



Business and Local Government  
Data Research Centre

Project Title

Data-Driven Insights for Maximizing  
Twitter Engagement

Organization

University of Essex

Data Science and AI Masters Project 2024  
Summary Report

EXPLORING DATA  
ENHANCING KNOWLEDGE  
EMPOWERING SOCIETY



University of Essex



Author: Nagadithya Bathala

Academic Supervisor: Amorim Renato

Date: 22-07-2024





## Table of Contents

1	Background .....	1
2	Research Objectives .....	2
3	Dataset .....	2
3.1	Data Source .....	2
3.2	Data Collection .....	2
3.3	Data Attributes .....	3
3.4	Data Collection Process .....	3
3.5	Data Preprocessing .....	4
4	Methods/Methodology.....	4
4.1	Data Exploration using NumPy and Pandas .....	4
4.2	Exploratory Data Analysis (EDA) for Uncovering Insights .....	4
4.2.1	Descriptive Statistics.....	4
4.3	Correlation Analysis.....	6
5	Analysis and Findings .....	6
5.1	Summary of Key Findings .....	6
6	Conclusions and Next Steps .....	21
6.1	Conclusions and Summary.....	21
6.2	Future Work and Next Steps .....	23





# 1 Background

Social media is digital technology that allows the sharing of ideas and information, including text and visuals, through virtual networks and communities. It typically features user-generated content that earns engagement via likes, shares, comments, and discussion. Social media started out as a way for people to interact with friends and family, but soon expanded to serve many different purposes. The usage and popularity of social media increased with the entry of Facebook and Twitter (now X). More than 5 billion people around the world use social media.

Businesses are increasingly using social media to connect with their customers in an innovative way. It is an important part of many companies marketing campaigns and plays a key role in many businesses marketing strategies given the sheer number of hours people spend each day on social websites and apps. Social media platforms are often divided into six categories such as social networking, social bookmarking, social news, media sharing, microblogging, and online forums. For businesses, social media has become a key marketing tool. Companies use it to find and engage with customers, drive sales through advertising and promotion, identify fast-moving consumer trends, provide customer service or support. Many businesses, cultural organizations and social institutions are leveraging social media to achieve their strategic goals. According to research that assessed the social media activity of the top 100 most valuable global brands, those that were most socially active saw an 18% increase in revenue during the previous year, while the least active experienced a 6% revenue decrease.

Among many social media platforms, Twitter has rapidly grown to a popular social network in recent years and provides many real-time messages for users. Twitter is one of the most popular microblogging services and an important social network with over 500 million registered users as of July 2014. In this research project twitter data was used to analyse how businesses have used twitter to engage with their customers and find useful insights that will help the businesses to maximise the user engagement. User engagement is a critical aspect of social media's impact. It refers to how users interact with content, including actions like likes, shares, comments, and views. High engagement levels indicate that content resonates with audiences, and increasing the visibility of posts through algorithms that prioritize popular content. For individuals, businesses, and influencers, understanding and leveraging user engagement is key to achieving communication goals.





## 2 Research Objectives

The main objective of this research is to focus on analysing user engagement on Twitter. By examining the factors that drive engagement, this study aims to provide insights that can help users and organizations optimize their content strategies, enhance their online influence, and better understand their audience's behaviour on the platform.

The primary objectives of this study are:

- Collect the data from twitter using available web scraping techniques from various businesses accounts.
- Calculate the overall user engagement and explore the factors affecting the engagement.
- To identify the influence of using hashtags, emojis and mentions on user engagement.
- To examine how different factors such as posting time, tweet length, and posting day affect engagement.
- To explore the relationship between the user sentiments and engagement levels.
- To assess the influence of an account's follower count on the engagement its tweets receive.

## 3 Dataset

### 3.1 Data Source

The primary step in any research project is data collection. The data for this study was collected from twitter social media platform using Selenium and NTScraper, two powerful tools for web scraping and data extraction from Twitter. Selenium is a web automation tool that simulates user interactions with web pages, enabling the collection of dynamic content. NTScraper is a specialized tool designed for efficiently scraping Twitter data, providing a streamlined approach to gather large volumes of tweets and related information.

### 3.2 Data Collection

**Period of Data Collection:** The data collection spanned from 2020 to 2024, ensuring that the dataset captures the tweets during different time periods, including seasonal trends and significant events that may influence user engagement.







**Sample Size:** The dataset comprises 23 thousand tweets. This sample size was selected to provide a representative cross-section of Twitter activity, ensuring diversity in terms of user accounts, content types, and engagement levels.

### 3.3 Data Attributes

The dataset includes a wide range of attributes, providing a comprehensive view of each tweet and its context:

- Link: A unique URL for each tweet.
- Tweet ID: A unique identifier for each tweet.
- Text: The text content of the tweet.
- Username: The handle of the user who posted the tweet.
- Profile ID: A unique identifier for the user who posted the tweet.
- Date: The timestamp of when the tweet was posted.
- Comments: Number of comments for each tweet.
- Retweets: Number of retweets for each tweet.
- Quotes: Number of times a tweet is retweeted with comment.
- Likes: Number likes for each tweet.
- Account name: Twitter account name of the user who posted the tweet.
- Followers: Followers count of the user who posted the tweet.

#### Engagement Metrics:

- Likes: The number of likes the tweet received.
- Retweets: The number of times the tweet was retweeted.
- Comments: The number of replies to the tweet.
- Quotes: The number of times the tweet was retweeted with comment.

### 3.4 Data Collection Process

**Using Selenium:** Selenium was used to automate web browsing tasks, such as logging into Twitter, navigating to user profiles, and scrolling through timelines to load more tweets. This approach allowed for the collection of dynamic content that is not readily available through static HTML. It was mainly used to scrape the following list of specific twitter handle.

**Using NTScraper:** NTScraper was used to the scrape the large volumes of tweets. This tool provided capabilities for extracting tweet content, user information, and engagement metrics in a structured format.





### 3.5 Data Preprocessing

**Dropping NaN values:** Rows with missing values could significantly affect the analysis. Therefore, dropped the rows containing these values.

**Removing Duplicates:** Dataset has duplicates which are not helpful for analysis. These duplicates can influence the analysis. Therefore, duplicate entries were removed from the data.

In summary, the dataset collected using Selenium and NTScraper offers a comprehensive view of user engagement on Twitter. The data provides a reliable basis for analysing the factors influencing tweet engagement and deriving actionable insights.

## 4 Methods/Methodology

### 4.1 Data Exploration using NumPy and Pandas

NumPy and Pandas are fundamental tools and powerful libraries tools in the Python ecosystem for data analysis and manipulation. The exploration aims to understand the dataset, clean and preprocess the data, and perform exploratory data analysis to uncover patterns and insights related to Twitter user engagement.

- **NumPy:** Primarily used for numerical computations on large, multi-dimensional arrays and matrices. It provides efficient functions for mathematical operations and linear algebra.
- **Pandas:** Built on top of NumPy, Pandas is used for data manipulation and analysis. It offers data structures like Data Frame (for 2-dimensional data) and Series (for 1-dimensional data), which allow easy indexing, slicing, grouping, and aggregating data.

### 4.2 Exploratory Data Analysis (EDA) for Uncovering Insights

Exploratory Data Analysis (EDA) is a critical phase in any data analysis research project, aimed at summarizing the main characteristics of the data and uncovering patterns and insights. Different EDA techniques were employed to analyse Twitter data and draw meaningful conclusions about user engagement and sentiment.

#### 4.2.1 Descriptive Statistics

Descriptive statistics play a crucial role in summarizing and understanding the basic features of the dataset. This section provides a detailed overview of the descriptive statistical analysis performed on the Twitter dataset.





#### *4.2.1.1 Calculate user engagement:*

User engagement of each tweet was calculated by adding the number of likes, number of retweets, number of comments and number of quotes. A utility function was created in python to calculate the user engagement of each tweet.

Each user has varying number of followers and number of followers has direct impact on the user engagement. To compare the tweets with each other, the user engagement was normalized based on the number of followers.

#### *4.2.1.2 Tweet Length*

Calculating tweet length is an important aspect of analysing Twitter data, as it can provide valuable insights into how the length of a tweet influences user engagement. Tweet length can significantly impact the level of user interaction, such as likes, retweets, and comments. Understanding this relationship helps in crafting more engaging tweets. The average length of the tweets was calculated to check the impact of tweet length of engagement metrics.

#### *4.2.1.3 Hashtags*

Hashtags are integral to Twitter's communication style, serving as a tool to categorize content, enhance visibility, and drive user engagement. They help categorize tweets, making it easier for users to find and participate in discussions on specific topics.

Hashtags increase the discoverability of tweets, enabling content to reach a broader audience beyond the user's followers. Analysis on hashtags was performed to examine their frequency, distribution, and impact on tweet engagement.

#### *4.2.1.4 Mentions*

Mentions are another important aspect of Twitter interactions, where users tag other accounts in their tweets using the "@" symbol followed by the username. Analysing mentions helps identify influential users and accounts that drive conversations and engagement.

#### *4.2.1.5 Emojis*

Emojis play a significant role in digital communication, adding emotional nuance and context to textual content. They provide additional emotional context that words alone may not fully convey. Certain emojis can carry cultural meanings or references that resonate with specific audiences. Analysis on emojis was carried out to examine their frequency, distribution, and impact on tweet engagement.







#### 4.2.1.6 *Analysis based on day of the week*

The day of the week can significantly influence Twitter user engagement and content dynamics. Analysing the tweet activity based on the day of the week is very important as user activity on Twitter can vary significantly across different days of the week. Understanding these patterns helps in identifying peak times for engagement. Knowing which days are most active or have the highest engagement can help in scheduling tweets for maximum visibility and interaction.

#### 4.2.1.7 *Analysis based on time of the day*

The time of day can significantly impact user engagement. It is very important to analyse tweet activity based on the time of day as user activity on Twitter can vary greatly throughout the day. Understanding these peaks can help identify optimal times for tweeting. Finding which hours are most active or have the highest engagement can aid in scheduling tweets for maximum impact.

### 4.3 Correlation Analysis

Examining the relationships between different variables such as tweet length, sentiment, number of hashtags, number of emojis and user engagement metrics helps to understand the dependency of user engagement on other variables.

## 5 Analysis and Findings

To understand the tweets and find actionable insights that helps to maximise the user engagement, the methods mentioned in the methodology section were used. This section presents the comprehensive analysis and key findings from the Twitter user engagement data. The analysis was performed using various techniques and tools to uncover insights into engagement patterns, sentiment trends, and content effectiveness.

### 5.1 Summary of Key Findings

#### Data Summary

A total of 23000 tweets were scraped from the Twitter using NTScraper and Selenium libraries. Basic data cleansing steps were performed to make sure data is suitable for data analysis. The data was collected from multiple Twitter businesses accounts each promotes different business to avoid bias in content types, and engagement levels.





## User Engagement

Calculating user engagement on Twitter involves quantifying the interactions that users have with tweets. This metric is crucial for understanding how effectively a tweet resonates with its audience. A utility function in python was created to calculate the user engagement by summing all the engagement metrics.

The user engagement values plotted using box plot to check the distribution of the values. It was observed that some tweets have very high engagement compared to most of the tweets. Initially, data analysis was carried out on the complete data and then removing the data anomalies referred here as outliers. These outliers had a greater impact on the results. Therefore, these outliers were filtered out to perform further analysis. There was a huge difference in average user engagement before and after filtering out the outliers. Average engagement with complete data was around 388, whereas after filtering out the outliers it came down to 78.

Data	Avg Engagement
Complete data	388
Data with outliers removed	78

**Figure1-** Average Engagement of Tweets with and without Outliers

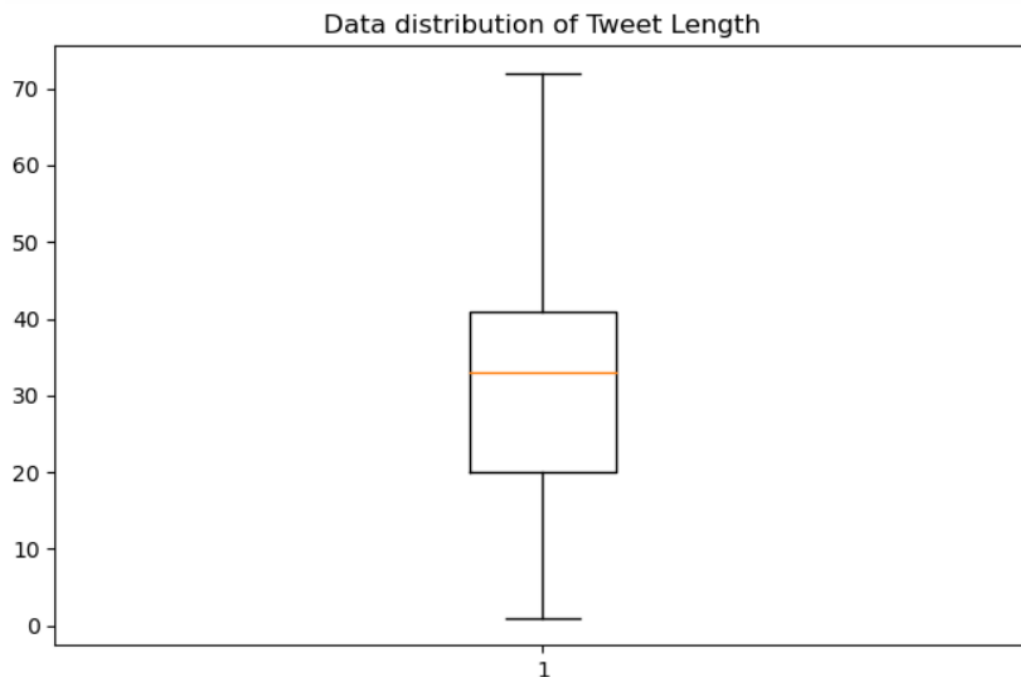
## Tweet Length

Tweet length plays a crucial role in user engagement and content effectiveness on Twitter. Identifying the ideal tweet length that generates the highest engagement helps in crafting effective tweets. It affects how clear and concise the message is, influencing user interaction. Length of the tweets was calculates using split and apply function in python.

The data collected contains the tweets with varying length between 1 to 72. It was observed that the average length of the tweets was 30. The box plot was used to check the data distribution.

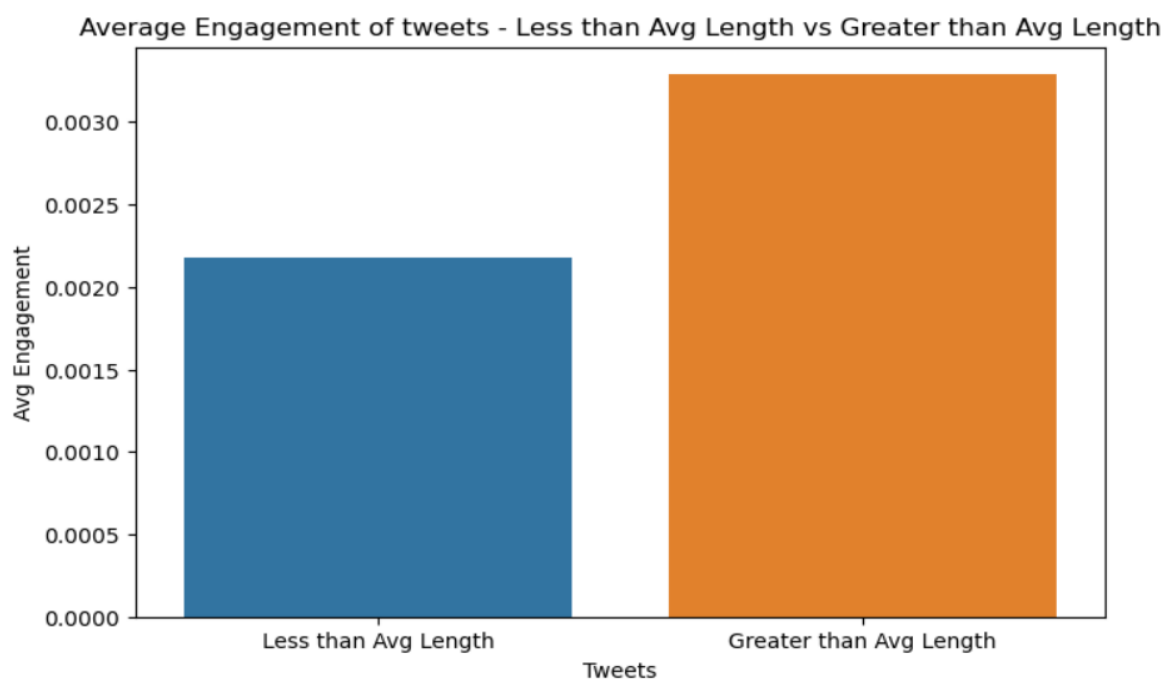
Tweet Length	Value
Miniumun	1
Maximum	72
Average	30

**Figure2-** Summary Statistics of Tweets Length



**Figure3** – Data distribution of Tweet Length

Further analysis was carried out to find out the user engagement of tweets with length more than average tweet length and tweets with length less than average tweet length. It was observed that tweets with length more than average tweet length has higher engagement.



**Figure4-** Average Engagement of tweets with respect to Length

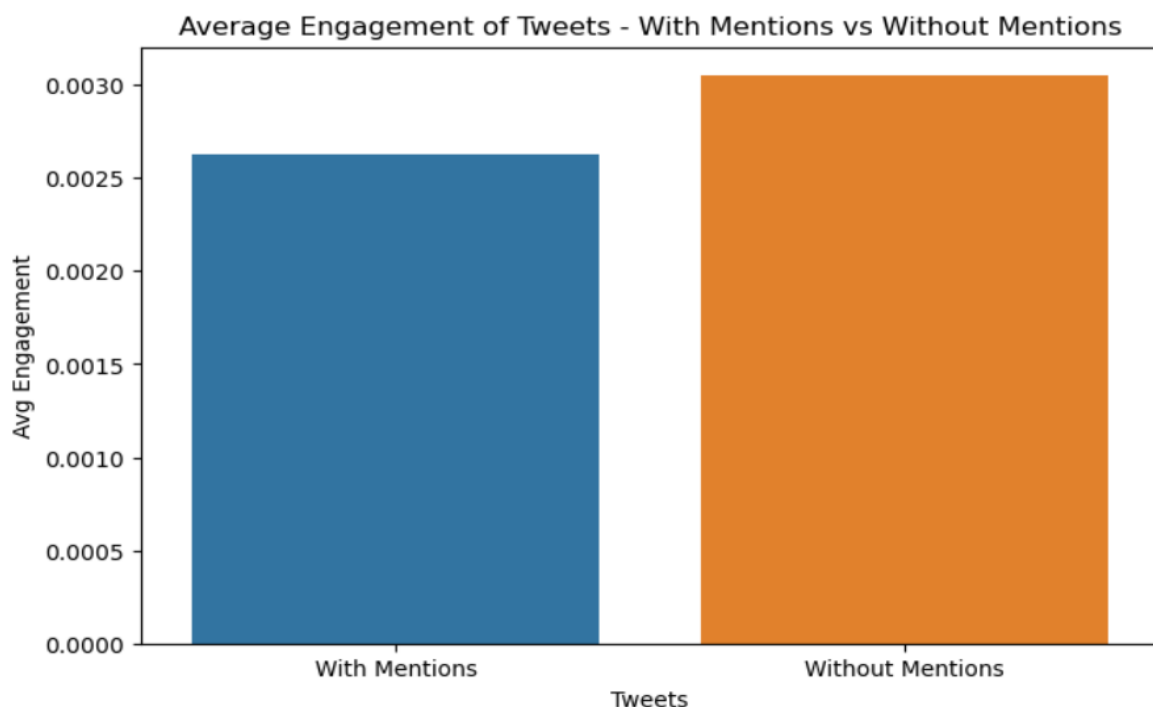


The results shows that tweets length has influence on user engagement. The above bar plot shows that Shorter tweets (less than 30 characters) tended to receive less engagement, possibly due to less content or context. Longer tweets often received higher engagement, possibly due to detailed content or storytelling.

### Analysis on mentions

Mentions on Twitter are indicated by the "@" symbol followed by a username. They are an important feature for engaging with other users, promoting interaction, and increasing visibility. The usage of mentions in tweets was analysed by exploring the impact on user engagement metrics such as likes, retweets, quotes, and comments.

Mentions from each tweet were extracted by defining a function using regular expressions in python. The function was applied on dataset to find Mentions in each tweet. The average user engagement of tweets that contains mentions and doesn't contain mentions were calculated by separating the data accordingly.



**Figure5-** Average Engagement of Tweets WITH MENTIONS vs WITHOUT MENTIONS

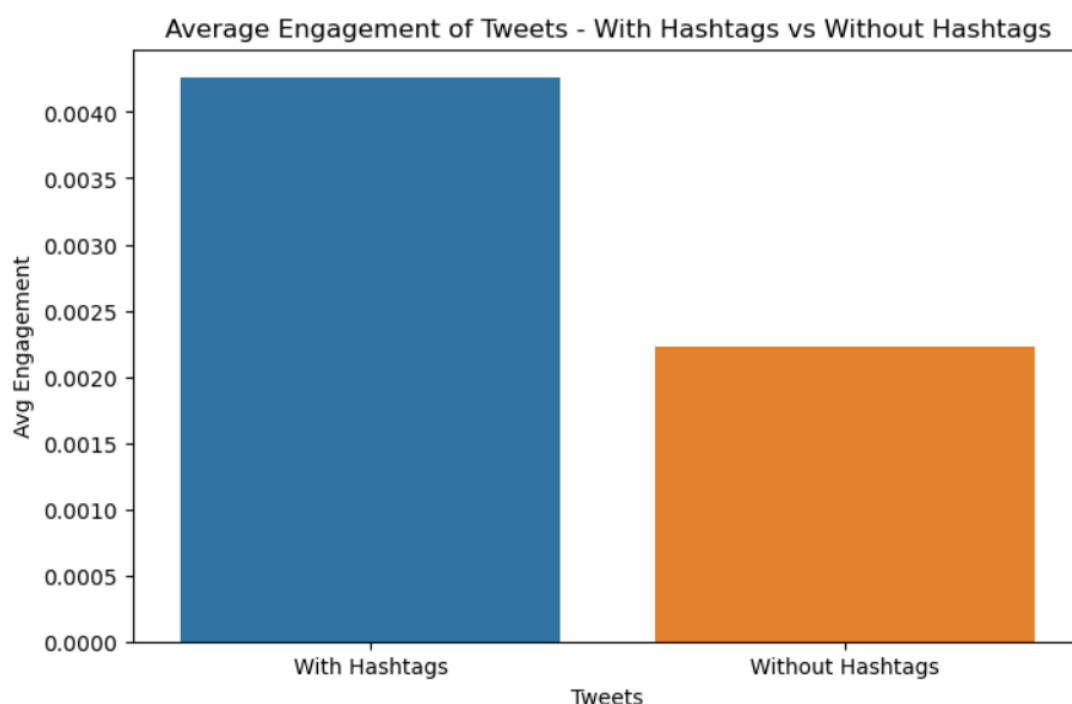
In general tweets with mentions may receive more immediate responses and interactions. But the data shows that overall average engagement of tweets without mentions exhibits slightly higher engagement compared to the tweets with mentions. Tweets without mentions often have a more general and neutral tone, appealing to a broader audience and might be perceived as more original and less promotional. Users often value original thoughts, opinions, and content that provide insights or entertainment without appearing to directly seek attention from specific individuals.



## Hashtag Analysis

Hashtags are a fundamental component of Twitter, serving to categorize content, increase visibility, and engage with broader conversations. Hashtags enable tweets to reach a wider audience beyond the followers, increasing discoverability. Hashtags can enhance engagement by connecting tweets to relevant topics and communities.

Hashtags from each tweet were extracted by defining a function using regular expressions in python. The function was applied on dataset to find hashtags in each tweet and count of hashtags in each tweet and stored them in a table. The average user engagement of tweets that contains hashtags and doesn't contain hashtags were calculated by separating the data accordingly.

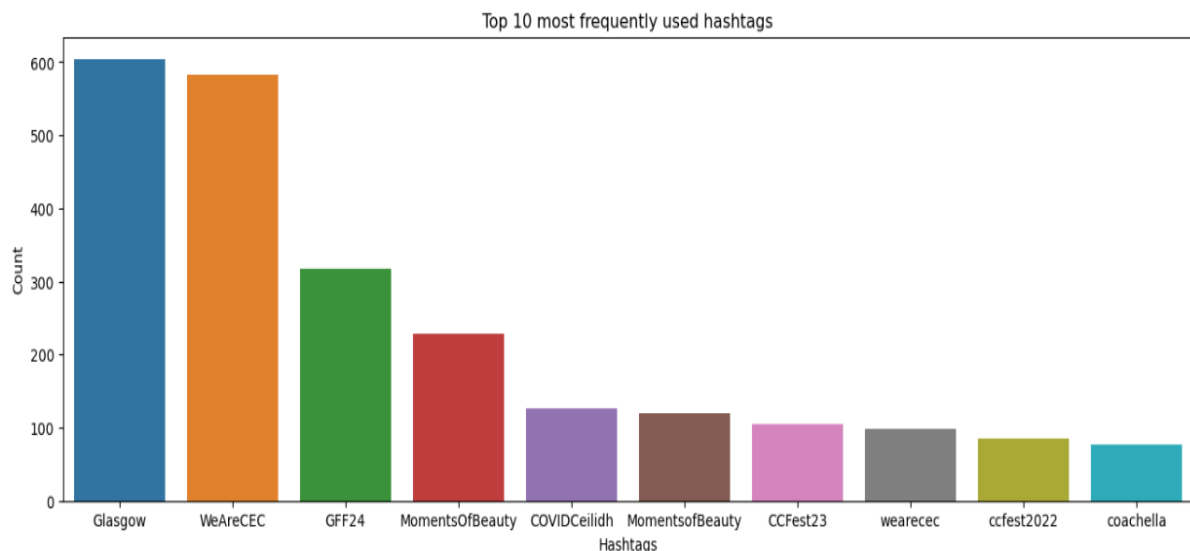


**Figure6-** Average Engagement of Tweets WITH HASHTAGS vs WITHOUT HASHTAGS

The above plot shows that the tweets that contain hashtags have higher user engagement compared to tweets without any hashtags. This is possibly due to tweets having hashtags having a higher chance of getting good reach in followers as well as outside of followers list which has a direct impact on increasing the user interactions.

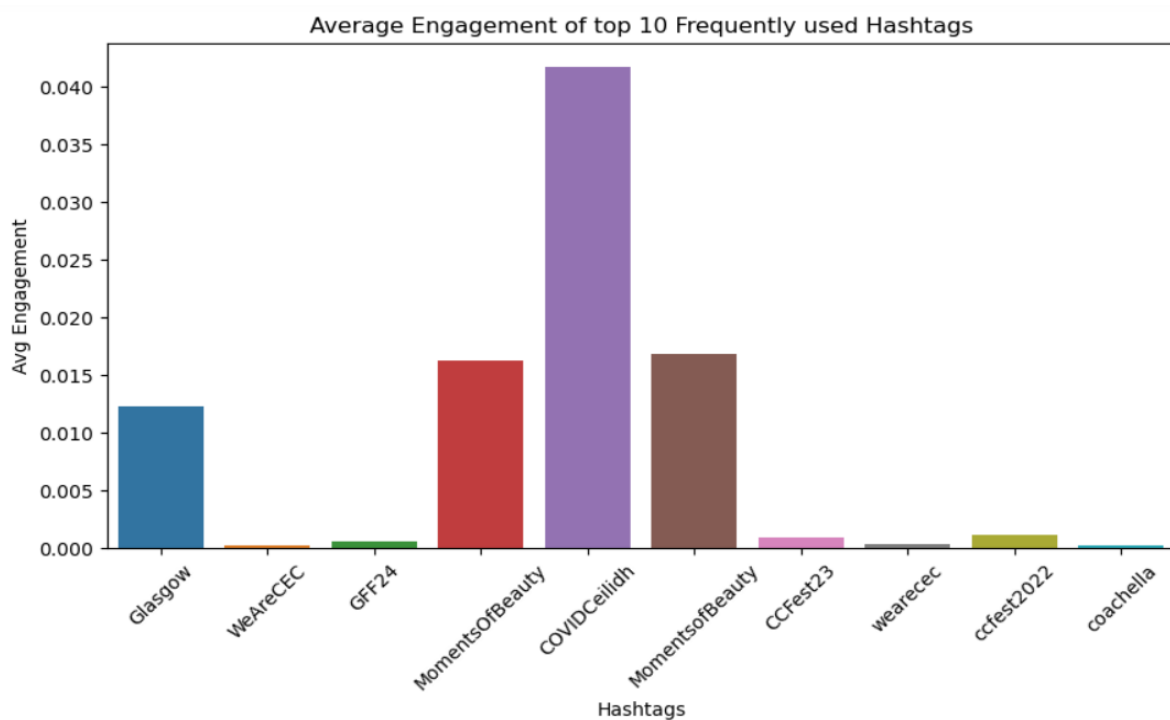
Further a table of hashtags and frequency of each hashtag was created to investigate the relation between most frequently used hashtags and user engagement. The dataset has around 4000 unique hashtags which were used in the tweets. It is difficult to investigate the engagement of all these hashtags. Therefore, top 10 most frequently used hashtags were selected for this analysis.





**Figure7-** Top 10 most Frequently used Hashtags

The above bar plot shows top 10 frequently used hashtags in the tweets. Glasgow hashtag was used around 600 times which is most compared other hashtags. The second most used hashtag was WeAreCEC.



**Figure8-** Average Engagement of top 10 Frequently used Hashtags

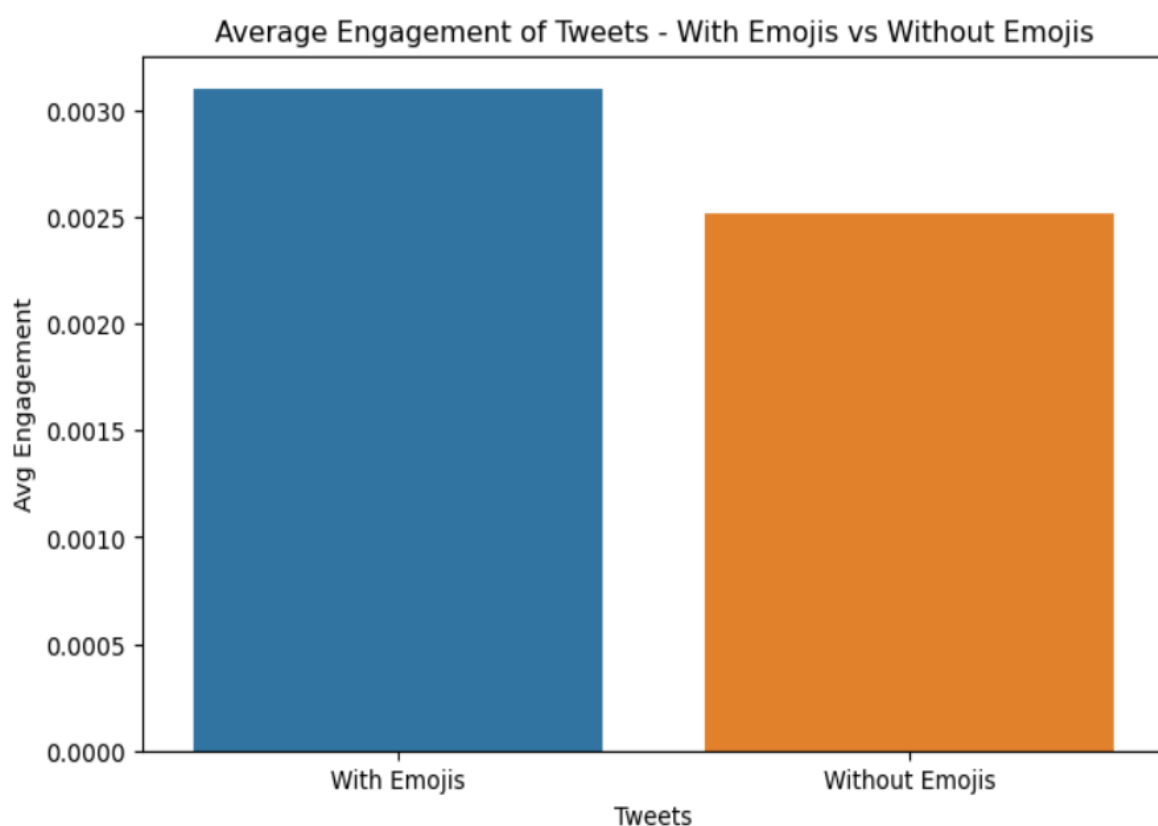
Upon filtering the top 10 most frequently used hashtags, the analysis revealed varying levels of user engagement. Surprisingly, the most frequently used hashtag did not yield the highest average engagement. Instead, the 5th most frequently used hashtag showed higher average engagement followed by 4th, and 6th. These hashtags exhibited higher average engagement, indicating that they have more reach in the audience, and they are used in more engaging content.



## Emoji Analysis

Emojis have become an integral part of communication on social media platforms, including Twitter. They add emotional context, enhance the visual appeal of tweets, and can significantly impact user engagement. Here analysis on emojis was carried out to understand the relation of user engagement with emojis.

Emojis from each tweet were extracted by defining a function using a list of predefined emojis in python. The function was applied on dataset to extract emojis in each tweet and count of emojis in each tweet and stored them in a table. The average user engagement of tweets that contains emojis and without emojis were calculated by separating the data accordingly.

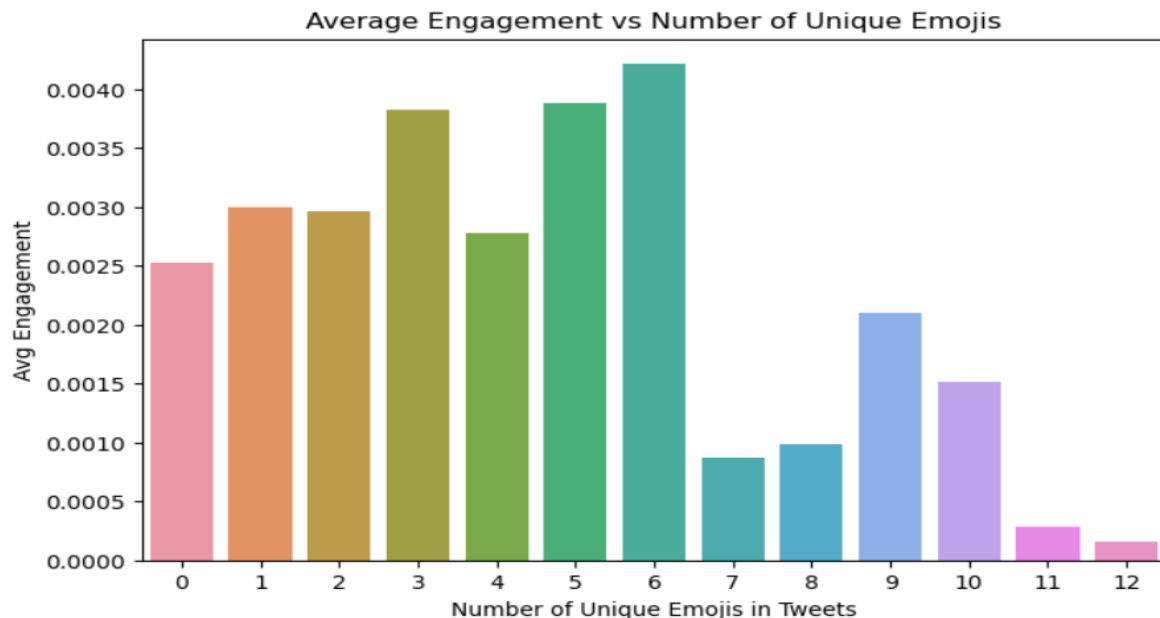


**Figure9-** Average Engagement of tweets WITH EMOJIS vs WITHOUT EMOJIS

The above plot shows that the tweets that contain emojis have slightly higher user engagement compared to tweets without any emojis. But the difference is very minimal. This is possibly due to emojis making tweets more visually engaging and can attract user attention which leads to higher engagement.



It was observed that there are duplicated emojis in many tweets. Therefore, unique emojis from each tweet were filtered out and calculated number of unique emojis in each tweet to carry out further analysis.



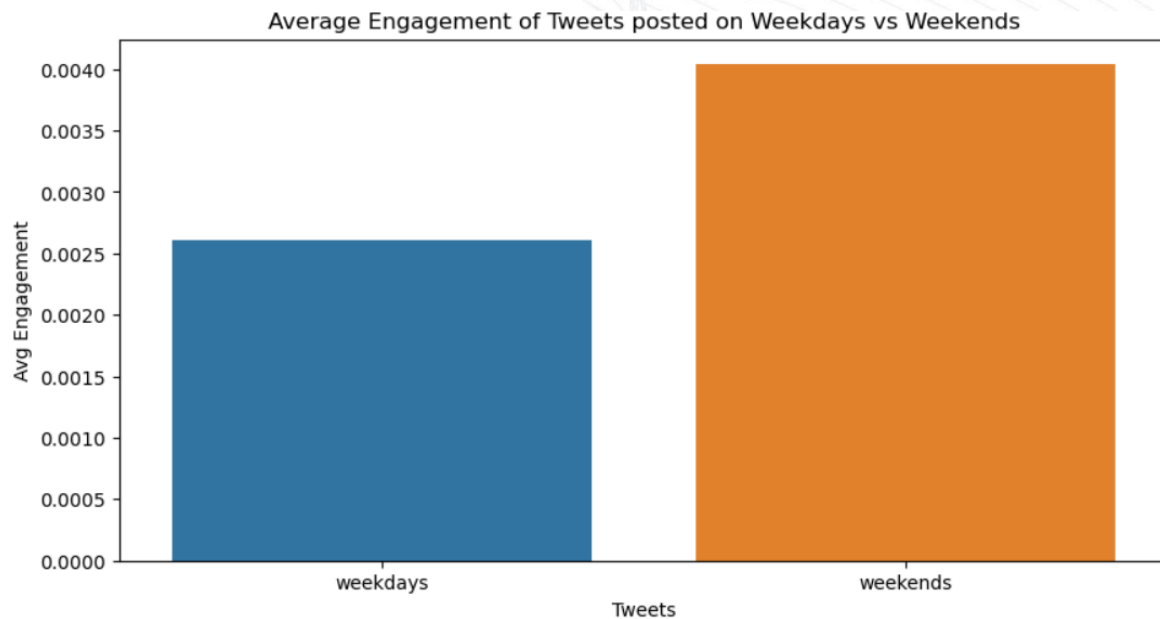
**Figure10-** Average Engagement of Tweets with respect to number of Emojis

The tweets were grouped based on number of unique emojis used and calculated average engagement of each group to understand the relation between them. It was observed that tweets with 1-6 emojis received the highest average user interaction. Excessive use of emojis (6+ emojis) showed a decline in likes, indicating potential overuse of emojis.

#### Analysis based on Weekdays vs Weekends

Understanding how engagement varies between weekdays and weekends can provide valuable insights into user's behaviour on Twitter. User routines differ significantly between weekdays and weekends, influencing their online activity and engagement levels. Analysing these patterns helps understand user behaviour, enabling more effective engagement strategies.

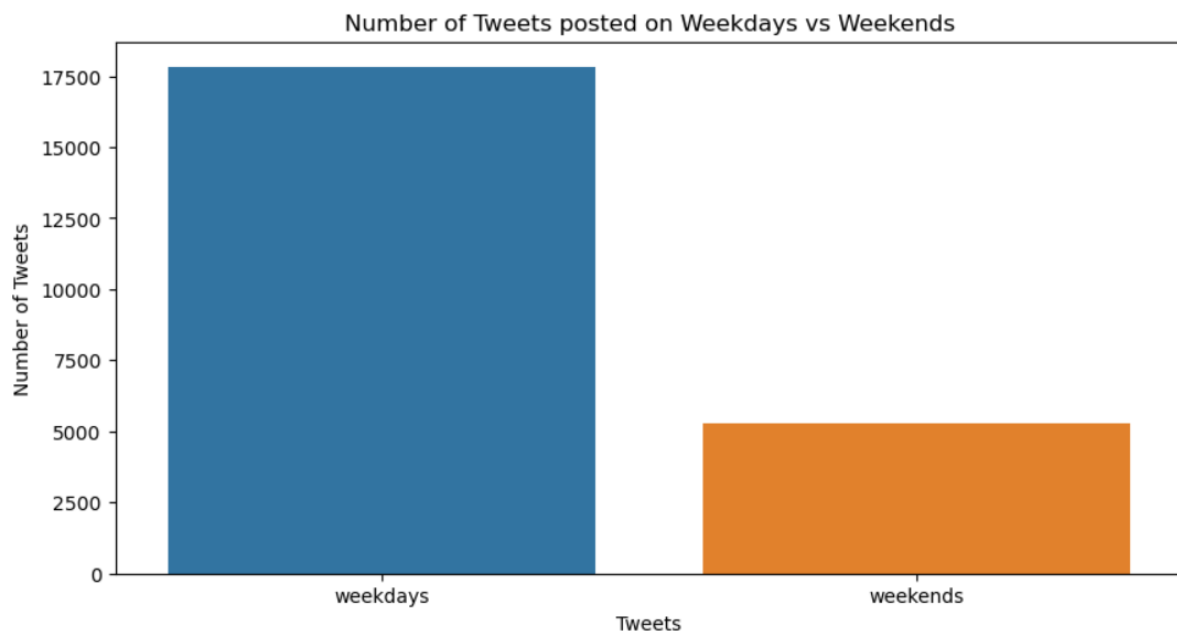
Analysis was performed to compare the engagement metrics of tweets posted during weekdays i.e., Monday to Friday with those posted on weekends i.e., Saturday and Sunday, helping to identify optimal days to post tweets to maximise the user engagement. Day of the week was extracted on which the tweets were posted and grouped the tweets based on weekdays and weekends. The average engagement of each group was calculated to check the relationship with engagement.



**Figure11-** Average Engagement of Tweets – Weekdays vs Weekends

The analysis on engagement of tweets posted on weekdays and weekends shows that the user interactions are higher for tweets posted on weekends compared to the tweets posted on weekdays. Engagement metrics are generally higher on weekends. This is possibly due to users getting more free time on weekends to interact with social media.

Furthermore, to investigate the user interactions on tweets, the number of tweets posted on weekdays and weekends were explored. Tweets were grouped based on weekdays and weekends and calculated number of tweets in each group.



**Figure12-** Number of Tweets posted on Weekdays vs weekends



The tweets used in the analysis are mostly related to businesses account and it is obvious that business accounts tend to post more tweets on weekdays compared to weekends. The bar plot above indicates the same trend i.e., tweets posted on weekdays are more compared to weekends.

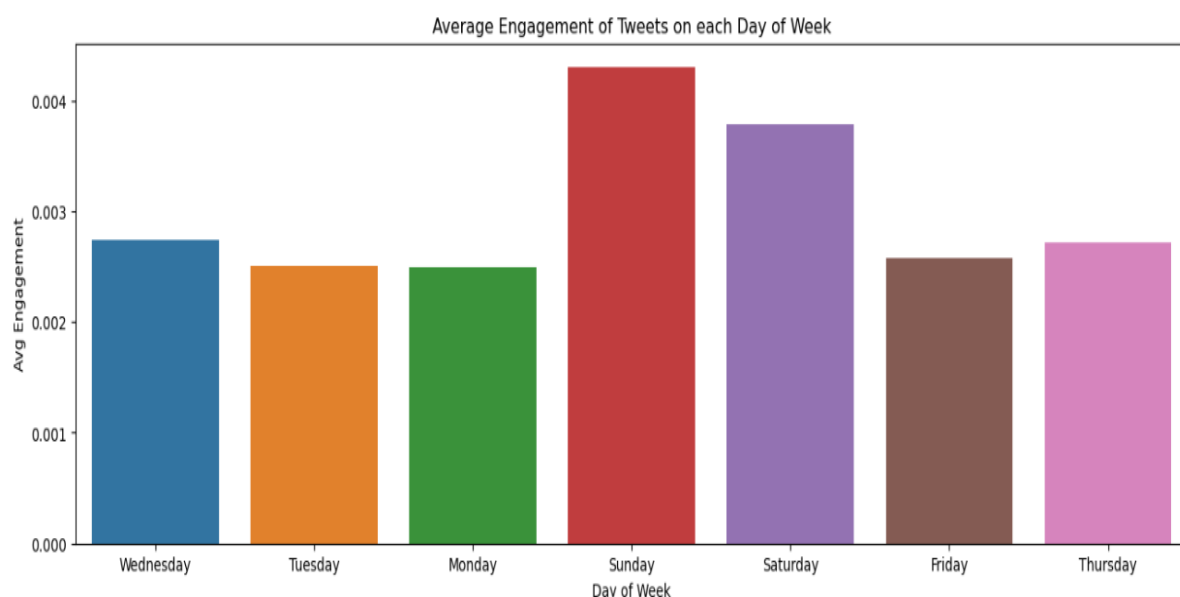
Businesses typically operate during standard work hours, which align with weekdays. As a result, their social media activities, including tweeting, are more concentrated during these periods. Social media teams and content creators are more likely to be active and available during weekdays, leading to a higher volume of posts.

### Analysis based on day of the week

The day of the week can significantly impact user engagement on Twitter just like any other social media platform. Different days may see varying levels of activity and interaction, influenced by factors such as work schedules, social habits, and online behaviour patterns. Here we explored how engagement metrics vary across different days of the week, providing insights into optimal posting days of the week.

User activity on Twitter fluctuates throughout the week, with certain days showing higher or lower engagement. People's routines and availability for interacting with content can vary based on the day of the week. Identifying days with higher engagement can help optimize posting schedules to maximize reach and user interaction. Understanding day specific engagement trends allows for strategic planning of content creation and posting.

The day of the week on which the tweets posted were extracted using parse function from dateutil library to identify the patterns in user interactions on each day of the week. The tweets were grouped based on day of the week and calculated the average engagement for each group.



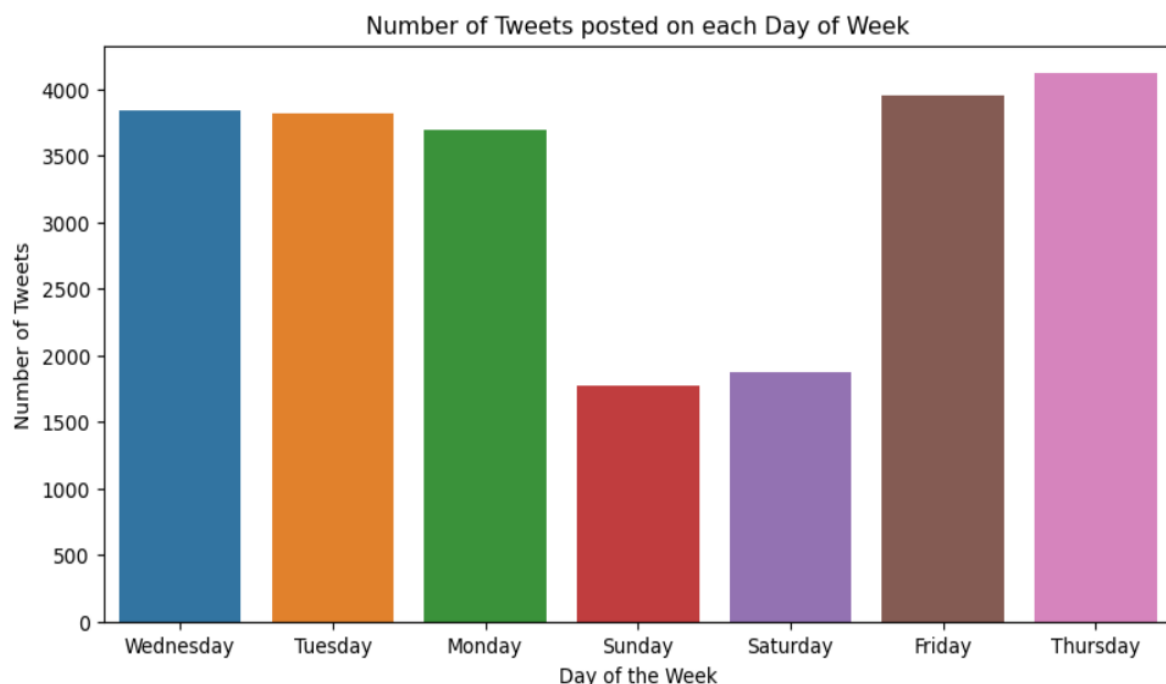
**Figure13-** Average engagement of Tweets on each Day of Week





The analysis shows that certain days, such as Saturday and Sunday, show higher average engagement, suggesting weekend peaks in user engagement supporting the previous analysis on engagement on weekdays and weekends. Saturday and Sunday are holidays for most of the people. On these days, users typically have more free time compared to weekdays, allowing them to spend more time on social media without the constraints of work or school, users can engage more with content.

To further investigate the number of tweets posted on each day of the week, tweets were grouped based on day of the week and calculated number of tweets posted on each day. Investigating this helps to find some actionable insight that helps to maximise user engagement.



**Figure14-** Number of Tweets posted on Each day of Week

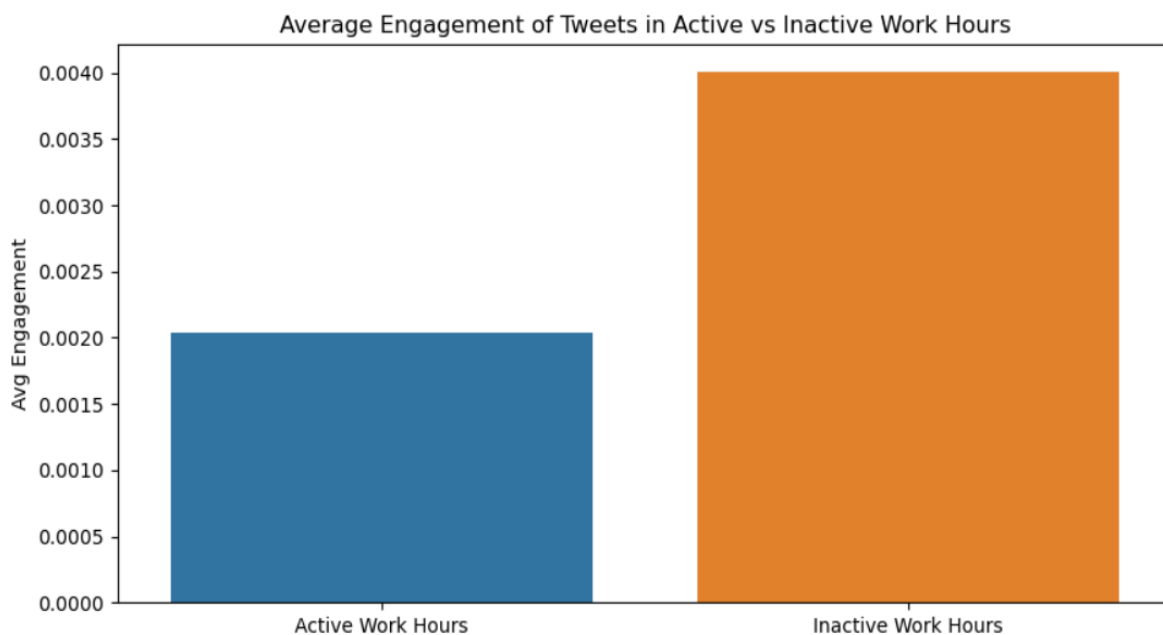
The bar plot above shows number of tweets posted on each day of the week. It shows number of tweets posted on Saturday and Sunday are less compared to other days of the week. But investigation on engagement showed that Saturday and Sunday have higher user interactions. Most businesses operate during standard working hours, which are typically Monday through Friday. Consequently, social media teams and content creators are less active on weekends. Businesses often have reduced staffing on weekends, leading to fewer social media activities, including posting tweets. They plan their content and social media activities around the workweek. They may schedule fewer posts on weekends to align with their operational schedules and the availability of their audience.



## Engagement analysis based on Work Hours vs Outside Work Hours

Understanding the impact of active work hours and inactive work hours on tweet engagement is crucial for optimizing content strategy. Users might be busier during typical work hours with work or school activities, affecting their ability to engage with content. They may have more free time to browse social media and interact with content outside of their work hours.

To better understand the patterns, investigation was carried out to examine how engagement metrics differ between tweets posted during typical work and those posted outside these hours. Here, 9Am to 5PM were considered as typical work hours. Understanding these patterns helps providing insights into optimal posting times. Hour of the day from the timestamp was extracted using `strptime` function from `datetime` module. Tweets were grouped based on the work hours and outside work hours and calculated average engagement.

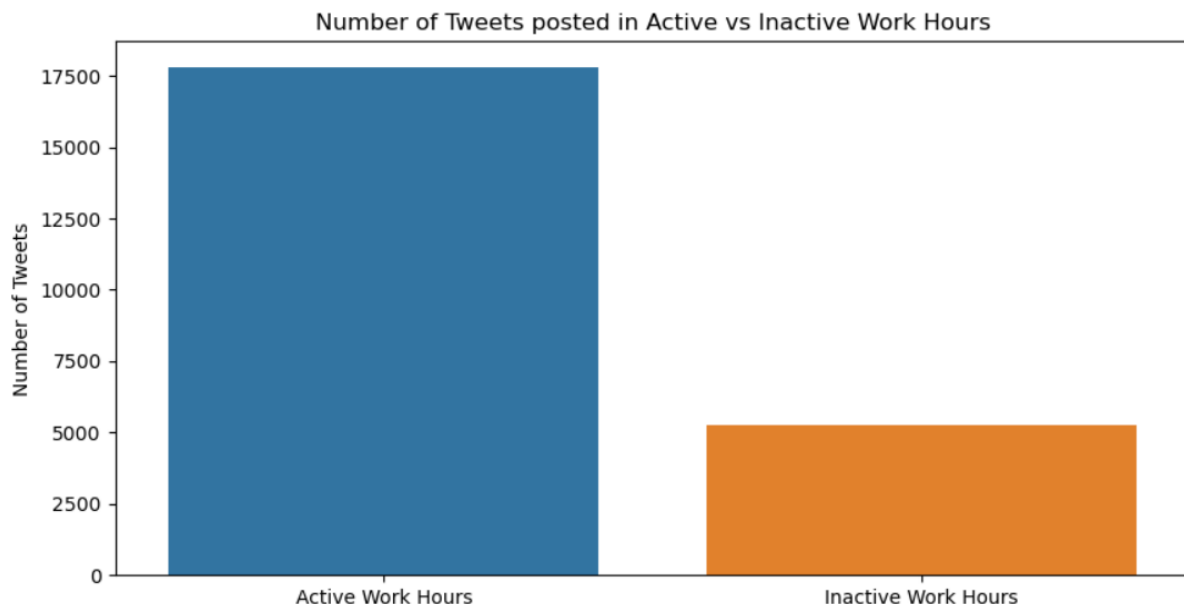


**Figure15-** Average Engagement of Tweets posted in WORK HOURS vs OUTSIDE WORK HOURS

The bar plot above shows that engagement of tweets posted during outside work hours is far higher compared to engagement of tweets posted during typical work hours. This trend suggests that users are more likely to engage with content when they have more free time outside of their work commitments. Engagement tends to be lower during active work hours, as users are often busy with professional or academic responsibilities. Posting high-value content during active work hours might result in lower immediate engagement and might get low engagement because of tweet might not stay visible throughout the day.



The further explore these patterns, number of tweets posted during work hours and outside work hours were calculated. Tweets were grouped based on the active and inactive work hours and calculated number of tweets in each group.



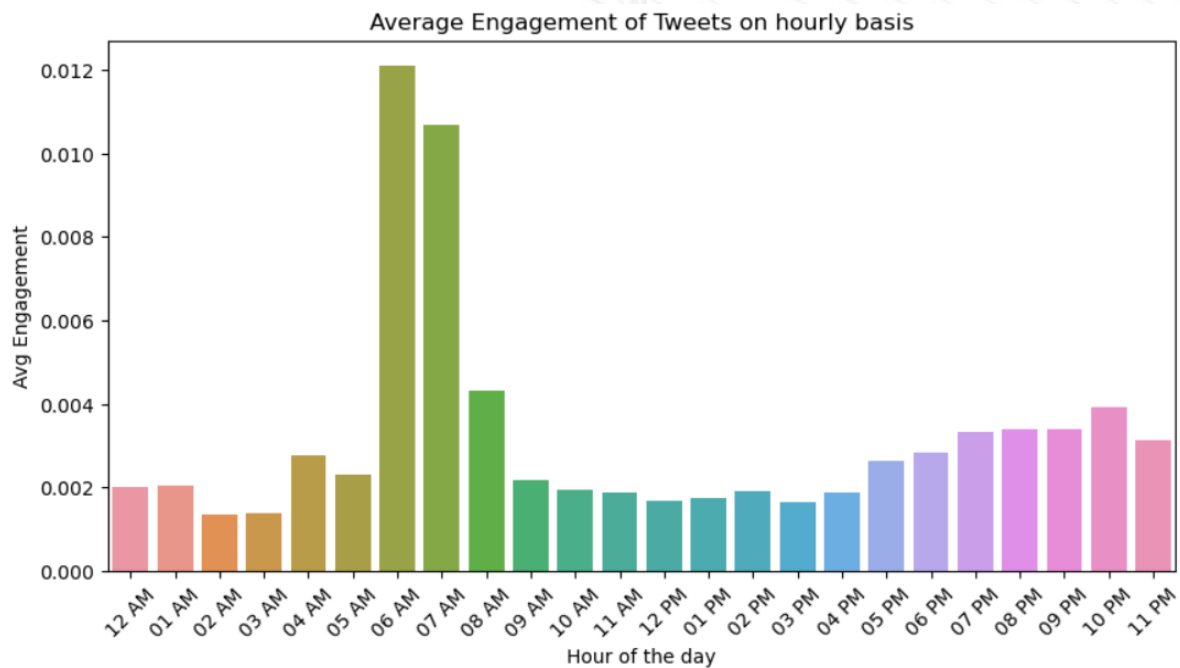
**Figure16-** Number of Tweets posted in WORK HOURS vs OUTSIDE WORK HOURS

Analysis shows that number of tweets posted during work hours were way more than the tweets posted during outside work hours. This is possibly due to content creators of the businesses are active during typical work hours. This might be the possible reason behind bulk tweets during work hours. To achieve maximum engagement, scheduling tweets during inactive work hours can be more effective. High-value content, promotional posts, and interactive tweets might perform better during this period.

#### Analysis based on Hour of the Day

Understanding how user engagement varies by hour of the day is crucial for optimizing the timing of tweets to maximize visibility and interaction. Users have varying levels of activity throughout the day, influenced by work, school, and personal schedules. Identifying hours with peak engagement can help in scheduling tweets for maximum impact. Aligning tweet schedules with hours of high user activity can enhance reach and interaction.

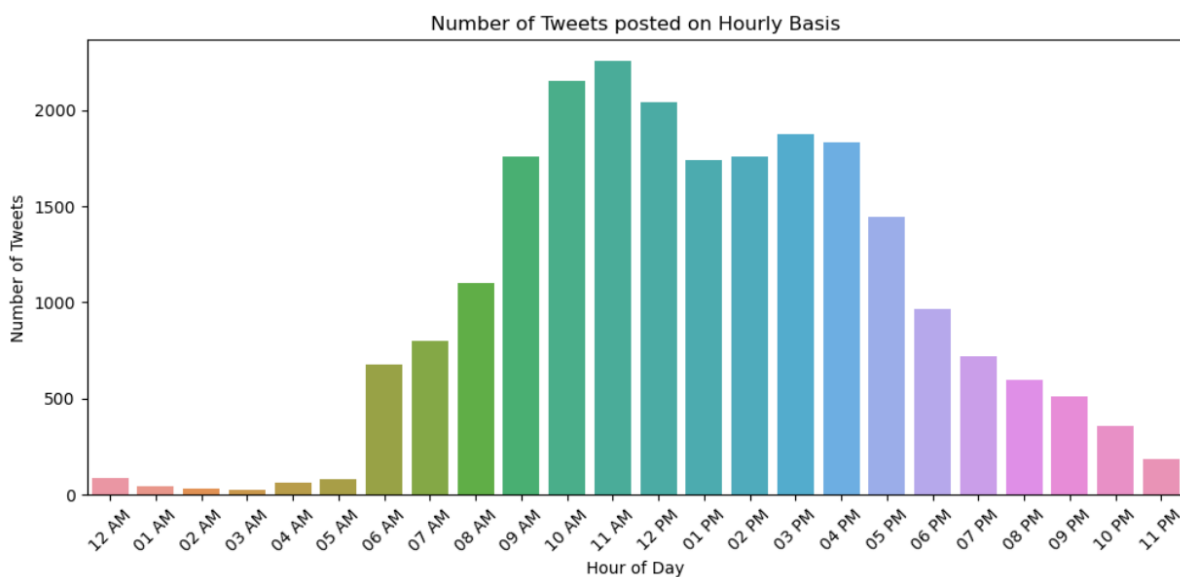
An investigation was carried out to further understand the engagement patterns of tweets posted at different hours throughout the day by examining how engagement metrics differ between tweets posted during different hours of the day. Tweets were grouped based on the hour and average user interaction was calculated for each group.



**Figure17-** Average Engagement of Tweets on Hourly Basis

The study above indicates that, the user interactions were higher for the tweets posted between 6AM to 8AM compared to tweets posted at other hours. This trend can be related to several factors related to user behaviour and daily routines. The data suggests that users are more likely to engage with content during this early morning period. It also indicates that users are more inclined to share content when they first wake up or start their day.

To further find out actionable insights to maximise engagement, number of tweets posted on hourly basis throughout the day were explored by grouping data accordingly and calculate number of tweets posted on each hour of the day.



**Figure18-** Number of Tweets posted on Hourly Basis





The above analysis shows that number of tweets are higher during 9 AM to 5PM and the count started to decrease post 5PM. This trend is possibly due to the content creators of businesses post tweets during typical work hours. Work hours of most of the businesses are ideally during 9 AM to 5 PM. This pattern suggests that businesses tend to post more during these periods.

Many users start their day early and check social media as part of their morning routine. This time often involves activities like commuting, having breakfast, or planning the day, during which users may browse and engage with content on Twitter. Early risers, including professionals and students, are likely to be active on social media during this hour, contributing to higher engagement.

## Correlation Analysis

Correlation analysis is a statistical method used to examine the strength and direction of relationships between two or more variables. In the context of Twitter user engagement, correlation analysis helps to identify how different factors such as tweet length, number of hashtags, mentions, and number of emojis relate to engagement metrics like likes, retweets, comments, and quotes. It helps in uncovering patterns that drives the user engagement. Data features were normalized, and correlation matrix was calculated to understand the relationship between different factors influencing user interactions.

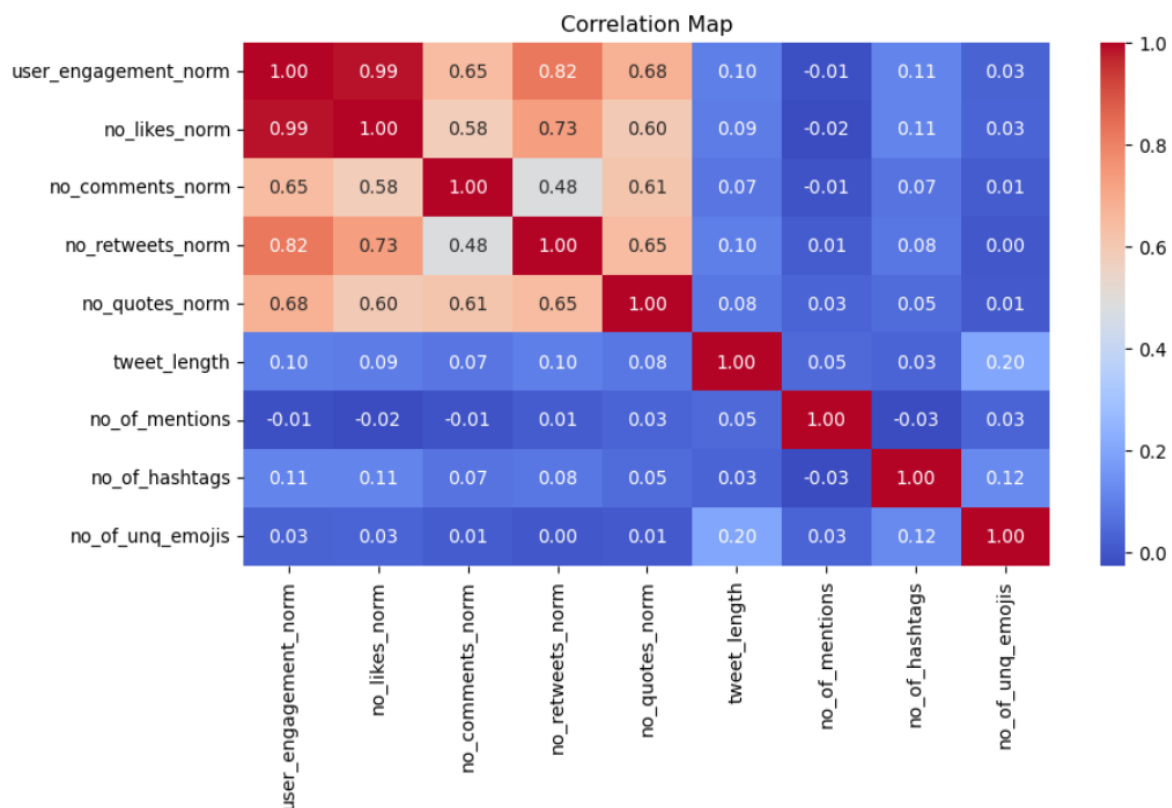


Figure19- Correlation Matrix





The correlation analysis shows the user engagement is very highly correlated with number of likes (0.99), number of retweets (0.82), number of quotes (0.68) and number of comments (0.65). It is also correlated with tweet length (0.10) and number of hashtags (0.11). The analysis also shows engagement has no relationship with number of emojis and number of mentions.

## 6 Conclusions and Next Steps

### 6.1 Conclusions and Summary

#### Key Findings

The comprehensive analysis of Twitter user engagement has provided valuable and actionable insights into the optimal strategies for maximizing user interaction with tweets. By examining various factors such as tweet length, hashtags, mentions, emojis, day of the week, and time of day, we can draw several key conclusions that can help to create effective content strategy on Twitter.

Tweet length affects how clear and concise the message is, influencing user interactions. Tweets that are brief and to the point tend to receive higher user engagement, highlighting the importance of clear and direct communication. Shorter tweets tend to have lower engagement, possibly due to lack of substantial content.

Hashtags are a fundamental component of Twitter. They enable tweets to reach a wider audience beyond the followers, increasing discoverability. While hashtags can enhance visibility, their effectiveness varies. The most frequently used hashtags showed lower average engagement compared to moderately used ones, indicating that strategic use of relevant hashtags is more beneficial than overloading tweets with them. Hashtags that are contextually relevant to the tweet content tend to drive higher engagement.

Mentions facilitate direct interaction with other users, promote conversations and connections. Tweets without mentions often have a more general and neutral tone, appealing to a broader audience without targeting specific users. These tweets may interact with a wider range of followers. Engagement on such tweets can be seen as more genuine, driven by the content itself rather than the involvement of mentioned users. This suggests that users might prefer content that is broadly relevant rather than content targeting specific individuals or groups.

Emojis convey emotions and tone, adding depth to the text. They make tweets more visually engaging and can attract user attention. Emojis that are contextually relevant to the tweet content tend to drive higher engagement. They complement the tweet's message enhance the emotional connection and user interaction. This suggests that inclusion of relevant emojis in tweets where required helps to maximise engagement, as they add visual appeal and emotional context to the content.



User activity on Twitter fluctuates throughout the week, with certain days showing higher or lower engagement. People's routines and availability for interacting with content can vary based on the day of the week. Analysis showed user engagement peaks on Saturdays and Sundays, even though the volume of tweets is lower on these days. This presents an opportunity for increased weekend posting to leverage higher user activity. Businesses can schedule important and critical tweets on weekends to maximise the user engagement.

Businesses tend to tweet more weekdays compared to weekends as the content creators and social media handlers work on weekdays. While there are more tweets posted on weekdays by business accounts, engagement is higher on weekends. This indicates that businesses can benefit from a more balanced posting schedule that includes weekends.

Users might be busier with work or school activities, affecting their ability to engage with content. They may have more free time to browse social media and interact with content outside of their work hours. Tweets posted between 6 AM and 8 AM experienced higher engagement. This is possibly due to users' morning routines and reduced competition for attention during these hours. Along with early mornings, early evenings also show increased engagement likely due to users get free time after finishing their work or school activities. To maximise engagement, scheduling tweets during inactive work hours can be more effective. High-value content, promotional posts, and interactive tweets might perform better during this period.

### Recommendations

User engagement can be maximized by creating brief and engaging tweets. Use relevant and targeted hashtags to boost visibility and avoid overused hashtags to maintain engagement. Use mentions selectively based on necessity to maintain broader appeal and avoid limiting engagement to specific groups. Use emojis to make tweets more engaging and expressive, capturing users' attention. Increase number of tweets on weekends to take advantage of higher user activity and engagement during these days. Schedule tweets for early morning hours between 6 AM to 8 AM and other high-engagement periods such as early evenings to maximize visibility and user interactions.

### Conclusion

This analysis brings out the importance of understanding user engagement patterns on Twitter to optimize content strategy on Twitter. By customizing tweet characteristics and posting schedules based on these insights, individuals and businesses can significantly enhance their social media strategy, driving higher engagement and can significantly enhance their reach and interaction on the platform. Implementing these recommendations will help optimize the use of Twitter as an effective platform for reaching and engaging with audiences, ultimately contributing to the success of business campaigns and initiatives on social media platform.



## 6.2 Future Work and Next Steps

Based on the comprehensive analysis conducted, several methods for future work can be outlined to further enhance understanding and optimization of Twitter user engagement. Data collection enhancement is necessary to collect more data which increases the robustness and reliability of the analysis. Explore the impact of different content types such as text, images, videos, links on engagement. Perform analysis of temporal patterns, considering seasonal and monthly variations in user engagement. Examine how external events such as holidays and news events influence engagement metrics. Analyse the impact of sentiment on engagement, identifying which types of emotional tones (positive, negative, neutral) drive the most interaction by using advanced sentiment analysis techniques, such as machine learning models such as BERT. Analyse how different topics influence engagement metrics and user interest by implementing topic modelling techniques such as Latent Dirichlet Allocation (LDA) to identify topics in tweets. Investigate posting same tweet multiple times by rephrasing during different periods of the day and day of the week impacts user engagement.

The outlined future work and next steps aim to deepen the understanding of Twitter user engagement and refine strategies for maximizing user interaction. By using advanced analytical techniques and expanding data collection, this project can provide actionable insights for optimizing content strategies. Implementing these recommendations will enhance the effectiveness of Twitter as a communication and marketing platform, driving higher engagement and achieving desired outcomes.





**Business and Local Government Data Research Centre**

Parkside 2C  
Knowledge Gateway  
University of Essex  
Wivenhoe Park  
Colchester CO4 3SQ

**T:** 01206 873859

**E:** [BLGDataResearch@essex.ac.uk](mailto:BLGDataResearch@essex.ac.uk)

**W:** [www.BLGdataresearch.org](http://www.BLGdataresearch.org)

 [@BLGDataResearch](https://twitter.com/BLGDataResearch)