

# What Do Consumers Want In A Boardgame?

GameStat Visionaries

Andrew Kulp, Talisa Manker, and Garrett Zimmerman

# Problem Statement

## Context

- Boardgames are expensive to produce
- To make a profit, we need to produce games consumers want to own

## Questions to Answer

- Can we use consumer game ratings to predict ownership?
- What features do games with high consumer ratings have?
- Can we predict if a game will be highly rated based on its features?



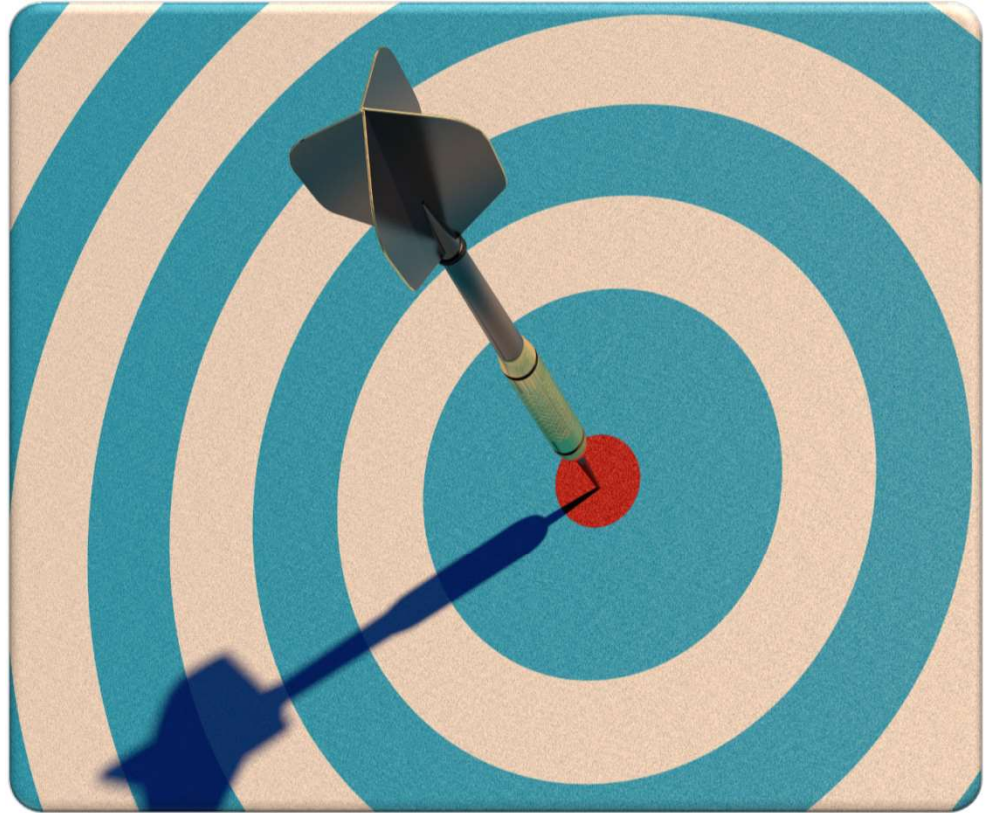
# Objectives and Goals

- **Project Objective**

- Determine the features that will make a newly designed board game highly rated and increase ownership by using data on existing board games

- **Expected Outcomes**

- Determine the features of a Board Game that will make the game popular and increase ownership
- Determine a highly rated value which correlates with high ownership
- Use Linear Regression to relate higher ratings with number owned
- Use Logistic Regression to classify if the newly designed board game will be highly rated based on already existing board games







# Data Sources

- **Data Overview**

- 9 Data Sources
  - All Data sources are linked by variable BGGId; the Board game ID
- Information Contained in Data Sources
  - Reviews (User Ratings, Bayes Ratings, Rankings)
  - Number Owned, Wanted, or Wished
  - Categories and Themes of Games
  - Artists and Publishers of Games

- **Data Relevance**

- Prediction: Data will help predict which features of a board game will lead to higher ownership of a newly designed board game
- Design: Categories and themes will guide the design of a board game which tend to have higher user ratings

# Methodology

- Determine threshold for high rating
  - What minimum rating would be considered “highly rated” based on business goals?
  - Explore if there is a rating threshold with an observable spike in ownership
- Variable selection
  - Several datasets are exclusively binary variables with as few as 11 variables to over 1,000
  - Use regularization techniques, such as LASSO regression, to select which game features to use in final model
- Predict probability of being highly rated
  - Use logistic regression to measure each feature’s impact on the likelihood of a board game being highly rated



# Project Plan and Timeline

## • Milestones

- Progress Update: November 4, 2024
- Mid Semester Update: November 25, 2024
- Final Presentation: December 8, 2024

Key	
Andrew	
Talisa	
Garrett	
Team	

Task	Oct 6- Oct 12	Oct 13 - Oct 19	Oct 20 - Oct 26	Oct 27 - Nov 2	Nov 3 - Nov 9	Nov 10 - Nov 16	Nov 17 - Nov 23	Nov 24 - Nov 30	Dec 1 - Dec 7	8-Dec
EDA										
Potential Threshold										
Address Challenges										
Linear Regression Model										
Log Model Threshold										
Variable Selection										
Choose Rating Thresholds										
Write Progress Update										
Edit Progress Update										
Submit Progress Update					Due Nov 4					
Mid Semester Update							Due Nov 25			
Final Variable Selection										
Final Log Model Threshold										
Final Linear Model										
Work on Slides for Final Pres										
Final Pres Submission										

# Expected Challenges

## Potential Obstacles

1. Selecting binary variables to use
2. Deciding on which rating variable to use
3. Choosing a probability threshold cut off for highly rated games
4. Using multiple features when relating rating and number owned could decrease the correlation

## Mitigation Strategies

1. Robust variable selection using multiple methods
2. Checking correlation of rating variables with number owned
3. Make multiple logistic regressions with different thresholds and look at evaluation metrics (sensitivity, specificity, etc.)
4. Using validation methods like k-fold validation



# Value Proposition

- Business Value
  - Intelligence on which features make a board game highly rated
  - A blueprint of what the next manufactured board game should include to drive ownership
  - A reputation for taking customer trends into account when designing games
- Consumer Value
  - Confidence in their purchase when buying our games
  - In the future, can adjust our model to create a recommender system based on customer preferences





# Team Roles and Responsibilities

## Andrew

- Determine Log Thresholds
- Midsemester Update Slides
- Create Final Log Model
- Work on Final Presentation Slides

## Talisa

- Variable Selection
- Write and Submit Progress Update
- Midterm Update Slides
- Select Variables for final logistic and linear regression
- Work on Final Presentation Slides

## Garrett

- Linear Regression Model
- Edit Progress Update
- Midterm Update Slides
- Create Final Linear Model
- Work on Final Presentation Slides
- Submit Final Presentation