Nullclass Internship Report - Data Analytics

This is report of completion of my internship successfully within given duration of 3 months with included training session in which I perform training and internship task on topic "**Learn to Build a Real-Time Twitter Analysis Project Report**".

Successfully Completed

By

Sammeta Vennela Harini

October-2024

Table of Contents

1. Introduction

This report summarizes my internship experience at Nullclass Online, where I was involved in a real-time Twitter analysis project. The goal was to collect and analyze Twitter data to provide insights into engagement metrics using Power BI. The report covers my key activities, skills learned, challenges faced, and overall impact on my professional growth.

2. Background

2.1 About Null Class

Nullclass Online is a cutting-edge educational platform that offers specialized instruction in a range of subjects, including data analytics. The firm is renowned for its methodical approach to instruction, which integrates theoretical understanding with realworld application. Throughout the internship, a focus on practical experience was maintained to make sure students could apply the principles they had learned.

The company offered a nurturing learning atmosphere with knowledgeable mentors on hand to offer advice and help whenever needed. These mentors were invaluable during the training sessions and the internship assignments, providing insightful criticism and guidance that enabled me to overcome obstacles and broaden my comprehension of the subject. I was able to gain self-assurance in my abilities and effectively finish the duties given to me during the internship because to this encouraging environment.

2.2 Project Overview

The core of my project was real-time Twitter analysis, where I worked on analyzing engagement metrics such as likes, retweets, replies, media views, and clicks. The data was processed, visualized, and presented through Power BI, enabling efficient analysis of Twitter interactions.

3. Learning Objectives

My learning objectives included:

- To gain proficiency in using Power BI for real-time data visualization.
- To understand and Data World Website for data extraction.
- Understand how to process and interpret large datasets for decision-making purposes.
- To develop critical analytical thinking by solving data-related challenges.

4. Activities and Tasks

Throughout my internship, I engaged in several key tasks, including:

4.1 Data Collection

I extracted real-time data from Twitter using the Data World Website to download the excel file. The data collected included tweet IDs, tweet content, timestamps, impressions and engagement metrics like retweets, likes, hashtags, replies and media interactions.

4.2 Data Cleaning and Preprocessing

The collected data was preprocessed to remove duplicates and null values. Missing data, such as some timestamps, were handled using interpolation and filtering.

4.3 Data Visualization

I used Power BI to create various charts and dashboards. Some key visualizations include:

1. **Line Chart:** Trend of average engagement rate over each month, with separate lines for tweets with media and those without.
2. **Donut Chart:** Visualization comparing replies, retweets, and likes for tweets that received media engagements greater than the median value. Include a filter for tweets posted in the last six months
3. **Dual-Axis Chart:** Displayed the number of media views and media engagements by day of the week for the last quarter. Highlight days with significant spikes in media interactions.
4. **Pie Chart:** Represented the proportion of total clicks (URL clicks, user profile clicks, hashtag clicks) for tweets with more than 500 impressions, with drill-down to view the specific types of clicks for each tweet.
5. **Clustered Bar Chart:** Displayed 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.
6. **Clustered Column 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.
7. **Clustered Column Chart:** 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.

5. Skills and Competencies

During this internship, I developed a range of technical and professional skills:

5.1 Technical Skills

- **Power BI**: Mastered creating advanced visualizations like clustered bar charts, dual-axis charts, and pie charts.
- **DAX (Data Analysis Expressions)**: Wrote complex DAX formulas to calculate engagement metrics, tweet category and filter data based on time intervals.
- **Creating Measures**: Wrote a Measure formulas to calculate average engagements with and without media and median media engagements.
- **Data world website**: Learned how to extract data from Twitter using Data world website.
- **Excel**: Used Excel for initial data extraction and transformation.

5.2 Professional Skills

- **Problem-solving**: Tackled challenges like missing data and performance issues with large datasets.
- **Time management**: Balanced multiple tasks and delivered results within tight deadlines.
- **Individual**: I worked individually in this twitter analytics project

6. Feedback and Evidence

My mentors provided me with crucial advice and support during my internship at Nullclass Online. Whenever I had questions or needed advice, the mentors were there to help, especially with technical difficulties. They gave me feedback on both my current tasks and how I may use these talents to apply to other projects in the future. They complimented my quick learning curve and my tenacity in solving problems on my own.

Evidence of Work

- Screenshots of Power BI dashboards.
- Excel for Twitter data extraction.
- DAX formulas used in Power BI.

7. Challenges and Solutions

7.1 Challenges

- **Missing Timestamp Data**: Some tweets lacked timestamp information, making it difficult to analyze trends.

- **Tasks**: Understanding the tasks is the biggest challenge I faced
- **Complex Visualizations**: Creating dual-axis charts and pie charts with drill-down features in Power BI posed initial difficulties.

7.2 Solutions

- **Handling Missing Data**: Used interpolation techniques and restricted analyses to complete datasets.
- **Power BI Learning**: Researched Power BI's advanced features and practiced DAX formulas to enhance the visualizations.

8. Outcomes and Impact

8.1 Technical Outcomes

- Successfully created and implemented real-time data dashboards using Power BI.
- Developed expertise in using DAX for data calculations and filtering.

8.2 Personal Growth

- Gained confidence in applying data analysis to real-world scenarios.
- Strengthened my problem-solving and time management skills by overcoming project obstacles.

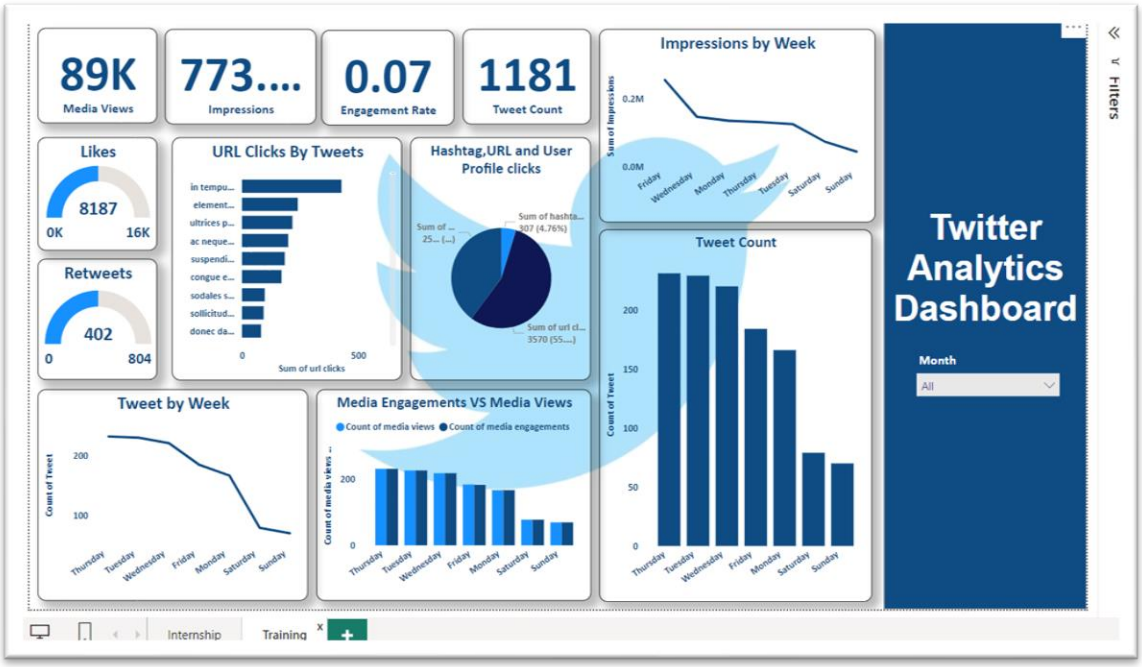
9. Conclusion

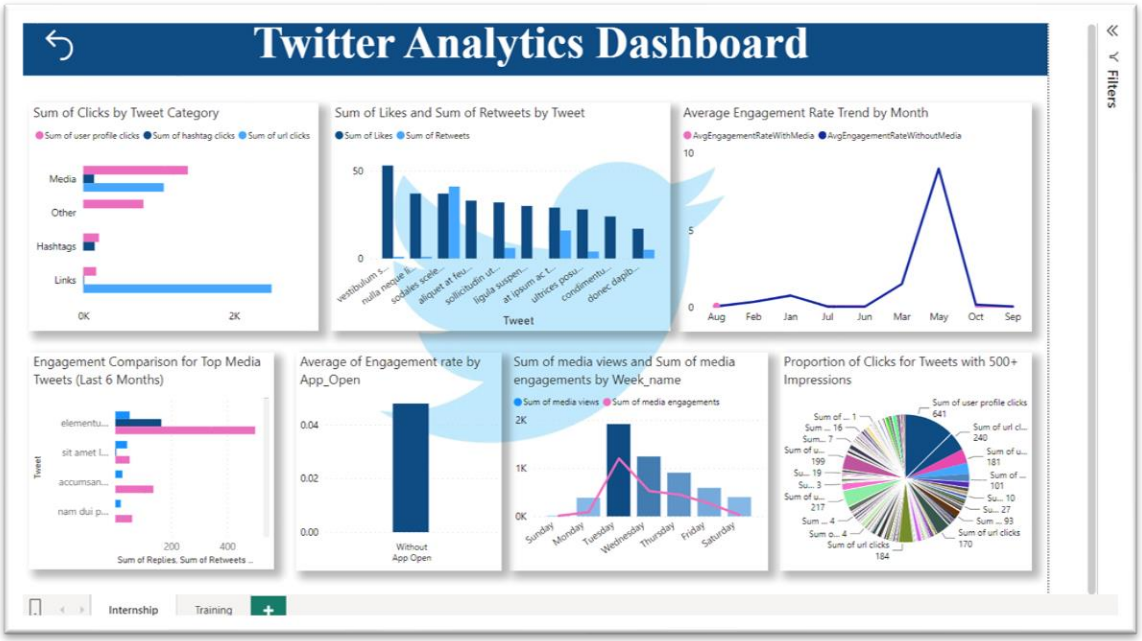
My internship at Null Class was a transformative experience, allowing me to gain hands-on experience in real-time data analysis and visualization. The skills I acquired in Power BI will undoubtedly serve me well in future roles as a data analyst.
Overall, the challenges I faced and the solutions I developed not only strengthened my technical competencies but also enhanced my problem-solving abilities and understanding of real-world data applications. This internship has provided a solid foundation for my future career, equipping me with the tools and confidence to succeed in the dynamic field of data analytics.

10.Appendices

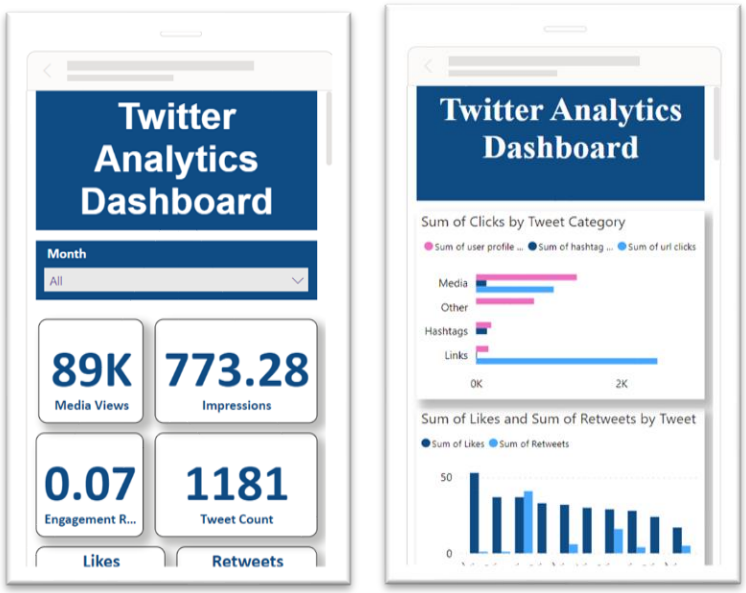
- **Appendix A**: Screenshots of Desktop View of Power BI visualizations.
- **Appendix B**: Mobile View of Power BI visualizations.
- **Appendix C**: Measure and DAX formulas used in Power BI calculations

Appendix A:


89K
Media Views
773....
Impressions
0.07
Engagement Rate
1181
Tweet Count
Impressions by Week
Likes
8187
0K
16K
URL Clicks By Tweets
Hashtag,URL and User Profile clicks
Retweets
402
0
804
Tweet Count
Twitter Analytics Dashboard
Month
All
Tweet by Week
Media Engagements VS Media Views
Filters
Internship
Training


Twitter Analytics Dashboard
Sum of Clicks by Tweet Category
Sum of Likes and Sum of Retweets by Tweet
Average Engagement Rate Trend by Month
Engagement Comparison for Top Media Tweets (Last 6 Months)
Average of Engagement rate by App_Open
Sum of media views and Sum of media engagements by Week_name
Proportion of Clicks for Tweets with 500+ Impressions
Filters
Internship
Training

Appendix B:



Appendix C:

Measure Formulas:

- MedianMediaEngagements = MEDIANX(ALL('SocialMedia'), SocialMedia[media engagements])
- AvgEngagementRateWithMedia =
  CALCULATE(
  AVERAGE([Engagement rate]),

  FILTER(SocialMedia, [media content] = "With Media")

)
- AvgEngagementRateWithoutMedia =

  CALCULATE(

  AVERAGE([Engagement rate]),

  FILTER(SocialMedia, [Media Content] = "Without Media")

)

DAX Formulas:

- Weekday = IF(WEEKDAY([Date], 2) <= 5, TRUE, FALSE)

- Media Content = IF([media views] > 0, "With Media", "Without Media")
- TweetCategory =

```
SWITCH(TRUE(),
    [media views] >0, "Media",
    [url clicks] >0,"Links",
    [hashtag clicks] >0,"Hashtags",
    "Other"
)
```

- Interaction = IF([URL Clicks] > 0 || [user profile clicks] > 0 || [Hashtag Clicks] > 0, TRUE, FALSE)
- RelevantTweet = IF(
  [media engagements] > [MedianMediaEngagements] &&
  [Date] >= DATE(2020,10,19) - 180,
  TRUE,
  FALSE
)
- App_Open = IF([app opens] > 0, "With App Open", "Without App Open")
- BusinessHourTweet =

```
IF(
    AND(
        [TweetHour] >= 9,
        [TweetHour] < 17
    ),
    TRUE,
    FALSE
)
```

- TotalClicks = [url clicks]+[user profile clicks]+[hashtag clicks]
- TotalEngagement = [Retweets] + [Likes]
- impressions>500 = IF(SocialMedia[Impressions] >500 ,TRUE,FALSE)
- TweetHour = HOUR([Time])