

# **IADB's Share of voice on twitter from Colombia**



# The Problem

There is a conversation in social network where people are mentioning our organization and our competitors. The problem is that we don't know how much of the conversation we owned compared to our competitors.



# The Question

- How many mentions we have and how does that compared to our competitors?



# The Data

April 20th to May 28th

```
"created_at": "Fri Jan 23 23:57:36 +0000 2015",
"id": 558775589612437504,
"id_str": "558775589612437504",
"text": "Thanks, polar vortex: Attendance dig",
"source": "\u003ca href=\"http://twitter.cc",
"truncated": false,
"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": 7313362,
  "id_str": "7313362",
  "name": "Chicago Tribune",
  "screen_name": "chicagotribune",
  "location": "Chicago, IL",
  "url": "http://www.chicagotribune.com/",
  "description": "Chicago Tribune news, featu",
  "protected": false,
  "verified": true,
  "followers_count": 321548,
  "friends_count": 523,
  "listed_count": 8074,
  "favourites_count": 34,
  "statuses_count": 47367,
  "created_at": "Sat Jul 07 14:10:07 +0000 20",
  "utc_offset": -21600,
  "time_zone": "Central Time (US & Canada)",
  "geo_enabled": false
```

Keys:

["@el\_BID", 'el bid', 'Banco Interamericano de Desarrollo', 'Banco Mundial', '@BancoMundial', 'BancoMundial', 'el Banco Mundial', '@AgendaCAF', 'AgendaCAF', 'banco de desarrollo de América Latina', 'CAF-Banco de Desarrollo de América Latina']])

API



Tweepy



# CSV File

Separator ::

27 Fields

tweetText, tweetId, tweetInReplyStatusId, tweetFavoriteCount, tweetRetweeted  
, tweetCoordinates, tweetTimestamp\_ms, tweetGeo, tweetLang, tweetDate  
, tweetUserScreen\_name, tweetUserProfile\_image\_url\_https  
, tweetUserFollowers\_count, tweetUserListed\_count, tweetUserFollowing\_count  
, tweetUserLocation, tweetUserGeo\_enabled, tweetUserName, tweetUserLang  
, tweetUserUrl, tweetUserCreated\_at, tweetUserTime\_zone, tweetUserIs\_translator

1	1429560528.77::RT @SinEmbargoMX: Las pensiones de 80 millones de latinoamericanos se encuentran en peligro
2	1429560563.56::RT @SinEmbargoMX: Las pensiones de 80 millones de latinoamericanos se encuentran en peligro
3	1429560581.3::RT @SinEmbargoMX: Las pensiones de 80 millones de latinoamericanos se encuentran en peligro
4	1429560636.03::RT @BancoMundial: Reducir a cero la quema rutinaria de gas para 2030. Apoye esta iniciativa. <a href="http://t.co/9mz78aF8bv">http://t.co/9mz78aF8bv</a> <a href="http://t.co/9mz78aF8bv">http://t.co/9mz78aF8bv</a>
5	1429560638.12::RT @BancoMundial: Reducir a cero la quema rutinaria de gas para 2030. Apoye esta iniciativa. <a href="http://t.co/9mz78aF8bv">http://t.co/9mz78aF8bv</a> <a href="http://t.co/9mz78aF8bv">http://t.co/9mz78aF8bv</a>
6	1429560648.67::RT @BancoMundial: Acceso a una #educación sin calidad es lo mismo a no acceder a ella. Únase a la lucha para poner fin a la pob
7	1429560649.71::RT @rubenromero685: @BancoMundial La caída de los precios del petróleo es inducida por los EEUU para frenar el avance de los p
8	1429560656.5::RT @rrsg55: @BancoMundial i Saboteo !::590246364812091392::None::0::False::None::1429560674275::None::tl::Mon Apr 20
9	1429560657.23::RT @BancoMundial: La dualidad de la caída de los precios del #petróleo ¿bonanza o #crisis? <a href="http://t.co/BjwmRZaY1e">http://t.co/BjwmRZaY1e</a> <a href="http://t.co/BjwmRZaY1e">http://t.co/BjwmRZaY1e</a>
10	1429560661.2::RT @BancoMundial: La dualidad de la caída de los precios del #petróleo ¿bonanza o #crisis? <a href="http://t.co/BjwmRZaY1e">http://t.co/BjwmRZaY1e</a> <a href="http://t.co/BjwmRZaY1e">http://t.co/BjwmRZaY1e</a>
11	1429560670.08::RT @NaranjaRA: 70 Bs el bidón de agua... todos los dias compro al menos 2... Régimen es tan incapaz que ni el Agua puede asegu
12	1429560674.88::RT @SinEmbargoMX: Las pensiones de 80 millones de latinoamericanos se encuentran en peligro
13	1429560676.86::Banco Mundial destaca estrategia económica de México <a href="http://t.co/akpV75QOA6">http://t.co/akpV75QOA6</a> ::590246449927041024::None::0::False::None::
14	1429560686.68::Las pensiones de 80 millones de latinoamericanos se encuentran en peligro
15	1429560691.38::Presentan al Banco Mundial Proyecto "Pase Universal" <a href="http://t.co/j1hV0II5Du">http://t.co/j1hV0II5Du</a> ::590246510757052419::None::0::False::None::14
16	1429560691.51::Banco Mundial destaca estrategia económica de México <a href="http://t.co/zVPds2xkUu">http://t.co/zVPds2xkUu</a> ::590246511759470592::None::0::False::None::1
17	1429560692.74::@BancoMundial Las razones en la caída de los precios del crudo son diversas
18	1429560710.61::RT @gobiernocolima: Presentan al Banco Mundial Proyecto "Pase Universal".

# Data cleaning

**Raw data:** 67,156 tweets

**After cleaning:** 53,987 valid tweets

**Creation** of 2 fields:

- 1) Brand
- 2) Country



# Brand

BRAND	BID	BANCO MUNDIAL	CAF	UNDEFINED
-------	-----	---------------	-----	-----------

Search for patterns in the tweet's text to assign a brand

## Example

### BID

RT @el\_BID 5 maneras de usar la #tecnología para mejorar la #movilidad urbana <http://t.co/UXvUw703N6> @BID\_Ciudades <http://t.co/VZB1bvpAEi>

RT @EFEnoticias: **El BID** aconseja reformas para salvar la pensión de 80 millones de personas en A.Latina <http://t.co/NKORj8qlzD>

### BANCO MUNDIAL

**Banco Mundial** destaca estrategia económica de México <http://t.co/zVPds2xkUu::590246511759470592>

@BancoMundial Las razones en la caída de los precios del crudo son diversas

**Regex** is the best and optimal option

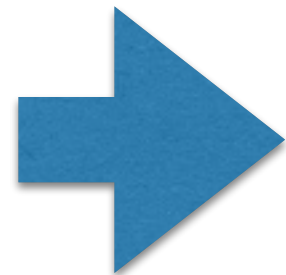
```
re.search(r"\b@el_BID\b",row.split('::')  
[1],flags=re.IGNORECASE)
```

# Country

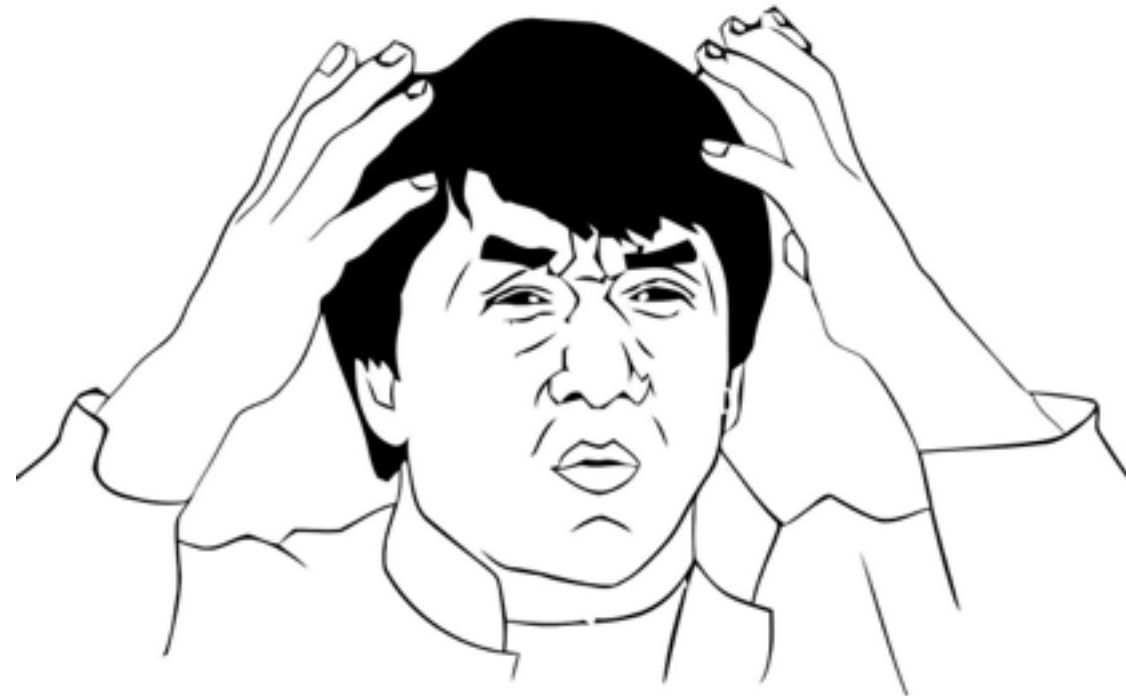
Twitter profile location field allows the user to input 'anything' they want :(

Example:

Bogot
Bogot
bogot
Bogot - Colombia
Bogot Colombia
Bogot Colombia
Bogot Colombia.
Bogot D.C
Bogot D.C - Colombia
Bogot D.C.
Bogot D.C. Colombia
Bogot-Colombia
Bogota
bogota
Bogota
BOGOTA
Bogota - Colombia
Bogota Colombia
Bogota D.C.
Bogota y Colombia Humana



COLOMBIA

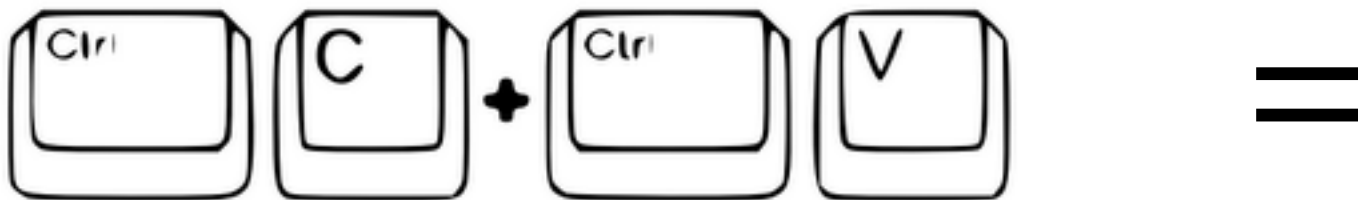




# First attempt (Make it manually)

```
213 twData['isColombia'] = twData.tweetUserLocation.str.strip().map({
214     'BARRANQUILLA':1, 'BARRANQUILLA - COLOMBIA':1
215     , 'BARRANQUILLA COLOMBIA':1, 'BOGOT':1, 'BOGOT - COLOMBIA':1, 'BOGOT COLOMBIA':1
216     , 'BOGOT COLOMBIA.':1, 'BOGOT D.C':1, 'BOGOT D.C - COLOMBIA':1, 'BOGOT D.C.':1
217     , 'BOGOT D.C. COLOMBIA':1, 'BOGOT-COLOMBIA':1, 'BOGOTA':1, 'BOGOTA - COLOMBIA':1
218     , 'BOGOTA COLOMBIA':1, 'BOGOTA D.C.':1, 'BOGOTA Y COLOMBIA HUMANA':1, 'BUCARAMANGA':1
219     , 'CALI':1, 'CALI - COLOMBIA':1
220     , 'CALI COLOMBIA':1, 'CALI- COLOMBIA':1
221     , 'CALI-COLOMBIA':1, 'CARTAGENA':1
222     , 'CCUTA COLOMBIA':1, 'COLOMBIA':1
223     , 'COLOMBIA | MXICO | ESPAA':1, 'MANIZALES':1
224     , 'MEDELLIN':1, 'MEDELLIN - COLOMBIA':1
225     , 'MEDELLIN COLOMBIA':1, 'MEDELLN':1
226     , 'MEDELLN - COLOMBIA':1, 'MEDELLN COLOMBIA':1
227     , 'PASTO':1, 'PASTO - NARIO - COLOMBIA':1
228     , 'PASTO COLOMBIA!!':1, 'PEREIRA':1
229     , 'SANTA MARTA - COLOMBIA':1, 'SANTA MARTA COLOMBIA':1
230     , 'VALLEDUPAR':1, 'VALLEDUPAR COLOMBIA':1
231 })
```

**Results:** 2,185 matches



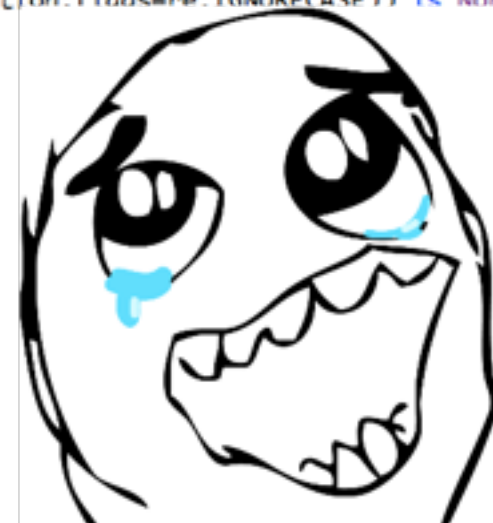


- Go to wikipedia
- Look for Colombia states, cities and counties
- Make a csv file
- Write code to match with the correct pattern

## Regex

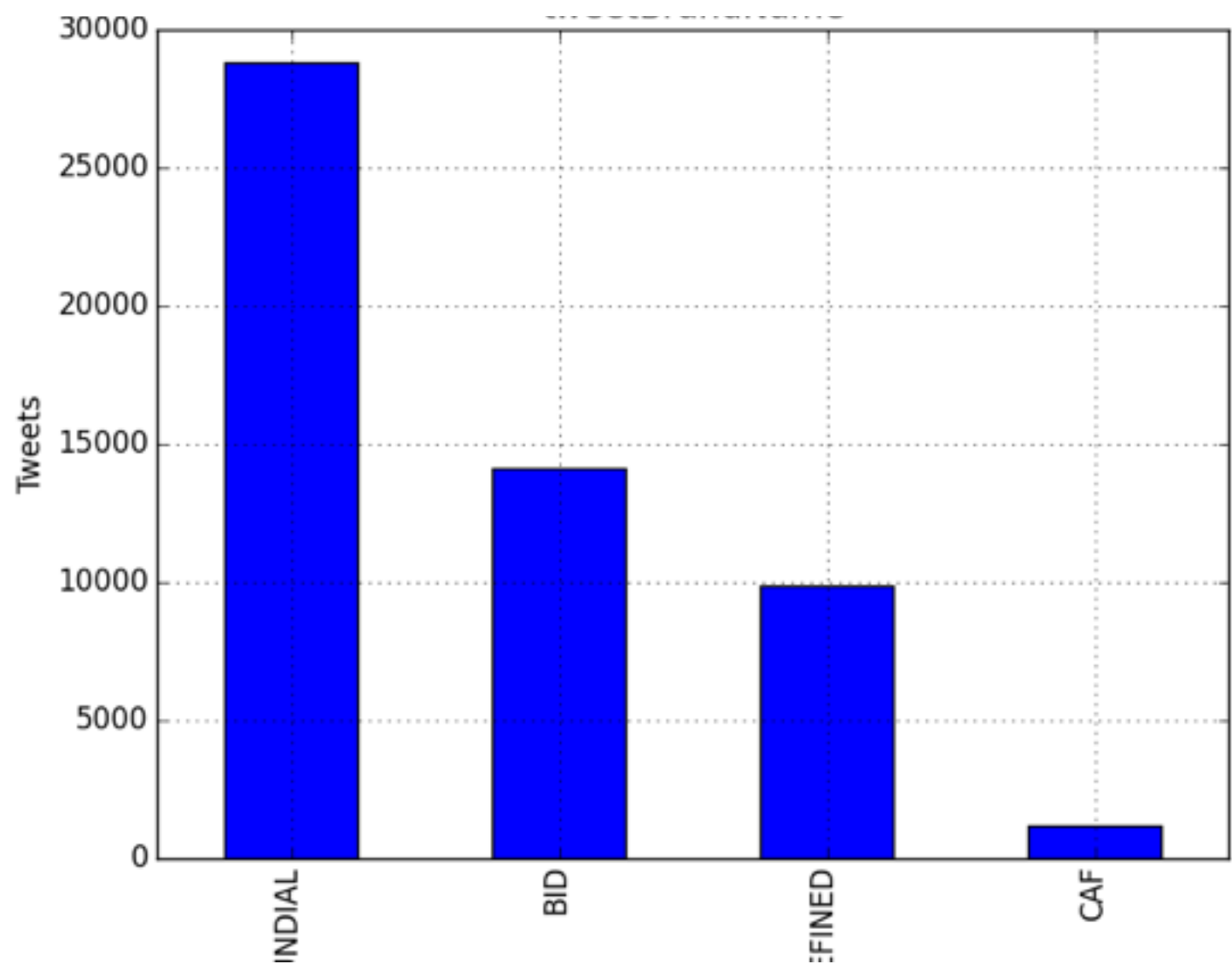
```
153 terms = pd.read_csv('Terms.csv', parse_dates=True)
154 #accent removed
155 termsCleaned = []
156 for term in terms['TERM']:
157     termsCleaned.append(term.encode("ascii", "ignore").strip())
158
159 #removing common places in other countris(the following list is irrelevant for our study)
160 #FLORENCIA, CORDOBA, AMAZONAS, LA PLATA, MADRID
161 termsCleaned.remove('FLORENCIA')
162 termsCleaned.remove('CRDOBA')
163 termsCleaned.remove('AMAZONAS')
164 termsCleaned.remove('LA PLATA')
165 termsCleaned.remove('FLORIDA')
166 termsCleaned.remove('MADRID')
167
168 #mapping twData.tweetUserLocation with terms cleaned
169
170 for location in tweetUserLocation:
171     flag = 0
172     op1 = '\\b'
173     op2 = '\\b'
174
175     for place in termsCleaned:
176         if not (re.search(r""+op1+place+op2, location.flags=re.IGNORECASE)) is None:
177             tweetisColombia.append("1")
178             flag = 1
179             break
180     if flag == 0:
181         tweetisColombia.append("0")
```

**Results:** 3,268 matches



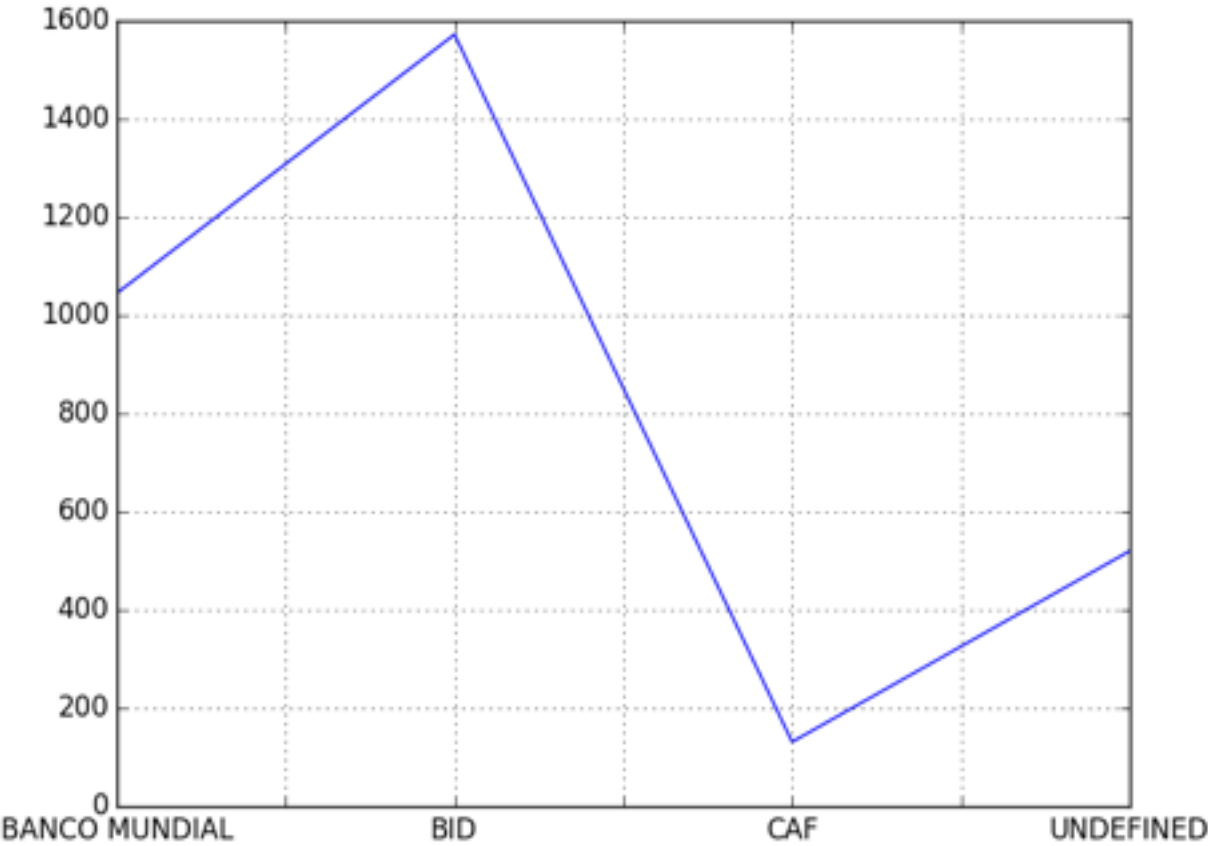
# Data exploration

Not filtered by location yet

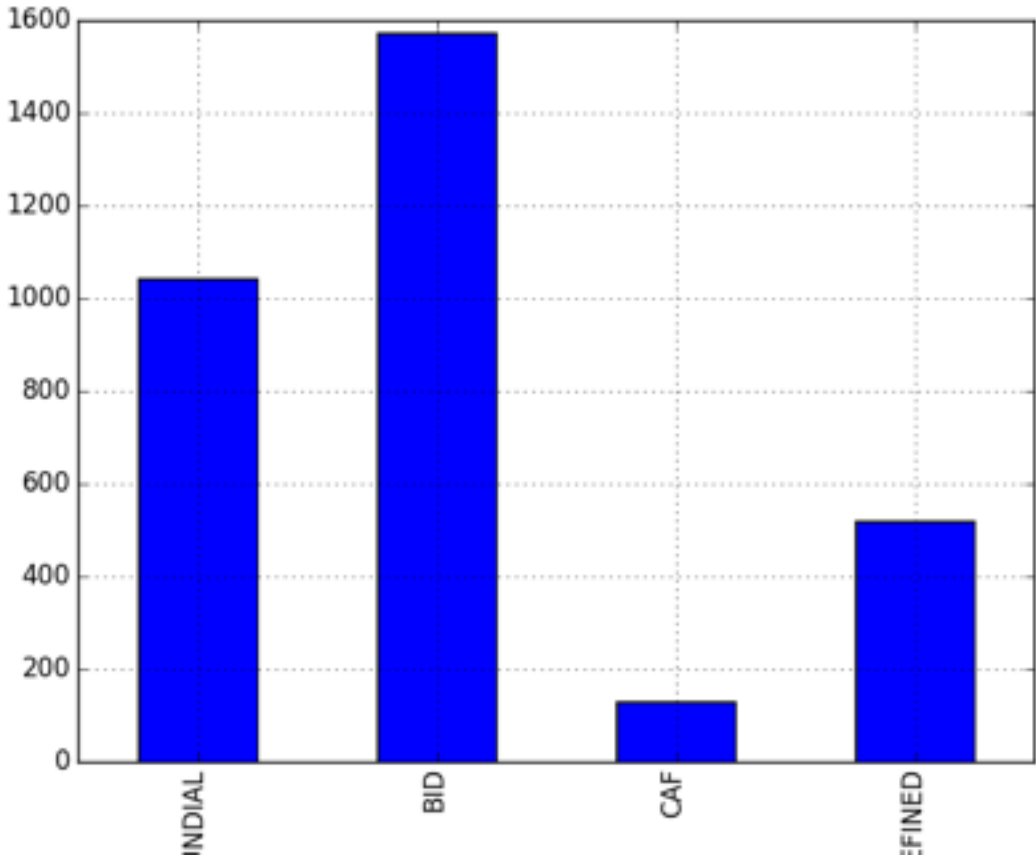


		TWEETS	%
1	BANCO MUNDIAL	28797	53
2	BID	14153	26
3	CAF	1190	2
	UNDEFINED	9847	18
	TOTAL:	53987	100

# Colombia



		TWEETS	%
1	BID	1572	48
2	BANCO MUNDIAL	1045	32
3	CAF	131	4
	UNDEFINED	520	16
	TOTAL:	3268	100



# Conclusion

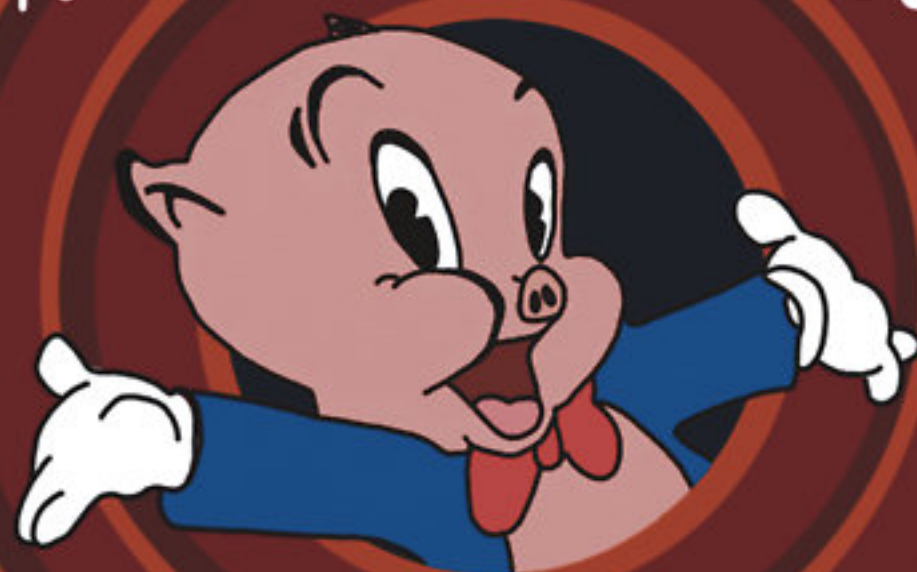
Collecting data with python is the best and faster option. Performance and usability is easier than other languages like php, c#, c++.

The challenge is to build the data-frame. Once that is complete you only have to analyze the data, make connections and write the code to look for matches.

# Recommendations

- Use a data base to collect data (don't use csv files)
- Run the algorithm every two days
- Getting real time data will prevent you from collecting the engagement information (Favs and RTs)
- Regular expression is a topic that you need to explore more deeply

*That's all Folks!*



kalilak