

Sentiment analysis for products launched using Twitter data

The aim is to use Twitter data to perform sentiment analysis on products that are being launched globally. This would help companies get the initial audience reception and also an insight into the features that are most talked about.

Acquiring and cleaning data:

- The source of the data is twitter, specifically the tweets with hashtags #iPhone11 and # GalaxyS10.
- Features contain the time of the tweet, country of origin, tweet text, and retweet count.
- Tweepy library was used to collect the tweets of interest after signing up for a dev account on Twitter.
- The cleaning process involved:
 - * Removing special characters from tweet text by using python's regex.
 - * With demoji library all emoji that are present in the text were cleaned up.
 - * As for country of origin which the location of the tweet, the empty values and arbitrary values like 'earth, somewhere in this planet were replaced' with the value 'unknown'. To clean up other values where only city has been mentioned or location is in native language of that county we use Google Maps API to replace it with the respective country name sin English.

Analysis and findings from the data:

- Initial EDA shows the top 5 countries of tweet origin. Interestingly India figures for iPhone as well, a trend that might be attributed to the massive campaign by resellers, mobile carries, banks and Apple themselves before the launch.

