Query based Sentiment Analysis using Python

In [1]:

```
# General:
import tweepy           # To consume Twitter's API
import pandas as pd     # To handle data
import numpy as np      # For number computing

# For plotting and visualization:
from IPython.display import display
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Creating a Twitter App

In [2]:

```
# Twitter App access keys for @user

# Consume:
CONSUMER_KEY    = 'yzxNyCyjkVJJrpNY0L3AphSVV'
CONSUMER_SECRET = 'e7xhbuIqTM520PCtwGrSrSIXWyrQbH7ckr9s0ToOlQE6tek4OP'

# Access:
ACCESS_TOKEN  = '2186056028-BPexxHqPrML7hcvwIuLfX3itD8IZPqguuksDHtn'
ACCESS_SECRET = 'vvhbgIqa7jK5nBAO9P4Mh1MTFvJdWH3Fl5bbzGhai9Dah'
```

In [3]:

```
# We import our access keys:
from credentials import *    # This will allow us to use the keys as variables

# API's setup:
def twitter_setup():
    """
    Utility function to setup the Twitter's API
    with our access keys provided.
    """
    # Authentication and access using keys:
    auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
    auth.set_access_token(ACCESS_TOKEN, ACCESS_SECRET)

    # Return API with authentication:
    api = tweepy.API(auth)
    return api
```

In [4]:

```
extractor = twitter_setup()
trends1 = extractor.trends_place(1) # from the end of your code
# trends1 is a list with only one element in it, which is a
# dict which we'll put in data.
data = trends1[0]
# grab the trends
trends = data['trends']
# grab the name from each trend
names = [trend['name'] for trend in trends]
# put all the names together with a ' ' separating them
trendsName = "\n".join(names)
print(trendsName)
```

```
dória
#ال_سعود_ساده_وقاده
#وش_بتسوي_لو_صرت_الملك
#díadelmédico
megyn kelly
joão doria
#23oct
#يامحمد_الهمه_شعبك_معك_للقمه
eli apple
parejo
chad kelly
lucas vázquez
young boys
baby one more time
javi martínez
odriozola
geraldo azevedo
#aekfcb
#obomdesseanofoi
#tempo_kai
#felizmartes
#٢٥_مليون_محمد_بن_سلمان
#وش_بتسوي_لو_صرت_مليونير
#yaşadıkçatürkçüyüz
#عموري_رباط_صليبي
#championsleague
#expectationsatmidnight
#غرد_وكانك_في_التسعينات
#bielnopânico
#atentadofakedopt
#tuesdaythoughts
#championsxespn
#يكرهك_الشخص_اللي
#venezuelapaíscomunal
#everybodyshouldget
#güneşile3yıldaha
#somethingwickedin3words
#نطالب_بتجديد_عقد_عموري
#munjuv
#herkesyoluna
#haddadnaglobo
#dhdl
#الغا_قناه_الجزيره
#پرسپولیس
#tuesdaymotivation
#haddadéfefeca
```

```
#unmédicoquemecure
#ولي_العهد
#mehmetçikkutluzafer
#ybvcf
```

In [5]:

```
query="#indvspak" # Enter user or query
```

In [6]:

```
# We create an extractor object:
extractor = twitter_setup()

# We create a tweet list based on query:

tweets = extractor.search(q=query, count=100)

# We create a tweet list based on user

# tweets = extractor.user_timeline(screen_name=query, count=100)

print("Number of tweets extracted: {}.\n".format(len(tweets)))

# We print the most recent 5 tweets:
print("5 recent tweets:\n")
for tweet in tweets[:5]:
    print(tweet.text)
    print()
```

```
Number of tweets extracted: 100.

5 recent tweets:

It's one thing to get inside someone's head. It's another to stay ther
e. #waqaryounis #pakvsind #indvspak #cricket… https://t.co/B0SqRDH7N5
 (https://t.co/B0SqRDH7N5)

RT @SnehalWele: One of the best places in India.#Vizag .And vizag remi
nds me of this #IndVsPak masterclass by my Favourite person MS Dhoni.…

RT @SnehalWele: One of the best places in India.#Vizag .And vizag remi
nds me of this #IndVsPak masterclass by my Favourite person MS Dhoni.…

One of the best places in India.#Vizag .And vizag reminds me of this #
IndVsPak masterclass by my Favourite person M… https://t.co/pduQ6gkVOB
 (https://t.co/pduQ6gkVOB)

RT @SirJadeja: Be It War
Be It GDP
Be It Economy
Be It Kabaddi
Be It Football
Be It Hockey
Be It Olympics
Be It Asian Games
Be It Cricket…
```

In [7]:

```
# We create a pandas dataframe as follows:
data = pd.DataFrame(data=[tweet.text for tweet in tweets], columns=['Tweets'])

# We display the first 10 elements of the dataframe:
display(data.head(10))
```

	Tweets
0	It's one thing to get inside someone's head. I...
1	RT @SnehalWele: One of the best places in Indi...
2	RT @SnehalWele: One of the best places in Indi...
3	One of the best places in India.#Vizag .And vi...
4	RT @SirJadeja: Be It War\nBe It GDP\nBe It Eco...
5	RT @atulsinghswaraj: Cameraman performing bett...
6	RT @iPoonampandey: Who will WIN today's Match?...
7	RT @virendersehwag: Koi Biwi ke liye kar raha ...
8	RT @SirJadeja: Quarter 1: Pakistan 1-0 India\n...
9	RT @Anshu_Vats1: Many Congratulations Team. 🇮🇳...

In [8]:

```
# Internal methods of a single tweet object:
print(dir(tweets[0]))
```

```
['__class__', '__delattr__', '__dict__', '__dir__', '__doc__', '__eq_
_', '__format__', '__ge__', '__getattribute__', '__getstate__', '__gt_
_', '__hash__', '__init__', '__le__', '__lt__', '__module__', '__ne_
_', '__new__', '__reduce__', '__reduce_ex__', '__repr__', '__setattr_
_', '__sizeof__', '__str__', '__subclasshook__', '__weakref__', '_ap
i', '_json', 'author', 'contributors', 'coordinates', 'created_at', 'd
estroy', 'entities', 'favorite', 'favorite_count', 'favorited', 'geo',
'id', 'id_str', 'in_reply_to_screen_name', 'in_reply_to_status_id', 'i
n_reply_to_status_id_str', 'in_reply_to_user_id', 'in_reply_to_user_id
_str', 'is_quote_status', 'lang', 'metadata', 'parse', 'parse_list',
'place', 'possibly_sensitive', 'retweet', 'retweet_count', 'retweete
d', 'retweets', 'source', 'source_url', 'text', 'truncated', 'user']
```

Adding relevant info to our dataframe

In [9]:

```
# We add relevant data:
data['len']  = np.array([len(tweet.text) for tweet in tweets])
data['ID']   = np.array([tweet.id for tweet in tweets])
data['Date'] = np.array([tweet.created_at for tweet in tweets])
data['Source'] = np.array([tweet.source for tweet in tweets])
data['Likes']  = np.array([tweet.favorite_count for tweet in tweets])
data['RTs']    = np.array([tweet.retweet_count for tweet in tweets])
data['Language'] = np.array([tweet.lang for tweet in tweets])
```

In [10]:

```
# Display of first 10 elements from dataframe:
display(data.head(10))
```

	Tweets	len	ID	Date	Source	Likes	RTs	Language
0	It's one thing to get inside someone's head. I...	139	1054802503408193537	2018-10-23 18:31:28	Instagram	0	0	en
1	RT @SnehalWele: One of the best places in Indi...	140	1054792814964355073	2018-10-23 17:52:58	Twitter Web Client	0	2	en
2	RT @SnehalWele: One of the best places in Indi...	140	1054791273645670402	2018-10-23 17:46:51	Twitter for Android	0	2	en
3	One of the best places in India.#Vizag .And vi...	140	1054790421346316289	2018-10-23 17:43:28	Twitter for Android	3	2	en
4	RT @SirJadeja: Be It War\nBe It GDP\nBe It Eco...	138	1054781686490251265	2018-10-23 17:08:45	Twitter Web Client	0	890	en
5	RT @atulsinghswaraj: Cameraman performing bett...	126	1054776585650401281	2018-10-23 16:48:29	Twitter Web Client	0	2	en
6	RT @iPoonampandey: Who will WIN today's Match?...	140	1054772591460642816	2018-10-23 16:32:37	Twitter for iPad	0	848	en
7	RT @virendersehwag: Koi Biwi ke liye kar raha ...	140	1054771354698829824	2018-10-23 16:27:42	Twitter for Android	0	948	hi
8	RT @SirJadeja: Quarter 1: Pakistan 1-0 India\n...	140	1054718910832885763	2018-10-23 12:59:18	Twitter for Android	0	120	et
9	RT @Anshu_Vats1: Many Congratulations Team. 🇮🇳...	139	1054690033024659457	2018-10-23 11:04:33	Twitter for Android	0	56	hi

In [11]:

```
# We extract the mean of lenghts:
mean = np.mean(data['len'])

print("The lenght's average in tweets: {}".format(mean))
```

```
The lenght's average in tweets: 131.2
```

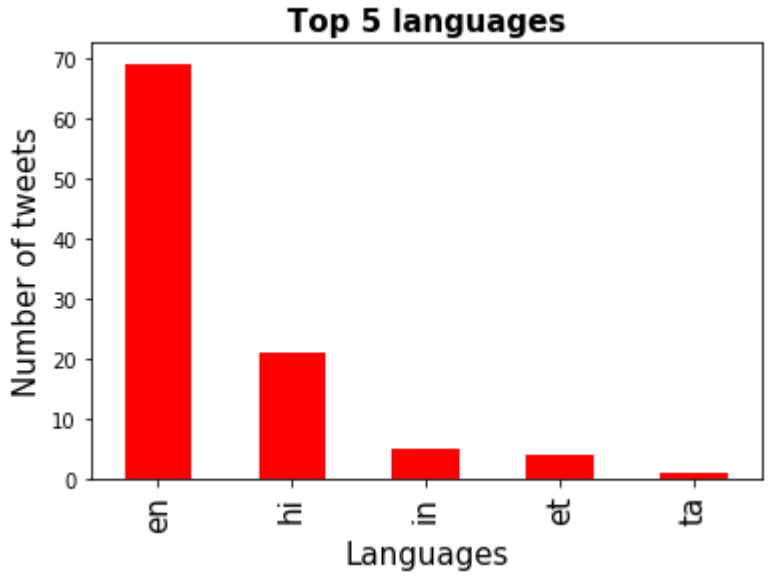
In [12]:

```
tweets_by_lang = data['Language'].value_counts()

fig, ax = plt.subplots()
ax.tick_params(axis='x', labelsize=15)
ax.tick_params(axis='y', labelsize=10)
ax.set_xlabel('Languages', fontsize=15)
ax.set_ylabel('Number of tweets' , fontsize=15)
ax.set_title('Top 5 languages', fontsize=15, fontweight='bold')
tweets_by_lang[:5].plot(ax=ax, kind='bar', color='red')
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x10b0f8fd0>
```



In [13]:

```
# We extract the tweet with more FAVs and more RTs:

fav_max = np.max(data['Likes'])
rt_max  = np.max(data['RTs'])

fav = data[data.Likes == fav_max].index[0]
rt  = data[data.RTs == rt_max].index[0]

# Max FAVs:
print("The tweet with more likes is: \n{}".format(data['Tweets'][fav]))
print("Number of likes: {}".format(fav_max))
print("{} characters.\n".format(data['len'][fav]))

# Max RTs:
print("The tweet with more retweets is: \n{}".format(data['Tweets'][rt]))
print("Number of retweets: {}".format(rt_max))
print("{} characters.\n".format(data['len'][rt]))
```

```
The tweet with more likes is:
@ICC @imVkohli Aur inko kasmir chahiye 😂😂
#INDvWI
#INDvsPAK https://t.co/BBOEtkrX1a (https://t.co/BBOEtkrX1a)
Number of likes: 41
85 characters.

The tweet with more retweets is:
RT @anjanaomkashyap: लेस भी नहीं बाँधके आते ठीक से । कोई भी मैदान हो, गाँठ बाँधकर हार
 का हमने विदा किया है। अगली बार थोड़ी तैयारी से आना ।…
Number of retweets: 1899
139 characters.
```

Time series of Tweets

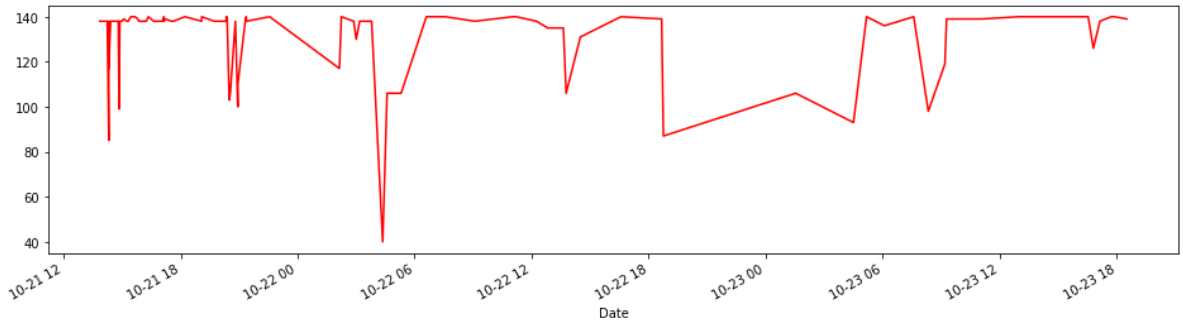
In [14]:

```
# We create time series for data:

tlen = pd.Series(data=data['len'].values, index=data['Date'])
tfav = pd.Series(data=data['Likes'].values, index=data['Date'])
tret = pd.Series(data=data['RTs'].values, index=data['Date'])
```

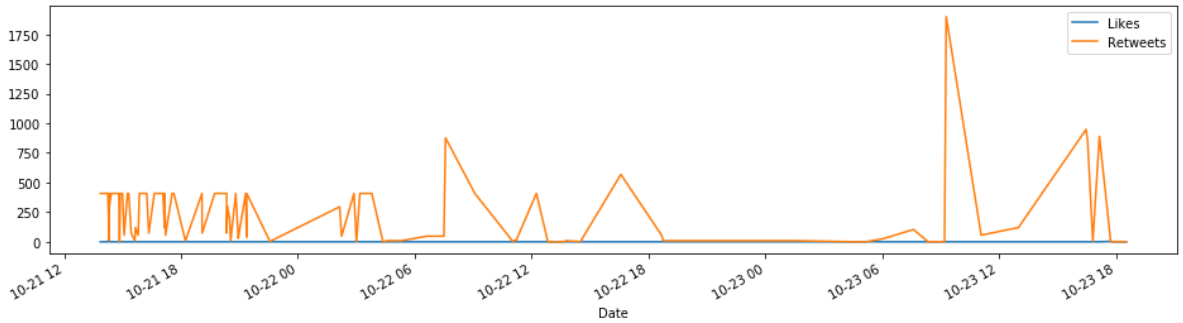
In [15]:

```
# Lenghts along time:
tlen.plot(figsize=(16,4), color='r');
```



In [16]:

```
# Likes vs retweets visualization:
tfav.plot(figsize=(16,4), label="Likes", legend=True)
tret.plot(figsize=(16,4), label="Retweets", legend=True);
```



Pie charts of sources

In [17]:

```
# We obtain all possible sources:
sources = []
for source in data['Source']:
    if source not in sources:
        sources.append(source)

# We print sources list:
print("Creation of content sources:")
for source in sources:
    print("* {}".format(source))
```

```
Creation of content sources:
* Instagram
* Twitter Web Client
* Twitter for Android
* Twitter for iPad
* Twitter Lite
* Twitter for iPhone
```

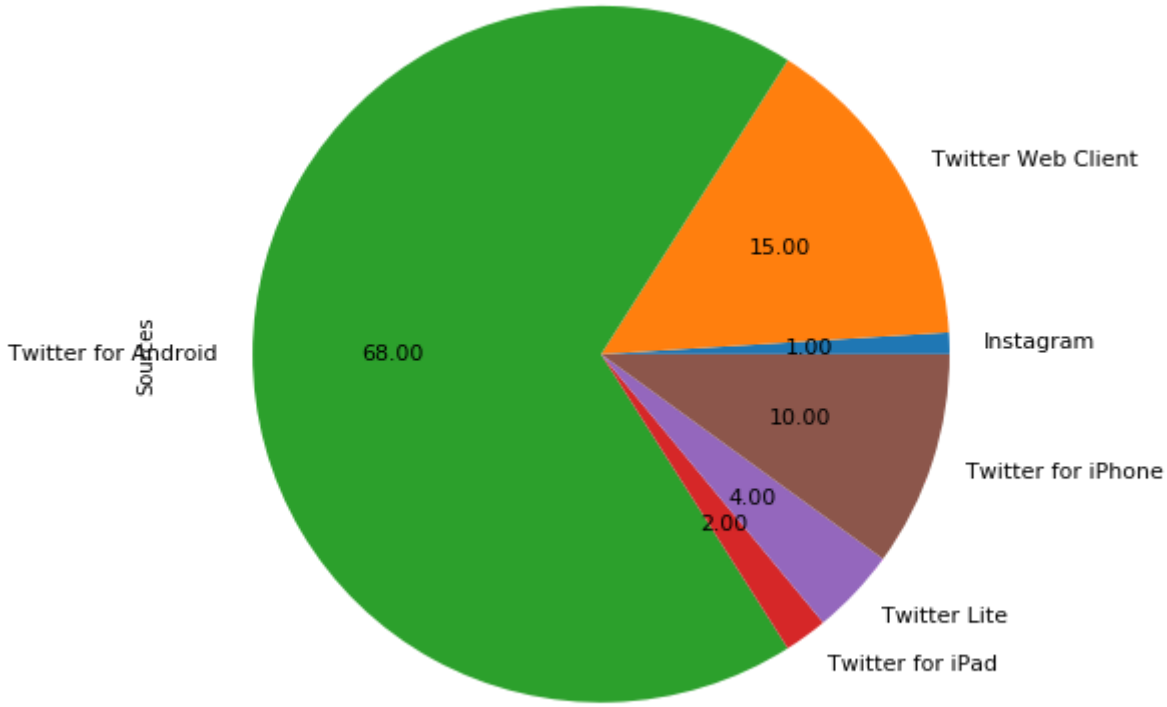
In [18]:

```
# We create a numpy vector mapped to labels:
percent = np.zeros(len(sources))

for source in data['Source']:
    for index in range(len(sources)):
        if source == sources[index]:
            percent[index] += 1
            pass

percent /= 100

# Pie chart:
pie_chart = pd.Series(percent, index=sources, name='Sources')
pie_chart.plot.pie(fontsize=11, autopct='%.2f', figsize=(10,8));
```



Sentiment analysis

In [19]:

```
from textblob import TextBlob
import re

def clean_tweet(tweet):
    '''
    Utility function to clean the text in a tweet by removing
    links and special characters using regex.
    '''
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)", " ", t

def analize_sentiment(tweet):
    '''
    Utility function to classify the polarity of a tweet
    using textblob.
    '''
    analysis = TextBlob(clean_tweet(tweet))
    if analysis.sentiment.polarity > 0:
        return 1
    elif analysis.sentiment.polarity == 0:
        return 0
    else:
        return -1
```

In [20]:

```
# We create a column with the result of the analysis:
data['SA'] = np.array([ analize_sentiment(tweet) for tweet in data['Tweets'] ])

# We display the updated dataframe with the new column:
display(data.head(10))
```

	Tweets	len	ID	Date	Source	Likes	RTs	Language	SA
0	It's one thing to get inside someone's head. I...	139	1054802503408193537	2018-10-23 18:31:28	Instagram	0	0	en	0
1	RT @SnehalWele: One of the best places in Indi...	140	1054792814964355073	2018-10-23 17:52:58	Twitter Web Client	0	2	en	1
2	RT @SnehalWele: One of the best places in Indi...	140	1054791273645670402	2018-10-23 17:46:51	Twitter for Android	0	2	en	1
3	One of the best places in India.#Vizag .And vi...	140	1054790421346316289	2018-10-23 17:43:28	Twitter for Android	3	2	en	1
4	RT @SirJadeja: Be It War\nBe It GDP\nBe It Eco...	138	1054781686490251265	2018-10-23 17:08:45	Twitter Web Client	0	890	en	0
5	RT @atulsinghswaraj: Cameraman performing bett...	126	1054776585650401281	2018-10-23 16:48:29	Twitter Web Client	0	2	en	1
6	RT @iPoonampandey: Who will WIN today's Match?...	140	1054772591460642816	2018-10-23 16:32:37	Twitter for iPad	0	848	en	1
7	RT @virendersehwag: Koi Biwi ke liye kar raha ...	140	1054771354698829824	2018-10-23 16:27:42	Twitter for Android	0	948	hi	0
8	RT @SirJadeja: Quarter 1: Pakistan 1-0 India\n...	140	1054718910832885763	2018-10-23 12:59:18	Twitter for Android	0	120	et	0
9	RT @Anshu_Vats1: Many Congratulations Team. 🇮🇳...	139	1054690033024659457	2018-10-23 11:04:33	Twitter for Android	0	56	hi	1

In [21]:

```
# We construct lists with classified tweets:

pos_tweets = [ tweet for index, tweet in enumerate(data['Tweets']) if data['SA'][inc
neu_tweets = [ tweet for index, tweet in enumerate(data['Tweets']) if data['SA'][inc
neg_tweets = [ tweet for index, tweet in enumerate(data['Tweets']) if data['SA'][inc
```

In [22]:

```
# We print percentages:

print("Percentage of positive tweets: {}%".format(len(pos_tweets)*100/len(data['Twee
print("Percentage of neutral tweets: {}%".format(len(neu_tweets)*100/len(data['Tweet
print("Percentage of negative tweets: {}%".format(len(neg_tweets)*100/len(data['Twee
```

```
Percentage of positive tweets: 19.0%
Percentage of neutral tweets: 77.0%
Percentage of negative tweets: 4.0%
```

In [23]:

```
# Data to plot
labels = 'Positive Tweets', 'Neutral Tweets','Negative Tweets'
sizes = [len(pos_tweets)*100/len(data['Tweets']), len(neu_tweets)*100/len(data['Twee
colors = ['gold', 'yellowgreen', 'lightcoral']
explode = (0.07, 0.07, 0.07)  # explode 1st slice

# Plot
plt.figure(figsize=(10,6))
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=140)
plt.axis('equal')
plt.show()
```