In [24]:

```
import pandas as pd
import numpy as np
import time
import re
import pyspark
import string
from pyspark.sql.types import *
from pyspark.sql.functions import *
```

In [2]:

```
tweets_df = spark.read.json('hdfs:///user/ivy2/Tweets/')
```

In [3]:

```
tweets_df.cache()
```

Out[3]:

```
DataFrame[contributors: string, coordinates: struct<coordinates:array<double>,type:string>,
created_at: string, display_text_range: array<bigint>, entities:
struct<hashtags:array<struct<indices:array<bigint>,text:string>>,media:array<struct<additional_medi
fo:struct<description:string,embeddable:boolean,monetizable:boolean,title:string>,description:strin
splay_url:string,expanded_url:string,id:bigint,id_str:string,indices:array<bigint>,media_url:string
ia_url_https:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigint>,medium:struct<h:bigi
esize:string,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:struct<h:bigint,resize:s
g,w:bigint>>,source_status_id:bigint,source_status_id_str:string,source_user_id:bigint,source_user_
tr:string,type:string,url:string>>,symbols:array<struct<indices:array<bigint>,text:string>>,urls:ar
struct<display_url:string,expanded_url:string,indices:array<bigint>,url:string>>,user_mentions:arra
ruct<id:bigint,id_str:string,indices:array<bigint>,name:string,screen_name:string>>>,
extended_entities:
struct<media:array<struct<additional_media_info:struct<description:string,embeddable:boolean,moneti
e:boolean,title:string>,description:string,display_url:string,expanded_url:string,id:bigint,id_str:
ng,indices:array<bigint>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bigint
ize:string,w:bigint>,medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resize:st
,w:bigint>,thumb:struct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status_id_
string,source_user_id:bigint,source_user_id_str:string,type:string,url:string,video_info:struct<asp
ratio:array<bigint>,duration_millis:bigint,variants:array<struct<bitrate:bigint,content_type:string
:string>>>>>>, extended_tweet:
struct<display_text_range:array<bigint>,entities:struct<hashtags:array<struct<indices:array<bigint>
t:string>>,media:array<struct<additional_media_info:struct<description:string,embeddable:boolean,mo
zable:boolean,title:string>,description:string,display_url:string,expanded_url:string,id:bigint,id_
string,indices:array<bigint>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bi
,resize:string,w:bigint>,medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resiz
ring,w:bigint>,thumb:struct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status
str:string,source_user_id:bigint,source_user_id_str:string,type:string,url:string,video_info:struct
ect_ratio:array<bigint>,duration_millis:bigint,variants:array<struct<bitrate:bigint,content_type:st
,url:string>>>>>,symbols:array<struct<indices:array<bigint>,text:string>>,urls:array<struct<display
:string,expanded_url:string,indices:array<bigint>,url:string>>,user_mentions:array<struct<id:bigint
str:string,indices:array<bigint>,name:string,screen_name:string>>>,extended_entities:struct<media:a
<struct<additional_media_info:struct<description:string,embeddable:boolean,monetizable:boolean,titl
ring>,description:string,display_url:string,expanded_url:string,id:bigint,id_str:string,indices:arr
igint>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bigint,resize:string,w:b
t>,medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thu
truct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status_id_str:string,source_
_id:bigint,source_user_id_str:string,type:string,url:string,video_info:struct<aspect_ratio:array<bi
>,duration_millis:bigint,variants:array<struct<bitrate:bigint,content_type:string,url:string>>>>>>,
_text:string>, favorite_count: bigint, favorited: boolean, filter_level: string, geo:
struct<coordinates:array<double>,type:string>, id: bigint, id_str: string,
in_reply_to_screen_name: string, in_reply_to_status_id: bigint, in_reply_to_status_id_str: string,
in_reply_to_user_id: bigint, in_reply_to_user_id_str: string, is_quote_status: boolean, lang: stri
ng, limit: struct<timestamp_ms:string,track:bigint>, place:
struct<bounding_box:struct<coordinates:array<array<array<double>>>,type:string>,country:string,cour
code:string,full_name:string,id:string,name:string,place_type:string,url:string>,
possibly_sensitive: boolean, quote_count: bigint, quoted_status:
struct<contributors:array<bigint>,coordinates:struct<coordinates:array<double>,type:string>,created
string,display_text_range:array<bigint>,entities:struct<hashtags:array<struct<indices:array<bigint>
t:string>>,media:array<struct<additional_media_info:struct<description:string,embeddable:boolean,mo
zable:boolean,title:string>,description:string,display_url:string,expanded_url:string,id:bigint,id_
string,indices:array<bigint>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bi
```

,resize:string,w:bigint>,medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resiz
ring,w:bigint>,thumb:struct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status
str:string,source_user_id:bigint,source_user_id_str:string,type:string,url:string>>,symbols:array<s
t<indices:array<bigint>,text:string>>,urls:array<struct<display_url:string,expanded_url:string,indi
array<bigint>,url:string>>,user_mentions:array<struct<id:bigint,id_str:string,indices:array<bigint>
e:string,screen_name:string>>>,extended_entities:struct<media:array<struct<additional_media_info:st
<description:string,embeddable:boolean,monetizable:boolean,title:string>,description:string,display
:string,expanded_url:string,id:bigint,id_str:string,indices:array<bigint>,media_url:string,media_ur
tps:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigint>,medium:struct<h:bigint,resize
ing,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:struct<h:bigint,resize:string,w:b
t>>,source_status_id:bigint,source_status_id_str:string,source_user_id:bigint,source_user_id_str:st
,type:string,url:string,video_info:struct<aspect_ratio:array<bigint>,duration_millis:bigint,variant
ray<struct<bitrate:bigint,content_type:string,url:string>>>>>,extended_tweet:struct<display_text_r
:array<bigint>,entities:struct<hashtags:array<struct<indices:array<bigint>,text:string>>,media:arra
ruct<additional_media_info:struct<description:string,embeddable:boolean,monetizable:boolean,title:s
g>,description:string,display_url:string,expanded_url:string,id:bigint,id_str:string,indices:array<
nt>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigi
medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:
ct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status_id_str:string,source_use
:bigint,source_user_id_str:string,type:string,url:string,video_info:struct<aspect_ratio:array<bigin
uration_millis:bigint,variants:array<struct<bitrate:bigint,content_type:string,url:string>>>>>,symb
array<struct<indices:array<bigint>,text:string>>,urls:array<struct<display_url:string,expanded_url:
ng,indices:array<bigint>,url:string>>,user_mentions:array<struct<id:bigint,id_str:string,indices:ar
bigint>,name:string,screen_name:string>>>,extended_entities:struct<media:array<struct<additional_me
info:struct<description:string,embeddable:boolean,monetizable:boolean,title:string>,description:str
display_url:string,expanded_url:string,id:bigint,id_str:string,indices:array<bigint>,media_url:stri
edia_url_https:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigint>,medium:struct<h:bi
,resize:string,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:struct<h:bigint,resize
ing,w:bigint>>,source_status_id:bigint,source_status_id_str:string,source_user_id:bigint,source_use
_str:string,type:string,url:string,video_info:struct<aspect_ratio:array<bigint>,duration_millis:big
variants:array<struct<bitrate:bigint,content_type:string,url:string>>>>>>,full_text:string>,favorit
unt:bigint,favorited:boolean,filter_level:string,geo:struct<coordinates:array<double>,type:string>,
igint,id_str:string,in_reply_to_screen_name:string,in_reply_to_status_id:bigint,in_reply_to_status_
tr:string,in_reply_to_user_id:bigint,in_reply_to_user_id_str:string,is_quote_status:boolean,lang:st
,place:struct<bounding_box:struct<coordinates:array<array<array<double>>>,type:string>,country:stri
ountry_code:string,full_name:string,id:string,name:string,place_type:string,url:string>,possibly_se
ive:boolean,quote_count:bigint,quoted_status_id:bigint,quoted_status_id_str:string,reply_count:bigi
etweet_count:bigint,retweeted:boolean,scopes:struct<followers:boolean,place_ids:array<string>>,sour
tring,text:string,truncated:boolean,user:struct<contributors_enabled:boolean,created_at:string,defa
profile:boolean,default_profile_image:boolean,description:string,favourites_count:bigint,follow_req
_sent:string,followers_count:bigint,following:string,friends_count:bigint,geo_enabled:boolean,id:bi
,id_str:string,is_translator:boolean,lang:string,listed_count:bigint,location:string,name:string,nc
cations:string,profile_background_color:string,profile_background_image_url:string,profile_backgrou
mage_url_https:string,profile_background_tile:boolean,profile_banner_url:string,profile_image_url:s
g,profile_image_url_https:string,profile_link_color:string,profile_sidebar_border_color:string,prof
sidebar_fill_color:string,profile_text_color:string,profile_use_background_image:boolean,protected:
ean,screen_name:string,statuses_count:bigint,time_zone:string,translator_type:string,url:string,utc
set:bigint,verified:boolean>,withheld_copyright:boolean,withheld_in_countries:array<string>>,
quoted_status_id: bigint, quoted_status_id_str: string, quoted_status_permalink:
struct<display:string,expanded:string,url:string>, reply_count: bigint, retweet_count: bigint, ret
weeted: boolean, retweeted_status:
struct<contributors:array<bigint>,coordinates:struct<coordinates:array<double>,type:string>,created
string,display_text_range:array<bigint>,entities:struct<hashtags:array<struct<indices:array<bigint>
t:string>>,media:array<struct<additional_media_info:struct<description:string,embeddable:boolean,mo
zable:boolean,title:string>,description:string,display_url:string,expanded_url:string,id:bigint,id_
string,indices:array<bigint>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bi
,resize:string,w:bigint>,medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resiz
ring,w:bigint>,thumb:struct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status
str:string,source_user_id:bigint,source_user_id_str:string,type:string,url:string>>,symbols:array<s
t<indices:array<bigint>,text:string>>,urls:array<struct<display_url:string,expanded_url:string,indi
array<bigint>,url:string>>,user_mentions:array<struct<id:bigint,id_str:string,indices:array<bigint>
e:string,screen_name:string>>>,extended_entities:struct<media:array<struct<additional_media_info:st
<description:string,embeddable:boolean,monetizable:boolean,title:string>,description:string,display
:string,expanded_url:string,id:bigint,id_str:string,indices:array<bigint>,media_url:string,media_ur
tps:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigint>,medium:struct<h:bigint,resize
ing,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:struct<h:bigint,resize:string,w:b
t>>,source_status_id:bigint,source_status_id_str:string,source_user_id:bigint,source_user_id_str:st
,type:string,url:string,video_info:struct<aspect_ratio:array<bigint>,duration_millis:bigint,variant
ray<struct<bitrate:bigint,content_type:string,url:string>>>>>,extended_tweet:struct<display_text_r
:array<bigint>,entities:struct<hashtags:array<struct<indices:array<bigint>,text:string>>,media:arra
ruct<additional_media_info:struct<description:string,embeddable:boolean,monetizable:boolean,title:s
g>,description:string,display_url:string,expanded_url:string,id:bigint,id_str:string,indices:array<
nt>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigi
medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:
ct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status_id_str:string,source_use
:bigint,source_user_id_str:string,type:string,url:string,video_info:struct<aspect_ratio:array<bigin
uration millis:bigint,variants:array<struct<bitrate:bigint,content type:string,url:string>>>>>,symb

array<struct<indices:array<bigint>,text:string>>,urls:array<struct<display_url:string,expanded_url:
ng,indices:array<bigint>,url:string>>,user_mentions:array<struct<id:bigint,id_str:string,indices:ar
bigint>,name:string,screen_name:string>>>,extended_entities:struct<media:array<struct<additional_me
info:struct<description:string,embeddable:boolean,monetizable:boolean,title:string>,description:str
display_url:string,expanded_url:string,id:bigint,id_str:string,indices:array<bigint>,media_url:stri
edia_url_https:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigint>,medium:struct<h:bi
,resize:string,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:struct<h:bigint,resize
ing,w:bigint>>,source_status_id:bigint,source_status_id_str:string,source_user_id:bigint,source_use
_str:string,type:string,url:string,video_info:struct<aspect_ratio:array<bigint>,duration_millis:big
variants:array<struct<bitrate:bigint,content_type:string,url:string>>>>>>,full_text:string>,favorit
unt:bigint,favorited:boolean,filter_level:string,geo:struct<coordinates:array<double>,type:string>,
igint,id_str:string,in_reply_to_screen_name:string,in_reply_to_status_id:bigint,in_reply_to_status_
tr:string,in_reply_to_user_id:bigint,in_reply_to_user_id_str:string,is_quote_status:boolean,lang:st
,place:struct<bounding_box:struct<coordinates:array<array<array<double>>>,type:string>,country:stri
ountry_code:string,full_name:string,id:string,name:string,place_type:string,url:string>,possibly_se
ive:boolean,quote_count:bigint,quoted_status:struct<contributors:array<bigint>,coordinates:struct<c
inates:array<double>,type:string>,created_at:string,display_text_range:array<bigint>,entities:struc
shtags:array<struct<indices:array<bigint>,text:string>>,media:array<struct<additional_media_info:st
<description:string,embeddable:boolean,monetizable:boolean,title:string>,description:string,display
:string,expanded_url:string,id:bigint,id_str:string,indices:array<bigint>,media_url:string,media_ur
tps:string,sizes:struct<large:struct<h:bigint,resize:string,w:bigint>,medium:struct<h:bigint,resize
ing,w:bigint>,small:struct<h:bigint,resize:string,w:bigint>,thumb:struct<h:bigint,resize:string,w:b
t>>,source_status_id:bigint,source_status_id_str:string,source_user_id:bigint,source_user_id_str:st
,type:string,url:string>>,symbols:array<struct<indices:array<bigint>,text:string>>,urls:array<struc
splay_url:string,expanded_url:string,indices:array<bigint>,url:string>>,user_mentions:array<struct<
igint,id_str:string,indices:array<bigint>,name:string,screen_name:string>>>,extended_entities:struc
dia:array<struct<additional_media_info:struct<description:string,embeddable:boolean,monetizable:boo
,title:string>,description:string,display_url:string,expanded_url:string,id:bigint,id_str:string,in
s:array<bigint>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bigint,resize:s
g,w:bigint>,medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resize:string,w:bi
>,thumb:struct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status_id_str:strin
urce_user_id:bigint,source_user_id_str:string,type:string,url:string,video_info:struct<aspect_ratio
ay<bigint>,duration_millis:bigint,variants:array<struct<bitrate:bigint,content_type:string,url:stri
>>>>,extended_tweet:struct<display_text_range:array<bigint>,entities:struct<hashtags:array<struct<i
es:array<bigint>,text:string>>,media:array<struct<additional_media_info:struct<description:string,e
dable:boolean,monetizable:boolean,title:string>,description:string,display_url:string,expanded_url:
ng,id:bigint,id_str:string,indices:array<bigint>,media_url:string,media_url_https:string,sizes:stru
arge:struct<h:bigint,resize:string,w:bigint>,medium:struct<h:bigint,resize:string,w:bigint>,small:s
t<h:bigint,resize:string,w:bigint>,thumb:struct<h:bigint,resize:string,w:bigint>>,source_status_id:
nt,source_status_id_str:string,source_user_id:bigint,source_user_id_str:string,type:string,url:stri
ideo_info:struct<aspect_ratio:array<bigint>,duration_millis:bigint,variants:array<struct<bitrate:bi
,content_type:string,url:string>>>>,symbols:array<struct<indices:array<bigint>,text:string>>,urls:
y<struct<display_url:string,expanded_url:string,indices:array<bigint>,url:string>>,user_mentions:ar
struct<id:bigint,id_str:string,indices:array<bigint>,name:string,screen_name:string>>>,extended_ent
s:struct<media:array<struct<additional_media_info:struct<description:string,embeddable:boolean,mone
ble:boolean,title:string>,description:string,display_url:string,expanded_url:string,id:bigint,id_st
ring,indices:array<bigint>,media_url:string,media_url_https:string,sizes:struct<large:struct<h:bigi
esize:string,w:bigint>,medium:struct<h:bigint,resize:string,w:bigint>,small:struct<h:bigint,resize:
ng,w:bigint>,thumb:struct<h:bigint,resize:string,w:bigint>>,source_status_id:bigint,source_status_i
r:string,source_user_id:bigint,source_user_id_str:string,type:string,url:string,video_info:struct<a
t_ratio:array<bigint>,duration_millis:bigint,variants:array<struct<bitrate:bigint,content_type:stri
rl:string>>>>>,full_text:string>,favorite_count:bigint,favorited:boolean,filter_level:string,geo:s
t<coordinates:array<double>,type:string>,id:bigint,id_str:string,in_reply_to_screen_name:string,in_
y_to_status_id:bigint,in_reply_to_status_id_str:string,in_reply_to_user_id:bigint,in_reply_to_user_
tr:string,is_quote_status:boolean,lang:string,place:struct<bounding_box:struct<coordinates:array<ar
array<double>>>,type:string>,country:string,country_code:string,full_name:string,id:string,name:str
place_type:string,url:string>,possibly_sensitive:boolean,quote_count:bigint,quoted_status_id:bigint
ted_status_id_str:string,reply_count:bigint,retweet_count:bigint,retweeted:boolean,scopes:struct<fo
ers:boolean,place_ids:array<string>>,source:string,text:string,truncated:boolean,user:struct<contri
rs_enabled:boolean,created_at:string,default_profile:boolean,default_profile_image:boolean,descript
string,favourites_count:bigint,follow_request_sent:string,followers_count:bigint,following:string,f
ds_count:bigint,geo_enabled:boolean,id:bigint,id_str:string,is_translator:boolean,lang:string,liste
unt:bigint,location:string,name:string,notifications:string,profile_background_color:string,profile
kground_image_url:string,profile_background_image_url_https:string,profile_background_tile:boolean,
ile_banner_url:string,profile_image_url:string,profile_image_url_https:string,profile_link_color:st
,profile_sidebar_border_color:string,profile_sidebar_fill_color:string,profile_text_color:string,pr
e_use_background_image:boolean,protected:boolean,screen_name:string,statuses_count:bigint,time_zone
ing,translator_type:string,url:string,utc_offset:bigint,verified:boolean>,withheld_copyright:boolea
thheld_in_countries:array<string>>,quoted_status_id:bigint,quoted_status_id_str:string,quoted_statu
rmalink:struct<display:string,expanded:string,url:string>,reply_count:bigint,retweet_count:bigint,r
eted:boolean,scopes:struct<followers:boolean,place_ids:array<string>>,source:string,text:string,tru
ed:boolean,user:struct<contributors_enabled:boolean,created_at:string,default_profile:boolean,defau
rofile_image:boolean,description:string,favourites_count:bigint,follow_request_sent:string,follower
unt:bigint,following:string,friends_count:bigint,geo_enabled:boolean,id:bigint,id_str:string,is_tra
tor:boolean,lang:string,listed_count:bigint,location:string,name:string,notifications:string,profil
ckground_color:string,profile_background_image_url:string,profile_background_image_url_https:string
file_background_tile:boolean,profile_banner_url:string,profile_image_url:string,profile_image_url_h

```
:string,profile_link_color:string,profile_sidebar_border_color:string,profile_sidebar_fill_color:st
,profile_text_color:string,profile_use_background_image:boolean,protected:boolean,screen_name:strin
atuses_count:bigint,time_zone:string,translator_type:string,url:string,utc_offset:bigint,verified:b
an>,withheld_copyright:boolean,withheld_in_countries:array<string>>, scopes:
struct<place_ids:array<string>>, source: string, text: string, timestamp_ms: string, truncated:
boolean, user:
struct<contributors_enabled:boolean,created_at:string,default_profile:boolean,default_profile_image
lean,description:string,favourites_count:bigint,follow_request_sent:string,followers_count:bigint,f
wing:string,friends_count:bigint,geo_enabled:boolean,id:bigint,id_str:string,is_translator:boolean,
:string,listed_count:bigint,location:string,name:string,notifications:string,profile_background_col
tring,profile_background_image_url:string,profile_background_image_url_https:string,profile_backgro
tile:boolean,profile_banner_url:string,profile_image_url:string,profile_image_url_https:string,prof
link_color:string,profile_sidebar_border_color:string,profile_sidebar_fill_color:string,profile_tex
lor:string,profile_use_background_image:boolean,protected:boolean,screen_name:string,statuses_count
int,time_zone:string,translator_type:string,url:string,utc_offset:bigint,verified:boolean>,
withheld_copyright: boolean, withheld_in_countries: array<string>]
```

In [5]:

```
pd.set_option("display.max_columns", 50)
pd.set_option("display.max_colwidth", 100)
tweets_df.limit(3).toPandas()
```

Out[5]:

	contributors	coordinates	created_at	display_text_range	entities	
0	None	None	Thu Jun 22 23:16:02 +0000 2017	None	([], [(None, None, pic.twitter.com/ly3fCiX1x5, https://twitter.com/millselle/status/875063995505...	([(None, None, https://twitter.cc
1	None	None	Thu Jun 22 23:16:02 +0000 2017	[0, 19]	([], None, [], [(twitter.com/politicalkathy…, https://twitter.com/politicalkathy/status/87802169...	None
2	None	None	Thu Jun 22 23:16:02 +0000 2017	None	([], None, [], [], [(21906952, 21906952, [3, 11], ACEP, ACEPNow)])	None

In [6]:

```
tweets_df.count()
```

Out[6]:

```
118168286
```

Identify tweets related to UChicago and 3-4 universities of your choice & Discard irrelevant tweets

In [7]:

```
chicago = ['%University of Chicago%', '%university of chicago%', '%UChicago%', '%Uchicago%',
'%uchicago%',
           '%UofChicago%', '%uofchicago%', '%U of Chicago%', '%u of chicago%']
harvard = ['%Harvard%', '%harvard%', '%Harvard University%', '%harvard university%']
stanford = ['%Stanford%', '%stanford%', '%Stanford University%', '%stanford university%']
northwestern = ['%Northwestern%', '%northwestern%', '%Northwestern University%', '%northwestern uni
versity%']
```

In [10]:

```
df = tweets_df.filter(
    tweets_df.text.like(chicago[0])      |
    tweets_df.text.like(chicago[1])      |
    tweets_df.text.like(chicago[2])      |
    tweets_df.text.like(chicago[3])      |
    tweets_df.text.like(chicago[4])      |
    tweets_df.text.like(chicago[5])      |
    tweets_df.text.like(chicago[6])      |
    tweets_df.text.like(chicago[7])      |
    tweets_df.text.like(chicago[8])      |
    tweets_df.text.like(harvard[0])      |
    tweets_df.text.like(harvard[1])      |
    tweets_df.text.like(harvard[2])      |
    tweets_df.text.like(harvard[3])      |
    tweets_df.text.like(stanford[0])     |
    tweets_df.text.like(stanford[1])     |
    tweets_df.text.like(stanford[2])     |
    tweets_df.text.like(stanford[3])     |
    tweets_df.text.like(northwestern[0])|
    tweets_df.text.like(northwestern[1])|
    tweets_df.text.like(northwestern[2])|
    tweets_df.text.like(northwestern[3])
)
```

In [17]:

```
df = df.\
withColumn("university",
           when(col('text').like(chicago[0])   |
                col('text').like(chicago[1])   |
                col('text').like(chicago[2])   |
                col('text').like(chicago[3])   |
                col('text').like(chicago[4])   |
                col('text').like(chicago[5])   |
                col('text').like(chicago[6])   |
                col('text').like(chicago[7])   |
                col('text').like(chicago[8]), 'Chicago').\
           when(col('text').like(harvard[0])   |
                col('text').like(harvard[1])   |
                col('text').like(harvard[2])   |
                col('text').like(harvard[3]), 'Harvard').\
           when(col('text').like(stanford[0])   |
                col('text').like(stanford[1])   |
                col('text').like(stanford[2])   |
                col('text').like(stanford[3]), 'Stanford').\
           when(col('text').like(northwestern[0])   |
                col('text').like(northwestern[1])   |
                col('text').like(northwestern[2])   |
                col('text').like(northwestern[3]), 'Northwestern'))
```

In [21]:

```
df.write.format("parquet").save("hdfs:///user/hlee22/final")
```

In [2]:

```
df = spark.read.parquet("hdfs:///user/hlee22/final")
```

Complete thorough EDA to identify which variables you can use to profile the tweet authors

In [3]:

```
pd.set_option("display.max_colwidth", -1)
df.limit(3).toPandas()
```

Out[3]:

	contributors	coordinates	created_at	display_text_range	entities	extended_entities	extended_tweet	favorite_c
0	None	None	Wed Mar 14 20:15:02 +0000 2018	None	([], None, [], [], [(777302125, 777302125, [3, 18], Kimberly Cotzias, kimbrolyclaire), (33639255, 33639255, [20, 34], Northwestern, NorthwesternU)])	None	None	0
1	None	None	Wed Mar 14 20:15:02 +0000 2018	None	([], None, [], [], [])	None	None	0

	contributors	coordinates	created_at	display_text_range	entities	extended_entities	extended_tweet	favorite_
2	None	None	Wed Mar 14 20:15:02 +0000 2018	None	([], None, [], [], [(87602778, 87602778, [3, 15], Katie Little, KatieLittle), (33639255, 33639255, [69, 83], Northwestern, NorthwesternU)])	None	None	0

3 rows × 41 columns

In [4]:

```
df.printSchema()
```

```
root
 |-- contributors: string (nullable = true)
 |-- coordinates: struct (nullable = true)
 |    |-- coordinates: array (nullable = true)
 |    |    |-- element: double (containsNull = true)
 |    |-- type: string (nullable = true)
 |-- created_at: string (nullable = true)
 |-- display_text_range: array (nullable = true)
 |    |-- element: long (containsNull = true)
 |-- entities: struct (nullable = true)
 |    |-- hashtags: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- text: string (nullable = true)
 |    |-- media: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- additional_media_info: struct (nullable = true)
 |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |-- description: string (nullable = true)
 |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |-- id: long (nullable = true)
 |    |    |    |-- id_str: string (nullable = true)
```

```
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- media_url: string (nullable = true)
|    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |-- type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |-- symbols: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- text: string (nullable = true)
|    |-- urls: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- display_url: string (nullable = true)
|    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- url: string (nullable = true)
|    |-- user_mentions: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- id: long (nullable = true)
|    |    |    |-- id_str: string (nullable = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- name: string (nullable = true)
|    |    |    |-- screen_name: string (nullable = true)
|-- extended_entities: struct (nullable = true)
|    |-- media: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |-- title: string (nullable = true)
|    |    |    |-- description: string (nullable = true)
|    |    |    |-- display_url: string (nullable = true)
|    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |-- id: long (nullable = true)
|    |    |    |-- id_str: string (nullable = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- media_url: string (nullable = true)
|    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
```

```
|    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |-- type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|-- extended_tweet: struct (nullable = true)
|    |-- display_text_range: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- entities: struct (nullable = true)
|    |    |-- hashtags: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |-- symbols: array (nullable = true)
```

```
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- urls: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |-- user_mentions: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |-- screen_name: string (nullable = true)
|    |-- extended_entities: struct (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |-- full_text: string (nullable = true)
|-- favorite_count: long (nullable = true)
|-- favorited: boolean (nullable = true)
|-- filter_level: string (nullable = true)
|-- geo: struct (nullable = true)
|    |-- coordinates: array (nullable = true)
|    |    |-- element: double (containsNull = true)
|    |-- type: string (nullable = true)
|-- id: long (nullable = true)
```

```
|-- id_str: string (nullable = true)
|-- in_reply_to_screen_name: string (nullable = true)
|-- in_reply_to_status_id: long (nullable = true)
|-- in_reply_to_status_id_str: string (nullable = true)
|-- in_reply_to_user_id: long (nullable = true)
|-- in_reply_to_user_id_str: string (nullable = true)
|-- is_quote_status: boolean (nullable = true)
|-- lang: string (nullable = true)
|-- limit: struct (nullable = true)
|    |-- timestamp_ms: string (nullable = true)
|    |-- track: long (nullable = true)
|-- place: struct (nullable = true)
|    |-- bounding_box: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: array (containsNull = true)
|    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- country: string (nullable = true)
|    |-- country_code: string (nullable = true)
|    |-- full_name: string (nullable = true)
|    |-- id: string (nullable = true)
|    |-- name: string (nullable = true)
|    |-- place_type: string (nullable = true)
|    |-- url: string (nullable = true)
|-- possibly_sensitive: boolean (nullable = true)
|-- quote_count: long (nullable = true)
|-- quoted_status: struct (nullable = true)
|    |-- contributors: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- coordinates: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- created_at: string (nullable = true)
|    |-- display_text_range: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- entities: struct (nullable = true)
|    |    |-- hashtags: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
```

```
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |-- symbols: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- urls: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |-- user_mentions: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |-- screen_name: string (nullable = true)
|    |-- extended_entities: struct (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |-- extended_tweet: struct (nullable = true)
|    |    |-- display_text_range: array (nullable = true)
|    |    |    |-- element: long (containsNull = true)
```

```
|    |    |-- entities: struct (nullable = true)
|    |    |    |-- hashtags: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- symbols: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- urls: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- user_mentions: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |    |-- screen_name: string (nullable = true)
|    |    |-- extended_entities: struct (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
```

```
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |-- full_text: string (nullable = true)
|    |-- favorite_count: long (nullable = true)
|    |-- favorited: boolean (nullable = true)
|    |-- filter_level: string (nullable = true)
|    |-- geo: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- id: long (nullable = true)
|    |-- id_str: string (nullable = true)
|    |-- in_reply_to_screen_name: string (nullable = true)
|    |-- in_reply_to_status_id: long (nullable = true)
|    |-- in_reply_to_status_id_str: string (nullable = true)
|    |-- in_reply_to_user_id: long (nullable = true)
|    |-- in_reply_to_user_id_str: string (nullable = true)
|    |-- is_quote_status: boolean (nullable = true)
|    |-- lang: string (nullable = true)
|    |-- place: struct (nullable = true)
|    |    |-- bounding_box: struct (nullable = true)
|    |    |    |-- coordinates: array (nullable = true)
|    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |    |-- element: double (containsNull = true)
|    |    |    |-- type: string (nullable = true)
|    |    |-- country: string (nullable = true)
|    |    |-- country_code: string (nullable = true)
|    |    |-- full_name: string (nullable = true)
|    |    |-- id: string (nullable = true)
|    |    |-- name: string (nullable = true)
|    |    |-- place_type: string (nullable = true)
|    |    |-- url: string (nullable = true)
```

```
|    |-- possibly_sensitive: boolean (nullable = true)
|    |-- quote_count: long (nullable = true)
|    |-- quoted_status_id: long (nullable = true)
|    |-- quoted_status_id_str: string (nullable = true)
|    |-- reply_count: long (nullable = true)
|    |-- retweet_count: long (nullable = true)
|    |-- retweeted: boolean (nullable = true)
|    |-- scopes: struct (nullable = true)
|    |    |-- followers: boolean (nullable = true)
|    |    |-- place_ids: array (nullable = true)
|    |    |    |-- element: string (containsNull = true)
|    |-- source: string (nullable = true)
|    |-- text: string (nullable = true)
|    |-- truncated: boolean (nullable = true)
|    |-- user: struct (nullable = true)
|    |    |-- contributors_enabled: boolean (nullable = true)
|    |    |-- created_at: string (nullable = true)
|    |    |-- default_profile: boolean (nullable = true)
|    |    |-- default_profile_image: boolean (nullable = true)
|    |    |-- description: string (nullable = true)
|    |    |-- favourites_count: long (nullable = true)
|    |    |-- follow_request_sent: string (nullable = true)
|    |    |-- followers_count: long (nullable = true)
|    |    |-- following: string (nullable = true)
|    |    |-- friends_count: long (nullable = true)
|    |    |-- geo_enabled: boolean (nullable = true)
|    |    |-- id: long (nullable = true)
|    |    |-- id_str: string (nullable = true)
|    |    |-- is_translator: boolean (nullable = true)
|    |    |-- lang: string (nullable = true)
|    |    |-- listed_count: long (nullable = true)
|    |    |-- location: string (nullable = true)
|    |    |-- name: string (nullable = true)
|    |    |-- notifications: string (nullable = true)
|    |    |-- profile_background_color: string (nullable = true)
|    |    |-- profile_background_image_url: string (nullable = true)
|    |    |-- profile_background_image_url_https: string (nullable = true)
|    |    |-- profile_background_tile: boolean (nullable = true)
|    |    |-- profile_banner_url: string (nullable = true)
|    |    |-- profile_image_url: string (nullable = true)
|    |    |-- profile_image_url_https: string (nullable = true)
|    |    |-- profile_link_color: string (nullable = true)
|    |    |-- profile_sidebar_border_color: string (nullable = true)
|    |    |-- profile_sidebar_fill_color: string (nullable = true)
|    |    |-- profile_text_color: string (nullable = true)
|    |    |-- profile_use_background_image: boolean (nullable = true)
|    |    |-- protected: boolean (nullable = true)
|    |    |-- screen_name: string (nullable = true)
|    |    |-- statuses_count: long (nullable = true)
|    |    |-- time_zone: string (nullable = true)
|    |    |-- translator_type: string (nullable = true)
|    |    |-- url: string (nullable = true)
|    |    |-- utc_offset: long (nullable = true)
|    |    |-- verified: boolean (nullable = true)
|    |-- withheld_copyright: boolean (nullable = true)
|    |-- withheld_in_countries: array (nullable = true)
|    |    |-- element: string (containsNull = true)
|-- quoted_status_id: long (nullable = true)
|-- quoted_status_id_str: string (nullable = true)
|-- quoted_status_permalink: struct (nullable = true)
|    |-- display: string (nullable = true)
|    |-- expanded: string (nullable = true)
|    |-- url: string (nullable = true)
|-- reply_count: long (nullable = true)
|-- retweet_count: long (nullable = true)
|-- retweeted: boolean (nullable = true)
|-- retweeted_status: struct (nullable = true)
|    |-- contributors: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- coordinates: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- created_at: string (nullable = true)
|    |-- display_text_range: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- entities: struct (nullable = true)
```

```
|    |    |-- hashtags: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |-- symbols: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- urls: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |-- user_mentions: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |-- screen_name: string (nullable = true)
|    |-- extended_entities: struct (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
```

```
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |-- extended_tweet: struct (nullable = true)
|    |    |    |-- display_text_range: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- entities: struct (nullable = true)
|    |    |    |    |-- hashtags: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |    |-- media: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
```

```
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- symbols: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |    |-- urls: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- user_mentions: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |    |    |-- screen_name: string (nullable = true)
|    |    |    |-- extended_entities: struct (nullable = true)
|    |    |    |    |-- media: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |    |-- video_info: struct (nullable = true)
```

```
|    |    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- full_text: string (nullable = true)
|    |    |-- favorite_count: long (nullable = true)
|    |    |-- favorited: boolean (nullable = true)
|    |    |-- filter_level: string (nullable = true)
|    |    |-- geo: struct (nullable = true)
|    |    |    |-- coordinates: array (nullable = true)
|    |    |    |    |-- element: double (containsNull = true)
|    |    |    |-- type: string (nullable = true)
|    |    |-- id: long (nullable = true)
|    |    |-- id_str: string (nullable = true)
|    |    |-- in_reply_to_screen_name: string (nullable = true)
|    |    |-- in_reply_to_status_id: long (nullable = true)
|    |    |-- in_reply_to_status_id_str: string (nullable = true)
|    |    |-- in_reply_to_user_id: long (nullable = true)
|    |    |-- in_reply_to_user_id_str: string (nullable = true)
|    |    |-- is_quote_status: boolean (nullable = true)
|    |    |-- lang: string (nullable = true)
|    |    |-- place: struct (nullable = true)
|    |    |    |-- bounding_box: struct (nullable = true)
|    |    |    |    |-- coordinates: array (nullable = true)
|    |    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |    |    |-- element: double (containsNull = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |-- country: string (nullable = true)
|    |    |    |-- country_code: string (nullable = true)
|    |    |    |-- full_name: string (nullable = true)
|    |    |    |-- id: string (nullable = true)
|    |    |    |-- name: string (nullable = true)
|    |    |    |-- place_type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |    |-- possibly_sensitive: boolean (nullable = true)
|    |    |-- quote_count: long (nullable = true)
|    |    |-- quoted_status: struct (nullable = true)
|    |    |    |-- contributors: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- coordinates: struct (nullable = true)
|    |    |    |    |-- coordinates: array (nullable = true)
|    |    |    |    |    |-- element: double (containsNull = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |-- created_at: string (nullable = true)
|    |    |    |-- display_text_range: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- entities: struct (nullable = true)
|    |    |    |    |-- hashtags: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |    |-- media: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
```

```
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- symbols: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |    |-- urls: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- user_mentions: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |    |    |-- screen_name: string (nullable = true)
|    |    |    |-- extended_entities: struct (nullable = true)
|    |    |    |    |-- media: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
```

```
|    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- extended_tweet: struct (nullable = true)
|    |    |    |    |-- display_text_range: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- entities: struct (nullable = true)
|    |    |    |    |    |-- hashtags: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |    |    |-- media: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- symbols: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |    |    |-- urls: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
```

```
|    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- user_mentions: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |    |    |    |-- screen_name: string (nullable = true)
|    |    |    |    |-- extended_entities: struct (nullable = true)
|    |    |    |    |    |-- media: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- full_text: string (nullable = true)
|    |    |    |-- favorite_count: long (nullable = true)
|    |    |    |-- favorited: boolean (nullable = true)
|    |    |    |-- filter_level: string (nullable = true)
|    |    |    |-- geo: struct (nullable = true)
|    |    |    |    |-- coordinates: array (nullable = true)
|    |    |    |    |    |-- element: double (containsNull = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |-- id: long (nullable = true)
|    |    |    |-- id_str: string (nullable = true)
|    |    |    |-- in_reply_to_screen_name: string (nullable = true)
|    |    |    |-- in_reply_to_status_id: long (nullable = true)
|    |    |    |-- in_reply_to_status_id_str: string (nullable = true)
|    |    |    |-- in_reply_to_user_id: long (nullable = true)
|    |    |    |-- in_reply_to_user_id_str: string (nullable = true)
|    |    |    |-- is_quote_status: boolean (nullable = true)
```

```
|    |    |-- lang: string (nullable = true)
|    |    |-- place: struct (nullable = true)
|    |    |    |-- bounding_box: struct (nullable = true)
|    |    |    |    |-- coordinates: array (nullable = true)
|    |    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |    |    |-- element: double (containsNull = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |-- country: string (nullable = true)
|    |    |    |-- country_code: string (nullable = true)
|    |    |    |-- full_name: string (nullable = true)
|    |    |    |-- id: string (nullable = true)
|    |    |    |-- name: string (nullable = true)
|    |    |    |-- place_type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |    |-- possibly_sensitive: boolean (nullable = true)
|    |    |-- quote_count: long (nullable = true)
|    |    |-- quoted_status_id: long (nullable = true)
|    |    |-- quoted_status_id_str: string (nullable = true)
|    |    |-- reply_count: long (nullable = true)
|    |    |-- retweet_count: long (nullable = true)
|    |    |-- retweeted: boolean (nullable = true)
|    |    |-- scopes: struct (nullable = true)
|    |    |    |-- followers: boolean (nullable = true)
|    |    |    |-- place_ids: array (nullable = true)
|    |    |    |    |-- element: string (containsNull = true)
|    |    |-- source: string (nullable = true)
|    |    |-- text: string (nullable = true)
|    |    |-- truncated: boolean (nullable = true)
|    |    |-- user: struct (nullable = true)
|    |    |    |-- contributors_enabled: boolean (nullable = true)
|    |    |    |-- created_at: string (nullable = true)
|    |    |    |-- default_profile: boolean (nullable = true)
|    |    |    |-- default_profile_image: boolean (nullable = true)
|    |    |    |-- description: string (nullable = true)
|    |    |    |-- favourites_count: long (nullable = true)
|    |    |    |-- follow_request_sent: string (nullable = true)
|    |    |    |-- followers_count: long (nullable = true)
|    |    |    |-- following: string (nullable = true)
|    |    |    |-- friends_count: long (nullable = true)
|    |    |    |-- geo_enabled: boolean (nullable = true)
|    |    |    |-- id: long (nullable = true)
|    |    |    |-- id_str: string (nullable = true)
|    |    |    |-- is_translator: boolean (nullable = true)
|    |    |    |-- lang: string (nullable = true)
|    |    |    |-- listed_count: long (nullable = true)
|    |    |    |-- location: string (nullable = true)
|    |    |    |-- name: string (nullable = true)
|    |    |    |-- notifications: string (nullable = true)
|    |    |    |-- profile_background_color: string (nullable = true)
|    |    |    |-- profile_background_image_url: string (nullable = true)
|    |    |    |-- profile_background_image_url_https: string (nullable = true)
|    |    |    |-- profile_background_tile: boolean (nullable = true)
|    |    |    |-- profile_banner_url: string (nullable = true)
|    |    |    |-- profile_image_url: string (nullable = true)
|    |    |    |-- profile_image_url_https: string (nullable = true)
|    |    |    |-- profile_link_color: string (nullable = true)
|    |    |    |-- profile_sidebar_border_color: string (nullable = true)
|    |    |    |-- profile_sidebar_fill_color: string (nullable = true)
|    |    |    |-- profile_text_color: string (nullable = true)
|    |    |    |-- profile_use_background_image: boolean (nullable = true)
|    |    |    |-- protected: boolean (nullable = true)
|    |    |    |-- screen_name: string (nullable = true)
|    |    |    |-- statuses_count: long (nullable = true)
|    |    |    |-- time_zone: string (nullable = true)
|    |    |    |-- translator_type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |    |    |-- utc_offset: long (nullable = true)
|    |    |    |-- verified: boolean (nullable = true)
|    |    |-- withheld_copyright: boolean (nullable = true)
|    |    |-- withheld_in_countries: array (nullable = true)
|    |    |    |-- element: string (containsNull = true)
|    |-- quoted_status_id: long (nullable = true)
|    |-- quoted_status_id_str: string (nullable = true)
|    |-- quoted_status_permalink: struct (nullable = true)
|    |    |-- display: string (nullable = true)
|    |    |-- expanded: string (nullable = true)
```

```
|    |    |-- url: string (nullable = true)
|    |-- reply_count: long (nullable = true)
|    |-- retweet_count: long (nullable = true)
|    |-- retweeted: boolean (nullable = true)
|    |-- scopes: struct (nullable = true)
|    |    |-- followers: boolean (nullable = true)
|    |    |-- place_ids: array (nullable = true)
|    |    |    |-- element: string (containsNull = true)
|    |-- source: string (nullable = true)
|    |-- text: string (nullable = true)
|    |-- truncated: boolean (nullable = true)
|    |-- user: struct (nullable = true)
|    |    |-- contributors_enabled: boolean (nullable = true)
|    |    |-- created_at: string (nullable = true)
|    |    |-- default_profile: boolean (nullable = true)
|    |    |-- default_profile_image: boolean (nullable = true)
|    |    |-- description: string (nullable = true)
|    |    |-- favourites_count: long (nullable = true)
|    |    |-- follow_request_sent: string (nullable = true)
|    |    |-- followers_count: long (nullable = true)
|    |    |-- following: string (nullable = true)
|    |    |-- friends_count: long (nullable = true)
|    |    |-- geo_enabled: boolean (nullable = true)
|    |    |-- id: long (nullable = true)
|    |    |-- id_str: string (nullable = true)
|    |    |-- is_translator: boolean (nullable = true)
|    |    |-- lang: string (nullable = true)
|    |    |-- listed_count: long (nullable = true)
|    |    |-- location: string (nullable = true)
|    |    |-- name: string (nullable = true)
|    |    |-- notifications: string (nullable = true)
|    |    |-- profile_background_color: string (nullable = true)
|    |    |-- profile_background_image_url: string (nullable = true)
|    |    |-- profile_background_image_url_https: string (nullable = true)
|    |    |-- profile_background_tile: boolean (nullable = true)
|    |    |-- profile_banner_url: string (nullable = true)
|    |    |-- profile_image_url: string (nullable = true)
|    |    |-- profile_image_url_https: string (nullable = true)
|    |    |-- profile_link_color: string (nullable = true)
|    |    |-- profile_sidebar_border_color: string (nullable = true)
|    |    |-- profile_sidebar_fill_color: string (nullable = true)
|    |    |-- profile_text_color: string (nullable = true)
|    |    |-- profile_use_background_image: boolean (nullable = true)
|    |    |-- protected: boolean (nullable = true)
|    |    |-- screen_name: string (nullable = true)
|    |    |-- statuses_count: long (nullable = true)
|    |    |-- time_zone: string (nullable = true)
|    |    |-- translator_type: string (nullable = true)
|    |    |-- url: string (nullable = true)
|    |    |-- utc_offset: long (nullable = true)
|    |    |-- verified: boolean (nullable = true)
|    |-- withheld_copyright: boolean (nullable = true)
|    |-- withheld_in_countries: array (nullable = true)
|    |    |-- element: string (containsNull = true)
|-- scopes: struct (nullable = true)
|    |-- place_ids: array (nullable = true)
|    |    |-- element: string (containsNull = true)
|-- source: string (nullable = true)
|-- text: string (nullable = true)
|-- timestamp_ms: string (nullable = true)
|-- truncated: boolean (nullable = true)
|-- user: struct (nullable = true)
|    |-- contributors_enabled: boolean (nullable = true)
|    |-- created_at: string (nullable = true)
|    |-- default_profile: boolean (nullable = true)
|    |-- default_profile_image: boolean (nullable = true)
|    |-- description: string (nullable = true)
|    |-- favourites_count: long (nullable = true)
|    |-- follow_request_sent: string (nullable = true)
|    |-- followers_count: long (nullable = true)
|    |-- following: string (nullable = true)
|    |-- friends_count: long (nullable = true)
|    |-- geo_enabled: boolean (nullable = true)
|    |-- id: long (nullable = true)
|    |-- id_str: string (nullable = true)
|    |-- is_translator: boolean (nullable = true)
|    |-- lang: string (nullable = true)
```

```
 |    |-- listed_count: long (nullable = true)
 |    |-- location: string (nullable = true)
 |    |-- name: string (nullable = true)
 |    |-- notifications: string (nullable = true)
 |    |-- profile_background_color: string (nullable = true)
 |    |-- profile_background_image_url: string (nullable = true)
 |    |-- profile_background_image_url_https: string (nullable = true)
 |    |-- profile_background_tile: boolean (nullable = true)
 |    |-- profile_banner_url: string (nullable = true)
 |    |-- profile_image_url: string (nullable = true)
 |    |-- profile_image_url_https: string (nullable = true)
 |    |-- profile_link_color: string (nullable = true)
 |    |-- profile_sidebar_border_color: string (nullable = true)
 |    |-- profile_sidebar_fill_color: string (nullable = true)
 |    |-- profile_text_color: string (nullable = true)
 |    |-- profile_use_background_image: boolean (nullable = true)
 |    |-- protected: boolean (nullable = true)
 |    |-- screen_name: string (nullable = true)
 |    |-- statuses_count: long (nullable = true)
 |    |-- time_zone: string (nullable = true)
 |    |-- translator_type: string (nullable = true)
 |    |-- url: string (nullable = true)
 |    |-- utc_offset: long (nullable = true)
 |    |-- verified: boolean (nullable = true)
 |-- withheld_copyright: boolean (nullable = true)
 |-- withheld_in_countries: array (nullable = true)
 |    |-- element: string (containsNull = true)
 |-- university: string (nullable = true)
```

In [3]:

```
# Drop duplicates
df = df.dropDuplicates(subset=['id','user'])
```

In [4]:

```
# Add column to add info on whether its retweet
df = df.withColumn('tweet_status',
                   when(col('retweeted_status').isNotNull(),'Retweet').\
                   when(col('text').like('RT %'),'Retweet').\
                   otherwise('Original'))
```

In [14]:

```
df2 = df.select(
    col('id').alias('tweet_id'),

    # add user info
    col('user').getItem('id').alias('user_id'),
    col('user').getItem('name').alias('user_name'),
    col('user').getItem('created_at').alias('user_created'),
    col('user').getItem('lang').alias('user_lang'),
    col('user').getItem('location').alias('user_location'),
    col('user').getItem('time_zone').alias('user_timezone'),
    col('user').getItem('statuses_count').alias('total_tweets'),
    col('user').getItem('followers_count').alias('followers'),
    col('user').getItem('friends_count').alias('followings'),
    col('created_at').alias('created_at'),

    # add geo info
    col('place').getItem('country').alias('country'),
    col('place').getItem('country_code').alias('country_code'),
    col('place').getItem('full_name').alias('full_name'),
    col('place').getItem('name').alias('name'),

    # add hashtags, text, University, and is_orig info
    col('entities').getItem('hashtags').getItem('text').alias('hashtags'),
    col('text').alias('text'),
    col('University').alias('University'),
    col('tweet_status').alias('tweet_status')
)
```

In [15]:

```
df2.write.format("parquet").save("hdfs:///user/hlee22/df2")
```

In [2]:

```
df2 = spark.read.parquet("hdfs:///user/hlee22/df2")
```

In [3]:

```
df2.cache()
```

Out[3]:

```
DataFrame[tweet_id: bigint, user_id: bigint, user_name: string, user_created: string, user_lang:
string, user_location: string, user_timezone: string, total_tweets: bigint, followers: bigint,
followings: bigint, created_at: string, country: string, country_code: string, full_name: string,
name: string, hashtags: array<string>, text: string, University: string, tweet_status: string]
```

In [18]:

```
df2.printSchema()
```

```
root
 |-- tweet_id: long (nullable = true)
 |-- user_id: long (nullable = true)
 |-- user_name: string (nullable = true)
 |-- user_created: string (nullable = true)
 |-- user_lang: string (nullable = true)
 |-- user_location: string (nullable = true)
 |-- user_timezone: string (nullable = true)
 |-- total_tweets: long (nullable = true)
 |-- followers: long (nullable = true)
 |-- followings: long (nullable = true)
 |-- created_at: string (nullable = true)
 |-- country: string (nullable = true)
 |-- country_code: string (nullable = true)
 |-- full_name: string (nullable = true)
 |-- name: string (nullable = true)
 |-- hashtags: array (nullable = true)
 |    |-- element: string (containsNull = true)
 |-- text: string (nullable = true)
 |-- University: string (nullable = true)
 |-- tweet_status: string (nullable = true)
```

Identify the most prolific / influential Twitter users

In [19]:

```
# Total mentions per university
df2.groupBy('University').count().show()
```

```
+------------+-------+
|  University|  count|
+------------+-------+
|Northwestern|1000739|
|     Chicago| 237240|
|    Stanford| 549221|
|     Harvard|1510674|
+------------+-------+
```

In [37]:

```
# User with most message volume
user_id = df2.groupBy(col('user_id')).count().orderBy(col('count').desc()).head(1)[0][0]
volume = df2.filter((col('user_id')==user_id)).groupBy(col('user_id')).count().collect()[0][1]
total = df2.filter(col('user_id')==user_id).select(max(col('total_tweets'))).collect()[0][0]
```

In [78]:

```
print("The user id with the most message volume : {}".format(user_id))
print("The proportion of tweets regarding university vs others : {}".format(volume/total))
print("Tweets about universities: ")
df2.filter(col('user_id')==user_id).groupBy(col('University')).count().show()
```

```
The user with the most message volume : 880005795974217728
The proportion of tweets regarding university vs others : 0.12467503267239077
Tweets about universities:
+----------+-----+
|University|count|
+----------+-----+
|   Chicago|   14|
|  Stanford| 2046|
|   Harvard| 6812|
+----------+-----+
```

In [104]:

```
# User with most message retweet
user_id2 = df2.filter(col('tweet_status')=='Retweet').groupBy(col('user_id')).count().orderBy(col('
count').desc()).head(1)[0][0]
retweets = df2.filter((col('tweet_status')=='Retweet') & (col('user_id')==user_id2)).\
groupBy(col('user_id')).count().select(max(col('count'))).collect()[0][0]
total2 = df2.filter(col('user_id')==user_id2).select(max(col('total_tweets'))).collect()[0][0]
```

In [105]:

```
print("The user id with the most message volume : {}".format(user_id2))
print("The proportion of tweets regarding university vs others : {}".format(retweets/total2))
print("Tweets about universities: ")
df2.filter(col('user_id')==user_id2).groupBy(col('University')).count().show()
```

```
The user id with the most message volume : 887182298340245504
The proportion of tweets regarding university vs others : 0.05979697375981613
Tweets about universities:
+----------+-----+
|University|count|
+----------+-----+
|  Stanford| 1529|
|   Harvard|   32|
+----------+-----+
```

Where are these authors located?

In [181]:

```
# For Uchicago
chi = df2.filter((col('University')=='Chicago') & (col('user_location').isNotNull())).\
groupBy(col('user_location')).count().orderBy(col('count').desc()).toPandas()
chi.head(5)
```

Out[181]:

	user_location	count
0	Chicago, IL	22274
1	Chicago	8194
2	Chicago, Illinois	4239
3	United States	3403
4	Washington, DC	2542

In [182]:

```
# For Northwestern
nor = df2.filter((col('University')=='Northwestern') & (col('user_location').isNotNull())).\
groupBy(col('user_location')).count().orderBy(col('count').desc()).toPandas()
nor.head(5)
```

Out[182]:

	user_location	count
0	Chicago, IL	26625
1	United States	14379
2	Evanston, IL	13982
3	Chicago	11198
4	Washington, DC	4642

In [183]:

```
# For Stanford
sta = df2.filter((col('University')=='Stanford') & (col('user_location').isNotNull())).\
groupBy(col('user_location')).count().orderBy(col('count').desc()).toPandas()
sta.head(5)
```

Out[183]:

	user_location	count
0	United States	10095
1	Stanford, CA	6972
2	USA	5655
3	San Francisco, CA	5246
4	California, USA	4944

In [184]:

```
# For Harvard
har = df2.filter((col('University')=='Harvard') & (col('user_location').isNotNull())).\
groupBy(col('user_location')).count().orderBy(col('count').desc()).toPandas()
har.head(5)
```

Out[184]:

	user_location	count
0	United States	25622
1	Boston, MA	12456
2	Cambridge, MA	10706
3	California, USA	8941
4	New York, NY	8552

Do you see any relationship between university locations and authors locations?

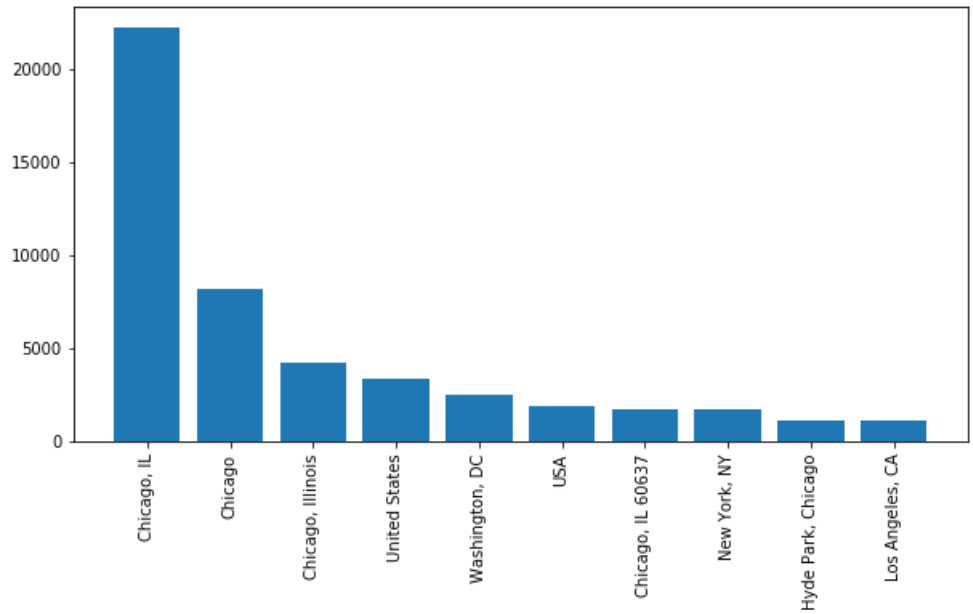
Yes, most users seem to live near where the school is located.

In [40]:

```
import matplotlib.pyplot as plt
```

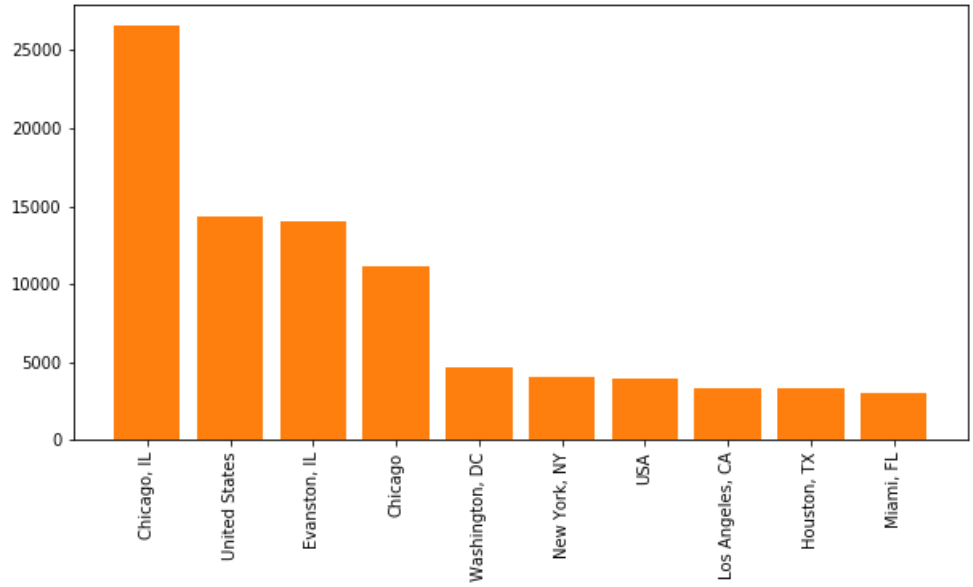
In [135]:

```
# Visualize the relationships (Chicago)

plt.figure(figsize=(10,5))
plt.bar(chi.iloc[0:10,0], chi.iloc[0:10,1])
plt.xticks(rotation='vertical',fontsize=10)
plt.show()
```



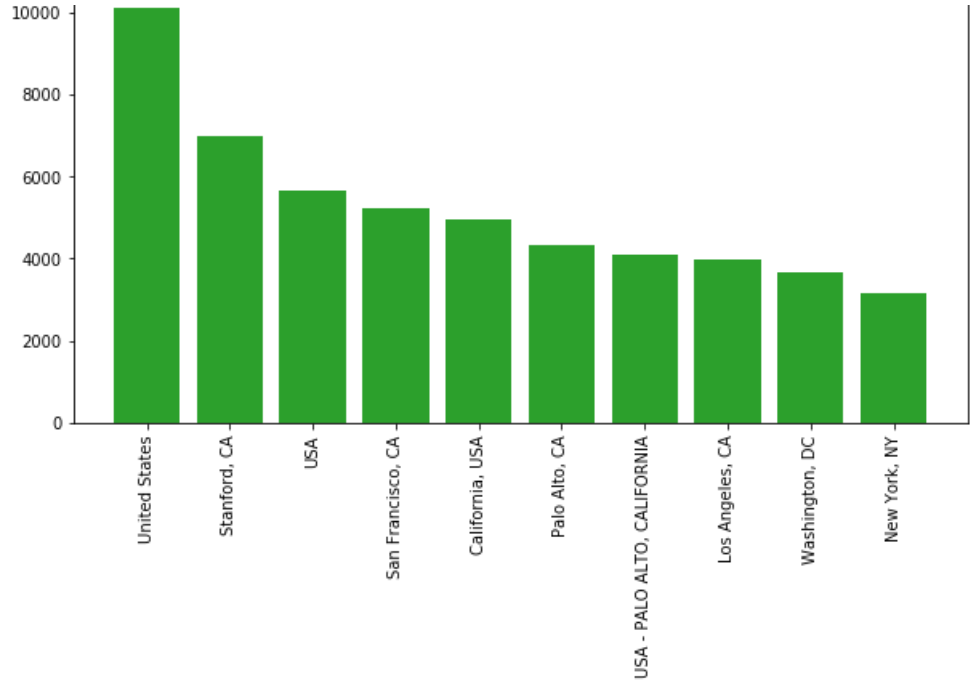
In [137]:

```
# Visualize the relationships (Northwestern)

plt.figure(figsize=(10,5))
plt.bar(nor.iloc[0:10,0], nor.iloc[0:10,1], color='C1')
plt.xticks(rotation='vertical',fontsize=10)
plt.show()
```



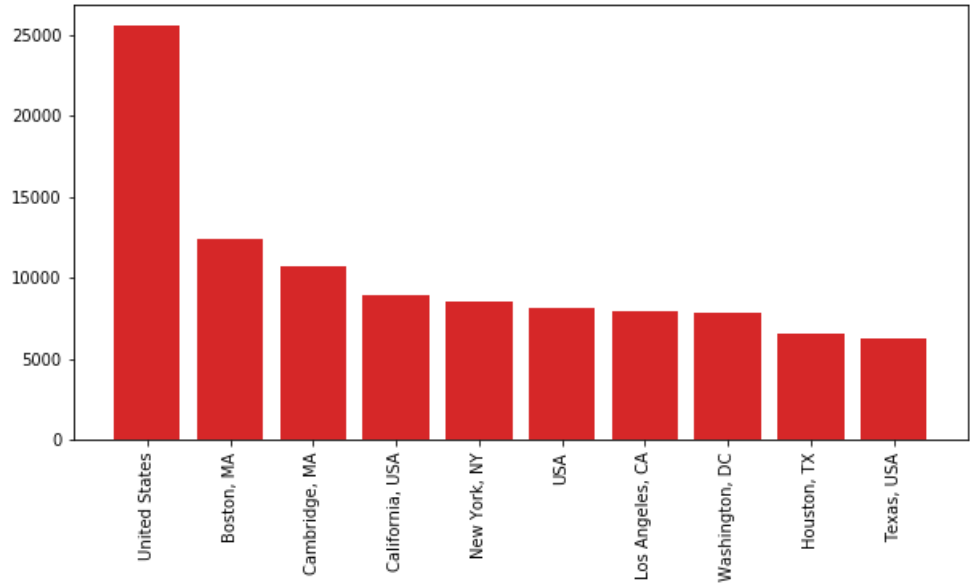
In [138]:

```
# Visualize the relationships (Stanford)

plt.figure(figsize=(10,5))
plt.bar(sta.iloc[0:10,0], sta.iloc[0:10,1], color='C2')
plt.xticks(rotation='vertical',fontsize=10)
plt.show()
```



In [139]:

```
# Visualize the relationships (Harvard)

plt.figure(figsize=(10,5))
plt.bar(har.iloc[0:10,0], har.iloc[0:10,1], color='C3')
plt.xticks(rotation='vertical',fontsize=10)
plt.show()
```



What distinguishes University of Chicago authors vs authors who tweet about other universities

In [185]:

```
chi_sub = chi.iloc[0:10,]
nor_sub = nor.iloc[0:10,]
sta_sub = sta.iloc[0:10,]
har_sub = har.iloc[0:10,]
```

In [186]:

```
# Grouping locations
# UChicago
chi_sub.loc[chi_sub['user_location'].str.contains('Chicago'),'user_location'] = 'Chicago'
chi_sub = chi_sub.groupby('user_location').agg('sum').sort_values('count', ascending=False)
```

```
# Northwestern
nor_sub.loc[nor_sub['user_location'].str.contains('Chicago'),'user_location'] = 'Chicago'
nor_sub.loc[nor_sub['user_location']=='USA','user_location'] = 'United States'
nor_sub = nor_sub.groupby('user_location').agg('sum').sort_values('count', ascending=False)

# Stanford
sta_sub.loc[sta_sub['user_location'].str.contains('PALO'),'user_location'] = 'Palo Alto, CA'
sta_sub.loc[sta_sub['user_location']=='USA','user_location'] = 'United States'
sta_sub = sta_sub.groupby('user_location').agg('sum').sort_values('count', ascending=False)

# Harvard
har_sub.loc[har_sub['user_location']=='USA','user_location'] = 'United States'
har_sub = har_sub.groupby('user_location').agg('sum').sort_values('count', ascending=False)
```

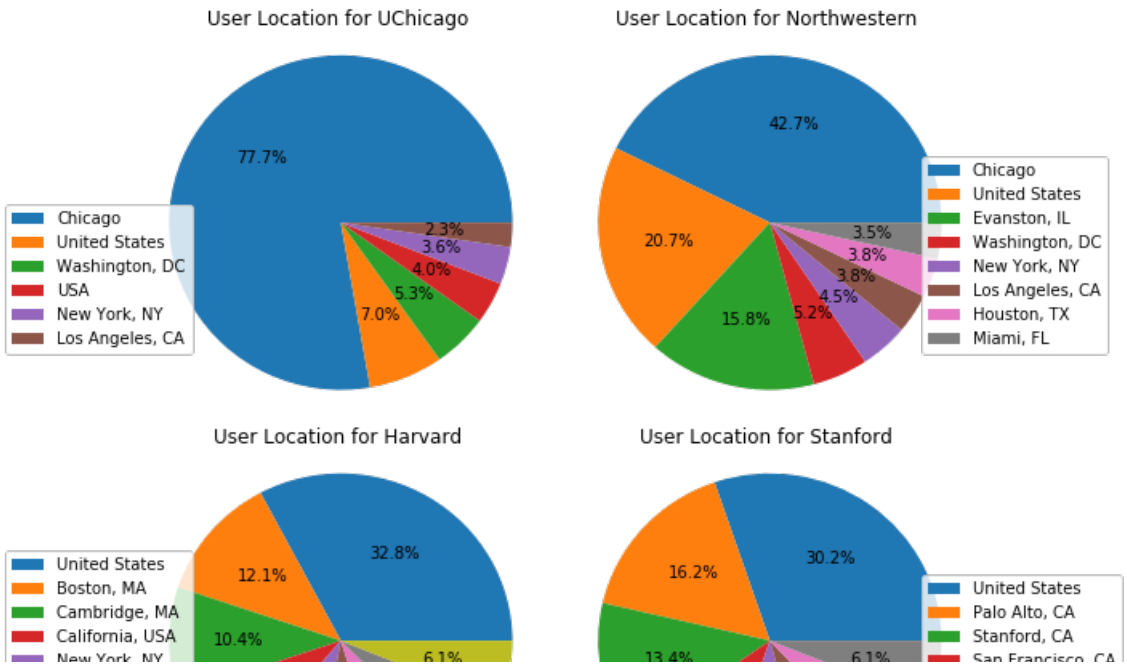
```
/software/Anaconda3-5.1.0-hadoop/lib/python3.6/site-packages/pandas/core/indexing.py:543:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-
docs/stable/indexing.html#indexing-view-versus-copy
  self.obj[item] = s
```

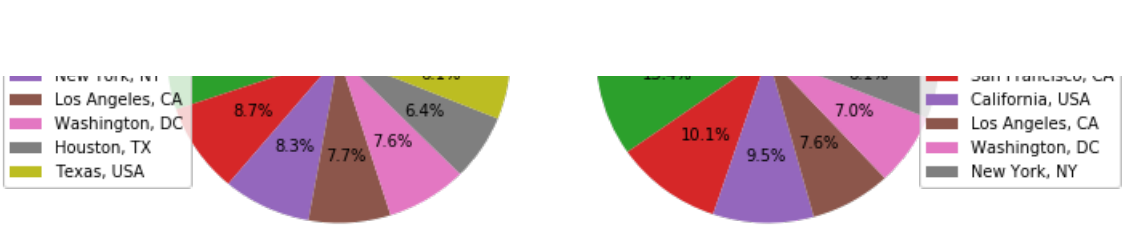
In [187]:

```
# Piecharts of locations of users per university
fig, ax = plt.subplots(2,2, figsize=(8,8))
plt.subplots_adjust(wspace=0.5, hspace=0.5)

ax[0,0].pie(chi_sub, autopct='%1.1f%%', radius=1.5, pctdistance=0.6)
ax[0,0].legend(labels=chi_sub.index, loc="lower left", bbox_to_anchor=(-0.7,0))
ax[0,0].set_title('User Location for UChicago \n\n')

ax[0,1].pie(nor_sub, autopct='%1.1f%%', radius=1.5, pctdistance=0.6)
ax[0,1].legend(labels=nor_sub.index,loc="lower left", bbox_to_anchor=(1,0))
ax[0,1].set_title('User Location for Northwestern \n\n')

ax[1,1].pie(sta_sub, autopct='%1.1f%%', radius=1.5, pctdistance=0.6)
ax[1,1].legend(labels=sta_sub.index,loc="lower left", bbox_to_anchor=(1,0))
ax[1,1].set_title('User Location for Stanford \n\n')

ax[1,0].pie(har_sub, autopct='%1.1f%%', radius=1.5, pctdistance=0.6)
ax[1,0].legend(labels=har_sub.index,loc="lower left", bbox_to_anchor=(-0.7,0))
ax[1,0].set_title('User Location for Harvard \n\n')
```

Out[187]:

```
Text(0.5,1,'User Location for Harvard \n\n')
```





The users who tweet about Uchicago are mostly located in Chicago area (lmost 78%) compared to other universities. About half of users who tweet about Northwestern are also located in Chicago area followed by Evanston where the univeristy is located. In comparison with these two universities, the users who tweet about Harvard and Stanford are quite evenly located in various area.

What are the timelines of these tweets? Do you see significant peaks and valleys? Do you see data collection gaps?

In [41]:

```
from datetime import datetime
from dateutil.parser import parse
```

In [42]:

```
dt =  udf(lambda x: datetime.strptime(x,'%a %b %d %H:%M:%S %z %Y'), DateType())
```

In [43]:

```
df2 = df2.withColumn('created_at', dt(col('created_at')))
```

In [44]:

```
df2.limit(3).toPandas()
```

Out[44]:

	tweet_id	user_id	user_name	user_created	user_lang	user_location	user_timezone	total_tweets	follo
0	877897647524200448	307514476	Andrew Campbell	Sun May 29 19:38:22 +0000 2011	en	None	None	3063	65
1	877910070482292736	77943492	Joel Engel	Mon Sep 28 06:27:55 +0000 2009	en	Southern Calif	Pacific Time (US & Canada)	47583	2678
2	877922072256696321	92093688	jerome henen	Mon Nov 23 19:28:35 +0000 2009	en	North America / Canada	None	9674	140

In [45]:

```
timedf = df2.groupBy(col('created_at'),
col('University')).count().orderBy(col('created_at')).toPandas()
```

In [46]:

```
timedf = timedf.pivot(index='created_at', columns='University', values='count')
timedf.head()
```

Out[46]:

University University	Chicago Chicago	Harvard Harvard	Northwestern Northwestern	Stanford Stanford
created_at created_at				
2017-06-22	178	614	570	749
2017-06-23	211	968	1277	889
2017-06-24	147	744	914	795
2017-06-25	249	915	758	790
2017-06-26	243	964	1209	1081

In [50]:

```
fig, ax = plt.subplots(figsize=(15,8))
plt.plot(timedf)
plt.legend(labels=timedf.columns)
plt.show()
```

There are significant peaks for especially harvard university followed by Northwestern, Stanford, and UChicago.

Harvard has the maximum peak at 2017-12-13, Northwestern at 2018-03-14, Stanford at 2018-05-16, and Uchicago at 2018-06-15. There is a gap between 2018-01-05 and 2018-02-13.

In [230]:

```
timedf.idxmax()
```

Out[230]:

```
University
Chicago          2018-06-15
Harvard          2017-12-13
Northwestern     2018-03-14
Stanford         2018-05-16
dtype: object
```

In [275]:

```
pd.DataFrame(timedf.index)[190:200]
```

Out[275]:

	created_at
190	2017-12-29

	created_at
191	2017-12-30
192	2017-12-31
193	2018-01-01
194	2018-01-02
195	2018-01-03
196	2018-01-04
197	2018-01-05
198	2018-02-13
199	2018-02-14

How unique are the messages about these universities?

In [4]:

```
import nltk as nltk
from nltk.corpus import stopwords
```

In [ ]:

```
stopwords
```

In [8]:

```
# Sample 50000 rows for other universities
chi_text = df2.filter((col('University') =='Chicago') &
                      (col('tweet_status')=='Original')).select(['text']).toPandas()
nor_text = df2.filter((col('University') =='Northwestern') &
                      (col('tweet_status')=='Original')).select(['text']).toPandas().sample(n=50000)
sta_text = df2.filter((col('University') =='Stanford') &
                      (col('tweet_status')=='Original')).select(['text']).toPandas().sample(n=50000)
har_text = df2.filter((col('University') =='Harvard') &
                      (col('tweet_status')=='Original')).select(['text']).toPandas().sample(n=50000)
```

In [63]:

```
def tokenize(df, univ) :

    words = set(nltk.corpus.stopwords.words('english') + list(string.punctuation) + list(univ.lower
().split()))

    n = df.shape[0]
    res = list()
    for i in range(n):
        a = df.iloc[i,0].lower()
        b = nltk.word_tokenize(a)
        b = [word for word in b if word not in words]
        res.append(b)
    return res
```

In [64]:

```
def freq_words(df, univ):

    words = set(nltk.corpus.stopwords.words('english') + list(string.punctuation) + list(univ.lower
().split()))

    n = df.shape[0]
    freq_df = pd.DataFrame()

    for i in range(n):
        a = df.iloc[i,0]
        b = [word for word in a.lower().split() if word not in words]
        if len(b) >= 3:
            b = nltk.ngrams(b,3)
```

```
            c = nltk.FreqDist(b).most_common()
            freq_df = freq_df.append(c)

    freq_df.columns = ['ngram', 'freq']


    freq_df = freq_df.groupby(['ngram']).agg('sum').sort_values(['freq'],ascending=False)

    return freq_df
```

In [28]:

```
chi_text['token'] = tokenize(chi_text, 'University of Chicago')
nor_text['token'] = tokenize(nor_text, 'Northwestern University')
sta_text['token'] = tokenize(sta_text, 'Stanford University')
har_text['token'] = tokenize(har_text, 'Harvard University')
```

In [32]:

```
chi_freq = freq_words(chi_text.iloc[0:,],'University of Chicago')
chi_freq.head(10)
```

Out[32]:

	freq
ngram	
(bioresearch, product, faire)	332
(booth, school, business)	293
(sat,, act, optional)	270
(nobel, prize, economics)	255
(richard, thaler, wins)	248
(make, sat,, act)	243
(becomes, first, elite)	240
(first, elite, college)	225
(elite, college, make)	221
(college, make, sat,)	219

In [34]:

```
nor_freq = freq_words(nor_text.sample(n=50000).iloc[0:,], 'Northwestern University')
nor_freq.head(10)
```

Out[34]:

	freq
ngram	
(severe, thunderstorm, warning)	333
(national, weather, service)	300
(mutual, wealth, management)	268
(thunderstorm, warning, for...)	263
(mutual, investment, management)	185
(significant, weather, advisory)	185
(investment, management, company)	182
(management, company, llc)	182
(wealth, management, co.)	169

(music, city, bowl)	166
	freq

In [66]:

```
sta_freq = freq_words(sta_text.sample(n=50000).iloc[0:,], 'Stanford University')
sta_freq.head(10)
```

Out[66]:

	freq
ngram	
(@stanford, u:, usa)	774
(u:, usa, □□)	734
(usa, □□, america!!)	334
(@stanford, @randihepstein, @yale)	280
(@chelseaclinton, @stanford, @randihepstein)	262
(security, chief, depart)	254
(#stanford, #cardinal, #cfb)	200
(going, save, us)	194
(save, us, sooo)	194
("this, going, save)	194

In [51]:

```
har_freq = freq_words(har_text.sample(n=50000).iloc[0:,], 'Harvard University')
har_freq.head(10)
```

Out[51]:

	freq
ngram	
(@xychelsea, @harvard, @cia)	809
(@foxnews, @hillaryclinton, @harvard)	462
(bear, keychain, written)	382
(@xychelsea, @cia, @harvard)	370
(@potus, @harvard, @brookingsinst)	358
(sociological, study, conducted)	315
(study, conducted, university.)	288
(month, sociological, study)	262
(created, 3d, model)	215
(24, month, sociological)	215

In [68]:

```
# Top 10
chi_top = set([j for i in chi_freq.head(10).index for j in i])
nor_top = set([j for i in nor_freq.head(10).index for j in i])
sta_top = set([j for i in sta_freq.head(10).index for j in i])
har_top = set([j for i in har_freq.head(10).index for j in i])
```

In [69]:

```
def jaccard(df, top):
    n = df.shape[0]
```

```
    n = df.shape[0]
    res = list()
    for i in range(n):
        a = df.iloc[i,1]
        intersection = len(list(set(a).intersection(set(top))))
        union = len(set(a)) + len(set(top)) - intersection
        jaccard_similarity = intersection / union
        res.append(jaccard_similarity)
    return res
```

In [70]:

```
# Jaccard similarity for each university
chi_text['jaccard'] = jaccard(chi_text,chi_top)
nor_text['jaccard'] = jaccard(nor_text,nor_top)
sta_text['jaccard'] = jaccard(sta_text,sta_top)
har_text['jaccard'] = jaccard(har_text,har_top)
```

In [73]:

```
print('UChicago : {:.4f}'.format(chi_text[chi_text['jaccard']>=0.05].shape[0]/chi_text.shape[0]))
print('Northwestern : {:.4f}'.format(nor_text[nor_text['jaccard']>=0.05].shape[0]/nor_text.shape[0]
))
print('Stanford : {:.4f}'.format(sta_text[sta_text['jaccard']>=0.05].shape[0]/sta_text.shape[0]))
print('Harvard : {:.4f}'.format(har_text[har_text['jaccard']>=0.05].shape[0]/har_text.shape[0]))
```

```
UChicago : 0.0368
Northwestern : 0.0337
Stanford : 0.0282
Harvard : 0.0199
```

Based on the frequent topics analysis, I would say small portion of people tweeted similar topics. Uchicago has the highest Jaccard similarity – the topics seem diverse but there are about 500 people talking about nobel prize winner, Richard Thaler. Northwestern has the second highest Jaccard similarity – over 500 people talked about weather.