Social Buzz Content Analysis

Today's agenda

Project Recap

Social Buzz is a fast growing technology unicorn that need to adapt quickly to it's global scale. Accenture has begun a 3 month POC focusing on these tasks:

- An audit of Social Buzz's big data practice
- Recommendation for a successful IPO
- Analysis to find Social Buzz's top 5 most Popular categories of content

Problem

Data Cleaning and Improved Data Quality: The dataset needs to undergo a thorough cleaning process to ensure data accuracy, consistency, and reliability. Enhancing the data quality will provide a solid foundation for meaningful analysis and decision-making.

Effective Data Merging: The relevant data from different sources should be merged appropriately to facilitate comprehensive analysis. The merging process should consider the analytical requirements and ensure seamless integration of data.

Comprehensive Analysis: The dataset should be analyzed across various factors, including content analysis, reaction type analysis, sentimental analysis, category analysis, descriptive analysis, and temporal analysis. Conducting these analyses will provide valuable insights into different aspects of the dataset and enable a deeper understanding of trends, patterns, and user preferences.

The Analytics team

Khushi Yadav
Data Analyst Intern

Marcus Rompton
Senior Principle

Andrew Fleming
Chief Technical Architect

Insights

Content Analysis

Top 5 Performing Categories

Category	Summed Score
animals	71242
healthy eating	69067
technology	68521
science	66549
culture	64952

Distribution of Content Types

- photo: 26.8%
- video: 25.4%
- GIF: 24.7%
- audio: 23.0%
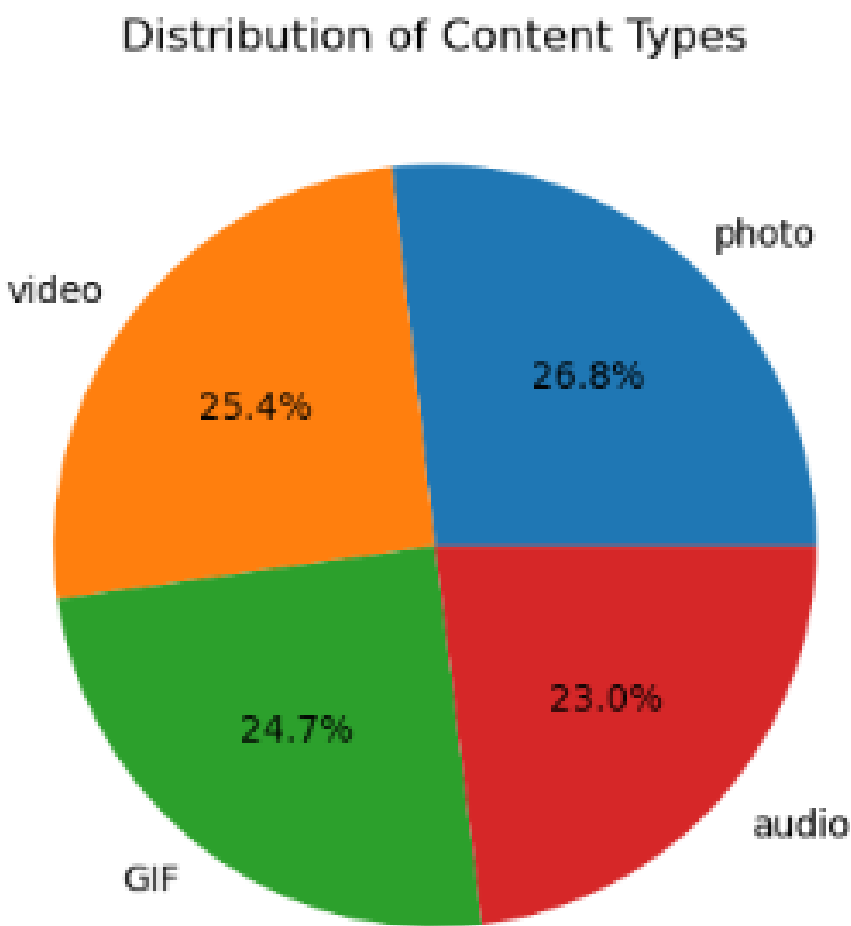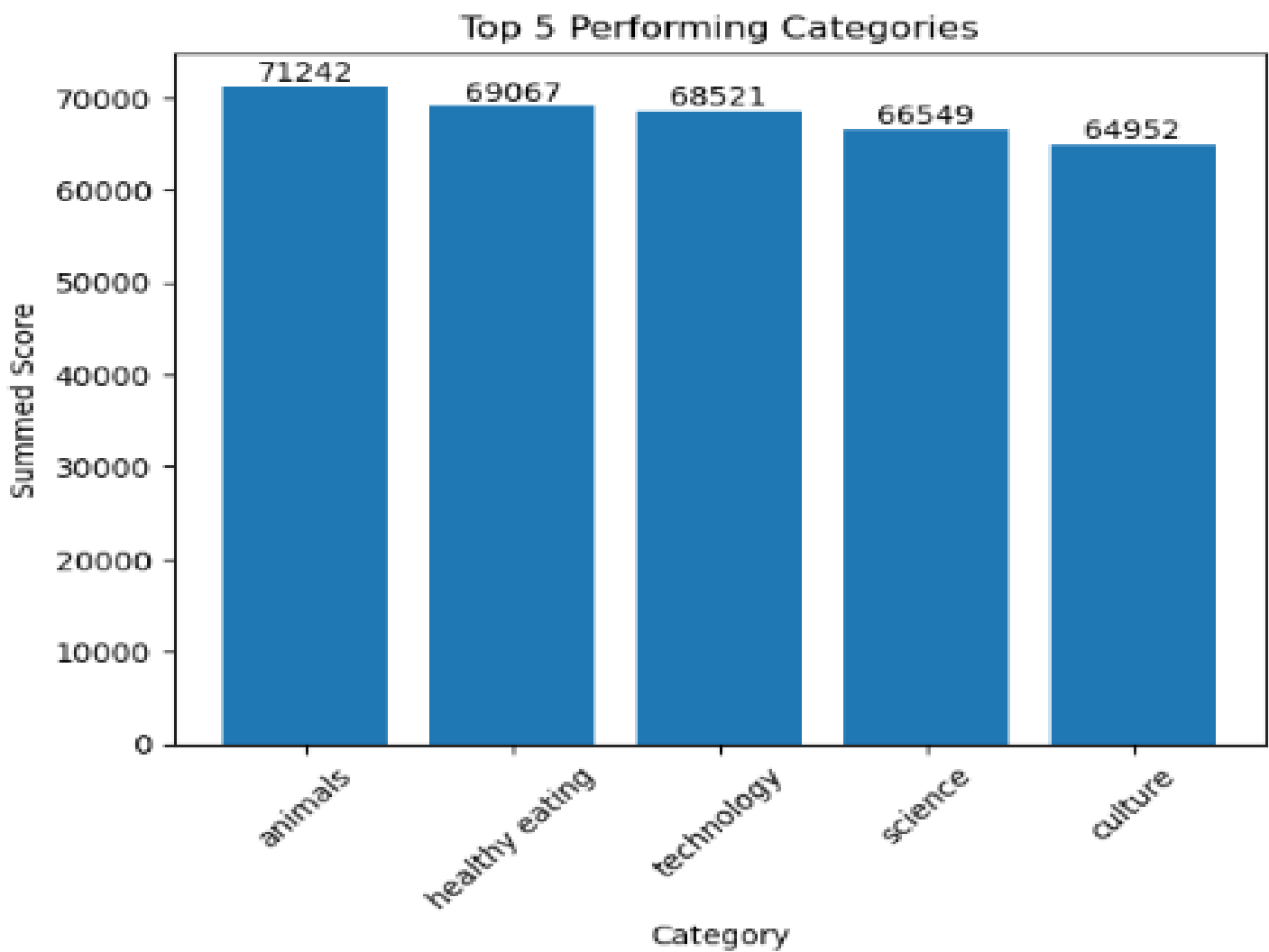
- The top score achieved in the animal category, healthy eating category, and technology category.
- The highest percentage of content is related to photos, accounting for 26.8% of the total.
- The lowest percentage of content is related to audio, accounting for 23.0% of the total.
- The content types, listed in ascending order based on their prevalence, are audio, Gif, video, and photo.
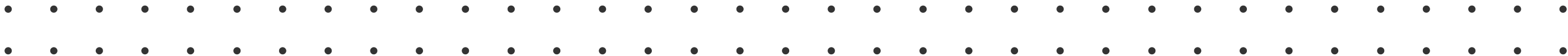
Sentimental Analysis
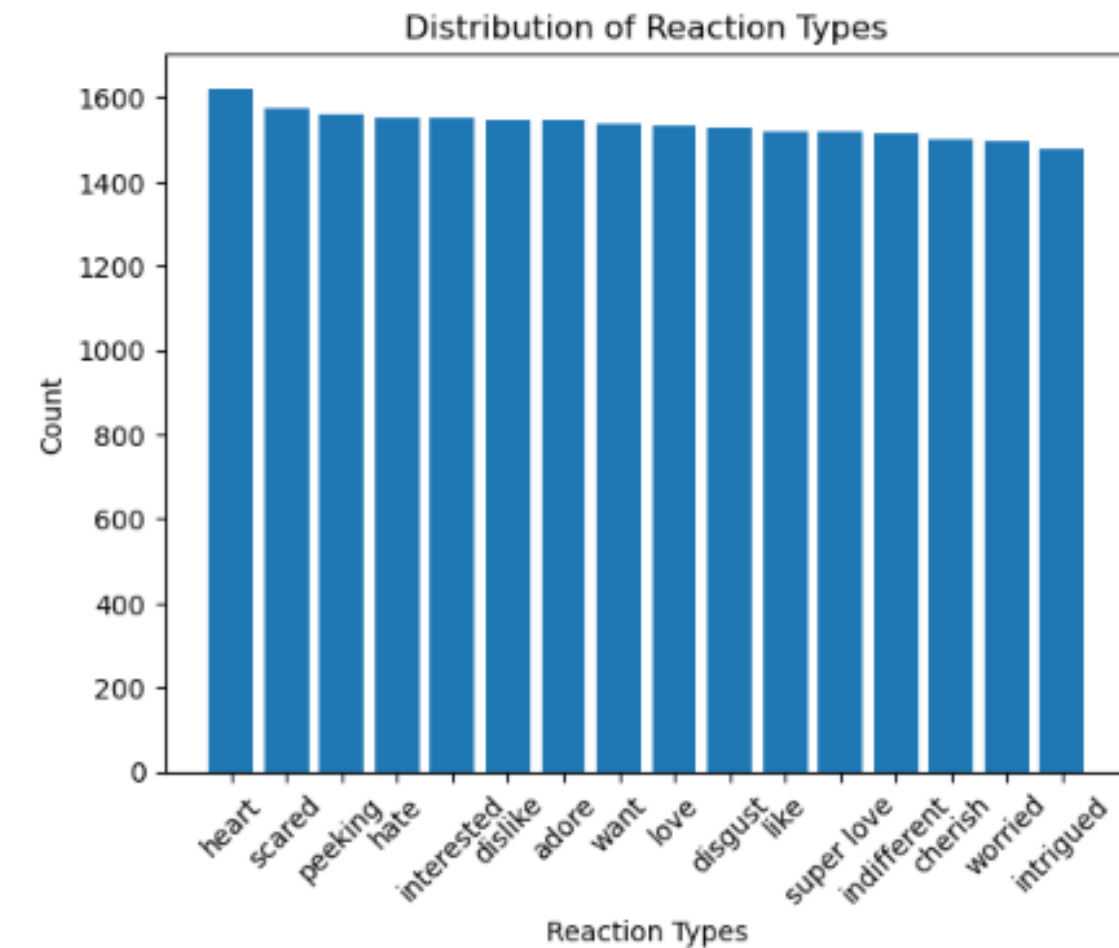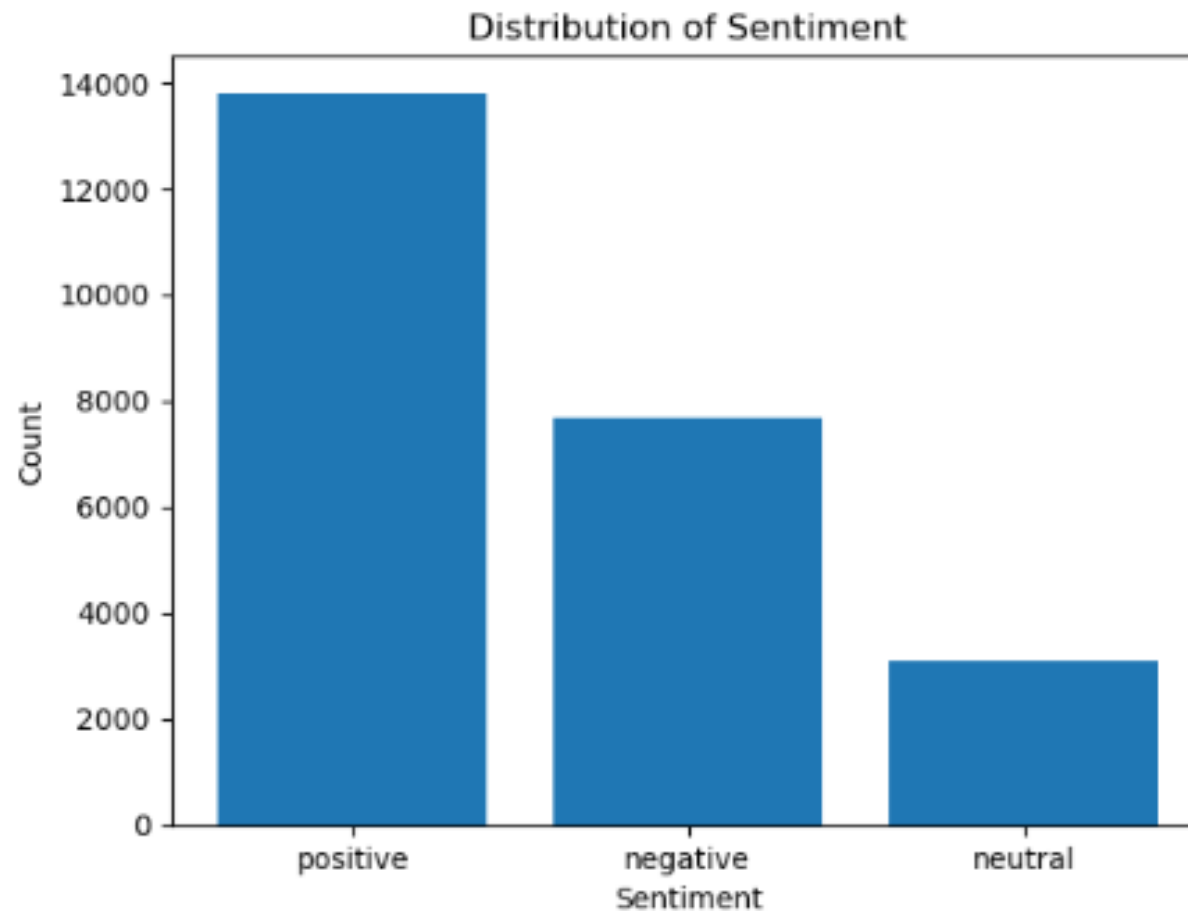
Distribution of Sentiment

Distribution of Reaction Types

- The category "Animals" has the highest mean score, indicating a positive sentiment, while "Veganism" has the lowest mean score, indicating a relatively lower sentiment.
- The top three categories with the highest mean scores are "Animals," "Culture," and "Education."
- The three categories with the lowest mean scores are "Tennis," "Travel," and "Veganism."
- Overall, the sentiment tends to be positive in different categories.
- Among different content types, GIF has the lowest mean score, while audio, photo, and video have slightly higher mean scores.
- The most commonly used reaction by people is the "Heart" reaction.
- The top three reactions based on frequency are "Heart," "Scared," and "Peeking."
- The least three reactions based on frequency are "Cherish," "Worried," and "Intrigued."
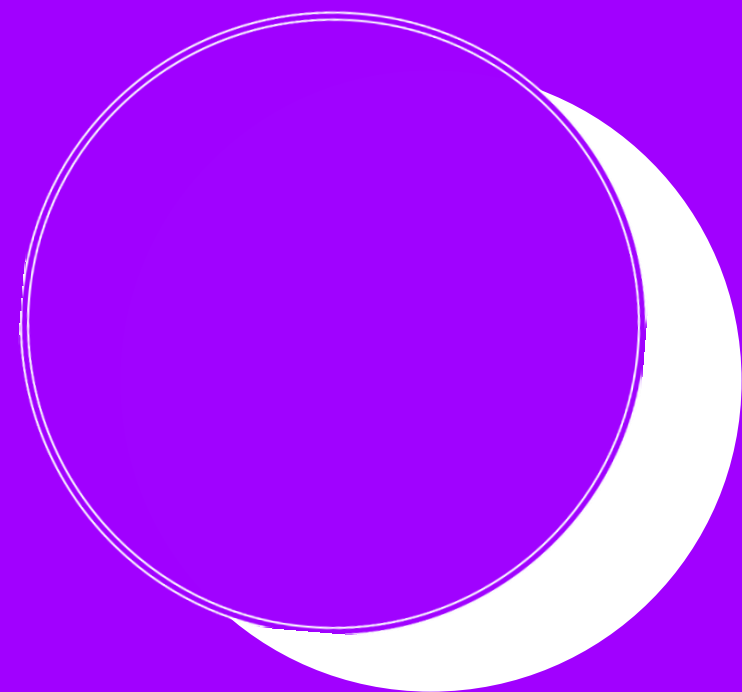
Monthly and Weekly Trend Analysis
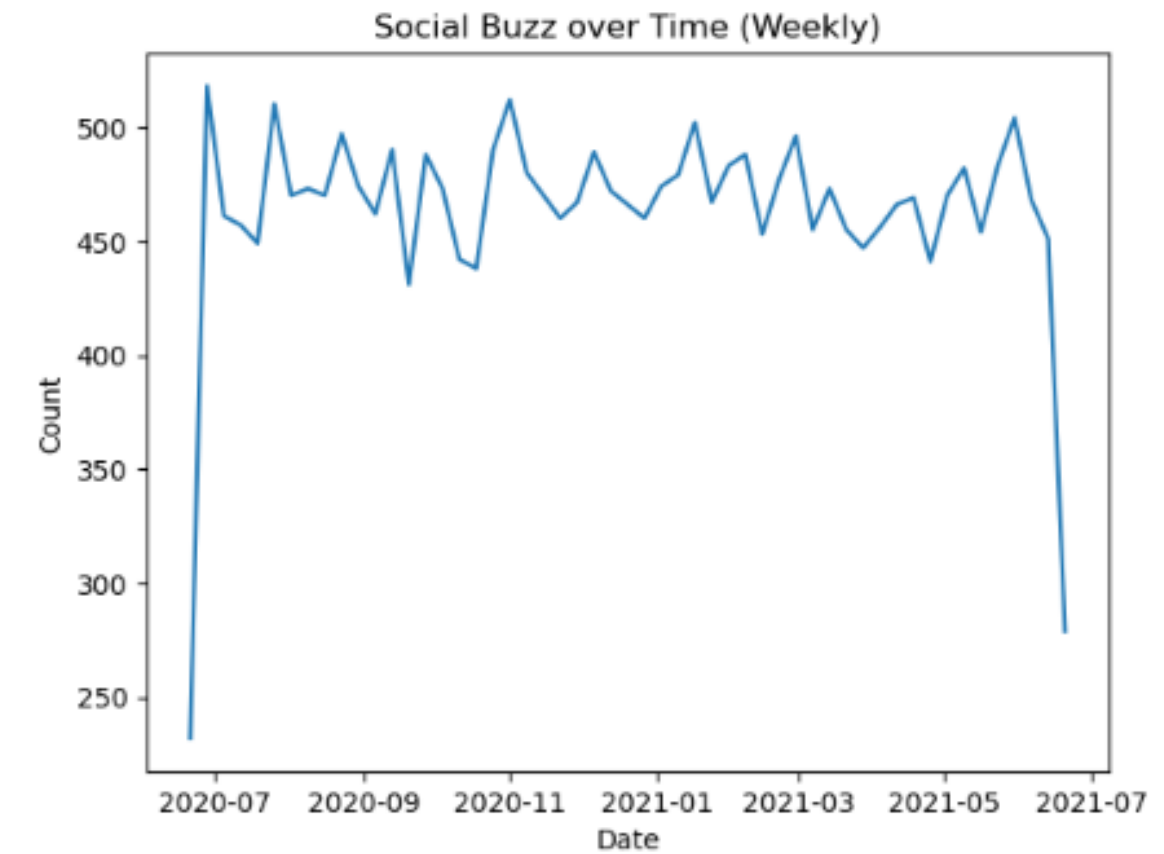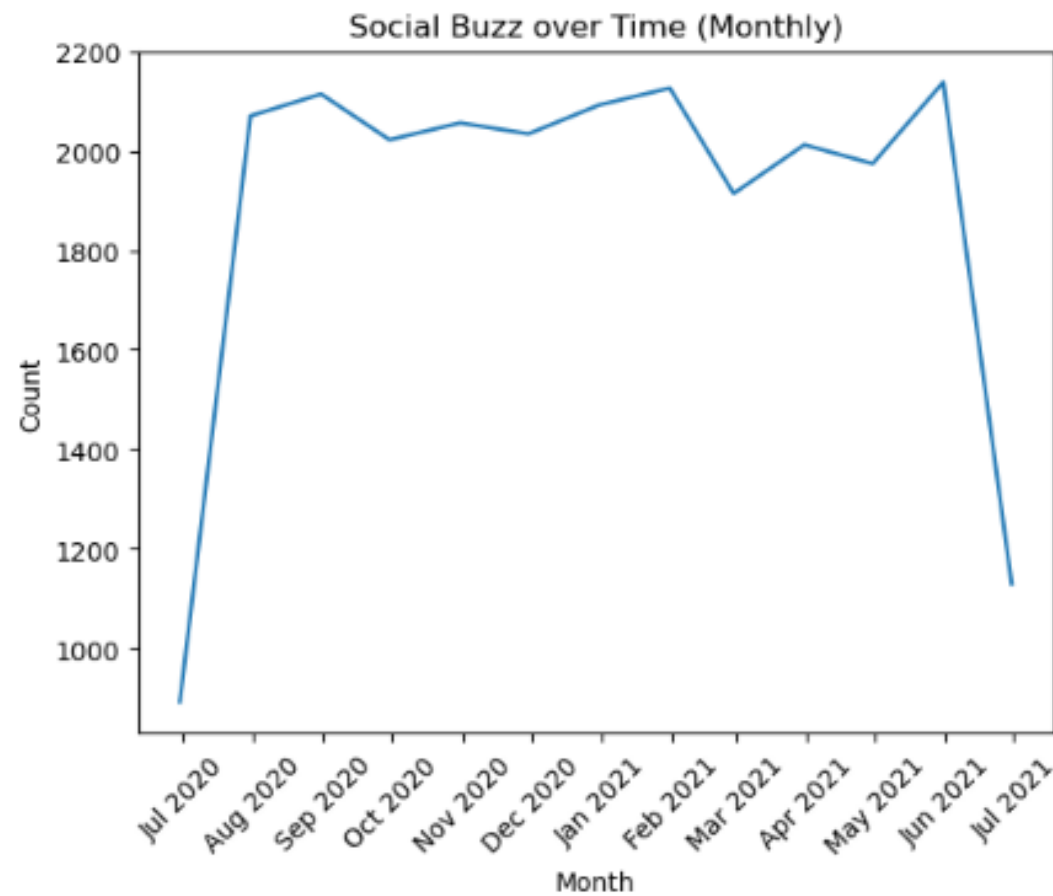
The highest count of social buzz occurred in the month of June 2021.
- The lowest count of social buzz occurred in the month of March 2021.
- On a monthly basis, the count of social buzz ranged between 1600 to 2000.
- On a weekly basis, the count of social buzz varied between 430 to 500.

Summary

- The dataset was prepared by merging relevant columns from the Reaction, Content, and Reaction Types datasets using pandas merge operation.
- The top 5 performing categories based on the summed scores were identified using pandas' nlargest function.
- The dataset contains a comprehensive collection of social media data with 24,573 unique values spanning from June 18, 2020, to June 18, 2021. This dataset provides a valuable resource for understanding trends and patterns in social media engagement.
- The content analysis reveals that photos are the most prevalent type of content, accounting for 26.8% of the total, while audio content is relatively less common, representing 23.0% of the total. The content types in ascending order of prevalence are audio, Gif, video, and photo.
- In terms of reaction types, the "Heart" reaction is the most commonly used by people, followed by "Scared" and "Peeking." On the other hand, reactions such as "Cherish," "Worried," and "Intrigued" are less frequently used.
- The temporal analysis shows that the highest count of social buzz occurred in June 2021, while the lowest count was observed in March 2021. Monthly counts ranged between 1600 and 2000, indicating a relatively consistent level of social media activity. Weekly counts varied between 430 and 500, suggesting regular engagement throughout the analyzed period.
- The sentimental analysis reveals interesting findings regarding sentiment scores across different categories. The category "Animals" has the highest mean score, indicating a positive sentiment, while "Veganism" has the lowest mean score. The top three categories with the highest mean scores are "Animals," "Culture," and "Education," while the lowest mean scores are observed in "Tennis," "Travel," and "Veganism."
- Overall, the dataset provides valuable insights into social media engagement, content types, reaction types, and sentiment. This information can be utilized to gain a deeper understanding of user preferences, popular topics, and trends within different categories. It also highlights the importance of visuals such as photos and videos in driving social media engagement.

Thank you!

ANY QUESTIONS?