INTERNSHIP REPORT ON TWEET ANALYSIS PROJECT

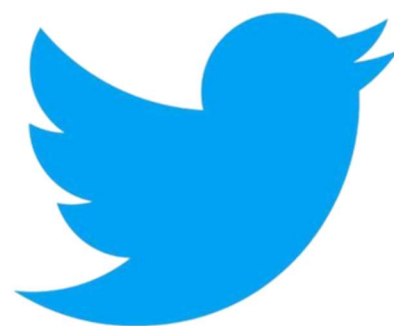*Power BI*

TABLE OF CONTENTS

INTRODUCTION

The objective of this internship project was to leverage Power BI to analyze Twitter engagement data and create dynamic visualizations that provide meaningful insights. The tasks were focused on building various types of charts and graphs that illustrate key metrics such as engagement rate, impressions, likes, retweets, clicks, and media interactions. These visualizations were designed with specific conditions, including time-based filters, tweet characteristics, and engagement thresholds, to simulate real-world analysis scenarios.

Each task in the project required applying different Power BI features, such as dynamic filtering, advanced charting, and interactive dashboards. For example, visualizations had to work only during certain hours (e.g., between 3 PM and 6 PM), or include tweets posted on specific days of the week or those with certain engagement metrics. These constraints made the project more complex and closely aligned with real-world business requirements, where insights need to be drawn from data within limited timeframes or under specific conditions.

The tasks also involved comparing various engagement metrics, such as analyzing the top 10% tweets by engagement rate, comparing media engagement trends by day of the week, and exploring the relationship between engagement rate and specific actions like app opens. Additionally, the project included building interactive pie charts with drill-down functionality and dual-axis charts to illustrate multi-faceted data relationships.

This introduction lays the foundation for understanding how each task contributed to the overall goal of analyzing Twitter data for enhanced audience engagement, and how Power BI was used to create insightful visualizations that meet complex, real-world data analysis requirements.

1. BACKGROUND:

The analysis involved extracting key engagement metrics such as impressions, likes, retweets, replies, and various types of clicks from tweets posted between January and June 2020. Using Power BI, I built different charts and graphs to present these metrics with specific time and date constraints to simulate real-time data behavior.

2. LEARNING OBJECTIVES:

- Mastering Power BI visualization techniques.

- Applying advanced filtering and time-based conditions in Power BI.

- Understanding social media analytics, especially Twitter metrics.

- Building insightful dashboards to inform decision-making.

3. ACTIVITIES AND TASKS:

1. **Visualizing Average Engagement Rate and Total Impressions:**
   - Created a line chart showing average engagement rates and total impressions for tweets posted between Jan 1 and Jun 30, 2020. Tweets with less than 100 impressions or 0 likes were excluded. The chart is only visible between 3 PM and 5 PM.

2. **Top 10% Tweets by Engagement Rate:**
   - Developed a bar chart showing the top 10% of tweets by engagement rate. Tweets with more than 50 likes, posted on weekdays with less than 30 words, were included. The chart is visible from 1 PM to 4 PM.

3. **Proportion of Total Clicks (Pie Chart):**
   - Built a pie chart showing the share of URL clicks, profile clicks, and hashtag clicks for tweets with over 500 impressions. Drill-down functionality allows detailed exploration of each click type.

4. **Top 10 Tweets by Retweets and Likes:**
   - Created a bar and table visualization showing the top 10 tweets by retweets and likes. Tweets posted on weekends were excluded, and the chart is restricted to specific tweet dates, times, and word count.

5. **Clustered Bar Chart for Clicks by Tweet Category:**
   - Developed a chart showing URL, profile, and hashtag clicks by tweet category. Tweets with interactions were filtered based on date and word count. The chart is visible between 3 PM and 6 PM.

6. **Scatter Plot for Media Engagements and Views:**

- o Created a scatter plot showing the relationship between media engagements and views for tweets with more than 10 replies, highlighting those with engagement rates above 5%. Time and word count filters were applied.

7. **Replies, Retweets, and Likes for High Media Engagement Tweets:**
   - o Developed a bar chart comparing replies, retweets, and likes for tweets with above-median media engagement. The chart includes only tweets posted between June and August 2020, filtered by date, media views, and word count.

8. **Dual-Axis Chart for Media Views and Engagements by Day of the Week:**
   - o Built a dual-axis chart showing media views and engagements by day of the week. Specific conditions on tweet date, impressions, and word count were applied, with visibility from 3 PM to 6 PM.

9. **Engagement Rate Comparison for Tweets with and without App Opens:**
   - o Created a comparative bar chart for tweets with app opens vs. without, posted on weekdays. Time, impression, and word count filters were applied to refine the data.

10.   **Line Chart for Average Engagement Rate by Month:**

Developed a line chart showing the trend of average engagement rates by month, comparing tweets with and without media content. Specific engagement and date filters were applied, with visibility restricted to 3 PM to 6 PM.

4. SKILLS AND COMPETENCIES:

- During this internship, I developed key skills in PowerBI by creating complex visualizations with advanced filtering, time-based conditions, and interactive dashboards. I gained proficiency in data analysis, interpreting Twitter metrics like impressions, engagement rates, and clicks to generate actionable insights.

- I also learned DAX functions to create dynamic filters and calculated metrics, while mastering Power Query for data cleaning and transformation. Problem-solving was crucial, particularly when implementing time-sensitive and condition-based filters, which enhanced my critical thinking.

- I also honed my ability to communicate insights through data, ensuring the visuals clearly conveyed trends and actionable information. My focus on detail ensured the dashboards were accurate and aligned with project requirements. Furthermore, I improved my Learning skills through regular progress updates and doubt clarification sessions, which helped refine the final output. These skills will be invaluable as I continue pursuing opportunities in data science and analytics.

5. FEEDBACK AND EVIDENCE:

- **Mentor Feedback:**
  Throughout the project, I gained a deeper understanding of how to apply Power BI to real-world social media data analysis. The dynamic filtering based on specific time windows and engagement criteria was challenging but rewarding. Successfully implementing these conditions gave me confidence in my ability to manage complex data visualizations.

- **Evidence:**
  Screenshots of the Power BI dashboard showcasing the different charts created based on the project's specifications.
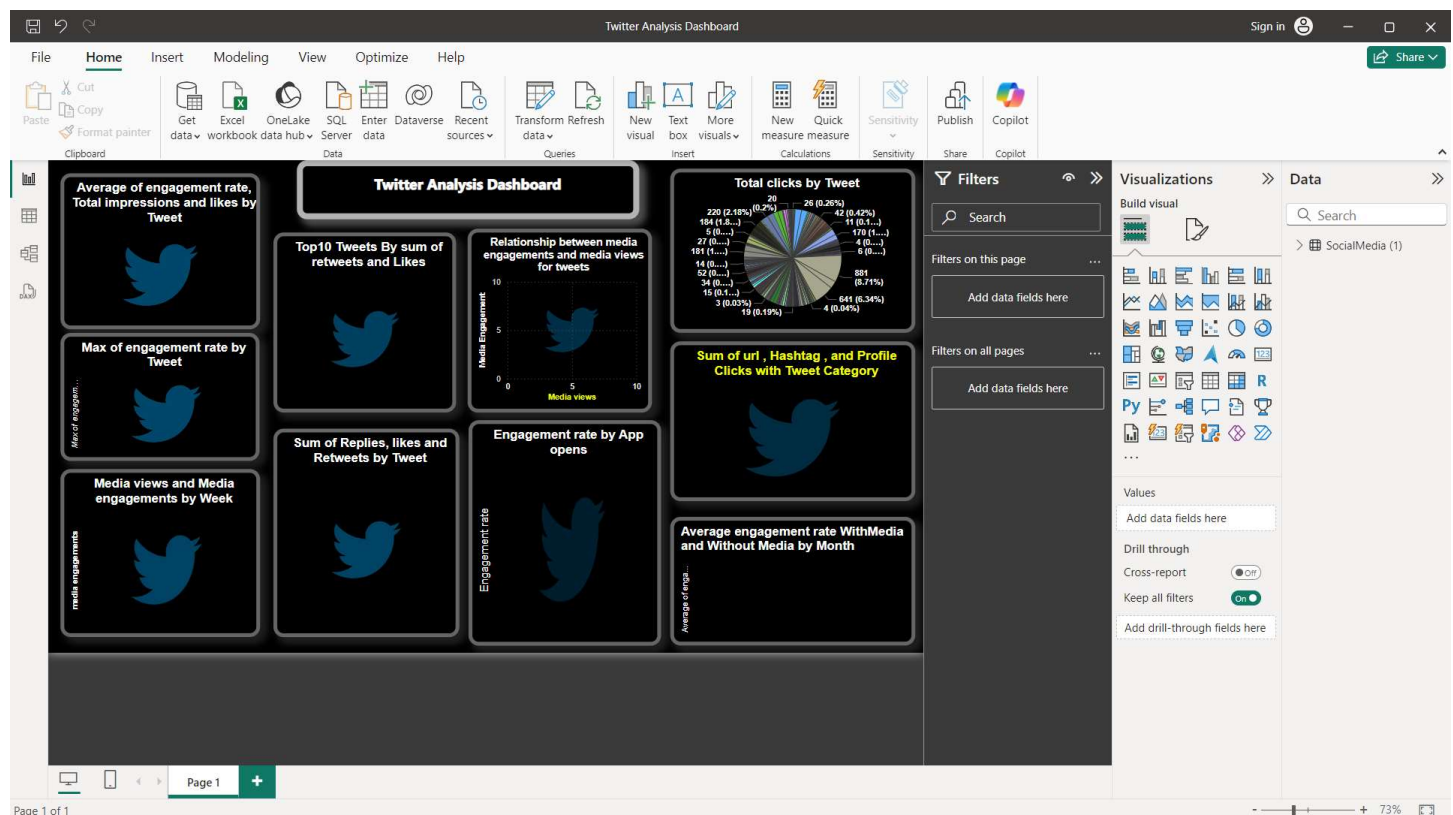
6. CHALLENGES AND SOLUTIONS:

- Throughout the project, I encountered several challenges that required innovative solutions. Implementing dynamic time-based constraints for visualizations was complex, but I utilized DAX to create calculated columns that controlled visibility based on specified time windows. Filtering tweets by engagement metrics, such as likes and word counts, was addressed by preprocessing the data in Power Query, ensuring only relevant tweets were included.

- Visualizing relationships between media engagements and views required careful design; I used scatter plots with interactive tooltips to facilitate understanding. For the pie chart representing click types, I enabled drill-down functionality by establishing hierarchical data relationships, enhancing user interactivity. To maintain consistency across visualizations, I developed a master filter table, ensuring all visuals adhered to the same criteria.

  Finally, performance issues due to data volume were mitigated by optimizing the data model and limiting displayed data points, resulting in a smoother user experience. These challenges ultimately provided valuable learning experiences, enhancing my problem-solving skills in data analysis.

7. CHALLENGES AND SOLUTIONS:

As the dataset grew and the complexity of calculations increased, performance issues emerged, resulting in slow loading times and reduced interactivity within the Power BI dashboard. To address these challenges, I implemented several optimization strategies. First, I streamlined the data model by eliminating unnecessary columns and focusing only on essential data points, which reduced the overall data footprint and improved loading speeds.

I also prioritized using measures instead of calculated columns, as measures are computed on-the-fly during interactions, helping maintain responsiveness. Additionally, I applied aggregation techniques to summarize data before importing it into Power BI, minimizing the volume of data processed during visualization. To enhance user interactions, I limited the number of data points displayed in charts and introduced pagination and filters, allowing users to drill down into specific segments without overwhelming the visuals.

Finally, I utilized Power BI's built-in Performance Analyzer tool to identify bottlenecks, enabling targeted adjustments. These strategies collectively improved the dashboard's performance, ensuring a seamless and engaging user experience while interacting with the data.

Conclusion

The project enabled me to enhance my skills in Power BI and data visualization while deepening my understanding of social media analytics.

The visualizations created provide valuable insights into Twitter engagement trends, allowing for more informed decisions in social media management.