Applying NLP for Topic Modelling in a Real-Life Context

Stakeholder Report

JULY 2025

PERFORMED BY
LAUREN BRIXEY

Table of Contents

1. Problem Statement & Objectives

Negative customer experiences can significantly impact customer retention and brand perception, especially in highly competitive domains such as the gym industry. The objective of this project was to analyse customer review data from Google and Trustpilot to uncover the key causes of negative sentiment and provide actionable insights for improving the customer experience at [redacted]. This was achieved using a combination of NLP techniques including BERTopic for topic modelling, LLMs for topic-extraction and insight generation, LDA as a baseline model, along with emotion detection to highlight emotionally charged complaints.

2. Data Preparation

2.1. Data Overview

This project used customer reviews datasets from Google Reviews and Trustpilot which included review text, star ratings (1-5), review time stamps and gym locations. Duplicate reviews and reviews with missing or non-English texts were removed to ensure consistency and to focus on customer pain points, only negative reviews were retained (< 3 stars). The datasets were then merged on common gym location names, excluding locations appearing in only one dataset as they likely represented isolated events which could skew the insights. To prioritise the most critical issues, two additional filtering steps were applied:

- **Worst performing locations** where the top 30 locations with the most negative reviews were extracted.
- **Emotion detection** was applied to flag reviews where the top emotion was anger.

2.2. Preprocessing

Review text was converted to lowercase and punctuation, special characters and the brand name was removed. Stopwords were initially retained for BERTopic and LLM analysis to preserve context, though CountVectorizer later removed them within BERTopic. For LDA, stopwords were removed during preprocessing to promote clear topic themes.

3. Results

Initial word cloud visualisations yielded limited insights so BERTopic, a topic modelling technique, was adopted as an approach for extracting coherent and interpretable topics from the review text. To uncover key pain points, BERTopic was applied to progressively concentrated subsets of the negative review data. The first pass analysed all negative reviews in common locations which identified broad issues such as broken or dirty equipment, poor cleanliness and limited gym access (**Figure 1**).

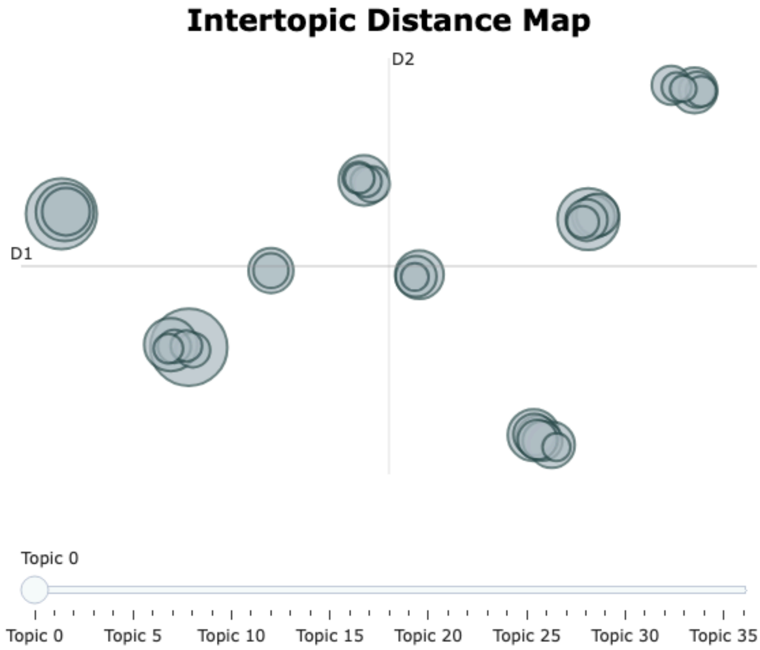

Figure 1: *Intertopic Distance Map of BERTopic run 1 (negative reviews from common locations). Identified 8 clusters with main topic themes: dirty equipment, poor gym facilities, gym classes, broken equipment, dirty and broken facilities, overcrowding, gym access and rude staff.*

Further passes were applied to the negative reviews from the worst-rated locations, followed by isolation of emotionally charged complaints. A fine-tuned BERT-based emotion classifier (bert-base-uncased) was used to identify the top emotion per review which allowed filtering for those flagged as ‘anger’. This revealed more specific and critical issues such as air-conditioning failures, cold showers and aggressive staff behaviour, consolidated into a smaller focused set of topics (**Figure 2**). .

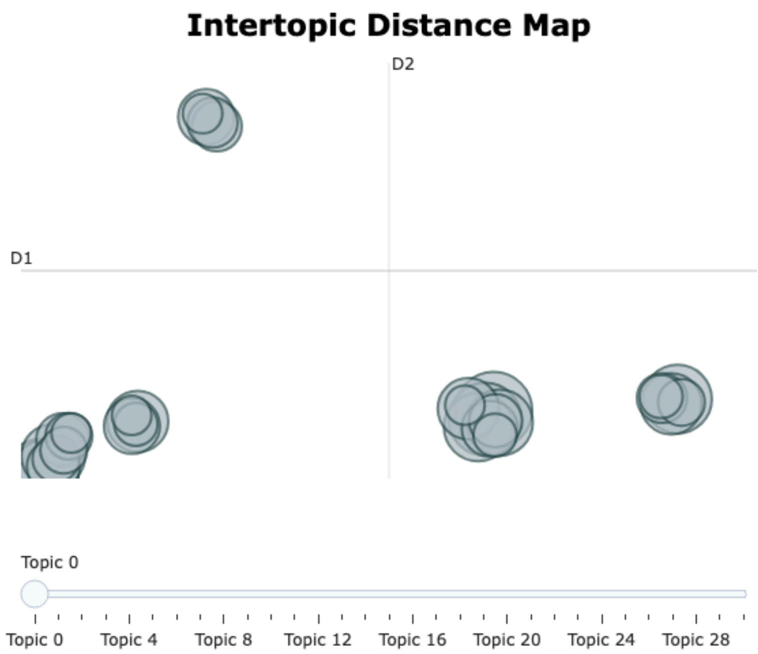

Figure 2: *Intertopic Distance Map of BERTopic run 3 (angry negative reviews from worst common locations). Identified 5 clusters with main topic themes: poor gym facilities, membership, overcrowding, staff behaviour and gym area.*

To gain a deeper, more actionable insight into customer complaints, a large language model (LLM) Mistral-7B Instruct was used in a hybrid approach alongside BERTopic. Due to model constraints, a subset of the 600 most recent negative reviews were selected as these were considered the most reflective of current customer sentiments. Each review was first passed through the LLM with a prompt to extract its top three topics, which were then clustered using BERTopic (**Figure 3**).

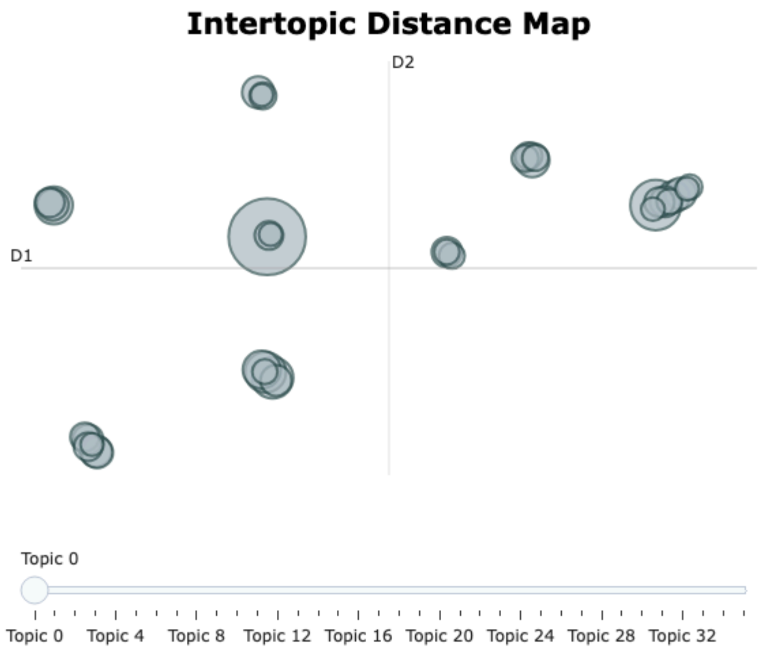

Figure 3: *Intertopic Distance Map of BERTopic run 4 (subset of most recent negative reviews). Identified 8 clusters with main topic themes: price and waiting times, equipment maintenance, equipment variety, classes, facilities, crowdedness, membership and customer service.*

In the second phases, the LLM was further prompted to consolidate the extracted topics generated by BERTopic and to output actionable insights. The extracted topics list was split into smaller batches to accomodate to LLM context limitations and the resulting suggestions were post-processed to eliminate redundancy and identify key themes. This yielded a consolidated list of actionable suggestions, summarised below:

1. **Accessibility and Convenience** to address parking, off-peak access and flexible membership options.
2. **Equipment Quality and Availability** to enhance maintenance routines, upgrade machines and reduce waiting times.
3. **Cleanliness and Hygiene** to implement stricter cleaning protocols
4. **Staff Behaviour and Customer Service** to improve training, professionalism and improve responsiveness.

5. **Class Scheduling and Space Planning** to optimise class time, reduce space conflicts and improve layout design
6. **Membership and Pricing** to simplify payment systems, address login issues and ensure transparent pricing
7. **Atmosphere and Member Experience** to encourage engagement, have clearer policies and promote a community environment.

To benchmark the performance of BERTopic and the LLM-based approach, LDA was used as a baseline topic modelling technique on the negative reviews from common locations dataset. While it partially validated some broad themes such as equipment, cleanliness and classes, its outputs were significantly less interpretable, with many topics containing vague or overlapping concepts and a high topic redundancy (**Figure 4)**. In addition it failed to isolate more specific complaints such as aggressive staff, cold showers and locker shortages.

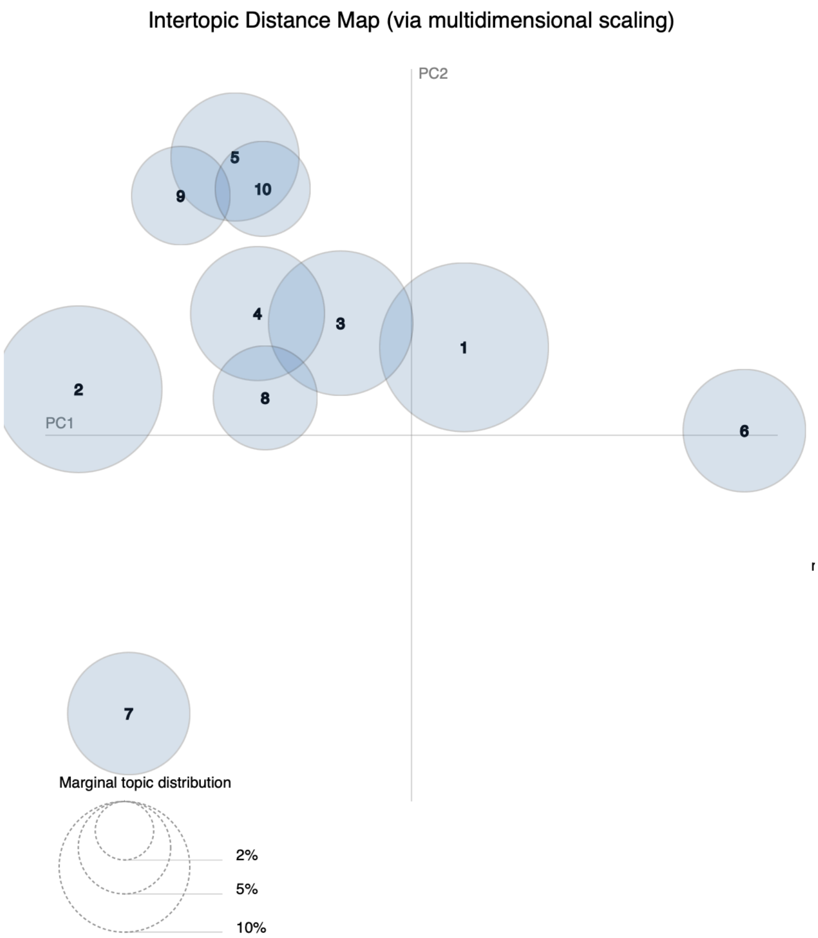

Figure 4: *Intertopic Distance Map of LDA run. IIdentified 10 clusters with main topic themes: equipment (x2), membership, parking, classes (x2), facilities, staff, people and weights, and machines.*

8. Conclusions and Recommendations

Across all modelling approaches, consistent customer pain points included poor hygiene, broken or unavailable equipment, rude staff and overcrowded facilities. Among the methods used, the LLM-enhanced BERTopic approach delivered the most actionable and clearly interpretable insights. Based on these findings, key recommendations include:

1. Implementing more rigorous maintenance and cleaning protocols.
2. Enhancing staff training and customer service responsiveness.
3. Improving membership systems and crowdedness, particularly around class schedules and facility access.

By combining BERtopic, LDA and LLM based analysis with emotion detection, this project achieved a clearer understanding of customer pain points, revealing actionable insights that can guide the future improvement of the customer's gym experience.