STM5IPL Final Report

Evaluating key performance metrics into Sports Analytics in AFL using Power BI Dashboards

Michael Le 21689299

November 2024

Abstract

AFL, commonly known as Australian Rules Football, where the sport was drafted in 1859 by Hammersley, Smith, Thompson, and Wills which has the sport has been famous across Australian culture for over 140 years, and the rules were established during the late 1880s. The first clubs were formed in 1860 and 1866 in South Australia and Queensland, respectively; later, they continued to expand in Victoria in 1877, forming eight-team leagues alongside South Australia, forming VFA/VFL in 1897 but was renamed to AFL in 1990, expanded more teams up to AFL 2011-2012 Season where two new teams GWS Giants and Gold Coast Suns were added into the AFL official roaster. Understanding sports analytics has been a main goal, as well as building relationships between teams and players and applying data analysis. Sports analytics collects, analyzes, and presents information frequently used in sporting events, teams, and athletics. At its core, when determining relevant indicators on players applied, the selection process requires scouting reports, and player assessments are required to enhance decision-making to resolve recruiting and drafting procedures. It requires understanding the relationships between big data and sports domain knowledge as sporting competitions evolve. To achieve this, our data team applied analytics to study player performance that could overall impact the team, to identify their strengths and weakness based on our data from 2022-2024 drafts from the Foundry Athletic Data consisting of Player and Team Data for males and females for the Sandringham Dragons Football Club. For instance, key performance metrics, such as physical fitness, intensive exercising, and various match results, can measure an athlete's progress over time during the 2023-2024 season. Sports analytics in AFL aimed to transform raw data into actionable insights up to the coach's standards. Hence, it will be valuable for the player and team's success and satisfaction with drafting outcomes. To ensure all players can produce the best data-driven results, they are guaranteed to stand out amongst other competitors to secure a player position recruited from potential AFL coaches.

Introduction

The purpose of applying Sports Analytics to Australian Rules Football is to determine based on "physical skills," "past performance in matches," "athletic abilities," and "psychological assessments." Evaluating applied statistics to include such as "speed," "growth," "endurance," and "strength" that are "quantifiable" for Game Analysis, crafting "Game Analysis," "Scouting," and "Training Performance." In forms of median-sized and complex datasets in big data analytics and convey the data into valuable insights for decision making by accessing real-time data on player and team performance improvements and "risk of injuries". To decide whether the player can cooperate under pressure on and off the field under time constraints. The data was collected by having athletes from lead coaches wear "devices and sensors" that tracks various performance metrics, gathering on and off field GPS data and monitoring an athlete's data. The data was uploaded and been automated into Microsoft Excel Spreadsheets through email to discuss data integrity and prevent data leakage from revealing sensitive information to be spread to other competitors and drawing inspiration from the "McKinsey Global Institute" to understand the capabilities of "traditional database software tools" for applying real-world sports data. It includes five features: Value, velocity, volume, veracity, and variety**, which** play a big role in big data in exercise performance, health data, training statistics, and analysis**.** Volume is the size and amount of big data organizations manage and analyze. Value comes from insights into data and patterns that lead to effective operations and building relationships to meet business expectations. Variety is the diversity and range of data types, including unstructured, semi-structured, and raw data. Velocity is the speed efficiency of how often the data is collected at a certain timeframe. Veracity is the truth or accuracy of data and information assets, determining how confident the data is well-represented. In addition, when analyzing big data containing player and team metrics, we must consider major factors such as fairness, accuracy, and reducing bias by understanding that the data analytics lifecycle consists of six phases "based on the CRISP-DM methodology" **(Figure 1).** As we analyze deeper into the 'applications and impacts of sports analytics and shift in how sports performance is approached and enhanced' that is manually 'been crafted' up to 'latest trends' in recent years. Keeping up to date with 'cutting-edge technologies, and innovative technologies shaping sports data analysis'. Defining the key requirements, technical tools, and questions needed to prepare for each phase during the Life Cycle of a Data Analysis Project from Northwestern University starting from "data preparation, exploratory analysis, validation, visualization, and presentation." For this purpose, performing a data analysis report is to transform raw data for player and team metrics into multiple dashboards to display valuable insights for key performance indicators and injury metrics on all players that present the overall data architecture and representation that is shared among data analysts, to draft potential players and help enable coaches at Foundry Athletic organization at the Sandringham Football Club to filter and enable data for concise decision making.

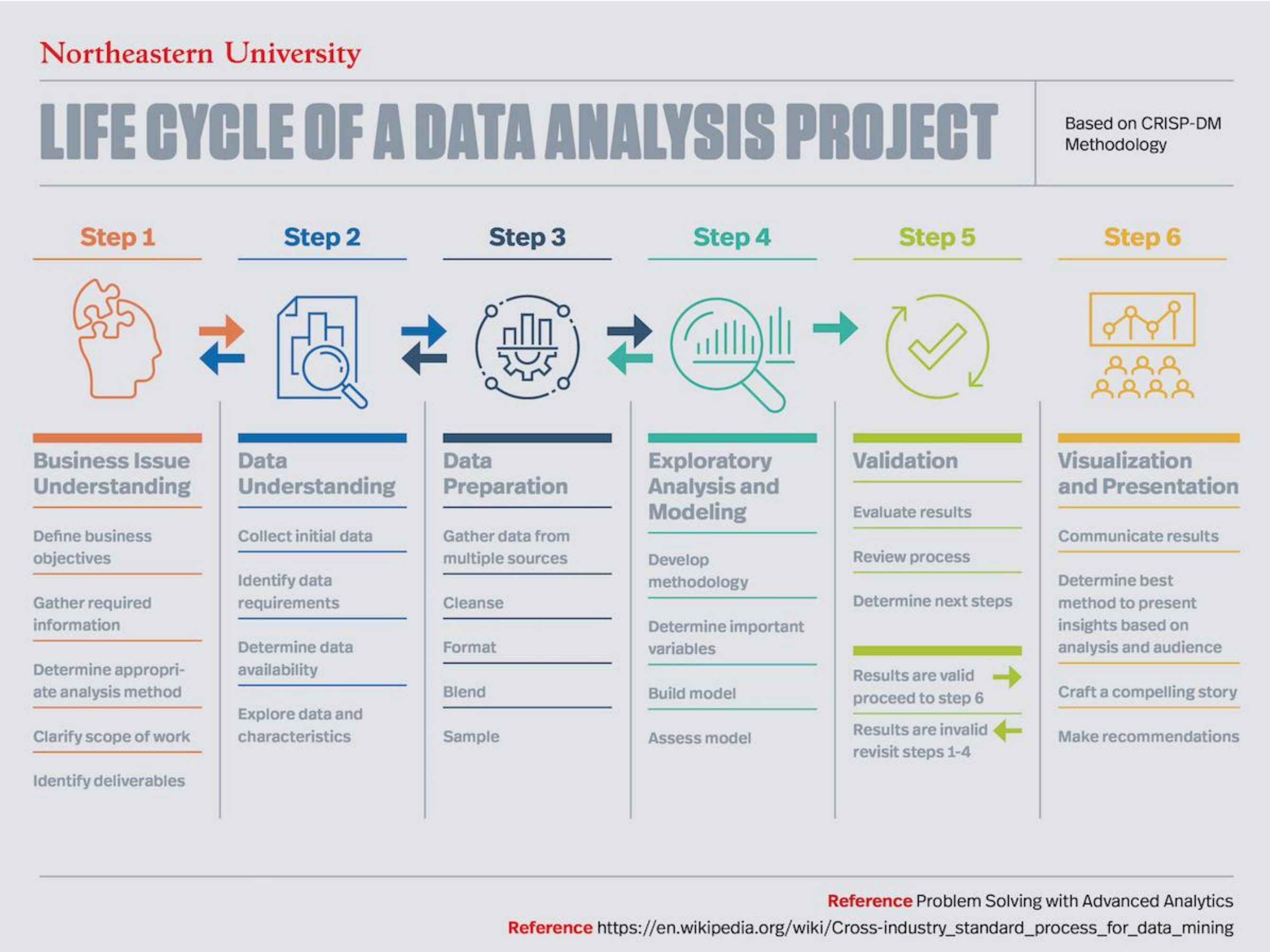


Figure 1. Life Cycle of a Data Analysis Project from Northeastern University, we will often refer to this process during the methodology section

Methodology

During the "Business Issue Understanding" phase for the data analysis project**,** I performed an explanatory analysis of the data collected on recent Player Performance, Statistics, and Sandringham games. To understand key performance indicators and injury prevention metrics when analyzing the data and applied statistics. The goal is to determine ongoing potential athlete talent and reduce injuries; this can be achieved by selecting only potential candidates securing player positions at the Australian Football League selected from recruiting coaches given the scope of the work from the Lead Football Coach. To achieve this, we must consider the business objective, information on the data that is required, a suitable analysis method, and identify our deliverables. In this case, the data collection contains folders containing Player and Team Data **(Figure**

2) contains 183,580 kilo-bytes which is roughly 0.2 giba-bytes in total for all the Microsoft Excel Spreadsheets during the placement tackling how we 'handle temporary data' while performing our data analysis. After further investigation by the data team and supervisor during the first meeting, we first needed some clarification before proceeding to the "Data Understanding" phase during the project scope. After further discussing the business proposal and direction, the lead coach insists on taking further caution and considering how well we manage our data during the analysis to explore some of the valuable data and characteristic types of player and team insights. During the "Data Understanding" phase, the following seven folders **(in Figure 3)** maintain all valuable information for player and team performance. For **2km Results** consisting of player names, 2km average times taken during 2022 Pre-season to 2024, Maximal Aerobic Speed, and time improvements taken from different periods. **GPS Data stands for Global Positioning System** and consists of Player data from the 2023-2024 Season, where features such as TOG%, Meterage per Minute, Total Distance in meters, HSR m/min (4m/s+), Sprint Distance, Sprint m/min (6m/s+), Maximum Velocity, Player Load, Total Acceleration Load, Distance (mts) to Vmax, Time (sec) to Vmax, 90% Vmax, D(90%), T(90%), Vmax convert to (km/h). Vmax converts to (km/h) 90% of the Maximum Velocity. **Gym Strength Testing Data** contains numerical and continuous variables NORDBOARD-ISO 30, known as the long-strength isometric strength test, are metrics that measure hamstring strength and legs to identify weakness and imbalances that can prevent injuries where the following data features consist of NORD LEFT, NORD RIGHT, and NORD ASSYM, IMPT consists Peak Vertical Force measured in Newtons and their respective accelerations in meters per second squared to demonstrate the maximum force exerted onto the perpendicular to the surface during stance phase. Features jump Height (Impulse-Momentum) in centimetres, Contract Time in seconds, and Speed in meters per second. SJ contains only the Jump Height Flight Time in centimetres, CMJ has Jump Height (Flight Time) in centimetres and Concentric Peak Force [N], and for the DASI, the ratio between CMJ Peak Force and IMPT Peak Force. In addition, the categorical (binary) and ordinal variable, BOTTOM AGE/TOP AGE (B, T), represents top and bottom agers. TYPE (Neural, Strong, and Powerful) measures the level of strength respectively, including additional strength data such as the Date that was recorded, the number of REPS and RMs, 'to customize training regimes that aligns with each athlete's physical capabilities and recovery needs'. **Height & Weight Data** The following numerical and continuous variables contain heights measured in meters and weights in kilograms for each of the players under 16 and 18 years old. 7 Day%, 14 Day%, and 28 Day% Contributions, HGTS (Old and New), WGTS (Old and New). Alongside some categorical variables and sensitive data features such as Photo URL, Date of birth, Status whether players are still in the program or not, Squad (U16, U18, Free Agent), Heights and Weights data from 2022-2024 period including their height and weight growths over the years, **Max Speed Data** which is already combined into the GPS Data similarly for the **Player Football Match Statistics.** Lastly, the **combined**

testing data contains key metrics for player and team matches during the 2022-2024 season. For instance, for Under 16 Boys and Girls, combined testing data, top results, and National % rankings data were collected during October 2023, including other vital metrics obtained from State Academy, Top 10s, Country, Metro, and U16 (ALL) for keeping track amongst other football clubs completing secure AFL positions across all the Victorian state alone.

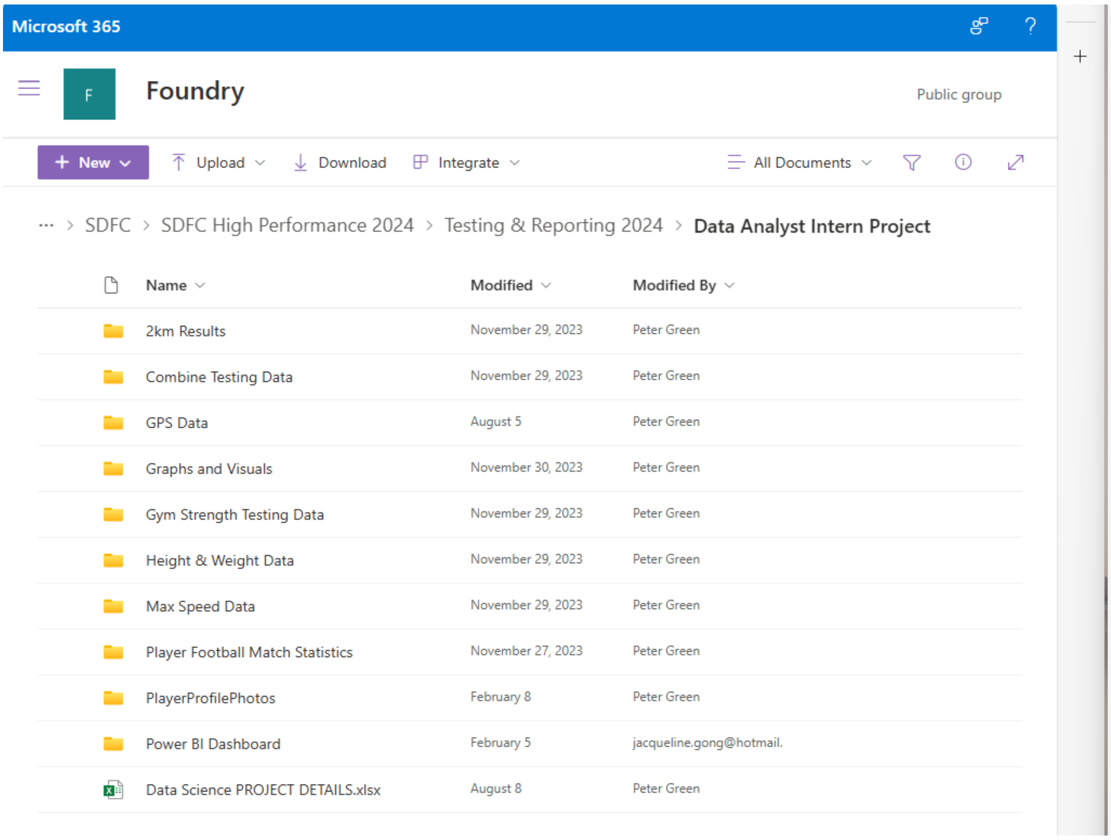


Figure 2. Lead Coach Folders at Foundry Athletic containing all the available data collected containing Player and Team Performance data from the Sandringham Football Club.

For the "Data Preparation" phase, we wanted to keep track of any relevant player and team records that are best suited for analysis before proceeding to the next stage of the data analysis project. The main goal for this phase was to analyze different rows and columns for each of the files contained in each of the folders to determine if duplicates, empty values, or misspellings were involved. The data team uses some of the technical skills to perform these tasks to transform and clean raw data by "standardizing data formats, enriching source data, or removing outliers." This can be done by transforming and cleaning data that contains duplicates, empty or Null values. Requires at least one of the programming languages, such as Python, to recognize and understand patterns of data using appropriate libraries using NumPy and Pandas to perform data cleansing and transforming onto Microsoft Excel Spreadsheets **in Figures 3,4,5 and 6** to maintain the five characteristics V's (Value, Velocity, Volume, Veracity, and Variety) of big data dealing with traditional database technology in Microsoft Excel procedures see **Appendixes C, D and E shows only for the 2km Results, Gym Strength Data and Heights and Weights Data** respectively where the GPS **Data folder** is the only folder that contains clean data that has no duplicates**,** empty or null values alongside with **Max Speed Data** and **Player Football Match Statistics**. To 'encompassing, structured, unstructured and semi-structured data' that has some features or values maintain and understanding data relationships for every player who participated during the program.

	A	B	C	D	E	F	G	H
1	Name	2KM TIME - 19/12/22	Min&Secs	Seconds	GROUP (Player Mas Groups)	MAS SCORE (m/s) (Player Mas Groups)	TARGET TIME (5x 80m (110% MAS) (20s rests)) 14_02_2024	TARGET TIME (5x 40m (120% MAS) (15s rests)) 14_02_2024
2	Aaron Taylor		0	0	6'30" - 6'45"	4.95049505	14.69090909	6.733333333
3	Adrian Cole		0	0	<6'15"	5.405405405	13.45454545	6.166666667
4	Alfie Lambert	06:22:00	622	382	6'15" - 6'30"	5.263157895	13.81818182	6.333333333
5	Angus Bowd		0	0	6'15" - 6'30"	5.128205128	14.18181818	6.5
6	Angus Phillips		0	0	6'30" - 6'45"	4.962779156	14.65454545	6.716666667
7	Angus Shepherdson		0	0	Unassigned	0	0	0
8	Angus Taylor	08:27:00	827	507	>7'10"	4.184100418	17.38181818	7.966666667
9	Archer Grant	06:55:00	655	415	6'15" - 6'30"	5.221932115	13.92727273	6.383333333
10	Archie Edwards	06:12:00	612	372	<6'15"	5.333333333	13.63636364	6.25
11	Archie Ludowyke		0	0	6'45" - 7'10"	4.784688995	15.2	6.966666667
12	Aris Moustakis		0	0	6'15" - 6'30"	5.167958656	14.07272727	6.45
13	Athlete 88					0	0	0
14	Athlete 88 (AVERAGES)							
15	Bailey McKenzie		0	0	6'15" - 6'30"	5.221932115	13.92727273	6.383333333
16	Benjamin Seers		0	0	<6'15"	5.586592179	13.01818182	5.966666667
17	Bennett Martin	06:45:00	645	405	6'15" - 6'30"	5.249343832	13.85454545	6.35
18	Brodie Findlay		0	0	>7'10"	4.310344828	16.87272727	7.733333333
19	Charlie Beaumont	06:28:00	628	388	<6'15"	5.39083558	13.49090909	6.183333333
20	Charlie Rozenes		0	0	6'15" - 6'30"	5.249343832	13.85454545	6.35
21	Chris Kellaway		0	0	<6'15"	5.376344086	13.52727273	6.2
22	Darby Heelis	06:51:00	651	411	Unassigned	0		
23	Dexter Prime		0	0	6'30" - 6'45"	5.063291139	14.36363636	6.583333333
24	Eddy Cooper	07:06:00	706	426	Unassigned	0		
25	Eiden Pitt	07:45:00	745	465	>7'10"	4.494382022	16.18181818	7.416666667
26	Emmanuel Ganas	07:17:00	717	437	6'45" - 7'10"	4.739336493	15.34545455	7.033333333
27	Ethan Lakman	07:14:00	714	434	Unassigned	0		
28	Fletcher Teelow		0	0	6'30" - 6'45"	5	14.54545455	6.666666667
29	Flynn Groves		0	0	6'45" - 7'10"	4.830917874	15.05454545	6.9
30	Frederick Brayshaw		0	0	6'15" - 6'30"	5.208333333	13.96363636	6.4
31	Grand Total							
32	Harrison Oliver		0	0	<6'15"	5.420054201	13.41818182	6.15
33	Harry Armstrong		0	0	6'30" - 6'45"	4.926108374	14.76363636	6.766666667
34	Harvey Allan		0	0	6'30" - 6'45"	5.050505051	14.4	6.6
35	Hudson Wright		0	0	>7'10"	0	0	0
36	Hunter Holmes	07:05:00	705	425	Unassigned	0		
37	Hunter Lynch		0	0	6'15" - 6'30"	5.235602094	13.89090909	6.366666667
38	Jack Cheep		0	0	6'30" - 6'45"	4.975124378	14.61818182	6.7
39	Jack Dalton		0	0	<6'15"	5.698005698	12.76363636	5.85
40	Jack Hayter		0	0	<6'15"	5.376344086	13.52727273	6.2
41	Jack Kerin		0	0	Unassigned	0	0	0
42	Jack Meredith		0	0	6'45" - 7'10"	4.87804878	14.90909091	6.833333333
43	Jack Walker	06:29:00	629	389	Unassigned	0		
44	Jake Matthews		0	0	6'45" - 7'10"	4.739336493	15.34545455	7.033333333
45	Jake Mehl	07:12:00	712	432	>7'10"	4.683840749	15.52727273	7.116666667
46	James Arnold	06:37:00	637	397	6'15" - 6'30"	5.249343832	13.85454545	6.35
47	James Cutler	06:15:00	615	375	<6'15"	5.494505495	13.23636364	6.066666667

Figure 3. Sample of Finalized 2km Results after transformation Microsoft Excel Spreadsheet

NOTE: Not all rows and columns are displayed in figures or appendixes)

Rou	Na	D	Total Durati	Total Distance (	Meterage Per Min	HSR Distance (4m/s+)	HSR m/min (4m/	Sprint Distance (6m/s+) (	Sprint m/min (6m/	Maximum Velocity (m	TO
14	Adrian Cole	29/07/2023	1:42:33	10866.88	106.0	2538	24.3	397	3.9	7.5	98
7	Archer May	20/05/2023	1:19:56	8409.61	105.2	1514	15.0	103	1.3	7.2	79
13	Archer May	8/07/2023	1:35:12	9738.29	102.3	1378	13.7	152	1.6	8.3	95
14	Archer May	29/07/2023	1:19:52	8323.14	104.2	1800	17.2	277	3.5	7.6	77
16	Archer May	19/08/2023	1:24:13	9320.53	111.0	1463	14.6	143	1.7	8.8	84
17	Archer May	26/08/2023	1:23:48	8970.38	107.0	1492	14.9	186	2.2	7.6	84
WC	Archer May	9/02/2023	1:18:52	7922.90	100.4	1521	15.1	334	4.2	7.3	78
QF	Archer May	9/10/2023	1:24:23	8612.60	102.0	2008	20.0	340	4.0	7.9	84
PF	Archer May	17/09/2023	1:28:14	7955.11	90.2	1183	11.8	180	2.0	7.5	88
GF	Archer May	24/09/2023	1:17:37	6541.80	84.3	1250	12.4	178	2.3	7.8	77
2	Archie Roberts	1/04/2023	1:27:15	10467.24	120.0	3417	34.1	707	8.1	8.0	87
3	Archie Roberts	7/04/2023	1:22:09	10161.41	123.7	3226	32.0	698	8.5	8.7	81
16	Archie Roberts	19/08/2023	1:28:20	10059.65	113.9	2950	29.4	525	5.9	8.3	88
17	Archie Roberts	26/08/2023	1:26:01	10062.74	117.0	3010	30.0	629	7.3	8.6	86
WC	Archie Roberts	9/02/2023	1:17:49	9568.14	123.0	3376	33.5	745	9.6	8.6	77
QF	Archie Roberts	9/10/2023	1:21:29	9664.51	118.6	3198	31.8	781	9.6	8.2	81
PF	Archie Roberts	17/09/2023	1:23:12	9664.76	116.2	2640	26.4	635	7.6	8.2	83
GF	Archie Roberts	24/09/2023	1:23:15	9644.20	115.8	2814	28.0	556	6.7	8.7	83
1	Billy McGee Galimberti	25/03/2023	1:14:35	8328.55	111.6	2412	24.1	435	5.8	8.1	74
2	Billy McGee Galimberti	1/04/2023	1:13:27	7593.28	103.4	1879	18.8	367	5.0	8.6	73
3	Billy McGee Galimberti	7/04/2023	1:18:07	8714.48	111.5	2544	25.2	611	7.8	8.5	77
4	Billy McGee Galimberti	15/04/2023	1:17:04	7974.86	103.5	1818	18.0	230	3.0	8.9	76
7	Billy McGee Galimberti	20/05/2023	1:20:44	10284.01	127.4	3831	38.1	794	9.8	8.2	80
16	Billy McGee Galimberti	19/08/2023	1:26:33	9710.54	112.2	2492	24.8	383	4.4	8.3	86
17	Billy McGee Galimberti	26/08/2023	1:19:02	8766.94	110.9	2441	24.3	363	4.6	9.0	79
WC	Billy McGee Galimberti	9/02/2023	1:20:01	9104.33	113.8	2237	22.2	263	3.3	7.8	79
QF	Billy McGee Galimberti	9/10/2023	1:22:23	8519.16	104.1	1255	12.5	198	2.4	8.6	82
PF	Billy McGee Galimberti	17/09/2023	1:20:46	8903.29	110.2	1868	18.7	173	2.1	7.8	81
GF	Billy McGee Galimberti	24/09/2023	1:19:59	7597.19	95.0	1700	16.9	588	7.4	8.8	80
1	Calsher Dear	25/03/2023	1:24:01	7549.98	89.9	1957	19.5	277	3.3	7.8	84
2	Calsher Dear	1/04/2023	1:14:01	8281.19	111.9	2645	26.4	399	5.4	8.0	74
4	Calsher Dear	15/04/2023	1:16:14	7406.84	97.1	1570	15.5	112	1.5	7.3	75
9	Calsher Dear	3/06/2023	1:40:09	7922.55	79.3	1346	13.3	265	2.6	8.1	99
13	Calsher Dear	8/07/2023	1:19:44	8308.02	104.2	2349	23.4	413	5.2	8.0	79
14	Calsher Dear	29/07/2023	1:20:05	7416.64	92.6	1581	15.2	186	2.3	7.3	77
16	Calsher Dear	19/08/2023	1:18:57	7366.18	93.3	1447	14.4	150	1.9	7.7	79
17	Calsher Dear	26/08/2023	1:17:04	7611.35	98.7	1717	17.1	200	2.6	8.4	77
WC	Calsher Dear	9/02/2023	1:14:20	7426.90	100.0	1488	14.8	196	2.6	7.5	74
QF	Calsher Dear	9/10/2023	1:19:34	7703.85	96.8	1020	10.1	122	1.5	7.7	79
PF	Calsher Dear	17/09/2023	1:23:03	7723.06	93.0	1544	15.4	236	2.8	7.8	83
GF	Calsher Dear	24/09/2023	1:12:44	8626.99	118.6	2674	26.6	303	4.2	7.6	72
7	Cameron Saultry	20/05/2023	1:19:08	10121.33	127.9	3549	35.3	429	5.4	8.6	79
14	Charlie Bruce	29/07/2023	1:24:42	9807.56	115.8	2387	22.9	254	3.0	8.4	81
2	Charlie Edwards	1/04/2023	1:15:03	7358.88	98.0	1435	14.3	218	2.9	7.6	75
3	Charlie Edwards	7/04/2023	1:14:23	8483.44	114.1	2181	21.6	454	6.1	8.1	74
4	Charlie Edwards	15/04/2023	1:19:08	8838.79	111.7	1941	19.2	200	2.5	8.2	78

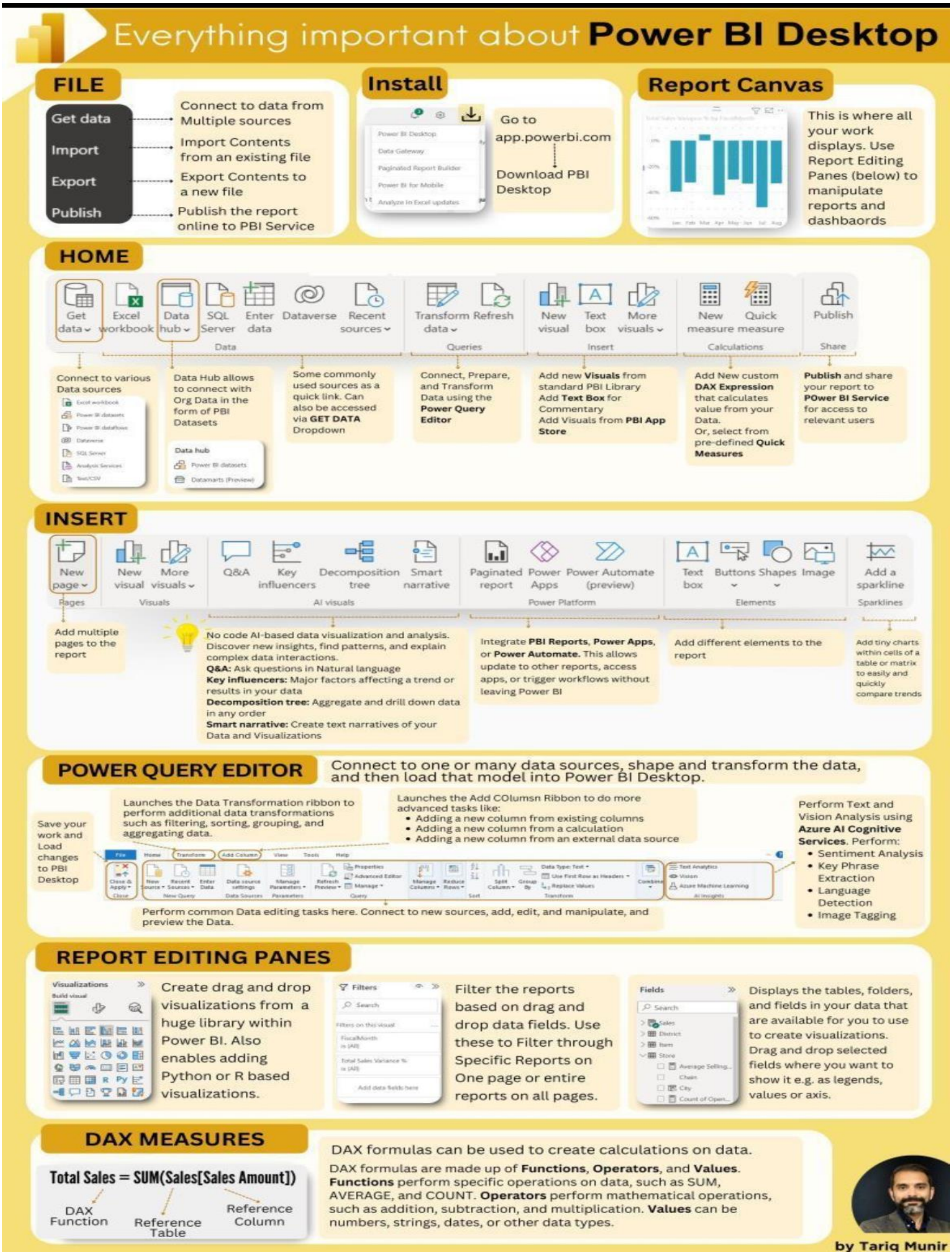
Figure 4. Sample of Finalized GPS 2023 Data for Sandringham Football Club for the Mans Team

Name	Date	Total Duration	Total D	Metera	HSR Di	HSR m	Sprint I	Sprint	Maxim	TOG%	Total P	Total A	Mod?	Max Speed (km/hr)
Aaron Taylor	16/02/2024	1:10:07	3935	56.1	361	2.9	111	1.6	7.6	56	396	945	Full	27.3
Adrian Cole	16/02/2024	1:09:44	3560	51.0	368	2.9	35	0.5	6.5	55	322	828	Full	23.4
Adrian Cole	21/02/2024	1:03:40	6303	99.0	1534	22.0	381	6.0	8.9	91	593	1319	MODIFIED	32.1
Adrian Cole	23/03/2024	1:18:03	8745	112.0	2312	22.1	534	6.8	7.8	75	800	1903	Full	28.2
Adrian Cole	30/03/2024	1:13:20	8249	112.5	1975	19.7	411	5.6	8.0	73	738	1705	Full	28.7
Angus Phillips	1/03/2024	1:39:33	9291	95.2	1909	16.7	157	1.6	7.4	87	1049	2284	Full	26.7
Archer May	16/02/2024	1:03:36	4083	64.2	229	1.8	17	0.3	6.4	50	331	1058	Full	23.0
Archie Ludowyke	19/02/2024	1:00:02	3557	61.7	260	4.3	0	0.0	5.9	100	425	1188	Full	21.3
Archie Ludowyke	1/03/2024	1:21:55	8328	101.7	1584	17.1	305	3.7	7.9	88	660	1766	Full	28.4
Bailey McKenzie	1/03/2024	1:21:45	8870	108.5	2182	23.5	346	4.2	7.5	88	797	2072	Full	27.1
Bailey McKenzie	4/03/2024	1:15:25	4919	65.2	577	7.6	0	0.0	5.9	100	491	1410	Full	21.4
Bailey McKenzie	15/03/2024	1:12:22	8529	117.8	3093	30.6	568	7.8	7.5	72	839	2108	Full	26.9
Bailey McKenzie	23/03/2024	1:17:00	7331	95.2	1572	15.1	267	3.5	7.4	74	812	2307	Full	26.5
Bailey McKenzie	30/03/2024	1:16:42	8593	112.0	2730	27.3	413	5.4	7.6	77	803	2182	Full	27.3
Benjamin Seers	29/01/2024	0:36:06	2718	75.3	418	11.6	15	0.4	6.2	100	249	771	Full	22.1
Benjamin Seers	31/01/2024	1:59:34	8656	72.4	1679	14.0	198	1.7	7.5	100	826	2159	Full	27.0
Benjamin Seers	2/02/2024	0:45:41	3432	75.1	644	14.1	187	4.1	8.4	100	346	923	Full	30.4
Benjamin Seers	5/02/2024	1:20:36	7039	87.6	2045	23.2	574	7.1	8.3	92	688	1789	Full	30.0
Benjamin Seers	7/02/2024	2:03:06	9277	75.6	1857	13.8	548	4.4	7.1	91	930	2656	Full	25.6
Benjamin Seers	9/02/2024	0:56:04	3341	59.6	398	6.6	40	0.7	6.7	94	311	922	Full	24.2
Benjamin Seers	12/02/2024	1:35:21	6392	67.4	1705	16.5	367	3.9	8.4	92	599	1398	Full	30.4
Benjamin Seers	19/02/2024	0:55:36	5348	96.2	1105	17.6	195	3.5	7.5	89	523	1304	Full	27.2
Benjamin Seers	26/02/2024	0:46:09	4074	88.3	770	16.7	67	1.4	7.1	100	358	885	Full	25.6
Benjamin Seers	1/03/2024	0:56:44	7345	129.5	2894	37.8	651	11.5	7.6	74	746	1659	MODIFIED	27.4
Benjamin Seers	6/03/2024	0:32:47	2061	62.9	590	14.2	62	1.9	6.6	79	222	614	MODIFIED	23.8
Benjamin Seers	15/03/2024	1:18:52	9367	118.7	2855	28.3	437	5.5	7.8	78	1044	2598	Full	28.1
Benjamin Seers	23/03/2024	1:27:22	10691	122.4	3776	36.2	647	7.4	7.7	84	1056	2468	Full	27.5
Benjamin Seers	27/03/2024	0:49:52	4571	91.7	1272	23.5	526	10.5	8.4	92	398	1077	MODIFIED	30.4
Benjamin Seers	30/03/2024	1:15:53	10217	134.6	4104	41.0	1065	14.0	8.4	76	985	2102	Full	30.4
Brodie Findlay	21/02/2024	1:40:06	9989	99.8	1046	10.5	44	0.4	6.7	100	887	1910	Full	24.2
Brodie Findlay	15/03/2024	1:14:18	7937	106.8	1729	17.1	50	0.7	6.9	74	793	1625	Full	24.8
Brodie Findlay	23/03/2024	1:20:44	8409	104.2	1184	11.3	58	0.7	6.6	77	865	2041	Full	23.9
Brodie Findlay	30/03/2024	1:12:09	8109	112.4	1319	13.2	53	0.7	6.5	72	805	1611	Full	23.4
Charlie Bruce	19/02/2024	1:14:59	5463	72.9	680	9.1	36	0.5	6.5	100	583	1522	Full	23.5

Name	7 Day%	14 Day%	28 Day%	DOB (MM-DD-YYYY)	HGTS_Old	WGTS_Old	Photo URL	Photo	School Details	Col4	HGTS_New	PRE-SEASON WEIGI
Aaron Taylor	57	57	46	3/25/2006	182.5	77.2	Not Applicable	No Picture	St Bede's College	177	182.5	77.2
Adrian Cole	0	0	7	1/05/2006	196.5	95.65	Not Applicable	No Picture	Uni - TBC	196.5	196.5	95.7
Albert Macgowan				7/27/2008	184	69.8	Not Applicable	No Picture	Not Applicable	189.5		
Alexander Carni				3/26/2008	181	68.7	Not Applicable	No Picture	Not Applicable	175		
Alexander Dreaver				5/24/2008	193.5	88.8	Not Applicable	No Picture	Not Applicable	196.5		
Alfie Lambert	43	71	43	11/02/2007	178	68.55	https://lh3.googleusercontent.com/MyXnSu749xd9KiH51PKkkZffTgAp35hMcml6X9KHpil19b7	Picture	Beaumaris Secondary College	190	178	0
Angus Bowd	0	36	68	4/14/2006	175	76.9	Not Applicable	No Picture	St Bede's College	175	175	76.9
Angus Phillips	29	29	18	3/18/2006	190	82.5	Not Applicable	No Picture	St Kevin's College	196.5	190	82.5
Angus Shepherdson				1/15/2007	197	80	https://lh3.googleusercontent.com/9ZqJ6YO6vYckJZdFc8tK5FYEsg0CKnHYDV5bbwlllEIN6I	Picture	Not Applicable	0	0	0
Angus Skinner				2/27/2008	193	69.9	Not Applicable	No Picture	Not Applicable			
Angus Taylor				6/14/2007			https://lh3.googleusercontent.com/fyuaWS0M-J5MPlNgOVNKrtHaFHnJpl_LQ10VDocSTdoNV	Picture	Xavier College	190	197	80
Archer Duke				6/04/2008	178	71.2	Not Applicable	No Picture	Not Applicable			
Archer Grant	0	0	0	8/18/2007	177	70.45	https://lh3.googleusercontent.com/5-dhs0ACbbnRtHTEPMVw7Zcts4wllcO1kQxf9EtUi4LT7iY	Picture	St Bede's College	177	177	0
Archie Bull				1/20/2008	176	70.8	Not Applicable	No Picture	Not Applicable			
Archie Edwards	0	0	0	7/16/2007	187.5	76.2	Not Applicable	No Picture	Melbourne Grammar School	0	187.5	76.2
Archie Hooper-Duffy				3/09/2008	183.5	82.1	Not Applicable	No Picture	Not Applicable			
Archie Ludowyke	29	21	25	11/19/2007	194.5	78.3	https://lh3.googleusercontent.com/OksV_7fUftFbgV2JQiAHqa7FsvA8B8Gy7qMo_P5ickxmlB	Picture	Brighton Grammar School	187	194.5	78.3
Archie May							Not Applicable	No Picture	Not Applicable	0	0	0
Archie Roberts							Not Applicable	No Picture	Not Applicable	0	0	0
Aris Moustakis	0	0	0	10/20/2006	175.5	69.3	Not Applicable	No Picture	Albert Park College	0	175.5	69.3
Arki Butler				4/23/2008	180	72.3	Not Applicable	No Picture	Not Applicable			
Bailey McKenzie	0	36	39	3/10/2006	190.5	81.1	Not Applicable	No Picture	Brighton Grammar School	195	190.5	81.1
Ben Forster				5/22/2008	184	66.3	Not Applicable	No Picture	Not Applicable			
Ben Baker				2/23/2008	182	77.5	Not Applicable	No Picture	Not Applicable			
Ben Challis				8/11/2008	190	77.2	Not Applicable	No Picture	Not Applicable			
Benjamin Seers	100	71	64	11/20/2006	191.5	78.05	Not Applicable	No Picture	Mentone Grammar School	170.5	180.5	71.4
Bennett Martin	0	0	0	1/26/2007	176.5	67.85	https://lh3.googleusercontent.com/UgjflsPmuTEvPhaNrzUhuyMD5V0ghAfaldpcTFitb7yjrloAg	Picture	St Bede's College	196	176.5	67.9
Billy McGee Galimberti							Not Applicable	No Picture	Not Applicable	0	0	0
Brodie Findlay	100	100	100	5/20/2005	196.5	106.6	Not Applicable	No Picture	Uni - TBC	196.5	196.5	106.6
Calsher Dear							Not Applicable	No Picture	Not Applicable	0	0	0
Cameron Saultry							Not Applicable	No Picture	Not Applicable	0	0	0
Charlie Beaumont				2/06/2007			https://lh3.googleusercontent.com/EC4GqNd5NwrQ9WLu7UR_xzl0yzJkfzuD4TgA-jh5b96Gu	Picture	Melbourne Grammar School	0	185	0
Charlie Brookes				2/22/2008	175	72	Not Applicable	No Picture	Not Applicable			
Charlie Bruce							Not Applicable	No Picture	Not Applicable	0	0	0
Charlie Rozenes	0	0	0	8/09/2006	178	75.5	Not Applicable	No Picture	Wesley College	177	178	75.5
Charlie Edwards							Not Applicable	No Picture	Not Applicable	0	0	0
Charlie Eerhard							Not Applicable	No Picture	Not Applicable	0	0	0
Charlie Gaiardo							Not Applicable	No Picture	Not Applicable	0	0	0
Charlton Harrop							Not Applicable	No Picture	Not Applicable	0	0	0
Chris Kellaway				11/07/2006	173.5	69.7	Not Applicable	No Picture	Caulfield Grammar School	195	191.5	0
Cody Goonewardene							Not Applicable	No Picture	Not Applicable	0	0	0
Cooper Lord							Not Applicable	No Picture	Not Applicable	0	0	0
Darby Heelis				5/13/2007			https://lh3.googleusercontent.com/a4ci8ANEnqc6PJvHs8UPnisBesbwCONDtx35UNoORFKlB	Picture	Not Applicable	0	0	0
Dexter Prime	0	0	0	9/06/2006	193.5	81.5	Not Applicable	No Picture	St Kevin's College	200.5	193.5	81.5
Eddy Cooper				6/28/2007	190	85.3	https://lh3.googleusercontent.com/aTSJL-xk98ZIm-gbXBs56IBhFOUmGg6_BpwWyKslJXg0L	Picture	Not Applicable	0	0	0
Eiden Pitt	100	93	93	6/01/2007	189	84.2	Not Applicable	No Picture	Mount Scopus Memorial College	194	189	0
Emmanuel Ganas				4/26/2007			https://lh3.googleusercontent.com/VbqbHRfc2ySoTSlli1Z-_8exLfCrc5frPBIH0rqA4RL2ONR5r	Picture	De La Salle College	190	190	0
Ethan Lakman				8/27/2007			Not Applicable	No Picture	Not Applicable	0	0	0
Ethan Williams							Not Applicable	No Picture	Not Applicable	0	0	0
Fletcher Teelow	0	0	0		182	81	Not Applicable	No Picture	Not Applicable	0	173.5	81
Flynn Groves				3/20/2006	185	76	Not Applicable	No Picture	St Bede's College	173.5	173	0
Frederick Brayshaw	0	0	0	10/31/2007	173	63.9	https://lh3.googleusercontent.com/8JXgshlbiSr_G2tLMWi73d34K6aBPyxphjCwg3J0ljp3oVx	Picture	Haileybury College	173	181.5	0
George Dimer				6/21/2008	170	64.4	Not Applicable	No Picture	Not Applicable			
Gus Teixeira				3/29/2008	178	70	Not Applicable	No Picture	Not Applicable			
Harrison Oliver	86	93	96	6/15/2006	181.5	73.5	Not Applicable	No Picture	Brighton Grammar School	193	195	73.5
Harry Armstrong	86	71	61	6/14/2006	195	85	Not Applicable	No Picture	Haileybury College	195	189	85
Harry Ockerby				1/10/2008	168	54.8	Not Applicable	No Picture	Not Applicable			
Harry Sullivan							Not Applicable	No Picture	Not Applicable	0	0	0
Harvey Allen	0	0	0	5/29/2007	189	80.15	Not Applicable	No Picture	Brighton Grammar School	189	0	0
Harvey Johnston							Not Applicable	No Picture	Not Applicable	0	0	0
Henry Parris				9/04/2008	191	70.5	Not Applicable	No Picture	Not Applicable			
Henry Thomas				11/04/2008	192	80.9	Not Applicable	No Picture	Not Applicable			
Hudson Wright				2/02/2008	182	74.2	Not Applicable	No Picture	Not Applicable	190	179	0
Hunter Holmes							https://lh3.googleusercontent.com/Mi3VvUQcV5GmhFiG6K_UfmcjFFm1o5by-Th_trLmBw9EB	Picture	Not Applicable	0	0	0
Hunter Lynch	0	7	4	10/28/2006	179	74.95	Not Applicable	No Picture	Not Applicable	0	194	75
Hunter Woodruff				4/19/2008	186	77.9	Not Applicable	No Picture	Not Applicable			
Isaac Dawson				4/25/2008	177	67.7	Not Applicable	No Picture	Not Applicable			
Jack Cheep	100	100	100	4/25/2006	194	86.95	Not Applicable	No Picture	Parkdale Secondary College	194	177	87

Figure 6. Sample of Finalized Heights and Weights Data during the 2022-2024 Season for the Sandringham Dragons Football Club

After transforming and cleaning irrelevant data during the "Data preparation" phase data, the folders **"2km Results", "GPS Data," "Heights and Weights Data," "Gym and Strength Data,"** and **"Combine Data"** were used to proceed with the "Exploratory Analysis and Modelling" phase, were such as **Max Speed Data and Player Football Match Statistics** were left out during the analysis confirmed with data team. To clarify any more data that needs cleaning, thoroughly check with the lead supervisor since we would need to model during the data analysis project if necessary. The supervisor proposed fundamental key statistics for computing players and team performance data metrics such as averages, minimums, maximums, totals, and standard derivations. However, during the exploration phase, additional modifications were made during the data analysis project, which included applying z-scores **(see Appendix F)** from the **Gym and strength data** for each of the critical features, which will later be used to visualize the "Validation" phase in the final dashboards. He proposed using the maximum value for each column as an observation value to enable and encourage players to optimize their results before draft proceedings. Qualitative and quantitative analysis takes place to demonstrate descriptive analysis, applying some of the descriptive statistics (Means et al.). Refer to the five characteristics when collecting big data from many different sources: Value, Velocity, Volume, Veracity, and Variety are essential to further enhancing the "data preparation" to derive more value from the data to indicate better decision-making." Maintain all quality and security issues and resolve fewer redundancies on data as long it has met the deliverables and business requirements and proposals. Before presenting our results to data visualization platforms, we must decide on Power BI. After the preparation and exploratory analysis stages were completed, the supervisor initialized what our final dashboards would be. During the placement, proposals change over time since we're still waiting for all data collection to be completely finalized, likely due to the expansion for recruits towards players under 16 to 18 years of age, including players who are not in the program that may be considered continuing playing into the program soon for the 2022-2024 Season. This will likely impact changes and improvements in critical data and metrics for crafting top talent players and their overall team performances.



Everything important about Power BI Desktop
FILE
Get data
Import
Export
Publish
Connect to data from Multiple sources
Import Contents from an existing file
Export Contents to a new file
Publish the report online to PBI Service
Install
Go to app.powerbi.com
Download PBI Desktop
Report Canvas
This is where all your work displays. Use Report Editing Panes (below) to manipulate reports and dashboards
HOME
Get data
Excel workbook
Data hub
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New measure
Quick measure
Calculations
Publish
Share
Connect to various Data sources
Data Hub allows to connect with Org Data in the form of PBI Datasets
Some commonly used sources as a quick link. Can also be accessed via GET DATA Dropdown
Connect, Prepare, and Transform Data using the Power Query Editor
Add new Visuals from standard PBI Library
Add Text Box for Commentary
Add Visuals from PBI App Store
Add New custom DAX Expression that calculates value from your Data. Or, select from pre-defined Quick Measures
Publish and share your report to POwer BI Service for access to relevant users
INSERT
New page
Pages
New visual
More visuals
Visuals
Q&A
Key influencers
Decomposition tree
Smart narrative
AI visuals
Paginated report
Power Apps
Power Automate (preview)
Power Platform
Text box
Buttons
Shapes
Image
Elements
Add a sparkline
Sparklines
Add multiple pages to the report
No code AI-based data visualization and analysis. Discover new insights, find patterns, and explain complex data interactions.
Q&A: Ask questions in Natural language
Key influencers: Major factors affecting a trend or results in your data
Decomposition tree: Aggregate and drill down data in any order
Smart narrative: Create text narratives of your Data and Visualizations
Integrate PBI Reports, Power Apps, or Power Automate. This allows update to other reports, access apps, or trigger workflows without leaving Power BI
Add different elements to the report
Add tiny charts within cells of a table or matrix to easily and quickly compare trends
POWER QUERY EDITOR
Connect to one or many data sources, shape and transform the data, and then load that model into Power BI Desktop.
Save your work and Load changes to PBI Desktop
Launches the Data Transformation ribbon to perform additional data transformations such as filtering, sorting, grouping, and aggregating data.
Launches the Add COlumsn Ribbon to do more advanced tasks like:
• Adding a new column from existing columns
• Adding a new column from a calculation
• Adding a new column from an external data source
Perform Text and Vision Analysis using Azure AI Cognitive Services. Perform:
• Sentiment Analysis
• Key Phrase Extraction
• Language Detection
• Image Tagging
Perform common Data editing tasks here. Connect to new sources, add, edit, and manipulate, and preview the Data.
REPORT EDITING PANES
Create drag and drop visualizations from a huge library within Power BI. Also enables adding Python or R based visualizations.
Filter the reports based on drag and drop data fields. Use these to Filter through Specific Reports on One page or entire reports on all pages.
Displays the tables, folders, and fields in your data that are available for you to use to create visualizations. Drag and drop selected fields where you want to show it e.g. as legends, values or axis.
DAX MEASURES
Total Sales = SUM(Sales[Sales Amount])
DAX Function
Reference Table
Reference Column
DAX formulas can be used to create calculations on data.
DAX formulas are made up of Functions, Operators, and Values. Functions perform specific operations on data, such as SUM, AVERAGE, and COUNT. Operators perform mathematical operations, such as addition, subtraction, and multiplication. Values can be numbers, strings, dates, or other data types.
by Tariq Munir

Figure 7. Power-BI Spreadsheet by Tariq Munir demonstrates all the functionalities to perform efficient dashboards

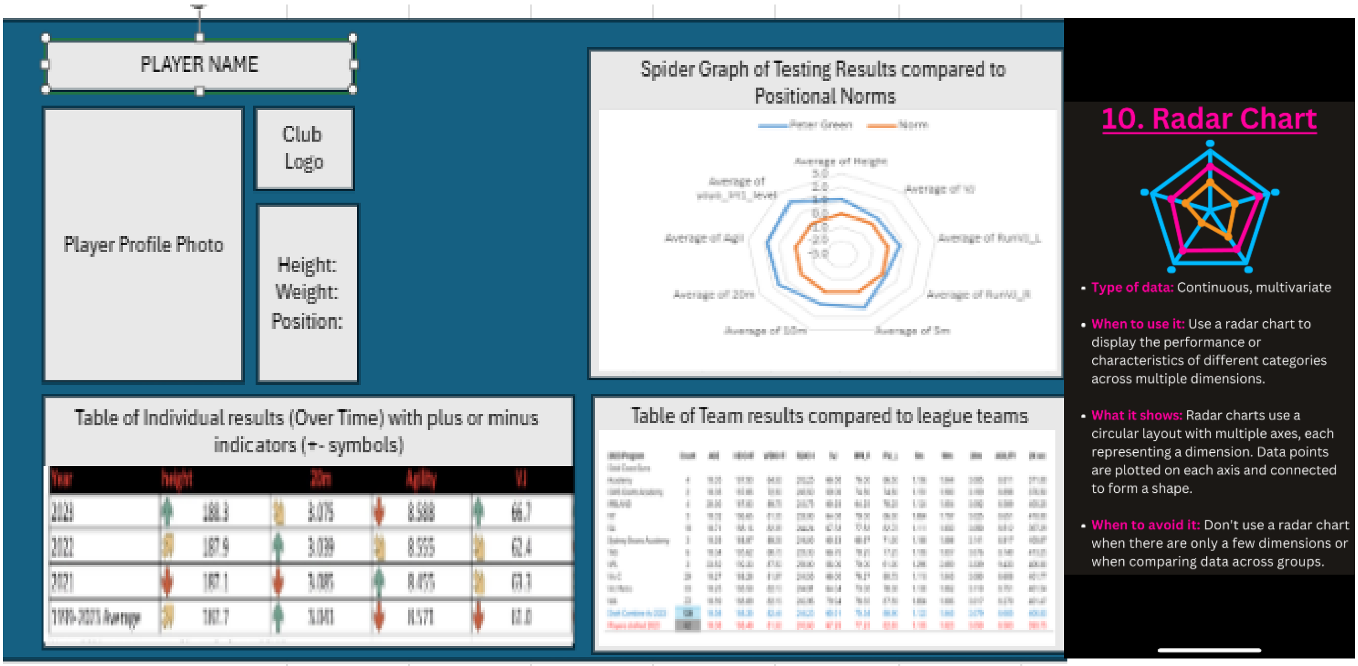


Figure 8. Supervisor's initial dashboard shown during via live chat during the placement including tables and Radar chart

Figure 9. Second initial Dashboard on Bar charts are commonly used in sports analytics displaying key metrics for player performance from GPS data using Power-BI.

Figure 10. The third dashboard depicts multiple line charts displaying key metrics for player performance and development during the July 2023 to January 2024.

They display the dashboards to match accurately the initial dashboards **(Figures 8, 9, and 10 above)** from the lead coach to distinguish the following data interpretations and "highlights" for the organization to secure all trust and integrity to prevent data leakage. Ensure that the data is shared internally within the organization to prevent "sensitive information" and "identify theft" from third-party organizations to accomplish their goals to prevent ‘financial’, ’reputational’ losses. Furthermore, the data team has to clarify better interpretations in terms of using the appropriate data features to enable decision-making to drive key performances more effectively without being misinterpreted in a kindly manner. The organization at Foundry Athletic proposed one of the two powerful data visualization tools to demonstrate data through dashboards to enable stakeholders and organizations to interact and interpret the data. Deciding from Tableau or Microsoft Power BI, data analysts can draft different visualizations through dashboards and reports to conclude our findings of what the data can impact for business decisions, direction, and value. After the first four phases of the data analysis project are completed to ensure the data team is done modifying the changes, we then proceed to the "Validation" phase of the data analysis project. During further discussion of finding the appropriate data visualization tool, the team proposed using Microsoft Power-BI. Since it is accessible and convenient to use, cost-effective, and supports multiple "data sources," it is suitable and easily shared with other users to reduce data leakage (i.e., revealing personal or sensitive information) and maintain data integrity within the organization without breaching to external third-party organizations. After carefully considering which data visualization platform is compatible with other tools required to demonstrate better dashboards, leaving minimal flaws in our data, power BI can "analyse and share large datasets" with large businesses and organizations. Enhances "sharing and collaboration aspects of business intelligence." It can integrate Microsoft Excel spreadsheets and samples of code in Python, R, and SQL, which is easily incorporated into Power BI Desktop. It is time-consuming nether-less, allows further adjustments with "the data being updated in real-time; businesses can identify and address issues promptly." This enables data analysts to make some modifications from previous phases to improve our data insights and adjust efficient" where experimenting different visualizations and implementing key metrics up to four to five significant figures with further context of the following values displayed; the lead coach has envisioned making the final product of our dashboards match that answers the business objectives and data architecture accordingly before moving onto the final phase during the "Visualization and Presentation" phase where we confirm our findings for the finalized dashboards down below in **Figures 11-30** showing twenty-one unique dashboards and visualizations including valuable player and team performance. This way we wanted to ensure we want to have finer details without confusing the audience that has shown onto their monitor or Desktop.

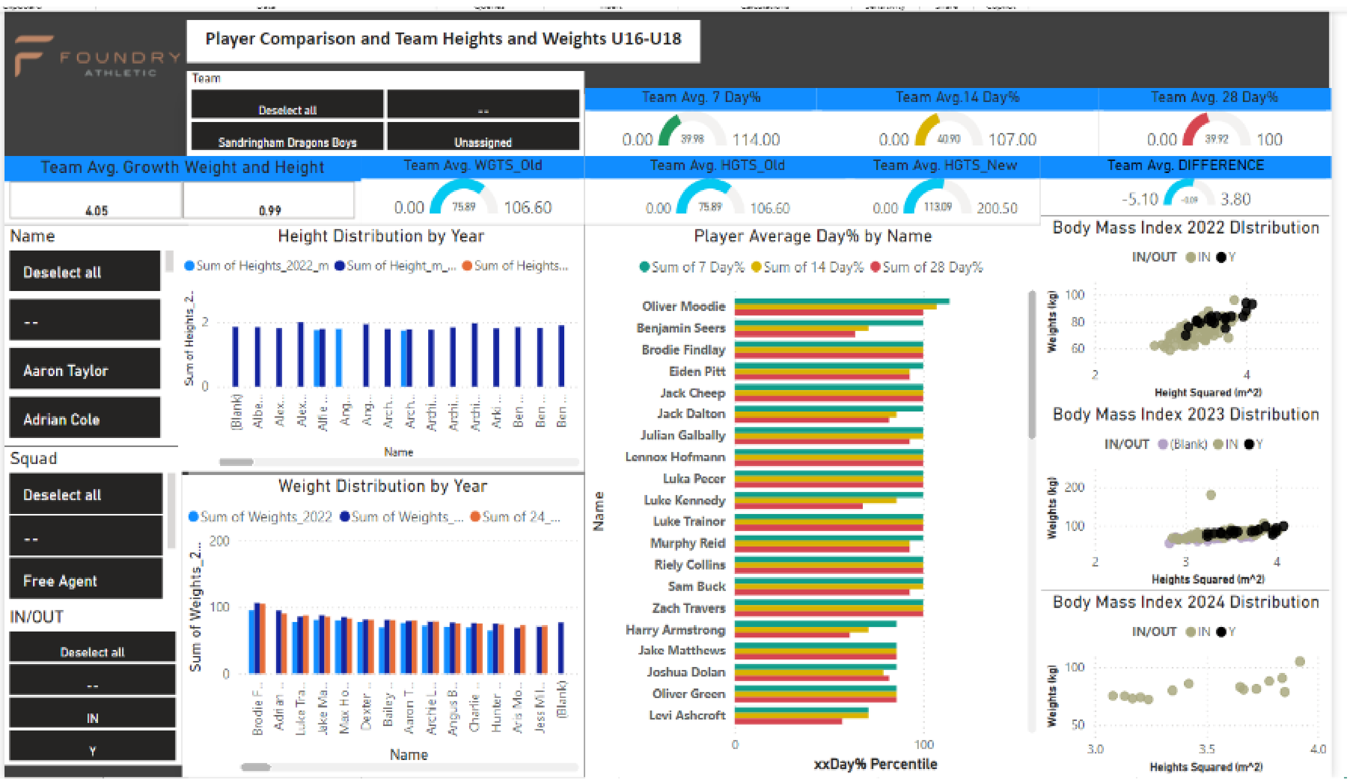


Figure 11. Player and Team Comparisons of Heights and Weights Dashboard

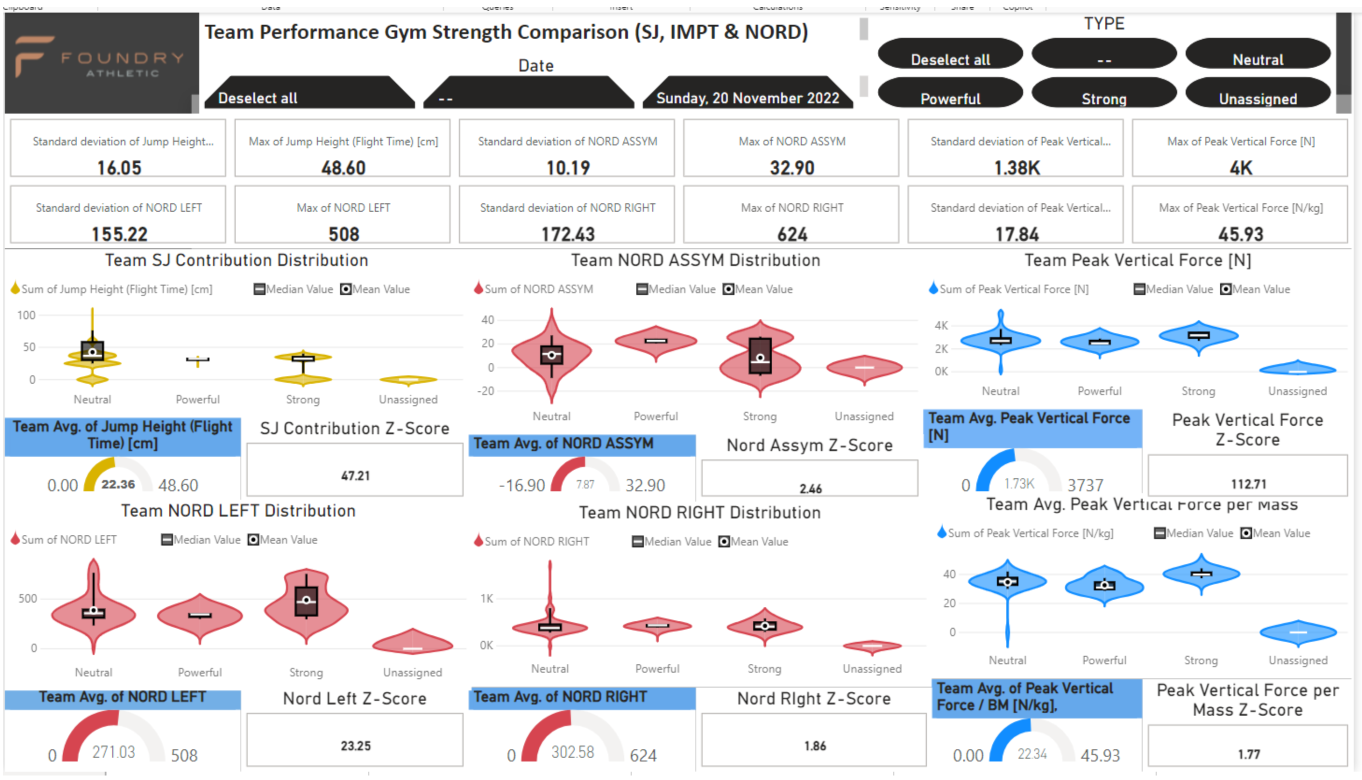


Figure 12. Team Performance Gym Strength Comparison (SJ, IMPT and NORD) Dashboard

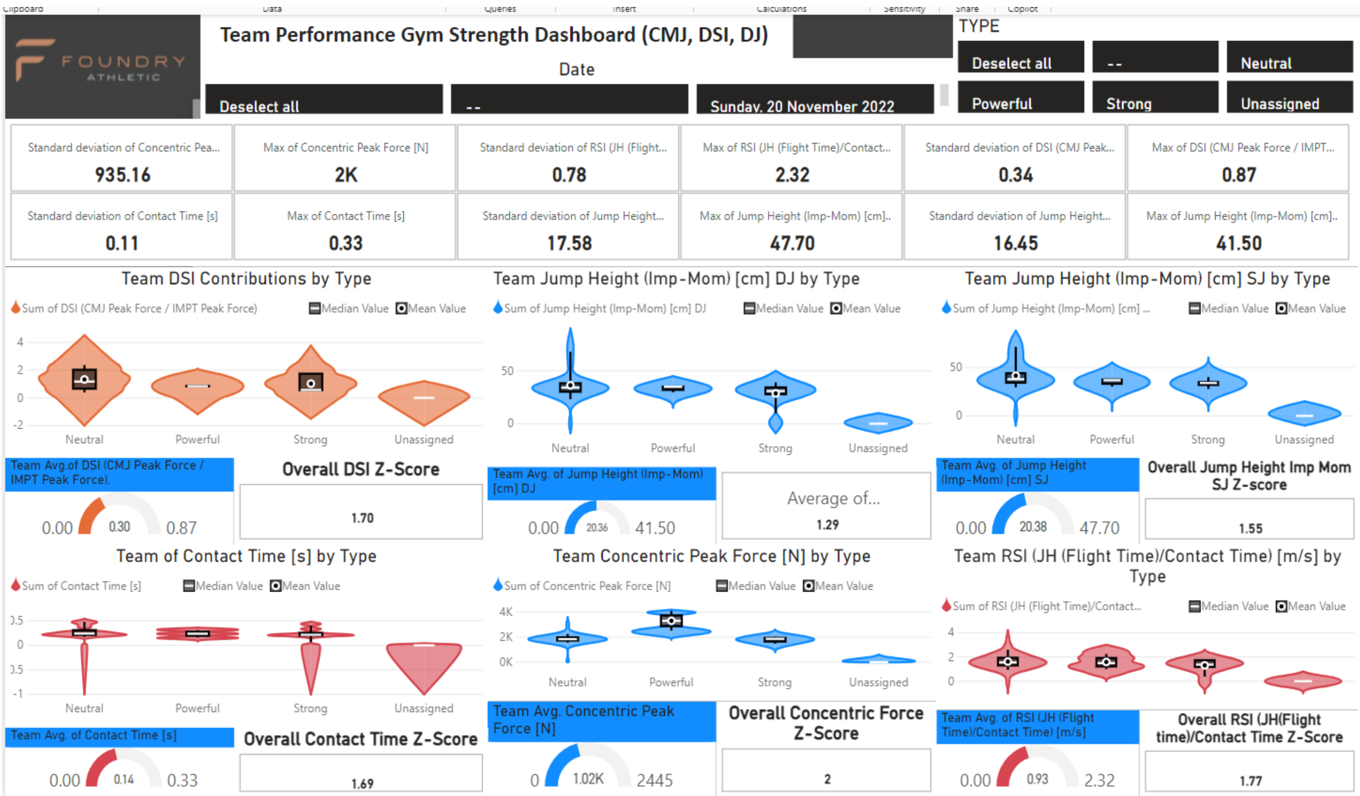


Figure 13. Team Performance Gym Strength Comparison (CMJ, DSI and DJ) Dashboard

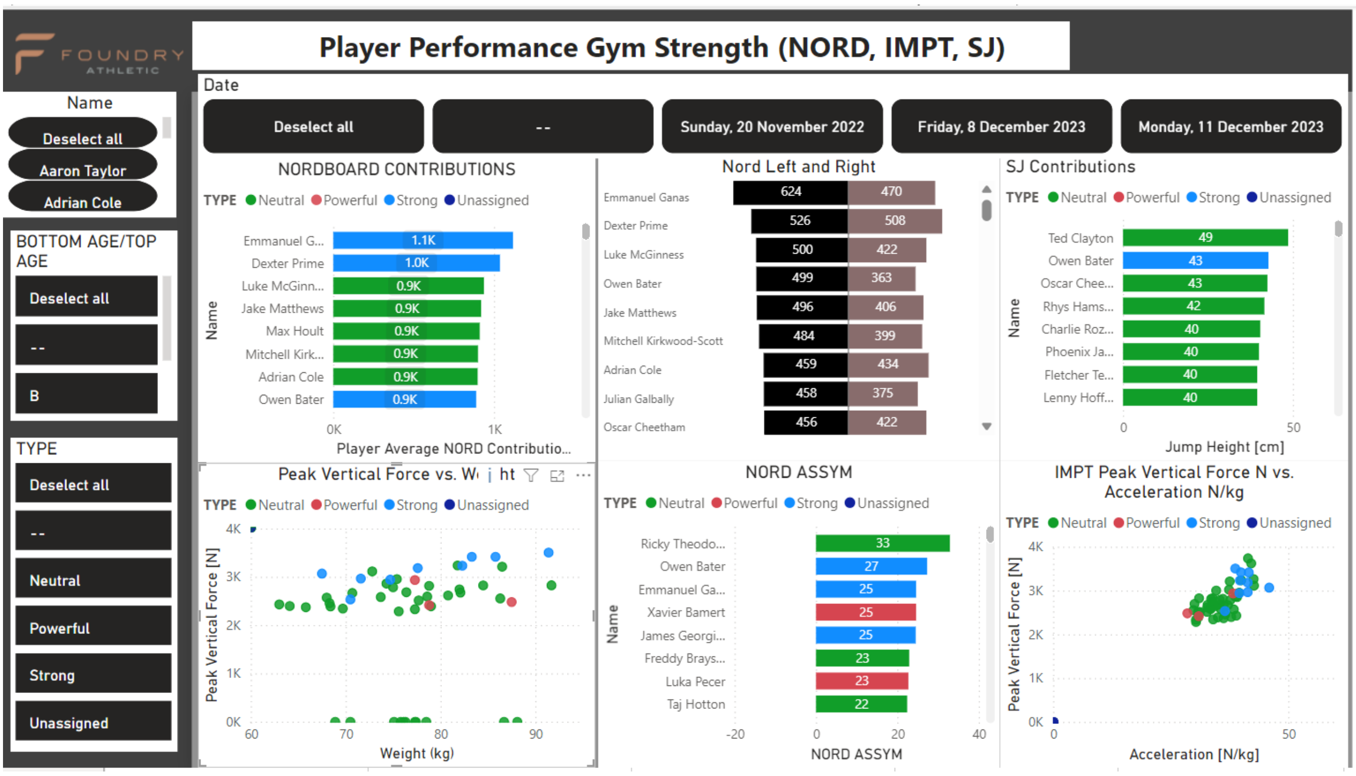


Figure 14. Team Performance Gym Strength Comparison (CMJ, DSI and DJ) Dashboard

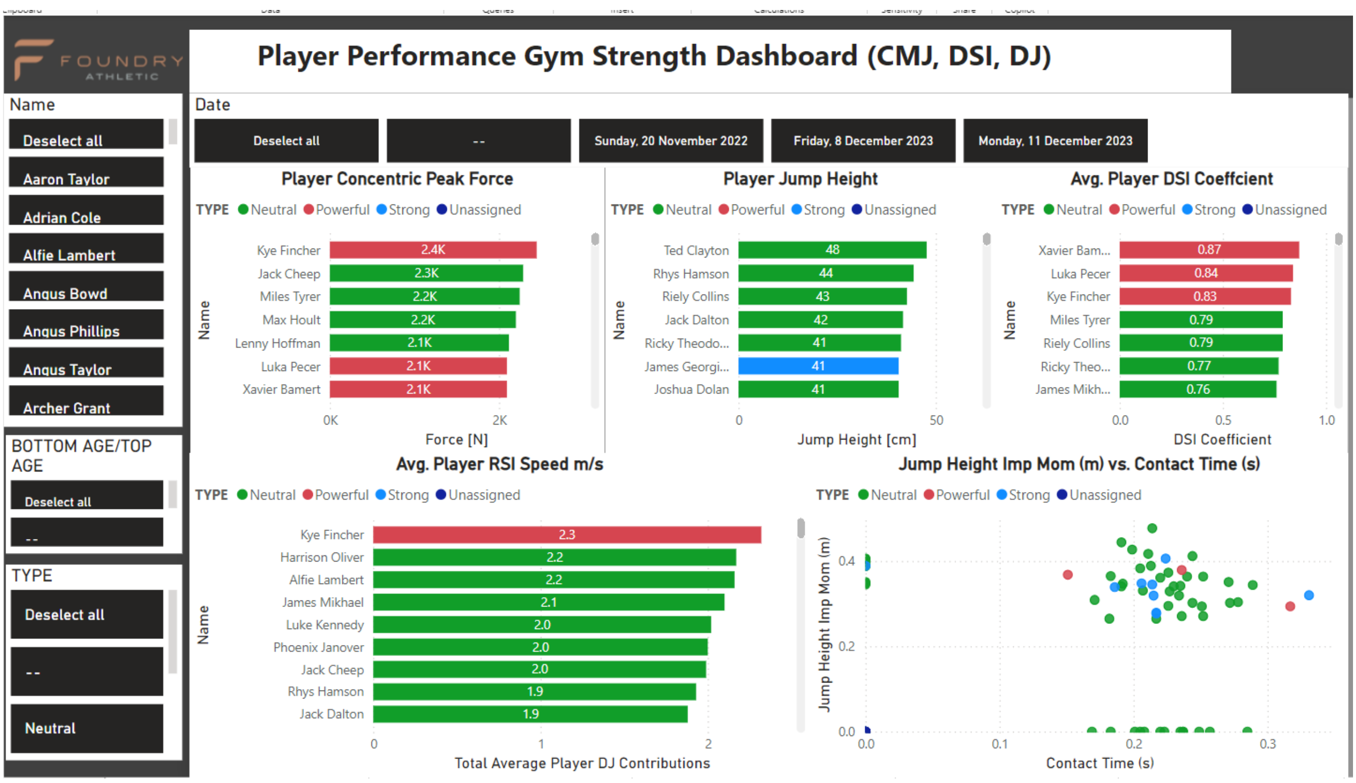


Figure 15. Player Performance Gym Strength Dashboard (CMJ, DSI, DJ)

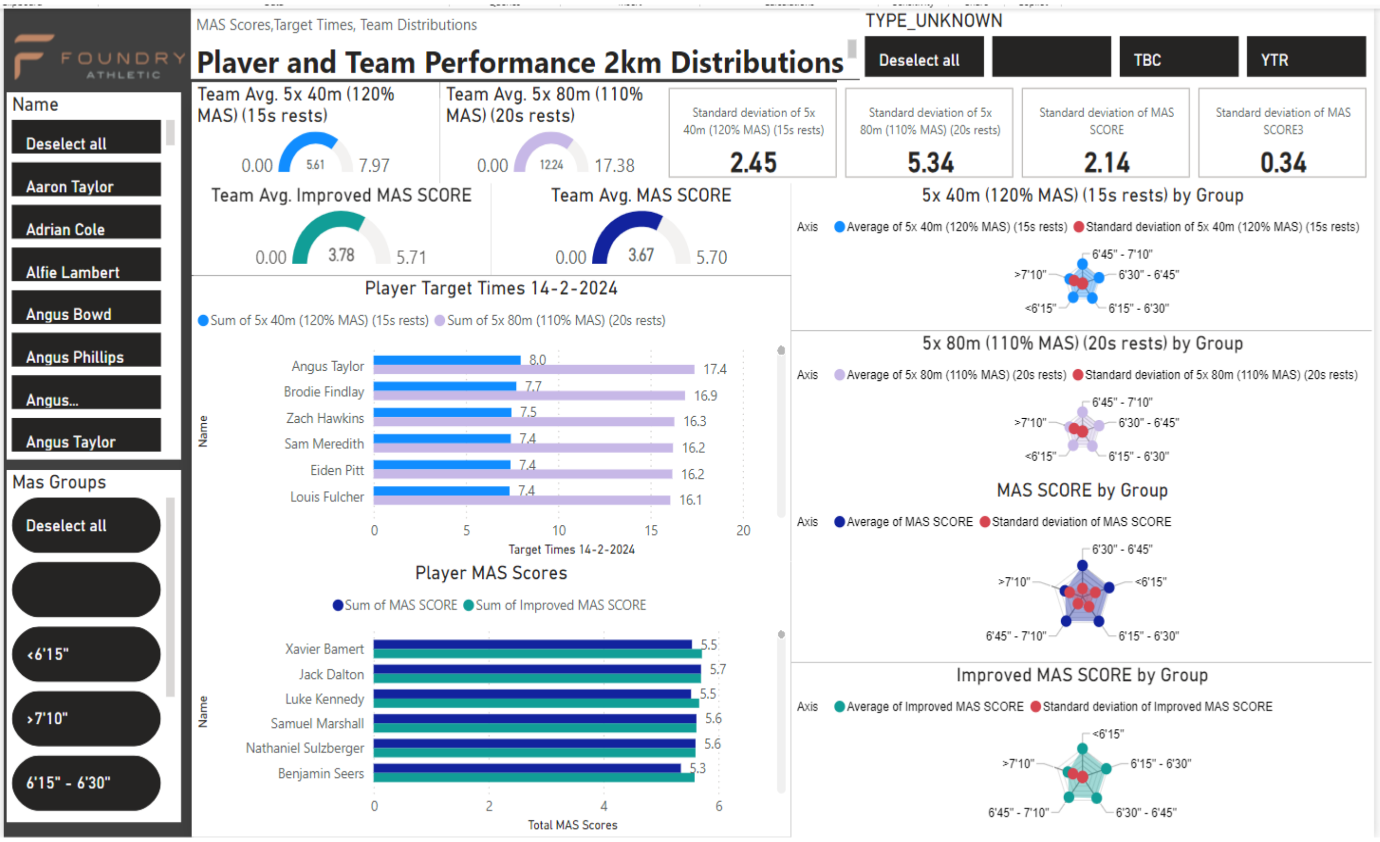


Figure 16. Player and Team Performance 2km Time Distributions comparing MAS Scores and Team Distributions.

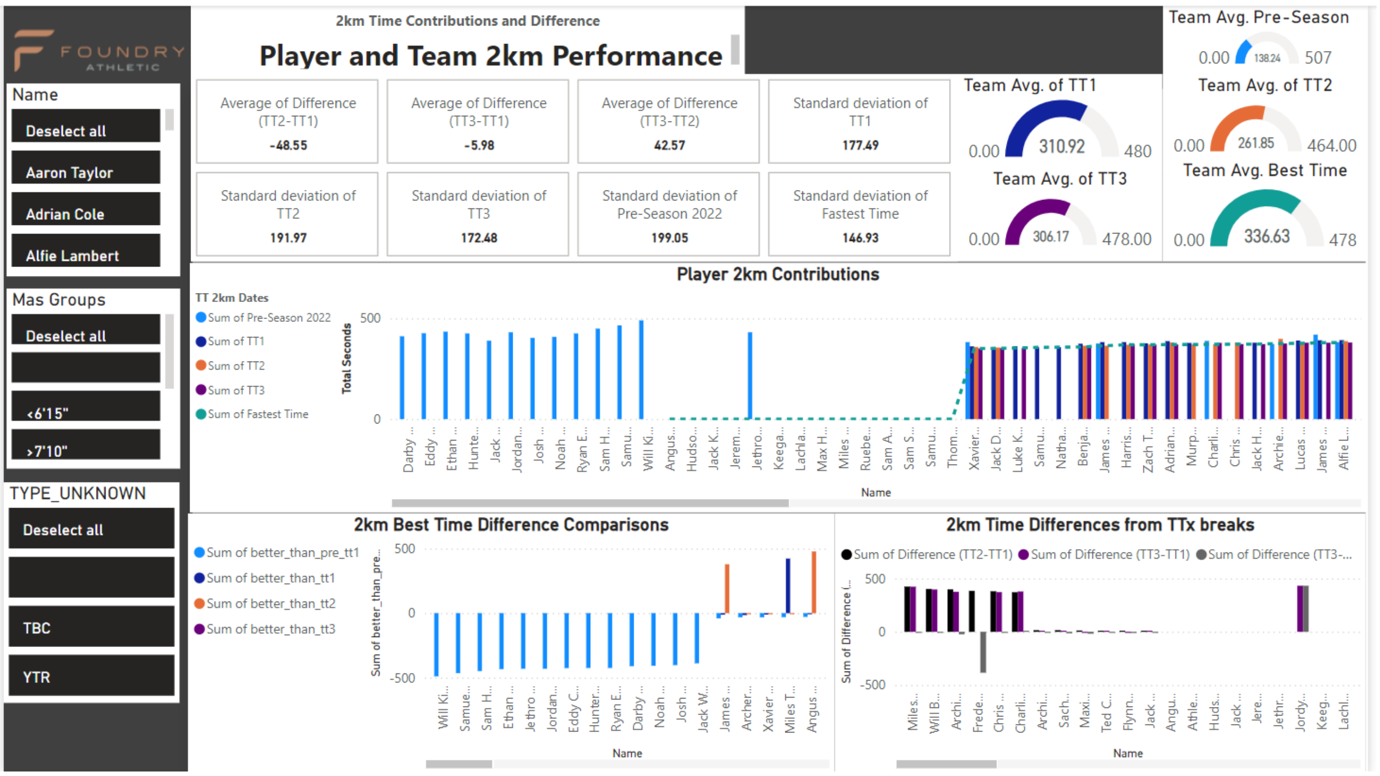


Figure 17. Player and Team Performance 2km Time Contributions and Differences

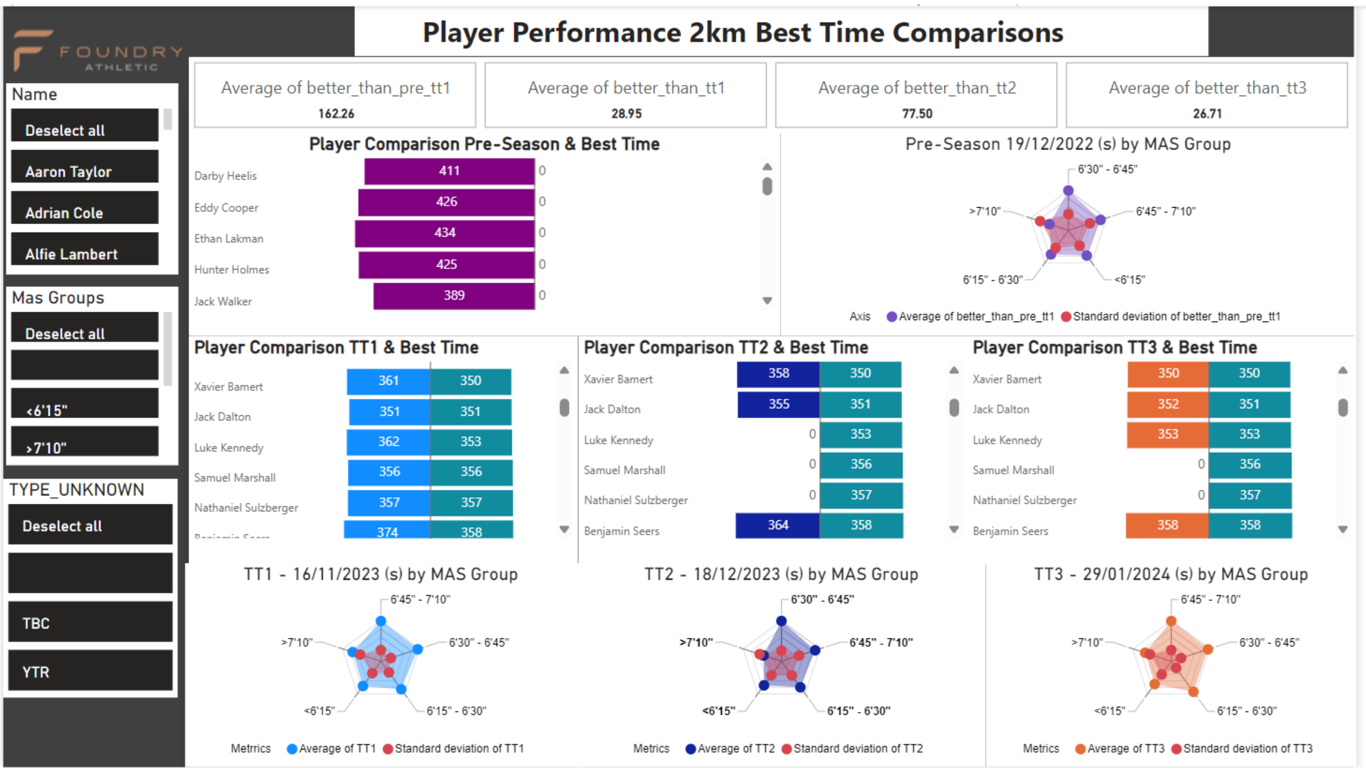


Figure 18. Player and Team Performance 2km Fastest Time Comparisons during 2022-2023 Season

Figure 19. Player and Team Performance Gym Strength RM and REPS Dashboard

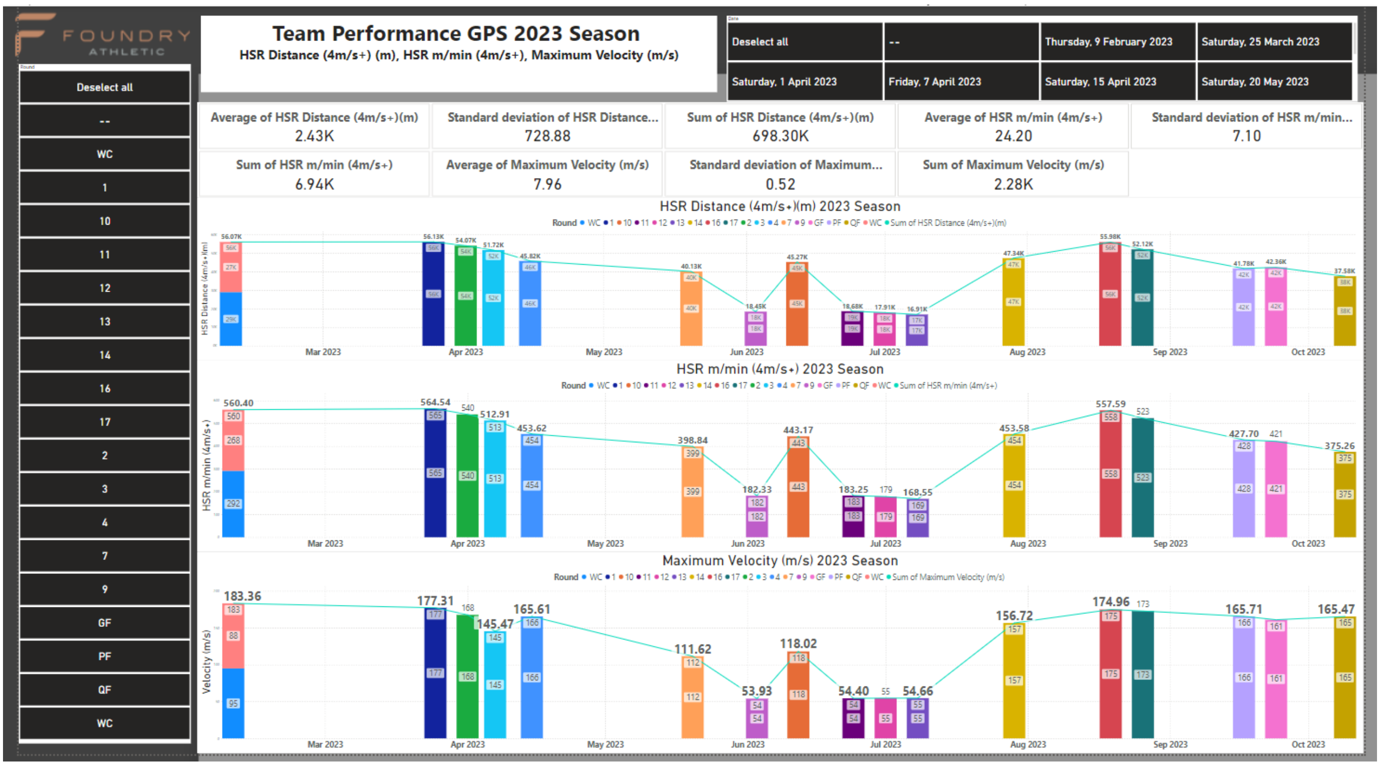


Figure 20. Team Performance GPS 2023 Season Dashboard containing HSR Distance (4m/s+) (m), HSR m/min (4m/s+) and Maximum Velocities

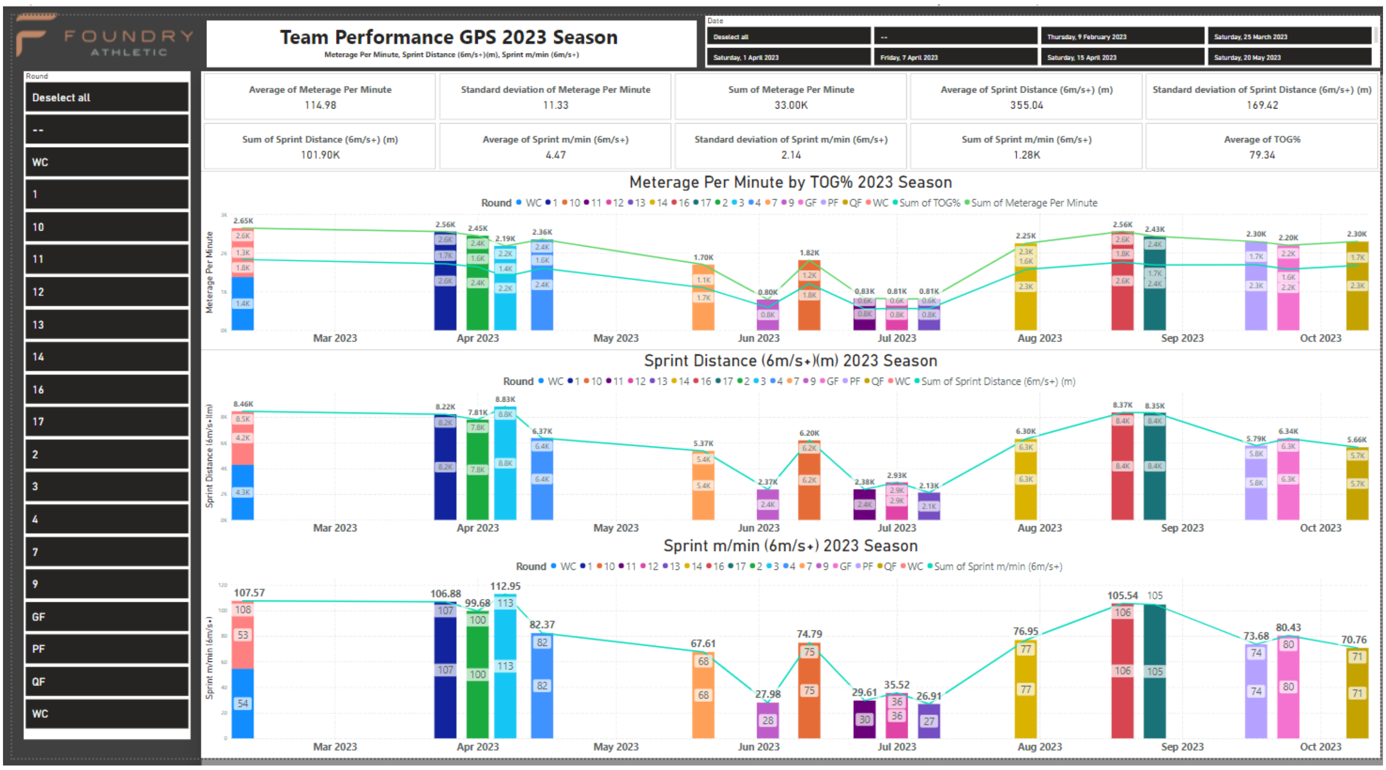


Figure 21. Team Performance GPS 2023 Season Dashboard containing Meterage Per Minute, Sprint Distance (6m/s+) (m) and Sprint m/min (6m/s+)

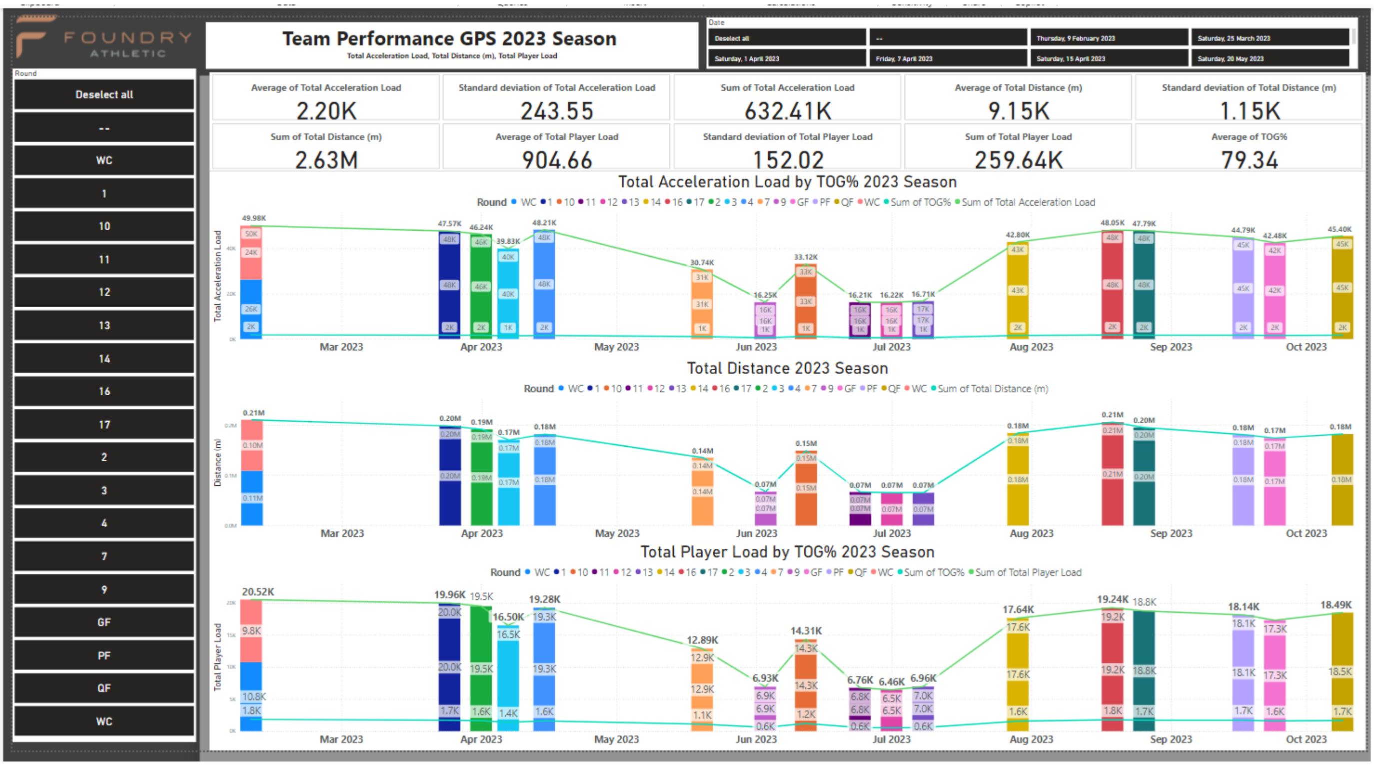


Figure 22. Team Performance GPS 2023 Season Dashboard containing Acceleration Load, Distance meters Taken and Player Load

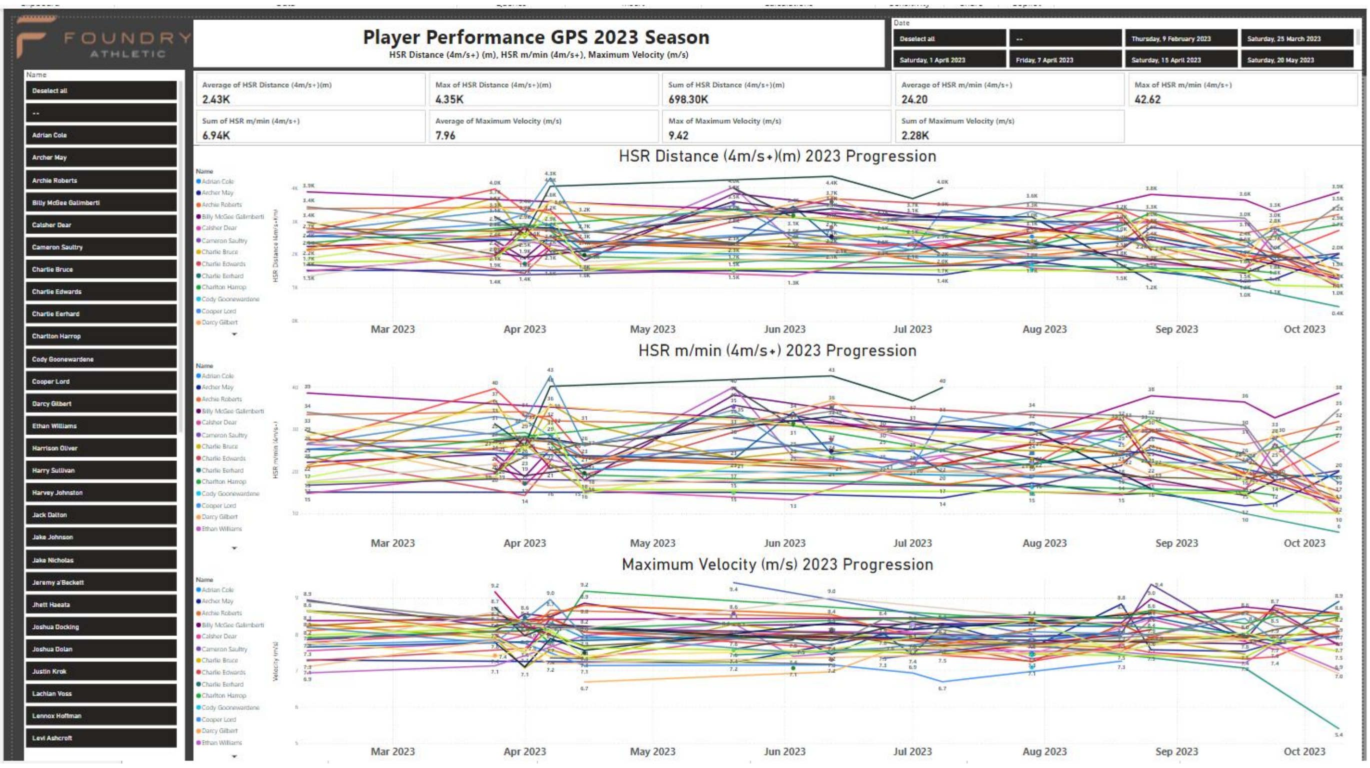


Figure 23. Player Performance GPS 2023 Season Dashboard containing HSR Distance (4m/s+) (m), HSR m/min (4m/s+) and Maximum Velocity

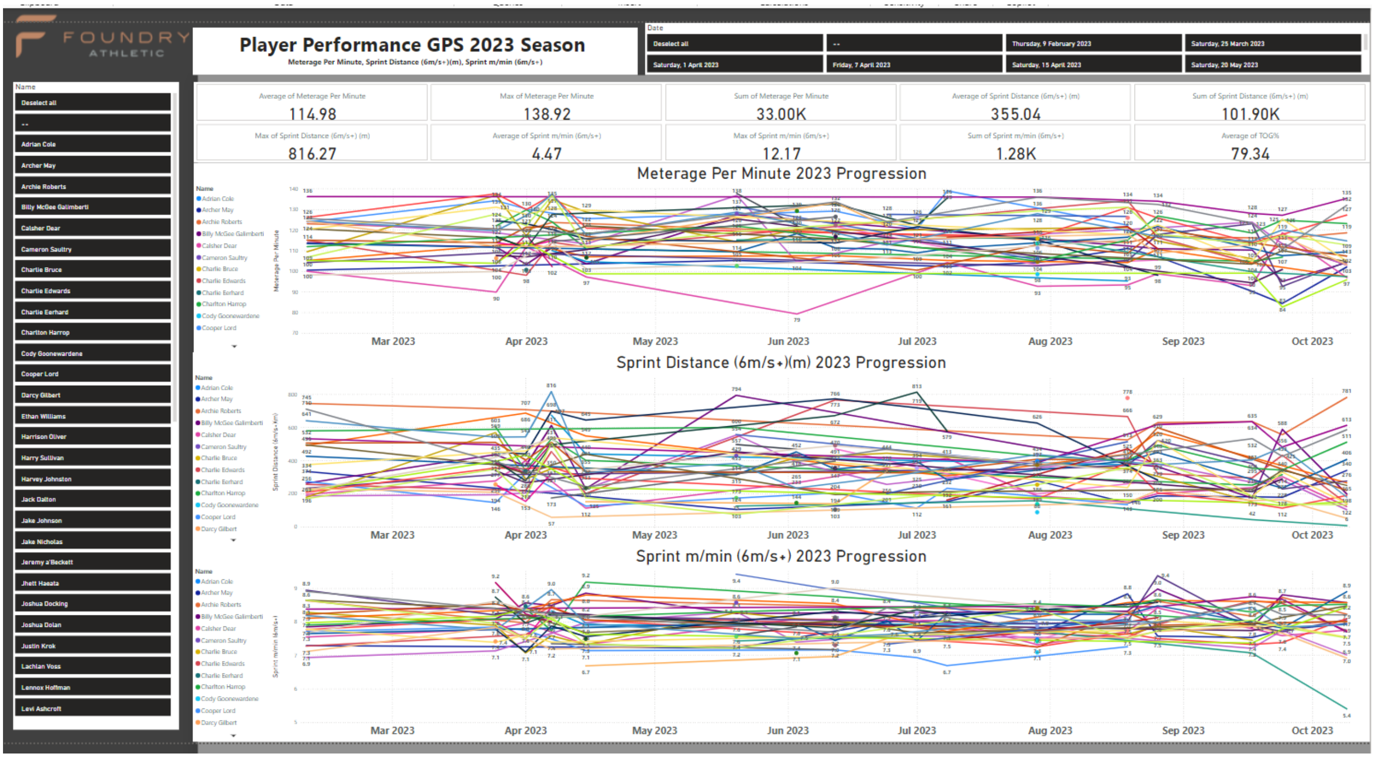


Figure 24. Player Performance GPS 2023 Season Dashboard containing Meterage per Minute, Sprint Distance (6m/s+) (m) and Sprint m/min (6m/s+)

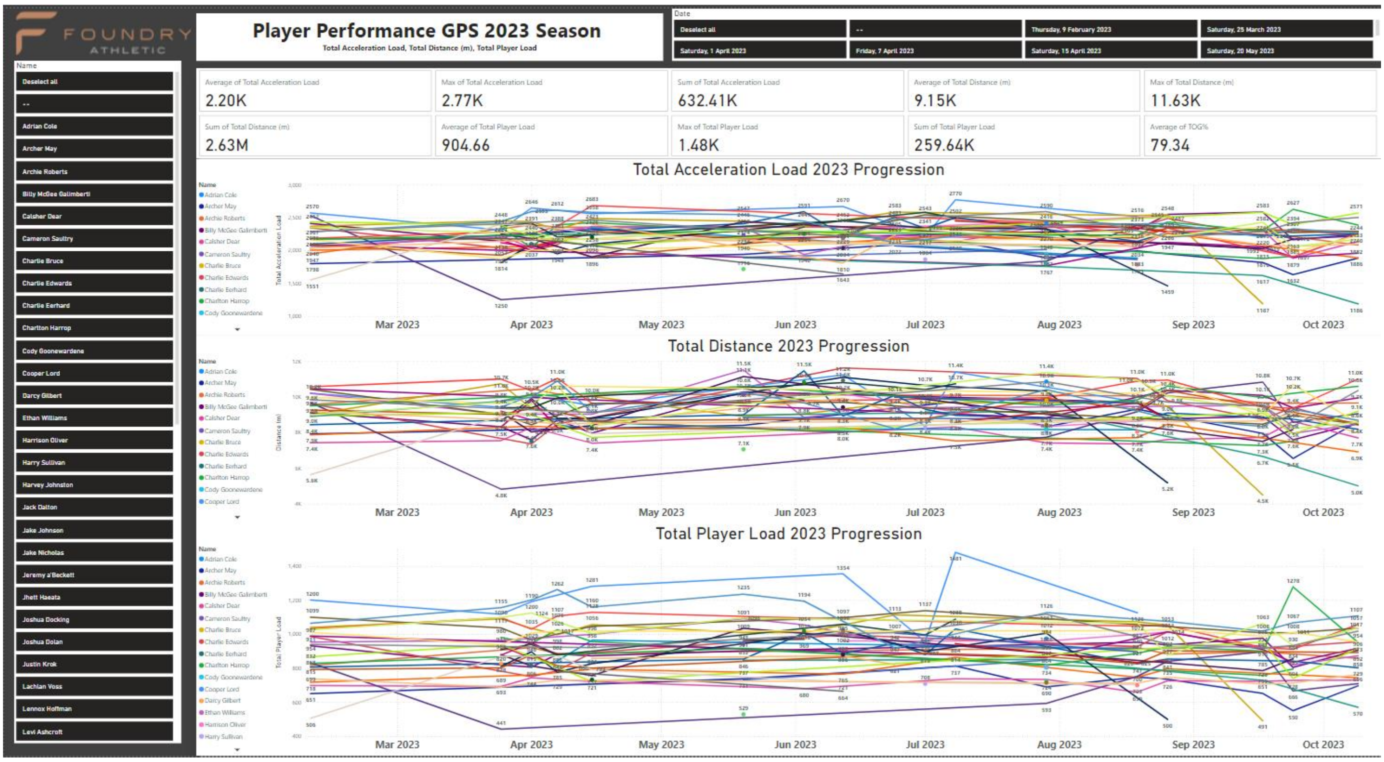


Figure 25. Player Performance GPS 2023 Season Dashboard containing Acceleration Load, Distance Taken and Player Load

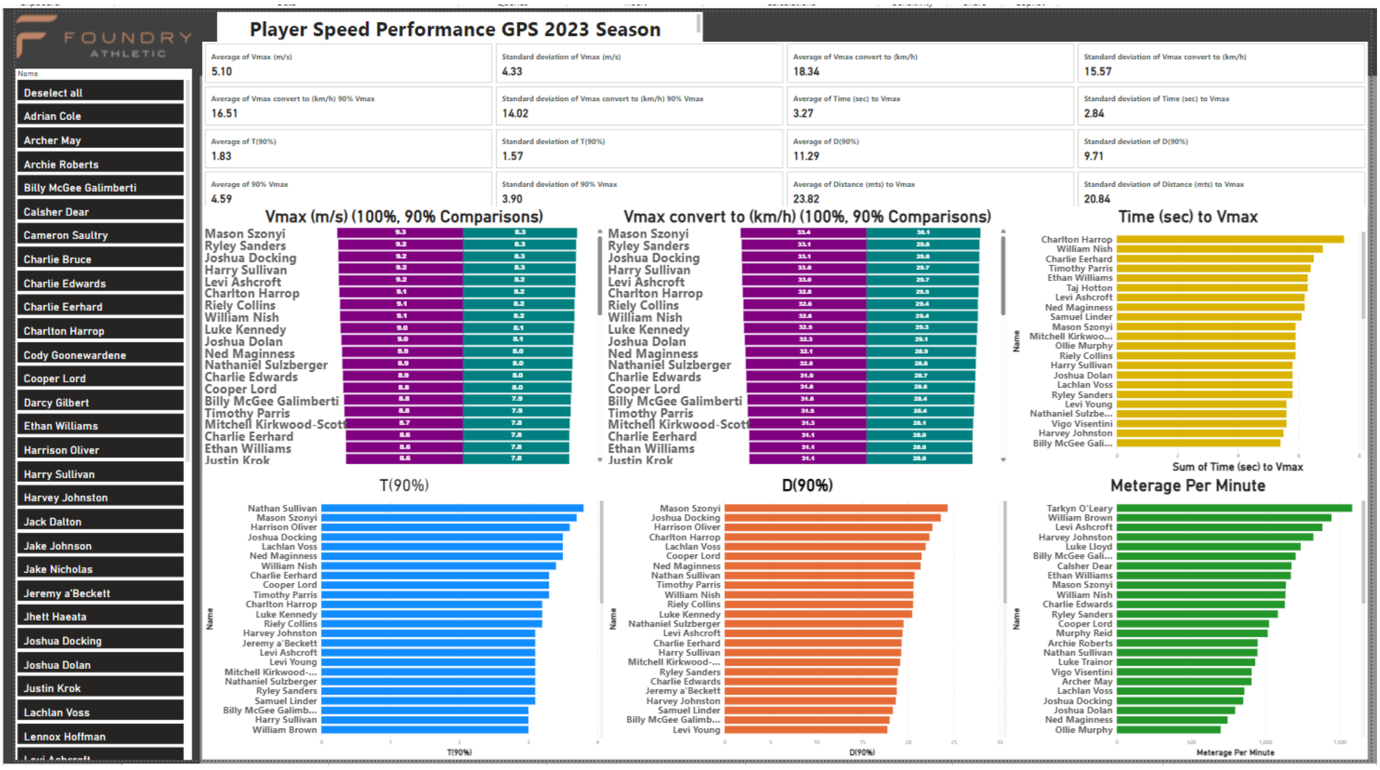


Figure 26. Player Speed Performance during the 2023 Season

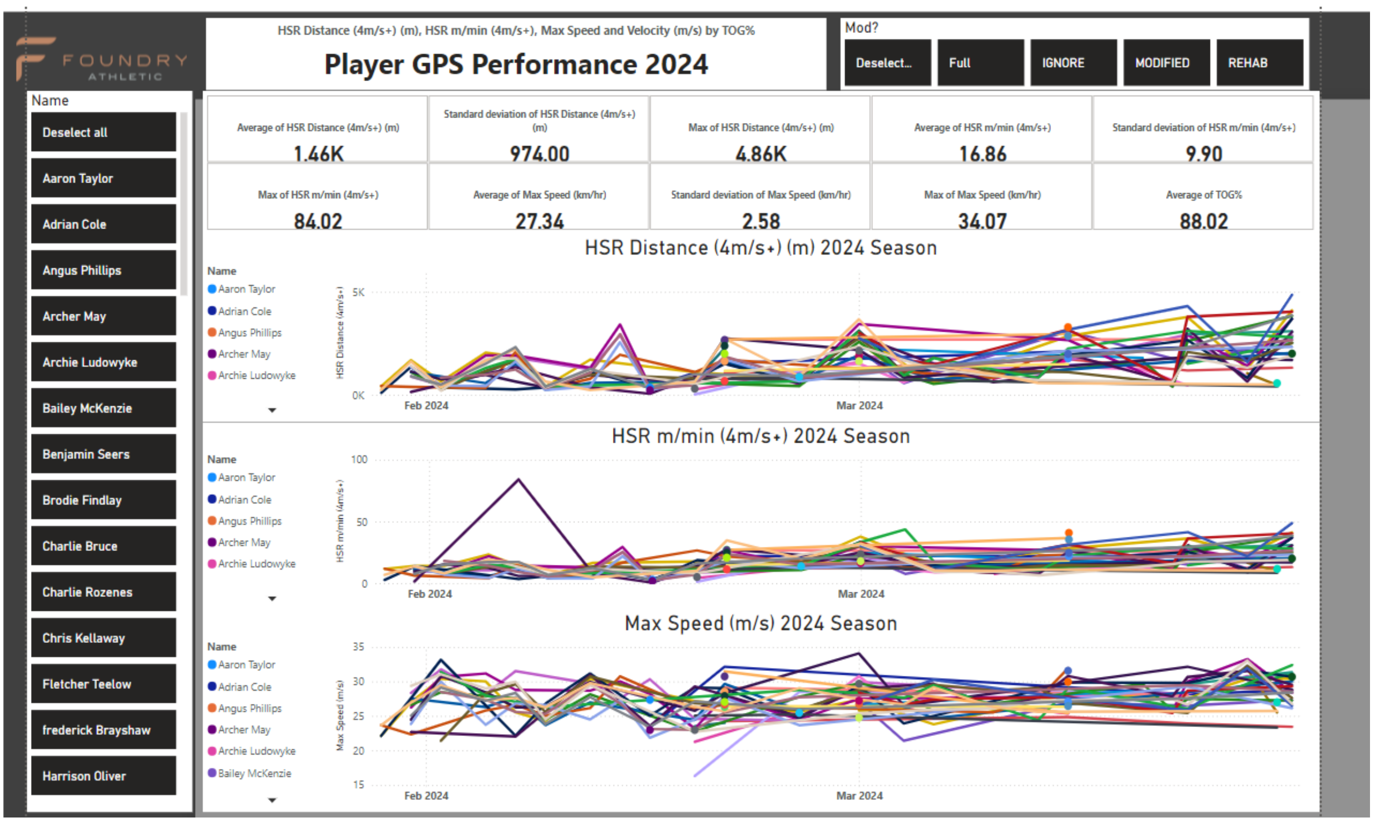


Figure 27. Player Performance GPS 2024 Season Dashboard containing HSR Distance (4m/s+) (m), HSR m/min (4m/s+) and Maximum Speed (m/s)

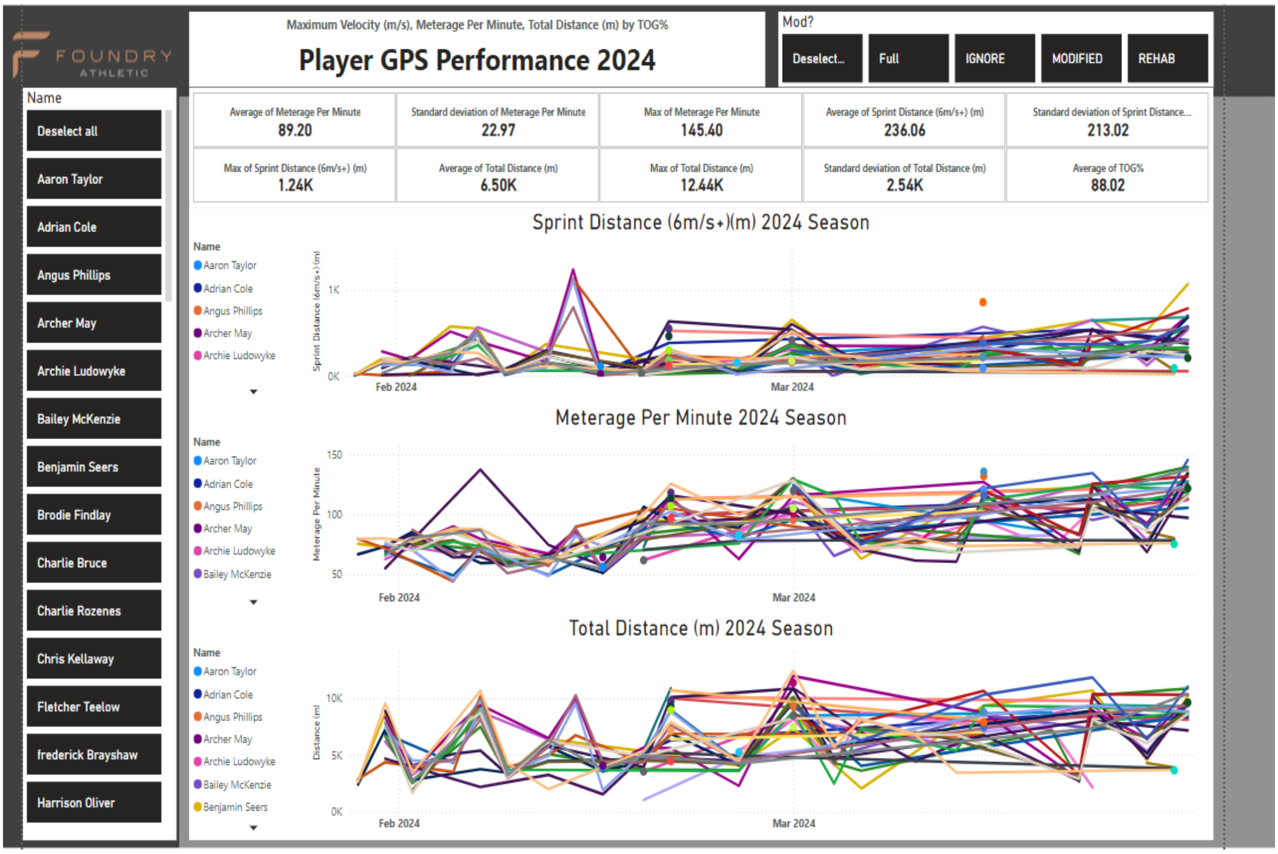


Figure 28. Player Performance GPS 2024 Season Dashboard containing Sprint Distance (6m/s+) (m), Meterage Per Minute and Distance in meters taken.

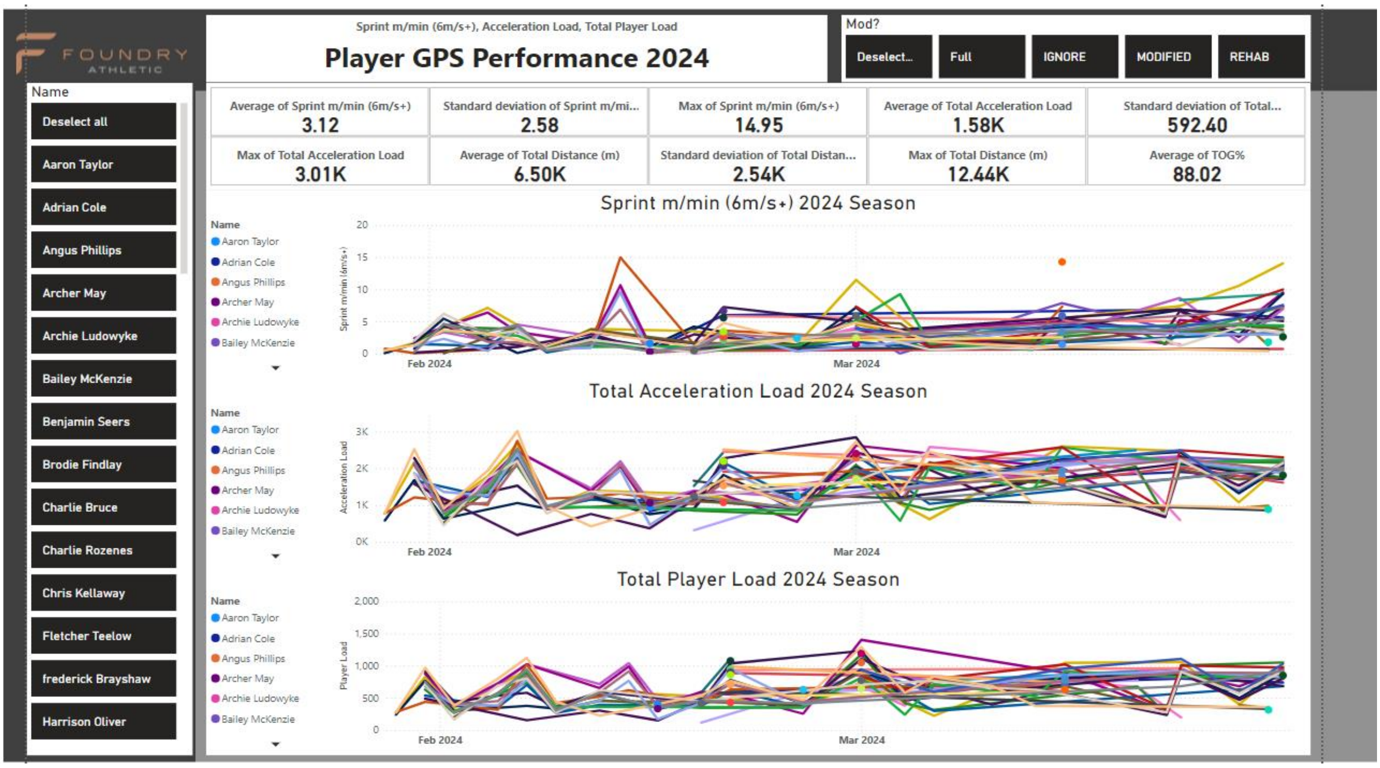


Figure 29. Player Performance GPS 2024 Season Dashboard containing Sprint m/min (6m/s+) (m), Acceleration and Player Load.

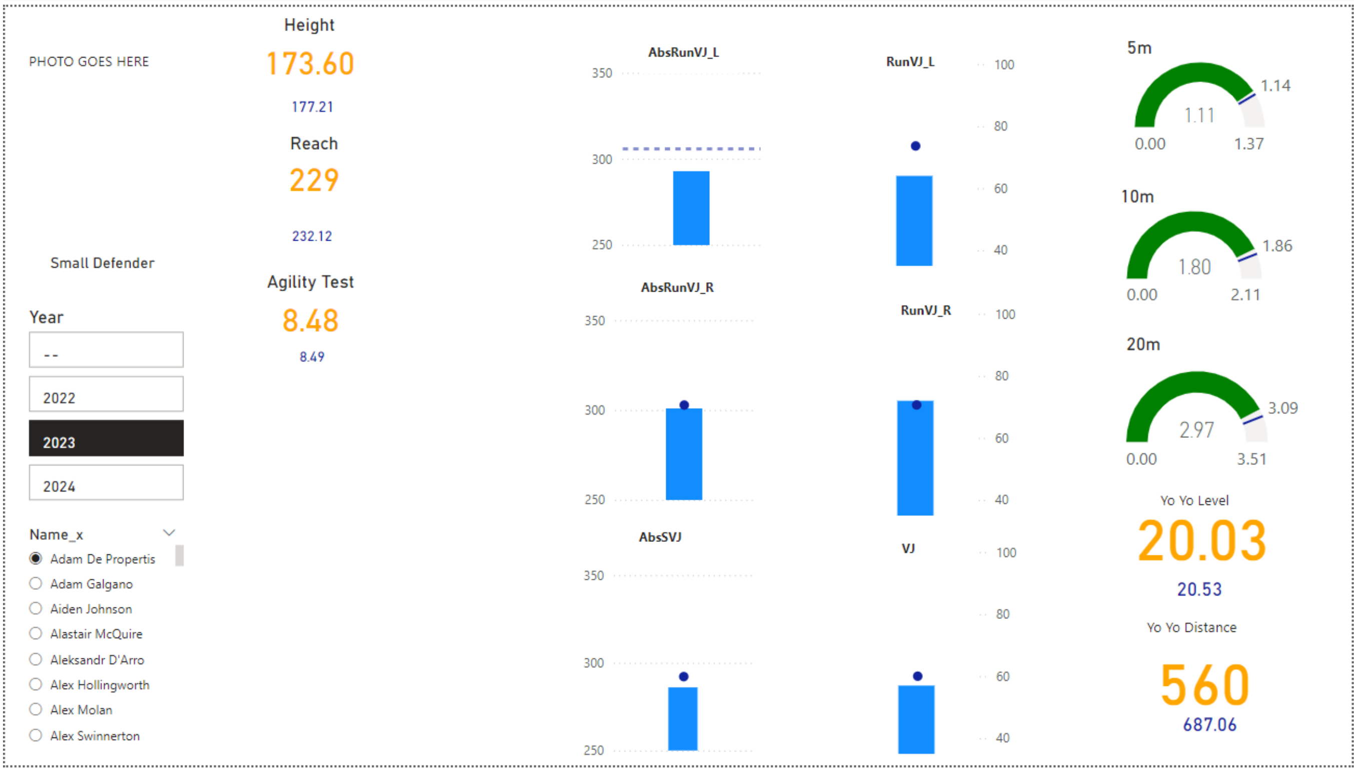


Figure 30. Player Performance from the Combine Data during the 2022-2024 Season

Player	Round	D	K	H	K:H	KE%	DE%	R50	CP	UP	CP%	M	IM	IP	SL	SI	SP	SP%	T	CM	LM	F50M	F50BG	FHCL	G	B	ACC%
Archie Roberts	1	18.00	11.00	7.00	1.57	81.80	88.90	2.00	4.00	11.00	26.70	4.00	1.00	3.00	3.00	3.00	0.00	0.00	3.00								
Billy Mcgee	1	14.00	8.00	6.00	1.33	87.50	64.30	3.00	1.00	11.00	8.30	3.00	1.00	3.00	2.00	4.00	2.00	50.00	0.00								
Calsher Dear	1	9.00	5.00	4.00		20.00			5.00	4.00		2.00				6.00			1.00	0.00	0.00	2.00	4.00	0.00	1.00	0.00	33.30
Cooper Lord	1	10.00	8.00	2.00	4.00	87.50	90.00		3.00	7.00	30.00	5.00		1.00		3.00			9.00						0.00		0.00
Harvey Johnston	1	19.00	9.00	10.00	0.90	77.80	79.00		8.00	10.00	44.40	3.00		0.00		4.00			2.00						2.00		100.00
Lachlan Voss	1	4.00	3.00	1.00	3.00	33.30	25.00	2.00	2.00	1.00	66.70	1.00	0.00	3.00	0.00	1.00	1.00	100.00	2.00								
Levi Ashcroft	1	24.00	13.00	11.00	1.18	84.60	83.30		8.00	16.00	33.30	5.00		1.00		9.00			5.00						2.00		66.70
Luke Trainor	1	10.00	7.00	3.00	2.33	57.10	60.00	0.00	3.00	7.00	30.00	2.00	0.00	0.00	0.00	1.00	0.00	0.00	3.00								
Matt Carrol	1	12.00	9.00	3.00	3.00	77.80	75.00	5.00	2.00	8.00	20.00	3.00	2.00	4.00	0.00	1.00	4.00	0.00	1.00								
Miles Enders	1	10.00	8.00	2.00		37.50			2.00	7.00		5.00				6.00			0.00	0.00	3.00	3.00	2.00	0.00	2.00	3.00	40.00
Murphy Reid	1	23.00	7.00	16.00	0.44	85.70	69.60		14.00	10.00	58.30	1.00		7.00		5.00			3.00						0.00		0.00
Ned Maginnes	1	18.00	10.00	8.00	1.25	30.00	50.00		4.00	14.00	22.20	4.00		2.00		4.00			3.00						0.00		0.00
Ollie Murphy	1	6.00	3.00	3.00	1.00	66.70	66.70	0.00	0.00	6.00	0.00	3.00	1.00	1.00	0.00	2.00	4.00	75.00	0.00								
Ryley Sanders	1	28.00	5.00	23.00	0.22	100.00	85.70		12.00	16.00	42.90	3.00		2.00		5.00			4.00						1.00		100.00
Samuel Marshall	1	30.00	15.00	15.00	1.00	73.30	76.70		12.00	19.00	38.70	7.00		4.00		4.00			5.00						0.00		0.00
Taj Hotton	1	19.00	7.00	12.00		42.90			11.00	8.00		4.00				8.00			3.00	1.00	0.00	0.00	5.00	1.00	1.00	0.00	100.00
Tarkyn O'leary	1	16.00	6.00	10.00	0.60	100.00	93.80		5.00	10.00	33.30	0.00		3.00		4.00			1.00						0.00		0.00

Player

Adrian Cole

Archie May

Player	Round	D	K	H	K:H	KE%	DE%	R50	CP	UP	CP%	M	IM	IP	SL	SI	SP	SP%	T	CM	LM	F50M	F50BG	FHCL	G	B	ACC%	SA	F50T	I50	GBG	C	CC
Adrian Cole	TOTAL:	13	9	4	2.25	66.70	69.20	7	5	8	38.50	2	1	8	0	1	16	75.00	3														
Archie May	TOTAL:	122	95	27		52.60			76	46		44				65			23	22	9	30	22	5	29	15	56.90	8	9				
Archie Roberts	TOTAL:	188	110	78	1.41	80.00	83.00	36	48	112	30.00	33	9	39	14	31	4	75.00	18														
Billy Mcgee	TOTAL:	216	131	85	1.54	73.30	73.20	40	62	134	31.60	57	23	66	19	45	8	37.50	25														
Calsher Dear	TOTAL:	175	105	70		42.90			107	69		45				95			32	16	6	29	48	10	21	24	36.80	17	20				
Charlie Edwards	TOTAL:	205	112	93	1.20	65.20	75.60		74	132	35.90	40		33		61			34						7		43.80	4		38	62	26	9
Cody Goon	TOTAL:	73	43	30		55.80			23	47		17				20			13	0	0	2	6	1	4	2	40.00	2	5				
Cooper Lord	TOTAL:	249	143	106	1.35	56.60	64.30		106	150	41.40	40		31		59			65						1		12.50	10		38	75	43	17
Dexter Prime	TOTAL:	13	8	5	1.60	87.50	92.30	2	5	8	38.50	5	3	7	0	0	1	100.00	0														

Figure 31. Match Performance from the Combine Data during the 2022-2024 Season

Limitations

During the "Exploratory Analysis and Modelling" phase there are some limitations when it comes to any gaps with data accuracy. For instance, the GPS Data may have limitations due to weather conditions or equipment, whether there are major concerns for players 'privacy and consent' hence it will be hard to adjust or justify the reason there are missing values from the Microsoft Excel Spreadsheets of our Foundry Data Folders **(see Figure 2)**. In addition, we should not account for historical data for the Combined Data since it would not be practical for current changes that can help coaches to strategize the 'reduce the likelihood of injuries. If the lead coach has any assumptions or additional conditions if all the players perform under time pressure on average each day during the season, maintaining consistent results for our metrics and keep track under long hours in various circumstances. Due to the 'Lack of Standardization' in which Australian Rules Football 'have different rules and scoring systems' which can be challenging to compare and analyze 'data across different metrics used in the 'Data Understanding' phase. Other limitations of the data were due to 'human element' since the following dashboards from **(figures shown in 11-37)** were focused on the males Sandringham Dragon Football Club instead investigating both males and females' data from the Combined Data. However, the decision was made to achieve reducing time costs and Labor to further investigate valuable sports data that counts to the lead coach fruition. Factors we need to consider such as 'player motivation', 'injuries' and 'team dynamics' assuming there are no inferences from other competitors can have a 'significant impact' on the outcome of a game and overall data metrics observed when exploring and modelling the data. This could have been further been clarified and adjusted with the lead coach and the data team could overall, can impact our findings and drafting procedures during the final phase of the data analysis project. That the lead coach can be able to use our dashboard to strategize player strengths and weaknesses before finalizing drafting proceedings. Thus, this can reduce and does not re-reinforce human bias and focus to data quality and fairness for every player candidate.

Further work

During the data analysis scope, some data had minimal rows that needed more information regarding player and team performance or injury reductions and metrics. During the "Data Understanding" and "Explanatory Analysis and Modelling," the Team Match Data folder contains files for the Boys ALL Team Match Averages 2023 Data containing only 12 teams. Additional features for the Team Match Folder include CTL Drafted Data Averages, which contain different Player Position Types Metrics that do not fit nor transition well during the data analysis since the data was transferred into the combine prior to the start of the data analysis. Hence, it was not drafted into the dashboards since the data team believed that it does not well represent the overall comparisons for all players from the Sandringham Dragons Football Club and could lead to bias and inconsistent results for the overall visualizations and misinterpretations displayed from our dashboards seen in **Figures 32,33,34,35,36 and 37** to enhance and improve business stakeholder requirements.

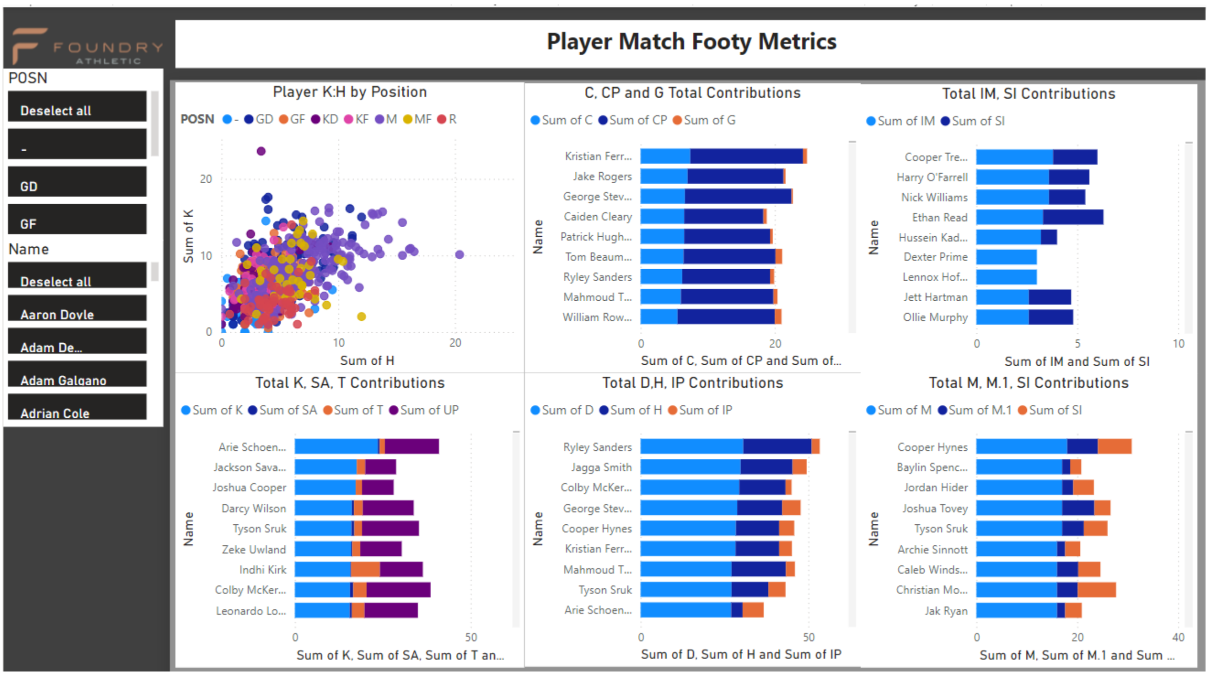


Figure 32. Player Match Footy Metrics

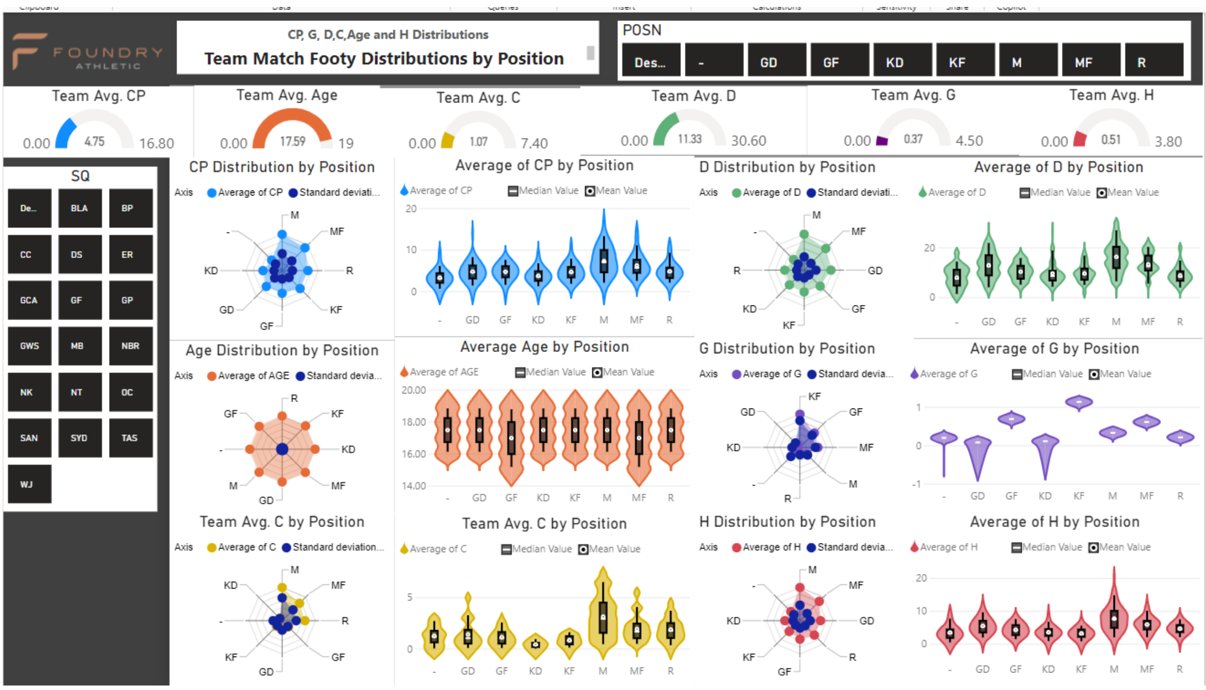


Figure 33. Team Match Footy Distributions by Position

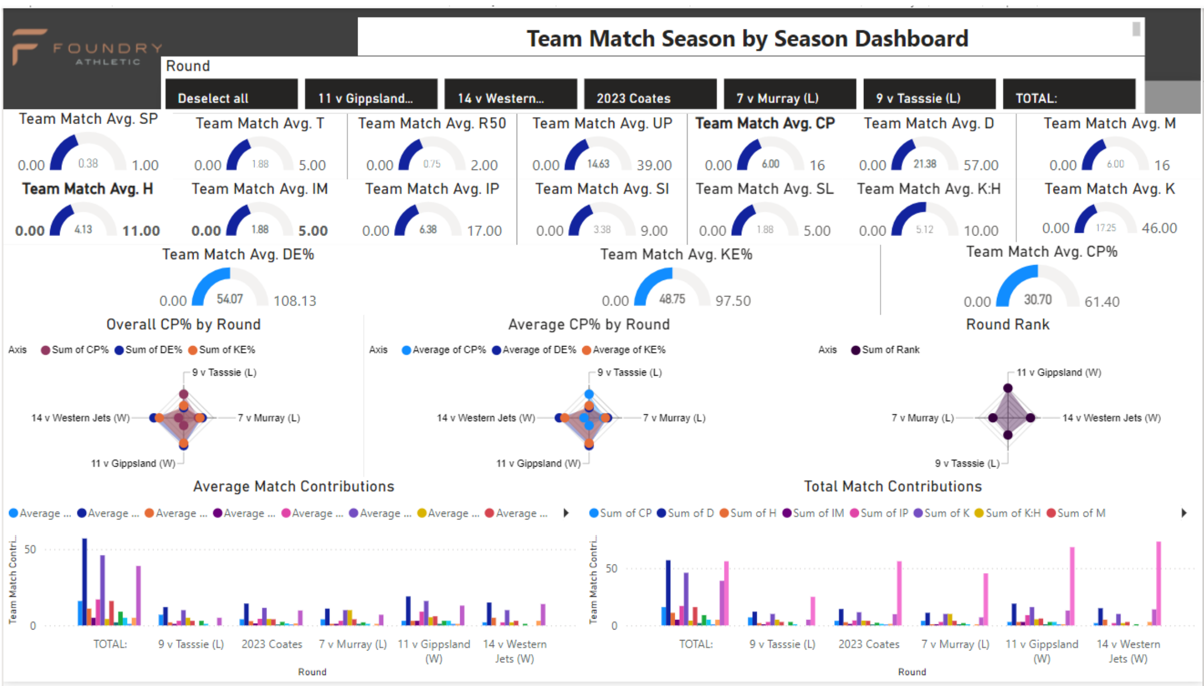


Figure 34. Team Match Season Dashboard

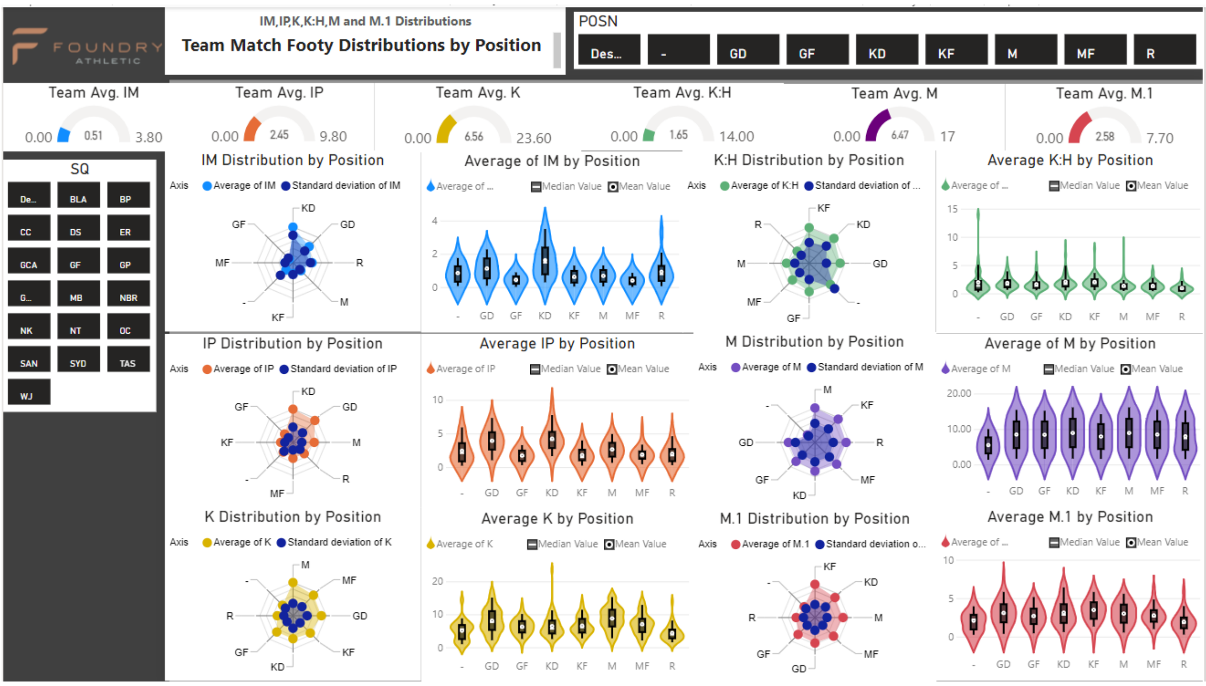


Figure 35. Team Match Footy Distributions by Position Dashboard displaying IM,IP,K:H,M and M.1 metrics

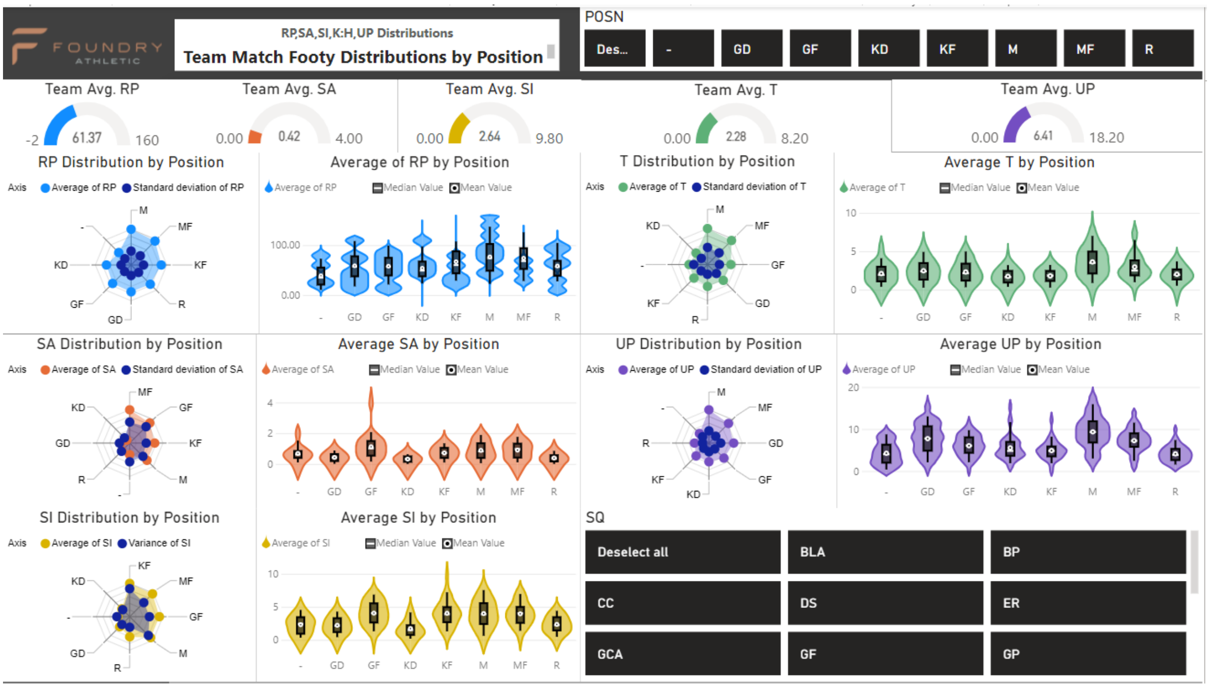


Figure 36. Team Match Footy Distributions by Position Dashboard displaying RP,SA,SI,K:H, and UP Distributions

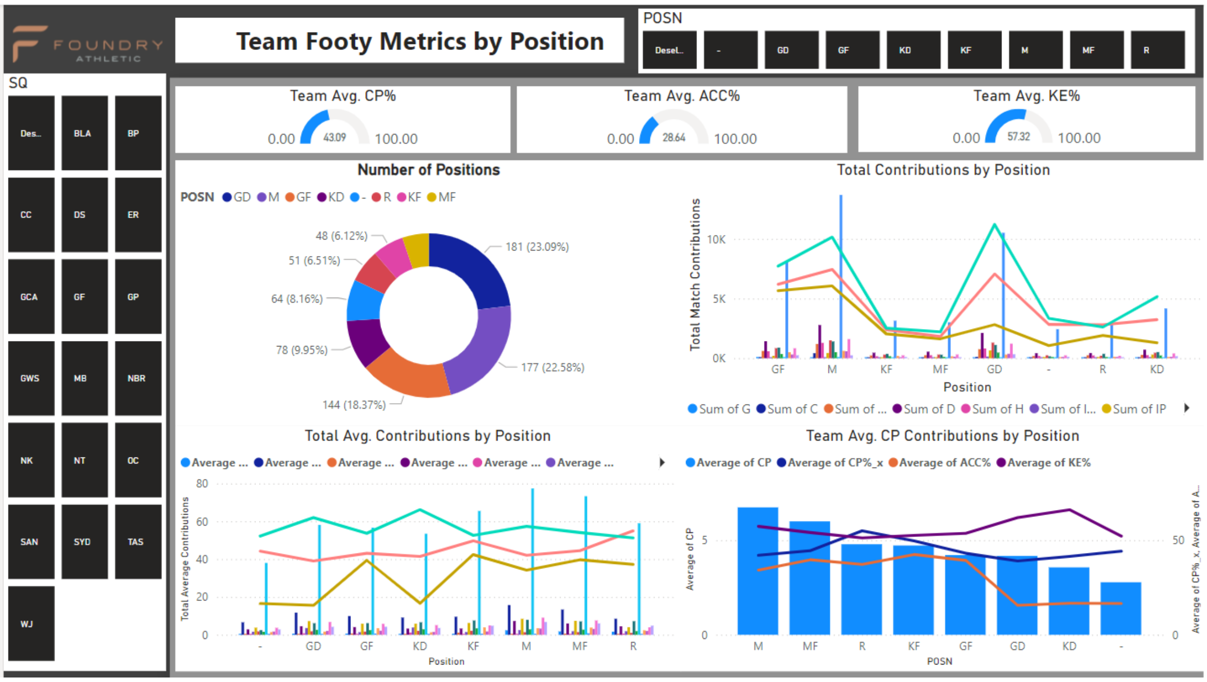


Figure 37. Team Footy Metrics determined by Position Dashboard

In other words, there were only four rounds were shown for the Footy Match Data on round matches given different positions; it would be better to investigate the **combined data** further to measure efficient team comparisons from other competitors. Furthermore, the Team Footy Distributions and Metrics folder itself does not actually represent the overall team. The data team decided it would be more accurate to overlook the combined data than the Team Player Football Statistics on certain players playing in the program during a certain time frame the season took place. Referring to the five characteristics, V's of Big Data would be applicable here to check for any "duplicates, missing data, or high latencies." The Excel spreadsheets of player match data and statistics feature testing data from the NAB League, Coates Talent League Testing Data, and historical data from the 2022 to 2024 Season. Similarly, for the GPS, 2km Results, Height and Weights, and Gym and Strength Data folders and files to reduce potential bias and errors to ensure fairness for every player that has contributed during training peak hours to improve 'Veracity' of our data during 'data preparation' phase. That could lead to misinterpretations of our data findings within our dashboards that have been presented during the final phase of the data with the final phase of the data analysis and Methodologies. To ensure that we wanted to contain information onto a single screen, to easily 'avoid scrolling bars and clustering', 'remove all but essential information'. Additional improvements can be made such as text and fonts are the same size, to highlight information and give enough context. Visualizations such as 'Bar and column charts' are 'better for comparing values' for side comparisons **(see from Figure 9)** formatted in descending order**.** 'Pie Charts' are useful when dealing with categorical data to determine ‘part-to-whole’ relationships other than focusing on features individually. ‘Gauge Charts’ are great for ‘great for displaying the status’ focusing on the ‘main context of a goal’. For the GPS Player Data for instance, during the early phase of the "Exploratory Analysis and Modelling,” Timeframes should be fixed for the GPS Data for Player Performance should be combined for the entire 2023-2024 Season, instead of putting 2023 and 2024 Dashboards separately to reduce redundancy. In addition, the team applied SQL efficiently to create tables to perform specific queries to perform data manipulation and aggregations from existing and accessible data **(refer from Appendix A and B)** to confirm and clarify key player and team metrics. However, the computations and data aggregations can be automated in Power-BI Desktop. In which is a valuable tool for future data analysts and professionals in case we want to secure "storage," "manipulation," and "analysis of large amounts of data" within Foundry's database for recruits of potential players who are eligible for the program. For their audience to figure out how to practically use our dashboards, key metrics and information. For the Sandringham Dragons Football Team to be successful team leading players to be drafted at the Australian Football League standards.

Conclusion

The overall feedback for our Power-BI dashboards covering data during the pre-season draft selection in 2022 and 2023-2024 was exceptional in leading coach standards and criteria of the business objectives including deliverables are met. During the experimenting for the ‘Exploratory analysis and modelling’ phase, different types of data visualizations were applied to identify problems and solve the dashboards. Notable visualizations that were finalized in the final stages of data analysis such as bar charts, scatterplots, and boxplots to understand data distributions, line plots, and gauge and radar charts, when appropriate, to match up to the business objectives. To be satisfied with high expectations with their results in finding key talent and recruitment from player and team performances. The data team prefers Microsoft Power BI over Tableau because we wanted to adapt as much of the familiarity as possible when figuring out and testing most of the capabilities the platform has to offer. To perform more efficient dashboards, we initially started with two dashboards, at least to showcase the player key statistics and team performance statistics. Available "in Power BI Desktop" since we can build and view reports. Giving the opportunity to able to "filter data" to showcase data manipulation and aggregations more effectively and efficiently. Where we can be able to code in different programming languages such as Python or R to test our programming level efficiencies, which can be integrated into Power-BI to enable programmers to cross-platform their technical expertise to update real-time data onto the dashboards. Until it is adequately optimized for stakeholders to easily maintain track requires additional material self-taught from courses and tutorials from Coursera, Codecademy, Data-Camp and Free Code Camp and certifications to further adapt real-world data. The purpose of developing a sports data analysis project was to allow more people to have more freedom in understanding and handling real-world data that is well-crated and interpreted by football coaches who are experts in their sports domain. To conclude, the data team can best tell stories to the clients about our data findings and how we can implement further strategies using player and team performances to improve better in future matches in sports analytics. To provide 'Value' 'without the proper processing, validation, and analytics frameworks' taken in place to achieve this. Thus, 'unlocking potential and taking' Australian Rules Football to the next level. Giving opportunities to allow recruitment coaches and professional coaches from lead Australian Football Clubs to find potential talent more effectively when drafting players into the Australian Football League that is well represented for the Sandringham Football Dragons Club.

References

- AFL. (n.d.). *History of football - AFL.com.au*. afl.com.au. https://www.afl.com.au/about-afl/history
- Collins, B. (2019, November 24). *How do I get drafted by an AFL club? Am I eligible?* afl.com.au. https://www.afl.com.au/news/149356/how-do-i-get-drafted-by-an-afl-club-am-i-eligible
- McLean, J. (n.d.). *The Science of Success: How AFL Teams Scout and Develop Emerging Talent*. https://preparelikeapro.com/the-science-of-success-how-afl-teams-scout-and-develop-emerging-talent

- Ambler, W. (2024, July 3). *Sports analytics: What is it & how it improves performance?* Catapult. https://www.catapult.com/blog/what-is-sports-analytics
- Avcontentteam. (2024, August 2). *All you need to know about sport analytics in 2023*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2023/08/sport-analytics/
- Bai, X., & Bai, Z. (2021, January 30). *Sports Big Data: Management, Analysis, Applications, and Challenges*. Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1155/2021/6676297
- *Big data characteristics: Recognize the 5 V's of big data*. (n.d.). twilio segment. https://segment.com/data-hub/big-data/characteristics/
- *Data analysis methods: 7 key methods you should know!* (2023, December 14). Atlan | Third-Gen Data Catalog. https://atlan.com/data-analysis-methods/
- FX Team. (2024, September 10). *Excel vs SQL data storage: Pros and cons*. eSoftware Associates Inc. https://www.eswcompany.com/micrsoft-excel-vs-azure-sql
- Metomic. (2024, September 23). Data leaks: The biggest risks, consequences, causes & how to prevent them. *Metomic | Data Security Software for SaaS, GenAI and Cloud*. https://www.metomic.io/resource-centre/what-are-the-biggest-risks-of-data-leaks
- Munir, T. (2023, September 25). *Tariq Munir - PepsiCo | LinkedIn*. LinkedIn. https://www.linkedin.com/in/tariq-munir/
- Pykes, K. (2022, November 24). *Sports Analytics: How Different Sports Use Data Analytics*. Learn Data Science and AI Online | DataCamp. https://www.datacamp.com/blog/sports-analytics-how-different-sports-use-data-analysis

- Sprinkle. (2023, October 25). *Tableau vs power BI: 15 key differences and comparison | Sprinkle data*. sprinkle | The most advanced platform for data pipelines and analytics. https://www.sprinkledata.com/blogs/tableau-vs-powerbi-15-differences
- Microsoft Ignite. (2023, June 30). *Tips for designing a great power BI dashboard*. Microsoft Learn: Build skills that open doors in your career. https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboards-design-tips
- Data Overload. (2023, March 27). *Challenges and limitations of sports analytics: What we still don't know*. Medium. https://medium.com/@data-overload/challenges-and-limitations-of-sports-analytics-what-we-still-dont-know-d6e1d34a445a

Appendix:

Appendix A: SQL queries from GPS Data 2023 Season Sandringham Dragons Football Club

```
SELECT [Name],

AVG([TOG%])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

SUM([Meterage Per Minute])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

AVG([Total Distance (m)])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

AVG([Meterage Per Minute])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

AVG([Meterage Per Minute])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

AVG([HSR m/min (4m/s+)])

FROM [DataTable$]

GROUP BY [Name];
```

```
SELECT [Name],
AVG([Sprint Distance(6m/s+) (m)])
FROM [DataTable$]
GROUP BY [Name];

SELECT [Name],
AVG([Maximum Velocity (m/s)])
FROM [DataTable$]
GROUP BY [Name];

SELECT [Name],
AVG([Total Player Load])
FROM [DataTable$]
GROUP BY [Name];

SELECT [Name],
AVG([Total Acceleration Load])
FROM [DataTable$]
GROUP BY [Name];

SELECT [Name],
MAX([Total Distance (m)])
FROM [DataTable$]
GROUP BY [Name];

SELECT [Name],
MAX([Sprint Distance (6m/s+) (m)])
FROM [DataTable$]
GROUP BY [Name];

SELECT [Name],
```

```
MAX([Maximum Velocity (m/s)])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

AVG([HSR Distance (4m/s+)])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

MAX([HSR Distance (4m/s+)])

FROM [DataTable$]

GROUP BY [Name];


SELECT [Name],

SUM([HSR Distance (4m/s+)])

FROM [DataTable$]

GROUP BY [Name];
```

Appendix B: SQL queries from GPS Data 2024 Season Sandringham Dragons Football Club, like Appendix A, but table was excluded during development

```
SELECT [Name],

AVG([TOG%])

FROM [Transformed_C_DataTable$]

GROUP BY [Name];


SELECT [Name],

SUM([Meterage Per Minute])

FROM [Transformed_C_DataTable$]]

GROUP BY [Name];
```

```sql
SELECT [Name],
AVG([Total Distance (m)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
AVG([Meterage Per Minute])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
AVG([Meterage Per Minute])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
AVG([HSR m/min (4m/s+)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
AVG([Sprint Distance(6m/s+) (m)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
AVG([Maximum Velocity (m/s)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];
```

```
SELECT [Name],
AVG([Total Player Load])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
AVG([Total Acceleration Load])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
MAX([Total Distance (m)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
MAX([Sprint Distance (6m/s+) (m)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
MAX([Maximum Velocity (m/s)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];

SELECT [Name],
AVG([HSR Distance (4m/s+)])
FROM [Transformed_C_DataTable$]]
GROUP BY [Name];
```

SELECT [Name],

MAX([HSR Distance (4m/s+)])

FROM [Transformed_C_DataTable$]]

GROUP BY [Name];

SELECT [Name],

SUM([HSR Distance (4m/s+)])

FROM [Transformed_C_DataTable$]]

GROUP BY [Name];

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Name	Average T(	Sum of Me	Average T(	Average M	Average H	Average Sp	Average Sp	Average M	Average T(	Average T(	Max Total	Maximum	Max Maxir	Average H	Max HSR D	Sum HSR D	Vmax (m/:	Distance (
2	Adrian Co	98	106	10867	106	24	397	4	7	974	2416	10867	397	7	2538	2538	2537.6	0	0
3	Archer Ma	83	906.7	8422	101	15	210	3	8	697	1893	9738	340	9	1512	2008	13608.6	0	0
4	Archie Rol	83	948	9912	119	31	659	8	8	883	2236	10467	781	9	3079	3417	24631.7	0	0
5	Billy McGe	79	1203.6	8682	109	22	401	5	8	890	2305	10284	794	9	2225	3831	24476.7	8.77	38.69
6	Calsher D	79	1175.3	7779	98	18	238	3	8	715	2169	8627	413	8	1778	2674	21337.5	0	0
7	Cameron	79	127.9	10121	128	35	429	5	9	943	2234	10121	429	9	3549	3549	3548.6	0	0
8	Charlie Br	81	115.8	9808	116	23	254	3	8	1012	2270	9808	254	8	2387	2387	2387.4	0	0
9	Charlie Ed	82	1130.9	9332	113	22	375	4	8	915	2275	11627	766	9	2172	3672	21717.3	8.86	39.6
10	Charlie Ee	75	100.6	7517	101	17	336	5	8	815	2097	7517	336	8	1716	1716	1715.6	8.64	46.61
11	Charlton H	76	697.9	8887	116	22	407	5	8	910	2164	10153	603	9	2235	2921	13409.4	9.12	58.25
12	Cody Goor	78	98.5	8030	98	16	88	1	7	804	1948	8030	88	7	1694	1694	1693.8	0	0
13	Cooper Lo	78	1024.4	9994	128	27	193	2	7	1213	2533	11398	323	8	2733	3320	21861.6	8.83	35.43
14	Darcy Gilb	85	106.3	9063	106	22	256	3	7	828	2057	9063	256	7	2161	2161	2161.4	0	0
15	Ethan Will	79	1171.2	9346	117	25	288	4	7	954	2145	11520	554	8	2516	4024	25163.9	8.63	46.96
16	Harrison (	81	330	9088	110	23	316	4	8	977	2427	10185	470	8	2366	3310	7098.9	8.6	31.22
17	Harry Sulli	66	125.8	8299	126	25	354	5	8	708	1864	8299	354	8	2497	2497	2496.8	9.17	44.35
18	Harvey Joh	79	1322.7	9124	120	26	338	4	8	997	2337	10741	569	8	2558	3701	28134.4	8.49	38.69
19	Jack Daltc	64	125.9	8047	126	30	778	12	8	700	1783	8047	778	8	3039	3039	3038.9	8.54	24.82
20	Jake John	70	113.5	8233	113	15	137	2	7	724	1800	8233	137	7	1537	1537	1536.9	0	0
21	Jake Nich	68	102.6	7058	103	15	173	3	8	529	1716	7058	173	8	1496	1496	1495.6	0	0
22	Jeremy a'E	80	309.9	8273	103	23	496	6	8	907	2119	8424	673	8	2267	2876	6802.2	8.61	29.08
23	Jhett Haea	73	236.9	8805	118	23	470	6	9	789	1955	8848	557	9	2368	2811	4736.1	0	0
24	Joshua Dc	74	850.6	7921	106	21	461	6	8	773	2120	8468	686	9	2143	2620	17148	9.19	37.94
25	Joshua Dc	74	796.3	9846	133	35	544	7	8	921	1981	10454	634	9	3579	3886	21472.5	8.97	42.8
26	Justin Kro	79	340.5	8949	113	21	243	3	8	891	2103	9207	321	8	2085	2317	6256.3	8.63	27.83
27	Lachlan V	83	858.4	8765	107	20	334	4	8	716	2131	10837	520	9	1998	2818	15986.1	8.59	41.12
28	Lennox Hc	77	104.5	8433	105	21	374	5	8	715	1913	8433	374	8	2209	2209	2209.4	0	0
29	Levi Ashcr	82	1383.6	10386	126	28	252	3	8	1040	2265	11044	385	8	2807	3969	30879.1	9.16	47.32
30	Levi Young	72	428.5	7462	107	14	82	1	7	749	1612	9199	156	8	1395	2406	5578.4	8.16	37.65
31	Luke Kenn	83	129.2	10826	129	31	144	2	7	1018	2254	10826	144	7	3172	3172	3172.4	9.03	37.1
32	Luke Lloyd	81	1237.7	9172	113	24	371	5	8	885	2322	11460	452	9	2408	3447	26492.8	0	0
33	Luke Trair	80	931.4	9180	116	27	579	7	8	842	2066	11110	773	9	2727	3370	21813.9	0	0
34	Mason Sz	81	1138	9235	114	25	421	5	8	846	2146	10367	598	9	2498	2975	24980	9.28	44.42

Figure 37. Final Table after computing the following queries from Appendix A, but it was concatenated later in Microsoft Excel

Appendix C: Method on Transforming 2km Results Data:

```
import pandas as pd

import numpy as np

# transformed_C_NEW_NEED_TO_ADD-U16_2km_Results_Dec_2022

df = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_NEW_NEED_TO_ADD-U16_2km_Results_Dec_2022\C_NEW_NEED_TO_ADD-U16_2km_Results_Dec_2022.xlsx')
```

```
#transformed_IP_Boys_2km_TT_Results_Nov_2023&MAS

#df2 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_Boys_2km_TT_Results_Nov_2023&MAS\BOYS2kmALLResults_Transformation_IP_Boys 2km_TT_Results_Nov_2023&MAS.xlsx')

#df3 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_Boys_2km_TT_Results_Nov_2023&MAS\Individual_Results_Transformed_IP_Boys_2km_TT_Results_Nov 2023&MAS.xlsx')

#df4 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_Boys_2km_TT_Results_Nov_2023&MAS\MAS_SESSION_CALCULATOR_Transformed_Boys_2km_TT_Results_Nov_2023&MAS.xlsx')

df2 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_Boys_2km_TT_Results_Nov_2023&MAS\merged folder\C_transformed_Boys_2km_TT_Results_Nov_2023&MAS.xlsx')



#transformed_IP_SDFC2km - RESULTS BOYS 2022

#df5 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_SDFC2km - RESULTS BOYS 2022\C_Transformed_SDFC2km - 2km TT Rankings _ver2.xlsx')

#df6 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_SDFC2km - RESULTS BOYS 2022\C_Transformed_SDFC2km -U16s 2km TT Rankings.xlsx')

#df7 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_SDFC2km - RESULTS BOYS 2022\C_Transformed_SDFC2km_2km Results 14Nov2022.xlsx')

df3 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed_C_SDFC2km - RESULTS BOYS 2022\merged\C_transformed_SDFC2km - RESULTS BOYS 2022.xlsx')

#transformed-IP-U16FUTUREBoys 2km TT Results Jan 2024

#df8 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed-C-U16FUTUREBoys 2km TT Results Jan 2024\C_Transformed_Boys2kmALLResults.xlsx')

#df9 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed-C-U16FUTUREBoys 2km TT Results Jan 2024\C_Transformed_MAS SESSION CALCULATOR.xlsx')

df4 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed-C-U16FUTUREBoys 2km TT Results Jan 2024\merged\C_transformed-U16FUTUREBoys 2km TT Results Jan 2024.xlsx')


#Transformed_IP_U16 Boys Futures 2km Results Dec 2023
```

```
df5 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\Transformed_C_U16 Boys Futures 2km Results Dec 2023.xlsx')


#transformed-IP-U16 Boys 2km TT Results - combined IGNORE THIS

#df11 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\transformed-IP-U16 Boys 2km TT Results - combined.xlsx')

merged_df  = df.merge(df2, how='outer', on='Name')

merged_df_2 = df3.merge(merged_df, how='outer', on='Name')

merged_df_3 = df4.merge(merged_df_2, how='outer', on='Name')

final_df = df5.merge(merged_df_3, how='outer', on='Name')


final_df.isnull().sum()

final_df = merged_df.drop_duplicates()

final_df.to_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\2km Results\Transformed_IP_2km_Results.xlsx', index=False, header=True)
```

Appendix D: Method on Transforming Gym and Strength Data:

```
import pandas as pd

import numpy as np

df_1= pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\RAW_Datasheets\COMBINED--U18-Chin_Ups_Bench-Press_(DONE)\U18 Boys Max Chin Ups and 3RM Bench Press - Dec23\U18MAXPULLUPTEST.csv')

df2 = pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\RAW_Datasheets\COMBINED--U18-Chin_Ups_Bench-Press_(DONE)\U18 Boys Max Chin Ups and 3RM Bench Press - Dec23\U183RMBENCHPRESS.csv')

df3 = pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\RAW_Datasheets\COMBINED--U18-Chin_Ups_Bench-Press_(DONE)\CLEANED---U18 Boys Max Chin Ups and 3RM Bench Press - Dec23 (1)\U18MAXPULLUPTEST_NR.csv')

df4 = pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\RAW_Datasheets\COMBINED--U18-Chin_Ups_Bench-Press_(DONE)\CLEANED---U18 Boys Max Chin Ups and 3RM Bench Press - Dec23 (1)\U183RMBENCHPRESS_NR.csv')

df5 = pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\RAW_Datasheets\COMBINED--U18-Chin_Ups_Bench-Press_(DONE)\CLEANED---U18 Boys Max Chin Ups and 3RM Bench Press - Dec23 (1)\U18MAXPULLUPTEST_BR.csv')
```

```
df6 = pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\RAW_Datasheets\COMBINED--U18-Chin_Ups_Bench-Press_(DONE)\CLEANED---U18 Boys Max Chin Ups and 3RM Bench Press - Dec23 (1)\U183RMBENCHPRESS_BR.csv')


merged_df  = df_1.merge(df2, how='outer', on='Name')

merged_df_2 = df3.merge(merged_df, how='outer',on='Name',suffixes=(' ', '_DROP')).filter(regex='^(?!.*_DROP)')

merged_df_3 = df4.merge(merged_df_2, how='outer',on='Name',suffixes=(' ', '_DROP')).filter(regex='^(?!.*_DROP)')

merged_df_4 = df5.merge(merged_df_3, how='outer',on='Name',suffixes=(' ', '_DROP')).filter(regex='^(?!.*_DROP)')

df = df6.merge(merged_df_4,how='outer',on='Name',suffixes=(' ', '_DROP')).filter(regex='^(?!.*_DROP)')

df.to_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\Transformed_SpreedSheets\Transformed_U18_Boys_Max_Chin_Ups_and_3RM_Bench_Press_Dec23.csv', index=False, header=True)

df7= pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\Transformed_SpreedSheets\FINAL_Gym_Strength_Testing_Data_Spreadsheet\Final_C_Transformed_Gym_Strength_Data_2023_2024.csv')

df8 = pd.read_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\Transformed_SpreedSheets\Transformed_COMBINED--U18-Chin_Ups_Bench-Press.csv')

df_final  = df7.merge(df8, how='outer', on='Name')

df_final = df_final.drop_duplicates()

df_final.to_csv(r'C:\Users\Michael Le\Desktop\Foundry_Data\Gym_Strength_Testing_Data\Transformed_SpreedSheets\FINAL_Gym_Strength_Testing_Data_Spreadsheet\Final_Transformed_Gym_Strength_Data_2023_2024_ver_2.csv', index=False, header=True)
```

Appendix E: Method on Heights and Weights Data:

```
import pandas as pd

import numpy as np

# transformed_C_NEW_NEED_TO_ADD-U16_2km_Results_Dec_2022

df = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_C_2023 - U16 Boys Player Heights & Weights\Transformed_C_2023_U16_Boys_Player_Heights_Weights_FINAL.xlsx')

df2 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_C_CLEANED---U16 & 18Boys Heights
```

```
and Weights Pre-Season 2023\Transformed_C_CLEANED---U16 & 18Boys Heights and Weights Pre-Season 2023.xlsx')

df3 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_IP_PlayerHeightsandWeightsMaster2023-2024\DF3.xlsx')

df4 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_IP_PlayerHeightsandWeightsMaster2023-2024\DF4.xlsx')

df5 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_IP_PlayerHeightsandWeightsMaster2023-2024\DF5.xlsx')

df6 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_IP_PlayerHeightsandWeightsMaster2023-2024\DF6.xlsx')

df7 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_IP_PlayerHeightsandWeightsMaster2023-2024\DF7.xlsx')

df8 = pd.read_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\Transformed_IP_PlayerHeightsandWeightsMaster2023-2024\DF8.xlsx')

merged_df  = df.merge(df2, how='outer', on='Name')

merged_df_2 = df3.merge(merged_df, how='outer', on='Name')

merged_df_3 = df4.merge(merged_df_2, how='outer', on='Name')

merged_df_4 = df5.merge(merged_df_3, how='outer', on='Name')

merged_df_5 = df6.merge(merged_df_4, how='outer', on='Name')

merged_df_6 = df7.merge(merged_df_5, how='outer', on='Name')

df_final = df8.merge(merged_df_6, how='outer', on='Name')

df_final = df_final.drop_duplicates()

df_final.to_excel(r'C:\Users\Michael Le\Desktop\Foundry_Data\Height_&_Weight_Data\FINAL_Transformed\Heights_Weights_final.xlsx', index=False, header=True)
```

Appendix F: Computations of Z-Scores on Gym Strength and Testing Data in Power-BI:

```
z_score_Concentric = (MAX(Final_Transformed_Gym_Strength[Concentric Peak Force [N]] ]) - AVERAGE(Final_Transformed_Gym_Strength[Concentric Peak Force [N]] ]))/STDEV.P(Final_Transformed_Gym_Strength[Concentric Peak Force [N]] ])
```

z_score_contact_time = (MAX(Final_Transformed_Gym_Strength[Contact Time [s]]])-AVERAGE(Final_Transformed_Gym_Strength[Contact Time [s]]]))/STDEV.P(Final_Transformed_Gym_Strength[Contact Time [s]]])

z_score_DSI = (MAX(Final_Transformed_Gym_Strength[DSI (CMJ Peak Force / IMPT Peak Force)]) - AVERAGE(Final_Transformed_Gym_Strength[DSI (CMJ Peak Force / IMPT Peak Force)]))/STDEV.P(Final_Transformed_Gym_Strength[DSI (CMJ Peak Force / IMPT Peak Force)])

Z_Score_Jump Height Flight Time = (MAX(Final_Transformed_Gym_Strength[Jump Height (Flight Time) [cm]]]) - AVERAGE(Final_Transformed_Gym_Strength[Jump Height (Flight Time) [cm]]])/STDEV.P(Final_Transformed_Gym_Strength[Jump Height (Flight Time) [cm]]]))

z_score_Jump_Height_Flight_Times = (MAX(Final_Transformed_Gym_Strength[Jump Height (Flight Time) [cm]]])-AVERAGE(Final_Transformed_Gym_Strength[Jump Height (Flight Time) [cm]]]))/STDEV.P(Final_Transformed_Gym_Strength[Jump Height (Flight Time) [cm]]])

z_score_Jump_Height_Imp_Mom = (MAX(Final_Transformed_Gym_Strength[Jump Height (Imp-Mom) [cm]] SJ]) - AVERAGE(Final_Transformed_Gym_Strength[Jump Height (Imp-Mom) [cm]] SJ]))/STDEV.P(Final_Transformed_Gym_Strength[Jump Height (Imp-Mom) [cm]] SJ])

z_score_Jump_Height_Imp_Mom_dj = (MAX(Final_Transformed_Gym_Strength[Jump Height (Imp-Mom) [cm]] DJ]) - AVERAGE(Final_Transformed_Gym_Strength[Jump Height (Imp-Mom) [cm]] DJ]))/STDEV.P(Final_Transformed_Gym_Strength[Jump Height (Imp-Mom) [cm]] DJ])

z_score_Left_NORD = (MAX(Final_Transformed_Gym_Strength[NORD LEFT])-AVERAGE(Final_Transformed_Gym_Strength[NORD LEFT]))/STDEV.P(Final_Transformed_Gym_Strength[NORD ASSYM])

z_score_NORD_ASSYM = (MAX(Final_Transformed_Gym_Strength[NORD ASSYM])-AVERAGE(Final_Transformed_Gym_Strength[NORD ASSYM]))/STDEV.P(Final_Transformed_Gym_Strength[NORD ASSYM])

z_score_NORD_RIGHT = (MAX(Final_Transformed_Gym_Strength[NORD RIGHT])-AVERAGE(Final_Transformed_Gym_Strength[NORD RIGHT]))/STDEV.P(Final_Transformed_Gym_Strength[NORD RIGHT])

z_score_peak_force = (MAX(Final_Transformed_Gym_Strength[Peak Vertical Force [N]]])-AVERAGE(Final_Transformed_Gym_Strength[Peak Vertical Force [N]]]))/STDEV.P(Final_Transformed_Gym_Strength[Peak Vertical Force [N/kg]]]) z_score_peak_vertical_force_mass = (MAX(Final_Transformed_Gym_Strength[Peak Vertical Force [N/kg]]]) - AVERAGE(Final_Transformed_Gym_Strength[Peak Vertical Force [N/kg]]]))/STDEV.P(Final_Transformed_Gym_Strength[DSI (CMJ Peak Force / IMPT Peak Force)])

z_score_RSI_Flight_Time/Contact_Time_m_s = (MAX(Final_Transformed_Gym_Strength[RSI (JH (Flight Time)/Contact Time) [m/s]]]) - AVERAGE(Final_Transformed_Gym_Strength[RSI (JH (Flight Time)/Contact Time) [m/s]]]))/STDEV.P(Final_Transformed_Gym_Strength[RSI (JH (Flight Time)/Contact Time) [m/s]]])

END OF FINAL REPORT