

# Internship Report

## Twitter Data Analysis Project - NullClass

Name: Ankita Selokar

Tool: Power BI

Date: 03-08-2024 to 03-11-2024

### 1. Introduction

This report provides an overview of my work as a Data Analyst Intern at NullClass, where I utilised Power BI to analyse Twitter engagement data. The project involved cleaning data, designing an interactive dashboard, and completing specific visualisations to uncover patterns and insights in tweet performance.

### 2. Background

The project aimed to explore and interpret engagement metrics (e.g., retweets, likes, replies, media engagements) on tweets. The primary objective was to identify patterns in engagement based on time, media content, and user interactions, offering strategic insights for content optimization.

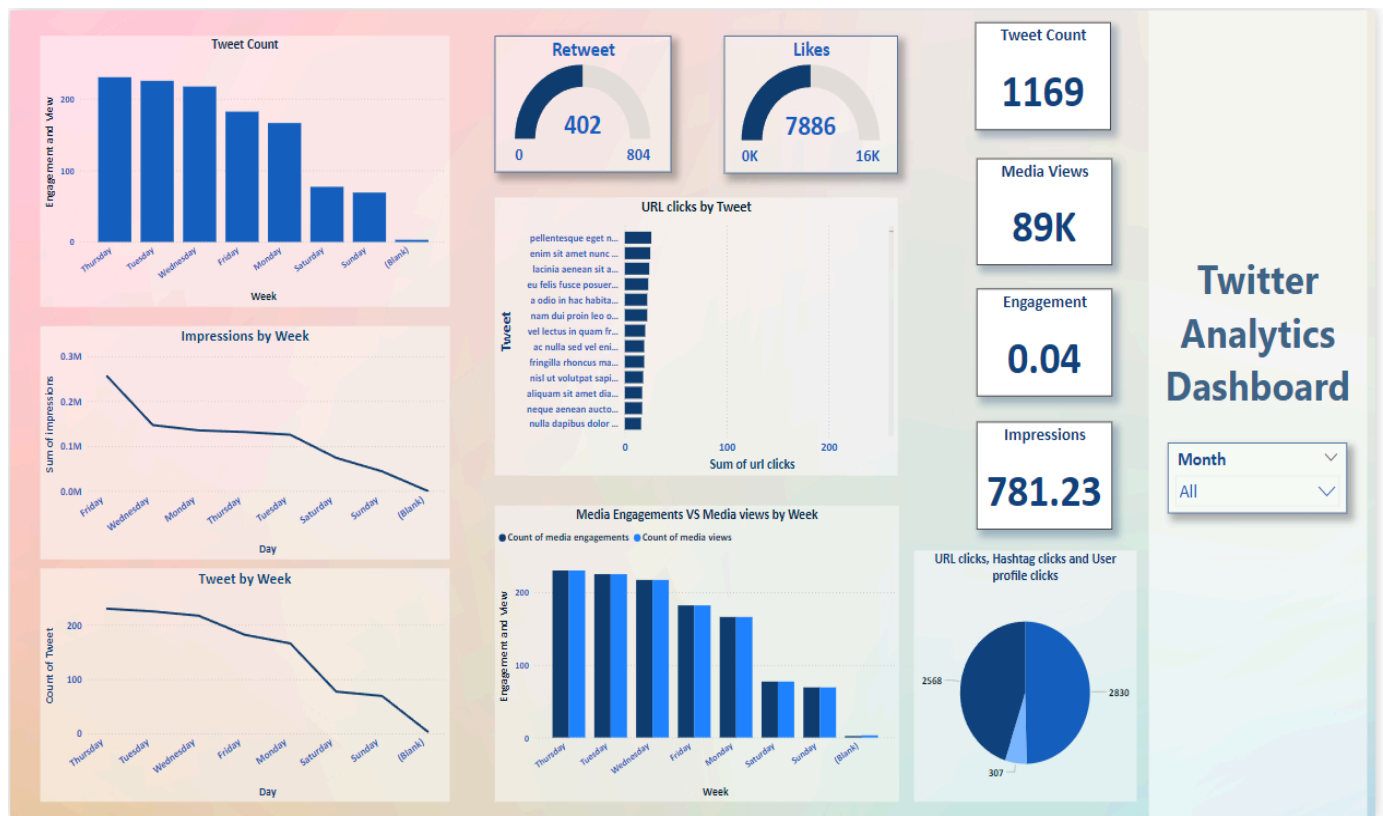
### 3. Learning Objectives

- Develop Power BI proficiency for data cleaning, visualisation, and analysis.
- Understand and analyse engagement trends across tweet content types.
- Generate insights for guiding content strategy through targeted data analysis.

### 4. Activities and Tasks

#### Class Activity:

I created this dashboard by following each step outlined in the tutorials, starting from data cleaning to dashboard creation using Power BI. This process has provided me with foundational knowledge of Power BI, enhancing my understanding of data visualization and analytics. The experience gained from these tasks will serve as a solid basis for future projects and analyses in my role as a Data Analyst Intern.



#### 4.1 Data Cleaning in Power Query Editor

- Ensured data consistency and accuracy by cleaning the dataset in Power Query Editor.
- Standardised data, filled missing values, and removed unnecessary fields, laying the foundation for reliable visualisations.

#### 4.2 Dashboard Creation

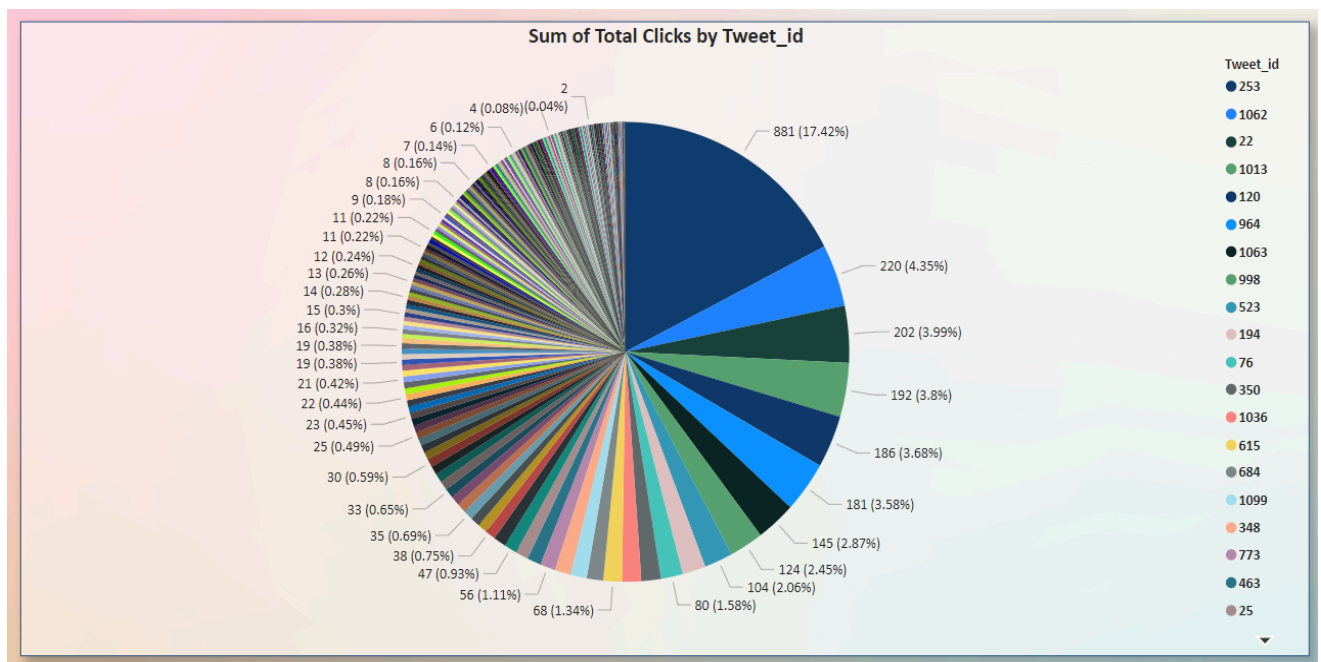
- Created an interactive Power BI dashboard featuring KPIs for total engagements, average engagement rates, and top-performing tweets.
- Designed a layout that allows users to easily navigate and interact with engagement insights.

#### 4.3 Visualisations

Completed the following seven tasks as specified:

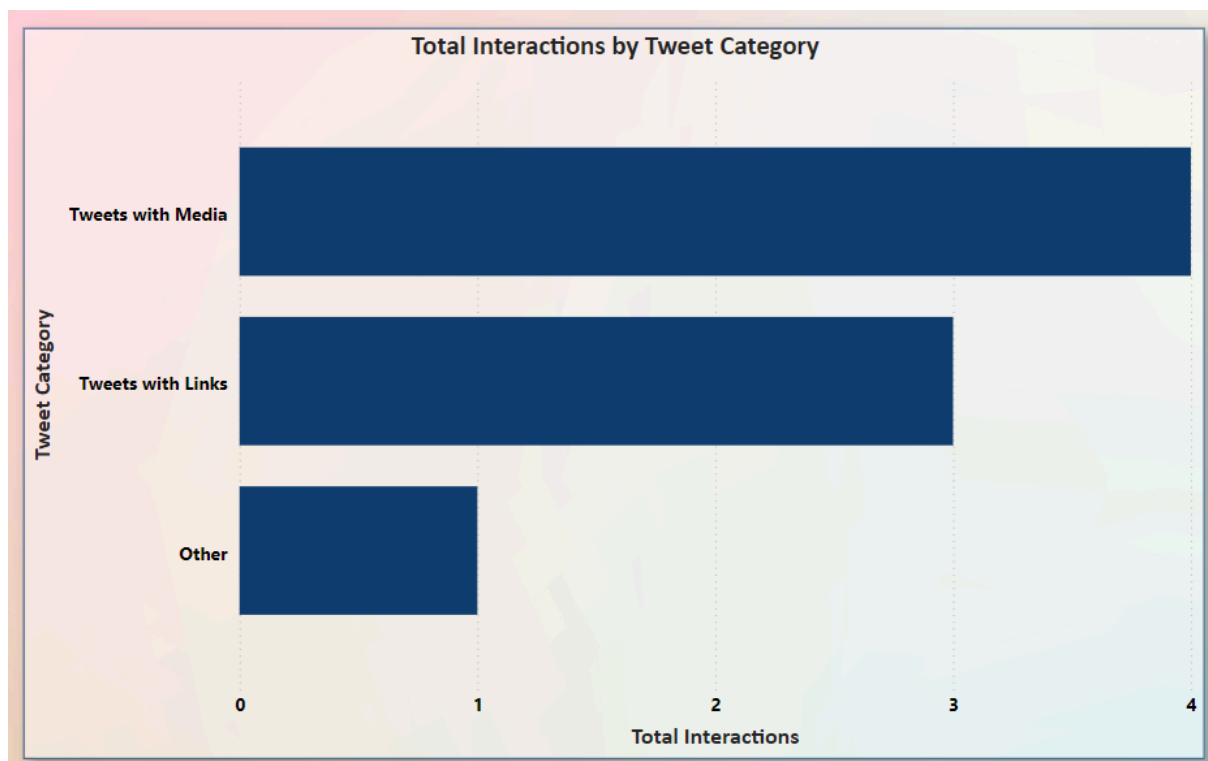
**1. Build a pie chart that represents the proportion of total clicks (URL clicks, user profile clicks, and hashtag clicks) for tweets with more than 500 impressions. Include a drill-down to view the specific types of clicks for each tweet**

I created a pie chart to display the proportion of URL clicks, user profile clicks, and hashtag clicks for tweets that have more than 500 impressions. I enabled a drill-down feature so that each segment can be clicked to reveal detailed click counts for individual tweets, helping to identify which types of engagements were most common within high-impression tweets. This visualisation provides an insightful breakdown of click interactions on popular tweets.



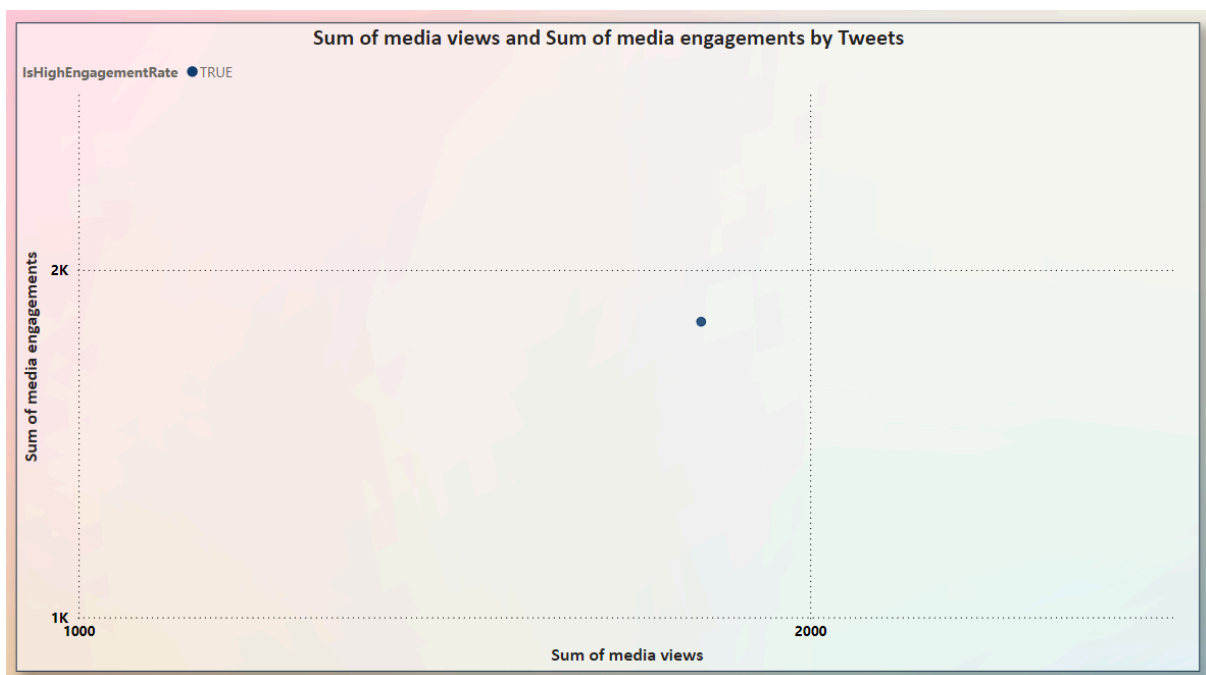
2. Create a clustered bar chart that breaks down the sum of URL clicks, user profile clicks, and hashtag clicks by tweet category (e.g., tweets with media, tweets with links, tweets with hashtags). Only include tweets that have at least one of these interaction types and this graph should work between 3 PM to 6 PM and the tweet date should be an even number as well as tweet word count be below 40.

I created a clustered bar chart to show the sum of URL clicks, user profile clicks, and hashtag clicks across tweet categories (media, links, hashtags). I applied filters for tweets posted between 3 PM and 6 PM on even-numbered dates with a word count below 40, focusing on high-engagement patterns across tweet types.



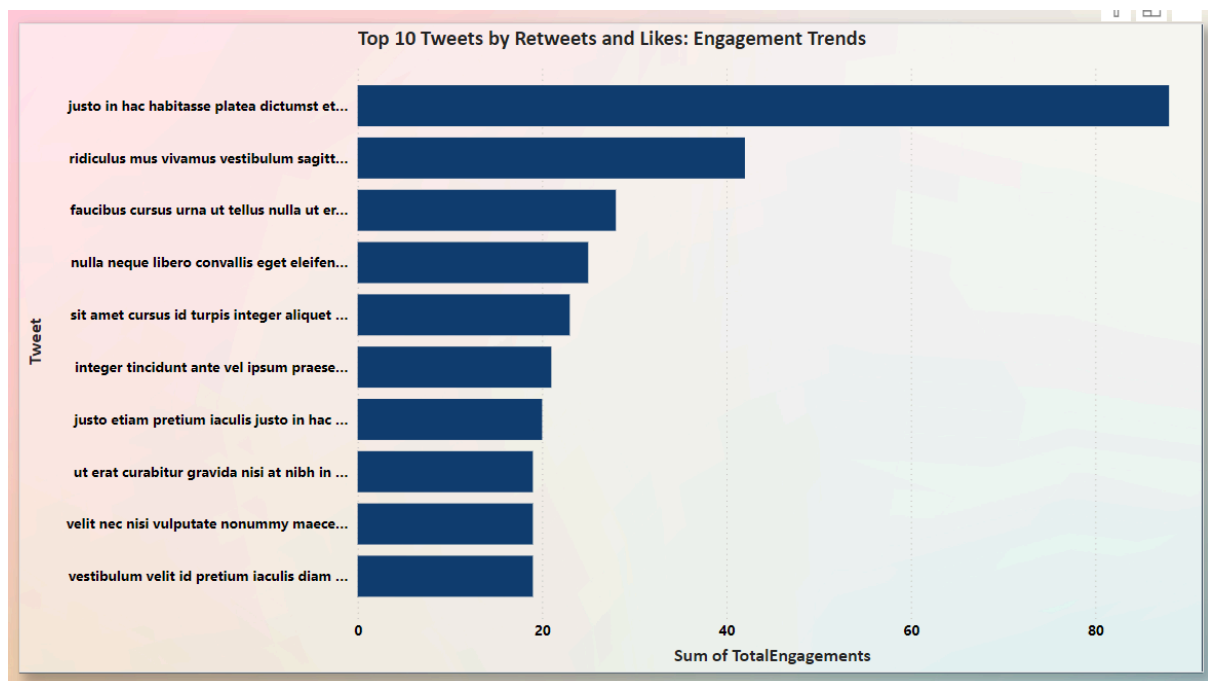
**3. Plot a scatter chart to analyse the relationship between media engagements and media views for tweets that received more than 10 replies. Highlight tweets with an engagement rate above 5% and this graph should work only between 12 PM to 6 PM and the tweet date should be odd number as well as tweet word count be below 50.**

I plotted a scatter chart to analyse the relationship between media engagements and media views for tweets with over 10 replies. I highlighted tweets with an engagement rate above 5% and applied filters for tweets posted between 12 PM and 6 PM on odd-numbered dates with a word count under 50, allowing insights into high-engagement tweets during specific times.



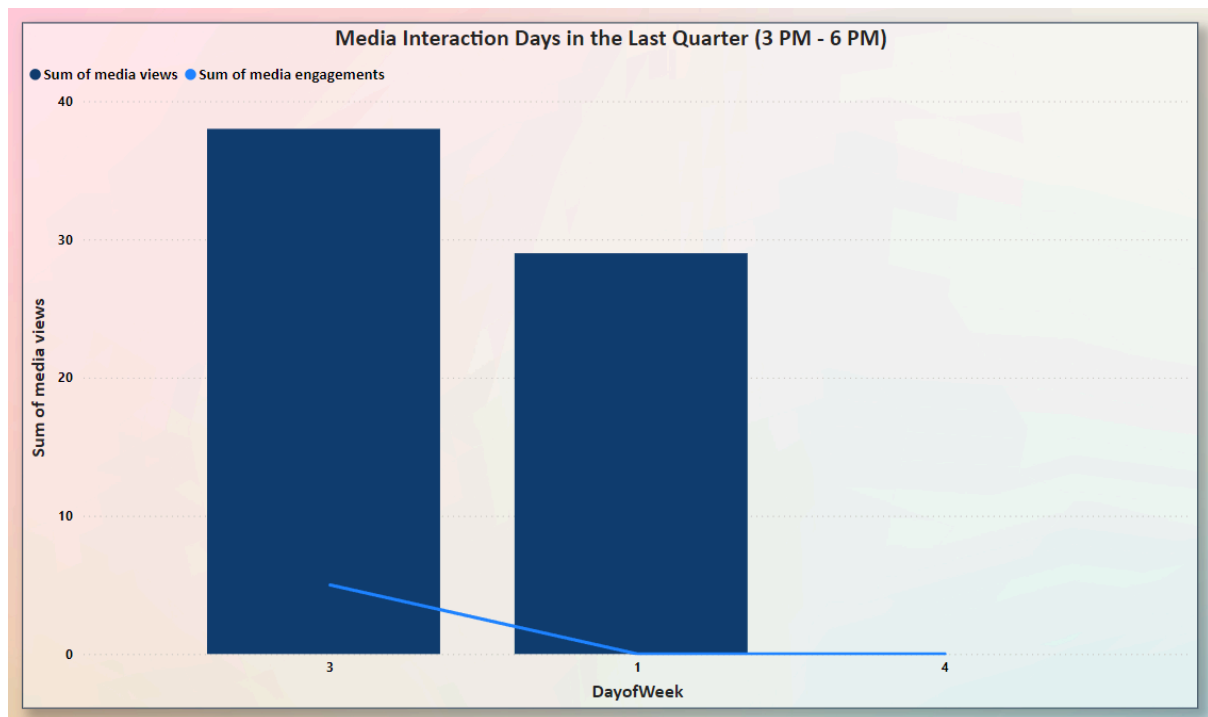
4. Build a chart to identify the top 10 tweets by the sum of retweets and likes. Filter out tweets posted on weekends and show the user profile that posted each tweet and this graph should work between 3 PM to 6 PM and the tweet impression should be even number and tweet date should be odd number as well as tweet word count be below 30

I built a chart to display the top 10 tweets based on the combined sum of retweets and likes, excluding tweets posted on weekends. The chart includes filters for tweets posted between 3 PM and 6 PM, with even impressions, odd-numbered tweet dates, and a word count below 30, enabling a focused view on high-engagement tweets within specific conditions.



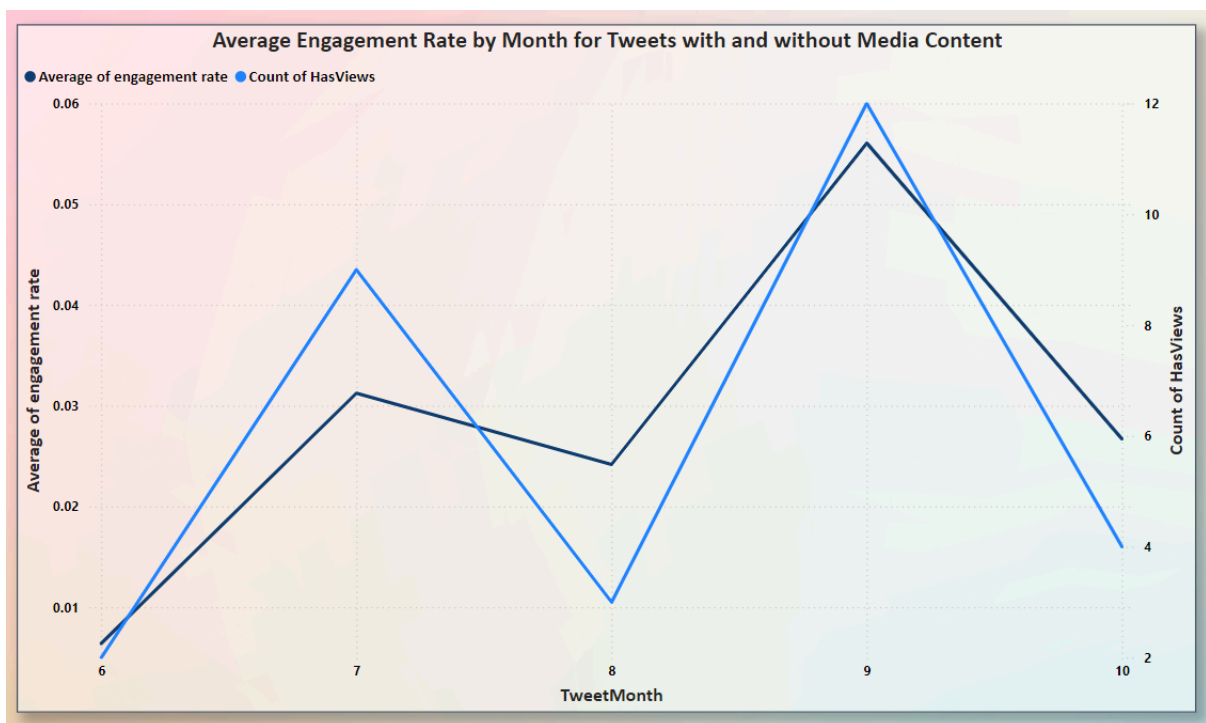
**5. Create a dual-axis chart that shows the number of media views and media engagements by the day of the week for the last quarter. Highlight days with significant spikes in media interactions. This graph should work between 3 PM to 6 PM and the tweet impression should be even number and tweet date should be odd number as well as the tweet word count be below 30.**

I created a dual-axis chart to show media views and media engagements by day of the week for the last quarter, highlighting days with significant spikes in interactions. The chart includes filters for tweets posted between 3 PM and 6 PM, with even-numbered impressions, odd-numbered tweet dates, and word counts below 30, allowing for a detailed analysis of engagement patterns across the week.



**6. Create a line chart showing the trend of the average engagement rate over each month of the year. Separate the lines for tweets with media content and those without and this graph should work between 3 PM to 6 PM and the tweet engagement should be even number and tweet date should be odd number**

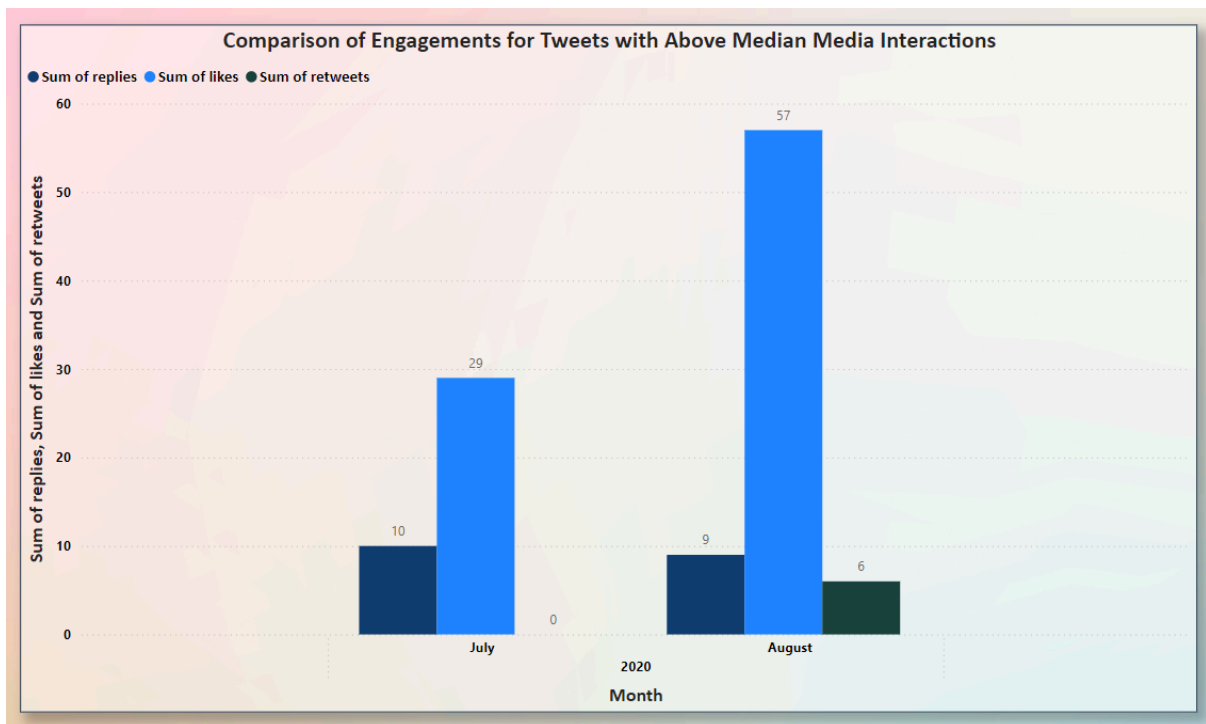
I developed a line chart to display the trend in average engagement rate over each month of the year, with separate lines distinguishing tweets with media content from those without. The chart is filtered to show data for tweets posted between 3 PM and 6 PM, with even-numbered engagements, odd-numbered tweet dates, and includes only the specified tweet criteria. This setup provides insight into engagement trends for both media and non-media tweets month by month.





**7. Develop a visualisation that compares the number of replies, retweets, and likes for tweets that have received media engagements greater than the median value. Include a filter for tweets posted in between June and August of 2020 and this graph should work between 3 PM to 6 PM and tweet date should be odd number and media views should be even number as well as tweet word count be below 50.**

I created a visualisation to compare the number of replies, retweets, and likes for tweets that received media engagements above the median value. This graph includes a filter for tweets posted between June and August of 2020 and is set to show only those posted between 3 PM and 6 PM, with odd-numbered tweet dates, even media views, and a word count of less than 50. This approach effectively highlights engagement metrics for high-performing tweets during the specified timeframe.



## 5. Skills and Competencies

**Data Cleaning and Preparation:** Refined Power Query skills, ensuring data accuracy.

**Dashboard Design:** Developed and optimised layouts for easy interaction and clear data storytelling.

**Complex Filtering:** Applied advanced filter conditions in Power BI, including specific tweet time and engagement criteria.

**DAX Expressions:** Created custom measures and calculations to support analysis, such as median engagement and engagement rate.

**Data Visualization:** Enhanced my ability to choose effective chart types for data presentation.

## 6. Feedback and Evidence

This internship has significantly strengthened my Power BI skills and my confidence in data analysis. I gained practical experience in managing complex filters and designing visuals that communicate insights clearly. The exposure to DAX for custom calculations was particularly beneficial, helping me build precise, effective measures to support analysis.

### **Evidence:**

Screenshots of each visualisation were submitted along with annotations to demonstrate functionality and adherence to task requirements.

## 7. Challenges and Solutions

**Complex Filtering Requirements:** Filtering multiple conditions simultaneously posed challenges. I resolved these by using calculated columns and custom measures to streamline the filter criteria.

**Data Aggregation Errors:** Encountered aggregation issues with certain metrics, which I addressed by adjusting DAX expressions and revisiting Power Query settings.

**Handling Missing Data:** Addressed missing values by filling in defaults or omitting specific records where necessary to maintain analysis consistency.

## 8. Outcomes and Impact

**Strategic Insights:** Provided clarity on optimal engagement times and the performance of tweets with media content, aiding in strategic content planning.

**Skill Development:** This project allowed me to build strong analytical and technical skills in Power BI, including data cleaning, visualisation, and DAX measures.

**Data-Driven Decision Support:** The dashboard offers valuable insights to guide engagement strategy and maximise tweet reach.

## **9. Conclusion**

The Twitter engagement project provided an in-depth experience in data analysis with Power BI, from data preparation to advanced visualisation techniques. By completing the internship, I gained the technical skills and practical experience needed for roles that require data-driven insights and effective communication of complex data.