

A2. Real-Time-Twitter-Listening

February 20, 2020

1 Twitter

Twitter assumes that someone using the Twitter API is most likely creating an app. So, in order to obtain API tokens, we'll have to go through their interface and create a dummy app. I've done this a while ago but here are the general steps:

Apply for a developer account [here](#). Click create an app and fill out the form app details. You can put any website in the website field.

Copy and paste the API keys and tokens into a safe place and never post them on the web! If you lose them, you can regenerate them in the twitter developer account dashboard.

```
In [1]: import twitter
```

```
CONSUMER_KEY = ''
CONSUMER_SECRET = ''
OAUTH_TOKEN = ''
OAUTH_TOKEN_SECRET = ''
```

```
auth = twitter.oauth.OAuth(OAUTH_TOKEN, OAUTH_TOKEN_SECRET,
                           CONSUMER_KEY, CONSUMER_SECRET)
```

```
twitter_api = twitter.Twitter(auth=auth)
```

```
print(twitter_api)
```

```
<twitter.api.Twitter object at 0x00000178F9BC2E10>
```

2 Retrieving World Wide Twitter Trends in Real Time

```
In [2]: WORLD_WOE_ID = 1
        US_WOE_ID = 23424977
```

```
# Prefix ID with the underscore for query string parameterization.
# Without the underscore, the twitter package appends the ID value
# to the URL itself as a special case keyword argument.
```

```

world_trends = twitter_api.trends.place(_id=WORLD_WOE_ID)
us_trends = twitter_api.trends.place(_id=US_WOE_ID)

print(world_trends)
print()
print(us_trends)

[{'trends': [{'name': '#MeCaigoEnVivo', 'url': 'http://twitter.com/search?q=%23MeCaigoEnVivo',

In [3]: for trend in world_trends[0]['trends']:
        print(trend['name'])

#MeCaigoEnVivo
#ReleaseTheContract
#AllInTheFamily
#TheJeffersons
#ElectionResults2019
Marco Ruben
Marla Gibbs
COAF
Morumbi
Casco
Roldán
Paranaense

Vamos River
Rajai Davis
All in the Family
Joplin
José Aldo
Jamie Foxx
George Jefferson
Juventude
Orioles

#LiveInFrontOfASudioAudience
#___
#__22
#1stVideoGameCrush
#_
#RIVERxESPN
#TheChallenge33
#Verdict2019
#BrazilIsExcitedForBTS

```

```
#-----  
#DetenteBenjamín  
#JokowiMundurlah  
#CLAMP  
#__18  
#___  
#WhereAreYouFrom  
#ChicagoFire  
#1stAnimeCrush  
#__7_  
#__  
#___5000_  
#GobackModi  
#RecopaSudamericana  
#  
#ChicagoPD  
#AmazingRace  
#ABPresults2019
```

```
In [4]: for trend in us_trends[0]['trends']:  
        print(trend['name'])
```

```
#AllInTheFamily  
#TheJeffersons  
#LiveInFrontOfASudioAudience  
#ReleaseTheContract  
#TheChallenge33  
Marla Gibbs  
Joplin  
All in the Family  
Jamie Foxx  
Rajai Davis  
George Jefferson  
Orioles  
Gleyber Torres  
Jennifer Hudson  
Norman Lear  
Marisa Tomei  
edith  
Turbo  
Golden City  
Juan Lagares  
Gary Thorne  
Harriet Tubman  
Charlie Tilson  
Archie Bunker  
Carl Junction
```

```
Mother Jefferson
Albert Almora
Woody Harrelson
kerry washington
Sherman Hemsley
New Apartment
#ChicagoFire
#ChicagoPD
#my600lblife
#AmazingRace
#1stAnimeCrush
#JEFFERSONS
#SaveWhiskeyCavalier
#TrumpTantrum
#PLLThePerfectionists
#ImpulsiveBuyingRegrets
#lastnightasaviking
#TheWeirdThingAboutMeIs
#WWENXT
#WhiskeyCavailer
#turbulentselfies
#SEALTeam
#ElectionResults2019
#OneChicago
#Stellaride
```

```
In [5]: world_trends_set = set([trend['name']
                                for trend in world_trends[0]['trends']])

    us_trends_set = set([trend['name']
                          for trend in us_trends[0]['trends']])

    common_trends = world_trends_set.intersection(us_trends_set)

    print(common_trends)

{'#TheJeffersons', 'Joplin', '#1stAnimeCrush', 'Orioles', 'George Jefferson', '#TheChallenge33
```

3 Anatomy of a Tweet

```
In [6]: import json

    q = '#AmazingRace'

    count = 100
```

```

# Import unquote to prevent url encoding errors in next_results
from urllib.parse import unquote

# See https://dev.twitter.com/rest/reference/get/search/tweets

search_results = twitter_api.search.tweets(q=q, count=count)

statuses = search_results['statuses']

# Iterate through 5 more batches of results by following the cursor
for _ in range(5):
    print('Length of statuses', len(statuses))
    try:
        next_results = search_results['search_metadata']['next_results']
    except KeyError as e: # No more results when next_results doesn't exist
        break

    # Create a dictionary from next_results, which has the following form:
    # ?max_id=847960489447628799&q=%23RIPSelena&count=100&include_entities=1
    kwargs = dict([ kv.split('=') for kv in unquote(next_results[1:]).split("&") ])

    search_results = twitter_api.search.tweets(**kwargs)
    statuses += search_results['statuses']

# Show one sample search result by slicing the list...
print(json.dumps(statuses[0], indent=1))

```

```

Length of statuses 65
Length of statuses 140
Length of statuses 227
Length of statuses 327
Length of statuses 415

```

```

{
  "created_at": "Thu May 23 02:57:00 +0000 2019",
  "id": 1131393564984918016,
  "id_str": "1131393564984918016",
  "text": "RT @AmazingRaceCBS: \ud83e\udd23\ud83e\udd23\ud83e\udd23 #AmazingRace https://t.co/8...",
  "truncated": false,
  "entities": {
    "hashtags": [
      {
        "text": "AmazingRace",
        "indices": [
          24,
          36
        ]
      }
    ]
  }
}

```

```

    }
  ],
  "symbols": [],
  "user_mentions": [
    {
      "screen_name": "AmazingRaceCBS",
      "name": "The Amazing Race",
      "id": 68792362,
      "id_str": "68792362",
      "indices": [
        3,
        18
      ]
    }
  ],
  "urls": [],
  "media": [
    {
      "id": 1131368162186817536,
      "id_str": "1131368162186817536",
      "indices": [
        37,
        60
      ],
      "media_url": "http://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
      "media_url_https": "https://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
      "url": "https://t.co/8JBE6f0Jie",
      "display_url": "pic.twitter.com/8JBE6f0Jie",
      "expanded_url": "https://twitter.com/AmazingRaceCBS/status/1131369650489401344/photo/1",
      "type": "photo",
      "sizes": {
        "thumb": {
          "w": 150,
          "h": 150,
          "resize": "crop"
        },
        "large": {
          "w": 500,
          "h": 280,
          "resize": "fit"
        },
        "medium": {
          "w": 500,
          "h": 280,
          "resize": "fit"
        },
        "small": {
          "w": 500,

```

```

        "h": 280,
        "resize": "fit"
    }
},
"source_status_id": 1131369650489401344,
"source_status_id_str": "1131369650489401344",
"source_user_id": 68792362,
"source_user_id_str": "68792362"
}
]
},
"extended_entities": {
    "media": [
        {
            "id": 1131368162186817536,
            "id_str": "1131368162186817536",
            "indices": [
                37,
                60
            ],
            "media_url": "http://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
            "media_url_https": "https://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
            "url": "https://t.co/8JBE6f0Jie",
            "display_url": "pic.twitter.com/8JBE6f0Jie",
            "expanded_url": "https://twitter.com/AmazingRaceCBS/status/1131369650489401344/photo/1",
            "type": "animated_gif",
            "sizes": {
                "thumb": {
                    "w": 150,
                    "h": 150,
                    "resize": "crop"
                },
                "large": {
                    "w": 500,
                    "h": 280,
                    "resize": "fit"
                },
                "medium": {
                    "w": 500,
                    "h": 280,
                    "resize": "fit"
                },
                "small": {
                    "w": 500,
                    "h": 280,
                    "resize": "fit"
                }
            }
        }
    ],
},

```

```

"source_status_id": 1131369650489401344,
"source_status_id_str": "1131369650489401344",
"source_user_id": 68792362,
"source_user_id_str": "68792362",
"video_info": {
  "aspect_ratio": [
    25,
    14
  ],
  "variants": [
    {
      "bitrate": 0,
      "content_type": "video/mp4",
      "url": "https://video.twimg.com/tweet_video/D7NtWWlU8AA4FiR.mp4"
    }
  ]
}
],
},
"metadata": {
  "iso_language_code": "und",
  "result_type": "recent"
},
"source": "<a href=\"http://twitter.com/download/android\" rel=\"nofollow\">Twitter for Android",
"in_reply_to_status_id": null,
"in_reply_to_status_id_str": null,
"in_reply_to_user_id": null,
"in_reply_to_user_id_str": null,
"in_reply_to_screen_name": null,
"user": {
  "id": 400946730,
  "id_str": "400946730",
  "name": "ainembabazi pearl",
  "screen_name": "ainepearl",
  "location": "uganda",
  "description": "",
  "url": null,
  "entities": {
    "description": {
      "urls": []
    }
  }
},
"protected": false,
"followers_count": 14,
"friends_count": 157,
"listed_count": 0,
"created_at": "Sat Oct 29 20:29:17 +0000 2011",

```



```

"favourites_count": 420,
"utc_offset": null,
"time_zone": null,
"geo_enabled": false,
"verified": false,
"statuses_count": 213,
"lang": "en",
"contributors_enabled": false,
"is_translator": false,
"is_translation_enabled": false,
"profile_background_color": "CODEED",
"profile_background_image_url": "http://abs.twimg.com/images/themes/theme1/bg.png",
"profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme1/bg.png",
"profile_background_tile": false,
"profile_image_url": "http://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png",
"profile_image_url_https": "https://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png",
"profile_link_color": "1DA1F2",
"profile_sidebar_border_color": "CODEED",
"profile_sidebar_fill_color": "DDEEF6",
"profile_text_color": "333333",
"profile_use_background_image": true,
"has_extended_profile": false,
"default_profile": true,
"default_profile_image": true,
"following": false,
"follow_request_sent": false,
"notifications": false,
"translator_type": "none"
},
"geo": null,
"coordinates": null,
"place": null,
"contributors": null,
"retweeted_status": {
  "created_at": "Thu May 23 01:21:59 +0000 2019",
  "id": 1131369650489401344,
  "id_str": "1131369650489401344",
  "text": "\ud83e\udd23\ud83e\udd23\ud83e\udd23 #AmazingRace https://t.co/8JBE6f0Jie",
  "truncated": false,
  "entities": {
    "hashtags": [
      {
        "text": "AmazingRace",
        "indices": [
          4,
          16
        ]
      }
    ]
  }
}

```

```

],
"symbols": [],
"user_mentions": [],
"urls": [],
"media": [
{
  "id": 1131368162186817536,
  "id_str": "1131368162186817536",
  "indices": [
    17,
    40
  ],
  "media_url": "http://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
  "media_url_https": "https://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
  "url": "https://t.co/8JBE6f0Jie",
  "display_url": "pic.twitter.com/8JBE6f0Jie",
  "expanded_url": "https://twitter.com/AmazingRaceCBS/status/1131369650489401344/photo/1",
  "type": "photo",
  "sizes": {
    "thumb": {
      "w": 150,
      "h": 150,
      "resize": "crop"
    },
    "large": {
      "w": 500,
      "h": 280,
      "resize": "fit"
    },
    "medium": {
      "w": 500,
      "h": 280,
      "resize": "fit"
    },
    "small": {
      "w": 500,
      "h": 280,
      "resize": "fit"
    }
  }
}
],
},
"extended_entities": {
  "media": [
    {
      "id": 1131368162186817536,
      "id_str": "1131368162186817536",

```

```

"indices": [
  17,
  40
],
"media_url": "http://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
"media_url_https": "https://pbs.twimg.com/tweet_video_thumb/D7NtWWlU8AA4FiR.jpg",
"url": "https://t.co/8JBE6f0Jie",
"display_url": "pic.twitter.com/8JBE6f0Jie",
"expanded_url": "https://twitter.com/AmazingRaceCBS/status/1131369650489401344/photo/1",
"type": "animated_gif",
"sizes": {
  "thumb": {
    "w": 150,
    "h": 150,
    "resize": "crop"
  },
  "large": {
    "w": 500,
    "h": 280,
    "resize": "fit"
  },
  "medium": {
    "w": 500,
    "h": 280,
    "resize": "fit"
  },
  "small": {
    "w": 500,
    "h": 280,
    "resize": "fit"
  }
},
"video_info": {
  "aspect_ratio": [
    25,
    14
  ],
  "variants": [
    {
      "bitrate": 0,
      "content_type": "video/mp4",
      "url": "https://video.twimg.com/tweet_video/D7NtWWlU8AA4FiR.mp4"
    }
  ]
}
],
},

```

```

"metadata": {
  "iso_language_code": "und",
  "result_type": "recent"
},
"source": "<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>",
"in_reply_to_status_id": null,
"in_reply_to_status_id_str": null,
"in_reply_to_user_id": null,
"in_reply_to_user_id_str": null,
"in_reply_to_screen_name": null,
"user": {
  "id": 68792362,
  "id_str": "68792362",
  "name": "The Amazing Race",
  "screen_name": "AmazingRaceCBS",
  "location": "",
  "description": "Watch The #AmazingRace every Wednesday at 8/7c on @CBS and @CBSAllAccess!",
  "url": "http://t.co/fbXOVssTeW",
  "entities": {
    "url": {
      "urls": [
        {
          "url": "http://t.co/fbXOVssTeW",
          "expanded_url": "http://cbs.com/amazingrace",
          "display_url": "cbs.com/amazingrace",
          "indices": [
            0,
            22
          ]
        }
      ]
    }
  },
  "description": {
    "urls": []
  }
},
"protected": false,
"followers_count": 389634,
"friends_count": 94,
"listed_count": 1589,
"created_at": "Tue Aug 25 20:01:42 +0000 2009",
"favourites_count": 1157,
"utc_offset": null,
"time_zone": null,
"geo_enabled": false,
"verified": true,
"statuses_count": 5999,
"lang": "en",

```

```

    "contributors_enabled": false,
    "is_translator": false,
    "is_translation_enabled": false,
    "profile_background_color": "161924",
    "profile_background_image_url": "http://abs.twimg.com/images/themes/theme9/bg.gif",
    "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme9/bg.gif",
    "profile_background_tile": false,
    "profile_image_url": "http://pbs.twimg.com/profile_images/1101657616814563328/IBIpXUyd_norma",
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/1101657616814563328/IBIpXU",
    "profile_banner_url": "https://pbs.twimg.com/profile_banners/68792362/1558037372",
    "profile_link_color": "2FC2EF",
    "profile_sidebar_border_color": "FFFFFF",
    "profile_sidebar_fill_color": "FFFCF",
    "profile_text_color": "666666",
    "profile_use_background_image": true,
    "has_extended_profile": false,
    "default_profile": false,
    "default_profile_image": false,
    "following": false,
    "follow_request_sent": false,
    "notifications": false,
    "translator_type": "none"
  },
  "geo": null,
  "coordinates": null,
  "place": null,
  "contributors": null,
  "is_quote_status": false,
  "retweet_count": 15,
  "favorite_count": 111,
  "favorited": false,
  "retweeted": false,
  "possibly_sensitive": false,
  "lang": "und"
},
  "is_quote_status": false,
  "retweet_count": 15,
  "favorite_count": 0,
  "favorited": false,
  "retweeted": false,
  "possibly_sensitive": false,
  "lang": "und"
}

```

```

In [7]: for i in range(10):
        print()
        print(statuses[i]['text'])

```

```
print('Favorites: ', statuses[i]['favorite_count'])
print('Retweets: ', statuses[i]['retweet_count'])
```

RT @AmazingRaceCBS: #AmazingRace <https://t.co/8JBE6f0Jie>
Favorites: 0
Retweets: 15

RT @MattOMalley: Two weeks from tonight: Join @BretLaBelle & me at Kellehers (1410 Centre S
Favorites: 0
Retweets: 3

The one nice thing about the head-to-head is that nobody has to be alone when they get eliminated
Favorites: 1
Retweets: 0

RT @AmazingRaceCBS: Fun Detours, thrilling challenges, and intense head-to-head battles. Preparation
Favorites: 0
Retweets: 35

RT @AmazingRaceCBS: \$1 million is still up for grabs, and the remaining teams are giving it their all
Favorites: 0
Retweets: 18

RT @Alexafan23: @RachelEReilly @CBSTweet @RachelEReilly @ElissaReillyS #AmazingRace #TAR31 <http://t.co/5R0whISjQG>
Favorites: 0
Retweets: 1

RT @AmazingRaceCBS: Dance break. #AmazingRace <https://t.co/5R0whISjQG>
Favorites: 0
Retweets: 12

Wow, these head-to-head portraits flatter nobody. #AmazingRace
Favorites: 0
Retweets: 0

RT @PhilKeoghan: #BurjKhalifa not only the world's tallest building; tallest freestanding structure
Favorites: 0
Retweets: 6

Making gay people travel to Uganda is not cool, #amazingrace
Favorites: 0
Retweets: 0

4 Extracting text, screen names, and hashtags from tweets

```
In [8]: status_texts = [ status['text']
                        for status in statuses ]

screen_names = [ user_mention['screen_name']
                 for status in statuses
                 for user_mention in status['entities']['user_mentions'] ]

hashtags = [ hashtag['text']
             for status in statuses
             for hashtag in status['entities']['hashtags'] ]

# Compute a collection of all words from all tweets
words = [ w
          for t in status_texts
          for w in t.split() ]

# Explore the first 5 items for each...

print(json.dumps(status_texts[0:5], indent=1))
print(json.dumps(screen_names[0:5], indent=1) )
print(json.dumps(hashtags[0:5], indent=1))
print(json.dumps(words[0:5], indent=1))

[
  "RT @AmazingRaceCBS: \ud83e\udd23\ud83e\udd23\ud83e\udd23 #AmazingRace https://t.co/8JBE6f0Ji",
  "RT @MattOMalley: Two weeks from tonight: Join @BretLaBelle & me at Kellehers (1410 Centre",
  "The one nice thing about the head-to-head is that nobody has to be alone when they get elimin",
  "RT @AmazingRaceCBS: Fun Detours, thrilling challenges, and intense head-to-head battles. Prep",
  "RT @AmazingRaceCBS: $1 million is still up for grabs, and the remaining teams are giving it t",
]

[
  "AmazingRaceCBS",
  "MattOMalley",
  "BretLaBelle",
  "AmazingRaceCBS",
  "AmazingRaceCBS"
]

[
  "AmazingRace",
  "AmazingRace",
  "AmazingRace",
  "AmazingRace",
  "TAR31"
]

[
  "RT",
```

```

"@AmazingRaceCBS:",
"\ud83e\udd23\ud83e\udd23\ud83e\udd23",
"#AmazingRace",
"https://t.co/8JBE6f0Jie"
]

```

5 Create a basic frequency distribution from the words in tweets

```
In [9]: from collections import Counter
```

```

for item in [words, screen_names, hashtags]:
    c = Counter(item)
    print(c.most_common()[:10]) # top 10
    print()

```

```

[('#AmazingRace', 382), ('the', 191), ('to', 161), ('and', 155), ('RT', 152), ('a', 146), ('is', 141), ('korey', 139), ('tyler', 138), ('koreykuhl', 137)]

[('tyleroakley', 57), ('AmazingRaceCBS', 51), ('PhilKeoghan', 29), ('koreykuhl', 22), ('Rachel', 21), ('TeamTylerandKorey', 19), ('TAR31', 14), ('TAR', 13), ('AmazingRace', 12), ('amazingrace', 11)]

[('AmazingRace', 421), ('amazingrace', 19), ('TAR31', 14), ('TeamTylerandKorey', 13), ('TAR', 9), ('tyler', 8), ('tyleroakley', 7), ('koreykuhl', 6), ('korey', 5), ('is', 4)]

```

6 Using prettytable to display tuples in a nice tabular format

```
In [11]: from prettytable import PrettyTable
```

```

for label, data in (('Word', words),
                   ('Screen Name', screen_names),
                   ('Hashtag', hashtags)):
    pt = PrettyTable(field_names=[label, 'Count'])
    c = Counter(data)
    [ pt.add_row(kv) for kv in c.most_common()[:10] ]
    pt.align[label], pt.align['Count'] = 'l', 'r' # Set column alignment
    print(pt)

```

```
Out[11]: [None, None, None, None, None, None, None, None, None, None]
```

```

+-----+-----+
| Word   | Count |
+-----+-----+
| #AmazingRace | 382 |
| the      | 191 |
| to       | 161 |
| and      | 155 |
| RT       | 152 |

```


a	146	
is	105	
I	89	
on	85	
of	76	
+-----+-----+		

Out[11]: [None, None, None, None, None, None, None, None, None, None]

+-----+-----+		
Screen Name	Count	
+-----+-----+		
tyleroakley	57	
AmazingRaceCBS	51	
PhilKeoghan	29	
koreykuhl	22	
RachelEReilly	13	
JanellePierzina	10	
britney_haynes	10	
BretLaBelle	9	
BB_timc	7	
HaymakerHattie	6	
+-----+-----+		

Out[11]: [None, None, None, None, None, None, None, None, None, None]

+-----+-----+		
Hashtag	Count	
+-----+-----+		
AmazingRace	421	
amazingrace	19	
TAR31	14	
TeamTylerandKorey	13	
TAR	9	
Survivor	7	
RacersRecap	7	
TeamTylerAndKorey	6	
BigBrother	6	
Head2Head	5	
+-----+-----+		

7 Calculating lexical diversity for tweets

```
In [12]: # A function for computing lexical diversity
def lexical_diversity(tokens):
```

```

        return len(set(tokens))/len(tokens)

# A function for computing the average number of words per tweet
def average_words(statuses):
    total_words = sum([ len(s.split()) for s in statuses ])
    return total_words/len(statuses)

print(lexical_diversity(words))
print(lexical_diversity(screen_names))
print(lexical_diversity(hashtags))
print(average_words(status_texts))

0.29327708219923787
0.28106508875739644
0.11525974025974026
14.43614931237721

```

8 Find most popular retweets

```

In [13]: retweets = [
    # Store out a tuple of these three values ...
    (status['retweet_count'],
     status['retweeted_status']['user']['screen_name'],
     status['retweeted_status']['id'],
     status['text'])

    # ... for each status ...
    for status in statuses

    # ... so long as the status meets this condition.
    if 'retweeted_status' in status.keys()
]

# Slice off the first 5 from the sorted results and display each item in the tuple

pt = PrettyTable(field_names=['Count', 'Screen Name', 'Tweet ID', 'Text'])
[ pt.add_row(row) for row in sorted(retweets, reverse=True)[:5] ]
pt.max_width['Text'] = 50
pt.align= 'l'
print(pt)

```

Out[13]: [None, None, None, None, None]

Count	Screen Name	Tweet ID	Text
149	tyleroakley	1131349357075111936	RT @tyleroakley: tonight on #AmazingRace, queer

			contestants were sent to locations where being
			queer is a crime. being there with protectio
149	tyleroakley	1131349357075111936	RT @tyleroakley: tonight on #AmazingRace, queer
			contestants were sent to locations where being
			queer is a crime. being there with protectio
149	tyleroakley	1131349357075111936	RT @tyleroakley: tonight on #AmazingRace, queer
			contestants were sent to locations where being
			queer is a crime. being there with protectio
149	tyleroakley	1131349357075111936	RT @tyleroakley: tonight on #AmazingRace, queer
			contestants were sent to locations where being
			queer is a crime. being there with protectio
149	tyleroakley	1131349357075111936	RT @tyleroakley: tonight on #AmazingRace, queer
			contestants were sent to locations where being
			queer is a crime. being there with protectio

9 Looking up users who have retweeted a status

```
In [28]: # Get the original tweet id for a tweet from its retweeted_status node
# and insert it here
```

```
_retweets = twitter_api.statuses.retweets(id=1131349357075111936)
print([r['user']['screen_name'] for r in _retweets])
```

```
['itstylernicole', 'mattjuniper', 'DarkSkyDestiny', 'BBvsSurvivor', 'd_m_elms', 'Pamela_Jocelyn']
```

10 Plotting frequencies of words

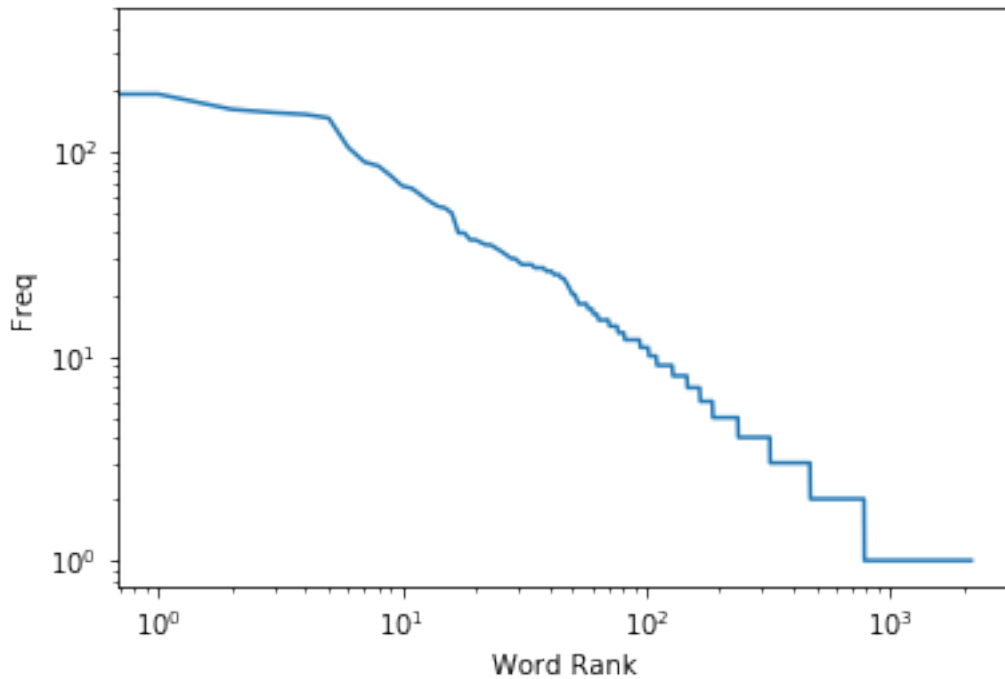
```
In [29]: import matplotlib.pyplot as plt
%matplotlib inline
word_counts = sorted(Counter(words).values(), reverse=True)

plt.loglog(word_counts)
plt.ylabel("Freq")
plt.xlabel("Word Rank")
```

```
Out[29]: [<matplotlib.lines.Line2D at 0x178ff0c5668>]
```

```
Out[29]: Text(0,0.5,'Freq')
```

```
Out[29]: Text(0.5,0,'Word Rank')
```



11 Generating histograms of words, screen names, and hashtags

```
In [30]: for label, data in (('Words', words),
                             ('Screen Names', screen_names),
                             ('Hashtags', hashtags)):

    # Build a frequency map for each set of data
    # and plot the values
    c = Counter(data)
    plt.hist(list(c.values()))

    # Add a title and y-label ...
    plt.title(label)
    plt.ylabel("Number of items in bin")
    plt.xlabel("Bins (number of times an item appeared)")

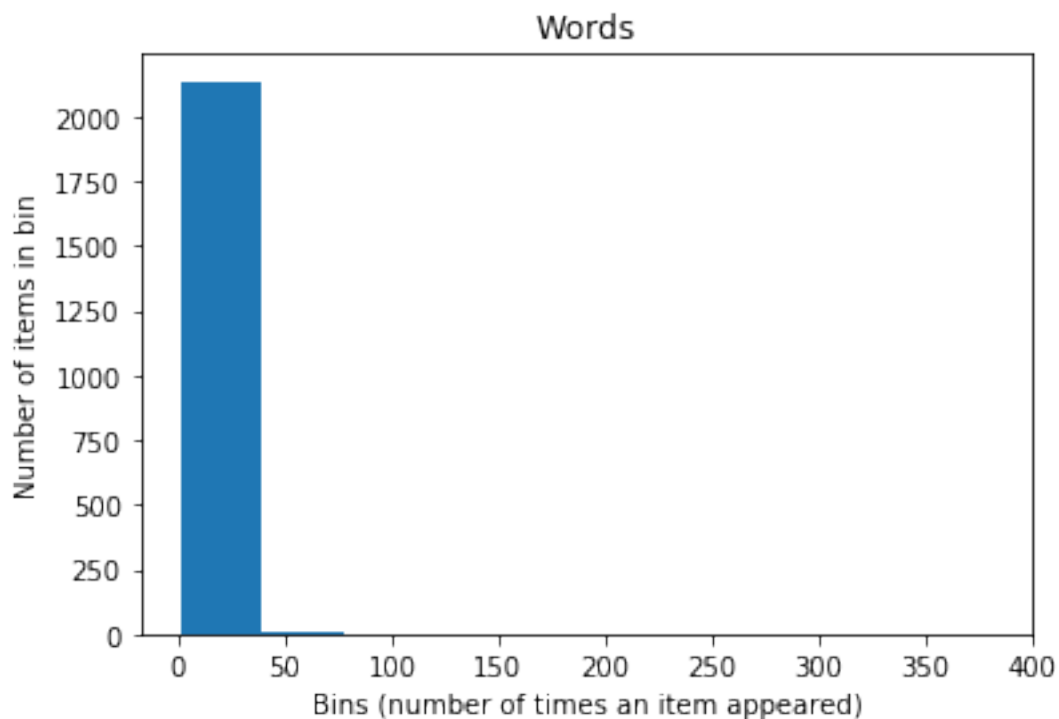
    # ... and display as a new figure
    plt.figure()

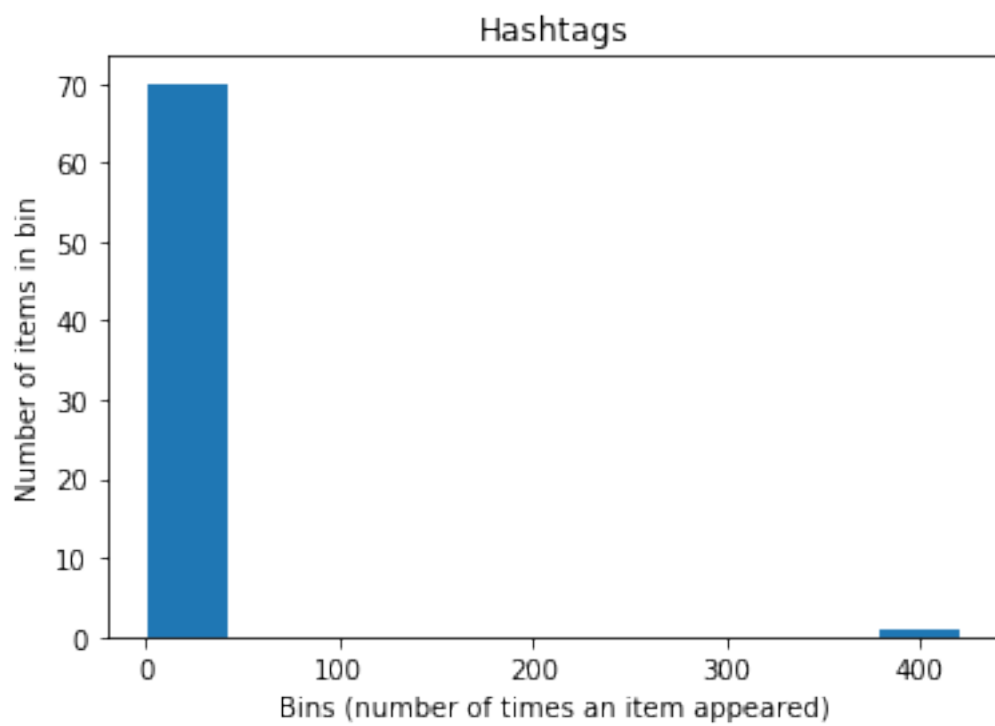
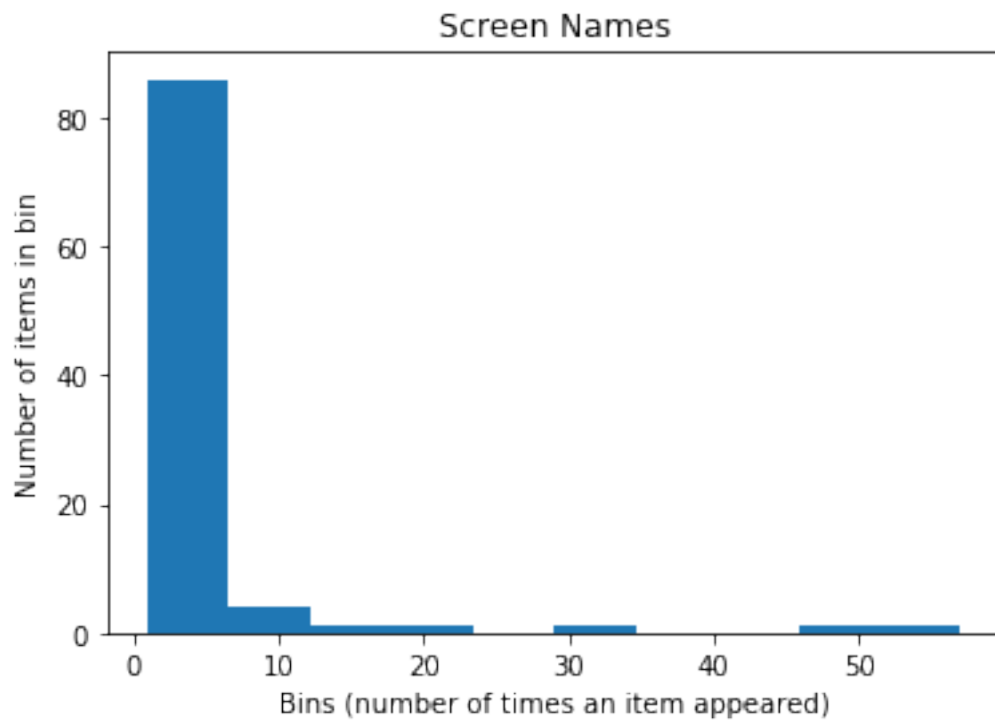
Out[30]: (array([2.136e+03, 1.000e+01, 3.000e+00, 2.000e+00, 3.000e+00, 0.000e+00,
                  0.000e+00, 0.000e+00, 0.000e+00, 1.000e+00]),
          array([ 1. , 39.1, 77.2, 115.3, 153.4, 191.5, 229.6, 267.7, 305.8,
                  343.9, 382. ]),
          <a list of 10 Patch objects>)
```

```

Out[30]: Text(0.5,1,'Words')
Out[30]: Text(0,0.5,'Number of items in bin')
Out[30]: Text(0.5,0,'Bins (number of times an item appeared)')
Out[30]: <Figure size 432x288 with 0 Axes>
Out[30]: (array([86.,  4.,  1.,  1.,  0.,  1.,  0.,  0.,  1.,  1.]),
          array([ 1. ,  6.6, 12.2, 17.8, 23.4, 29. , 34.6, 40.2, 45.8, 51.4, 57. ])),
          <a list of 10 Patch objects>)
Out[30]: Text(0.5,1,'Screen Names')
Out[30]: Text(0,0.5,'Number of items in bin')
Out[30]: Text(0.5,0,'Bins (number of times an item appeared)')
Out[30]: <Figure size 432x288 with 0 Axes>
Out[30]: (array([70.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  1.]),
          array([ 1.,  43.,  85., 127., 169., 211., 253., 295., 337., 379., 421.])),
          <a list of 10 Patch objects>)
Out[30]: Text(0.5,1,'Hashtags')
Out[30]: Text(0,0.5,'Number of items in bin')
Out[30]: Text(0.5,0,'Bins (number of times an item appeared)')
Out[30]: <Figure size 432x288 with 0 Axes>

```





<Figure size 432x288 with 0 Axes>

12 Generating a histogram of retweet counts

```
In [16]: # Using underscores while unpacking values in  
# a tuple is idiomatic for discarding them
```

```
counts = [count for count, _, _, _ in retweets]
```

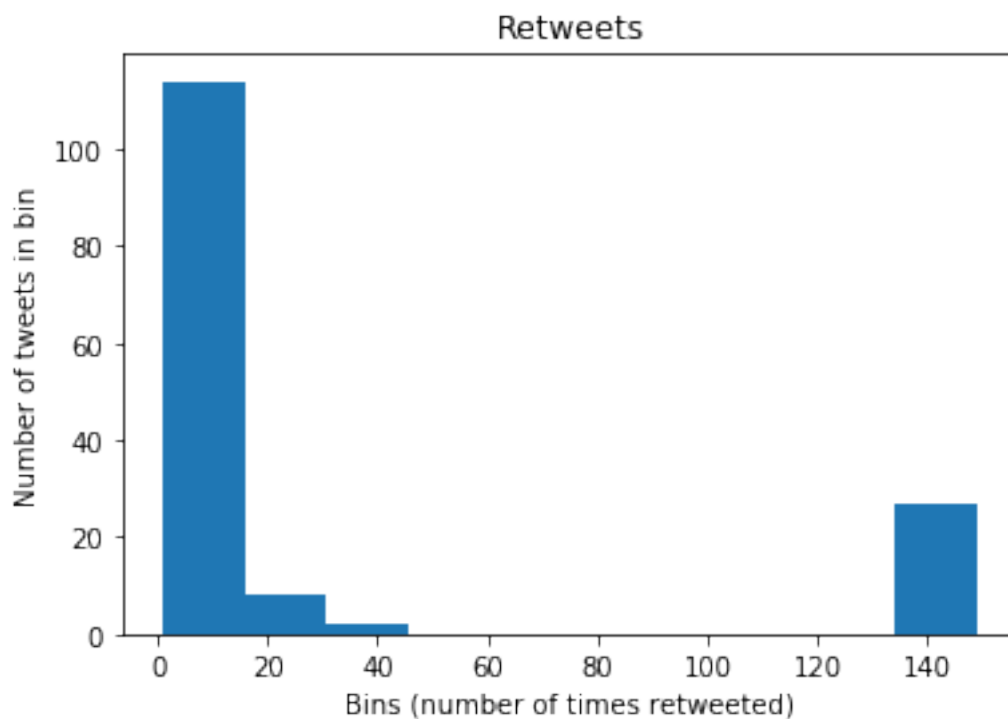
```
plt.hist(counts)  
plt.title('Retweets')  
plt.xlabel('Bins (number of times retweeted)')  
plt.ylabel('Number of tweets in bin')
```

```
Out[16]: (array([114.,  8.,  2.,  0.,  0.,  0.,  0.,  0.,  0., 27.]),  
array([ 1., 15.8, 30.6, 45.4, 60.2, 75., 89.8, 104.6, 119.4,  
134.2, 149. ]),  
<a list of 10 Patch objects>)
```

```
Out[16]: Text(0.5,1,'Retweets')
```

```
Out[16]: Text(0.5,0,'Bins (number of times retweeted)')
```

```
Out[16]: Text(0,0.5,'Number of tweets in bin')
```



13 Sentiment

```
In [17]: # pip install nltk
import nltk
nltk.download('vader_lexicon')

import numpy as np
from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data] C:\Users\emyhr\AppData\Roaming\nltk_data...
```

```
Out[17]: True
```

```
C:\Users\emyhr\Anaconda3\lib\site-packages\nltk\twitter\__init__.py:20: UserWarning: The twython
warnings.warn("The twython library has not been installed. "
```

```
In [18]: twitter_stream = twitter.TwitterStream(auth=auth)
iterator = twitter_stream.statuses.sample()
```

```
In [19]: tweets = []
for tweet in iterator:
    try:
        if tweet['lang'] == 'en':
            tweets.append(tweet)
    except:
        pass
if len(tweets) == 100:
    break
```

```
In [20]: analyzer = SentimentIntensityAnalyzer()
```

```
In [21]: analyzer.polarity_scores('Hello')
```

```
Out[21]: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

```
In [23]: scores = np.zeros(len(tweets))

for i, t in enumerate(tweets):
    # Extract the text portion of the tweet
    text = t['text']

    # Measure the polarity of the tweet
    polarity = analyzer.polarity_scores(text)
```



```

        # Store the normalized, weighted composite score
        scores[i] = polarity['compound']

In [24]: most_positive = np.argmax(scores)
        most_negative = np.argmin(scores)

In [25]: print('{0:6.3f} : "{1}"'.format(scores[most_positive], tweets[most_positive]['text']))

0.891 : "@kingkanono Im all for freeing the nip and living your life. Honestly I laughed , lil

In [26]: print('{0:6.3f} : "{1}"'.format(scores[most_negative], tweets[most_negative]['text']))

-0.932 : "RT @MrMouthPiece_: hold up... sirius blacks death was sad asf BUT professor snapes d

```

14 Part Two

```

In [1]: from collections import Counter
        import tweepy
        import pandas as pd
        import time
        import sys
        import os
        import numpy as np
        import emoji
        import emot as emo

In [2]: # Write out Tweets in case they are needed later.
        df = pd.read_csv('data/amazingrace/amazingrace_400.csv',encoding='utf-8')

In [3]: df

Out[3]:
   timestamp                sn \
0  2019-05-23 03:59:58      MrUNotReady
1  2019-05-23 03:59:49      kevintoy888
2  2019-05-23 03:59:06    Pinkjellybean3
3  2019-05-23 03:58:49    iowanliveintexa
4  2019-05-23 03:58:34    Loveleelee_xoxo
5  2019-05-23 03:58:24    Loveleelee_xoxo
6  2019-05-23 03:58:22    abdulmannanbutt
7  2019-05-23 03:58:08    Loveleelee_xoxo
8  2019-05-23 03:57:56    Loveleelee_xoxo
9  2019-05-23 03:57:55          d0nnieeeee
10 2019-05-23 03:57:37      JayBarnes77
11 2019-05-23 03:57:21      maddiea_xo
12 2019-05-23 03:57:20        richksu
13 2019-05-23 03:57:11    iowanliveintexa

```

14	2019-05-23 03:57:09	Luvzparanoid
15	2019-05-23 03:56:27	ShawnReynolds_
16	2019-05-23 03:56:26	cindraloo
17	2019-05-23 03:56:18	jordychamberlin
18	2019-05-23 03:55:49	Luvzparanoid
19	2019-05-23 03:55:37	ashrose31
20	2019-05-23 03:55:03	kailakay_17
21	2019-05-23 03:54:34	marie81411
22	2019-05-23 03:54:22	SanJoYouKnow
23	2019-05-23 03:54:21	sjn1922
24	2019-05-23 03:54:19	cindraloo
25	2019-05-23 03:53:36	Phil_okk
26	2019-05-23 03:53:14	rayraycja
27	2019-05-23 03:52:38	geenahd
28	2019-05-23 03:52:29	sjn1922
29	2019-05-23 03:51:59	KiranSanthanam
..
365	2019-05-23 02:00:47	realityTVstud
366	2019-05-23 02:00:45	Survality
367	2019-05-23 02:00:44	KatieSecrets
368	2019-05-23 02:00:43	JanLapatha
369	2019-05-23 02:00:34	CalebWH25
370	2019-05-23 02:00:11	HaymakerHattie
371	2019-05-23 02:00:10	_cfoxx90
372	2019-05-23 02:00:09	ALovelierGirl
373	2019-05-23 02:00:06	katm0512
374	2019-05-23 01:59:57	EvermanJeffery
375	2019-05-23 01:59:56	Srahspillz
376	2019-05-23 01:59:54	fjdinardo
377	2019-05-23 01:59:50	RachelEReilly
378	2019-05-23 01:59:48	KristelleJoySa4
379	2019-05-23 01:59:46	itswessums
380	2019-05-23 01:59:45	ericaeff
381	2019-05-23 01:59:36	Divine_Angel89
382	2019-05-23 01:59:30	katm0512
383	2019-05-23 01:59:28	andreafan111
384	2019-05-23 01:59:26	Holly_Homa
385	2019-05-23 01:59:11	lordsntee
386	2019-05-23 01:59:09	tayloroliviq
387	2019-05-23 01:59:08	FightYouForIt1
388	2019-05-23 01:59:06	CulturalRights1
389	2019-05-23 01:58:55	musicsmith14
390	2019-05-23 01:58:54	planetadenegoci
391	2019-05-23 01:58:46	hiliejozz
392	2019-05-23 01:58:37	ajgraves14
393	2019-05-23 01:58:31	delsina363
394	2019-05-23 01:58:30	doco86

	text	EST
0	Yeaaaaah!!! @RachelEReilly...You Are Teaching ...	2019-05-22 22:59:58
1	Why do these reality shows keep saving the Rei...	2019-05-22 22:59:49
2	RT @AmazingRaceCBS: Dubai. Dinosaurs. Glowing ...	2019-05-22 22:59:06
3	RT @PhilKeoghan: A picture says 1,000 words ...	2019-05-22 22:58:49
4	RT @Space_Nerd_at_C: Hey Phil, put US out of o...	2019-05-22 22:58:34
5	RT @DrkLovely1: I TEAM FUN!! #AmazingRace h...	2019-05-22 22:58:24
6	A memorable group foto with Ceo. perennial_clo...	2019-05-22 22:58:22
7	RT @FightYouForIt1: Corrine and Eliza after se...	2019-05-22 22:58:08
8	RT @nynngaile: Phil NOOOOO why a non eliminatio...	2019-05-22 22:57:56
9	RT @tyleroakley: tonight on #AmazingRace, quee...	2019-05-22 22:57:55
10	Rachel and Elissa finish last\r\n*n*Non-Eliminat...	2019-05-22 22:57:37
11	RT @SrslyChris: Elissa getting fed up with Rac...	2019-05-22 22:57:21
12	Rachel does everything good- #TAR #AmazingRace	2019-05-22 22:57:20
13	RT @PhilKeoghan: Don't go anywhere ... brace y...	2019-05-22 22:57:11
14	RT @CalebWH25: I love Rachel and Elissas dyna...	2019-05-22 22:57:09
15	Phil waiting for the last team to arrive #Amaz...	2019-05-22 22:56:27
16	RT @chrisrhahn: Excited for a double dose of #...	2019-05-22 22:56:26
17	RT @SrslyChris: Elissa getting fed up with Rac...	2019-05-22 22:56:18
18	RT @Steffiesays1: Elissa's face while Rachel i...	2019-05-22 22:55:49
19	Team fun is the best team coming from the blin...	2019-05-22 22:55:37
20	RT @tyleroakley: tonight on #AmazingRace, quee...	2019-05-22 22:55:03
21	@tjlvain #MTVChallenge and @PhilKeoghan from #...	2019-05-22 22:54:34
22	Rach, please stop. @ElissaReillyS saying wha...	2019-05-22 22:54:22
23	@AmazingRaceCBS Rach!!! THERE IS NO CRYING IN ...	2019-05-22 22:54:21
24	#AmazingRace	2019-05-22 22:54:19
25	#Uganda thru #AmazingRace 's eyes #Dubai t...	2019-05-22 22:53:36
26	Who knew that @PhilKeoghan had such a sick flo...	2019-05-22 22:53:14
27	RT @tyleroakley: tonight on #AmazingRace, quee...	2019-05-22 22:52:38
28	@AmazingRaceCBS This is what Team Fun is all a...	2019-05-22 22:52:29
29	RT @tyleroakley: tonight on #AmazingRace, quee...	2019-05-22 22:51:59
..
365	@Jessica65481190 @mainlybravo @jesusjugs2 @Jes...	2019-05-22 21:00:47
366	Favorite team thats representing #AmazingRace...	2019-05-22 21:00:45
367	Rachel just shut your big mouth for once #Am...	2019-05-22 21:00:44
368	Janelle & Britney deserves better\r\n#Amaz...	2019-05-22 21:00:43
369	The preview for next weeks #AmazingRace has m...	2019-05-22 21:00:34
370	Who was less impressed this week: Sansa Stark ...	2019-05-22 21:00:11
371	Showdown? #AmazingRace	2019-05-22 21:00:10
372	I need Becca and Floyd to get their own realit...	2019-05-22 21:00:09
373	Routing for Nichole and Victor #AmazingRace	2019-05-22 21:00:06
374	Big congrats on the W #TeamFun in Dubai! #Team...	2019-05-22 20:59:57
375	Sad my big brother girls are being split up! #...	2019-05-22 20:59:56
376	The #AmazingRace is going to Switzerland next ...	2019-05-22 20:59:54
377	RT @BB_timc: Now we need the mom squad on a BB...	2019-05-22 20:59:50
378	RT @tyleroakley: tonight on #AmazingRace, quee...	2019-05-22 20:59:48
379	Im...upset #AmazingRace	2019-05-22 20:59:46
380	The #AmazingRace just got a little less entert...	2019-05-22 20:59:45

```

381 I'm sorry but if you don't like team fun, you ... 2019-05-22 20:59:36
382 Becca and Floyd are super annoying af #Amazing... 2019-05-22 20:59:30
383 RT @KiranSanthanam: Thoughts on Janelle and Br... 2019-05-22 20:59:28
384 Im so disappointed My Favorite team. @Janell... 2019-05-22 20:59:26
385      #AmazingRace Becca and Floyd are so much fun 2019-05-22 20:59:11
386 So proud of @tyleroakley and @koreykuhl on the... 2019-05-22 20:59:09
387 Having to watch a @CBSBigBrother team have to ... 2019-05-22 20:59:08
388 THANK YOU @PhilKeoghan AND @AmazingRaceCBS for... 2019-05-22 20:59:06
389 At least with this head to head, the order of ... 2019-05-22 20:58:55
390 #boatdrinks #vacationmodeon #cruisetravel #dri... 2019-05-22 20:58:54
391 @AmazingRaceCBS did yall atleast get a chance... 2019-05-22 20:58:46
392 Noooooo! Regardless, Im so proud of my 2 of 3... 2019-05-22 20:58:37
393 RT @BB_timc: Now we need the mom squad on a BB... 2019-05-22 20:58:31
394 Soooo glad u made it.!!! @RachelEReilly goo... 2019-05-22 20:58:30

```

[395 rows x 4 columns]

```

In [4]: # Create a list of the unique usernames in order to see which users we need to retrieve
allNames = list(df['sn'].unique())

```

Now having a list of all of the users we would like to include in our community, we can retrieve additional necessary information such as who their are following in order to build our social graph.

```

In [6]: allNames

```

```

Out[6]: ['MrUNotReady',
        'kevintoy888',
        'Pinkjellybean3',
        'iowanliveintexa',
        'Loveleelee_xoxo',
        'abdulmannanbutt',
        'dOnnieeeee',
        'JayBarnes77',
        'maddiea_xo',
        'richksu',
        'Luvzparanoid',
        'ShawnReynolds_',
        'cindraloo',
        'jordychamberlin',
        'ashrose31',
        'kailakay_17',
        'marie81411',
        'SanJoYouKnow',
        'sjn1922',
        'Phil_okk',
        'rayraycja',
        'geenahd',
        'KiranSanthanam',

```

'janet_rooker',
'alissaMlewis',
'belrosegr',
'Alicestar13',
'PixelGamer_',
'JessicaRunion2',
'SaltwaterVal',
'MarisolT1971',
'Jordanzahhutt',
'4realseottraffic',
'westcoast_246',
'thearashow',
'hive_babes',
'SuperCoolCarly',
'riikabob',
'BrianHiam',
'heybrittrob',
'Spyglas08492298',
'mssrfcali',
'thrllsnsplls',
'realityTVstud',
'justinlosacco',
'ObiWOKE_Kenobi',
'peiffer_tanner',
'OCannaday32',
'XoBabyDollXo',
'oshawotter',
'Green4Lori',
'LeslieGzz',
'jojoanello',
'Team_DMP',
'kam_isforever',
'AlexSeifer',
'ErinEMalone',
'CaitlinNetzer',
'TVGal_Chele',
'TravelWithJared',
'TheNewsIntern1',
'RedRobz',
'kierashoyos',
'Quest_fk',
'SarahQuider',
'dryedmangoez',
'DebiJune2',
'nerdgirl67',
'randyknup',
'majingir',
'JwatsJackie',

'sam_barton123',
'McLovin_0228',
'globalboliek',
'feministbeetch',
'TheShonBailey',
'IGIF_',
'evett_chiu',
'gfisherz',
'TeamElissa1',
'FelixsFosterSis',
'JayHartDMP',
'sassy_mel',
'Reddzz730',
'Theater_Poverty',
'justoofthings',
'TopsyMic',
'itssavannahjade',
'SprinkleofSchoee',
'mattjuniper',
'RealityRecaps',
'morfair71',
'SolaceBiersack',
'DarkSkyDestiny',
'joparke',
'wtfjag',
'DrWillHatch',
'AnneEllerson',
'Pamela_Jocelyn',
'JoelBBCAN',
'CowieKaunweigh',
'ainepearl',
'BurgerSlamMann',
'HaymakerHattie',
'ElissaReillyS',
'JenniferCole224',
'erniebufflo',
'XmanIce',
'elana_danielle',
'MyLifeAs_Ashley',
'kyrababy27',
'Hayven_C',
'M_A_Schilling',
'EARTHDOGGBB',
'OGBigSprite',
'AprilinAlberta4',
'sneppydooda',
'MattSchmidtLA',
'KathrynNiforos',

'CoCoSpice66',
'JesseGarcia321',
'Cc87962961',
'catmandu13',
'steph9222',
'mike_913',
'bbqueenaleigh',
'TvWizards',
'rglick13',
'AllanBSportsFan',
'smartcorgi',
'juscallsizzles',
'ZacccIsHere',
'yoshifan331',
'BrennanSwain',
'juliaakathrynnn',
'ShawnPooley',
'amysposito',
'maybeasa',
'coletto',
'BiglowRoselia',
'Bentley20000',
'jtobin33',
'catmom007',
'CindyLouWho116',
'MeLovesTV',
'ChristyClark814',
'AndrewSimanski',
'CrazyPackersFan',
'FabsCulture',
'AshsDreamboard',
'colkmiren',
'wendyispresent',
'SanMakinami',
'Tremble26',
'kale_chipps',
'markpillls',
'stephensroczyk',
'djlabelleklein',
'cassidythomp94',
'JustinClassic1',
'theslaycademy',
'cheylouis',
'julyheat7',
'gingabear31',
'Stephyo4',
'GamerGirl12330',
'PL83M',

'Glenn_Weisinger',
'ASweetHeart123',
'lexiecenter',
'RachelEReilly',
'KeithZingraf',
'CrossLy_er',
'heyitsmyopinion',
'TaraKfoury',
'absmile28',
'cloudyyllaurenn',
'lopezgovlaw',
'ShanMR42',
'BBfan4life4sure',
'LilMel13',
'fpaez2178',
'TrevonIi',
'AriFerarri',
'mashmartin92',
'Ryan__Kaiser',
'Natesmb64',
'mickeyshy12',
'JohnG500',
'deidra_dwallace',
'NeffStartups',
'KatelynKozma',
'MattOMalley',
'SurvivingTribal',
'Vaden83',
'Survivor_on_DVD',
'OminousTomato',
'PhilKeoghan',
'DoraTheShutIn',
'Srahspillz',
'TamGarrettCasey',
'Seapoeschl',
'alyleanne01',
'Mariat3',
'rallyssa11',
'LC3_TDSB',
'GamerKevin94',
'InTheDrunkTank',
'IconicQueenAri',
'br_infinity_ke',
'Divine_Angel89',
'Tweettweetter',
'planetadenegoci',
'BB_timc',
'Survality',