



THE UNIVERSITY OF
CHICAGO

BIG DATA: Course Project

**Social Media Analysis using Twitter:
University of Chicago**

Prepared for: Nick Kadocnikov

December 2018

Submitted by: Ashish Mahadik

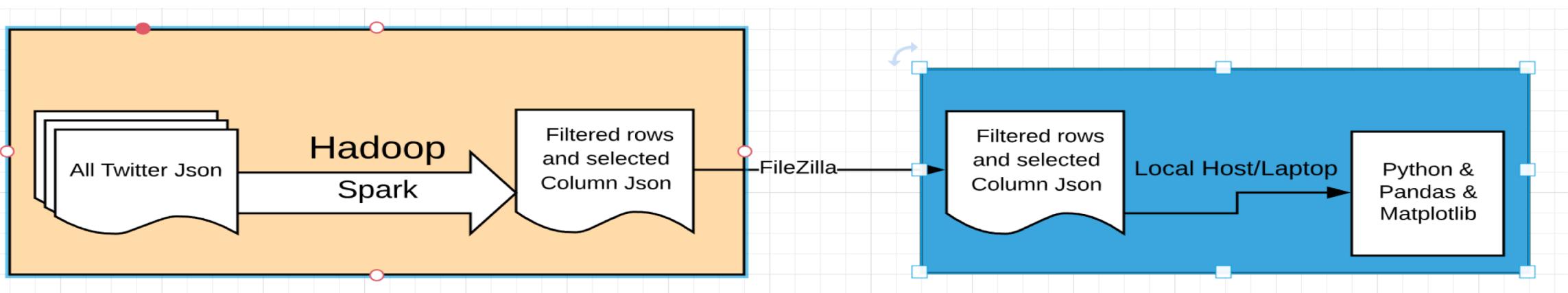
Executive Summary

Background & Objective: University of Chicago (UChicago) wants to improve its social media presence. The objective is to make recommendations through a comparative study of UChicago's Twitter presence vis-à-vis 3 other universities – Purdue university, Stanford university and Yale university.

Key Findings:

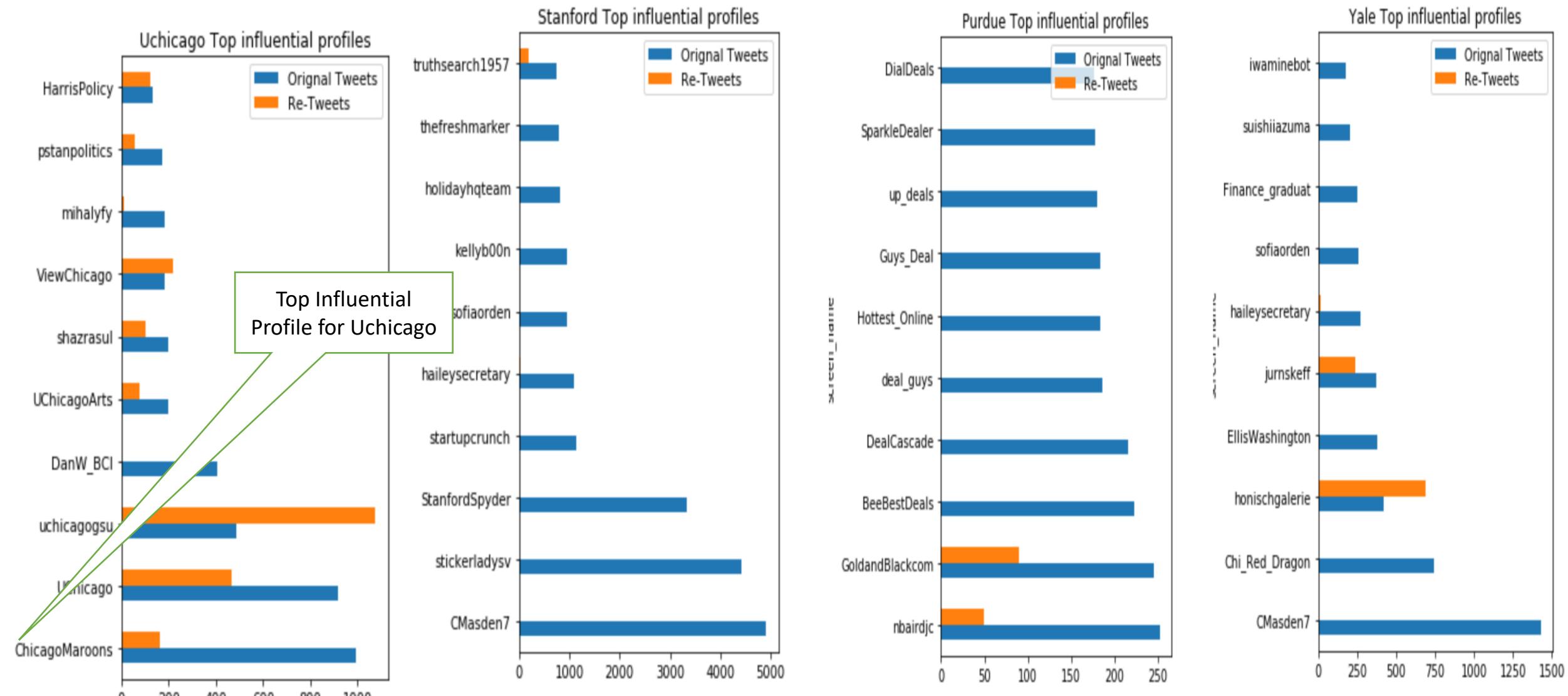
- Projects clearly identified pattern in UChicago tweets, which is largely related to full time and part hiring post
- UChicago should try to encourage more posts about current news, social & political topics or even sports related to university to increase social media footprint on Twitter.
- UChicago should try to increase Social media followers throughout USA rather than just concentrating on Chicago or Midwest region
- UChicago should not limit their tweets just from official handler but encourage individual students/professors to write about university's popular events such as prestigious award nominations and winners as well as current social and sports news related to University.

Project Workflow & Source Data Overview



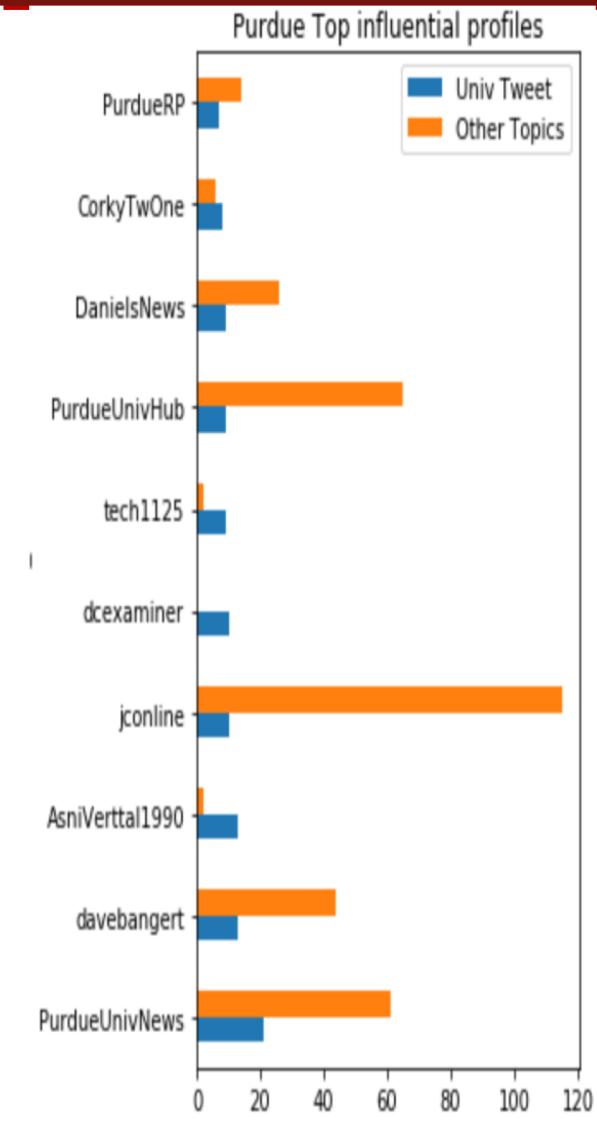
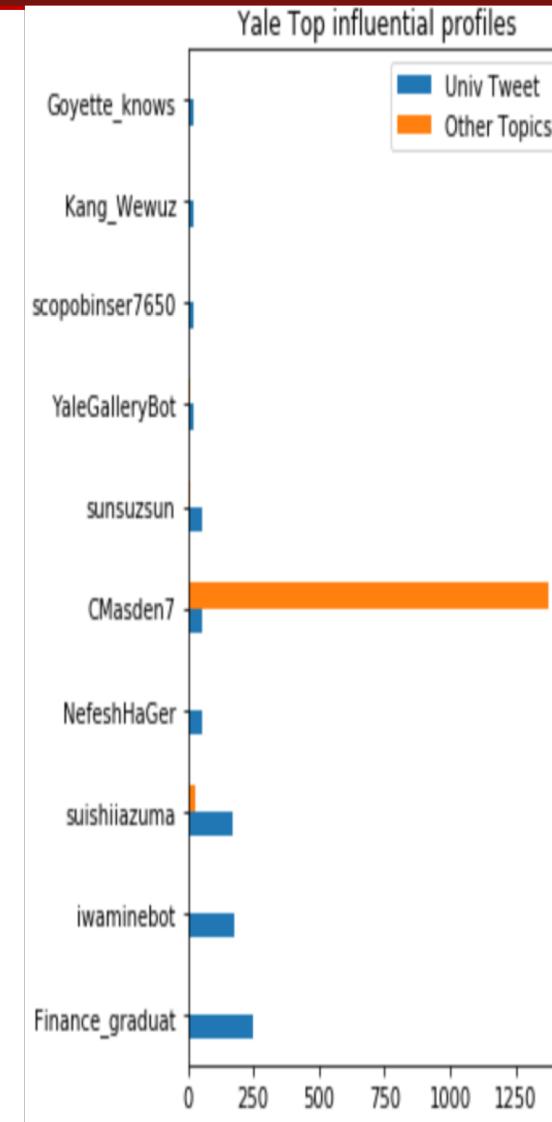
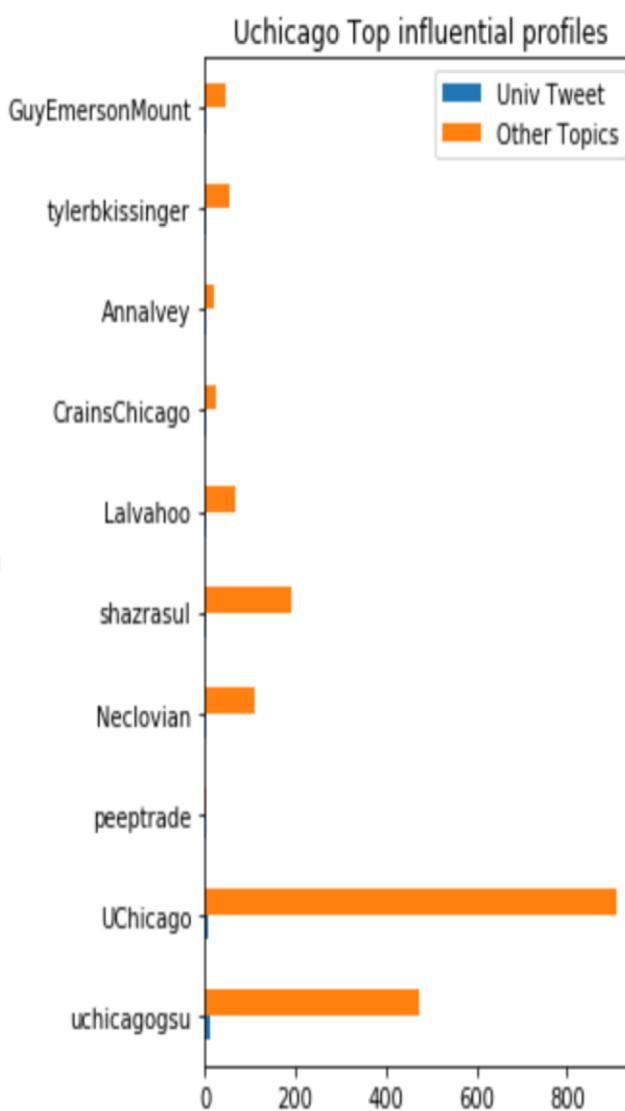
- **Data Source:** A live twitter collection of various topics was provided in the form of jason files;. Project scanned around 60M tweet to select tweets related to the University of Chicago, Stanford, Yale and Purdue universities.
- **Data Filtering:** From the data provided, tweets related to 4 universities of interest were filtered using predetermined keywords based on secondary research. Total 1.2 M tweets of 4 universities were downloaded to local laptop in form for json and then converted to Pandas Dataframe.
- Important columns from twitter json useful are - user.id, screen_name, text, user_location, created_at (tweet's date)
- Project created following variables to increase insights and generate meaningful recommendations
 - a) University name of Tweets
 - b) Flag - Other-topics tweet(if text does not contain word “University”)
 - c) Tweet Count after omitting Retweets
 - d) Latitude and longitude from User Location
- Original tweets vs retweets identified by presence of “RT @” in text column. Influential profiles were identified by counts of original tweets counts.
- University tweets vs OtherTopics tweet identified by presence of word “university” in text of column.

Compare Original Tweets vs Retweets Of Most Influential Profiles



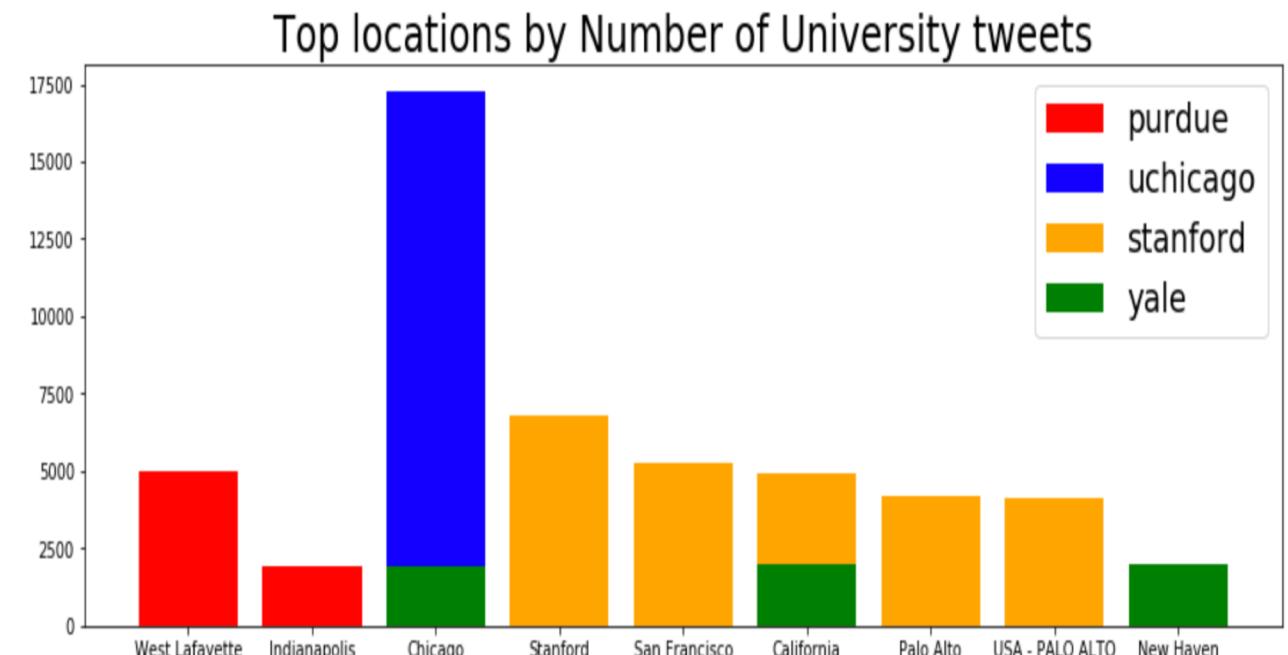
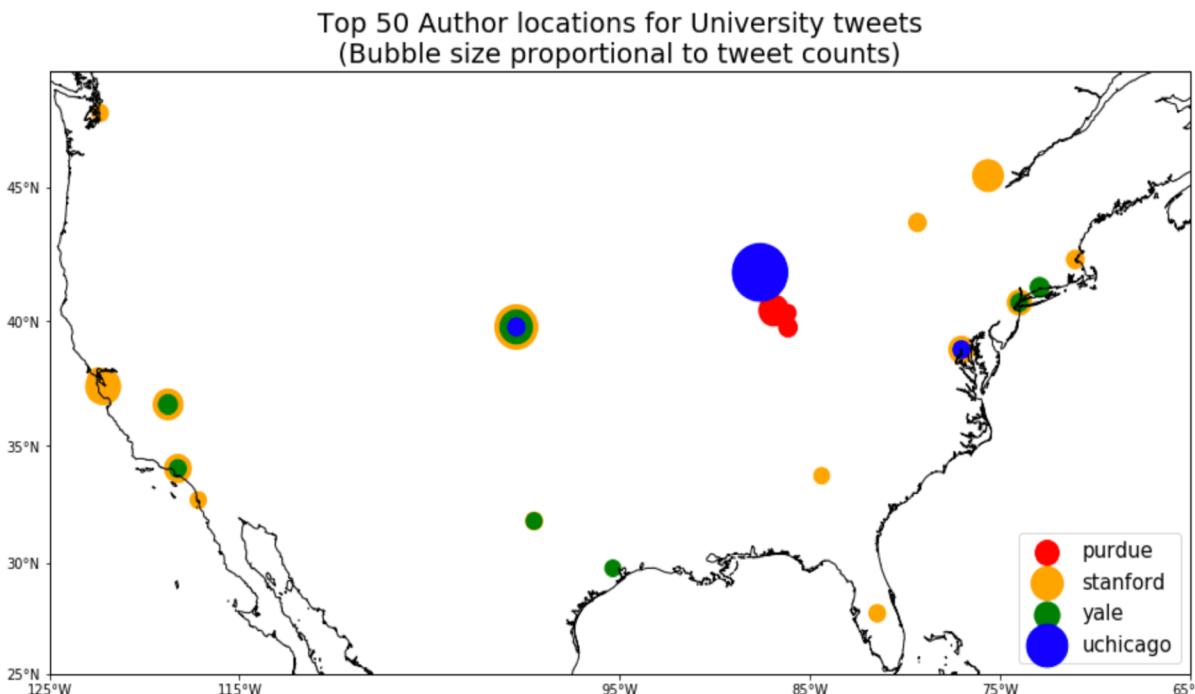
UChicago influencers are retweeting more compare to other universities influencers

Compare University Tweets vs Other Topic tweets Of Most Influential Profiles



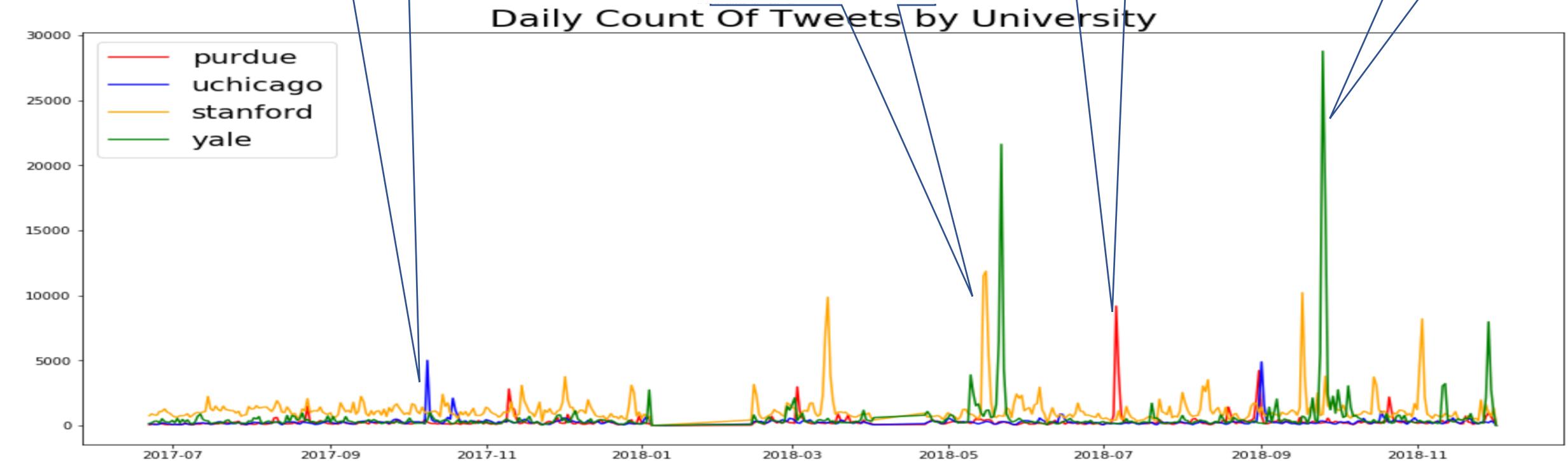
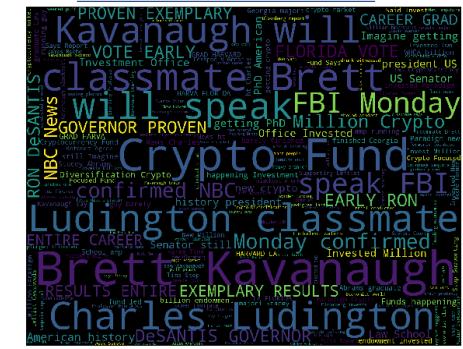
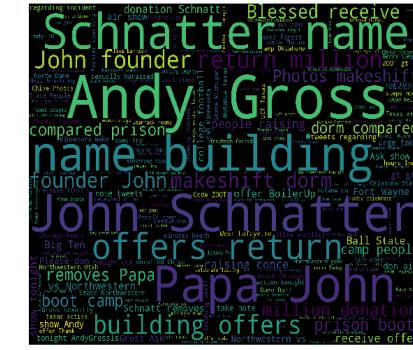
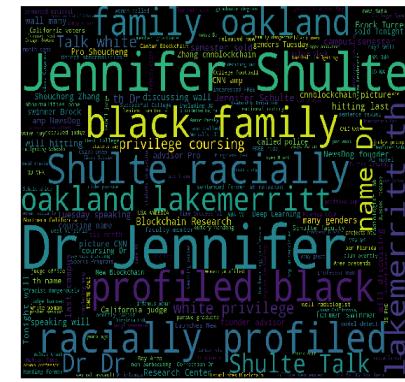
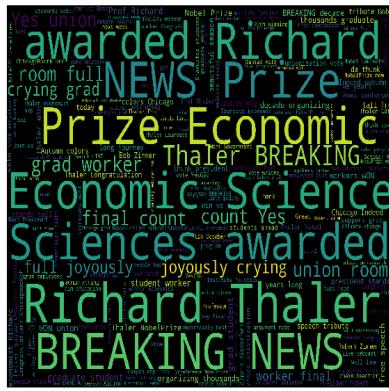
Most Influential profiles are tweeting about Other Topics like Sport & social and not about university topics

Author's Location For All Four Universities

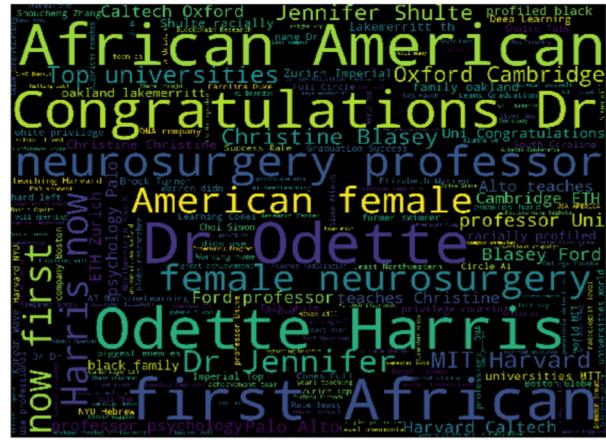
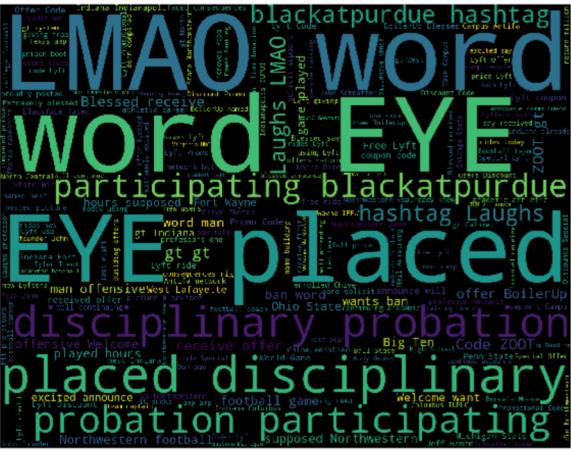
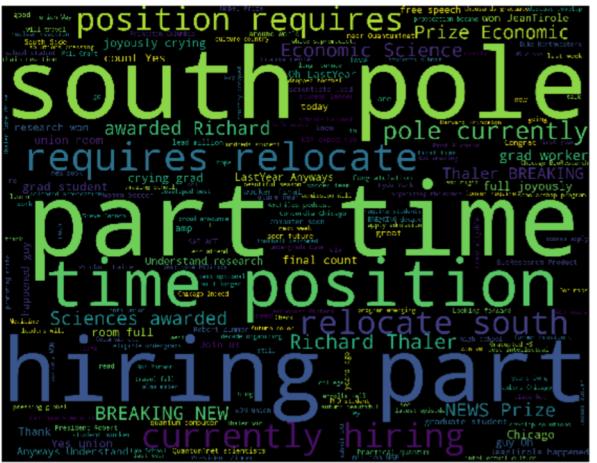


- UChicago and Purdue university tweets are concentrated at their respective city locations
- Stanford and Yale authors are well spread across USA

- There is significant increase in twitter activity in 2018 compared to 2017. March 2017, June 2018 and October 2018 are showing high activity
- Please see wordcloud for those month to see what each universities are tweeting for those months
- Timeline shows missing data around Jan 2018 and Late April 2018



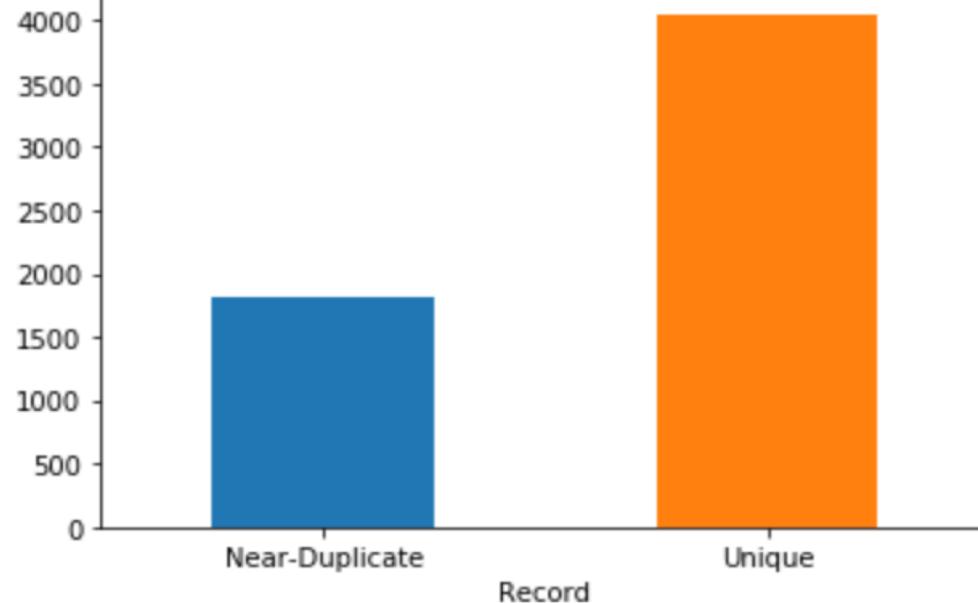
WordCloud Comparisons Of Uchicago and Other Universities



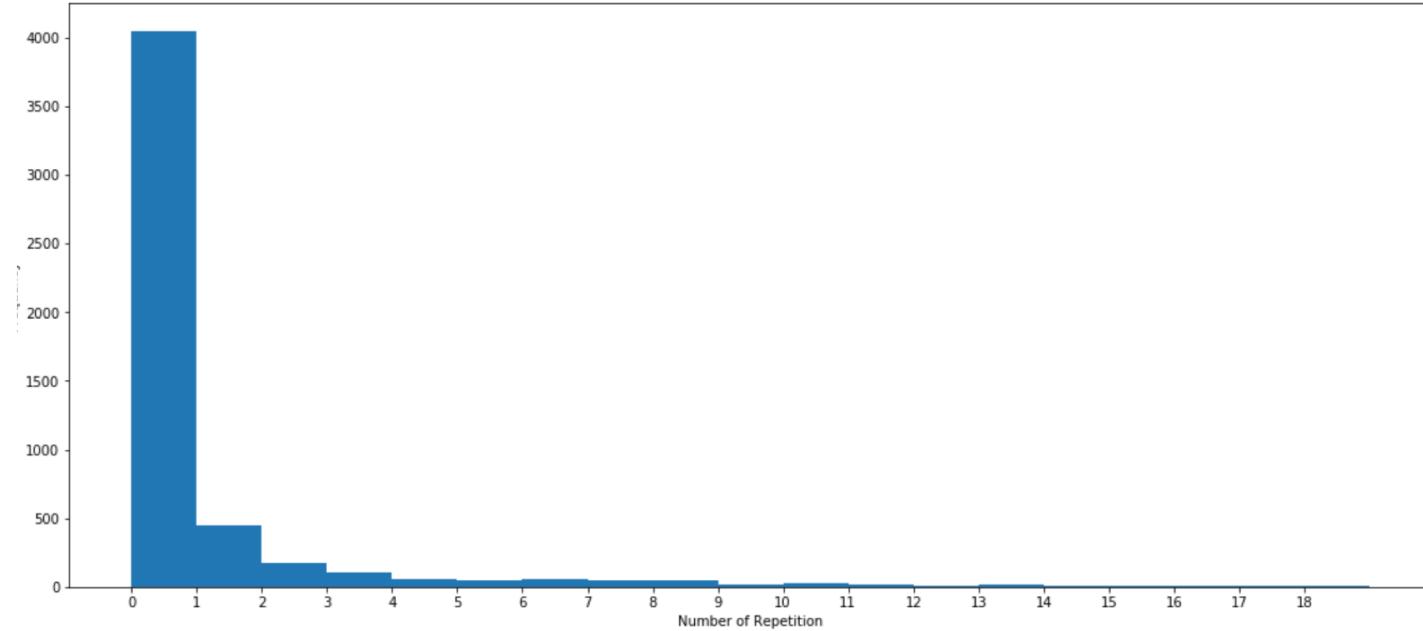
- UChicago accounts are tweeting more about hiring part time and full time positions
- Stanford is tweeting more about famous professors and persons from the university
- Yale twitter accounts are active about current news topics and student protests.
- Histogram on left side shows that more profiles are adding original and different content, if we just consider 15K original retweets .

Similarity of Tweets

Near-Duplicate vs Unique for all Universities



Histogram of Near Duplicate for Universities



- Bar chart on left side is showing more original and different content in tweets, provided we ignore retweets from analysis
- Histogram on left side shows that more profiles are adding original and different content, if we just consider 15K original retweets .