Capgemini

GAME CHANGERS: PYTHON, VIDEO ANALYSIS, AND PROPERTY GRAPHS IN SPORTS ANALYTICS

Dr. Abi Giles-Haigh,
Oracle Analytics and Innovation Director

INTRODUCTIONS

- PhD in ML – Computation Modelling of the human heart
- Oracle DBA for 5 years
- Worked in public health care, private banking
- Oracle ACE Director

Oracle ACE

400+ technical experts helping peers globally

The Oracle ACE Program recognizes and rewards community members for their technical and community contributions to the Oracle community

3 membership tiers

Oracle ACE Director | Oracle ACE Pro | Oracle ACE Associate

For more details on Oracle ACE Program:
ace.oracle.com

Oracle ACE

Nominate

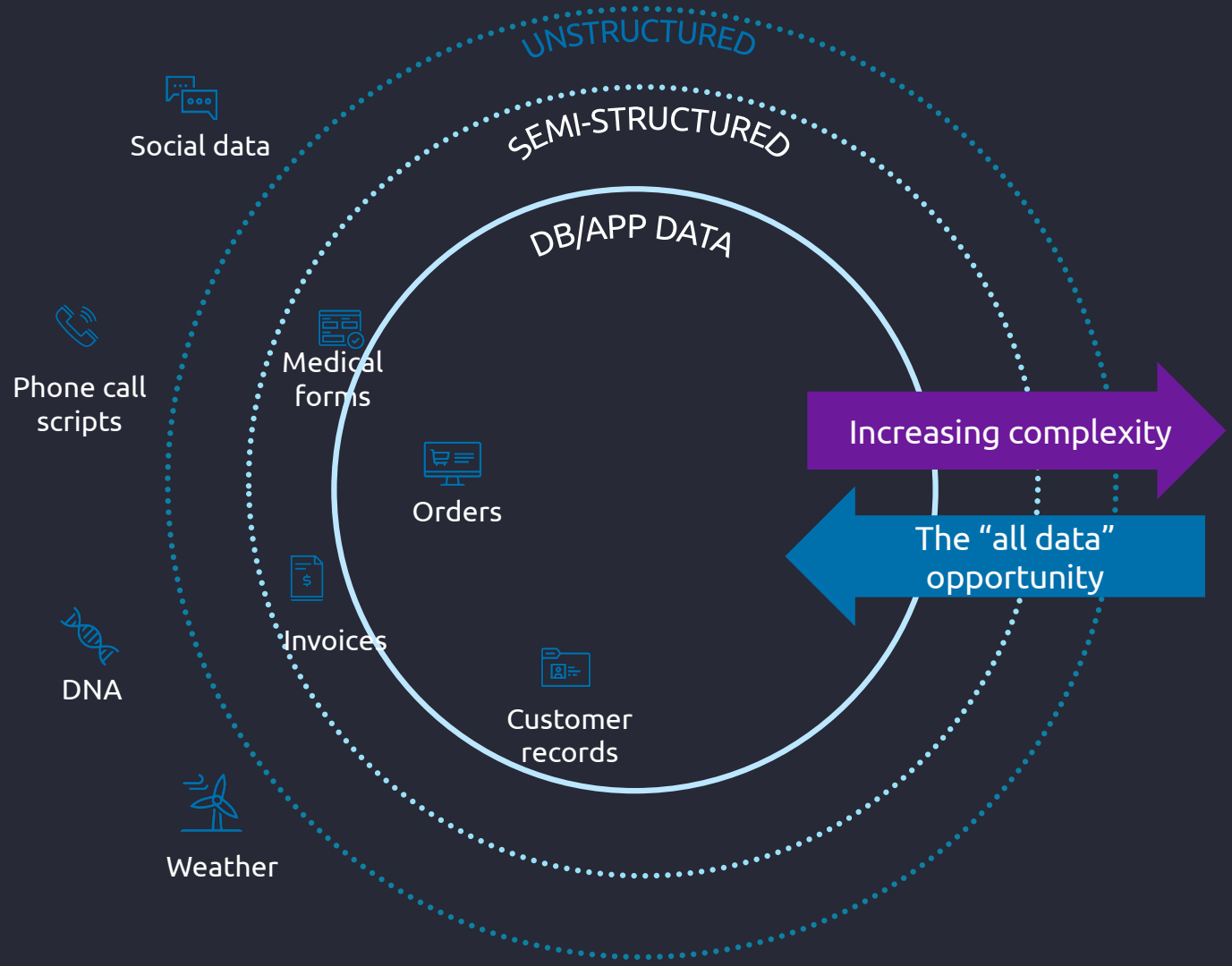
yourself or someone you know:
ace.oracle.com/nominate

Connect: aceprogram_ww@oracle.com Facebook.com/OracleACEs @oracleace

AGENDA

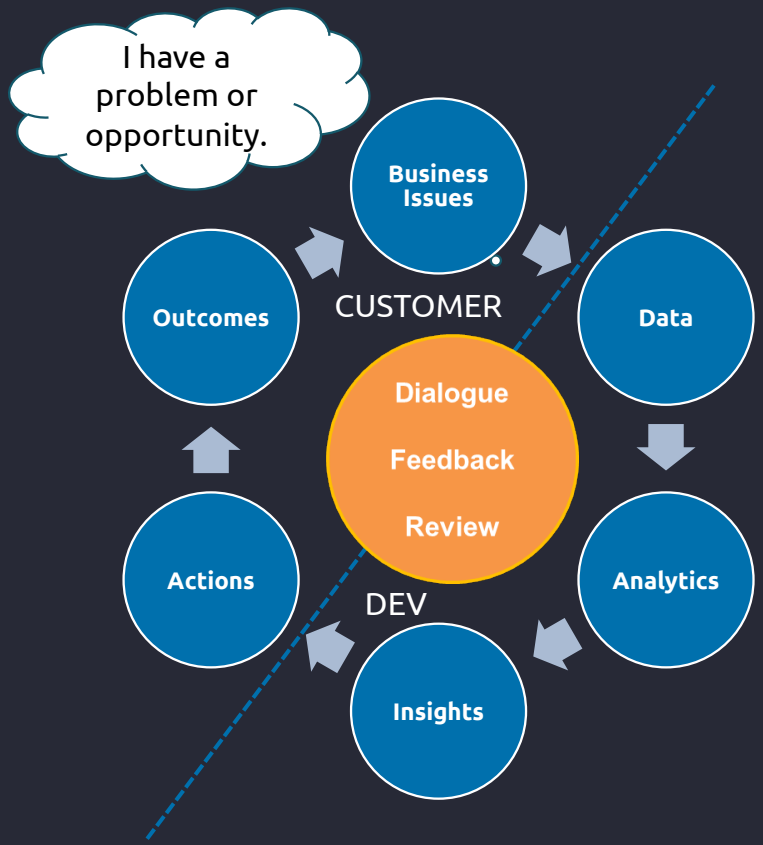
- History of this project
- A quick review of our Property Graph
- How we are using it now
- What the future holds

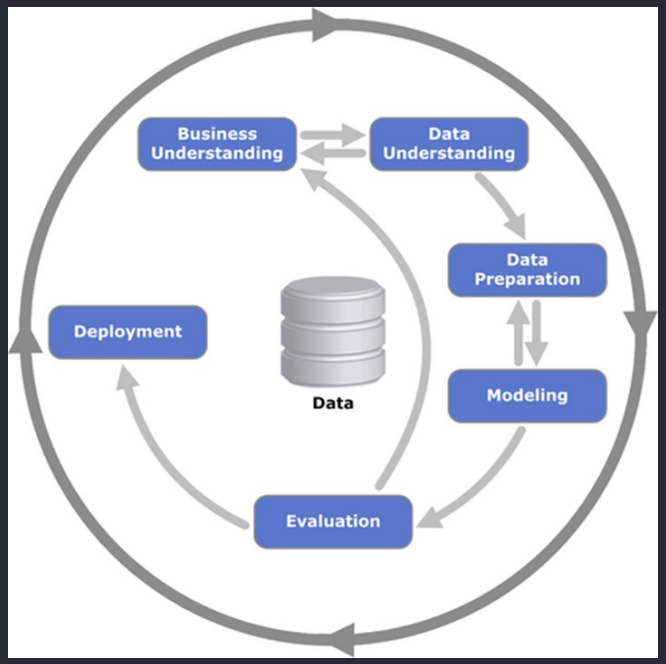
IT ALL STARTS WITH DATA....



IT ALL STARTS WITH DATA....

BUSINESS NEED AND OPERATIONAL PROCESS

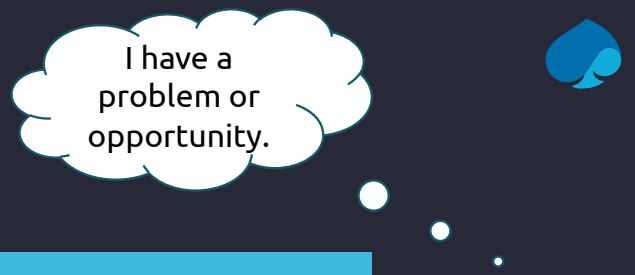




- The operating model must be flexible to respond to the needs of the customers

- Cross-industry standard process for data mining (CRISP-DM)
- Data mining is a process of extracting and discovering patterns in large data sets

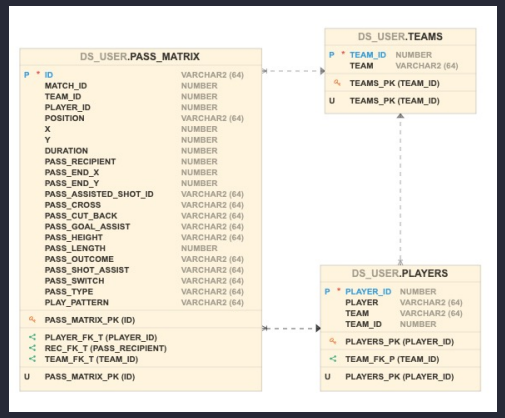
DATA SCIENCE MEANS ...
...BEING ABLE TO ANSWER QUESTIONS SUCH AS

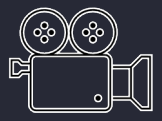


How can I detect fraudulent claims/actions early?	How to detect indicators and trends for managing risk?	How to optimize delivery/production cost?
How to set the right price to maximize profitability?	How can I improve efficiency of product recommendation engines?	How to decrease the number of false alerts?
How can I predict future demand?	How to improve productivity by keeping the right talent?	How to find specific patterns and anomalies?

CATCH UP ON THE STORY SO FAR

- Started during Covid lockdown
- Could I bring analytics to grassroots football
- The whole solution needs to run for less than $500 per year
- Using one camera, some Python and a database (free ADW)
- Along the way, could I inspire new ways of thinking…





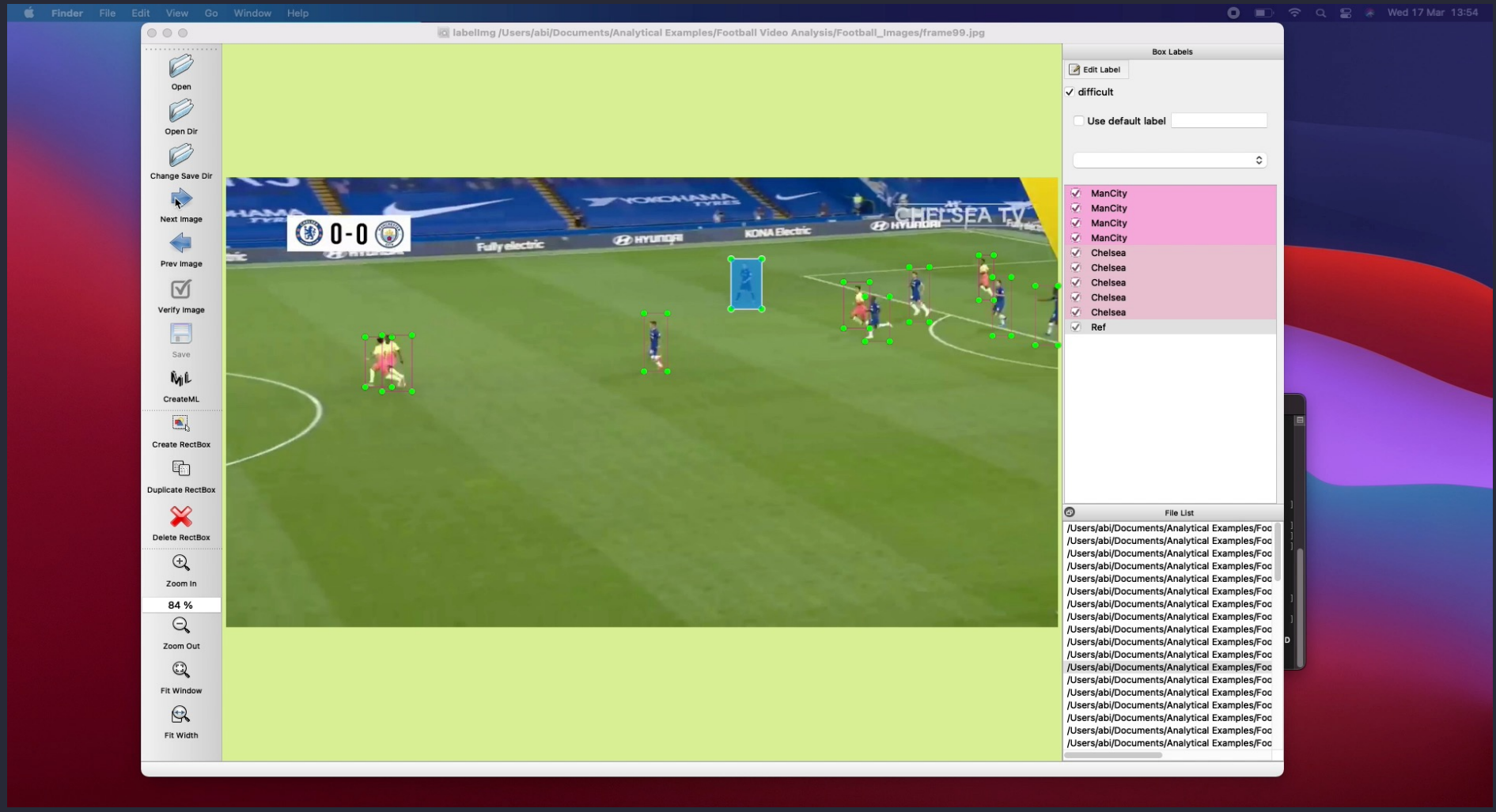
Video Analytics:
Open Source

Structured Data:
ADW

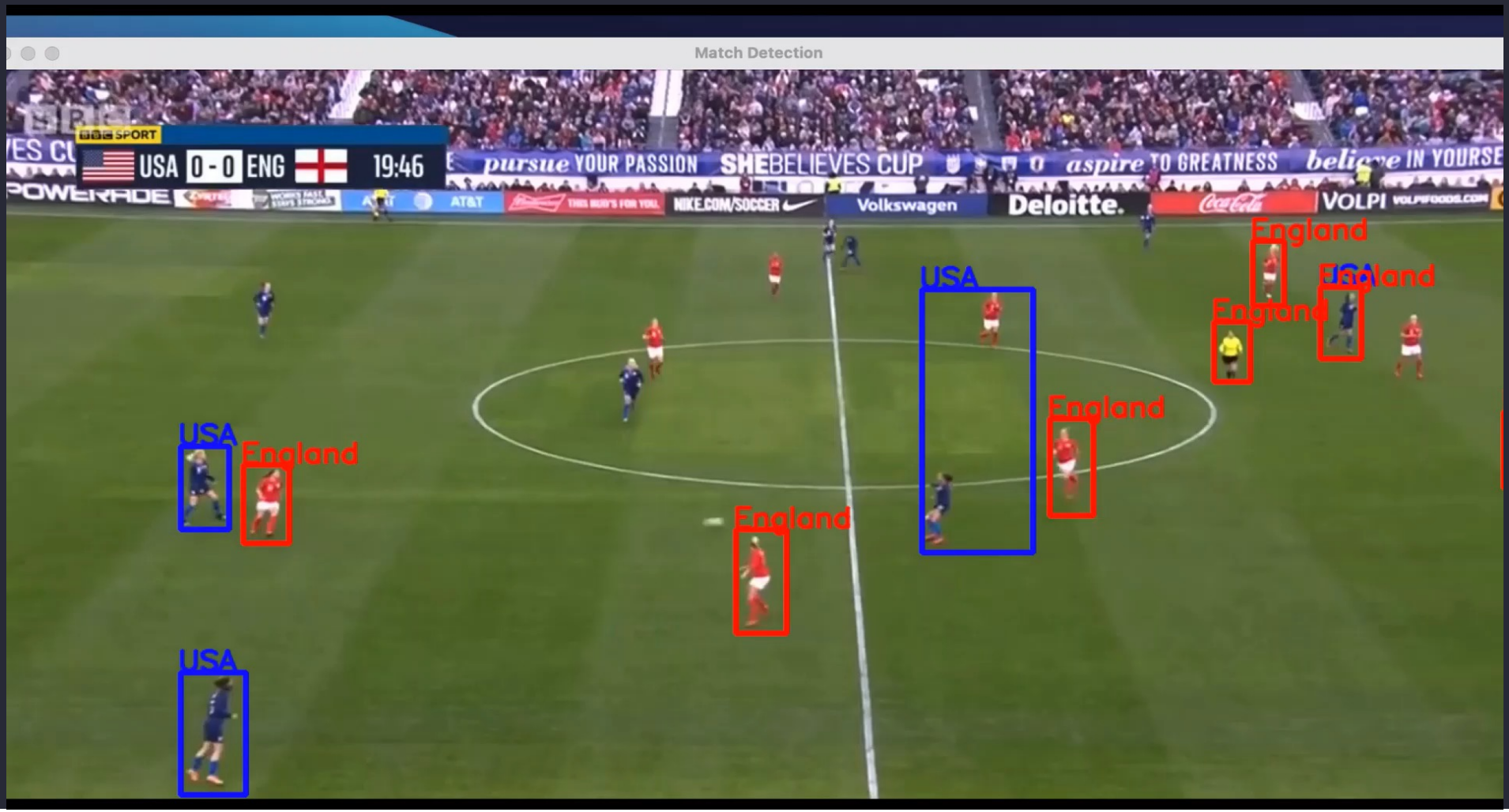
Analytics:
APEX and OA

Enrichment:
Graph

THE ORIGINAL VERSION



THE ORIGINAL IN MOTION

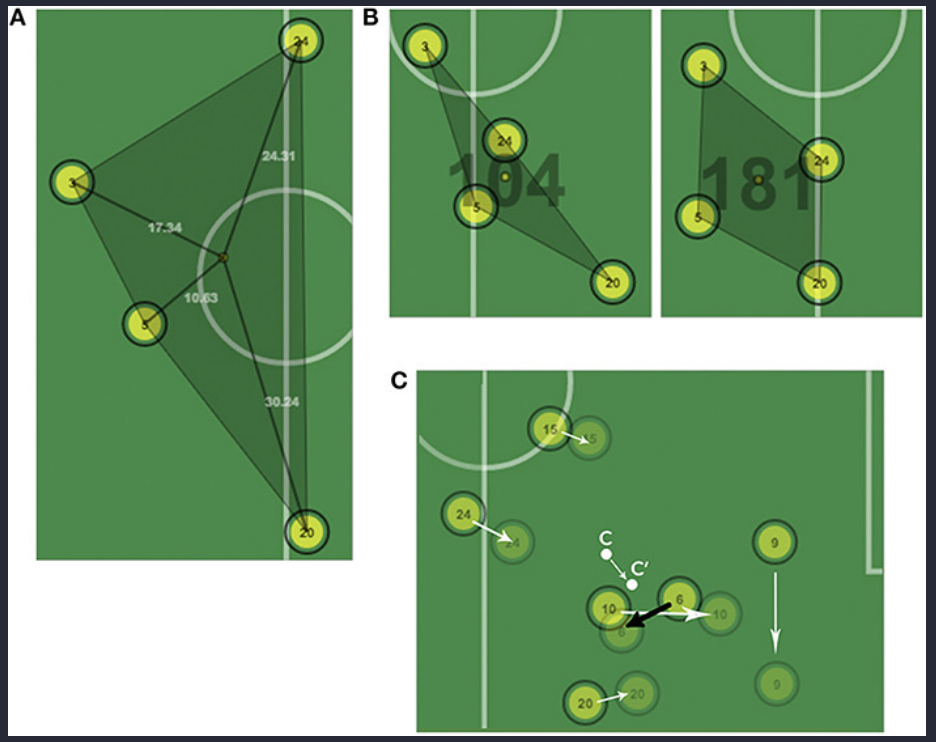


CATCH UP ON THE STORY SO FAR

FOOTBALL ANALYTICS

Several elements can be measured:

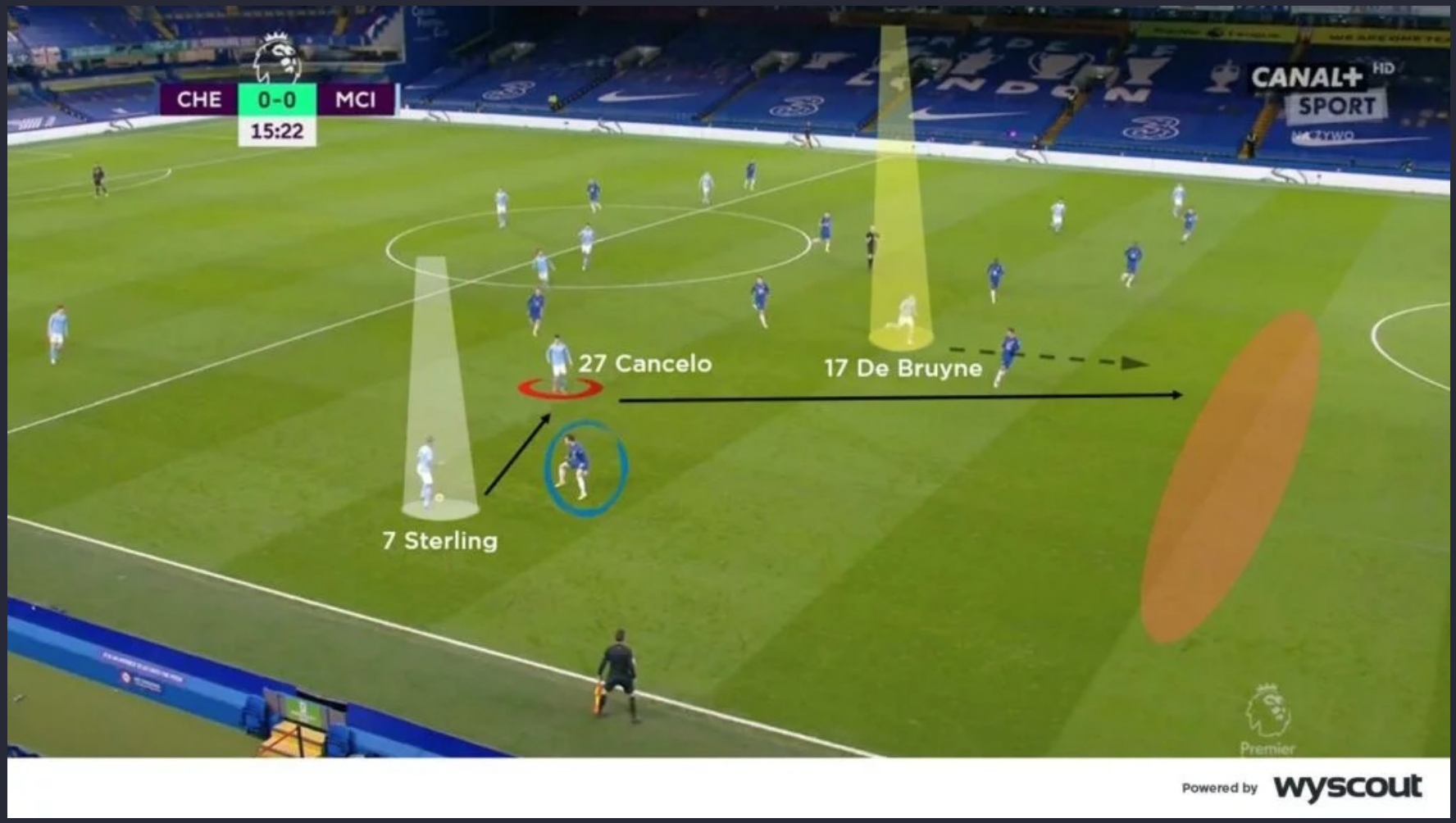
- Body position
- Position of the player on the pitch
- Position of the ball
- Number of attacks
- Number of corners
- Distance dribbled
- Number of Crosses
- The number of throw in's…..



THE ART OF PASSING

- Where on the pitch did the pass start?
- Which player started the passing pattern?
- Which position does that player represent?
- Where did the pass end?
- Did the pass result in a shot?
- Was the shot successful?

WHAT IT LOOKS LIKE



WHAT DOES THE PROPERTY GRAPH LOOK LIKE

Player A (Node)

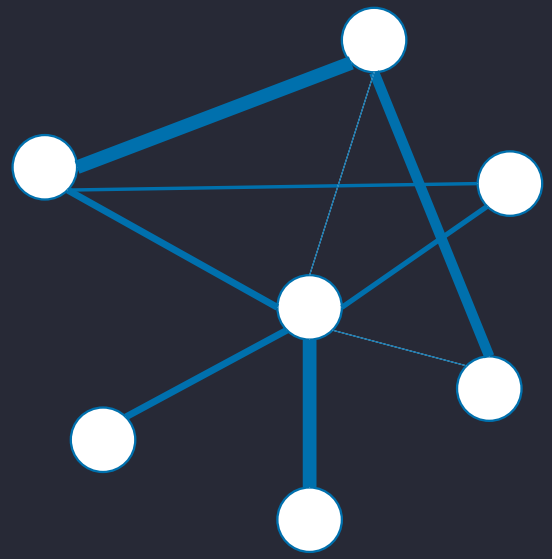
Number of Passes between players (Weight)

Passing Line (Edge)

Player B (Node)

PUT THE PASSES TOGETHER

- Nodes for each player
- Edges for each pass
- Weights for the edges:
  - number of passes between players
  - Distance of the pass
- Degrees for the number of players a player receives or passes too
- All combined, it makes a passing graph



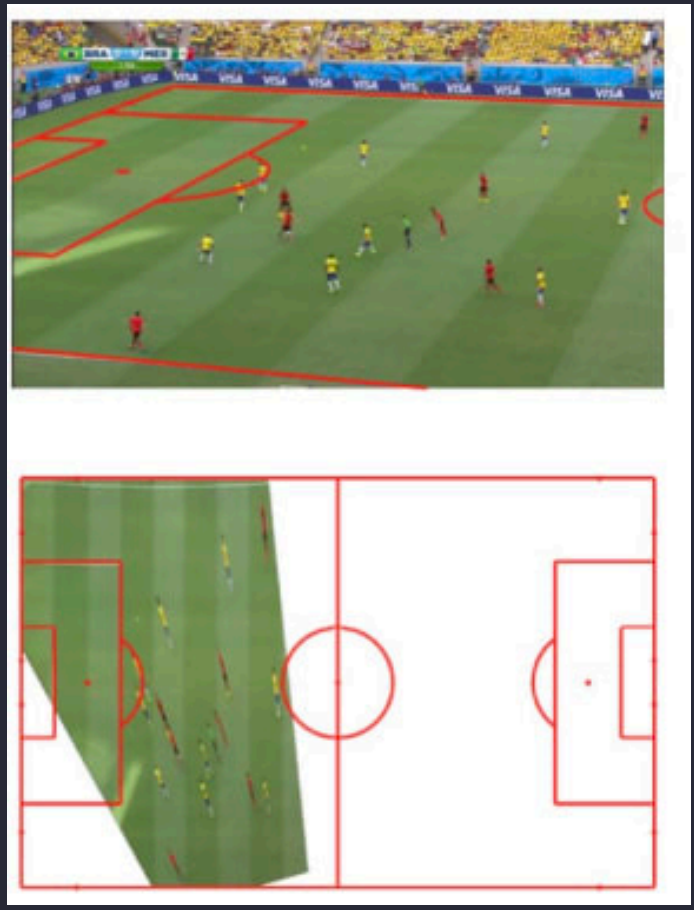
LOCATION IS CONTEXT

Teach the machine to look for edge of the pitch

Use the lines detected to map to a 2D pitch

Use some clever maths (homography matrices) to figure out the location on the pitch

Plot the player each frame of the video



OLD VS NEW: ORACLE 23C (NOW 23AI)

Previously:

- Oracle ADW + built-in Graph Studio
- Utilised Euro's data
- Identified the team that did the most passes was England
- Leah Williamson passed the most
- There were a lot of Triangles
- England had a compact shape/graph
- The algorithms were great at giving us some descriptive analytics….

Current:

- Utilising Oracle 23c Developer edition on VirtualBox
- Not Graph Studio – what can we achieve using 23c
- For demo purposes, we are using Women's World Cup 2023 data…
- … but applying real coaching tactics.

WOMEN'S WORLD CUP 2023: DATA SET

StatsBomb
Data Champions.

Column	Type	Child (/s)	Child (/s) Type	Description	Values	Value Description
id	uuid			The unique identifier for each event	e.g. "0052d1b5-e2b0-4629-bbea-c18c884ab103"	
index	integer			Sequence notation for the ordering of events within each match.	e.g. 1-# of events	
period	integer			The part of the match the timestamp relates to (1 = first half, 2 = second half).	1	1st Half
					2	2nd Half
					3	3rd Period
					4	4th Period
					5	Penalty Shootout
timestamp	timestamp			Time in the match the event takes place, recorded to the millisecond.	e.g., 00:00:06.293	
minute	integer			The minutes on the clock at	e.g., 40	

https://github.com/statsbomb/open-data

YOU CAN STILL DO THE CLASSICS

```
SELECT match_id,player_id, player, team, COUNT(1) AS Num_Passes_Recieved
   FROM graph_table ( WWC23_Passing_Graph
      MATCH (src is WWC23_Players) - [e IS WWC23_Pass_Transactions] -> (dst is WWC23_Players)
      COLUMNS (dst.player_id, dst.player, dst.team, e.match_id)
   ) GROUP BY match_id, player_id, player, team ORDER BY Num_Passes_Recieved DESC
FETCH FIRST 10 ROWS ONLY;
```

MATCH_ID	PLAYER_ID	PLAYER	TEAM	NUM_PASSES_RECIEVED
3893799	10261	Sara Doorsoun-Khajeh	Germany Women's	113
3893799	10399	Kathrin Julia Hendrich	Germany Women's	103
3893799	32355	Felicitas Rauch	Germany Women's	90
3901797	4999	Lindsey Michelle Horan	United States Women's	88
3904628	10161	María Francesca Caldentey Oliver	Spain Women's	85
3893828	4642	Millie Bright	England Women's	84
3893828	10252	Alex Greenwood	England Women's	82
3901734	25638	Moeka Minami	Japan Women's	81
3893820	10395	Maren Nævdal Mjelde	Norway Women's	81
3893833	10399	Kathrin Julia Hendrich	Germany Women's	79

YOU CAN FIND PLAYERS INVOLVED IN A CHAIN OF PASSES

```
SELECT player_id, player, team, COUNT(1) AS Num_In_Middle
   FROM graph_table (WWC23_Passing_Graph
      MATCH (src is WWC23_Players) - [e IS WWC23_Pass_Transactions] -> (via is
WWC23_Players) - [e2 IS WWC23_Pass_Transactions] -> (dst is WWC23_Players)
       where e.match_id = 3904629
       and e.match_id = e2.match_id
       and (e2.event_index - e.event_index) < 4
      COLUMNS ( via.player_id, via.player, via.team)
   ) GROUP BY player_id, player, team ORDER BY Num_In_Middle DESC FETCH
FIRST 10 ROWS ONLY;
```

YOU CAN FIND PLAYERS INVOLVED IN A CHAIN OF PASSES

```
where e.match_id = 3904629
```

Specific Match

```
and e.match_id = e2.match_id
```

Ensure the passes are the same match

```
and (e2.event_index - e.event_index) < 4
```

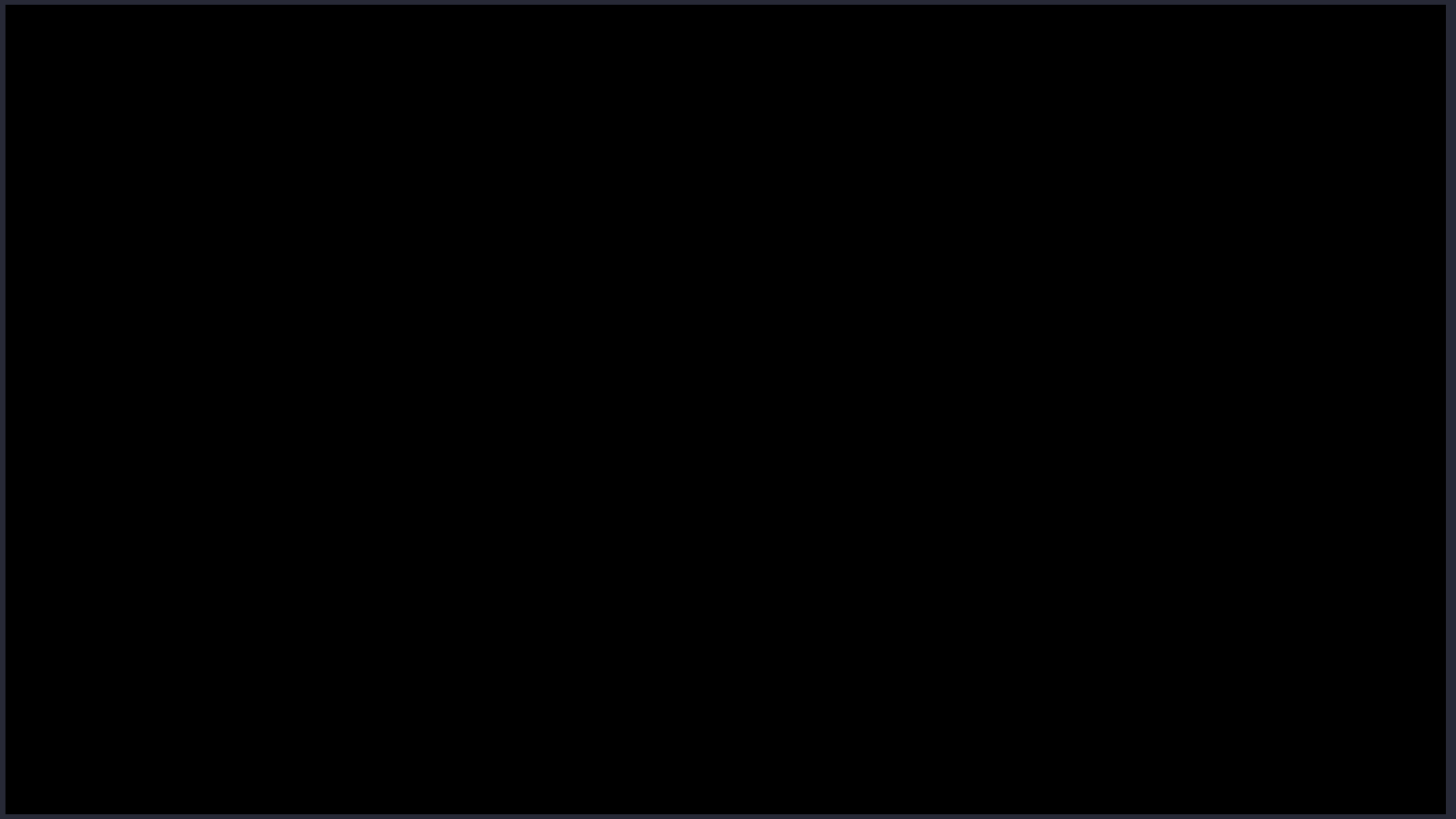
The events happen close to each other

YOU CAN FIND PLAYERS INVOLVED IN A CHAIN OF PASSES

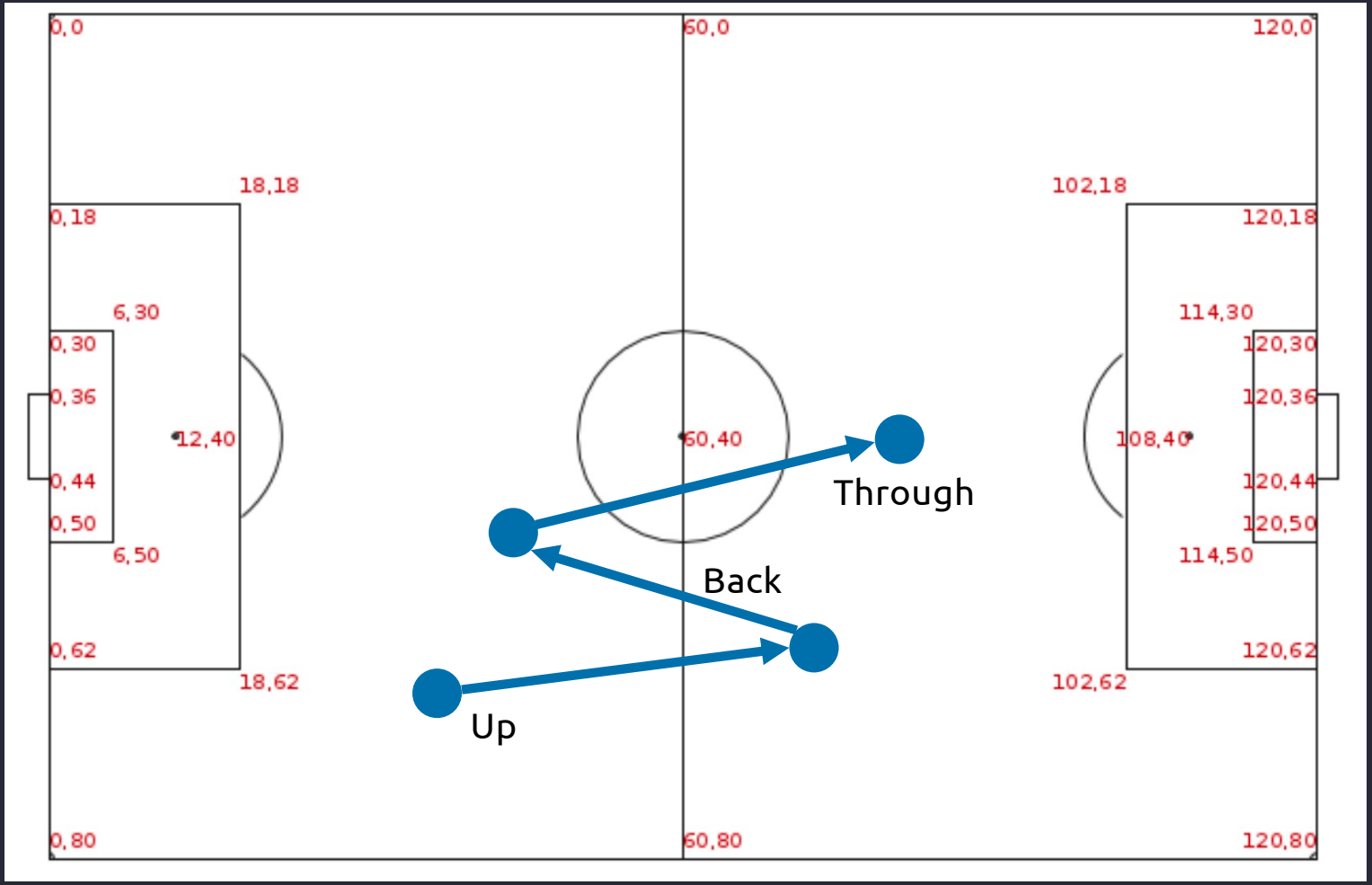
```
SELECT player_id, player, team, COUNT(1) AS Num_In_Middle
   FROM graph_table (….)
   ) GROUP BY player_id, player, team ORDER BY Num_In_Middle DESC FETCH
FIRST 10 ROWS ONLY;
```

PLAYER_ID	PLAYER	TEAM	NUM_IN_MIDDLE
10252	Alex Greenwood	England Women's	2578
4642	Millie Bright	England Women's	2401
4658	Keira Walsh	England Women's	1712
5000	Stephanie-Elise Catley	Australia Women's	1508
19422	Jessica Carter	England Women's	1182
4643	Georgia Stanway	England Women's	997
5058	Rachel Daly	England Women's	944
5095	Ellie Madison Carpenter	Australia Women's	819
47521	Alessia Russo	England Women's	687
10178	Lucy Bronze	England Women's	676

PASSING PATTERN: UP, BACK, THROUGH



DEFINING AN "UP, BACK, THROUGH"



FIND IT IN OUR GRAPH

```
SELECT distinct ev_id, ev_id2, ev_id3, period, minute, second, team,
                srcID, srcplayer, x1,   viaID, viaPlayer, x2 dstID, dstPlayer, x3 dstID2, dstPlayer2, x4, (x4-x1) as total_distance
 FROM graph_table ( WWC23_Passing_Graph
MATCH (src is WWC23_Players) - [e IS WWC23_Pass_Transactions] -> (via is WWC23_Players) - [e2 IS WWC23_Pass_Transactions] -
(dst is WWC23_Players) - [e3 IS WWC23_Pass_Transactions] -> (dst2 is WWC23_Players)
      where e.match_id = 3904629  and e.loc_x < e.end_loc_x
      and e.pass_recipient_id = e2.player_id and e2.pass_recipient_id = e3.player_id
      and e.period = e2.period and e.period = e3.period
      and e.match_id = e2.match_id and e.match_id = e3.match_id
      and e.event_index < e2.event_index and e2.event_index < e3.event_index
      and (e3.event_index - e.event_index) < 10
      COLUMNS (e.event_index as ev_id, e.period, e.minute, e.second,
            src.player_id as srcID, src.player as srcplayer, src.team as team,
            e.loc_x as x1, e.end_loc_x as x2,
            via.player_id as viaID, via.player as viaPlayer,
            dst.player_id as dstID, dst.player as dstPlayer,
            e2.event_index as ev_id2, e2.end_loc_x as x3,
            dst2.player_id as dstID2, dst2.player as dstPlayer2,
            e3.event_index as ev_id3, e3.end_loc_x as x4
            )
      ) order by minute, second asc;
```

FIND IT IN OUR GRAPH

```
and e.loc_x < e.end_loc_x
and e2.end_loc_x < e.end_loc_x
```

”Up, Back…”

```
and e3.end_loc_x > e2.end_loc_x
and src.player_id != dst.player_id
```

”…through”

…that are linked passes

```
and (e2.event_index - e.event_index) < 10
```

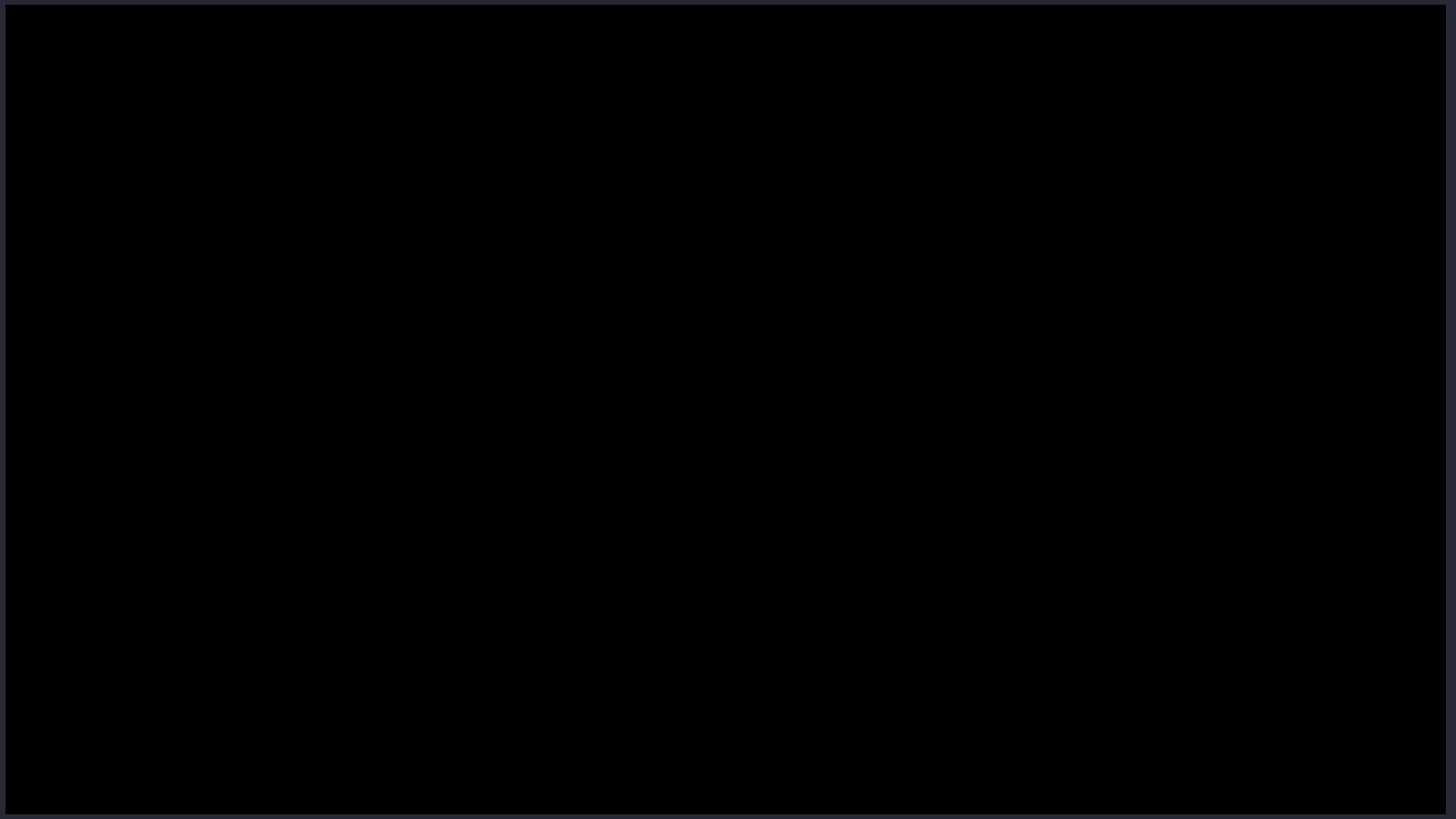
WHAT WE FOUND: UP, BACK, THROUGH

	EV_ID	EV_ID2	EV_ID3	PERIOD	MINUTE	SECOND	TEAM	SRCID	SRCPLAYER	X1	VIAID	VIAPLAYER	DSTID	DSTPLAYER	DSTID2	DSTPLAYER2	X4
1	24	27	30	1	0	19	England Women's	31538	Mary Alexandra Earps	15.9	10252	Alex Greenwood	21.2	Millie Bright	24.5	Lauren Hemp	87.
2	45	47	49	1	1	9	Australia Women's	401635	Clare Hunt	58.4	4979	Katrina Gorry	67.9	Clare Hunt	53.6	Clare Elizabeth Polkinghorne	45.
3	84	86	88	1	2	41	Australia Women's	401635	Clare Hunt	20.4	5095	Ellie Madison Carpenter	27.8	Katrina Gorry	29.3	Hayley Emma Raso	41.
4	86	88	90	1	2	41	Australia Women's	5095	Ellie Madison Carpenter	28.4	4979	Katrina Gorry	29.3	Hayley Emma Raso	41.5	Ellie Madison Carpenter	38.
5	88	90	96	1	2	43	Australia Women's	4979	Katrina Gorry	29.5	6818	Hayley Emma Raso	41.5	Ellie Madison Carpenter	38.9	Mary Boio Fowler	58.
6	96	98	100	1	3	7	Australia Women's	5095	Ellie Madison Carpenter	50.3	35693	Mary Boio Fowler	58.9	Ellie Madison Carpenter	51.2	Hayley Emma Raso	68.
7	113	116	119	1	3	25	England Women's	4642	Millie Bright	26.5	4643	Georgia Stanway	44.8	Jessica Carter	27.9	Millie Bright	34.
8	122	125	128	1	3	37	England Women's	4642	Millie Bright	38.9	19422	Jessica Carter	44.4	Millie Bright	38.3	Alex Greenwood	39.
9	128	131	134	1	3	44	England Women's	4642	Millie Bright	39	10252	Alex Greenwood	39.5	Ella Toone	62.6	Rachel Daly	57.
10	131	134	136	1	3	47	England Women's	10252	Alex Greenwood	43.1	31534	Ella Toone	62.6	Rachel Daly	57.5	Lauren Hemp	75.
11	141	145	149	1	4	5	Australia Women's	5095	Ellie Madison Carpenter	45.7	4961	Samantha May Kerr	66.2	Katrina Gorry	60.7	Hayley Emma Raso	69.
12	156	159	161	1	4	14	Australia Women's	401635	Clare Hunt	33.7	131586	Kyra Lillee Cooney-Cross	42	Stephanie-Elise Catley	38.6	Caitlin Jade Foord	72.
13	167	169	172	1	4	31	England Women's	4643	Georgia Stanway	39	10178	Lucy Bronze	51.2	Jessica Carter	36.4	Millie Bright	33.
14	175	178	180	1	4	38	England Women's	4642	Millie Bright	34.4	19422	Jessica Carter	39.8	Alessia Russo	87.3	Lauren Hemp	10
15	184	188	191	1	4	53	Australia Women's	42787	Mackenzie Arnold	12	401635	Clare Hunt	20.3	Mackenzie Arnold	7.8	Hayley Emma Raso	64.
16	193	197	200	1	5	2	England Women's	10252	Alex Greenwood	55.3	31534	Ella Toone	69.7	Rachel Daly	63.9	Ella Toone	55.
17	211	214	217	1	5	16	England Women's	19422	Jessica Carter	22.6	4642	Millie Bright	26.7	Jessica Carter	35.6	Millie Bright	34.
18	214	217	220	1	5	20	England Women's	4642	Millie Bright	29.8	19422	Jessica Carter	35.6	Millie Bright	34.8	Alex Greenwood	3
19	220	224	227	1	5	27	England Women's	4642	Millie Bright	36.2	10252	Alex Greenwood	39	Millie Bright	32.9	Rachel Daly	51.
20	247	250	254	1	5	48	England Women's	31538	Mary Alexandra Earps	16.7	4642	Millie Bright	26.4	Jessica Carter	31.1	Mary Alexandra Earps	9.

WHAT WE FOUND: UP, BACK, THROUGH

EV_ID	EV_ID2	EV_ID3	PERIOD	MINUTE	SECOND	TEAM	SRCPLAYER	X1	VIAPLAYER	X2	DSTPLAYER	X3	DSTPLAYER2	X4	TOTAL_DISTANCE
131	134	136	1	3	47	England Women's	Alex Greenwood	43.1	Ella Toone	62.6	Rachel Daly	57.5	Lauren Hemp	75.8	32.7

PASSING PATTERN: UP, BACK, THROUGH



WHAT WE FOUND: UP, BACK, THROUGH

```
select team, count(*) as num_UPBACKTHROUGH
From

(SELECT distinct ev_id, ev_id2, ev_id3, period, minute, second, team,
                srcID, srcplayer, x1,   viaID, viaPlayer, x2 dstID…..

) group by team;
```

	TEAM	NUM_UPBACKTHROUGH
1	Australia Women's	30
2	England Women's	49

WHAT WE FOUND: UP, BACK, THROUGH

All the teams in the world cup….

	TEAM	NUM_UPBACKTHROU...
1	Spain Women's	243
2	England Women's	241
3	Japan Women's	171
4	Sweden Women's	164
5	Netherlands Women's	164
6	Germany Women's	158
7	Norway Women's	142
8	United States Women's	102
9	Switzerland Women's	100
10	Brazil Women's	98
11	Australia Women's	93
12	Italy Women's	93
13	France Women's	88
14	Argentina Women's	74
15	Colombia Women's	71

Top matches….

	TEAM	NUM_UPBACKTHROU...
1	Spain Women's	243
2	England Women's	241
3	Japan Women's	171
4	Sweden Women's	164
5	Netherlands Women's	164
6	Germany Women's	158
7	Norway Women's	142
8	United States Women's	102
9	Switzerland Women's	100
10	Brazil Women's	98
11	Australia Women's	93
12	Italy Women's	93
13	France Women's	88
14	Argentina Women's	74
15	Colombia Women's	71

NICE TOUCH....

EXPLAIN PLAN FOR
SELECT distinct match_id, ev_id, ev_id2, ev_id3, period....

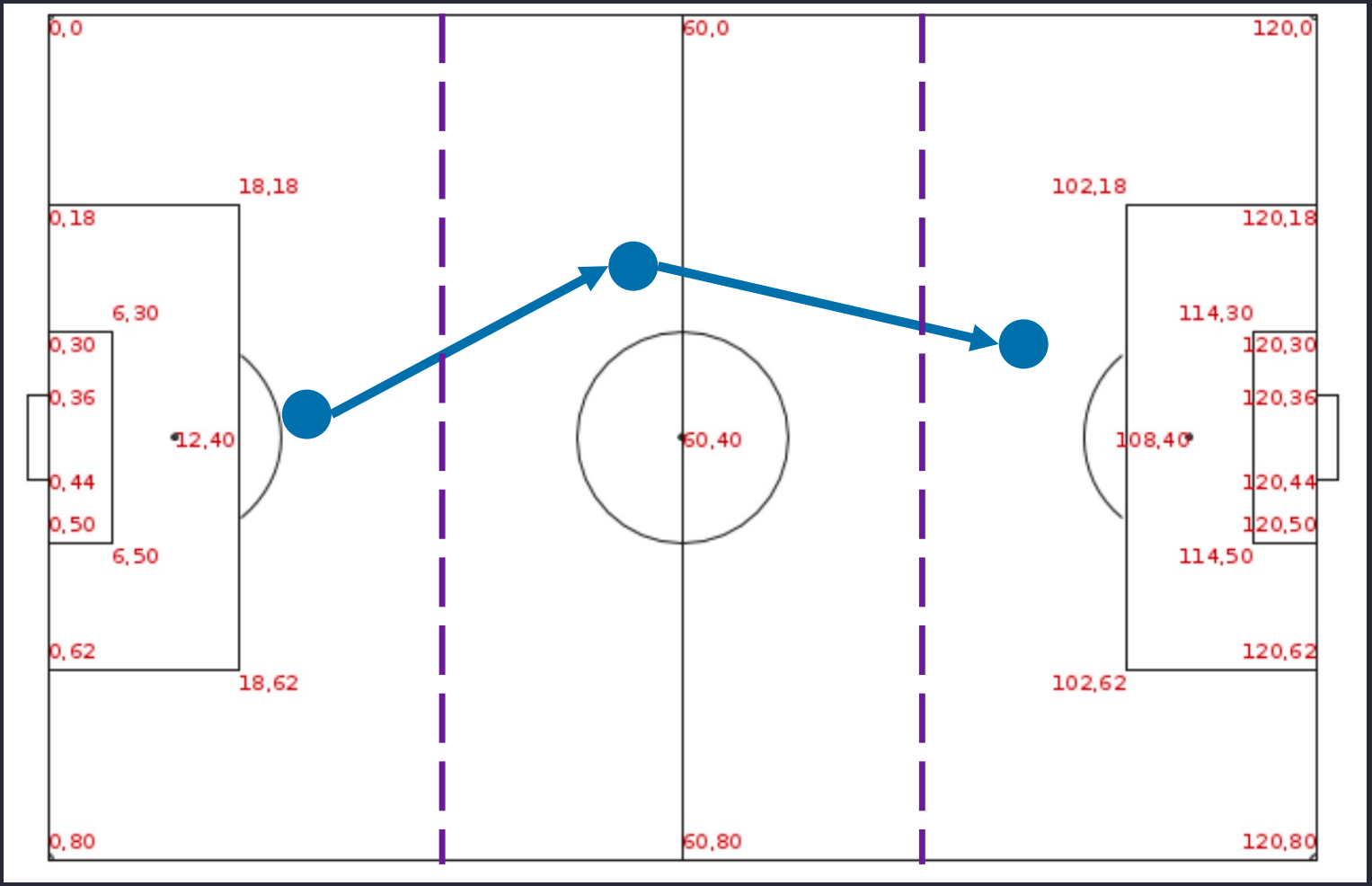
SELECT * FROM TABLE(DBMS_XPLAN.DISPLAY(format=>'ALL'));

PLAN_TABLE_OUTPUT

```
Plan hash value: 472747197

-----------------------------------------------------------------------------------------------------------------
| Id  | Operation                    | Name                    | Rows  | Bytes | Cost (%CPU)| Time     |
-----------------------------------------------------------------------------------------------------------------
|   0 | SELECT STATEMENT             |                         |     4 |   960 |   911   (1)| 00:00:01 |
|   1 |  SORT ORDER BY               |                         |     4 |   960 |   911   (1)| 00:00:01 |
|   2 |   HASH UNIQUE                |                         |     4 |   960 |   910   (1)| 00:00:01 |
|*  3 |    HASH JOIN                 |                         |     4 |   960 |   909   (1)| 00:00:01 |
|*  4 |     HASH JOIN                |                         |     4 |   860 |   905   (1)| 00:00:01 |
|*  5 |      HASH JOIN               |                         |     4 |   760 |   901   (1)| 00:00:01 |
|*  6 |       HASH JOIN              |                         |     4 |   660 |   897   (1)| 00:00:01 |
|*  7 |        HASH JOIN             |                         |     4 |   488 |   893   (1)| 00:00:01 |
|*  8 |         HASH JOIN            |                         |    47 |  3995 |   595   (1)| 00:00:01 |
|*  9 |          TABLE ACCESS FULL   | WWC23_PASS_TRANSACTIONS |    91 |  4368 |   298   (1)| 00:00:01 |
|* 10 |          TABLE ACCESS FULL   | WWC23_PASS_TRANSACTIONS |   182 |  6734 |   298   (1)| 00:00:01 |
|* 11 |         TABLE ACCESS FULL    | WWC23_PASS_TRANSACTIONS |   182 |  6734 |   298   (1)| 00:00:01 |
|  12 |        TABLE ACCESS FULL     | WWC23_PLAYERS           |   616 | 26488 |     4   (0)| 00:00:01 |
|  13 |       TABLE ACCESS FULL      | WWC23_PLAYERS           |   616 | 15400 |     4   (0)| 00:00:01 |
|  14 |      TABLE ACCESS FULL       | WWC23_PLAYERS           |   616 | 15400 |     4   (0)| 00:00:01 |
|  15 |     TABLE ACCESS FULL        | WWC23_PLAYERS           |   616 | 15400 |     4   (0)| 00:00:01 |
-----------------------------------------------------------------------------------------------------------------
```

BUILDING PLAY THROUGH THE THIRDS



IT ALL STARTS WITH DATA....

We all have to make trades.....

Time.....

Money....

Processes....

Next trade... what data do you need?

Dr. Abi Giles-Haigh

Oracle Analytics and Innovation Director, Capgemini

Abi.Giles-Haigh@capgemini.com

@Abi_Giles_Haigh

About Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of 270,000 team members in nearly 50 countries. With its strong 50 year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fuelled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2020 global revenues of €16 billion.

Get the Future You Want | www.capgemini.com