

**MSBA 305 – Data processing framework**

**Final Project Report**

**Prepared for:**

**Dr. Elie Nasr**

**Prepared by:**

**Mazen Hajj Diab**

**Samer Haidar**

**Maryam Ahmed Ali**

**Ibrahim Al Bahri**

**December 14, 2020**

**Introduction:**

With the evolution of web technologies, a huge amount of data is being generated. Data expressing people’s opinions, preferences, and behaviors are now available in huge amounts on platforms such as Facebook, Twitter, Instagram, and others. One of social media applications that generate the highest amount of data is twitter with more than 500 million tweet per day. Unlike other social media applications, Twitter’s tweets are public and accessible to everyone. Additionally, Twitter’s API allows you to filter the tweets by specifying a date and topic for your search. Twitter data can be leveraged in many ways and by many people. Marketers for example could leverage twitter data to get a sense of whether people are responding positively or negatively to their marketing campaigns. Moreover, governments and decision makers could also use twitter data to keep track of peoples’ concerns, political views, and opinions regarding certain issues. As for our study, we are going to extract tweets that expresses people’s opinions about the COVID-19 vaccine which is expected to be released soon using the twitter API in the developer’s account. Then we are going to analyze these tweets to get a sense of whether people are confident and willing to take this vaccine or not. This can be done using machine learning algorithms such as NLP and sentimental analysis.

**Problem Statement:**

With over 69.1 million confirmed covid-19 cases which resulted nearly 1.57 million deaths worldwide, the covid-19 vaccine is finally expected to be distributed and used as preventive measure for the coronavirus. After about a year of public debate about the Covid-19 vaccine and the doubts regarding its safety, side effects and effectiveness especially that it is being released to the market a in record time, the question that most governments are facing is whether the public will be convinced to take the vaccine or not. Another concern for the governments is the measures that they should take to build up people’s confidence in the vaccine.

Now human challenging trials for the vaccine are starting to show good results and many countries started vaccination, but polls across the world showed that the public opinion still hesitant and it’s not a surprise that anti-vaccination campaigns impacted the public sentiment, which led to initiating awareness-raising campaigns that quickly reversed this trend.

A CNN poll conducted by SSRS found that in the US, only about half of the American population is willing to get the Covid-19 vaccine. Also, a new Ipsos survey conducted on behalf of the World Economic Forum found different opinions among EU member states. The percentage of people who were surveyed and said that they are willing get vaccinated whenever the vaccine becomes available was lowest in Russia (54%), Poland (56%), Hungary (56%), and France (59%). The highest percentages were in China (97%), Brazil (88%), Australia (88%), and India (87%).

**Tweets Analysis:**

So, in order to further investigate the public sentiment regarding the newly approved Covid-19 vaccine, we analyzed people tweets on Twitter to see how positive, negative, or neutral their opinions on the vaccine are and how they differ between locations, we also tried to measure popularity of the accounts that tweeted most regarding this topic by ranking them due to number of followers or re-tweets.

The following steps were done:

* We signed up for Twitter developer account so we can get access to Twitter API by using Tweepy library.
* Then we searched tweets by #Covid-19 vaccine
* We got more than 20000 tweets, saved them to csv file and then read them to data frame using Pandas library.
* We cleaned data to be prepared for analysis.
* Then, conducted different analysis to discover public sentiment regarding the vaccine worldwide and by city as well, we imported machine learning libraries for natural language processing like nltk and TextBlob libraries.
* We saved the data frame to MySQL database to save the tweets, and then we analyzed them using Python
* Then we were able to query the data we saved in MySQL to make assure of our work.

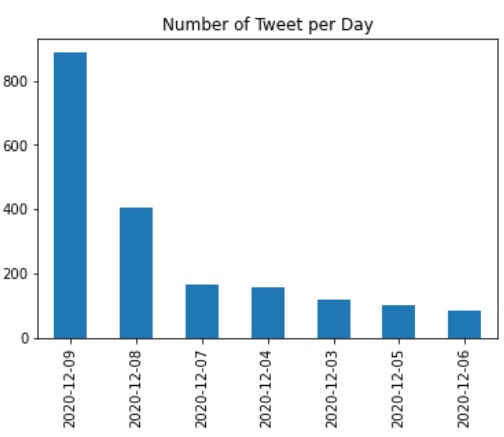
**Questions and Analysis:**

**General Questions:**

**Question 1: What are the total number of cleaned tweets do we have?**

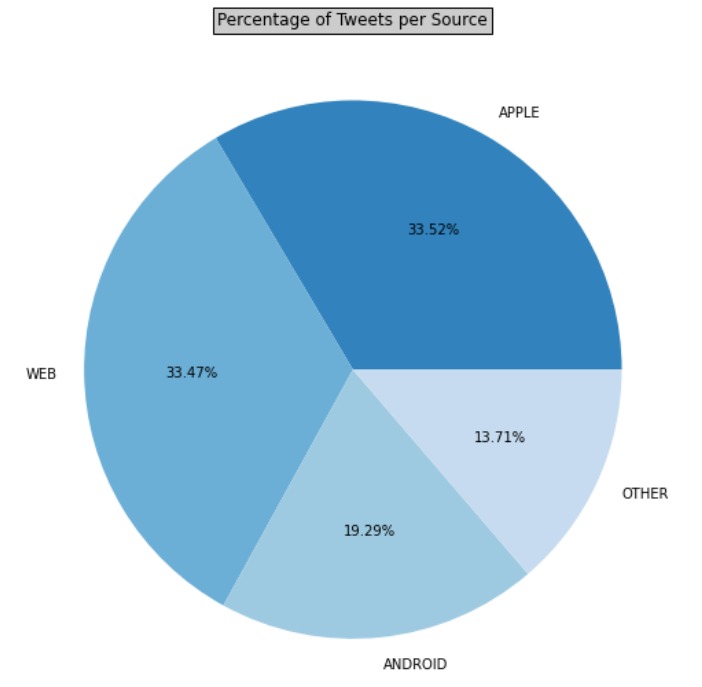
We have 1918 tweets.

**Question 2: What is the date the data is ranging from?**



As shown in the graph, the data ranges from December 3 2020 till December 9, 2020 with most tweets appearing on the days of December 9 and 8, and the least appearing on December 5 and December 6.

**Question 3: From where are the users tweeting?**

****

As shown in Fig. X,

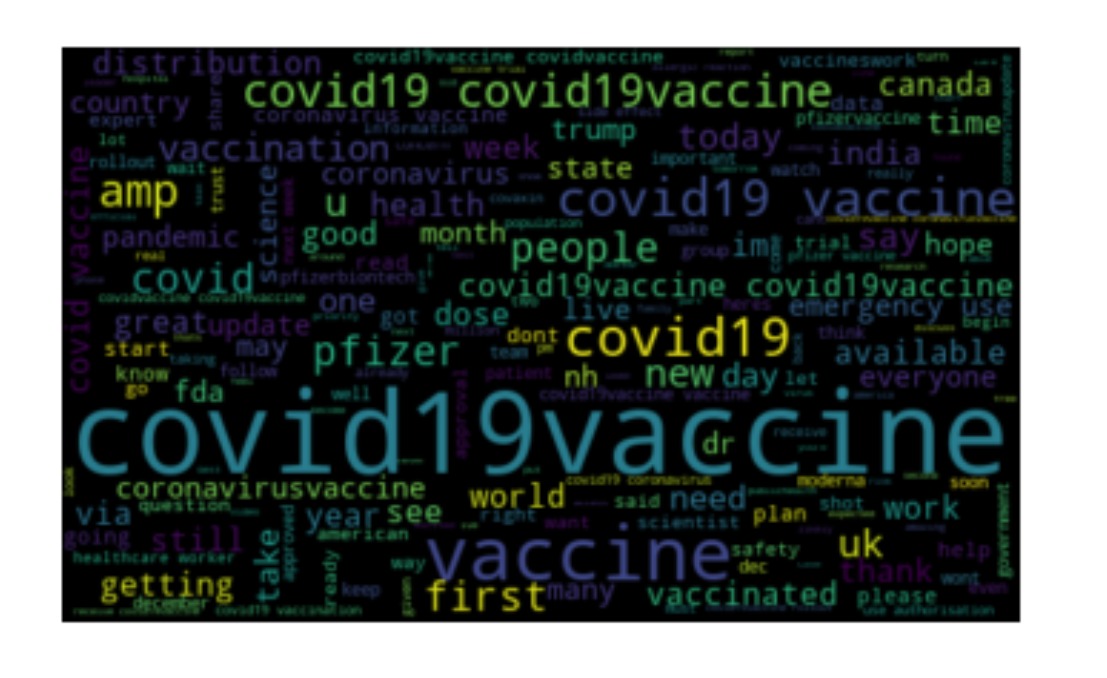
33.52% of the users tweeted from an Apple device

33.47% of the users tweeted from Web

19.29% of the users tweeted from an Android device

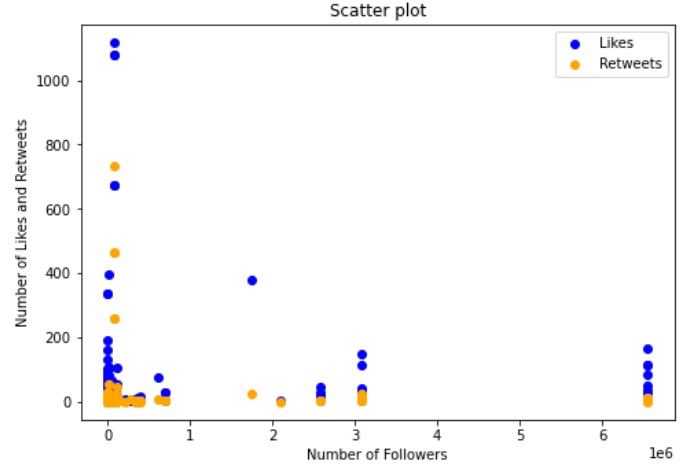
13.71% of the users tweeted from other others devices

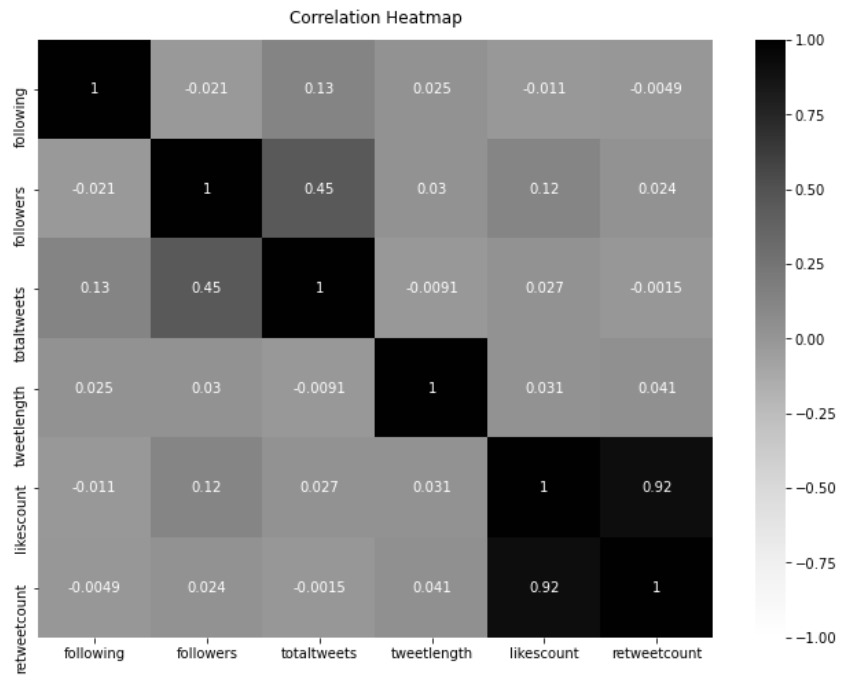
**Question 4: What are the most used words in these tweets?**



As show in Fig. X, the most common words in the following tweets are: covid19vaccine, covid19, pandemic, first, vaccinated, please, work, world, good etc..

**Question 5: What is the correlation between the number of likes and the number of followers of the users?**

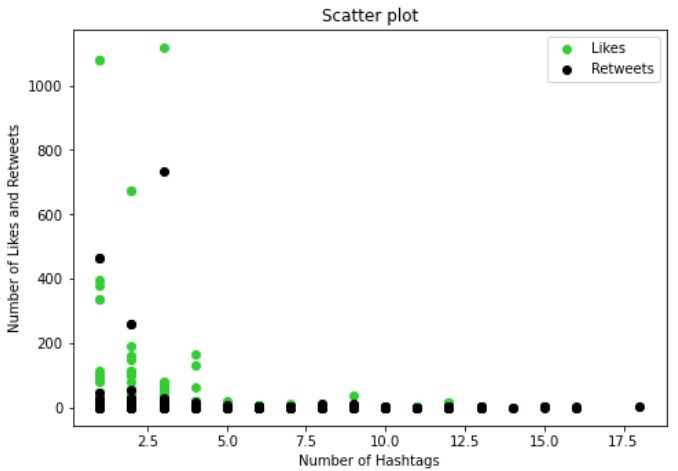




As we can see in Fig. X, there is no clear correlation between the number of followers and the number of likes or retweets.

As show in Fig. Y, the only significant correlation occurs between the number of likes and the number of retweets of a specific tweet.

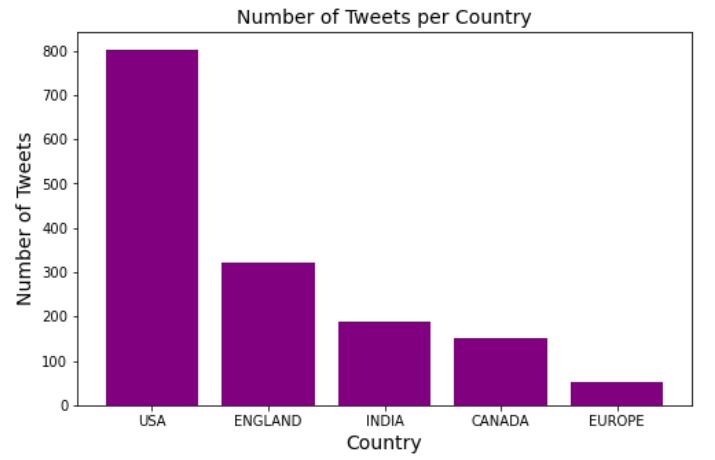
**Question 6: What is the correlation between the number of likes and the number of hashtags the users used?**



As we can see in Fig. X, there is no clear correlation between the number of hashtags used and the number of likes or retweets

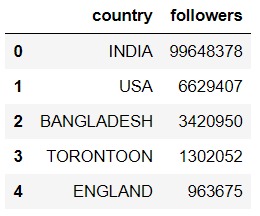
**Per Country Questions:**

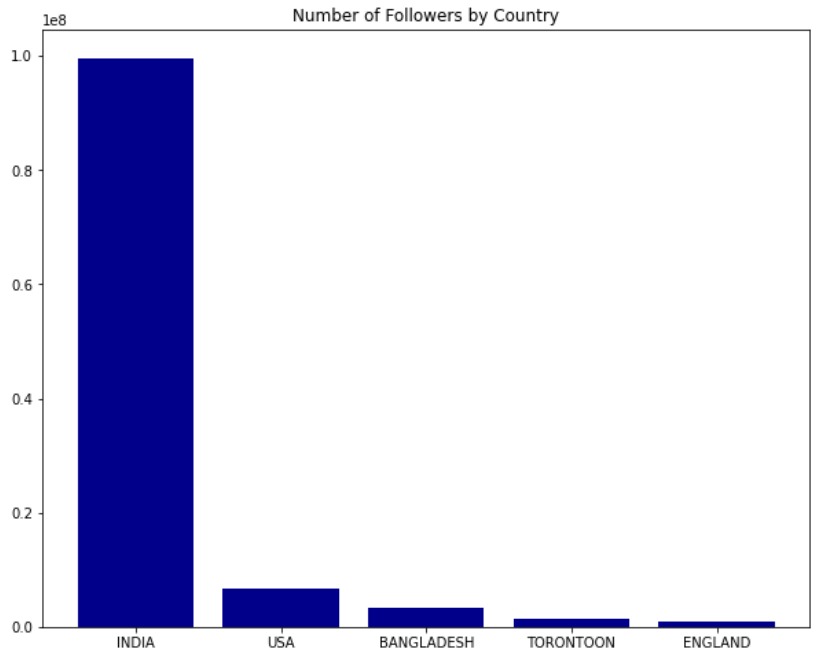
**Question 7: What are the top 5 countries that tweeted the most?**

****

As shown in Fig. 1, the top 5 countries that tweeted about the covid vaccine were USA, England, India, Canada and Europe. Note that we separated England from Europe since at the beginning of December, England started the vaccination campaign which consequently increased the number of tweets concerning this issue. Including it would make it biased.

**Question 8: What are the top 5 countries where its users hav the most followers?**

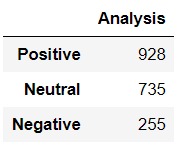


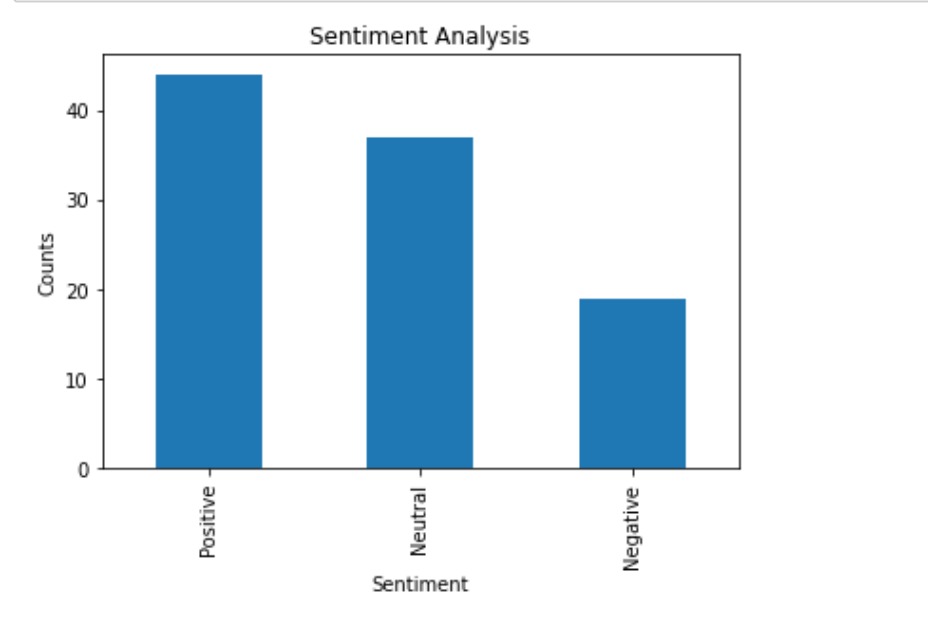


As shown in Fig. X and Fig, Y , countries with the most reach, as in, the user that tweeted in this country’s followers are India, USA, Bangladesh Toronto (Canada) and England.

**Per Analysis Category Questions:**

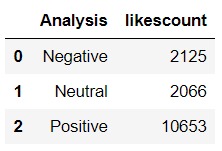
**Question 9: What is the number of tweets for each Analysis category (whether the tweet is Negative, Neutral or “Positive”)?**

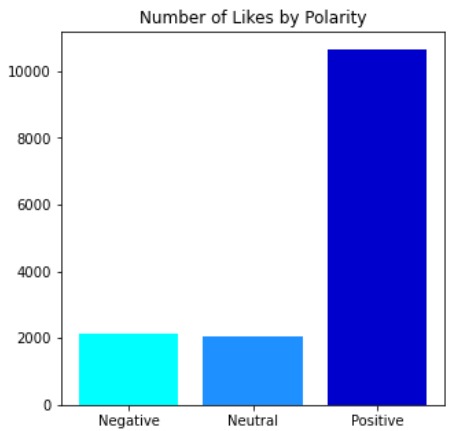
****

****

As shown in Fig. X, Fig. Y, most of the tweets are positive tweets (928) and the least are negative tweets (255)

**Question 10: What is the sum of likes tweets about each Analysis (whether the tweet is Negative, Neutral or “Positive”)?**

****

****

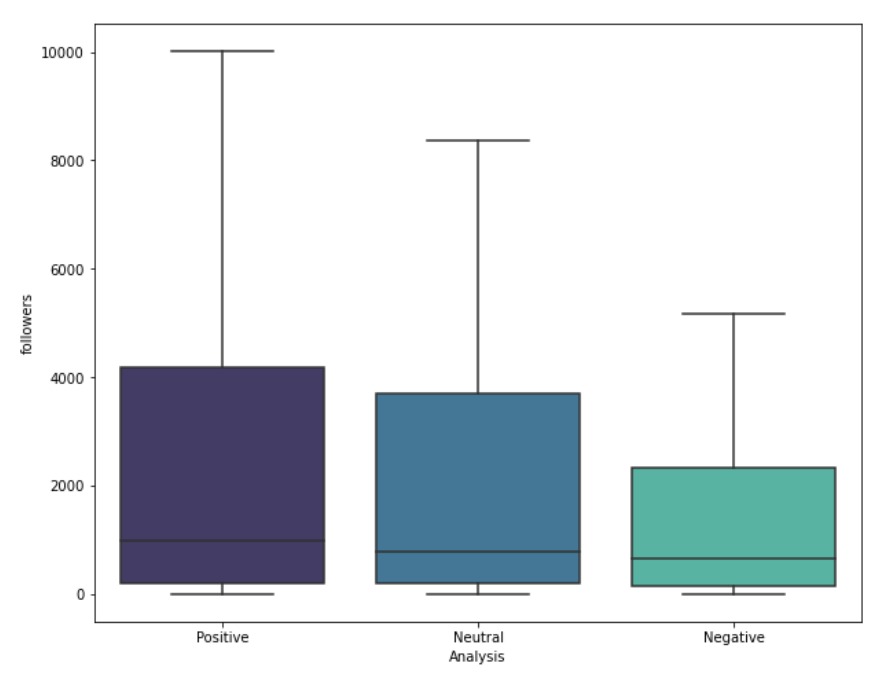
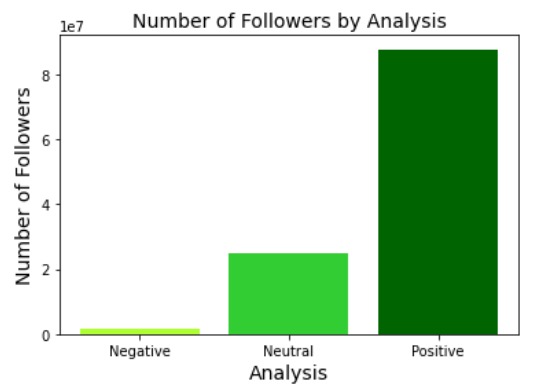
As shown in Fig. X and Fig, Y , the most liked tweets are those of a positive attitude.

**Question 11: What is the average likes per tweet for each category of Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| Analysis | Sum of Likes | Number of Tweets | Average likes |
| Positive | 10653 | 928 | 11 |
| Neutral | 2066 | 735 | 3 |
| Negative | 2125 | 255 | 8 |

As show in the table, positive tweets have the highest average likes (11 likes)

**Question 12: What is the reach of tweets (number of followers of the users tweeting) about each Analysis (whether the tweet is Negative, Neutral or Positive)?**



As shown in Fig. X, the number of followers of users with positive tweets is the highest. Thus, positive tweets covid vaccine are reach more people, which consequently portrays an overall positive idea over the vaccine.

**Popular Users Questions:**

**Question 13: Who are the top 10 users in terms of frequency of tweets? What is their overall analysis (Negative, Neutral or Positive)?**

|  |  |  |
| --- | --- | --- |
| Username | Frequency | Analysis Category |
| latestly | 38 | Positive |
| JagranEnglish | 14 | Positive |
| Timethief | 12 | Negative |
| Editorji | 11 | Positive |
| Thebuddhistgoth | 10 | Negative |
| Dhakatribune | 10 | Positive |
| The\_hindu | 9 | Positive |
| MicheleD2020 | 8 | Negative |
| Juliettejenne03 | 8 | Positive |
| HolisticZombie | 8 | Negative |

As shown in the Table, 6 out of the top 10 most frequent users in the data had a positive attitude towards the vaccine while 4 of them had a negative attitude towards the vaccine

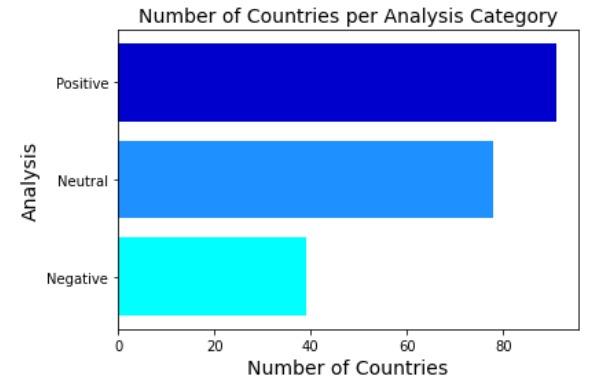
**Question 14: What is the Analysis (Negative, Neutral or Positive) of the top 10 users in terms of number of followers?**

|  |  |  |
| --- | --- | --- |
| Username | Number of followers | Analysis Category |
| MumbaiMirror | 1915 | Negative |
| firstpost | 1897 | Neutral |
| HarvardChanSPH | 1851 | Neutral |
| DDNewslive | 1752 | Positive |
| airnewsalets | 1534 | Positive |
| FinancialXpress | 1526 | Positive |
| The\_hindu | 1375 | Positive |
| otvnews | 1364 | Neutral |
| Katieecouric | 681 | Positive |
| DhakaTribune | 463 | Positive |

As shown in the table, 6 out of the top 10 most followed users in the data had a positive attitude towards the vaccine while 2 of them had a negative attitude towards the vaccine

**Country and Analytical Categories Questions:**

**Question 15: What is the number of countries that tweeted about each Analysis (whether the tweet is Negative, Neutral or Positive)?**



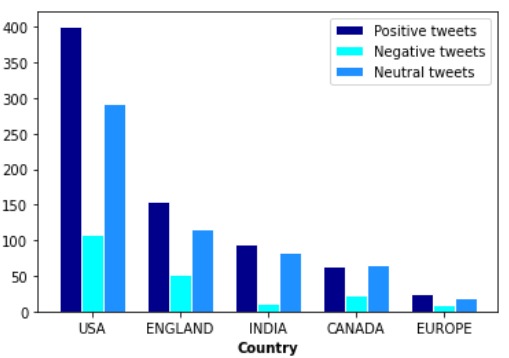
As show in Fig. X

There are more than 80 countries that have tweeted positive tweets about the covid vaccine.

There are around 80 countries that have tweeted neutral tweets about the covid vaccine.

There are around 40 countries that have tweeted negative tweets about the covid vaccine.

**Question 16: Number of tweets for the top 5 countries for each Analysis category (Negative, Neutral or Positive)?**



As shown in the graph, in all countries the positive tweets are more than the negative tweets. So we can conclude that the overall attitude towards the vaccine in the world is a positive one.

**Question 17: What are the top 5 countries with the most Negative tweets?**

**Question 18: What are the top 5 countries with the most Neutral tweets?**

**Question 19: What are the top 5 countries the most Positive tweets?**