NBA Player Archetype Clustering

Zulmei Eshaqzaie

Project Objective

The goal is to identify and group NBA players into archetypes based on their statistical performance during the 2022 season.

I do this by applying K-Means Clustering to categorize players into different groups and using PCA for simplifying the data and making it easy to see the data in a graph.

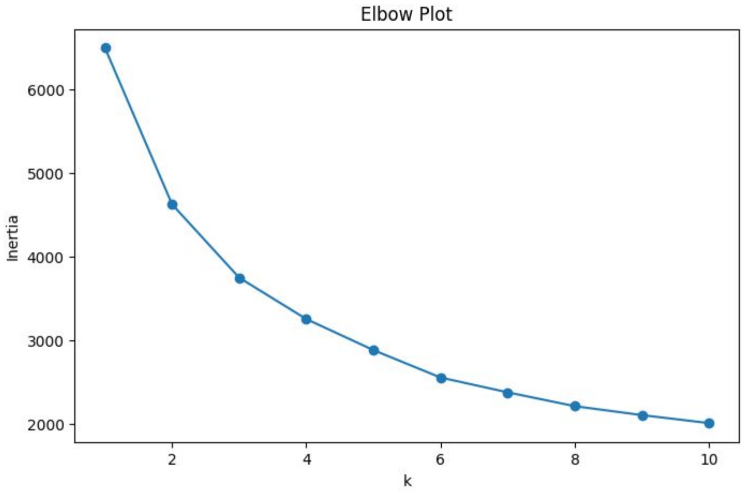
This form of analysis can help NBA coaches and scouts get a better understanding of the type of a player and how they contribute on the court.

Data Preparation

- Loading the dataset
  - Imported 2022 season stats from Kaggle (link)
- Cleaned the data
  - Checked for any missing value
  - Fixed some inconsistencies in player names
- Selected key stats used for clustering
  - Points (PTS), Assists (AST), Rebounds (TRB), Steals (STL), Blocks (BLK), Field Goal Percentage (FG%), Three Point Percentage (3P%), Free Throw Percentage (FT%)
- Standardized the stats
  - Used StandardScaler to put all the stats on the same scale

	PTS	AST	TRB	STL	BLK	FG%	3P%	FT%
0	0.209896	-0.385559	1.347449	-0.194639	0.683406	0.085997	0.523628	-0.223310
1	-0.148744	0.866518	2.835943	0.746117	1.238055	0.813597	-1.756004	-0.406850
2	1.840075	0.866518	2.878472	1.922062	1.238055	0.880967	-1.756004	0.334371
3	-0.605194	-0.603311	-0.268631	-0.900206	-0.148567	-0.163274	-0.962260	-0.117421
4	0.829364	-0.494435	0.922165	-0.665017	1.792703	0.833808	0.174381	0.757926

Determining Optimal Clusters



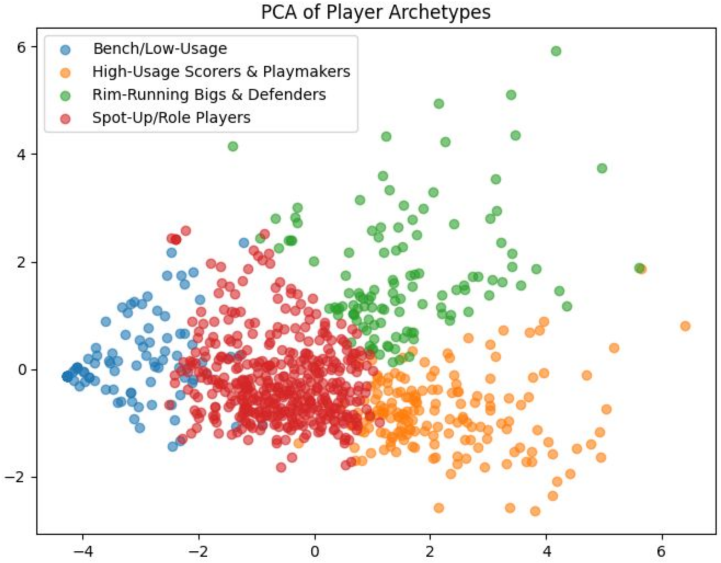
- How did we choose the number of clusters?
  - The elbow of the curve appears at k=4
  - This is the point where inertia started to flatten out and adding more clusters wouldn't improve the grouping that much more
- Cluster Assignment
  - Each player gets assigned to a cluster
  - Calculated cluster centroids to understand the average stats of each group

	PTS	AST	TRB	STL	BLK	FG%	3P%	FT%	Cluster
0	5.366745	1.123585	2.404717	0.432075	0.233491	0.440868	0.304033	0.719120	0
1	15.898837	4.453488	4.708140	1.080233	0.395930	0.443448	0.348110	0.800762	1
2	10.351261	1.478992	6.527731	0.698319	0.951261	0.548378	0.246899	0.695983	2
3	1.049485	0.514433	1.021649	0.217526	0.070103	0.181907	0.065804	0.093330	3

PCA Visualization

	PC1	PC2
PTS	0.462050	-0.098530
AST	0.381484	-0.330306
TRB	0.409756	0.381617
STL	0.393104	-0.198248
BLK	0.316913	0.554431
FG%	0.255006	0.384108
3P%	0.239795	-0.415895
FT%	0.309169	-0.261447

- Interpreting the archetypes
  - Based on the cluster centroids, we have the average stats of each group and could assign each cluster an archetype
    - Spot-Up/Role Players
    - High-Usage Scorers & Playmakers
    - Rim-Running Bigs & Defenders
    - Bench/Low-Usage
- PCA to visualize
  - Reduced the dataset to down to 2 dimensions using PCA to make the clu easier to visualize
  - Players are on a scatter plot, grouped by cluster and indicated by color
- Principal Components
  - PC1 captures general offensive contribution and all-around activity
  - PC2 captures big man vs. guard style



Profiling Clusters

Average Stats by Archetype:

Role	PTS	AST	TRB	STL	BLK	FG%	3P%	FT%
Bench/Low-Usage	1.05	0.51	1.02	0.22	0.07	0.18	0.07	0.09
High-Usage Scorers & Playmakers	15.90	4.45	4.71	1.08	0.40	0.44	0.35	0.80
Rim-Running Bigs & Defenders	10.35	1.48	6.53	0.70	0.95	0.55	0.25	0.70
Spot-Up/Role Players	5.37	1.12	2.40	0.43	0.23	0.44	0.30	0.72

- How did we profile the clusters?
  - Calculated the average stats for each archetype to understand the average performance of players under each role
  - Identified the top 5 scorers within each archetype to give real player examples
- Key findings per archetype
  - High-Usage Scorers & Playmakers
    - Lead in points and assists
    - LeBron James, Giannis Antetokounmpo, Kevin Durant, Luka Dončić, Trae Young
  - Rim-Running Bigs & Defenders
    - Lead in rebounds, blocks, field goal percentage
    - Joel Embiid, Karl-Anthony Towns, Anthony Davis, Kristaps Porziņģis
  - Spot-Up/Role Players
    - Lead in three point percentage and free throw percentage
    - De'Andre Hunter, Eric Gordon, Coby White, Lonnie Walker IV, Malik Beasley
  - Bench/Low-Usage
    - Lower across all stats
    - Isaiah Thomas, Greg Monroe, Scotty Hopson, Gabriel Lundberg, Tremont Waters

OOP Wrappers

- Why use OOP?
  - Makes it easier to manage player data and cluster assignments using objects instead of raw dataframes
- Player Class
  - Stores player details: name, team, position, stats, and role
  - Has a simple assign_role method to update the player's role after clustering
- PlayerCluster Class
  - Represents each cluster with:
    - An archetype
    - Centroid stats
    - A list of player members
  - Has methods to add players to clusters and summarize cluster info

Conclusion & Future Work

- Conclusion
  - Successfully used K-Means Clustering to group NBA players into different archetypes
  - Applied PCA to visualize these roles and confirm clear separation between player types
  - Used Object Oriented Programming (OOP) to neatly organize players and clusters for easy analysis.
  - The statistical profiles that were created aligned well with actual NBA roles and players, demonstrating that evaluating players through data analysis was effective.
- Future Work
  - Testing a higher number of clusters to see if the data can find more specific archetypes
  - Using advanced stats such as PER, BPM, Win Shares to deeper evaluate players