University of Münster

Department of Information Systems

Popularity and Controversy: A Location-Based Event and Sentiment Analysis of Donald Trump and Boris Johnson

Seminar Thesis

in the context of the seminar

Urban Analytics

submitted by

Leo Richard Hermann Giesen

Chairs of Data Science:
Statistics and Optimization &
Machine Learning and Data Engineering

Principal Supervisor	Prof. Dr. Tobias Brandt
Supervisor	Supervisor, M.Sc. Ann-Kathrin Meyer Chairs for Data Science: Statistics and Optimization ML and Data Engineering
Student Candidate	Leo Richard Hermann Giesen
Matriculation Number	462502
Field of Study	Information Systems
Contact Details	leo.giesen@uni-muenster.de
Submission Date	18.02.2024

Abstract

This seminar thesis explores the influence of geographic location on public perceptions of the political leaders Donald Trump and Boris Johnson between 2018 and 2022 through a detailed analysis of Twitter/X posts from Los Angeles, New York City, Birmingham, and London. It demonstrates that local cultural, economic, and political contexts significantly affect public sentiment, with a notable homogeneity within countries that highlights the national context's impact. Key events, including elections and scandals, were found to provoke vital, often critical, reactions, particularly within the leaders' home countries, indicating that significant occurrences have the power to transcend geographical boundaries. This research contributes to urban analytics by highlighting the importance of geographic context in shaping political perceptions, offering valuable insights into how global political events are locally interpreted and responded to.

Contents

1 Motivation and Relevance

X (formerly Twitter) had over 368 million users in 2022, and the US and United Kingdom make up more than 30% of its traffic (EMarketer & Insider Intelligence, 2022). 78% of posts from X users above 50 are political, underscoring X's prominent role in facilitating political discourse and engagement (Bestvater et al., 2021; Demand Sage, 2024). The online posts of users have crucial real-life consequences. For instance, former President Donald Trump incited the tumultuous Capitol riot with the help of posts on X (Naylor, 2021; NBC News, 2021; Steve Holland & Landay, 2021). Additionally, geographic-based sentiment analysis can reflect on-ground public opinions, which further underscores the importance of investigating social media data to comprehend the emotions conveyed by the public (Yaqub et al., 2020).

The current research focuses on event analysis and the relationship between social media sentiment and election results (see Chapter 2). While previous studies have extensively explored sentiment analysis within political discourse, the incorporation of urban analytics to understand the geographic dimensions of public sentiment is notably less common.

This study aims to delve into the dynamics of public perception by analyzing X data pertaining to Donald Trump and Boris Johnson from 2018 to 2022 to understand the political discourse surrounding them before, during, and after their term of office. Since most social media data is georeferenced, the X data from users from Los Angeles, New York City, Birmingham, and London is used to investigate the public sentiment and attention towards their own and other political leaders (Singleton et al., 2017). By examining the normalized post count (NPC) and compound sentiment score (CSS) at both city, country, and overall levels, this study is able to fill the research gap of an event-based comparative local, national, and cross-national political sentiment analysis with temporal and spatial insights. Hence, this study focuses on the factor and effect of **geographic location** on **public sentiment and attention** concerning the **political leaders** Donald Trump and Boris Johnson.

The subsequent sections are structured as follows: First, a closer look is taken at the theoretical background and related work in the literature. Second, the applied methods inspired by the related work are outlined. Third, the analysis results are analyzed and explained. Fourth, these findings are discussed and interpreted. Moreover, the limitations and implications of the results are reflected. Fifth, the findings are concluded, and an outlook on future work is presented.

2 Related Work and Theoretical Background

2.1 Related Work in Political Sentiment Analysis

Political analysis has increasingly recognized the role of social media in shaping public opinion and political discourse. Digital platforms have facilitated a direct and immediate communication channel between political figures and the public, enabling researchers to gauge public sentiment and perception towards political leaders in real-time. The literature covers event-specific polarization, such as the effect of the capitol riot (Norgaard & Walbert, 2023; Pradipta et al., 2023). Recent methodological innovations have underlined the effectiveness of different approaches in capturing public sentiment (Endsuy, 2021). Furthermore, research on location-based X sentiment analysis has illuminated the potential biases inherent in social media datasets, including the non-representativeness of X users and the challenges in accurately capturing geographical distribution (Heredia et al., 2018; Yaqub et al., 2020). The significance of neutral sentiments in these datasets also points to the complexity of public opinion, which extends beyond a binary positive and negative dichotomy (Endsuy, 2021).

Although prior research has thoroughly investigated sentiment analysis in political discourse, including areas like hate speech, as highlighted by Solovev and Pröllochs, the integration of urban analytics to examine the geographical aspects of public sentiment remains relatively rare. The current state of the research in the analysis of political discourse on social media focuses on the correlation between location-based social media sentiment and voting decisions during elections (Endsuy, 2021; Fachrie & Ardiani, 2021; Heredia et al., 2018; Oikonomou & Tjortjis, 2018; A. Singh et al., 2021; S. Singh et al., 2022; Yaqub et al., 2020). However, there are inconsistencies, such as Yaqub et al. contradicting Heredia et al. by concluding that sentiment correlates with election results. Moreover, the literature primarily investigated nationwide patterns on a state level in the US and did not fully compare local and international sentiments.

This study addresses this research gap with an **event-based comparative analysis** of the normalized post count and compound sentiment scores in the selected cities and countries to identify how the **location-specific public perceptions** about Donald Trump and Boris Johnson are affected by political events. This investigation is timely and essential for comprehending the complexities of modern political communication and its ramifications on public opinion. Through comparative analysis, we endeavor to shed light on how local, national, and international factors shape the perception of prominent political figures, thereby contributing to a deeper understanding of contemporary socio-political dynamics.

2.2 Sentiment Analysis Approaches

To guarantee the findings' reliability, validity, and accuracy, it is crucial to carefully select a suitable approach for the sentiment based on the research question. Three types of sentiment analysis could be applied to the X dataset in various ways. First, **Aspect-Based Sentiment Analysis** (ABSA) is a complex approach that offers a granular understanding of sentiment by identifying detailed opinions on specific aspects of a subject, such as "Trump is offensive" (MonkeyLearn, 2023; Selvaraj, 2020; Thematic, 2023). This approach could dissect the multifaceted nature of public opinions toward political figures, enabling the analysis of sentiments related to particular policies, behaviors, or events (Do et al., 2019; Endsuy, 2021). ABSA was excluded because it does not fit the research question of grasping public sentiment and attention.

Second, **Sentiment Scoring** or **Graded Sentiment Analysis** quantifies the level of positivity or negativity expressed in text, enabling the analysis of the research question (Hutto, 2014). It can be implemented with a machine learning or lexicon-based approach (Selvaraj, 2020). The effectiveness of the VADER approach was established by related work (Anwar et al., 2021; Bekmanova et al., 2023; Pai et al., 2022; Rita et al., 2023; Thakur, 2023). VADER stands for Valence Aware Dictionary and sEntiment Reasoner and is particularly adept at handling social media content because it understands slang terms and captures the emotion of emojis. It employs a lexicon- and rule-based approach to generate a compound sentiment score, which is utilized for the analysis. "The compound score is computed by summing the valence scores of each word in the lexicon [n], adjusted according to the rules, and then normalized [with the normalization factor α] to be between -1 (most extreme negative) and $+1$ (most extreme positive)" as visualized in equation 2.1 (Keita, 2022; Swarnkar, 2020).

$$\text{CSS} = \frac{\sum_{i=1}^{n} S_i}{\sqrt{\sum_{i=1}^{n} S_i^2 + \alpha}} \tag{2.1}$$

2.1 Calculation of the Compound Sentiment Score (CSS)

Third, **emotion detection** addresses the research question by identifying specific emotions conveyed in a text, such as happiness, frustration, anger, and sadness (Pennebaker et al., 2001). Nevertheless, the approach was disregarded here because it is not as corroborated in the literature, and implementing a second approach is out of scope for this seminar thesis.

3 Methods

3.1 Research Method

The research gap outlined in the previous chapter is addressed by employing a structured procedure inspired by and adapted from Pang and Lee, Liu, Bird et al., Mikolov, Sutskever, et al., and Hutto and Gilbert. At first, the data source is selected and data is collected and preprocessed. After that, the relevant features are extracted, providing a solid foundation for the subsequent sentiment analysis. Finally, the analysis results are interpreted based on the research about political developments.

Since X is a prominent platform for political discourse, it offers a rich data source for analyzing public opinion (Bestvater et al., 2021). Relevant **data** was **collected** by selecting posts, which included the full names of Trump and Johnson, with specific considerations for informal mentions such as 'Boris' and relevant hashtags, such as 'MakeAmericaGreatAgain' for Trump and 'UKPrimeMinister' for Johnson. These posts still included data from non-targeted geographic locations, such as posts from Rotterdam and Amsterdam mistakenly included in the London dataset up to April 2019. Therefore, the correct country code was selected to clean the dataset. Reposts were retained under the assumption that they echoed the sentiment and opinion of the reposter. Language filtering was deemed unnecessary due to VADER's capability to assign sentiment scores regardless of language nuances (Hutto, 2014). Further **data preprocessing** and **cleaning** were achieved by detecting spam in X to ensure the sentiment analysis's integrity and relevance. This study adopted a user behavior analysis approach complemented by a bigram similarity algorithm with a similarity threshold of 90%, as recommended by Varol and Tareq Abdulhadi. The bigram similarity algorithm measures the similarity of two posts by comparing every combination of two consecutive words (bigrams) within the posts as presented in equation 3.1 (Manning & Schütze, 1999; Mikolov, Chen, et al., 2013; Pennington et al., 2014).

$$\text{similarity}(S1, S2) = \frac{Bigrams(S1) \cap Bigrams(S2)}{Bigrams(S1) + Bigrams(S2)} \tag{3.1}$$

3.1 Calculation of the Bigram Similarity for Spam Detection

The study **extracted** the **features** of the compound sentiment score and the post count. The sentiment score was calculated with VADER because of its previously mentioned effectiveness. The post count is affected by a fluctuating population size. This variability can introduce biases in the analysis. This bias is addressed by employing normalization to the post counts by dividing the posts of the presidents in the specific location by the total number of posts for the corresponding month in the location (see equation 3.2). So the normalized post count expresses the percentage of the public speaks about the president compared to all posts.

$$\text{Normalized Post Count: } NPC_{L,D,G}(L,G,D,M) = \frac{\text{Post Count about } L \text{ in } G \text{ on } D}{\text{Post Count in } G \text{ in } M} \tag{3.2}$$

3.2 NPC Calculation (Leader L, Geographic Location G, Day D, and Month M)

3.2 Method of Analysis

Two types of analysis contribute to the overarching **sentiment analysis**. First, an **event analysis** captures reactions to specific events or news. Moreover, the relationships between sentiment scores and the engagement metric post counts are analyzed to identify if the public's reaction is positive or negative. Second, a **comparative analysis** examines the sentiment scores across different locations. The analysis were conducted by data aggregation to identify overall baselines and compare the mean sentiment. Moreover, data visualizations aided in identifying patterns and anomalies and testing hypotheses. Further, specific data instances were selected to investigate outliers. The observed patterns were interpreted based on research, focusing on understanding the impact of geographic location on public sentiment toward political figures.

These analysis consider five factors influenced by unique regional, national, and international contexts. First, the **awareness level** of presidents is gauged by the number of posts about each president. Second, the **popularity** examines the mean sentiment score of posts regarding the heads of state. Third, the **event impact** is evaluated by inspecting patterns in the data and seeking explanations for observed sentiment shifts and peaks in post counts. Fourth, the comparison of sentiments across cities uncovers **regional differences**. Fifth, **temporal trends**, such as a higher baseline during the term of office, are identified.

Considering both quantitative post-data and qualitative sentiment analysis, this comprehensive approach offers a nuanced understanding of public opinion toward political figures, highlighting the complex interplay of factors that shape public sentiment. This methodology provides a comprehensive framework for analyzing the nuanced interplay between geographic location and public perception of political leaders through sentiment analysis of X data. The approach combines data preprocessing, sentiment analysis, and various forms of data analysis to uncover insights into public opinion dynamics.

4 Results

4.1 Comparative Analysis

Great Britain has a higher CSS **baseline** than the U.S., which could be attributed to cultural differences in communication, where the stereotypical directness of Americans may result in more explicit negativity, which is more easily picked up from VADER (see Figure 4). Another factor is the heightened political division in the U.S., potentially leading to a more critical view of politicians. This explains why Trump has an overall negative CSS of **-0.05** compared to Johnson's relatively positive **0.06** (see Figure 3).

The overall sentiments in Great Britain and the U.S. differ, whereas the CSS has a very similar development within the cities of each country, emphasizing how **national contexts** shape public perception. For instance, national media may overshadow local concerns (see Figures 9-12, 16-19, 22 and 25).

Expectedly, Trump has an overall higher NPC of **0.042%** compared to Johnson's **0.013%** (Jackson, 2019). Both leaders are treated more critically and receive more attention in their home country compared to other countries (see Table 1, figures 4, 21 and 24). The less critical viewpoint in other countries is potentially due to their more negligible direct influence through policies on the daily lives of individuals in foreign nations (Center, 2023; UK, 2021).

Political Leader	**NPC U.S.**	**NPC G.B.**	**CSS U.S.**	**CSS G.B.**
Donald Trump	0.054	0.016	-0.056	0.023
Boris Johnson	0.004	0.032	0.175	0.032

Table 1 NPC (in %) and CSS Means of Political Leaders in the U.S. and G.B.

Despite Donald Trump's controversial status, he exhibits relatively small variance in CSSs compared to Johnson, which might be the case because the mean sentiment calculation could neutralize the strong opinions from both his followers and opponents, which is still present after polarizing events, such as the Capitol riot (see Figures 6 and 13; Norgaard and Walbert, 2023). For Boris Johnson, a slightly higher mean CSS could reflect a less intense public opinion compared to Trump (see Figure 3). Both leaders show mean CSSs significantly below the general mean, potentially due to critical media coverage, controversial policies, and political or legal scandals affecting public sentiment.

The comparative analysis of NPCs for President Donald Trump between the selected **countries** reveals a consistently higher baseline in the U.S., likely reflecting Trump's direct political influence as the president. Concurrently, both nations demonstrate correlative surges in NPCs during critical events, indicating global responsiveness to significant political moments, exemplified by the collective attention during the Capitol riot on January 6th-7th, 2021 (see Figure 21).

In both countries, both leaders exhibit a lower CSS and higher NPC baseline during their term of office than before or after it[1]. The observation suggests that they received more attention during their presidency due to their position of power. However, the public is ultimately disappointed in them or at least viewed them more critically. The effect can be observed in the U.S. and Great Britain, so the location is not a factor here.

The absence of seasonal trends in discussion volume suggests that political sentiment on social media is primarily event-driven, with specific actions by political figures and significant political events catalyzing public engagement (Theocharis et al., 2015).

4.2 Event Analysis

4.2.1 Donald Trump

The NPC data for President Donald Trump shows three significant peaks towards the end of his presidency. These events affected the NPC in all selected locations, so their importance **transcended local and national borders**. The first peak occured from November 4th-8th, 2020, where 0.294% of all posts were about Trump, which correlates with the aftermath of the U.S. presidential election, a period marked by extensive political discourse and media coverage (BBC News, 2020b; Bovet & Makse, 2019). The official withdrawal from the Paris Climate Agreement also happened on November 4th, 2020, but it is not considered influential at this time because it was announced on June 1st, 2017 (School, 2017). The second peak was from September 30th to October 2nd, 2020, averaging a NPC of 0.238%, linked to the first **presidential debate** and Trump's **Covid-19 diagnosis**, events that garnered considerable public attention (The New York Times, 2020). Finally, the third location from January 6th to January 7th, 2021, aligns with the U.S. Capitol riot, where 0.309% of the U.S. X discourse was about Trump with a negative CSS of −0.15. Trump's account

[1] Trump has a CSS of −0.05 and NPC of 0.059% during his presidency from 20.01.2017 to 20.01.2021, and after it, he is viewed similarly −0.03 but loses almost 80% of his NPC down to 0.013%. Before Johnson's position as Prime Minister, he achieved a CSS of 0.13 and NPC of 0.007% up to his term and had half the positive perception with a CSS of 0.05 and double the NPC of 0.015% during it (24.07.2019 until the end of the dataset).

on X was suspended because he encouraged his supporters to march to the Capitol (Naylor, 2021; NBC News, 2021; Reuters, 2021; Steve Holland & Landay, 2021). Unexpectedly, there is some stability in NPCs and CSSs on X throughout Donald Trump's impeachment trials[2], which might be traced back to the fact that the second impeachment trial was the aftermath of the Capital Riot and occurred at the end of Trump's term. This period did not markedly alter public discourse or sentiment on social media, likely due to entrenched political alignments, which did not represent a shift in public opinion and may buffer against significant sentiment shifts in response to political controversiesConover et al., 2011; Galston, 2020; Saad and Brenan, 2021.

The particularly low CSS of −0.31 in Great Britain and not in the U.S. on January 5th, 2020, can be traced back to the assassination of the Iranian major general Qasem Soleimani. This **international difference** might be because of the high focus on foreign affairs exemplified by the high U.S. defense spendings.

The effect of **local events** can be observed in Great Britain on June 3rd-4th, 2019, where the spike in post activity aligns with President Trump's state visit (NPC of 0.165%), indicating a heightened local interest, which is not mirrored in the U.S. (NPC of 0.060%; Hallemann, 2019).

Unexpectedly, **Covid**-related events, such as the declaration of a public health and national emergency on January 31st, 2020, and March 13th, 2020, were minor NPC peaks of 0.048% and 0.090% and insignificant CSS deviation from Trump's mean (−0.04 and −0.10), but not as significant as in Boris Johnson's case (see Figures 20 and 23; White House, 2020).

4.2.2 Boris Johnson

The focus of the British public on the pandemic over politics might have led to a dip in posts about Johnson starting from mid-December 2019 until Johnson's hospitalization due to Covid-19 on April 6th, 2020 (see Figure 24; BBC News, 2020a; Crace, 2020). Unexpectedly, the event sparked an intense positive debate with a sentiment score of 0.21 and a NPC of 0.057% (see Figure 13). The hospitalization might have given people hope that Johnson recognizes the severity of Covid. There are several further Covid-related NPC peaks in the Johnson discussion on X. For instance, Johnson's statement about Covid measures and lockdown on May 10th, 2020, and October 31st, 2020, resulted in a rather neural widespread public discourse with NPCs of 0.068 and

[2] In the first impeachment (December 18th, 2019, until February 5th, 2020), Trump experienced a mean NPC of 0.068% and a CSS of −0.05. The second impeachment, from January 13th to February 13th, 2021, exhibited a mean NPC of 0.042% and CSS of −0.04. In both cases, all values are close to the overall mean.

0.060 and a CSS of 0.060 and −0.05 (Blackall, 2020; Mason, 2020; UK Government, 2020a, 2020b). On July 12th, 2021, Johnson gave a press conference discussing the step towards lifting Covid restrictions in England, which was followed by a significant adverse public reaction, expressed by a mean CSS of −0.18 and NPC of 0.023% for the following two days (see Figure 13; UK Government, 2021).

Similar to Trump, Johnson's **election campaign** represented a significant boost in popularity with numerous CSS peaks between 0.28 and 0.50 between November 21st, 2018, and July 24th, 2019 (see Figure 13). This period coincides with Theresa May's resignation, Boris Johnson's campaign for the leadership of the Conservative Party, his subsequent election as party leader, and the positive public reaction to his strong stance on Brexit, leading to his appointment as Prime Minister on July 24th, 2019 (AP News, 2020; Stewart & Mason, 2019).

Johnson had multiple **scandals** during his office, which took form through a peak in NPC and CSS. These were mainly present in Great Britain because Johnson primarily had national importance and impact. For instance, on September 24th, 2019, Britain's Supreme Court ruled against Prime Minister Boris Johnson's suspension of Parliament and deemed it illegal. The ruling came with a surge in public attention and an expected slightly more negative perception of Johnson (NPC of 0.060% and CSS of 0.04; see Figure 13; Bowcott et al., 2019; Landler, 2019). There is a high concentration of negative sentiment from November 20th, 2020 until April 16th, 2022 (see Figure 13). On the start day, Johnson ignored findings that Home Secretary Priti Patel had bullied staff, breaching the ministerial code and leading to the resignation of his ethics adviser (NPC of 0.015% and CSS of −0.10; see Figure 13; Stewart and Murphy, 2020). Additionally, Johnson had to face criticism for the Partygate scandal in April 2022, where he got fined for attending a party during lockdown (NPC of 0.018% and CSS of −0.02; see Figure 13; Osborne, 2022).

5 Discussion

5.1 Synthesis of Findings

The research question focuses on the factor and effect of geographic location on public sentiment and attention concerning the political leaders Donald Trump and Boris Johnson. The analysis yielded that Trump and Johnson both garnered more attention and criticism domestically, suggesting that the national relevance of political figures and national political orientation significantly affects public discourse leading to a high homogeneity of NPCs and CSSs within a country. The study also observes temporal trends, highlighting a peak in attention and critical sentiment during their respective terms in office, reflecting the public's heightened scrutiny of leaders in power. Event analyses indicate global responsiveness to significant political events, transcending local biases. This research underscores the complex interplay between location, event significance, and temporal context in public political sentiment and attention, providing valuable insights for understanding modern political communication dynamics.

5.2 Implications for Sentiment Analysis and Political Analysis

The findings from this thesis have profound implications for urban analytics and political sentiment analysis research by introducing different levels of granularity of geospatial data. The previously outlined insights can be transferred to future sentiment and political analysis. For example, the identification of differences between national narratives and homogeneity within a country can help generalize local analysis to a nation-wide context. Here, it is necessary to consider cultural and political context, nation-wide baselines of CSS and NPC and their variance during analysis. The insights of this study transcend findings of approval ratings and yield more comprehensive insights into the nuances of location-specific political discourse.

Moreover, the higher sentiment score baseline for Boris Johnson in Great Britain, compared to Donald Trump in the U.S., may indicate cultural differences in communication and political division, affecting public sentiment towards politicians. This highlights the need for sentiment analysis tools to account for cultural and demographic factors when interpreting social media data (Hargittai, 2015).

The similarity in sentiment baselines within cities like Los Angeles and New York City, and London and Birmingham underscores the influence of national media and the overshadowing of local concerns by nationwide issues. This observation challenges the assumption that sentiment analysis can easily discern local from national sentiment, pointing to the complexity of media influence on public opinion (Barberá, 2015a).

5.3 Limitations and Future Research Directions

While the analysis provides valuable insights into the sentiment towards Donald Trump and Boris Johnson, it also reveals limitations inherent in using social media data for political sentiment analysis. The **representativeness** of cities like London, Birmingham, Los Angeles, and New York City may only partially capture the national sentiment, suggesting a need for broader geographic sampling in future studies as the sentiment in rural areas may differ significantly (Malik et al., 2015).

While providing insightful findings on the sentiment towards political figures in specific urban contexts, this seminar thesis acknowledges several limitations that must be considered when interpreting the results. For example, despite multiple sources and the source repository claiming that VADER captures the sentiment of **emojis** (Hutto, 2014; Malde, 2020; Todi, 2019), it could not be verified in the dataset. For instance, "#BorisJohnson Survives" followed by two enraged emojis has a compound sentiment of 0.0. Moreover, the algorithm might misinterpret sarcasm, humor, or context, leading to inaccurate values.

The **representativeness** of cities like London, Birmingham, Los Angeles, and New York City may not fully capture the national sentiment, suggesting a need for broader geographic sampling in future studies (Malik et al., 2015). The unique **demographic** and **socio-economic makeup** of these cities introduces potential biases. London's diversity and Birmingham's distinct cultural identity may not fully capture the sentiment prevalent in rural or less urbanized areas of Great Britain (Florida, 2002). Similarly, Los Angeles and New York City, with their significant influence and distinct cultural landscapes, may not represent the diverse political beliefs and sentiments found across the United States (Glaeser, 2011). The liberal **political leanings** of these cities contrast with more conservative regions, potentially skewing sentiment analysis results. The political heterogeneity of Great Britain and the United States necessitates consideration of a more comprehensive array of urban and rural perspectives to achieve national representativeness (Rodden, 2006). The sentiment can also be heavily influenced by **media coverage** and local events. The dominant role of national media might overshadow local concerns, yet local events can significantly impact sentiment within these cities, necessitating a nuanced approach to analyzing sentiment drivers (Iyengar & Kinder, 2010). Moreover, higher **education levels** and **historical political affiliations** in these cities might influence public sentiment in ways not indicative of broader national trends (Highton, 2006).

The analysis is also subject to limitations inherent to **social media dynamics**. X's user **demographics skew** towards younger, more technologically adept individuals, which may differ from the general population's sentiment (Demand Sage, 2024; Mislove et al., 2011). Future research could explore methods to mitigate these biases and enhance the representativeness of social media-based sentiment analysis. Additionally, the nature of online discourse, characterized by the amplification of minority opinions and the impact of social media influencers or political groups, introduces further complexity to sentiment analysis (Barberá, 2015b). Moreover, undetected spam might have skewed results.

The economic situation, in the selected cities may also impact the sentiment because poor **economic conditions** are likely to lead to a more negative overall sentiment. The economic situation including the job market and the gravity of the disparities between the rich and poor, differ from other parts of each country, potentially affecting sentiment toward political figures (Dunn et al., 2017).

The thesis advocates for a more encompassing approach that includes diverse geographic regions to understand public opinion better. Acknowledging the intricate role of geographic, cultural, economic, and political factors, the conclusion calls for future research to consider broader contexts for a nuanced political sentiment analysis. As sentiment analysis continues to evolve, incorporating these considerations will be crucial for accurately gauging public opinion and its implications for political dynamics in the digital age.

6 Conclusion and Future Work

6.1 Research Objective and Findings

This seminar thesis has explored the impact of geographic location on public sentiment and attention towards Donald Trump and Boris Johnson, utilizing X's social media data from Los Angeles, New York City, Birmingham, and London. The findings indicate that geographic location, with its unique cultural, economic, and political features, significantly influences public sentiment toward political leaders. The analysis shows a homogeneity in public perception within countries, underscoring the national context's role in shaping sentiment. Significant events, such as elections and scandals, transcend national borders. During their time in office, the leaders experience more scrutiny, leading to a more critical view and higher attention, primarily prevalent in their home country.

6.2 Future Research Directions and Implications

This thesis opens new avenues for further research and significantly contributes to urban analytics and political sentiment analysis. **Segmentation of sentiment scores** into positivity level intervals for a nuanced analysis of public sentiment towards political leaders could uncover strong sentiments negate each other in mean sentiment calculations, providing deep insights into the public's emotional intensity regarding political issues and leaders (Endsuy, 2021; Norgaard & Walbert, 2023).

A detailed exploration of **Linguistic Inquiry and Word Count** data to identify regional differences in specific emotions elicited by Boris Johnson and Donald Trump. Additionally, a borough-level analysis offers a promising avenue to map political sentiment with greater detail, identifying regions with pronounced preferences for or against political leaders and connecting them to local characteristics. Moreover, data from more regions could be explored to validate the representativeness of the data here. It is also worth investigating the public perception of **Johnson's resignation** as the dataset did not extend until September 6th, 2022.

The contributions of this study lay a solid foundation for future work, elucidating the significant role of geographic location in shaping public perception of political figures. It is crucial to understand the local developments of the public's political attitude, which can be fortunately performed by applying the outlined methodology to future evaluations of politicians in social media discourse.

A Appendix

A.1 Data Overview

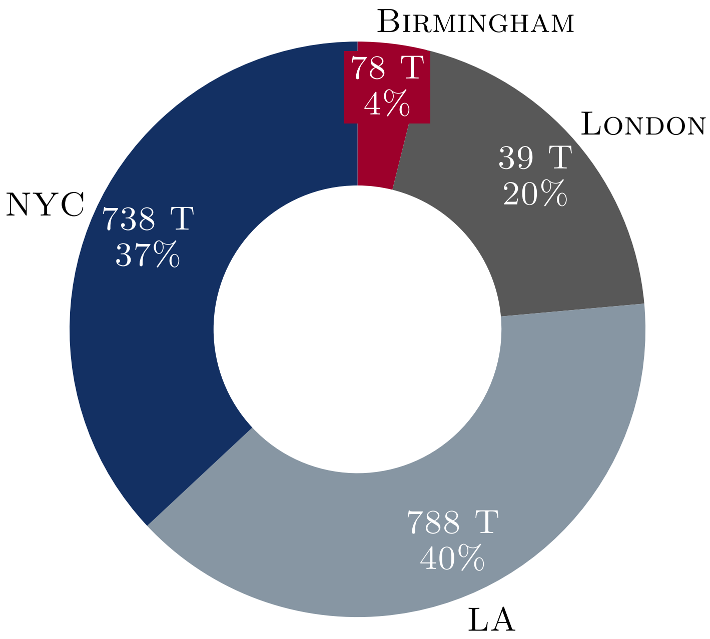
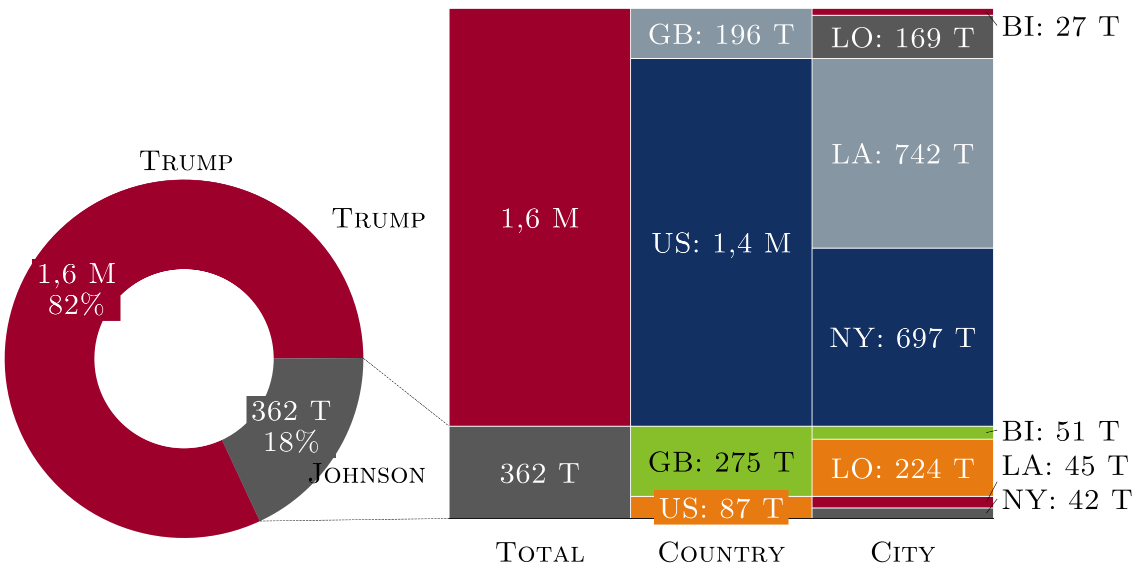


Figure 1 Location-Specific Number of Posts (in Thousands (T))



Birmingham (BI), London (LO), Los Angeles (LA), New York City (NY)

Figure 2 Location-Specific Number of Posts (in Millions (M) or Thousands (T))

A.2 Aggregated Sentiment

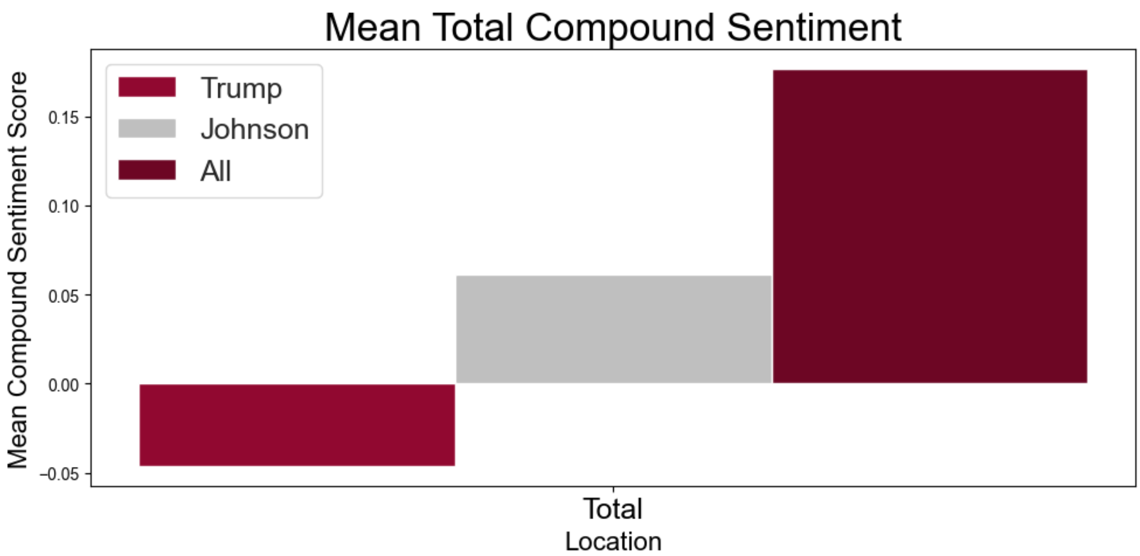


Figure 3 Mean Compound Sentiment

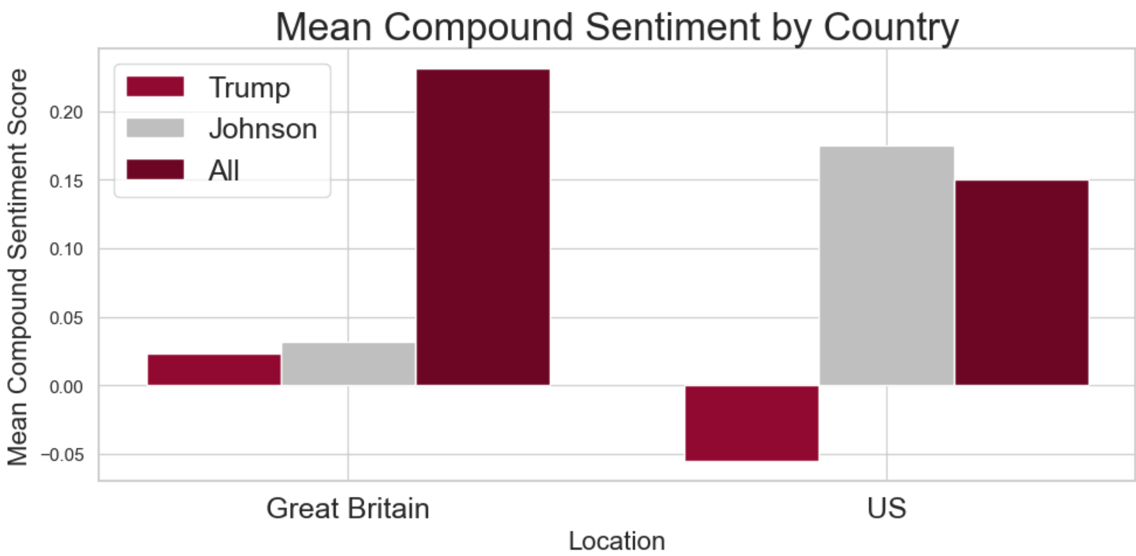


Figure 4 Mean Compound Sentiment Score by Country

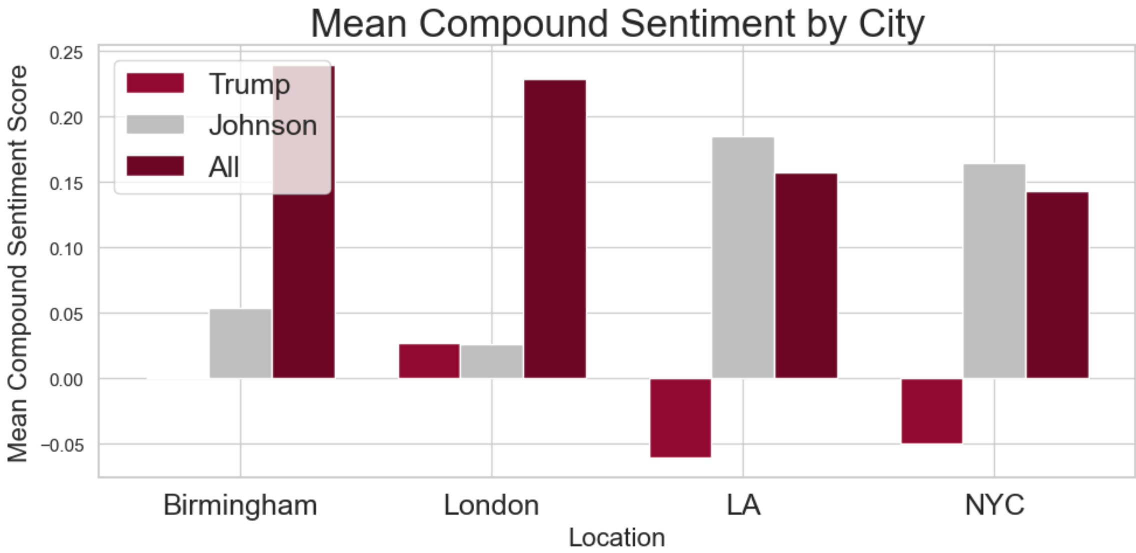


Figure 5 Mean Compound Sentiment Score by City

A.3 Sentiment Score Development

A.3.1 Donald Trump

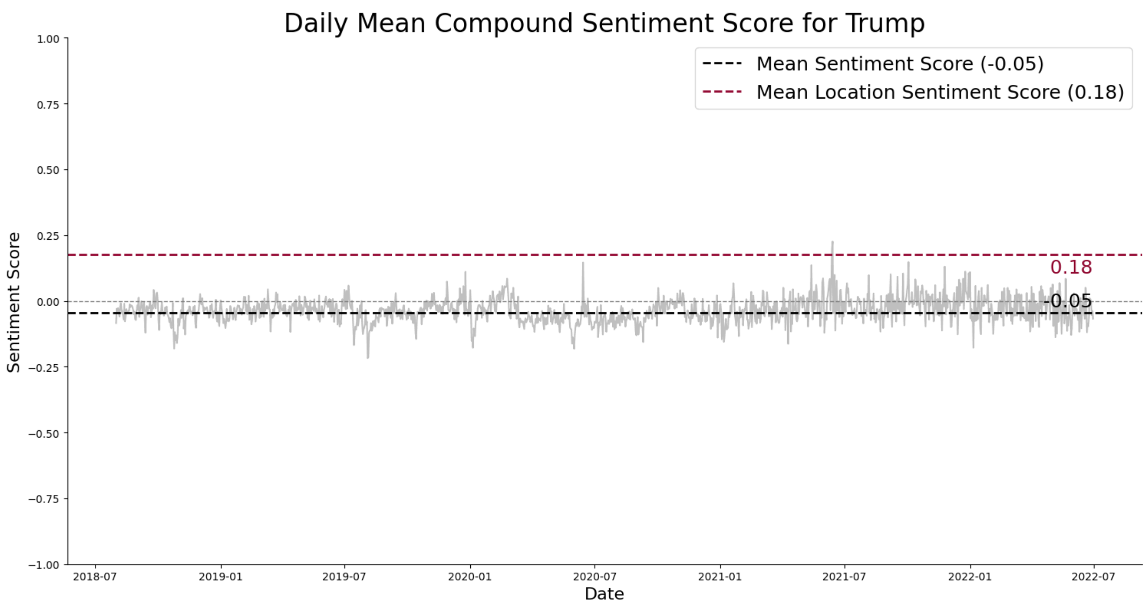


Figure 6 Sentiment Score Development of Trump

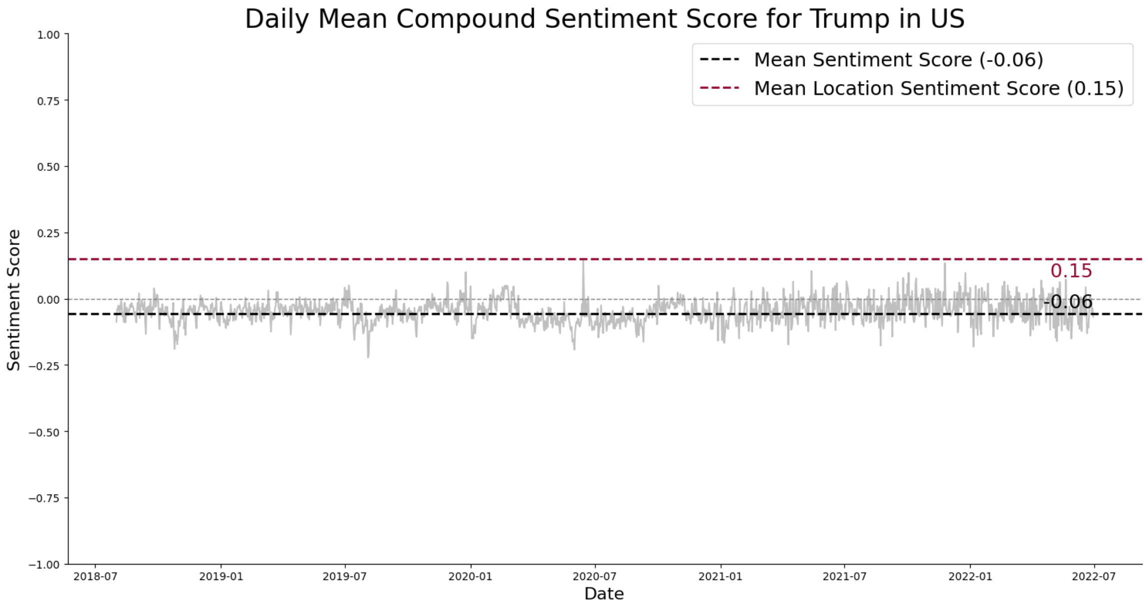


Figure 7 Sentiment Score Development of Trump in the United States

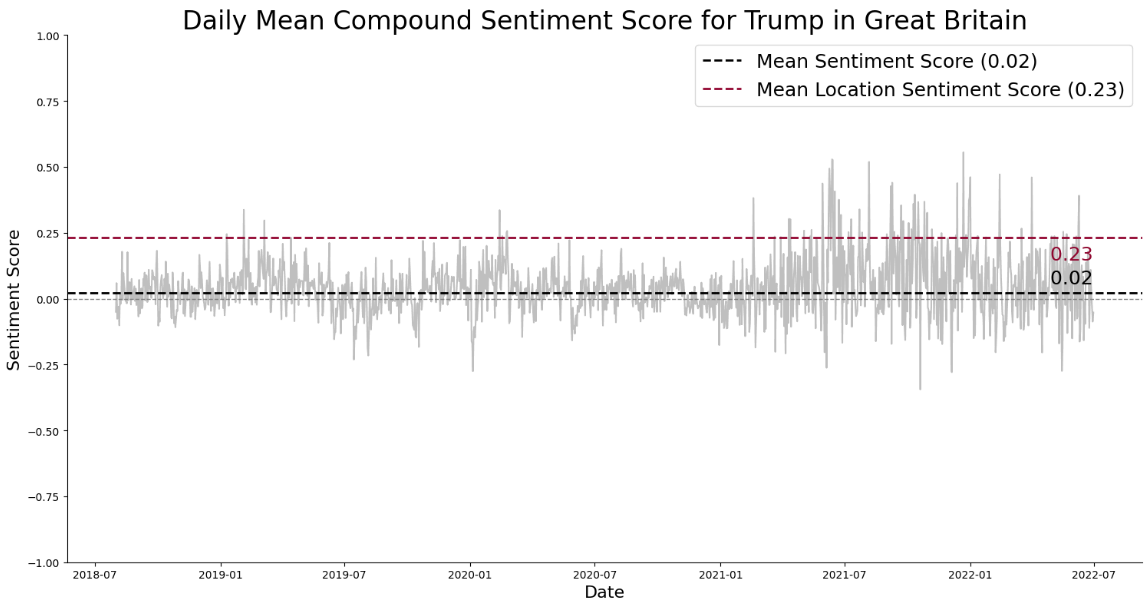


Figure 8 Sentiment Score Development of Trump in Great Britain

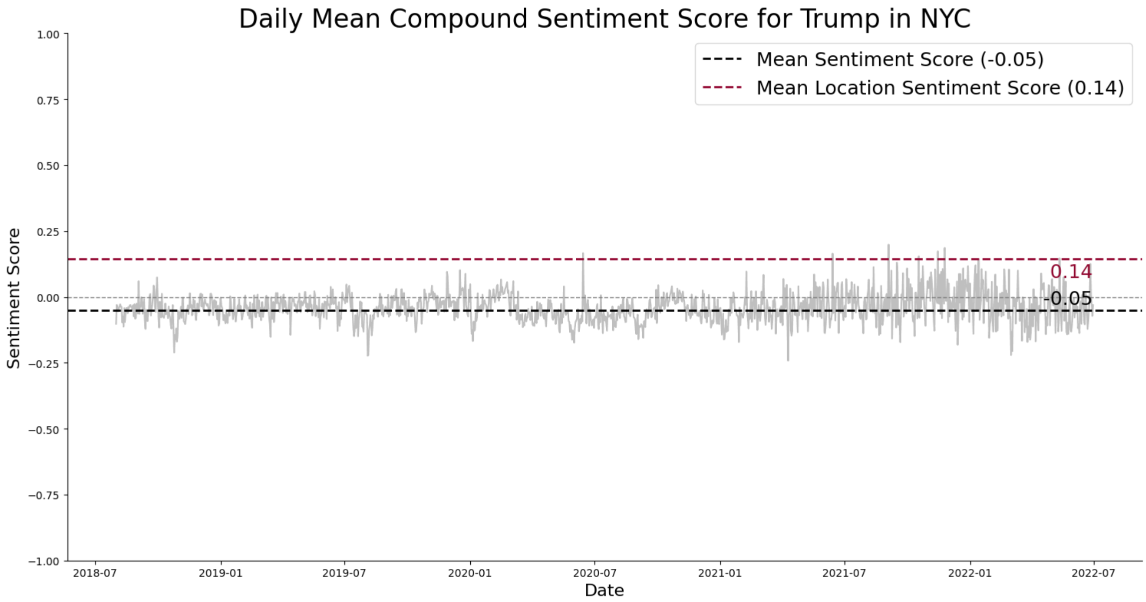


Figure 9   Sentiment Score Development of Trump in New York City

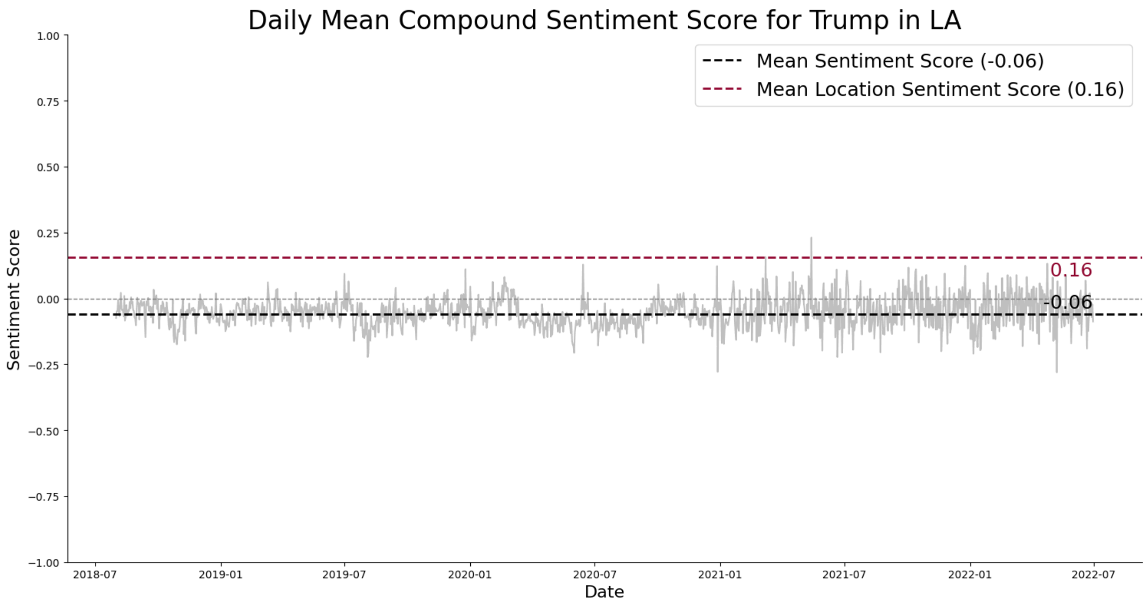


Figure 10   Sentiment Score Development of Trump in Los Angeles

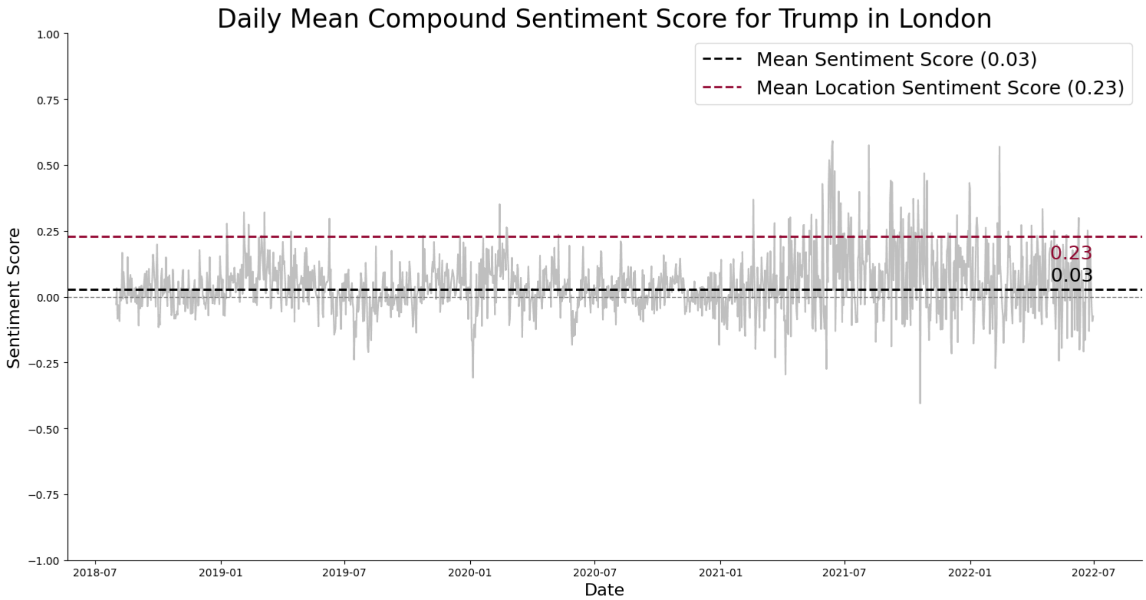


Figure 11 Sentiment Score Development of Trump in London

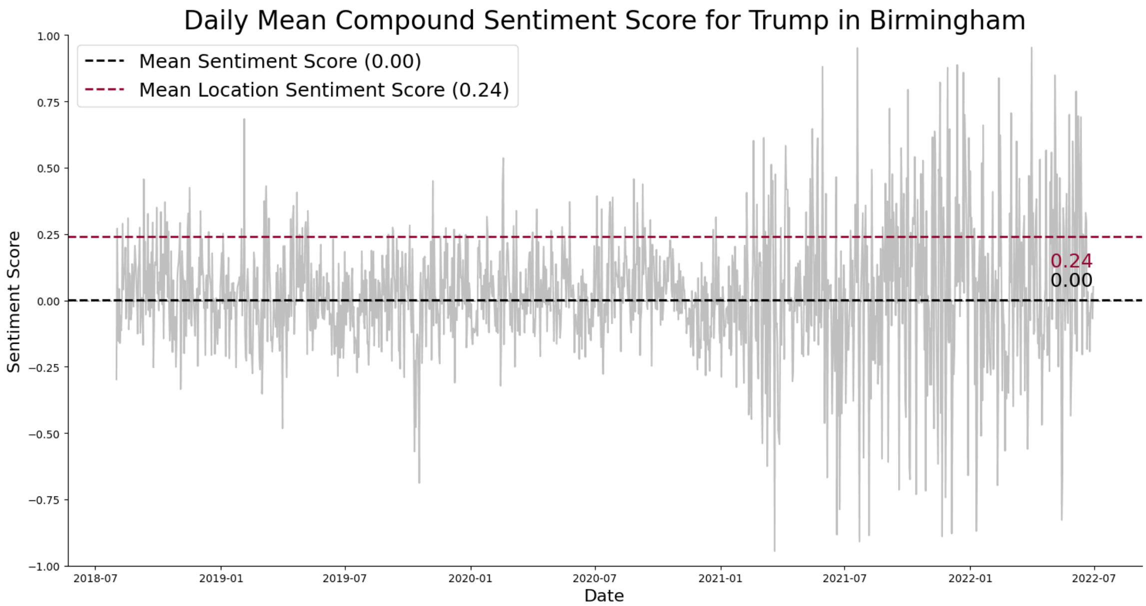


Figure 12 Sentiment Score Development of Trump in Birmingham

A.3.2 Boris Johnson

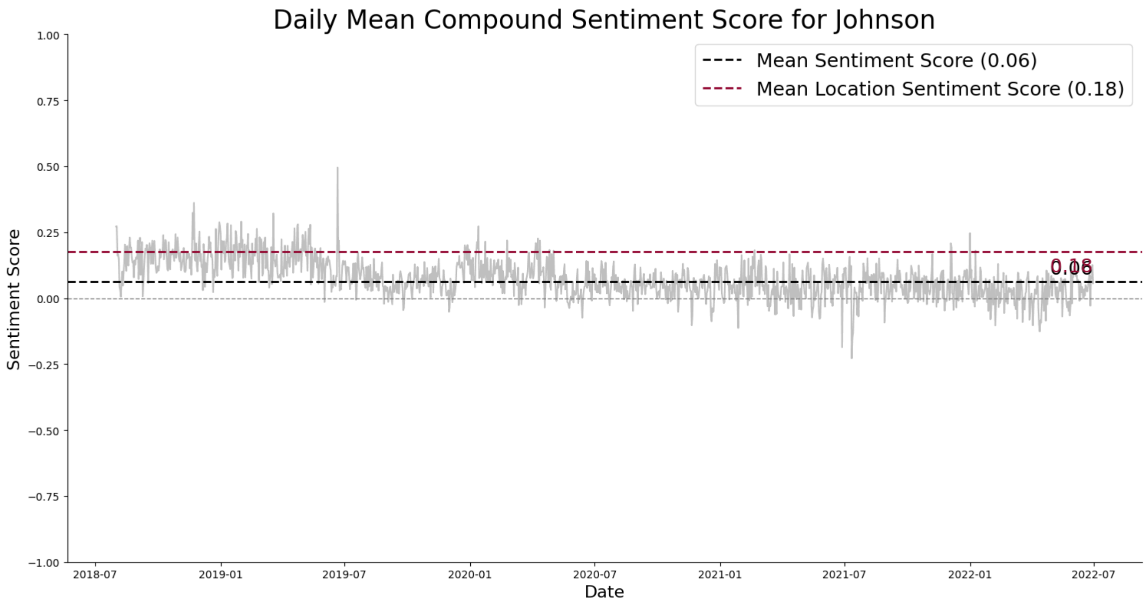


Figure 13 Sentiment Score Development of Johnson

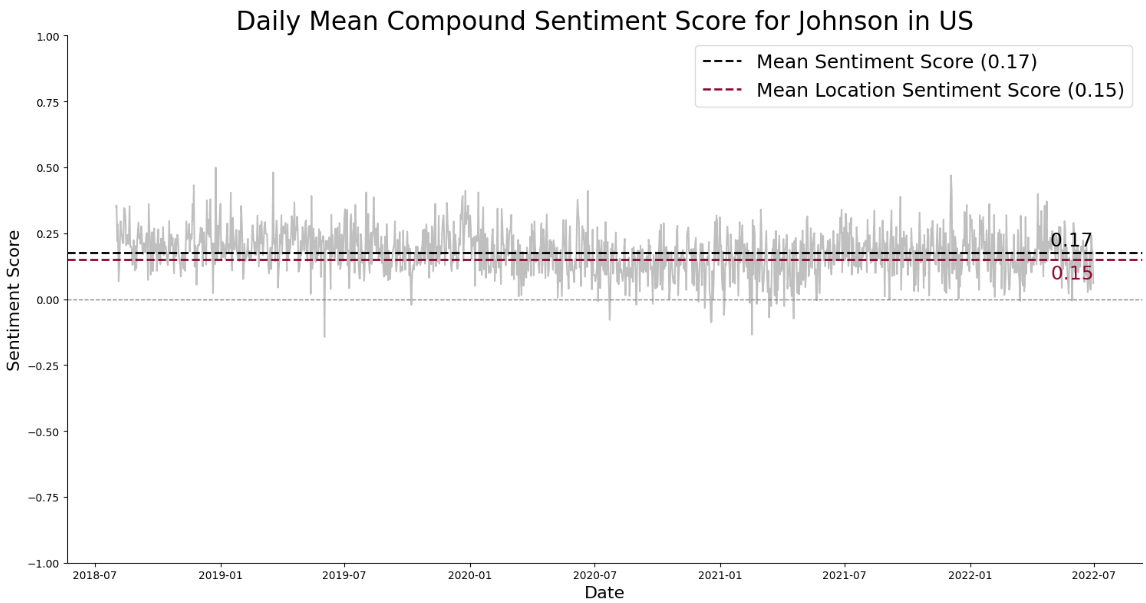


Figure 14 Sentiment Score Development of Johnson in the United States

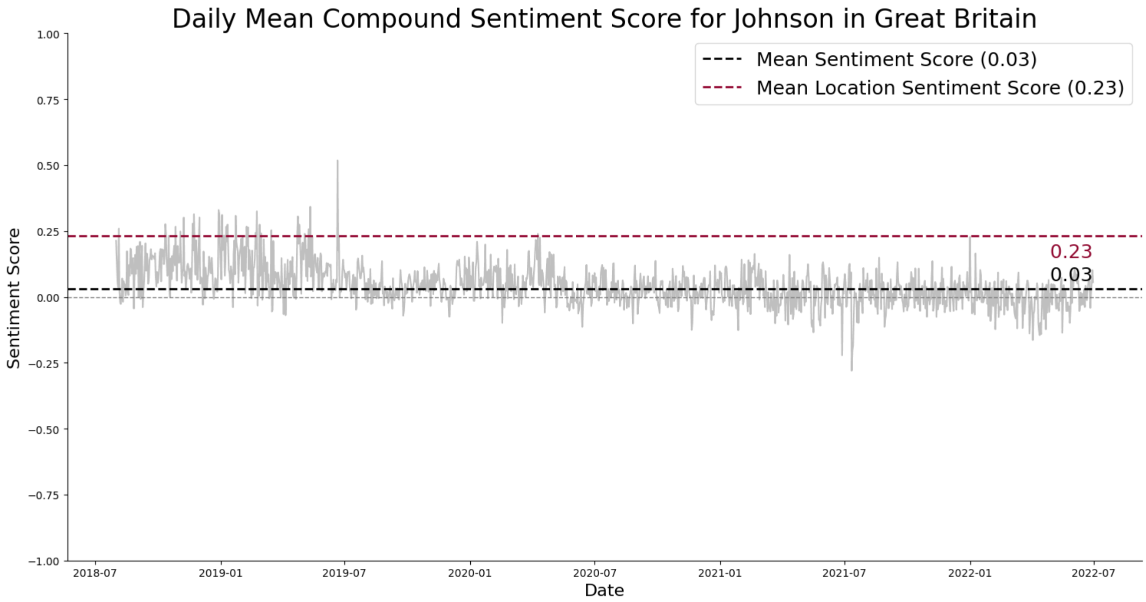


Figure 15 Sentiment Score Development of Johnson in Great Britain

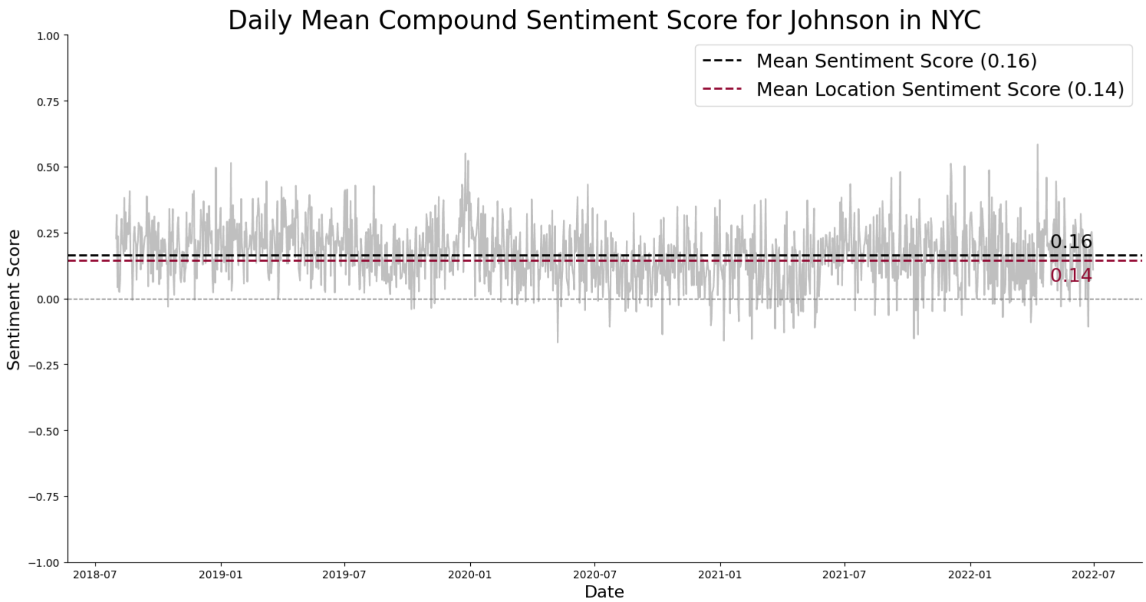


Figure 16 Sentiment Score Development of Johnson in New York City

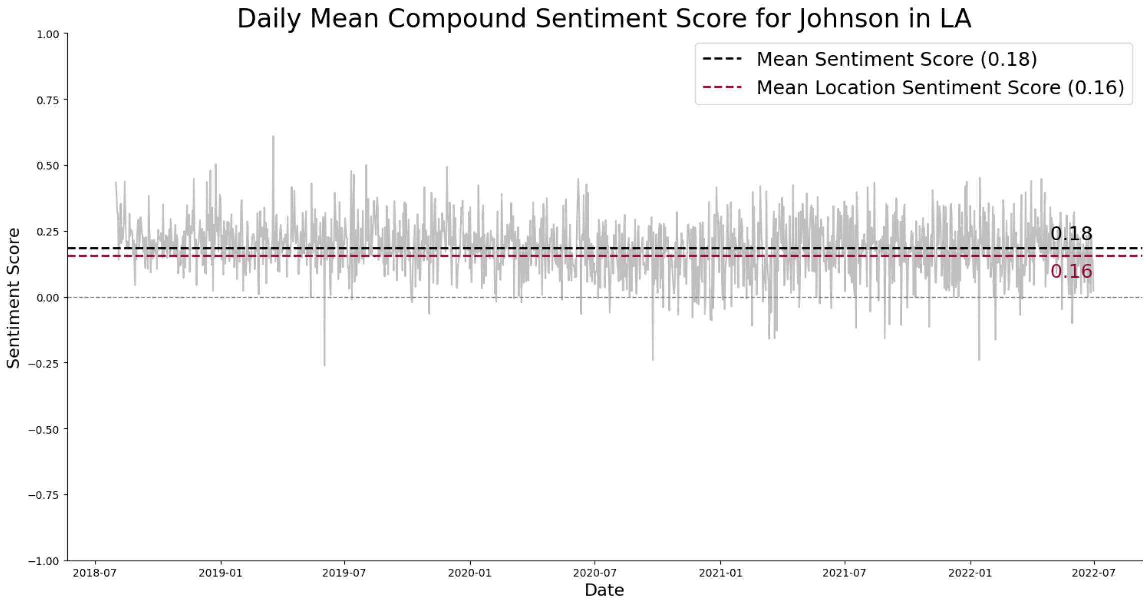


Figure 17 Sentiment Score Development of Johnson in Los Angeles

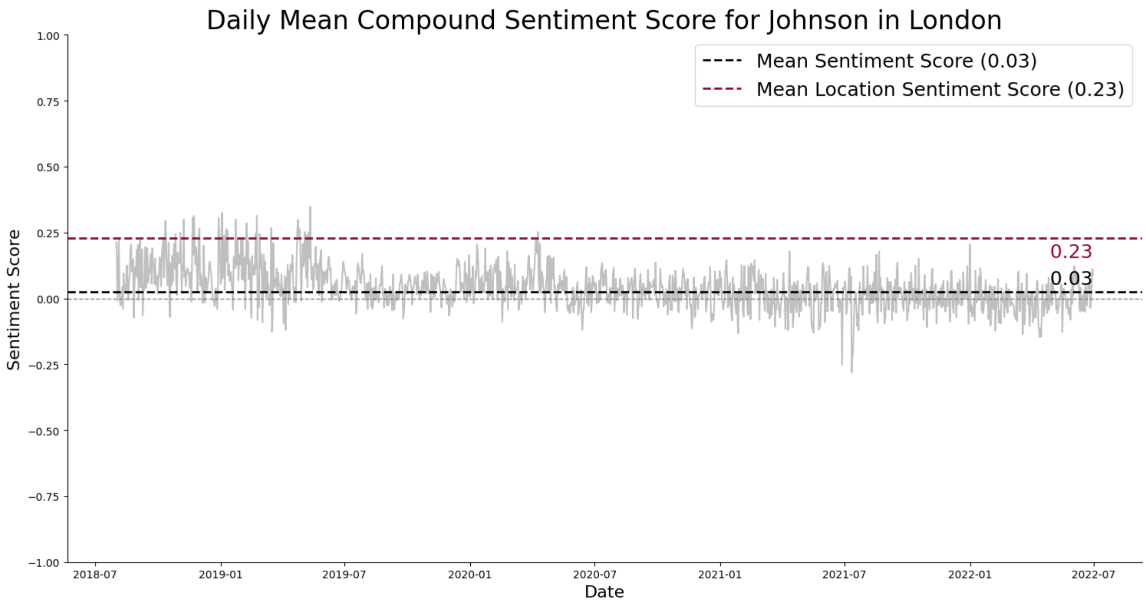


Figure 18 Sentiment Score Development of Johnson in London

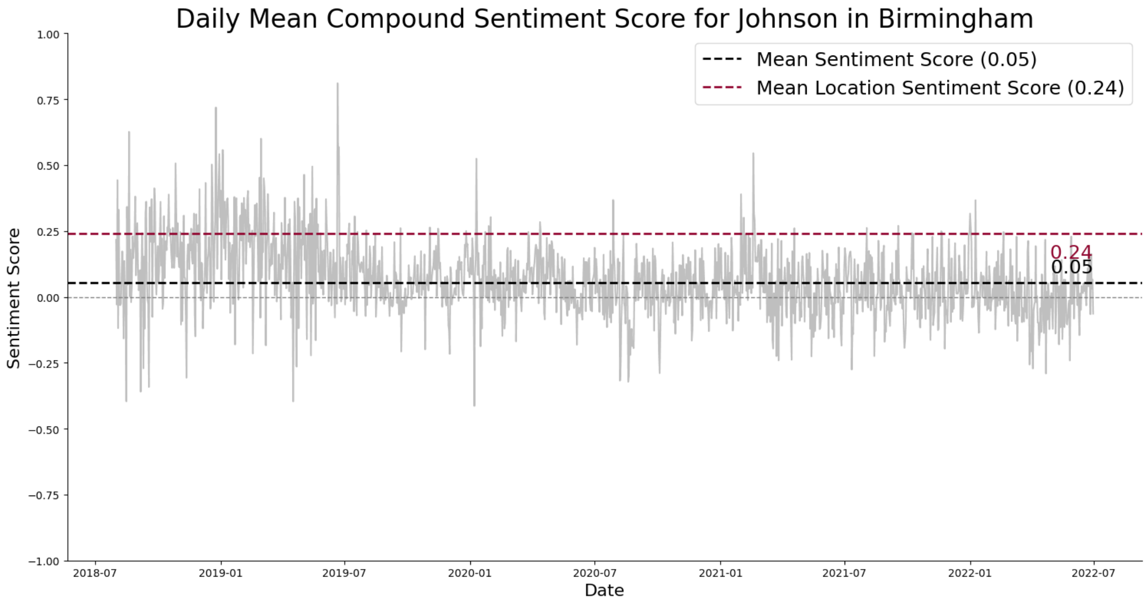


Figure 19 Sentiment Score Development of Johnson in Birmingham

A.4 Post Count

A.4.1 Donald Trump

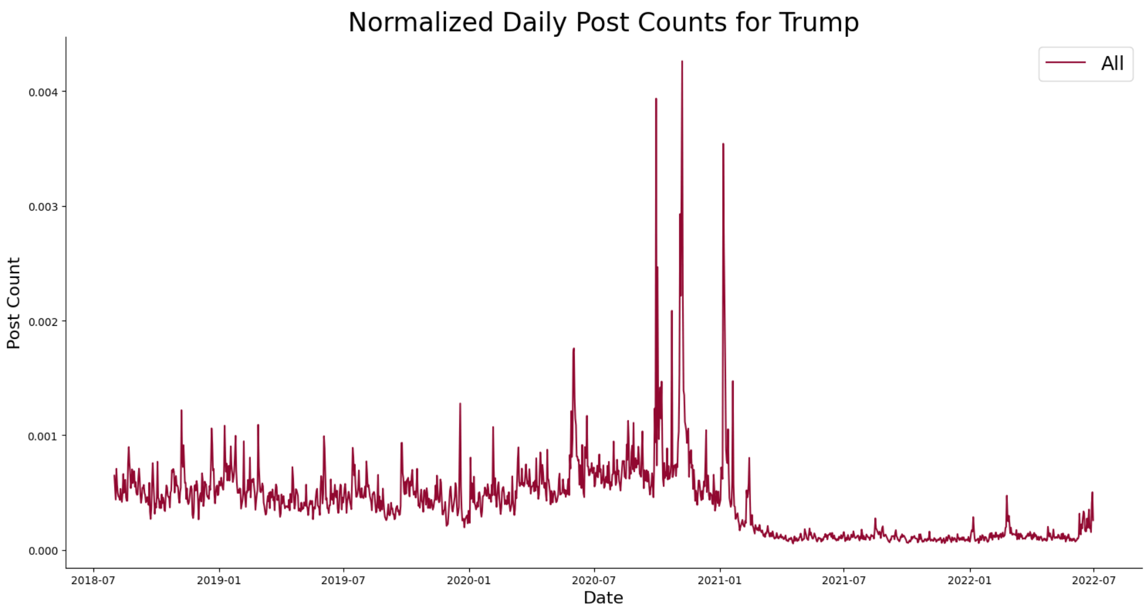


Figure 20 Normalized Post Count About Donald Trump

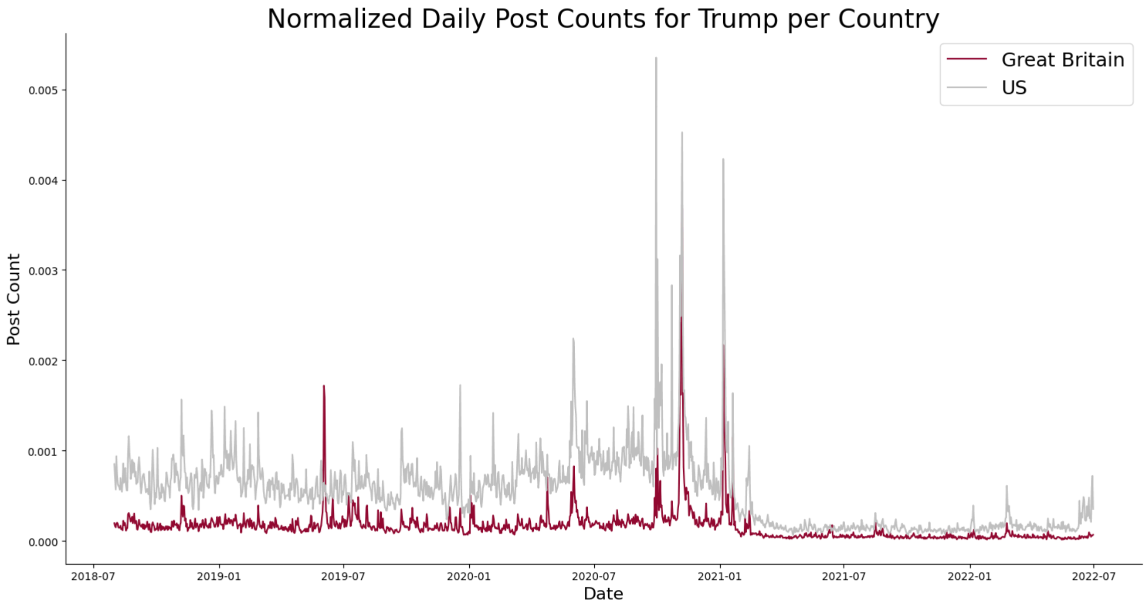


Figure 21 Normalized Post Count About Donald Trump per Country

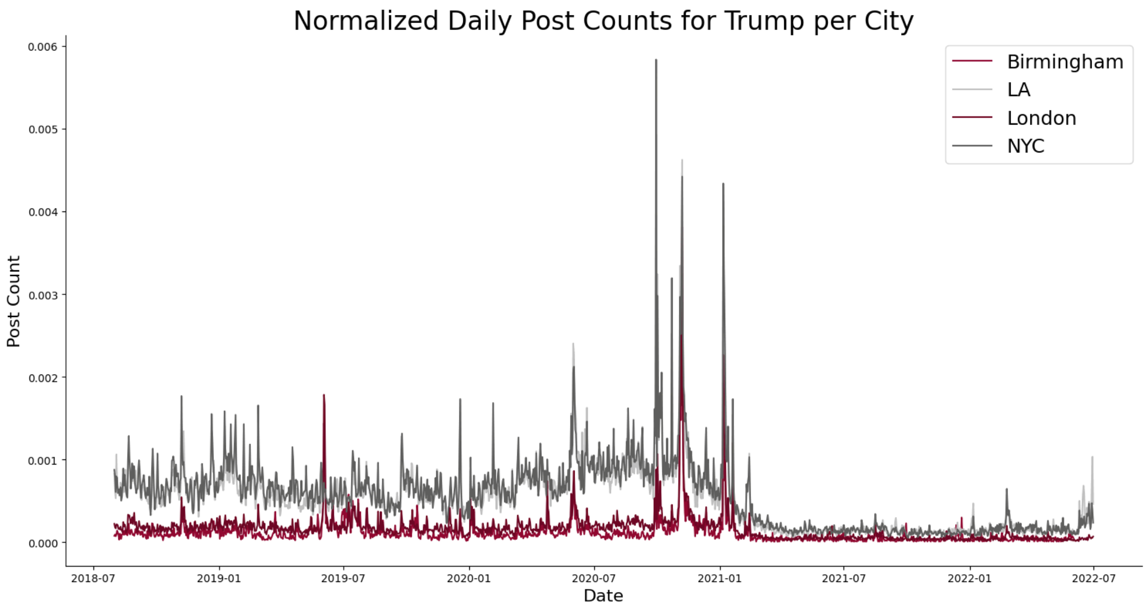


Figure 22 Normalized Post Count About Donald Trump per City

A.4.2 Boris Johnson

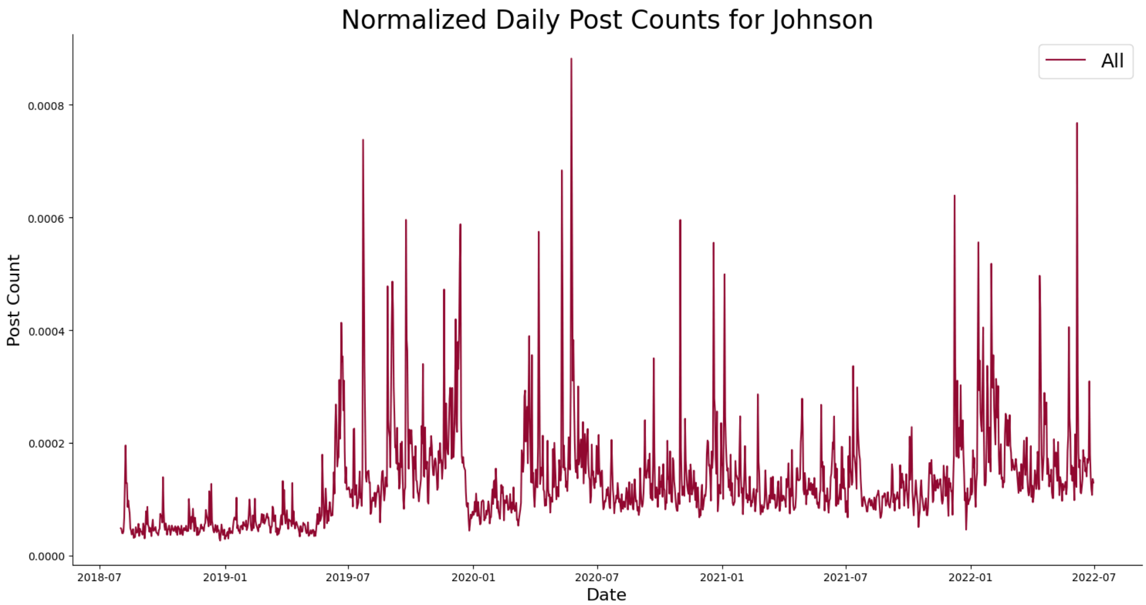


Figure 23 Normalized Post Count About Boris Johnson

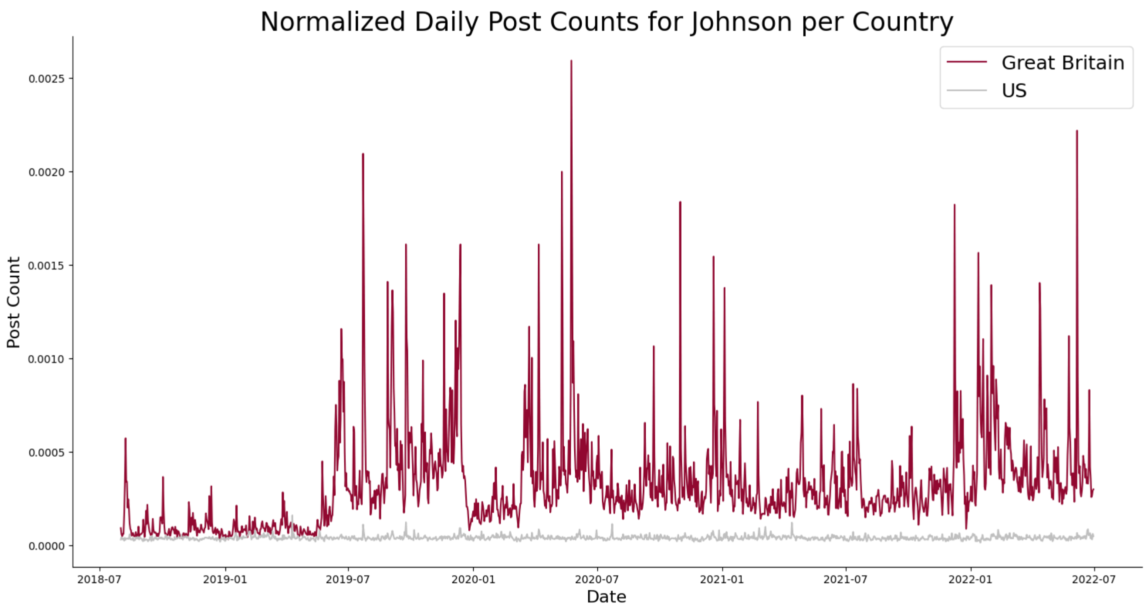


Figure 24 Normalized Post Count About Boris Johnson per Country

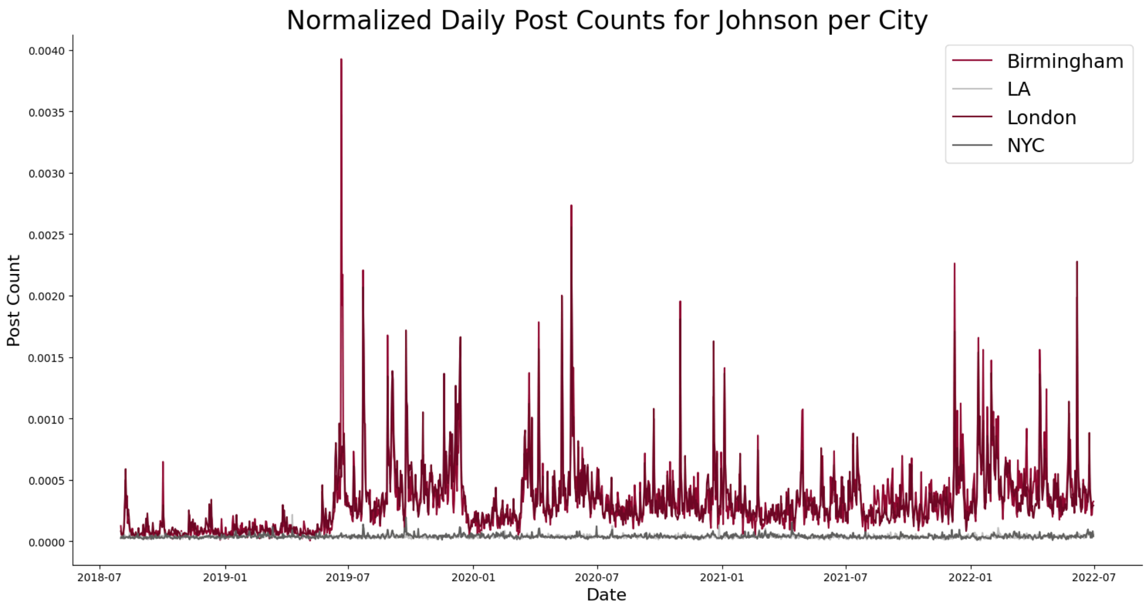


Figure 25 Normalized Post Count About Boris Johnson per City

Bibliography

Anwar, A., Ilyas, H., Yaqub, U., & Zaman, S. Analyzing QAnon on Twitter in Context of US Elections 2020: Analysis of User Messages and Profiles Using VADER and BERT Topic Modeling. In: In *Association for computing machinery international conference*. Association for Computing Machinery, 2021, 82–88.

AP News. (2020). Boris Johnson: A Career Timeline. https://apnews.com/article/boris-johnson-career-timeline-e741ed418c1c36f8430df83de0ed03b7

Barberá, P. (2015a). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, *23*(1), 76–91.

Barberá, P. (2015b). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, *23*(1), 76–91.

BBC News. (2020a). Coronavirus: Boris Johnson Admitted to Hospital Over Virus Symptoms. https://www.bbc.com/news/uk-52192604

BBC News. (2020b). US Election 2020: A Guide to the Final Results. https://www.bbc.com/news/election-us-2020-55016029

Bekmanova, G., Yergesh, B. Z., Omarbekova, A., Mukanova, A. S., & Ongarbayev, Y. (2023). Sentiment Processing of Socio-Political Discourse and Public Speeches. *Lecture Notes in Computer Science*.

Bestvater, S., Shah, S., Rivero, G., & Smith, A. (2021). Political Behavior on Twitter Among U.S. Adults. *Pew Research Center*. https://www.pewresearch.org/politics/2022/06/16/politics-on-twitter-one-third-of-tweets-from-u-s-adults-are-political/

Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing With Python: Analyzing Text With the Natural Language Toolkit. https://www.nltk.org/book/

Blackall, M. (2020). Boris Johnson Announces National Lockdown for England. https://www.theguardian.com/world/2020/oct/31/boris-johnson-announces-national-lockdown-england-coronavirus

Bovet, A., & Makse, H. (2019). Influence of News Media on the US Presidential Election. *Nature Communications*, *10*(1), 1–14.

Bowcott, O., Quinn, B., & Carrell, S. (2019). Boris Johnson's Suspension of Parliament Unlawful, Supreme Court Rules. https://www.theguardian.com/law/2019/sep/24/boris-johnsons-suspension-of-parliament-unlawful-supreme-court-rules-prorogue

Center, P. R. (2023). Little Change in Americans' Views of Trump Over the Past Year. https://www.pewresearch.org/topic/politics-policy/leaders/more-leaders/donald-trump/

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political Polarization on Twitter. *5th International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media*.

Crace, J. (2020). Boris Johnson Admitted to Hospital With Coronavirus. https://www.theguardian.com/politics/2020/apr/05/boris-johnson-admitted-to-hospital-with-coronavirus

Demand Sage. (2024). Twitter Statistics for 2024 - (Facts After "X" Rebranding). https://www.demandsage.com/twitter-statistics/

Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Systems With Applications*, *118*, 272–299. https://www.sciencedirect.com/science/article/pii/S0957417418306456

Dunn, E. W., Weidman, A. C., & Kasser, T. (2017). The Happiness of Cities. *Social Psychological and Personality Science*, *8*(5), 443–451.

EMarketer & Insider Intelligence. (2022). X/Twitter: Number of Users Worldwide 2024. Retrieved February 9, 2024, from https://www.statista.com/statistics/303681/twitter-users-worldwide/

Endsuy, R. D. (2021). Sentiment Analysis Between VADER and EDA for the US Presidential Election 2020 on Twitter Datasets. *Journal of Applied Data Sciences*, *2*(1), 8–18.

Fachrie, M., & Ardiani, F. Predictive Model for Regional Elections Results based on Candidate Profiles. In: In *International conference on electrical engineering, computer science and informatics*. Institute of Electrical; Electronics Engineers Inc., 2021, 247–252.

Florida, R. (2002). The Rise of the Creative Class. *Basic Books*.

Galston, W. A. (2020). *Impeachment and Public Opinion: Three Key Indicators to Watch*. https://www.brookings.edu/blog/fixgov/2019/09/27/impeachment-and-public-opinion-three-key-indicators-to-watch/

Glaeser, E. (2011). *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. Penguin.

Hallemann, C. (2019). Donald Trump to Meet with Queen Elizabeth in State Visit to the UK in June of 2019. Retrieved February 5, 2024, from https://www.townandcountrymag.com/society/politics/a27073402/donald-trump-queen-elizabeth-uk-state-visit-2019/

Hargittai, E. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived From Social Network Sites. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 63–76.

Heredia, B., Prusa, J. D., & Khoshgoftaar, T. M. Location-Based Twitter Sentiment Analysis for Predicting the U.S. 2016 Presidential Election (R. V. & B. K., Eds.). In: *31st international florida artificial intelligence research society conference* (R. V. & B. K., Eds.). Ed. by V., R., & K., B. Association for the Advancement of Artificial Intelligence Press, 2018, 265–270.

Highton, B. (2006). Long Lines, Voting Machine Availability, and Turnout: The Case of Franklin County, Ohio in the 2004 Presidential Election. *Political Science and Politics*, *39*(1), 65–68.

Hutto, C. J. (2014). Commit to vaderSentiment: Added Support for Emoji Recognition (Utf-8 Encoded). https://github.com/cjhutto/vaderSentiment/commit/b16e5f25d3a8d54dd919ad3215441913f206f4c2

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *8th International Conference on Weblogs and Social Media*.

Hutto, C. (2014). VADER Sentiment Analysis. Retrieved January 9, 2024, from https://github.com/cjhutto/vaderSentiment

Iyengar, S., & Kinder, D. R. (2010). *News That Matters: Television and American Opinion*. University of Chicago Press.

Jackson, S. (2019). HashtagActivism: The Promise and Risk of "Attention". *Sociology Compass*, *13*(1).

Keita, Z. (2022). Social Media Sentiment Analysis In Python With VADER. Retrieved February 4, 2024, from https://towardsdatascience.com/social-media-sentiment-analysis-in-python-with-vader-no-training-required-4bc6a21e87b8

Landler, M. (2019). U.K. Supreme Court Rules Suspension of Parliament Unlawful, in Blow to Boris Johnson. https://www.nytimes.com/2019/09/24/world/europe/uk-supreme-court-brexit.html

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan Claypool Publishers.

Malde, R. (2020). A Short Introduction to VADER. Retrieved February 4, 2024, from https://towardsdatascience.com/an-short-introduction-to-vader-3f3860208d53

Malik, M. M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). Population Bias in Geotagged Tweets. *9th International Association for the Advancement of Artificial Intelligence Conference on Web and Social Media*.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press. https://nlp.stanford.edu/fsnlp/

Mason, R. (2020). Boris Johnson Sets Out Plan to Reopen Shops and Schools in England. https://www.theguardian.com/politics/2020/may/10/boris-johnson-coronavirus-lockdown-shops-schools-june-reopening

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *International Conference on Learning Representations.*

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. *Advances in Neural Information Processing Systems*, *26*. https://papers.nips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. N. (2011). Understanding the Demographics of Twitter Users. *5th International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media.*

MonkeyLearn. (2023). Sentiment Analysis Guide. Retrieved December 5, 2023, from https://monkeylearn.com/sentiment-analysis/

Naylor, B. (2021). Read Trump's Jan. 6 Speech, a Key Part of Impeachment Trial. https://www.npr.org/2021/02/10/966396848/read-trumps-jan-6-speech-a-key-part-of-impeachment-trial

NBC News. (2021). Trump Encourages Those at His Rally to March to the Capitol. https://www.youtube.com/watch?v=5fiT6c0MQ58

Norgaard, J. R., & Walbert, H. (2023). Group Polarization?: An Analysis of Parler Data in the Wake of the Capitol Riot. *Journal of Entrepreneurship and Public Policy*, *12*(2), 145–171.

Oikonomou, L., & Tjortjis, C. A Method for Predicting the Winner of the USA Presidential Elections Using Data Extracted From Twitter. In: In *South-east europe design automation, computer engineering, computer networks and social media conference*. Institute of Electrical; Electronics Engineers Inc., 2018.

Osborne, S. (2022). *Boris Johnson Fined: Prime Minister Apologises After Receiving Fixed Penalty Notice for Lockdown-Breaking Party.* https://news.sky.com/story/boris-johnson-fined-prime-minister-apologises-after-receiving-fixed-penalty-notice-for-lockdown-breaking-party-12588712

Pai, A. R., Prince, M., & Prasannakumar, C. V. (2022). Real-Time Twitter Sentiment Analytics and Visualization Using Vader. *2nd International Conference on Intelligent Technologies.*

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic Inquiry and Word Count. *Mahway: Lawrence Erlbaum Associates*, *71*(2001).

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Conference on Empirical Methods in Natural Language Processing*, 1532–1543. https://nlp.stanford.edu/pubs/glove.pdf

Pradipta, E. P., Rahman, T., Sukmono, F. G., & Junaedi, F. (2023). Analysis of Political Polarization Discourse on Social Media Ahead of the 2024 Election. *International Conference on Human-Computer Interaction*, 95–102.

Reuters. (2021). Timeline: How a Trump Mob Stormed the U.S. Capitol, Forcing Washington Into Lockdown. https://www.reuters.com/article/us-usa-election-capitol-events-idUSKBN29B2PU

Rita, P., Antonio, N., & Afonso, A. P. (2023). Social Media Discourse and Voting Decisions Influence: Sentiment Analysis in Tweets During an Electoral Period. *Social Network Analysis and Mining*.

Rodden, J. (2006). Geographic Distribution of Political Preferences. *Annual Review of Political Science*, *9*, 297–340.

Saad, L., & Brenan, M. (2021). Americans' Views of Impeachment, Trump's Record on Issues. *Gallup*. https://news.gallup.com/poll/329996/americans-views-impeachment-trump-record-issues.aspx

School, C. L. (2017). President Trump Announces Withdrawal From Paris Agreement. https://climate.law.columbia.edu/content/president-trump-announces-withdrawal-paris-agreement-0

Selvaraj, N. (2020). A Beginner's Guide to Sentiment Analysis with Python. Retrieved December 5, 2023, from https://towardsdatascience.com/a-beginners-guide-to-sentiment-analysis-in-python-95e354ea84f6

Singh, A., Kumar, A., Dua, N., Mishra, V. K., Singh, D., & Agrawal, A. Predicting Elections Results using Social Media Activity A Case Study: USA Presidential Election 2020. In: In *7th international conference on advanced computing and communication systems*. Institute of Electrical; Electronics Engineers Inc., 2021, 314–319.

Singh, S., Singhania, S., Pandya, V., Singal, A., & Biwalkar, A. (2022). East Meets West: Sentiment Analysis for Election Prediction. *Studies in Computational Intelligence*, *1027*, 9–20.

Singleton, A. D., Spielman, S., & Folch, D. (2017). *Urban Analytics*. Sage.

Solovev, K., & Pröllochs, N. (2022). Hate Speech in the Political Discourse on Social Media: Disparities Across Parties, Gender, and Ethnicity. *Association for Computing Machinery Web Conference*, 3656–3661.

Steve Holland, J. M., & Landay, J. (2021). Trump Summoned Supporters to Wild Protest, and Told Them to Fight. https://www.reuters.com/article/idUSKBN29B2CY/

Stewart, H., & Mason, R. (2019). Boris Johnson Leads Tories to Historic General Election Win. https://www.theguardian.com/politics/2019/dec/13/boris-johnson-leads-tories-historic-general-election-win

Stewart, H., & Murphy, S. (2020). Priti Patel: Bullying Inquiry Report Criticises Home Secretary. *The Guardian*. https://www.theguardian.com/politics/2020/nov/20/priti-patel-boris-johnson-bullying-report-findings

Swarnkar, N. (2020). VADER Sentiment Analysis: A Complete Guide. Retrieved February 4, 2024, from https://blog.quantinsti.com/vader-sentiment/

Thakur, N. (2023). Sentiment Analysis and Text Analysis of the Public Discourse on Twitter About Covid-19 and MPox. *Big Data and Cognitive Computing*.

The New York Times. (2020). Highlights From the First 2020 Presidential Debate. https://www.nytimes.com/live/2020/09/29/us/presidential-debate

Thematic. (2023). Sentiment Analysis - Comprehensive Beginners Guide. Retrieved December 5, 2023, from https://getthematic.com/sentiment-analysis/

Theocharis, Y., Lowe, W., van Deth, J. W., & García-Albacete, G. (2015). Using Twitter to Mobilize Protest Action: Online Mobilization Patterns and Action Repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi Movements. *Information, Communication and Society*, *18*(2), 202–220.

Todi, M. (2019). Sentiment Analysis Using the Vader Library.

UK, E. (2021). Boris to the US! New Poll Shows Departing PM Is Still the Most Popular in America. https://www.express.co.uk/news/politics/1655243/Boris-Johnson-poll-USA-UK-Prime-Minister-popularity

UK Government. (2020a). PM Address to the Nation on Coronavirus: 10 May 2020. https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-10-may-2020

UK Government. (2020b). Prime Minister's Statement on Coronavirus (Covid-19): 31 October 2020. https://www.gov.uk/government/speeches/prime-ministers-statement-on-coronavirus-covid-19-31-october-2020

UK Government. (2021). PM Statement at Coronavirus Press Conference: 12 July 2021. https://www.gov.uk/government/speeches/pm-statement-at-coronavirus-press-conference-12-july-2021

Varol, C., & Tareq Abdulhadi, H. (2018). Comparision of String Matching Algorithms on Spam Email Detection.

White House. (2020). Proclamation on Declaring a National Emergency Concerning the Novel Coronavirus Disease (Covid-19) Outbreak. https://trumpwhitehouse.archives.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novel-coronavirus-disease-covid-19-outbreak/

Yaqub, U., Sharma, N., Pabreja, R., Chun, S. A., Atluri, V., & Vaidya, J. (2020). Location-Based Sentiment Analyses and Visualization of Twitter Election Data. *Digital Government: Research and Practice*, *1*(2).

Declaration of Authorship

I hereby declare that, to the best of my knowledge and belief, this thesis titled *Popularity and Controversy: A Location-Based Event and Sentiment Analysis of Donald Trump and Boris Johnson* is my own, independent work. I confirm that each significant contribution to and quotation in this thesis that originates from the work or works of others is indicated by proper use of citation and references; this also holds for tables and graphical works.

Münster, 18.02.2024

Leo Richard Hermann Giesen

Consent Form

Name: Leo Richard Hermann Giesen
Title of Thesis: Popularity and Controversy: A Location-Based Event and Sentiment Analysis of Donald Trump and Boris Johnson

What is plagiarism? Plagiarism is defined as submitting someone else's work or ideas as your own without a complete indication of the source. It is hereby irrelevant whether the work of others is copied word by word without acknowledgment of the source, text structures (e.g. line of argumentation or outline) are borrowed or texts are translated from a foreign language.

Use of plagiarism detection software. The examination office uses plagiarism software to check each submitted bachelor and master thesis for plagiarism. For that purpose the thesis is electronically forwarded to a software service provider where the software checks for potential matches between the submitted work and work from other sources. For future comparisons with other theses, your thesis will be permanently stored in a database. Only the School of Business and Economics of the University of Münster is allowed to access your stored thesis. The student agrees that his or her thesis may be stored and reproduced only for the purpose of plagiarism assessment. The first examiner of the thesis will be advised on the outcome of the plagiarism assessment.

Sanctions Each case of plagiarism constitutes an attempt to deceive in terms of the examination regulations and will lead to the thesis being graded as "failed". This will be communicated to the examination office where your case will be documented. In the event of a serious case of deception the examinee can be generally excluded from any further examination. This can lead to the exmatriculation of the student. Even after completion of the examination procedure and graduation from university, plagiarism can result in a withdrawal of the awarded academic degree.

I confirm that I have read and understood the information in this document. I agree to the outlined procedure for plagiarism assessment and potential sanctioning.

Münster, 18.02.2024

Leo Richard Hermann Giesen