Twitter Sentiment and Engagement Analysis: April–June 2023

This project explored tweets about "Elon" and "Tesla" posted between April 1, 2023, and June 30, 2023. Understanding public opinion on these subjects, examining sentiment patterns over time, and determining how sentiment affects engagement metrics like likes and impressions were the objectives. This study offered a thorough examination of public opinion and interaction on Twitter by fusing sentiment analysis, visuals, and engagement data.

Approach

To get started, I loaded the dataset first, concentrating on the most pertinent fields, including the timestamp, tweet content, and engagement indicators like likes, replies, impressions, and bookmarks. I transformed timestamps into a format that could be used for time-based analysis after cleaning the data to eliminate missing values.

For sentiment analysis, I used `TextBlob`, a straightforward tool for measuring the polarity of text. Tweets with scores greater than 0.1 were classified as positive, those with scores less than 0.1 as negative, and all other tweets were classified as neutral. I was able to quickly and efficiently categorize sentiment using this strategy, which I subsequently utilized to produce insights.

I used Plotly to build interactive infographics that clearly conveyed the findings. These included a grouped bar chart to compare engagement metrics (likes, replies, etc.) across sentiment categories, a line chart displaying sentiment trends over time, and a bar chart for sentiment distribution. In order to concentrate on important content, I also made word clouds to emphasize frequently used words in both good and negative tweets, excluding phrases like "elon," "tesla," and URLs.

Findings

From the analysis, a few key insights stood out:

- **Sentiment Breakdown**: Neutral tweets made up the majority, followed by Positive tweets, with Negative tweets being the smallest group. This suggests that while some users express strong opinions, most tweets are either factual or neutral in tone.
- **Trends Over Time**: Positive and neutral tweets showed similar trends over the three-month period, with noticeable spikes on certain days. Negative sentiment, while less common, also had specific peaks, likely tied to events or controversies.

- **Engagement by Sentiment**: Positive tweets generally had higher engagement metrics (likes and replies) than Neutral or Negative tweets. However, "Impressions" dominated across all sentiment categories, so I normalized the metrics to make the smaller ones more visible in the charts.

Challenges

Like any data project, there were a few challenges along the way:

1. **Imbalanced Metrics**: Impressions were significantly larger than likes, replies, or bookmarks. This disparity made it difficult to visualize all metrics together, so I normalized the data to scale everything down proportionally.
2. **Sentiment Classification**: While TextBlob worked well for a first pass, it struggled with nuances like sarcasm or subtle language. In future iterations, tools like VADER or transformer-based models (e.g., BERT) could offer better accuracy.

This experiment demonstrated how people interact with Elon and Tesla on Twitter and provided insightful information about public opinion regarding these subjects. I obtained practical experience in handling typical issues like noise and outliers, working with Plotly visualizations, and cleaning and analyzing real-world data. The results provided a thorough picture of how people view these subjects, and the insights were simple to understand thanks to the visualizations.

Next Steps

To build on this work, there are several opportunities for improvement and deeper exploration:

- Using advanced tools like VADER or transformer-based models to improve sentiment classification.
- Exploring the relationship between sentiment trends and specific events or announcements.
- Analyzing hashtags, mentions, and other tweet metadata to uncover additional patterns and insights.

Deliverables

This project is packaged with:

- **A Jupyter Notebook**: Includes all the code, outputs, and markdown explanations for a step-by-step walkthrough of the analysis.
- **Interactive Visualizations**: Charts created with Plotly that can be explored to gain deeper insights.
- **Processed Dataset**: A clean version of the dataset with sentiment scores and categories added.
- **GitHub Repository**: All resources, including the notebook and visualizations, are organized in a repository for easy access.

Working on this fascinating and enlightening project gave me a good grasp of the sentiment trends and interaction patterns for "Elon" and "Tesla." The results were both understandable and significant due to the integration of data analysis and visualization. With more time, the investigation might become even more in-depth by utilizing sophisticated models like BERT.