APPLYING NLP FOR TOPIC MODELLING TO ANALYSE PUREGYM REVIEWS

Stakeholder Report

Prepared by Matt Cocker

Contents

- Introduction
- Data
- Method
- Results
- Conclusions
- References

Introduction

The PureGym Group offers high-quality, low-cost, and flexible fitness facilities at over 600 locations. PureGym needs to analyse its customer reviews to see what their key concerns are. That knowledge can then be leveraged to enable PureGym to implement targeted solutions to these issues, improving customer satisfaction, therefore reducing customer churn and attracting new customers.

Figure 1: *Project Flow Chart*

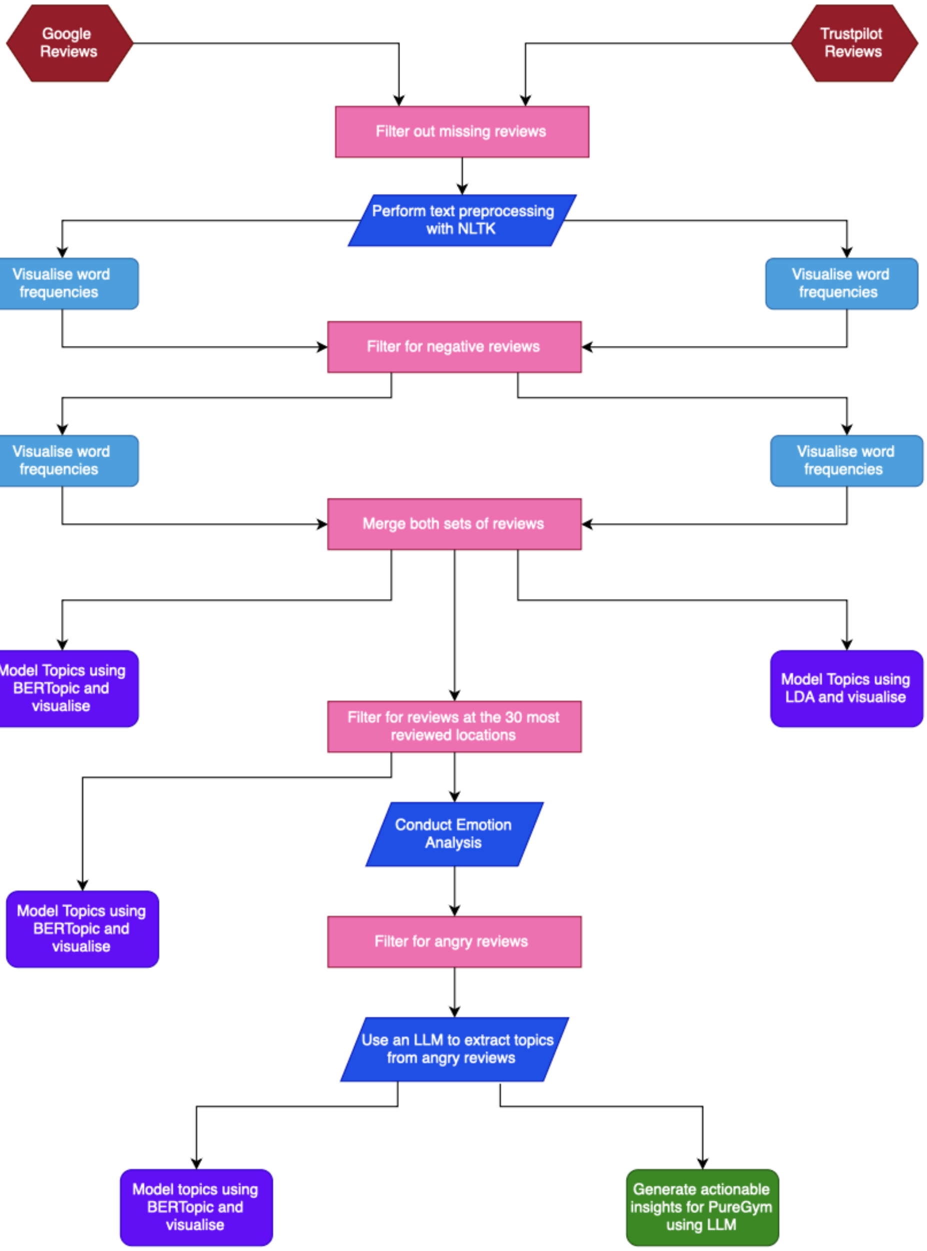

Figure 1 describes the process of analysing the reviews. Missing values were removed, and text preprocessing was conducted to ensure the reviews were in a format that models could easily derive conclusions from. Topic modelling was conducted on negative reviews for an initial view, then emotion analysis was done to identify angry reviews. An LLM was used to extract topics from these reviews and topic modelling was conducted, before an LLM was used to generate actionable insights for PureGym to address its customers' concerns.

Data

The initial data was obtained from two sources, Google reviews and Trustpilot reviews. After filtering out reviews with no text written, a total of 30,571 reviews were left, 13,898 in the Google dataset and 16,673 in the Trustpilot dataset.

Following an initial analysis of all reviews, the dataset was filtered for negative reviews (< 3 stars), leaving 2,453 Google reviews and 3,440 Trustpilot reviews. These were again filtered to only include reviews from locations common to both datasets, generating a consistent dataset with a good sample size (3,886 reviews). Topic modelling using BERTopic and LDA was conducted on these reviews.

This list was further filtered down to only include reviews from the top 30 most reviewed locations, leaving 990 reviews. Emotion analysis was conducted on these reviews, and the 420 angry reviews were filtered for topic modelling using LLMs and BERTopic, before an LLM was used to generate actionable insights based on the topics in the angry reviews.

This process ensured that the most relevant insights would be obtained, as the focus of the later analysis was on the angry reviews at the locations with the most negative reviews, which are a priority for PureGym to address.

Methods

Several methods were employed across this project to derive insights from these reviews:

- **Bag of Words:** Initial analysis used word clouds, a methodology used to represent sentiment in reviews by counting the most frequently used words.

- **BERTopic:** A model that uses sentence embeddings and clustering to group similar documents into topics. Unlike frequency-based models, it captures contextual meaning and produces topics from textual data.

- **Latent Dirichlet Allocation (LDA):** Like BERTopic, this is a model used to group words into topics and identify topics in documents. The model assigns these topics according to probability distributions.

- **Emotion Classification Model:** These models fine-tune transformer-based language models (such as BERT) to classify text according to a set of predefined emotional categories. They use contextual understanding to detect emotional tone, enabling the model to identify emotions in reviews.

- **Large Language Models (LLMs):** These are general purpose models trained on huge corpora of text, designed to understand and generate human-like text. They are adaptable for a wide range of natural language tasks.

Results

This section will be broken down into 3 stages. First will be the results of the negative reviews using Bag of Words. Second will be the results of LDA and BERTopic on negative reviews at common locations to provide a direct comparison. Finally, the results of running BERTopic on the list of topics generated by the LLM from the angry reviews will be displayed, and those will be compared to the previous results to show the improvement in insights.

Figure 2: *Word Cloud for Negative Google Reviews*

Figure 3: *Word Cloud for Negative Trustpilot Reviews*

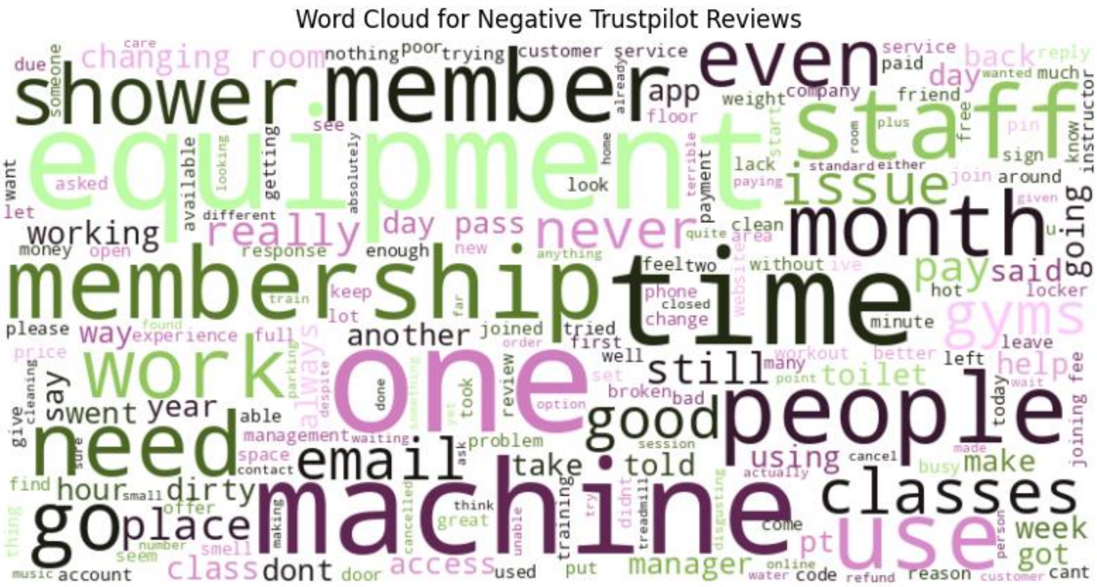

Figure 2 and *Figure 3* above show Word Clouds for the negative Google and Trustpilot reviews respectively. We can see that the most common words across both sets are equipment, machine, time, member and people. This can begin to give some insights into what customers' concerns are, but clearly individual words without context can only show so much.

Figure 4: *BERTopic Results on Negative Reviews*

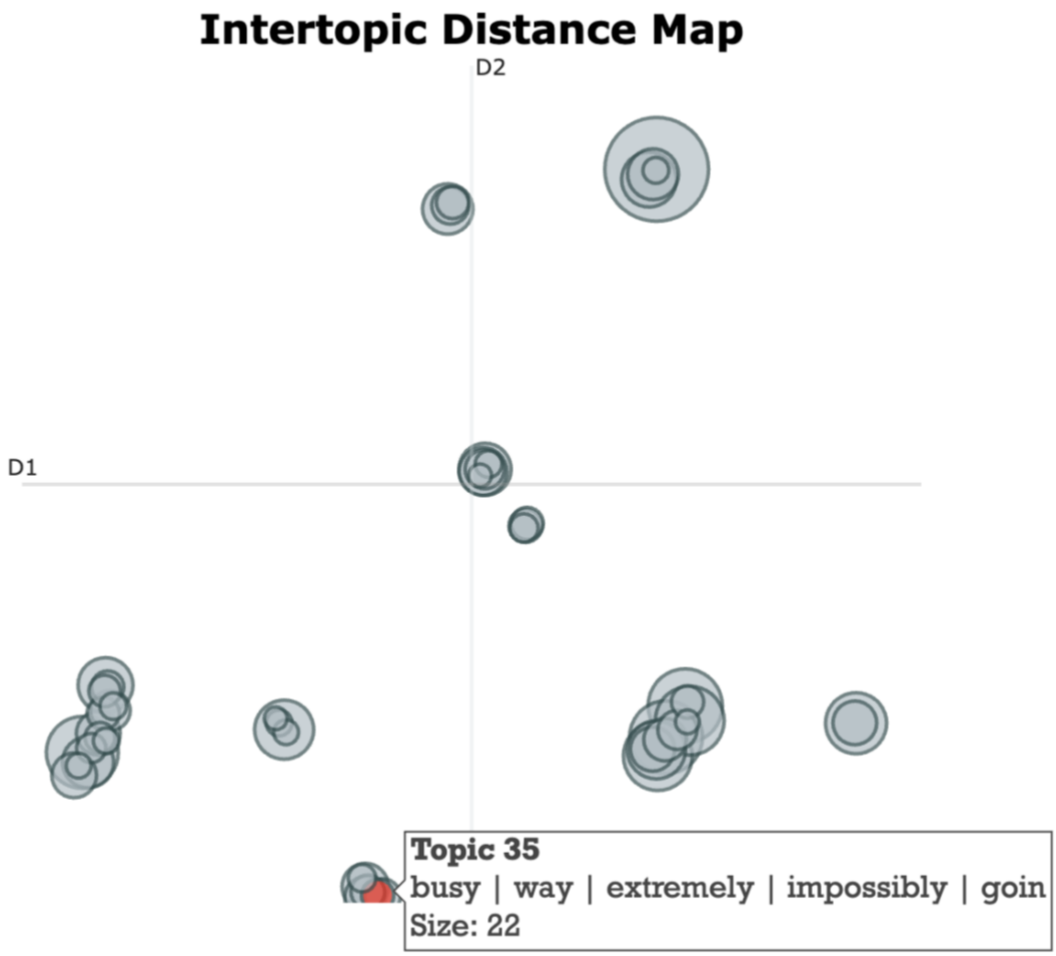

Figure 5: *LDA Results on Negative Reviews*

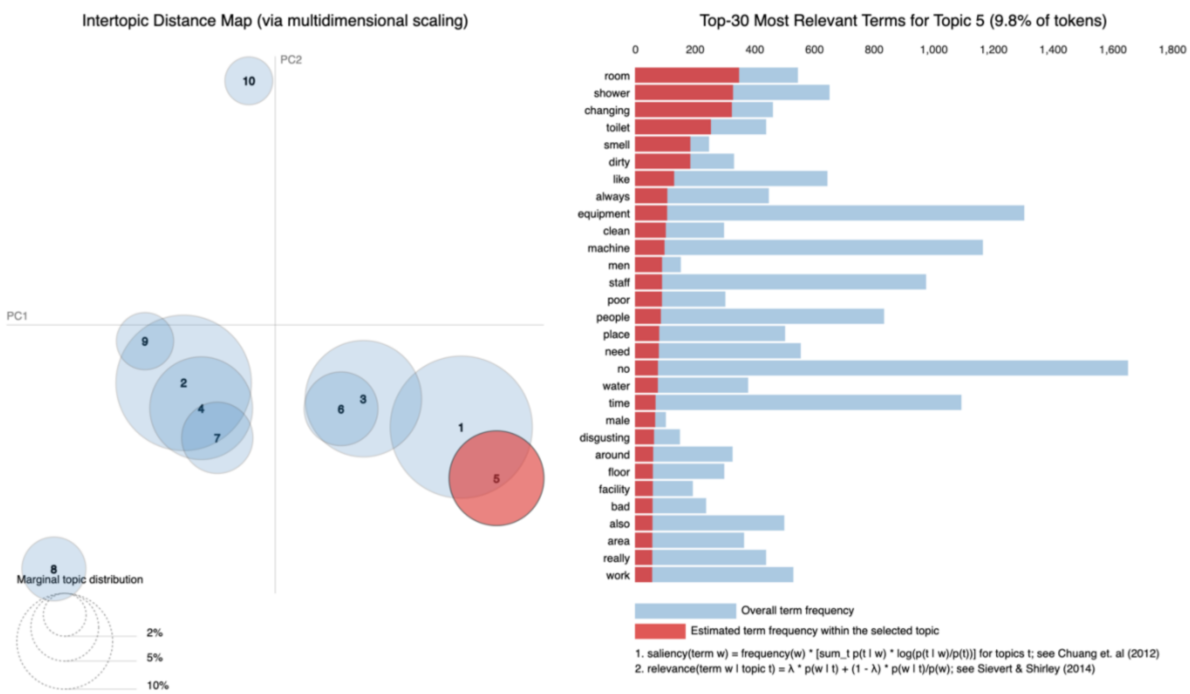

Figure 4 above shows the Intertopic Distance Map for the merged set of negative reviews with topics modelled using BERTopic. There are 9 distinct clusters of topics – with the one around Topic 35 themed around the gym being too busy, as shown. *Figure 5* shows the Intertopic Distance Map and a bar chart of the Top 30 most relevant terms for the selected topic (Topic 5 in this case) for the same set of negative reviews, with topics modelled using LDA.

There are clear differences between these two methods – BERTopic generates more distinct clusters, albeit some are very small, but this ought to give more clarity to separate themes of complaints. However, this is not always the case, with some small clusters being a significant distance apart when they could be grouped together. This is not an issue with LDA, however the lower number of topics leads to some generic clusters – although this could be fixed by tuning the model to split the reviews among more topics.

Figure 6: *BERTopic Results on Angry Reviews at the Top 30 Locations*

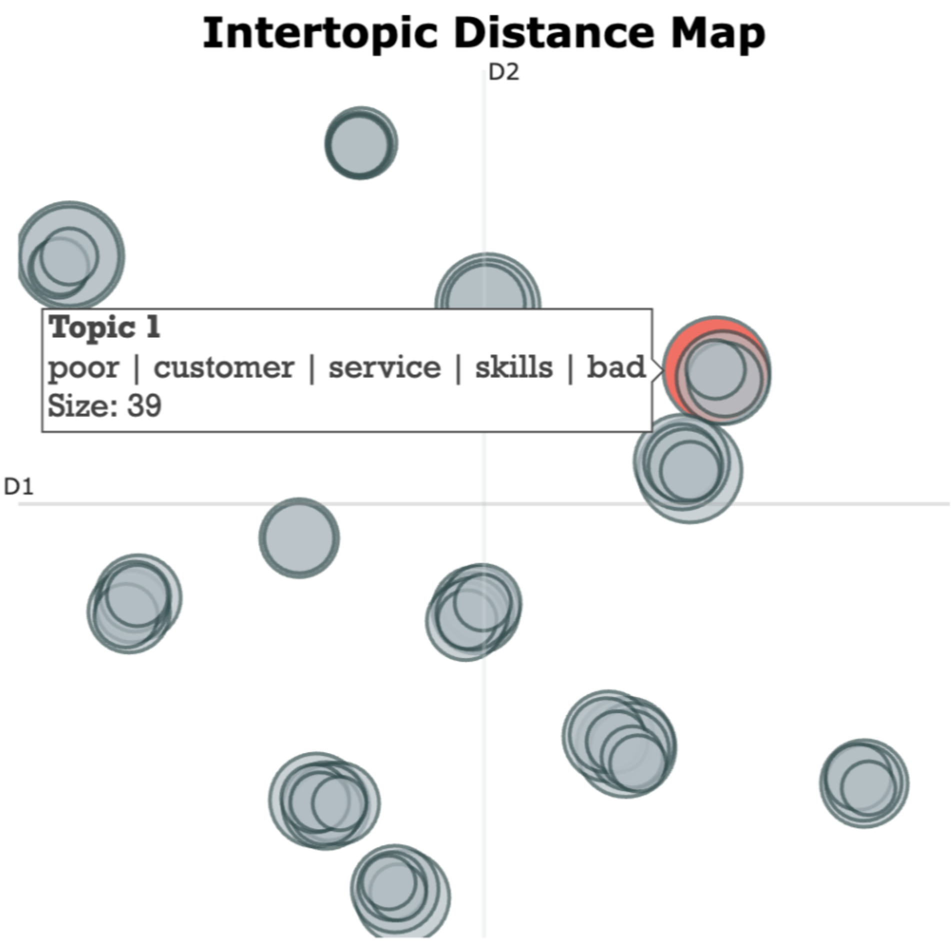

Figure 6 above shows the Intertopic Distance Map for the set of angry reviews at the 30 locations with the most negative reviews across the two datasets, with topics modelled using BERTopic. There are clear, distinct clusters and the distances make sense, with the two clusters closest to the Topic 1 cluster also having themes of poor customer service interactions. It is worth noting that the model was fine-tuned to get this visualisation, as opposed to the *Figure 4* model, which was not.

Conclusions & Next Steps

Through a combination of topic modelling and LLM analysis, we can conclude that angry customers' main concerns revolve around staff attitude, equipment maintenance and cleanliness of facilities. The focus on angry customers at the locations with the most negative reviews allows us to focus on areas which are a priority for PureGym. This brief would recommend PureGym take the following actions to improve its offering to customers:

1. **Facility Cleanliness:** Establish a strict cleaning schedule and assign specific staff members to responsibility, ensuring that the gym remains clean and hygienic at all times.

2. **Staff Training:** The gym should invest in ongoing staff training programs focused on customer service to ensure all employees are equipped to provide a positive experience and are knowledgeable about equipment usage.

3. **Equipment Maintenance:** Implement a regular maintenance schedule and routine checks for all equipment, reducing the risk of malfunctions and ensuring a safe workout environment for members.

References

Praveen S.V. et al. (2024), Crafting clarity: Leveraging large language models to decode consumer reviews., Journal of Retailing and Consumer Services., doi:10.1016/j.jretconser.2024.103975.