03.17.2025

CSC 466
Analysis of Board Games

By Lucas Summers, Braeden Alonge, Nathan Lim, and Xiuyuan Qiu

Agenda

1. The Dataset
2. Collaborative Filtering
3. Clustering
4. Regression
5. Conclusion

Part 1
The Dataset

Overview

- Data source: BoardGameGeek (online forum & community)
- Covers ~22,000 games, ~411,000 unique users, ~19M user ratings
- Split into 9 different .csv files, covering a total of almost 100 features for the games, game designers, users, etc.

boardgamegeek.com

BGG Browse Forums GeekLists Shopping Community Help Sign In Join (it's free)! Search

Explore Dashboard

Designer Diary: Takeover

by Ron Sierra • BoardGameGeek News

BGG.CON 2025 Registration Opens March 14th

by CaptainQwyx • BGG.CON

Top 10 90 Minute Games

by boardgamegeektv

Designer Diary: 23 Knives

by T. Brown • BoardGameGeek News

Cities - GameNight!

by heccubus

THE HOTNESS

SEE ALL ›

Top 50 trending games today

The Tables

- **GAMES**: core information (47 features) about each game
- **USER_RATINGS**: individual ratings (~19M total) from ~411K users
- **RATINGS_DISTRIBUTION**: rating counts (0.0 - 10.0) for each game
- **THEMES**: game themes represented by binary flags
- **MECHANICS**: game mechanics (e.g. dice rolling) represented by binary flags
- **SUBCATEGORIES**: game subcategories (e.g. fantasy) represented by binary flags
- **ARTISTS_REDUCED, DESIGNERS_REDUCED, PUBLISHERS_REDUCED**: information about artists, designers, and publishers of the games

Research Questions

1. Which attributes of a game tend to produce the highest user ratings?
2. Can we cluster board games into meaningful categories based on mechanics, themes, and categories?
3. How effectively can we predict a “value” metric (e.g., average rating) for a board game from its attributes?
4. Given a user's past game ratings, can we predict their rating of new games based on ratings of similar users?

Part 2

Collaborative Filtering

Data Preprocessing

400k users x 22k games
= 9B cells

Filter:
- Most popular 1% of games (219)
- Users w/50+ ratings for this 1%

35k users x 219 games
= 7.5M cells

Train on this!

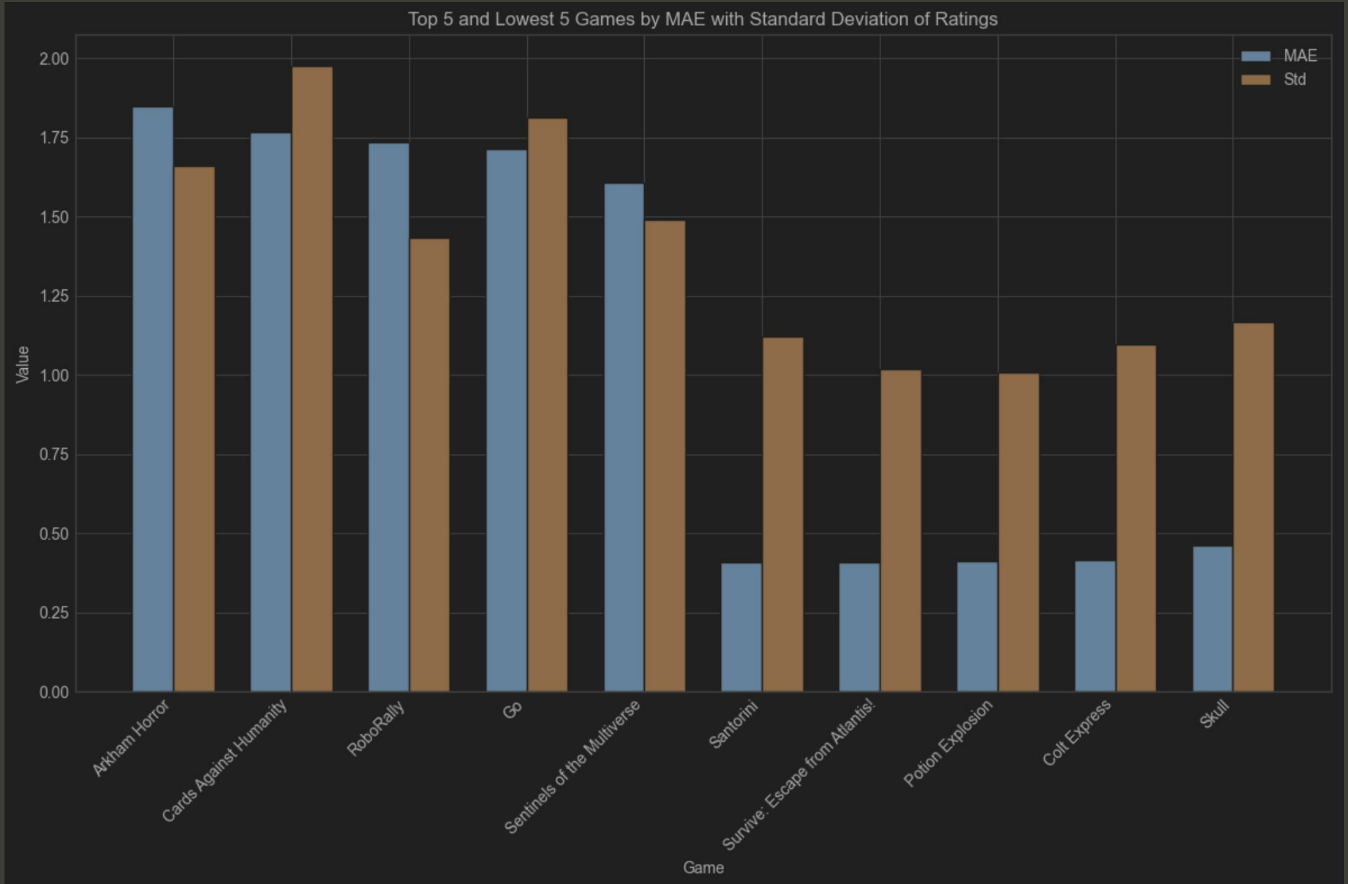
Methods

- Reused much of our implementations from lab 5 (CF lab)
  - KNN w/adjusted weighted sum + cosine similarity
- Mean Absolute Error with randomly selected sampling and tests
- Examined results overall and by game

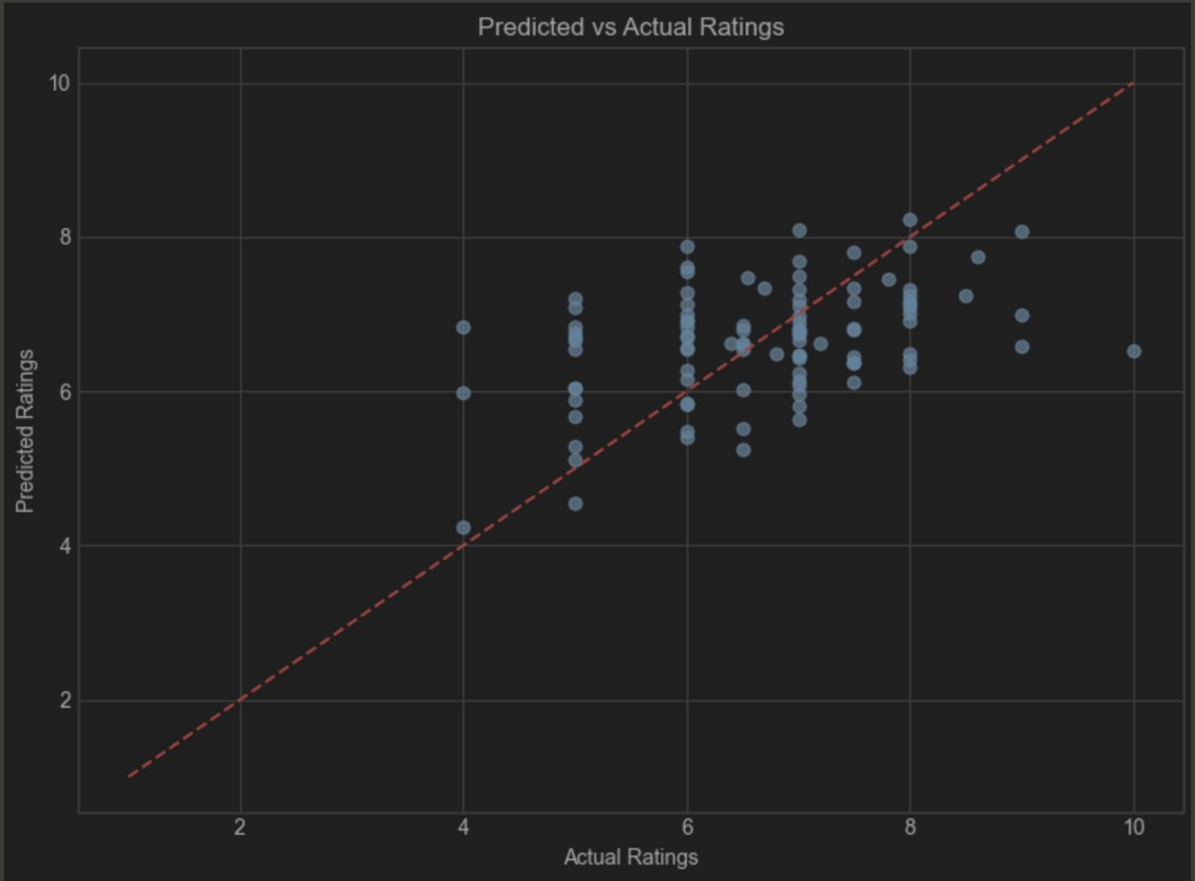
Results

- 5 test samples of size 100:
    - Average MAE of 0.92 with a std. deviation of 0.08
- On a per-game basis, some games had noticeably higher/lower performance ranging from ~0.4 to ~1.9 MAE

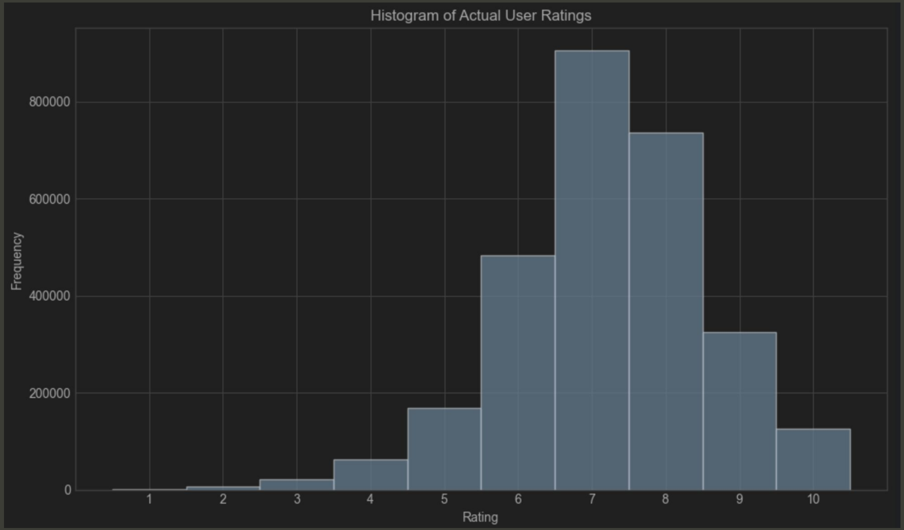
Results (cont.)

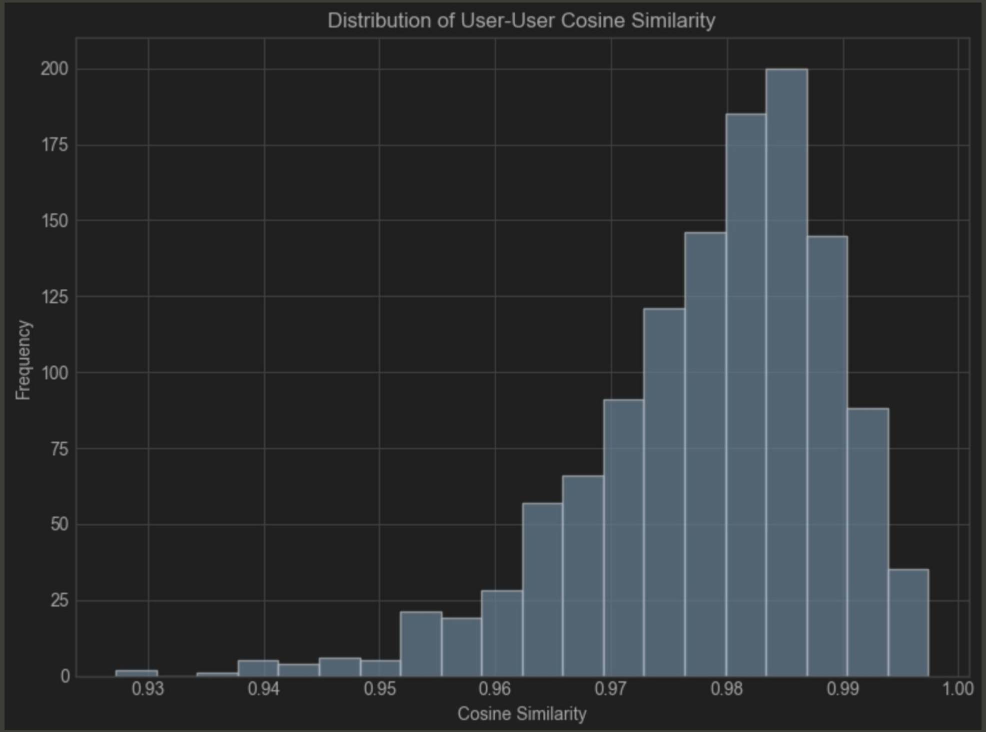


Results (cont.)



Results (cont.)





Initial Conclusions

- Predicting ratings performed effectively with the filtered data
- High similarity scores among users reflected similar rating patterns
  - Very common rating pattern amongst users (~7/10)
- Variability in user perception of specific games directly correlated with model performance

Baseline Comparison

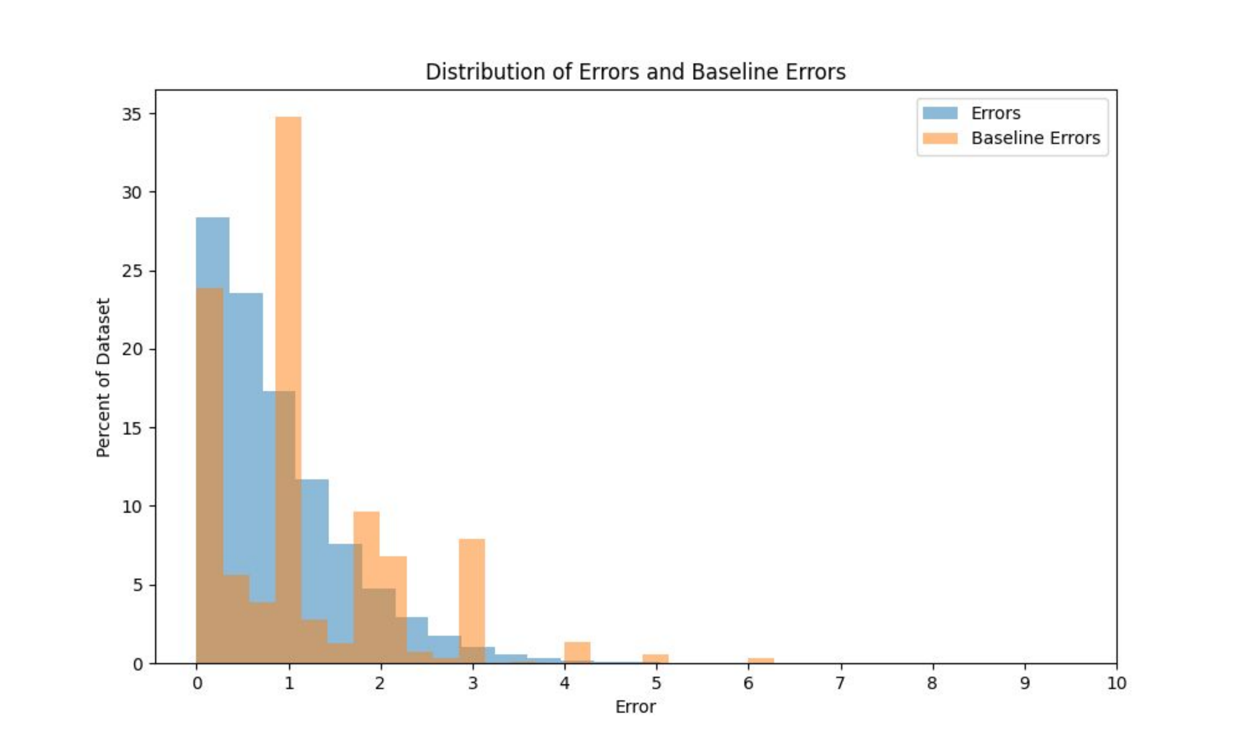
Baseline: predict the average game score (7.1) for every query

- Error with $\mu = 1.069$, $\sigma = 0.857$
- RMSE: 1.54

Collaborative filtering on all 2,836,563 ratings

- Error with $\mu = 0.879$, $\sigma = 0.748$
- RMSE: 1.15

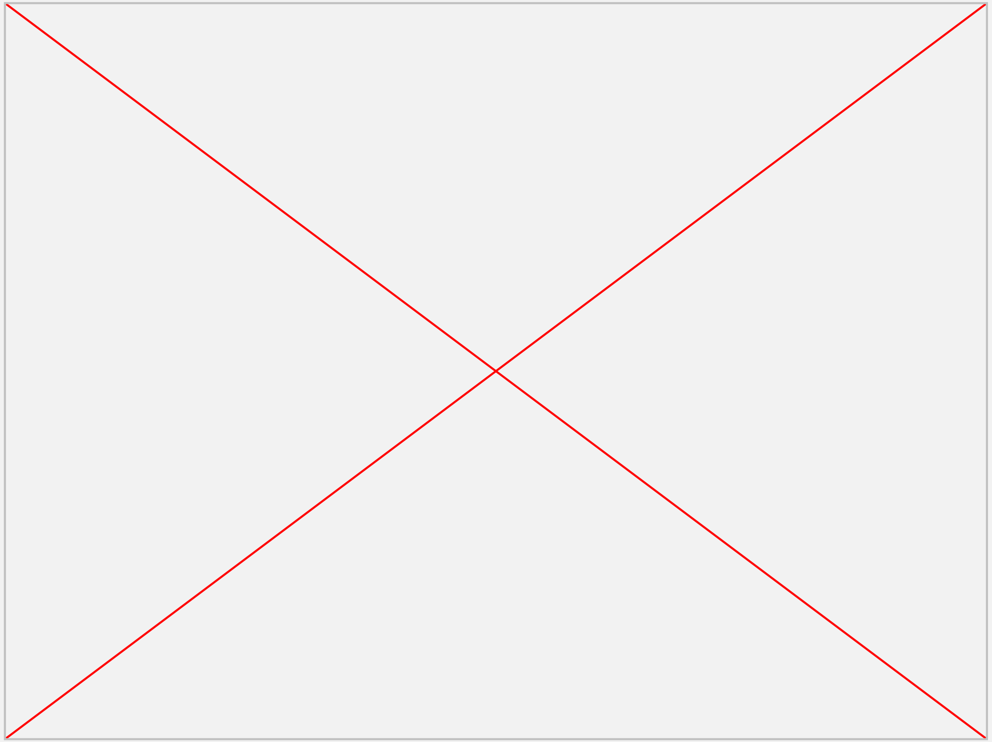
sidenote: CF took 2500+ cpu credit-hours (3 months)



Colab Filtering does something! Most errors are ≤ 1.

Colab Filtering Applet

Rate 10 random games to get personalized suggestions!



Conclusions 2

We beat a baseline:

1.069	→	0.879	MAE
1.54	→	1.15	RMSE

Qualitative Evaluation

Using the applet, pretended to be a user that likes:

minifigures	(2/3 games with minifigures suggested)
card games	(1/3 games with a lot of cards suggested)
social deduction games	(1/3 games with social deduction suggested)
history	(0/3 games with history suggested)

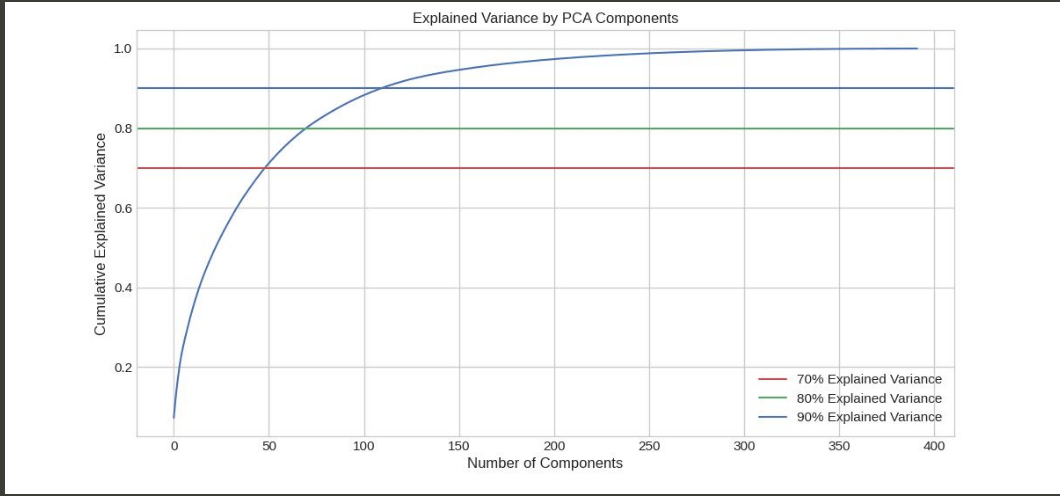
Part 3
Clustering

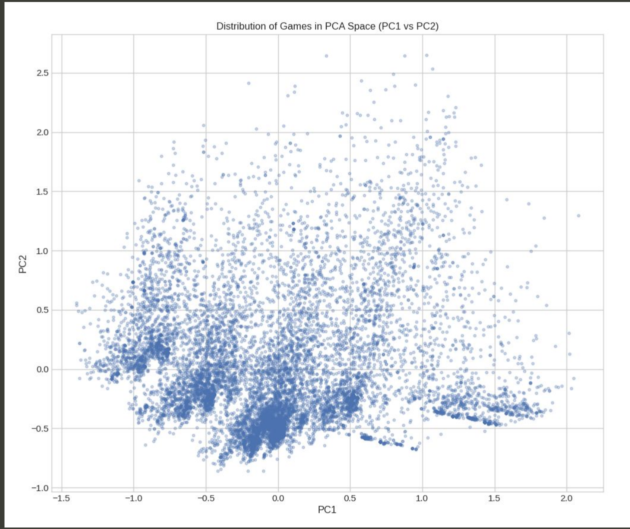
Data Preprocessing

- Merged GAMES data with MECHANICS, THEMES, and SUBCATEGORIES based on id
- Filtered games to only those with >=100 user ratings
- Final dataset reduced from 21,925 to 12,239 games
- 432 total features

Methods

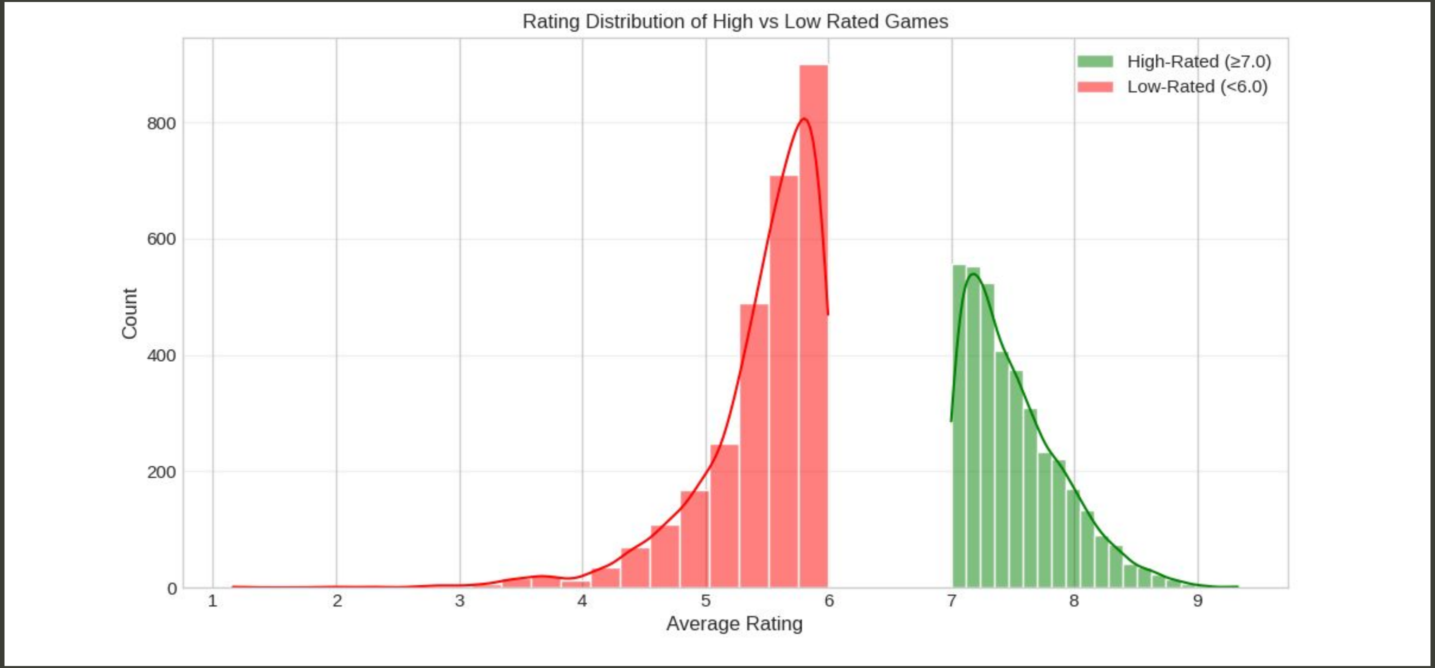
- Feature selection & PCA
  - Only used binary attributes (392 in total)
    - 1: the game has the feature, 0: the game doesn’t
  - PCA applied to reduce dimensionality
    - 71 components for 80% variance





Methods (cont.)

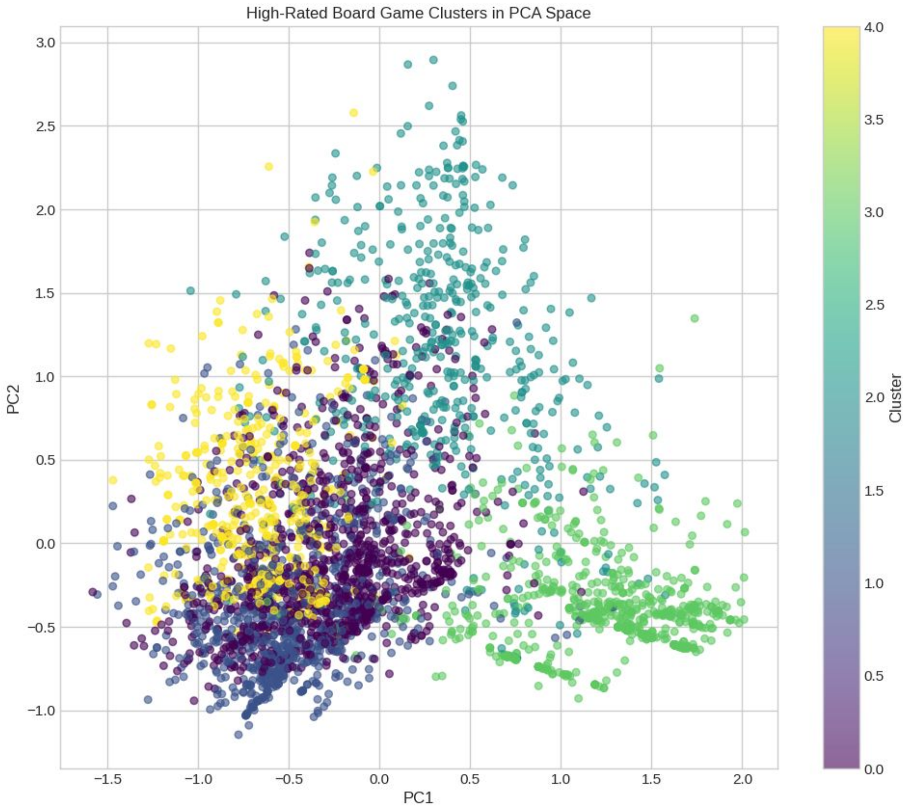
- Split dataset into two groups:
  - High-rated games (>=7) and low-rated games (<6)
  - Further removed 5677 games

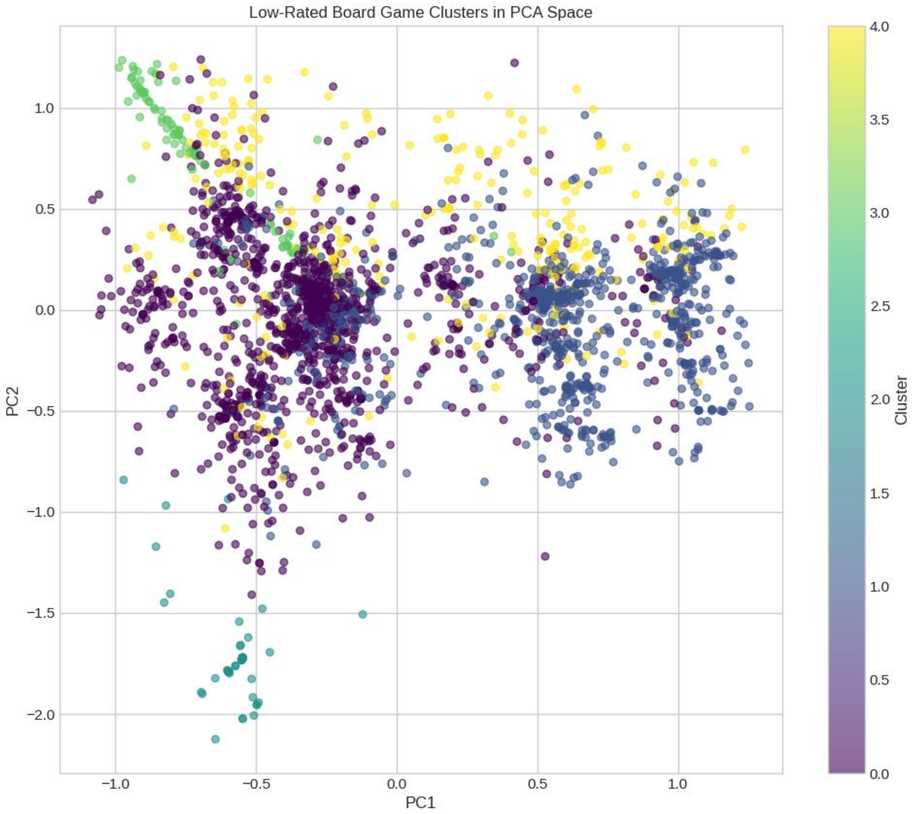


Methods (cont.)

- Agglomerative hierarchical clustering used
  - Grid search over n_clusters from 5-15, various linkage options, and distance metrics
  - Grid search optimized using Calinski-Harabasz score
  - Clusters then examined manually for feature prevalences
    - Decision Tree also trained to see which features distinguish a cluster from other clusters

Results


High-Rated Board Game Clusters in PCA Space
PC1
PC2
Cluster
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0


Low-Rated Board Game Clusters in PCA Space
PC1
PC2
Cluster
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Results (high-rated games)

- Cluster 0: Card-Based Games (21%)
  - Defined by hand management, drafting, deckbuilding, along with fantasy and fighting themes
- Cluster 1: Classic Strategy Games (31.4%)
  - Defined by area majority/influence, tile placement, worker placement, along with economic & city building themes
- Cluster 2: Dice-Based Adventure Games (14.5%)
  - Defined by dice rolling, variable player powers, cooperative gameplay, along with fantasy, adventure, & fighting themes
- ...

Results (high-rated games)

- Cluster 3: War Simulation Games (20%)
  - Defined by dice rolling, simulation, hexagon grid, along with WWII, modern warfare, and aviation themes
- Cluster 4: Party & Deduction Games (13%)
  - Defined by deduction, cooperative gameplay, dice rolling, along with science fiction, murder/mystery, and humor themes

Results (high-rated games)

- No one path to excellence in terms of rating
  - varied approaches, from themed adventures to economic strategy to card games
- Successful mechanic integration for each cluster
  - **Balanced dice mechanics** in Cluster 2 (Dice-Based Adventure Games)
  - **Coherent themes** in Cluster 3 (War Simulation Games)
  - **Social interaction value** in Cluster 4 (Party and Deduction Games)

Results (low-rated games)

- Cluster 0: Family & Children's Games (49.2%)
  - Defined by dice rolling, roll/spin and move, and dexterity, along with themes of Movies/TV, economic, & trivia
- Cluster 1: Card Games (34.6%)
  - Defined by hand management, set collection, betting/bluffing, with themes of humor and animals
- Cluster 2: Monopoly-Style Games (1.9%)
  - Defined by roll/spin and move, negotiation, and trading, with economic and movies/TV themes
- ...

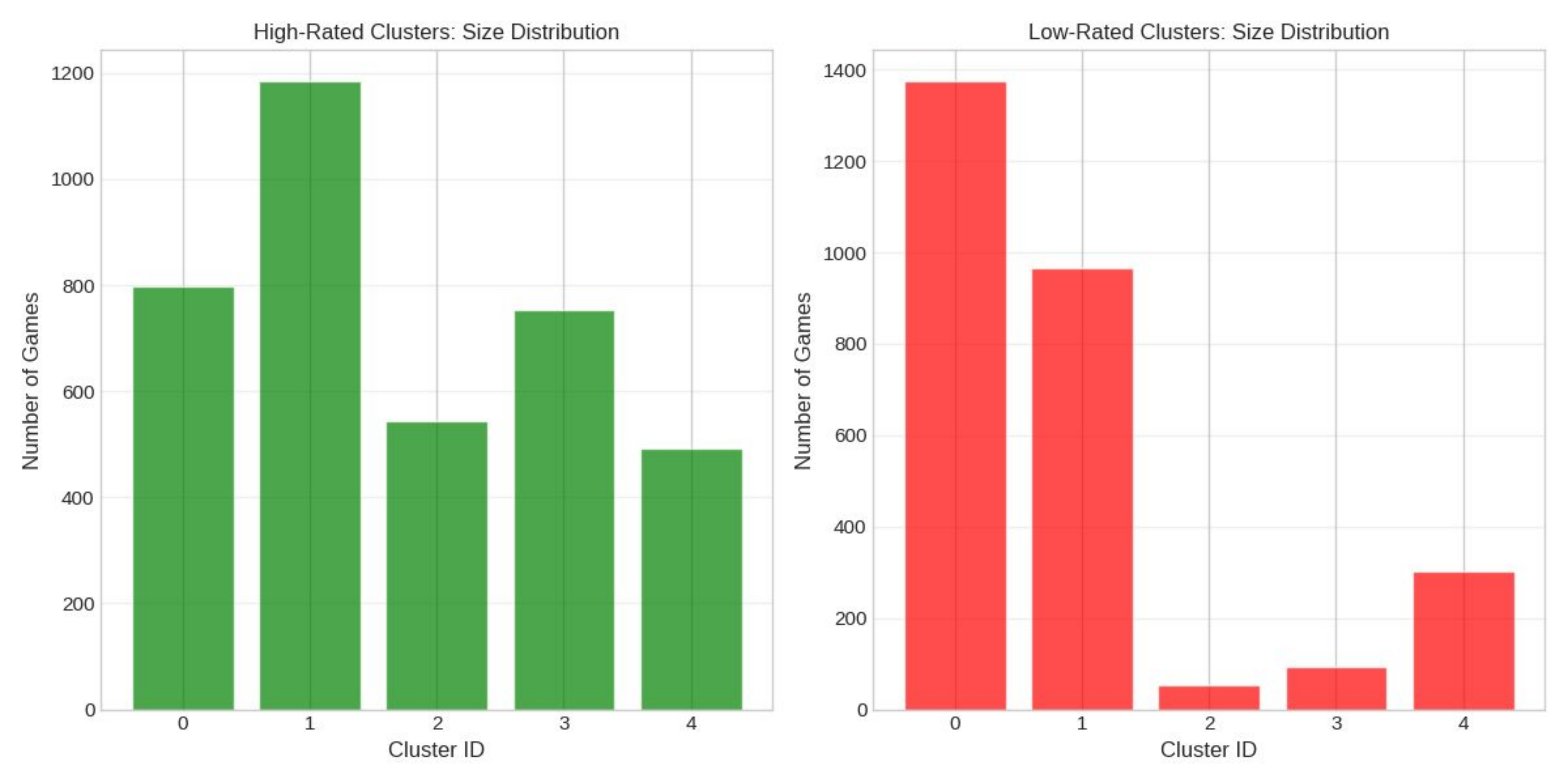
Results (low-rated games)

- Cluster 3: War Simulation Games (3.4%)
  - Defined by dice rolling, hexagon grid, and simulation, with themes of WWII and modern warfare
- Cluster 4: Themed Fantasy Games (10.9%)
  - Defined by dice rolling and hand management, with themes of fantasy and fighting

Results (low-rated games)

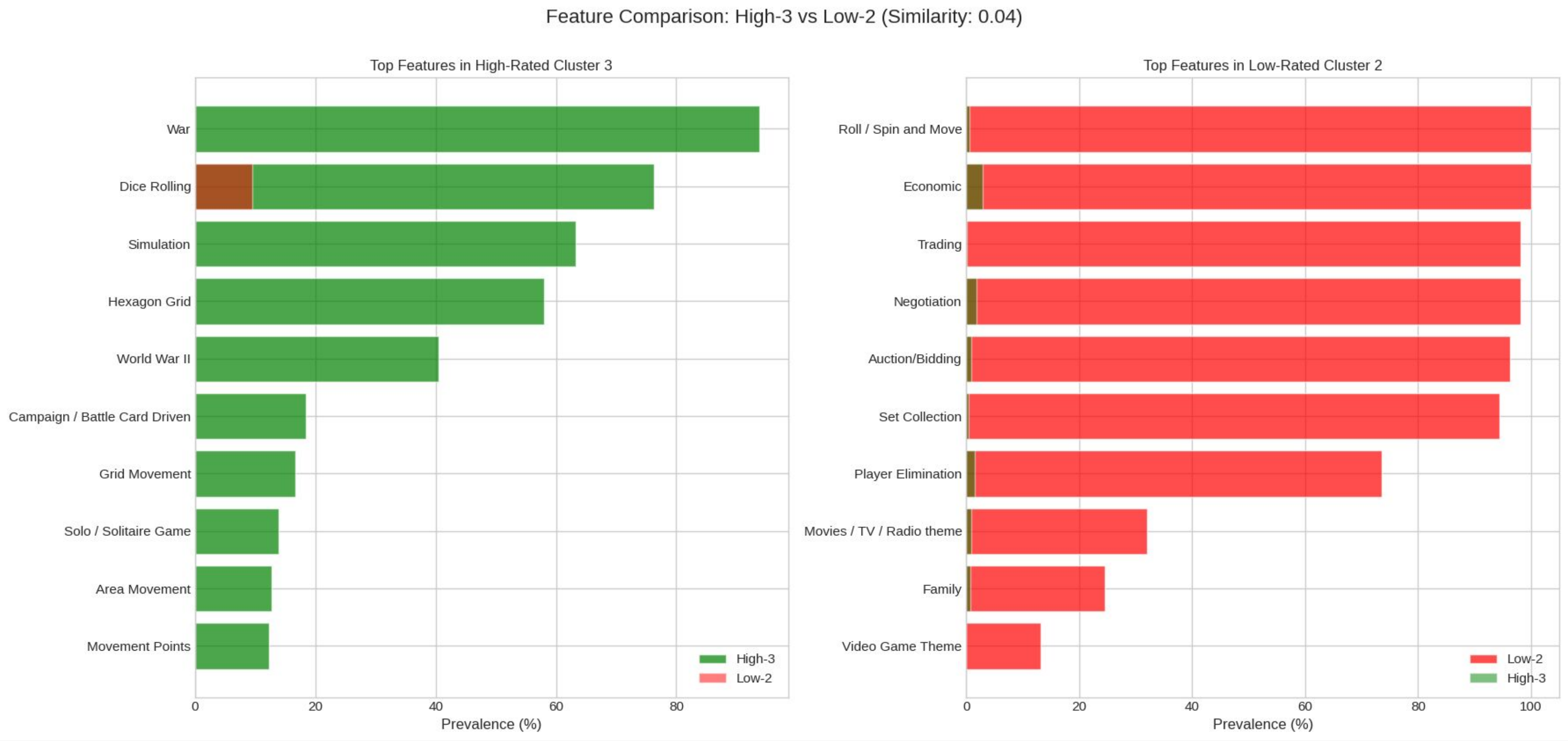
- Games with an over reliance on luck do not seem to be well received (ex. roll/spin move feature)
- Themes without substance (fantasy, movies/TV, video games) in Cluster 2 and 4 show that rehashing existing games is not a recipe for success
- Outdated designs (Cluster 2: Monopoly-Style Games) do not perform well
- Games (Cluster 0: Children′s Games) that are too childish for strategy gamers or too complex for families

Results (cont.)



Results (cont.)

Results (cont.)



Conclusions

- Execution, quality, and effective integration of features was more important than features alone
  - ex. War Games
- Players prefer games with strategic depth; excessive reliance on luck-driven mechanics led to lower ratings
  - ex. Card Games

Part 4
Regression

Data Preprocessing

- Filtered attributes to include only numerics
- Filtered games to only those with >=100 user ratings
- Final dataset reduced from 21,925 to 12,239 games

Methods: Linear Regression

- Assumes linear relationship
- Simplicity allows for easy interpretation
    - In contrast, complex relationship representation is limited

Methods: Random Forest Regressor

- Accounts for nonlinear relationship

- Predictions aggregated from multiple decision trees

  - Aim to improve accuracy and reduce overfitting

Methods: Random Forest Regressor (cont.)

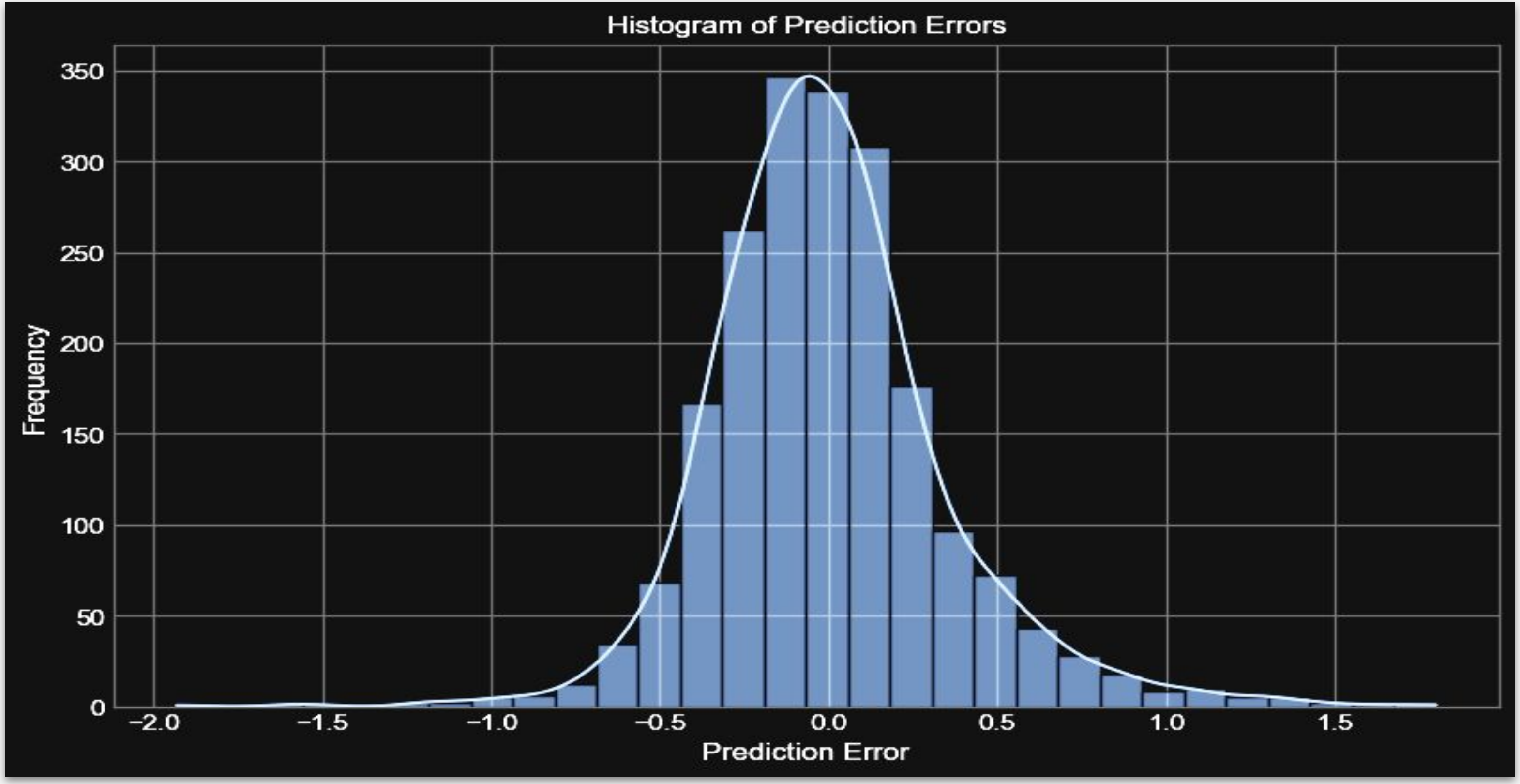
- Implemented Grid Search with the following parameters:
    - *n_estimators*: The number of trees in the forest
    - *max_depth*: The maximum depth of the forest, and
    - *min_samples_split*: The minimum number of samples required to split an internal node.
- 5-fold cross-validation ➡ 36 param combinations, 180 fits

Results: Linear Regression

- Evaluated Mean Squared Error (MSE), Mean Average Error (MAE), and coefficient of determination (R^2)

- ***Model Performance:***
  - Mean Squared Error (MSE): 0.1219
  - Mean Absolute Error (MAE): 0.2557
  - R^2 Score: **0.8166**

Results: Linear Regression



Results: Random Forest Regressor

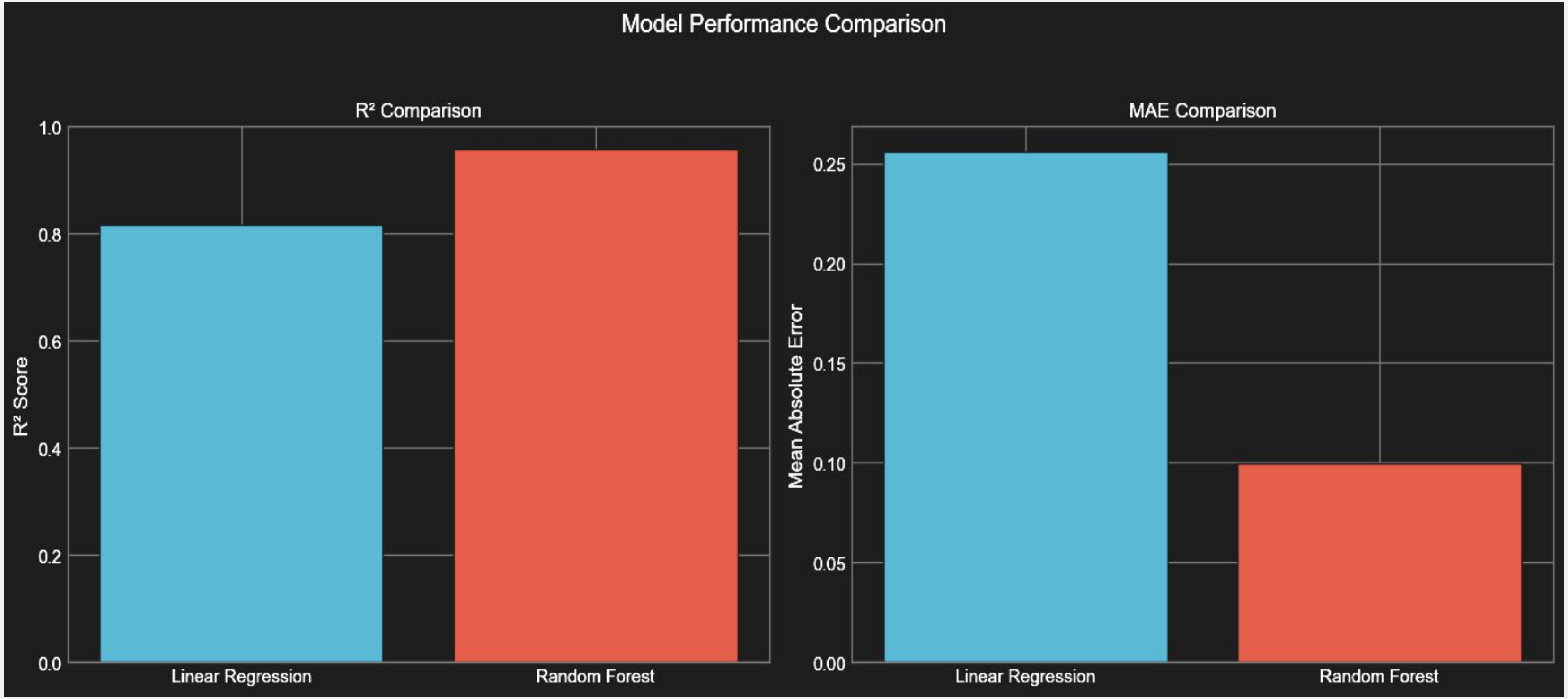
- Evaluated Mean Squared Error (MSE), Mean Average Error (MAE), and coefficient of determination (R^2)

- ***Model Performance:***
    - Mean Squared Error (MSE): 0.02795
    - Mean Absolute Error (MAE): 0.09931
    - R^2 Score: **0.95797**

Results: Comparative Analysis

- Linear Regression
  - Mean Squared Error (MSE): 0.1219
  - Mean Absolute Error (MAE): 0.2557
  - R^2 Score: **0.8166**
- Random Forest Regressor
  - Mean Squared Error (MSE): 0.02795
  - Mean Absolute Error (MAE): 0.09931
  - R^2 Score: **0.95797**

Results: Comparative Analysis



Conclusions

- We can build a model to accurately predict a value metric, average rating, given game attributes.

- There is a correlation between game attributes and overall rating of game

    - Better represented as a nonlinear relationship

Part 5
Conclusion

Summary

- Explored board game data from BoardGameGeek
- Performed analysis with **Collaborative Filtering**, **Clustering**, and **Regression Models**
- All three analytic approaches were successful with their own insights

Our Experience

- Data analysis is iterative—not a one-step process
- Initial findings led us to continually ask **“why?”** for deeper insights
- Multiple analytical methods gave us a complete dataset perspective

Societal Impact

- Findings can help designers to create engaging games aligned with consumer preference
- Over-reliance on predictive models risks reduction in creative diversity and can lead to formulaic games
- Niche games become more niche, mainstream becomes more mainstream
- Communal and global economy boost for mainstream game developers