

Exploratory Data Analysis on Social Media Data

Introduction

This report presents an exploratory data analysis (EDA) on a dataset related to social media posts. The dataset contains various attributes such as text, sentiment, timestamp, user, platform, hashtags, retweets, likes, country, and temporal details (year, month, day, hour). The analysis aims to uncover patterns, trends, and insights from the data.

Data Overview

The dataset is loaded from a CSV file named `sentimentdataset_1.csv`. It contains 732 entries with the following columns:

- Unnamed: 0.1, Unnamed: 0:** Index columns.
- Text:** The content of the social media post.
- Sentiment:** The sentiment of the post (Positive, Negative, Neutral).
- Timestamp:** The time when the post was made.
- User:** The user who made the post.
- Platform:** The social media platform (e.g., Twitter, Instagram, Facebook).
- Hashtags:** Hashtags used in the post.
- Retweets:** Number of retweets.
- Likes:** Number of likes.
- Country:** Country of the user.
- Year, Month, Day, Hour:** Temporal details of the post.

Data Cleaning

- Handling Missing Values:** The dataset was checked for missing values, and all missing values were filled with 0.
- Removing Duplicates:** Duplicate entries were removed to ensure data integrity.
- Dropping Unnecessary Columns:** The index columns (Unnamed: 0.1 and Unnamed: 0) were dropped as they were not needed for analysis.

Summary Statistics

Likes: The average number of likes per post is approximately 42.90, with a minimum of 10 and a maximum of 80.

Retweets: The average number of retweets per post is approximately 21.51, with a minimum of 5 and a maximum of 40.

Total Likes: The cumulative likes across all posts are 31,404.

Total Retweets: The cumulative retweets across all posts are 15,744.

Temporal Analysis

Engagement by Year: The dataset spans from 2010 to 2023. The highest engagement (likes and retweets) was observed in 2023.

Engagement by Month: Engagement peaks in February and June, with the lowest engagement in December.

User Analysis

Top Users by Likes: The top users based on the total number of likes include WinterWarmth, CosmosExplorer, and CarnivalDreamer.

Top Users by Retweets: The top users based on the total number of retweets include WinterWarmth, CosmosExplorer, and CarnivalDreamer.

Post Analysis

Top Posts by Likes: The top posts by likes include content related to joy, nature, and celebrations.

Top Posts by Retweets: The top posts by retweets include content related to dance, sunshine, and books.

Visualization

Engagement Trends Over Time: A line plot shows the trends of likes and retweets over the years, indicating a significant increase in engagement in 2023.

Top Users by Likes: A bar plot visualizes the top users based on the total number of likes.

Conclusion

The analysis reveals several key insights:

Engagement Trends: There is a noticeable increase in engagement (likes and retweets) over the years, with a significant peak in 2023.

User Influence: Certain users, such as WinterWarmth and CosmosExplorer, consistently receive high engagement, indicating their influence on the platform.

Content Popularity: Posts related to positive emotions, nature, and celebrations tend to receive higher engagement.

Recommendations

Content Strategy: Focus on creating content that resonates with positive emotions and nature, as these themes tend to attract more engagement.

User Engagement: Collaborate with influential users like WinterWarmth and CosmosExplorer to boost engagement.

Temporal Posting: Consider posting more frequently during months with higher engagement, such as February and June.