



Inter Miami CF Data Project: MLS Center Backs

David Goldstein



Prompt

Inter Miami CF seeks the best center backs (CB's) in MLS

Traits particularly relevant to the team's playing style include:

- Possession retention
- Consistent defending

The club will assign less weight to:

- Aggressive interception of passes

The club will not evaluate:

- Contribution to goals scored
- Player Salary





Table of Contents

Process

- Data Capture
- Model Development
- Validation

Selections

- Guram Kashia
- Graham Smith
- Matt Hedges
- Maxime Chanot
- Daniel Steres



Analytical Process





Data Capture: A Unique Challenge

Challenges:

- Unavailability of MLS player defensive data
- Lack of access to Opta Provision
- WhoScored.com had no option to download its MLS data and does not allow auto-clicking of the “next” button to view more players

Creative Solution:

- Utilize advanced computer programming to scrape data from WhoScored.com by automatically scrolling through search filters

Major League Soccer Player Statistics								
Summary		Defensive		Offensive		Passing		Detailed
Category:	Shots	Sub category:	Zones					Accumulation Per game
View:	Overall	Home	Away	Appearances	All	Minutes	0 - 90	
Age:	All	Positions	1	Nationality	All			
R	Player	Apps	Mins	Total	OutOfBox	SixYardBox	PenaltyArea	Rating
1	Joseph Mora DC United, 26, D(L)	1	90	-	-	-	-	7.29
2	Dejan Jakovic Los Angeles FC, 34, D(C)	2	180	-	-	-	-	7.28
3	Donovan Pines DC United, 21, D(C)	9	772	0.2	-	0.1	0.1	7.19
4	Francisco Calvo Chicago Fire, 27, D(CL)	16	1440	1.3	0.4	0.3	0.6	7.14
5	Ike Opara Minnesota United, 30, D(C)	23	2064	0.9	-	0.2	0.6	7.13
6	Miles Robinson Atlanta United, 22, D(C)	27	2430	0.6	-	-	0.5	7.13
7	Auro Toronto FC, 23, D(R),M(R)	1	90	-	-	-	-	7.12
8	Keegan Rosenberry Colorado Rapids, 25, D(R)	3	270	0.7	0.3	-	0.3	7.09
9	Tristan Blackmon Los Angeles FC, 23, D(CR)	5	450	-	-	-	-	7.05
10	Graham Smith Sporting Kansas City, 23, D(C)	8	691	0.6	-	0.1	0.5	7.04

© WhoScored

Page 1/13 | Showing 1 - 10 of 123 first | prev | next | last





Data Capture: Automated Program Components Pt. 1



- Step 1 - Confirm WhoScored.com security compliance
- Step 2 - Connect to the WhoScored.com 2019 MLS player stats webpage
- Step 3 - Exclusively filter the search table for center backs
- Step 4 - Change metrics to their "Per 90 Minutes" values





Data Capture: Automated Program Components Pt. 2

- Step 5 – Scroll through combinations of ages and nationalities to find each CB
- Step 6 – For each CB, scroll through all statistical categories and store the web script for each metric
- Step 7 – Convert stored information into database
- Step 8 – Filter out players from database with fewer than 500 minutes played in 2019





Model Development: Choosing metrics to use for model

- Remove overlapping metrics
 - i.e. including both “assists” and “corner assists” double-counts “corner assists,” so one of them must go
 - In this case, it is better to keep “assists” rather than the subcategories of assists - the simpler the inputs, the better
- Remove insufficiently relevant metrics
 - Although “red cards earned” may describe some aspect of CB play, it does not say nearly as much as “passing accuracy.” Including red cards in the model would reduce the relative importance of passing accuracy.





Model Development: Creating Success Rate Metrics

- For example, a CB may total a large number of tackles due to playing-style or game flow
 - i.e. deep defending, lack of possession
- The % of tackles a CB wins measures defensive skill
 - So, the model should include success rate metrics for actions such as tackling, passing, aerial duels, etc. to measure ability
- However, a CB who enters and wins just one tackle in a match is likely playing poorly
 - Number of successful actions still factor into CB play – do not discard such metrics





Model Development: Percentiles for An Overall Performance Metric

- Scale and compare metrics using percentiles
 - Leandro González Pírez example:
 - Is 2.4 tackles won per 90 min good?
 - 95th percentile of all CB's: very active tackler
 - How about 90.4% short pass accuracy?
 - 49th percentile of all CB's: average short passer
- Percentiles show the uniqueness of a CB relative to his peers
 - According to supply and demand, scarce resources have a higher price
 - A high-performing CB is even more valuable if there are few others like him
 - Percentiles adjust value to account for this phenomenon





Model Development: Deciding to Use the Average

- Defining successful individual defending is more difficult to do than with attacking
 - On-ball actions of attacking players can largely explain goals
 - A greater % of successful defending is done off-ball and through communication
- Attacking and defensive skills both factor into CB quality
 - I averaged percentiles for 6 defensive metrics, 6 offensive metrics, and 2 aerial metrics to holistically evaluate CB performance
 - Analysts and technical staff members can adjust the weights of the metric inputs based on their interpretation of CB play





Model Development: Overall Performance Score Input Metrics

- Long Pass Accuracy
- Long Passes Completed
- Short Pass Accuracy
- Short Passes Completed

- Aerial Duel Success Rate
- Aerial Duels Won
- Tackle Success Rate
- Tackles Won
- Interceptions

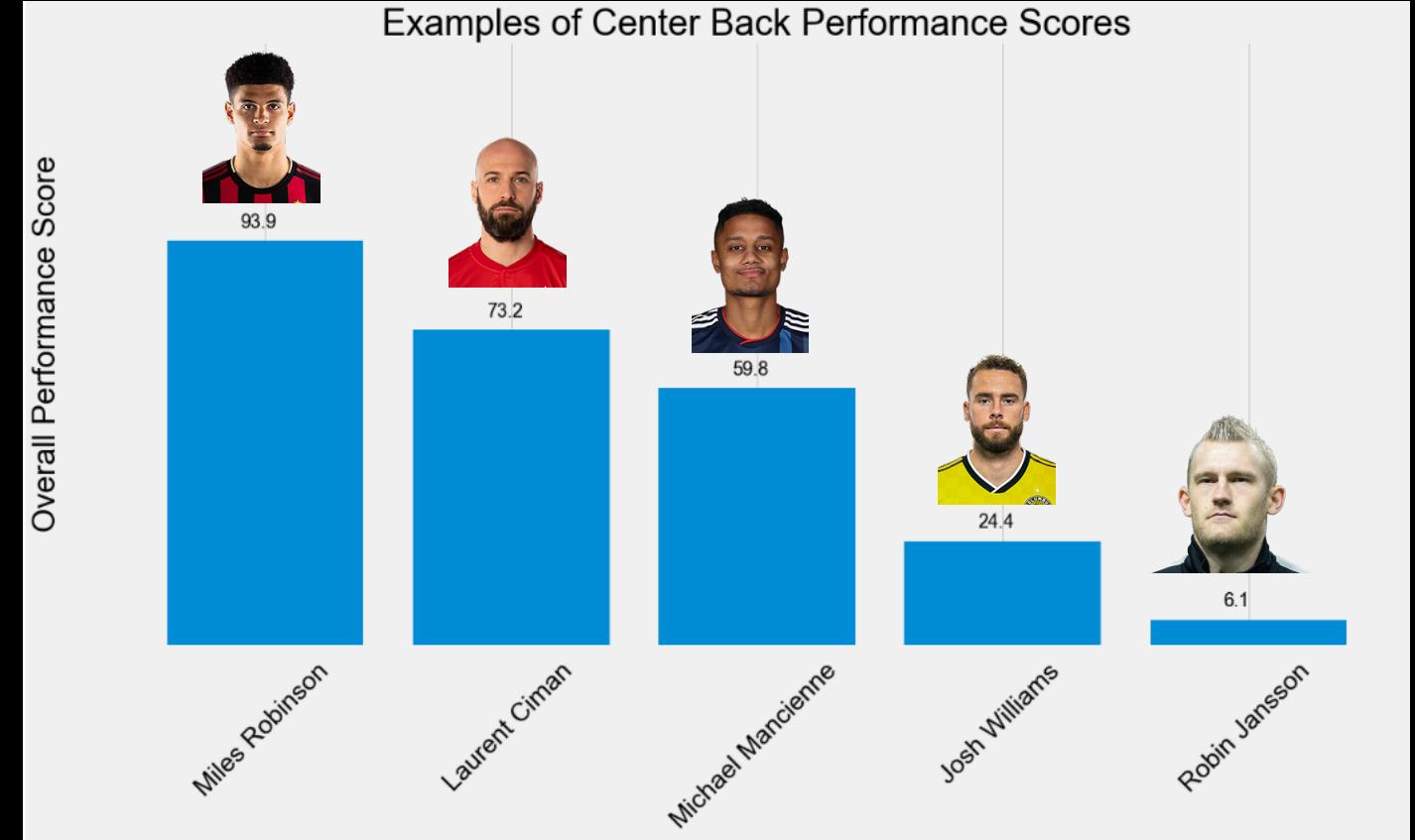
- Turnovers*
- Times Dispossessed*
- Fouls Committed*
- Shots Blocked
- Passes Blocked

* This metric's percentile goes up as its value goes down





Model Development: Overall Performance Metric Visualization





Model Development: Generating a Team-Specific Ranking

- Filter out players who lack in key metrics for Inter Miami's playing style
 - Tackling and Aerial Duel Success Rates
 - Limiting Turnovers and Dispossessions
 - Short Passing Accuracy
- Identify threshold for key metric minimum
 - Objective: find players who have overall quality and also fit our system
 - Minimum for key metrics: 34th percentile (bottom of the middle 3rd)
 - Top 5 results are at least around average for Inter Miami's KPI's
 - Top 5 results are all in the 85th percentile or higher overall – elite CB's





Model Validation: Intuitive Tweaking

- I tested to see if older players avoided turnovers more successfully than younger ones.
- Age-weighting the turnover metric could predict how a talent like Miles Robinson will develop his possession game with more experience
- Testing, however, revealed no age-turnover correlation
- Proper analysis will retain and retest such intuitions based on in-depth understanding of soccer





Model Summary:

- Convert metrics relevant to CB quality to percentiles
- Find the average of these percentiles
- Adjust average into an Overall Performance Score
- Filter out players that fail to meet Inter Miami CF's playing philosophy criteria for CB's
- Filter out players transferred out of MLS or retired
- Select the top 5 remaining CB's



5 Selections

FUTBOLMIAMI.COM

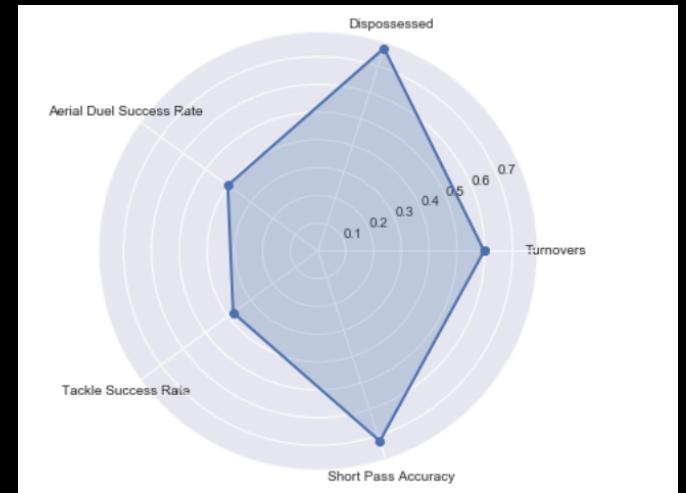
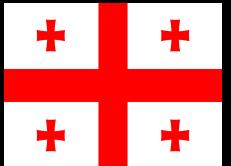
MIAMI





Guram Kashia - San Jose Earthquakes

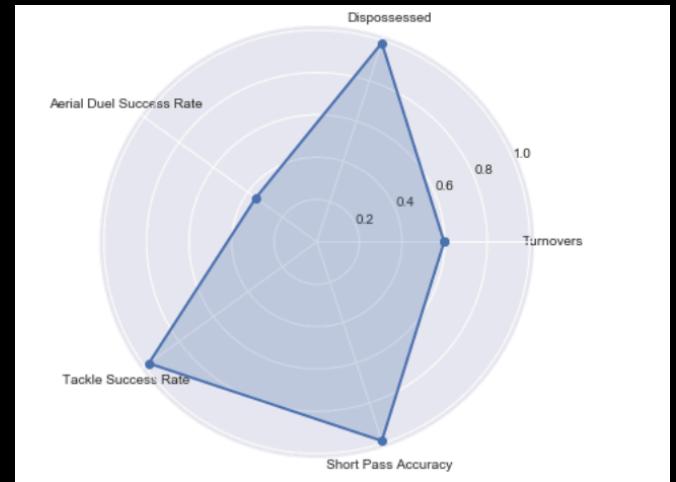
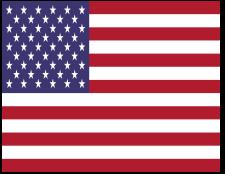
- Age: 32
- Height: 6' 1"
- Minutes Played: 1350
- Status: Senior Roster
- Overall Performance Score: 100
- Tackle Success Rate: 70.6% - 37th percentile
- Aerial Success Rate: 56.6% - 40th percentile
- Short Pass Accuracy: 92.2% - 72nd percentile
- Dispossessed: 0.1 per 90 - 76th percentile
- Turnovers: 0.4 per 90 - 59th percentile





Graham Smith - Sporting Kansas City

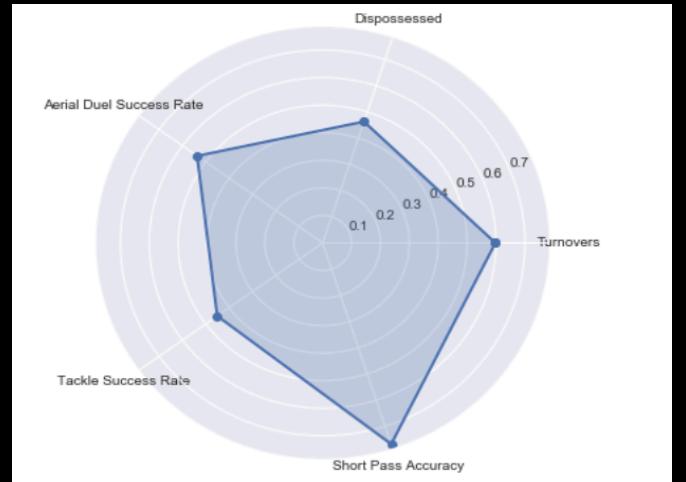
- Age: 23
- Height: 6' 3"
- Minutes Played: 691
- Status: 8 starts since May 5 call-up from USL
- Overall Performance Score: 98.78
- Tackle Success Rate: 92.3% - 97th percentile
- Aerial Success Rate: 55.3% - 35th percentile
- Short Pass Accuracy: 94.9% - 98th percentile
- Dispossessed: 0.0 per 90 - 98th percentile
- Turnovers: 0.4 per 90 - 59th percentile





Matt Hedges – FC Dallas

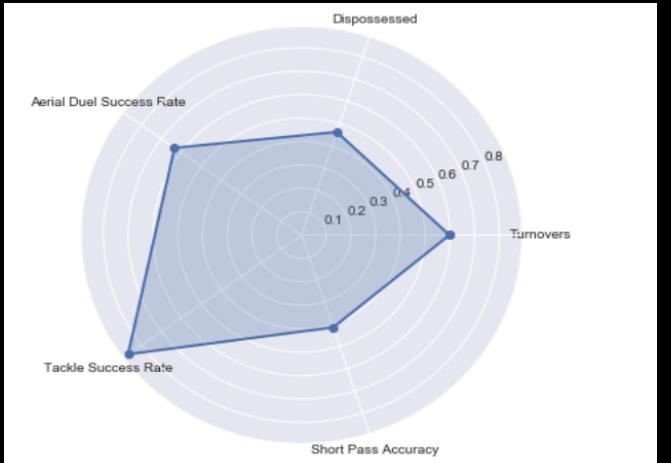
- Age: 29
- Height: 6' 4"
- Minutes Played: 2435
- Status: Senior Roster
- Overall Performance Score: 97.56
- Tackle Success Rate: 72.2% - 45th percentile
- Aerial Success Rate: 62.8% - 53rd percentile
- Short Pass Accuracy: 92.6% - 76th percentile
- Dispossessed: 0.2 per 90 - 46th percentile
- Turnovers: 0.4 per 90 - 59th percentile





Maxime Chanot - NYCFC

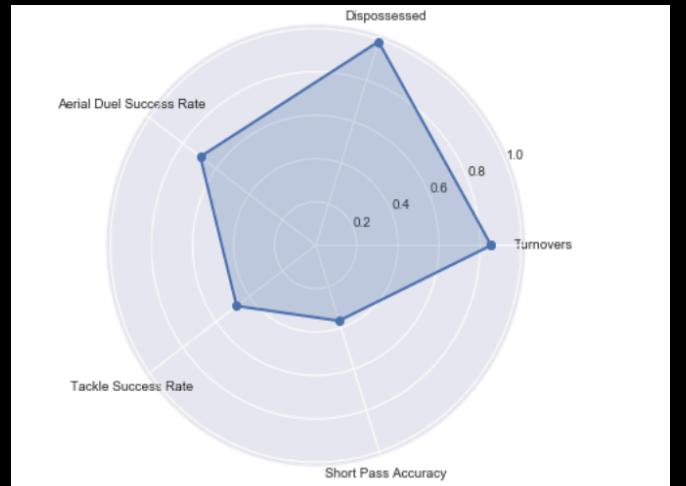
- Age: 29
- Height: 6'1"
- Minutes Played: 2340
- Status: Senior Roster
- Overall Performance Score: 96.34
- Tackle Success Rate: 84.6% - 86th percentile
- Aerial Success Rate: 63.6% - 63rd percentile
- Short Pass Accuracy: 89.9% - 41st percentile
- Dispossessed: 0.2 per 90 - 46th percentile
- Turnovers: 0.4 per 90 - 59th percentile





Daniel Steres- LA Galaxy

- Age: 28
- Height: 6'0"
- Minutes Played: 1985
- Status: Senior Roster
- Overall Performance Score: 90.24
- Tackle Success Rate: 72.7% - 47th percentile
- Aerial Success Rate: 64.4% - 69th percentile
- Short Pass Accuracy: 89.5% - 36th percentile
- Dispossessed: 0.0 per 90 - 98th percentile
- Turnovers: 0.3 per 90 - 85th percentile



In Closing





Concluding Note:

It was an honor to create my own database from the web and turn it into CB recommendations for my new hometown team. I believe the rationale for my model is sound and shows that I can take an abstract soccer question and turn it into actionable information. I am confident that I can combine my data science proficiency, soccer knowledge, and communications skills to have a positive impact as a member of the Inter Miami CF Technical Staff. I appreciate your time and consideration, and I look forward to talking with you about this analysis project and career opportunities.

Best Regards,
David Goldstein



A photograph of a city skyline at night, featuring several illuminated skyscrapers. In the foreground, there is a large mural on a building wall. The mural includes a circular logo for "CIBAO INTERNACIONAL DE FUTBOL MIAMI FC" with the year "MMXX" and a stylized bird or dragon-like creature. Below the logo, there is a figure in a dynamic pose. To the right of the mural, a street sign indicates "ONE WAY" with an arrow pointing left.

Thank You

All data provided by WhoScored.com