

**Proposing a Business Intelligence Solution for X – Driving Engagements to Increase Value Through Improved Customer Experience**

By: Group 3

Date: April 17, 2024

Professor: Prof. Daniel Treku

**Section 1: Executive Summary**

X, formally known as Twitter, was acquired by Elon Musk in 2023 for $44 Billion. Since taking over as the CEO of X and making the company private, Musk has received much criticism for many of the problems with the app, leading to the value of the company diminishing by more than 70% since his acquisition. (Bratton) The goal of this project was to increase the value of X as a company and to use data warehousing and business intelligence strategies to provide a monitor diagnosis of the problem X is facing through historical tweets, as well as real time engagement tracking, stock value visibility, competitor analysis. We believed that an overall improvement of the customer experience would be able to drive more engagements and activity on the app, improving the valuation of X as a company overall.

First, a data warehouse using a multidimensional data model was created to be able to store the data that we will be using for dashboards in a centralized location. A snowflake schema was used to combine various data sources into one centralized database. We created four main dashboards, each for various purposes. We made a dashboard to look at tweets about different CEOs from before Elon Musk took over the company. People had all kinds of feelings about them, some good and some not so good, which shows why it's important to keep a good reputation. We looked at the percentage of good, neutral, and bad comments for each CEO to see what they're good at and what they need to work on. This shows us how what people say online can affect how people see tech leaders, helping us make better decisions in the tech world. The second dashboard created was a strategic dashboard comparing historical engagement on twitter with other social media competitors. The KPIs represented on this dashboard were specified for two age ranges: 20-29- and 30–39-year-old. They were Average twitter engagement, average total engagement, distance from highest engagement, share of twitter in total engagement and share of twitter engagement among YouTube and LinkedIn. The third dashboard created was a stock analysis dashboard that serves as a tool for investors and analysts to examine Twitter's stock performance trends, providing insights into its volatility, trading activity, and how the stock opens and closes each trading day over a spanning range of 10 years. Even though X is now a private company, this dashboard will be able to provide insight into what was making the company more valuable in the past. The KPI of this dashboard is total Volume. The last dashboard created was an operational dashboard to be utilized by X business analysts and software engineers to gain insight into tweet activity in real time and monitor performance and activity trends. Thai dashboard uses interactive filters to filter by date range, tweet sentiment, or tweet source. The KPIs on this dashboard are total tweets, total engagements, total likes, total retweets, and negative tweet sentient percentage.

In addition to the dashboard, we utilized predictive analysis to attempt to make the X user experience better. One of the main complaints about X is the number of bots, or fake accounts, generating activity on the app. We ran four different classification models to attempt to classify which twitter accounts are spam or bot accounts and which are not. We ultimately decided that a Random Forest Classification model was the best tool to successfully classify accounts as spam or not due to this model having the highest test accuracy and the least number of false positives, implying that this would lead to the least number of real users being classified as spam. The main goal of this predictive analysis was to get spam users off the app, making users of the app happier, and hopefully generate more interaction time and engagements while users are on the app.

Overall, business intelligence and predictive analysis were used to diagnose the problem X is currently facing, analyze ways that they can fix the problem using historical data, make the customer experience more pleasurable, and provide insight into real time engagement data to check if our initiative to generate more engagements on the app is being successful or not. We believe the best way to increase engagements and activity on the app is through improving the user experience, and that the solution we propose will be able to create a better user experience, driving engagement and value.

**Section 2: Introduction**

**Client Company Background**

Twitter, now known as X, was an American social media company based in San Francisco, California. Twitter was created in 2006 by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams. Twitter’s main product was the app itself, Twitter which has over 500 million users on it. Twitter’s main function was to allow users to “tweet” or share their thoughts in an up to 140-character post. Twitter added hyperlinks, hashtags and introduced “retweets” (reposts) in 2010. Twitter also started to allow for pictures and videos to be shared starting in 2012, GIFs to be shared in 2014, and added polls, direct messages, and live streams in 2015. In 2017, the character limit was changed from 140 to 280 (Murphy, 2019). With the ability to share videos and pictures, the extended character limit, and the ability to track engagements and engagement data, many companies use X as one of their digital marketing tools.

 In April 2023, Twitter was bought out by Elon Musk for $44 Billion, becoming a part of X Corp. During this transaction, Elon Musk also bought all of twitter's public shares, making it privately owned rather than publicly traded. Elon Musk also rebranded Twitter to be named X, while keeping the headquarters in the San Francisco, California and the app itself relatively the same. After Musk bought twitter, he laid off about 80% of the employees, leaving X with approximately 1,500 full time employees. Most of the current employees are software engineers and other quality assurance engineers. These engineers’ main job is to make sure that X is working and functioning smoothly and dropping any bugs that may arise, leaving users with the best experience possible (Liu & Toh, 2023).

X has a market cap of $41.09 Billion, making it the 468th most valuable company in the world. The largest markets for X are the United States, United Kingdom, and Germany. As twitter itself grew, the revenue of the company has also increased going from $1.2 Billion in 2014 to $4.8 Billion in 2023. Although X will have about 335 million users monthly in 2024, the app growth has plateaued in the last couple of years (*Twitter (TWTR) - Market Capitalization*, n.d.).

While X has been a successful company in the past, it has also had its fair number of challenges to this point. One of the challenges that X faces is monetization as it is often a challenge to find companies that want to make an advertising deal with X, especially after Musk’s acquisition. Another challenge that X faces is misinformation as users can say whatever they want, whenever they want. In addition to misinformation, one of the biggest criticisms of X since Musk has taken over the company is the increasing number of “bots” that are active on the app. These “bots” act as spam in many cases, using sophisticated algorithms to reply to tweets trying to promote some kind of content on the app. X also faces competition from other social media outlets that emphasize visual content and other companies have looked to copy Twitter’s interface such as Instagram with the creation of Threads.

On X, there are 500 million tweets a day and 200 billion tweets a year, totaling up to 4.3 petabytes of data per year. X makes efficient use of Big Data concepts using Hadoop as the first technology to manage the data at a large scale, using the concept of distributed storage. Originally, X used MySQL to store data but as the amount of data increased, X introduced Gizzard, a framework for creating distributed data stores. Tweets are stored in an internal system called T-bird, built on top of Gizzard, and secondary indexes are stored in another Gizzard-based system called T-flock. Unique IDs for each tweet that is sent out are generated by Snowflake. Flock DB is another database used for ID-to-ID mapping, efficiently storing the relationship between IDs by using Gizzard. A Gizzard is X’s distributed data store framework that is built on top of InnoDB as InnoDB doesn’t corrupt data and Gizzard can only act as a data store. Now, Twitter relies heavily on Hadoop for its internal analytics. To store more data that is a part of tweets such as images and videos, X uses a storage system called Blob store (Chavan, 2020). Separate from storage, X also uses a database management system called Manhattan. Manhattan acts as an in-house, real time, multi-tenant distributed database that is designed to handle high volumes of both structured and unstructured data, ensuring that data such as tweets, user profiles, and timelines are delivered in real time. Manhattan was also to be used as the metric storage database for X but as they faced scalability issues, they developed Metrics DB, an in-house database to store all metrics X is collecting (Shivang, 2019). Due to all these different integrated storage systems, X has teams in place to ensure all these systems are working up to date and using Hadoop to provide valuable insights for the company.

In addition to their internal systems, X also provide X Analytics, a powerful tool that businesses and other users can use to take a deep dive into their X campaigns. X provides an analytics dashboard that allows businesses to track the progress of ad campaigns and better understand their audience, helps that businesses can get from using the analytics dashboard are tracking ROI (Return on Investment), tracking audience growth and engagement, and refining strategies (*Twitter Analytics: How to Track and Measure Twitter Campaigns*, n.d.). Overall, X likely lies in the adult section of the BI maturity model as its metrics allow it to be analytic and performance driven, and they have an extremely sophisticated data management process in place.

**Case Studies:**

We believe that the customer experience of X as an app can be improved by Business Intelligence as Business Intelligence has been used to enhance the customer experience and generate revenue int he two following case studies:

**Starbucks:**  
 Starbucks uses its loyalty program to gather individual customer data, which is then analyzed to understand preferences and behaviors.This data powers their BI systems, leading to highly personalized marketing tactics. Offers and promotions are not sent out at random; they are the result of predictive analytics that assess a customer's earlier choices to forecast future purchases. This level of personalization could be used by Twitter to enhance its user experience. For example, by adopting similar BI strategies, Twitter could develop more precise algorithms for content delivery, ensuring that users receive tweets, topics, and ads that align with their interests and earlier interactions. This targeted approach could increase the platform's engagement rates and advertising effectiveness (Hyperight, 2021, Marr, 2021).

**Netflix:**

Netflix uses BI to turn viewing patterns into a robust recommendation engine, directly influencing viewer retention by suggesting relevant latest content. This is possible because of the deep analytical dive into the data that Netflix performs, understanding not only what was watched but when, how often, and what was searched. Twitter could adopt a parallel BI approach by analyzing tweet interactions and user activity to perfect the content displayed to each user. It could also help in the curation of the 'Explore' tab, making it more dynamic and reflective of real-time global or local trends, thus keeping users engaged and on the platform for longer periods (Niwate, 2021).

By integrating advanced BI solutions, X can generate useful insights from massive amounts of real-time data. These insights can guide the platform's content strategy, making it possible to not only personalize and better the user experience but also to predict and influence trending topics. Moreover, BI could help Twitter offer enhanced tools for advertisers, providing them with more specified control over their target demographics and the effectiveness of their campaigns, much like how Starbucks and Netflix fine-tune their customer engagement strategies.

The value of BI solutions is justified through the enhanced ROI and customer satisfaction visible in Starbucks and Netflix cases. Both companies have seen significant returns from their investment in BI through increased sales and subscriber numbers, respectively. For Twitter, similar BI adoption could lead to an increase in ad revenue and user growth. The company can present this potential by referencing the measurable successes that Starbucks and Netflix have achieved through BI, aligning it with Twitter's specific goals and metrics.

In summary, the Starbucks and Netflix cases provide blueprints for successful BI implementations that Twitter can adapt. By focusing on personalized user experiences and using real-time data analytics, Twitter can enhance its platform’s attractiveness to both users and advertisers, thereby driving growth and revenue.

**Section 3: The Proposed BI Solution**

Business intelligence uses business analytics, data mining, data visualizations, data tools and infrastructure, and best practices to help organizations make better informed, data-driven decisions. Combining these tools also allows you to have a complete and comprehensive view of your organization’s data, letting you use the data to drive change, drop inefficiencies, quickly adapt to market changes, and track key KPIs. when it comes to strategic planning. Overall, Business Intelligence is an umbrella term that covers all the process and methods in place for collecting, storing, and analyzing data from various business operations and activities to gain better insight and improve performance. (Tableau)

The main components of a BI solution are people, processes, management, and governance. People have the role of being the ones to set up the system to be able to track various KPIs that the business is interested in. People will likely consist of IT professionals, business analysts, and any other technical people that may be involved in the implementation of the BI solution. Processes have the role of being put in place to make the extraction, transforming, and reporting of data automated and easy to understand. It is important that processes are put in place that track important KPIs and truly bring value to the company. Management has the role of overseeing and assuring that people are implementing processes that can bring value to the company. Management should be a clear communicator when it comes to their expectations of the system put in place and should be able to understand all the various capabilities that the BI framework has when it comes to the data being reported and visualized. Data Governance has the role of being the glue to put people, process, and technology all together. Data governance involves setting up internal standards and data polices to govern how the data is gathered, stored, and processed and who has access to what kinds of data. Data governance plays an essential role in keeping data protected. All these components have various underlying frameworks and systems to ensure that internal and external business data is used in a way that can provide value to the company through a business intelligence solution.

Within each of these components, there are various frameworks in place for a BI solution to be applicable. First, a company will need external data as well as operational data. In our business case of Twitter, external stock data and external customer data can be bought through various sources and can be used to assess the value of twitter as a company, and what threats there are for X when it comes to competition with other social media outlets. Second, a company needs to have a process in place that can extract, load, and transform data so that it able to be stored in Data Warehouses and Data Marts. From the introduction, X already has processes and software in place such as Gizzard, Flock DB, Snowflake, T-floc, InnoDB, Hadoop, Blob store, Manhattan, and Metrics DB. Lastly, there has to be a system in place to take all of this stored data, and query and report it so that it can be analyzed and visualized. A data warehouse can be used with the following Star Schema to set up the data that can potentially solve the problem that X has been facing with diminishing value ever since Elon Musk took over the company.

A diagram of a computer

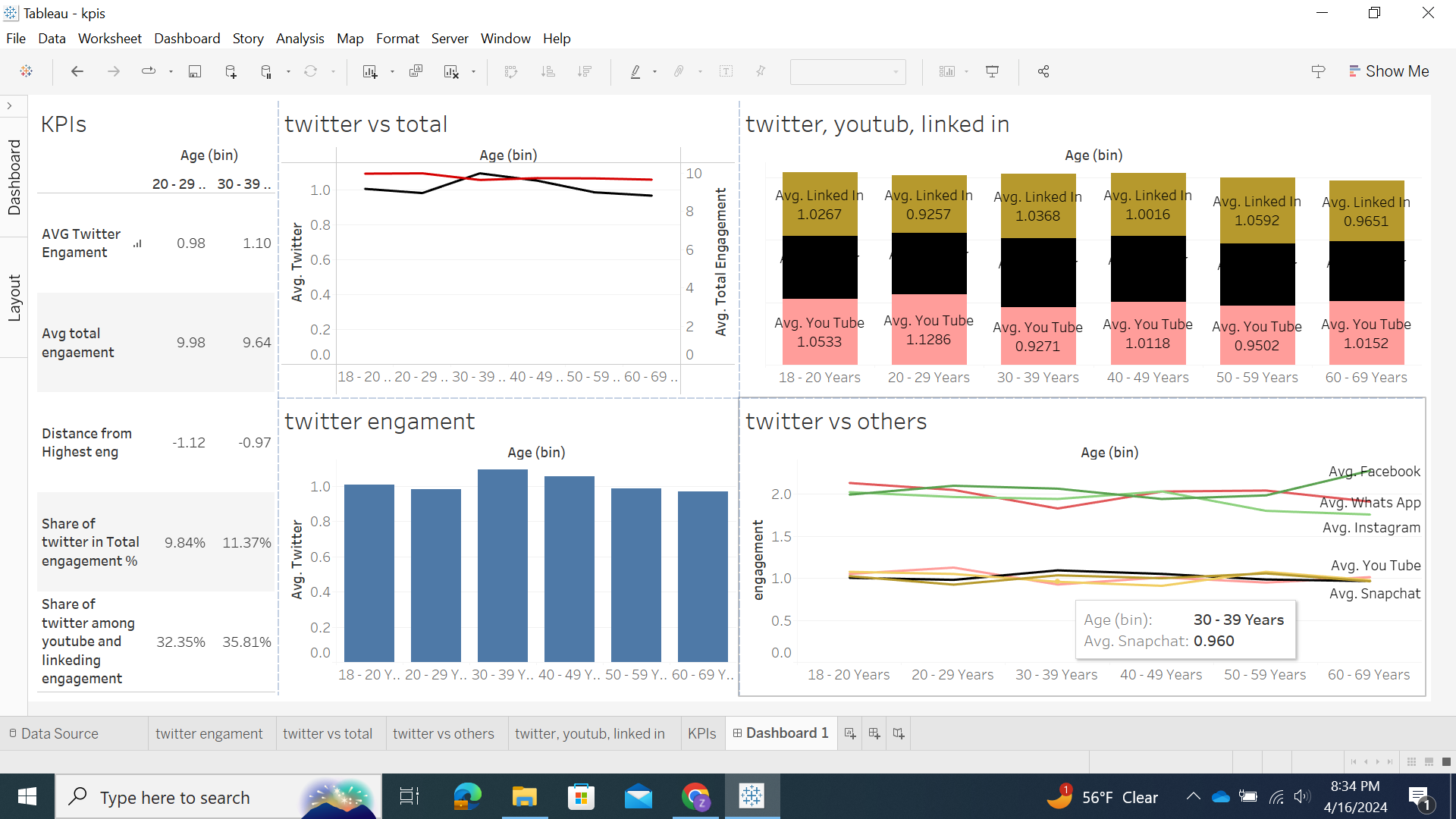
Description automatically generated with medium confidence

**Star Schema to Centralize Data used for Dashboards**

This data warehouse allows data from various sources to be integrated and shown across various dashboards. The dashboards that our BI solution will include a CEO sentiment analysis dashboard, competitive analysis strategic dashboard, a stock market analysis competitive dashboard, and a tweet activity operational dashboard. External data will also be utilized for the CEO analysis and consumer engagement dashboard as data will have to be scraped straight off the app and bought for the competitive analysis dashboard. The competitive analysis dashboard will be mainly composed of external data and used to be able to track user engagements across various social media apps and strategically track what kinds of customer markets twitter should be attacking based on age groups. This dashboard will also be able to track how well X is staying competitive with other companies based on user engagements. The stock market analysis dashboard will be used to gage when X has had the most value as a company and predict the direction that the value of the company is going. This dashboard can also be helpful when presenting to executives about the performance of twitter as a company in reference to its stock. This dashboard will be able to help Twitter answer questions about what it has done well in times of stock share increase and what it has done wrong in times where the stock price decreases. X will be able to use this data to make decisions that will help raise the value for the company overall. The operational dashboard will be used to track twitter data and see what types of tweets are coming from various sources at a given time. This operational dashboard will be able to track twitter engagement data in real time and analyze it, allowing for users to be able to see trends in sentiment of tweets, source of tweets, and engagement levels in tweets. This dashboard will be critical in assessing if the strategic goal of improving engagement is working in real time. Alongside the dashboards, a predictive analysis will be done, using historical data to try to classify spam accounts and remove them from the app. We hope that with less spam accounts on the app, the customer experience will be better as many unpaid ads and spam will not be shown on their feed anymore. Overall, the business solution through Business intelligence looks to bring more value to X by enhancing the customer experience. We believe that a mix of operational, strategic, and tactical dashboards can be used to solve the problem that X is currently facing: The diminishing value of X as a company since the acquisition of the company by Elon Musk. We believe that dashboards and predictive analysis could help X to improve their overall value as a company, through enhancing the customer experience with the goal of creating more engagements.

**Section 4: Use Cases/Prototypes**

**Dashboards:**



The Strategic dashboard we propose offers a comprehensive analysis of user engagement across various social media platforms, providing valuable insights for managers to discern the company's position relative to competitors. By delving into the engagement patterns of different age groups, this dashboard offers a nuanced understanding of audience preferences and behaviors.

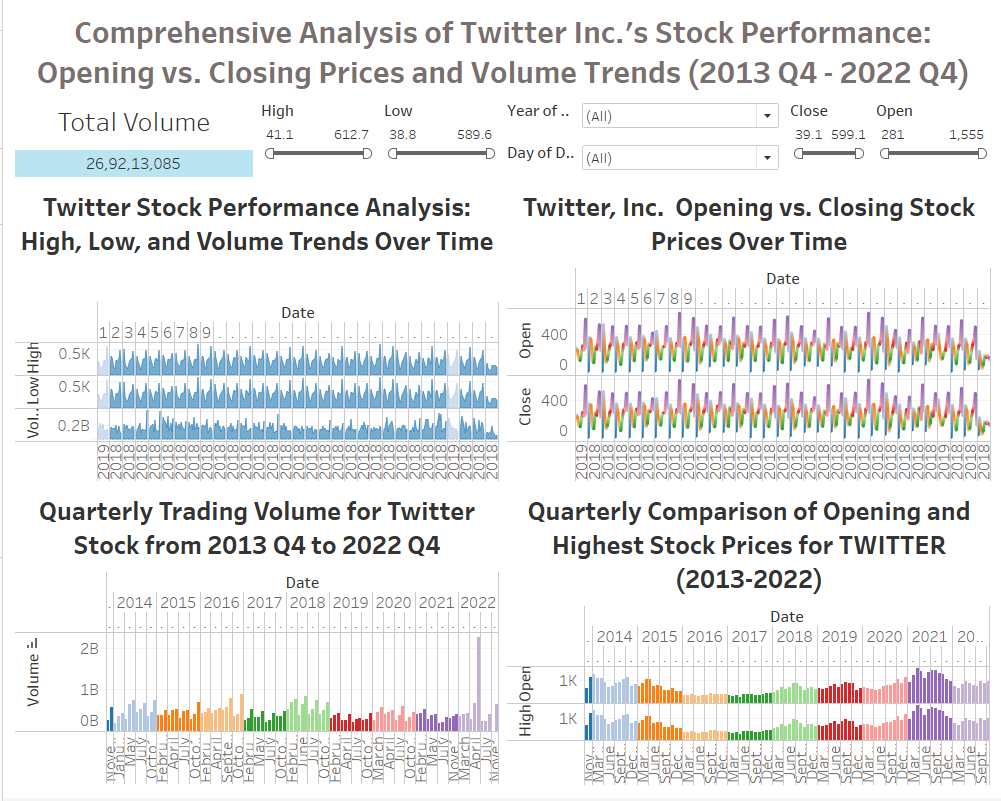
In the initial sheet, we conducted a comparative analysis of the average engagement levels among different age groups on Twitter. Notably, our findings revealed that individuals aged 30-39 exhibit the highest engagement rate, while those in the 20-29 age bracket demonstrate the lowest level of engagement.

Moving forward, in the second sheet, we explored the dynamics of engagement on Twitter in comparison to the overall engagement across all social media platforms. Surprisingly, while the total engagement rate with social media declines for individuals in their 30s, it experiences an upward trend specifically concerning Twitter. This divergence highlights Twitter's unique appeal and growing significance within the social media landscape.

In the subsequent sheet, we examined the engagement metrics across a spectrum of social media platforms, including Facebook, WhatsApp, Instagram, Twitter, YouTube, LinkedIn, and Snapchat, segmented by age groups. Our analysis revealed a distinct clustering of engagement patterns, wherein platforms like Facebook, Instagram, and WhatsApp constituted a high-engagement cluster, contrasting with Twitter, YouTube, LinkedIn, and Snapchat, which formed a cluster with comparatively lower engagement rates. This categorization underscores Twitter's role as an educational and intellectually stimulating platform, akin to YouTube and LinkedIn. Consequently, our analysis identifies LinkedIn and YouTube as Twitter's primary competitors in this sphere.

Further, in the fourth sheet, we delved deeper into the engagement dynamics within the lower-engagement cluster, focusing on Twitter, YouTube, and LinkedIn. Consistent with previous findings, our analysis reaffirmed Twitter's dominance in the engagement realm among individuals aged 30-39, while indicating a diminishing competitive edge relative to YouTube among the 20-29 age demographic.

In essence, our Strategic dashboard offers invaluable insights into the intricacies of user engagement, enabling managers to make informed decisions regarding strategic positioning and competitive strategies within the dynamic realm of social media.

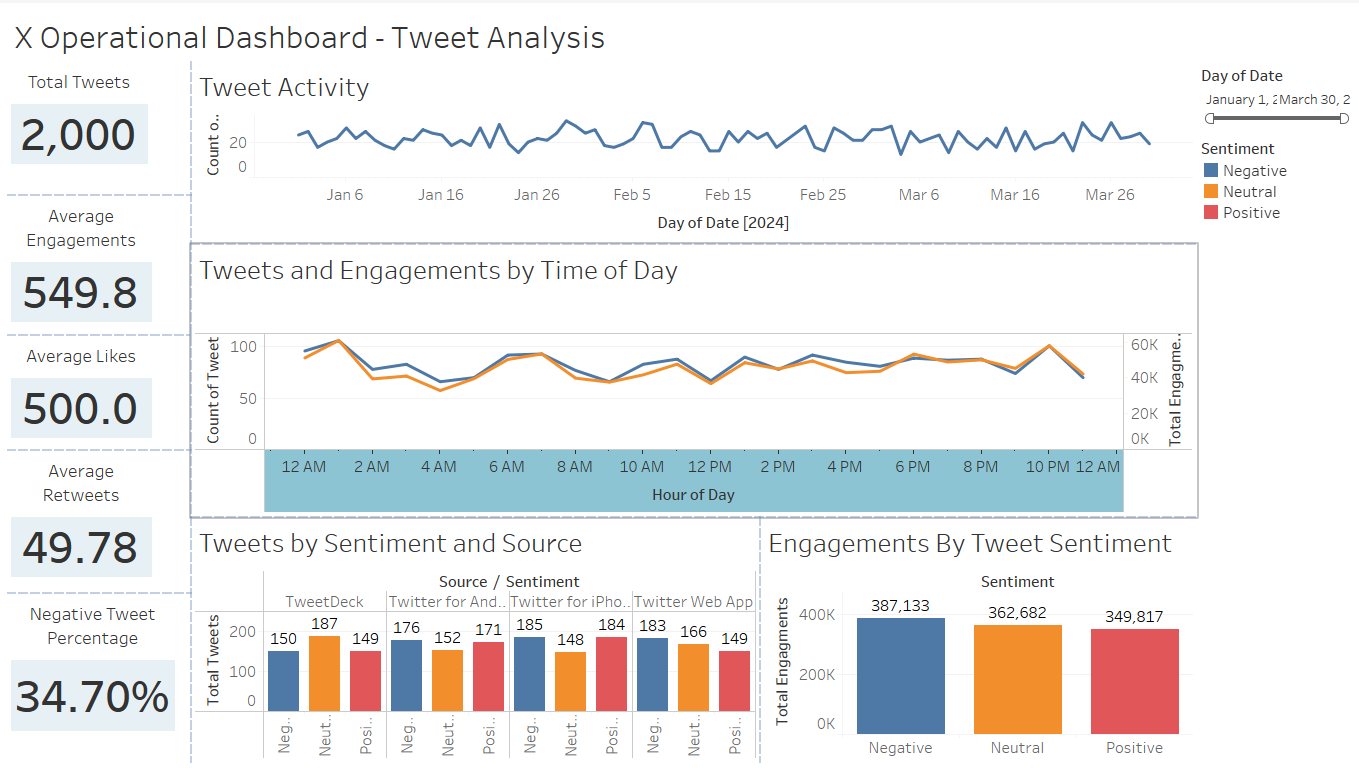
**Dashboard: Twitter Inc.'s stock performance from Q4 2013 to Q4 2022**

The dashboard provides a comprehensive analysis of Twitter Inc.'s stock performance from Q4 2013 to Q4 2022, highlighting various data visualizations related to stock prices and trading volumes. The top left graph, "Twitter Stock Performance Analysis," plots high, low, and volume trends over time, using a layered area chart where the highs and lows are shaded in blue, and the volume is represented in a deeper shade at the bottom, showing fluctuations over each day.

The top right graph compares opening and closing stock prices over time, with two overlapping line graphs in distinct colors—opening prices in green and closing prices in purple. This graph includes a control panel to filter by day of the date and year of the date, allowing for a more specific analysis of the stock's daily performance.  
 The bottom left graph, "Quarterly Trading Volume for Twitter Stock," uses a bar chart to represent trading volumes on a quarterly basis, with each year designated by a specific color, illustrating the trend and seasonal patterns in trading activity.

Finally, the bottom right graph offers a "Quarterly Comparison of Opening and Highest Stock Prices," again using a bar chart with each quarter color-coded to correspond to a specific year. It depicts the range between the opening and highest stock prices for each quarter, giving an idea of the intra-quarter price volatility.

The entire dashboard serves as a tool for investors and analysts to examine Twitter's stock performance trends, providing insights into its volatility, trading activity, and how the stock opens and closes each trading day over a large period.

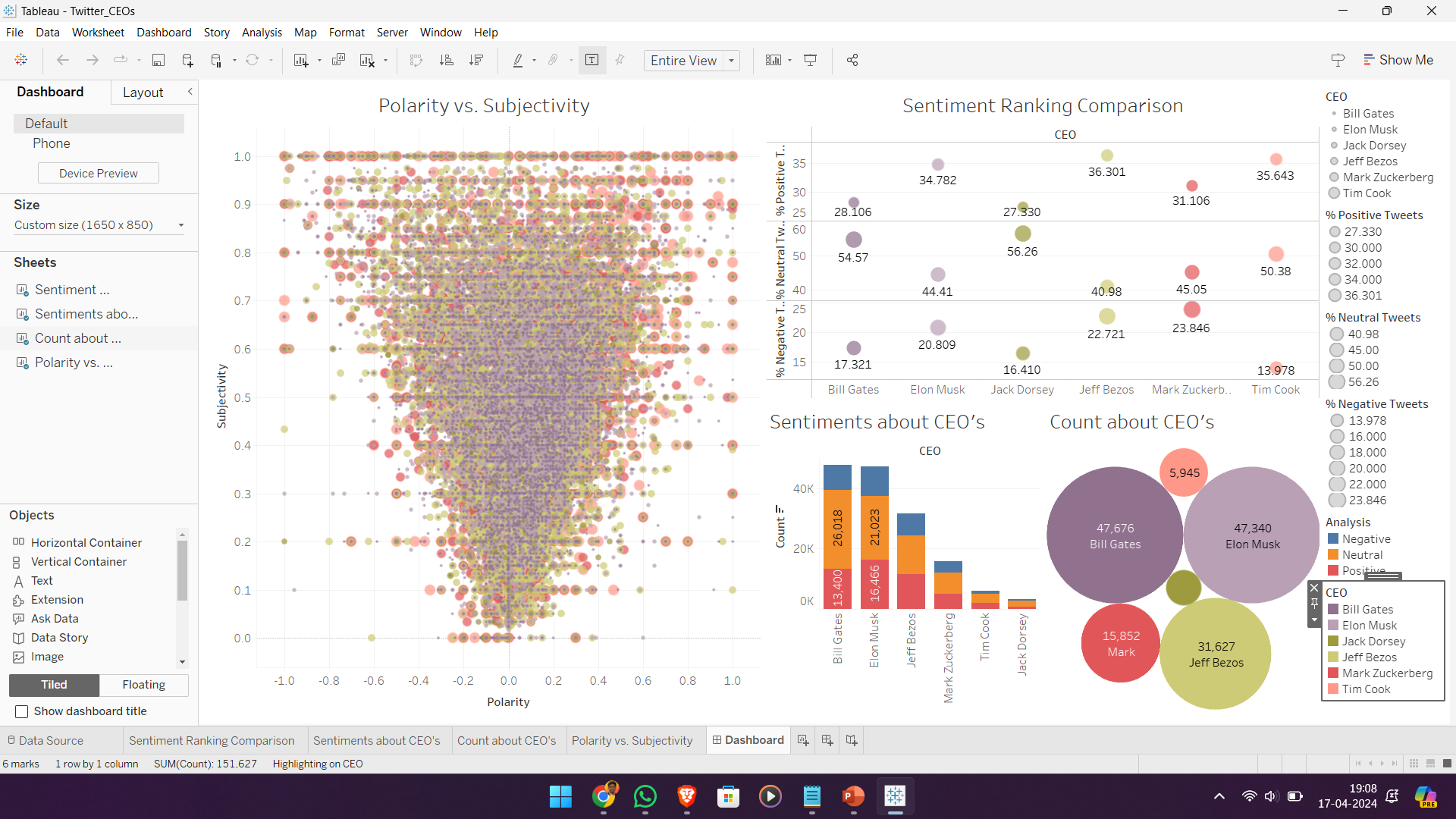
 **Operation Dashboard to Track Engagements and Activity**

The operational dashboard that we propose X uses was on a much smaller scope than X’s actual data would be. We took a dataset of 2000 values and set up a dashboard that assesses tweet activity and engagements during a specified period. This dashboard would mainly act as an operational dashboard, tracking tweet activity, how many engagements tweets are getting, and the sentiment of the tweets being sent out by users. An assumption of this problem is that X already has a program in place that can take raw twitter text and classify sentiment to it.

The KPIs on this dashboard are Total Tweets, Average Engagements by Tweet, Average Retweets by Tweet, Average Likes by Tweet, and Negative Tweet Percentage. The graphs that are shown on this dashboard are Tweets and Engagements by Time-of-Day Line Graph, Tweet Activity by Date Line Chart, Tweets by Sentiment and Source Bar Chart, and Engagements by Tweet Sentiment Bar Chart. There is a filter in place to be able to query at different dates.

The main users of this dashboard would be the software engineers that are concerned about ensuring that the app is running smoothly and efficiently. Inefficiencies in how the app is performing would likely be able to be seen through the total amount of tweets each time and if it is differing from what the norm is. When looking to enhance the user experience to create value, twitter engineers can look at the most common sources for tweets to focus on their user experience while using the app on those interfaces. All sheets and graphs on the dashboard are used as filters meaning the engineers at twitter can focus in on user bases by time of day, source of tweet, time, and tweet sentiment.

**Additional Analysis:**

**CEO’s Tweet Sentiment Analysis Dashboard**

The scatter plot in the bottom-left corner shows how positive or negative tweets are Polarity and how opinionated they are Subjectivity. The top-right quadrant has a bubble chart comparing sentiments about different CEOs, with bubble size indicating tweet volume and position showing positive, neutral, or negative sentiment. Below this is a bar chart showing the count of positive, neutral, and negative tweets for each CEO. To the right of this is another bubble chart showing the total tweet count for each CEO.

This dashboard helps analyze public sentiment toward these CEOs on Twitter. It considers tweet volume, sentiment polarity, and subjectivity to provide a detailed understanding of public opinion.

In addition to dashboards, predictive analysis can also be used by X to improve its overall business and app performance. As mentioned earlier in the paper, spammers or bots are one of the largest challenges that X faces as many people do not have the ability to decipher what accounts are real and which aren’t, potentially leading to human users trusting false information. In addition to false information from bots, many other bots are used to artificially reply to tweets and gain engagements in that way. X can use predictive analytics to classify what accounts are bots on the apps and which are real, human users.

To prove the power that predictive analytics could have for twitter in classifying bot accounts, we utilized a dataset that had the features of UserID, When the account was Created At, when the user was Collected At, Number Of Followings, Number of Followers, Number of Tweets, Length of Screen Name, Length of Description In User Profile, Account Age, Followers to Following Ratio, Average Tweets per Day, Followings Series, Average number of words in a Tweet, Average number of mentions in a tweet, average number of hashtags in a tweet, average number of links in a tweet, and a binary variable to represent whether the account was found to be a spam account or not with 1 indicating spammer and 0 indicated the account is not a spammer. This dataset has over 20,000 instances of users. Using this dataset, we applied four different machine learning classification models to predict whether an account is a spam account using all features except for UserID, Created At, Collected At, and Followings Series. The four machine learning classification methods we used were Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine. A train test split of 85% Training and 15% Testing was used to split the data into training and test sets. The test accuracy and confusion matrices of the machine learning classification models can be seen below:

**A blue squares with white text

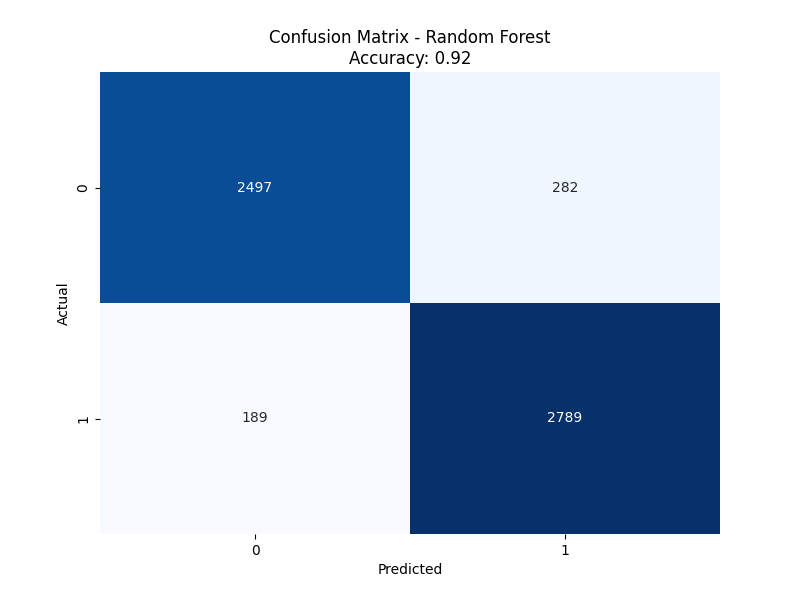
Description automatically generated**

**Logistic Regression Classification Model**

**A blue squares with white text

Description automatically generated**

**Decision Tree Classification Model**

****

**Random Forest Classification Model**

**A blue squares with white numbers

Description automatically generated**

**Support Vector Machine Classification Model**

Using the results from this predictive analysis, it is likely very worth it for X to invest in predictive modeling to classify spam accounts and ban their accounts. The best model for twitter to use is likely the random forest classifier model, as it uses multiple decision trees and a majority voting system whether to classify a user as a spammer or not, leading to less of a chance for overfitting. Also, the random forest classifier model had the highest test accuracy of 91.82%. In addition to the highest test accuracy, the random forest classifier also had the least false positives, meaning in practice this model would limit the number of real twitter users classified as spam the best. Overall, random forest classification could be used by X to classify spam accounts at an extremely high accuracy and banning them from the app, leading to less spam accounts on the app and an overall better user experience for human users.

**Section 5: Implementation**

Although we believe that BI can play a major role in the solution for the problems that X is facing right now, there will be managerial, technical, and ethical implications as well as challenges that the BI implementation process will face. To enable the organizational transformation needed to build a data warehouse needed for the BI solution, Kotter's eight-step model will be used as the framework. Kotter’s eight-step model consists of the following three steps with eight sub steps:

**Prepare:** Create a Sense of Urgency, form a Powerful Coalition, and Create a Vision

**Implement:** Communicate the Vision, Remove Obstacles, and Create Short-Term Wins

**Manage:** Build on the Change, Embed the Change into the Culture

To prepare for the change that will take place when implementing a new BI system at X, it is essential to first create a sense of urgency. In this case, the sense of urgency is simple: If X as a company keeps losing value, the company may have to start cutting costs and laying off employees. When displaying this message, it should create a sense of urgency but also a sense of hope in the workers. When it comes to forming a powerful coalition for this organizational change, it is important to find a powerful leader and someone who employees will listen to. This is not always a top end executive, which can be a common mistake during this step. It is important to utilize the respect that this person’s coworkers have for them and have them lead the change and try to convert employees into believers of the change to create a data warehouse and BI solution. When it comes to creating a vision for X, it is important to have a justification process in which projects and tasks should be checked to ensure that they will bring value to the company ad that they align overall with the hierarchy business objective which in this case is to track tweets, stock prices, and engagement data to help monitor the success of attempting to create a better customer experience for X.

During the implement stage, it is first important to communicate the vision and the overall business objectives clearly to the team who will be responsible for implementation of the BI solution which will likely be a mix of software engineers, business analysts, and various executives within X. Clearly stating the business objectives, vision, and strategy in place to executive the plan is essential to give employees direction on projects, lowering the risk that they do work unrelated to the problem at hand. In addition to communicating the vision, it is important to be continuously removing obstacles. The future is so unknown now as technology moves at a faster rate than ever and because of this, it is essential for the organizational change leader to be highly proactive when removing obstacles through realigning resources, restructuring processes, and addressing any culture issues that may arise during the implementation of the data warehouse and BI solution. During the implementation, it is also critical to provide employees with short-term wins to ensure that they continue to be satisfied and motivated. By breaking the entire implementation of the data warehouse and BI solution into smaller subtasks and goals, it will help keep the organizational change on a steady schedule and build momentum for the team.

To manage the organizational change caused by the implementation of a data warehouse and BI solution, it is essential to build on the change, and truly embed the change into the culture of X. We propose a very minor organizational change when compared to all the data that X has access too, and we believe that this data warehouse and BI solution can lead to other projects aimed to bring more value to X as a company through other business objectives and goals. After employees realize that data warehouses and BI solutions make the tracking of business objectives so much easier, they will be much more likely to use them and even diagnose ways that BI can be used to solve other business problems. Through this organizational change, there will likely be many employees who do not buy in fully right away, but all employees buying-in is necessary for the change to be embedded into the company's culture. Also, fully embedding the changes into the company culture of X may require updates to company policies, procedures, and practices.

With respect to data quality, strict data governance procedures and processes will be implemented into our BI solution to ensure that the data we are using is providing valuable and reliable insights. The most common data quality issues are incompleteness, inaccuracy, inconsistency, and redundancy. Incompleteness will be assessed and handled either through listwise deletion, which is deletion of the entire data row, or through imputation methods. Listwise deletion will only be used when there are many values from the row missing, and these values are essential to analysis. An imputation method will be used when the row is missing one or two values, since these values will be able to be predicted based on the other data present in the row. Inaccuracy will be handled by listwise deletion, as if there is no way to verify that the data is accurate, we will not want it to influence our end analysis and solution. Inconsistency is a problem that usually arises when gathering data from different sources or during any potential mergers. This issue will be mitigated by using software engineers to recode the data, so it matches the data warehouse's standard. Redundancy or duplicate data in the data warehouse will be handled by a listwise deletion after ensuring that one of the datapoints has all the current information. The Data Governance plan will include rules and policies in place to mitigate these problems as much as possible, but when they arise, the following solutions will be a part of the Data Governance plan too.

The main ethical challenge that will arise during the implementation of the data warehouse and BI solution will be customer data privacy and ensuring that their privacy is being kept secure. X as a company should look to encode usernames into unique customer IDs prior to analysis to ensure none of the customers personal data is easily accessible. As we are looking to purchase customer interaction data using IP addresses and cookies, it will be essential to also encode this data once we have received it to ensure that customers' personal information is not a part of any of the final analysis. Another important aspect of keeping all this data private will be having a strong cyber security system and data governance policies in place to ensure that no internal or external threat will be able to directly access a customer’s personal data.

**Section 6: Conclusion**

The implementation of a Business Intelligence (BI) solution at X, focusing on Twitter Inc., aims to address the company's diminishing value post-acquisition by Elon Musk. The proposed solution incorporates various components, including external and operational data integration, strategic dashboards for competitive analysis and stock market assessment, operational dashboards for real-time tweet activity monitoring, and predictive analytics for spam account identification.

The strategic dashboard provides insights into user engagement across social media platforms, indicating Twitter's positioning against competitors like YouTube and LinkedIn. The stock market analysis dashboard offers a comprehensive view of Twitter's stock performance over time, aiding executives in decision-making and performance evaluation. The operational dashboard, though on a smaller scale, tracks tweet activity and engagement, facilitating real-time monitoring for software engineers to enhance user experience.

Predictive analysis plays a crucial role in identifying and eliminating spam accounts, thus improving user trust and experience. The analysis suggests employing the Random Forest classifier model for its high accuracy and low false positive rate.

To successfully implement the BI solution, creating a sense of urgency, forming a powerful coalition, and communicating a clear vision are essential steps. Continuous obstacle removal, short-term wins, and embedding the change into the company culture are crucial during implementation. Data governance procedures ensure data quality and ethical challenges, particularly concerning privacy, are addressed through encoding and cybersecurity measures.

In conclusion, the proposed BI solution offers a holistic approach to enhancing Twitter's value by leveraging data-driven insights to improve user experience, track business objectives, and mitigate challenges such as spam accounts. By implementing these strategies effectively, X can revitalize its market position and foster sustainable growth in the digital landscape.

# **References**

Bratton, L. (2024, January 3). *See X’s value plummet 72% in one chart*. Quartz. <https://qz.com/elon-musk-x-twitter-value-drops-2023-1851136401>

Chavan, A. (2020, September 16). *Twitter data storage and processing*. Medium. <https://ankush-chavan.medium.com/twitter-data-storage-and-processing-dd13fd0fdb30>

Hyperight. (2021, June 30). *Deep Brew: Transforming Starbucks into an AI & data-driven company*. Hyperight. <https://hyperight.com/deep-brew-transforming-starbucks-into-a-data-driven-company/>

Liu, J., & Toh, M. (2023, April 12). *Elon Musk says he’s cut about 80% of Twitter’s staff | CNN Business*. CNN. <https://www.cnn.com/2023/04/12/tech/elon-musk-bbc-interview-twitter-intl-hnk/index.html#:~:text=Elon%20Musk%20has%20laid%20off>

Marr, B. (2021, July 23). *Starbucks: Using Big Data, Analytics And Artificial Intelligence To Boost Performance*. Bernard Marr. <https://bernardmarr.com/starbucks-using-big-data-analytics-and-artificial-intelligence-to-boost-performance/>

Murphy, C. (2019, November 25). *Tweet Twist: The Evolution of Twitter*. [Www.bostondigital.com](http://Www.bostondigital.com). <https://www.bostondigital.com/insights/tweet-twist-evolution-twitter>

Niwate, S. (2021, April 28). *How Netflix uses Data Analytics: A Case Study*. Plainenglish.io/Blog/How-Netflix-Uses-Data-Analytics-Data-Science-General-Research-Case-Study-4d525b881038. <https://plainenglish.io/blog/how-netflix-uses-data-analytics-data-science-general-research-case-study-4d525b881038>

Shivang. (2019, May 30). *What Database Does Twitter Use? - A Deep Dive - Scaleyourapp*. Scaleyourapp.com. <https://scaleyourapp.com/what-database-does-twitter-use-a-deep-dive/#:~:text=Twitter%20started%20with%20MySQL%20as>

*Twitter (TWTR) - Market capitalization*. (n.d.). Companiesmarketcap.com. <https://companiesmarketcap.com/twitter/marketcap/>

*Twitter (TWTR) Stock Price, Quote, News & Analysis - TipRanks.com*. (n.d.). TipRanks | Know Who to Trust. <https://www.tipranks.com/stocks/twtr>

*Twitter Analytics: How to track and measure Twitter campaigns*. (n.d.). Business.x.com. <https://business.x.com/en/blog/twitter-analytics.html#:~:text=Twitter%20Analytics%20is%20a%20powerful>

*What is Data Governance?* (n.d.). Google Cloud. <https://cloud.google.com/learn/what-is-data-governance#:~:text=Data%20governance%20means%20setting%20internal>

# **Dataset References**

*TWITTER DATA*. (n.d.). [Www.kaggle.com](http://Www.kaggle.com). Retrieved April 17, 2024, from <https://www.kaggle.com/datasets/smmmmmmmmmmmm/twitter-data>

*Twitter Stocks Dataset*. (n.d.). [Www.kaggle.com](http://Www.kaggle.com). Retrieved April 17, 2024, from <https://www.kaggle.com/datasets/maharshipandya/twitter-stocks-dataset?resource=download>

*Twitter(X) Spammer Classification*. (n.d.). [Www.kaggle.com](http://Www.kaggle.com). <https://www.kaggle.com/datasets/vinaykumar52/twitter-spammer-classification>