NullClass Internship Report for Task 2

By Chinta Anagha Sree

Introduction

Social media platforms have become an integral part of contemporary communication, influencing public opinion, brand strategies, and personal interactions. These platforms generate vast amounts of data daily, offering valuable insights into user behavior and content performance. Analyzing this data effectively can reveal trends, measure engagement, and guide strategic decisions. During my internship, I was tasked with a project aimed at visualizing social media data to identify high-engagement tweets. The objective was to build a chart to pinpoint the top 10 tweets based on the sum of retweets and likes. To ensure the relevance of the data, the analysis required filtering out tweets posted on weekends, including only those with even impressions, odd tweet dates, and word counts below 30. Additionally, the graph needed to display user profiles and work only between 3 PM and 6 PM based on the system's local time. This task involved a combination of data transformation, visualization techniques, and time-based conditions to create a meaningful and actionable dashboard.

Background

Social media analytics has become essential for organizations and individuals seeking to measure the effectiveness of their online content. Twitter, being a platform with global reach, generates data at an enormous rate, which requires tools like Power BI to analyze and derive insights. Tracking engagement rates, such as likes and retweets, is crucial in understanding which tweets perform best and why. This helps in identifying trends, improving content strategies, and engaging more effectively with the target audience.

I worked on this task aimed at optimizing the analysis of social media engagement metrics. The primary focus was to develop a dashboard that could identify high-engagement tweets while applying multiple filters. The data was sourced from a CSV file containing various engagement metrics such as retweets, likes, impressions, and more. The challenge was to create a dynamic and interactive dashboard in Power BI that could not only display the top-performing tweets but also adhere to several specific conditions. These conditions included filtering out tweets based on their post date, impression count, and word count. Moreover, the dashboard needed to be operational only within a specified time frame (3 PM to 6 PM) relative to the local system time. This task required a deep understanding of Power BI's capabilities, including advanced filtering, custom measures, and dynamic time-based visualizations.

Learning Objectives

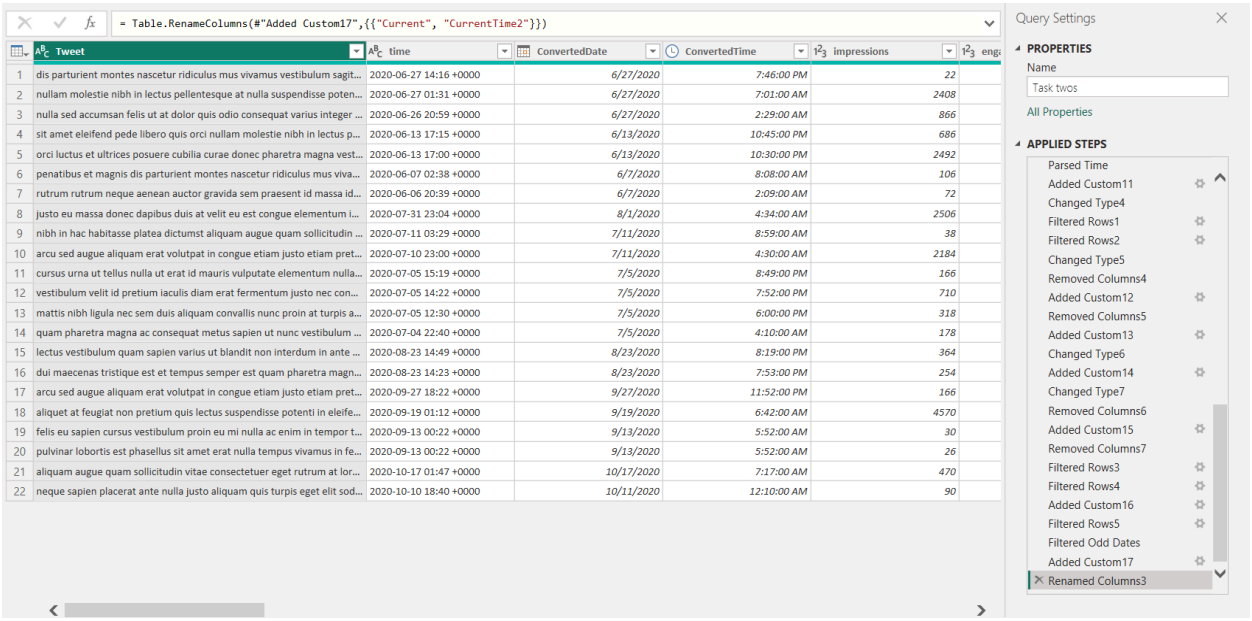
The primary learning objective of this task was to apply advanced data analysis techniques using Power BI, with a focus on building dynamic and interactive dashboards. By completing this task, I sought to strengthen my skills in several key areas, including:

1. **DAX Formulas and Logic:** Understanding how to write complex DAX to calculate and rank tweets based on engagement metrics. This involved creating measures to sum up retweets and likes, which was crucial for identifying the top 10 tweets.

2. **Real-Time Data Handling:** Learned to implement time-based functionality, ensuring the dashboard operates only between 3 PM and 6 PM based on the system's local time. This involved integrating dynamic time-based filters and displaying current system time to indicate dashboard usage.
3. **Visualizing Data:** Learning to choose the right types of visualizations (bar charts, pie charts, KPI cards) to represent social media engagement data clearly and effectively. The goal was to create visuals that would make it easy to compare top-performing tweets with others and present meaningful insights.
4. **Applying Filters:** Developed skills in applying complex filters in Power BI, including excluding tweets posted on weekends, ensuring even impressions, odd tweet dates, and a word count below 30. This required a thorough understanding of data transformation and filtering logic.

Activities and Tasks

The activities for this task were divided into multiple stages, each requiring a specific set of actions and considerations. The first major task was to design and build a chart that identified the top 10 tweets based on the sum of retweets and likes. To achieve this, I began by developing a calculated column, "Total Engagement," which summed the retweets and likes for each tweet. This involved using DAX to create a measure that accurately reflected the engagement metrics.



	Tweet	time	ConvertedDate	ConvertedTime	impressions
1	dis parturient montes nascetur ridiculus mus vivamus vestibulum sagit...	2020-06-27 14:16 +0000	6/27/2020	7:46:00 PM	22
2	nullam molestie nibh in lectus pellentesque at nulla suspendisse poten...	2020-06-27 01:31 +0000	6/27/2020	7:01:00 AM	2408
3	nulla sed accumsan felis ut at dolor quis odio consequat varius integer ...	2020-06-26 20:59 +0000	6/27/2020	2:29:00 AM	866
4	sit amet eleifend pede libero quis orci nullam molestie nibh in lectus p...	2020-06-13 17:15 +0000	6/13/2020	10:45:00 PM	686
5	orci luctus et ultrices posuere cubilia curae donec pharetra magna vest...	2020-06-13 17:00 +0000	6/13/2020	10:30:00 PM	2492
6	penatibus et magnis dis parturient montes nascetur ridiculus mus viva...	2020-06-07 02:38 +0000	6/7/2020	8:08:00 AM	106
7	rutrum rutrum neque aenean auctor gravida sem praesent id massa id...	2020-06-06 20:39 +0000	6/7/2020	2:09:00 AM	72
8	justo eu massa donec dapibus duis at velit eu est congue elementum i...	2020-07-31 23:04 +0000	8/1/2020	4:34:00 AM	2506
9	nibh in hac habitasse platea dictumst aliquam augue quam sollicitudin ...	2020-07-11 03:29 +0000	7/11/2020	8:59:00 AM	38
10	arcu sed augue aliquam erat volutpat in congue etiam justo etiam pret...	2020-07-10 23:00 +0000	7/11/2020	4:30:00 AM	2184
11	cursus urna ut tellus nulla ut erat id mauris vulputate elementum nulla...	2020-07-05 15:19 +0000	7/5/2020	8:49:00 PM	166
12	vestibulum velit id pretium iaculis diam erat fermentum justo nec con...	2020-07-05 14:22 +0000	7/5/2020	7:52:00 PM	710
13	mattis nibh ligula nec sem duis aliquam convallis nunc proin at turpis a...	2020-07-05 12:30 +0000	7/5/2020	6:00:00 PM	318
14	quam pharetra magna ac consequat metus sapien ut nunc vestibulum ...	2020-07-04 22:40 +0000	7/5/2020	4:10:00 AM	178
15	lectus vestibulum quam sapien varius ut blandit non interdum in ante ...	2020-08-23 14:49 +0000	8/23/2020	8:19:00 PM	364
16	dui maecenas tristique est et tempus semper est quam pharetra magn...	2020-08-23 14:23 +0000	8/23/2020	7:53:00 PM	254
17	arcu sed augue aliquam erat volutpat in congue etiam justo etiam pret...	2020-09-27 18:22 +0000	9/27/2020	11:52:00 PM	166
18	aliquet at feugiat non pretium quis lectus suspendisse potenti in eleife...	2020-09-19 01:12 +0000	9/19/2020	6:42:00 AM	4570
19	felis eu sapien cursus vestibulum proin eu mi nulla ac enim in tempor t...	2020-09-13 00:22 +0000	9/13/2020	5:52:00 AM	30
20	pulvinar lobortis est phasellus sit amet erat nulla tempus vivamus in fe...	2020-09-13 00:22 +0000	9/13/2020	5:52:00 AM	26
21	aliquam augue quam sollicitudin vitae consectetuer eget rutrum at lor...	2020-10-17 01:47 +0000	10/17/2020	7:17:00 AM	470
22	neque sapien placerat ante nulla justo aliquam quis turpis eget elit sod...	2020-10-10 18:40 +0000	10/11/2020	12:10:00 AM	90

Next, I implemented a range of filters to refine the data. I applied conditions to exclude tweets posted on weekends, ensuring that only tweets from weekdays were considered. Additionally, I filtered the data to include only tweets with even impressions, odd tweet dates, and a word count of fewer than 30 characters. These filters were crucial for meeting the assignment's specific criteria. In Power Query Editor, I used M code to transform the data accordingly, applying these filters meticulously.

The dashboard included several visualizations to present the data effectively. I utilized a stacked bar chart to illustrate engagement trends over time, showing how retweets and likes varied. A

pie chart was used to highlight the contribution of different user profiles to overall engagement. The area chart provided a visual representation of engagement trends, while the slicer for 'DayWeek' allowed for filtering the data by specific days. I also added cards to display current metrics, including the total number of retweets, likes, and user profile clicks.
To ensure the dashboard functioned correctly within the specified time frame of 3 PM to 6 PM, I created a dynamic time-based measure that adjusted the dashboard's visibility based on the system's local time. This involved setting up a measure to display "Active" or "Inactive" status depending on whether the current time fell within the specified range. Each of these activities required careful attention to detail and iterative testing to ensure the final dashboard met all requirements and provided accurate insights into social media engagement.

Skills and Competencies

This task honed a variety of technical and analytical skills essential for a career in data analytics:

1. **Advanced Power BI Skills:** I deepened my knowledge of Power BI, particularly in creating complex DAX queries and calculated columns. Handling conditions like time-based functionality and multi-level filters improved my ability to work with real-time data and build responsive dashboards.
2. **Data visualization:** Improved skills in selecting and applying appropriate visualizations to represent data effectively. This included using bar charts to rank tweets, pie charts to show user profile contributions, and KPI cards for key metrics.
3. **Time-Based Filtering:** Learned to implement time-based functionality, ensuring the dashboard operates only between 3 PM and 6 PM based on the system's local time. This involved integrating dynamic time-based filters and displaying current system time to indicate dashboard usage.
4. **Problem-Solving and Debugging:** Enhanced problem-solving skills through debugging DAX formula errors and addressing data type mismatches, which was critical for ensuring accurate data representation and functionality in the dashboard.
5. **Project Management:** Managing multiple aspects of the dashboard—from data preparation to visualization—taught me how to structure and organize a data analytics task efficiently. I learned how to break down complex requirements into manageable tasks and execute them systematically.

Feedback and Evidence

Since I undertook this task independently, there was no formal feedback provided by a supervisor or mentor. The entire process—from planning to implementation—was self-driven. I encountered several challenges, particularly with the time-based functionality, but through trial and error, I was able to develop effective solutions. This task was a valuable exercise in self-assessment and problem-solving, as I had to rely on my judgment to determine the effectiveness of the visualizations and the accuracy of the filters.
Despite the lack of external feedback, I took proactive steps to evaluate the success of the task. For instance, I continuously tested the dashboard under various conditions (different times of the day, filtering out tweets, etc.) to ensure all components worked as expected. After several

iterations, I implemented a robust solution that dynamically updated the dashboard based on system time and other filtering criteria.
Evidence of my work includes the completed Power BI file, which features fully functional filters, time-based constraints, and accurate visualizations. I also captured screenshots at various times during the day to demonstrate how the dashboard transitioned between "Active" and "Inactive" states based on system time. Through this process, I gained valuable hands-on experience, further improving my technical proficiency in Power BI and data analytics.

Challenges and Solutions

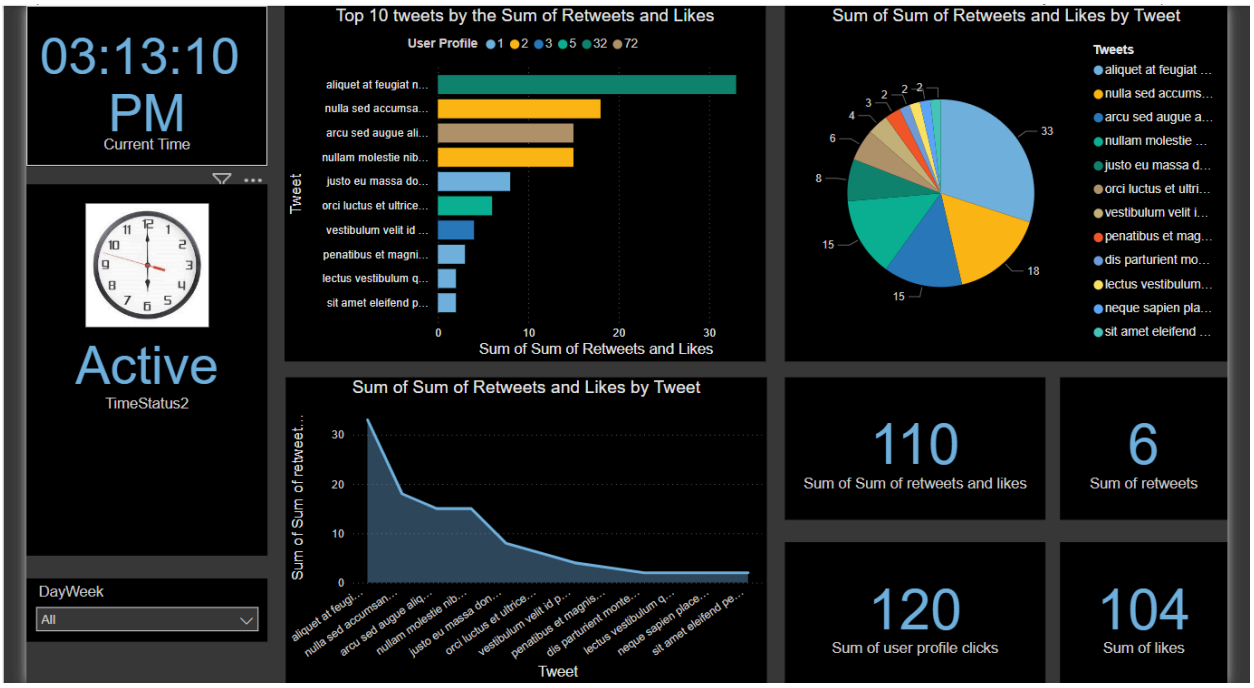
One of the key challenges in this task was ensuring that the dashboard's functionality was restricted to the time frame between 3 PM and 6 PM based on the system's local time. Power BI does not natively support real-time dynamic time-based conditions, so implementing this required creating a custom DAX measure. The solution involved crafting a measure that compared the system's current time to the 3 PM to 6 PM window. This measure used DAX functions like NOW() and conditional logic to determine if the current time fell within the specified range and display either "Active" or "Inactive" accordingly.
Another significant challenge was applying multiple data filters, including even impressions, odd tweet dates, and short word counts. This involved complex data transformations in Power Query. I had to write custom M code to filter out tweets with even impression counts and odd dates while ensuring that the tweet word count was below 30 characters. This required precise handling of data types and validation to ensure that only tweets meeting all criteria were included.
Ensuring that the dashboard displayed only the top 10 tweets after applying these filters was another hurdle. The complexity arose from the need to accurately rank tweets based on the sum of retweets and likes, considering the filters applied. Implementing this required the creation of a calculated column for engagement metrics and applying a "Top N" filter in Power BI. The solution involved thorough testing and validation to confirm that the top 10 tweets were correctly identified and displayed, ensuring that the aggregation and ranking were accurate.

Outcomes and Impact

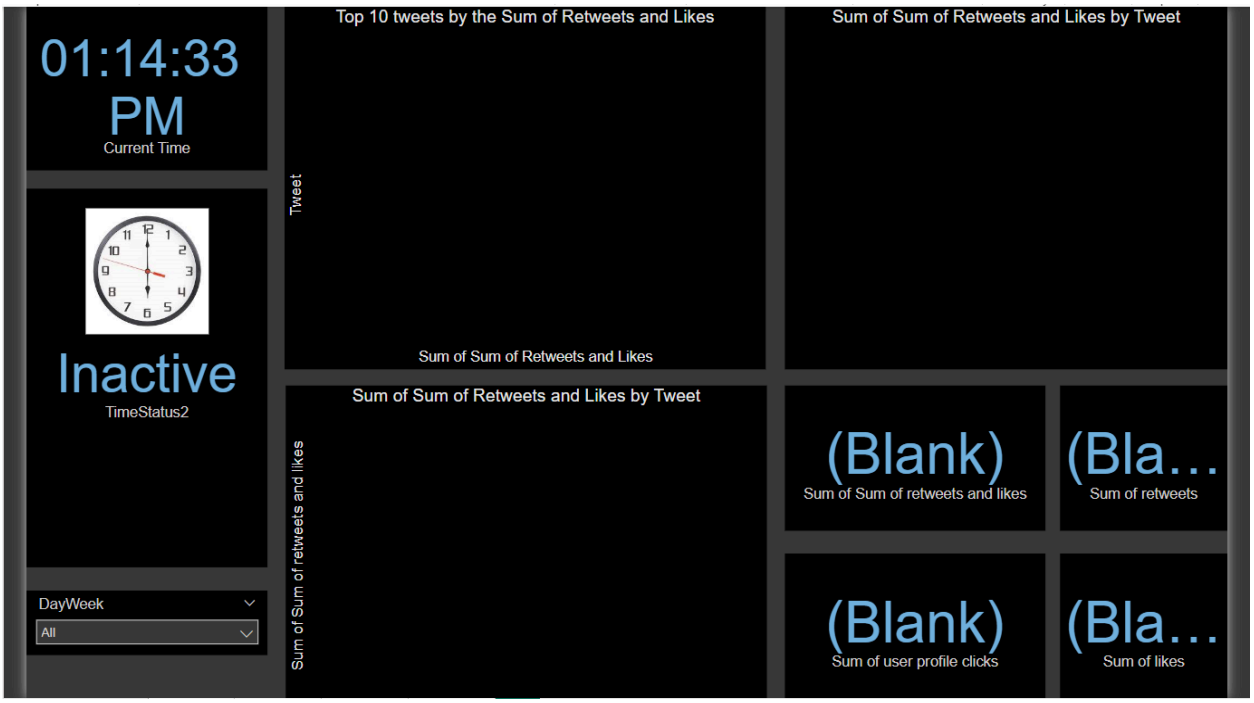
The final dashboard effectively addressed the task requirements by identifying the top 10 tweets based on engagement metrics and applying all specified filters. The stacked bar chart provided a clear visualization of engagement trends over time, highlighting variations in retweets and likes. This chart was instrumental in understanding how engagement fluctuated across different periods.

- The pie chart offered insights into the distribution of engagement among different user profiles, showcasing which users were most influential in terms of retweets and likes. This helped in assessing the relative impact of each user on overall tweet engagement.
- The area chart illustrated trends in engagement metrics, revealing patterns and anomalies in the data. This visualization was useful for tracking changes in engagement over time and identifying periods of high or low activity.
- The slicer for DayWeek allowed users to filter data based on the day of the week, making it easier to analyze engagement patterns for specific days.

- The cards provided quick, at-a-glance information on current metrics, such as the total number of retweets, likes, and user profile clicks. The dynamic card displaying the current status ("Active" or "Inactive") based on the time window further ensured that users only interacted with relevant data.



The implementation of time-based filtering ensured that the dashboard functioned correctly within the specified 3 PM to 6 PM window, aligning with the real-time conditions.



This task demonstrated the successful integration of complex filtering and visualization techniques, delivering a functional and insightful dashboard that enhances the understanding of social media engagement patterns and supports effective data-driven decision-making.

Conclusion

The task successfully achieved its objective of developing a dynamic and interactive dashboard for analyzing tweet engagement metrics. By applying various filters and visualizations, the dashboard provided valuable insights into the performance of tweets under specific conditions. The use of advanced PowerBI features, such as time-based filtering and custom measures, enabled the creation of a tool that meets the task's requirements while offering a clear and actionable view of the top-performing tweets. Despite challenges related to data filtering and time-based conditions, the task underscored the importance of precise data handling and visualization techniques. The final dashboard serves as a powerful resource for analyzing social media engagement, demonstrating the effective application of data analytics skills in a practical context.