##### Hajj And Umrah Tweets Data Analysis

**Special Topics I**

**Group Members:**

**Khalil Alsulaimani (Group Leader) ID:439007816**

**Nawaf Noor Aldeen ID:439003406**

**Yasser Alharbi ID:439003070**

**Supervisor : Dr. Mohammed Algamidi**

Table Of Content

[Chapter 1: Introduction 3](#_Toc122034379)

[Introduction: 4](#_Toc122034380)

[Purpose Of This Document: 4](#_Toc122034381)

[Data Analysis: 4](#_Toc122034382)

[Tools: 5](#_Toc122034383)

[Chapter 2: Data Gathering 7](#_Toc122034384)

[Data Gathering: 8](#_Toc122034385)

[Method 1: Twitter API: 8](#_Toc122034386)

[References: 11](#_Toc122034387)

# Chapter 1: Introduction

## Introduction:

Social media has taken over the world in every aspect from our daily lives to business, therefore, the world has evolved and changed from the past where governments and components depended on surveys to know the public opinion of a certain subject to now depending on social media.

The advantage of this evolution is a larger sample population which will give a better more detailed public opinion, moreover, with the advancements in Artificial Intelligence ( AI ) help make the process automated with grater statistics and numbers.

However, with every evolution there are some new challenges, these vairy from social media platform, however, some common problems occur and one of the main ones is filtering out useless meaningless social media responses that don’t benefit the study.

These advantages and disadvantages are what made the new field of data analysis such an important and integral part of society as it handles all the data and gives the results researchers need and are looking for.

As for this project, we will be using data analysis on hajj and umrah to learn the techniques and process of data analysis to find the public opinion about hajj and umrah our goal is to take an initial set of one hundred thousand tweets to analysis.

## Purpose Of This Document:

This document details the whole process of analyzing tweets about hajj and umrah from the first stages to the end product.

## Data Analysis:

Is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis [[1]](#ref1).

### Tools:

Data analysis uses many fields to achieve the desired results, therefore, it depends on a lot of different tools and frameworks, we will be using the following tools in our project:

#### Python:

Python is a general-purpose language, which means it’s designed to be used in a range of applications, including data science, software and web development, automation, and generally getting stuff done[[2](#ref2)].

Which makes it the perfect tool for this project as python has a lot of sophisticated libraries for AI and Data analysis which will be integral to our project and save us a lot of time with better results.

#### Libraries:

#### SnScrape:

SnScrape is a scraper for social networking services (SNS). It scrapes things like user profiles, hashtags, or searches and returns the discovered items, e.g. the relevant posts[[3](#ref3)] , which makes it a powerful library which we will be using to gather data.

#### Pandas:

Pandas is a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language[[4]](#ref4).

#### XlsxWriter:

XlsxWriter can be used to write text, numbers, formulas, and hyperlinks to multiple worksheets, and it supports features such as formatting and many more[[5]](#ref5).

#### 

#### PyCharm And VSCode:

PyCharm and VSCode are both IDEs which are Integrated Development Environments which are the places where you can write Python code and test.

#### TweetBinder:

Tweet Binder is an excellent and cost-effective tool for measuring campaign hashtag performance. It is especially handy when it comes to historical reports. Tweet Binder is the ideal tool for tracking, facilitating, and amplifying the online conversation[[7](#ref7)].

## Summary:

To conclude, we discussed the change of the world in the sense of general opinion metrics and how they have changed in recent years, how the new advantages have their disadvantages and disadvantages, lastly we listed all the tools we will be using in the project even.

# Chapter 2: Data Gathering

## Data Gathering:

To analysis data firstly you need the data this is where data gathering comes in, there are many different approaches to gathering data from twitter we have tried two methods both worked however had varying results.

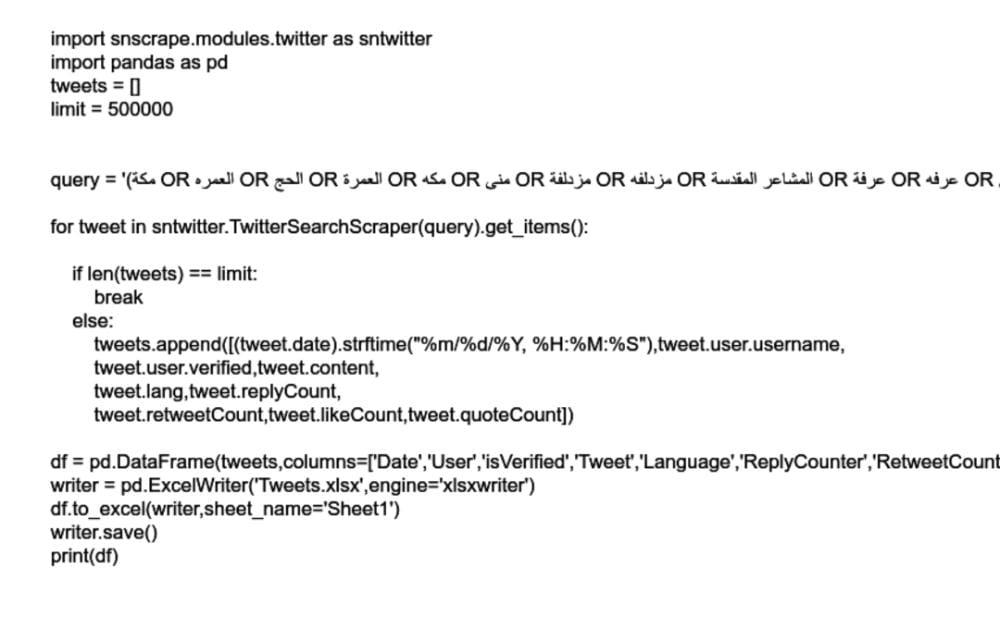
### Method 1: Twitter API:

Firstly, we made a twitter developers account to be able to access the twitter API, afterwards twitter asked us some questions such as what we are going to use the API for, are we students or a private company etc., after completing the sign up and questions we waited for twitters approval for us to use the API which came days later, however, we would come to find out that there is a limit of three thousand tweets we can gather, which would make reaching our goal of one hundred thousand tweets impossible, therefore, we abandoned this method and started looking for alternatives.

#### Method 2: SnScrape Library:

The second method we used which gave us positive results and met the requirements was using the SnScrape library, as it gave us more control over the number of tweets we can search for, as well as the ability to choose what information from the tweet we want to save.

#### Code:



The code set up starts of by importing “SnScrape” twitter module, then importing the panda and remaining it as pd for ease of use in the code, lastly we set a empty array called tweets which we will use to save the tweets information in and set a maximum value of tweets to read to make sure the code doses not run for infinity.

Our method of gathering the data depend on using keywords which ae mentioned in tweets, instead, of hashtags as it will guarantee more data this was done with the following code :

query = '(مكة OR العمره OR الحج OR العمرة OR مكه OR منى OR مزدلفة OR مزدلفه OR المشاعر المقدسة OR عرفة OR عرفه OR الحرم OR المسجد الحرام OR الكعبة OR المشاعر المقدسه OR الكعبه OR الصفا OR المروة OR المروه OR الجمرات OR الاضحى OR نمرة OR نمره)'

The query included the major keywords that are related to the hajj and umrah topic, furthermore, we used the advanced search twitter feature to achieve better results, now everything is ready we used the query and collected the tweets data by looping over them in a for each loop and before saving any data we check if the array has met the limit we set if it has we end the code and start the data transfer to an excel file.

If the limit has not been reached we add a tweet to the array and add the following information from the tweet:

1. Date of tweet
2. Username
3. Verification status true if verified false if not
4. Tweet language
5. Replay count
6. Retweet count
7. Like count
8. Quote count

The majority of the information is straightforward to gather, however, the date of tweets needed further processing to enable python to handle the date format which we did by using “strftime” command to parse the date to the accepted format, this concludes the scraping process.

Afterwards , we need to transfer the tweets to excel for further processing later, firstly, the date format needed further processing to the accepted excel date format which was achieved by the “dateframe” command from the “panda” library, moreover, the “dateframe” was critical to the process by allowing us to transform the data from a 1-Diminsoal array which is not useful to a 2-Dimsoinal array which is what we need.

Lastly ,the tweets where ready to be transferred to the excel file and this was achieved by the “pandas” command “ExcelIWriter” which we received the tweet and using the “XlsxWriter” library we set the engine to be used in the transfer, to close the code we used the “writer. Save” command to save the work and end the code, our end result was 202 thousand tweets.

## Summary:

To summarize, data gathering is the first step in data analysis which is what we achieved by the end of this chapter, furthermore , we explained the code we used in detailed and provided a code snippet which delivers the requirements for the project.

# [References:](#_top)

[1] <https://www.guru99.com/what-is-data-analysis.html>

[2] <https://www.coursera.org/articles/what-is-python-used-for-a-beginners-guide-to-using-python>

[3] <https://github.com/JustAnotherArchivist/snscrape>

[4] <https://pandas.pydata.org/>

[5] <https://pypi.org/project/XlsxWriter/>

[6]

[7] <https://www.tweetbinder.com/>