

Trending Topics on Twitter

Importing Modules

```
In [1]: import json
import tweepy
import pandas as pd
import plotly.express as px

from keys import *
from tqdm import tqdm
from datetime import datetime
from collections import Counter
```

Authentication

```
In [2]: auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)
api = tweepy.API(auth, wait_on_rate_limit=True, wait_on_rate_limit_notify=True)
woeid_Delhi = 20070458
setup = False
```

Data Collection

Most Prominent Hashtag

```
In [3]: if setup:

    result = api.trends_place(id = woeid_Delhi)
    trends = pd.DataFrame(result[0]['trends'])[['name', 'tweet_volume']]
    trends.dropna(inplace=True)
    trends.reset_index(inplace=True, drop=True)
    trends = trends[trends.name.str.contains('#')]
    trends['tweet_volume'] = trends['tweet_volume'].astype(int)
    trends.sort_values(by=['tweet_volume'], ascending=False, ignore_index=True, inplace=True)
    top_hashtag = trends['name'][0]
    print(trends)
```

Tweets Around Most Prominent Hashtag

```
In [4]: if setup:

    Files = {}
    Files['Topic'] = top_hashtag
    Files['Top_10'] = []
    Query = top_hashtag
    Tweets = []
    cursor = tweepy.Cursor(api.search, q = Query, tweet_mode = "extended", lang = 'en').items(2000)

    for i in tqdm(cursor):
        Tweets.append(i._json)

    json.dump(Tweets, open('./Dump/'+top_hashtag+'_tweets.json', 'w'))
```

Top 10 Hashtags In Tweets Around Most Prominent Hashtag

```
In [5]: if setup:

    Tweets = json.load(open('./Dump/'+top_hashtag+'_tweets.json', 'r'))
    Hashtags = []

    for t in Tweets:
        if 'hashtags' in t['entities']:
            for x in t['entities']['hashtags']:
                H = "#"+x['text']
                H1=top_hashtag
                Hashtags.append(H)

    Top_10_Hashtags = Counter(Hashtags).most_common(10)

    print(Top_10_Hashtags)
```

Tweets Around Top 10 Hashtags

```
In [6]: if setup:

    for x in Top_10_Hashtags:
        Query = x[0]
        Tweets = []
        cursor = tweepy.Cursor(api.search, q = Query, tweet_mode = "extended", lang = 'en').items(2000)

        for i in tqdm(cursor):
            Tweets.append(i._json)

        json.dump(Tweets, open('./Dump/'+Query+'_tweets.json', 'w'))
        Files['Top_10'].append(Query)
    json.dump(Files, open('./Dump/hashtags.json', 'w'))
```

Merging Data

```
In [7]: if setup:

    Files = json.load(open('./Dump/hashtags.json', 'r'))
    Data = []
    Data += json.load(open('./Dump/'+Files['Topic']+'_tweets.json', 'r'))
    for x in Files['Top_10']:
        Data += json.load(open('./Dump/'+x+'_tweets.json', 'r'))
```

Removing Duplicates

```
In [8]: if setup:

    IDs = set()
    New_Data = []

    for x in Data:
        if x['id'] not in IDs:
            New_Data.append(x)
            IDs.add(x['id'])

    json.dump(New_Data, open('./Dump/Data.json', 'w'))
```

Analysis

Loading Data

```
In [9]: Data = json.load(open('./Dump/Data.json', 'r'))
```

Top 10 Hashtags Based on Number of Occurrences

```
In [10]: Hashtags = []

for tweet in Data:
    H = set()
    if 'hashtags' in tweet['entities']:
        for x in tweet['entities']['hashtags']:
            H.add("#"+x['text'])
    for i in H:
        Hashtags.append(i)

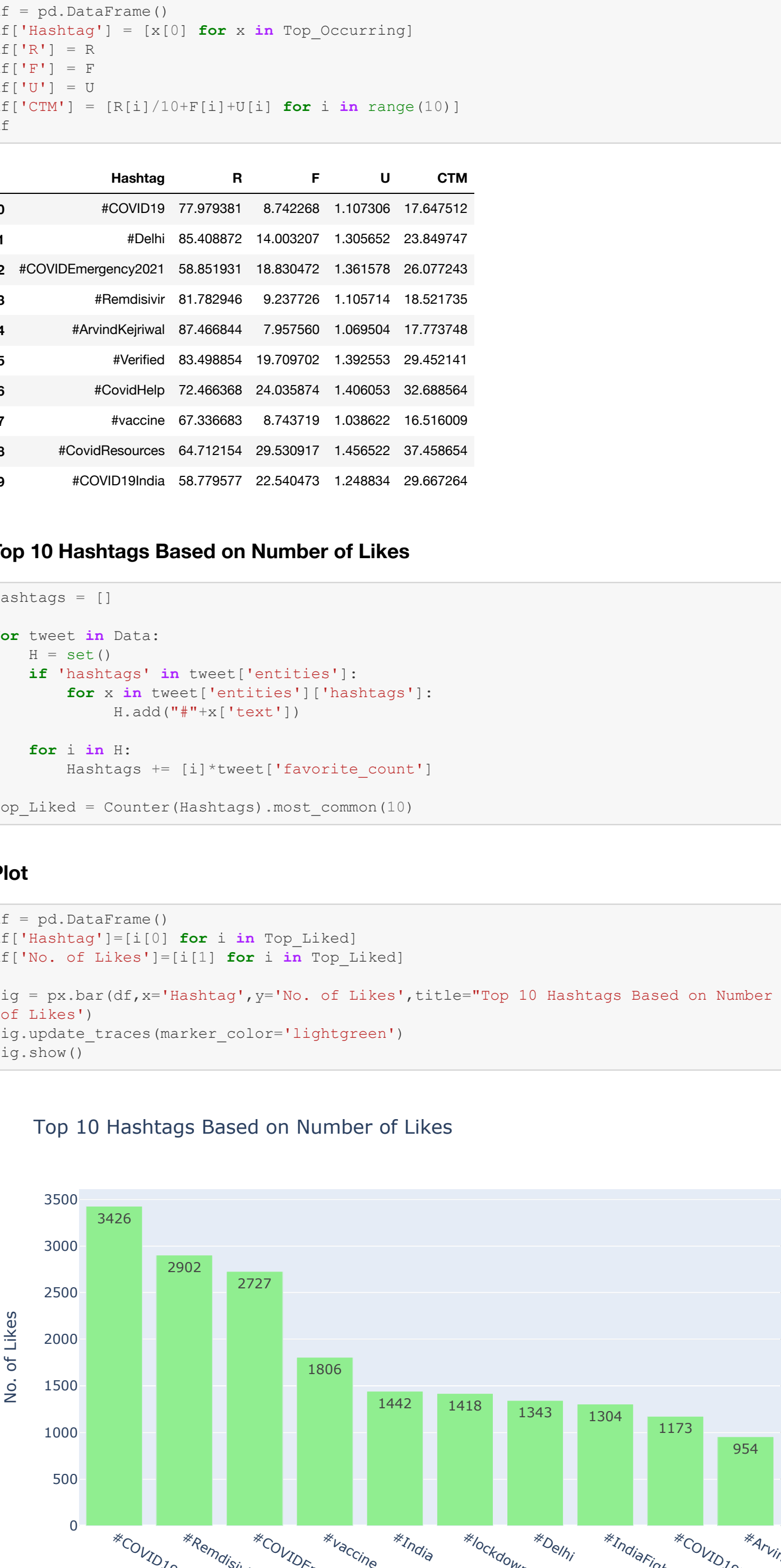
Top_Occurring = Counter(Hashtags).most_common(10)
```

Plot

```
In [11]: df = pd.DataFrame()
df['Hashtag']=[i[0] for i in Top_Occurring]
df['No. of Occurrences']=[i[1] for i in Top_Occurring]

fig = px.bar(df, x='Hashtag', y='No. of Occurrences', title="Top 10 Hashtags Based on Number of Occurrence", text='No. of Occurrences')
fig.update_traces(marker_color='skyblue')
fig.show()
```

Top 10 Hashtags Based on Number of Occurrences



CTM

```
In [12]: R = []
F = []
U = []

for x in Top_Occurring:

    H = x[0]
    r = 0
    t = 0
    Users = {}

    for tweet in Data:
        flag = False

        if 'hashtags' in tweet['entities']:
            for y in tweet['entities']['hashtags']:
                if y['text'] == H[1:]: flag = True

        if flag:
            UID = tweet['user']['id']
            if UID in Users:
                Users[UID]+=1
            else:
                Users[UID]=1

            r+=int(tweet['full_text'][:2].lower()=='rt')
            t+=1

    R.append(r*100/t)
    F.append(sum(sorted(list(Users.values()), reverse=True)[:50])*100/t)
    U.append(t/len(Users))
```

```
In [13]: df = pd.DataFrame()
df['Hashtag'] = [x[0] for x in Top_Occurring]
df['R'] = R
df['F'] = F
df['U'] = U
df['CTM'] = [R[i]/10+F[i]+U[i] for i in range(10)]
df
```

```
Out[13]:
```

	Hashtag	R	F	U	CTM
0	#COVID19	77.979381	8.742268	1.107306	17.647512
1	#Delhi	85.408872	14.003207	1.305652	23.849747
2	#COVIDEmergency2021	58.851931	18.830472	1.361578	26.077243
3	#Remdisivr	81.782946	9.237726	1.105714	18.521735
4	#ArvindKejriwal	87.466844	7.957560	1.069504	17.773748
5	#Verified	83.498854	19.709702	1.392553	29.452141
6	#CovidHelp	72.466368	24.035874	1.406053	32.688564
7	#vaccine	67.336683	8.743719	1.038622	16.516009
8	#CovidResources	64.712154	29.530917	1.456522	37.458654
9	#COVID19India	58.779577	22.540473	1.248834	29.667264

Top 10 Hashtags Based on Number of Likes

```
In [14]: Hashtags = []

for tweet in Data:
    H = set()
    if 'hashtags' in tweet['entities']:
        for x in tweet['entities']['hashtags']:
            H.add("#"+x['text'])

    for i in H:
        Hashtags += [i]*tweet['favorite_count']

Top_Liked = Counter(Hashtags).most_common(10)
```

Plot

```
In [15]: df = pd.DataFrame()
df['Hashtag']=[i[0] for i in Top_Liked]
df['No. of Likes']=[i[1] for i in Top_Liked]

fig = px.bar(df, x='Hashtag', y='No. of Likes', title="Top 10 Hashtags Based on Number of Likes", text='No. of Likes')
fig.update_traces(marker_color='lightgreen')
fig.show()
```

Top 10 Hashtags Based on Number of Likes



CTM

```
In [16]: R = []
F = []
U = []

for x in Top_Liked:

    H = x[0]
    r = 0
    t = 0
    Users = {}

    for tweet in Data:
        flag = False

        if 'hashtags' in tweet['entities']:
            for y in tweet['entities']['hashtags']:
                if y['text'] == H[1:]: flag = True

        if flag:
            UID = tweet['user']['id']
            if UID in Users:
                Users[UID]+=1
            else:
                Users[UID]=1

            r+=int(tweet['full_text'][:2].lower()=='rt')
            t+=1

    R.append(r*100/t)
    F.append(sum(sorted(list(Users.values()), reverse=True)[:50])*100/t)
    U.append(t/len(Users))
```

```
In [17]: df = pd.DataFrame()
df['Hashtag'] = [x[0] for x in Top_Liked]
df['R'] = R
df['F'] = F
df['U'] = U
df['CTM'] = [R[i]/10+F[i]+U[i] for i in range(10)]
df
```

```
Out[17]:
```

	Hashtag	R	F	U	CTM
0	#COVID19	77.979381	8.742268	1.107306	17.647512
1	#Remdisivr	81.782946	9.237726	1.105714	18.521735
2	#COVIDEmergency2021	58.851931	18.830472	1.361578	26.077243
3	#vaccine	67.336683	8.743719	1.038622	16.516009
4	#India	75.655431	25.468165	1.072289	34.105997
5	#lockdown	88.652482	37.588652	1.021739	47.475640
6	#Delhi	85.408872	14.003207	1.305652	23.849747
7	#IndiaFightsCorona	71.392723	15.307403	1.099310	23.545985
8	#COVID19India	58.779577	22.540473	1.248834	29.667264
9	#ArvindKejriwal	87.466844	7.957560	1.069504	17.773748

Top 10 Hashtags Based on Number of Retweets

```
In [18]: Hashtags = []

for tweet in Data:
    H = set()
    if 'hashtags' in tweet['entities']:
        for x in tweet['entities']['hashtags']:
            H.add("#"+x['text'])

    if tweet['full_text'][:2].lower()!='rt':
        for i in H:
            Hashtags += [i]*tweet['retweet_count']

Top_Retweeted = Counter(Hashtags).most_common(10)
```

Plot

```
In [19]: df = pd.DataFrame()
df['Hashtag']=[i[0] for i in Top_Retweeted]
df['No. of Retweets']=[i[1] for i in Top_Retweeted]

fig = px.bar(df, x='Hashtag', y='No. of Retweets', title="Top 10 Hashtags Based on Number of Retweets", text='No. of Retweets')
fig.update_traces(marker_color='pink')
fig.show()
```

Top 10 Hashtags Based on Number of Retweets

CTM

```
In [20]: R = []
F = []
U = []

for x in Top_Retweeted:

    H = x[0]
    r = 0
    t = 0
    Users = {}

    for tweet in Data:
        flag = False

        if 'hashtags' in tweet['entities']:
            for y in tweet['entities']['hashtags']:
                if y['text'] == H[1:]: flag = True

        if flag:
            UID = tweet['user']['id']
            if UID in Users:
                Users[UID]+=1
            else:
                Users[UID]=1

            r+=int(tweet['full_text'][:2].lower()=='rt')
            t+=1

    R.append(r*100/t)
    F.append(sum(sorted(list(Users.values()), reverse=True)[:50])*100/t)
    U.append(t/len(Users))
```

```
In [21]: df = pd.DataFrame()
df['Hashtag'] = [x[0] for x in Top_Retweeted]
df['R'] = R
df['F'] = F
df['U'] = U
df['CTM'] = [R[i]/10+F[i]+U[i] for i in range(10)]
df
```

```
Out[21]:
```

	Hashtag	R	F	U	CTM
0	#COVIDEmergency2021	58.851931	18.830472	1.361578	26.077243
1	#Remdisivr	81.782946	9.237726	1.105714	18.521735
2	#CovidHelp	72.466368	24.035874	1.406053	32.688564
3	#Delhi	85.408872	14.003207	1.305652	23.849747
4	#COVID19	77.979381	8.742268	1.107306	17.647512
5	#CovidResources	64.712154	29.530917	1.456522	37.458654
6	#Verified	83.498854	19.709702	1.392553	29.452141
7	#COVID19India	58.779577	22.540473	1.248834	29.667264
8	#vaccine	67.336683	8.743719	1.038622	16.516009
9	#IndiaFightsCorona	71.392723	15.307403	1.099310	23.545985

Tweet IDs

```
In [22]: with open('section_II.txt', 'w') as file:
    for tweet in Data:
        file.write(str(tweet['id'])+'\n')
```