

# Player Similarity Modeling: A Case Study on Sergio Busquets

Prabesh Timilsina

April 2025

## Abstract

This study applies data-driven techniques to identify football players statistically similar to Sergio Busquets. Using player attribute data extracted from a public dataset and applying an ETL (Extract, Transform, Load) pipeline in Python, key performance indicators (KPIs) were normalized and analyzed. By establishing a tolerance-based similarity model, the study identifies players whose technical and tactical profiles most closely resemble that of Busquets. Results highlight a small cohort of candidates—notably Axel Witsel, Ivan Rakitić, and Toni Kroos—who demonstrate close alignment across multiple domains. The findings provide a replicable framework for player scouting and comparative analytics.

## 1 Introduction

Sergio Busquets is widely recognized for his unique midfield role, combining elite vision, composure, and positional intelligence. While traditional metrics often underrepresent players with Busquets’ profile, modern data analytics offers new methods for evaluating and comparing player attributes. This study seeks to identify players statistically similar to Busquets by constructing a feature-based similarity model, drawing on normalized attribute data for interpretability and precision.

## 2 Methodology

### 2.1 Data Acquisition and Preprocessing

Player data was sourced from the Kaggle dataset *Football Players Data* (maso0dahmed). The dataset contains comprehensive performance metrics for professional footballers.

An ETL process was implemented using Python:

- **Extract:** Dataset downloaded and unzipped using the Kaggle API.
- **Transform:**
  - Filtered for defensive and central midfield roles (CDM, CM, CB).

- Selected 17 key attributes across offensive, defensive, and spatial dimensions.
- Non-numeric entries were coerced, and missing values imputed via column-wise mean.
- Features were normalized using MinMaxScaler to a 0–1 scale.
- **Load:** Final dataset stored as a CSV for downstream analysis.

## 2.2 Similarity Modeling

Sergio Busquets’ normalized attribute vector was extracted as a baseline. A tolerance threshold of 0.3 was defined across all features, and players whose feature-wise differences remained within this threshold were selected.

The similarity function was defined as:

```
def is_similar(row, busquets_stats, tolerance):
    for stat in busquets_stats.index:
        if abs(row[stat] - busquets_stats[stat]) > tolerance:
            return False
    return True
```

This approach preserves multidimensional comparison without reducing the data to a single distance metric, enhancing interpretability.

## 3 Results

### 3.1 Attribute Comparison: Busquets vs. Average of Similar Players

#### Offensive Profile

Metric	Busquets	Similar Avg
Short Passing	0.94	0.79
Long Passing	0.85	0.78
Vision	0.91	0.75
Composure	0.97	0.74
Dribbling	0.82	0.72
Ball Control	0.90	0.74

### Average vs Sergio Busquets offensive stats

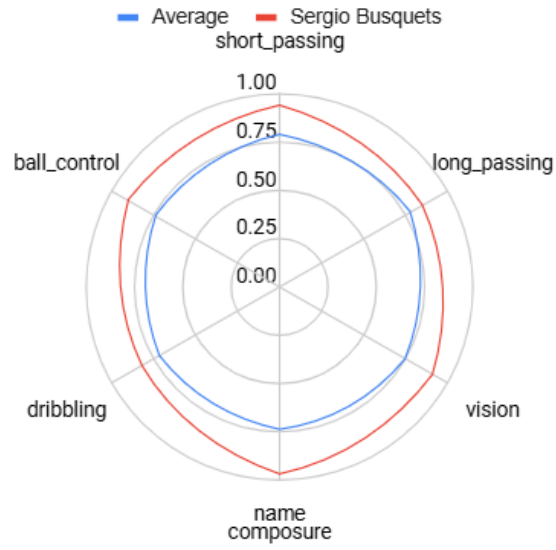


Figure 1: Offensive profile

### Defensive Metrics

Metric	Busquets	Similar Avg
Interceptions	0.94	0.81
Marking	0.95	0.77
Standing Tackle	0.91	0.80

## Average and Sergio Busquets defensive

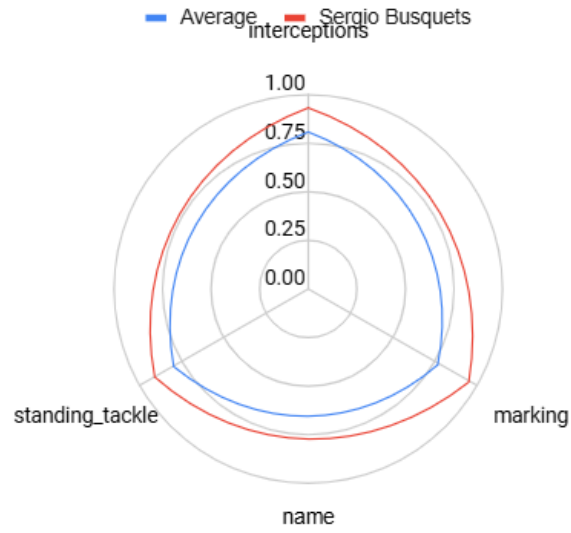


Figure 2: defensive profile

## On the Ball Attributes

Metric	Busquets	Similar Avg
Ball Control	0.90	0.74
Agility	0.61	0.61
Reactions	0.89	0.73
Balance	0.42	0.60

### Average and Sergio Busquets on the ball

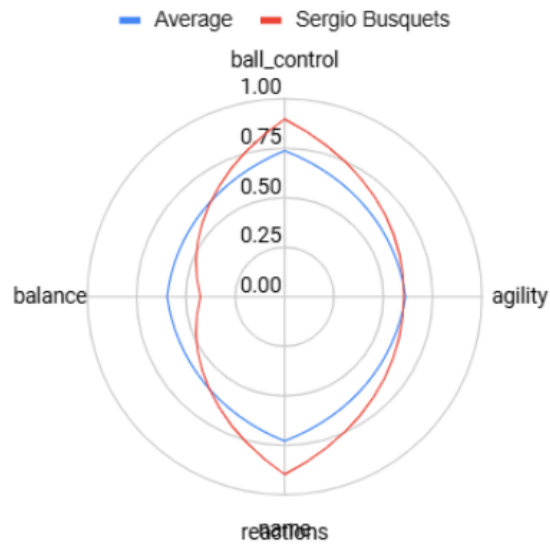


Figure 3: On the ball profile

### Off the Ball Attributes

Metric	Busquets	Similar Avg
Positioning	0.85	0.72
Stamina	0.84	0.75
Marking	0.95	0.77
Interceptions	0.94	0.81

### Average and Sergio Busquets off the ball



Figure 4: off the ball profile

### 3.2 Identified Similar Players

These are the metrics where Busquets seems to be on a league of his own compared to players in similar position and with similar role. A rough comparison is below:

Player	Short Passing	Composure	Ball Control	Marking
A. Witsel	0.94	0.95	0.84	0.87
I. Rakitić	0.91	0.85	0.86	0.79
T. Kroos	0.99	0.89	0.93	0.74

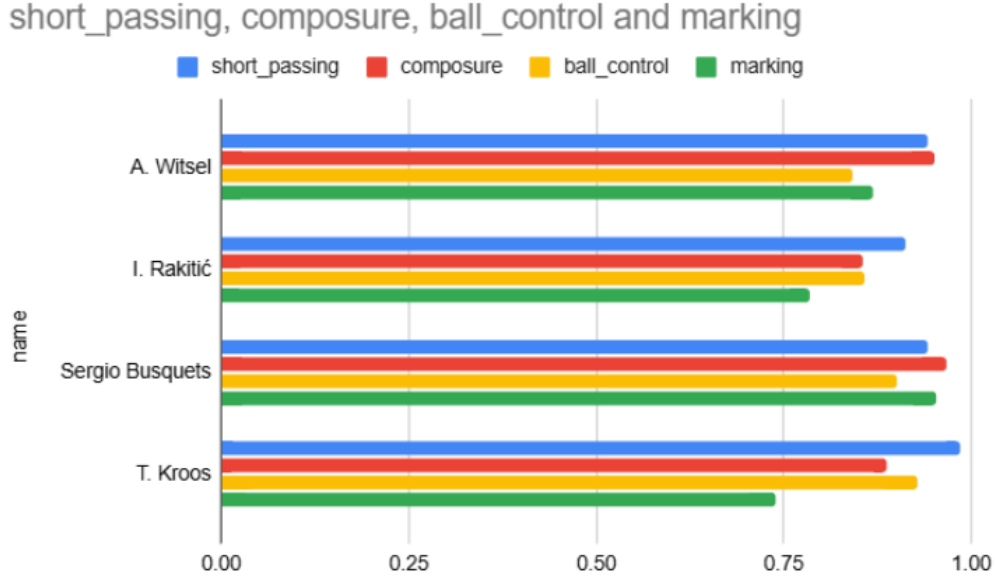


Figure 5: similar profile

## 4 Discussion

The findings confirm the uniqueness of Busquets’ profile—particularly in attributes like composure, vision, and defensive intelligence—while still revealing a set of midfielders who emulate aspects of his playstyle. The similarity approach used here preserves interpretability and avoids the opacity of black-box models, making it useful for real-world scouting and tactical planning.

Interestingly, while players like Toni Kroos show higher values in short passing and ball control, they fall short defensively. Conversely, Axel Witsel offers a more balanced resemblance, albeit slightly below Busquets in composure. The divergence in "balance" for Busquets also points to a stylistic rather than technical variance.

## 5 Conclusion

This study presents a systematic method for player similarity analysis using normalized performance metrics. Sergio Busquets serves as a compelling case study due to his complex and balanced skill set. The methodology can be adapted to various player archetypes and extended to real-time scouting or transfer target identification.

Future work may involve dynamic weight assignments, positional clustering, or the inclusion of match context (e.g., opponent strength, league level) for deeper insights.

## References

- Kaggle Dataset: *Football Players Data* - <https://www.kaggle.com/datasets/maso0dahmed/footba>

- Pedregosa et al., *Scikit-learn: Machine Learning in Python*, JMLR 2011.
- FIFA Ratings Data (Various Sources)