Let Twitter Talk

An analysis of tweets for major universities to identify users' profile

By: Jenny Jiang

Executive Summary

- The analysis is based on 0.88 million tweets, whose content is about selected universities (University of Chicago, Northwestern University, Boston University, and Northeastern University).

- The goal of this analysis is to distinguish the profiles of Twitterers, who are tweeting about the University of Chicago, to the profiles of Twitterers, who are tweeting about other universities, so that actionable business recommendations could be provided to the University of Chicago to improve its social media outreach programs.

- Tools for analysis: PySpark and Python.

- Recommendations for University of Chicago based on the analysis results:

    Take users with the most retweets as supplement channels to deploy the social media outreach programs; focus on the twitters in Chicago area; Reach out to twitterers on Friday and Saturday nights; Promote academically related topics; Initiate topics that can evoke heated discussion among U Chicago's twitterers

Methodology & Data Overview

Methodology

- Filter twitter source data by selected universities and by selected tweets features.

  Selected universities:

  University of Chicago   Northwestern University

  Boston University   Northeastern University

- Only the full names of universities and user-related tweets features have been used to filter the data.

- Run below analysis across all universities:
  - Who are the most prolific / influential users?
  - Where do most users locate?
  - What are the timelines of these tweets?
  - Identify distinguishing characteristics of users
  - How unique are the tweets?

Source data:

- The source data contains approximate 300 million records, most of which are related to either "Chicago" or "University".
- The data ranged from 2017 to 2021

Filtered data:

- 877,081 records together with 18 features have been filtered

Universities	# of records
University of Chicago	299,651
Northwestern University	129,090
Boston University	352,727
Northeastern University	96,366

Findings - Prolific / Influential Users

The users who tweet the most about a given university:

- do not necessarily receive the most retweets

- are usually not individual account
  - e.g., chiftf_uchicago is a twitter account that tweets about food truck information in University of Chicago

The users who receive most retweets:

- are tweeting about big events

- are not tweeting about certain university that much
 - e.g., Only 0.82% of the twitter account MarkoSilberhand's tweets are about the Boston University

Universities	Users	% of tweets about University
University of Chicago	Goal58014091	100%
Northwestern University	davidgura	4.4%
Boston University	MarkoSilberhand	0.8%
Northeastern University	DreeProvince	2.5%

Findings - Location & Timeline

Where do most users locate?

- Most users who are tweeting about these four universities are in the United States

- 6%~8% of users are local. Northeastern University has a high rate, about 18%.

- Below is the top 10 user locations for University of Chicago
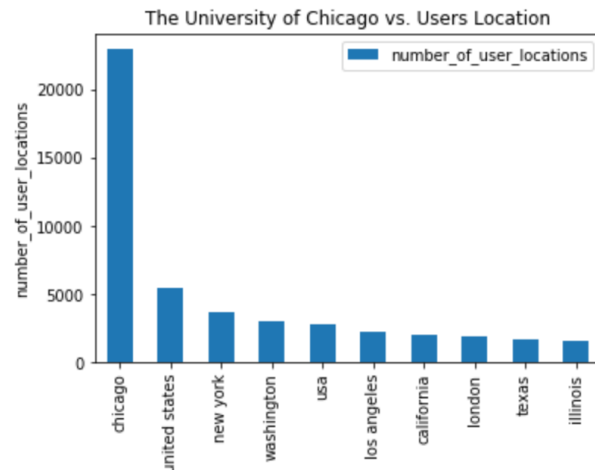
What are the timelines of these tweets?

- # of tweets are growing at steady speed from 2018 to 2020

- February, March, and September are the three months with most tweets

- In a week, users are more active on Friday and Saturday

- In a day, the peak occurs at 08:00 PM, while the valley occurs at 09:00 AM

- Two data collection gaps have been identified:
    - 2018-01-06 ~ 2018-01-31
    - 2018-04-03 ~2018-04-21

Findings - Distinguishing Characteristics of Users

What distinguish University of Chicago Twitterers to Other Universities Twitters

- University of Chicago twitterers on average have more friends and more favorites count than other universities twitterers

- Top 10 tweets receives the most retweet counts in different universities:
    - University of Chicago: more than half of them are academic related topics
    - Northwestern University: 4 out of 10 are about the news that Joseph Epstein urges Jill Biden to drop Dr. title
    - Boston University: 4 of 10 most retweeted tweets are politics topics or topics related politician. 2 topics are related to the research about CTE diagnosis
    - Northeastern University: 4 out of 10 are about academic researches. 3 topics are related the disciplines in the university

Findings - How unique are the tweets?

- MinHashLSH and threshold = 0.3 are used to determine the uniqueness of tweets

- All four universities have more duplicates tweets than unique tweets based on our criteria
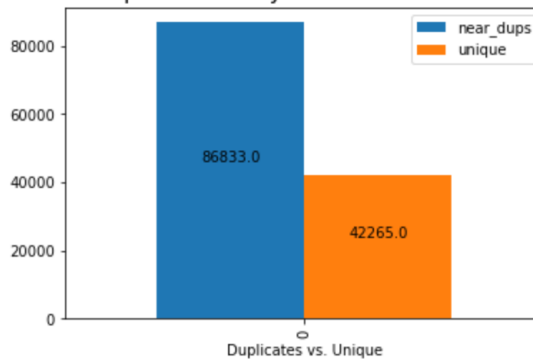
Duplicates vs. Uniqueness
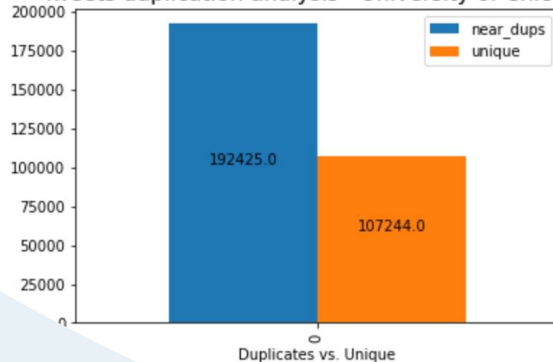
| Duplicate Ratio: | 64% | 67% | 59% | 80% |
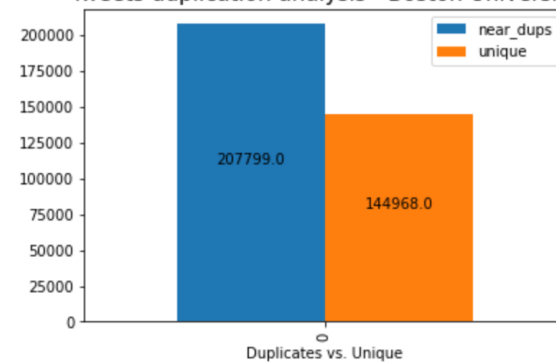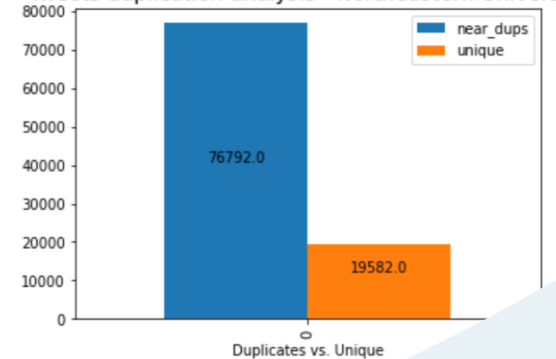
Actionable Recommendations

- Identify users with the most retweets count and take them as supplement channels to deploy the social media outreach programs

- 7.6% University of Chicago's twitterers are in local area, some offline outreach programs can be organized to get people more engaged

- It's better to reach out people on Friday and Saturday nights, since that's the time that twitterers are most active

- Promote academically related topics, since that's the interest of UChicago's twitterers

- Apart from advocating topics triggering retweets, initiate topics that can evoke heated discussion among U Chicago's twitterers

Thank you!