

## DATA COLLECTION:

The twitter dataset of corona virus is collected from kaggle. The data contains the tweets from various phases of lockdown.

Phase 1 : 25 March – 14 April

Phase 2 : 15 April – 3 May

Phase 3 : 4–17 May

Phase 4 : 18–31 May

Phase 5 : 1–30 June

Out[46]:

	status_id	user_id	created_at	screen_name	text	source	reply_to_status_id	reply_to_user_id	reply_to_s
0	1250574628726214658	817072420947247104	2020-04-16T00:00:00Z	Tu_IMSS_Coah	Para complementar la higiene de tus manos, el ...	TweetDeck	NaN	NaN	NaN
1	1250574627127975938	44728980	2020-04-16T00:00:00Z	ANCALETS	PWDs from Tahanang Walang Hagdan get aid amid ...	TweetDeck	NaN	NaN	NaN
2	1250574628671467521	1235017593649418241	2020-04-16T00:00:00Z	julycio	#France ได้รับ ความทุกข์ทรมานจาก #ชุดการแยกตัว...	Twitter for Advertisers	NaN	NaN	NaN
3	1250574628575010816	92174767	2020-04-16T00:00:00Z	DZMMTeleRadyo	Target ng Department of Health na magkaroon ng...	TweetDeck	NaN	NaN	NaN
4	1250574628201717760	1042498579909107712	2020-04-16T00:00:00Z	ContraReplicaMX	La @SSalud_mx lanzó una nueva	TweetDeck	NaN	NaN	NaN

## DATA CLEANING:

All the cleaning and pre processing is done.

TextBlob: TextBlob, one of the popular Python libraries for processing textual data, stands on the NLTK. TextBlob has some advanced features like –Sentiment

Extraction, Spelling Correction. Other tasks involves : Removing

Twitter Handles, Removing punctuations, numb

ers, and special character, removing short word, tokenization, stemming

Out[48]:

	created_at	text	favourites_count	retweet_count	country_code	lang
0	2020-04-16T00:00:00Z	Para complementar la higiene de tus manos, el ...	1828	2	NaN	es
1	2020-04-16T00:00:00Z	PWDs from Tahanang Walang Hagdan get aid amid ...	5449	1	NaN	tl
2	2020-04-16T00:00:00Z	#France ได้รับความช่วยเหลือจาก # มาตรการกักตัว...	27	0	NaN	th
3	2020-04-16T00:00:00Z	Target ng Department of Health na magkaroon ng...	578	3	NaN	tl
4	2020-04-16T00:00:00Z	La @SSalud_mx lanzó una nueva convocatoria esp...	5743	0	NaN	es

In [49]: `# filtering data with 'country_code = IN' and 'Language = en'`  
`tweet = tweet[(tweet.country_code == "IN") & (tweet.lang == "en")].reset_index(drop = True)`  
`tweet.drop(['country_code', 'lang'], axis=1, inplace=True)`  
`tweet.head()`

Out[49]:

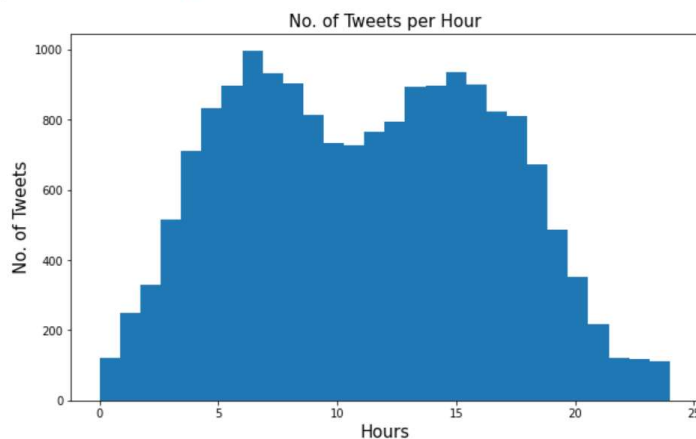
	created_at	text	favourites_count	retweet_count
0	2020-04-16T00:22:06Z	Interesting Read. A Four point strategy beyond...	841	0
1	2020-04-16T00:31:50Z	@Imamofpeace China sneezed - world freezed.\n...	664	0
2	2020-04-16T00:33:14Z	The poor in India are starving, the public is ...	5134	1
3	2020-04-16T00:40:40Z	We are not stuck at home we are safe at home.....	230	0
4	2020-04-16T00:50:40Z	I wish to contribute a little amount to our lo...	5	0

## Data Visualisation To Get Insights About Tweets.

The visualisations are done which involve:

- Number of tweets per hour

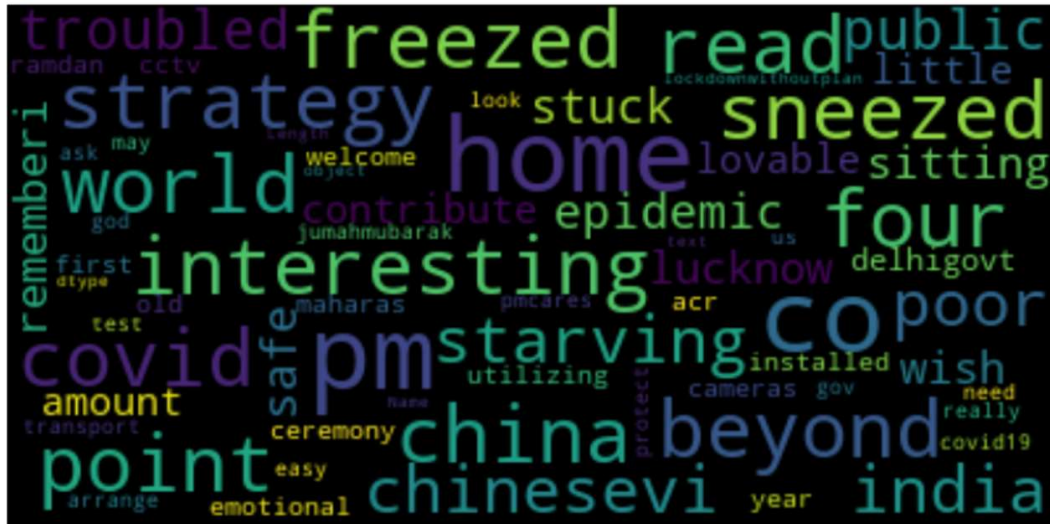
In [80]: `plt.figure(1, figsize=(10,6))`  
`plt.hist(tweet["created_at"],bins = 28);`  
`plt.xlabel('Hours',size = 15)`  
`plt.ylabel('No. of Tweets',size = 15)`  
`plt.title('No. of Tweets per Hour',size = 15)`  
`plt.savefig('Histo.png')`



- Most commonly used words in tweet using wordcloud:

```
plt.axis('off')
plt.title(title, size = 25)
plt.imshow(wordcloud, interpolation='bilinear')
plt.show()

show_wordcloud(tweet['text'])
cloud.to_file('WordCloud1.png')
```



## Analyzing:

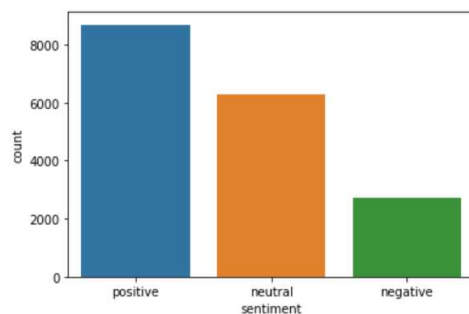
Following visuals were drawn after the analysis

- Countplot for total tweet count

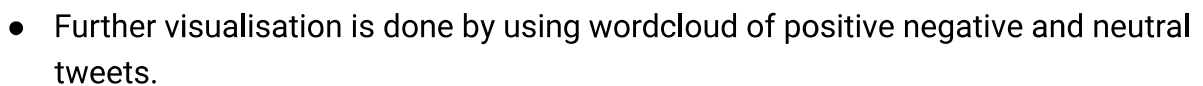
3	0.666667	stuck home safe home sitting lucknow rememberi...	230	0	positive	0.675
4	0.833333	wish contribute little amount lovable pm s pm ...	5	0	positive	0.304167

```
In [75]: print(tweet.sentiment.value_counts())
sns.countplot(x='sentiment', data = tweet);
plt.savefig("Count_for_Sentiments.png")
```

```
positive    8689
neutral     6265
negative    2715
Name: sentiment, dtype: int64
```



```
In [62]: #Sentiment Distribution
plt.figure(figsize=(10,6))
sns.distplot(tweet['polarity'], bins=30)
plt.title('Sentiment Distribution',size = 15)
plt.xlabel('Polarity',size = 15)
plt.ylabel('Frequency',size = 15)
plt.show();
```

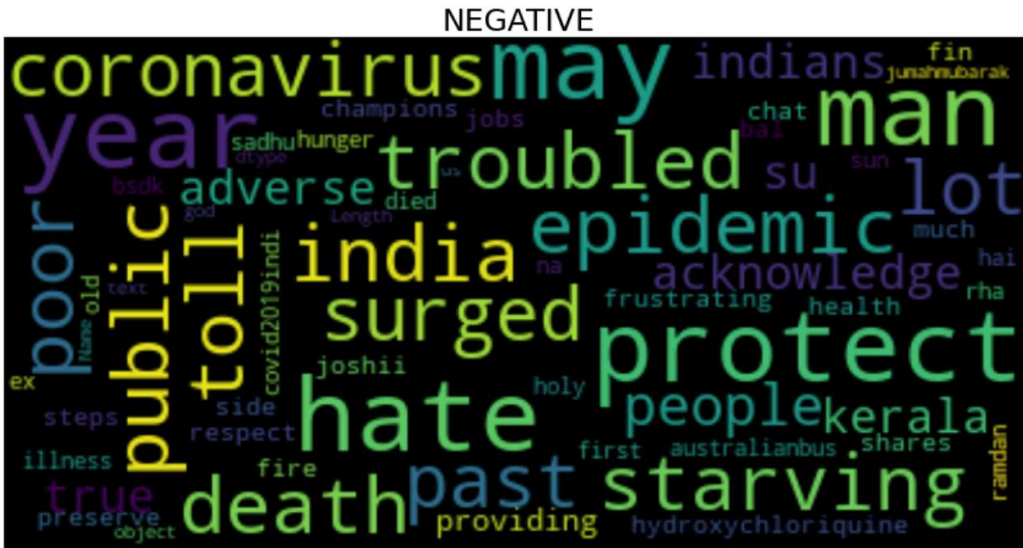


```
In [76]: #Using Word Clouds to see the higher frequency words from each sentiment
pos = tweet['text'][tweet['sentiment'] == 'positive']
show_wordcloud(pos, 'POSITIVE')
plt.savefig('WordCloud_for_positive_sentiments.png')
```

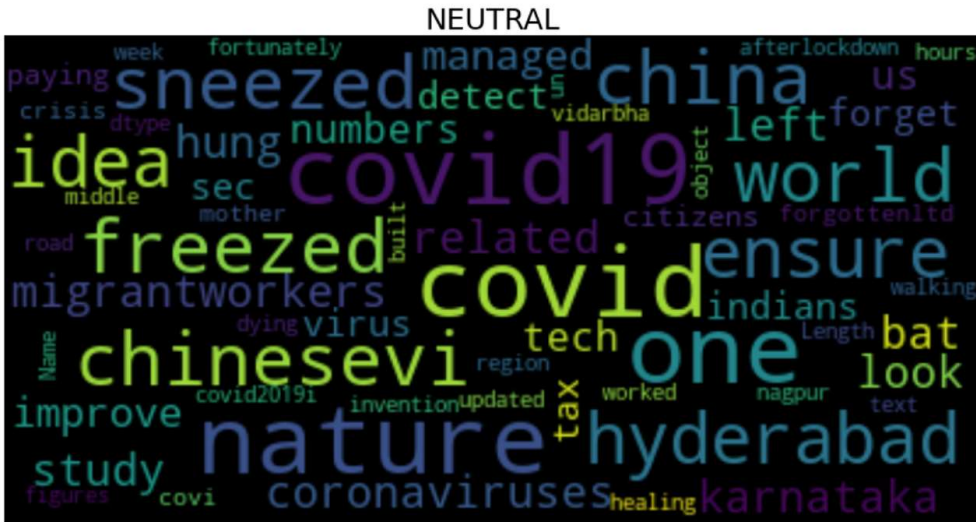


<Figure size 432x288 with 0 Axes>

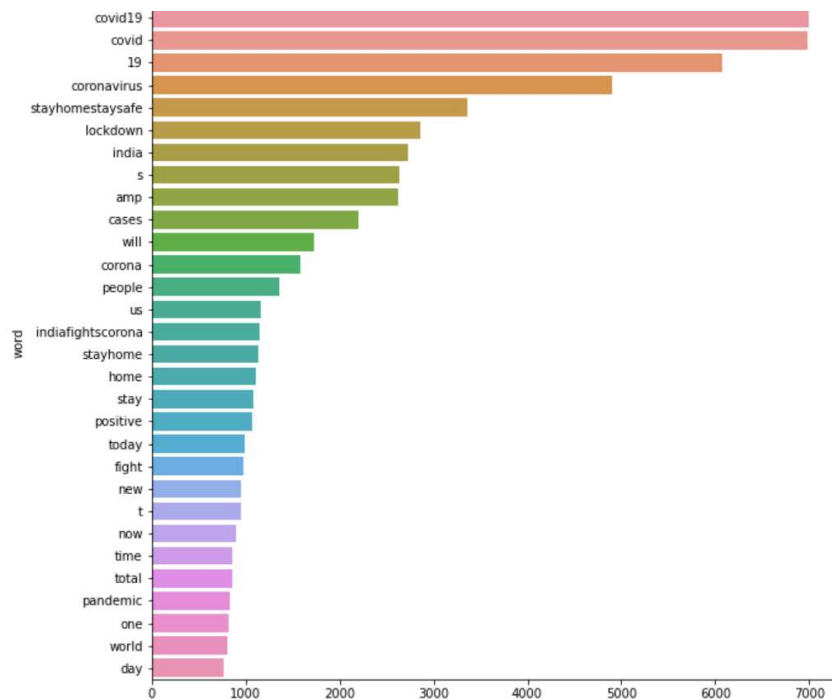
```
[77]: neg = tweet['text'][tweet['sentiment'] == 'negative']
      show_wordcloud(neg , 'NEGATIVE')
      plt.savefig('WordCloud_for_Negative_sentiments.png')
```



```
n [78]: neutral = tweet['text'][tweet['sentiment'] == 'neutral']
        show_wordcloud(neutral , 'NEUTRAL')
        plt.savefig('WordCloud_for_Neutral_sentiments.png')
```



- The following diagram has the frequency of most used words through which we can conclude



**Conclusion is that the tweets are mostly positive and neutral regarding lockdown extension.**