

4.1 Mini-project: Applying Topic Modelling in a Real-life Context

Prepared By: Mohamed Nuri

Date - 22/04/2025

Client - PureGym Group



Table of Contents

Introduction.....	3
Approach & Results.....	4
Conclusions:.....	15
Evaluation:	15

Introduction

Problem Statement:

PureGym, a leading fitness brand aims to uncover actionable insights from customer reviews posted on platforms such as Google Reviews and Trustpilot. The goal is to identify the drivers of negative reviews. By analysing these insights, PureGym seeks to enhance the overall customer experience and improve member satisfaction across its locations.

Datasets:

- Two csv files on real-world data, from PureGym containing 12 months' worth of customer review data.
- The first Dataset is from Google Reviews and the other is from Trustpilot.

Google Dataset:

- 23250 Customer reviews
- The Dataset has 7 features, among them are:
 - o Club's Name: Name/location of the gym
 - o Comment: Customer review
 - o Overall Score: Number of stars given out of 5

Trustpilot Dataset:

- 16673 Customer Reviews
- The Dataset has 15 features, among them are:
 - o Review Content: Customer review
 - o Review Stars: Number of stars given out of 5
 - o Review Language: ID for language
 - o Location Name: Name/location of the gym

Approach & Results

Data Cleaning Approach

In our initial data preparation phase, we focused on cleaning and refining both datasets to ensure consistency and quality for analysis.

For both datasets redundant columns, duplicate entries and rows with null values in the review column were dropped and to ensure consistency in text analysis, we filtered the dataset to retain only English-language reviews, removing all non-English entries

Initial Investigation

To identify common themes in customer feedback, we created word clouds using trigrams from the customer reviews. Please see the figure 1 and 2 below:

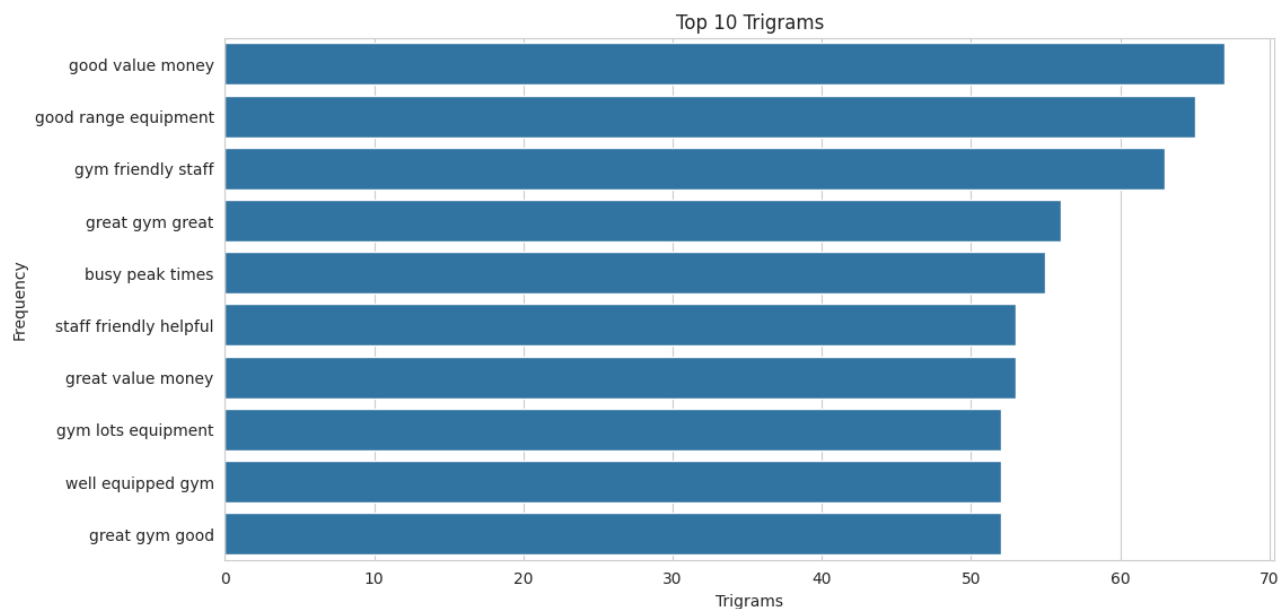


Figure 1- Top 10 trigrams in the Google customer reviews Dataset

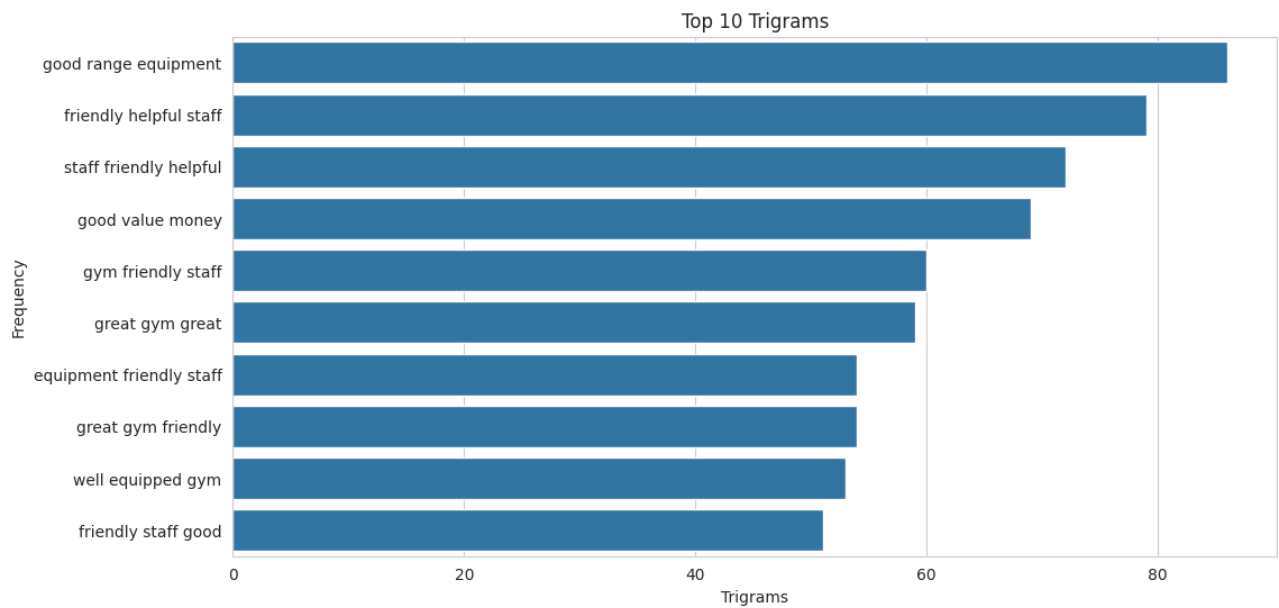


Figure 2 - Top 10 trigrams in the Trustpilot customer reviews Dataset

We then filtered the reviews to keep only those with a score below 3, which we considered to be bad reviews. For these, we created word clouds using trigrams to show the most common phrases in negative reviews from both datasets. The results are shown in the figures below:

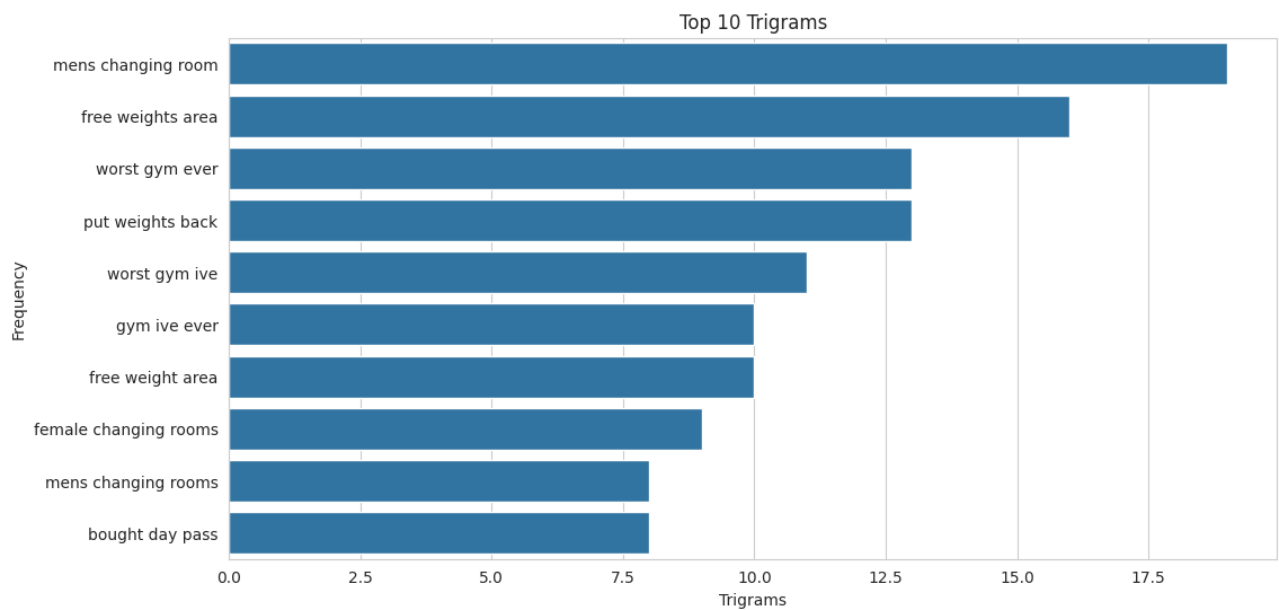


Figure 3 - Top 10 Trigrams Found in Google's Negative Customer Reviews

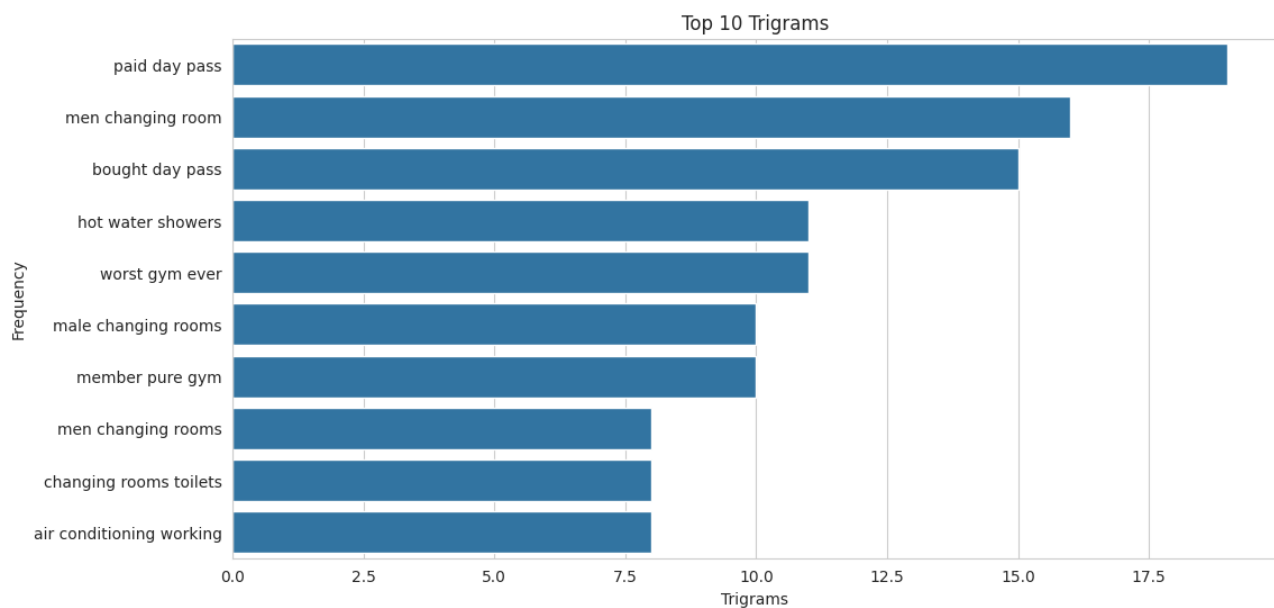


Figure 4 - Top 10 Trigrams Found in Trustpilot Negative Customer Reviews

Conducting Initial Topic Modelling - Using BERT

Since our goal was to identify areas for improvement based on customer feedback, we focused on negative from both datasets and created a combined list of reviews from locations that appeared in both datasets.

After cleaning this combined dataset, we used the combined list as input into Bertopic for topic modelling. Please see figure 6 to see the hierarchical clustering for the topics from Bertopic.

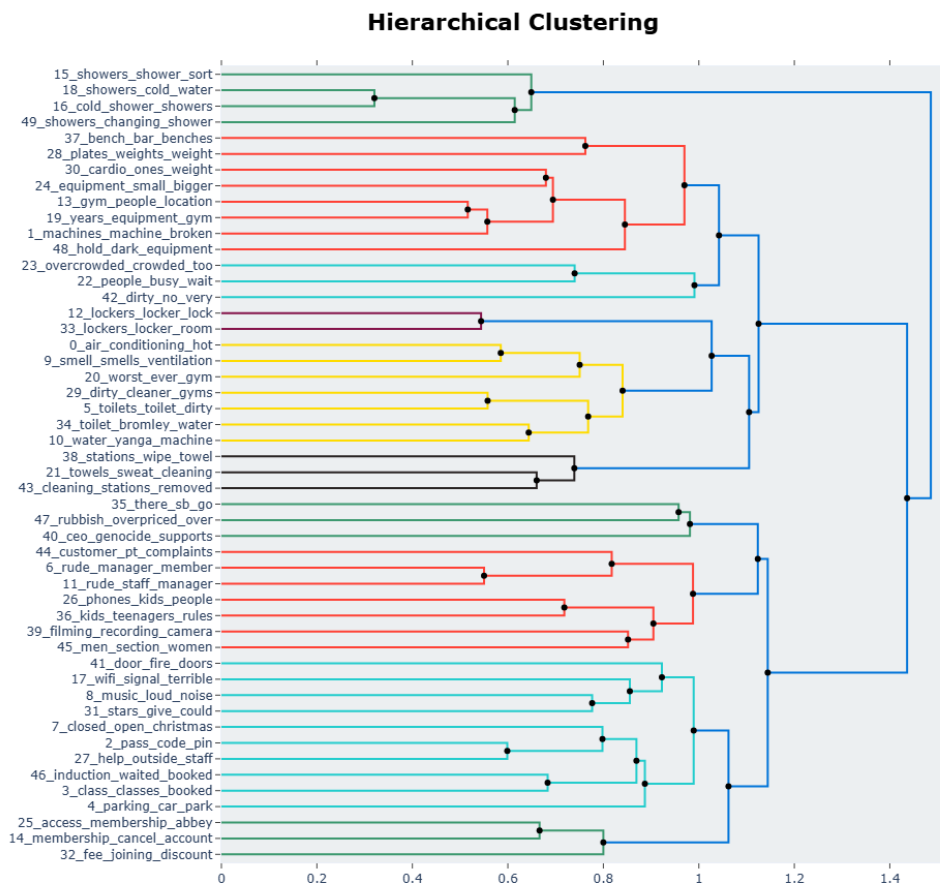


Figure 6 – hierarchical clustering of the topics from the Bertopic model

The themes of the highlighted topics in the hierarchical cluster are:

- The first cluster topic is complaints regarding the showers,
- The second cluster topic is about complaints regarding overcrowding.
- The third cluster topic is on issues with equipment in the gym.
- The fourth cluster topic is on issues with cleanliness and smell.
- The fifth cluster topic is on issues with aircon and a smell coming from the ventilation
- The sixth cluster topic is on issues with regards to loud noise
- The seventh cluster topic is on issues with regards to the car park
- The eight-cluster topic is on issues regarding memberships such as joining fees, membership emails and issues with cancellation
- The ninth cluster topic is regarding staff and issues with regards to poor customer interactions with issues such as rudeness
- The tenth cluster topic is with regards to issues with entrance with regards to pincode and door

Performing further data Investigation

To identify which gym locations received the most negative feedback, we counted the number of negative reviews per location. The top 30 locations with the highest number of negative reviews are shown in figure 7.

	Location Name	Count_Negative_reviews_Google	Count_Negative_reviews_trustpilot	Total_Count_Negative_reviews
0	London Stratford	59	22.0	81.0
59	Leicester Walnut Street	11	50.0	61.0
3	London Enfield	25	23.0	48.0
4	London Swiss Cottage	22	15.0	37.0
7	Birmingham City Centre	20	14.0	34.0
15	London Seven Sisters	17	16.0	33.0
20	London Bermondsey	15	18.0	33.0
8	Bradford Thornbury	19	14.0	33.0
14	London Hayes	17	16.0	33.0
2	London Canary Wharf	26	6.0	32.0
9	New Barnet	18	14.0	32.0
16	Walsall Crown Wharf	17	13.0	30.0
97	Burnham	9	20.0	29.0
29	London Hammersmith Palais	14	15.0	29.0
21	London Muswell Hill	15	13.0	28.0
49	London Finchley	12	16.0	28.0
43	Birmingham Beaufort Park	12	14.0	26.0
41	Basildon	12	14.0	26.0
71	Maidenhead	10	16.0	26.0
17	Nottingham Colwick	16	10.0	26.0
12	London Hoxton	18	8.0	26.0
10	Wakefield	18	7.0	25.0
18	London Park Royal	15	9.0	24.0
6	London Leytonstone	21	3.0	24.0
13	Manchester Exchange Quay	17	6.0	23.0
25	Chatham	14	9.0	23.0
65	Leicester St Georges Way	10	13.0	23.0
30	London Holborn	14	9.0	23.0
86	London Wembley	9	13.0	22.0
158	York	6	16.0	22.0

Figure 7 – The top 30 locations with the highest number of negative reviews

Using negative reviews from the top 30 locations with most negative reviews, we used the combined list on Bert topic modelling. Please see figure 8.

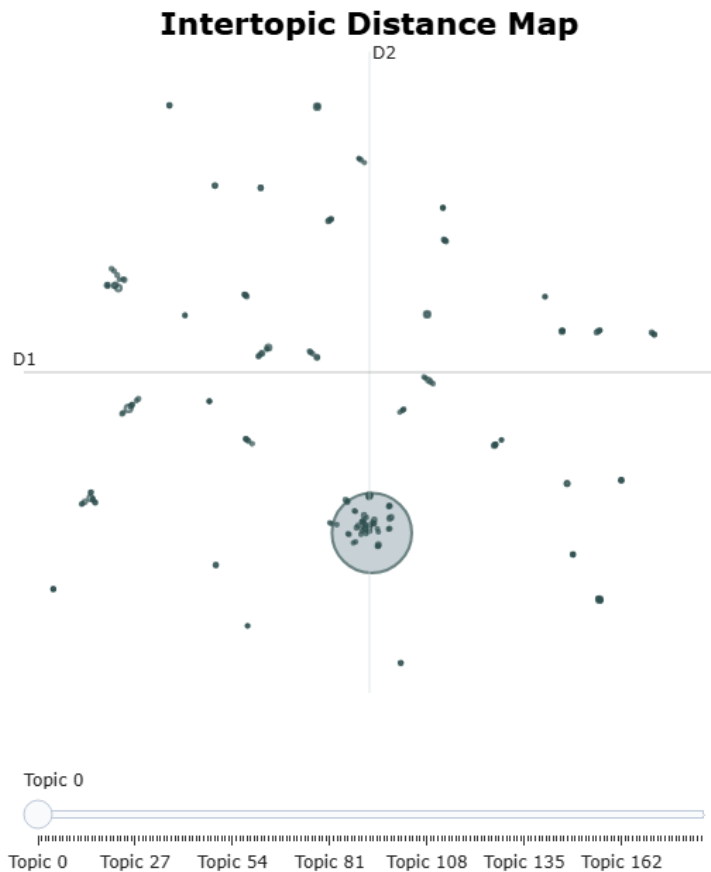


Figure 8 – Intertopic Distance map for the topic from Bertopic Model

Compared with the topic modelling in the initial investigation without filtering based on top 30 locations with negative reviews, this run of bertopic provided more fine-grained clusters and less outliers. The reason for the more fine grained clusters is due to more diverse opinions on locations, with more unique wording which means less shared vocabulary and so harder for BERTopic to merge topics.

Emotional Analysis

We then conducted emotional analysis on the negative reviews from the Google and Trustpilot Datasets. The Top emotion distribution for the Google and Trustpilot Datasets in Figure 9,10.

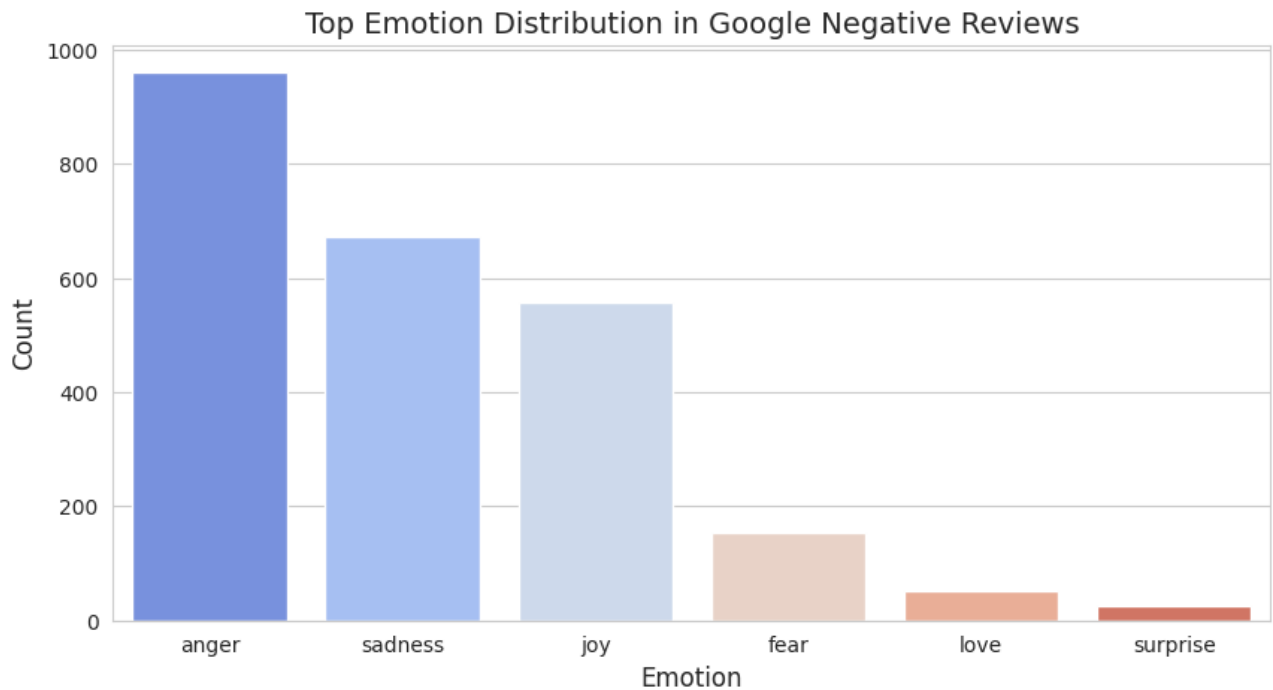


Figure 9

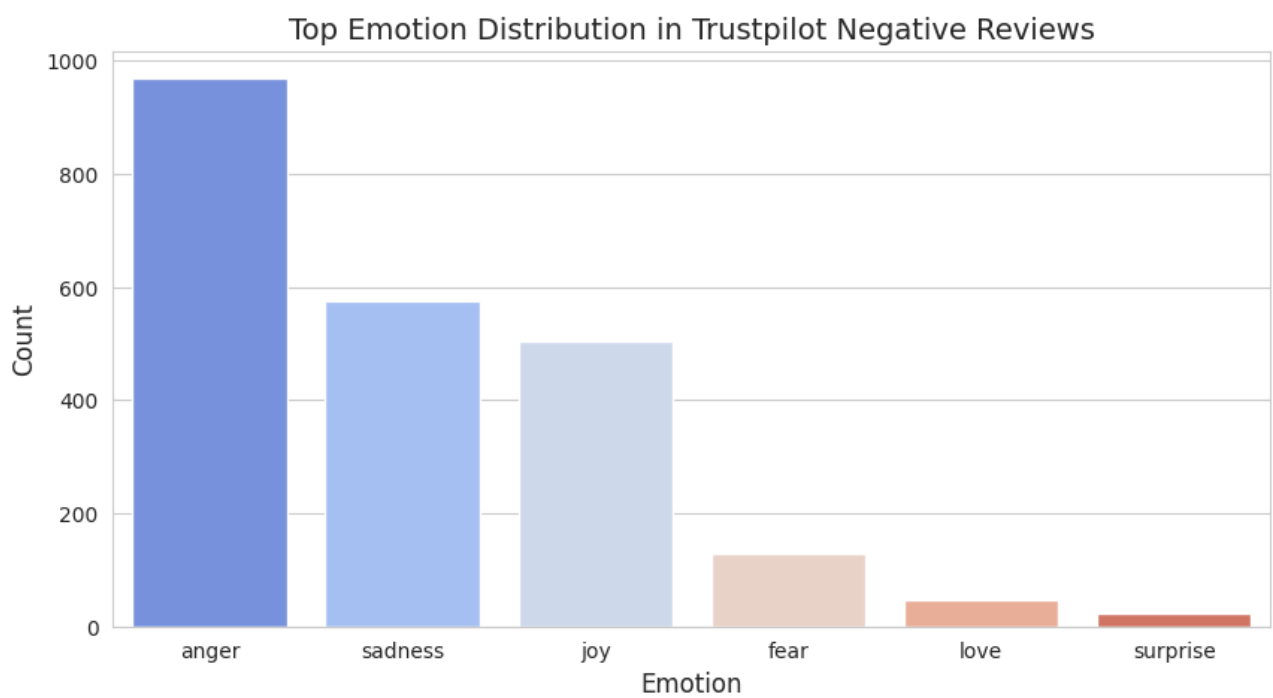


Figure 10

We then filtered the negative reviews where anger was the top emotion and ran Bertopic modelling using those reviews.

In the initial topic modelling, Bert failed to generalise the reviews well and was unable to cluster the topics well, we had to lemmatize the review lists to get the clustering seen below in figure 11.



Figure 11 – Intertopic Distance map for topics

Compared to the clustering in the initial investigation We now have fewer topics 27 in total which is expected, as this analysis focuses specifically on a smaller subset of the data: reviews expressing anger.

Initially, when we ran BERTopic on this subset, the model struggled to cluster effectively and produced a high number of outliers. To improve performance, we lemmatized the review text before feeding it into BERTopic, which helped group similar content more accurately.

Interestingly, the topics identified in this round such as staff behaviour, membership issues, equipment maintenance, access problems, and class cancellations match those found in the previous analysis of all negative reviews.

Using a large language model from Hugging Face:

We used the Microsoft Phi LLM from Hugging Face, prompting it to extract up to three main topics from each review in the combined negative reviews from both datasets.

We then ran the topics extracted on Bertopic and got the results below as seen in the visualisations in figure 12.



Figure 12 – Intertopic Distance map for topics

Bertopic clusters the topics well into 7 distinct topics, however, it's important to note that this analysis was only based on the first 100 negative reviews. In contrast, the earlier BERTopic analyses in both the initial and deeper investigation sections were run on the full set of negative reviews.

Actionable insights:

We then used a prompt to group similar semantic topics into broader categories and generate three actionable insights based on those groupings. Below are the first five categories along with the actionable insights provided by the LLM.

1) Safety and Security:

- Actionable insight: Install security cameras in locker areas and common spaces.
- Actionable insight: Implement a strict locker security policy and provide lockers with combination locks.
- Actionable insight: Increase security personnel presence during peak hours and in areas with previous incidents.

2) Equipment and Maintenance:

- Actionable insight: Schedule regular maintenance checks for all gym equipment.
- Actionable insight: Replace broken and rusty equipment promptly to ensure safety and functionality.
- Actionable insight: Provide adequate cleaning supplies and assign staff to clean equipment after each use.

3) Staff Training and Behaviour:

- Actionable insight: Conduct regular staff training on customer service and conflict resolution.
- Actionable insight: Establish a code of conduct for staff and enforce it consistently.
- Actionable insight: Implement a system for staff to report and address unprofessional behaviour.

4) Customer Experience:

- Actionable insight: Improve gym layout to reduce overcrowding and enhance the workout experience.
- Actionable insight: Offer a variety of classes and equipment to cater to different fitness levels and preferences.
- Actionable insight: Address customer complaints promptly and take corrective action to prevent recurrence.

5) Hygiene and Cleanliness:

- Actionable insight: Increase the frequency of cleaning gym facilities, including changing rooms and bathrooms.
- Actionable insight: Provide adequate cleaning supplies and assign staff to clean equipment and facilities regularly.
- Actionable insight: Implement a policy for customers to report hygiene issues and take immediate action.

LDA Modelling:

We then used the combined negative reviews and ran it on a LDA model, we defined the number of topics as 10.

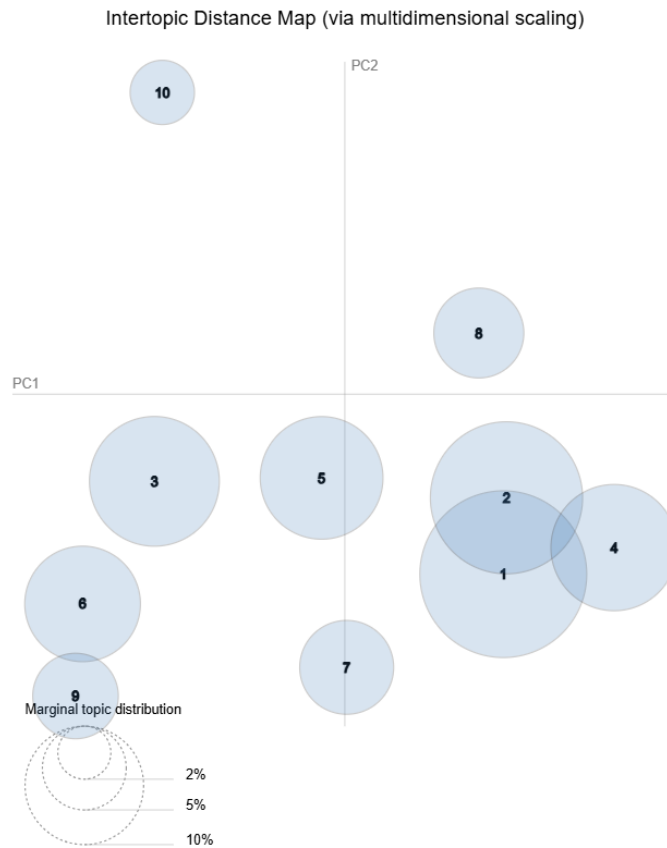


Figure 13 – Intertopic Distance map for topics

The Intertopic Distance Map displays the 10 clusters we predefined in the LDA model. For each cluster, we examined the top 30 most salient terms to interpret the dominant themes:

- Cluster 8: The terms most frequent in cluster 8 are related with temperature air and air conditioning,
- Cluster 3, 9 & 6: are closely clustered around similar regarding membership and access
- Cluster 10: is related to interactions with terms such as manger, rude said etc
- Clusters 1, 2, and 4: These clusters are linked by themes related to gym infrastructure, with salient terms like equipment, machines, and showers. The reviews highlight issues with maintenance & cleanliness of the facilities.
- Cluster 5: This topic is centred around time and staff, suggesting concerns regarding scheduling, and staff availability
- cluster 7: the most salient terms in this cluster is class, music, loud

Conclusions:

- All the techniques utilised for topic modelling were useful in highlighting the key themes in the negative reviews and provided similar topics.
- The most highlighted themes were staff behaviour, equipment maintenance, cleanliness, class scheduling/cancellations, ventilation, overcrowding and membership issues
- The Microsoft phi LLM was the most useful as it was able to not only group the topics but provided actionable insights.
- Emotional analysis on the negative reviews did not provide much insight but further emphasised on the topics Bertopic had extracted in the initial investigation
- In our investigation
- Figure 7 helps show that certain gyms are more prone to negative reviews than others and have high negative reviews from both platforms.

Evaluation:

- For more nuances, review individual gyms that had a high count of negative reviews, and see the distribution of the negative reviews that came from the same gym, so that customer feedback from one geographic area is not generalised to all gyms
- Check the review date for the customer reviews, to understand temporal context did the bulk of the reviews happen around the same time (due to a temporary issue) or is this an unresolved problem.