Analyzing Opponent Attacking Tendencies

FCB-MIT Final Presentation

Massachusetts Institute of Technology

Central Question

*Can the **clustering** of **opponent attacks** based on **observable features** provide accurate **classifications** of **offensive tendencies**?*

Data Set Summary

- Working with on ball event data

- Type, Success, Start/End Time, Players, Start/End Position, and Distance are known for each event

- Parsed sequences of events into attacks

- Defined an attack as a sequence of events by 1 team that ends when possession is lost, the ball goes out of play, or a shot is taken

Feature Engineering

Methodology:

- We propose an EM clustering model with 21 features in 3 aspects:
- Basic Macro-Event Features (understand basic football insights)
- Flow Motifs (understand passing structure and ball movement)
- Re-designed Spatial Region Features (understand pitch position and attack tendency)

Basic Macro-Event Features

- Number of Passes
- Time Duration
- Total Vertical Distance Traveled
- Total Vertical Distance
- Total Horizontal Distance
- Vertical/Horizontal Ratio
- Average Attacking Speed
- Number of Successful Long/Short Passes

Extended Features

Flow Motif Features

- Count of occurrences for each flow motif during the attack
  - ABAB
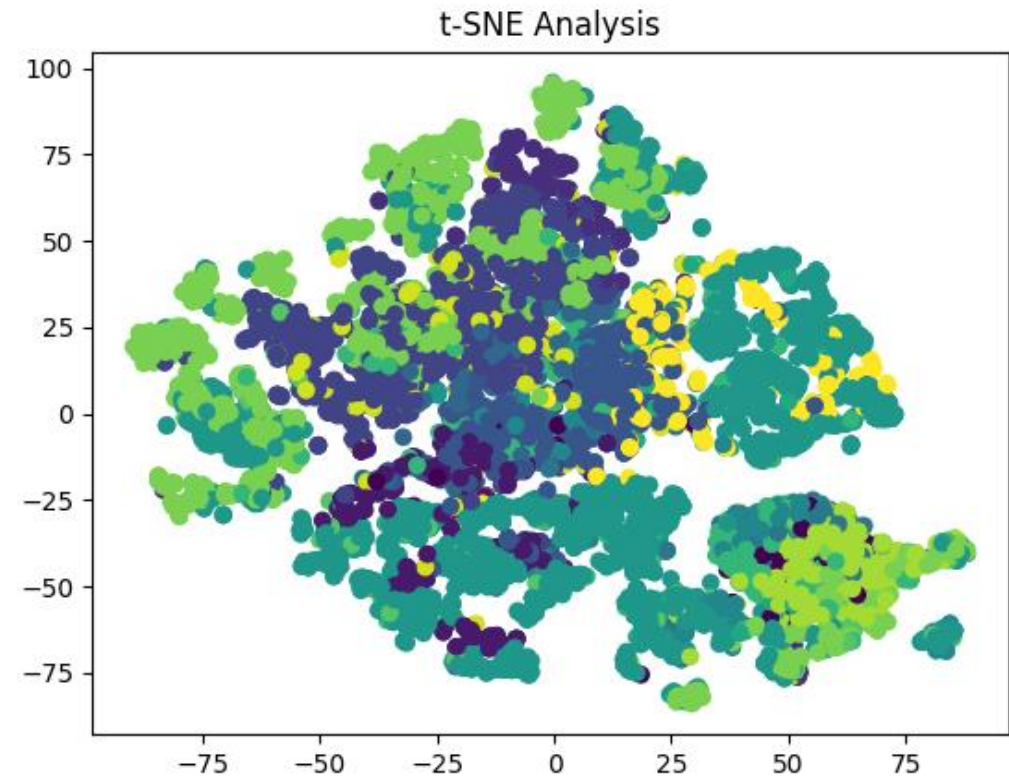 - ABAC
  - ABCA
  - ABCB
  - ABCD

Spatial Region Features

- Percentage of attacking time inside a specific spatial region
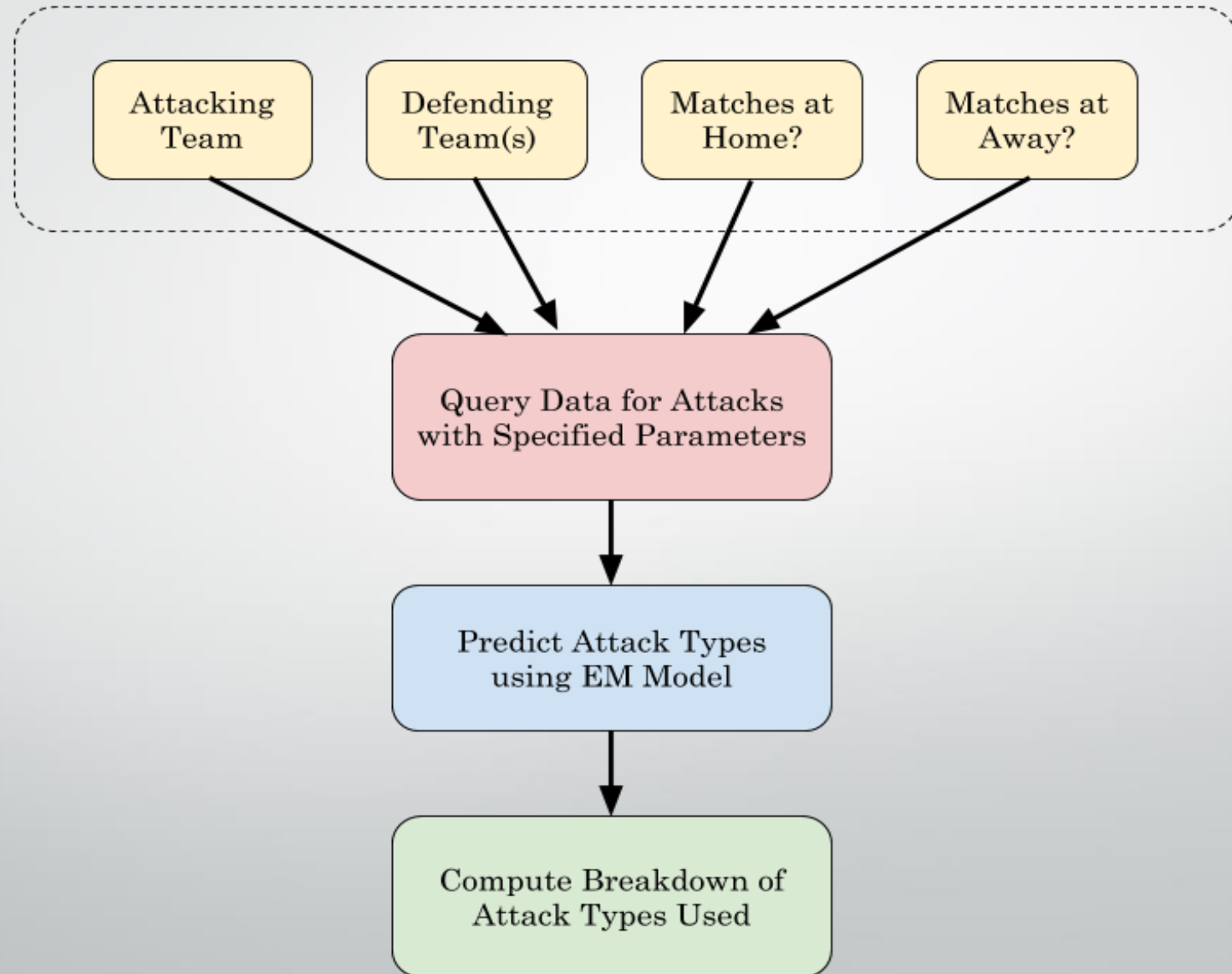- Region of field where ball was lost

Final Clustering Model

- Algorithm: Expectation-Maximization (EM) Algorithm

- Optimal Number of Clusters: 16

- Trained on 60,000+ attacks with the 21 engineered features from the 2016-17 La Liga Season

t-SNE Analysis

Filtering Analysis
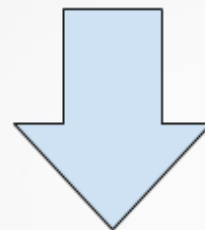
Filtering Parameters

| Attacking Team | Defending Team(s) | Matches at Home? | Matches at Away? |

Query Data for Attacks with Specified Parameters

Predict Attack Types using EM Model

Compute Breakdown of Attack Types Used
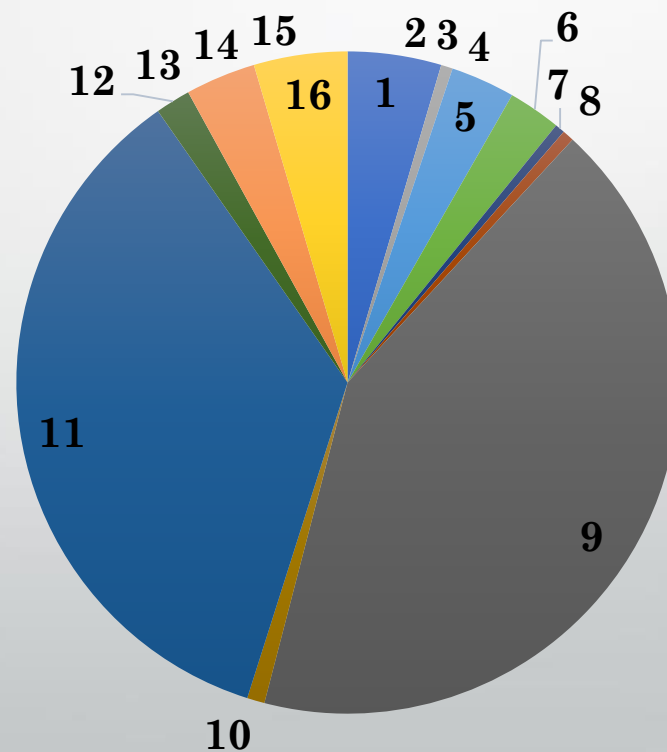
Filtering Parameters

| Real Madrid | FC Barcelona, Atletico de Madrid, Valencia CF, Sevilla | True | False |

Filtering Example

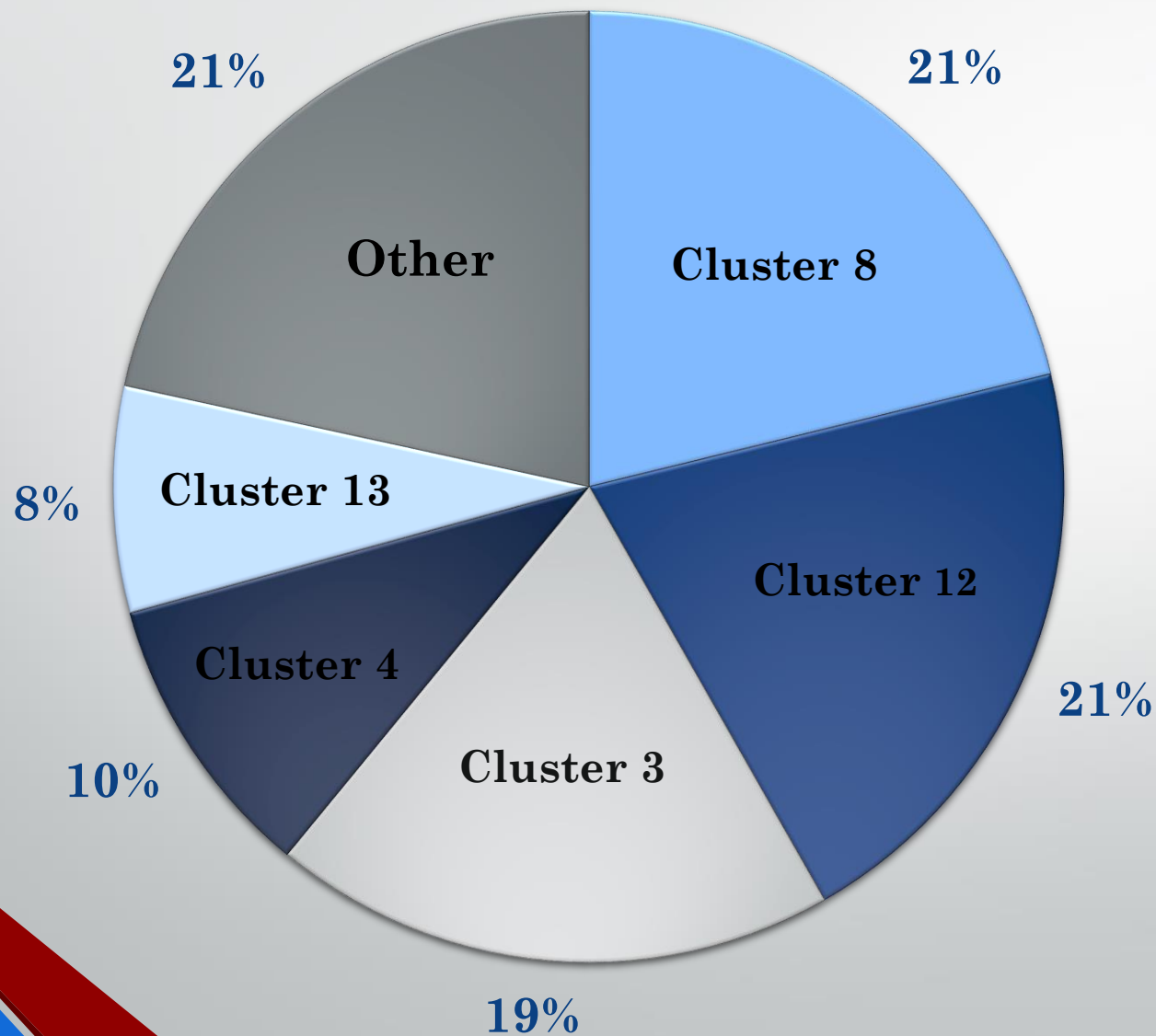Real Madrid Attacking Breakdown Based On Given Parameters

CLUSTER	CHARACTERISTICS	LOCATION	RESULT
1	Many Passes	Defensive and Middle Third	Long Ball
2	Short, sloppy	Defensive Third	Long Clearance
3	Quick	Right Flank	Cross
4	Quick	Left Flank	Cross
5	Short passes	Defensive Third	Dispossessed in Midfield
6	Patient, Many Passes	Midfield	Never Reaches Final Third
7	Long, Many Passes	Midfield	Enters Final Third
8	Patient, Many Passes	Center and Left of Defensive and Middle Thirds	Never Reaches Final Third
9	Long Distribution From Goalkeeper/Defender	Final Third	Fail to Reach Second Ball
10	Long Distribution From Goalkeeper/Defender	Right Side	Shot or Cross
11	Clearance to Sideline	Defensive Third	Ball Goes Out Before Half Field
12	Dribble through lines	Left Flank	Cross
13	Set Pieces, Throws, Corners	N/A	Cross
14	Patient, Pass Through Lines	Defensive Third	Cross
15	Patient, Back Passes	Midfield	Breaking Lines Into Final Third
16	Patient, Break Through Lines	Defensive and Middle Third	Rarely Reach Final Third

Feature Engineering Results

Real Madrid Attacking Breakdown

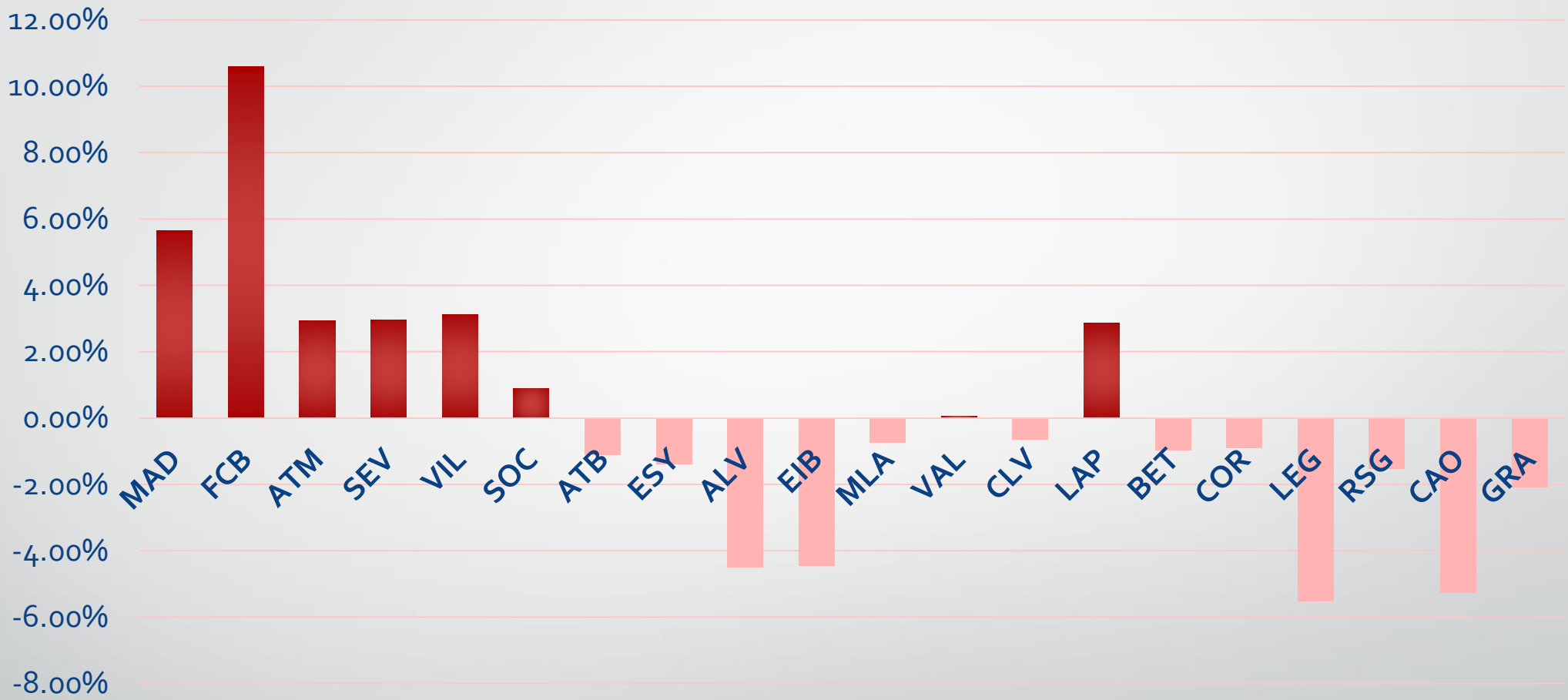| CLUSTER	CHARACTERISTICS	LOCATION	RESULT
3	Quick	Right Flank	Cross
4	Quick	Left Flank	Cross
8	Patient, Many Passes	Center and Left of Defensive and Middle Thirds	Never Reaches Final Third
12	Dribble through lines	Left Flank	Cross
13	Set Pieces, Throws, Corners	N/A	Cross

Pie chart labels: Cluster 8 (21%), Cluster 12 (21%), Cluster 3 (19%), Cluster 4 (10%), Cluster 13 (8%), Other (21%)

% Difference From Cluster 3 League Avg

CLUSTER	CHARACTERISTICS	LOCATION	RESULT
3	Quick	Right Flank	Cross

% Difference From Cluster 8 League Avg

Percentage Difference From League Average
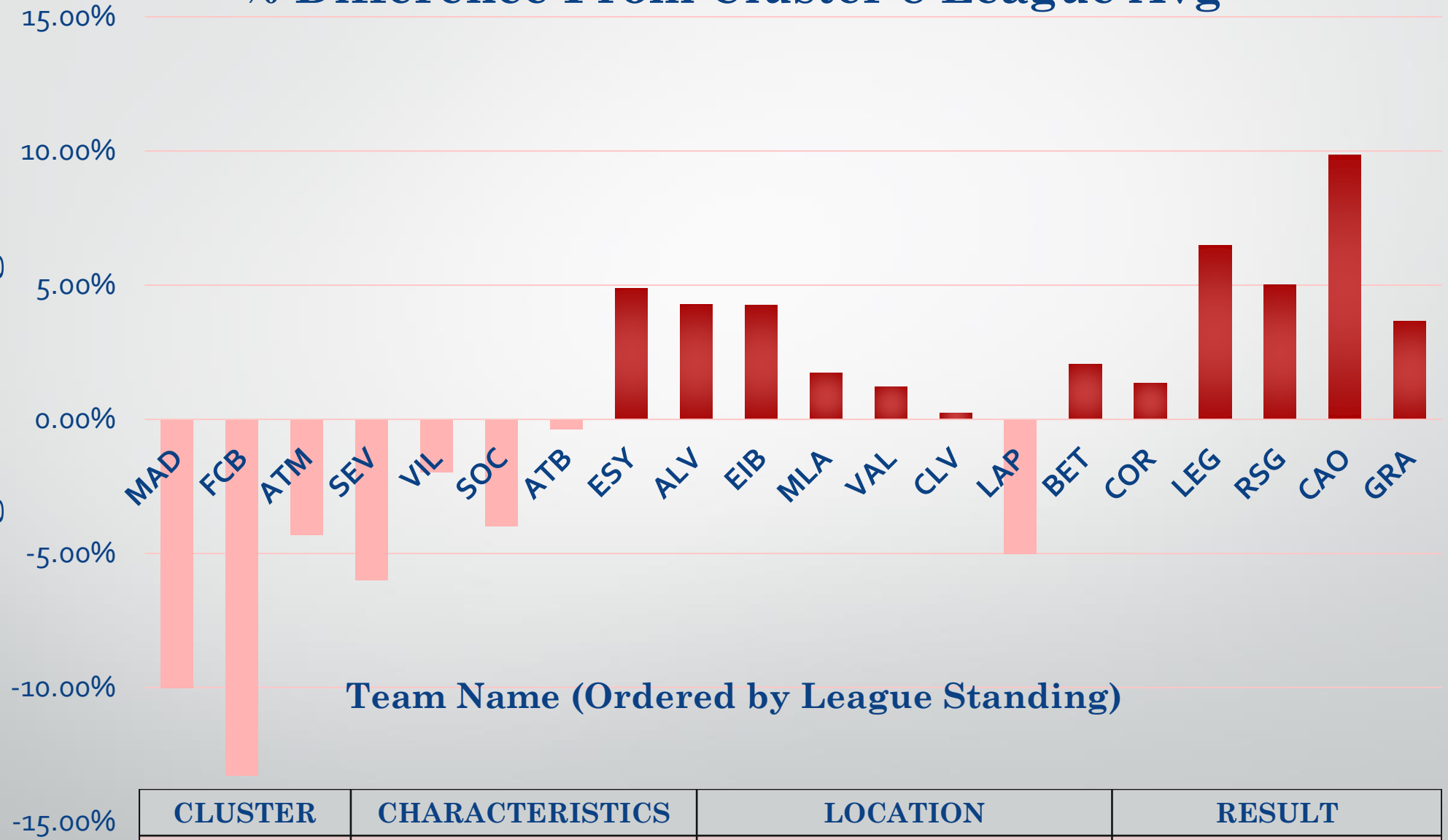
Team Name (Ordered by League Standing)

MAD FCB ATM SEV VIL SOC ATB ESY ALV EIB MLA VAL CLV LAP BET COR LEG RSG CAO GRA

CLUSTER	CHARACTERISTICS	LOCATION	RESULT
8	Patient, Many Passes	Center and Left of Defensive and Middle Thirds	Never Reaches Final Third

CLUSTER	CHARACTERISTICS	LOCATION	RESULT
1	Many Passes	Defensive and Middle Third	Long Ball
2	Short, sloppy	Defensive Third	Long Clearance
3	Quick	Right Flank	Cross
4	Quick	Left Flank	Cross
5	Short passes	Defensive Third	Dispossessed in Midfield
6	Patient, Many Passes	Midfield	Never Reaches Final Third
7	Long, Many Passes	Midfield	Enters Final Third
8	Patient, Many Passes	Center and Left of Defensive and Middle Thirds	Never Reaches Final Third
9	Long Distribution From Goalkeeper/Defender	Final Third	Fail to Reach Second Ball
10	Long Distribution From Goalkeeper/Defender	Right Side	Shot or Cross
11	Clearance to Sideline	Defensive Third	Ball Goes Out Before Half Field
12	Dribble through lines	Left Flank	Cross
13	Set Pieces, Throws, Corners	N/A	Cross
14	Patient, Pass Through Lines	Defensive Third	Cross
15	Patient, Back Passes	Midfield	Breaking Lines Into Final Third
16	Patient, Break Through Lines	Defensive and Middle Third	Rarely Reach Final Third

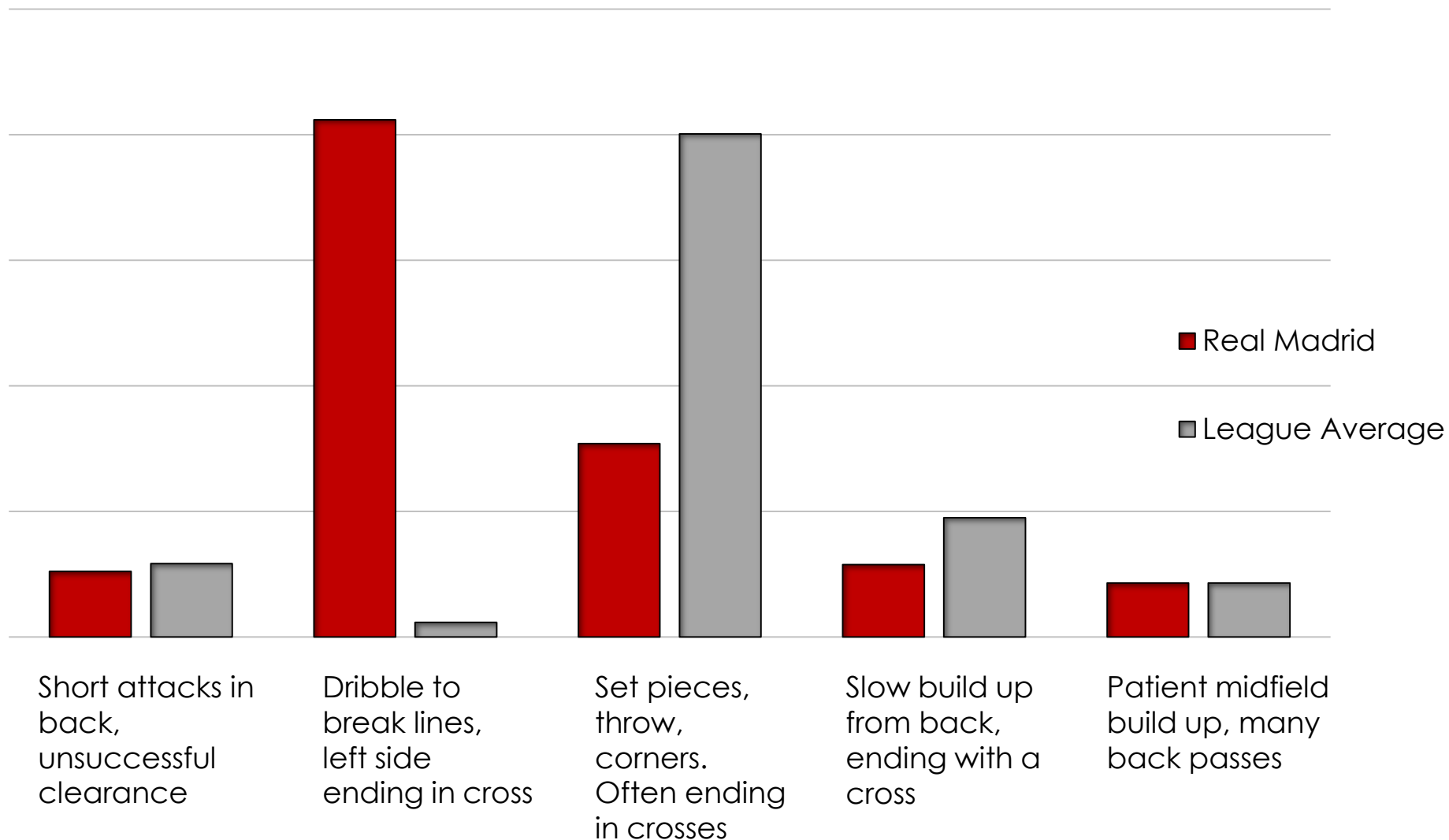Conclusions From League Trends

Scouting Report

Relative Divergence From League Average Cluster Percentage for Real Madrid
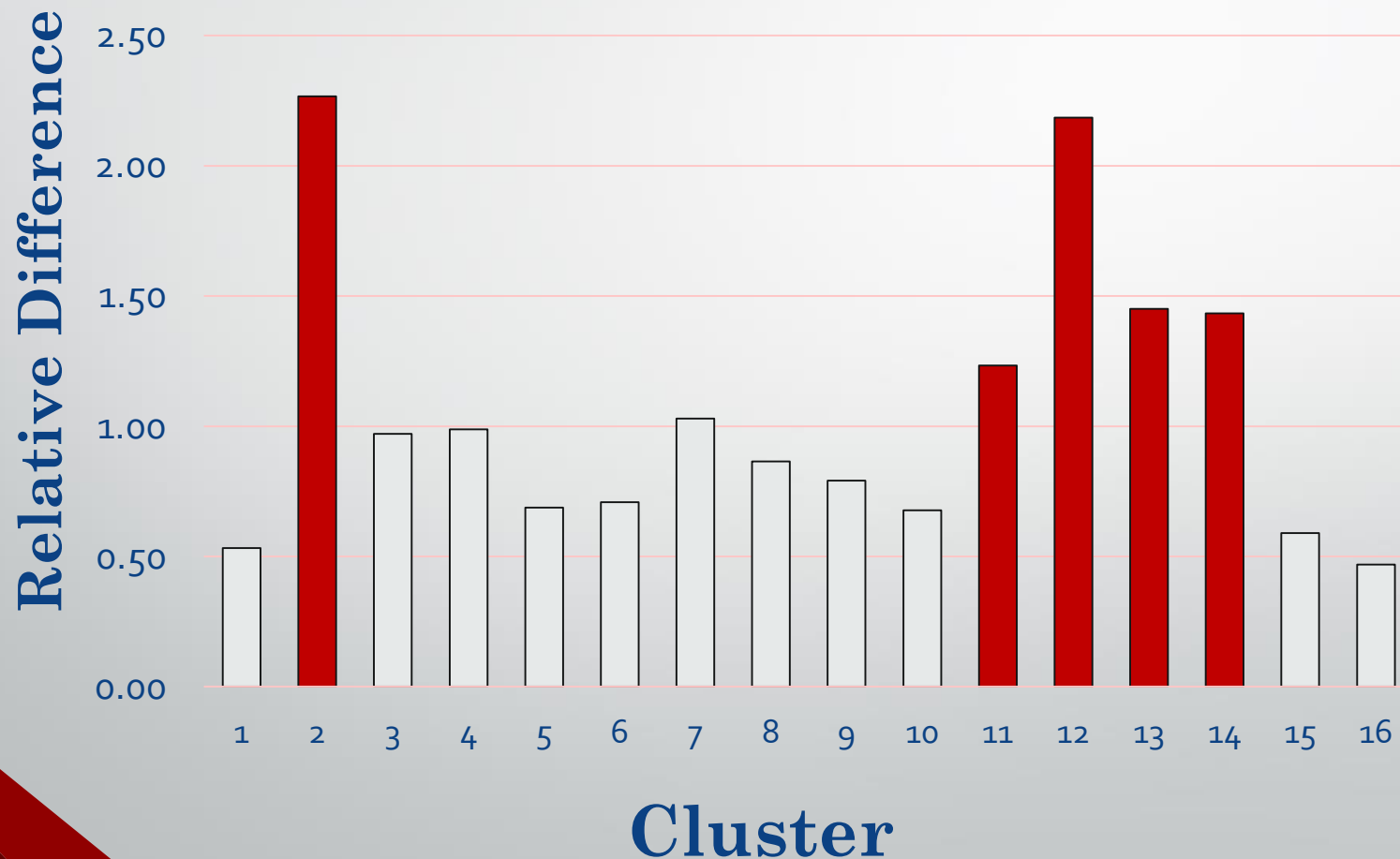
- Extract Clusters which stray most from league average
- Account for less common attacks by utilizing standard deviation of clusters

REAL MADRID SCOUTING REPORT

Questions?