Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Technologies

# Master's Thesis

## Relationship of COVID-19 Twitter hashtag with the number of cases

Solomon Erkineh

# CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

# DIPLOMA THESIS ASSIGNMENT

Solomon Erkineh

Systems Engineering and Informatics
Informatics

Thesis title

**Relationship of COVID-19 Twitter hashtags with the number of cases**

---

**Objectives of thesis**

The main objective of the thesis is to identify relationships between COVID-19 related hashtags and the daily increases of COVID-19 cases.

Specific objectives:
- To determine the relationship between COVID-19 Twitter hashtags and the number of daily increases of cases;
- To ascertain the strength of association between COVID-19 Twitter hashtags and the daily increases of COVID-19 cases;
- To interpret pertinent findings related to COVID-19 Twitter hashtags and the daily increases of COVID-19 cases in contrast to other studies and thus formulate conclusions.

**Methodology**

Methodology of this thesis is based on the study and analysis of COVID 19 hashtags on Twitter for selected countries. Practical part will consist of statistical analysis of Twitter dataset and other open sources datasets related to COVID-19. The results will be discussed and contrasted with other studies.

**The proposed extent of the thesis**

60-80 pages

**Keywords**

COVID-19, SARS-CoV-2, daily cases, Twitter, hashtag, statistical analysis.

**Recommended information sources**

CARTY, Victoria. Arab Spring in Tunisia and Egypt: The impact of new media on contemporary social movements and challenges for social movement theory. 2014.

FARHAN, Adnan Abdulrahman Naef; VARGHESE, P. A. Social media utilisation among youth. International Journal of Research in Social Sciences, 2018, 8.4: 88-96.

GROVER, Sangeeta; AUJLA, Gagangeet Singh. Prediction model for influenza epidemic based on Twitter data. International Journal of Advanced Research in Computer and Communication Engineering, 2014, 3.7: 7541-7545.

KHATUA, Apalak; KHATUA, Aparup. Leave or remain? Deciphering Brexit deliberations on Twitter. In: 2016 IEEE 16th international conference on data mining workshops (ICDMW). IEEE, 2016. p. 428-433.

SINGH, Lisa, et al. A first look at COVID-19 information and misinformation sharing on Twitter. arXiv preprint arXiv:2003.13907, 2020.

**Expected date of thesis defence**

2021/22 WS – FEM

**The Diploma Thesis Supervisor**

Ing. Miloš Ulman, Ph.D.

**Supervising department**

Department of Information Technologies

Electronic approval: 20. 7. 2020

**Ing. Jiří Vaněk, Ph.D.**

Head of department

Electronic approval: 19. 10. 2020

**Ing. Martin Pelikán, Ph.D.**

Dean

Prague on 26. 10. 2021

**Declaration**

I declare that have worked on my master's thesis titled "Relationship of COVID-19 Twitter hashtag with the number of cases" by myself and have used only the sources mentioned at the end of the thesis as the author of the master's thesis, declare that the thesis does not break any copyrights.

In Prague on 30 November 2021 ____________________________

## Acknowledgment

This project would not have been possible without the support of many people. Many thanks to my supervisor, Ing. Milos Ulman, Ph.D. for being an excellent mentor and patiently listening to the developing concepts during my work on this thesis. Also, I would like to thank the Czech Republic's government for providing me with a scholarship to pursue my master's degree.

I am thankful to my parents for their unwavering love and support, which keeps me motivated and confident. My successes and success are a result of their faith in me. My heartfelt gratitude goes to my brothers, who keep me grounded, remind me of what matters in life, and are always supportive. Finally, I owe my sincerest appreciation to Ethiopia for her unconditional love and support during the whole thesis process and every day of my life. I am forever grateful.

# Relationship of COVID-19 Twitter hashtag tags with the number of cases

**Abstract**

The global health care infrastructure, as well as the social, economic, and psychological well-being of mankind, are suffering as a result of the current coronavirus illness (COVID-19) pandemic that began in late 2018. The main objective of this study was to identify the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases. The study used daily time-series data for 12 months, from 2020 March to 2021 February. This data was obtained from Worldometers and GitHub (Lopez Bec) websites, bulletins, and other publications. Data was analyzed by the use of correlation analysis as well as bivariate regression analysis. The finding of the study revealed that there existed a positive association between daily, weekly, and monthly increases in COVID-19 and those of Twitter hashtags, also based on daily, weekly, and monthly datasets. Though a strong and positive association for monthly data existed between the two datasets. Additionally, the regression analysis revealed that a positive and statistically significant relationship between Twitter hashtags and the daily increase of COVID-19 existed. Finally, the research noticed that on-the-ground and online public health crisis response operations are becoming more concurrent and linked. Social media platforms enable the direct communication of health information to the general audience. Through social media monitoring, health systems should aim to develop national and worldwide illness detection and surveillance systems.

**Keywords:** COVID-19, SARS-CoV-2, Daily cases, Twitter, Hashtag, Statistical analysis.

# Vztah mezi Twitter hashtagy COVID-19 a počtem případů


**Abstrakt**

Globální infrastruktura zdravotní péče, stejně jako sociální, ekonomická a psychická pohoda lidstva, trpí v důsledku současné pandemie koronavirového onemocnění (COVID-19), která začala na konci roku 2018. Hlavním cílem této studie bylo identifikovat vztah mezi hashtagy COVID-19 na Twitteru a každodenním nárůstem případů COVID-19. Studie používala denní časové řady dat za 12 měsíců, od března 2020 do února 2021. Tato data byla získána z webových stránek, bulletinů a dalších publikací Worldometers a GitHub (Lopez Bec). Data byla analyzována a analyzována pomocí korelační analýzy a také regresní analýzy. Zjištění studie odhalilo, že existuje pozitivní souvislost mezi denním, týdenním a měsíčním nárůstem COVID-19 a nárůstem hashtagů na Twitteru, a to rovněž na základě denních, týdenních a měsíčních souborů dat. I když mezi těmito dvěma soubory dat existovala silná a pozitivní souvislost pro měsíční data. Regresní analýza dále odhalila, že existuje pozitivní a statisticky významný vztah mezi hashtagy Twitteru a denním nárůstem COVID-19. A konečně, výzkum si všiml, že operace reakce na krize veřejného zdraví na místě a online jsou stále více souběžné a propojené. Platformy sociálních médií umožňují přímou komunikaci informací o zdraví širokému publiku. Prostřednictvím monitorování sociálních médií by se zdravotnické systémy měly zaměřit na rozvoj národních a celosvětových systémů detekce a sledování nemocí.



**Klíčová slova:** COVD-19, SARS-CoV-2, Denní případy, Twitter, Hashtag, Statistická analýza.

## Table of Contents

**1 Introduction ........ 1**

**2 Objectives and Methodology ........ 4**

2.1 Objectives ........ 4

2.2 Methodology ........ 4

**3 Literature Review ........ 6**

3.1 Theoretical Background ........ 6

3.1.1 Twitter ........ 7

3.1.2 Basic Function of Twitter ........ 8

3.1.3 Twitter Hashtag ........ 10

3.1.3.1 Use of Twitter Hashtag ........ 11

3.1.4 Daily increases of the Novel Coronavirus (COVD-19) and History of Pandemics 12

3.1.5 COVID-19 Hashtag Tweets Dataset ........ 14

3.1.6 Social Media Networking ........ 16

3.1.7 Prediction model for Influenza Epidemic Based on Twitter Data ........ 17

3.2 Theories of information propagation in social networks ........ 20

3.3 Empirical Studies ........ 21

3.4 Chapter Summary ........ 24

**4 Practical Part ........ 26**

4.1 Research Philosophy and Design ........ 26

4.1.1 Research Philosophy ........ 26

4.1.2 Research Approach and Strategy ........ 29

4.1.3 Conceptual Framework ........ 29

4.1.4 Research Design ........ 30

4.2 Data Frequency, Data Choice, and Collection ........ 32

4.2.1 Data Frequency and Sample Size ........ 32

4.2.2 Type of Data ........ 32

4.2.3 Data Choice ........ 35

4.2.4 Data Collection ........ 36

4.3 Analysis Methods ........ 37

4.3.1 Statstical Analysis ........ 37

4.3.2 Descriptive Statistics ........ 38

4.3.3 Correlational Analysis ........ 39

4.3.4 Regression Equation ........ 39

**5 Results and Discussion .......................................... 42**
5.1 Research validity and Reliability .......................................... 42
5.1.1 Descriptive Statistics .......................................... 44
5.1.2 Correlation Analysis .......................................... 45
5.2 Regression Analysis .......................................... 49
5.2.1 Normality Test of Residual .......................................... 50
5.2.2 Serial Correlation LM Test .......................................... 52
5.2.3 Model Misspecification Test .......................................... 52
5.3 Pertinent Issues Arising from the Study .......................................... 53
5.3.1 Research Implications .......................................... 54
5.3.2 Strengths and Limitations .......................................... 55

**6 Conclusion .......................................... 57**

**7 References .......................................... 59**

**8 Appendix .......................................... 64**

# List of Pictures

Figure 2: Pearson’s Correlation Scatter Plot for Daly Data 46
Figure 3: Pearson’s Correlation Scatter Plot for Weekly Data 47
Figure 4: Pearson’s Correlation Scatter Plot for Monthly Data 48
Figure 5: Jarque – Bera Normality Test Results 50
Figure 6: Regression Residual Graph 51

# List of Tables

Table 1 Basic Function of Twitter 8
Table 2: Descriptive Statistics of the Variables COVID-19 TWIT#TAG 44
Table 3: Pearson’s Correlation Matrix Daly Data 46
Table 4: Pearson’s Correlation Matrix Weekly Data 46
Table 5: Pearson’s Correlation Matrix Monthly Data 47
Table 6: Summary of Regression Estimate Results 49
Table 7: Model statistics 49
Table 8: Ramsey’s RESET Test Results 53
Table 9: Daily Data 64
Table 10: Weekly Data 72
Table 11: Monthly Data 73
Table 12: Major Pandemics: Historical Overviews 74
Table 13: Regression Model 75

# List of abbreviations

| | |
|---|---|
| API | Application Programming Interface |
| COVID-19 | Coronavirus Disease 2019 |
| CZU | Czech University of Life Sciences Prague |
| ICDMW | International Conference on Data Mining Workshops |
| IEEE | Institute of Electrical and Electronics Engineers |
| NGO’s | Non-Governmental Organization |
| Ph.D. | Doctor of Philosophy |
| SARA-CoV-2 | Severe Acute Respiratory Syndrome Coronavirus-2 |
| SARS | Severe Acute Respiratory Syndrome |
| UK | United Kingdom |
| US | United States |
| WHO | World Health Organization? |

# 1 Introduction

Coronavirus disease 2019 or sometimes referred to as COVID-19, is a contagious disease often caused by a severe acute respiratory syndrome coronavirus 2 (SARA-CoV-2). The first known case was identified in Wuhan, China in December 2019 (Page et al., 2021). From the time this disease was discovered, it started to spread like a world fire across most parts of the globe, which has led to an ongoing pandemic(Zimmer, 2021). Furthermore, the outbreak of this pandemic has caused immense societal and economic disruption across the world. The COVID-19 epidemic has prompted governments throughout the globe to implement social distancing strategies, which have had the unintended result of highlighting and strengthening the importance of social media in bringing people together (Viral lies, 2020; Wajahat Hussain, 2020). Thus, the study of social media data such as Twitter hashtags data in relation to the daily increase in cases during the current COVID-19 pandemic can provide some unique insight into online social behaviour.

Global health care infrastructure, as well as the social, economic, and psychological well-being of mankind, are suffering as a result of the current coronavirus illness (COVID-19) pandemic that began in late 2018. On a variety of topics connected to the COVID-19 epidemic, individuals, organizations, and governments are utilizing social media to interact with one another and to share information. There is little information available on the subjects being discussed on social media platforms in relation to COVID-19. Analyzing this kind of information may assist policymakers and health-care organizations in determining the needs of their constituents and responding to those needs in an appropriate manner.With the worldwide spread of the COVID-19 infection, individual activity on social media platforms such as Facebook, Twitter, and YouTube began to increase. Studies have demonstrated that social media may play an essential role in identifying epidemics, as well as in analyzing public attitudes and behaviors during a crisis as a method to improve crisis communication and health promotion messages (Jordan et al., 2018). Public health experts rely on modern surveillance systems to better understand and monitor public health information on a worldwide scale, and these systems are designed to filter through massive volumes of real-time data from social media (Jordan et al., 2018).

The COVID-19 pandemic's impact on people's thoughts, emotions, and themes may be swiftly analyzed using publicly available data shared on social media platforms by individuals throughout the globe. Such data can help policymakers, health care professionals, and the public identify primary issues that are the concern and address them in a more appropriate manner.

Twitter, a service that publishes short news, has grown exponentially over the last two years. In January 2010, Twitter traded more than 600 messages per second and became a cultural phenomenon around the world. This success is due to the simplicity of the system and the cleanliness of websites and APIs. The ease of publication also means that Twitter has inspired contributions in a timely manner and has become an important source of information about the news already used by major search engines. Publishers benefit from Twitter's ease of use, while those that utilize and analyse Twitter data, particularly when tweets are aggregated suffer. Twitter mining applications including message and tag identification, brand management, and customer support all begin with aggregation. This is a critical first step in distinguishing personal conversations from forums. The purpose of this study was to identify the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases worldwide. This study serves as a first step in understanding the existence of a relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases worldwide. Thus, it is hoped that the knowledge from this study will help stimulate further research in this area of study and will also help policy formulators or key decision-makers about the pandemic by having a detailed understanding of how social media relates to the daily increase in COVID-19 cases worldwide. This study, on the other hand, will also provide a basis for further research by the academic fraternity and other scholars. Last but not least, this study will contribute a fair share of knowledge to the scantly available literature on the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases worldwide, both for scholars and academicians.

This study is of paramount importance because it identifies the nexus between the Daily Increases of COVID-19 and the Twitter Hashtags globally. The study focuses on the reality of Twitter Hashtags via the use regression as well as correlation analysis. The significance of the study lies on the premise that Twitter Hashtags play a relevant role on the Daily

Increases in COVID-19. Thus this study is expected to provide a new brand of information to policy makers especially those in health care which can turn out to be beneficial in their day to day management of their operations and the general public as a whole. Additionally, this study is expected to generate literature that may be of practical use to the world at large especially men and women who would use the findings to improve their operations. Furthermore, Twitter Hashtags are proved to be a significant concept in both social media as well as health and community sensitization. Future researchers specializing on social media fundamentals can find reference materials for their studies. The study is expected to enhance knowledge to the existing research. It will explore the various gaps and prompt further research by scholars and other stakeholders.

This study can also be a benchmarking tool for governments globally that are challenged by COVID-19. They can benefit from the findings of the study which can enable them to understand better on how to improve their management of crisis amidst social media. Last, but not least, this report will further serve as a partial fulfilment for the award of a Master's Degree in System Engineering and Informatics.

# 2 Objectives and Methodology

## 2.1 Objectives

The main objective of this thesis is to identify the relationship of COVID-19 Twitter hashtags with the daily increases of COVID-19 cases.

Following the main objective of the study, the following are specific objectives arising from the study:

i. To determine the relationship between COVID-19 Twitter hashtags and the number of daily increases of cases.
ii. To ascertain the strength of associations between COVID-19 Twitter hashtags and the daily increases of COVID-19 cases.
iii. To interpret pertinent findings related to COVID-19 Twitter hashtags and the daily increases of COVID-19 cases in contrast to other studies and thus formulate conclusions.

## 2.2 Methodology

According to (Saunders, 2012), research methodology is the backbone of successful research. Based on the objectives of the study, the study approach used was quantitative in nature. The study used time series data because the study investigated the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases over a period of twelve (12) months (from March 2020 to February 2021). As earlier stated, since the research used time-series data and following the purpose of the study, this study pursued an explanatory study inquiry.

Thus, according to (Saunders et al., 2003), the main strength of longitudinal research is its ability to study change and development over time. Henceforth, notwithstanding the time constraints, the series data analysis was made possible on account of the availability of data on COVID-19 Twitter hashtags and the daily increase in COVID-19 cases through the following sources; Worldometer and Github (Emily Chen, 2020) websites, bulletins, and other publications. Since the study is quantitative in nature, descriptive statistics, multiple regression analysis, as well as correlation analysis were used as empirical methods to identify the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases from selected countries. Using EViews under the practical section of the study, the results drawn were discussed.

# 3 Literature Review

The literature review presents a facet of existing literature depicted by commendable authors. Chapter 3, therefore, gathers organized and appropriate literature concerning the topic of research. It gives a thorough discussion of empirical reviews, theoretical frameworks, and exposes knowledge gaps in the prevailing literature.

## 3.1 Theoretical Background

Social media use rises during times of crisis, regardless of whether the crisis is natural or man-made. Information obtained through social media channels, such as Facebook and Twitter, is quickly disseminated in times of crisis since these platforms provide breaking news and relevant emergency response information before official sources do (Imran et al., 2020). On these occasions, individuals tend to exchange information regarding their own safety as well as the safety of their loved ones (Castillo, 2016; Imran et al., 2015). As a result, a significant quantity of socially produced data is created by having discussions continuously in public forums. The total data varies, according to (Kalyanam et al., 2016), from hundreds of thousands to millions of records. When social media data is correctly collected and used, it may be used to gather situational information that can then be combined with intelligence that can be utilized to carry out an effective reaction to a crisis. For first responders and decision-makers, situational information may be very useful in both the short-term and long-term when devising responses to a crisis.

Social media platforms that are most often used nowadays for casual interactions include Facebook, Twitter, Reddit, and so on. The application programming interface (API) for accessing Twitter data (tweets) is well-documented. As a result, it is a significant resource for academics who are doing research in the Socal Computing area. Crises have demonstrated that Twitter messages about them may offer greater insights into the events than previous similar studies conducted in the past, such as those by (Carley et al., 2016; Chatfield et al., 2013; Landwehr et al., 2016).

There have been many crises in the past when millions of tweets have been gathered and made accessible, including the Nepal Earthquake, India Floods, Pakistan Floods, the Palestine

Conflict, Flight MH370, and so on (Imran et al., 2016). This data was used in the development of machine learning models for categorizing tweets that were not seen into a variety of categories such as community needs, volunteer efforts, deaths, and infrastructure damage. This data was used in the classification of tweets that were not seen into numerous categories such as community needs, volunteer efforts, deaths, and infrastructure damage (Imran et al., 2014). According to  (Purohit et al., 2013), which used Twitter corpora, classified tweets may be trimmed or summarized and sent to relevant departments for further analysis. Last, they can also be used for sketching heat maps that display alert-level fake news sentiment (Bondielli & Marcelloni, 2019). If users are alerted to misinformation and unsubstantiated rumors before they become viral, they may be classified as spam or removed. Even further, text-based in-depth studies of Twitter data may reveal, respectively, if a geographic area is being favorably or adversely textual with regard to a crisis, as well as how information is being spread during a crisis.

When the COVID-19 outbreak began, individuals utilized social media to check up on the latest facts. The content is referred to as an infodemic in this instance due to its inaccuracy and confusion. When individuals are experiencing a crisis, they spend more time on social media than normal. Social media platforms such as Twitter have gained prominence as a source of information due to their ability to disseminate information faster than traditional news sources and emergency response organizations during times of disaster. (Imran et al., 2020). For this reason, it is based on this that the argument for carrying out a study such as this rests. In other words, the goal of this study is to find or examine the reasons why an infection or pandemic is transmitted via a face-to-face friendship network and to learn how information on the network, such as Twitter hashtags on COVID-19, such as the COVID-19 related Twitter hashtags "COVID 19 Unveiling," helps predict the daily increases in COVID-19 cases.

### 3.1.1 Twitter

As (Edosomwan et al., 2011) tried to explain in the paper, Twitter was launched in 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams. After Facebook opened the door for multiple social networking sites to be launched, one of them was Twitter. It is an online networking website, and its basic role is to interface individuals and make individuals influence the major crowds. Twitter enables users to find stories about the latest trending news and occasions, follow individuals or organizations that tweet about content they appreciate spending, or speak with their associates, But as (Edosomwan et al., 2011; Jasra, 2010) explained in the paper, what makes Twitter different from the others is that it offers more different options, such as microblogging. It was mostly used by celebrities, and it is the most widely used microblogging platform, with 218 million monthly active users and an average of 500 million tweets each day and 335.

As per the research nowadays, numerous individuals use Twitter, and this enables individuals to get the chance to access what they are saying and doing. As a result of this, it makes Twitter well known by youngsters. Given that youngsters are normally conceived as multi-taskers, Twitter is famous because it suits the cutting-edge, face-paced world we live in. As laid out, tweets fly around the internet.

### 3.1.2 Basic Function of Twitter

As (Maclean et al., 2013) explain well the basic functions of Twitter below

***Table 1 Basic Function of Twitter***

| **Term** | **Definition** |
| --- | --- |
| Follow | we will see all of their tweets in our feed if we follow them.<br><br>we may follow them by clicking on their username, which will bring up their profile on the right side of your screen with a green Follow button. To get started, just click here. |
| Who to follow list | This is a list of Twitter's recommendations for persons or organizations that you may be interested in following based on similarities between our profile and theirs. Select anybody we wish to follow by clicking on the green Follow button next to their name as we go down the list. |

| | |
|---|---|
| Unfollow | Go to following list and identify the person we wish to unfollow, then hover our mouse over their green Following button until it is replaced by the red Unfollow button, and then click. |
| Block | Sometimes a spammer or other unsavory persona may show up in our Followers list. Unwanted followers may be unfollowed by clicking the head and shoulders symbol next to their name, which brings up an option to "Block [their name]." Clicking this will remove them from your Followers list.<br><br>It's a good idea to additionally click 'Report [their name] for spam' if we're dealing with a spammer or malware user.<br><br>Our 'Followers' list should be reviewed often. New followers appear at the top of the list on Twitter. |
| Retweet | Hover over a tweet in our feed and choose retweet to share it with our own followers. A little arrow indicator denotes that it wasn't initially our tweet, and it then goes out to all of our followers. |
| Reply | Using the @Mentions column, mouse over the tweet we want to react to and pick the Reply option. Check our @Mentions column to see whether they've responded to us. |
| At (@) | When mentioning another user in a tweet, the @ symbol is used. First of all, the first portion of every Twitter account name – @someone |
| Mentions | In this case, it is when we are mentioned in a comment or a post so that we can share what we think about, and we know it when we get a notification that says that someone mentioned us in their post or some article that we need to see. |

| | |
|---|---|
| Direct Message( DM ) | This is used when we want to send a personal message for a specific person or in a group it is only applicable if the person, we are sending follow us or if we follow them |
| URL’s | URL shortening services like shorturl.at and cutt.ly allow us to post shorter URLs into our tweets since a standard web address is fairly lengthy and clunky. Once we've copied and pasted the URL of the page we want to share, it'll be given a short link that will send anybody who clicks on it back to the website you want to share. |
| # (hashtag) | To classify tweets, use a hash tag. Trending subjects, such as #UnityforEthiopia and #NoMore, are often accompanied by hashtags. Many different users' tweets will be seen when clicking on one of the topics mentioned on the homepage. Including already-popular hashtags in a tweet may help it get more exposure.<br><br>Also known as "backchannel communications," hashtags may be used in conjunction with events such as conferences and television shows, as well as worldwide events. Attendees of an event may interact with each other during the presentation by posting comments, questions, and links. |

### 3.1.3 Twitter Hashtag

A hashtag is a word or watchword expressed before a hash, in any case, called the pound sign (#). It's used inside a post through electronic systems administration media to help the people who may be enthusiastic about your point have the choice to find it when they search for a catchphrase or explicit hashtag. It helps by causing us to see the posts and invigorate our coordinated effort.

When using an articulation as a hashtag, use it without spaces, for example, # COVID19. It can combine numbers but not images or highlights. The hashtag can be set close to the beginning, focus, or end of your web-based life post or comment, and it allows what you have written to be recorded by the web-based systems administration framework. With this system, people who are not your fans or followers can still find your substance.

#### 3.1.3.1 Use of Twitter Hashtag

Twitter, as a service that publishes short news , has grown exponentially over the years. In January 2010, Twitter traded more than 600 messages per second and became a cultural phenomenon around the world. This success is due to the simplicity of the system and the cleanliness of websites and APIs. The ease of publication also means that Twitter has inspired contributions in a timely manner and has become an important source of information about the news already used by major search engines. As a community solution to these problems, Twitter users have introduced rules that add a hash to a hash at the beginning of a word.A hashtag is an identifier for a discussion that focuses on the same topic (Sajjad, 2017). By including a hashtag in the message, the user can view the conversation to which the message is associated.

In (MacDonald, 2017) research, he explained that hashtags help increase web-based life nearness as they make one’s substance distinguishable by any individual who has an enthusiasm for hashtag as it goes past simply one’s adherents. Hashtags can assist with building a brand for a business that is, by drawing in clients and joining the discussion about what's slanting.

Furthermore, individuals are encouraged to join in conversations about a similar issue via the use of hashtags on Twitter, and to draw in with one another regardless of whether or not they are following one another. Additionally, using hashtags on Twitter is also deemed to increase one’s internet-based life nearness, using lengthy and longwinded hashtags in a tweet, such as # Impact of Influenza in the Economy, can detract from the substance of your message, because hashtags are supposed to make things easier to find and draw in with others. Using an excessive number of hashtags in a single post, in addition to using too long hashtags in a single post, may be detrimental to online social networking presence. Tweets which have more hashtags than actual text, will be seen as edgy by followers, and the message sought to be conveyed somehow gets weakened as a result (MacDonald, 2017).

### 3.1.4 Daily increases of the Novel Coronavirus (COVD-19) and History of Pandemics

As of July 2020, the daily increase of novel coronavirus (COVD-19) had reached already alarming thirteen million cases, and the death toll had since then crossed over half a million (*Worldometers*, n.d.). However, as of January 2021, the total number of cases had escalated to a huge number of more than a hundred million, with approximately three million deaths (*Worldometers*, n.d.).

Meanwhile, states and countries worldwide are trying their best to contain the spread of the virus by initiating lockdowns, taking newly introduced vaccines, and even curfews in some regions. Social distance has now become the new normal for individuals who work from home. As the number of cases has increased, individuals have been increasingly engaged on social media with regard to the epidemic. Several unique words associated with the epidemic have recently emerged on social media. Researchers working in the theme areas of Social Computing, including but not limited to sentiment analysis, topic modeling, behavioral analysis, fact-checking, and analytical visualization, may benefit greatly from Twitter data.

To train machine learning models or conduct any sort of analysis, large-scale datasets are needed. Extracting information from tiny datasets and region-specific datasets is limited by the size of the datasets. So, in this way, our study incorporates a dataset of almost two million COVID-19-specific English language tweets.

A worldwide epidemic has seized the world in the early part of 2020. A new coronavirus (Coronavirus Disease-19, or COVID-19) was discovered, and it was given the name SARS-CoV-2 (Mu et al., 2020; Qiu, Yang,Chen et al., 2020). The significant growth of the COVID-19 virus across the globe has been accompanied by economic devastation and human sorrow, the worst consequences of which occurred in Wuhan, the Hube province of China. The global incidence of COVID-19 by February 28th had surpassed 114 million, with approximately 2.6 million fatalities.
Other nations have taken various public health steps to prevent the spread of COVD-19, including social distance (Fong et al., 2020). Businesses, schools, community centers, and

NGOs were all forced to close, as well as large meetings. All movement was permitted solely for necessities only. It is the aim of this project to help nations "flatten the curve" by "reducing the number of new cases linked to COVID-19 from one day to the next" (University, n.d.).

In the history of the world, disease epidemics have never been new (Bartsch et al., 2020; Ferguson et al., 2020). You can see a chronology of some of the biggest pandemics in the world in Table 9 in the appendix.

Following the year 2000, the frequency of pandemics has increased significantly, rather than the number of outbreaks and human catastrophes, which has been a notable feature of the period. Because of the increasing development of viral diseases in animals, this is becoming more common (Madhav et al., 2017). As a consequence of increasingly frequent outbreaks of illness, according to researchers such as Garrett (2007; Keogh-Brown et al., 2008), the newest additions include (Madhav et al., 2017) and (Fan et al., 2018), the probability of a large-scale global pandemic is increasing. Following the publication of a recent article in the Imperial College London COVID-19 Response Team report, COVID-19 has been dubbed "the most devastating epidemic since the 1918 Spanish influenza pandemic," which resulted in the deaths of tens of thousands of people worldwide.

However, as far as decreasing total mortality during the 1918 Spanish flu pandemic is concerned, (Barro, 2020) reaches the opposite conclusion; the non-pharmaceutical measures that were undertaken in 1918 did not succeed. As interventions were not sustained for an adequate length of time, this occurred. The range for the overall average of the various numbers that he calculated was between 22 and 48 days, with the highest number in his study being 36 days (0.05 years). Compared to the length of time the 1918 Spanish influenza pandemic was active, these figures were relatively modest.

From the time COVID-19 announced its presence and based on the cumulative cases and deaths experienced for a period of almost one year, approximately 114 million and over 2.6 million deaths were recorded across the world by October 23, 2021. Some countries, such as the US, UK, Brazil, and some European Union countries, such as Italy, France, Spain, and Germany, were deemed to be the worst hit by the number of cases and deaths.

According to health officials, COVID-19 has a disproportionately negative effect on the elderly when compared to past pandemics. Unlike its predecessors, however, the lockdown measures have a more global reach and scale, and they have disrupted international supply networks as well as aggregate demand and consumption patterns, according to the World Bank. Because of this, financial market volatility has increased, and the economic shock has been magnified even more. Furthermore, increased borrowing and higher debt levels among businesses and families during this period make the short-term shocks more powerful when compared to past pandemics, as previously stated (Boissay et al., 2020).

### 3.1.5 COVID-19 Hashtag Tweets Dataset

Numerous additional studies have also compiled and shared large-scale statistics to aid researchers in comprehending the public conversation about COVID-19. Several of the freely accessible datasets are global, while others are language-specific. (Alqurashi et al., 2020; Banda et al., 2020). In recent times, hashtag tweets have emerged as a preferred means of relaying information via social media.

The number and variety of hashtags being used in the data is both rising and becoming more prominent with from time to time (Cruickshank & Carley, n.d.). Hashtags typically grow throughout the period of data collection, and their usage also shows cyclical patterns that tend to increase and decrease on a weekly basis (i.e. slight drops in the number of unique hashtags being used on weekend days). The number of unique hashtags counted is important since it does not represent the entire number of hashtags that were used. So, it is possible that, as the scope of the COVID-19 pandemic expanded, hashtags that were originally not associated with the COVID-19 pandemic ended up becoming a part of the conversation. It is also possible that, as the scope of the pandemic expanded, new hashtags were invented to better address the changing needs of the conversation about the pandemic. Although there is a high ratio of hashtags being used in tweets across all the data, the numbers are much higher when it comes to Instagram posts (greater than 40 percent of tweets have a hashtag). Other researchers have noticed a growing trend of hashtag use on social media platforms (Sheldon et al., 2020; Zhang, 2019).

Twitter was launched in July of 2006 and features a microblogging platform in which users communicate in real-time by using "tweets" (a.k.a. tweets) which range from a maximum of 280 characters, and which include hashtags (or #) to help organize the tweets. Direct messaging, status updates, responses, retweets, and likes are ways users interact with one another (retweeting is the reposting of specific messages). Following a user does not require authorization, and reciprocation is not required. The latest Twitter user statistics report for Q1 of 2019 states that there are an estimated 330 million active monthly users (Kullar et al., 2020).

Twitter serves as a vehicle for individuals to express themselves, including their ideas, emotions, and health conditions, in real-time (Panuganti et al., 2020). The top tweets on the Internet featured "likes," "retweets," and "comments." When you ‘like' a tweet, you're showing that you like the content, while when you retweet a tweet, you've shared it on your own Twitter profile, and when you "comment" on a tweet, you have contributed more written content. Increasing these efforts will expand the spread of the particular thriving posts (Rufai & Bunce, 2020). Many healthcare experts use Twitter, a daily part of their lives, to communicate information and warnings about a specific topic in real-time to a wide global audience, especially those thought to be experts in a particular field. They also use Twitter to respond to people in their respective fields (Kullar et al., 2020).

It is at times of crisis and tragedy that Twitter is the most popular social media. This is because of its capability to constantly republish relevant information (Vieweg et al., 2010). Twitter is an effective way to spread information, but it is also a kind of social media that has a major impact on how people connect with each other. It has recently created a new feature known as hashtags, which helps to search for important phrases or subjects that are currently in use. For our purposes, # is shorthand for a term or subject on Twitter. The following feature was built on Twitter and makes it easy for individuals to follow certain subjects that interest them. To classify their tweets, users first use the #(hashtag) before a relevant term or phrase in their tweets. You may see all tweets that use a hashtag by clicking or pressing on a hashtag in a message. Let's say you use a hashtag in a tweet, for example. A result of this is that famous hashtagged phrases often become trending topics, such as COVID-19.

### 3.1.6 Social Media Networking

Web-based technologies like social media are helpful for two-way communication between people, enabling two-way conversation on the web. Different kinds of social media arose thanks to the creation of the first type of social media, and for many individuals, using these networks is a daily habit (Boyd & Ellison, 2007; Dastani et al., 2019; Dastani & others, 2016). Some of the most significant websites and popular social networking apps include Instagram, Telegram, Facebook, Twitter, and WhatsApp (Amani et al., 2020; Dastani et al., 2019).

They provide greater utility in the process of creating and sustaining new connections, managing existing and previous relationships, enabling communication, boosting social engagement, and enhancing know-how and practical abilities (Dastani et al., 2019; Pempek et al., 2009). (Dastani & Ramezani, 2017). Alternatively, disseminating incorrect information and rumors, as well as inadequate supervision of social media networks (Dastani et al., 2019), is one of the major dangers posed by social media.

According to many studies, people may use internet information while facing a public health emergency, such as a pandemic (Gui et al., 2017; Rizo et al., 2005). Individuals' emotions and ideas (ideas and emotions exchanged amongst internet users) lead people's worries and views (concerns and opinions that shape people's awareness and actions) to grow. Even if these ideas and expertise aren't relevant on the Internet, they are still important. There is a risk that inaccurate medical information is distributed, which may lead to wrong treatment or an increase in stress (Auter et al., 2016; Barros et al., 2020; Wang et al., 2019).

Millions of individuals on social media during the COVID-19 epidemic exchanged and gained knowledge about the virus. However, public opinions on social media are associated with traits of severe ineffectiveness, irrationality, and uniformity. Web-based networking and social technologies have dramatically expanded the number of citizens who may get involved in disaster response, due to the widespread use of social media (Goodchild & Glennon, 2010; Liu et al., 2014).

As location-based social networks such as Twitter and Facebook gain in popularity, more individuals use them to capture real-time geo-located data and transmit information about the area around them (Chae et al., 2014). Social and technical systems are dependent on platforms such as Twitter, which enable individuals to be linked even during crises (Chen et al., 2020b).

For instance (Barton, 2021) Matt Zeller is one of "hundreds of thousands" of Afghan translators being evacuated as part of a network. Taliban checkpoints near Kabul's airport are being identified using satellite footage by security experts. At least 20,000 translators and members of their families are believed to be stranded in Afghanistan. One interpreter's brother said that the Taliban would "eliminate" Afghans with ties to U.S. soldiers. In order to get past the Taliban, Afghan translators had to wait up to 10 hours in "horrendously hot, humid circumstances."

The interpreter was able to leave Afghanistan thanks to Zeller's assistance. He claims that failing to remove the Afghan interpreters will cause veteran suicide as a result of the moral hurt. The sound of gunfire reverberated throughout the voice recordings that a female Fox News correspondent submitted. American soldiers at Kabul's airport could have heard the gunfire because of where they were stationed.

### 3.1.7 Prediction model for Influenza Epidemic Based on Twitter Data

A reaserch of (Grover & Aujla, 2014) says that the social media provides real-time information, as opposed to the traditional media, which only provides information when a major event occurs, the social media plays a more important role in inhibiting and mitigating the spread of influenza, such as H1N1, than the traditional media (Grover & Aujla, 2014). In this case, even a small delay is dangerous, as an example, if we solely depend on the traditional media, the disease will spread on a global scale. Early discovery is only feasible via the use of social media, namely micro-blogging services such as Twitter, Facebook, and MySpace. Twitter has become a popular social media platform for individuals to broadcast their current state (tweets) based on their mood, health difficulties, relationships, and other factors(Grover & Aujla, 2014). These tweets are also updated by mobile phones, allowing us to determine the user's actual position as well as the current weather conditions.

Specifically, the modeling technique addresses the emergence of any epidemic, with the emphasis on swine flu, which must be stopped at an early stage for control measures to be successful if it spreads beyond that point. Twitter data was selected for this purpose since social media provides real-time information, but conventional media only provides information after a major event has occurred, which cannot be utilized for early warning reasons and is thus ineffective. Most people believe that early discovery is only feasible via social media, namely microblogging websites such as Twitter and Facebook, among others (Grover & Aujla, 2014). All information (status updates, tweets, and so on) kept in social media communities may aid in the discovery of data pertaining to epidemics. It is possible to get both healthy (correct) and false (inaccurate, inconsistent, or inadequate) information from this source. It is necessary to guarantee that the information provided by the user is true and that it can be utilized to encourage good health information and enhance the health outcomes of individuals rather than misleading information that may endanger public safety and should be avoided.

The data utilized in the (Grover & Aujla, 2014) research was gathered from Twitter since the social media platform offers free APIs that may be used to gain a sampling perspective of the population . As discovered by, Twitter provides untested information and data sources from which one might extract information on the commencement of flu outbreaks and the spread of the virus. Users that tweet about illnesses such as "high fever," "I got FLU," "Swine Flu," "H1N1," and other such terms are given special consideration. It is possible to assess the model's correctness by comparing its findings to those received from the CDC (Centers for Disease Control and Prevention) data, which is the real data gathered through the manual registration of influenza cases.

(Grover & Aujla, 2014) research report address the development of a prediction model for influenza based on Twitter data.  The researchers give a general review of various works the work of (Huang et al., 2013). This prediction model development addressed using data from the social media.

The major gaps and limitations of the models for the detection of epidemics are using social communities like Twitter, Facebook, etc., to collect real-time data that can help us prevent and detect epidemics.

Major limitations of the previous modeling works according to Grover and Aujla included lack of real time information, machine learning tools that can be used for predicting and detecting the epidemic.

In 2013, (Huang et al., 2013) developed a model for detecting influenza transmissions, which was published in Science. The idea is based on "Sina Weibo," a Chinese micro-blogging website that is similar to a cross between Twitter and Facebook in its functionality. They displayed the findings of their detection and transmission experiments at the city level. According to (Lee et al., 2012), tweets from Twitter were compared to Influenza-like Illness (ILI) data sets, meteorological conditions, and flu predictions in order to see whether there was a link between the two. As a consequence of their research, they created comparison graphs between flu signals on Twitter and meteorological parameters and flu prediction. In (Achrekar et al., 2011) and in (Chen et al., 2010), researchers proposed a SNEFT architecture model that included crawler, predictor, and detector components to predict influenza activity using information gathered from micro-blogging websites such as Twitter and Facebook, among other sources. According to (Sakaki et al., 2010), they achieved a 95 percent connection between their findings and national health statistics and developed an earthquake reporting system to identify seismic activity. That framework investigated the real-time aspect of Twitter, primarily for the purpose of event detection and identification. (Tang & Yang, 2010) proposed a model in which each Twitter client was treated as an earthquake sensor, and based on those sensory observations, an earthquake event was predicted. They also introduced an algorithm called User Rank that took into account the link structure, content similarity, responding order, and time of repliers to calculate the number of users affected by influenza (swine flu) within a social networking community.

As a result, (Grover & Aujla, 2014) provide a new framework for resolving the shortcomings of prior modeling methods, in which their model incorporates a variety of techniques such as machine learning algorithms, time series classification and prediction, and other techniques. An epidemic stage identification model based on a probabilistic model of language (BOWs) used in various phases of the epidemic and communicated online via the act of tweeting was developed for this model, which included time series classifications and predictions for each stage of the epidemic.

The focus of this research effort was on a related study of algorithms and methodologies, modeling the development of any epidemic with a particular emphasis on swine flu, which must be stopped at the earliest opportunity if it is to be contained.

This newly developed model incorporates a machine learning technique in order to create a model trained with the new idea of the Swine Epidemic Hint algorithm, which will monitor epidemic activities occurring on Twitter, as well as a Markov Chain state model, which will categorize epidemic activities into three stages (Beginning of Epidemic, Spread of Epidemic, and Decay of Epidemic).

## 3.2 Theories of information propagation in social networks

Numerous network scientists have examined the issue of dampening or eradicating erroneous information spread via social networks (Cho et al., 2019). False information is broadly classified into two types: disinformation and misinformation. Disinformation is intentionally created false information that is exchanged and spread with malice (Cho et al., 2019). By contrast, misinformation is erroneous information that is spread unintentionally and without malice. Numerous current strategies for mitigating or eliminating erroneous information in networks focus on ways of determining which seeded nodes (or agents) to address based on their network characteristics (for instance centrality attributes) (Cho et al., 2019). The purpose of these techniques is to transmit accurate information in the most effective manner possible. However, little research has been conducted on the role of uncertainty in agents' opinion formation. Uncertainty-aware agents may generate a variety of views and eventually beliefs regarding accurate or erroneous information, resulting in a variety of information dissemination patterns in networks.

According to Cho et al. (2019), they selected a model of opinion called Subjective Logic (SL) since it clearly addresses the degree of ambiguity in an opinion. The degree of uncertainty is simply understood as a person's confidence in a certain belief or denial. (Cho et al., 2019). However, SL analyzes uncertainty only as a result of a lack of knowledge (i.e., ignorance), not as a result of other factors, such as contradictory evidence. In the age of Big Data, when we are inundated with information, contradictory information may exacerbate uncertainty (or ambiguity) and have a bigger impact on public opinion than a lack of knowledge (or ignorance). To help SL cope with ambiguity and ignorance, Cho et al. (2019) present an SL-based opinion model that incorporates a degree of uncertainty drawn from both sources. Cho et al. (2019) capture the development of agents' views over time by constructing a variation of the Susceptible- Infected-Recovered epidemic model in which the condition of an agent's beliefs may affect their status. Cho et al. (2019) investigate the essential modification outcomes as a function of key design parameters such as the scaling factor of correct or incorrect information propagation, the centrality metrics used to seed false and true informers, a viewpoint degeneration factor, the extent of agents' prior belief, and the percentage of true informers. Cho et al. (2019) validated the opinion model by comparing it to both synthetic and real-world network conditions. The real-world network settings comprised a genuine network design, user behaviors, and the quality of news articles. The suggested agent opinion model and associated tactics for dealing with erroneous information might be used to counteract the spread of fake news across several social media platforms (e.g., Facebook).

## 3.3 Empirical Studies

Multiple studies have performed social network analysis on Twitter data related to the COVID-19 pandemic. A case study conducted by (Gruzd & Mai, 2020) examined the propagation of the # FilmYourHospital hashtag using social network analysis techniques to understand whether the hashtag virality was aided by bots or coordination among Twitter users. Their findings revealed that the spread of misinformation can be potentially mitigated by fact-checking and directing people to credible sources of information from public health agencies. False and misleading assertions that are motivated by politics and backed by strong convictions rather than scientific evidence, on the other hand, are considerably more difficult to disprove.

Another study conducted by (Ahmed et al., 2020) Between March 27, 2020, and April 4, 2020, they gathered tweets with the hashtag # 5G Coronavirus and did network analysis to better understand the drivers of the 5G COVID-19 conspiracy theory and techniques for dealing with such material. In their results, they discovered that social network analysis indicated that the 2 biggest network topologies were made up of an isolated group and a broadcast group, respectively. Furthermore, the analysis also revealed that there was a lack of an authority figure who was actively combating such information.

A regional study by (Park et al., 2020) concerning South Korea used network analysis to investigate the information transmission networks and news-sharing behaviors regarding COVID-19 on Twitter. In the paper, Park et al. (2020) investigated the information transmission networks and news-sharing behaviors related to COVID-19 on Twitter using network analysis, regional research conducted by Park et al., 2020. Their results found that the transmission of information in the Coronavirus network was quicker than in any of the other networks (Corona19, Shincheon, and Daegu). In February 2020, the Korean government planned to move its citizens from Wuhan, and thus also experienced an increase in the number of Twitter users in South Korea. A similar study done by (Kim, 2020) focused on social network size and incivility and investigated the Twitter messages sent by South Koreans between February 10, 2020, and February 14, 2020, when the government was moving people. Income and social networks have a correlation, but the research found that they have distinct roles when it comes to incivility.

In a recent study, researchers used sentiment analysis on a variety of distinct sets of COVID-19 Twitter data. Using frequencies of unigrams and bigrams, as well as a sentiment analysis and topic modeling process, (Abd-Alrazaq et al., 2020) examined 2.8 million COVID-19-specific tweets gathered between February 2, 2020, and March 15, 2020, to determine who tweeted about each subject often. Researchers found that ground and online crisis response operations are becoming more integrated. Social media gives the public a chance to directly engage in health discussions. However, national and international disease detection and surveillance systems should strive to improve illness detection by monitoring social media.

Another study (Lwin et al., 2020) analyzed tweets from January 28, 2020, to April 9, 2020, to see which emotions were expressed globally throughout the global pandemic (fear, anger, sorrow, and joy). Researchers discovered that during the COVID-19 pandemic, emotion-driven communal problems were emerging around common public distress experiences, including mass social isolation and loss of human life.

According to a Spanish study conducted by (de Las Heras-Pedrosa et al., 2020), the communication of risk influenced emotions in Spanish society during the 2009 H1N1 pandemic, and in order to find out more about this relationship, sentiment analysis was used to collect conversations from various social media platforms, including Twitter and Instagram. Using the same type of media resulted in peaks of sadness, disdain, anger, and terror being experienced.

In a similar study conducted in China and Italy (Su et al., 2020), a follow-up study on individuals' psychological states that took place after the implementation of the COVID-19 lockdown looked at the conversations on Webo (for China) and Twitter (for Italy) by analyzing the two weeks' worth of posts published before and after the lockdown. According to the findings, people were more concerned with "home" during the lockdown, and after being confined, they exhibited better levels of cognitive functioning in both Wuhan and Lombardy. Meanwhile, stress levels dropped, and people in Lombardy were free to devote more time to leisure after the lockdown. Prior to the lockdown, Wuhan paid greater attention to the needs of groups, religions, and emotions. Findings let decision-makers use reliable information on public responses and psychological states in the COVID-19 setting.

Since the COVID-19 illness outbreak created a major problem, new information concerning this has rapidly become common knowledge and has also been referred to as an extraordinary surge of information (Kouzy et al., 2020). COVD-19-related tweets had the potential to reach as many as 6.737 million people in January, and as many as 110.2 million people in March, according to Banda et al. in their study.

The number of people's discussions linked to their experiences and symptoms of COVID-19 illness, the lack of access to testing, and the disease recovery of those exposed or suspected also rose significantly throughout the study period (Golder et al., 2020). (March 3 to 20).

Social media, such as Twitter, is used to trade, inform, search, and transmit information to many individuals (Aguilar-Gallegos et al., 2020). There are other people that use this social media platform, and they too want to share information and warnings with their friends and followers regarding COVID-19 (Abd-Alrazaq et al., 2020). The increased dissemination of

COVID-19 in any geographic location has fueled the rise in the number of tweets and hashtags pertaining to this illness, as well as an increase in the amount of online research into symptoms (Bisanzio et al., 2020; Chen et al., 2020; Golder et al., 2020; Klein et al., 2020). It includes information on government actions related to Corona, the number of individuals affected in the nation, and treatment and preventive options for COVID-19 (Yum, 2020).

Twitter users use various hashtags alongside tweets and information about COVD-19, depending on the type of information shared and the role of information sources (politicians, health care providers, and regular users), including hashtags used in related tweets about COVD-19 and other hashtags such as Coronavirus, Pandemic, Corona, COVD-19, Wuhan. (Aguilar-Gallegos et al., 2020; Emily Chen, 2020)

According to (Aguilar-Gallegos et al., 2020), multilingual hashtags regarding the coronavirus subject are very diverse and spread throughout a 23-day period between January 21 and February 12, 2020. SARS, nCOV, and Coronavirus were widely used by Twitter users prior to the formal release of the name of the new coronavirus, COVID-19, by the World Health Organization (WHO). On February 11, 2020, when the formal announcement was made, Twitter users began using this name, the announcement of which was made on January 24th.

## 3.4 Chapter Summary

Most recent studies into online social behavior during the COVID-19 pandemic have largely focused on how information operates during a pandemic. This is due to the fact that accurate knowledge is a crucial factor in battling the pandemic's impacts (Hussain et al., 2020). Thus, if we accept that Twitter is an important platform for communication, then there are good reasons for focusing research on the relationship and interactions between Twitter and the COVID-19 pandemic cases. In trying to relate Twitter discussions and emotions about COVID-19 with its prevailing daily increasing cases, most studies that have been conducted so far by scholars such as (Lamsal, 2020; Xue et al., 2020) and many others have mainly focused on how social media such as Twitter are a home ground for misinformation, hence creating what is commonly known as an infodemic in relation to the spread of pandemics such as COVD-19. However, in contrast, this study's focus is mainly to try and address a question about how Twitter hashtags are related to the daily increase in COVID-19 cases worldwide.

Much research on the relationship between social media and COVD-19 has been performed, according to the literature that was examined. In other words, though, most of the research looked at how information about the epidemic has been handled. One of the goals of this project is to learn more about the connection between COVID-19 hashtags and the rise in COVID-19 instances. Thus, this research is focused on investigating how human movement may be determined by means of using Twitter social media to forecast the spatiotemporal distribution of COVID-19 throughout the world (Bisanzio et al., 2020).

# 4 Practical Part

This section starts with the overall research approach and strategy in order to provide an overview of the foundation of the methodology. Thereafter, the data choice were examined, its frequency, and collection method used in line with the research strategy and design. This chapter also explains the kind of data that was used and where the data was sourced from. Furthermore, a detailed explanation is given of which tools were used to analyze the data and how the analysis was done. This chapter, in short, explains how the research questions were answered and also provides the basis for the validity and reliability of the research.

## 4.1 Research Philosophy and Design

### 4.1.1 Research Philosophy

What should be examined, how it should be done, and how the findings should be understood are all dictated by philosophical assumptions/paradigms (Bryman, 2008). In a nutshell, these are the researcher's overall views about the world (Creswell, 2009). They claim that the researcher's assumptions about how an investigation should be conducted (methodology), as well as his or her definition of truth or reality (ontology) and how the investigator comes to know that truth or reality ( epistemology) are contained in a paradigm (Lincoln and Guba, 1985). Due to these philosophical assumptions regarding ontology and epistemology, the methodological choice of a researcher is defined (Collis and Hussey, 2003).

With regard to "reality's nature and researchers' assumptions" and "commitment adhered to a certain perspective," we are talking about "ontology" (Saunders et. al., 2007). Because of the ontological presupposition, the researcher must answer this question: What is reality? (Creswell, 1994). It is the study of "the presence and interaction between humans, society, and all things in general" that constitutes ontology (Eriksson and Kovalainen, 2008). Ontology has two polarizing views: objectivism and constructionism or subjectivism (Grey, 2014). There is a single objective reality to every study occurrence or scenario regardless of the researcher's viewpoint or belief, according to an objectivist ontology. Carson et al. (2001) believe that social reality has an existence independent of social actors, hence the world is external (Hudson and Ozanne, 1988). As a result, it is possible to talk about social entities in the same way physical scientists talk about physical events (Johnson and Onwuegbuzie, 2007). In this school of thinking, people, who are products of the external reality to which they are exposed, only act as responsive mechanisms with minimal engagement as social reality investigators (Morgan and Smircich, 1980).

There is no external universe in which truth and meaning may be found; rather, they are formed by a subject's interactions with it (constructivism) or imposed on it (subjectivism) (Grey, 2014) When it comes to subjectivism or constructivism, they reject the idea that social reality is a product of the human mind (Morgan and Smircich, 1980). As investigators, human beings are supposed to be able to make sense of the events and phenomena that surround them, and to be able to modify the world in accordance with their own views and experiences (Gill et.el., 2010). Because of this, the two extremes of these viewpoints on reality and human beings allow for differing ontological assumptions. Some examples of ontology are: concrete structure; concrete process; contextual field of knowledge; realm of symbolic discourse; social construction; and projection of human structure (Collis and Hussey, 2003), among others.

It is the study of knowledge and what we recognize as true knowledge that is the focus of epistemology (Collis and Hussey, 2003). Epistemological issues are concerned with what kinds of knowledge are considered acceptable in a subject (Bryman,2004). Positive (realism) epistemology and phenomenological (or normative, interpretative) epistemology are two fundamentally separate but opposing ideas in epistemological endeavors (Bryman, 2004).

As a research paradigm, positivism is concerned with finding practical solutions, identifying generalizations, and uncovering exact causal linkages using statistical analysis (Kim, 2003). To positivists, the social world is something that can be seen from the outside, and its qualities can only be assessed through methods based on objective measurement. Because there is only one reality, it is possible to represent and quantify this reality using the variables (Onwuegbuzie, 2002). Because of this, the researcher should concentrate on the facts, find causation between variables, establish and test hypotheses (deductive methodology), operationalize ideas so that they can be quantified, and use quantitative techniques (Easterby-Smith et.al. , 2002).

To comprehend social reality, phenomenologists believe that it must be based on people's own experiences of that reality (Grey, 2014). Understanding what is occurring and developing hypotheses and models from facts (the inductive approach) will be the emphasis of this study (Easterby-Smith et.al. ,2002). Interaction between researchers and the subject matter they are studying is a primary goal in this scenario (Collis and Hussey, 2003).

An ontological assumption is coupled with an epistemological one in this study's epistemological position. In the research, it is acknowledged that knowledge is a fabrication based on the reality that human beings encounter and live (Johnson et. el., 2007). In reality, knowledge is obtained through exploring the nature of links between phenomena as well as comprehending the function of human people in the social world (Morgan and Smircich, 1980). Positivists, then, believe that cause-and-effect linkages may be used to build knowledge.

It is assumed in this study that there are certain external factors influencing bank performance. In order to determine the nature of the connection between industry concentration and bank performance, this study primarily looks at the relationship between those two variables. Furthermore, the phenomenologist's beliefs on the need of searching for meanings via a variety of perspectives on a phenomena are pertinent. Rather than just evaluating hypotheses, the research also aims to explain the "why" behind the causal link and provide suggestions for how it may be strengthened in future iterations. It is primarily aimed at developing meaning from the proven causal link via an in-depth investigation of the viewpoints of bank specialists and regulatory staff members.

The research philosophy used in this research is positivism in nature. Hypotheses that can be tested and theories that can be compared to the accepted knowledge of the world we live in may be generated by positivism, which provides research questions. The study investigated the relationship between COVID-19 Twitter hashtags and the daily increase in COVD 19 cases and was basically guided by the following research questions:

(i) What is the relationship between the COVID-19 Twitter hashtag and the daily increase in COVID-19 cases?

(ii) What sort of relationship exists between the COVID-19 Twitter hashtag and the daily increase in COVID-19 cases?

Thus, based on the research questions at hand, this study adopted a quantitative research method.

### 4.1.2 Research Approach and Strategy

This study followed a deductive approach process. According to Dawson (2013), researchers that draw from deductive technique rely heavily on existing and substantive prior knowledge to conceptualize specific situations. This research examined the relationship that exists between Daily Increases of COVID-19 and Twitter Hashtags. Data was collected and used to validate the existing theory on the relationship that Twitter Hasthtags has with Daily Inreases of COVID-19. The approach taken in this study is opposite to the 'bottom-up' inductive approach where theory develops from data as data is being collected or as data is being analyzed (Saunders *et al.*, 2000).

Founded on the research questions and the problem at hand, this study adopted a quantitative research method or strategy. The research questions for this study sought to investigate if there exists a relationshp betweem Daily Increases of COVID-19 and Twitter Hashtags.. Therefore, all the variables of interest used in this study are all quantifiable and the causal relationship between these variables can statistically be determined. Though the variables of interest in this study can be analysed qualitatively but in this study, it may not be most appropriate to analyse the topic of study. This is because qualitative method is best suited for data sets that are based on meaning, expressed through words, collection results in non-standardized data requiring classification into categories and analysis conducted through the use of conceptualization (Saunders *et al.*, 2000).

### 4.1.3 Conceptual Framework

Analytical tools such as conceptual frameworks may be used in a variety of ways, depending on the situation. It may be used in a variety of contexts when an overall picture is required, such as in the workplace. Conceptual distinctions are made and thoughts are organized with the use of this tool. Furthermore, a conceptual model is a tool used in research to help the person conducting the research achieve understanding and consciousness of the condition under study and be able to communicate it. It helps illustrate the causal relationships between the explanatory variable (s) and the explained variable. The study is based on a bivariate relationship between COVID-19 Twitter Hashtags, which is the explanatory variable, while the number of COVID-19 cases is regarded as the explained variable. Since the pandemic is a deadly virus that impacts the world on a daily basis, I focused on the daily cycle it has.

Since the study is deductive in nature, the following are hypotheses arising from the study:

***The Main Hypothesis****:*

$H_0$: There is no relationship between COVID-19 Twitter hashtags and the daily increases of COVID-19 cases

$H_1$: There is a relationship between COVID-19 Twitter hashtags and the daily increases of COVID-19 cases

***The Specific Hypothesis***

$H_0$: There is no correlation between COVID-19 Twitter hashtags and the daily increases of COVID-19 cases

$H_1$: There is a correlation between COVID-19 Twitter hashtags and the daily increases of COVID-19 cases

### 4.1.4 Research Design

A research design is the 'procedures for collecting, analyzing, interpreting and reporting data in research studies' (Creswell & Plano, 2007). Connecting conceptual research questions with relevant and feasible empirical research is the overarching aim. Thus, the study design determines how to gather and analyze the necessary data and how to answer the research question using these methodologies and data collection and analysis procedures (Grey, 2014). There are three types of study designs: exploratory, descriptive, and explanatory, according to Robson (2002). His categorization system is based on the study area's aim, since each design serves a distinct end-goal. The goal of descriptive research is to paint a picture of a scenario, person, or event or demonstrate how things are connected and occur naturally (Blumberg, Cooper and Schindler, 2005). It's true that descriptive studies can't explain why an event happened, but they are ideal for fresh or undiscovered areas of inquiry (Punch, 2005). Explanatory or exploratory research strategies should be used when there is a lot of descriptive data available.

It is necessary to do exploratory research when there is a lack of information regarding the phenomena or the issue that has to be solved (Saunders et al., 2007). It does not try to offer definitive solutions to the research questions, but rather investigates the study subject at various degrees of detail.. Therefore, its focus is on fresh issues that have received little or no attention in the past (Brown, 2006). Even in the worst-case scenario, exploratory research establishes the first study design, sampling strategy, and data gathering strategy for more definitive research (Singh, 2007).

An explanatory research, on the other hand, aims to explain and account for the descriptive facts. Explanatory studies, on the other hand, ask 'why' and 'how' inquiries rather than 'what' queries (Grey, 2014). It relies on exploratory and descriptive research and identifies the true causes of a phenomena. An explanation or forecast is supported or refuted by information gathered via explanatory research, which searches for the underlying causes and conditions. Discovering and reporting on correlations between various components of the phenomena under investigation is its primary purpose.

The study used time series data because the study investigated the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases over a period of eleven (12) months. That is, from March 2020 to February 2021. As earlier stated, since the research used time-series data and following the purpose of the study, this study pursued an explanatory study inquiry. In this regard, time-series data makes it possible to analyze or predict changes and new developments over a stipulated period of time. (Saunders et al., 2003), argues that the main strength of longitudinal research is its ability to study change and development over time. Further, notwithstanding the time constraints, time-series data analysis was made possible on account of the availability of data on COVID-19 Twitter hashtags and the daily increase in COVID-19 cases through the following sources: Worldometer and Github websites, bulletins, and other publications.

## 4.2 Data Frequency, Data Choice, and Collection

### 4.2.1 Data Frequency and Sample Size

A frequency distribution is an overview of all different values of a variable and the number of times they occur. To put it another way, a frequency distribution shows you how often a certain value occurs. When summarizing categorical data, frequency distributions are most often utilized.

The researcher relied on secondary data on COVID-19 Twitter hashtags and the daily increase in COVID-19 cases. This was readily available and extracted from Worldometer and Github websites for the period of 12 months, that is, from 2020 March to 2021 February. It suffices to state that this period guaranteed a sufficient sample size to be able to run a regression model, correlation analysis, and other statistical tests. On the other hand, this research focused on data that was generated in March 2020, just after COVID-19 was declared a pandemic.

### 4.2.2 Type of Data

Two basic categories are used to arrange data: quantitative and qualitative. Most qualitative data is non-numerical, descriptive or nominal in character. This indicates that the data obtained is in the form of phrases and words. Sometimes such data captures subjective impressions or sentiments about a topic. The 'how' and 'why' of a program are the focus of qualitative techniques, which tend to employ unstructured methods of data collecting. When it comes to qualitative issues, there is no right or wrong answer. Focus groups, group discussions, and interviews all fall under the category of qualitative approaches. Qualitative methods may help researchers better understand the program's intended and unanticipated impacts. However, the implementation costs money and takes time. It is also important to note that the results cannot be extrapolated to individuals outside of the program.

It is crucial to understand the processes that led to observed outcomes and to analyse changes in people's views of their well-being via the use of qualitative data gathering methodologies. The quality of quantitative assessments may also be improved by using qualitative approaches to assist establish evaluation hypotheses, enhance the design of survey questionnaires, and broaden or clarify quantitative evaluation results. Following are the characteristics of these methods:

- They're more flexible and less regimented in their methods (i.e., researchers may change the data collection strategy by adding, refining, or dropping techniques or informants);
- interactive interviews are increasingly common; To ensure that a problem is being followed up on, ideas are being clarified, or data is being checked for accuracy, respondents may be contacted more than one time;
- In order to boost the credibility of their conclusions, they use triangulation (i.e., researchers rely on multiple data collection methods to check the authenticity of their results);
- A typical case study provides a single piece of data that may be compared to other studies on the same topic in order to identify commonalities and trends.

It is technically possible to calculate quantitative data because quantitative data is numerical in nature. Different scales are used in quantitative data measurement, and they may be divided into four categories: nominal scale, ordinal scale, interval scale, and ratio scale. This kind of data often (but not always) contains measurements of some kind of object. Programs that use quantitative techniques are those that address the "what" of the program. They have a methodical and uniform approach, and they apply tools such as surveys and inquiries to gather information. Quantitative techniques have the benefit of being less expensive to adopt, being standardized so that comparisons can be made with ease, and being able to assess the extent of the impact in the majority of cases. Quantitative techniques, on the other hand, are restricted in their ability to investigate and explain similarities and differences that are unforeseen to the investigator. However, it is important to note that for peer-based programs, quantitative data collection approaches are frequently difficult to implement for agencies due to a lack of necessary resources to ensure rigorous implementation of surveys and a high incidence of low participation and loss to follow-up rates, which are both frequently encountered factors. A combination of random sampling and structured data collecting devices are used to categorize a wide range of events into preset response categories in quantitative data gathering approaches. These researchers provide findings that are simple to describe, compare, and extrapolate. To pick participants for the study, the researcher will use probability sampling if the goal is to generalize the findings from the research participants to a wider population at large.

Typical quantitative data gathering strategies include:

- Experiments/clinical trials are being conducted.
- Paying attention to and documenting clearly defined occurrences (e.g., counting the number of patients waiting in emergency at specified times of the day).
- Management information systems (MIS) are used to gather useful information.
- Using closed-ended questions in surveys that are administered (e.g., face-to face and telephone interviews, questionnaires etc.).
- Interviews in quantitative research (survey research) are more structured than interviews in qualitative research (case study research). A structured interview is one in which the researcher asks just a predetermined set of questions and nothing else. One particular benefit of face -to-face interviews is that they allow the researcher to create a relationship with prospective participants, which in turn increases the likelihood of their cooperating.
- Paper-and-pencil questionnaires may be sent to a large number of individuals, saving both time and money for the researcher. It has been shown that people are more honest when answering to surveys, and this is especially true when it comes to sensitive matters, since their comments are anonymous.

Designing a research study using mixed methodologies means include both qualitative and quantitative research data, procedures, and methods in a single research framework. Mixed methods approaches may refer to a variety of things, including the use of a variety of various kinds of techniques in research or at different stages within a study, as well as a combination of qualitative and quantitative approaches. Mixed methods are a collection of multiple ways that are used in conjunction to maximize on strengths while simultaneously reducing shortcomings that result from the use of a single research design. It is possible that using this strategy to collect and assess data will help to improve the validity and reliability of the study.

Some of the most popular domains in which mixed-method techniques are used are as follows:

- establishing, creating, developing, and scaling up interventions;

- evaluating interventions;
- establishing, creating, developing, and scaling up interventions;
- Findings that are consistent with one another, data triangulation, or convergence.

Some of the difficulties associated with using a mixed methods approach are as follows:

- Identifying and defining compatible qualitative and quantitative research topics.
- Data collecting and analysis that is time-consuming; and
- Determining which research methodologies to use and how to combine them.

The use of mixed methods can be beneficial in drawing attention to complex research problems such as health disparities, but they can also be transformative when it comes to addressing issues affecting vulnerable or marginalized populations or when it comes to research that involves community participation. A mixed-methods approach to research and evaluation may be used to generate innovative alternatives to standard or single design approaches to research and evaluation.

There is a plethora of classification schemes available. A popular categorization is based on the person who gathered the information.

### 4.2.3 Data Choice

This study relied on secondary data because it is the best data source for longitudinal studies as it cuts down on the cost and time of collecting primary data over a long period of time, which can run into decades. Further, secondary data in this regard was also readily available and made it easy to analyze the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases. Lastly, secondary data is permanent, and this makes it easy for others to validate this study. However, secondary data may also have its own limitations. These include the data may have been collected for a purpose which does not match this research, technically making the data not match the model and purpose. Further, access to data may be difficult or costly.

Finally, the initial purpose may affect how data is presented and aggregations may result in data not being suitable for other researchers. This study relied on secondary data because it is the best data source for longitudinal studies as it cuts down on the cost and time of collecting primary data over a long period of, which can run into decades. Tha dataFurther, secondary data in this regard was also readily available and made it easy to analyze the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases. Lastly, secondary data is permanent, and this makes it easy for others to validate this study.

### 4.2.4 Data Collection

It is the act of acquiring and evaluating information on variables of interest in a systematic manner that allows one to answer stated research questions, test hypotheses, and assess results. The data gathering component of research is common to all disciplines of study, including the physical and social sciences, the humanities, business, and other realms of endeavor. While the techniques used by different disciplines differ, the focus on ensuring accurate and honest collecting is the same across the board. The ultimate purpose of any data gathering is to get high-quality evidence that can be used to conduct in-depth data analysis and ultimately to provide a persuasive and credible response to the issues that have been given to researchers. Exact data collecting is critical to the integrity of research, regardless of the topic of study or choice for defining data (quantitative vs qualitative). In order to limit the possibility of mistakes happening, it is important to pick acceptable data gathering instruments (whether current, modified, or newly invented) and to provide clear instructions for their right usage.

The collecting of data is one of the most crucial steps in the process of performing a study. Having the greatest study design in the world will not ensure that you are able to finish your project if you are unable to gather the necessary data. Data collecting is a difficult process that takes meticulous preparation, long hours, patience, tenacity, and other qualities in order to be completed properly. The process of data collecting begins with the determination of the kind of information that is needed, followed by the selection of a sample from a certain population. A specific instrument must then be used by the researcher in order to gather data from the sample that has been chosen.

A main piece of data is information that has been gathered through first-hand observation or experience. Primary data has not yet been made public, yet it is more trustworthy, genuine, and impartial than secondary data. The validity of primary data is better than that of secondary data since primary data has not been edited or altered by human beings. In order to conduct statistical surveys, it is required to get information from primary sources and to conduct analyses based on primary information. For example, a country's statistical data on the female population cannot be dependent only on information obtained from newspapers, magazines, and other written media. A research project may be carried out without the use of secondary data, however a research project that relies only on secondary data is the least credible and most likely to include biases, since secondary data has already been modified by humans. In addition to being outdated, they often provide limited information and have the potential to be misleading and prejudiced.

Secondary data is data that has been acquired from a source that has previously been published in some way or another. The review of literature in any study is based on secondary data, which is collected from other sources. Somebody else has obtained it for some other reason than yours (but being utilized by the investigator for another purpose). Examples include the use of Census data to examine the relationship between education and occupational choice and earnings. Censuses, organizational records, and data obtained via qualitative techniques or qualitative research are all examples of secondary data sources in the social sciences that are often used. It is necessary to use secondary data because it is difficult to conduct a fresh survey that will fully reflect prior change and/or advancements.

The study objectives and research questions were explored using quantitative empirical techniques appropriate for time series data. In line with the strategy selected above, this study used a longitudinal case study approach and used secondary data. Secondary data made it possible to undertake longitudinal studies over a short period of time. In this case, data on COVID-19 Twitter hashtags and the daily increases of COVID-19 cases were made readily available on Github (Emily Chen, 2020) and Worldometer (*Worldometers*, n.d.) websites respectively. As such, data collection utilized a purposive sampling, this being a case study research (Saunders et al., 2003).

## 4.3 Analysis Methods

### 4.3.1 Statstical Analysis

Being guided by proceedings from previous chapters, such as the general objective and specific objectives of the study (chapter two), the research hypothesis (chapter three), as well as the statement of the problem and the research methodology adopted, this chapter of the study thus serves as a platform via which the research findings, data analysis, and discussion of the study are covered. The statistical software E-Views 10 was used for statistical analysis of the gathered secondary data. A series of diagnostic quantitative tests or deductive tests were carried out so as to address the aforementioned targets on which this study hinges. To further support the reports in the analysis, some of the output printouts emanating from E-Views are included in the appendix of this entire research report.

### 4.3.2 Descriptive Statistics

Statistical methods such as descriptive statistics are used to organize and summarize data by explaining the connection between variables in a sample or population. When doing research, the calculation of descriptive statistics is an essential initial step that should always be completed before performing inferential statistical comparisons across groups. Types of variables (nominal, ordinal, interval and ratio) as well as measures of frequency, central tendency, dispersion/variation, and position are all included in descriptive statistics, as is the use of a t-test. Because descriptive statistics compress data into a more digestible summary, they allow health-care decision-makers to examine particular populations in a more manageable manner. Table 1: Descriptive statistics

Initial data analysis is incomplete without descriptive statistics, which serve as the basis for comparing variables using inferential statistical tests (Laerd Statistics, 2018). As a consequence, it is critical to report the most relevant descriptive statistics using a systematic manner as part of good research practice in order to limit the chance of providing misleading data (Huebner, 2016). Since the results of statistical analysis are fundamental in influencing the future of public health and health sciences, the appropriate use of descriptive statistics allow health-care administrators and providers to more effectively weigh the impact of health policies and programs (Peace & Hsu, 2018).

Prior to the commencement of analysis, the descriptive statistics of the data were produced. The descriptive statistics informed the research on the mean, median, skewness, and kurtosis values of our variables and also guided the research on whether the variables, that is, COVID-19 Twitter hashtags and the daily increases in COVID-19 cases, are symmetric or not. Further, the descriptive statistics pointed out the characteristics of the residual of the two research variables in the research.

### 4.3.3 Correlational Analysis

There is a correlation between variables if one variable changes in value, the other changes in the same direction. The value of one variable depends, to some degree, on the value of another variable. Correlation measures the strength of that association. Understanding that link is helpful because we can use one variable's value to anticipate the other's value. There are several correlations, such as height and weight. When a result, weight tends to rise as a person's height rises. When a person is taller than the norm, we may infer that he or she is heavier than the average.Correlation coefficients in statistics are a way of quantifying the intensity and direction of this propensity to change together. There are a variety of correlations that may be applied to a variety of data sets. In this study, Pearson's correlation was used to measure the strength of the association between daily, weekly, and monthly data of COVID-19 cases and Twitter hashtags.

### 4.3.4 Regression Equation

To assess the connection between a dependent variable and one or more independent variables, regression analysis is performed. Analysis of the connection between one or more independent variables and one or more dependent variables is known as regression analysis in statistical modeling, and it may be used to estimate the link between the two variables. It is possible to identify a line (or a more sophisticated linear combination) that best fits the data according to a given mathematical criteria using linear regression analysis. When using the least squares approach for example, the unique line (or hyperplane) that minimizes the sum of squared discrepancies between the genuine data and that line is computed by the algorithm (or hyperplane).

This permits the researcher to estimate the dependent variable's conditional expectation (or population average value) when the independent variables take on a specified set of values for precise mathematical reasons (see linear regression). To estimate alternative parameters (e.g., quantile regression or Necessary Condition Analysis) or to estimate the conditional expectation over a wider array of non-linear models, less frequent regressions employ somewhat different approaches (e.g., nonparametric regression).

For the most part, regression analysis is employed for two main goals. It is important to note that regression analysis has a lot in common with machine learning when it comes to predicting and forecasting future events. It is possible to derive causal correlations between variables using regression analysis in certain cases, as will be discussed below. The links between a dependent variable and a group of independent variables in a fixed dataset can only be discovered using regressions. To utilize regressions for prediction or to infer causal linkages, a researcher must carefully explain why previous correlations have predictive value in a new context or why a relationship between two variables has a causal meaning. When researchers are trying to determine causal linkages based on observational data, the latter is very crucial (Dennis & Sanford, 1982).

After analyzing the descriptive statistics, the regression equation was estimated in order to establish co-integration of the research variables. The estimated regression model was subjected to the following diagnostic tests; serial correlation, heteroscedasticity, and normality tests. These tests are important as they help in determining how good the regression equation is. Thus, as earlier stated, the study adopted a simple or bivariate regression model. That is, the model tries to predict the extent to which the independent variable relates to the dependent variable. Henceforth, the proposed regression model of the study is a bivariate one which was as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where:

$Y =$ the daily increases of COVID-19 cases

$X =$ COVID-19 Twitter Hashtags

$\beta_0 =$ is a constant which represents the value of daily increases of COVID-19 cases when the independent variable under consideration is zero.

$\beta_1 =$ Represents the coefficient of $X$

$\varepsilon =$ Represents the error term

Therefore, according to Earl (2009), a bivariate analysis is one of the simplest forms of quantitative (statistical) analysis. Two variables (commonly labeled as X,Y) are analyzed in order to discover the empirical connection between them in this kind of study (Earl, 2009). The independent variable in this research is Twitter Hashtags, while the dependent variable is COVID-19 Daily Increases.

Simple hypotheses of connection may be tested using bivariate analysis. For example, a bivariate analysis may assist assess how much simpler it is to forecast one variable's value when we know its value for another (perhaps a dependent variable) (see also correlation and simple linear regression).

Two variables are used to make comparisons between bivariate and monovariate analyses (Earl, 2009). Both descriptive and inferential bivariate analysis may be performed. In this case, it's a look at the connection between the two variables (Earl, 2009). Bivariate analysis is a specific instance of multivariate analysis that has just two variables (where multiple relations between multiple variables are examined simultaneously)

# 5 Results and Discussion

## 5.1 Research validity and Reliability

As defined earlier validity is the extent to which an instrument measures what it purports to measure. Attempting to convey the reality of study results, as defined by Zohrabi, is what validity is all about (2013). Is it true that an IQ test measures intellect, for example? The validity of a claim is determined by both theoretical and empirical evidence. It is in the process of translating or representing a notion of a construct into an operational measure that theoretical assessment occurs. Typically, this is done by a panel of experts, who are either judges or university lecturers, who score the acceptability of each item and evaluate its fit in relation to the construct's description. When validity is determined by quantitative analysis utilizing statistical methods, this is referred to as empirical evaluation.

Type's validity in educational research is comprised of the following factors;
*Construct validity* – Specifically, it refers to the process through which a constructed notion, idea, or action gets translated or turned into a functional and operational reality (Trochim, 2006).
*Face validity* – The situation occurs when, "on its face," an indication seems to be an appropriate measure of the underlying concept.
*Content validity* – This is an assessment on how well a set of scale of items matches with the relevant content domain of the construct that it is trying to measure.
*Convergent and Discriminant validity* – In this context, "convergent" refers to the degree to which a measure corresponds directly to the concept that it purports to measure. In the context of discriminant validity, the extent to which a measure does not measure or discriminates against the concept it is not meant to measure.
*Criterion-related validity* - When two or more independent variables are compared to see how well they match up, the result is called the correlation coefficient (Mohajan, 2017).
Conceptualization refers to how an idea or activity that is a construct is brought to life in the actual world by means of translation or transformation (Trochim, 2006)
Measurement stability is described as the capacity to get findings across a wide range of situations (Nunnally, 1978). Reproducibility is essentially the capacity of research results to

be replicated. This means that data is considered dependable when researchers do the same experiment for the second time and get the same findings. Drost (2011) states that two types of mistakes might impair the accuracy of data collected by research instruments: random error and systematic error.

If an external element has an impact on some observations but not others, it is ascribed to random error. When it comes to notions like self-esteem, happiness, and contentment, those who are in a better mood could answer more favorably than those who are in a negative mood. Noise in measurements is sometimes referred to as random error, which is why it's often overlooked. Errors of this kind may be attributed to variables that have a disproportionate effect on the whole sample. The sample's findings will be more accurate if systematic errors are eliminated from the process. The best method for determining reliability is to calculate the reliability coefficients for tests, items, and raters (Rosnow and Rosenthal, 1991).

The following are the type’s reliability;

*Test-retest reliability* – It is a measure of consistency between measurements of the same construct administered to the same sample at two different points in time (Drost, 2011).

*Split-half reliability* – as a measure of consistency between two halves of a construct measure (Heale & Twycross, 2015).

*Inter-rater reliability* – It is also called inter-observer rating or an agreement. It involves rating of observations using a specific measure but by different judges.

*Internal consistency reliability* – It is a measure of consistency between different items of the same construct. It measures the consistency within the instrument and questions on how well a set of items measures a particular characteristic of the test.

Following study methodologies such as (Gangal & Gupta, 2013; Garba & Abdullahi, 2013) as well as (Muhammad et al., 2015) which have been done in this vein before, this study is focused on global data which none of the studies done focused on. Therefore, this study did not reinvent the new cycle. However, the methodology applied has been tested and proven in more than one study, as stated above. Thus, the choice of this approach is most important based on its analytical ability to resolve the research questions.

To put it another way, dependability is the degree to which outcomes are consistent across time (Joppe, 2020). Similarly, (Orodho, 2009) states that reliability is concerned with the extent to which a measuring procedure produces similar results when repeated several times. (Cronbach, 1951) recommends Cronbach's alpha of 0.7 to establish reliability. Thus, in this study, the process of maintaining validity and reliability was considered at the stage of selecting the tools of analysis and in the use of the tools of analysis itself. Also, this research study used secondary data which was readily available, which can help in ensuring the reliability of this study. Additionally, the data for this study was sourced from well-respected sources such as the Worldometer databank.

### 5.1.1 Descriptive Statistics

A critical initial step in doing research is the calculation of descriptive statistics, which should always precede any inferential statistical comparisons. Different kinds of variables (nominal, ordinal, interval and ratio) and measures of frequency, central tendency, dispersion/variation, and location are included in the descriptive statistics section. A simplified overview of the data is provided by descriptive statistics, particularly in this study, they are very important because they help key layers or stakeholders to identify or assess specific ways in which daily increases in COVID-19 cases can be managed. Table1 below shows the results of the descriptive statistics of the data under study.

The mean is a measure of central tendency, particularly in this study, the daily increase of COVID-19 cases averaged 1242.01 per day while that of the Twitter hashtags was 357649.50 per day. To be precise, a standard deviation is a measure of the average distance between the values of the data in the set and the mean. Low standard deviation implies that the data points are often around the mean, whereas a high standard deviation suggests that there is a wide range of values covered by the data points. According to the findings in Table 1, both variables have substantial standard deviations, indicating that the data points span a wide range of values.

The maximum and minimum values of all the variables shown rule out the possibility of outliers in the data used. The kurtosis of the variable daily increases of COVID-19 cases and Twitter hashtags are bigger than three and less than three, indicating that their probability distribution functions are short-tailed and long-tailed respectively. By use of the rule of thumb, the skewness values for the daily increases of COVID-19 cases and Twitter hashtags lie between-1 and 1, implying that the datasets are moderately skewed.

***Table 2: Descriptive Statistics of the Variables COVID-19 TWIT#TAG***

| | **COVID-19** | **TWIT#TAG** |
|---|---|---|
| Mean | 1242.01 | 357649.50 |
| Median | 2422.00 | 351290.50 |
| Maximum | 390286.0 | 613154.0 |
| Minimum | -296949.0 | 35694.00 |
| Std. Dev. | 48796.10 | 101320.4 |
| Skewness | 0.63 | -0.06 |
| Kurtosis | 19.05 | 2.84 |
| Jarque-Bera | 3627.97 | 0.58 |
| Probability | 0.00 | 0.75 |
| **Observations** | 336 | 336 |

*Source: Author's Compilation, 2021*

### 5.1.2 Correlation Analysis

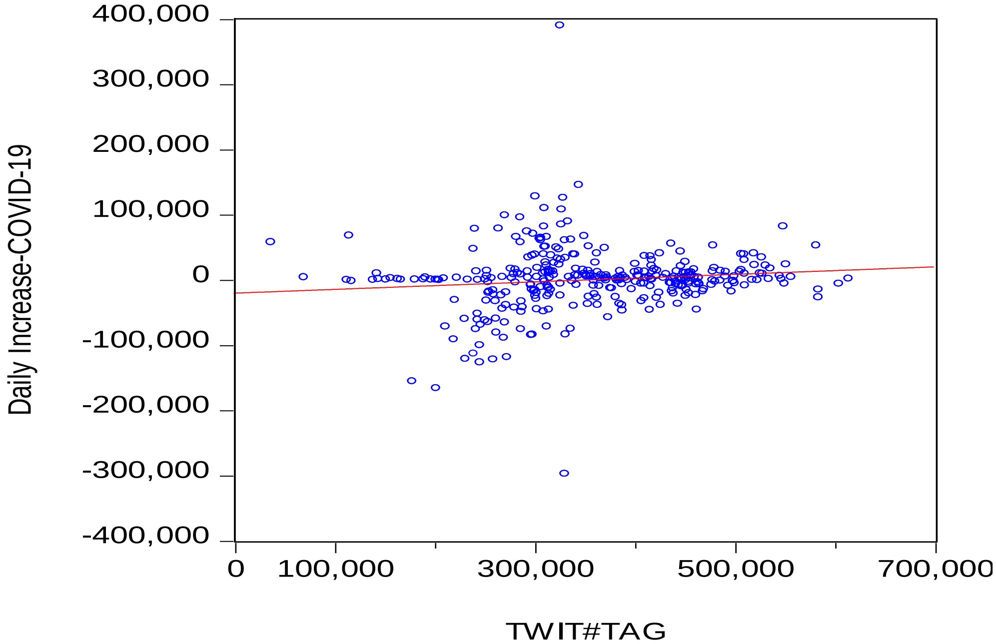
Table 2 below shows the Pearson's pairwise correlation between daily increases in COVID-19 and Twitter hashtags. Correlation coefficients are a statistical measure of the strength of the link between two variables' relative movements. The values range from -1 to 1. Correlation coefficients less than one imply an ideal negative link, whilst correlation coefficients greater than one suggest an ideal positive relationship. A correlation coefficient of zero shows that there is no linear relationship between the movements of the two variables. A correlation coefficient of  shows that there is no linear relationship between the movements of the two variables. Correlation coefficients are a quantitative measure of the direction and intensity of this propensity to fluctuate in tandem in statistics. Correlations come in a variety of forms and may be used to a variety of different sorts of data. However, Pearson's correlation coefficient was chosen in this investigation because to its emphasis on the strength of the association between two continuous variables. Additionally, correlation analysis was performed on daily, weekly, and monthly data for the two variables studied in this research, and the findings are shown in tables 2–4 below.

***Table 3: Pearson's Correlation Matrix Daly Data***

| | COVID-19 | TWIT#TAG |
|---|---|---|
| **COVID-19** | 1.00 | |
| **TWIT#TAG** | 0.12 | 1.00 |

*Source: Author's Compilation, 2021*

***Figure 1: Pearson's Correlation Scatter Plot for Daly Data***




According to table 2 above, the correlation between the explained and the explanatory variables is 0.12 and the direction of the relationship is positive. This implies that as one variable, such as the daily increase of COVID-19, increases, the value of the other variable, such as Twitter hashtags, also increases by approximately 12%. Furthermore, the scatterplot in figure 2 above confirms an upwards positive scatterplot.

***Table 4: Pearson's Correlation Matrix Weekly Data***

| | COVD-19 | TWIT#TAG |
|---|---|---|
| **COVID-19** | 1.00 | |
| **TWIT#TAG** | 0.14 | 1.00 |

*Source: Author's Compilation, 2021*

According to table 3 above, which shows the correlation analysis results for weekly data for the two variables under study, the correlation between the weekly increases of COVID-19 and the Twitter hashtag variable is 0.14, and the direction of the relationship is positive. This implies that as one variable, such as the weekly increase in COVID-19, increases, the value of the other variable, such as Twitter hashtags, also increases by approximately 14%. Furthermore, the scatterplot in figure 3 below confirms an upwards positive scatterplot.

***Figure 2: Pearson's Correlation Scatter Plot for Weekly Data***

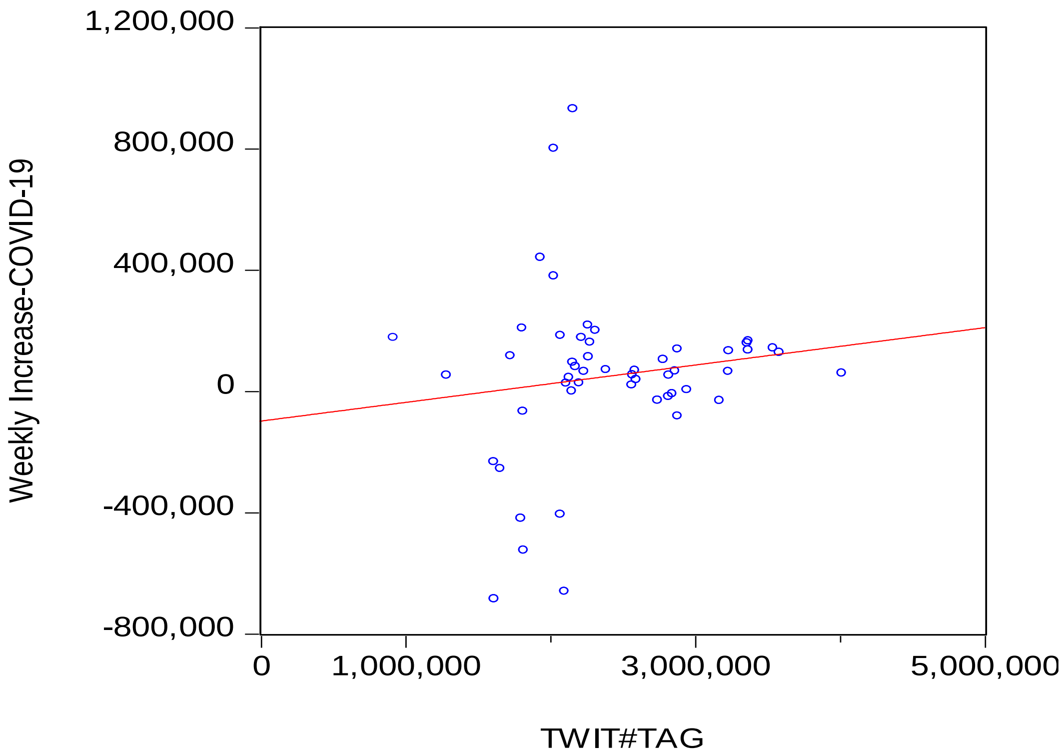


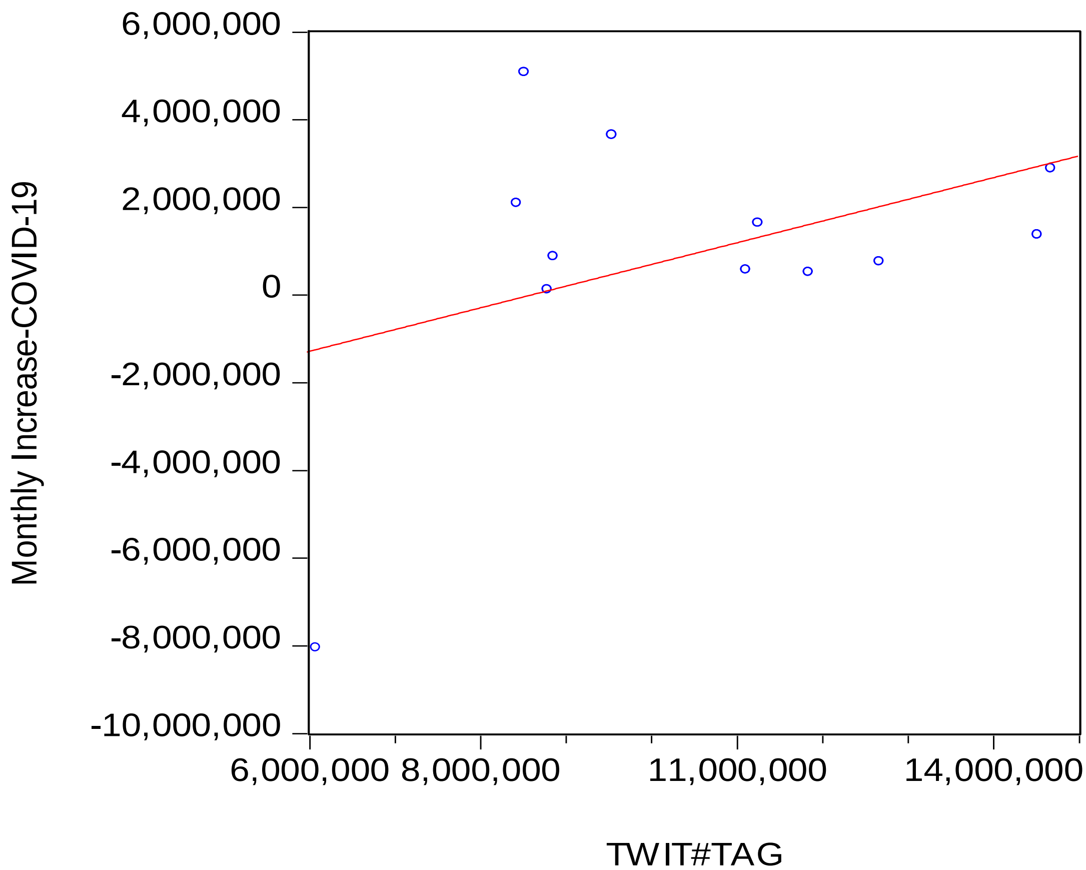

Table 5: Pearson's Correlation Matrix Monthly Data

| | **COVID-19** | **TWIT#TAG** |
|---|---|---|
| **COVID-19** | 1.00 | |
| **TWIT#TAG** | 0.41 | 1.00 |

*Source: Author's Compilation, 2021*

Furthermore, according to table 4 above, it shows the correlation analysis results for the monthly data for the two variables under study. The correlation between the monthly increases of COVID-19 and the Twitter hashtag variable is 0.41, and the direction of the relationship is positive. This implies that as one variable, such as the weekly increase in COVID-19, increases, the value of the other variable, such as Twitter hashtags, also increases by approximately 41%. Furthermore, the scatterplot in figure 4 below confirms an upwards positive scatterplot.

***Figure 3: Pearson's Correlation Scatter Plot for Monthly Data***



## 5.2 Regression Analysis

Regression analysis is a strong statistical technique used in statistical analysis that enables us to investigate the connection between two or more variables of interest. While regression analysis comes in a variety of flavors, they all focus on the effect of one or more independent variables on a dependent variable. In particular, in this study, regression analysis was used to examine the relationship between daily increases in COVID-19 cases and Twitter hashtags. Below is an equation representation of the regression model where Daly increases in COVID-19 cases is a dependent variable and Twitter Hashtags is the independent variable.

$Y = \beta_0 + \beta_1 X + \varepsilon$ where:

$$COVID - 19 = -19366.76 + 0.06TWIT\#TAG \qquad t = (-1.99) \qquad (2.20)$$
$$\text{P}-value = (0.05) \qquad (0.03)$$

According to the equation above, the explanatory variable or independent variable has a positive and statistically significant effect on the daily increase of COVID-19 cases over the period 2020 March to 2021 February. This implies that for every one-unit increase in the Twitter hashtags, daily increases in COVID-19 cases vary by upwards. The p-value for the constants is equally significant at 5% level though negative, implying that the expected value of the dependent variable will be less than 0 when the independent/predictor variable is set to 0. However, a negative value for the constant/intercept in this study is no cause for concern.

***Table 6: Summary of Regression Estimate Results***

Dependent Varable: COVD-19

Method of Estmaton: Least Squares

| Variable | Coefficient | Std. Error | t-Statistic | Prob.* |
|---|---|---|---|---|
| C | -19366.76 | 9724.26 | -1.99 | 0.05 |
| TWIT#TAG | 0.06 | 0.026 | 2.20 | 0.03 |

*Source: Author's compilation from EVews 10.*

***Table 7: Model statistics***

| *Number of Observations* | *F-Statistics* | *Prob.** | *R-Squared* | *Adjusted R-Squared* | *Durbin Watson Statistics (DW)* |
|---|---|---|---|---|---|
| 336 | 4.85 | 0.03 | 0.01 (1%) | 0.01 (1%) | 1.83 |

*Source: Author's compilation from EVews 10.*

In answering the research hypothesis, the model is correctly specified as the probability (0.03) of the F-statistic critical value is less than 5%. This further means that the influence of the explanatory variable on the daily increase of COVID-19 cases is statistically significant at a 5 percent level of significance. This means that we reject the overall model null hypothesis that Twitter hashtags have no impact on the daily increase of COVID-19 cases.

Furthermore, the Durbin Watson statistic value of 1.83 is higher than the R-squared value. Thus, this implies that the adopted model is non-spurious, hence a good fit to explain the relationship between the two variables.

### 5.2.1 Normality Test of Residual

The Jarque-Bera test for normality was done to test whether the error term was normally distributed because this is the condition required for performing subsequent diagnostic tests. The null hypothesis of this test suggests that residues are normally distributed and the alternative hypothesis claims that the residues are not normally distributed. Further interpretation of the results shown in the figure below is done in chapter five of this study. Following the shape of the histogram and the residual graph, it is noted that the residuals of the fitted regression model are normally distributed.

***Figure 4: Jarque – Bera Normality Test Results***

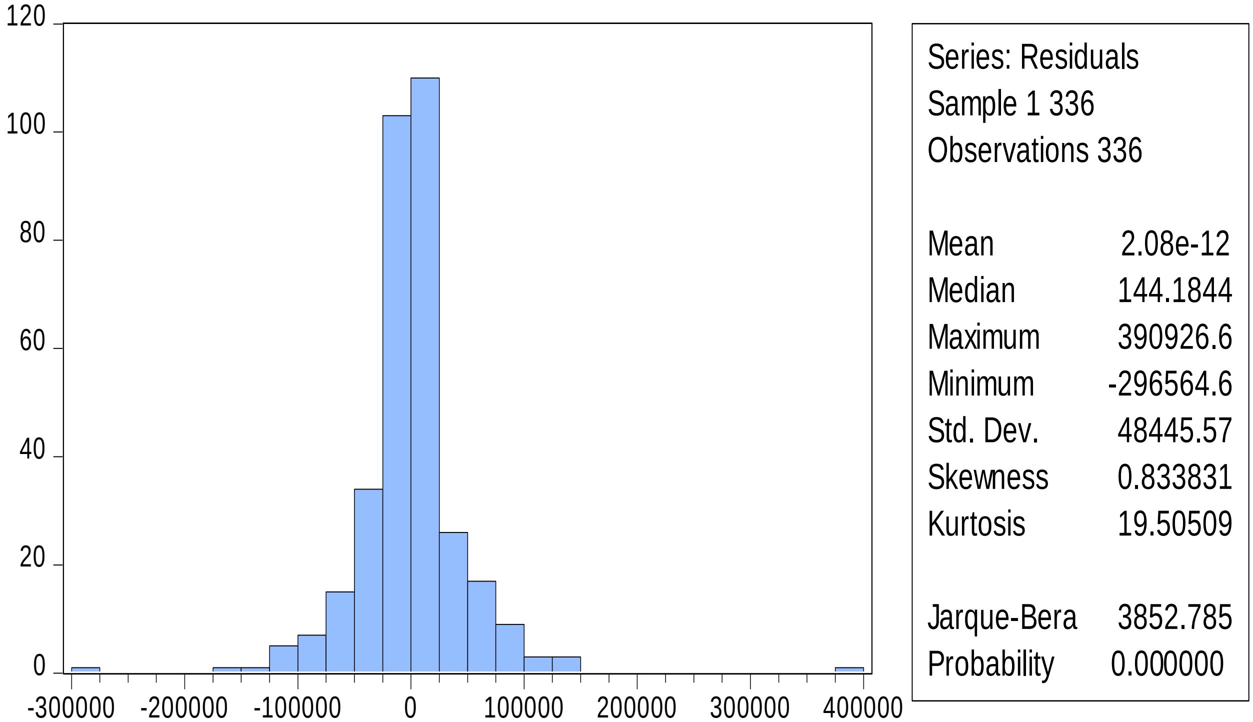
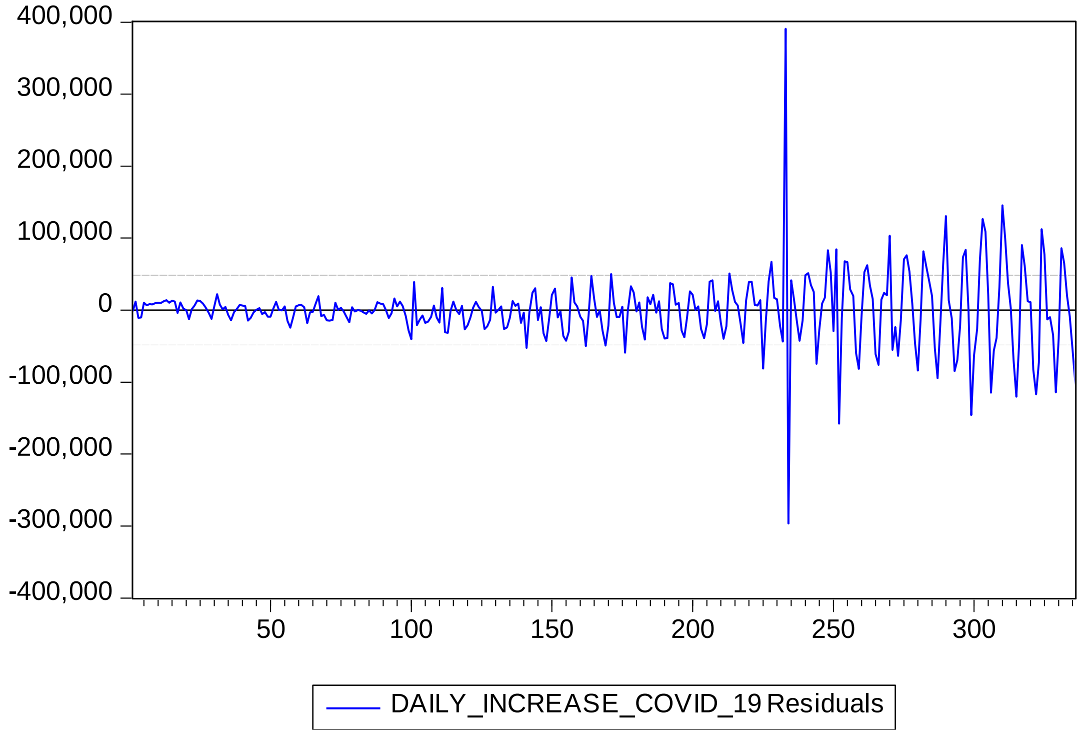



***Figure 5: Regression Residual Graph***



### 5.2.2 Serial Correlation LM Test

Serial correlation is used in statistics to describe the relationship between observations of the same variable over specific periods. If the serial correlation of a variable is 0, then there is no correlation between the observations, and each observation is independent of the previous data. As an alternative, if the serial correlation of a variable is biased toward one, the observations are serially correlated and future observations are influenced by previous values. To put it simply, a variable that is serially connected follows a pattern rather than being random.

Error terms occur when a model is not completely accurate and results in differing results during real-world applications. When error terms from different (usually adjacent) periods (or cross-section observations) are correlated, the error term is serially correlated. During time-series analysis, serial correlation arises when the mistakes associated with a specific period spill over into subsequent periods. When estimating the rise of stock dividends, for example, an overestimation in one year will result in overestimations in subsequent years, as seen in the chart.

The Breusch-Godfrey test was used to determine the Lagrange Multiplier in this research, and the findings are reported in Table 7 below. These results indicate that there is no serial correlation in the residuals of the specified model and this is a desirable outcome. This is evidenced by a p-value of 0.1581 which is more than 5% standard significance level and this implies that we failed to reject the null hypothesis of no serial correlation in our residuals.

***Table 7: Breusch-Godfrey Serial Correlation LM Test***

| | | | |
|---|---|---|---|
| F-statistic | 1.99 | Prob. F (1,333) | 0.1597 |
| Obs*R-squared | 1.99 | Prob. Chi-Square (1) | 0.1581 |

*Source: Author's compilation from EVews 10.*

### 5.2.3 Model Misspecification Test

Model Misspecification is where the model you made with regression analysis is in error. In other words, it doesn't account for everything it should. Models that are misspecified can have biased coefficients and error terms, and tend to have biased parameter estimations.

It is a broad specification test for the linear regression model that is used in statistics. The Ramsey Regression Equation Specification Error Test (RESET) is one such test. The test especially looks for the presence of non-linear combinations of the fitted values that may aid in predicting the response variable. It is deemed mis specified if non-linear combinations of explanatory variables have any power in describing the response variable, according to intuition, and a polynomial or similar non-linear functional form may better resemble the data producing process.

Therefore, the model stability test adopted in this model was the Ramsey RESET test, and below are the results of the model stability test in table 8. The Ramsey RESET test is used to determine whether the model is misspecified or not, with the null hypothesis being that the model has no omitted variables, while the alternate hypothesis says otherwise, or the opposite of the null hypothesis.

As can be seen in Table 8., the P-values of all the models are greater than the 5% level of significance for all the different measures of the daily increase in COVID-19 cases, indicating that the model was not misspecificied. This is a desirable outcome.

***Table 8: Ramsey's RESET Test Results***

Specification : *DALY_INCREASE_COVID_19, TWIT_TAG*
Omitted Variables: Squares of fitted values

| | Value | df | Probability |
|---|---|---|---|
| t-statistic | 0.31 | 333 | 0.7557 |
| F-statistic | 0.1 | (1, 333) | 0.7557 |
| Likelihood ratio | 0.1 | 1 | 0.7545 |

*Source: Author's compilation from EViews 10.*

## 5.3 Pertinent Issues Arising from the Study

According to the findings of the study, crisis response actions in the real world and on the internet are becoming more "simultaneous and interconnected" (Veil et al., 2011), Social media is a profitable opportunity to communicate and disseminate public health knowledge and information directly to members of the general population. (Breland et al., 2017). Social media, on the other hand, may be a strong tool that, if not utilized properly, can be detrimental to public health initiatives, particularly during a public health crisis.

Though social media platforms could also be a source of misinformation especially in times of crisis, however, the study findings play a significant role on the sensitization as well as provide statistics for the prevailing daily increases of the COVID-19 pandemic cases. Therefore, it is recommended that various strategies to enhance the spread of information via Twitter by use of hashtags be looked into. By doing so, it would provide a basis for sieving misinformation from true information in relation to the daily increases of the pandemic cases.

Drawn from other research findings conducted by Veil et al. (2011), it is further noted that social media such as Twitter provides lucrative opportunities to spread and disseminate public health knowledge and information directly to the public. In order to do this, a larger and more aggressive public health presence on social media is required. Instead, health systems should focus their efforts on developing national and worldwide illness detection and surveillance systems, which may be accomplished via the use of social media monitoring.

Furthermore, results emanating from the study revealed that correlation between Daily Increases of COVID-19 and Twitter Hashtags was highest when using monthly data. This result is similar to the findings of Kim (2020) who focused on social network size and incivility. In this study, the high correlation results for monthly data could be alluded to the fact that monthly data has the highest volume of both tweets and increases in the number of COVID-19 cases.

### 5.3.1 Research Implications

Since a holistic approach is necessary to effectively identify the relationship between the daily increases of the COVID-19 cases and the Twitter hashtags, it is important to further carry out an investigation on how our results as well as future results on the propagation of the relationship between the daily increases of the COVID-19 cases and the Twitter hashtags with how other social media platforms such as Facebook relate to the daily increases of the COVID-19 cases. Thus, further, studies are recommended to be conducted by way of adding more other explanatory variables and analysis methods as well other than the ones used in this study.

Due to the fact that the majority of research efforts have been focused to English-language data (Dashtipour et al., 2016), as well as this study, further work is required for multilingual sentiment analysis across social media platforms. Considering longitudinal, multilingual sentiment analysis in addition to contemporaneous study of infectious disease outbreaks on multiple social media platforms such as Facebook, if possible, might be beneficial for future research.

### 5.3.2 Strengths and Limitations

This research, which examined tweets connected to the current COVID-19 epidemic, has a number of advantages and disadvantages, which are discussed below. Considering that the virus has spread over the globe, there were no geographical constraints placed on the tweets that were analyzed in this research. However, since the research only looked at tweets in the English language, it is possible that the conclusions regarding this widespread pandemic would be limited in their applicability. We were also unable to collect COVID-19-related tweets that were made before February 2, 2020, due to the fact that the Twitter standard search API does not enable researchers to access tweets that were posted more than one week ago. As a result, the results may not be applicable outside of that time period.

Furthermore, this research was unable to gather tweets from accounts that had been designated as private. In order to avoid bias, it is possible that the results may not reflect all of the themes addressed by users on Twitter about COVID-19. Due to the fact that only Twitter tweets were examined in this research, our results may not be applicable to posts made on other social media sites.

Furthermore, the results of this research are confined to just individuals who have access to and use Twitter, which limits the generalizability of the findings. Consequently, care is suggested in drawing conclusions about the generalizability of the findings, given that Twitter is not utilized by everyone in the general population.

Additionally, just like many other researches, this study is also subjected to some limitation, such as, this study does not comprehensively capture all aspects related to daily increases of COVID-19 cases as its main focus was mainly on the relationship between the daily increases of the COVID-19 cases and the Twitter hashtags.

# 6 Conclusion

During a crisis, whether natural or man-made, people tend to spend more time on social media than normal. As a crisis unfolds, social media platforms such as Facebook and Twitter become an active source of information because these platforms break the news faster than official news channels and emergency response agencies (Imran et al., 2020). Therefore, the main goal or objective of this study was to identify the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases. The study used daily time-series data for 12 months, that is, from 2020 March to 2021 February. This data was obtained from Worldometer and Github (Lopezbec) websites, bulletins, and other publications. Data were analyzed using the statistical software Eviews 10. Following the study objectives, the following conclusions were drawn:

*Specific Objectives I: To determine the relationship between COVID-19 Twitter hashtags and the daily increase in COVID-19 cases.*

Emanating from the study results drawn from chapter five, it is evident that the findings of the study revealed that there existed a relationship between the daily increases of COVID-19 and Twitter hashtags. Evidence was first drawn from the Pearson's correlation matrix, which revealed a positive associateship between the two variables under study. The regression model further revealed that the null hypothesis that there exists no relationship between daily increases in COVID-19 cases and Twitter hashtags was rejected as the p-value of the overall model was highly significant at 5% level. Hence, it was concluded that Twitter hashtags have a significant effect on the daily increase in COVID-19 cases across some selected countries.

*Specific Objective II*: To ascertain the strength of association between COVID-19 Twitter hashtags and the daily increases of COVID-19 cases.

In order to ascertain the strength of the association between the daily increases in COVID-19 cases and Twitter hashtags, Pearson's correlation analysis was used. The study results revealed that a positive association existed, though not very strong, but the results meant that there existed an association between the two variables under study.

*Specific Objective III*: *To highlight the pertinent findings related to COVID-19 Twitter hashtags and the daily increases of COVID-19 cases in contrast to other studies and thus, formulate conclusions.*

This study focused mainly on the relationship that existed between the daily increases of COVID-19 and the Twitter hashtags. However, further analysis was conducted based on correlation analysis so as to draw other pertinent results. The results of both the weekly and monthly correlations between the two variables of understudy revealed a positive association with the monthly exhibiting a relatively stronger association compared to the daily and weekly associations.

# 7 References


Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *Journal of Medical Internet Research*, *22*(4), e19016.

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., & Liu, B. (2011). Predicting flu trends using twitter data. *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 702–707.

Aguilar-Gallegos, N., R.-G., L. E., M.-G., E. G., G.-S., I., E., & Aguilar-Ávila, J. (2020). *Dataset on dynamics of Coronavirus on Twitter*. 105684. https://www.sciencedirect.com/science/article/pii/S2352340920305783

Ahmed, W., Vidal-Alaball, J., Downing, J., Segu\'\i, F. L., & others. (2020). COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of Medical Internet Research*, *22*(5), e19458.

Alqurashi, S., Alhindi, A., & Alanazi, E. (2020). Large arabic twitter dataset on covid-19. *ArXiv Preprint ArXiv:2004.04315*.

Amani, F., Aghaie, B., Zeynizadeh, S., Tabrizian, S., Aslanian, R., & Jafarizadeh, R. (2020). *Using social network rates among Ardabil city women over 25 years old*.

Auter, P. J., Douai, A., Makady, H., & West, C. (2016). Circulating health rumors in the 'Arab World': a 12-month content analysis of news stories and reader commentary about Middle East Respiratory Syndrome from two Middle Eastern news outlets. *International Communication Gazette*, *78*(5), 411–431.

Banda, J. M., Tekumalla, R., Wang, G., Yu, J., Liu, T., Ding, Y., Artemova, K., Tutubalina, E., & Chowell, G. (2020). A large-scale COVID-19 Twitter chatter dataset for open scientific research--an international collaboration. *ArXiv Preprint ArXiv:2004.03688*.

Barros, J. M., Duggan, J., & Rebholz-Schuhmann, D. (2020). The application of internet-based sources for public health surveillance (infoveillance): systematic review. *Journal of Medical Internet Research*, *22*(3), e13680.

Barton, E. (2021). *Massive veterans group uses intel, satellite images to direct Afghan interpreters around Taliban checkpoints*. Fox News. https://www.foxnews.com/world/afghan-interpreters-veterans-roadmaps-satellite-imagery-taliban-checkpoints

Bisanzio, D., Kraemer, M. U. G., Bogoch, I. I., Brewer, T., Brownstein, J. S., & Reithinger, R. (2020). Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale. *Geospatial Health*, *15*(1).

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, *497*, 38–55.

Boyd, D., & Ellison, N. (2007). *Social network sites: Definition, history, and scholarship, Computer-Mediated Communication, 13 (1), Article 11*. Retrieved.

Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J., & Kowalchuck, M. (2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. *Safety Science*, *90*, 48–61.

Castillo, C. (2016). *Big crisis data: social media in disasters and time-critical situations*. Cambridge University Press.

Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., & Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers \& Graphics*, *38*, 51–60.

Chatfield, A. T., Scholl, H. J. J., & Brajawidagda, U. (2013). Tsunami early warnings via Twitter in government: Net-savvy citizens' co-production of time-critical public information services. *Government Information Quarterly*, *30*(4), 377–386.

Chen, K, L., & E, F. (2020). *Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set JMIR Public Health Surveill*. *2*. https://doi.org/10.2196/19273

Chen, L., Achrekar, H., Liu, B., & Lazarus, R. (2010). Vision: towards real time epidemic vigilance through online social networks: introducing SNEFT--social network enabled flu trends. *Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond*, 1–5.

Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. *16*, 297–334. https://doi.org/10.1007/BF02310555

Cruickshank, I. J., & Carley, K. M. (n.d.). *Characterizing communities of hashtag usage on twitter during the 2020 COVID-19 pandemic by multi-view clustering. Appl Netw Sci 5, 66 (2020)*.

Dastani, M., Mohammadpour, A., & Bagheri, J. (2019). The Opportunities and Damages of Virtual Social Networks from Students' Perspectives; the Experience of Iranian Users. *Library Philosophy and Practice*, 1.

Dastani, M., & others. (2016). The role of Social Networks on Academic Achievement of Gonabad University of Medical Science'students. *The Journal of Medical Education and Development*, *11*(2), 153–160.

Dastani, M., & Ramezani, A. (2017). Role of membership in scientific and specialized groups of virtual social networks in increasing knowledge, professional skills and e-learning: A case study. *International Research: Journal of Library and Information Science*, *7*(4).

de Las Heras-Pedrosa, Carlos, Pablo, S.-N., & Peláez Ignacio, J. (2020). *Sentiment analysis and emotion understanding during the COVID-19 pandemic in Spain and its impact on digital ecosystems. 17*(15), 5542. https://www.mdpi.com/1660-4601/17/15/5542

Edosomwan, S., Prakasan, S. K., Kouame, D., Watson, J., & Seymour, T. (2011). The history of social media and its impact on business. *Journal of Applied Management and Entrepreneurship*, *16*(3), 79–91.

Emily Chen. (2020). *Lopezbec*. COVID-19-TweetIDs. https://github.com/lopezbec/COVID19_Tweets_Dataset

Gangal, V. L. N. & Gupta, H. (2013). Public expenditure and economic growth: A case study of India. *Global Journal of Management and Business Studies*, *3*(2), 191–196.

Garba, T. & Abdullahi, S. Y. (2013). *Public expenditure and economic growth; An application of co-integration and Granger causality test on Nigeria. Journal of Economics and Social Research*. *15*(1), 1–30.

Golder, Su, Klein, A. Z., Magge, A., O'Connor, K., Cai, H., & Davy Weissenbacher, G. G.-H. (2020). *Extending A chronological and geographical analysis of personal reports of COVID-19 on Twitter to England, UK.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7273260/

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, *3*(3), 231–241.

Grover, S., & Aujla, G. S. (2014). Prediction model for influenza epidemic based on Twitter data. *International Journal of Advanced Research in Computer and Communication Engineering*, *3*(7), 7541–7545.

Gruzd, A., & Mai, P. (2020). Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data \& Society*, *7*(2), 2053951720938405.

Gui, X., Kou, Y., Pine, K. H., & Chen, Y. (2017). Managing uncertainty: using social media for risk assessment during a public health crisis. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 4520–4533.

Huang, J., Zhao, H., & Zhang, J. (2013). Detecting flu transmission by social sensor in China. *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, 1242–1247.

Hussain, Akhtar, Bhowmik, B., & Moreira, N. C. do V. (2020). *COVID-19 and diabetes: Knowledge in progress*. *1*(162), 108142. https://www.sciencedirect.com/science/article/pii/S0168822720303922

Imran, M., Castillo, C., Diaz, F., & Vieweg, S. (2015). Processing social media messages in mass emergency: A survey. *ACM Computing Surveys (CSUR)*, *47*(4), 1–38.

Imran, M., Castillo, C., Lucas, J., Meier, P., & Vieweg, S. (2014). AIDR: Artificial intelligence for disaster response. *Proceedings of the 23rd International Conference on World Wide Web*, 159–162.

Imran, M., Mitra, P., & Castillo, C. (2016). Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages. *ArXiv Preprint ArXiv:1605.05894*.

Imran, M., Ofli, F., Caragea, D., & Torralba, A. (2020). *Using ai and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions*. Elsevier.

Jasra, M. (2010). *The History of Social Media [Infographic]*. Web Analytics World. https://www.researchgate.net/deref/http%3A%2F%2Fwww.webanalyticsworld.net%2F2010%2F11%2Fhistory-of-social-media-infographic.html

Joppe, M. (2020). *The Research Process*. http://www.ryerson.ca/~mjoppe/rp.htm

Kalyanam, J., Quezada, M., Poblete, B., & Lanckriet, G. (2016). Prediction and characterization of high-activity events in social media triggered by real-world news. *PloS One*, *11*(12), e0166694.

Kim, B. (2020). Effects of social grooming on incivility in COVID-19. *Cyberpsychology, Behavior, and Social Networking*, *23*(8), 519–525.

Klein, A. Z., Magge, A., M., O. K., Cai, H., & Weissenbacher, D. Gonzalez-Hernandez, G. (2020). *A chronological and geographical analysis of personal reports of COVID-19 on Twitter*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7276035/

Kullar, R., Marcelin, J. R., Swartz, T. H., Piggott, D. A., Macias Gil, R., Mathew, T. A., & Tan, T. (2020). Racial disparity of coronavirus disease 2019 in African American communities. *The Journal of Infectious Diseases*, *222*(6), 890–893.

Lamsal, R. (2020). *Coronavirus (covid-19) tweets dataset*.

Landwehr, P. M., Wei, W., Kowalchuck, M., & Carley, K. M. (2016). Using tweets to support disaster planning, warning and response. *Safety Science*, *90*, 33–47.

Lee, B., Yoon, J., Kim, S., & Hwang, B.-Y. (2012). Detecting social signals of flu symptoms. *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 544–545.

Liu, Q., Gao, Y., & Chen, Y. (2014). Study on disaster information management system

compatible with VGI and crowdsourcing. *2014 IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA)*, 464–468.
Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. *JMIR Public Health and Surveillance*, *6*(2), e19447.
MacDonald, J. (2017). *The Importance of Hashtags: Know Where, Why, and How to Use Them*. Business2community.Com. https://www.business2community.com/social-media/importance-hashtags-know-use-01837644
Maclean, F., Jones, D., Carin-Levy, G., & Hunter, H. (2013). Understanding Twitter. *British Journal of Occupational Therapy*, *76*, 295. https://doi.org/10.4276/030802213X13706169933021
Muhammad, Xu, T., & Karim, R. (2015). Impact of expenditure on economic growth in Pakistan. *International Journal of Academic Research in Business and Social Sciences*, *5*(2), 231–236.
Orodho, J. A. (2009). *Techniques of writing research proposals and reports in education and social sciences*.
Page, J., Hinshaw, D., & McKay, B. (2021). In Hunt for Covid-19 Origin, Patient Zero Points to Second Wuhan Market-The man with the first confirmed infection of the new coronavirus told the WHO team that his parents had shopped there. *The Wall Street Journal. Retrieved*, *27*.
Panuganti, B. A., Jafari, A., MacDonald, B., & DeConde, A. S. (2020). <? covid19?> Predicting COVID-19 Incidence Using Anosmia and Other COVID-19 Symptomatology: Preliminary Analysis Using Google and Twitter. *Otolaryngology--Head and Neck Surgery*, *163*(3), 491–497.
Park, H. W., Park, S., & Chong, M. (2020). Conversations and medical news frames on twitter: Infodemiological study on covid-19 in south korea. *Journal of Medical Internet Research*, *22*(5), e18897.
Pempek, T. A., Yermolayeva, Y. A., & Calvert, S. L. (2009). College students' social networking experiences on Facebook. *Journal of Applied Developmental Psychology*, *30*(3), 227–238.
Purohit, H., A, H., VL, S., AP, S., J, F., & S, B. (2013). *What kind of# conversation is twitter? mining# psycholinguistic cues for emergency coordination. 29*(6), 2438–2447.
Rizo, C. A., Lupea, D., Baybourdy, H., Anderson, M., Closson, T., & Jadad, A. R. (2005). What Internet services would patients like from hospitals during an epidemic? Lessons from the SARS outbreak in Toronto. *Journal of Medical Internet Research*, *7*(4), e46.
Rufai, S. R., & Bunce, C. (2020). World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. *Journal of Public Health*, *42*(3), 510–516.
Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. *Proceedings of the 19th International Conference on World Wide Web*, 851–860.
Saunders. (2012). Research methods forbusiness students. *Essex: Prentice Hall: Financial Times*.
Saunders, M., Lewis, P., & Thornhill, A. (2003). Research methods forbusiness students. *Essex: Prentice Hall: Financial Times*.
Sheldon, P., Herzfeldt, E., & Rauschnabel, P. A. (2020). Culture and social media: the relationship between cultural values and hashtagging styles. *Behaviour \& Information*

*Technology*, *39*(7), 758–770.
Su, Yue, Xue, J., Liu, X., Wu, P., Chen, J., Chen, C., Liu, T., Gong, W., & Zhu, T. (2020). *Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: a psycholinguistic analysis on Weibo and Twitter*. *17*(12), 4552. https://www.mdpi.com/1660-4601/17/12/4552
Tang, X., & Yang, C. C. (2010). Identifing influential users in an online healthcare social network. *2010 IEEE International Conference on Intelligence and Security Informatics*, 43–48.
Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging during two natural hazards events: what twitter may contribute to situational awareness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1079–1088.
Viral lies. (2020). No Title. *Misinformation and the Coronavirus. Technical Report*. https://www.article19.org/wp-content/uploads/2020/03/Coronavirus-briefing.pdf.
Wajahat Hussain. (2020). Role of Social Media in COVID-19 Pandemic. *The International Journal of Frontier Sciences*. https://doi.org/https://doi.org/10.37978/tijfs.v4i2.144
Wang, W., Ma, Y., Wu, T., Dai, Y., Chen, X., & Braunstein, L. A. (2019). Containing misinformation spreading in temporal social networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(12), 123131.
*Worldometers*. (n.d.). https://www.worldometers.info/coronavirus/
Xue, J., Y, S., X, L., P, W., J, C., C, C., T, L., W, G., & T, Z. (2020). *Examining the impact of covid-19 lockdown in wuhan and lombardy: a psycholinguistic analysis on weibo and twitter*. *17*(12), 4552.
Yum, S. (2020). Social Network Analysis for Coronavirus (COVID-19) in the United States. *Social Science Quarterly*, *101*(4), 1642–1647. https://doi.org/10.1111/ssqu.12808
Zhang, Y. (2019). Language in our time: An empirical analysis of hashtags. *The World Wide Web Conference*, 2378–2389.
Zimmer, C. (2021). *The Secret Life of a Coronavirus-An oily, 100-nanometer-wide bubble of genes has killed more than two million people and reshaped the world. Scientists don't quite know what to make of it*. Retrieved.

# 8 Appendix

Table 9: Daily Data

| DAY | Daily increase-COVID-19 | TWIT#TAG |
|---|---|---|
| 1 | 2648.00 | 409683 |
| 2 | 302.00 | 137836 |
| 3 | 50.00 | 521958 |
| 4 | 200.00 | 516909 |
| 5 | 1121.00 | 179629 |
| 6 | 822.00 | 232732 |
| 7 | 651.00 | 200671 |
| 8 | 151.00 | 203841 |
| 9 | 1347.00 | 195921 |
| 10 | 1558.00 | 188392 |
| 11 | 2537.00 | 208811 |
| 12 | 2020.00 | 162450 |
| 13 | 3301.00 | 155593 |
| 14 | 607.00 | 165771 |
| 15 | 2061.00 | 143163 |
| 16 | 1332.00 | 150803 |
| 17 | 2324.00 | 446438 |
| 18 | 4084.00 | 221591 |
| 19 | 4850.00 | 383148 |
| 20 | 6823.00 | 454765 |
| 21 | -5164.00 | 463779 |
| 22 | 3059.00 | 370145 |
| 23 | 7920.00 | 365915 |
| 24 | 429.00 | 111568 |
| 25 | 4164.00 | 190197 |
| 26 | 12586.00 | 399444 |
| 27 | 4900.00 | 360536 |
| 28 | 837.00 | 416823 |
| 29 | -6295.00 | 440916 |
| 30 | 3876.00 | 314568 |
| 31 | 10708.00 | 141622 |
| 32 | 2995.00 | 256577 |
| 33 | 2655.00 | 364600 |
| 34 | 5311.00 | 355120 |
| 35 | -2883.00 | 400725 |
| 36 | -9447.00 | 415218 |
| 37 | 3594.00 | 449615 |

| | | |
|---|---|---|
| 38 | 6288.00 | 420892 |
| 39 | 5174.00 | 301758 |
| 40 | 1415.00 | 249855 |
| 41 | 7560.00 | 370946 |
| 42 | -12251.00 | 375139 |
| 43 | -9157.00 | 364295 |
| 44 | 60.00 | 391722 |
| 45 | 3337.00 | 386105 |
| 46 | 6606.00 | 403688 |
| 47 | 743.00 | 446578 |
| 48 | 1984.00 | 415856 |
| 49 | -4896.00 | 405983 |
| 50 | -6059.00 | 386764 |
| 51 | 4169.00 | 371892 |
| 52 | -1147.00 | 116393 |
| 53 | 2376.00 | 369705 |
| 54 | 2455.00 | 384458 |
| 55 | 11931.00 | 455707 |
| 56 | -7018.00 | 476309 |
| 57 | -16823.00 | 467848 |
| 58 | -2941.00 | 462685 |
| 59 | 6427.00 | 355258 |
| 60 | 3269.00 | 276472 |
| 61 | 7797.00 | 351073 |
| 62 | 6983.00 | 387607 |
| 63 | -12458.00 | 437726 |
| 64 | -644.00 | 381961 |
| 65 | -570.00 | 370511 |
| 66 | 953.00 | 202516 |
| 67 | 13681.00 | 240978 |
| 68 | -25.00 | 477347 |
| 69 | -124.00 | 450479 |
| 70 | -7902.00 | 443590 |
| 71 | -8209.00 | 447543 |
| 72 | -6444.00 | 460810 |
| 73 | 14110.00 | 403788 |
| 74 | 3430.00 | 390524 |
| 75 | 6964.00 | 402240 |
| 76 | 3491.00 | 428065 |
| 77 | -5258.00 | 409239 |
| 78 | -13613.00 | 396484 |

| | | |
|---|---|---|
| 79 | 9488.00 | 430973 |
| 80 | 4956.00 | 459411 |
| 81 | 6959.00 | 456179 |
| 82 | 5261.00 | 439230 |
| 83 | -549.00 | 383761 |
| 84 | -7503.00 | 295793 |
| 85 | -3539.00 | 280315 |
| 86 | -5210.00 | 325211 |
| 87 | 2389.00 | 380763 |
| 88 | 13860.00 | 385003 |
| 89 | 9817.00 | 349108 |
| 90 | 9480.00 | 355219 |
| 91 | -2083.00 | 309739 |
| 92 | -15732.00 | 257828 |
| 93 | -6664.00 | 296079 |
| 94 | 13519.00 | 292423 |
| 95 | 4966.00 | 333897 |
| 96 | 10352.00 | 307755 |
| 97 | 306.00 | 242845 |
| 98 | -2343.00 | 440731 |
| 99 | -14299.00 | 582951 |
| 100 | -26218.00 | 582951 |
| 101 | 53302.00 | 580923 |
| 102 | -5425.00 | 603498 |
| 103 | 2367.00 | 613154 |
| 104 | 5052.00 | 555504 |
| 105 | -8588.00 | 492913 |
| 106 | -9820.00 | 446690 |
| 107 | 1754.00 | 533673 |
| 108 | 17759.00 | 534963 |
| 109 | 1946.00 | 546140 |
| 110 | -5181.00 | 549243 |
| 111 | 41332.00 | 518686 |
| 112 | -24299.00 | 450263 |
| 113 | -27485.00 | 409229 |
| 114 | 9824.00 | 527048 |
| 115 | 24118.00 | 550554 |
| 116 | 10401.00 | 524715 |
| 117 | 6491.00 | 544305 |
| 118 | 14581.00 | 485246 |
| 119 | -17554.00 | 496508 |

| | | |
|---|---|---|
| 120 | -13650.00 | 468514 |
| 121 | -1249.00 | 497688 |
| 122 | 12880.00 | 504698 |
| 123 | 22741.00 | 530347 |
| 124 | 11133.00 | 450746 |
| 125 | 5381.00 | 451483 |
| 126 | -19731.00 | 453807 |
| 127 | -16833.00 | 436842 |
| 128 | -6193.00 | 463858 |
| 129 | 36751.00 | 415146 |
| 130 | 5368.00 | 489883 |
| 131 | 10232.00 | 509344 |
| 132 | 13446.00 | 477470 |
| 133 | -20438.00 | 438349 |
| 134 | -19116.00 | 423828 |
| 135 | -1238.00 | 498068 |
| 136 | 23357.00 | 519174 |
| 137 | 15558.00 | 505993 |
| 138 | 13117.00 | 409574 |
| 139 | -7921.00 | 509769 |
| 140 | 5671.00 | 499200 |
| 141 | -45155.00 | 461681 |
| 142 | 3857.00 | 463389 |
| 143 | 34902.00 | 526415 |
| 144 | 39993.00 | 506063 |
| 145 | -4306.00 | 499425 |
| 146 | 13025.00 | 490651 |
| 147 | -27450.00 | 421328 |
| 148 | -37691.00 | 425375 |
| 149 | -5387.00 | 453829 |
| 150 | 30690.00 | 509227 |
| 151 | 39877.00 | 508753 |
| 152 | -1891.00 | 480060 |
| 153 | 4336.00 | 444284 |
| 154 | -31952.00 | 406377 |
| 155 | -36321.00 | 442473 |
| 156 | -22655.00 | 460532 |
| 157 | 53386.00 | 477837 |
| 158 | 18703.00 | 479044 |
| 159 | 11401.00 | 448211 |
| 160 | -613.00 | 484775 |

| | | |
|---|---|---|
| 161 | -12563.00 | 376467 |
| 162 | -45598.00 | 414638 |
| 163 | -6899.00 | 341309 |
| 164 | 49435.00 | 369610 |
| 165 | 21592.00 | 445582 |
| 166 | -2702.00 | 453421 |
| 167 | 3441.00 | 411479 |
| 168 | -25636.00 | 380488 |
| 169 | -46515.00 | 386916 |
| 170 | -14937.00 | 451505 |
| 171 | 55970.00 | 435889 |
| 172 | 16856.00 | 459025 |
| 173 | -4067.00 | 434798 |
| 174 | -8302.00 | 358510 |
| 175 | 5624.00 | 351508 |
| 176 | -56865.00 | 373171 |
| 177 | 5894.00 | 451780 |
| 178 | 37165.00 | 409593 |
| 179 | 23663.00 | 324066 |
| 180 | -736.00 | 358714 |
| 181 | 12326.00 | 362332 |
| 182 | -24619.00 | 312482 |
| 183 | -38153.00 | 386787 |
| 184 | 22193.00 | 416378 |
| 185 | 12814.00 | 417501 |
| 186 | 25034.00 | 399804 |
| 187 | 2360.00 | 437709 |
| 188 | 17116.00 | 419133 |
| 189 | -28349.00 | 301189 |
| 190 | -37819.00 | 362347 |
| 191 | -36367.00 | 384333 |
| 192 | 43841.00 | 445089 |
| 193 | 41121.00 | 424464 |
| 194 | 13665.00 | 441659 |
| 195 | 14931.00 | 421773 |
| 196 | -26951.00 | 361744 |
| 197 | -36802.00 | 352420 |
| 198 | -4242.00 | 436189 |
| 199 | 30847.00 | 416577 |
| 200 | 28025.00 | 450159 |
| 201 | 4962.00 | 414220 |

| | | |
|---|---|---|
| 202 | 11970.00 | 454277 |
| 203 | -25162.00 | 353477 |
| 204 | -42267.00 | 279155 |
| 205 | -21123.00 | 313884 |
| 206 | 39702.00 | 338171 |
| 207 | 39642.00 | 308654 |
| 208 | -1298.00 | 336817 |
| 209 | 10980.00 | 313666 |
| 210 | -21748.00 | 258459 |
| 211 | -43740.00 | 267270 |
| 212 | -23220.00 | 325217 |
| 213 | 51695.00 | 353729 |
| 214 | 26457.00 | 319214 |
| 215 | 8681.00 | 285266 |
| 216 | 3897.00 | 292818 |
| 217 | -22979.00 | 266120 |
| 218 | -48520.00 | 285966 |
| 219 | 14289.00 | 352933 |
| 220 | 39390.00 | 339645 |
| 221 | 38272.00 | 315790 |
| 222 | 7303.00 | 339851 |
| 223 | 6868.00 | 342916 |
| 224 | 10748.00 | 282791 |
| 225 | -83472.00 | 296972 |
| 226 | -14830.00 | 315362 |
| 227 | 41414.00 | 361886 |
| 228 | 67882.00 | 349032 |
| 229 | 17380.00 | 340906 |
| 230 | 15839.00 | 347773 |
| 231 | -24114.00 | 300064 |
| 232 | -45151.00 | 313704 |
| 233 | 390286.00 | 324979 |
| 234 | -296949.00 | 329425 |
| 235 | 39248.00 | 299733 |
| 236 | 13193.00 | 314015 |
| 237 | -18543.00 | 270761 |
| 238 | -44529.00 | 301853 |
| 239 | -16623.00 | 298717 |
| 240 | 47585.00 | 323585 |
| 241 | 50403.00 | 321304 |
| 242 | 34238.00 | 330282 |

| | | |
|---|---|---|
| 243 | 26756.00 | 360201 |
| 244 | -74374.00 | 335625 |
| 245 | -31761.00 | 260194 |
| 246 | 8065.00 | 318485 |
| 247 | 15877.00 | 313297 |
| 248 | 79026.00 | 263239 |
| 249 | 50977.00 | 309199 |
| 250 | -32154.00 | 286174 |
| 251 | 78721.00 | 239689 |
| 252 | -165444.00 | 200793 |
| 253 | -11901.00 | 312749 |
| 254 | 66317.00 | 311772 |
| 255 | 64902.00 | 305953 |
| 256 | 27248.00 | 310227 |
| 257 | 16446.00 | 280295 |
| 258 | -64201.00 | 252826 |
| 259 | -84099.00 | 295868 |
| 260 | -8143.00 | 312275 |
| 261 | 51336.00 | 310295 |
| 262 | 62407.00 | 335916 |
| 263 | 33070.00 | 322727 |
| 264 | 14025.00 | 315290 |
| 265 | -64828.00 | 269821 |
| 266 | -80213.00 | 261276 |
| 267 | 13477.00 | 317644 |
| 268 | 22704.00 | 311462 |
| 269 | 18597.00 | 302249 |
| 270 | 99471.00 | 269742 |
| 271 | -60719.00 | 242384 |
| 272 | -30242.00 | 219775 |
| 273 | -70943.00 | 209936 |
| 274 | -18922.00 | 253443 |
| 275 | 82649.00 | 547974 |
| 276 | 58593.00 | 35694 |
| 277 | 48004.00 | 238542 |
| 278 | 5176.00 | 267499 |
| 279 | -51530.00 | 242523 |
| 280 | -90831.00 | 218552 |
| 281 | -17669.00 | 253973 |
| 282 | 68677.00 | 113896 |
| 283 | 58108.00 | 285541 |

| | | |
|---|---|---|
| 284 | 37611.00 | 297114 |
| 285 | 14336.00 | 251648 |
| 286 | -59068.00 | 229506 |
| 287 | -99746.00 | 244440 |
| 288 | -19540.00 | 301580 |
| 289 | 61566.00 | 305545 |
| 290 | 128435.00 | 299980 |
| 291 | 12178.00 | 310209 |
| 292 | -13860.00 | 296715 |
| 293 | -88469.00 | 268858 |
| 294 | -70726.00 | 311576 |
| 295 | -21301.00 | 359451 |
| 296 | 71278.00 | 298142 |
| 297 | 82036.00 | 309063 |
| 298 | -3072.00 | 252927 |
| 299 | -154960.00 | 176884 |
| 300 | -68383.00 | 245401 |
| 301 | -31158.00 | 251377 |
| 302 | 66068.00 | 281009 |
| 303 | 126157.00 | 328017 |
| 304 | 108147.00 | 326176 |
| 305 | 17472.00 | 275616 |
| 306 | -120933.00 | 230350 |
| 307 | -61505.00 | 249692 |
| 308 | -41449.00 | 287439 |
| 309 | 31449.00 | 325652 |
| 310 | 146129.00 | 343757 |
| 311 | 96378.00 | 285100 |
| 312 | 35231.00 | 293183 |
| 313 | 988.00 | 308614 |
| 314 | -75178.00 | 240722 |
| 315 | -125679.00 | 244374 |
| 316 | -48006.00 | 308093 |
| 317 | 90015.00 | 332771 |
| 318 | 61188.00 | 329824 |
| 319 | 8800.00 | 278591 |
| 320 | 6048.00 | 251648 |
| 321 | -83104.00 | 330343 |
| 322 | -121514.00 | 257941 |
| 323 | -75369.00 | 285752 |
| 324 | 110715.00 | 309426 |

| | | |
|---|---|---|
| 325 | 74887.00 | 291770 |
| 326 | -15266.00 | 299733 |
| 327 | -11323.00 | 314015 |
| 328 | -38372.00 | 270761 |
| 329 | -118112.00 | 271860 |
| 330 | -39163.00 | 338604 |
| 331 | 85391.00 | 325809 |
| 332 | 63430.00 | 304448 |
| 333 | 4708.00 | 68704 |
| 334 | -10521.00 | 305611 |
| 335 | -58790.00 | 260730 |
| 336 | -112762.00 | 238262 |

***Table 10: Weekly Data***

| WEEK | Weekly increase-COVID-19 | TWIT#TAG |
|---|---|---|
| 1 | 27056 | 2199418 |
| 2 | 53848 | 1280779 |
| 3 | 113708 | 2263687 |
| 4 | 176892 | 2214628 |
| 5 | 161551 | 2274128 |
| 6 | 68839 | 2583423 |
| 7 | -17683 | 2814227 |
| 8 | 20647 | 2561228 |
| 9 | -28927 | 2738669 |
| 10 | 54411 | 2567382 |
| 11 | 5006 | 2942209 |
| 12 | 67264 | 2861831 |
| 13 | 71238 | 2385358 |
| 14 | 81994 | 2171558 |
| 15 | 60178 | 4011894 |
| 16 | 128496 | 3579658 |
| 17 | 143076 | 3537605 |
| 18 | 159558 | 3357283 |
| 19 | 133173 | 3230892 |
| 20 | 135900 | 3365606 |
| 21 | 166260 | 3368952 |
| 22 | 64687 | 3227905 |
| 23 | -29993 | 3169339 |
| 24 | 53490 | 2816527 |
| 25 | -81806 | 2878151 |

| | | |
|---|---|---|
| 26 | 39148 | 2592138 |
| 27 | 105441 | 2778501 |
| 28 | -7491 | 2841409 |
| 29 | 139220 | 2877319 |
| 30 | 1042 | 2148806 |
| 31 | 26088 | 2109634 |
| 32 | 218401 | 2259892 |
| 33 | 200943 | 2311995 |
| 34 | 931545 | 2154470 |
| 35 | 64890 | 2229908 |
| 36 | 441984 | 1930876 |
| 37 | 184119 | 2069690 |
| 38 | 45586 | 2127600 |
| 39 | -66110 | 1808991 |
| 40 | 208625 | 1804757 |
| 41 | 116404 | 1723725 |
| 42 | 95513 | 2152334 |
| 43 | -523522 | 1814803 |
| 44 | 380214 | 2022942 |
| 45 | 801576 | 2023843 |
| 46 | -405824 | 2066870 |
| 47 | -660368 | 2096169 |
| 48 | -418851 | 1796133 |
| 49 | -232400 | 1606772 |
| 50 | -684794 | 1609273 |
| 51 | -254911 | 1653853 |
| 52 | 177526 | 913513 |

***Table 11: Monthly Data***

| MONTH | Monthly increase-COIVD-19 | TWIT#TAG |
|---|---|---|
| March | 876211 | 8855618 |
| April | 1641429 | 11249236 |
| May | 515978 | 11839941 |
| June | 1372958 | 14513787 |
| July | 2889293 | 14673361 |
| August | 758396 | 12665697 |
| September | 571636 | 11108300 |
| October | 3647071 | 9540275 |
| November | 5079538 | 8514995 |
| December | 2091705 | 8425428 |
| January | 118396 | 8783579 |

| | | |
|---|---|---|
| February | -8049145 | 6075980 |

***Table 12: Major Pandemics: Historical Overviews***

| Name | Time Period | Type/Pre-human host | Estimated Death Toll |
|---|---|---|---|
| Antoine Plague | 165-180 | Believed to be ether smallpox or measles | 5 million |
| Japanese smallpox epidemic | 735-737 | Variola major virus | 1 million |
| Plague of Justinian | 541-542 | Yersinia pests bacteria/rats, fleas | 30 to 50 million |
| Black Death | 1347-1351 | Yersinia pests bacteria/rats, fleas | 200 million |
| New World Smallpox Outbreak | 1520-onwards | Variola major virus | 56 million |
| Great Plague of London | 1665 | Yersinia pests bacteria/rats, fleas | 100,000 |
| Italian Plague | 1629-1631 | Yersinia pests bacteria/rats, fleas | 1 million |
| Cholera Pandemics 1-6 | 1817-1923 | V. cholera bacteria | 1 million+ |
| Third Plague | 1885 | Yersinia pests bacteria/rats, fleas | 12 million (China & India) |
| Yellow Fever | Late 1800s | Virus/Mosquitoes | 100,000-150,000 (US) |
| Russian Flu | 1889-1890 | H2N2 (avan orgn) | 1 mllon |
| Spanish Flu | 1918-1919 | H1N1 Virus/pgs | 40 to 50 mllon |
| Asan Flu | 1957-1958 | H2N2 Virus | 1.1 mllon |
| Hong Kong Flu | 1968-1970 | H3N2 Virus | 1 mllon |
| HIV/AIDS | 1981-present | Virus/chimpanzees | 25 to 35 mllon |
| Swine Flu | 2009-2010 | H1N1 Virus/pgs | 200,000 |
| SARS | 2002-2003 | CoronaVirus/bats, cvets | 770 |
| Ebola | 2014-2016 | EbolaVirus/ wild animals | 11,000 |
| MERS | 2015-present | CoronaVirus/bats, camels | 850 |

*Source: World Economc Forum (2020)*

***Table 13: Regression Model***

Dependent Variable: DALY_INCREASE_COVID_19
Method: Least Squares
Date: 07/21/21 Time: 22:09
Sample: 1 336
included observations: 336

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -19366.76 | 9724.264 | -1.991591 | 0.0472 |
| TWIT_TAG | 0.057623 | 0.026163 | 2.202473 | 0.0283 |

| | | | |
|---|---|---|---|
| R-squared | 0.014316 | Mean dependent var | 1242.009 |
| Adjusted R-squared | 0.011365 | S.D. dependent var | 48796.10 |
| S.E. of regression | 48518.04 | Akaike info criterion | 24.42319 |
| Sum squared resid | 7.86E+11 | Schwarz criterion | 24.44591 |
| Log-likelihood | -4101.097 | Hannan-Quinn criteri. | 24.43225 |
| F-statistic | 4.850887 | Durbin-Watson stat | 1.832923 |
| Prob(F-statistic) | 0.028315 | | |