



APPLYING NLP FOR TOPIC MODELLING IN A REAL-LIFE CONTEXT



Candidate: Kazi Md Hasanul Hoque

Course Code: CAM_C301

Week: 4 & 5

Date: 18/01/2026

CONTENTS

1.	Problem context and objectives.....	2
<u>2.</u>	Data Preparation and Exploration	2
3.	Topic modelling and further investigation	5
4.	Emotion analysis and anger focused topics.....	14
5.	LDA and model comparison.....	15
6.	Results, business interpretation and recommendations	16
7.	Conclusion.....	17

1. Problem context and objectives

PureGym is a large low cost gym operator with around 2 million members and hundreds of locations, competing on affordability, flexibility and 24/7 access. In such a high volume, low margin business, understanding and acting on customer feedback is critical to maintain satisfaction, reduce churn and protect reputation. This project analyses 12 months of customer reviews from Google and Trustpilot to uncover key drivers of negative experiences and derive actionable recommendations.

The main objectives are:

- To clean and preprocess review text from Google and Trustpilot and perform basic exploratory analysis using word frequencies and word clouds.
- To apply topic modelling using BERTopic and Gensim LDA to identify recurrent themes in negative reviews, including a focused analysis of anger driven feedback.
- To apply emotion analysis with a BERT based classifier and, optionally, large language model prompts (Falcon 7b instruct) to refine topics and generate business oriented improvement suggestions.
- To synthesise these findings into clear business insights that can guide PureGym in prioritising improvements at both network and location level.

2. Data Preparation and Exploration

Two datasets were imported: Google_12_months.xlsx and Trustpilot_12_months.xlsx, each containing textual reviews and rating scores. Rows with missing review text were removed using the Comment column for Google and Review Content for Trustpilot, ensuring that subsequent NLP steps operated only on valid text. For location-level analysis, the gym identifiers Club's Name (Google) and Location Name (Trustpilot) were used, and a common location key was created by stripping and standardising names.

Example rows are shown in Table 1

Table 1: Example rows from Google and Trustpilot review datasets

Source	Customer / Title	Location / Club	Date / Created (UTC)	Text (truncated)	Rating
Google	** (Customer Name hidden)	Leeds City Centre North	2024-05-09 23:49:18	<i>Comment missing (NaN).</i>	4
Google	**	Cambridge Leisure Park	2024-05-09 22:48:39	Too many students from two local colleges go here...	1
Google	**	London Holborn	2024-05-09 22:08:14	Best range of equipment, cheaper than regular gyms...	5
Google	**	Cheshunt Brookfield Shopping Park	2024-05-09 21:58:07	Good gym when it's not busy, tend to get too busy at times...	4
Google	**	Bristol Union Gate	2024-05-09	Gym is quite dirty, more	1

Source	Customer / Title	Location / Club	Date / Created (UTC)	Text (truncated)	Rating
			21:48:00	obvious grime than previous months...	
Trustpilot	A very good environment	Solihull Sears Retail Park	2024-05-09 23:29:00	A very good environment.	5
Trustpilot	I love to be part of this gym	Aylesbury	2024-05-09 23:11:00	I love to be part of this gym. Superb value for money...	5
Trustpilot	Extremely busy	Sutton Times Square	2024-05-09 22:51:00	Extremely busy, no fresh air.	1
Trustpilot	Great vibes	London Finchley	2024-05-09 22:35:00	Great vibes, fantastic gym.	5
Trustpilot	Everything it needs to be	Crayford	2024-05-09 22:30:00	Clean, well managed, classes are good.	5

An initial investigation counted the number of unique locations in each dataset and the overlap between them, revealing that many clubs receive feedback on both platforms. Text preprocessing followed a consistent pipeline: reviews were converted to lowercase, numbers and punctuation removed with regular expressions, NLTK stopwords were removed, and tokens were generated using word_tokenize before being joined into a cleaned text string. This produced both token lists and cleaned text fields, reused across word frequency analysis, topic models and emotion analysis.

Word frequency distributions were computed separately for Google and Trustpilot using nltk.FreqDist, and bar charts of the top 10 tokens highlighted recurring vocabulary around equipment, cleanliness, staff, crowding and value.

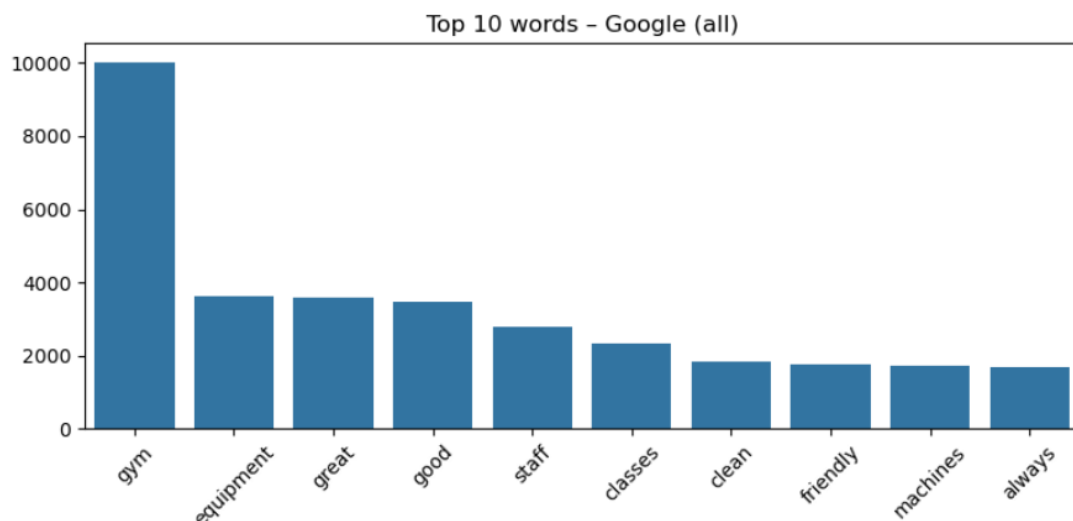
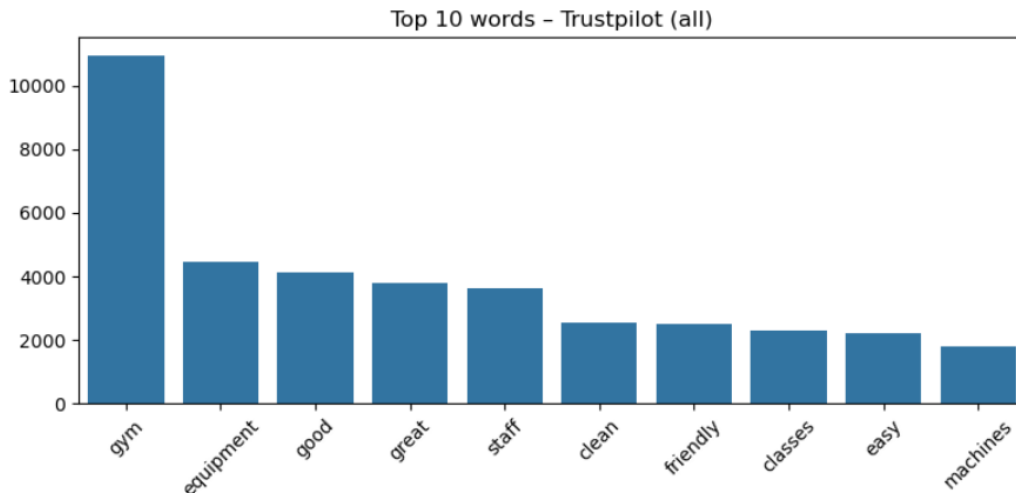


Figure 1: bar chart of top 10 words – Google (all reviews).



Word clouds for both platforms visually reinforced these patterns, with terms related to “clean”, “equipment”, “busy”, “staff”, “classes” and “app” appearing prominently in both positive and negative contexts.

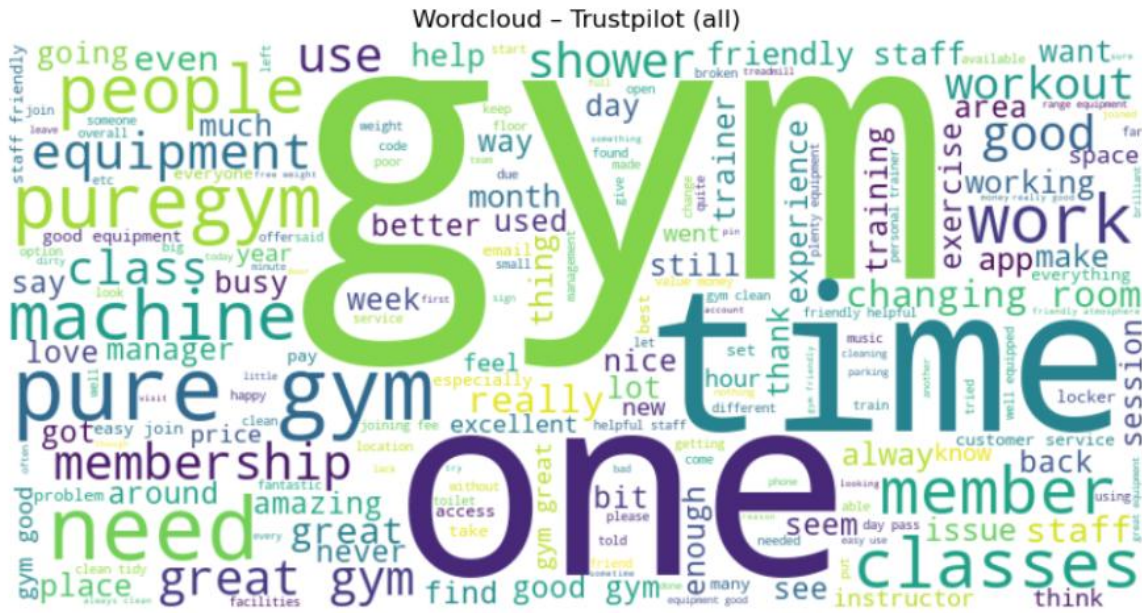


Figure 4: Word Cloud – Trustpilot (all reviews).

These early explorations indicated that operational factors and perceived value are central to the PureGym member experience.

3. Topic modelling and further investigation

Negative reviews were defined as those with rating scores below 3: Overall Score < 3 for Google and Review Stars < 3 for Trustpilot. After filtering, separate word frequency plots and word clouds on the negative subsets showed a stronger emphasis on terms associated with dissatisfaction, such as “dirty”, “broken”, “crowded”, “cancel”, “charge” and “queues”.

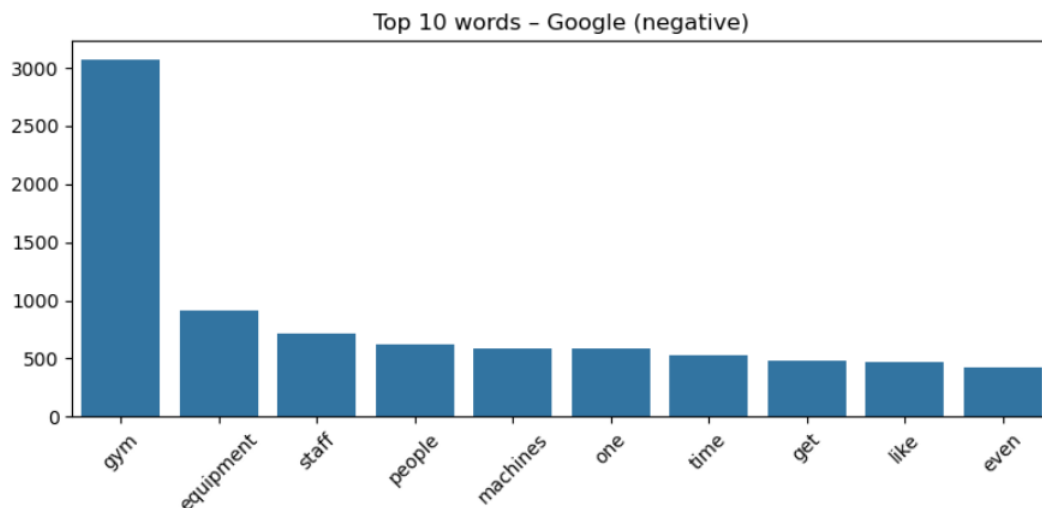


Figure 5: bar chart of top 10 words – Google (negative reviews).

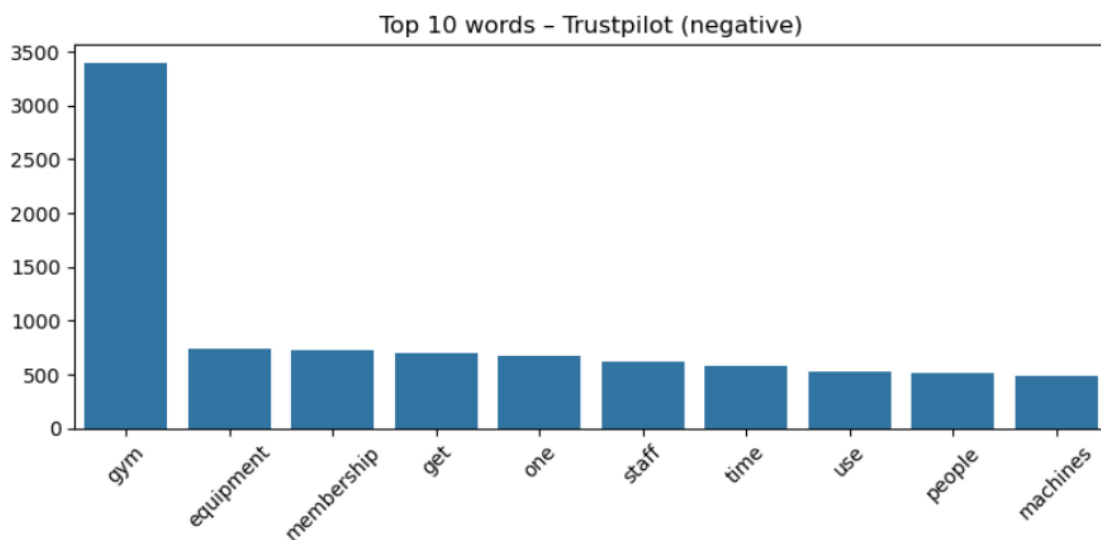


Figure 6: bar chart of top 10 words – Trustpilot (negative reviews).



Figure 7: word cloud – Google (negative reviews).

Topic ID	Document count	Topic label (short name)	Top words (representation)	Example representative document (truncated)
2	140	2_code_pass_pin_day	code, pass, pin, day, access, bought, get, pay	sent pin late waited hour outside gym come sa...
3	118	3_toilets_toilet_changing_dirty	toilets, toilet, changing, dirty, rooms, smell, showers, clean	toilets shower area changing rooms disgusting...
4	116	4_parking_car_park_fine	parking, car, park, fine, free, fines, reg, ticket	free parking times, parking hours, get free p...
5	108	5_smell_dirty_smells_gym	smell, dirty, smells, gym, smelly, ventilation, sweat	wembley gym ongoing problems last three month...
6	90	6_closed_open_days_christmas	closed, open, days, christmas, opening, gym, holiday	visited gym today closed notification closed...
7	79	7_music_loud_noise_hear	music, loud, noise, hear, volume, headphones, speakers	noise gym absolutely atrocious managers owner...
8	60	8_gym_equipment_good_machines	gym, equipment, good, machines, people, location, weights	feeling gym neglected compared puregyms used...

The top topics, described in terms of their highest probability words, reflected themes such as:

- Cleanliness and hygiene in changing rooms, showers and toilets.
- Overcrowding, lack of space and queues for popular machines.
- Broken or poorly maintained equipment and air conditioning.
- Staff behaviour, including rudeness, lack of support and poor enforcement of etiquette.
- Billing, cancellations, access issues and app or PIN problems.

Interactive BERTopic visualisations (similarity heatmap) showed clusters of related topics, allowing grouping into around ten broader themes, for example: cleanliness and facilities; equipment and maintenance; crowding and capacity; staff and customer service; and membership, pricing and access processes.

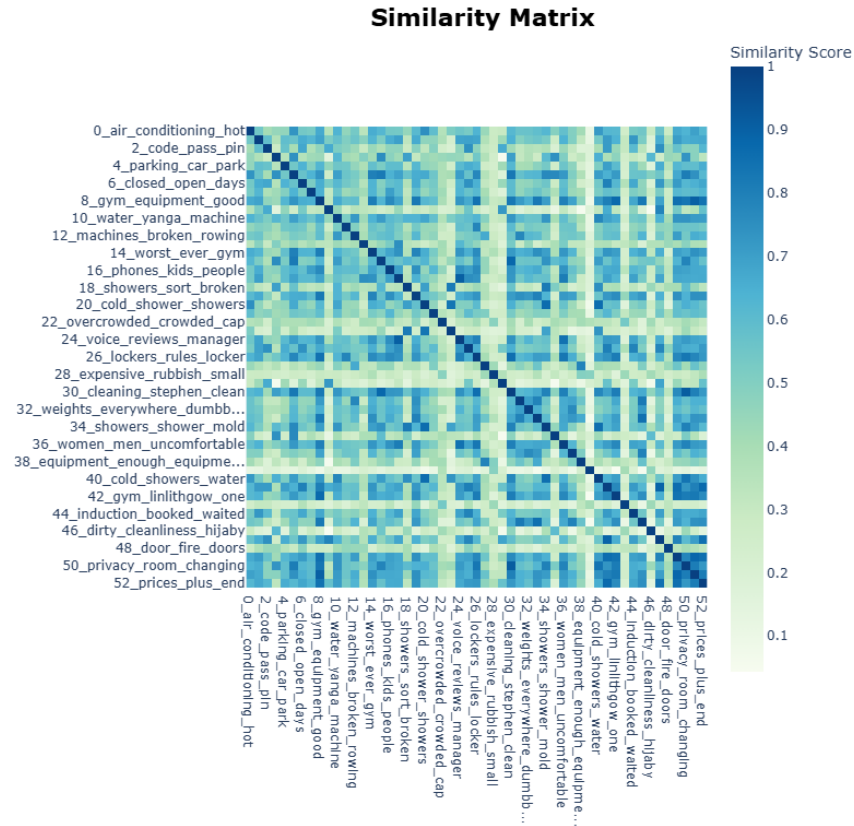


Figure 9: BERTopic heatmap – topic similarity matrix.

A brief description of each cluster was added in the notebook, linking top words to intuitive business themes, which later supported the report’s interpretation.

Further investigation at location level examined the top 20 clubs with the highest number of negative reviews separately for Google and Trustpilot.

Table 3: Top 20 locations by number of negative Google reviews

Rank	Club’s Name	Number of negative Google reviews
1	London Stratford	59
2	London Canary Wharf	26
3	London Woolwich	26
4	London Enfield	25
5	London Swiss Cottage	24
6	London Palmers Green	22

Rank	Club's Name	Number of negative Google reviews
7	Birmingham City Centre	21
8	London Leytonstone	21
9	New Barnet	20
10	Peterborough Serpentine	19
11	Wakefield	19
12	Bradford Thornbury	19
13	London Hoxton	18
14	London Seven Sisters	18
15	Walsall Crown Wharf	18
16	London Hayes	17
17	Manchester Exchange Quay	17
18	Bachenbülach	17
19	London Bermondsey	16
20	Nottingham Colwick	16

Table 4: Top 20 locations with most negative reviews – Trustpilot.

Rank	Location Name	Number of negative Trustpilot reviews
1	Leicester Walnut Street	50
2	345	45
3	London Enfield	23
4	London Stratford	22
5	Burnham	20
6	London Ilford	18
7	London Bermondsey	18
8	Maidenhead	16

Rank	Location Name	Number of negative Trustpilot reviews
9	London Seven Sisters	16
10	London Hayes	16
11	London Finchley	16
12	York	16
13	Northwich	15
14	London Swiss Cottage	15
15	London Bromley	15
16	London Hammersmith Palais	15
17	Dudley Tipton	14
18	Bradford Thornbury	14
19	Watford Waterfields	14
20	Telford	14

While the rankings were not identical, there was noticeable overlap, suggesting that some locations consistently underperform across channels. An aggregated table after merging on a common location key reported, for each club, the number of Google reviews, Trustpilot reviews and the combined total, sorted by total volume.

Table 5: Merged locations – location name, number of Google reviews, number of Trustpilot reviews, total reviews.

Rank	Location (loc_key)	Trustpilot reviews	Google reviews	Total reviews
1	London Park Royal	6 713	6 713	13 426
2	London Stratford	5 432	5 432	10 864
3	Leicester Walnut Street	5 395	5 395	10 790
4	London Finchley	4 794	4 794	9 588
5	Peterborough Brotherhood Retail Park	4 125	4 125	8 250
6	London Enfield	4 081	4 081	8 162
7	Manchester Market Street	4 020	4 020	8 040

Rank	Location (loc_key)	Trustpilot reviews	Google reviews	Total reviews
8	Purley	3 654	3 654	7 308
9	Halifax	3 348	3 348	6 696
10	London Bermondsey	3 180	3 180	6 360
11	London Swiss Cottage	3 127	3 127	6 254
12	Leeds Bramley	2 940	2 940	5 880
13	Wolverhampton Bentley Bridge	2 814	2 814	5 628
14	Burnham	2 808	2 808	5 616
15	Maidenhead	2 805	2 805	5 610
16	Port Talbot	2 800	2 800	5 600
17	London Hammersmith Palais	2 750	2 750	5 500
18	Manchester Stretford	2 700	2 700	5 400
19	Stoke on Trent North	2 573	2 573	5 146
20	Caerphilly	2 530	2 530	5 060
21	Tyldesley	2 457	2 457	4 914
22	New Barnet	2 268	2 268	4 536
23	Glasgow Giffnock	2 244	2 244	4 488
24	Stockport South	2 236	2 236	4 472
25	Northwich	2 220	2 220	4 440
26	Warrington Central	2 210	2 210	4 420
27	Altrincham	2 178	2 178	4 356
28	Birmingham City Centre	2 160	2 160	4 320
29	Sheffield Crystal Peaks	2 135	2 135	4 270
30	London Southgate	2 128	2 128	4 256

For the top 30 locations by total reviews, word frequency and word clouds were recomputed. In many cases, high volume locations showed a stronger emphasis on crowding, equipment availability and cleanliness compared with the overall corpus, indicating that busy urban clubs face particular operational pressures.——

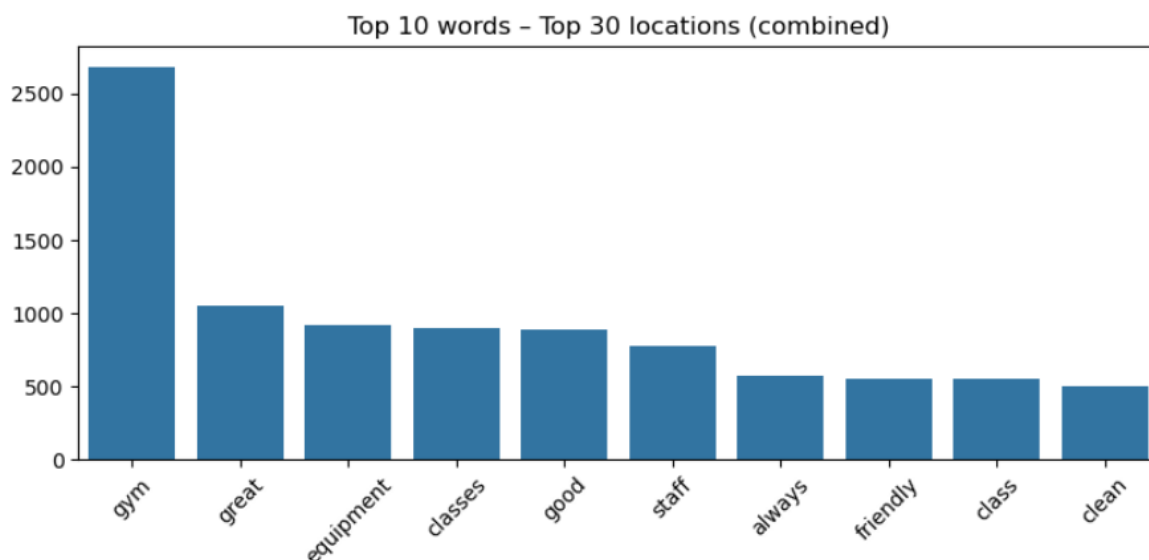


Figure 10: bar chart of top words – combined reviews for top 30 locations.



Figure 11: Word Cloud – combined reviews for top 30 locations.

Where BERTopic was rerun on the top 30 location subset, the high volume environment sharpened themes related to overcrowding and waiting times, providing more targeted insight for site specific interventions.

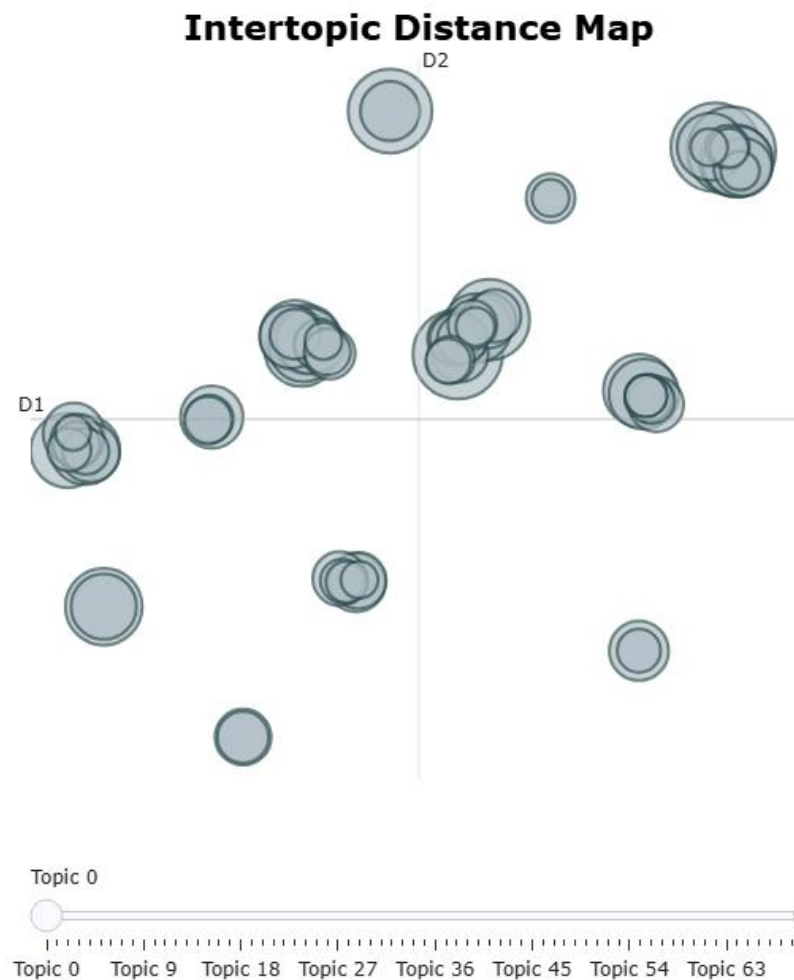


Figure 12: BERTopic intertopic map – top 30 locations subset .

4. Emotion analysis and anger focused topics

Emotion analysis used the Hugging Face model `bhadresh-savani/bert-base-uncased-emotion` via a text classification pipeline, assigning each review a top emotion label such as joy, anger, sadness, fear, love or surprise. An example sentence demonstrated that clearly negative text about dirt and rude staff is predominantly classified as anger or sadness, providing confidence that the model captures intuitive emotional cues. The classifier was then applied to all Google and Trustpilot reviews, and an additional `top_emotion` column was stored for each row.

Focusing on negative reviews only (rating < 3), bar plots of emotion distributions showed that anger and sadness were the dominant emotions on both platforms, with joy and love largely confined to higher rated reviews. Anger was particularly prominent in reviews describing billing problems, unexpected closures, overcrowding and poor cleanliness, while sadness more often appeared in narratives of disappointment or feeling unsafe or unwelcome. To narrow down the most acute issues, all negative reviews whose top emotion was anger were extracted, and BERTopic was rerun on this anger subset. Compared with the original model,

the anger focused topics became more concentrated around a few key failures: cancellations and overcharging; dirty or poorly maintained facilities; lack of staff presence or support; and extreme overcrowding and broken equipment at peak times. This confirms that the most emotionally charged complaints align with operational and administrative pain points that directly undermine perceived value in a low cost gym proposition.

5. LDA and model comparison

To validate the BERTopic findings with a more traditional topic model, a Gensim LDA model with 10 topics was fitted on the tokenised negative reviews from both datasets combined. The preprocessing steps mirrored those used earlier (lowercasing, stopwords removal, tokenisation), and the resulting dictionary and bag-of-words corpus were used to estimate the LDA model.

Table 6: Gensim LDA topics on combined negative reviews

Topic ID	Top words in topic
0	gym, membership, get, month, cancel, pay, use, service, app, email
1	gym, changing, cleaning, like, dirty, rooms, toilets, machine, one, place
2	und, die, nicht, ich, das, ist, der, ger, ein, sind
3	classes, class, instructor, cancelled, one, booked, instructors, gym, first, spin
4	media, social, message, unhelpful, debit, miss, report, branch, back, customers
5	bot, muswell, ten, lessons, failed, weighing, wandsworth, comfort, verry, news
6	gym, equipment, machines, people, water, time, showers, broken, use, enough
7	air, parking, conditioning, con, hot, working, gym, temperature, summer, free
8	gym, staff, member, puregym, members, people, manager, one, pure, like
9	det, ikke, der, har, jeg, man, til, med, kan, min

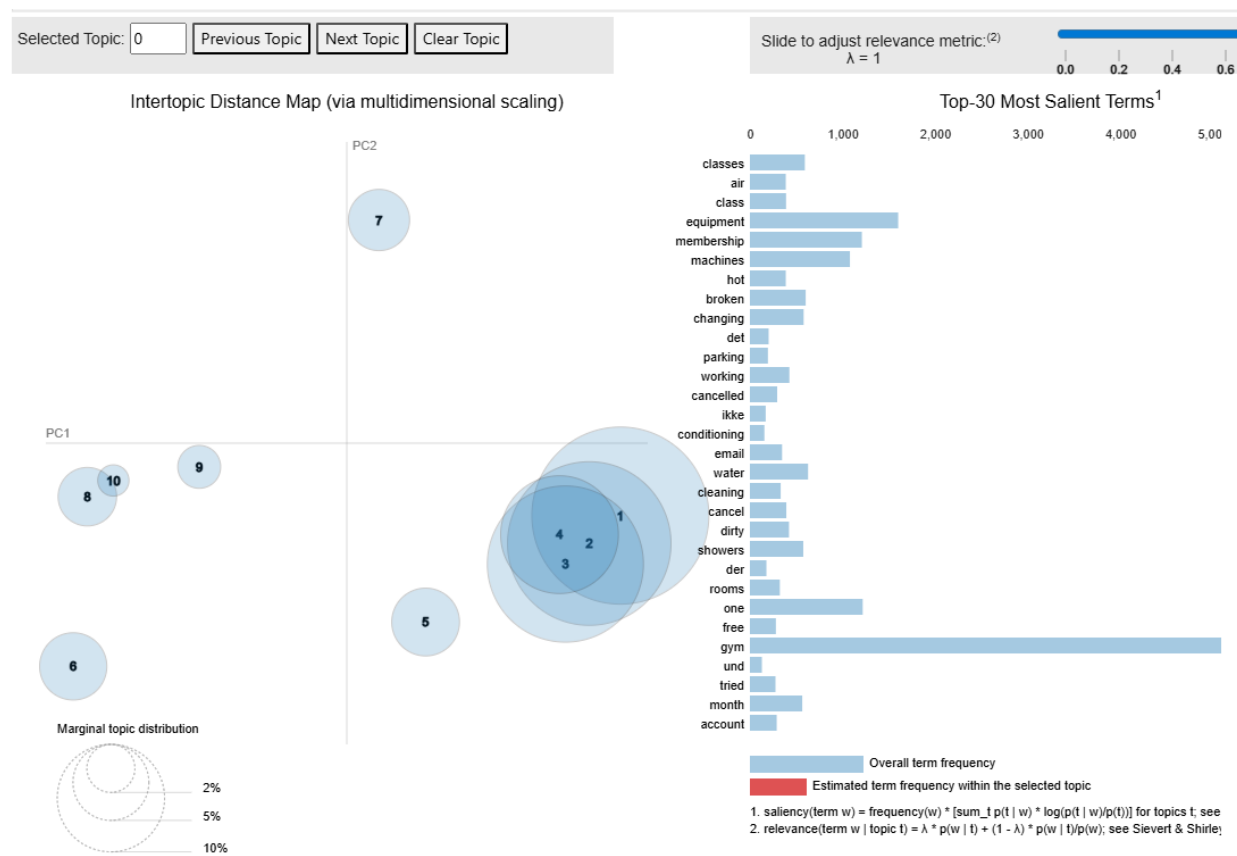


Figure 13: pyLDavis screenshot

Inspection of the top words per topic showed themes that broadly matched the BERTopic clusters, including cleanliness, equipment, staff, crowding and pricing or membership issues. However, LDA topics were generally broader and less sharply separated than BERTopic’s dense clusters, which benefited from contextual embeddings.

6. Results, business interpretation and recommendations

Across all analytic layers—word frequencies, topic models, emotion analysis and location level breakdown—several consistent drivers of negative experience emerge:

- Cleanliness and hygiene problems in changing rooms, showers and toilets, often described in angry language and prominent in both global and anger only topics.
- Overcrowding and insufficient equipment or space at busy locations, generating complaints about queues, “too many people” and “no machines free”, especially in high volume urban clubs.
- Broken or poorly maintained equipment and climate control, including out of order machines, faulty showers and inadequate ventilation or air conditioning.
- Staff behaviour and presence, covering rudeness, lack of help, poor enforcement of gym etiquette, and feelings of an unwelcoming or hostile environment.
- Administrative and digital issues around billing, cancellations, access (PIN and app problems) and communication about opening hours and closures.

Based on these findings, several actionable recommendations arise:

- Prioritise cleaning and maintenance at high risk locations
- Focus additional cleaning resources and regular audits on the top 20–30 locations identified with the highest volume of negative reviews, particularly where cleanliness and facilities topics dominate.
- Address overcrowding and equipment availability
- For busy clubs where topics around crowding and queues are strong, consider capacity management measures such as membership caps, targeted off peak pricing, clearer live busyness indicators in the app, or additional key pieces of equipment.
- Strengthen staff presence and customer service training
- Use topic and emotion insights to design training focused on friendly engagement, enforcing re rack and hygiene etiquette, supporting vulnerable members, and proactively intervening when the environment feels unsafe or hostile.
- Simplify billing, cancellation and access journeys
- The anger focused topics highlight frustration with cancellations, unexpected charges and access issues; reviewing app flows, communications and self service options can reduce these pain points and associated reputational damage.
- Monitor emotions and topics over time as KPIs
- Integrate the emotion classifier and BERTopic into a periodic monitoring pipeline so that shifts in dominant topics and emotional tone at specific locations can trigger early operational interventions.

By systematically mining customer reviews with NLP, PureGym can move beyond anecdotal feedback to a data driven understanding of what matters most to members, where issues are concentrated, and which interventions will most effectively support its mission of providing affordable access to the benefits of being healthy.

7. Conclusion

The analysis of 12 months of Google and Trustpilot reviews shows that PureGym’s main customer pain points are consistently operational and process-driven, rather than strategic misalignment with its value-for-money positioning. Across word frequencies, BERTopic, LDA and emotion analysis, negative reviews concentrate on cleanliness and hygiene in changing areas, equipment reliability, overcrowding, staff behaviour and the transparency of billing and access processes. Locations with the highest volume of negative feedback on both platforms tend to be busy urban clubs, which reinforces the conclusion that crowding and maintenance pressures are localised and therefore need targeted intervention at specific sites rather than generic, network-wide messaging.

The emotion model confirms that anger and sadness dominate the most critical reviews, especially where customers experience dirty facilities, broken equipment, overcharging, difficulty cancelling or unexpected closures, indicating a strong link between these issues and perceived breach of the brand promise of simple, flexible, affordable fitness. BERTopic run on anger-only reviews produces more focused clusters around cancellations, cleanliness, and overcrowding, supporting the view that these are the primary drivers of strong negative sentiment, while LDA recovers similar themes using a completely different modelling approach. Together, these results provide converging evidence that the most impactful improvements for PureGym lie in tightening cleaning and maintenance standards, actively managing capacity at high-risk locations, training staff to handle issues more proactively and empathetically, and simplifying digital journeys for billing and access, all of which are directly grounded in how members describe their experiences in their own words.