NET VALUE

An interactive visualization tool for clustering Soccer Playstyles

MOTIVATION AND OBJECTIVES

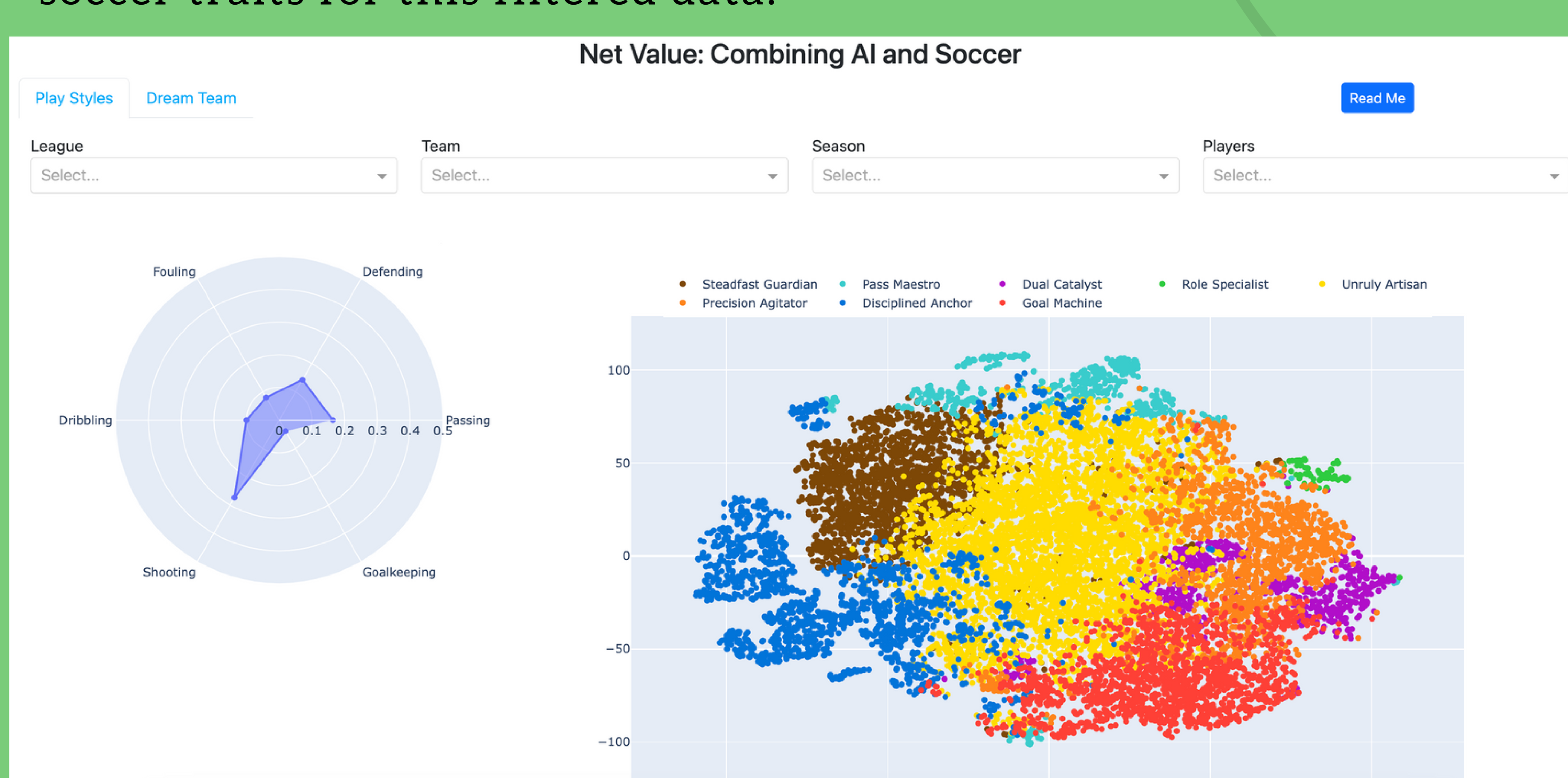
- Traditional soccer positions do not adequately describe the playstyles of soccer players.
- There is no easy way for scouts or managers to visually compare players to find alternatives in case their target players are not available.
- Most sports clustering analyses do not incorporate temporal changes over many seasons, and thus tend to have a risk of recency bias. These analyses tend to cover only a handful of leagues and teams, which limits how much their findings can be generalized.

DATA

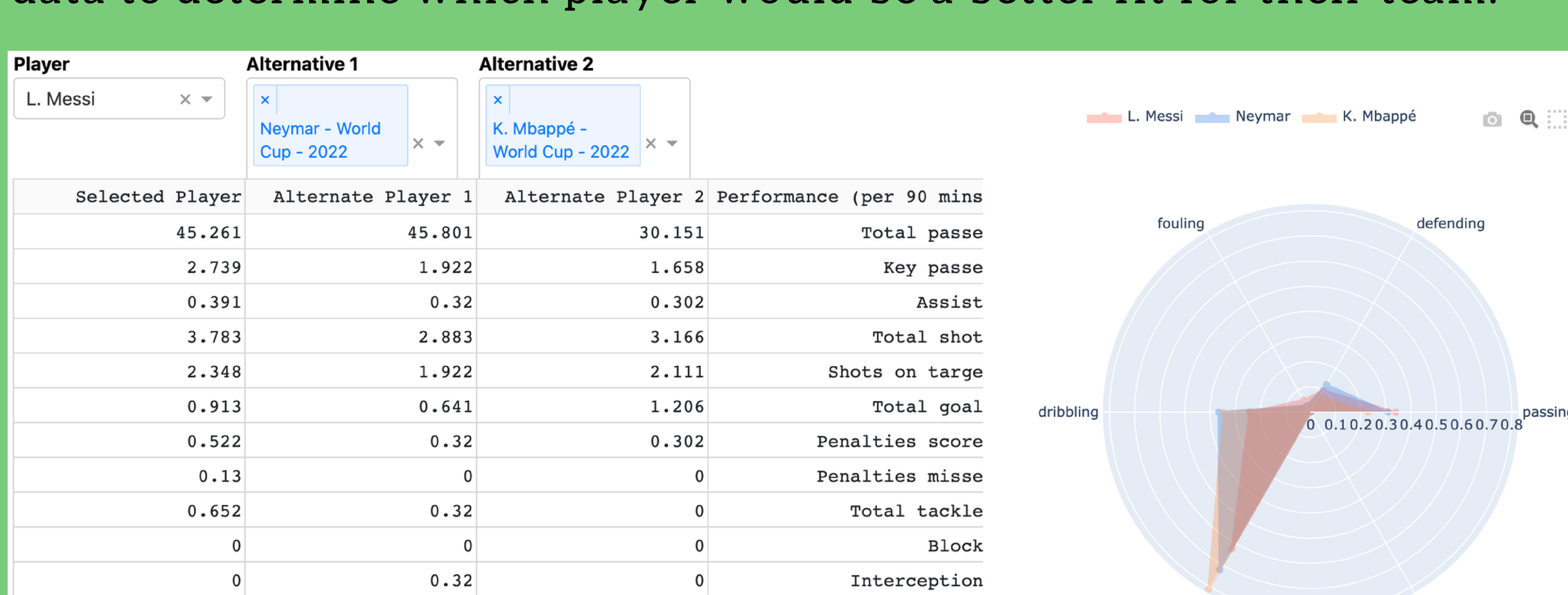
- 5 continents
- 30 countries
- 53 competitions
- 1,217 teams
- 30,498 players
- 10 seasons
- 27 player attributes

APPROACH

- **Data Collection & Processing:** We collected our initial dataset, consisting of 760k rows and 59 columns, from the Football API. We used total stats and minutes played to compute stats per 90 minutes. We removed records with missing values and imputed for outliers. Lastly, we used a Robust Scaler to convert all variables to the same order of magnitude.
- **Clustering Analysis:** We used PCA for dimension reduction, bringing the numbers of our features from 27 down to 15, while retaining 99.5% of the variance of the original data. We then ran k-Means clustering on the PCA results to find play style groupings. Finally, we used t-SNE to visualize our multi-dimensional data into a 2-dimensional plane.
- **Visualization:** Our tool displays a 2-D scatter plot of players clustered by play styles. Users are able to filter the data in this plot for any combination of seasons, leagues, teams, and players, to observe player progression over time. Users can also generate a radar plot of the major soccer traits for this filtered data.



Additionally, our tool lets users compare the attributes of different players to help in the team creation process. Users can easily select the player they wish to compare, choose two other players to compare them against, and analyze the corresponding radar plots and feature data to determine which player would be a better fit for their team.



Selected Player	Alternate Player 1	Alternate Player 2	Performance (per 90 mins
45.261	45.801	30.151	Total passe
2.739	1.922	1.658	Key passe
0.391	0.302	0.302	Assist
3.783	2.883	3.166	Total shot
2.348	1.922	2.111	Shots on targe
0.913	0.641	1.206	Total goal
0.522	0.32	0.302	Penalties score
0.13	0	0	Penalties misse
0.652	0.32	0	Total tackle
0	0	0	Block
0	0.32	0	Interception
12.783	17.616	11.608	Total duel

EXPERIMENTS AND RESULTS

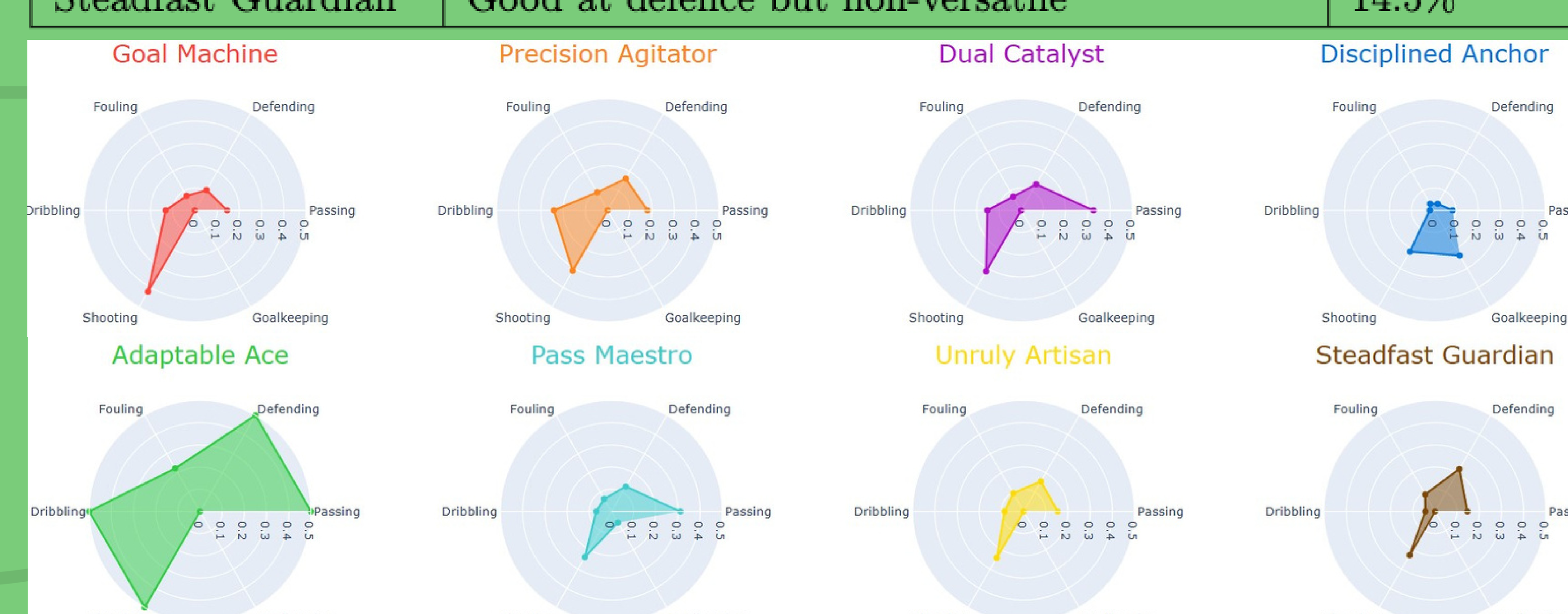
Selecting the number of clusters

- We evaluated the effectiveness of our clustering model using metrics like silhouette score and within-cluster sum of squares.
- We selected **8 clusters** since this had a large silhouette score and low within-cluster sum of squared. This number of clusters also allowed us to find enough unique play styles, while maximizing the relevant clustering metrics.

Play style clustering results

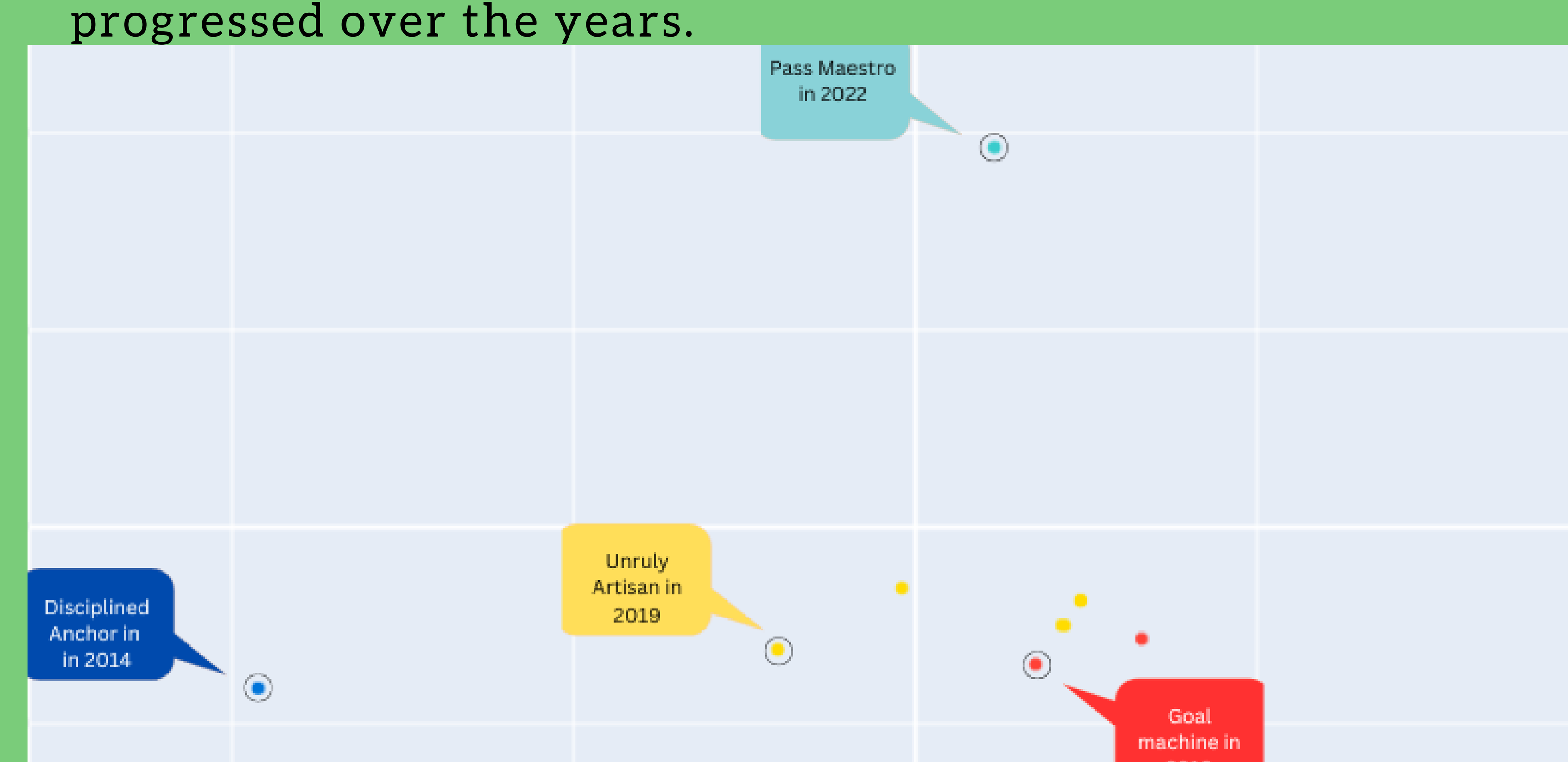
- We interpreted the clusters in terms of player attributes and play style, and assigned each a label and description to define the type of player they describe.

Cluster	Description	Percent of Players
Goal Machine	Excellent in attack with high goal contributions	13.3%
Disciplined Anchor	Low performing but good at game discipline	15.4%
Precision Agitator	Good shooting, excellent dribbling, committing and drawing a lot of fouls	11.3%
Unruly Artisan	Jack of all trades, but poor discipline	34.4%
Adaptable Ace	Extremely versatile performances	0.8%
Dual Catalyst	Good attacking and midfield contributions	5.4%
Pass Maestro	Good passers but mediocre performance level	4.9%
Steadfast Guardian	Good at defence but non-versatile	14.5%



Use Case

- In the below plot for Jonjo Shelvey, we can observe how his playstyle progressed over the years.



Evaluation

- We performed a survey targeting fantasy soccer enthusiasts to evaluate the tool's effectiveness and user experience.
- We ran stress tests to gauge the scalability of the website.

FUTURE EXTENSIONS

- A recommendation tool to suggest similar players to users when creating teams. These can be filtered by league, nationality, season, etc, based on their distance on the scatter plot.
- Include “off-the-ball movement” metrics to capture how players contribute when they are not directly interacting with the ball.

Yash Bhole, Sunil Ravilla, Vamsi Ravilla,
Alberto Debes, Sai Raksith Potluri, Nikolaos Kavouras