Exploratory Analysis: A Case of Netball Australia

Table of Contents

Executive Summary

This report demonstrates exploratory analysis of the dataset provided by the Netball Australia which includes the information about matches, players, squads and their performance measures since 2009. Representing the best team of the world and being the most willing destination for the best player all over the globe for Netball, Netball Australia is interested in digging information for the past data which can be useful for increasing the competitiveness of Netball in Australia.

The report aims to explore the following issues on Australian national team:

- Efficiency of benchmark for choosing National Team Players
- Performance impact of split of Australia and New Zealand Championship
- Gaps (if any) in the new player substitution rules

To achieve exploratory analysis, Excel, RapidMiner and Tableau has been used for cleansing, analyzing and visualizing the data followed by formulating six hypotheses to address:

i. The confidence level of players to attempt goal in International tournaments is most likely to decrease by 50% than that in SSN.
ii. The confidence level of players to attempt goal in International tournament is most likely to decrease by 35% than that in ANZC.
iii. After the split of Australia and New Zealand Championship the performance quality of Australian Diamonds has improved with an increase in average gain around 5 times.
iv. After the split, the number of errors of Australian Diamond has decreased possibly by 20%.
v. Most of the coaches tend to substitute player in the mid of the period in the range of 350-600 seconds with an aim to win game by replacing tired players.
vi. Squad “Thunderbirds” tend to do maximum substitution from Goal attacker to Goal shooter while “WBOP magic” tends to substitute to center forward from attacking and wing defense tactics during their game.

The report suggests not to use SSN as a selection measure for Australian Diamonds. Similarly, substitution of players after quarters have shown significance for winning and top teams are suggested to keep interchanging their players.

These findings can be helpful for Netball Australia to overview their selection benchmark, improve performance and encounter exploitation of substitution rules if any exists.

1. Introduction and Approach

Started as Women's Basketball on Australia in 1897, the acceptance, growing demand, and unavoidable changes has given birth to existing unisex ball sport naming Netball Australia. Organizing the most advanced and professional domestic Competition in the World, Australia has succeeded to be most willing destination for best players over the globe. Before 2017, Australia and New Zealand used to share the domestic league. Australian National Team, the Australian Diamonds, stands out to be the best of world and has victory around 82% of all its International Matches.

To lead in the growing global competition and improve the level of performance, Netball Australia is curious about the following three issues and this report provides insights on:

1. What level of the confidence in performance can Netball Australia have in the domestic competition to indicate the performance of those players at international tournaments?
2. How is the impact on the performance of Australian Diamonds after the split of Australia and New Zealand Championship?
3. What would be the new strategy of the coach to make changes in the court player?

This report mainly aims to analyze and explore data provided by Netball Australia to comprehend its concerns. Statistics provided about matches, players, squads and their performance measures since 2009 is analyzed, cleansed and visualized using Excel, Tableau and RapidMiner. The report further develops hypothesis to address the issues presented based on some assumptions, draw conclusions and provide recommendations based on the findings. Furthermore, a path for future predictions has been provided.

2. Assumptions

The data exploration in this report are subject to the following assumptions made:

1. The year mentioned in the column 'Match_id" stands for the year the tournament was held.
2. Squad represented with the name "Australian", "Australia", "Diamonds" and "Australia Diamonds" stands for the national team of Australia which is "Australian Diamonds" with squad ID "811"
3. Intercepts, penalties, gain, feeds, goal attempts, bad passes and missed goals are Key Performance Indicator of players and teams.
4. Missing values (as "-" in "frompos" and "topos" columns) were removed to increase accuracy of evaluation.

3. Data Analysis

This section provides insight for Netball Australia's 3 concerns. Hypotheses are presented supported by exploratory analysis followed by steps to be undertaken for advanced analytics.

3.1 Exploratory Data Analysis

For the 3concerns presented 2 hypotheses have been developed which has been further evaluated.

3.1.1 Performance confidence of players in Domestic and International Tournaments

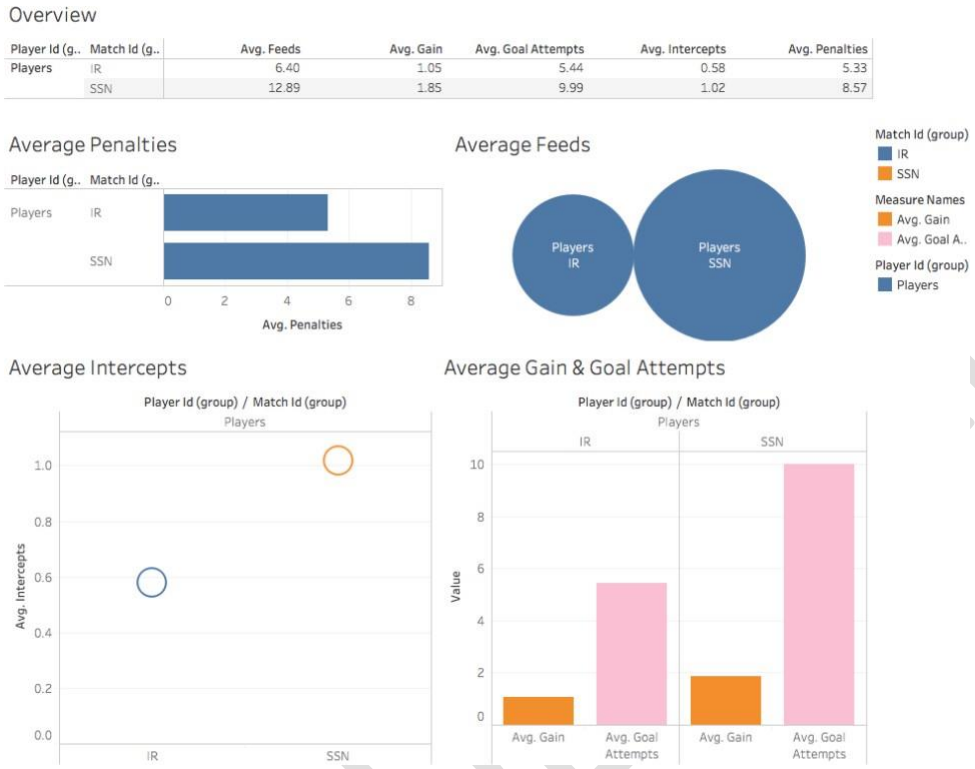


Fig 1: Performance Comparison of Players on SSN and International (IR) Championship

The dashboard demonstrates the performance of 27 squad member of Australian Diamonds who have performed in International Tournament and SSN. On observation, it can be clearly seen that average goal attempt of those players in SSN is 50% more than that in IR. Moreover, average gain, feeds, intercept and penalties also stands to be less in IR.

Hypothesis 1: Based on the exploratory analysis, we can hypothesize that in comparison to SSN, the confidence of players to attempt goal in IR decreases by 50%.

Fig 2: Performance Comparison on ANZC and International (IR) Championship

The above figure displays performance of the 27 squad members of Australian Diamonds in Australia and New Zealand Championship compared to International Championship. The average Feeds of those players stands out to be 8.532 in ANZC while that in IR is only 6.403. Likewise, other key performance Indicator such as average gain, goal attempts, penalties and intercepts also appear to be less in IR.

Hypothesis 2: Corresponding to the above finding we hypothesize that in comparison to ANZC, the confidence of players to attempt goal in IR decreases by 35%.

3.1.2 Impact on the Performance of Australian Diamonds after domestic competition split

Fig 3: Performance Evaluation of Australian Diamonds after split

The above dashboard compares the average Goals, Gains, Penalties, Feeds and Goal Attempts of Australian Diamonds before and after the split of domestic championship. On the observation, it can be clearly seen that all considered performance indicators has been progressed. The average Feeds of Diamonds from 2014 to 2016 was 74.32 which has increased to 78.17 in 2017-18. Similarly, for the same period the average goal attempts has also risen from 63.63 to 64.75. The inclination in the average gain and penalties appears to be impressive with increase of more than 5 times. A small increase can also be observed in average goals.

We can clearly see that the increase in the number of Goal Attempt and Feeds results to the increase in the number of goals.

Hypothesis 3: Average Gain has been increased by 5times after the split of domestic championship of Australia and NewZealand

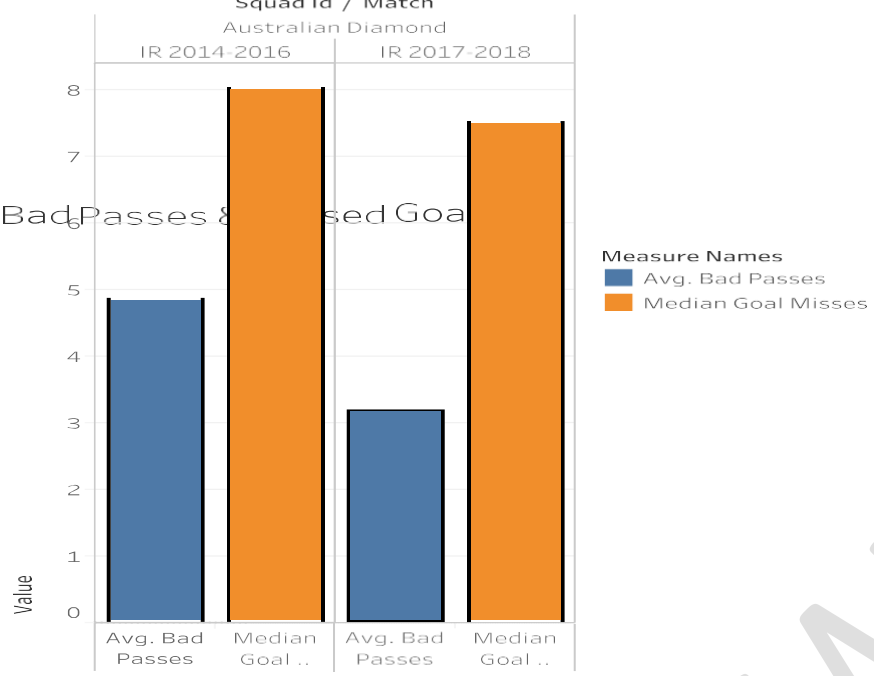


Fig 4: Changes in Number of Errors in the Performance of Australian Diamonds After Split

The diagram above demonstrates the changes in average Bad passes and Missed Goals before and after 2017. The median goal missed has decreased from 8 to 7.5 while average bad passes also declined to 3 from 5.

Hypothesis 4: With Reference to the observation, we have hypothesized that the Number of Errors of Australian Diamonds has decreased by 20% after the split of domestic championship of Australia and NewZealand.

The performance of Australian Diamonds has been improved after split; this may be because of relaxation in restriction in having players from outside the clubs of the competing nation in SSN.

3.1.3 Player's Substitution Trend

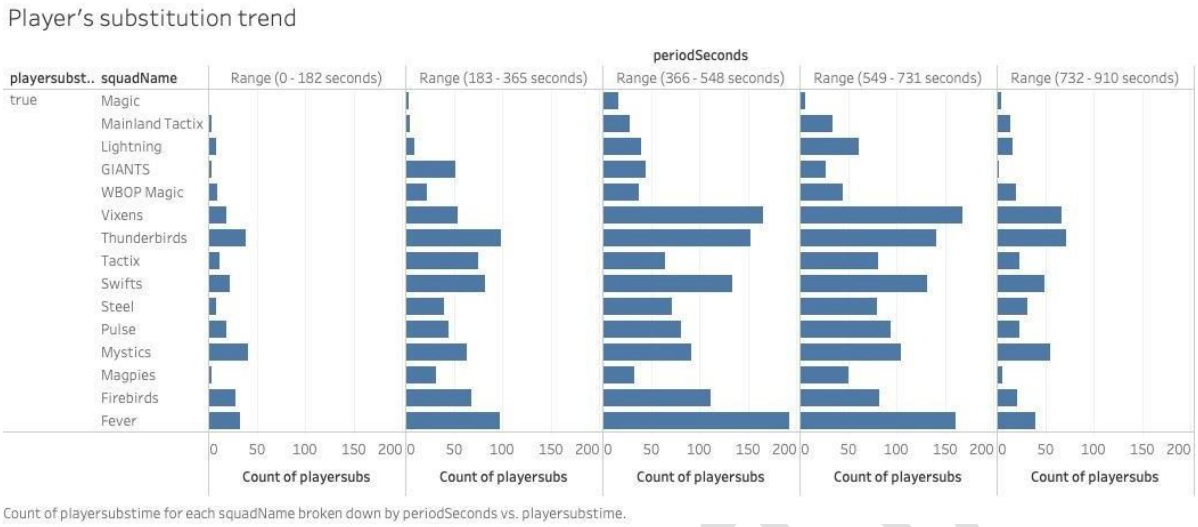


Fig 5: Distribution of Player Substitution according to periods

Above graph shows trend about how coaches are substituting players in different netball competitions. A clear indication of during which period and at what duration coaches are substituting their players to get advantage in game is seen. Different bins have been created for duration of period seconds to get more clear view about during which moment players are getting substituted in the game to show that the squad fever has made highest substitution between 366-548 seconds and lowest in first 182 seconds.

Hypothesis 5: We can hypothesize that most of the coaches are mainly substituting their players in mid of the period in the range of 350-600 seconds.

Main reason for such substitution can be to introduce fresh players in the game so that they can create more chances to win. Also, it can be hypothesized from the above trend is coach of squad "Fever" is mostly exploiting the rule of "interchanging" the players for an injury followed by squad "Vixens". While coach of squad "Magpies" seems to make lowest number of substitutions throughout the match.

<Substitution of players based on game position>

fromPos	Max. squadNa..	GA	WA	GS	C	GK	GD	WD	I	S
GA	Tactix				1					
	Thunderbirds		15				1			
	Vixens								11	
	WBOP Magic			67						125
WA	Pulse							3		
	Tactix		2							
	Thunderbirds	14		2					18	
	Vixens									110
	WBOP Magic				112					
GS	Thunderbirds		3				2			
	Vixens			1						103
	WBOP Magic	76							10	
C	Thunderbirds	1								
	Vixens								15	124
	WBOP Magic		112		3			52		
GK	Thunderbirds			2		1			11	
	Vixens									83
	WBOP Magic						116	9		
GD	Thunderbirds	1					2			
	Vixens									139
	WBOP Magic					81		54	25	
WD	Mystics							3		
	Thunderbirds		6							
	WBOP Magic				86	11	49		17	91
I	Thunderbirds					16				
	Vixens	12	14		15					
	WBOP Magic			9			21	20		
S	Vixens		107	108		102		104		
	WBOP Magic	119			107		128			

toPos: 1 – 139

Sum of Number of Records (color) broken down by toPos vs. fromPos and maximum of squadName.

Fig 6: Distribution of Player Substitution According to Position

Above graphs gives insights about how coaches of netball team are substituting players based on their "position" in game. Here we are analyzing the team's strategy of how they are taking advantage of substitution rule by interchanging the players with changing the position of players in the game. For instance, coaches are substituting players from attacking region and making more player available in defense area for the game.

Hypothesis 6– Based on the analysis it was observe that there are very less teams who are substituting players with the same position they are playing in the game.

From the graph we hypothesize that overall "Thunderbirds" did maximum substitution from Goal attacker to Goal shooter whereas "WBOP magic" doing most of substitution to center forward from attacking and wing defense tactics during their game.

It was also observed that at least one player of each team got substituted from Wing attack or Wing defense to the center part of game to create more chances of high passes and scoring more goals for the team.

3.2 Advanced Analytics

Data cleaning is to be performed to remove missing values presented mainly in "position stats" to understand data in depth followed by creating new attribute called "years" to understand the year the match was played. Moreover, observation of matches based on home and away ground's will be made as team playing in their home ground generally have an advantage over team playing in away ground.

Period seconds can be divided into five bins to help in predicting more outcome about data like player substitution or in which period players are scoring most. Furthermore, different model for prediction of data can be applied like Naive Bayes, simplistic model to estimate team abilities where we going to model the score difference of each score in a game. To model the score differential, we will use regression model with t distribution for analysis. Later the use of bagging and bootstrapping can be done to reduce variance without any noticeable effect of bias, and it will help in running the model faster with good results.

4. Recommendation

According to fig1,2 and Appendix I, the performance of squad member has declined in International tournament thus, considering performance history of player would be better than only focusing on SSN performance. Similarly, the relaxation in restriction in having team members from the clubs competing nations must be continued in SSN as the performance of Australian Diamonds tend to progress after split.

The correct rotation/substitution of players during the game and changing the position of players after every quarter can create more chances of winning the game. For instance, a player playing as goalkeeper in first half must transfer to defense or center in second half.

Finally, teams like Vixens, Giants who are top teams and score a greater number of goals in every game must use the rule of "interchanging" player more often so that they can ensure fitness of their players on right level. This will help teams to keep fitness level up-to data for all players and will assist in identifying new players that can contribute to their team.

5. General Analytics Issue

In the process of cleaning and analyzing data, encounter with ethical and analytical issues related to data is inevitable. Thus, some of the general analytical issues faced during this exploratory analysis are as follows:

- **Data Quality Issues**: In the provided dataset, various number the data were semi-structured and unstructured which hamper the data quality and can also lead to the inaccurate findings or predictions (Wang, Kon and Madnick, 1993).
- **Missing Values Issues:** There were null and missing values in the data which were tackled through the assumption. However, this issue can result to the biasness in predictions and results.
- **Human Error:** Starting from the data extraction continuing with data preparation, data analysis and visualization human work has been involved thus, the prediction cannot be 100% correct.
- **Privacy and Ethical Issues:** As per the signed contract, the disclosure of data has been the main concern. In case of failure to so, would lead to breach of contract. Thus, the data and information must be shared on the bias of privacy norms and be altered to present the unanimous findings (GS1 Australia Privacy Policy - GS1 Australia, 2020).

- **Misinterpretation of data:** The understanding of attributes and elements of data is key otherwise the proper insight and relevant approach may not be possible (Keim, Lee, Thuraisinghaman, and Wittenbrink, 1995). Here assumptions have been made which may lead to biased results.

6. References

Wang, R.Y., Kon, H.B. and Madnick, S.E., 1993, April. Data quality requirements analysis and modeling. In *Proceedings of IEEE 9th International Conference on Data Engineering* (pp. 670-677). IEEE.

Keim, D.A., Lee, J.P., Thuraisinghaman, B. and Wittenbrink, C., 1995, October. Database issues for data visualization: supporting interactive database exploration. In *Workshop on Database Issues for Data Visualization* (pp. 12-25). Springer, Berlin, Heidelberg.

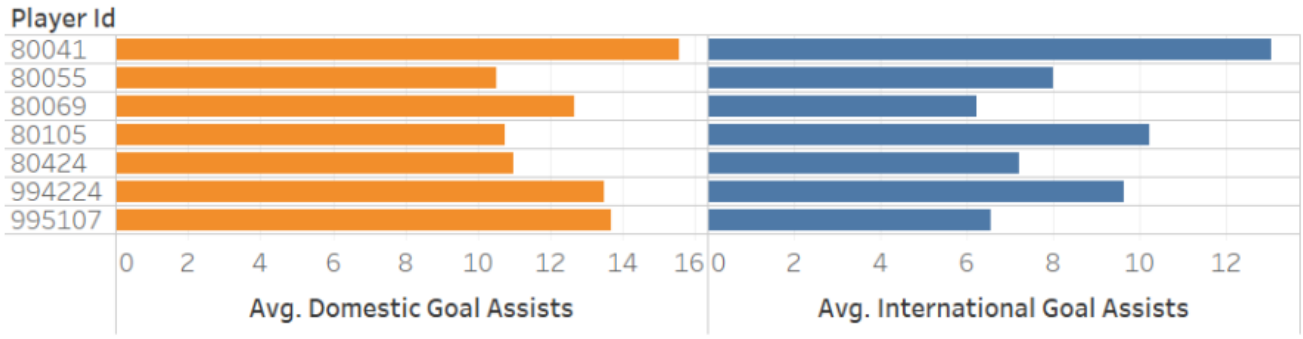
Gs1au.org. 2020. *GS1 Australia Privacy Policy - GS1 Australia*. [online] Available at: <https://www.gs1au.org/gs1-australia-privacy-policy/> [Accessed 12 April 2020].

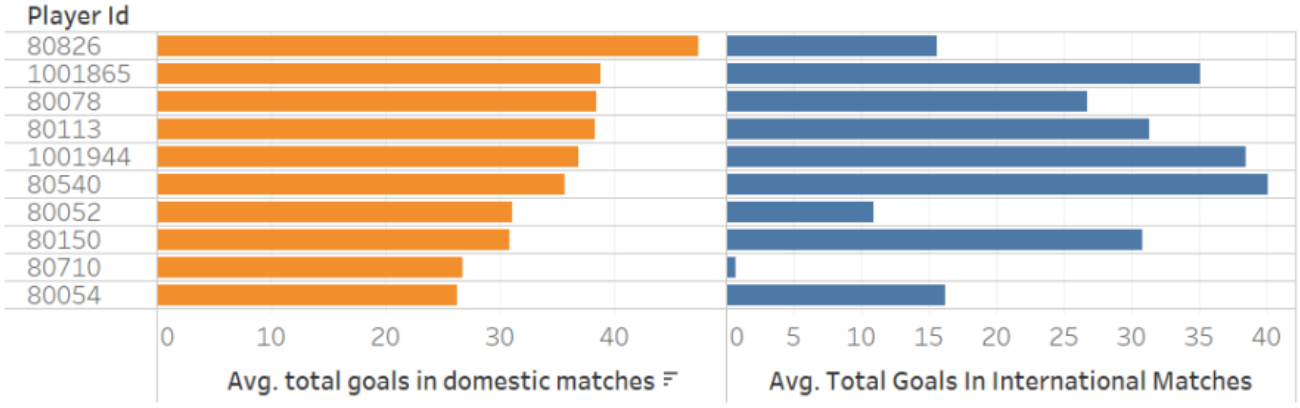
7. Appendix A

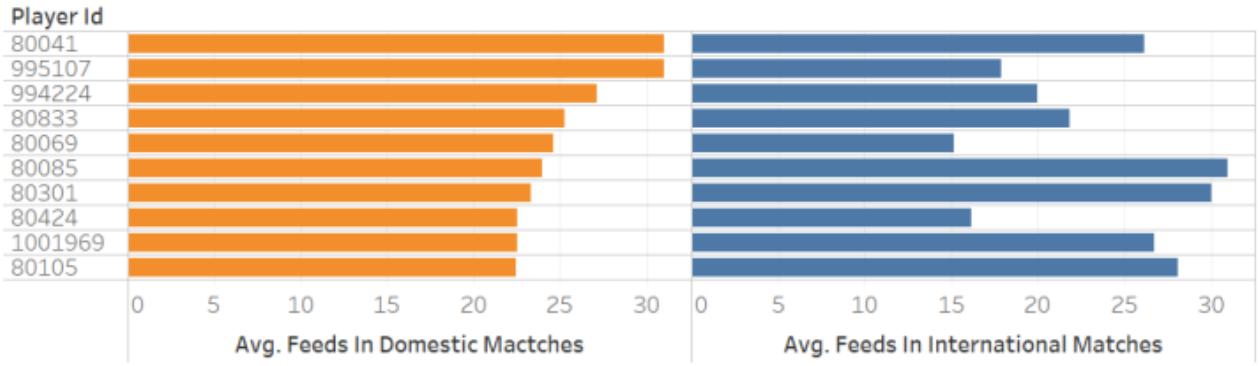
An overview of the players in Domestic and International Match has been presented below. Here selection of players is limited to 10 while in the hypothesis presented above the performance of all the players in Australian Diamond were evaluated with domestic competitions.

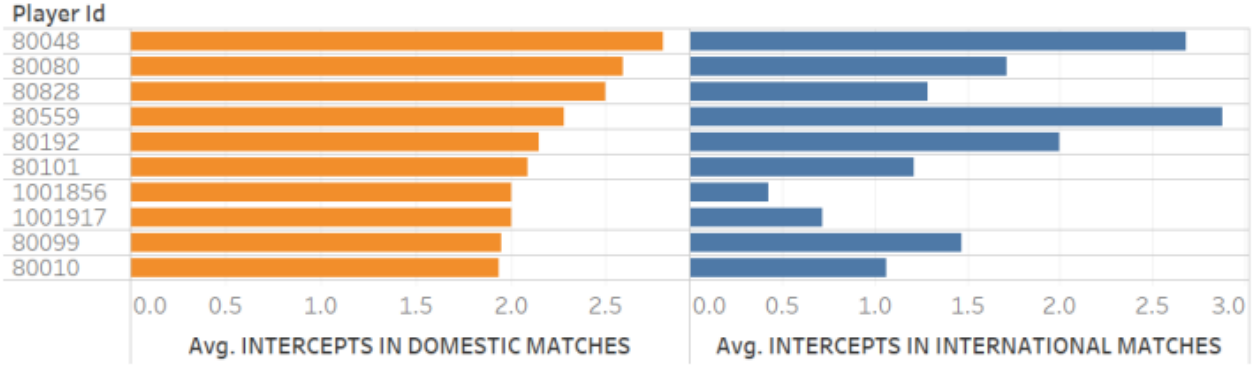
Overall Performance Averages of the top 10 players

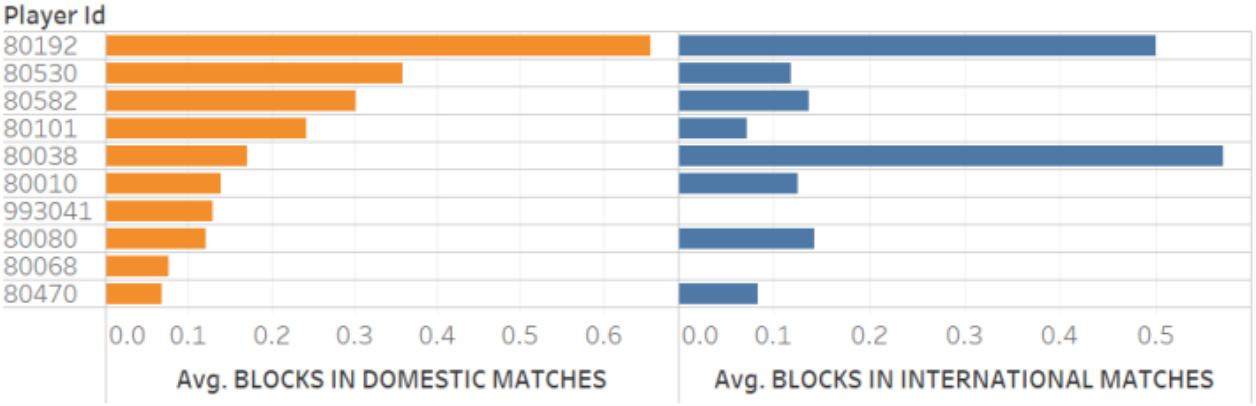
	Domestic	International
Goals Scored per match	35.04	24.59
Goal Assists per match	4.5	3.4
Feeds per match	25.4	23.2
Intercepts per match	2.2	1.5
Blocks per match	0.22	0.17






Player Id
80041
995107
994224
80833
80069
80085
80301
80424
1001969
80105
0 5 10 15 20 25 30
Avg. Feeds In Domestic Mactches
0 5 10 15 20 25 30
Avg. Feeds In International Matches


Player Id
80048
80080
80828
80559
80192
80101
1001856
1001917
80099
80010
0.0 0.5 1.0 1.5 2.0 2.5
Avg. INTERCEPTS IN DOMESTIC MATCHES
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Avg. INTERCEPTS IN INTERNATIONAL MATCHES


Player Id
80192
80530
80582
80101
80038
80010
993041
80080
80068
80470
0.0 0.1 0.2 0.3 0.4 0.5 0.6
Avg. BLOCKS IN DOMESTIC MATCHES
0.0 0.1 0.2 0.3 0.4 0.5
Avg. BLOCKS IN INTERNATIONAL MATCHES