Management

FORM FOR THE SUBMISSION OF ASSESSED COURSEWORK

THIS FORM MUST BE COMPLETED AND ATTACHED TO **ALL** ASSESSED WORK.

Student Number (7 digit number)	2	6	3	8	7	9	5		
Username (2 letters & 5 digit number ab12345)	e	z	2	4	5	7	9		
Unit Code	E	F	I	M	M	0	1	3	9
Unit Title	Social Media and Web Analytics								

By submitting this work online using my unique log-in and password, I hereby confirm that this work is entirely my own. I understand that all marks are provisional until ratified by the Faculty Examination Board.

Please remember to save **your work** either as a **Word document (.doc or** as a **PDF (.pdf)** unless another format is specified by the Unit Director with the filename [Unit Code_Student Number] **e.g. MGRCM0000_1234567.docx**.

Please start writing your coursework on the next page.

FITNESS, FEEDBACK, AND THE FUTURE:
UNDERSTANDING PUREGYM USERS THROUGH SOCIAL MEDIA ANALYTICS

by 2638795

EFIMM0139 - Social Media and Web Analytics
Dr. Wen Zhang and Dr. Tian Han
University of Bristol
Bristol, United Kingdom
April 1, 2025

Table of Contents

1. Introduction

In recent years, the fitness industry has grown substantially due to increased consumer health consciousness and a heightened emphasis on wellness. However, the COVID-19 lockdown and the subsequent rapid rise of digital fitness solutions have posed significant challenges to the traditional fitness industry (Sevilmiş & Şirin, 2022; Li et al., 2023). Within this competitive context, fitness chain brands face continual pressure not only to maintain operational excellence but also to improve customer satisfaction and ensure long-term brand loyalty.

Online customer reviews, particularly those shared spontaneously through social media platforms, have become an important source for understanding consumer experiences. Over 90% of consumers report that online reviews significantly influence their purchasing decisions, placing trust in these reviews comparable to personal recommendations from friends or family (Igniyte, 2019; Kaemingk, 2019). Typically, these reviews are unsolicited and authentically reflect customer sentiment. Nevertheless, the huge volume and varying structure of online reviews pose considerable challenges to effective analysis (Chernikova et al., 2020). Therefore, advanced text analytics methods are required to leverage this valuable data. Techniques such as sentiment analysis, topic modeling, and pre-trained language model-based approaches have been demonstrated as effective tools in identifying operational strengths and weaknesses, thereby offering systematic and scientific foundations for strategic decision-making (Anusuya, 2023; Rodríguez-Ibánez et al., 2023; Boluki et al., 2024).

To address these challenges, this study focuses on PureGym, the largest fitness chain in the UK, and examines customer reviews from its acquisition by a US private equity firm in 2017 (Curry, 2017) to the present. This research employs advanced social media analysis methods, specifically multimodal sentiment analysis and dynamic topic modeling, to interpret consumer feedback collected from Google Maps. First, multimodal sentiment analysis is used to quantify the emotional attitudes expressed by customers. Subsequently, dynamic topic modeling is utilized to identify prominent themes and recurring operational concerns consistently reported in customer reviews. By integrating these analytical methods, this study aims to bridge the gap between large amounts of consumer-generated data and strategic corporate decision-making, thereby providing PureGym with data-driven insights to enhance customer satisfaction, optimize operational effectiveness, and sustain its competitive advantage within the increasingly challenging fitness marketplace.

2. Literature review

In the contemporary business environment, understanding user needs has become an essential step to sustain company growth. While traditional analytics are effective for analyzing structured data, such methods often fail to capture consumer sentiment, preferences, and emerging trends. Relying primarily on historical data and predefined analytical frameworks, traditional approaches are limited in their ability to reflect the dynamic and nuanced nature of consumer behaviour. As a consequence, rapidly evolving trends that emerge in real time may go undetected (Kumar & Nanda, 2023; Vasilopoulou et al., 2023). Consumer sentiments are often conveyed through unstructured data formats such as social media posts, likes, and images—types of data that traditional tools struggle to process and interpret accurately. In response to these limitations, an increasing number of studies have turned to social media analytics as a more effective approach for capturing and analyzing consumer insights. By systematically analyzing user-generated content across various social media platforms, companies could gain valuable insights into consumer sentiment, preferences, and emerging trends (Durau, 2022; Yin et al., 2024; Sang et al., 2024).

Sentiment analysis, a core component of social media analytics, focuses on identifying the emotional tone inherent in textual data. However, inaccurate results are often produced when users use ambiguous language, sarcasm, or when the interpretation of emotional expressions varies from person to person. Although numerous methodological improvements have been proposed to address these limitations (Medhat et al., 2014; Wei et al., 2024), this issue is far from being resolved. In response, multimodal sentiment analysis integrates diverse data forms—including text, images, or videos—to overcome the constraints of single textual analysis, thereby enabling a more comprehensive and accurate interpretation of user emotions (D'mello & Kory, 2015; Soleymani et al., 2017; Poria et al., 2018). In the health and wellness sector, user-generated content frequently includes multimodal elements, highlighting the particular relevance and value of multimodal sentiment analysis. For example, Shah et al. (2019) recently developed a sentiment classification model combining textual reviews with corresponding image data, achieving improved sentiment recognition accuracy through capturing complex interactions between these multimodal components.

Topic modeling, another widely used social media analytical technique, is employed to uncover latent thematic structures within large textual datasets (Asmussen & Møller, 2019). By applying topic modeling to social media data, companies can effectively identify prevalent discussion topics and potential consumer needs. For example, analysis of discourse in online fitness communities enables firms to better understand user priorities, allowing them to tailor products and services to consumer preferences (Dessart & Duclou, 2019; Fan et al., 2023; Wang et al., 2024). Nevertheless, traditional topic modeling techniques typically provide static snapshots of popular topics, ignoring temporal dynamics and failing to capture topic evolution trends (Wang & Andrew, 2006). To address this limitation, dynamic topic modeling techniques were proposed to analyze changes in topics over time, allowing the identification of evolving consumer interests and emerging market shifts promptly. By examining trends in user-generated content across multiple operiods, companies can effectively gauge the impact of seasonal factors or specific marketing initiatives on consumer attention and engagement (Zhang et al., 2015; Swaminathan et al., 2022).

3. Data and Methods

The data used in this study were sourced from the official PureGym website and the Google Maps platform. The specific data collection process was conducted as follows. First, basic information regarding PureGym locations across the UK was extracted using the “Find a Gym” feature provided by PureGym’s official website, resulting in a comprehensive list of PureGym stores. This preliminary dataset comprised 442 gyms, including both operating locations and those planned for future opening.

Subsequently, further data retrieval was performed for each identified gym location through the Google Maps search functionality. This step facilitated the collection of detailed gym information and corresponding customer reviews. However, due to inherent limitations associated with the Google Maps search algorithm, such as inaccurate text matching and inconsistent naming conventions, geographic annotation errors occasionally emerged (Bandy & Hecht, 2021). These limitations caused non-PureGym facilities or permanently closed locations to appear erroneously in the search results. Moreover, a considerable proportion of collected review data only consisted of numerical ratings without accompanying textual reviews, thus reducing the utility for further analysis.

To ensure data reliability and analytical accuracy, this study adopted rigorous data-cleaning procedures informed by the method proposed by Li et al. (2022). This process entailed removing irrelevant entries, incomplete data, and reviews falling outside the specified time frame of the research. Consequently, a refined and verified dataset was established, consisting of 47,398 valid reviews from 415 PureGym locations (see Table 1). This curated dataset provides a robust foundation for subsequent analyses, facilitating a precise and comprehensive exploration of PureGym user experiences.

Table 1. Variables for Social Media Analytics.

Variable Name	Description	Data Type
Search Query	The gym name obtained from the PureGym website was also used as a search query in Google Maps.	Text
CSV Address	The gym address obtained from the PureGym official website	Text
Merchant Address	The gym address obtained from Google Maps is used to compare with the CSV Address to ensure the search is correct.	Text
Username	The username obtained from Google Maps, converted using the SHA-256 algorithm to ensure privacy.	Text
Review	User reviews from Google Maps.	Text
Rating	User ratings obtained from Google Maps.	Number
Reply	Replies to the review from other users, pulled from Google Maps.	Text
Likes	Likes of reviews by other users from Google Maps	Number
Adj Review Time	The time of the review retrieved from Google Maps, converted to years.	Number
Since Open	The review time obtained from Google Maps, converted to the opening time of the address.	Number
Reply Status	Whether reviews obtained from Google Maps are replied to.	Number

3.1 Web Scraping

This study employed web scraping techniques to systematically collect store location data from PureGym's official website, as well as review data from the Google Maps platform. Contemporary web platforms frequently incorporate dynamically loaded content, regularly changing page structures, and mechanisms such as IP address blocking, posing significant challenges to traditional static web scraping approaches. Such methods have proven inadequate in addressing these complexities and may inadvertently infringe upon the terms of service or copyright provisions of respective platforms (Foerderer, 2023; Brown et al., 2024). Both PureGym and Google Maps utilize substantial amounts of dynamically generated content, review lists that load incrementally upon scrolling, and interactive elements, including consent dialogs and buttons that require user engagement.

To effectively address these technical challenges, this study utilized Selenium WebDriver, a dynamic scraping technology known for significantly reducing development cycles and improving the efficacy of data extraction (Leotta et al., 2025). Selenium provides an automated, browser-based environment capable of simulating authentic user interactions, such as button clicks, page navigation, and scrolling. Additionally, this research integrated Selenium with BeautifulSoup, a specialized parsing tool, to accurately process the dynamically loaded HTML content and extract relevant data fields, including usernames, review texts, ratings, and timestamps.

Furthermore, following the website privacy protection policies outlined by the European Commission (European Commission, 2025), necessary anonymization procedures were implemented to safeguard user identities. Specifically, usernames extracted during data collection were anonymized using the SHA-256 transfer, ensuring compliance with privacy regulations and preventing inadvertent exposure of personally identifiable information.

3.2 Multimodal Sentiment Analysis

In user-generated reviews on the Google Maps platform, consumers typically provide textual reviews accompanied by numerical ratings, this is equivalent to providing a direct emotional score. However, due to inherent differences in subjective interpretation, these numerical ratings sometimes diverge from the actual emotional tone conveyed in the corresponding textual content (different users may give different ratings to similar text content based on their own emotional benchmarks or personal preferences). To address this inconsistency, this study employs a multimodal sentiment analysis approach, incorporating user interaction behaviors as supplementary features alongside textual reviews, thereby facilitating a more nuanced and accurate interpretation of sentiment. Compared to text-only analysis methods, feature-level fusion enables early-stage exploration of inter-modal correlations, consequently enhancing the performance of sentiment classification tasks (Poria et al., 2017).

This study proposes that user "likes" reflect collective sentiment or consensus among a broader audience and thus serve as an external benchmark for calibrating individual textual expressions of emotion. Specifically, the sentiment classification model assigns greater weight to reviews receiving higher numbers of "likes," thus prioritizing user-recognized reviews to establish a more objective and consistent sentiment benchmark during model training. Subsequently, a multimodal neural network model based on BERT was trained using this enhanced dataset, combining text embeddings with additional numerical features, such as the number of user likes, reply statuses, and the store opening year when the review was made. This approach recalibrates the originally subjective numerical ratings, yielding more accurate, stable, and consistent sentiment evaluation outcomes.

3.3 Dynamic Topic Modelling

Before constructing the dynamic topic model, this study conducted rigorous preprocessing of the review data obtained from Google Maps to ensure data integrity and standardization. This phase involved tokenization, lemmatization, and the removal of extraneous content, including URL links, punctuation, and stop words. Subsequently, the cleaned textual reviews were transformed into high-quality semantic representations using a text embedding approach based on the SentenceTransformer model. This method was selected for its proven effectiveness in accurately capturing semantic nuances and enhancing topic identification and differentiation capabilities (Mersha et al., 2024).

Following the embedding step, the BERTopic model was utilized to conduct a topic clustering analysis, thereby revealing latent thematic structures within the review data. Importantly, this study incorporated temporal segmentation of the dataset by employing the timestamps associated with individual reviews. This approach enabled the intuitive visualization of how user opinions and areas of concern evolved naturally in response to environmental changes, specific events, or market trends within defined time intervals (Bollen et al., 2021). Consequently, this approach allowed for a more comprehensive and nuanced exploration of the evolving patterns in consumer sentiment and opinions over time.

4. Analysis and Results

4.1 Descriptive Analysis

An analysis of star rating data indicates that the overall customer experience tends to be positive. Specifically, approximately 70% of reviewers provided high ratings of four or five stars (see Figure 1). This distribution aligns with a commonly observed pattern within online review research, where user ratings typically concentrate at the positive end of the rating scale, exhibit fewer occurrences in the middle ratings, and present only a small proportion at the negative end (Hu et al., 2017).

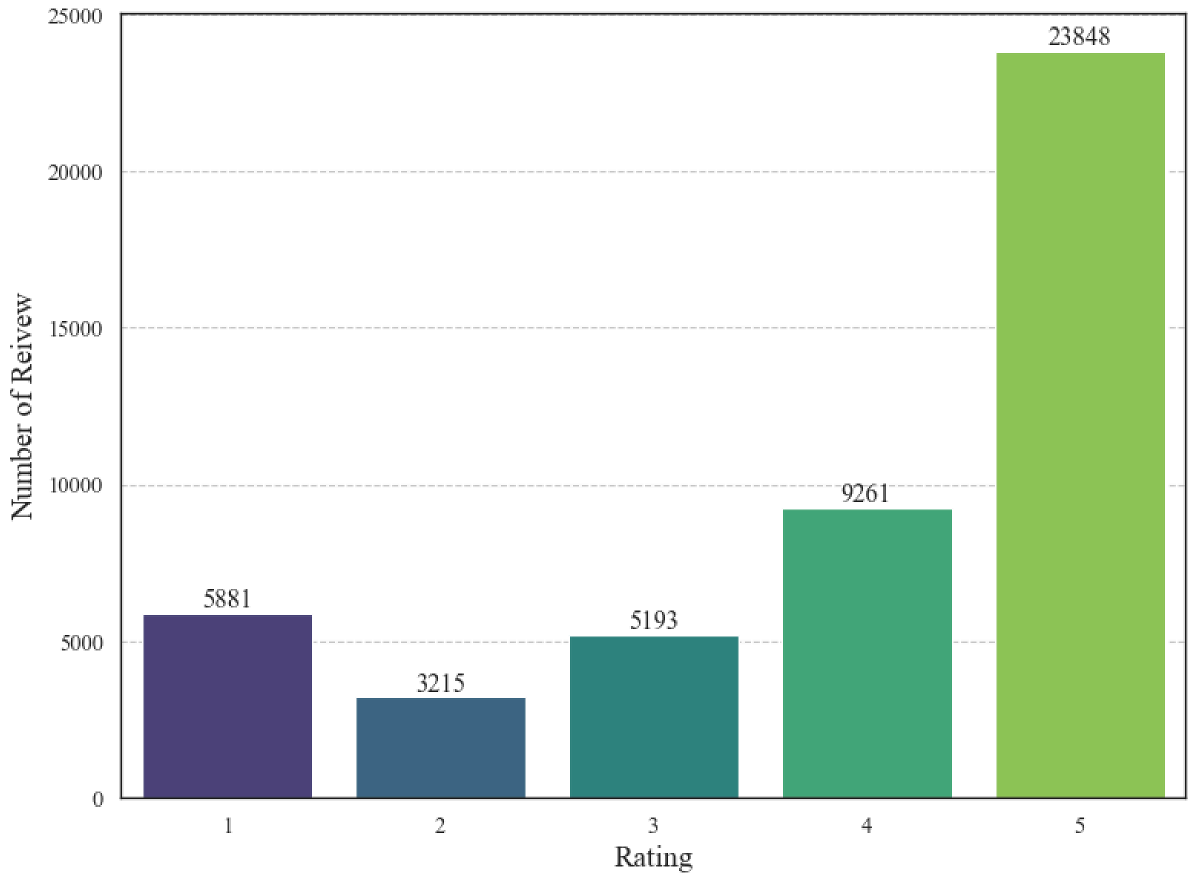

Figure 1. Number of Reviews by Rating.

The temporal analysis of the review data reveals a clear annual growth trend in the total number of reviews. Over the past few years, PureGym reviews have generally increased year by year, rising from fewer than 2,000 annually to more than 10,000 in recent years (see Figure 2). However, a notable decline occurred in 2020-2021, this short-term decrease likely correlates with the COVID-19 pandemic, during which individuals spent more time at home and engaged less frequently with physical gym locations (Rada & Szabó, 2022).

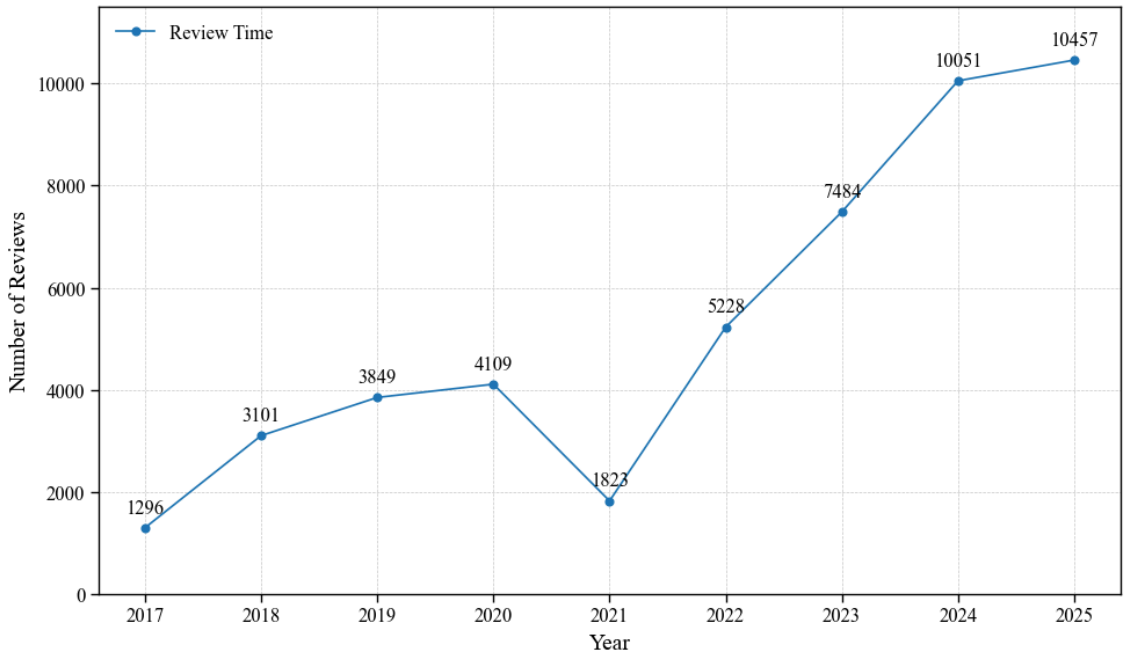

Figure 2. Number of Reviews by Years.

Additionally, an examination between store opening duration and review volume revealed no significant correlation. Regardless of whether the gym locations were newly opened or had an established operational history, the number of reviews remained relatively stable, around 5,000 per year (see Figure 3). This finding suggests that the length of time a gym has been operational does not significantly influence the frequency of consumer reviews.

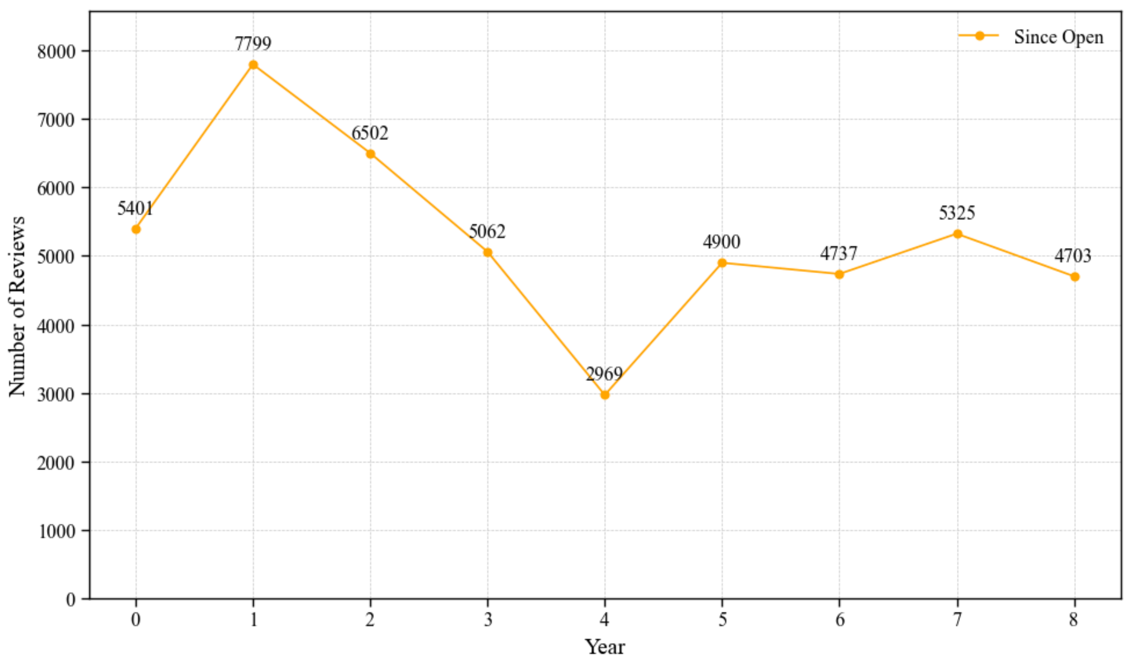

Figure 3. Number of Reviews by Since Open Years.

Finally, from the perspective of user engagement and interaction, the dataset shows that 27.1% of reviews received likes from other users, while an even greater proportion, 53.4%, prompted replies (see Figure 4). These metrics indicate substantial interactive participation among users, suggesting that reviews on the platform transcend simple one-way feedback and instead foster interactive two-way communication between reviewers and their peers.

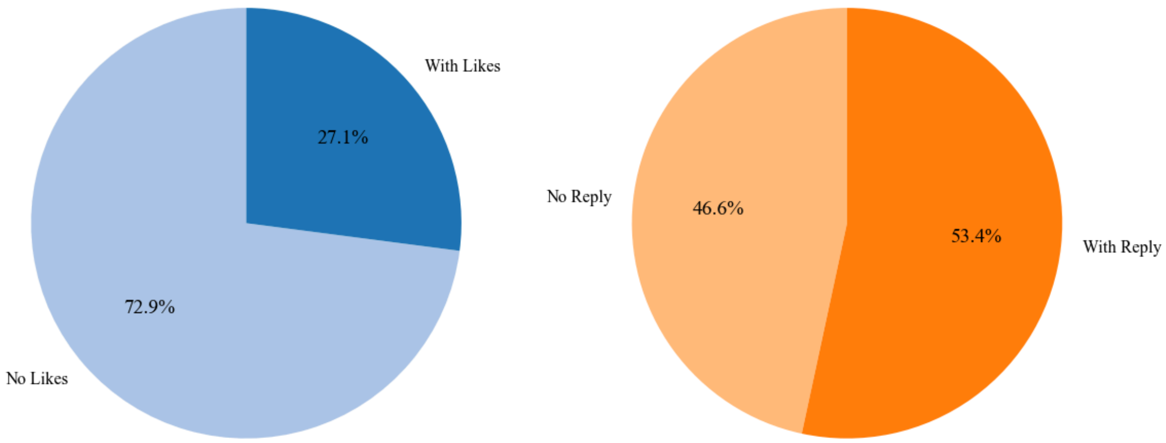

Figure 4. Distribution of Likes and Reply Status.

4.2 Model Comparison

This study compares sentiment recognition results from user-generated reviews using three different approaches: user-provided ratings (Rating Sentiment Score), the traditional sentiment analysis model VADER (Sentiment Score), and a customized BERT-based sentiment analysis model (Adj Sentiment Score) (see Figure 5). The findings reveal that, compared to the original user ratings, the sentiment scores produced by the VADER model exhibit a more balanced distribution. However, the number of reviews categorized as negative is significantly reduced. This suggests that the traditional model may exhibit an overly optimistic bias, potentially overlooking negative sentiments. In contrast, the customized BERT model effectively preserves the original polarity distribution of user sentiments while also identifying nuanced reviews that fall between standard integer rating categories. This demonstrates the customized model's enhanced ability to capture sentiment with greater accuracy.

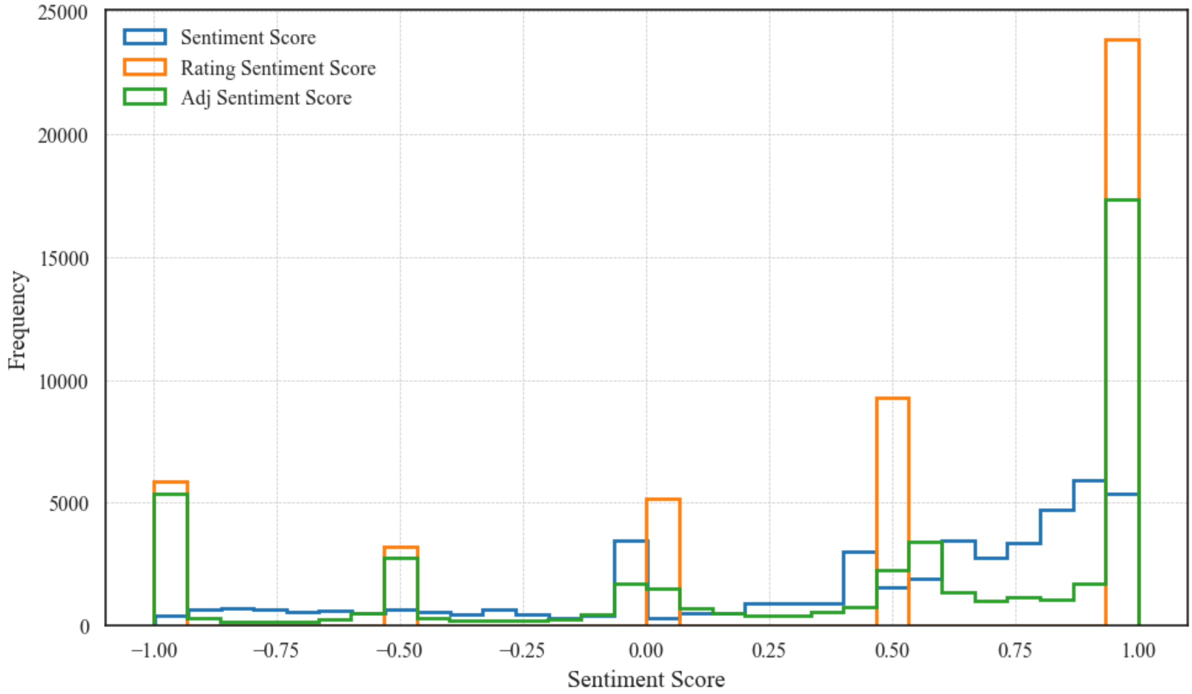

Figure 5. Distribution of Sentiment Scores.

Further analysis of specific review texts provides deeper insight into the discrepancies among the three rating methods (see Table 2). For instance, user 028b3132a6 gave a high rating of 5 stars, which was interpreted as a Rating Sentiment Score of 1. The VADER model rated the sentiment at 0.66 (moderately positive), while the customized BERT model assigned a score of -0.97 (strongly negative). Upon review of the review text, it was evident that the user expressed strong dissatisfaction regarding the smell in the locker room—a sentiment that is clearly negative.

Table 2. Review Samples with Sentiment Score Comparison.

Username	Review	Rating Sentiment Score	Sentiment Score	Adj Sentiment Score
028b3132a6	The odour in the men's locker room is unrecognisable, truly idiosyncratic â" I half expect to find penicillin in one of its corners. However, this acts as an incentive to get in and get out swiftly. No dilly dally round here.	1	0.66	-0.76
16cca6ebdc	I love Rafael class and always looking forward to his classes. His motivation is exquisite and his classes are always full. I'm not the only one, the class mates say the same thing about him. Thank you Rafael, you're the best. Helena	-1	0.92	1
e86ec43035	Cheap gym for cheap price and return you recive cheap service and they make sure yoy feel cheap every munite you speand in this gym, Super rude staff the gym manager completely unHelpful and useless more intrested in internet browsing them helping people out, half of machines not working most of time, its 24 hours but if u go late night they advise you to change in Toilet and if its your lucky day the last person use toilet might use flush after doing business other wise nice charming smell will always welcome you, when come to shower moment u push button you can clearly smeall drain must be magic trick, and obe last thing dont bother to take lock because lockers are broken nearly half of them so if you plan to this place good luck	-1	0.99	-0.98
d0c6afc887	This is an excellent and you get it at a brilliant price. This gym has 24 hour access to members who gain access via their personal PIN. In terms of atmosphere this is a very laid back gym and you don't feel any form of pressure to get off of a machine. It can get busy but that is only a problem at the beginning of the year and Mondays around 6pm. But it isn't so busy to make a visit unpleasant. There is changing rooms and showers inside there is also a shower that is open overnight from 10-5 for late night gym goers. High quality gym would advise anyone to go there.	1	-0.52	1

Conversely, the review from user 16cca6ebdc, associated with a low user rating of 1 star (Rating Sentiment Score of -1), received scores of 0.92 and 1 (highly positive) from the VADER and customized BERT models, respectively. Manual inspection of the review text revealed that the user expressed enthusiastic approval of PureGym's courses, suggesting that the original low rating may have been an anomaly or misrepresentation. Again, the customized BERT model correctly interpreted the sentiment, demonstrating its robustness in recognizing key semantic features and adjusting for inconsistencies in subjective user ratings.

Further comparison between the traditional sentiment analysis model and the customized BERT model reveals that the customized BERT model not only effectively corrects sentiment misclassifications but also aligns more closely with the actual content of user evaluations. For example, user e86ec43035 expressed dissatisfaction with the gym's services, restrooms, lockers, and other facilities. The original user rating was -1 (one star), and the customized BERT model generated a sentiment score of -0.96 (extremely negative), accurately reflecting the user's sentiment. In contrast, the traditional model incorrectly assigned a highly positive score of 0.99, likely due to its inability to detect sarcasm or contextual cues in the text.

Similarly, user d0c6afc887 stated that PureGym was a high-quality gym. Both the original rating and the customized BERT model scored this review as 1 (extremely positive). However, the traditional model misclassified the sentiment as -0.52 (moderately negative). This misjudgment may be attributed to the model's overemphasis on isolated negative words such as "pressure," "busy," and "unpleasant" within the sentence. While these words typically carry negative connotations, their usage in this context was clearly positive, highlighting the traditional model's limitations in capturing nuanced meanings. The customized BERT model, by contrast, more accurately interpreted the broader context.

4.3 Sentiment Trends

Incorporating the temporal dimension of the reviews into the analysis reveals that, over the past nine years, users' overall sentiment ratings of PureGym have remained generally positive (see Figure 6). However, when sentiment data from the COVID-19 pandemic period is excluded, a gradual downward trend in user sentiment becomes evident. This may suggest a potential decline in user satisfaction over time, posing a long-term risk to the company's service quality and brand perception.

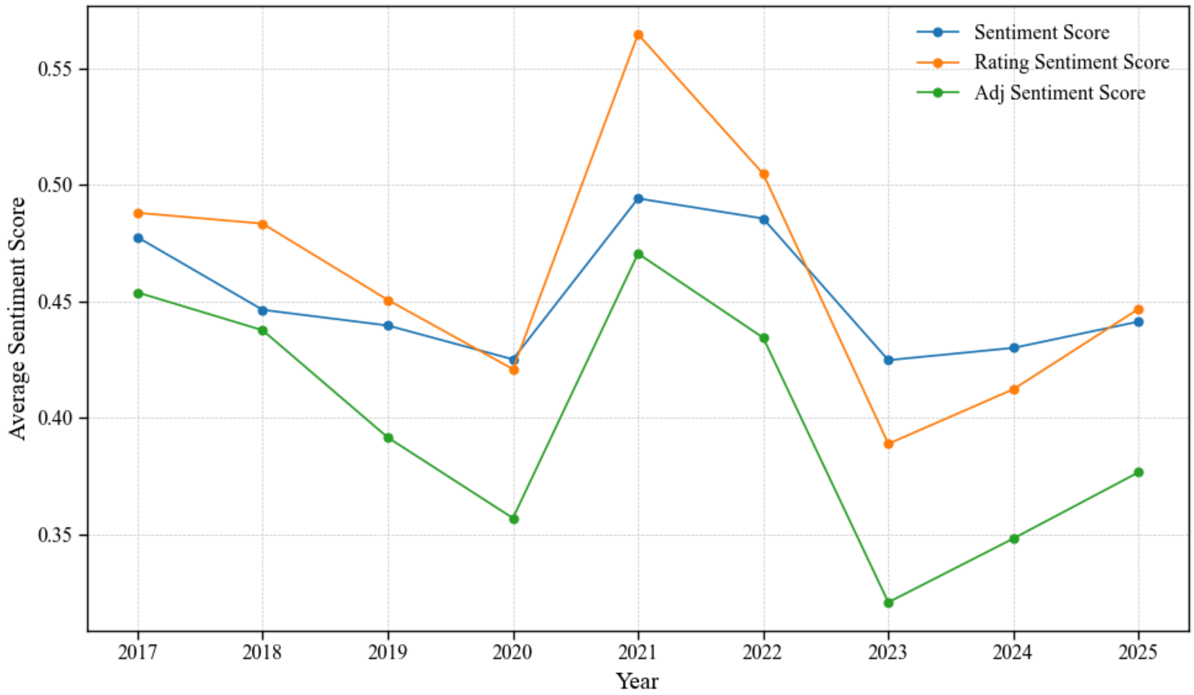

Figure 6. Average Sentiment Scores by Years.

This downward trend becomes more pronounced when analyzed for each gym's years of operation. The result indicates that PureGym locations tend to receive the highest sentiment scores within their first year of operation. Over time, however, user sentiment tends to decline (see Figure 7). While overall ratings remain positive, the persistent downward trajectory warrants attention from the company. Notably, a modest rebound in user sentiment is observed after four years of operation, which may be linked to PureGym's policy of periodic renovations or equipment upgrades. These interventions appear to enhance the user experience temporarily, contributing to an improvement in sentiment scores.

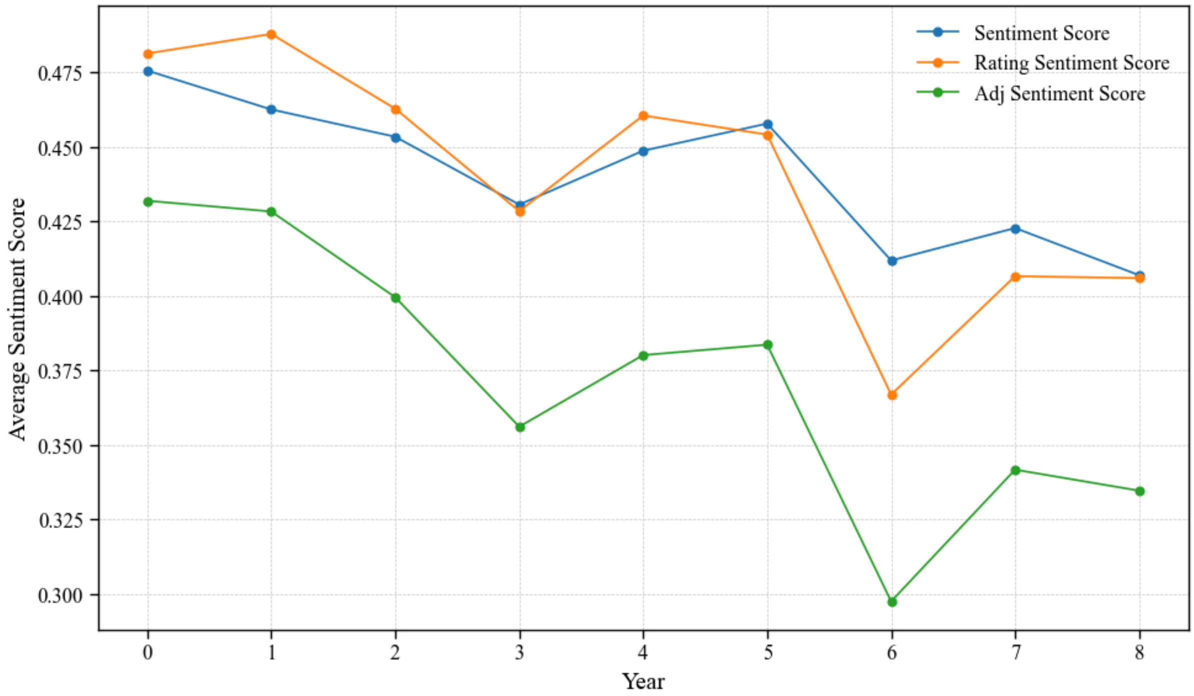

Figure 7. Average Sentiment Scores by Since Open Years.

4.4 Event Detection

While sentiment analysis offers an intuitive understanding of the overall emotional tendencies of users toward PureGym, it remains insufficient for identifying the specific operational and managerial areas that require improvement. To address this limitation, this study incorporates topic analysis as a complementary approach to sentiment analysis, enabling a deeper exploration of the specific content and concerns reflected in user reviews.

Figure 8 presents a sentiment-topic visualization in the form of word clouds, generated using the BERTopic model to extract thematic content from individual user reviews. The word clouds are categorized by sentiment polarity—positive and negative. In the positive sentiment word cloud, terms such as “staff,” “clean,” and “parking” frequently appear. This indicates that users generally appreciate the staff service attitudes, the cleanliness of facilities, and the availability of adequate parking spaces, highlighting areas where PureGym has received widespread recognition. Conversely, in the word cloud corresponding to negative sentiment, terms such as “smell,” “toilet,” and “shower” are prominently featured. This suggests that certain users are dissatisfied with the air circulation, locker room environment, or quality. These recurring themes point to operational shortcomings that warrant serious attention and targeted managerial intervention.

Figure 8. Word Cloud of Positive (left) and Negative (right) Sentiment Reviews.

To further enhance the analysis, a temporal dimension was introduced through dynamic topic modeling, enabling the examination of how users' concerns have evolved over time. Results indicate that in the early years of PureGym's operation, "parking park" emerged as a primary concern, with "shower cold" also receiving considerable attention (see Figure 9). During the COVID-19 pandemic, "covid measure" and "clean staff" became the dominant topic of concern—reflecting heightened public sensitivity to hygiene and safety in shared environments. Notably, topics such as "smell toilt" and "shower cold" have remained persistent in user feedback, with dissatisfaction related to "music loud" and "air hot" increasing in prominence over time. This trend suggests a gradual decline in user satisfaction with the environment, highlighting a specific area for operational optimization.

Figure 9. Dynamic Topic Table.

5. Conclusion and Limitations

This study investigates the evolving consumer sentiment and operational dynamics of PureGym, the largest fitness chain in the United Kingdom, through a comprehensive analysis of user-generated content on the Google Maps platform. The findings reveal that although the overall sentiment toward PureGym has remained generally positive over time, there is a discernible gradual decline in user satisfaction as gym branches continue to operate. This trend may indicate operational fatigue or inconsistent service quality, both of which could undermine long-term customer loyalty if left unaddressed. Emerging concerns—such as unpleasant odors in restrooms and dissatisfaction with cold showers—have gained prominence in recent years. These insights offer actionable implications for PureGym's corporate strategy, such as enhancing air quality, optimizing in-gym environments, and maintaining high standards of cleanliness. Such improvements could foster sustained customer satisfaction and loyalty.

Despite its contributions, this study has several limitations. First, the data is exclusively sourced from Google Maps reviews. While these reviews are content-rich, they may not fully capture the diversity of PureGym's customer base. Users on platforms such as Instagram, TikTok, or the PureGym mobile application may express different concerns and communication styles. To reduce platform bias and enhance representativeness, future research should incorporate data from multiple platforms (Olteanu et al., 2019). Second, the analysis treats user comments as isolated entries, without accounting for interactions among users. Integrating social network analysis could uncover influential users, identify sub-communities, and highlight socially salient topics within the review ecosystem. This would allow for the identification of opinion leaders and localized sentiment shifts, thereby enriching the analytical framework (Haythornthwaite, 1996). Finally, the study did not consider the broader transformation of the fitness industry, particularly the integration of digital and physical services. As online classes, virtual coaching, and remote interactions become increasingly prevalent, reviews based solely on physical locations may fail to capture the full spectrum of user experiences. Future research should explore the interplay between online and offline channels to assess how hybrid service models align with evolving consumer expectations (Rada & Szabó, 2022).

6. Reference

Anusuya, K. (2023) “Exploring the Power of Topic Modeling Techniques in Analyzing Customer Reviews: A Comparative Analysis”. *arXiv*. 2308.11520.

Asmussen, C.B. and Møller, C. (2019) “Smart literature review: a practical topic modelling approach to exploratory literature review,” *Journal of Big Data*, 6(1), pp. 1–18.

Bandy, J. and Hecht, B. (2021) “Errors in Geotargeted Display Advertising Good News for Local Journalism?,” *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), pp. 1–19.

Bollen, J. et al. (2021) “Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena,” P*roceedings of the International AAAI Conference on Web and Social Media*, 5(1), pp. 450–453.

Boluki, A. et al. (2024) “Evaluating the Effectiveness of Pre-trained Language Models in Predicting the Helpfulness of Online Product Reviews,” *in Intelligent Systems and Applications*, pp. 15–35.

Brown, A. et al. (2024). “Web Scraping for Research: Legal, Ethical, Institutional, and Scientific Considerations.” *arXiv*. 2410.23432.

Chernikova, V. et al. (2020) “Challenges in Usage of Unstructured Data in Marketing Decision Making: An Abstract,” *Marketing Opportunities and Challenges in a Changing Global Marketplace: Proceedings of the 2019 Academy of Marketing Science (AMS) Annual Conference*, pp. 67–68.

Curry, R. (2017). “US private equity firm with Topshop stake snaps up Pure Gym.” *Telegraph Media Group Holdings Limited*. Available at: https://www.telegraph.co.uk/business/2017/11/03/us-private-equity-firm-topshop-stake-snaps-pure-gym/. (Accessed: 30 March 2025).

D’mello, S.K. and Kory, J. (2015) “A Review and Meta-Analysis of Multimodal Affect Detection Systems,” *ACM Computing Surveys (CSUR)*, 47(3), pp. 1–36.

Dessart, L. and Duclou, M. (2019) “Health and fitness online communities and product behaviour,” *Journal of Product & Brand Management*, 28(2), pp. 188–199.

Durau, J. et al. (2022) “Motivate me to exercise with you: The effects of social media fitness influencers on users’ intentions to engage in physical activity and the role of user gender,” *DIGITAL HEALTH*, 8.

European Commission. (2025). “Privacy policy for websites managed by the European Commission.” *European Commission*. Available at: https://commission.europa.eu/privacy-policy-websites-managed-european-commission_en. (Accessed: 30 March 2025).

Foerderer, J. (2023) “Should we trust web-scraped data?” *arXiv*. 2308.02231.

Haythornthwaite, C. (1996) “Social Network Analysis: An Approach and Technique for the Study of Information Exchange,” *Library & Information Science Research*, 18(4), pp. 323–42.

Hu, N. et al. (2017) “On Self-Selection Biases in Online Product Reviews” *MIS Quarterly*, 41 (2), 449–71.

Igniyte. (2019) “30 Essential Online Review Facts and Stats.” *Igniyte Limited*. Available at: https://www.igniyte.co.uk/blog/30-online-review-facts-and-stats/. (Accessed: 30 March 2025).

Fan, J. et al. (2023) “Research on Users’ Exercise Behaviors of Online Exercise Community Based on Social Capital Theory,” *Systems*, 11(8), p. 411.

Kaemingk, D. (2019), “20 Online Review Stats to Know in 2019,” *Qualtrics*. Available at: https://www.qualtrics.com/blog/online-review-stats/. (Accessed: 30 March 2025).

Kumar, V. and Nanda, P. (2023) “Product innovation: traditional marketing analytics vs. social media analytics approach,” *International Journal of Services Operations and Informatics*, 12(4), pp. 307–324.

Leotta, M., García, B. and Ricca, F. (2025) “A family of experiments to quantify the benefits of adopting WebDriverManager and Selenium-Jupiter,” *Information and Software Technology*, 178.

Li, S. et al. (2022) “Text Mining of User-Generated Content (UGC) for Business Applications in E-Commerce: A Systematic Review,” *Mathematics*, 10(19), p. 3554.

Li, W. et al. (2023) “The Impact of Fitness Influencers on a Social Media Platform on Exercise Intention during the COVID-19 Pandemic: The Role of Parasocial Relationships,” *International journal of environmental research and public health*, 20(2).

Mersha, M.A., Gemeda yigezu, M. and Kalita, J. (2024) “Semantic-Driven Topic Modeling Using Transformer-Based Embeddings and Clustering Algorithms,” *Procedia Computer Science*, 244, pp. 121–132.

Olteanu, A. et al. (2019) “Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries,” *Frontiers in Big Data*, 2.

Poria, S. et al. (2017) “A review of affective computing: From unimodal analysis to multimodal fusion,” *Information Fusion*, 37, pp. 98–125.

Poria, S. et al. (2018) “Combining Textual Clues with Audio-Visual Information for Multimodal Sentiment Analysis,” *Multimodal Sentiment Analysis*.

Rada, A. and Szabó, Á. (2022) "The impact of the pandemic on the fitness sector – The general international situation and a Hungarian example," *Society and Economy*, 44(4), pp. 477–497.

Rodríguez-Ibánez, M. et al. (2023) "A review on sentiment analysis from social media platforms," *Expert Systems With Applications*, 223.

Sang, V.M. et al. (2024) "Impact of user-generated content in digital platforms on purchase intention: the mediator role of user emotion in the electronic product industry," *Cogent Business & Management*, 11(1).

Sevilmiş, A. and Şirin, E.F. (2022), "Fitness and digital transformation in Turkey", in García-Fernández, J., Valcarce-Torrente, M., Mohammadi, S. and Gálvez-Ruiz, P. (Eds), *The Digital Transformation of the Fitness Sector: A Global Perspective*, Emerald Publishing, Bingley, pp. 39-48.

Shah, A.M. et al. (2019) "Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach," *Journal of Ambient Intelligence and Humanized Computing*, 11(7), pp. 2925–2942.

Soleymani, M. et al. (2017) "A survey of multimodal sentiment analysis," *Image and Vision Computing*, 65, pp. 3–14.

Swaminathan, V. et al. (2022) "The Language of Brands in Social Media: Using Topic Modeling on Social Media Conversations to Drive Brand Strategy," *Journal of Interactive Marketing*, 57(2), pp. 255–277.

Vasilopoulou, C. et al. (2023) "Big Data and Consumer Behavior: The Power and Pitfalls of Analytics in the Digital Age," *Technium Social Sciences Journal*, 45, pp. 469–480.

Medhat, M. et al. (2014) "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, 5(4), pp. 1093–1113.

Wang, L. et al. (2024). Fitness or socializing - A multi-dimensional analysis of online fitness communities users. *iScience*, 27(7), 109753.

Wei, Y. et al. (2024) "Towards multimodal sarcasm detection via label-aware graph contrastive learning with back-translation augmentation," *Knowledge-Based Systems*, 300.

Wang, W and McCallum, A. (2006) "Topics over time: a non-Markov continuous-time model of topical trends." In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '06). *Association for Computing Machinery*, New York, NY, USA, 424–433.

Yin, H. et al. (2024) “An Empirical Investigation into the Impact of Social Media Fitness Videos on Users’ Exercise Intentions,” *Behavioral Sciences* (Basel, Switzerland), 14(3).

Zhang H. et al. (2015) “Dynamic topic modeling for monitoring market competition from online text and image data,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015-August, pp. 1425–1434.

7. Appendix

https://github.com/SigaoLi/UB_SM_PUREGYM