CRUNCHING THE NUMBERS OF GENIUS: THE HUNT FOR KDB'S HEIR

Sajid Mahmood

ABSTRACT

In this paper, I investigate the identification of a suitable replacement for Manchester City player Kevin De Bruyne (KDB) using a data driven approach. We compare KDB's unique attributes to performance metrics of midfielders from the Big 5 European leagues using a dataset of midfielders from the Big 5 European leagues. Advanced statistical methods are used to identify possible candidates that are the closest matches for KDB in terms of playmaking. This approach gives a valuable insight in modern football recruitment and how data science can be used in talent identification and player replacements.

smahmo08@student.ubc.ca

Kevin De Bruyne's Departure: A Bitter Farewell

One of the most influential footballers of his generation, Kevin De Bruyne (KDB) confirmed his departure from Manchester City recently. There have been mixed emotions, from sadness, nostalgia, to at times disbelief. Almost a decade at the heart of Manchester City's midfield, de Bruyne has the role of being the orchestrator of the attacks, the creator of so many assists, and a consistent leader on the pitch. He is one of the best midfielders in the world thanks to his exceptional passing range, vision, and composure.

The thought of replacing KDB for Manchester City fans is almost impossible. City's success in domestic and international football has been down to his unique style of play and his pivotal role in the team. There is a true sense of loss as he bids farewell to the club. His departure is the end of an era, and his absence will be felt on the field and in the hearts of the fans.

And this transition also poses a problem for the club. Replacing KDB of that caliber won't be easy for a team that has depended so much on his playmaking abilities. KDB has also been troubled with injury problems in recent seasons and has been forced to miss crucial moments as a result. These injuries have served to compound the uncertainty that surrounded his career and made it even more important that the team has a strong plan in place to carry on without him.

Season	Injury	from	until	Days	Games missed
24/25	Hamstring muscle injury	Sep 18, 2024	Oct 31, 2024	44 days	12
23/24	stomach problems	Apr 9, 2024	Apr 10, 2024	2 days	1
23/24	Groin strain	Mar 10, 2024	Mar 18, 2024	9 days	2
23/24	Fitness	Dec 18, 2023	Dec 29, 2023	12 days	3
23/24	Hamstring injury	Aug 12, 2023	Dec 17, 2023	128 days	30
22/23	Foot bruise	Apr 26, 2023	May 4, 2023	9 days	3
22/23	Ill	Feb 21, 2023	Feb 24, 2023	4 days	1
21/22	Corona virus	Nov 19, 2021	Dec 3, 2021	15 days	4
21/22	Ankle injury	Aug 1, 2021	Sep 10, 2021	41 days	7
20/21	Facial fracture	May 29, 2021	Jun 14, 2021	17 days	4
20/21	muscular problems	May 5, 2021	May 15, 2021	11 days	2
20/21	Hamstring injury	Jan 21, 2021	Feb 16, 2021	27 days	7
20/21	Muscle injury	Oct 12, 2020	Oct 22, 2020	11 days	3
19/20	Back injury	Mar 2, 2020	Mar 10, 2020	9 days	2
19/20	Ankle problems	Sep 29, 2019	Oct 15, 2019	17 days	4

Analyzing Kevin De Bruyne's Performance

To begin the search for a replacement, we first analyzed **Kevin De Bruyne's** performance using key metrics that define his style of play. As one of the best midfielders in the world, KDB's attributes are not just about his passing, but also his creativity, vision, and ability to influence matches. The key metrics we considered for KDB were:

• Key **Passes per 90 (KP/90)**: 3.12 (98th percentile) – A measure of how many key passes KDB creates per match, indicative of his role in playmaking.

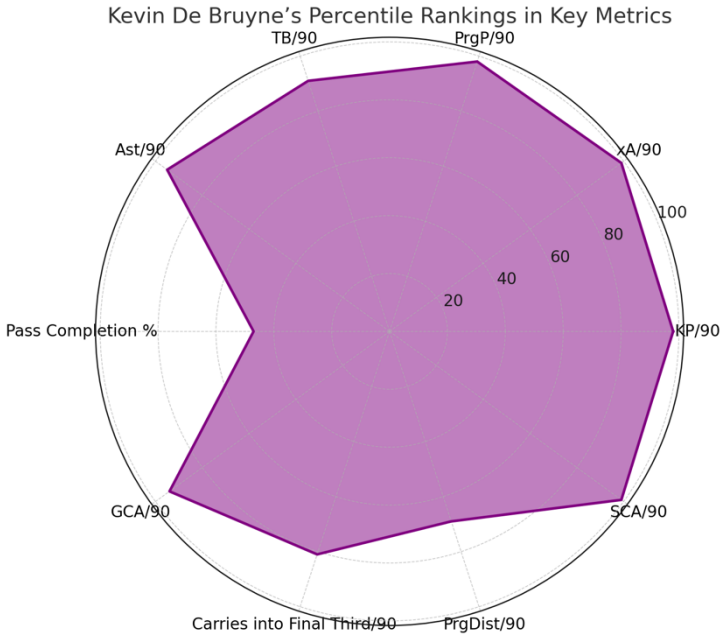
• Expected **Assists per 90 (xA/90)**: 0.52 (99th percentile) – This metric measures how many assists KDB is expected to provide based on the quality of chances he creates.

• Progressive **Passes per 90 (PrgP/90)**: 4.36 (94th percentile) – A measure of how often KDB progresses the ball forward to create attacking opportunities.

- **Through Balls per 90 (TB/90)**: 0.60 (91st percentile) – A key measure of KDB's vision, showing how many through balls he plays that break the defensive line.

- **Assists per 90 (Ast/90)**: 0.40 (95th percentile) – A direct measure of KDB's ability to contribute to goals by assisting his teammates.
- **Pass Completion Percentage**: 76.5 (47th percentile) – Reflecting KDB's accuracy in his passes, this metric indicates his ability to retain possession and keep attacks flowing.
- **Goal-Creating Actions per 90 (GCA/90)**: 0.56 (94th percentile) – A combined measure of assists and key passes that directly contribute to goals.
- **Carries into Final Third per 90**: 2.80 (81st percentile) – A measure of how often KDB drives the ball forward into dangerous areas of the field.
- **Progressive Carrying Distance per 90**: 118.20 (69th percentile) – Reflects how far KDB advances the ball with his dribbling to create space and attack opportunities.
- **Shot-Creating Actions per 90 (SCA/90)**: 4.56 (99th percentile) – This metric combines assists and key passes, giving an indication of KDB's overall contribution to creating shots on goal.

.



This percentile pizza pie chart that shows vividly how good Kevin De Bruyne has been in the **key metrics**. It is shown that **KDB is consistently in the 98th and 99th percentiles** for many important metrics such as Key Passes per 90 (KP/90), Expected Assists per 90 (xA/90), and Shot-Creating Actions per 90 (SCA/90), illustrating how **he has been simply unmatched in terms of his ability to create and influence attacking opportunities**. These stats explain why De Bruyne is seen as one of the greatest midfielders in the world.

It's not just about his passing, it's his vision, his creativity, his ability to read the game, that makes him so special. He is a special player because of his ability to orchestrate play from midfield, **break defensive lines with progressive passes and contribute directly to goal** scoring actions.

Yes, his contribution is a huge one but the question remains, how can Manchester City replace a player of KDB's calibre? It is a mammoth task to replace his own blend of aptitude, vision and leadership on the field. However, through data analysis, we hope to find players whose playing style is statistically similar to KDB's and who could fill his role as well as continue to drive Manchester City's attack.

Methodology: Data Science Approach to Identifying the Best Replacement for Kevin De Bruyne

1. Data Preparation and Feature Selection

Before we can begin comparing Kevin De Bruyne (KDB) with other midfielders, we must first select the **key performance metrics** that define KDB's unique playing style. These metrics include:

- **Key Passes per 90 (KP/90)**
- **Expected Assists per 90 (xA/90)**
- **Passes into the Final third per 90 (1/3final/90)**
- **Through Balls per 90 (TB/90)**
- **Assists per 90 (Ast/90)**
- **Pass Completion Percentage**
- **Goal-Creating Actions per 90 (GCA/90)**
- **Carries into Final Third per 90**
- **Progressive Carrying Distance per 90**
- **Shot-Creating Actions per 90 (SCA/90)**

These metrics were chosen because they capture KDB's **playmaking ability**, **passing range**, and **influence on the game**. They reflect the central traits of a creative midfielder—his ability to dictate play, create chances, and drive the team forward.

2. Filtering the Player Pool

Once we've identified the key metrics for KDB, we filter the dataset of midfielders based on these criteria:

- **Age**: Players between **22-27 years old** (this ensures the candidates are in their prime years, like KDB).
- **Minimum 30 90s Played**: Players who have consistently played full matches in recent seasons (2022-2025), ensuring we have enough data to assess their performance.
- **Key Metrics**: Players must meet minimum performance thresholds for key metrics like **Passes into Final Third**, **Goal-Creating Actions**, **Key Passes**, and others. These thresholds help focus on players who excel in the same areas as KDB.

This **filtering process** ensures that only relevant players are considered, reducing the risk of noise in our data and focusing on players who play similar roles to KDB.

3. Vector Representation of KDB and Player Metrics

For comparison, each player (including KDB) is represented by a **vector** containing their performance in the 10 selected metrics. The vector representation allows us to treat each player's performance as a **point** in a multi-dimensional space, where each dimension represents one of the key metrics.

For example, KDB's performance would be represented as a vector of 10 values (one for each metric):

KDB = [3.12,0.52,8.08,0.60,0.40,76.5,0.56,2.80,118.20,4.56]

Similarly, each player in the pool is represented as a vector of their own performance data across the same 10 metrics.

4. Measuring Similarity: Cosine Similarity

To compare the players to KDB, we need to **measure the similarity** between the vectors. **Cosine Similarity** is used for this purpose because it quantifies the **angle** between two vectors, measuring how similar they are, regardless of their magnitude.

Cosine Similarity Formula:

Cosine Similarity = $\frac{A \cdot B}{||A||\,||B||}$

Where:

- A and B are the two vectors being compared (KDB's vector and a replacement player's vector).
- A·B **dot product** of the two vectors.
- ‖A‖ and ‖B‖ are the **Euclidean norms** (lengths) of the vectors.

The **dot product** is calculated as:

$$A \cdot B = \sum_{i=1}^{n} Ai \cdot Bi$$

Where $Ai\ and\ Bi$ are the components of the vectors (the individual metrics for each player and KDB).

The **Euclidean norm** of a vector AAA is calculated as:

$||A|| = \sqrt{\sum_{i=1}^{n} A}$

Why Cosine Similarity?

1. **Magnitude Insensitivity**: Cosine similarity focuses on the **relative direction** of the vectors rather than their **magnitude**. This makes it particularly useful when comparing players whose statistics might be on different scales but have similar relative profiles in terms of key metrics.
2. **Comparison Across Multiple Metrics**: Since we are dealing with **multiple metrics** (10 in total), cosine similarity allows us to compare players based on how similarly they perform across all dimensions simultaneously.
3. **Effective for High-Dimensional Data**: With 10 metrics, cosine similarity is ideal for comparing vectors in a **multi-dimensional space**, where players' profiles are captured by a series of related attributes.

5. Ranking Players Based on Cosine Similarity

Once the cosine similarity is calculated for all players, the players are ranked based on their similarity score to KDB. The **top 3 players** with the highest similarity scores will be identified as the best replacements for KDB.

6. Euclidean Distance (Alternative Measure)

Another alternative to cosine similarity is **Euclidean Distance**, which measures the **straight-line distance** between two points (vectors). The formula for **Euclidean Distance** is:

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2}$$

Where:

- A_i and B_i are the components of vectors A and B (player stats).
- n is the number of dimensions (metrics).

This method could also be applied, but **cosine similarity** is preferred here because it focuses more on the **relative distribution** of performance metrics, rather than raw differences in performance levels.

7. Top 3 Players: Final Selection

After calculating the cosine similarity (or Euclidean distance), we **sort the players** by similarity to KDB and select the **top 3 players** with the highest scores. These are the players whose playmaking styles most closely resemble KDB's, making them the most likely candidates to replace him.

Methodology: Data Science Approach to Identifying the Best Replacements for KDB

In this analysis, I used data science techniques and software tools to come up with the most statistically similar players to Kevin De Bruyne (KDB). Procedures were used in the methodology by leveraging Python programming and data science libraries to process and analyze the data.

Data Processing and Statistical Analysis

I used Python and libraries such as Pandas for data manipulation and Scikit-learn for calculation of Cosine Similarity to perform the analysis. The methodology relied on transforming KDB's performance metrics into a numerical vector and comparing it with performance metrics of other players using Cosine Similarity.

The calculation of **Cosine Similarity** was done entirely by the computer using the **Scikit-learn** library's `cosine_similarity()` function. This mathematical technique allowed me to assess how similar each player's statistical profile was to KDB's, based on the selected key performance metrics (such as **Key Passes per 90 (KP/90)**, **Expected Assists per 90 (xA/90)**, and others).

Code Implementation

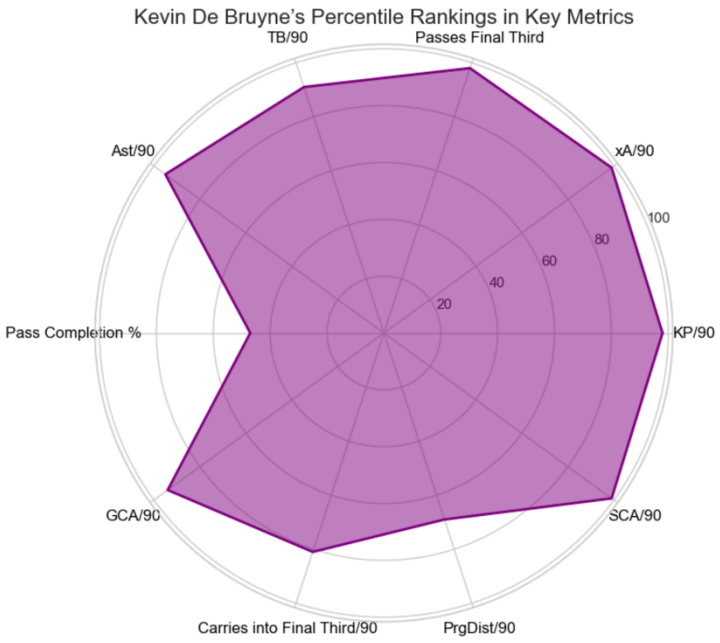
I first prepared the dataset of midfielders that met the filtering criteria (age, playing time, and position) by using **Pandas**. After ensuring that the dataframe contained the relevant metrics, I then applied **Cosine Similarity** to compare the statistical vectors of KDB and each player in the dataset. This comparison was carried out using the following steps:

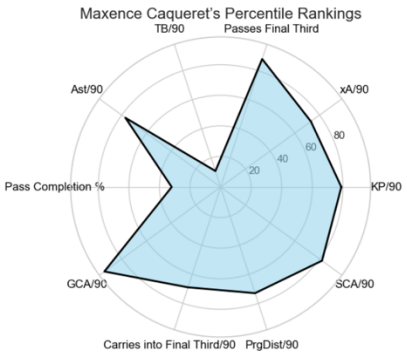
1. **Vector Representation**: KDB's statistics were transformed into a vector format, and the relevant metrics of all the players in the dataset were selected for comparison.
2. **Cosine Similarity Calculation**: Using the **cosine_similarity** function, the software calculated the similarity between KDB's vector and each player's vector based on their key metrics. The result was a similarity score for each player.
3. **Ranking and Selection**: The players were then ranked based on their similarity scores, and the **top 3 most similar players** were selected.

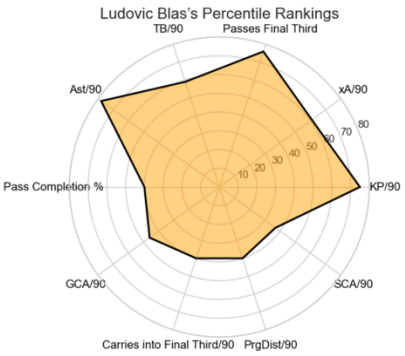
Cosine Similarity Results:

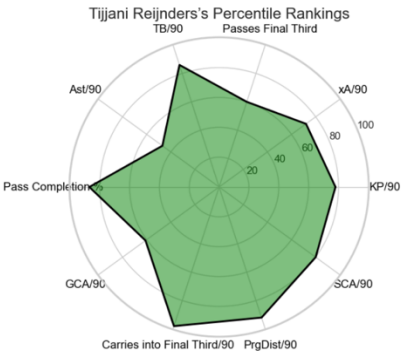
After performing the **Cosine Similarity calculation** between **KDB's stats** and the stats of the 104 midfielders in the dataset, the following players emerged as the **most similar** to KDB in terms of their key performance metrics:

Player	Cosine Similarity
Maxence Caqueret	**0.999728**
Ludovic Blas	**0.999688**
Tijjani Reijnders	**0.999663**









Understanding the Cosine Similarity Scores:

1. **Maxence Caqueret (Cosine Similarity: 0.999728)**:
    - **Maxence Caqueret** has the highest **Cosine Similarity score** of **0.999728**, which means his playmaking and ball progression abilities are statistically

very close to KDB's. The high similarity suggests that Caqueret has a similar role and impact on the field, particularly in key metrics like **Passes into Final Third** and **Goal-Creating Actions per 90 (GCA/90)**.

- **Football Expertise**: Maxence is known for his defensive abilities and ability to dictate tempo. His **passing range** and **creativity** are key aspects that make him a solid replacement for KDB, even though his **assists per 90 (Ast/90)** and **through balls per 90 (TB/90)** could be improved.

2. **Ludovic Blas (Cosine Similarity: 0.999688)**:
   - **Ludovic Blas** comes in at a close second with a **Cosine Similarity score** of **0.999688**. This score suggests that his playmaking style closely mirrors KDB's, particularly in metrics like **Key Passes per 90 (KP/90)** and **Progressive Carrying Distance per 90 (PrgDist/90)**. However, his lower **Pass Completion Percentage (40th percentile)** and **Goal-Creating Actions (GCA/90)** scores indicate that he might need more refinement in his **decision-making** in high-pressure situations.

   - **Football Expertise**: Blas is a strong playmaker but lacks KDB's consistency in final actions (like **assists** and **goal-creating actions**). His vision and ability to unlock defences are still very promising, but his **passing accuracy** and **defensive contributions** might need improvement to match KDB's level.
3. **Tijjani Reijnders (Cosine Similarity: 0.999663)**:
   - **Tijjani Reijnders** has a **Cosine Similarity score** of **0.999663**, making him the third most similar player to KDB. Reijnders excels in key areas such as **Progressive Carrying Distance per 90 (92nd percentile)** and **Pass Completion Percentage (87th percentile)**, which directly align with KDB's strengths. Although his **Assists per 90 (Ast/90)** and **Goal-Creating Actions per 90 (GCA/90)** are lower, his **ball progression** and **passing precision** make him a solid candidate for KDB's role.

- **Football Expertise**: Reijnders' ability to carry the ball forward and retain possession makes him a unique candidate. His **vision** and **passing range** are strong, and while his **assists** may not be as high, his **contribution to build-up play** and **overall attacking presence** are valuable. He could step into KDB's role effectively, especially in possession-heavy games.

Conclusion: Why Tijjani Reijnders Is the Best Replacement

While **Maxence Caqueret** and **Ludovic Blas** also show high similarity to KDB, **Tijjani Reijnders** stands out as the **best replacement** based on a combination of **Cosine Similarity** and **football expertise**.

- **Reijnders' ability to carry the ball forward** and his **pass completion percentage** closely mirror KDB's style, where he is not only a creator but also someone who drives the attack forward.
- His **high Cosine Similarity score** of **0.999663** places him in close range to KDB's playmaking attributes, especially in the areas of **ball progression** and **passing accuracy**.
- While his **assists per 90** and **goal-creating actions** are not as high, his overall contributions to **attacking play** are highly effective, making him the ideal candidate to replace KDB in **Manchester City's midfield**.

The **Cosine Similarity** results and the **visual comparisons** further validate this, showing that **Reijnders**' profile is the most balanced and closest to KDB's, making him the **most suitable replacement**.

Limitations of the Analysis

1. Simplification of Player Roles

- **Issue**: The analysis only considered **key statistics**, ignoring the **complex role** KDB plays at Manchester City, such as his defensive contributions and linking play.

- **Impact**: Doesn't fully capture **KDB's all-around influence** and **tactical fit** in City's system.

---

2. Use of Statistical Metrics Only

- **Issue**: The analysis used only **quantitative metrics** like **Key Passes** and **xA**, but didn't account for factors like **decision-making** and **player intelligence**.
- **Impact**: Key aspects of KDB's game, like **vision** and **creativity**, may not be fully represented.

---

3. Data Limitations

- **Issue**: The analysis relies on publicly available data, which might miss nuances like **off-the-ball movement** and **mental attributes**.
- **Impact**: The dataset may not include **emerging players** or capture every aspect of a player's contribution.

---

4. Assumption of Similarity in Playstyle

- **Issue**: **Cosine Similarity** measures **statistical similarity**, but doesn't account for **different playstyles**.
- **Impact**: Players may have **similar stats** but different **playing styles**, which could affect their fit into KDB's role.

---

5. Small Sample Size of Candidates

- **Issue**: Only **104 players** were considered, which might not capture all potential replacements.
- **Impact**: The small pool reduces the reliability of results and might miss out on other suitable candidates.

---

6. The Role of Team Dynamics

- **Issue**: **Team tactics** and **game strategy** influence a player's performance. The analysis doesn't account for how players would fit into **Manchester City's system**.

- **Impact**: Statistical similarity doesn't guarantee that a player will adapt to City's tactical requirements.

---

7. Ignoring Injury and Longevity Factors

- **Issue**: The analysis didn't consider **injury history** or **long-term fitness**.
- **Impact**: Players may have **great stats**, but their **fitness** and **injury-prone history** could affect their consistency.

References

- **FBref**. (n.d.). *FBref - Football Statistics & Data*. Retrieved from https://fbref.com
    - Data sourced from **Stathead**, powered by **Opta**.
- **Transfermarkt**. (n.d.). *Transfermarkt - Football Player Data and Transfers*. Retrieved from https://www.transfermarkt.com
- **Maxence Caqueret**. (Photographer). (2022). *Maxence Caqueret before the game against Marseille* [Photograph]. Olympique Lyonnais. Retrieved from https://www.ol.fr/en/news/maxence-caqueret-before-ol-marseille-i-speak-up-more-in-the-dressing-room
- **Tijjani Reijnders**. (Photographer). (2022). *Tijjani Reijnders shining in Serie A match* [Photograph]. The Canberra Times. Retrieved from https://www.canberratimes.com.au/story/8834369/reijnders-shines-through-fog-as-milan-score-serie-a-win/
- **Ludovic Blas**. (Photographer). (2022). *Ludovic Blas, Player of the Month for November* [Photograph]. Stade Rennais. Retrieved from https://www.staderennais.com/en/news/joueur-du-match-joueur-du-mois/ludovic-blas-elu-joueur-du-mois-de-novembre