Natural Language Processing: Understanding Customer Reviews

1 Introduction

This report presents a natural language processing (NLP) analysis of customer reviews related to PureGym, a major low-cost fitness operator. The objective is to identify key themes, emotional drivers, and actionable insights that can inform improvements to the customer experience. The analysis draws on user-generated reviews from Google and Trustpilot, using a combination of traditional NLP techniques, such as topic modelling and emotion classification, along with large language models for topic extraction and suggestion generation. By comparing results across methods, the report aims to uncover consistent patterns in customer dissatisfaction and propose data-driven areas for improvement.

2 Methodological Overview

This project used a combination of NLP techniques to analyse customer reviews. Initial preprocessing included tokenisation, stopword removal, and lowercasing, followed by frequency analysis and word clouds to visualise dominant terms (Appendices A and B).

Topic modelling was performed using BERTopic and validated with Gensim's latent Dirichlet allocation (LDA). BERTopic, a transformer-based model, clustered semantically similar reviews and extracted interpretable topics. LDA offered a probabilistic comparison based on word distributions.

To analyse emotional tone, the *bert-base-uncased-emotion* BERT model was used to classify reviews by emotion, with a focus on anger to isolate high-frustration content.

Finally, the *falcon-7b-instruct* language model was prompted to extract key topics from negative reviews and generate improvement suggestions, providing deeper insight and actionable guidance.

3 Results & Key Findings

This section presents the findings from each method applied throughout the project. While some techniques require more computational resources than others, this does not necessarily indicate higher reliability. Instead, the most valuable insights emerged from areas where results were consistent across different models.

3.1 Initial Review Insights

An initial exploration of the review content provided early direction for the analysis. Frequency distributions (Appendix A) of the most commonly occurring words revealed prominent themes related to equipment, staff, machines, classes, and people. These were especially prevalent in the negative reviews. The same themes were clearly illustrated in the word clouds presented in Appendix B.

3.2 Topic Modelling with BERTopic

By focusing on reviews rated below three stars, the BERTopic model helped surface the key pain points that likely motivated customer dissatisfaction. The most common topics, along with their top keywords, are displayed in Appendix C.

The ten primary themes included cold showers, rude or unhelpful staff, poor temperature regulation, issues with classes, problems accessing day passes, unclean facilities, parking difficulties, inconsistent opening hours, excessively loud music, and unresolved membership or payment issues, often worsened by inadequate customer service.

When narrowing the analysis to the 30 locations with the highest volume of negative reviews, the most frequently used words remained similar. However, the BERTopic model produced only two topic clusters within this subset. This suggests that complaints in high-traffic locations are more concentrated, pointing to fewer but more prominent issues in those gyms. (The clusters visualised in Appendix D)

3.3 Emotion Analysis

The BERT-based emotion analysis pipeline revealed that anger was the most dominant emotion across negative reviews (Appendix E). Filtering for angry reviews provided clearer insight into the most emotionally charged customer concerns.

Several topics from the earlier BERTopic run reappeared, including staff behaviour, class cancellations, issues with day passes, parking frustrations, loud music, and payment concerns. In this analysis, topics were more detailed. Staff-related complaints split into general rudeness and specific problems with management. Similarly, payment concerns were divided into subtopics such as double charges, cancellation difficulties, and missing discounts.

Additional themes also emerged. These included complaints about faulty or poorly maintained equipment, overcrowded gym spaces, and limited access to machines. These findings highlight how customer frustration is driven by both the service experience and the physical environment of the gyms.

3.4 Falcon-7B-Instruct Model

The Falcon model identified a mixture of recurring and new topics when compared to previous analyses. Running BERTopic on the LLM-generated outputs revealed a stronger focus on customer service-related issues. These were more specifically categorised, providing deeper insight into the nature of service complaints.

Familiar issues, such as broken machines, cold showers, class disruptions, and management behaviour, remained prominent. Their persistence across methods reinforces the need to address them as core concerns.

When prompted to generate suggestions for improvement, the Falcon model returned a diverse list of potential actions that PureGym could consider. Common suggestions included upgrading equipment, enhancing staff and customer service training, and implementing consistent cleaning and maintenance routines. These recommendations are intended to serve as inspiration rather than definitive solutions.

3.5 Gensim LDA Model

The LDA model revealed three main topic clusters (Appendix D). The first cluster focused on core gym-related issues, particularly the poor condition or limited availability of equipment. The second cluster related to problems with facilities and amenities, including broken air conditioning and malfunctioning showers. The third cluster captured more peripheral concerns, such as confusing parking policies and inadequate climate control.

These clusters closely resemble those found in the initial BERTopic analysis. However, they differ from the emotion-filtered results, where gym overcrowding, staff behaviour, and equipment issues were grouped together, and external factors like parking and membership payments were separated into a distinct cluster.

4 Insights and Recommendations

The analysis consistently highlighted four key areas of concern: cleanliness, unhelpful or rude staff, hidden or incorrect charges, and unresolved membership issues. Anger was the dominant emotion in negative reviews, particularly around poor customer service, equipment problems, and overcrowded gyms.

To address these issues, PureGym should prioritise staff and management training with a focus on professionalism and responsiveness. Upgrading and maintaining gym equipment, alongside improving the cleanliness of changing rooms and toilets, would directly target frequent complaints. Additionally, clearer communication around membership terms, cancellation policies, and day pass access would help reduce confusion and frustration.

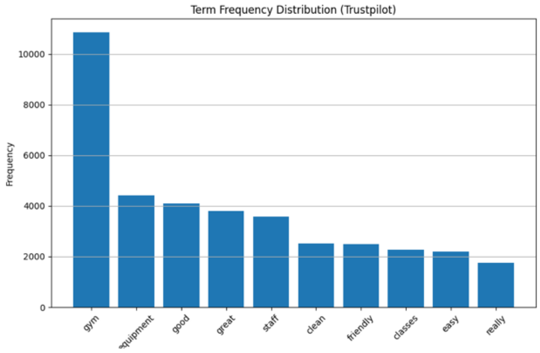
Combining traditional NLP methods with large language models like Falcon enabled a more nuanced and actionable understanding of customer feedback, supporting data-driven improvements across gym locations.

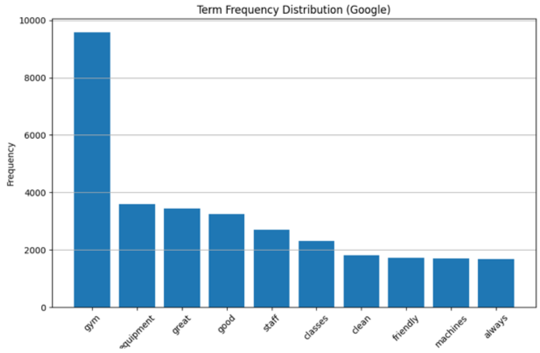
5 Conclusion

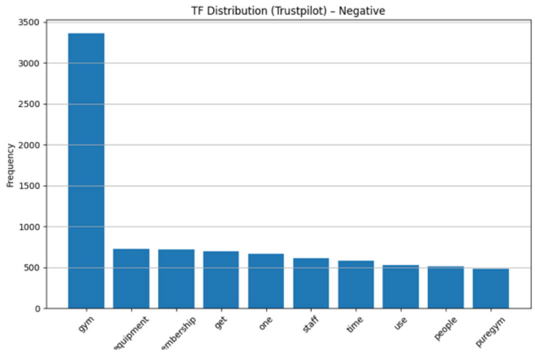
This analysis demonstrates how NLP can effectively uncover key drivers of customer dissatisfaction and generate actionable insights to support customer-focused improvements. By combining traditional techniques with large language models, the approach provides both depth and scalability. These findings can be applied across PureGym locations to systematically enhance the member experience and guide data-informed decisions for ongoing service improvement.

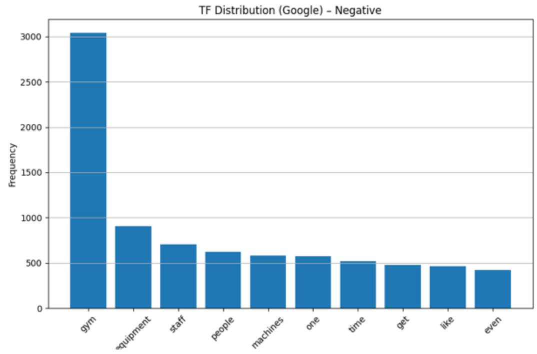
Appendix

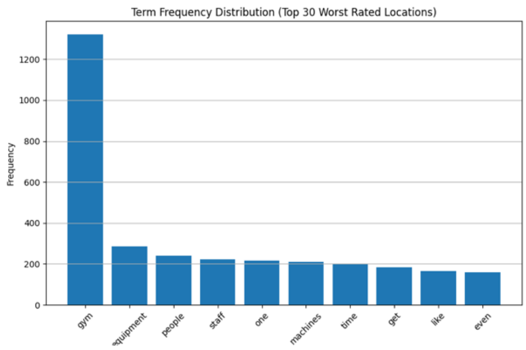
A – Frequency Distributions











B – Word Clouds

Google Reviews

Trustpilot Reviews

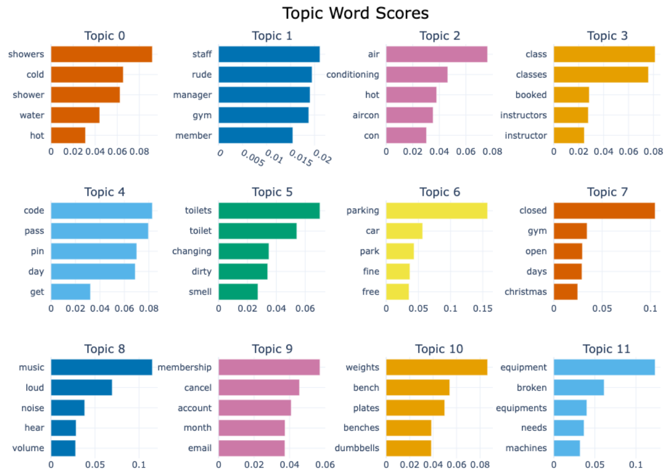
Trustpilot Negative Reviews

Google Negative Reviews

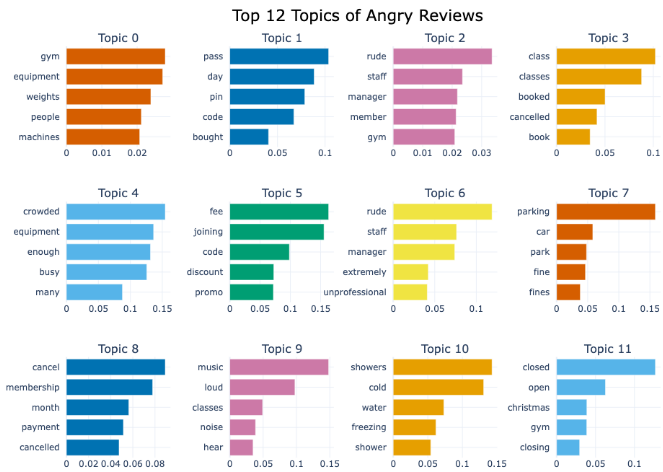
Top 30 Worst Rated Locations

C – Topic Word Scores

BERTopic

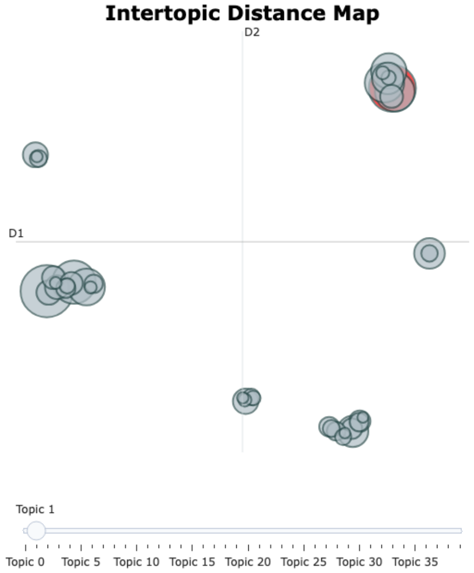


BERTopic top 30 worst reviewed locartions



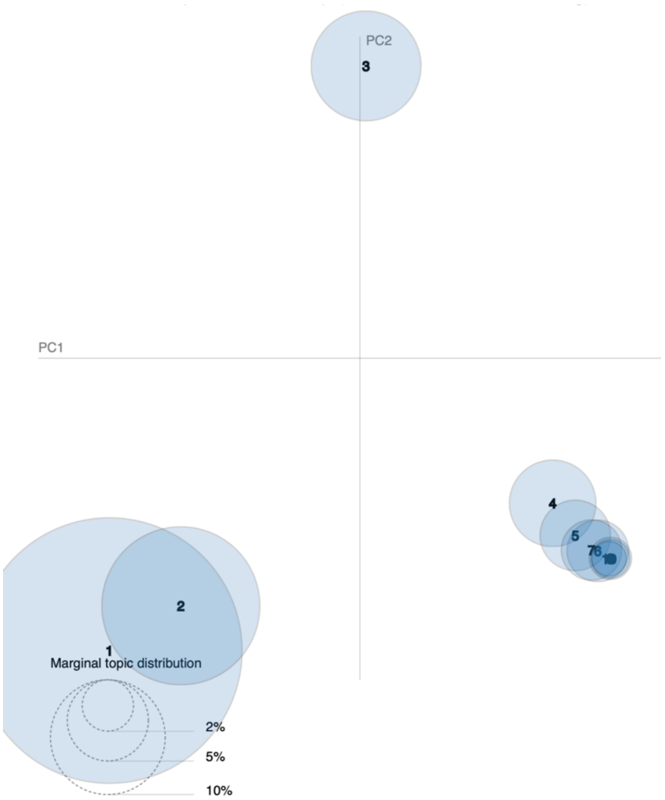
D – Intertopic Distance Maps

BERTopic



Emotion analysis + BERTopic

Gensim LDA Intertopic Distance Map



E – Emotion Distributions