DATA ANALYSIS PROJECT

PLAYER SCOUTING – MLS FRANCHISE

&

TRACKING DATA VISUALIZATION

CONTENTS

PART-1 : PLAYER SCOUTING FOR MLS FRANCHISE - PROCESS

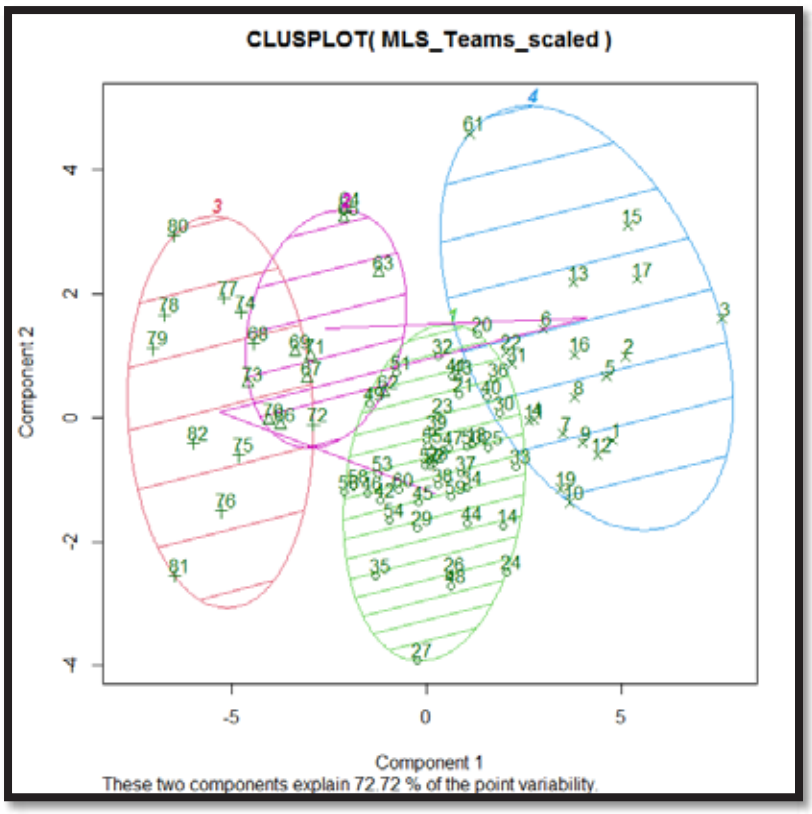
PART-2 : TRACKING DATA VISUALIZATION

PART-1 : PLAYER SCOUTING FOR MLS FRANCHISE - PROCESS

1. Identification of MLS Teams with similar playing style:

1.1 Classifying MLS teams into clusters based on K-means clustering Algorithm:

The entire set of MLS teams from "Intern_Teams.xlsx" file are applied with K-means Clustering algorithm to group them into clusters based on similar playing style. The output clusters are as follows:

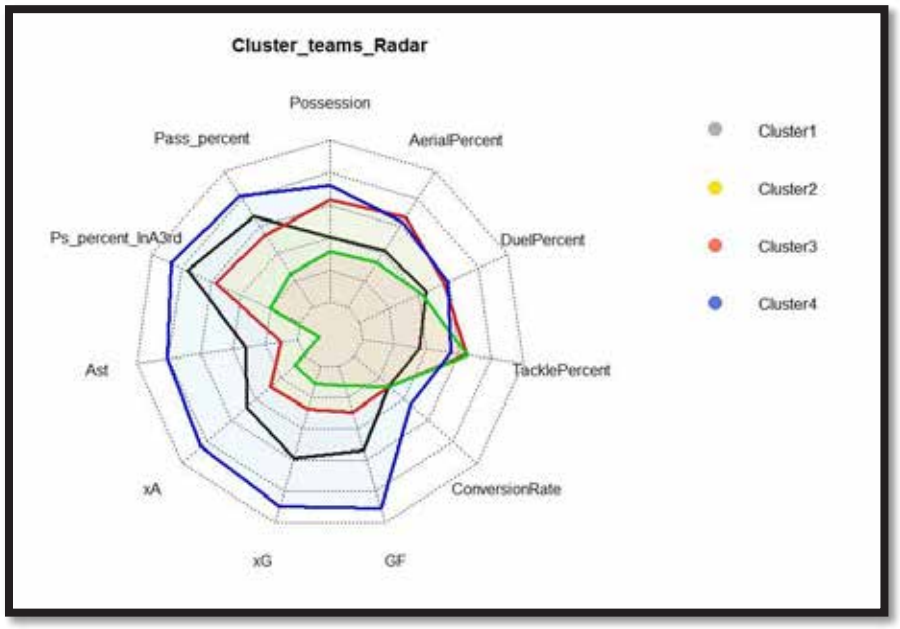


Thus, the cluster model has grouped the MLS teams into **four** clusters based on their similar attributes. About **72.72%** of variations in data points are explained by this model. The aggregates measures of every attribute for each cluster are given below:

Cluster	Poss%	Pass%	xA	xG	Tackle%	Duel%	Aerial%
1	0.480854	0.783878	29.27463	51.47463	0.385659	0.495171	0.491659
2	0.5264	0.7571	23.562	37.429	0.4149	0.5071	0.5159
3	0.463	0.699818	17.46364	30.12455	0.415545	0.491818	0.483
4	0.5443	0.81375	40.422	65.311	0.40485	0.50955	0.51225

1.2 Identifying playing style of MLS teams' clusters through Radar plots & Visualization:

The clusters identified are visualized through radar plots as follows:



- **Playing styles – Interpretation:**

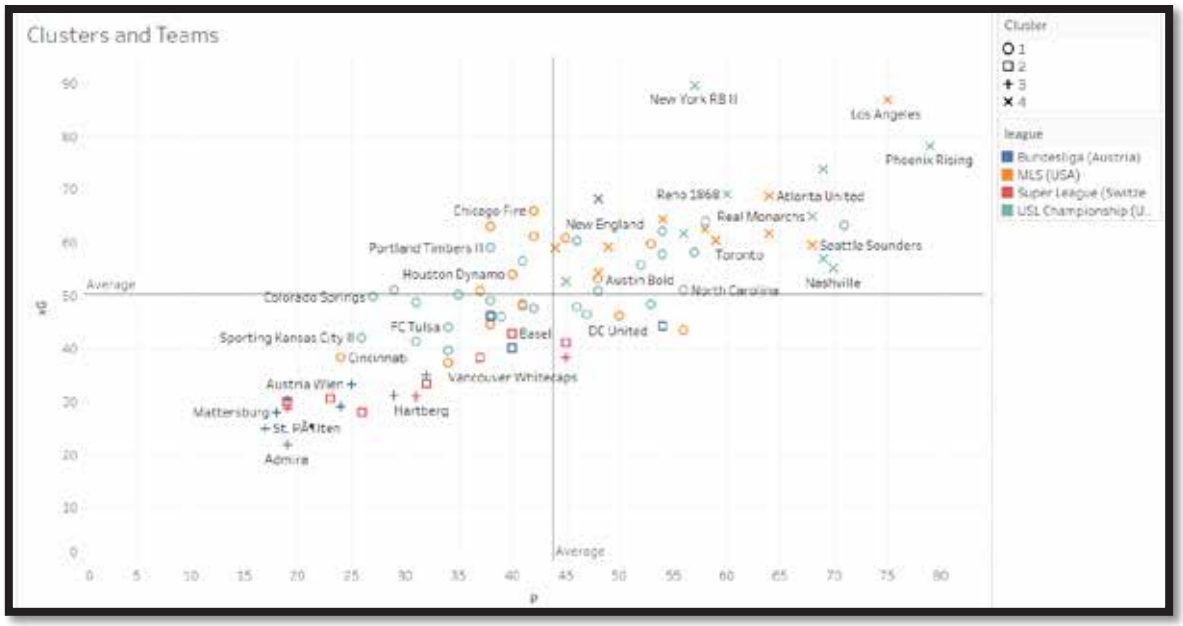
Cluster – 1 **-- >** High passing team with more focus on attack

Cluster – 2 **-- >** Teams with Defensive, counter-attacking style of football.

Cluster – 3 -- > Teams with Possession-based football with fair balance of attack and Defense

Cluster – 4 -- > Teams with Solid-attacking style of football

- **Visualization of Teams within each cluster:**



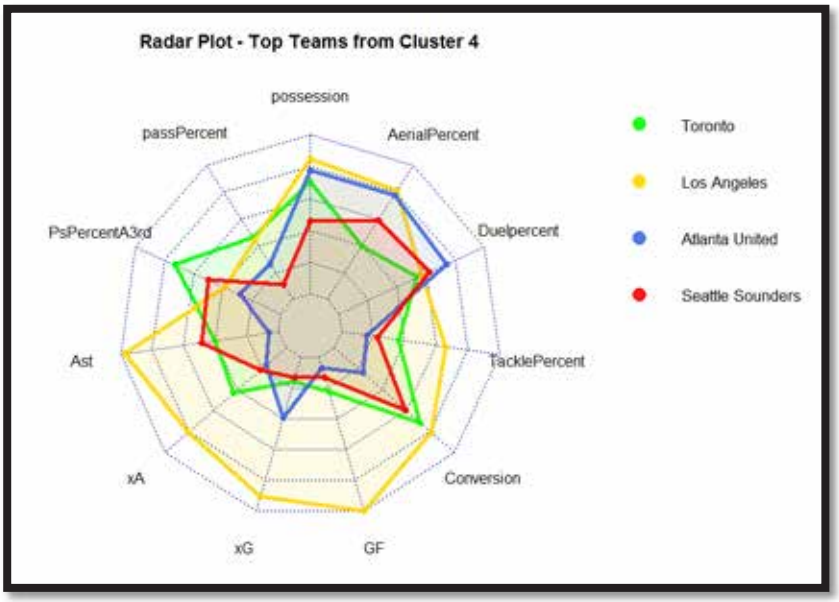
Teams within Cluster 2 & 3 - Low on points and xG metrics.

Teams within Cluster 1 & 4 – High on Points and xG metrics

1.3 Finalizing on the MLS Franchise to scout players:

My personal favourite playing style in football would be "Solid attacking style". So, naturally I would want to choose an MLS franchise from Cluster – 4.

Visualizing the top 4 MLS teams from Cluster – 4:

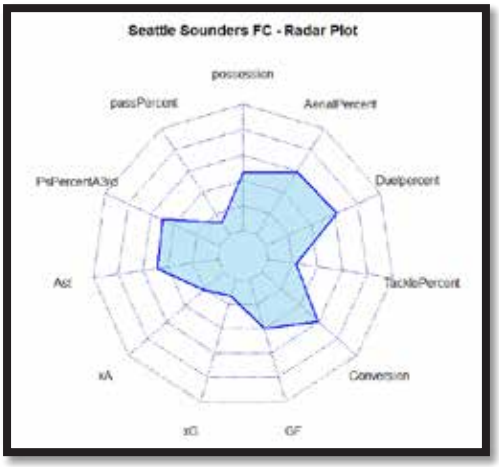


Teams		Top Strengths
Toronto	-	Pass % ; Pass % in Att 3rd ; Conversion rate & GF.
Los Angeles	-	Possession ; Pass % ; GF & Ast.
Atlanta United	-	Possession ; Aerial & Duel % ; GF.
Seattle Sounders	-	GF & Assists; Duel & Aerial % ; Pass % in Att 3rd ; Conversion rate.

.

Selection of MLS Team

Based on above visualizations, I would choose to scout for **Seattle Sounders FC.**



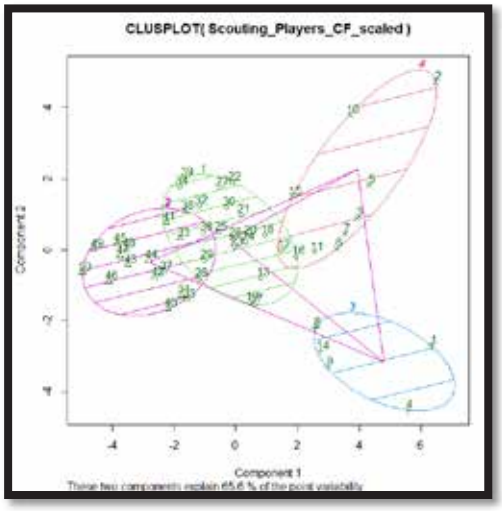
Reasons for Selection:

- ✓ High Conversion rate with higher GF & Assists than xG and xA respectively – > emphasizing their attacking style of football
- ✓ Their average age of their roster is 29 and especially the age of their players for the required scouting positions (RB, RW/RM and ST) are above 28 - > opportunity for addition of young, talented players to their rosters.

2. Scouting of Centre Forwards / Strikers

2.1 Filtering & Classification of CF/ST using Cluster algiorithm:

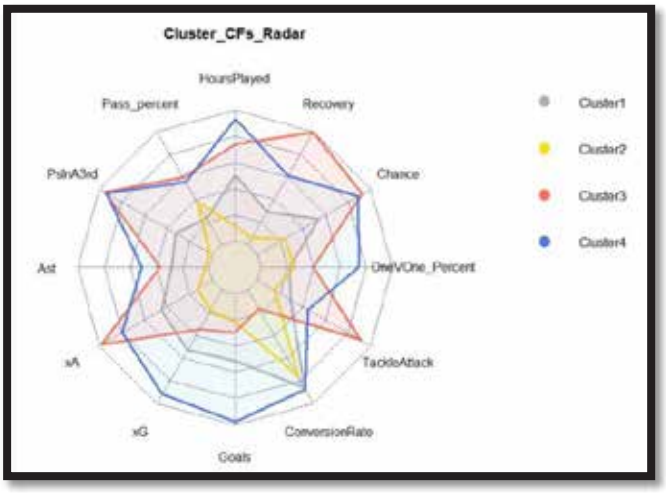
The Strikers / Centre Forwards from the same cluster as that of Seattle Sounders (Cluster 4) are filtered out and K-means clustering algorithm is again applied on them to segregate the different players based on their similar attributes:



Thus, the cluster model has grouped the MLS teams into **four** clusters based on their similar attributes. About **65.6%** of variations in data points are explained by this model. The aggregates measures of every attribute for each cluster are given below:

Cluster	Min	xA	xG	Goal	Conversion	TcklAtt	Ast	Chance
1	1667.10	2.09	11.01	10.81	18.64	21.62	3.57	22.05
2	848.88	0.97	4.47	5.13	17.67	15.88	1.25	10.50
3	2162.20	3.92	7.53	6.60	11.66	46.20	4.40	37.00
4	2564.00	3.29	18.40	19.63	18.74	27.63	5.63	35.63

Visualizing the Clusters of CF/ST:



Key Points:

- ✓ Top 2 clusters are Cluster 4 and Cluster 3.
- ✓ Cluster 4 contains Strikers with highest conversion rates ,Goals and passing rates in Att 3rd with less assists and defensive capabilities.
- ✓ Cluster 3 contains Strikers with almost similar attacking capabilities as Cluster 4 but they are also very much useful in defense.

2.2 Comparison of top players from a Clusters – 3 & 4:

The top players from the Clusters 3 and 4 are filtered out and are compared as below:

D.Rios

Age	25
Country	Mexico
Club	Nashville SC
Wages	$91,625

J.Martinez

Age	26
Country	Venezuela
Club	Atlanta United
Wages	$3,058,333

V. Castellanos

Age	21
Country	Argentina
Club	New York City
Wages	$274,806

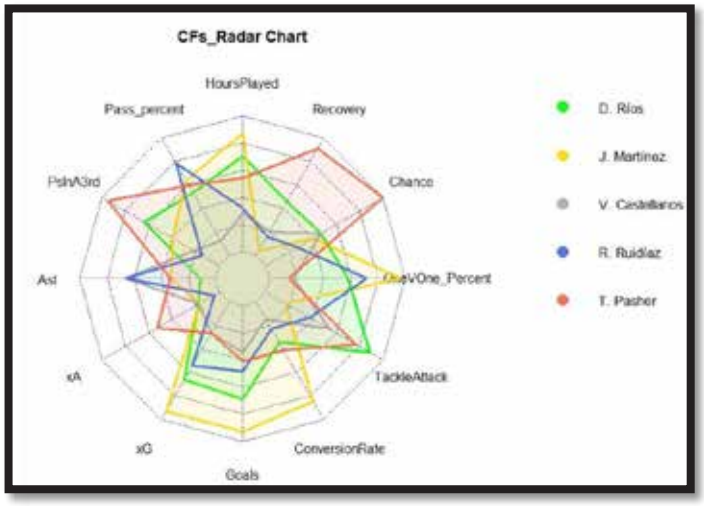
T. Pasher

Age	25
Country	Canada
Club	Indy Eleven
Wages	-

R. Ruidíaz

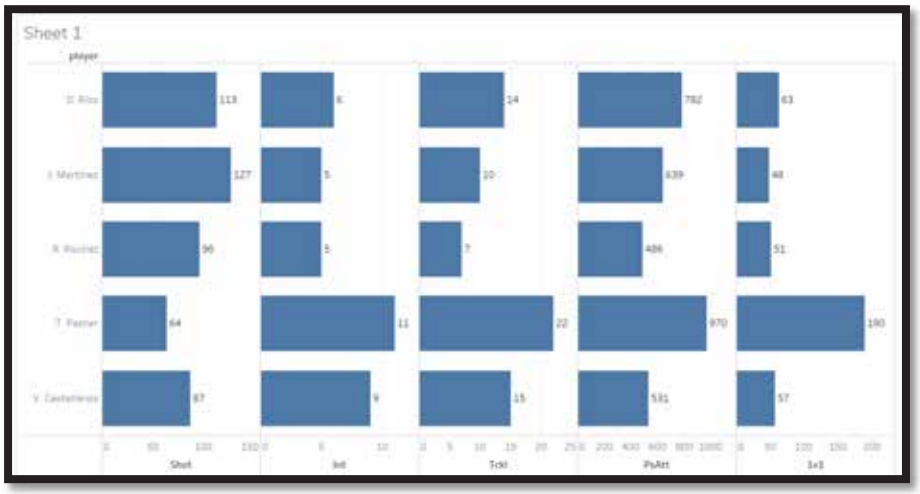
Age	29
Country	Peru
Club	Seattle Sounders
Wages	$2,000,004

Radar Plot – CF:



:

Comparison of additional attributes:



2.3 Final selection of Right-fit player from the pool of players

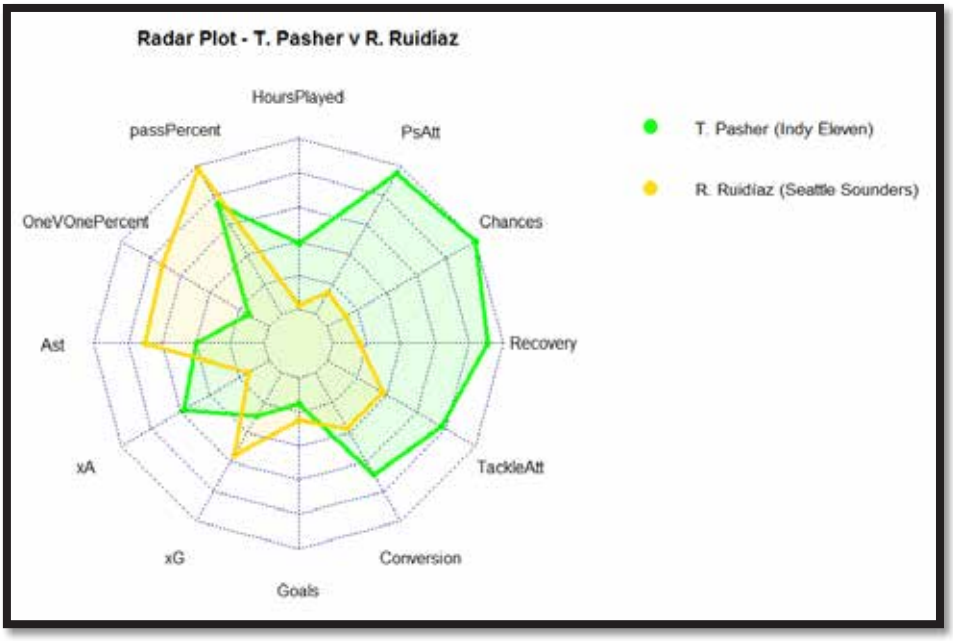
From the above list of Centre Forwards, the most suited player for Seattle Sounders' style of play would be

"Tyler Pasher"

Reasons for Selection:

- ✓ As mentioned previously, Seattle's top strengths include **GF & Assists; Duel & Aerial % ; Pass % in Att 3rd & Conversion rate** --> indicating their all round strengths in Attack & Defense.
- ✓ Given the Seattle Sounders' forward **Raul Ruidiaz** is a well-established Striker at the club and has excellent goal-scoring capabilities, he lacks work rate and he **doesn't track back** well enough to help the defense. So there is a **need for someone to complement his capabilities**.
- ✓ On the other hand, **Tyler Pasher** – inspite of having a **slightly lesser goal scoring records**, he compensates for it with his **excellent Conversion rate and Passing records in Att 3rd**, together with his **excellent defensive capabilities**. So, he is regarded as a all-round Striker who also tracks back and helps in defense.

Radar Plot – Tyler Pasher V Ruidiaz



Key Points:

As evident from this plot,

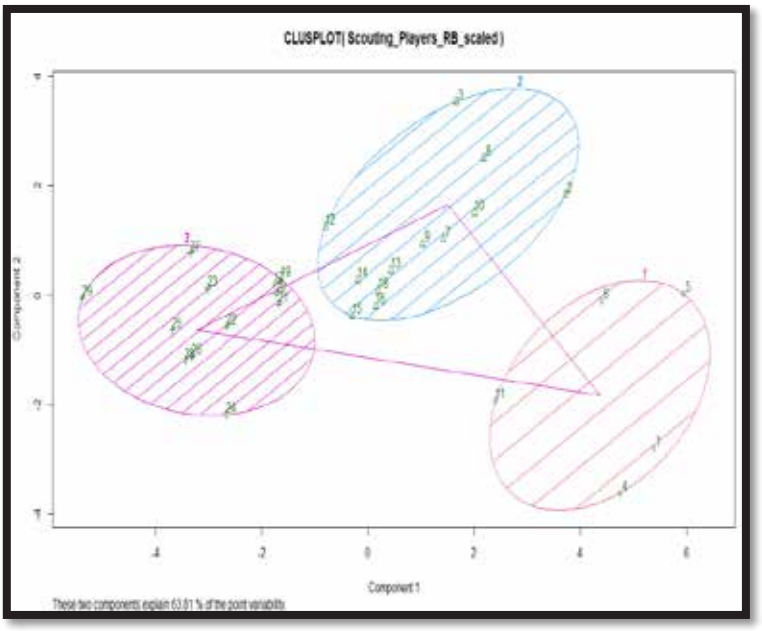
- ✓ Tyler Pasher clearly has the edge over Ruidiaz in Passes in Att 3rd, Chances created, Ball Recoveries and tackles in Att 3rd and also Conversion rate - > **indicative of his work rate and creative abilities**

- ✓ Although Pasher is slightly weak on his Goal scoring and Assists, his strenghts would definitely complement Ruidiaz together making them a solid partnership.

3. Scouting of Right back players

3.1 Filtering & Classification of RB players using Cluster algiorithm:

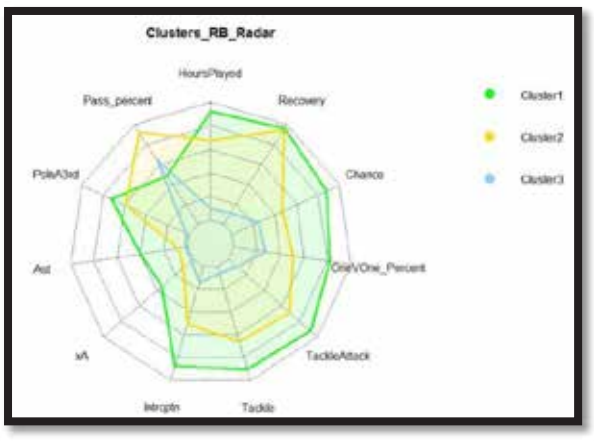
The RB players from the same cluster as that of Seattle Sounders (Cluster 4) are filtered out and K-means clustering algorithm is again applied on them to segregate the different players based on their similar attributes:



Thus, the cluster model has grouped the MLS teams into three clusters based on their similar attributes. About 63.81% of variations in data points are explained by this model. The aggregate measures of every attribute for each cluster are given below:

Clusters	Min	xA	Ast	Int	Tckl	TcklAtt	Chances	Pass%
1	2561.60	4.10	4.60	45.00	60.80	85.00	31.00	77.28
2	2121.17	1.95	1.67	30.33	49.75	69.83	17.00	83.875
3	1075.83	1.14	1.17	15.83	22.83	31.25	9.33	79.7

Visualizing the Clusters of RB:



Key Points:

- ✓ Top RB players are contained in Cluster 1.
- ✓ Except Pass %, the players in Cluster 1 are excellent in all other attributes including Sounder's areas of strengths like Pass% in Att 3rd, higher Assists than xA metric.

3.2 Comparison of top RB players from a Clusters -1:

The top RB players from the Cluster 1 are filtered out and are compared as below:

Auro

Age	24
Country	Brazil
Club	Toronto FC
Wages	$276,666

T.Thompson

Age	24
Country	USA
Club	San Jose Earthquakes
Wages	$174,999

R.Cannon

Age	21
Country	USA
Club	FC Dallas
Wages	$80,250

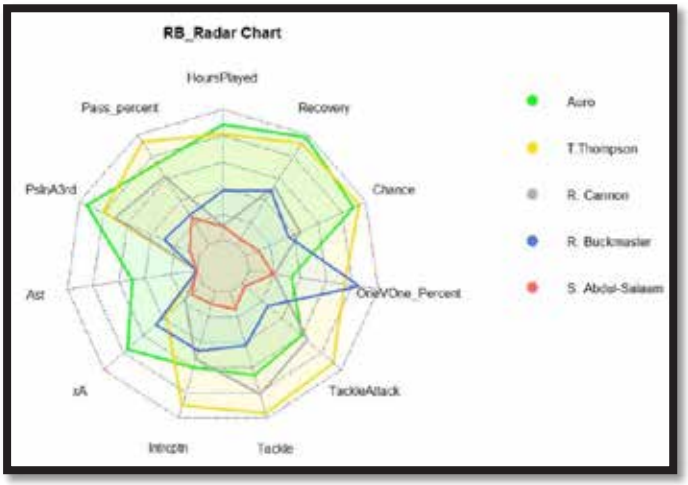
R. Buckmaster

Age	23
Country	USA
Club	NewYork RedBulls
Wages	$56,250

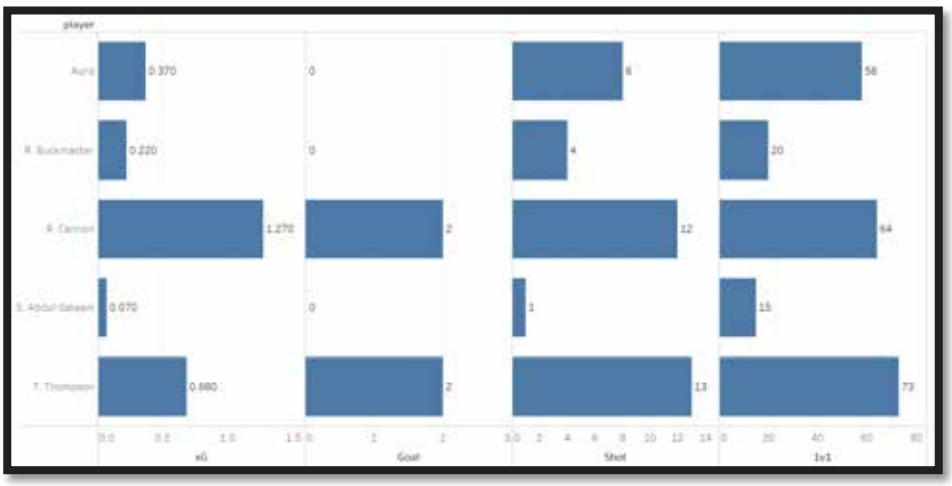
S. Salaam

Age	28
Country	USA
Club	Seattle Sounders
Wages	$70,250

Radar Plot –RB:



Comparison of additional attributes:



3.3 Final selection of Right-fit player from the pool of players

From the above list of RB players, the most suited player for Seattle Sounders' style of play would be

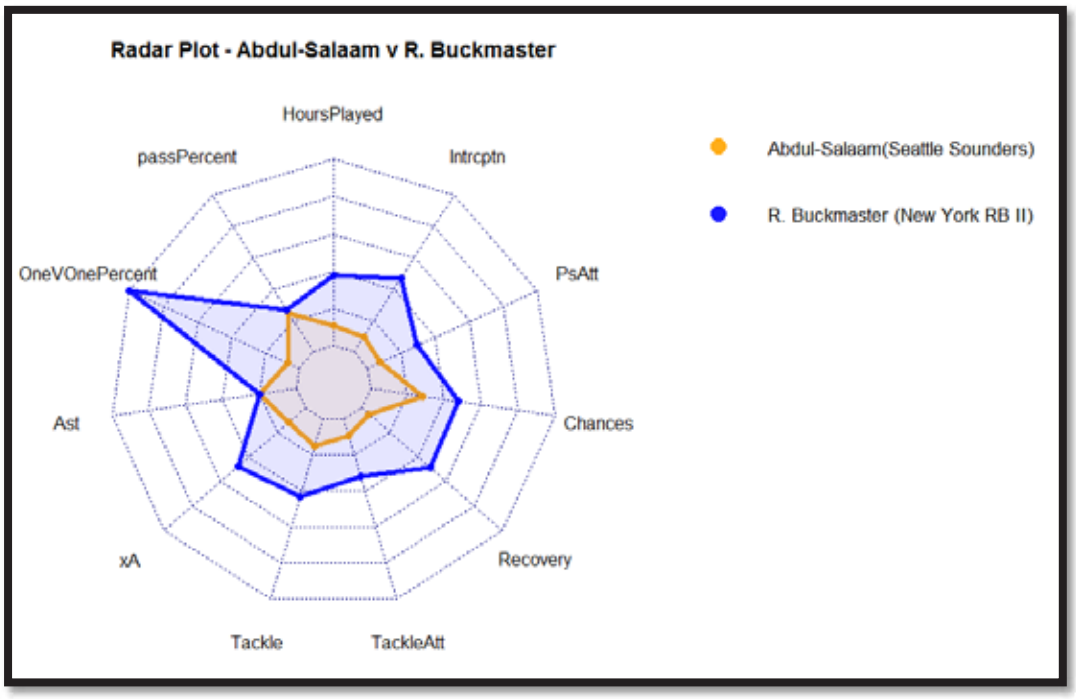
"Rece Buckmaster"

Reasons for Selection:

- ✓ Seattle Sounders already have two right-backs in their roster at the moment – Kelvin Leerdam and S.Abdul Salaam (both in their late 20s). So, it's a fantastic opportunity to add fresh, young (U-24) and talented youth to their Roster to add to their strengths who is expected to play the role of a back-up.
- ✓ Also, R. Buckmaster's guaranteed compensations are lesser compared to S.Abdul Salaam.

Technical Reasons:

Radar Plot – R.Buckmaster v S.Abdul Salaam



Key Points:

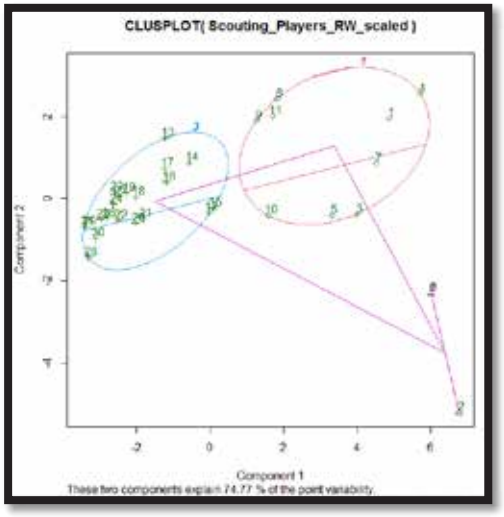
The above plot clearly indicates:

- ✓ R.Buckmaster is better than S.Salaam in almost all attributes .
- ✓ Together with his younger age and lesser compensation, he also adds great value to the side through his excellent ratings in all the above attributes.
- ✓ Thus, he is definitely a very good-fit to play the role of a squad player, who can be used on rotation with first choice RB, developing himself and the team's value.

4. Scouting of Right Winger/ Right midfield players

4.1Filtering & Classification of RB players using Cluster algiorithm:

The RB players from the same cluster as that of Seattle Sounders (Cluster 4) are filtered out and K-means clustering algorithm is again applied on them to segregate the different players based on their similar attributes:

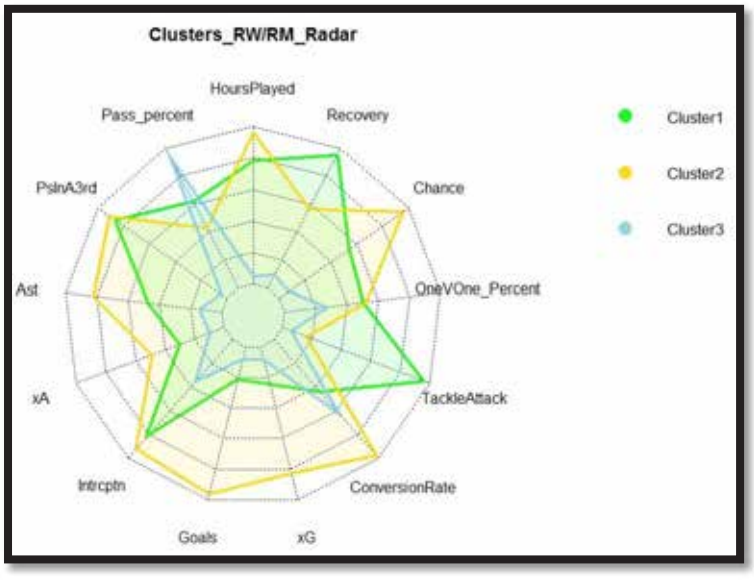


`

Thus, the cluster model has grouped the MLS teams into **three** clusters based on their similar attributes. About 74.77% of variations in data points are explained by this model. The aggregate measures of every attribute for each cluster are given below:

Cluster	Min	xA	xG	Goal	Ast	Chances	Pass%	Conversion
1	2548.667	6.043333	7.372222	6.333333	9.444444	58.77778	0.780778	8.850932
2	2923	9.78	24.885	29	16.5	96.5	0.7715	19.59459
3	1009.7	1.9025	2.407	2.55	2.9	16.8	0.79675	12.62376

Visualizing the Clusters of RW/RM:



Key Points:

- ✓ The clusters containing top RW/RM players are Clusters 1 & 2.
- ✓ Players in Cluster 1 are excellent in attacking abilities as evident from their higher Goals, assists, Chances created and conversion rate
- ✓ Players in Cluster 2 are well equipped with defensive ablitites as evident from their higher ball recoveries, tackle rates.
- ✓ Players in Cluster 2 also possess high ratings in Sounder's strengths of pass % in Att 3rd and higher assists than xA metric -- > indicates their value to the attacking threat of their sides.

4.2 Comparison of top RW/RM players from a Clusters -1:

Cluster 2 contains only two players with both of them above 29 years of age. Since Seattle's current RW player is 28 years old, it wouldn't make sense to bring an older player than him. Thus, comparing top players from Cluster 1:

J.Gressel

Age	26
Country	Germany
Club	Atlanta United
Wages	$132,999

C.Espinoza

Age	25
Country	Argentina
Club	SJ Earthquakes
Wages	$550,000

B.Gomez

Age	26
Country	Argentina
Club	San Antonio
Wages	-

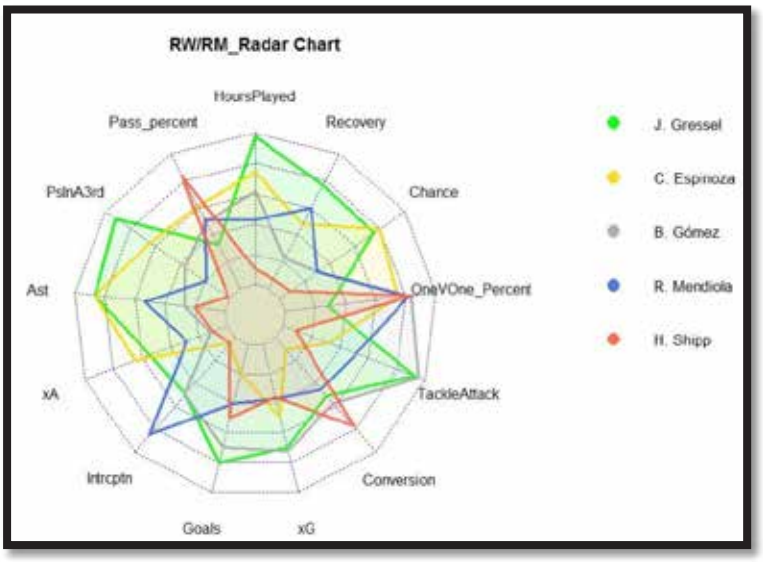
R.Mendiola

Age	25
Country	Mexico
Club	Reno 1868
Wages	-

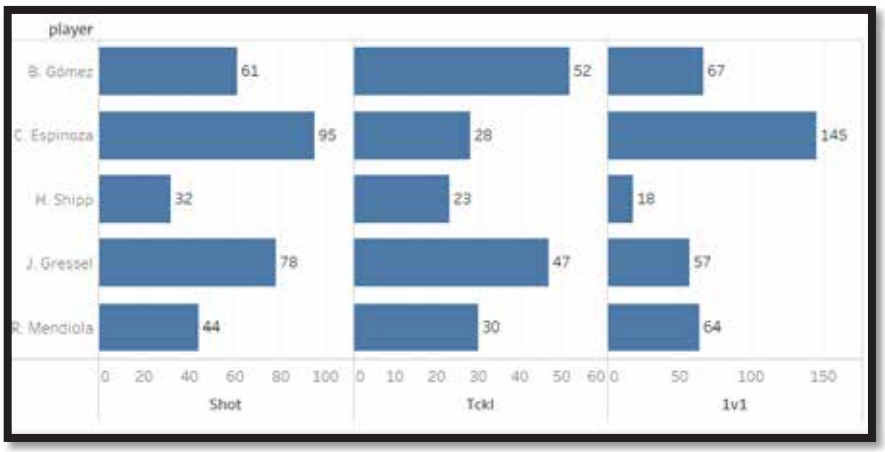
H.Shipp

Age	28
Country	USA
Club	Seattle Sounders
Wages	$234,999

Radar Plot –RW/RM:



Comparison of additional attributes:



4.3 Final selection of Right-fit player from the pool of players

From the above list of RB players, the most suited player for Seattle Sounders' style of play would be

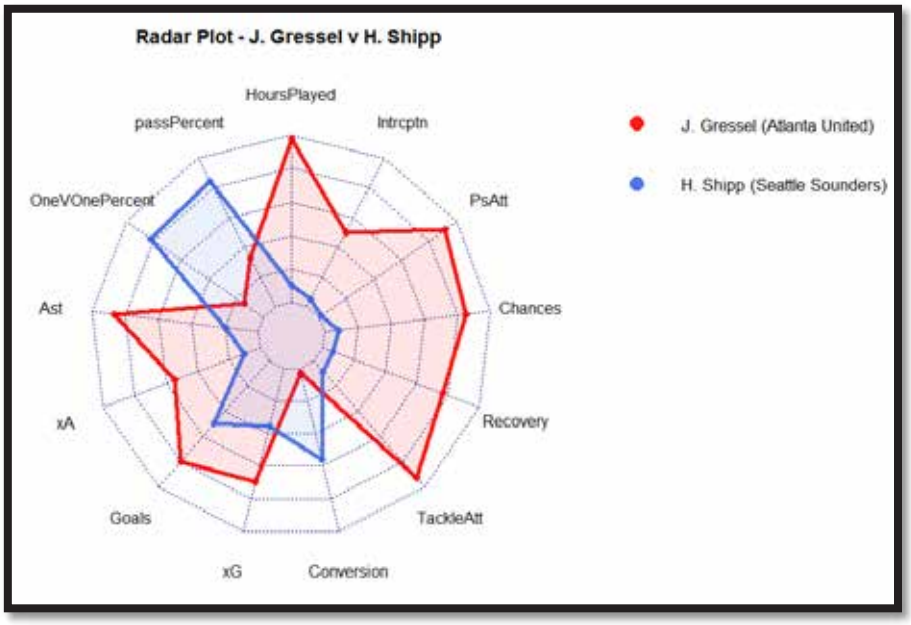
"J.Gressel"

Reasons for Selection:

- ✓ J.Gressel is **younger** compared to the current RW at Sounders – Harry Shipp
- ✓ J.Gressel **out performs H.Shipp in majority of the attributes** as shown below
- ✓ J.Gressel's **guaranteed compensation is lesser** compared to H.Shipp

Technical Reasons:

Radar Plot – J.Gressel v H.Shipp



Key Points:

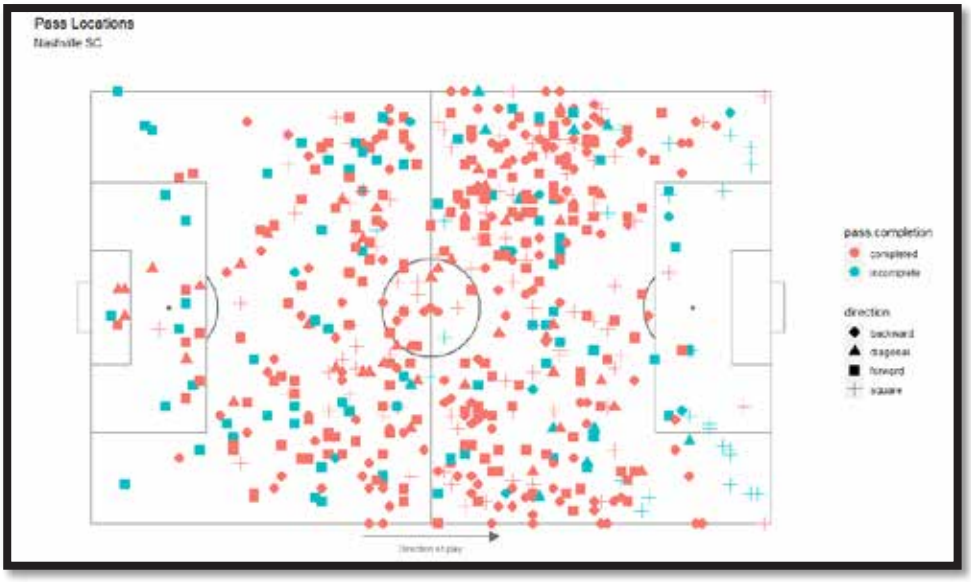
- ✓ J.Gressel possesses higher ratings than H.Shipp in key areas of Sounder's strengths as in higher goals and assists than the xA and xG metric -- > indicates his better finishing abilities.
- ✓ He also boasts of exceptional **work rate** as evident from his high ratings of ball recoveries, chances creation and passes in Att 3rd.
- ✓ Also, his high pressing ability is very evident from his high rating in tackles in Att 3rd.
- ✓ Thus, J.Gressel is an overall complete midfielder with right balance of attacking and defensive abilities which is very much lacking currently at Sounders' side.

PART-2 : TRACKING DATA VISUALIZATION

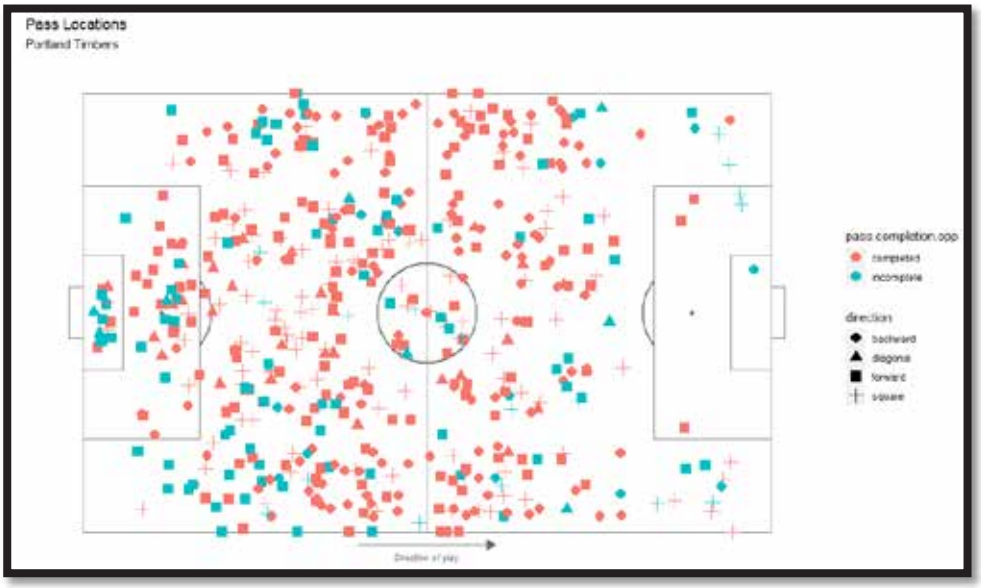
5. Visualization of Pass Locations

5.1 Plotting of overall passes – Nashville SC:

The passes made by Nashville SC and Portland are filtered from given tracking dataset and are visualized as follows:



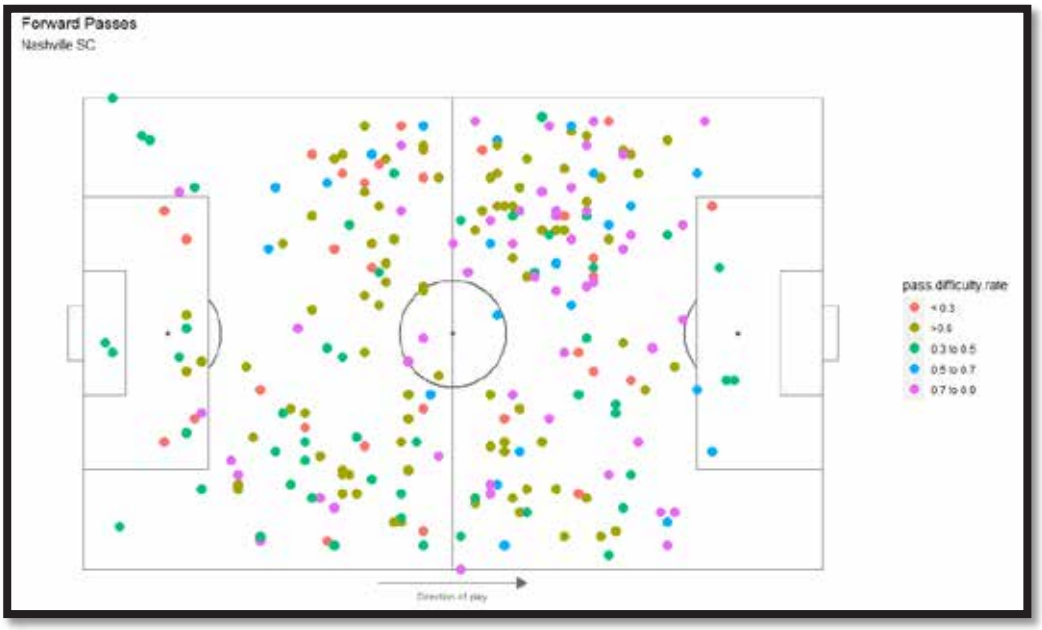
5.2 Plotting of overall passes – Portland Timbers:



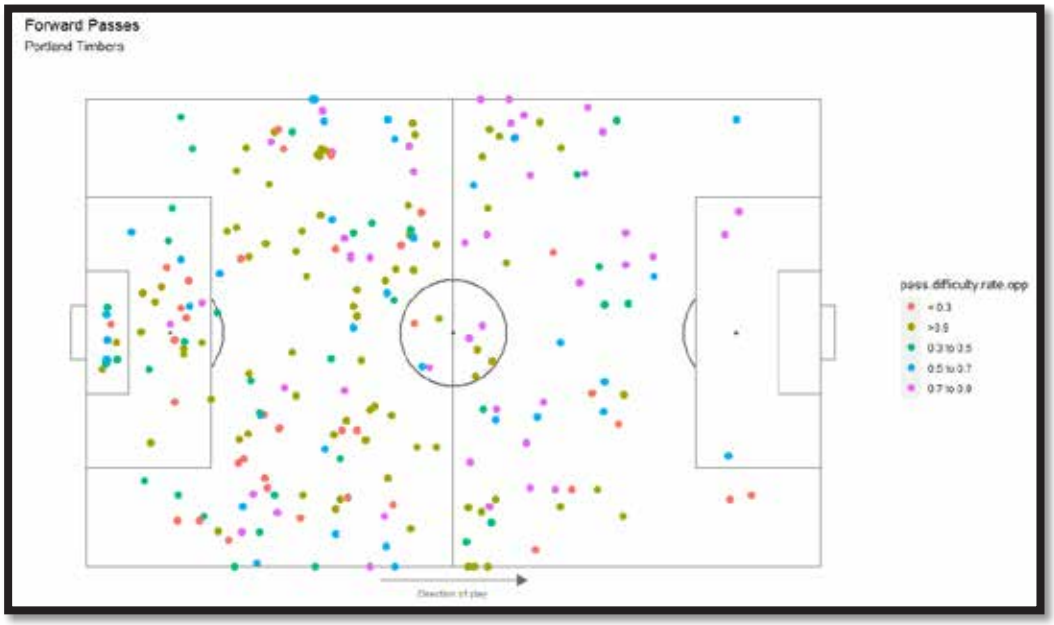
Key Points:

- ✓ It can be clearly seen from the Tracking data that Nashville SC has made more passes in their attacking half compared to Portland -- > **indicating that Nashville dominated the attacking play**
- ✓ Also, Portland made more incomplete passes within their defensive penalty box and majority of their passes are made in their defensive half -- > **indicating that Portland were defensive**.

➢ Plotting of forward passes – Nashville SC:



➢ Plotting of Forward passes – Portland Timbers:



Key Points:

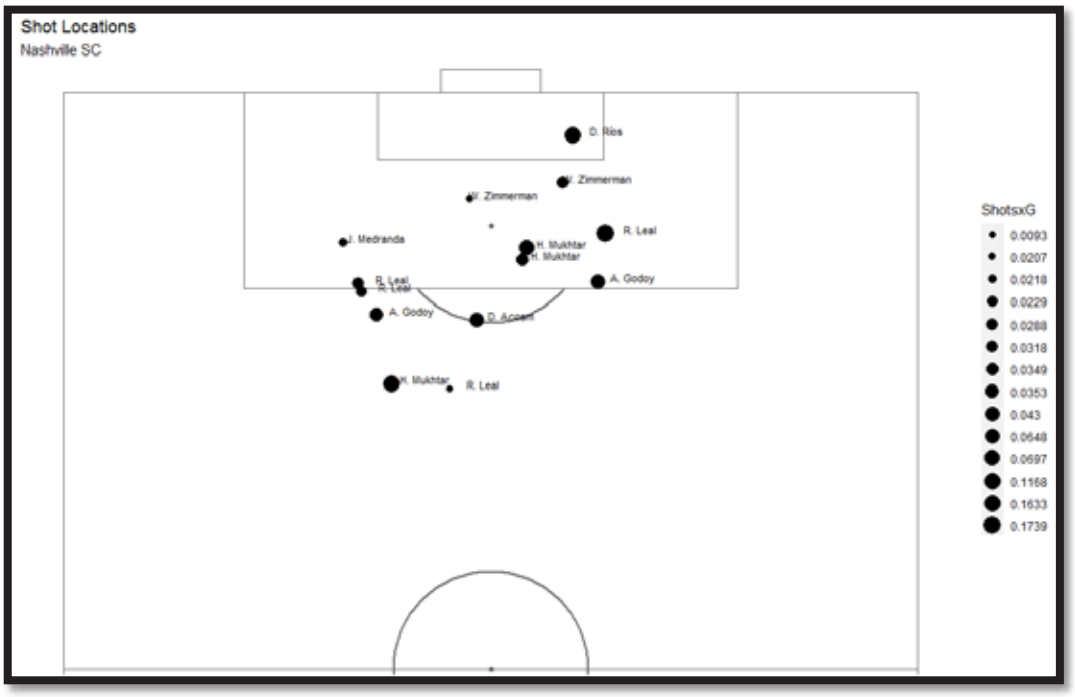
- ✓ Nashville FC have attempted majority of their forward passes from their left channel (attack & defensive half) with most of the highly difficult passes (prob < 0.3) coming through from their left defensive half -- > indicative of their efforts to **attack from their left channel**.

- ✓ Majority of Portland's forward passes were made from their defensive half with large percent of highly difficult passes (prob < 0.3) coming from defensive half as well

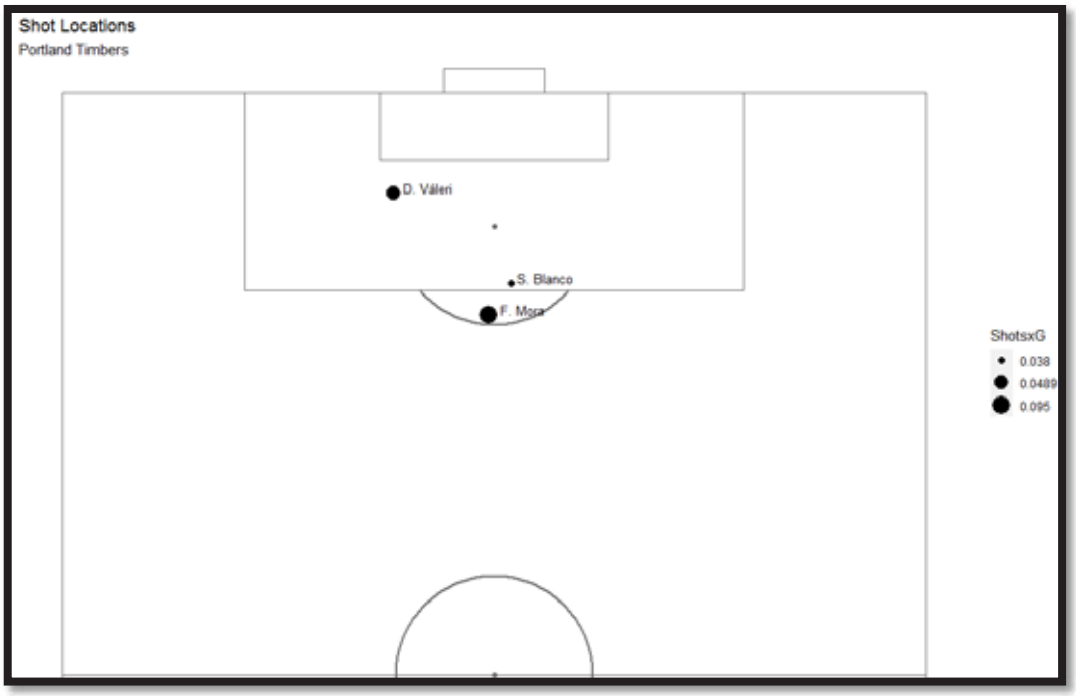
[Note: The lesser the pass difficulty rate, more difficult will be the the pass]

6. Visualization of Shot Locations:

6.1 Plotting of Shot locations – Nashville SC:



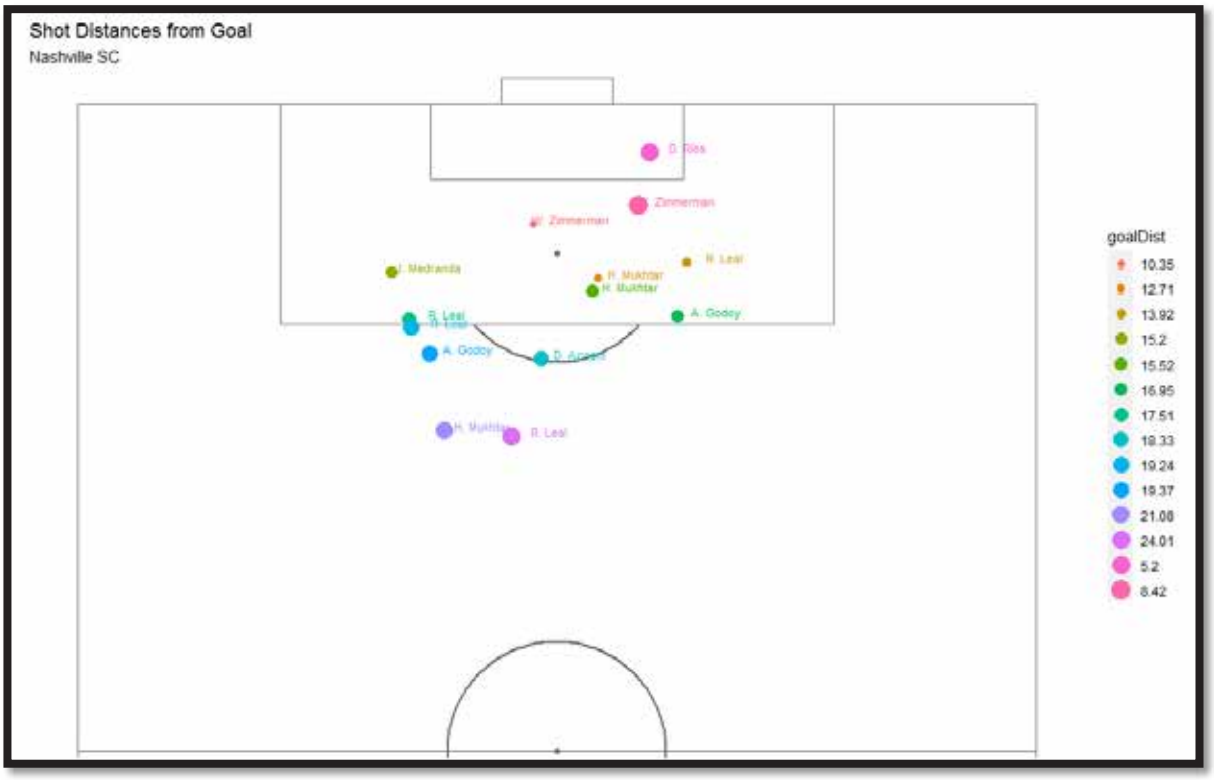
6.2 Plotting of Shot locations – Portland Timbers;



Key Points:

- ✓ Nashville SC have attempted 14 shots in total with majority of them coming from inside their penalty box (9).
- ✓ Portland have attempted just 3 shots in the entire match -- > indicating their lack of attacking gameplay in this match.

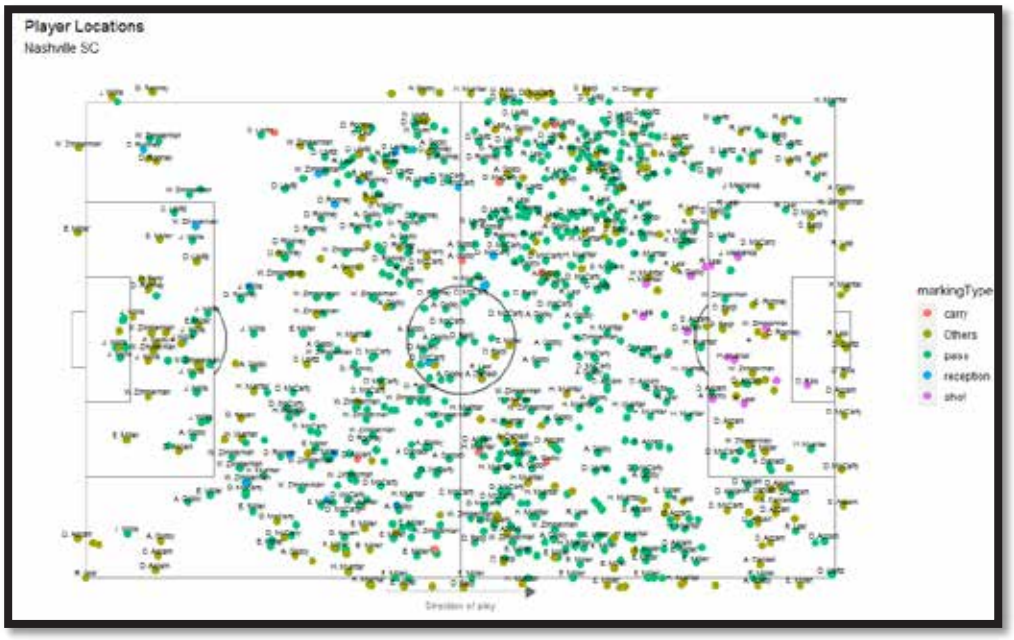
➢ Plotting of Shot Distances – Nashville SC:



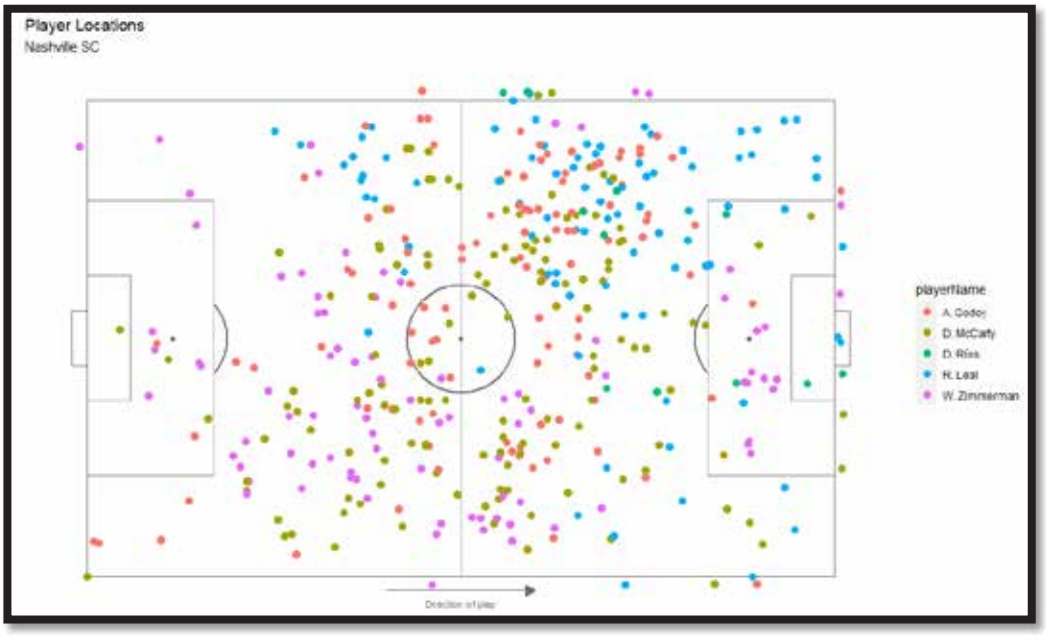
➢ Plotting of Shot Distances – Portland Timbers;

7. Visualization of Player Positions

7.1 Plotting of all Player positions – Nashville SC:



7.2 Plotting of selected Player positions – Nashville SC (McCarty, Godoy, Zimmerman, Leal & Rios)



Key Points:

- ✓ A.Godoy's majority of positions are on the mid-third on the left channel – > indicates he is a **Left midfielder**
- ✓ D.McCarty's majority of positions are on the mid-third of pitch - > indicates he is a **Central Midfielder**
- ✓ D.Rios has mainly positioned himself in attacking third (central channel) - > indicates he is a **Forward**
- ✓ R.Leal's positions are mainly on the left channel of attacking half - > indicates he is a **Left winger**
- ✓ W.Zimmerman's positions are recorded mainly on the defensive half - > indicates he is a **Centre Back** (along with few marks in the attacking penalty box – for set-pieces probably)

[Note: The positions outside the touchline represent their locations during throw-ins, Corners etc]