

Approach

# Sentiment Analysis on Twitter data for BRExit

---

Neha Pai

Apurva Mokal

19 August, 2016

## Introduction

In the Internet Age, the anonymity the platform provides makes Social media the most forthcoming and impactful in terms of information, opinions, and sentiments. This implies increasing need for understanding public voice, and thus demands ways for Analysing Sentiments. This documentation presents one of these ways where Tweets are used for analysing what the people have to say about 'BRExit', and present their views as Statistical data for further Data Analytics.

The approach is explained as: -

1. Extracting Tweets
  - Creating an App on twitter
  - Using '#brexit' to search tweets (Corpus Creation)
  - Other tweet processing methods
2. Sentiment Analysis
  - Importance of Opinion Lexicon
  - Using Opinion Lexicon for Scoring tweets
  - Classifying processed tweets into Sentiments
  - Preparing the data for further Data Analysis
3. Data Analysis
  - Introduction to Tableau
  - Representing the Data in charts
  - Creating Dashboard

## Extracting Tweets

**Creating an App on twitter:** For extracting tweets from twitter, Authentication must be done. For this we must create an app in twitter by logging into dev.twitter.com and apps.twitter.com

Following is the app we created and used in our project.

### Twitter Apps

Create New App



On creation of the app we are provided with Consumer Keys and Access Tokens. These keys will duly authenticate our R script, enabling us to use 'twitterR' package API for extracting tweets by using 'searchTwitter()' function.

**Using '#brexit' to search tweets (Corpus Creation):** A corpus can be defined as a collection of machine-readable authentic texts (including transcripts of spoken data) which is sampled to be representative of a particular natural language or language variety. Corpora provide a material basis and a test bed for building NLP systems. By using #brexit tag we have retrieved the data about brexit from twitter. This data is transformed into corpus.

```
s <- searchTwitter('#brexit', n = 15000, since = '2016-07-23', locale = 'en', geocode =  
'51.507351,-0.127758,100mi')
```

This function will create a Comma Separated Value File (.csv) file which will be used for further processing.

```
#brexit_stack - Notepad
File Edit Format View Help

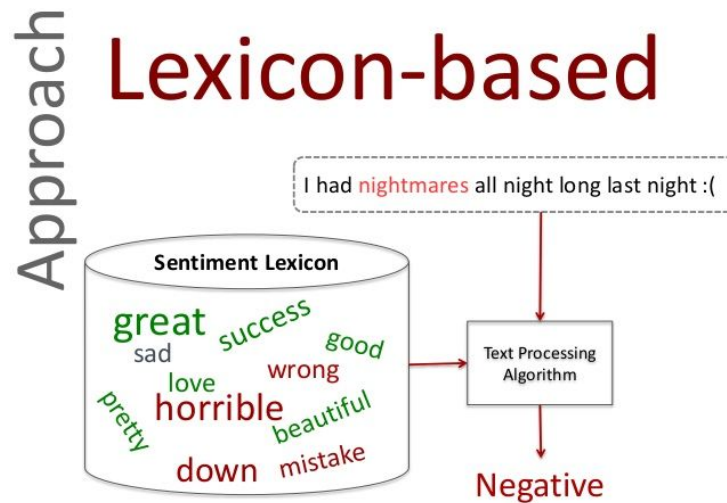
{"created","favoriteCount","favorited","id","isRetweet","language","latitude","location","longitude","profileImageURL","reply"
"2016-08-19",0,FALSE,"766587942408220672",TRUE,"en-gb",NA,"",NA,"http://pbs.twimg.com/profile_images/759747455944536065/Vqt
"2016-08-19",0,FALSE,"766587937689600000",TRUE,"en",NA,"Cambridge, England",NA,"http://pbs.twimg.com/profile_images/6873239
"2016-08-19",0,FALSE,"766587932786434048",TRUE,"en",NA,"Speaker's Chair",NA,"http://pbs.twimg.com/profile_images/5205655713
"2016-08-19",0,FALSE,"766587924339167232",FALSE,"en",NA,"London, England",NA,"http://pbs.twimg.com/profile_images/758936820
"2016-08-19",0,FALSE,"766587893926268928",TRUE,"en",NA,"Lurgan, Ireland",NA,"http://pbs.twimg.com/profile_images/7366274829
"2016-08-19",0,FALSE,"766587874087231488",FALSE,"en",NA,"Marylebone London",NA,"http://pbs.twimg.com/profile_images/5205513
are losing theirs and blaming it on you." - Your #brexit erudition is ace!",FALSE
"2016-08-19",0,FALSE,"766587851224051712",TRUE,"en",NA,"EU",NA,"http://pbs.twimg.com/profile_images/473743397983760384/CIBL
"2016-08-19",0,FALSE,"766587811462017024",FALSE,"en",NA,"Manchester, UK",NA,"http://pbs.twimg.com/profile_images/1712369846
https://t.co/mnkJsMUJu6",FALSE
"2016-08-19",0,FALSE,"766587776674430976",FALSE,"en-gb",NA,"Alba / Ecosse / Scotia",NA,"http://pbs.twimg.com/profile_image
"2016-08-19",0,FALSE,"766587773688156160",TRUE,"en",NA,"Nuneaton England",NA,"http://pbs.twimg.com/profile_images/76417417
"2016-08-19",0,FALSE,"766587771893022720",TRUE,"en",NA,"UK",NA,"http://pbs.twimg.com/profile_images/752512181791436800/xZnh
"2016-08-19",0,FALSE,"766587697561538560",TRUE,"en-gb",NA,"Europe",NA,"http://pbs.twimg.com/profile_images/7288750103587061
#Brexit is a bad decision...",FALSE
"2016-08-19",0,FALSE,"766587666028646400",TRUE,"en",NA,"Canada",NA,"http://pbs.twimg.com/profile_images/761476957821345792/!
"2016-08-19",0,FALSE,"766587640254754816",FALSE,"en",NA,"Peckham",NA,"http://pbs.twimg.com/profile_images/74339629856969113
"2016-08-19",0,FALSE,"766587637050273792",TRUE,"en",NA,"London, United Kingdom",NA,"http://pbs.twimg.com/profile_images/70
"2016-08-19",0,FALSE,"766587633724166144",FALSE,"it",NA,"",NA,"http://pbs.twimg.com/profile_images/632187818794741760/1KdJh
"2016-08-19",0,FALSE,"766587631547416576",TRUE,"en",NA,"United Kingdom",NA,"http://pbs.twimg.com/profile_images/72143649527
https://t...",FALSE
"2016-08-19",0,FALSE,"766587630683381760",TRUE,"en",NA,"in the sunshine",NA,"http://pbs.twimg.com/profile_images/7546958800
"2016-08-19",0,FALSE,"766587605332987904",TRUE,"en-gb",NA,"England, United Kingdom",NA,"http://pbs.twimg.com/profile_images
"2016-08-19",0,FALSE,"766587596159856640",FALSE,"en",NA,"Lakes and Mountains",NA,"http://pbs.twimg.com/profile_images/74601
"2016-08-19",0,FALSE,"766587592330579968",TRUE,"en",NA,"",NA,"http://pbs.twimg.com/profile_images/576056919942696961/gH7pW6
"2016-08-19",0,FALSE,"766587585426821120",FALSE,"en",NA,"Glasgow",NA,"http://pbs.twimg.com/profile_images/68411425231619277
"2016-08-19",0,FALSE,"766587585397481472",TRUE,"en",NA,"MIDDLE ENGLAND, UNITED KINGDOM",NA,"http://pbs.twimg.com/profile_im
"2016-08-19",0,FALSE,"766587574458654720",FALSE,"en",NA,"Brussels, Belgium",NA,"http://pbs.twimg.com/profile_images/7484945
"2016-08-19",0,FALSE,"766587564975394816",TRUE,"en",NA,"",NA,"http://pbs.twimg.com/profile_images/765663461611298816/EmFAkg
"2016-08-19",0,FALSE,"766587558218207232",TRUE,"en",NA,"Australia",NA,"http://pbs.twimg.com/profile_images/7367093036088442
```

**Other tweet processing methods:** The Corpus is then treated with certain Cleaning Functions, i.e. Removing 'at(@) mentions', deleting duplicate tweets, removing Punctuations, followed by converting text into Lowercase, and Tokenization of the entire corpus.

## Sentiment Analysis

**Importance of Opinion Lexicon:** Opinion Lexicon is a collection of words which have been determined as either positive or negative, with scores of each and every word in terms of polarity. The general idea is to calculate a sentiment score for each tweet so we can know how positive or negative is the posted message.

 negative-words	3/12/2011 11:46 AM	Text Document	46 KB
 positive-words	3/12/2011 11:46 AM	Text Document	21 KB



### Using Opinion Lexicon for Scoring tweets:

Tokens, previously generated, will all be judged against the lexicon and each tweet will be scored on this basis. This will allow better classification in terms of statistics since tweets with negative impact on the sentiment will have lower scores compared to positively oriented tweets. These scores are then saved into a .csv file as well.

```
write.csv(scores, file = paste(searchterm, '_scores.csv'))
```

```
#brexit_scores - Notepad
File Edit Format View Help

["", "score", "text"
"1", 1, "RT @LouiseMensch: Good #brexit italexit? https://t.co/JdNLSw8QBQ"
"2", -1, "RT @sutherla: Oddly relevant for post #Brexit UK https://t.co/ahkS2Exkjl"
"3", 1, "RT @UKIP2Win: Migrants: Why do we have to work? https://t.co/VCOQI35y80 #ukip #brexit #farage"
"4", 1, "UK hopes #FinTech will help #banking #reform https://t.co/Ng801wk0Fv #Brexit #EUref"
"5", 0, "RT @UpperBannSFry: #Brexit meeting with @M_AndersonSF - Erasmus, apprenticeships & education on the agenda. https://t.co/..."
"6", -1, "@davidallengreen "If you can keep your head when all about you
are losing theirs and blaming it on you..." - Your #brexit erudition is ace!"
"7", -1, "RT @euromove: More than 100 Tory MPs "desperately" want #UK to stay in the #EU, says Ken Clarke https://t.co/AE3pT8wvAw"
"8", 0, "Read and reach your own conclusions. ""Boom or gloom? The economic verdict on #Brexit - so far"" The Guardian
https://t.co/mnkJsMUJu6"
"9", 0, "#Brexit forced Scotland & @scotgov in2 expedited negotiations w/ Europe, the EU, EUBA/EUBU & other members of crum..."
"10", 0, "RT @standust193: #Brexit British Shoppers Unfazed By Brexit Vote And Project Fear As Spending Jumps In July https://t.co/l..."
```

**Classifying processed tweets into Sentiments:** Once all the scores are obtained, the corpus is classified with respect to Negative tweets, Positive tweets and Neutral tweets, each with individual scores, which is stored in opin.csv.

```
#brexit_opin - Notepad
File Edit Format View Help

["", "tweet", "created", "number"
"1", "negative", 2016-08-16, 440
"2", "negative", 2016-08-17, 527
"3", "negative", 2016-08-18, 614
"4", "negative", 2016-08-19, 759
"5", "neutral", 2016-08-16, 893
"6", "neutral", 2016-08-17, 982
"7", "neutral", 2016-08-18, 1057
"8", "neutral", 2016-08-19, 1576
"9", "positive", 2016-08-16, 401
"10", "positive", 2016-08-17, 466
"11", "positive", 2016-08-18, 515
"12", "positive", 2016-08-19, 626
```

## Data Analysis

**Introduction to Tableau:** Tableau is a tool that can be described as an instrument that combines interactive visual analytics, sophisticated visualization methods, an easy to use graphical user-interface and a lot of connectors for dozens of different data sources ranging from standard Excel to more “Big Data” like things such as different Hadoop distributions and NoSQL databases. What is really impressing is that you can create interactive visualizations in just no time.



**Representing the Data in charts:** A picture is worth more than a thousand words. So it is better to represent data in forms of charts, as it makes data more expressive and efficient to read.

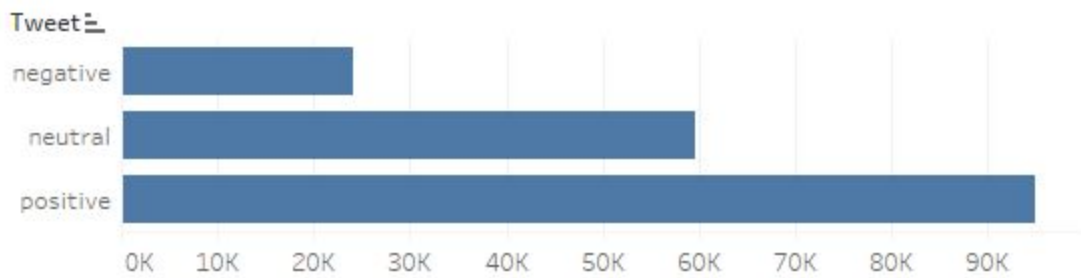
We have used tableau, an excellent visual analytics tool.

We have depicted sentiments using bar graph, pie chart and bubble chart as shown by creating individual sheets.

Further to analyse tweets geographically by location, we used a tree map and map analysis



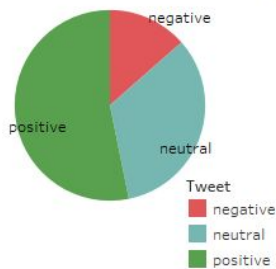
## Bar graph



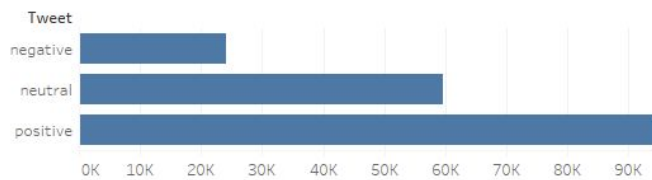
Creating Dashboard:

# BREXIT ANALYSIS

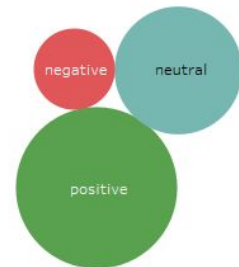
Pie Chart



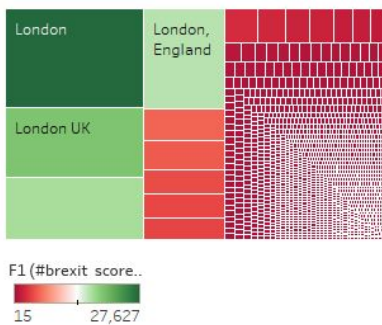
Bar graph



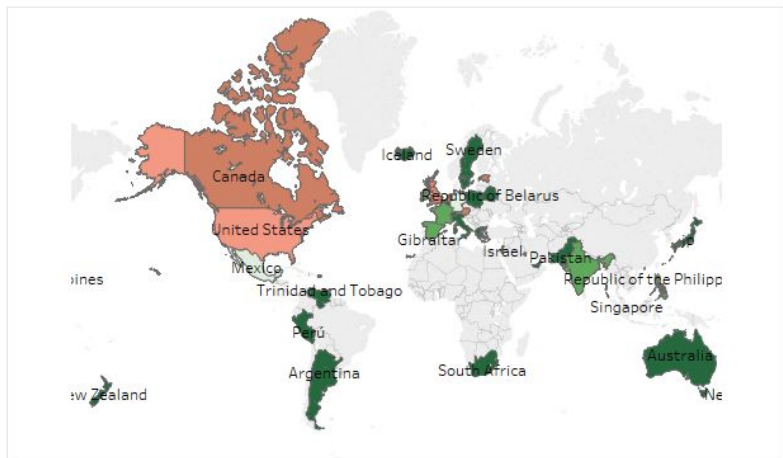
Bubble Analysis



Tree Map



Analysis Of Sentiments Country-Wise





## Conclusion

- 53.17% and 13.49% of tweets exhibit a positive and negative response respectively about BRExit and its impact
- 33.33% tweets are analyzed to be neutral
- Countries like USA, Canada, and parts of Europe including UK itself and Netherlands etc scale more negative responses over positive

