

GYM REVIEW ANALYSIS



THE BODY ACHIEVES
WHAT THE MIND
BELIEVES

CONTEXT ANSWER THE ‘W’

Where the problem stems from?

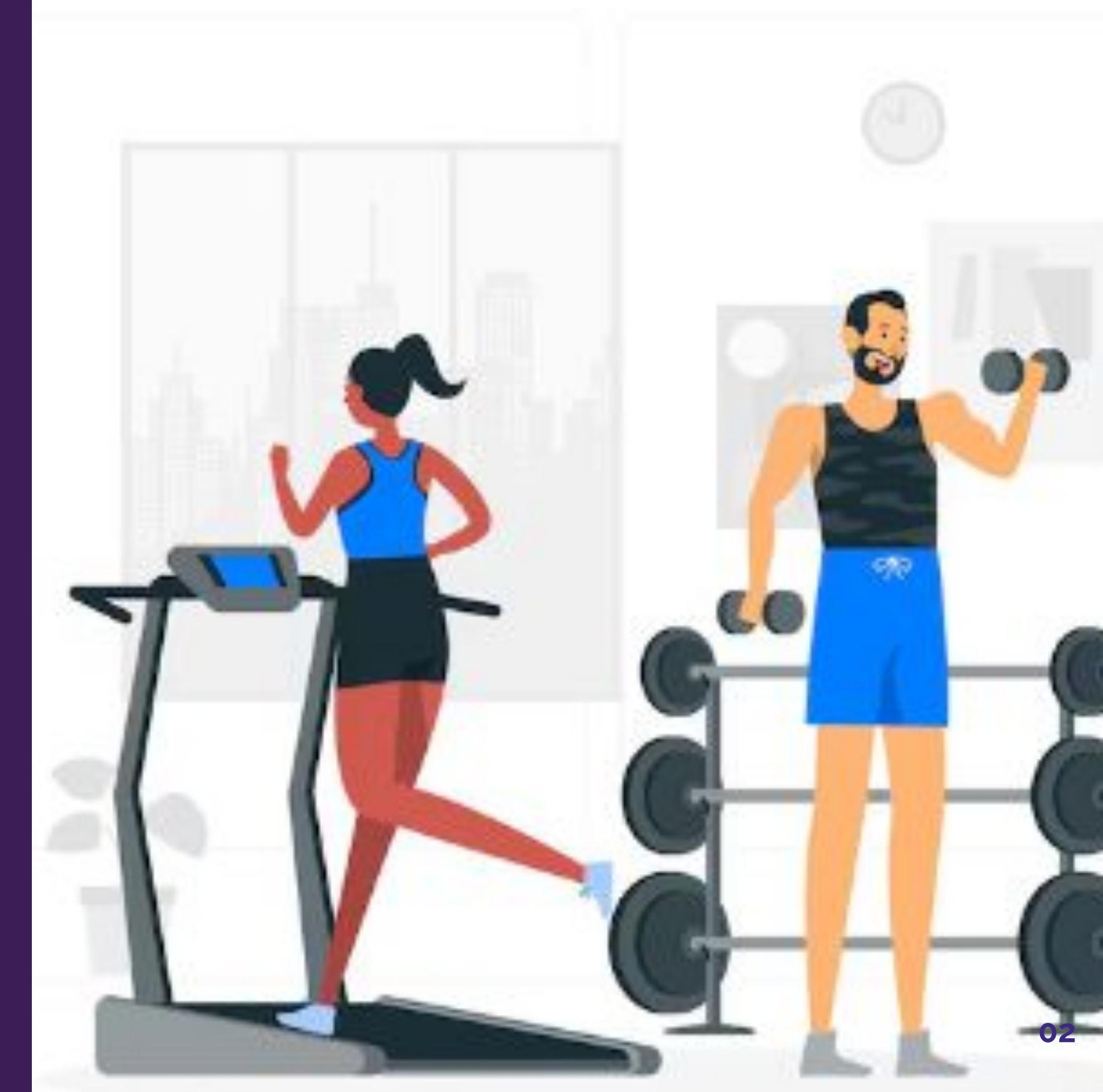
- *Online reviews shape a gym's reputation and customer decisions.*

What is the problem?

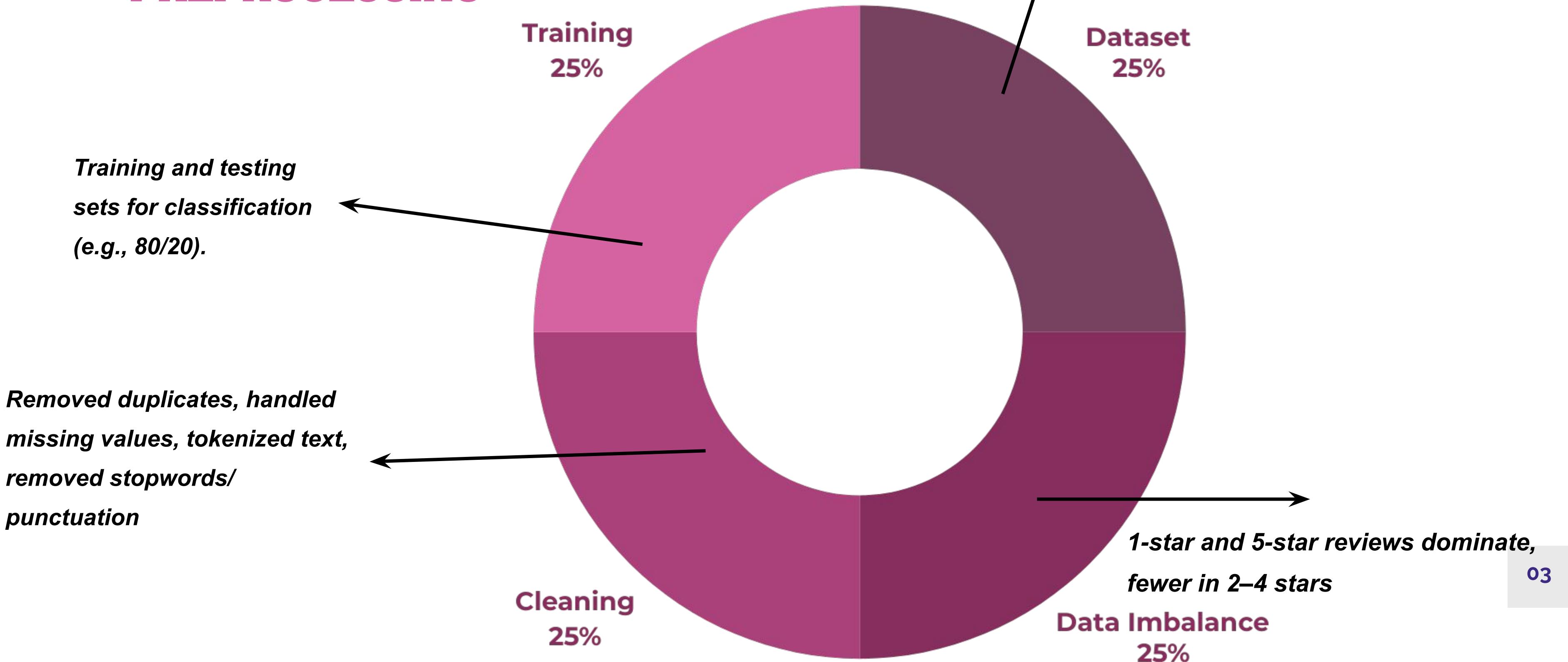
- *Identify factors that drive **1-star vs. 5-star** ratings and predict review sentiment.*

How do we approach the problem?

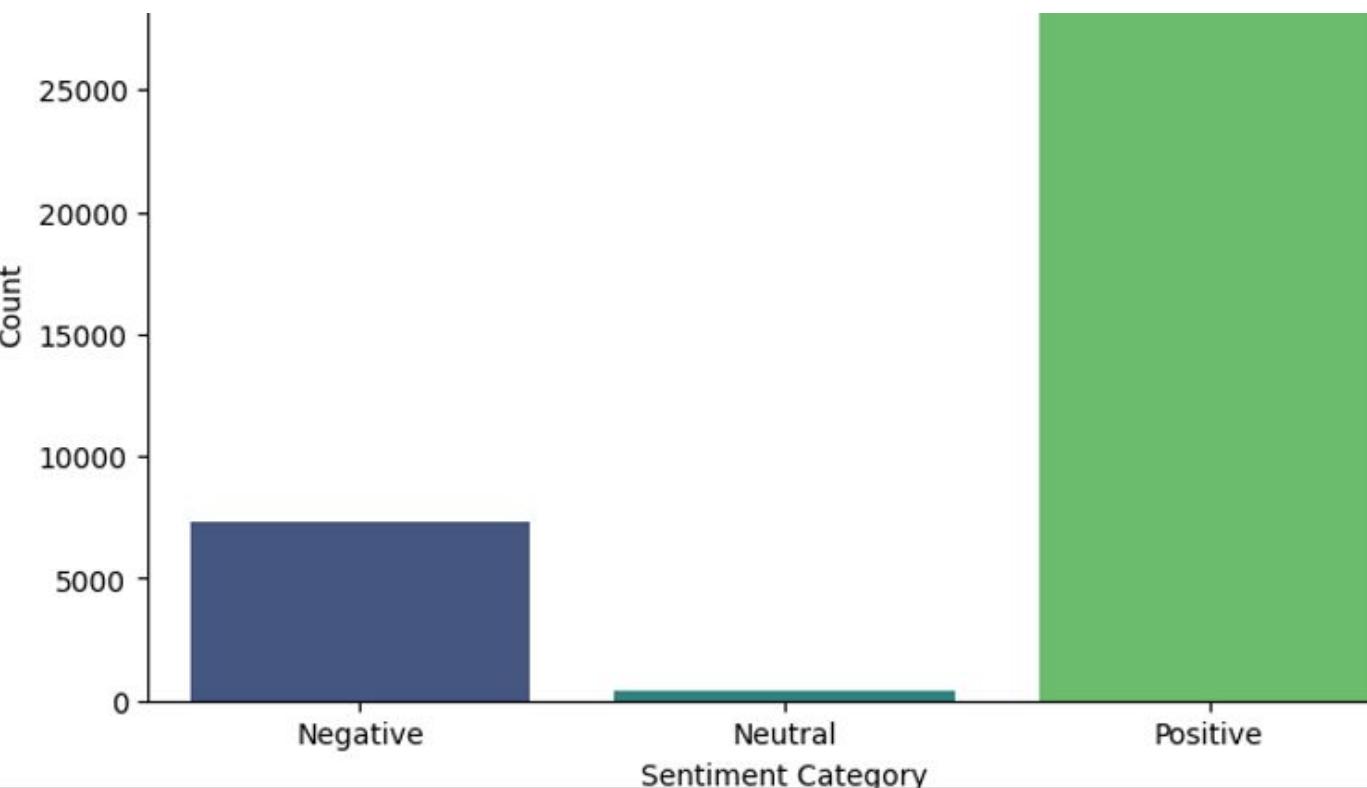
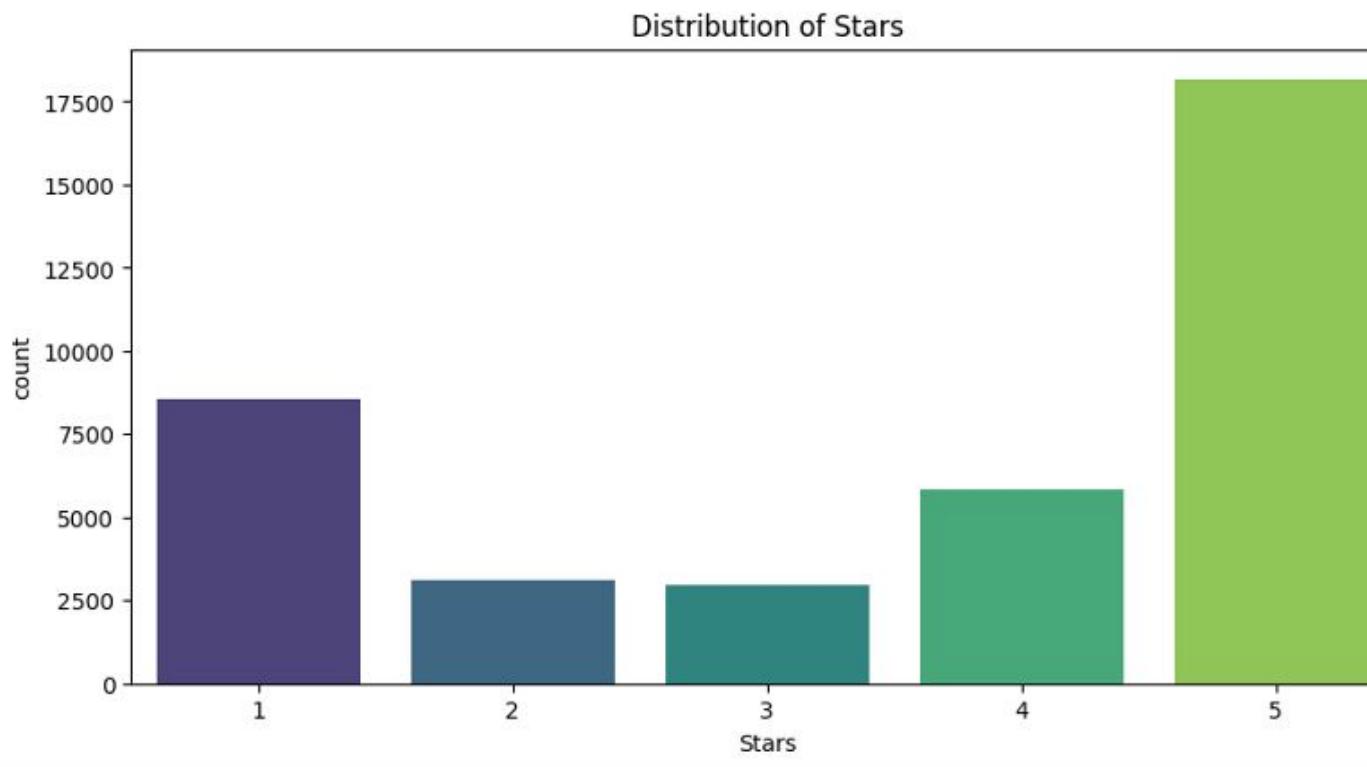
- *By applying Text analytics, word clouds, topic modeling (LDA), and classification (Logistic Regression, Random Forest, XGBoost*



DATA OVERVIEW & PREPROCESSING



EXPLORATORY DATA ANALYSIS



KEY TAKEAWAYS

Star Rating Distribution: Large proportion of 5-star reviews, significant 1-star chunk.

Review Length: Most are relatively short; a few are very long.

Initial Sentiment: Majority positive, fewer negative, small neutral.

LDA Topic Exploration: Tested 2–9 topics; coherence scores suggest fewer topics may capture main themes.

TEXTUAL ANALYSIS

KEY TAKEAWAYS

- 1-Star reviews center on billing issues, staff problems, facility concerns.
- 5-Star reviews highlight positive staff interactions, cleanliness, and variety of classes.
- LDA reaffirms core themes (equipment, staff, membership, etc.)



CLASSIFICATION ANALYSIS

(LOGISTIC, RF, XGBOOST)



| Model | Accuracy | Precision | Recall | F1 Score |
|----------------------------|-------------|-------------|-------------|-------------|
| <i>Logistic Regression</i> | 0.91 | 0.83 | 0.94 | 0.94 |
| Random Forest | 0.89 | 0.85 | 0.81 | 0.85 |
| XG Boost | 0.89 | 0.87 | 0.85 | 0.86 |

KEY TAKEAWAY

Logistic Regression achieves the highest overall accuracy (~92%),
though all models perform reasonably well on the Positive class.

ACTIONABLE INSIGHTS FOR GYM OWNERS

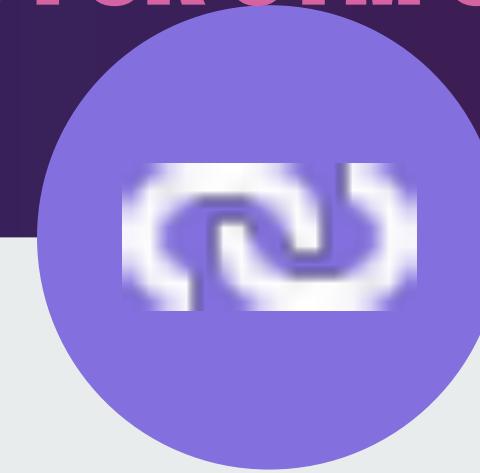


Negative Drivers (1-Star)

Billing & Contracts: Cancellation difficulties, hidden fees.

Staff & Service: Rude or unhelpful interactions.

Facility: Outdated equipment, cleanliness issues.



Positive Drivers (5-Star)

Friendly Staff: Helpful, welcoming environment.

Clean, Well-Maintained Facilities: Modern equipment, tidy areas.

Class Variety: Diverse workouts, training sessions, group classes



Recommendations

Provide transparent contracts, simplify cancellations.

Invest in staff training and customer service.

Keep facilities clean and up-to-date.

CONCLUSION & FUTURE WORK

Summary

1-Star reviews often cite contract/staff/equipment issues.

5-Star reviews emphasize positive staff experiences, cleanliness, and class variety.

Logistic Regression is the top performer (92% accuracy), with RF and XGBoost close behind.

Future Directions

Data Balancing: Address star rating imbalance (especially 2–4 stars).

More Topics or Larger Corpus: Refine LDA or try advanced topic modeling for deeper insights.

Real-Time Feedback: Deploy sentiment classification to track reviews continuously.

Closing

Focus on billing transparency, customer service, and facility upgrades for the biggest impact.

TEAM 110

RYAN KARIM,

SAGAR PASAD

ANWESHA BANERJEE

ZIFAN WANG

KEVIN MAREPAL

WWW.REALLYGREATSITE.COM



THANK YOU!

