

## **Moisés Caicedo Transfer Valuation Analysis**

### **MGEC45 – Applied Econometrics Project**

**Author:** Mohammed Tajudeen Onigbanjo

---

## **1. Introduction**

In the summer of 2023, Chelsea FC completed the record-breaking £115 million transfer of Moisés Caicedo from Brighton & Hove Albion, making him the most expensive midfielder in Premier League history at the time. The size of this fee immediately sparked debate among fans, analysts, and economists alike: does Caicedo's on-pitch performance truly justify such a valuation?

This project aims to approach that question from an econometric and data-driven perspective. Rather than relying on subjective narratives or isolated performance metrics, I construct a systematic framework that evaluates Caicedo relative to a large population of midfielders across Europe's top five leagues. Using regression-based residual analysis, composite scoring methods, and interactive visualizations, the project seeks to determine whether Caicedo meaningfully outperforms a statistically defined baseline and whether that performance aligns with his market price.

Beyond answering a single valuation question, this project reflects my broader interest in using economic theory and statistical modelling to support data-driven decision-making in real-world contexts. In particular, it emphasizes how carefully constructed empirical frameworks can move discussions away from intuition and narrative toward evidence-based evaluation, allowing data to meaningfully inform decisions with real economic impact.

---

## **2. Data Description and Preparation**

The dataset used in this analysis consists of midfielder-level performance statistics from Europe's top five leagues between the 2022-2023 seasons. Players were filtered to ensure comparability in terms of position and minutes played, resulting in a final sample of approximately 1,000 midfielders.

Key variables include offensive, defensive, and overall contribution metrics, along with contextual controls such as playing time, league and team effects. Extensive data cleaning was required, including handling missing values, standardizing player names across sources, and removing duplicate observations. Additional derived variables, such as regression residuals and composite scores were created and stored separately to preserve transparency and reproducibility.

All my personal cleaned datasets, scripts, and intermediate outputs are documented in the project repository to ensure that the analysis can be replicated or extended.

---

### **3. Methodology**

#### **3.1 Regression Framework and Residual Analysis**

To evaluate player performance relative to expectations, I estimate a series of linear regression models that predict offensive, defensive, and overall contributions based on relevant explanatory variables. The residuals from these regressions represent the portion of a player's performance not explained by observable characteristics effectively capturing whether a player over- or under-performs relative to the model's baseline.

Residual analysis is particularly well-suited to this context because it allows for direct comparison across players while controlling for structural differences in role and opportunity. Positive residuals indicate overperformance, while negative residuals suggest underperformance. Here are the regressions that were ran:

##### **Overall Impact model:**

$$(+/-)\text{per90} \sim \text{Recoveries}/90 + \text{KeyPasses}/90 + \text{ProgressiveCarries}/90 + \text{Touches}/90 + \text{PassesCompletedPercentage} + \text{Team Fixed Effect}$$

##### **Defensive contribution model:**

$$(\text{xGA when on})/90 \sim \text{TacklesAndIntercepted}/90 + \text{Recoveries}/90 + \text{TacklesInDefensive3rd}/90 + \text{TacklesInAttack3rd}/90 + \text{TacklesInMidfield3rd}/90 + \text{TacklePercentage} + \text{Team Fixed Effects}$$

##### **Offensive Contribution Model:**

$$(\text{xG when on})/90 \sim \text{KeyPasses}/90 + \text{PassesCmp}/90 + \text{ProgressiveCarries}/90 + \text{PassCompletedPercentage} + \text{Team Fixed Effects}$$

##### **Market Value model:**

$$(\text{MarketValue}) \sim \text{OverallImpactModelResidues} + \text{DefensiveImpactModelResiduals} + \text{OffensiveImpactModelResiduals}$$

##### **Market Value model (controlled for age):**

$$(\text{MarketValue}) \sim \text{Age} + \text{MInutesPlayed} + \text{Recoveries}/90 + \text{TacklesAndIntercepted}/90 + \text{ProgressiveCarries}/90 + \text{KeyPasses}/90 + \text{Team Fixed Effects}$$

### **3.2 Scoring and Ranking**

To complement the residual analysis, I construct a composite scoring system that aggregates performance across multiple dimensions. Players are ranked based on these scores, allowing Caicedo's relative position within the broader midfielder population to be assessed.

Importantly, the scoring framework is not intended to produce a single "true" ranking, but rather to provide an additional lens through which relative performance can be evaluated. Here is how I produced the scores value:

#### **Defensive Score:**

TacklesAndInterception/90 , Recoveries/90

#### **Progressive Score:**

ProgressiveCarries/90 , ProgressivePasses/90

#### **Onball Score:**

PassesCompleted/90, PassesCompletedPercentage

#### **Creative score:**

KeyPasses/90 , xGWhenOnThePitch/90

---

## **4. Results**

The residual analysis suggests that Moisés Caicedo performs broadly in line with the modeled baseline across offensive, defensive, and overall impact dimensions. As illustrated by the residual distributions, Caicedo is positioned near the center of the sample in both offensive and overall impact residuals, indicating performance close to what would be expected given his observable characteristics. His defensive residual is modestly positive, suggesting slight overperformance relative to the baseline; however, this advantage is not large enough to place him among the extreme upper tail of midfielders. Taken together, the residual results imply that Caicedo is a dependable and effective player, but not a statistical outlier when performance is evaluated relative to expectations.

The composite score analysis provides important additional context and helps reconcile these findings with Chelsea's willingness to pay a substantial transfer fee. While residuals capture deviations from an expected baseline, the score-based framework evaluates absolute performance across multiple dimensions simultaneously. Under this approach, Caicedo ranks extremely highly in several key areas: he places in the 91.96th percentile for defensive score, the 94.94th percentile for on-ball contribution, the 77.78th percentile for creativity, and the

73.04th percentile for progressive actions. When compared across the full population of midfielders, only three players outperform Caicedo across all score dimensions simultaneously.

These results suggest that Caicedo's valuation is not driven by exceptional overperformance relative to expectations, but rather by his rare combination of consistently high-level contributions across defensive, on-ball, and transitional metrics. In this sense, the score analysis complements the residual findings by highlighting that while Caicedo may not dramatically exceed statistical baselines, he offers a balanced and elite skill profile that is scarce within the market. This distinction helps explain why his transfer fee appears difficult to justify using residuals alone, yet more defensible when viewed through a multidimensional performance lens.

---

## 5. Visualization and Power BI Dashboard

To enhance interpretability and accessibility, I developed an interactive Power BI dashboard that visualizes composite scores, the score percentile and the market residual. The dashboard uses Caicedo as a fixed baseline, allowing users to select another player for direct comparison across multiple performance dimensions.

Rather than embedding all interpretation within the dashboard itself, the visuals are designed to complement the written analysis. This separation allows the report to provide nuanced explanation, while the dashboard serves as an exploratory tool for users interested in player-to-player comparisons.

---

## 6. Discussion

The central takeaway from this analysis is not that Moisés Caicedo is an underperforming player, but rather that his transfer fee cannot be fully justified by observable performance metrics alone.

This gap highlights an important feature of football labor markets: prices often reflect factors beyond current statistical output. Age, positional scarcity, contract length, home-grown status, resale potential, and club-specific strategic considerations all play significant roles in shaping transfer fees. Caicedo's valuation may therefore be better understood as an investment in future potential and stability rather than a payment for past performance.

From an econometric perspective, this underscores the limitations of purely performance-based valuation models and the importance of interpreting results within a broader institutional context.

---

## **7. Personal Reflection**

One of the most significant challenges in this project was data acquisition. Unlike many other sports, football lacks a centralized and readily accessible performance and valuation dataset, requiring extensive filtering, cleaning, and merging across multiple sources. This process highlighted how easily measurement inconsistency and data harmonization issues can arise, reinforcing the importance of transparency and careful validation in empirical work.

Market value data posed an additional challenge, as available measures are often noisy proxies rather than true transaction prices. This limitation forced me to think more critically about how transfer fees reflect not only player performance, but also market expectations, bargaining power, and institutional context.

From a methodological standpoint, I came to realize that while regression models are effective tools for isolating conditional relationships between performance metrics and outcomes, they were not always sufficient for capturing a player's overall impact. In contrast, the composite score evaluation provided a clearer and more intuitive picture of Caicedo's value by aggregating performance across multiple dimensions, which helped rationalize Chelsea's willingness to pay a premium.

If I were to extend this analysis, I would explore alternative model specifications, such as non-linear or multilevel regressions, to better capture heterogeneous effects across players and leagues. I would also place greater emphasis on contextual variables that influence transfer values, particularly club-level strategies. For example, clubs such as Brighton specialize in acquiring young talent, developing players, and selling them at substantial premiums, suggesting that transfer fees reflect institutional reputation and market signaling in addition to on-pitch performance.

---

## **8. Conclusion**

Using econometric modelling and comparative performance analysis, this project evaluates whether Moisés Caicedo's record-breaking transfer fee aligns with his measurable on-pitch contributions. While Caicedo emerges as a competent and reliable midfielder, the results suggest that his valuation is driven more by market dynamics and future expectations than by exceptional statistical performance.

Ultimately, this analysis demonstrates both the power and the limits of data-driven valuation in football markets, and highlights the importance of combining quantitative tools with economic intuition.