AI Odyssey Hackathon

Trend Analysis and Content Generation

9^{st} February- 16^{th} February, 2025

Elaborated by:

DS Girls Team

Rahma Aroua

Dhouha Meliane

Organized by:

Dione Protocol  **X**  **GDG Carthage**

Academic Year

2024 - 2025

Contents

List of Figures

Abstract

In today's rapidly evolving digital landscape, staying ahead of social media trends is essential for maximizing engagement and reach. This project harnesses the power of artificial intelligence to identify and capitalize on trending topics across TikTok and X (Twitter), enabling the seamless generation of viral content. By integrating AI-driven trend analysis with automated content creation, the system extracts real-time trending topics and hashtags using platform APIs and employs machine learning models to analyze patterns in viral content. The insights gained from this analysis inform the creation of engaging video and image posts using AI-powered media tools, ensuring that content remains relevant and impactful. This approach not only optimizes content strategies through AI-driven insights but also enhances user engagement by leveraging real-time trends. Additionally, the system offers potential applications in financial trend analysis, including the correlation of social media trends with market price movements or even the creation of digital assets based on sentiment analysis. By integrating AI and data-driven insights, this project provides an innovative solution for content creators, marketers, and businesses looking to stay at the forefront of social media dynamics.

1 Overview and Objectives

1.1 Solution Description and Purpose

The proposed solution leverages **Artificial Intelligence (AI)** to detect and exploit **trending topics** across **TikTok** and **X (Twitter)**, enabling the seamless creation of **viral content**. The system follows a multi-step approach that begins with **real-time data collection**, where **trending hashtags** and topics are extracted using **TikTok** and **X APIs**. This raw data is then processed through **AI-powered trend analysis** to identify patterns in **viral content** and predict emerging trends. Based on these insights, an **automated content creation module** generates engaging **video** and **image posts** using **AI-driven media tools**.

By combining **machine learning** and **natural language processing**, the system enhances **content relevance**, optimizes **user engagement**, and maximizes **visibility** in the ever-changing **social media ecosystem**.

The primary objective of this project is to empower **content creators**, **marketers**, and **businesses** with an **AI-driven strategy** to stay ahead of **social media trends**. In an era where **virality** is key to **digital success**, the ability to quickly identify and respond to **trending topics** can significantly boost **engagement** and **brand visibility**. By automating **trend detection** and **content creation**, this solution reduces the time and effort required for **manual analysis**, allowing users to focus on **strategy** and **creativity** rather than **data processing**. Furthermore, the system has broader applications beyond **content marketing**, such as **financial trend analysis**, where **social media sentiment** can be correlated with **market price movements** or even leveraged for creating **digital assets**. Ultimately, this project bridges the gap between **AI-driven insights** and **real-world content engagement**, enabling users to make **data-informed decisions** and achieve maximum **outreach** and **impact** in the **digital space**.

1.2 Objectives and Expected Outcomes

The objectives of this project are as follows:

- Develop an **AI-driven framework** for extracting and analyzing trending topics on social media platforms.
- Enhance engagement and visibility by leveraging **real-time trend analysis**.
- Explore applications in financial markets by analyzing **social media sentiment** and its correlation with market trends.
- Facilitate **content generation** by providing insights that guide the creation of relevant and engaging digital media.

By implementing this solution, the following outcomes are expected:

- Increased engagement rates for digital content through **AI-optimized trends**.
- Enhanced strategic decision-making with **data-driven insights** from AI trend analysis.
- Potential financial market insights derived from **social media trend correlations**.
- Improved **content generation** strategies based on AI-driven trend analysis.

2 Data Collection and Processing

2.1 X (Twitter) Data Collection

2.1.1 Overview

The data collection process involves web scraping trending tweets from X (formerly Twitter) using Selenium. The script automates the browser interactions, extracts tweet information, and saves structured data for further analysis.

2.2 Implementation Details

1. Setting Up the Environment: The script begins by installing necessary dependencies, including Selenium, Pandas, and WebDriver Manager. These libraries facilitate automated web scraping and data handling.
2. Cookie Management: A function `load_cookies()` loads cookies from a JSON file and adds them to the browser session, ensuring seamless login and bypassing authentication barriers.
3. Tweet Extraction: The function `extract_tweet_data()` processes individual tweets, extracting details such as:

- User handle
- Timestamp
- Content (cleaned from unwanted characters)
- Hashtags and mentions
- Engagement metrics: likes, retweets, replies, and views

The script employs regular expressions to convert Twitter's shorthand numerical formats (e.g., `1K`, `2.5M`) into integer values.
4. Scraping Trending Tweets: The function `fetch_trending_tweets()` launches a Chrome browser session, navigates to Twitter's search page with predefined trending keywords, and scrapes tweets iteratively by scrolling down the page until the required number of tweets is collected.
5. Data Storage: Once the tweets are collected, the script saves them in a structured CSV file using the function `save_tweets_to_csv()`. The dataset contains essential tweet information formatted for further analysis.

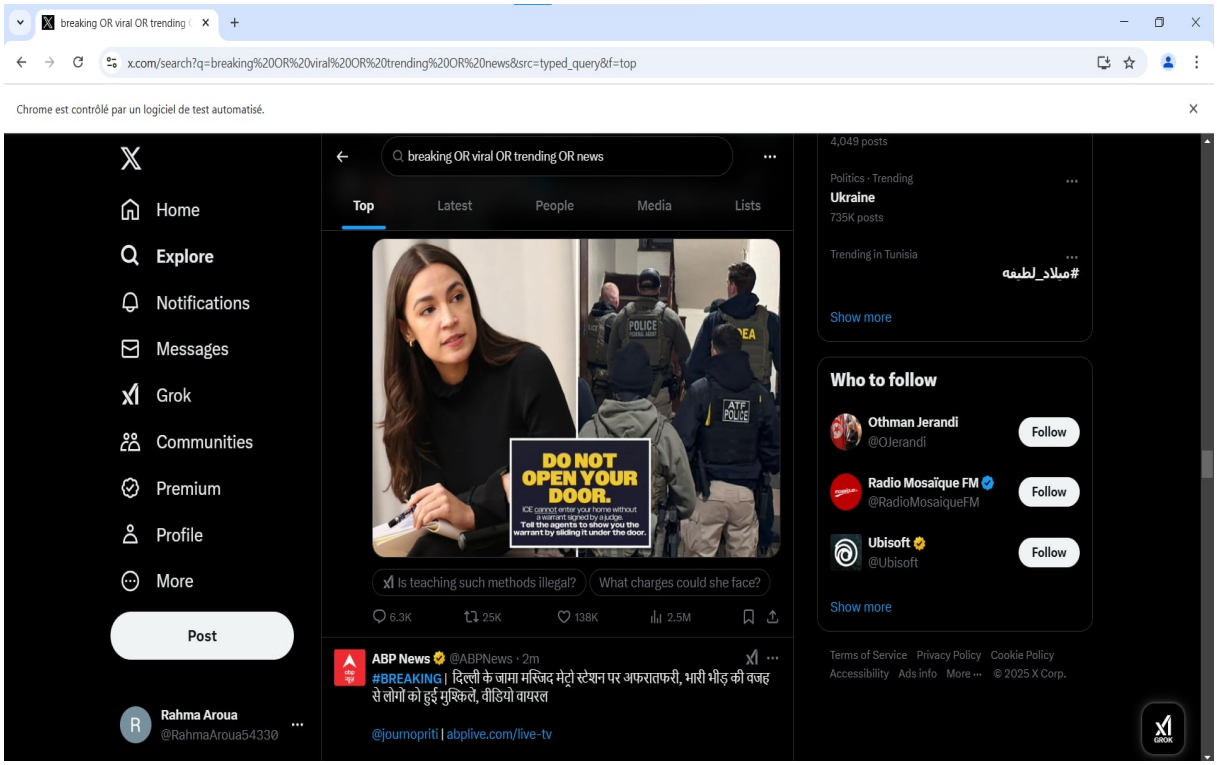
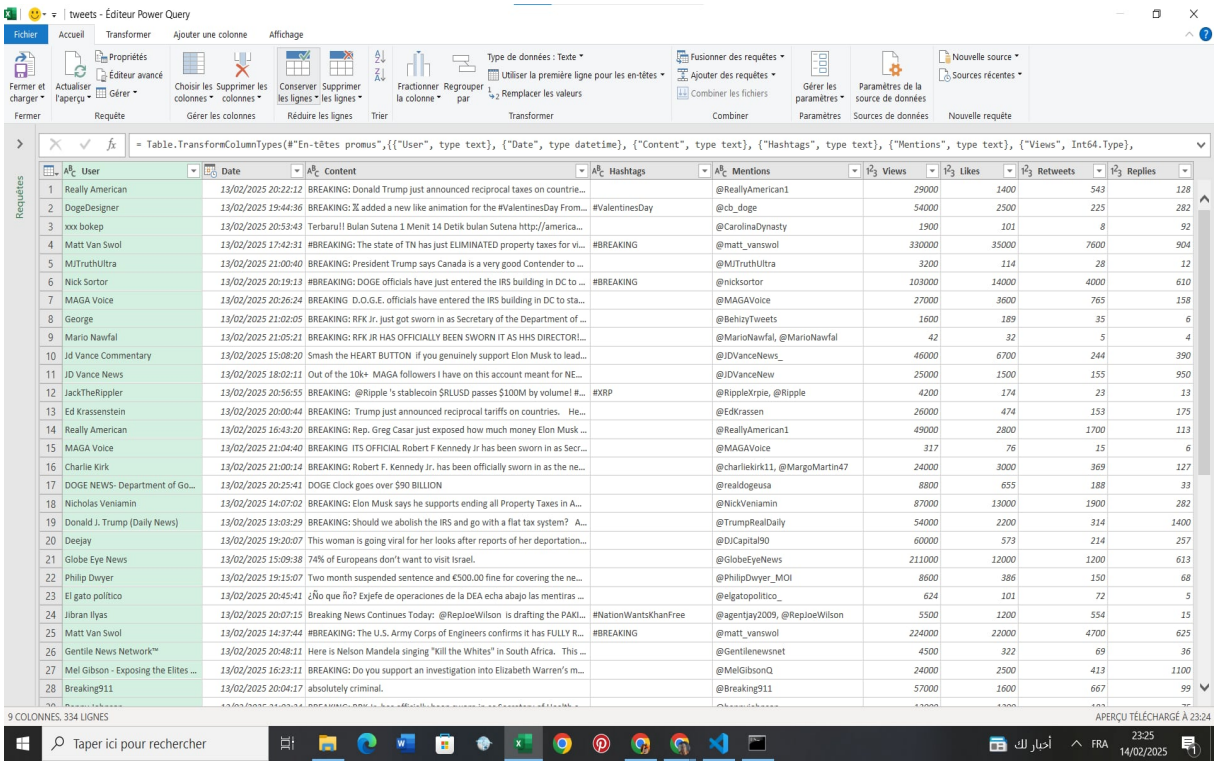


Figure 1: X scraping

2.3 Dataset



	User	Date	Content	Hashtags	Mentions	Views	Likes	Retweets	Replies
1	Really American	13/02/2025 20:22:12	BREAKING: Donald Trump just announced reciprocal taxes on countrie...		@ReallyAmerican1	29000	1400	543	128
2	DogeDesigner	13/02/2025 19:44:36	BREAKING: 𝕏 added a new like animation for the #ValentinesDay From...	#ValentinesDay	@cb_doge	54000	2500	225	282
3	xxx bokep	13/02/2025 20:53:43	Terbaru!! Bulan Sutena 1 Menit 14 Detik bulan Sutena http://america...		@CarolinaDynasty	1900	101	8	92
4	Matt Van Swol	13/02/2025 17:42:31	#BREAKING: The state of TN has just ELIMINATED property taxes for vi...	#BREAKING	@matt_vanswol	330000	35000	7600	904
5	MJTruthUltra	13/02/2025 21:00:40	BREAKING: President Trump says Canada is a very good Contender to ...		@MJTruthUltra	3200	114	28	12
6	Nick Sortor	13/02/2025 20:19:13	#BREAKING: DOGE officials have just entered the IRS building in DC to ...	#BREAKING	@nicksortor	103000	14000	4000	610
7	MAGA Voice	13/02/2025 20:26:24	BREAKING D.O.G.E. officials have entered the IRS building in DC to sta...		@MAGAVoice	27000	3600	765	158
8	George	13/02/2025 21:02:05	BREAKING: RFK Jr. just got sworn in as Secretary of the Department of ...		@BehizyTweets	1600	189	35	6
9	Mario Nawfal	13/02/2025 21:05:21	BREAKING: RFK JR HAS OFFICIALLY BEEN SWORN IT AS HHS DIRECTOR!...		@MarioNawfal, @MarioNawfal	42	32	5	4
10	Jd Vance Commentary	13/02/2025 15:08:20	Smash the HEART BUTTON if you genuinely support Elon Musk to lead...		@JDVanceNews_	46000	6700	244	390
11	JD Vance News	13/02/2025 18:02:11	Out of the 10k+ MAGA followers I have on this account meant for NE...		@JDVanceNew	25000	1500	155	950
12	JackTheRippler	13/02/2025 20:56:55	BREAKING: @Ripple 's stablecoin $RLUSD passes $100M by volume! #...	#XRP	@RippleXrpie, @Ripple	4200	174	23	13
13	Ed Krassenstein	13/02/2025 20:00:44	BREAKING: Trump just announced reciprocal tariffs on countries. He...		@EdKrassen	26000	474	153	175
14	Really American	13/02/2025 16:43:20	BREAKING: Rep. Greg Casar just exposed how much money Elon Musk ...		@ReallyAmerican1	49000	2800	1700	113
15	MAGA Voice	13/02/2025 21:04:40	BREAKING ITS OFFICIAL Robert F Kennedy Jr has been sworn in as Secr...		@MAGAVoice	317	76	15	6
16	Charlie Kirk	13/02/2025 21:00:14	BREAKING: Robert F. Kennedy Jr. has been officially sworn in as the ne...		@charliekirk11, @MargoMartin47	24000	3000	369	127
17	DOGE NEWS- Department of Go...	13/02/2025 20:25:41	DOGE Clock goes over $90 BILLION		@realdogeusa	8800	655	188	33
18	Nicholas Veniamin	13/02/2025 14:07:02	BREAKING: Elon Musk says he supports ending all Property Taxes in A...		@NickVeniamin	87000	13000	1900	282
19	Donald J. Trump (Daily News)	13/02/2025 13:03:29	BREAKING: Should we abolish the IRS and go with a flat tax system? A...		@TrumpRealDaily	54000	2200	314	1400
20	Deejay	13/02/2025 19:20:07	This woman is going viral for her looks after reports of her deportation...		@DJCapital90	60000	573	214	257
21	Globe Eye News	13/02/2025 15:09:38	74% of Europeans don't want to visit Israel.		@GlobeEyeNews	211000	12000	1200	613
22	Philip Dwyer	13/02/2025 19:15:07	Two month suspended sentence and €500.00 fine for covering the ne...		@PhilipDwyer_MOI	8600	386	150	68
23	El gato político	13/02/2025 20:45:41	¿Ño que ño? Exjefe de operaciones de la DEA echa abajo las mentiras ...		@elgatopolitico_	624	101	72	5
24	Jibran Ilyas	13/02/2025 20:07:15	Breaking News Continues Today: @RepJoeWilson is drafting the PAKI...	#NationWantsKhanFree	@agentjay2009, @RepJoeWilson	5500	1200	554	15
25	Matt Van Swol	13/02/2025 14:37:44	#BREAKING: The U.S. Army Corps of Engineers confirms it has FULLY R...	#BREAKING	@matt_vanswol	224000	22000	4700	625
26	Gentile News Network™	13/02/2025 20:48:11	Here is Nelson Mandela singing "Kill the Whites" in South Africa. This ...		@Gentilenewsnet	4500	322	69	36
27	Mel Gibson - Exposing the Elites ...	13/02/2025 16:23:11	BREAKING: Do you support an investigation into Elizabeth Warren's m...		@MelGibsonQ	24000	2500	413	1100
28	Breaking911	13/02/2025 20:04:17	absolutely criminal.		@Breaking911	57000	1600	667	99

Figure 2: X Dataset

3 TikTok Data Collection

3.1 Overview

This section describes the methodology used to collect trending TikTok videos and extract relevant metadata. The scraping process leverages extttSelenium to automate web interactions and retrieve structured information from TikTok's Explore page and individual video pages.

3.2 Scraping Methodology

The data collection is performed using the following steps:

- **Web Driver Setup:** The script initializes a `Chrome` WebDriver using `Selenium` and applies browser configurations to reduce detection.
- **Fetching Trending Videos:** The bot navigates to the TikTok Explore page (https://www.tiktok.com/explore) and dynamically scrolls to load more content. It extracts video URLs and their respective view counts.
- **Extracting Video Metadata:** The script visits each extracted video URL and retrieves key information, including:
  - Number of likes, shares, and comments.
  - Hashtags used in the video description.
  - Date of posting (either relative, e.g., "3 days ago", or absolute, e.g., "Jan 14, 2024").
- **Data Storage:** The collected data is saved in a CSV file (`tiktok_trendings.csv`) for further analysis.

3.3 Implementation Details

The scraping process includes mechanisms to mimic human interaction, such as:

- Random delays between interactions to avoid detection.
- JavaScript execution for extracting dynamically loaded elements.
- Error handling for missing elements due to TikTok's varying page structures.

The collected data can be used for trend analysis, content recommendation, and engagement pattern studies. The methodology follows ethical scraping guidelines by limiting the request frequency and avoiding unauthorized access.

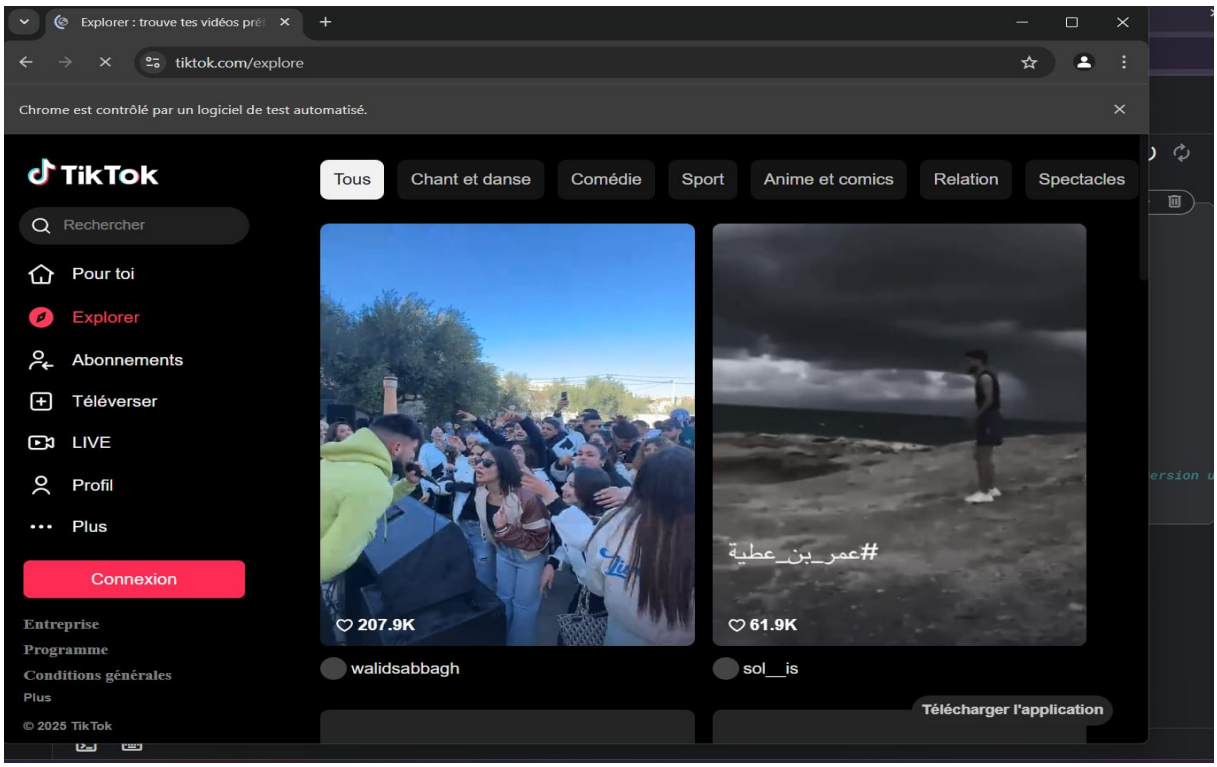


Figure 3: Tiktok scraping

3.4 Dataset

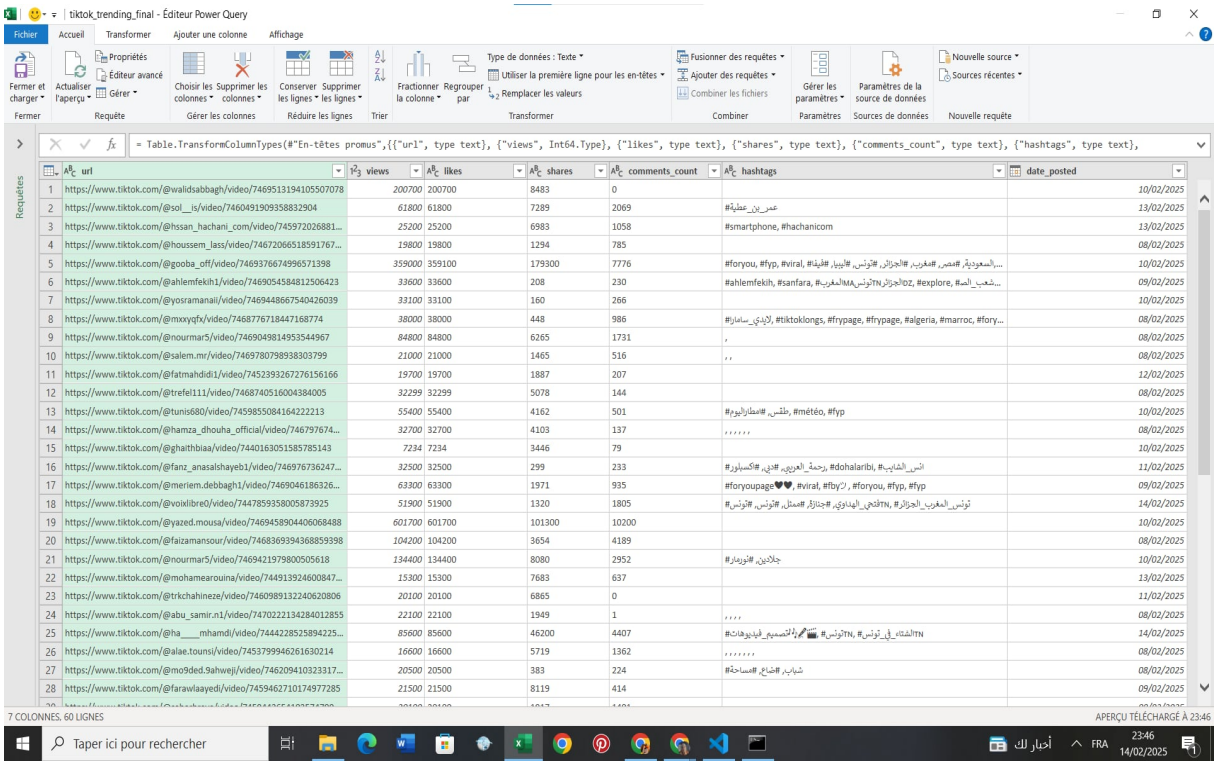


Figure 4: TikTok Dataset

3.4.1 Date Format Standardization

The `date_posted` column undergoes a comprehensive standardization process to ensure uniform date formatting in the `day-month-year` format. This process involves:

1. Converting relative dates (such as "3 days ago") to absolute dates by:
    - Identifying dates containing the "ago" keyword
    - Calculating the actual date based on the current timestamp
    - Extracting the number of days from relative date expressions
    - Transforming the calculated date into the standardized DD-MM-YYYY format
2. Processing absolute dates by:
    - Parsing the existing date string
    - Reformatting it to match the required DD-MM-YYYY standard

3.4.2 Hashtag Processing

The handling of the `hashtags` column follows a systematic approach to ensure data consistency and usability:

1. Null Value Management:
    - Identifying missing or null hashtag entries
    - Replacing null values with empty lists to maintain data structure
2. Hashtag Standardization:
    - Removing special characters and extraneous symbols
    - Converting hashtags to lowercase for consistency
    - Stripping the '' symbol from hashtag strings
    - Creating a clean, standardized list of hashtags for each entry

3.4.3 Numeric Value Conversion

The process of converting abbreviated numeric values (such as "1.5k", "2M") into their full numerical representations involves:

1. Value Recognition and Processing:
    - Identifying numeric values with suffixes (k, M, B)
    - Preserving existing integer and float values
    - Converting string representations to standardized numeric format
2. Multiplication Factor Application:
    - Thousand (k) multiplication by 1,000
    - Million (M) multiplication by 1,000,000

- Billion (B) multiplication by 1,000,000,000

3. Implementation across relevant columns:
    - Views count standardization
    - Likes count conversion
    - Shares metric normalization
    - Comments count standardization

3.5 Data Quality Assurance

The data validation process encompasses multiple layers of quality checks:

1. Format Validation:
    - Ensuring all dates follow the DD-MM-YYYY format
    - Verifying the correct structure of processed hashtag lists
    - Confirming proper numeric conversions across all relevant columns
2. Data Integrity Checks:
    - Monitoring for any remaining null values
    - Validating the consistency of numeric data types
    - Ensuring all processed values fall within expected ranges
3. Statistical Verification:
    - Analyzing the distribution of processed numeric values
    - Identifying potential outliers or anomalies
    - Verifying the consistency of temporal data

4 X Data Analysis

4.1 Most Frequent Words

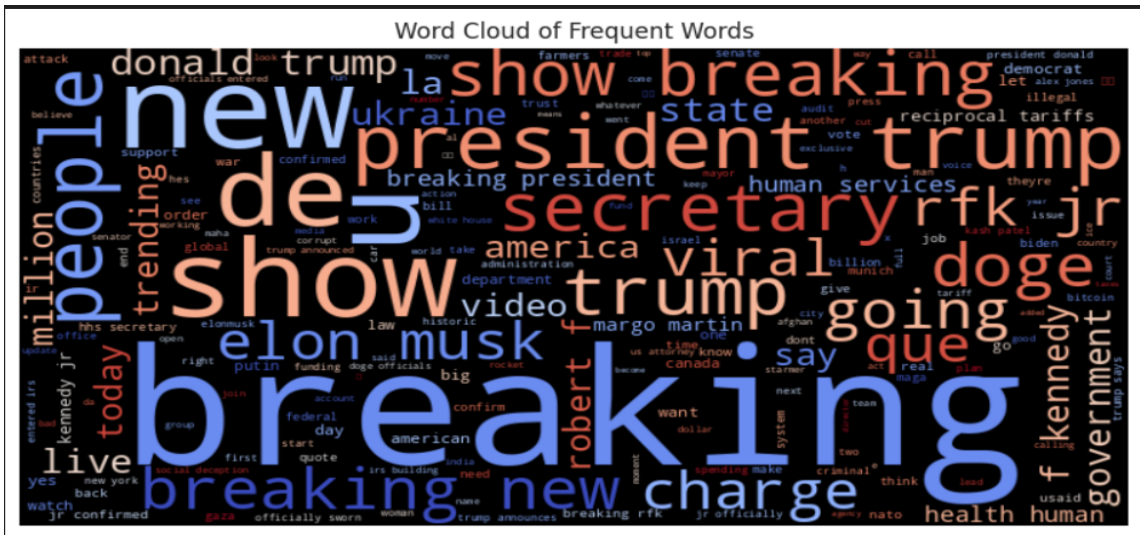


Figure 5: Most Frequent words

This visualization highlights the most commonly used words in trending tweets. Identifying these words helps understand key topics driving engagement on X. Frequently appearing words indicate prevalent discussions, enabling content strategists to align their posts accordingly.

4.2 Total Engagement by Hour

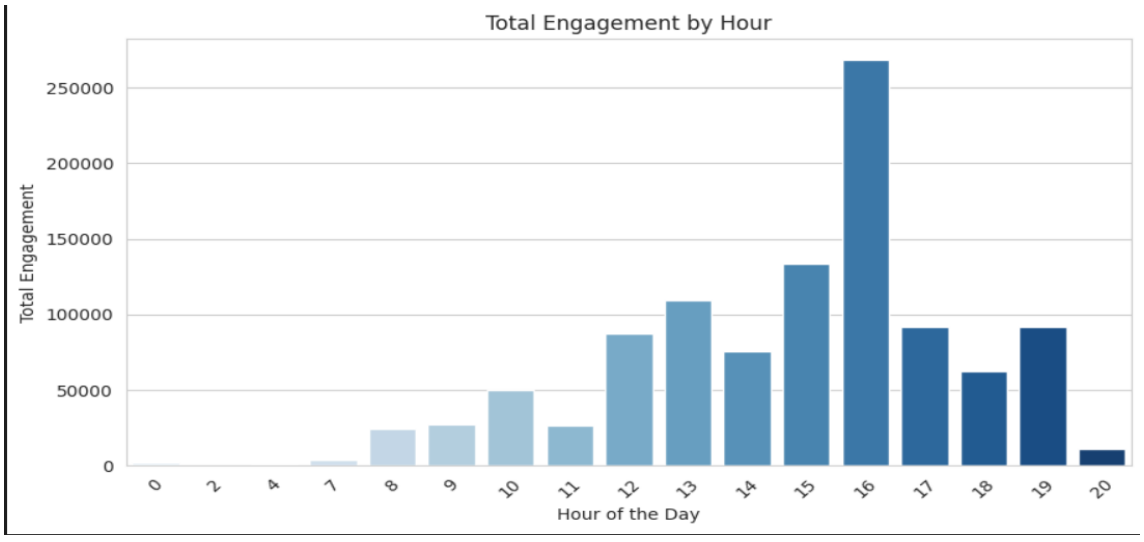


Figure 6: Total Engagement by hour

This graph shows when users are most active in engaging with tweets. Peaks in engagement suggest optimal posting times to maximize visibility. Understanding this pattern enables content creators to schedule tweets strategically.

4.3 Top Trending Hashtags

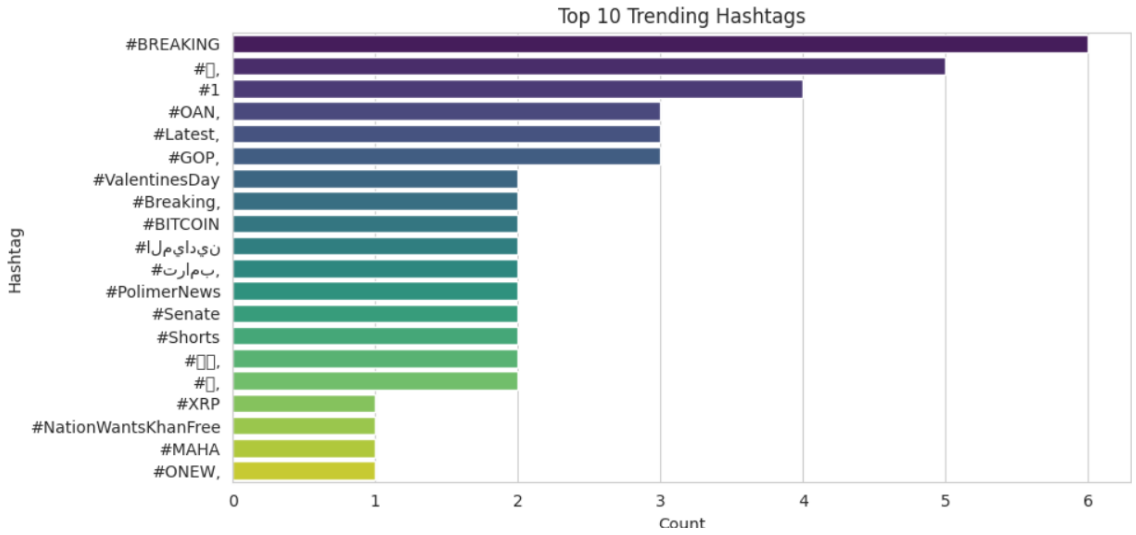


Figure 7: TOP Trending hashtags

Trending hashtags provide insights into the topics that resonate most with users. Using these hashtags in posts can significantly improve visibility and engagement.

4.4 Top Influential Users

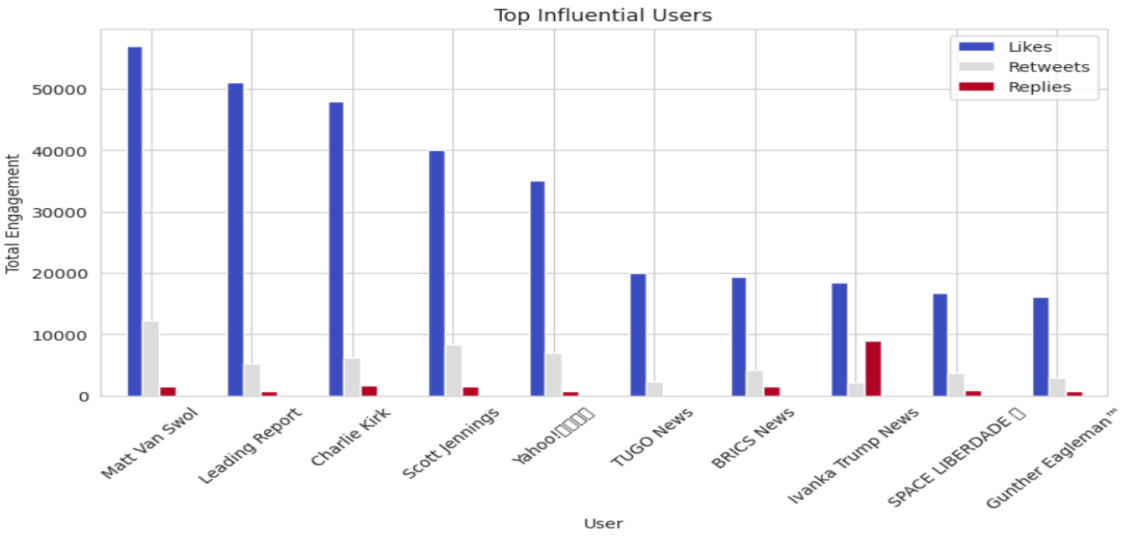


Figure 8: TOP Influential users

Identifying influential users who drive discussions helps in understanding key opinion leaders. Engaging with or collaborating with these users can boost content visibility and credibility.

4.5 Top Viral Tweets

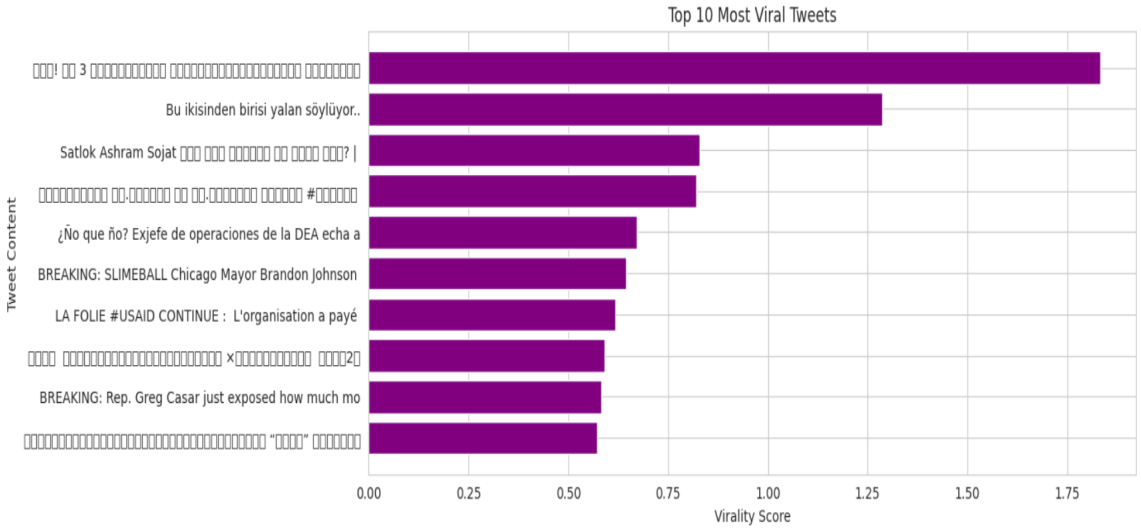


Figure 9: TOP Viral tweets

Analyzing viral tweets helps identify what kind of content resonates most with audiences. Common patterns in these tweets, such as tone, structure, or multimedia use, can be replicated for better engagement.

5 TikTok Data Analysis

5.1 Most Frequent Words in Hashtags

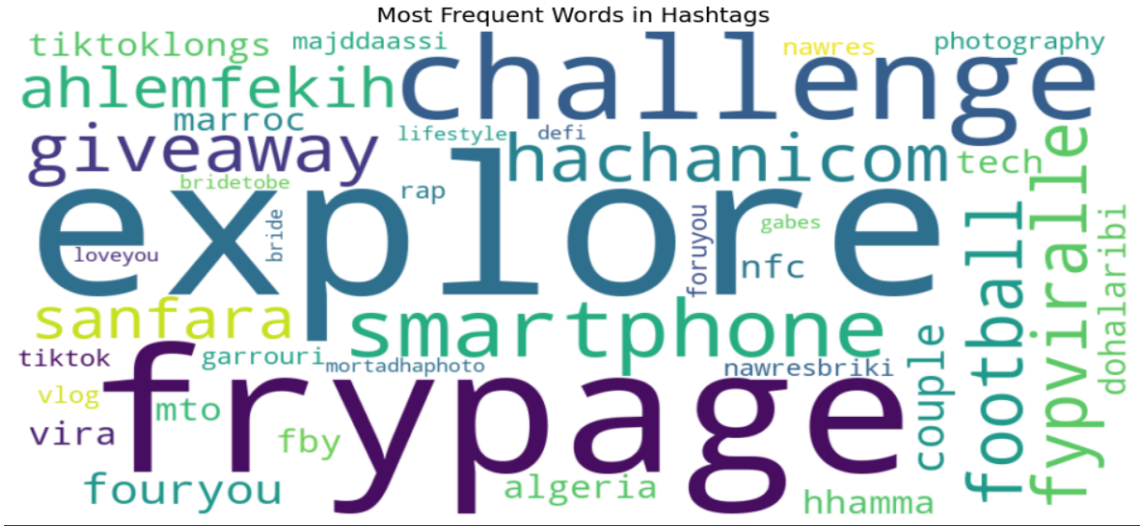


Figure 10: Most Frequent words in Hashtags

The most frequently used words in TikTok hashtags indicate popular trends. Leveraging these keywords can help boost content visibility on the platform.

5.2 Engagement Trend Over Time

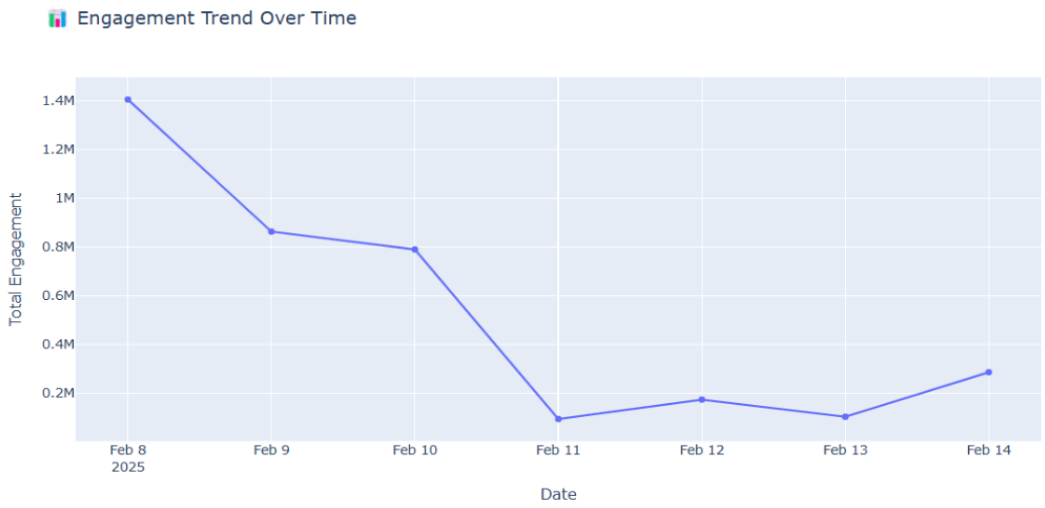


Figure 11: Engagement trend over time

This time series analysis reveals how engagement levels fluctuate over time. Spikes in engagement suggest periods of high user activity, guiding optimal content posting times.

5.3 Top Influential Hashtags

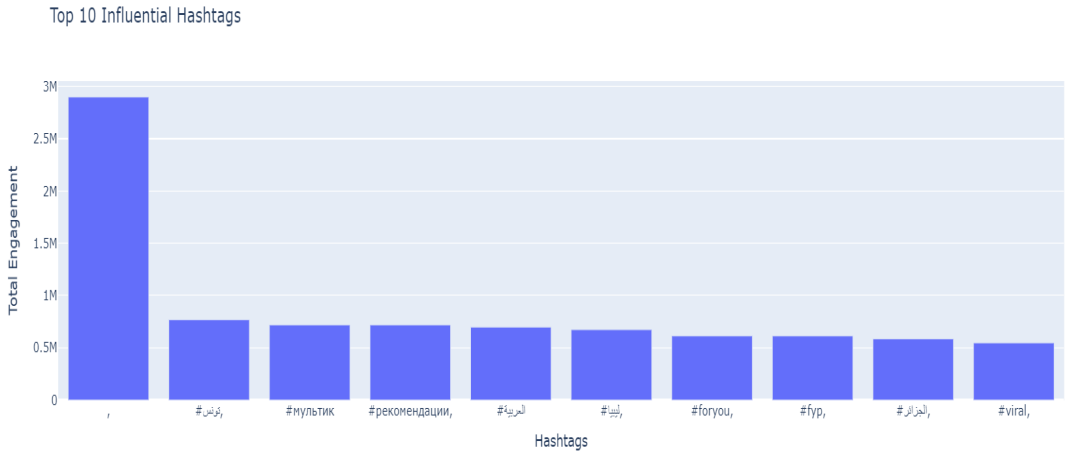


Figure 12: TOP Influential hashtags

Identifying impactful hashtags enables creators to optimize their content strategy. Using these hashtags can improve discoverability and engagement.

5.4 Top Influential Users

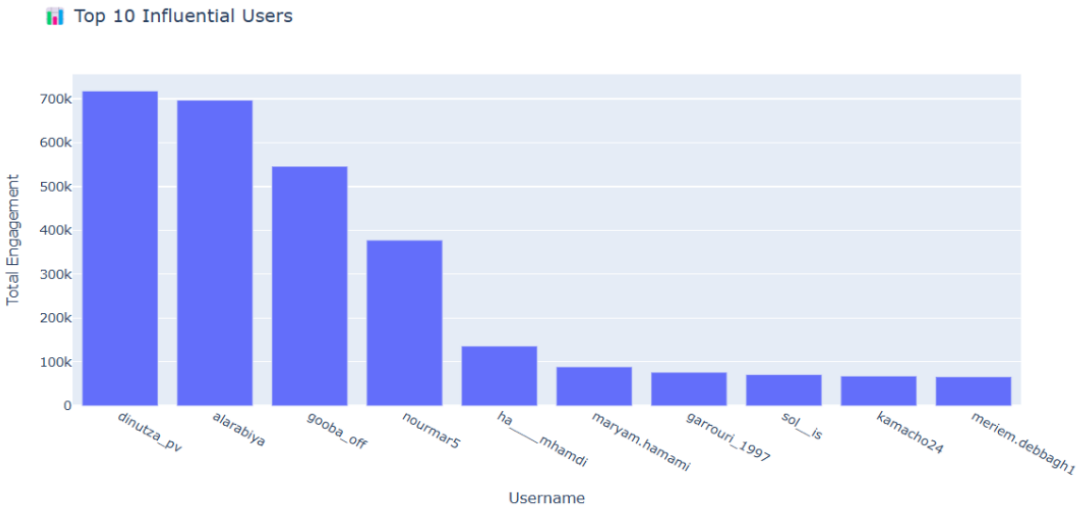


Figure 13: TOP Influential Users

This bar chart visualizes the Top 10 Influential Users based on Total Engagement (likes, retweets, replies, etc.). The usernames are displayed on the x-axis, while total engagement is on the y-axis.

5.5 Correlation Matrix

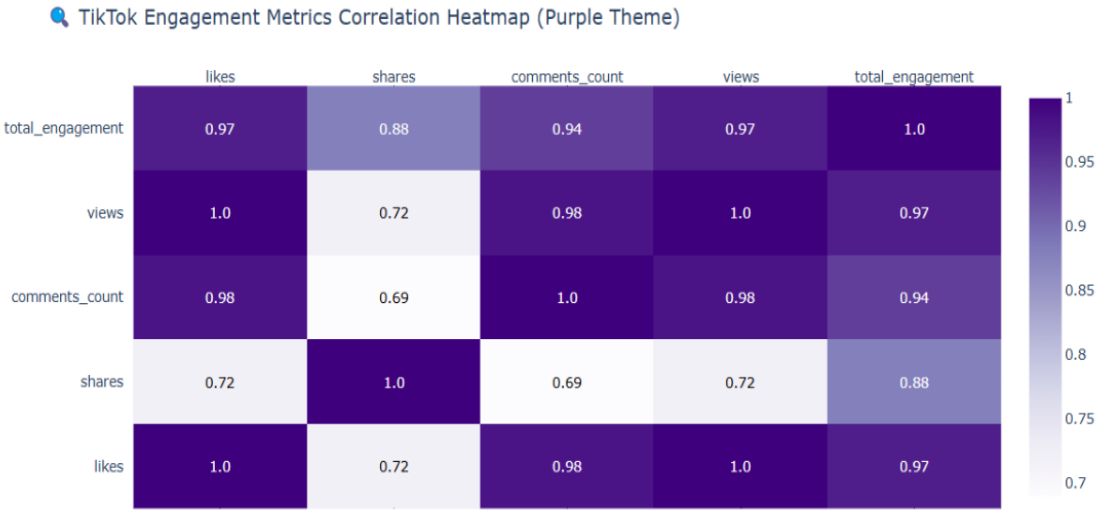


Figure 14: Correlation Matrix

This heatmap visualizes the correlation between Likes, views, and comments , they are highly correlated with total engagement (above 0.94), meaning they contribute significantly to overall engagement.

5.6 Views vs. Likes

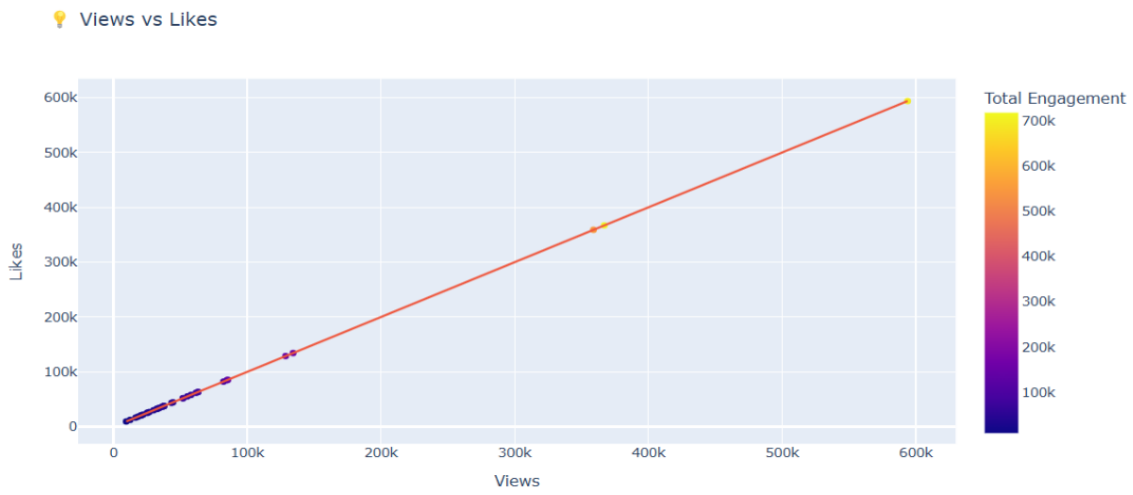


Figure 15: Views VS Likes

This scatter plot highlights the relationship between views and likes. If a high number of views leads to a proportional increase in likes, it suggests effective content that resonates with the audience.

5.7 3D Engagement Analysis

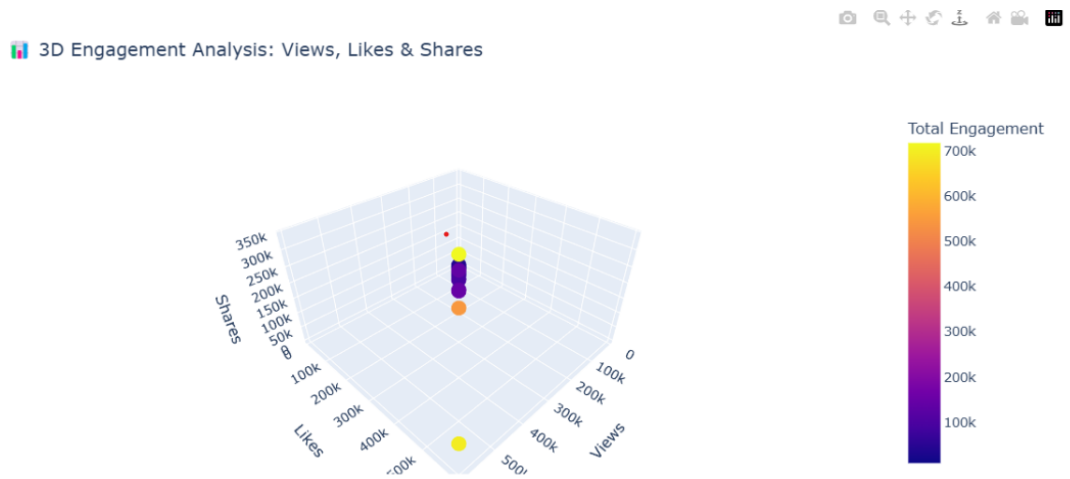


Figure 16: 3D Engagement Analysis

By analyzing engagement metrics in three dimensions, this visualization provides a holistic view of content performance. Identifying trends in multi-metric relationships can enhance content strategy.

5.8 Multi-Feature Comparison: TikTok Engagement Trends

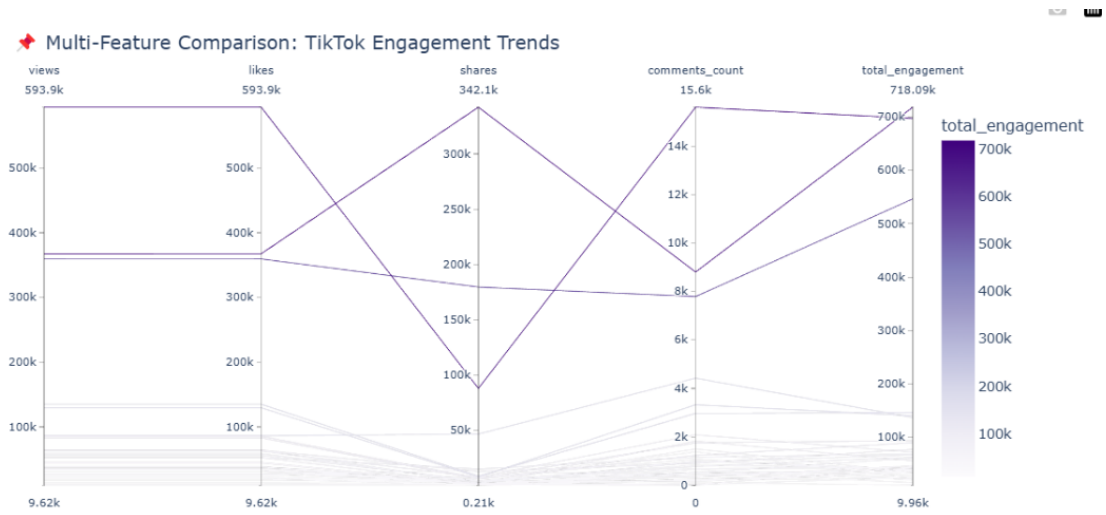


Figure 17: Multi-Feature Comparaison : TikTok engagement trends

This comparative analysis evaluates multiple engagement factors simultaneously, offering deeper insights into what drives content success on TikTok.

6 Data Generation

6.1 Text Generation

Overview

Text generation is a crucial component in automating content creation. In this project, we leverage Llama 2, a powerful AI language model, to generate textual content based on trends identified in social media data. This approach ensures that the generated text aligns with trending topics, making it more relevant and engaging.

Implementation Details

6.1.1 Model Setup

- We use **Llama 2 (7B parameter model)** for text generation.
- The model is loaded in a local environment using the **llama-cpp-python** library, allowing efficient inference.
- Dependencies are installed via **pip install**, and the model weights are downloaded from Hugging Face.

6.1.2 Initialization

- The model is loaded using the Llama class, specifying the model path.
- This ensures that the system can generate text based on input prompts efficiently.

6.1.3 Text Generation Process

- The model generates text based on input prompts extracted from trending topics on **X (formerly Twitter).**
- By fine-tuning prompts and adjusting parameters, we can control the creativity, coherence, and relevancy of the generated content.

6.1.4 Generated Output

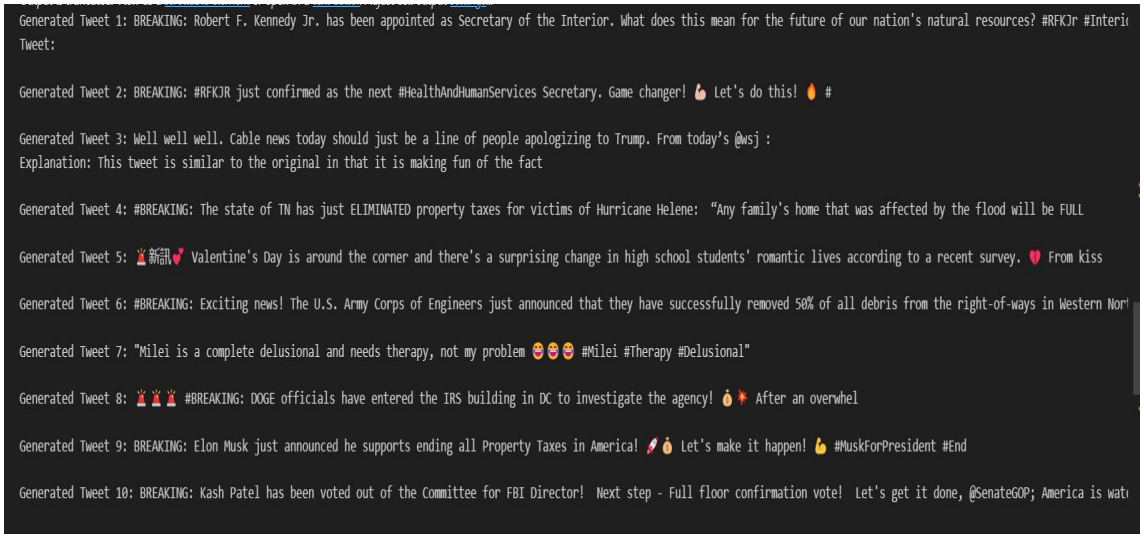


Figure 18: Generated Tweets

6.2 Image Generation

Overview

In addition to text generation, AI-driven image creation enhances content engagement by visualizing trending topics. We use Stable Diffusion, a state-of-the-art deep learning model, to generate images based on extracted social media trends. This enables automatic content generation tailored to viral discussions.

Implementation Details

6.2.1 Data Preparation

- The dataset of trending tweets is loaded, and engagement metrics (likes, retweets, replies) are analyzed.
- Trending posts are selected based on their engagement score.

6.2.2 Generating Text Prompts for Images

- For each trending topic, a descriptive text prompt is formulated (e.g., "Illustration of [trending topic] trending on social media").

6.2.3 Text Generation Process

- The **Stable Diffusion v1-4** model from Hugging Face is used.
- The model runs on a **CUDA GPU** (if available) for efficient image processing.
- Images are generated and saved as .png files for further use.

6.2.4 Generated Output

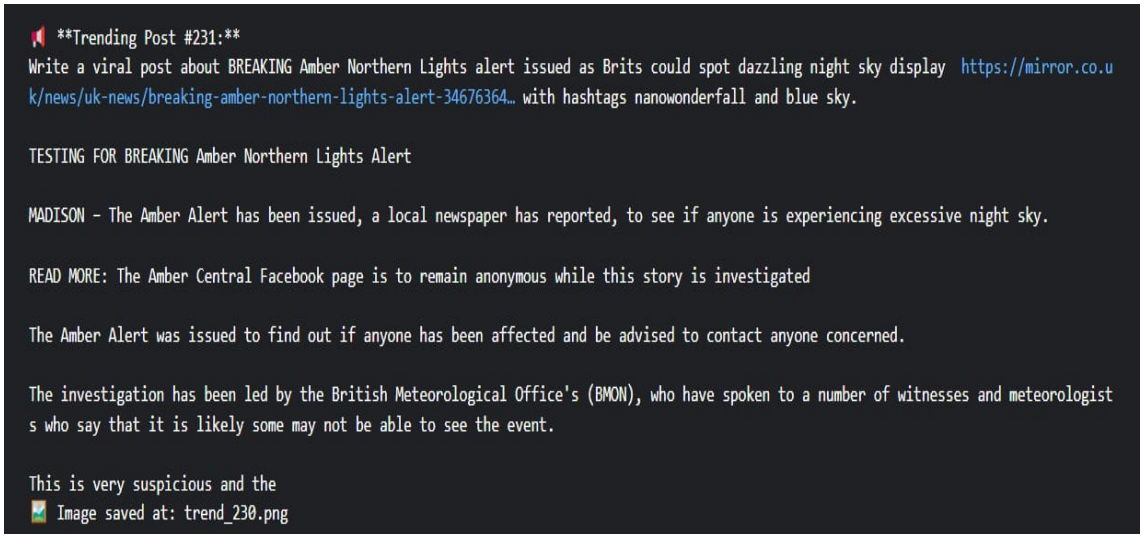


Figure 19: Generated Image caption

Figure 20: Generated Image

7 Conclusion

This project showcases the power of **AI-driven trend analysis and content generation** by leveraging **machine learning** and **automation** to extract, analyze, and create engaging digital content. Through the combination of **text generation** using Llama 2 and **image generation** via Stable Diffusion, the system effectively transforms trending social media discussions into impactful, AI-generated posts.

Key Takeaways

- **Real-time trend analysis**: Identifying viral topics using data from X (Twitter) and TikTok.
- **Automated text generation**: Producing contextually relevant posts using state-of-the-art language models.
- **AI-powered image creation**: Enhancing engagement with visuals generated from trending topics.
- **Optimized content strategies**: Leveraging AI insights to schedule posts at peak engagement times.

By integrating **AI, data automation, and deep learning**, this project provides a scalable and efficient solution for **content creators, marketers, and businesses** looking to stay ahead in the fast-paced world of social media. The methodology demonstrated here can be expanded to **financial trend analysis, digital marketing, and automated media production**, opening new opportunities for AI-enhanced content strategies.

As AI technology continues to evolve, future improvements could include **fine-tuning models for better contextual accuracy** and **incorporating multi-modal content generation (video, GIFs, etc.)** This project serves as a foundation for **next-generation AI-driven content creation**, making digital engagement smarter, faster, and more impactful.