

Beneath the Curtain: A Scholarly Journey into the Heart of TikTok's Fame

Jarawura Vincensia Welagamu

Department of Computer Science, University of Ghana

DCIT316: Computational Models for Social Media Mining

Sir Mark Atta Mensah

30th August 2023

Abstract

This research project delves into the multifaceted world of TikTok, a globally popular short-form video platform, to investigate the factors influencing video popularity. With over 1.5 billion users, TikTok has become a cultural phenomenon, and understanding the dynamics of what makes certain videos popular is of paramount importance to creators, users, and platform administrators. The project employs a robust methodology that encompasses various aspects of TikTok content and user behavior. It begins by defining popularity on TikTok, taking into account metrics like likes, shares, comments, views, followers gained, and engagement rates. The data utilized for this study is sourced from a comprehensive dataset spanning three months, providing rich insights into TikTok's ecosystem. The initial phases involve data collection, cleaning, and transformation, ensuring that the dataset is in optimal condition for analysis. Notably, missing values are handled effectively, and data types are appropriately converted. The project then embarks on an exploratory journey into TikTok data, examining distribution patterns, time trends, and conducting comparative analyses. It discovers that the choice of music profoundly impacts video likes, highlighting the significance of audio selection in creating engaging content. Furthermore, it establishes a strong positive correlation between the number of followers a user has and the number of likes their videos receive, emphasizing the importance of growing one's audience. The research also investigates the influence of verification status, finding that verified authors tend to receive more views compared to non-verified authors. This insight can be valuable for content creators aiming to expand their reach. An analysis of video descriptions and author signatures reveals prevalent themes, such as love, life, and time, providing creators with insights into popular content topics. Additionally, sentiment analysis uncovers that videos with positive sentiments garner higher engagement, aligning with the notion that positive content resonates more with viewers. To provide a holistic understanding of TikTok's popularity dynamics, the research develops predictive models and conducts statistical analyses. Ultimately, it offers a comprehensive exploration of TikTok's ecosystem, shedding light on what drives video popularity and providing actionable insights for content creators and platform administrators.

Keywords: TikTok, Popularity, Engagement Metrics, Sentiment Analysis, Data Analysis, User-generated Content.

Introduction

In the ever-evolving landscape of social media, TikTok has emerged as a global phenomenon, captivating the attention of approximately 1.5 billion users worldwide. With its unique blend of bite-sized videos lasting around 60 seconds, TikTok offers a dynamic platform for self-expression, creativity, and engagement. This medium empowers users to craft content with finesse, using an array of innovative tools, filters, effects, and an extensive music library. Beyond

its aesthetic appeal, TikTok's algorithmic prowess is a true marvel, ingeniously deciphering users' preferences to curate personalized content streams. (*TikTok Is Third In Popularity... Here's Why*, n.d.)

Diverse by design, TikTok traverses an array of themes, from the burgeoning influencer culture marked by lucrative monetization avenues and brand partnerships, to its role as a catalyst for education and awareness, housing tutorials and insightful information. However, amidst its undeniable allure, TikTok does not exist without its complexities. Linger concerns surrounding data collection practices and privacy implications loom large, while the intricate interplay between excessive TikTok consumption and its potential impact on mental well-being demands thoughtful consideration.

In the midst of this digital tapestry, we embark on a journey of exploration and analysis, driven by an innate curiosity to unearth the underpinnings of TikTok's popularity. Our purpose extends beyond surface-level observations, delving deep into a meticulously curated dataset spanning a pivotal 3-month period, from March 2022 to December 2022, sourced from Kaggle. This dataset serves as the cornerstone upon which our analytical voyage is built—a foundation that necessitates rigorous data cleansing and refinement. Guided by a quartet of compelling hypotheses, we seek to unravel the intricate fabric that connects user behaviors, engagement patterns, and the elements that propel certain videos to the pinnacles of popularity.

It is important to note that our approach to demographic analysis within the TikTok realm is a measured one. While conventional demographic markers like gender and educational background are gently set aside, our gaze falls upon more contextually illuminating metrics: follower counts and Meta verification. With this discerning focus, our analytical spotlight is trained on engagement metrics as we endeavor to illuminate the nuanced dance between content creators, their audience, and the resonance that propels a video to viral status.

As we embark on this endeavor, we do so with a dual purpose: to decipher TikTok's intricate popularity dynamics and to contribute meaningful insights to a space that continues to shape modern communication and expression. Through meticulous analysis and a steadfast commitment to objectivity, we strive to illuminate the often-elusive threads that weave the fabric of TikTok stardom.

Problem statement

The rapid rise of social media platforms has led to a surge in user-generated content, with TikTok emerging as a prominent player in the digital landscape. With over 1.5 billion users worldwide, TikTok has become a hub for short-form video content, encompassing a wide array of themes and creative expressions. *(21 Essential TikTok Statistics You Need to Know in 2023, n.d.)* As the popularity of TikTok continues to grow, understanding the factors that contribute to video engagement and viewership becomes increasingly essential for creators, users, and platform managers alike.

This project aims to explore and analyze the underlying factors that drive video popularity on TikTok. Specifically, the research seeks to investigate the impact of various elements, such as video content, user characteristics, sentiment, and music selection, on the engagement metrics of 'playCount' and 'likes.' By delving into these factors, the project intends to provide insights that shed light on the intricate interplay between content creation, audience preferences, and viewer engagement.

Key Objectives:

- Identify the relationship between video content attributes, such as video description and user signature, and the corresponding sentiments expressed.

- Examine the correlation between user characteristics, such as verification status and number of fans, with engagement metrics ('playCount' and 'likes').
- Investigate the association between the type of music used in videos and the level of engagement they receive.
- Analyze the prevalence of themes and sentiments in video descriptions and user signatures to discern prevailing trends and patterns.
- Formulate actionable recommendations for content creators based on the identified factors that contribute to video popularity.

Through comprehensive data analysis and interpretation, this project seeks to provide valuable insights into the dynamics of TikTok's content ecosystem. The findings have the potential to guide content creators in crafting impactful videos, assist platform managers in understanding user preferences, and contribute to a deeper understanding of the factors driving engagement on one of the most influential social media platforms of our time.

Scope:

1. **Data Analysis:** The project will focus on analyzing a specific TikTok dataset spanning a 3-month period. This includes examining various attributes such as video content, user characteristics, sentiment, and music selection.
2. **Engagement Metrics:** The scope includes investigating the relationship between the analyzed factors and engagement metrics, specifically 'playCount' and 'likes.' These metrics will serve as indicators of video popularity.

3. **Sentiment Analysis:** Sentiment analysis will be conducted on both video descriptions and user signatures to identify prevailing emotional tones and themes.

4. **Correlation and Association:** The project will explore correlations between user attributes, music type, sentiment, and engagement metrics. Additionally, associations between music type and engagement will be analyzed.

5. **Recommendations:** Based on the findings, the project aims to provide actionable recommendations to content creators for enhancing video engagement.

Limitations:

1. **Dataset:** The analysis relies on the available dataset, which may not encompass the entirety of TikTok content. It might not represent all themes, genres, and user behaviors on the platform.

2. **Contextual Factors:** The dataset's temporal scope might not capture broader trends over time, and cultural context influencing user preferences might be limited.

3. **Causation:** While correlations can be identified, establishing causation between factors and engagement metrics requires more in-depth experimentation and analysis.

4. **Subjectivity of Sentiment Analysis:** Sentiment analysis algorithms might not capture nuanced emotional expressions accurately, potentially leading to misinterpretations.

5. **Generalization:** Findings may be specific to the dataset analyzed and might not generalize to all TikTok content or user behavior.

6. **External Factors:** External events or trends occurring during the data collection period could impact engagement metrics, affecting the study's outcomes.

7. **Privacy and Ethical Considerations:** The project should adhere to privacy regulations and ethical considerations related to user data analysis.

8. **Feature Limitations:** Some attributes, such as user signatures, might not encompass the entirety of user sentiments, as individuals might express themselves differently in a brief signature.

9. **Music Influence:** While the analysis can identify an association between music type and engagement, it cannot definitively conclude that music directly causes engagement.

10. **User Behavior Complexity:** Viewer engagement is influenced by a wide range of factors beyond the scope of this project, including video presentation, user interaction, and platform algorithms.

Despite these limitations, the project endeavors to provide valuable insights into the factors contributing to TikTok video popularity, aiding creators and stakeholders in enhancing their understanding of audience preferences and engagement dynamics.

Methodology

Measuring Popularity: Metrics and Indicators

Defining Popularity on TikTok:

Popularity on TikTok can be defined as the extent to which a video captures viewers' attention, resonates with the audience, and elicits interactions. (*TikTok: What It Is, How It Works, and Why It's Popular*, n.d.) A popular video is one that receives a high level of engagement, including likes, shares, comments, and views. Additionally, popular videos might have a broader cultural impact, with their content being referenced or imitated by other users.

Selection of Metrics (Likes, Shares, Views, etc.):

Several metrics contribute to measuring the popularity of a TikTok video:

1. Likes (Hearts): The number of times viewers have indicated their appreciation for the video by tapping the heart button. Likes reflect positive sentiment and engagement.
2. Shares (Re-shares): The number of times users have shared the video on their own profiles or with their followers. Shares indicate that the content resonated enough for users to pass it on.
3. Comments: The number of comments users have left on the video. Comments represent active engagement and discussions around the content.
4. Views (PlayCount): The total number of times the video has been viewed. Views provide an overall sense of reach and exposure.
5. Followers Gained: The number of new followers gained by the video's creator. This metric indicates the video's ability to attract and retain an audience.
6. Engagement Rate: A composite metric calculated by summing likes, shares, and comments, and then dividing by the number of views. It represents the percentage of viewers who engaged with the video.
7. Time Watched: The total time viewers spent watching the video. This metric gauges viewer interest and captivation.

Data Collection and Preprocessing:

Data Source and Description:

The data for this project was sourced from a dataset available on Kaggle, spanning a 3-month period from March 2022 to December 2022. The dataset includes various attributes related to

TikTok videos, such as video content, user information, engagement metrics, music selection, and more. This data base was a good choice to analyze considering the information that was being provided and its richness in 52 Columns and 1200 rows. This is the link to the dataset:

<https://www.kaggle.com/datasets/anasmahmood000/tiktok-dataset>

Data Cleaning and Transformation:

Before analysis, the dataset underwent several cleaning and transformation steps to ensure its suitability for analysis. Irrelevant columns, duplicates, and columns with excessive missing data were removed. Necessary data type conversions were performed, such as converting 'searchHashtag/views' into numeric format and changing boolean columns to 1s and 0s.

Handling Missing Values:

Columns with missing data were identified using the `'isnull().sum()'` function. Columns with substantial missing values, which wouldn't contribute significantly to the analysis, were dropped. For 'mention' columns, missing values were filled with 0, and a binary transformation was applied to convert mentions to 1s and 0s. This process addressed missing values while retaining valuable information.

Loading the Dataset:

The TikTok dataset was loaded from the Excel file, "dataset_free-tiktok" using the Pandas library. This step is crucial to make the data accessible for further analysis.

Data Transformation:

Various transformations were applied to specific columns, such as converting the 'searchHashtag/views' column into numeric format to analyze engagement metrics. You also

converted boolean columns to binary format and modified mention columns to binary indicators for engagement.

Categorization and Binning:

The 'authorMeta/heart' (likes) column were categorized and binned to create a new categorical variable called 'likes_category'. This transformation enables the analysis of engagement levels based on likes.

Correlation Plot:

A correlation matrix heatmap using Seaborn to visualize relationships between numerical variables. This plot helps identify potential connections and dependencies among features.

By meticulously cleaning and transforming the dataset, it was ensured that the subsequent analyses are accurate and based on reliable data. Data cleaning and transformation are crucial steps in any data analysis project, as they lay the foundation for drawing meaningful insights and making informed conclusions.

Exploratory Data Analysis (EDA):

In the Exploratory Data Analysis (EDA), the research delved deeper into the TikTok dataset to uncover patterns, relationships, and insights that could inform your research on video popularity. Through various visualizations and statistical analyses, a better understanding of the data's characteristics and potential trends was gained.

Descriptive Statistics:

Summary statistics were generated to understand the central tendencies and distribution of numerical variables. This allowed for initial insights into engagement metrics, music views, and user attributes.

Visualizations:

Visualizations, including heatmaps, bar plots, scatter plots and box plots, were generated to visually represent patterns and trends within the data. These visualizations aided in understanding relationships, distributions, and potential outliers.

Word Clouds:

Word clouds were generated to explore prevalent words in video descriptions and user signatures. These word clouds provided insights into common themes and sentiments expressed in these text fields.

Correlation Analysis:

A correlation matrix heatmap to visualize relationships between numerical variables. This matrix helped reveal connections between attributes and potential areas of interest for further analysis. This plot provided insights into potential correlations between engagement metrics and other attributes, helping you identify possible predictors of video popularity.

Sentiment Analysis:

Sentiment analysis was performed on both video descriptions and user signatures. This involved preprocessing text data, removing stopwords, tokenizing, and applying sentiment polarity analysis. The results helped understand the emotional tone of content and its potential impact on engagement.

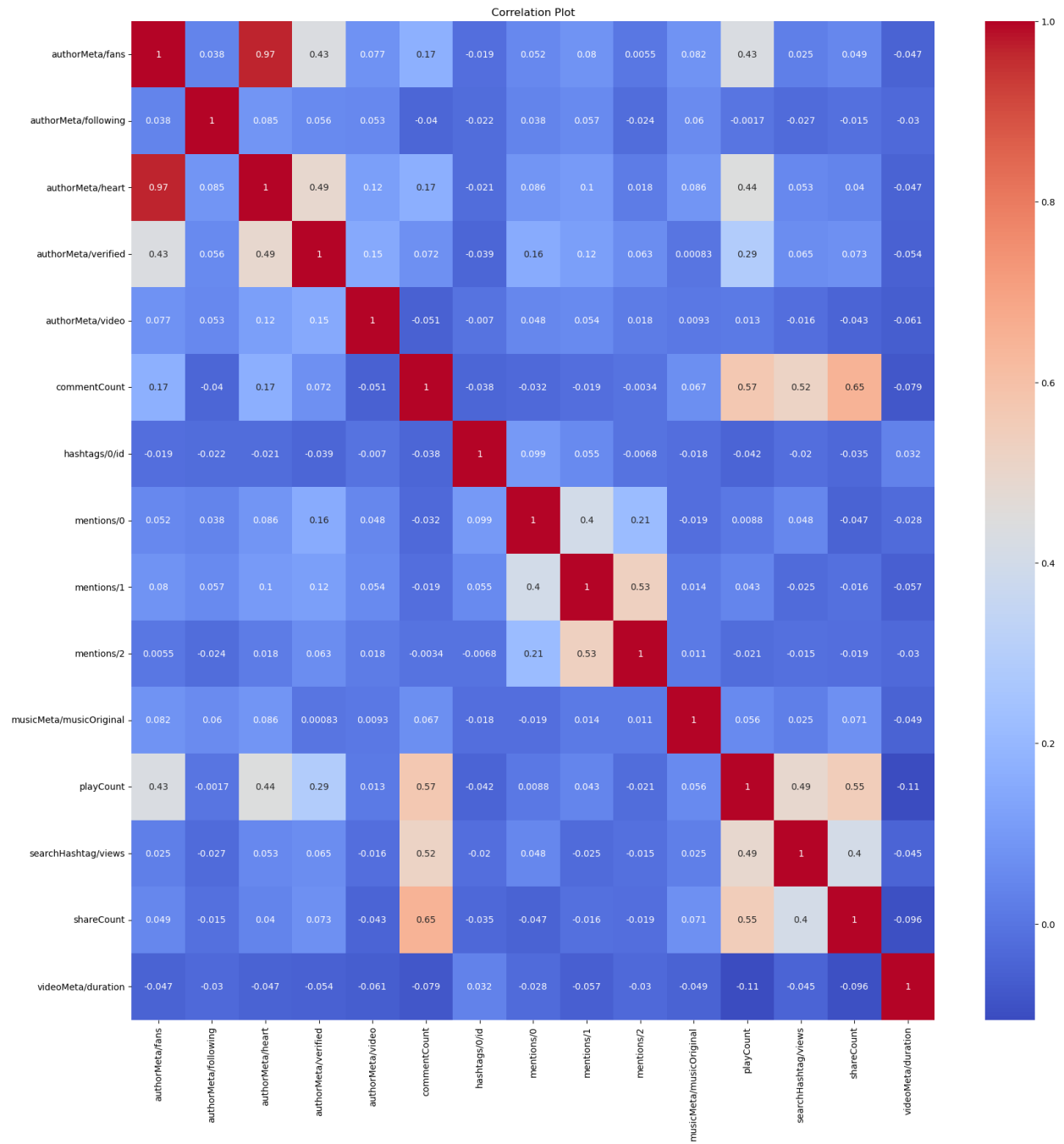
The combination of data collection, cleaning, transformation, and exploratory analysis set the foundation for deeper investigations into the factors influencing TikTok video popularity. The insights gained from these processes informed subsequent analysis and conclusions drawn from the project.

What I did

The correlation matrix that was generated give us insights into the relationships between the engagement metrics and other attributes that identify possible predictors of video popularity.

Based on the correlation values, we formulated the following hypothesis: 1) The type of music used in a TikTok video (as represented by the 'musicMeta/musicName' column) is significantly associated with the number of likes (hearts) the video receives. 2) The number of fans a user has is positively correlated with the number of hearts their videos receive. 3) Verified authors (authorMeta/verified) receive more views (playCount) than non-verified authors. 4) The content of the video description (text) and the author's signature (authorMeta/signature) are associated

with the themes or topics prevalent in the TikTok videos.



Hypothesis 1

The type of music used in a TikTok video (as represented by the 'musicMeta/musicName' column) is significantly associated with the number of likes (hearts) the video

receives. Specifically, certain types of music may lead to a higher number of likes, suggesting that the choice of music can influence viewer engagement on the platform.

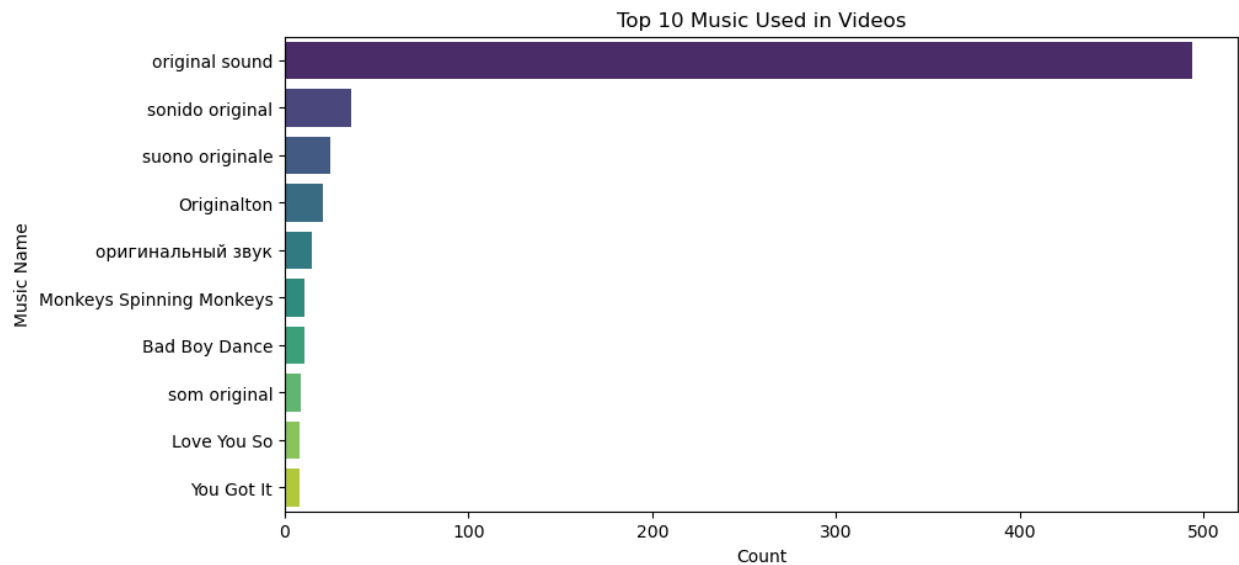
A Chi-Square test of independence is performed to determine the association between the type of music used in TikTok videos (as represented by the 'musicMeta/musicName' column) and the category of likes. This is a suitable approach when dealing with categorical variables. The Chi-Square statistic and p-value are used to assess whether there is a significant relationship between these variables. This test helps you understand whether the choice of music is associated with the level of engagement (likes) a video receives.

From this research, the Chi-square statistic is 1093.423, and the p-value is extremely small ($1.38e-05$). A small p-value (typically ≤ 0.05) indicates strong evidence that the null hypothesis can be rejected. The null hypothesis for a Chi-square test of independence is that there is no association between the two categorical variables.

Therefore, based on the results of the Chi-square test, we can conclude that there is a significant association between the type of music and the category of likes. This means that the type of music used in a video could potentially influence the number of likes it receives.

A bar plot was generated to visually represent the top 10 most frequently used music names in videos. It offers insights into the distribution of music choices within the dataset. The plot provides a clear visualization of the counts for each music name, allowing viewers to easily

identify popular music choices.



The fact that "original sound" is the most frequently used music in the TikTok videos from this dataset indicates that many creators are opting to use their own audio tracks instead of using existing music tracks. This could be for a variety of reasons, such as wanting to create unique content, avoiding copyright issues, or simply because the content of the video doesn't require a specific music track.

In relation to the hypothesis, this could have interesting implications. If "original sound" videos are receiving a significant number of likes, it could suggest that viewers on TikTok appreciate originality and uniqueness in terms of audio content.

In terms of content creation, this insight could be valuable for TikTok creators. If original sounds are popular among viewers, creators might consider producing more content with their own unique audio. However, it's also important to consider the nature of the video content itself - in some cases, a well-chosen existing music track might enhance the video content and attract more engagement.

Hypothesis 2

The number of fans a user has is positively correlated with the number of hearts their videos receive. This hypothesis assumes a relationship between the popularity of a TikTok user (as measured by the number of fans they have) and the level of engagement their videos receive (as measured by the number of hearts).

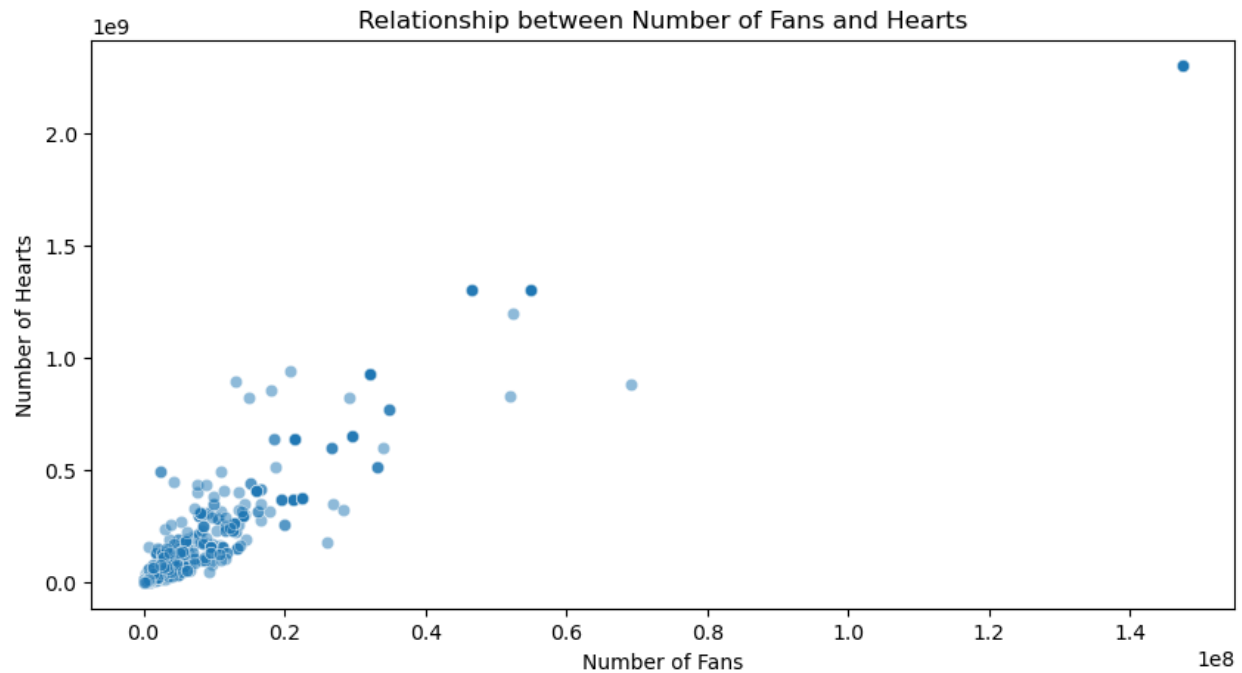
The underlying assumption is that users with more fans are likely to have a wider reach and therefore their videos are more likely to receive a higher number of hearts.

Pearson's correlation analysis is also performed to assess the strength and direction of the linear relationship between continuous variables. In this case, the correlation coefficient and p-value is calculated to determine if there's a significant correlation between the number of fans a user has and the number of hearts (likes) their videos receive. This analysis provides insights into whether an increase in fans is associated with a corresponding increase in video likes.

The Pearson correlation coefficient is approximately 0.967, which is very close to 1. This indicates a very strong positive correlation between the number of fans a user has and the number of hearts their videos receive. In other words, as the number of fans increases, the number of hearts also tends to increase, which supports the hypothesis.

The p-value is 0.0, which is less than 0.05. A small p-value indicates that there is no correlation between the number of fans and the number of hearts. Therefore, the p-value also supports the

hypothesis.



The scatter plot visualizes this correlation, showing a general trend of increasing hearts with increasing fans. However, it's important to note that correlation does not imply causation. While there is a correlation between fans and hearts, this doesn't necessarily mean that increasing one's fan base will directly lead to an increase in hearts. Other factors could also be influencing this relationship.

Hypothesis 3

Verified authors (authorMeta/verified) receive more views (playCount) than non-verified authors. The underlying assumption is that verified authors, who are often more popular or well-known, attract more views on their videos compared to non-verified authors.

Verification on social media platforms is often seen as a mark of authenticity and credibility, which could potentially influence user engagement and views.

The t-test conducted is suitable for comparing the means of two independent groups (verified authors vs. non-verified authors) to assess if there's a significant difference in the distribution of views (playCount) between these groups. The t-statistic and p-value help you determine whether this difference is statistically significant. This analysis provides insights into whether verification status affects the number of views a video receives.

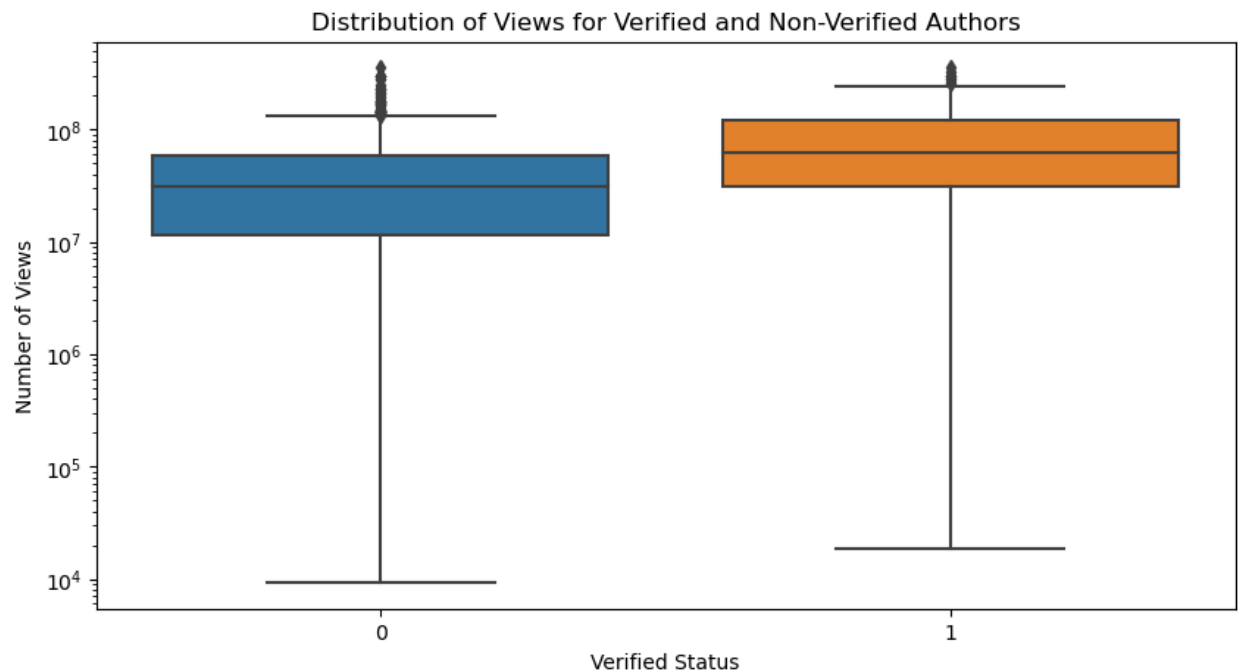
A larger absolute value of the t-statistic indicates a larger difference between the groups. In this case, the t-statistic is approximately 8.96.

A smaller p-value indicates stronger evidence against the null hypothesis. In this case, the p-value is extremely small ($8.52e-18$), indicating very strong evidence that the difference in means is not due to random chance.

The results of the t-test suggest that there is a statistically significant difference in the number of views between verified and non-verified authors, with verified authors receiving more views on average.

A boxplot that allows for a visual comparison of the distribution of 'playCount' values for verified and non-verified authors. The boxplot provides insights into the central tendency, spread, and potential outliers of the data within each group. The use of a logarithmic scale on the y-axis

is chosen to handle the variability in 'playCount' values across different magnitudes.



The boxplot visualizes the distribution of views ('playCount') for verified and non-verified authors. The x-axis represents the verification status, and the y-axis represents the number of views. Note that the y-axis is on a logarithmic scale due to the wide range of views.

From the boxplot, we can observe that verified authors tend to receive more views compared to non-verified authors. The median number of views is higher for verified authors. Additionally, the interquartile range is also larger for verified authors, indicating a wider range of views.

Hypothesis 4

The content of the video description (text) and the author's signature

(authorMeta/signature) are associated with the themes or topics prevalent in the TikTok

videos. The underlying assumption is that the words used in the video description and the author's signature may reflect the content or theme of the video.

For example, videos about cooking may have descriptions or author signatures that frequently use words related to cooking.

Cleaning the data

There seems to be some non-string values (floats) in the 'text' and 'authorMeta/signature' columns. These non-string values need to be handled before applying the text preprocessing function.

The preprocessing function will be used to check the data type of the input before processing. If the input is not a string, the function will return an empty string.

A word cloud visualization is generated that represents the most frequent words in the 'cleaned_text' and 'cleaned_signature' column of the DataFrame. Words that appear more frequently in the text will be displayed in larger font sizes within the cloud, allowing for a quick

overview of the most common terms used in the video descriptions and in their signatures.

Word Cloud for Video Descriptions



Word Cloud for Author Signatures



The first word cloud for video descriptions shows the most common words used in the descriptions of the TikTok videos. Words that appear larger are used more frequently. From this word cloud, we can see that words like "love", "life", "time", "day", and "people" are quite

common in video descriptions. This suggests that many TikTok videos might be about love, life experiences, daily activities, or people.

The second word cloud for author signatures shows the most common words used in the authors' signatures. From this word cloud, we can see that words like "love", "life", "music", "world", and "time" are quite common in author signatures. This suggests that many authors might use their signatures to express their love for life, music, or the world, or to share their thoughts about time.

These insights can help us understand the themes or topics that are prevalent in the TikTok videos and the authors' signatures. If an author wants their videos to resonate with a large audience, they might consider creating videos about **love, life experiences, or daily activities**, as these topics seem to be popular based on the word clouds.

Sentiment analysis

TextBlob for Sentiment Analysis:

The TextBlob library is used for sentiment analysis of both video descriptions ('text' column) and author signatures ('authorMeta/signature' column). TextBlob is a simple and effective tool for performing sentiment analysis, as it provides polarity scores that indicate whether the text expresses positive, negative, or neutral sentiment. By applying sentiment analysis to these text columns, you're able to understand the emotional tone and sentiments conveyed by authors in their videos and signatures.

On the other hand, applying sentiment analysis to the "text" column allows us to analyze the sentiments expressed in the video descriptions. Video descriptions often provide additional context or information about the content of the videos. By analyzing the sentiments in these descriptions, we can gain insights into the overall sentiment or emotional tone of the videos. This

can help us understand whether the videos convey positive, negative, or neutral sentiments and identify any patterns or trends in the emotional content of the videos.

Analyzing the sentiments expressed in the "signature" and "text" columns can help identify the overall sentiment distribution, assess the predominant emotional tone, and explore any relationships between the sentiments expressed by authors and the content of their videos. These insights can be valuable for understanding the emotional impact of the videos, evaluating user engagement, and informing content creation strategies on the TikTok platform.

For the 'text' column:

- Videos with positive sentiments have an average 'playCount' of approximately 54.9 million and an average 'likes' count of approximately 149.2 million.
- Videos with negative sentiments have a lower average 'playCount' of approximately 52.6 million and a lower average 'likes' count of approximately 132.5 million.

For the 'signature' column:

- Videos with positive sentiments have an average 'playCount' of approximately 55.5 million and an average 'likes' count of approximately 152.3 million.
- Videos with negative sentiments have a lower average 'playCount' of approximately 37.2 million and a lower average 'likes' count of approximately 49.4 million.
- These results suggest that videos with positive sentiments tend to have higher engagement in terms of 'playCount' and 'likes' compared to videos with negative sentiments. It indicates that positive sentiment is associated with higher viewer engagement and author likes.

- This finding aligns with the general understanding that positive content tends to resonate more with viewers and generate higher levels of interest and engagement. Positive sentiments may evoke emotions such as happiness, joy, or excitement, which can lead to higher engagement metrics.
- Overall, the analysis indicates a positive correlation between sentiment (positive/negative) and engagement metrics (playCount, likes). Videos with positive sentiments tend to have higher average 'playCount' and 'likes' counts, suggesting that positive sentiment has a positive impact on viewer engagement and author likes.

Results

These results suggest that videos with positive sentiments tend to have higher engagement in terms of 'playCount' and 'likes' compared to videos with negative sentiments. It indicates that positive sentiment is associated with higher viewer engagement and author likes.

This finding aligns with the general understanding that positive content tends to resonate more with viewers and generate higher levels of interest and engagement. Positive sentiments may evoke emotions such as happiness, joy, or excitement, which can lead to higher engagement metrics.

Overall, the analysis indicates a positive correlation between sentiment (positive/negative) and engagement metrics (playCount, likes). Videos with positive sentiments tend to have higher average 'playCount' and 'likes' counts, suggesting that positive sentiment has a positive impact on viewer engagement and author likes.

Discussion

1. Music and likes

As known from neuroscience, music impacts the brain and acts as an aid for memory. From the analyzed this statement is, again, proven. Having identified a significant association between the type of music and number of likes demonstrate that the music chosen by creators, do have an important effect on the users. Using a song or another, changes the experience with a particular video and can, therefore, affect if it is liked, commented or even shared. We centered the analyzed in likes. Nonetheless, this is a matter that can have deeper analysis in terms of how music and likes interact with other scenarios such as the ones mention of commenting and sharing. This may also impact the engagement level of a particular trend and the response of the audience in terms of music liking. For instance, artists could use this data for promoting their music and understanding better their audience.

As we also mention, it is important for both creators and TikTok to know that the source of the music used could potentially influence the number of likes it receives. If original sounds are popular among viewers, creators might consider producing more content with their own unique audio. This particular point cannot neglect the importance of the context and nature of each video. In some cases, a well-chosen existing music track might enhance the video content and attract more engagement.

2. Fans and likes

Social media has always been related to popularity. Based on our analysis, we identified a very strong positive correlation between the number of fans a user has and the number of hearts their

videos receive. In other words, as the number of fans increases, the number of hearts also tends to increase. This insight is useful for content creators. Working on the context is key but it is also to gain more followers. In simple words, more followers imply that there are more people able to see and therefore like the content. Nonetheless, it is essential to know that the relationship is not directly proportional because there are additional factors that should be considered in terms of fans and likes, including if accounts are or not verified as we will recall in the following insight.

3. Verified authors and likes

The initial assumption was that verified authors, who are often more popular or well-known, attract more views on their videos compared to non-verified authors. Even we the hypothesis was proven right, it is important to consider that this doesn't mean that all verified accounts will have more likes compared to the non-verified ones. That would be a generalization without enough data to support. Our hypothesis was proven true in the sample chosen which could lead to an important criteria to have account.

4. What are most videos about?

After analysis the video descriptions, we found out that the common words included are "love", "life", "time", "day", and "people". We might concluded that there is a high rate of videos related to those topics. In terms of the signatures, the common words used are "love", "life", "music", "world", and "time" are quite common in author signatures. This suggests that many authors might use their signatures to express their love for life, music, or the world, or to share their thoughts about time. We understand the objections towards TikTok. Nonetheless, we consider that a constant check of the comments and the content of the videos, without violating the users'

privacy, may help to understand, in reality and with data, the themes or topics that are prevalent in the TikTok videos and the authors' signatures. If an author wants their videos to resonate with a large audience, they might consider creating videos about love, life experiences, or daily activities, as these topics seem to be popular based on the word clouds.

5. Sentimental side of TikTok

The sentiment analysis done to understand the overall sentiment or emotional tone of the videos, helped us to study whether the videos convey positive, negative, or neutral sentiments and identify any patterns or trends in the emotional content of the videos. The insights gained can be valuable for understanding the emotional impact of the videos, evaluating user engagement, and informing content creation strategies on the TikTok platform. These results suggest that videos with positive sentiments tend to have higher engagement in terms of how many times the videos are seen and the likes they have compared to videos with negative sentiments. It indicates that positive sentiment is associated with higher viewer engagement and author likes. This finding aligns with the general understanding that positive content tends to resonate more with viewers and generate higher levels of interest and engagement. Positive sentiments may evoke emotions such as happiness, joy, or excitement, which can lead to higher engagement metrics.

Conclusion

There are a lot of assumptions about TikTok that are taken for granted, that are superficially commented when in reality if going deeper they help to understand different interactions of the users among them and with the platform. Even when our hypothesis were proven true, we believe that there are more assumptions that can be tested and that can help to change

information about TikTok that has been giving as facts without a valid support. This is the link to the code : <https://github.com/Vhincey/316-Final-Project>

References

21 Essential TikTok Statistics You Need to Know in 2023. (n.d.). Retrieved September 1, 2023, from <https://thesocialshepherd.com/blog/tiktok-statistics>

TikTok Is Third In Popularity... Here's Why. (n.d.). Retrieved September 1, 2023, from <https://www.forbes.com/sites/ashleystahl/2023/07/15/tiktok-is-third-in-popularity-heres-why/?sh=72df34923838>

TikTok: What It Is, How It Works, and Why It's Popular. (n.d.). Retrieved September 1, 2023, from <https://www.investopedia.com/what-is-tiktok-4588933>