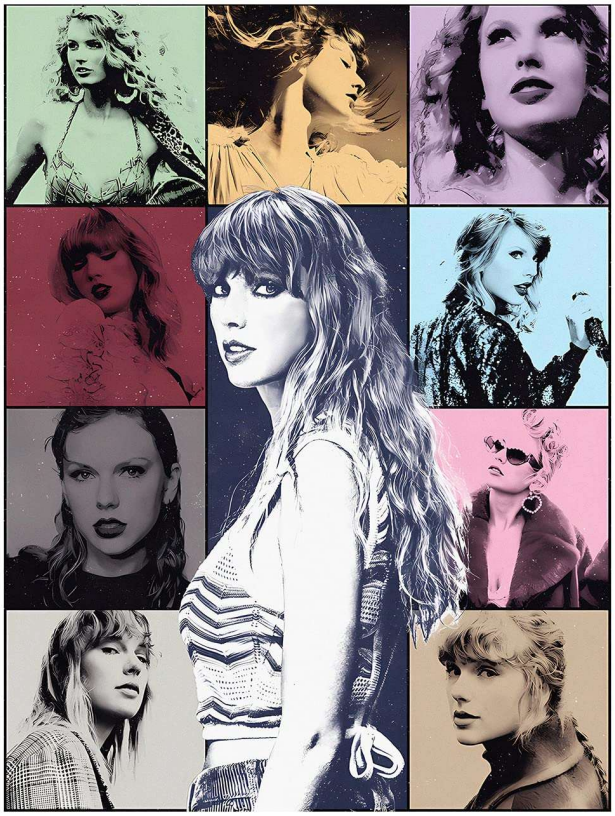
Ca' Foscari University of Venice



The evolution of Taylor Swift's songwriting through data

(*camilla's version*)

A.Y. 2023–2024

Social Media and Web Analytics – Project Report

Camilla Benagli, mat. 895238

Table of contents

1.0 Introduction

Taylor Swift, born on December 13th, 1989, is a prominent American singer and songwriter renowned for her contributions to pop, country, folk, and alternative music genres. Beginning her songwriting journey in her teenage years, Swift released her debut album in 2006 and has since produced ten full-length albums, including her latest work, "Midnights," unveiled in October 2022.

At the moment, she is touring the world with *The Eras Tour*, where she revisits songs from each of her past albums, offering fans a nostalgic voyage spanning nearly two decades of her musical evolution. This work is inspired by it.

Swift's songwriting, often likened to a personal diary, intimately chronicles her life experiences from adolescence to international stardom. This thematic continuity shows her growth from an unknown high schooler to a cultural icon, in a one of a kind coming-of-age style journey. An analysis of her discography seeks to highlight the stylistic shifts and thematic patterns that have defined her career, assessing whether her songwriting has undergone evolution over time.

2.0 Data

This section will report information about the data that was used in the project, where it was sourced from and the initial design choices regarding which of the albums will be included in the analysis.

2.1 Datasets

The analysis in this report is divided into two parts. The first one, delves into statistics related to the general characteristics of the songs, analysing some measures that Spotify calculates for each track. The dataset was published on Kaggle by Jarred Priester and uses Spotify's API to collect information about every Taylor Swift album on Spotify. This dataset updates monthly, so the version currently available at the linked Kaggle page may differ from the one that I worked with for this project. Link.

The second part of the analysis leverages the lyrics of all the songs of the studio albums to study how Taylor Swift's songwriting style and topics changed through time. The lyrics were published in .txt format on Kaggle by Ishika Johari, and they were collected using Genius Lyrics' API. Link.

2.2 Album selection

Both datasets, when considered in full, contain a large number of Swift's album, because they take into consideration all the "alternative" versions of each album, on top of the regular edition. These alternative versions are essentially the same album as the original, only with some marketing differences, but most of the time they include the same exacts songs; for these reasons, we will include only the "normal" version of the albums; specifically, we will consider the ten studio albums that Swift has produced so far, that are: "Taylor Swift" (also occasionally referred to as "Debut" to avoid confusion), "Fearless", "Speak Now", "Red", "1989", "reputation", "Lover", "folklore", "evermore", "Midnights".

A note regarding three albums: *Fearless*, *Speak Now*, and *Red*. In 2019, due to a contractual dispute with her previous record label, Taylor Swift chose to re-record all of her previous albums to gain legal ownership of the associated rights. These re-recordings are commonly referred to as "Taylor's Versions". At the time of analysing the data, the three aforementioned albums had already been re-released: each of these re-recorded albums includes the original songs as well as some new tracks written during the same period as the original release. Consequently, I decided to analyse the Taylor's Version of each album when available, both to incorporate additional songs into the analysis and as a matter of personal ethical consideration.

3.0 Analysis

This section will provide an in-depth explanation of the analyses performed and the findings discovered.

3.1 Spotify data

We first start by analysing the popularity of each album. According to Spotify[1], this value, ranging from 0 to 100, is based on the number of times a song has been played and how recent those plays were, meaning that a song that has been played a lot in last few days will be more popular than a song that has been played the same amount, but a year ago. We compute average popularity of each album and visualize the results on a scatterplot (figure 1). A note on the colours used in all the plots of the report; each Taylor Swift's "era" is characterized by a distinctive colour (the ones in the image used

[1] Spotify for Developers, *Get Artist's Top Tracks*, link (last consulted on March 22nd, 2024).

in the title page): we will be using them throughout the analysis to make it more understandable and as a way to keep a coherent colour scheme.

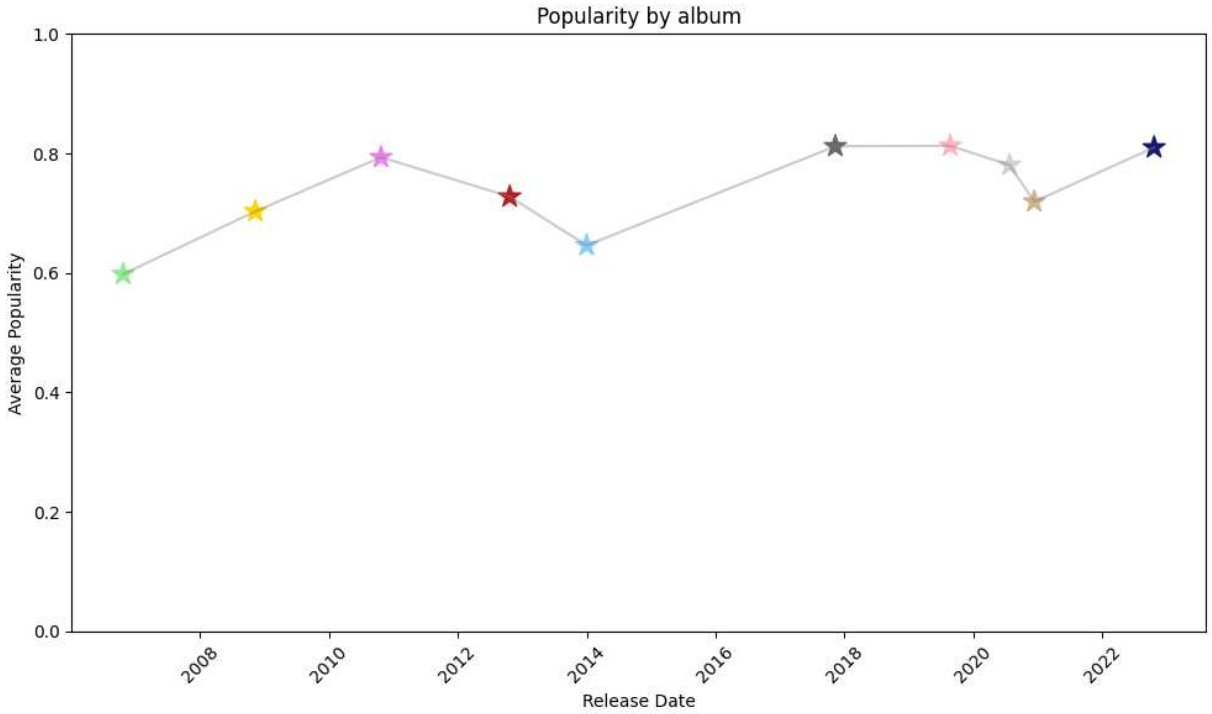


Figure 1 – Average popularity by album

In general, we can see that older albums are less popular, both because of how the metric is calculated and because Swift was at the beginning of her career so her fame, back then, was not as big of an artist as she is right now. We notice, however, a big peak for *Speak Now*, almost on par with the newest albums: this is probably because it is the most recent "Taylor's Version" record, so its popularity was boosted by this newfound wave of fans listening to it after the re-release. As per *folklore* and *evermore*, which show a lower level of popularity, they are the records where the American singer experimented with new genres, such as folk and alternative, far from the more mainstream tunes the audience was used to; also, they were released during the pandemic with very little marketing, so that also needs to be taken into account.

We then look at what Spotify refers to as “danceability” (figure 2), meaning «how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity»[2].

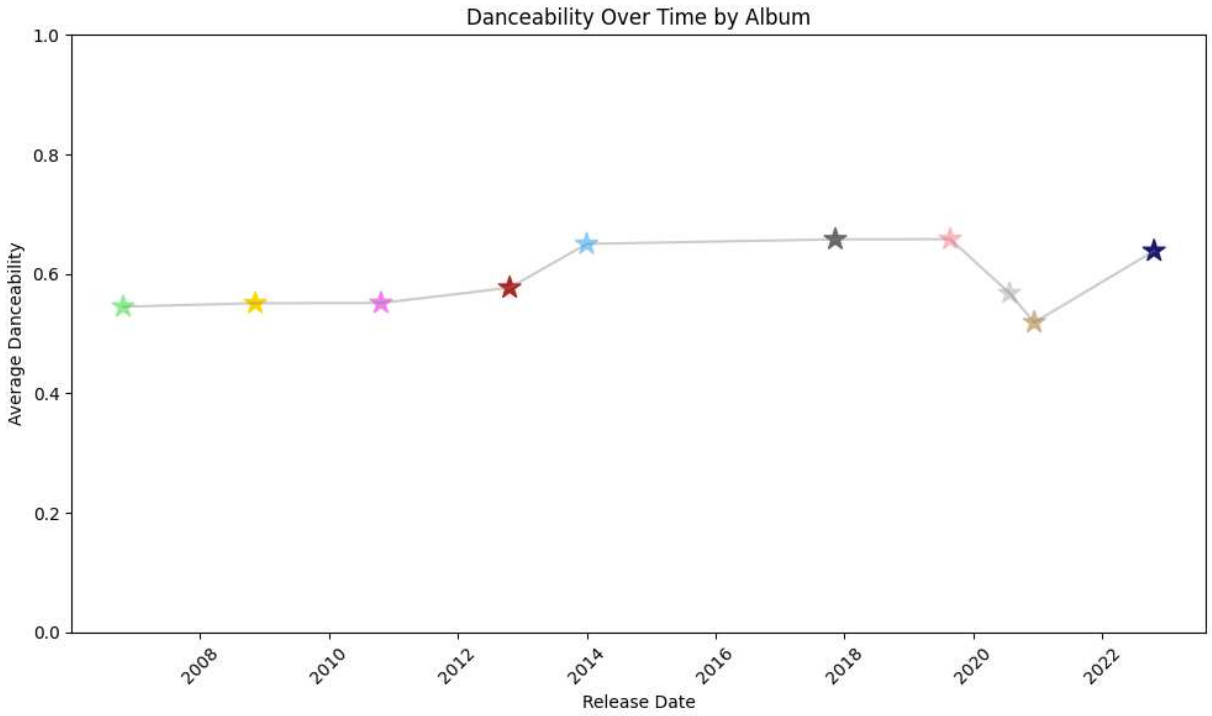


Figure 2 - Average danceability by album

Swift career, which started in the country genre, progressively got more and more pop: this is also reflected in the albums’ danceability, which is increasing as time goes on. The reflection we previously made on *folklore* and *evermore* also stands here: these two albums are to be considered the “odd men out” of the singer’s discography as they clearly distance themselves from anything Taylor Swift has done, especially in the second part of her career.

We took a more in-depth at this (figure 3) and saw that a third of the top 30 songs with the highest danceability are from the *Lover* album.

[2] Spotify for Developers, *Get Track’s Audio Features,* link (last consulted on March 22nd, 2024).

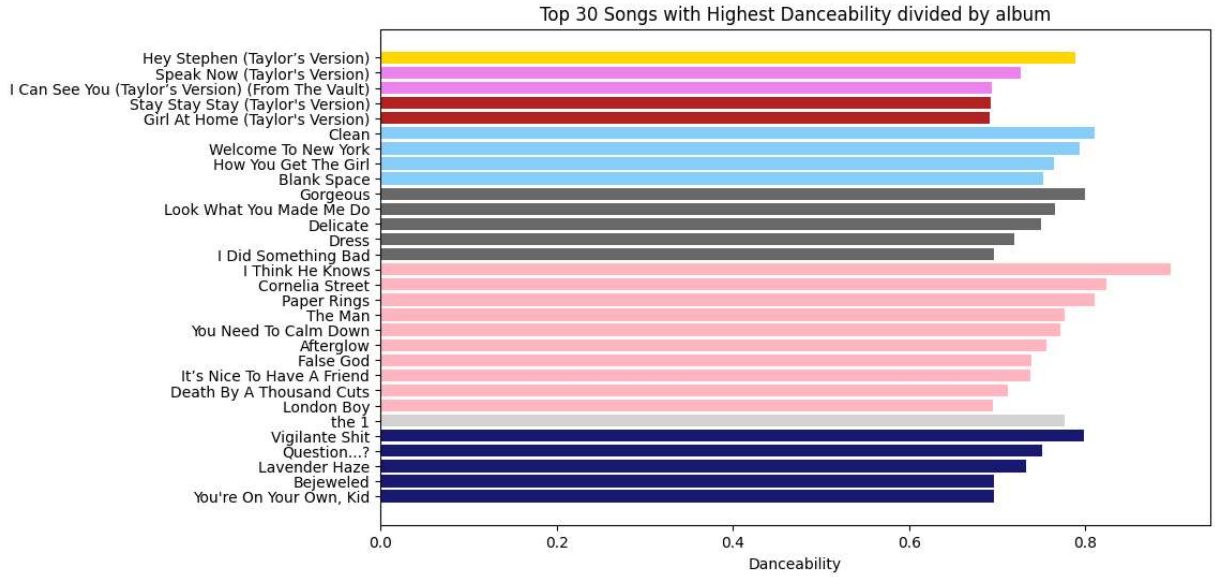


Figure 3 - Top 30 songs with highest danceability divided by album

A question arises: are the songs that are easier to dance to also the most popular ones? Looking at the two plots together (figure 4), we notice that the curves follow a very similar trend in the second part of Swift's career, while the similarity isn't so strong for the first five albums. It's important to note, however, that while the danceability is not affected by the re-releases (because the songs are the same), popularity certainly is; this could explain the increased distance between the two values for *Speak Now* and the curves' behaviour for *1989* (it would be interesting to see if after *1989 Taylor's Version* is published the popularity will stay the same). The correlation between the two variables is 0.387: it's a positive correlation, meaning there is a tendency for more danceable albums – and thus songs - to be more popular, but it's also quite weak, so there are probably other factors that contribute to a track popularity.

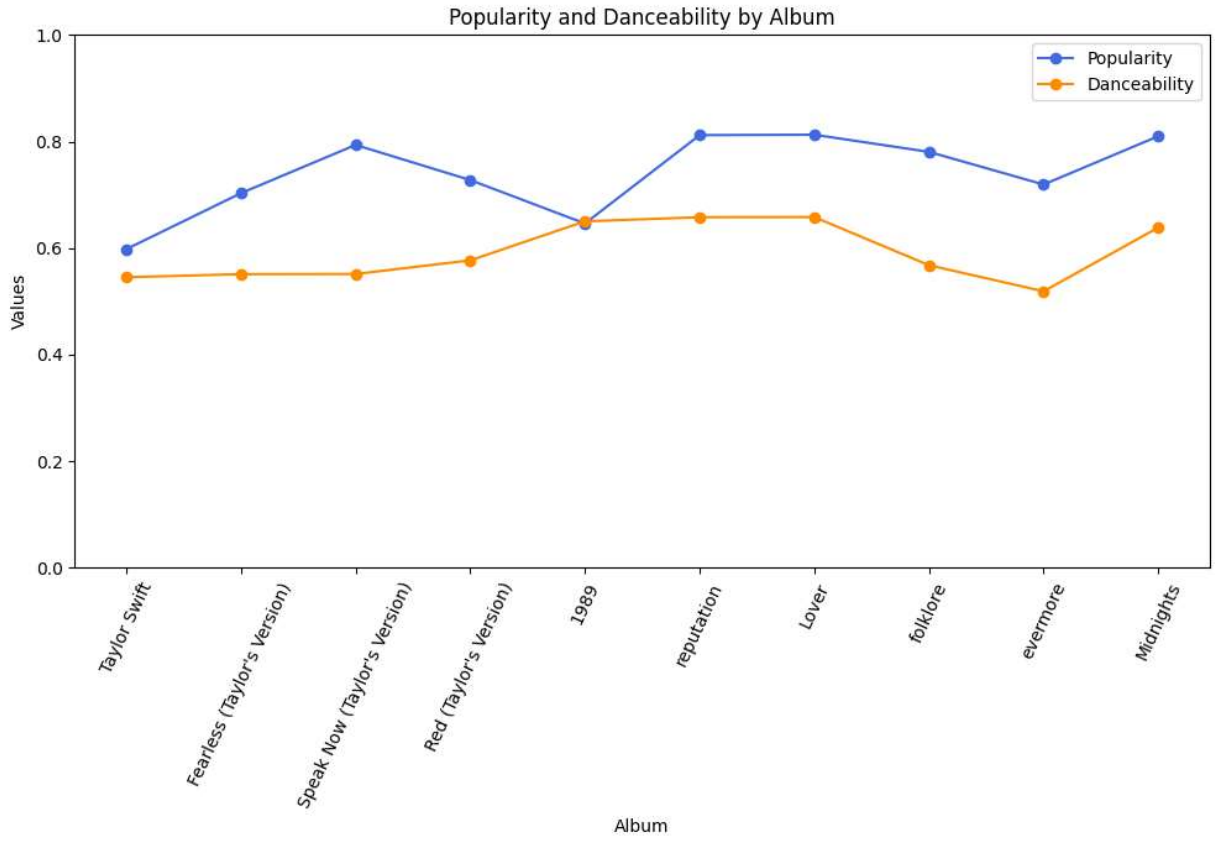


Figure 4 - Popularity vs Danceability

The third value that we looked at was valence (figure 5), which describes «the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry) »[3].

[3] Spotify for Developers, *Get Track's Audio Features,* link (last consulted on March 22nd, 2024).

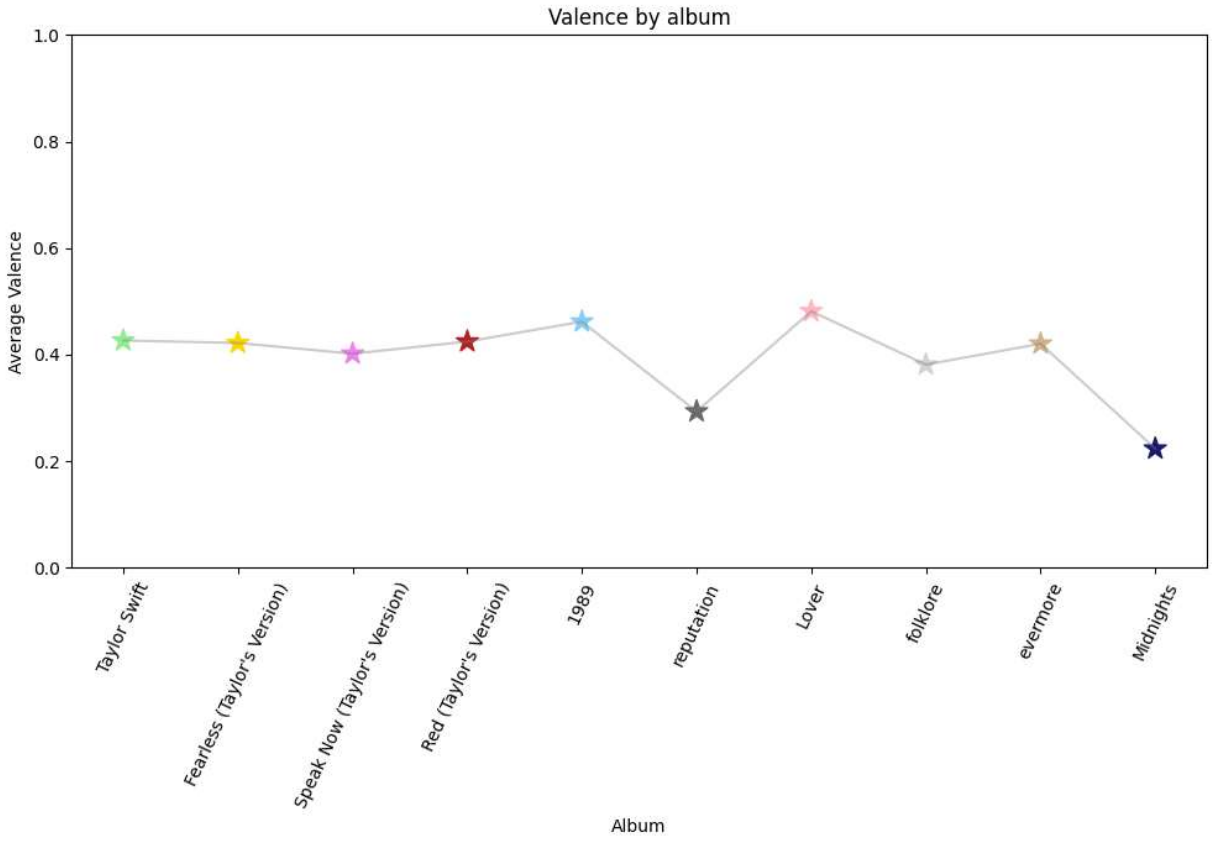


Figure 5 - Valence by album

We can see that, in general, Taylor Swift's albums report a mid-range valence (the average value is 0.397), with the highest value for *Lover*, which, as the name suggest, is a love album mainly written for Swift's then boyfriend, and lowest values for *Reputation* and *Midnights*, by many considered a break-up album from said boyfriend.

Looking at the relationship between valence and danceability (figure 6). It seems that, once again, the first half of the albums follows a similar trend, while the second half doesn't; in particular, *reputation* and *Midnights* are the most interesting data points, because they are generally sad albums (low valence) but with very high danceability. The correlation between the two variables is -0.262: it is a weak negative correlation, meaning that some highly danceable songs might convey emotions that are not necessarily positive or uplifting, but it's not a strong enough relationship to make broad generalizations.

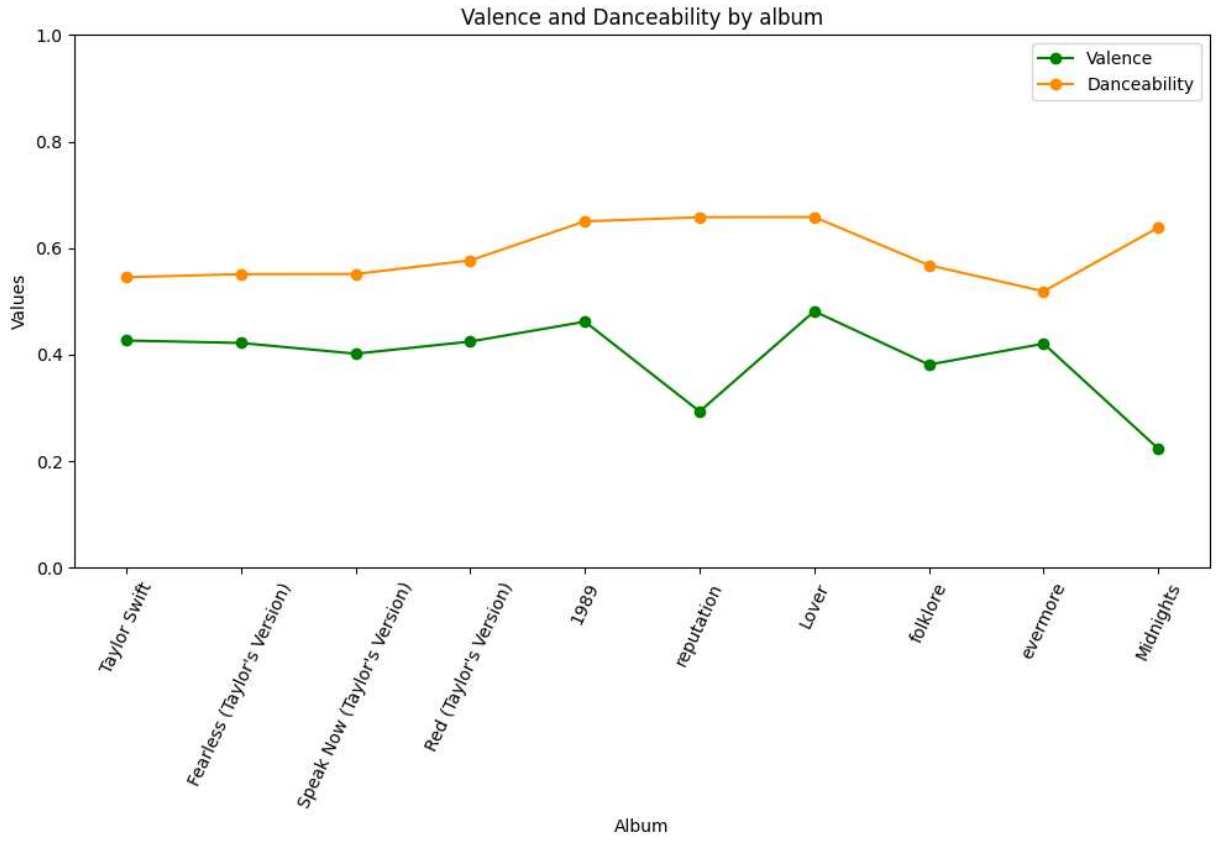


Figure 6 - Valence vs Danceability

Lastly, we looked at the evolution of two other metrics: the average duration of the songs (figure 7) and their tempo (figure 8), meaning their speed and pace, measured in beats per minute.

Taylor Swift's songs' duration span a wide range, and it doesn't seem to be following a particular trend; there are some albums, like *Speak Now*, *folklore* and *evermore* that reports a longer duration, while others, like *Lover* and *Midnights* are much shorter. Five of the albums report outliers, songs that deviate significantly from the median duration; in particular, the furthest outlier is in *Red*, and the song is *All Too Well (10 Minutes Version)*, which, as the name suggests is ten minutes long.

Regarding the average tempo for each album, we can see that in the first part of her discography, Swift seemed to favour higher-paced songs. However, in her more recent albums, the tempo has notably slowed down, with the last three records being her slowest to date. This shift in tempo may reflect Swift's inclination towards exploring deeper and more delicate themes, for which a slower rhythm may be more suitable.

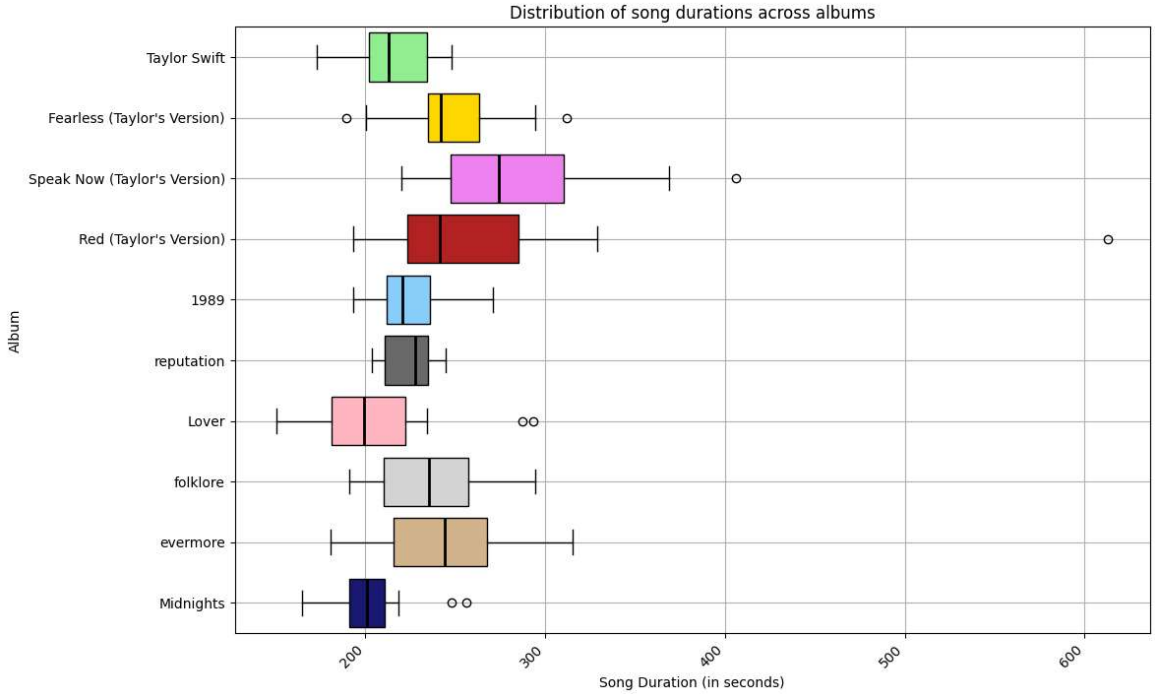


Figure 7 - Average song length by album

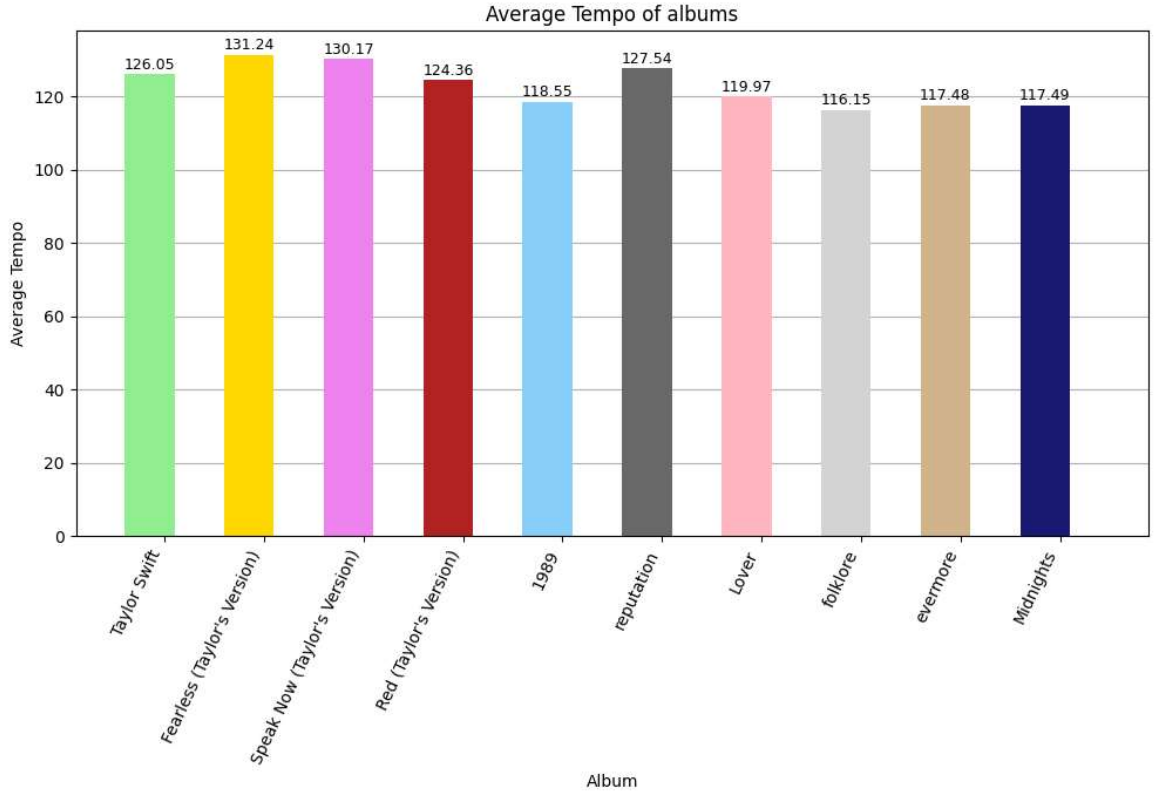


Figure 8 - Average tempo by album

3.2 Lyrics Data

We continue the analysis by taking a deeper look into the lyrics of each song. By looking at the exact words that Taylor Swift's chooses for her work, we hope to get an understanding of the themes that she covers in her songwriting and how they have changed through the years. Based on the lyrics, we will also perform a sentiment analysis and colour analysis. Finally, we will apply a network analysis.

3.2.1 Lyrics analysis

The analysis was performed by looking at the lyrics of each song individually, and then analysing them within their respective album. For each album, we created a histogram of the frequency of the most used fifteen words. At the end, we joined all these analyses into one and created a word cloud of the 150 most used words of Taylor Swift's discography.

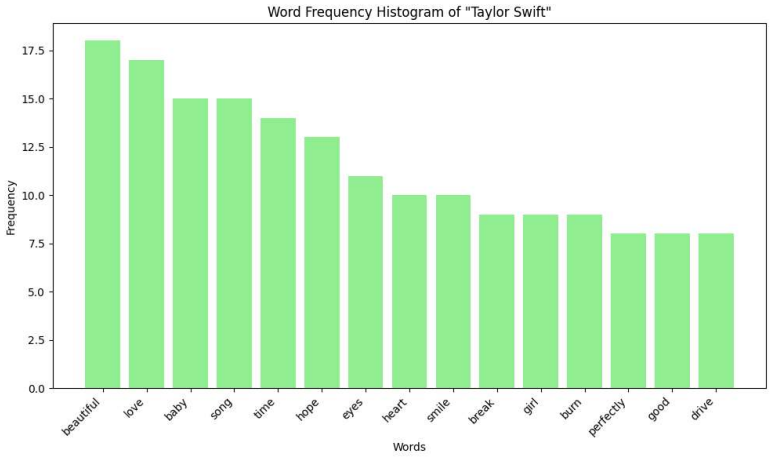


Figure 9 - Word Frequency histogram of "Taylor Swift"

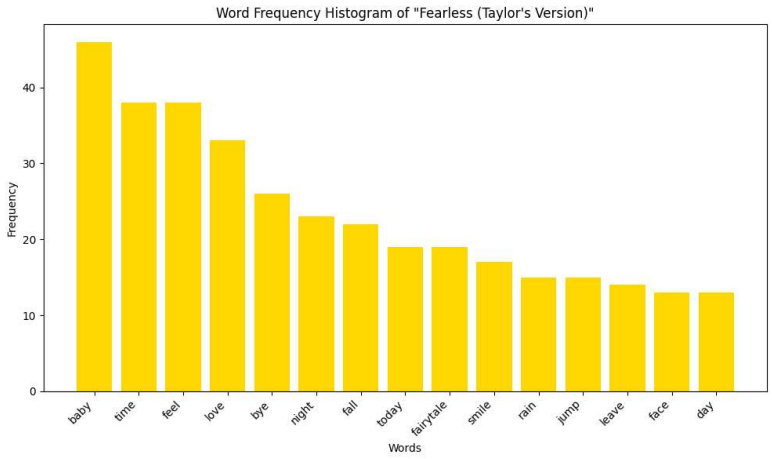


Figure 10 - Word frequency histogram of "Fearless (Taylor's Version)"

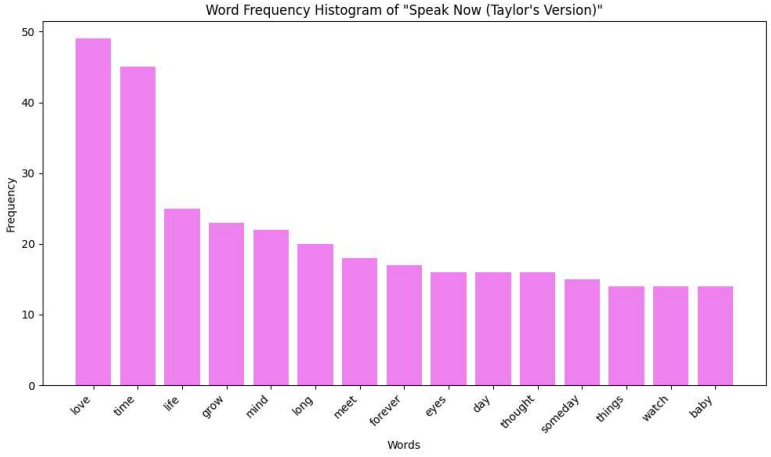


Figure 11 - Word frequency histogram of "Speak Now (Taylor's Version)"

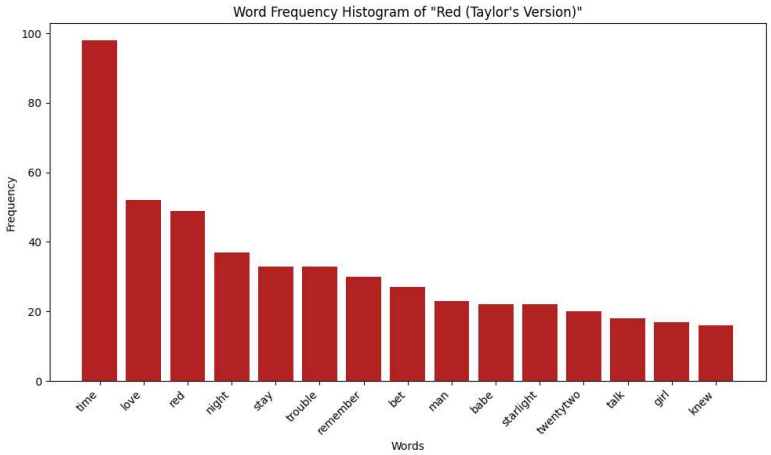


Figure 12 - Word frequency histogram of "Red (Taylor's Version)"

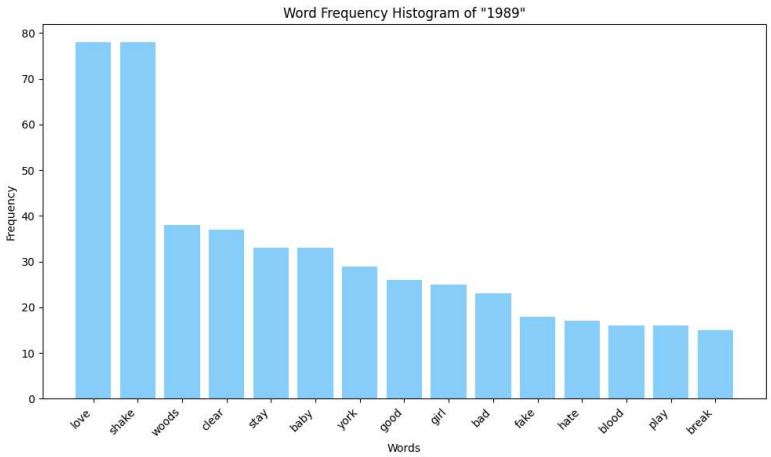


Figure 13 - Word frequency histogram of "1989"

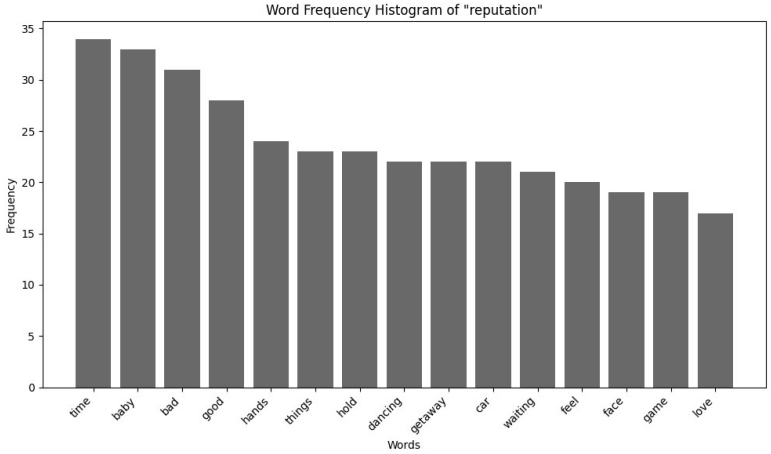


Figure 14 - Word frequency histogram of "reputation"

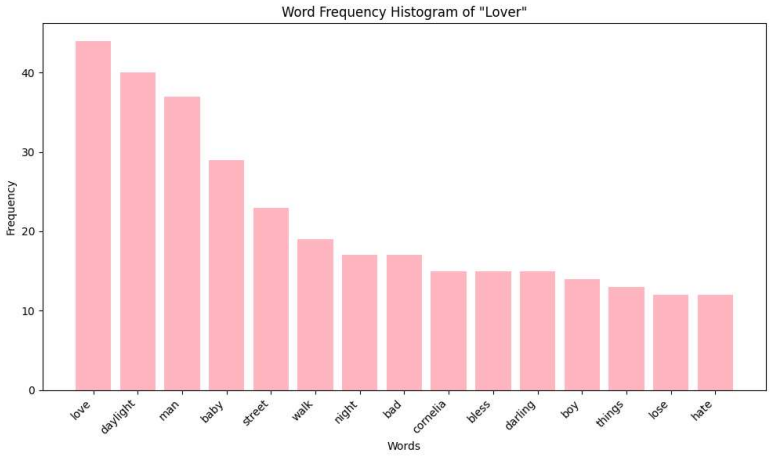


Figure 15 - Word frequency histogram of "Lover"

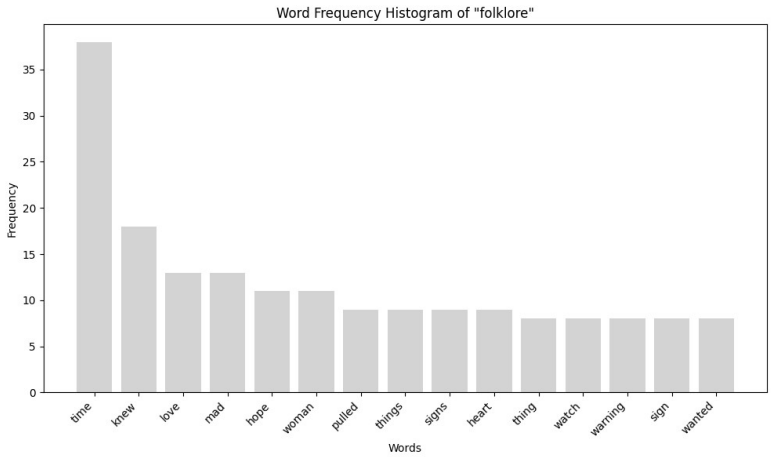


Figure 16 - Word frequency histogram of "folklore"

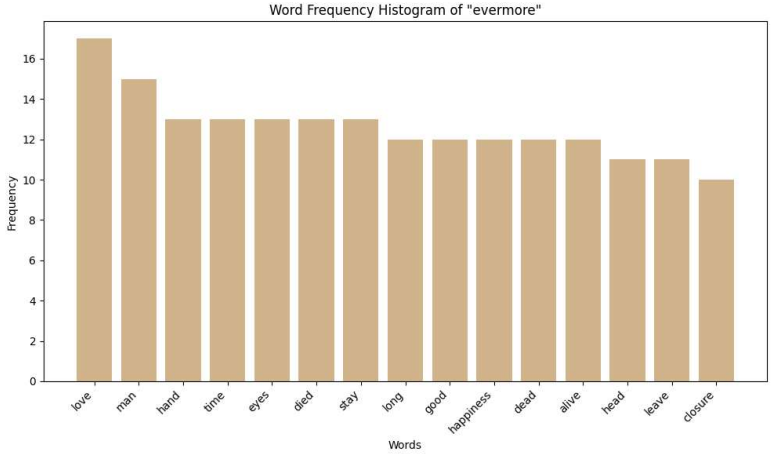


Figure 17 - Word frequency histogram of "evermore"

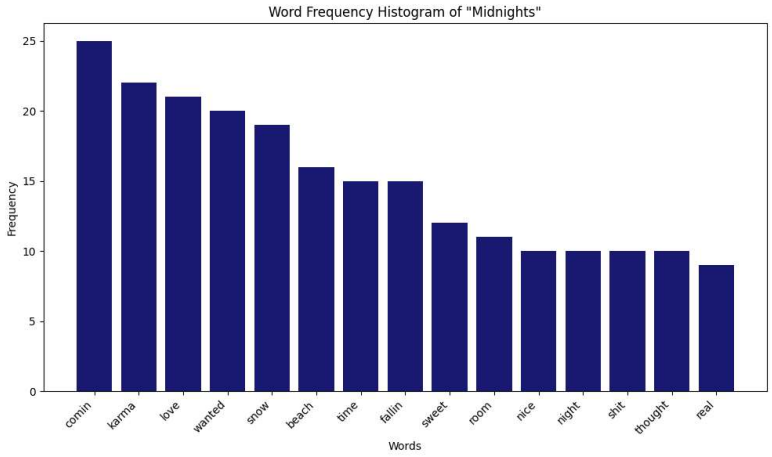


Figure 18 - Word frequency histogram of "Midnights"

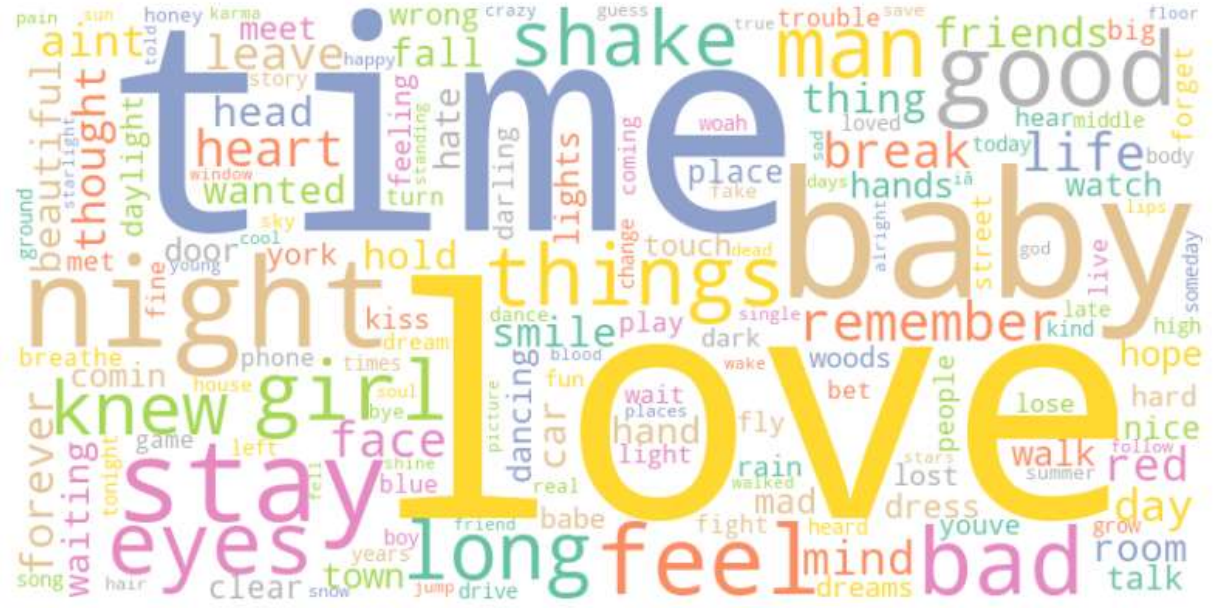


Figure 19 - Word cloud of all the albums' lyrics

We can see that there are some recurring words present in pretty much all the albums, highlighting the cornerstone themes of Taylor Swift's discography. Notably, the most prevalent term is "love". Swift, renowned for her candid portrayal of relationships and heartaches, aligns with a long tradition of songwriters who gravitate towards this universal subject matter. It's however interesting to see how this word is only ranked 15th in the *reputation* album; we can say that this record differs from all the others as it was written at a very particular time in Swift's life, during which the public opinion of her persona was generally unfavourable and negative. The album delves into themes like Swift's public image and how it has been shaped by media scrutiny and public perception, revenge in the face of betrayal, self-reflection, and growth.

The second most used word is "time", which the American singer uses with different connotations through her career. She uses it to explore the passage of time and its impact on relationships and emotions, as she does, for example in "All Too Well", where she regrettably remembers an ended relationship; time is also a vessel for nostalgia, to reminisce about past experiences and triumphs, and "Long Live" from *Speak Now* is the perfect example of this; and it is also used to express self-growth and coming-of-age, illustrating the transformative power of maturity and experience, like Swift does in "Clean" and "Daylight".

While in the first two albums Taylor Swift naively sings about young love and innocence, using words like "hope" and "fairytale", in *Speak Now* she crafts narratives and creates stories around the

emotions that she felt. In *Red*, a more mature Swift goes deeper in her emotional experiences, often evoking painful memories, and using words like "trouble" and "remember". The most common words of 1989 reflect the themes of transformation and reinvention ("shake", "clear") that the singer delves into as she definitively transitions music genres from country to pop. And if *reputation*, as we said before, is the darkest of the albums, *Lover* is the most positive one, analysing love in its different shapes: romantic (like in "Cornelia Street"), friendships ("It's Nice to Have a Friend") and family ("Soon You'll Get Better"); Swifts also sings about empowerment and feminism, touching these themes in "The Man" and "Miss Americana and the Heartbreak Prince". In *folklore* and *evermore* the singer leverages once more her storytelling abilities and transport the listeners to dreamlike landscapes, exploring themes like nostalgia and longing, but also escapism and wilderness. In her latest album, the second most used word is "karma", indicating her new attitude towards her rivals, which doesn't focus anymore on revenge and sabotage like it did in her past work.

3.2.2 Colour analysis

Another intriguing aspect we wanted to explore is the colours mentioned in Taylor Swift's lyrics. We aim to identify any specific shades she uses more frequently and uncover any hidden patterns within them. Once again, the analysis was performed album by album, and then looking at all the discography together.

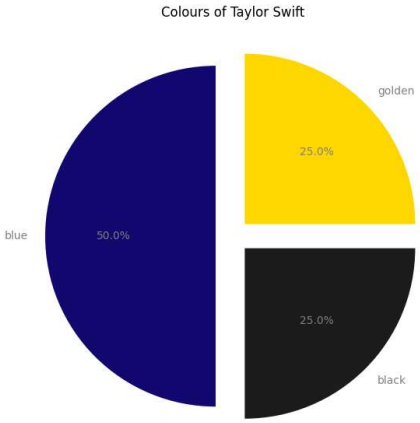


Figure 20 - Colours of "Taylor Swift"

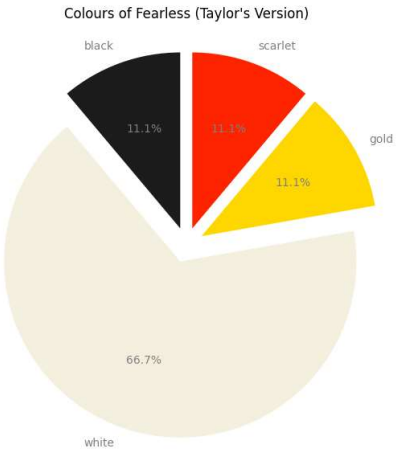


Figure 21 - Colours of "Fearless (Taylor's Version)"

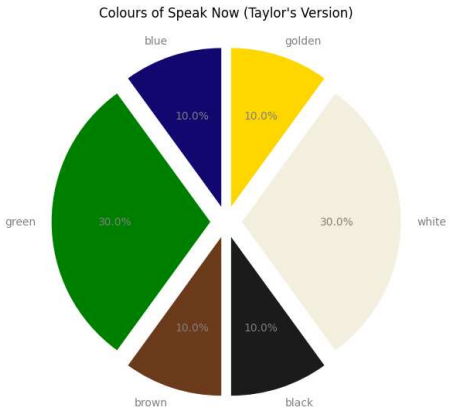


Figure 22 - Colours of "Speak Now (Taylor's Version)"

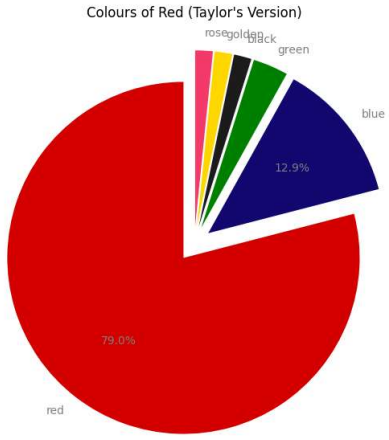


Figure 23 - Colours of "Red (Taylor's Version)"

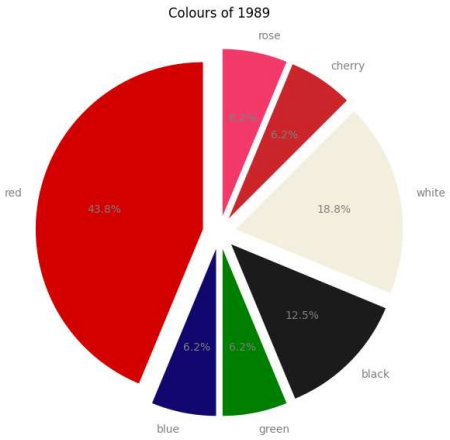


Figure 24 - Colours of "1989"

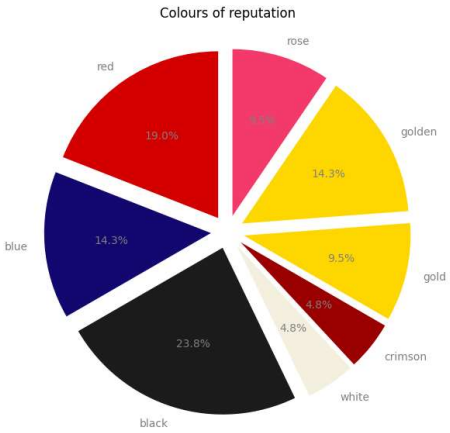


Figure 25 - Colours of "reputation"

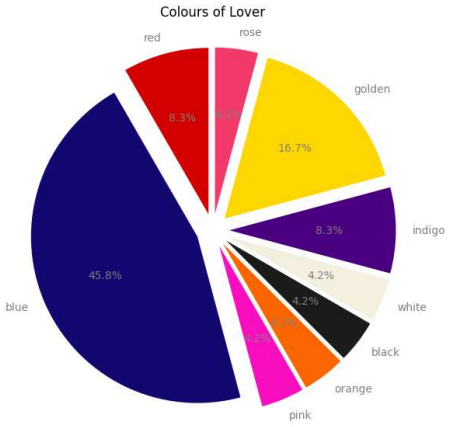


Figure 26 - Colours of "Lover"

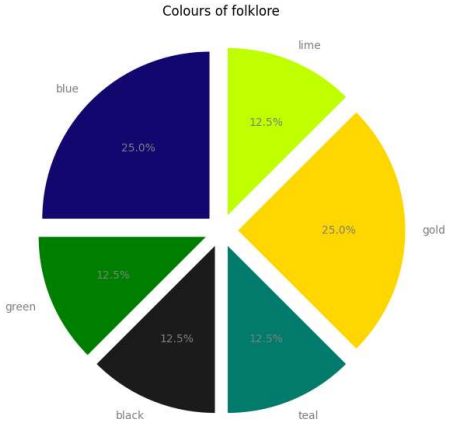


Figure 27 - Colours of "folklore"

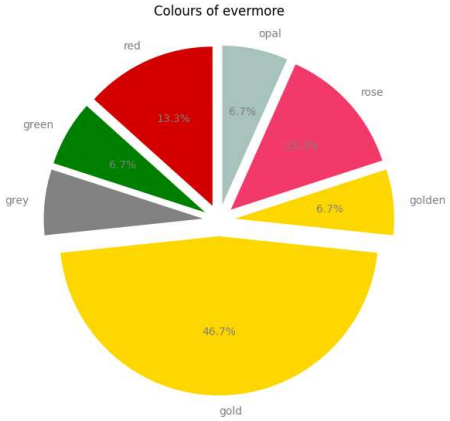


Figure 28 - Colours of "evermore"

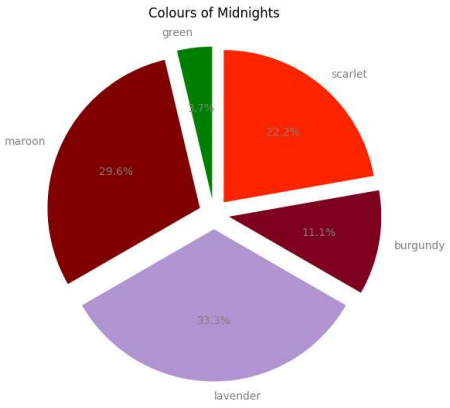


Figure 29 - Colours of "Midnights"

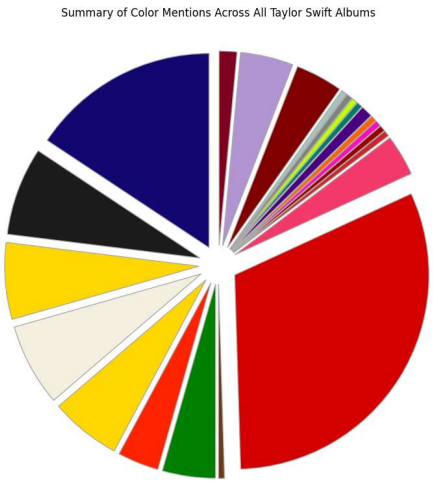


Figure 30 - Colours of all the albums

It's immediately noticeable how the number and variety of colours increases as Swift's discography progresses: while for the firsts she mentions basic colours, she uses more unique shades in more recent albums, such as teal, opal and lavender. In almost every album three colours are present: black, white and gold/golden. Black it's often mentioned as the colour of elegant and chic clothes («The ties were black, the lies were white», in "Getaway Car") or with a negative connotation («The sky turned black like a perfect storm», in "Clean"), white it's often linked to the themes of innocence and fairytale-like scenarios («It's too late for you and your white horse», in "White Horse"), and gold or golden are used to signify something precious («I once believed love would be black and white, but it's golden», from "Daylight"). Other colours that are often mentioned are red and its variations, such as scarlet and crimson, which are used as a metaphor for love (notably in the song "Red", when she sings «Loving him was red») and blue, used the describe the colour of someone's eyes («The way my blue eyes shined put those Georgia stars to shame» from "Tim McGraw") or to signify sadness («Losing him was blue» from "Red").

3.2.3 Sentiment Analysis

We performed a sentiment analysis on the songs, by applying an analyser that, starting from the words of the text, assigns a positive, a negative and a compound score, an overall score that represents the combined sentiment of the text. The compound score ranges from -1 (most negative) to +1 (most positive), and this is how we assigned a "positive" or "negative" connotation to each song.

	Title	Sentiment	Compound Score	Album
89	This Love	Positive	0.999900	1989
117	London Boy	Positive	0.999600	Lover
91	Welcome to New York	Positive	0.999500	1989
90	Out of the Woods	Positive	0.999500	1989
61	Better Man	Positive	0.999200	Red (Taylor's Version)
102	I Did Something Bad	Positive	0.999000	reputation
14	Fearless	Positive	0.998800	Fearless (Taylor's Version)
103	Call It What You Want	Positive	0.998800	reputation
0	Stay Beautiful	Positive	0.998800	Taylor Swift
151	Gold Rush	Positive	0.998600	evermore
118	False God	Positive	0.998500	Lover
166	Snow on the Beach	Positive	0.998500	Midnights
104	This Is Why We Can't Have Nice Things	Positive	0.998400	reputation
62	Red	Positive	0.998100	Red (Taylor's Version)
15	The Best Day	Positive	0.998100	Fearless (Taylor's Version)
1	A Perfectly Good Heart	Positive	0.998000	Taylor Swift
152	Happiness	Positive	0.997900	evermore
39	Ours	Positive	0.997800	Speak Now (Taylor's Version)
63	Message in a Bottle	Positive	0.997700	Red (Taylor's Version)
2	Tim McGraw	Positive	0.997600	Taylor Swift

Figure 31 - Top 20 most positive songs

We subsequently computed the positive rate for each album, representing the percentage of positive songs within an album relative to all tracks on that album.

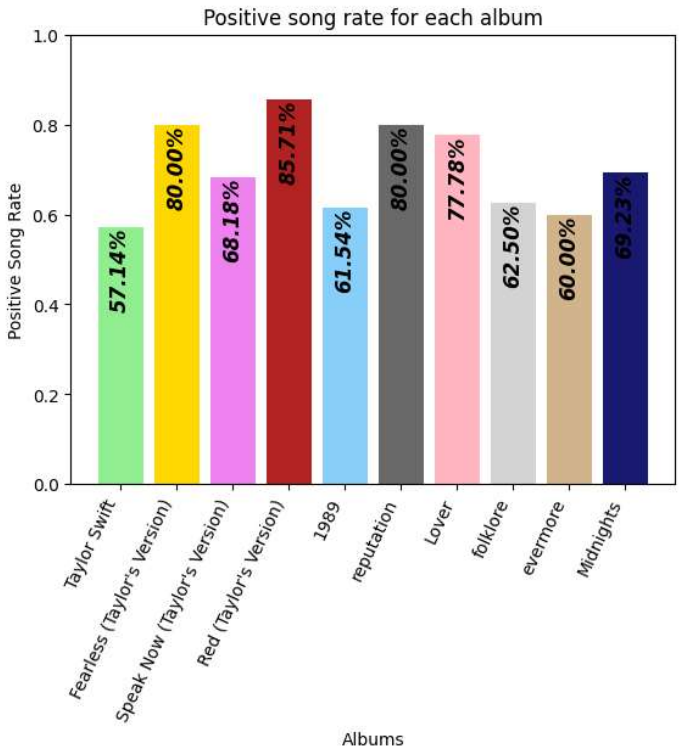
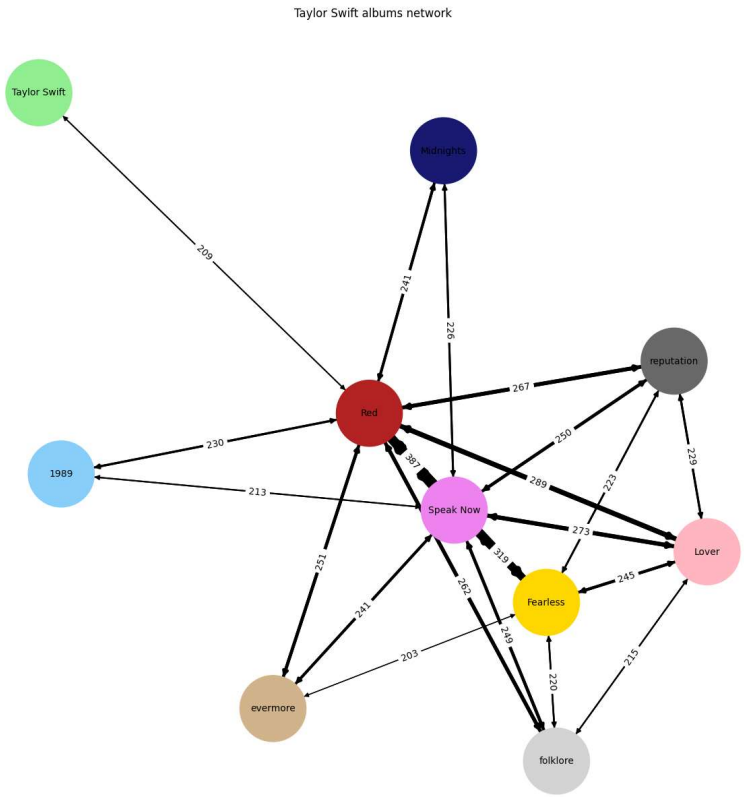


Figure 32 - Album's positive rate

This analysis tells us that Taylor Swift's albums are generally positive, with *Red* being the most positive one and *Taylor Swift* being the least positive one, but still scoring well above the halfway mark. However, we had already performed an exploration of the positiveness of the records when we looked at what Spotify refers to as "valence". There are some notable discrepancies between the two analyses, however, we must remember that Spotify's valence doesn't seem to consider the lyrics of the songs, but just sound variables. While the company is not very transparent about how the valence is calculated, GitHub's user Panagiotis Giannopoulos dissected the value[4] and concluded that it's most likely a computation between danceability, energy and "speechiness" (presence of spoken words). The sentiment analysis we conducted only considers the lyrics. Consequently, a song with predominantly positive lyrics but a melancholic tune might receive differing classifications between the two analyses.

3.2.3 Network Analysis



The last analysis we performed is a network analysis. Every Taylor Swift's album is assigned to a node of the network; an edge is present between two nodes if the albums have at least 200 words

[4] P. Giannopoulos, *Dissecting Spotify Valence* (2022), GitHub repository, link (last consulted on March 26th, 2024).

in common. The number on each edge is the number of words that the two albums share, and boldness of the edge also shows the same information: the more words two albums have in common, the bolder will the line be. The album with the most edges is *Red*, which is connected to all the other records except *Fearless*. It's important to note, however, that *Red* contains 28 songs, the most of any other album, therefore, it's much more likely to share numerous words in common with the others. *Red* shares the strongest connection with its predecessor, *Speak Now*, and *Speak Now* also shares the strongest connection with its predecessor, *Fearless*; this highlights how the second, third and fourth records are deeply connected to each other. Two other very well connected albums are *reputation* and *folklore*, which each have four links. Surprisingly, *folklore* and *evermore*, considered by Swift herself "sister albums", do not share any edges: of course, this does not mean that they don't have anything in common at all, but, at least from a lyrical point of view, the similarities are limited.

This is an undirected network; the total number of links, L, is 20 and the average degree is 4. The graph is connected, meaning that every node can reach another node through a path, and its diameter d_{max} is 2.

4.0 Conclusion

Throughout this analysis, we have noticed how the discography often seemed to be divided in half, with the first part following a similar pattern, while the second appears to be more variable and differentiated. We can see that, for example, in the danceability and the variance, as well as in the network analysis, where *Fearless*, *Speak Now* and *Red* are very connected to each other. This corresponds to reality, since we know that for the first few years of her career, Taylor Swift focused on the country genre, slowly transitioning into pop, while in the second part she also experimented with other genres such as folk and alternative, bringing different sounds and rhythms to her music.

We also have noticed trends that are presents in the whole discography, as we've seen with the words choice, telling us that Swift tends to sing about the same themes, but declined in different ways according to which phase of life she finds herself in; we have also noticed this in the colour analysis, seeing that the singer keeps coming back to the same shades, but with slightly different meanings, going from more naïve and innocent interpretations to more mature and adult thoughts.

Bibliography and sitography

Get Artist's Top Tracks. (n.d.). Retrieved from Spotify for Developers: https://developer.spotify.com/documentation/web-api/reference/get-an-artists-top-tracks

Get Track's Audio Features. (n.d.). Retrieved from Spotify for Developers: https://developer.spotify.com/documentation/web-api/reference/get-audio-features

Giannopoulos, P. (2022, April 5). *Dissecting Spotify Valence*. Retrieved from GitHub: https://developer.spotify.com/documentation/web-api/reference/get-audio-features