Sentiment analysis of tweets about debut sports during the Tokyo 2020 Olympics

Corinne Medeiros
Project 3 Proposal
Summer 2021
https://corinnemedeiros.github.io/

Domain

The domain this data is coming from is the sports industry, specifically the Olympics, with a focus on social media. Over 20 new sports have been added to the Tokyo Olympics schedule, including BMX Freestyle, surfing, climbing, skateboarding, and more. Fans on Twitter express their feelings towards these new events, providing important insight for the International Olympic Committee and sports sponsors through sentiment analysis.

Annotated References:

Adwan, O., Al-Tawil, M., Huneiti, A., Shahin, R., Zayed, A. & Al-Dibsi, R. (2020). Twitter sentiment analysis approaches: A survey. International Journal of Emerging Technologies in Learning, 15, 79–93. https://doi-org.ezproxy.bellevue.edu/10.3991/ijet.v15i15.14467

This journal article goes over different approaches and future research for analyzing sentiment in tweets, which helps bring awareness to different methods to consider for this project.

Constantin Răzvan, B., Bogdan, B., Laurenţiu, D., Cătălin, P., Daniel, P. & Paula, A. (2021). The role of social media on sponsorship activation. Studia Universitatis Babes-Bolyai, Educatio Artis Gymnasticae, 66(1), 111–126.

Authors explain how sponsorship in the sports industry works and how social media affects it. This helps me understand more detail on the business use cases for this project.

Day, M. (2021, July 23). The New Olympic Events Debuting in Tokyo for 2021. NBC Chicago. Retrieved from https://www.nbcchicago.com/news/sports/tokyo-summer-olympics/tokyo-olympics-new-sports-and-events-debuting-in-2021/2563173/

This is an overview of the new sports included in the Tokyo 2020 Olympics, which will give me options of topics to search within the data.

Kennedy, M., Fadel, L. & Goldman, T. (2021). Olympic opening ceremony is a delicate mix of celebration and solemnity. NPR. Retrieved from https://www.npr.org/sections/tokyo-olympics-live-updates/2021/07/23/1019622003/tokyo-olympics-opening-ceremony

This news article reports on the opening ceremony of the Olympics in Tokyo, helping provide background to the controversy and people's feelings towards the games from the start.

Kortemeier, T. (2021, July 31). What you need to know as BMX Freestyle makes its Olympic debut. Retrieved from https://www.teamusa.org/News/2021/July/31/What-You-Need-To-Know-As-BMX-Freestyle-Makes-Its-Olympic-Debut

This article describes the historical background and current world of BMX Freestyle and will help me understand the context of the sport.

Morrissey, M., Wasser, L., Diaz, J. & Palomino, J. (2020, September 11). Analyze word frequency counts using Twitter data and Tweepy in Python. Earth Lab. Retrieved from https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/calculate-tweet-word-frequencies-in-python/

This lesson provides helpful examples of tweets analysis, specifically text clean up and calculating word frequencies.

Preda, G. (2021). Tokyo Olympics 2020 Tweets. Kaggle. Retrieved from https://www.kaggle.com/gpreda/tokyo-olympics-2020-tweets

This is the dataset and a description of the attributes and collection methods.

Rivenburgh, N. K. (2002). The Olympic games: Twenty-first century challenges as a global media event. Culture, Sport, Society, 5(3), 31. https://doi-org.ezproxy.bellevue.edu/10.1080/911094208

The author discusses behavior and challenges surrounding media events, giving useful insight on societal impact during the Olympics.

Seilsepour, A., Ravanmehr, R. & Sima, H. (2019). 2016 Olympic games on Twitter: Sentiment analysis of sports fans tweets using big data framework. Journal of Advances in Computer Engineering and Technology, 5(3), 143–160.

This journal article demonstrates sentiment analysis during the 2016 Olympics, which is a good example to learn from for this project. It compares data processing tools to other keyword-based methods.

Zimmerman, M. & Huang, G. (2021, August 1). Tokyo Olympics: Tone shifts in Japan as medals outshine Covid-19 concerns. Stuff Limited. Retrieved from https://www.stuff.co.nz/sport/olympics/300371436/tokyo-olympics-tone-shifts-in-japan-as-medals-outshine-covid19-concerns

This news article reports on the currently shifting sentiment towards the Olympics in Tokyo, helping provide background to the ups and downs of locals' feelings towards the games.

Zuccarini, D. (2020, July 21). Now, more than ever, the sponsorship industry must demand data. SportsPro. Retrieved from https://www.sportspromedia.com/opinion/sports-sponsorship-analytics-strategy-liverpool-manchester-city

This article explains the importance of social media data on sports sponsorship, giving helpful information about the business benefit for this project.

Data Source

Tokyo Olympics 2020 Tweets
https://www.kaggle.com/gpreda/tokyo-olympics-2020-tweets

This dataset from Kaggle contains one csv file with over 150,000 tweets pulled from Twitter using the topic #Tokyo2020. Additional data about each tweet include username, user location, user description, hashtags, date, source, and more. A full list of attributes can be found in the dataset description in the link

above. The data are collected continuously using the Twitter API and tweepy and are merged frequently with the main dataset.

Research Questions and Benefits

Which new sports, like BMX Freestyle, are most tweeted about?

What is the sentiment of the tweets about these new sports?

Which new sports should brands focus on investing in?

Which country are the majority of these tweets about new sports coming from?

Analyzing the data this way would be beneficial for brands deciding on marketing plans. Depending on where tweets are coming from, targeted marketing can be applied. Sponsorship and brand awareness are important factors in generating more interest from consumers, and more support for the athletes and teams. For example, Nike could sponsor and focus on clothing for BMX Freestyle riders or skateboarders, depending on which sport garnered the most attention on Twitter. It would also allow for optimal scheduling within the Olympic Committee. Members could decide to include surfing on the schedule for the next Olympic games if it received positive sentiment from fans on Twitter, and they could also make sure it received a prime spot on the schedule.

Method

In this project, I'll be using Python in Jupyter Notebook to clean and process the data with the NLTK package, perform NLP tasks to prepare the text for analysis, and generate exploratory visualizations. I'll calculate word frequencies to determine which words are most common, remove stop words and collection words, and finally run polarity tests using the Textblob package. Then I'll use Tableau to create final visualizations plotting the sentiment trends.

Potential Issues

Some of the challenges I foresee in this project include text mining, bias, and unforeseen evolving sentiment from outside factors like Covid-19. Through this program, I've learned a lot about text mining and different techniques, but I still need a lot of practice. This project will be a good challenge. Sentiment could shift as more events take place, unexpected controversies unfold, and different wins occur. The dataset I'm using is updated frequently so it could be difficult to keep up with. Due to the global nature of these tweets, there could also be limited representation from non-English speakers and countries without the same access to watch certain events or comment on Twitter.

Concluding Remarks

Gaining insight on how the world is feeling about particular sports during the Covid-19 pandemic is important for understanding what to focus on in a marketing sense globally. In the sports industry, there is an expectation for results when it comes to sponsorship. Sponsors need to know they'll have a good chance on return on investment, especially with tighter budgets brought on by the pandemic. Knowing which newly introduced Olympic sports generate the most positive reactions and overall popular interest can help brands leverage current trends and feel good about where they're spending their money. Analyzing the characteristics and sentiment of these tweets will also provide a framework to analyze future data.