<u>Report</u>

<u>Question 1:</u>

1. **Class A → #RestInPeace:**
   - Number of tweets and number of users for class : 416
   - Number of verified users: 6

2. **ClassB → #MissYouYovi**
   - Number of tweets and number of users for class : 5664
   - Number of verified users: 34
3. **ClassC → #MissYouYuvi** and **#RestInPeace**
   - Number of tweets and number of users for class : 269
   - Number of verified users: 0

It can be seen that there exists no verified user in case of class C. This might be due to the fact that they used both tags in order to make their tweet trending.

<u>Question 2:</u>

Consider time in hours since date is same in all the tweets.

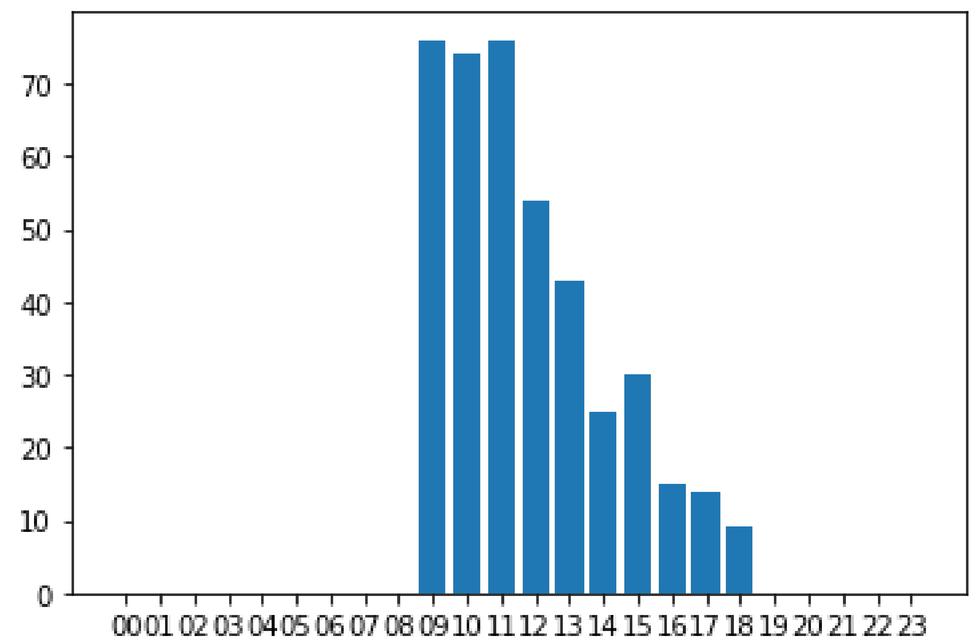
1. #RestInPeace: Make a list storing 24hours and and number of tweets in an hour.

   Time in hours:
   ```
   {'a': ['00', '01', '02', '03', '04', '05', '06', '07', '08',
   '09', '10', '11', '12', '13', '14', '15', '16', '17', '18',
   '19', '20', '21', '22', '23']}
   ```

   Tweets in hours:

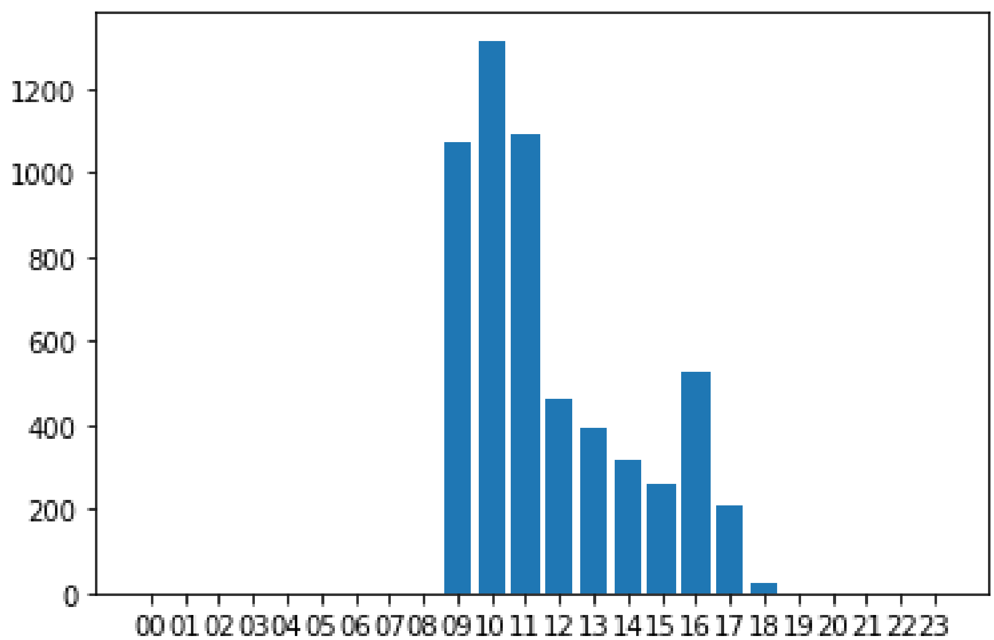
   ```
   {'a': [0, 0, 0, 0, 0, 0, 0, 0, 0, 76, 74, 76, 54, 43, 25, 30,
   15, 14, 9, 0, 0, 0, 0, 0]}
   ```

`Plot(time vs tweets)`

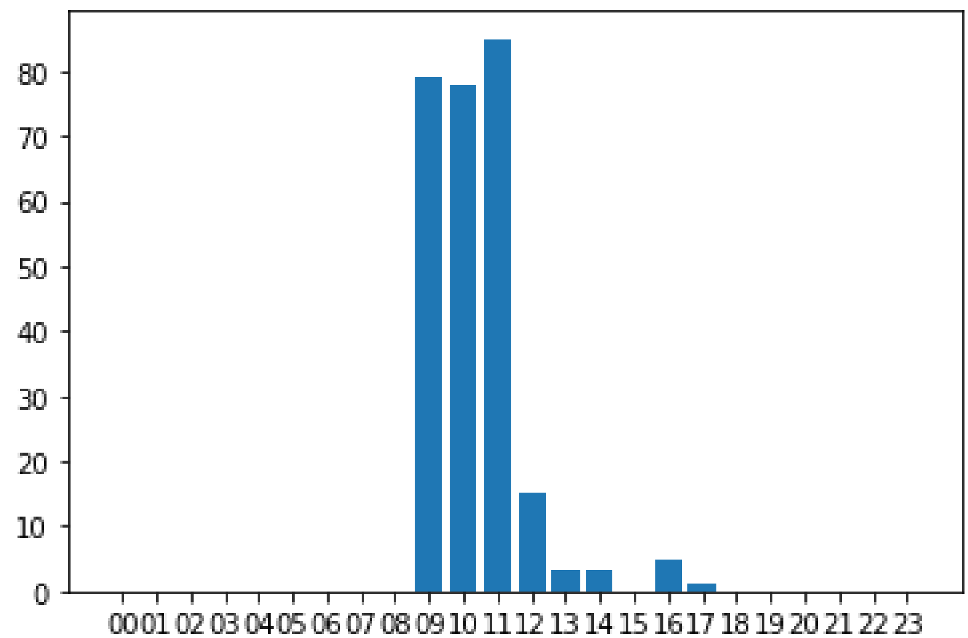


In similar way above below plots are plotted

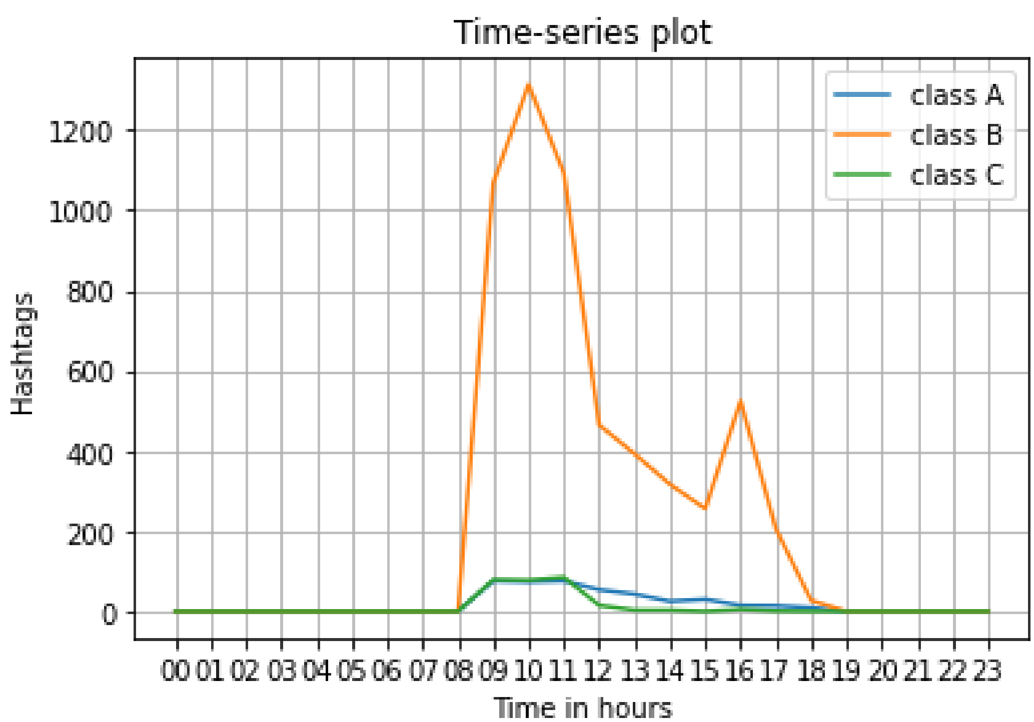
2. #MissYouYuvi :



3. #Both :



4. Combined Plot:



__Question 3:__

Do Topic Modelling: Used LDA

Parameters: max_df = 0.9 , min_df = 25, number of topics = 10

Few wordclouds is build for better analysis of topics.

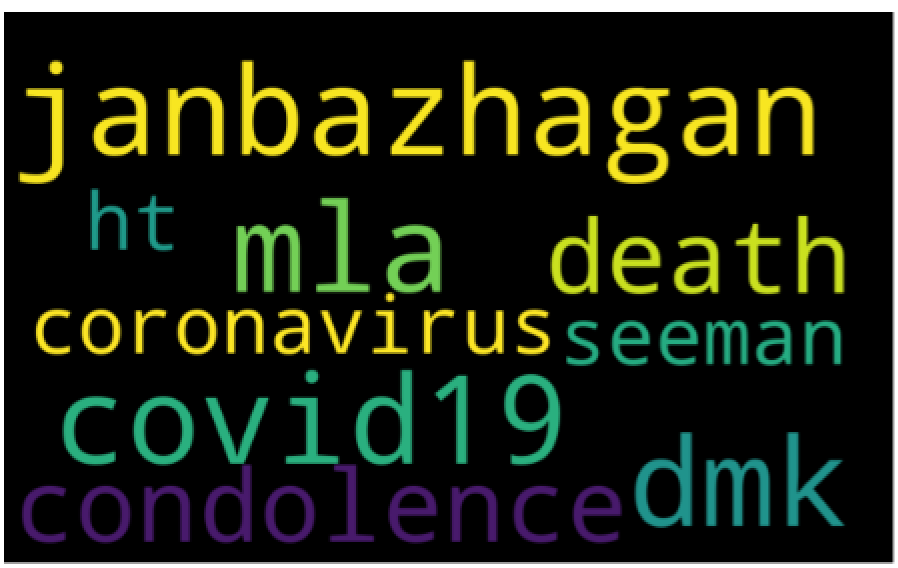
1. Class A → RestInPeace

   - Displaying topic words:

Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words
covid19	28.1	dmk	15.1	condolence	10.9	janbazhagan	0.1	janbazhagan	30.3	mla	20.1	seeman	139.1	janbazhagan
janbazhagan	8.0	death	10.1	janbazhagan	0.1	...	0.1	dmk	13.1	dmk	18.1	dmk	136.1	...
dmk	0.1	janbazhagan	6.1	dmk	0.1	dmk	0.1	...	0.1	janbazhagan	12.9	janbazhagan	69.4	dmk
...	0.1	mla	5.1	seeman	0.1	mla	0.1	condolence	0.1	...	0.1	condolence	69.3	mla
mla	0.1	condolence	0.1	covid19	0.1	condolence	0.1	seeman	0.1	coronavirus	0.1	...	69.1	condolence
condolence	0.1	covid19	0.1	...	0.1	death	0.1	mla	0.1	covid19	0.1	ht	68.1	death
death	0.1	...	0.1	mla	0.1	coronavirus	0.1	death	0.1	condolence	0.1	death	68.1	coronavirus
coronavirus	0.1	coronavirus	0.1	death	0.1	seeman	0.1	coronavirus	0.1	death	0.1	mla	68.1	seeman
seeman	0.1	seeman	0.1	coronavirus	0.1	ht	0.1	ht	0.1	seeman	0.1	coronavirus	68.1	ht
ht	0.1	ht	0.1	ht	0.1	covid19	0.1	covid19	0.1	ht	0.1	covid19	0.1	covid19

   - Word Cloud for few topics

Topic 0:



Topic 5:

Thus from the topic words analysis of RestInPeace it can been that most of the words are related to the deaths like condolence , mla who died in Chennai due to corona virus, covid19, georgefloydd who died in USA.

2. Class B → MissYouYuvi
    - Displaying topic words

Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Top wei
missing	148.1		248.7	yuvi	460.7	world	270.1	champion	151.1	tweet	367.1	year	139.3	cricket	456.9	six	
rt	135.1	redheart	225.3	miss	307.0	legend	202.9	nt	97.1	speed	166.1	love	135.7	india	153.5	dt	
heartsuit	123.1	master	129.9	u	269.5	cup	175.2	keep	70.2	retweet	154.1	one	134.2	...	130.0	trending	
badly	79.1	yuvraj	124.7	love	191.1	never	162.9	miss	65.5	completed	149.0	retirement	130.9	indian	129.4	give	
	76.8	singh	99.2	always	92.2	heart	158.0	today	63.1	hour	141.1	u	110.7	day	125.1	...	
12ykp12	56.1	yuvrajsingh	93.5	amp	92.2	retire	137.1	please	57.1	minute	135.9	field	102.9	one	106.4	life	
dhandsfoldedhands	54.1	retirement	82.4	paji	87.1	six	130.6	like	56.4	tag	118.1	life	96.6	best	94.2	never	
biharkacmgayabhai	53.1	world	81.8	...	85.0	always	107.5		50.8	cult	100.2	thalaivaaaaa	75.1	team	90.1	game	
love	53.1	great	71.4	thank	75.8	man	95.1	...	47.8	yuvians	86.1	missiooov	71.1	year	87.5	everything	
thalaivaaaa	47.1	year	66.7	sir	75.1	remain	87.1	made	42.8	common	78.1	happy	64.0	follow	67.1	top	

- Word cloud of topic words

Topic 1

Similarly in this case also all tweets are related to yuvraj singh retirement.

3. Class C → Both
    - Displaying Topic Words

	Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights	Topic 7 words	Topic 7 weights	Topic 8 words	Topic 8 weights	Topi wor
0	together	39.1	seeing	16.8	people	26.1	scared	10.1	time	5.1	amp	51.1	trending	51.1	together	6.1	trending	0.1	mome
1	trending	26.1	trending	0.1	scared	7.1	amp	0.1	indian	2.1	trending	46.1	time	15.1	seeing	6.1	time	0.1	scare
2	seeing	8.4	people	0.1	trending	0.1	time	0.1	trending	0.1	time	41.1	moment	0.1	people	6.1	amp	0.1	fu
3	amp	0.1	together	0.1	seeing	0.1	trending	0.1	amp	0.1	every	40.1	people	0.1	trending	0.1	scared	0.1	trendir
4	moment	0.1	time	0.1	time	0.1	together	0.1	scared	0.1	indian	40.1	seeing	0.1	time	0.1	together	0.1	an
5	people	0.1	amp	0.1	amp	0.1	seeing	0.1	together	0.1	rn	39.1	scared	0.1	amp	0.1	seeing	0.1	togeth
6	scared	0.1	scared	0.1	together	0.1	moment	0.1	seeing	0.1	scared	0.1	amp	0.1	scared	0.1	moment	0.1	tin
7	time	0.1	moment	0.1	moment	0.1	indian	0.1	moment	0.1	together	0.1	together	0.1	moment	0.1	indian	0.1	seeir
8	indian	0.1	indian	0.1	indian	0.1	people	0.1	people	0.1	moment	0.1	indian	0.1	indian	0.1	people	0.1	indi
9	every	0.1	every	0.1	every	0.1	every	0.1	every	0.1	people	0.1	every	0.1	every	0.1	every	0.1	peop

From the topic words, to make tweet trending these words are used as in all the topics similar words appear.

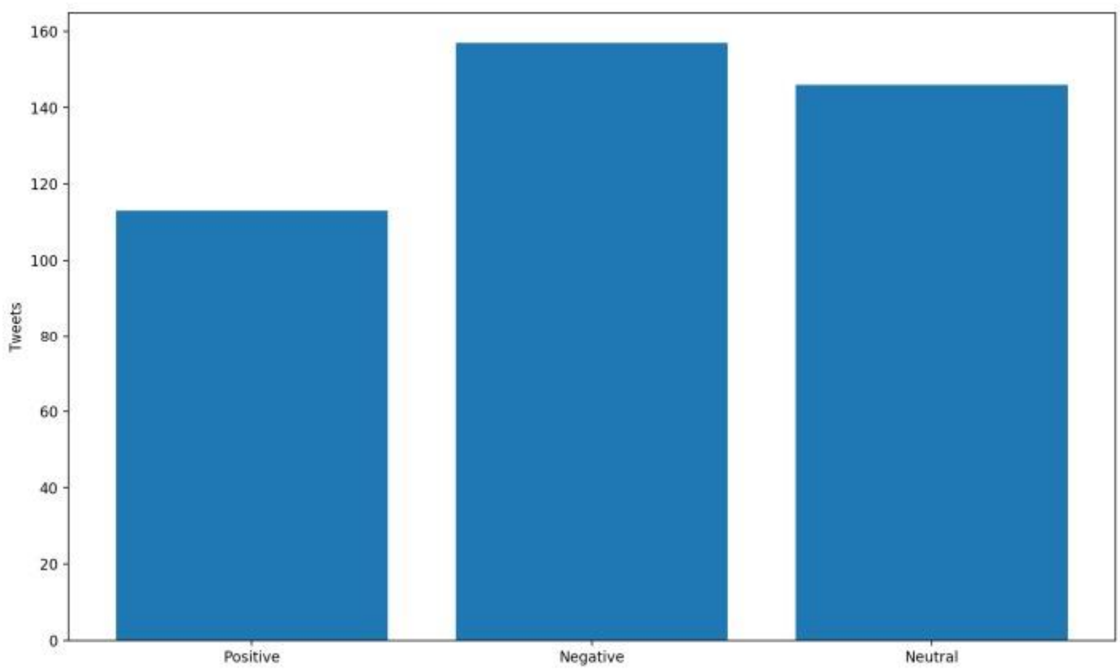
Thus for this a word cloud on all the words was generated to know which hashtag was used most among both.

Chatter Plot: Vader Sentiment Analyzer is used as it works best for social media sites.

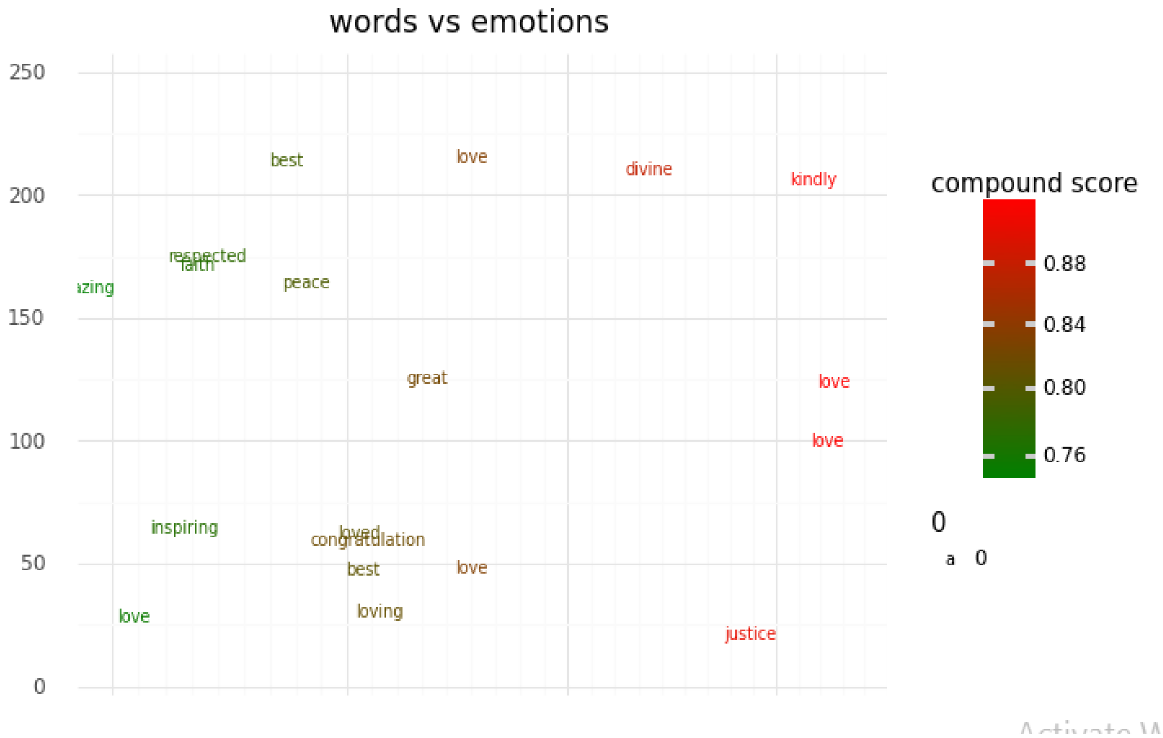
Tweets are classified into 3 categories: Positive, negative and neutral.

1. Class A

    - Showing number of tweets vs sentiments related to tweets:



- Chatter Plot :
  First Applied sentiment analysis on Tweets and then divided tweet into tokens and again applied sentiment analysis to found out which was emphasis the most to make it negative, positive and neutral and thus through those words we can get information related to tweet also.

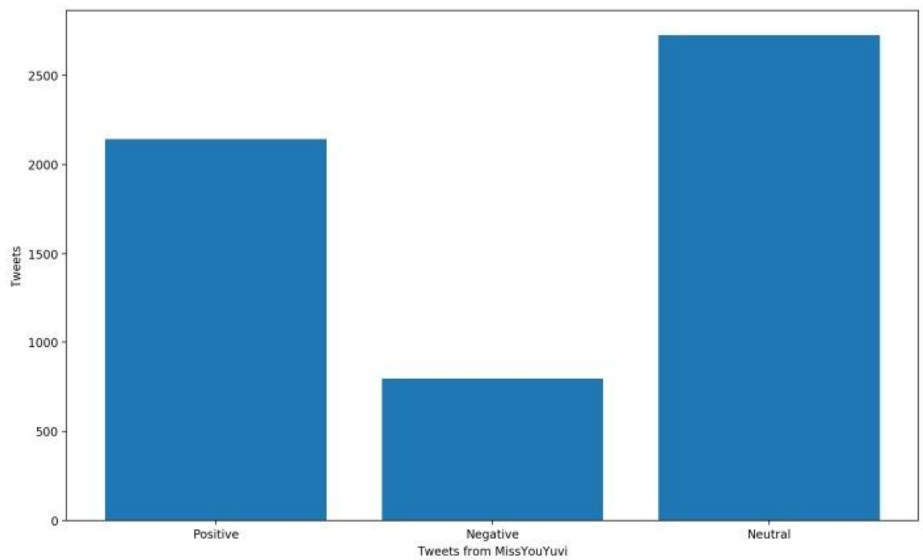


Colour intensiy is changing according to the class, here compound scores signifies the sentiments for the corresponding tweet.

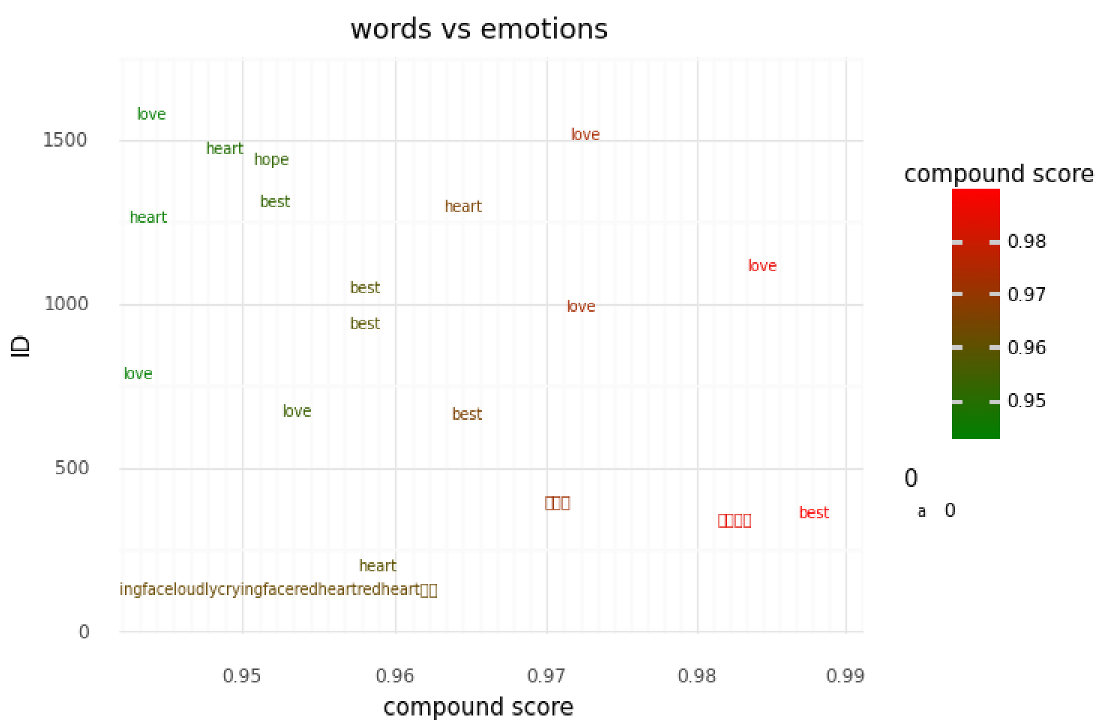
Similar approach was used in other classes also.

2. Class B

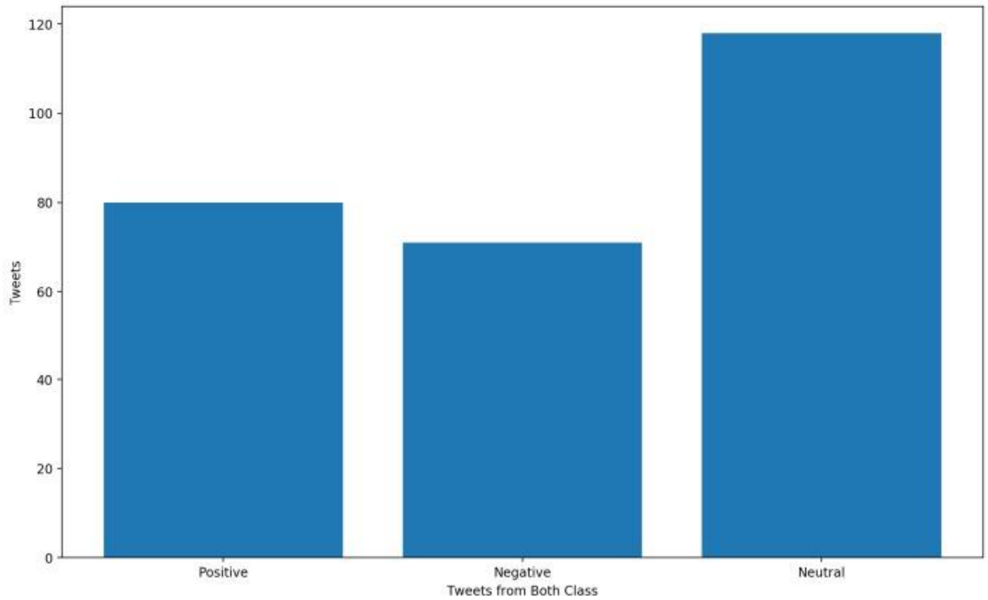
- Showing number of tweets vs sentiments related to tweets



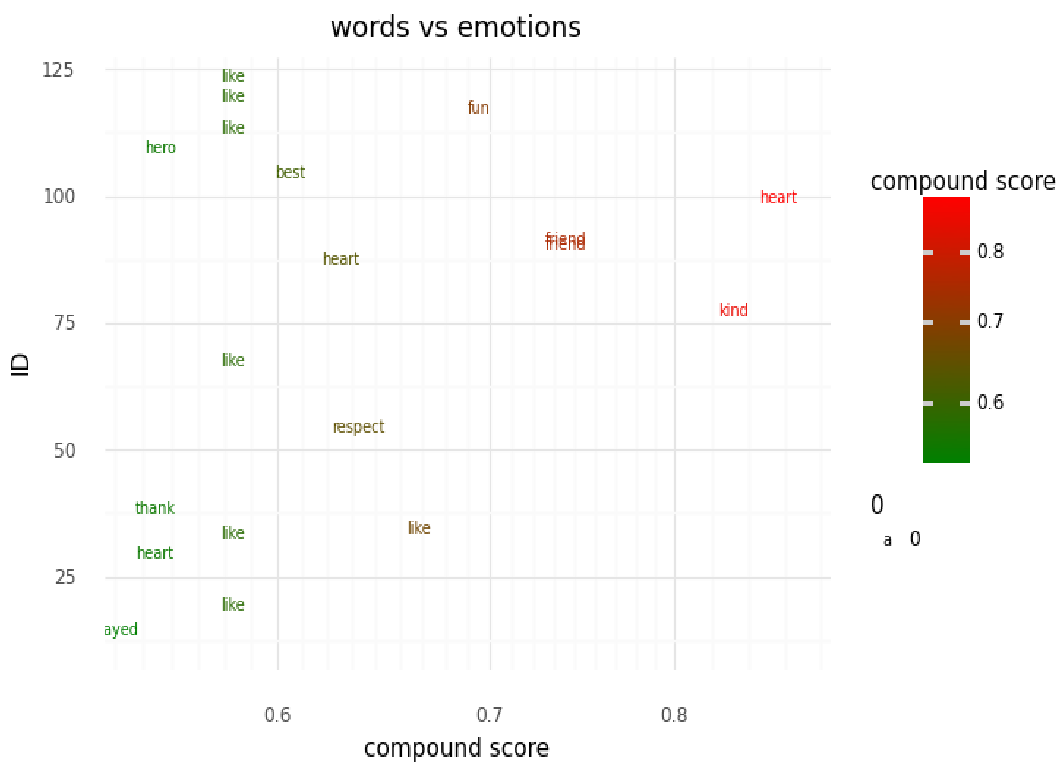
- Chatter Plot :



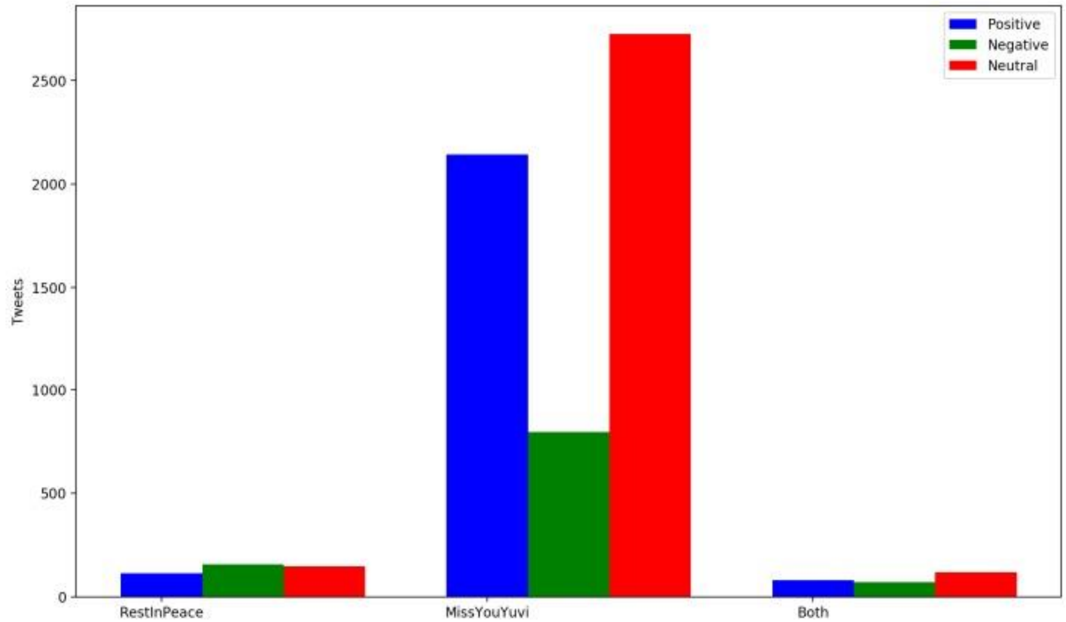
3. Class C
    - Showing number of tweets vs sentiments related to tweets:



- Chatter Plot :



- Combined plot of all classes:



Question 4:

1. Missyouyuvi are significant as compared to RestInPeace.
2. Combined tweets shows similar sentimental words which can be seen in chatter plot.
3. While applying topic modelling on class C i.e. (MissYouYuvi and RestInPeace), it was found that the topics were unrelated to the hashtags but in other two classes topics were related to hashtags.
4. After performing sentiment analysis, it was found that missyouyuvi and combined tweets follows the same distribution since in both count of neutral sentiment is higher than the other two's (positive and negative) while in case of rest in peace negative sentiment score was high.
5. Even though Missyouyuvi tweets were at large, Restinpeace were used in intersection tweets because it was a trending topic so that Missyouyuvi can reach to the masses. This is just an inference.
6. No verified users were seen to be using combined tweets.

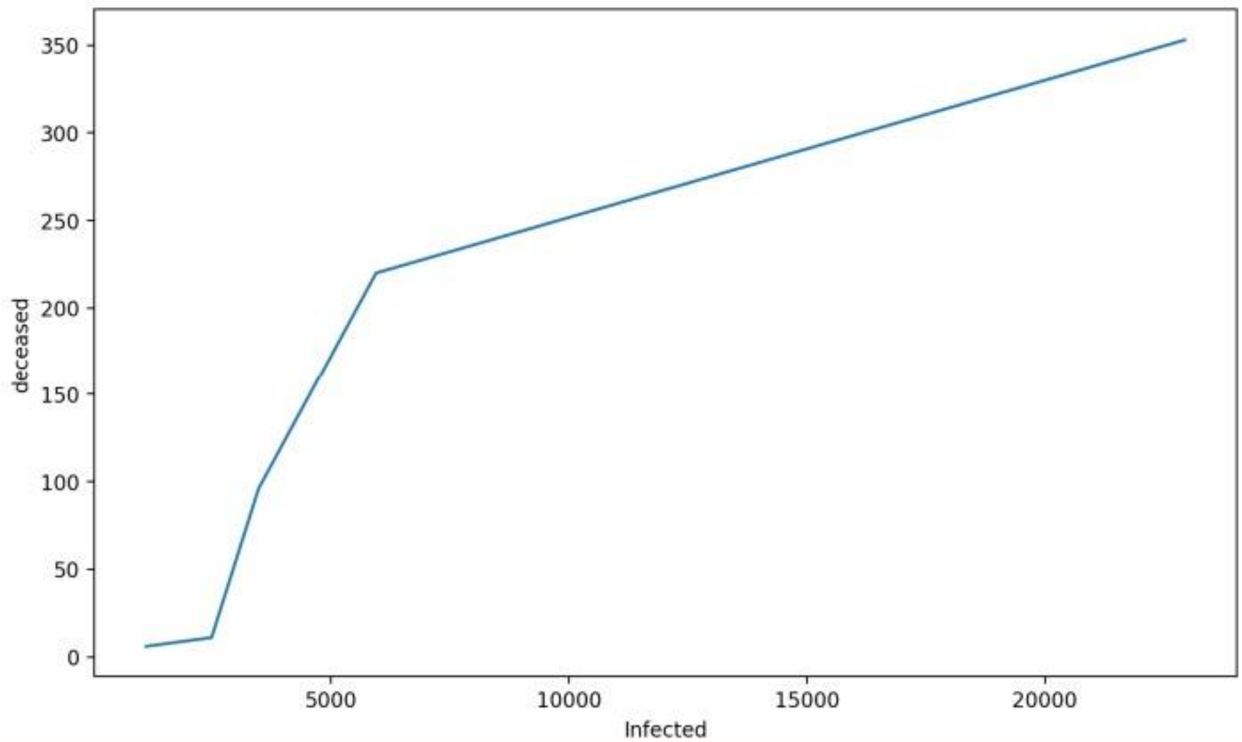
Question 5:

For finding correlation Karl Pearson Correlation is used.

For finding location through NER spacy is used.

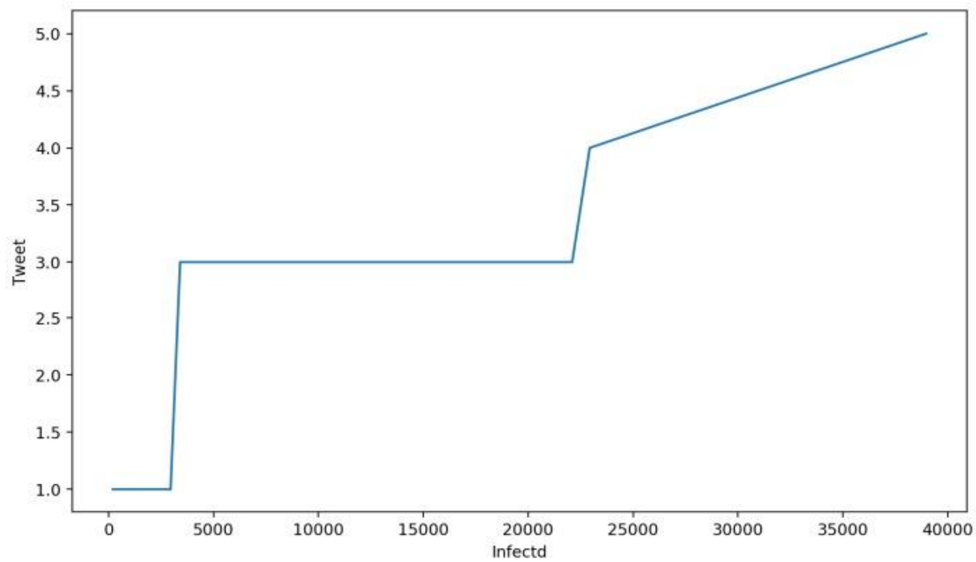
Found Correlation on the basis of states for RestInPeace hashtag only.

1. Actual Scenario plot of 10th june
   - number of cases in a particular state vs number of death in that state



Correlation: 0.9172702816273612 (There exists a positive correlation)

2. On the basis of tweets
   - Number of tweets vs number of infected in a particular state
     As the number of cases in an area increases, the number of tweets also increases. In case of Bihar numbers of tweets are less than the number of cases and deaths.



Here also positive correlation existed. Correlation: 0.42652163981000174

- Number of Tweets vs Number of Deaths.

  Here in this case actual value of death is way more than number of tweets. This can be easily represented through graph. In case of Delhi and Mumbai actual deaths are more and In case of Himachal Pradesh number of tweet and actual death is same on that day.