

|  |  |
| --- | --- |
| **COURSEWORK COVER PAGE** | |
| **Module Number:** | LD7091 |
| **Module Title:** | Research Methods for Professional Practices |
| **Module Tutor Name:** | Ning Tse [NT] |
| **Coursework Title:** | Assignment |
| **Student Name:** | Busra Ecem Sakar |
| **Student ID:** | w22065255 |
| **Programme of Study:** | MSc Big Data & Data Science Technology [BD] |
| **Word count:** | Part A: 1013 Part C: 3428 |

**Declaration**

*I confirm that this assessment is my own work and that I have duly acknowledged and correctly referenced the work of others. I am aware of and understand that any breaches to the Code of Academic Conduct will be investigated and sanctioned in accordance with the Academic Conduct Regulation.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Your signature:** | ECEM | **Date:** | 11.05.2023 |

**Contents**

[Part 1 – Reflective CPD 4](#_Toc134687910)

[1.1 Self-reflection 4](#_Toc134687911)

[1.2 Personal SWOT analysis 5](#_Toc134687912)

[1.3 Research personal development plan 6](#_Toc134687913)

[Part 2 – Research Logs 8](#_Toc134687914)

[2.1. Skill Audit Tools 8](#_Toc134687915)

[2.2. Personal SWOT Analysis 8](#_Toc134687916)

[2.3. Research Personal Development Plan 9](#_Toc134687917)

[2.4. My first literature search 10](#_Toc134687918)

[2.5. Mind Map of My research 11](#_Toc134687919)

[2.6. Draft my project title, aims, research questions and objectives 12](#_Toc134687920)

[2.7. Research Methodology Plan 12](#_Toc134687921)

[2.8. My project plan using Gantt Chart 13](#_Toc134687922)

[2.9. Draft Student Project Approval Form 14](#_Toc134687923)

[Part 3 – Research Proposal 15](#_Toc134687924)

[3.1. Introduction 15](#_Toc134687925)

[3.1.1 Project Title 16](#_Toc134687926)

[3.1.2 Background 16](#_Toc134687927)

[3.1.3 Aim 16](#_Toc134687928)

[3.1.4 Research Questions 16](#_Toc134687929)

[3.1.5 SMART Objectives 17](#_Toc134687930)

[3.1.6 Research Scope 17](#_Toc134687931)

[3.2. Literature Review 17](#_Toc134687932)

[3.3. Research Methodology and Planning 19](#_Toc134687933)

[3.3.1 Research Methodology 19](#_Toc134687934)

[3.3.2 Research Plan and Task Lists 20](#_Toc134687935)

[3.3.3 Risk Assessment and Management 22](#_Toc134687936)

[3.4. Conclusions 22](#_Toc134687937)

[Appendix: Draft Student Project Approval Form 26](#_Toc134687938)

[References 32](#_Toc134687939)

**Tables**

[Table 1. SWOT Anaysis about Research 5](#_Toc134687943)

[Table 2. Research Personal Development Plan 7](#_Toc134687944)

**Figures**

[Figure 1 Researcher Development Assessment tool, Overall Anaysis 8](#_Toc134041835)

[Figure 2 Personal SWOT Analysis 9](#_Toc134041836)

[Figure 3 Research Personal Development Plan 10](#_Toc134041837)

[Figure 4 My first literature search 11](#_Toc134041838)

[Figure 5 Mind Map of My Research 11](#_Toc134041839)

[Figure 6 Draft my project title, aims, research questions and objectives 12](#_Toc134041840)

[Figure 7 Research Methodology Plan 13](#_Toc134041841)

[Figure 8 My project plan using Gantt Chart 13](#_Toc134041842)

[Figure 9 Draft Student Project Approval Form 14](#_Toc134041843)

# Part 1 – Reflective CPD

## 1.1 Self-reflection

Self-reflection is an important part of professional life. It helps individuals to recognise their own strengths and flaws and enhance their abilities. Self-reflection can assist students in building a better knowledge of their methods of learning as well as proposing solutions for strengthening academic achievement (Lew and Schmidt, 2011).

My name is Busra Ecem, I have a university background in Statistics. During my university education, I learned to plan and manage my time well as I was working part-time while completing my education. At the same time, I developed myself in SPSS and analytical thinking by providing statistical support to the projects of medical school students.

After graduating, I first worked as a CRM assistant specialist for 2 years at Adeka Pharmaceutical Company. In this position, I developed myself in the field of Excel and analytical thinking in the field of marketing.

After working at Adeka Pharmaceutical Company for 2 years, I started working at Akbank as a campaign management assistant. While working at Akbank, I developed myself in SQL and SAS programs. While working here, I wanted to analyze the effect of the words in the subject in email submissions on the number of reads and clicks. While working on this project, I applied classical statistical methods. As a result of the study, I analyzed that personalized texts attract more attention.

After working at Akbank for 2 years, I aimed to develop myself further in statistics by doing a master's degree and started my master's degree in Biostatistics at Ege University. In my master's degree, I worked on a project comparing classical statistical methods and machine learning techniques in survival analysis. In the study, I studied random forest survival analysis and coz regression analysis. During this master's degree, I developed myself in the R programming language. At the beginning of the research project, it was hard for me to do literature research, but I developed myselft in this field. In addition, I can mention that, the risks are so important. Because in the middle of my research, I did not get a data from my proffesor even he said that he would give me. That is why, I had to find open source and it was so hard to get permission from data holders.

As a statistics graduate, I learned to manage my time effectively and how to use SPSS while providing statistical support to medical school students. Working as a CRM assistant specialist at Adeka Pharmaceutical Company, I improved my analytical skills in Excel and SQL. My role as a campaign management assistant at Akbank enabled me to enhance my programming skills in SQL and SAS. While working on a project to analyze the impact of personalized texts in the email subject lines on engagement, I applied classical statistical methods. Pursuing a master's degree in Biostatistics at Ege University helped me develop my knowledge of R programming, Machine Learning and literature research. However, I also discovered the value of risk management as I struggled to find the data I needed for my research assignment.

I have made efforts to develop my abilities in order to address these areas for improvement. For instance, I have taken classes to hone my abilities in literature research, and I've looked for more chances to obtain knowledge in data administration and analysis.

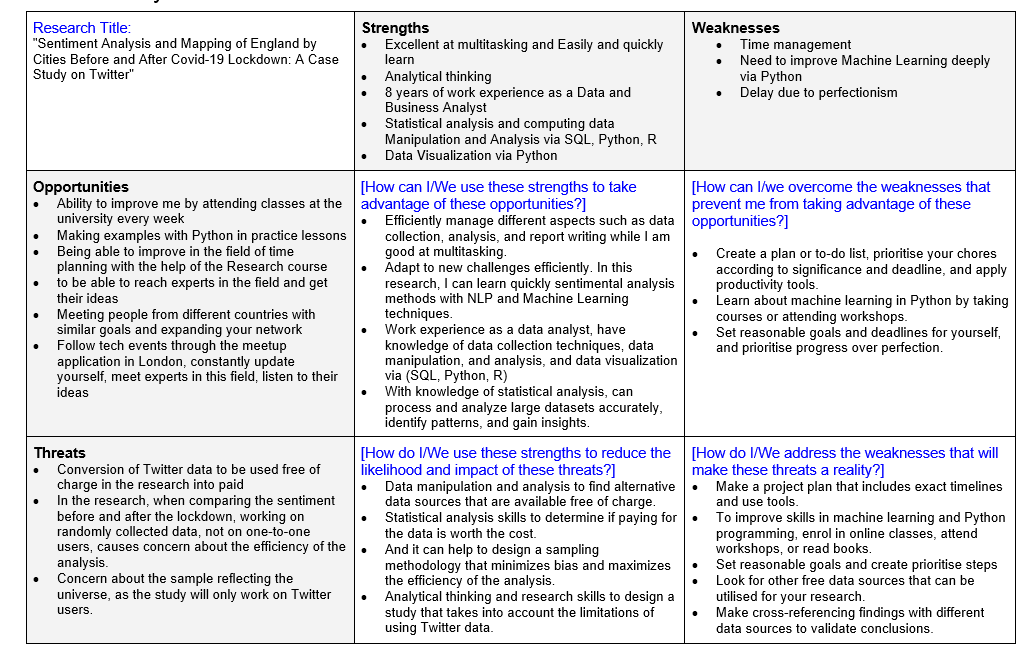
In the end, self-reflection has allowed me to identify my areas of strength and need for improvement. I have stated, for instance, that I do well at analysis, particularly when it comes to statistics and programming languages like R, SAS, and SQL. However, I've also come to the realisation that I need to strengthen both my capacity for conducting literature searches and my ability to handle unforeseen challenges like issues with data accessibility.

In general, I think that self-reflection is an essential part of developing professionally. I can continue to develop my skills and meet my professional objectives by constantly evaluating my strengths and potential improvement areas.

## 1.2 Personal SWOT analysis

SWOT analysis (Strengths, Weaknesses, Opportunities, and Threats) is an important paradigm in business, marketing, and strategic planning. SWOT analysis (Strengths, Weaknesses, Opportunities, Threats) is a well-known strategic planning tool that assists both individuals and organisations to determine their own strengths and weaknesses, as also threats and opportunities from the outside, and then creating a strategy (Namugenyi et al., 2019) Table 2 contains the SWOT analysis that created by researcher for the research title.

**Table 1. SWOT Anaysis about Research**

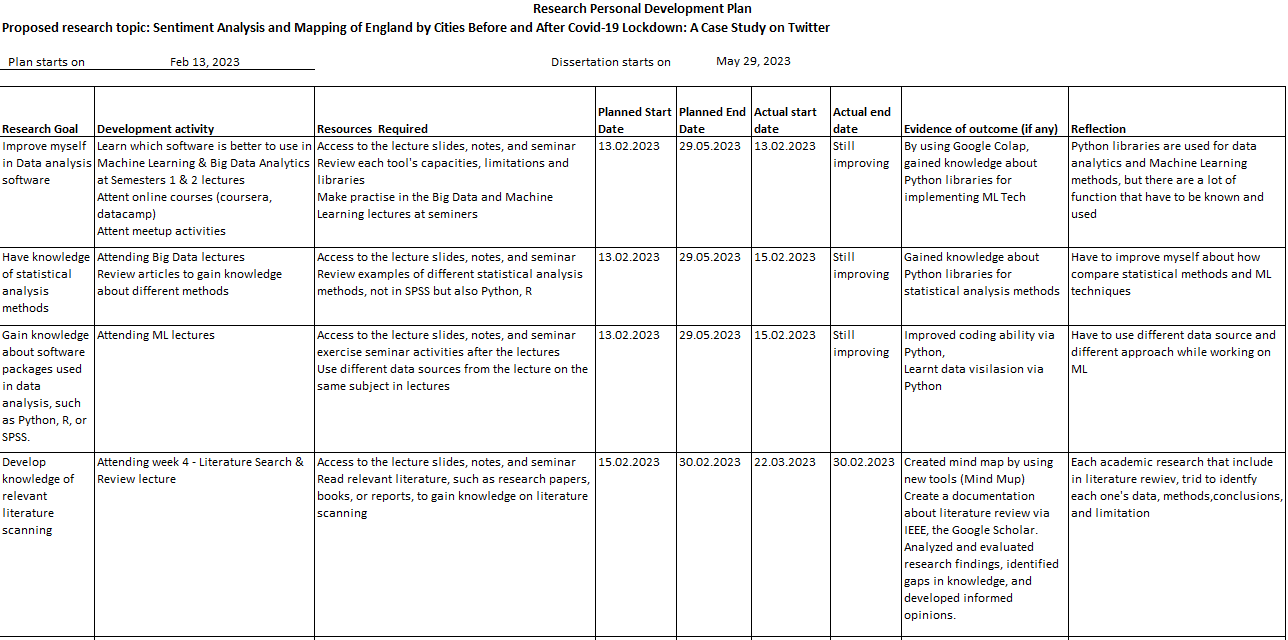


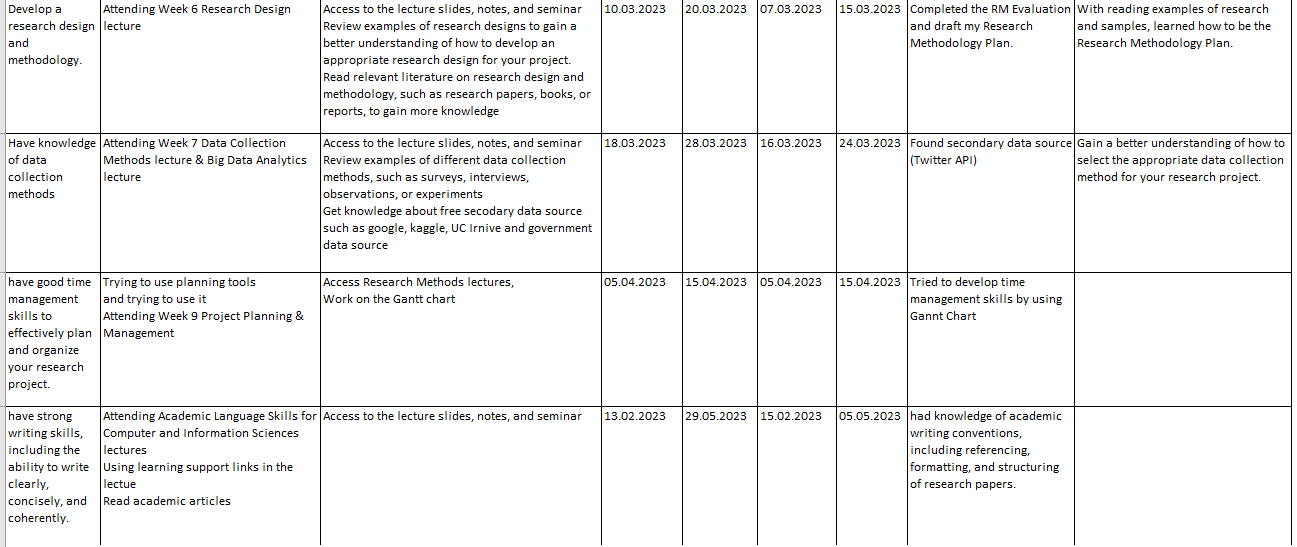
The study "Sentiment Analysis and Mapping of England by Cities Before and After Covid-19 Lockdown: A Case Study on Twitter" looks at how Twitter users felt before and after the Covid-19 lockdown in England. Excellent multitasking abilities, analytical thinking, and professional expertise as a data and business analyst are just a few of the researcher's strengths. The researcher is also proficient in data management, visualisation, and statistical analysis utilising SQL, Python, and R. The researcher has admitted that she has several weaknesses including poor time management and the need to develop her machine learning abilities. The study offers the researcher the chance to attend college lectures, take part in hands-on training, and connect with professionals in the field. The researcher has also noted potential dangers, such as the transformation of Twitter data into a paid service and worries about how accurately the sample represents the entire population. The researcher suggests using analytical thinking and research techniques, establishing realistic goals, and cross-referencing findings with many data sources to validate conclusions in order to overcome these challenges.

## 1.3 Research personal development plan

Personal Development Planning (PDP) is the sole regulatory approach for studying in UK higher education that is actively encouraged. It is an organized and encouraged method that a learner uses to gain insight into how they have learned, performed, and succeeded and to plan for their professional, academic, and personal development life (Jackson, 2010). In Table 2, the topics that the researcher wants to improve herself before academic research and what she has to do for them are presented as a PDP.

**Table 2. Research Personal Development Plan**

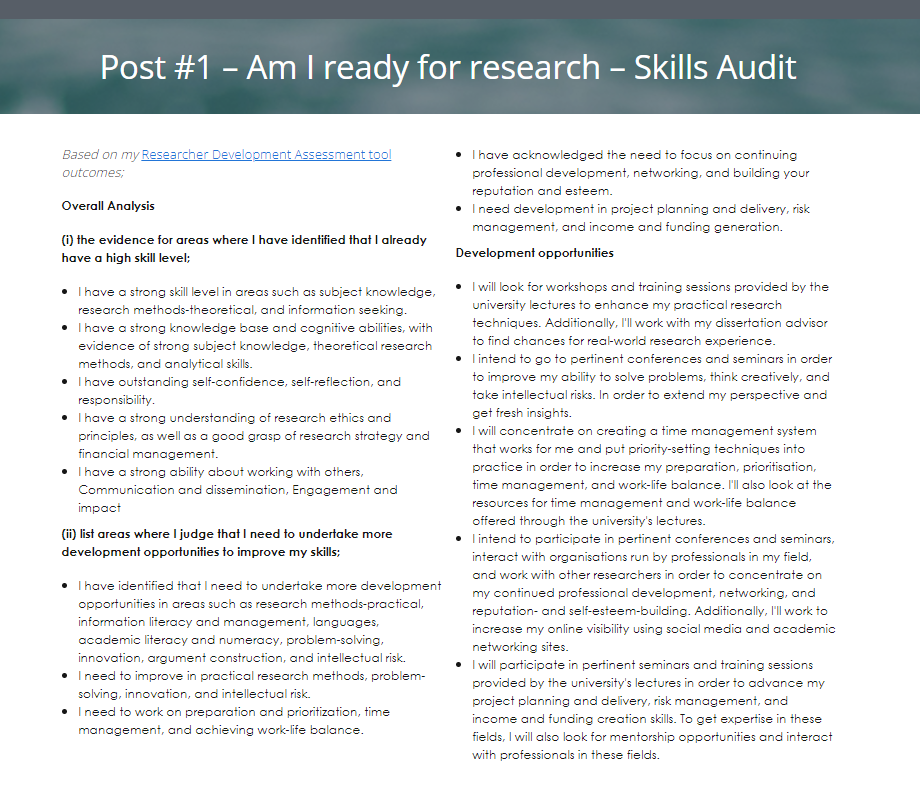




# Part 2 – Research Logs

## Skill Audit Tools

The Researcher Development Framework (RDF) is a significant new approach to researcher development that will improve our ability to develop the UK workforce, produce world-class scholars, and expand our research base and the RDF is a framework for academic growth that describes the skills, attitudes, and qualities of successful researchers and motivates them to reach their full potential (The Vitae, 2017)

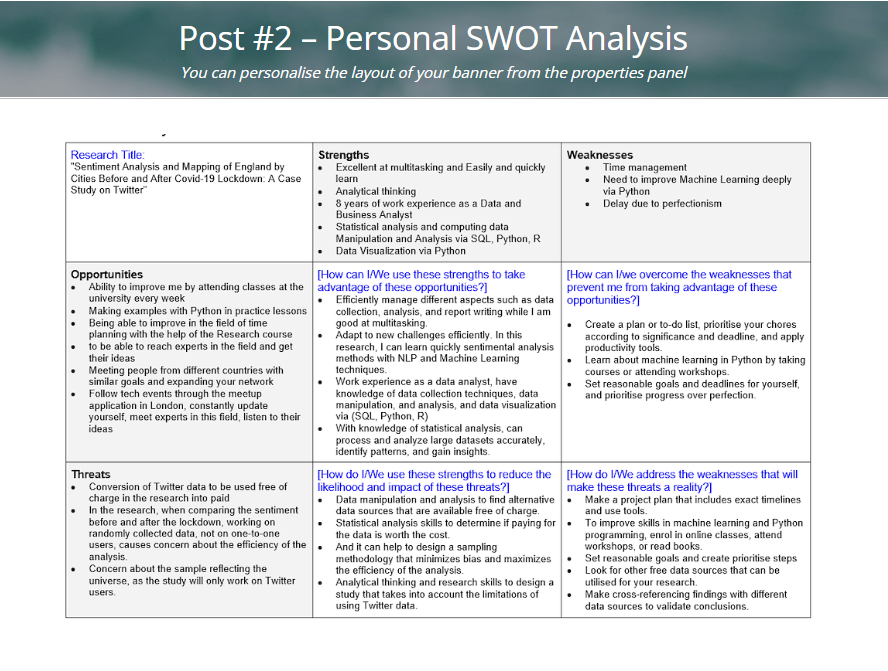


**Figure 1 Researcher Development Assessment tool, Overall Anaysis**

For more details please flow the link: [Post #1 – Am I ready for research – Skills Audit](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666ts45Hm5Mhn8gk7h)

## Personal SWOT Analysis

A personal SWOT analysis can be a beneficial tool for academic research because it can reveal strengths, weaknesses, opportunities, and threats in connection to study goals. By doing this, researchers can better plan their research operations and gain a more thorough understanding of research techniques and information.

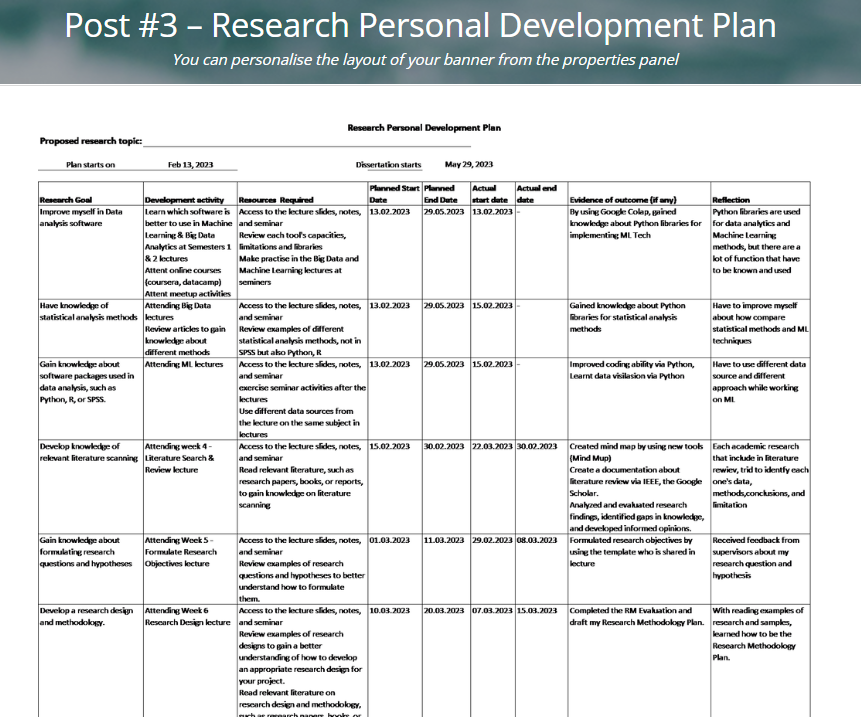


**Figure 2 Personal SWOT Analysis**

For more details please flow the link: [Post #2 – Personal SWOT Analysis](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666ts5s6y3s5y4mhnc)

## Research Personal Development Plan

Academic researchers can define specific goals, pinpoint areas for growth, and plan steps to reach these goals with the aid of a research personal development plan (RPDP). It offers a methodical way to acquire the competences, skills, and information essential for productive research.

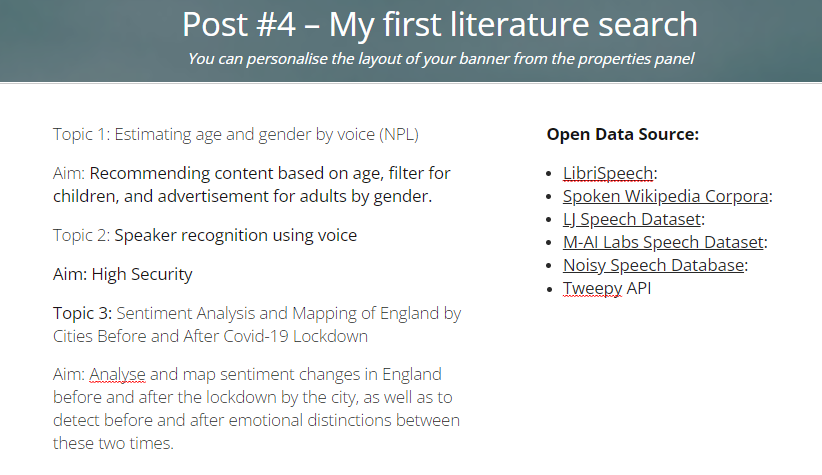
****

**Figure 3 Research Personal Development Plan**

For more details please flow the link: [Post #3 – Research Personal Development Plan](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666ts66kbr8hHRkdHW)

## My first literature search

To propose content based on age, screen out youngsters, and advertise to adults based on gender, Topic 1 aims to estimate age and gender by voice. This might be done by using open data sources like Spoken Wikipedia Corpora and LibriSpeech. Speaker recognition utilising voice to achieve high security is the objective of Topic 2, and data sources like the LJ Speech Dataset and M-AI Labs Speech Dataset may be helpful in pursuing this goal. The third topic, "Sentiment Changes in England by City Before and After COVID-19 Lockdown," uses open data sources including the Tweepy API and the Noisy Speech Database to map and analyse sentiment changes in England by city before and after COVID-19 lockdown. This analysis aims to identify emotional differences between the two timeframes.

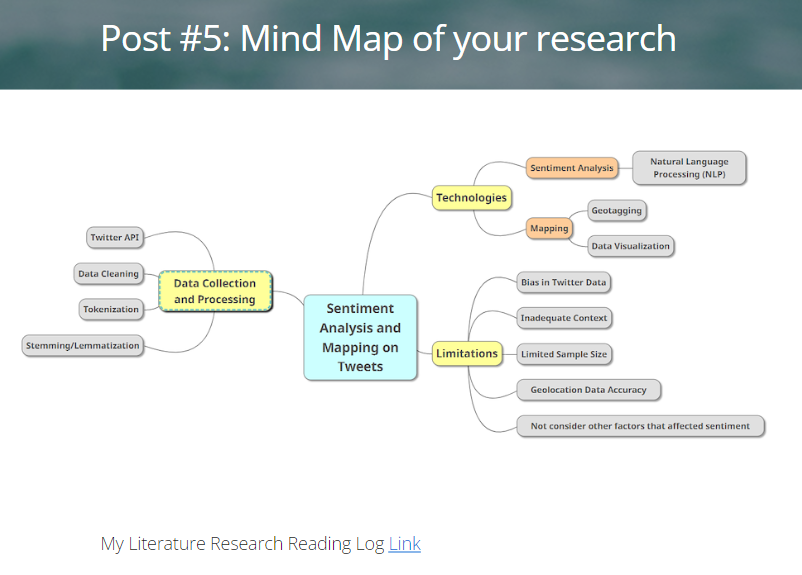


**Figure 4 My first literature search**

For more details please flow the link: [Post #4 – My first literature search](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666tsrj6rRbpGmg7fc)

## Mind Map of My research

After choosing "Sentiment Changes in England by City Before and After COVID-19 Lockdown", the researcher created a Mind Map of research. And while doing the literature review (the documentation linked into the log), the researcher decided to use Twitter API for sentiment analyses.

****

**Figure 5 Mind Map of My Research**

For more details please flow the link: [Post #5: Mind Map of your research](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666ts4w4qsRsxpc89h)

## Draft my project title, aims, research questions and objectives

After the literature review, while determining the project title, the researcher also determined its objectives, research questions and objectives.

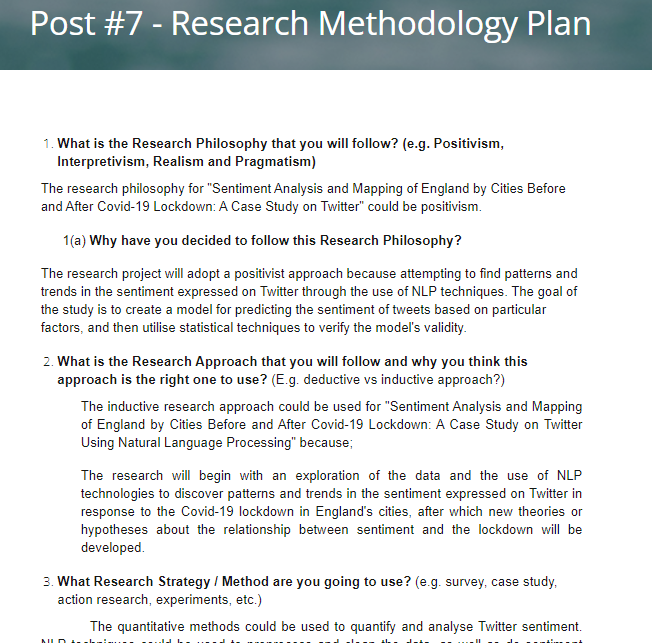
****

**Figure 6 Draft my project title, aims, research questions and objectives**

For more details please flow the link: [Post #6: Draft your project title, aims, research questions and objectives.](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666tzdpf79j8Mrrgzh)

## Research Methodology Plan

The researcher’s methodology plan includes research design, Data collection methods, Data Analysis, and more detail please follow the link shared below the figure 7.

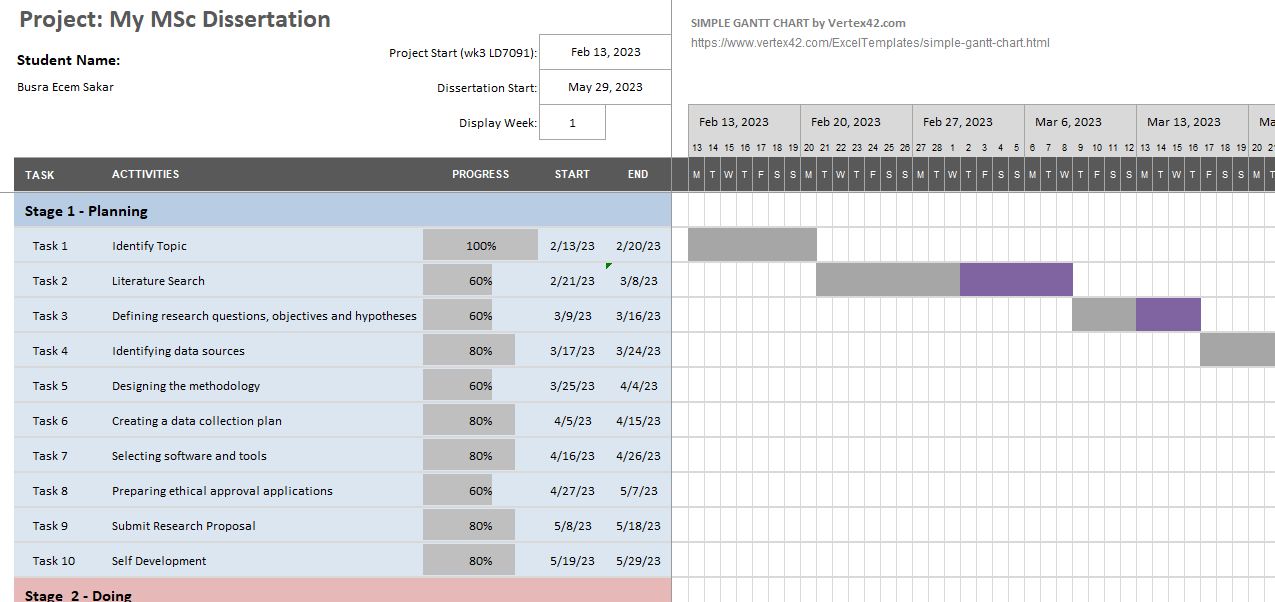


**Figure 7 Research Methodology Plan**

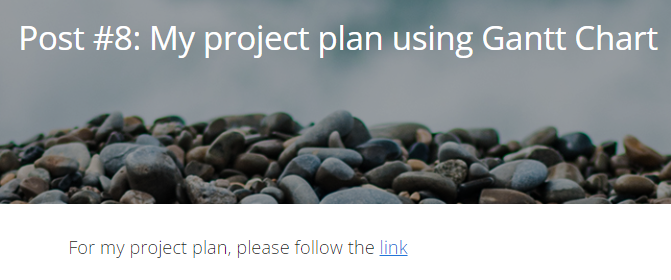
For more details please flow the link: [Post #7 - Research Methodology Plan](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666tsc4tdMdMjRGmZr)

## My project plan using Gantt Chart

Gantt chart is a diagram that shows the workflow breakdown of a project on the rows and the timeline for finishing the project's activities and phases on the columns. To reach the Gannt chart for my research, please follow the link below the figure.



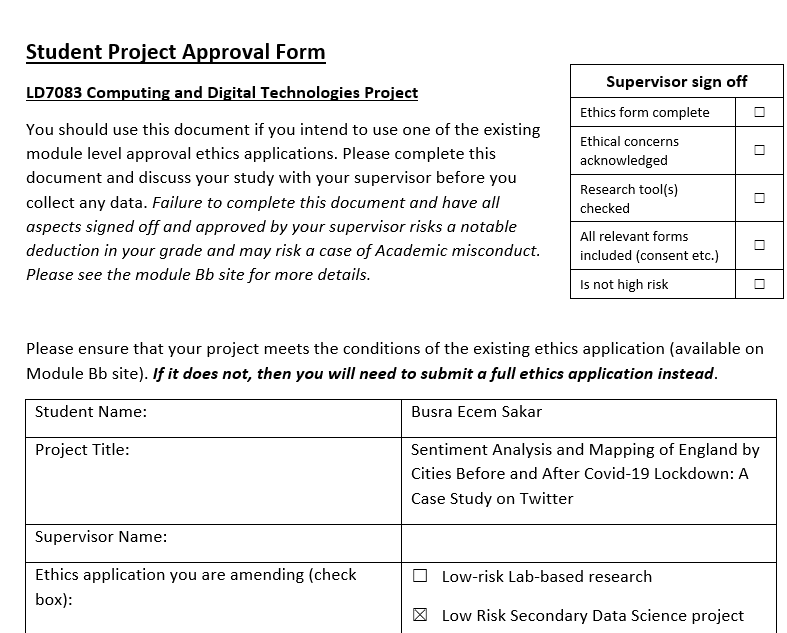
**Figure 8 Gantt Chart**

****

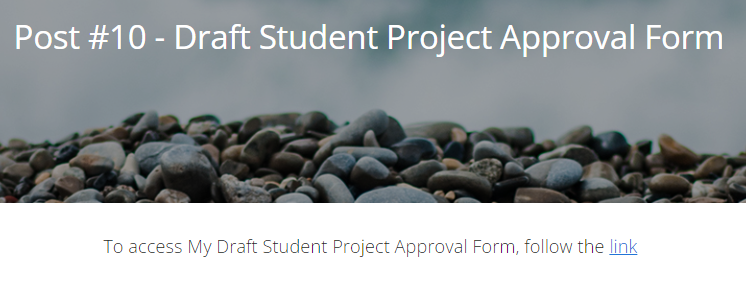
**Figure 9 My project plan using Gantt Chart**

For more details please flow the link: [Post #8: Your project plan using Gantt](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666tsssfpbM5qRG5Mr)

## Draft Student Project Approval Form

****

**Figure 10 Student Project Approval Form**

****

**Figure 11 Draft Student Project Approval Form**

For more details please flow the link: [Post #10 - Draft Student Project Approval Form](https://v3.pebblepad.co.uk/spa/#/public/xkHk3kkR666tzr8jmbMMm398bh)

# Part 3 – Research Proposal

## 3.1. Introduction

The Centres for Disease Control and Prevention launched its Emergency Operations Centre on January 21, 2020, and the World Health Organisation issued its first situation report on the coronavirus illness 2019, caused by the novel coronavirus SARS-CoV-2 (Banks and Xu, 2020). On March 11, 2020, the World Health Organisation announced the newly identified coronavirus outbreak as a global pandemic. The Covid-19 epidemic has caused vast changes in human behaviour and has interrupted daily routines throughout the world. People's actions and emotions have been altered as a result of the lockdown measures contributed forward in order to prevent the virus's spread (Rajkumar, 2020).

Twitter, a major social media platform, was significantly utilised during the pandemic to express people's views and emotions. This research's aim is to conduct a sentiment analysis of tweets written before and after the Covid-19 lockdown in England. And also intend to map sentiments by the cities in order to detect regional disparities in the emotional expression on Twitter. Natural Language Processing (NLP) techniques will be applied to evaluate tweets and generate insights into people's feelings prior to and after the lockdown (Xue et al.,2020) This research will help politicians, entrepreneurs, and health and social scientists understand the emotional impact of the epidemic and lockdown and devise appropriate strategies to support them.

### **3.1.1 Project Title**

This research's title would be "Sentiment Analysis and Mapping of England by Cities Before and After Covid-19 Lockdown: A Case Study on Twitter", and will implement quantitative methods in order to measure and analyse sentiment expressed on Twitter in a systematic and objective manner.

### **3.1.2 Background**

The background for this research is the Covid-19 pandemic, which has been affecting the world since its emergence in early 2020. To combat the epidemic, several interventions including as lockdowns and social distancing measures have been implemented. People's life, especially their mental health and well-being, have been significantly impacted by the measures and people have used social media channels, particularly Twitter, to convey their ideas, feelings, and experiences during the pandemic (Qi, Shabrina, 2023). As a result, sentiment analysis of Twitter data can provide insights into people's sentiments before and after lockdowns in various places across England.

### **3.1.3 Aim**

The aim of this project is to analyse and map sentiment changes in England before and after the lockdown by the city, as well as to detect before and after emotional distinctions between these two times.

### **3.1.4 Research Questions**

The research questions would be these;

1. How has the sentiment of tweets from various English cities evolved before and after the Covid-19 lockdown?
2. What are the differences in emotion in English cities before and after the Covid-19 lockdown?

### **3.1.5 SMART Objectives**

The SMART objectives for the research question "Sentiment Analysis and Mapping of England by Cities Before and After Covid-19 Lockdown: A Case Study on Twitter" could be:

The SMART objective for this research question is;

Specific: To analyze the sentiment of tweets about England by city before and after the Covid-19 lockdown, using NLP techniques like NLTK and spaCy (Smelyakov et al., 2015).

Measurable: To gather 2,000 tweets from each city in England before and after the Covid-19 lockdown using the Tweepy Twitter API.

Achievable: To complete data collection and analysis within 2 months.

Relevant: to provide insights into changes in sentiment in English cities before and after the Covid-19 lockdown, which policymakers, entrepreneurs, and health and social scientists can use.

Time-bound: To prepare and submit the study paper for publication within 4 months after the completion of data collection and analysis.

### **3.1.6 Research Scope**

This study's research scope aims to undertake sentiment analysis and mapping of England by city before and after the Covid-19 lockdown using Twitter data and natural language processing techniques. The project will analyse the sentiment of tweets collected from chosen towns in England to discover changes in public mood and opinion before and after the lockdown. In addition, mapping techniques will be used to visualise the results and discover any spatial trends in the data. The research will take place before and after the Covid-19 lockdown in England.

## 3.2. Literature Review

The COVID-19 pandemic has had a significant impact on people, communities, and the economy all across the world. To prevent the spread of the virus, the government of England enforced a nationwide lockdown in March 2020, which radically impacted people's lives and activities. The lockdown also had an impact on people's emotions and feelings, which were mirrored on social media platforms such as Twitter (Kausar et al., 2021).

The popularity of sentiment analysis on social media has grown to determine public sentiment in events that impact society deeply with natural catastrophes, political campaigns, and health emergencies such as the COVID-19 pandemic. Several researchers have worked on sentiment analysis of Twits which are an enormous source of real information. In addition to this, tweets are open-source data that are reachable by using Tweeter API for focus content created by users.

Gupta et al. (2021) used Twitter data by using Tweepy API to analyse sentiment during the COVID-19 lockdown in India. A total of 12.741 tweets with the keywords "Indialockdown" were extracted from the dataset gathered between April 5 and April 17, 2020. TextBlob and VADER sentiment analysis Python modules were used to annotate the data, and the natural language tool kit offered by Python was used to preprocess it. Researchers used eight supervised machine-learning algorithms to analyse the given data (Multinomial NaiveBayes, Bernoulli NaiveBayes, LogisticRegression, LinearSVC, AdaBoostClassifier, RidgeClassifier, PassiveAggressiveClassifier, and Perceptron). The linearSVC classifier and unigram performed the best, with an accuracy of 84.4%. According to the study, sentiment analysis of public lockdown tweets using this combination revealed that over half of the population (48.69%) is positive about the lockdown, 29.81% are neutral, and 21.5% are negative.

Another work by Darad and Krishnan (2023) uses deep learning and machine learning models to analyse the sentiment of COVID-19-related tweets that were tweeted during the peak of COVID-19 in April 2021. From the 16th of April 2021 to the 26th of April 2021, there were 2,00,000 tweets data acquired from Twitter using its API and tweeps the dates between the 16th and 26th of April 2021. In addition, the researcher worked on a labelled sentiment analysis dataset that was collected for tweets by the blog (Gujral,2021). Sentiment analysis was performed using a deep learning model known as Bidirectional Encoder Representations from Transformers (BERT) and various ML models for text analysis and performance (Nave Bayes, Logistic Regression, Random Forest, Support Vector Machines, Stochastic Gradient Descent, and Extreme Gradient Boosting), which were then compared. In order to identify the sentiment of aggregated tweets, the NB classifier was trained on NLTK (Natural Language Tool Kit). Classical ML approaches had an acceptable accuracy rate of 70%, however, the deep learning model using BERT had a remarkable accuracy rate of 84.2%. According to the researcher, the majority of people who tweet take both positive and neutral views.

Ahmad, Jun (2021) focused on understanding the sentiment of cancer patients regarding their treatment during the peak of the COVID-19 pandemic and tried to investigate the potential of natural language processing (NLP) in healthcare management. The dataset was gathered using "Twitterscraper" command line tool to retrieve the 150000 tweets (after cleaning 42000) related to coronavirus and cancer, published from Canada between January 1 and April 30, 2020, and specified the period and location in the Twitterscaper before downloading the tweets. The "Twitterscraper" retrieves public tweets uploaded from that address within a given time frame. Researchers conducted the following 15 searches (1.'Covid-19 and cancer'). 2.'Cancer and Coronavirus' 'Covid-19 and Oncology,' 3. 4.'Oncology and Coronaviruses' 'Cancer and Surgery,' 5th. 'Cancer and Radiation Therapy' 'Coronavirus with radiation,' page 7 'Coronavirus and Cancer Therapy' 'Cancer therapy consultation' 'Chronic illnesses and COVID-19,' page 10 'Chronic illness and coronaviruses,' page 11 'Coronavirus and cancer surgery,' 12th 'Chemotherapy and covid-19,' page 13 'Chemotherapy appointment has been cancelled.' 15.'Radiotherapy appointment cancelled.') to look for the relevant tweets. The researchers discovered that interaction should be improved between cancer patients and healthcare professionals during the pandemic to decrease fear and anxiety while providing necessary support based on three criteria, 52% positive sentiment tweets, tweet polarity and subjectivity distribution, and words cloud of tweets demonstrate that traditional cancer treatments have not been significantly altered from the patients' perspective.

According to the review of the literature, the potential gap in research on the topic "Sentiment Analysis and Mapping of England by Cities Before and After Covid-19 Lockdown: A Case Study on Twitter" could be the lack of research studies that specifically focus on the during Covid-19 pandemic time frame. This research aim will be to analyse the sentiment of Twitter users in different cities of England before and after the Covid-19 lockdown. While several researchers have explored sentiment analysis on Twitter during the Covid-19 pandemic, they either focus on a different geographical region or on the Covid-19 pandemic time period. This research may have a beneficial effect by giving an understanding of how the sentiment of Twitter users changed in various cities in England has altered as a result of the Covid-19 pandemic and the lockdown that followed. This information could assist policymakers, health institutions, and businesses in better understanding the consequences of the epidemic on people's attitudes and behaviours. Furthermore, this study could serve as an empirical foundation for future research in this sector, potentially leading to more specialised treatment choices and legislation focused at addressing the mental and physical well-being of persons afflicted by the global epidemic.

## 3.3. Research Methodology and Planning

### **3.3.1 Research Methodology**

The research project will adopt a positivist approach because attempting to find patterns and trends in the sentiment expressed on Twitter through the use of NLP techniques. The goal of the study is to create a model for predicting the sentiment of tweets based on particular factors, and then utilise statistical techniques to verify the model's validity.

The inductive research approach could be used because; the research will begin with an exploration of the data and the use of NLP technologies to discover patterns and trends in the sentiment expressed on Twitter in response to the Covid-19 lockdown in England's cities, after which new theories or hypotheses about the relationship between sentiment and the lockdown will be developed.

The research will implement quantitative methods in order to measure and analyse sentiment expressed on Twitter in a systematic and objective manner. The practice of finding and extracting views, emotions, and attitudes conveyed in text data is known as sentiment analysis. This study could use quantitative methods to apply algorithms and mathematical models to process enormous amounts of data efficiently and accurately.

Natural Language Processing (NLP) techniques are used in sentiment analysis to preprocess and sanitise data before doing sentiment analysis and topic modelling (Chong et al., 2014). These tools can aid in identifying patterns and trends in Twitter sentiment regarding the Covid-19 shutdown in England's cities. The use of quantitative approaches can aid in providing empirical proof for the researcher's conclusions as well as increasing the reliability and reliability of the research's findings.

Finally, sentiment mapping could enable the visualisation and analysis of location-based information. The analysis could uncover geographic patterns and trends in the sentiment expressed on Twitter about the Covid-19 lockdown in England by mapping the sentiment scores across different locations of England.

### **3.3.2 Research Plan and Task Lists**

The researcher create a Gannt Chart and will follow the steps created in this chart. Gantt chart is a diagram that shows the workflow breakdown of a project on the rows and the timeline for finishing the project's activities and phases on the columns, follow the [link](file:///C:\Users\info\Desktop\Big%20Data%20and%20Data%20Science%20Tech\Semester%202\Research%20Methods%20for%20Professional%20Practice\Gantt%20Chart%20for%20my%20Research.xlsx) to reach this research’s Gantt chart.

Overall the study will be carried out in the following stages:

**Literature review:** A complete study of relevant research will be conducted at this stage to acquire knowledge about recent studies on sentiment analysis, mapping, and natural language processing techniques with social media platforms.

**Data collection:** The Python Tweepy API will be used to obtain data from Twitter for the purpose of the research. User information, geolocation, timestamps, and other specific metadata will include tweets sent previous to and following the lockdown in England, and it will also include tweets from particular cities in England.

According to Twitter's developer documentation's instruction on analysing the sentiment of your own tweets (Twitter, n.d.), the steps that follow can be taken to use Twitter API for sentiment analysis:

*Apply for a Twitter Developer account:*To access the Twitter API, the researcher will create a developer account on the Twitter Developer page.

*Create a New Project and App:*The researcher will establish a new Twitter App once have a developer account by filling out a form with information about the project.

*Generate API Key and Secret:*Following the creation of an app, a set of API keys, comprising a consumer key and a secret key, will be provided. These keys will be used to authenticate the researcher's Twitter API calls.

*Choose API Endpoint:*Twitter API supports multiple endpoints for various activities such as retrieving tweets, searching for tweets, obtaining user information, and so on. The researcher will select an endpoint based on the demands of the investigation.

*Write code:*The researcher will write code in a programming language like Python to make calls to the Twitter API using the API keys generated, and will use libraries like Tweepy for Python to make the process easier.

*Analyze and Visualize Twitter Data:* Once the data has been obtained, the researcher will analyse it using sentiment analysis or topic modelling methods. And also will use charts, graphs, and other visualisation tools to visualise the data.

**Preprocessing of obtained data:** The data gathered will be extracted by eliminating any

unrelated data and sanitising the data for analysis.

**Sentiment analysis:** It will be carried out with NLP techniques to identify the sentiment reflects in tweets.

**Sentiment mapping:** The research project will apply Python to map emotions expressed in tweets to several cities in England.

**Data analysis:** The obtained data will be analysed using statistical analysis methods and Machine Learning techniques to uncover patterns and trends in the sentiment expressed in tweets before and following the lockdown in various places across England.

**Reporting and presentation:** The results of the research will be reported and provided in a clear and straightforward manner, highlighting major results and conclusions.

A dataset will be collected in this study utilising the Python Twitter API. **Steps that have to be followed to achieve the aim are:**

**1.** Using the Tweepy API, create a tweet dataset comprising tweets from Twitter that were tweeted within England in two periods (before and after the lockdown), by location.

**2.** Classifying the sentiment of each tweet in these two datasets by city.

**3.** Prepare the dataset for data analysis.

**4.** Examining the dataset's sentiment and mapping.

### **3.3.3 Risk Assessment and Management**

Some of the potential risks are:

Data collection: Twitter data to be used for free, in case of it could charge, data manipulation and analysis to find alternative data sources that are available free of charge.

Bias in data: Because the data collected may not be an accurate representation of the population, there is a potential for bias. To reduce this risk, the project will collect a significant number of tweets from a varied spectrum of individuals in various places across England. And a sampling methodology can help to design that minimizes bias and maximizes the efficiency of the analysis.

Concerns about privacy and ethics: The usage of Twitter data creates privacy and ethical concerns. To reduce this potential risk, the research project will guarantee that all data obtained is anonymized, and ethical concerns will be addressed.

Technical concerns: such as software failure, internet access challenges, and data loss, are possible. To mitigate this danger, the study will use dependable software and hardware and perform regular data backups.

Issues with data quality: There is a possibility that the accuracy of the data collected will be insufficient to draw significant conclusions. To mitigate this risk, the study will use trustworthy and valid methods to clean, process, and analyse the data.

## 3.4. Conclusions

In conclusion, the study will present an analysis of sentiments expressed on Twitter by users in different cities across England before and after the COVID-19 lockdown. The research will implement NLP techniques to preprocess and sanitise the data. After this process, the research will use statistical and machine learning methods. Research findings will be presented by mapping across England.

As the COVID-19 pandemic continues to affect people's lives, social media platforms such as Twitter remain essential channels for people to express their emotions and experiences. Understanding individuals' opinions and carrying them out in different places will give useful insights into how individuals deal with the crisis.

The experiments discovered that the security precautions in lockdown had a negative impact on community sentiments, and it has an important rise in negative feelings such as worry, despair, and irritation. This researh take this analysis further and try to analyze before and after lockdown differences in individual sentiment.

Overall, this study highlights the significance of sentiment analysis and mapping in understanding the impact of significant events such as the COVID-19 pandemic on people's emotions and behaviour. By employing advanced techniques such as NLP, researchers can gain valuable insights into user sentiment and behaviour that can inform decision-making and contribute to improving people's lives. This study’s result creates an impact on policymakers, public health officials, and marketers, among others. Policymakers can make more informed decisions about public health measures and resource allocation if they understand the people's emotional condition throughout the epidemic and how it has altered over time and across different areas. And they can give psychological support in these areas. Marketers may use this data to refine their marketing tactics.

**3.4.1 Ethics, Legal, Social, Security and Professional Consideration**

The use of Twitter data raises concerns about privacy and ethical considerations. The researcher will be certain that any personal information received from participants is kept private and secure. When collecting and analysing Twitter data, the researcher should follow Twitter's terms of service. This includes requesting authorization from Twitter before collecting data and according to their API usage standards (Twitter, n.d).

*Ethics Consideration;*

To preserve participants' privacy, the researcher will remove any identifying information from the data collection or will replace it with anonymous identities. In this study, getting informed consent via a tweet can be challenging because users might not react or might no longer be maintaining their account. Additionally, since Twitter profiles and tweets can be set by default to be publicly visible, Twitter could be regarded as a public space. (Ahmed, et al., 2017). After the study, the researcher will delete the data from all sources to preserve the privacy of the individuals.

*Legal Consideration;*

Twitter's Twitter Privacy Policy agreement, users consent for their information to be collected and used by third parties (Twitter, 3.1). For example, the privacy policy notes that:

"Through our APIs. We use technology like APIs and embeds to make public Twitter information available to websites, apps, and others for their use, for example, displaying Tweets on a news website or analyzing what people say on Twitter. We generally make this content available in limited quantities for free and charge licensing fees for large-scale access. We have standard terms that govern how this information can be used, and a compliance program to enforce these terms. But these individuals and companies are not affiliated with Twitter, and their offerings may not reflect updates you make on Twitter.”

*Social Consideration;*

The study's authors will make sure that it does not discriminate against, offend, or hold prejudices towards any specific group of people, and caution should be exercised to prevent the spread of any inaccurate or misleading material that might be harmful to people or communities.

*Security* *Consideration;*

Data should be stored on password-protected devices or servers, and only researchers and leaders can have access to it. To prevent unauthorised access or exposure, the researchers will make sure that the Twitter data is stored and handled securely. Any sensitive or personal data will be gathered, safeguarded, and managed in line with the relevant laws and regulations.

*Personal Consideration;*

Without misrepresenting or manipulating the data, the researcher will transparently provide the findings. The researcher will abide by Twitter's API rules, which include limitations on data usage, sharing, and dissemination, when using the Twitter API for the study. Additionally, the researchers will ensure that the study complies with all pertinent laws and policies and obtain the necessary approvals and permissions from Twitter before using the API.

## Appendix: Draft Student Project Approval Form

**Student Project Approval Form**

|  |  |
| --- | --- |
| **Supervisor sign off** | |
| Ethics form complete |  |
| Ethical concerns acknowledged |  |
| Research tool(s) checked |  |
| All relevant forms included (consent etc.) |  |
| Is not high risk |  |

**LD7083 Computing and Digital Technologies Project**

You should use this document if you intend to use one of the existing module level approval ethics applications. Please complete this document and discuss your study with your supervisor before you collect any data. *Failure to complete this document and have all aspects signed off and approved by your supervisor risks a notable deduction in your grade and may risk a case of Academic misconduct. Please see the module Bb site for more details.*

Please ensure that your project meets the conditions of the existing ethics application (available on Module Bb site). ***If it does not, then you will need to submit a full ethics application instead***.

|  |  |
| --- | --- |
| Student Name: | Busra Ecem Sakar |
| Project Title: | Sentiment Analysis and Mapping of England by Cities Before and After Covid-19 Lockdown: A Case Study on Twitter |
| Supervisor Name: |  |
| Ethics application you are amending (check box): | Low-risk Lab-based research  Low Risk Secondary Data Science project  Medium Risk Secondary Data Science project from the private domain required membership  Questionnaire/ survey Study  Interview Study or other Usability Study |

**Introduction to the project:** *Treat it like an introduction to the study. Why is your proposed study important? What has already been done on the topic? How does your proposed study ‘fit’ with the current literature and what does it add? What is the aim of the proposed study? Make reference to appropriate studies.*

|  |
| --- |
| The purpose of the study is to analyse the sentiment of tweets sent before and after the COVID-19 lockdown in England. The study aims to analyse tweets and produce insights into people's emotions both before and after the lockdown using Natural Language Processing (NLP) techniques. In order to identify geographical differences in emotional expression on Twitter, the study also seeks to visualise feelings by cities. The planned study is crucial because it lets policymakers, businesspeople, health experts, and social scientists recognise the emotional effects of the epidemic and lockdown and develop supportive solutions for them.  The study also makes reference to earlier investigations that examined tweets about COVID-19 using sentiment analysis. Twitter data was used by Gupta et al. (2021) to study public opinion during the COVID-19 lockdown in India. Eight supervised machine-learning algorithms were employed to extract 12,741 tweets with the keyword "Indialockdown" from the dataset collected between April 5 and April 17, 2020. With an accuracy of 84.4%, the linearSVC classifier and unigram performed best. According to the study, 48.69% of people felt positive about the lockdown, 29.81% felt neutral, and 21.5% felt negative.  During the peak of COVID-19 in April 2021, Darad and Krishnan (2023) used deep learning and machine learning models to analyse the sentiment of tweets related to COVID-19. Bidirectional Encoder Representations from Transformers (BERT), a deep learning model for text analysis and performance, was used for sentiment analysis of 200,000 tweets that were obtained from Twitter utilising its API and tweeps. The BERT-based deep learning model obtained an astounding accuracy rate of 84.2%. The majority of Twitter users, according to the researcher, have both positive and neutral opinions.  When the COVID-19 pandemic was at its worst, Ahmad, Jun (2021), concentrated on understanding cancer patients' attitudes towards their treatments and attempted to look into the potential of natural language processing (NLP) in healthcare administration. The dataset was created by retrieving 150000 tweets about coronavirus and cancer from Canada between January 1 and April 30, 2020, using the "Twitterscraper" command-line programme. Based on three criteria—52% positive sentiment tweets, tweet polarity and subjectivity distribution, and words clouds of tweets showing that conventional cancer treatments have not significantly changed—the study concluded that communication between cancer patients and healthcare professionals during the pandemic needs to be improved in order to reduce fear and anxiety while providing necessary support.  In conclusion, the proposed project seeks to undertake a sentiment analysis of tweets prior to and following the COVID-19 lockdown in England and apply NLP methods to produce insights into people's feelings. The study will aid in the development of effective people-supporting solutions by assisting social scientists and policymakers in comprehending the emotional effects of the pandemic and lockdown.  **References:**  Gupta, P. *et al.* (2021) “Sentiment analysis of lockdown in India during COVID-19: A case study on Twitter,” *IEEE Transactions on Computational Social Systems*, 8(4), pp. 992–1002. Available at: https://doi.org/10.1109/tcss.2020.3042446.  Darad, S. and Krishnan, S. (2023) “Sentimental analysis of COVID-19 Twitter data using Deep Learning and machine learning models,” *Ingenius*, (29), pp. 108–117. Available at: https://doi.org/10.17163/ings.n29.2023.10.  Ahmad, B. and Jun, S. (2021) “Sentiment analysis of cancer patients about their treatment during the peak time of pandemic covid-19,” *2021 4th International Conference on Computing & Information Sciences (ICCIS)* [Preprint]. Available at: https://doi.org/10.1109/iccis54243.2021.9676393. |

**Methodology:** *Please complete the table below, using the following info to guide you.**Write this as a future tense method. Describe the* ***participants*** *that you will recruit, how many you are going to recruit, and indicate if you have any additional exclusion criteria. Include the* ***research design*** *(e.g. randomised/repeated measures/quantitative/qualitative/case study etc) and detail of your proposed* ***procedures*** *(i.e., how are you collecting the data?). Include information on all of the equipment you plan to use. If this is a low-risk study, outline how you will extract data and list the criteria you will use to do this. Somebody should be able to read this and replicate it. Describe all planned* ***data analysis*** *for both quantitative (e.g. t-tests, ANOVA, correlation etc.) and qualitative (content analysis, thematic analysis etc.) data. If doing a low-risk study explain how you intend to analyse the data you have collected. Use literature to justify your method.*

|  |  |
| --- | --- |
| 1. Is this a low-risk secondary data or lab-based study? If Yes please go to questions 6 and 7. | YES  NO |
| 1. Who are your participants and what is the inclusion criteria? |  |
| 1. How many will you recruit and from where? |  |
| 1. Are there any exclusion criteria (reasons why people should not participate)? |  |
| 1. Research design: |  |
| 1. Procedures (describe what you will do to collect data, including all equipment/methods you plan to use). | A researcher must first register as a developer on the Twitter Developer page in order to access the Twitter API for research purposes. The researcher will construct a new Twitter App by filling out a form with project details after creating the account. A set of API keys, comprising a consumer key and a secret key, will be given when the app is created and used to authenticate the researcher's Twitter API calls. The researcher will next choose an API endpoint in accordance with the needs of the study and develop code in a programming language like Python utilising tools like Tweepy to facilitate the process.  After gathering the data, the researcher will analyse it using sentiment analysis or topic modelling techniques and visualise it using tools like charts and graphs. When gathering and examining Twitter data, ethical issues like privacy and consent should also be taken into account. |
| 1. Data analysis methods: | The research will implement NLP techniques to preprocess and sanitise the data. After this process, the research will use statistical and machine learning methods. Research findings will be presented by mapping across England. |
| 1. Additional information: |  |

**Health and Safety:**  *Relevant risk assessments are listed in the ethics application. You must submit a new ethics application if your project needs additional risk assessments. Please identify the elements of the listed risk assessment that are relevant to your study and the risk assessment(s) you are working with.*

Please check the relevant boxes\*:

HL\_RISK\_173 Testing in an external environment

HL\_RISK\_722 face-to-face interview

HL\_RISK\_727 Group interview

|  |  |  |
| --- | --- | --- |
| **Areas of potential risk**  *Please indicate how you will eliminate, or as a minimum ameliorate, the following areas of potential risks throughout the processes of research design, data generation, data analysis and dissemination* | | |
| **Area of risk** | **Questions relating to this risk** | **How will you mitigate against this risk?** |
| Avoiding harm to all involved in or potentially affected by the research | How will you ensure that your participants/ respondents come to no harm (psychological; emotional; physical). e.g. not subjecting them to questioning about sensitive issues without advance agreement? | This research will not come to any harm because the research will follow TwitterAPI ethics and regulations, and all the tweets will not use personal information. |
| How will you ensure your own safety (beyond just physical) in undertaking the Enquiry? | I will strong passwords, two-factor authentication, and encryption to protect my data and ensure it is not accessed by unauthorized individuals.  I will follow TwitterAPI ethical guidelines and regulations set by my research to avoid any ethical violations.  I will plan my research schedule in a way that ensures I am not overworking and have enough time for rest and relaxation.  I will seek out the support and guidance of other researchers or experts in the field to ensure I am conducting the research safely and effectively. |
| Ensuring the anonymity of all participants/respondents | How will you ensure anonymity in collecting/generating data | To protect users' identities, I won't use the participants' real names, I'll use unique, random ids. Before starting the analysis, I will delete all personal information from the data. I'll use data encryption and save all information and create it in a safe place that only authorised personnel may access. I will follow ethical guidelines when gathering data to keep participants' privacy safeguarded. |
| How will you ensure anonymity in reporting the data? | To protect users' identities, I won't use the participants' real names, I'll use unique, random ids. |
| Gaining informed consent from all participants / respondents | How will you ensure respondent/participant consent in advance? You should provide a copy of the necessary consent form/s with this document | I will use only publicly available tweets and I will remove any personal information from tweets. Also, when I gathering data, I will follow TwitterAPI ethics and regulations. |
| (How) might participants/respondents be able to withdraw their data? | I can provide public notice at the beginning of my research explaining that I will be using tweets as data for academic research purposes. I can also share an email address for individuals to request the removal of their tweets from the dataset. |
| Avoiding deception | How will you how you promote accuracy in recording, analysis, reporting of the data/findings? | I will clearly define my research question and hypotheses.  I will use reliable data sources.  I will establish clear and consistent data collection procedures that follow TwitterAPI ethics and regulations.  I will use appropriate data analysis techniques.  I will validate my findings which are robust and reliable.  I will report findings accurately and transparently and it will enable others to assess the validity and reliability of my findings. |
| Data storage and destruction | How will you transport and store your data securely (e.g. password protected; cloud storage) | I will use encrypted cloud storage services like Google Drive or Dropbox, which provide both resting and transaction encryption, to keep the data safe. I will make sure to save the passwords and use strong passwords to protect the info. For further security, I will also utilise two-factor authentication. |
| How will you destroy the data and when? | I will safely remove any files and data associated with the study, including any backups or copies, in order to eliminate the information gathered for the study. I shall delete the data in accordance with any institutional or ethical rules for data destruction, and I will fully record the procedure.  Once all of the analyses and results have been released or finalised and there are no open requests for access to the data, I will delete the data. |
| Secondary data sets | *Is your data set(s) from a domain requires membership?* | Yes |
| *Does this data set can be used for educational or academic research purpose?* | Yes |

Please check this box after you have read and understood ethics and health and safety information.

I confirm I have read the University’s health and safety policy and ethics policy. I have read and understood the requirement for the mandatory completion of risk assessments and that my study does not deviate from the module level approval ethics forms on Blackboard.

**Further information (add below, if applicable)**

* Consent forms
* Participant information sheet
* Debrief form
* Recruitment materials
* Permission letters
* Data collection tools

|  |  |
| --- | --- |
| **Student’s Name and sign**  **Busra Ecem Sakar**  **(Name)** | **Date** |
| **Supervisor’s name and sign**  **(Name)** | **Date** |

## References

Ahmad, B. and Jun, S. (2021) “Sentiment analysis of cancer patients about their treatment during the peak time of pandemic covid-19,” *2021 4th International Conference on Computing & Information Sciences (ICCIS)* [Preprint]. Available at: https://doi.org/10.1109/iccis54243.2021.9676393.

Ahmed, W., Bath, P.A. and Demartini, G., 2017. Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. *The ethics of online research*, *2*, pp.79-107.

Chong, W.Y., Selvaretnam, B. and Soon, L.-K. (2014) “Natural language processing for sentiment analysis: An exploratory analysis on Tweets,” *2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology* [Preprint]. Available at: https://doi.org/10.1109/icaiet.2014.43.

Darad, S. and Krishnan, S. (2023) “Sentimental analysis of COVID-19 Twitter data using Deep Learning and machine learning models,” *Ingenius*, (29), pp. 108–117. Available at: https://doi.org/10.17163/ings.n29.2023.10.

Gupta, I., Chatterjee, I. and Gupta, N. (2022) “Sentiment analysis of covid-19 tweets,” *2022 1st International Conference on Informatics (ICI)* [Preprint]. Available at: https://doi.org/10.1109/ici53355.2022.9786887.

Gupta, P. *et al.* (2021) “Sentiment analysis of lockdown in India during COVID-19: A case study on Twitter,” *IEEE Transactions on Computational Social Systems*, 8(4), pp. 992–1002. Available at: https://doi.org/10.1109/tcss.2020.3042446.

Jackson, N., 2010. Learning to be a self-regulating professional: the role of personal developing planning (PDP). *Learning to be professional through a Higher Education*.

Kausar, M.A., Soosaimanickam, A. and Nasar, M. (2021) “Public sentiment analysis on Twitter data during COVID-19 outbreak,” *International Journal of Advanced Computer Science and Applications*, 12(2). Available at: https://doi.org/10.14569/ijacsa.2021.0120252.

Le, Y. and Nanjun, J. (2023) “‘A study of the relationship between students’ self-regulation, academic motivation, and academic performance ",” *The Journal of Humanities and Social sciences 21*, 14(1), pp. 325–340. Available at: https://doi.org/10.22143/hss21.14.1.24.

Namugenyi, C., Nimmagadda, S.L. and Reiners, T. (2019) “Design of a SWOT analysis model and its evaluation in Diverse Digital Business Ecosystem Contexts,” *Procedia Computer Science*, 159, pp. 1145–1154. Available at: https://doi.org/10.1016/j.procs.2019.09.283.

Qi, Y. and Shabrina, Z. (2023) “Sentiment analysis using Twitter data: A Comparative Application of lexicon- and machine-learning-based approach,” *Social Network Analysis and Mining*, 13(1). Available at: https://doi.org/10.1007/s13278-023-01030-x.

Rajkumar, R.P. (2020) “Covid-19 and mental health: A review of the existing literature,” *Asian Journal of Psychiatry*, 52, p. 102066. Available at: https://doi.org/10.1016/j.ajp.2020.102066.

Smelyakov, K. *et al.* (2020) “Effectiveness of preprocessing algorithms for Natural Language Processing Applications,” *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)* [Preprint]. Available at: https://doi.org/10.1109/picst51311.2020.9467919.

The Vitae. (2017) *RDF conditions of use* (2017). The Vitae. Available at: https://www.vitae.ac.uk/researchers-professional-development/about-the-vitae-researcher-development-framework/rdf-cou (Accessed: March 30, 2023).

Twitter (n.d). *Developer agreement and policy – twitter developers | twitter developer platform* (no date) *Twitter*. Twitter. Available at: https://developer.twitter.com/en/developer-terms/agreement-and-policy (Accessed: April 23, 2023).

Twitter (n.d). *How to analyze the sentiment of your own tweets | docs | twitter developer platform* (no date) *Twitter*. Twitter. Available at: https://developer.twitter.com/en/docs/tutorials/how-to-analyze-the-sentiment-of-your-own-tweets (Accessed: April 23, 2023).

Twitter (n.d). *Twitter privacy policy* (no date) *Twitter*. Twitter. Available at: https://twitter.com/en/privacy (Accessed: April 23, 2023).

Xue, J. *et al.* (2020) “Twitter discussions and emotions about the covid-19 pandemic: Machine Learning Approach,” *Journal of Medical Internet Research*, 22(11). Available at: https://doi.org/10.2196/20550.