Applying NLP for Topic Modeling

23rd Sep 2025
Vipin Badoni

Contents

1. Executive Summary

- **Data analysed:** 12 months of reviews from Google (**23,250**) and Trustpilot (**16,673**).
- **Techniques used:** Word frequency analysis, topic modelling (**BERTopic & LDA**), emotion detection, and generative language models.
- **Positive drivers:** Affordability, flexibility, accessibility.
- **Negative drivers:** Equipment maintenance, cleanliness, overcrowding at peak times, staff responsiveness.
- **Location insights:** Some gyms had consistently higher negative reviews, showing clear targets for operational improvement.

2. Business Context & Objectives

- **Founded in 2008:** PureGym has grown into one of the world's largest value fitness operators with 2 million members and 600+ gyms across the UK and Europe.
- **Business model:** Focus on affordability, flexibility, and 24/7 access → making high-quality fitness accessible to more people.
- **Mission:** "Inspiring a healthier world" through a customer-centric approach.
- **Why reviews matter:**
  - Provide direct feedback on what members value and where challenges exist.
  - Influence both brand reputation and member retention.
  - Act as a key source of business intelligence.
- **Project objective:** Analyse Google and Trustpilot reviews to uncover key themes, emotions, and drivers of dissatisfaction, and turn them into actionable recommendations to enhance member experience.

3. Data Overview & Preparation

The analysis used two datasets comprising 12 months of customer reviews:
- **Google Reviews:** 23,250 entries with gym location, review date, rating, and review text.
- **Trustpilot Reviews:** 16,673 entries with similar attributes.

Dataset	Reviews	Fields Used
Google	23,250	Location, Rating, Review text
Trustpilot	16,673	Location, Rating, Review text

Table 1: Data Overview

Number of Reviews by Platform

Figure 1: Number of reviews by platform

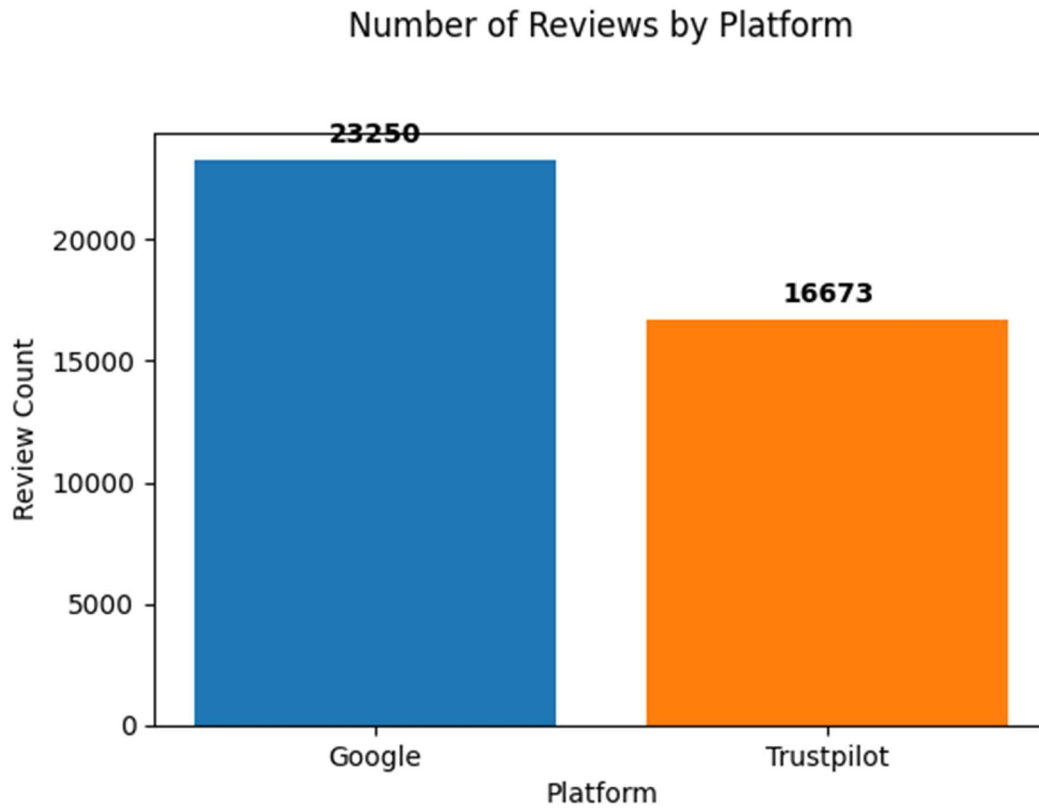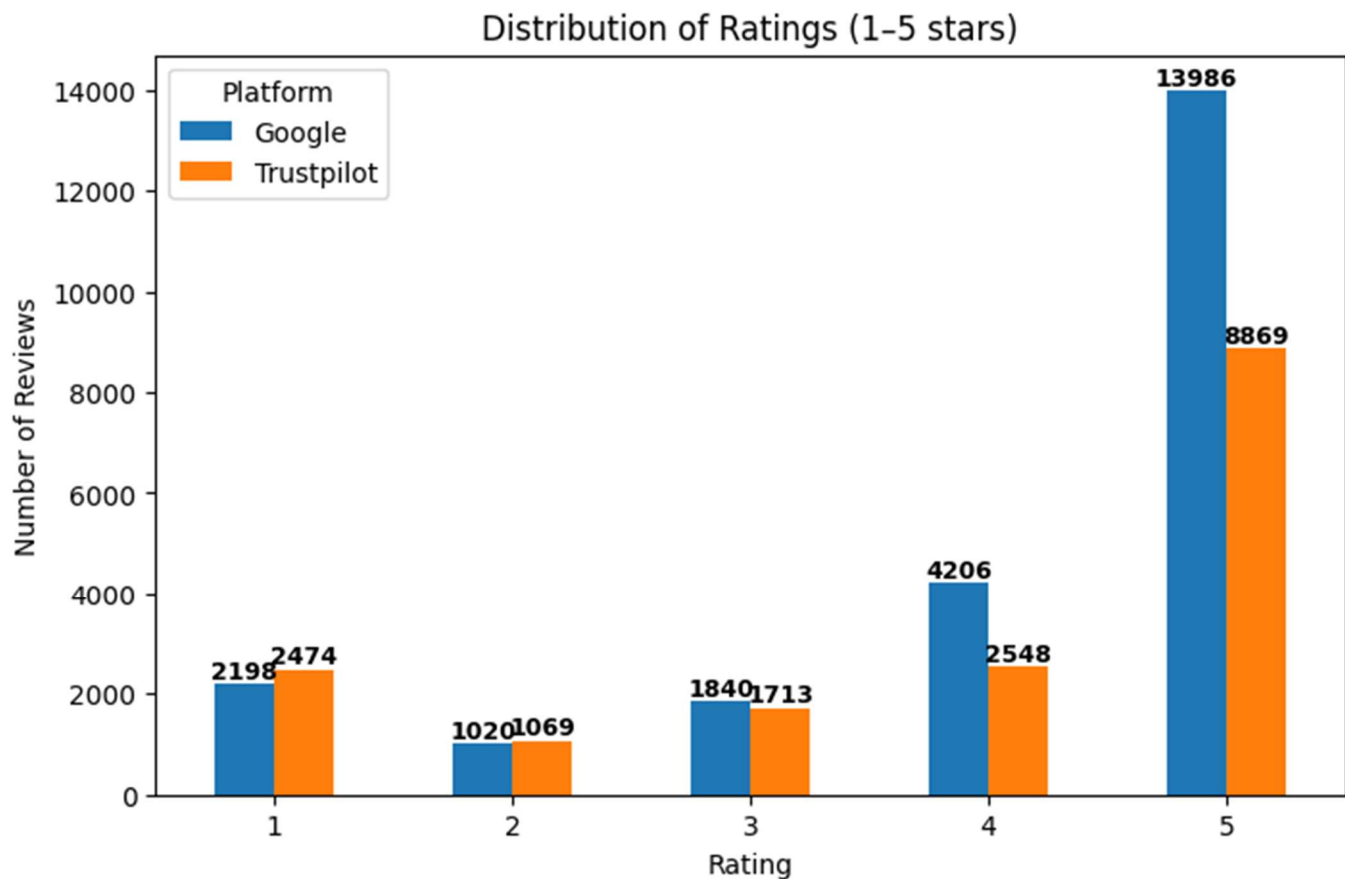## Distribution of Ratings (1–5 stars)

Figure 2: Distribution of rating by platform

4.Methodology

The analysis followed a structured methodology designed to uncover themes and insights from customer reviews. The process began with exploratory text analysis to identify common terms and patterns. This was followed by the application of multiple topic modelling approaches (BERTopic, LDA, and LLMs) to group reviews into coherent themes. Finally, the performance of these models was assessed to compare their strengths, limitations, and the unique insights each provided.

4.1   Exploratory Text Analysis

- For this project, only the key analytical fields—**location, rating, and review text**—were retained. Records with missing or empty comments and missing location were removed to ensure text quality.
- Standard preprocessing steps included **lowercasing, removal of punctuation and numbers, stopword filtering, and tokenisation**.
- To better understand drivers of dissatisfaction, negative reviews were isolated (scores below 3 stars).
- The final prepared dataset enabled word frequency analysis, topic modelling, sentiment/emotion detection, and location-based comparisons across both review platforms.
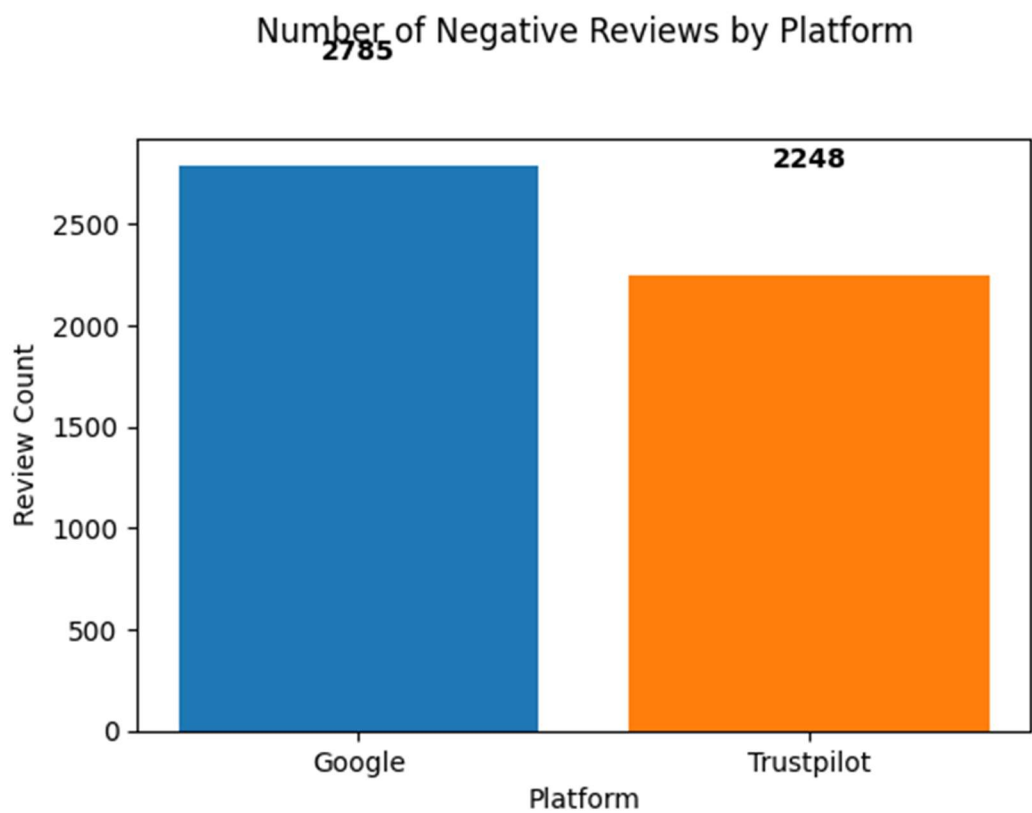

Figure 3: Total number of negative reviews by platform used in this project

The following figures show the 10 most common words used in negative reviews from both datasets.
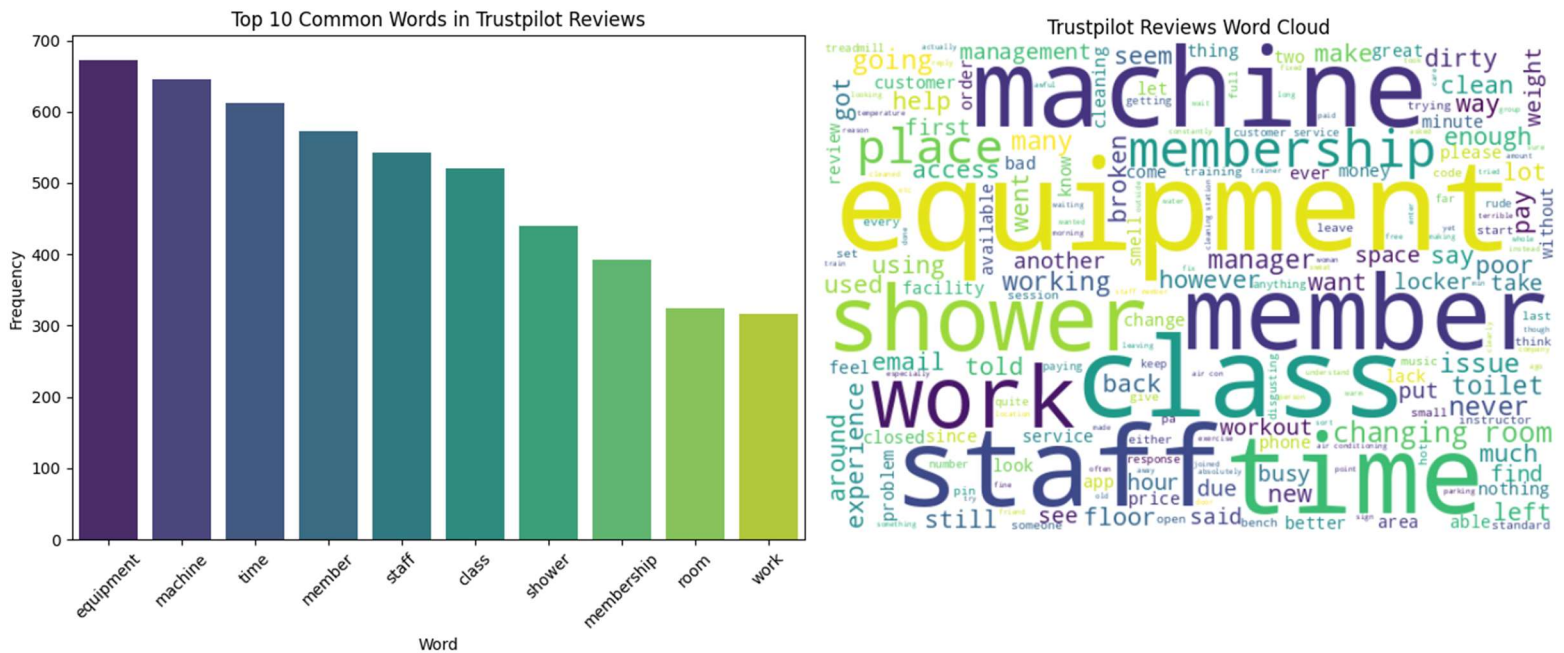

Figure 4: Top 10 most common words and word cloud from Trustpilot negative reviews
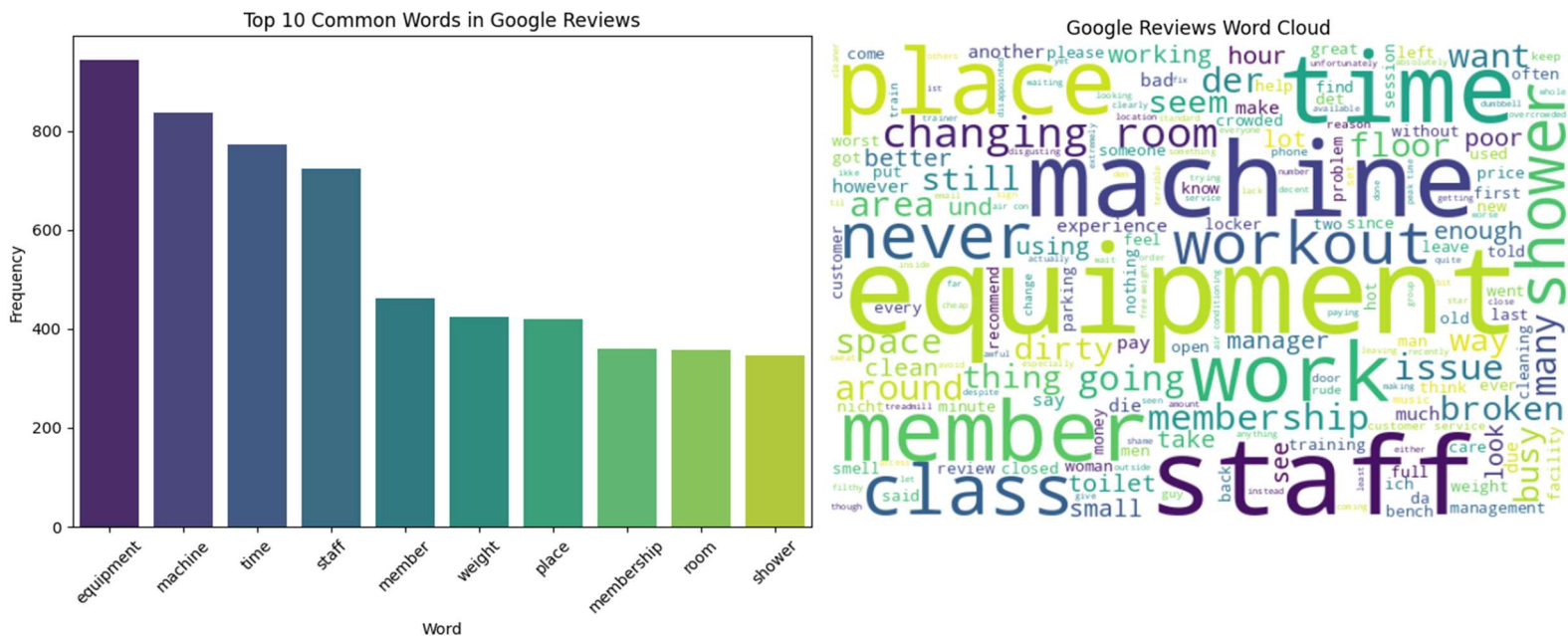

Figure 5: Top 10 most common words and word cloud from Google negative reviews

4.2   Modelling Approach

Topic modelling was performed using BERTopic, which grouped semantically similar reviews into coherent clusters. The analysis followed a structured sequence:

1) **Initial Run (BERTopic):** Topic modelling was first applied on the combined dataset of negative reviews to identify recurring themes.
2) **Location-Specific Analysis (BERTopic)**: To investigate further, BERTopic was re-run on reviews from the Top 30 common locations and results were compared with the initial run.
3) **Emotion-Analysis (BERTopic)**: The BERT emotion model was applied to classify reviews into categories such as anger, joy, and sadness. Clusters were re-analysed using only anger-driven reviews to highlight urgent issues.
4) **LLM Insights(Phi-4-mini-instruct)**: A generative language model (microsoft/Phi-4-mini-instruct) was prompted to extract key topics per review and generate actionable improvement suggestions, reframing operational clusters into strategic themes.
5) **Gensim**: To validate the robustness of BERTopic outputs, a parallel Latent Dirichlet Allocation (LDA) model was run using Gensim. LDA produced broader clusters (e.g., grouping cleanliness and air-conditioning together), while BERTopic offered more granular separation. The overlap between the two approaches added confidence in the identified themes.

4.3   Model Performance

To assess the performance of the modelling pipeline, each stage was evaluated for its **strengths, limitations, and unique insights**. As shown in Table 2, BERTopic provided granular clusters, location-specific analysis highlighted branch-level issues, emotion-aware modelling added urgency weighting, the LLM reframed insights into actionable recommendations, and LDA validated the robustness of topics.

Model	Purpose / Focus	Strengths	Limitations	Key New Insights
BERTopic - Initial Run	Broad clusters from all negative reviews	High-level overview of pain points	Some issues mixed together (e.g., cleanliness + AC)	10 broad clusters (equipment, cleanliness, staff, etc.)
BERTopic - Location-Specific	Focus on Top 30 location	Granular, location-level insights	Smaller data → narrower clusters	Staff rudeness, PIN failures, water machine complaints
Emotion-Aware (BERT)	Highlight emotion-driven issues	Prioritises urgent "anger" topics	Less granular than BERTopic	Anger linked to billing, PIN, staff
LLM (Phi-4)	Strategic reframing for management	Actionable, big-picture themes	Some redundancy, less detail	Safety, Communication, Digital Usability

LDA (Gensim)	Validate robustness	Confirms overall structure	Broad themes, low granularity	Reinforces core pain points (equipment, cleanliness, billing)

Table 2: Comparison of model performance

5. Findings & Insights

Analysis of negative reviews across multiple models (BERTopic, Emotion Analysis, LLMs, LDA) revealed consistent pain points, additional operational details, and broader themes for management action.

5.1 Core Pain Points (Consistently Observed Across Models)

- **Equipment availability & maintenance** – broken/missing machines, long waits, slow repairs.
- **Cleanliness & hygiene** – dirty toilets, showers, changing rooms, unpleasant odours.
- **Showers & facilities** – cold/broken showers, poor pressure, lack of maintenance.
- **Membership & billing** – cancellation issues, ongoing charges, poor email support.
- **Staff behaviour & service** – rudeness, lack of professionalism, dismissive attitude.
- **Safety & security** – locker thefts, faulty locks, unsafe facilities.

5.2 Additional Operational Issues (granular models)

- PIN/passcode access failures.
- Paid Yanga water machines and broken fountains.
- Loud music in classes and gym floors.
- Poor communication on closures/holiday schedules.

5.3 Emotional Triggers

- Anger was strongly linked to billing disputes, PIN failures, and staff behaviour — signalling high-priority risks to retention.

5.4 LLM

- Customer safety & security.
- Website and digital usability.
- Communication and transparency.

Overall Insights:

Core issues are robust across models. Addressing basics (cleanliness, equipment, showers) alongside service improvements (billing clarity, staff training, communication) offers the fastest path to better member experience.
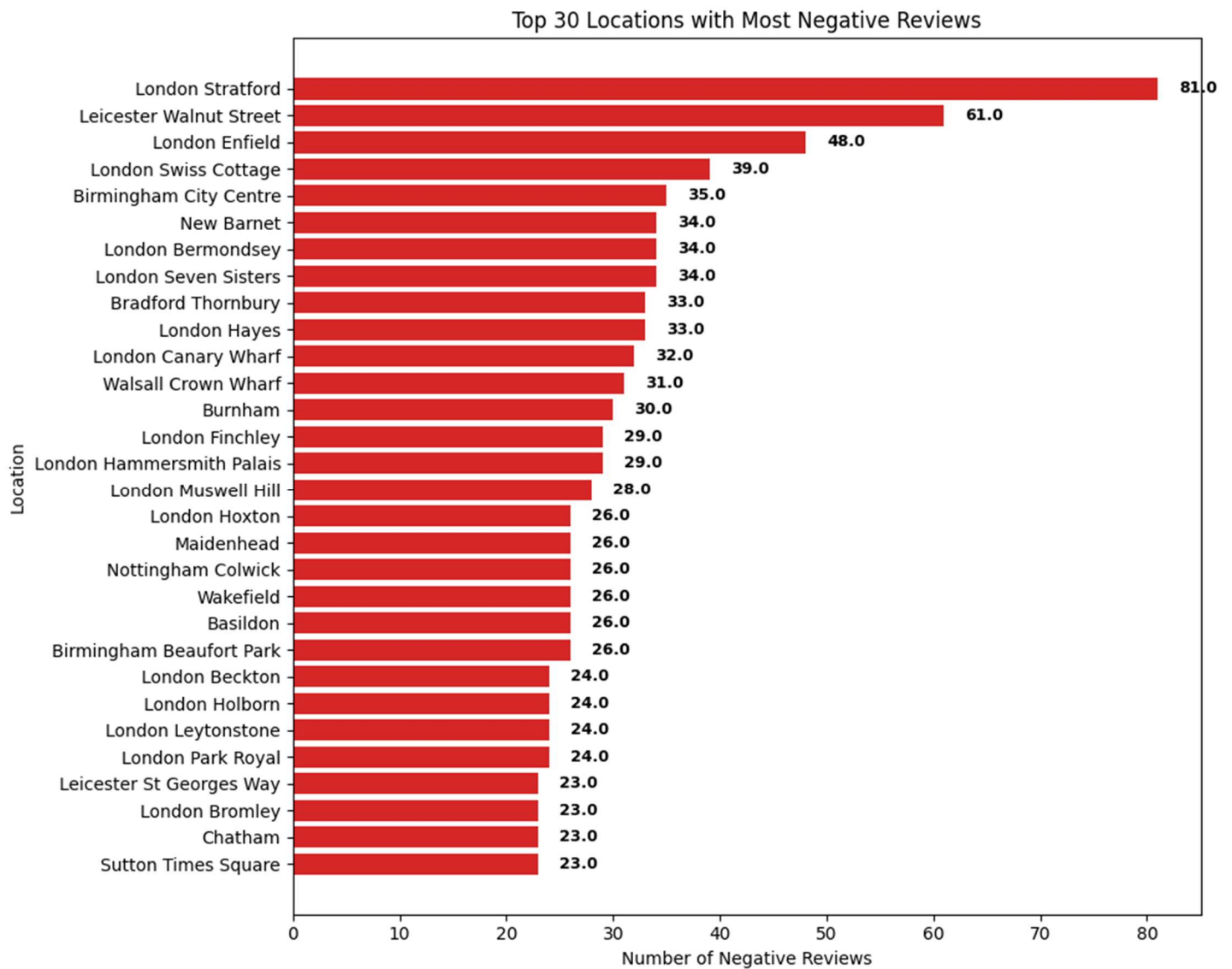
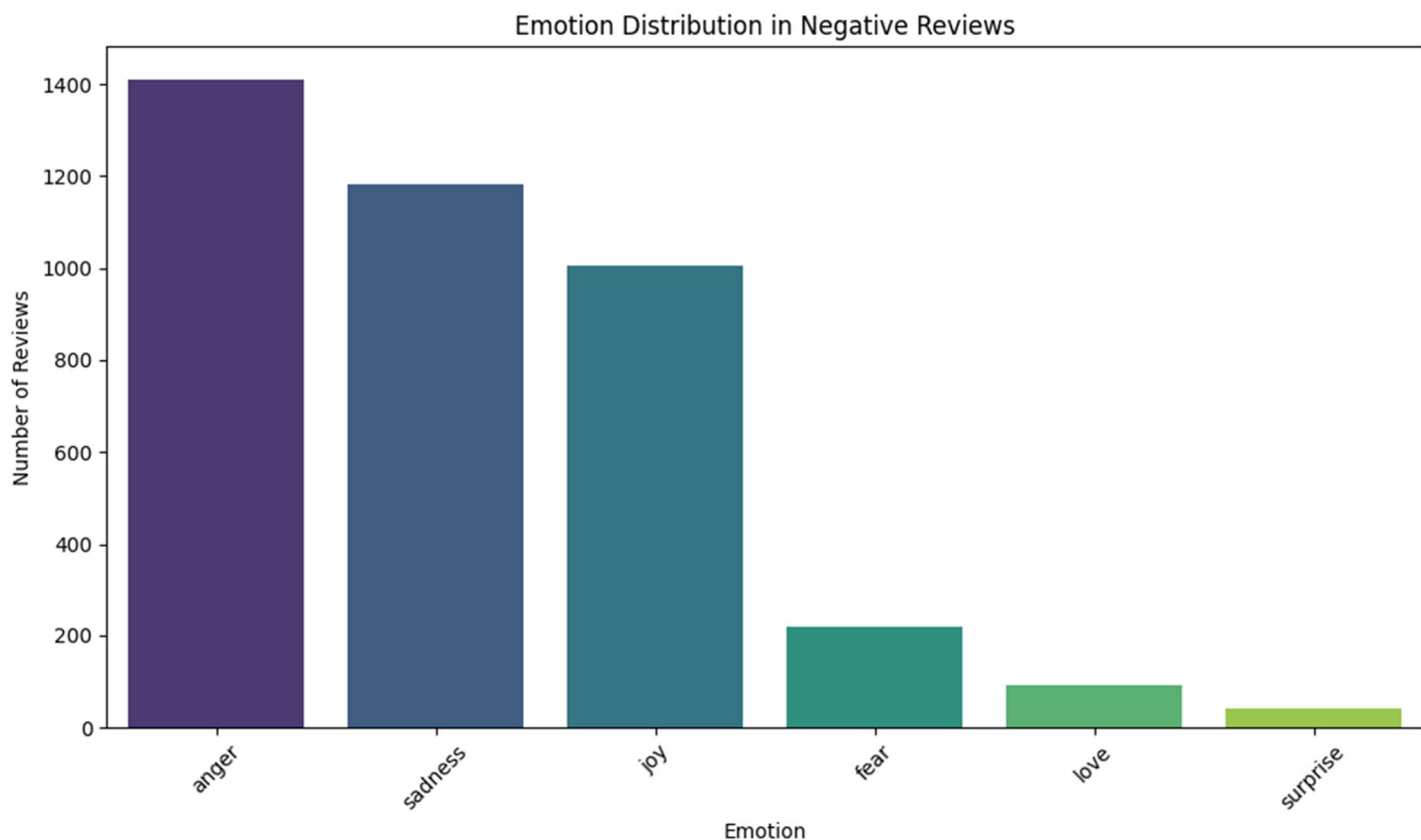*Figure 6: Top 30 Locations with most negative reviews*

Figure 7: Top Emotions in negative reviews

Findings and insights – summary:

Theme	Example Issues Raised by Customers
Equipment Availability & Maintenance	Broken machines, missing benches/dumbbells, long waits, slow repairs
Cleanliness & Hygiene	Dirty toilets/changing rooms, unpleasant odours, lack of cleaning supplies
Showers & Facilities	Cold/broken showers, poor water pressure, mould in shower areas
Membership & Billing	Difficulty cancelling, ongoing charges after cancellation, unresponsive email service
Staff Behaviour & Service	Rudeness, unprofessional behaviour, lack of accountability
Safety & Security	Locker thefts, broken locks, customers feeling unsafe
Access Failures (PIN/Passcodes)	PIN not working, customers locked out, poor support for access issues
Water Machine Complaints	Paid Yanga refills, broken fountains, time limits
Noise & Atmosphere	Music too loud in classes/floor, customers unable to focus
Closures & Communication	Gyms closed without notice, poor communication about holiday/maintenance closures

Table 3: Findings and insights – summary of themes and issues

6. Recommendations

Based on the analysis, recommendations are grouped into high-impact operational fixes, service improvements, safety/accessibility, and strategic priorities.

- **High Priority (Immediate Fixes):** Equipment & Maintenance, Cleanliness, Showers, Billing Transparency, Staff Training.
- **Medium Priority:** Locker Security, Access Reliability, Closure Communication, Water Machine Policy.
- **Low Effort / Quick Wins:** Noise Management, proactive hygiene checks.
- **Strategic Focus:** Customer Safety, Digital Experience, Communication & Transparency.

Area	Key Issues	Recommended Actions
Equipment & Maintenance	Broken machines, long waits	Preventive maintenance, faster repairs, add benches/dumbbells
Cleanliness & Hygiene	Toilets, showers, odours	Stronger cleaning schedules, hygiene stations, staff checks
Membership & Billing	Cancellations, ongoing charges	Simplify process, real-time billing in app
Staff Behaviour	Rudeness, poor service	Mandatory training, manager coaching, audits
Locker Security	Thefts, faulty locks	Upgrade lockers, increase CCTV
PIN / Access	Entry failures	Improve PIN system reliability, better onboarding
Water Machines	Paid refills, faults	Review policy, ensure working fountains
Noise Management	Loud music	Standardise levels, allow feedback
Closure Comms	No advance notice	Proactive updates via app, SMS, email
Strategic Themes	Safety, Digital, Communication	Position safety as core value, improve app/website, proactive comms

Table 4: Recommended actions by area

7. Conclusion

- **Advanced NLP methods** (BERTopic, Emotion Modelling, LLMs, LDA) transformed customer reviews into clear, actionable insights.
- **Recurring issues across models**: equipment, cleanliness, showers, membership/billing, and staff behaviour.
- **Granular models** revealed new details: PIN access failures, dissatisfaction with paid water machines.
- **Confidence strengthened** as different models consistently reinforced the same themes.
- **Balanced perspective**: BERTopic gave diagnostic detail, LLMs reframed insights into strategic priorities.
- **Business value**: Acting on hygiene, equipment reliability, staff professionalism, and communication can:
  - Improve customer satisfaction
  - Reduce churn
  - Protect PureGym's competitive reputation

Key Takeaway: Fixing operational basics while improving service quality offers the fastest path to boosting member experience.