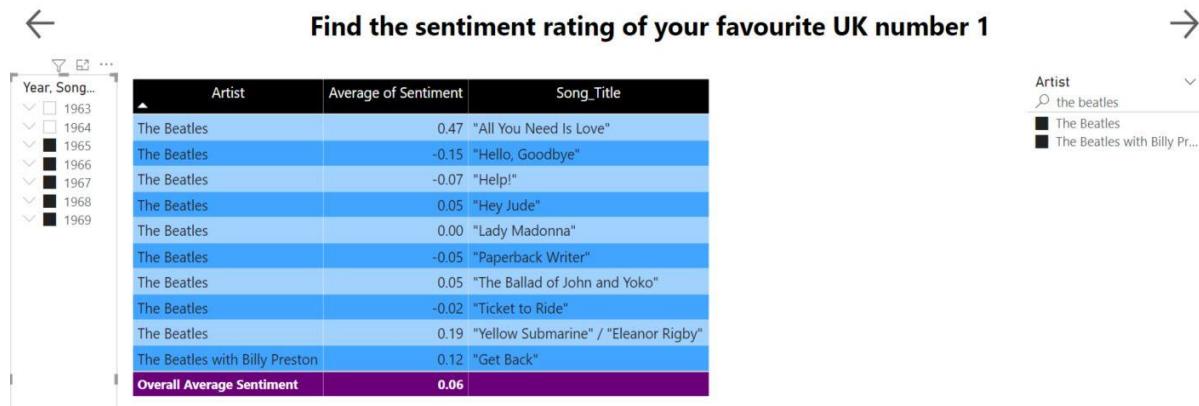


Figure 74 - Proof of how a sentiment score can be notably varied over time. The first two years of Beatles' number ones scored highly (a) which is in stark contrast to their number ones which followed over the next five years (b). A considerable dip in polarity.

At first, I was a tad over-excited by (Figure 75) because it just seemed too convenient that the average sentiment of UK artists was the only one which matched exactly with the average sentiment of all 50 nations combined. But when I consider the statistics, it is really not a surprise. The UK has 438 artists who have had an aggregate of 654 number ones between them. Considering that there is a total of 1248 number ones in the dataset, it means that 35.1% of UK artists have produced 52.4% of the number ones (LLC, 2021b). The UK is also just one position off the median in the chart, so it makes a lot of sense, although still rather quirky.

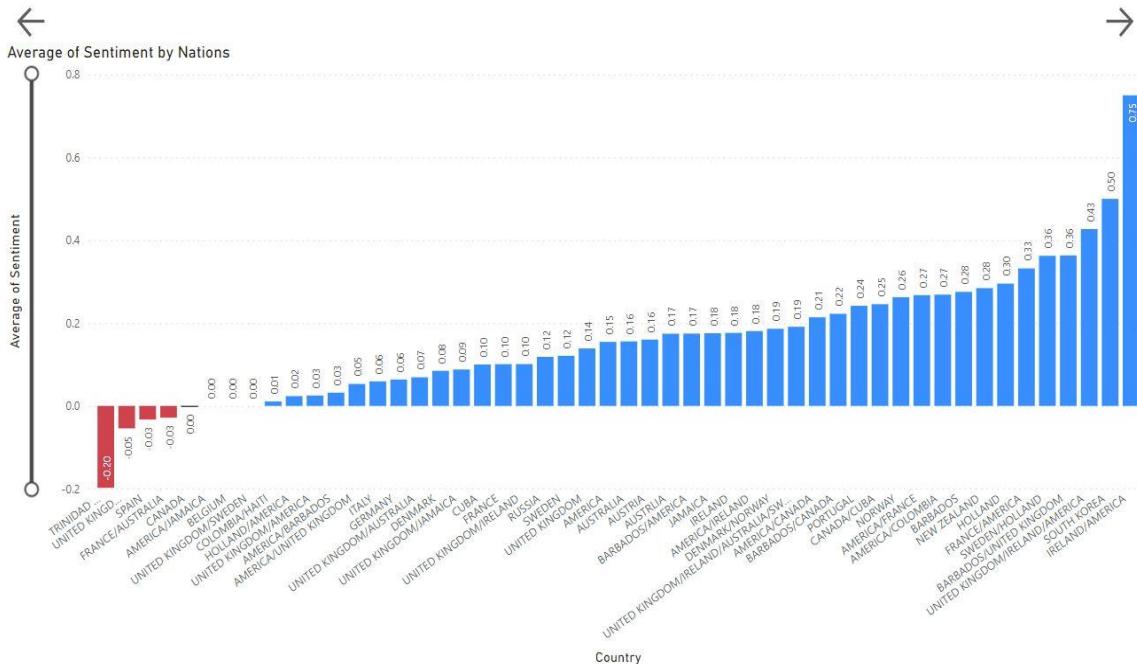


Figure 75 - Average sentiment by nationality of artists.

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

(Figure 76) is, in my opinion, irrefutable proof that the lyrical content of number ones has become increasingly negative over the past seven decades of the UK singles chart. On average, there is a decrease of 0.67% annually.

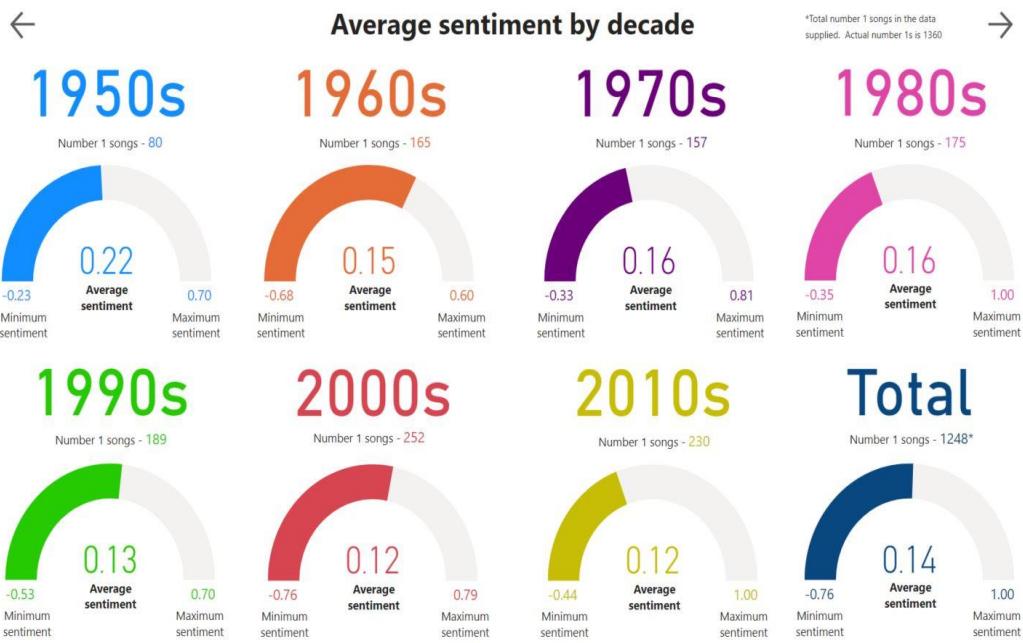


Figure 76 - The average sentiment for each decade of the UK charts.

Chapter 7: Discussion & Conclusions

7.1 Introduction

At the beginning of this project, I was, in all honesty, terrified that I would not possess the required skills and ability to even attempt it. There have been issues with code, barriers to finding data, and the everlasting battle with time which has cursed me through the past four years of my degree. But it has become a labour of love for me. In fact, it is now conceivable that I will carry on exploring this data until I finally obtain the answers to the many questions I now have in my head. I am so glad I took on the challenge.

7.2 Discussion

There are many areas of interest to discuss, and there are probably more questions than answers is what I will take away from this process.

To begin with, the ‘N’ word (It has been seen in full too many times for my liking). There could be an argument to suggest that the majority of the artists who use the ‘N’ word do not intend it as a racial slur. But I do feel that it is certainly implied in a derogatory sense. I am still a little baffled as to how The Fugees ended up with a polarity of 0.14 when containing that word. I *do* understand that it is only used once in the entire song, and so, the impact is not going to make a huge difference to the average sentiment, but it still does not quite sit right with me. It has also brought to my attention that I need to develop a better understanding of polarity by going away and doing far more research into how it is calculated, and which words score what?

On the topic of The Fugees, I would also take this opportunity to disclose that I believe that a number of the songs in this dataset have alternative lyrics for radio airplay. I began my deejaying career in 1996, the same year as “Ready or Not” was released, and I certainly do not remember hearing the ‘N’ word on Capital FM. I would go as far to say that I was such a radio addict back then, that I would guarantee there was a radio edit played instead. This also applies to “Don’t Stop” by The Outhere Brothers, and many others in the dataset too. This would suggest that the data could be skewed towards negative sentiment polarity, and I feel it is important to point it out.

Another issue which arose was the amount of erroneous lyrical data pulled in from Genius.com. I would have liked to have had more time to go through the data with a fine-

tooth comb in order to either fix or erase these concerns. It is definitely something to be wary of if attempting a project such as this again.

There is a tendency to, sometimes, present the negatives over the positives. It happens in all walks of life, but it is important to note that, although on the decrease, the overall average sentiment of UK number ones has remained at a consistently neutral level. The current polarity rating of 0.14 is very neutral in the grand scheme of things. Another point I would make is that, while the data does not lie, it can be witnessed through different perspectives. For example, yes, the average sentiment has evidently decreased between 1952-2020; but it also displayed a major decrease for The Beatles in the much smaller timeframe of 1964-1970.

7.3 Answering the Research Question

Can sentiment analysis of UK number ones (1952-2020) detect correlation between lyrics and society?

Although I detected some similarities between the first two decades of the charts and the corresponding GDP data, and also some vague parallels with divorce rates, I cannot honestly, and conclusively, determine that there is significant correlation between lyrical sentiment and societal trends.

As I stated in my introduction to this paper, music lyrics are powerful, and I truly believe that they have an amazing ability to make us feel happy, sad, peaceful, and angry. Not everybody likes football, or drinking. Some people like to read, some do not. But I honestly have never met one person who does not have a favourite song. I remember living in London during the World Cup in 1990, and Euro 1996. The atmosphere around the town was electric as people partied and celebrated. But the whole experience of both were immeasurably enhanced by New Order's, "World in Motion", and Skinner, Baddiel, & The Lightning Seeds singing, "Three Lions". So, I may not have the statistics to back up my hypothesis, but I lived through over half of the UK chart history and witnessed the correlation myself.

Without any doubt though, I will maintain that sentiment polarity of number one songs has decreased during the same period in which the level of offensive language such as profanities, misogyny, homophobia, and racism has increased.

References

- Anon 2008. Stereotyping musical genres. *PopMatters*. Available at: <<https://www.popmatters.com/stereotyping-musical-genres-2496131174.html>> [Accessed 10 Mar. 2021].
- Anon 2015. *How Hip-Hop Holds Blacks Back*. [online] City Journal. Available at: <<https://www.city-journal.org/html/how-hip-hop-holds-blacks-back-12442.html>> [Accessed 10 Mar. 2021].
- Anon 2017. *Charting the Rise of Song Collaborations with Scrapy and Pandas*. [online] dashee87.github.io. Available at: <<https://dashee87.github.io/data%20science/python/charting-the-rise-of-song-collaborations-with-scrapy-and-pandas/>> [Accessed 15 May 2021].
- Anon 2017. Divorce Rates Continue To Decline As Couples Look To Cohabitation. *Sinclair Law*. Available at: <<https://www.sinclairlaw.co.uk/blog/divorce-rates-continue-to-decline-as-couples-look-to-cohabitation/>> [Accessed 27 May 2021].
- Anon 2017. Removing stop words with NLTK in Python. *GeeksforGeeks*. Available at: <<https://www.geeksforgeeks.org/removing-stop-words-nltk-python/>> [Accessed 11 May 2021].
- Anon 2020. *Country music reckons with racial stereotypes and its future*. [online] PBS NewsHour. Available at: <<https://www.pbs.org/newshour/arts/country-music-reckons-with-racial-stereotypes-and-its-future>> [Accessed 10 Mar. 2021].
- Anon 2020. *Everything There Is to Know about Sentiment Analysis*. [online] MonkeyLearn. Available at: <<https://monkeylearn.com/sentiment-analysis/>> [Accessed 15 Dec. 2020].
- Anon 2020. Python - Lemmatization Approaches with Examples. *GeeksforGeeks*. Available at: <<https://www.geeksforgeeks.org/python-lemmatization-approaches-with-examples/>> [Accessed 14 May 2021].
- Anon 2020. *Singer and the Song / Learning English / BBC World Service*. [online] Available at:

<<https://www.bbc.co.uk/worldservice/learningenglish/music/singersong/noddyholder.shtml>>
[Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - the Sixties*. [online] Available at:
<<https://www.officialcharts.com/who-we-are/sixties-history/>> [Accessed 2 Dec. 2020].

Anon 2021. *All The Official Singles Chart Number 1s*. [online] Available at:
<https://www.officialcharts.com/chart-news/all-the-number-1-singles_7931/> [Accessed 22 May 2021].

Anon 2021. *AllMusic / Record Reviews, Streaming Songs, Genres & Bands*. [online]
AllMusic. Available at: <<https://www.allmusic.com/>> [Accessed 23 May 2021].

Anon 2021. *Alt shortcuts list*. [online] Available at:
<<https://fsymbols.com/keyboard/windows/alt-codes/list/shortcuts/>> [Accessed 23 May 2021].

Anon 2021. *Chart Archive / Haven*. [online] Available at:
<<https://fatherandy2.proboards.com/thread/14/chart-archive>> [Accessed 15 May 2021].

Anon 2021. *Chart Histories*. [online] UKMIX Forums. Available at:
<<https://www.ukmix.org/forum/chart-discussion/chart-analysis/12769-chart-histories>>
[Accessed 15 May 2021].

Anon 2021. *Crimes of the century*. [online] Available at:
<<https://www.parliament.uk/business/publications/research/olympic-britain/crime-and-defence/crimes-of-the-century/>> [Accessed 26 May 2021].

Anon 2021. *everyHit.com - UK Top 40 Chart Archive, British Singles & Album Charts*. [online] Available at: <<http://www.everyhit.com/>> [Accessed 15 May 2021].

Anon 2021. *Examples of Ambiguity*. [online] Available at:
<<https://examples.yourdictionary.com/reference/examples/examples-of-ambiguity.html>>
[Accessed 3 Feb. 2021].

Anon 2021. *Frank Sinatra – My Way*. Available at: <<https://genius.com/Frank-sinatra-my-way-lyrics>> [Accessed 26 May 2021].

Anon 2021. *Genius – How Genius Works*. [online] Genius. Available at: <<https://genius.com/Genius-how-genius-works-annotated>> [Accessed 11 May 2021].

Anon 2021. *Gensim: topic modelling for humans*. [online] Available at: <<https://radimrehurek.com/gensim/intro.html#gensim-generate-similar>> [Accessed 11 May 2021].

Anon 2021. *Gross Domestic Product: Year on Year growth: CVM SA % - Office for National Statistics*. [online] Available at: <<https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyp/pn2>> [Accessed 26 May 2021].

Anon 2021. *How It Works — lyricsgenius documentation*. [online] Available at: <https://lyricsgenius.readthedocs.io/en/master/how_it_works.html> [Accessed 11 May 2021].

Anon 2021. *Installing Packages — Python Packaging User Guide*. [online] Available at: <<https://packaging.python.org/tutorials/installing-packages/>> [Accessed 11 May 2021].

Anon 2021. *Investor Brief - Death And Taxes*. [online] Available at: <<http://www.habershamcapital.com/brief-no5-death-taxes>> [Accessed 27 May 2021].

Anon 2021. *Language Processing Pipelines · spaCy Usage Documentation*. [online] Available at: <<https://spacy.io/usage/processing-pipelines>> [Accessed 14 May 2021].

Anon 2021. *lyric*. [online] Available at: <<https://dictionary.cambridge.org/dictionary/english/lyric>> [Accessed 26 May 2021].

Anon 2021. *Lyrics containing the term: best — Page #9*. [online] Available at: <<https://www.lyrics.com/serp.php?st=best&p=9&genre=Pop>> [Accessed 15 May 2021].

Anon 2021. *Marriages in England and Wales - Office for National Statistics*. [online] Available at: <<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/marriagecohabitationandcivilpartnerships/bulletins/marriagesinenglandandwalesprovisional/2012-02-29>> [Accessed 27 May 2021].

Anon 2021. *Official Charts - Home of the Official UK Top 40 Charts*. [online] Available at: <<https://www.officialcharts.com/>> [Accessed 15 May 2021].

Anon 2021. *Pandas vs NumPy - javatpoint*. [online] www.javatpoint.com. Available at: <<https://www.javatpoint.com/pandas-vs-numpy>> [Accessed 11 May 2021].

Anon 2021. *polyhex / UK singles chartruns*. [online] Available at: <<https://polyhex.me.uk/uksingles/chartruns-uksingles.cfm>> [Accessed 15 May 2021].

Anon 2021. *Positive and Negative Connotations: Example Sentences*. [online] Available at: <<https://examples.yourdictionary.com/positive-and-negative-connotations-example-sentences.html>> [Accessed 3 Feb. 2021].

Anon 2021. *Public Spending Chart for United Kingdom 1950-2019 - Central Government Local Authorities*. [online] Available at: <https://www.ukpublicspending.co.uk/spending_chart_1950_2019UKp_17c1li0181536_722cs_F0t10t20t40t60t> [Accessed 26 May 2021].

Anon 2021. *QuickGraph #4: UK Official Singles Chart 2019 · Mark Needham*. [online] Mark Needham. Available at: <<https://www.markhneedham.com/blog/2020/01/04/quick-graph-uk-official-charts>> [Accessed 15 May 2021].

Anon 2021. *Scrap UK top 40 singles chart from BBC - DZone*. [online] dzone.com. Available at: <<https://dzone.com/articles/scrap-uk-top-40-singles-chart>> [Accessed 15 May 2021].

Anon 2021. *Sentiment Analysis in Python With TextBlob*. [online] Stack Abuse. Available at: <<https://stackabuse.com/sentiment-analysis-in-python-with-textblob>> [Accessed 11 May 2021].

Anon 2021. *spaCy 101: Everything you need to know · spaCy Usage Documentation*. [online] spaCy 101: Everything you need to know. Available at: <<https://spacy.io/usage/spacy-101>> [Accessed 11 May 2021].

Anon 2021. *spaCy Word Lemmatize - API & Demo / Text Analysis Online / TextAnalysis*. [online] Available at: <<http://textanalysisonline.com/spacy-word-lemmatize>> [Accessed 15 May 2021].

Anon 2021. *Text Analysis - API & Demo / Text Analysis Online / TextAnalysis*. [online] Available at: <<http://textanalysisonline.com/>> [Accessed 14 May 2021].

Anon 2021. *TextBlob Sentiment Analysis - API & Demo / Text Analysis Online / TextAnalysis*. [online] Available at: <<http://textanalysisonline.com/textblob-sentiment-analysis>> [Accessed 14 May 2021].

Anon 2021. *UK Healthcare Spending Since World War II for United Kingdom 1952-2020 - Central Government Local Authorities*. [online] Available at: <https://www.ukpublicspending.co.uk/spending_chart_1952_2020UKp_17c1li111lcn_10t_UK_Healthcare_Spending_Since_World_War_II> [Accessed 26 May 2021].

Anon 2021. *United Kingdom Long Term Unemployment Rate / 1992-2020 Data / 2021-2023 Forecast*. [online] Available at: <<https://tradingeconomics.com/united-kingdom/long-term-unemployment-rate>> [Accessed 26 May 2021].

Anon 2021. *Urban Dictionary: metalhead*. [online] Urban Dictionary. Available at: <<https://www.urbandictionary.com/define.php?term=metalhead&page=21>> [Accessed 10 Mar. 2021].

Anon 2021. *What is Beautiful Soup?* [online] Educative: Interactive Courses for Software Developers. Available at: <<https://www.educative.io/edpresso/what-is-beautiful-soup>> [Accessed 15 May 2021].

Anon 2021. *What is the MoSCoW Method?* [online] SearchSoftwareQuality. Available at: <<https://searchsoftwarequality.techtarget.com/definition/MoSCoW-method>> [Accessed 13 Apr. 2021].

Anon 2021. *What's with Gay Men and Musical Theater? | by James Finn | Th-Ink Queerly / Medium*. [online] Available at: <<https://medium.com/th-ink/whats-with-gay-men-and-musical-theater-a8bc6f71a739>> [Accessed 10 Mar. 2021].

Anon 2021. *Who We Are - Our charts & data*. [online] Available at: <<https://www.officialcharts.com/who-we-are/our-charts-and-data/>> [Accessed 15 May 2021].

Anon 2021. *Why do girls like listening to love songs?* - Quora. [online] Available at: <<https://www.quora.com/Why-do-girls-like-listening-to-love-songs>> [Accessed 10 Mar. 2021].

Anon 2021. *word.* [online] Available at: <<https://dictionary.cambridge.org/dictionary/english/word>> [Accessed 26 May 2021].

Anon n.d. Diagnosing Teenager's Problems: Listen to Their Music. *Understanding Teenagers.* Available at: <<https://understandingteenagers.com.au/diagnosing-teenagers-problems-listen-to-their-music/>> [Accessed 10 Mar. 2021].

Anon n.d. *salimzubair/lyric-sentiment: Sentiment analysis of all songs that entered the Billboard Hot 100 charts from 1958-2019.* Available at: <<https://github.com/salimzubair/lyric-sentiment>>.

Anon n.d. Sentiment Analysis: Types, Tools, and Use Cases. *AltexSoft.* Available at: <<https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/>> [Accessed 16 Dec. 2020].

Baker, F. and Bor, W., 2008. Can Music Preference Indicate Mental Health Status in Young People? *Australasian Psychiatry*, 16(4), pp.284–288.
<https://doi.org/10.1080/10398560701879589>.

Bannister, Kristian, 2020. *Sentiment Analysis: How Does It Work? Why Should We Use It?* [online] Brandwatch. Available at: <https://www.brandwatch.com/blog/understanding-sentiment-analysis> [Accessed 16 Dec. 2020].

Bates, V.C., Talbot, B.C. and Tobias, E.S., 2014. *A refereed journal of the Action for Change in Music Education Flipping the Misogynist Script: Gender, Agency, Hip Hop and Music Education. Criticism & Theory for Music Education ISN 1*, pp.4–5.

Chakraborty, A., Durgesh, K., Upadhyay, M., Gandhi, K., Vidyapith, V. and Agrawal, M., 2018. *Young Adults' Music Preferences and its Relation to their Attitude towards Women and Sexuality Convener of 3rd International Conference on Human Behaviour and Development Issues with theme "Growth and Sustainability View project A Critical Analysis of Psy.* [online] Available at: <www.printspublications.com>.

Christenson, P.G., de Haan-Rietdijk, S., Roberts, D.F. and ter Bogt, T.F.M., 2019. What has America been singing about? Trends in themes in the U.S. top-40 songs: 1960–2010. *Psychology of Music*, 47(2), pp.194–212. <https://doi.org/10.1177/0305735617748205>.

Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J.M., Castillo, J.C., Lozano-Monasor, E., López, M.T., Zangróniz, R., Latorre, J.M. and Fernández-Sotos, A., 2016. Smart environment architecture for emotion detection and regulation. *Journal of Biomedical Informatics*, 64, pp.55–73. <https://doi.org/10.1016/j.jbi.2016.09.015>.

Forums, B.E., 2021. *Music forums focusing on chart music with chart and entertainment discussion*. [music,charts,lyrics,chart music,midweeks,albums,singles,bands,downloads,music charts,entertainment discussion,music forums,music talk] Available at: <<http://www.buzzjack.com/forums/UK-Charts-f2.html>> [Accessed 15 May 2021].

Fried, C.B., 2003. *Stereotypes of Music 1 Running head: STEREOTYPES OF MUSIC FANS Stereotypes of Music Fans: Are Rap and Heavy Metal Fans a Danger to Themselves or Others? Journal of Media Psychology*, .

GayBoiBlue, 2019. Stereotypical “Gay Music”. [Reddit Post] *r/gaybros*. Available at: <www.reddit.com/r/gaybros/comments/b8ayzi/stereotypical_gay_music/> [Accessed 10 Mar. 2021].

Goyal, Kechit, 2020. Top 4 Types of Sentiment Analysis & Where to Use. *upGrad blog*. Available at: <<https://www.upgrad.com/blog/types-of-sentiment-analysis/>> [Accessed 16 Dec. 2020].

H, S. and el, 2011. Classification of Emotions. *The Emotion Machine*. Available at: <www.theemotionmachine.com/classification-of-emotions/> [Accessed 18 Mar. 2021].

Hayes, J., 2021. *The top 10 this week in 1973: Battle of the giants in a classic Christmas chart*. [online] independent. Available at: <www.independent.ie/regionals/corkman/lifestyle/the-top-10-this-week-in-1973-battle-of-the-giants-in-a-classic-christmas-chart-39875113.html> [Accessed 16 May 2021].

Kuckartz, U., 2019. Qualitative Inhaltsanalyse: von Kracauers Anfängen zu heutigen Herausforderungen. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 20, p.No 3 (2019): Qualitative Content Analysis I.
<https://doi.org/10.17169/FQS-20.3.3370>.

LLC, C., 2021a. *Mean, Median, Mode Calculator*. [online] CalculatorSoup. Available at: <<https://www.calculatorsoup.com/calculators/statistics/mean-median-mode.php>> [Accessed 25 May 2021].

LLC, C., 2021b. *Percentage Calculator*. [online] CalculatorSoup. Available at: <<https://www.calculatorsoup.com/calculators/math/percentage.php>> [Accessed 27 May 2021].

Mäntylä, M.V., Graziotin, D. and Kuutila, M., 2018. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, pp.16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>.

mihart, 2021a. *Export data from a Power BI visualization - Power BI*. [online] Available at: <<https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-export-data>> [Accessed 22 May 2021].

mihart, 2021b. *What is Power BI? - Power BI*. [online] Available at: <<https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>> [Accessed 17 May 2021].

Morgan, N., Shaw, O., Mailley, J., Channing, R., Sweiry, A., Kruithof, K., Bloom, J., Kent, A., Smith, K. and Romaniuk, A., n.d. Trends and drivers of homicide: Main findings. p.80.

Osborne, C., 2017. *Analysing UK Chart history — 1956 to 2017*. [online] Medium. Available at: <<https://medium.com/@caineosborne/analysing-uk-chart-history-1956-to-2017-6fec0ecc991b>> [Accessed 15 May 2021].

Pandey, A., Rajpoot, D. and Saraswat, M., 2017. Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53, pp.764–779.
<https://doi.org/10.1016/j.ipm.2017.02.004>.

Parkinson, Justin, 2016. The significance of Sarah Baartman. *BBC News*. [online] 7 Jan. Available at: <<https://www.bbc.com/news/magazine-35240987>> [Accessed 1 Dec. 2020].

Pettijohn, T.F. and Sacco, D.F., 2009. The language of lyrics: An analysis of popular Billboard songs across conditions of social and economic threat. *Journal of Language and Social Psychology*, 28(3), pp.297–311. <https://doi.org/10.1177/0261927X09335259>.

Prabhakaran, S., 2018. Lemmatization Approaches with Examples in Python. *ML+*. Available at: <<https://www.machinelearningplus.com/nlp/lemmatization-examples-python/>> [Accessed 14 May 2021].

Prasad, B.D., 2008. *Content Analysis A method in Social Science Research*.

Rentfrow, P., McDonald, J. and Oldmeadow, J., 2009. You Are What You Listen To: Young People's Stereotypes about Music Fans. *Group Processes & Intergroup Relations - GROUP PROCESS INTERGROUP RELA*, 12, pp.329–344.
<https://doi.org/10.1177/1368430209102845>.

Schoell, E.D., 2017. *How do I use the slice notation in Python?* [online] O'Reilly Media. Available at: <<https://www.oreilly.com/content/how-do-i-use-the-slice-notation-in-python/>> [Accessed 15 May 2021].

Sharma, V.B., n.d. *rake-nltk: Python implementation of the Rapid Automatic Keyword Extraction algorithm using NLTK*. [Python] Available at: <<https://github.com/csurfer/rake-nltk>> [Accessed 11 May 2021].

Shelley, M. and Krippendorff, K., 1984. Content Analysis: An Introduction to its Methodology. *Journal of the American Statistical Association*, 79(385), p.240.
<https://doi.org/10.2307/2288384>.

Songfacts, 2021. *Lyrics for What A Wonderful World by Louis Armstrong - Songfacts*. [online] Available at: <<https://www.songfacts.com/lyrics/louis-armstrong/what-a-wonderful-world>> [Accessed 14 May 2021].

Williams, M., 2002. *Obituary: Percy Dickins*. [online] the Guardian. Available at: <<http://www.theguardian.com/news/2002/feb/19/guardianobituaries1>> [Accessed 2 Dec. 2020].

Works Cited

Bibliography

- Adriaanse, L. and Rensleigh, C., 2011. Comparing Web of Science, Scopus and Google Scholar from an Environmental Sciences perspective. *South African Journal of Library & Information Science*, 77, pp.169–178. <https://doi.org/10.7553/77-2-58>.
- Agbo-Quaye, S. and Robertson, T., 2010. The motorway to adulthood: Music preference as the sex and relationships roadmap. *Sex Education*, 10(4), pp.359–371. <https://doi.org/10.1080/14681811.2010.515094>.
- Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. and Ortega, F., 2016. A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*, 236(1), pp.197–213. <https://doi.org/10.1007/s10479-013-1449-6>.
- Aljanaki, A., Yang, Y.-H. and Soleymani, M., 2017. Developing a benchmark for emotional analysis of music. *PLOS ONE*, 12(3), pp.e0173392–e0173392. <https://doi.org/10.1371/journal.pone.0173392>.
- Almatarneh, S. and Gamallo, P., 2018. A lexicon based method to search for extreme opinions. *PLoS ONE*, 13(5), pp.1–19. <https://doi.org/10.1371/journal.pone.0197816>.
- Anglada-Tort, M., Krause, A.E. and North, A.C., 2019. Popular music lyrics and musicians' gender over time: A computational approach. *Psychology of Music*. <https://doi.org/10.1177/0305735619871602>.
- Anon 1937. Foreword. *The Public Opinion Quarterly*, 1(1), pp.3–5.
- Anon 2008. Stereotyping musical genres. *PopMatters*. Available at: <<https://www.popmatters.com/stereotyping-musical-genres-2496131174.html>> [Accessed 10 Mar. 2021].
- Anon 2015. *How Hip-Hop Holds Blacks Back*. [online] City Journal. Available at: <<https://www.city-journal.org/html/how-hip-hop-holds-blacks-back-12442.html>> [Accessed 10 Mar. 2021].

- Anon 2017. *Charting the Rise of Song Collaborations with Scrapy and Pandas*. [online] dashee87.github.io. Available at: <<https://dashee87.github.io/data%20science/python/charting-the-rise-of-song-collaborations-with-scrapy-and-pandas/>> [Accessed 15 May 2021].
- Anon 2017. Divorce numbers for opposite-sex couples highest since 2009. *BBC News*. [online] 18 Oct. Available at: <<https://www.bbc.com/news/uk-41669400>> [Accessed 26 May 2021].
- Anon 2017. Divorce Rates Continue To Decline As Couples Look To Cohabitation. *Sinclair Law*. Available at: <<https://www.sinclairlaw.co.uk/blog/divorce-rates-continue-to-decline-as-couples-look-to-cohabitation/>> [Accessed 27 May 2021].
- Anon 2017. Removing stop words with NLTK in Python. *GeeksforGeeks*. Available at: <<https://www.geeksforgeeks.org/removing-stop-words-nltk-python/>> [Accessed 11 May 2021].
- Anon 2018. Introduction to sentiment analysis: What is sentiment analysis? *Algorithmia Blog*. Available at: <<https://algorithmia.com/blog/introduction-sentiment-analysis>> [Accessed 15 Dec. 2020].
- Anon 2018. Lasswell's Communication Model. *Businessstopia*. Available at: <<https://www.businessstopia.net/communication/lasswell-communication-model>> [Accessed 6 Apr. 2021].
- Anon 2019. Sentiment Analysis Using Python: A Hands-on Guide. *upGrad blog*. Available at: <<https://www.upgrad.com/blog/sentiment-analysis-using-python/>> [Accessed 16 Dec. 2020].
- Anon 2020. *All you need is love: Metaphors of love in 1946–2016 Billboard year-end number-one songs in: Text & Talk - Ahead of print*. [online] Available at: <<https://www.degruyter.com/view/journals/text/ahead-of-print/article-10.1515-text-2019-0209/article-10.1515-text-2019-0209.xml>> [Accessed 25 Nov. 2020].
- Anon 2020. *Beautiful Soup Documentation — Beautiful Soup 4.9.0 documentation*. [online] Available at: <<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>> [Accessed 25

Nov. 2020].

Anon 2020. *Country music reckons with racial stereotypes and its future*. [online] PBS NewsHour. Available at: <<https://www.pbs.org/newshour/arts/country-music-reckons-with-racial-stereotypes-and-its-future>> [Accessed 10 Mar. 2021].

Anon 2020. *Digging Your Scene Lyrics*. [online] Available at: <<https://www.lyrics.com/lyric/14866167/The+Blow+Monkeys/Digging+Your+Scene>> [Accessed 2 Dec. 2020].

Anon 2020. *Everything There Is to Know about Sentiment Analysis*. [online] MonkeyLearn. Available at: <<https://monkeylearn.com/sentiment-analysis/>> [Accessed 15 Dec. 2020].

Anon 2020. *Examining Billboard Hot 100 Lyrics from 1987 – 2016 / Data Science Blog*. [online] Available at: <<https://nycdatascience.com/blog/student-works/billboard-hot-100-lyrics-1987-2017/>> [Accessed 25 Nov. 2020].

Anon 2020. *How @Chartdata Became One of Music's Most Important Twitter Accounts*. [online] Complex. Available at: <<https://www.complex.com/music/2020/07/chartdata-interview>> [Accessed 11 Nov. 2020].

Anon 2020. Python - Lemmatization Approaches with Examples. *GeeksforGeeks*. Available at: <<https://www.geeksforgeeks.org/python-lemmatization-approaches-with-examples/>> [Accessed 14 May 2021].

Anon 2020. *Sentiment Analysis / Lexalytics*. [online] Available at: <<https://www.lexalytics.com/technology/sentiment-analysis#machine-learning>> [Accessed 16 Dec. 2020].

Anon 2020. *Sentiment Analysis with Python / End Point*. [online] Available at: <<https://www.endpoint.com/blog/2018/05/18/sentiment-analysis-with-python>> [Accessed 15 Dec. 2020].

Anon 2020a. *Singer and the Song / Learning English / BBC World Service*. [online] Available at: <<https://www.bbc.co.uk/worldservice/learningenglish/music/singersong/noddyholder.shtml>> [Accessed 2 Dec. 2020].

Anon 2020b. *Singer and the Song / Learning English / BBC World Service*. [online]

Available at:

<<https://www.bbc.co.uk/worldservice/learningenglish/music/singersong/jimmysomerville.shtml>> [Accessed 2 Dec. 2020].

Anon 2020c. *Singer and the Song / Learning English / BBC World Service*. [online]

Available at:

<<https://www.bbc.co.uk/worldservice/learningenglish/music/singersong/drrobert.shtml>> [Accessed 2 Dec. 2020].

Anon 2020d. *Singer and the Song / Learning English / BBC World Service*. [online]

Available at:

<<https://www.bbc.co.uk/worldservice/learningenglish/music/singersong/karlhyde.shtml>> [Accessed 2 Dec. 2020].

Anon 2020. *Tweets Dataset for Detection of Cyber-Trolls*. [online] Available at:

<<https://kaggle.com/dataturks/dataset-for-detection-of-cyber-trolls>> [Accessed 25 Nov. 2020].

Anon 2020. *Who We Are - History of the Official Charts*. [online] Available at:

<<https://www.officialcharts.com/who-we-are/history-of-the-official-charts>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - The Eighties*. [online] Available at:

<<https://www.officialcharts.com/who-we-are/eighties>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - the Fifties*. [online] Available at:

<<https://www.officialcharts.com/who-we-are/fifties-history>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - The Nineties*. [online] Available at:

<<https://www.officialcharts.com/who-we-are/nineties>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - The Noughties*. [online] Available at:

<<https://www.officialcharts.com/who-we-are/noughties>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - the Seventies*. [online] Available at:

<<https://www.officialcharts.com/who-we-are/seventies-history>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - the Sixties*. [online] Available at: <<https://www.officialcharts.com/who-we-are/sixties-history/>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - History of the Official Charts - The Teens*. [online] Available at: <<https://www.officialcharts.com/who-we-are/teens/>> [Accessed 2 Dec. 2020].

Anon 2020. *Who We Are - How the Official Charts are compiled*. [online] Available at: <<https://www.officialcharts.com/getting-into-the-charts/how-the-charts-are-compiled/>> [Accessed 2 Dec. 2020].

Anon 2021. (4) *What is difference between stemming and lemmatization? - Quora*. [online] Available at: <<https://www.quora.com/What-is-difference-between-stemming-and-lemmatization>> [Accessed 14 May 2021].

Anon 2021. *About*. [online] Oxford Academic. Available at: <<https://academic.oup.com/poq/pages/About>> [Accessed 6 Apr. 2021].

Anon 2021. *All The Official Singles Chart Number 1s*. [online] Available at: <https://www.officialcharts.com/chart-news/all-the-number-1-singles_7931> [Accessed 22 May 2021].

Anon 2021. *AllMusic / Record Reviews, Streaming Songs, Genres & Bands*. [online] AllMusic. Available at: <<https://www.allmusic.com/>> [Accessed 23 May 2021].

Anon 2021. *Alt shortcuts list*. [online] Available at: <<https://fsymbols.com/keyboard/windows/alt-codes/list/shortcuts/>> [Accessed 23 May 2021].

Anon 2021. *Chart Archive / Haven*. [online] Available at: <<https://fatherandy2.proboards.com/thread/14/chart-archive>> [Accessed 15 May 2021].

Anon 2021. *Chart Histories*. [online] UKMIX Forums. Available at: <<https://www.ukmix.org/forum/chart-discussion/chart-analysis/12769-chart-histories>> [Accessed 15 May 2021].

Anon 2021. *Crimes of the century*. [online] Available at: <<https://www.parliament.uk/business/publications/research/olympic-britain/crime-and-defence/crimes-of-the-century/>> [Accessed 26 May 2021].

Anon 2021. *Daniel Droba: Methods Used for Measuring Public Opinion*. [online] Available at: <https://brocku.ca/MeadProject/Droba/Droba_1931.html> [Accessed 25 Mar. 2021].

Anon 2021. *Demo - Snowball*. [online] Available at: <<https://snowballstem.org/demo.html>> [Accessed 14 May 2021].

Anon 2021. *everyHit.com - UK Top 40 Chart Archive, British Singles & Album Charts*. [online] Available at: <<http://www.everyhit.com/>> [Accessed 15 May 2021].

Anon 2021. *Examples of Ambiguity*. [online] Available at: <<https://examples.yourdictionary.com/reference/examples/examples-of-ambiguity.html>> [Accessed 3 Feb. 2021].

Anon 2021. *Frank Sinatra – My Way*. Available at: <<https://genius.com/Frank-sinatra-my-way-lyrics>> [Accessed 26 May 2021].

Anon 2021. *Genius – How Genius Works*. [online] Genius. Available at: <<https://genius.com/Genius-how-genius-works-annotated>> [Accessed 11 May 2021].

Anon 2021. *Gensim: topic modelling for humans*. [online] Available at: <<https://radimrehurek.com/gensim/intro.html#gensim-generate-similar>> [Accessed 11 May 2021].

Anon 2021. *Google Trends*. [online] Google Trends. Available at: <<https://trends.google.com/trends/explore?cat=14&date=all&q=%2Fm%2F0g57xn>> [Accessed 20 Mar. 2021].

Anon 2021. *Gross Domestic Product: Year on Year growth: CVM SA % - Office for National Statistics*. [online] Available at: <<https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyp/pn2>> [Accessed 26 May 2021].

Anon 2021. *How does sentiment analysis work, generally? - Quora*. [online] Available at: <<https://www.quora.com/How-does-sentiment-analysis-work-generally>> [Accessed 20 Mar. 2021].

Anon 2021. *How It Works — lyricsgenius documentation*. [online] Available at:

<https://lyricsgenius.readthedocs.io/en/master/how_it_works.html> [Accessed 11 May 2021].

Anon 2021. *I am a beginner in this field and want to do sentiment analysis on a given data set. Where is a step-by-step tutorial for the same?* - Quora. [online] Available at: <<https://www.quora.com/I-am-a-beginner-in-this-field-and-want-to-do-sentiment-analysis-on-a-given-data-set-Where-is-a-step-by-step-tutorial-for-the-same>> [Accessed 20 Mar. 2021].

Anon 2021. *Infographic: Population growth in the UK.* [online] Statista Infographics. Available at: <<https://www.statista.com/chart/18515/population-growth-in-the-uk/>> [Accessed 27 May 2021].

Anon 2021. *Installing Packages — Python Packaging User Guide.* [online] Available at: <<https://packaging.python.org/tutorials/installing-packages/>> [Accessed 11 May 2021].

Anon 2021. *Investor Brief - Death And Taxes.* [online] Available at: <<http://www.habershamcapital.com/brief-no5-death-taxes>> [Accessed 27 May 2021].

Anon 2021a. *Language Processing Pipelines · spaCy Usage Documentation.* [online] Language Processing Pipelines. Available at: <<https://spacy.io/usage/processing-pipelines>> [Accessed 14 May 2021].

Anon 2021b. *Language Processing Pipelines · spaCy Usage Documentation.* [online] Available at: <<https://spacy.io/usage/processing-pipelines>> [Accessed 14 May 2021].

Anon 2021c. *Language Processing Pipelines · spaCy Usage Documentation.* [online] Available at: <<https://spacy.io/usage/processing-pipelines>> [Accessed 14 May 2021].

Anon 2021. *lyric.* [online] Available at:

<<https://dictionary.cambridge.org/dictionary/english/lyric>> [Accessed 26 May 2021].

Anon 2021. *Lyrics containing the term: best — Page #9.* [online] Available at: <<https://www.lyrics.com/serp.php?st=best&p=9&genre=Pop>> [Accessed 15 May 2021].

Anon 2021. *LyricsGenius: a Python client for the Genius.com API — lyricsgenius documentation.* [online] Available at: <<https://lyricsgenius.readthedocs.io/en/master/>> [Accessed 11 May 2021].

Anon 2021. *Marriages in England and Wales - Office for National Statistics*. [online]

Available at:

<<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/marriag ecohabitationandcivilpartnerships/bulletins/marriagesinenglandandwalesprovisional/2012-02-29>> [Accessed 27 May 2021].

Anon 2021. *NLTK Lancaster Stemmer - API & Demo / Text Analysis Online / TextAnalysis*.

[online] Available at: <<http://textanalysisonline.com/nltk-lancaster-stemmer>> [Accessed 14 May 2021].

Anon 2021. *Official Charts - Home of the Official UK Top 40 Charts*. [online] Available at:

<<https://www.officialcharts.com/>> [Accessed 15 May 2021].

Anon 2021. *Pandas vs NumPy - javatpoint*. [online] www.javatpoint.com. Available at:

<<https://www.javatpoint.com/pandas-vs-numpy>> [Accessed 11 May 2021].

Anon 2021. *polyhex / UK singles chartruns*. [online] Available at:

<<https://polyhex.me.uk/uksingles/chartruns-uksingles.cfm>> [Accessed 15 May 2021].

Anon 2021. *Porter Stemmer Online*. [online] Available at:

<https://9ol.es/porter_js_demo.html> [Accessed 14 May 2021].

Anon 2021. *Positive and Negative Connotations: Example Sentences*. [online] Available at:

<<https://examples.yourdictionary.com/positive-and-negative-connotations-example-sentences.html>> [Accessed 3 Feb. 2021].

Anon 2021. *Public Revenue Chart for United Kingdom 1950-2020 - Central Government Local Authorities*. [online] Available at:

<https://www.ukpublicrevenue.co.uk/revenue_chart_1950_2020UKb_17c1li0181536_722cs_10t30t> [Accessed 26 May 2021].

Anon 2021. *Public Spending Chart for United Kingdom 1950-2019 - Central Government Local Authorities*. [online] Available at:

<https://www.ukpublicspending.co.uk/spending_chart_1950_2019UKp_17c1li0181536_722cs_F0t10t20t40t60t> [Accessed 26 May 2021].

Anon 2021. *Public Spending Chart for United Kingdom 1952-2020 - Central Government*

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

Local Authorities. [online] Available at:

<https://www.ukpublicspending.co.uk/spending_chart_1952_2020UKp_17c1li011lcn_60t> [Accessed 26 May 2021].

Anon 2021. *python - Convert floats to ints in Pandas?* [online] Stack Overflow. Available at: <<https://stackoverflow.com/questions/21291259/convert-floats-to-ints-in-pandas>> [Accessed 18 May 2021].

Anon 2021. *python - Exporting Data from google colab to local machine.* [online] Stack Overflow. Available at: <<https://stackoverflow.com/questions/49394737/exporting-data-from-google-colab-to-local-machine>> [Accessed 18 May 2021].

Anon 2021. *Python Array of Numeric Values.* [online] Available at: <<https://www.programiz.com/python-programming/array>> [Accessed 15 May 2021].

Anon 2021. *QuickGraph #4: UK Official Singles Chart 2019 · Mark Needham.* [online] Mark Needham. Available at: <<https://www.markhneedham.com/blog/2020/01/04/quick-graph-uk-official-charts/>> [Accessed 15 May 2021].

Anon 2021. *Scrap UK top 40 singles chart from BBC - DZone.* [online] dzone.com. Available at: <<https://dzone.com/articles/scrap-uk-top-40-singles-chart>> [Accessed 15 May 2021].

Anon 2021. *Sentiment Analysis in Python With TextBlob.* [online] Stack Abuse. Available at: <<https://stackabuse.com/sentiment-analysis-in-python-with-textblob/>> [Accessed 11 May 2021].

Anon 2021. *spaCy 101: Everything you need to know · spaCy Usage Documentation.* [online] spaCy 101: Everything you need to know. Available at: <<https://spacy.io/usage/spacy-101>> [Accessed 11 May 2021].

Anon 2021. *spaCy Word Lemmatize - API & Demo / Text Analysis Online / TextAnalysis.* [online] Available at: <<http://textanalysisonline.com/spacy-word-lemmatize>> [Accessed 15 May 2021].

Anon 2021. *Text Analysis - API & Demo / Text Analysis Online / TextAnalysis.* [online] Available at: <<http://textanalysisonline.com/>> [Accessed 14 May 2021].

Anon 2021. *TextBlob Sentiment Analysis - API & Demo / Text Analysis Online / TextAnalysis*. [online] Available at: <<http://textanalysisonline.com/textblob-sentiment-analysis>> [Accessed 14 May 2021].

Anon 2021. *The top 10 this week in 1973: Battle of the giants in a classic Christmas chart*. [online] independent. Available at:
<<https://www.independent.ie/regionals/corkman/lifestyle/the-top-10-this-week-in-1973-battle-of-the-giants-in-a-classic-christmas-chart-39875113.html>> [Accessed 11 Mar. 2021].

Anon 2021. *U.K. Crime Rate & Statistics 1990-2021*. [online] Available at:
<<https://www.macrotrends.net/countries/GBR/united-kingdom/crime-rate-statistics>>
[Accessed 26 May 2021].

Anon 2021. *UK Healthcare Spending Since World War II for United Kingdom 1952-2020 - Central Government Local Authorities*. [online] Available at:
<https://www.ukpublicspending.co.uk/spending_chart_1952_2020UKp_17c1li111lcn_10t_UK_Healthcare_Spending_Since_World_War_II> [Accessed 26 May 2021].

Anon 2021. *United Kingdom Long Term Unemployment Rate / 1992-2020 Data / 2021-2023 Forecast*. [online] Available at: <<https://tradingeconomics.com/united-kingdom/long-term-unemployment-rate>> [Accessed 26 May 2021].

Anon 2021. *United Kingdom Youth Unemployment Rate / 1992-2021 Data / 2022-2023 Forecast*. [online] Available at: <<https://tradingeconomics.com/united-kingdom/youth-unemployment-rate>> [Accessed 26 May 2021].

Anon 2021. *Urban Dictionary: metalhead*. [online] Urban Dictionary. Available at:
<<https://www.urbandictionary.com/define.php?term=metalhead&page=21>> [Accessed 10 Mar. 2021].

Anon 2021. *What is Beautiful Soup?* [online] Eduative: Interactive Courses for Software Developers. Available at: <<https://www.educative.io/edpresso/what-is-beautiful-soup>>
[Accessed 15 May 2021].

Anon 2021. *What is the ACL and what is Computational Linguistics? / ACL Member Portal*. [online] Available at: <<https://www.aclweb.org/portal/what-is-cl>> [Accessed 25 Mar. 2021].

Anon 2021. *What is the MoSCoW Method?* [online] SearchSoftwareQuality. Available at: <<https://searchsoftwarequality.techtarget.com/definition/MoSCoW-method>> [Accessed 13 Apr. 2021].

Anon 2021. *What's with Gay Men and Musical Theater? | by James Finn | Th-Ink Queerly / Medium.* [online] Available at: <<https://medium.com/th-ink/whats-with-gay-men-and-musical-theater-a8bc6f71a739>> [Accessed 10 Mar. 2021].

Anon 2021. *Who We Are - Our charts & data.* [online] Available at: <<https://www.officialcharts.com/who-we-are/our-charts-and-data/>> [Accessed 15 May 2021].

Anon 2021. *Why do girls like listening to love songs? - Quora.* [online] Available at: <<https://www.quora.com/Why-do-girls-like-listening-to-love-songs>> [Accessed 10 Mar. 2021].

Anon 2021. *word.* [online] Available at: <<https://dictionary.cambridge.org/dictionary/english/word>> [Accessed 26 May 2021].

Anon n.d. *2015.6337.Language-Of-Politics.pdf.* Available at: <<https://ia801609.us.archive.org/30/items/in.ernet.dli.2015.6337/2015.6337.Language-Of-Politics.pdf>> [Accessed 6 Apr. 2021].

Anon n.d. *AES E-Library » The Beat Goes Static: A Tempo Analysis of U.S. Billboard Hot 100 #1 Songs from 1955—2015.* Available at: <<https://www.aes.org/e-lib/browse.cfm?elib=19246>>.

Anon n.d. *All Rock and Roll Is Homosocial: The Representation of Women in the British Rock Music Press on JSTOR.* Available at: <<https://www.jstor.org/stable/853623?seq=1>>.

Anon n.d. *Are Love Songs Lyrically Gendered? A Content Analysis of Gender-Specific Speech Features in Song Lyrics / Semantic Scholar.* Available at: <<https://www.semanticscholar.org/paper/Are-Love-Songs-Lyrically-Gendered-A-Content-of-in-Gallée/6c4fff62ea6c141b22d4c7b3276ea4e6e86efa98?sort=relevance&citedPapersSort=relevance&citedPapersLimit=10&citedPapersOffset=50>>.

Anon n.d. *Billboard Hot weekly charts - dataset by kcmillersean / data.world.* Available at: <https://data.world/kcmillersean/billboard-hot-weekly-charts>
Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

<<https://data.world/kcmillersean/billboard-hot-100-1958-2017>>.

Anon n.d. *Bye, Bye, Miss American Pie? The Supply of New Recorded Music Since Napster*. Available at: <<https://www.nber.org/papers/w16882>>.

Anon n.d. *Climbing the Charts: What Radio Airplay Tells Us about the Diffusion of ... - Gabriel Rossman - Google Books*. Available at:

<https://books.google.co.uk/books?hl=en&lr=&id=RX3BCHwxDlxC&oi=fnd&pg=PR7&dq=top+peer+reviewed+papers+music+chart&ots=wFmqwLyBQl&sig=Y3PVSDnb4cGao0IurzZ_SzKdE60&redir_esc=y#v=onepage&q=top peer reviewed papers music chart&f=false>.

Anon n.d. *CONTEMPORARY URRAN MUSIC: Controversial Messages in Hip-Hop and Rap Lyrics on JSTOR*. Available at: <<https://www.jstor.org/stable/42577630?seq=1>>.

Anon n.d. Diagnosing Teenager's Problems: Listen to Their Music. *Understanding Teenagers*. Available at: <<https://understandingteenagers.com.au/diagnosing-teenagers-problems-listen-to-their-music/>> [Accessed 10 Mar. 2021].

Anon n.d. *Diffusion Of Innovations (4th edition)*, Rogers, 1995. Available at: <<http://www.lamolina.edu.pe/postgrado/pmdas/cursos/innovacion/lecturas/Obligatoria/17%20-20Rogers%201995%20cap%206.pdf>> [Accessed 5 Apr. 2021].

Anon n.d. *Fredrik Stiernstedt - Google Scholar*. Available at: <https://scholar.google.com/citations?user=moB5RFkAAAAJ&hl=sv#d=gs_md_cita-d&u=%2Fcitations%3Fview_op%3Dview_citation%26hl%3Dsv%26user%3DmoB5RFkAA%26AA%26citation_for_view%3DmoB5RFkAAAAJ%3AdhFuZR0502QC%26tzom%3D-60>.

Anon n.d. *IDEALS @ Illinois: Improving music mood classification using lyrics, audio and social tags*. Available at: <<https://www.ideals.illinois.edu/handle/2142/18435>>.

Anon n.d. *Introduction To Sentiment Analysis*. Available at: <<https://lct-master.org/files/MullenSentimentCourseSlides.pdf>>.

Anon n.d. *MullenSentimentCourseSlides.pdf*. Available at: <<https://lct-master.org/files/MullenSentimentCourseSlides.pdf>> [Accessed 5 Apr. 2021].

Anon n.d. (*No Title*). Available at:

<<https://journals.sagepub.com/doi/pdf/10.1177/0305735617748205>>.

Anon n.d. *olympicbritain.pdf*. Available at:

<<https://www.parliament.uk/contentassets/90b7f09a39a74dbcaa34acdfe7a210cb/olympicbritain.pdf#page=159>> [Accessed 26 May 2021].

Anon n.d. [PDF] *Expressions of love, sex, and hurt in popular songs: a content analysis of all-time greatest hits / Semantic Scholar*. Available at:

<<https://www.semanticscholar.org/paper/Expressions-of-love%2C-sex%2C-and-hurt-in-popular-a-of-Dukes-Bisel/11b0f8fee09abcdab039165c60ade44676dcad09>>.

Anon n.d. [PDF] *Male and female roles in the lyrics of three genres of contemporary music / Semantic Scholar*. Available at: <<https://www.semanticscholar.org/paper/Male-and-female-roles-in-the-lyrics-of-three-genres-Freudiger-Almquist/86829ab8a3840c9e8cb22ffb863165556986ddf4>>.

Anon n.d. [PDF] *Music and Aggression: The Impact of Sexual-Aggressive Song Lyrics on Aggression-Related Thoughts, Emotions, and Behavior Toward the Same and the Opposite Sex / Semantic Scholar*. Available at: <<https://www.semanticscholar.org/paper/Music-and-Aggression%3A-The-Impact-of-Song-Lyrics-on-Fischer-Greitemeyer/f1ae2011530f41e3f273709b21fe231e76eaea2c>>.

Anon n.d. [PDF] *Popular music and society / Semantic Scholar*. Available at:
<<https://www.semanticscholar.org/paper/Popular-music-and-society-Nowak/8566c15514edb3a26e1e08a43485ce9ce129941c>>.

Anon n.d. *Popular Music and Society - Brian Longhurst - Google Books*. Available at:
<https://books.google.co.uk/books?hl=en&lr=&id=PxnOFDDMZOU&oi=fnd&pg=PR4&dq=top+peer+reviewed+papers+MUSIC&ots=MYPSVU0wrw&sig=dAYKb5ZImjiE491hqRtBcyuZpUs&redir_esc=y#v=onepage&q&f=false>.

Anon n.d. *Psychological Effects of Music with Lyrics: A Methodological Study - ProQuest*. Available at:
<<https://search.proquest.com/openview/838f38d39ae7ee2876ad99e1d40d4cd3/1?pq-origsite=gscholar&cbl=18750&diss=y>>.

Anon n.d. *Radio Automation*. Available at:
<<https://repository.najah.edu/handle/20.500.11888/12075>>.

Anon n.d. *Rebellion in the top music charts: Defiant messages in rap/hip-hop and rock music 1993 and 2003*. - PsycNET. Available at:
<<https://psycnet.apa.org/search/display?id=e7ffad39-3f10-af8a-f7e5-b62eff198bb3&recordId=6&tab=PA&page=1&display=25&sort=PublicationYearMSSortdesc,AuthorSort asc&sr=1>>.

Anon n.d. *Salim Zubair – Medium*. Available at: <<https://medium.com/@salim.zubair100>>.

Anon n.d. *salimzubair/lyric-sentiment: Sentiment analysis of all songs that entered the Billboard Hot 100 charts from 1958-2019*. Available at:
<<https://github.com/salimzubair/lyric-sentiment>>.

Anon n.d. Sentiment Analysis: Types, Tools, and Use Cases. AltexSoft. Available at:
<<https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/>>
[Accessed 16 Dec. 2020].

Anon n.d. *Sexualization in Lyrics of Popular Music from 1959 to 2009: Implications for Sexuality Educators / Semantic Scholar*. Available at:
<<https://www.semanticscholar.org/paper/Sexualization-in-Lyrics-of-Popular-Music-from-1959-Hall-West/82bffb437619c61acaa4a27118250693b4fee36a>>.

Anon n.d. *Sexually Degrading Music Videos and Lyrics / Journal of Media Psychology / Vol 24, No 1*. Available at: <<https://econtent.hogrefe.com/doi/abs/10.1027/1864-1105/a000060?journalCode=zmp>>.

Anon n.d. *The Effect of Music and Lyrics on Personality / Semantic Scholar*. Available at:
<<https://www.semanticscholar.org/paper/The-Effect-of-Music-and-Lyrics-on-Personality-Djikic/c35d6362fd7cca1ccae8a360527919391a8088>>.

Anon n.d. *The effect of music and lyrics on personality*. - PsycNET. Available at:
<<https://psycnet.apa.org/record/2011-05678-001>>.

Anon n.d. *THE EFFECT OF SONGS WITH VIOLENT LYRICS ON INTERPERSONAL AGGRESSION*. Available at: <<https://minds.wisconsin.edu/handle/1793/46734>>.

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

Anon n.d. *The Effects of Sexism on American Women: The Role of Norms vs. Discrimination*. Available at: <<https://www.nber.org/papers/w24904>>.

Anon n.d. *The Language of Lyrics / Semantic Scholar*. Available at: <<https://www.semanticscholar.org/paper/The-Language-of-Lyrics-Pettijohn-Sacco/98f93799f7ea4e85d4605ceeb29dadbf03b090>>.

Anon n.d. *The Psychology of Music in Multimedia - Google Books*. Available at: <https://books.google.co.uk/books?hl=en&lr=&id=89m5EyX59ugC&oi=fnd&pg=PP1&dq=top+peer-reviewed+papers+MUSIC&ots=b3R4xMWU8y&sig=0FD9D8sRsWxZaFQfaJW3x1oHLY&redir_esc=y#v=onepage&q&f=false>.

Anon n.d. *The sociology of popular music: A selected and annotated bibliography: Popular Music and Society: Vol 7, No 1*. Available at: <<https://www.tandfonline.com/doi/abs/10.1080/03007767908591148?journalCode=rpms20>>.

Anon n.d. “*The Whole Feminist Taking-Your-Clothes-off Thing*”: Negotiating the Critique of Gender Inequality in UK Music Industries / Bennett / IASPM Journal. Available at: <https://iaspmjournal.net/index.php/IASPM_Journal/article/view/882>.

Anon n.d. *Undressing the Words: Prevalence of Profanity, Misogyny, Violence, and Gender Role References in Popular Music from 2006–2016-Indian Journals*. Available at: <<http://www.indianjournals.com/ijor.aspx?target=ijor:mw&volume=10&issue=1&article=001>>.

Anon n.d. *What has America been singing about? Trends in themes in the U.S. top-40 songs: 1960–2010 - Peter G. Christenson, Silvia de Haan-Rietdijk, Donald F. Roberts, Tom F.M. ter Bogt, 2019*. Available at: <<https://journals.sagepub.com/doi/full/10.1177/0305735617748205>>.

Anon n.d. *Women and Rock Music / Atlantis: Critical Studies in Gender, Culture & Social Justice*. Available at: <<https://journals.msvu.ca/index.php/atlantis/article/view/4451>>.

Apte, A., 2019. *3 Ways to Load CSV files into Colab*. [online] Medium. Available at: <<https://towardsdatascience.com/3-ways-to-load-csv-files-into-colab-7c14fcdb92>> [Accessed 12 May 2021].

Arnot, Chris, 1994. *Hit machine on the road*. [online] The Independent. Available at: <<http://www.independent.co.uk/news/business/hit-machine-on-the-road-1395301.html>> [Accessed 2 Dec. 2020].

Baker, F. and Bor, W., 2008. Can Music Preference Indicate Mental Health Status in Young People? *Australasian Psychiatry*, 16(4), pp.284–288. <https://doi.org/10.1080/10398560701879589>.

Balahur, A., Hermida, J.M. and Montoyo, A., 2012. Detecting implicit expressions of emotion in text: A comparative analysis. *Decision Support Systems*, 53(4), pp.742–753. <https://doi.org/10.1016/j.dss.2012.05.024>.

Bannister, Kristian, 2020. *Sentiment Analysis: How Does It Work? Why Should We Use It?* [online] Brandwatch. Available at: <<https://www.brandwatch.com/blog/understanding-sentiment-analysis/>> [Accessed 16 Dec. 2020].

Bates, V.C., Talbot, B.C. and Tobias, E.S., 2014. *A refereed journal of the Action for Change in Music Education Flipping the Misogynist Script: Gender, Agency, Hip Hop and Music Education. Criticism & Theory for Music Education ISN 1*, pp.4–5.

Benson, K., 2008. *Musical key to unlocking teenage wasteland*. [online] The Age. Available at: <<https://www.theage.com.au/national/musical-key-to-unlocking-teenage-wasteland-20080804-3pxy.html>> [Accessed 10 Mar. 2021].

de Boise, S., 2019. Tackling gender inequalities in music: a comparative study of policy responses in the UK and Sweden. *International Journal of Cultural Policy*, 25(4), pp.486–499. <https://doi.org/10.1080/10286632.2017.1341497>.

Boso, M., Emanuele, E., Minazzi, V., Abbamonte, M. and Politi, P., 2007. Effect of Long-Term Interactive Music Therapy on Behavior Profile and Musical Skills in Young Adults with Severe Autism. *The Journal of Alternative and Complementary Medicine*, 13(7), pp.709–712. <https://doi.org/10.1089/acm.2006.6334>.

Breuer, J., Kowert, R., Festl, R. and Quandt, T., 2015. Sexist Games=Sexist Gamers? A Longitudinal Study on the Relationship Between Video Game Use and Sexist Attitudes. *Cyberpsychology, Behavior, and Social Networking*, 18(4), pp.197–202.

<https://doi.org/10.1089/cyber.2014.0492>.

Cambria, E., Das, D., Bandyopadhyay, S. and Feraco, A. eds., 2017. *A Practical Guide to Sentiment Analysis*. Socio-Affective Computing. [online] Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-55394-8>.

Chakraborty, A., Durgesh, K., Upadhyay, M., Gandhi, K., Vidyapith, V. and Agrawal, M., 2018. *Young Adults' Music Preferences and its Relation to their Attitude towards Women and Sexuality Convener of 3rd International Conference on Human Behaviour and Development Issues with theme "Growth and Sustainability View project A Critical Analysis of Psy.* [online] Available at: <www.printspublications.com>.

Chakraborty, Koyel, 2020. *Sentiment Analysis - an overview / ScienceDirect Topics*. [online] Available at: <<https://www.sciencedirect.com/topics/computer-science/sentiment-analysis>> [Accessed 16 Dec. 2020].

Charles, K.K., Guryan, J. and Pan, J., 2018. *The Effects of Sexism on American Women: The Role of Norms vs. Discrimination*. [online] Cambridge, MA. <https://doi.org/10.3386/w24904>.

Christenson, P.G., de Haan-Rietdijk, S., Roberts, D.F. and ter Bogt, T.F.M., 2019. What has America been singing about? Trends in themes in the U.S. top-40 songs: 1960–2010. *Psychology of Music*, 47(2), pp.194–212. <https://doi.org/10.1177/0305735617748205>.

Cole, R.R., 1971. Top Songs in the Sixties: A Content Analysis of Popular Lyrics. *American Behavioral Scientist*, 14(3), pp.389–400. <https://doi.org/10.1177/000276427101400311>.

Cougar Hall, P., West, J.H. and Hill, S., 2012. Sexualization in Lyrics of Popular Music from 1959 to 2009: Implications for Sexuality Educators. *Sexuality and Culture*, 16(2), pp.103–117. <https://doi.org/10.1007/s12119-011-9103-4>.

Devi Prasad, B., 2019. Qualitative Content Analysis: Why is it Still a Path Less Taken? *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 20, p.No 3 (2019): Qualitative Content Analysis I. <https://doi.org/10.17169/FQS-20.3.3392>.

Dill, K.E. and Thill, K.P., 2007. Video game characters and the socialization of gender roles: Young people's perceptions mirror sexist media depictions. *Sex Roles*, 57(11–12), pp.851–864. <https://doi.org/10.1007/s11199-007-9278-1>.

Dissertation, X.H., 2010. |376|377|000X|000i|000a|000o|000H|000u|000-|000d|000r|000a|000f|000t|000-|0002|0000|0001|0000|0001|0001|0002|0002. [online] Available at: <<http://moodydb.com>>.

Djamba, Y.K. and Neuman, W.L., 2002. Social Research Methods: Qualitative and Quantitative Approaches. *Teaching Sociology*, 30(3), p.380. <https://doi.org/10.2307/3211488>.

Djikic, M., 2011. The effect of music and lyrics on personality. *Psychology of Aesthetics, Creativity, and the Arts*, 5(3), pp.237–240. <https://doi.org/10.1037/a0022313>.

Dukes, R.L., Bisel, T.M., Borega, K.N., Lobato, E.A. and Owens, M.D., 2003. Expression of love, sex, and hurt in popular songs: A content analysis of all-time greatest hits. *Social Science Journal*, 40(4), pp.643–650. [https://doi.org/10.1016/S0362-3319\(03\)00075-2](https://doi.org/10.1016/S0362-3319(03)00075-2).

Duric, A. and Song, F., 2012. Feature selection for sentiment analysis based on content and syntax models. *Decision Support Systems*, 53(4), pp.704–711. <https://doi.org/10.1016/j.dss.2012.05.023>.

Dwaikat, I. and Abu-Higleh, M., 2008. *Radio Automation*. [online] Available at: <<https://repository.najah.edu/handle/20.500.11888/12075>>.

Esuli, A., 2021. *aesuli/SentiWordNet*. [online] Available at: <<https://github.com/aesuli/SentiWordNet>> [Accessed 6 Apr. 2021].

Fischer, P. and Greitemeyer, T., 2006. Music and aggression: The impact of sexual-aggressive song lyrics on aggression-related thoughts, emotions, and behavior toward the same and the opposite sex. *Personality and Social Psychology Bulletin*, 32(9), pp.1165–1176. <https://doi.org/10.1177/0146167206288670>.

Flynn, M.A., Craig, C.M., Anderson, C.N. and Holody, K.J., 2016. Objectification in Popular Music Lyrics: An Examination of Gender and Genre Differences. *Sex Roles*, 75(3–4), pp.164–176. <https://doi.org/10.1007/s11199-016-0592-3>.

Forums, B.E., 2021. *Music forums focusing on chart music with chart and entertainment discussion*. [music,charts,lyrics,chart music,midweeks,albums,singles,bands,downloads,music charts,entertainment discussion,music forums,music talk] Available at: <[http://www.buzzjack.com/forums/UK-Sentiment analysis of UK number ones \(1952-2020\) to detect correlation between lyrics and society](http://www.buzzjack.com/forums/UK-Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society)>

Charts-f2.html> [Accessed 15 May 2021].

Freudiger, P. and Almquist, E.M., 1978. Male and female roles in the lyrics of three genres of contemporary music. *Sex Roles*, 4(1), pp.51–65. <https://doi.org/10.1007/BF00288376>.

Fried, C.B., 2003. *Stereotypes of Music 1 Running head: STEREOTYPES OF MUSIC FANS Stereotypes of Music Fans: Are Rap and Heavy Metal Fans a Danger to Themselves or Others? Journal of Media Psychology*, .

Gallée, J., 2016. Are Love Songs Lyrically Gendered? A Content Analysis of Gender-Specific Speech Features in Song Lyrics. *undefined*.

GayBoiBlue, 2019. Stereotypical “Gay Music”. [Reddit Post] *r/gaybros*. Available at: <www.reddit.com/r/gaybros/comments/b8ayzi/stereotypical_gay_music/> [Accessed 10 Mar. 2021].

Gladkykh, T., Hnot, T. and Grubnyk, R., 2018. Music Content Selection Automation. In: *Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018*. Institute of Electrical and Electronics Engineers Inc. pp.599–604. <https://doi.org/10.1109/DSMP.2018.8478468>.

González-Bailón, S. and Paltoglou, G., 2015. *Signals of Public Opinion in Online Communication: A Comparison of Methods and Data Sources by Sandra González-Bailón, Georgios Paltoglou :: SSRN*. [online] Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2558788&download=yes> [Accessed 7 Apr. 2021].

Goyal, Kechit, 2020. Top 4 Types of Sentiment Analysis & Where to Use. *upGrad blog*. Available at: <<https://www.upgrad.com/blog/types-of-sentiment-analysis/>> [Accessed 16 Dec. 2020].

Griffin, M. and Fournet, A., 2020. F**k B*tches Raw on the Kitchen Floor: A Feminist Examination of Condom Messages in Hip Hop and Rap Music, 1991–2017. *Sexuality and Culture*, 24(1), pp.291–304. <https://doi.org/10.1007/s12119-019-09638-y>.

Grimes, Seth, n.d. 11 things to know about Sentiment Analysis. *KDnuggets*. Available at: <<https://www.kdnuggets.com/11-things-to-know-about-sentiment-analysis.html>> [Accessed Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society]

16 Dec. 2020].

Hargreaves, D.J. and North, A.C., 1999. The Functions of Music in Everyday Life: Redefining the Social in Music Psychology. *Psychology of Music*, 27(1), pp.71–83. <https://doi.org/10.1177/0305735699271007>.

Hayes, J., 2021. *The top 10 this week in 1973: Battle of the giants in a classic Christmas chart*. [online] independent. Available at: <<https://www.independent.ie/regional/corkman/lifestyle/the-top-10-this-week-in-1973-battle-of-the-giants-in-a-classic-christmas-chart-39875113.html>> [Accessed 16 May 2021].

Heller, G.M., n.d. *ABSTRACT THE EFFECT OF SONGS WITH VIOLENT LYRICS ON INTERPERSONAL AGGRESSION*.

Herd, D., 2009. Changing images of violence in Rap music lyrics: 1979-1997. *Journal of Public Health Policy*, 30(4), pp.395–406. <https://doi.org/10.1057/jphp.2009.36>.

Hostcomm, 2021. *The Evolution of Natural Language Processing: Infographic*. [text/html] Hostcomm. Available at: <<https://www.hostcomm.co.uk/blog/2019/evolution-natural-language-infographic/>> [Accessed 10 Feb. 2021].

Hu, Y. and Ogiara, M., 2011. *NEXTONE PLAYER: A MUSIC RECOMMENDATION SYSTEM BASED ON USER BEHAVIOR*.

Keith Norambuena, B., Lettura, E.F. and Villegas, C.M., 2019. Sentiment analysis and opinion mining applied to scientific paper reviews. *Intelligent Data Analysis*, 23(1), pp.191–214. <https://doi.org/10.3233/IDA-173807>.

Kian, E.M., Clavio, G., Vincent, J. and Shaw, S.D., 2011. Homophobic and sexist yet uncontested: Examining football fan postings on internet message boards. *Journal of Homosexuality*, 58(5), pp.680–699. <https://doi.org/10.1080/00918369.2011.563672>.

Knobloch-Westerwick, S., Musto, P. and Shaw, K., 2008. Rebellion in the Top Music Charts: Defiant Messages in Rap/Hip-Hop and Rock Music 1993 and 2003. *Journal of Media Psychology*, 20(1), pp.15–23. <https://doi.org/10.1027/1864-1105.20.1.15>.

Kohlbacher, F., n.d. The Use of Qualitative Content Analysis in Case Study Research. p.31.

Kubrin, C.E., 2015. “Come Along and Ride on a Fantastic Voyage”1: My Journey Through Rap Music Lyrics. In: *Envisioning Criminology*. [online] Springer International Publishing. pp.77–87. https://doi.org/10.1007/978-3-319-15868-6_8.

Kubrin, C.E. and Weitzer, R., 2010. Rap music’s violent and misogynistic effects: Fact or fiction? *Sociology of Crime Law and Deviance*, 14, pp.121–143. [https://doi.org/10.1108/S1521-6136\(2010\)0000014009](https://doi.org/10.1108/S1521-6136(2010)0000014009).

Kuckartz, U., 2019. Qualitative Inhaltsanalyse: von Kracauers Anfängen zu heutigen Herausforderungen. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 20, p.No 3 (2019): Qualitative Content Analysis I. <https://doi.org/10.17169/FQS-20.3.3370>.

Lafrance, M., Scheibling, C., Burns, L. and Durr, J., 2018. Race, gender, and the Billboard Top 40 charts between 1997 and 2007. *Popular Music and Society*, 41(5), pp.522–538. <https://doi.org/10.1080/03007766.2017.1377588>.

Lasswell, H.D., 1949. *Language of politics: studies in quantitative semantics*. [online] New York,. Available at: <<http://hdl.handle.net/2027/mdp.39015002398272>>.

Lasswell, H.D., 1952. *The comparative study of symbols :an introduction /*. [online] Stanford : Available at: <<http://hdl.handle.net/2027/uc1.32106001017414>>.

Lena, J.C., 2006. Social Context and Musical Content of Rap Music, 1979-1995. *Social Forces*, 85(1), pp.479–495. <https://doi.org/10.1353/sof.2006.0131>.

Lewis, G.H., 1979. The sociology of popular music: A selected and annotated bibliography. *Popular Music and Society*, 7(1), pp.57–68. <https://doi.org/10.1080/03007767908591148>.

LLC, C., 2021a. *Mean, Median, Mode Calculator*. [online] CalculatorSoup. Available at: <<https://www.calculatorsoup.com/calculators/statistics/mean-median-mode.php>> [Accessed 25 May 2021].

LLC, C., 2021b. *Percentage Calculator*. [online] CalculatorSoup. Available at: <<https://www.calculatorsoup.com/calculators/math/percentage.php>> [Accessed 27 May 2021].

- LLC, C.T., 2020. *What Are the Different Types of Sentiment Analysis?* [online] Medium. Available at: <<https://medium.com/cogitotech/what-are-the-different-types-of-sentiment-analysis-808f36ef89ee>> [Accessed 16 Dec. 2020].
- Lyons, F., Sun, H., Collopy, D.P., Curran, K. and Ohagan, P., 2019. Music 2025 – The Music Data Dilemma: Issues Facing the Music Industry in Improving Data Management. *SSRN Electronic Journal*. [online] <https://doi.org/10.2139/ssrn.3437670>.
- M. FRISBY, C. and BEHM-MORAWITZ, E., 2019. Undressing the Words: Prevalence of Profanity, Misogyny, Violence, and Gender Role References in Popular Music from 2006–2016. *Media Watch*, 10(1). <https://doi.org/10.15655/mw/2019/v10i1/49562>.
- Macnamara, J., 2018. Content Analysis. In: *Mediated Communication*. pp.191–212. <https://doi.org/10.1515/9783110481129-012>.
- Madanikia, Y. and Bartholomew, K., 2014. Themes of lust and love in popular music lyrics from 1971 to 2011. *SAGE Open*, 4(3). <https://doi.org/10.1177/2158244014547179>.
- Maks, I. and Vossen, P., 2012. A lexicon model for deep sentiment analysis and opinion mining applications. *Decision Support Systems*, 53(4), pp.680–688. <https://doi.org/10.1016/j.dss.2012.05.025>.
- Mäntylä, M.V., Graziotin, D. and Kuutila, M., 2018a. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, pp.16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>.
- Mäntylä, M.V., Graziotin, D. and Kuutila, M., 2018b. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, pp.16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>.
- Markert, J., 2001. Sing a Song of Drug Use-Abuse: Four Decades of Drug Lyrics in Popular Music?From the Sixties through the Nineties. *Sociological Inquiry*, 71(2), pp.194–220. <https://doi.org/10.1111/j.1475-682X.2001.tb01108.x>.
- Mayring, P., 2000. Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal]*, <http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm>, 1.

Mays, K.L., Clark, D.L. and Gordon, A.J., 2008. Treating Addiction with Tunes: A Systematic Review of Music Therapy for the Treatment of Patients with Addictions. *Substance Abuse*, 29(4), pp.51–59. <https://doi.org/10.1080/08897070802418485>.

mihart, 2021a. *Export data from a Power BI visualization - Power BI*. [online] Available at: <<https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-export-data>> [Accessed 22 May 2021].

mihart, 2021b. *What is Power BI? - Power BI*. [online] Available at: <<https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>> [Accessed 17 May 2021].

Montoyo, A., Martínez-Barco, P. and Balahur, A., 2012. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*, 53(4), pp.675–679. <https://doi.org/10.1016/j.dss.2012.05.022>.

Morgan, N., Shaw, O., Mailley, J., Channing, R., Sweiry, A., Kruithof, K., Bloom, J., Kent, A., Smith, K. and Romaniuk, A., n.d. Trends and drivers of homicide: Main findings. p.80.

Mulder, J., Bogt, T.T., Raaijmakers, Q. and Vollebergh, W., 2007. Music taste groups and problem behavior. *Journal of Youth and Adolescence*, 36(3), pp.313–324. <https://doi.org/10.1007/s10964-006-9090-1>.

Napier, K. and Shamir, L., 2018. Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*, 30(4), pp.161–176. <https://doi.org/10.1525/jpms.2018.300411>.

Nowak, R., 2015. Popular music and society. *Information, Communication & Society*, 18(12), pp.1452–1454. <https://doi.org/10.1080/1369118x.2015.1046895>.

O'harat, T., Wiebet, J. and Bruce, R., 2002. Development and Use of a Gold-Standard Data Set for Subjectivity Classifications. <https://doi.org/10.3115/1034678.1034721>.

Osborne, C., 2017. *Analysing UK Chart history — 1956 to 2017*. [online] Medium. Available at: <<https://medium.com/@caineosborne/analysing-uk-chart-history-1956-to-2017-6fec0ecc991b>> [Accessed 15 May 2021].

Parkinson, Justin, 2016. The significance of Sarah Baartman. *BBC News*. [online] 7 Jan. Available at: <<https://www.bbc.com/news/magazine-35240987>> [Accessed 1 Dec. 2020].

Patodkar, V.N. and I.R, S., 2016. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *IJARCCE*, 5(12), pp.320–322. <https://doi.org/10.17148/IJARCCE.2016.51274>.

Pettijohn, T.F. and Sacco, D.F., 2009. The language of lyrics: An analysis of popular Billboard songs across conditions of social and economic threat. *Journal of Language and Social Psychology*, 28(3), pp.297–311. <https://doi.org/10.1177/0261927X09335259>.

Philip E. Converse, 1987. *CHANGING CONCEPTIONS OF PUBLIC OPINION IN THE POLITICAL PROCESS.*: EBSCOhost. [online] Available at: <<http://web.b.ebscohost.com.mtu.idm.oclc.org/ehost/pdfviewer/pdfviewer?vid=2&sid=6af3733e-819a-4b55-8c71-fe257424642b%40pdc-v-sessmgr01>> [Accessed 6 Apr. 2021].

Power, Ed and O'Connor, Roisin, 2019. *The 40 best song lyrics of all time*. [online] The Independent. Available at: <<https://www.independent.co.uk/arts-entertainment/music/features/best-song-lyrics-ever-rolling-stones-beatles-beyonce-nirvana-bowie-kate-bush-a8758411.html>> [Accessed 2 Dec. 2020].

Prabhakaran, S., 2018. Lemmatization Approaches with Examples in Python. *ML+*. Available at: <<https://www.machinelearningplus.com/nlp/lemmatization-examples-python/>> [Accessed 14 May 2021].

Prasad, B.D., 2008. *Content Analysis A method in Social Science Research*.

Primack, B.A., Gold, M.A., Schwarz, E.B. and Dalton, M.A., 2008. Degrading and non-degrading sex in popular music: A content analysis. *Public Health Reports*, 123(5), pp.593–600. <https://doi.org/10.1177/003335490812300509>.

Ransom, P.F., n.d. Message in the Music: Do Lyrics Influence Well-Being? p.45.

Read, H., 2020. An Academic Guide to Content Analysis. *Heather Read*. Available at: <<https://heatherread.com/2020/03/25/an-academic-guide-to-content-analysis/>> [Accessed 6 Apr. 2021].

Rentfrow, P., McDonald, J. and Oldmeadow, J., 2009. You Are What You Listen To: Young

People's Stereotypes about Music Fans. *Group Processes & Intergroup Relations - GROUP PROCESS INTERGROUP RELA*, 12, pp.329–344.
<https://doi.org/10.1177/1368430209102845>.

Rentfrow, P.J., 2012. The Role of Music in Everyday Life: Current Directions in the Social Psychology of Music. *Social and Personality Psychology Compass*, 6(5), pp.402–416.
<https://doi.org/10.1111/j.1751-9004.2012.00434.x>.

Roessler, S., 2017. The Beat Goes Static: A Tempo Analysis of U.S. Billboard Hot 100 #1 Songs from 1955–2015.

Rush, B., 2018. *A visual analysis of UK number 1s: getting down and dirty with data*. [online] Medium. Available at: <<https://towardsdatascience.com/a-visual-analysis-of-uk-number-1s-getting-down-and-dirty-with-data-a663cee021c4>> [Accessed 15 May 2021].

Sapienza, Z., Iyer, N. and Veenstra, A., 2015. Reading Lasswell's Model of Communication Backward: Three Scholarly Misconceptions. *Mass Communication & Society*, 18, pp.599–622. <https://doi.org/10.1080/15205436.2015.1063666>.

Schneider, L. and Gros, C., 2019. Five decades of US, UK, German and Dutch music charts show that cultural processes are accelerating. *Royal Society Open Science*, 6(8), pp.190944–190944. <https://doi.org/10.1098/rsos.190944>.

Schoell, E.D., 2017. *How do I use the slice notation in Python?* [online] O'Reilly Media. Available at: <<https://www.oreilly.com/content/how-do-i-use-the-slice-notation-in-python/>> [Accessed 15 May 2021].

Serge Denisoff, R. and Bridges, J., 1983. The sociology of popular music: A review. *Popular Music and Society*, 9(1), pp.51–62. <https://doi.org/10.1080/03007768308591206>.

Sharma, V.B., n.d. *rake-nltk: Python implementation of the Rapid Automatic Keyword Extraction algorithm using NLTK*. [Python] Available at: <<https://github.com/csurfer/rake-nltk>> [Accessed 11 May 2021].

Sharpe, C., 2018. *49 Years of Lyrics: Why so Angry?* [online] Medium. Available at: <<https://towardsdatascience.com/49-years-of-lyrics-why-so-angry-1adf0a3fa2b4>> [Accessed 25 Nov. 2020].

Sharpe, C., 2019. *sharpie-007/dataAndMusic*. [Jupyter Notebook] Available at: <<https://github.com/sharpie-007/dataAndMusic>> [Accessed 25 Nov. 2020].

Shelley, M. and Krippendorff, K., 1984. Content Analysis: An Introduction to its Methodology. *Journal of the American Statistical Association*, 79(385), p.240. <https://doi.org/10.2307/2288384>.

Songfacts, 2021. *Lyrics for What A Wonderful World by Louis Armstrong - Songfacts*. [online] Available at: <<https://www.songfacts.com/lyrics/louis-armstrong/what-a-wonderful-world>> [Accessed 14 May 2021].

SongMeanings, 2020. *Slade - Merry Xmas Everybody Lyrics*. [online] SongMeanings. Available at: <<https://songmeanings.com/songs/view/3530822107858616247/>> [Accessed 2 Dec. 2020].

Sprankle, E.L., End, C.M. and Bretz, M.N., 2012. Sexually degrading music videos and lyrics their effects on males' aggression and endorsement of rape myths and sexual stereotypes. *Journal of Media Psychology*, 24(1), pp.31–39. <https://doi.org/10.1027/1864-1105/a000060>.

Stagner, R., 1940. The Cross-Out Technique as a Method in Public Opinion Analysis. *The Journal of Social Psychology*, 11(1), pp.79–90. <https://doi.org/10.1080/00224545.1940.9918734>.

Steinberger, J., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., Steinberger, R., Tanev, H., Vázquez, S. and Zavarella, V., 2012. Creating sentiment dictionaries via triangulation. *Decision Support Systems*, 53(4), pp.689–694. <https://doi.org/10.1016/j.dss.2012.05.029>.

Timmerman, L.M., Allen, M., Jorgensen, J., Herrett-Skjellum, J., Kramer, M.R. and Ryan, D.J., 2008. A Review and Meta-Analysis Examining the Relationship of Music Content with Sex, Race, Priming, and Attitudes. *Communication Quarterly*, 56(3), pp.303–324. <https://doi.org/10.1080/01463370802240932>.

Varghese, R. and M, J., n.d. *A Survey on Sentiment Analysis and Opinion Mining*.

Vinodhini, G. and Chandrasekaran, R., 2012. Sentiment Analysis and Opinion Mining: A Survey. 2(6), p.11.

Vryzas, N., Tsipas, N. and Dimoulas, C., 2020. Web Radio Automation for Audio Stream Management in the Era of Big Data. *Information*, 11(4), pp.205–205.
<https://doi.org/10.3390/info11040205>.

Waldfogel, J., 2011. Bye, Bye, Miss American Pie? The Supply of New Recorded Music Since Napster. *National Bureau of Economic Research*. <https://doi.org/10.3386/w16882>.

Weitzer, R. and Kubrin, C.E., 2009. Misogyny in rap music: A content analysis of prevalence and meanings. *Men and Masculinities*, 12(1), pp.3–29.
<https://doi.org/10.1177/1097184X08327696>.

Werner, V., 2012. Love is all around: a corpus-based study of pop lyrics. *Corpora*, 7(1), pp.19–50. <https://doi.org/10.3366/cor.2012.0016>.

Williams, M., 2002. *Obituary: Percy Dickins*. [online] the Guardian. Available at: <<http://www.theguardian.com/news/2002/feb/19/guardianobituaries1>> [Accessed 2 Dec. 2020].

Wright, C.L. and Rubin, M., 2017. “Get lucky!” Sexual content in music lyrics, videos and social media and sexual cognitions and risk among emerging adults in the USA and Australia. *Sex Education*, 17(1), pp.41–56. <https://doi.org/10.1080/14681811.2016.1242402>.

Yadollahi, A., Shahraki, A.G. and Zaiane, O.R., 2017. Current State of Text Sentiment Analysis from Opinion to Emotion Mining. *ACM Computing Surveys*, 50(2), p.25:1-25:33. <https://doi.org/10.1145/3057270>.

Zealand, N. and Nett, E., 1984. *Women and Rock Music 1 Deborah Harding*. [online] *Atlantis*, pp.60–76. Available at: <<https://journals.msvu.ca/index.php/atlantis/article/view/4451>>.

Appendices

Appendix A - Prototype screenshots

```

1  from textblob import TextBlob
2  #install modules
3  !pip install lyricsgenius
4  !pip install -U gensim
5  !pip install -U spacy
6  !pip install textblob
7  !python -m textblob.download_corpora
8  !pip install rake-nltk
9
10 #import packages
11 from gensim.summarization import keywords
12 import spacy
13 import nltk
14 from nltk.corpus import stopwords
15 nltk.download('stopwords')
16 stopwords = set(stopwords.words('english'))
17 from spacy.lang.en import English
18 nlp = English()
19 nlp.max_length = 10000000
20 import lyricsgenius
21 import pandas as pd
22 from textblob import TextBlob

```

Figure 77 - Installing libraries and packages.

```

Requirement already satisfied: lyricsgenius in d:\anaconda3\lib\site-packages (2.0.2)
Requirement already satisfied: beautifulsoup4>=4.6.0 in d:\anaconda3\lib\site-packages (from lyricsgenius) (4.7.1)
Requirement already satisfied: requests>=2.20.0 in d:\anaconda3\lib\site-packages (from lyricsgenius) (2.22.0)
Requirement already satisfied: soupsieve>=1.2 in d:\anaconda3\lib\site-packages (from beautifulsoup4>=4.6.0->lyricsgenius) (1.8)
Requirement already satisfied: certifi>=2017.4.17 in d:\anaconda3\lib\site-packages (from requests>=2.20.0->lyricsgenius) (2019.6.16)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in d:\anaconda3\lib\site-packages (from requests>=2.20.0->lyricsgenius) (1.24.2)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in d:\anaconda3\lib\site-packages (from requests>=2.20.0->lyricsgenius) (3.0.4)
Requirement already satisfied: idna<2.9,>=2.5 in d:\anaconda3\lib\site-packages (from requests>=2.20.0->lyricsgenius) (2.8)
Requirement already up-to-date: gensim in d:\anaconda3\lib\site-packages (3.8.3)
Requirement already satisfied, skipping upgrade: scipy>=0.18.1 in d:\anaconda3\lib\site-packages (from gensim) (1.2.1)
Requirement already satisfied, skipping upgrade: Cython=>0.29.14 in d:\anaconda3\lib\site-packages (from gensim) (0.29.14)
Requirement already satisfied, skipping upgrade: smart-open>=1.8.1 in d:\anaconda3\lib\site-packages (from gensim) (4.0.1)
Requirement already satisfied, skipping upgrade: six>=1.5.0 in d:\anaconda3\lib\site-packages (from gensim) (1.12.0)
Requirement already satisfied, skipping upgrade: numpy>=1.11.3 in d:\anaconda3\lib\site-packages (from gensim) (1.16.4)
Requirement already up-to-date: spacy in d:\anaconda3\lib\site-packages (2.3.4)
Requirement already satisfied, skipping upgrade: plac<1.2.0,>=0.9.6 in d:\anaconda3\lib\site-packages (from spacy) (1.1.3)
Requirement already satisfied, skipping upgrade: thinc<7.5.0,>=7.4.1 in d:\anaconda3\lib\site-packages (from spacy) (7.4.4)
Requirement already satisfied, skipping upgrade: requests<3.0.0,>=2.13.0 in d:\anaconda3\lib\site-packages (from spacy) (2.22.0)
Requirement already satisfied, skipping upgrade: murmurhash<1.1.0,>=0.28.0 in d:\anaconda3\lib\site-packages (from spacy) (1.0.5)
Requirement already satisfied, skipping upgrade: numpy>=1.15.0 in d:\anaconda3\lib\site-packages (from spacy) (1.16.4)
Requirement already satisfied, skipping upgrade: srsly<1.1.0,>=1.0.2 in d:\anaconda3\lib\site-packages (from spacy) (1.0.5)
Requirement already satisfied, skipping upgrade: catalogue<1.1.0,>=0.0.7 in d:\anaconda3\lib\site-packages (from spacy) (1.0.0)
Requirement already satisfied, skipping upgrade: blis<0.8.0,>=0.4.0; python_version >= "3.6" in d:\anaconda3\lib\site-packages (from spacy) (0.7.4)
Requirement already satisfied, skipping upgrade: preshed<3.1.0,>=3.0.2 in d:\anaconda3\lib\site-packages (from spacy) (3.0.5)
Requirement already satisfied, skipping upgrade: setuptools in d:\anaconda3\lib\site-packages (from spacy) (41.0.1)
Requirement already satisfied, skipping upgrade: wasabi<1.1.0,>=0.4.0 in d:\anaconda3\lib\site-packages (from spacy) (0.8.0)
Requirement already satisfied, skipping upgrade: cymem<2.1.0,>=2.0.2 in d:\anaconda3\lib\site-packages (from spacy) (2.0.5)
Requirement already satisfied, skipping upgrade: tqdm<5.0.0,>=4.38.0 in d:\anaconda3\lib\site-packages (from spacy) (4.54.1)
Requirement already satisfied, skipping upgrade: idna<2.9,>=2.5 in d:\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.8)
Requirement already satisfied, skipping upgrade: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in d:\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (1.24.2)
Requirement already satisfied, skipping upgrade: certifi>=2017.4.17 in d:\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2019.6.16)
Requirement already satisfied, skipping upgrade: chardet<3.1.0,>=3.0.2 in d:\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.0.4)
Requirement already satisfied, skipping upgrade: importlib-metadata>=0.20; python_version < "3.8" in d:\anaconda3\lib\site-packages (from catalogue<1.1.0,>=0.0.7->spacy) (3.1.1)
Requirement already satisfied, skipping upgrade: zipp>=0.5 in d:\anaconda3\lib\site-packages (from importlib-metadata>=0.20; python_version < "3.8">catalogue<1.1.0,>=0.0.7->spacy) (0.5.1)
Requirement already satisfied: textblob in d:\anaconda3\lib\site-packages (0.15.3)
Requirement already satisfied: nltk>=3.1 in d:\anaconda3\lib\site-packages (from textblob) (3.4.4)
Requirement already satisfied: six in d:\anaconda3\lib\site-packages (from nltk>=3.1->textblob) (1.12.0)
Finished.

```

Figure 78 - Libraires and packages installing (1)

```
[nltk_data] Downloading package brown to
[nltk_data]   C:\Users\colmg\AppData\Roaming\nltk_data...
[nltk_data]   Package brown is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\colmg\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]   C:\Users\colmg\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   C:\Users\colmg\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]     date!
[nltk_data] Downloading package conll2000 to
[nltk_data]   C:\Users\colmg\AppData\Roaming\nltk_data...
[nltk_data]   Package conll2000 is already up-to-date!
[nltk_data] Downloading package movie_reviews to
[nltk_data]   C:\Users\colmg\AppData\Roaming\nltk_data...
[nltk_data]   Package movie_reviews is already up-to-date!

Requirement already satisfied: rake-nltk in d:\anaconda3\lib\site-packages (1.0.4)
Requirement already satisfied: nltk in d:\anaconda3\lib\site-packages (from rake-nltk) (3.4.4)
Requirement already satisfied: six in d:\anaconda3\lib\site-packages (from nltk>rake-nltk) (1.12.0)

[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\colmg\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Figure 79 - Libraires and packages installing (2)

```
1 #define Genius API authentication
2 genius = lyricsgenius.Genius('_g20PhiAv3u_FpHXcs_vPjwUFlainNCXpbfiYgaRSwpLt4FiDFUDpe5NSOIq8ppA')
3
4 lyric_test = genius.search_song('Shake It Off', 'Taylor Swift').lyrics
5 print(lyric_test)
```

Figure 80 - Testing connection between lyricsgenius and Genius API

Searching for "Shake It Off" by Taylor Swift...

Done.

[Verse 1]

I stay out too late
Got nothing in my brain
That's what people say, mmm, mmm
That's what people say, mmm, mmm
I go on too many dates
But I can't make them stay
At least that's what people say, mmm, mmm
That's what people say, mmm, mmm

[Pre-Chorus 1]

But I keep cruisin'
Can't stop, won't stop movin'
It's like I got this music in my mind
Sayin' it's gonna be alright

[Chorus]

'Cause the players gonna play, play, play, play, play
And the haters gonna hate, hate, hate, hate, hate
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off
Heartbreakers gonna break, break, break, break, break
And the fakers gonna fake, fake, fake, fake, fake
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off

[Verse 2]

I never miss a beat
I'm lightning on my feet
And that's what they don't see, mmm, mmm
That's what they don't see, mmm, mmm
I'm dancin' on my own (Dancin' on my own)
I make the moves up as I go (Moves up as I go)
And that's what they don't know, mmm, mmm
That's what they don't know, mmm, mmm

[Pre-Chorus 2]

But I keep cruisin'
Can't stop, won't stop groovin'
It's like I got this music in my mind
Sayin' it's gonna be alright

[Post-Chorus]

Shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off
Shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off (Yeah)
Shake it off, I shake it off
I, I, I shake it off, I shake it off (You got to)
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off

[Video directed by Mark Romanek]

[Chorus]

'Cause the players gonna play, play, play, play, play
And the haters gonna hate, hate, hate, hate, hate
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off
Heartbreakers gonna break, break, break, break, break
And the fakers gonna fake, fake, fake, fake, fake
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off

[Post-Chorus]

Shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off
I, I, I shake it off, I shake it off

[Interlude]

Hey, hey, hey
Just think while you've been gettin' down and out about the liars
And the dirty, dirty cheats of the world
You could've been getting down
To this sick beat

[Bridge]

My ex-man brought his new girlfriend
She's like, "Oh my God," but I'm just gonna shake
And to the fellas over there with the hella good hair
Won't you come on over, baby?
We can shake, shake, shake
Yeah, oh, oh, oh

[Chorus]

'Cause the players gonna play, play, play, play, play
And the haters gonna hate, hate, hate, hate, hate
(Haters gonna hate)
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off
Heartbreakers gonna break, break, break, break, break (Mmhmm)
And the fakers gonna fake, fake, fake, fake, fake
(And fake, and fake, and fake)
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off

Figure 81 - Taylor Swift's "Shake It Off" proving successful connection between lyricsgenius and Genius API (a), (b), (c)

```
1 #import billboard hot100 dataset
2 hot100_df = pd.read_csv('https://query.data.world/s/qf6et5c7dh23kglnvjcoztlmom62it')
3 hot100_df.drop_duplicates(subset='SongID', inplace = True) #remove duplicate occurrences of songs
4 hot100_df = hot100_df.head(1000)
5 hot100_df.reset_index()
```

Figure 82 - Importing csv file from data.world

index		url	WeekID	Week Position	Song	Performer	SongID	Instance	Previous Week Position	Peak Position	Weeks on Chart
0	0	http://www.billboard.com/charts/hot-100/1958-0...	8/2/1958	1	Poor Little Fool	Ricky Nelson	Poor Little FoolRicky Nelson	1	NaN	1	1
1	1	http://www.billboard.com/charts/hot-100/1995-1...	12/2/1995	1	One Sweet Day	Mariah Carey & Boyz II Men	One Sweet DayMariah Carey & Boyz II Men	1	NaN	1	1
2	2	http://www.billboard.com/charts/hot-100/1997-1...	10/11/1997	1	Candle In The Wind 1997/Something About The Wa...	Elton John	Candle In The Wind 1997/Something About The Wa...	1	NaN	1	1
3	3	http://www.billboard.com/charts/hot-100/2006-0...	7/1/2006	1	Do I Make You Proud	Taylor Hicks	Do I Make You ProudTaylor Hicks	1	NaN	1	1
4	4	http://www.billboard.com/charts/hot-100/2009-1...	10/24/2009	1	3	Britney Spears	3Britney Spears	1	NaN	1	1
...
27	27	https://www.billboard.com/charts/hot-100/2019-...	2/23/2019	2	Break Up With Your Girlfriend, I'm Bored	Ariana Grande	Break Up With Your Girlfriend, I'm BoredAriana...	1	NaN	2	1
28	28	https://www.billboard.com/charts/hot-100/2019-...	7/6/2019	2	Senorita	Shawn Mendes & Camila Cabello	SenoritaShawn Mendes & Camila Cabello	1	NaN	2	1
29	29	http://www.billboard.com/charts/hot-100/1958-0...	8/2/1958	3	Splish Splash	Bobby Darin	Splish SplashBobby Darin	1	NaN	3	1
...
970	1158	http://www.billboard.com/charts/hot-100/2016-0...	5/14/2016	91	Company	Justin Bieber	CompanyJustin Bieber	2	58.0	53	12
971	1159	https://www.billboard.com/charts/hot-100/2018-...	9/29/2018	91	Solo	Clean Bandit Featuring Demi Lovato	SoloClean Bandit Featuring Demi Lovato	2	80.0	58	12
972	1160	https://www.billboard.com/charts/hot-100/2017-...	12/2/2017	91	Questions	Chris Brown	QuestionsChris Brown	2	78.0	78	12
995	1185	http://www.billboard.com/charts/hot-100/2016-0...	5/28/2016	7	Don't Let Me Down	The Chainsmokers Featuring Daya	Don't Let Me DownThe Chainsmokers Featuring Daya	2	7.0	7	13
996	1186	http://www.billboard.com/charts/hot-100/2011-0...	1/15/2011	8	Black And Yellow	Wiz Khalifa	Black And YellowWiz Khalifa	2	6.0	6	13
997	1187	http://www.billboard.com/charts/hot-100/1959-0...	1/3/1959	8	A Lover's Question	Clyde McPhatter	A Lover's QuestionClyde McPhatter	2	9.0	8	13
998	1188	http://www.billboard.com/charts/hot-100/2008-0...	5/24/2008	9	Pocketful Of Sunshine	Natasha Bedingfield	Pocketful Of SunshineNatasha Bedingfield	2	8.0	8	13
999	1189	http://www.billboard.com/charts/hot-100/2013-0...	2/16/2013	10	Sweet Nothing	Calvin Harris Featuring Florence Welch	Sweet NothingCalvin Harris Featuring Florence Welch	2	14.0	10	13

Figure 83 - Imported csv table displayed in notebook.

```

1 #Function to return lyrics of each song using Genius API
2 def get_lyrics(title, artist):
3     try:
4         return genius.search_song(title, artist).lyrics
5     except:
6         return 'not found'
7
8 def get_lyric_sentiment(lyrics):
9     analysis = TextBlob(lyrics)
10    return analysis.sentiment.polarity
11
12 # Function to preprocess text
13 def preprocess(text):
14     # Create Doc object
15     doc = nlp(text, disable=['ner', 'parser'])
16     # Generate Lemmas
17     lemmas = [token.lemma_ for token in doc]
18     # Remove stopwords and non-alphabetic characters
19     a_lemmas = [lemma for lemma in lemmas
20                 if lemma.isalpha() and lemma not in stopwords]
21
22     return ' '.join(a_lemmas)
23
24 """Extract Keywords from text"""
25 def return_keywords(texts):
26     xkeywords = []
27     values = keywords(text=preprocess(texts), split='\n', scores=True)
28     for x in values[:10]:
29         xkeywords.append(x[0])
30     try:
31         return xkeywords
32     except:
33         return "no content"

```

Figure 84 - Functions being defined for use further on

```

1 #Use get_lyrics function to get lyrics for every song in dataset
2 lyrics = hot100_df.apply(lambda row: get_lyrics(row['Song'], row['Performer']), axis =1)
3 hot100_df['Lyrics'] = lyrics
4 hot100_df = hot100_df.drop(hot100_df[hot100_df['Lyrics'] == 'not found'].index) #drop rows where lyrics are not found on Genius API
5 hot100_df

```

Figure 85 - get_lyrics function connecting with Genius API.

	url	WeekID	Week Position	Song	Performer	SongID	Instance	Previous Week Position	Peak Position	Weeks on Chart
0	http://www.billboard.com/charts/hot-100/1958-0...	8/2/1958	1	Poor Little Fool	Ricky Nelson	Poor Little FoolRicky Nelson	1	NaN	1	I used he...
1	http://www.billboard.com/charts/hot-100/1995-1...	12/2/1995	1	One Sweet Day	Mariah Carey & Boyz II Men	One Sweet DayMariah Carey & Boyz II Men	1	NaN	1	[Verse 1: ...
3	http://www.billboard.com/charts/hot-100/2006-0...	7/1/2006	1	Do I Make You Proud	Taylor Hicks	Do I Make You ProudTaylor Hicks	1	NaN	1	I beeninT...
28	https://www.billboard.com/charts/hot-100/2019-...	7/6/2019	2	Senorita	Mendes & Camila Cabello	SenoritaCamila Mendes & Camila Cabello	1	NaN	2	[Chorus: Camila Cabello]I love it when you c...
29	http://www.billboard.com/charts/hot-100/1958-0...	8/2/1958	3	Splish Splash	Bobby Darin	Splish SplashBobby Darin	1	NaN	3	Splish splash, I was taking bathinLong about...
30	http://www.billboard.com/charts/hot-100/1998-1...	10/3/1998	3	One Week	Barenaked Ladies	One WeekBarenaked Ladies	1	NaN	3	[Chorus 1]N'ts been on week since you look...
31	http://www.billboard.com/charts/hot-100/2008-1...	12/20/2008	3	Circus	Britney Spears	CircusBritney Spears	1	NaN	3	[Verse 1]InThere's only tw...
...
1157	http://www.billboard.com/charts/hot-100/2003-1...	11/8/2003	90	Pon De River, Pon De Bank	Elephant Man	Pon De River, Pon De BankElephant Man	2	90.0	86	[Intro]InUp, good t...
1158	http://www.billboard.com/charts/hot-100/2016-0...	5/14/2016	91	Company	Justin Bieber	CompanyJustin Bieber	2	58.0	53	[Chorus]InCan we, w...
1185	http://www.billboard.com/charts/hot-100/2016-0...	5/28/2016	7	Don't Let Me Down	The Chainsmokers Featuring Daya	Don't Let Me DownThe Chainsmokers Featuring Daya	2	7.0	7	[Verse 1]InCrashing, hit wallInRight now I n...
1186	http://www.billboard.com/charts/hot-100/2011-0...	1/15/2011	8	Black And Yellow	Wiz Khalifa	Black And YellowWiz Khalifa	2	6.0	6	[Intro]InYeah, uh-huh, yo...
1187	http://www.billboard.com/charts/hot-100/1959-0...	1/3/1959	8	A Lover's Question	Clyde McPhatter	A Lover's QuestionClyde McPhatter	2	9.0	8	Does she love me, with a her heartInShould I...
1188	http://www.billboard.com/charts/hot-100/2008-0...	5/24/2008	9	Pocketful Of Sunshine	Natasha Bedingfield	Pocketful Of SunshineNatasha Bedingfield	2	8.0	8	[Intro]InOh, yeah, yeah yeahIn[Chorus]In g...
1189	http://www.billboard.com/charts/hot-100/2013-0...	2/16/2013	10	Sweet Nothing	Calvin Harris Featuring Florence Welch	Sweet NothingCalvin Harris Featuring Florence ...	2	14.0	10	[Verse 1: Florenc...Welch]InYou took my hear...

Figure 86 - Lyrics being imported from Genius API (a), and being added as a new 'Lyrics' column (b), (c), (d)

```

1 #Use get_lyric_sentiment to get sentiment score for all the song lyrics
2 sentiment = hot100_df.apply(lambda row: get_lyric_sentiment(row['Lyrics']), axis =1)
3 hot100_df[ 'Sentiment' ] = sentiment

```

Figure 87 - *get_lyric_sentiment* function being called to classify polarity on lyrics and create 'Sentiment' column.

```

1 #Set the index of the dataframe to the WeekID. This sets us up to resample dataframe based on time
2 hot100_df[ 'WeekID' ] = pd.to_datetime(hot100_df[ 'WeekID' ],infer_datetime_format=True)
3 hot100_df = hot100_df.sort_values(by='WeekID')
4 hot100_df = hot100_df.reset_index(drop=True)
5 hot100_df = hot100_df.set_index('WeekID')

```

Figure 88 - Setting the index of the dataframe to be sorted by 'WeekID' column.

```

1 #Resample dataframe sentiment column by year. This arranges the mean Lyric sentiment for every year
2 hot100_resample = hot100_df[ 'Sentiment' ].resample('Y').mean()

```

Figure 89 - Resampling the dataframe to calculate the mean sentiment of the lyrics for each year.

```

1 #Plot a Line graph of lyric sentiment over time
2 hot100_resample.plot(label='Sentiment', legend=True)

```

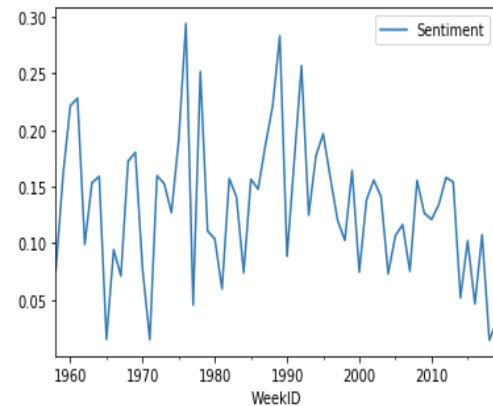


Figure 90 - Plotting a line graph to display the sentiment of each year.

```

1 #Calculate the mean percentage change of Lyric sentiment
2 hot100_resample.pct_change().mean()

```

0.4078566286381956

Figure 91 - Resampling the lyric sentiment (a) to calculate the mean percentage change (b)

```

1 #Resample daraframe lyrics by year. Get all the lyrics for every song for each year
2 lyrics_resample = hot100_df[ 'Lyrics' ].resample('Y').sum()

```

Figure 92 - Resampling the 'Lyrics' column to collect the lyrics from each year.

```

1 #Use return_keywords function on lyrics_resample to get the top 20 keywords for each year
2 lyric_keywords = [return_keywords(x[1]) for x in lyrics_resample.iteritems()]
3
4 #Lyric_keywords = ""
5 #for x in lyrics_resample:
6 #    #print(x)
7 #    if x != "0":
8 #        #print(x)
9 #        lyric_keywords = lyric_keywords + str(x)
10 #lyrics_resample
11
12 lyric_keywords

```

['love',	'knows',	'likes',
'like',	'yeah'],	'yeah',
'lovely things',	['tell',	'shits',
'hearts',	'let',	'got',
'heart',	'night like',	'nigga',
'humpty',	'baby',	'niggas',
'biz',	'chorus',	'gon hit like',
'kiss',	'dance',	'chorus',
'kissed',	'dancing',	'know'],
'got'],	'yeah',	['yeah',
['great',	'damn time',	'likes',
'near',	'know'],	'shit like',
'come',	['come',	'chorus',
'comes',	'comes',	'verse',
'kings',	'like',	'know',
'loves',	'baby',	'knows',
'place',	'liked loved',	'money got',
'people',	'feeling',	'ooh',
'blood',	'chorus talk',	'shits']]

Figure 93 - Calling the return_keywords function (a) to display an array of the top keywords for each year (b), (c), (d)

```

1 #Put lyric_keywords into a DataFrame and include lyrics_resample and hot100_resample
2 keywords_df = pd.DataFrame(lyric_keywords, columns = ['Keywords', 'Keywords', 'Keywords', 'Keywords', 'Keywords', 'Keywords'])
3 #keywords_df = pd.DataFrame(lyric_keywords)
4 #keywords_df = pd.DataFrame(list(lyric_keywords), columns = ['Keywords'])
5 #keywords_df = keywords_df.head(62)
6 keywords_df = keywords_df.set_index(lyrics_resample.index)
7 keywords_df['Lyrics'] = lyrics_resample
8 keywords_df['Sentiment'] = hot100_resample
9 keywords_df

```

Figure 94 - Creating a dataframe from the keywords yielded.

WeekID	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Keywords	Lyrics	Sentiment
1958-12-31	love	like	lovely things	hearts	heart	humpty	biz	kiss	kissed	got	I used to play around with hearts\nThat hasten...	0.076854
1959-12-31	great	near	come	comes	kings	loves	place	people	blood	borne	Does she love me, with all her heart\nShould I...	0.160398
1960-12-31	lovely	loved	alley	like	oop	ways	way	think love	know	going	Each night as I go wandering\nDown the street ...	0.221558
1961-12-31	love	loving	nigga	niggas	baby	dance step	nai	like	got	coming	[Intro: KRS-One]\nEvery once in a while\nYou g...	0.228065
1962-12-31	featuring	rap	rapping	def	jive	boy	boys	rock	rocking	rocks	A little bitty tear let me down, spoiled my ac...	0.098939
1968-12-31	black	like	baby	let	yeah	sweet	stay	good	love	coming	Baby, now that I've found you, I can't let you...	0.172540
1969-12-31	love	loves	lennon	songs	song	musical	musically	know	knows	way	When the morning sun\nStreaks across my room\n...	0.180237
1970-12-31	rivers	river	hey	feel	spot	spots	said	come	bad place	wanna	The changing of sunlight to moonlight\nReflect...	0.077246
1971-12-31	nadia	liked	like started	verse	verses	maxine	fucked	mike	got	cut	[Nicki Minaj]\nWho the hell is this?\nCalling ...	0.015035
1972-12-31	come	coming	play	playing	love	know loves	knows	good	city	got wife	I was slippin' into darkness\nWhen they took m...	0.159588
1973-12-31	lovely	love	saturday	makes	man	little extra	leave	alright	talk	looks	Write a letter, write another letter\nWrite a ...	0.152638
1977-12-31	blooms	hands	handed	handing	stephen	stephens	bloom appears flushed panting	zoe	head	heads	Do What You Want, Be What You Are\nIn Do what y...	0.045471
1978-12-31	love	lost	lola	chair	chairs	chorus	music	tony	need	feathers	Precious love, I'll give it to you\nBlue as th...	0.251662
1979-12-31	love	loving	want	verse	like	chorus	shine	shining	talk	till song night	Too many long conversations\nAnd no one is hea...	0.110870
1980-12-31	jeremy	james uses	patient	richard tests van	know	knows	knowing	time	ambulances getting	going	[Verse 1]\nEvery now and then I cry\nEvery nig...	0.103816
1981-12-31	verse	like	nah	away let	shine	chorus	know	night	angel	pre	Just yesterday\nI cast my eyes upon your lovel...	0.059668
1982-12-31	lovin	love	ont	like	talk	talks	talkinn	chorus	awav	kiss	I still can remember\nThe last	0.157072
1986-12-31	far	taken	heart	verse	chorus	love	loved	old	wall	walls	Headin' out on some uncharted path\nYou soon t...	0.147804
1987-12-31	loving	love baby	know	times	want life time	god knows	chorus	tell	feel	feeling	[Verse 1]\nWhen we met, always knew\nI would f...	0.186476
1990-12-31	love	swingin	chorus	time	times	step	alright	verse	like	baby	[Verse 1]\nMississippi\nin the middle of a dr...	0.088443
1991-12-31	come	coming	play like	love	loves	girls	got	time	times	know	[Verse 1]\nYou were so blind to let me go\nYou...	0.172874
1992-12-31	loved	loving	chorus	man	need love	like	know	brothers	comes	look	[Verse 1]\nYou said that you still loved me	0.256808
1993-12-31	like	love	got	girls	know	knows	verse	baby	cause	come	[Intro: Shabba Ranks & Johnny Gill]\nSh...	0.124866

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

1997-12-31	featuring	gots	tay	tsu	love	loving	loves	loved	got know	time	So alone\nAnd this room feels so cold\nNo one ...	0.120186
1998-12-31	got	gots	like tears	verse	chorus	know	nigga	man	love	loving	[Verse 1]\nYou taught me everything\nEverythin...	0.102551
1999-12-31	nigga	niggas	got	gots	chorus	verse	love	loved	knows	yeah	[Intro: Spoken]\nTo my heart\nTo my car\nTo my...	0.164161
2000-12-31	tell	let	night like	baby	chorus	dance	dancing	yeah	damn time	know	15 minutes-marc nelson\nDamn, what time is i...	0.074506
2001-12-31	come	comes	like	baby	liked loved	feeling	chorus talk	way	watch	watched	[Intro]\nExcuse me love, can I come dance with...	0.138253
2002-12-31	way	ways	come	coming	baby	yeah	love	loves	loving	verse	[Verse 1]\nHe is my king, he is my one\nYes, h...	0.155745
2007-12-31	like	got live	chorus	verse	yeah	know	things	thing	right	sorry times	Yeah\nCome on now\nHere we go\nLet's do it\nYe...	0.075261
2008-12-31	verse like	girl	got	chorus	knows	yeah	verses	know love	baby	little girls	[Intro]\n1, 2, 3, 4\nYeah (uh huh)\nPolo the D...	0.155396
2009-12-31	chorus	verse	got	love	loving	loves	wanna hold like	babies	knows	knowing	[Intro]\nMum mum mum mahn\nMum mum mum mahn\nMum...	0.126555
2010-12-31	chorus	love feels like	verse	verses	knows got	babies	girl	gone yeah	know	loved	[Verse 1]\nSo this is how it ends\nThis is who...	0.121034
2011-12-31	like	loved	sing likes	chorus	gots	yeah	verse	got girl	girls	niggas	[Verse 1]\nI remember saying I don't care either...	0.133970
2012-12-31	chorus	verse	love	loved	girl	let	loving yeah	like party hey	baby	know	[Verse 1: Taylor Swift]\nI remember tears stre...	0.158021
											[Intro: Nicki	
2015-12-31	got	nigga	like	yeah	chorus	verse	long niggas	ayy	know	knows	[Produced by Mike Dean & Paul McCartney]\nV...	0.102153
2016-12-31	featuring	gots	boy	niggas	knowing	knows	chorus	def	rapping	got hoes nigga	[Intro]\nNah, nah, nah\nCake by the ocean\n[...]	0.046511
2017-12-31	like	liked	yeah	verse	chorus let	jobs got	cabello	knowing	knows	time	[Verse 1]\nRight now we both know\nWe're Marin...	0.107435
2018-12-31	liking	likes	yeah	shits	got	nigga	niggas	gon hit like	chorus	know	[Intro: Cardi B & 21 Savage]\nBardi in a 'Rari...	0.014456
2019-12-31	yeah	likes	shit like	chorus	verse	know	knows	money got	ooh	shits	[Intro: 21 Savage]\nI love you\nTurn my headph...	0.030666

Figure 95 - Keywords dataframe displayed in tabular form including 'WeekID', 'Lyrics', and 'Sentiment' columns (a), (b), (c), (d), (e), (f), (g)

Appendix B – Lemmatisation vs Stemming

As well as the Glenn Medeiros song, there are many other examples of why lemmatisation, over stemming, is so important to this project. Another variation of ‘change’ is ‘changing’, which happens to be a 2014 number one for Sigala featuring Paloma Faith. Lemmatisation will view both of these words as the base word, ‘change’. Retired software developer, Martin Porter’s *Snowball* version of stemming, on the other hand, would trim both words to be left only with the stump, ‘chang’, which would obviously make no sense (Text Analysis - API & Demo | Text Analysis Online | TextAnalysis, 2021).

Text Analysis Result -- NLTK Porter Stemmer	
Original Text	Analysis Result
change changing	chang chang
Text Analysis Result -- spaCy Word Lemmatize	
Original Text	Analysis Result
change changing	change change

Figure 96 – Porter stemming (a), and spaCy lemmatisation (b) of ‘change’ and ‘changing’.

U2’s “Beautiful Day”, from the year 2000, would lose its positive polarity sentiment if used with Porter’s stemming (Text Analysis - API & Demo | Text Analysis Online | TextAnalysis, 2021).

Text Analysis Result -- NLTK Snowball Stemmer	
Original Text	Analysis Result
beautiful	beauti
Text Analysis Result -- spaCy Word Lemmatize	
Original Text	Analysis Result
beautiful	beautiful

Figure 97 – Porter stemming (a), and spaCy lemmatisation (b) of ‘beautiful’.

And it is not just song titles which would impact the sentiment analysis. Delving into the actual lyrics would show that “What A Wonderful World”, a 1968 chart-topper for Louis Armstrong, would lose context just from the usage of the word, ‘skies’, if using the aggressive Lancaster stemming algorithm.

“I see skies of blue and clouds of white” (Songfacts, 2021).

Text Analysis Result -- NLTK Lancaster Stemmer

Original Text

I see skies of blue and clouds of white

Analysis Result

i see **ski** of blu and cloud of whit

Text Analysis Result -- spaCy Word Lemmatize

Original Text

I see skies of blue and clouds of white

Analysis Result

i see **sky** of blue and cloud of white

Figure 98 – Lancaster stemming (a), and spaCy lemmatisation (b) of ‘skies’.

I doubt that Mr. Armstrong was signalling his intention to ‘ski’ when singing the song (spaCy Word Lemmatize - API & Demo | Text Analysis Online | TextAnalysis, 2021).

To add a modicum of balance, lemmatisation can also have its own flaws. There is proof of the word ‘best’ being downgraded, to just ‘good’ (Python - Lemmatization Approaches with Examples, 2020).

Code:

```

 import spacy
 nlp = spacy.load('en_core_web_sm')
 # Create a Doc object
 doc = nlp(u'the bats saw the cats with best stripes hanging upside down by their feet')
 # Create list of tokens from given string
tokens = []
for token in doc:
    tokens.append(token)

print(tokens)
#> [the, bats, saw, the, cats, with, best, stripes, hanging, upside, down, by, their, feet

lemmatized_sentence = " ".join([token.lemma_ for token in doc])

print(lemmatized_sentence)
#> the bat see the cat with good stripe hang upside down by -PRON- foot

```

In the above code, we observed that this approach was more powerful than our previous approaches as :

- Even **Pro-nouns** were detected. (identified by **-PRON-**)
- **Even best was changed to good.**

Figure 99 – spaCy lemmatisation of ‘best’ to ‘good’.

This scenario would mean that any songs in the dataframe including the word, ‘best’, would have their polarity score significantly skewed (TextBlob Sentiment Analysis - API & Demo | Text Analysis Online | TextAnalysis, 2021).

Text Analysis Result -- TextBlob Sentiment Analysis

Original Text	Analysis Result
best	Polarity = 1.000000, Subjectivity = 0.300000
Text Analysis Result -- TextBlob Sentiment Analysis	
Original Text	Analysis Result
good	Polarity = 0.700000, Subjectivity = 0.600000

Figure 100 – TextBlob polarity score of ‘best’ (a) and ‘good’ (b).

I know of at least five UK number ones containing the word, ‘best’, which would therefore have a lower polarity sentiment rating than they should have (Lyrics containing the term: best — Page #9, 2021).

- Abba – “Knowing Me, Knowing You” (1977)
- Abba – “Take A Chance On Me” (1978)
- Ronan Keating – “When You Say Nothing At All” (1999)
- Michael Andrews featuring Gary Jules – “Mad World” (2003)
- Jessie J – “Domino” (2012)

Although in Ronan Keating’s case, his pronunciation of the word as, “beshhht”, should probably justify his downgrading!

Appendix C – Web scraping and cleaning data in Power BI



Figure 101 - Power BI's simple method to retrieve data (a) from a website (b).

The screenshot shows the Power BI Navigator interface. On the left, the 'Navigator' pane lists various tables and suggested tables. 'Table 3' is selected and highlighted with a yellow box. On the right, the 'Table View' pane displays the data for 'Table 3' in a tabular format. The columns are labeled 'Column1', 'Column2', 'Column3', and 'Column4'. The data consists of 15 rows of song information, including artist names and song titles. At the bottom of the table view are 'Load', 'Transform Data', and 'Cancel' buttons.

Column1	Column2	Column3	Column4
No.	Artist[nb 2]	Single[nb 2]	Re
1952	1952	1952	15
1	Al Martino	"Here in My Heart"	Ce
1953	1953	1953	15
2	Jo Stafford	"You Belong to Me"	Cc
3	Kay Starr	"Comes A-Long A-Love"	Ce
4	Eddie Fisher	"Outside of Heaven"	Hf
5	Perry Como	"Don't Let the Stars Get in Your Eyes"	Hf
6	Guy Mitchell	"She Wears Red Feathers"	Cc
7	The Stargazers	"Broken Wings"	Dt
8	Lita Roza	"(How Much Is) That Doggie in the Window?"	Dt
9	Frankie Laine	"I Believe" †	Pf
10	Eddie Fisher	"I'm Walking Behind You"	Hf
re	Frankie Laine	"I Believe" †	Pf
11	Mantovani	"The Song from Moulin Rouge"	Dt
re	Frankie Laine	"I Believe" †	Pf
12	Guy Mitchell	"Look At That Girl"	Pf
13	Frankie Laine	"Hey Joe"	Pf
14	David Whitfield	"Answer Me"	Dt
15	Frankie Laine	"Answer Me"	Pf
re	David Whitfield	"Answer Me"	Dt

Figure 102 - Wikipedia's 1950s data in tabular form.

Figure 103 shows a screenshot of the Power BI desktop interface. A table named 'Table 3' is displayed with 31 rows of data. The columns are labeled A_C_Column1 through A_C_Column6. The data includes columns for 'No.', 'Artist[nb 2]', 'Single[nb 2]', 'Record label[nb 3]', 'Week ending date[nb 4]', and 'Weeks at number one[nb 2]'. Rows 14, 16, 17, 18, 19, 20, 21, 28, 29, and 30 are highlighted in yellow.

Figure 103 - Wikipedia data brought into Power BI before cleaning.

Figure 104 shows a screenshot of the Power BI ribbon. The 'Table' tab is selected. The ribbon includes various data transformation and analysis tools like Transpose, Reverse Rows, Detect Data Type, Pivot Column, and Run R/Python script.

Figure 104 - Giving meaningful heading names.

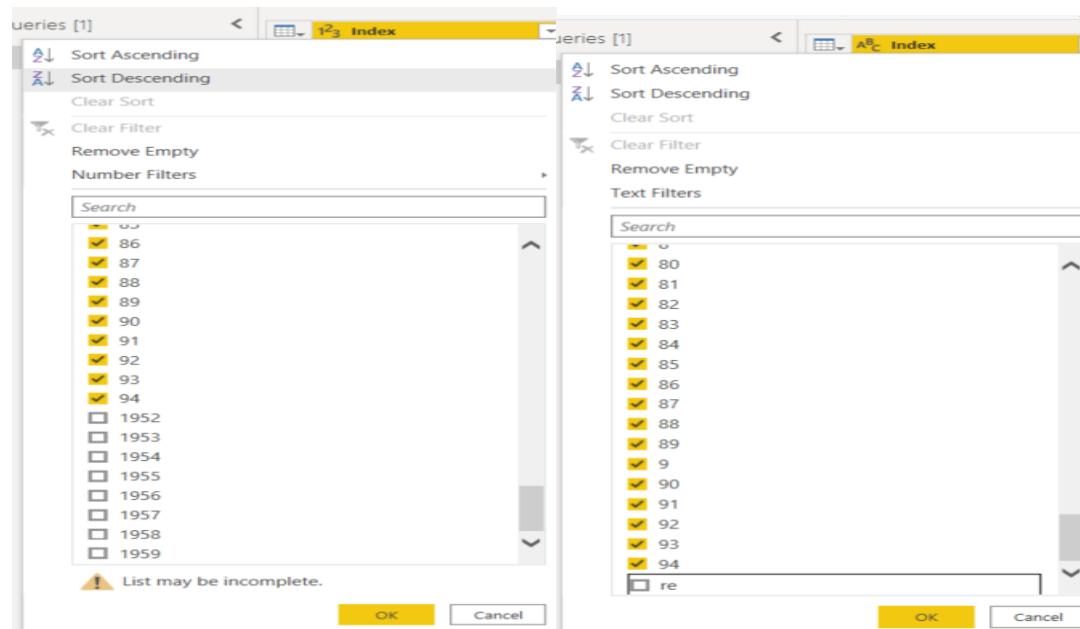


Figure 105 - Omitting the years (a) to ensure correct index numbers (b).

Cross Symbol Quick Guide

Symbol Name	Cross Sign
Symbol	†
Alt Code	0134
Shortcut for Windows	Alt+0134

Figure 106 - Using Google to find the correct keyboard code for the 'Cross sign' symbol (Alt shortcuts list, 2021).

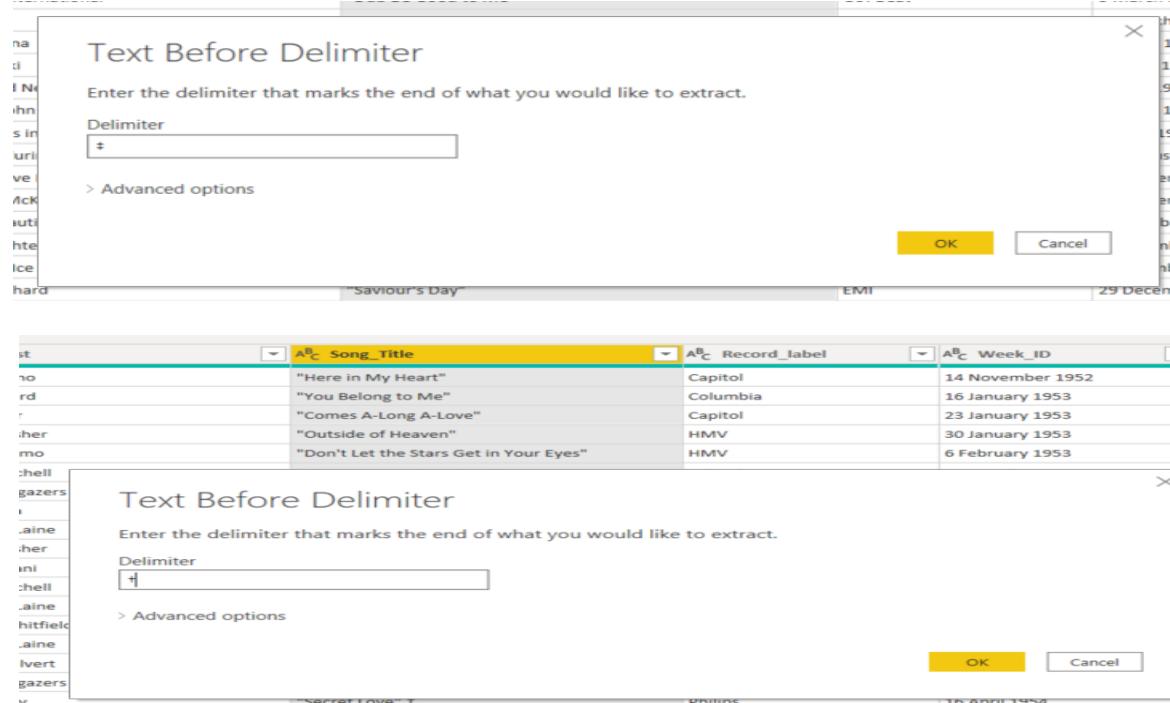


Figure 107 - Using 'Text before delimiter' function to remove 'double-cross' (a) & 'cross' (b) symbols.



A B C Song_Title
"Here in My Heart"
"You Belong to Me"
"Comes A-Long A-Love"
"Outside of Heaven"
"Don't Let the Stars Get in Your Eyes"
"She Wears Red Feathers"
"Broken Wings"
"(How Much Is) That Doggie in the Window?"
"I Believe" ↑
"I'm Walking Behind You"
"The Song from Moulin Rouge"
"Look At That Girl"
"Hey Joe"
"Answer Me"
"Answer Me"
"Oh Mein Papa"
"I See the Moon"
"Secret Love" ↑
"Such a Night"

Figure 108 - Two songs before (a) and after (b) removal of symbols.

Replace Values

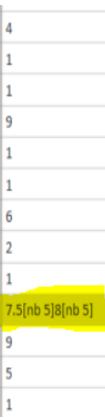
Replace one value with another in the selected columns.

Value To Find: 7.5[nb 5]8[nb 5]

Replace With: 8

Advanced options

OK Cancel



A B C Song_Title
"Here in My Heart"
"You Belong to Me"
"Comes A-Long A-Love"
"Outside of Heaven"
"Don't Let the Stars Get in Your Eyes"
"She Wears Red Feathers"
"Broken Wings"
"(How Much Is) That Doggie in the Window?"
"8" ↑
"I'm Walking Behind You"
"The Song from Moulin Rouge"
"Look At That Girl"
"Hey Joe"
"Answer Me"
"Answer Me"
"Oh Mein Papa"
"I See the Moon"
"Secret Love" ↑
"Such a Night"

Replace Values

Replace one value with another in the selected columns.

Value To Find: 910

Replace With: 9

Advanced options

OK Cancel



A B C Song_Title
"Such a Night"
Philips
50 April 1954
2 July 1954
910
1

Figure 109 - Replacing floating-point numbers with an integer (a), and an erroneous value (b).

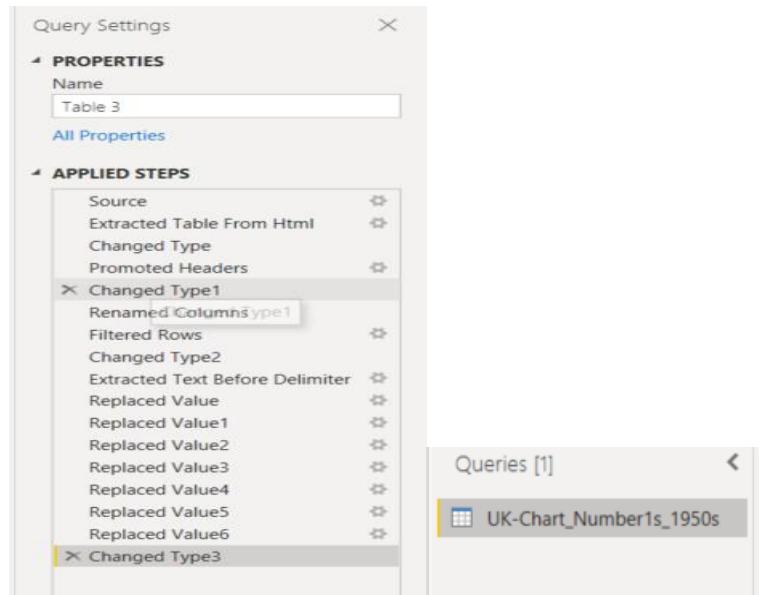


Figure 110 - All applied modifications to the original data table (a). Labelling the dataset with a meaningful name (b).

Appendix D – Google Colab run of UK dataset

```

1 from textblob import TextBlob
2 #install modules
3 !pip install lyricsgenius
4 #!pip install -U gensim
5 !pip install gensim==3.8.3
6 !pip install -U spacy
7 !pip install textblob
8 !python -m textblob.download_corpora
9 !pip install rake-nltk
10
11 #import packages
12 from gensim.summarization import keywords
13 import spacy
14 import nltk
15 from nltk.corpus import stopwords
16 nltk.download('stopwords')
17 stopwords = set(stopwords.words('english'))
18 from spacy.lang.en import English
19 nlp = English()
20 nlp.max_length = 10000000
21 import lyricsgenius
22 import pandas as pd
23 from textblob import TextBlob

```

Figure 111 - Installing libraries and packages.

```

[ ] 1 #define Genius API authentication
2 genius = lyricsgenius.Genius('Y98TmRsxVPIMhJm4z47BPxkzzGqbKBjbwN8Pgqdv0r76K3VITQGULjm8pnTSOHIB')
3 lyric_test = genius.search_song('Shake It Off', 'Taylor Swift').lyrics
4 print(lyric_test)

[Chorus]
'Cause the players gonna play, play, play, play, play
And the haters gonna hate, hate, hate, hate, hate
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off
Heartbreakers gonna break, break, break, break, break
And the fakers gonna fake, fake, fake, fake, fake
Baby, I'm just gonna shake, shake, shake, shake, shake
I shake it off, I shake it off

```

Figure 112 - Testing connection between lyricsgenius and Genius API

```

1 from google.colab import files
2 uploaded = files.upload()

C Choose File: UK-Chart...020_FYP.csv
• UK-Chart_Number1s_1952_2020_FYP.csv(application/vnd.ms-excel)- 113751 bytes, last modified: 5/17/2021 - 100% done
Saving UK-Chart_Number1s_1952_2020_FYP.csv to UK-Chart_Number1s_1952_2020_FYP.csv

[ ] 1 import io
2 #import UK Chart No.1s dataset
3 UK1s_df = pd.read_csv(io.BytesIO(uploaded['UK-Chart_Number1s_1952_2020_FYP.csv']))
4 # Dataset is now stored in a Pandas Dataframe
5 #UK1s_df.drop_duplicates(subset='Song title' and 'Artist(s)', inplace = True) #remove duplicate occurrences of songs
6 #UK1s_df = UK1s_df.head(40)
7 UK1s_df.reset_index()

   index  Index          Artist        Song_Title  Record_label  Week_ID  Year  Country  Weeks_at_number_one
0       0      1    Al Martino  "Here In My Heart"    Capitol  14/11/1952  1952  AMERICA            9
1       1      2    Jo Stafford  "You Belong to Me"  Columbia  16/01/1953  1953  UNITED KINGDOM         1
2       2      3      Kay Starr  "Comes A-Long A-Love"    Capitol  23/01/1953  1953  AMERICA            1
3       3      4    Eddie Fisher  "Outside of Heaven"     HMV  30/01/1953  1953  AMERICA            1
4       4      5    Perry Como  "Don't Let the Stars Get in Your Eyes"    HMV  06/02/1953  1953  AMERICA           5
...
1355  1355  1356  Shawn Mendes and Camila Cabello  "Señorita"  EMI, Syco  18/07/2019  2019  CANADA/CUBA         1
1356  1356  1357  Ed Sheeran featuring Khalid  "Beautiful People"  Asylum  25/07/2019  2019  UNITED KINGDOM         1
1357  1357  1358  Ed Sheeran featuring Stormzy  "Take Me Back to London"  Asylum  05/09/2019  2019  UNITED KINGDOM           5
1358  1358  1359  Tones and I  "Dance Monkey"  Bad Batch  10/10/2019  2019  AUSTRALIA          11
1359  1359  1360  LadBaby  "I Love Sausage Rolls"  Frtyfve  26/12/2019  2019  UNITED KINGDOM         1
1360 rows x 9 columns

```

Figure 113 - Importing csv file from local machine.

```
[ ] 1 #Function to return lyrics of each song using Genius API
2 def get_lyrics(title, artist):
3     try:
4         |  return genius.search_song(title, artist).lyrics
5     except:
6         |  return 'not found'
7
8 def get_lyric_sentiment(lyrics):
9     analysis = TextBlob(lyrics)
10    return analysis.sentiment.polarity
11
12 # # Function to preprocess text
13 # def preprocess(text):
14 #     # Create Doc object
15 #     doc = nlp(text, disable=['ner', 'parser'])
16 #     # Generate lemmas
17 #     lemmas = [token.lemma_ for token in doc]
18 #     # Remove stopwords and non-alphabetic characters
19 #     a_lemmas = [lemma for lemma in lemmas
20 #                 if lemma.isalpha() and lemma not in stopwords]
21
22 #     return ' '.join(a_lemmas)
23
24 # """Extract Keywords from text"""
25 # def return_keywords(texts):
26 #     xkeywords = []
27 #     values = keywords(text=preprocess(texts), split='\n', scores=True)
28 #     for x in values[:5]:
29 #         xkeywords.append(x[0])
30 #     try:
31 #         return xkeywords
32 #     except:
33 #         return "no content"
```

Figure 114 - Functions being defined for use further on (note that 'preprocess' and 'keywords' functions commented out from prototype as this will be done in Power BI on this occasion).

```
▶ 1 #Use get_lyrics funcion to get lyrics for every song in dataset
2 lyrics = UK1s_df.apply(lambda row: get_lyrics(row['Song_Title'], row['Artist']), axis =1)
3 UK1s_df['Lyrics'] = lyrics
4 UK1s_df = UK1s_df.drop(UK1s_df[UK1s_df['Lyrics'] == 'not found'].index) #drop rows where lyrics are not found on Genius
5 UK1s_df
Done.
⇒ Searching for ""Promises"" by Calvin Harris and Sam Smith...
Done.
Searching for ""Funky Friday"" by Dave featuring Fredo...
Done.
Searching for ""Shallow"" by Lady Gaga and Bradley Cooper...
Done.
Searching for ""Thank U, Next"" by Ariana Grande...
Done.
Searching for ""We Built This City"" by LadBaby...
Done.
```

Figure 115 - get_lyrics function connecting with Genius API.

Searching for ""Señorita"" by Shawn Mendes and Camila Cabello...
 Done.
 Searching for ""Beautiful People"" by Ed Sheeran featuring Khalid...
 Done.
 Searching for ""Take Me Back to London"" by Ed Sheeran featuring Stormzy...
 Done.
 Searching for ""Dance Monkey"" by Tones and I...
 Done.
 Searching for ""I Love Sausage Rolls"" by LadBaby...
 Done.

Index	Artist	Song_Title	Record_label	Week_ID	Year	Country	Weeks_at_number_one	Lyrics
0	Al Martino	"Here in My Heart"	Capitol	14/11/1952	1952	AMERICA	9	Here in my heart I'm alone, I'm so lonely\nHer...
1	Jo Stafford	"You Belong to Me"	Columbia	16/01/1953	1953	UNITED KINGDOM	1	See the pyramids along the Nile\nWatch the sun...
2	Kay Starr	"Comes A-Long A-Love"	Capitol	23/01/1953	1953	AMERICA	1	The last tall son of Lot and Bellcent,\nAnd t...
3	Eddie Fisher	"Outside of Heaven"	HMV	30/01/1953	1953	AMERICA	1	I pass your house with misty eyes\nThere stand...
4	Perry Como	"Don't Let the Stars Get in Your Eyes"	HMV	06/02/1953	1953	AMERICA	5	Don't let the stars get in your eyes\nDon't le...
...
1355	Shawn Mendes and Camila Cabello	"Señorita"	EMI, Syco	18/07/2019	2019	CANADA/CUBA	1	[Chorus: Camila Cabello]\nI love it when you c...
1356	Ed Sheeran featuring Khalid	"Beautiful People"	Asylum	25/07/2019	2019	UNITED KINGDOM	1	[Intro: Ed Sheeran]\nWe are, we are, we are\n...
1357	Ed Sheeran featuring Stormzy	"Take Me Back to London"	Asylum	05/09/2019	2019	UNITED KINGDOM	5	[Chorus: Ed Sheeran & Stormzy]\nJet plane head...
1358	Tones and I	"Dance Monkey"	Bad Batch	10/10/2019	2019	AUSTRALIA	11	[Verse 1]\nThey say, 'Oh my God, I see the way...
1359	LadBaby	"I Love Sausage Rolls"	FiftyFive	26/12/2019	2019	UNITED KINGDOM	1	[Verse 1: Roxanne]\nSaw him standing there in ...

Figure 116 - Lyrics being imported from Genius API and being added as a new 'Lyrics' column.

```
[ ] 1 #Use get_lyric_sentiment to get sentiment score for all the song lyrics
2 sentiment = UK1s_df.apply(lambda row: get_lyric_sentiment(row['Lyrics']), axis = 1)
3 UK1s_df['Sentiment'] = sentiment
4 print(UK1s_df.columns.to_list())

['Index', 'Artist', 'Song_Title', 'Record_label', 'Week_ID', 'Year', 'Country', 'Weeks_at_number_one', 'Lyrics', 'Sentiment']
```

Figure 117 get_lyric_sentiment function being called to classify polarity on lyrics and create 'Sentiment' column. (Using 'print' to check all fields are entered).

Index	Artist	Song_Title	Record_label	Week_ID	Year	Country	Weeks_at_number_one	Lyrics	Sentiment
0	Al Martino	"Here in My Heart"	Capitol	14/11/1952	1952	AMERICA	9	Here in my heart I'm alone, I'm so lonely\nHer...	0.055000
1	Jo Stafford	"You Belong to Me"	Columbia	16/01/1953	1953	UNITED KINGDOM	1	See the pyramids along the Nile\nWatch the sun...	0.214286
2	Kay Starr	"Comes A-Long A-Love"	Capitol	23/01/1953	1953	AMERICA	1	The last tall son of Lot and Bellcent,\nAnd t...	0.153604
3	Eddie Fisher	"Outside of Heaven"	HMV	30/01/1953	1953	AMERICA	1	I pass your house with misty eyes\nThere stand...	0.195382
4	Perry Como	"Don't Let the Stars Get in Your Eyes"	HMV	06/02/1953	1953	AMERICA	5	Don't let the stars get in your eyes\nDon't le...	0.489784
5	Guy Mitchell	"She Wears Red Feathers"	Columbia	13/03/1953	1953	UNITED KINGDOM	4	(She wears red feathers and a hoo-hooy skirt...)	0.249237
6	The Stargazers	"Broken Wings"	Decca	10/04/1953	1953	UNITED KINGDOM	1	[Hook]\nNo more pencils, no more books\nI'll buil...	-0.074474
7	Lita Roza	"(How Much Is) That Doggie in the Window?"	Decca	17/04/1953	1953	AMERICA	1	How much is that doggy in the window\nWoof woof...	0.151136
8	Frankie Laine	"I Believe"	Philips	24/04/1953	1953	AMERICA	9	I believe for every drop of rain that falls\nA...	0.259091
9	Eddie Fisher	"I'm Walking Behind You"	HMV	26/06/1953	1953	AMERICA	1	I'm walking behind you on your wedding day\nAn...	-0.096364
11	Guy Mitchell	"Look At That Girl"	Philips	11/09/1953	1953	UNITED KINGDOM	6	[Round 1: Charron]\nOver the years, we've part...	-0.004034
12	Frankie Laine	"Hey Joe"	Philips	23/10/1953	1953	AMERICA	2	Hey Joe\nWhere'd you find that pretty girly\nW...	0.207692
13	David Whitfield	"Answer Me"	Decca	06/11/1953	1953	UNITED KINGDOM	1	Letter from Mr. Abel James, with Notes of my L...	0.138547
14	Frankie Laine	"Answer Me"	Philips	13/11/1953	1953	AMERICA	8	Answer me, Lord above:\nJust what sin have I b...	0.150000
15	Eddie Calvert	"Oh Mein Papa"	Columbia	08/01/1954	1954	AMERICA	9	Oh, my pa-pa, to me he was so wonderful\nOh, m...	0.470000
16	The Stargazers	"I See the Moon"	Decca	12/03/1954	1954	UNITED KINGDOM	5	Over the mountain over the sea\nBack where my ...	0.295789
17	Doris Day	"Secret Love"	Philips	16/04/1954	1954	UNITED KINGDOM	1	[Verse 1]\nOnce I had a secret love\nThat live...	0.165476
18	Johnnie Ray	"Such a Night"	Philips	30/04/1954	1954	AMERICA	1	Such a night\nIt was a night, ooo what a night...	0.197727
19	David Whitfield	"Cara Mia"	Decca	02/07/1954	1954	UNITED KINGDOM	9	Cara mia why, must we say goodbye?\nEach time ...	0.487500

Figure 118 - Double checking all fields are implemented in table.

```
[ ] 1 UK1s_df.dtypes
```

Index	int64
Artist	object
Song_Title	object
Record_Label	object
Week_ID	object
Year	int64
Country	object
Weeks_at_number_one	int64
Lyrics	object
Sentiment	float64
dtype:	object

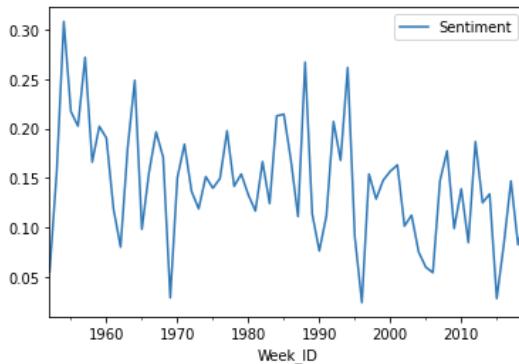
Figure 119 - Checking data types of columns are correct.

```
[ ] 1 #Set the index of the dataframe to the WeekID. This sets us up to resample dataframe based on time
2 UK1s_df['Week_ID'] = pd.to_datetime(UK1s_df['Week_ID'],infer_datetime_format=True)
3 UK1s_df = UK1s_df.sort_values(by='Week_ID')
4 UK1s_df = UK1s_df.reset_index(drop=True)
5 UK1s_df = UK1s_df.set_index('Week_ID')
```

```
[ ] 1 #Resample dataframe sentiment column by year. This arranges the mean lyric sentiment for every year
2 UK1s_resample = UK1s_df['Sentiment'].resample('y').mean()
```

```
[ ] 1 #Plot a line graph of lyric sentiment over time
2 UK1s_resample.plot(label='Sentiment', legend=True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f9a90cb2e50>



```
[ ] 1 #Calculate the mean percentage change of lyric sentiment
2 UK1s_resample.pct_change().mean()
```

0.2537757864974681

Figure 120 - Setting the index of the dataframe to be sorted by 'WeekID' column (a), resampling the dataframe to calculate the mean sentiment of the lyrics for each year (b), plotting a line graph to display the sentiment of each year (c), and resampling the lyric sentiment to calculate the mean percentage change (d).

```
[ ] 1 # #Use return_keywords function on lyrics_resample to get the top 20 keywords for each year
2 # lyric_keywords = [return_keywords(x[1]) for x in lyrics_resample.items()]
3
4 # #lyric_keywords = [return_keywords(x[1])]
5 # #lyric_keywords = ""
6 # #for x in lyrics_resample.items():
7 # #    #print(x)
8 # #if x != "0":
9 # #    #print(x)
10 # #lyric_keywords = lyric_keywords + str(x)
11 # #lyric_keywords = lyric_keywords.append(x)
12 # #    #lyric_keywords = [return_keywords(x[1])]
13 # #lyrics_resample
14
15 # lyric_keywords
16
17 # #print(lyric_keywords)
18
19 # #UK1s_df['Year']

[ ] 1 # #Put lyric_keywords into a DataFrame and include lyrics_resample and hot100_resample
2 # keywords_df = pd.DataFrame(lyric_keywords, columns = ['Keywords', 'Keywords', 'Keywords'],
3 # #keywords_df = pd.DataFrame(lyric_keywords)
4 # #keywords_df = pd.DataFrame(list(lyric_keywords), columns = ['Keywords'])
5 # #keywords_df = keywords_df.head(62)
6 # keywords_df = keywords_df.set_index(lyrics_resample.index)
7 # keywords_df['Lyrics'] = lyrics_resample
8 # keywords_df['Sentiment'] = hot100_resample
9 # keywords_df
```

Figure 121 - An essential section of coding for the prototype is commented out for the UK data run as Power BI will perform this function.

```
[ ] 1 from google.colab import files
2 UK1s_df.to_csv('UK_Number1s_Lyrics_FYP.csv')
3 files.download('UK_Number1s_Lyrics_FYP.csv')
```

Figure 122 - Saving amended dataset back to csv on local drive.

Appendix E – Sentiment Analysis in Power BI

The full sentiment analysis report is intended to be used as an interactive tool for the reader and can be found at this link [Sentiment Analysis in Power BI](#).

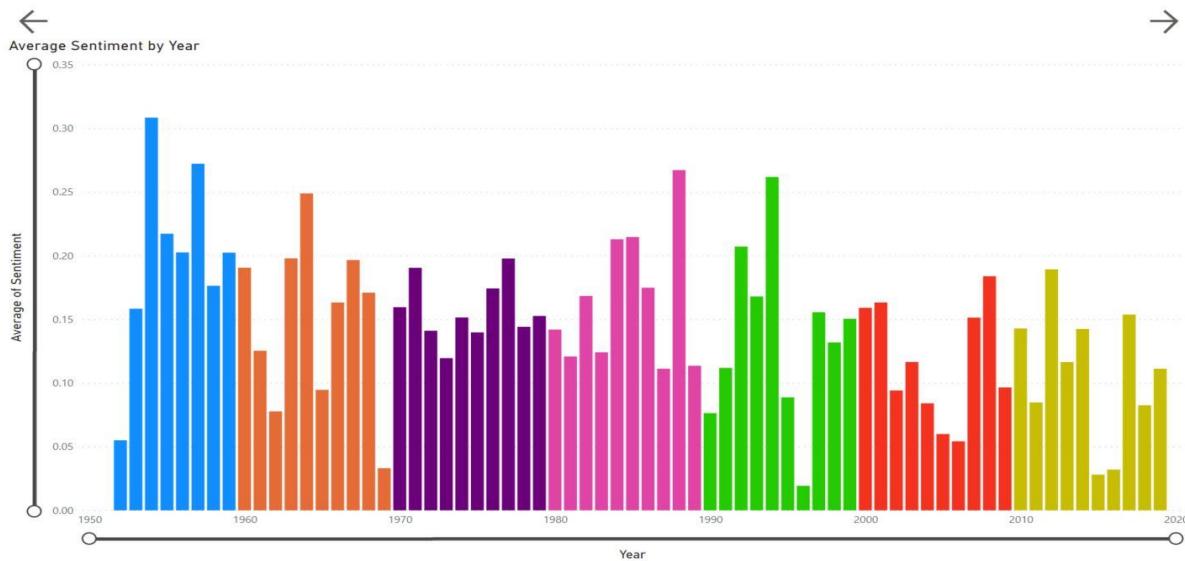


Figure 123 - Average sentiment by year (stacked column chart).

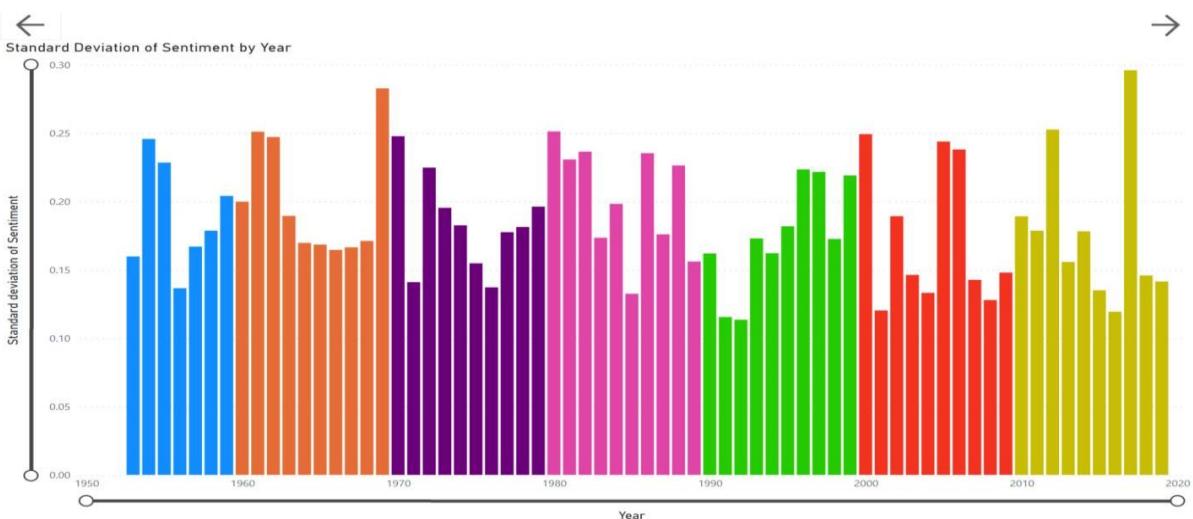


Figure 124 - Standard deviation of sentiment by year (stacked column chart).

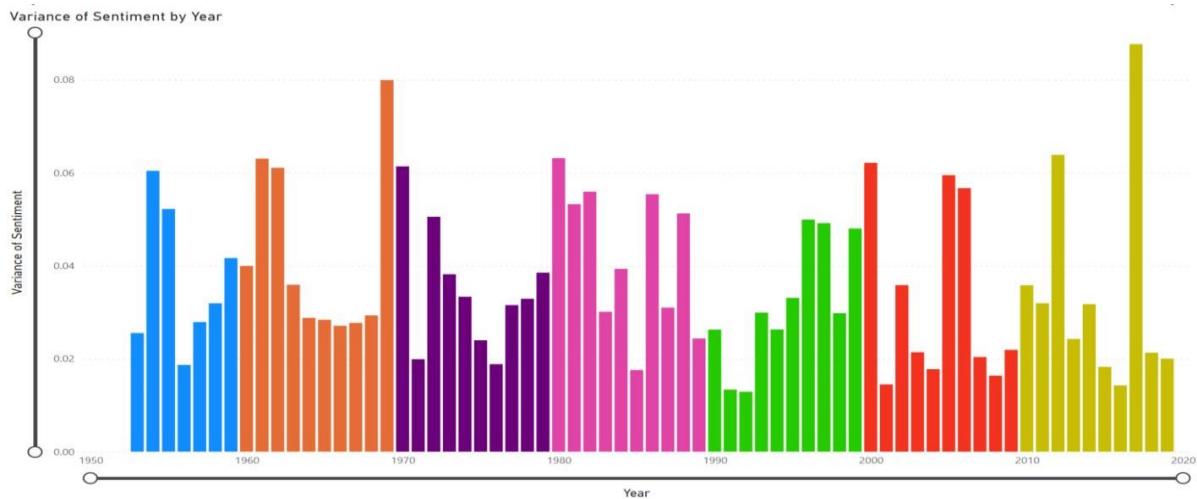


Figure 125 - Variance of sentiment by year (stacked column chart).

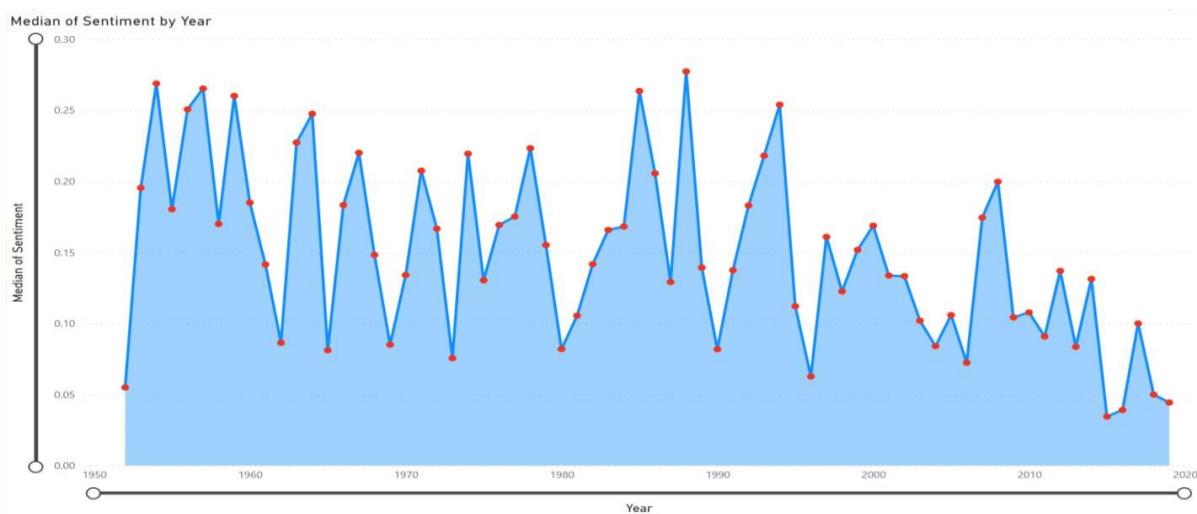


Figure 126 - Median of sentiment by year (area chart).

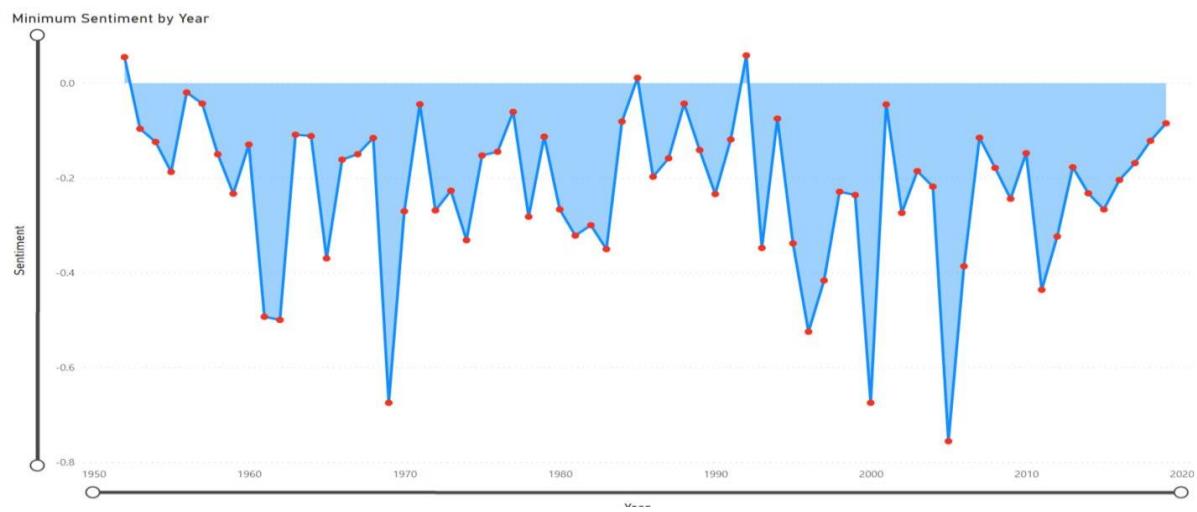


Figure 127 - Minimum sentiment by year (area chart).

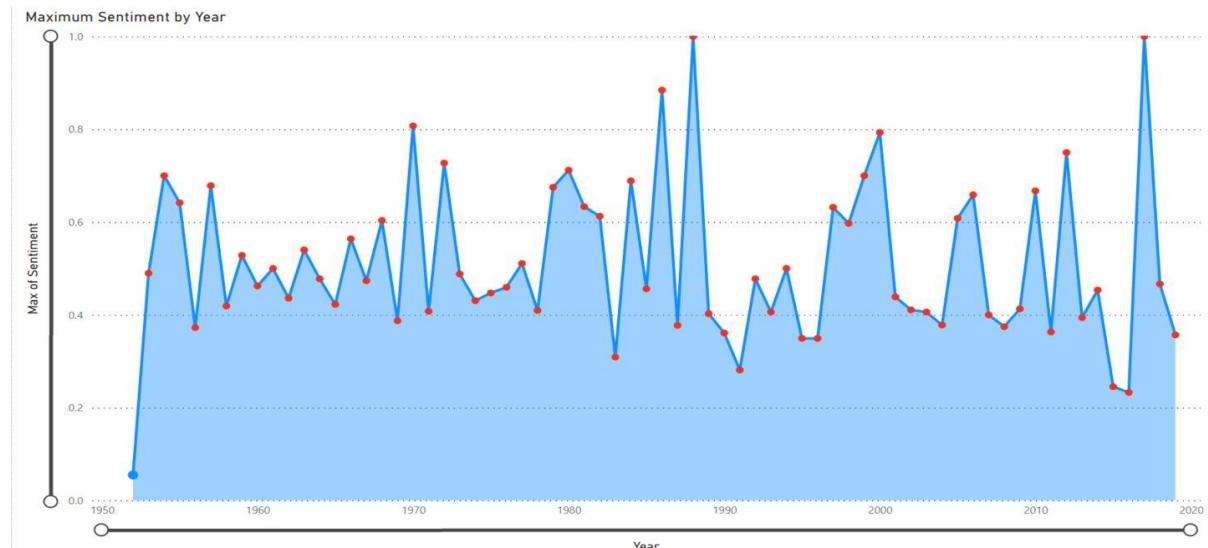


Figure 128 - Maximum sentiment by year (area chart).

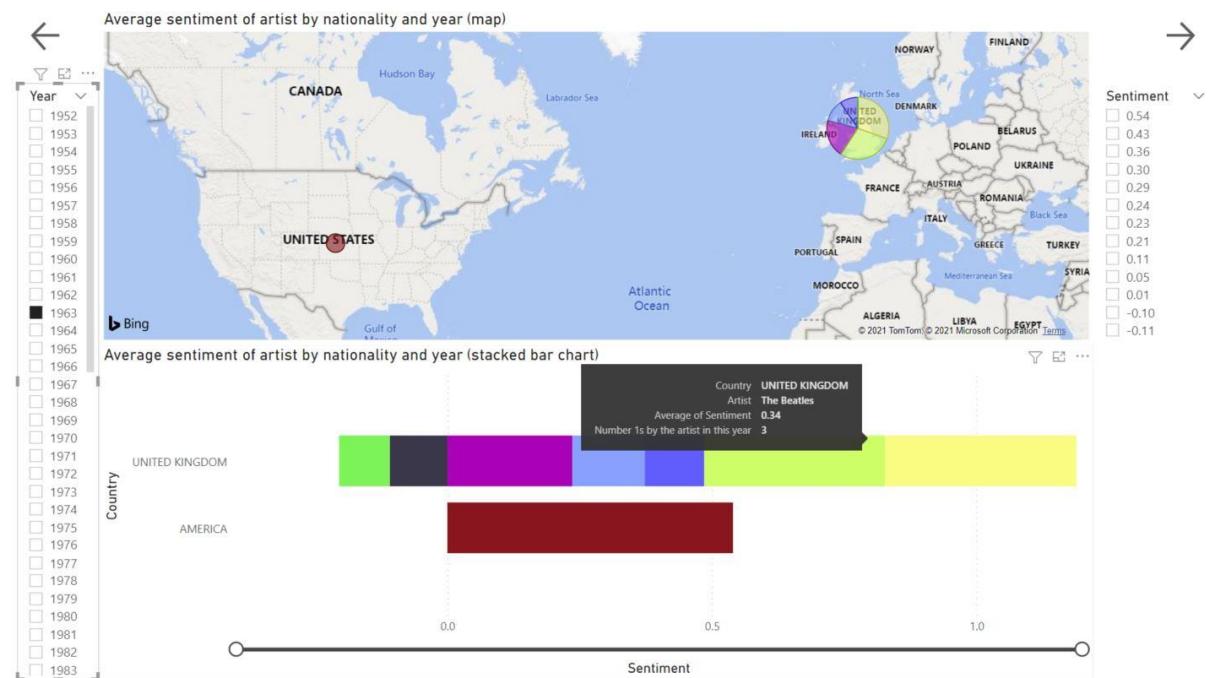


Figure 129 - Average sentiment of artists by year and nationality. The Beatles' 1963 seen here as an example (Map) and (stacked bar chart).

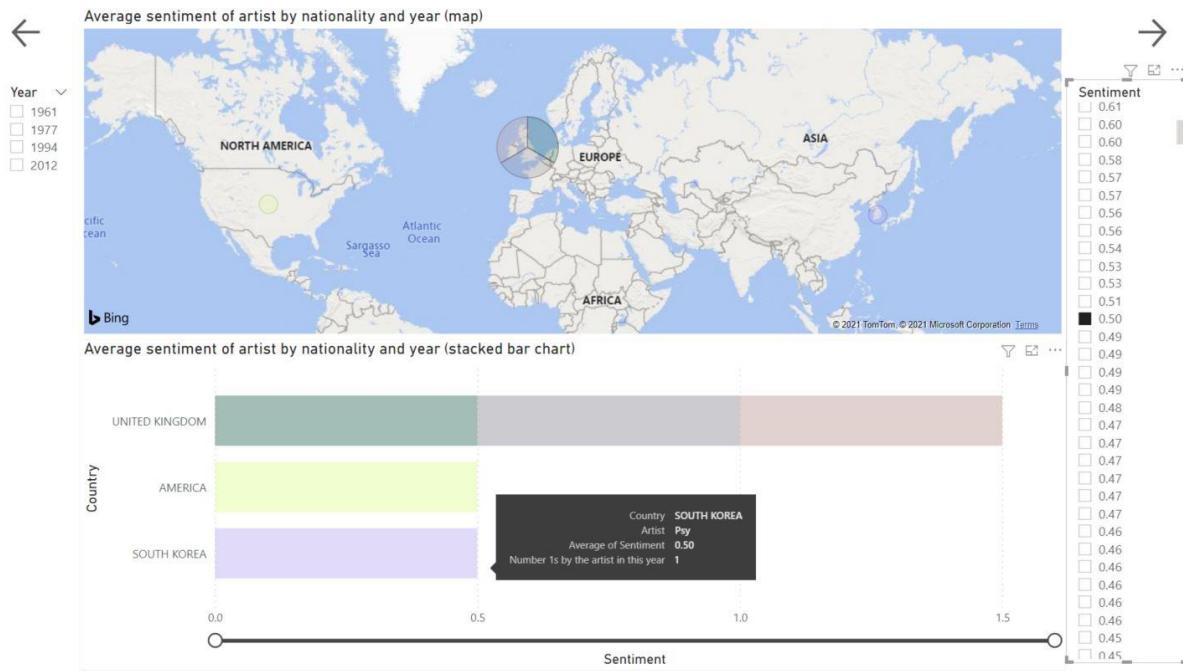


Figure 130 - Using the filter to show all artists with a sentiment of 0.50 by year and nationality. South Korean, Psy, is the example shown here through his 'one-hit-wonder' global success, "Gangnam Style" (Map) and (stacked bar chart).

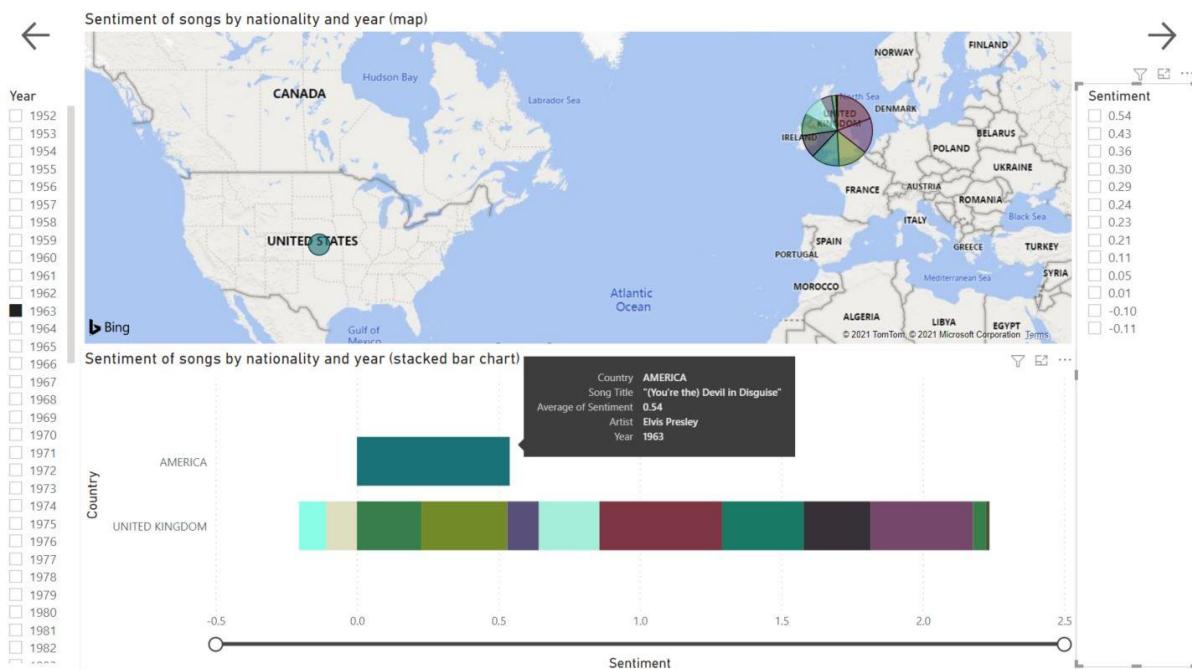


Figure 131 - Sentiment of songs by artist, year, and nationality. Elvis Presley had the highest polarity score of 1963 with, "(You're the) Devil in Disguise" (Map) and (stacked bar chart).

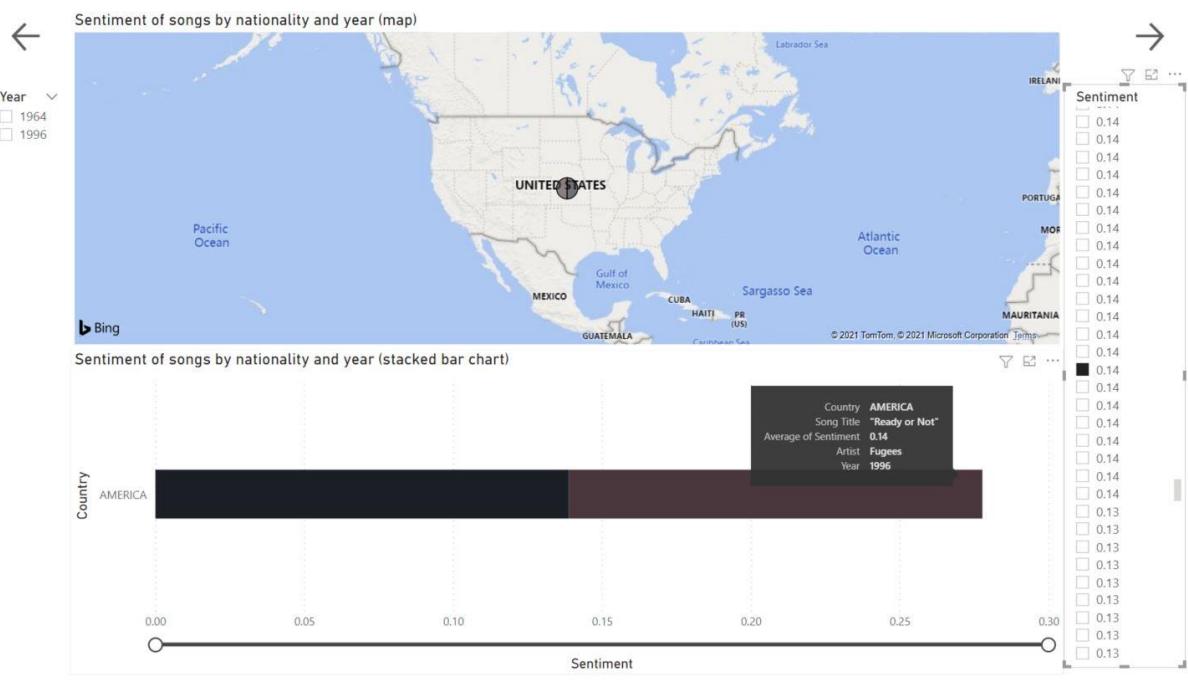


Figure 132 - Sentiment of songs by artist, year, and nationality. 1996 hit, "Ready or Not" receiving a sentiment rating of 0.14 (Map) and (stacked bar chart).

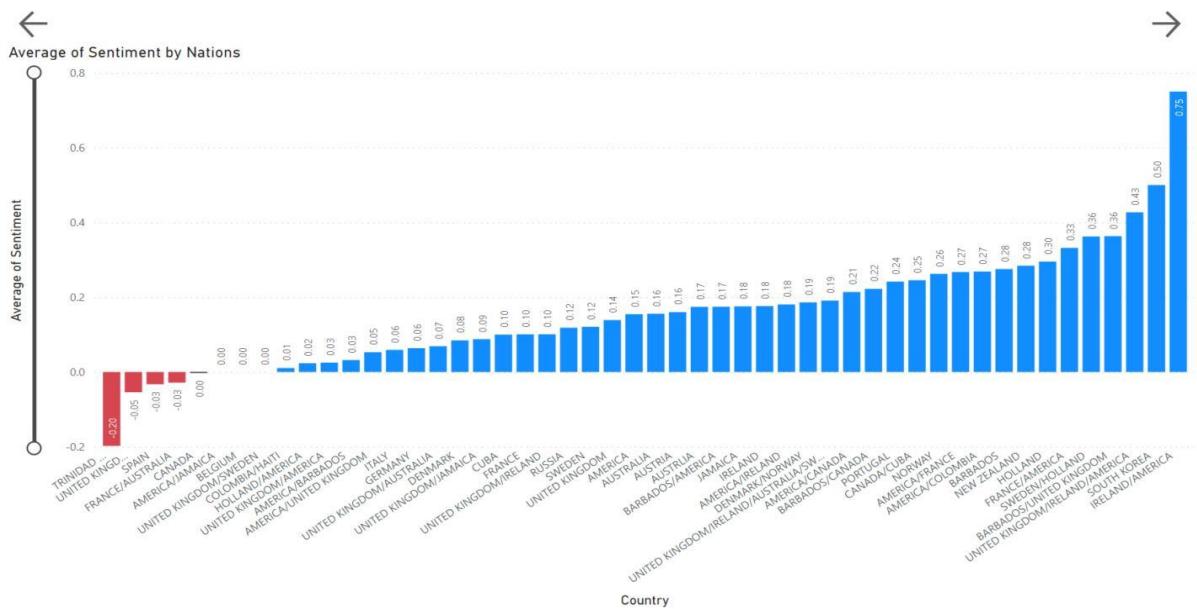


Figure 133 - Average sentiment of each nation to have a UK number one (stacked column chart)



Figure 134 - 1954 was the year which reached the highest average sentiment (word cloud).

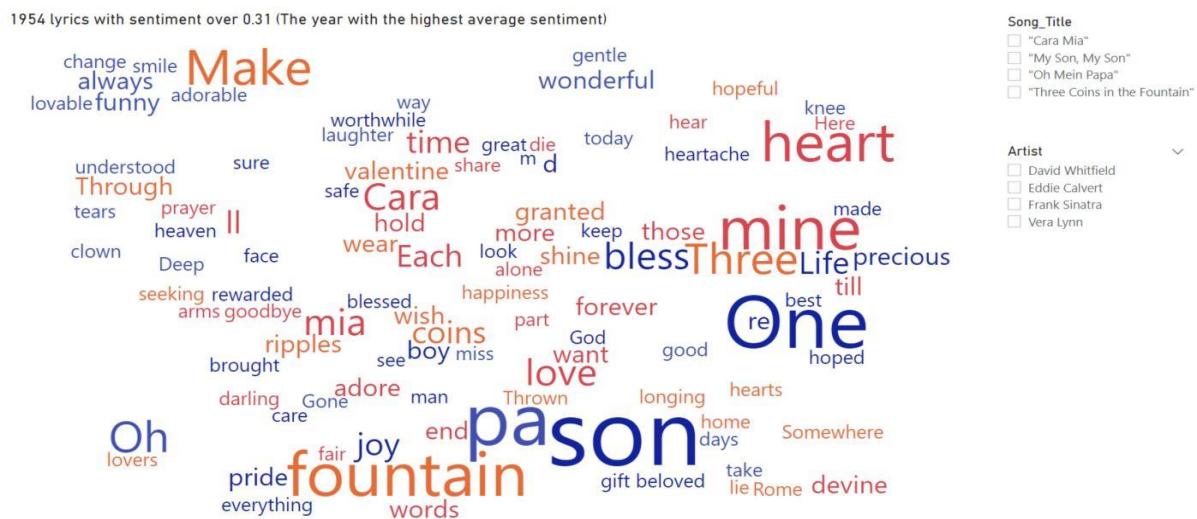


Figure 135 - Songs and their lyrics from 1954 which at least matched the 0.31 sentiment score (word cloud).

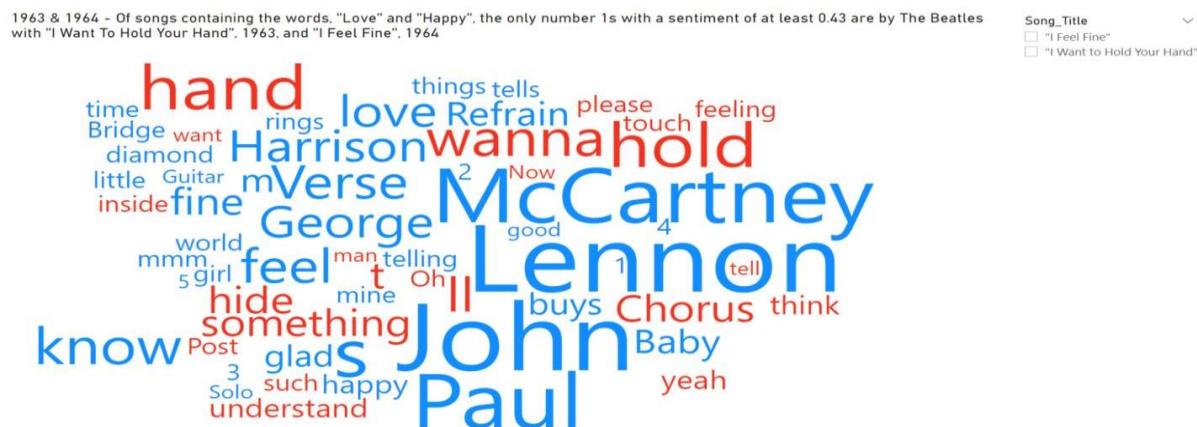


Figure 136 - The highest sentiment scoring songs containing the lyrics, 'Love' and 'Happy' both came from The Beatles in 1963, and 1964, respectively (word cloud).

2012 - Of songs containing the words, "Love" and "Happy", the only number 1 in the past decade with a sentiment of at least 0.2 is Gotye's "Somebody That I used to Know" from 2012

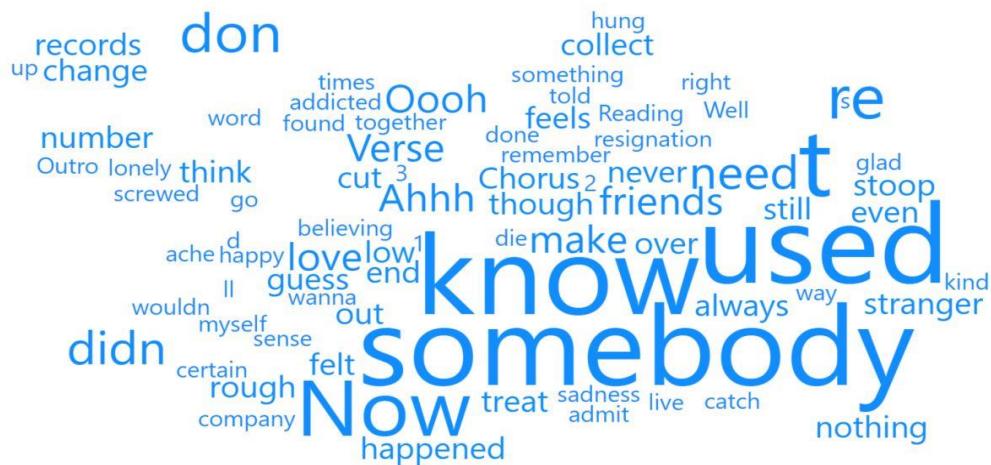


Figure 137 - Almost 50 years later, and Gotye is the highest scoring artist with a song containing, "Love" and "Happy" in the past decade. His polarity score is not even half of The Beatles previously mentioned efforts (word cloud).

2012 – The only year in the past decade with an average sentiment of at least 0.18 (*Word Cloud includes the other years which match this average)

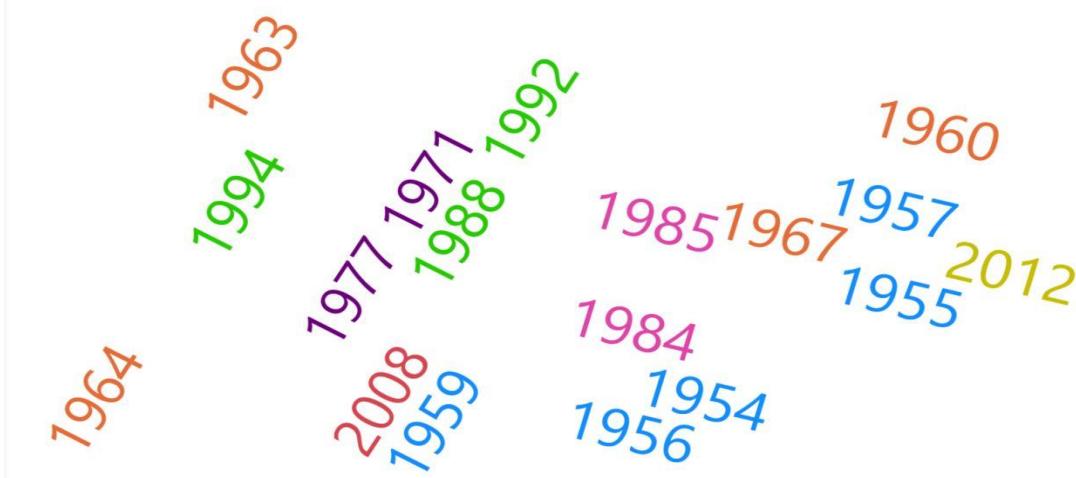


Figure 138 - 2012 is also the only year of the past decade to reach an average sentiment of at least 0.18 (word cloud).

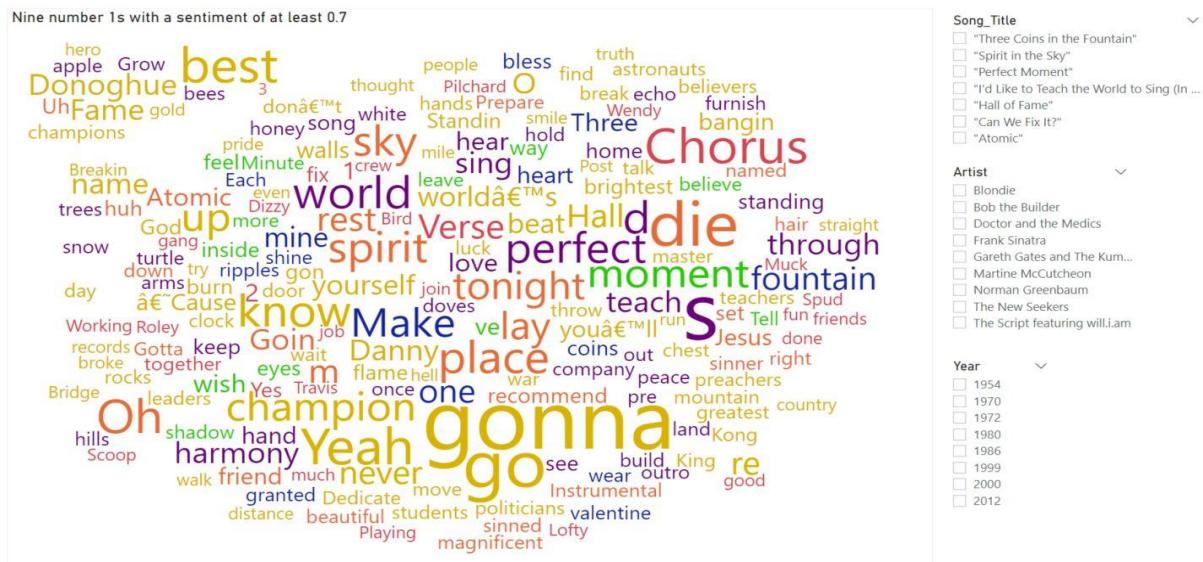


Figure 139 - Nine number one hits with a, relatively, high sentiment score (word cloud)



Figure 140 - 1996 was the lowest rating year of sentiment (word cloud).



Figure 141 - Song lyrics matching or less than the 0.02 sentiment score of 1996 (word cloud).

1996 - Of songs containing the word, "nigga", the first number 1 with a POSITIVE sentiment value is "Ready or Not" by The Fugees

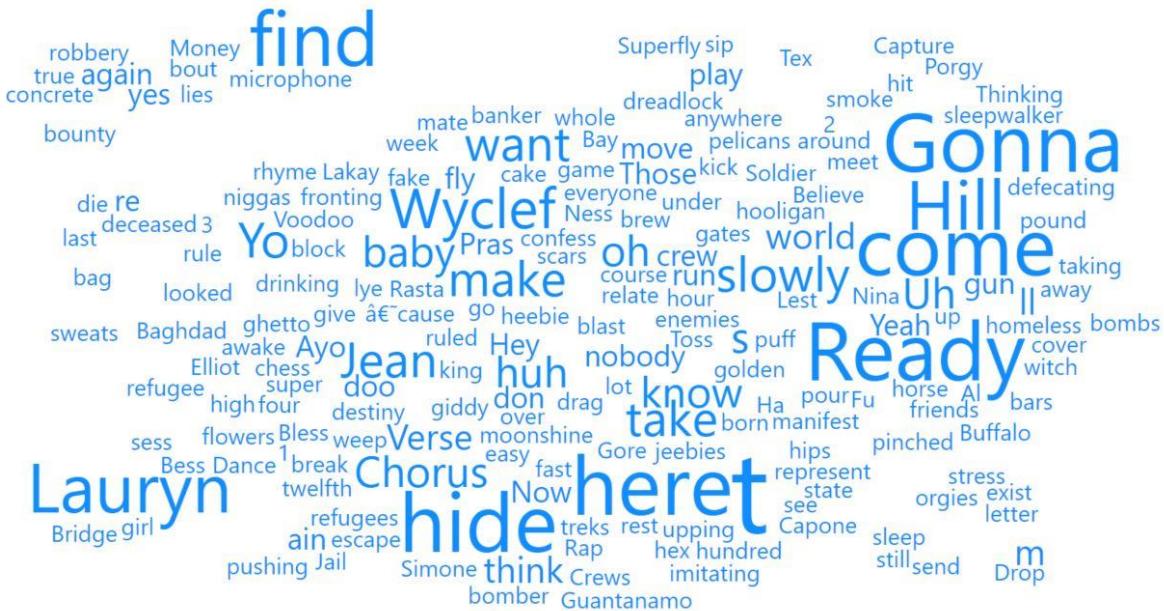


Figure 142 - 1996 included the first case of the word, “nigga”, appearing in a number one of, surprisingly, high polarity (word cloud).

2006 - Of songs containing, "nigga", the first number 1 with a NEGATIVE sentiment is Notorious B.I.G with "Nasty Girl". 2006

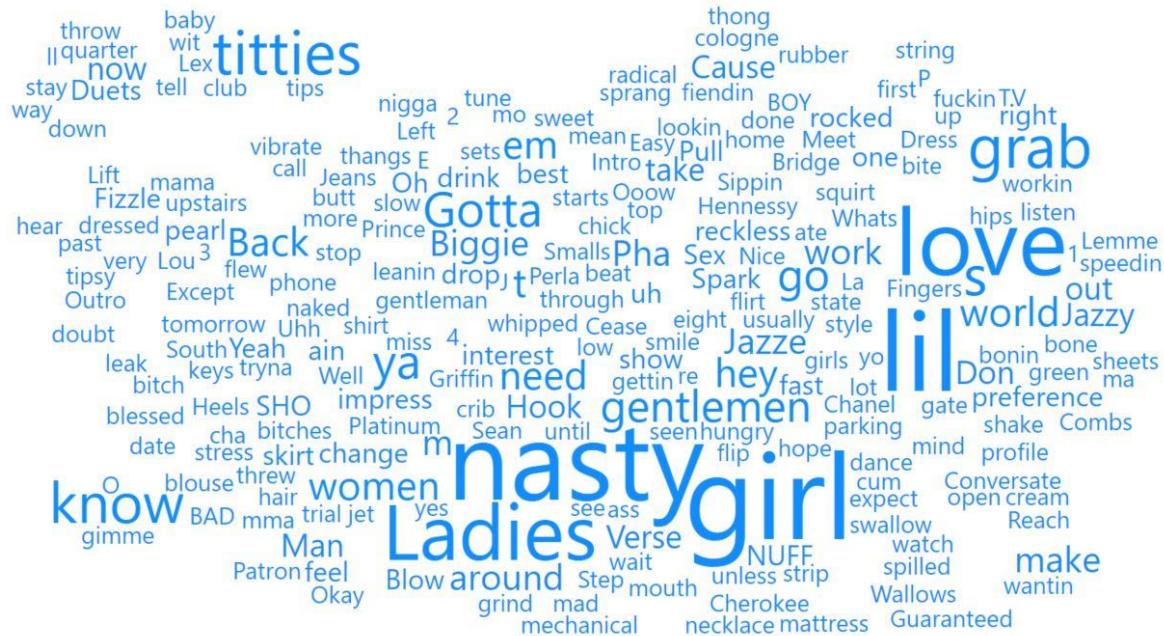


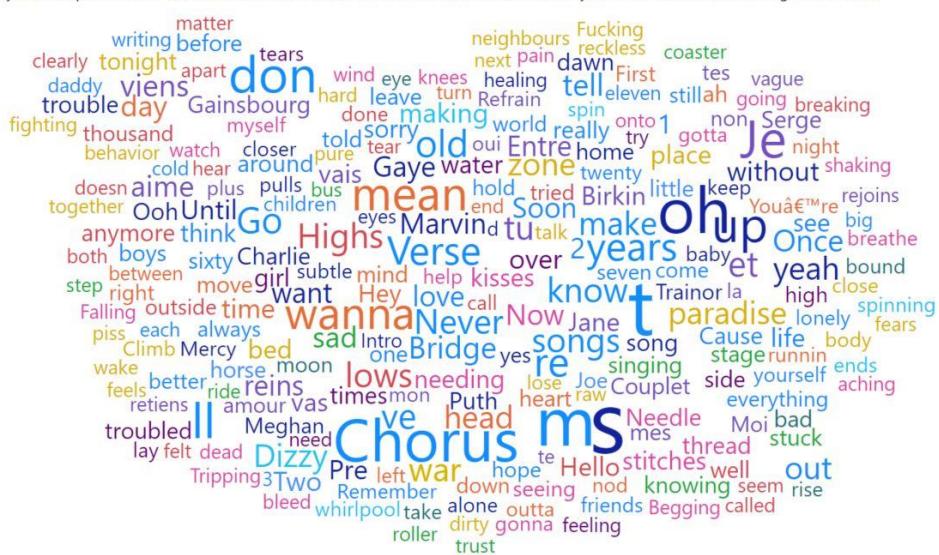
Figure 143 - "Nasty Girl", by Notorious B.I.G., became the first number one to use, "nigga", which had a negative polarity score (word cloud).

The four years with the lowest average sentiment (between 0.02-0.04)



Figure 144 - The four lowest scoring years over the history of the charts, of which two have occurred very recently (word cloud).

Lyrics sample of 1969, 1996, 2015, and 2016 with sentiment under 0.04 (The four years with the lowest average sentiment)



Song_Title
"7 Years"
"A Bridge over You"
"Bad Moon Rising"
"Dizzy"
"Hello"
"I Took a Pill in Ibiza"
"Je t'aime... moi non plus"
"Marvin Gaye"
"Pillowtalk"
"Stitches"
"Two Little Boys"
"What Do You Mean?"

Artist
Adele
Charlie Puth featuring Meghan Trainor
Creedence Clearwater Revival
Jane Birkin and Serge Gainsbourg
Justin Bieber
Lewisham and Greenwich
Lukas Graham
Mike Posner
Rolf Harris
Shawn Mendes
Tommy Roe
ZAYN

Year
1969
2015
2016

Figure 145 - Song lyrics from the four lowest sentiment rating years (word cloud).

1995 - Of songs containing the words, "Ass" and "Pussy", the first number 1 with a negative sentiment is "Don't Stop" by The Outhere Brothers



Figure 146 - "Don't Stop" by The Outhere Brothers, from 1994, was the first number one to contain both of the words, "Ass" and "Pussy" (word cloud).

2000 - Of songs containing the words, "Shit" and "Fuck", the first case of both words appearing in a number 1 was in the year 2000, Eminem achieved this feat twice, with "Stan" and "The Real Slim Shady.



Figure 147 - Eminem's year 2000 included two number ones containing the words, "Shit" and "Fuck" (word cloud).

2018 - Another song containing "Shit" and "Fuck" ("Freaky Friday" by Lil Dicky ft Chris Brown is full of expletives)

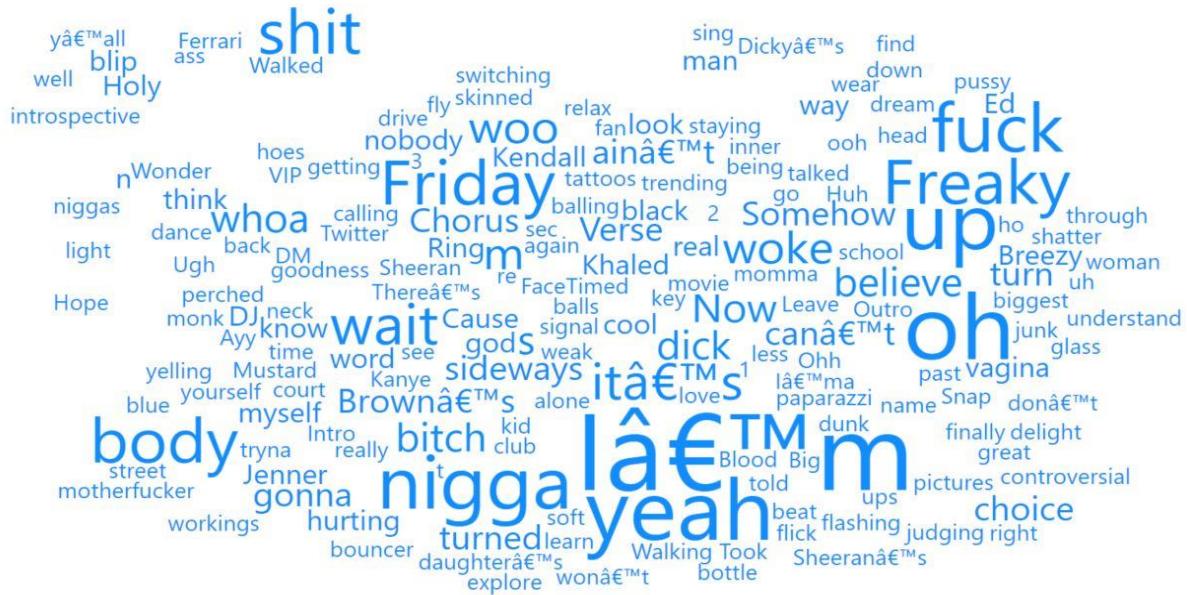


Figure 148 - "Shit and "Fuck" were amongst many obscene words littered throughout "Freaky Friday", a 2018 number one for Lil' Dicky and Chris Brown (word cloud).

2018 - Another number 1 from this year containing "Shit" and "Fuck" (2018's "Nice For What" by Drake is full of expletives)



Figure 149 - Drake's, "Nice For What" was another 2018 hit using bad language (word cloud).

Sentiment comparison of milestone years in the UK charts

← 1954 lyrics with sentiment over 0.31 (The year with the highest average sentiment)



→ 1996 lyrics with sentiment under 0.02 (The year with the lowest average sentiment)



2015 & 2016 lyrics with sentiment under 0.04 (The years with the lowest average sentiment in the past decade, and in the bottom four years for sentiment of all time)



2018 lyrics with sentiment under 0.08 (A recent year with clear evidence of containing explicit language)

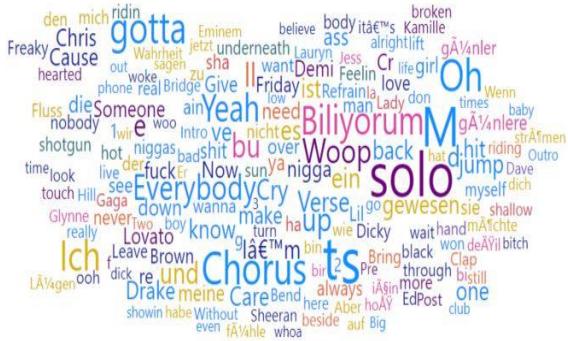
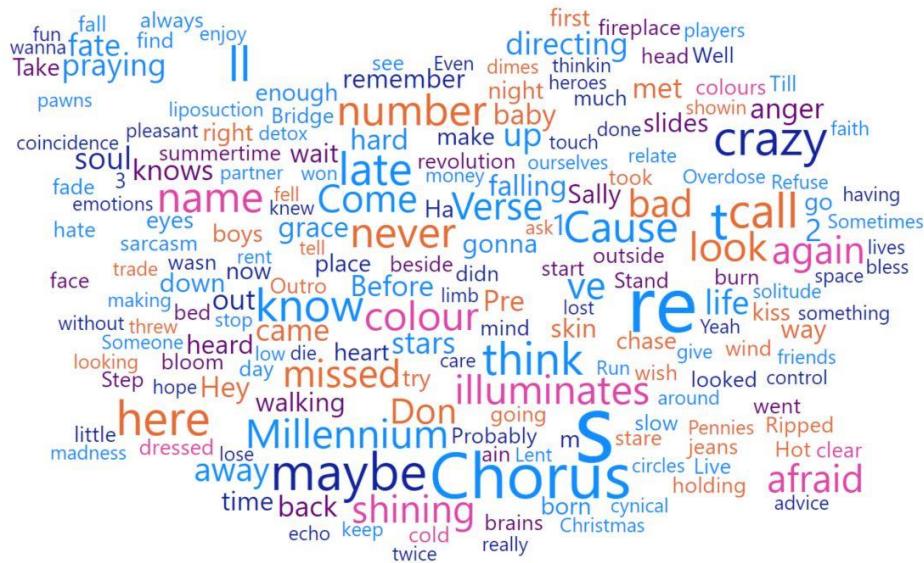


Figure 150 - A snapshot of the evolution of song lyrics over the history of the UK charts (word cloud).

Five number 1s with a surprising sentiment of less than -0.2



Song_Title
<input type="checkbox"/> "Call Me Maybe"
<input type="checkbox"/> "Crazy"
<input type="checkbox"/> "Don't Look Back in Anger"
<input type="checkbox"/> "Millennium"
<input type="checkbox"/> "Spectrum (Say My Name) ...

Artist
<input type="checkbox"/> Carly Rae Jepsen
<input type="checkbox"/> Florence and the Machine
<input type="checkbox"/> Gnarls Barkley
<input type="checkbox"/> Oasis
<input type="checkbox"/> Robbie Williams

Year
<input type="checkbox"/> 1996
<input type="checkbox"/> 1998
<input type="checkbox"/> 2006
<input type="checkbox"/> 2012

Figure 151 - Five number ones with a surprisingly negative rating (word cloud).

Sentiment analysis of UK number ones (1952-2020) to detect correlation between lyrics and society

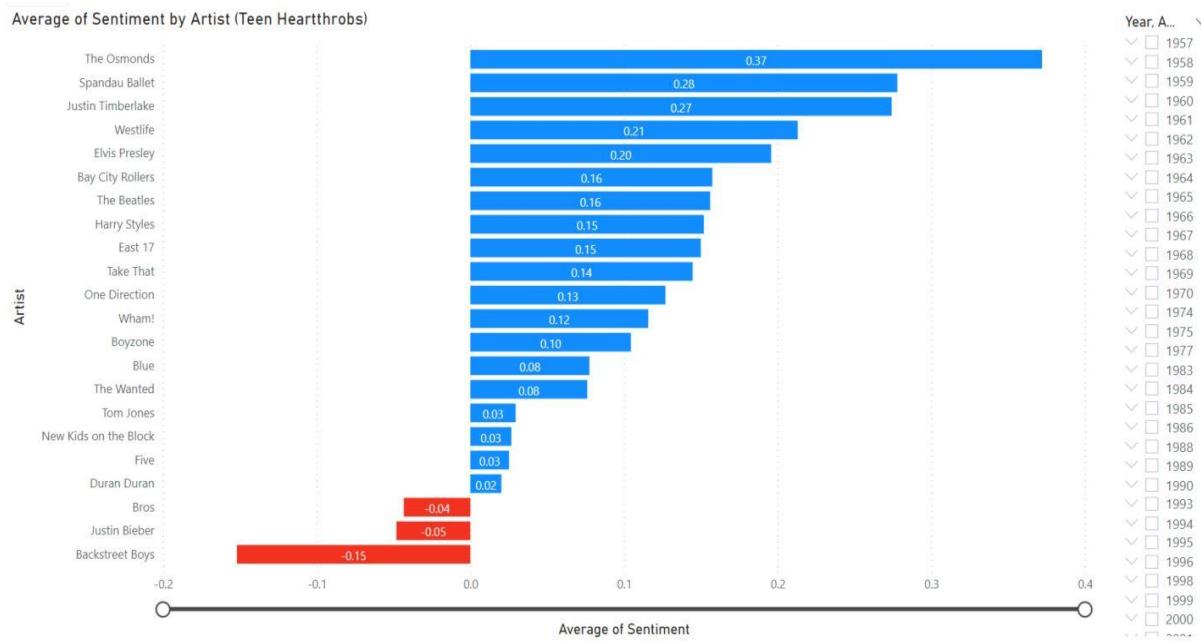


Figure 152 - A closer look at the average sentiment of artists deemed as "Heartthrobs" by many (stacked bar chart).

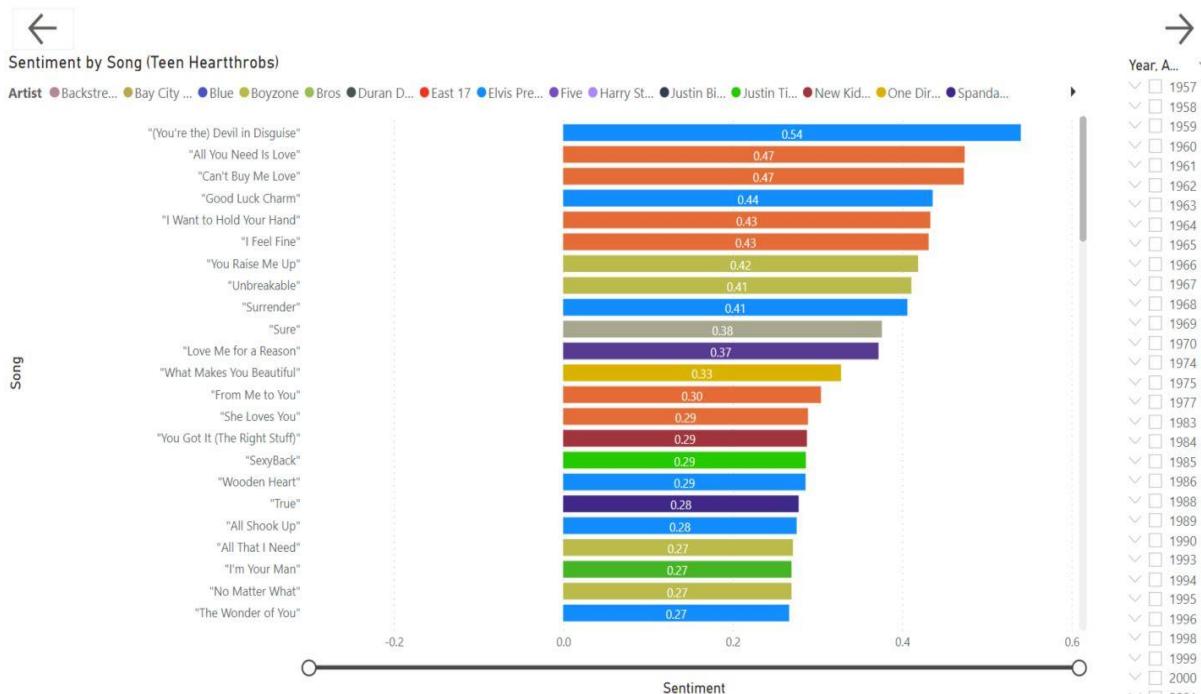


Figure 153 - "Heartthrob" songs with a positive sentiment value (stacked bar chart).

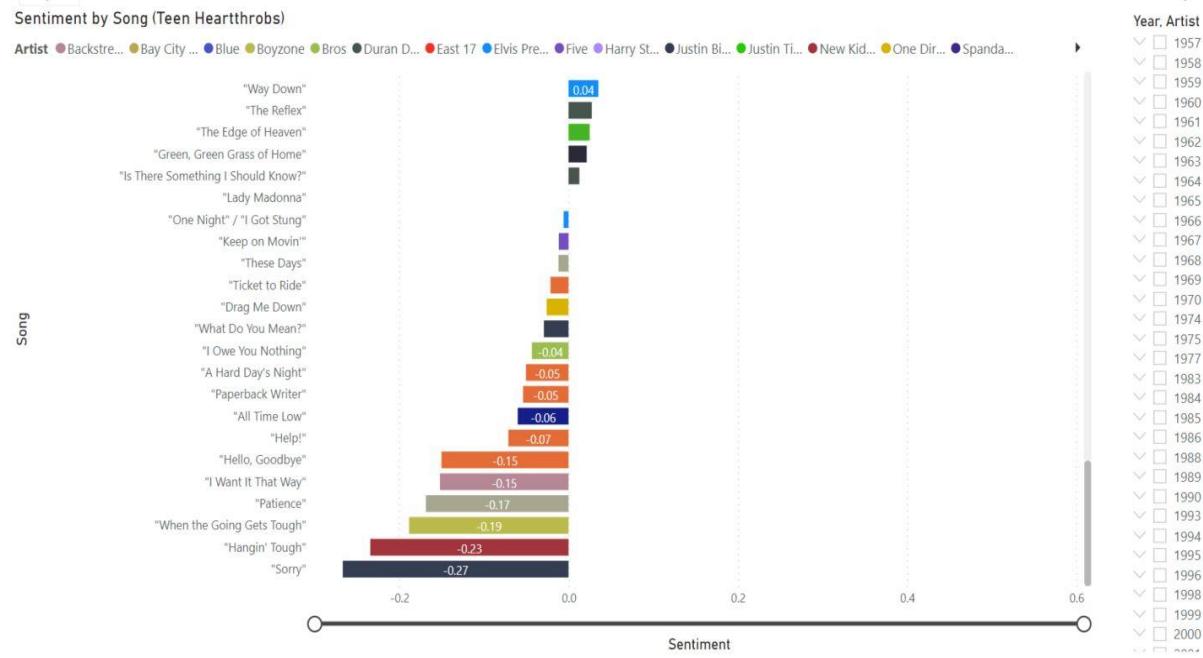


Figure 154 - "Heartthrob" songs with a negative sentiment value (stacked bar chart).

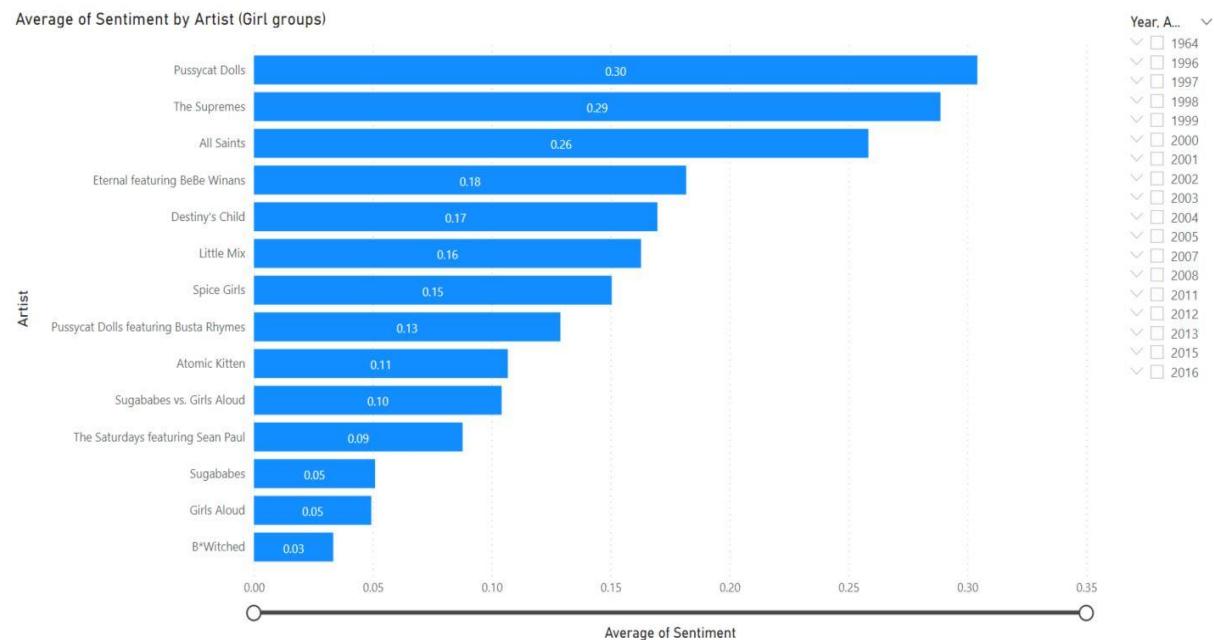


Figure 155 - A closer look at the average sentiment of artists deemed as "Girl groups" (stacked bar chart).

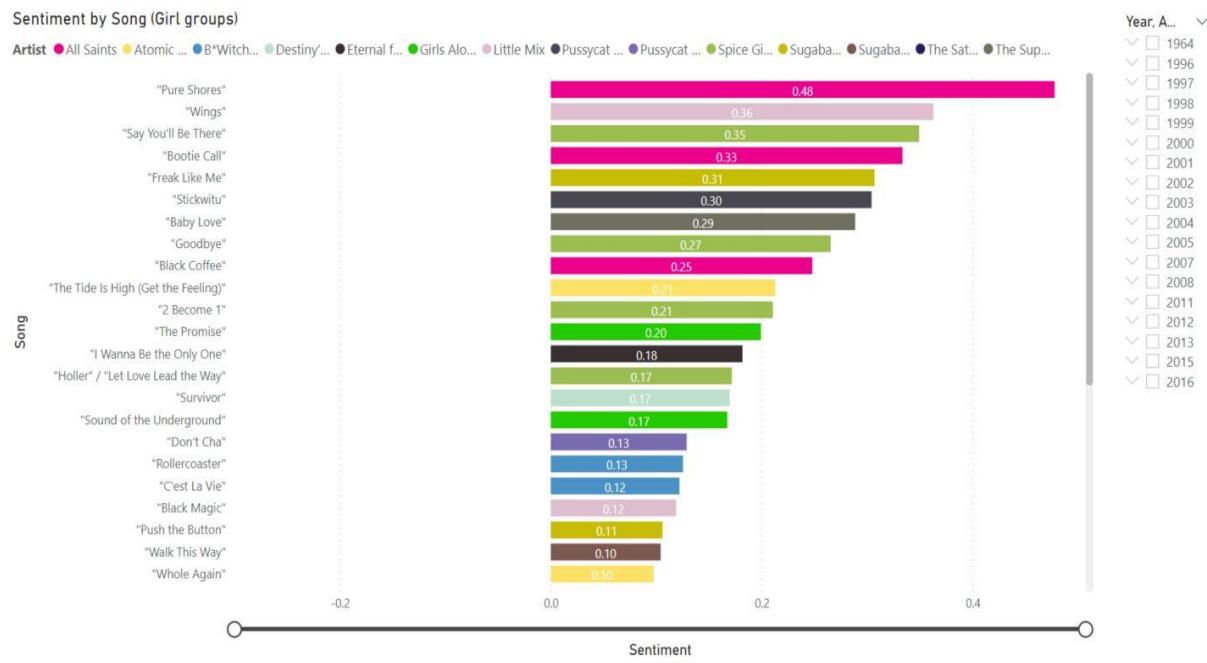


Figure 156 - "Girl group" songs with a positive sentiment value (stacked bar chart).

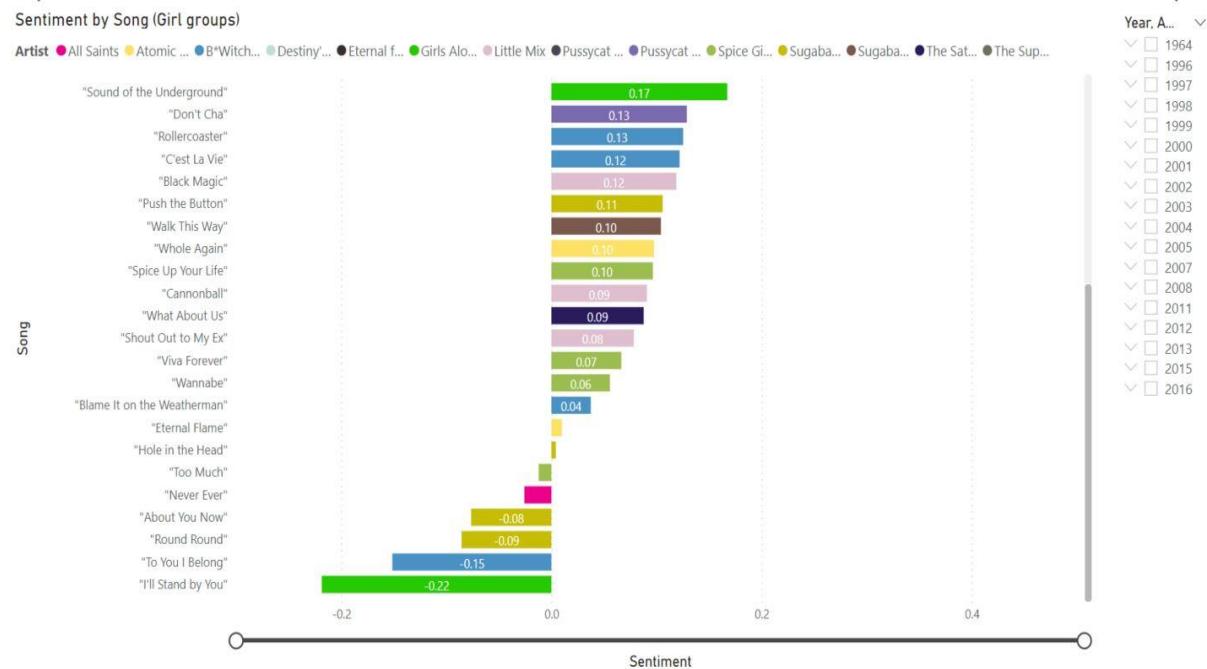


Figure 157 - "Girl group" songs with a negative sentiment value (stacked bar chart).

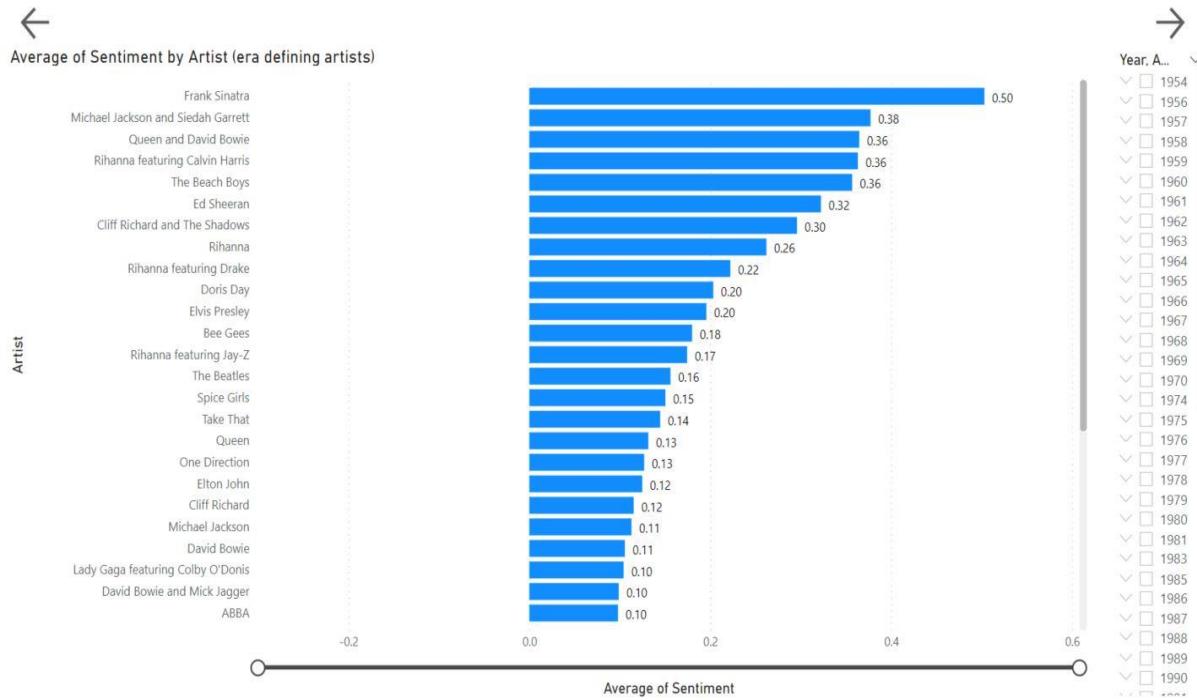


Figure 158 - A closer look at the average sentiment of artists deemed as "Era-defining" (stacked bar chart).

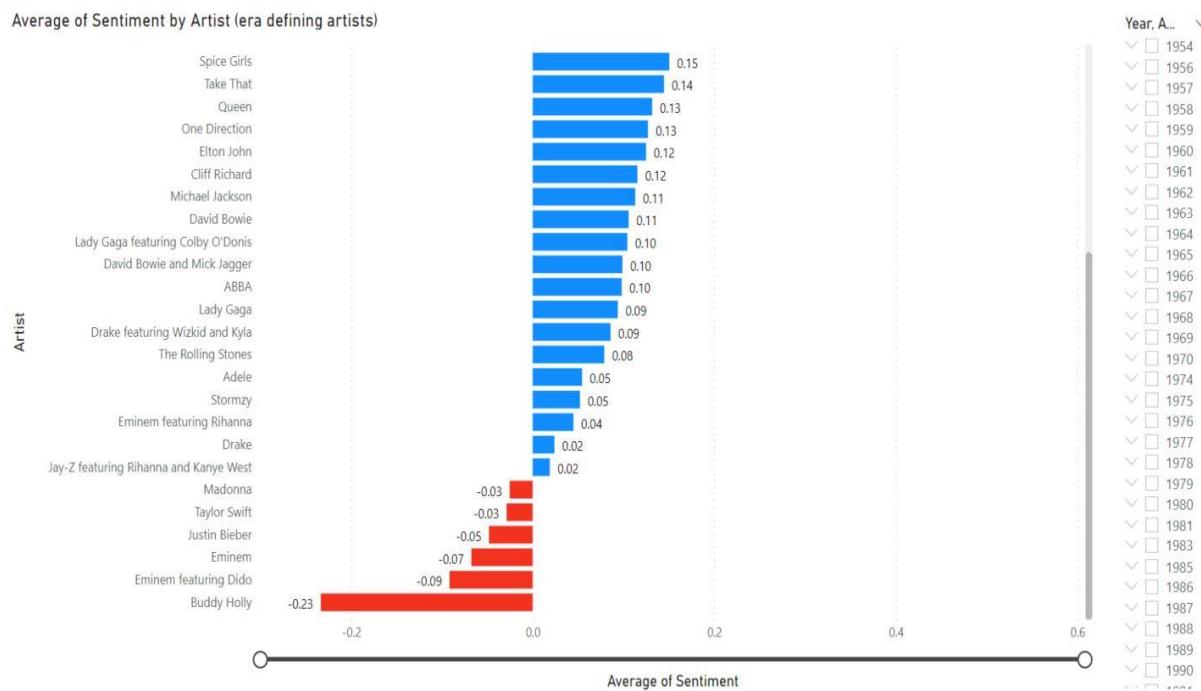


Figure 159 - A closer look at the average sentiment of artists deemed as "Era-defining" (stacked bar chart).

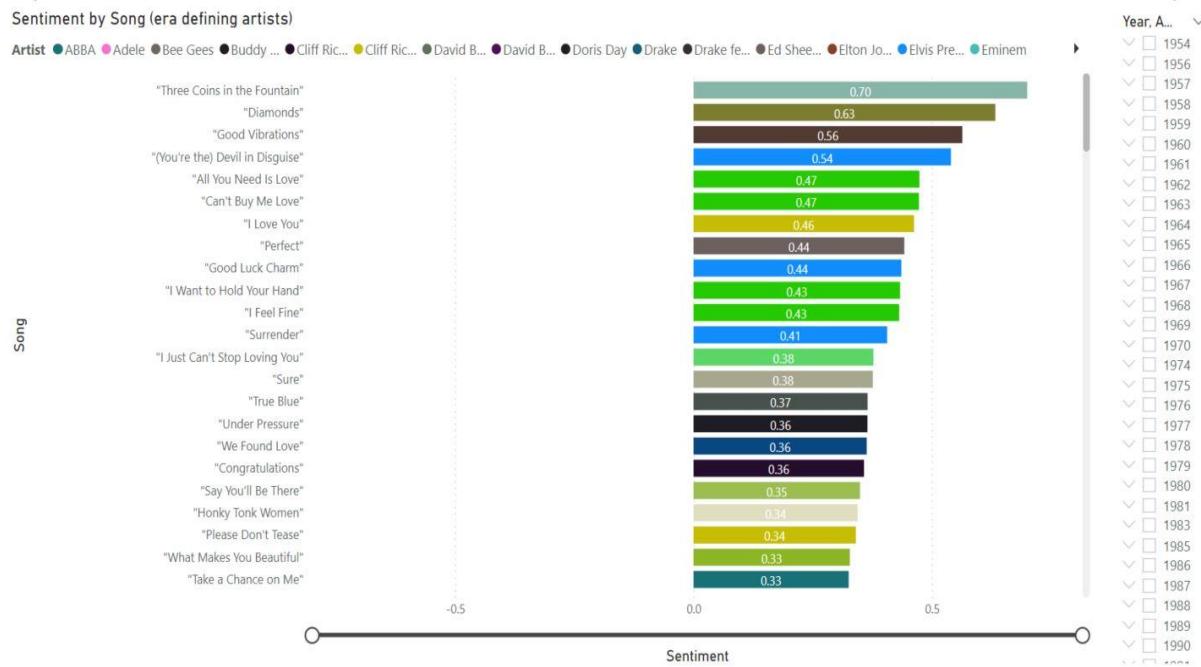


Figure 160 - "Era-defining" artists' songs with a positive sentiment value (stacked bar chart).

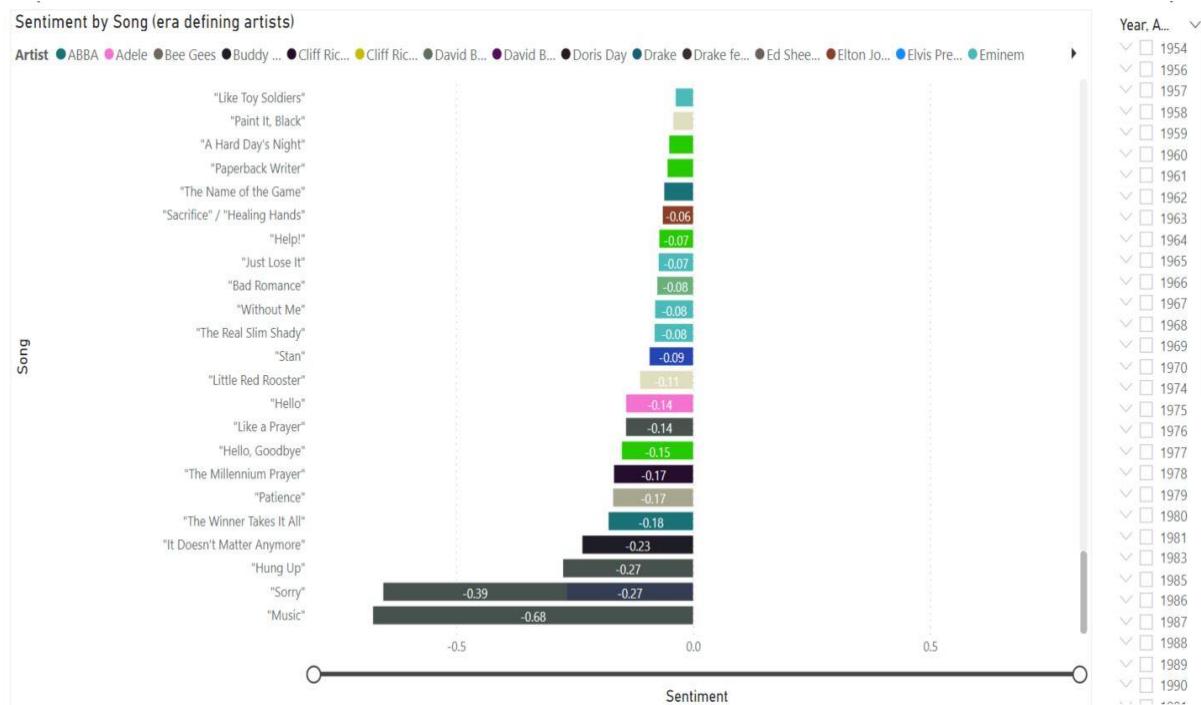


Figure 161 - "Era-defining" artists' songs with a negative sentiment value (stacked bar chart).

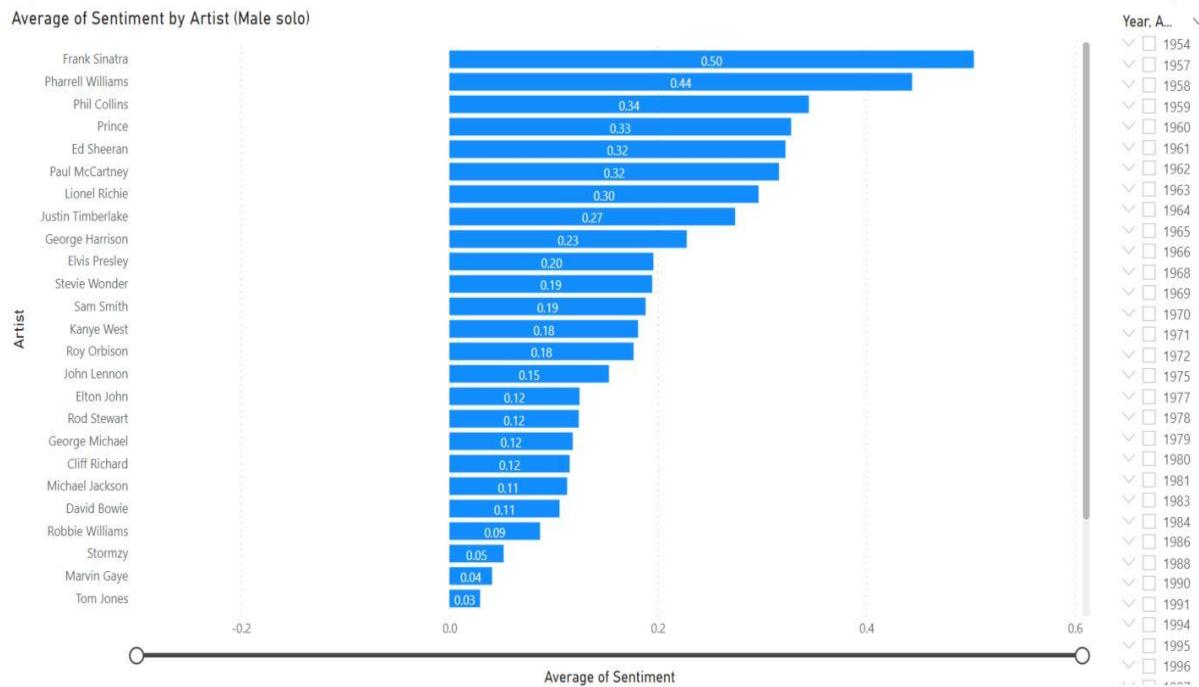


Figure 162 - A closer look at the average sentiment of male solo artists (stacked bar chart).

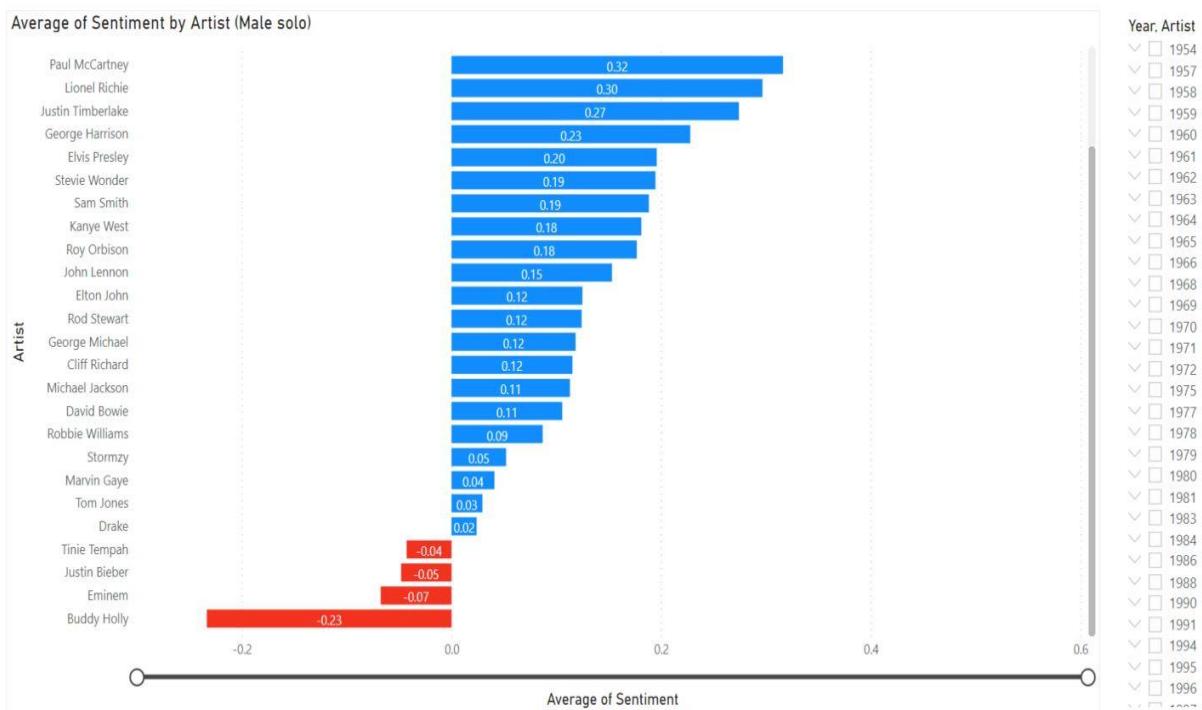


Figure 163 - A closer look at the average sentiment of male solo artists (stacked bar chart).

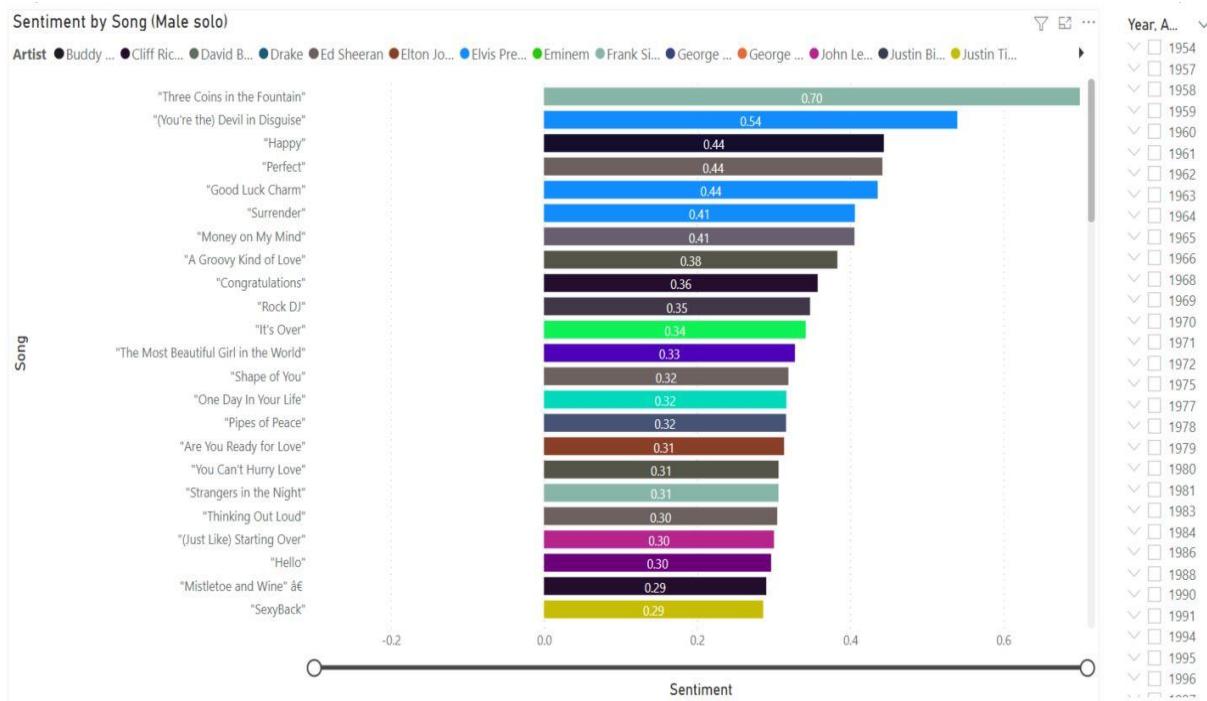


Figure 164 - Male solo artists' songs with a positive sentiment value (stacked bar chart).

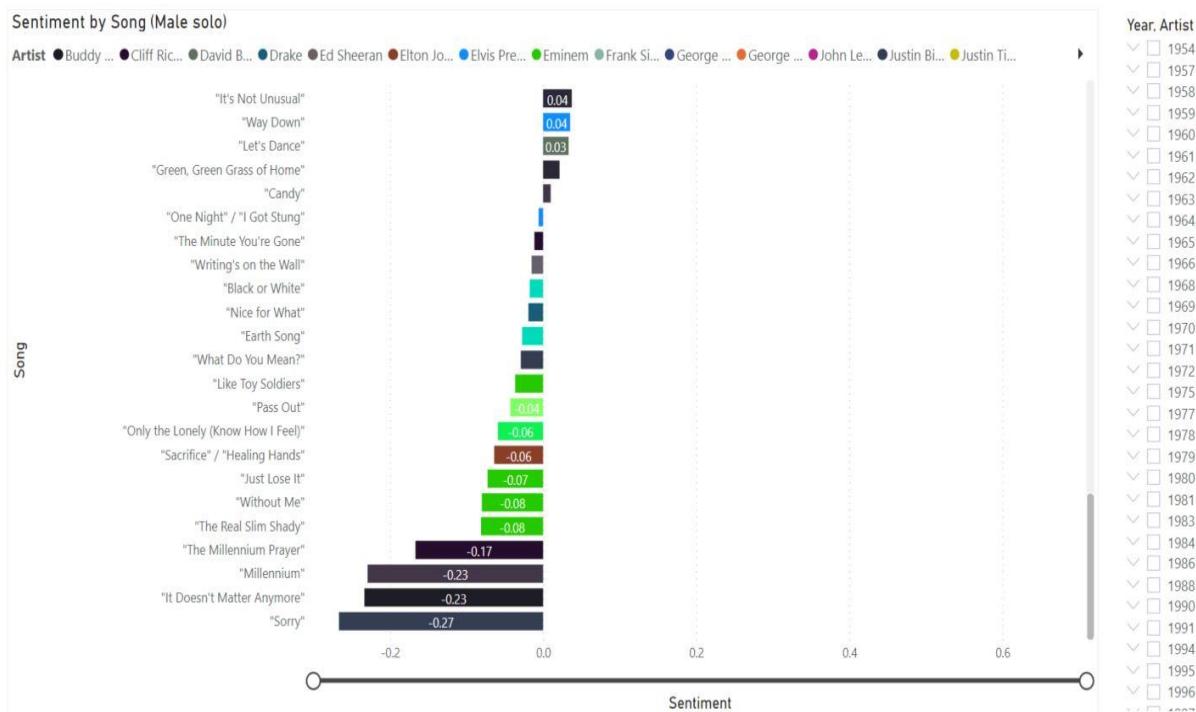


Figure 165 - Male solo artists' songs with a negative sentiment value (stacked bar chart).

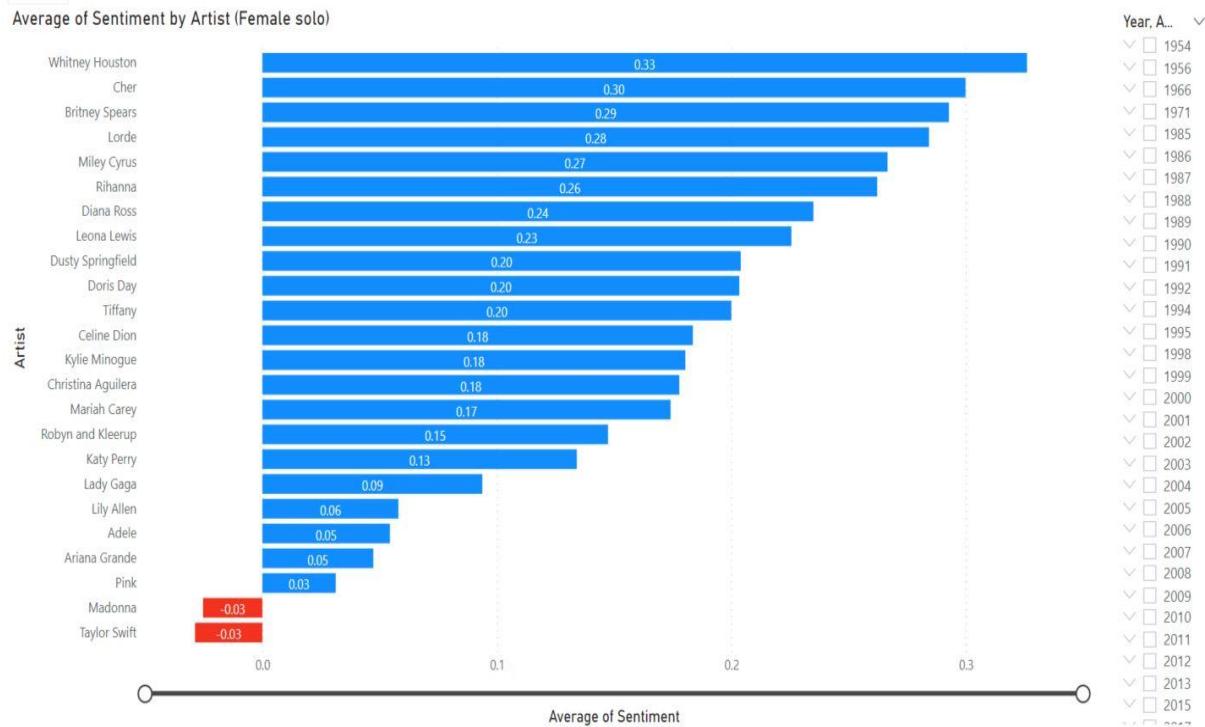


Figure 166 - A closer look at the average sentiment of female solo artists (stacked bar chart).

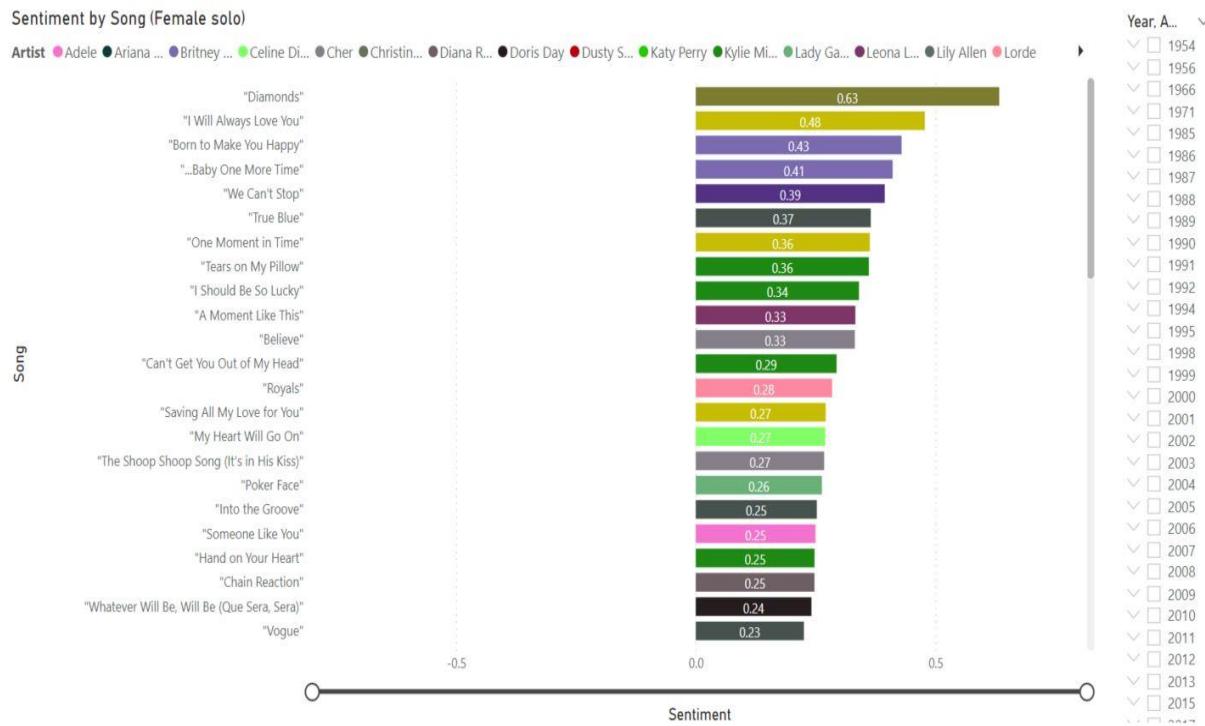


Figure 167 - Female solo artists' songs with a positive sentiment value (stacked bar chart).

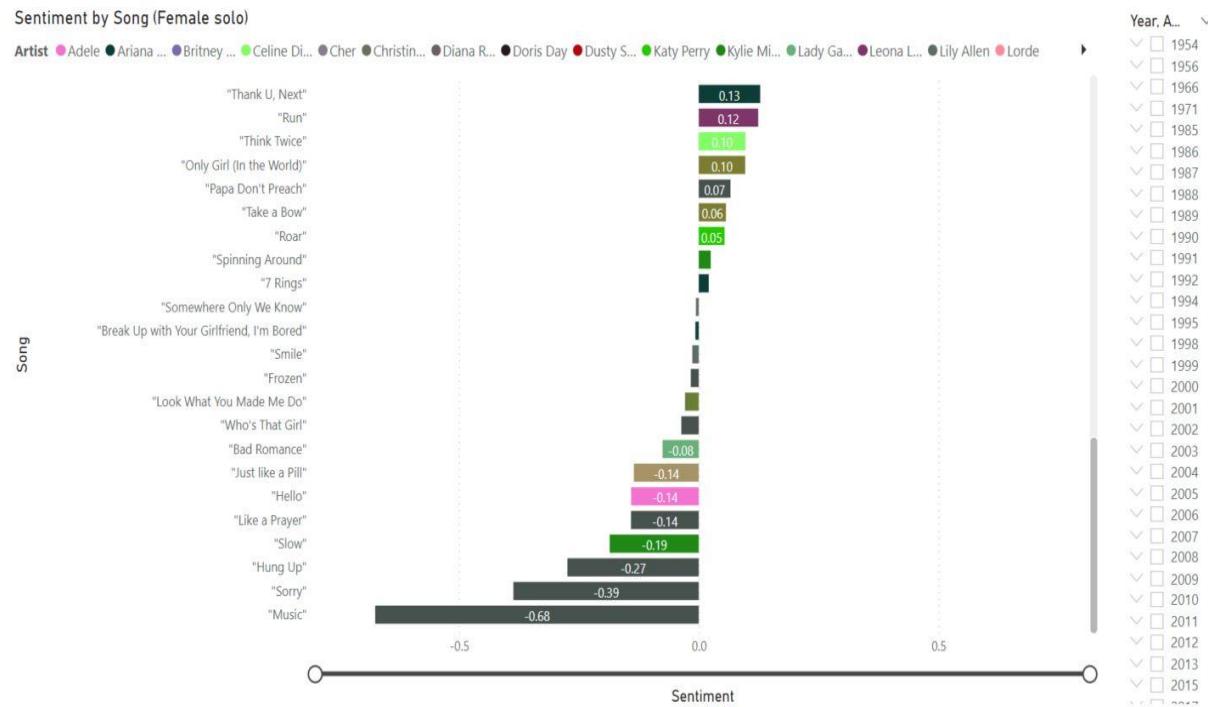


Figure 168 - Female solo artists' songs with a negative sentiment value (stacked bar chart).

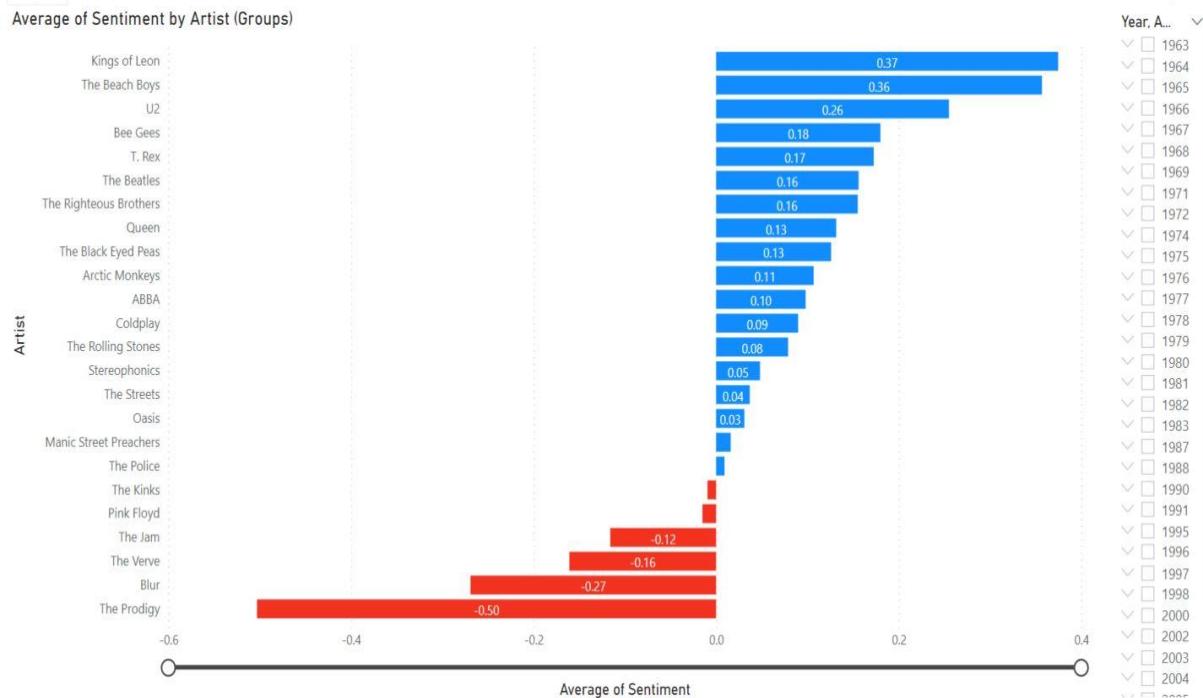


Figure 169 - A closer look at the average sentiment of groups (stacked bar chart).

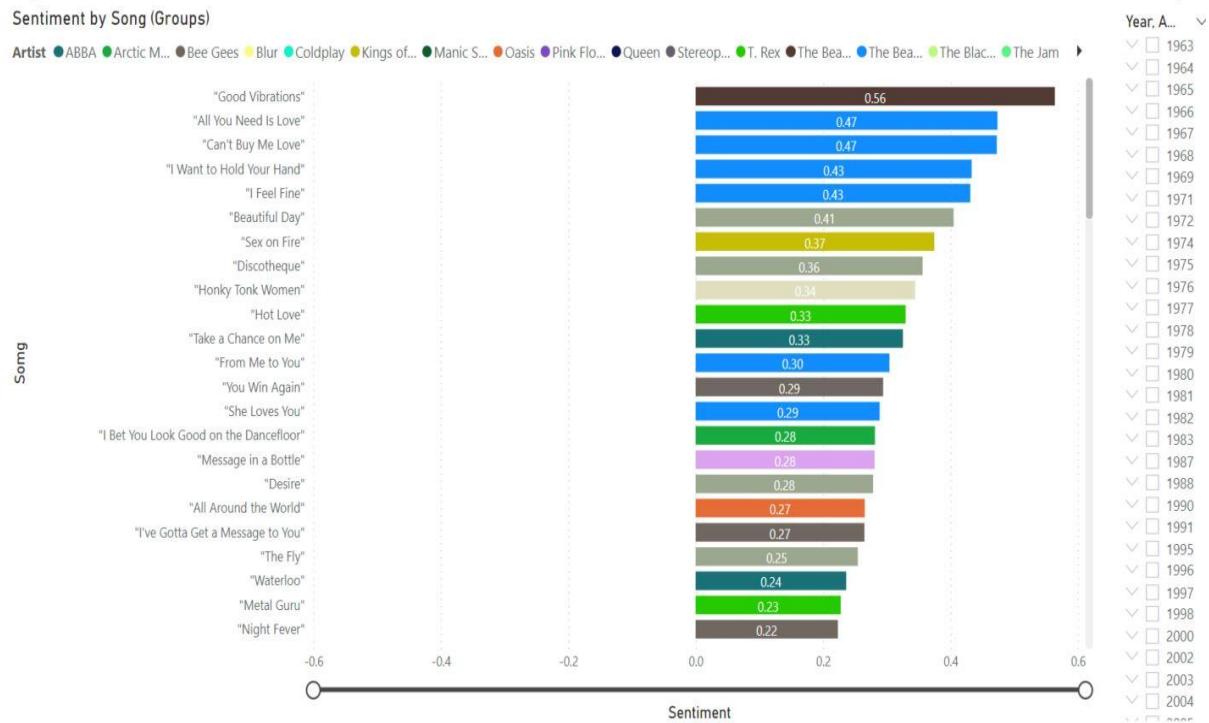


Figure 170 - Groups' songs with a positive sentiment value (stacked bar chart).

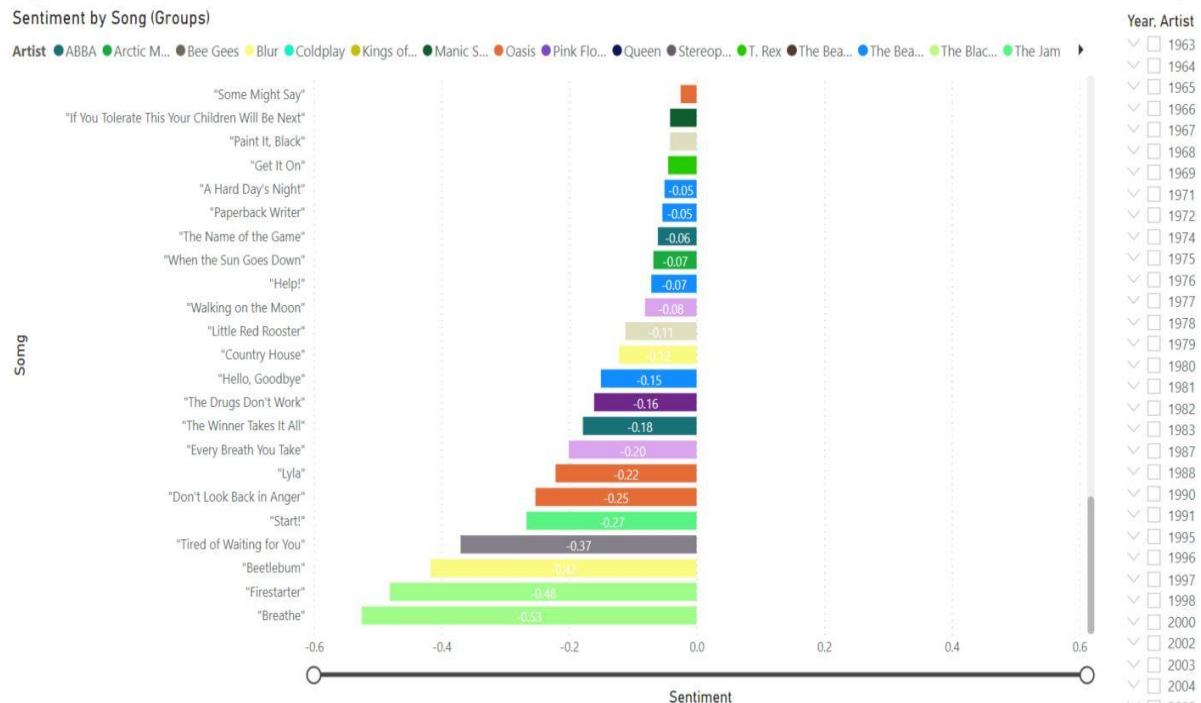


Figure 171 - Groups' songs with a negative sentiment value (stacked bar chart).

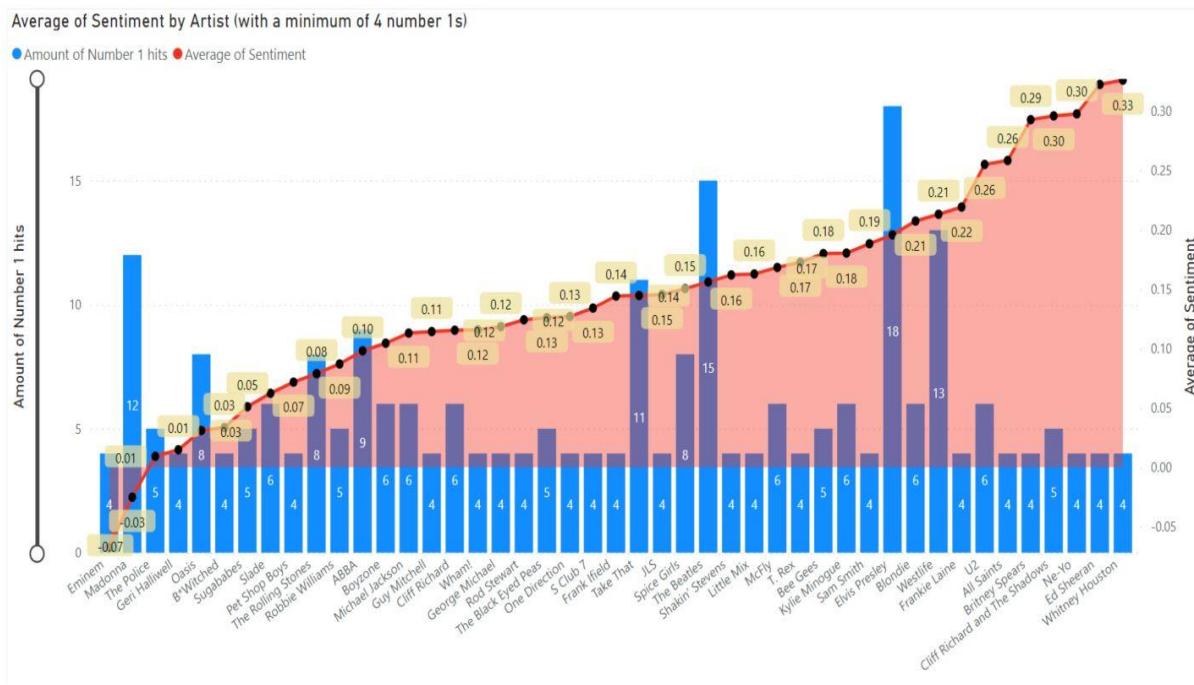


Figure 172 - The average sentiment of artists who claimed at least four number ones (line and clustered column chart).

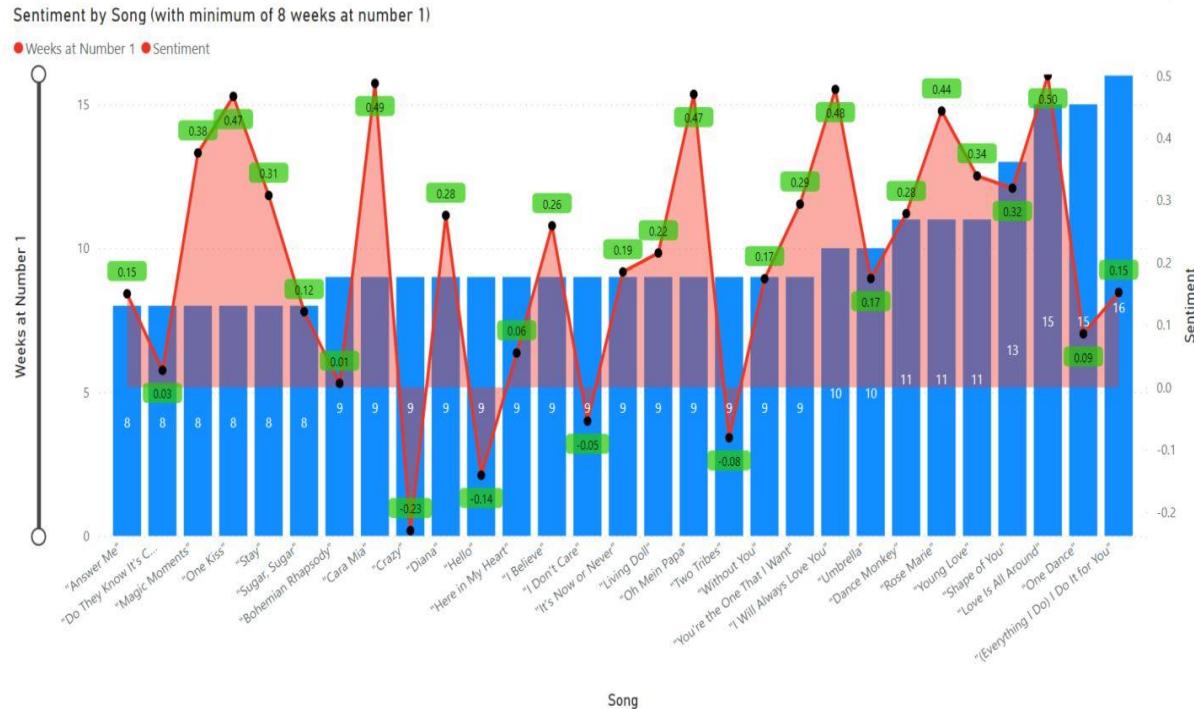


Figure 173 - Sentiment score of songs which topped the UK charts for at least eight weeks (line and stacked column chart).

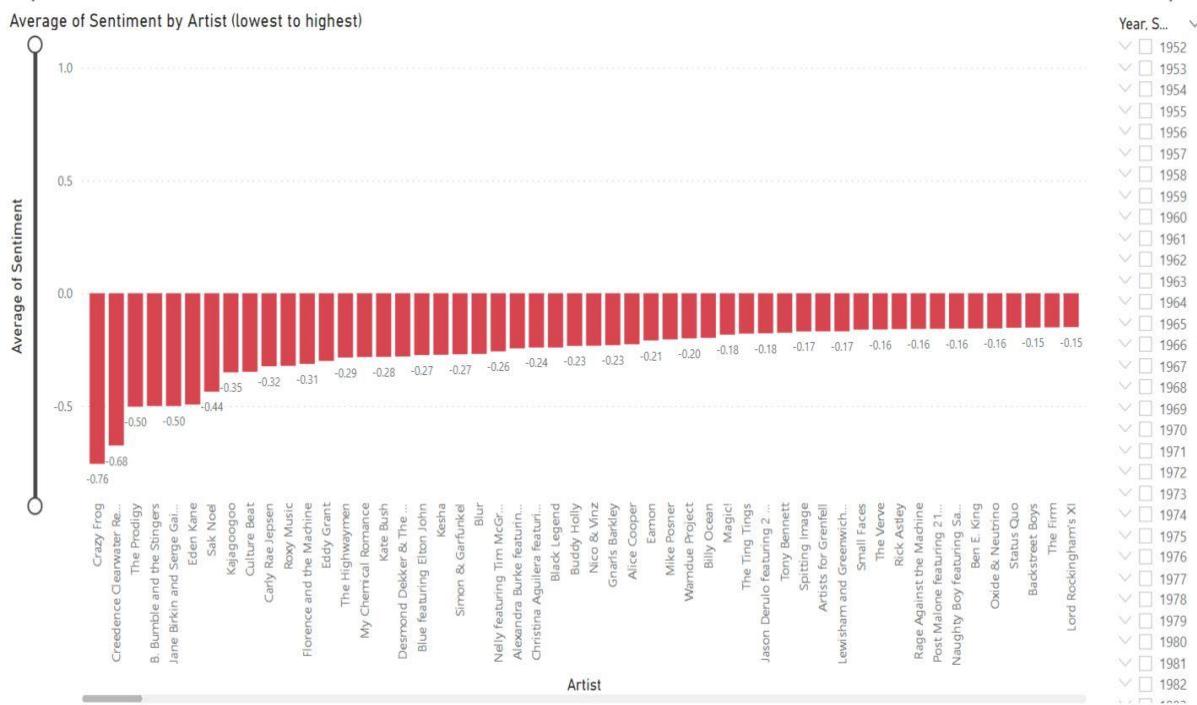


Figure 174 - The sentiment rating of every artist from negative to positive (line and clustered column chart).

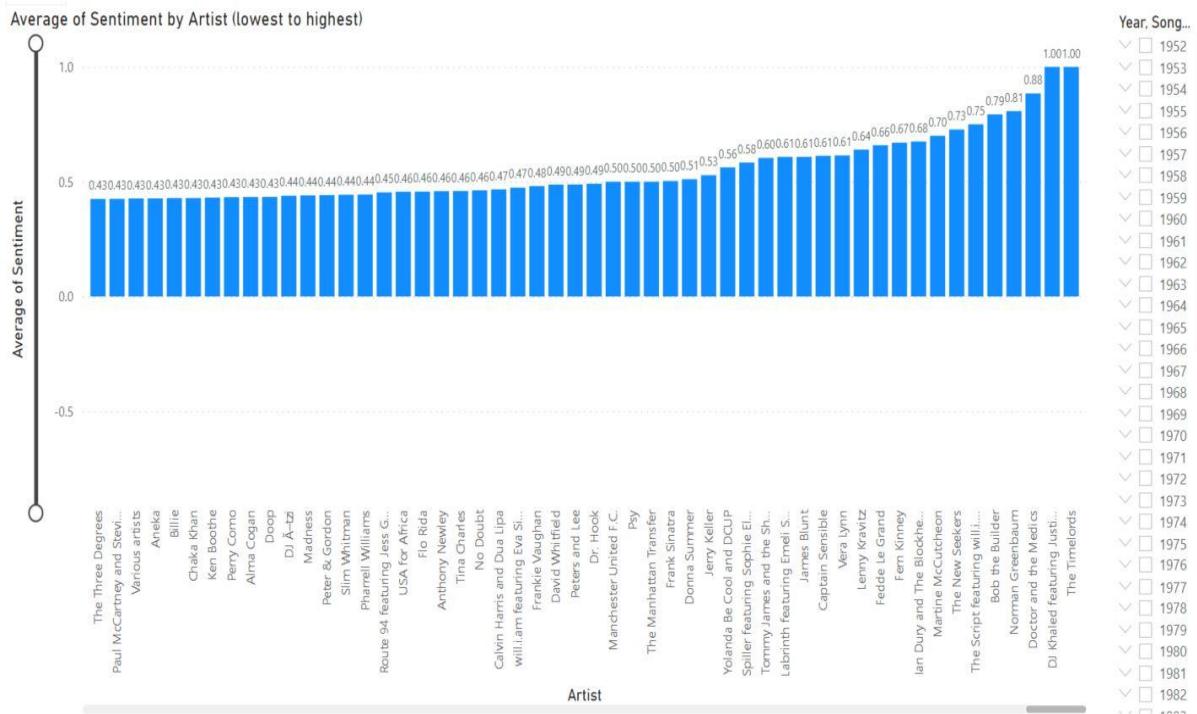


Figure 175 - The sentiment rating of every artist from negative to positive (line and clustered column chart).

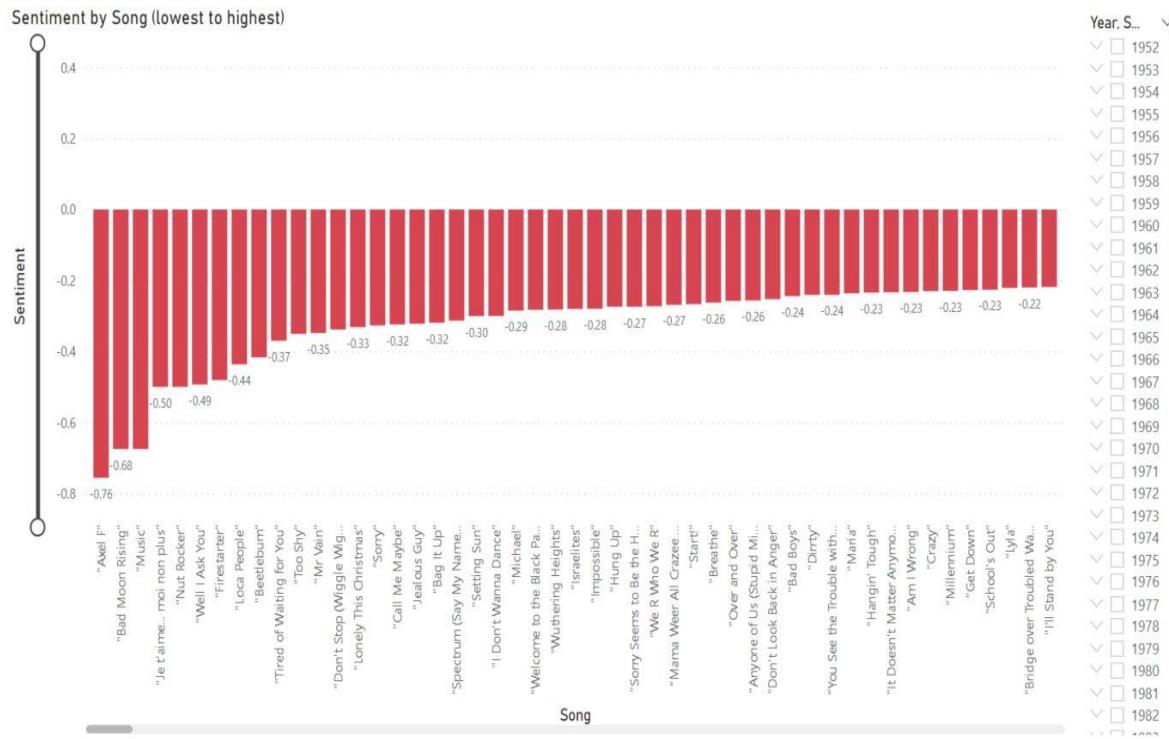


Figure 176 - The sentiment rating of every number one song from negative to positive (line and clustered column chart).

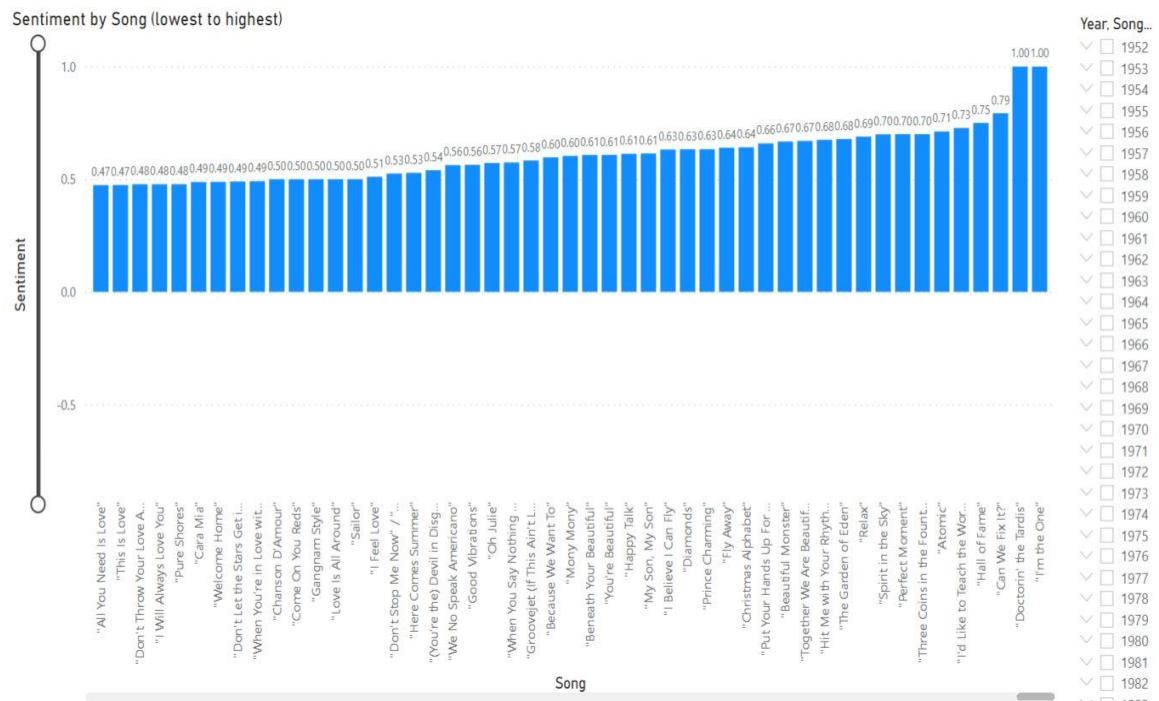


Figure 177 - The sentiment rating of every number one song from negative to positive (line and clustered column chart).

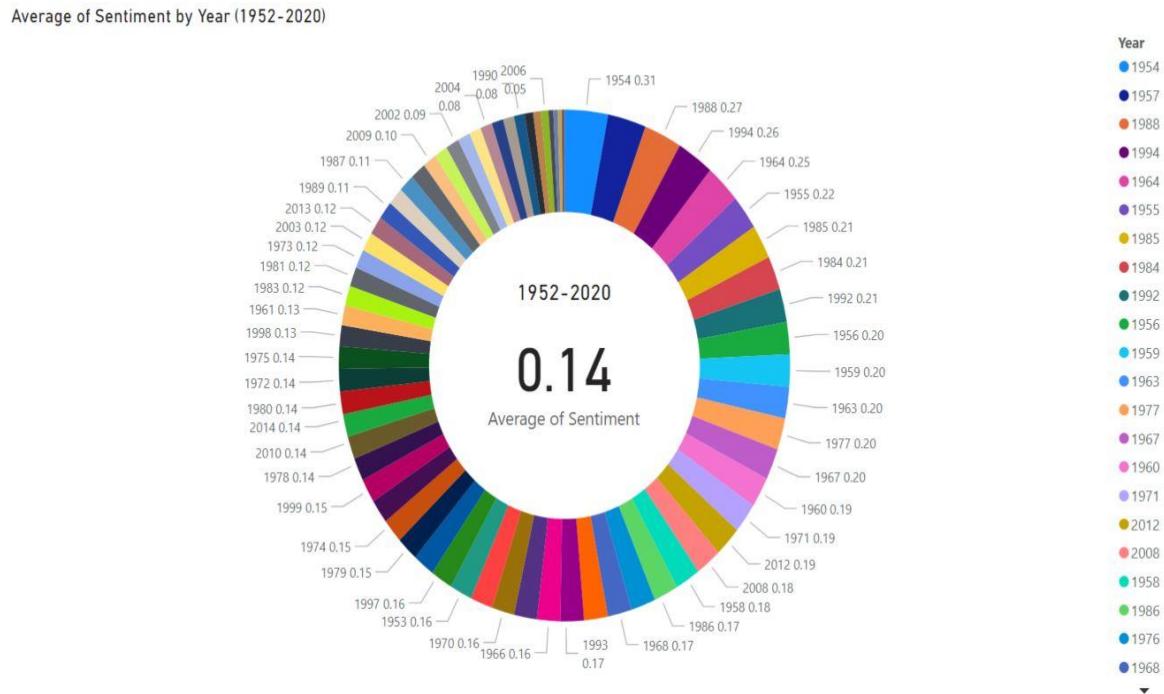


Figure 178 - The average sentiment of every year of the UK chart, 1952-2020 (donut chart).

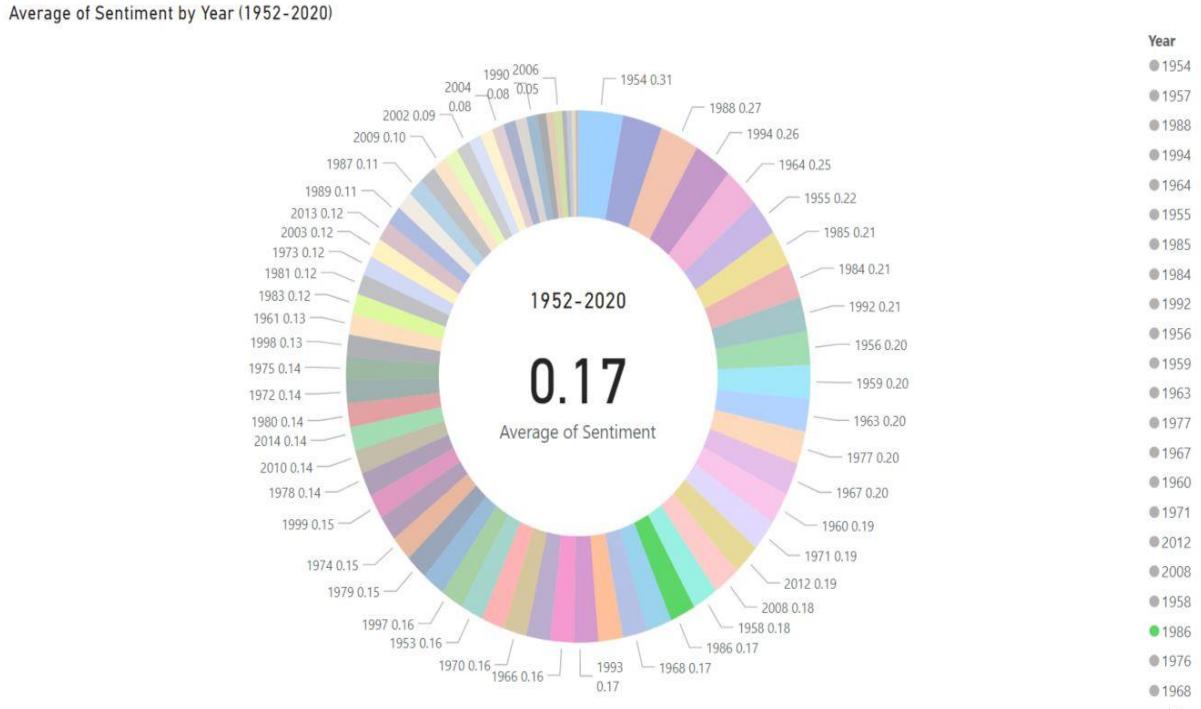


Figure 179 - 1986, randomly selected, as an example of the sentiment being updated interactively (donut chart).

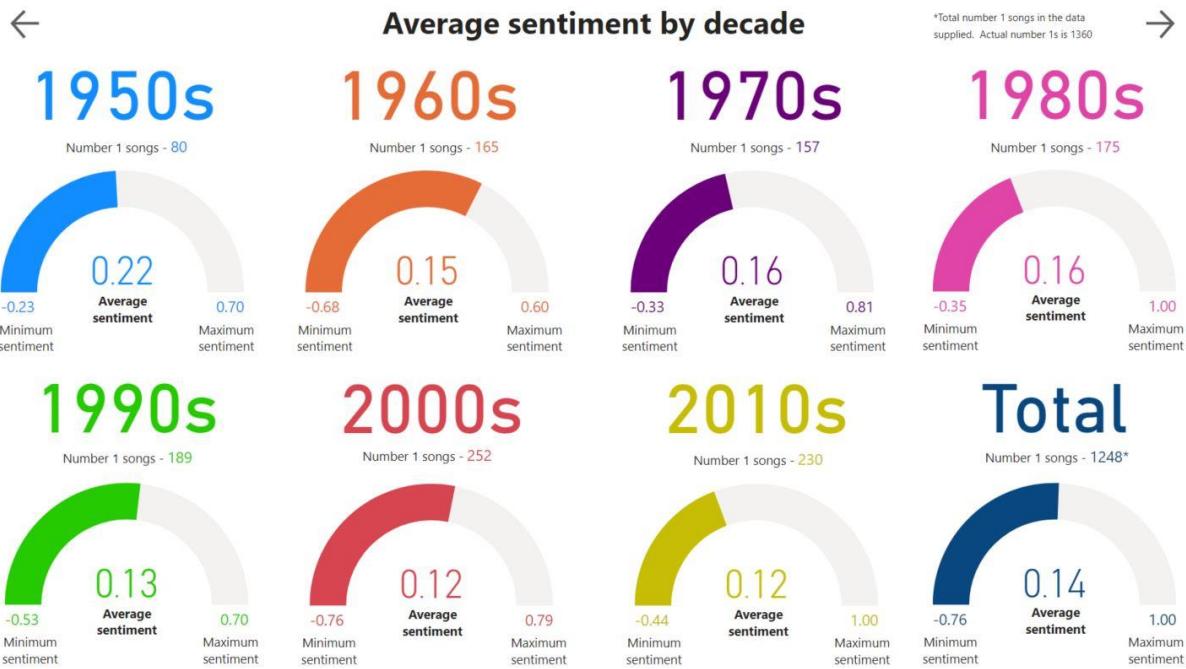


Figure 180 - The average sentiment for each decade of the UK charts (gauge)



Figure 181 - An interactive tool for the reader to search the average sentiment of their favourite artist or number one song (table).

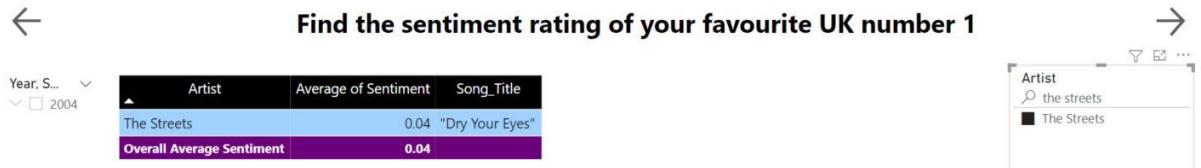


Figure 182 - An example of using the 'search bar' to look at the sentiment of "Dry Your Eyes", by The Streets (table).

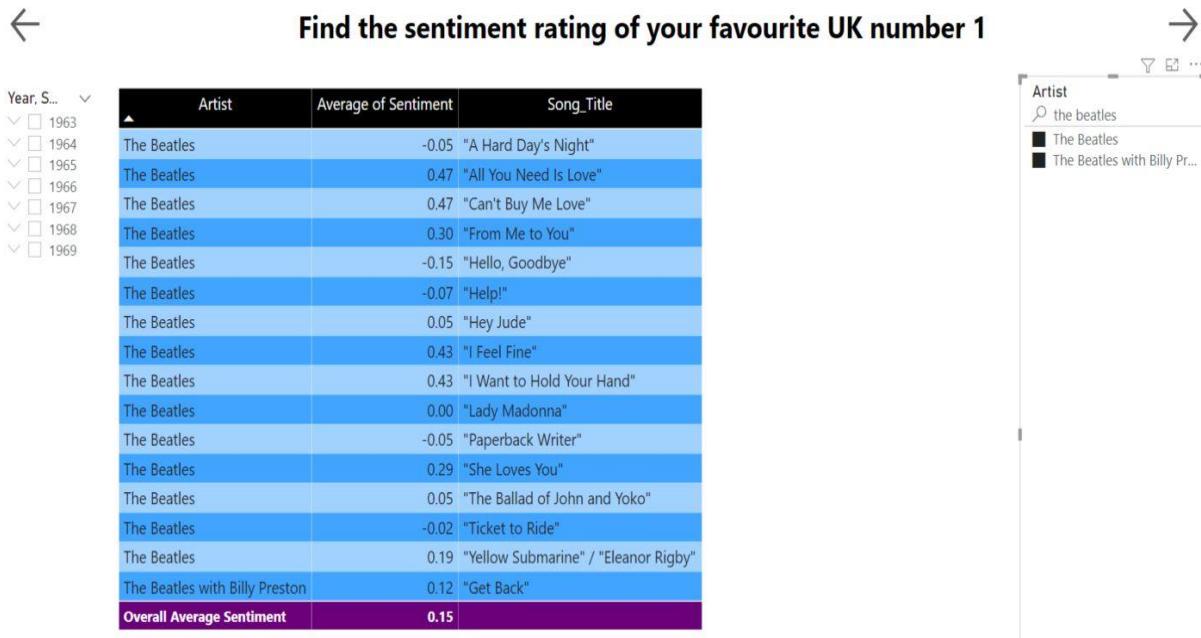
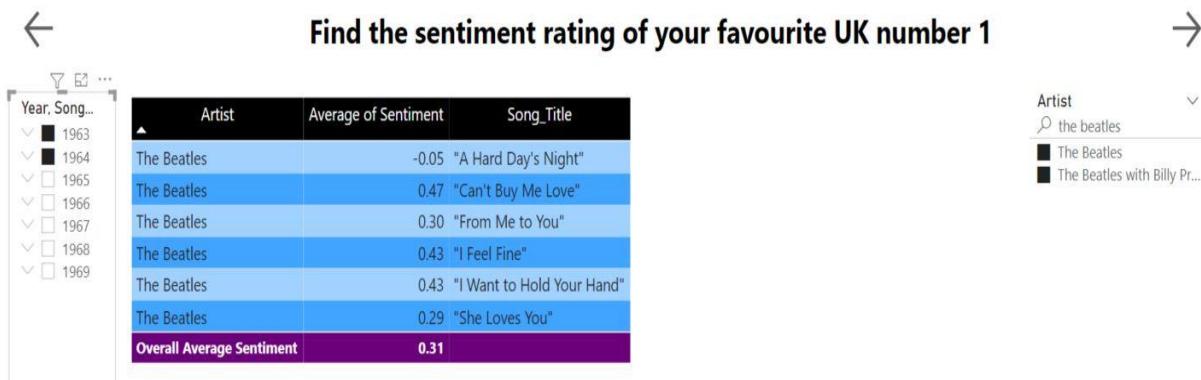


Figure 183 - As demonstrated, more than one artist can be selected from the 'search bar' by holding the 'Ctrl' button on the keyboard (table).



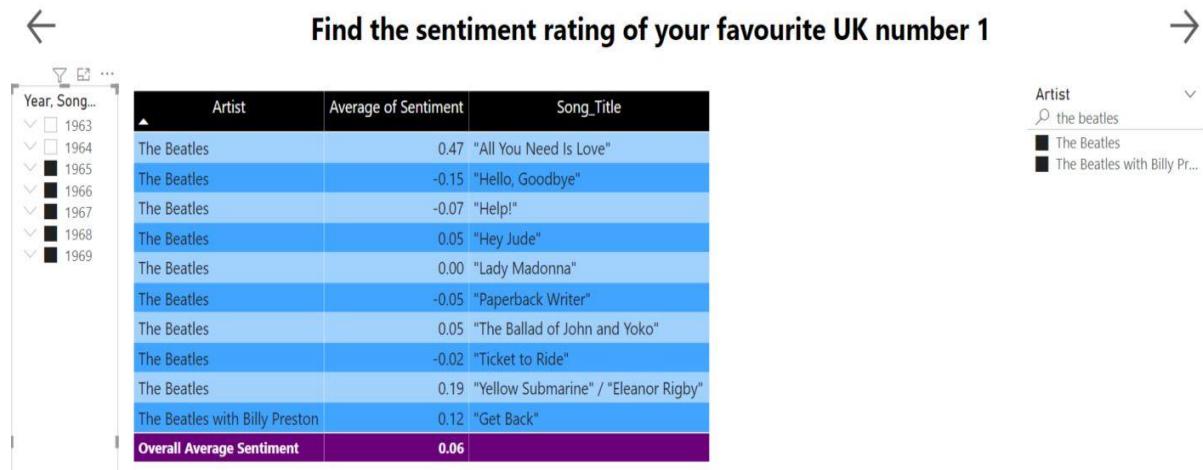


Figure 184 - Proof of how a sentiment score can be notably varied over time. The first two years of Beatles' number ones scored highly (a) which is in stark contrast to their number ones which followed over the next five years (b). A considerable dip in polarity (table).

Figure 1- Example of Positive analysis (Everything There Is to Know about Sentiment Analysis, 2020)	3
Figure 2 - Example of Neutral analysis (Everything There Is to Know about Sentiment Analysis, 2020)	4
Figure 3 - Example of Negative analysis (Everything There Is to Know about Sentiment Analysis, 2020)	4
Figure 4 – Ekman’s words associated with emotion detection (Fernández-Caballero et al., 2016)	10
Figure 5 - Parrot's emotion detection theory (H and el, 2011).....	10
Figure 6 - Sentiment Analysis tree diagram (Pandey, Rajpoot and Saraswat, 2017)	11
Figure 7 - Australian study on teen music listening stereotypes.....	15
Figure 8 - UK study on musical taste by ethnicity.....	16
Figure 9 - UK study on musical taste by social class.....	16
Figure 10 - Installing libraries and packages.	33
Figure 11 - Testing connection between lyricsgenius and Genius API (a), and Taylor Swift's "Shake It Off" proving successful connection between lyricsgenius and Genius API (b). ..	34
Figure 12 - Importing csv file from data.world (a), then displayed in tabular form in notebook (b).	34
Figure 13 - Calling the return_keywords function (a) to display an array of the top keywords for each year (b), (c), (d).	35
Figure 14 - Example of nltk 'Stop Words' in the English language	38
Figure 15 - Obtaining the client access token from genius.com API.	39
Figure 16 - Testing Genius API with Adele's "Hometown Glory".....	39
Figure 17 - Defining 'get_lyrics' function.	40
Figure 18 - code implemented to drop any songs where lyrics cannot be found.....	40
Figure 19 - Defining 'get_lyric_sentiment' function.	41
Figure 20 – Defining ‘preprocess’ function.	41
Figure 21 - Defining 'return_keywords' function.....	42
Figure 22 - Cain Osborne's code snippet referring to Beautiful Soup.	44
Figure 23 - Osborne's code for scraping UK albums chart (a), (c) & my modifications for scraping UK singles chart (b), (d).....	45
Figure 24 - Power BI's simple method to retrieve data (a) from a website (b).	48
Figure 25 - Wikipedia data brought into Power BI before cleaning.	49

Figure 26 - Setting the index of the dataframe to be sorted by 'WeekID' column. Although this worked for Zubair, it had caused errors on the UK code test run.	50
Figure 27 - UK dataset created with added columns. First five number 1s (a) and last five number 1s (b)	50
Figure 28 - Altering genism install code.....	50
Figure 29 - Importing csv file from local drive.	51
Figure 30 - 'Lyrics' and 'Sentiment' columns added to UK dataframe.....	51
Figure 31 - Plot graph of UK chart (1952-1955).	51
Figure 32 - Amended dataset saved as a csv file.	52
Figure 33 - error-laden csv file.	52
Figure 34 - Erroneous lyrics assigned to Hale & Pace (a), Guy Mitchell (b), and Band Aid 2020 (c)	53
Figure 35 - Chart (a) showing 0.03 as the average sentiment of 1969, while the median of 0.09 is displayed in (b).....	54
Figure 36 - 1969 dataframe in Power BI.....	55
Figure 37 - Arranging the values in order (a) before calculating the median value (b).....	55
Figure 38 - Example of how to calculate average.....	56
Figure 39 - Variance of sentiment in 1969.	56
Figure 40 - 1969 outliers (a). How they compare to the full dataset (b).....	57
Figure 41 - 1996, the lowest ranked year in the dataset with an average sentiment of just, 0.02.....	58
Figure 42 - 1996 data showing songs with a significantly low polarity (a), and those songs at the lowest end of the sentiment scale overall (b).	59
Figure 43 - The four lowest scoring years over the history of the charts, of which two have occurred very recently.....	59
Figure 44 - Variances of the four lowest rated years; 1969, 1996, 2015, and 2016, are bolded.....	60
Figure 45 - Song lyrics from the four lowest sentiment rating years (word cloud).....	60
Figure 46 - "Shit and "Fuck" were amongst many obscene words littered throughout 2018 number ones, "Freaky Friday", by Lil' Dicky and Chris Brown.(a) and Drake's, "Nice For What" (b).....	61
Figure 47 - Clean Bandit ft Demi Lovato's, "Solo" received the lowest polarity of 2018; - 0.12.....	62

Figure 48 - Songs and their lyrics from 1954 which at least matched the 0.31 sentiment score.....	62
Figure 49 - 1954 outliers.....	63
Figure 50 - 1957 outlier.....	64
Figure 51 - 1988 outlier.....	64
Figure 52 - The 50 songs with the highest average sentiment.....	65
Figure 53 - 1994's lack of outliers.....	65
Figure 54 - Sentiment of songs which topped the UK charts for at least eight weeks. The 1950s perform particularly well.....	67
Figure 55 - The average sentiment of every year of the UK chart, 1952-2020	67
Figure 56 - The average sentiment for each decade of the UK charts.....	68
Figure 57 - Line graphs detailing the annual UK GDP of the 1950s (a), 1960s (b), 1970s (c), 1980s (d), 1990s (e), 2000s (f), and 2010s (g)	70
Figure 58 - Chart displaying significant growth relating to Income Tax and National Insurance.....	71
Figure 59 - Public spending figures for healthcare (a), transport (b), education (c), and welfare (d).....	72
Figure 60 - Unemployment rates in general terms (a), and youth unemployment (b).....	72
Figure 61 - Crime figures across the 20th century.....	73
Figure 62 - Homicide rates, 1950s-2017.....	73
Figure 63 - Alcohol consumption rates 1900-2010.	74
Figure 64 - Marriage and divorce rates 1930-2010.....	74
Figure 65 - Closer look at the period of 1950-2014 for divorce rates.	74
Figure 66 - "Don't Stop" by The Outhere Brothers, from 1995, was the first number one to contain both of the words, "Ass" and "Pussy".....	78
Figure 67 - 1996 included the first case of the word, "nigga", appearing in a number one of, surprisingly, high polarity.....	79
Figure 68 - Sentiment of songs by artist, year, and nationality. 1996 hit, "Ready or Not" receiving a sentiment rating of 0.14.....	80
Figure 69 - "Nasty Girl", by Notorious B.I.G., became the first number one to use, "nigga", which had a negative polarity score.....	81
Figure 70 - Eminem's year 2000 included two number ones containing the words, "Shit" and "Fuck"	81

Figure 71 - Gotye is the highest scoring artist with a song containing, "Love" and "Happy" in the past decade	82
Figure 72 - 2012 is also the only year of the past decade to reach an average sentiment of at least 0.18.	82
Figure 73 - The highest sentiment scoring songs containing the lyrics, 'Love' and 'Happy' both came from The Beatles in 1963, and 1964, respectively.	83
Figure 74 - Proof of how a sentiment score can be notably varied over time. The first two years of Beatles' number ones scored highly (a) which is in stark contrast to their number ones which followed over the next five years (b). A considerable dip in polarity.	84
Figure 75 - Average sentiment by nationality of artists.....	84
Figure 76 - The average sentiment for each decade of the UK charts.	85
Figure 77 - Installing libraries and packages	124
Figure 78 - Libraires and packages installing (1)	124
Figure 79 - Libraires and packages installing (2)	125
Figure 80 - Testing connection between lyricsgenius and Genius API.....	125
Figure 81 - Taylor Swift's "Shake It Off" proving successful connection between lyricsgenius and Genius API (a), (b), (c)	126
Figure 82 - Importing csv file from data.world	127
Figure 83 - Imported csv table displayed in notebook.....	127
Figure 84 - Functions being defined for use further on	128
Figure 85 - get_lyrics function connecting with Genius API.	128
Figure 86 - Lyrics being imported from Genius API (a), and being added as a new 'Lyrics' column (b), (c), (d)	129
Figure 87 - get_lyric_sentiment function being called to classify polarity on lyrics and create 'Sentiment' column.	130
Figure 88 - Setting the index of the dataframe to be sorted by 'WeekID' column.....	130
Figure 89 - Resampling the dataframe to calculate the mean sentiment of the lyrics for each year.....	130
Figure 90 - Plotting a line graph to display the sentiment of each year.....	130
Figure 91 - Resampling the lyric sentiment (a) to calculate the mean percentage change (b)	130
Figure 92 - Resampling the 'Lyrics' column to collect the lyrics from each year.....	130

Figure 93 - Calling the return_keywords function (a) to display an array of the top keywords for each year (b), (c), (d)	131
Figure 94 - Creating a dataframe from the keywords yielded.	131
Figure 95 - Keywords dataframe displayed in tabular form including 'WeekID', 'Lyrics', and 'Sentiment' columns (a), (b), (c), (d), (e), (f), (g)	133
Figure 96 – Porter stemming (a), and spaCy lemmatisation (b) of ‘change’ and ‘changing’.	134
Figure 97 – Porter stemming (a), and spaCy lemmatisation (b) of ‘beautiful’.....	134
Figure 98 – Lancaster stemming (a), and spaCy lemmatisation (b) of ‘skies’.	135
Figure 99 – spaCy lemmatisation of ‘best’ to ‘good’.....	135
Figure 100 – TextBlob polarity score of ‘best’ (a) and ‘good’ (b).	136
Figure 101 - Power BI's simple method to retrieve data (a) from a website (b).....	137
Figure 102 - Wikipedia's 1950s data in tabular form.....	137
Figure 103 - Wikipedia data brought into Power BI before cleaning.	138
Figure 104 - Giving meaningful heading names.....	138
Figure 105 - Omitting the years (a) to ensure correct index numbers (b).....	138
Figure 106 - Using Google to find the correct keyboard code for the 'Cross sign' symbol (Alt shortcuts list, 2021).....	139
Figure 107 - Using 'Text before delimiter' function to remove 'double-cross' (a) & 'cross' (b) symbols.	139
Figure 108 - Two songs before (a) and after (b) removal of symbols.	140
Figure 109 - Replacing floating-point numbers with an integer (a), and an erroneous value (b).	140
Figure 110 - All applied modifications to the original data table (a). Labelling the dataset with a meaningful name (b).	141
Figure 111 - Installing libraries and packages.	142
Figure 112 - Testing connection between lyricsgenius and Genius API	142
Figure 113 - Importing csv file from local machine.	143
Figure 114 - Functions being defined for use further on (note that 'preprocess' and 'keywords' functions commented out from prototype as this will be done in Power BI on this occasion).....	144
Figure 115 - get_lyrics function connecting with Genius API.	144

Figure 116 - Lyrics being imported from Genius API and being added as a new 'Lyrics' column.....	145
Figure 117 get_lyric_sentiment function being called to classify polarity on lyrics and create 'Sentiment' column. (Using 'print' to check all fields are entered).	145
Figure 118 - Double checking all fields are implemented in table.	145
Figure 119 - Checking data types of columns are correct.....	146
Figure 120 - Setting the index of the dataframe to be sorted by 'WeekID' column (a), resampling the dataframe to calculate the mean sentiment of the lyrics for each year (b), plotting a line graph to display the sentiment of each year (c), and resampling the lyric sentiment to calculate the mean percentage change (d).....	146
Figure 121 - An essential section of coding for the prototype is commented out for the UK data run as Power BI will perform this function.	147
Figure 122 - Saving amended dataset back to csv on local drive.	147
Figure 123 - Average sentiment by year (stacked column chart).	148
Figure 124 - Standard deviation of sentiment by year (stacked column chart).	148
Figure 125 - Variance of sentiment by year (stacked column chart).	149
Figure 126 - Median of sentiment by year (area chart).....	149
Figure 127 - Minimum sentiment by year (area chart).	149
Figure 128 - Maximum sentiment by year (area chart).	150
Figure 129 - Average sentiment of artists by year and nationality. The Beatles' 1963 seen here as an example (Map) and (stacked bar chart).....	150
Figure 130 - Using the filter to show all artists with a sentiment of 0.50 by year and nationality. South Korean, Psy, is the example shown here through his 'one-hit-wonder' global success, "Gangnam Style" (Map) and (stacked bar chart).	151
Figure 131 - Sentiment of songs by artist, year, and nationality. Elvis Presley had the highest polarity score of 1963 with, "(You're the) Devil in Disguise" (Map) and (stacked bar chart).	151
Figure 132 - Sentiment of songs by artist, year, and nationality. 1996 hit, "Ready or Not" receiving a sentiment rating of 0.14 (Map) and (stacked bar chart).	152
Figure 133 - Average sentiment of each nation to have a UK number one (stacked column chart)	152
Figure 134 - 1954 was the year which reached the highest average sentiment (word cloud).	153

Figure 135 - Songs and their lyrics from 1954 which at least matched the 0.31 sentiment score (word cloud).	153
Figure 136 - The highest sentiment scoring songs containing the lyrics, 'Love' and 'Happy' both came from The Beatles in 1963, and 1964, respectively (word cloud).	153
Figure 137 - Almost 50 years later, and Gotye is the highest scoring artist with a song containing, "Love" and "Happy" in the past decade. His polarity score is not even half of The Beatles previously mentioned efforts (word cloud).....	154
Figure 138 - 2012 is also the only year of the past decade to reach an average sentiment of at least 0.18 (word cloud).....	154
Figure 139 - Nine number one hits with a, relatively, high sentiment score (word cloud)	155
Figure 140 - 1996 was the lowest rating year of sentiment (word cloud).	155
Figure 141 - Song lyrics matching or less than the 0.02 sentiment score of 1996 (word cloud).	155
Figure 142 - 1996 included the first case of the word, "nigga", appearing in a number one of, surprisingly, high polarity (word cloud).....	156
Figure 143 - "Nasty Girl", by Notorious B.I.G., became the first number one to use, "nigga", which had a negative polarity score (word cloud).....	156
Figure 144 - The four lowest scoring years over the history of the charts, of which two have occurred very recently (word cloud).....	157
Figure 145 - Song lyrics from the four lowest sentiment rating years (word cloud).....	157
Figure 146 - "Don't Stop" by The Outhere Brothers, from 1994, was the first number one to contain both of the words, "Ass" and "Pussy" (word cloud).....	158
Figure 147 - Eminem's year 2000 included two number ones containing the words, "Shit" and "Fuck" (word cloud).....	158
Figure 148 - "Shit and "Fuck" were amongst many obscene words littered throughout "Freaky Friday", a 2018 number one for Lil' Dicky and Chris Brown (word cloud).	159
Figure 149 - Drake's, "Nice For What" was another 2018 hit using bad language (word cloud).	159
Figure 150 - A snapshot of the evolution of song lyrics over the history of the UK charts (word cloud).....	160
Figure 151 - Five number ones with a surprisingly negative rating (word cloud).....	160
Figure 152 - A closer look at the average sentiment of artists deemed as "Heartthrobs" by many (stacked bar chart).	161

Figure 153 - "Heartthrob" songs with a positive sentiment value (stacked bar chart).....	161
Figure 154 - "Heartthrob" songs with a negative sentiment value (stacked bar chart).....	162
Figure 155 - A closer look at the average sentiment of artists deemed as "Girl groups" (stacked bar chart).....	162
Figure 156 - "Girl group" songs with a positive sentiment value (stacked bar chart).	163
Figure 157 - "Girl group" songs with a negative sentiment value (stacked bar chart).	163
Figure 158 - A closer look at the average sentiment of artists deemed as "Era-defining" (stacked bar chart).....	164
Figure 159 - A closer look at the average sentiment of artists deemed as "Era-defining" (stacked bar chart).....	164
Figure 160 - "Era-defining" artists' songs with a positive sentiment value (stacked bar chart).	165
Figure 161 - "Era-defining" artists' songs with a negative sentiment value (stacked bar chart).	165
Figure 162 - A closer look at the average sentiment of male solo artists (stacked bar chart).	166
Figure 163 - A closer look at the average sentiment of male solo artists (stacked bar chart).	166
Figure 164 - Male solo artists' songs with a positive sentiment value (stacked bar chart). 167	167
Figure 165 - Male solo artists' songs with a negative sentiment value (stacked bar chart).167	167
Figure 166 - A closer look at the average sentiment of female solo artists (stacked bar chart).	168
Figure 167 - Female solo artists' songs with a positive sentiment value (stacked bar chart).	168
Figure 168 - Female solo artists' songs with a negative sentiment value (stacked bar chart).	169
Figure 169 - A closer look at the average sentiment of groups (stacked bar chart).....169	169
Figure 170 - Groups' songs with a positive sentiment value (stacked bar chart).....170	170
Figure 171 - Groups' songs with a negative sentiment value (stacked bar chart).....170	170
Figure 172 - The average sentiment of artists who claimed at least four number ones (line and clustered column chart).	171
Figure 173 - Sentiment score of songs which topped the UK charts for at least eight weeks (line and stacked column chart).	171

Figure 174 - The sentiment rating of every artist from negative to positive (line and clustered column chart).....	172
Figure 175 - The sentiment rating of every artist from negative to positive (line and clustered column chart).....	172
Figure 176 - The sentiment rating of every number one song from negative to positive (line and clustered column chart)	173
Figure 177 - The sentiment rating of every number one song from negative to positive (line and clustered column chart)	173
Figure 178 - The average sentiment of every year of the UK chart, 1952-2020 (donut chart).	174
Figure 179 - 1986, randomly selected, as an example of the sentiment being updated interactively (donut chart).....	174
Figure 180 - The average sentiment for each decade of the UK charts (gauge).....	175
Figure 181 - An interactive tool for the reader to search the average sentiment of their favourite artist or number one song (table).....	175
Figure 182 - An example of using the 'search bar' to look at the sentiment of "Dry Your Eyes", by The Streets (table).....	176
Figure 183 - As demonstrated, more than one artist can be selected from the 'search bar' by holding the 'Ctrl' button on the keyboard (table).	176
Figure 184 - Proof of how a sentiment score can be notably varied over time. The first two years of Beatles' number ones scored highly (a) which is in stark contrast to their number ones which followed over the next five years (b). A considerable dip in polarity (table). 177	

End Notes

ⁱ “Hottentot Venus” refers to South African woman, Sarah Baartman, who was a 19th century “freak show” attraction. Baartman was famed for her unusually large buttocks, a result of a condition she had called “Steatopygia”. She was duly paraded around Europe by typically wealthy white men in what would now be viewed as a form of slavery and overtly racist.

Baartman sadly passed away at the young age of 26, in 1815, with parts of her anatomy, lewdly, showcased to the public in a Paris museum until as recently as 1974. She was repatriated and buried in South Africa at the request of President, Nelson Mandela, in 2002.

Her influence can still be seen presently, some 200 years later, as figures in the public eye such as Kim Kardashian adopt traits of her figure and style (Parkinson, Justin, 2016).

My Power BI report can be found at the following link:

[Sentiment Analysis in Power BI](#)