Overview

The Football Development Index (FDI) is a weighted composite score on a scale from 0 to 100 that quantitatively reflects the overall level of football development in a city, with higher values indicating stronger infrastructure, participation, and industry maturity. It combines multiple football-related factors, ranging from youth participation and government spending to market engagement and professional exports, into a single score.

Each factor represents a measurable aspect of football growth. Since factors use different units (e.g., dollars, number of clubs, percentage of population), they are first normalized to a comparable 0–1 scale, and then combined using weighted averaging to reflect their relative importance.

Why Machine Learning?

The Football Development Index (FDI) is calculated using machine learning instead of manual methods to achieve greater fairness, accuracy, and adaptability. Rather than relying on human judgment to combine and weigh each factor, machine learning allows data to be processed in a consistent and unbiased way. It can handle complex datasets, automatically normalize values, and adjust the influence of different factors based on new information. As more data about football development becomes available—such as changes in participation, investment, or league performance—the model can learn and refine itself over time. This makes the FDI not just a static measure, but a dynamic and evolving reflection of how football is growing across cities around the world.

Weighing Procedure

Each factor contributes differently to football development, so a weighted system is applied to reflect their relative importance.

For example, structural and performance-related indicators such as *Government Spending on Football Development* or *National Top-Tier League Ranking* may have greater influence than spectator behavior or media engagement.

Weights are assigned on a scale typically between 1.0 and 2.0, based on expert evaluation of their impact.

To ensure comparability across all indicators, the raw data for each factor is normalized to a scale of 0–1 using the Min–Max normalization method:

$$\chi_{normalized} = \frac{\chi - \chi_{min}}{\chi_{max} - \chi_{min}}$$

This process prevents units (such as dollars or percentages) from biasing the outcome and ensures each factor contributes proportionally.

For factors where lower raw values are better (if applicable, e.g. league ranking number where a lower rank is better), we invert after normalization:

$$\chi_i^{inverted} = 1 - x_i$$

Categorical/qualitative factor (Education) is mapped to numeric values: High = 1.0, Moderate = 0.5, Low = 0.2.

Why weighted average?

Weights let us reflect the relative importance of each factor (for example, youth participation and registered players are heavily weighted because they reflect long-term supply of talent; public or private investment and transfer market value reflect financial backbone). The denominator ΣW_i ensures the final score is comparable regardless of how weights are scaled.

Missing data & robustness

If a city lacks a value for a factor, that factor is excluded from both numerator and denominator for that city (i.e., divide by the sum of weights of available factors only).

Recommend flagging results where more than 20% of factors are missing.

Limitations & caveats

The FDI is an index . It aggregates diverse measures into a single score. It is useful for comparison but not a substitute for deeper qualitative study.

Normalization depends on the sample of cities: adding/removing cities will change min/max and therefore normalized values. For stable comparisons, define a fixed reference range or update consistently.

FDI Index Formula

After normalization and weighting, the Football Development Index is calculated as a weighted average of all selected factors:

$$FDI = \frac{\sum_{i=3}^{n} (W_i \times S_i)}{\sum_{i=3}^{n} W_i} \times 100$$

Where:

$W_i = weight\ of\ factor\ i$

$S_i = normalized\ score\ (0–1)\ of\ factor\ i$

$n = number\ of\ factors\ included$

The final value is multiplied by 100 to convert it into a percentage-based scale (0–100), making the score intuitive:

Range	Category	Interpretation
0–30	Emerging Markets	Limited infrastructure and participation, football in early development.
30–50	Moderate Development	Growing ecosystem with increasing investment and youth engagement.
50-100	High Development	Mature, balanced football environment with sustainable performance.

Why this method?

The weighted normalization method is widely used in composite indices (such as HDI and GDI) because it:

- Allows integration of diverse data types under a unified scale.
- Ensures comparability across cities with different baselines.
- Reflects relative importance via customizable weights.
- Maintains transparency and reproducibility, as every step can be verified and replicated.

Simplified Example

Assume a city has normalized values:

Factor	Normalized (x_i)	Weight (w_i)
Government spending	0.8	2
Youth participation	0.7	2
Number of clubs	0.65	1.5
Market share of football content	0.9	1.2

Then:

$$FDI = \frac{(0.8\times2.0)+(0.7\times2.0)+(0.65\times1.5)+(0.9\times1.2)}{2.0+2.0+1.5+1.2} \times 100 = 75.4$$

The city's FDI = 75.4, placing it in the *High Development* football ecosystem category.

Key note on interpretation

A higher FDI index reflects a city or region that has effectively integrated football into its economic, social, and institutional frameworks—supporting not only elite performance but also grassroots engagement and sustainable growth.

The FDI Index does not measure “success” in football, but rather the structural capacity for sustained football development — a city’s ability to grow talent, infrastructure, and market presence over time.

Factors used and weights

Factor (column name)	Weight (w_i)	Short description
Government spending on football-related development ($)	2	Public capital and operating investment in stadiums, training centres, grants.
Investment scale from private/commercial sectors	1.2	Private stadiums, club-owned facilities, commercial projects.
Registered football players	2	Total registered players (youth + adult).
Youth football participation rate	2	% of youth regularly participating in football programs.
Number of football clubs in the city	1.5	Clubs at professional, semi-pro and organized amateur levels.
Percentage of population watching football regularly (TV/online)	1.2	Local viewership demand for football content.
Average ticket sales for local events (USD)	1.4	Commercial viability and matchday engagement.
Number of professional players exported abroad	1.5	Talent export as a proxy for player quality & development.
National top-tier league international ranking	2	Strength/competitiveness of the country’s top division.

Registered football coaches and referees	1.8	Local coaching/refereeing capacity and quality.
Total market value of domestic transfers	2	Financial activity and circulation within the domestic market.
Frequency of major football events hosted annually	1.5	Event hosting capacity and exposure.
Market share of football content	1.2	The share of local sports media consumed is football.
Presence of football leagues in education system	1.8	Degree of formal school/academy integration (High→1.0, Moderate→0.5, Low→0.2).