Name: Karen Qiu

NetID: ktq7

Team members: N/A

Abstract

The rapid growth of dog-related content on social media platforms such as TikTok, Instagram, and YouTube significantly increased since the COVID-19 pandemic, providing viewers comfort and entertainment amid greater digital consumption and pet adoptions. This project analyzes factors influencing canine content virality, including optimal video length, popular breeds, emotional content (joy, humor, surprise, empathy, sadness), music genres, and hashtag usage. Findings will offer practical insights for content creators, marketers, pet businesses, and animal shelters, while also contributing to understanding viewer psychology and digital culture in a post-pandemic context.

1 Research Question and Motivation

1.1 Research Question

What are the common characteristics and patterns present in viral dog videos shared on social media platforms, specifically focusing on factors such as but not limited to video length, breed popularity, emotional content, music choice, hashtags, camera quality?

1.2 Context and Motivation

In recent years, especially during and after the COVID-19 pandemic, increased pet adoptions have coincided with the rise of dog-focused content on social media. Having adopted a dog myself during this period, I became actively engaged with this growing digital community, following several new creators who rapidly gained popularity through their canine companions. Observing this firsthand sparked my curiosity about factors influencing dog-video virality. Understanding these viral characteristics (optimal video length, popular breeds, emotional appeal, music choice, strategic hashtags) holds academic and personal significance; as a content creator regularly sharing dog videos, identifying these patterns would help me produce more engaging content and assist others in building successful pet-focused platforms, positively impacting society through data-backed strategies for audience engagement.

Data Sources Data identification:

The primary data sources for this project will be dog-focused video content sourced from popular social media platforms—TikTok, Instagram Reels, and YouTube Shorts—using available APIs, scraping tools, and public data archives.

- **APIs and Webscraping:** Videos will be collected through public APIs offered by platforms (e.g., YouTube Data API, TikTok unofficial APIs) and web-scraping techniques (using Python libraries such as BeautifulSoup and Selenium for automated content retrieval).
- **Video Metadata:** Key metadata such as view count, likes, comments, hashtags, and upload timestamps will be collected alongside video characteristics.
- **Cleaning and Annotation:** Videos will undergo manual and semi-automated annotation to classify dog breeds, detect emotional context (manual tagging combined with sentiment analysis tools), identify background music (using audio recognition libraries like Shazam API or Acoustic Fingerprinting libraries), and tag video content features (humorous, playful, dramatic).
- **Possible Challenges:** API rate limits, incomplete metadata, copyright restrictions on music identification, ambiguous dog breed identification, and consistency in annotating subjective emotional content.

2 Methodology and Analysis Plan

Data Exploration Techniques

EDA(Exploratory data analysis) includes statistical distribution analysis of video lengths, breed frequencies, emotional content tags, music genre distributions, and hastag popularity through visualizations(histograms, bar chars, word clouds) and description of statistics. As practice, I would be able to make this myself.

Modeling Approach

- **Regression Analysis:** identify significant relationships between video characteristics(length, breed, emotions, hashtags, etc) and virality metrics(views, shares, likes).
- **Sentiment and Emotion Analysis**: Utilizing natural language processing(NLP) for automated detection and quantification of emotional content.
- **Cluster Analysis**: K-means clustering and hierarchical clustering will identify common groups or archetypes of viral dog videos based on extracted features.
- **Audio and Music Analysis:** Leveraging audio recognition tools and libraries to classify prevalent music choices and their correlation to video popularity.

Tools and Libraries:

- **Programming Languages:**Python
- **Data Management:** Pandas, Numpy
- **Visualization:** Matplotlib, Seaborn
- **Web Scraping/API:** BeautifulSoup, Selenium, Requests
- **Audio Recognition:** Acoustic Fingerprinting Libraries(Dejavu)
- **Statistical Analysis and Modeling** Scikit-learn, Statsmodels
- **NLP and Sentiment Analysis** NLTK, TextBlob, VADER

3 Expected Outcomes and Evaluation

3.1 Anticipated Findings

I hypothesize shorter videos(10-30 seconds), emotionally humorous or joyful contexts, certain breeds(e.g. top breeds like French Bulldog, Labrador Retriever, Golden Retriever, German Shepherd, Poodle), top trending music tracks, those that have a storyline, and hashtags that of both very general to very specific(trending hashtags at the time, dog breed, etc).

3.1.1 Evaluation Metrics

- **Virality Score:** Views, likes, comments and shares
- **Predictive Modeling Accuracy:** Evaluating regression, classification model performance using standard metrics(R-squared, MAE, confusion matrices)
- **Cluster Stability:** Internal validity metrics like Silhouette scores for clustering analyses.

4 Broader Implications and Extensions

This research can help marketers, pet influencers, animal shelters, and general content creators optimize engagement strategies. It also offers theoretical contributions by providing empirical evidence about online user behavior and emotional engagement. Possible extensions include a comparative analysis across cultures, long-term virality forecasting, and examining other animal content. Furthermore, this project could extend into exploration of societal emotional stability and coping mechanisms during periods of hardship.