**Real-time Insights from Social Media Data**

**Sai Krishna Kandula**

**Department of Computer Science, Clark University, Worcester, MA**

**Abstract**

It is very important to understand the viewpoints underlying user-generated material. It seems like social media has transformed our basic human needs; these hooks are keeping us interested and hooked. Fear of missing out, curiosity, self-esteem, and speed. Twitter is a pro at this game, too. Because of the nature of its users, Twitter has such a significant impact on the world that Elon Musk's tweets keep Wall Street on its toes and Trump’s have the ability to ignite wars. Near real-time data from Twitter storms is accessible. This implies that we can pick up on the global thinking and mood waves as they emerge. So, it seems to reason that we won't pass up the chance to examine this treasure trove. We will use pre-downloaded datasets in this project to learn the ins and outs of Twitter Data. We will specifically do a thorough examination of a popular trend.

**Real-time Insights from Social Media Data**

Twitter, like any other location known to house valuables, employs security personnel who prevent us from obtaining the data as quickly as possible. Calling their APIs for data collection requires going through some authentication steps, which aren't very complicated. Because gaining knowledge on how to derive insights from data is today's primary objective, we have already been given the all-clear by security. Since our data is already formatted and organized in the datasets folder, we can now focus on the more enjoyable part of the process.

**Objectives:**

1. Regional and global patterns of thought
2. Polishing the result
3. Recognizing patterns
4. Examining the current craze
5. Going further
6. Frequency examination
7. Movement related to the trend
8. A table that can convey a thousand words
9. Investigating languages in use
10. Concluding remarks
11. Code

**Tools:** Google Colab, Scikit learn, Pandas, Matplotlib

**Results Achieved:** From making comparisons based on location to examining the activity surrounding a tweet to detecting patterns from languages and devices, artificial intelligence can do it all.

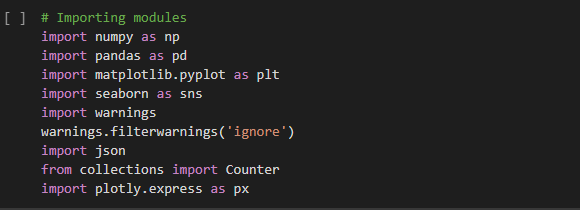
**5 Best Published papers in the same domain:**

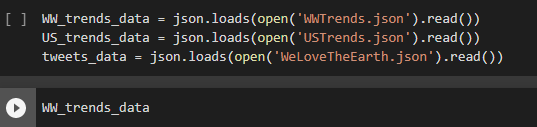
|  |
| --- |
| * **Literature Review:** |
| 1. Sulthana, A. R., Jaithunbi, A. K., & Ramesh, L. S. (2018, April 1). Sentiment analysis in twitter data using data analytic techniques for predictive modelling - IOPscience. Retrieved December 12, 2022, from <https://iopscience.iop.org/article/10.1088/1742-6596/1000/1/012130>   A sentiment analysis of tweets on Twitter is performed in order to determine which candidate is more interesting to users. The approach of linear regression was the one that we utilized in order to predict the polarity of the tweets. The performance of this data analysis method is superior to that of the support vector machine and the naive bayes method. |
| 1. Hebbar, N., & HebbarFollow, N. (2021). *Twitter Sentiment Analysis Using Python for Complete Beginners | by Nachiketa Hebbar*. <https://medium.com/swlh/tweet-sentiment-analysis-using-python-for-complete-beginners-4aeb4456040>   Classic Twitter Sentiment Analysis problem is solved using the NLTK library in Python. |
| 1. Mishra, R., & MishraFollow, R. (2021). *Sentiment Analysis of Tweets using Logistic Regression in Python | by Rishi Mishra*. <https://medium.datadriveninvestor.com/sentiment-analysis-of-tweets-using-logistic-regression-in-python-dba10a682556>   NLTK aka Natural Language Toolkit is the python library for performing Natural Language Processing (NLP) tasks. Which is to predict the sentiment of various tweets using the nltk library in python. |
| 1. Ramírez, H. (2020). Twitter JSON data processing. Cleaning and polishing a dataframe of… | by Héctor Ramírez, Ph.D. <https://towardsdatascience.com/twitter-json-data-processing-3f353a5deac4>   Twitter allows collecting tweets using tweepy, a Python library for accesing the Twitter API. Cleaning and polishing a dataframe of tweets for social media analysis, using python libraries. |
| 1. Mameli, M., Paolanti, M., Morbidoni, C., Frontoni, E., & Teti, A. (2022). Social media analytics system for action inspection on social networks. *Social Network Analysis and Mining*, *12*(1). <https://doi.org/10.1007/s13278-021-00853-w>   It is possible to forecast the results of an election by extracting and visually representing the results of opinion polls that are conducted among users of online social networks. In this article, a novel SDSS is introduced with the goal of assessing the general public's perception of a certain topic or situation. The candidates' use of social media platforms like Facebook, Instagram, and Twitter for political communication during the election cycle is the primary subject of this report. |

**Best Paper:**

From above reviewed paper 4. Twitter JSON data processing, which is cleaning and polishing a dataframe of tweets for social media analysis, using python libraries, is the best as it is a new data to work on after scrapping twitter data in JSON format and performing data cleaning and analyzing gives the best result yet allows us to work on complicated libraries to achieve the results. This is the best method because it is a new data to work on after scrapping twitter data in JSON format.

1. **Task-1: Regional and global patterns of thought (Load and inspect the data).**

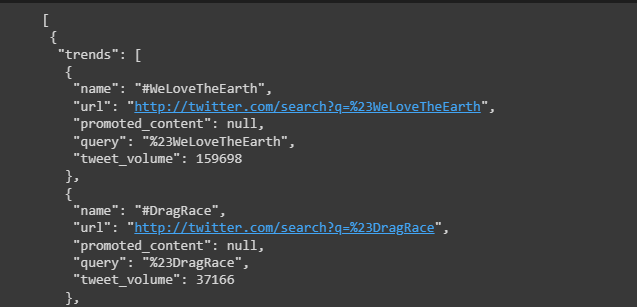


****

1. **Task-2: Polishing the result**

Our data was quite difficult to read! Fortunately, we can make use of the json.dumps() method in order to have it formatted as an attractive JSON string.

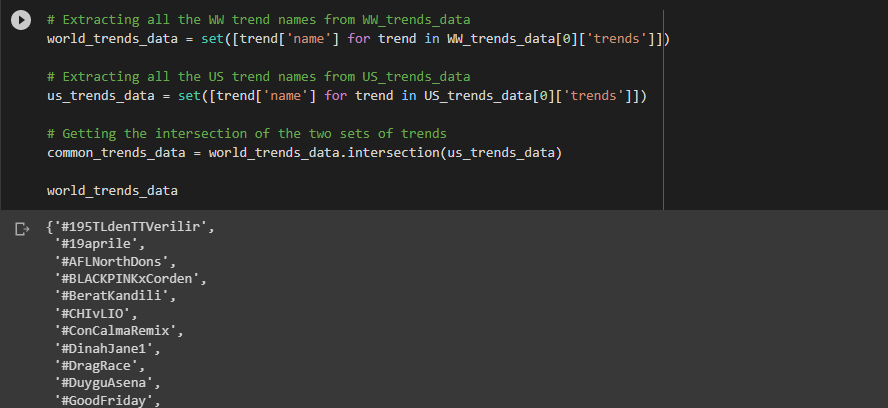
****

****

1. **Task-3: Recognizing patterns**

We can see from the attractively printed results (the output of the preceding task) that, we have an array of trend objects that includes the name of the trending topic, the Twitter-Search query parameter that can be used to look up the topic, the search URL, and, if available, the volume of tweets for the last 24 hours. The trends are updated ever five minutes.#BeratKandili, #GoodFriday, and #WeLoveTheEarth were trending on Twitter at the time of the search.

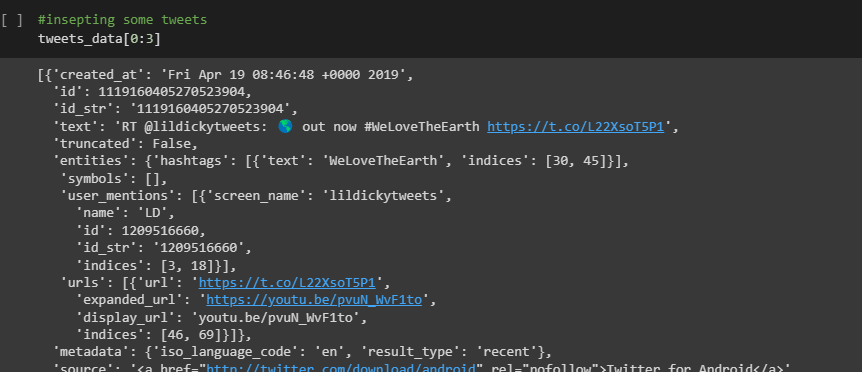
According to "tweet volume," #WeLoveTheEarth was the most well-liked hashtag out of the three.Tweet volume is not used to order the results.There are several tendencies that are specific to the United States.The two sets of trends can be quickly skimmed over to identify shared trends, but let's avoid doing "manual" effort. We may iterate through the two trends objects, turn the lists of names to sets, and then execute the intersection method to obtain the names that are shared by the two sets using Python's set data structure.



1. **Task-4: Examining the current craze**

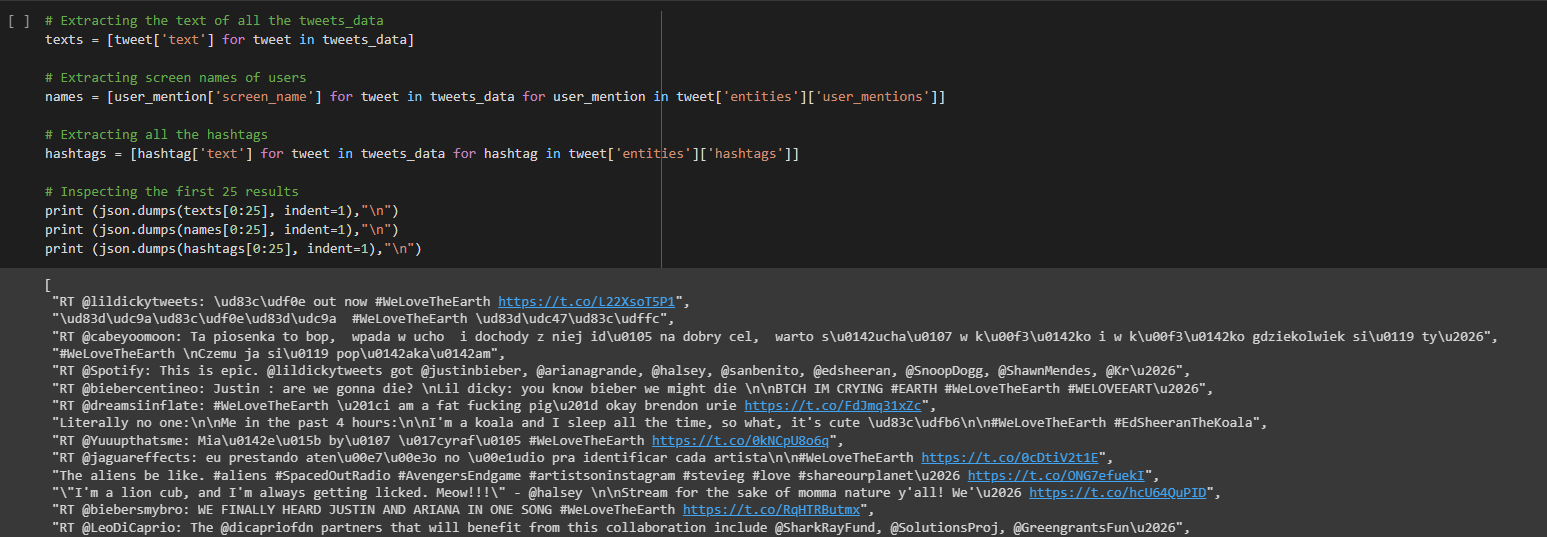
We've discovered a trend: #WeLoveTheEarth. Let's check out the narrative it is yelling at us to tell now. With this hashtag as a query parameter, we can retrieve actual tweets that are related to it through Twitter's search API. We've saved the search API's result as "WeLoveTheEarth.json" in the datasets folder. So let's load this dataset and investigate this trend in more detail.

The data is already imported above for the topic “WeLoveTheEarth” and stored in “tweets\_data”.



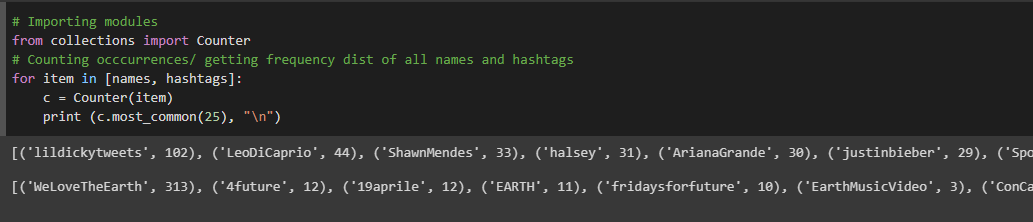
1. **Task-5: Going further**

Printing the first two tweet items helps us recognize that a tweet is much more than simply a short text message, which is how we typically think of them. But let's avoid being confused by the volume of data in a tweet object! See if we can uncover any undiscovered ideas by concentrating on a few fascinating subjects.



1. **Task-6: Frequency examination**

We can infer the following from the final extraction's initial few results alone:We're discussing a song that encourages love for the planet.This Twitter craze is being driven by many well-known musicians, including Lil Dicky. The hashtag "EdSheeranTheKoala" refers to Ed Sheeran as an adorable koala in the song.

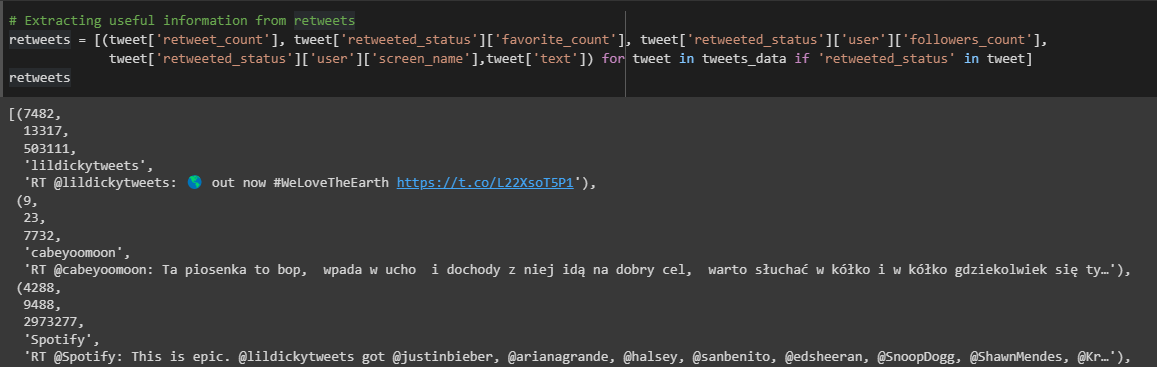


We got a sense of the data by looking at the first 10 items in the interesting fields. Now that we've completed a straightforward but crucial exercise, computing frequency distributions, we can take a closer look. In general, it's a good idea to start out simply with frequencies because it gives us ideas on how to move forward.

1. **Task-7: Movement related to the trend**

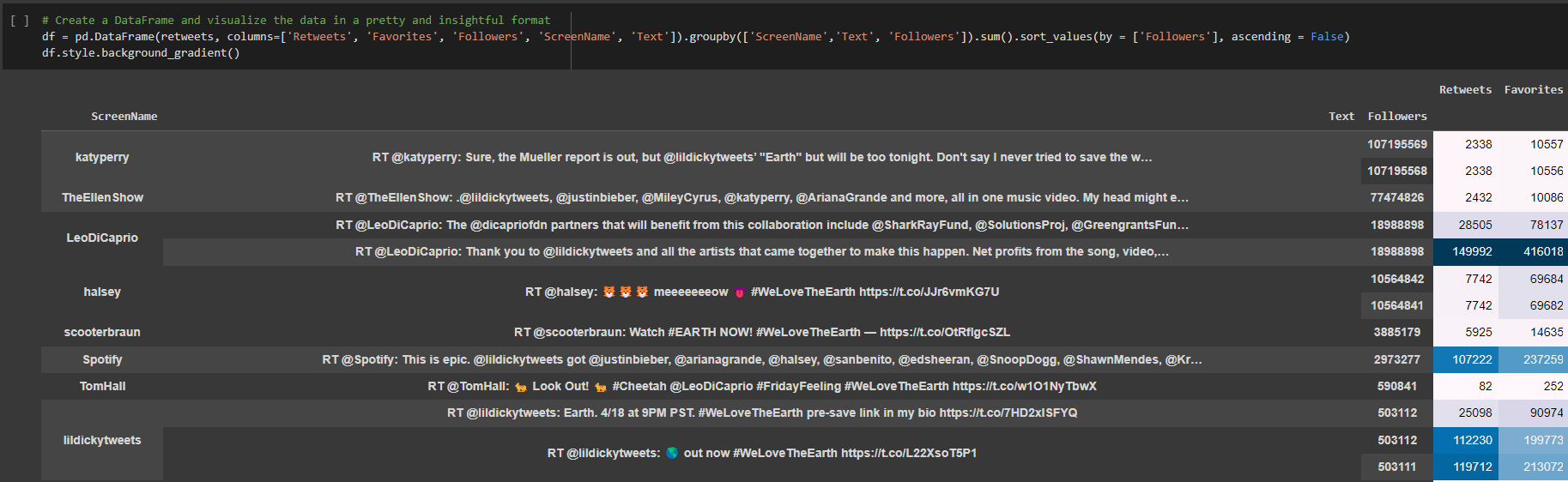
Do all retweets happen around a specific tweet? Let's further examine the data to look for trends in the activity surrounding the tweets.

The "retweeted status" column, which appears if a tweet has been retweeted, contains a wealth of information on the original tweet's content and author. By examining the retweetcount and favoritecount data, we may determine how popular a tweet is.



1. **Task-8: A table that can convey a thousand words**

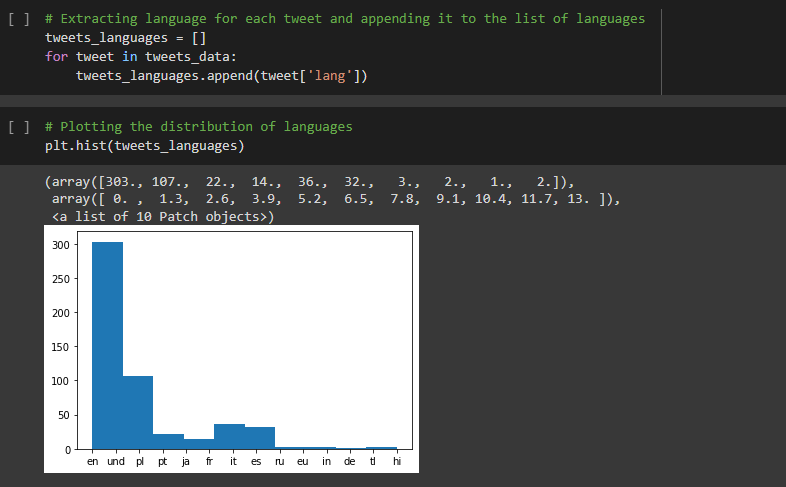
Let's make the visualization a little bit better so that the data can be shown in a tabular fashion.



1. **Task-9: Investigating languages in use**

According to our table, Lil Dicky's followers responded the most; 42.4% of them liked his first tweet. Even while celebrities like Ellen and Katy Perry have sizable Twitter followings, their fans seldom engaged with them; for instance, only 0.0098% of Katy's followers liked her tweet.

Leo received the most likes and retweets overall, but only 2.19% of his followers liked his first tweet. The fact that this was Lil Dicky's music video may have contributed to the stark contrasts in responses. Due to his significant contribution to this campaign, Leo still gained more traction than Katy or Ellen.



1. **Concluding remarks**

* Understanding the data
* The majority of tweets were in English.
* Italian, Polish, and Spanish came next.

1. **Code**

****

**Comparison with best published paper:**

In comparison to the results shown earlier in the published work, I chose to focus on Twitter data that was centered on topic discussions rather than sentiment analysis. This decision was made since it enables us to understand the "kind" of people that are interested in this issue (clustering).

**Future works to improve:**

In order to answer questions such as "Does owning an Apple as opposed to Andorid influence people's attitude towards this trend? ", we may additionally analyze the device type utilized by the Twitteratis, tweet['source']. Also work on better visualization with machine learning models.

**1.** Data collection should be more accurate.

**2.** Quality of tweets chosen should be more versatile.

**References:**

1. McKinney, W. (2017). Python for Data Analysis. In *Revision History for the Second Edition*.
2. <https://developer.twitter.com/en/docs/twitter-api/v1/trends/trends-for-location/api-reference/get-trends-place>
3. Dadhich, A., & Thankachan, B. (2021). Sentiment Analysis of Amazon Product Reviews Using Hybrid Rule-based Approach. *International Journal of Engineering and Manufacturing*, *11*(2), 40–52. <https://doi.org/10.5815/ijem.2021.02.04>
4. Furini, M., Mandreoli, F., Martoglia, R., & Montangero, M. (2022). A Predictive Method to Improve the Effectiveness of Twitter Communication in a Cultural Heritage Scenario. *Journal on Computing and Cultural Heritage*, *15*(2), 1–18. https://doi.org/10.1145/3470786