Board Game Prediction & Clustering

Paksi Yudha Mahendra

Overview Dataset

This dataset contains comprehensive information about board games scraped from BoardGameGeek (BGG), the world's largest board game database and community

# Overview Dataset

What's Included

- Game Details   : Names, publication years, player counts, play times, age recommendations
- Quality Metrics   : Average ratings, complexity scores, user review counts
- Popularity Data   : Ownership statistics, wishlist counts, play frequencies
- Community Rankings : Overall and category-specific rankings
- Designer Credits   : Game designers, artists, publishers, and other contributors
- Rating Distributions   : Breakdown of user ratings from 1-10
- Amazon Pricing   : Current prices for games available on Amazon
- Game Classifications : Categories, mechanics, and game families

OBJECTIVE

Average Rating Board Game Prediction and Clustering Board Game using machine learning.

EDA

Heatmap of Correlation Matrix for Selected Columns
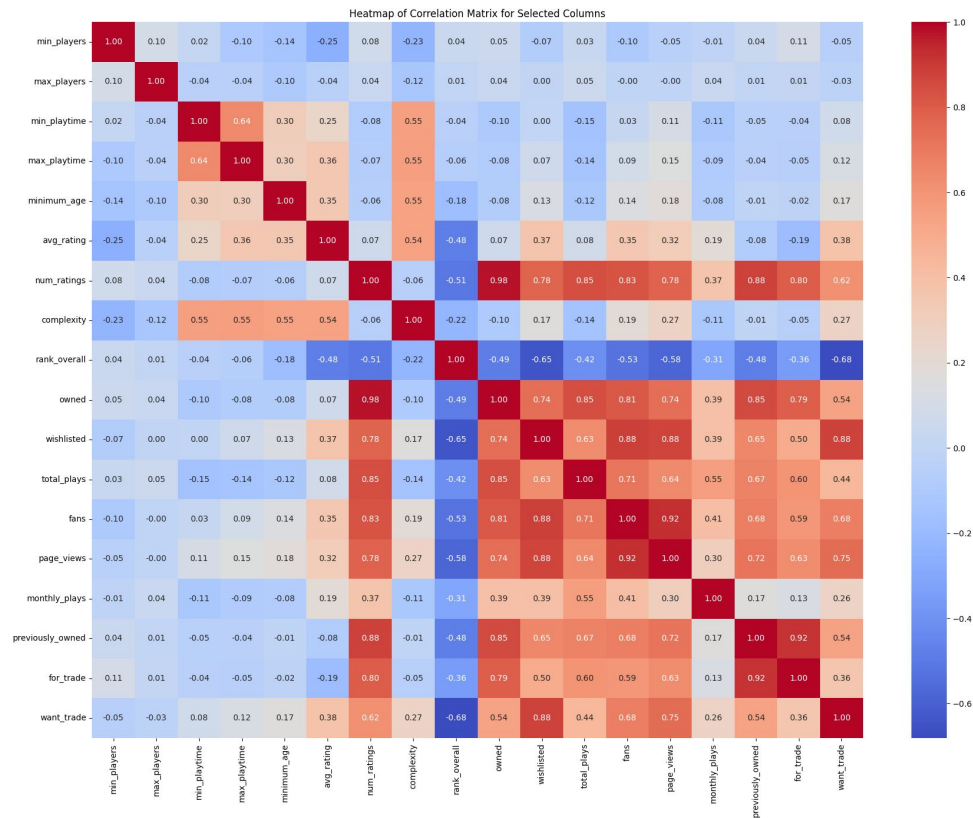
Most of the data has correlation with other data.
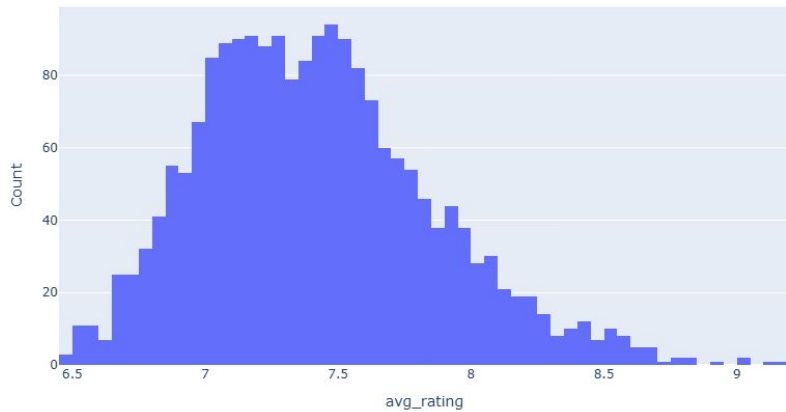
EDA

Interactive Boxplot for avg_rating

The majority of the data contains quite a lot of extreme values.

EDA

Interactive Histogram for avg_rating

And the majority of the data has a distribution that is skewed to the left.

EDA

- The majority of the datasets exhibit correlations.
- The dataset distribution is skewed to the left and contains quite a few extreme values.

MACHINE LEARNING MODEL IMPLEMENTATION

Prediction:

Regression Linear, Random Forest

Clustering:

K-Means, Hierarchy Clustering

MACHINE LEARNING MODEL IMPLEMENTATION

Evaluasi model Prediction

Random Forest Regressor

MSE: 0.00184

R-squared : 0.935694

Regression Linear

MSE: 0.00912

R-squared : 0.655323

MACHINE LEARNING MODEL IMPLEMENTATION

Evaluasi model Clustering

K-Means (n = 3)

Silhouette Score: 0.2846

Hierarchy Clustering (n = 3)

Silhouette Score: 0.2351

MACHINE LEARNING MODEL IMPLEMENTATION

Conclusion Evaluasi Model Prediction

Random Forest provides a smaller MSE and a larger R-squared than Linear Regression, meaning this model can model predictions well and with small errors.

MACHINE LEARNING MODEL IMPLEMENTATION

Conclusion Evaluasi Model Clustering
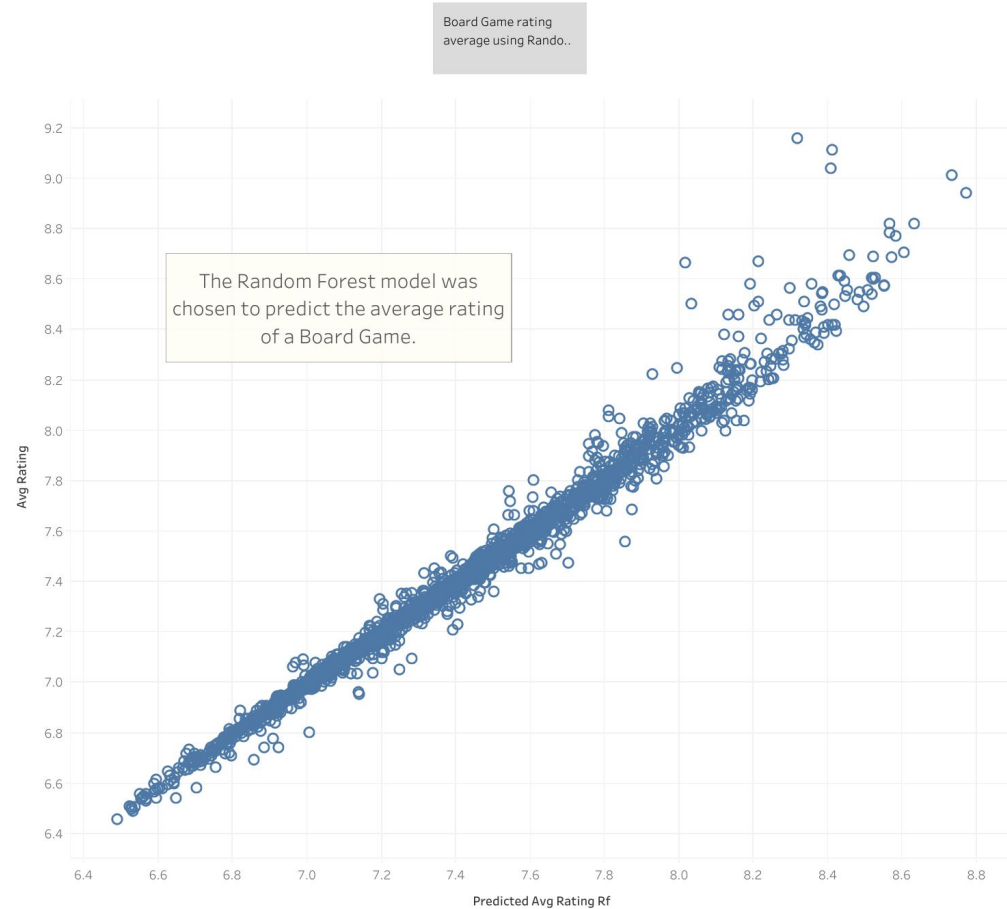
The K-Means model has a larger Silhouette Score value than the hierarchy model even though the value is close to 0, which means that there are still data points that overlap with other cluster data.

CONCLUSION MACHINE LEARNING MODEL

BoardGame rating average

Board Game rating
average using Rando..

The Random Forest model was
chosen to predict the average rating
of a Board Game.
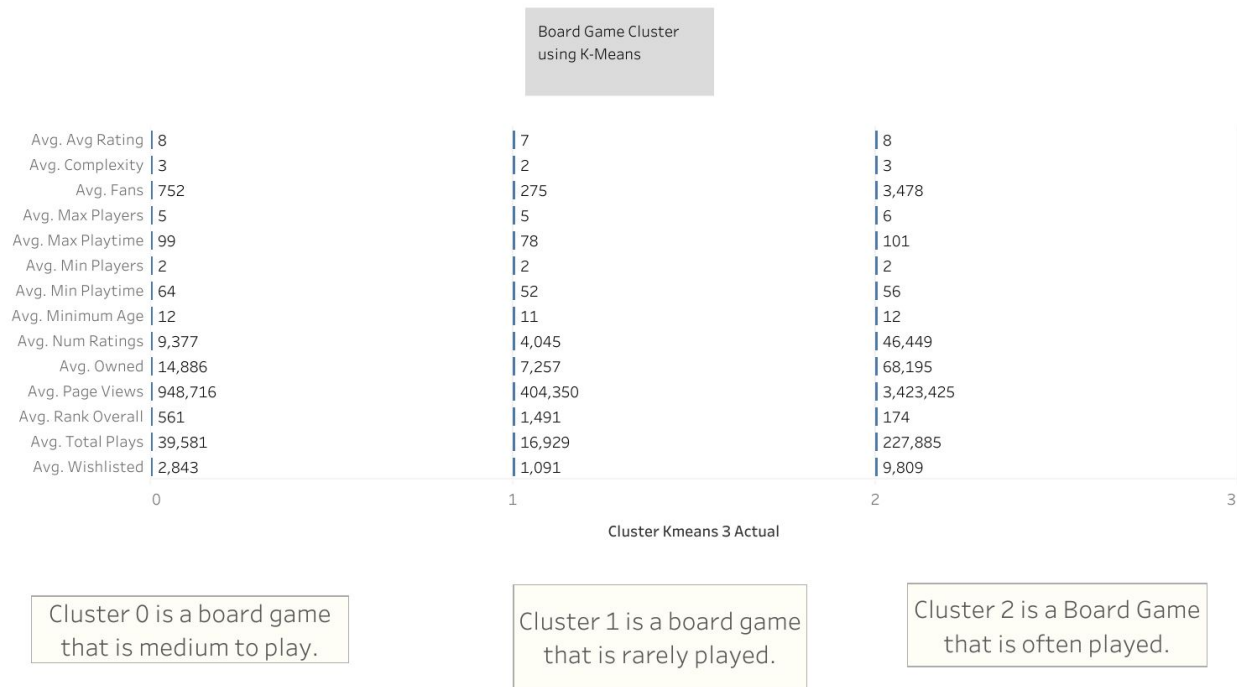
CONCLUSION MACHINE LEARNING MODEL

K-Means Clustering with 3 Clusters (PCA 2D)
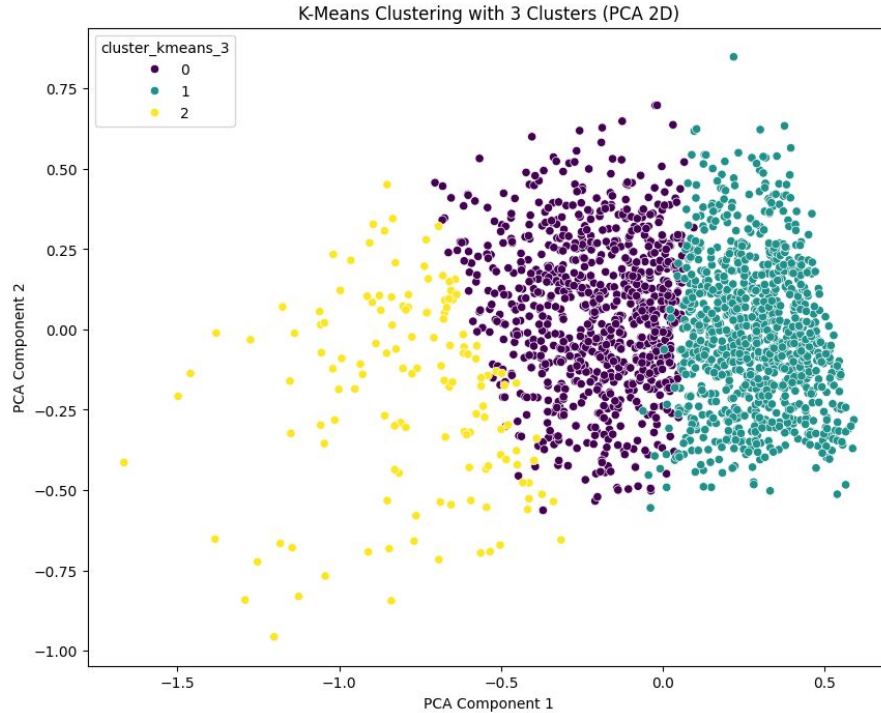
The K-Means model was chosen for clustering Board Games based on features.

CONCLUSION MACHINE LEARNING MODEL

BoardGame Clustering

	Board Game Cluster using K-Means	

	0	1	2
Avg. Avg Rating	8	7	8
Avg. Complexity	3	2	3
Avg. Fans	752	275	3,478
Avg. Max Players	5	5	6
Avg. Max Playtime	99	78	101
Avg. Min Players	2	2	2
Avg. Min Playtime	64	52	56
Avg. Minimum Age	12	11	12
Avg. Num Ratings	9,377	4,045	46,449
Avg. Owned	14,886	7,257	68,195
Avg. Page Views	948,716	404,350	3,423,425
Avg. Rank Overall	561	1,491	174
Avg. Total Plays	39,581	16,929	227,885
Avg. Wishlisted	2,843	1,091	9,809

Cluster Kmeans 3 Actual

Cluster 0 is a board game that is medium to play.

Cluster 1 is a board game that is rarely played.

Cluster 2 is a Board Game that is often played.

[Link Tableau](#)

[Link Google Colab](#)