SOCIAL MEDIA ANALYTICS

Case Project

Abstract
Analysis on the tweets data of the Pakistani and Indian Top news channels to have a view on their reaction towards the #Kashmir Issue

Osman Manzoor Ahmed - 721347
osman.ahmed@aalto.fi
Aalto University

Contents

Research Question:

Q: How the News channels from Pakistan and India have reacted towards the #Kashmir issue?

Background

Kashmir is a territorial conflict between India and Pakistan. In August 2019, Indian government revoked Article 370 from the Indian constitution that granted a special status to the Kashmir Valley. Following this move, both the countries came on the verge of war. ([Source](#))

The purpose of this analysis is to gather a view on how the Media channels from both countries are reacting to this issue.

Data Collection:

The data was collected using [TAGS](#) which is a free Google sheet template to get twitter data. The data was collected for **#kashmir** hashtag for the top News channels from both the countries (India and Pakistan).

The list of channels (10) from Pakistan and the script that was ran on TAGS is:

- ARYNEWSOFFICIAL
- BOLNETWORK
- dawn_com
- Dawn_News
- DunyaNews
- ExpressNewsPK
- geonews_english
- gnnhdofficial
- humnewspakistan
- PTVNewsOfficial

```
(#kashmir AND From:geonews_english) OR (#kashmir AND from:DunyaNews) OR
(#kashmir AND from:ARYNEWSOFFICIAL) OR (#kashmir AND from:PTVNewsOfficial)
OR (#kashmir AND from:BOLNETWORK) OR (#kashmir AND from:ExpressNewsPK) OR
(#kashmir AND from:dawn_com) OR (#kashmir AND from:Dawn_News) OR (#kashmir
AND from:humnewspakistan) OR (#kashmir AND from:gnnhdofficial)
```

The list of channels (9) from India and the script that was ran on TAGS is:

- aajtak
- ABPNews
- CNNnews18
- IndiaToday
- indiatvnews
- MirrorNow
- ndtv
- TimesNow
- WIONews

```
(#kashmir AND from:WIONews) OR (#kashmir AND from:Indiatvnews)OR (#kashmir
AND from:Aajtak) OR (#kashmir AND from:ABPNews) OR (#kashmir AND
from:IndiaToday) OR (#kashmir AND from:CNNnews18) OR (#kashmir AND
```

```
from:Ndtv) OR (#kashmir AND from:MirrorNow) OR (#kashmir AND from:TimesNow)
OR (#kashmir AND from:Republic)
```

The data set was collected for a period of 2 month starting from **28 september** till **27 november**.

Analysis and Findings

The analysis of the collected data was done on **Power BI** and **jupyter notebook** that uses the python language.

Power BI

Pakistan Tweets

In Figure 1 we can see the analysis of the Pakistani Tweets that were collected from the Pakistani news channels.

The total number of tweets are **501** for a 60 days count. On the Tweets breakdown section, we can see the number of tweets per individual news channels and we can see that **Dawn_News** is more active than the others with a tweet count of **109.**

In case of outreach, we can examine the number of followers each individual channel has. **ExpressNewsPk** is on top in case of followers with outreach of **2,564,740. Dawn_News** that has the most tweets is on number 4 in terms of its outreach with number **1,480,183.** While **ExpressNewsPk** is at the last spot with 2 tweets in case of Tweet Count which shows that the number one channel of Pakistan in case of outreach has no interest in this matter of Kashmir.

The collective global outreach of all channels combine is **11.36M.**

In case of Temporal analysis, we can see that there is a steady graph and they have been tweeting on the daily basis on this issue. The trend has seem to go down in the month of November but it's a steady one.
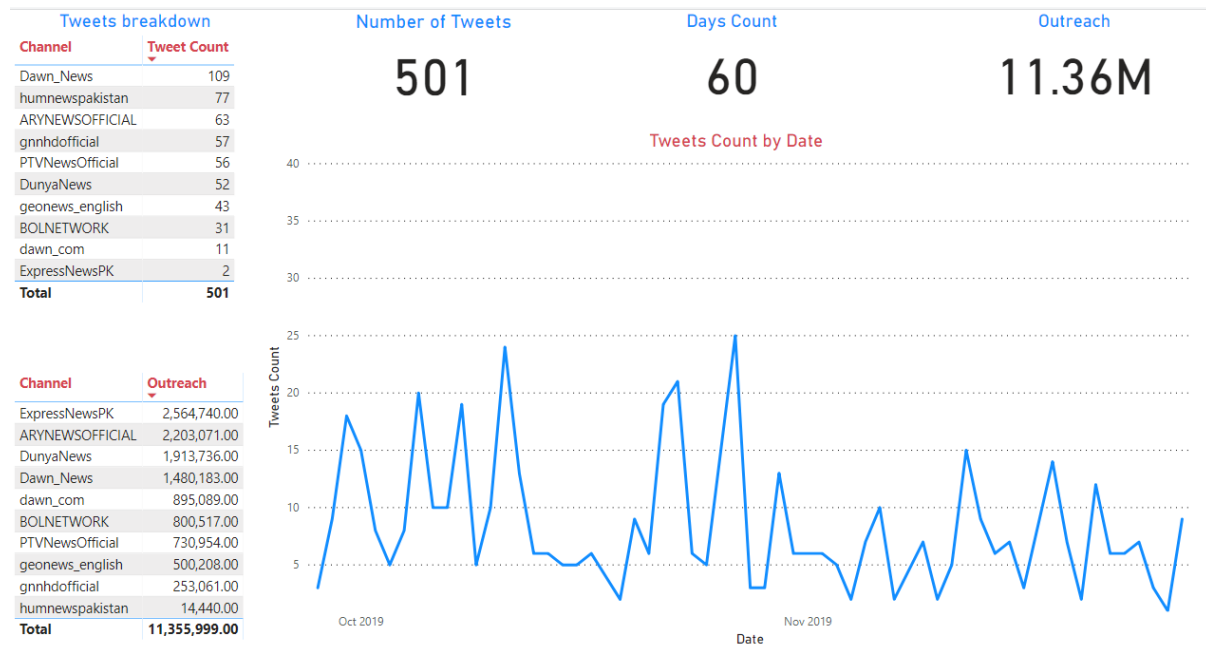
| Tweets breakdown | |
Channel	Tweet Count
Dawn_News	109
humnewspakistan	77
ARYNEWSOFFICIAL	63
gnnhdofficial	57
PTVNewsOfficial	56
DunyaNews	52
geonews_english	43
BOLNETWORK	31
dawn_com	11
ExpressNewsPK	2
Total	**501**

Channel	Outreach
ExpressNewsPK	2,564,740.00
ARYNEWSOFFICIAL	2,203,071.00
DunyaNews	1,913,736.00
Dawn_News	1,480,183.00
dawn_com	895,089.00
BOLNETWORK	800,517.00
PTVNewsOfficial	730,954.00
geonews_english	500,208.00
gnnhdofficial	253,061.00
humnewspakistan	14,440.00
Total	**11,355,999.00**

Figure 1: Analysis of Pakistani Tweets

In Figure 2, we can see the breakdown of news channels on the daily basis to see who has been more active on daily basis and how every channel has been contributing.
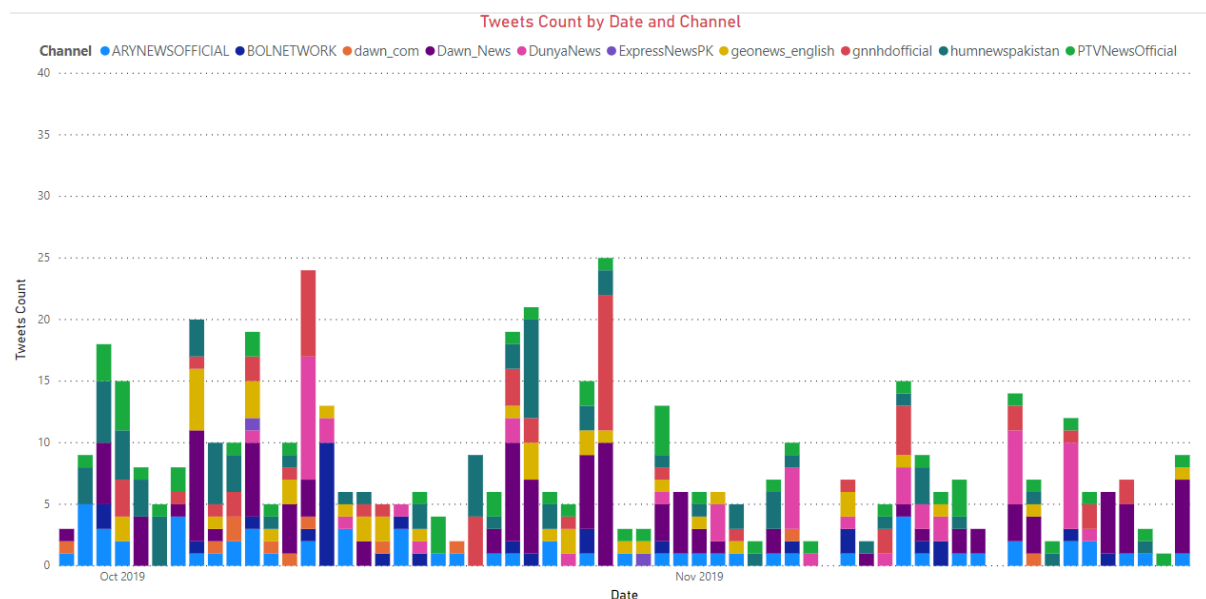

Figure 2: Active channels from Pakistan

Indian Tweets

In Figure 3 we can see the analysis of the Indian Tweets that were collected from the Indian news channels. I was just able to find 9 top news channels for India that were on Twitter.

The total number of tweets are **300** for a 60 days count. On the Tweets breakdown section, we can see the number of tweets per individual news channels and we can see that **IndiaToday** is more active than the others with a tweet count of **111.**

In case of outreach, we can examine the number of followers each individual channel has, **ndtv** is on top in case of followers with outreach of **11,799,652. IndiaToday** that has the most tweets is on

number 5 in terms of its outreach with number **5,277,382.** While **ndtv** is on the 7th spot with 11 tweets in case of Tweet Count which shows that the number one channel of India in case of outreach has no interest in this matter of Kashmir also as he has just tweeted 11 times.

The collective global outreach of all channels combine is **48.59M.**

In case of Temporal analysis, we can see that they have been slightly active till mid of October and the it declined with a boom at the end of October. After that the trend has declined significantly and it seems that Indian channels don't have any interest overall on the Kashmir Issue.
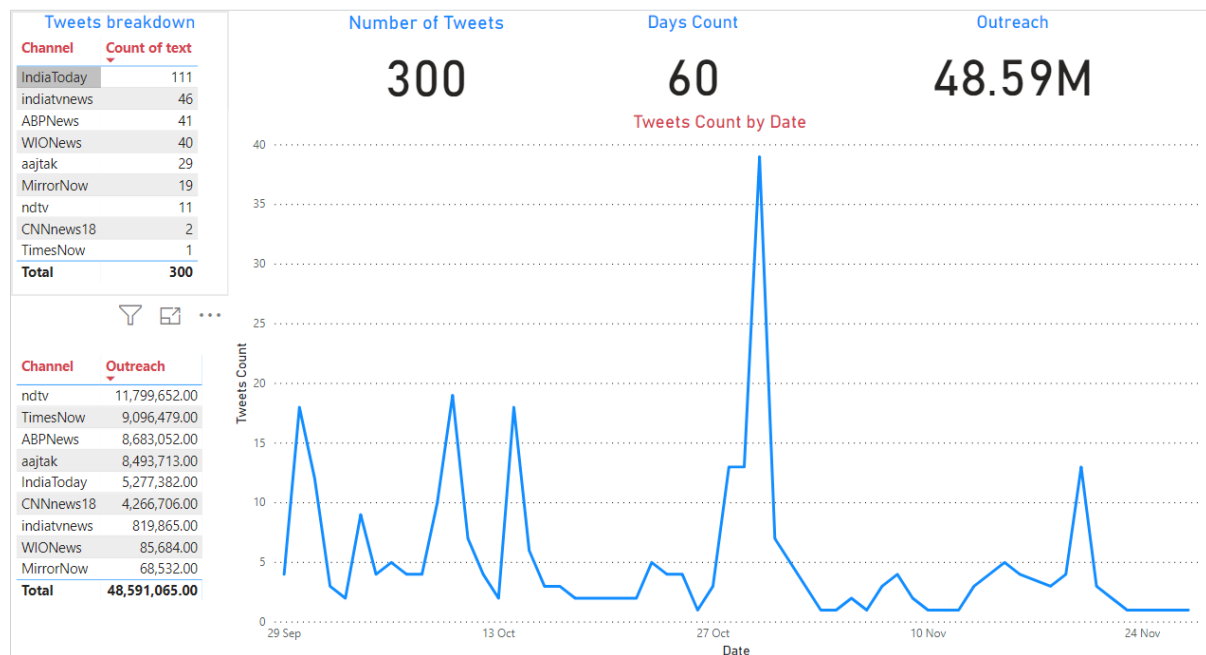

Figure 3: Analysis of Indian Tweets

In Figure 4, we can see the breakdown of news channels on the daily basis to see who has been more active on daily basis and how every channel has been contributing.
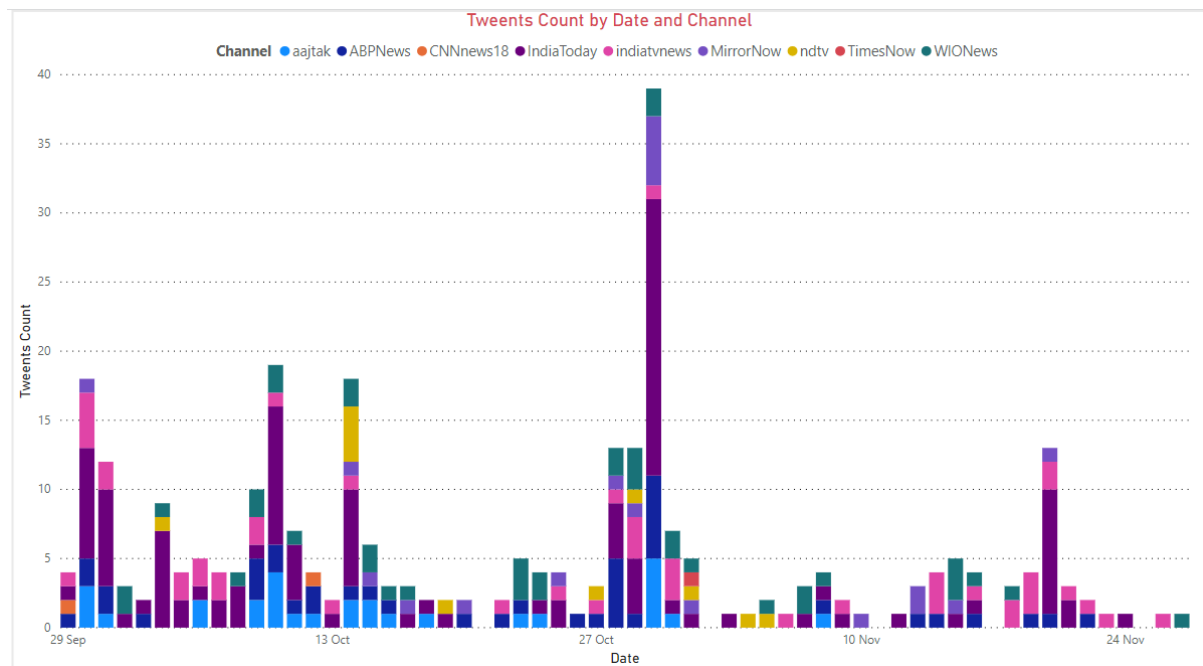
Figure 4: Active channels from India

Comparison between India and Pakistan tweets

In terms of comparison, clearly, we can see that outreach of Indian channels is way bigger than the Pakistani channels. The simple fact ([source](#)) might be that India is the 2nd most populated (1,368,737,513) country in the world after China and Pakistan lies on the 6th number (204,596,442) and the difference is significantly huge between the two countries in terms of population.

In Figure 5, we can see a comparison of retweet and favorite count for both the countries, this analysis will help us to understand the that how the people from both countries are reacting towards this issue and who is more active.

Clearly, we can see that in Retweet count Pakistani channels out beat India with a difference of more than 1000. And the same goes in terms of Favorite count. Even with a very big outreach for Indian channels, we can see that the people of India are not much active on this issue as compared to the Pakistani people.

Data from 04 November 2019			
Pakistan		**India**	

Pakistan

Channel	retweet_count
PTVNewsOfficial	544
ARYNEWSOFFICIAL	410
DunyaNews	102
Dawn_News	83
BOLNETWORK	56
gnnhdofficial	43
geonews_english	23
dawn_com	8
humnewspakistan	2
Total	**1271**

Channel	favorite_count
DunyaNews	672
ARYNEWSOFFICIAL	594
PTVNewsOfficial	351
Dawn_News	269
gnnhdofficial	215
BOLNETWORK	173
geonews_english	94
dawn_com	25
humnewspakistan	10
Total	**2403**

Total Retweet Cont
1271

Total Favorite Cont
2403

India

Channel	retweet_count
IndiaToday	61
ABPNews	47
indiatvnews	44
WIONews	29
aajtak	8
ndtv	8
MirrorNow	4
Total	**201**

Channel	favorite_count
ABPNews	416
IndiaToday	231
indiatvnews	210
WIONews	119
ndtv	63
aajtak	60
MirrorNow	6
Total	**1105**

Total Retweet Count
201

Total Favorite Cont
1105

Figure 5: Retweet and Favourite Count

Even India has the biggest outreach, the number of tweets, retweet count and favorite count by Pakistani channels are higher than the Indian channels. If we focus on the temporal analysis, we see that Pakistani channels has been consistent on day to day tweeting on the matter than the Indian channels who had a significant rise of tweets on a particular day that might be because of some reasons.

Python Analysis

Python was used to make analysis on the tweets text itself. For this purpose, we filtered it down to English Tweets only.

- Pakistan tweets in English: 130
- Indian Tweets in English: 257

Pakistan Tweets

The top 20 words according to their frequency can be seen in Table 1. Main words that can be highlighted are occupi, peopl, pm, militari, lockdown etc.

Word	Freq
occupi	65
read	41
kashmir	29
indian	27
day	24
continu	20
peopl	16
pm	13
militari	12
today	12
issu	12

jammu	11
urg	10
lockdown	10
consecut	9
remain	9
valley	8
enter	8
say	8
india	8

Table 1: Word Frequency Pakistani Channels

The top 20 user mentions can be seen in Table 2. The main mentions are ungeneva, unhumanrights, imrankhanpti, nytimes and realdonaldtrump etc.

User	Freq
ungeneva	3
unhumanrights	3
imrankhanpti	2
nytimes	2
uscirf	2
barribaat	1
commrbhargava	1
dunyablogs	1
humnewspakistan	1
mustafa_mfa	1
realdonaldtrump	1
repmcgovern	1
ruptly	1

Table 2: User Mentions Frequency Pakistani Channels

The top 20 hash tags used can be seen in Table 3. Main hashtags other than Kashmir would be Kashmirbleeds, unga, blackday and kashmirstillundercurfew etc.

Hash Tag	Freq
kashmir	130
arynews	54
geonews	40
india	17
kashmirbleeds	8
indian	7
pakistan	7
iok	6
us	5
pmimran	4
malaysia	4
unga	3
blackday	3
china	3

kartarpurcorridor	2
modi	2
eu	2
kashmirstillundercurfew	2
un	2
turkey	1

Table 3: Hash Tags Frequency Pakistani Channels

Indian Tweets

The top 20 words used based on frequency can be seen in Table 4. In Indian tweets, the main words that can be highlighted are watch, issu, eu and deleg etc

Word	Freq
watch	57
visit	48
kashmir	45
issu	34
india	34
live	31
report	26
valley	26
video	26
thi	24
ha	24
pm	24
day	23
eu	20
say	19
mp	17
deleg	16
detail	16
servic	16
today	15

Table 4: Word Frequency Indian Channels

The top 20 user mentions can be seen in Table 5. The only two mentions worth noticing are of narendramodi and imrankhanpti.

User	Freq
palkisu	25
shujauh	15
gauravcsawant	12
ashraf_wani	10
indiatvnews	9
rahulkanwal	9
rajatsharmalive	9
mirfareed2	8
wionews	7

ankit_tyagi01	6
loveenatandon	5
narendramodi	4
amitshah	4
sardesairajdeep	3
shivaroor	3
kamaljitsandhu	3
preetichoudhry	3
imrankhanpti	2
poulomimsaha	2
news18dotcom	2

Table 5: User Mentions Frequency Indian Channels

The top 20 hash tags used can be seen in Table 6. Main hashtags would be india, pakistan, article370, jammuandkashmir, indiafirst and europeanunion etc.

Hash Tag	Freq
kashmir	256
india	24
pakistan	24
itvideo	21
article370	17
gravitas	15
jammuandkashmir	14
aajkibaat	9
newstrack	9
indiafirst	8
turkey	7
newstoday	7
reporterdiary	6
europeanunion	6
amitshah	6
eu	5
itlivestream	5
srinagar	5
china	5
exclusive	5

Table 6: Hash Tag Frequency Indian Channels

Sentiment Analysis

For sentiment analysis, Python library **textblob** was used to find the polarity of the tweets and then to analyze it using a histogram plot. The sentiment analysis is performed on the English tweets.

Pakistan Tweets

For the Pakistani channel tweets we can see in the figure that there is more on positive side than the negative side. The negative side also holds a significance number but majority of the tweets have positive sentiment. We can see that there are tweets with the extremes on both positive and negative sides.
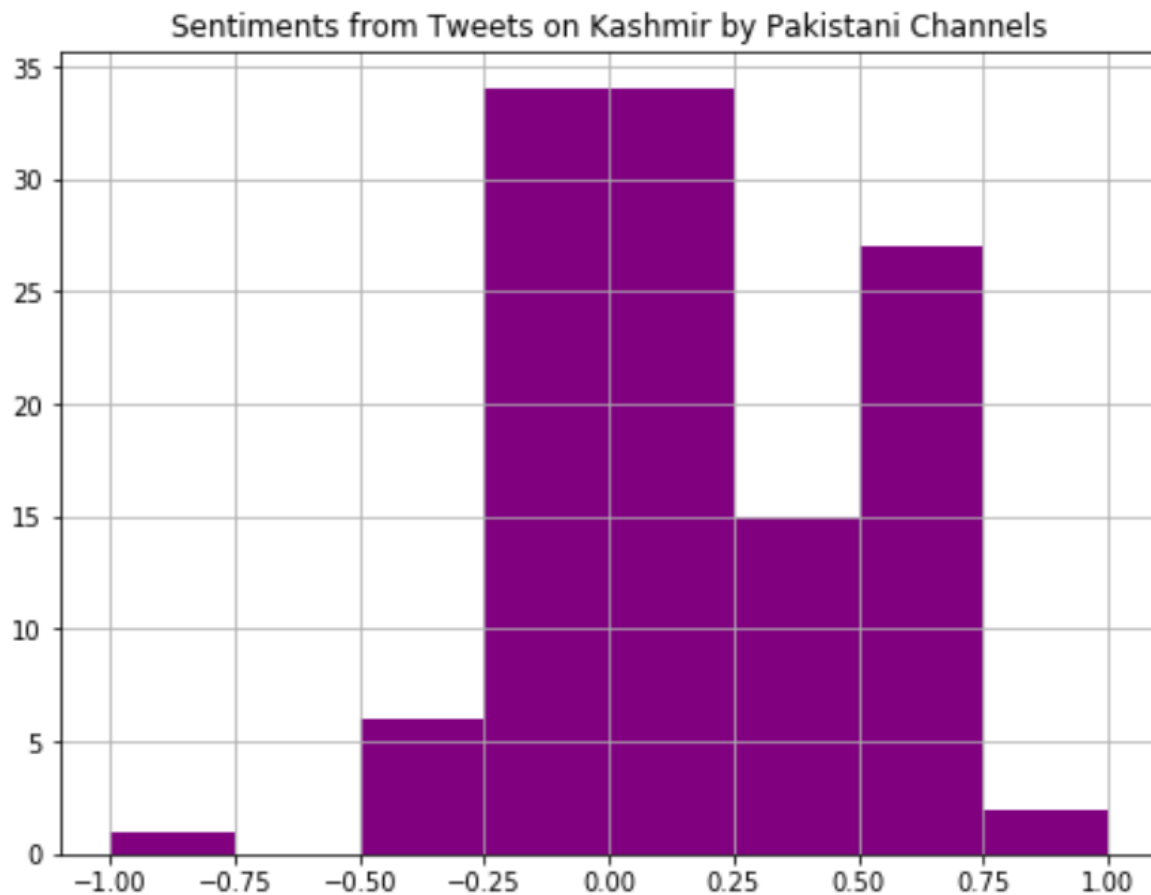
Figure 6: Sentiment on Pakistani Channels Tweets

Indian Tweets

For the Indian channel tweets we can see in the figure that the tweets polarity lies between 0.75 and -0.75 with no extremes. Majority of it is on the positive side and negative side is quite small in numbers.
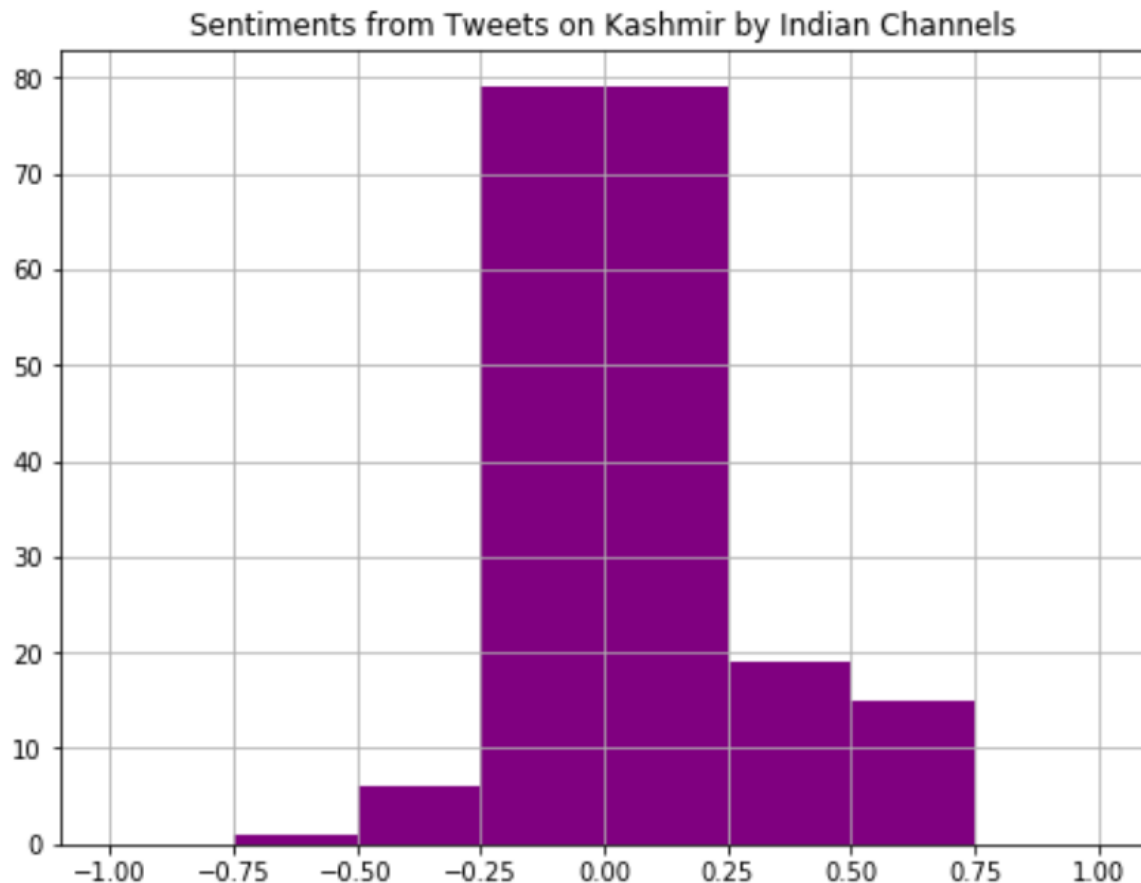
Figure 7: Sentiment on Indian Channel Tweets

Conclusion / Implications of the Findings

Q: How the News channels from Pakistan and India have reacted towards the #Kashmir issue?

A: The answer to this question based on the tweets gathered for a time period of 2 months can be summarized as follows:

- The tweets from Pakistani Channels are greater than that of Indian Channels. Meaning that Pakistani channels are more active on the #kashmir issue.
- Irrespective of the greater outreach of the Indian Channels, the retweets and favorite of Pakistani channels are more than that of Indian Channels.
- Both the countries channels are tweeting daily based on the temporal graphs.
- The frequent word counts, user mentions and the Hash Tags of both are entirely different from each other.
- The sentiment analysis shows that both are on the positive side in majority but there are negative sentiments also.

Challenges

The main challenge here is the collection of data, the analysis is done based on 2 months data that was collected on a daily basis. It would have been better if we can collect historical data also.

Second challenge is about the sentiment analysis, we cannot say that it is going to be 100 percent accurate, nor we have any means to analyze that too much extent its true without the involvement of a human being for manual checking.

 Also, the same goes for other libraries. We are analyzing human created text that can contains typos and other mistakes that the libraries will not be able to get and this makes this kind of analysis very difficult in case we have raw data.