

The background is a light gray with a complex arrangement of geometric shapes. There are several triangles in various colors (red, green, orange, black, gray, white) and sizes. Some are solid, while others are outlines. There are also circles in blue, white, orange, red, purple, and green. A large orange circle in the top right contains the name 'Ana Morais'. A purple banner across the middle contains the title 'Board Games ML Clustering'.

# Board Games ML Clustering

Ana Morais

# overview

- **Objective:** Build a recommender system for board games based on the BoardGameGeek database
- **Features (numerical continuous):**
  - Game weight
  - Min players
  - Max players
  - Playtime
  - Age recommendation
  - # expansions
- **Features (categorical):**
  - Game categories
  - Game subcategories
  - Game mechanics
  - Game themes
- **(19504, 396)**

# Implementation details

## Data Processing

Normalization of continuous variables:

- log-transform
- Robust-Scaler
- Min-Max scaler (0-1)

## Estimating best parameters

Elbow method – optimal number of clusters = 4

K-means for PCA with n\_components in [2, 3, 4] & choosing best metrics

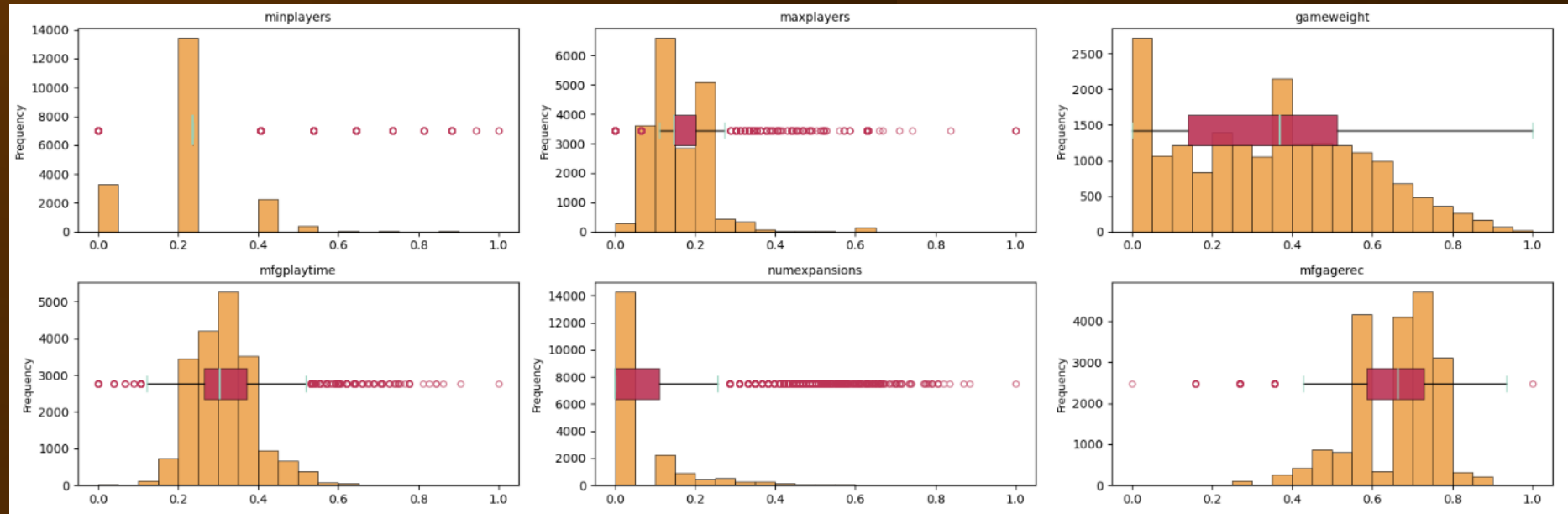
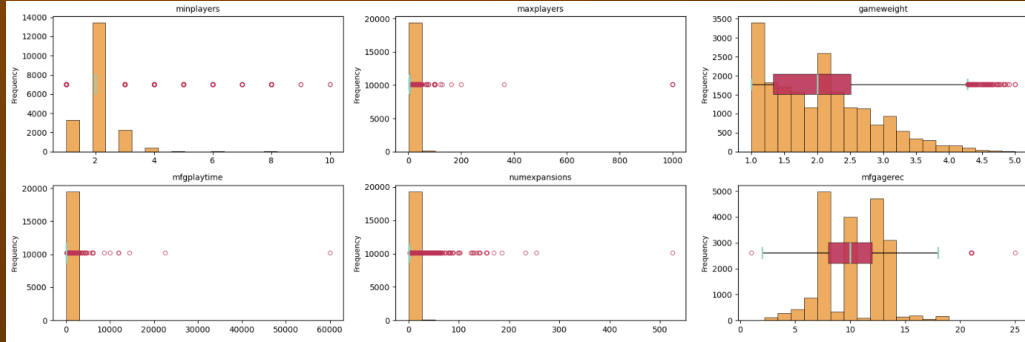
## Data Modeling

K-means Clustering

Metrics:

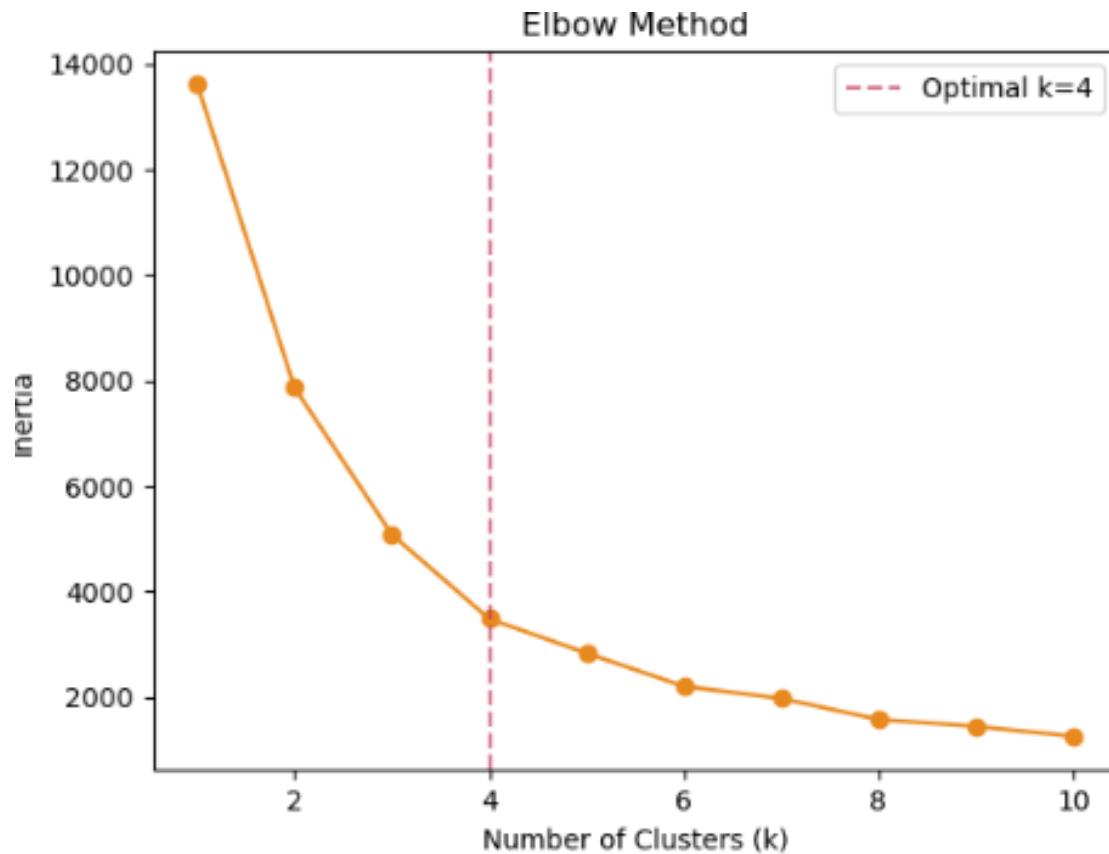
- Silhouette Score = 0.4806
- Explained variance ratio = 0.1707

# Normalized data



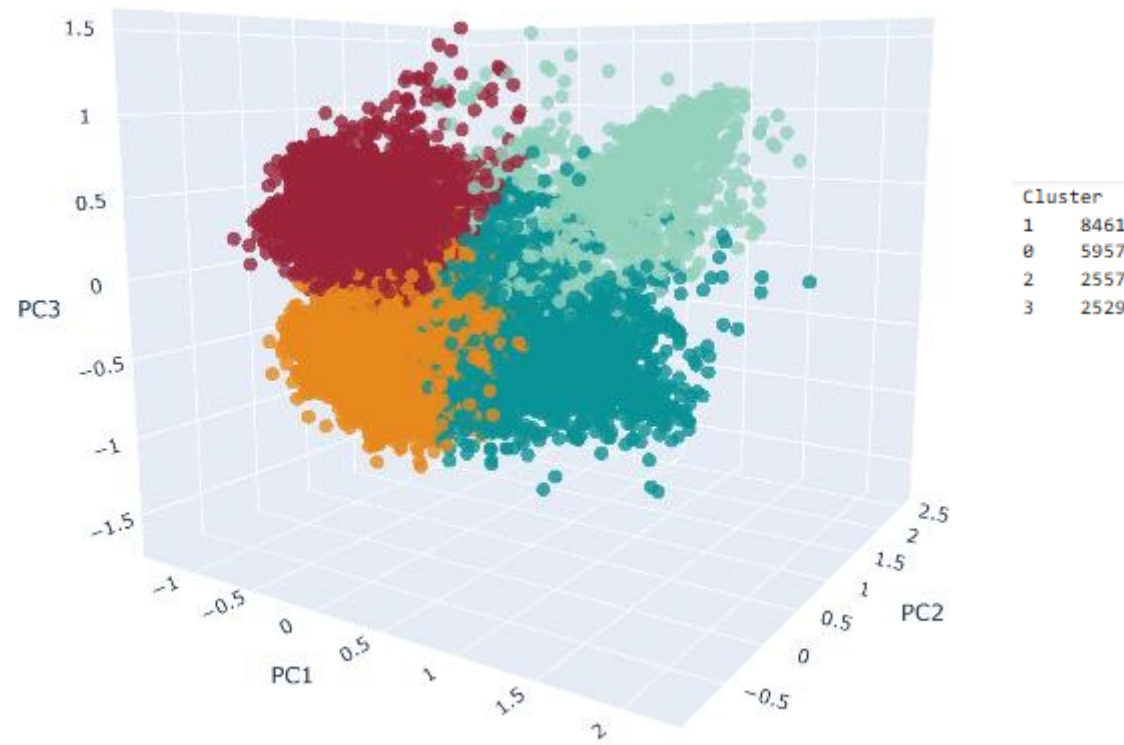
Numerical continuous features

# Best Parameters for K-means with PCA

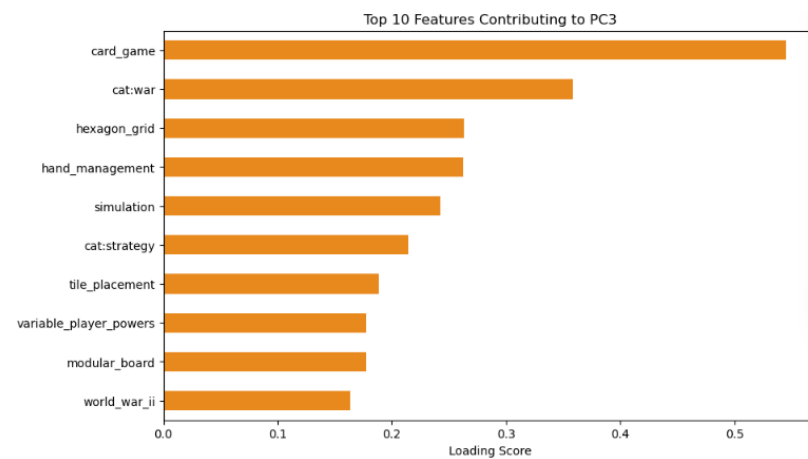
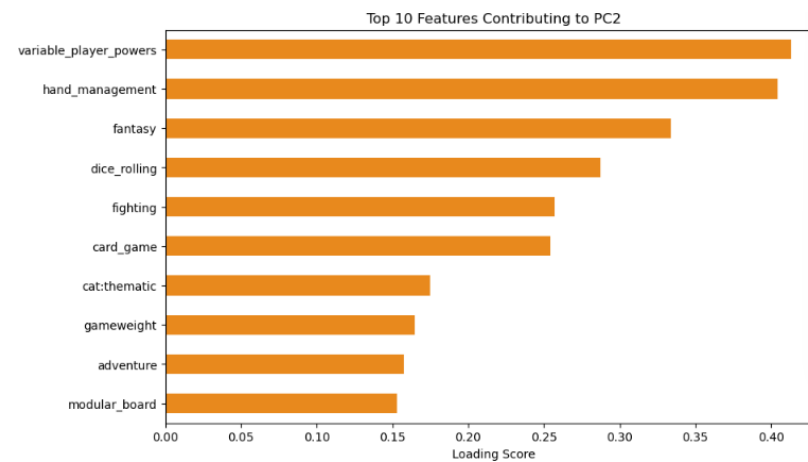
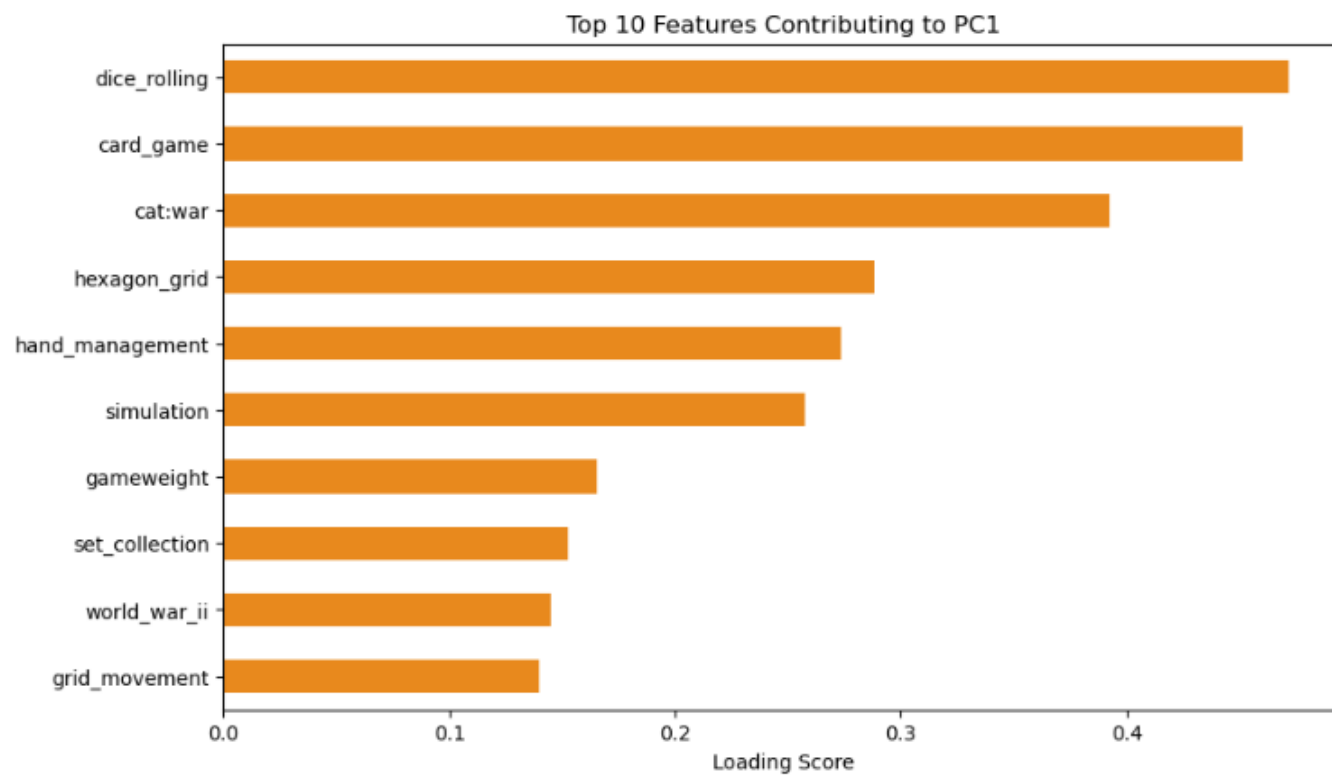


K-means metrics for PCA reduced data with  $n=[2,3,4]$ :

	n_components	explained_variance_ratio	silhouette_score
1	3	0.170690	0.480610
0	2	0.129425	0.457911
2	4	0.205596	0.394283



Scatter plot of K-means clusters





What is your  
favorite board  
game?





# Conclusions



## Model Performance:

**Best Parameters:** 3 PCA components and 4 K-means clusters

**Key Metrics:** explained variance ratio = 0.1707; silhouette score = 0.4806

## Future Directions:

Explore cumulative explained variance to get optimal n PCA components

Explore other clustering models like Agglomerative Clustering, DBSCAN and HDBSCAN

Refine the recommender system with extra steps to measure similarity within clusters

## Main Takeaway:

With a bit more effort, we can build a reasonably good recommender system for board games! It is possible!!

**Thank you!**

