KCON-USA 2015 Twitter Data Analysis

Following are some macro level details about data analysis carried out on the tweets collected around the time period of KCON USA (<http://www.kconusa.com/>) event held in Los Angeles and New York during the months of July and August in 2015. The data collection and analysis was geared towards Korean pop culture and the artists and groups of artists at the centre of K-Pop culture.

## Event Dates:

* LA KCON : July 31, 2015 – August 01, 2015
* NY KCON: August 08, 2015

## Duration of data collection:

* Total : July 27, 2015 – August 09, 2015
* for LA KCON : July 27, 2015 – August 04, 2015
* for NY KCON : August 04, 2015 – August 09, 2015

## Size of data collected:

Total tweets collected : 916,031 [17 GB]

LA KCON : 469,510 [09 GB]

NY KCON : 446,521 [08 GB]

## Tools and Technologies used:

* Platform : Windows / Ubuntu
* Language : Python 3.4 / Python 2.7
* Tweet collection : Twitter's Public Streaming API (https://dev.twitter.com/streaming/public/)
* Libraries used for data analysis:
  + Pandas
  + NLTK (Natural Language Processing)
  + Text Blob (for sentiment analysis) (<https://textblob.readthedocs.org/en/dev/>)
  + Plotly (for plotting graphs) (https://plot.ly/)

## Analysis carried out:

* Tracking tweet volume trend per hour for
  + Total tweets related to K-POP in general
  + Total tweets specifically about KCON
  + Tweet volume comparison for groups of K-POP artists
* Sentiment analysis
  + Plotting sentiment polarity per hour for various artists
* Word Cloud analysis
  + Visualizing the top keywords used during the entire period
* Tweet volume analysis for user location based on "time zone" attribute of a twitter user's profile