TikTok

A report of data visualisation to explain what makes TikTok' videos (super) popular

Table of content:

1. Introduction

TikTok undoubtedly is the hottest social media nowadays. What is the lure of these trendy short videos that made people so addicted? Understanding the secrets of how TikTok videos became viral can perhaps uncover user psychology and preferences and help companies steer their marketing strategies.

To complement the New York Times article “what makes TikTok' videos (super-)popular”, three corresponding statistical charts have been designed to guide the audience. Each graph reveals why TikTok videos are so popular: music, video duration and sentiment.

2. Visualisation

Across the 3 statistical charts, the title font and text font are picked to match the style of the New York Times as much as possible. The fonts are Cheltenhem and Franklin Gothic respectively. The colours are chosen in accordance with the TikTok logo palette: #FE2C55 (red), #25F4EE (turquoise), and #000000(black).

2.1 Music

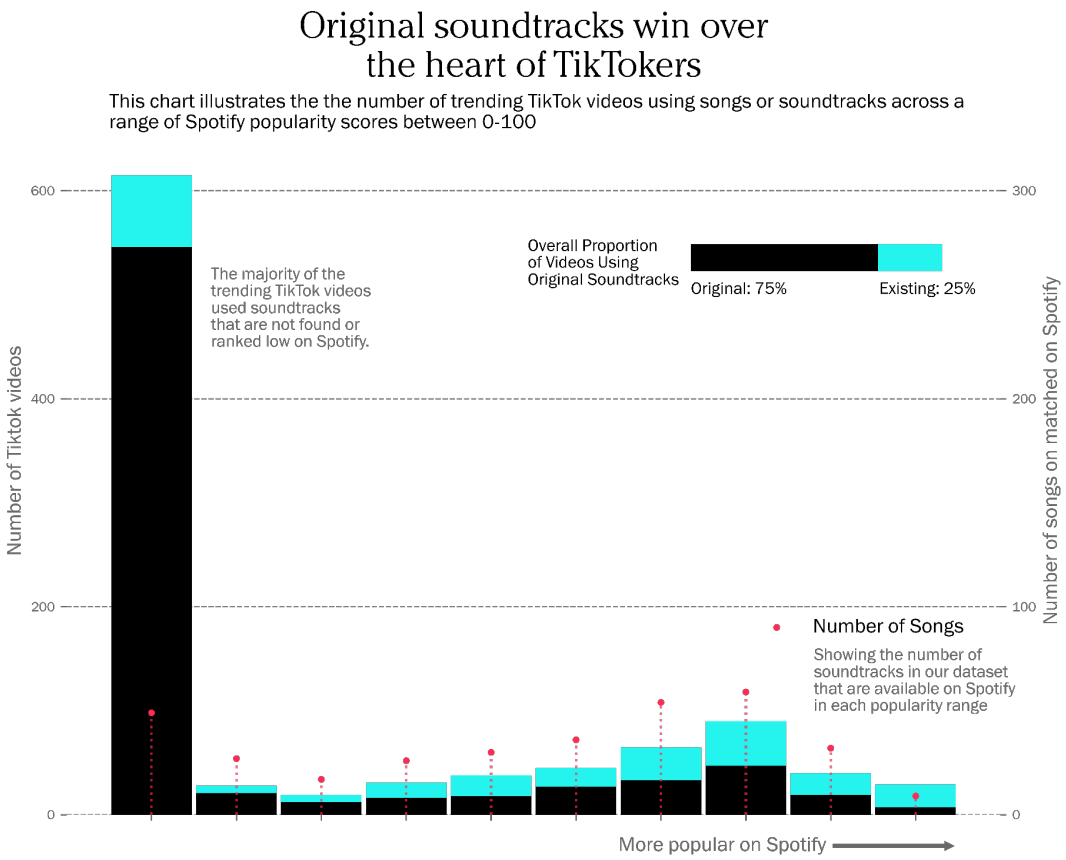

Figure 1. Original soundtracks win over the heart of TikTokers

2.1.1 Information Delivery

The soundtracks of TikTok videos can be divided into two main categories: existing and original soundtracks. Existing soundtracks are mainly commercially produced music, their remix versions, or sound clips from other online videos; original soundtracks can be a completely new soundtrack (music or background of a video) or sound clips from other TikTok videos.

To understand the sound choice behaviour on TikTok, we merged the Spotify popularity data with the trending TikTok video data. Around 75% of the trending TikTok videos used original soundtracks, and the majority of these videos used music not found or ranked low on Spotify. Considering that a limited number of songs have popularity over 90, trending videos sometimes also employ popular commercial songs. Still, the number is trivial compared to the number of videos that use soundtracks with low popularity or unknown on Spotify.

Viral TikTok songs often are a mix of huge commercial hits, remixes from smaller independent producers, and comedic sound clips from other TikTok videos. TikTok has functionality that facilitates reusing the original sound of another video, which explains the disproportionately high number of trending videos using unmatched soundtracks.

2.1.2 Design Choice and Rationale

The chart utilised a stacked histogram to represent the distribution of Spotify popularity of soundtracks used in trending TikTok videos and is separated by original vs existing using black and turquoise to understand how the split evolves across popularity.

The red lollipop plot shows the number of songs matched on Spotify for each popularity range. It provides the context of proportion between the number of songs matched and the number of trending videos made. Without this context, the audience could be misled into thinking popular songs beyond the popularity score of 90 are neglected by TikTokers. In contrast, the truth is that there are only a handful of extremely popular songs, hence by proportion, the number of trending videos is lower. The size of this plot is relatively small as what it represents is not the main message of the chart. Dual axes are also used to facilitate the comparison between the number of songs matched and the number of trending videos.

The bar chart on the top-right provides an overview of the data by showing the overall proportion of original and existing soundtracks used by trending videos. It also acts as a legend for the stacked histogram.

2.2 Duration

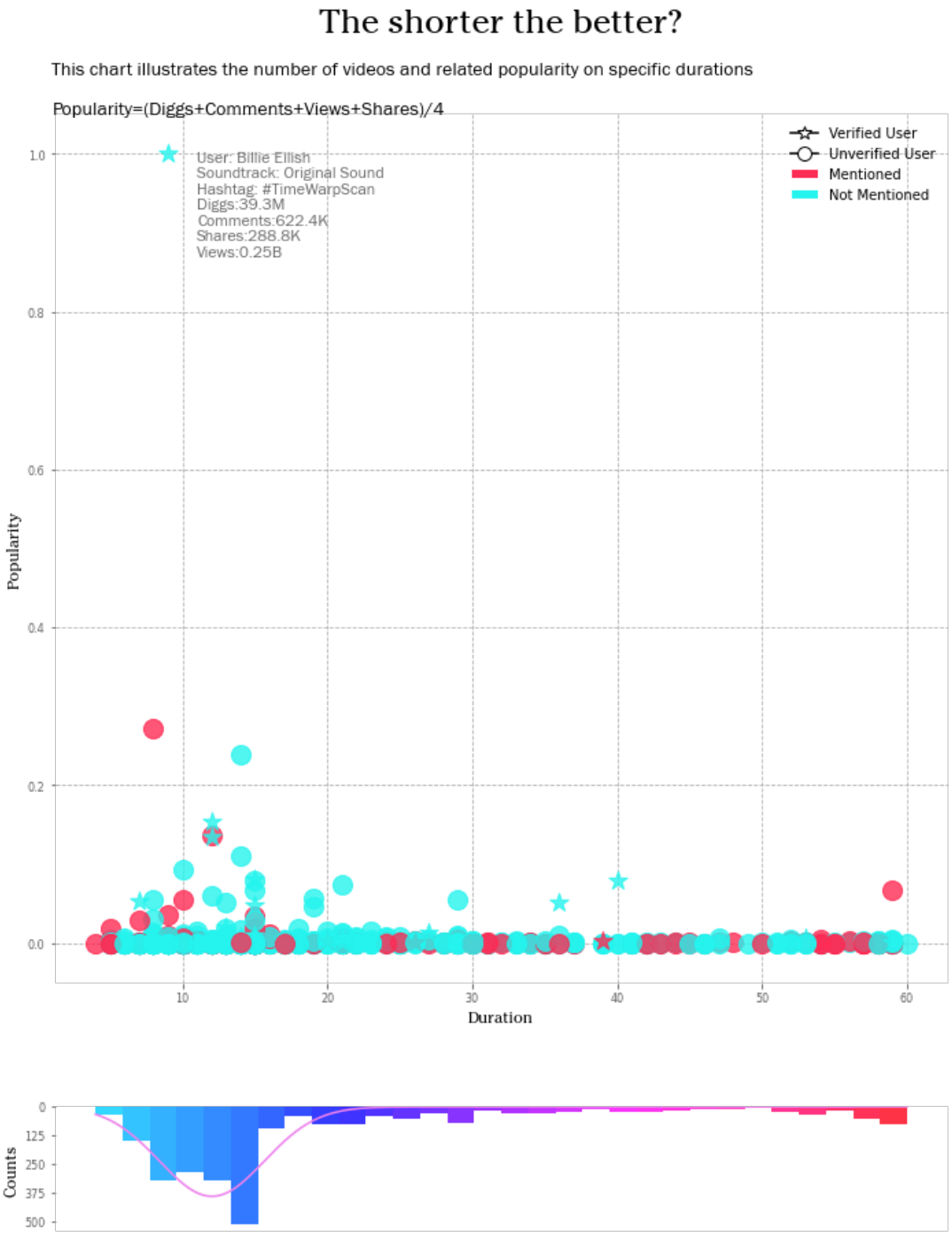

Figure 2 The shorter the better?

2.2.1 Information Delivery

The graph demonstrates that for the 1,000 popular TikTok videos in the dataset. The majority duration is within 5-20s, and the shorter videos are relatively far more popular. The most

popular video is from Billie Ellish and is less than 10s. The short duration of TikTok videos enables users to access information and entertain themselves in a very short time. Most video creators are non-verified, reflecting that everyone could gain great attention on TikTok. However, verified users' videos can be more popular. The mention feature allows TikTok videos to be interactive, with half of the videos shown mentioning others in the video. Thus, the figure reveals that a) TikTok videos are short, b) ordinary people can easily gain attention, and c) videos are interactive.

2.2.2 Design Choice and Rationale

This graph combines a scatter plot and a histogram to demonstrate the relationship between the popularity of a TikTok video and the video's duration. The scatter plot reflects the popularity and duration of each video and its distribution. Gridline serves as an aid to audience targeting. Additionally, the graph also contains other dimensions: whether the user is verified and whether the video mentions others or not. This provides alternative information about the video while delivering the main message.

The popularity is calculated according to the four measures within the dataset: diggs, comments, shares and views. As diggs, shares, comments and views all respond to the popularity of a video, we assume that the four variables are equally weighted. The popularity is the average of the sum over the four dimensions, and higher popularity suggests the video is more popular in the dataset. We introduced this parameter to differentiate the popularity of the 1000 TikTok videos, and thus the audience could more intuitively perceive the association between the level of popularity and the characteristics of the video.

To differentiate each data point, in the scatter plot, verified users are represented by stars and non-verified by circles, and videos with others mentioned in description marked in red and vice versa in blue. By distinguishing the colours and shapes, the audience can understand the characteristics of the video at a glance.

The histogram illustrates the distribution of video lengths. Videos of shorter duration are represented as cooler toned colour (blue), and longer duration videos as warmer toned (red). The duration can also be reflected in the horizontal coordinates of the theme. The bar length of the histogram reflects the extent to which the video is aggregated at that duration. The data are also represented by curves after the histogram has been fitted through Gaussian distribution, thus presenting a fat-tailed distribution of the video duration.

Notably, the features of the most popular points are marked on the graph to provide more intuitive insight into the characteristics of the video. The audio track and hashtag used for this video are marked, thus linking the information to Figure 1. music and Figure 3. text sentiment.

2.3 Sentiment analysis of videos in the top 45 hashtags

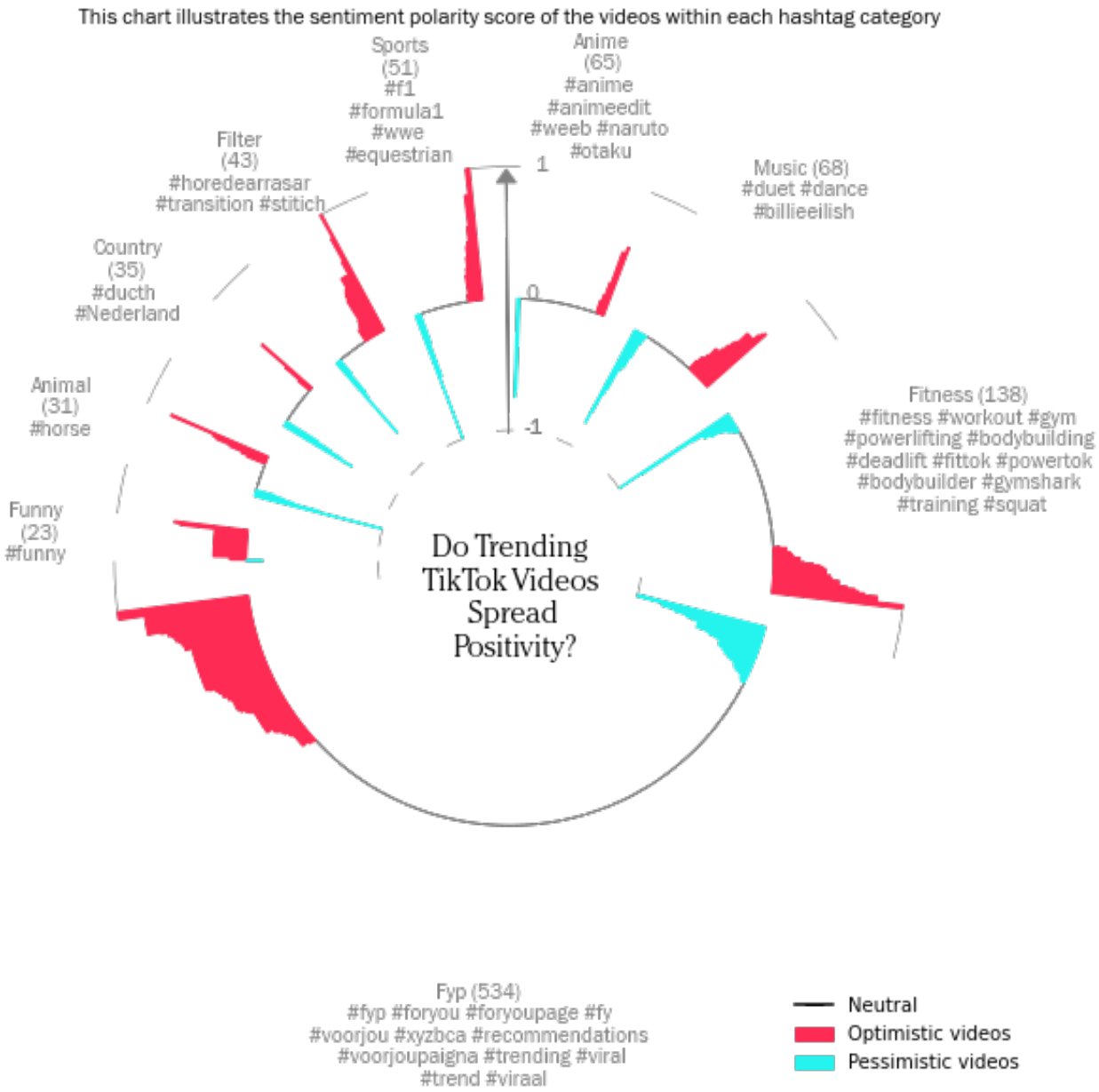

Figure3. Do trending TikTok videos spread positivity?

2.3.1 Information Delivery

Evaluation of popular TikTok videos can help us understand the content that attracts views. One method is to analyse the video's sentiment around hashtags. Hence, we have filtered the hashtags in the TikTok dataset and picked the top 45 hashtags among 1000 popular videos and categorised them into 9 groups: *funny, anime, sports, fyp, sports, country, filter, music and fitness.* The text of each video was given a sentiment score of polarity between -1 and 1, with -1 being the most negative sentiment and the latter being the most positive sentiment.

The highest video count consists of hashtags under fyp, followed by fitness, music, anime, sports, filter, country, animal and funny, as illustrated in figure X, indicating that popular TikTok

videos are not limited to specific content. In addition, many popular Tiktok videos have neutral sentiment, which explains that neutral content can still lead to high popularity in a video. Nevertheless, there is an overall greater number of videos with positive sentiment than negative sentiment. Therefore, the sentiment analysis suggests that popular TikTok videos have more positive content than negative.

Overall, a wide range of sentiments is found, but positive videos appear more frequently in viral TikTok videos.

2.3.2 Design Choices and Rationale

The dataset was presented as a circular bar plot, showing the polarity of each video under different hashtag categories. The chart was chosen as it allows information of different hashtag categories to be consolidated into a place, providing a more straightforward category comparison as a part or a whole. As we have 9 categories and a large dataset of polarity (bars) to display, a circular bar chart can maximise the use of space than a standard bar chart. The title has been designed to place at the center of the graph as this encourages the readers to focus the attention on our core message before interpreting the in-depth information in the chart.

The videos within the same hashtag category were arranged from the lowest polarity to the highest polarity to enable a clear understanding of the distribution of positive and negative videos. The different hashtag categories were also organised into ascending order from the lowest number of videos to the highest. This approach can further emphasise the popularity ranking of each hashtag category. For the reader to have a clear idea of what hashtags are included in each category, the hashtags have been listed out under the category name in the graph. This could prevent readers from misunderstanding that each category has the same quantity of hashtags and give them a better overview of the actual 45 popular hashtags on TikTok. Instead of putting the hashtags in a table format, this method can minimise the density of information display.

Finally, red has been chosen to represent a positive sentiment, and blue has been selected for negative sentiment as these colours represent TikTok colours. The axes label of 1, 0, and -1 is the scale of the polarity score and aims to act as a reference line. Placing it at the centre allows the reader to spot this scale before looking at the scores in detail, enabling a more understandable representation of the bars.

3. Conclusion

In summary, the 3 graphs provide a visualisation to explain why Tiktok' videos are popular from the perspectives of music, duration and sentiment. With the aim of utilising the data-ink ratio, all the graphs consist of several dimensions to provide in-depth data analysis results. To aid the understanding of the article about TikTok, the design choices behind the three charts aim to give the best data result comprehension to the audience of the New York Times.