

A Report on Twitter User Segmentation for African Influencers

Title

10 Academy July 2020 Training - Weekly Challenge: Week1

Topic

African influencers: Twitter users segmentation

Introduction

Background

Nike Inc is planning a new digital campaign in Africa, and wants to understand who the social media influencers are in Africa. And in which area they fall into (e.g. politics, fashion, art, etc.). This is key to Nike as they don't want to be associated with the wrong influencers (ethnic and religious polarisation figures for example).

Objective

This analysis is aimed at identifying top twitter influencer and most influential government official accounts on Twitter.

Top Influencers would be identified based on three different metrics:

1. Reach (Followers count)
2. Popularity (Likes & Retweets)
3. Relevance (Mentions counts & Reply counts)

The findings from this analysis would be used to recommend a certain line of actions to Nike Inc in its proposed expansion to Africa. The data would be gathered using the Twitter API.

Research Questions

1. Who are the Top Influencers & what can be learned about them.
Among the top influencers, is there a correlation between:
 - reach and relevance

- reach and popularity
 - popularity and relevance
2. What is the popular narrative(s) among these influencers (governmental & non-governmental) :
 - What are the top 5 (most used) hashtags among the influencers
 3. What is the spatial and temporal setting of these influencers
 - What profession/domain is most influential among the non-government influencers
 - What is the average age of the top influencers account?
 - Is there a correlation between account age and popularity score, reach score and relevance score

Methodology

Data Gathering Effort:

1. I started by gathering twitter handles of influential personalities and brands in Africa from the following sites:
 - [100 Most Influential African Startups](#)
 - [Africa Leader's Response to Corona-virus](#)
2. The scraped handles from these sites were then cleaned, and stored into a list and csv file.
3. I went further to use the cleaned handles to query twitter databases using the twitter API.

Some of the details gathered about each user include the following:

(The gathered data for all user was stored into a table and later csv file)

- Twitter Id
- User description
- Follower's count
- Friend's count
- Account Age
- Hashtags
- Retweets Counts

Data Wrangling Effort

Upon loading the data into the analysis notebook, I assessed the data visually and programmatically and I identified certain quality and structural issues about the data. This formed the basis of my cleaning activities on the data. Some of the issues identified include multiple index, wrong column naming, etc. I also had to concatenate the hashtag tables into one because they were loaded separately.

Analysis

Observations & Insights

The analysis was done in line with the objective of the project which is to help Nike Inc identify key influencers in Africa that are best fit for media partnership or hire for its proposed expansion to Africa.

In an attempt to answering the above research questions, the following insights were cleaned from the wrangled and analysed data:

1. There is strong positive correlation among the various measures of influence i.e popularity (retweets, count and tweets likes), reach (followers), relevance (mention counts, reply counts). An increase in one is strongly likely to increase in the other.

However emphasis was placed in the correlation between popularity and relevance this is to avoid the [million follower fallacy](#)

- Spearman's Rank Correlation Coefficient was used to calculate these data points.
2. The following are the top influential african government on twitter across various measures of influence:
 - Macky Sall
 - Muhammadu Buhari
 - Nana Akufo Addo
 - Presidency South Africa
 - President of Zimbabwe
 - Yoweri K Museveni
 3. The following are the top influential african personalities on twitter across various measures of influence:
 - Julius Sello Malema
 - Loyiso Gola
 - Trevor Noah

4. **#Covid19** is the most used hashtag among both sets of identified top influencers:

Africa Govt	2069
Africa Influencers	5301

5. There is no correlation between the age of a user account and its effect on any measure of the influence except for reach which has a weak positive correlation and this is reasonably so because there are potentials of a long existing account to attract more followers than new ones.
6. The rise to becoming influential Africans has been as a result of consistency in audience engagement over a long period of time.

7. Most of the influencers in Africa function in the entertainment industry especially.

Limitations

1. I could not gather the reply counts to effectively measure relevance scores. This is because it is a premium feature of the Twitter AP.
2. I could not gather all the needed hashtags for Research Question 2. This is both due to poor internet constraints and time constraints

Recommendations & Conclusions

The following is recommended to Nike Inc in its proposed social media marketing campaign and expansion in Africa.

- Nike Inc should consider partnership with the following influencers (who have no political affiliation):
 1. Loyiso Gola
 2. Trevor Noah
- Nike Inc Should consider framing their marketing campaign & CSR programmes around covid19 - the most popular hashtag.

References

Spearman Rank Correlation Coefficient

- <https://towardsdatascience.com/discover-the-strength-of-monotonic-relation-850d11f72046>
- <https://machinelearningmastery.com/how-to-calculate-nonparametric-rank-correlation-in-python/>