Nottingham University Business School (NUBS)
The University of Nottingham
Jubilee Campus
Nottingham, NG8 1BB
United Kingdom

Individual Report

Module Title: Analytics Specializations and Applications

Module Code: BUSI4370

Assignment Title: Brand Analysis Using Twitter

Word Count: 2972

Student Number: 20680080

Module Convenor: Prof James Goulding

1. Executive Summary

This report analyzes Adidas and Lululemon on Twitter using social network and text analysis. The goal was to understand how users interact with each brand, identify micro-influencers, and uncover key themes and public sentiment.

A directed graph approach was used for the network analysis, with centrality metrics (in-degree, betweenness, closeness) to detect influential users. A custom Influencer Score helped highlight strong micro-influencers with authentic engagement. Text cleaning and lemmatization enabled effective topic modeling using LDA, while sentiment analysis with VADER provided insight into public attitudes.

Results show Adidas is associated with sports, product launches, and digital trends, while Lululemon focuses on wellness, sustainability, and community. Both brands received mostly positive sentiment, with Adidas showing more emotional variation.

Candace Parker and Nicholas Ferroni were recommended as micro-influencers due to their alignment with brand values and strong positions in the network.

This analysis supports better influencer selection and content strategies based on user behavior and brand perception on social media.

2. Approach Breakdown

The social network analysis was initiated by constructing directed graphs where users represented nodes and interactions, mentions extracted via regex, from tweet content were modeled as edges. Outliers were removed by filtering users with exceptionally high follower counts (3 std above mean), and brand owned accounts were excluded to focus on organic interactions.

Structural metrics were calculated on a subgraph of the top 2000–4000 most mentioned users, depending on the brand, to ensure computational efficiency and focus on relevant nodes. In-degree, betweenness, and closeness centrality were computed using NetworkX. These metrics were standardized and combined into a custom Influencer Score, weighted to emphasize betweenness and closeness. This allowed for the identification of key micro-influencers within each brand's interaction network based on structural positioning rather than raw activity levels.

Following the social network analysis, a comprehensive text analysis was conducted to uncover key themes associated with each brand. The process began with a tailored cleaning pipeline to eliminate noise elements such as mentions, emojis, URLs, retweet tags, numbers, and low-information words. Custom stopwords were added to better reflect the informal language used on social media.

To capture multi-word expressions, bigrams were constructed using Gensim's Phrases model, ensuring the representation of frequent collocations. Texts were then lemmatized using spaCy, retaining only nouns, verbs, adjectives, and adverbs to reduce noise while preserving meaningful linguistic structures.

The cleaned and lemmatized tokens were transformed into Bag-of-Words format, forming the input for topic modeling. Latent Dirichlet Allocation (LDA) was applied to identify recurring themes within each brand's tweet set. To determine the optimal number of topics, coherence scores (c_v) were computed over multiple models with varying topic counts, and the model with the highest score was selected for analysis.

Building on the identified topics, a sentiment analysis was conducted to better understand public attitudes towards the brands and their core themes. The sentiment pipeline used the VADER algorithm from NLTK, selected for its suitability in analyzing informal and short texts like tweets without requiring labeled data.

A separate preprocessing routine was applied to the sentiment input, focusing on reducing noise while retaining emotional content. Compared to the cleaning used for topic modeling, this process additionally removed hashtags and adjusted punctuation handling to better align with VADER's lexicon.

The sentiment polarity of each tweet was calculated and categorized as **positive**, **neutral**, or **negative** using VADER's compound score thresholds. Sentiment distributions were then visualized at both the general brand level and by dominant topic, allowing for comparison of emotional tone across different areas of discussion.

3. Data Description

For this analysis, the **Adidas** and **Lululemon** datasets were selected. Both contain the same 15 features but differ notably in size: **Adidas** includes **38,812** tweets, while **Lululemon** has **6,190**. This size difference may affect the comparability and statistical reliability of the results.

All tweets mention the brand names, suggesting that the datasets were collected using brand-specific keywords or hashtags. Due to a high number of null or inconsistent entries in the **user.location** field, this feature was excluded. Similarly, features such as **user.id**, **user.name**, **id** and **is_quote_status** were removed for simplification, as they provided no relevant insights.

The tweets cover a period of 4 months. The larger volume of Adidas tweets suggests greater online activity or engagement compared to Lululemon during that time.

Figure 1

```
Adidas Tweet Date Range: ('2021-10-01', '2022-01-01')
Lululemon Tweet Date Range: ('2021-10-01', '2022-01-01')
```

Additionally, the percentage of unique users is relatively high in both datasets, suggesting that most tweets come from different individuals, which adds diversity to the data as shown in Figure 2.

Figure 2

```
Percentage of unique users in Adidas dataset: 76.3
Percentage of unique users in Lululemon dataset: 72.1
```

Moreover, the percentage of verified users is low in both datasets, with Lululemon having a slightly higher proportion (**4.2%**) compared to Adidas (**1.8%**). This may suggest a greater presence of public figures, influencers, or official accounts engaging with Lululemon content.

4. Analysis Section

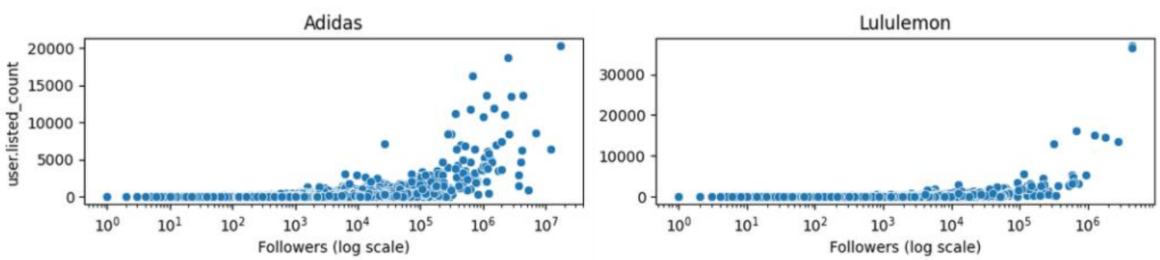
4.1. Social Network Analysis

4.1.1. Exploratory Analysis

The analysis started by examining key metrics to better understand different types of influencers. In particular, the number of **followers**, **listings**, and **posts** were used to identify users who might be suitable for brand promotion. Due to the wide range of follower counts, a logarithmic scale was applied, allowing clearer insights when compared with listed and post counts.

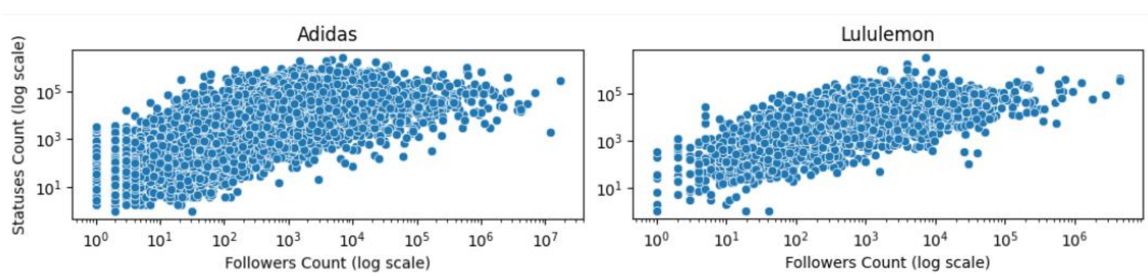
The scatter plot (Figure 3) shows that being listed does not grow proportionally with the number of followers. Instead, listed counts tend to increase only after certain thresholds (1k or 10k), suggesting that relevance and expertise play a key role. Some users with many followers are rarely listed, which may indicate a lack of clear niche or consistent thematic content.

Figure 3



Using a log-log scale (Figure 4) reveals that both Adidas and Lululemon show a positive trend between followers and post activity, but it's not linear. For Adidas, there's a wider spread, especially at higher follower counts, suggesting more variability in influencer behavior. In contrast, Lululemon's pattern is tighter, possibly indicating more consistent engagement strategies among their influencers. Interestingly, the most active users for both brands seem to cluster around the mid-follower range (1k-10k). This might indicate that mid-tier influencers are more involved and valuable.

Figure 4



4.1.2. In-Degree Analysis

To deepen the analysis, in-degree centrality was used to identify the most frequently mentioned users. Prior to this, outliers with follower counts more than three standard deviations above the mean were removed to exclude celebrity accounts and focus on more typical users. As noted earlier, mid-tier influencers seem more engaged, which is further supported by the presence of many such users among the most mentioned accounts.

For **Adidas**, top mentions include large corporate and entertainment accounts such as **Xbox, Nike, EA Sports FIFA, Hyperkin**, and **Adidas** itself. Celebrities, athletes, and tech influencers are also prominent, suggesting that Adidas-related conversations are shaped by high-profile figures in sports, gaming, and digital culture.

In contrast, **Lululemon's** most mentioned accounts include social campaigns, organizations like **Team Canada** and **CDNParalympics**, and public figures such as **Calvin McDonald**. Mentions of wellness brands like **OnePeloton** indicate that Lululemon's Twitter discourse is more centered around wellness, social values, and national identity, pointing to a distinct brand narrative and influencer ecosystem (Figure 5).

Figure 5

Adidas Top 10 In-Degrees

	User	In-degree
1181	Xbox	7428
3134	AustinEkeler	6938
5	Nike	1817
10792	Hyperkin	986
19601	BoredApeYC	624
20606	richsignorelli	416
1434	DashieXP	344
3355	Michael_Fabiano	339
1033	spidadmitchell	289
9974	EASPORTSFIFA	285

Lululemon Top 10 In-Degrees

	User	In-degree
114	TeamCanada	227
94	standearth	89
93	calvinmcdonald	78
230	Nike	76
1509	hockeynight	69
1654	DeezeFi	64
210	onepeloton	52
115	CDNParalympics	50
1062	Devin_Heroux	48
1746	orthokimchipac	45

4.1.3. Structural Analysis

To move beyond a basic analysis based only on mention frequency or follower count, additional network centrality metrics were applied: in-degree, betweenness, and closeness. Metrics like closeness and betweenness centrality are more useful in identifying influential individuals than indegree, which simply measures how many followers someone has" (Libert, 2016). These measures help to identify users who are not only visible but also structurally important within the network. Based on this approach, several types of users were identified for both Lululemon and Adidas: Corporate accounts, brand-affiliated employees, independent content creators, community-driven curators and lifestyle micro-influencers.

Figure 6

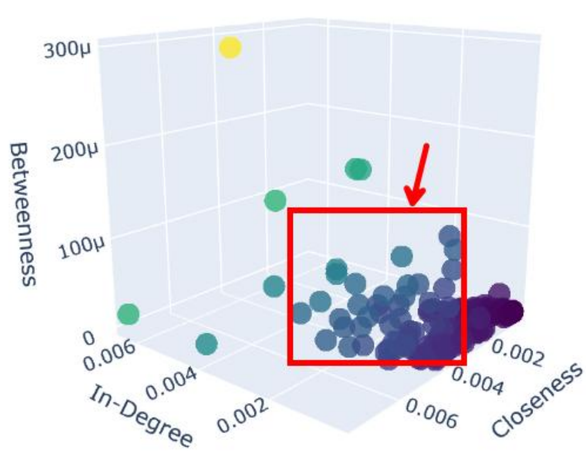
	In-Degree	Betweenness	Closeness	Influencer Score	listed	followers	friends	publications
turtlepace5	0.004751	0.000299	0.005639	36.626231	5.0	2794.000000	4861.000000	57172.000000
iAmTMCII	0.003501	0.000151	0.005674	22.437955	70.0	22518.000000	68.000000	96725.000000
snkr_twitr	0.006502	0.000015	0.007644	22.001508	1491.0	415324.500000	308.000000	216876.000000
trudyspeaks	0.002001	0.000187	0.004723	20.335622	14.0	2019.000000	2870.000000	44703.526316
Moonman989	0.001750	0.000190	0.004923	20.232388	0.0	1114.411765	1176.411765	22797.000000
Hyperkin	0.004251	0.000007	0.007330	15.654606	142.0	15489.000000	1150.000000	5684.000000
aarongreenberg	0.004251	0.000039	0.004822	14.571737	1338.0	243248.800000	2177.000000	28796.000000
sftwofive	0.002251	0.000085	0.005085	13.640078	3.0	738.277778	1885.777778	37003.000000
beiberlove69	0.002251	0.000080	0.005085	13.267852	2.0	453.000000	1017.000000	6485.000000
Reebok	0.002751	0.000043	0.005005	11.525444	2379.0	704554.000000	5125.000000	37767.000000
	In-Degree	Betweenness	Closeness	Influencer Score	listed	followers	friends	publications
DeezeFi	0.004502	2.804205e-05	0.004502	27.625546	2087.0	137697.857143	3386.000000	41102.428571
getthemirror	0.005503	4.506758e-06	0.006483	20.189561	53.0	7679.000000	58.000000	1840.000000
AFineBlogger	0.002001	2.478717e-05	0.002001	15.973546	114.0	4720.000000	4591.846154	50707.000000
jeffieruth	0.004502	5.007509e-07	0.004002	10.826206	72.0	1148.000000	1199.000000	67583.000000
BillGarlandSpkr	0.004002	1.084960e-06	0.004002	10.268246	280.0	11863.000000	11882.000000	24995.000000
Paul_Kandavalli	0.004002	1.084960e-06	0.004002	10.268246	498.0	7663.000000	6180.000000	20075.000000
coleenlou1	0.004002	1.084960e-06	0.004002	10.268246	61.0	820.000000	648.000000	2792.000000
MrSneet	0.003502	2.503754e-06	0.003557	9.238705	5.0	1359.000000	2000.000000	11999.000000
thejenweg	0.003502	5.007509e-07	0.003557	8.094525	244.0	2180.000000	3073.000000	8961.000000
JohnKnopfPhotos	0.001501	6.259386e-06	0.002701	5.991806	695.0	75360.600000	10360.000000	12034.000000

In addition to individual users, several types of companies were identified as active participants (Figure 6). These include retailers and resale platforms such as **snkr_twitr**, which highlight product releases, and consumer electronics brands like **Hyperkin**, supporting Adidas-related content through gaming and tech promotion. **Sportswear brands** such as **Reebok** also appear. Moreover, corporate representatives like **aarongreenberg** from **Xbox** contribute by sharing branded content and announcements. These organizations help shape brand narratives through direct marketing and strategic partnerships.

While metrics like the **Influencer Score** are useful to detect users with strong positions in the network, it is also important to balance this with **follower count**, **list inclusion, friends** and **likes**. To support this, a visual analysis was carried out combining **in-degree**, **betweenness**, and **closeness**. The influencer score was represented through color, where yellow indicates high values and purple low ones.

In Figure 7, sweet spots were identified for both brands in areas with intermediate Influencer Score values (blue). Users in these zones may not hold the highest centrality, but they still occupy strong positions in the network and show notable follower and list metrics. Being added to public lists reflects how others perceive them as relevant sources within specific communities or topics.

Figure 7



The selected influencer should have strong reach within their community, communicate with a personal and authentic voice, and show consistent engagement. They should not be affiliated with large brands, maintaining a sense of independence aligning with brand's value and identity.

4.2. Text Analysis

4.2.1. Key Descriptive Words

To begin the text analysis, all hashtags were extracted and filtered using a frequency threshold, only those mentioned at least 50 times for Adidas and 10 times for Lululemon were included, ensuring the focus remained on the most prominent and resonant terms in each brand's conversation.

The conversation around Adidas highlights product promotions, branded campaigns, and emerging tech trends, as illustrated in Figure 8. Hashtags indicate engagement with both athletic culture and digital innovation. The focus remains on performance, and community engagement through modern platforms.

Figure 8

For Lululemon, the discussion seems focused on **wellness, sustainability, and fitness culture**. Mentions reflect a focus on mindful living and eco-conscious values (Figure 9). The brand appears to be part of a lifestyle movement, connecting with users through community-based initiatives and holistic health themes.

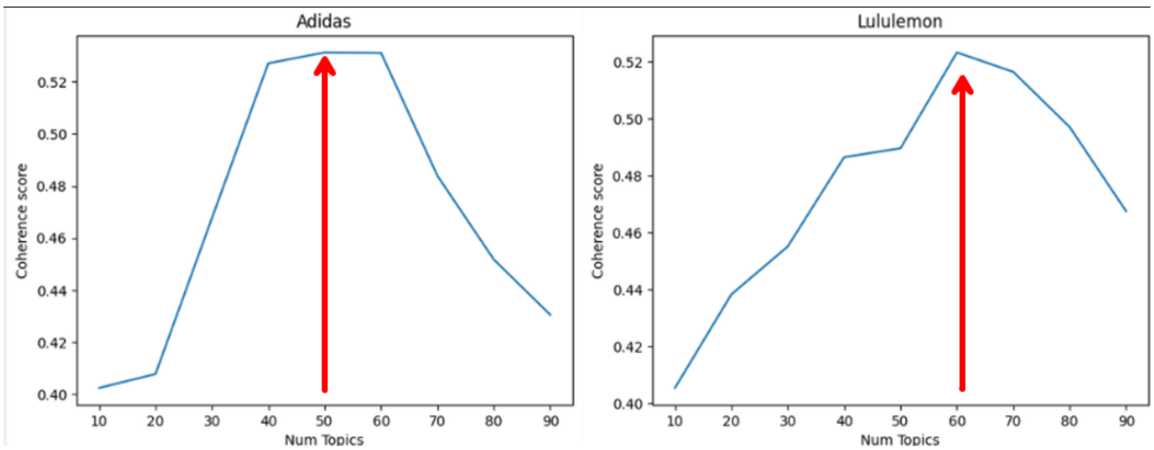
Figure 9

This analysis aligns with the profile of the most mentioned influencers for each brand. Adidas operates in an environment where the narrative centers on performance, pop culture, and commercial momentum. In contrast, Lululemon inhabits a more intimate space, where principles of wellness, sustainability, and community take precedence.

4.2.2. Topic Analysis

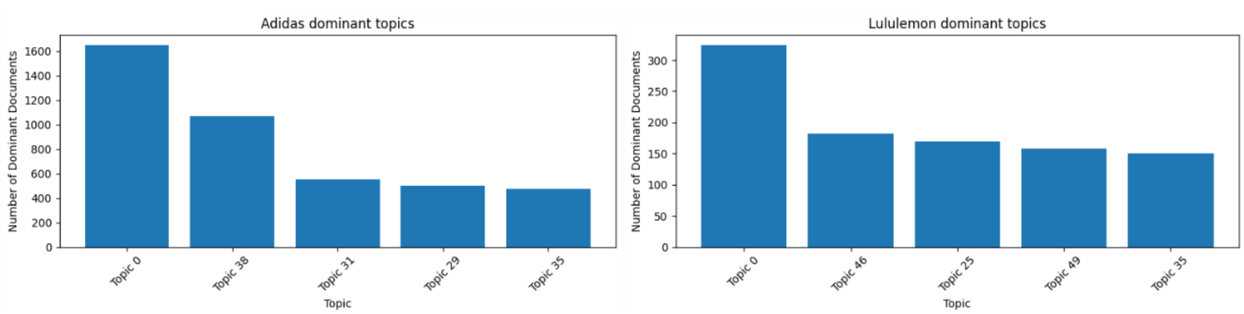
The tweets were processed and prepared as explained in the methodology. Then, LDA models were trained to identify the main topics discussed around each brand. To find the best number of topics, multiple models were tested and compared using the *c_v* coherence score, which measures how well the top words in each topic relate to each other. Based on the results, 50 topics were selected for Adidas and 60 for Lululemon. These models helped uncover the most common themes in user discussions.

Figure 10



Next, the five most frequently mentioned topics were identified for each brand. The bar charts in Figure 11 illustrate these topics, based on the number of documents (tweets) in which each topic was most prevalent. These topics represent the most frequently occurring themes surrounding Adidas and Lululemon on Twitter. For each brand, Topic 0 appears as the most dominant, suggesting a strong central theme in the conversation.

Figure 11



The following section provides a brief description of each of these key topics, based on their most representative keywords.

ADIDAS TOPICS

Topic 0 – Player Development and Classic Moments

Centers around athletes, highlighting player growth, missed opportunities, and classic performances. Carries a nostalgic tone with mentions of goals, youth, and emotional moments in sports.

Topic 38 – Adidas Brand Identity and Promotions

Focuses on brand recognition, prominently mentioning Adidas and associated giveaways. Includes references to running shoes and brand positioning in the athletic space.

Topic 31 – Shoe Launches and Consumer Interest

Captures market excitement around new shoe releases, buyer intent, and product announcements. Reflects strategic messaging and consumer focus in footwear marketing.

Topic 29 – Thank You Messages and Delivery Issues

Combines expressions of gratitude with mentions of sneakers and service problems, including refund requests and delivery delays. Touches on logistics challenges and customer recovery efforts.

Topic 35 – Style Praise and Product Satisfaction

Highlights positive sentiments toward product style, quality, and comfort. Emphasizes love for items like pants and compliments on aesthetics, with minor notes on fit and updates.

LULULEMON TOPICS

Topic 0 – Sustainability and Brand Responsibility Highlights concerns around environmental impact, including coal-powered production and factory emissions. Includes calls to action and brand messaging around taking responsibility and aligning practices with values.

Topic 46 – In-Store Experience and Product Exchanges Captures shopping experiences related to leggings, store visits, sizing, and returns.

Topic 25 – Customer Appreciation and Positive Shopping Reflects gratitude and satisfaction from customers, often tied to sales or yoga-related purchases. Carries a warm, thankful tone emphasizing brand loyalty and enjoyable spending.

Topic 49 – Shipping Delays and Customer Service Issues Focuses on online shopping frustrations, including long wait times, missing items, and poor communication via chat. Highlights pain points in post-purchase service and logistics.

Topic 35 – Healthcare and Hero Discounts Acknowledges exclusive offers for nurses, healthcare workers, and military personnel.

Adidas mainly attracts people who are passionate about sports, sneakers, and the brand itself. These users are interested in athletic performance, new product releases, and have a strong emotional connection to the brand, especially through nostalgic sports moments. At the same time, they expect high product quality and reliable service.

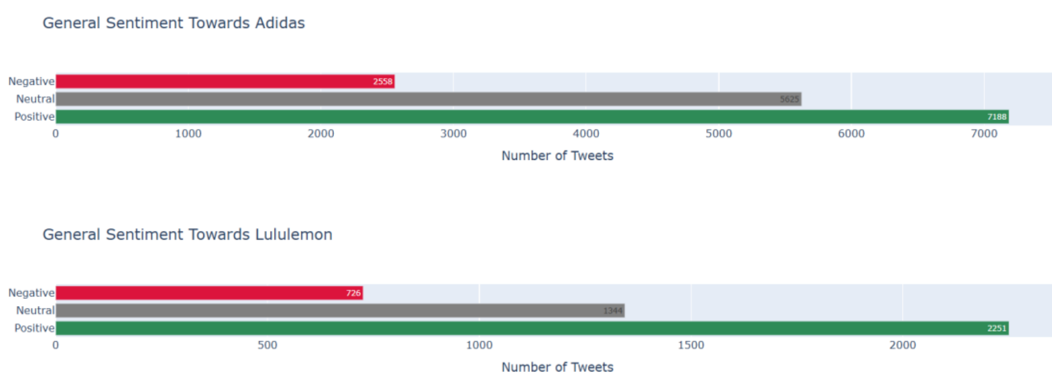
On the other hand, Lululemon appeals to a lifestyle-oriented audience that cares about wellness, sustainability, and ethical choices. Conversations focus on eco-friendly values, in-store shopping, and community support programs like healthcare discounts. This shows a customer base that looks for both good products and strong brand values.

4.3. Sentiment Analysis

Sentiment analysis was performed using VADER, applied to tweets that were preprocessed using a simplified cleaning approach (as described in the methodology). This process focused on removing noise, such as retweets, mentions, and excessive punctuation, while preserving the structure needed for effective sentiment scoring.

As shown in Figure 12, both Adidas and Lululemon were generally perceived positively by users. Positive sentiment was the most common category for both brands, followed by neutral, while negative sentiment appeared least frequently. These results suggest an overall favorable attitude toward each brand within the sample analyzed.

Figure 12



While general sentiment analysis provides an initial overview of public opinion, it is important to move beyond overall polarity and explore the specific emotions associated with each brand's dominant topics. This allows for a more nuanced understanding of how users feel about particular aspects of the brand, helping to identify which themes evoke positive, negative, or mixed emotional responses.

Figure 13

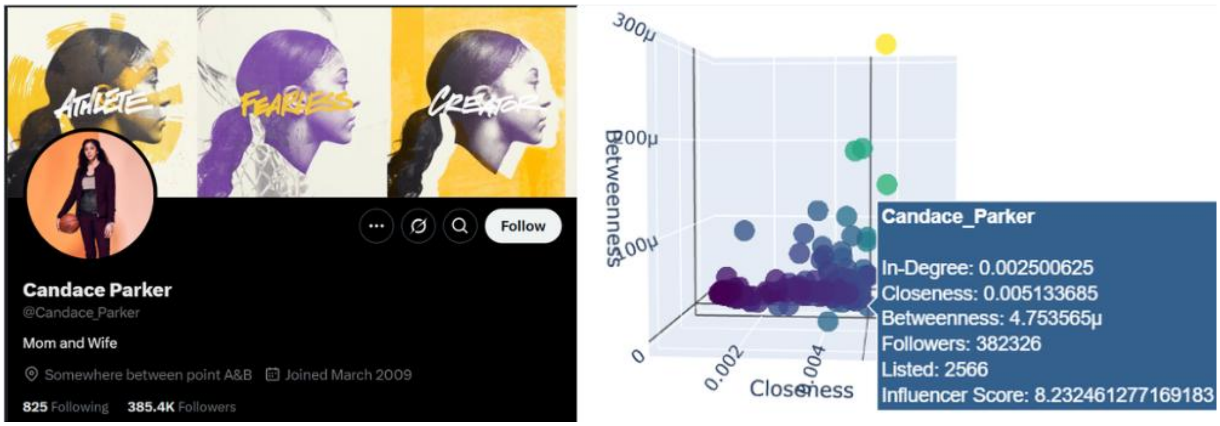
Topic-level sentiment analysis reveals for Adidas notable polarization, particularly around "player development and classic moments" and "thank you messages and delivery issues", which may point to controversial reactions to brand content and dissatisfaction with the delivery system. While for Lululemon, the sentiment analysis of dominant topics offers deeper insight into brand perception, showing a highly positive response to the company's responsible initiatives and a generally favorable view of customer service interactions. This can be observed in Figure 13.

5. Micro-Influencer Recommendation

Two micro-influencers were selected based on network centrality metrics, as well as follower count and list inclusions. Retweets and favorites were excluded from the selection criteria to avoid bias against influential users with low activity in the sample.

For Adidas, Candace Parker is recommended as a micro-influencer. As a former athlete with strong community ties and consistently positive engagement, she aligns well with Adidas' sports-oriented audience. Despite a relatively low betweenness score, indicating limited bridging between user groups, her high **in-degree** shows she receives frequent mentions, and her strong **closeness centrality** suggests she can efficiently reach others within the network. Combined, these factors result in a high influencer score. Additionally, her inclusion in numerous public Twitter lists indicates that her content is valued and widely monitored, further reinforcing her potential impact for the brand.

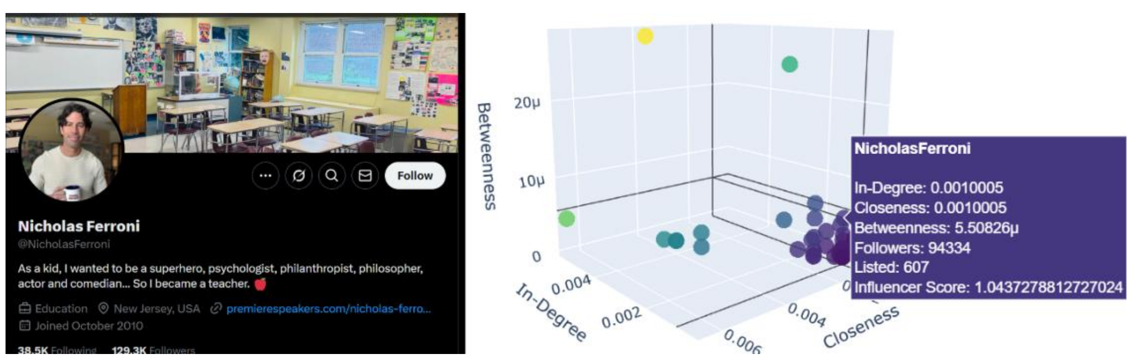
Figure 14



For Lululemon, Nicholas Ferroni is recommended as a potential micro-influencer. With 94k followers, he promotes themes such as community, values, and well-being—closely aligned with the key topics found in Lululemon's network. While his **in-degree** and **closeness centrality** are relatively low, indicating fewer direct mentions and less proximity to others in the network, his **betweenness centrality** is high. This suggests that he acts as a bridge between different groups, making him effective at spreading content across otherwise disconnected audiences.

He is actively engaged on social media, with tweets that consistently receive retweets and likes, showing strong audience interaction. His presence on many public Twitter lists further indicates that users recognize his relevance and monitor his content. Altogether, these factors support his suitability for future brand engagement.

Figure 15



6. Conclusions and Recommendations

Conclusions

This exploratory analysis of Adidas and Lululemon highlights two contrasting but complementary brand identities, Adidas being dynamic, innovation-driven, and associated with sports and digital campaigns, while Lululemon emphasizes wellness, sustainability, and community values. By combining Adidas's energetic promotional tactics with Lululemon's ethical and lifestyle-centered approach, the company can craft a brand that is both engaging and trustworthy. The selection of micro-influencers also shows that authentic voices with strong community ties (like Candace Parker for Adidas and Nicholas Ferroni for Lululemon) are key for building meaningful connections.

Recommendations

The company should implement a temporal sentiment analysis to identify when shifts in user opinion occur and what events, such as product launches, social posts, or service issues, are driving them. This will help adjust communication strategies in real time and anticipate potential risks or opportunities.

Working with mid-tier micro-influencers who are deeply connected to specific communities will help build trust and authenticity. These influencers don't need massive reach but should align with the brand's values and show consistent engagement with their audience.

For sentiment analysis, the use of RoBERTa-based models could offer more context-aware interpretations. These models can recognize subtle emotional tones and language complexity, which would enhance the overall accuracy and depth of the sentiment insights drawn from Twitter data.

References

Libert, K. (2016). *Your Network's Structure Matters More than Its Size*. Harvard Business Review.

Zankadi, H. (2023) *Topic Modeling with Latent Dirichlet Allocation (LDA) using Gensim and NLP techniques (Part II)*. [online] *Medium*. Available at: https://medium.com/@hajar.zankadi/using-latent-dirichlet-allocation-lda-and-nlp-techniques-to-predict-interest-tags-from-tweets-d0e275b1032d

Kashyap, P. (2024) *Topic Modeling with Latent Dirichlet Allocation (LDA)*. [online] *Medium*. Available at: https://medium.com/@piyushkashyap045/topic-modeling-with-latent-dirichlet-allocation-lda-d2ab3fcfba68

Akladyous, B. (2021) *Sentiment Analysis using VADER*. [online] *Medium*. Available at: https://akladyous.medium.com/sentiment-analysis-using-vader-c56bcffe6f24

Abdullayev, A. (2022) *Sentiment Analysis with VADER and Twitter-roBERTa: Benchmarking of two different algorithms for short social media text analysis*. [online] *Medium*. Available at: https://medium.com/@amanabdulla296/sentiment-analysis-with-vader-and-twitter-roberta-2ede7fb78909