Debajo, se importan los tweets y se les agrega un campo nuevo 'longitud'

In [1]:
```python
1  import pandas as pd
2  import numpy as np
3  import seaborn as sns
4  import matplotlib.pyplot as plt
5  #import geopy as gp
6  from geopy.geocoders import Nominatim
7  from geopy.extra.rate_limiter import RateLimiter
8  import datetime
9
10 %pylab inline
11
12 tweets = pd.read_csv('train.csv', encoding='latin-1')
13 tweets['longitud'] = tweets['text'].str.len()
14 tweets.head(10)
```

Populating the interactive namespace from numpy and matplotlib

Out[1]:

	id	keyword	location	text	target	longitud
0	1	NaN	NaN	Our Deeds are the Reason of this #earthquake M...	1	69
1	4	NaN	NaN	Forest fire near La Ronge Sask. Canada	1	38
2	5	NaN	NaN	All residents asked to 'shelter in place' are ...	1	133
3	6	NaN	NaN	13,000 people receive #wildfires evacuation or...	1	65
4	7	NaN	NaN	Just got sent this photo from Ruby #Alaska as ...	1	88
5	8	NaN	NaN	#RockyFire Update => California Hwy. 20 closed...	1	110
6	10	NaN	NaN	#flood #disaster Heavy rain causes flash flood...	1	95
7	13	NaN	NaN	I'm on top of the hill and I can see a fire in...	1	59
8	14	NaN	NaN	There's an emergency evacuation happening now ...	1	79
9	15	NaN	NaN	I'm afraid that the tornado is coming to our a...	1	52

Ahora tomamos los tweets reales, como muestra para demostrar el desarrollo del algoritmo y los problemas encontrados Luego se hará el mismo procesamiento pero para todos los tweets cuya "location" no sea nula

In [2]:
```python
1  true_tweets = tweets[tweets.target==1]
```

In [3]:
```
1  true_tweets.count()
```

Out[3]:
```
id          3271
keyword     3229
location    2196
text        3271
target      3271
longitud    3271
dtype: int64
```

Ahora, de esos tweets reales, se toman los que tengan datos de localización

In [4]:
```
1  true_tweets_with_loc = true_tweets[true_tweets.location.isnull()==False]
```

In [5]:
```
1  true_tweets_with_loc.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2196 entries, 31 to 7572
Data columns (total 6 columns):
id          2196 non-null int64
keyword     2196 non-null object
location    2196 non-null object
text        2196 non-null object
target      2196 non-null int64
longitud    2196 non-null int64
dtypes: int64(3), object(3)
memory usage: 120.1+ KB
```

In [6]:
```
1  true_tweets_with_loc_sub = true_tweets_with_loc.head(10)
2  geolocator = Nominatim(user_agent = "TP1 1C2020 7506 FIUBA - Barbijo_de_un
3  geocode = RateLimiter(geolocator.geocode, min_delay_seconds = 0.1)
4  true_tweets_with_loc_sub = true_tweets_with_loc_sub.assign(geolocation = t
5                                                      apply(geocode))
```

In [7]:   1   true_tweets_with_loc_sub

Out[7]:

	id	keyword	location	text	target	longitud	geolocation
31	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	(Birmingham, West Midlands Combined Authority,...
33	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	(Africa, (11.5024338, 17.7578122))
37	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	(環球大廈 World-Wide House, 中環 Central, 西環 Sai Wan...
46	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	(Greensboro, Guilford County, North Carolina, ...
50	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	(Sheffield Township, Lorain County, Ohio, Unit...
51	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	(भारत - India, (22.3511148, 78.6677428))
53	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	(Anaheim, Orange County, California, United St...
55	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	(United States, (39.7837304, -100.4458825))
56	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	(South Africa, (-28.8166236, 24.991639))
59	83	ablaze	Edmonton, Alberta - Treaty 6	How the West was burned: Thousands of wildfire...	1	138	None

In [8]:
```
1 true_tweets_with_loc_sub = true_tweets_with_loc_sub[true_tweets_with_loc_s
2 true_tweets_with_loc_sub
```

Out[8]:

	id	keyword	location	text	target	longitud	geolocation
31	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	(Birmingham, West Midlands Combined Authority,...
33	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	(Africa, (11.5024338, 17.7578122))
37	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	(環球大廈 World-Wide House, 中環 Central, 西環 Sai Wan...
46	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	(Greensboro, Guilford County, North Carolina, ...
50	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	(Sheffield Township, Lorain County, Ohio, Unit...
51	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	(भारत - India, (22.3511148, 78.6677428))
53	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	(Anaheim, Orange County, California, United St...
55	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	(United States, (39.7837304, -100.4458825))
56	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	(South Africa, (-28.8166236, 24.991639))

In [9]:
```
1  true_tweets_with_loc_sub.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9 entries, 31 to 56
Data columns (total 7 columns):
id            9 non-null int64
keyword       9 non-null object
location      9 non-null object
text          9 non-null object
target        9 non-null int64
longitud      9 non-null int64
geolocation   9 non-null object
dtypes: int64(3), object(4)
memory usage: 576.0+ bytes
```

Ahora se arma una lista de países, tomando el campo "address" de cada geolocalización

In [10]:
```
1  country_list = []
2  for row in true_tweets_with_loc_sub.iterrows():
3      try:
4          country_list.append(geolocator.geocode(row[1].location).address.sp
5      except:
6          country_list.append(None)
```

La lista de países así obtenida produce resultados distintos para un mismo país (ver caso de USA debajo)

In [11]:
```
1  country_list
```

Out[11]:
```
['United Kingdom',
 'Africa',
 'China 中国',
 'United States of America',
 'United States of America',
 'भारत - India',
 'United States of America',
 'United States',
 'South Africa']
```

In [12]:
```
1  true_tweets_with_loc_sub['country'] = country_list
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWi
thCopyWarning:
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/st
able/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas
-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.
```

In [13]:
```
1  true_tweets_with_loc_sub
```

Out[13]:

	id	keyword	location	text	target	longitud	geolocation	country
31	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	(Birmingham, West Midlands Combined Authority,...	United Kingdom
33	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	(Africa, (11.5024338, 17.7578122))	Africa
37	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	(環球大廈 World-Wide House, 中環 Central, 西環 Sai Wan...	China 中国
46	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	(Greensboro, Guilford County, North Carolina, ...	United States of America
50	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	(Sheffield Township, Lorain County, Ohio, Unit...	United States of America
51	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	(भारत - India, (22.3511148, 78.6677428))	भारत - India
53	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	(Anaheim, Orange County, California, United St...	United States of America
55	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	(United States, (39.7837304, -100.4458825))	United States
56	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	(South Africa, (-28.8166236, 24.991639))	South Africa

In [14]:
```
1  true_tweets_with_loc_sub.reset_index(drop=True, inplace=True)
2  true_tweets_with_loc_sub
```

Out[14]:

	id	keyword	location	text	target	longitud	geolocation	country
0	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	(Birmingham, West Midlands Combined Authority,...	United Kingdom
1	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	(Africa, (11.5024338, 17.7578122))	Africa
2	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	(環球大廈 World-Wide House, 中環 Central, 西環 Sai Wan...	China 中国
3	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	(Greensboro, Guilford County, North Carolina, ...	United States of America
4	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	(Sheffield Township, Lorain County, Ohio, Unit...	United States of America
5	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	(भारत - India, (22.3511148, 78.6677428))	भारत - India
6	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	(Anaheim, Orange County, California, United St...	United States of America
7	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	(United States, (39.7837304, -100.4458825))	United States
8	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	(South Africa, (-28.8166236, 24.991639))	South Africa

Debajo se ve que en vez de resultar 4 tweets de USA, hay 3 por un lado, y 1 por otro, ya que el "país" obtenido es distinto por cómo está cargado en el campo "address" de la geolocalizaicón correspondiente

In [15]:
```python
1  true_tweets_with_loc_sub.country.value_counts()
```

Out[15]:
```
United States of America    3
United Kingdom              1
भारत - India                1
Africa                      1
South Africa                1
United States               1
China 中国                     1
Name: country, dtype: int64
```

Para obtener un nombre de país único, se nota que convirtiendo la "location" a coordenadas primero, y luego haciendo un reverse lookup de esas coordenadas de la cual se extrae la "address" resultante, el país ahora sí tiene nombre único

In [16]:
```python
1  lat_list = []
2  long_list = []
3  for row in true_tweets_with_loc_sub.iterrows():
4      try:
5          point = (geolocator.geocode(row[1].location))
6          lat_list.append(point.latitude)
7          long_list.append(point.longitude)
8      except:
9          lat_list.append(None)
10         long_list.append(None)
```

In [17]:
```python
1  lat_list
```

Out[17]:
```
[52.4796992,
 11.5024338,
 22.2823025,
 36.0726355,
 41.4283035,
 22.3511148,
 33.8347516,
 39.7837304,
 -28.8166236]
```

In [18]:
```python
1  long_list
```

Out[18]:
```
[-1.9026911,
 17.7578122,
 114.15786063390414,
 -79.7919754,
 -82.15437852366313,
 78.6677428,
 -117.911732,
 -100.4458825,
 24.991639]
```

In [19]:
```python
1 true_tweets_with_loc_sub['latitude'] = lat_list
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: 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 (http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.

In [20]:
```python
1 true_tweets_with_loc_sub['longitude'] = long_list
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: 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 (http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.

In [21]: 1 true_tweets_with_loc_sub

Out[21]:

	id	keyword	location	text	target	longitud	geolocation	country
0	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	(Birmingham, West Midlands Combined Authority,...	United Kingdom
1	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	(Africa, (11.5024338, 17.7578122))	Africa
2	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	(環球大廈 World-Wide House, 中環 Central, 西環 Sai Wan...	China 中国
3	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	(Greensboro, Guilford County, North Carolina, ...	United States of America
4	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	(Sheffield Township, Lorain County, Ohio, Unit...	United States of America
5	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	(भारत - India, (22.3511148, 78.6677428))	भारत - India
6	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	(Anaheim, Orange County, California, United St...	United States of America
7	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	(United States, (39.7837304, -100.4458825))	United States
8	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	(South Africa, (-28.8166236, 24.991639))	South Africa

In [22]:
```python
true_tweets_with_loc_sub['coordinates'] = true_tweets_with_loc_sub['latitu
                                          true_tweets_with_loc_sub['longit
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWi
thCopyWarning:
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/st
able/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas
-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.
```

```python
true_tweets_with_loc_sub['coordinates'] = true_tweets_with_loc_sub['latitu
                                          true_tweets_with_loc_sub['longit
```

In [23]:　　`1　true_tweets_with_loc_sub`

Out[23]:

	id	keyword	location	text	target	longitud	geolocation	country
0	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	(Birmingham, West Midlands Combined Authority,...	United Kingdom
1	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	(Africa, (11.5024338, 17.7578122))	Africa
2	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	(環球大廈 World-Wide House, 中環 Central, 西環 Sai Wan...	China 中国
3	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	(Greensboro, Guilford County, North Carolina, ...	United States of America
4	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	(Sheffield Township, Lorain County, Ohio, Unit...	United States of America
5	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	(भारत - India, (22.3511148, 78.6677428))	भारत - India
6	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	(Anaheim, Orange County, California, United St...	United States of America
7	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	(United States, (39.7837304, -100.4458825))	United States
8	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	(South Africa, (-28.8166236, 24.991639))	South Africa

In [24]:　　`1　true_tweets_with_loc_sub['coordinates'][0]`

Out[24]:　`'52.4796992, -1.9026911'`

En el ejemplo debajo, se ve que ahora "country" en vez de ser "United States" como antes, ahora sí es "United Sates of America", lo cual hará que el conteo resulte 4 como debía ser

In [25]:
```
1  geolocator.reverse(true_tweets_with_loc_sub['coordinates'][7]).address.spl
```

Out[25]: 'United States of America'

In [26]:
```
1  country_from_coord_list = []
2  for row in true_tweets_with_loc_sub.iterrows():
3      try:
4          country_from_coord_list.append(geolocator.reverse(\
5                          row[1].coordinates).address.split(", ")[-1])
6      except:
7          country_from_coord_list.append(None)
```

In [27]:
```
1  country_from_coord_list
```

Out[27]: ['United Kingdom',
 'Tchad / تشاد',
 'China 中国',
 'United States of America',
 'United States of America',
 'India',
 'United States of America',
 'United States of America',
 'South Africa']

In [28]:
```
1  true_tweets_with_loc_sub['country from coordinates'] = country_from_coord_
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWi
thCopyWarning:
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/st
able/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas
-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.

In [29]:   `1  true_tweets_with_loc_sub`

Out[29]:

	id	keyword	location	text	target	longitud	geolocation	country
0	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	(Birmingham, West Midlands Combined Authority,...	United Kingdom
1	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	(Africa, (11.5024338, 17.7578122))	Africa
2	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	(環球大廈 World-Wide House, 中環 Central, 西環 Sai Wan...	China 中国
3	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	(Greensboro, Guilford County, North Carolina, ...	United States of America
4	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	(Sheffield Township, Lorain County, Ohio, Unit...	United States of America
5	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	(भारत - India, (22.3511148, 78.6677428))	भारत - India
6	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	(Anaheim, Orange County, California, United St...	United States of America
7	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	(United States, (39.7837304, -100.4458825))	United States
8	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	(South Africa, (-28.8166236, 24.991639))	South Africa -

Ahora sí, la cuenta de países de la muestra da bien

In [30]: `true_tweets_with_loc_sub['country from coordinates'].value_counts()`

Out[30]:
```
United States of America    4
United Kingdom              1
South Africa                1
China 中国                    1
Tchad / 1              تشاد
India                       1
Name: country from coordinates, dtype: int64
```

Ahora se hará el mismo procesamiento pero para todos los tweets cuya "location" no sea nula, independientemente de si los tweets son reales o falsos

In [31]: `tweets_with_loc = tweets[tweets.location.isnull()==False]`

In [32]:
```
tweets_with_loc.reset_index(drop=True, inplace=True)
tweets_with_loc.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5080 entries, 0 to 5079
Data columns (total 6 columns):
id          5080 non-null int64
keyword     5080 non-null object
location    5080 non-null object
text        5080 non-null object
target      5080 non-null int64
longitud    5080 non-null int64
dtypes: int64(3), object(3)
memory usage: 238.2+ KB
```

In [33]:

```
1  tweets_with_loc.head(20)
```

Out[33]:

	id	keyword	location	text	target	longitud
0	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55
1	49	ablaze	Est. September 2012 - Bristol	We always try to bring the heavy. #metal #RT h...	0	67
2	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82
3	52	ablaze	Philadelphia, PA	Crying out for more! Set me ablaze	0	34
4	53	ablaze	London, UK	On plus side LOOK AT THE SKY LAST NIGHT IT WAS...	0	76
5	54	ablaze	Pretoria	@PhDSquares #mufc they've built so much hype a...	0	122
6	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55
7	57	ablaze	Paranaque City	Ablaze for you Lord :D	0	22
8	59	ablaze	Live On Webcam	Check these out: http://t.co/rOI2NSmEJJ http:/...	0	114
9	62	ablaze	milky way	Had an awesome time visiting the CFC head offi...	0	120
10	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97
11	68	ablaze	Live On Webcam	Check these out: http://t.co/rOI2NSmEJJ http:/...	0	114
12	71	ablaze	England.	First night with retainers in. It's quite weir...	0	138
13	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73
14	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75
15	76	ablaze	Barbados	SANTA CRUZ Â□Ã□Ã□ Head of the St Elizabeth Pol...	0	137
16	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97
17	78	ablaze	Abuja	Noches El-Bestia '@Alexis_Sanchez: happy to se...	0	130
18	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135
19	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97

In [34]:
```python
1  print(datetime.datetime.now())
```

2020-05-19 12:56:21.047428

Para obtener un nombre de país único, se nota que convirtiendo la "location" a coordenadas primero, y luego haciendo un reverse lookup de esas coordenadas de la cual se extrae la "address" resultante, el país ahora sí tiene nombre único

In [35]:
```python
 1  lat_list = []
 2  long_list = []
 3  for row in tweets_with_loc.iterrows():
 4      try:
 5          point = (geolocator.geocode(row[1].location))
 6          lat_list.append(point.latitude)
 7          long_list.append(point.longitude)
 8      except:
 9          lat_list.append(None)
10          long_list.append(None)
```

In [36]:
```python
1  print(datetime.datetime.now())
```

2020-05-19 14:20:37.965400

In [37]:
```python
1  tweets_with_loc['latitude'] = lat_list
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: 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 (http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.

In [38]:
```python
1  tweets_with_loc['longitude'] = long_list
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: 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 (http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.

In [39]:
```
1  tweets_with_loc.head(20)
```

Out[39]:

	id	keyword	location	text	target	longitud	latitude	longit
0	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1	55	52.479699	-1.902
1	49	ablaze	Est. September 2012 - Bristol	We always try to bring the heavy. #metal #RT h...	0	67	NaN	N
2	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set a...	1	82	11.502434	17.757
3	52	ablaze	Philadelphia, PA	Crying out for more! Set me ablaze	0	34	39.952724	-75.163
4	53	ablaze	London, UK	On plus side LOOK AT THE SKY LAST NIGHT IT WAS...	0	76	51.507322	-0.127
5	54	ablaze	Pretoria	@PhDSquares #mufc they've built so much hype a...	0	122	-25.745937	28.187
6	55	ablaze	World Wide!!	INEC Office in Abia Set Ablaze - http://t.co/3...	1	55	22.282303	114.157
7	57	ablaze	Paranaque City	Ablaze for you Lord :D	0	22	14.500702	120.991
8	59	ablaze	Live On Webcam	Check these out: http://t.co/rOI2NSmEJJ http:/...	0	114	NaN	N
9	62	ablaze	milky way	Had an awesome time visiting the CFC head offi...	0	120	-7.128039	52.744
10	66	ablaze	GREENSBORO,NORTH CAROLINA	How the West was burned: Thousands of wildfire...	1	97	36.072635	-79.791
11	68	ablaze	Live On Webcam	Check these out: http://t.co/rOI2NSmEJJ http:/...	0	114	NaN	N
12	71	ablaze	England.	First night with retainers in. It's quite weir...	0	138	52.795479	-0.540
13	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	41.428303	-82.154
14	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	22.351115	78.667
15	76	ablaze	Barbados	SANTA CRUZ Â□Ã□Ã□ Head of the St Elizabeth Pol...	0	137	13.150033	-59.525

	id	keyword	location	text	target	longitud	latitude	longit
16	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	33.834752	-117.911
17	78	ablaze	Abuja	Noches El-Bestia '@Alexis_Sanchez: happy to se...	0	130	9.064331	7.489
18	79	ablaze	USA	#Kurds trampling on Turkmen flag later set it ...	1	135	39.783730	-100.445
19	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	-28.816624	24.991

In [40]:
```
1  tweets_with_loc['coordinates'] = tweets_with_loc['latitude'].astype('str')
2                           tweets_with_loc['longitude'].ast
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWi
thCopyWarning:
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/st
able/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas
-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.
```

In [41]:
```
1  tweets_with_loc
```

	id	keyword	location	text	target	longitud	latitude
18	79	ablaze	USA	Turkmen flag later set it ...	1	135	39.783730
19	80	ablaze	South Africa	TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE O...	1	97	-28.816624
20	81	ablaze	Sao Paulo, Brazil	Set our hearts ablaze and every city was a gif...	0	126	-23.550651
21	82	ablaze	hollywoodland	They sky was ablaze tonight in Los Angeles. I'...	0	119	33.808198
22	83	ablaze	Edmonton, Alberta - Treaty 6	How the West was burned: Thousands of wildfire...	1	138	NaN

In [42]:
```python
1  print(datetime.datetime.now())
```

2020-05-19 14:20:38.180446

In [43]:
```python
1  country_from_coord_list = []
2  for row in tweets_with_loc.iterrows():
3      try:
4          country_from_coord_list.append(geolocator.reverse(\
5                      row[1].coordinates).address.split(", ")[-1])
6      except:
7          country_from_coord_list.append(None)
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:4: Deprecati
onWarning: Unable to parse the string as Point: "Failed to create Point insta
nce from string: unknown format.". Using the value as-is for the query. In ge
opy 2.0 this will become an exception.
  after removing the cwd from sys.path.

In [44]:
```python
1  print(datetime.datetime.now())
```

2020-05-19 15:44:15.925647

In [45]:
```python
1  tweets_with_loc['country from coordinates'] = country_from_coord_list
```

C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWi
thCopyWarning:
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/st
able/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas
-docs/stable/indexing.html#indexing-view-versus-copy)
  """Entry point for launching an IPython kernel.

In [46]:

```
1  tweets_with_loc
```

13	73	ablaze	Sheffield Township, Ohio	Deputies: Man shot before Brighton home set ab...	1	73	41.428303
14	74	ablaze	India	Man wife get six years jail for setting ablaze...	1	75	22.351115
15	76	ablaze	Barbados	SANTA CRUZ Â☐Ã☐Ã☐ Head of the St Elizabeth Pol...	0	137	13.150033
16	77	ablaze	Anaheim	Police: Arsonist Deliberately Set Black Church...	1	97	33.834752
17	78	ablaze	Abuja	Noches El-Bestia '@Alexis_Sanchez: happy to se...	0	130	9.064331

Ahora se cuentan los países

In [47]:

```
1  tweets_with_loc['country from coordinates'].value_counts()
```

```
Eesti                      1
Uruguay                    1
Nepal                      1
Tanzania                   1
Беларусь                   1
Sénégal                    1
Togo                       1
1                    البحرين
السودان al-Sūdān           1
Guernsey                   1
Cameroon;Cameroun          1
Armenia                    1
Zimbabwe                   1
Mozambique                 1
Afghanistan                1
Србија (Serbia)            1
Sūriyya 1            سوريا
Seychelles                 1
Saint Lucia                1
Name: country from coordinates, Length: 123, dtype: int64
```

In [48]:

```
1  tweets_with_loc.to_csv(r'F:\Documentos\Pablo\facultad\orga_de_datos\TP1-mi
```