SENTIMENT ANALYSIS DURING COVID-19 DUE TO VACCINATION ROLL-OUTS

PROJECT REPORT

Serhan TÜRKAN
MSc. Data Science
Koç University
sturkan20@ku.edu.tr

***Abstract*— For more than 1.5 years, the new coronavirus epidemic has been continuing on a global scale. The number of new cases, deaths and lockdowns affect people's psychology, mostly in a bad way. In this study, the effect of vaccination on people's emotions was examined.**

I. INTRODUCTION

People use social media to express their opinions and feelings about a particular product, organization or situation. Twitter is one of them. COVID-19 first appeared in Wuhan, China in December 2019, and people have been tweeting aggressively on Twitter ever since.

The attitude and feelings of an individual play an important role in analyzing their behavior, which is referred to as sentiments. These sentiments can then be examined in relation to a specific entity, a process known as sentiment analysis. This analysis is based on the predicted outcomes, such as analyzing the text based on its polarity. In this study, I will analyze people's emotions during the covid-19 process and examine how the vaccination process changes these emotions by using the information gathered from twitter.

Research question: Are people's positive thoughts increasing as the number of people vaccinated around the world increases?

II. DATASET

In this project "All COVID-19 Vaccines Tweets" [1] and "COVID-19 World Vaccination Progress" [2] datasets taken from Kaggle were used. All COVID-19 Vaccines Tweets contains 88860 tweets about vaccines from 2020-12-20 to 2021-06-01. COVID-19 World Vaccination Progress consists of the vaccination numbers of all countries from 2020-12-02 to 2021-05-31.

III. PRE-PROCESS AND VISUALIZATION

Since tweets have a character limit, people avoid using unnecessary expressions when tweeting. So, preprocessing must be done very sensitively for these data. The data cleaned in this study are: URLs, punctuation, emojis and some stopwords.

Figure 1. Most Frequently Used Words

Sentiment Analysis

After cleaning the data, Sentiment analysis was performed with natural language processing technique to determine whether the data were positive, negative or neutral.

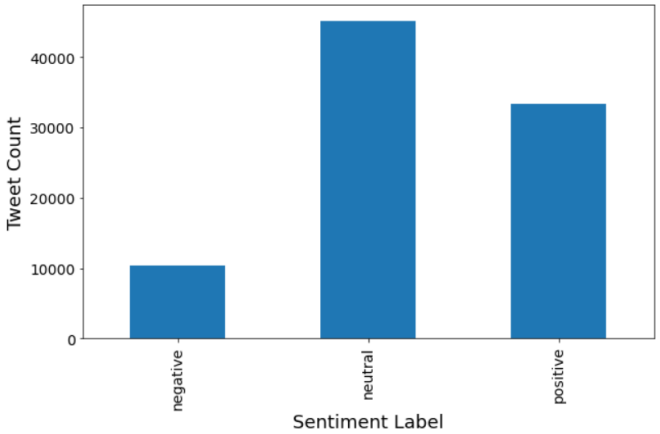
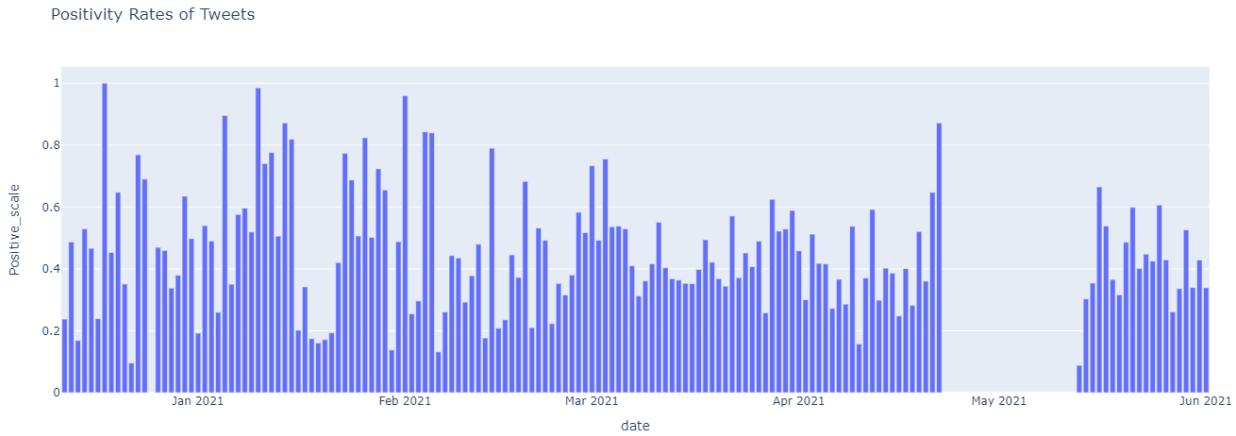
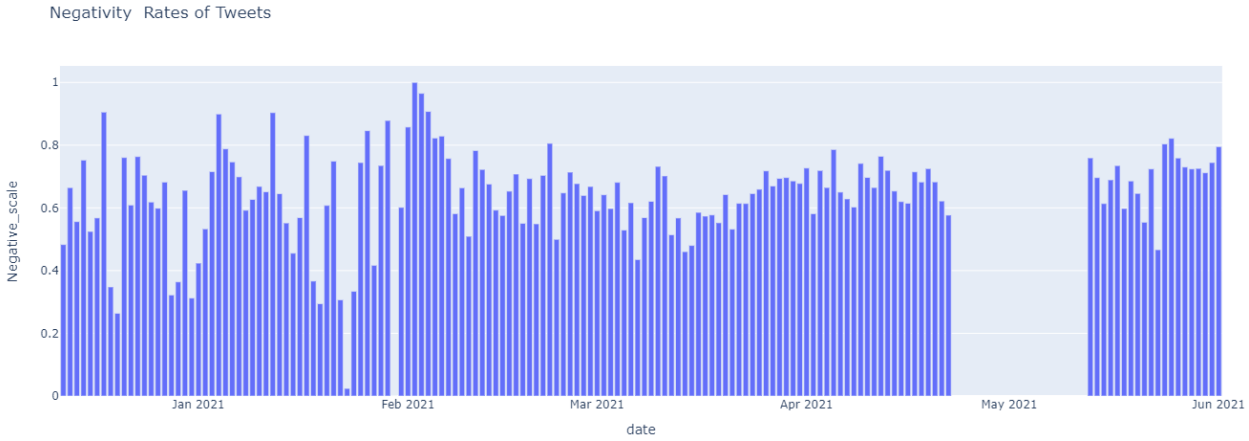
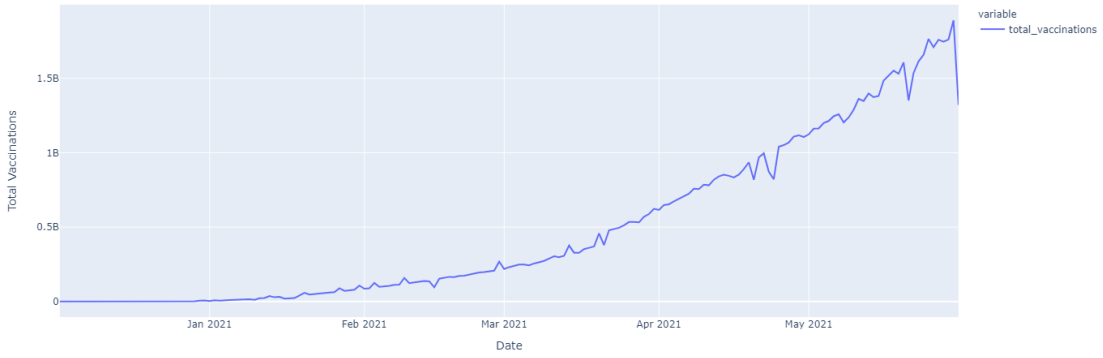


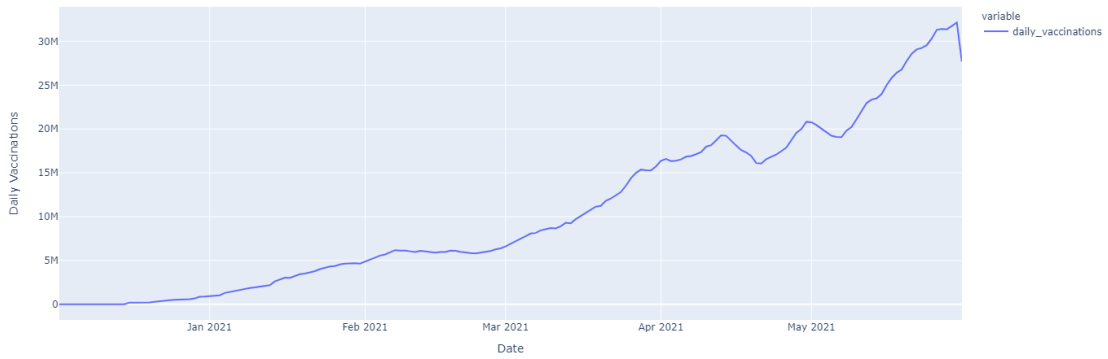
Figure 2. Sentiment Label of Tweets

After calculating the change in polarity (positivity and negativity) of tweets according to days, it was scaled for easy processing. The same procedure was performed for the daily and total number of vaccinations data.


Positivity Rates of Tweets
Positive_scale
1
0.8
0.6
0.4
0.2
0
Jan 2021
Feb 2021
Mar 2021
Apr 2021
May 2021
Jun 2021
date


Negativity Rates of Tweets
Negative_scale
1
0.8
0.6
0.4
0.2
0
Jan 2021
Feb 2021
Mar 2021
Apr 2021
May 2021
Jun 2021
date


variable
total_vaccinations
Total Vaccinations
1.5B
1B
0.5B
0
Jan 2021
Feb 2021
Mar 2021
Apr 2021
May 2021
Date


variable
daily_vaccinations
Daily Vaccinations
30M
25M
20M
15M
10M
5M
0
Jan 2021
Feb 2021
Mar 2021
Apr 2021
May 2021
Date

IV. Model Implementation

In this project, regression discontinuity design tools were used to calculate the regression between vaccination progress and emotions.

Result is as follow:

	Total vac.	Daily vac.
Positivity	0.47-0.07x	0.48-0.08x
Negativity	0.61+0.1x	0.6+0.11x

V. Results

In this study, it was investigated whether the progress of the vaccination process has a positive or negative effect on humans. As a result of the study conducted with 6 months vaccination data and vaccination related tweets, no significant link was found between these two events.

References

[1] *See* https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets

[2] *See* https://www.kaggle.com/gpreda/covid-world-vaccination-progress