**INTERNSHIP REPORT**

Real Time Twitter Analytics Dashboard

**Name:** Mohan S

**Project:** Real Time Twitter Analytics Dashboard

**Internship Duration:** 18-04-2025 to 18-06-2025

**INTRODUCTION**

This internship report highlights the key contributions, learning, and experience gained during my internship focused on data analytics and dashboard development using Microsoft Power BI. The primary objective of the internship was to build interactive and dynamic visualizations based on specific business logic and real-world analytical conditions using DAX formulas and Power BI functionalities.

Over the course of the internship, I worked with real-time social media datasets derived from Twitter, which contained valuable metrics such as likes, retweets, replies, impressions, app opens, and media engagements. These datasets provided the foundation to analyze audience behavior, post performance, and engagement trends using advanced data modeling techniques.

The core aim of the project was to design a fully functional, logic-based Power BI dashboard that adapts to various time constraints (especially Indian Standard Time), tweet characteristics, and conditional filters. This involved rigorous application of DAX measures to control what visualizations appear under which conditions, such as specific time intervals, content length thresholds, and text-based filters (e.g., removing tweets containing specific characters like 'D' or 'S').

I also gained hands-on experience in **Power Query**, where I performed essential data cleaning tasks including handling null values, transforming text, extracting dates, and building new computed fields like tweet word count, character count, and time zones. These transformations enabled precise filtering and meaningful data storytelling throughout the report.

A major focus of the internship was not just on static reporting, but on implementing **smart dashboards**—charts and visuals that react dynamically to live conditions such as current time and contextual content rules. This allowed me to simulate real-world business reporting scenarios where decision-makers often require filtered, contextualized, and timely insights.

This internship allowed me to build a strong foundation in data analytics, DAX programming, conditional visual logic, and Power BI dashboarding techniques. It also enhanced my problem-solving abilities and helped me understand the intersection of business logic and data visualization.

**BACKGROUND AND DATA INFO**

The internship involved working with a real-world social media dataset that included tweet-level metrics such as likes, retweets, replies, engagement rates, impressions, and time-based data. I was required to develop custom visualizations, apply advanced DAX filtering, and use Power Query to clean and preprocess data.

**TABLE - SocialMedia**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| app installs | Numeric (Sum) | No. of app installs from tweet |
| app opens | Numeric (Sum) | App opens driven by the tweet |
| CleanedTweet\_NoD | Text (Calc.) | Tweet text with words containing 'D' removed |
| CleanedTweet\_NoS | Text (Calc.) | Tweet text with words containing 'S' removed |
| ContainsWordWithD | Boolean | TRUE if tweet contains word with 'D' |
| ConvertedDateTime | DateTime | Timestamp converted to local IST time |
| Date (Subtable) | Hierarchy | Includes Year, Quarter, Month, Day |
| detail expands | Numeric (Sum) | No. of detail view expands |
| dial phone | Numeric (Sum) | Dial phone clicks |
| email tweet | Numeric (Sum) | Tweet shares via email |
| engagement rate | Numeric | Calculated engagement per tweet |
| engagements | Numeric (Sum) | Total engagements on tweet |
| follows | Numeric (Sum) | New followers via tweet |
| HasAppOpens | Boolean | TRUE if tweet has any app opens |
| hashtag clicks | Numeric (Sum) | No. of hashtag clicks |
| id | Numeric | Unique tweet ID |
| impressions | Numeric (Sum) | Number of tweet views |
| IsEvenImpression | Boolean | TRUE if impression count is even |

**Custom Logic / Calculated Columns**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Type** | **Description** |
| IsOddDate | Boolean | TRUE if tweet date is odd |
| IsTweetBetween9to5 | Boolean | TRUE if tweet time is between 9 AM to 5 PM |
| media engagements | Numeric | Engagements on media content |
| media views | Numeric | Views on media |
| Month | Numeric/Text | Month of tweet |
| Original Time | DateTime | Original UTC time |
| permalink clicks | Numeric | Clicks on permalink |
| Quarter | Numeric/Text | Quarter of the year |
| RemoveIfContainsWordWithD | Text | Tweet with words containing "D" removed |
| replies | Numeric | Reply count |
| retweets | Numeric | Retweet count |
| ShowChart, ShowChart2, ShowChart3 | Boolean | Dynamic chart visibility flags |
| time | Time | Extracted time field |
| Top10PercentFlag | Boolean | Flag if tweet is in top 10% engagement |
| TotalRetweetsLikes | Numeric | Sum of retweets + likes |

**SOFTWARE AND HARDWARE SPECIFICATION**

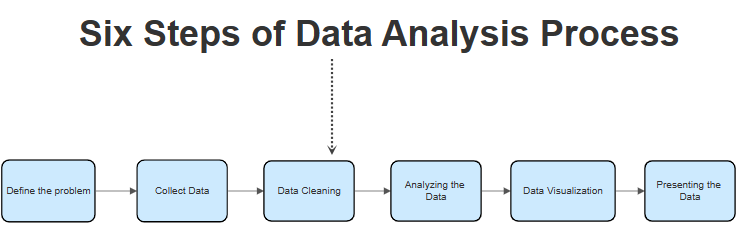
|  |  |
| --- | --- |
| **Hardware Requirements** | |
| Processor | Intel i5 or higher |
| Ram | 8 GB |
| Storage | 256 GB |

|  |  |
| --- | --- |
| **Software Requirements** | |
| OS | Windows 11 |
| Tool Used | Excel, Power BI |
| Data file | CSV |

**LEARNING OBJECTIVES**

* Understand how to build conditional dashboards using Power BI that respond dynamically to time, data, and user-defined filters.
* Learn to write advanced DAX formulas for time filtering, aggregation, conditional visibility, and logic-based chart control.
* Explore Power Query for text transformations, such as filtering tweets based on word content and removing unwanted characters or phrases.
* Gain hands-on experience in building dashboards that automatically hide or show visuals depending on business rules, such as time of day or tweet metadata.
* Develop the ability to handle real-time datasets and apply business logic to extract actionable insights.
* Learn to implement data validation rules and calculated columns that improve report interactivity and usability.
* Understand the structure of the Power BI data model, including relationships, fact and dimension tables, and column optimization.
* Build reusable DAX measures to drive KPIs and visual elements with maximum performance and clarity.
* Practice creating professional, presentation-ready dashboards with enhanced UI/UX principles and clear storytelling.
* Improve problem-solving skills by translating complex business rules into precise data logic using Power BI's features.

**STEPS IN DATA ANALYSIS PROCESS**



1. Data Collection: Imported tweet-level metrics from a CSV dataset into Power BI.

2. Data Cleaning: Used Power Query to handle missing values, duplicate rows, and convert date/time formats.

3. Time Zone Handling: Converted UTC time to IST using DAX logic (UTC + 5:30).

4. Feature Engineering: Created columns such as word count, character count, odd/even date checks, and impression filters.

5. Filtering & Transformation: Applied Power Query transformations for text filtering and Power BI filters for visual-level filtering.

6. DAX Measures: Developed dynamic DAX formulas to restrict visuals by current time and conditions.

7. Dashboard Layout: Organized the dashboard with bookmarks and visibility controls to display charts based on logic.

8. Testing & Validation: Cross-checked results with problem requirements and fine-tuned based on feedback.

**ACTIVITIES AND TASKS**

**PROBLEM STATEMENTS**

1. Develop a chart that displays tweets with the highest engagement rates (top 10%). Include only those tweets that have received more than 50 likes and were posted on weekdays and this graph should work only between 3PM IST to 5 PM IST apart from that time we should not show this graph in dashboard itself as well as tweet character count should be below 30.  
  
2. 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 6PM IST to 11 PM IST apart from that time we should not show this graph in dashboard itself and the tweet date should be odd number as well as tweet word count be above 50.  
  
3. 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 only between 3PM IST to 5 PM IST apart from that time we should not show this graph in dashboard itself and the tweet impression should be even number and tweet date should be odd number as well as tweet word count be below 30  
  
4. Analyse tweets to show a comparison of the engagement rate for tweets with app opens versus tweets without app opens. Include only tweets posted between 9 AM and 5 PM on weekdays andthis graph should work only between 12PM IST to 6PM IST and 7 AM to 11AM apart from that time we should not show this graph in dashboard itself and the tweet impression should be even number and tweet date should be odd number as well as tweet character count should be above 30 and need to remove tweet word which has letter 'D'.  
  
5. Develop a visualization 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 only between 3PM IST to 5 PM IST and 7 AM to 11AM apart from that time we should not show this graph in dashboard itself and tweet date should be odd number and media views should be even number as well as tweet character count should be above 20 and need to remove tweet word which has letter 'S'.

**SOLUTIONS**

I developed five main visualizations:

1. A chart showing top 10% tweets by engagement rate filtered by time (3 PM to 5 PM IST), likes > 50, weekday filter, and tweet character count < 30.

2. A scatter chart comparing media engagements vs media views for tweets with >10 replies and word count > 50, limited to 6 PM to 11 PM IST.

3. A chart showing top 10 tweets by sum of likes and retweets filtered for weekdays, tweet impression being even, date being odd, and word count < 30.

4. A bar chart comparing engagement rate for tweets with and without app opens, filtered for tweets posted between 9 AM to 5 PM on weekdays and displayed only between 12 PM to 6 PM IST and 7 AM to 11 AM IST. Also included filters for tweet date being odd, impressions even, and character count > 30, while removing words containing “D”.

5. A comparison chart of replies, retweets, and likes for tweets with media engagements greater than the median value. Filters included date range (June to August 2020), media views even, tweet date odd, tweet character count > 20, and removal of words with “S”.

**SKILLS AND COMPETENCIES**

Proficiency in Power BI report creation - Writing advanced DAX queries - Power Query for string and word manipulation - Data cleaning and transformation techniques - Time-based filtering using IST conversions.

**CHALLENGES AND SOLUTIONS**

Challenge: Applying time filters dynamically in IST timezone Solution: Created DAX measures using UTC + TIME (5,30,0) to reflect IST. - Challenge: Word-based filtering inside tweets in DAX Solution: Switched to Power Query for more robust string manipulation. - Challenge: Implementing complex multi-condition filters Solution: Broke down requirements into reusable calculated columns and measures for clarity and efficiency.

**All Dax Measures Created:**

* Cleaned Tweet Without D

CleanedTweet\_NoD =

VAR CleanText = SUBSTITUTE([Tweet], ".", "")

RETURN

    CONCATENATEX(

        FILTER(

            ADDCOLUMNS(

                SELECTCOLUMNS(

                    GENERATESERIES(1, PATHLENGTH(SUBSTITUTE(CleanText, " ", "|"))),

                    "WordIndex", [Value]

                ),

                "Word",

                PATHITEM(SUBSTITUTE(CleanText, " ", "|"), [WordIndex])

            ),

            SEARCH("d", LOWER([Word]), 1, 0) = 0

        ),

        [Word], " "

    )

* Cleaned Tweet Without S

CleanedTweet\_NoS =

VAR CleanText = SUBSTITUTE([Tweet], ".", "")

RETURN

    CONCATENATEX(

        FILTER(

            ADDCOLUMNS(

                SELECTCOLUMNS(

                    GENERATESERIES(1, PATHLENGTH(SUBSTITUTE(CleanText, " ", "|"))),

                    "WordIndex", [Value]

                ),

                "Word",

                PATHITEM(SUBSTITUTE(CleanText, " ", "|"), [WordIndex])

            ),

            SEARCH("s", LOWER([Word]), 1, 0) = 0

        ),

        [Word], " "

    )

ContainsWordWithD =

IF(

    CONTAINSSTRING(LOWER([Tweet]), "d"),

    1,

    0

)

* Median Media Engagement

IsMediaEngagementAboveMedian =

VAR MedianVal = MEDIANX(ALL('SocialMedia'), [media engagements])

RETURN IF([media engagements] > MedianVal, TRUE(), FALSE())

* Even Impressions

IsEvenImpression = IF(MOD([impressions], 2) = 0, 1, 0)

* Odd Date

IsOddDate =

IF(

    MOD(DAY('SocialMedia'[Date]), 2) = 1,

    1,

    0

)

* Post Time between 9 to 5

IsTweetBetween9to5 =

VAR PostTime = [Original Time]  -- Now a real TIME type

RETURN

    IF(PostTime >= TIME(9, 0, 0) && PostTime <= TIME(17, 0, 0), 1, 0)

* Graph Display Time Settings

ShowChart =

VAR CurrentIST = NOW()

VAR CurrentTime = TIME(HOUR(CurrentIST), MINUTE(CurrentIST), SECOND(CurrentIST))

RETURN

    IF(

        CurrentTime >= TIME(15, 0, 0) && CurrentTime <= TIME(17, 0, 0),

        1,

        0

    )

* Top 10 Percentage Post

Top10PercentFlag =

VAR Threshold =

    PERCENTILEX.INC(ALL('SocialMedia'), [Engagement Rate], 0.9)

RETURN

    IF([Engagement Rate] >= Threshold, 1, 0)

* Tweet Character Count

TweetCharCount = LEN(TRIM([Tweet]))

* Tweet Word Count

TweetWordCount =

IF(

    ISBLANK([Tweet]),

    0,

    LEN(TRIM([Tweet])) - LEN(SUBSTITUTE(TRIM([Tweet]), " ", "")) + 1

)

* Even Media Views

IsEvenMediaViews = MOD([media views], 2) = 0

**FEEDBACK AND EVIDENCE**

Throughout the internship, I received continuous feedback on query optimization, use of custom measures, and better data modeling. All visualizations were built and tested based on functional requirements and screenshots were shared for evaluation.

Dashboard 1:

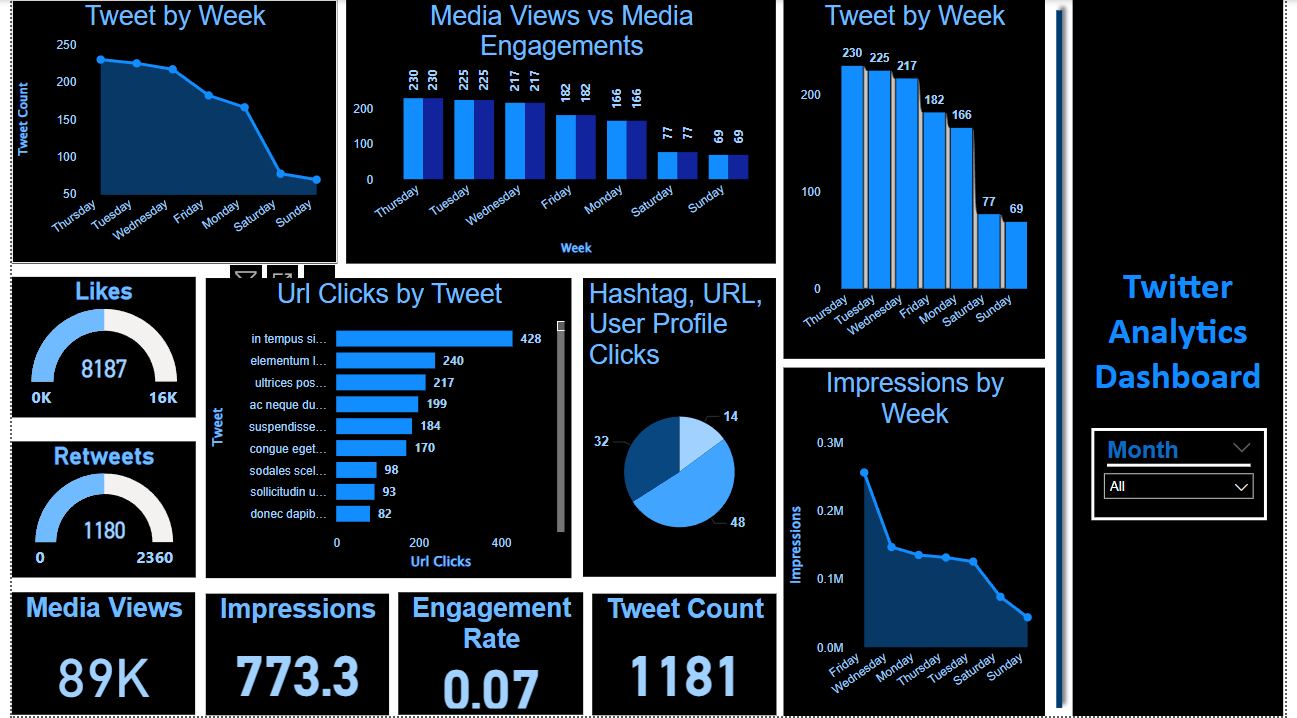


Chart 1: Top 10% Tweets by Engagement Rate (Visible 3 PM – 5 PM IST)

- Filters: Likes > 50, weekdays, tweet character count < 30.

- Displayed only in the allowed time window using a DAX filter.

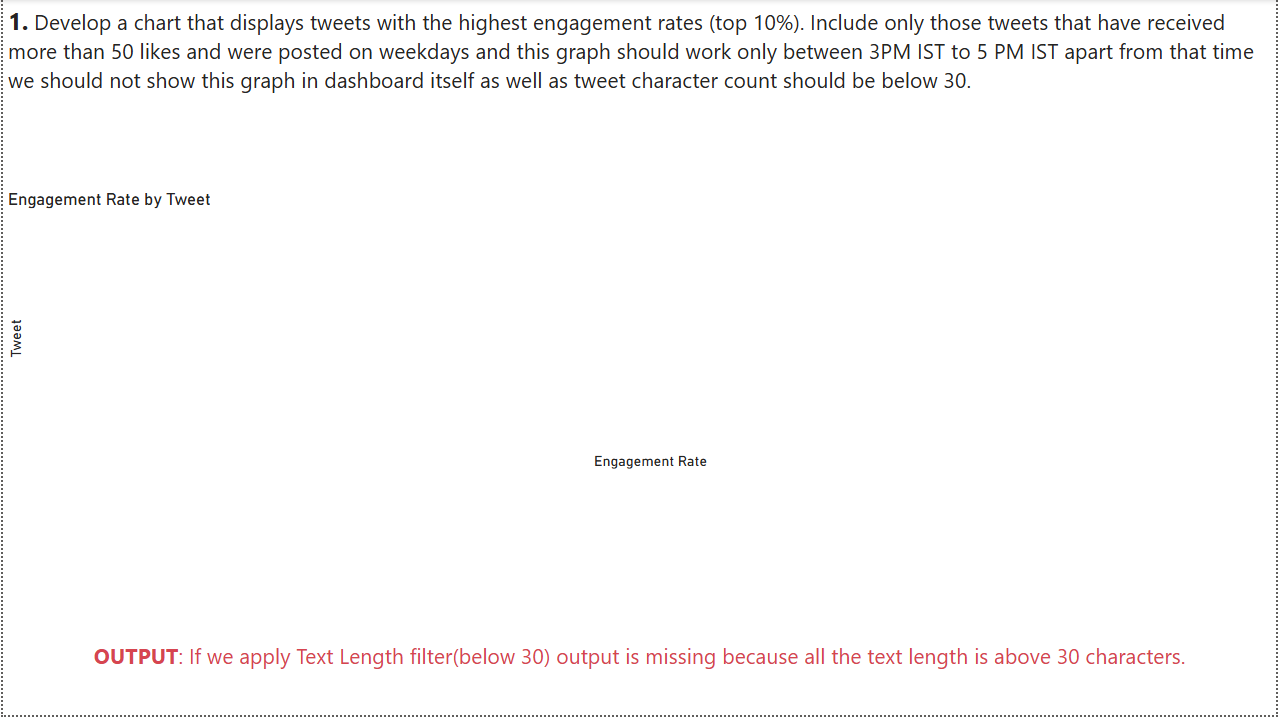


Chart 2: Media Engagement vs Media Views (Visible 6 PM – 11 PM IST)

- Filters: Tweets with >10 replies, word count > 50, odd dates.

- Points above 5% engagement rate are highlighted.

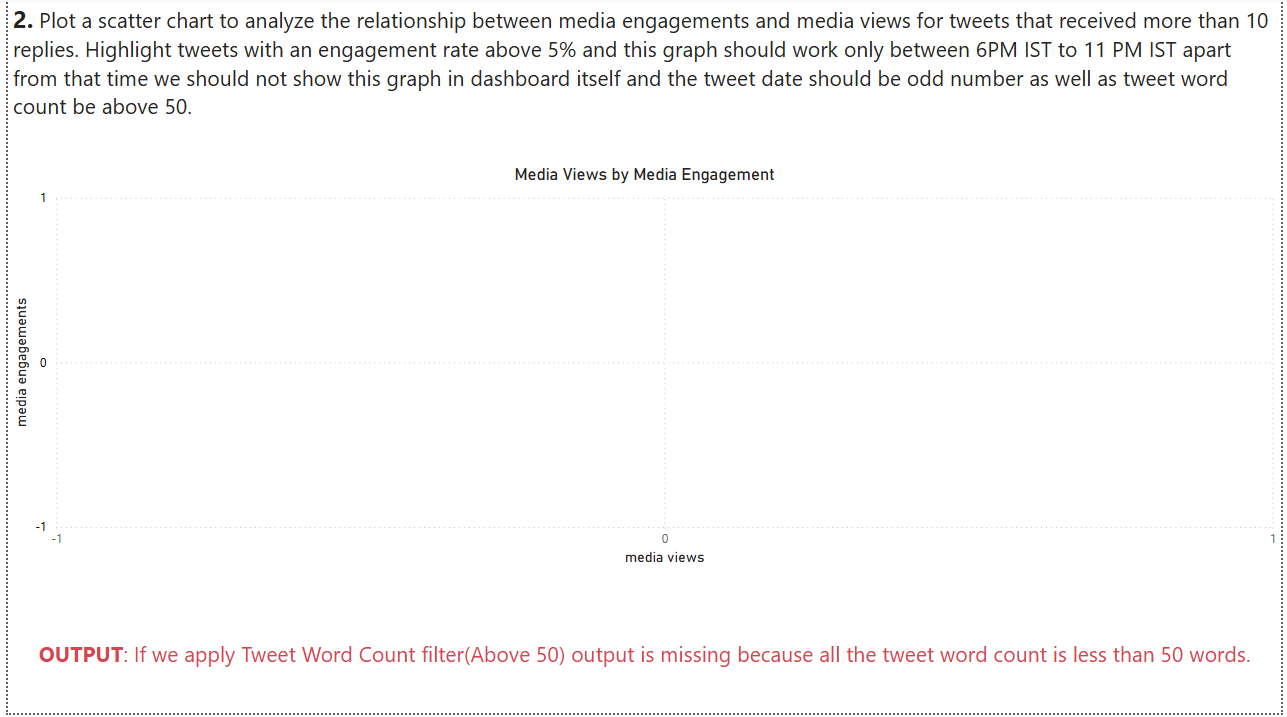


Chart 3: Top Tweets by Retweets + Likes (Visible 3 PM – 5 PM IST)

- Filters: Weekday tweets, impressions even, date odd, word count < 30.

- Each tweet linked to its user profile for identification.

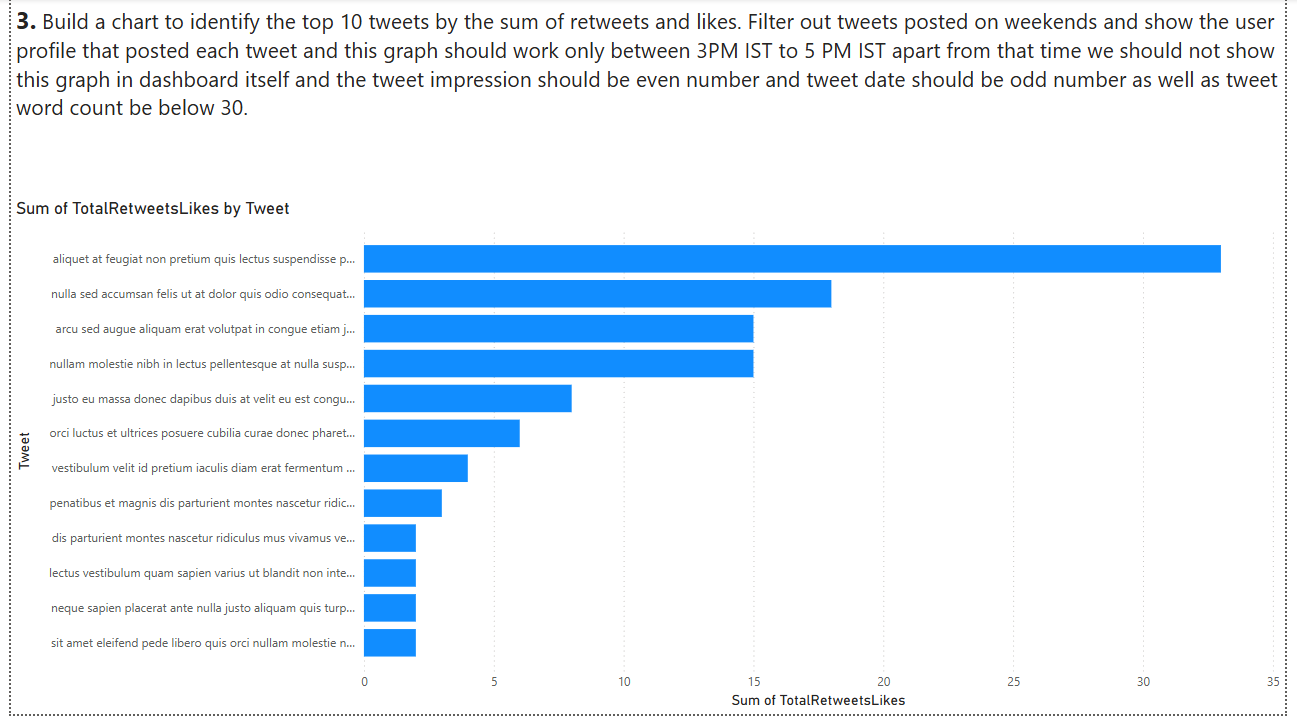


Chart 4: App Open Comparison (Visible 12 PM – 6 PM & 7 AM – 11 AM IST)

- Filters: Weekday tweets 9 AM – 5 PM only, impressions even, date odd, char count > 30.

- Removed tweet words containing ‘D’.

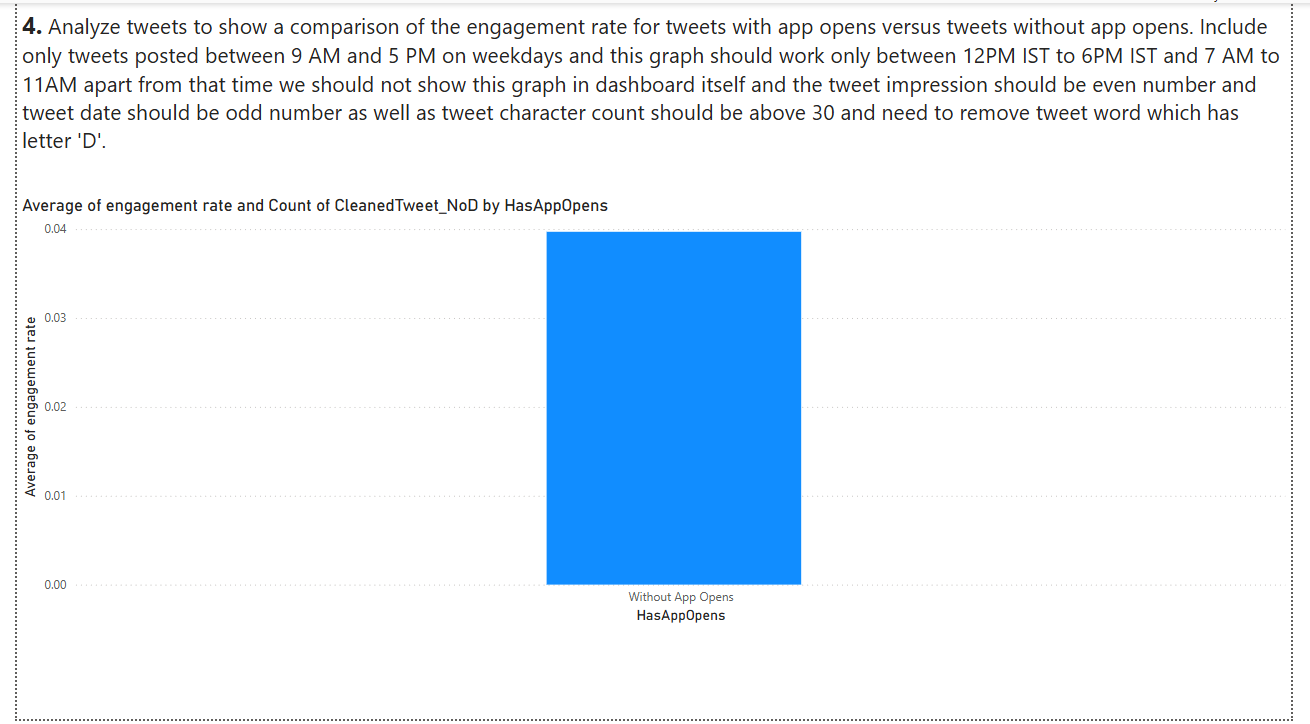
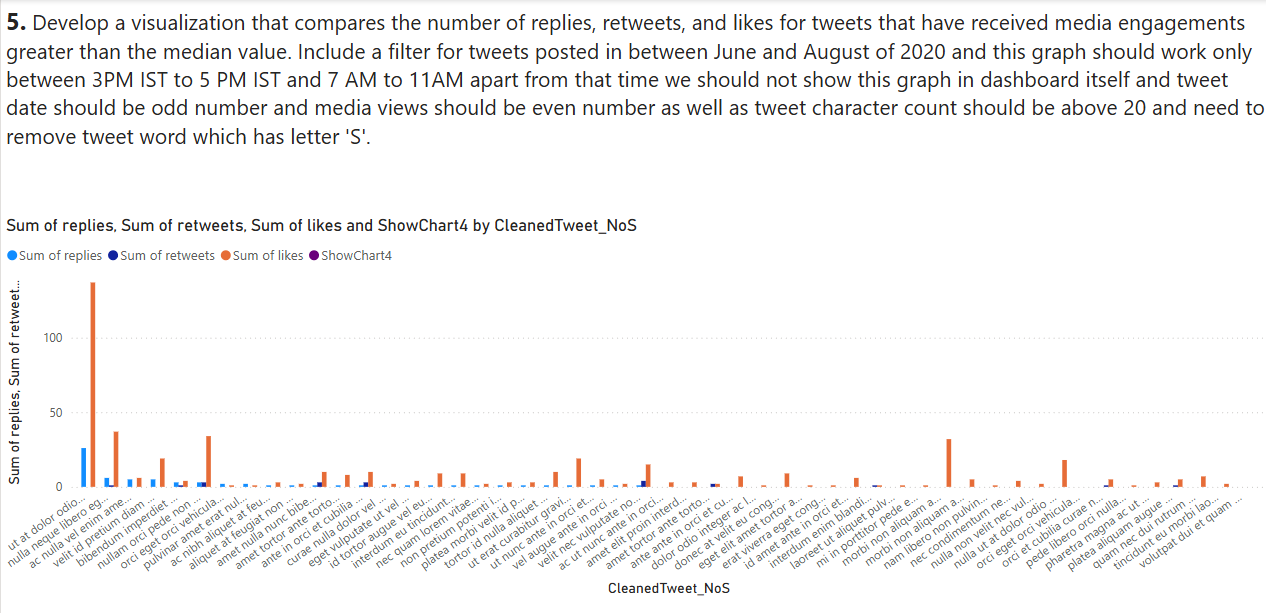


Chart 5: Engagement Comparison for High Media Engagement Tweets (June–August 2020)

- Filters: Media engagements > median, media views even, odd date, char count > 20.

- Words containing ‘S’ were removed using Power Query.



**OUTCOMES AND IMPACT**

The outcome of the internship was a fully interactive, condition-based Power BI dashboard capable of dynamically displaying or hiding charts based on the current time, tweet content, and user-defined conditions. The solutions are scalable and reusable in real-world business environments.

* Created complex time-based conditional visualizations using DAX.
* Mastered Power Query for advanced string and date transformations.
* Gained expertise in building custom measures and logic-driven dashboards.
* Learned to apply IST filtering by dynamically calculating current time and matching conditions.
* Developed reusable calculated columns for filtering and logic clarity.

**CONCLUSION**

This internship enhanced my technical capabilities in Power BI, DAX, and data modeling. I now feel more confident in handling complex dashboard requirements, creating efficient and clean code, and applying business logic into visual storytelling through data analytics.