Data Visualisation Assignment 3

Overview: **Student Id**: 17230755

This assignment is to retrieve data from twitter and visually analyse the data (Exploratory Data Analysis). The data retrieval is done through R Twitter Tool box and Processed the data in python to match the structure of data required for high chart and leaflet.js.

Selected Tag*: #IrishWater – Its water utility company in Ireland.*

Twitter Data Extraction:

I have used R to extract the twitter data – using the library twitter and base64enc. Extracted a total of 700 tweets corresponding to the tag #IrishWater. Twitter Data Extraction Code is also included in zip file.
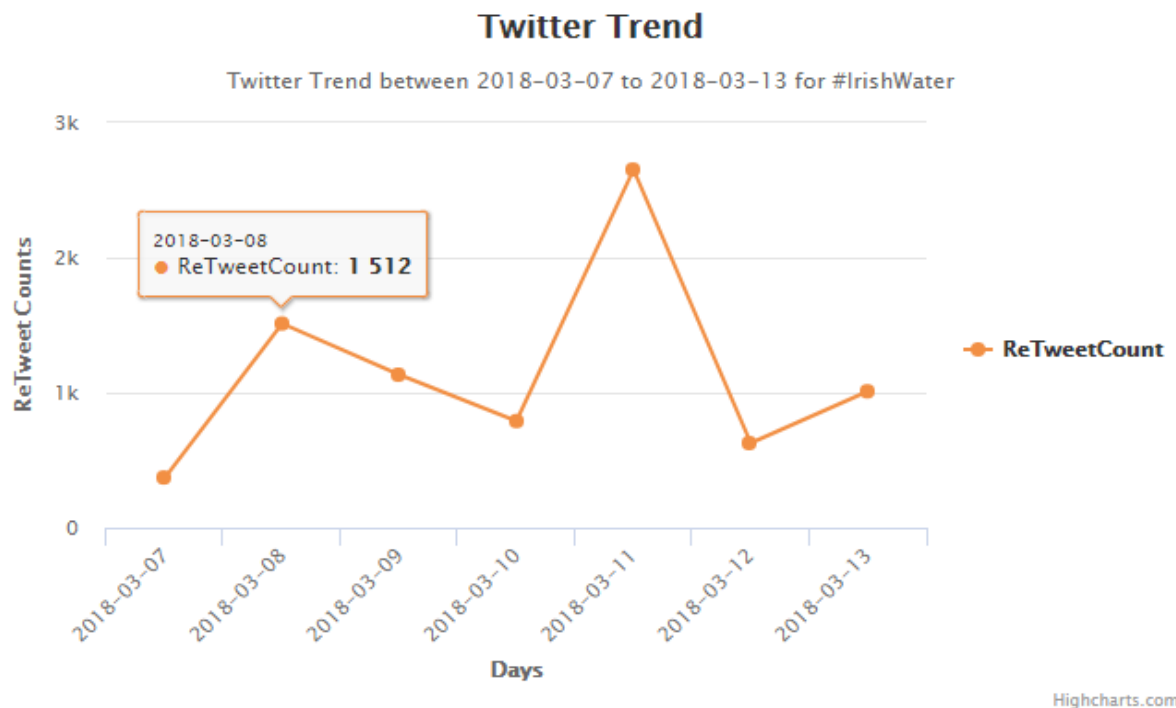
Data Pre - Processing:

Several stages are involved in this stage depending on the chart used. Because all the charts need data in specific format. Hence, pre-processing is done in python. Python code is also included in zip folder.

Extracted Data:

1. *Extracting the Retweet Count group by date – This allows us to capture the trend for a given tag #IrishWater for a given period. Here, all the tweet data are from 2018 Mar 07 to 2018 Mar 13.*
2. *Extracting the Top 20 tweets and its corresponding text is captured. Top 20 tweets are calculated based on the no of shares or re tweet for a given hash tag.*
3. *Extracted Tweets count for a selected counties and rest of the counties is mentioned as others and calculated the percentage.*
4. *Extracted the user name (ScreeName) who is having most followers. This is used in word cloud to display the users.*
5. *Collected the external geojson for polygon representation of Ireland Counties.*
6. *Extracted top 50 tweets to show the marker and its location in map. I have not mapped every tweet in the map is because map would be crowded with markers. Hence selected only top 50 tweets.*
7. *Calculated the percent of tweets shape for few counties and rest is mentioned as others.*

Visualisation / Graphs:

Twitter Trend

Twitter Trend between 2018-03-07 to 2018-03-13 for #IrishWater

2018-03-08
● ReTweetCount: 1 512

— ReTweetCount

Highcharts.com

Description - Fig1:

This is a line chart describing the trend in #IrishWater based on Re tweets from 2018-03-07 to 2018-03-13.

Interpretation:

Made Title, x-label and y-label as bold for better visibility. In addition, added a subtitle and legend for understanding. I have selected this graph because of its simplicity and can understand easily. Moreover, I believe this is the better option for understanding the trend of tweets and how does it vary for a period.
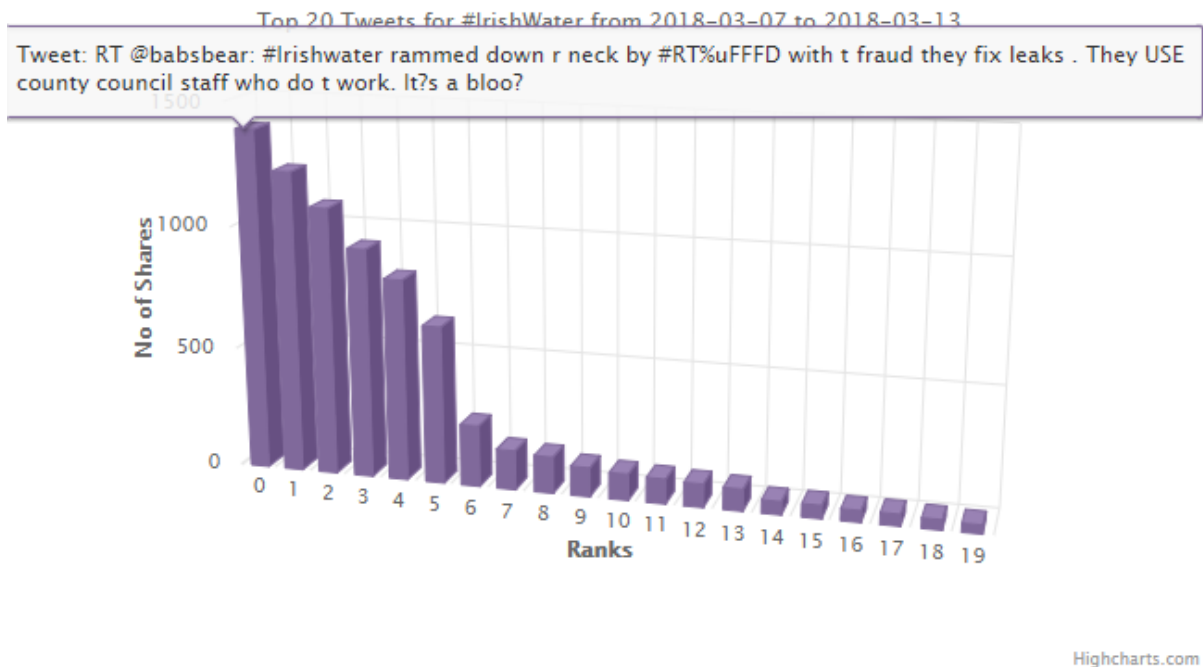
*This graph shows how the Re tweet varies. For date **2018-03-11** has showed highest number of tweets among the 7 days which indicates some news/some changes in Irish Water agency.*

There was a news regarding Water shortage on March 8 2018.

https://www.independent.ie/business/irish/water-shortage-shows-why-we-must-turn-on-investment-tap-and-fix-leaky-system-36680958.html

***Note**: Quality of graph is poor because I have taken a screenshot.*

Top 20 Tweets

Top 20 Tweets for #IrishWater from 2018-03-07 to 2018-03-13

Tweet: RT @babsbear: #Irishwater rammed down r neck by #RT%uFFFD with t fraud they fix leaks . They USE county council staff who do t work. It?s a bloo?

Highcharts.com

Description - Fig2:

Interactive Bar Chart Indicating the top 20 tweets and tweet descriptions on click based on number of shares.
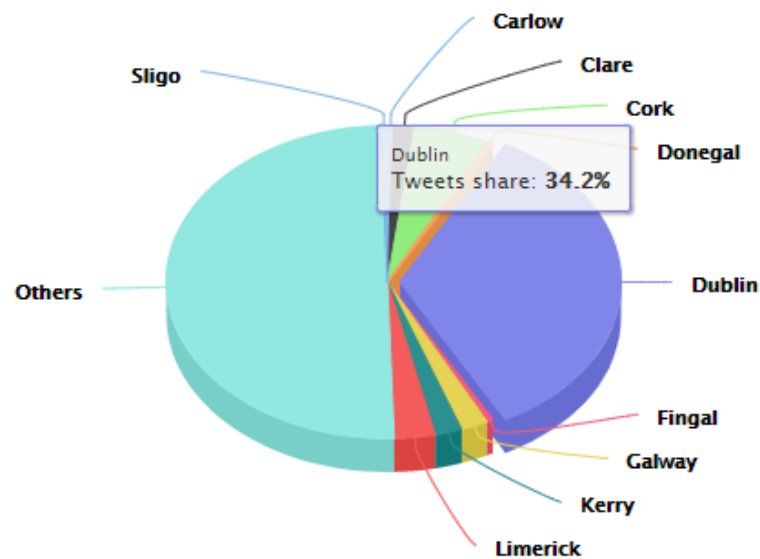
Interpretation:

This bar chart gives the most 20 popular tweets and its tweet description on mouse hover. The x axis Ranks – 0 indicates most popular tweets till now and 19 indicates least popular. The y axis indicates number of shares.

I have made this graph as 3d for better appealing visualisation. Moreover, it gives the clear size of each bar.

Tweets Share Breakdown

Number of Tweets Breakdown for every county

Dublin
Tweets share: 34.2%

Highcharts.com

Description - Fig3:

Interactive Pie Chart Indicating the Number Tweet Share from selected County and rest of the county is mentioned as Others.

Interpretation:

I have taken few counties – Galway, Dublin, Carlow and calculated the percentage of tweets from these counties. Note: Rest all counties comes under others. It is observed from the graph that around 34% tweets for #IrishWater comes from Dublin County for a given period. This highest among all the counties.

I have used 3d visualisation technique for better visualisation and on mouse hover it displays percentage of tweets from corresponding locations.

Most Followed Users

Users with highest followers among the displayed users.

Description - Fig4:

Interactive Word Cloud - Mentioning the Most followed Users based on the follower. Note - Displayed words are User Screen Name.

Interpretation:

This word cloud represents the user name (screen name) and on mouse hover displays the number of followers that user has. I believe word cloud is better to represents the users and their corresponding followers than any other chart. Moreover, it also gives user with highest follower among the displayed users.

E.g.: user name – damienmulley has 17498 followers.

Top 50 Tweet Location

Tweet share break down and top 50 tweets location

Tweet: RT @caulmick: #TBT One year on... The water committee agreed, unanimously without a vote, that it is in favour of placing #IrishWater in?
ReTweet Counts: 6
Location: (53.28, -6.43)

Carlow - 0.50%
Clare - 1.49%
Cork - 5.45%
Donegal - 0.50%
Dublin - 34.16%
Fingal - 0.50%
Galway - 1.98%
Kerry - 1.98%
Limerick - 2.97%
Sligo - 0.50%
Others - 50.00%

Leaflet | Data: CSO.ie | Tiles courtesy of OpenStreetMap Sweden — Map data © OpenStreetMap contributors, CC-BY-SA
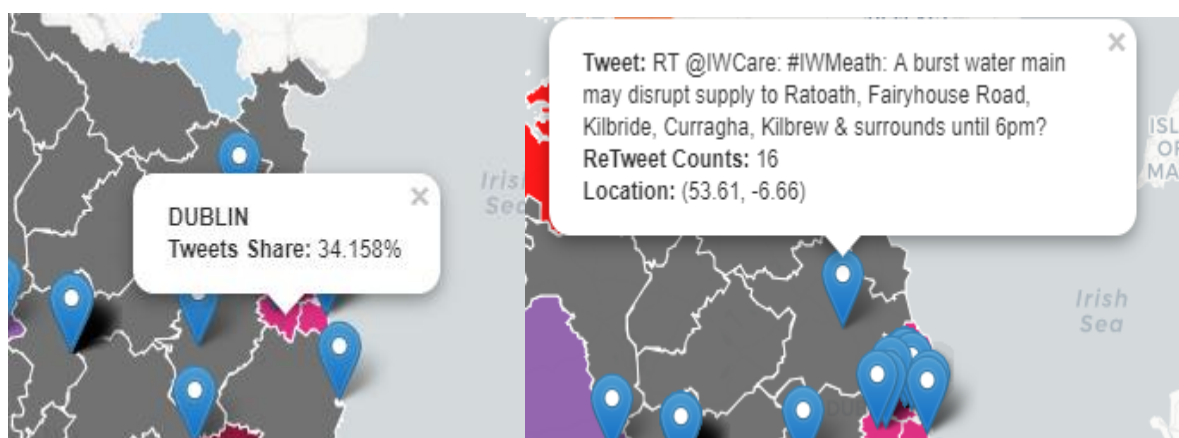
Interpretation:

This map gives the location of top 50 tweets. Also, it gives information about the tweet, its number of shares and users coordinates on click on marker.
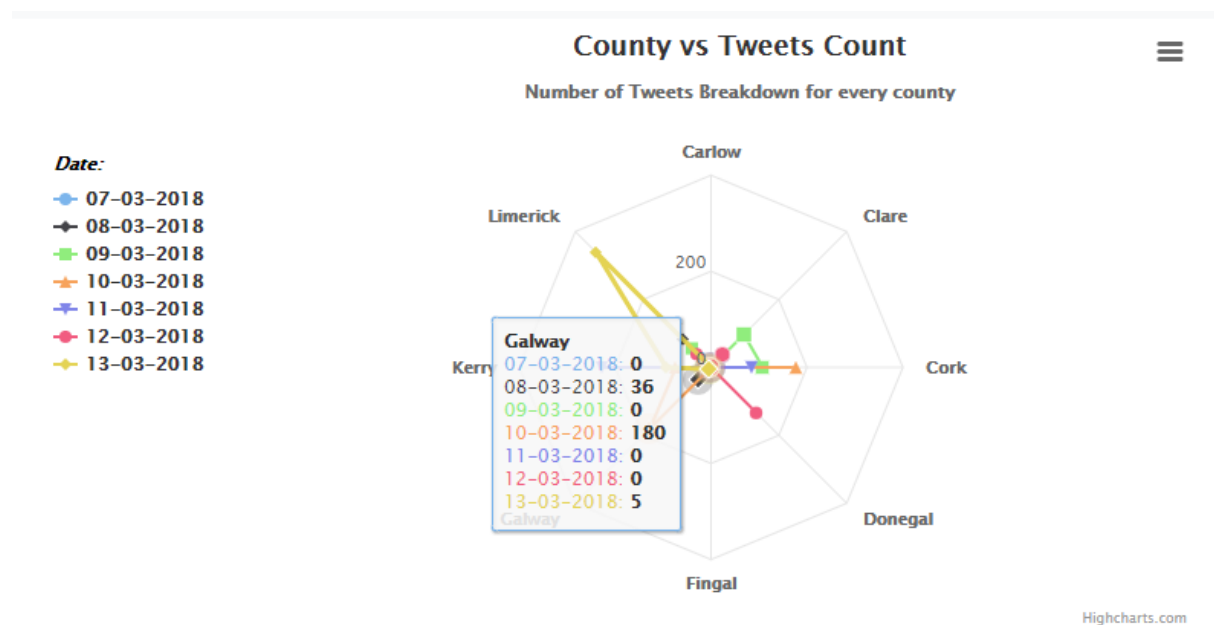
In addition, if we click on polygon it gives tweets share from that corresponding county.

e.g.: If we click on polygon representing Dublin county we get the percentage of tweets from Dublin i.e. 34%. Also, I have changed the map type to suite the white background goes well with the web page.

Note: Majority of grey regions represents others – which includes all counties other than selected.

DUBLIN
Tweets Share: 34.158%

Tweet: RT @IWCare: #IWMeath: A burst water main may disrupt supply to Ratoath, Fairyhouse Road, Kilbride, Curragha, Kilbrew & surrounds until 6pm?
ReTweet Counts: 16
Location: (53.61, -6.66)

Extras:

County vs Tweets Count

Number of Tweets Breakdown for every county

Date:
- 07–03–2018
- 08–03–2018
- 09–03–2018
- 10–03–2018
- 11–03–2018
- 12–03–2018
- 13–03–2018

Carlow

Clare

Limerick

200

Kerry

Galway
07–03–2018: **0**
08–03–2018: **36**
09–03–2018: **0**
10–03–2018: **180**
11–03–2018: **0**
12–03–2018: **0**
13–03–2018: **5**
Galway

Cork

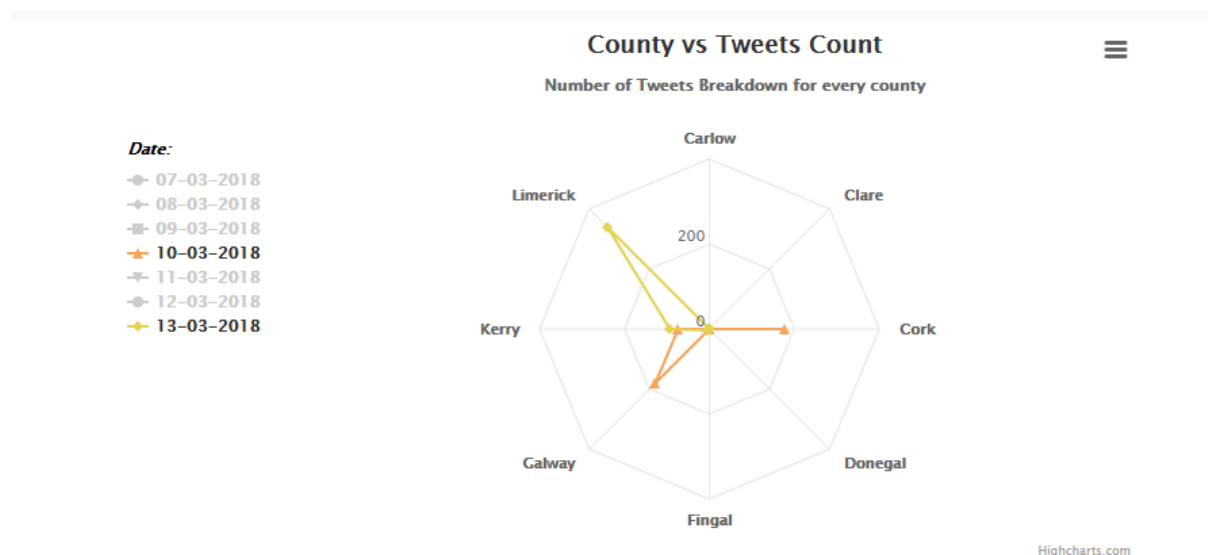Donegal

Fingal

Highcharts.com

<u>Description - Fig1:</u>

This is a interactive spider chart - which describes number of tweets from different counties for a selected date.
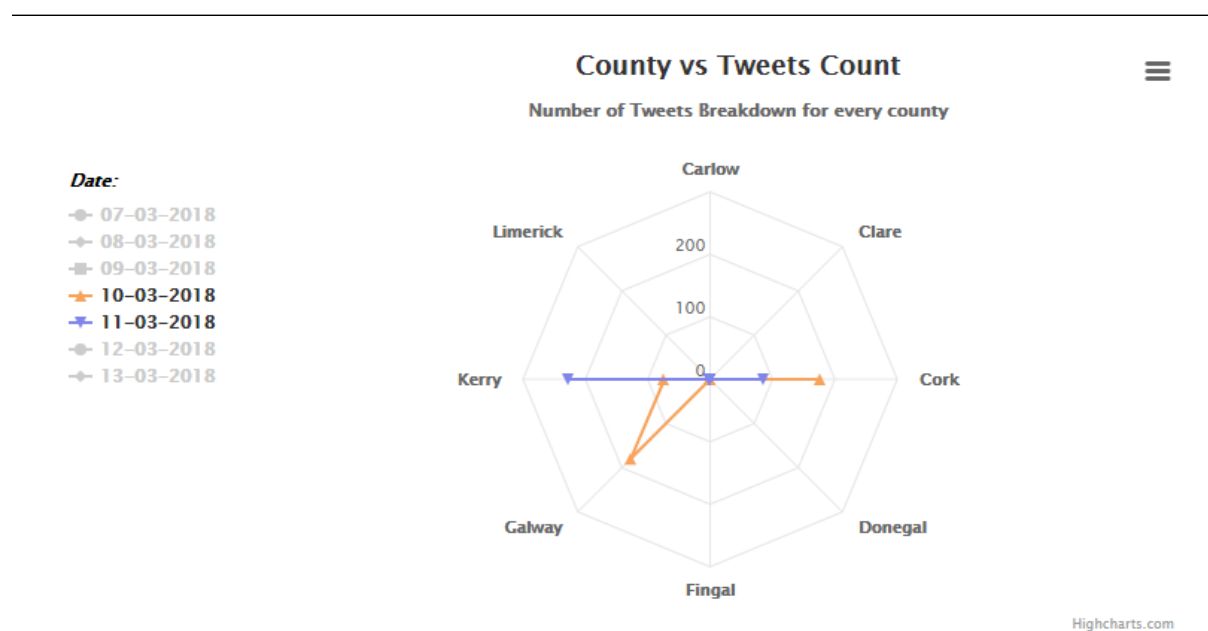_Please click on the Date legend to see the tweets from selected counties for a given day._

Interpretation:

This is an interesting map where we can see the number of tweets for each location for a given date. More over this plot also allows to select the date and tweets corresponding to this date is displayed.

County vs Tweets Count

Number of Tweets Breakdown for every county

Date:
- 07-03-2018
- 08-03-2018
- 09-03-2018
- **10-03-2018**
- 11-03-2018
- 12-03-2018
- **13-03-2018**

Carlow
Clare
Cork
Donegal
Fingal
Galway
Kerry
Limerick

200
0

Highcharts.com

Description - Fig1:

This is a interactive spider chart - which describes number of tweets from different counties for a selected date.
Please click on the Date legend to see the tweets from selected counties for a given day.

County vs Tweets Count

Number of Tweets Breakdown for every county

Date:
- 07-03-2018
- 08-03-2018
- 09-03-2018
- **10-03-2018**
- **11-03-2018**
- 12-03-2018
- 13-03-2018

Carlow
Clare
Cork
Donegal
Fingal
Galway
Kerry
Limerick

200
100
0

Highcharts.com

Description - Fig1:

This is a interactive spider chart - which describes number of tweets from different counties for a selected date.
Please click on the Date legend to see the tweets from selected counties for a given day.