Clustering Methods for Identifying NBA Player Archetypes.

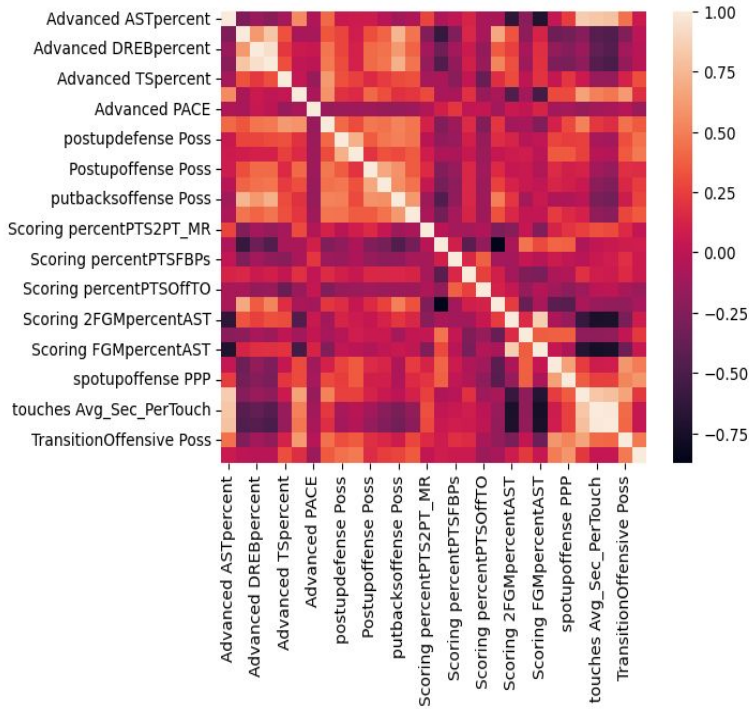
Anand Addepalli
Sahithi Bathini
Vaishnavi Jangili

Data Sources and Collection

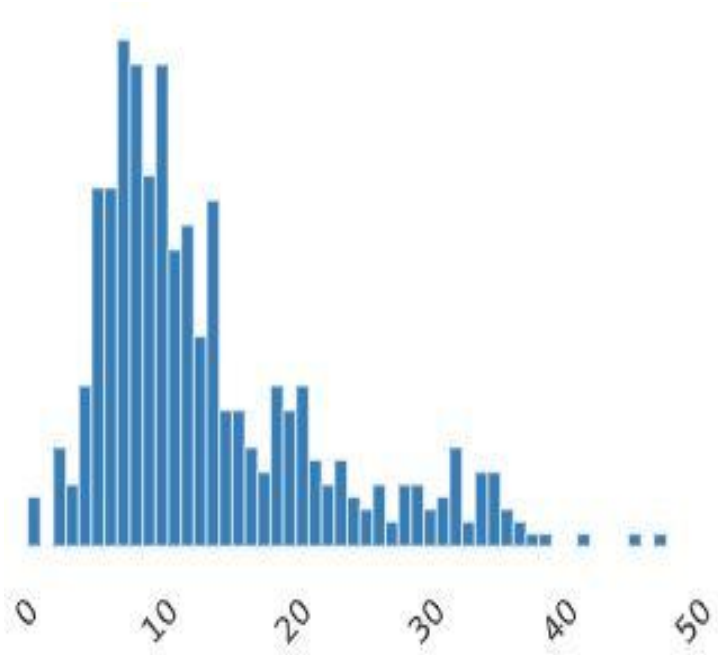
- We are analysing the advanced statistics of 2019 NBA archetypes Stats for different players based on their position.
- https://www.nba.com/stats
- The data set has 30 attributes and 488 instances.

Exploratory Data Analysis

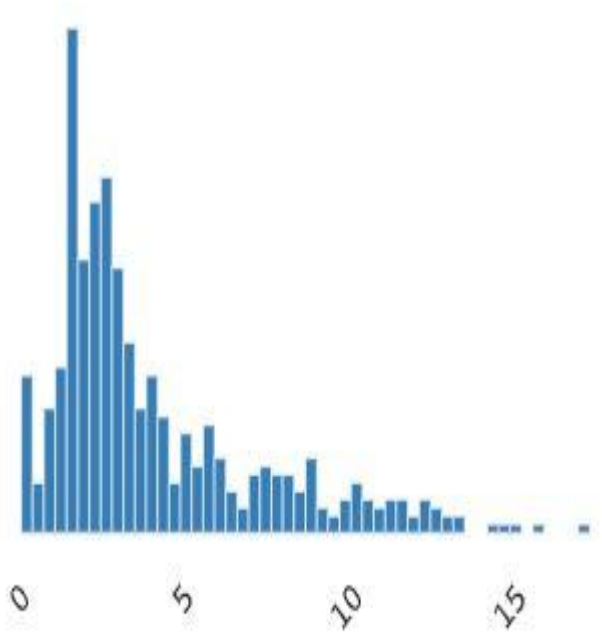
HeatMap



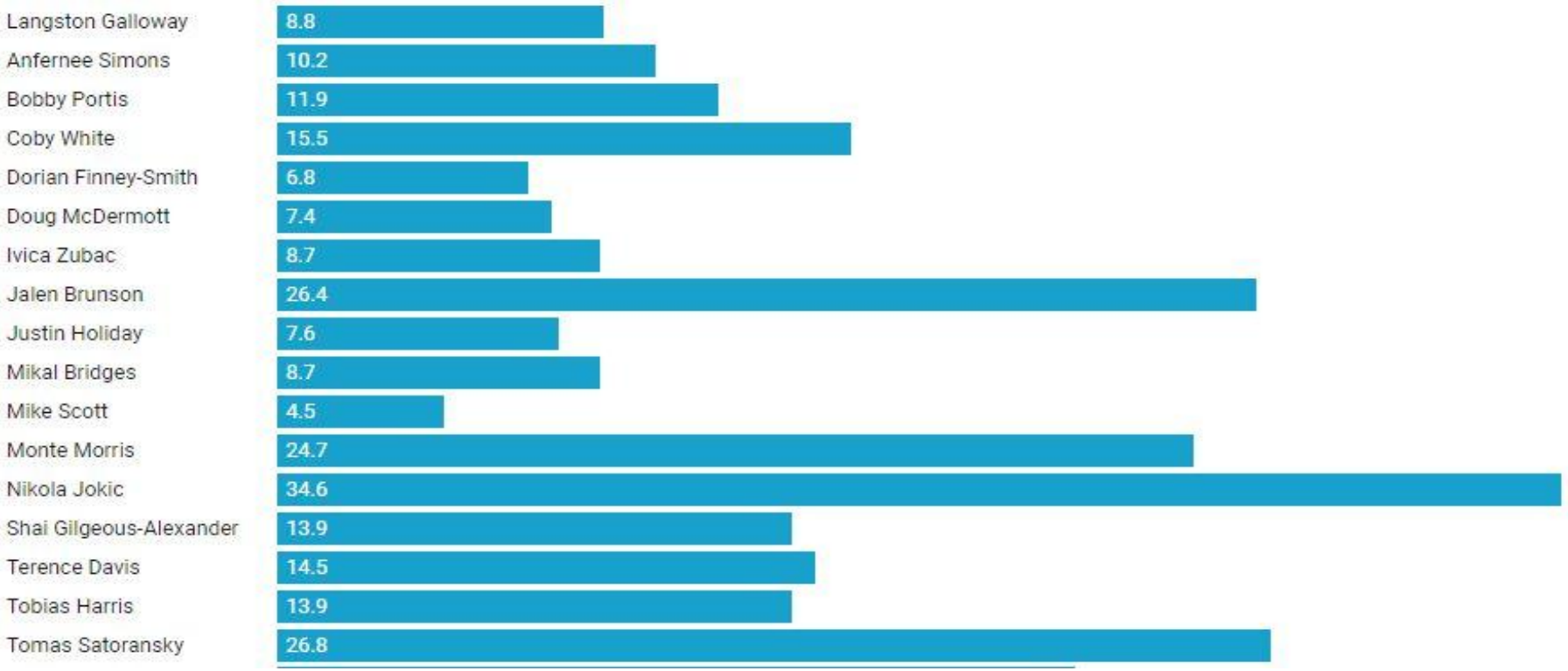
Advanced ASTpercent



Advanced OREBpercent

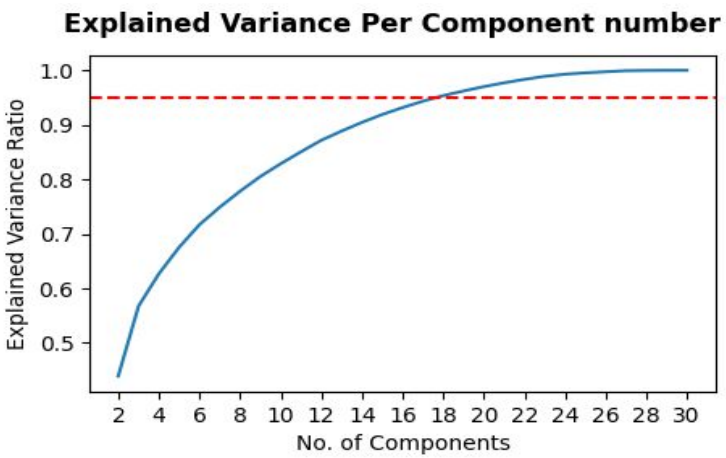


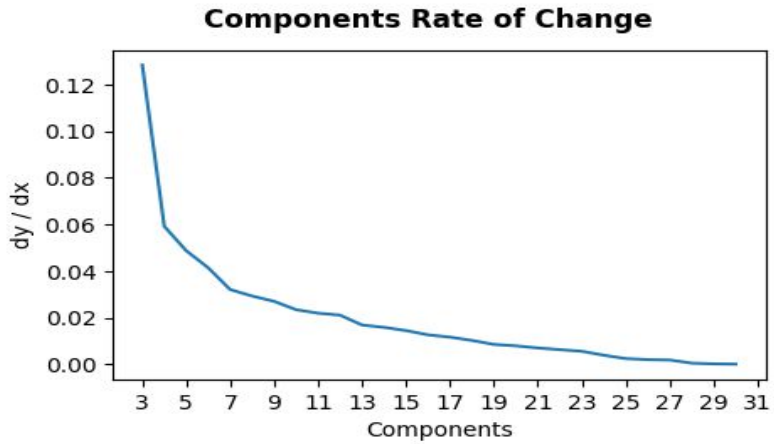
Cont...


Langston Galloway 8.8
Anfernee Simons 10.2
Bobby Portis 11.9
Coby White 15.5
Dorian Finney-Smith 6.8
Doug McDermott 7.4
Ivica Zubac 8.7
Jalen Brunson 26.4
Justin Holiday 7.6
Mikal Bridges 8.7
Mike Scott 4.5
Monte Morris 24.7
Nikola Jokic 34.6
Shai Gilgeous-Alexander 13.9
Terence Davis 14.5
Tobias Harris 13.9
Tomas Satoransky 26.8

Preprocessing Plan

- Perform normalization using standard scaler.
- Principal component analysis (PCA) was used for dimensionality reduction and was useful in clustering the objects into different groups based on the elbow curve.





Experimental Design

Built various clustering models like:

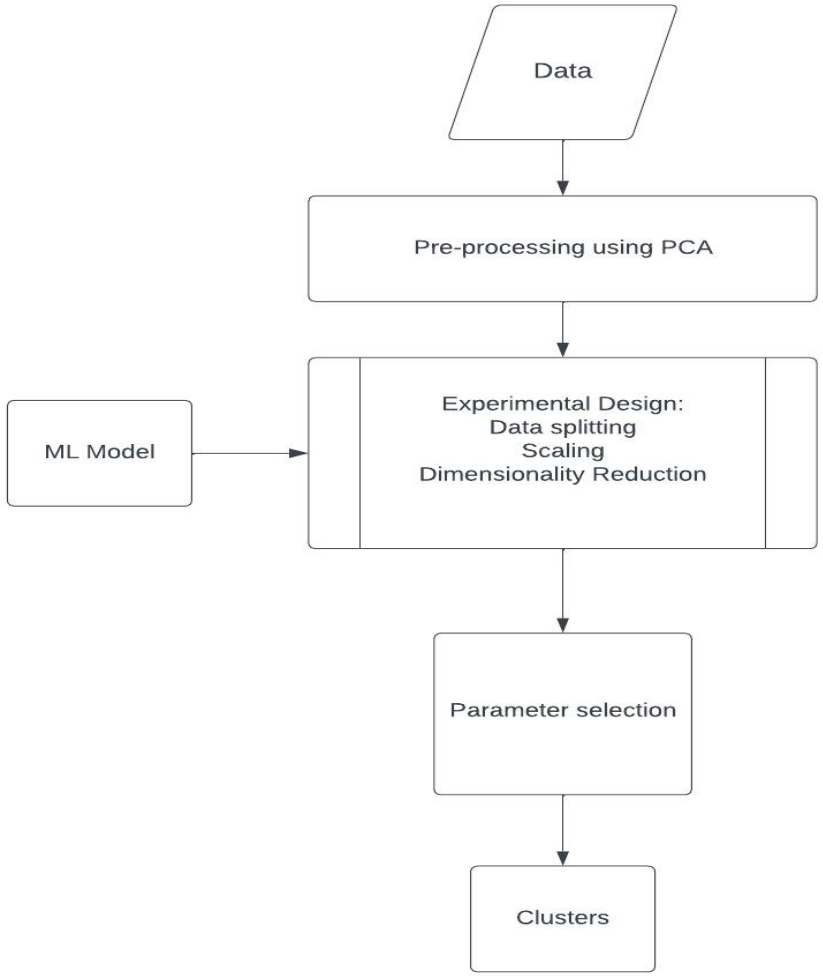
- K-means clustering
- Gaussian Clustering
- DBSCAN

Hyper parameter selection

- elbow method

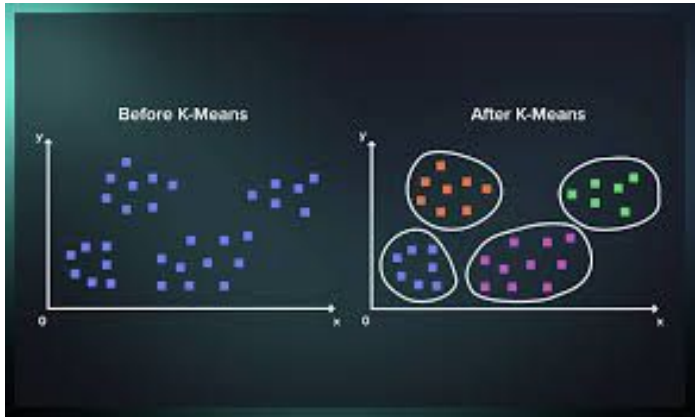
Evaluation Metrics

- Silhouette score

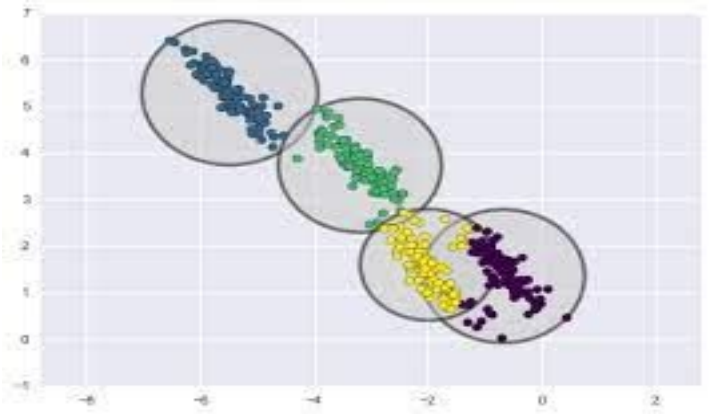


Clustering Methods

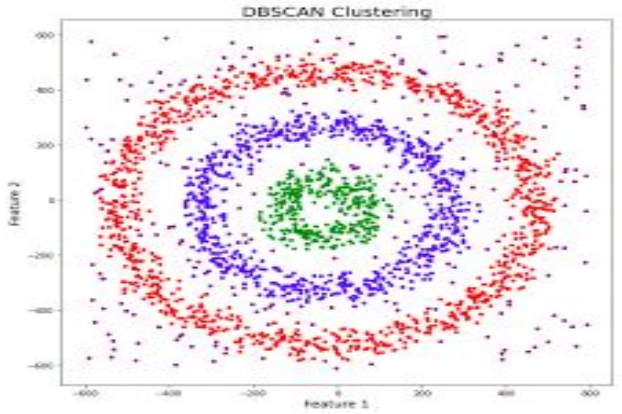
K-Means Clustering



Gaussian Clustering

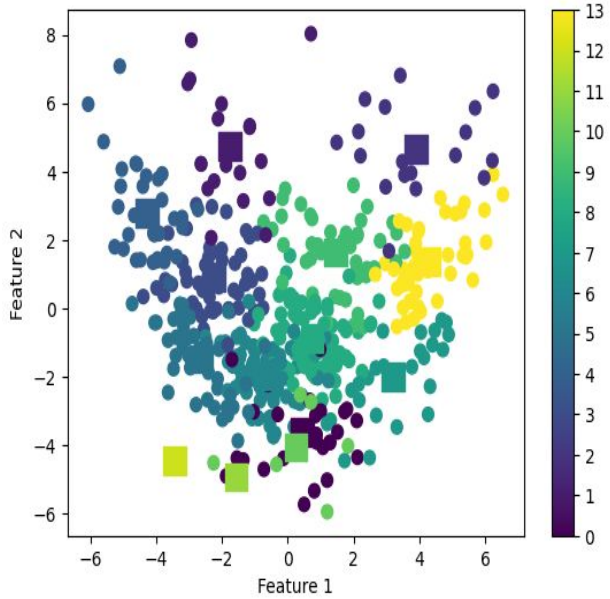


DBSCAN

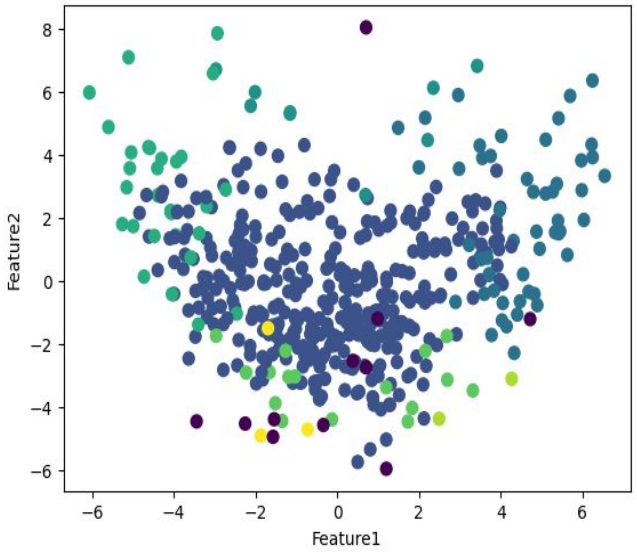


Results and Discussion

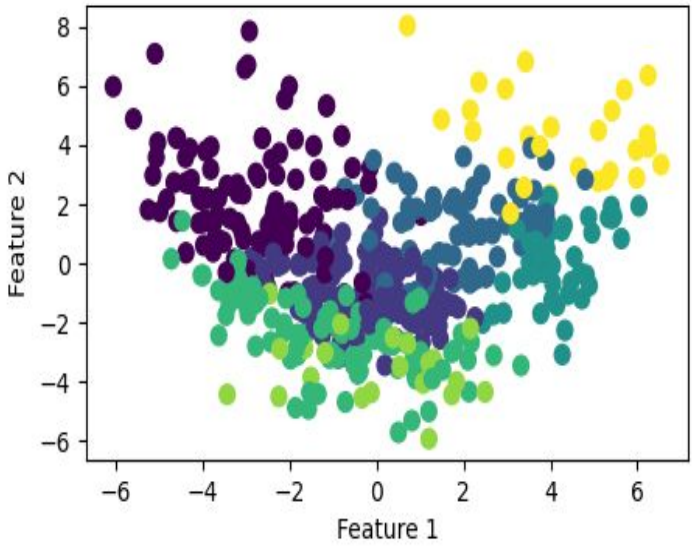
K-Means Clustering



Gaussian Clustering



DBSCAN



Summary

- After the three clustering models are tried and tested, we can interpret that K-Means clustering and DBSCAN have performed better than Gaussian clustering .
- Further, we would also like to use other clustering methods such as Hierarchical clustering for much improvement

References

- https://www.nicksniche.net/machine-learning/identifying-nba-player-archetypes-using-k-means-clustering-part-one
- https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1
- https://scikit-learn.org/stable/modules/mixture.html
- https://www.analyticsvidhya.com/blog/2021/01/a-simple-guide-to-centroid-based-clustering-with-python-code/