

Sentiment Analysis

...

of the twitter responses to the
recently Korean summit

Jimmy Pulido

27th April, 2018 inter-Korean Summit



Polarization of responses:

Example of a positive response:

“반드시 통일되는 날이 올거야”

Seguramente llegará el día de la reunificación

Example of a neutral response:

“지금은 통일이 힘듭니다.”

La reunificación será difícil.

Example of a neutral response:

“미팅끝나는 동시에 남침 ”

Al tiempo de terminar la reunión habrá una gran invasión al sur.

but... How to start?

KOSAC: A Full-fledged Korean Sentiment Analysis Corpus

Type: direct-explicit, direct-action...

Subjectivity-type: emotion-{pos, neg, complex, neutral}, argument

Polarity positive, negative, complex, neutral

Intensity: low, medium, high

Hayeon Jang, Munhyong Kim, and Hyopil Shin
Department of Linguistics, Seoul National University
Gwanak-no Gwanak-gu, Seoul 151-741

```
anchor: morpheme id(s)
id: tag id
nested-source: w-(morpheme id(s)
               |implicit|out)-...-
               (morpheme id(s)|implicit|out)
target: morpheme id(s)
type: direct-explicit,direct-speech,
      direct-action,indirect,
      writing-device
subjectivity-type: emotion-{pos,neg,
                           complex,neutral},judgment-{pos,
                           neg,complex,neutral},argument-
                           {pos,neg,complex,neutral},
                           agreement-{pos,neg,neutral},
                           intention-{pos,neg},
                           speculation-{pos,neg}, others
polarity: positive,negative,complex,
          neutral
intensity: low,medium,high
```

Table 1: The list of SEED tag attributes

Table of Expressive and Subjectivity Type

Emotion	ratio	Agreement	ratio	Argument	ratio
두렵- twulyep- 'fear'	1.00	합의하- hapuyha- 'agree'	0.86	주장하- cwucangha- 'insisit'	0.98
분노 pwunno 'anger'	0.93	인정하- incengha- 'admit'	0.90	지적하- cicekha- 'point out'	0.90
사랑하- salangha- 'love'	0.94	반대하- pantayha- 'disagree'	1.00	제시하- ceysiha- 'suggest'	0.82
행복하- hayngpokha- 'happy'	0.94	거부하- kepwuha- 'deny'	0.90		
Intention	ratio	Judgement	ratio	Speculation	ratio
-고 싶- -ko siph- 'want'	0.88	인기 inki 'popular'	0.87	-는 것 같- -nun kes kath- 'might'	0.50
-기 위하- -ki wiha- 'purpose'	0.63	재미 caymi 'fun'	0.59	-을 것 -ul kes 'would'	0.20
-도록 -tolok 'purpose'	0.52	중요하- Cwungyoha- 'important'	0.90	예상 되- yeysang toy- 'expected'	1.00
예정 yeyceng 'plan'	0.61	풍부하- Phwungpwuha- 'plentiful'	0.91		

Simple Twitter Scraper - JavaScript

```
simple-twitter-scraper <keyword> <first date(YYYY-MM-DD)> <last  
date(YYYY-MM-DD)> <output directory>
```

```
{  
  "id": "123456789012345678",  
  "permalink":  
    "https://twitter.com/username/status/123456789012345678",  
  "username": "username",  
  "text": "@abcd @efgh #This is #tweet text",  
  "datetime": "2017-01-01 11:11",  
  "retweets": 30,  
  "favorites": 20,  
  "mentions": "@abcd @efgh",  
  "hashtags": "#This #tweet",  
  "geo": ""  
}
```

KoNLpy - Korean NLP in Python

Kkma POS tagging: <http://kkma.snu.ac.kr/>

KOSAC Corpus used Kkma so as to tag ngrams

ngram, freq, COMP, NEG, NEUT, None, POS, max.value, max.prop

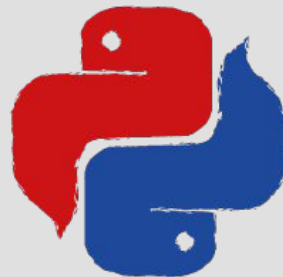
가슴/NNG,5,0.2,0.2,0.2,0,0.4,POS,0.4

가슴/NNG;은/JX,1,0,0,0,0,1,POS,1

가슴/NNG;은/JX;녹아내리/VV,1,0,0,0,0,1,POS,1

가슴/NNG;을/JKO,1,0,1,0,0,0,NEG,1z

가슴/NNG;을/JKO;치밀/VV,1,0,1,0,0,0,NEG,1



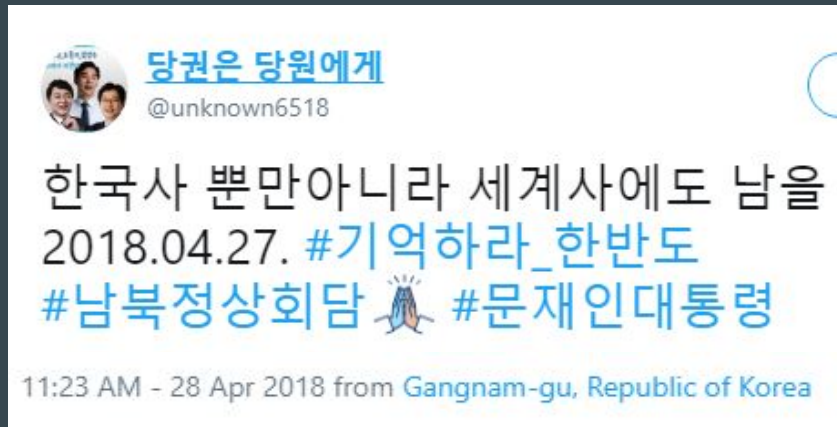
KoNLpy

Coding....

Json file of a already analyzed tweet:

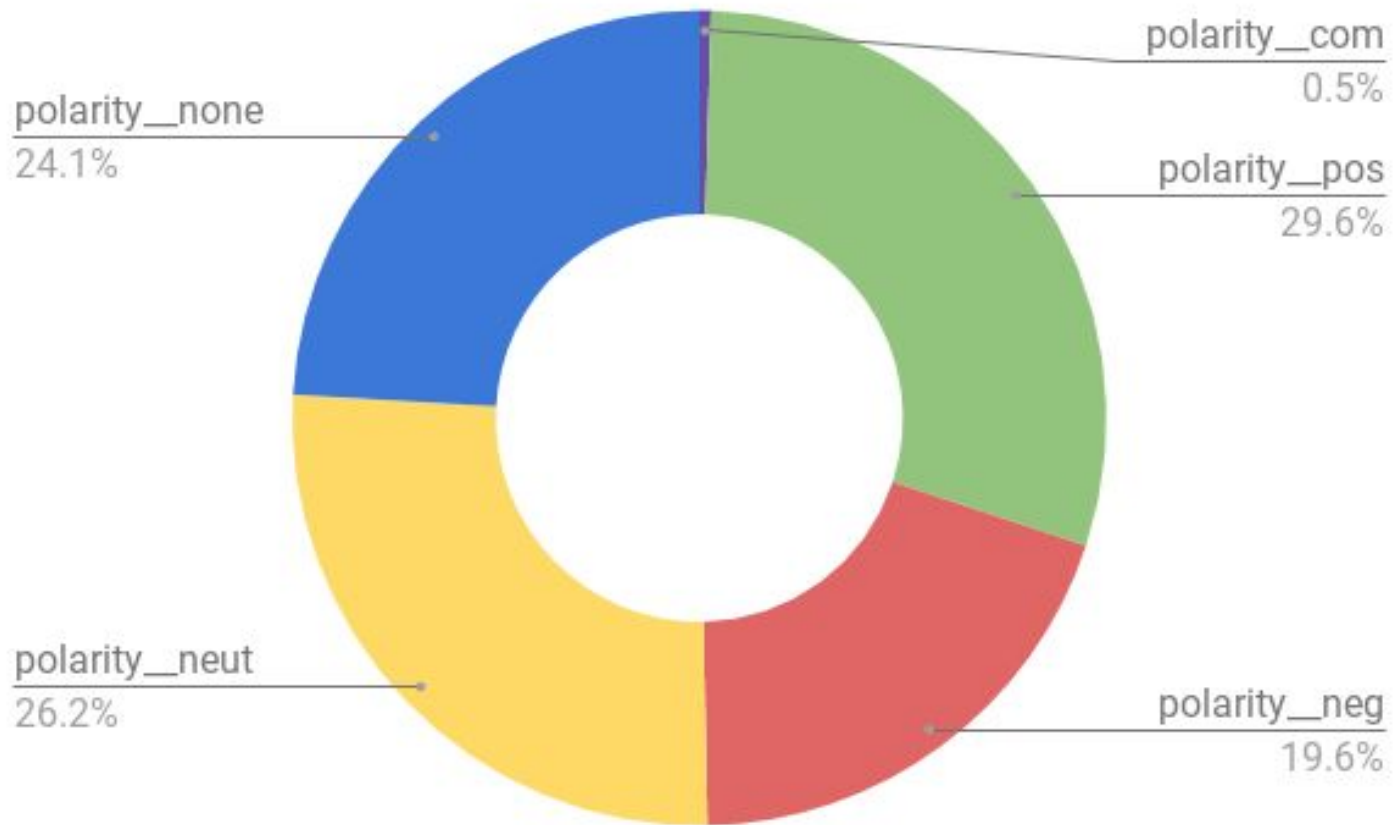
```
{'id': '990191034377236480', 'permalink':  
'https://twitter.com/eechu/status/990191034377236480', 'username': 'eechu', 'text':  
'혼자 뉴스 보면서 헤벌레 웃고있다. 누가 보면 바보같아 보이겠지? ≡  
#남북정상회담', 'date': '2018-04-28 06:28', 'retweets': 0, 'favorites': 1, 'mentions': '',  
'hashtags': '#', 'geo': '', 'raw': '혼자 뉴스 보면서 헤벌레 웃고있다. 누가 보면  
바보같아 보이겠지? ≡ #남북정상회담', 'polarity': {'com': 0.14880952383588436,  
'pos': 0.2827380952453231, 'neg': 0.25595238096343537, 'neut':  
0.3065476189562074, 'none': 0.005952380999149658}, 'intensity': {'high':  
0.03273809528571429, 'low': 0.2470238094285714, 'meium':  
0.7142857142857142, 'none': 0.005952381}, 'expressive': {'dir-action': 0,  
'dir-explicit': 0.1369047618767007, 'dir-speech': 0.47321428578188784, 'indirect':  
0.3779761904825681, 'writing-device': 0.011904761858843538}}
```

Some “random” tweets results: Positive sample:



“No solo la historia de Corea, sino también la historia mundial”

polarity__com	polarity__pos	polarity__neg	polarity__neut	polarity__none
0.00469397	0.29597454	0.19615329	0.26239096	0.24078721
96	6	62	23	58

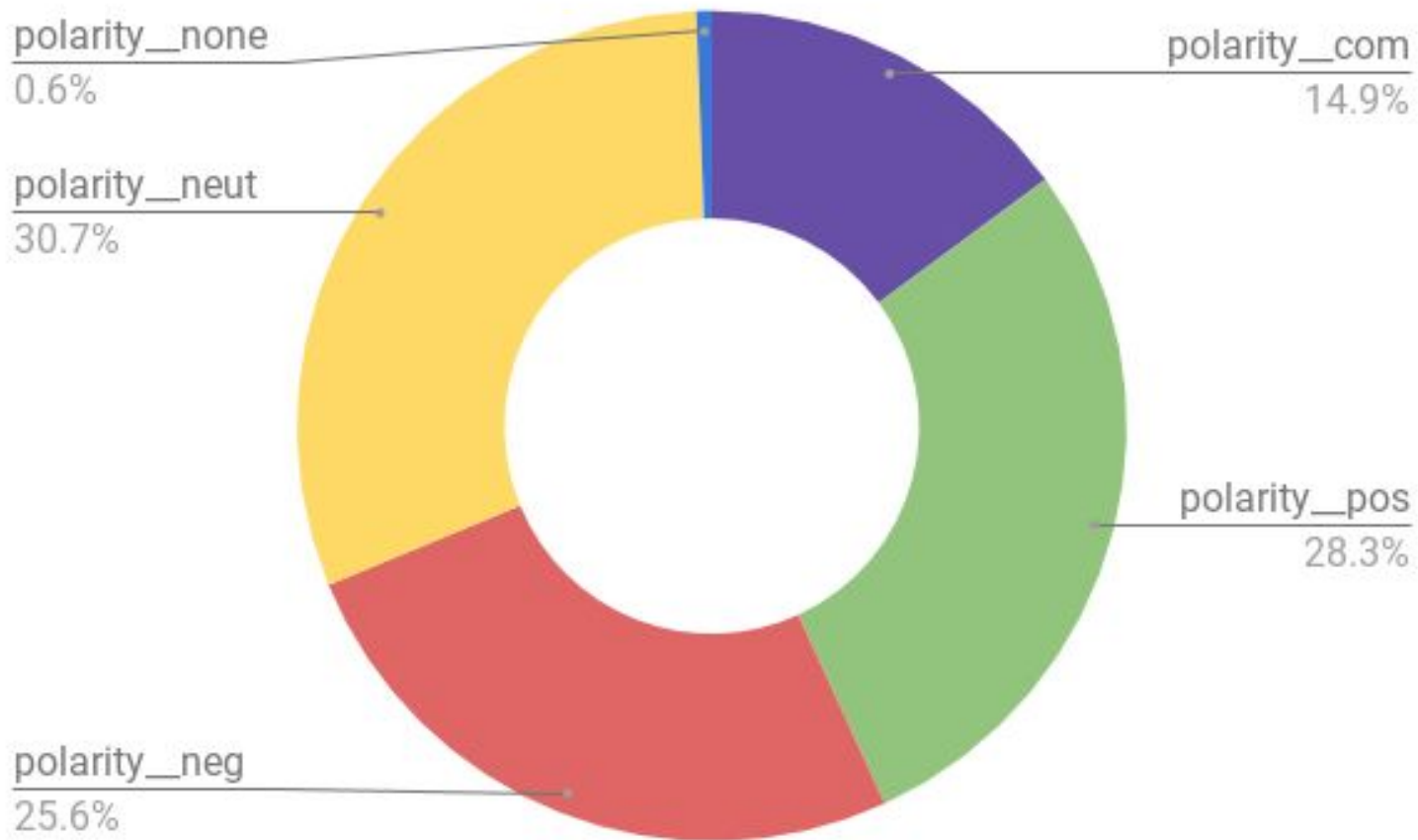


Neutral sample:



Me río mientras veo las noticias solo. ¿Quién se vería estúpido?
jaja

polarity__com	polarity__pos	polarity__neg	polarity__neut	polarity__none
0.1488095238	0.2827380952	0.255952381	0.306547619	0.005952380999



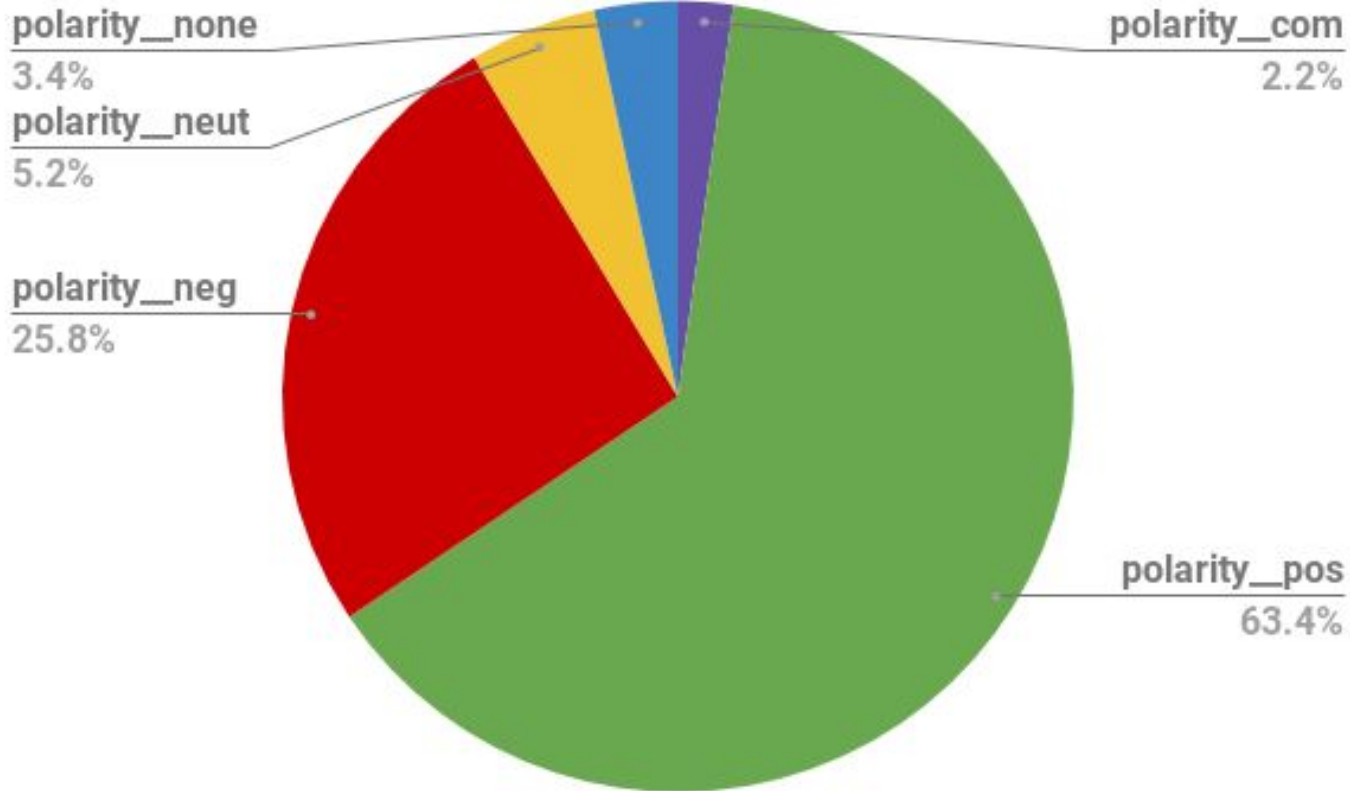
Negative sample:



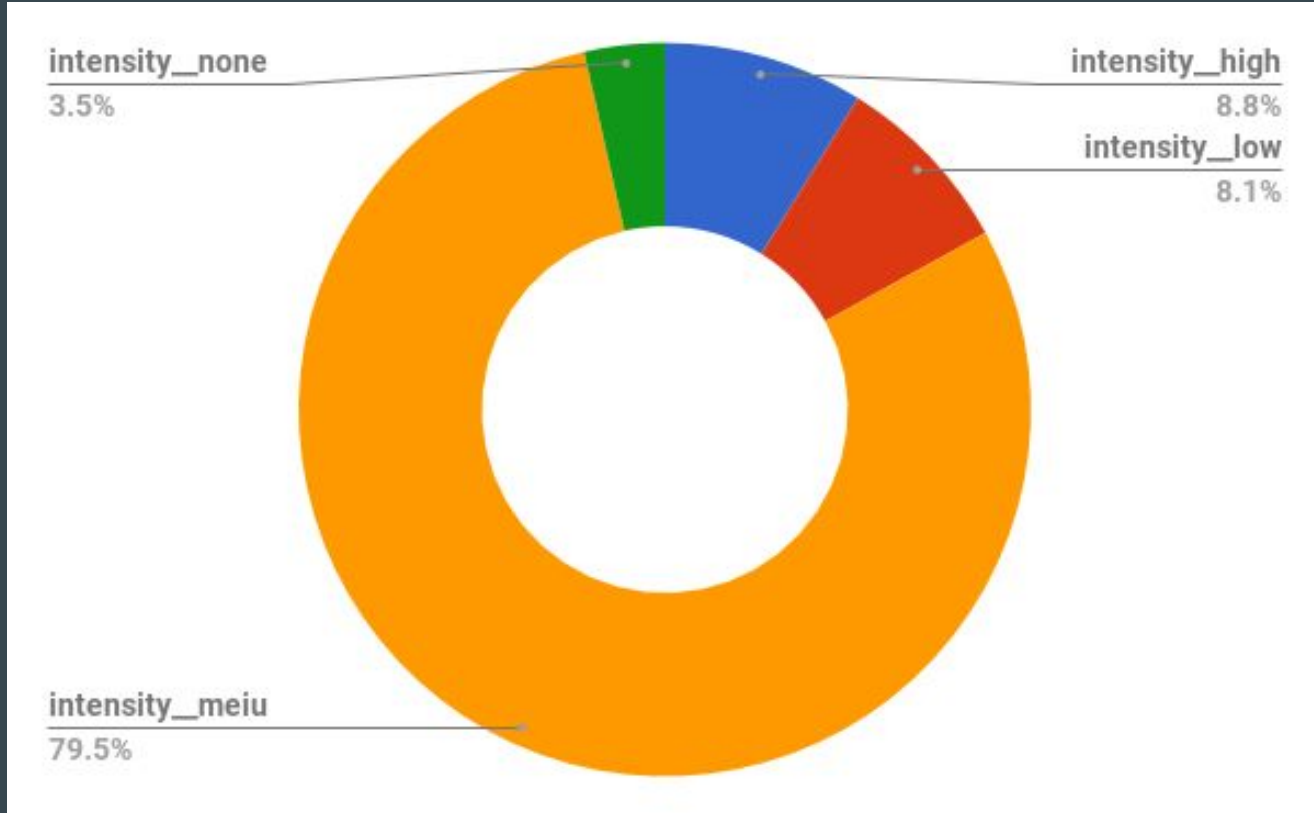
Que
miedo

100% negative polarity

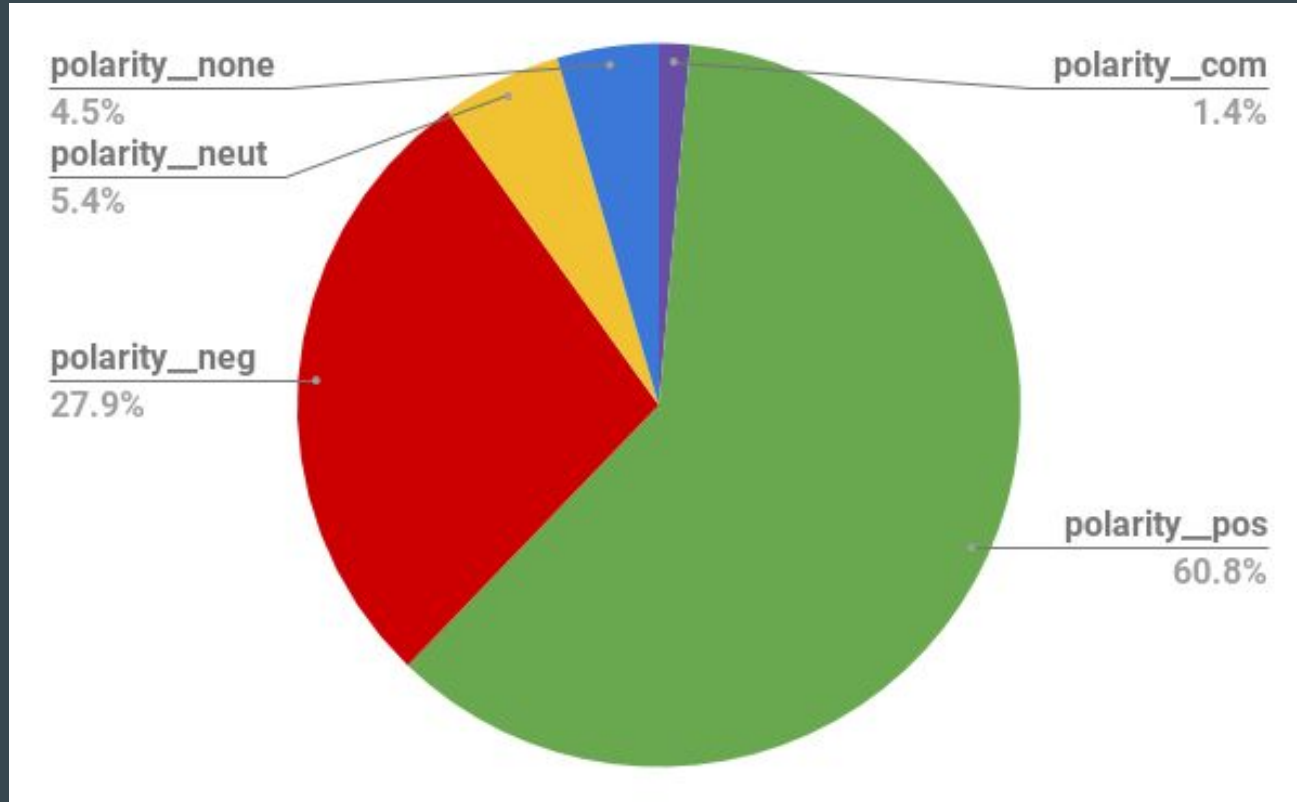
Polarity Average of a population of 3000 tweets



Intensity Average of a population of 3000 tweets



Weighted average of a population of 37596 tweets



Positive Nouns:



Negative Nouns:

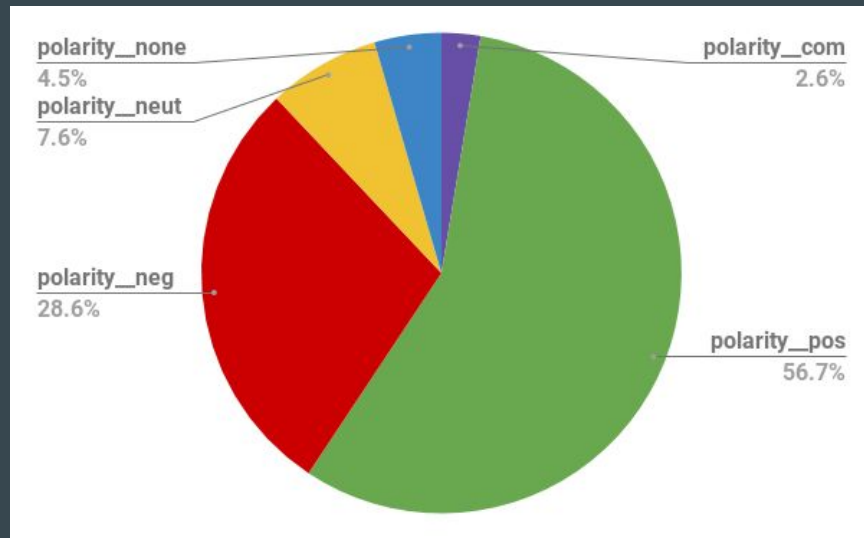
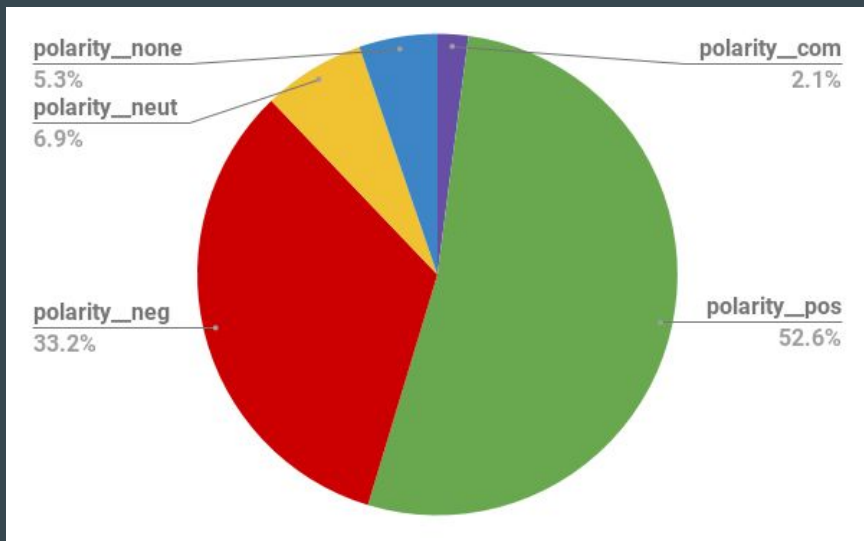


Geolocated data

Seoul

vs

Busan



Geolocated data

Seoul

vs

New York

